

Département des affaires économiques et sociales
Division de la statistique

Études méthodologiques

Serie F, n° 103

Manuel sur l'infrastructure géospatiale à l'appui des activités de recensement



Nations Unies
New York, 2010

Département des affaires économiques et sociales

Le Département des affaires économiques et sociales du Secrétariat de l'Organisation des Nations Unies sert de relais entre les orientations arrêtées au niveau international dans les domaines économiques, sociaux et environnementaux et les politiques exécutées à l'échelon national. Il intervient dans trois grands domaines liés les uns aux autres : i) il compile, produit et analyse une vaste gamme de données et d'éléments d'information sur des questions économiques, sociales et environnementales dont les États Membres de l'Organisation se servent pour examiner des problèmes communs et évaluer les options qui s'offrent à eux; ii) il facilite les négociations entre les États Membres dans de nombreux organes intergouvernementaux sur les orientations à suivre de façon collective afin de faire face aux problèmes mondiaux existants ou en voie d'apparition; iii) il conseille les gouvernements intéressés sur la façon de transposer les orientations politiques arrêtées à l'occasion des conférences et sommets des Nations Unies en programmes exécutables au niveau national et aide à renforcer les capacités nationales au moyen de programmes d'assistance technique.

Note

Les cotes des documents de l'Organisation des Nations Unies se composent de lettres majuscules et de chiffres. La simple mention d'une cote dans un texte signifie qu'il s'agit d'un document de l'Organisation

ST/ESA/STAT/SER.F/103

ISBN 978-92-1-261221-8

Publication des Nations Unies

Numéro de vente : F.09.XVII.8

Copyright © Nations Unies, 2009

Tous droits réservés

Préface

Les Nations Unies ont publié le *Manuel des systèmes d'information géographique et de cartographie numérique* aux fins des recensements de la population et de l'habitation de 2000. Ce *Manuel* s'est avéré être une source de directives utiles dans le domaine de la cartographie de recensement. Il convient toutefois de le mettre à jour et de le réviser à la lumière des récents développements dans les technologies géospatiales et leurs applications dans le cadre des opérations statistiques, en particulier pour les recensements de la population et de l'habitation.

S'agissant du Programme mondial de recensements de la population et des logements de 2010, la Commission de statistique a demandé à l'occasion de sa trente-sixième session que la Division de statistique des Nations Unies poursuive la révision et la mise à jour des Principes et recommandations concernant les recensements de la population et de l'habitat. La Commission a invité ladite Division à se pencher sur des points spécifiques connexes, y compris l'utilisation des systèmes d'information géographique (SIG) aux fins de la collecte et de la diffusion des données. Ces développements transparaissent dans les *Principes et recommandations concernant les recensements de la population et des logements, deuxième révision*, adoptés par la Commission de statistique en mars 2007.

Après l'adoption des *Principes et recommandations, deuxième révision*, la Division de statistique des Nations Unies a engagé une série d'activités visant à promouvoir et faciliter le Programme mondial de recensements de la population et des logements de 2010, ainsi qu'à réviser le *Manuel* de 2000, compte tenu des récentes avancées dans le domaine des technologies géospatiales. En mai 2007, la Division de statistique a organisé une réunion d'experts à New York sur les pratiques contemporaines en matière de cartographie de recensement et l'utilisation de systèmes d'information géographique. Elle entendait contribuer à la révision du *Manuel* de 2000. Au cours du quatrième trimestre 2007 et début 2008, la Division statistique a par ailleurs mis sur pied cinq ateliers sur les technologies géospatiales au service des opérations de cartographie de recensement. Deux ateliers se sont déroulés en Afrique, respectivement pour les pays anglophones et francophones; un autre a été organisé pour les régions de l'Asie, des Caraïbes et des îles du Pacifique. La Division de statistique a engagé un consultant, David Rain, pour faciliter la préparation d'une version révisée du *Manuel*. En avril 2008, une deuxième réunion d'experts s'est tenue à New York afin d'étudier le projet de *Manuel* révisé.

Le nouveau *Manuel sur l'infrastructure géospatiale à l'appui des activités de recensement*, révisé et rebaptisé, tient compte des recommandations des réunions du groupe d'expert et des ateliers régionaux sur les SIG et la cartographie de recensement. Ces groupes ont insisté sur le fait que ces pays devaient considérer le recours aux programmes de géographie du recensement comme un processus continu et non comme une simple suite d'opérations de cartographie et de diffusion. Ils ont en outre recommandé que le *Manuel* montre en quoi l'utilisation et l'application de technologies géospatiales contemporaines ainsi que de bases de données géographiques sont bénéfiques pour

chacune des étapes du recensement de la population et des logements. Le *Manuel* devrait par exemple indiquer de quelle manière ces technologies permettent une préparation, un dénombrement, un traitement et une diffusion plus efficaces du recensement.

À et égard, il est primordial qu'à travers le *Manuel*, les planificateurs du recensement et le personnel connexe disposent d'un guide technique sur les méthodes, outils et meilleures pratiques contemporaines leur permettant de mieux structurer leurs besoins et de gérer plus efficacement les opérations de cartographie de recensement. En bref, le *Manuel* traite de manière exhaustive les besoins opérationnels et de gestion. Il répond à des questions d'ordre organisationnel et institutionnel préoccupant les responsables d'agences statistiques et autres. Il se penche par ailleurs sur des questions techniques et pratiques intéressant les cartographes du recensement et les recenseurs.

Le présent *Manuel* comporte six chapitres et sept annexes. Après une brève introduction au chapitre I, le chapitre II aborde des questions de gestion que les responsables d'agences statistiques doivent considérer lors de la réorganisation de bureaux nationaux de statistique de manière à tirer pleinement parti de l'infrastructure géospatiale. Le chapitre III propose au gestionnaire informatique ou au responsable SIG/cartographie les éléments techniques nécessaires pour constituer une base de données géographique numérique des districts de recensement. Le chapitre IV, lui aussi à caractère technique, détaille le processus d'élaboration d'une base de données géographique des districts de recensement et l'utilisation d'innovations technologiques telles que les systèmes de positionnement universel (GPS) et la télédétection afin d'effectuer des corrections, en les intégrant à des travaux sur le terrain lorsque cela s'avère nécessaire. Le chapitre V traite du processus de création des cartes requises pour le dénombrement, en poursuivant l'analyse du point de vue opérationnel là où la discussion sur les bases de données géographiques s'est arrêtée. Au chapitre VI, il est question de l'exploitation de l'infrastructure géospatiale pour la diffusion des résultats des recensements. Les annexes I à VII constituent un guide pratique pour les responsables de la planification et de la mise en œuvre de solutions géospatiales dans le cadre de projets de recensement.

Au cours de la révision, le Secrétariat des Nations Unies a consulté des cartographes du recensement ainsi que des experts en SIG issus de toutes les régions du monde afin de passer en revue et finaliser le *Manuel*. Ce dernier propose en outre des exemples de pratiques nationales fournis par certains de ces experts sur l'emploi des SIG, du GPS et de la cartographie numérique aux fins des recensements. Le *Manuel* a été préparé par David Rain, consultant pour la Division de statistique des Nations Unies, en collaboration avec le personnel de la Division de statistique.

Table des matières

Préface	iii
<i>Chapitre</i>	
I. Introduction	1
A. Avant-propos et raison d'être du <i>Manuel</i>	1
B. Portée, objet et résumé du <i>Manuel</i>	3
C. Résumés chapitre par chapitre	4
II. Considérations managériales pour les responsables des services nationaux de statistique et autres décideurs	7
A. Introduction	7
B. Le rôle des cartes dans le recensement.	13
C. Des cartes aux bases de données géographiques : la révolution cartographique continue	14
D. Augmentation de la demande en données statistiques ventilées	16
E. Coûts et avantages d'un investissement dans la technologie géospatiale.	18
F. Facteurs déterminant le succès d'une mise en œuvre de technologies géospatiales dans le bureau de statistique national.	20
G. Planification des opérations de recensement à l'aide d'outils géospatiaux	21
H. Évaluation des besoins et définition des solutions géographiques	23
1. Évaluation des besoins des utilisateurs	23
2. Détermination des données de recensement.	23
3. Solutions géographiques	24
4. Ressources humaines et renforcement des capacités	25
I. Coopération institutionnelle : infrastructure nationale de données spatiales : dispositions garantissant la compatibilité avec d'autres organismes publics.	29
J. Normes	34
K. Collaboration	34
L. Résumé et conclusions.	35
III. Développement d'une base de données du niveau des districts de recensement pour le recensement	37
A. Introduction	37
B. Définition du territoire géographique national aux fins du recensement.	39

1. Hiérarchie administrative	39
2. Rapports entre les circonscriptions administratives et les autres unités de déclaration statistique ou de gestion	40
3. Critères et processus de délimitation des districts de recensement	42
4. Délimitation des zones de supervision	44
5. Codage géographique ou « géocodage » des districts de recensement	44
6. Éléments d'une base de données de recensement	46
7. Concordance des districts de recensement avec les recensements antérieurs	47
C. Sources de données géographiques pour la délimitation des districts de recensement	51
1. Types de cartes nécessaires.	51
2. Inventaire des sources existantes	51
3. Importation de données numériques existantes.	52
4. Conversion des données géographiques : de l'analogique au numérique	56
a) Scannage	58
i) Considérations supplémentaires	62
ii) Avantages et inconvénients du scannage	63
b) Numérisation	64
Avantages et inconvénients de la numérisation.	65
c) Édition	66
d) Élaboration et gestion de la topologie.	67
e) Intégration de la carte numérique.	68
f) Géoréférencement.	68
g) Changement de projection et de système de référence . . .	70
h) Intégration de segments de carte séparés.	71
D. Mise en œuvre d'une base de données de districts de recensement .	72
1. Bases de données relationnelles	73
Variétés de bases de données relationnelles et structure de la base de données géographique	74
2. Définition du contenu de la base de données (modélisation des données)	76
E. Problèmes de qualité des données	76
1. Exigences concernant l'exactitude	76
2. Contrôle de la qualité	79
3. Partage du territoire national en unités de traitement	81
4. La carte de base administrative numérique	81
5. Prise en charge d'unités spatiales séparées.	82
6. Superficies utilisées pour le calcul.	83
F. Élaboration de métadonnées	84
G. Résumé et conclusions.	87

IV. Intégration des activités sur le terrain à l'aide de GPS et de données télédétectées	89
A. Systèmes de positionnement universel (GPS)	90
1. Fonctionnement des systèmes de positionnement universel . . .	91
2. Précision du système de positionnement universel	91
3. Systèmes mondiaux de localisation différentielle	92
4. Autres systèmes mondiaux de navigation par satellite	94
5. Les systèmes de positionnement universel dans les applications cartographiques de recensement	95
6. Certaines tâches cartographiques spécifiques liées aux GPS. . .	96
7. Besoins en formation à l'utilisation de GPS	98
8. Résumé : avantages et inconvénients des systèmes de positionnement universel.	98
B. Systèmes intégrés de levé de terrain à l'aide d'ordinateurs portables	99
C. Télédétection satellitale	101
1. Utilisation de l'imagerie pour vérifier sur le terrain les cartes de districts de recensement produites au siège du recensement. . .	101
2. Principes de la télédétection satellitale	102
3. Résolution de données de télédétection.	104
4. Sources en ligne de données de télédétection satellitale	106
5. Applications de données de télédétection pour l'analyse de la population	108
6. Avantages et inconvénients des données obtenues par télédétection satellitale	110
D. Photographie aérienne.	110
1. Aperçu de la photographie aérienne	110
2. Utilisation de la photographie aérienne pour établir les cartes de recensement	113
3. Mise en œuvre et questions institutionnelles liées à la photographie aérienne.	114
4. Avantages et inconvénients des photographies aériennes	115
E. Résumé et conclusions.	116
V. Utilisation de bases de données géographiques (cartes) pendant le recensement	117
A. Introduction : utilisation d'outils géospatiaux pendant le recensement	117
B. Assurance de la qualité, établissement de cartes de districts de recensement et tenue à jour de la base de données	118
1. Aperçu	118
2. Établissement du projet de carte et procédures d'assurance de la qualité	118
3. Appariement des limites et des fichiers d'attributs et impression de cartes d'ensemble	118
4. Assurance de la qualité	120
5. Vérification par les autorités locales et contrôle final de l'unité administrative	121

6. Production de la carte de district de recensement (dont l'impression de la carte)	122
C. Utilisation d'une infrastructure géospatiale pendant le recensement	127
1. Utilisation de cartes numériques pour la logistique du recensement	127
2. Surveillance de la progression des opérations de recensement	128
3. Directives pour l'utilisation de la carte par les agents recenseurs pendant le recensement.	129
4. Mise à jour et correction des cartes de districts de recensement au cours du recensement	129
D. Résumé et conclusions	131
VI. Bases de données géographiques pour la diffusion des résultats du recensement, produits et services	133
A. Introduction	133
B. Tâches à accomplir après le recensement et pendant la période intercensitaire.	134
Tâches immédiates.	134
1. Inclusion de mises à jour et de changements suggérés par les agents de recensement	134
2. Regroupement des unités de collecte et des unités statistiques ou de dépouillement	135
3. Archivage de la base de données	137
4. Tenue à jour de la base de données : avantages d'un programme cartographique permanent.	137
C. Diffusion des produits géographiques de recensement	138
1. Planification de la diffusion des données.	139
2. Considérations relatives à la confidentialité et à la divulgation : problème du calcul de différences.	141
3. Commercialisation de produits et de services de recensement géographique.	142
4. Élargissement du cercle des utilisateurs et sensibilisation	144
5. Liste de produits potentiels.	144
a) Fichiers d'équivalence et de comparabilité.	144
b) Cartothèque de référence.	144
c) Nomenclatures toponymiques et fichiers de centroïdes	145
6. Cartes thématiques destinées à être publiées	145
a) L'avantage des cartes	145
b) Cartographie thématique des données de recensement	146
c) Atlas numériques de recensement.	149
i) Atlas statiques	149
ii) Atlas dynamiques	151
d) Techniques d'analyse spatiale	152
e) Questions relatives à l'établissement et à la publication de cartes : types de produits	157
f) Outils/logiciels cartographiques.	158

7. Options de sortie : fichiers numériques.	159
a) Formats des données	159
b) Formats d'images matricielles.	161
c) Formats de données de SIG	163
d) Données relatives aux coordonnées.	163
e) Données présentées sous forme de tableaux	165
D. Impression	165
1. Aperçu	165
2. Types d'imprimantes	166
3. Impression commerciale	167
E. Données géographiques numériques destinées à être utilisées	169
1. Stratégies de diffusion des données numériques pour atteindre des utilisateurs potentiels.	170
a) Définition du contenu des données.	170
i) Jusqu'à quel niveau les données seront-elles diffusées ?	171
ii) Une grande base de données de SIG ou un ensemble de bases de données de recensement ?	171
iii) Dans quelle mesure doit-on intégrer étroitement les limites et la base de données	172
iv) Quelle quantité de métadonnées doit être fournies .	172
b) Conventions relatives aux noms de fichiers	173
c) Compression.	173
d) Documentation et dictionnaires de données	173
e) Contrôle et assurance de la qualité des produits de données livrables.	175
2. Problèmes juridiques et questions de commercialisation.	176
a) Droit d'auteur sur les données.	176
b) Options concernant la commercialisation des données géographiques	176
c) Questions de responsabilité	178
3. Cartographie sur Internet	179
a) Approches côté serveur.	180
b) Approches côté client.	181
c) Approches hybrides.	182
d) Possibilités en matière de diffusion des données de recensement, y compris le serveur cartographique . . .	182
F. Résumé et conclusions.	186
Bibliographie et références	187
<i>Annexes</i>	
I. Systèmes d'information géographique.	191
A. Introduction aux systèmes d'information géographique	191
1. Matériel, logiciel et données	192

2. Couches de données géographiques.	194
B. Modèles de données de SIG.	195
1. Le système vectoriel.	196
2. Le système matriciel.	198
3. Avantages et inconvénients des modèles de données vectoriel et matriciel.	200
4. Précision ou exactitude.	200
C. Possibilités des SIG.	201
II. Systèmes de coordonnées et projections cartographiques.	207
A. Introduction	207
B. Coordonnées	208
C. Propriétés des projections cartographiques.	211
D. Établissement de cartes plus précises : systèmes de référence géodésique.	214
E. Échelle cartographique	217
F. Exemple de géoréférencement	218
G. Considérations pratiques	221
III. Modélisation de données	223
A. Introduction	223
B. Définition des termes essentiels	223
C. Exemple de modèle.	224
IV. Exemple de dictionnaire de données établi à des fins de diffusion	227
V. Élaboration d'une carte thématique	231
A. Introduction	231
B. Principes de la conception des cartes	231
1. Éléments d'une carte thématique	233
2. Niveaux de mesure et variables graphiques	236
a) Dimensions spatiales	236
b) Niveaux de mesure	238
c) Variables graphiques	238
3. Types de cartes thématiques	240
a) Représentation cartographique d'objets discontinus.	240
b) Données ponctuelles nominales.	242
c) Symboles ponctuels proportionnels.	243
d) Cartes-diagrammes.	246
e) Cartes de flux	249
f) Représentation de phénomènes continus.	250
C. Classification des données	253
1. Classification de données en série.	254
2. Classification statistique	256
3. Points d'inflexion naturels.	261
4. Cartes choroplèthes sans intervalles de catégories	261

5. Classification des données extérieures	263
6. Remarques générales	263
D. Choix de la couleur.	264
E. Conception de la légende cartographique.	267
F. Cartes qui racontent des histoires	269
1. Cartes multivariées	269
2. Petites cartes multiples	271
VI. Glossaire	273
VII. Adresses utiles, URL	293
Logiciels de SIG	293
Logiciels spécialisés	293
Systèmes de traitement d'images de télédétection	293
Images satellitaires à haute résolution et orthophotographie numérique .	294
Systèmes de positionnement universel	294
Revues.	294
Divers	295

Encadrés

II.1. Quatre études de cas de pays	10
II.2. Obstacles technologiques et financiers.	19
II.3. Trois exemples de collaboration en matière d'échange de données.	30
II.4. Participation et coordination des agences internationales.	33
III.1. Critères de sélection d'un logiciel géospatial : LDSM, analyse des images et options FOSS.	53
IV.1. Étude de cas : expérience de Fidji en matière de GPS	100
IV.2. Établissement d'orthophotoplans numériques	112
V.1. Production de cartes de recensement en Inde.	130
VI.1. Étude de cas de la cartographie Web pour la diffusion de données de recensement : Canada	185

Figures

II.1. Étapes de planification des travaux cartographiques de recensement . . .	22
III.1. Hiérarchie géographique générique pour le recensement	39
III.2. Exemple d'une hiérarchie administrative imbriquée.	40
III.3. Système de codage générique des districts de recensement	45
III.4. Éléments d'une base de données de recensement numérique et spatiale. .	47
III.5. Étapes de la mise en place d'une base de données géographique en vue d'un recensement	50
III.6. Corrélations négatives dans le processus de conversion de données	56
III.7. Photo d'un scanner feuille à feuille.	59
III.8. Vectorisation semi-automatique	61
III.9. Vectorisation et lissage de données image scannées	62
III.10. Table à numériser	64

III.11.	Numérisation frontale	66
III.12.	Quelques erreurs courantes de numérisation	67
III.13.	Translation, mise à l'échelle et rotation	69
III.14.	Carte en unités de numérisation; carte en coordonnées du monde réel . .	69
III.15.	Assemblage de feuilles adjacentes de carte numérique.	71
III.16.	Ajustement d'objets après raccordement de coupures de cartes adjacentes	71
III.17.	Exemple d'un tableau d'entité — district de recensement	73
III.18.	Tableaux de base de données relationnelle	75
III.19.	Exactitude logique	77
III.20.	Difficultés lorsque l'exactitude de localisation n'est pas assurée.	78
III.21.	Prise en charge de circonscriptions administratives constituées de plusieurs polygones	83
III.22.	Lac occupant une grande superficie dans plusieurs circonscriptions administratives	84
IV.1.	Systèmes mondiaux de localisation différentielle	92
IV.2.	Délimitation des limites de districts de recensement sur une image satellitale panchromatique	102
IV.3.	Processus de télédétection.	103
IV.4.	Le spectre électromagnétique	104
IV.5.	Exemples de dimensions de pixels dans des photographies aériennes et des images satellitales	105
IV.6.	Film photographique ou image numérisée par balayage.	112
IV.7.	Déformation due au terrain	112
V.1.	Différentes étapes concernant l'assurance de la qualité, l'établissement des produits cartographiques et la tenue à jour de la base de données	119
V.2.	Échantillons de composants d'une carte numérique de district de recensement	124
V.3.	Exemple de carte de district de recensement urbain	125
VI.1.	Exemples d'unités de tabulation censitaire et d'unités déclarantes	136
VI.2.	Le problème du calcul de différences en matière de divulgation statistique.	142
VI.3.	Diverses possibilités de présentation pour un atlas numérique statique de recensement	150
VI.4.	Instantané d'un atlas dynamique ukrainien	153
VI.5.	Création de corridor pour un objet polygone	155
VI.6.	Exemple d'interpolation linéaire créant des courbes de niveau	155
VI.7.	Illustration d'un polygone de Thiessen	156
VI.8.	Exemple de cartogramme	157
VI.9.	Le processus d'impression numérique	168
VI.10.	Cartographie par Internet — l'approche côté serveur	180
VI.11.	Cartographie par Internet — l'approche côté client	182
A1.1.	Bases du SIG	192
A1.2.	Types d'informations stockées dans un SIG.	194
A1.3.	Couches de données — le concept d'espace comme système d'indexation	195
A1.4.	Points, lignes et polygones.	196
A1.5.	Modèles de données vectorielles : mode spaghetti et mode topologique. .	197

A1.6.	Données spatiales et non spatiales stockées dans un SIG vectoriel	198
A1.7.	Exemple de fichier de données matricielles	199
A1.8.	Mode vectoriel et mode maillé peuvent être utilisés l'un et l'autre pour présenter des données discontinues et continues.	199
A2.1.	Systèmes de coordonnées planaires et polaires	208
A2.2.	Coordonnées sur la sphère : le système de référence latitude/longitude . .	208
A2.3.	Exemple de processus de projection cartographique (projection azimutale)	209
A2.4.	Familles de projections cartographiques.	210
A2.5.	Différents modes d'établissement de la projection	211
A2.6.	Projections cartographiques courantes	213
A2.7.	Sphère et ellipsoïde.	214
A2.8.	Le système UTM	215
A2.9.	La situation du bâtiment du Siège de l'ONU dans le système de référence UTM	216
A2.10.	Points de contrôle sur une feuille de carte	218
A5.1.	Établissement d'une hiérarchie visuelle par le choix des couleurs ou des nuances de gris.	232
A5.2.	Échantillon de carte des unités administratives et des principaux centres urbains	235
A5.3.	Exemple de carte thématique représentant la densité démographique . . .	236
A5.4.	Effets de la généralisation sur l'affichage d'objets spatiaux.	237
A5.5.	La complexité du monde réel doit parfois être simplifiée pour faciliter sa représentation dans un SIG	237
A5.6.	Mesure des variables.	238
A5.7.	Variables graphiques correspondant à des polygones, des lignes et des points	239
A5.8.	Carte de densité par points	241
A5.9.	Combinaison de cartes de points et de cartes choroplèthes	242
A5.10.	Représentation d'objets ponctuels distincts	243
A5.11.	Symboles proportionnels correspondant à des objets ponctuels et surficiels	244
A5.12.	Pictogrammes ou symboles graphiques simples	244
A5.13.	Présentation de l'importance et de la direction des flux au moyen de symboles graphiques simples.	245
A5.14.	Représentation de valeurs de données faisant varier le nombre de symboles cartographiques pour chaque objet.	245
A5.15.	Carte-diagramme à secteurs	246
A5.16.	Combinaison de cartes choroplèthes et de cartes-diagrammes à secteurs .	247
A5.17.	Carte montrant l'évolution dans le temps au moyen d'histogrammes . . .	248
A5.18.	Combinaison de cartes et de pyramides de population	248
A5.19.	Représentation des rapports de masculinité sur une carte.	249
A5.20.	Autres présentations des flux entre régions	250
A5.21.	Représentation de l'immigration et de l'émigration	251

A5.22.	Différentes méthodes cartographiques utilisables pour présenter des données continues	252
A5.23.	Intervalles égaux	255
A5.24.	Établissement de cartes par quantiles (fréquence égale)	257
A5.25.	Écart-type	258
A5.26.	Affectation de couleurs aux catégories déterminées par les écarts-types . .	260
A5.27.	Points d'inflexion naturels.	262
A5.28.	Différents types de légendes pour des cartes ombrées	267
A5.29.	Légendes montrant la répartition des données statistiques	268
A5.30.	Combinaison de symboles ombrés pleins et hachurés pour présenter deux variables sur la même carte	269
A5.31.	Carte équivalant à un tableau à double entrée.	270
A5.32.	Petites cartes multiples — représentation de l'évolution dans le temps	271

Tableaux

IV.1.	Produits de télédétection de satellites civils à résolution spatiale très élevée	106
VI.1.	Comparaison entre d'anciennes et de nouvelles unités de districts de recensement	135
VI.2.	Suggestions de cartes thématiques pour un atlas de recensement.	147
A2.1.	Coordonnées projetées du bâtiment du Secrétariat de l'ONU à New York utilisant différents ellipsoïdes de référence	215
A2.2.	Paramètres de transformation	220
A3.1.	Informations réunies pour définir un modèle de données spatiales	224
A3.2.	Exemple : unités administratives pour un pays comportant trois niveaux subnationaux	225
A5.1.	Évaluation des différentes techniques de classification.	263
A5.2.	Choix des nuances de gris et des couleurs.	266

Chapitre I

Introduction

A. Avant-propos et raison d'être du *Manuel*

1.1. Le nouveau *Manuel sur l'infrastructure géospatiale à l'appui des activités de recensement* tient compte des réalisations de son prédécesseur, le *Manuel des systèmes d'information géographique et de cartographie numérique*, publié en 2000, et cherche à les développer.

1.2. Au rang des innovations technologiques majeures, citons la large disponibilité des ordinateurs personnels, des ordinateurs portables, des systèmes de positionnement universel (GPS), des systèmes d'information géographique (SIG) ainsi que de l'imagerie satellite et aérienne à bas coût. Grâce à ces avancées, les organismes nationaux de statistique ont disposé de nouveaux outils de collecte d'informations plus précises, opportunes et impartiales sur leurs populations. L'émergence de nouvelles technologies est de fait l'un des éléments moteurs du nouveau *Manuel*. Dans le même temps, il est admis que l'adoption de ces nouvelles méthodes remettra en question le leadership des organismes nationaux de statistique et amènera des modifications dans leur organisation. De nouvelles capacités géospatiales sont mises en œuvre tout au long du programme de recensement.

1.3. Le *Manuel* explique que les plus grands défis auxquels les organismes nationaux de statistique sont confrontés sont non seulement d'ordre technique mais également à caractère institutionnel et managérial. La plupart des pays membres ont commencé à utiliser la science et la technologie dans le domaine de l'information géographique en fonction des besoins de leurs programmes. Le *Manuel* vise notamment à satisfaire ces besoins. Le mandat juridique ou constitutionnel qui sous-tend un recensement démographique national dans de nombreux pays est toujours d'application. Dans le même temps, les données des recensements sont de plus en plus utilisées dans le cadre de la gestion des désastres et à bien d'autres fins (pour un aperçu exhaustif, voir chapitre II). On constate par ailleurs une plus grande disponibilité des données sous diverses formes et la mise en œuvre de nouvelles technologies simplifiant la collecte, l'analyse et le stockage de données.

1.4. Dans le cadre de la production de données à ces fins, les pays prennent conscience qu'ils peuvent tirer parti des atouts d'autres agences gouvernementales à travers une infrastructure nationale de données spatiales (NSDI)¹. Une NSDI est un accord institutionnel permettant un échange de données et une collaboration intergouvernementaux notamment aux niveaux national, régional et local. Bien qu'elles ne revêtent pas en soi un caractère technique, d'autres questions institutionnelles, telles que les principes de base du financement, de la dotation en personnel et de la gestion de projets, seront déterminantes pour le succès des projets de recensement géospatial. Le principe « Créer une fois, utiliser plusieurs fois » vaut pour la mise en place d'une

¹ Une infrastructure nationale de données spatiales allie la technologie, les politiques, les normes et les ressources humaines nécessaires pour acquérir, traiter, stocker, distribuer et améliorer l'utilisation des données géospatiales. Les éléments conceptuels d'une telle infrastructure comprennent : a) un cadre institutionnel précisant les politiques ainsi que l'aide administrative et juridique requise pour la création, la gestion et l'application de normes aux ensembles de données fondamentales; b) les normes définissant les caractéristiques techniques applicables aux ensembles de données fondamentales; c) les ensembles de données fondamentales exigeant un cadre géodésique, les bases de données topographiques et cadastrales; et d) un cadre technologique permettant aux utilisateurs d'identifier et d'accéder à des ensembles de données fondamentales (voir *GSDI Cookbook*, 2000).

base de données géospatiale de recensement nationale, étant donné qu'une fois créée, elle peut servir dans divers contextes nationaux et à diverses fins.

1.5. Le présent *Manuel* fournit des pistes constructives pour réorganiser la cartographie de recensement ainsi que les tâches analytiques inhérentes aux bases de données numériques géographiquement référencées. Afin de structurer l'organisation statistique autour de telles « bases de données géographiques », une planification soignée s'impose, non seulement en raison des investissements nécessaires en amont pour les SIG, mais également compte tenu du travail demandé pour développer les capacités d'analyse des données de recensement et livrer des produits au public de façon opportune. Le succès des SIG en tant qu'industrie repose sur la capacité des informations géospatiales à résoudre les problèmes et à faciliter les décisions.

1.6. Le développement des capacités des SIG implique une modification de la structure organisationnelle traditionnelle d'un organisme statistique national, et plus particulièrement l'élargissement de l'« unité cartographique » actuelle à un noyau géographique beaucoup plus large et versatile d'une capacité suffisante pour satisfaire l'ensemble des besoins en cartographie censitaire. Une telle réorganisation exigera éventuellement un financement continu des opérations, en plus d'un personnel disposant de l'expertise nécessaire pour épauler les organismes statistiques nationaux dans leurs tâches tout au long d'une décennie. Pour un noyau géographique spécifique, il faut une équipe disposant non seulement de compétences géospatiales, mais ayant également sur le plan opérationnel la volonté de moderniser les recensements. Le personnel qualifié respectera un calendrier strict garantissant que les recenseurs sur le terrain disposeront à temps de cartes détaillées corrigées pour le recensement. Une base de données de districts de recensement numériques consignera des estimations démographiques de localités ainsi que des limites de districts de recensement définies géospaialement de manière à représenter de petites zones. Les opérations nécessiteront une gestion continue, notamment en combinant les activités sur le terrain (ce qui peut s'avérer onéreux lorsque les équipes sont dotées de véhicules tout-terrain) et la collecte de données de télédétection, pour ainsi disposer d'un recensement plus précis et rentable.

1.7. Il n'est jamais trop tôt pour planifier la mise en œuvre opérationnelle d'un programme géographique. Compte tenu des longs délais qu'impose une restructuration organisationnelle complexe, la planification doit être entamée des années à l'avance. Quant au recensement, les recenseurs et chefs d'équipe devront consulter des cartes à grande échelle très détaillées. La production de données géographiques détaillées à une échelle suffisante à utiliser avec d'autres couches de données d'un SIG, dont les unités de diffusion sont composées de districts de recensement ou d'ensembles de population, imposera une planification considérable au bureau de statistique national. La définition d'un noyau géographique ou d'un concentrateur de données spécifique, à l'aide d'un système de base de données relationnelle à l'échelle de l'entreprise, peut exiger une restructuration organisationnelle. Les couches de données fondamentales, telles que les frontières démographiques et administratives définies par les organismes statistiques nationaux, peuvent être communiquées à de nombreux utilisateurs, évitant ainsi les coûts occasionnés par un chevauchement des efforts. Ces deux couches sont des composants de base de toute infrastructure nationale de données spatiales. En effet, les organismes statistiques nationaux pourraient contribuer aux initiatives des infrastructures nationales de données spatiales de la façon suivante :

- a) Une base de données géographique spatiale, contenant des informations sur les polygones et les attributs spécifiques aux districts de recensement du pays (à savoir les unités pour lesquelles le territoire est alloué aux

démarcheurs pendant le recensement). Comme en témoignent de récentes expériences nationales, une base numérique commune peut faciliter les recensements de l'agriculture et de la population. Les données des recensements peuvent être publiées au niveau des districts de recensement ou rassemblées afin de constituer de nouvelles unités de diffusion spécifiques à de petites zones, telles que des ensembles démographiques;

- b) Une bibliothèque de frontières administratives numériques, allant du niveau provincial au niveau municipal (voire peut-être même au niveau de la parcelle de terrain). Si les limites sont totalement recensées et corrigées sur le terrain par l'organisme national de statistique, des économies substantielles seront alors réalisées puisque le territoire ne devra pas faire l'objet d'une nouvelle étude du territoire à d'autres fins, telles qu'un redécoupage électoral;
- c) Un répertoire toponymique national, renseignant notamment des noms géographiques et des coordonnées des habitats (dénommés «codes-p» au sein de la communauté humanitaire des Nations Unies). Couplés aux estimations de la population, ces répertoires peuvent être exploités par les organisations humanitaires à des fins de développement et d'intervention en cas d'urgence. Des données supplémentaires sur les habitations au format vectoriel peuvent être utilisées de la même manière.

1.8. Quant aux données démographiques et géographiques au format SIG à partager entre d'autres organisations, les organismes statistiques nationaux respecteront les normes applicables au référencement et aux métadonnées géographiques définies au niveau national. Il convient d'accorder une attention particulière au système de classification administratif à utiliser pour l'organisation du territoire visé par le recensement. Le présent *Manuel* parle de «géocodage», étant donné qu'il permet de relier les informations démographiques à leur emplacement sur la surface de la Terre.

1.9. Le *Manuel* met surtout en exergue la nécessité d'élaborer des plans réalistes afin de tirer parti du potentiel des SIG ainsi que d'autres technologies géospatiales pour la modernisation des recensements et l'amélioration des résultats et analyses. Il est par ailleurs souligné qu'il est indispensable d'étendre la portée de la cartographie censitaire au cadre spatial national du pays.

B. Portée, objet et résumé du *Manuel*

1.10. L'évolution rapide ces derniers temps des techniques de la cartographie numérique et de l'analyse géographique, et l'augmentation de la demande de données démographiques intéressant une petite zone géoréférencée sont les principales raisons de la rédaction du présent *Manuel*. Tout pays qui entreprend un projet de recensement doit évaluer les possibilités existantes pour réduire au minimum les coûts et tirer le plus grand avantage possible des activités géospatiales requises. Outre le fait qu'il informe les principaux partisans d'initiatives géographiques au sein des organismes statistiques nationaux, le présent *Manuel* a pour objet de fournir les renseignements techniques et méthodologiques pour faciliter le choix d'un ensemble approprié d'outils et de procédures pour un pays donné.

1.11. Les choix à opérer seront évidemment différents dans chaque cas en raison du nombre de possibilités offertes et aussi parce que la situation et les ressources disponibles diffèrent d'un pays à l'autre. Le présent *Manuel* est donc conçu comme un document de référence pratique, un «livre de cuisine» illustrant le rôle des technolo-

gies géospatiales à chaque étape du recensement. Chaque pays doit déterminer dans quelle mesure les solutions de cartographie existantes sont adaptées au contexte de son propre programme de recensement et à sa planification nationale. Des questions telles que les ressources géographiques du pays, les ressources technologiques et le personnel disponibles, les moyens financiers et les délais autorisés pour réaliser les tâches géographiques de recensement détermineront la meilleure combinaison de techniques et de méthodes dans chaque cas particulier.

1.12. Le présent document n'est pas un manuel de SIG général. Il ne contient pas les commandes et les routines nécessaires à l'utilisation d'un logiciel spécifique et ne recommande nullement un tel logiciel. Il ne s'agit par ailleurs pas d'un guide général sur la réalisation d'un recensement. Le *Manuel* respecte la grande tradition de la cartographie de recensement, alléguant que les techniques traditionnelles de cartographie analogique utilisées avec succès dans de nombreux pays restent pertinentes. La principale référence en la matière — *Mapping for Censuses and Surveys* (BUCEN, 1978) — continue d'être une ressource précieuse pour les cartographes débutants ou expérimentés. En particulier les chapitres relatifs à l'organisation et au contrôle d'un programme cartographique, la délimitation des districts de recensement et les zones délimitées à des fins statistiques sont encore pertinents. Le *Manuel* recourt fréquemment aux données de son prédécesseur. Mais avec les progrès des techniques, on dispose aujourd'hui de meilleurs moyens pour effectuer un grand nombre de travaux cartographiques de recensement. Le présent *Manuel* a donc pour objet de compléter des directives antérieures en fournissant des informations sur les technologies récentes.

1.13. Il vise deux public cibles : les gestionnaires préoccupés par les coûts et les bénéfices d'un investissement dans les technologies géospatiales et les techniciens mettant en œuvre les aspects géographiques d'un plan de recensement. Les cinq chapitres du *Manuel* présupposent une connaissance de base des SIG et des concepts cartographiques. À l'intention des lecteurs qui ne sont pas familiers avec ces sujets, les annexes I et II en fournissent un bref aperçu. En particulier, les projections cartographiques et les systèmes de coordonnées sont un sujet plus important dans un projet utilisant les SIG que dans le cas d'une méthode traditionnelle basée sur des croquis cartographiques.

C. Résumés chapitre par chapitre

1.14. Les chapitres II à VI du présent *Manuel* traitent de domaines auxquels les technologies géospatiales peuvent contribuer avant, pendant et après le recensement. La structure du *Manuel* a été révisée pour tenir compte de la volonté dans de nombreux États Membres de passer à des opérations de recensement numériques.

1.15. Le chapitre II propose des considérations managériales à l'attention des responsables d'agences statistiques et présente quelques questions à considérer dans le cadre de la réorganisation de l'organisme statistique national afin de permettre de tirer pleinement parti de l'infrastructure géospatiale. Toute planification visant à satisfaire à la nouvelle demande en données statistiques désagrégées, dont celles pour la gestion des désastres, exige de travailler à reculons, de la livraison de nouveaux produits aux exigences requises à chacune des étapes du processus. Des arguments sur le rendement, dont l'analyse coûts-avantages, sont avancés mettant en lumière le rôle des technologies géospatiales en tant qu'investissement stratégique. Les facteurs déterminant le succès de la mise en œuvre d'un projet géospatial sont ensuite traités, suivis d'une discussion détaillée sur le processus de planification. Elle porte notamment sur la réalisation d'une évaluation des besoins, la définition des produits et options, le re-

crutement du personnel et d'autres considérations relatives aux ressources humaines ainsi que la coopération institutionnelle avec d'autres producteurs et utilisateurs de données à travers des NSDI.

1.16. Le chapitre III propose au gestionnaire informatique ou au responsable SIG/cartographie les éléments techniques nécessaires pour établir de manière pragmatique une base de données géographique numérique. Il passe en revue chacune des étapes de l'élaboration d'une base de données du niveau des districts de recensement, en commençant par l'importance du codage géographique, les hiérarchies administratives et les composants d'une base de données censitaire. Les questions de délimitation des districts de recensement sont ensuite abordées. S'ensuit une discussion sur les sources actuelles de données géographiques utilisées pour une telle délimitation, dont les sources numériques et analogiques. Il est ensuite question de la saisie numérique par le biais du scannage et de la numérisation, puis d'une discussion sur l'intégration des cartes, la construction topologique et le géoréférencement. Lors de l'élaboration d'une base de données géographique, de nouveaux problèmes de qualité des données peuvent se poser tant sur le plan de la précision que de l'exactitude et des procédures d'assurance qualité. Une rubrique détaillée sur la mécanique des métadonnées vient clore le chapitre. Elle insiste sur l'importance de la documentation et suggère des stratégies pour les utilisateurs de données à la fois externes et internes.

1.17. Le chapitre IV, lui aussi à caractère technique, détaille le processus d'élaboration d'une base de données géographique du niveau du district de recensement et l'utilisation des innovations technologiques, telles que les GPS et la télédétection, afin d'effectuer des corrections, en les intégrant au besoin aux travaux sur le terrain. Un tri a ici été effectué afin de mettre l'accent sur les domaines ayant changé depuis le recensement précédent. Les principes de base d'un GPS sont expliqués, en plus de nouvelles améliorations techniques d'un GPS différentiel, des besoins de formation et de l'utilisation des ordinateurs portables. Viennent ensuite une section sur la télédétection satellitaire, répertoriant les sources de télédétection disponibles ainsi que les outils de visualisation, tels que Google Earth, et une autre rubrique sur les applications de la photographie aérienne aux fins des travaux de recensement.

1.18. Le chapitre V traite du processus de création des cartes requises pour le dénombrement, en poursuivant l'analyse du point de vue opérationnel là où la discussion sur les bases de données géographiques s'est arrêtée. Certaines de ses rubriques sont consacrées à l'assurance qualité, l'élaboration de cartes pour un usage sur le terrain, les éléments et la conception des cartes, l'impression et la distribution de ces dernières ainsi que leur utilisation pour les principes de base de la logistique du recensement. Le chapitre insiste sur le fait que pour parer aux mésaventures, les plans pour imprévus doivent faire partie intégrante de la gestion de projets.

1.19. Au chapitre VI, il est question de l'exploitation de l'infrastructure géospatiale pour la diffusion des résultats des recensements. L'approche adoptée est similaire à celle appliquée à la phase post-censitaire des chapitres précédents, mais s'appuie sur la notion de base de données géographique (géocodée en fonction d'emplacements spécifiques) comme interface de données centrale pour la publication des produits de recensement. Les sujets traités sont entre autres l'agrégation des unités de collecte et unités de diffusion, la maintenance et l'archivage de bases de données géographiques ainsi que la planification de produits censitaires. Une rubrique sur l'analyse spatiale propose des exemples et des options pour une interprétation visuelle des données de recensement. D'autres questions sont également abordées : la divulgation et le respect de la vie privée (point pertinent dans le cas de données relatives à de petites zones) et la possibilité pour une agence statistique de tirer des revenus de la vente de produits à

valeur ajoutée, tels que des CD-ROM ou des DVD de données. Il est en outre question de produits de données géographiques, tels que des solutions de visualisation de cartes, des fichiers de données spatiales attribués réservés aux progiciels commerciaux ou aux produits en ligne de SIG.

1.20. Les annexes constituent un guide pratique pour les responsables de la planification et de la mise en œuvre de solutions géospatiales dans le cadre de projets de recensement. L'annexe I fournit des renseignements de base sur les modèles de données de SIG ainsi que leur degré de pertinence et de précision. L'annexe II présente les systèmes de coordonnées et les projections cartographiques, en s'arrêtant en particulier sur les questions liées à la saisie de données dans un SIG. L'annexe III concerne la modélisation des données et propose un exemple de modèle. L'annexe IV fournit un exemple de répertoire de données. L'annexe V propose plusieurs conseils en matière de conception de cartes thématiques pour l'analyse et l'affichage des données censitaires, dont les cartes choroplèthes conventionnelles. L'annexe VI contient un glossaire de la terminologie SIG communément utilisée. L'annexe VII reprend une liste de coordonnées et d'URL utiles pour obtenir un complément d'information.

Chapitre II

Considérations managériales pour les responsables des services nationaux de statistique et autres décideurs

A. Introduction

2.1. Ce chapitre s'adresse aux organismes nationaux de statistique, aux directeurs du recensement et aux responsables des services géographiques. Il traite essentiellement des questions institutionnelles (c'est-à-dire le contenu technique non-SIG), en mettant l'accent sur diverses considérations liées à l'utilisation de technologies géospatiales. Les organismes nationaux de statistique peuvent produire des données géoréférencées plus précises en moins de temps, mais uniquement sous réserve d'une planification soignée des activités. Des exemples d'expériences de pays illustrant l'utilité de technologies géospatiales pour des travaux de recensement sont proposés.

2.2. Afin de planifier un recensement recourant aux fonctions géospatiales, il peut être nécessaire de réorganiser l'organisme national de statistique dans une certaine mesure, autour de la géographie — y compris la cartographie et les SIG, mais également de thèmes tels que le codage géographique et les limites administratives. Il n'existe aucune solution technologique permettant aux organismes nationaux de statistique de moderniser leur recensement de la même manière; le *Manuel* présente donc toute une série d'options grâce auxquelles un organisme national de statistique pourra adopter de nouvelles technologies en disposant d'un niveau de compétence et d'expérience adéquat. Le *Manuel* présente ces options comme un éventail de possibilités pour l'organisme national de statistique, proposant notamment des tâches spécifiques pouvant être adaptées et adoptées en tout ou en partie, selon la taille du pays et de l'organisme. Tout au long du processus, il est important d'insister sur la nécessité de commencer à planifier tôt et de demander de l'aide au besoin.

2.3 Utilisés par les gouvernements nationaux pour obtenir des dénombrements de population exacts, les recensements sont également l'un des outils les plus importants de la société pour comprendre les populations humaines sur un plan scientifique. La technologie géospatiale a modifié la manière de présenter les informations pour le bien de la société et pour promouvoir le développement durable, économique ou social. Elle promet des gains de productivité sensibles dans presque n'importe quel secteur, grâce au principe d'organisation centralisée de la géographie accordant de l'importance au lieu. Une organisation géographique des informations selon un affichage ordonné des données en fonction de l'emplacement permet effectivement d'établir une corrélation entre les observations sociales et leur localisation. Visant à collecter des données, un recensement à référence géographique, est, pour

un pays, une opportunité sans pareil d'obtenir des informations détaillées sur sa population.

2.4. Le rôle dévolu aux SIG et aux technologies géospatiales en général dans le cadre des opérations gouvernementales, en ce compris l'imagerie satellite de télédétection et le GPS, évolue. Les pays se rendent de plus en plus compte de tout l'intérêt d'une organisation des informations gouvernementales autour d'un modèle spatial implicitement axé sur la géographie. Le recours aux bases de données géographiques promet d'être fructueux, et s'est bien souvent avéré l'être, pour des aspects aussi importants que l'efficacité et le service à la clientèle, lorsque les agences gouvernementales s'échangent les résultats et les demandes à l'aide d'outils géospatiaux.

2.5. De nombreuses études des services publics en ligne ont illustré les avantages procurés par l'utilisation de ces technologies. O'Looney (2002) met en avant une plus grande perméabilité des frontières entre le gouvernement et les entreprises, les groupes civiques et les citoyens. Le recours aux technologies de l'information pour prester des services permet une interaction plus forte avec un personnel relativement réduit. Garson (2003) souligne la responsabilité accrue qui en découle, en particulier lorsque les usagers de données peuvent vérifier l'arithmétique et remettre en question les résultats. Khosrow-Pour (2005) se penche sur les nouveaux défis administratifs systémiques exigeant un plus grand engagement aux niveaux supérieurs ; un « gouvernement 24h sur 24 » a des répercussions évidentes sur la fourniture de produits et services, offrant aux citoyens/clients davantage de possibilités de libre-service. White (2007) met en exergue la nécessité de politiques d'utilisation de l'internet internes transparentes, invitant par ailleurs les principaux décideurs à gérer plus efficacement les fournisseurs pour satisfaire aux objectifs à court terme en matière de ressources humaines, avec un impact à long terme moindre.

2.6. L'investissement initial en matériel et logiciels et au niveau de la réorganisation étant important, il convient de planifier soigneusement l'utilisation des technologies géospatiales aux fins des activités de recensement. David Rhind, chef de file de l'industrie des SIG au Royaume-Uni, reconnaît que ces derniers sont non seulement une technologie mais que les commerces, les autorités publiques et le milieu universitaire — agissant tous d'une certaine manière comme des entreprises, y recourent de plus en plus. À l'instar de n'importe quel investissement stratégique, il convient de considérer sérieusement les coûts et avantages des SIG. Roger Tomlinson (Canada) soutient qu'une bonne planification est la clé du succès, que ce soit pour déployer des SIG ou toute autre technologie.

2.7. Un recensement est une initiative nationale minutieuse qui aura des répercussions politiques au cours des années à venir. Les chefs d'un organisme national de statistique et autres gestionnaires sont invités à y réfléchir, en s'attardant sur les résultats que l'organisme national de statistique obtiendra au cours des prochaines années. En remontant de la publication à la planification et aux besoins de données, il leur appartient de déterminer les produits de recensement conventionnels — dont les tableaux, totaux et ventilations en fonction de l'âge et du sexe — à diffuser. Ils devraient également considérer de nouvelles offres, telles que des atlas, des DVD proposant des données désagrégées, des données régionales ou des micro-données, ainsi que des cartes électroniques capables de satisfaire aux besoins de toute une myriade de nouveaux utilisateurs de données, contribuant ainsi à renforcer la satisfaction globale de la clientèle. Ces nouvelles capacités sont à portée de main grâce au potentiel de la géographie. Comme le souligne par ailleurs Tomlinson, un SIG est une technologie particulièrement horizontale proposant une large palette d'applications à travers le paysage intellectuel et industriel. Il conviendrait de l'adopter pour

répondre aux besoins spécifiques. L'objectif stratégique des organismes nationaux de statistique ne doit pas être perdu de vue. Il a plus que probablement pour mission de réaliser un recensement précis, en respectant les délais impartis et les limites budgétaires. Il s'agira de clarifier les objectifs ministériels ou mandats spécifiques confiés à l'organisation, au même titre que leur impact sur les plans de recensement. Tomlinson demande aux organisations de déterminer comment améliorer leurs objectifs stratégiques grâce aux nouvelles technologies, telles que le GPS ou les bases de données géographiques. Il recommande de procéder à des analyses coûts-avantages approfondies liées à l'adoption de technologies pour l'horizon de planification spécifié du recensement. Avant de planifier les produits que leur agence peut mettre au point, les gestionnaires sont invités dans un premier temps à discuter avec les utilisateurs et à élaborer une description du produit d'information pouvant répondre aux besoins. Il leur est ensuite recommandé de créer un concept de données tenant compte des besoins identifiés, de choisir un modèle logique applicable aux données planifiées, de déterminer les exigences du système et enfin de planifier la mise en œuvre.

2.8. Dans le domaine du recensement, les organismes nationaux de statistique doivent envisager de réorganiser les systèmes de traitement et de diffusion de manière à proposer les produits attendus par les clients. Les conséquences sur la structure organisationnelle risquent d'être vastes, à commencer par la possibilité d'organiser les opérations autour d'un centre de données ou d'un noyau d'informations stockant ou fournissant des données au sein de l'organisation.

2.9. Un «recensement géocentrique», consiste à organiser le recensement autour de la géographie. Pour de nombreux organismes nationaux de statistique sur le point de procéder à une intégration totale des capacités numériques, il s'agit souvent d'investir dans la conversion analogique/numérique de cartes de districts de recensement, laquelle exige des procédures de numérisation et de correction minutieuses afin de pouvoir utiliser les districts de recensement pour les nouvelles bases de données géographiques numériques. Une comparaison est alors possible entre ces nouvelles bases de données et des images de télédétection, telles que des photos aériennes ou images satellites, avant leur correction à l'aide d'un GPS. Grâce à l'intégration de technologies telles que la numérisation, l'imagerie satellite et le GPS, l'unité cartographique peut axer ses efforts sur les régions ayant le plus besoin d'une mise à jour depuis le recensement précédent.

2.10. Au cours du dénombrement, les technologies géospatiales peuvent simplifier la planification de la logistique et tenir les décideurs informés de l'évolution du processus. Une fois le recensement terminé, les technologies peuvent aider la société à en prendre connaissance à travers la diffusion des informations, en particulier à l'aide de l'internet. Une approche géospatiale unique peut permettre de coordonner plus facilement la dissémination des produits du recensement à des milliers de nouveaux utilisateurs de données et inciter les utilisateurs existants à demander de nouveaux produits et services.

2.11. L'un des facteurs clés de toute nouvelle initiative technique est le renforcement des capacités. Bon nombre de directeurs d'organisme national de statistique ne pensent pas disposer du budget ou de la capacité institutionnelle nécessaires à la réorganisation de leurs agences. Il s'agit là moins d'une question budgétaire que de l'étendue de l'horizon de planification. Les gestionnaires devraient planifier pour cinq ou dix ans; pour l'organisme national de statistique, le degré de détail des produits d'information devra être bien plus précis qu'il ne l'est actuellement.

Encadré II.1

Quatre études de cas de pays**1. Namibie**

La Namibie a entamé son programme de SIG et de cartographie numérique en vue de son recensement de 2001, l'objectif principal étant une production efficace de cartes de base nécessaires au travail sur le terrain. L'infrastructure du SIG a été mise en place avec le concours d'une société de conseils, les données spatiales requises pour le recensement étant saisies et numérisées à l'aide d'un GPS. La superficie de la Namibie avoisine 824 000 km² et sa population compte 1,8 million d'habitants; la couverture de la totalité de la superficie du pays posait donc un défi logistique. Le Bureau central de statistique (BCS) a délimité des districts de recensement pour 13 régions et 107 circonscriptions, en les géocodant à l'aide d'identifiants uniques à 9 chiffres. Il a par ailleurs créé des couches limites supplémentaires, incluant notamment des parcs nationaux, des exploitations agricoles, des communes, des superficies déterminées et des localités. L'utilisation d'une structure organisationnelle qui, à l'époque, ne disposait pas de SIG, l'absence de personnel qualifié, le manque de formations, le manque initial de données spatiales, des problèmes frontaliers et l'inaccessibilité de certaines zones comptaient parmi les défis à relever. Afin de combler les lacunes en matière de ressources humaines, la Namibie a scellé un partenariat avec l'institut polytechnique pour développer des programmes d'études spécifiques dans le domaine des SIG et des technologies de l'information et des communications (TIC). Pour son recensement de 2011, la Namibie établira un SIG en ligne à l'aide du logiciel libre Postgress grâce auquel les utilisateurs pourront créer leurs propres cartes. Les données seront stockées et consultées de manière centralisée avec Oracle. Le Bureau central de statistique (BCS) envisage de saisir des unités d'habitation et de les utiliser pour démarquer les districts de recensement (pour de plus amples informations, veuillez contacter Otilie Mwazi : omwazi@npc.gov.na).

2. Bhoutan

Le Bhoutan a procédé à son recensement de la population et de l'habitation de 2005 en recourant à un SIG. Le Bureau national de statistique a assuré une couverture complète en délimitant 6 800 districts de recensement pour un pays de 47 000 km² et une population totale d'approximativement 2,3 millions d'habitants. Lors du recensement des habitations, toutes les structures, dont les abris, les logements provisoires et les grottes, ont été localisées par GPS. Des données spatiales ont été extraites de cartes topographiques à une échelle de 1/50 000 et des districts de recensement créés de manière à les imbriquer dans des unités administratives sous-nationales. S'agissant de la diffusion des données, le Bureau national de statistique a installé un serveur Web et établi une série de limites administratives. Une partie des cartes indicatrices diffusées avait trait à l'emploi, la santé, le logement, l'accès à l'eau, l'énergie et l'assainissement. Des projets appellent à l'utilisation des bases numériques pour la planification du prochain recensement du Bhoutan (pour un complément d'information, contacter Thinley Jyamtshow Wangdi à : thinly_j@yahoo.com)

3. Sainte-Lucie

Sainte-Lucie a commencé à développer ses capacités en matière de SIG en 1995 pour le recensement de son agriculture. Le Bureau central de statistique s'est rendu compte que les cartes de niveaux réalisées à la main utilisées par le passé n'étaient pas adaptées à la localisation d'unités d'habitation nécessaire au recensement de la population. Grâce à une coopération avec le service des recherches et de l'aménagement, l'unité en charge de l'aménagement du territoire du ministère de la planification, le département des forêts et le ministère de l'agriculture, ainsi que, pour le secteur privé, avec Cable and Wireless et St. Lucia Electricity Services, le Bureau central de statistique a numérisé des cartes de districts de recensement et employé des GPS haut de gamme pour la saisie des données spatiales sur le terrain. En 2004, le Bureau central de statistique a pu exploiter des photos aériennes de l'île grâce au concours du département des levés et de la cartographie du ministère de l'aménagement du territoire. Il a établi les limites des zones d'habitat en se basant sur des latitudes et longitudes pour chacune des unités d'habitation; Sainte-Lucie a ainsi pu être délimitée dans presque chaque recoin. Les habitats ont été géocodés à l'aide d'une combinaison de neuf chiffres. Dans le cadre de la préparation aux activités de conversion en produits cartographiques numériques, Sainte-Lucie a été confrontée à des défis au niveau des ressources humaines, dont une érosion des effectifs. Pourtant, le pays est parvenu à former son personnel aux SIG et mener à bien ses tâches (pour un complément d'information, contacter Sherma Lawrence : sherma_l@slucia.com).

4. Brésil

En 2007, le Brazilian Institute for Geography and Statistics (IBGE) a réalisé un recensement sur trois axes : le recensement de l'agriculture, le dénombrement de la population et le registre national d'adresses à des fins statistiques. Il a travaillé sur ces trois plans à la fois de manière simultanée et intégrée. La base de données territoriale utilisée pour cette opération repose sur une série de cartes et de fichiers d'enregistrement (cadastres) représentant la division du territoire brésilien en petites zones géographiques ou « districts de recensement ».

La base de données territoriale elle-même totalise 249 068 districts de recensement. 162 770 d'entre eux ont été visités lors du recensement de 2007, 70 085 étant des espaces ruraux et 92 685 des zones urbaines.

L'IBGE compte 27 bureaux d'État dont les sièges se situent dans les capitales de chaque État. Les bureaux d'État coordonnent les activités de 530 bureaux locaux. Chaque bureau local couvre un groupe de municipalités. De cette manière, les activités de recensement peuvent couvrir l'ensemble des 5 564 municipalités brésiliennes. Le siège de chaque bureau d'État dispose d'un département bases de données territoriales responsable de la cartographie censitaire de l'État correspondant. De plus, six des bureaux d'État sont dotés d'unités régionales de gestion de la cartographie et de la géodésie dans tout le pays, aidant les départements bases de données territoriales à produire les cartes (statistiques) numériques des municipalités.

Cette infrastructure a permis d'élaborer les cartes de districts de recensement en 2007. Des cartes imprimées sur papier ont été actualisées dans les 530 bureaux locaux de l'IBGE et traitées numériquement dans les départements bases de données territoriales avec l'aide des amas globulaires galactiques susmentionnés. Deux unités managériales de la direction des géosciences de l'IBGE, chargées respectivement de la cartographie censitaire rurale et urbaine, ont assuré la coordination centrale de toutes ces activités. Les cartes ont ensuite été envoyées pour révision et approbation aux commissions de recensement des municipalités, dont les membres sont des représentants des autorités locales et de la société.

Afin d'assurer l'uniformité du processus de mise à jour de la base de données territoriale à travers le pays, des systèmes automatisés de réseaux informatiques ainsi que des manuels d'exploitation détaillés ont été utilisés. Des programmes de formation régionaux et locaux ont été mis sur pied. Par ailleurs, les coordinateurs techniques ont réalisé des activités de supervision. L'évolution du projet a été contrôlée conformément au calendrier établi au moyen d'un système de contrôle et d'accompagnement de la production.

La mise en place de la base de données territoriale pour le recensement de 2007 ne s'est pas faite sans efforts. Exigeant en main-d'œuvre, ce projet a nécessité un inventaire approfondi des cartes disponibles aux échelles cadastrales et topographiques ainsi que d'autres documents connexes, l'établissement de partenariats avec des tiers pour échanger des documents et la consolidation des données territoriales, la mise à jour et la numérisation des documents cartographiques, la création de plans cadastraux, des municipalités et localités, la production de fichiers numériques relatifs aux limites des districts de recensement, et de cartes compatibles avec des assistants numériques, tout en facilitant la publication des résultats de recensement.

La création de cartes de districts de recensement pour les municipalités/espaces ruraux recourt à des cartes topographiques systématiques produites par et disponibles auprès de l'IBGE ainsi que de la direction du service géographique de l'armée brésilienne. À partir de ces cartes topographiques en guise de données d'entrée, la production est ensuite lancée grâce au système d'élaboration de cartes municipales semi-automatique (SisCart), développé spécialement pour l'IBGE en Visual Basic 6.0 sur la plate-forme graphique MicroStation/MGE de Bentley/Intergraph et l'outil de gestion des données alphanumériques Access97 de Microsoft.

SisCart a grandement facilité l'établissement des cartes municipales. Opérant de manière décentralisée, ce système propose des fonctions simplifiant plusieurs tâches importantes, dont l'homogénéisation de l'échelle et la projection des cartes, le géocodage des feuilles de carte topographique comportant la carte municipale, l'ajustement d'objets entre des feuilles assemblées, le cadrage de feuilles suivant le périmètre des municipalités et l'édition de données-cadres et de bas de page.

La création de cartes de districts de recensement pour les zones urbaines fait appel à des plans cadastraux à des échelles comprises entre 1:2 000 et 1:10 000. Ces cartes sont le fruit d'agences gouvernementales locales et d'État, d'entreprises du service public et d'autres sources. Les plans urbains ont été produits sur la plate-forme MicroStation (totalement personnalisée pour satisfaire aux exigences de l'IBGE). Le système peut gérer toutes les transformations requises pour ingérer les données cartographiques des sources les plus diverses et coordonner les systèmes. Il peut également contribuer aux tâches de mise à jour des cartes alimentées en données à la fois par le travail sur le terrain et les activités des bureaux.

Le recensement réalisé par le Brésil en 2007 a apporté une innovation majeure : le registre national des adresses à des fins statistiques, développé à l'origine à partir des données des districts de recensement de 2000 et mis à jour grâce au travail d'investigation effectué dans le cadre du recensement de 2007. Les innovations technologiques introduites en 2007, notamment l'utilisation d'assistants numériques avec GPS intégré, ont permis à l'IBGE de créer de nouveaux produits de la base de données territoriale, tels que :

- Des cartes de 70 085 districts de recensement ruraux et 92 685 districts de recensement urbains au format PDF.
- Une description des districts de recensement ruraux et urbains au format PDF;
- Des cartes de districts de recensement ruraux au format JPG;
- 70 085 images géocodées Google Earth;
- Des limites numériques de districts de recensement/municipalités au format vectoriel shape-file, englobant les périmètres urbains ainsi que les zones urbaines isolées dans les 27 unités de fédération, avec environ 77 000 polygones.

Tous les éléments liés à la base de données territoriale ont abouti aux recenseurs dans les bureaux locaux et les postes de collecte, les points de convergence au début et à la fin des activités de collecte de données sur le terrain. Les cartes des districts de recensement de 2007 ont été diffusées au format numérique par le biais d'assistants numériques et rendues sur les cahiers du recenseur, avec les descriptions des périmètres des districts de recensement.

La manipulation d'un tel volume d'informations généré par la collecte de données de recensement a de tout temps été source de préoccupation pour les départements bases de données territoriales. Grâce à l'adoption des assistants numériques, les données de terrain à afficher sur les cartes ont pu être transmises et stockées sur les serveurs de l'IBGE en vue d'autres traitements et analyses et pour les intégrer dans les nouvelles éditions de cartes, enlevant de cette façon une épine du pied des départements bases de données territoriales. Au cours de la collecte, les modifications apportées aux cartes imprimées ont été relevées par les recenseurs, qui ont également apporté leur contribution en analysant la pertinence des mises à jour attendues de SisCart.

Les coûts de l'actualisation de la base de données territoriale dans le cadre du recensement de 2007, et notamment leur impact sur le budget total de l'activité de recensement, sont détaillés dans le tableau ci-après.

Coûts du projet de mise à jour de la base de données territoriale					
Année	Mise à jour de la base de données	Recensement de 2007	Mise à jour de la base de données	Recensement de 2007	Pourcentage pour la mise à niveau de la base de données
	Reais brésiliens		Dollars des États-Unis. ^a		
2004	10 774 885,19	–	3 468 273,47	–	
2006	4 041 240,44	179 200 904,00	1 867 917,93	82 829 167,55	
2007	834 380,58	428 919 454,00	433 354,41	222 769 011,11	
2008	–	1 457 000,00	–	833 292,09	
Total	15 650 506,21	609 577 358,00	5 769 545,81	306 433 470,76	1,88

^a Valeurs de référence : Taux de change du réal brésilien (R\$) en dollar des États-Unis au 30 juin de chaque année : 2008, 1,7443 US\$; 2007, 1,9254 US\$; 2006, 2,1635 US\$; et 2004, 3,10670 US\$.

Le personnel temporaire a été recruté à travers la passation de marchés publics, organisée par des sociétés de services de sous-traitance et dans le respect de la législation régissant le recrutement de personnel relevant du gouvernement fédéral. En 2004, 500 assistants en cartographie et levés, qui travaillaient dans des bureaux locaux de l'IBGE dans les départements bases de données territoriales et aux sièges de la DGC, ont été embauchés pour exécuter plusieurs étapes du processus de mise à jour de la base de données territoriale.

Approximativement 68 000 agents recenseurs et 18 000 superviseurs du recensement ont été recrutés en 2007 pour la collecte de données. Tous les agents recenseurs ont bénéficié d'une formation sur le terrain et de cours sur des concepts de la base de données territoriale, sur les caractéristiques pertinentes de chaque district de recensement et sur les procédures de révision de la base de données territoriale (pour en savoir plus, contacter : Rafael Castaneda : rafael.march@ibge.gov.br).

B. Le rôle des cartes dans le recensement

2.12. Depuis des siècles, les hommes ont utilisé les cartes pour représenter leur environnement. Les cartes servent à localiser les détails et à préciser les distances, les directions et la superficie des zones. Elles révèlent aussi les relations géographiques, les différences, les ensembles et les constantes. Les cartes sont utilisées à des fins de navigation, d'exploration, d'illustration et de communication dans les secteurs public et privé. La quasi-totalité des secteurs de la recherche scientifique utilise des cartes sous une forme ou une autre. Les cartes sont donc un outil indispensable pour de nombreux types d'activités professionnelles et théoriques.

2.13. La cartographie fait depuis longtemps partie intégrante des opérations de recensement. Traditionnellement, on se servait de cartes lors des recensements pour faciliter le dénombrement et présenter sous forme cartographique les résultats regroupés du recensement. Au cours des derniers recensements, très peu de dénombrements ont été réalisés sans que l'on ait eu recours à des cartes détaillées.

2.14. En général, la cartographie est utilisée à différentes fins lors du recensement :

- a) *Les cartes garantissent la couverture et facilitent les opérations de recensement (avant le dénombrement).* Le bureau de recensement doit veiller à ce que chaque ménage et chaque habitant du pays soit compté, et qu'aucun d'eux ne soit compté deux fois. À cette fin, les membres du personnel cartographique subdivisent le territoire national en petites unités déclarantes. Les cartes présentant des districts de recensement sont ainsi un instrument de contrôle essentiel qui garantit la couverture du recensement;
- b) *Les cartes facilitent la collecte des données et peuvent aider à contrôler les activités du recensement (pendant le dénombrement)* Pendant le recensement, les cartes permettent aux agents chargés du dénombrement de localiser facilement les zones géographiques qui leur sont attribuées, dans lesquelles ils recenseront les ménages. Des cartes sont également distribuées aux responsables du recensement assignés aux agents de recensement pour faciliter les activités de préparation et de contrôle. Elles peuvent aussi être utilisées pour contrôler la marche des opérations de recensement, ce qui permet aux responsables de planifier stratégiquement, d'indiquer des affectations, de localiser les secteurs qui donnent lieu à des difficultés et de prendre rapidement des mesures correctrices;

- c) *Les cartes facilitent la présentation, l'analyse et la diffusion des résultats du recensement (après le dénombrement).* La présentation cartographique des résultats d'un recensement constitue un moyen efficace pour donner une idée des résultats d'un recensement. Elle aide à définir les caractéristiques locales des indicateurs démographiques et sociaux importants. Ainsi les cartes font partie intégrante de l'analyse des grandes orientations dans les secteurs public et privé.

C. Des cartes aux bases de données géographiques : la révolution cartographique continue

2.15. Aujourd'hui, les cartes ne sont que l'une des formes de présentation d'informations modulées sous le terme général d'informations géographiques; le plus souvent, ces données géographiques prennent la forme de base de données géoréférencée (ou base de données géographique). La cartographie a été touchée par la révolution informatique un peu plus tard que les autres secteurs. Les premiers ordinateurs emmagasinaient correctement des chiffres et du texte. Mais les cartes sont plus compliquées et la cartographie numérique exige une grande capacité de stockage de données et une grande vitesse de calcul. De plus, la cartographie est essentiellement une application graphique et les premiers ordinateurs n'avaient que des capacités graphiques limitées. Les premières applications graphiques réalisées sur des ordinateurs au cours des années 1960 ne trouvèrent donc guère d'applications au-delà de quelques projets officiels et universitaires. Il fallut attendre les années 1980 pour que les systèmes d'information géographique commerciaux aient des capacités suffisantes justifiant leur adoption rapide, par exemple, par les administrations locales et régionales, les services d'urbanisme, les agences de protection de l'environnement, les services d'exploration de minéraux, les services publics de distribution, les entreprises commerciales de distribution et les agences immobilières. Cela a influencé l'utilisation initiale de technologies cartographiques par les bureaux de statistique nationaux.

2.16. De nouvelles sources d'information raccourcissent aussi les délais entre planification du projet et établissement d'une base de données opérationnelle. Les perfectionnements récents les plus importants ont porté sur la navigation, la télédétection, l'analyse des images, la manipulation des données et la cartographie par Internet. Le GPS a révolutionné la collecte de données sur le terrain dans divers domaines, de la géodésie à la surveillance de l'environnement et à la gestion des transports. Une nouvelle génération de satellites commerciaux à haute résolution peut donner des images de la quasi-totalité de la surface terrestre avec suffisamment de détails pour prendre en charge de nombreuses applications de cartographie, dont celles utilisées pour les recensements. Le coût de la cartographie numérique de précision a fortement diminué du fait de l'intégration étroite des techniques d'utilisation du GPS et de la saisie de données à distance.

2.17. Les SIG ont tiré de grands avantages des progrès survenus dans divers secteurs de l'informatique. L'amélioration des logiciels de gestion de base de données permet de gérer de grands volumes d'information se référant à des cartes numériques. Les utilisateurs de données censitaires développent des modèles de données pour le stockage, la recherche et l'affichage d'objets géographiques. Des techniques élaborées de visualisation permettent de créer des représentations de plus en plus complexes de notre environnement. Les fonctions d'affichage de données de SIG vont bien au-delà de l'affichage statique à deux dimensions et offrent des possibilités d'animation et de modélisation tridimensionnelle. De même que la saisie de texte

facilitée par la reconnaissance optique de caractères, l'exploration rapide par balayage à haute résolution et des logiciels évolués accélèrent la conversion de données cartographiques, opérée jusque-là exclusivement par numérisation manuelle. Cette technologie habilitante a permis aux producteurs de données de passer d'une base de cartes sur papier à une base numérique, en mettant à jour les informations existantes. Des renseignements sont cependant encore nécessaires pour apporter une valeur ajoutée aux produits et services ultérieurs rendus possibles grâce à l'adoption de cette technologie.

2.18. On constate des progrès comparables dans le secteur de la diffusion des données géographiques. Tous les grands distributeurs de SIG fournissent aujourd'hui les instruments nécessaires à la création de bases de données géographiques ou géospatiales accessibles via Internet. Des organismes publics à tous les niveaux adoptent cette technique pour permettre au public d'accéder rapidement et aux moindres frais à de très grands volumes d'informations spatiales. Internet devrait remplacer les cartes imprimées et les supports numériques comme principal moyen de diffusion de données. De ce fait, les cartes elles-mêmes ne sont plus des objets statiques, mais des instantanés dynamiques d'une base de données géographique en mutation.

2.19. Les programmes cartographiques disponibles sur Internet montrent que les instruments disponibles pour exploiter les informations spatiales numériques sont de moins en moins coûteux et de plus en plus faciles à utiliser. Les logiciels de SIG de haut de gamme exigent encore une formation poussée pour pouvoir être utilisés efficacement, mais les programmes de cartographie numérique ne sont pas plus compliqués à utiliser que les logiciels commerciaux courants. La cartographie numérique est de plus en plus étroitement intégrée à des applications informatiques classiques comme les tableurs, grapheurs ou logiciels de gestion commerciale.

2.20. Les services de statistique ont été parmi les premiers à adopter les SIG. Les statistiques démographiques, sociales et économiques sont à la base de la planification et de la gestion administrative. La répartition spatiale des indicateurs socio-économiques guide les décisions politiques relatives au développement régional, aux prestations de services et à beaucoup d'autres questions. Les techniques numériques permettent de mieux gérer, de trouver plus rapidement ces données et d'en améliorer la présentation. Il a donc toujours existé un lien étroit entre la géographie et les statistiques — confirmé par le fait que dans de nombreux pays d'Amérique latine service cartographique et service de statistique sont logés sous le même toit. Cette intégration étroite des SIG et des applications statistiques présente de grands avantages pour les services nationaux de statistique, car elle diminue le coût et les délais requis pour collecter, grouper et diffuser les informations. Les SIG permettent au bureau de statistique de fournir un grand nombre de services, ce qui augmente beaucoup le rendement des investissements consacrés à la collecte de données.

2.21. L'automatisation de la cartographie, les SIG et d'autres outils géospatiaux ont permis d'établir plus efficacement les cartes utilisées par les agents de recensement et les cartes thématiques des résultats de recensement. En outre, les progrès technologiques et les nouvelles fonctions des SIG utilisant de nouvelles sources de données, telles que la télédétection et l'enregistrement en fonction du lieu avec GPS, ont donné davantage de place à la représentation géographique au sein des services nationaux de statistique. Néanmoins, les nouvelles technologies doivent être utilisées de manière judicieuse, à la lumière des objectifs stratégiques déterminés par les responsables de l'organisation de recensement. Loin d'être une balle magique pouvant résoudre tous les problèmes organisationnels, elles doivent être planifiées avec soin.

D. Augmentation de la demande en données statistiques ventilées

2.22. Les avantages procurés par l'informatisation des données géographiques dans le domaine statistique sont partagés par les utilisateurs de données du recensement et de levés topographiques. Les fonctions d'intégration de données des SIG, qui permettent de confronter des données provenant de nombreux secteurs différents, ont beaucoup élargi l'utilisation des informations statistiques, ce qui a incité les services de statistique à préparer des informations géoréférencées de haute qualité pour de petites unités géographiques. Grâce à une planification méthodique (une collecte d'informations en petites unités par le service de statistique avant de les regrouper comme il se doit), il devrait être possible de satisfaire les besoins de nombreux nouveaux consommateurs de données. Voici quelques exemples d'utilisation de ces données :

- a) **Planification des interventions d'urgence et réponse humanitaire.** Les agences peuvent se préparer à l'éventualité de catastrophes naturelles en repérant des zones densément peuplées et difficiles à évacuer en cas d'incendies, de séismes, d'éruptions volcaniques ou de tsunamis. Après une catastrophe naturelle majeure, les premières questions posées sont : Quels villages ont été touchés ? Quelle est la taille de leur population ? Quel est le nombre de morts, de blessés et de sans-abris ? Quel est l'état des infrastructures, en particulier des routes et des ponts, des centres de santé, des systèmes d'approvisionnement en eau, des écoles et des bâtiments publics ? Si les cartes numériques renseignant la répartition de la population et les caractéristiques de l'habitat pouvaient être enrichies de données relatives à l'altitude et aux pentes, aux réseaux de transport et à d'autres informations géographiques sur la zone touchée par la catastrophe, on pourrait obtenir des estimations fiables sur le nombre de personnes concernées, leurs besoins en assistance médicale, en nourriture et abris et surtout sur leur emplacement. Les « codes-p » pour les zones peuplées facilitent la localisation des zones touchées ayant besoin d'une assistance;
- b) **Modélisation des plaines inondables.** Les inondations de grande envergure paraissent constituer dans le monde entier un risque croissant dans de nombreux bassins hydrographiques. Les données numériques hypsométriques et hydrologiques combinées aux statistiques de recensement relatives à de petites zones permettent aux planificateurs de procéder à des évaluations détaillées en vue de réduire les risques pour les populations dans les zones exposées aux inondations et d'organiser la gestion des interventions d'urgence. Dans certains pays, les compagnies d'assurance utilisent les mêmes instruments pour évaluer les niveaux de risque pour les propriétaires, ce qui permet une évaluation plus équitable des primes;
- c) **Planification des services sociaux et d'enseignement.** Une des principales tâches des autorités locales et régionales est de veiller à ce que toutes les parties du pays aient également accès à des services publics tels que la santé et l'éducation. Les données de recensement sur l'âge et les caractéristiques sociales de la population de petites zones permettent aux planificateurs de prévoir la demande concernant divers services. Associées aux données de SIG relatives à l'infrastructure des transports, ces informations permettent de mieux répartir les ressources entre les centres de service existants et de prendre des décisions plus rationnelles concernant l'implantation de nouveaux aménagements;

- d) **Analyse de la pauvreté.** Dans les pays où les données relatives au revenu ou à la consommation ne sont pas collectées lors d'un recensement, les caractéristiques des ménages sont un indicateur important du bien-être des divers groupes de population. Les données du recensement relatives à de petites zones, combinées à des informations géoréférencées concernant l'infrastructure et la situation agroécologique, peuvent être utilisées pour évaluer l'incidence de la pauvreté et localiser les communautés défavorisées. Ces renseignements permettent de mieux cibler les programmes de lutte contre la pauvreté en affectant les ressources aux zones les plus pauvres tout en évitant l'affectation de subventions à des communautés non défavorisées;
- e) **Planification des services publics de distribution.** Les entreprises publiques et privées de distribution d'eau, de gaz, d'électricité et de télécommunications ont recours aux SIG pour gérer leurs infrastructures, mais elles utilisent aussi l'analyse spatiale des données démographiques pour évaluer la demande actuelle et future de services. Les données du recensement numérique — complétées par les modèles de terrain numériques — ont joué un rôle essentiel dans la conception des systèmes fondés sur l'emplacement dans le monde entier;
- f) **Analyse de la main-d'œuvre.** Qu'il s'agisse d'une entreprise privée recherchant un site approprié pour implanter une usine ou d'un organisme public qui s'efforce d'harmoniser l'offre et la demande de main-d'œuvre, les données de recensement relatives à une petite zone sont un élément important de l'analyse relative à l'emploi. L'analyse du déplacement domicile-travail, qui met en rapport le lieu du travail et le lieu de résidence des employés, joue un rôle essentiel dans la planification des transports;
- g) **Analyse de marché.** Les entreprises utilisent les données de recensement relatives à de petites zones pour préparer l'implantation de nouveaux magasins et entrepôts, gérer les données nécessaires au service après-vente et cibler la publicité. Ainsi est apparu un nouveau secteur d'application des SIG — qualifié de géographie commerciale ou géodémographie. De fait, la forte demande de ce type d'analyse a été un puissant facteur de développement de logiciels de cartographie bon marché et faciles à utiliser pour ordinateur de bureau;
- h) **Délimitation des circonscriptions électorales.** Dans les démocraties représentatives, la représentation parlementaire se fonde sur le principe de l'égalité des votes. Pour garantir l'application de ce principe, on utilise les données démographiques des petites zones pour délimiter des circonscriptions électorales de dimensions approximativement égales. En fait, aux États-Unis d'Amérique, elles constituent la base essentielle du recensement décennal requis par la Constitution, les SIG et les données de recensement étant utilisées pour définir les limites des circonscriptions électorales;
- i) **Analyse épidémiologique.** Les données de recensement de petites zones, combinées aux données relatives à l'incidence des maladies et à des données biophysiques, permettent aux responsables des services sanitaires d'évaluer la population exposée à certaines maladies infectieuses ou à vecteur. Connaissant le nombre d'habitants du pays susceptibles d'être atteints par le paludisme ou la bilharziose, par exemple, les planificateurs peuvent évaluer les ressources requises pour prendre des mesures d'éradication. Le

repérage des zones habitées par ces groupes à risque aide à définir les mesures d'intervention par ordre de priorité et à les mettre en œuvre;

- j) **Agriculture.** Les informations géographiques relatives à la situation agro-écologique et les données de production combinées aux données de petites zones concernant la demande de produits alimentaires facilitent l'analyse des problèmes de sécurité alimentaire. Des systèmes d'alerte précoce à la pénurie alimentaire ont été mis en place dans de nombreux pays caractérisés par des écosystèmes fragiles en vue de prévenir de graves crises alimentaires.

E. Coûts et avantages d'un investissement dans la technologie géospatiale

2.23. La présente section concerne les coûts en jeu et les avantages potentiels procurés par l'adoption d'une technologie géospatiale pour les recensements. Il existe une gamme de solutions, allant d'une infrastructure géospatiale interne parfaitement intégrée à l'utilisation, par exemple, de la cartographie sur ordinateur individuel utilisée uniquement pour présenter et diffuser les résultats. En d'autres termes, il n'existe pas de solution universelle pour appliquer la technologie géospatiale au travail de recensement. Cependant, compte tenu des innovations de pointe et de la baisse des coûts à l'entrée, la conversion est certainement à portée de main à ce stade; en effet, les technologies GPS deviennent abordables et les systèmes d'information géographique haut de gamme complexes peuvent être installés sur des ordinateurs personnels peu coûteux. L'adaptation à la tâche est le principe prioritaire de toute analyse coûts-avantages.

2.24. Comme nous l'avons souligné, l'adoption de toute nouvelle technologie pose des défis pour les organisations existantes. Ce choix ne doit être retenu qu'après une analyse minutieuse des coûts et des avantages. Les éléments de coûts sont entre autres liés à la conception des systèmes, à l'acquisition de matériel et de logiciels, au développement d'un prototype, à la planification des ressources humaines, aux formations, à l'appui contractuel, à la conception de bases de données géographiques, aux frais de transition, à l'acquisition de données, à la saisie de données, au contrôle de la qualité, à la maintenance des systèmes et au développement de produits. Le passage aux données géospatiales est assorti de coûts associés à la numérisation, à savoir la conversion des cartes de districts de recensement analogiques aux formats géospatiaux.

2.25. La migration vers des systèmes géospatiaux aux fins des recensements offre deux types d'avantages : les avantages d'efficacité et les avantages d'utilité. Les premiers sont liés à la quantité de résultats pouvant être obtenus par unité d'intrants. En termes de recensement, cela signifie que l'on peut en faire plus à moindre coût. Les avantages d'efficacité englobent des réductions de coûts ou des gains de productivité, une économie de temps, une plus grande crédibilité et autorité des produits de données géospatiaux, un meilleur service, une plus grande exactitude, une meilleure cohérence et la production d'un revenu. Sous réserve d'une planification adéquate, l'organisme de recensement lui-même peut en bénéficier.

2.26. L'utilité fait quant à elle référence aux effets des politiques ou des programmes qui bénéficient d'informations améliorées, y compris les avantages sociaux dont jouissent les utilisateurs de données statistiques générées par les recensements. L'utilité a pour avantage de garantir notamment une meilleure analyse à une échelle

mieux adaptée aux études locales et régionales, une élaboration de politiques plus éclairée, un partage plus vaste des données avec d'autres agences gouvernementales ou organisations non gouvernementales ainsi qu'une meilleure diffusion auprès du public. Certains avantages sociaux, comme une intervention humanitaire et une planification des infrastructures de santé, peuvent être déduits des exemples présentés dans la section précédente. En fait, une utilisation efficace des technologies géospatiales peut sauver des vies et améliorer le bien-être général des êtres humains.

2.27. Pour différentes raisons, il est aussi difficile d'évaluer les coûts et avantages de l'utilisation des SIG sur une base quantitative. Par exemple, il se peut que de nombreux avantages profitent non à l'agence qui a investi dans les SIG, mais à des tiers qui accéderont à des produits moins coûteux ou d'une plus grande exactitude ou qui pourront obtenir des produits qui jusque-là n'étaient absolument pas disponibles. Cela illustre bien la différence entre un produit bon marché et un produit rentable. La solution la moins coûteuse pour établir des cartes de recensement peut encore consister dans l'approche manuelle traditionnelle, surtout dans les pays où le coût de la main-d'œuvre est faible. Mais du point de vue social, il peut s'avérer plus rentable d'investir davantage au départ dans une approche numérique, car les produits numériques seront beaucoup plus avantageux à long terme aussi bien à l'intérieur qu'à l'extérieur du service de recensement ou du service de statistique, donnant ainsi lieu à une véritable initiative nationale. Par ailleurs, rien ne garantit que les coûts de main-d'œuvre resteront faibles.

Encadré II.2

Obstacles technologiques et financiers

1. Cambodge

Dans le cadre de son enquête de santé de 2003, le Cambodge a procédé à une cartographie des districts de recensement et des limites des villages. Les activités de SIG de recensement relèvent du Département des statistiques démographiques de l'Institut national des statistiques. La cartographie censitaire a été coordonnée avec le ministère de gestion des terres et le ministère de l'intérieur. 30 personnes ont été chargées de délimiter des districts de recensement en unités de maximum 120 foyers. Un questionnaire sur les villages a également été utilisé pour élaborer une base de données de districts de recensement. Cette base de données proposait des fonctions permettant entre autres de libeller, de contrôler, de coder, d'éditer, de saisir ainsi que de vérifier et nettoyer les données relatives aux limites. Le suivi de cartes de districts de recensement a été facilité par l'utilisation de codes-barres. Pour son recensement de 2008, l'institut national des statistiques envisage de diffuser les résultats en ligne et de recourir pour ce faire à un système CamInfo basé sur DevInfo. Le Cambodge estime le coût de la cartographie et de la délimitation des districts de recensement, associé à celui du traitement des données, approximativement à 15 % du coût total du recensement. Du point de vue des ressources, les véhicules de terrain et les ordinateurs (ainsi que les dispositifs périphériques) s'avèrent particulièrement coûteux, bien qu'ils soient souvent mis à disposition par les donateurs.

Source : Présentation du Cambodge, atelier de Bangkok, 2007.

2. Lesotho

Le Lesotho a effectué son dernier recensement de la population et du logement en 2006. La cartographie censitaire préalable au dénombrement a été effectuée en recourant aux dernières technologies de télédétection et de SIG. Citons notamment l'imagerie satellitale en couleurs naturelles SPOT 2,5 m à couverture totale, le GPS et les SIG. Cinq membres du personnel ont effectué le travail administratif préparatoire dont la délimitation préliminaire des districts de recensement à travers une interprétation des images (et des connaissances locales) et la préparation de cartes d'images

satellitaires en vue du travail de démarcation des districts de recensement. Ce dernier a été réalisé par 8 équipes de 4 travailleurs sur le terrain. Chacune d'entre elles disposait d'un véhicule et d'un GPS. Les limites des districts de recensement ont été vérifiées et modifiées au besoin. Un complément d'information sur les attributs, tels que les noms de lieux et les noms des responsables locaux, a été recueilli. Les informations glanées sur le terrain ont été saisies au format numérique et les cartes des zones de supervision et des districts de recensement ont en fin de compte été créées et imprimées à l'aide du SIG. Au total, 4 500 districts de recensement ont été délimités. La cartographie censitaire préalable au dénombrement a été effectuée en l'espace de 9 mois. Elle a monopolisé approximativement 5 % du budget total du recensement. Près de 40 % du budget de cette cartographie censitaire ont été utilisés pour l'imagerie satellitaire. Ce chiffre est important mais cela a permis d'économiser plus de 60 % du travail sur le terrain puisque les croquis cartographiques individuels pour chaque district de recensement étaient désormais superflus. En fin de compte, cette approche moderne a amélioré la qualité du produit, en permettant une économie d'environ 20 % par rapport au budget initial prévu pour l'approche des croquis cartographiques.

Source : présentation Geospace Inc., atelier de Lusaka, 2007.

2.28. Les investissements budgétaires consacrés aux SIG sont surtout importants pendant la phase initiale. Cela signifie qu'il faut engager les dépenses les plus élevées dès le début du projet mais que les avantages tangibles risquent de n'apparaître que beaucoup plus tard au cours de l'exécution du projet. En confrontant les coûts et les avantages de l'approche cartographique traditionnelle et ceux de la cartographie numérique, une différence saute aux yeux : dans le premier cas, les cartes sont établies à la main pour chaque recensement. Il en a résulté des doubles emplois absolument inutiles. Par le passé, la cartographie censitaire a trop souvent été fondée uniquement sur un projet. Quelques années avant le recensement, une équipe était formée pour établir rapidement à la main des cartes croquis utilisées uniquement pour le dénombrement. Plusieurs années plus tard, l'opération était recommencée pour le recensement suivant.

2.29. Développer une stratégie à long terme pour l'adoption de la technologie géospatiale contraint à entrevoir la cartographie censitaire et le développement de bases de données géographiques comme un processus continu impliquant une gestion régulière des bases de données par un noyau de personnel permanent régulièrement formé. Dans le cas de l'analogique, les coûts ont tendance à dépasser les avantages, les cartes originales n'étant utiles qu'aux fins du recensement. Dans le cas du numérique, un investissement élevé au départ se traduit par des coûts de maintenance et de mise à jour plus bas et par des avantages durables à long terme. Les avantages à long terme sont sensiblement plus élevés car l'opération permet de créer une base de données numérique à des fins multiples pour le bureau de statistique national et d'autres agences gouvernementales.

F. Facteurs déterminant le succès d'une mise en œuvre de technologies géospatiales dans le bureau de statistique national

2.30. Au-delà des coûts que l'on peut évidemment calculer pour une application géospatiale donnée dans le cadre du recensement, divers obstacles peuvent empêcher un projet d'aboutir ou de donner tous les résultats que l'on peut en attendre. Pour la plupart, ces problèmes sont liés au manque de préparation, au choix de matériel et de logiciels inappropriés et à diverses erreurs d'organisation. L'analyse de projets de SIG

effectivement réalisés peut révéler un ensemble de caractéristiques communes aux SIG mis en œuvre avec succès. La liste suivante de facteurs déterminants, établie par ordre de priorité, est adaptée spécifiquement aux travaux des services nationaux de statistique :

1. Planification stratégique, opérationnelle et de gestion détaillée fondée sur une évaluation réaliste des coûts et des opérations en cause, avec des buts et objectifs bien définis pour les opérations géospatiales. Procédures consacrées de contrôle/assurance de qualité.
2. Décision d'investir dans une technologie géospatiale en raison des besoins et des problèmes rencontrés et non pour des motifs techniques. Technologie considérée non comme un complément indépendant, mais comme un élément faisant partie intégrante de la stratégie d'ensemble de gestion de l'information.
3. Rôle essentiel de la personne (ou « champion ») préconisant le développement géospatial dans le cadre de l'organisme intéressé et appui du personnel de direction à un niveau élevé.
4. Sur le plan des ressources humaines, capacité à fournir des formations et un soutien aux employés et cadres concernés. Capacité à rencontrer les besoins en personnel, et notamment de conserver le personnel qualifié et d'embaucher des fournisseurs au besoin. Contrats écrits bien définis avec les vendeurs, consultants, partenaires et clients dans le cadre des organismes publics ou en dehors de ceux-ci.
5. Évaluation des besoins des utilisateurs et définition à l'avance des produits à obtenir avec un calendrier précis pour l'exécution du projet. Programme de financement à long terme bien défini et prévoyant des stratégies de recouvrement des coûts et de facturation des données, avec des évaluations exactes des coûts de maintenance et des coûts connexes. Échelonnement fréquent des étapes et des fournitures de produits afin d'encourager le respect de délais préétablis.
6. Élaboration d'accords de coopération avec d'autres parties intéressées, y compris des contrats en matière d'infrastructures de données spatiales. Facilitation du partage des données grâce à une documentation et des métadonnées détaillées sur les réseaux utilisateur.
7. Facilité d'utilisation des systèmes et produits.
8. Respect des normes géographiques.
9. Utilisation de méthodologies d'intégration des données.
10. Utilisation de protocoles clairs, par exemple avec le GPS pour la collecte et le traitement des données.

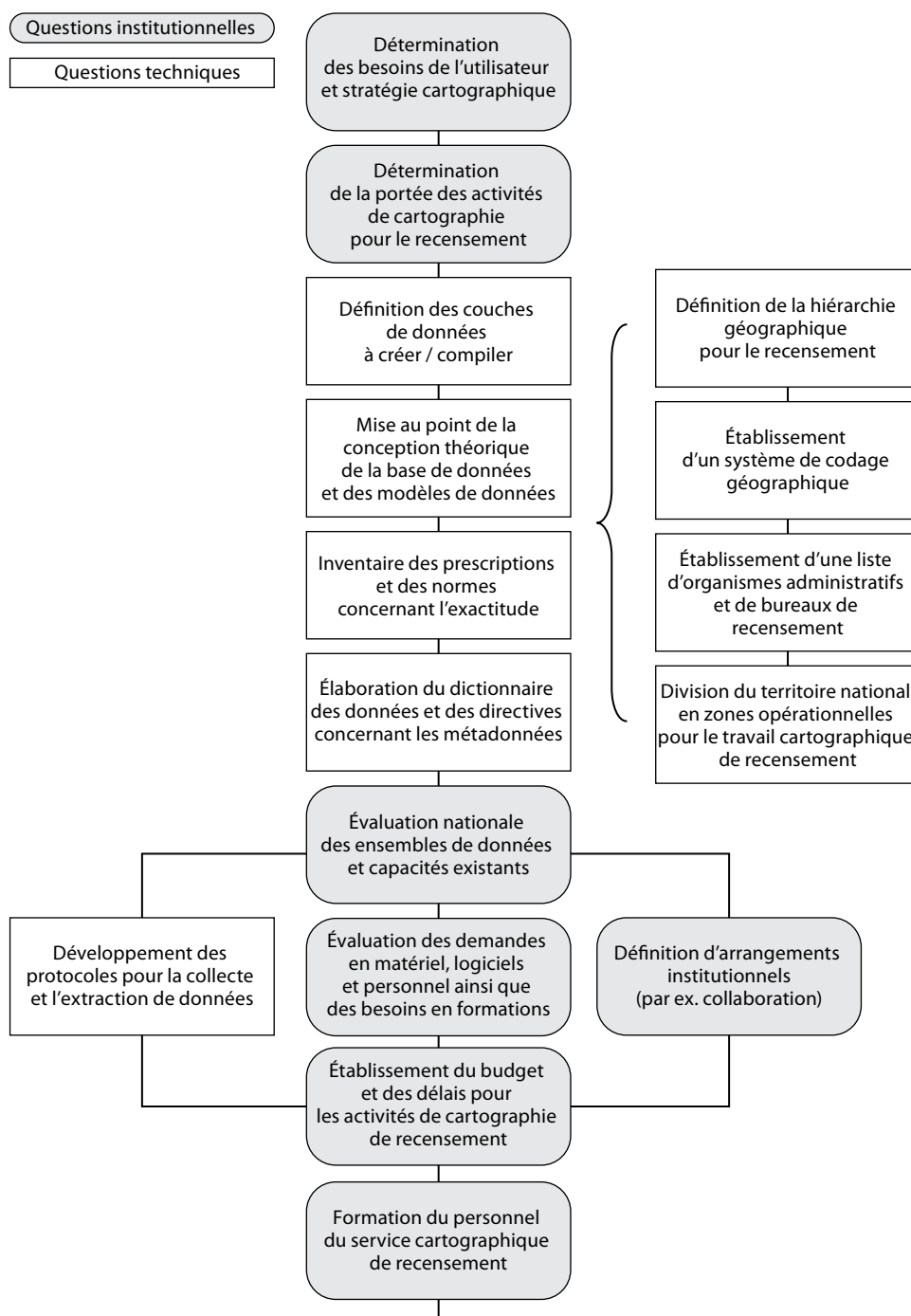
G. Planification des opérations de recensement à l'aide d'outils géospatiaux

2.31. La présente section concerne les tâches d'organisation préliminaires d'un projet d'élaboration de cartes de recensement et des questions essentielles de conception qui déterminent la nature des bases de données obtenues et par conséquent l'éventail d'applications que le projet prend en charge. Le succès du processus de conversion de données proprement dit dépend d'une conception appropriée de l'environnement institutionnel et d'une stratégie opérationnelle bien préparée. On distinguera ici les étapes de planification suivantes : questions institutionnelles telles que la structure organisationnelle pour un soutien géographique (dont le recrutement du personnel et

la coopération avec d'autres institutions), définition de la géographie du recensement et conception de la base de données géospatiale. Comme l'indique la figure II.1, ces étapes peuvent se dérouler plus ou moins simultanément à l'aide de méthodologies approuvées sur le plan institutionnel; dans de nombreux cas, les choix dépendent aussi de la stratégie de saisie de données retenue.

Figure II.1

Étapes de planification des travaux cartographiques de recensement



(continuité avec la figure III.5)

H. Évaluation des besoins et définition des solutions géographiques

1. Évaluation des besoins des utilisateurs

2.32. L'une des premières étapes d'un projet de cartographie de recensement consiste dans l'évaluation détaillée des besoins, suivie de l'analyse des solutions géographiques possibles. Le service chargé d'établir les cartes de recensement doit faire coïncider les attentes de l'utilisateur et ce qui est réalisable compte tenu des ressources disponibles, ce qui implique de remonter des produits et des services finaux jusqu'aux exigences.

2.33. Pour donner de bons résultats, une opération de planification de recensement nécessite des consultations poussées avec les principaux utilisateurs des informations qui seront récoltées au cours du recensement. Ce processus doit passer par une consultation axée sur le *contenu géographique*, à savoir les structures géographiques, en ce compris les hiérarchies administratives ou les niveaux sommaires, et également sur les *produits géographiques de base* à l'appui de l'analyse des données censitaires. Ce travail devrait être intégré au programme général de consultation pour le recensement. Au fur et à mesure que la demande de données de recensement spatialement référencées augmente, les consultations relatives aux produits cartographiques se verront accorder un rôle plus important. Ainsi, les institutions qui utilisent des cartes statistiques devraient être incluses dans les groupes consultatifs qui apportent leur contribution au processus de préparation du recensement. Le bureau de recensement doit engager des consultations avec trois groupes principaux au cours des étapes de préparation :

- a) **Les personnes et institutions qui participent aux opérations de recensement.** Afin d'être pleinement renseigné sur les ressources et les goulets d'étranglement éventuels, le service de recensement devra procéder à une enquête approfondie sur les ressources humaines disponibles dans le pays, l'équipement disponible et utilisable, les produits cartographiques numériques et analogiques et les activités cartographiques actuelles ou prévues d'autres organismes publics et privés. Éviter les activités en double est essentiel pour réduire le coût de la cartographie numérique et fournir en temps utile les données cartographiques de recensement;
- b) **Les utilisateurs de logiciels géographiques de recensement.** Ils proviendront essentiellement d'autres départements du gouvernement, de la communauté des chercheurs universitaires et du secteur privé;
- c) **Le grand public.** En fait, compte tenu de l'accès aux ordinateurs et de solutions cartographiques offertes par Internet, les utilisateurs privés devraient devenir un groupe d'utilisateurs important. Il se pourrait par exemple que les citoyens veuillent se procurer des informations statistiques sur leur quartier ou sur le quartier où ils ont l'intention de s'établir. Aujourd'hui, en raison de l'évolution rapide des techniques, le bureau de recensement doit établir avec soin des plans pour anticiper la demande de données.

2. Détermination des données de recensement

2.34. Les besoins des utilisateurs détermineront la gamme de produits à obtenir à la fin du cycle des opérations cartographiques de recensement. Les produits obtenus par le système cartographique de recensement, examinés plus en détail au chapitre VI, doivent toujours inclure une documentation adéquate, notamment sur le codage et

les métadonnées, de manière à ce que les utilisateurs puissent en tirer le meilleur parti. Voici quelques exemples:

- Un ensemble de cartes numériques des districts de recensement ou des unités de diffusion dérivées conçues de façon à permettre l'élaboration de tous les produits qui seront diffusés auprès des organismes administratifs et du public.
- Des fichiers de limites géographiques sous forme numérique pour toutes les unités déclarantes utilisées à des fins statistiques, auxquels des indicateurs de recensement seront communiqués sous forme de tableaux.
- Des listes de toutes les unités déclarantes statistiques et administratives, y compris les villes et les villages, leurs allonymes et coordonnées géographiques.
- Des fichiers d'équivalences géographiques indiquant comment les unités déclarantes actuelles se rattachent aux unités utilisées lors des précédents recensements, ou comment un ensemble d'unités déclarantes se rattache à un autre ensemble.
- Des couches vectorisées contenant des données d'objets, tels que des bâtiments, des routes, des écoles, des hôpitaux et des cliniques, utilisables en cas d'analyse spatiale des données démographiques.
- Des listes d'index de rues pour toutes les grandes zones urbaines.
- Des fichiers centroïdes fournissant pour chaque unité déclarante un point de référence géographique représentatif.
- Des nomenclatures toponymiques fournissant des coordonnées géographiques pour toutes les zones d'habitat et autres détails géographiques importants d'un pays.

2.35. Les besoins des utilisateurs sont les plus importants facteurs à prendre en compte pour concevoir la réalisation des cartes de recensement. Mais ils doivent être mis en balance avec les ressources disponibles du cycle budgétaire. Divers autres facteurs déterminent le choix de la stratégie cartographique, notamment les suivants :

- Les ressources financières et humaines disponibles.
- Les produits cartographiques numériques et analogiques actuels.
- Le degré d'intégration entre les services cartographiques et les services de la statistique ou autres agences pertinentes du pays.
- Les capacités techniques du bureau de statistique et des organismes qui collaborent avec lui.
- L'arbitrage entre l'utilisation de la technologie, qui peut exiger des devises étrangères et créer une situation de dépendance à l'égard des technologies extérieures, et un recours accru à une main-d'œuvre utilisant des technologies simples, qui peut stimuler l'économie locale.
- Les dimensions du pays, tant la taille de la démographie que l'étendue aréale, en plus de l'accessibilité des étendues d'eau et de terrain.
- Le temps disponible pour préparer et effectuer les opérations d'établissement des cartes de recensement.

3. Solutions géographiques

2.36. Les divers pays entreprennent leur travail de cartographie de recensement à partir de bases d'informations existantes, de budgets, de capacités techniques et de délais différents dans chaque cas. Il existe donc toute une série d'options pour créer

une base de données cartographiques totalement numérique en vue de collecter et de diffuser les données de recensement. La liste partielle des options disponibles proposée ci-après, par ordre de complexité croissante, varie selon les budgets et les délais :

- Production de cartes numériques créées sur la base de croquis cartographiques existants.
- Couches géoréférencées de districts de recensement ou fichiers spatiaux, avec le codage et les métadonnées adéquats, pouvant être intégrés de façon satisfaisante à d'autres bases de données géographiques numériques.
- Inclusion de couches de référence géographique montrant par exemple les routes, les cours d'eau, les sites, les détails ponctuels et d'autres détails; elles peuvent être incluses sous la forme de simples images tirées de cartes numérisées ou conçues comme une base de données géographique structurée.
- Un registre numérique d'adresses postales où les adresses sont mises automatiquement ou semi-automatiquement en rapport avec les bases de données numériques du réseau routier.
- Une base de données numérique des unités d'habitation localisées avec précision, créée à l'aide de systèmes de positionnement géographique.

2.37. La liste ci-dessus est incomplète et donnée à titre d'exemple. Toutes les questions qui se posent seront examinées en détail dans la suite du présent *Manuel*. La meilleure stratégie de cartographie de recensement pour un pays consiste dans une approche sur mesure, prenant en compte les besoins et les ressources du pays. Le présent *Manuel* examinera la gamme des options techniques et logistiques disponibles. Parmi celles-ci, le bureau de recensement pourra choisir le sous-ensemble de techniques et de procédures correspondant le mieux aux besoins du pays.

4. Ressources humaines et renforcement des capacités

2.38. La disponibilité d'un personnel motivé et qualifié est un facteur essentiel dont dépend le succès ou l'échec d'un projet de cartographie de recensement. Les objectifs d'un projet de cartographie sont semblables, que les cartes soient établies à la main ou à l'aide d'un ordinateur. Mais l'utilisation de l'ordinateur suppose que le personnel chargé d'établir les cartes de recensement possède plusieurs qualifications nouvelles, car des produits semblables sont créés au moyen de techniques différentes. Par ailleurs, une base de données géographique numérique est utile à beaucoup d'autres fins. C'est ainsi qu'un bureau de recensement pourra avoir à satisfaire des demandes supplémentaires de produits et de services qui n'étaient pas disponibles jusque-là. Chaque membre du service cartographique de recensement devra donc posséder un certain niveau de compétence informatique.

2.39. Les compétences requises pour la cartographie de recensement traditionnelle, à la main, sont en grande partie utilisables pour exécuter un projet de cartographie numérique. Il n'y a pas lieu de remplacer complètement les qualifications existantes, mais l'approche de la cartographie numérique demande des connaissances supplémentaires concernant les méthodes informatiques, portant notamment sur les bases de données, les feuilles de calcul, la gestion de fichiers et l'exploitation de base d'un réseau. Seule une part relativement faible des connaissances des cartographes et géographes de l'équipe est dépassée, mais le niveau des qualifications requises par leur emploi a augmenté. Par exemple, les cartographes ayant reçu une formation traditionnelle n'auront plus besoin d'appliquer certaines techniques de cartographie manuelle telles que la transcription toponymique, le tracé négatif ou le dessin à la plume ou au

crayon. Par contre, après avoir reçu une formation informatique, ils pourront utiliser leurs connaissances en matière de dessin de cartes et de communication cartographique pour établir des cartes de district de recensement ou des cartes thématiques bien conçues en utilisant un logiciel de SIG ou de cartographie numérique. Il est souvent plus facile de former aux techniques informatiques un spécialiste d'une discipline que de donner à un expert informatique une formation dans un secteur d'application concret.

2.40. Les paragraphes suivants décrivent en détail le profil des tâches qui exigent le recrutement de personnel dans un projet de cartographie numérique de recensement. Les collaborateurs d'un bureau de recensement pourront accomplir plusieurs des tâches requises au cours des différentes phases du projet de recensement.

2.41. **Planification.** Aux premiers stades du projet, on devra constituer l'équipe qui élaborera la stratégie d'ensemble de la cartographie numérique de recensement. Il faudra pour cela du personnel formé en géographie, en applications informatiques et géospatiales, et ayant l'expérience de la cartographie numérique. Outre le personnel du bureau de recensement, le groupe de planification peut inclure des représentants du service cartographique national et d'autres organismes publics, groupes d'utilisateurs de données ou consultants extérieurs intéressés. Des conseillers techniques d'organismes nationaux de statistique dans des pays déjà passés à la cartographie numérique de recensement ou d'organisations internationales devraient participer au travail de préparation, car ils peuvent apporter des contributions utiles et aider à déterminer le coût des alternatives.

2.42. **Conduite du projet.** Le responsable du projet de cartographie de recensement, qui supervise aussi la mise en œuvre de la stratégie de cartographie numérique de recensement, dirige le travail de planification. Le responsable du projet doit avoir les qualifications requises en géographie, en informatique ou dans une spécialité semblable, ainsi qu'une formation en science de l'information géographique et en cartographie numérique. Une expérience en cartographie du recensement, idéalement glanée dans le cadre d'un précédent recensement dans le pays, est hautement souhaitable. L'expérience de la gestion ou une formation en ce domaine est nécessaire pour superviser l'établissement du budget, la gestion du personnel et l'établissement du calendrier. Des compétences adéquates en communication faciliteront la coopération avec d'autres participants au projet de recensement et avec les organismes qui y collaborent. Le responsable du projet doit aussi se tenir au courant de l'évolution et des tendances dans le domaine des technologies géospatiales et il doit être prêt à adapter la stratégie de cartographie de recensement si la situation évolue ou si de meilleures solutions deviennent disponibles. Il est primordial de ne pas élaborer de stratégie une fois le processus engagé et de ne pas perdre de vue un éventuel « glissement de la mission ». Des mécanismes de contrôle du changement fournissent un outil essentiel pour relever les défis inattendus dans le cadre du plan.

2.43. **Conversion des données géospatiales.** Les spécialistes de la conversion de données sont chargés de la mise en œuvre de la conversion des informations cartographiques au format de la base de données numériques. Ils ont été formés aux techniques pertinentes telles que la numérisation, le scannage et la vérification des bases de données géospatiales ainsi que l'élaboration de bases de données d'attributs utilisant des systèmes de gestion de base de données relationnelle. Les spécialistes de la conversion de données doivent définir le moyen le plus efficace pour élaborer la base de données numériques et diriger le personnel technique. La connaissance des nouvelles sources de données et des implications technologiques de l'obtention de

nouveaux supports pour faciliter la création de bases de données géographiques est également essentielle.

2.44. **Scannage et numérisation des cartes.** Si le scannage a fait de nouveaux adeptes et occupé une place de premier choix dans les stratégies de saisie de données de nombreux services de statistiques, la numérisation reste une solution éprouvée. Les personnes qui n'ont suivi aucune formation professionnelle en géographie ou dans un domaine similaire peuvent acquérir assez rapidement le savoir-faire technique. Cependant, la numérisation est une tâche extrêmement répétitive; même si elle demande une grande concentration, une vigilance sur des points de détail et une bonne compréhension de la structure des bases de données géographiques, elle peut être confiée à des employés de bureau. Le personnel le plus performant devrait en outre bénéficier d'une formation en assurance/contrôle de la qualité. Cette exigence de formation concerne également la numérisation frontale.

2.45. **Réalisation de la carte.** Les cartographes seront chargés de concevoir toutes les cartes et produits dérivés, y compris les cartes de district de recensement, les cartes de supervision et les cartes thématiques des résultats de recensement. Ils doivent connaître l'élaboration de cartes, la communication cartographique et avoir une formation en matière de SIG et de cartographie numérique. Les cartographes possédant une formation ont déjà la plupart des qualifications requises, mais ils devront être suffisamment formés aux techniques informatiques.

2.46. **Travail sur le terrain.** Les exigences du travail de cartographie de recensement sur le terrain ont changé en même temps que les techniques utilisées pour produire les cartes numériques. Les systèmes de positionnement universel sont devenus un outil essentiel pour la collecte de données sur le terrain et le personnel de terrain doit maintenant apprendre à utiliser ces systèmes et éventuellement aussi les ordinateurs portables qui servent à télécharger et à afficher ces données sur le terrain. Le personnel de terrain n'a pas besoin d'avoir des compétences professionnelles en géographie ou en géodésie mais il doit recevoir une formation pour pouvoir utiliser correctement ces nouveaux outils.

2.47. **Administration de systèmes.** Il a déjà été souligné que les organismes nationaux de statistique devaient investir dans des plates-formes de technologies d'information (TI) dédiées pouvant servir pour les SIG ainsi que pour d'autres besoins de traitement des données/TI. L'achèvement en temps utile d'un projet de base de données géographique de recensement dépend du bon fonctionnement du matériel informatique. Un administrateur de systèmes a pour fonctions d'assurer la maintenance du matériel et des logiciels informatiques en vue de réduire au minimum la durée d'indisponibilité, fournir une assistance au personnel qui réalise les cartes de recensement et assurer la sécurité des données (par exemple sauvegarde des données). Même s'ils ne participent pas directement aux activités de cartographie de recensement, les administrateurs de systèmes sont des membres essentiels de l'équipe cartographique, car la plupart des aspects de ce travail dépendent du bon fonctionnement du système informatique. L'administration des systèmes informatiques pour le service cartographique du bureau de recensement peut dans certains cas être assurée par le personnel normal d'assistance informatique de l'organisme.

2.48. **Exigences particulières.** Suivant la stratégie adoptée en matière de cartographie de recensement, il faudra peut-être compléter le personnel de l'organisme chargé d'élaborer les cartes de recensement en recrutant d'autres experts. Par exemple, si la mise à jour des cartes de recensement exige un nombre élevé de produits obtenus par les techniques de la télédétection, l'équipe devra disposer d'un analyste ayant reçu une formation en matière d'analyse d'images numériques. Il faudra éventuellement

recourir aux services d'autres experts : opérateurs d'un système de numérisation de cartes en grande série ou collaborateurs ayant l'expérience des logiciels de gestion de bases de données et de la programmation informatique. Les autres peuvent encore avoir des compétences graphiques, en déploiement Web ou en relations avec la clientèle. Toutes ces compétences sont utiles pour le développement de bases de données et l'optimisation de systèmes logiciels.

2.49. Niveaux de formation. Dans de nombreux pays, il peut exister un manque d'experts qualifiés dans le domaine géospatial susceptibles d'être recrutés à titre permanent ou temporaire pour le projet de cartographie de recensement. Le bureau de recensement doit donc évaluer les possibilités de formation pour que le personnel ancien et nouveau puisse acquérir les connaissances appropriées nécessaires à la réalisation correcte du projet. En général, un personnel formé aux techniques géographiques traditionnelles et possédant des notions d'informatique n'aura guère de difficulté à s'adapter aux techniques numériques après avoir reçu des cours de formation. Divers types de formation seront requis en fonction des besoins :

- a) Des séminaires de courte durée pour mieux faire connaître le programme numérique de cartographie de recensement doivent être organisés à l'intention de tous les collaborateurs du bureau de recensement, y compris du personnel d'autres sections et de responsables. Cela favorise l'intégration du projet de cartographie numérique au processus d'ensemble du recensement. Une large diffusion des informations présente en outre l'avantage de favoriser une meilleure utilisation des produits cartographiques de recensement par d'autres agences du bureau de recensement. Ces séminaires peuvent être organisés par le responsable du projet ou les spécialistes de la géographie du recensement;
- b) Une formation à des tâches répétitives telles que le scannage, la numérisation ou la révision peut comporter des séminaires internes de courte durée, suivis d'une formation en cours d'emploi. Les produits élaborés par les nouveaux collaborateurs devront être examinés de près pour vérifier si ces derniers ont besoin de recevoir des instructions ou une formation supplémentaires ou s'ils doivent être éventuellement réaffectés à d'autres tâches;
- c) Le principal groupe de cartographes chargés d'établir les cartes de recensement devra recevoir une formation complémentaire en matière de SIG et de techniques de cartographie numérique. Mais en raison du coût de la formation, seuls des collaborateurs permanents seront envoyés dans le pays ou à l'étranger pour y suivre des cours organisés par des universités, des distributeurs ou d'autres organismes. Les collaborateurs ainsi formés joueront un rôle essentiel dans l'information et la formation du personnel supplémentaire. Un grand nombre de personnes peuvent être formées en appliquant la méthode hiérarchique de la « formation de formateurs », particulièrement adaptée à une approche décentralisée de la cartographie de recensement;
- d) Les applications de techniques spécialisées comme le traitement d'images numériques ou les applications avancées en bases de données informatisées exigent généralement un diplôme professionnel ou une expérience pratique équivalente. Si l'on ne peut pas recruter de personnel compétent, le bureau de recensement devra envisager, bien avant d'entreprendre le projet de cartographie proprement dit, d'envoyer un collaborateur dans une université pour y recevoir une formation. Aujourd'hui dans le monde plusieurs universités et centres de formation ont des programmes d'ensei-

gnement professionnel spécialisé d'un ou deux ans en SIG, télédétection et techniques apparentées (pour une liste des institutions proposant une formation en informations et techniques géospatiales, voir annexe VII).

I. Coopération institutionnelle : infrastructure nationale de données spatiales : dispositions garantissant la compatibilité avec d'autres organismes publics

2.50. Dans de nombreux pays, plusieurs organismes publics établissent des bases de données géographiques numériques. Les organismes nationaux de cartographie utilisent de plus en plus des techniques pleinement numériques dans toutes les opérations d'établissement des cartes. Mais d'autres organismes publics, y compris les services des transports, de la santé, de l'environnement et des ressources en eau, utilisent aussi les technologies géospatiales pour gérer les informations qu'ils collectent et utilisent à des fins d'analyse et de planification. En outre, des entreprises du secteur privé, par exemple dans les secteurs des services collectifs, des télécommunications et des mines, ont compris les avantages que leur confère la gestion de leurs besoins d'informations sous la forme de cartes géographiques numériques.

2.51. Les États Membres se rendent de plus en plus compte que le développement d'une infrastructure d'informations géographiques est tout aussi important que la construction de routes, de réseaux de télécommunications et la fourniture d'autres services de base. Il est à présent largement admis que grâce au développement d'infrastructures de données nationales, les organisations gouvernementales, le secteur privé, les universités et la société civile en général disposeront plus facilement de données spatiales. De fait, les infrastructures de données spatiales permettent aux organismes nationaux de statistique d'accéder à l'expertise présente dans le pays, ainsi qu'aux sources numériques existantes, telles que les données de base sous forme géospatiale utilisées dans les logiciels de SIG. Dans ce cadre, les organismes nationaux de statistique sont perçus comme des nœuds d'information dans des réseaux de données et de prestataires de contenu couvrant tout le pays.

2.52. Les infrastructures nationales de données spatiales sont désormais présentes dans plus de 100 pays au monde. Les autorités de cartographie nationales, les ministères en charge de l'environnement et de l'urbanisme ainsi que l'armée sont autant de sources d'informations et d'expertise susceptibles d'aider les organismes nationaux de statistique à développer leur infrastructure nationale de données spatiales. Une structure organisationnelle de base d'une infrastructure nationale de données spatiales repose notamment sur un ministère responsable, une agence principale, un forum ou un réseau de producteurs et utilisateurs de données, un comité pilote et des groupes de travail techniques.

2.53. Pour entrer en contact avec l'infrastructure nationale de données spatiales, il convient d'identifier au préalable la personne de contact de l'agence principale. Cette dernière est en général l'agence de cartographie nationale. Au rang des données pouvant être échangées, citons une base de données de cartes topographiques numérisées ou des couvertures vectorielles numériques à une échelle de 1/100 000 ou plus fine, pouvant servir à la définition des frontières des districts de recensement, des données hypsométriques, des réseaux hydrographiques et de transport — tous utilisés pour délimiter les zones de recensement.

Encadré II.3

Trois exemples de collaboration en matière d'échange de données**1. Partage de matériel entre le Fonds des Nations Unies pour la population (FNUAP) et Fidji**

Le développement de nouvelles technologies dans le domaine de la géographie du recensement fournit des outils puissants permettant de renforcer l'efficacité des recensements. À Fidji, le recensement a été supervisé par une équipe du Bureau de statistique de Fidji qui géoréférençait chaque habitation occupée par un ménage à l'aide de dispositifs GPS. Cet exemple illustre l'une des manières d'apporter une valeur ajoutée au recensement, puisque cette procédure a permis une cartographie beaucoup plus précise et qu'elle facilitera à l'avenir les recensements et exercices statistiques. Le Fonds des Nations Unies pour la population (FNUAP) a facilité le financement des exercices de GPS et proposé d'étendre l'utilisation des dispositifs GPS à d'autres pays des îles du Pacifique désireux de procéder à des exercices identiques ou similaires. Cet accord de partage est primordial pour le renforcement des capacités et la collecte de précieuses informations pendant le recensement (pour un complément d'information, contacter Scott Pontifex, Secrétariat de la Communauté du Pacifique).

2. Le Secrétariat de la Commission du Pacifique Sud pour les géosciences appliquées (PACSU) et l'accord de traitement et de partage des données dans les pays insulaires du Pacifique

Les coûts d'acquisition de l'imagerie satellitaire permettant l'utilisation d'applications de SIG utilisées pour la géographie censitaire, telles que la délimitation des districts de recensement, sont considérablement inférieurs à ceux d'il y a quelques années. Toutefois, la majorité des services nationaux de statistique a encore du mal à y faire face. Les agences nationales et internationales peuvent être d'une grande utilité en mettant à disposition leurs ensembles de données afin de faciliter les opérations de recensement. Dans ce contexte, le Secrétariat de la Commission du Pacifique Sud pour les géosciences appliquées (PACSU) a eu accès à l'imagerie satellitaire, qu'elle peut proposer gratuitement à tous les États insulaires du Pacifique. Le PACSU peut également numériser et géoréférencer des cartes existantes. Cela constitue un excellent point de départ pour le développement de la géographie du recensement sur une plate-forme moderne. Les services nationaux de statistique envisagent de recourir aux solutions proposées par le PACSU dans le cadre de l'évaluation de l'utilisation des SIG aux fins du recensement de la population et du logement (pour un complément d'information, contacter Scott Pontifex, Secrétariat de la Communauté du Pacifique).

3. Les États-Unis et le partage de données

Le Bureau du recensement des États-Unis collabore avec des responsables tribaux, locaux, régionaux et nationaux ainsi qu'avec des agences de coordination, telles que des commissions de planification régionales, afin de mettre en œuvre plusieurs programmes de partage d'informations géographiques. Les connaissances fournies par les partenaires locaux permettent au Bureau du recensement des États-Unis de combler les besoins en matière de données spatiales et statistiques qu'il est tenu de satisfaire dans le cadre de sa mission : être une source statistique pour une meilleure compréhension de la nation.

Le recensement décennal des États-Unis et l'enquête communautaire américaine annuelle sont réalisés essentiellement à travers l'envoi de questionnaires par courrier aux ménages du pays. Une étroite collaboration avec le service postal des États-Unis fournit des mises à jour régulières des adresses au fichier principal d'adresses de la division géographie qui sert de source pour l'envoi de questionnaires. Conformément à un accord de protection de la confidentialité, le Bureau du recensement des États-Unis partage sa liste d'adresses avec des responsables tribaux, nationaux, régionaux et locaux afin de veiller à une expédition correcte et donc de garantir un comptage adéquat.

Il est primordial de disposer de l'emplacement correct des limites en vue de la collecte, de la tabulation et de la diffusion des données statistiques. Le Bureau du recensement des États-Unis effectue une enquête annuelle sur les limites et annexions, dans le cadre de laquelle les gouvernements lo-

caux passent en revue leurs limites et font part de toute modification. Dans le cadre de la préparation du recensement décennal, les organisations de planification régionales et les gouvernements ont la possibilité de passer en revue les secteurs statistiques et de recommander des changements afin de mieux répondre à leurs besoins locaux en données.

Grâce au recours croissant aux systèmes de SIG au sein des gouvernements locaux aux États-Unis, une agence fédérale telle que le Bureau du recensement des États-Unis a la possibilité d'acquérir, à travers des partenariats, des données géospatiales très précises et actuelles, l'aidant à gérer son codage et son référencement géographiques topologiquement intégrés (MAF/TIGER). MAF/TIGER est la source de toutes les activités géographiques au sein de l'agence statistique, telles que le support d'adresses, l'utilisation de données spatiales, le géocodage et la cartographie.

La coordination et le partenariat avec d'autres agences fédérales visant à mettre sur pied une infrastructure nationale de données spatiales sont engagés à travers la participation active du Bureau du recensement des États-Unis avec des groupes tels que le Federal Geographic Data Committee. Des accords avec des agences telles que le National Geodetic Service permettent à l'agence de disposer d'informations plus précises sur la position des habitations. Enfin, une collaboration étroite avec des organisations nationales telles que le National States Geographic Information Council et la National Association of Counties facilite les partenariats nécessaires avec les gouvernements locaux.

Les tendances actuelles en matière de partenariats sont critiques pour préserver l'infrastructure géospatiale nécessaire aux recensements et enquêtes. Aux États-Unis, il ne s'avère plus rentable de développer et gérer quantitativement et qualitativement les adresses et données spatiales lesquelles, au cours des décennies précédentes, dépendaient d'un effort centralisé au niveau fédéral. Une collaboration étroite avec des partenaires au niveau local offre de nouvelles opportunités d'améliorer les données alimentant l'infrastructure géospatiale d'une agence statistique (pour un complément d'information, contacter Tim Trainor à l'adresse suivante : timothy.f.trainor@census.gov).

* * *

Une coopération au niveau national est primordiale en ce qui concerne l'acquisition, l'utilisation et le développement d'applications de SIG. L'instauration de correspondants des SIG et de groupes d'utilisateurs est la meilleure solution pour garantir des synergies et éviter la duplication et le gaspillage de ressources humaines autrement insuffisantes. On ne peut sous-estimer l'importance d'une assistance technique en bonne et due forme fournie par des agences nationales et internationales. Dans de nombreux pays, elle est une condition essentielle du développement de leur capacité à mener à bien des exercices complexes, tels qu'un recensement. La réussite d'un recensement repose bien souvent sur l'efficacité des communications publiques et de la campagne de sensibilisation qui le précède. Les ressources nécessaires à une telle campagne sont souvent obtenues auprès de donateurs extérieurs et d'agences internationales et il convient d'en tirer pleinement parti.

2.54. De nombreux utilisateurs au sein et en dehors des agences gouvernementales ont besoin de consulter des bases de données géographiques de base. Un grand nombre de ces utilisateurs ont besoin d'accéder à plusieurs bases de données ou d'utiliser une couche de données géographiques de base en guise de modèle pour collecter leurs propres données spatiales. Ces couches de données standard, qui constituent la base de nombreuses activités de cartographie et de collecte de données, sont qualifiées de « données-cadres ». Les couches de données de base qui forment le cadre national de données spatiales sont les suivantes [voir Commission économique pour l'Afrique (2007)] :

- a) **Points géodésiques.** Système de points géographiques définis avec précision qui servent de référence pour toutes les activités cartographiques d'un pays; fréquemment qualifiés également de points de référence;
- b) **Géographique de base.** Inclut l'imagerie (photos aériennes ou images satellites haute résolution), l'hypsographie (terrain) et l'hydrographie (détails hydrographiques superficiels; soit naturels tels que les rivières et les lacs, soit artificiels, comme les canaux);

- c) **Organisation administrative et spatiale.** Comprend des unités gouvernementales sous-nationales, telles que des provinces et des districts, délimités au sol, les noms géographiques et les unités/zones de gestion du territoire;
- d) **Infrastructure.** Inclut les routes, les voies d'eau intérieures, les voies ferrées ainsi que toute infrastructure de transport de personnes ou de marchandises, les services publics et les services ;
- e) **L'environnement naturel.** Comprend les types de sols, les zones de végétation, les noms géographiques des zones d'habitat.

2.55. Les couches les plus importantes pour le bureau de recensement sont celles de divisions administratives. De toute évidence, les districts de recensement doivent concorder avec les limites définissant la hiérarchie des divisions administratives du pays et la répartition de la population. Mais les couches de données comme celles des transports et de l'hydrographie sont aussi importantes pour établir les cartes de recensement, car les routes et les cours d'eau constituent une délimitation naturelle des districts de recensement.

2.56. Une fois un recensement terminé, le bureau de statistique national sera invité à partager ses résultats et produits avec d'autres protagonistes de l'infrastructure nationale de données spatiales. Les limites des districts de recensement, avec les données de recensement, telles que les informations démographiques de base et éventuellement les caractéristiques spécifiques au logement et aux communautés, sont une importante source de données pour d'autres organismes publics et privés. L'organisme national de statistique pourra également fournir un ensemble de données contenant les points centraux (centroïdes) des districts de recensement. L'analyse du secteur de la santé, par exemple, nécessite des renseignements détaillés concernant la population exposée à un risque. Pour la planification dans le secteur des transports, on a besoin de données concernant la demande de services de transports publics. Et les services collectifs publics et privés doivent être renseignés sur les zones où ils devront développer leur capacité d'alimentation en électricité, en eau ou en fournitures de services de télécommunications.

2.57. La notion d'infrastructure nationale de données spatiales constituée de grandes bases de données géographiques géoréférencées a trois conséquences pour les activités cartographiques de recensement :

- a) Le bureau de recensement doit mettre à la disposition de l'infrastructure nationale de données spatiales un ensemble cohérent d'unités déclarantes correspondant à la hiérarchie administrative et auxquelles des renseignements socioéconomiques et des données connexes peuvent être associés. Pour faire en sorte que ces cartes de recensement puissent être intégrées à d'autres sources de données, l'organisme qui établit les cartes de recensement doit respecter toutes les normes nationales existantes applicables aux données géographiques;
- b) Pour assurer la compatibilité avec d'autres ensembles de données et faciliter l'élaboration de cartes de recensement, les services chargés de la cartographie de recensement doivent collaborer étroitement avec les autres organismes publics participant à l'élaboration des cartes. Non seulement, cette collaboration assurera la cohérence des normes et des définitions, mais elle entraînera une baisse des coûts en évitant d'entreprendre des activités en double;
- c) Les organismes nationaux de statistique doivent accorder une attention particulière aux questions relatives au référencement géographique. Afin de superposer différents ensembles de données couvrant des zones situées

dans la même étendue géographique, il est nécessaire de savoir comment la position des éléments a été définie. En d'autres termes, on devra connaître la projection ainsi que la date et les détails du système de coordonnées afin de garantir des relations spatiales correctes entre les objets dans différents ensembles de données.

2.58. Une agence de statistique pourrait également contribuer aux efforts de l'infrastructure nationale de données spatiales à travers d'autres formes de participation, dont un engagement à participer aux réunions de planification et à rester informée des développements au sein du pays. L'organisme national de statistique sera également contraint de fournir des métadonnées.

Encadré II.4

Participation et coordination des agences internationales

1. Le projet de cartographie mondiale

Le projet de cartographie mondiale a été proposé en 1992 à l'occasion de la Conférence des Nations Unies sur l'environnement et le développement, à Rio de Janeiro (Brésil) ; En 1996, le Comité directeur international de la cartographie mondiale a été mis sur pied pour coordonner le réseau d'organisations de cartographie officielles des pays participants. En mars 2007, 172 pays et régions du monde avaient fourni ou élaboré des données pour la cartographie mondiale (« Global Map »). Ce projet a pour objectif d'assurer la couverture de toute la surface de la Terre à une échelle de 1/1 000 000, avec une résolution spatiale de 1 km. Les données seront mises à jour à des intervalles de cinq ans environ afin de superviser les changements au fil du temps. Global Map est une plate-forme gérant quatre cartes de base au format raster (l'utilisation des terres, l'occupation des sols, la végétation et l'élévation) et quatre cartes de données au format vectoriel (agglomérations, drainage, transports et frontières). Les données de la cartographie mondiale sont consultables gratuitement via Internet à des fins non lucratives (pour un complément d'information, voir <http://www.globalmap.org>).

2. DevInfo

DevInfo du Fonds des Nations Unies pour l'enfance (UNICEF) est un outil logiciel destiné à aider les pays à superviser les objectifs du Millénaire pour le développement et favoriser leur concrétisation en recourant à des mesures politiques, des stratégies multisectorielles et au développement des mesures adéquates. DevInfo est une solution générale pour la compilation et la présentation de données. Outre un référentiel de données, DevInfo met à disposition des outils simples permettant de consulter la base de données et d'élaborer des graphiques (tableaux, graphiques et cartes) destinés à des rapports et présentations. DevInfo devrait être un puissant outil de mobilisation qui contribuera à une meilleure prise de conscience et une meilleure connaissance des objectifs au niveau national ainsi qu'au processus décisionnel éclairé (pour un complément d'information, voir <http://www.devinfo.org/>).

3. Infrastructure nationale de données spatiales des Nations Unies

L'infrastructure de données géospatiales des Nations Unies a été proposée en 2005 afin de promouvoir et mettre en œuvre des activités de développement durable et parfaire les opérations humanitaires et de maintien de la paix. Elle ambitionne une vaste infrastructure d'informations géospatiales décentralisée facilitant la prise de décisions à divers niveaux en permettant un accès, une extraction et une diffusion de données et d'informations géospatiales rapides et sécurisés. L'infrastructure de données géospatiales des Nations Unies permet une interopérabilité entre les infrastructures de données spatiales conçues à des fins spécifiques au sein d'agences des Nations Unies, entre les groupes d'agences des Nations Unies ayant des intérêts communs ainsi qu'entre les États Membres des Nations Unies et leurs groupes/partenaires thématiques et régionaux. L'infrastructure de données géospatiales des Nations Unies permettra d'accéder, d'extraire et de diffuser des données et services géospatiaux, en évitant toute duplication au sein des Nations Unies (pour un complément d'information, voir <http://www.ungwg.org/unsdi.htm>).

J. Normes

2.59. Pour faciliter l'échange de données entre utilisateurs de données, il est évidemment nécessaire de coordonner l'élaboration des bases de données géographiques. Dans plusieurs pays, des comités des données géographiques ont été constitués à cette fin, réunissant les principaux responsables de l'élaboration des données spatiales. En outre, des organisations supranationales comme l'Organisation européenne pour l'information géographique (EUROGI), l'Infrastructure mondiale de données géospatiales (GSDI), l'Organisation internationale de normalisation–Information géographique/Géomatique (ISO-TC/211) et le Open GIS Consortium jouent un rôle actif dans la définition de normes de données géographiques.

K. Collaboration

2.60. Lorsqu'il élabore une base de données géographique numérique, l'organisme de recensement peut choisir de collaborer avec d'autres organismes publics ou avec le secteur privé. Les deux possibilités ont été utilisées avec succès dans divers pays. Comme indiqué ci-dessus, parmi les organismes publics, le service cartographique national est naturellement le premier point de contact, en particulier pour les organismes nationaux de statistique qui entament leurs activités. Mais d'autres organismes (tels que des organismes gouvernementaux locaux, environnementaux et cadastraux) peuvent aussi mettre leurs ressources à disposition ou avoir intérêt à partager le coût de la mise en place d'une base de données de recensement de qualité élevée. Parmi les organismes du secteur privé, les distributeurs de logiciels et de matériel peuvent appuyer les aspects techniques des opérations cartographiques de recensement, soit en passant un contrat avec le bureau de recensement, soit dans le cadre d'un système de partage des coûts dans lequel l'entreprise privée récupérera son investissement en vendant des bases de données de recensement géoréférencées. Les prestataires de données, départements universitaires et autorités chargées des routes comptent parmi les autres collaborateurs. Il faut cependant noter que la collaboration avec d'autres organismes est recommandée, mais non obligatoire. Le service cartographique de recensement ayant pour tâche principale d'établir une base géographique pour le recensement, il doit éviter de dépendre d'un fournisseur extérieur d'informations.

2.61. Tout partenariat ou collaboration doit être fondée sur une intention commune ou un accord précis. Les éléments suivants de l'accord de coopération ou de la lettre d'intention doivent être spécifiés :

- a) **Régularisation.** Une collaboration mal définie est-elle suffisante ou faut-il régulariser les arrangements dans le détail ? Il faudra beaucoup de temps pour mettre en place un accord rédigé dans des termes plus précis, mais cela permettra d'éviter par la suite des différends relatifs aux droits et aux responsabilités en matière d'élaboration et d'utilisation des données obtenues. Dans la plupart des cas, il faudra donc prévoir un protocole d'accord officiel, juridiquement contraignant, entre le bureau de recensement et l'organisme qui collabore avec lui, couvrant tous les aspects pertinents du partenariat. Il convient de prévoir ce type d'accord contractuel si l'on a affaire à des fournisseurs de données ou de services du secteur privé. Au rang des questions importantes, citons la nécessité de définir des normes pour un usage externe et interne, la valeur d'objectifs clairement définis ainsi que le besoin d'une aide à la gestion.

- b) **Portée du partenariat.** Les accords de collaboration peuvent ne concerner que l'utilisation des données d'une autre agence ou peuvent prévoir l'élaboration à partir de zéro d'une vaste base de données spatiales très complète. Des protocoles doivent être définis pour exploiter les données fournies par des municipalités et d'autres entités administratives ou géographiques d'ordre inférieur.
- c) **Responsabilités.** Qui s'acquittera de telle ou telle tâche ou de telle ou telle fonction ? Les questions à examiner concernent notamment l'élaboration des données, la maintenance, l'accès aux données, la supervision du projet et l'utilisation des ressources.
- d) **Avantages.** Il est évident que l'accord doit être avantageux pour tous les participants, sauf si un organisme se contente d'acheter les services d'un autre organisme. Il est utile de préciser comment les divers partenaires tireront avantage des accords, de façon à répartir équitablement les tâches et les responsabilités.
- e) **Besoins en ressources.** Les ressources comprennent le personnel, le contexte informatique, les documents et les communications. Il convient également de tenir compte des ressources nécessaires à la gestion et la supervision du projet.
- f) **Partage des coûts.** Tous les coûts directs et indirects liés aux activités de partenariat doivent être répartis équitablement. La gestion comptable peut être complexe car les contributions peuvent prendre la forme d'argent en espèces, de données, de main-d'œuvre, d'utilisation de matériel ou toute autre forme.
- g) **Récupération des coûts.** Si la diffusion des produits obtenus procure des recettes, elles doivent être partagées compte tenu des dépenses effacées au titre de la gestion et des opérations de diffusion de données. Cela suppose aussi une détermination précise des utilisations convenues et des droits d'auteur sur les produits obtenus.
- h) **Résolution des conflits.** En cas de différends surgissant pendant la durée du projet, il est utile de prévoir à l'avance la ligne de conduite à suivre en vue de les résoudre.

L. Résumé et conclusions

2.62. Pour planifier un recensement géospatial, il conviendra de se pencher sur les questions du financement, des effectifs et de la gestion de projets. Comme montré ci-dessus, les organismes nationaux de statistique ont la chance de recourir à de nouvelles technologies géospatiales, et bénéficient d'ordinateurs personnels, d'appareils portables, de GPS et d'une imagerie satellitaire et aérienne à bas coût, afin d'obtenir des informations sur leurs populations. De nombreux organismes nationaux de statistique doivent engager de nouveaux talents et réorganiser l'entreprise de manière à s'accommoder de la situation. Pour ce faire, il est nécessaire de prendre conscience que le défi à relever consiste à attirer et fidéliser le personnel formé; à cet égard, de nombreux pays ont fait part de leurs préoccupations quant à la manière de résoudre ces questions.

2.63. Les organismes nationaux de statistique ont besoin d'informations sur les initiatives des infrastructures nationales de données spatiales, que ce soit pour consulter de précieuses données de base réservées à la planification des recensements ou pour partager des données et informations après le recensement. Il s'agit ici d'amener

les organismes nationaux de statistique à passer de la phase des engagements au stade opérationnel en commençant par des questions pratiques. Ils pourront dès lors obtenir les réponses dont ils ont besoin pour élaborer un plan de numérisation géocentrique, adapté à l'échelle du pays, qui les aidera à développer leur efficacité et leur rentabilité. La planification est primordiale, il convient de ne pas perdre cela de vue. Les responsables d'agences doivent définir des objectifs clairs et contribuer au plan stratégique afin de garantir le succès.

Chapitre III

Développement d'une base de données du niveau des districts de recensement pour le recensement

A. Introduction

3.1. Comme indiqué au chapitre II, l'utilisation de la technologie géospatiale pour créer de meilleures données relève à certains égards d'une question organisationnelle, nécessitant entre autres de définir des objectifs et de se doter des compétences adéquates en ressources humaines. Réorganiser l'organisme national de statistique autour d'un noyau d'informations géographiques revient à comprendre le lien entre la géographie d'un pays et les divers jeux d'informations que l'organisme utilise et produit. La relation entre la géographie et les bases de données prend forme à travers le processus de codage. La première étape consiste à établir un lien entre les supports de gestion décrits au chapitre II et le contenu technique, en montrant de quelle manière la base de données de recensement géographique en arrive à être au centre d'activités dans le cadre desquelles des formulaires d'informations du recensement sont stockés et consultés.

3.2. Le présent *Manuel* poursuit un objectif essentiel : élaborer un plan opérationnel pour construire une base de données du niveau des districts de recensement aux fins du recensement, laissant transparaître l'évolution de la méthode de réalisation d'un recensement. Un recensement est bien plus qu'un simple exercice de collecte des données. Il permet aux pays membres de se doter de capacités en matière de technologies de l'information (TI) tout en progressant dans leurs objectifs de développement. Alors que le recensement était auparavant une opération ponctuelle généralement réalisée tous les 10 ans, la gestion d'une base géographique précise en vue du recensement est désormais un processus constant.

3.3. Il sera très bénéfique à l'avenir de contribuer aux opérations de recensement en optant pour un géoréférencement numérique puisque une nouvelle génération de produits devient plus accessible au public. S'efforcer d'élaborer un recensement précis à travers l'adoption de technologies géospatiales est un service public louable. Compte tenu de l'engagement nécessaire et des investissements en jeu pour restructurer les opérations de cartographie de recensement, l'organisme national de statistique peut s'en remettre au travail de ses prédécesseurs et employer des cartes établies les années précédentes.

3.4. Le présent chapitre présente un concept de codage géographique légèrement différent de celui de l'appariement d'adresses traditionnellement utilisé par l'industrie des SIG. Les Nations Unies proposent une définition plus vaste du « géocodage ». Il s'agit du rapport entre les observations statistiques et les emplacements du monde réel exprimé en termes de latitude et de longitude ou d'autres attributs de

localisation. Plus simplement, le codage géographique est une méthode garantissant que les données savent où elles se trouvent. Si le présent chapitre met en avant cette définition plus large du géocodage, il souligne également toute l'utilité du codage traditionnel (à savoir le codage des attributs) pour un recensement.

3.5. Le géocodage est conçu pour englober un continuum d'échelles spatiales, allant d'habitations individuelles aux niveaux national et administratif supérieurs en passant par le niveau des districts de recensement. Pour une mise en œuvre fructueuse, un pays doit définir un ensemble de circonscriptions administratives avec des territoires connus et une représentation numérique sous la forme d'un codage informatique. Le système de codage doit laisser transparaître une qualité importante propre à un plan de géocodage : il doit être à la fois souple et extensible tout en permettant une couverture exhaustive et une marge de manœuvre pour son développement.

3.6. Les hiérarchies administratives reposent sur l'idée que le territoire du pays comporte des limites servant à démarquer l'étendue territoriale effective au niveau de l'état, de la province ou des districts, voire à des fins électorales, de contrôle de la santé ou de distribution postale. Ces diverses géographies peuvent être stockées dans une base de données avec le code de niveau administratif et le nombre d'unités. Ainsi, les unités au deuxième niveau administratif (ADM 2s) sont des provinces, alors que les unités au troisième niveau administratif (ADM 3s) sont des districts. Idéalement, toute opération géospatiale doit disposer de ces unités au format SIG afin de les utiliser dans ses divers projets.

3.7. Le présent chapitre s'attardera ensuite sur les normes et sur une question relevant de la gestion, la coordination. Les normes nationales en exigent une. Lorsque l'agence statistique est l'autorité (c'est-à-dire le responsable des codes), l'organisme national de statistique a besoin d'élaborer un plan transparent et justifiable. Une coordination régionale s'impose alors, en particulier dans les grands pays.

3.8. Le rôle d'un organisme national de statistique au niveau des délimitations de frontières administratives variera d'un pays à l'autre. L'organisme national de statistique disposera éventuellement d'une autorisation légale pour définir des unités sous-nationales sur le terrain, ou peut-être sera-t-il supposé qu'il jouit d'une autorité de facto. Si l'organisme n'est pas le responsable des frontières sous-nationales, il devra collaborer avec d'autres agences au sein du gouvernement pour utiliser le plan de géocodage en vigueur, en l'adaptant éventuellement à des besoins de recensement spécifiques. Un certain degré de coordination régionale peut également s'avérer nécessaire.

3.9. Le présent chapitre traite notamment des thèmes suivants : la définition de la géographie de recensement nationale, en ce compris les hiérarchies et les critères administratifs ainsi que le processus de délimitation des districts de recensement; le codage des zones géographiques, dont les districts de recensement, en veillant à leur compatibilité avec des recensements précédents; les sources de données géographiques utilisées pour la délimitation des districts de recensement ainsi que les procédures d'importation; la conversion des données géographiques à travers le scannage et la numérisation; l'élaboration et la gestion de la topologie; le déploiement d'une base de données de districts de recensement; les problèmes de qualité des données; et le développement de métadonnées.

3.10. Comme toujours, il convient de faire preuve de prudence dans le cadre de l'élaboration des projets. Qui plus est, il est primordial d'insister sur le fait que compte tenu de la demande croissante en données de petites zones, les besoins des pays doivent être évalués de manière réaliste.

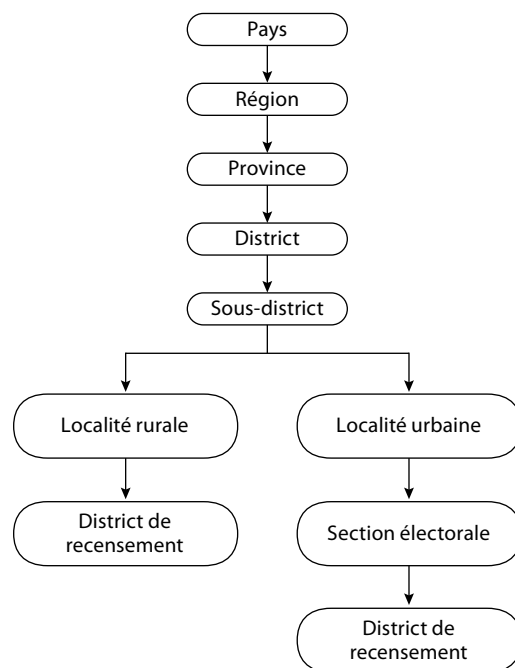
B. Définition du territoire géographique national aux fins du recensement

1. Hiérarchie administrative

3.11. L'une des premières décisions à prendre lorsque l'on prépare un recensement concerne les circonscriptions administratives auxquelles les données du recensement se rapporteront. Ces circonscriptions peuvent être constituées par n'importe quelle unité géographique spéciale, mais il s'agit essentiellement d'unités administratives, à savoir des autorités publiques compétentes sur le territoire. La préparation du recensement nécessite l'établissement d'une liste de toutes les unités déclarantes administratives et statistiques du pays. Les relations entre tous les types de limites de circonscriptions administratives et d'unités déclarantes doivent être définies. Chaque pays a sa hiérarchie administrative propre, c'est-à-dire le système par lequel le pays puis chaque ensemble de circonscriptions administratives de niveau inférieur (à l'exception du niveau le plus bas) sont subdivisés pour former le prochain niveau inférieur. Par exemple, aux fins du recensement, un pays peut avoir été subdivisé en sept niveaux hiérarchiques dans les zones urbaines et six dans les zones rurales (voir figure III.1).

Figure III.1

Hiérarchie géographique générique pour le recensement

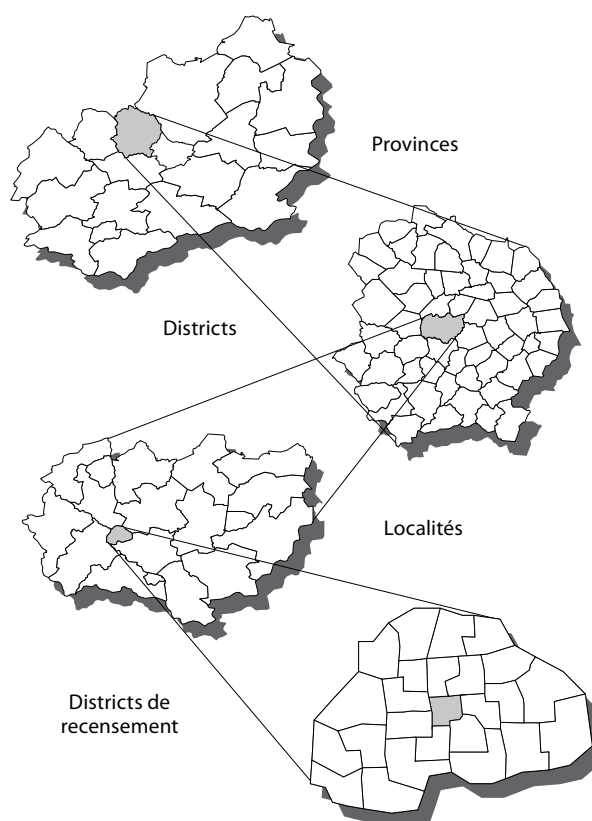


3.12. Quelques-uns seulement de ces niveaux hiérarchiques peuvent jouer un rôle administratif, par exemple la province, le district et la localité peuvent avoir des chefs-lieux, avec des services d'administration locale au niveau de ces circonscriptions. D'autres unités peuvent remplir uniquement un rôle statistique, étant conçues pour l'affichage des données et non pour l'administration du territoire. La figure III.2 montre l'imbrication des circonscriptions administratives et des unités de recensement au moyen d'un exemple simple avec seulement quatre niveaux hiérarchiques.

Cependant, dans certains cas, les circonscriptions administratives peuvent ne pas être complètement emboîtées. En particulier lorsqu'il prend en compte les unités administratives et les autres unités statistiques déclarantes, le bureau de recensement peut avoir affaire à un système très complexe de régions géographiques.

Figure III.2

Exemple d'une hiérarchie administrative imbriquée



3.13. Tous les niveaux n'ont pas la même importance. Par exemple, de nombreux pays subdivisent leur territoire en grandes régions, souvent définies du point de vue géographique, comme la région « Nord-Sud-Sud-Ouest-Est » ou la région « Montagne-Plaines-Zone côtière ». Souvent, ces régions ne jouent aucun rôle administratif, mais elles peuvent cependant être utilisées pour fournir des informations statistiques.

2. Rapports entre les circonscriptions administratives et les autres unités de déclaration statistique ou de gestion

3.14. En plus des circonscriptions administratives, la plupart des pays comportent divers autres ensembles de zones, utilisées à différentes fins, et pour lesquelles des données de recensement devront être réunies. Ces zones ont des finalités spécifiques et peuvent être définies à une échelle approximative. On peut citer :

- Les régions sanitaires.
- Les zones d'emploi.
- Les circonscriptions électorales.

- Les zones postales.
- Les zones culturelles ou tribales.
- Les agglomérations urbaines ou les zones métropolitaines.
- Les unités de recensement agricoles ou économiques.
- Les circonscriptions d'enregistrement des titres fonciers et du cadastre.
- Les zones des services d'utilité publique (distribution d'eau ou d'électricité).

3.15. L'interaction avec les agences responsables du pays ou les activités relatives à l'infrastructure de données spatiales permettent d'en apprendre davantage sur ces géographies particulières. Au niveau de leur rendu spatial, certaines zones peuvent ne pas s'imbriquer parfaitement dans la hiérarchie administrative du pays. En définissant les districts de recensement, le service cartographique de recensement devrait tenir compte le plus possible de ces unités déclarantes afin de faciliter la tabulation des données de recensement pour ces régions. L'analyse des besoins des usagers effectuée aux stades de la préparation du recensement devrait fournir des indications sur les zones non administratives que l'on devra le plus prendre en considération. En général, pour éclairer la définition des districts de recensement, le service cartographique de recensement devra diviser tous les ensembles de zones en unités pour lesquelles la compatibilité est obligatoire, souhaitable ou improbable et il devra prendre une décision en conséquence.

3.16. Pour certaines zones déclarantes ou de gestion à l'intérieur du pays, les données numériques relatives aux limites ont peut-être déjà été enregistrées par les organismes responsables. Par exemple, plusieurs pays qui ont entrepris des programmes de réforme agraire utilisent les SIG pour gérer des bases de données sur les titres fonciers (renseignements cadastraux) et de nombreux organismes postaux utilisent des bases de données de SIG relatives aux codes postaux pour accélérer la distribution du courrier. Lorsque les bases de données numériques de ces unités sont disponibles, elles peuvent appuyer l'élaboration des bases de données géographiques du recensement. Si un degré élevé de compatibilité peut être assuré, cela présentera en outre l'avantage que les statistiques relatives à d'autres zones, concernant par exemple la demande d'eau ou les résultats électoraux, pourront être combinées plus facilement à des statistiques démographiques et sociales.

3.17. À l'office de statistique, d'autres opérations de recensement supposent aussi que des unités de collecte de données sont définies. Il est très important de noter que des recensements agricoles et économiques sont effectués à intervalles réguliers dans de nombreux pays. Nombre d'applications analytiques tirent avantage de l'analyse conjointe des informations tirées des recensements de la population et des données agricoles et économiques. Un degré élevé de concordance entre les unités géographiques utilisées pour grouper ces types de données contribuera beaucoup à accroître leur utilité dans des applications publiques et non gouvernementales.

3.18. Il est par ailleurs nécessaire de préserver une coordination entre l'office de statistique et d'autres agences ou niveaux du gouvernement afin d'établir la responsabilité des limites administratives et les moments auxquels des changements sont possibles. Avant un recensement, il convient de geler les limites (il est recommandé de le faire au moins six mois avant le recensement) de manière à ce que leur modification n'engendre pas de divergence dans le codage effectif des zones. Les organismes nationaux de statistique peuvent également opter pour un « horodatage » de versions spécifiques des limites administratives de manière à ce qu'elles reflètent les divisions internes du pays au moment du recensement.

3. Critères et processus de délimitation des districts de recensement

3.19. Les districts de recensement sont les unités géographiques opérationnelles servant à la collecte de données de recensement et sont définies dans les premiers stades du processus de recensement. La délimitation des districts de recensement est la même, que les techniques cartographiques utilisées soient manuelles ou numériques. On devra tenir compte de divers critères pour définir les districts de recensement. Dûment délimités, les districts de recensement :

- Doivent être mutuellement exclusifs (ne pas se superposer) et exhaustifs (couvrir l'intégralité du pays).
- Leurs limites doivent être faciles à identifier sur le terrain.
- Doivent être compatibles avec la hiérarchie administrative.
- Doivent être homogènes sans présenter de poches ni d'éléments disjoints.
- Doivent être d'importance à peu près égale du point de vue de la population.
- Doivent être assez petits et faciles d'accès pour être traités par un seul enquêteur pendant la durée du recensement.
- Doivent être suffisamment petits et adaptables pour qu'il soit possible d'établir les tableaux les plus variés pour diverses unités statistiques déclarantes.
- Doivent répondre aux besoins des administrations et des autres utilisateurs de données.
- Doivent être utiles pour d'autres types de recensement et d'activités de collecte de données.
- Doivent être suffisamment grands pour garantir la confidentialité des données.

3.20. Certains de ces critères sont de nature à faciliter la collecte des données de recensement, tandis que d'autres concernent l'utilité des districts de recensement pour obtenir les produits recherchés — c'est-à-dire les rapports entre la collecte des données et les unités de tabulation. Il faut avoir présent à l'esprit que l'objet du recensement est de fournir des données utiles aux administrateurs, aux responsables politiques et aux autres utilisateurs de données de recensement. La flexibilité et l'adaptabilité maximales pour obtenir les meilleurs produits possibles doivent donc prendre le pas sur la facilité de dénombrement lors du recensement. Toutefois, la délimitation des districts de recensement doit également être cohérente du point de vue logistique pour le travail sur le terrain.

3.21. Les dimensions des districts de recensement peuvent être définies de deux manières : sur la base de la superficie ou sur celle de la population. Pour établir les cartes de recensement, l'importance de la population est le critère le plus important, mais la superficie et l'accessibilité doivent aussi être prises en compte pour qu'un enquêteur puisse desservir un district de recensement pendant le délai qui lui est imparti. Le chiffre de population retenu varie d'un pays à l'autre et est déterminé en général à partir des résultats du test préalable. La taille moyenne de la population peut aussi varier entre zones rurales et zones urbaines, car le dénombrement peut progresser plus vite dans les villes ou les agglomérations qu'à la campagne. Dans certains cas, il peut être recommandé de définir des districts de recensement plus grands ou plus petits que la moyenne. À toutes fins utiles, la population d'un district de recensement sera comprise entre 100 et 500 habitants.

3.22. Avant de définir les limites d'un district de recensement, on devra évaluer le nombre de personnes qui y vivent et leur répartition géographique. Sauf si

l'on dispose de renseignements provenant d'une enquête récente d'un système d'enregistrement ou d'une autre source d'informations, ce nombre devra être déterminé en comptant les habitations, en calculant le nombre de ménages correspondant et en multipliant par le nombre moyen de membres d'un ménage. Le nombre d'habitations peut être déterminé par un travail cartographique sur le terrain, une coopération avec des responsables gouvernementaux, une extrapolation des résultats de précédents recensements ou à l'aide de photographies aériennes voire d'images satellitales comme on l'expliquera au chapitre suivant.

3.23. Les limites des districts de recensement doivent être clairement observables sur le sol. Tous les enquêteurs, même s'ils n'ont pas de grandes connaissances géographiques, doivent pouvoir trouver les limites du district dont ils sont chargés. On pourra donc retenir un nombre de population plus ou moins grand suivant le district pour avoir une délimitation facilement identifiable. Les détails naturels utilisables à cette fin sont les routes, les voies ferrées, les ruisseaux et les cours d'eau, les lacs, les clôtures ou tout autre élément définissant une limite précise. Les détails aux contours plus flous, comme les broussailles, les forêts ou des courbes de niveau telles que des arêtes conviennent moins bien. Les limites administratives sont souvent invisibles. Dans certains cas, on est amené à utiliser des limites de districts de recensement qui n'apparaissent pas clairement sur le sol. Dans ce cas, on devra les décrire par écrit avec précision et porter une annotation appropriée sur la carte de district. On pourra par exemple avoir des lignes décalées ou prolongées. Ainsi, une limite de district de recensement peut suivre une route parallèlement à une certaine distance de celle-ci. Ou une partie d'une limite de district peut être définie comme le prolongement d'une route nettement repérable jusqu'à un autre détail clairement défini tel qu'un cours d'eau ou une voie ferrée.

3.24. Dans de nombreux pays, la délimitation des districts de recensement pose des problèmes particuliers. Par exemple, si un village peut être attribué à telle ou telle circonscription administrative, la limite effective circonscrivant la zone du village peut ne pas être définie. D'autre part, des groupes particuliers de populations, comme du personnel itinérant, des nomades ou des militaires, doivent être rattachés à une localité géographique. Les marins le sont souvent aussi à leur port d'attache. Dans le cadre de la planification de la localisation de populations difficiles à recenser, il convient de ne pas perdre de vue que les coûts opérationnels sont parfois 10 à 20 fois supérieurs à ceux applicables aux populations résidentielles des zones urbanisées.

3.25. Une délimitation des districts de recensement repose sur plusieurs critères, dont la taille idéale d'un district de recensement, laquelle repose sur le nombre de personnes qu'un recenseur peut compter au cours de la période de la collecte de données. Le plan de ladite délimitation doit tenir compte du plan global de recensement, qui varie selon le nombre de jours alloués au dénombrement. Préalablement au recensement, un test peut déterminer le nombre d'habitations qu'un recenseur peut couvrir par jour. À titre d'exemple, si 16 habitations peuvent être recensées par jour dans les zones urbaines mais seulement 10 dans les zones rurales et si la période de recensement ne dure que 10 jours, le district de recensement urbain idéal comporterait 160 habitations et le district de recensement rural en compterait 100. Si un logement compte cinq personnes en moyenne, la taille idéale de la population devrait être de 800 personnes pour une zone urbaine et de 500 pour une zone rurale. D'autres facteurs influencent la taille d'un district de recensement : les limites des districts administratifs, la visibilité des limites des districts de recensement, la présence de lieux de vie collective tels que des casernes, des hôtels et des dortoirs d'école ainsi que le mode et la disponibilité des moyens de transport.

3.26. Les évaluations de la population sont primordiales pour délimiter correctement les districts de recensement. Il peut être demandé aux responsables locaux de fournir des estimations de petites zones; les zones en question peuvent également être visitées par le personnel de terrain de l'organisme national de statistique. Pour les zones n'ayant fait l'objet d'aucune modification importante, un ajustement des estimations du recensement précédent peut être effectué en fonction du temps écoulé.

4. Délimitation des zones de supervision

3.27. Les zones de supervision permettent une gestion efficace des équipes de recenseurs. Une fois les districts de recensement délimités, la conception des cartes de supervision est en général simple. Les zones de supervision sont en général constituées de groupes de 8 à 12 districts de recensement contigus qui partagent un certain nombre de caractéristiques avec les districts de recensement. Les districts attribués à une même zone de supervision doivent être compacts pour réduire au minimum les durées de transport et de taille approximativement égales. Ils devront être inclus dans la zone relevant d'un même bureau sur le terrain, généralement définie sur la base des circonscriptions administratives.

3.28. Suivant les dimensions du pays, on pourra ajouter éventuellement des niveaux supplémentaires de zones de gestion du recensement. Dans les grands pays, ces zones coïncideront souvent avec les services provinciaux ou régionaux de statistique.

5. Codage géographique ou « géocodage » des districts de recensement

3.29. Une base de données géographique numérique au format vectoriel consiste en un ensemble structuré de points, de lignes et de polygones. Chaque objet géographique — c'est-à-dire chaque point, ligne ou surface — possède un identificateur unique, utilisé intérieurement par le système. Cet identificateur interne n'est en général pas accessible pour l'utilisateur et il ne doit pas être modifié de l'extérieur. Par contre, il faut avoir un identificateur plus significatif qui puisse être utilisé pour relier les entités géographiques aux attributs enregistrés pour chacune d'elles. Pour les districts de recensement et les unités administratives, cette liaison est assurée par l'identificateur unique du district de recensement ou de l'unité administrative, qui figure dans le fichier maître de toutes les zones géographiques prises en considération lors du recensement.

3.30. Le mode de saisie de cet identificateur dépend ici encore du logiciel utilisé. Il peut être ajouté pendant le travail de numérisation en entrant l'identificateur avant de numériser l'entité. Ou il peut être ajouté plus tard en sélectionnant l'objet de façon interactive et en ajoutant l'identificateur par l'interface d'un menu. Pour les objets polygonaux, certains systèmes exigent que l'utilisateur ajoute un point étiquette figurant dans chaque unité de surface. Bien que simple en théorie, le codage peut exiger beaucoup de temps et de ressources.

3.31. En effet, un numéro de code unique doit être attribué à chaque district de recensement. Ce numéro est utilisé dans le traitement des données pour grouper les données de dénombrement relatives aux ménages dans chaque district de recensement et pour agréger ces informations par zone administrative ou statistique en vue de leur publication. Le code numérique permet aussi d'établir un lien entre les données de recensement agrégées et la base de données numériques relative aux limites de districts de recensement emmagasinée dans un SIG. Le système de codage idéal

doit être défini pays par pays. Mais les règles appliquées pour affecter des numéros de code doivent être sans ambiguïté et conçues en collaboration avec le personnel de l'organisme national de statistique, en particulier entre ceux responsables des données géospatiales/cartographiques et ceux qui gèrent le noyau de données. Les principes les plus importants à respecter lorsqu'on définit un système de codage sont la souplesse, le caractère intuitif et la compatibilité avec d'autres systèmes de codage utilisés dans le pays. L'office de la statistique est souvent chargé d'assurer la garde des systèmes de codage du pays et il devra aussi être chargé de coordonner l'élaboration des codes cartographiques de recensement.

3.32. Un système de codage hiérarchique accroîtra en général la cohérence et la clarté des identificateurs numériques. Selon cette approche, les unités géographiques sont numérotées à chaque niveau de la hiérarchie administrative — en laissant habituellement des intervalles entre les nombres pour permettre ultérieurement l'insertion de zones nouvellement créées à ce niveau. Autrement dit, les intervalles laissés seront fonction du nombre d'unités ajoutées. Par exemple, au niveau de la province, on pourra choisir les numéros 5, 10, 15, etc. On pourra utiliser un système semblable pour les unités administratives de niveau inférieur et pour les districts de recensement. Étant donné qu'il y a souvent plus de districts dans une province que de provinces dans un pays, on pourra avoir besoin d'un plus grand nombre de chiffres aux niveaux inférieurs. L'identificateur unique de chaque unité située au niveau le plus bas — à savoir le district de recensement — est constitué simplement des identificateurs concaténés des circonscriptions administratives dont fait partie l'unité considérée.

3.33. Par exemple, un petit pays pourra utiliser le système suivant de codage :

Figure III.3

Système de codage générique des districts de recensement

1	2	0	3	5	0	1	7	5	0	0	2	3
Province		District			Localité				District de recensement			

3.34. Un numéro de code 12 035 0175 00023 de district de recensement signifie que le district de recensement 23 fait partie de la province 12, du district 35 et de la localité 175. Le numéro de code unique est enregistré dans la base de données sous la forme d'un nombre entier de grande longueur ou d'une variable en chaîne de 13 caractères. Le type de variable doit être le même dans la base de données du recensement et dans la base de données géographique. Le stockage sous la forme d'un nombre entier présente l'avantage que les sous-ensembles d'enregistrements peuvent être choisis facilement, en utilisant des commandes d'interrogation dans un système quelconque de gestion de la base de données ou son logiciel de SIG.

3.35. L'enregistrement du code sous la forme d'une variable de type caractère peut améliorer la cohérence, par exemple grâce à l'utilisation de zéros de gauche. Dans ce cas, le code est considéré comme un nom composé de caractères du Code standard américain pour l'échange d'informations (ASCII) et non comme un nombre ordinal.

3.36. Dans le cas où les unités administratives et déclarantes ne sont pas hiérarchiques, on doit élaborer des conventions de codage spéciales. Dans tous les cas, il importe d'être complètement cohérent pour définir et utiliser les identificateurs de circonscriptions administratives, car ils constituent le lien entre les limites du SIG et les données tabulaires. Le bureau de recensement devra donc tenir à jour une liste

maîtresse des districts de recensement et des circonscriptions administratives avec leurs codes respectifs et consigner tout changement apporté à la liste maîtresse dans les bases de données du SIG et du recensement. L'organisme national de statistique peut également envisager de publier une liste de districts de recensement, renseignant le codage adéquat, les latitudes et longitudes des centroïdes (points centraux), et éventuellement les populations dénombrées.

6. Éléments d'une base de données de recensement

3.37. Une base de données géographique détaillée de recensement est constituée d'une carte numérique des districts de recensement et, dans la plupart des cas, d'une série de couches de carte de base fournissant le contexte et l'orientation des cartes finales. Les couches de données de base peuvent concerner les routes de tous types, les cours d'eau, les bâtiments ou les habitats. Chacune de ces couches est contenue dans une base de données géographique en tant qu'entité distincte. Ainsi, les routes et les cours d'eau, bien que représentés par des lignes, ne seront pas enregistrés dans le même fichier numérique. Afin de garantir la cohérence, il convient de fournir des spécifications écrites précisant les produits, tailles et autres exigences.

3.38. Avant d'entreprendre l'entrée et la conversion des données, le personnel chargé d'établir les cartes de recensement devra définir la structure de tous les ensembles de données géographiques à établir. La définition de la structure consistera en une description détaillée de toutes les conventions et directives que les cartographes doivent respecter pour assurer la cohérence entre les produits finalement obtenus. Un processus de planification et une documentation appropriés éviteront des risques de confusion et d'incompatibilité au cours des opérations ultérieures.

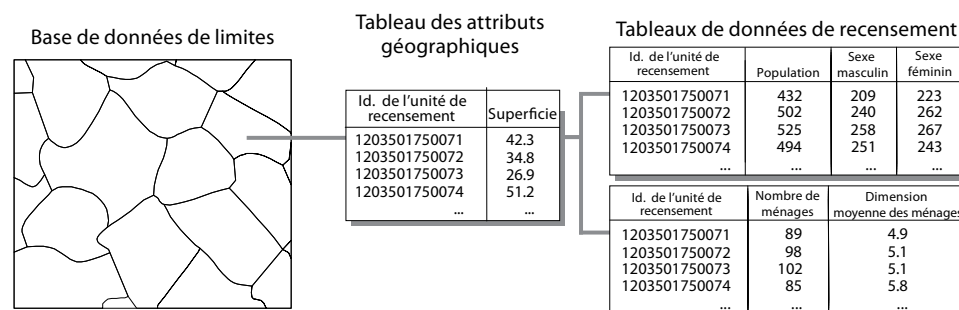
3.39. En premier lieu, on doit réfléchir à quoi ressembleront les cartes qui seront utilisées par les recenseurs. La base de données numérique complète des districts de recensement, par exemple, comportera probablement des éléments tels que des objets et attributs, qui peuvent être représentés comme suit (voir également figure III.4) :

- La *base de données des limites spatiales*, constituée d'éléments de surface (polygones) représentant les unités de recensement.
- Le *tableau des attributs géographiques*. Un fichier de base de données relié intérieurement à la base de données spatiale et contenant un enregistrement pour chaque polygone. Ce tableau contient l'identificateur particulier de chaque unité de recensement et éventuellement d'autres variables statiques, c'est-à-dire permanentes, telles que la superficie de l'unité en kilomètres carrés.
- Les *tableaux de données de recensement contenant des attributs non spatiaux* — à savoir les indicateurs de recensement pour les unités spatiales de recensement. Chacun de ces fichiers doit contenir l'identificateur particulier de l'unité de recensement, qui fournit le lien avec les enregistrements correspondants du tableau d'attributs des polygones. On aura un enregistrement pour chaque unité de recensement.
- Par ailleurs, *d'autres objets vectoriels (points ou zones)*, tels que des sites, des routes, des voies d'eau, des écoles, des installations de santé voire d'autres bâtiments peuvent guider les travailleurs sur le terrain pendant le recensement. De tels objets, enregistrés au cours du démarchage sur le terrain ou de l'établissement de la liste des foyers préliminaire, peuvent s'avérer utiles par la suite pour d'autres agences gouvernementales ou non gouvernementales, permettant une économie de temps et d'argent. La collaboration

avec d'autres utilisateurs de données procure de nombreux avantages sur le plan de l'efficacité et de l'efficacité et doit être engagée chaque fois que cela s'avère possible.

Figure III.4

Éléments d'une base de données de recensement numérique et spatiale



7. Concordance des districts de recensement avec les recensements antérieurs

3.40. Un recensement donne une image représentative de la taille et des caractéristiques de la population d'un pays. L'une des applications les plus importantes d'un recensement est l'analyse des changements de la composition démographique dans le temps. Souvent cette analyse n'est effectuée qu'à des niveaux très agrégés, par exemple au niveau national ou provincial. Mais des variations survenues dans des zones locales sont également importantes car l'évolution des petites zones influe sur les décisions de planification au niveau local. L'analyse des changements au niveau local est très facilitée si les unités de dénombrement demeurent compatibles entre recensements. Le problème posé par le changement de base géographique entre recensements est au moins aussi important que les changements de définition des rubriques du questionnaire de recensement.

3.41. En définissant la géographie du recensement, le bureau de recensement devra donc s'efforcer le plus possible de préserver les limites du précédent recensement. Du fait de l'augmentation du nombre de la population, on pourra être amené à définir de nouveaux districts de recensement. Dans ces cas, on aura toujours intérêt à subdiviser un district de recensement existant plutôt que d'en modifier les limites. Un analyste peut simplement regrouper une zone de dénombrement subdivisée pour rendre les données du nouveau recensement compatibles avec les informations d'un dénombrement antérieur. Si les limites sont modifiées, des méthodes d'adaptation plus complexes seront nécessaires.

3.42. La compilation de fichiers de compatibilité ou d'équivalence est un élément de la délimitation de districts de recensement qui peut faciliter l'analyse de changements. Ces fichiers répertorient les codes de chaque district de recensement du recensement en cours et le code correspondant tiré d'un dénombrement antérieur. Toute division ou agrégation des unités est indiquée dans les fichiers.

3.43. La base de données des limites et le tableau des attributs géographiques sont étroitement liés — en fait, ils représentent un même ensemble de données. Pendant la préparation du recensement, certaines informations liées au recensement de base comme des évaluations d'habitat et de population et des informations documentaires seront regroupées pour chaque district de recensement. Ces informations ex-

ternes concernant les unités de recensement seront enregistrées dans des tableaux de données séparées dans un système général de gestion de base de données. À partir de là, elles pourront être reliées le cas échéant aux données relatives aux limites par l'identificateur commun — le numéro de code du district de recensement — dans le tableau des attributs géographiques. Pour assurer que les bases de données de recensement obtenues au moyen du programme d'entrée et de tabulation des données correspondent bien aux fichiers de limites géographiques, il faudra établir une coopération étroite entre la section des cartes de recensement et la section informatique.

3.44. Généralement, on établit des bases de données séparées pour chaque niveau de la hiérarchie administrative ou chaque ensemble de zones statistiques pour lesquels des données de recensement sont publiées. Lorsque l'on met à jour les limites à un niveau quelconque, des changements doivent naturellement être apportés à toutes les autres bases de données contenant ces limites. La meilleure méthode consiste à apporter toutes les modifications dans la base de données maîtresse des limites au niveau d'agrégation le plus bas (à savoir la base de données du niveau des districts de recensement) et à établir chaque base de données d'unités administratives ou statistiques de niveau plus élevé en utilisant les fonctions d'agrégation courantes de SIG et de bases de données.

3.45. Certaines couches de base de données peuvent être beaucoup plus simples que la carte numérique des districts de recensement. Par exemple, pour une base de données routière, seuls quelques attributs — nom ou identificateur de la route et, le cas échéant, type de revêtement et nombre de voies — pourront être collectés. Dans ce cas, il ne sera pas forcément nécessaire d'enregistrer les informations concernant des attributs descriptifs dans un tableau séparé. Dans un souci de simplification, tous les attributs peuvent être enregistrés sur le tableau des attributs géographiques proprement dit.

3.46. À certains stades, entre les cycles de recensement et pendant les cycles, des ensembles de données de référence devront être créées. Par exemple, il devrait y avoir une version unique de la base de données des cartes de recensement du pays correspondant à chaque opération de collecte de données ou application statistique connexe. Des ensembles séparés de données agrégées relatives aux limites peuvent être établis pour chaque unité statistique déclarante pour laquelle des données sont requises. Ces ensembles de données de référence devraient être horodatés et faire l'objet d'un archivage permanent. C'est ainsi qu'on pourra avoir des ensembles de données de référence créés à partir de la même base de données maîtresse pour un recensement en 2010, une grande enquête en 2012 et une élection en 2015.

3.47. Les bases de données numériques existantes, par exemple les produits créés par un autre organisme public et les coordonnées collectées sur le terrain au moyen de systèmes de positionnement universel peuvent être importés dans une base de données de recensement géographique. On peut avoir à convertir les coordonnées de GPS de localisations de points en lignes et limites représentant des détails linéaires et polygonaux tels que des routes et des quartiers de ville. Après avoir rattaché des codes d'attributs à tous les détails de la base de données, on pourra ajouter des feuilles de cartes élaborées séparément pour créer une base de données homogène pour toute une région. La base de données obtenue montrera — en fonction du contenu des activités cartographiques — des détails physiques importants, des sites, des infrastructures, des habitats et des bâtiments individuels. À partir de ces informations, le personnel de recensement pourra délimiter interactivement les districts de recensement en utilisant comme cadre les informations géographiques de référence.

3.48. En tant que source de données pour l'établissement des cartes et l'évaluation des districts de recensement, le GPS est traité plus en détail dans le chapitre IV et à l'annexe II. Un complément d'information est par ailleurs proposé sur l'utilisation du GPS et des données de télédétection afin de corriger la base de données géographique.

3.49. Parallèlement, pendant tout le travail d'élaboration des données, le personnel de recensement devra tenir à jour une liste de toutes les circonscriptions administratives et de tous les districts de recensement délimités dans la base de données, en ce compris le nom, le niveau géographique et la référence de localisation. Cette liste informatisée est le tableau des attributs géographiques et il sera rattaché à la base de données de SIG achevée.

3.50. L'organigramme de la figure III.5 montre seulement une des nombreuses séquences possibles de la conversion de données. Les limites de districts de recensement, en particulier, peuvent être définies en divers points pendant ce processus. Par exemple, des photographies aériennes numérisées par balayage et correctement géoréférencées sont suffisamment détaillées pour qu'un opérateur puisse définir des limites numériques de districts de recensement sur l'écran, en utilisant comme fond les photographies aériennes. Les limites de districts de recensement peuvent aussi être tracées à la main sur des cartes papier appropriées et numérisées avec d'autres renseignements provenant de ces sources imprimées. D'autres activités peuvent aussi être menées dans un ordre différent. Par exemple, la plupart des logiciels de SIG prennent en charge le géoréférencement au début du processus de numérisation, ce qui rend inutile une activité supplémentaire à un stade ultérieur.

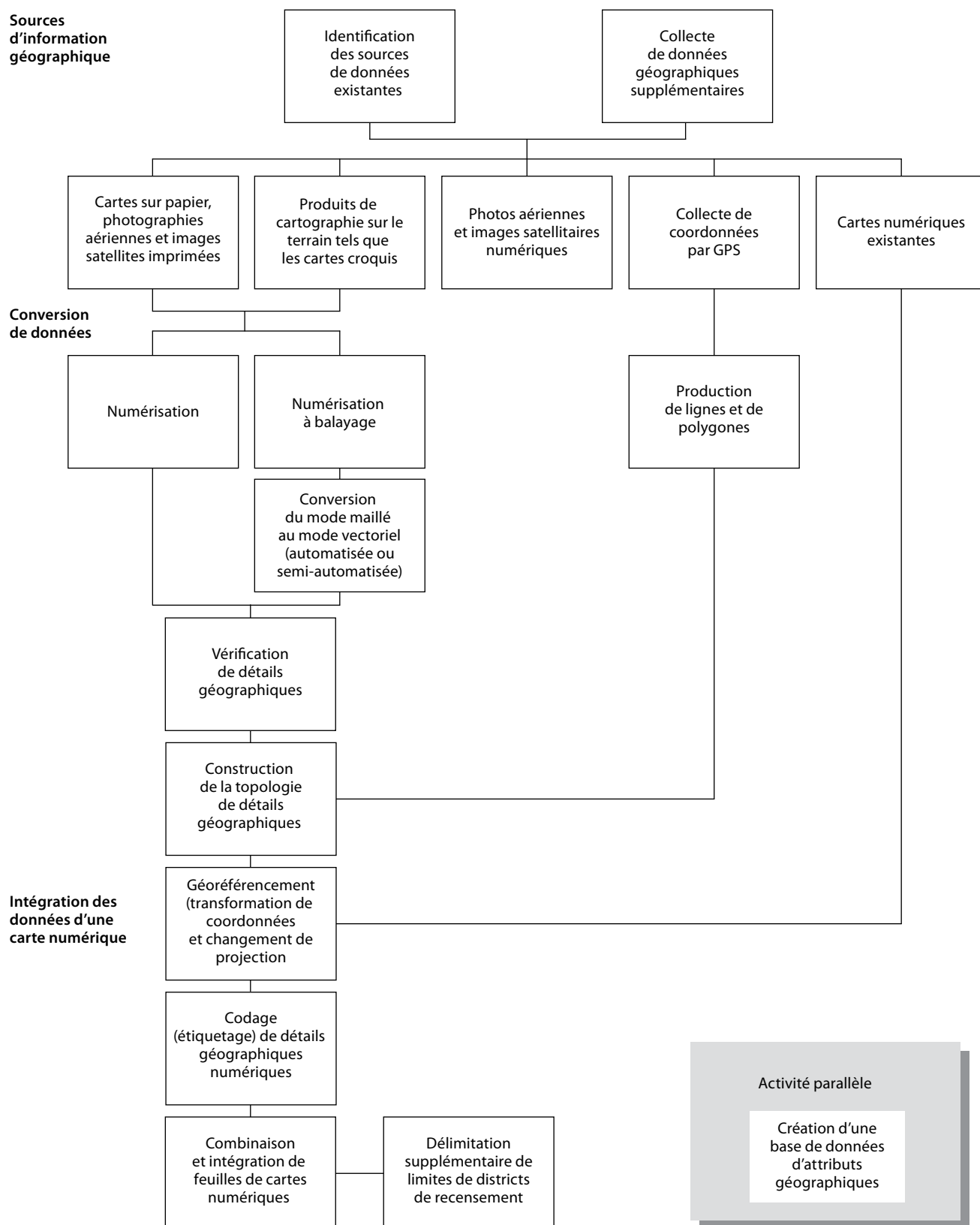
3.51. Quel que soit le processus choisi, le bureau de recensement devra évaluer la faisabilité de la méthode en effectuant une étude pilote. Cette étude comporte habituellement un test de la méthode suivie sur une petite zone échantillon. Elle permet de repérer les difficultés suffisamment tôt, de telle sorte que la technologie et les procédures puissent être mises au point, modifiées ou, dans le pire des cas, abandonnées. Les informations fournies par les tests expérimentaux aident aussi à établir le calendrier et le budget, car elles permettent de mieux évaluer les besoins en personnel et en matériel ainsi que le temps requis pour mener à bien toutes les activités.

3.52. La zone pilote devrait pouvoir représenter le plus grand nombre possible de régions du pays. En d'autres termes, elle devrait comporter un degré de variété élevé, couvrant des zones rurales et urbaines, des régions présentant des structures caractéristiques de l'habitat, des terres agricoles ainsi que des zones de végétation dense ou d'autres éléments qui empêchent la collecte de données sur le terrain.

3.53. Les vendeurs de logiciels de SIG et de matériel sont souvent prêts à apporter leur concours pour une étude pilote, car ils espèrent tirer un bénéfice de la vente de leurs produits si ceux-ci conviennent pour le projet de cartographie de recensement. Les vendeurs fournissent aussi habituellement des données de référence, ce qui est important pour des applications exigeant une grande capacité comme la production d'un grand volume de cartes ou l'accès à une base de données. Certaines techniques peuvent facilement être essayées sur une partie du territoire d'un pays. Par exemple, les récepteurs de positionnement universel sont peu coûteux et le personnel de recensement peut procéder à des évaluations des techniques de collecte de données sur le terrain. Cependant, l'achat de photographies aériennes numériques pour un site expérimental pilote de petites dimensions peut s'avérer trop coûteux. Dans ce cas, on pourra se procurer des produits plus anciens ou des échantillons de photographies aériennes concernant un pays où les conditions sont semblables.

Figure III.5

Étapes de la mise en place d'une base de données géographique en vue d'un recensement



C. Sources de données géographiques pour la délimitation des districts de recensement

1. Types de cartes nécessaires

3.54. Dans presque tous les cas, pour établir un programme cartographique de recensement, on devra consulter des cartes imprimées existantes en vue de créer une base de données cartographiques numériques ou pour mettre à jour une base de données de SIG existante. Le personnel du service cartographique doit se procurer toutes les cartes à jour du territoire du pays, y compris :

- a) Des cartes nationales d'ensemble, généralement à des échelles comprises entre 1/250 000 et 1/5 000 000, en fonction des dimensions du pays. Ces cartes doivent montrer les grandes divisions du territoire, l'emplacement des zones urbaines et des détails physiques importants comme les routes principales, les cours d'eau, les lacs, les altitudes et des points de référence particuliers. Ces cartes sont utilisées à des fins de planification;
- b) Des cartes topographiques à des échelles cartographiques élevées et moyennes. La disponibilité de cartes à ces échelles varie d'un pays à l'autre. Si certains pays ont une couverture complète au 1/25 000 ou au 1/50 000, la plus importante série de cartes complètes dans d'autres pays ne dépasse pas le 1/100 000 ou le 1/250 000;
- c) Des cartes de villes grandes et moyennes à de grandes échelles cartographiques, montrant les routes, les quartiers, les parcs, etc. Les échelles proposées peuvent varier, allant de 1/5 000 à 1/20 000, et provenir de sources différentes, dont éventuellement des cartes coloniales n'étant plus d'actualité et des plans urbains;
- d) Des cartes de circonscriptions administratives à tous les niveaux de la hiérarchie administrative;
- e) Des cartes thématiques montrant la répartition de la population à des dates de recensement précédentes ou tout détail qui pourrait être utile pour la cartographie de recensement.

3.55. Pour être incorporées à une base de données de SIG, ces cartes devraient de préférence être complétées par une documentation exhaustive. Celle-ci comprendra les renseignements géographiques de référence, y compris l'échelle de la carte, les caractéristiques géographiques et de projection, la date de compilation de la carte, le nom de l'organisme qui l'a établie et une légende complète. Cependant, même si elles ne sont pas géoréférencées correctement, les cartes sont utiles si elles fournissent des informations pertinentes pour l'élaboration des cartes de recensement, en particulier lorsqu'elles peuvent être aisément numérisées et intégrées dans un projet sous la forme d'une couche traçable à l'écran grâce au nouveau concept de « numérisation frontale ». Dans ces cas, les avantages apportés par des informations supplémentaires valent largement les ressources requises pour intégrer ces données dans la base de données du SIG de recensement et les défauts d'exactitude pouvant résulter de l'utilisation de tels produits.

2. Inventaire des sources existantes

3.56. Toutes les cartes ainsi obtenues doivent être accompagnées des documents nécessaires et structurées conformément à l'organisation du programme de cartographie de recensement, c'est-à-dire par région ou district de recensement. Outre

les sources de cartes imprimées, on disposera de plus en plus de cartes numériques provenant de nombreuses sources. Naturellement, les cartes numériques présentent l'avantage d'être plus faciles à manipuler et à adapter aux objectifs de la cartographie de recensement. Mais cela n'est pas toujours très simple. Si les documents ou méta-données font défaut, il sera souvent impossible de déterminer la projection correcte et il sera difficile d'évaluer la qualité des données (pour un complément d'information sur l'élaboration des métadonnées, voir section F ci-dessous).

3.57. Grâce aux activités soutenues par l'infrastructure nationale de données spatiales ou à des contacts personnels, il est possible de contacter les organismes et institutions suivants pour vérifier s'ils ne pourraient pas fournir des cartes imprimées ou numériques utiles :

- a) L'institut géographique/service cartographique national. Il s'agit souvent de l'organisme principal du pays en matière de cartographie. Il peut déjà avoir lancé un programme de numérisation pour des cartes topographiques. Cependant, dans certains pays, le service cartographique peut manquer des ressources ou du mandat légal nécessaires pour engager une collaboration approfondie avec l'organisme national de statistique;
- b) Les services cartographiques militaires. Dans certains pays, le principal service cartographique fait partie des forces armées. Les services cartographiques militaires ont souvent une grande expérience de la photographie aérienne et de l'interprétation des données de télédétection;
- c) Les administrations de province, de district et administrations municipales. Il peut s'agir de bureaux de planification urbains ou municipaux. Les autorités locales font de plus en plus appel à la technologie géospatiale pour gérer l'information relative aux transports, aux services sociaux, aux services d'utilité publique ainsi que les données intéressant la planification;
- d) Divers organismes publics ou privés traitant de données spatiales:
 - i) Le service d'études géologiques ou hydrologiques;
 - ii) Le service de protection de l'environnement;
 - iii) L'office des transports;
 - iv) L'autorité responsable de l'électrification rurale;
 - v) Les sociétés du secteur des services collectifs et des communications;
 - vi) Les agences d'enregistrement des biens fonciers;
- e) Les activités d'assistance. Les activités d'organismes d'aide multinationale ou bilatérale au niveau des projets comportent parfois des éléments cartographiques. Ces projets disposent souvent des moyens nécessaires pour acheter et analyser des données de télédétection ou des photographies aériennes, qui peuvent être très utiles pour le service cartographique.

3. Importation de données numériques existantes

3.58. L'importation directe de données numériques est dans la plupart des cas la forme la plus simple de conversion numérique de données spatiales. L'industrie des SIG est passée des tableaux autonomes, le plus souvent au format DBASE (.dbf), à des bases de données relationnelles, telles qu'Oracle ou Microsoft Access. Elle recourt également à des bases de données géographiques personnelles ou de fichiers. Le transfert de données est donc tributaire de l'échange de données dans des formats de fichier le plus souvent propres au fournisseur, utilisant les fonctions d'importation/exportation de logiciels commerciaux de SIG.

Encadré III.1

**Critères de sélection d'un logiciel géospatial :
LDSM, analyse des images et options FOSS****1. Options de sélection pour le logiciel SIG commercial**

Compte tenu de la pléthore de fournisseurs commerciaux, il est nécessaire que les organismes nationaux de statistique évaluent leurs objectifs opérationnels et adoptent les technologies en conséquence. L'interopérabilité logicielle peut être essentielle pour satisfaire aux besoins actuels et à venir. Diverses agences et personnes doivent être consultées pour déterminer la plate-forme logicielle privilégiée. Un choix éclairé est le meilleur choix.

Il existe deux types de logiciels commerciaux normalisés (COTS) : ceux permettant une intégration en mode maillé/vectoriel et les logiciels servant essentiellement à l'analyse des images.

2. Intégration en mode maillé/vectoriel

Parmi les logiciels, citons ArcGIS de l'Environmental Systems Research Institute (ESRI), Geomedia d'Intergraph, IDRISI, Maptitude, Geographic Resources Analysis Support System (GRASS), MapInfo de Pitney Bowes, AutoCAD et Microstation.

ESRI est le chef de file du secteur et propose une prise en charge étendue de formats, des fonctions souples adaptées à différents utilisateurs, un ensemble d'outils d'analyse très riche, une gestion de bases de données de divers types, un support étendu, des formations ainsi qu'une base de connaissances impressionnante.

GeoMedia prend en charge une multitude de types de données, propose un vaste ensemble d'outils d'analyse ainsi qu'un accès direct aux principaux formats de données géospatiales/CAO. GeoMedia incorpore des bases de données relationnelles conventionnelles et est régulièrement mis à jour avec un support complet et des formations.

IDRISI est un produit autonome prêt à l'emploi ; le code source est fourni à des fins de personnalisation. Il permet une analyse sophistiquée à base matricielle et met un ensemble d'outils d'analyse matricielle à disposition en permettant une saisie de données vectorielles exhaustive et leur exportation à travers CartaLinX.

Mapitude est à la fois un logiciel de CAO et de SIG. Il est doté d'options d'importation/exportation, d'une fonction d'appariement des adresses et est idéal pour générer rapidement des ensembles de données.

GRASS propose des fonctions vectorielles/matricielles. Il repose sur un nouveau moteur vectoriel 2D/3D topologique et permet des analyses vectorielles en réseau. GRASS était le premier SIG d'UNIX ; il offre les fonctionnalités UNIX, un code ouvert et une base globale d'utilisateurs.

MapInfo (un logiciel de Pitney Bowes) dispose de fonctions cartographiques, mais les fonctionnalités SIG sont limitées. Il s'appuie traditionnellement sur Visual Basic et des applications souples. Il est doté d'une base globale et d'un puissant outil de diffusion.

AutoCAD Map offre une intégration SIG/CAO sophistiquée, avec grille, projection et support topologique. Il met à disposition de riches fonctions de base de données ainsi qu'une analyse et un affichage en mode maillé/vectoriel. Il repose sur un système de menus mais est coûteux.

Microstation compte essentiellement des utilisateurs américains et européens. Doté d'une cartographie basée sur la CAO, ses fonctions d'analyse sont toutefois limitées au même titre que l'intégration de formats de données et le géoréférencement. Il propose une simple interface Google Earth/Google SketchUp.

3. Analyse des images

L'analyse à base matricielle est de plus en plus répandue dans les activités de cartographie de recensement. Bon nombre de logiciels d'analyse d'images peuvent désormais intégrer des données

vectérielles pour une analyse plus fine des images. Ces dernières devenant moins coûteuses et plus faciles à traiter et à intégrer à des ensembles de données vectorielles existants, davantage d'organismes nationaux de statistique achèteront des licences d'analyse d'images.

Les principales plates-formes sont : Leica Geosystems Imagine, Geomatica (PCI Geomatics), ENVI (ITT visuals), Definiens Professional et Google Earth.

Leica Geosystems Imagine est un logiciel à base matricielle conçu pour extraire des informations d'images. Il gère des collections de données géospatiales impressionnantes et offre une interaction côté client, avec des bases de données compatibles spatialement. Imagine étend les fonctionnalités à l'édition topologique de bases de données spatiales et propose un support pour couches vectorielles.

Geomatica (PCI Geomatics) traite d'importantes collections de formats de données géospatiales et offre une interaction côté client, avec des bases de données compatibles spatialement; il permet en outre de disposer de fonctions intégrales d'établissement de cartes, d'outils de gestion d'attributs pour les visualiser, éditer, interroger et analyser ainsi que de riches fonctions de traitement des images.

ENVI 4.5 (ITT Solutions) intègre les images matricielles à des systèmes d'information géographique. Les couches vectorielles peuvent être superposées sur des données d'images pour faciliter la comparaison des informations vectorielles et matricielles. ENVI offre un support relativement vaste et une grande base de connaissances. En outre, il gère de nombreux formats vectoriels (y compris les shapefiles ArcView, les fichiers d'échange ARC/INFO, DXF, les fichiers Microstation/Intergraph DGN, USGS DLG et plus encore). Son extracteur d'objet linéaire numérise automatiquement tout ce qui se trouve entre les points-graines, en suivant fidèlement les courbes, en sautant les trous et en raccordant les vecteurs au besoin.

Definiens garantit une extraction hautement sophistiquée et automatisée d'objets définis par l'utilisateur, permettant d'extraire des données géographiques de tout type d'image de télédétection. Il permet une connexion au serveur ArcGIS pour gérer des données, charger et enregistrer des données vectorielles de et vers des bases de données, avec des mises à jour simultanées de différents endroits d'un grand ensemble de données. Qui plus est, il est possible de mettre à jour des portions de vastes séries de données vectorielles. Une extension ArcGIS permet à ArcCatalog de définir un éventail de carte comme espace de travail de Definiens; les utilisateurs peuvent ainsi réviser et éditer les entrées.

L'imagerie gratuite est l'argument de vente le plus fort de Google Earth. Cependant, les mises à niveau des fonctions impliquent des coûts, comme dans le cas de Google Earth Pro. Google Earth propose une couverture mondiale et démocratise l'accès pour de nombreux utilisateurs. Cet outil convivial dispose d'une vaste base de connaissances et est utile pour un affichage rapide de données matricielles et vectorielles.

Les questions à considérer sont notamment le coût initial du logiciel, de la maintenance et des mises à niveau, la configuration LAN, les besoins en formation, la facilité d'installation, la maintenance, la documentation et les manuels, la ligne d'assistance et l'aide aux fournisseurs, les possibilités de création de programmes de correction pour la main-d'œuvre et les effectifs.

4. Logiciels libres et gratuits pour la cartographie sur ordinateur

Les systèmes d'information géographique libres (FOSS) sont une alternative aux logiciels commerciaux. Ils permettent une familiarisation à faible coût — voire gratuite — aux SIG. Ils peuvent être téléchargés librement par internet et sont dotés de fonctions similaires aux logiciels commerciaux. Avec ce type de logiciel, les utilisateurs peuvent accéder au code source de l'application, ce qui signifie que les BSN disposant d'une expertise en programmation peuvent la personnaliser en fonction de leurs besoins particuliers. Ces logiciels gagnent en convivialité, offrant la possibilité de disposer d'applications personnalisées pour des besoins particuliers. Il peut s'avérer intéressant d'essayer un logiciel de ce type même si le BSN opte finalement pour un produit commercial. Les logiciels FOSS de cartographie par ordinateur ont traditionnellement été utilisés par des programmeurs ou d'autres personnes expérimentées dans les TI. Heureusement, la base d'utilisateurs plus importante et les développements consécutifs des produits ont pu changer cela. Plus conviviaux, ces logiciels permettent

une formation en ligne et s'accompagnent d'un support technique. Ils sont garants d'une « interopérabilité », définie par la Global Spatial Data Infrastructure Association comme la « capacité à communiquer, exécuter des programmes ou transférer des données parmi diverses unités fonctionnelles de telle manière que l'utilisateur n'ait pas besoin de connaître leurs caractéristiques uniques ».

Voici quelques exemples de logiciels du genre :

Quantum GIS (<http://qgis.org>). QGIS est le logiciel FOSS vectoriel/matriciel le plus évolué ; il compte des utilisateurs sur six continents et sa série de fonctionnalités évolue rapidement. QGIS propose GRASS permettant de numériser et d'éditer des objets vectoriels. Avec ses multiples mises à niveau annuelles, les utilisateurs devancent les développements logiciels. Sa puissante communauté d'utilisateurs offre un support en ligne pour de nombreuses questions techniques.

Thuban (<http://thuban.intevation.org>). Thuban est un autre exemple de logiciel du genre. Il est doté d'une vaste base d'utilisateurs et bénéficie d'une assistance technique en ligne. Mis en œuvre à l'aide de Python, Thuban est multiplateforme. Il peut gérer à la fois des données vectorielles et matricielles et propose des fonctions de SIG complètes, dont l'identification d'objets et l'annotation, un éditeur de légende et une classification, des tableaux de requêtes et d'adhésions, une aide à la projection et un support multilingue.

Open EV (<http://openev.sourceforge.net>). OpenEV est à la fois une logithèque et une application de visualisation et d'analyse des données géospatiales matricielles et vectorielles. OpenEV permet un affichage 2D et 3D et peut reprojecter à la volée. Elle offre une puissante capacité d'analyse des images. La communauté virtuelle des utilisateurs d'OpenEV se sert des listes de discussion et des astuces pratiques pour venir en aide aux novices.

3.59. Tous les systèmes de logiciels comportent des liaisons avec d'autres formats, mais le nombre et la fonctionnalité des sous-programmes d'importation varient d'un logiciel à l'autre. On se heurte souvent à des difficultés parce que les créateurs de logiciels ne souhaitent pas publier avec précision les formats de fichiers utilisés par leurs systèmes. Par conséquent, la conversion des données peut s'avérer problématique. Les concurrents utilisent alors une forme ou une autre de rétro-technique pour découvrir les formats de fichiers exacts afin de permettre à leurs clients d'importer des fichiers extérieurs. En conséquence, les sous-programmes sont parfois instables et ils perdent fréquemment une partie des informations contenues dans les fichiers de données initiaux. Dans certains cas, il peut être préférable de passer par un troisième format de données plutôt que d'essayer d'importer directement le fichier d'échange d'un autre logiciel. Par exemple, le format d'échange de dessin d'Autocad (DXF) est pris en charge par la plupart des logiciels de SIG et assorti d'une documentation abondante. Les fonctions d'importation et d'exportation de DXF d'autres logiciels commerciaux sont donc généralement très fiables.

3.60. Les difficultés peuvent être réduites si le service cartographique ou de SIG de recensement emploie un progiciel de SIG complet et largement utilisé. Les systèmes de haut de gamme ont toutes chances d'offrir des fonctions d'importation pour un grand nombre de formats d'échange. Il est aussi plus probable que d'autres producteurs de données pourront fournir des données de SIG dans le format d'origine du logiciel de SIG. Les capacités d'importation sont un critère important pour choisir un logiciel de SIG. Une autre solution consiste à utiliser un logiciel de conversion fourni par un tiers.

3.61. À côté des problèmes de conversion de fichiers de données d'un format dans un autre, la difficulté rencontrée le plus souvent pour utiliser les données numériques existantes est l'insuffisance ou le manque de métadonnées. Sans ces informations, il est difficile d'évaluer la qualité de l'information numérique. Pire, le manque d'informations sur le cadre de référence géographique risque de rendre impossible la conver-

sion des données provenant du système extérieur de coordonnées de l'ensemble de données vers le système utilisé par l'organisme de recensement. De même, l'absence d'un livre de code ou d'un dictionnaire de données rendra difficile l'interprétation des attributs géographiques et de données inclus dans les tableaux d'attributs de l'ensemble de données du SIG. Lorsqu'il se procure des données en s'adressant à des sources extérieures, le bureau de recensement doit donc toujours insister pour qu'une documentation détaillée et des métadonnées soient également fournies.

3.62. D'autres problèmes devront peut-être être examinés : les différences de définitions et de systèmes de codage, l'utilisation de systèmes de référence cartographique différents, les échelles spatiales incompatibles et diverses normes d'exactitude qui peuvent amener à déplacer des objets qui devraient se correspondre dans deux bases de données. La recherche d'une solution à ces problèmes pourra exiger un travail considérable de traitement et de vérification pour utiliser pleinement les cartes numériques existantes.

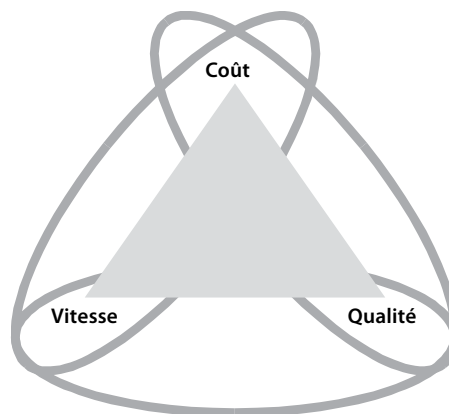
4. Conversion des données géographiques : de l'analogique au numérique

3.63. Pour mettre en place la base de données numérique de recensement, on fera appel à deux sources de données : la conversion et l'intégration de produits cartographiques existants, qui peuvent être sous forme imprimée ou sous forme numérique, et la collecte de données supplémentaires, grâce au travail sur le terrain, à des photographies aériennes ou des images satellitales. Collectivement, on utilise le terme de conversion de données pour désigner toutes ces opérations. En matière de conversion des données, il est sage de procéder à une planification prudente. En étant davantage vigilant au début du processus, moins de problèmes se poseront par la suite.

3.64. La meilleure stratégie pour convertir des données dépend de nombreux facteurs, y compris la disponibilité des données et les contraintes au niveau du temps et des ressources. Il existera toujours une corrélation négative entre le coût d'un projet, la durée requise pour réaliser la conversion des données et la qualité du produit final (voir figure III.6). On ne peut en général optimiser que deux de ces trois objectifs aux dépens du troisième. Par exemple, il est possible de créer rapidement une base de données de grande qualité, mais cela sera coûteux. De bonnes données peuvent être obtenues pour un faible coût, mais cela prendra longtemps. Sinon, on pourra créer une base de données rapidement et à moindre coût, mais la qualité du produit obtenu sera faible.

Figure III.6

Corrélations négatives dans le processus de conversion de données (D'après Hohl, 1998)



3.65. La figure III.7 décrit sommairement les étapes essentielles du travail de conversion des données qui permet d'obtenir une base numérique complète de données de recensement. L'examen de sources existantes sous forme numérique et imprimée permettra de déceler des lacunes dans les données. Les cartes existantes peuvent être périmées ou l'échelle des cartes topographiques disponibles peut être insuffisante aux fins du recensement. Pour toutes les zones où les documents existants sont de qualité insuffisante, on doit élaborer une stratégie de cartographie sur le terrain ou une autre méthode de collecte de données.

3.66. Les limites et les emplacements ponctuels de détails géographiques requis pour le recensement — emplacement des bâtiments et localisation des villages, infrastructure routière, cours d'eau et tout autre renseignement utilisé pour délimiter les districts de recensement — doivent être délimités numériquement à partir de cartes imprimées sur papier, de croquis cartographiques, de photographies aériennes imprimées ou d'images satellitales. Cela est effectué par scannage suivi de la conversion des données tramées de l'image en vecteurs ou par numérisation — tracé des détails au moyen d'un curseur du type souris. Bien que la numérisation et le scannage progressent sans cesse, c'est encore la partie la plus fastidieuse du processus de conversion des données. La capture des données est suivie d'une phase de vérification, de la construction de la topologie de la base de données du SIG et du référencement de toutes les coordonnées dans une projection cartographique appropriée (cette étape peut parfois être intégrée à des activités de numérisation).

3.67. Le processus de conversion d'objets visibles sur une carte imprimée en ponts, lignes, polygones et attributs numériques est appelé traitement automatique des données ou conversion de données. Dans de nombreux projets de SIG, c'est l'opération qui continue à exiger des ressources considérables et en particulier du temps.

3.68. La conversion de cartes imprimées ou d'informations extraites de photographies aériennes ou d'images de télédétection imprimées en une base de données numérique de SIG nécessite toute une série d'opérations. Bien que l'ordre des opérations puisse varier, les procédures requises sont semblables dans chaque cas. Les objets ponctuels et linéaires sélectionnés une fois convertis sur l'ordinateur en coordonnées numériques, il reste en général à accomplir un travail de vérification considérable pour remédier à toutes les erreurs ou omissions qui subsistent. Après cette opération, les coordonnées cartographiques initialement enregistrées dans des unités utilisées par le numériseur ou le scanner doivent être converties en coordonnées réelles correspondant à la projection cartographique de la carte source. Certains systèmes permettent de déterminer la projection avant la numérisation. Dans ce cas, les coordonnées sont converties automatiquement pendant l'opération de numérisation. Le résultat est naturellement le même.

3.69. L'opération suivante consiste à affecter des numéros de code cohérents aux objets numérisés. Par exemple, chaque ligne représentant une route recevra un numéro de code indiquant la catégorie de route (route en terre battue, route à une voie, route principale à deux voies, etc.) ou un code unique qui peut être associé par exemple à une liste de noms de rues. Dans les progiciels SIG de haut de gamme, cette opération est suivie de la structuration de la base de données (appelée aussi construction de la topologie). Au cours de cette opération, le SIG détermine les relations entre les objets dans la base de données. Par exemple, pour une base de données routières, le système déterminera les intersections entre deux routes ou plus et créera des nœuds à ces intersections. Pour les données polygonales, le système déterminera quelles lignes définissent le contour de chaque polygone. Lorsqu'il aura été constaté que la base de données numérique complète ne comportait pas d'erreurs, l'opération finale consis-

tera à lui ajouter des attributs additionnels. Ces derniers peuvent être rattachés en permanence à la base de données, sinon les informations additionnelles relatives à chaque objet de la base de données peuvent être stockées dans des fichiers séparés, rattachés le cas échéant à la base de données géographique.

3.70. Les deux principales méthodes de conversion des informations figurant sur les cartes imprimées en données numériques sont la numérisation manuelle et la numérisation par balayage optique ou scannage. Le scannage est une opération automatique consistant à convertir une carte en une image numérique matricielle qui peut être convertie par la suite en vecteurs numériques. La numérisation par contre suppose que l'on trace tous les objets ponctuels et linéaires requis sur une carte au moyen d'un curseur ou d'une souris. Les techniques de numérisation à l'écran servent à dessiner de nouvelles couches de cartes, à l'aide de cartes ou images numérisées. Le contenu peut également être numérisé à partir de feuilles de cartes annotées. Les deux approches sont examinées plus bas de façon détaillée.

a) Scannage

3.71. Pour de nombreuses tâches d'introduction de données, on peut dire que le scannage a dépassé la numérisation comme principale méthode de saisie de données spatiales. Cela s'explique essentiellement par sa capacité à automatiser certaines étapes fastidieuses liées à l'entrée des données, à l'aide de scanners à feuilles grand format et de logiciels de vectorisation interactifs. Il existe différents types de scanners mais pour l'essentiel tous fonctionnent de la même façon. La carte est placée à l'envers sur la surface de balayage optique où la lumière est dirigée sur la carte sous un certain angle. Un dispositif photosensible enregistre l'intensité de la lumière reflétée pour chaque cellule ou pixel d'une trame très serrée. Dans une échelle de gris, l'intensité lumineuse est convertie directement en une valeur numérique, par exemple un nombre compris entre 0 (noir) et 255 (blanc). En mode binaire, l'intensité lumineuse est convertie en valeur de pixel blanc (0/1) ou noir en fonction d'un seuil d'intensité lumineuse. Dans les scanners en couleurs, le dispositif photosensible comporte trois éléments sensibles respectivement au rouge, au vert et au bleu. L'intensité relative des trois signaux colorés combinés détermine la couleur du pixel. Le scannage donne une image matricielle de la carte initiale, image qui peut être enregistrée dans un format d'image standard tel que GIF ou TIFF. Après avoir géoréférencé l'image — ce qui suppose que l'on précise les coordonnées d'un angle de l'image et les dimensions de pixels en unités du monde réel — on peut l'afficher dans de nombreux logiciels de SIG en tant que toile de fond pour les données vectorielles existantes. Mais, en général, les objets géographiques figurant sur l'image sont extraits manuellement ou automatiquement et convertis en données vectorielles.

3.72. Trois types principaux de scanners sont d'un usage courant :

- Les **scanners feuille à feuille** sont actuellement le type de scanner le plus couramment utilisé pour des applications SIG à grande échelle. Dans ces scanners, le système de capteurs est immobile. C'est au contraire la carte qui est déplacée devant la barrette de capteurs. Leur précision est moindre que celle obtenue avec les scanners à tambour, car il est plus difficile de guider avec précision l'introduction de la carte que de contrôler le déplacement des capteurs du scanner. Mais leur précision est en général suffisante pour des applications de SIG, leur coût est moindre et ils produisent couramment des images en moins de cinq minutes. Il ne faut pas oublier que des documents anciens ou fragiles peuvent être endommagés par les rouleaux du scanner feuille à feuille (voir figure III.7).

Figure III.7

Photo d'un scanner feuille à feuille

Source : Ideal.com.

- Les **scanners à plat** ou **scanners de bureau** sont couramment rencontrés dans de nombreux bureaux. Leur format est relativement restreint et par conséquent les grandes cartes doivent être scannées en plusieurs morceaux, que l'on doit ensuite réunir dans l'ordinateur. Le document étant posé à l'envers sur une plaque de verre, les capteurs et la source lumineuse se déplacent le long du document sous la plaque. Ce qui fait l'intérêt des scanners à plat, c'est leur faible coût et le fait qu'ils soient faciles à installer et à entretenir. Ils sont utiles pour scanner des documents textuels — par exemple des tableaux de données — qui sont ensuite interprétés au moyen d'un logiciel de reconnaissance optique de caractères. Ils permettent aussi d'introduire de petits graphiques et des cartes dans l'ordinateur. Ils conviennent moins bien pour des opérations de conversion de cartes à grande échelle, qui nécessitent le scannage de nombreuses cartes topographiques et thématiques de grand format. Le scannage par section de ces cartes et l'assemblage ultérieur des éléments dans l'ordinateur prennent beaucoup de temps et peuvent occasionner de nombreuses erreurs.
- Les **scanners à tambour** sont plus chers et sont utilisés pour des applications professionnelles exigeant une très grande précision (par exemple photogrammétrie ou applications médicales). La carte est fixée sur le tambour rotatif. Un système de capteurs se déplace ensuite le long de la carte et enregistre l'intensité lumineuse ou la couleur de chaque pixel. Les scanners à tambour assurent une très grande précision, mais sont aussi très coûteux et très lents. Un seul scannage peut prendre de 15 à 20 minutes.

3.73. Les paramètres du scanner choisis par l'opérateur ont de grandes incidences sur les caractéristiques de l'image de sortie. Le choix des paramètres optimaux exige une certaine expérimentation, car il dépend des options du scanner, des caractéristiques des cartes ou photographies de base que l'on scanne et des opérations ultérieures prévues. Les paramètres les plus importants sont :

- **Le mode de scannage.** Le mode binaire ou « dessin au trait » convient pour les dessins ou croquis monochromes ainsi que pour les séparations de couleurs où toutes les entités sont essentiellement du même type. Le mode échelle de gris permet d'enregistrer les variations sur une carte et on peut ensuite manipuler l'image pour en extraire seulement les entités présentant un certain pouvoir réflecteur dans un système de traitement de graphismes ou d'images. Cela est même encore plus facile lorsque les cartes sont scannées en mode couleur, dans lequel, par exemple, toutes les entités dessinées en vert sur la carte peuvent être extraites au moyen de quelques commandes simples.
- **La résolution de l'image** est mesurée en points par pouce (ppp). Les résolutions de scannage courantes sont comprises entre 100 et 400 ppp (cependant les photographies aériennes sont en général scannées à des résolutions plus élevées sur des scanners spécialisés). Une résolution de scannage élevée préserve davantage de détails sur la carte initiale et donne des lignes plus lissées dans le lot de données vectorisées du SIG. Mais les images résultantes sont plus grandes et exigent plus de mémoire et d'espace disque; le doublement de la résolution de scannage se traduit par un quadruplement des dimensions de l'image. Le choix dépend des caractéristiques du document source, du matériel disponible et de l'utilisation prévue de l'image obtenue.
- **Luminosité, contraste et seuil.** Ces paramètres déterminent l'apparence de l'image obtenue. La luminosité détermine les valeurs de blanc et de noir de l'image. Le contraste est utilisé pour déterminer comment les valeurs de gris ou les fines nuances de couleurs sont reproduites. Un contraste élevé rend l'image plus nette, mais peut aussi entraîner une perte de différences et de détails. Le seuil est un paramètre utilisé en mode binaire pour déterminer comment les valeurs de gris de l'original sont converties en pixels noirs ou blancs. Le choix des paramètres peut être très différent selon que le scannage est destiné à fournir une représentation visuellement attrayante et exacte du document source ou à permettre sa vectorisation ultérieure. Dans ce dernier cas, un supplément de contraste ou de luminosité peut mettre en lumière des entités sur la carte et faciliter ainsi la conversion ultérieure en format vectoriel.
- **Correction de gamma.** Le contrôle de la luminosité et du contraste donne de bons résultats si les valeurs de pixels de l'image sont réparties très régulièrement sur toute la gamme des gris. Mais souvent ce n'est pas le cas. Par exemple, l'image peut comporter surtout des zones très claires et très sombres. La correction de gamma est une technique qui prend en compte la répartition des valeurs de gris dans l'image et assure la correction automatique, éclaircissant ou assombrissant certaines zones ou élargissant la gamme de valeurs de gris des pixels. Cette technique peut souvent aider à préserver de faibles différences dans l'image.

3.74. Le scannage du document source n'est qu'une première opération et une opération très simple. Le résultat final du processus de conversion étant une base de données géographique numérique de points et de lignes, les informations scannées contenues dans l'image matricielle doivent être converties en coordonnées. Ce processus est appelé « conversion de mode maillé en mode vecteur ». Jusqu'à une date récente, cette opération était le point faible du scannage et c'est pourquoi on donnait la préférence à la numérisation en tant que mode d'entrée de données. Les perfectionne-

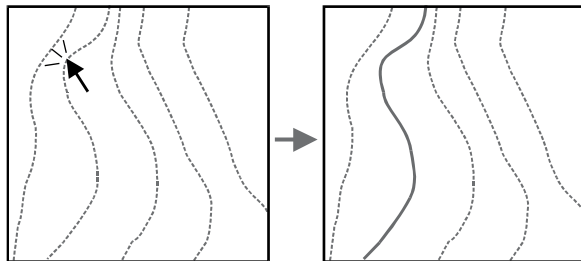
ments récents concernant l'élaboration des logiciels, les techniques de reconnaissance des formes et les vitesses de traitement en ont incité beaucoup à privilégier le scannage comme méthode de saisie des données.

3.75. La conversion de mode maillé en mode vecteur peut être effectuée en mode automatique, semi-automatique ou manuel. En mode automatique, le système convertit automatiquement toutes les lignes de l'image matricielle en suite de coordonnées. Étant donné que les lignes épaisses sur la carte donnent sur l'image matricielle des lignes larges de plusieurs pixels, le processus automatisé de conversion commence par un algorithme d'amincissement de ligne. L'opération suivante consiste à déterminer les coordonnées de chaque pixel qui définit la ligne et à éliminer ensuite éventuellement les coordonnées redondantes — c'est-à-dire les lignes droites qui peuvent être représentées par un plus petit nombre de coordonnées. Le logiciel de conversion permet aussi en général à l'utilisateur de spécifier des niveaux de tolérance. Par exemple, des entités constituées seulement d'un ou de plusieurs pixels peuvent en fait correspondre à des taches sur les cartes sources et peuvent être supprimées automatiquement. De même, si l'image a été scannée au moyen d'un scanner couleur, le logiciel de conversion mode maillé/mode vecteur permet souvent à l'utilisateur de spécifier les codes de ligne à affecter aux couleurs. Cela est utile pour extraire divers types d'entités dans des couches séparées de données de SIG. Par exemple, les cours d'eau peuvent être représentés en bleu sur la carte source, tandis que les routes sont tracées en noir et les limites d'unités administratives en rouge.

3.76. En mode semi-automatique, l'opérateur clique sur chaque ligne à convertir (voir figure III.8). Le système trace alors cette ligne jusqu'aux intersections les plus proches et la convertit en une représentation vectorielle. Cela présente l'avantage que l'opérateur peut se limiter à sélectionner un seul sous-ensemble d'entités sur la carte, toutes les routes par exemple, mais pas les cours d'eau. Enfin, en mode manuel, l'image matricielle scannée est simplement utilisée comme arrière-plan sur l'écran de l'ordinateur. Les coordonnées sont créées en traçant les entités sur l'image scannée au moyen d'une souris, de la même façon que pour la numérisation frontale mentionnée plus haut.

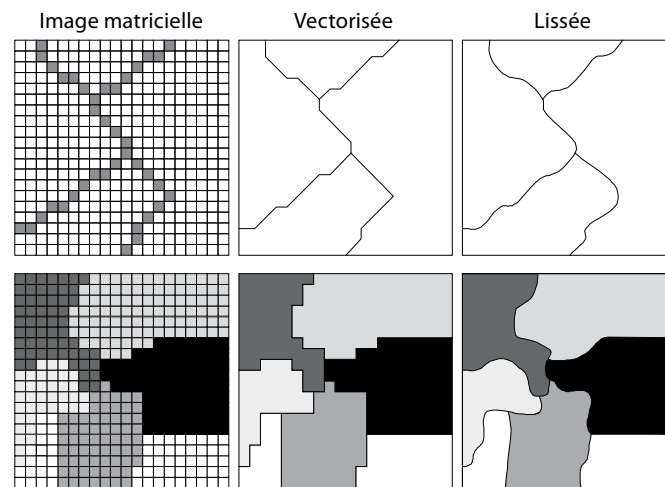
Figure III.8

Vectorisation semi-automatique



3.77. Si l'on convertit automatiquement en format vectoriel des entités linéaires ou surfaciques à partir d'images matricielles de résolution relativement faible, le trait obtenu peut présenter des discontinuités peu naturelles. Il est pratique courante de lisser les données vectorielles en utilisant des fonctions de « spline » ou de généralisation disponibles dans les logiciels de SIG. La figure III.9 montre des exemples d'un lot de données linéaires et polygonales.

Figure III.9

Vectorisation et lissage de données image scannées*i) Considérations supplémentaires*

3.78. Divers éléments doivent être pris en considération lors de la préparation d'un projet de conversion de données fondé sur le scannage de cartes. Une préparation appropriée de la carte de base avant le scannage peut améliorer sensiblement la qualité des produits obtenus. Les cartes doivent être planes et propres. Tout reste de ruban adhésif qui pourrait se trouver sur la carte doit être enlevé, car il risque de laisser des traces à la surface du scanner. Les traits peu apparents sur la carte peuvent être renforcés à la plume ou au marqueur. De même, l'opérateur peut retracer des symboles de ligne sélectionnés et remplir les polygones marqués de hachures croisées pour obtenir des lignes pleines et des remplissages qui faciliteront la vectorisation automatique. Mais ces changements peuvent aussi être apportés sur l'image scannée avant la vectorisation. N'importe quel logiciel graphique à base matricielle peut être utilisé à cette fin. Cependant, il est souvent plus facile d'apporter ces modifications à la main. On devra utiliser un marqueur à l'eau ou un crayon à la cire car les marqueurs à base d'huile minérale risquent de salir la surface de la vitre du scanner et les marques de crayons au graphite reflètent la lumière de telle sorte qu'elles peuvent devenir invisibles. En ce qui concerne les photographies, le papier mat donne de meilleurs résultats que le papier brillant.

3.79. Il est souvent prévu une opération supplémentaire pour la conversion de cartes relativement complexes présentant de nombreux objets différents (par exemple des cartes topographiques) ou de cartes de mauvaise qualité. En ce qui concerne ces sources de données cartographiques, la précision peut être améliorée et le travail après traitement peut être réduit si l'on trace d'abord tous les objets cartographiques requis sur un support translucide comme du mylar. Bien qu'il augmente la charge de travail de l'opérateur, le calquage s'avère souvent plus rapide en fin de compte, car il réduit la durée du travail de vérification et de correction des erreurs. Le document source calqué et ensuite scanné est plus clair et ne contient que les objets réellement nécessaires. Cette technique est employée dans la plupart des applications de scannage professionnel à grande échelle. Le calquage peut être évité si les clichés originaux de sélections de couleurs des cartes sources publiées sont disponibles. On pourra souvent se les procurer pour les séries de cartes topographiques nationales. Chaque séparation ne contient qu'un sous-ensemble des objets de la carte imprimée, ce qui facilite beaucoup la séparation des objets dans différentes couches de données.

3.80. Malgré ces opérations préalables, il peut s'avérer nécessaire d'effectuer un traitement complémentaire avant de passer à l'exécution des programmes de vectorisation. Ce traitement peut consister à améliorer encore l'image en précisant les contours, en renforçant le contraste ou encore en éliminant les petites taches voire en effectuant des modifications interactives au niveau des pixels. Un logiciel graphique orienté raster ou le logiciel de vectorisation lui-même offriront les fonctions nécessaires.

3.81. Les logiciels de SIG qui prennent en charge les données matricielles comportent des programmes de conversion mode maillé-mode vectoriel. Ils sont surtout destinés à convertir les données en mode maillé de SIG en données vectorielles de SIG et non pas à convertir des images scannées complexes en objets vectoriels simples. Pour un projet de vectorisation à grande échelle, un logiciel spécialisé est plus approprié. Il existe actuellement plusieurs logiciels commerciaux et non commerciaux de conversion mode maillé-mode vectoriel (par exemple Vextractor, AbleVector et PTracer), ainsi que des extensions logicielles telles qu'ESRI ArcScan. Les options disponibles diffèrent d'un produit à l'autre. Certains produits offrent des fonctions de réaligement des images scannées ou de reconnaissance optique de caractères des annotations cartographiques qui peuvent être enregistrées comme attributs des objets vectoriels obtenus. Les prix varient dans de grandes proportions. Le personnel chargé d'effectuer la conversion devra donc comparer soigneusement les possibilités et fonctions disponibles et les exigences des tâches de conversion de données.

ii) *Avantages et inconvénients du scannage*

3.82. Les avantages du scannage sont les suivants :

- Les cartes scannées peuvent être utilisées comme arrière-plan pour les informations vectorielles. Par exemple, des cartes topographiques scannées peuvent être utilisées en combinaison avec des limites de districts de recensement numérisées pour établir des cartes pour agent recenseur, à travers la numérisation frontale.
- Les fonds de cartes transparents ou les sélections de couleurs originales peuvent être vectorisés assez facilement en utilisant des logiciels de conversion mode maillé-mode vectoriel.
- Les scanners de petit format sont relativement peu coûteux et assurent une saisie rapide des données.

Les inconvénients sont les suivants :

- La conversion de grandes cartes avec des scanners de petit format rend nécessaire un réassemblage fastidieux des divers éléments.
- Si les scanners de grand format à grand débit sont coûteux, ce coût peut être justifié s'ils sont utilisés pour un scannage de cartes et une vectorisation à grande échelle afin de créer une base de données géographique numérique.
- En dépit des perfectionnements apportés récemment aux logiciels de vectorisation, il peut encore y avoir à effectuer à la main un travail considérable de correction et d'étiquetage d'attributs.
- Le scannage de volumes importants de cartes imprimées compliquera l'archivage des fichiers sur bon nombre de systèmes informatiques de bureau. Les organismes nationaux de statistique envisageant de scanner toutes leurs cartes de districts de recensement devraient penser à investir dans des disques durs séparés avec un système de sauvegarde afin de gérer les volumes de fichiers importants générés.

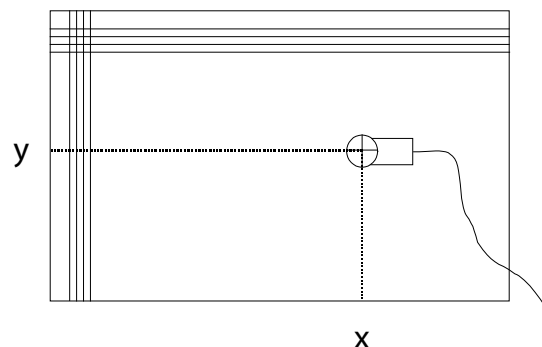
b) Numérisation

3.83. La numérisation manuelle a été la première méthode la plus courante de conversion de données spatiales. Elle nécessite une table à numériser de dimensions variables : de petites tables de 30 x 30 cm aux grandes tables de 120 x 180 cm. Les tables à numériser de grandes dimensions facilitent la numérisation de grandes feuilles de cartes. Sur une petite tablette, une grande carte doit être numérisée en plusieurs morceaux qui doivent être combinés par la suite. Au cours de la numérisation, la carte est fixée sur la table au moyen de ruban de masquage. De préférence, la carte doit être plate, sans déchirures ni plis. Souvent le papier rétrécit, surtout dans une ambiance humide et ce rétrécissement provoque des déformations qui vont être reportées dans la base de données cartographique numérique.

3.84. La première opération consiste à déterminer un nombre de points de repère précisément définis sur la carte (en général au moins quatre). Ces points de repère ont un double objet. En premier lieu, si une grande carte est numérisée en plusieurs opérations et si la carte doit être de temps à autre retirée de la table à numériser, les points de repère permettent d'enregistrer la carte avec exactitude sur la table à numériser. En second lieu, on choisit des points de repère dont on connaît les coordonnées réelles dans le système de projection de la carte de base. On peut adopter comme points de repère appropriés les intersections du quadrillage des latitudes et des longitudes figurant sur de nombreuses cartes topographiques. Lors de l'opération de géoréférencement qui précède ou qui suit la numérisation des objets ponctuels et linéaires, ces informations sont utilisées pour convertir les coordonnées mesurées en centimètres ou en pouces sur la tablette à numériser en coordonnées réelles — exprimées en général en mètres ou en pieds — de la projection cartographique. Après avoir choisi les points de repère, l'opérateur trace les objets linéaires sur la carte en utilisant un curseur qui communique avec la table à numériser. La table contient un réseau de fils (dont une partie est représentée sur la figure III.10). Ce réseau crée un champ électromagnétique. Le curseur contient une bobine métallique de sorte que la table à numériser et le curseur se comportent comme un émetteur et un récepteur. Cela permet au curseur de déterminer les fils les plus proches dans les directions x et y. La position exacte est trouvée avec un grand degré de précision par interpolation. Les objets numérisés sont immédiatement tracés sur l'écran de l'ordinateur. Cela permet à l'opérateur de vérifier quelles limites ont été saisies et si des erreurs importantes ont été introduites.

Figure III.10

Table à numériser



3.85. Les coordonnées sont enregistrées en mode point, en mode distance ou en mode contrôlé. En mode point, l'opérateur presse un bouton sur le curseur chaque fois qu'une ligne change de direction. Pour les lignes courbes, le nombre de coordon-

nées enregistrées déterminera dans quelle mesure la courbe apparaîtra lissée dans la base de données du SIG. En mode distance, une coordonnée est automatiquement enregistrée lorsque l'opérateur a déplacé le curseur d'une certaine distance. Enfin, en mode contrôlé, le curseur enregistre automatiquement les coordonnées à des intervalles de temps spécifiés à l'avance. Avec le mode distance et le mode contrôlé, on court le risque d'enregistrer des segments de lignes complexes comportant de nombreuses courbes avec trop peu de coordonnées. Par contre, les segments longs et rectilignes peuvent donner de nombreux points redondants. Le mode point, qui laisse le choix de la densité des coordonnées, est généralement le mode de numérisation préféré des opérateurs expérimentés.

3.86. La numérisation est pour les opérateurs un travail fastidieux et fatigant. Il faut donc non seulement avoir des opérateurs très qualifiés, mais aussi leur assurer un environnement de travail approprié, comportant un système de table à numériser de conception ergonomique. Des macro-instructions cohérentes de logiciel de SIG guidant l'opérateur et des procédures de contrôle de la qualité permettront de réduire au minimum les erreurs en cours de numérisation et de gagner du temps lors des corrections ultérieures.

3.87. Pendant la numérisation, l'opérateur peut affecter des codes objets à chaque ligne ou à chaque point saisi. Par exemple, il est possible d'affecter des codes aux divers types de limites administratives, de un pour les limites provinciales à trois pour les limites de districts. Dans certains systèmes de SIG à structure topologique, l'utilisateur doit aussi ajouter un point étiquette à chaque polygone numérisé. Cela peut être fait manuellement pendant la numérisation ou automatiquement avant que la topologie ne soit construite. Ce point étiquette établit le lien entre le polygone et la table des attributs géographiques qui contient des données relatives au polygone (voir annexe I).

3.88. On appelle parfois numérisation frontale une méthode particulière d'entrée de données n'exigeant pas de tablette à numériser. La numérisation frontale fait désormais référence à deux concepts. Selon l'ancienne méthode, l'opérateur traçait les entités cartographiques sur un transparent et attachait la carte obtenue sur l'écran de l'ordinateur. En utilisant un module d'entrée de données de SIG ou simplement un logiciel graphique prenant en charge un format graphique compatible avec le SIG, les lignes ou les points peuvent être numérisés à l'aide de la souris (voir figure III.11). Avec la nouvelle méthode de numérisation frontale, une image de carte scannée est utilisée numériquement pour tracer les limites dans une couche du SIG. L'opérateur utilise une carte scannée, une photographie aérienne ou une image satellitale comme arrière-plan. L'image a été géoréférencée en ce sens qu'elle a été convertie dans un format à l'aide du même système de coordonnées réelles que pour les autres couches du projet de SIG. L'analyste convertit l'image à l'aide de points de contrôle et en « attachant » l'image à des emplacements connus à la fois dans les couches restantes et dans le monde réel. Les intersections de rues et les points de repère font figure de points de contrôle corrects. L'analyste trace ensuite les entités avec une souris sur l'image scannée, créant ainsi une nouvelle couche.

Avantages et inconvénients de la numérisation

3.89. Les avantages de la numérisation sont les suivants :

- L'apprentissage de la numérisation étant facile, il n'est pas nécessaire d'avoir une main-d'œuvre spécialisée d'un coût élevé.
- Les attributs peuvent être ajoutés pendant la numérisation.

- La numérisation manuelle permet une grande précision; il n'y a en général aucune perte de précision par rapport à la carte source.

3.90. Les inconvénients sont les suivants :

- La numérisation étant fastidieuse, la fatigue de l'opérateur peut causer des problèmes de qualité rendant nécessaire un travail considérable de vérification et de correction.
- La numérisation manuelle est très lente. Pour effectuer un travail de conversion de données à grande échelle, il faudra éventuellement prévoir un grand nombre d'opérateurs et de tables à numériser.
- Contrairement à ce qui se passe lors de la collecte de données primaires au moyen du GPS ou de photographies aériennes, la précision des cartes numérisées est limitée par la qualité des documents sources.

Figure III.11

Numérisation frontale



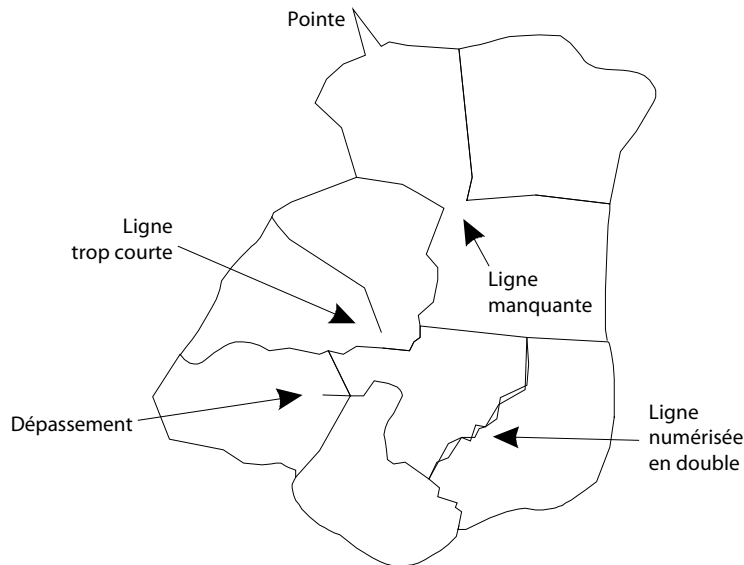
c) Édition

3.91. L'objet de la conversion analogique — numérique des informations géographiques est d'obtenir une représentation exacte des données de la carte d'origine. Cela signifie que toutes les lignes qui se raccordent sur la carte doivent aussi se raccorder dans la base de données numérique. Il ne doit y avoir ni omission d'objets ni lignes en double. La numérisation manuelle est souvent source d'erreurs. Les types d'erreurs les plus courants sont représentés sur la figure III.12. En cas de tracés trop courts et de lignes manquantes, l'effet net est de créer un polygone lorsque deux zones existent. Lorsque des lignes sont numérisées deux fois, l'effet net consiste à créer un ou plusieurs polygones s'il n'en existe pas. De même après la conversion de mode maillé en mode vectoriel, les segments de ligne interrompus doivent être reliés à la main. Cela se produit par exemple lorsque de petites routes ou de petits cours d'eau représentés par une ligne étroite traversent de grandes routes représentées par des lignes épaisses.

Si les routes ou cours d'eau secondaires sont extraits dans une couche cartographique séparée, il y aura des lacunes dans le réseau routier aux intersections avec les routes principales.

Figure III.12

Quelques erreurs courantes de numérisation



3.92. Certaines erreurs courantes de numérisation présentées sur la figure III.12 peuvent être évitées en utilisant les tolérances du logiciel de numérisation dites tolérances de raccordement, qui sont définies par l'utilisateur. Par exemple, l'utilisateur peut spécifier que tous les points extrêmes d'une ligne qui se trouvent à moins de 1 mm d'une autre ligne seront automatiquement raccordés à cette ligne. Les petits polygones minces créés lorsqu'une ligne est numérisée deux fois peuvent aussi être automatiquement éliminés. Cependant, quelques-uns seulement de ces problèmes peuvent être résolus de cette façon. La correction manuelle des erreurs de numérisation après une comparaison approfondie de l'original et de la carte numérisée reste un élément nécessaire du processus de conversion de données.

d) Élaboration et gestion de la topologie

3.93. Afin de transformer des données géographiques pour que les points deviennent des nœuds de polygones, il est nécessaire de les définir de manière à ce qu'elles connaissent leur relation avec les autres objets. C'est le concept de « topologie », défini comme l'étude des propriétés de figures géométriques non altérées par une déformation. Des données topologiques correctes sont différentes de tableaux ou d'objets graphiques dénués de toute topologie (souvent appelés « spaghetti »). Les objets ayant une topologie connaissent leurs emplacements dans l'espace absolu ainsi que leurs voisins les plus proches. La construction de la topologie de la carte numérique contribue à faciliter le processus de correction. Par exemple, elle permet à l'utilisateur de repérer des problèmes comme celui des polygones qui ne sont pas complètement clos. La topologie des objets décrit les relations spatiales entre des objets géographiques solidaires ou adjacents tels que des routes qui se raccordent à des points d'intersection (voir annexe I sur les SIG). La structuration topologique d'une base de données de SIG

comporte l'identification de ces relations spatiales et leur description dans la base de données. Ce processus dépend du type de logiciel utilisé. La mise en mémoire des informations topologiques facilite l'analyse, car de nombreuses opérations utilisant les SIG n'exigent pas en fait de coordonnées, mais sont fondées uniquement sur la topologie. Par exemple, les voisins d'un district peuvent être déterminés à partir d'un tableau de base de données indiquant pour chaque ligne le polygone de droite et le polygone de gauche (voir annexe, paragraphe A1.5).

3.94. L'utilisateur n'a pas en général à se préoccuper de la façon dont le SIG enregistre les informations topologiques. À condition que la base de données numériques ne comporte pas d'erreurs ou de lacunes — toutes les lignes étant reliées entre elles et les polygones étant correctement identifiés — on utilise une fonction de SIG pour construire la topologie et créer tous les fichiers de données internes nécessaires. Cette fonction ne donnera de résultats satisfaisants que si la base de données cartographique ne contient pas d'erreurs. La construction de la topologie permet ainsi de vérifier l'intégrité de la base de données.

e) Intégration de la carte numérique

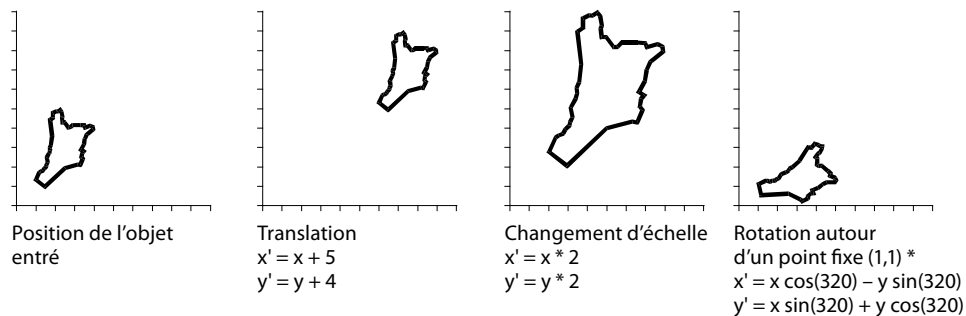
3.95. Pour établir un projet cartographique à des fins de recensement, on devra faire appel à toutes les sources de données géospatiales disponibles. Celles-ci seront probablement stockées sous différents formats, en utilisant différentes échelles et projections cartographiques. L'intégration de ces sources de données hétérogènes exige une connaissance approfondie des méthodes d'intégration de données de SIG si l'on veut obtenir une base de données cartographique numérique complète et homogène. On examinera dans les sections suivantes les méthodes les plus importantes qui facilitent l'intégration des données cartographiques numériques.

f) Géoréférencement

3.96. Le géoréférencement consiste à attribuer des coordonnées d'un système de référence connu, telles que les latitudes-longitudes, aux coordonnées de page d'une carte (image) matricielle. Les logiciels de SIG comportent presque tous les fonctions nécessaires au géoréférencement. L'utilisateur doit spécifier un certain nombre de points de contrôle dont on connaît les coordonnées du monde réel. À partir des coordonnées d'entrée des dispositifs de numérisation et des coordonnées de sortie du monde réel, le système calcule un ensemble de paramètres permettant d'effectuer les transformations ci-après (voir figure III.13) :

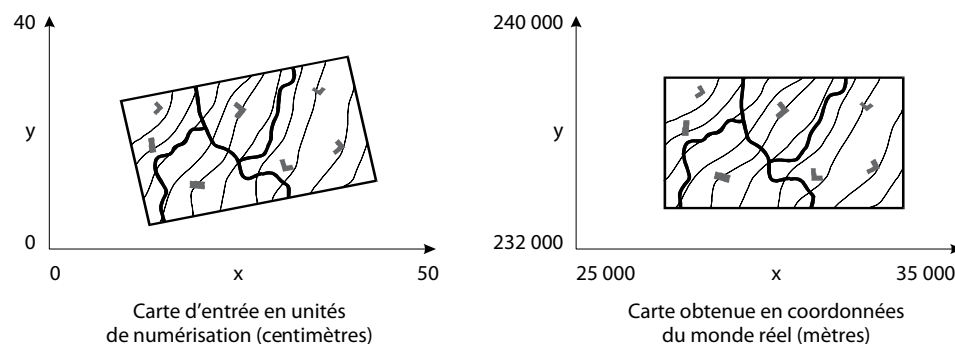
- **Translation.** L'objet géographique est déplacé vers une nouvelle position en ajoutant (ou en retranchant) simplement des valeurs constantes aux coordonnées x et y. Le décalage est en général différent pour les coordonnées x et y.
- **Changement d'échelle.** Les dimensions de l'objet sont agrandies ou réduites en multipliant les coordonnées x et y par un facteur approprié pour chacune d'elles. Le changement d'échelle est en général effectué à partir de l'origine du système de coordonnées.
- **Rotation.** On fait tourner d'un certain angle l'objet géographique autour de l'origine du système de coordonnées. Cette rotation donne à la carte numérique l'orientation appropriée même si la carte sur papier n'a pas été correctement alignée sur la table à numériser.

Figure III.13

Translation, mise à l'échelle et rotation

3.97. Il faut noter que la forme des objets numérisés n'est pas modifiée par cette transformation comme elle le serait en cas de changement de projection. Seules les dimensions relatives et l'orientation des objets sont modifiées. Lorsque les paramètres corrects de translation, de mise à l'échelle et de rotation ont été calculés, le système applique ces paramètres à toutes les coordonnées de points et de lignes dans la base de données. Il en résulte une carte qui ressemble beaucoup à la précédente, mais qui est maintenant enregistrée dans le système de coordonnées approprié qui avait été utilisé pour établir la carte de base initiale (voir figure III.14). Il importe de veiller à réduire au minimum les erreurs au cours de cette opération. Le système fournit en général des informations sur l'erreur commise dans l'estimation des paramètres de transformation pour chaque point, ce qui aide à déceler les erreurs lorsqu'on définit les coordonnées du monde réel des points de contrôle. Des renseignements techniques plus complets sont fournis à l'annexe II.

Figure III.14

Carte en unités de numérisation; carte en coordonnées du monde réel

3.98. On se trouve placé devant un problème difficile lorsqu'on ignore quel était le système de projection et de coordonnées de la carte source sur papier. C'est malheureusement un problème fréquent car de nombreuses cartes imprimées, surtout des cartes thématiques, ne comportent pas ces informations. On peut alors envisager deux solutions : essayer un grand nombre de projections cartographiques possibles (une possibilité intéressante est offerte par la projection normale utilisée dans les programmes cartographiques du pays) ou effectuer le type de correction spatiale appelé « étirement élastique ».

3.99. L'étirement élastique exige une grande quantité de points de contrôle bien répartis sur la carte. Une carte numérique d'un pays et les limites administratives ou tout autre point clairement défini également présents sur la carte numérisée peuvent parfois être utilisés pour établir des liens entre les points correspondants. Le système utilise alors les coordonnées d'entrée et de sortie pour calculer des transformations polynomiales d'ordre supérieur. En général, l'erreur apportée par l'étirement élastique est assez importante et cette opération devra donc être évitée lorsque c'est possible. Cependant, dans certains cas, lorsqu'il est évident que les cartes d'entrée ne correspondent à aucune projection bien définie, l'étirement élastique est une solution valable pour utiliser les informations géographiques disponibles. Le géoréférencement de croquis cartographiques établis à la main en offre un bon exemple dans le contexte de la cartographie de recensement. Le paragraphe A2.6 de l'annexe II présente un exemple pratique illustrant l'opération de conversion d'une carte numérique en une base de données numérique correctement référencée.

3.100. Les couches géoréférencées de la même zone ou les coordonnées collectées au moyen d'un système de positionnement universel ne seront pas compatibles avec les cartes numérisées car elles sont référencées dans un système de coordonnées réelles. C'est pourquoi l'on exprime en mètres ou en pieds les coordonnées ponctuelles et linéaires numérisées à convertir de coordonnées de périphériques de numérisation en coordonnées du monde réel (voir aussi annexe II). Comme indiqué plus haut, cette opération peut être effectuée dans la plupart des systèmes soit au début de la numérisation, soit après la saisie des données spatiales.

g) Changement de projection et de système de référence

3.101. On est amené à modifier la projection lorsque l'on doit assembler dans une base de données homogène des cartes numérisées à partir de feuilles de cartes différentes. Le changement de projection est lié au processus de transformation qui convertit les coordonnées des objets de la carte numérique sans modifier leur forme. Lorsque l'on convertit d'une projection à une autre, les formes et les déformations des objets cartographiques sont bien modifiées, mais les changements peuvent être pratiquement imperceptibles aux grandes échelles cartographiques.

3.102. Les cartes publiées à différentes échelles cartographiques utilisent parfois des projections différentes. Dans d'autres cas, l'organisme cartographique peut avoir modifié la projection normale utilisée à des fins cartographiques dans le pays, et par conséquent les feuilles de cartes anciennes peuvent utiliser une projection différente des feuilles de cartes révisées à une date plus récente. De même, l'organisme cartographique peut avoir modifié le système de références géodésiques déterminant le cadre de référence pour les travaux cartographiques du pays, de sorte que les anciennes cartes topographiques, par exemple, utilisent un système de coordonnées différent de celui des nouvelles cartes.

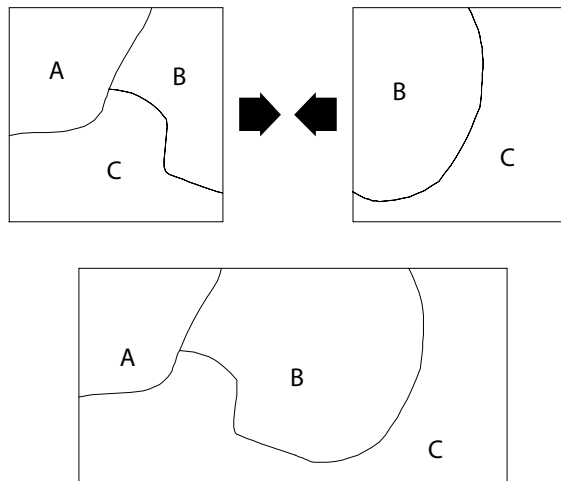
3.103. Les projections et les systèmes de référence géodésiques seront examinés de façon plus détaillée à l'annexe II. Il serait utile que le service cartographique de recensement puisse compter dans son personnel un cartographe qualifié ou puisse recourir aux services d'experts de l'agence cartographique nationale qui pourront lui conseiller la stratégie la plus appropriée en vue d'harmoniser les projections et les problèmes connexes afin d'établir une base nationale cohérente de cartes de recensement. Les opérations techniques effectives visant à modifier la projection exigeront relativement peu d'efforts, car tous les logiciels commerciaux de SIG possèdent les fonctions requises de changement de projection.

h) Intégration de segments de carte séparés

3.104. L'objectif d'un projet de cartographie numérique est de produire une base de données homogène pour une grande région ou tout un pays. Aux échelles cartographiques grandes et moyennes (par exemple 1/250 000 ou plus), les informations de la carte de base sont contenues dans des feuilles de carte topographique séparées. Ces feuilles sont numérisées séparément et les feuilles de carte numérique obtenues sont assemblées dans un SIG (voir figure III.15).

Figure III.15

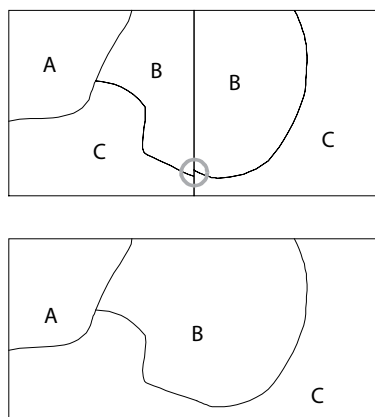
Assemblage de feuilles adjacentes de carte numérique



3.105. Ce procédé est simple en général. Il arrive néanmoins que les feuilles de carte ne correspondent pas parfaitement. Des objets s'étendant sur deux feuilles — par exemple des routes ou des limites — peuvent se trouver décalés aux limites de la carte (voir figure III.16). Des erreurs auront pu être introduites pendant la numérisation ou elles peuvent en fait se trouver déjà sur les feuilles de la carte source. Par exemple, des coupures de carte adjacentes peuvent avoir été établies à des dates différentes, de sorte que des objets nouveaux comme de nouvelles routes ne se prolongent pas à travers les limites des coupures de carte ou sont représentées par des symboles différents.

Figure III.16

Ajustement d'objets après raccordement de coupures de cartes adjacentes



3.106. Le problème est particulièrement difficile lorsqu'il n'existe pas de couverture complète du pays à l'échelle cartographique souhaitée, ce qui oblige à intégrer des coupures de cartes à différentes échelles avec des densités d'entités différentes. On rencontre souvent ce problème lorsqu'on intègre des coupures de carte à l'interface ville/campagne, les cartes urbaines à grande échelle devant être appariées à des cartes rurales à plus petite échelle. En raison des différences de généralisation cartographique, des objets peuvent ou non être présents sur les cartes à plus petite échelle ou les symboles correspondants peuvent être différents dans les deux séries de cartes. L'intégration de ces cartes exige beaucoup de jugement et d'expérience. Il est recommandé de faire preuve de prudence en cas de mélange des échelles.

3.107. Le travail de correction de ces erreurs est appelé ajustement d'objets. Il est en général effectué à la main, ce qui suppose un travail considérable de vérification. Si le décalage n'est pas trop important et si les objets sont compatibles d'une coupure de carte à l'autre, on peut les raccorder en utilisant des fonctions automatiques d'ajustement d'objets présentes dans certains logiciels de SIG.

D. Mise en œuvre d'une base de données de districts de recensement

3.108. L'élaboration d'une base de données géographique permet à l'organisme national de statistique de prendre conscience de tout l'intérêt d'une organisation autour d'un modèle géographique. Afin de concevoir des modèles de données de recensement, un organisme national de statistique doit tenir compte des besoins de ses utilisateurs de données ainsi que des dispositions légales et constitutionnelles applicables au travail de recensement. En général, il est possible de développer un modèle conceptuel afin d'établir un lien entre les données géographiques de base collectées à partir des couches de recensement, telles que les districts de recensements, les objets physiques (dont l'altitude, l'inclinaison et l'aspect) et d'autres couches ayant un impact sur le recensement, y compris les objets ponctuels et surfaciques, tels que les points de repère. Il a été indiqué plus haut que la topologie vectorielle joue un rôle déterminant, en particulier pour la définition de circonscriptions administratives et de limites.

3.109. Si ce type de topologie est important, d'autres formats de données devront toutefois éventuellement être intégrés dans un projet de SIG, et notamment des cartes scannées sous la forme de fichiers matriciels, des images satellitales, des photos aériennes, des fichiers de CAO (qui peuvent être incorporés avec une topologie et des attributs, voire sous la forme d'un « spaghetti » dénué de tout objet) ainsi que des données d'un système de positionnement universel dans des flux de points.

3.110. Les innovations informatiques à la fin des années 1990 et au début des années 2000 ont notamment vu le développement de données basées sur les objets dans le domaine informatique. Grâce à la programmation orientée objets, il est possible de définir des objets ayant littéralement des milliers de caractéristiques différentes qui, une fois importées dans un SIG, leur permettent de se « comporter » de manière prévisible dans des conditions simulées et de procéder à une modélisation sophistiquée, comme l'analyse d'un itinéraire.

3.111. Les bases de données géographiques sont bien plus que des feuilles de calcul. Des types d'entité peuvent être définis en leur attribuant des propriétés spécifiques déterminant leur comportement dans le monde réel. En tant qu'unité géographique, le district de recensement est un type d'objet servant à délimiter un territoire aux fins d'une prospection pour un recensement. Sur le plan morphologique, le district de

recensement est contigu; il s'imbrique dans des unités administratives et est composé d'unités démographiques.

1. Bases de données relationnelles

3.112. Avant d'aborder les structures particulières de la base de données du SIG pour le recensement, on examinera les notions de bases de données relationnelles. Tous les SIG opérationnels importants reposent sur des bases de données géographiques; on peut dire que celles-ci constituent la partie la plus importante du SIG. Ce type de base de données est au cœur de toutes les interrogations, analyses et prises de décisions. C'est dans un système de gestion de base de données que sont stockées les bases de données.

3.113. Un objet géographique peut être défini comme un ensemble intégré regroupant la géométrie, les propriétés et les méthodes. Les objets d'un même type général sont associés dans des classes d'objets, les objets individuels d'une classe étant appelés des « instances ». Dans de nombreux systèmes de SIG, chaque classe d'objets est enregistrée physiquement sous la forme d'un tableau de la base de données; chaque ligne est un objet et chaque propriété une colonne. Ainsi, la classe d'objets « éclairage public » peut inclure des instances telles que « lampe à gaz », « lampe à vapeur de sodium » et « lampe à vapeur de mercure ».

3.114. Le modèle de la base de données relationnelle est utilisé pour emmagasiner, rechercher et manipuler des tables de données concernant les détails géographiques enregistrés dans la base de données sur les coordonnées. Il est fondé sur le modèle entité-relation.

3.115. Dans un contexte géographique, une « entité » peut être une unité administrative ou une unité de recensement voire encore un autre détail spatial pour lequel des caractéristiques seront regroupées. Par exemple, une entité peut représenter le détail « district de recensement » (figure III.17). Les divers districts de recensement dans un district ou un pays offrent des exemples de cette entité et ils seront représentés par des lignes sur le tableau de l'entité. Le type d'entité, par contre, se réfère à la structure du tableau de la base de données : les attributs de l'entité qui sont enregistrés dans les colonnes du tableau. Pour un district de recensement, ce peut être l'identificateur spécifique, la superficie, la population, le code du secteur de supervision auquel le district de recensement est rattaché, etc. Il convient de noter que le type d'entité se réfère seu-

Figure III.17

Exemple d'un tableau d'entité — district de recensement

Entité : districts de recensements
Type (attributs)

Code de district de recensement	Sup.	Pop.	Code de superviseur
723101	32,1	763	88
723102	28,4	593	88
723103	19,1	838	88
723201	34,6	832	88
723202	25,7	632	89
723203	28,3	839	89
723204	12,4	388	89
...	...	...	...

Districts

Critère principal

lement à la définition générique du tableau de la base de données, et non aux valeurs enregistrées dans chaque cas. Un ou plusieurs attributs (colonnes) dans le type d'entité sont utilisés comme critères ou comme identificateurs. L'un deux est le critère principal, qui sert d'identificateur unique pour un type d'entité. Pour une base de données de districts de recensement, ce sera le numéro de code du district de recensement.

Variétés de bases de données relationnelles et structure de la base de données géographique

3.116. Il existe divers types de système de gestion de base de données, notamment les systèmes relationnels, objet et objet-relationnels. Un système de gestion de base de données relationnelle consiste en un ensemble de tableaux dont chacun est une matrice bidimensionnelle d'enregistrements comportant des attributs sur les objets étudiés. Bien que souples et utiles, ils ne sont pas conçus pour gérer des types de données riches telles que la géographie, pour lesquels la topologie d'un objet et les relations peuvent être compliquées. Microsoft Access et Oracle comptent parmi les systèmes de gestion de base de données relationnelle commerciaux et libres.

3.117. Les systèmes de gestion de base de données objet servent à combler une faiblesse majeure des systèmes de gestion de base de données relationnelle, à savoir leur incapacité à enregistrer des objets complets directement dans la base de données. Les systèmes de gestion de base de données objet peuvent enregistrer continuellement des objets et proposer des outils d'interrogation orientés objet. Les systèmes de gestion de base de données objet-relationnels forment une base de données objet-relationnels hybride, intégrant un moteur de base de données relationnelle associé à un cadre extensible pour traiter les objets. Idéalement, un système de gestion de base de données objet-relationnels se compose des éléments suivants : un analyseur d'interrogations pour les interrogations SQL (langage d'interrogation structurée), un optimiseur d'interrogations, un langage d'interrogation, une indexation, une gestion du stockage, des services de transaction et une réplication.

3.118. Les entreprises logicielles ont réagi au besoin de capacités spatiales dans leurs bases de données relationnelles en utilisant des extensions de système de gestion de base de données géographique. Il s'agit de systèmes de gestion de base de données importants, avec des extensions pour bases de données spatiales. Le DB 2 Spatial Extender d'IBM, l'Informix Spatial Datablade et l'Oracle Spatial en sont des exemples. Ces logiciels traitent les points, les lignes et les polygones comme des types d'objet pouvant être groupés avec des types plus riches, à l'aide de fonctions de topologie et de référencement linéaire, et une indexation avec les méthodes de l'arbre-R et de l'arbre de quadrants.

3.119. L'extension de système de gestion de base de données en vue du stockage de données géographiques pose des défis spécifiques. Les modèles de données objet sont axés sur la géométrie, en ce sens qu'ils modélisent le monde sous la forme d'un ensemble d'objets, tels que des points, des lignes, des polygones ou des matrices. Les opérations sont effectuées au niveau de la géométrie dans le cadre de procédures séparées, à l'aide de programmes ou de scripts. Mais cette méthode est trop simple pour les systèmes géographiques, en particulier lorsqu'ils se caractérisent par un grand nombre d'entités dotées d'une kyrielle de propriétés, des relations complexes et un comportement sophistiqué.

3.120. Les relations désignent l'association entre plusieurs entités. Par exemple, un tableau décrivant des districts de recensement peut être relié à un tableau correspondant à l'entité zone de supervision. Ce tableau comporte des attributs tels que le nom du superviseur, le responsable du bureau régional et des renseignements sur

la personne à contacter. Le critère principal pour ce tableau est le numéro de code du superviseur qui figure aussi sur le tableau des districts de recensement. Un système de gestion de base de données relationnelle peut ainsi relier les deux tableaux de telle sorte que chaque district sur le tableau des districts de recensement soit rattaché au district correspondant sur le tableau des secteurs de supervision.

3.121. L'opération consistant à élaborer une structure de base de données relationnelle en une série d'étapes s'appelle « *normalisation* ». Le résultat est une base de données présentant le moins possible de redondances. En d'autres termes, les données sont organisées dans un certain nombre de tableaux de façon à éviter de répéter plusieurs fois les mêmes valeurs. Cela réduit l'espace nécessaire au stockage et évite les erreurs qui pourraient être introduites au cours d'opérations courantes concernant la base de données, telles que l'insertion, la suppression ou la mise à jour de données.

3.122. La figure III.18 montre la différence entre un simple tableau de données et sa forme normalisée à partir de l'exemple d'une base de données de district. Dans le premier cas, les informations en provenance d'une province sont répétées pour chaque district de la province. Non seulement on gaspille ainsi de l'espace mémoire, mais il devient plus difficile de mettre à jour ou de modifier les informations concernant les provinces. Les valeurs devraient être remises en place pour chaque district individuellement. Dans la structure normalisée de la base de données, le nom de la province a été remplacé par un numéro de code numérique plus compact, qui établit un lien avec un second tableau. Ici, le code de la province devient le critère principal pour les informations relatives à la province, comprenant le nom de la province, sa population et le taux de fécondité total. Après avoir associé temporairement les deux bases de données au moyen du numéro de code de la province, on peut accéder aux renseignements relatifs à la province pour chaque district figurant sur le tableau des districts.

Figure III.18

Tableaux de base de données relationnelle

Provinces et districts

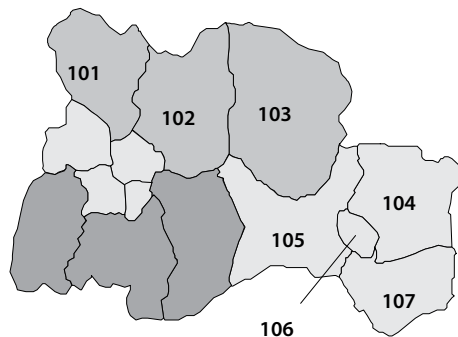


Tableau de données initiales

N°	Code Province	District	P_Pop	P_TFR	D_Pop	D_TFR
101	Mérida	Palma	214 084	3,2	89 763	3,4
102	Mérida	S. María	214 084	3,2	45 938	2,9
103	Mérida	Veralo	214 084	3,2	78 383	3,2
104	La Paz	Bolo	397 881	3,7	98 302	3,9
105	La Paz	José	397 881	3,7	67 352	4,2
106	La Paz	Malabo	397 881	3,7	102 839	3,7
107	La Paz	Chilabo	397 881	3,7	129 388	2,8
...	...	...	...	...	...	...

Tableau de données après normalisation

Districts					Provinces		
N°							
Code	Province	District	D_Pop	D_TFR	Province	P_Pop	P_TFR
101	Mérida	Palma	89 763	3,4	Mérida	214 084	3,2
102	Mérida	S. María	45 938	2,9	La Paz	397 881	3,7
103	Mérida	Veralo	78 383	3,2	...	...	...
104	La Paz	Bolo	98 302	3,9			
105	La Paz	José	67 352	4,2			
106	La Paz	Malabo	102 839	3,7			
107	La Paz	Chilabo	129 388	2,8			
...	...	...	...	...			

2. Définition du contenu de la base de données (modélisation des données)

3.124. Une fois déterminée la portée des activités géographiques de recensement, le bureau de recensement doit définir et étayer par des documents la structure des bases de données géographiques de façon plus détaillée. Ce travail est parfois qualifié de modélisation des données et consiste à définir les détails géographiques à inclure dans la base de données, leurs attributs et leurs relations avec d'autres détails. Le produit obtenu est un dictionnaire de données détaillé, qui guide le processus d'élaboration de la base de données et sert de documentation aux stades ultérieurs.

3.125. Il convient de noter que beaucoup de bases de données de SIG sont créées sans une modélisation détaillée des données. Cette phase exige en effet du temps et, dans une certaine mesure, la connaissance de concepts relatifs aux bases de données. L'investissement supplémentaire se justifie pour un projet détaillé d'élaboration de cartes de recensement. Le processus de modélisation de données nécessite un niveau de rigueur et de cohérence suffisant pour obtenir une base de données de grande qualité et facile à tenir à jour. Pour un service de cartographie de recensement qui entreprend cette activité la première fois, il peut être souhaitable de recruter un consultant ayant l'expérience des bases de données de SIG pour guider l'équipe de cartographes pendant ce travail.

3.126. Comme on l'a dit plus haut, de nombreux organismes nationaux et internationaux ont déjà entrepris d'élaborer des modèles généraux de données pour les informations spatiales dans le cadre de l'infrastructure nationale de données géographiques. Souvent un bureau de recensement pourra simplement adapter une norme nationale de données spatiales aux besoins particuliers de la collecte de données statistiques. Dans le cas où ces informations ne sont pas disponibles, un modèle de données devra être mis au point par l'organisme lui-même. Des modèles communiqués par des services cartographiques ou statistiques d'autres pays serviront de référence utile à cette fin.

3.127. À l'annexe III, un exemple indique à quoi pourrait ressembler une description de modèle de données dans un dictionnaire de données. Des normes de métadonnées, examinées dans la prochaine section, et des dictionnaires de bases de données plus simples, accompagnant les bases de données diffusées auprès du public, sont associées à ce modèle de données (voir annexe IV).

E. Problèmes de qualité des données

1. Exigences concernant l'exactitude

3.128. Une des tâches les plus importantes pour la préparation d'un projet de création de base de données numériques est peut-être la mise au point de normes définissant une exactitude suffisante des données. Dans de nombreux domaines comme la gestion d'installations et de services essentiels, l'établissement de cartes de terrain ou hydrologiques, il existe des normes d'exactitude de base de données qui peuvent être adoptées pour tout nouveau projet. Par contre, les cartes de recensement ont été jusque-là souvent établies au coup par coup en utilisant des techniques manuelles et des croquis cartographiques, sans s'inquiéter beaucoup d'exactitude géographique. Cela pouvait convenir tant que les cartes de recensement étaient utilisées aux fins du recensement seulement. Mais avec les SIG, les cartes de recensement font désormais partie intégrante de nombreuses applications analytiques dans les secteurs public,

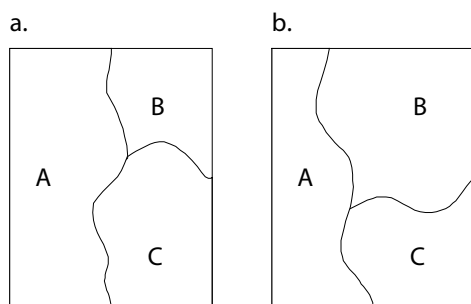
privé et universitaire. C'est un facteur important qui justifie au départ l'investissement dans la cartographie numérique de recensement. Lorsque les cartes de recensement sont combinées à d'autres sources de données géographiques numériques, les défauts d'exactitude sautent immédiatement aux yeux. Les exigences en matière d'exactitude sont donc plus grandes pour la cartographie numérique de recensement que pour les techniques de cartographie de recensement traditionnelles.

3.129. L'exactitude d'un SIG concerne à la fois les données de recensement qui peuvent lui être rattachées — et les données géographiques. Les questions relatives à l'exactitude des données sont les mêmes que celles concernant les activités d'entrée et de traitement de données liées à un recensement. Elles ne seront donc examinées que brièvement. L'exactitude des données géographiques concerne les points, les lignes et les zones enregistrées dans la base de données du SIG et décrivant des détails à la surface de la Terre.

3.130. On peut distinguer deux types d'exactitude de données géographiques : l'exactitude *logique* et l'exactitude de *localisation*. L'exactitude de localisation est quelquefois qualifiée d'exactitude absolue. L'exactitude logique se réfère à l'intégrité des relations entre détails géographiques. Par exemple, une route dans une couche de base de données de SIG doit se rattacher à un pont dans une autre couche. Un cours d'eau enregistré dans une base de données hydrologiques qui définit la limite entre deux circonscriptions administratives doit coïncider avec la limite entre ces circonscriptions. Et une ville représentée par un point dans une base de données de SIG doit se rattacher à l'unité administrative correspondante dans une autre couche du SIG. Les mêmes relations logiques peuvent être représentées correctement sur différentes cartes d'apparence très différente. Par exemple, sur la figure III.19, les deux cartes représentent correctement les relations de voisinage entre trois circonscriptions administratives.

Figure III.19

Exactitude logique



3.131. L'exactitude de localisation, par contre, garantit que les coordonnées des détails figurant dans la base de données du SIG sont correctes par rapport à leurs positions vraies sur la surface terrestre. Cela signifie que les mesures cartographiques doivent être effectuées avec un degré suffisant de précision, en utilisant des instruments de mesure exacts comme les systèmes de positionnement universel. Naturellement, un ensemble de données sans erreur de localisation représentera aussi avec exactitude les relations logiques entre les détails géographiques.

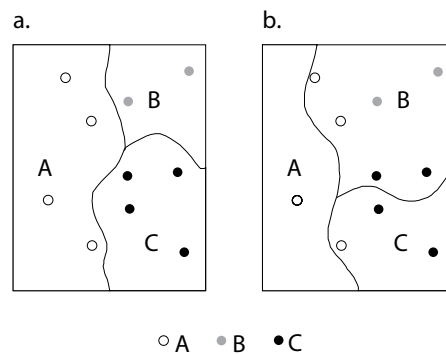
3.132. Dans certaines applications, l'exactitude logique est plus importante que l'exactitude de localisation. Pour une base de données de recensement, il peut être plus important de savoir qu'une certaine rue définit la limite d'un district de recensement que de savoir que les coordonnées exactes représentent l'emplacement de la rue dans le monde réel avec un degré d'exactitude élevé. En fait, les croquis cartographi-

ques établis au cours des activités cartographiques traditionnelles de recensement sont généralement exactes du point de vue logique, mais leur exactitude de localisation est faible. Cela ne présente pas de difficulté quand les cartes sont utilisées seulement pour étayer le dénombrement censitaire, tant que des distorsions n'empêchent pas de s'orienter dans le district de recensement. Mais si les cartes de recensement sont utilisées par la suite à d'autres fins, on pourra rencontrer de sérieuses difficultés.

3.133. La figure III.20a, par exemple, montre un ensemble de sites d'enquête par sondage qui ont été déterminés au moyen d'un système de positionnement universel de grande précision. La carte de base sous-jacente présente un degré élevé d'exactitude de localisation et par conséquent les points sont correctement placés dans la circonscription administrative correspondante. Au contraire, la carte de base sur la figure III.20b, bien qu'exacte sur le plan logique, l'est peu du point de vue du positionnement. En conséquence, certains points correctement localisés grâce au GPS ne se trouvent pas placés dans les circonscriptions administratives correctes, ce qui se traduira par des résultats erronés lorsque les réponses de l'enquête seront agrégées par unité administrative.

Figure III.20

Difficultés lorsque l'exactitude de localisation n'est pas assurée



3.134. Lors de l'élaboration des cartes numériques de recensement, on doit donc s'efforcer d'assurer un degré suffisant d'exactitude de positionnement, si les limites obtenues sont utilisées à d'autres fins que le dénombrement proprement dit. Naturellement, on ne rencontre que rarement des ensembles de données géographiques exacts à 100 %. Dans toute activité cartographique, manuelle ou numérique, il existe une corrélation négative entre l'exactitude que l'on peut obtenir et le temps et les ressources nécessaires pour atteindre ce niveau de qualité des données. Généralement, un gain incrémentiel d'exactitude dépassant 90 ou 95 % exige une dépense plus que proportionnelle de temps et d'autres ressources. De fait, d'après certaines évaluations, les mesures à prendre pour porter le degré d'exactitude de 95 à 100 % absorberaient 95 % du budget total d'un projet (Hohl, 1998).

3.135. Il est de pratique courante en cartographie topographique de définir des normes d'exactitude fondées sur la localisation de points. Les cotes d'altitude, par exemple, doivent se situer à x mètres de leur position réelle dans y % des cas. L'erreur admissible augmente au fur et à mesure que l'échelle cartographique diminue. Par exemple, sur une carte au 1/25 000, l'erreur doit être plus faible que sur une carte au 1/100 000. Étant donné que les cartes de recensement seront basées dans une grande mesure sur les cartes topographiques disponibles, les normes d'exactitude pour la cartographie de recensement devraient être établies en étroite collaboration avec des

experts des services cartographiques nationaux. Cela assurera aussi la compatibilité entre la qualité des produits du projet de cartographie de recensement et celle d'autres séries de cartes numériques nationales.

3.136. Bien qu'un niveau élevé d'exactitude de localisation soit souhaitable, l'application de normes d'exactitude trop élevées se traduira par une augmentation des coûts, des attentes exagérées des utilisateurs et le cas échéant des frustrations pour le personnel cartographique, qui risque de ne pouvoir atteindre des objectifs trop élevés. Des normes d'exactitude trop basses peuvent se traduire par des produits de qualité inférieure. Les utilisateurs risquent soit de rejeter le produit s'ils se rendent compte de ses limitations, soit de l'utiliser avec un degré de confiance exagéré qui pourra entraîner de graves erreurs dans les résultats d'analyses. L'« adaptation aux besoins » est une notion courante en matière d'élaboration de base de données de SIG. Elle tient compte du fait que les données spatiales numériques ne sont jamais parfaites. Elles peuvent être toutefois appropriées pour une tâche, mais de qualité insuffisante pour une autre.

3.137. En définissant les normes de qualité, l'organisme de recensement doit prendre en considération non seulement ses besoins internes, mais aussi les besoins des utilisateurs extérieurs de cartes numériques de recensement. Des directives concernant l'exactitude des données pourront ainsi être élaborées avec la collaboration de toutes les parties prenantes dans le cadre de l'évaluation des besoins des utilisateurs. Les normes dépendront aussi des ressources disponibles, de la qualité des documents sources — les informations pour différentes couches de données peuvent être de qualité variable — et de la technique choisie pour collecter les données sur le terrain.

2. Contrôle de la qualité

3.138. Le contrôle de la qualité consiste dans l'ensemble des opérations et des conventions garantissant que les bases de données mises en place lors de l'élaboration des cartes de recensement sont conformes aux normes d'exactitude adoptées. La version révisée des *Principes et recommandations concernant les recensements de la population et de l'habitat* (Nations Unies, 2008) souligne l'importance du contrôle de la qualité et donne un aperçu de ces questions dans le processus du recensement. Ces notions générales s'appliquent aussi à l'élaboration des bases de données géographiques

3.139. L'ensemble du recensement doit faire l'objet d'un contrôle de la qualité, lequel est tout aussi important pour les programmes géographiques. De manière générale, les meilleurs modèles de programmes intègrent des protocoles garantissant un contrôle de la qualité à chaque étape du recensement.

3.140. Les tests et les procédures de vérification des erreurs constituent l'essentiel du processus de contrôle de la qualité. Mais le contrôle de la qualité est aussi une question de comportement du personnel cartographique qui s'efforce de limiter les erreurs à chaque étape du processus de conversion des données. Le personnel de recensement devrait être encouragé à signaler les difficultés rencontrées au niveau des produits. Les problèmes à répétition peuvent attirer l'attention sur des procédures inappropriées ou un manque de formation et exiger que l'on modifie les affectations de membres du personnel ou que l'on change de matériel ou de techniques. C'est pourquoi il importe que les membres du personnel ne craignent pas de signaler les problèmes rencontrés au cours de leur travail et qu'ils comprennent clairement l'objectif général des procédures de contrôle de la qualité.

3.141. Si dans la plupart des cas la spécialisation du personnel dans différentes tâches peut améliorer la qualité générale des données, beaucoup de tâches requises par la mise en place d'une base de données de SIG sont assez répétitives. Une affectation

de travail monotone peut se traduire par une augmentation du nombre d'erreurs causées par un défaut de concentration. Une modification des affectations par roulement pourra contribuer à y remédier. Les membres du personnel se familiariseront ainsi avec divers aspects du processus général de conversion des données; ils pourront ainsi mieux comprendre leur travail, d'où une amélioration de la qualité générale des produits. On devrait aussi demander aux membres du personnel de suggérer des modifications de procédures débouchant sur une amélioration de la qualité des données. Ces suggestions devraient être évaluées dans un environnement contrôlé — en dehors des activités normales — avant de mettre en œuvre les changements. Ainsi l'obtention de la meilleure qualité possible devient un processus continu.

3.142. Les procédures de contrôle de la qualité font appel à des méthodes informatisées et manuelles. Les procédures informatisées sont préférables, car elles sont rapides et fiables. Cependant, de nombreux aspects de la conversion des données ne peuvent être évalués que par inspection et comparaison visuelles. Les techniques automatisées concernant les données relatives aux attributs géographiques sont semblables à celles utilisées pour l'entrée des données de recensement. Des contrôles d'intervalles et de numéros de code garantissent que les champs d'attributs ne contiennent que des valeurs autorisées. Des tests statistiques peuvent être effectués afin d'épingler des valeurs hors norme. Le nombre de circonscriptions administratives ou de recensement dans la base de données numérique doit être identique au nombre correspondant dans la liste maîtresse des zones géographiques. L'identificateur de zone géographique est le champ le plus important dans la base de données de recensement du SIG, car il assure la correspondance entre les cartes de base numériques et les données agrégées de recensement. Les ressources les plus importantes pour le contrôle — automatisé ou manuel — des données attributs devraient donc être utilisées pour faire en sorte que les attributs ne comportent aucune erreur.

3.143. Les possibilités de contrôle automatisé de la qualité sont assez limitées. Certains progiciels de SIG vérifieront l'exactitude de la topologie de la base de données : ils vérifieront par exemple si toutes les zones sont fermées et si toutes les lignes se rattachent les unes aux autres. Une base de données d'un village peut être combinée avec un ensemble de données de qualité connue relatif aux limites d'une circonscription administrative pour garantir que les identificateurs administratifs dans la base de données du village soient corrects (vérification de points dans un polygone). Certaines erreurs sont évidentes, lorsque par exemple les limites de deux circonscriptions administratives numérisées séparément ne correspondent pas entre elles. D'autres sont moins faciles à repérer, par exemple lorsque certaines limites intérieures ou routes manquent dans l'ensemble de données d'un SIG. Pour l'essentiel, le contrôle de la qualité de produits cartographiques doit donc être fondé sur la comparaison visuelle de documents sources (cartes, photographies aériennes, etc.) avec les données numérisées. À cette fin, les cartes numériques sont imprimées, de préférence à la même échelle que les cartes sources. Le document source et le produit sont ensuite comparés soit juxtaposés, soit superposés sur une table lumineuse. Toute erreur systématique appelle l'attention sur un problème concernant les procédures de conversion de données, qui doit être examiné immédiatement. Le travail de contrôle manuel des erreurs ne doit jamais être effectué par le collaborateur qui a produit les données.

3.144. Les mesures de contrôle de la qualité doivent être étayées en détail. Un formulaire de registre imprimé constitue en général le moyen le plus approprié pour prouver la qualité des données, mais des formulaires numériques informatisés peuvent aussi être utilisés. Le formulaire doit indiquer la procédure de contrôle de la qualité utilisée, quand et par qui elle a été mise en œuvre, qui a établi les données contrôlées et le résultat des tests. Des registres doivent être établis pour les tests manuels comme

pour les tests informatisés. Ces registres attestent non seulement l'exactitude d'un ensemble de données et son historique mais ils peuvent aussi indiquer les collaborateurs ayant besoin d'un complément de formation.

3.145. Un ensemble cohérent de procédures de contrôle de la qualité devrait permettre d'obtenir un produit final présentant un degré d'exactitude acceptable. Mais la plupart des projets ajoutent en général une étape finale d'« assurance de la qualité », consistant en une nouvelle série de contrôles et une dernière opération destinée à résoudre les problèmes. L'assurance de la qualité est examinée ci-après.

3. Partage du territoire national en unités de traitement

3.146. Une base de données numérique complète des districts de recensement est constituée de milliers d'unités. Dans les grands pays, il n'est en général pas commode d'enregistrer tous les polygones des districts de recensement dans la même couche de données physiques. Par contre, on peut diviser le territoire national en zones opérationnelles. Dans une structure administrative de recensement décentralisée, divers bureaux régionaux et divers opérateurs dans chaque bureau régional peuvent ainsi travailler sur des éléments séparés de la base de données. À condition d'avoir assuré la concordance entre les limites des sous-sections de la base de données, les éléments séparés peuvent être combinés à un stade ultérieur pour produire des cartes au niveau des districts, des provinces ou au niveau national. Ce travail exige cependant un alignement latéral qui consiste à raccorder manuellement des détails liés entre eux traversant deux blocs ou plus.

3.147. Dans les grands pays, le travail cartographique est probablement décentralisé. Dans ce cas, les zones opérationnelles sont naturellement définies par le secteur relevant de chaque bureau régional de recensement. Par exemple, un pays peut répartir le travail cartographique de recensement entre quatre bureaux régionaux, le bureau principal jouant à la fois le rôle de l'organe principal de coordination et celui d'un bureau régional. Dans chaque bureau régional, les bases de données peuvent être encore partagées en zones plus petites. Le travail sur des bases de données de moindres dimensions exige en général moins de calculs. La division en éléments plus restreints permet aussi à plusieurs opérateurs de travailler simultanément sur différentes parties de la base de données.

4. La carte de base administrative numérique

3.148. Le choix du concept opérationnel dépend de la procédure ainsi que de l'environnement organisationnel. Si l'approche décentralisée est retenue, le bureau national de recensement doit en premier lieu créer un modèle national de limite pour les principaux niveaux de la hiérarchie administrative du pays. Par exemple, le bureau de recensement devra créer, se procurer ou commander un ensemble de limites spatiales numériques de provinces, de districts et, si possible, aussi de sous-districts. Ces limites devraient être très exactes et présenter un degré de détail qui les rendent utiles pour la cartographie de districts de recensement à de grandes échelles cartographiques (par exemple à une échelle d'au moins 1/250 000). Il conviendrait de les utiliser pendant tout le processus d'élaboration des cartes de recensement, ainsi que pour la distribution d'informations agrégées de recensement à référence spatiale à ces niveaux de la hiérarchie administrative.

3.149. Ces limites ont pu déjà être définies sous forme numérique par l'organisme cartographique national. Dans ce cas, elles représenteront la carte de base

administrative numérique officiellement reconnue du pays (voir ci-dessus l'examen de la question des infrastructures nationales de données spatiales). Les numéros de code utilisés dans la base administrative devraient correspondre aux numéros de code utilisés dans la base de données du recensement.

3.150. Les limites officielles de district de chaque zone opérationnelle devront être distribuées aux bureaux chargés de délimiter les districts de recensement. Les limites de districts de recensement sont ensuite introduites dans ces polygones officiels d'unités administratives. On aura ainsi la garantie que dans toute agrégation ultérieure les limites de districts voisins concorderont parfaitement. Si des limites de district ont été numérisées séparément par chaque bureau local, il est peu probable qu'elles coïncident parfaitement. Il faudra alors un important travail de vérification. De plus, il y aura aussi beaucoup de travail en double, car les mêmes limites auront été numérisées deux fois — l'ayant été par chaque bureau ou opérateur régional voisin.

5. Prise en charge d'unités spatiales séparées

3.151. Les unités administratives sont souvent subdivisées en unités spatiales ou polygones séparés et distincts. Par exemple, un district peut être constitué d'une zone englobant une partie de terre ferme et un certain nombre d'îles. Aux fins du traitement des données de recensement, cela ne cause pas de difficultés car il n'y aura dans chaque tableau de données de recensement qu'un seul enregistrement concernant ce district. Par contre, dans la base de données des attributs géographiques, ce district fera l'objet de deux enregistrements ou plus — un pour chaque polygone. Cela donnera lieu à des difficultés lorsqu'on voudra rattacher les attributs de recensement aux polygones au moyen du tableau des attributs géographiques. Dans un système de base de données relationnelle, l'enregistrement des données censitaires est relié à chaque polygone de la base de données du SIG possédant le même identificateur de district. L'élaboration de cartes de valeurs ou de densités moyennes ne présente pas de difficultés. Le revenu ou la densité démographique sont les mêmes dans tout le district. Par contre, le décompte de données telles que la population totale ou le nombre de ménages occasionnera des difficultés si l'utilisateur veut additionner la population totale de tous les districts. Les enregistrements étant répétés pour chaque polygone appartenant au même district, il y aura des doubles comptages et le total final sera exagéré. Il existe deux méthodes pour résoudre ce problème.

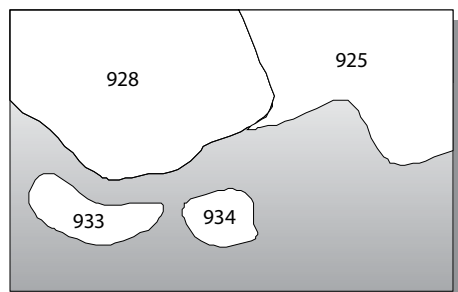
3.152. Certains logiciels perfectionnés de SIG permettent de définir des « régions ». Les régions peuvent consister en un ou plusieurs polygones, mais il n'y a qu'un enregistrement par région sur le tableau des attributs géographiques. Le logiciel garde la trace de chaque polygone considéré individuellement et de la région à laquelle il se rattache. Dans certains logiciels, les régions peuvent même se superposer, bien que ce ne soit pas là une caractéristique utile pour les applications censitaires où les districts de recensement doivent être mutuellement exclusifs.

3.153. De nombreux logiciels de SIG de bas de gamme ne comportent pas cette option. Dans ce cas, une solution simple consiste à ajouter un champ de données supplémentaire (« une valeur indicative ») au tableau des attributs géographiques (voir figure III.21). Ce champ prendra une valeur de un pour le plus grand polygone appartenant au district et de zéro pour les polygones moins étendus. Avant de totaliser ou d'établir la moyenne de valeurs d'attributs quelconques, l'utilisateur peut ne choisir d'abord que les polygones ayant une valeur égale à un dans ce champ. Un champ supplémentaire pourra être ajouté, contenant le nombre de polygones appartenant à la

même unité. Ces renseignements peuvent être obtenus rapidement, grâce à la fonction de fréquence ou de tableau à entrées multiples du progiciel de SIG.

Figure III.21

Prise en charge de circonscriptions administratives constituées de plusieurs polygones



Identificateur interne	Identificateur administratif	Nombre administratif	Contrôle
925	02015	Kalana	1
928	02012	Bagor	1
933	02012	Bagor	0
934	02012	Bagor	0

6. Surfaces utilisées pour le calcul

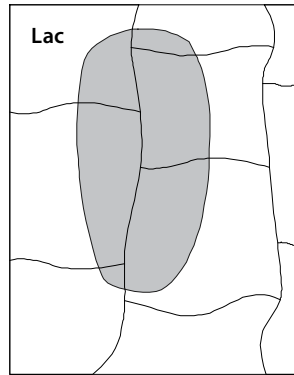
3.154. L'utilité des bases de données censitaires sera accrue si elles comprennent diverses variables géographiques courantes. La plus importante est la superficie de chaque district de recensement ou de chaque circonscription administrative. N'importe quel logiciel de SIG peut calculer la superficie d'un polygone, à condition que la base de données soit correctement référencée dans une projection équivalente de référence. Cependant, en fonction de la résolution et de l'exactitude des limites numérisées, on pourra constater dans les mesures de SIG des erreurs importantes dues aux limites très schématiques et aux îles non comptées parce qu'elles étaient trop petites pour figurer sur une carte à petite échelle. Lorsque c'est possible, il est donc préférable d'utiliser des valeurs de superficie plus exactes établies par l'organisme cartographique national.

3.155. Les valeurs de superficie sont utilisées pour obtenir des évaluations de densité, et surtout des densités démographiques. Les valeurs de superficie publiées se réfèrent en général à l'étendue comprise dans le périmètre total de la circonscription administrative défini par la loi — c'est-à-dire à la superficie totale. Parfois, cela peut donner des évaluations de densité de nature à induire en erreur. Dans un cas par exemple, une publication nationale relative au recensement a indiqué la superficie de plusieurs districts bordant un grand lac. Les superficies mentionnées contenaient la portion des districts qui s'étendait du rivage du lac à la ligne médiane du lac (voir illustration sur la figure III.22). Cette inclusion de la surface du lac doublait la superficie totale de certains districts. En conséquence, les densités démographiques réelles étaient sous-estimées par un facteur de deux. Si les statistiques officielles de la densité démographique, par exemple, sont utilisées comme critère pour affecter des ressources

ou pour déterminer la population susceptible de bénéficier de programmes publics, la définition de la densité démographique peut avoir des conséquences très sérieuses.

Figure III.22

Lac occupant une grande superficie dans plusieurs circonscriptions administratives



3.156. Dans les pays où cela présente des difficultés, le bureau de recensement peut décider de présenter deux champs consacrés à la superficie : le premier indiquant la *superficie totale* d'une circonscription administrative et le second la *superficie terrestre* c'est-à-dire la superficie totale moins la superficie couverte par des étendues d'eau et éventuellement celle d'autres zones inhabitées telles que des réserves naturelles protégées. Certains pays indiquent aussi la superficie des terres agricoles. Cela permet aux utilisateurs de calculer des densités de population agricole ou inversement le nombre d'hectares de terres agricoles disponibles par habitant dans le district. Ces valeurs de superficie peuvent être calculées très facilement dans un SIG, en utilisant les couches de données géographiques appropriées, et en tenant compte des risques d'erreur dus à la schématisation cartographique mentionnée plus haut. Dans tous les cas, il importe que les définitions des superficies nettes soient correctement étayées par des documents.

3.157. Étant donné que la plupart des logiciels de SIG considèrent chaque polygone figurant dans la base de données comme un enregistrement séparé, les chiffres de superficie calculés à l'aide du SIG pour des circonscriptions administratives ou des unités de recensement comprenant plus d'un polygone ne seront pas utilisables pour les calculs de densité. Au contraire, les superficies de tous les polygones appartenant à la même circonscription administrative ou à la même unité de recensement doivent être agrégées. Cela peut être fait dans le cadre du SIG en utilisant des fonctions appropriées de tableaux à entrées multiples.

F. Élaboration de métadonnées

3.158. Il est recommandé dans le présent *Manuel* de considérer le travail d'élaboration d'une base de données géographique comme une activité de longue durée et non comme une activité ponctuelle. Sur une longue période de temps, les éléments d'une base de données seront consultés à plusieurs reprises, parfois après de longs intervalles. Le personnel étant appelé à changer fréquemment, la mémoire institutionnelle ne peut se référer uniquement aux éléments dont se souviennent les analystes de

données géospatiales ayant participé à l'élaboration des données initiales. Il faut donc pouvoir disposer d'une documentation détaillée sur toutes les phases de l'élaboration de la base de données numériques et spatiales de recensement.

3.159. Les informations relatives à la qualité des données, aux formats, aux phases du traitement et à toutes les autres informations concernant un ensemble de données sont appelées métadonnées, c'est-à-dire « données relatives aux données ». Les métadonnées ont plusieurs rôles :

- Faciliter la sauvegarde et la mise à jour des ensembles de données numériques détenus par un organisme.
- Faciliter la diffusion des données en fournissant des informations sur l'aptitude d'un ensemble de données à être consulté par des utilisateurs extérieurs.
- Faciliter l'intégration d'ensembles de données produits à l'extérieur aux ensembles de données détenus par l'organisme considéré.

3.160. Il peut évidemment exister de grandes différences entre les métadonnées que les divers producteurs de données considèrent comme essentielles. De nombreux pays ont donc commencé à élaborer des normes générales pour les métadonnées géographiques. Ces normes visent à unifier les conventions relatives aux éléments destinés à étayer les informations spatiales. Les pays appuient donc la mise en place d'une infrastructure nationale de données spatiales en facilitant les échanges et l'intégration des données spatiales. Sur le plan international, plusieurs organisations s'efforcent de coordonner l'élaboration de normes de métadonnées spatiales entre divers groupes de pays. On peut citer le Groupe de travail de l'ISO (ISO/TC211) sur l'information géographique/la géomatique (www.statkart.no/isotc211/), le service d'échange ouvert d'informations de la Commission européenne (www2.echo.lu/oii/en/oii-home.html) et le Comité régional permanent chargé de l'infrastructure des SIG pour l'Asie et le Pacifique (www.permcom.apgis.gov.au).

3.161. Les données de recensement spatialement référencées faisant partie intégrante d'une infrastructure nationale de données spatiales, l'élaboration de cartes numériques de recensement devrait être le plus possible intégrée aux autres activités de cartographie numérique du pays. En ce qui concerne les métadonnées, cela signifie qu'une norme nationale ou régionale de métadonnées, si elle existe, devrait être adoptée par un organisme national de recensement. Une étroite collaboration avec l'autorité nationale responsable — en général le service cartographique national ou un conseil consultatif interdépartemental — facilitera la mise en place de ces normes. S'il n'existe pas de norme nationale, l'organisme de recensement gagnera du temps et économisera des ressources en adaptant une norme appropriée d'un autre pays plutôt que d'élaborer intégralement une nouvelle norme de métadonnées.

3.162. Les Content Standards for Digital Geospatial Metadata (CSDGM) créées par le Federal Geographic Data Committee aux États-Unis (www.fgdc.gov) constituent un exemple de norme de métadonnées très élaborée et largement utilisée. Elle illustre les types d'informations contenus dans une base de données consacrée aux métadonnées. La norme complète est très détaillée et divers comités spécialisés établissent des directives pour des types particuliers de données. Le Sous-Comité pour les données culturelles et démographiques, par exemple, est installé dans les locaux du United States Census Bureau (<http://www.census.gov/geo/www/standards/scdd/>) (voir également FGDC, 1997b). Seuls les principaux éléments de la définition des métadonnées seront examinés ici.

3.163. Les CSDGM comportent sept sections principales et on peut les considérer comme un modèle de base de données avec des champs décrivant les divers

aspects d'un ensemble de données spatiales. Certains champs ne comprendront qu'un ensemble prédéfini de codes ou d'attributs. Mais de nombreux éléments sont des champs de texte, dans lesquels le producteur de données décrit des caractéristiques de la base de données, tels que des informations sur la qualité ou l'historique des données. Les éléments les plus importants sont considérés comme obligatoires et ils doivent donc être sélectionnés pour chaque ensemble de données. Cet ensemble obligatoire de champs est un point de départ approprié pour définir un modèle de métadonnées pour un organisme de recensement. Les autres sont qualifiés « d'obligatoires le cas échéant » ou de « facultatifs ».

3.164. Les principaux éléments de la norme sont les suivants :

- **Informations concernant l'identification**, y compris le titre de l'ensemble des données, la zone couverte, les mots clés, l'objet, le résumé et les restrictions d'accès et d'utilisation.
- **Informations concernant la qualité des données**, telles que l'évaluation de l'exactitude horizontale et verticale, la cohérence logique, l'exactitude sémantique, l'information temporelle, la complétude et l'historique des ensembles de données. L'historique inclut les sources de données utilisées pour produire l'ensemble de données ainsi que les étapes du traitement et les produits intermédiaires.
- **Informations concernant l'organisation des données spatiales** qui concernent la façon dont les données sont stockées, par exemple les informations relatives aux points, au mode maillé et au mode vectoriel ainsi qu'à la juxtaposition des feuilles de carte numérique.
- **Informations de référence spatiale** comprenant les paramètres de projection et tous autres paramètres pertinents définissant le système de coordonnées.
- **Informations relatives aux entités et aux attributs** contenant des définitions détaillées des attributs de l'ensemble de données, y compris les types de données attributs, les valeurs admissibles et les définitions. Il s'agit pour l'essentiel des mêmes informations que celles figurant dans le dictionnaire de données décrit au paragraphe 3.127 ci-dessus.
- **Informations sur la distribution des données** y compris le distributeur, le format du fichier de données, les types de médias hors ligne, les liaisons en ligne avec les données, la facturation et la procédure de commande.
- **Informations de référence sur les métadonnées** renseignant sur les métadonnées elles-mêmes et ce qui est plus important, leur auteur et la date de leur élaboration.

3.165. Outre les sept sections principales, la norme de contenu comprend trois éléments secondaires. Ceux-ci sont fréquemment mentionnés dans les sections principales. Au lieu de répéter ces éléments plusieurs fois, il suffit de les stocker en un seul lieu. Les trois sections secondaires sont les suivantes :

- Informations relatives aux citations : elles renvoient systématiquement à l'auteur, au titre, à la date de publication et à l'éditeur.
- Informations relatives aux dates, incluant une date simple, des dates multiples ou une série de dates.
- Informations sur les personnes à contacter : personne et/ou organisation à contacter, adresse, numéro de téléphone et adresse électronique.

3.166. La normalisation des informations relatives aux métadonnées entre gouvernements et autres producteurs de données a l'avantage de permettre l'élaboration

de systèmes génériques qui gèrent et utilisent les métadonnées. Par exemple, il existe une série d'outils pour gérer les CSDGM. Il s'agit notamment de formulaires d'entrée sous forme de texte, de base de données ou de logiciel de navigation sur le Web (via Internet ou un intranet) et de lecteurs de métadonnées qui peuvent être utilisés par des bibliothèques ou des systèmes de diffusion de données via Internet. Les vendeurs de logiciels commerciaux ont aussi ajouté à leurs logiciels des outils de documentation qui facilitent l'élaboration de métadonnées selon le format des CSDGM.

3.167. La définition du modèle de métadonnées utilisé pour le projet de cartographie de recensement n'est qu'un aspect de la gestion des métadonnées. L'autre concerne la mise en œuvre des procédures de tenue à jour des métadonnées. L'organisme de recensement doit décider quand et par qui les métadonnées sont enregistrées et sous quel format elles sont stockées — formulaires papier ou fichiers numériques — et qui est chargé de vérifier la complétude, l'exactitude et la facilité d'utilisation des informations ainsi obtenues. L'élaboration des métadonnées devra accompagner chaque phase de la création d'une base de données. Elle ne doit pas être considérée simplement comme la phase ultime de l'élaboration de la documentation. Dans l'intérêt des utilisateurs futurs ou extérieurs des données, les métadonnées devraient être considérées comme aussi importantes que les bases de données spatiales elles-mêmes.

G. Résumé et conclusions

3.168. Le chapitre III a abordé l'aspect technique de chacune des étapes de l'élaboration d'une base de données du niveau des districts de recensement, y compris les principes fondamentaux des bases de données géographiques, les techniques d'introduction de données, le codage géographique (« géocodage ») et la délimitation des districts de recensement.

3.169. À l'issue de cette phase opérationnelle, l'organisme national de statistique a réalisé une couverture cohérente et exhaustive des districts de recensement en se basant sur le recensement précédent. Il est à présent prêt à actualiser les données sur le terrain. Il a pour ainsi dire fait tout ce qu'il pouvait depuis ses locaux et il doit se rendre sur le terrain.

3.170. L'ultime tâche consiste à étayer la décision d'établir une base de données du niveau des districts de recensement à des fins de prospection pour un recensement. Elle permettra de le rendre encore plus précis et de disposer de bien meilleures données en vue des analyses et de la diffusion ultérieures. Cette démarche facilitera en outre la diffusion des résultats après le recensement, en particulier à des fins humanitaires (ex. : préparation et gestion des catastrophes). C'est ce dont traite le chapitre VI.

Chapitre IV

Intégration des activités sur le terrain à l'aide de GPS et de données télédéteectées

4.1. Ce chapitre poursuit le processus d'élaboration progressive de la base de données géographique du niveau des districts de recensement, introduit au chapitre III. Conscients de la valeur de nouveaux outils et sources de données qui ont vu le jour grâce aux satellites — à savoir, les systèmes de positionnement universel et la télédétection (en ce compris la photographie aérienne) — nous traitons ici d'emblée de ces nouveaux outils et sources de données.

4.2. Le chapitre IV a pour thème principal l'utilisation du GPS et de la télédétection aux fins de la délimitation de districts de recensement. Cette démarche vise essentiellement à valider sur le terrain les limites des districts de recensement créées au sein du laboratoire de SIG de l'organisme national de statistique à partir des cartes du recensement précédent. Toutefois, lorsqu'aucune carte précise n'est disponible, il s'agit de l'approche de base pour délimiter des districts de recensement au sein du bureau principal de recensement, préalablement aux activités effectuées sur le terrain pour compléter et valider les données. Les données de télédétection permettent aux analystes géographiques d'identifier le territoire ayant le plus besoin d'une actualisation, qu'ils distinguent de zones nécessitant une légère mise à jour. Certains principes de base des GPS seront présentés au même titre que des directives spécifiques à l'utilisation de données de GPS aux fins d'un recensement (en plus d'exemples sur l'utilisation du GPS pour la délimitation de limites administratives et de districts de recensement ainsi que la localisation d'unités d'habitation et de quartiers collectifs) et à l'extraction d'autres caractéristiques. Il sera question de l'informatique mobile et des assistants numériques. La télédétection sera abordée sous l'angle de l'imagerie satellitaire et de la photographie aérienne, en s'attardant sur certains principes fondamentaux ainsi que des directives à l'attention des organismes nationaux de statistique.

4.3. À ce stade du processus, l'agence statistique aura idéalement scanné les cartes des districts de recensement du recensement précédent et les aura exploitées pour constituer une base de données géographique numérique des districts de recensement. Aussi fastidieuse et exhaustive qu'elle puisse paraître, la base de données n'est en réalité qu'une ébauche, les cartes des districts de recensement délimitées dans le bureau principal de recensement n'ayant pas encore été mises à jour sur le terrain. Le degré de collaboration entre les sièges et les bureaux de recensement locaux pour ce faire sera fonction du degré de centralisation des opérations de recensement, ainsi que de la structure de communication et de l'accessibilité du pays. Le présent *Manuel* suppose une intégration des activités sur le terrain et du siège à travers des opérations sur le terrain et le partage de données.

4.4. Globalement, la numérisation géographique sert avant tout à tirer parti des nouvelles technologies pour réaliser plus rapidement de meilleures cartes et améliorer la qualité globale des données de recensement. Une fois de plus, nous nous

concentrerons sur les conditions spécifiques du pays, en prêtant attention aux régions administratives constituant le territoire et aux districts de recensement dont l'agence statistique se sert pour le recensement. Grâce à l'imagerie satellitaire, les analystes et planificateurs du recensement peuvent identifier des zones nécessitant un travail de terrain supplémentaire, notamment pour tenir compte de la croissance récente de zones avoisinant les villes. À des fins de planification et de logistique, il est tout à fait opportun d'identifier ces domaines prioritaires à l'avance pour localiser les zones à mutation rapide depuis le dernier recensement et se concentrer sur celles-ci. C'est ce qu'évoque l'approche de « détection des changements », dont une synthèse des données de terrain, de laboratoire et de télédétection peut garantir l'efficacité.

4.5. La télédétection est définie comme l'utilisation de capteurs imageurs pour collecter des informations sur une zone ou un objet donné. Outil puissant, elle propose des « aperçus » du paysage à recenser pouvant sensiblement renforcer la précision du dénombrement; pour les organismes nationaux de statistique il s'agit, par ailleurs, d'un investissement potentiellement conséquent. La télédétection peut être bien plus qu'un complément pour la cartographie sur le terrain; en réalité, pour en tirer la quintessence, il est recommandé d'utiliser d'autres sources de données, telles que les plans cadastraux, les descriptions des limites et les districts de recensement de recensements précédents. Compte tenu de l'investissement considérable exigé par la télédétection, les organismes nationaux de statistique doivent soigneusement réfléchir avant de s'engager à faire le nécessaire pour se procurer une imagerie et de lancer des formations. Leur plan proposera des analyses de données et des produits au préalable afin que l'organisme national de statistique ne crée pas de données extrinsèques n'ayant pas, voire peu, d'utilité pour le public. Cela est particulièrement primordial lorsque la technologie apparaît comme un complément onéreux plutôt qu'une fin en soi. Ce n'est que lorsque l'organisme national de statistique a tranché sur la finalité des données qu'il sera possible de déterminer les répercussions en termes de ressources et plus particulièrement de ressources humaines.

A. Systèmes de positionnement universel (GPS)

4.6. Jadis un gadget exotique, la technologie du GPS a révolutionné la navigation au point de devenir courante. Le GPS a également transformé la cartographie censitaire de terrain ces dernières années. En consignait les latitudes et longitudes dans un format facile d'emploi, le GPS permet à n'importe quelle application de préciser la localisation sans aucun problème. Les prix des récepteurs de GPS ayant fortement baissé — un modèle fiable coûte moins de 100 dollars —, les GPS ont été intégrés dans de nombreux domaines d'application. Ils sont largement utilisés à des fins personnelles, avec comme secteur phare le marché grand public des voitures, des bateaux, du matériel de construction et des équipements d'exploitation agricole, sans compter leur intégration aux assistants numériques et ordinateurs portables. Les principaux groupes d'utilisateurs professionnels se situent dans les secteurs de la gestion des services collectifs, des télécommunications, de la géodésie et de la navigation. Mais les GPS ont aussi contribué à améliorer la recherche sur le terrain dans des domaines tels que la biologie, la foresterie et la géologie et ils sont de plus en plus utilisés dans les études épidémiologiques et démographiques. Les GPS deviennent un outil important pour les applications cartographiques de recensement.

4.7. La présente analyse concerne essentiellement le système des États-Unis appelé couramment GPS. C'est le système le plus largement utilisé. Il a suscité l'apparition d'un vaste marché auquel participent les fabricants de récepteurs et servi-

ces géodésiques. D'autres systèmes de positionnement par satellite, le système russe GLONASS, le système Galileo de l'Union européenne et le système chinois Beidou, sont également décrits brièvement et examinés plus bas.

1. Fonctionnement des systèmes de positionnement universel

4.8. Le GPS est un système de radionavigation mondial s'appuyant sur des satellites sur orbite. Les récepteurs GPS reçoivent les signaux émis par un système de 24 satellites — 21 satellites actifs et trois en réserve — et leurs stations au sol. Ce système, connu sous le nom de NAVSTAR, est géré par le Département de la Défense des États-Unis. Les satellites tournent autour de la Terre dans six plans orbitaux, à une altitude d'environ 20 000 km. À tout moment, cinq à huit satellites GPS se trouvent dans le champ de vision d'un récepteur GPS à la surface terrestre.

4.9. La position du récepteur GPS sur la surface terrestre est déterminée en mesurant la distance à partir de plusieurs satellites, en trois (X, Y et Z) dimensions. Le satellite GPS et le récepteur produisent l'un et l'autre un signal synchronisé avec précision (« code pseudo-aléatoire »). La synchronisation est rendue possible par des horloges très précises sur le satellite et dans le récepteur. Le récepteur peut mesurer le décalage entre le signal interne et le signal reçu du satellite. Ce décalage correspond au temps qu'il faut au signal pour parcourir la distance du satellite au récepteur. Étant donné que le signal se déplace à la vitesse de la lumière (299,338 km/sec), il suffit de multiplier la valeur du décalage par la vitesse de la lumière pour obtenir la distance.

4.10. Une fois connue la distance de plusieurs satellites, la position peut être déterminée par triangulation. Si nous obtenons une mesure de la distance d'un second satellite, nous pouvons circonscrire notre position aux deux points où les deux cercles se recoupent. Pour confirmer notre position exacte, nous devons déterminer la distance d'un troisième satellite. Les cercles ayant pour rayon cette distance autour des trois satellites se recoupent en un seul point, qui est notre vraie position. Naturellement, nous vivons en réalité dans un monde à trois dimensions. Avec une seule mesure de distance du satellite, nous pourrions nous trouver n'importe où à la surface d'une sphère entourant le satellite. Avec deux mesures de distance, nous pourrions nous trouver n'importe où sur un cercle formé par l'intersection des deux sphères. Enfin, la sphère entourant un troisième satellite coupe ce cercle en deux points. Ici encore, l'un des deux points seulement est conforme à la réalité. Cependant, pour améliorer l'évaluation de la position il est procédé à une quatrième mesure. Celle-ci aide également à corriger tout manque de précision de l'horloge interne du récepteur. L'horloge atomique du satellite, quant à elle, est extrêmement précise.

2. Précision du système de positionnement universel

4.11. Des récepteurs GPS bon marché peuvent fournir des informations raisonnablement exactes sur la latitude, la longitude et l'altitude du point où se trouve l'utilisateur en un point quelconque du monde et à n'importe quel moment. La plupart des vendeurs de GPS bon marché indiquent que la position enregistrée est exacte à 15-20 mètres près pour les applications civiles. Les informations relatives à l'altitude sont un peu moins fiables que celles concernant la latitude et la longitude.

4.12. L'exactitude est influencée par plusieurs facteurs. L'un d'eux est le nombre et la position des satellites. Théoriquement, ils sont répartis dans tout le ciel pour permettre un calcul géométrique optimal. La dilution de la précision du positionnement est liée à la répartition des satellites dans le ciel et elle peut être chiffrée. D'autres

facteurs affectent la qualité du signal, comme les irrégularités orbitales. Elles peuvent être vérifiées avec les éphémérides. Au rang des facteurs, citons encore l'influence de la lune ainsi que les perturbations atmosphériques qui modifient le signal lorsqu'il traverse l'atmosphère, et les erreurs dues à la dispersion des signaux en trajets multiples par des bâtiments ou d'autres objets solides. Ces erreurs représentent plus ou moins un bruit aléatoire — c'est-à-dire une fluctuation aléatoire de courte durée de la position. Jusqu'à ce que cette pratique ait été éliminée progressivement en 2000, la plus grande source d'erreurs était l'accès sélectif, introduit par le Département de la Défense des États-Unis pour réduire la précision du signal. Même si l'accès sélectif a été suspendu, il pourrait être réinstauré en temps de guerre.

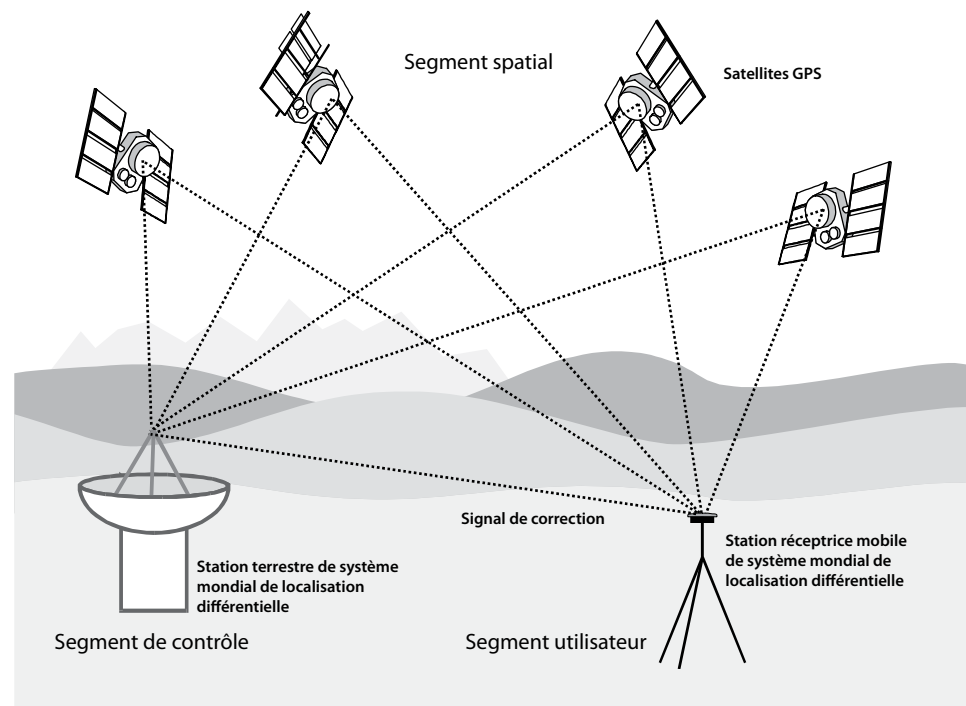
4.13. L'évaluation des coordonnées n'est pas nécessairement améliorée par des lectures répétées de coordonnées de GPS. Pour obtenir des positions plus exactes, il faudrait faire une moyenne des lectures de coordonnées sur une longue période — de plus de 24 heures. En pratique, il existe de meilleures solutions pour améliorer les coordonnées de GPS.

3. Systèmes mondiaux de localisation différentielle

4.14. Pour des applications exigeant une plus grande exactitude, les systèmes mondiaux de localisation différentielle utilisent des informations de correction transmises par une station de base dont on connaît avec précision les coordonnées pour corriger les signaux de satellite (voir figure IV.1). Ces systèmes sont de plus en plus populaires et accessibles. Ils exigent la coopération de deux récepteurs, l'un stationnaire et l'autre itinérant. Les signaux reçus par la station de base de ce système et l'unité mobile de GPS sont sujets aux mêmes erreurs. Le récepteur de la station de base mesure les erreurs de synchronisation, puis transmet les informations correctives aux unités

Figure IV.1

Systèmes mondiaux de localisation différentielle



mobiles de GPS. La station de référence reçoit les mêmes signaux de GPS, puis calcule les temps de déplacement des signaux GPS, en les comparant avec la réalité, générant un facteur de correction des erreurs. La précision qui peut être obtenue avec les systèmes mondiaux de localisation différentielle dépend du système et de la procédure de collecte de coordonnées. Une précision de 2 mètres peut être obtenue au moyen d'un matériel d'un prix tout à fait abordable et avec des temps d'observation plus courts; des résultats encore plus satisfaisants sont enregistrés en position stationnaire. Grâce à des systèmes plus onéreux et une collecte plus longue des données pour chaque lecture de coordonnées, la précision peut être de moins d'un mètre.

4.15. Il existe différentes possibilités de correction des données de GPS en temps réel. Dans de nombreux pays des organismes publics mettent actuellement en place des stations de base de systèmes mondiaux de localisation différentielle diffusant en permanence des informations de correction. Ces stations sont en général situées près de zones côtières où elles contribuent à aider la navigation maritime. Des stations de base relativement bon marché sont parfois installées par des groupes d'utilisateurs, par exemple pour des travaux agricoles très précis. De même, des unités de GPS portables de haut de gamme peuvent être converties en stations de base de systèmes mondiaux de localisation différentielle qui diffusent des informations de correction, de plus en plus par l'Internet. L'utilisateur doit trouver un emplacement connu avec précision à proximité duquel on puisse établir des données cartographiques précises. Enfin, des informations de corrections sont aussi diffusées par des satellites géostationnaires, par exemple pour la navigation aérienne.

4.16. Si des données en temps réel ne sont pas nécessaires, le traitement ultérieur des coordonnées de GPS est souvent une solution moins compliquée et utile. Dans ce cas, l'utilisateur collecte des coordonnées avec un récepteur normal de GPS. Pour chaque coordonnée, l'heure et les satellites utilisés sont enregistrés dans la mémoire du récepteur. De retour au bureau, l'utilisateur peut télécharger les informations de correction pour la période considérée, et appliquer les facteurs de correction à toutes les coordonnées collectées. Des fichiers de données de correction sont fournis par diverses sources commerciales ou publiques dans de nombreux pays. Lorsque ces informations ne peuvent être fournies par des sources secondaires, une station de base d'un système mondial de localisation différentielle peut être installée dans un emplacement central. Pour appuyer les opérations cartographiques de recensement, par exemple, une station de base pourrait être installée dans la capitale afin que les données relatives aux coordonnées collectées sur le terrain au moyen de récepteurs ordinaires bon marché puissent être corrigées après coup. Dans les grands pays, il faudra peut-être installer plusieurs stations de base.

4.17. Certains nouveaux systèmes nationaux dignes d'intérêt se servent des nouvelles techniques de satellites et de communication par l'Internet. Le système de renforcement à couverture étendue (WAAS) est un système continental de localisation différentielle mis en place aux États-Unis par la Federal Aviation Administration. Il est doté d'un satellite géostationnaire (ce terme faisant référence à la position fixe du satellite, par opposition à un satellite sur orbite) diffusant des informations de correction sur une fréquence GPS, à travers 24 stations actives aux États-Unis. Des stations de référence en continu ou CORS (« Continuously Operating Reference Stations ») ont également été conçues pour l'aviation, bien qu'ayant de nombreuses autres applications. Chaque site propose des mesures de GPS, corrigeant les erreurs et positionnements à quelques centimètres près. À la date de la publication de la présente, ce service n'est disponible qu'en Amérique du Nord, bien qu'un système similaire doive être accessible sous peu pour l'Afrique.

4.18. Dans d'autres régions, des gouvernements développent des systèmes différentiels similaires basés sur le satellite. En Asie, les Japonais mettent au point un système plurifonctionnel satellitaire d'aide renforcée aux transports (MSAS) qui opérera dans toute l'Asie avec une précision de 3m. L'Europe dispose d'un système européen de navigation par recouvrement géostationnaire (EGNOS) composé de trois satellites géostationnaires et d'un réseau de 34 stations terrestres. EGNOS a débuté en 2005 et sera certifié en 2008 comme étant d'une précision inférieure à 2 mètres. L'Afrique et l'Amérique du Sud seront couvertes.

4. Autres systèmes mondiaux de navigation par satellite

4.19. Il existe plusieurs alternatives au système de GPS américain NAVSTAR. De tels systèmes sont désignés par un terme plus générique : GNS ou système de navigation global. L'équivalent russe du GPS est GLONASS, système exploité par le Ministère de la défense de la Fédération de Russie. Entamé en 1976, sa constellation est complète depuis 1995. Depuis lors, la situation s'est détériorée avec seulement 7 satellites sur orbite en 2007. L'Agence spatiale russe envisage un redéploiement complet de GLONASS d'ici 2011, avec 24 satellites. Grâce à un accord de coopération avec le Gouvernement indien, deux satellites seront lancés depuis le territoire indien en contrepartie de signaux de grande précision.

4.20. Le système européen Galileo, développé par l'UE et l'Agence spatiale européenne, devrait être pleinement opérationnel d'ici 2013. Galileo disposera d'une constellation de 30 satellites, lancés entre 2006 et 2010, et deux stations terrestres à Munich et Rome. Il présentera deux atouts par rapport au système GPS géré par les États-Unis : l'ajout d'un « message d'intégration », informant sans délai l'utilisateur en cas d'erreurs de signaux, et une capacité à opérer sous des latitudes extrêmes. Au niveau de l'utilisateur, Galileo sera interopérable avec le système de GPS américain.

4.21. Le système Beidou proposé par la Chine comptera 35 satellites, dont cinq géostationnaires et 30 sur orbite. En Chine, les habitants bénéficieront d'un service gratuit, avec une précision de 10m; en dehors du territoire, les abonnés jouiront d'un service plus précis moyennant paiement. Deux satellites supplémentaires Beidou II ont été lancés en 2007.

4.22. Les récepteurs de GPS disponibles sur le marché diffèrent par leur prix et leurs prestations. Les spécifications techniques déterminent la précision avec laquelle les positionnements peuvent être effectués. Plus un récepteur est puissant et plus il est coûteux. L'utilisateur doit décider si le gain de précision obtenu justifie la dépense supplémentaire. Pour de nombreuses applications cartographiques, la précision des systèmes courants est tout à fait suffisante. Les récepteurs diffèrent aussi du point de vue de la convivialité, des capacités de poursuite, utiles pour la navigation — de nombreux récepteurs pouvant tracer des cartes simples — et aussi des systèmes de projection cartographique et de référence géographique susceptibles d'être pris en charge. D'autres considérations interviennent dans le choix des récepteurs de GPS : la robustesse des appareils, la consommation d'électricité (les batteries étant coûteuses, les adaptateurs pour allume-cigares d'automobile sont utiles), la capacité de stockage de coordonnées et la facilité de transfert des coordonnées vers un ordinateur portable ou de bureau.

4.23. La plupart des vendeurs offrent des produits intégrés combinant un récepteur GPS et un ordinateur format calculette ou bloc-notes pour permettre de tracer immédiatement les coordonnées sur l'écran soit isolément soit sur une carte de base numérique. Ces technologies seront traitées plus loin, avec les systèmes intégrés de cartographie de terrain.

5. Les systèmes de positionnement universel dans les applications cartographiques de recensement

4.24. Les techniques d'utilisation des GPS trouvent des applications dans toutes les formes d'activités cartographiques, y compris l'établissement et la correction de cartes d'agent recenseur pour les activités de recensement. Comme indiqué plus haut, nous devons insister sur le fait qu'il est primordial de prévoir l'utilisation des nouvelles technologies dans le cadre d'un plan global détaillé et robuste. Grâce aux systèmes mondiaux de localisation différentielle, on peut déterminer avec précision l'emplacement géographique des limites de districts de recensement avec les GPS et il est possible de localiser de façon économique des détails ponctuels tels que des équipements de service et des centres de villages. Les coordonnées peuvent être téléchargées ou entrées manuellement dans un système de cartographie numérique ou un SIG et elles peuvent être combinées avec des informations existantes géoréférencées. Les paragraphes suivants proposent des directives sur la démarcation des districts de recensement sur le terrain, suivies d'exemples d'activités de recensement spécifiques pouvant tirer parti du GPS (pour un complément d'information sur la délimitation des districts de recensement, voir chapitre III).

4.25. Comme nous le savons, les districts de recensement sont l'unité géographique opérationnelle pour la collecte de données de recensement. Ils peuvent également faire office d'unité de diffusion des données de recensement, mais ils servent principalement à collecter des données. Les districts de recensement se caractérisent par leur couverture exhaustive du territoire d'un pays. Ils permettent de représenter des zones de « population d'importance égale » et se basent sur des objets tels que des étendues d'eau et des routes pour les limites pouvant être observées au sol. En exploitant des données de télédétection en sus de cartes élaborées à partir du recensement précédent, un nombre incalculable d'heures de travail sera épargné. Par ailleurs, il y aura peu de vérifications sur le terrain et les ressources pourront être affectées aux zones à mutation rapide.

4.26. En délimitant des districts de recensement, les agents du recensement doivent comprendre que leur taille idéale est fonction de la superficie et de la population concernée. Un district de recensement représente l'étendue d'un territoire qu'un recenseur peut traiter sur la période de collecte de données de recensement. Le seuil de population pour chaque district dépend du plan développé par le service national de statistique, compte tenu des résultats d'un test de recensement préalable visant à calculer le nombre de jours requis pour le dénombrement.

4.27. Les estimations de la population sont le volet le plus critique de la délimitation de ces districts. Étant donné qu'ils dépendent à la fois de la superficie et de la démographie, il convient d'établir un système estimant la population de chaque district de recensement procurant les résultats les plus précis possibles en fonction des ressources disponibles. À défaut de bonnes estimations préalables de la démographie desdits districts, le dénombrement sera ralenti et la qualité des résultats menacée.

4.28. Les estimations démographiques des districts de recensement peuvent être obtenues grâce au concours des autorités locales. Bien que les unités du territoire de recensement elles-mêmes ne soient peut-être pas connues, les responsables locaux des zones rurales peuvent estimer la taille des hameaux et villages. Il est parfois plus facile d'évaluer les unités d'habitation que la population de petites surfaces. À défaut de la participation des autorités locales, le personnel de l'organisme national de statistique peut réaliser les estimations en profitant de visites sur le terrain ou sur la base d'informations existantes, telles que des photos aériennes, des images satellite, des cartes

routières ou de planification, des dossiers administratifs, des registres de la population, des dossiers d'entreprises de service public ou des résultats du recensement précédent. Dans ce dernier cas, il convient d'ajuster les chiffres afin de tenir compte de la croissance de la population dans la région.

4.29. Les directives relatives à la délimitation des districts de recensement sont fondées sur un objectif primordial, celui d'une couverture complète, laquelle permettra de minimiser les chevauchements et les lacunes. En général, un symbole spécial, tel qu'une ligne sinueuse, est mis en évidence dans un souci de visibilité. Les limites des districts de recensement suivent des détails visibles, tels que des routes, des ruisseaux, des lacs et des voies ferrées, tous repérables sur la carte et au sol. Certains éléments du paysage, tels que des arêtes et des forêts, sont des limites plus floues et ils ne devraient être employés qu'en l'absence de toute alternative. Des « lignes décalées » peuvent être utilisées pour préciser qu'il est nécessaire d'inclure les logements des deux côtés d'une route.

4.30. En général, la délimitation des districts de recensement devrait équilibrer la taille de la population d'un côté ainsi que la superficie et la facilité de déplacement de l'autre. Si les conditions de déplacement obligent un recenseur à passer beaucoup trop de temps entre les unités d'habitation, la taille des districts de recensement doit être réduite. Des détails naturels tels que des falaises, des rivières, des marécages et bois peuvent entraver les déplacements, au même titre que des conditions d'origine humaine, telles que des structures de l'habitat dispersées, un système routier fragmenté et de piètres infrastructures globales.

6. Certaines tâches cartographiques spécifiques liées aux GPS

4.31. Voici plusieurs tâches cartographiques spécifiques liées aux GPS :

- a) **Délimitation des districts de recensement.** Les limites des districts de recensement sont polygonales et basées sur des éléments naturels du paysage. Compte tenu de la taille restreinte des districts de recensement, il pourrait être difficile de délimiter complètement un territoire à l'aide d'un GPS sans procéder à un tri qui permettra de se concentrer sur les zones ayant récemment changé. En matière de télédétection, cette approche s'appelle la « détection des changements ». Si un pays de 20 millions d'habitants nécessite 40 000 districts de recensement comptant chacun environ 500 personnes, le simple enregistrement des limites à l'aide de GPS nécessiterait sans doute de nombreuses années en soi. Il serait préférable de numériser les limites des districts du recensement précédent et de privilégier une approche GPS au sol uniquement lorsque cela s'avère nécessaire, en particulier en cas de modification des limites (par exemple la création de nouveaux districts ou l'annexion de territoires);
- b) **Délimitation des limites administratives.** Dans la plupart des pays, les limites administratives (telles que les provinces, districts et sous-districts) ont déjà été définies à petite échelle (de façon généralisée). Il est très probable que ces unités ne soient pas suffisamment précises pour les recensements. Les organismes nationaux de statistique doivent considérer le bénéfice potentiel de limites administratives détaillées au moment du recensement par rapport au temps et au travail requis. Lorsque cela est possible, les organismes nationaux de statistique doivent pouvoir consulter les fichiers numériques existants sur les limites administratives. L'agence de cartographie nationale devrait pouvoir les fournir. Les métadonnées complémentaires devraient être examinées, dont la date et les informations de projection, avant d'essayer d'utiliser ces données dans un projet de SIG;

- c) **Localisation des unités d'habitation.** Certains pays sont allés jusqu'à enregistrer une latitude/longitude pour chaque unité d'habitation du pays, photographiant même parfois chaque logement. Mener à bien une telle tâche à l'échelle du pays requiert des ressources considérables. Si le recenseur s'en charge dans le cadre du dénombrement, cela ne prendra peut-être pas trop de temps, mais le stockage et l'indexation des fichiers doivent être exhaustifs afin d'éviter toute duplication coûteuse des tâches;
- d) **Localisation des lieux de vie collectifs.** Les lieux de vie collectifs sont toutes sortes de logements communaux ou institutionnels, dont des hôtels, des casernes, des orphelinats, des camps de travailleurs, des monastères, des couvents, des maisons de repos, des hôpitaux, des dortoirs et des établissements pénitentiaires. La population des collectivités peut parfois être extrêmement vulnérable aux catastrophes naturelles; les planificateurs humanitaires demandent d'ailleurs de plus en plus l'emplacement géographique des lieux de vie collectifs afin de planifier efficacement les interventions en cas de catastrophe. De plus, il devrait être moins ardu de localiser ces lieux de vie avec des systèmes GPS — étant donné qu'ils sont moins nombreux — que de mesurer toutes les unités d'habitation du pays;
- e) **Autres aspects importants (dont les routes).** Des éléments tels que les routes peuvent être utiles pour délimiter des districts de recensement ou fournir des données de navigation. Les points d'eau sont également utiles pour orienter les recenseurs. Par ailleurs, les paysages peuvent servir de points de contrôle pour le géoréférencement d'images satellitales ou de cartes auxiliaires. L'organisme national de statistique doit s'enquérir auprès des autorités publiques de l'existence de versions numériques de données de base, telles que des routes, afin d'économiser du temps et de l'argent en vue du dénombrement.

4.32. Pour les applications de recensement, le recours au GPS à grande échelle doit faire l'objet d'une analyse minutieuse. Pour de nombreuses tâches, dont la consignation des relevés de chaque unité d'habitation et la délimitation de districts de recensement, le matériel nécessaire pour équiper un grand nombre d'enquêteurs sur le terrain dépasserait sans doute les ressources prévues pour un projet de recensement.

4.33. La façon exacte dont les coordonnées obtenues par GPS sont utilisées dans la cartographie de recensement varie en fonction de la stratégie retenue pour cette application. Les GPS peuvent être utilisés en mode point pour enregistrer une coordonnée, par exemple, pour chaque bâtiment d'un village ou chaque intersection dans le réseau de la voirie municipale. Les cartes ou croquis cartographiques disponibles tracés pendant la collecte des données aideront l'agent, de retour au bureau, à interpréter les informations relatives aux coordonnées. Une seconde possibilité consiste à collecter des coordonnées de GPS en mode contrôlé, mode dans lequel le système enregistre les coordonnées à intervalles réguliers. Ainsi, des objets linéaires peuvent être enregistrés automatiquement en marchant le long d'une route ou en se déplaçant dans un véhicule ou à bicyclette. On dispose ainsi d'un moyen économique pour créer la base de données d'un réseau de rues ou de routes, mais ce sont les normes adoptées en matière de qualité des données qui déterminent si la précision des lignes obtenues est suffisante. Par souci de sécurité et pour permettre une sauvegarde dans les zones ne bénéficiant pas d'un approvisionnement régulier en électricité, une solution moins coûteuse consistera à stocker les coordonnées dans le système et à les enregistrer à la main sur des feuilles de données utilisées en guise de copies de sauvegarde. Les unités de GPS ne peuvent fonctionner sans alimentation par batterie; il convient donc de veiller à une alimentation

adéquate, y compris d'effectuer plusieurs sauvegardes. D'autres fonctionnalités des GPS doivent être verrouillées pour empêcher leur reprogrammation ou toute modification de leurs paramètres par le personnel de terrain. Enfin les unités doivent être clairement identifiées afin de ne pas faciliter leur revente en cas de vol.

7. Besoins en formation à l'utilisation de GPS

4.34. Pour garantir la réussite d'un projet de GPS, l'organisme national de statistique doit soigneusement gérer les achats d'équipements, répondre aux besoins en formations et en personnel et développer des protocoles applicables aux points de collecte (Montana et Spencer, 2004). Un coordinateur de GPS est recommandé pour de telles activités. Il devrait au minimum comprendre le fonctionnement des unités de GPS ainsi que les informations pertinentes relatives aux points de collecte et de stockage. Il lui appartiendra en outre de superviser les enquêteurs afin de s'assurer de la cohérence de leur travail. Un programme de formation pour les agents recenseurs peut porter sur la compréhension du fonctionnement des unités de GPS, ainsi que sur le mode de calcul d'une position par un récepteur et le dépannage des problèmes rencontrés par les unités.

4.35. Idéalement, l'utilisation du GPS doit être prévue des années à l'avance dans le cadre de la planification globale d'un recensement. Les projets portant sur les systèmes GPS doivent au moins être planifiés six mois avant la mise en œuvre, les stratégies de collecte de données étant déjà mises au point et publiées. À ce stade du processus de planification, les besoins en matériel doivent être identifiés. Les tableaux de données peuvent être conçus en respectant des conventions cohérentes pour les noms de diverses unités et géographies, de telle manière que dans la base de données, les districts de recensement ou identifiants des objets correspondent aux codes des points de cheminement GPS.

4.36. Les coordinateurs doivent instaurer un système de sauvegarde pour parer à toute perte catastrophique éventuelle de matériel GPS ou de sauvegardes. Pour ce faire, ils pourront tout simplement mettre par écrit les latitudes et longitudes sur les formulaires d'enquête papier et les copier par la suite dans un tableur.

4.37. Pour charger des points de cheminement du GPS sur un ordinateur portable ou de bureau, un logiciel du domaine public tel que GPS Utility, EasyGPS et le logiciel gratuit GPSTabel, voire des logiciels commerciaux, peuvent être utilisés. Les opérateurs ont la possibilité de reformater les fichiers chargés dans Excel ou un autre tableau destinés à un projet de SIG; à ce stade, des attributs supplémentaires peuvent être prévus. Dans certains programmes, un GPS peut être ajouté à un projet de SIG sous la forme d'un tableau dûment formaté ou en tant que « données événement ».

8. Résumé : avantages et inconvénients des systèmes de positionnement universel

4.38. Les avantages des GPS sont notamment les suivants :

- Collecte de données sur le terrain facile et peu coûteuse. Les récepteurs modernes ne nécessitent pas d'apprentissage particulier pour pouvoir être utilisés.
- La précision obtenue est suffisante pour de nombreuses applications cartographiques de recensement — une grande précision pouvant être obtenue moyennant une correction différentielle.
- Les données collectées peuvent être lues directement dans les bases de données des SIG, ce qui rend inutiles les étapes intermédiaires d'entrée ou de conversion de données.

- Les GPS sont disponibles partout dans le monde.
- De nouveaux systèmes GPS feront leur apparition en ligne au cours des cinq prochaines années.

4.39. Les inconvénients sont les suivants :

- En tant que composant physique, le GPS est bon marché, mais les opérations sur le terrain recourant aux GPS peuvent prendre du temps (et donc s'avérer coûteuses) si elles ne sont pas correctement planifiées.
- Une planification exhaustive doit déterminer les produits qui résulteront d'un usage intensif du GPS.
- La transmission du signal peut être entravée dans des zones urbaines à forte densité ou des zones forestières (erreur de trajets multiples).
- La précision des récepteurs GPS courants peut être insuffisante dans des zones urbaines ou pour capturer des objets linéaires, exigeant dès lors des techniques différentielles.
- Les systèmes mondiaux de localisation différentielle sont plus coûteux et peuvent ne pas être disponibles sur des sites distants. Ils exigent plus de temps pour collecter les données sur le terrain et un traitement ultérieur plus complexe pour obtenir de informations plus précises.
- Un très grand nombre de GPS sera éventuellement nécessaire pendant une brève période de collecte de données, rendant l'utilisation de GPS à grande échelle potentiellement hors de prix.
- Plus l'unité de GPS procède à des mesures complexes, plus des formations seront nécessaires.

4.40. Des problèmes peuvent se poser lors de l'utilisation du système GPS. Dans un cadre urbain dense, une éventuelle erreur de trajets multiples peut rendre impossible la définition de districts de recensement adjacents. La présence de tours d'habitations ou de bureaux ou une forte densité d'arbres le long des rues peuvent rendre difficile la réception de signaux provenant d'un nombre suffisant de satellites, le signal de satellite ne pouvant traverser des objets solides. Un collecteur de données qualifié peut cependant obtenir des coordonnées en gagnant à pied un espace plus découvert et en apportant une correction à la coordonnée enregistrée. Dans certains cas, on doit recourir à un système mondial de localisation différentielle ou recouper les relevés de GPS avec des sources de données supplémentaires, telles que des cartes publiées, des photographies aériennes ou même des croquis cartographiques établis au cours du travail sur le terrain. Certains pays ont mis en place un système de stations de base GPS qui permettent d'établir des cartes très précises au moyen du système mondial de localisation différentielle. Mais dans certains pays en développement ce type de réseau n'existe pas encore.

B. Systèmes intégrés de levé de terrain à l'aide d'ordinateurs portables

4.41. Dans certains pays, une nouvelle technologie combinant les fonctionnalités d'un ordinateur personnel à un GPS a été intégrée aux opérations de recensement. La « saisie directe » — à savoir la possibilité d'enregistrer directement les informations sans transcription et ainsi à sauter plusieurs étapes intermédiaires — est l'un des avantages résultant de l'utilisation d'ordinateurs portables ou d'assistants numériques (PDA). L'actualisation des cartes est immédiate. Les coordonnées sont saisies et immédiatement affichées sur l'écran de l'ordinateur portable. Si une carte de base numéri-

que est disponible, les coordonnées peuvent être affichées par-dessus. Le personnel sur le terrain peut ajouter tout attribut demandé et enregistrer ces données dans une base de données géographique. Ces informations peuvent ensuite être insérées dans la base de données géographique au bureau principal. Compte tenu de la baisse continue des ordinateurs bloc-notes et d'autres dispositifs de calcul portables, les systèmes intégrés de levé de terrain deviennent une solution intéressante pour la collecte de données sur le terrain à des fins de recensement.

Encadré IV.1

Étude de cas : expérience de Fidji en matière de GPS

Le recensement de Fidji en 2007 était le premier de la région du Pacifique à recourir aux techniques du GPS afin de lier les questionnaires de recensement à des emplacements géoréférencés pour tous les ménages du pays. Comme dans de nombreux pays en développement, Fidji ne disposait d'aucune base de données reprenant l'emplacement précis des habitations et, en l'absence de noms de rues et de numéros, il était impossible de fournir des listes d'adresses. Le GPS a procuré un avantage spécifique : il a augmenté la précision et la couverture des emplacements des ménages et a servi d'outil de gestion et de suivi (aux fins du contrôle et de la validation des données), simplifiant ainsi les activités des agents et cadres de recensement. La localisation par GPS a permis de regrouper les données de recensement dans une série d'unités administratives, dont celles liées aux applications relatives à la santé, à l'éducation et à l'environnement.

Environ 200 GPS Garmin eTrex et 20 ordinateurs portables ont été achetés pour ces activités et 10 agents ont été formés à la manipulation du GPS et d'ESRI Arcview; ils ont en outre été chargés de procéder aux contrôles de validité et de télécharger des points de cheminement. Qui plus est, 200 opérateurs de GPS ont suivi une formation sur une période de trois semaines. Les opérateurs et superviseurs de systèmes GPS ont reçu un aide-mémoire, un récapitulatif étape par étape de la configuration du GPS, de l'enregistrement des points de cheminement, du téléchargement de ces derniers d'un GPS vers l'ordinateur, de l'affichage des points de cheminement sur l'image et de l'exportation ainsi que de l'impression de fichiers image JPEG.

Des activités sur le terrain ont été organisées afin de permettre aux opérateurs de GPS de suivre les agents de recensement et de regrouper les points de cheminement. Des superviseurs leur rendaient régulièrement visite pour vérifier et télécharger ces données. Si des images étaient disponibles, les points de cheminement étaient superposés dans Google Earth afin de valider les emplacements. Un système de « trois fois le numéro de point de cheminement » a été mis au point pour établir une corrélation entre les emplacements des foyers et les questionnaires. L'agent recenseur plaçait d'abord un autocollant comportant un code à six chiffres sur la maison où l'entretien était effectué ainsi qu'une autre vignette sur le portillon dans le cas de maisons clôturées. Un petit autocollant avec le même chiffre était apposé sur le formulaire attribué à ce ménage. Une vignette de réserve était posée sur le portillon des foyers dont la porte frontale et le portillon étaient éloignés l'un de l'autre, facilitant ainsi le relevé des points de cheminement. Ensuite, l'opérateur de GPS rendait visite aux ménages, en saisissant le même numéro que celui correspondant à l'identification des points de cheminement afin de le lier par la suite au questionnaire.

Une fois le travail de terrain achevé, deux bases de données étaient créées pour les points de cheminement, l'une renseignant leurs latitudes et longitudes et l'autre reprenant les questionnaires. Les deux bases de données pouvaient alors être reliées à l'aide du numéro unique des points de cheminement/questionnaires. À ce stade, des contrôles étaient effectués pour relever les questionnaires manquants et les points de cheminement incorrects. Une équipe se rendait alors sur le terrain pour rectifier ces problèmes.

Plusieurs difficultés ont été rencontrées et des enseignements peuvent éventuellement être tirés pour les futurs recensements :

- Aux premiers stades de la collecte de points GPS, les opérateurs relevaient des points de cheminement avant que des niveaux de précision acceptables ne soient disponibles. Pour résoudre ce problème, le délai d'attente a été augmenté.

- Perte de vignettes en raison de deux festivals religieux (les maisons étant repeintes pour l'occasion) — les opérateurs ont donc dû revenir au bureau pour prendre les noms et adresses des résidents avant de retourner sur le terrain pour collecter les points de cheminement.
- Saisie parfois incorrecte des codes à six chiffres dans le GPS en tant que point de cheminement et dans la base de données à partir des questionnaires. Pour pallier ce problème sur le terrain, des codes-barres attachés aux récepteurs de GPS pourraient être utilisés à l'avenir.
- Autre source d'erreur : le relevé des points de cheminement du GPS était à la traîne par rapport au dénombrement, le retard atteignant parfois plusieurs mois. Ces erreurs n'auraient plus lieu d'être si l'équipe en charge du recensement relevait également les points de cheminement.

Pour Fiji, ce type de données GPS devrait notamment lui permettre de se préparer aux catastrophes naturelles et de les gérer lorsqu'elles surviennent. La superposition des données de base relatives aux foyers sur un modèle de terrain numérique fait très clairement ressortir les populations affectées par une certaine catastrophe.

Source : présentation de Fidji lors d'un atelier organisé à Nouméa 2008.

4.42. Les progrès technologiques, dont le GPS, les communications sans fil et la miniaturisation des ordinateurs, ont permis l'émergence de nombreuses nouvelles applications pour les SIG portables, en particulier le développement de logiciels spécialisés pour le recensement sur le terrain. Une multitude de dispositifs de poche prennent en charge de nombreuses applications d'affichage, de requête et d'analyse simples, les programmes et données étant stockés dans la mémoire puisque les dispositifs portables sont dépourvus de disques durs. Au rang des autres fonctionnalités, citons la communication via Bluetooth et/ou une connexion sans fil Wifi, ainsi que la synchronisation avec un PC pour des chargements de données et des mises à jour rapides, tout en assurant des sauvegardes de manière à prévenir la perte de données. Des logiciels de SIG ont été mis au point pour des « téléphones intelligents ». Des assistants numériques renforcés peuvent être employés pour collecter des données mobiles dans des conditions extrêmes. Les logiciels destinés aux ordinateurs de poche sont dotés de versions allégées d'applications populaires de bureautique. Parmi les logiciels de SIG utilisés pour les ordinateurs de poche, citons Autodesk OnSite, ESRI ArcPad et Intergraph Intellipoint.

4.43. Les organismes nationaux de statistique qui envisagent de recourir à des ordinateurs de poche soit pour la cartographie préalable au recensement soit pour le dénombrement à proprement parler doivent en étudier le coût. Les ordinateurs de poche équipés d'un GPS peuvent vite atteindre 750 dollars — voire plus pour les ordinateurs renforcés. Pour les gestionnaires d'activités géographiques, la question clé est la suivante : quelle est la valeur ajoutée d'une unité de 750 dollars (ou plus) par rapport à un récepteur GPS de 100 dollars ? Certains critères sont la lisibilité des cartes à l'écran, les exigences en alimentation, en particulier dans les zones sans énergie sûre, ainsi que d'autres conditions environnementales nuisibles pour les ordinateurs, même ceux qui ont été renforcés.

C. Télédétection satellitale

1. Utilisation de l'imagerie pour vérifier sur le terrain les cartes de districts de recensement produites au siège du recensement

4.44. Depuis la publication du *Manuel* de 2000, les données de télédétection ont gagné en volume, en popularité et en facilité d'utilisation. La télédétection a plus particulièrement révolutionné la cartographie depuis l'avènement de l'imagerie sa-

tellitale avec une résolution spatiale élevée (1 mètre ou mieux). L'heure est venue de mettre cette ressource précieuse au service des travaux de recensement. L'un des défis pour les organismes nationaux de statistique tient à la quantité de territoire à cartographier. Utilisées de manière pragmatique, les images satellitales peuvent économiser d'innombrables heures-personnes en permettant à l'organisme national de statistique de se concentrer sur les zones critiques. Les données de télédétection peuvent être utilisées pour un contrôle indépendant du processus de vérification sur le terrain. Le présent *Manuel* préconise une approche consistant à effectuer un tri au niveau de la superficie du pays, en la divisant en fonction des zones ayant besoin de plus ou moins d'attention. Cette approche, la « détection des changements », s'avère particulièrement utile pour établir les périmètres de zones peuplées.

4.45. Conformément au schéma présenté au chapitre III, à ce stade, le bureau de statistique national a numérisé les anciennes cartes de districts de recensement, a superposé d'autres données géographiques puis a saisi le tout dans une base de données géographique. Toutefois, il n'a pas encore corrigé les résultats sur le terrain. En superposant les limites des districts de recensement provisoires sur les images de télédétection (voir figure IV.2), les zones d'habitat peuvent être rapidement localisées et les sections prioritaires identifiées. Ces activités doivent néanmoins être planifiées de manière détaillée et réaliste.

Figure IV.2

**Délimitation des limites de districts de recensement
sur une image satellitale panchromatique**



2. Principes de la télédétection satellitale

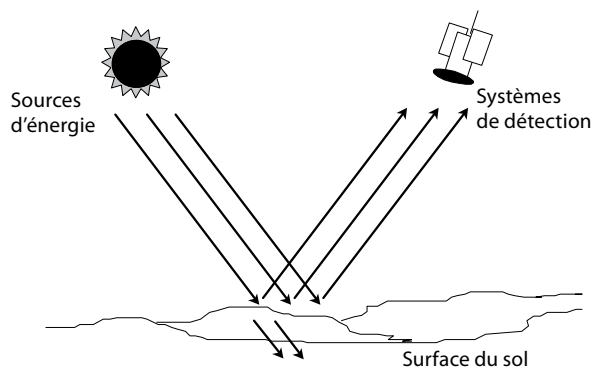
4.46. La science de l'information géographique représente l'imagerie satellitale comme l'une des nombreuses formes de données géographiques utilisables à des fins d'analyse et de présentation des résultats. L'imagerie est formatée comme un champ continu, également appelé matrice, dont les rangées de données correspondent à des pixels représentant des valeurs. La télédétection est considérée comme une saisie de données primaires; il s'agit en réalité de la saisie de données de trame primaires la

plus courante. Depuis l'apparition des satellites sur orbite, leur énorme potentiel a été concrétisé grâce au développement d'applications permettant d'effectuer des relevés de la surface de la Terre, bon nombre d'entre elles, actuelles et à venir, permettant des recensements. L'un des principaux atouts de la télédétection pour les recensements, tient à la possibilité de couvrir des zones dangereuses ou inaccessibles et donc d'économiser un nombre incalculable d'heures de travail sur le terrain pour autant qu'elle soit réalisée comme il se doit.

4.47. Les images satellitaires sont collectées par des systèmes basés dans l'espace, dont la plupart utilisent des capteurs optiques dits passifs pour mesurer le rayonnement réfléti par les objets à la surface du sol dans le spectre électromagnétique visible et invisible (figures IV.3 et IV.4). La collecte de données satellitaires est pour l'essentiel considérée comme étant passive en ce sens qu'elle reçoit une énergie émise par la Terre, par opposition aux capteurs dits actifs, tels que les radars, qui soit dit en passant peuvent également pénétrer la couverture nuageuse. Les systèmes à satellites n'utilisent pas de film photographique pour enregistrer l'énergie réfléti. Au lieu d'un film, un ensemble de détecteurs électro-optiques — semblable à une caméra CCD — mesure l'intensité du rayonnement électromagnétique et l'enregistre numériquement sous la forme d'une trame ou d'une image ordinaire constituée de lignes et de colonnes.

Figure IV.3

Processus de télédétection



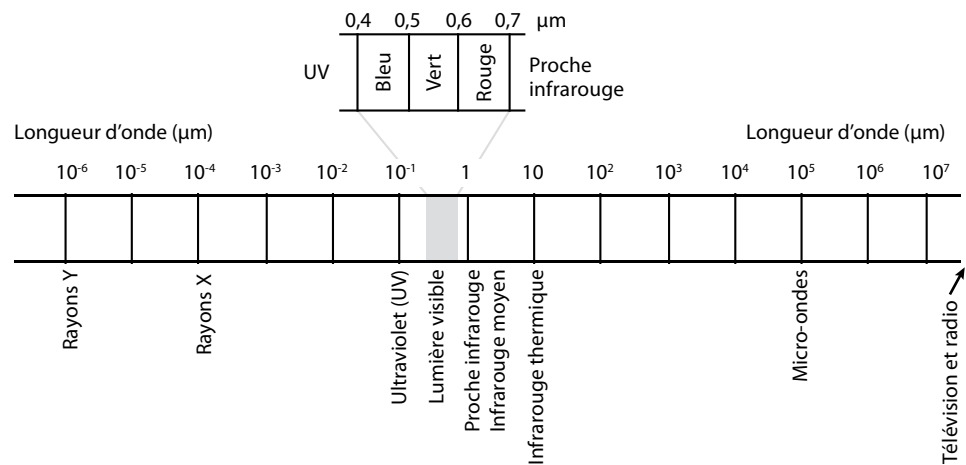
4.48. Les détecteurs de satellites opèrent en mode multispectral ou panchromatique. Le mode multispectral signifie que le satellite collecte plusieurs images (ou bandes) dont chacune mesure l'énergie réfléti dans une partie différente du spectre électromagnétique, en général dans le visible et le proche infrarouge. Pouvoir diviser une image en différentes bandes spectrales et combiner des bandes spécifiques pour analyser l'image facilite la classification d'objets au sol en fonction de leurs caractéristiques de réflectance. Par exemple, les rizières pourront envoyer un signal intense dans une bande particulière, tandis que les zones bâties apparaîtront très clairement dans une autre. Les détecteurs *panchromatiques* de satellite capturent l'énergie réfléti dans un large intervalle du spectre. Les images obtenues sont semblables à des photographies en noir et blanc. Elles présentent aussi une résolution plus élevée que les images multispectrales et constituent donc la base à retenir pour des applications cartographiques.

4.49. Les données numériques fournies par le système de capteurs sont constituées d'un ensemble de nombres qui indiquent le niveau d'énergie réfléti à l'emplacement correspondant à la surface terrestre. Le satellite envoie ces données à une station

réceptrice d'un système de stations terriennes, où elles subissent une correction géométrique et sont géoréférencées. Les images numériques ou imprimées ainsi obtenues peuvent être interprétées visuellement, de la même façon que les photographies aériennes examinées plus haut. Les images satellitales numériques peuvent être affichées dans un SIG et les objets sur l'image peuvent être délimités par un opérateur qualifié. Pour de nombreuses applications telles que les enquêtes sur l'utilisation des sols ou la gestion des ressources naturelles, cependant, les images multispectrales sont classées au moyen de techniques statistiques. Celles-ci permettent de définir des catégories de couverture des sols sur la base d'une relation calibrée entre des sites de référence d'une catégorie connue et leur signature spectrale.

Figure IV.4

Le spectre électromagnétique



3. Résolution de données de télédétection

4.50. La résolution spatiale d'une image satellitale est mesurée par la dimension d'un pixel au sol. Il existe également d'autres mesures de la résolution, dont les mesures temporelles, radiométriques et spectrales. Les dimensions d'un pixel pour un satellite commercial sont diverses et commencent à moins d'un mètre pour les systèmes haute résolution les plus populaires, tels que Quickbird et Ikonos. Le satellite de télédétection de l'Inde, le capteur panchromatique de SPOT et les images multispectrales de Landsat sont également considérés comme des systèmes haute résolution qui permettent d'établir des cartes à des échelles de 1/25 000 à 1/50 000 ou moins. À compter de mars 2008, l'entreprise de télédétection GeoEye envisage le lancement d'un capteur ayant des pixels de 0,41 mètre de dimension.

4.51. La figure IV.5 compare des dimensions de pixels simulées par agrégation à partir d'une photographie aérienne numérique d'une résolution de 0,5 mètre. L'image couvre une superficie au sol de 100 x 150 mètres. On peut distinguer des maisons individuelles ou même des automobiles à une résolution de 2 mètres, mais cela n'est plus possible pour des dimensions de pixels plus grandes. On peut extraire plus d'informations des données de télédétection en utilisant des méthodes perfectionnées de traitement d'image, y compris des algorithmes de détection de contours et des algorithmes de filtrage spéciaux. Ces techniques ont été utilisées avec succès pour établir les cartes et détecter les changements concernant des zones construites dans certaines grandes villes en croissance rapide du monde en développement.

4.52. Parmi les variétés d'images disponibles, d'un METEOSAT et d'un radar à faible résolution spatiale/haute résolution temporelle à une imagerie AVHRR de résolution spatiale et temporelle modérée, les images satellites sont utilisées à bien des fins. Seul un éventail relativement étroit de produits de télédétection est également employé pour les travaux de recensement, avec des données spatiales de haute résolution, tels que Quickbird (0,82 mètre), Ikonos (1 mètre), IRS (5,8 mètres pan), Orbimage 3 et 4 (1 mètre), et SPOT 5 (2,5 mètres). Ikonos a été lancé en 1999 et Quickbird en 2001.

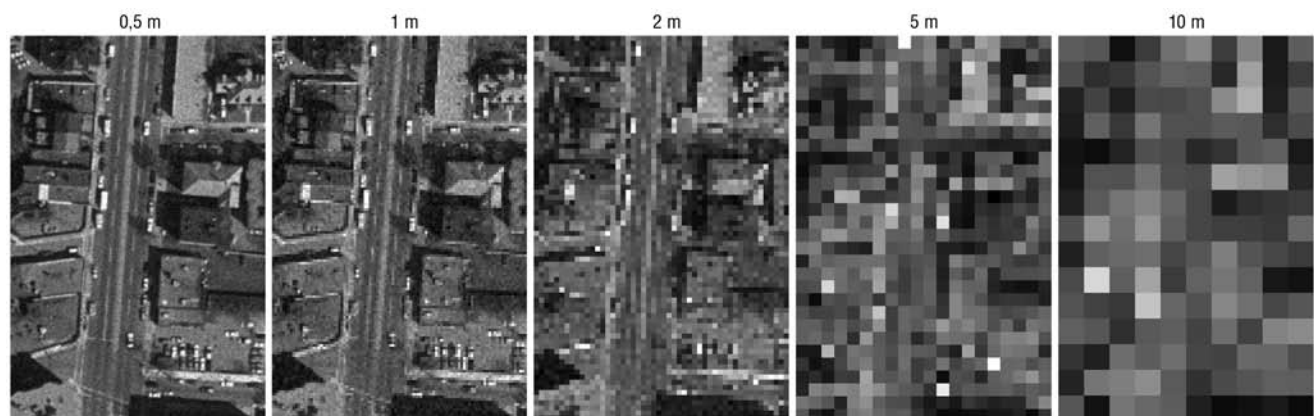
4.53. Pour la plupart des applications de recensement, une résolution spatiale de 5 mètres ou mieux s'impose pour identifier les unités d'habitation et la répartition des zones d'habitat, une imagerie multispectrale s'avérant moins indispensable. Ces sources de données ont pour inconvénient d'avoir des images haute résolution ayant une empreinte relativement réduite, nécessitant dès lors de se procurer de nombreuses scènes pour couvrir ne fut-ce qu'une fraction d'un territoire national, une couverture exhaustive étant donc une perspective hors de prix.

4.54. Avec une résolution spatiale de 30 ou 15 mètres pour l'appareil de cartographie thématique amélioré, Landsat est capable d'identifier des objets linéaires, tels que des routes et des rivières, ainsi que d'autres couches accessoires, tels que des lacs et d'autres étendues d'eau. Il présente un autre avantage majeur : il est disponible gratuitement. Dans certains cas, Landsat et ASTER peuvent être utiles pour documenter l'évolution de la couverture terrestre et l'anthropisation; ces sources de données sont d'une moins grande utilité pour les recensements. S'agissant des produits de télédétection à utiliser pour le dénombrement, le cas échéant, un organisme national de statistique doit soigneusement étudier les objectifs et les coûts. Il conviendra, notamment, surtout pour les grands pays, de considérer l'emploi d'une mosaïque d'images de 1 ou 5 mètres pour des zones très peuplées telles que les villes, avec une couverture Landsat pour les zones rurales avoisinantes. Dans ce cas, les zones rurales devront tout de même encore être cartographiées au sol pour une planification idéale de la logistique pendant le dénombrement.

4.55. La plupart des opérateurs commerciaux proposent plusieurs possibilités pour l'acquisition d'images satellitaires. En général, le prix est déterminé selon que les images sont « archivées » ou collectées pour la première fois. L'option la plus coûteuse concerne des demandes spéciales visant à acquérir d'urgence des images concernant une zone particulière. Grâce à leur résolution plus élevée, ces satellites couvrent une zone plus restreinte sur le sol et ils peuvent par conséquent offrir des images de ré-

Figure IV.5

Exemples de dimensions de pixels dans des photographies aériennes et des images satellitaires



gions sélectionnées sur la trajectoire du satellite. Une option moins coûteuse consistera à obtenir des images dans des délais moins restrictifs, en utilisant des archives d'images dont une partie pourra être achetée à un coût nettement inférieur. Le prix des images dépendra aussi du degré de traitement des données brutes. Il peut s'agir de correction radiométrique, de correction géométrique et de géoréférencement, sans ou avec des points de contrôle au sol. Ainsi, les images archivées disponibles auprès de Digital Globe (à partir de mars 2008) sont facturées 16 dollars par kilomètre carré (km²), pour une commande minimale de 25 km². Les images Ikonos sont généralement moins onéreuses, environ 8 dollars par km². Les prix seront fonction de la superficie couverte par les images achetées — les zones plus étendues coûtant moins au km². Les données brutes des images coûteront beaucoup moins cher qu'une carte orthophotographique numérique produite à partir d'images satellitales. Cependant, les images achetées sont normalement totalement traitées. Le tableau IV.1 propose une liste de produits de télédétection de satellites civils à très haute résolution spatiale. Le Bureau des affaires spatiales du Secrétariat des Nations Unies gère une liste plus exhaustive (voir www.oosa.unvienna.org). Parmi les autres capteurs, citons ALOS, le système japonais utilisé en Amérique, l'Alaska Satellite Facility (ASF) et CBS26, ainsi que l'IMPE, le système chinois/brésilien gratuit pour l'Afrique.

Tableau IV.1

Produits de télédétection de satellites civils à résolution spatiale très élevée

Produit	Entreprise	Lancement	Mode	Dimensions de pixels au nadir	Hauteur (km)
Quickbird	Digital Globe	2001	Pan/4ms	0,61/2,44	450
Ikonos 2	GeoEye	1999	Pan/4ms	0,82/3,28	680
OrbView 3	OrbImage	2003	Pan/4ms	1,0/4,0	470
Spot 5	SpotImage	2002	Pan/4ms	5(2,5)/10	830
Cartosat-1	NASDA, Japon	2004	Pan	2,5	617
Cartosat-2	NASDA, Japon	2004/5	Pan	1	630

4.56. Les images peuvent être commandées en ligne ou auprès d'un revendeur local ou régional. Autre solution sans doute moins coûteuse : l'utilisation de l'infrastructure de données spatiales du pays pour accéder à des archives d'images communes. Les représentants d'organismes nationaux de statistique peuvent contacter d'autres agences, en particulier l'agence de cartographie nationale, afin de s'enquérir de la consultation des archives de télédétection existantes aux fins des recensements, en signant au besoin une convention d'utilisation stipulant exactement de quelle manière les images peuvent être exploitées.

4. Sources en ligne de données de télédétection satellitale

4.57. On peut dire que les nouvelles ressources d'images satellitales en ligne sont tout aussi importantes pour les organismes nationaux de statistique. Elles peuvent procurer une aide visuelle dans le cadre du recensement tout en évitant le coût et les démarches liés à l'achat d'images. Les données de télédétection sont désormais consultables en ligne, à l'aide d'applications telles que Google Earth, ArcGIS Explorer, Microsoft Virtual Earth et d'autres sources de données en ligne. Par rapport à l'achat d'images, la récupération de données en ligne a pour avantage de permettre aux organismes nationaux de statistique de tester l'applicabilité des images sans procéder aux investissements en amont. Le revers de la médaille, c'est que la résolution et la qualité

globale de l'image ne se prêteront peut-être pas à une cartographie détaillée des cartes de recensement.

4.58. Google Earth est un programme virtuel réalisant la cartographie de la Terre en cataloguant et en affichant des images satellitaires. Cette solution a eu un impact sur la communauté géospatiale et développe l'intérêt du public pour la technique et les cartes satellitaires. L'application a été développée par Keyhole, ensuite rachetée par Google en 2004. La plupart des images de Google Earth proviennent de l'application Quickbird de Digital Globe, même si certaines images aériennes et bâtiments en trois dimensions sont désormais repris. Google Earth propose une résolution spatiale de 15 mètres ou mieux pour la plupart des régions du monde, grâce à une projection géographique et aux caractéristiques du WGS84. Les images de Google Earth sont protégées par copyright. Les utilisateurs ne peuvent accéder à la source réelle des données mais seulement la visualiser. Ils ont toutefois la possibilité d'ajouter leurs propres données.

4.59. À l'heure actuelle, Google Earth propose trois niveaux de licence : la version gratuite, une version Plus pour un abonnement annuel de 20 dollars et une version Pro pour des applications commerciales à raison de 400 dollars par an (prix de mars 2008 — vérifier la tarification la plus récente). La version Plus permet l'intégration d'un GPS. Grâce à cette option, l'utilisateur est en mesure de lire les traces et points de cheminement d'un dispositif GPS. Par ailleurs, Google Earth Plus propose une assistance directe pour les gammes de produit GPS Magellan et Garmin, une meilleure résolution d'impression, une aide à la clientèle par courriel et un importateur de données pouvant lire des points d'adresse à partir d'un tableur, en utilisant des valeurs séparées par des virgules. Cette option est toutefois limitée à 100 points/adresses.

4.60. La version Pro comprend des logiciels complémentaires, notamment pour réaliser des films, et permet à l'utilisateur de représenter les données basées sur l'emplacement à l'aide d'outils de dessin en 3D. Il est par ailleurs possible de transférer jusqu'à 2 500 localisations de points à partir d'une feuille de calcul. Grâce à un module d'importation de données SIG, des données géographiques peuvent être ajoutées, dont des informations démographiques, aux formats shapefile et .tab. Avec les outils de mesure intégrés dans la version Pro, l'utilisateur sait calculer des superficies et des distances linéaires; des images pouvant atteindre 11 x 17 pouces ou 4 800 pixels sont exportables. Google Earth Pro n'est pas commercialisé en ligne, mais doit être acheté par l'intermédiaire d'un représentant commercial.

4.61. Quelle que soit la version de Google Earth, les données géographiques au format .kml (Keyhole Markup Language) peuvent être exportées vers Google Earth. Un script téléchargeable gratuitement peut convertir des points et polygones au format adéquat, bien que l'importation des polygones puisse être plus difficile. Une topologie rudimentaire (par exemple pas de tableaux d'attributs de polygone ou d'emplacements de nœuds) rend impossible l'importation de données problématiques. La version Pro peut importer des shapefiles mais pas les exporter. Les cartes scannées pour un recensement, telles que des cartes de districts de recensement, peuvent poser problème en cas d'importation dans Google Earth puisque la taille des fichiers est relativement fort limitée (18 000 x 18 000 pixels). Un script Python « regionator » permet de surmonter ce problème mais peut compliquer la gestion des fichiers.

4.62. Google Earth, en particulier la version Pro et ses fonctions d'exportation d'images, peut exécuter plusieurs tâches utiles pour les organismes nationaux de statistique, bien que ces fonctions soient particulièrement limitées par rapport à une imagerie indépendante. Google Earth a pour atouts notamment son faible coût et sa facilité d'utilisation pour certaines tâches peu spécialisées. La résolution spatiale de

15 mètres de la plupart des images propose un certain degré de détail du paysage aux délinéateurs de districts de recensement mais celui-ci n'est toutefois pas suffisant pour comptabiliser les unités d'habitation. Les points faibles de Google Earth ? Notamment l'absence d'une résolution pour une délimitation des districts de recensement, le difficile transfert des images dans un programme SIG, la nécessité d'une connexion Internet ultra-rapide pour le téléchargement des images, ainsi que les métadonnées et l'authenticité. D'autres sources de données satellitaires en ligne, telles que celles obtenues par l'application gratuite ArcGIS Explorer de l'Environmental Systems Research Institute (ESRI), peuvent contourner certains problèmes rencontrés avec Google Earth en permettant l'importation directe d'images dans des projets de SIG. Toutefois, la résolution spatiale des images disponibles pourrait ne pas être suffisante pour certaines applications de recensement.

4.63. Les images satellitaires de haute résolution fournissent un niveau de détail géographique comparable à celui de cartes orthophotographiques numériques créées à partir de photographies aériennes. Cependant, on se heurte à une difficulté importante, à savoir qu'il est plus difficile d'obtenir des images sans nuages à partir de satellites qu'à partir d'avions volant à basse altitude et opérant selon un horaire souple. Les images à haute résolution sans nuages permettent de dénombrer les habitations, d'évaluer la population et de délimiter les districts de recensement. La photographie aérienne est souvent opérée de manière ponctuelle et peut être plus appropriée pour des projets de cartographie et des levés détaillés. Les nouvelles techniques de photographie aérienne numérique gagnent en popularité et peuvent offrir une qualité d'image supérieure même par rapport aux images satellitaires haute résolution (voir par. 4.73 à 4.89 ci-dessous).

5. Applications de données de télédétection pour l'analyse de la population

4.64. Les techniques de télédétection peuvent identifier des zones connaissant une croissance rapide ou des changements, ce qui permet aux organismes nationaux de statistique d'affecter les ressources aux endroits où elles sont le plus nécessaires. L'analyse démographique à l'aide de la télédétection en est à ses balbutiements, mais elle s'avère de plus en plus bénéfique. Un rapport du National Research Council aux États-Unis (NRC, 2007) a indiqué que les méthodes alternatives d'estimation de la population n'étaient pas encore suffisamment robustes pour les interventions humanitaires. Des capteurs à haute résolution, tels qu'Ikonos et Quickbird, n'offrent pas le même niveau de détail pour les données d'archives que Landsat; qui plus est, la couverture de zones étendues peut être très onéreuse. Toutefois, comme indiqué, Landsat pose problème pour l'estimation de la taille de la population et souffre depuis longtemps de défaillances au niveau du capteur, sans compter la variabilité des financements. Les chercheurs devraient dans un avenir proche disposer d'autres alternatives permettant de remplacer Landsat par d'autres données à haute résolution, comme SPOT.

4.65. Les caractéristiques de la population qu'il est possible d'évaluer à l'aide d'images satellitaires sont entre autres le nombre d'habitations, les surfaces terrestres urbanisées (taille de l'habitat) ainsi que les estimations de la couverture des sols/l'utilisation des terres en lieu et place de l'expansion résidentielle et de la densité de la population (Jensen et Cowen, 1999). Dans certaines zones sujettes aux catastrophes naturelles, les photos aériennes sont plus avantageuses que les images satellitaires grâce à leur capacité à capturer des scènes en-dessous de la couverture nuageuse. Le radar

doit encore jouer un rôle significatif dans l'analyse de la population en dépit de sa capacité à pénétrer les nuages.

4.66. La « détection des changements » peut être une solution pour mesurer l'évolution démographique sur un plan spatial, en particulier pour identifier les zones de croissance rapide, en saisissant deux images ou plus du même endroit sur une période de cinq ans voire plus. Pour quantifier l'expansion de l'urbanisation, les analystes classent chaque image en recourant à une technique de classification brute, considérant chaque pixel soit comme urbain soit comme non urbain en fonction de la signature spectrale. L'évolution de la couverture des sols peut alors être calculée en superposant les images et en mesurant la croissance des zones habitées.

4.67. Une étude de cas (Yankson, 2004) citée par Antos s'est basée sur les données Landsat TM de 1985, 1991 et 2002 pour calculer les taux annuels d'expansion de la superficie d'Accra au cours de cette période. Yankson a découvert qu'entre 1984 et 1991, Accra s'était développée d'environ 10 km² par an. Ce chiffre est passé à 25 km² par an entre 1991 et 2001. L'étude portait uniquement sur l'expansion de la superficie. Pour mettre le doigt sur des changements internes majeurs susceptibles de résulter du développement, il convient d'adopter une approche de classification élastique mesurant le développement urbain en tant que variable continue. Plutôt que de classer un pixel comme urbain ou non, l'analyste peut également se servir des données des sous-pixels pour catégoriser chaque pixel comme ayant un certain degré d'urbanisation, traduit par un pourcentage. Cela plaide pour une utilisation systématique des SIG dans le cadre des activités de cartographie de recensement.

4.68. Dans le domaine de la santé, des études ont été basées sur des images télédéteectées afin de se concentrer sur des disparités intraurbaines dans des phénomènes tels que la prévalence de la maladie. Castro (2004) s'est servi de la photographie aérienne et des polygones de voisinage pour identifier des sites de reproduction de la malaria potentiellement nuisibles. La détection des zones d'habitat informelles et des voisinages peut être délimitée à travers une reconnaissance structurelle de la signature particulière desdites zones. Il peut s'agir de toits denses, d'une faible croissance végétative voire de routes et chemins majoritairement en terre battue. En général, les voisinages informels tendent à avoir une texture minimale, qui se traduit par une faible variabilité de la luminosité et de fortes concentrations de surfaces imperméables (Weeks, 2007).

4.69. Enfin, les applications démographiques les plus avancées recourant aux données de télédétection reposent sur une analyse axée sur les objets plutôt que sur les pixels. Pellika (2006) a illustré la segmentation d'une image haute résolution en zones de taille, forme et couleur similaires grâce à une application automatisée, puis identifié chaque zone comme un type de surface particulier (ex. : un « toit »). Une telle technique permettrait de grouper et d'identifier tous les toits d'un type particulier dans le cadre d'une zone d'habitat informelle. Appliquant cette technique à plusieurs domaines d'étude, Pellika a pu obtenir une précision de 97 %. L'analyse d'images basée sur des objets (OBIA) présente certains inconvénients : bien qu'automatisée, elle continue à être fastidieuse; qui plus est, elle dépend du site et de données détaillées haute résolution. Pour le moment, elle requiert un logiciel coûteux et une expertise bien trop souvent hors de portée de la majeure partie du personnel d'un organisme national de statistique.

4.70. Comme pour les photographies aériennes, l'acquisition d'images satellitales — moins coûteuse cependant que des missions aériennes — reste chère. Les données satellitales de haute résolution devraient donc être obtenues dans le cadre d'un accord de partage de coûts avec d'autres agences ou elles devraient être utilisées sélectivement pour des zones dont la couverture cartographique reste insuffisante.

6. Avantages et inconvénients des données obtenues par télédétection satellitale

4.71. Les avantages des données obtenues par télédétection satellitale sont notamment les suivants :

- On peut assurer la couverture actualisée de zones très étendues à un coût relativement faible au moyen d'images de faible résolution.
- Les images de haute résolution permettent de couvrir des zones à un niveau de détail suffisant pour une délimitation des districts de recensement, à condition que des estimations de la population existent pour les zones délimitées.
- Les images permettent de cartographier les zones inaccessibles.
- Les images peuvent servir à effectuer un contrôle indépendant d'un processus de vérification sur le terrain.
- Les images ont plusieurs finalités; une fois achetées, elles peuvent être utilisées dans d'autres applications.
- Les sources d'images en ligne peuvent présenter un certain degré de fonctionnalité à faible coût (ou gracieusement).
- Il est possible de mettre à jour les cartes topographiques dans les zones rurales; par exemple, on pourra identifier de nouveaux habitats ou de nouveaux villages qui ne figurent pas sur les cartes.

4.72. Les inconvénients sont les suivants :

- Avec de nombreux systèmes, en particulier les moins coûteux, la résolution est insuffisante pour des applications censitaires.
- Dans le cas de capteurs optiques, la couverture nuageuse et le couvert végétal rendent difficile l'interprétation de l'image.
- Le manque de contraste entre divers objets — par exemple les routes en terre battue ou les matériaux de construction traditionnels dans les zones rurales — rendent leur délimitation particulièrement difficile dans les pays en développement.
- Le traitement des images exige des connaissances considérables dont les organismes nationaux de statistique ne disposent peut-être pas. Parmi les autres défis qui se posent pour les organismes nationaux de statistique en matière de ressources humaines, les agences peuvent décider de faire appel à une expertise externe pour mener à bien leurs tâches cartographiques. Une telle approche de collaboration institutionnelle permet de partager les dépenses et l'expertise.

D. Photographie aérienne

1. Aperçu de la photographie aérienne

4.73. En dépit de l'engouement pour les images satellitales haute résolution, la photographie aérienne garde toute son utilité pour les applications de cartographie exigeant une grande précision et une exécution rapide des tâches. Les photographies aériennes sont semblables aux cartes et aux images satellitales en ceci qu'elles offrent une vue d'en haut d'objets situés à la surface terrestre. Elles en diffèrent en ceci qu'elles montrent seulement les objets effectivement visibles au sol. Les limites

artificielles, les informations thématiques et annotations sont naturellement absentes. Sans autre traitement, les photographies aériennes n'offrent pas non plus l'exactitude géométrique d'une carte. L'angle de prise de vue et les différences de relief du terrain induisent des déformations sur la photographie aérienne. Il faut donc procéder à un traitement supplémentaire pour obtenir des cartes appelées orthophotoplans, combinant l'exactitude d'une carte topographique à la foule de détails d'une photographie (voir encadré IV.2).

4.74. La photogrammétrie — technique qui consiste à obtenir des mesures à partir d'images photographiques — sert à créer et mettre à jour des cartes de base topographiques, réaliser des levés agricoles, des études de sols et de nombreuses tâches de planification urbaine et régionale. Pour les projets de recensement, on a aussi fréquemment recours à la photogrammétrie pour créer rapidement des cartes de zones pour lesquelles il n'existe pas de cartes à jour ou dont on peut difficilement dresser le levé avec les méthodes traditionnelles sur le terrain. Une mission de photographie aérienne effectuée peu de temps avant un recensement offre la base la plus complète pour délimiter les districts de recensement dans des délais raisonnables.

4.75. Les photographies aériennes ont été déjà utilisées à des fins cartographiques peu de temps après l'invention de l'avion. Les premières applications faisaient appel à des appareils photographiques ordinaires. Mais très vite, des appareils spécialement construits pour réduire au minimum les déformations géométriques furent montés sur des avions adaptés spécialement pour permettre au système photographique de faire face au sol à travers un orifice dans le plancher. D'importants perfectionnements furent apportés en peu de temps au matériel utilisé pour interpréter les photographies aériennes et convertir en cartes les informations extraites de ces photographies. Par exemple, l'interprétation de paires d'images stéréoscopiques devint la méthode principale pour établir des cartes hypsométriques.

4.76. La photographie aérienne fait appel à des appareils de prise de vues installés à bord d'avions volant à basse altitude. L'appareil enregistre l'image soit sur du film photographique, soit numériquement. Traditionnellement, le film a offert une résolution (capacité de distinguer de petits détails) supérieure à celle obtenue à l'aide de systèmes à capteurs numériques, même si les récents développements dans le domaine de l'imagerie numérique ont changé la donne. D'après la documentation de l'entreprise de photographie aérienne MJ Harden, les nouveaux capteurs de pointe peuvent saisir des images de 12 bits avec une résolution au sol de seulement 1½ pouces par pixel d'image, et 4 096 nuances de niveaux de gris contre 256 pour un film. L'entreprise sud-africaine Rob Wooding and Associates marque sa préférence pour les images aériennes numériques par rapport aux images satellitales de 1 m, jugeant les images aériennes moins coûteuses et plus précises. Cela dépendra de l'emplacement de la zone survolée. Les besoins doivent donc être évalués au cas par cas.

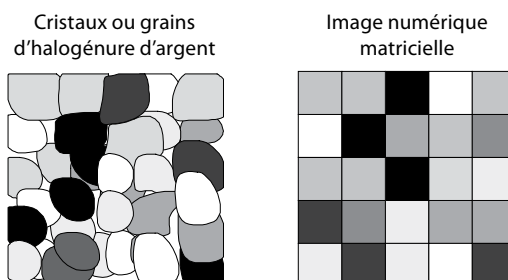
4.77. Habituellement, les produits finaux d'un projet de photographie aérienne sont des photos imprimées d'une zone au sol. La mission photographique est conçue de telle sorte que les photos obtenues se recouvrent de 30 à 60 %. L'opérateur de photogrammétrie peut combiner ces photographies pour produire une mosaïque continue couvrant la région toute entière. Les mosaïques de photographies aériennes imprimées peuvent être utilisées de la même façon que des cartes. Elles peuvent être annotées, fournir une référence pour le travail sur le terrain et permettre la numérisation d'objets en vue de créer ou de compléter des bases de données géographiques.

4.78. Les récents progrès dans le traitement d'images numériques ont changé la manière de transformer les images aériennes en produits utiles. Dans le cas des systèmes analogiques, la photo est généralement un produit intermédiaire. La méthode

la plus courante consiste à convertir le négatif en une épreuve transparente qui est numérisée à l'aide d'un scanner haute résolution. Il en résulte une image numérique pouvant être affichée et traitée sur un ordinateur. La figure IV.6 illustre la différence microscopique entre la photographie et l'image numérique. Le film photographique noir et blanc, par exemple, est constitué d'une suspension de petits cristaux photosensibles d'halogénure d'argent dans une couche de gélatine. Ces cristaux ou grains ont des dimensions et des formes irrégulières. Par contre, cette image numérisée par balayage est formée d'une trame régulière de pixels (images élémentaires).

Figure IV.6

Film photographique ou image numérisée par balayage



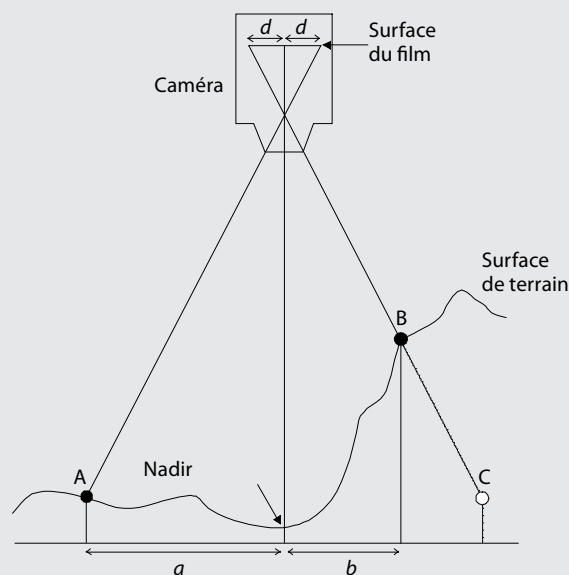
Encadré IV.2

Établissement d'orthophotoplans numériques

Pour obtenir des orthophotoplans numériques, on doit éliminer les déformations de l'image dues à l'angle de prise de vue et aux irrégularités du terrain. Ce dernier type de déformation est illustré par la figure IV.7 (d'après Jones, 1997). La photographie est essentiellement une projection en perspective de la surface terrestre. Le point B est situé à une altitude supérieure à celle du point A. En réalité, B se trouve à une distance b du nadir, qui est le point situé verticalement au-dessous du centre de perspective de l'objectif de l'appareil photo. Mais dans l'appareil la projection en perspective donne une fausse impression. Le point B paraît situé à l'emplacement du point C et est donc projeté à la même distance du centre de la surface du film que le point A.

Figure IV.7

Déformation due au terrain



Pour corriger les déformations sur la photographie aérienne, nous avons donc besoin de connaître l'altitude de chaque point du sol. L'altitude peut être déterminée à partir de paires stéréoscopiques de photographies aériennes. Il s'agit de photographies couvrant à peu près la même surface au sol, mais écartées l'une de l'autre par un faible espace. Des appareils de stéréorestitution analytique permettent à l'opérateur d'enregistrer conjointement la paire d'images stéréoscopiques et d'en extraire des emplacements d'objets en trois dimensions. Les systèmes de visualisation cartographique sur écran les plus récents assurent une automatisation poussée de l'enregistrement des images et de l'élimination des déformations. Tous les paramètres pertinents, comme l'inclinaison de l'appareil de prise de vues pendant le vol et les distorsions de l'objectif, peuvent être pris en compte. L'opérateur peut ainsi extraire des données correctement géoréférencées des photos aériennes. Les produits sont notamment des données vectorielles de SIG directement obtenues à partir des photos aériennes, des cartes à maillage graphique montrant le terrain ou des modèles numériques de terrain (MNT) — images matricielles correspondant à la photographie aérienne, où chaque valeur de pixel indique l'altitude du point considéré sur le sol. Si un MNT n'est que modérément utile pour des applications cartographiques de recensement, les lots de données ainsi obtenus sont par contre très utiles pour des applications environnementales ou des études de ressources naturelles, particulièrement en hydrologie.

Après ces opérations d'enregistrement dans un système de référence géographique appropriée et une fois les déformations éliminées, les photographies aériennes initiales ont été converties en orthophotoplans numériques. Ces cartes sont en général établies aux échelles de 1/2 000 et 1/20 000 suivant l'altitude de l'avion et le traitement effectué. Les orthophotographies voisines peuvent être numériquement combinées pour créer des bases de données images continues pour une ville entière, une région ou même tout un pays. Les cartographes peuvent extraire ou délimiter des objets sur des cartes orthophotographiques par numérisation sur écran. Ou ils peuvent simplement les utiliser comme fond pour servir de contexte aux couches de données existantes du SIG.

2. Utilisation de la photographie aérienne pour établir les cartes de recensement

4.79. Les cartes orthophotographiques sont tout à fait appropriées pour effectuer des comptages d'unités d'habitation et des évaluations de population. Les dénombrements d'habitations ou de population au moyen de photographies aériennes sont parfois appelés recensements de toits. Dans un milieu rural, où les habitats sont faciles à distinguer sur la photographie aérienne et où les maisons sont plus ou moins dispersées, le nombre d'habitations peut être déterminé assez facilement. Une évaluation fiable du nombre moyen de personnes par ménage permet alors d'évaluer avec une précision suffisante la population à des fins de recensement. En zone urbaine, les maisons peuvent être très rapprochées les unes des autres. Le nombre de familles vivant dans des habitations à plusieurs étages peut aussi être difficile à déterminer. Mais même ainsi, moyennant un peu de formation et une certaine connaissance de la zone, on pourra encore effectuer des évaluations de la population avec un degré de précision suffisant. Le personnel de recensement pourra alors définir des limites de districts de recensement englobant un nombre déterminé d'habitations. Les orthophotographies étant correctement géoréférencées, les districts de recensement obtenus seront aussi enregistrés sur une projection cartographique appropriée avec des paramètres connus. Cela signifie qu'il sera inutile de procéder à un géoréférencement parfois fastidieux pour rendre les limites numériques compatibles avec d'autres données de SIG.

4.80. L'interprétation des photographies aériennes est le plus souvent fondée sur l'interprétation visuelle. Le personnel cartographique de recensement n'a donc pas besoin de recevoir une formation en matière de techniques avancées de traitement d'image. Les limites de districts de recensement peuvent être tracées sur la photographie aérienne. Des détails géographiques supplémentaires fournissant aux agents recenseurs des références géographiques peuvent aussi être extraits des photographies.

Il est possible de délimiter ces objets de façon interactive sur l'ordinateur à l'aide d'une souris ou d'un dispositif de pointage semblable. Le personnel de recensement peut également imprimer les photos et tracer les objets sur des feuilles de film plastique (acétate ou mylar). Celles-ci pourront alors être numérisées par balayage et vectorisées. Ce processus nécessite une opération supplémentaire et davantage de matériaux, mais il améliore souvent la précision du produit obtenu (voir aussi les sections sur la numérisation et le scannage).

4.81. Il convient de souligner qu'un fichier de limites des districts de recensement reposant soit sur une télédétection satellitale soit sur des photographies aériennes n'a aucun sens en l'absence de données de référence adéquates — des repères et noms de rue par exemple — puisque les agents recenseurs ne seront peut-être pas en mesure de se situer.

3. Mise en œuvre et questions institutionnelles liées à la photographie aérienne

4.82. La construction d'orthophotoplans numériques exige de vastes compétences en matière de méthodes photogrammétriques, compétences que l'organisme de recensement ne possède généralement pas. Il lui faut donc passer un accord de coopération avec un autre organisme national, le plus souvent le service cartographique ou une unité de reconnaissance de l'armée de l'air. Sinon, le travail peut être confié à une entreprise commerciale de cartographie aérienne. Il existe plusieurs entreprises de cartographie actives sur le plan international qui fournissent l'avion, l'appareil de prises de vues et le matériel de traitement.

4.83. Mais ces services ne sont pas précisément bon marché. Heureusement, les photographies aériennes sont utilisables pour de nombreuses applications différentes, y compris la planification de fournitures de services, la mise à jour de cartes urbaines et des projets d'enregistrement de titres fonciers. Le partage des coûts entre plusieurs départements ministériels et éventuellement avec le secteur privé peut sensiblement réduire les dépenses pour l'organisme de recensement. Si les ressources limitées ne permettent pas de réaliser des clichés couvrant la totalité du territoire national, on pourra cependant réaliser des photographies aériennes pour certaines zones. Par exemple, l'Office de la statistique de Hong-Kong utilise la photographie aérienne pour évaluer le nombre de personnes vivant sur des bateaux (Netherlands Interdisciplinary Demographic Institute, 1996). On voit ainsi que ces techniques peuvent être utilisées pour dénombrer des populations difficiles à recenser. On peut citer également les populations nomades ou les populations de réfugiés, les zones urbaines connaissant une croissance rapide ou les régions inaccessibles pendant certaines saisons.

4.84. Ainsi qu'on l'a indiqué aux paragraphes précédents, l'élaboration des cartes orthophotographiques exige des connaissances techniques considérables et un matériel spécialisé. Par contre, l'utilisation de ce type de carte n'exige pas de formation poussée complémentaire, même si elle doit cadrer avec le plan global de recensement. La base de données d'une grande ville, par exemple, consistera uniquement en une mosaïque de plusieurs images sur un support de stockage portable, tel qu'un DVD, qui peut être visualisé sans solution de continuité dans un SIG standard ou un logiciel de cartographie informatisée. Les cartes orthophotographiques numériques peuvent être obtenues dans des formats graphiques courants [tels que TIFF (format de fichier image étiqueté)]. L'utilisateur n'a donc pas besoin d'un logiciel spécialisé de traitement d'image. En fait, on peut utiliser n'importe quel logiciel graphique pour extraire des objets des images, même si les informations de géoréférencement sont perdues. Ces

informations concernent les dimensions et les coordonnées réelles de l'image numérique et sont en général contenues dans un petit fichier d'en-tête. Avec ces informations, la plupart des progiciels de cartographie informatisée peuvent enregistrer les images ainsi que tout autre lot de données de SIG stocké dans le même système de référence géographique.

4.85. Les cartes orthophotographiques sont aussi utilisées comme fond de carte offrant un contexte approprié pour visualiser des emplacements de points collectés au moyen d'un GPS ou des objets numérisés tels que des équipements sanitaires ou des réseaux de transport. Par le passé, outre les cartes de district de recensement, des orthophotoplans numériques imprimés montrant les limites des districts de recensement ont pu être fournis aux agents recenseurs pour les aider à s'orienter dans la zone qui leur est affectée. Aujourd'hui, il y a néanmoins plus de chances que les images de télédétection soient intégrées dans un projet de SIG que dans une carte séparée.

4.86. Le volume considérable de données nécessaires pour travailler avec des cartes orthophotographiques numériques de haute résolution représentant des zones étendues freine l'application de cette technologie dans les bureaux de recensement. Pour un bureau de recensement, il peut ainsi être préférable d'obtenir des photographies aériennes numériques de moindre résolution qui présentent suffisamment de détails pour les applications censitaires et qui seront plus faciles à traiter et à stocker. Les orthophotoplans numériques ont souvent une très haute résolution avec des dimensions de pixels au sol de l'ordre du centimètre (en général, 5 à 30 cm). Des images orthophotographiques numériques obtenues par échantillonnage répété avec des dimensions de pixels comprises entre 0,5 et 2 mètres sont suffisantes pour délimiter des districts de recensement en zone urbaine.

4.87. L'avenir de la photographie aérienne est un processus purement numérique rendant inutile la production intermédiaire de photographies imprimées. Des systèmes utilisant un contrôle GPS en vol et des appareils de prises de vues à trame numérique sont déjà opérationnels. Les appareils de prises de vues à trame numérique utilisent des ensembles de dispositifs à couplage de charge (CCD) qui peuvent fournir des images de 9 216 sur 9 216 pixels, avec une exactitude de position de 1 à 4 centimètres. Les opérations intermédiaires de production de tirages photographiques puis de scannage étant supprimées, cette technologie est beaucoup moins coûteuse et beaucoup plus rapide que la technologie photographique traditionnelle. La résolution des appareils de photo numériques va augmenter régulièrement de même que les vitesses de traitement informatique. Il est donc probable que la photographie aérienne classique sera remplacée dans un proche avenir par une cartographie aérienne précise, pleinement numérique et en temps réel.

4. Avantages et inconvénients des photographies aériennes

4.88. Les avantages des photographies aériennes sont notamment les suivants :

- Les photographies aériennes fournissent une grande quantité de détails et peuvent être interprétées visuellement. On dispose simultanément d'informations sur de nombreux types d'objets, routes, cours d'eau et bâtiments.
- La collecte des données est plus rapide et les données cartographiques peuvent donc être obtenues beaucoup plus vite qu'au moyen de levés cartographiques sur le terrain. Les photographies aériennes récentes offrent donc pour l'élaboration de cartes de recensement une base plus fiable que des cartes rarement mises à jour.

- Les photographies aériennes peuvent servir à établir des cartes pour des zones d'accès difficile ou des zones dans lesquelles le travail sur le terrain est difficile ou dangereux.
- Le recours aux photographies aériennes pour établir des cartes topographiques peut revenir moins cher que l'utilisation des techniques traditionnelles de levé de terrain. Cependant, les cartes de recensement n'exigent pas une aussi grande précision que les cartes topographiques, et les coûts élevés ne se justifient pas nécessairement si les produits sont utilisés uniquement pour la cartographie de recensement.
- Les tirages de photographies aériennes sont utiles pour le travail sur le terrain parce qu'ils fournissent une image d'ensemble. L'agent recenseur peut replacer le terrain visible de son point de vue dans le contexte plus large de la zone environnante. Les photos aériennes numériques sont utiles comme cartes de fond pour l'affichage de lots de données de SIG.

4.89. Les inconvénients sont les suivants :

- Le traitement des photographies aériennes exige un matériel coûteux et des connaissances spécialisées et les bureaux de recensement sont donc obligés de collaborer avec d'autres agences pour accéder à des orthophotographies ou de faire appel à une aide extérieure.
- Les photographies aériennes peuvent être protégées par des droits d'auteur avec des droits de distribution limités.
- Les photographies aériennes doivent encore être complétées par des renseignements concernant les noms d'objets qui doivent être extraits de cartes éventuellement périmées. La photographie aérienne ne rend pas nécessairement inutile le travail sur le terrain. Elle ne conviendra peut-être pas aux zones distantes mais peut servir de solution de secours pour les zones difficiles à recenser.
- L'interprétation des photographies aériennes peut être difficile lorsque des objets sont cachés par une végétation dense ou une couverture nuageuse ou lorsqu'un contraste limité ne permet pas de distinguer clairement des objets adjacents (par exemple des habitations en matériaux naturels et le terrain avoisinant).
- Les photographies aériennes comprennent de très grandes quantités de données numériques et exigent par conséquent des ordinateurs très puissants pour l'affichage et le traitement ultérieur.

E. Résumé et conclusions

4.90. Le chapitre IV a passé en revue l'utilisation d'outils tels que les systèmes de positionnement universel (GPS) et la télédétection pour des activités de recensement en les intégrant à des travaux sur le terrain. Ces outils géospatiaux permettront aux organismes nationaux de statistique de concentrer leurs efforts sur les zones à évolution rapide du pays. À ce stade du processus de recensement, l'organisme a développé une base de données géographique des districts de recensement. Il s'agira ensuite de concevoir, imprimer et distribuer des cartes de la base de données géographique à utiliser pour le recensement proprement dit.

Chapitre V

Utilisation de bases de données géographiques (cartes) pendant le recensement

A. Introduction : utilisation d'outils géospatiaux pendant le recensement

5.1. Les cartes servent à planifier diverses étapes du recensement, immédiatement avant et pendant celui-ci. Elles sont notamment utilisées pour l'affectation d'agents de recensement au territoire; l'identification de zones accidentées ou inaccessibles; la gestion logistique pour le transport du personnel de terrain et de fournitures; la localisation de populations difficiles à recenser et de lieux de vie collectifs; la définition de limites administratives à divers niveaux; le suivi des progrès du recensement ainsi que la création de cartes de positionnement.

5.2. Le présent chapitre traite de la création de telles cartes de recensement et s'attarde sur la façon de tirer parti de la polyvalence de la base de données géographique afin de fournir les bonnes informations aux agents de recensement, de manière opportune et bien organisée. Plus généralement, il explique comment la technologie géospatiale peut faciliter les opérations en cours de recensement. La compilation de la carte, la détermination des couches pertinentes pour les agents de recensement et les superviseurs ainsi que les préceptes de base de l'impression et de la distribution sont autant de thèmes abordés ici.

5.3. De manière générale, c'est sous l'angle de la gestion de projet que ce chapitre traitera du recensement. Une planification détaillée de la logistique minimise les erreurs et évite les retards. Le recensement est une opération territoriale, divisant le pays en unités opérationnelles qui peuvent être prospectées. Une carte d'agent de recensement numérique a pour avantage d'être adaptable au contexte spécifique du recensement dans le pays. En définissant numériquement des districts de recensement, l'organisme national de statistique dispose d'un document vivant, tirant parti des recensements précédents et de la valeur ajoutée apportée par la télédétection et le GPS.

5.4. À ce stade du processus, la compilation de la carte à partir de recensements antérieurs et l'ensemble des travaux de terrain ont fait appel au géoréférencement. Les planificateurs du recensement ont épinglé les zones ayant le plus besoin d'attention, en procédant à la détection des changements dont il est question au chapitre IV. S'il a travaillé correctement, l'organisme national de statistique disposera de tout ce dont il a besoin à ce stade, exception faite du dénombrement. Le principe ici, c'est qu'il s'agit de disposer d'une base de données soigneusement éditée et mise à jour qui servira de base aux cartes utilisées par les agents de recensement sur le terrain. Dès lors, ce qu'il faut pour le recensement, c'est extraire les cartes de la base de données des districts de recensement et les remettre aux agents de recensement.

B. Assurance de la qualité, établissement de cartes de districts de recensement et tenue à jour de la base de données

1. Aperçu

5.5. L'exactitude et la complétude des données de recensement dépendent pour beaucoup de la qualité des cartes de base utilisées par les agents de recensement. Le travail permanent de contrôle et d'amélioration de la qualité pendant la conversion des données s'achève par un examen approfondi de tous les documents cartographiques avant la distribution des cartes de districts de recensement aux agents recenseurs. Cela suppose que les administrateurs locaux vérifient l'exactitude des limites administratives. On doit résoudre les problèmes et éliminer les incohérences qui pourraient subsister avant de générer les produits terminaux.

5.6. À ce stade, des bureaux locaux auront été mis sur pied pour le recensement. Le degré de centralisation de la structure organisationnelle du recensement aura un impact direct sur les procédures de distribution des cartes des districts de recensement et de supervision sur le terrain. Dans le bureau local, une carte à plus petite échelle compilée à partir des cartes des districts de recensement peut être affichée afin de faire état des progrès accomplis. Le bureau local peut également faciliter la planification en estimant les charges de travail ainsi que les frais de déplacement, en prenant des dispositions en vue de la distribution et la réception de matériaux, en identifiant les zones problématiques et en organisant des visites du personnel du bureau sur le terrain.

5.7. L'établissement de cartes de districts de recensement est simple en théorie, à condition que la qualité de la base de données géographique numérique soit satisfaisante. Il s'agit là d'une tâche logistique ardue car des milliers de cartes doivent être distribuées ainsi que les instructions relatives à la lecture de la carte et d'autres directives.

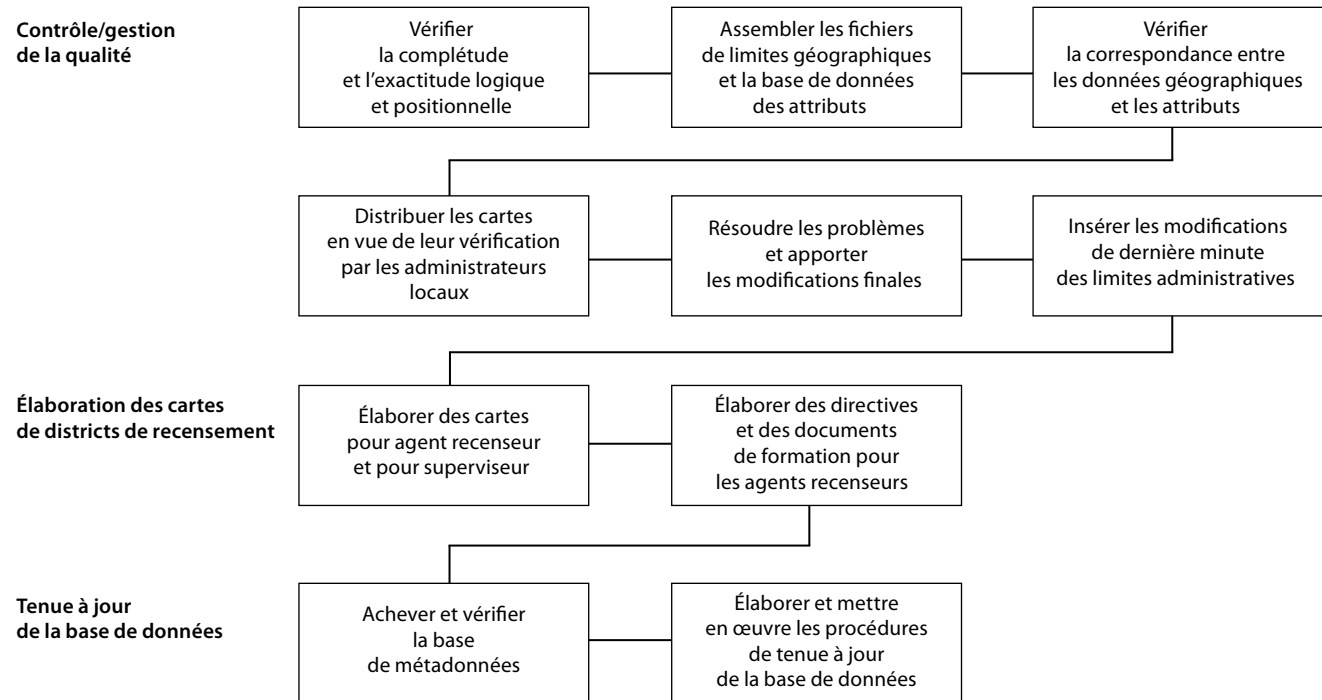
2. Établissement du projet de carte et procédures d'assurance de la qualité

5.8. Les sections suivantes passent en revue le processus de création de cartes pour le dénombrement. Les étapes sont entre autres l'intégration de bases de données géographiques, la compilation d'objets cartographiques, la vérification par les autorités locales ainsi que l'impression et la distribution de la carte.

3. Appariement des limites et des fichiers d'attributs et impression de cartes d'ensemble

5.9. Pour préparer le dessin de la carte finale et son impression, les lots de données relatives aux limites et le fichier des attributs géographiques doivent être appariés s'ils ne sont pas déjà intégrés dans une base de données cohérente. Au cours de cette opération, on doit aussi vérifier l'exactitude de l'appariement entre les données relatives aux limites et les données relatives aux attributs géographiques. Si les deux ensembles sont exacts, il devra y avoir au moins une entité cartographique (un point, une ligne ou un polygone) pour chaque enregistrement dans le fichier des attributs géographiques. Si ce n'est pas le cas, il y a une erreur dans la base de données cartographiques — un district de recensement est manquant — ou bien la table des attributs contient un enregistrement erroné ou en double. S'il y a deux polygones ou plus pour l'enregistrement d'un attribut, les agents du contrôle de la qualité doivent confirmer que les conventions définies pour ces cas sont bien respectées.

Figure V.1

Différentes étapes concernant l'assurance de la qualité, l'établissement des produits cartographiques et la tenue à jour de la base de données

5.10. Les responsables chargés de l'élaboration des cartes de districts de recensement ne doivent pas perdre de vue que les cartes de districts de recensement doivent couvrir la totalité du territoire du pays. Pour ce faire, il peut être utile de considérer la limite du district de recensement comme une « cloison » entourant la zone d'affectation de l'agent de recensement. Il y a moins d'erreurs lorsque les limites des districts de recensement empruntent les lignes médianes des rues et routes de telle manière que les agents de recensement tiennent compte des habitations d'un côté de la rue/route en laissant l'autre côté au recenseur du district de recensement suivant. Les chapitres III et IV proposent d'autres règles pour la délimitation des districts de recensement.

5.11. Lorsque les données géographiques et les informations relatives aux attributs sont correctement appariées, des étiquettes doivent être ajoutées à la carte, et on doit choisir des symboles cartographiques définissant les objets sur la carte de base (pour la diffusion des informations à travers des cartes, voir aussi chapitre VI). L'étiquetage peut être effectué de façon interactive (manuellement), semi-automatique ou automatique, en utilisant un logiciel de SIG ou un logiciel de dessin cartographique plus spécialisé. Dans un projet d'établissement de cartes de recensement de très grande envergure, l'étiquetage des objets est une tâche longue et fastidieuse. En particulier lorsque le dessin de cartes de districts de recensement est très complexe — par exemple, de nombreuses couches de cartes numériques sont combinées pour établir chaque carte de district de recensement — les ressources requises pour l'insertion correcte d'étiquettes — temps de travail et ressources informatiques — peuvent être très importantes.

5.12. La plupart des logiciels de SIG et de cartographie sur ordinateur comportent des fonctions d'insertion automatisée d'étiquettes. L'utilisateur précise simplement la zone des attributs dans la table des attributs de la base de données qui doit être utilisée pour l'étiquetage, par exemple un nom de rue ou un identificateur de bâtiment. Le système applique alors quelques règles simples pour placer les étiquettes sur chaque objet ou à proximité. L'utilisateur peut en général déterminer la dimension des étiquettes et décider si les étiquettes doivent être placées les unes sur les autres si les objets sont trop rapprochés. Cependant, dans tous les cas sauf les plus simples, il faudra encore prévoir une modification manuelle des étiquettes.

5.13. Pour les programmes d'établissement de cartes de districts de recensement de très grande ampleur, le bureau de recensement peut envisager d'acheter un progiciel spécialisé de positionnement des noms tel que Maplex. Ces programmes possèdent des algorithmes plus évolués pour vérifier que les règles les plus importantes de positionnement des étiquettes sont respectées, y compris les suivantes :

- Absence de recouvrement ou recouvrement minimal des étiquettes.
- Absence de recouvrement ou recouvrement minimal des objets et étiquettes.
- Affectation précise d'étiquettes à des objets (absence d'ambiguïté).
- Aspect d'ensemble attrayant, par exemple en ce qui concerne le type et les dimensions de la police choisie.

5.14. Les logiciels soumettent le positionnement des étiquettes à un certain nombre de règles heuristiques qui peuvent être modifiées par l'utilisateur à des fins particulières. L'utilisateur peut sauvegarder les étiquettes conçues pour une couche de données particulière dans une couche séparée de données d'annotation et les superposer à des couches d'objets géographiques le cas échéant.

4. Assurance de la qualité

5.15. De nombreux contrôles de cohérence peuvent être effectués de façon interactive sur l'écran de l'ordinateur, mais il est préférable d'effectuer la vérification finale de la qualité en utilisant des cartes imprimées sur papier. Il convient donc d'établir des cartes de grand format contenant toutes les informations qui figureront aussi sur les cartes définitives de districts de recensement. Ces cartes sont établies pour la vérification et le contrôle finals de la qualité et ce travail doit être organisé par l'unité administrative. Si elles sont imprimées à la même échelle que les cartes finales de districts de recensement, on devra disposer de plusieurs coupures de cartes pour chaque district.

5.16 L'assurance de la qualité consiste en un contrôle final de la base de données cartographique numérique avant que les produits soient mis en circulation pour l'opération de recensement. L'assurance de la qualité est semblable au contrôle de la qualité, examiné au chapitre III. Elle consiste en des vérifications opérées au moyen du logiciel ou manuellement. Certaines de ces vérifications seront opérées sur tous les produits, tandis que les vérifications plus complexes et plus longues seront réalisées sur un sous-ensemble de produits en appliquant une stratégie appropriée d'échantillonnage pour acceptation.

5.17. Le contrôle de la qualité au cours de la conversion des données porte essentiellement sur l'exactitude topologique et de position des limites et des coordonnées. Il importe de vérifier qu'il existe une concordance sans solution de continuité entre des limites numérisées et enregistrées séparément. Par exemple, les limites entre districts voisins doivent être identiques si les cartes de districts sont emmagasinées dans des fichiers cartographiques numériques séparés. Pour vérifier la qualité, on met

l'accent sur l'adéquation des produits cartographiques finals aux fins du recensement. Cela suppose que l'on vérifie plusieurs aspects de l'intégrité de la base de données, décrits dans les paragraphes qui suivent. L'assurance de la qualité n'est pas une tâche facile. Elle nécessite beaucoup de temps et de ressources et le bureau de recensement doit établir en conséquence un calendrier et un budget. Si cela est correctement effectué, le recensement sera plus précis en fin de compte.

5.18. La vérification par les cartographes du bureau de recensement nécessite l'inspection des critères d'acceptation suivants :

- **Lisibilité.** Toutes les annotations figurant sur la carte doivent être clairement lisibles. Parfois les objets tracés sur la carte sont trop nombreux, ce qui rend difficile la lecture des noms de rues et d'autres données textuelles. Certaines étiquettes de texte non essentielles peuvent être omises pour améliorer la clarté de la carte. Il faut aussi vérifier à quel objet l'étiquette de texte se réfère. Dans certains cas, il pourra être nécessaire d'ajouter des flèches pour préciser son affectation.
- **Couches de données.** L'ordre dans lequel se succèdent les couches de données tracées sur une carte est important car les couches supérieures peuvent occulter des objets importants sur une couche de données géographiques inférieure.
- **Échelle de la carte.** Par exemple, un district de recensement qui est très étendu, mais qui contient une zone de population très dense relativement petite peut nécessiter un encart ou une carte séparée pour que tous les détails puissent être identifiés.
- **Informations concernant les sources et le droit d'auteur.** Chaque carte doit énumérer toutes les sources de données exclusives qui ont été utilisées pour créer la base de données numériques utilisée pour établir la carte de district de recensement. Il s'agit notamment de la date à laquelle la carte a été produite afin que les recenseurs puissent établir si les données sont actualisées ou si des corrections supplémentaires doivent être apportées.

5. Vérification par les autorités locales et contrôle final de l'unité administrative

5.19. Pour une vérification critique de la cohérence, les cartes de districts de recensement doivent être adressées aux autorités locales pour être vérifiées. Les administrateurs locaux — à l'intérieur et à l'extérieur des services de recensement — doivent confirmer que tous les habitats et quartiers de villes moyennes et grandes sont inclus dans la base de données géographique. La participation des autorités locales à ce travail présente l'avantage que les cartes sont examinées par des personnes connaissant bien l'environnement local. Les règles d'affectation des noms et d'orthographe peuvent varier dans les pays où plusieurs langues ou dialectes sont utilisés. L'approbation des cartes par les fonctionnaires locaux réduira les risques d'erreurs d'interprétation des cartes que pourraient commettre les agents de recensement recrutés sur le plan local.

5.20. Le processus de vérification comporte la confirmation des limites d'unités administratives incluses dans les cartes de districts de recensement. Ces limites sont souvent modifiées, avec l'ajout régulier de nouveaux États, provinces et districts, quels que soient les efforts en matière de recensement. Cela peut parfois poser des problèmes à l'organisme national de statistique, qui doit pouvoir établir des statistiques sommaires pour ces unités. Les procédures de gestion des limites administratives

sont traitées au chapitre III. Idéalement, les limites administratives doivent être gelées par décision du gouvernement plusieurs mois avant le recensement. Cela assure la stabilité du cadre de référence pour la durée du recensement. La structure des limites applicable pendant cette période est celle pour laquelle les tableaux de recensement seront établis. Il existe plusieurs solutions pour traiter ces problèmes à l'approche de la période du recensement :

- a) Si les limites administratives continuent à être modifiées, une solution consiste à les suivre en permanence avant le recensement. Lorsque des changements se produisent, ils sont immédiatement insérés dans la base de données cartographique numérique. Ainsi, les limites seront celles en vigueur au moment du recensement. Cependant, il faut prévoir des ressources supplémentaires pour suivre en permanence les changements et modifications des bases de données consacrées aux limites. L'organisme national de statistique doit envisager de recourir à un moteur de base de données spatiales afin d'enregistrer la date de la création des diverses limites administratives.
- b) Dans certains pays les changements de limites sont annoncés à l'avance. Le service chargé d'établir les cartes pour le recensement peut ainsi programmer le travail consacré à ces zones pour une date avancée dans le processus d'élaboration des cartes de recensement.
- c) La dernière solution consiste pour le service cartographique à déterminer une date de gel des limites et à réviser toutes les limites à une date ultérieure, éventuellement après le recensement. Si des limites d'unités administratives modifiées recoupent des districts de recensement existants, les questionnaires pour les ménages de ces unités devront être réaffectés aux unités appropriées. Il faudra alors effectuer une opération supplémentaire après le dénombrement, ce qui ralentira la diffusion des résultats du recensement.

6. Production de la carte de district de recensement (dont l'impression de la carte)

5.21. Une fois achevées, les activités de vérification et de contrôle de la qualité pour toutes les cartes de base et délimitations des districts de recensement, les cartographes du service de recensement vont imprimer les cartes définitives de districts de supervision et de recensement. Cette opération peut être effectuée au siège de l'organisme national de statistique ou dans ses bureaux extérieurs régionaux, en fonction du type de structure organisationnelle de l'organisme. Les cartes de districts de supervision comprendront plusieurs districts de recensement et seront imprimées à une plus petite échelle cartographique (à savoir plus généralisée). La définition du tracé de la carte pour des districts de recensement pris individuellement est semblable aux procédures d'incision dans les opérations d'établissement de cartes de recensement. Les cartes de districts de recensement doivent être simples, car elles vont être utilisées par des agents de recensement qui n'ont guère l'expérience des cartes. D'autre part, elles doivent contenir suffisamment d'informations pour permettre de s'orienter facilement. Elles doivent comporter les informations suivantes :

- a) La superficie complète de la zone à recenser, définie par une ligne de limite clairement indiquée;
- b) Certains éléments des zones voisines (par exemple les zones périphériques) pour faciliter l'orientation;

- c) Toute information géographique et textuelle contenue dans la base de données cartographique de recensement de nature à faciliter l'orientation dans le district de recensement, à l'aide des conventions cartographiques courantes applicables à la symbolisation (par exemple des lignes en pointillés pour indiquer des chemins, des traits bleus pour l'eau, etc.). Les cartes de districts de recensement doivent comporter les éléments suivants :
- i) Des rues et routes;
 - ii) Des bâtiments;
 - iii) Des points de repère;
 - iv) Des détails hydrologiques;
 - v) D'autres détails notables ou pertinents tels que la topographie, les étendues d'eau, etc.;
 - vi) Une légende de carte cohérente comportant les noms et codes exacts des zones administratives et de recensement, une flèche d'orientation, une échelle graphique et une légende expliquant les symboles utilisés pour les objets géographiques.

5.22. La figure V.2 montre les éléments d'une carte hypothétique de district de recensement urbain. Tous les objets sont enregistrés dans des couches de carte séparées dans le même système de référence spatiale sous la forme de modèles graphiques. Les principaux éléments sont le réseau de voirie, les bâtiments et la couche des limites de district de recensement. En outre, les annotations, symboles, étiquettes et numéros de bâtiments sont stockés dans des couches de données séparées mais celles-ci peuvent aussi être ajoutées en cours d'exécution. Le dernier élément est un modèle constitué de limites de coupures (un encadré marquant le prolongement de la carte vers l'extérieur) et d'une légende qui est régulièrement utilisée pour tous les districts de recensement. La figure V.3 montre la carte de district de recensement complète avec tous les éléments superposés sur une présentation cartographique. Suivant le but des activités cartographiques menées à des fins de recensement et la complexité de la zone recensée, les cartes de district de recensement peuvent contenir plus ou moins d'informations que cette carte échantillon.

5.23. Dans de nombreux pays qui ne sont pas encore complètement passés au numérique en vue du cycle de recensement à venir, le tracé des cartes de districts de recensement peut être plus simple que celui proposé à la figure V.3. Par exemple, au lieu d'une carte de base numérique pleinement intégrée en format vectoriel, on peut utiliser des images matricielles de cartes topographiques comme fond pour des limites de district de recensement. Dans certains cas, les objets cartographiques peuvent être plus généralisés, par exemple en utilisant seulement les lignes centrales pour les rues et des polygones pour des quartiers urbains entiers plutôt que pour des maisons individuelles.

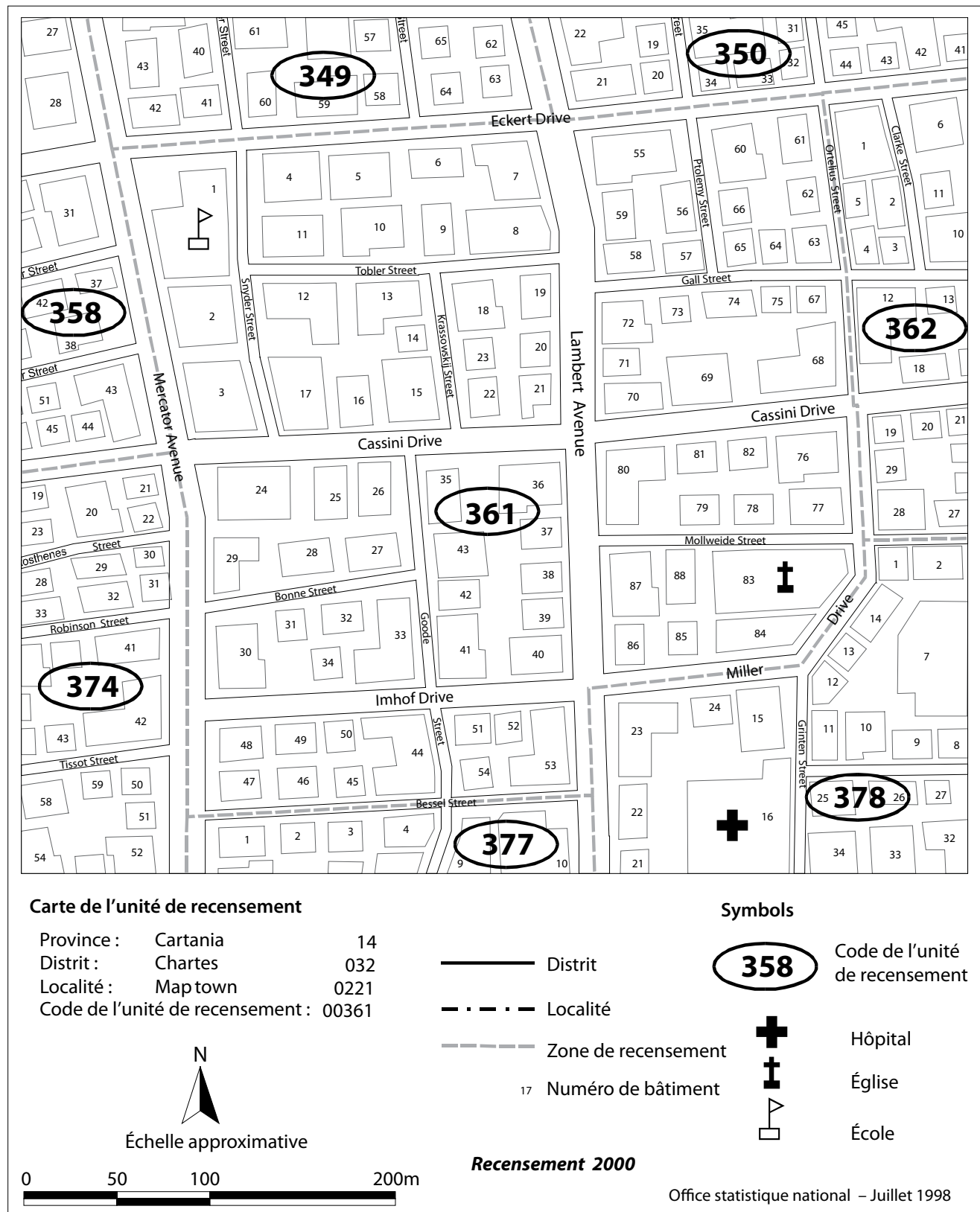
5.24. Des décisions doivent être prises concernant le format et la couleur des cartes de districts de recensement imprimées (pour les critères de sélection d'une imprimante, voir chapitre VI). Les imprimantes laser offrent une résolution élevée et par conséquent les cartes de districts de recensement peuvent en général être établies sur du papier de format A3 (420 x 297 mm, la taille de deux feuilles A4) ou sur du papier de 11 x 14 pouces, si possible. Par rapport aux imprimantes ou tables de numérisation de plus grand format, les imprimantes de format standard représentent l'avantage d'un coût moindre et d'une plus grande vitesse d'impression. Si l'on se base sur un coût à la page, encre et toner compris, les imprimantes laser sont bien moins onéreuses que les imprimantes à jet d'encre. Étant donné que l'on doit produire des milliers de cartes de districts de recensement, il s'agit de considérations importantes. Des problèmes

Figure V.2
Échantillons de composants d'une carte numérique de district de recensement



Figure V.3

Exemple de carte de district de recensement urbain



peuvent se poser dans des secteurs où un district de recensement de très grande dimension contient de petites zones à grande densité de population. Pour ces zones, on devra imprimer des cartes de plus grand format ou inclure dans le tracé des cartouches montrant des détails dans les parties denses du district de recensement.

5.25. Pour garder la trace d'un inventaire de cartes papier avant le début du recensement, les organismes nationaux de statistique doivent penser à ajouter des codes-barres sur leurs cartes. Les codes-barres peuvent aisément être mis en œuvre — un lecteur laser et un simple logiciel de base de données suffisent. Les codes-barres ne contiennent aucune donnée descriptive, mais simplement un numéro de référence aléatoire que l'ordinateur utilise pour rechercher des enregistrements connexes. Le lecteur de codes-barres détecte les zones claires et sombres de la lumière réfléchie, qu'il convertit en un signal électronique (élevé pour le noir et faible pour le blanc). De nombreux lecteurs de codes-barres sont désormais dotés d'une interface USB et d'un décodeur soit intégré dans l'unité soit indépendant. L'organisme national de statistique peut employer des étiquettes à codes-barres préimprimées ou recourir à un logiciel qui crée un code à barres pour chaque carte produite (pour un complément d'information, consulter l'adresse suivante : www.barcodehq.com/primer).

5.26. Une carte de district de recensement bien conçue donnera en général de bons résultats en noir et blanc. Bien que les imprimantes couleurs soient relativement bon marché, leur débit est limité et les fournitures sont souvent assez chères. De bonnes cartes en noir et blanc peuvent aussi être photocopiées sans pertes d'informations, ce qui permet au personnel local de produire le cas échéant des copies supplémentaires de cartes de districts de recensement. Lorsque les ressources nécessaires sont disponibles cependant, la couleur peut contribuer à rendre plus lisible le tracé de la carte. Par exemple, les limites d'un district de recensement peuvent être indiquées par une ligne brillamment colorée sur la carte. Une carte noir et blanc peut produire le même effet en passant sur les lignes des limites des districts de recensement avec un surligneur de couleur claire (l'utilisation de la couleur est débattue à l'annexe 5).

5.27. Les responsables de l'organisme national de statistique doivent envisager un essai de convivialité afin de s'assurer de la lisibilité des cartes. Sur le plan de la conception, des décisions telles que la taille, la couleur, l'échelle de la carte ainsi que les objets qui figureront, peuvent alors être tranchées en s'en remettant à l'expérience réelle de l'utilisateur et non au seul jugement d'une équipe de conception interne.

5.28. Plusieurs exemplaires de chaque carte de district de recensement doivent être établis, outre les copies de sauvegarde qui sont conservées au bureau central du service de cartographie pour le recensement. Chaque carte de district de recensement sera distribuée aux responsables locaux du recensement, superviseurs et agents de recensement, ce qui peut exiger quatre ou cinq exemplaires. La décision de distribuer les cartes sous forme numérique ou sur papier dépend du degré de centralisation du recensement. Si les activités cartographiques sont centralisées dans un ou quelques bureaux de recensement, une solution peut consister à distribuer des fichiers de cartes numériques plutôt que des cartes imprimées. Ces fichiers peuvent être communiqués aux bureaux locaux de recensement sur CD-ROM, DVD ou par Internet sur un site FTP dédié. Le bureau local n'aura pas besoin d'accéder au logiciel cartographique si les cartes sont exportées dans un format générique de fichier tel que le format de document portable (PDF) ou sous la forme d'un fichier graphique intégré à un format générique de traitement de texte. Ces fichiers peuvent être imprimés sur n'importe quel système informatique générique. Cette méthode permet au bureau local de produire autant d'exemplaires de cartes de districts de recensement que nécessaire et de

résoudre rapidement des problèmes tels que la perte de cartes imprimées. Les responsables du recensement doivent privilégier l'approche la plus fiable et rentable parmi les options disponibles.

5.29. Si la base de données géographique de recensement est cohérente et bien organisée, la production de cartes de districts de recensement doit être rapide et il n'est en général pas nécessaire de sous-traiter ce travail. L'impression de cartes de districts de recensement n'exige pas un logiciel SIG de haut de gamme, mais peut être effectuée à l'aide de logiciels de cartographie sur ordinateur relativement peu coûteux. Une partie du travail peut être automatisée en utilisant le macro-langage intégré du logiciel. Par exemple, une liste de districts de recensement peut être accompagnée des coordonnées de limites du district de recensement (l'aire de la carte) dans les unités cartographiques. On peut alors charger le logiciel de parcourir cette liste, d'inclure le contenu des couches de cartes dans un gabarit préparé à l'avance comprenant la légende et d'autres informations marginales et d'imprimer un nombre déterminé d'exemplaires.

C . Utilisation d'une infrastructure géospatiale pendant le recensement

5.30. C'est avant et après le dénombrement proprement dit que les bases de données géographiques contribuent le plus au succès du recensement. L'infrastructure géospatiale intervient toutefois aussi pendant le recensement pour appuyer la planification logistique et suivre l'avancement du recensement. En même temps, le travail de dénombrement donne au bureau de recensement la possibilité d'effectuer une nouvelle série d'opérations de contrôle de la qualité de la base de données numérique du recensement. Ces deux aspects sont examinés ci-après.

1. Utilisation de cartes numériques pour la logistique du recensement

5.31. L'utilisation de cartes est nécessaire à de nombreuses fins pendant le recensement. Notamment, les technologies géospatiales peuvent aussi jouer un rôle actif dans le travail préparatoire de planification et la logistique du dénombrement. L'affectation d'unités administratives aux zones opérationnelles, l'implantation des bureaux extérieurs et la planification des déplacements des travailleurs sur le terrain et des agents de recensement sont quelques-unes des tâches où la base de données géographique peut être utile. Si des cartes numériques doivent être utilisées à ces fins et si les organismes nationaux de statistique ont choisi de ne pas développer une base de données géographique numérique, l'unité cartographique du bureau de recensement peut utiliser une base de données géospatiale à faible résolution, composée de cartes numériques à petite échelle concernant les habitats, les routes, les cours d'eau ainsi que les circonscriptions administratives. Les échelles de ces cartes varieront de 1/50 000 à 1/500 000; 1/1 000 000 sera peut-être trop schématique pour être utile. La plupart du temps, il est possible de se procurer ces ensembles de données auprès de sources existantes, dont l'agence de cartographie nationale ou le département des levés. L'organisme national de statistique peut par ailleurs développer ses propres ensembles de données ou utiliser des données « de série » fournies avec le logiciel de SIG.

5.32. De nombreux logiciels de SIG comportent des outils d'analyse de réseau permettant au personnel chargé de la préparation de déterminer les distances et le

coût des déplacements sur un réseau routier. La qualité des données routières varie fortement en fonction de l'emplacement et il convient d'être prudent avec les données à petite échelle (à savoir très schématiques). Dans les zones urbanisées, les déplacements sur les sites de recensement ne posent pas grand problème. Mais dans les zones rurales, les grandes distances et les obstacles naturels qui rendent difficiles les déplacements augmentent le coût des activités sur le terrain. C'est aussi un facteur déterminant pour l'implantation des bureaux extérieurs qui s'occupent d'un certain nombre de zones ou secteurs de supervision. L'emplacement des bureaux extérieurs doit être choisi de façon à réduire au minimum les temps de déplacement et alléger ainsi les fonctions de supervision des administrateurs régionaux de recensement. Les caractéristiques d'agrégation de zones d'un SIG peuvent être utilisées pour déterminer et indiquer des affectations régionales éventuelles.

5.33. Le recours aux bases de données géographiques à des fins logistiques n'est pas aussi essentiel que leur utilisation pour effectuer une délimitation des districts de recensement. De nombreuses tâches peuvent être exécutées tout aussi bien en étudiant les cartes publiées bien que cela réduise l'utilité de la base de données géographique. Pour répondre à ces objectifs, une base de données géographique présente l'avantage de donner des évaluations plus exactes de la distance et du temps de déplacement et de permettre au personnel de recensement d'établir rapidement des cartes montrant divers aspects du processus de préparation du recensement. De plus, la mise en place d'une petite base de données géographique à faible résolution pour le pays est un exercice utile en vue de la tâche difficile que représente l'établissement d'une base de données de recensement géoréférencée et détaillée.

2. Surveillance de la progression des opérations de recensement

5.34. Pendant le recensement et les activités qui suivent immédiatement le dénombrement, le personnel du siège surveillera l'avancement du dénombrement et de la collecte des données. En général, les bureaux de recensement régionaux réunissent les informations concernant l'achèvement des activités de dénombrement et les premiers résultats. Le siège collecte ces informations et détermine les points où les opérations progressent normalement et ceux où des problèmes pourraient se poser.

5.35. Certains pays appliquent une stratégie de comptage rapide dans laquelle les nombres totaux de population sont rapidement compilés et comparés aux évaluations précédentes. Les zones dans lesquelles les nombres présentés sont exceptionnellement élevés ou faibles devront éventuellement être examinées en priorité. Habituellement, ces évaluations sont compilées sous formes de tableaux. Cependant, s'il existe une base de données cartographiques de recensement détaillée et numérique, ces informations pourront être affichées géographiquement. Cela facilitera la détection des zones où se posent des problèmes.

5.36. Dans la pratique, toute statistique sommaire peut être compilée dans un système de base de données relationnelle standard. Par exemple, on peut avoir un indicateur qui montre si le recensement a été ou non achevé dans la zone déclarante ou le pourcentage de districts de recensement dans chaque secteur où le travail est achevé. Le personnel du bureau de recensement peut alors lier normalement ces informations à la base de données géographique et établir un imprimé cartographique en vue de son évaluation par les responsables assurant la supervision d'ensemble du recensement.

5.37. Le facteur essentiel dans cette opération rapide de contrôle de la qualité est la transmission rapide des informations des superviseurs aux bureaux régionaux

et ensuite au siège. Le moyen le plus rapide pour échanger ces informations est Internet. Si les superviseurs locaux et régionaux ont accès à Internet, les informations peuvent même être transmises par une interface de base de données protégée par mot de passe.

3. Directives pour l'utilisation de la carte par les agents recenseurs pendant le recensement

5.38. Le programme de formation des agents de recensement doit inculquer les notions indispensables pour la lecture des cartes et les techniques de navigation. Sur le terrain, il leur faudra localiser leur district de recensement à la fois au sol et sur la carte. Ils devront ensuite veiller à utiliser leur carte, même si leur position leur semble évidente. Pour orienter la carte, il s'agira de poser la carte à plat de manière à en voir toutes les parties. Le recenseur doit commencer à s'orienter à une intersection de routes majeure, en repérant l'emplacement à la fois sur la carte et sur le terrain, en identifiant les détails — maisons, lieux de culte, voies ferrées, rivières — et en les comparant aux symboles sur la carte.

5.39. Aidés des superviseurs, les recenseurs doivent planifier le trajet autour du district de recensement afin d'éviter autant que faire de devoir revenir sur leurs pas. Un district de recensement peut être divisé en sections, îlots ou secteurs. Il est recommandé que le recenseur indique un X au point de départ de l'opération de prospection du jour et inscrive la date sur la carte, à l'endroit où il s'arrête en fin de journée. Les recenseurs doivent « tenir la droite », en procédant dans le sens des aiguilles d'une montre à partir du point de départ de manière à ce que les maisons recensées soient à la droite du recenseur. Il faudra par ailleurs qu'il oriente la carte de sorte qu'elle pointe dans la direction du déplacement. La carte devra être réorientée à chaque tournant de manière à aligner les détails sur ceux au sol. Les agents de recensement identifieront les habitations sur la carte en perspective d'un retour éventuel. Les obstacles sur le terrain tels que le relief du terrain, l'eau et les routes en mauvais état, seront notés sur la carte pour que ces informations puissent servir aux futurs recenseurs.

4. Mise à jour et correction des cartes de districts de recensement au cours du recensement

5.40. Même s'il a été procédé à un contrôle approfondi de la qualité pendant la préparation des cartes d'agents de recensement, il est probable que beaucoup de cartes contiendront des erreurs, parfois des erreurs significatives. Par exemple, pendant le travail initial sur le terrain, des bâtiments ou des rues peuvent avoir été omis ou enregistrés de façon incorrecte sur les cartes. Le travail cartographique sur le terrain devant être effectué plusieurs mois ou même plusieurs années avant le recensement, des constructions ou des infrastructures nouvelles n'auront pas été prises en considération sur les cartes des agents de recensement.

5.41. Le bureau de recensement doit non seulement veiller à former ses agents à la collecte des données et leur enseigner les notions indispensables pour la lecture des cartes, il doit aussi leur apprendre à annoter les cartes de district de recensement pendant le dénombrement afin de relever toute erreur ou omission éventuelle. Le personnel cartographique du service de recensement devra collecter les cartes de district de recensement après le recensement et donner suite à toute proposition de révision par les agents du recensement, en respectant les procédures établies pour l'incorporation des corrections dans la base de données principale (cet aspect est

traité plus exhaustivement dans le chapitre VI). Cela n'exigera que d'apporter les corrections correspondantes à la base de données numérique du recensement ou de procéder à d'autres vérifications sur le terrain, voire d'accéder à des données de photographies aériennes ou satellitaires pour vérifier les changements. Cette opération est peut-être très complexe, mais elle garantira que le bureau de recensement dispose des informations les plus récentes sur les districts de recensement, ce qui réduira la charge de travail en matière d'activités cartographiques avant les prochains recensements ou enquêtes. Cela ne doit toutefois pas retarder la publication des résultats du recensement.

Encadré V.1

Production de cartes de recensement en Inde

1. En Inde, le recensement est une opération gigantesque mise sur pied pour dénombrer et recueillir des informations sur plus d'un milliard de personnes dans le pays. La cartographie est une solution incontournable pour garantir une couverture complète de la zone géographique du pays, sans omission ni chevauchement.
2. La préparation de cartes renseignant les limites administratives des États, districts, sous-districts, villages et villes commence près de trois ans à l'avance. Les notifications précisant les changements de juridiction sont recueillies auprès des gouvernements provinciaux et des autorités communales. Les cartes préparées dans le cadre du dernier recensement sont actualisées, en indiquant les modifications rapportées; les autorités compétentes les certifient avant qu'elles ne soient utilisées dans le dénombrement. Tout autre changement au niveau des juridictions est appliqué jusqu'au gel des limites administratives pour le recensement. L'organisme de recensement gère une base de données de fichiers des limites qui sera utilisée pour l'actualisation, à l'aide de SIG.
3. Lorsque vient le moment de l'inventaire des habitations, six à huit mois avant le dénombrement démographique, des quartiers sont définis en attente d'une attribution à quelque deux millions d'agents recenseurs. La base de données géographique, agencant les bâtiments, maisons, etc., de manière détaillée, n'existant pas au format numérique, des dessins réalisés à la main, reprenant les bâtiments, maisons, routes et principaux repères, provenant du recensement précédent, sont utilisés pour les zones rurales. Dans les zones urbaines, les quartiers d'habitations sont délimités à l'aide des cartes les plus récentes disponibles auprès des autorités locales. Grâce aux informations recueillies au cours de l'inventaire, de nouveaux quartiers de recensement sont définis, couvrant des zones comptant en moyenne 500 à 750 personnes en vue du recensement. Les dessins à la main des districts de recensement, appelés « cartes conceptuelles », sont utilisés par les agents recenseurs au cours du recensement démographique et sont également transmis à d'autres agences pour des levés sur le terrain.
4. Si l'on peut se procurer des cartes numériques précisant les circonscriptions administratives jusqu'au niveau des villes et villages, il n'est pas encore possible d'obtenir des cartes numériques détaillées indiquant l'agencement de bâtiments, maisons, voies, voies secondaires, réseaux routiers et principaux repères dans ceux-ci. Une initiative vise actuellement à préparer une base de données géographique numérique détaillée des principales villes; des images satellitaires seront utilisées à cette fin. Par ailleurs, la Survey of India, la première agence cartographique publique du pays, fournira des cartes exhaustives des circonscriptions électorales au format numérique, compilées à partir des images satellitaires. Des enquêtes spéciales sur le terrain sur les numéros de maison, les types de bâtiment, l'utilisation prévue, la population, etc., sont proposées afin de faire le lien avec les cartes numériques et, par la suite, pour délimiter des quartiers de recensement. Cet exercice aura pour objectif de préparer des cartes géoréférencées destinées au recensement. Pour mener à bien ce dernier, les agents de recensement se verront remettre des cartes renseignant l'emplacement des quartiers de recensement ainsi que l'agencement détaillé des bâtiments, voies, voies secondaires, etc. Elles seront également remises à d'autres agences publiques pour les aider à mettre en œuvre leurs programmes respectifs (pour un complément d'information, contacter Chinmoy Chakravorty à l'adresse suivante : cchakravorty.rgi@censusindia.gov.in).

D. Résumé et conclusions

5.42. Le chapitre V a traité du processus de création des cartes nécessaires au recensement, en poursuivant l'analyse du point de vue opérationnel là où la discussion sur les bases de données géographiques s'est arrêtée. Certaines rubriques se sont par ailleurs penchées sur la marche à suivre pour les cartes des recenseurs, sur la compilation, la définition des couches utiles aux recenseurs et superviseurs ainsi que sur les principes fondamentaux de l'impression et de la distribution; le recensement a en outre été abordé sous l'angle de la gestion de projet, en proposant par ailleurs une solution pour faire face aux imprévus en cas de retards.

Chapitre VI

Bases de données géographiques pour la diffusion des résultats du recensement, produits et services

A. Introduction

6.1. Au chapitre V, l'infrastructure géospatiale a été présentée comme une solution pour faciliter le recensement. Dans le présent chapitre, il sera question des tâches géographiques que l'organisme national de statistique effectuera après le dénombrement ainsi que de la diffusion et de l'utilisation de données censitaires géoréférencées. À ce stade du processus, les résultats du recensement devraient être disponibles. Tout au long du présent *Manuel*, nous avons souligné que l'ensemble des plans géographiques devaient se conformer au plan de recensement global. Il est un autre facteur déterminant qui incitera un organisme national de statistique à investir dans une base de données géographique : la possibilité de l'utiliser à chacun des stades du processus de recensement. Une fois ce dernier achevé, les résultats permettront de peaufiner davantage la base de données. Utiliser la base de données géographique pour créer des produits d'information et de sensibilisation du public soulèvera des questions d'agrégation et d'échelle qui exigeront l'intervention de personnes extérieures aux organismes nationaux de statistique.

6.2. Si une base de données numérique géographique complète existe pour le recensement, des bases de données statistiques pourront être rapidement établies pour les unités administratives ou statistiques, par agrégation. Les pays qui n'utiliseront pas encore les techniques numériques pour établir les cartes de districts de recensement pendant la série des recensements de 2010 auront encore le loisir de décider de mettre au point une base de données numérique de recensement géoréférencées afin d'établir des cartes de qualité appropriée pour la publication destinées à accompagner les rapports de recensement et à être distribuées à des utilisateurs extérieurs souhaitant effectuer l'analyse spatiale des données de recensement ou pour des applications internes. Cette base de données pourra être compilée pour un niveau approprié de la hiérarchie administrative ou d'autres grandes régions statistiques. À ce niveau d'agrégation, les ressources requises pour établir une base de données numérique sont bien moindres que celles nécessaires pour avoir une base de données numérique complète de cartes de district de recensement.

6.3. Pour l'essentiel, cependant, le présent chapitre part de l'hypothèse qu'une base de données numérique complète de districts de recensement ou d'unités d'habitation a été créée aux fins du dénombrement censitaire. Pour justifier l'investissement nécessaire à l'élaboration de cette base de données, l'organisme national de statistique doit adopter une perspective à long terme. Les tâches immédiates à entreprendre après le recensement ne sont donc que les premières phases de l'élaboration des documents cartographiques pour le prochain recensement.

6.4. Les sections principales du présent chapitre sont consacrées aux tâches de gestion concernant les bases de données géographiques après et entre les recensements ainsi que l'élaboration et la diffusion des produits cartographiques. D'autres thèmes seront abordés et notamment l'intégration des changements relevés sur le terrain; le processus d'agrégation des données vers des unités de diffusion; la gestion des bases de données; la diffusion de produits et services; des considérations sur la divulgation et la confidentialité des données; la commercialisation; l'élargissement du cercle des utilisateurs; les produits géographiques; les cartes de recensement et les bases de données destinées à la publication; certaines méthodes d'analyse des données de recensement sur un plan spatial ainsi que des questions relatives à la cartographie en ligne et à la distribution de bases de données de recensement.

B. Tâches à accomplir après le recensement et pendant la période intercensitaire

6.5. C'est après le recensement qu'une bonne planification s'avérera payante. Si l'organisme national de statistique s'est conformé à un plan global défini en début de processus, le choix des produits finals aura alors orienté les options envisageables au cours du recensement. Si le plan demande une publication des résultats au niveau des districts de recensement, aucun rapprochement ne sera sans doute nécessaire. Si les résultats ne doivent pas être publiés à ce niveau, un rapprochement s'imposera toutefois dans une certaine mesure. Celui-ci devra être approfondi si des changements géographiques sont constatés en cours de recensement. L'organisme national de statistique doit convenir à l'avance des types d'erreurs qui exigent une attention immédiate et de celles qui peuvent attendre car susceptibles de ralentir la publication des résultats.

Tâches immédiates

1. Inclusion de mises à jour et de changements suggérés par les agents de recensement

6.6. Le processus de recensement, prévoit notamment une formation des agents de recensement à la détection d'erreurs ou d'incohérences lors de la délimitation des districts de recensement ou dans les détails de fond de carte du territoire étudié. Le bureau de cartographie censitaire doit inciter les agents de recensement à enregistrer ces erreurs; les superviseurs locaux pourront alors les collecter après le recensement et les envoyer au service cartographique de recensement. Ce dernier pourra alors, selon une procédure imaginée par l'organisme national de statistique, corriger la base de données cartographique ayant servi à établir la carte de district de recensement fondée sur ces informations. Cette opération présente un double avantage.

6.7. En premier lieu, elle garantit que les tableaux et les produits cartographiques numériques et imprimés sont bel et bien fondés sur la délimitation du district de recensement ayant effectivement servi pour le recensement. En second lieu, l'enregistrement de la modification des limites du district de recensement dans la base permanente de données géographiques facilitera à l'avenir les activités de recensement ou les autres activités de collecte de données statistiques fondées sur les mêmes unités de collecte de données géographiques ou des unités semblables. Une telle approche a un avantage évident : la base de données géographique restera aussi actualisée que possible et les participants sur le terrain seront incités à contribuer plus activement à l'amélioration de la qualité globale des données. Cependant, le rapprochement ne peut

être effectué sans avoir évalué dans une certaine mesure les répercussions sur l'intégrité du plan de recensement.

2. Regroupement des unités de collecte et des unités statistiques ou de dépouillement

6.8. La principale responsabilité dans la période post-censitaire consiste à appuyer l'élaboration des tableaux de données statistiques obtenues à partir des résultats du recensement. On doit disposer de données censitaires pour de nombreux types différents de zones agrégées, car les utilisateurs de données censitaires appartenant à différents secteurs utilisent en général comme base de planification et pour leurs opérations des zones géographiques différentes. Il faut donc agréger les districts de recensement au niveau de ces diverses unités déclarantes en fonction des besoins d'élaboration de produits cartographiques de recensement très variés. Il est peu probable que tous les utilisateurs aient besoin du niveau de détail — sans compter la gestion complexe des fichiers — d'une base de données du niveau des districts de recensement couvrant le pays.

6.9. La mise en concordance des unités de collecte de données et de tabulation exige que l'on établisse des *fichiers d'équivalence* ou de *comparabilité*. Un tableau rapprochant ces deux géographies différentes est parfois également appelé une « concordance ». Pour chaque unité de tabulation, les fichiers d'équivalence répertorient les districts de recensement correspondants faisant partie de cette unité. Une fois ces listes établies, l'agrégation peut commencer en procédant aux opérations habituelles sur la base de données. Le tableau VI.1 illustre les changements dont les districts de recensement ou d'autres unités font l'objet d'un recensement à l'autre; il montre qu'un district de recensement (secteur de recensement) de 2000 a été divisé en trois nouveaux districts de recensement (secteurs de recensement) pour 2010, avec une révision du codage.

6.10. L'établissement de fichiers d'équivalence est facilité si l'on a mis en œuvre un système cohérent de codage tout au long du processus. Le géocodage des unités et attributs de recensement est traité en profondeur dans le présent *Manuel*, en particulier au chapitre III. Cela confirme qu'il importe d'élaborer des systèmes de référence souples et intuitifs pour affecter des codes numériques ou alphanumériques à chaque district de recensement unique aux premiers stades d'un projet cartographique censitaire.

Tableau VI.1

Comparaison entre d'anciennes et de nouvelles unités de districts de recensement

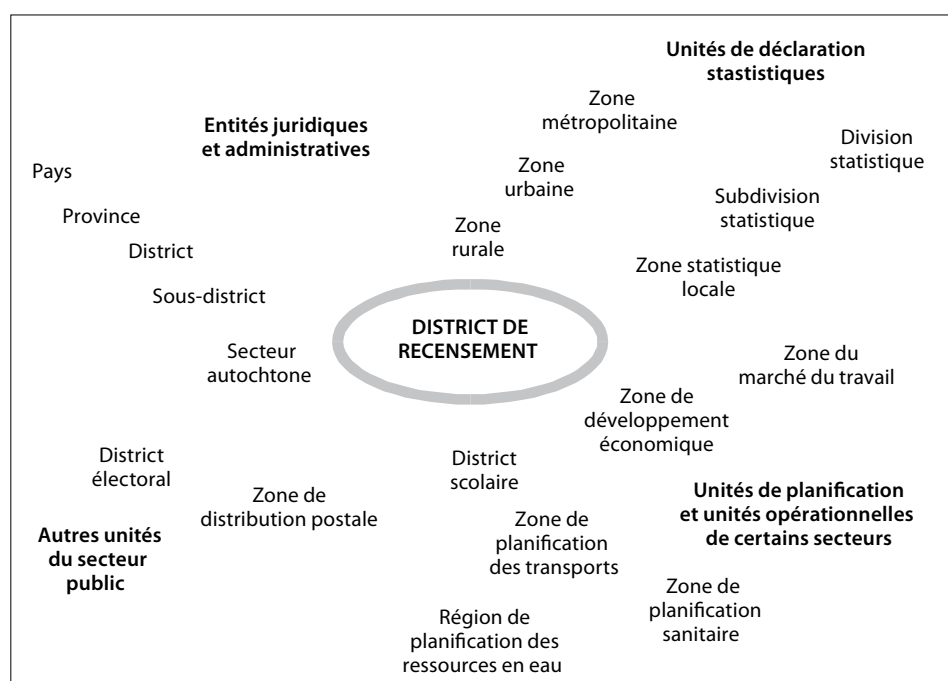
Longueur– Description
1-11 Code de secteur de recensement.
12 Indicateur de partie (P).
13-18 Kilométrage côté rues du secteur de recensement de 1990.
19-22 Pourcentage du kilométrage côté rues du secteur de recensement de 1990.
23-31 Code de secteur de recensement de 2000.
32-33 Suffixe de secteur de recensement de 2000.
34 Indicateur de partie (P) de secteur de recensement de 2000.
35-40 Kilométrage côté rues du secteur de recensement de 2000.
41-44 Pourcentage du kilométrage côté rues du secteur de recensement de 2000.

Longueur– Description					
45-50 Kilométrage côté rues de la zone couverte par l'enregistrement.					
51-64 14 Superficie terrestre de l'enregistrement (1 000 mètres carrés).					
65-66 2 Abréviation du nom d'État en 2000.					
67-126 60 Nom de comté en 2000.					
L'enregistrement ci-dessous montre que le secteur de recensement 402 de 1990 est divisé en trois secteurs de recensement en 2000 : 402.01, 402.02, et 402.03					
10001040200P	11735029610001040201	34771000	3477	9796DEKent	
10001040200P	11735044210001040202	51891000	5189	32059DEKent	
10001040200P	11735026210001040203	30691000	3069	59822DEKent	

6.11. Le nombre des unités de sortie pour lesquelles des fichiers d'équivalence doivent être établis peut être très élevé. On peut avoir à grouper des données de recensement, non seulement pour des unités juridiques et administratives telles que des districts ou des provinces, mais aussi pour toute une série d'unités de planification ou d'unités opérationnelles. On peut citer des unités sanitaires, des districts scolaires, des régions de planification des transports, des districts électoraux, des zones de service d'utilité collective, des zones postales et des unités de planification environnementale (voir figure VI.1). Celles-ci peuvent dans certains cas coïncider avec des circonscriptions administratives, mais souvent elles ne seront pas compatibles avec les unités déclarantes habituelles. En outre, les secteurs privé et universitaire peuvent avoir des exigences particulières en matière de tabulation. L'élaboration d'une procédure cohérente pour établir et tenir à jour les fichiers d'équivalence est donc une tâche importante pour le service cartographique de recensement. Comme indiqué tout au long du présent *Manuel*, une planification efficace procurera une kyrielle d'avantages, dont l'organisme national de statistique aura conscience ou non.

Figure VI.1

Exemples d'unités de tabulation censitaire et d'unités déclarantes



6.12. Des fichiers supplémentaires de comparabilité devront être élaborés pour harmoniser les districts de recensement ou de déclaration statistique anciens et les districts actuels. Étant donné que les unités de collecte de données et de tabulation sont en général modifiées régulièrement, il est difficile pour les utilisateurs de données censitaires de constater les changements de variables censitaires dans le temps. L'unité géographique du bureau de recensement doit donc garder la trace de ces modifications de la géographie censitaire du pays et fournir aux utilisateurs de données des fichiers de comparabilité permettant d'harmoniser les données de recensement précédentes et actuelles.

6.13. Un tel effort s'avérera particulièrement utile lorsque l'organisme national de statistique participera à une activité d'infrastructure de données spatiales au sein du pays; en effet, les personnes ayant besoin de données censitaires disposeront ainsi des spécifications exactes requises pour localiser les fichiers de recensement corrects de manière à satisfaire leurs besoins particuliers.

3. Archivage de la base de données

6.14. Après avoir pris en considération les erreurs et les incohérences dans la base de données géographique permanente, des copies de référence de tous les lots de données géographiques doivent être établies et archivées. Cette base de données, pour laquelle les données géographiques de recensement ont été gelées afin de refléter la situation au moment du recensement, servira de base pour tous les produits cartographiques, y compris les cartes de référence, les cartes thématiques de résultats de recensement et les extractions numériques de cette base permanente destinées à être distribuées. Tous les résultats de recensement présentés sous forme de tableau après le dénombrement se référeront aux unités de référence de cette base de données. Cela suppose aussi que tous les documents et les métadonnées soient soigneusement vérifiés afin que le bureau de recensement puisse répondre à toute question relative aux données qui pourrait se poser à l'avenir. Des exemplaires de cette base de données de référence devront être archivés en lieu sûr dès la fin de son élaboration.

6.15. Pour les organismes nationaux de statistique ayant un programme cartographique permanent, un exemplaire de cette base de données servira de base à une mise à jour régulière pendant les activités intercensitaires. Les avantages d'un programme cartographique permanent sont examinés ci-dessous.

4. Tenue à jour de la base de données : avantages d'un programme cartographique permanent

6.16. Comme indiqué précédemment dans le présent *Manuel*, les avantages d'un programme géographique numérique de recensement ne contrebalanceront les coûts que si la base de données de recensement obtenu est utilisée pour de nombreuses applications allant au-delà des tâches essentielles du recensement. On ne pourra mettre à profit les divers avantages de ce programme que si la base de données est gérée de telle sorte que les mises à jour pour de futures applications censitaires n'exigent que relativement peu de ressources. Mettre en œuvre la base de données cartographiques de recensement pour le plus grand nombre possible d'utilisations pour le recensement et veiller à utiliser au maximum les données numériques existantes dans les dénombrements ultérieurs n'est possible que s'il existe un degré de continuité élevé dans le programme national de cartographie pour le recensement. La continuité des activi-

tés géographiques de recensement assurera ainsi la préservation de l'investissement consacré à l'élaboration de la base de données.

6.17. Il est essentiel de ne pas perdre de vue qu'il est recommandé que les services cartographiques de recensement entreprennent les opérations de maintenance de la base de données immédiatement après un recensement. À cette fin, on doit mettre à jour en permanence les limites et d'autres objets au fur et à mesure que de nouvelles informations deviennent disponibles. Pendant la période intercensitaire, un système précis de contrôle de version devrait être mis en œuvre. Par ailleurs, il est préférable de documenter et de publier les changements apportés dans la base de données. Par exemple, un agent ou un petit groupe de collaborateurs devront être chargés d'enregistrer les modifications dans la base de données permanente. En accordant ces droits administratifs spéciaux uniquement à certains membres du personnel, on évitera que différents collaborateurs apportent à différentes versions de la base de données des modifications qui devront être harmonisées par la suite.

6.18. Pendant la période intercensitaire, le service cartographique de recensement devra suivre l'évolution de l'industrie ainsi que les nouvelles approches adoptées par d'autres services cartographiques de recensement nationaux. Ainsi le service sera tenu au courant des décisions relatives aux investissements consacrés à des améliorations du logiciel et du matériel. Compte tenu de l'évolution rapide des techniques, des investissements périodiques dans ces secteurs pourront être requis pour assurer une qualité élevée des opérations de recensement pendant la période intercensitaire.

6.19. L'élaboration de données cartographiques numériques exige des connaissances spécialisées concernant l'utilisation de l'ordinateur, les concepts géographiques et les logiciels spécialisés. Il s'avère coûteux d'assurer la formation du personnel dès que l'on va au-delà des tâches et des concepts géographiques essentiels. Pour assurer le succès d'un programme à long terme de cartographie de recensement, la stabilité du personnel est par conséquent un facteur essentiel. Le bureau de recensement doit désigner un personnel de base qui assurera la tenue à jour de la base de données pendant la période intercensitaire, fournira des services géospatiaux pour d'autres applications statistiques telles que des enquêtes par sondage et gèrera la mémoire institutionnelle. Cela facilitera le bon déroulement des applications SIG de recensement lors du prochain dénombrement. Le personnel de base peut par exemple assurer la formation du personnel temporaire recruté pour veiller à la numérisation ou travailler sur le terrain. Le maintien en fonctions du personnel de base réduira aussi les coûts de démarrage que l'on devrait autrement engager pour recruter des experts géographiques, ces derniers ayant besoin de quelque temps pour s'intégrer pleinement au processus cartographique de recensement (pour une étude détaillée des questions liées aux ressources humaines, voir chapitre II).

6.20. On ne saurait à nouveau trop insister sur l'importance d'une conception à long terme des activités cartographiques de recensement. Les avantages d'une stratégie à long terme valent certainement les ressources supplémentaires requises pour maintenir une capacité cartographique aux organismes nationaux de statistique pendant la période intercensitaire.

C. Diffusion des produits géographiques de recensement

6.21. Les organismes nationaux de statistique négligent souvent la diffusion; celle-ci représente pourtant la solution la plus importante pour sensibiliser le public et appuyer d'autres activités. Il convient de souligner une fois encore que l'organisme

national de statistique doit déterminer l'éventail de produits et services tôt dans le processus, en ce compris la structure et le format. La planification de la diffusion des informations, les questions de divulgation et de respect de la confidentialité des données, la commercialisation, l'élargissement du cercle des utilisateurs et l'éducation sont passés en revue ci-après.

1. Planification de la diffusion des données

6.22. La définition des produits et services cartographiques et l'échelonnement de leur mise en circulation doivent être étroitement coordonnés avec le calendrier d'ensemble du projet de recensement. Pour établir les tableaux de données de recensement, on peut avoir besoin d'informations du service géographique du recensement; en outre, les cartes thématiques et les bases de données géographiques numériques ne peuvent être établies que si le traitement des données de recensement est terminé.

6.23. La sélection de produits cartographiques appropriés devrait être guidée par une évaluation détaillée des besoins des utilisateurs — autrement dit une étude de marché — qui devra être effectuée dès le début de la planification du recensement. Il est primordial que ces plans relatifs à la diffusion des produits soient établis très tôt et largement publiés pour que la communauté des utilisateurs puisse les commenter.

6.24. Une initiative utile consiste à créer un groupe consultatif de représentants des principales communautés d'utilisateurs de données censitaires, susceptible de fournir des orientations aux organismes de recensement. Le groupe ne doit pas nécessairement limiter son action à la période de préparation du recensement, il peut aussi devenir un mécanisme officiel ou informel permanent assurant des échanges d'idées entre le bureau de recensement et les utilisateurs. Les exemples d'utilisation de statistiques de recensement de petites zones fournies dans l'introduction au présent *Manuel* montrent quels peuvent être les utilisateurs de données très divers que le bureau de recensement devra prendre en compte dans son évaluation des besoins des utilisateurs.

6.25. La connaissance des applications ayant précédemment suscité l'intérêt des utilisateurs de données de recensement n'offre qu'une aide limitée pour définir les produits et services à obtenir. Les exigences varient notamment en fonction de l'évolution des capacités techniques des utilisateurs de données. Les produits et services que l'on peut tirer d'une base de données cartographique numérique n'ont été que rarement disponibles après la dernière série de recensements, alors qu'ils seront l'un des principaux résultats de la présente série. Dans de nombreux pays, la demande de cartes imprimées peut dépasser celles des informations numériques, mais cette situation change étant donné que les produits cartographiques sont de plus en plus distribués électroniquement. Le service cartographique du recensement doit faire preuve de souplesse pour s'adapter à l'évolution des besoins des clients et à des demandes particulières.

6.26. Il est conseillé d'établir des prévisions plusieurs années à l'avance lorsqu'on planifie la stratégie applicable en matière de résultats. Par exemple, Internet est devenu un mode de diffusion de données incontournable, même dans les pays pauvres, avec l'amélioration des infrastructures de communications dans le monde entier. De nouvelles communautés d'utilisateurs voient le jour au fur et à mesure que de nouveaux types de données sont créés. Ainsi, les entreprises de téléphonie mobile ont besoin des données démographiques géoréférencées pour planifier l'installation de nouvelles tours et d'autres infrastructures. Pour accroître les avantages sociaux de la collecte de données censitaires, le bureau de recensement peut rechercher activement de nouveaux groupes de clients et leur présenter ses produits. À cette fin, la participa-

tion à une infrastructure nationale de données spatiales sera efficace pour ce qui est des utilisateurs de données gouvernementaux et non gouvernementaux.

6.27. Le bureau de recensement devra aussi tenter d'évaluer le volume de la demande éventuelle pour ses produits et ses services, ce qui lui permettra aussi dans une certaine mesure d'évaluer la capacité requise pour répondre aux demandes des utilisateurs. Cette évaluation sera elle aussi difficile, car la demande pourrait s'accroître avec le lancement de nouveaux produits et services, lorsque de nouveaux utilisateurs auront compris en quoi ils pourraient répondre à leurs propres besoins. Ainsi, le bureau de recensement devra être prêt à répondre à une demande croissante lorsque les produits et services seront disponibles. Des plans efficaces peuvent être adaptés en fonction de la demande. Il est souhaitable de définir clairement et suffisamment tôt les besoins des utilisateurs de données censitaires que l'on **devra** satisfaire (qu'il est légalement demandé de satisfaire), ceux que l'on **devrait** satisfaire (avec des notions de service à la clientèle acceptable) et ceux que l'on **ne** prendra **pas** en compte (en raison de ressources limitées). La définition précise d'un ensemble de priorités facilitera aussi l'élaboration d'un calendrier concernant la diffusion des produits et services de recensement.

6.28. L'organisme national de statistique est vivement encouragé à diffuser le plus largement possible les résultats, les produits et les services en s'appuyant sur une politique ouverte de diffusion des données. Une telle politique — garantissant un accès gratuit ou à faible coût aux données — pourra aider à réduire la charge de travail du bureau de recensement. Dans les pays où les données censitaires sont disponibles gratuitement, les fournisseurs de services du secteur privé seront peut-être en mesure de satisfaire les requêtes spécifiques de certains utilisateurs de données de recensement. Cela permettra au bureau de recensement de mieux concentrer ses efforts sur les utilisateurs de données qu'il a pour mission de servir en vertu de la législation.

6.29. Certaines données géographiques de recensement seront requises en vue de leur utilisation interne et officielle. Il peut s'agir des fichiers d'équivalence ou des cartothèques de référence, ainsi que de produits spécialisés tels que des cartes de districts électoraux. Dans certains pays, la loi pourra exiger que le bureau de recensement fournisse certains produits cartographiques. Il lui faudra peut-être préparer ces produits à intervalles réguliers ou en réponse à une demande spéciale, émanant par exemple des ministères ou du parlement.

6.30. D'autres produits plus courants seront élaborés en vue d'être largement diffusés auprès des utilisateurs des secteurs public et privé et du grand public. Le bureau de recensement devrait tenter d'utiliser le plus grand nombre possible de canaux de distribution et de satisfaire de multiples besoins des utilisateurs en publiant des analyses ainsi que les données et résultats. Plusieurs exemples de ces analyses spéciales sont proposés ci-après.

6.31. Les sections suivantes porteront sur les produits de recensement et les possibilités de diffusion, y compris les produits requis, les cartes thématiques qui peuvent être distribuées sur papier ou sous forme numérique, la diffusion de bases de données cartographiques numériques, les atlas numériques de recensement et les stratégies de diffusion d'informations par Internet. Une solide formation aux techniques cartographiques pour les cartes thématiques est requise pour un grand nombre de ces produits. Dans les paragraphes ci-dessous, on examinera seulement les questions plus générales concernant la cartographie thématique. L'annexe IV donne un aperçu plus complet de la conception des cartes thématiques.

2. Considérations relatives à la confidentialité et à la divulgation : problème du calcul de différences

6.32. L'organisme national de statistique doit avoir conscience que la préoccupation du public quant à la confidentialité des données personnelles sur le recensement peut avoir des répercussions sur la diffusion des données. Les directives régissant la vie privée des individus figurent dans les principes directeurs standard des Nations Unies (voir par exemple Nations Unies, 2008).

6.33. Divers organismes publics aux utilisateurs extérieurs de données peuvent avoir besoin de données censitaires pour divers ensembles de petites unités géographiques. Par exemple, certains organismes utilisent comme unités déclarantes primaires des zones postales ou des districts sanitaires de petites dimensions. Pour répondre aux besoins de ces utilisateurs de données, le bureau national de recensement peut souhaiter distribuer les renseignements censitaires relatifs à plusieurs ensembles de petites zones géographiques dont les limites sont indépendantes les unes des autres. Si des limites et des tableaux de données sont publiés pour deux groupes de zones ou plus, l'utilisateur pourra effectuer des opérations géospatiales ou une manipulation simple de tableaux de données pour obtenir des statistiques de recensement relatives à de très petites zones géographiques. Les dénombrements concernant ces nouvelles unités peuvent se situer au-dessous du seuil de divulgation de l'organisme. Ce problème est appelé problème de la différenciation en matière de divulgation statistique (voir Duke-Williams et Rees, 1998).

6.34. Le problème ne se pose pas si les limites se chevauchent irrégulièrement, à moins qu'une des zones qui se recouvrent présente des valeurs nulles. Dans la plupart des cas, l'utilisateur ne peut être certain qu'une valeur nulle soit en fait exacte. Cela tient au fait que la plupart des bureaux de recensement utilisent une méthode de perturbation ou codage élargi (indication d'un intervalle de données sous forme de « <10 » plutôt que le petit nombre indiquant la valeur exacte) pour éviter que les utilisateurs ne puissent obtenir des caractéristiques exactes concernant de petits groupes de personnes dans des zones à faible population.

6.35. Le problème du calcul de différence peut se poser cependant si une zone d'un ensemble de secteurs géographiques est emboîtée dans une zone d'un autre ensemble et si l'utilisateur a des tableaux de données concernant les deux ensembles de secteurs. Par exemple, le secteur postal B sur la figure VI.2 est emboîté dans le district de recensement A. En superposant les deux ensembles de limites, nous pouvons déterminer le secteur géographique qui se trouve dans A et non dans B. En utilisant les tableaux de données, nous pouvons obtenir des données de recensement concernant les personnes qui se trouvent dans ce petit secteur en soustrayant simplement le dénombrement relatif au secteur postal B de celui du district de recensement A. Les nombres obtenus peuvent être inférieurs aux seuils de divulgation, même si les dénombrements concernant A et B ne le sont pas.

6.36. Pour éviter les problèmes de divulgation de données, l'organisme national de statistique devra examiner avec soin les limites des divers secteurs géographiques utilisés pour le recensement. Dans les cas où le calcul de différences paraît possible, on devra adopter des mesures complémentaires pour protéger les données. Duke-Williams et Rees (1998) analysent le problème du calcul de différences de façon très détaillée. Sur la base de leurs expérimentations, ils donnent quelques recommandations d'ordre général pour résoudre ce problème :

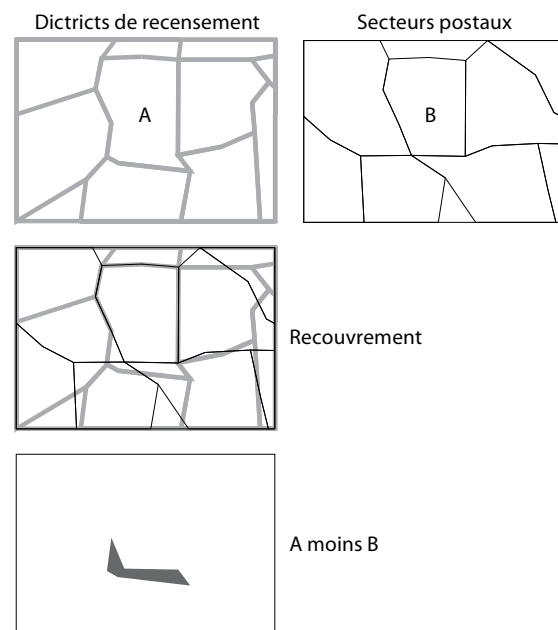
- On utilisera des niveaux de seuil minimaux pour les tableaux. On peut obtenir une protection supplémentaire en introduisant de légères perturbations

de valeurs de données pour de très petites zones ou en utilisant des intervalles plutôt que des valeurs exactes pour les petits dénombrements. Cela réduit les risques de publication de données de recensement concernant plus d'un ensemble de petites unités de zones.

- La circonscription géographique primaire utilisée pour le recensement devrait être en général aussi utile que possible. Par exemple, si la plupart des organismes du pays utilisent de petites circonscriptions administratives en tant qu'unités primaires de référence, les données de recensement devront être publiées pour ces unités.
- Le risque de publication d'autres cartes géographiques dont les zones sont beaucoup plus étendues que celle des unités de recensement primaires est très faible. Même s'il est possible de calculer des différences dans ces cas, il est peu probable que les dénombrements obtenus seront inférieurs au seuil de sécurité.
- Si deux cartes géographiques de résolution à peu près égale sont très semblables — c'est-à-dire si de nombreuses limites sont les mêmes — le risque de calcul de différences sera plus grand que si les limites sont très différentes.
- Les problèmes de différences peuvent être atténués si l'organisme national de statistique recourt à des coordonnées et à un niveau de schématisation des limites connus pour définir les limites administratives ou statistiques.

Figure VI.2

Le problème du calcul de différences en matière de divulgation statistique



3. Commercialisation de produits et de services de recensement géographique

6.37. Compte tenu de la demande importante et croissante en données de recensement de petites zones, les organismes nationaux de statistique devraient réfléchir à des solutions simplifiant l'accès des utilisateurs à leurs produits et services. Des stra-

tégies de renforcement des capacités internes peuvent être envisagées pour satisfaire aux besoins des utilisateurs ou exploiter l'expertise extérieure. Un organisme national de statistique devrait uniquement développer ses capacités internes de diffusion de ses propres données et analyses en recourant à des entités extérieures telles que des entreprises commerciales, lorsque la demande potentielle dépasse ses capacités de diffusion.

6.38. Quant à la question de savoir s'il convient de faire payer ou non l'accès aux données, il appartient aux organismes nationaux de statistique de trouver un équilibre entre la génération de recettes et l'expansion de la base d'utilisateurs de données. Les pays qui entendent récupérer en partie les coûts d'élaboration de bases de données géographiques de recensement et dans lesquels il existe une forte demande commerciale de données statistiques concernant de petites zones pourront souhaiter examiner la possibilité de conclure un contrat de commercialisation avec un vendeur privé de données. Au nombre des collaborateurs potentiels figurent les distributeurs locaux des principaux producteurs de logiciels de SIG, les fournisseurs de systèmes d'imagerie et de cartographie en ligne ainsi que les vendeurs de données géodémographiques. La plupart des vendeurs principaux de SIG produisent et vendent des lots de données relatifs à de nombreux sujets. C'est pour eux d'une part une source de revenu et d'autre part un moyen de faciliter l'utilisation de leurs logiciels grâce à la fourniture de lots de données présentées dans le format de données du logiciel. Ces vendeurs privés collaborent parfois avec les instituts nationaux de cartographie et de statistique pour établir les bases de données de SIG conçues de façon professionnelle.

6.39. Pour l'office national de la statistique, cela présente quelques avantages. Les vendeurs de logiciels et de données peuvent mettre à disposition leur savoir-faire technique et éventuellement leurs ressources informatiques pour élaborer le logiciel de diffusion de la base de données géographique en échange d'une part des recettes procurées par les ventes de base de données. Les vendeurs de logiciels exerçant leur activité sur le plan international peuvent aussi développer la diffusion des données géographiques nationales. Dans d'autres pays la demande peut émaner d'entreprises actives sur le plan international ou d'universitaires effectuant des études sur le pays.

6.40. La collaboration avec un vendeur de logiciel commercial peut parfois donner lieu à des difficultés parce que le vendeur peut souhaiter distribuer des données dans son propre format exclusif. Le bureau de recensement devra s'assurer que les utilisateurs de données qui souhaitent utiliser un autre format pourront accéder également aux données. Les inconvénients de la distribution commerciale ont été précédemment mentionnés. Les organismes nationaux de statistique devraient être prudents lorsqu'ils confèrent les droits de commercialisation et de distribution à une entreprise privée. Ce faisant, l'organisme de statistique ne peut plus distribuer de données gratuitement ou à un très faible coût. Si l'objectif est d'assurer la distribution la plus large possible, l'organisme devra de préférence établir et distribuer lui-même les bases de données.

6.41. Il existe d'autres partenaires potentiels pour distribuer les produits : les universités ou d'autres organismes publics qui diffusent des informations. Dans tous les cas, un contrat précis de commercialisation et de partage des recettes devra être établi pour éviter des problèmes ultérieurs. Le bureau de recensement devra procéder à une évaluation détaillée de la valeur de marché de ses données par rapport aux coûts de production, de publicité et de vente des données afin qu'un contrat équitable et mutuellement avantageux constitue la base d'un partenariat public-privé ou public-public.

4. Élargissement du cercle des utilisateurs et sensibilisation

6.42. Pour faire largement connaître les disponibilités des données et assurer la diffusion la plus large possible des données de recensement à référence spatiale, l'office national de la statistique souhaitera peut-être établir un plan publicitaire. Ce plan pourrait prévoir notamment l'établissement de brochures imprimées et d'affiches comportant des cartes de recensement. Ces brochures pourront être largement distribuées aux écoles, aux universités, aux entreprises commerciales et aux organismes publics nationaux et locaux. Une campagne de marketing et de distribution peut précéder le recensement pour sensibiliser à la nécessité de données statistiques actualisées et garantir la pleine participation du public.

Le bureau de recensement pourra aussi organiser dans tout le pays une série de séminaires à l'intention des utilisateurs régionaux. Dans ces ateliers, le personnel de recensement pourra présenter l'utilisation de logiciels cartographiques gratuits ou plus coûteux à un vaste cercle d'utilisateurs potentiels en vue de l'analyse des données de recensement, en informant par ailleurs le public de la gamme de produits à venir résultant du recensement.

5. Liste de produits potentiels

6.43. Une description des divers produits que l'organisme national de statistique devrait publier une fois le recensement terminé est proposée ci-dessous. Il peut s'agir de fichiers d'équivalence et de comparabilité, de cartothèques de référence, de nomenclatures toponymiques et de fichiers de centroïdes, ainsi que de cartes thématiques.

a) Fichiers d'équivalence et de comparabilité

6.44. Les fichiers d'équivalence ont été considérés précédemment comme l'une des responsabilités prioritaires du bureau de recensement après le dénombrement. Ils sont non seulement utilisés pour établir immédiatement les tableaux de données de recensement, mais ils figurent aussi parmi les produits élaborés. Les utilisateurs de données peuvent avoir besoin de savoir quels districts de recensement appartiennent à une région statistique ou administrative donnée ou quelles unités statistiques de petites dimensions constituent une unité déclarante plus agrégée.

6.45. Les fichiers d'équivalence devraient être mis à disposition sous forme imprimée ou sous forme numérique. La plupart des utilisateurs qui travaillent avec des données numériques de recensement — à référence géographique ou sous forme de tableaux — tireront avantage de la disponibilité de ces fichiers sous une forme consultable sur ordinateur. Cela permettra d'utiliser directement ces fichiers au cours des opérations sur les bases de données.

b) Cartothèque de référence

6.46. Outre les fichiers d'équivalence, le bureau de recensement doit aussi établir les cartes de référence de toutes les unités déclarantes. Dans certains pays, le service cartographique du recensement est tenu en vertu de la loi d'établir ce type de cartes à l'intention des fonctionnaires et du grand public.

6.47. Les cartes de référence peuvent être diffusées sous forme numérique en tant que fichiers graphiques simples ou de fichiers PostScript ou PDF. Cependant, tous les utilisateurs ne sont pas en mesure d'utiliser des fichiers numériques. Il faudra donc mettre à disposition sur demande des ensembles complets de cartes de référence imprimées.

6.48. Les cartes de référence doivent être accompagnées d'une description détaillée des définitions de chaque zone géographique de recensement. Un bon exemple d'une documentation cartographique complète de référence avec le *Geographic Areas Reference Manual* (Manuel de référence des zones géographiques) établi par le Bureau du recensement des États-Unis et disponible sur Internet est proposé à l'adresse suivante : www.census.gov/geo/www/garm.html.

c) Nomenclatures toponymiques et fichiers de centroïdes

6.49. Bien qu'il incombe normalement au service cartographique national d'établir des nomenclatures toponymiques — liste des noms de lieux et de leur emplacement géographique — un programme cartographique national à grande échelle mis en œuvre à des fins de recensement peut offrir une base d'informations améliorée ou mise à jour pour une nomenclature toponymique nationale. Dans certains pays, où d'autres sources ne sont pas disponibles pour ce type de données, une nomenclature toponymique peut être l'un des produits demandés à un projet de cartographie de recensement. Si un tel projet a été réalisé en grande partie à partir de données de GPS, il devrait être facile d'élaborer une nomenclature toponymique énumérant tous les noms de lieux géographiques. Une telle nomenclature peut inclure les noms de tous les lieux peuplés ainsi que les variations de noms, les latitudes et longitudes, les codes administratifs et les populations.

6.50. Une nomenclature toponymique doit être stockée et diffusée sous forme numérique, ce qui permet d'utiliser directement les coordonnées et les informations relatives aux noms dans un SIG. Il est également utile d'élaborer un système simple d'interrogation qui permettra aux utilisateurs de demander des coordonnées d'un lieu particulier tel qu'un village dans telle ou telle province. Ces données peuvent être fournies par l'intermédiaire de l'Internet en utilisant une interface utilisateur simple pour accéder à la base de données. Les fichiers des limites numériques sont également des produits de recensement éventuels. Il en sera question plus loin.

6. Cartes thématiques destinées à être publiées

a) L'avantage des cartes

6.51. Avant d'examiner les types de cartes thématiques susceptibles d'être produites afin d'être intégrées dans des publications sur le recensement, il est utile d'examiner en quoi ces cartes sont efficaces pour présenter les résultats de recensement :

- Les cartes communiquent une notion ou une idée.
- Les cartes sont souvent considérées comme un complément aux informations textuelles. Les thèmes ou questions qu'il est difficile d'expliquer à l'aide de mots peuvent être illustrés plus efficacement avec des cartes ou des images.
- Les cartes suscitent la curiosité du lecteur. Elles constituent des points d'attraction sur les pages d'un rapport. Ainsi elles retiennent l'attention du lecteur et les encouragent à lire le texte qui les accompagne.
- Les cartes résument de grandes quantités d'informations avec concision. Il serait difficile d'évaluer la capacité des cartes à représenter non seulement de grandes quantités de nombres mais aussi des informations sur les relations spatiales entre les observations. Une carte des densités démographiques de comtés en Chine ou aux États-Unis, par exemple, comporte plus de 3 000 valeurs de données. Cette carte peut être imprimée sur une page de

format commercial sans grande perte de clarté. Il serait difficile de faire tenir 3 000 nombres sur une page de format commercial et cela fournirait moins d'informations, notamment sur les zones où les valeurs faibles et élevées se trouvent concentrées dans le pays.

- Les cartes peuvent être utilisées à des fins de description, d'exploration, de confirmation, de tabulation et même de décoration. Les cartes peuvent servir à de nombreuses fins. Les cartes de présentation dans les rapports de recensement sont généralement descriptives de par leur nature. Elles présentent simplement les résultats du recensement avec ou sans commentaire. Par contre, un démographe ou un géographe utilisant des données de recensement pourra utiliser les cartes pour explorer des relations entre différentes variables, par exemple l'espérance de vie et le taux d'alphabétisation. Dans un rapport final, les cartes représentant ces facteurs pourront être utilisées en sus du texte et des graphiques pour appuyer les résultats de l'analyste. La carte devient ainsi un outil pour confirmer des résultats qui pourraient ou non être obtenus en consultant la carte seule. Les cartes peuvent aussi être utilisées simplement à des fins d'inventaire, par exemple pour montrer toutes les écoles ou les cliniques existantes dans un pays. Naturellement, l'inventaire suscite l'analyse, par exemple, en attirant l'attention sur des zones insuffisamment desservies par les services publics. Enfin, les cartes sont populaires parce qu'elles sont souvent belles et accrocheuses. Le nombre considérable de cartes affichées sur des murs de bureaux le confirme. Par contre, rares sont ceux qui affichent des diagrammes statistiques ou des tableaux de chiffres.
- Les cartes encouragent les comparaisons. Qu'elles soient descriptives ou exploratoires, les cartes thématiques visent surtout à comparer différentes choses à travers l'espace géographique. De nombreux types de comparaisons sont possibles :
- Entre différentes zones sur la même carte : où les densités démographiques sont-elles les plus élevées ?
- Entre différentes cartes : la mortalité infantile est-elle plus élevée dans les districts de la province A que dans ceux de la province B ?
- Entre différentes variables dans la même zone : où et dans quelle mesure les taux d'alphabétisation des personnes de sexe masculin et de sexe féminin diffèrent-ils dans les districts ?
- Entre des cartes concernant différentes périodes : les taux de fécondité ont-ils diminué depuis le dernier recensement ?

b) Cartographie thématique des données de recensement

6.52. L'adoption d'une technologie géospatiale dans le cadre du processus de recensement favorise une conception des cartes très différente de la cartographie traditionnelle. Sur un ordinateur, des cartes peuvent être établies rapidement sur l'écran. Cela favorise un mode de travail optimisé pour la validation des données, l'exploration des systèmes de données et l'analyse des données. La géovisualisation est le fruit des principes établis en matière d'affichage et de production de cartes. Elle est définie comme la création et l'utilisation de représentations visuelles afin de faciliter la réflexion, la compréhension et le développement de connaissances sur les environnements physique et humain, à des échelles de mesure géographiques (Longley et autres, 2005).

6.53. Les cartes créées sur un écran d'ordinateur sont parfois appelées « cartes virtuelles » pour les distinguer des cartes imprimées ou dessinées sur papier, bien que

la frontière entre les cartes informatisées et sur papier soit de plus en plus floue. Dans le processus du recensement, on a assez peu à se préoccuper de la conception cartographique traditionnelle aux premiers stades d'un projet cartographique numérique de recensement. Comme indiqué au chapitre III, l'accent est mis sur l'élaboration et la vérification de la base de données. Même l'établissement de cartes de districts de recensement indiquant les principaux objets de la zone de travail de l'agent recenseur ne nécessite en général qu'un travail cartographique relativement simple.

6.54. Mais à partir du moment où les données de recensement auront été rassemblées, le bureau de recensement souhaitera généralement établir des cartes de qualité appropriée pour la publication illustrant les résultats du recensement et accompagnant les rapports de recensement publiés. Ces cartes seront proposées à un public plus large de non-spécialistes. Elles devront donc être établies avec beaucoup plus de soin — que le produit final soit imprimé sous forme de livre, publié sur un CD-ROM ou DVD ou affiché sur un site Web sur Internet.

6.55. Le tableau VI.2. présente une liste des cartes thématiques susceptibles d'être incluses dans un atlas de recensement ou un site Web du bureau de recensement. On pourrait aussi envisager de nombreux autres types de cartes en vue de publications sur des thèmes spécialisés ou pour mettre en lumière des aspects intéressants des résultats de recensement dans les différentes régions du pays. De même que les tableaux de données de recensement peuvent être ventilés par sexe, groupe d'âge ou zones urbaines/rurales, les cartes de recensement peuvent être subdivisées par groupes de population. D'autres cartes informatives présentent des comparaisons dans le temps, à condition de disposer d'indicateurs comparables tirés de recensements antérieurs (pour une vue d'ensemble des principes cartographiques thématiques, voir annexe V).

Tableau VI.2

Suggestions de cartes thématiques pour un atlas de recensement

Catégorie/titre de la carte thématique
<i>Dynamique et répartition de la population</i>
Variation de la population en pourcentage pendant la période intercensitaire
Taux de croissance annuel moyen
Densité démographique (habitants au kilomètre carré)
Population urbaine en pourcentage de la population totale
Répartition et dimensions des grandes villes
Taux d'immigration, d'émigration et de migration nette
Population née dans le pays ou à l'étranger
Population née dans une autre partie du pays
<i>Caractéristiques démographiques</i>
Rapport de masculinité (nombre de personnes de sexe masculin pour 100 personnes de sexe féminin), éventuellement par groupes d'âge
Pourcentage de la population d'âge compris entre 0 et 14 ans
Pourcentage de la population d'âge compris entre 15 et 64 ans
Pourcentage de la population âgée de 65 ans et plus
Pourcentage de la population féminine en âge de procréer entre 15 et 49 ans
Rapport total de dépendance (pourcentage de la population âgée de 0 à 14 ans et de 65 ans et plus par rapport à la population d'âge compris entre 15 et 64 ans)

Catégorie/titre de la carte thématique
Situation matrimoniale
Taux de natalité
Taux de fécondité total
Âge moyen au premier mariage
Taux de mortalité
Taux de mortalité infantile
Espérance de vie à la naissance
Pourcentage de personnes handicapées
<i>Caractéristiques socioéconomiques</i>
Pourcentage d'enfants ne suivant pas l'enseignement primaire
Taux d'alphabétisation des 15 ans et plus
Nombre moyen d'années de scolarisation (des 25 ans et plus)
Taux d'analphabétisme des 15 ans et plus
Population illettrée âgée de 15 ans et plus
Niveau d'éducation de la population âgée de 10 ans et plus
Population active en pourcentage de la population totale
Proportion de femmes actives
Pourcentage de la population active par secteur économique, type d'emploi et statut
<i>Ménages et habitations</i>
Nombre moyen de personnes par ménage
Pourcentage de ménages dirigés par des femmes
Nombre moyen de pièces d'habitation par ménage
Régime foncier (propriété, location, etc.)
Type de matériau de construction
Pourcentage de la population bénéficiant d'un abri adéquat
Pourcentage de la population ayant accès à l'eau potable
Pourcentage de la population alimentée en électricité
Pourcentage de la population ayant accès aux services sanitaires
Pourcentage de la population ayant accès aux services de santé

Source : Nations Unies (2008).

6.56. Les cartes de recensement destinées à être publiées sont généralement établies seulement pour des unités statistiques déclarantes très agrégées. Un service de recensement peut établir des cartes nationales d'ensemble montrant la répartition des indicateurs par province ou par district, ainsi que des cartes plus détaillées pour chaque province. Pour les grandes zones urbaines, des cartes très détaillées peuvent être établies en utilisant des données au niveau des quartiers ou des districts de recensement.

6.57. Les logiciels de SIG et de cartographie sur ordinateur offrent des fonctions cartographiques très variées et de nombreux cartographes commerciaux ont adopté des techniques de production entièrement numériques. Cependant, des produits cartographiques de haute qualité exigent beaucoup d'expérience et de savoir-faire. Les outils mis à disposition par les systèmes de cartographie sur ordinateur ne peuvent se substituer à une formation de cartographe. En fait, l'accès à des logiciels cartographi-

ques faciles d'emploi a favorisé la prolifération de cartes allant à l'encontre de nombreux principes du dessin cartographique traditionnel. Cela s'expliquait au début par l'absence de fonctions cartographiques dignes de ce nom dans les premiers logiciels de SIG. Mais, les programmes de SIG ayant gagné en puissance et en polyvalence, les personnes disposant de peu ou d'aucune formation aux techniques cartographiques auront éventuellement du mal à établir des cartes décentes.

6.58. Dans la plupart des agences de recensement, des cartographes professionnels sont habituellement chargés d'établir des cartes en vue de leur publication et de leur distribution. Ces collaborateurs n'éprouveront guère de difficultés à produire des cartes de grande qualité sur l'ordinateur après avoir reçu une formation en matière de techniques de cartographie numérique.

6.59. Les logiciels de SIG et de cartographie sur ordinateur étant aujourd'hui largement répandus, les cartes thématiques sont de plus en plus établies par des spécialistes de certains sujets qui ne connaissent pas ou insuffisamment les principes du dessin cartographique. Les informations contenues dans l'annexe V devraient intéresser le personnel cartographique permanent ainsi que les personnes à l'intérieur et à l'extérieur de l'organisme de recensement qui peuvent à l'occasion établir des cartes à partir de bases de données spatiales numériques. On trouvera aussi d'autres références de qualité sur la cartographie et les cartes thématiques dans Robinson et autres (1995), Kraak et Ormeling (1997) et Dent (1999). MacEachren (1994) a rédigé un premier manuel utile sur la cartographie thématique, destinée spécialement aux utilisateurs de SIG n'ayant reçu qu'une formation limitée en cartographie. Les guides de conception spécifiquement créés pour les cartes de SIG sont de plus en plus répandus (voir par exemple Krygier et Wood, 2005; et Brewer, 2005).

c) Atlas numériques de recensement

6.60. Si une base de données géographique plus générale s'adresse aux utilisateurs ayant une grande expérience des SIG, un atlas numérique de recensement vise le grand public, les écoles et les autres utilisateurs non experts. Dans les paragraphes suivants, on examinera deux méthodes permettant d'établir un atlas numérique de recensement. Un atlas de recensement *statique* est constitué d'une collection de cartes et d'autres documents établis par le bureau du recensement. Il s'agit essentiellement d'une présentation où l'utilisateur peut modifier l'ordre dans lequel le contenu est visualisé, sans pouvoir modifier le contenu lui-même. Par contre, un atlas de recensement *dynamique* combine une base de données géographique numérique et des données de recensement dans un logiciel cartographique simple. L'utilisateur peut utiliser les données pour établir des cartes personnalisées, qui peuvent être imprimées ou copiées dans d'autres logiciels d'applications.

i) *Atlas statiques*

6.61. Un atlas de recensement numérique statique peut rassembler des cartes, des tableaux, des graphiques et éventuellement des produits multimédia tels que des photographies ou des extraits de films dans un environnement visuellement attrayant et convivial. La présentation peut être préparée dans un logiciel de présentation standard, tel qu'un diaporama PowerPoint. Certains logiciels graphiques de présentation donnent au développeur la possibilité de créer une version autonome de présentation graphique qui peut être distribuée en même temps qu'un logiciel gratuit de visualisation. La plupart des présentations ou des produits graphiques peuvent être exportés en format PDF, qui peut être diffusé sur un support lisible par ordinateur ou via Internet.

démographique (2) où l'une des diapositives (3) montrera une carte, ainsi que des tableaux et graphiques indiquant la proportion de la population de moins de 15 ans. À partir de là, des liens pourront être établis avec des cartes montrant les indicateurs de santé infantile (4), les établissements d'enseignement (5) etc.

6.65. L'utilisation de concepts graphiques orientés vers les hyperliens exige une étude très soignée de la présentation, car les utilisateurs sont facilement perdus après avoir suivi un certain nombre de liens. Il est primordial d'inclure sur chaque page des outils de navigation précis, dont des « fils d'Ariane », à savoir des liens destinés à permettre à l'utilisateur de quitter la page en cours de consultation, et de poursuivre sa navigation.

6.66. Les hyperliens sont naturellement familiers pour tous ceux qui utilisent l'Internet. En fait, plutôt que d'utiliser un progiciel de présentation, un atlas de recensement statique pourra aussi être mis en œuvre dans le langage HTML, langage standard de lecture sur Internet. Les outils de conception de pages Web donnent au développeur une grande souplesse dans la conception de la base de données de recensement. Une carte cliquable par exemple est un outil susceptible de rendre la présentation plus intéressante. L'écran d'entrée pourra montrer par exemple une carte d'ensemble du pays et une invitation à cliquer sur la province qui intéresse le lecteur pour lui permettre d'obtenir des cartes plus détaillées au niveau subnational. La technologie du Web permet aussi d'inclure un contenu multimédia et des liens avec des informations hors présentation, par exemple avec d'autres parties de la page Web du bureau de recensement ou avec d'autres organismes publics. Ne pourront y accéder naturellement que les utilisateurs ayant accès à Internet.

6.67. L'utilisation d'outils de conception pour le Web présente l'avantage que le même atlas de recensement statique peut être distribué sur CD-ROM ou disquette pour un usage individuel et qu'il peut être inscrit sur le site Web du bureau de recensement pour des utilisateurs partout dans le monde.

ii) *Atlas dynamiques*

6.68. Au lieu d'établir un atlas de recensement statique, on pourra publier une carte numérique et une base de données, avec les logiciels cartographiques permettant à l'utilisateur d'établir des cartes personnalisées des indicateurs de recensement. Cela exige bien sûr quelques connaissances de la cartographie de la part de l'utilisateur. Un atlas de recensement dynamique comprendra des fichiers de limites numériques à une résolution inférieure à celle de la base de données de recensement complète, pour que l'on puisse rapidement établir un tracé sans recourir beaucoup au disque. Le tableau des attributs étroitement intégré ne devrait contenir qu'un certain nombre d'indicateurs de recensement sélectionnés. Les densités et les rapports appropriés pour la cartographie devraient déjà être calculés.

6.69. Cette méthode répondra aux besoins des utilisateurs n'ayant pas l'expérience des SIG ni les capacités requises pour utiliser la base géospatiale complète de données censitaires numériques, mais qui souhaitent avoir plus de souplesse dans l'exploration et l'utilisation des informations géographiques de recensement que ce qu'autorise un atlas de recensement statique préétabli.

6.70. On se heurte naturellement à une difficulté, le fait que ces utilisateurs ne disposent peut-être pas d'un logiciel de SIG bureautique qui puisse être utilisé pour créer des cartes. Le fournisseur de données devra donc fournir un logiciel facile d'emploi, avec les limites et les données. L'utilisation de ce logiciel ne devrait exiger qu'un minimum de formation et d'expérience. Pour l'essentiel, l'application devrait être du

type « plug-and-play » — l'utilisateur devant pouvoir établir des cartes immédiatement après l'installation.

6.71. Certains bureaux de recensement ont mis au point en interne leurs logiciels de visualisation de cartes et ils les distribuent avec leurs produits utilisant les données de recensement. Cependant, la tenue à jour de ces programmes est coûteuse et elle mobilise des ressources qui seraient autrement consacrées à la mise au point des données ou à leur distribution. Certains vendeurs de SIG vendent maintenant des boîtes à outils pour logiciels de SIG qui peuvent être assemblées pour produire des applications personnalisées ou pour intégrer des fonctions de SIG dans d'autres produits logiciels (par exemple des tableurs ou des applications de bases de données).

6.72. Sinon, il existe aujourd'hui plusieurs logiciels cartographiques gratuits pouvant être distribués avec une base de données (pour en savoir plus, voir chapitre III).

6.73. Certains vendeurs commerciaux de SIG fournissent aussi gratuitement des logiciels de visualisation et permettent aux utilisateurs de distribuer ces logiciels cartographiques simples dans un programme de distribution de base de données. On peut citer le logiciel ArcGIS Explorer produit par ESRI Inc. de Redlands (Californie). ArcGIS Explorer est une interface cartographique pour les données créées avec le logiciel ArcGIS.

6.74. L'interface ArcGIS Explorer est très facile à utiliser et le système offre des fonctions cartographiques de base pour établir des cartes thématiques qui peuvent être exportées sous forme de métafichiers bitmap ou Windows. ArcGIS Explorer peut lire des données provenant d'un disque dur local ou d'un CD-ROM. Sur les ordinateurs connectés à Internet, il peut aussi afficher des données qui se trouvent sur des sites Web éloignés. Ses fonctions analytiques sont limitées, mais le système prend en charge divers types d'interrogations de données — interactifs ou utilisant des commandes du type SQL — et l'assortiment d'adresses.

6.75. La documentation établie pour un atlas de recensement dynamique doit inclure beaucoup d'informations semblables à celles qui devraient accompagner une base de données censitaires de SIG plus détaillée. Cependant, le texte devra être conçu à l'intention d'utilisateurs qui ne sont pas des experts. Le jargon technique des SIG devraient être évité. Étant donné que les utilisateurs n'utiliseront probablement pas la base de données dans des applications plus avancées, la documentation devra mettre l'accent davantage sur les attributs et moins sur les détails géographiques techniques.

6.76. La figure VI.4, préparée par le comité d'État ukrainien de la statistique, illustre l'utilisation d'éléments graphiques et statistiques dans la même présentation cartographique.

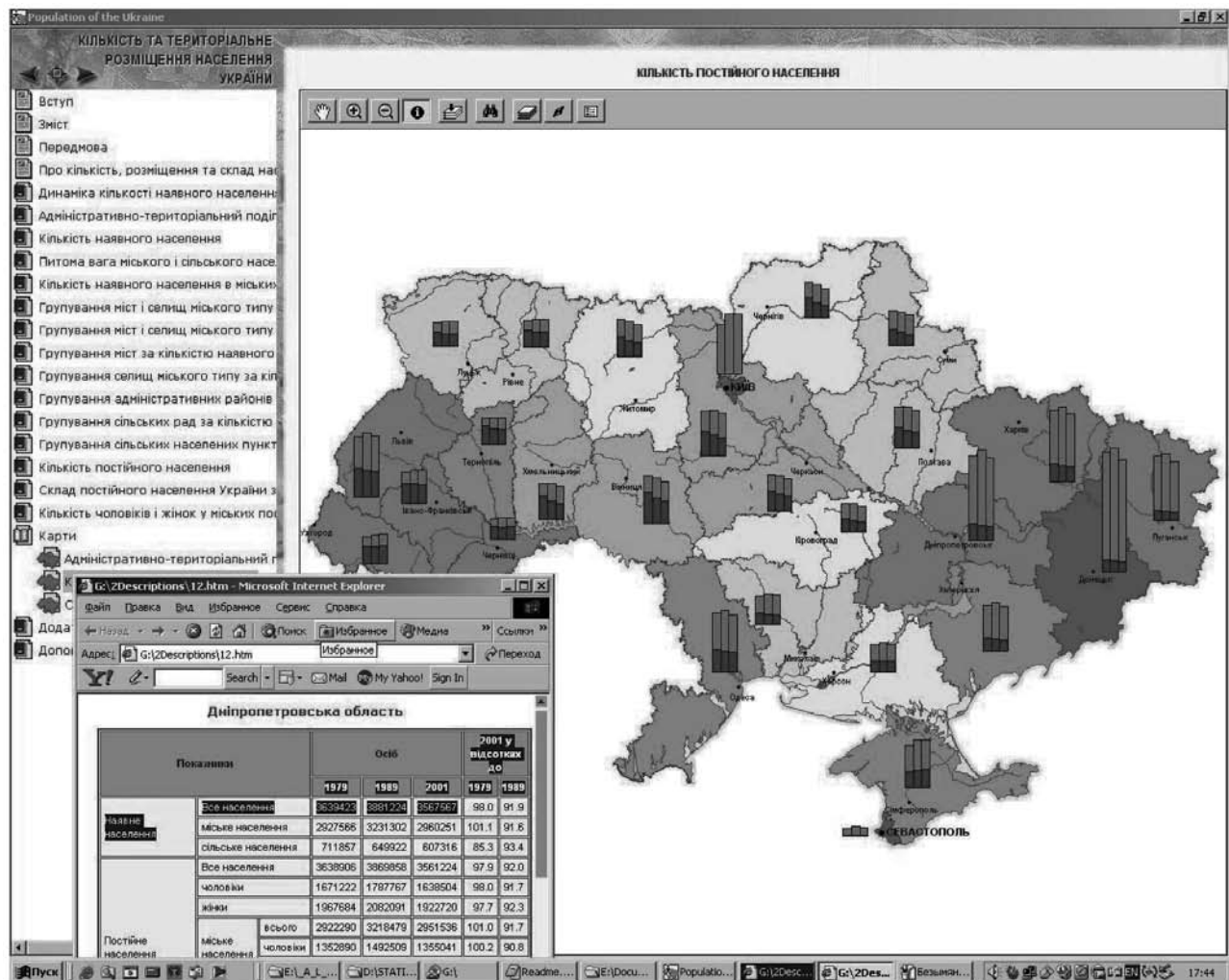
d) Techniques d'analyse spatiale

6.77. Bien qu'il soit parfois procédé à une analyse spatiale en cours du recensement (le regroupement par exemple peut faciliter l'identification d'habitations à démarcher), elle concerne principalement les produits et services du recensement. Une série de techniques, notamment la création de corridors, l'interpolation linéaire, l'analyse des structures de points et les cartogrammes, offrent une fonctionnalité allant au-delà de la cartographie thématique traditionnelle (choroplèthe), de nombreux outils étant désormais proposés sous forme de logiciels commerciaux et libres.

6.78. L'utilisation de nouvelles méthodes d'analyse spatiale présuppose la disponibilité de données démographiques d'une « granularité » (ou spécificité spatiale) supérieure, au niveau des districts de recensement, des agglomérations ou d'autres petites zones. Si les analystes ou d'autres utilisateurs de SIG souhaitent se servir de

Figure VI.4

Instantané d'un atlas dynamique ukrainien



la distribution spatiale de la population ou cartographier entre autres des variables démographiques en fonction d'autres variables, ils disposent à présent d'un éventail de techniques allant de simples requêtes à des transformations, en passant par des résumés descriptifs, des modèles et des mesures.

6.79. Longley et autres (2005) définissent l'analyse spatiale comme un ensemble de méthodes dont les résultats varient en cas de déplacement des objets analysés. L'analyse spatiale est parfois appelée le « cœur des SIG » car elle recourt à des méthodes permettant de transformer les données en informations. Elle peut aider à répondre à des questions telles que : quelle est la relation spatiale entre X et Y ? Où peut-on trouver des caractéristiques similaires ? Ou encore, peut-on trouver un modèle pour décrire un schéma général et où sont les anomalies ?

6.80. Certaines formes très répandues d'analyse spatiale sont particulièrement utiles pour des données démographiques et notamment :

- Les « requêtes » sont considérées comme l'analyse la plus basique. Elles passent par un programme SIG afin de répondre à de simples questions posées par l'utilisateur, sans modifier la base de données ni générer de nouvelles

données. Il s'agit souvent de la première étape d'une analyse dans le cadre de laquelle on cherche à créer un sous-ensemble d'unités, tel que des endroits peuplés, dotés de certaines caractéristiques, grâce auquel l'utilisateur pourra apprécier le caractère typique d'une observation par rapport à d'autres. Voici un exemple de requête utilisant des données censitaires géocodées : « Sélectionner toutes les villes dont la population est supérieure à 1000 personnes ». Les attributs de ces villes peuvent ensuite être résumés notamment pour comparer leurs taux de fécondité totaux par rapport à de plus petites villes et villages. La cartographie des résultats serait l'étape suivante. L'expression « analyse des données exploratoires » fait référence aux études des caractéristiques et tendances au niveau des données à l'aide de techniques telles que des requêtes.

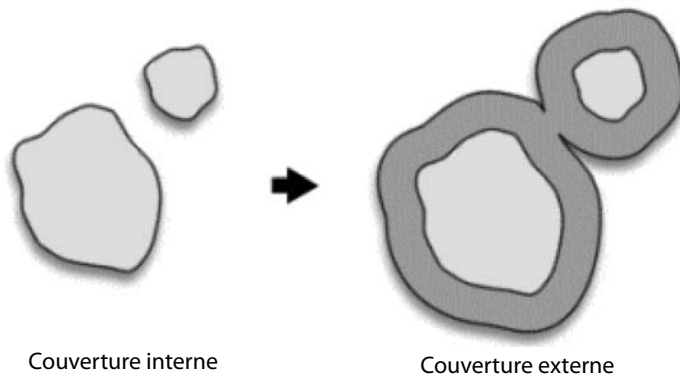
- Les analyses de mesure utilisent les propriétés géographiques des données de population, dont les propriétés simples d'objets telles que la longueur, la zone ou la forme, ainsi que les relations entre les paires d'objets, comme la distance ou la direction, pour décrire certains aspects de telles données. Il est facile de réaliser des « mesures de la distance » avec tous les programmes de SIG grâce aux centroïdes (points centraux) des villes et villages. Une analyse servira à sélectionner les villages situés à plus d'un kilomètre d'une école, d'une clinique ou d'une source d'eau. L'analyse de ces derniers peut être encore plus poussée en utilisant les informations d'attributs spécifiques aux zones peuplées.

6.81. Les deux premiers exemples ont porté sur la consultation ou la mesure d'ensembles de données, mais en l'absence de toute production de nouvelle donnée. Les transformations comptent parmi les méthodes d'analyse spatiale et servent de règles logiques, arithmétiques ou géométriques simples qui sous-tendent la création de nouveaux ensembles de données. Les transformations peuvent donner lieu à des opérations convertissant des données du mode maillé au mode vectoriel voire un flux de coordonnées GPS en une route ou une limite.

6.82. De toutes les techniques de transformation, la « création de corridors » est la plus connue et la plus importante. Elle génère une nouvelle couche de données en identifiant toutes les zones se trouvant à une distance précise des données originales. La création de corridor peut être effectuée sur des points, lignes et polygones (voir figure VI.5) et être pondérée par des valeurs d'attributs. Elle peut servir à modéliser les temps de déplacement, notamment en créant un « bassin » autour d'un objet particulier, tel qu'une école ou une clinique. Il en résulte une mesure de l'accessibilité cartographiable à l'échelle du pays. Les données démographiques peuvent aussi être utilisées avec d'autres ensembles de données, tels qu'une couche de données indiquant le territoire susceptible d'être inondé. L'analyse ainsi obtenue pourrait par extension identifier les populations et habitats menacés qui pourraient bénéficier de programmes d'atténuation des dégâts.

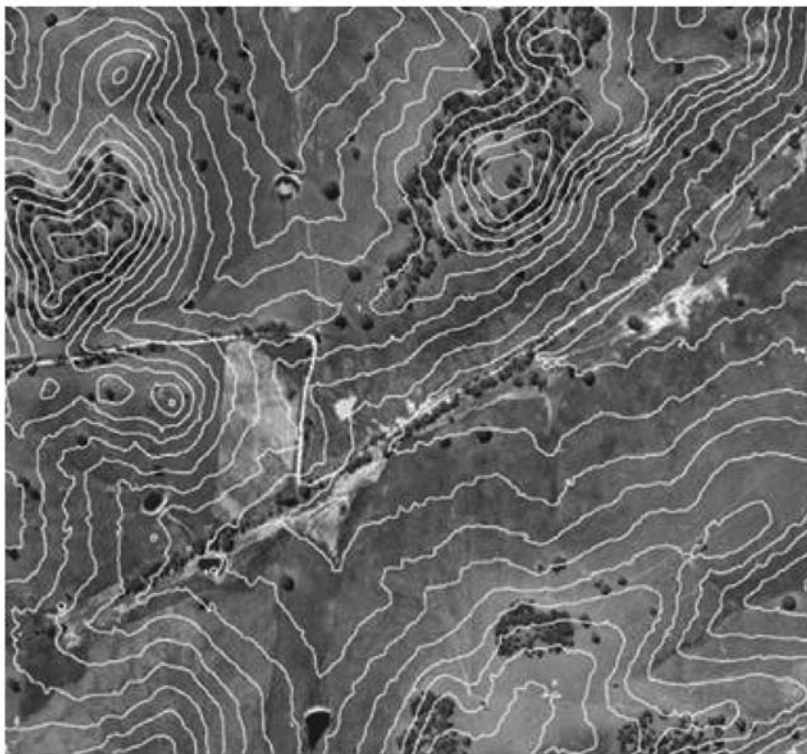
6.83. L'analyse « inscription d'un point dans un polygone » est un autre type de transformation. Elle établit si un point se situe à l'intérieur ou à l'extérieur d'un polygone. Elle sert entre autres à comparer des centroïdes de villages géocodés se trouvant à l'intérieur et à l'extérieur de zones dangereuses (ex. : sur la trajectoire d'une tempête tropicale ou dans des zones de tremblements de terre). La « superposition de polygones » est une analyse qui consiste à comparer les emplacements de deux couches de données polygonales différentes. Ainsi, les limites de deux circonscriptions administratives peuvent être comparées aux erreurs de dépannage dans le processus de dénombrement sur le terrain.

Figure VI.5

Création de corridor pour un objet polygone

6.84. L'« interpolation spatiale » est une méthode d'analyse spatiale destinée à ajouter les valeurs qui se trouvent entre les observations. Un ensemble de méthodes, dont la pondération par l'inverse des distances et le krigeage, servent à estimer les valeurs de sites non échantillonnés, sur la base de la première loi de Tobler. Celle-ci veut que la similarité de tous les objets avoisinants est plus importante que celle des objets distants. Le krigeage modélise les caractéristiques générales d'une surface de manière à en estimer les parties manquantes (voir figure VI.6 pour visualiser comment il est possible d'obtenir les courbes de niveau par interpolation linéaire à partir d'une image satellite).

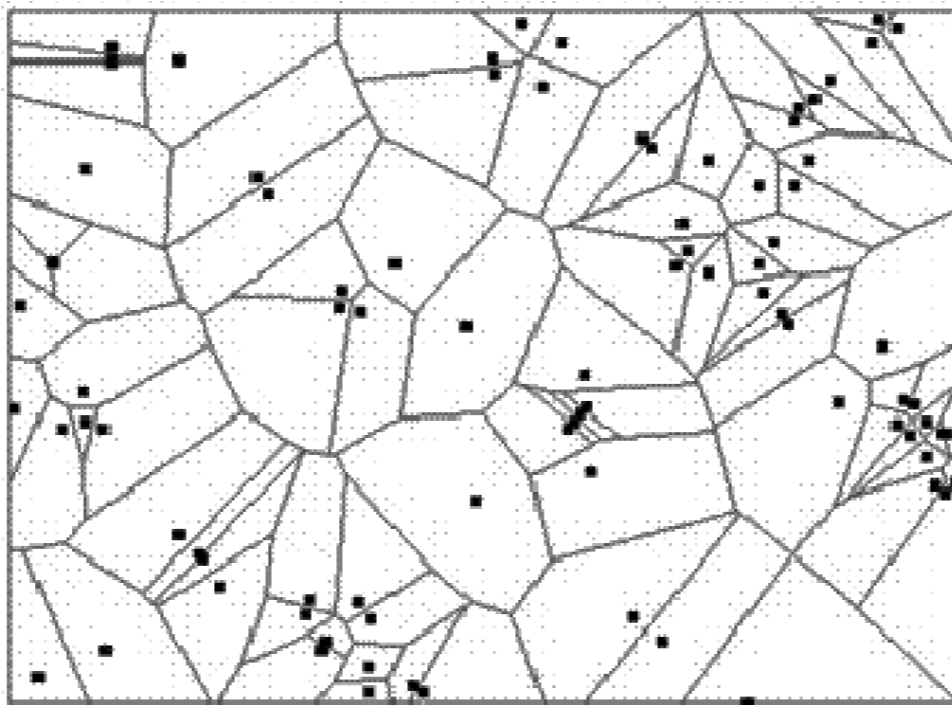
Figure VI.6

Exemple d'interpolation linéaire créant des courbes de niveau

6.85. Les polygones de Thiessen sont des objets spatiaux permettant de créer des zones autour de données ponctuelles, en fonction des distances entre les points déployés dans un espace bidimensionnel. Cette méthode suppose que les valeurs des données non échantillonnées équivalent à celles des points échantillonnés (illustration des polygones de Thiessen à la figure VI.7).

Figure VI.7

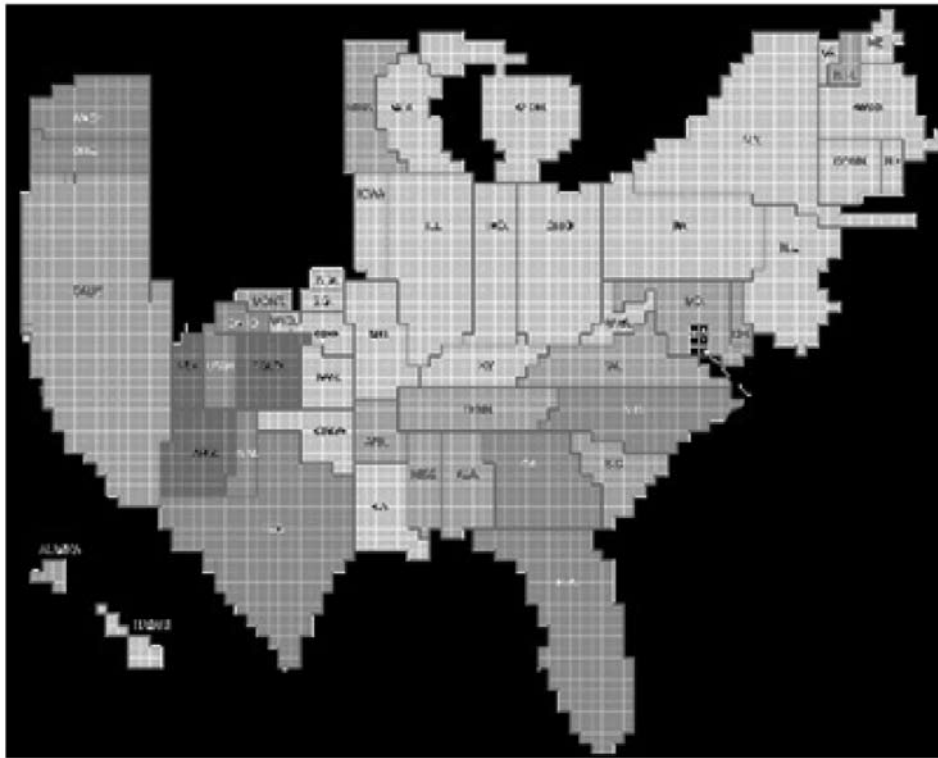
Illustration d'un polygone de Thiessen



6.86. Les états descriptifs sont le pendant spatial des statistiques descriptives (telles que la moyenne et l'écart-type); à 1 ou 2 chiffres, ils sont l'essence même d'une donnée. Les « centres de population » sont l'équivalent bidimensionnel d'une moyenne statistique; ils sont souvent utilisés pour afficher le centre habité, en s'appuyant sur la moyenne pondérée des coordonnées x et y des points peuplés. La « structure de points » ou « analyse en grappes » analyse la répartition des points dans l'espace, indépendamment de leur position réelle afin de déterminer si les patrons sont aléatoires, en grappes ou dispersés. La statistique de Moran (entre 0 et 1) renseigne les propriétés générales de schémas d'attributs; elle peut servir à identifier des « points chauds », où des valeurs élevées sont entourées de valeurs élevées, ou des points froids où de faibles valeurs sont voisins de valeurs faibles. Ils s'avèrent particulièrement utiles pour identifier des populations menacées.

6.87. L'affichage des résultats des recensements est parfois réalisé à l'aide de « cartogrammes » (voir figure VI.8). Dans un cartogramme, les zones des polygones originaux sont soit étendues soit réduites en fonction de leurs valeurs d'attributs (taille de la population ou habitudes électorales par exemple). Les analystes peuvent à présent obtenir des scripts et des extensions auprès de l'ESRI et ailleurs (voir par exemple MAPresso : www.mapresso.com) afin de créer les cartogrammes à l'aide de leurs couches polygonales.

Figure VI.8

Exemple de cartogramme

e) Questions relatives à l'établissement et à la publication de cartes :
types de produits

6.88. Une fois le recensement achevé, l'office de la statistique va créer des produits cartographiques destinés à la publication à différentes fins. On peut citer par exemple :

- Des cartes de référence standard décrivant chaque unité de diffusion statistique définie au cours de l'établissement des tableaux de données de recensement.
- Des cartes destinées à illustrer des rapports imprimés sur les résultats ou les méthodes de recensement. Dans ce cas, les cartes ne font pas l'objet principal de la publication, mais elles en complètent le texte. Elles sont souvent imprimées en noir et blanc, ce qui est plus facile que l'impression en pleines couleurs. Pour une large diffusion, le nombre d'exemplaires imprimés sera relativement élevé. L'impression sera donc effectuée par l'imprimerie de l'organisme de recensement ou un imprimeur extérieur engagé par contrat.
- Des atlas de recensement imprimés allant d'une brève publication de type brochure aux atlas détaillés sur papier comportant des douzaines, voire des centaines de cartes.
- Des atlas numériques de recensement remplaçant économiquement les versions imprimées dans les pays où l'ordinateur est très répandu. Les atlas de recensement peuvent être basés soit sur des cartes statiques préparées à l'avance soit sur une interface simple de cartographie thématique où l'utili-

sateur peut choisir les variables à représenter, le système de classification, les symboles et couleurs cartographiques et un schéma de base.

- La plupart des cartes peuvent être publiées sur Internet ou sous forme imprimée. Ici encore on peut choisir entre des cartes statiques qui ne sont pas différentes des autres images ou photographies publiées sur Internet ou des interfaces cartographiques dynamiques donnant à l'utilisateur la maîtrise du processus de conception thématique.
- Des cartes spécialisées en divers formats seront établies sur demande pour les utilisateurs de données de recensement à l'intérieur ou à l'extérieur du service de recensement. Ces produits seront imprimés en petit nombre sur des périphériques de sortie du service tels que des imprimantes laser ou à jet d'encre.
- Les documents de présentation tels que les diapositives ou les affiches grand format sur des thèmes relatifs aux recensements tireront avantage de l'inclusion de cartes.

f) Outils/logiciels cartographiques

6.89. De nouveaux logiciels de SIG sophistiqués ont grandement facilité la production de cartes de qualité. Il convient néanmoins de rester fidèle aux mêmes principes de conception que les cartographes ont traditionnellement appliqués pour créer des cartes visuellement efficaces.

6.90. La première génération de logiciels de SIG n'était pas assortie d'outils cartographiques appropriés. Les produits cartographiques étaient obtenus au moyen d'interfaces à ligne de commande ou de macro-langages. Pour ajouter du texte sur une carte, l'utilisateur devait spécifier la coordonnée sur la page de la carte pour déterminer l'emplacement, les dimensions et le style du texte sous forme de commandes distinctes. La nouvelle génération de logiciels cartographiques sur ordinateur offre des fonctions de dessin cartographique très améliorées. Les utilisateurs accèdent à de nombreux modèles de polices, de lignes et de remplissages, ainsi qu'à des images graphiques qui peuvent être intégrées au modèle de cartes. Les logiciels sont aussi assortis d'ensembles de symboles cartographiques spéciaux fournissant des symboles de points ou de lignes couramment utilisés dans les cartes topographiques et thématiques. L'interface utilisateur des logiciels de cartographie sur ordinateur ressemble en général beaucoup à celle des logiciels graphiques courants où l'utilisateur peut sélectionner des styles dans des menus interactifs et où des éléments de cartes peuvent être déplacés et redimensionnés en utilisant la souris de l'ordinateur. L'affichage de la carte sur l'écran montre de façon très réaliste comment la carte apparaîtra sur la page imprimée.

6.91. Les fonctions de dessin cartographique des logiciels modernes de cartographie sur ordinateur et de SIG répondent à la plupart des besoins des utilisateurs. Mais pour certaines applications, les cartographes professionnels préfèrent exporter la carte de base du SIG et l'importer dans un logiciel de graphisme ou de publication assistée par ordinateur ou un logiciel graphique. Ces logiciels possèdent des fonctions graphiques perfectionnées telles que des effets tridimensionnels, des remplissages graduels ou des effets de transparence qui donnent au cartographe une plus grande latitude en matière de dessin. Pour copier à partir du SIG dans le logiciel graphique, deux options sont disponibles. La première consiste à utiliser les options courantes de copier-coller dans l'environnement Windows. La seconde consiste à passer par un fichier intermédiaire dans un format courant qui peut être importé par un logiciel graphique. Les options de sortie sont décrites ci-après.

7. Options de sortie : fichiers numériques

6.92. Les divers formats pour la diffusion de données de fichiers numériques sont abordés dans les rubriques qui suivent. Les analystes doivent être au fait de la différence entre la diffusion de formats graphiques (formats finis destinés à la production) et des formats de données (version plus brute).

6.93. Tous les logiciels de SIG et les logiciels graphiques permettent à l'utilisateur d'exporter le tracé de la carte vers divers formats de fichiers graphiques. Cette option est utile pour différentes raisons. Elle permet l'échange de fichiers entre différents logiciels. Par exemple, une carte de base provenant d'un SIG et des diagrammes tirés de logiciels statistiques peuvent être exportés vers un logiciel graphique utilisé pour dessiner le tracé final de la page. Le graphique obtenu peut être importé dans un logiciel de traitement de texte pour être intégré dans un rapport ou une publication. La plupart des graphiques du présent *Manuel* ont été établis de cette façon. Les fichiers graphiques peuvent être incorporés à des sites Web en tant qu'images cartographiques statiques et peuvent aussi être échangés en tant qu'annexes de fichiers transmis par courrier électronique.

6.94. Les formats de fichiers graphiques — semblables aux structures de données des SIG — peuvent être subdivisés en formats prenant en charge les graphiques vectoriels et en formats constitués de fichiers matriciels ou de fichiers image. Les images matricielles représentent des objets graphiques sous forme de variations de couleurs ou de nuances de gris ou de petites taches ou pixels disposés dans un réseau régulier. Des tonalités de couleurs ou des échelles de gris continues sont utilisées pour les images de type photographique. On a besoin d'un plus petit nombre de couleurs pour représenter les objets plus discontinus que l'on trouve généralement sur les cartes thématiques.

6.95. Les formats de graphiques vectoriels représentent les objets graphiques sous forme de points, de lignes et de surfaces au moyen d'un système de coordonnées interne qui peut être soit amovible, soit lié à une dimension de page de sortie. Certains formats de fichiers peuvent traiter aussi bien des images matricielles que des objets vectoriels. Ces formats sont utiles pour les cartes SIG qui combinent par exemple des images satellitales et des couches de données linéaires et polygonales. Que l'on utilise un format graphique matriciel ou vectoriel, le contenu graphique doit être converti en images point par point avant que la formation puisse être affichée sur l'écran ou l'imprimante, qui sont tous deux essentiellement des périphériques d'affichage matriciels. Cela est réalisé automatiquement par le système d'exploitation de l'ordinateur et les pilotes d'imprimante.

6.96. On trouvera ci-après une brève description des formats de fichiers les plus couramment utilisés. Cette liste n'est nullement exhaustive, car il existe des dizaines de formats différents.

a) Formats des données

6.97. Les formats de fichiers vectoriels sont plus étroitement associés aux données vectorielles de SIG. Ils peuvent représenter des données linéaires ou polygonales sous une forme plus compacte et conservent toute la résolution des couches de données initiales du SIG. On peut citer divers formats graphiques vectoriels courants :

- **WMF.** Ce métafichier Windows est un format de fichier graphique destiné à être utilisé dans un environnement Windows. Il est le plus souvent utilisé pour des données vectorielles, mais peut aussi enregistrer des images en mode point. Les fichiers EMF (Enhanced WMF) sont une variante plus

complète du format WMF élaborée pour l'environnement Windows 32 bits. WMF est l'un des formats les plus stables pour exporter et importer des fichiers graphiques entre différentes applications Windows. WMF est aussi l'un des formats utilisés par Windows quand un objet graphique est copié sur le presse-papiers, puis collé dans une autre application.

- **VML.** Le VML (Vector Markup Language) est un langage XML utilisé pour produire des graphiques vectoriels. Microsoft, Macromedia et d'autres l'ont proposé en 1998 au W3C en guise de norme, mais il a été rejeté en tant que norme Web, Adobe, Sun et d'autres ayant soumis une proposition rivale, le PGML. Les deux normes ont été fusionnées pour créer le SVG. Bien que refusé en tant que norme par le W3C et largement ignoré par les développeurs, Microsoft a continué à mettre en œuvre le langage VML dans Internet Explorer 5.0 et les versions supérieures ainsi que dans Microsoft Office 2000 et supérieur. Google Maps se sert actuellement du VML pour le rendu des vecteurs sur Internet Explorer 5.
- **CGM.** Les métafichiers d'infographie sont une norme internationale pour enregistrer des données graphiques en deux dimensions. Ils ont été élaborés d'abord en tant que norme purement vectorielle, mais les versions ultérieures prennent aussi en charge les images matricielles. Il existe trois types de formats CGM : le premier est un codeur de caractères qui réduit la dimension de fichier et accroît la vitesse de transmission, le second un code binaire pour la rapidité d'accès et le troisième un mode texte en clair pour la mise en forme à partir d'un fichier.
- **HPGL.** Le langage graphique de Hewlett-Packard est un format de fichier utilisé à l'origine pour les traceurs à plume. Avant la mise en service des imprimantes à jet d'encre grand format et des imprimantes électrostatiques, les traceurs à plume étaient le périphérique de sortie le plus largement utilisé pour les projets de SIG qui exigeaient l'impression de grandes cartes.
- **DXF.** Le DrawingXchange Format a été mis au point par Autodesk, producteur de logiciels spécialisé dans les logiciels de conception assistée par ordinateur (CAO) et de SIG. Établi au départ pour échanger les dossiers en mode naturel d'Autodesk entre différentes plates-formes, DXF est devenu un format d'échange courant qui est pris en charge par la plupart des logiciels de SIG et de nombreux logiciels graphiques.
- **PS et EPS.** Postscript est essentiellement un langage de programmation utilisé pour décrire des données vectorielles dans un fichier texte en clair. C'est le logiciel le plus largement utilisé pour la description de mise en page. Postscript a été mis au point par Adobe, entreprise de logiciels graphiques. Optimisés pour les graphismes vectoriels indépendamment de l'échelle, les fichiers PostScript peuvent aussi inclure des images matricielles. PostScript est principalement utilisé comme un format de sortie pour envoyer des documents et des graphismes à des imprimantes PostScript. Postscript est donc essentiellement un format de sortie. De nombreux logiciels graphiques prennent en charge l'importation de fichiers PostScript mais parce que les codes PostScript ne sont pas complètement normalisés, il est souvent impossible d'importer des fichiers PostScript en vue de leur mise en forme ultérieure s'ils ont été créés dans un programme informatique différent. Cela est particulièrement vrai lorsque le fichier PostScript circule entre des plates-formes matérielles. Parfois il n'est même pas possible d'importer un fichier PostScript créé dans le même logiciel.

S'il est souvent impossible de modifier un fichier PostScript importé, la plupart des progiciels pourront inclure un fichier PostScript dans un document. Au lieu du contenu du fichier, seule une case étiquetée apparaîtra à l'affichage sur écran. Une fois envoyé à une imprimante PostScript, le contenu effectif du fichier PostScript sera imprimé. Étant donné que les fichiers PostScript sont indépendants de l'échelle, le document graphique importé peut être redimensionné pour remplir l'espace voulu (pour les PDF (Portable Document Format), voir par. 6.102 ci-dessous).

b) Formats d'images matricielles

6.98. Des images matricielles peuvent être créées directement au moyen de logiciels de SIG ou de logiciels graphiques. Dans certains cas, deux autres options concernant la création d'images sont utiles. L'une consiste à utiliser la commande de saisie d'écran dans un logiciel graphique orienté matriciel. Ces utilitaires de saisie d'écran donnent parfois de meilleurs résultats pour préserver les couleurs originales d'affichage que les fonctions d'exportation des logiciels de SIG ou des logiciels graphiques. Une seconde option consiste à utiliser un élément spécialisé de logiciel ou de matériel pour convertir les objets graphiques en images matricielles. Ces processeurs d'images matricielles peuvent par exemple produire des images de très haute résolution conservant tous les détails d'un format vectoriel. Les fichiers de sortie obtenus peuvent cependant devenir très volumineux.

6.99. Les dimensions des fichiers dépendent de deux facteurs : le nombre de couleurs disponibles dans l'image et le degré de compression de l'image. Par exemple, un format d'images qui ne prend en charge que deux couleurs (noir et blanc) n'exige qu'un bit pour représenter chaque pixel. Huit bits (un octet) par pixel peuvent enregistrer jusqu'à 256 couleurs et les affichages ou formats d'images de qualité utilisant 24 ou 32 bits par pixel peuvent enregistrer plus de 16 millions de couleurs. Pour les cartes thématiques, un nombre relativement faible de couleurs différentes est en général suffisant. Pour des photographies ou des images graphiques photoréalistes, des formats d'image en 16 ou 24 bits sont plus utiles.

6.100. La plupart des formats d'image utilise une forme de compression qui réduit la dimension du fichier. Le système de compression le plus simple est le codage par longueur de ligne, technique également utilisée dans certains systèmes matriciels de SIG. S'il y a de nombreux pixels de la même couleur dans une ligne d'image, le système n'enregistre qu'une seule fois le nombre de répétitions et la couleur de pixel. Par exemple, cinq pixels avec le nombre de couleur quatre seront représentés par une paire de nombres 5,4 plutôt que sous la forme 4,4,4,4,4. Le numéro de couleur représente en fait un indice sur une table de couleurs figurant dans un petit label de début de fichier et contient la spécification de couleur dans un modèle de couleur courant tel que RVB.

6.101. On peut citer quelques formats de fichiers matriciels courants :

- **BMP.** Le format mode point transportable de Microsoft Windows. Il permet à Windows d'afficher l'image en mode point sur pratiquement n'importe quel périphérique d'affichage. C'est l'un des formats de fichiers matriciels les plus élémentaires. Il prend en charge le codage par longueur de ligne, mais les dimensions de fichiers sont en général plus importantes que pour d'autres formats d'image.
- **TIFF.** Ce format de fichier image étiqueté est l'un des formats image les plus largement pris en charge. Il accepte des couleurs plus ou moins variées et un certain nombre de systèmes de compression. Les images TIFF peuvent être importées par la plupart des progiciels prenant en charge des graphis-

mes, bien que les problèmes puissent parfois se poser pour importer des images créées sur une plate-forme matérielle différente. Le format TIFF est particulièrement important pour les applications géographiques, car il est souvent utilisé comme format pour afficher des images satellitaires, des photographies aériennes, des cartes scannées ou d'autres données matricielles dans un logiciel de SIG ou de cartographie sur ordinateur. Comme on avait besoin d'un format de fichier courant indépendant de la plate-forme utilisée pour les images géospatiales, on a mis au point la norme GeoTIFF. Cette norme fournit la spécification nécessaire pour les informations contenues dans l'en-tête d'image TIFF qui décrit toutes les informations géographiques associées à l'image, telles que la projection, les coordonnées du monde réel, l'aire de la carte etc. tout en répondant aux spécifications TIFF courantes. GeoTIFF est utilisé par la plupart des vendeurs principaux de SIG, des organismes publics et des institutions universitaires.

- **GIF.** Le fichier d'échange graphique a été créé en vue d'échanger des graphismes en mode image matricielle entre les plates-formes matérielles. Il prend en charge un système de compression qui réduit sensiblement les dimensions de fichiers et qui est donc optimal pour assurer l'échange par l'intermédiaire des réseaux informatiques. En fait, ce format a été mis au point par CompuServe en vue de son utilisation dans ses premiers services de babillard. GIF, qui prend en charge jusqu'à 256 couleurs, est un des deux formats d'image matricielle pris en charge par les programmes de lecture sur le Web. La plupart des images matricielles non photographiques sur les pages Web sont au format GIF.
- **JPEG.** Créé par le Joint Photographic Experts Group, le format JPEG a été mis au point en tant que système de compression d'images comportant un grand nombre de couleurs ou de nuances de gris telles que des photographies ou des images graphiques photoréalistes. Le format JPEG est aussi pris en charge par les programmes de lecture sur le Web et est utilisé pour afficher des photographies sur les pages Web. JPEG possède une option de compression variable, qui n'est pas pleinement inversible. Cela signifie qu'une photographie qui a été exportée avec un degré élevé de compression ne peut être restaurée pour montrer tous les détails de la photographie d'origine. Les logiciels de SIG prennent désormais en charge un nouveau format, le JPEG 2000.
- **PNG.** Le format de fichier (Portable Network Graphics) a été créé pour être le successeur gratuit et libre du format de fichier GIF. Le format PNG prend en charge la couleur naturelle (16 millions de couleurs), le fichier GIF ne supportant que 256 couleurs. Le PNG excelle lorsque l'image comporte de grandes zones de couleur uniforme. Le format PNG « sans perte » est plus adapté à l'édition d'images, tandis que les formats impliquant une « perte » comme le JPG sont plus efficaces au stade de la distribution finale d'images de type photographie compte tenu de la taille plus petite des fichiers. De nombreux anciens navigateurs ne prennent pas en charge le format de fichier PNG, mais avec la sortie d'Internet Explorer 7, il est totalement compatible avec tous les navigateurs populaires modernes. L'entrelacement Adam7 procure un aperçu rapide lorsque seul un faible pourcentage des données de l'image a été transmis. Le langage de balisage géographique (GML) a été recommandé par l'OpenGIS Consortium (www.opengis.org) et est actuellement mis en œuvre (pour en savoir plus, voir www.opengeospatial.org/standards/gml).

- **DGN.** Le format de fichier de dessin de MicroStation (.dgn) est utilisé par les logiciels d'environnement modulaire de SIG de Bentley (MGE) et des logiciels géographiques de SIG. Ce format ne prend pas directement en charge les données attributaires, mais il fournit des liens avec des tableaux de bases de données extérieures. Un format d'exportation séparé combine fichiers géographiques et fichiers d'attributs.

6.102. Le format de fichier PDF est véritablement unique en ce sens qu'il ne relève ni vectoriel ni matriciel mais est sans aucun doute largement utilisé.

- **PDF.** Le Portable Document Format (format de document portable) a été élaboré par Adobe. Il a servi au départ à diffuser des documents complexes — contenant du texte et des graphiques — sur Internet. Des fichiers PDF peuvent être créés à partir de n'importe quel logiciel de traitement de texte ou logiciel graphique en utilisant le pilote d'imprimante Adobe Acrobat une fois celui-ci installé. Le lecteur PDF peut être téléchargé gratuitement à partir du site Web d'Adobe. Certains experts prévoient que le format PDF remplacera les fichiers PostScript comme norme principale pour l'impression d'images graphiques de haute qualité. Le langage PDF est plus simple que PostScript, ce qui rend les fichiers PDF plus faciles à convertir en images point par point. La conversion en mode matriciel d'un fichier graphique est nécessaire pour permettre l'affichage sur un écran d'ordinateur et pour imprimer avec une résolution élevée.

c) Formats de données de SIG

6.103. Il existe toute une série de formats pour diffuser des données spatiales brutes à des utilisateurs de SIG spécialisés et des analystes spatiaux. Il y a entre autres ceux applicables aux données de coordonnées géographiques et ceux spécifiques aux données présentées sous forme de tableaux, lesquelles peuvent être unies ou liées en tant qu'attributs à des données spatiales dans un SIG.

d) Données relatives aux coordonnées

6.104. Il existe de grandes différences entre les logiciels de SIG du point de vue des formats de données. Chaque logiciel commercial a son format de données propre. En outre, des fonctions d'importation et d'exportation permettent à l'utilisateur de convertir des données à partir d'un certain nombre de données extérieures. Dans certains cas, ces fonctions de conversion doivent être achetées séparément. Pour l'essentiel, de nombreux logiciels de fournisseurs peuvent lire les logiciels d'autres fournisseurs.

6.105. En dépit de certains efforts de groupes de technologie géospatiale commerciaux et publics (voir par exemple www.opengis.org pour prendre connaissance des discussions en cours sur le format de données GML), il n'existe toujours pas de format d'échanges de données générales universellement accepté et largement utilisé pour les SIG. Par contre, un certain nombre de formats d'échange élaborés par les principaux vendeurs de logiciels de SIG sont devenus des normes de fait qui sont aussi pris en charge par les autres systèmes de logiciels. Les plus importants sont décrits brièvement ci-après :

- Le **format d'exportation d'Arc/Info (.e00)** a été élaboré en tant que format d'échange entre plates-formes pour les bases de données de SIG créées au moyen du SIG Arc/Info de l'Environmental Systems Research Institute (ESRI). Les fichiers d'exportation peuvent être comprimés pour intégrer des

fichiers de plus petites dimensions. Cependant, pour une compatibilité maximale, le format d'exportation non comprimé est en général préférable. Les fichiers obtenus peuvent ensuite être comprimés à l'aide d'un programme courant de compression et d'archivage tel que PKZIP. Le format .e00 n'est pas publié, mais de nombreux autres logiciels de SIG ont élaboré des routines d'importation.

- Les **fichiers « shape »** (.shp) d'ESRI sont un format plus simple utilisé par les logiciels ArcView et ArcGIS d'ESRI. Une base de données de fichiers « shape » est constituée de plusieurs fichiers contenant les données relatives aux coordonnées, un répertoire spatial et des données attributaires. Leurs formats de fichiers sont publiés et de nombreux autres systèmes de SIG peuvent importer des fichiers « shape ».
- Le **format File Geodatabase** est le nouveau format de données d'origine recommandé utilisé avec ArcGIS. Les File Geodatabases sont stockés en tant que dossiers dans un système de fichier et sont conçus pour être compatibles avec la totalité du modèle d'information de la base de données géographique, en ce compris les topologies, les catalogues matriciels, les ensembles de données en réseau, les ensembles de données de terrain et les localisateurs d'adresse. Chaque fichier File Geodatabase peut aller jusqu'à 1 téraoctet. Les bases de données géographiques personnelles utilisent le système de fichier Microsoft Access (.mdb) et peuvent atteindre jusqu'à 2 gigaoctets. Le format File Geodatabase est également employé par le moteur de base de données spatiale ArcSDE.
- Le **format KML** (Keyhole Markup Language) de Google est un schéma de langage basé sur le XML. Il permet de rendre les annotations géographiques et les visualisations sur des cartes en ligne (2D) et des navigateurs terrestres (3D) actuels ou à venir. Le KML a été développé à l'origine pour être utilisé avec Keyhole, la société remplacée acquise par Google en 2004. Le fichier KML spécifie un ensemble de caractéristiques (repères, images, polygones, modèles 3D, descriptions textuelles, etc.) pour un affichage dans Google Earth, Maps et Mobile ou tout autre navigateur terrestre 3D (géonavigateur) faisant appel à un codage KML. Chaque endroit est toujours caractérisé par une latitude et une longitude. D'autres données permettent un affichage plus spécifique (ex. : inclinaison, cap, altitude) qui, ensemble, définissent une « vue de caméra ». Les fichiers KML sont très souvent distribués sous la forme de fichiers KMZ, en réalité des fichiers KML zippés, portant une extension .kmz (pour le GML, voir par. 6.101 et 6.105 ci-dessus).
- Le **format d'échange MapInfo** (.mif) est utilisé pour échanger des fichiers produits par MapInfo, l'un des principaux systèmes de cartographie sur ordinateur. Les fichiers MIF sont établis en format ASCII et peuvent être lus par de nombreux programmes.
- Le **format DXF** (.dxf) d'AutoCAD a été créé dans un environnement de CAO. Il est tout à fait approprié pour transférer les coordonnées géographiques, mais il est moins efficace pour convertir des informations attributaires.

6.106. Tous ces formats prennent en charge les informations relatives aux limites et aux attributs. Les logiciels commerciaux de SIG comportent tous une fonction d'importation pour au moins un ou deux de ces formats. Idéalement, un bureau de recensement devrait offrir ses bases de données de SIG destinées à être publiées dans différents formats pour répondre aux besoins d'utilisateurs très variés possédant des compétences variables en matière de SIG et disposant de plates-formes logicielles dif-

férentes. Pour choisir un format de diffusion, on devra prendre en compte les informations relatives aux systèmes cartographiques les plus largement utilisés par les groupes d'utilisateurs de données censitaires ainsi que la souplesse et la robustesse du format de données.

6.107. La diffusion de données de SIG dans leur format interne d'origine — par exemple un répertoire contenant une couverture Arc/Info ou un espace de travail MapInfo — n'est pas généralement une option viable. Souvent les données en format originel ne peuvent être transférées vers un autre système d'exploitation, on peut rencontrer des incompatibilités de nom de parcours et les autres logiciels de SIG ne peuvent pas en général importer des formats originels de données de SIG. Il est donc toujours préférable d'utiliser un format d'échange de données robuste, du type mis en œuvre par la plupart des logiciels commerciaux de SIG.

e) Données présentées sous forme de tableaux

6.108. Les nouveaux développements logiciels de SIG ont été moins orientés sur l'importation de tableaux pour privilégier les bases de données relationnelles, telles qu'Oracle et Access. La plupart des logiciels de SIG prennent encore en charge plusieurs formats de fichiers pour les données attributaires. Certains ont aussi des fonctions permettant de relier la base de données relative aux coordonnées à un système externe de gestion de base de données. Cependant, pour diffuser les données, il est préférable d'utiliser un format de fichier simple et largement utilisé pour les tableaux de données. Le format le plus couramment utilisé est le format dBase (.dbf), qui peut être produit par la plupart des logiciels de gestion de bases de données et de tableurs ainsi que par des logiciels pour l'établissement de tableaux de données de recensement comme REDATAM (REtrieval of DATa for small Areas by Microcomputer) et CSPro (Census and Survey Processing System), une suite d'outils d'analyse démographique disponible gratuitement auprès du United States Census Bureau. Le format CSV (valeurs séparées par des virgules) est également utilisé dans des tableaux et n'est pas propre aux vendeurs.

6.109. Si la diffusion de données sous forme de tableaux dans le format dBase assure une large compatibilité avec les logiciels de SIG, ce format comporte plusieurs limitations. Par exemple, les noms de champs énumérés sur la première ligne du tableau sont limités à 10 caractères. La documentation du tableur ou du logiciel de gestion de base de données fournira des détails sur les questions de compatibilité. Dans la présentation du tableau, le champ le plus important est l'identificateur commun utilisé pour relier les attributs aux limites de l'unité déclarante. Il devrait figurer dans la première colonne de chaque tableau d'attributs. Il est en général souhaitable de trier les lots de données dans un ordre cohérent, par exemple au moyen de leurs identificateurs géographiques.

D. Impression

1. Aperçu

6.110. Pour effectuer des tirages limités ou réaliser des lots en vue du contrôle de la qualité, un bureau de recensement doit disposer d'une ou de plusieurs imprimantes. Les techniques d'impression ne cessent d'évoluer et la gamme de produits disponibles est très large. Pour choisir des imprimantes appropriées, un bureau de recensement doit prendre en considération les critères suivants :

- Le coût du matériel, de l'entretien et de l'impression par page.
- Le débit (pages par minute).
- La résolution des produits obtenus en points par pouce (ppp) et le nombre de couleurs et de nuances de gris qui peuvent être obtenus.
- La dimension du support.
- Les types de supports utilisables (papier ordinaire, papier couché spécial, transparents, etc.).

2. Types d'imprimantes

6.111. Les paragraphes qui suivent décrivent les types d'imprimantes les plus courants (la meilleure solution en termes de rentabilité et de fiabilité pour la plupart des besoins des organismes nationaux de statistique est l'imprimante laser, mais d'autres options sont possibles) :

- Les **imprimantes laser** utilisent un rayon laser et un système de dispositifs optiques pour décharger sélectivement une surface photoconductrice. Un toner chargé d'électricité statique de sens opposé est alors mis en contact avec cette surface et est attiré dans les zones qui retiennent la charge. Le toner est ensuite transféré sur la page et fixé. Un processus semblable à celui de la photocopie électrostatique est ensuite utilisé pour appliquer l'image du tambour sur le papier. Les imprimantes laser monochromes peuvent donner une qualité d'impression proche de celle obtenue avec les systèmes professionnels de composition. Les imprimantes laser couleur ont atteint un niveau de prix qui permet d'en envisager l'emploi dans la plupart des environnements d'applications graphiques.
- Les **imprimantes à jet d'encre** produisent des documents imprimés en projetant sur la page des gouttes de couleur électriquement chargées à travers une buse. Les imprimantes à jet d'encre liquide utilisent une encre liquide qui sèche par évaporation. De l'encre est envoyé à travers la buse en utilisant la pression hydraulique produite par un procédé piézo-électrique. Par contre l'imprimante à jet d'encre thermique utilise la chaleur pour créer une bulle d'encre dans la buse. Lorsqu'elle est assez grande, la bulle est chassée à travers la buse sur le papier. Les imprimantes à jet d'encre solide utilisent une encre qui doit être fondue à partir de l'état solide pour pouvoir être projetée sur le papier où elle se solidifie rapidement. Les imprimantes à jet d'encre solides produisent des points plus fins sur la page que les imprimantes à jet d'encre liquide. Les imprimantes à jet d'encre utilisent du papier ordinaire, mais pour obtenir des produits de la plus haute qualité possible, il est en général recommandé d'utiliser un papier couché spécialement conçu pour être utilisé avec les imprimantes à jet d'encre. En raison de leur coût raisonnable et de leur facilité d'utilisation, les imprimantes à jet d'encre, disponibles pour une large gamme de dimensions de pages, sont actuellement les périphériques d'impression en couleur les plus largement utilisés.
- Les **imprimantes à transfert thermique** utilisent un papier spécial et des rubans encres qui se déplacent devant une tête thermique. L'encre est thermofixée sur le papier là où il est chauffé par la tête thermique. Les rubans de couleur sont couverts de trois [cyan, magenta, jaune (CMJ)] ou de quatre [cyan, magenta, jaune, noir (CMJN)] couleurs ce qui nécessite trois ou quatre passages de la tête thermique d'un côté à l'autre du papier. Dans les impriman-

tes à transfert thermique utilisant un ruban imprégné de cire, la chaleur fixe une couche de cire colorée sur le papier. Dans les procédés à sublimation, le colorant est diffusé dans la surface à imprimer. Les imprimantes à sublimation donnent une plus grande résolution et une plus grande variété de couleurs que les imprimantes à transfert thermique utilisant un ruban imprégné de cire.

- Les **imprimantes électrostatiques** utilisent un toner qui est transféré par décharge électrique sur une surface non conductrice. Le toner est soit attiré, soit repoussé. Les imprimantes électrostatiques directes appliquent la charge directement sur un papier spécialement enduit. Un toner est appliqué pour chaque couleur par passages séparés. Le toner est ensuite thermofixé sur le papier lorsque toutes les couleurs ont été appliquées. Un autre procédé électrostatique est la xérographie couleur utilisant un tambour ou un ruban qui se charge lorsqu'on l'expose à la lumière.

6.112. De nombreuses cartes en mode rapide n'ont pas besoin d'être imprimées en couleur. En fait, il est plus facile de photocopier les cartes de petit format en noir et blanc. Les imprimantes laser qui prennent en charge le papier A4 ou de format commercial associent rapidité d'impression et très haute résolution (600 ppp et plus). Elles sont idéales pour imprimer des rapports et d'autres documents constitués surtout de texte, avec quelques illustrations graphiques et des cartes.

6.113. Les imprimantes couleur sont utiles pour imprimer des cartes complexes pour lesquelles la symbolisation et l'ombrage monochrome seraient insuffisants. Les imprimantes à jet d'encre sont actuellement les imprimantes couleur les plus couramment utilisées — des imprimantes de bureau format A4 aux imprimantes grand format (par exemple 60 x 90 cm ou 24 x 36 pouces) — même si les imprimantes laser couleur sont susceptibles de les remplacer. Elles produisent des cartes de grande qualité à 600 ppp. Les vitesses d'impression sont encore relativement lentes pour des imprimantes à jet d'encre.

6.114. Lors du choix d'une imprimante appropriée pour un projet de SIG, le coût est un élément essentiel. Il faut avoir présent à l'esprit que le coût d'achat d'une imprimante n'est qu'un élément du coût — souvent relativement secondaire. Les prix des imprimantes ont fortement baissé, mais le coût des cartouches d'encre et du papier spécial est resté plutôt élevé. Dans certains cas, on constate que les prix du matériel sont maintenus très bas par les fabricants d'imprimantes qui espèrent tirer surtout profit de la vente des fournitures correspondant au matériel. Outre le prix d'achat, les coûts d'impression par page standard devraient être aussi comparés (par exemple, lorsque 5 % de la page sont couverts d'encre). Les magazines consacrés au commerce informatique publient souvent des comparaisons.

3. Impression commerciale

6.115. Pour de grands tirages, les imprimantes personnelles sont trop lentes et les coûts par page trop élevés. Les brochures, affiches ou atlas de recensement seront, par conséquent, imprimés dans une imprimerie interne ou une imprimerie commerciale. Si le volume à imprimer est élevé, les procédés analogiques d'impression, dans lesquels des planches d'impression sont préparées et utilisées dans des machines d'imprimerie lithographiques ou comparables, sont actuellement encore moins coûteux et plus rapides que les procédés d'impression numériques.

6.116. Cependant, le processus d'impression jusqu'à l'établissement des planches d'impression est déjà presque intégralement numérique. Le procédé d'impression type pour un atlas de recensement numérique ressemblera à la figure ci-dessous (voir

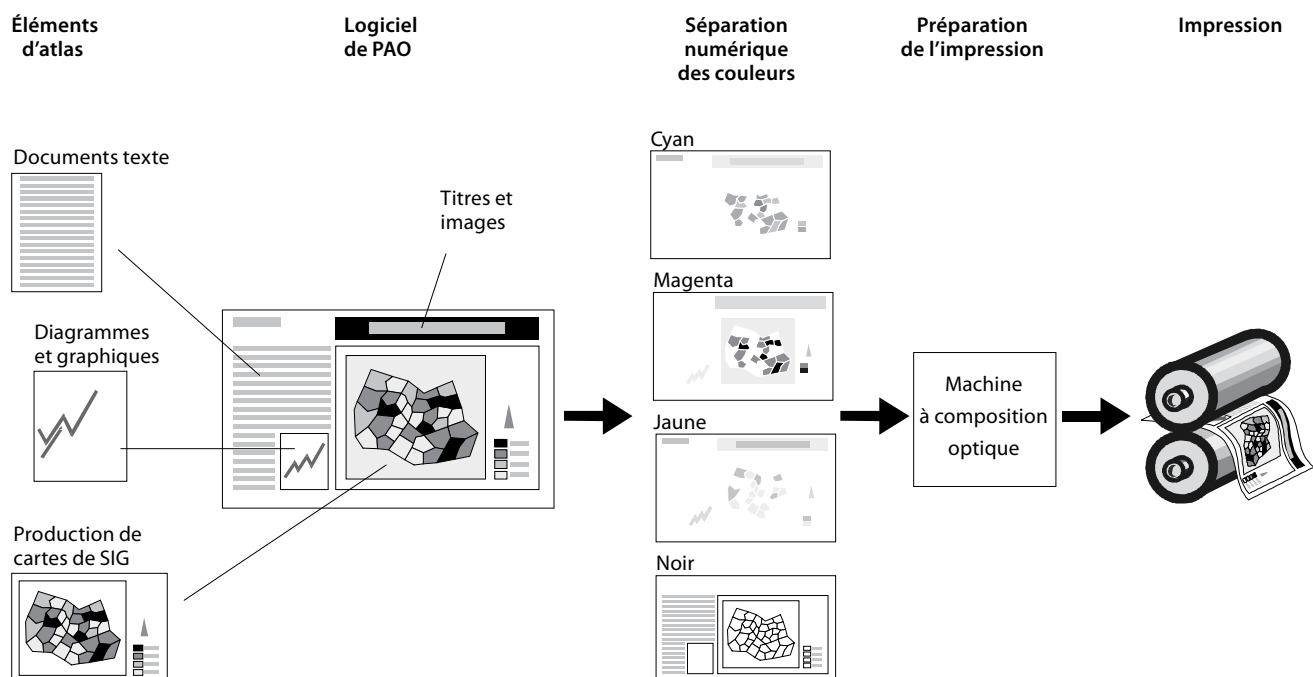
figure VI.9). Après une étape initiale de préparation au cours de laquelle le texte, les graphiques et le contenu cartographique sont précisés, les cartographes du service de recensement produisent toutes les cartes à inclure dans l'atlas. Ces cartes sont enregistrées en format PostScript prêt à imprimer. Pour des cartes complexes comportant des graphiques produits dans des logiciels externes ou des photographies, les tracés peuvent être établis au moyen d'un logiciel graphique de haut de gamme. D'autres collaborateurs du bureau de recensement écriront le texte destiné à accompagner les cartes, les tableaux, les références et autres documents textuels à l'aide de logiciels standard de traitement de texte.

6.117. Au cours d'une seconde étape, tous les éléments d'atlas sont combinés dans un programme de publication assistée par ordinateur. Les en-têtes de texte, les légendes des figures, les images, le texte et les éléments graphiques sont formatés et arrangés dans un tracé visuellement attrayant correspondant exactement à la dimension des pages du produit imprimé. Ce travail peut être effectué en interne ou par un bureau de service externe. La décision de sous-traiter ou non le travail nécessite de prendre en compte des critères tels que la qualité, le coût, le nombre de copies requises ainsi que le temps nécessaire.

6.118. Une fois le tracé final de l'atlas établi, il est enregistré dans un dossier de sortie numérique. Le format de fichier le plus courant est un fichier PostScript intégré, mais certains formats de fichiers spécifiques de tel ou tel logiciel peuvent aussi être utilisés par les imprimantes commerciales. La plupart des programmes graphiques et de publication assistée par ordinateur de haut de gamme peuvent aussi fournir des sélections de couleurs qui sont soit stockées dans des fichiers séparés, soit réunies dans un même fichier. L'imprimante proprement dite utilise quelques planches d'impression, une pour chacune des couleurs cyan, magenta, jaune et noir (modèle CMJN). Les couleurs sur les cartes et les documents graphiques sont produites en tant que

Figure VI.9

Le processus d'impression numérique



combinaisons additives de divers pourcentages de ces quatre couleurs. Les fichiers numériques sont ensuite envoyés à la machine à composition optique qui crée le film à partir duquel les planches d'impression sont produites. C'est en utilisant des fichiers numériques pour produire le film qu'on obtient en général les meilleurs résultats. Des exemplaires prêts à être photographiés imprimés sur une imprimante laser, reproduits par des techniques photographiques, pourront être moins chers, mais ils ne donneront pas la même résolution. À moins qu'une ligne de production ait déjà été établie et essayée, il est généralement souhaitable d'obtenir et d'évaluer une épreuve de contrôle en couleurs provenant de l'imprimante avant la production définitive. De nombreux vendeurs de matériel et de logiciel fournissent aussi des informations détaillées et d'autres ressources sur leurs sites Web.

E. Données géographiques numériques destinées à être utilisées

6.119. Les logiciels de plus en plus sophistiqués et le recours accru à l'Internet dans le monde ont permis la diffusion des données géographiques numériques. La demande de bases de données numériques constituées d'extractions de la base maîtresse de données géographiques numériques de l'agence de recensement va continuer à s'accroître. Les données de recensement sont un intrant important dans la planification des politiques et l'analyse théorique dans de nombreux domaines. Les fournitures de services sanitaires, l'affectation des ressources à l'éducation, la conception des infrastructures et services collectifs et la planification électorale sont quelques applications pour lesquelles les organismes publics ont besoin de statistiques de population sur de petites zones à référence spatiale. Les utilisateurs commerciaux utilisent ces données pour des applications de commercialisation et pour prendre des décisions d'implantation.

6.120. Compte tenu de l'amélioration du traitement des données et des SIG, l'élaboration d'une base de données géographique numérique au niveau des districts de recensement ou à des niveaux désaggrégés similaires est de plus en plus à la portée de nombreux organismes statistiques nationaux. Pour satisfaire aux besoins des utilisateurs, l'organisme national de statistique devrait appliquer une stratégie de diffusion des données numériques. Il est recommandé que les organismes nationaux de statistique considèrent les coûts et avantages de la fourniture de données à un niveau aussi détaillé.

6.121. Au rang des avantages procurés, citons un niveau de détail et de précision inégalé, l'utilisation potentielle de données de recensement dans de nombreuses applications, en particulier lorsqu'elles sont superposées à d'autres données géographiques, telles que les données de terrain, et la facilité relative de gestion et de stockage de milliers d'unités.

6.122. Quant aux coûts, ils sont liés notamment aux dépenses consenties pour le traitement et la gestion des données, aux éventuelles questions de divulgation des données et au contrôle de la qualité. La publication des résultats de recensement au niveau de ce qui constitue une unité de collecte des données peut occasionner des erreurs de processus imprévues. La diffusion des données géographiques doit toujours s'accompagner de métadonnées et les coûts de leur production sont à prendre en compte dans l'équation.

6.123. Il y a une alternative pour publier une base de données de diffusion des districts de recensement : créer un produit dérivé à une échelle tout aussi détaillée.

1. Stratégies de diffusion des données numériques pour atteindre des utilisateurs potentiels

6.124. Comme indiqué plus haut, il est recommandé que l'organisme national de statistique considère les coûts et avantages d'une diffusion plus vaste des données lorsqu'il planifie ses produits et services. Les utilisateurs potentiels de données de recensement relatives à de petites zones sont très divers, ce qui signifie que l'organisme national de statistique doit suivre une stratégie de diffusion de données numériques à plusieurs niveaux. Dans l'ensemble, on peut distinguer les divers utilisateurs suivants :

- Les **utilisateurs évolués de SIG** peuvent travailler aisément avec de grands ensembles de données et utiliser le FTP (protocole de transfert de fichiers) pour y accéder. Ils ont besoin de métadonnées exhaustives. Les utilisateurs évolués combinent souvent les données censitaires d'une petite zone avec leurs propres données de SIG sur les installations sanitaires, les circonscriptions scolaires ou les secteurs de vente par exemple. Ces utilisateurs sont parfois appelés « extracteurs de données » ou « utilisateurs importants ». Ils souhaiteront accéder à des attributs et des informations spatiales dans un format géographique numérique complet. Le bureau de recensement doit fournir des documents détaillés (voir chapitre III. F sur l'élaboration de métadonnées) sur les paramètres géographiques utilisés pour établir la base de données géographique, ainsi que sur diverses variables de recensement. Les informations spatiales seront diffusées dans un format géographique ouvert, qui pourra facilement être converti dans les formats commerciaux de SIG les plus divers.
- Les **utilisateurs possédant des notions d'informatique** dans l'administration, le secteur commercial ou le secteur privé, souhaiteraient pouvoir parcourir spatialement les informations thématiques dans une base de données de recensement. Ils souhaiteront aussi produire des cartes thématiques et doivent donc pouvoir effectuer des manipulations simples de paramètres cartographiques. Des opérations analytiques simples, telles que l'agrégation des unités de recensement dans des régions établies sur mesure devraient aussi être possibles. La meilleure prise en charge pour ce groupe d'utilisateurs consistera en un programme d'application complet, prêt à l'emploi, conçu pour un progiciel de cartographie de bureau commercial ou gratuit. Les besoins de documentation sont moindres, car il y a peu de chances que les utilisateurs modifient les paramètres géographiques de la base de données ou qu'ils effectuent des opérations plus avancées au moyen du SIG.
- Les **nouveaux utilisateurs** souhaitent pour la plupart visualiser des cartes préparées à l'avance sur un ordinateur et peut-être effectuer des demandes d'information de base. Pour ce troisième groupe d'utilisateurs, la meilleure stratégie de diffusion de données consistera à établir un atlas numérique de recensement autonome. Cet atlas pourra consister en une série d'images cartographiques statiques, par exemple sous la forme d'une projection de diapositives. Ou ce pourrait être une simple interface cartographique avec des vues de cartes préétablies permettant d'effectuer des demandes d'informations de base. Les cartes statiques et une interface cartographique simple pourront être mises à disposition par l'intermédiaire d'Internet.

a) Définition du contenu des données

6.125. La première étape de l'élaboration des bases de données géographiques destinées à une large diffusion consiste à définir le contenu des données à diffuser.

L'organisme national de statistique doit planifier la publication d'ensembles de données séparés, et pas les données opérationnelles utilisées à des fins de gestion interne. En effet, en maintenant la séparation entre les deux types de données, l'organisme national de statistique pourra constamment assurer la maintenance et la mise à niveau de ses données internes. Plusieurs questions se posent avant de diffuser des ensembles de données, comme indiqué ci-dessous.

i) *Jusqu'à quel niveau les données seront-elles diffusées ?*

6.126. Pour développer au maximum les avantages d'ensemble procurés par la collecte de données censitaires, l'organisme de recensement devra se fixer pour objectif de diffuser des données de recensement à référence géographique au niveau le plus bas ne portant pas atteinte à la confidentialité des données ou à la validité statistique. Même au niveau du district de recensement, il peut y avoir des zones déclarantes particulières qui ne comportent que quelques ménages et pour lesquelles les données de recensement ne peuvent être diffusées. Si nécessaire, les données d'un certain nombre de zones déclarantes devront être supprimées ou recodées.

ii) *Une grande base de données de SIG ou un ensemble de bases de données de recensement ?*

6.127. Une base de données censitaires de SIG à haute résolution comportera plusieurs milliers d'unités déclarantes. De tels volumes de données dépassent la capacité de calcul de l'utilisateur de données moyen. L'organisme national de statistique doit en outre tenir compte de la taille physique des fichiers de données ainsi que de l'impact des téléchargements de fichiers sur les serveurs. Ainsi, télécharger un fichier de données de 3 gigaoctets en utilisant un serveur de 56 kilooctets prend environ trois minutes.

6.128. Au lieu de diffuser une grande base de données, l'organisme de recensement devrait envisager de créer une famille de bases de données censitaires. À un niveau de résolution moyen — par exemple celui des districts — une base de données nationale résumée peut donner un aperçu suffisamment détaillé de la situation socioéconomique du pays. Pour chaque grande circonscription administrative ou même chaque district, des bases de données présentant des indicateurs au niveau du sous-district et du district de recensement pourront être créées. Des bases de données séparées peuvent aussi être utiles pour les grandes zones urbaines.

6.129. Enfin, une base de données ponctuelle des habitats du pays avec les données de recensement correspondantes répondra aux besoins des utilisateurs qui n'ont pas besoin de la résolution spatiale d'une base de données d'unités déclarantes de SIG. Cette base de données devrait au minimum contenir tous les habitats classés dans la catégorie urbaine et les indicateurs agrégés de recensement pour chaque ville grande et moyenne. Idéalement, il sera aussi préférable de développer une base de données au niveau du village à l'intention des planificateurs travaillant dans les secteurs de la santé, de l'éducation ou de l'agriculture. Une base de données de villages pourra être fondée sur un répertoire toponymique de noms et d'emplacements de lieux si ces informations ont été collectées pendant les opérations de terrain préalables au recensement.

6.130. La mise à disposition de bases de données relatives à des subdivisions du pays encouragera l'utilisation des données. De nombreux utilisateurs n'ont besoin de renseignements censitaires que pour une région relativement restreinte. Un sous-ensemble de la base de données censitaires nationale est plus facile à prendre en charge

par des utilisateurs possédant des notions limitées de la mise en œuvre des SIG. De même, dans les pays où les coûts d'accès aux données sont supérieurs au coût de leur reproduction, de plus petits ensembles de données seront à la portée d'un plus grand nombre d'utilisateurs non commerciaux.

6.131. Si l'on diffuse des bases de données séparées, on devra veiller à ce que les partis ou blocs individuels soient compatibles. Cela signifie que les limites partagées entre des sous-ensembles de base de données doivent se raccorder exactement. Des éléments séparés de la base de données devraient avoir le même système de référence géographique et les mêmes définitions de base de données d'attributs. Si la base de données maîtresse utilisée par le bureau de recensement est très détaillée, il pourra être utile pour certains utilisateurs de données de disposer également d'une version plus généralisée de cartes numériques de recensement. Certains pays proposent des cartes numériques de recensement à des échelles cartographiques nominales différentes ou avec une précision différente des coordonnées. Le tarif peut être plus élevé pour les utilisateurs qui demandent un degré élevé d'exactitude et de détail.

6.132. De nombreux producteurs commerciaux de données de SIG diffusent leurs données avec les coordonnées de latitude/longitude (c'est-à-dire les coordonnées géographiques) plutôt que dans une projection spécifique. Les coordonnées géographiques sont le système de référence le plus courant et elles peuvent facilement être converties dans d'autres systèmes de projection, si l'utilisateur veut utiliser les limites de recensement en combinaison avec d'autres couches de données. Par contre, le logiciel de SIG risque de ne pas prendre en charge des projections et des systèmes de coordonnées nationaux spécifiques. Les utilisateurs auront alors des difficultés à employer la base de données censitaires dans des applications d'analyse géographique.

iii) *Dans quelle mesure doit-on intégrer étroitement les limites et la base de données ?*

6.133. Les bases de données censitaires de SIG sont caractérisées par le grand nombre de leurs champs d'attributs. Les questionnaires de recensement fournissent des informations qui peuvent être enregistrées dans des centaines de champs variables. En général, il n'est pas commode d'enregistrer tous les champs dans le même tableau de données. Il est préférable d'enregistrer un petit nombre d'indicateurs parmi les plus importants dans le tableau des attributs géographiques et de reprendre les informations restantes dans une série de tableaux séparés. Ces tableaux extérieurs pourront être organisés par thème — données démographiques, données sur les ménages, etc. L'utilisateur pourra ensuite relier les tableaux au SIG, au besoin, en se servant de l'identificateur géographique commun. Il est capital d'insister une fois encore sur la nécessité d'un codage cohérent pour garantir le caractère unique des diverses entités géographiques définies.

iv) *Quelle quantité de métadonnées doit être fournie ?*

6.134. S'il est primordial de produire des métadonnées et des documents en tant que compléments à toute diffusion des données, le volume d'informations nécessaires variera en fonction de l'utilisateur. Il est important de faire la distinction entre les utilisateurs internes, qui ont normalement besoin de métadonnées exhaustives, et les utilisateurs externes pour qui des données moins détaillées suffisent en général. Voir la dernière rubrique du présent chapitre ainsi que les rubriques précédentes (pour en savoir plus sur les métadonnées, voir le chapitre II et la section F du chapitre III, plus haut, et les paragraphes 6.137 à 6.139 ci-dessous).

b) Conventions relatives aux noms de fichiers

6.135. Bien que les systèmes d'exploitation Windows, Macintosh, UNIX, et LINUX acceptent tous des noms de fichiers de grande longueur, il est souhaitable d'appliquer les conventions relatives aux noms de fichier de DOS 8.3 (à savoir 8 caractères) pour tous les fichiers de données et de documentation à diffuser, et de développer des normes internes pour assurer la cohérence. Certains utilisateurs peuvent travailler sous DOS ou Windows 3.1 ou des progiciels de SIG plus anciens. Les noms de fichiers courts peuvent réduire au minimum les incompatibilités, par exemple avec un logiciel de réseau ancien. Grâce aux conventions cohérentes en matière de noms de fichiers expliquées dans la documentation, les utilisateurs pourront rapidement trouver les données dont ils ont besoin.

c) Compression

6.136. Les fichiers de SIG sont souvent très imposants et couplés aux données sous forme de tableaux, l'ensemble des fichiers à distribuer peut devenir très volumineux. La compression des fichiers facilitera beaucoup la diffusion des données, en particulier pour la fourniture de données par Internet ou leur diffusion sur disquette ou CD-ROM. Les logiciels de compression les plus largement utilisés dans l'environnement Windows sont les programmes utilitaires PKZIP et Winzip. Ils sont disponibles sur la plupart des ordinateurs et des utilitaires permettant d'extraire des fichiers d'archives comprimées existent aussi pour le système d'exploitation UNIX. Il convient toutefois de noter qu'il est impossible de comprimer certains formats, dont le File Geodatabase. Les fichiers autoextractibles sont plus commodes pour les utilisateurs inexpérimentés et ils n'exigent pas d'utilitaire de décompression. Cependant, ils sont propres à un système d'exploitation et ne devraient être utilisés que si l'on connaît la plate-forme matérielle cible.

d) Documentation et dictionnaires de données

6.137. La documentation qui accompagne la diffusion d'un lot de données ne doit pas être aussi complète que les informations internes compilées pour toutes les bases de données (voir chapitres II-V ci-dessus). Les fichiers texte simples en ASCII (« lisez-moi ») peuvent être lus par n'importe quel utilisateur. Les utilisateurs de données n'ont pas en général besoin d'informations détaillées sur l'origine des données ou les étapes du traitement et la facilité d'interprétation est plus importante pour les utilisateurs extérieurs. Ainsi, la documentation devrait contenir une description claire, concise et complète des aspects de la base de données pertinents pour un utilisateur. À condition que le bureau de recensement tienne à jour une base de métadonnées complète, la documentation relative aux données ciblée sur l'utilisateur peut être compilée très rapidement. La documentation relative aux données peut comporter les informations suivantes :

- Noms de lots de données et information de référence, y compris toutes les sources de données.
- Contenu descriptif des lots de données.
- Description de la hiérarchie des unités administratives et des unités déclarantes et de leurs rapports avec d'autres entités (par exemple des habitats). Elle devrait inclure une description claire des définitions statistiques utilisées pour chaque type d'unité déclarante. Il sera utile d'établir une liste complète de toutes les unités déclarantes et de leurs codes géographiques.

- Besoins de matériel et de logiciel.
- Directives générales concernant le format des données, la décompression et l'installation.
- Informations à référence géographique (tous les lots de données géographiques devraient être dans le même système de référence) :
 - "Ω Projection cartographique avec tous les paramètres requis, tels que parallèle sécant ou méridien standard, abscisse et ordonnée fictives, etc.
 - "Ω Unités de coordonnées (par exemple, degrés décimaux, mètres, pieds).
 - "Ω Échelle de la carte source (c.-à-d. échelle des cartes imprimées à partir desquelles les limites ont été numérisées).
 - "Ω Informations relatives à la précision géographique. Par exemple, on pourra communiquer toute information disponible sur la précision numérique des cartes sources. S'il n'est pas possible d'établir une évaluation quantitative de la qualité des données, la précision pourra être décrite en termes plus généraux.
 - "Ω Des cartes imprimées des lots de données de SIG constituent un complément utile de la documentation. Par exemple, elles permettent à l'utilisateur de vérifier que l'importation des cartes a été effectuée correctement.
- Conventions concernant le traitement d'unités déclarantes non consécutives.
- Informations relatives à des produits connexes, par exemple des bases de données censitaires de SIG plus détaillées ou des fichiers de données additionnels qui peuvent être utilisés avec les limites.
- Bibliographie des publications pertinentes relatives aux recensements.
- Informations relatives aux personnes à contacter pour aider les utilisateurs.
- Notes concernant la responsabilité, informations sur le droit d'auteur, etc.

6.138. En outre, chaque lot de données de SIG devrait être accompagné d'un dictionnaire de données fournissant des informations sur chaque couche de données ou tableau de données du SIG. L'expression « dictionnaire de données » est plus ancienne et fait référence au formatage spécifique et aux noms de champs dans le lot de données. Ce dictionnaire devrait fournir les informations suivantes :

- Noms et formats de fichiers.
- Types d'objets (points, lignes ou polygones).
- Relation entre les fichiers de données concernant les coordonnées et les tableaux associés de données attributaires extérieures.
- Pour chaque champ du tableau des attributs et dans les tableaux extérieurs supplémentaires :
 - "Ω Nom du champ.
 - "Ω Description du contenu du champ (par exemple, population totale en 1995) et la définition statistique exacte utilisée. Pour les indicateurs démographiques dérivés, la formule utilisée peut être fournie, par exemple, en utilisant les noms de champ des variables employées comme numérateur et dénominateur.
 - "Ω Définitions de champs, y compris le type de variable (par exemple champ de nombres réels, entiers ou de caractères), l'intervalle de valeurs admissibles et les conventions concernant le traitement des valeurs manquantes. Pour les données classifiées, le système de codage doit être expliqué

en détail. Par exemple, dans une base de données d'habitats, un champ numérique appelé TYPE peut utiliser le « 1 » pour la capitale nationale, le « 2 » pour les chefs-lieux de province, le « 3 » pour les centres administratifs de district, etc.

”Ω Toute information disponible sur la qualité des données qui permet aux utilisateurs d'évaluer l'adéquation des données à une tâche donnée.

6.139. La documentation relative aux données et les dictionnaires de données peuvent être inclus dans un guide détaillé de l'utilisateur. Un tel guide peut contenir une explication plus détaillée du contenu de la base de données, de l'origine et de la qualité des données. Des explications progressives d'applications types ou des exemplaires de cartes de recensement créés avec la base de données peuvent aussi être inclus. Un dictionnaire de données échantillons est présenté à l'annexe IV.

e) Contrôle et assurance de la qualité des produits de données livrables

6.140. Comme indiqué plus haut, le contrôle de la qualité est une opération importante précédant la mise en circulation du produit finale en vue de sa reproduction. Après l'établissement de la version définitive de toutes les bases de données sous la forme dans laquelle elle sera distribuée (par exemple comprimée), la base de données doit être essayée sur toutes les plates-formes cibles (par exemple environnement Windows, UNIX, Macintosh et LINUX).

6.141. Pour de nombreux utilisateurs, le CD-ROM reste le support le plus approprié pour la diffusion de grands ensembles de données. Un CD-ROM peut contenir jusqu'à 700 mégaoctets et la plupart des ordinateurs sont équipés de lecteurs de CD-ROM. Les graveurs de CD sont également peu coûteux, ce qui permet au service de graver ses disques originaux numériques. Cela permet aussi de diffuser des lots de données personnalisés pour lesquels quelques exemplaires suffisent. Pour une diffusion plus large de grands ensembles de données, le CD-ROM présente plusieurs avantages : faible coût de production unitaire, durabilité et lisibilité sur de multiples plates-formes matérielles.

6.142. La technique d'enregistrement de DVD (vidéodisque numérique/disque numérique polyvalent) a remplacé le CD-ROM dans certains domaines et les graveurs de DVD sont de plus en plus répandus sur les ordinateurs de bureau. Actuellement, un DVD standard une face peut contenir 4,7 gigaoctets de données.

6.143. À plus long terme, la diffusion de données s'effectuera pour l'essentiel via Internet. Actuellement, la largeur de bande — déterminant la quantité de données pouvant être transférée dans un intervalle de temps donné — est limitée, ce qui freine la diffusion des très grands fichiers dans bon nombre de pays. Les temps de téléchargement sont souvent inacceptables du fait des insuffisances de l'infrastructure d'Internet dans de nombreux pays. Mais le principal goulet d'étranglement est la connexion par modem aux câbles principaux d'Internet depuis le bureau ou le domicile. De grands fichiers peuvent être envoyés à des utilisateurs universitaires, administratifs ou commerciaux disposant d'un accès spécialisé à grande vitesse à Internet.

6.144. La diffusion de données par Internet élimine dans une grande mesure le coût de reproduction pour l'organisme de recensement. Les coûts restants concernent la mise au point de l'interface de logiciel, la tenue à jour du site Web et l'utilisation incrémentielle des ressources du serveur Web. Les bases de données géographiques pour le recensement peuvent ainsi être communiquées à l'utilisateur à un coût très faible ou même gratuitement. Cependant, certains organismes peuvent décider de facturer les données en ligne et cela pour subventionner ainsi un programme de publication

destiné aux utilisateurs non connectés à Internet ou encore pour recouvrer en partie le coût de la collecte et de la compilation des données de recensement.

2. Problèmes juridiques et questions de commercialisation

6.145. Les questions que les organismes nationaux de statistique doivent considérer en matière de production et de propriété des données géographiques sont abordées ci-après. Il s'agit notamment des droits d'auteur sur les données, des options concernant la commercialisation des données géographiques et des questions de responsabilité.

a) Droit d'auteur sur les données

6.146. Le droit d'auteur est le droit exclusif et garanti par la loi de publier, reproduire ou vendre une œuvre — dans ce contexte, une base de données géographique numérique. Les données numériques étant faciles à reproduire, les questions de droit d'auteur concernant les bases de données géographiques ont un caractère plus urgent que dans le cas des cartes sur papier. Le bureau de recensement a donc besoin d'élaborer une politique d'accès aux données pour les informations sous forme de tableaux comme pour les informations cartographiques de recensement.

6.147. La question du droit d'auteur comporte deux aspects : les droits moraux et les droits matériels. Les droits moraux protègent l'intégrité du travail en interdisant toute altération du produit originel. Les droits matériels concernent le droit à une rétribution monétaire lorsque le produit a été mis en circulation en vue de sa reproduction, de son utilisation ou de sa transformation. Tout droit accordé par le titulaire du droit d'auteur doit être précisé dans un accord de licence.

6.148. La question du droit d'auteur est liée à la politique de fixation des prix des produits à base de données numériques. Un organisme de recensement a le choix entre plusieurs options pour adopter une stratégie de fixation de prix des données spatiales numériques. L'organisme peut décider :

- De supporter la totalité du coût de la collecte et de la diffusion des données.
- De facturer le coût de diffusion des données (coût du support et de l'expédition).
- De recouvrer tout ou partie du coût de la collecte et de la compilation des données.
- De s'assurer des recettes dépassant le coût effectif de l'investissement consacré au SIG et de l'élaboration des données.

b) Options concernant la commercialisation des données géographiques

6.149. Les législations relatives au droit d'auteur varient d'un pays à l'autre. Dans un cas extrême, certains gouvernements n'exercent pas de droit d'auteur sur les informations produites par des organismes publics. La raison en est que les contribuables ont déjà eu à payer la collecte des données et qu'ils ne doivent pas être taxés à nouveau pour l'utilisation des données. Par conséquent, les données géographiques produites par des organismes publics sont diffusées gratuitement ou au coût de reproduction. D'autre part, n'importe quelle entreprise commerciale peut utiliser les informations publiques, les reconditionner et les vendre avec bénéfice.

6.150. Aux États-Unis, par exemple, l'accès gratuit aux données publiques a donné naissance à une importante industrie de services qui produit des données de

recensement à référence spatiale dans divers formats en vue de les vendre aux utilisateurs privés, commerciaux et publics. Bien que les entreprises fassent payer les données, l'utilisation non exclusive des données censitaires a attiré sur ce marché de nombreuses entreprises. Cette concurrence a maintenu bas le prix des données de recensement reconditionnées, tout en élargissant la gamme des produits spécialisés. Les utilisateurs qui veulent procéder eux-mêmes à la conversion des données ont toujours accès aux données gratuites.

6.151. Cette évolution a eu l'avantage d'élargir dans de grandes proportions l'utilisation des données de recensement dans des applications géographiques. L'augmentation du nombre des utilisateurs a encouragé à son tour la mise au point commerciale de logiciels de cartographie sur ordinateur faciles d'emploi et la fourniture de services à valeur ajoutée. Les avantages économiques d'ensemble de cette évolution sont élevés. En effet, les recettes fiscales ont augmenté et l'accès plus facile aux informations s'est traduit par des gains de productivité et de meilleures prises de décisions dans les secteurs public et privé. Ces avantages ont justifié la diffusion de données sans redevance, qui a été essentiellement une subvention publique aux entreprises privées.

6.152. Dans d'autres pays, les dépenses budgétaires étant comprimées, les organismes publics ont été de plus en plus incités à dégager des recettes en vue de financer leurs opérations. En conséquence, les prix des informations de recensement à référence géographique sont parfois très élevés, limitant ainsi l'usage des informations. Ces prix peuvent refléter par exemple la valeur commerciale de ces données pour les institutions financières et les entreprises commerciales. Cependant, de tels prix peuvent exclure les petites entreprises et les utilisateurs non commerciaux du marché des informations censitaires et ils peuvent limiter l'utilisation générale et par conséquent les avantages procurés par les données censitaires géographiques. Comme Prévost et Gilruth (1997) le font remarquer, les tentatives de récupération des coûts qui mettent les produits censitaires géographiques hors de portée pour les utilisateurs non commerciaux entraînent souvent le copiage illégal de lots de données, la duplication de l'élaboration des données à partir des documents sources initiaux, qui prend un temps considérable, ou l'utilisation de données de remplacement moins coûteuses et de moindre qualité.

6.153. Des contrats de licence restrictifs empêchent ou freinent la distribution de produits de recensement et de services dérivés, ce qui réduit l'avantage social procuré par la collecte des données de recensement. Cette réduction des effets économiques généraux due à l'absence de retombées risque d'être plus importante que l'augmentation des recettes pour l'organisme de recensement. En fait, dans certains pays les politiques de diffusion de données produites par les organismes publics reprennent une approche de gratuité ou de faible coût car on s'est rendu compte que les avantages d'une facturation plus élevée ne compensaient pas le coût de l'application du droit d'auteur ou les avantages sociaux perdus en raison d'une utilisation réduite d'informations essentielles.

6.154. Par ailleurs, l'accès aux données et les utilisations secondaires sont souvent restreints lorsque le bureau de recensement collabore avec un producteur de données privé ou lorsque des données provenant de producteurs publics ou privés sont utilisées pour produire des cartes de recensement. Par exemple, l'agence de recensement peut passer un contrat avec une entreprise cartographique privée qui prend en charge une partie du coût de la production de cartes numériques pour le recensement. L'entreprise ne pourra récupérer son investissement que s'il lui est accordé un droit exclusif de commercialisation des données géographiques (cela ne pose naturellement

aucun problème lorsque l'agence se contente d'acheter les services de l'entreprise, tous les produits demeurant la propriété de l'organisme de recensement).

6.155. Si des données fournies par d'autres organismes — par exemple le service cartographique national ou des autorités locales — sont utilisées pour produire des cartes de recensement, il faudra apporter des précisions sur la facturation, le droit d'auteur, la définition de la source et l'indication de la provenance figurant sur les cartes de recensement. Il faudra particulièrement éviter les différends relatifs aux droits d'auteur car l'organisme de recensement devra probablement encore collaborer à l'avenir avec ces autres organismes dans d'autres activités cartographiques relatives au recensement.

6.156. Dans la plupart des pays, l'arbitrage entre un accès le plus large possible aux données de recensement et les pressions exercées pour récupérer en partie les coûts de la collecte de données aboutit en général à un compromis entre les deux positions extrêmes qui viennent d'être décrites. Par exemple, des arrangements particuliers peuvent être conclus entre des organismes publics qui souhaitent insérer réciproquement leurs données dans leurs produits. L'organisme de recensement peut conclure des contrats avec le service cartographique national en vue de distribuer aux utilisateurs de données censitaires de SIG des cartes numériques de base représentant les routes, les cours d'eau, etc. De même, des réductions peuvent être accordées aux utilisateurs universitaires et aux autres utilisateurs non commerciaux. Une autre possibilité peut consister à fournir des produits génériques gratuitement, tout en facturant les produits à valeur ajoutée exigeant un traitement plus poussé.

c) Questions de responsabilité

6.157. Dans plusieurs cas les tribunaux ont estimé que les producteurs de données pouvaient être tenus pour responsables lorsque des erreurs d'information géographique provoquaient des accidents ou d'autres dommages. Jusqu'ici la plupart des affaires concernaient des accidents résultant de l'absence d'informations ou de la présence d'informations erronées sur les cartes topographiques. Des cas dans lesquels des accidents d'avions ou des accidents en mer ont été causés par des informations erronées sur les cartes de navigation ont été présentés. La conception des cartes et leur contenu informatif dépendent de l'utilisation prévue, mais les cartes sont parfois utilisées à des fins que le producteur de données n'avait pas prévues. Par exemple, un organisme de recensement peut publier des données sur des unités déclarantes avec une base de données du réseau de voirie. Les informations relatives aux routes n'étant pas essentielles pour l'utilisation des données de recensement, le contrôle de la qualité de ces informations peut très bien avoir été moins rigoureux que si les informations relatives aux routes avaient été compilées pour un service d'acheminement des services d'urgence. Si les données imparfaites sont utilisées à de telles fins non prévues à l'origine, des dommages peuvent très bien survenir.

6.158. Un autre exemple relatif aux questions de responsabilité qui concerne particulièrement la diffusion des données de recensement est celui de la violation de la confidentialité des informations. En général, un organisme de recensement publie seulement des données agrégées à un niveau qui ne révèle pas les informations relatives à un individu, un ménage ou un très petit groupe de personnes. Si l'organisme de recensement réagrège les microdonnées concernant la géographie de plusieurs petites zones — par exemple, des districts de recensement, des secteurs postaux, des districts sanitaires ou des divisions administratives dans le domaine de l'éducation — il peut en utilisant habilement les SIG isoler des informations concernant des groupes de personnes plus restreints que ce que prévoit le niveau de divulgation le plus bas (voir

par. 6.32 à 6.36 ci-dessus sur le problème du calcul de différences). Dans certains pays, cela peut justifier des poursuites intentées par les intéressés.

6.159. On notera que d'après Johnson et Onsrud (1995), le fait de vendre des données de SIG et de limiter les utilisations secondaires des données peut accroître la responsabilité du fournisseur de données. La rémunération sous-entend que le fournisseur de données garantit que le document est exempt d'erreur et adapté aux buts poursuivis. Par contre le fait de rendre les données publiques pourrait mettre l'organisme à l'abri de telles réclamations.

6.160. Avant de diffuser des données à référence spatiale, l'agence devrait donc consulter des experts juridiques et rédiger une note accompagnant les produits établis à partir des données. Cet avertissement pourrait inclure les points suivants :

- Une déclaration selon laquelle on estime que l'information était exacte au moment de la collecte et qu'elle a été obtenue à partir de sources fiables, sans qu'il soit possible d'en garantir l'exactitude.
- Des avertissements selon lesquels les informations sont susceptibles de varier et la notification des modifications effectives.
- Si des éléments de la base de données géographique ont été créés par un organisme extérieur, on doit l'indiquer clairement.
- On doit mentionner que l'utilisation des données sous-entend l'acceptation des avertissements et des conditions fixées.

3. Cartographie sur Internet

6.161. De nombreux organismes nationaux de statistique ont adopté l'Internet comme moyen pour diffuser des informations et des données. Les pages Web peuvent aller des simples listes et tableaux de résultats de recensement à des interfaces d'interrogation élaborées, dans lesquelles l'utilisateur peut demander des tabulations croisées spécialisées.

6.162. La diffusion des données par l'Internet doit être encouragée pour de nombreuses raisons. Cependant, le choix d'une telle diffusion représente un investissement pour le bureau de statistique national, en particulier compte tenu des contraintes imposées aux serveurs, pour le stockage, la mise à jour et les sauvegardes. Les organismes nationaux de statistique sont invités à étudier les coûts totaux du cycle de vie d'un tel investissement dans le cadre de leur planification censitaire globale.

6.163. L'Internet est totalement approprié pour présenter et diffuser des informations géographiques. L'option la plus simple consiste à présenter des images cartographiques statiques établies par le bureau de recensement. Par exemple, une série de cartes montrant des variables de recensement peut être produite en utilisant un logiciel cartographique de bureau. La plupart des logiciels permettent à l'utilisateur d'enregistrer des cartes dans un format d'image standard telles que GIF ou JPEG. Ces images peuvent ensuite être intégrées à des pages Web comme n'importe quel autre document graphique ou photographie. Ces sites Web peuvent fournir aux utilisateurs de données un accès à des informations utiles. Cependant, ils ne permettent pas à l'utilisateur de manipuler les données et de produire des cartes personnalisées pour certaines zones géographiques. Les sections ci-après mettront l'accent sur les méthodes qui autorisent un degré d'interaction élevé entre l'utilisateur et la base de données géographique de recensement.

6.164. La plupart des entreprises productrices de logiciels de SIG et de cartographie sur ordinateur ont mis au point des outils indépendants de la plate-forme utilisée

pour la cartographie sur Internet, qui recourent à des protocoles courants d'échange de données. Ces outils permettent à l'organisme statistique d'installer des informations géographiques sur un serveur et à l'utilisateur de cartographier et d'interroger ces données de façon interactive en utilisant des logiciels courants de lecture sur Internet. Les utilisateurs d'Internet peuvent ainsi accéder aux applications des SIG sans avoir à acheter de logiciel exclusif de SIG. Toutes les données qui peuvent être stockées ou manipulées avec un SIG peuvent aussi être distribuées de cette façon — y compris les cartes vectorielles, les images matricielles et les tableaux de données.

6.165. Les logiciels de cartographie sur Internet sont également utiles comme outil utilisé en interne pour mettre les données spatiales à la disposition du personnel du bureau de statistique sur un intranet. Plutôt que d'acheter des permis d'utilisation sur site de logiciels commerciaux de SIG exploités à partir d'un serveur central, les membres du personnel peuvent accéder aux informations géographiques avec leur logiciel de lecture.

6.166. L'architecture orientée service est un terme général faisant référence à des méthodes qui incluent des approches côté serveur, côté client et hybrides (chacun de ces concepts est décrit plus en détail ci-après) :

- Dans les stratégies côté serveur, l'utilisateur adresse une requête de carte au serveur détenant la base de données. Le logiciel cartographique du serveur traite la requête, produit une carte — par exemple au format GIF — et la renvoie à l'utilisateur.
- Dans les stratégies côté client, au contraire, la plupart des tâches de traitement sont exécutées localement sur l'ordinateur de l'utilisateur (client).
- Enfin les approches hybrides combinent les approches côté serveur et côté client.

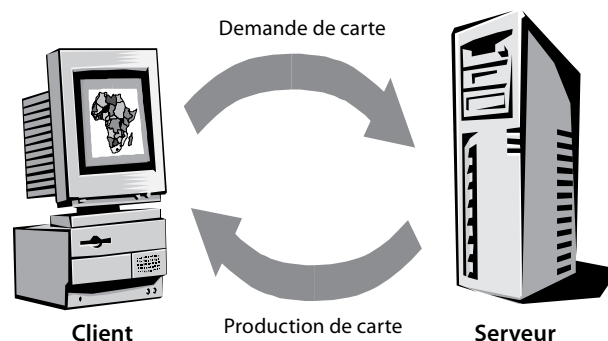
a) Approches côté serveur

6.167. Parfois qualifiées d'architecture de « client léger/gros serveur », ces stratégies confient au serveur situé dans les locaux de l'organisme distributeur de données le plus gros du travail de traitement de données. Cette architecture ressemble à l'architecture traditionnelle fondée sur un ordinateur central, dans laquelle un ordinateur central puissant effectue la gestion, le stockage et le traitement des données pour un certain nombre d'utilisateurs connectés par des terminaux passifs.

6.168. Le principe de la stratégie côté serveur est résumé sur la figure VI.10. L'utilisateur se connecte à un site Web et entre une demande de carte. Les spécifications définies par l'utilisateur pour la carte à fournir comprennent la région qui l'inté-

Figure VI.10

Cartographie par Internet — l'approche côté serveur



resse — spécifiée soit par le nom de la région, par exemple le nom du district, soit par les coordonnées formant le rectangle qui l'entoure — la variable pour laquelle la carte doit être établie, la classification et le système de couleurs et les couches de données supplémentaires qui forment le contexte, par exemple les routes, les cours d'eau ou les limites administratives.

6.169. La demande de l'utilisateur est adressée par Internet au serveur et acheminée jusqu'à un logiciel de SIG. Ce logiciel peut être situé soit dans le serveur du Web, soit sur un ordinateur distinct connecté au serveur. Le logiciel de SIG peut être soit un logiciel commercial de cartographie par Internet, soit un logiciel de cartographie sur Internet personnalisé et fondé sur des modules de logiciel cartographique vendus dans le commerce. Le logiciel cartographique accède aux bases de données demandées, établit la carte et renvoie ce produit à l'utilisateur sous la forme d'une page Web. Les cartes sont envoyées en général sous la forme d'images graphiques ordinaires au format GIF ou JPEG, car les logiciels de lecture du Web ne peuvent prendre en charge les formats de données vectorielles. Si l'utilisateur souhaite modifier le tracé de la carte, il adresse une nouvelle requête au serveur.

6.170. L'approche côté serveur présente plusieurs avantages : l'utilisateur n'a pas besoin d'un ordinateur puissant pour accéder à des bases de données géographiques qui peuvent être assez importantes. Même des opérations de SIG assez complexes comme l'appariement d'adresses ou l'acheminement sur le réseau peuvent être effectuées rapidement si l'on dispose d'un serveur puissant. L'utilisateur n'a besoin que d'un simple logiciel de lecture sur Internet et d'une connexion Internet. L'intégrité des données est préservée, car l'utilisateur ne peut manipuler la base de données elle-même. L'utilisateur est donc toujours assuré d'avoir accès aux informations les plus récentes. Le fournisseur de données exerce un contrôle plus étroit sur ce que les utilisateurs peuvent visionner et comment ils le peuvent. Les choix de dessin cartographique peuvent être préétablis pour assurer que même des utilisateurs non experts obtiennent des produits cartographiques acceptables. Cette approche a notamment pour inconvénient la charge du trafic en réseau sur les serveurs occupés.

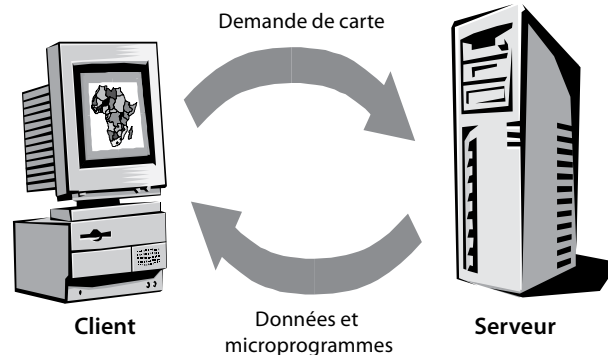
b) Approches côté client

6.171. Les approches côté client — architectures de gros client — transfèrent une grande partie du traitement demandé à l'ordinateur de l'utilisateur. Le serveur est utilisé surtout pour stocker la base de données et envoyer à l'utilisateur les éléments demandés de la base de données, éventuellement avec des modules cartographiques. Deux variantes de l'approche côté client sont disponibles.

6.172. Dans la première, l'ordinateur de l'utilisateur ne dispose d'aucune capacité cartographique. Une fois présentée la demande de l'utilisateur, le serveur envoie les données géographiques ainsi qu'un programme court ou miniprogramme qui permet l'établissement de cartes ou l'analyse géographique (figure VI.11). Un miniprogramme est un élément de logiciel indépendant de toute plate-forme, écrit en langage de programmation Java, qui peut être exécuté par les logiciels courants de lecture sur le Web. L'utilisateur peut ensuite travailler avec les données indépendamment du serveur. Le survol des couches de cartes ou le changement du tracé cartographique n'exige pas d'adresser de nouvelles requêtes au serveur.

6.173. Selon une autre approche côté client, un logiciel cartographique, un miniprogramme ou un module d'extension pour la lecture est installé de façon permanente sur l'ordinateur de l'utilisateur. Un module d'extension est un programme qui élargit les capacités du logiciel de lecture sur Internet, par exemple, pour lui permettre d'afficher des fichiers d'un certain format. Cette approche présente l'avantage que le

Figure VI.11

Cartographie par Internet — l'approche côté client

logiciel cartographique n'a pas besoin d'être téléchargé chaque fois que l'utilisateur accède au serveur cartographique.

6.174. Les données et les programmes une fois téléchargés, l'utilisateur n'a pas besoin de communiquer davantage avec le serveur de cartes. La cartographie ou l'analyse peut être effectuée en mode déconnecté. Les ressources de l'ordinateur de l'utilisateur peuvent être utilisées, ce qui en général accélère le traitement. Les approches côté client peuvent donner à l'utilisateur plus de souplesse et de liberté pour analyser et afficher les données spatiales. Cependant, les fichiers de données et de programmes peuvent être très importants, exigeant une connexion rapide à Internet. Les utilisateurs disposant d'ordinateurs moins puissants risquent de ne pas pouvoir exécuter des tâches de cartographie et d'analyse plus complexes. Les approches côté client peuvent permettre aux utilisateurs de sauvegarder sur leur ordinateur les données géographiques brutes demandées au serveur. Cela sera difficile si tout ou partie des données géographiques enregistrées sur le serveur du bureau de recensement sont protégées par un droit d'auteur.

c) **Approches hybrides**

6.175. Les approches côté serveur permettent à un cercle d'utilisateurs non experts d'accéder facilement à des cartes relativement simples. Elles seront donc particulièrement appropriées pour un bureau de recensement qui souhaite présenter au grand public des cartes de recensement. Inversement, les stratégies côté client sont préférables pour les intranets où un petit nombre d'utilisateurs possédant des connaissances assez complètes des SIG et de la cartographie accède à des bases de données complexes. Elles permettront donc au personnel du bureau de recensement d'accéder aux données de SIG à l'intérieur du service.

6.176. Les approches hybrides combinent les avantages des stratégies côté serveur et côté client. Elles offrent à l'utilisateur une grande souplesse pour interroger et manipuler les cartes localement, mais transfèrent au serveur l'essentiel du lourd travail d'analyse. Cela exige un certain degré de communication entre le client et le serveur concernant la puissance de traitement disponible.

d) **Possibilités en matière de diffusion des données de recensement, y compris le serveur cartographique**

6.177. Les logiciels cartographiques actuellement disponibles sur Internet sont modulables. Les fournisseurs de données peuvent acheter un logiciel standard qui fonctionne avec des lots de données standard. L'établissement de cartes de données

de recensement étant une application assez courante, les organismes nationaux de statistique ne devraient pas avoir de difficultés à trouver une solution appropriée. Pour les applications plus complexes, on pourra se procurer une boîte à outils de modules logiciels permettant au fournisseur de données de définir sur mesure l'interface du serveur cartographique.

6.178. Le concept de MapServer a été conçu comme un environnement de développement libre d'applications Internet dotées de fonctions spatiales. Mapserver est la technologie de serveur en ligne libre la plus utilisée à l'heure actuelle (voir www.opengeospatial.org) et l'un des principaux éléments nécessaires pour traduire les données de SIG en une image de carte qui sera visualisée par un client Web. En effet, MapServer prend en charge les normes de l'Open GIS Consortium (OGC) et se caractérise par des produits cartographiques avancés, dont l'automatisation de la cartographie thématique et des éléments de la carte, le dessin d'objets basé sur l'échelle et l'étiquetage d'objets, une production personnalisable et basée sur un modèle, la prise en charge de scripts populaires et des environnements de développement, une aide à la projection de cartes, en plus d'une multitude de formats matriciels et vectoriels (pour en savoir plus, voir <http://mapserver.gis.umn.edu>).

6.179. Avec l'accroissement des capacités du réseau, des lots de données et des modules de programmes plus importants peuvent être envoyés aux utilisateurs et un plus grand nombre d'utilisateurs peuvent être servis simultanément. Les problèmes inhérents aux solutions côté client et côté serveur devraient trouver une solution grâce à des connexions plus rapide à Internet. Les ordinateurs clients pourront établir avec les serveurs des communications fréquentes, sans retards, ce qui se traduira par l'exécution quasi instantanée des demandes des utilisateurs.

6.180. Pour les données de recensement, la meilleure stratégie d'accès aux données et de diffusion des données par Internet dépendra des capacités et des connaissances de l'utilisateur. Un système souple fournira des services à l'utilisateur à n'importe quel niveau, même pour les utilisateurs importants, actifs et passifs. Ces catégories cadrent globalement avec la typologie des utilisateurs décrits plus haut :

- Les **utilisateurs importants** qui souhaitent obtenir la totalité de la base de données pour l'utiliser sur leur propre ordinateur au moyen d'un logiciel commercial de SIG. Ces utilisateurs sont desservis par les méthodes classiques de distribution de données, l'achat de CD-ROM, DVD ou le téléchargement via Internet de lots « bruts » de données censitaires de SIG.
- Les **utilisateurs actifs** ayant une certaine connaissance des SIG, mais qui ne disposent pas de capacité locale de SIG. Ces utilisateurs désirent télécharger des éléments de la base de données avec des modules de programmes de SIG (miniprogrammes) qui peuvent effectuer les tâches demandées.
- Les **utilisateurs passifs** qui veulent seulement obtenir une carte prédéterminée. La demande de l'utilisateur est exécutée par le serveur et les résultats obtenus sont envoyés à l'utilisateur via Internet dans un format approprié — par exemple des fichiers d'images matricielles ou des fichiers PostScript pour les cartes et des fichiers de tableurs ou bases de données pour les données.

6.181. Un système souple de diffusion de données de recensement sur Internet pourrait prendre la forme suivante :

- Les utilisateurs déterminent l'étendue de la région qui les intéresse en vue de télécharger les données ou simplement de demander une carte. La région géographique en question peut être spécifiée en utilisant l'une des adresses géographiques suivantes :

- ”Ω Le nom de la région géographique, par exemple une ville, un district ou une province.
- ”Ω Un rectangle limite déterminé par des coordonnées géographiques.
- ”Ω Une région spécifiée de façon interactive par l'utilisateur au moyen des fonctions de lecture et d'agrandissement. Par exemple, l'interface peut commencer par une carte du pays. L'utilisateur peut ensuite agrandir une région qui l'intéresse et sélectionner la zone géographique en traçant un rectangle ou un polygone sur l'écran. Lorsque l'utilisateur agrandit l'image, des détails plus nombreux apparaissent sur l'interface cartographique. Au début, la carte montre seulement les limites du pays et des provinces. Lorsque l'utilisateur agrandit une province, les limites de district et les emplacements des villes apparaissent. S'il choisit une ville, les rues principales et des limites des districts de recensement urbain apparaissent. Le niveau de détail dépend de l'échelle de la carte, correspondant à l'étendue de la carte actuellement affichée sur l'écran de l'utilisateur.
- ”Ω Une région qui est définie par une requête géographique. Par exemple, un utilisateur commercial qui souhaite obtenir des informations sur les caractéristiques démographiques de clients potentiels pourra demander des informations démographiques pour une zone circulaire de cinq kilomètres de rayon autour de l'emplacement d'un centre commercial. Un organisme public de planification pourra demander des données sur la population résidant à moins de cinq kilomètres du tracé proposé d'une autoroute.

6.182. L'utilisateur spécifie les variables qui l'intéressent et le type de produit souhaité. Les options peuvent inclure des cartes pour lesquelles l'utilisateur peut spécifier des présentations cartographiques de base comme le nombre de catégories, le type de classification et les couleurs d'ombrage. Le produit souhaité peut aussi être un simple tableau de données montrant les variables choisies pour la région en question. L'utilisateur précise aussi s'il a besoin d'une base de données et de modules d'interrogations et d'analyse géographique ou s'il souhaite avoir une carte ou les résultats tirés d'une base de données.

6.183. Le serveur de base de données interprète la demande de l'utilisateur et crée le sous-ensemble approprié de la base de données. Pour les régions définies au moyen de noms géographiques, cela nécessite seulement un choix logique, par exemple de tous les districts de recensement dans une zone donnée. Pour les zones demandées qui ne correspondent pas à la hiérarchie géographique normalement utilisée pour le recensement, des opérations de traitement supplémentaires sont requises. Dans certains pays, des bases de données géospatiales pour zones d'habitation sont maintenant disponibles ou en construction, chaque immeuble à usage d'habitation étant associé à une coordonnée géographique. Un SIG sur le serveur peut alors compiler un tableau sur mesure en sélectionnant tous les ménages qui résident dans la zone géographique définie par l'utilisateur. Dans les cas où cela n'est pas possible, la base de données géospatiale basée sur le serveur doit procéder à une interpolation surfacique au moyen de techniques telles que celles décrites plus haut.

6.184. Le résultat de la demande est renvoyé à l'utilisateur soit sous forme de données de base qui peuvent être encore manipulées par l'utilisateur en utilisant de miniprogrammes de SIG, soit sous la forme d'une carte ou d'un relevé de base de données qui peut être utilisé directement par l'utilisateur. Naturellement, outre la base de données ou les cartes, des documents relatifs aux données et d'autres informations pertinentes doivent être également disponibles.

6.185. En fonction des politiques de diffusion de données du pays, ces services peuvent être gratuits ou rétribués. Les réponses fournies aux demandes d'informations essentielles déjà compilées pourront être gratuites mais des requêtes plus complexes pourront être facturées.

6.186. On doit particulièrement prendre en considération la confidentialité des données si les tableaux individualisés sont établis à partir de microdonnées. Les questions de sécurité sur Internet sont aussi importantes pour la gestion des données de recensement sur les réseaux qu'elles le sont pour les applications commerciales d'Internet. Le réseau interne qui peut fournir un accès aux microdonnées de recensement doit donc être séparé par un pare-feu du domaine d'Internet qui permet aux utilisateurs extérieurs d'accéder à des données de recensement agrégées.

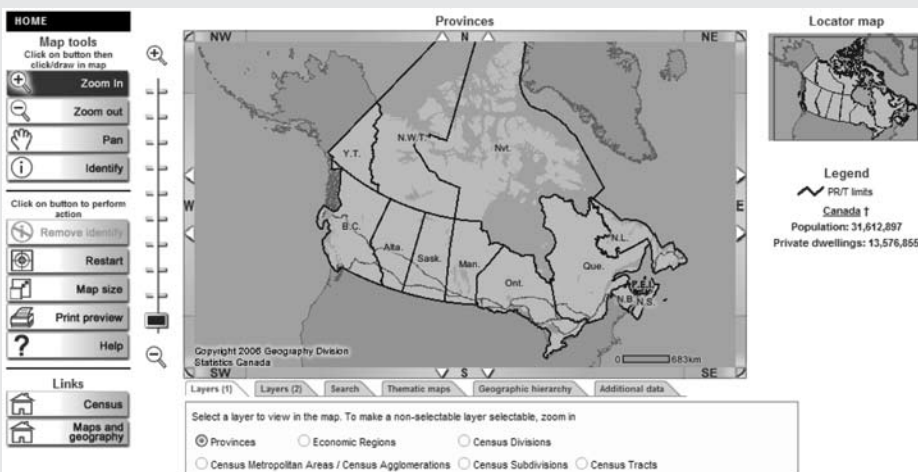
6.187. L'interface envisagée pour la diffusion des données paraît très ambitieuse. Elle exige des connexions rapides à Internet et ne permettra d'atteindre beaucoup d'utilisateurs que si les ménages, les entreprises et les organismes publics ont largement accès à Internet. Dans de nombreux pays ces conditions ne sont pas encore réunies, mais vu le développement rapide de la technologie, beaucoup pourront répondre à la plupart des demandes de données via Internet dans un proche avenir. Certains organismes de recensement appliquent activement des stratégies de diffusion de données qui englobent certains éléments que l'on vient de décrire.

Encadré VI.1

Étude de cas de la cartographie Web pour la diffusion de données de recensement : Canada

Depuis la fin des années 1990, l'Internet est le principal support de diffusion privilégié par Statistics Canada. La diffusion en ligne recèle un énorme potentiel, en offrant à l'utilisateur une grande souplesse de consultation des données statistiques. La clé pour exploiter ce potentiel ? Mettre à disposition des outils permettant à l'utilisateur de parcourir aisément les données existantes. Statistics Canada a mis au point un outil de cartographie en ligne grâce auquel l'utilisateur peut explorer des données de recensement géographiquement à travers une interaction avec une carte. Grâce à GeoSearch, l'utilisateur a le loisir de rechercher, d'identifier, de visualiser et d'obtenir des données démographiques et géographiques de base spécifiques à certaines régions du Canada.

GeoSearch2006



Afin de trouver un lieu spécifique au Canada, les utilisateurs peuvent interagir avec un ensemble commun d'outils cartographiques à gauche de l'écran permettant d'agrandir, de réduire, de se déplacer autour de la carte, voire d'identifier la région qui les intéresse. Une série d'onglets sous la fenêtre de la carte donne accès à diverses fonctions, telles que la sélection d'un type standard de zone géographique, la recherche d'une zone d'intérêt à partir de plusieurs types d'information (lieu, nom, adresse ou code postal), la cartographie thématique de données de recensement, l'identification des relations hiérarchiques géographiques ou l'accès à des données complémentaires. Une fois une zone identifiée, les utilisateurs peuvent afficher les dénombrements de population et d'habitations, choisir parmi plus de 35 variables de recensement pour créer des cartes thématiques dynamiques, faire le lien avec plus de 10.000 cartes de référence au format PDF ou consulter une kyrielle de données démographiques ou socioéconomiques des profils communautaires de 2006.

Outre GeoSearch, Statistics Canada développe toute une série de produits de données géographiques censitaires disponibles sur son site Web. Ces produits sont répartis en trois grandes catégories. La gamme de produits liés aux cartes (format PDF) propose notamment des cartes de référence illustrant la superficie de zones géographiques standard ainsi que des cartes thématiques affichant des statistiques relatives à plusieurs thèmes. La gamme de produits sur les informations spatiales englobe les fichiers de limites pour toutes les zones géographiques standard, le fichier du réseau routier et le fichier des données hydrographiques. À l'aide de ces produits spatiaux, les utilisateurs sont en mesure de développer leurs propres bases de données géospatiales à des fins d'analyse ou d'intégrer les zones géographiques de recensement à leur base de données géospatiale existante. La gamme des produits se rapportant aux données d'attributs fournit des données sur les caractéristiques des zones géographiques (ex. : dénombrement de population et d'habitations, superficie et densité démographique) ainsi que les relations entre des zones géographiques. Ces produits peuvent être utilisés indépendamment de ou conjointement avec des produits de données spatiales (pour en savoir plus, contacter Joe Kresovic—Joe.Kresovic@statcan.gc.ca).

F. Résumé et conclusions

6.188. Le chapitre VI a passé à la loupe les tâches géographiques post-recensement, dont la diffusion des données. Une fois que les agents de terrain ont fait part des modifications à effectuer sur la base géographique, un processus d'édition continue de la base de données géographique est engagé en perspective des levés et recensements futurs. Les données peuvent être diffusées au niveau d'un district de recensement modifié ou à travers l'agrégation en de nouvelles unités de diffusion, telles que des grappes. Plusieurs produits et services statiques et dynamiques peuvent être développés dans le sillage du recensement, dont des cartes, rapports, atlas, CD-ROM et sites Web interactifs.

6.189. La mise en circulation de données au niveau désagrégé soulève une série de questions quant à la divulgation et à la confidentialité des données. Elle suscite en outre une autre question : les données doivent-elles être fournies gracieusement ou faut-il en tirer des revenus en vendant des produits à valeur ajoutée, tels que des CD-ROM ou des DVD de données ? Le présent chapitre a également proposé une description des produits de données géographiques, tels que des solutions de visualisation de cartes ou des fichiers de données spatiales attribués destinés aux progiciels commerciaux de SIG ou aux produits cartographiques en ligne.

Bibliographie et références

- Aronoff, S. (1991). *Geographic Information Systems: A Management Perspective*. Ottawa : WDL Publications.
- ASCE (1994). *The Glossary of the Mapping Sciences*. Bethesda, Maryland : American Society for Photogrammetry and Remote Sensing, and American Society for Civil Engineers.
- Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. Madison, Wisconsin : University of Wisconsin Press. Original en français : *Sémiologie Graphique*. Paris, 1977.
- Brewer, C. (1994). Colour use guidelines for mapping and visualization. In A. M. MacEachren et D. R. F. Taylor, eds. *Visualization in Modern Cartography*. Londres : Pergamon.
- _____ (2005). *Designing Better Maps: A Guide for GIS Users*. Redlands, California : ESRI Press.
- Bugayevskiy, L. M., et J. P. Snyder (1995). *Map Projections : A Reference Manual*. Londres : Taylor and Francis.
- Canter, F., et H. Deleir (1989). *The World in Perspective: A Directory of World Map Projections*. New York : John Wiley and Sons.
- Castro, M., et al. (2004). Integrated urban malaria control: a case study in Dar es Salaam, Tanzania. *American Journal of Medical Hygiene*, 103 à 117.
- Commission économique pour l'Afrique (2007). Détermination des ensembles de données géographiques fondamentales pour l'Afrique : Géoinformation pour le développement économique et social. Document ECA/ISTD/GEO/2007/02E. Addis-Abeba.
- Dent, B. D. (1999). *Cartography: Thematic Map Design*, 5^e édition. Dubuque, Iowa : Wm. C. Brown Publishers.
- Duke-Williams, O., et Rees, P. H. (1998). Can census offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. *International Journal of Geographical Information Science*, vol. 12, n° 6.
- États-Unis d'Amérique, Census Bureau (1978). *Mapping for censuses and surveys*, Statistical Training Document ISP-TR-3. Washington, D.C. : United States Department of Commerce, Bureau of the Census.
- États-Unis d'Amérique, Federal Geographic Data Committee (1997a). *Framework Introduction and Guide*. Washington, D.C.
- _____ (1997b). The subcommittee on cultural and demographic data. Washington, D.C. (www.census.gov/geo/www/standards/scdd/index.html).
- États-Unis d'Amérique, National Center for Health Statistics (1997). *Atlas of United States Mortality*. Washington, D.C. : Center for Disease Control and Prevention.

- États-Unis d'Amérique, National Research Council (2007). *Tools and Methods for Estimating Populations at Risk from Natural Disasters and Complex Humanitarian Crises*. Washington, D.C. : National Academy Press.
- Fothergill, S., et J. Vincent (1985). *The State of the Nation: An Atlas of Britain in the Eighties*. Londres : Pan Books.
- Garson, G. D. (2003). *Public Information Technology: Policy and Management Issues*. Hershey, Pennsylvanie : Idea Publishing Group.
- Hohl, P., ed. (1998). *GIS Data Conversion: Strategies, Techniques, Management*. Santa Fe, Nouveau Mexique : Onword Press.
- Hurskainen, P., et P. Pellikka. (2006). Change detection of informal settlements using multi-temporal aerial photographs: the case of Voi, SE Kenya. Proceedings of the Fifth Conference of the African Association of Remote Sensing of the Environment. Helsinki.
- Jensen, J. R. (1996). *Introductory Digital Image Processing: A Remote Sensing Perspective*, 2^e édition. New York : Prentice-Hall.
- _____, et D. Cowen (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, vol. 65, n° 5, p. 611 à 622.
- Johnson, J., et H. J. Onsrud (1995). Is Cost Recovery Worthwhile? *Proceedings of the Annual Conference of the URISA*, San Antonio, Texas, juillet.
- Khosrow-Pour, M. (2005). *Practicing e-Government: A Global Perspective*. Hershey, Pennsylvanie : Idea Publishing Group.
- Kraak, M. J., et F. J. Ormeling (1997). *Cartography: Visualization of Spatial Data*. Harlow, Essex, Royaume-Uni : Longman.
- Krygier, J., et D. Wood. (2005). *Making Maps: A Visual Guide to Map Design for GIS*. New York : Guilford Press.
- Longley, P., M. Goodchild, D. Maguire et D. Rhind. (2005). *Geographic Information Systems and Science*. New York : John Wiley and Sons.
- Lynch, M., et K. E. Foote (1997). Legal issues relating to GIS: the geographer's craft project. Austin : University of Texas.
- MacEachren, A. M. (1994). *Some Truth with Maps: A Primer on Symbolization and Design*. Washington, D.C. : Association of American Geographers.
- _____. (1995). *How Maps Work: Representation, Visualization and Design*. New York : Guilford Press.
- McDonnell, R., et K. Kemp (1995). *International GIS Dictionary*. Cambridge, Royaume-Uni: GeoInformation International.
- Monmonier, M. (1993). *Mapping It Out. Expository Cartography for the Humanities and Social Sciences*. Chicago : University of Chicago Press.
- Montana, L., et J. Spencer (2004). *Incorporating Geographic Information into Measure Surveys: A Field Guide to GPS Data Collection*. Chapel Hill, Caroline du Nord : Carolina Population Center of the University of North Carolina.
- Nations Unies (2007). *Geographical Information Systems for Population Statistics*. Numéro de vente : E.97.XVII.30.
- _____. (2008). *Principles and Recommendations for Population and Housing Censuses, Revision 2*, publication des Nations Unies, numéro de vente : E.07.XVII.8.

- NIDI (Netherlands Interdisciplinary Demographic Institute) (1996). *Proceedings of the Expert Group Meeting on Innovative Techniques for Population Censuses and Large-Scale Demographic Surveys*, 22-26 avril 1996. La Haye : Institut démographique interdisciplinaire des Pays-Bas et Fonds des Nations Unies pour la population.
- Niebert, D. D., ed. (2004). *Developing Spatial Data Infrastructures: The SDI Cookbook*. Global Spatial Data Infrastructure.
- O'Looney, J. A. (2002). *Wiring Governments: Challenges and Possibilities for Public Managers*. Westport, Connecticut : Quorum Books.
- Padmanabhan, G., J. Yoon et M. Leipnik (1992). *A Glossary of GIS Terminology*, Technical Report, n° 92-13. Santa Barbara, Californie : National Center for Geographic Information and Analysis.
- Prévost, Y., et P. Gilruth (1997). *Environmental Information Systems in Sub-Saharan Africa*. Building Blocks for Africa 2025, Paper n° 12. Banque mondiale, Washington, D.C. et PNUD/BNUS, New York.
- Rhind, D., ed. (1997). *Framework for the World*. Cambridge, Royaume-Uni : GeoInformation International.
- Robinson, A. H., J. L. Morrison, P. C. Muehrcke, A. J. Kimerling et S. C. Gupitill (1995). *Elements of Cartography*, 6^e édition. New York : John Wiley and Sons.
- Snyder, J. P. (1982). *Map Projections Used by the United States Geological Survey*, Washington, D.C. : Government Printing Office.
- _____ (1993). *Flattening the Earth: Two Thousand Years of Map Projections*. Chicago : University of Chicago Press.
- Tomlinson, R. (2007). *Thinking about GIS: Geographic Information System Planning for Managers*. Redlands, Californie : ESRI Press.
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, Connecticut : Graphics Press.
- _____ (1990). *Envisioning Information*. Cheshire, Connecticut : Graphics Press.
- Weeks, J. (2007). Can we spot a neighbourhood from the air? Defining neighbourhood structure in Accra, Ghana. *GeoJournal*, 69:9-22.
- White, J. D. (2007). *Managing Information in the Public Sector*. Armonk, New York : M. E. Sharpe.
- Yankson, P., et al. (2004). A flexible model for urban vegetation cover measurement based on remote-sensing images. Chine : Université de Wuhan.

Annexe I

Systèmes d'information géographique

A. Introduction aux systèmes d'information géographique

A1.1. Un système d'information géographique (SIG) est un outil informatique permettant d'entrer, de stocker, de gérer, de rechercher, de mettre à jour, d'analyser et de produire des informations. Les informations contenues dans un SIG concernent les caractéristiques d'objets ou de zones géographiques. En d'autres termes, un SIG permet de répondre à des questions concernant le lieu où se trouvent les objets ou l'objet situé en un certain lieu.

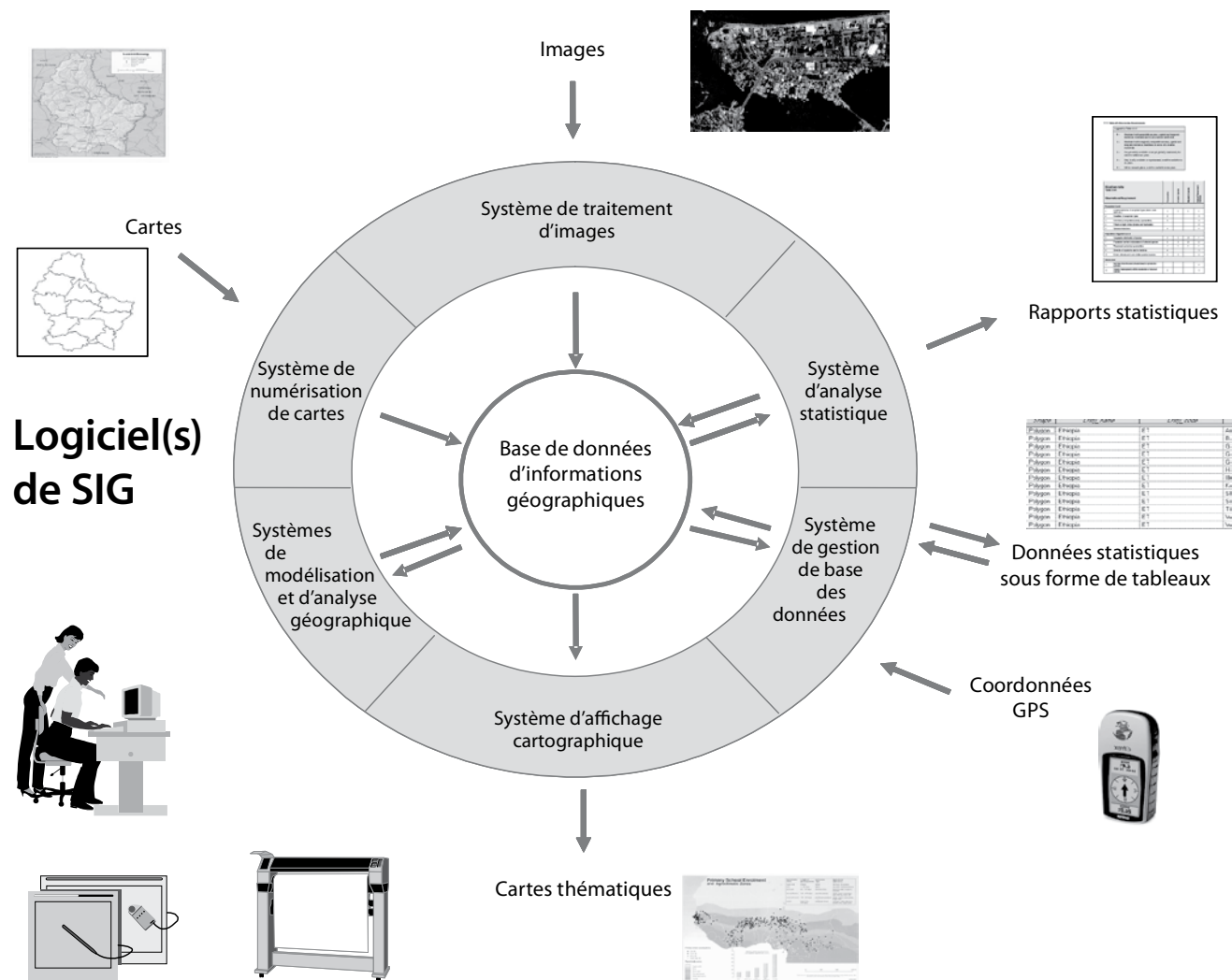
A1.2. Le terme « SIG » a différentes significations dans différents contextes. Il peut désigner dans son ensemble le système de matériel et de logiciel utilisé pour travailler sur des informations spatiales. Il peut se référer à un logiciel particulier conçu pour traiter des informations relatives à des objets géographiques. Il peut concerner une application, par exemple, une base de données géographique d'ensemble concernant un pays ou une région. Enfin, il sert parfois à décrire le secteur d'étude consacré aux méthodes, aux algorithmes et aux procédures utilisés pour travailler avec des données géographiques. Par exemple, il existe aujourd'hui des cours consacrés aux SIG dans plusieurs universités. Le terme de « science de l'information géographique » est de plus en plus utilisé pour désigner la recherche universitaire relative aux logiciels et procédures géographiques mis en œuvre au moyen de l'ordinateur. La technologie géospatiale est un terme connexe plus large englobant les SIG, la télédétection et les systèmes de positionnement universel (GPS).

A1.3. Plusieurs secteurs ont contribué à jeter les bases du SIG, comme l'indique la figure A1.1. Les traditions géodésiques et cartographiques ont fourni les règles et les instruments utilisés pour mesurer et représenter des objets du monde réel. L'informatique offre le cadre nécessaire pour stocker et gérer les informations géographiques et, avec les mathématiques, fournit les instruments qui permettent de manipuler les objets géométriques représentant des entités géographiques du monde réel. Alimenté en données par les enquêtes socioéconomiques et environnementales et par les levés topographiques, le SIG prend en charge des applications intéressantes des secteurs très variés; celles-ci vont des applications plutôt théoriques, comme l'archéologie ou l'océanographie à des applications commerciales pratiques, y compris la commercialisation ou le secteur de la gestion immobilière.

A1.4. D'autres applications du type inventaire intéressent le secteur des services collectifs, lorsque, par exemple, une compagnie de téléphone gère et entretient son infrastructure physique à l'aide d'une base de données géographique. Les systèmes d'enregistrement des titres fonciers gérés par des organismes publics locaux et régionaux en offrent un autre exemple. Dans certains domaines, le SIG sert à faciliter la collecte de données. L'utilisation de la cartographie numérique pour les opérations de recensement et de diffusion de données est naturellement l'exemple le plus pertinent dans le

contexte du présent manuel. On trouve d'autres applications analytiques dans le secteur universitaire ainsi que dans de nombreux domaines pratiques tels que la gestion des ressources naturelles ou la commercialisation. Les entreprises forestières, par exemple, utilisent des SIG pour optimiser l'exploitation durable des arbres et les entreprises de commercialisation ou de distribution emploient des techniques perfectionnées d'analyse spatiale pour cibler la clientèle ou implanter une nouvelle installation.

Figure A1.1
Bases du SIG



Source : Ebener/OMS

1. Matériel, logiciel et données

A1.5. Les questions de matériel et de logiciel sont examinées dans le présent *Manuel* dans le contexte de l'élaboration et de la maintenance de bases de données géographiques pour les opérations de recensement. En général, le matériel nécessaire n'est pas différent de celui utilisé dans d'autres applications utilisant des documents graphiques qui se caractérisent par de grands volumes de données : un ordinateur personnel ou une station de travail haut de gamme, un moniteur de grandes dimensions à haute

résolution et les périphériques d'entrée habituels, le clavier et la souris. On utilise des tables à numériser ou un scanner pour convertir les cartes sur papier en bases de données numériques. Ces outils sont aussi utilisés par les architectes ou les concepteurs graphistes. On se sert également de numériseurs et d'imprimantes de bureau grand format pour obtenir des produits cartographiques destinés à l'affichage et à l'analyse visuelle.

A1.6. Les logiciels de SIG ont évolué rapidement ces dernières années des systèmes commandés ligne par ligne et difficiles à apprendre aux progiciels faciles d'emploi pouvant être utilisés par tous avec un minimum de formation. Les progiciels haut de gamme sont employés par les analystes de SIG qui créent de nouvelles bases de données et effectuent des opérations avancées d'analyse spatiale. À un niveau intermédiaire, il existe aujourd'hui de nombreux logiciels de cartographie sur ordinateur associant une interface Windows courante à de nombreuses prestations en matière d'entrée, de gestion, d'analyse et de sortie de données. Enfin, on trouve en bas de gamme des logiciels de lecture de données géographiques. Ils ne permettent pas à l'utilisateur de modifier les données, mais assurent de nombreuses fonctions d'affichage. Ces logiciels, dont certains sont distribués gratuitement, sont un excellent moyen de diffusion de données.

A1.7. Récemment sont apparues des boîtes à outils de scripts ou de routines logicielles commercialisés par plusieurs vendeurs de SIG. Ils permettent à l'utilisateur de construire des applications cartographiques sur mesure dans des environnements de programmation orientés objet de qualité industrielle. Il peut s'agir de systèmes autonomes ou ils peuvent être intégrés à des progiciels. Certains de ces produits comportent aussi les outils nécessaires pour mettre au point des applications cartographiques sur Internet.

A1.8. Les tendances actuelles en matière de logiciels de SIG paraissent orientées dans deux directions : la cartographie sur Internet et la conception modulaire qui permet d'intégrer des fonctions de SIG dans n'importe quelle application. Les utilisateurs pourront bientôt demander et analyser des données sur des bases de données géospatiales éloignées en utilisant leur logiciel de lecture sur le Web et des logiciels téléchargés à la demande. Pour des applications haut de gamme, il peut exister une convergence plus poussée entre les SIG et les systèmes de gestion de bases de données relationnelles. De même que les logiciels de SIG utilisent des systèmes de gestion de bases de données relationnelles pour stocker et manipuler des attributs, certains systèmes de gestion de bases de données comportent déjà des fonctions permettant de stocker et de manipuler des objets géographiques. La distinction entre les SIG et les autres systèmes d'information pourrait ainsi s'estomper graduellement.

A1.9. Les données alimentent les applications de SIG (voir figure A1.2). De nombreux ensembles de données de SIG parmi les plus courants sont des équivalents numériques de cartes sur papier telles que des cartes topographiques montrant les routes, les cours d'eau, les courbes de niveau et les habitats sous forme de couches distinctes. Les informations thématiques comprennent les attributs socioéconomiques référencés par unités administratives, des cartes décodées montrant le couvert végétal ou l'utilisation des sols et des indicateurs dérivés tels que les limites de bassins hydrographiques. Tout objet géographique figurant sur une carte numérique peut être décrit de façon très détaillée sur un tableau de données lié à la base de données spatiales numérique. Parfois, quelques attributs suffiront pour caractériser un ensemble d'objets. Dans d'autres cas, par exemple pour une base de données censitaires, les attributs stockés dans le système peuvent être nombreux.

A1.10. La télédétection est une autre source d'informations provenant de l'imagerie de télédétection (aériennes ou satellitaires par exemple). Les photographies

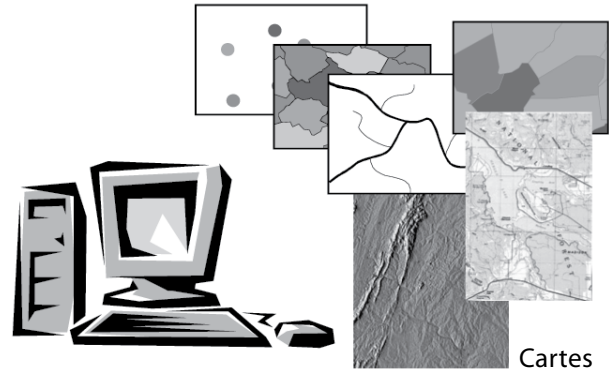
Figure A1.2

Types d'informations stockées dans un SIGDonnées sous
forme de tableaux

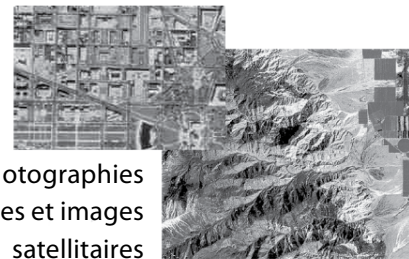
Pop Id	Popdens
103	23.7
104	110.5
105	35.7
201	96.8
	73.4

LU Id	Land use
2308	Forest
2712	Urban
2487	Agric.
3102	
2402	

New Id	Popdens	Land use
23	110.5	Forest
24	110.5	Agric.
25	73.8	Forest
26	96.8	Urban
27	73.4	Agric.



Cartes

Images et
multimédiaPhotographies
aériennes et images
satellites

ou images prises à partir d'un avion volant à basse altitude ou à partir de satellites peuvent être intégrées à d'autres informations à référence spatiale. Parfois ces images offrent simplement un arrière-plan pour des informations cartographiques thématiques ou topographiques. Mais le plus souvent des informations interprétées et extraites de ces images sont stockées sous forme de données cartographiques numériques. Enfin, des informations multimédia telles que des photographies, de la vidéo, du texte ou même du son peuvent être intégrées au SIG. Souvent, l'intégration s'effectue au moyen de liaisons rapides. L'utilisateur peut cliquer de façon interactive sur un objet pour voir des photographies ou un film vidéo de l'emplacement géographique. Cela a été facilité dans une large mesure par l'émergence de programmes de cartographie en ligne simples d'utilisation tels que Google Earth.

2. Couches de données géographiques

A1.11. Une base de données SIG est une représentation informatique du monde réel. Le logiciel de SIG fournit les outils pour organiser les informations relatives à des objets spatialement définis. Le principe fondamental pour l'organisation d'un SIG est la couche de données, comme l'illustre la figure A1.3. Plutôt que de stocker tous les objets spatiaux au même endroit, comme sur une carte topographique, des groupes d'objets semblables peuvent être combinés dans une de ces couches de données.

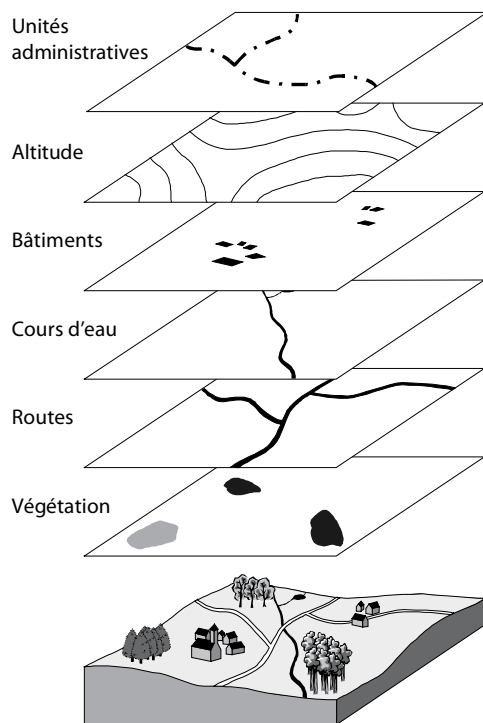
A1.12. Une base de données géographique peut comporter des couches d'objets physiques tels que des routes, des cours d'eau et des bâtiments, ainsi que des couches d'objets définis comme des limites administratives ou des zones postales qui ne peuvent être observées sur le terrain. En outre, le logiciel de SIG nous permet de créer de nouvelles couches de données basées sur des couches existantes. Par exemple, une nouvelle couche de données pourra montrer des bassins hydrographiques obtenus à

partir de données numériques concernant l'altitude ou toutes les zones qui se trouvent à une distance déterminée d'un hôpital.

A1.13. Lorsque l'on crée une base de données géographique multicouches, des objets peuvent être extraits de toute une série de sources différentes topographiques et thématiques. En outre, des observations sur le terrain et des données de télédétection par satellite ou de photographies aériennes sont souvent intégrées aux données cartographiques. Le SIG fournit les outils qui intègrent tous ces lots de données différents dans un cadre de référence commun défini par le système de coordonnées géographiques. Cela permet à l'utilisateur de combiner divers types de données, de créer de nouvelles informations ou d'effectuer des requêtes complexes concernant plusieurs couches de données. Cette possibilité d'intégrer des données provenant de sources hétérogènes en utilisant la localisation géographique comme lien est parfois qualifiée « d'utilisation de l'espace comme système d'indexation ». C'est en fait l'un des principaux avantages procurés par le système d'information géographique.

Figure A1.3

Couches de données — le concept d'espace comme système d'indexation



B. Modèles de données de SIG

A1.14. Bien que les informations qui peuvent être stockées dans une base de données géographique aient un caractère hétérogène, seul un petit nombre de méthodes communes permet de représenter les informations spatiales dans une base de données. Pour élaborer une application de SIG, les objets du monde réel doivent être traduits en représentations simplifiées qui peuvent être stockées et manipulées dans un ordinateur. Deux modèles de données — représentations numériques internes des

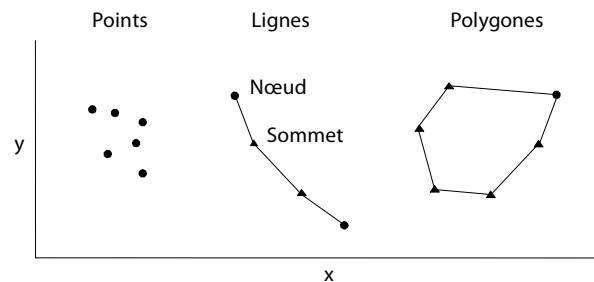
informations — prédominant actuellement dans les logiciels commerciaux de SIG, bien que de nombreux programmes gèrent désormais aisément ces deux modèles. Le modèle de données « vectoriel » utilisé pour symboliser des objets individualisés tels que des maisons, des routes ou de districts et le modèle « matriciel », qui est plus souvent utilisé pour représenter des phénomènes présentant des variations continues comme l'altitude ou le climat, mais qui est aussi utilisé pour stocker des photographies ou des images provenant de satellites ou de caméras embarquées à bord d'avions. Pour des applications de recensement, le modèle de données vectorielles est en général plus utile, bien qu'il soit plus approprié de stocker de nombreux lots de données auxiliaires au moyen du modèle matriciel.

1. Le système vectoriel

A1.15. Les systèmes vectoriels de SIG représentent des objets du monde réel au moyen d'un ensemble de primitives géométriques : points, lignes et polygones (voir figure A1.4). Un point est représenté dans une base de données d'ordinateur par ses coordonnées x,y . Une ligne est une succession de coordonnées x,y , les points extrêmes étant appelés généralement des nœuds et les points intermédiaires des sommets. Les polygones ou les surfaces sont représentés par une série fermée de lignes telles que le premier point est le même que le dernier du périmètre. Des points peuvent être utilisés pour représenter des maisons, des puits ou des points de contrôle géodésique; les lignes décrivent des objets tels que des routes et des cours d'eau; et les zones ou districts de recensement, par exemple, sont représentés par des polygones.

Figure A1.4

Points, lignes et polygones



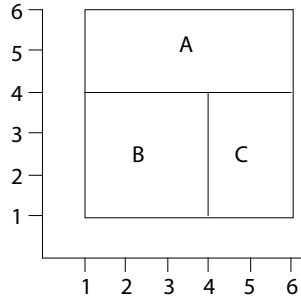
A1.16. Les modèles de données vectorielles les plus simples stockent les données sans établir de relations entre les objets géographiques (voir figure A1.5). C'est ce que l'on appelle parfois le « modèle spaghetti » (par exemple, Aronoff, 1991) car les lignes dans la base de données se recouvrent sans se recouper, comme des spaghettis dans une assiette. Des *modèles de données topologiques* plus perfectionnés stockent les relations entre les divers objets dans une base de données. Par exemple, les lignes qui se croisent sont fractionnées et un nœud supplémentaire est ajouté à l'intersection. Au lieu de définir deux fois la limite entre des polygones voisins — une fois pour chaque polygone en boucle fermée — la ligne est stockée seulement une fois, avec des informations précisant quel polygone est situé respectivement à droite et à gauche de la ligne. Les informations relatives aux relations entre nœuds, lignes et polygones sont stockées dans des tables d'attributs.

A1.17. Les avantages du modèle topologique deviennent évidents si l'on pense aux requêtes que l'on pourra adresser à une base de données spatiales. Une base de données à structure topologique permet d'obtenir rapidement des informations sur des ob-

Figure A1.5

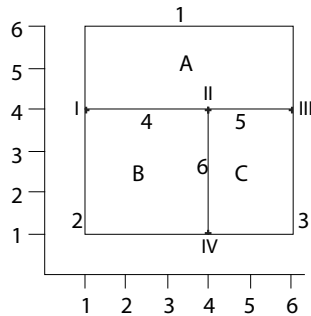
Modèles de données vectorielles : mode spaghetti et mode topologique

Structure de données « spaghetti »



Polygone	Coordonnées
A	(1,4), (1,6), (6,6), (6,4), (4,4), (1,4)
B	(1,4), (4,4), (4,1), (1,1), (1,4)
C	(4,4), (6,4), (6,1), (4,1), (4,4)

Structure de données topologique



Nœud	X	Y	Lignes
I	1	4	1,2,4
II	4	4	4,5,6
III	6	4	1,3,5
IV	4	1	2,3,6

Polygone	Lignes
A	1,4,5
B	2,4,6
C	3,5,6

Ligne	À partir du nœud	Vers le nœud	Polygone gauche	Polygone droit
1	I	III	O	A
2	I	IV	B	O
3	III	IV	O	C
4	I	II	A	B
5	II	III	A	C
6	II	IV	C	B

O = Polygone « extérieur »

jets de données individuels et leurs rapports avec d'autres objets de données. Par exemple, pour identifier rapidement tous les districts voisins d'un district de recensement particulier, le système parcourra simplement la liste des lignes qui définissent ce district et trouvera tous les autres districts de recensement limités également par ces lignes.

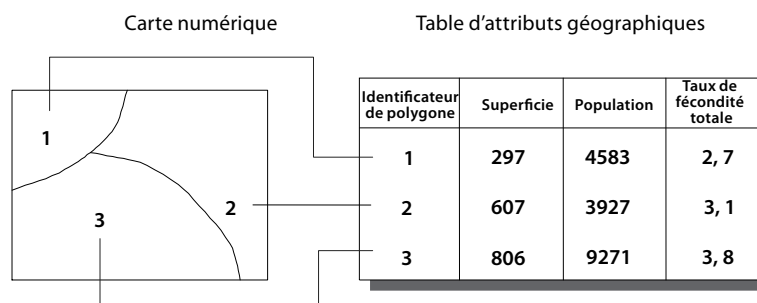
A1.18. La plupart des logiciels de SIG utilisent à présent des structures de données entièrement topologiques qui autorisent des opérations complexes telles que la superposition de polygones. Au cours de cette opération, deux ensembles de données vectorielles sont combinés — par exemple des districts administratifs et des limites de bassins hydrographiques. De nouveaux polygones plus petits sont créés en recoupant des polygones obtenus à partir des deux ensembles de données d'entrée. La plupart des systèmes de cartographie sur ordinateur utilisent des structures de données plus simples. Dans celles-ci, tous les polygones sont définis sous la forme de boucles fermées de telle sorte que les lignes définissant la limite entre deux districts sont stockées deux fois dans la base de données.

A1.19. Chaque objet figurant dans la base de données est étiqueté à l'intérieur au moyen d'un identificateur unique qui lie l'objet géométrique à une entrée correspondante dans un tableau de données ou d'attributs (voir figure A1.6). L'utilisateur peut ajouter les informations relatives à chaque objet dans l'enregistrement correspondant de la base de données. Pour les points représentant des maisons individuelles, l'utilisateur pourra énumérer l'adresse postale, le type de maison et indiquer si la

maison dispose d'électricité et d'installations sanitaires. Dans une base de données des districts de recensement, l'utilisateur pourra ajouter le code administratif officiel, le nombre d'unités d'habitation et toutes les données de recensement qui ont été réunies pour le district de recensement. À des fins pratiques, la plupart des SIG utilisent un modèle de base de données relationnelle pour stocker séparément les attributs ou les informations non spatiales dans une base de données. Les fichiers d'attributs sont étroitement intégrés aux données géographiques numériques et on peut y accéder par l'intermédiaire du SIG ou d'un système de gestion de base de données relationnelle.

Figure A1.6

Données spatiales et non spatiales stockées dans un SIG vectoriel



A1.20. Entre les deux extrêmes — le modèle spaghetti simple et le modèle complexe entièrement topologique — certains logiciels de cartographie sur ordinateur ont établi un compromis. Sans être entièrement topologiques, ces systèmes permettent un calcul rapide des informations de voisinage et de connexité. Ils combinent ainsi la facilité de vérification du modèle de données simple à certains éléments des puissantes capacités analytiques d'un modèle vectoriel topologique de données de SIG.

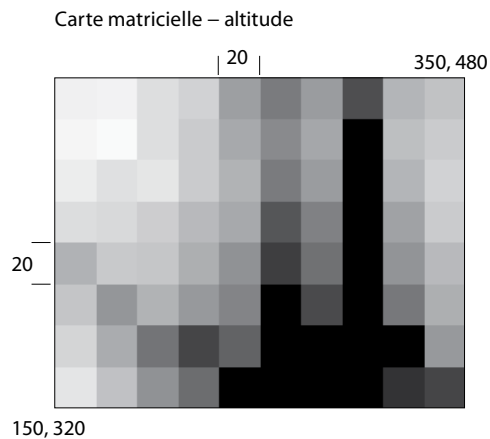
2. Le système matriciel

A1.21. Les logiciels de SIG en mode maillé divisent l'espace en un tableau ordonné de lignes et de colonnes. Une maille de cette matrice ou réseau est parfois appelée un pixel, c'est-à-dire un élément d'image. Cela révèle que ce modèle de données trouve son origine dans la télédétection et le traitement d'images. Dans la plupart des systèmes matriciels, la valeur d'un attribut en un lieu donné, par exemple son altitude, est stockée dans la maille correspondante de la matrice. La base de données matricielle de l'altitude est ainsi simplement une longue série de valeurs d'altitude. La seule information supplémentaire requise par le système est le nombre de lignes et de colonnes dans l'image matricielle, les dimensions des mailles (généralement carrées) exprimées en unités du monde réel (par exemple mètres ou pieds) et les coordonnées de l'un des angles de la grille toute entière (voir figure A1.7). Ces informations sont généralement stockées dans un en-tête ou un petit fichier séparé. Ces éléments d'information permettent au système de calculer les dimensions de la grille. Par exemple, la coordonnée x de l'angle droit supérieur est $150 + 10 \times 20 = 350$. Le système peut utiliser cette information pour enregistrer correctement la grille avec d'autres couches de données géographiques, par exemple pour tracer des objets vectoriels au-dessus de la grille.

A1.22. Cette méthode d'enregistrement de données est naturellement très inefficace si de nombreuses mailles de la grille présentent des valeurs semblables. Par exemple, des objets distincts sont parfois enregistrés en mode maillé. Une carte de district dans ce format indiquera dans chaque maille l'identificateur de district ou la population totale du district dans lequel se trouve cette maille. Il y aura évidemment beaucoup de cellules

contiguës présentant la même valeur. C'est pourquoi la plupart des systèmes matriciels de SIG utilisent telle ou telle forme de compression de données. La plus simple consiste dans le codage par longueur de ligne, dans lequel le système enregistre des paires constituées de deux nombres : la valeur de la donnée et le nombre de fois où cette valeur est répétée. Cela peut réduire sensiblement les dimensions des fichiers.

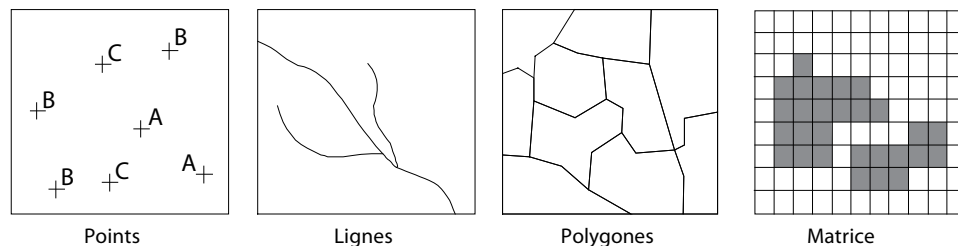
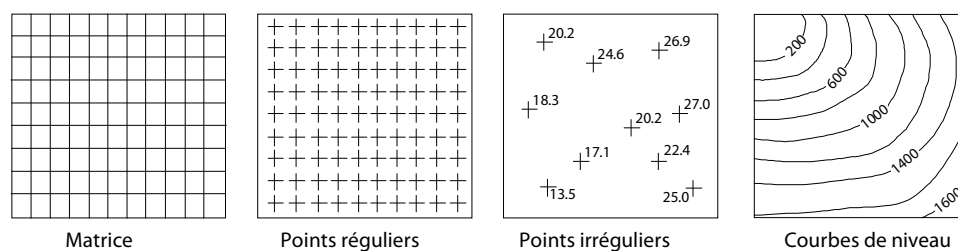
Figure A1.7

Exemple de fichier de données matricielles**Fichier matriciel ASCII**

Nombre de colonnes	10								
Nombre de lignes	8								
Angle inférieur gauche x	150								
Angle inférieur gauche y	320								
Dimension d'une maille (carrés)	20								
219	313	407	462	681	783	689	877	595	540
297	274	407	501	642	744	650	955	556	501
336	391	368	501	603	783	689	994	595	462
414	430	485	579	642	861	767	1072	673	501
609	508	524	618	720	900	806	1267	712	579
531	703	602	696	759	978	884	1189	790	618
453	625	797	891	837	1173	1079	1111	985	696
375	547	719	813	1032	1095	1001	1033	907	891

A1.23. Les données matricielles sont le plus souvent utilisées pour enregistrer des données qui varient constamment ou des images présentant de nombreuses nuances de gris continues. De même que des objets discontinus peuvent être présentés en mode maillé, de même des données continues peuvent aussi être représentées en utilisant des structures de données vectorielles. Le meilleur exemple est celui des courbes de niveau qui indiquent l'altitude sur les cartes topographiques. D'autres exemples sont présentés sur la figure A1.8.

Figure A1.8

Mode vectoriel et mode maillé peuvent être utilisés l'un et l'autre pour présenter des données discontinues et continues**Objets discontinus****Données continues**

3. Avantages et inconvénients des modèles de données vectoriel et matriciel

A1.24. Le modèle de données matriciel a l'avantage de la simplicité. De nombreuses opérations sur des données géographiques sont plus faciles à mettre en œuvre et s'exécutent plus rapidement dans un SIG en mode maillé. La modélisation de données continues, comme c'est souvent le cas pour les données hydrologiques ou d'altitude, est en général effectuée avec un SIG matriciel. Mais ce modèle présente un inconvénient : il existe une relation inverse entre la dimension des lots de données matricielles obtenues et la précision avec laquelle des entités spatiales peuvent être représentées. Une grille très fine représente toutes les courbes d'une limite avec suffisamment de détails, mais elle exige une grande quantité d'espace disque. Pour une analyse géographique utilisant des « objets », tels que des points, des lignes et des polygones, les modèles vectoriels conviennent mieux.

A1.25. La plupart des opérations d'un SIG peuvent être effectuées avec les deux modèles de données. Le choix d'un modèle approprié dépendra de l'application. Pour le recensement et de nombreuses autres applications socioéconomiques, le modèle vectoriel est plus approprié. Les structures de données vectorielles permettent une représentation plus compacte de points et de polygones qui définissent des objets socioéconomiques. La relation étroite avec les systèmes de gestion de base de données facilite des applications socioéconomiques caractérisées par un grand nombre d'informations attributaires — par exemple des centaines de variables pour un recensement ou une enquête — liées à un nombre fixe d'objets spatiaux tels que des districts de recensement, des villages ou des groupes en cas d'enquête. Enfin, les produits imprimés obtenus à partir des bases de données vectorielles de SIG ressemblent en général davantage aux cartes établies au moyen des techniques cartographiques traditionnelles.

A1.26. Même ainsi, la possibilité de manipuler des données en mode maillé présente une importance croissante pour les applications démographiques. Certaines données d'entrée utiles pour définir les limites des districts de recensement sont disponibles en mode maillé. Le chapitre 4 du présent *Manuel*, par exemple, traite de l'utilisation d'images télédétections pour créer ou mettre à jour une cartographie censitaire. Heureusement, le choix entre les modèles de données ne s'exprime pas en général sous la forme du « ou bien..., ou bien... ». De nombreux logiciels de SIG prennent aujourd'hui en charge les deux types de données spatiales. Cela permet par exemple d'utiliser des données en mode maillé comme fond sur lequel des objets linéaires et polygonaux peuvent être tracés. Ainsi, des images de télédétection ou des surfaces d'altitude peuvent être affichées sur un écran d'ordinateur en même temps que d'autres informations pertinentes pour aider à définir les limites des districts de recensement.

4. Précision ou exactitude

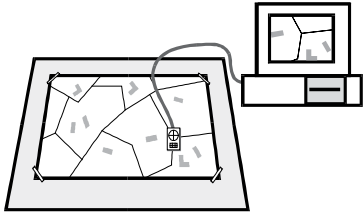
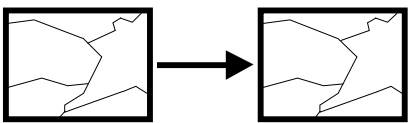
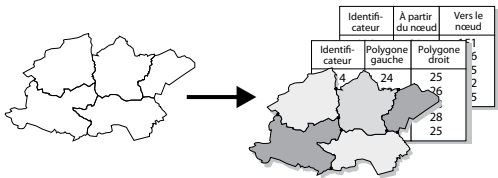
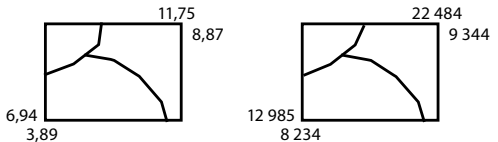
A1.27. On utilise souvent indifféremment les termes de précision et d'exactitude, malgré leur sens différents. Dans un contexte spatial, par exemple, une coordonnée de point exacte dans une base de données de SIG est enregistrée à l'endroit approprié du point de vue de la localisation exacte du point sur la surface terrestre. Par contre, la précision se réfère à la possibilité de distinguer de petites quantités ou de petites distances lors d'une mesure. Par exemple, si nos instruments de topographie ne mesurent les coordonnées qu'en mètres, les localisations de points dans notre SIG ne seront exactes qu'au mètre près. Si nous disposons d'un outil de mesure plus précis, nous pouvons obtenir des coordonnées d'un point d'une précision équivalent au centimètre ou au millimètre près.

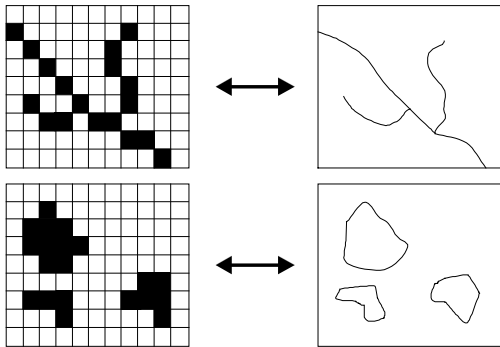
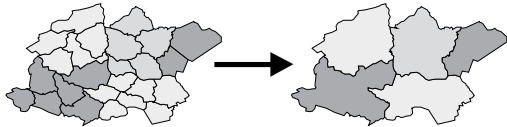
A1.28. En pratique, la précision avec laquelle les coordonnées peuvent être enregistrées dans un SIG vectoriel est virtuellement infinie, parce qu'elles utilisent des types de données double précision (8 octets pour chaque nombre à virgule flottante) pour enregistrer les coordonnées géographiques. Mais l'exactitude des coordonnées spatiales dépend en grande mesure des outils utilisés pour collecter les données. Les meilleurs instruments de topographie utilisés pour des applications d'études techniques ou de recherche sur la tectonique des plaques permettent d'atteindre une exactitude de moins d'un millimètre. Cependant, la plupart des données utilisées dans un SIG proviennent de sources de données beaucoup moins exactes comme les cartes imprimées, les systèmes de positionnement universel tenus à la main ou même des croquis cartographiques esquissés pendant le travail sur le terrain. Ici, l'exactitude sera probablement mesurée en mètres plutôt qu'en millimètres.

C. Possibilités des SIG

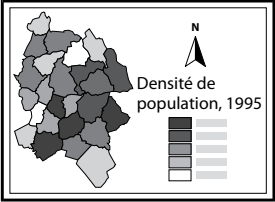
A1.29 Le tableau ci-après fournit un aperçu des possibilités d'un SIG. Cette liste n'est pas exhaustive, car les logiciels de SIG de haut de gamme et même les logiciels de cartographie sur ordinateur offrent de nombreuses fonctions spécialisées d'entrée, de manipulation, d'analyse et d'affichage de données.

Entrée et gestion des données

Tracé de lignes, entrée de coordonnées	La table à numériser demeure le moyen le plus couramment utilisé pour entrer des coordonnées. Des lignes sont tracées sur la carte papier avec un curseur et elles sont saisies dans les logiciels de SIG ou de numérisation. Sinon, les cartes peuvent être scannées pour créer des images matricielles en mode point qui sont ensuite converties en format vectoriel.																									
Édition	Après avoir numérisé les lignes, on doit vérifier si les données ne comportent pas d'erreurs. On rencontre souvent des problèmes de lignes non raccordées (tracés trop courts et dépassements), de lignes manquantes ou de lignes qui ont été numérisées deux fois. Certaines de ces opérations sont automatisées dans les SIG.																									
Construction d'une topologie	Les lignes numérisées ou vectorisées n'ont aucun rapport entre elles. Le logiciel de SIG peut calculer les relations de voisinage et la connectivité entre objets dans le lot de données.	 <table border="1" data-bbox="1234 1499 1421 1648"> <thead> <tr> <th>Identificateur</th> <th>À partir du nœud</th> <th>Vers le nœud</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>4</td> <td>5</td> </tr> <tr> <td>2</td> <td>5</td> <td>2</td> </tr> <tr> <td>3</td> <td>2</td> <td>5</td> </tr> <tr> <td>4</td> <td>24</td> <td>25</td> </tr> <tr> <td>5</td> <td>25</td> <td>26</td> </tr> <tr> <td>6</td> <td>26</td> <td>28</td> </tr> <tr> <td>7</td> <td>28</td> <td>25</td> </tr> </tbody> </table>	Identificateur	À partir du nœud	Vers le nœud	1	4	5	2	5	2	3	2	5	4	24	25	5	25	26	6	26	28	7	28	25
Identificateur	À partir du nœud	Vers le nœud																								
1	4	5																								
2	5	2																								
3	2	5																								
4	24	25																								
5	25	26																								
6	26	28																								
7	28	25																								
Géoréférencement et changement de projection	Les lignes numérisées sont exprimées en centimètres ou en pouces. Elles doivent être converties en unités du monde réel correspondant au système de coordonnées de la carte source par exemple en mètres ou en pieds. Pour l'intégration des données, on pourra aussi avoir à modifier la projection des cartes numériques.	 Avant transformation (carte en unités numériques) Après transformation (Carte en unités du monde réel)																								

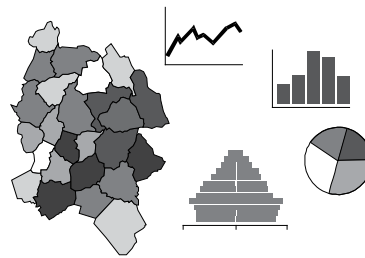
Conversion mode maillé – mode vectoriel	La plupart des logiciels de SIG prennent aujourd’hui en charge les images matricielles sous telle ou telle forme. Chaque modèle de données étant approprié pour des tâches différentes, on doit disposer de fonctions pour convertir un mode dans l’autre. La conversion de mode maillé - mode vectoriel est aussi utilisée pour la conversion automatique des cartes scannées. L’opération inverse — mode vectoriel-mode maillé — doit être effectuée à des fins d’analyse et de modélisation dans un SIG en mode maillé.																																																													
Gestion des données attributaires	Chaque objet de la base de données est étiqueté à l’aide d’un identificateur unique. Cet identificateur sert de lien pour accéder à des informations extérieures relatives à des objets géographiques. Pour permettre la manipulation et l’analyse des tables d’attributs, le SIG est généralement intégré à un système de gestion de base de données relationnelle.	Districts <table><tr><th>Identifi- cateur</th><th>District</th><th>D_Pop</th><th>Identificateur de province</th></tr><tr><td>0101</td><td>Palma</td><td>89763</td><td>01</td></tr><tr><td>0102</td><td>S. Maria</td><td>45938</td><td>01</td></tr><tr><td>0103</td><td>Veralo</td><td>78383</td><td>01</td></tr><tr><td>0201</td><td>Bolo</td><td>98302</td><td>02</td></tr><tr><td>0202</td><td>Jose</td><td>67352</td><td>02</td></tr><tr><td>0203</td><td>Malabo</td><td>102839</td><td>02</td></tr><tr><td>0204</td><td>Chilabo</td><td>129388</td><td>02</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td></tr></table> Provinces <table><tr><th>Identificateur de province</th><th>P_Pop</th><th>Région</th></tr><tr><td>01</td><td>214084</td><td>112</td></tr><tr><td>02</td><td>397881</td><td>113</td></tr><tr><td>...</td><td>...</td><td>...</td></tr></table> <table><tr><th>Identifi- cateur_F</th><th>Ménage_R</th><th>Région</th></tr><tr><td>23</td><td>1249</td><td>112</td></tr><tr><td>24</td><td>2458</td><td>113</td></tr><tr><td>...</td><td>...</td><td>...</td></tr></table>	Identifi- cateur	District	D_Pop	Identificateur de province	0101	Palma	89763	01	0102	S. Maria	45938	01	0103	Veralo	78383	01	0201	Bolo	98302	02	0202	Jose	67352	02	0203	Malabo	102839	02	0204	Chilabo	129388	02	...	...	...	...	Identificateur de province	P_Pop	Région	01	214084	112	02	397881	113	...	...	...	Identifi- cateur_F	Ménage_R	Région	23	1249	112	24	2458	113	...	...	...
Identifi- cateur	District	D_Pop	Identificateur de province																																																											
0101	Palma	89763	01																																																											
0102	S. Maria	45938	01																																																											
0103	Veralo	78383	01																																																											
0201	Bolo	98302	02																																																											
0202	Jose	67352	02																																																											
0203	Malabo	102839	02																																																											
0204	Chilabo	129388	02																																																											
...	...	...	...																																																											
Identificateur de province	P_Pop	Région																																																												
01	214084	112																																																												
02	397881	113																																																												
...	...	...																																																												
Identifi- cateur_F	Ménage_R	Région																																																												
23	1249	112																																																												
24	2458	113																																																												
...	...	...																																																												
Reclassification, agrégation	Le SIG permet d’agréer des objets basés sur un identificateur commun. Par exemple, des districts de recensement peuvent être groupés en zones opérationnelles de recensement de population à peu près égale.																																																													
Création de sous-ensembles, découpage	Parallèlement à la sélection de sous-ensembles établis à partir de requêtes, le SIG peut aussi créer des sous-ensembles individualisés par des opérations de découpage.																																																													

Affichage

Fonctions cartographiques	L'établissement de produits cartographiques à des fins de présentation n'est qu'une application de cartographie dans un SIG. La symbolisation cartographique est aussi importante pour distinguer des objets lors de l'édition et de l'analyse sur écran.	
Affichage combiné d'images et de données vectorielles	Les images ou les données en mode maillé proviennent de différentes sources : cartes scannées, images de télédétection et données matricielles de SIG sont toutes stockées dans un format de grille. L'affichage des données vectorielles et matricielles combinées peut fournir un contexte très utile pour l'analyse et il permet l'extraction sélective d'objets à partir des données en mode maillé.	

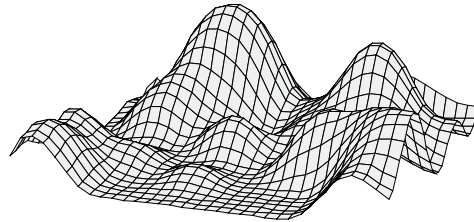
Liens avec les graphiques statistiques

L'analyse, dirigée vers les données, de données spatiales est en général une combinaison de cartographie et d'examen d'attributs. Les graphiques statistiques sont précieux, surtout s'ils peuvent être affichés sur les cartes.



Affichage de surfaces en trois dimensions

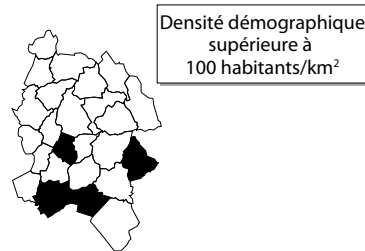
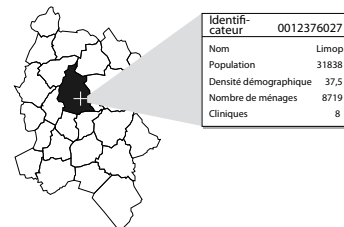
Des données continues telles que l'altitude ou les précipitations — et dans une certaine mesure aussi la densité démographique — peuvent être affichées dans différents formats : matrices, courbes de niveau ou visualisations simulées en trois dimensions, au moyen de modèles fil de fer sur lesquels les autres objets peuvent être drapés.



Consultation

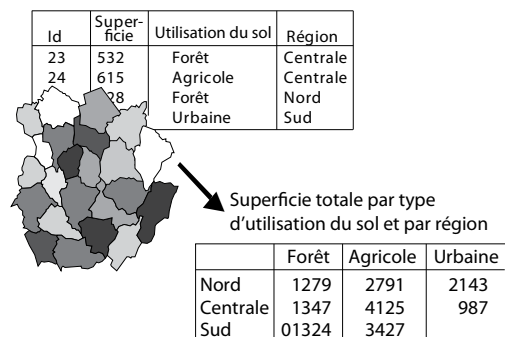
Consultation de base de données spatiale

« Qu'est-ce qui se trouve à... ? » et « Où est ... ? » sont deux questions géographiques essentielles auxquelles un SIG peut répondre. En simple mode de lecture, un utilisateur peut choisir des objets sur une carte numérique et obtenir des informations les concernant. Inversement, l'utilisateur peut choisir des objets correspondant à un ensemble de critères et les afficher sur la carte. Les SIG sont en général associés à des logiciels de gestion de base de données et les opérations de consultation sont fondées sur le concept du SQL. Les SIG permettent aussi de procéder à des consultations fondées sur des relations géographiques, telles que la distance (Qu'est-ce qui se trouve à x km de cet endroit ?) ou des requêtes fondées sur deux couches de données de SIG ou plus (Quels bâtiments se trouvent dans ce district de recensement ?)



Récapitulation des attributs

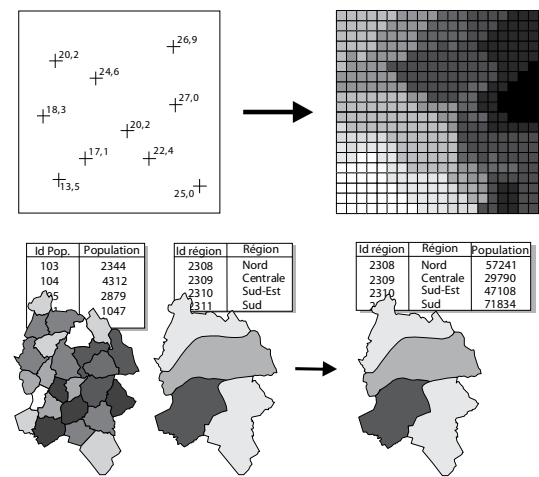
Les opérations sur les bases de données nous permettent d'extraire des statistiques sommaires utiles ou des tableaux à multiples entrées du tableau des attributs géographiques d'un ensemble de données de SIG. Par exemple, nous pouvons calculer la valeur minimum, maximum et moyenne d'un champ dans le tableau. Ou bien nous pouvons dresser un tableau à multiples entrées à partir de deux champs ou plus du tableau et établir des totaux récapitulatifs d'un troisième champ pour chaque combinaison de catégories d'attributs. Cela nous permet, par exemple, de calculer la superficie totale de chaque catégorie d'utilisation des terres dans les différentes régions d'un pays. Les tableaux à multiples entrées sont souvent utilisés lorsque deux couches de SIG ou plus ont été combinées par une opération de superposition de polygones (voir plus bas).



Transformation de données spatiales

Interpolation

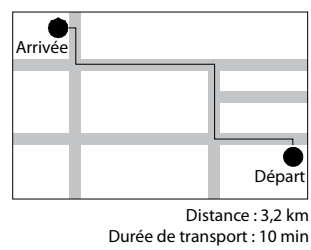
L'interpolation nous permet de créer une couverture complète à partir de données échantillons. Par exemple, à partir d'un ensemble de surfaces de précipitations à la station, nous pouvons créer une surface matricielle qui montre les précipitations dans l'ensemble de la région. L'interpolation aréolaire est encore plus importante pour les applications socioéconomiques. Par exemple, à partir de la population par district, nous pourrions vouloir évaluer la population dans des régions de surveillance environnementale dont les limites ne correspondent pas à celles des districts.



Opérations concernant les distances et les durées

Calculs simples de distance

Le calcul des distances est l'une des opérations essentielles effectuées à l'aide d'un SIG. Les distances (et les durées) peuvent être calculées en ligne droite ou dans un réseau. À partir d'une base de données routières d'un SIG, par exemple, on pourra évaluer les distances et les temps de déplacement.



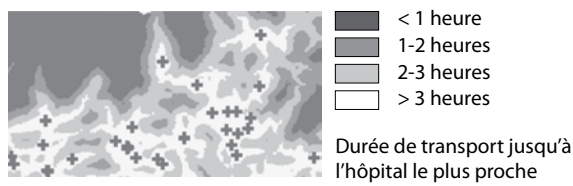
Zones enveloppes

La création de régions enveloppes constitue un type particulier d'opération relative aux distances. Des zones enveloppes peuvent être créées autour de points, de lignes ou de polygones et peuvent être pondérées par des valeurs d'attributs. Par exemple, des routes bitumées pourront avoir une zone enveloppe plus large que les routes en terre. Les zones enveloppes sont souvent utilisées dans les demandes d'informations spatiales. Par exemple, pour déterminer le nombre de cas de bilharziose à une distance de 3 km d'un cours d'eau, la demande d'information portera successivement sur une zone enveloppe, un point dans un polygone et une base de données.



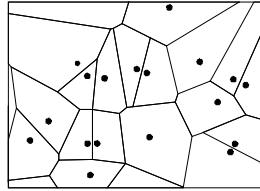
Recherche de l'objet le plus proche

On a recours à une combinaison de la consultation de base de données et du calcul de distances lorsqu'on a besoin de découvrir le plus proche de divers objets d'une catégorie donnée. Par exemple, nous voudrions calculer pour toutes les localités d'un district la distance de l'hôpital le plus proche. L'ensemble de données de SIG obtenu est souvent qualifié de surface d'accessibilité.



Polygones de Thiessen

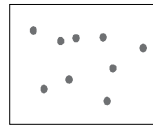
La fonction « découvrir l'objet le plus proche » comporte une variante, à savoir une opération consistant à subdiviser toute la région en polygones affectés à l'installation la plus proche. Les éléments de surface obtenus sont appelés polygones de Thiessen. Cette fonction est souvent utilisée pour définir de simples aires de chalandise ou de service.



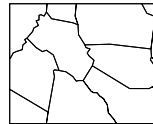
Combinaison de couches de données

Opération portant sur un objet ponctuel ou linéaire

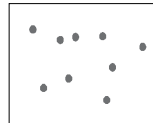
De nombreuses questions qu'un SIG peut aider à résoudre exigent la combinaison de plusieurs ensembles de données. Par exemple, nous pouvons avoir un ensemble de coordonnées ponctuelles représentant les groupes d'une enquête démographique et nous souhaitons combiner les informations de l'enquête avec des données de recensement disponibles par district de recensement. Le SIG permettra de définir pour chaque point le district de recensement dans lequel il se situe et rattachera les données de recensement à l'enregistrement des attributs de ce point. La même opération nous permet de récapituler les attributs d'objets ponctuels ou linéaires pour un groupe de régions. Par exemple, nous pouvons déterminer le taux de fécondité moyen pour chaque circonscription sanitaire en utilisant un échantillon de ménages recensés (points).



Id de grappe
12
13
14
15
16



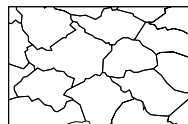
Id de l'unité de recensement	Dimension moyenne des ménages
507	4,3
508	3,8
601	2,9
602	5,3
603	4,6



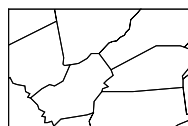
Id de grappe	Dimension moyenne des ménages
12	4,3
13	3,8
14	2,9
15	5,3
16	4,6

Superposition de polygones

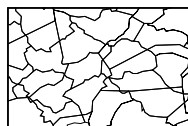
La combinaison de deux lots de données de SIG relatives à des éléments aréolaires est appelée superposition de polygones. Le système fusionne les ensembles de données et crée de nouveaux éléments aréolaires à partir des zones de superposition. Le nouvel ensemble de données ainsi obtenu contient les attributs des deux lots de données. On décidera en fonction des types de données si l'attribut doit rester inchangé (par exemple, renseignements sur des catégories ou pourcentages) ou s'il doit être divisé entre les nouveaux polygones (par exemple, données de dénombrement). La superposition de polygones est souvent utilisée en combinaison avec des tableaux à entrées multiples, par exemple pour calculer des données de recensement par zone d'utilisation des sols.



Id. de population	Densité démographique
103	23.7
104	110.5
105	35.7
201	96.8
202	73.4



Id. d'utilisation du sol	Utilisation du sol
2308	Forêt
2712	Urbaine
2487	Agricole
3102	Agricole
2402	Urbaine



Nouvel identificateur	Densité démographique	Utilisation du sol
23	110.5	Forêt
24	110.5	Agricole
25	73.8	Forêt
26	96.8	Urbaine
27	73.4	Agricole

Annexe II

Systèmes de coordonnées et projections cartographiques

A. Introduction

A2.1. La précédente analyse des principes du SIG à l'annexe I a fait ressortir les avantages de l'intégration des données spatiales. En organisant divers types d'informations géographiques sous la forme de couches de données, des mesures, des interrogations, des opérations de modélisation et d'autres types d'analyses peuvent être réalisés en utilisant des données provenant de nombreux domaines différents. Ainsi, les données de recensement peuvent être analysées en combinaison avec des données sur l'utilisation des terres ou des données agroécologiques, ou bien des résultats d'enquête socioéconomique peuvent être associés à des données à référence géographique sur les risques de maladies. C'est l'intégration verticale des différentes couches de données qui permet de relier entre elles des données provenant de nombreuses sources différentes. Cela signifie simplement que tous les ensembles de données géographiques sont référencés en utilisant le même système de coordonnées, de telle sorte que différentes couches de données soient correctement alignées lorsqu'on les superpose.

A2.2. En constituant une base de données géographique — par exemple pour un recensement — l'agent d'élaboration des données doit veiller à ce que les coordonnées spatiales et les limites saisies à partir de sources de données imprimées, de répertoires toponymiques numériques ou pendant le travail sur le terrain soient enregistrées dans un système de coordonnées approprié selon un processus appelé « géoréférencement ». Cela garantit également que les cartes numériques mises au point séparément pour des régions voisines se raccordent parfaitement lorsqu'elles sont affichées ensemble sur un écran d'ordinateur ou une page imprimée.

A2.3. Le problème était moins préoccupant lorsqu'on établissait les cartes de recensement à l'aide de techniques traditionnelles, car les cartes sur papier — souvent des croquis cartographiques esquissés sur le terrain — n'étaient utilisées qu'aux fins du recensement. Elles n'étaient ni intégrées à d'autres données ni utilisées pour un type quelconque d'analyse spatiale. La connaissance des systèmes de coordonnées et des projections cartographiques revêtait donc beaucoup moins d'importance qu'aujourd'hui lorsqu'on construit une base de données numérique destinée à servir à de nombreuses fins différentes. La présente annexe analyse brièvement des concepts cartographiques importants. On pourra trouver beaucoup d'autres informations dans des manuels de cartographie tels que ceux de Robinson et autres (1995), Kraak et Ormeling (1996) et Dent (1998). Des analyses plus spécialisées sur ce thème peuvent être trouvées dans Canters et Decler (1989), Snyder (1993) et Bugayevskiy et Snyder (1995).

B. Coordonnées

A2.4. En cartographie, la méthode par laquelle on mesure les positions des objets à la surface du sol est appelée le système de coordonnées géographiques. Elle est également parfois appelée « système de référencement géographique ». Dans la géométrie à deux dimensions, le système de coordonnées le plus courant est le système de coordonnées cartésiennes, du nom du savant français René Descartes (1596-1650). Les coordonnées se présentent comme des distances perpendiculaires sur deux axes fixes (x et y) mesurées à partir d'une origine fixe. C'est un système utilisé dans les SIG et aussi dans les applications graphiques plus générales sur ordinateur. On peut utiliser une autre méthode pour définir des positions, le système de coordonnées polaires, qui mesure l'angle et la distance à partir d'un point d'origine fixe (voir figure A2.1).

Figure A2.1

Systèmes de coordonnées planaires et polaires

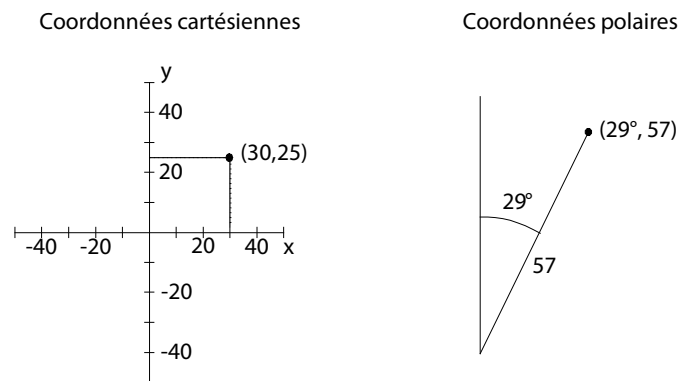
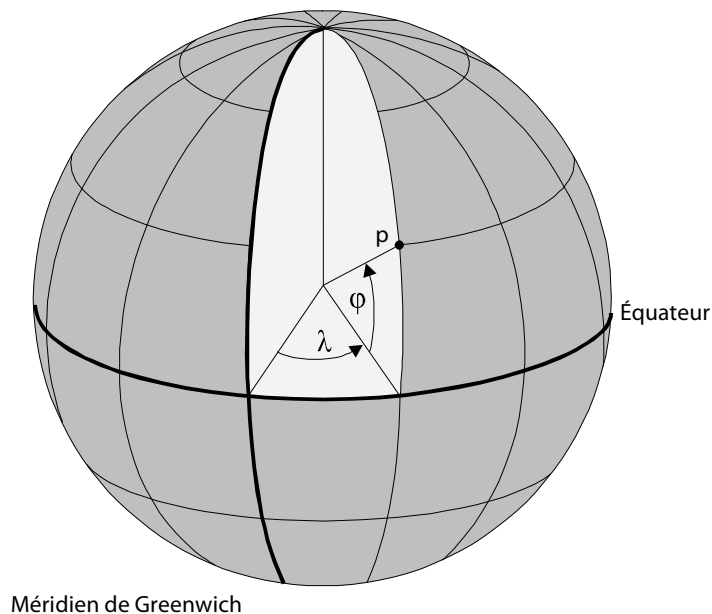


Figure A2.2

Coordonnées sur la sphère : le système de référence latitude/longitude



A2.5. Une carte plane, sur papier ou sur l'écran d'un ordinateur, montre les coordonnées dans un système de coordonnées planaires, à deux dimensions, dans lequel les coordonnées sont mesurées en unités standard telles que des mètres ou des pieds. Les coordonnées sont en général appelées coordonnées x et y , bien que l'on utilise aussi souvent les termes « abscisse » et « ordonnée » dans les textes cartographiques. Mais les objets figurant sur une carte sont la représentation d'entités situées à la surface terrestre. La Terre étant une sphère, les coordonnées à la surface terrestre sont mesurées dans un système de coordonnées sphériques. Plus précisément, on utilise généralement les coordonnées de latitude et de longitude pour référencer des positions. Il s'agit du système de coordonnées polaires sphériques dans lequel un point p quelconque est défini par l'angle de la latitude ϕ par rapport au plan défini par l'équateur et par l'angle de la longitude λ , mesurée par rapport au plan défini par le méridien zéro ou de Greenwich (voir figure A2.2).

A2.6. Pour établir des cartes sur papier du monde ou d'une partie de celui-ci, ces coordonnées sphériques de latitude et de longitude doivent être converties d'une façon ou d'une autre dans un système de coordonnées planaires. Un livre récent consacré aux projections cartographiques qualifie d'« aplatissement de la Terre » cette opération qui consiste à établir une représentation bidimensionnelle d'une partie du globe tridimensionnel (Snyder, 1993).

Projections cartographiques

A2.7. L'opération mathématique par laquelle les coordonnées de latitude et de longitude sphériques sont converties en coordonnées planaires est appelée projection cartographique. Nous pouvons effectivement considérer cette opération comme une projection en imaginant une source de lumière placée par exemple au centre de la Terre. Supposons que la surface terrestre est transparente et que seuls les objets présentant de l'intérêt y sont esquissés, nous pourrions simplement poser une feuille de papier plane sur la partie supérieure de la Terre et reproduire les objets projetés sur cette surface, dite développable. Par exemple, un objet situé en un point p à la surface de la Terre sera placé au point p' sur la carte. Comme nous pouvons le constater sur la figure A2.3, plus un point est éloigné de l'endroit où la carte touche le globe, plus sa distance relative par rapport à des points plus rapprochés du point tangent sera faussée. Par exemple, la distance entre p et q sur le globe est très inférieure à la distance entre p' et q' sur la carte. Les points situés sur l'équateur ne peuvent pas être projetés du tout avec ce système, car les rayons lumineux passant par l'équateur sont parallèles à la carte. Cette méthode de projection particulière n'est donc utile que pour les zones relativement proches du point tangent.

Figure A2.3

Exemple de processus de projection cartographique (projection azimutale)

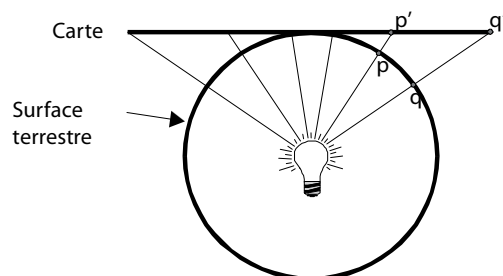
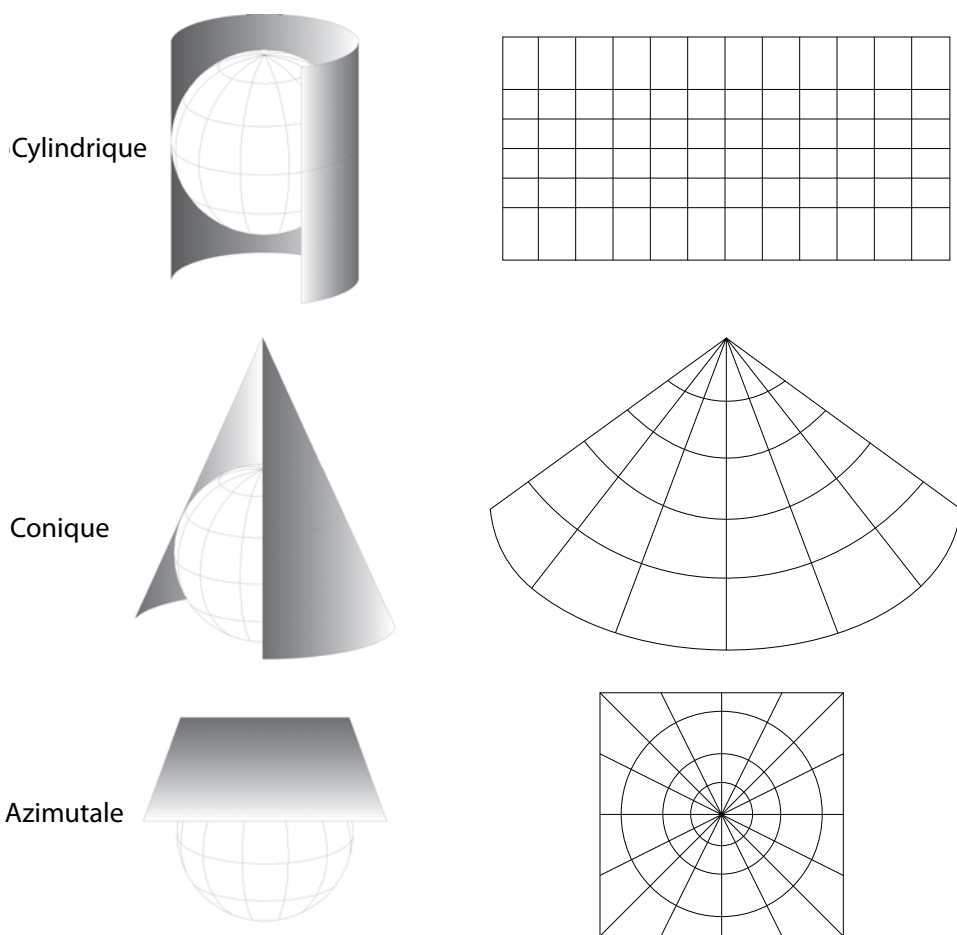


Figure A2.4
Familles de projections cartographiques



Aspects

Normale (équatoriale)

Transversale

Oblique



A2.8. Au cours des siècles, les cartographes ont élaboré beaucoup de projections cartographiques différentes, que l'on peut classer suivant la façon dont la carte est placée sur le globe ou autour de celui-ci. La figure A2.4 en donne un aperçu. Elle montre comment trois types de projections cartographiques — cylindrique, conique et azimutale — sont établis. Comme l'indique le canevas géographique à droite, chaque

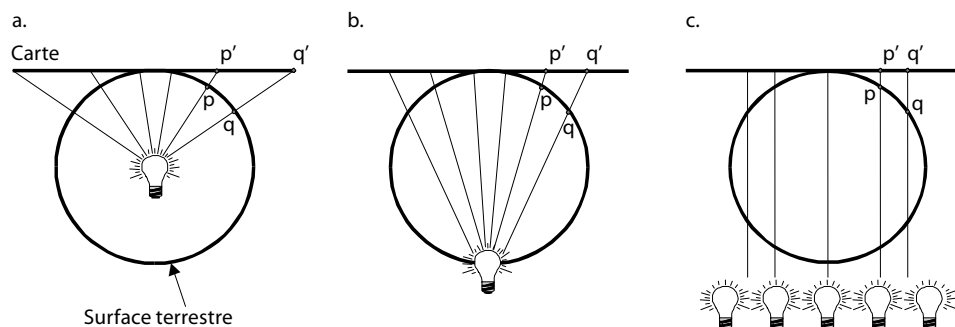
groupe de projections cartographiques se traduit par une configuration caractéristique de lignes de quadrillage correspondant à la latitude et à la longitude.

A2.9. Un cartographe peut aussi choisir l'endroit où la surface développable — cylindre, cône ou plan — entre en contact avec le globe. Cette ligne et ce point tangent est en général l'endroit où les dimensions sont les plus réduites. Si nous établissons des cartes pour une région particulière du monde, nous pouvons donc choisir l'aspect de la projection cartographique pour optimiser la représentation cartographique de la zone qui nous intéresse.

A2.10. La source lumineuse hypothétique n'est pas toujours placée au centre du globe (voir figure A2.5a.). En fait, elle peut être placée au pôle opposé (voir figure A2.5b) ou bien nous pouvons imaginer une série de sources lumineuses émettant de la lumière à partir d'une base plane parallèle à la carte, plutôt qu'à partir d'un point (voir figure A2.5c.). Selon la terminologie cartographique, ces méthodes de projection sont appelées respectivement *gnomoniques*, *stéréographiques* et *orthographiques*. Comme nous pouvons le constater en regardant où les points projetés p' et q' se situent sur la carte, chacune de ces options entraîne un type différent de distorsion de la position relative des lieux représentés sur la carte.

Figure A2.5

Différents modes d'établissement de la projection



C. Propriétés des projections cartographiques

A2.11. Bien que la source lumineuse imaginaire soit un bon moyen d'illustrer le principe des projections cartographiques, celles-ci sont en pratique définies mathématiquement. À partir de la latitude et de la longitude d'un lieu, une formule est utilisée pour obtenir le point correspondant dans le système de coordonnées planaires projeté. Le cartographe dispose de nombreuses options différentes pour créer une projection cartographique qui présentera des caractéristiques particulières. La façon dont la surface développable est disposée autour du globe, l'aspect et la position de la source lumineuse imaginaire ne sont que quelques-uns des paramètres possibles.

A2.12. Malheureusement, il n'existe pas de moyen parfait pour représenter des coordonnées sphériques sur une carte plane. Par conséquent, aucune projection cartographique ne peut répondre à tous les besoins. Chacune peut enregistrer correctement certaines caractéristiques tout en enregistrant d'autres de façon erronée. En fonction de la méthode de projection choisie, divers types de distorsions sont créés. Les projections cartographiques sont donc classées en fonction des caractéristiques qu'elles préservent. Les plus importantes sont les suivantes :

Reproduction correcte des superficies. La plupart des systèmes de projection étirent les éléments aréolaires sur carte. En général, cet étirement n'est pas constant sur toute la surface de la carte, si bien que des éléments proches des pôles sur une carte mondiale, par exemple, paraissent relativement plus étendus que des éléments plus proches de l'équateur. Par exemple, la superficie de la péninsule arabique dépasse de plusieurs centaines de milliers de kilomètres carrés celle de l'île du Groenland. Cependant, sur de nombreuses cartes, le Groenland paraît plusieurs fois plus grand que la péninsule arabique. Les systèmes cartographiques donnant une représentation correcte de la superficie relative de tous les éléments sont qualifiés de systèmes équivalents de projection. *Exemple* : la projection de Mollweide.

Calcul de distances égales. Aucun système de projection cartographique ne peut représenter correctement les distances entre tous les points de la carte. Il convient de s'en souvenir, car le calcul des distances est une application courante des bases de données de SIG. Lors de l'établissement d'une carte à grande échelle d'une petite région géographique, les erreurs apportées seront généralement négligeables. Mais pour des applications nationales ou continentales utilisant des cartes à petite échelle, les distances calculées par un SIG ne seront pas fiables à moins que le logiciel ne compense l'erreur introduite par le calcul de distances euclidiennes à cette échelle. Même des projections équidistantes ne montrent pas toutes les distances correctement, mais elles peuvent représenter exactement toutes les distances entre un ou deux points et tous les autres points, ou le long d'une ou de plusieurs lignes. *Exemple* : la projection conique équidistante. Il faut noter que l'on effectue en général des calculs de distance très exacts en utilisant des formules géométriques exactes plutôt que de simples distances euclidiennes. Ces calculs fondés sur des coordonnées de latitude et de longitude permettent d'obtenir la distance dite « orthodromique ».

Angles corrects. Les projections conformes préservent les angles autour de tous les points et formes sur de petites surfaces. Les méridiens et les latitudes se recoupent à angle droit. Ces projections sont particulièrement utiles pour la navigation. *Exemple* : la projection de Mercator.

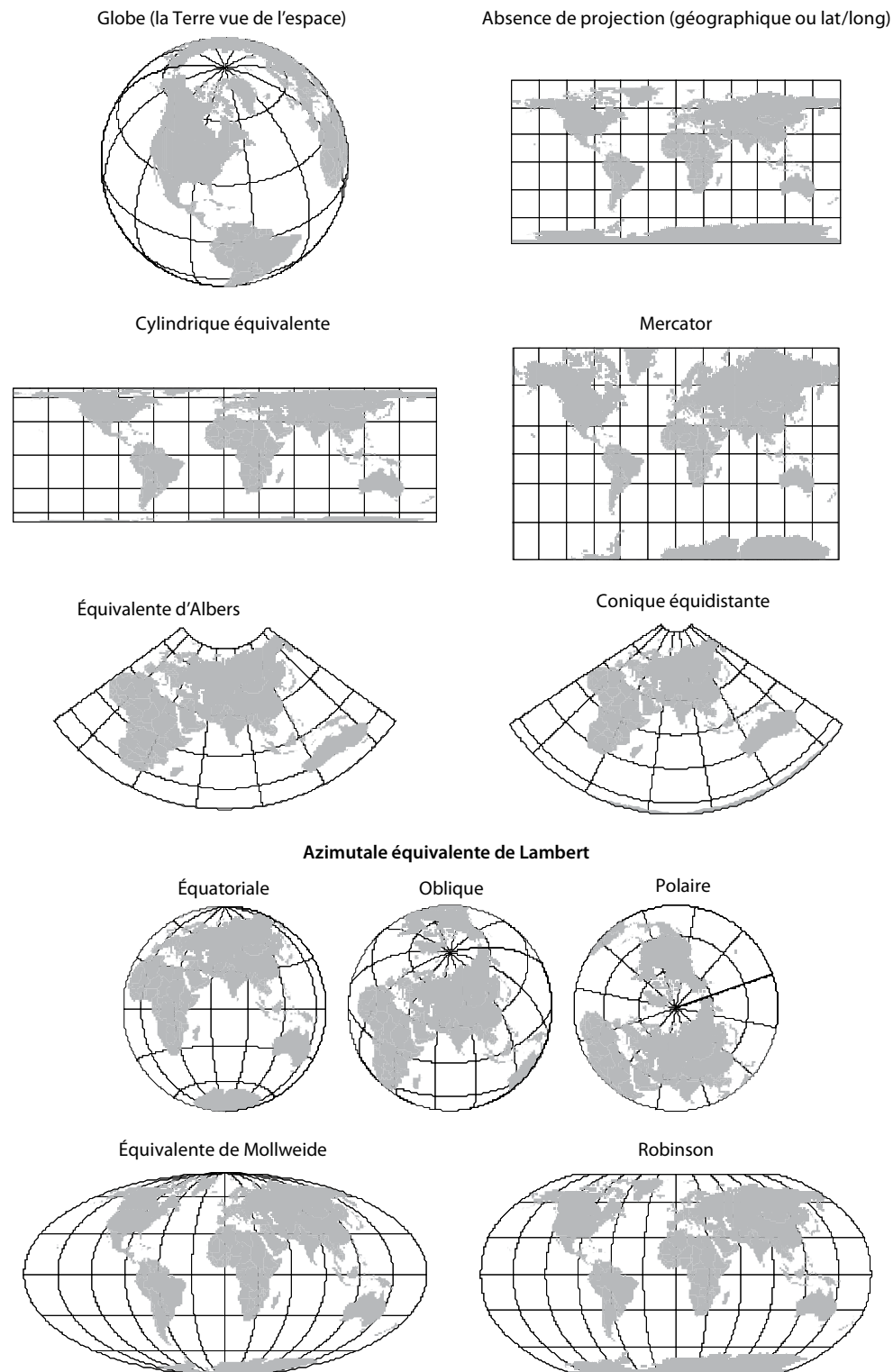
A2.13. Ainsi, toute projection cartographique représente un compromis entre des caractéristiques géographiques souhaitables. Pour une application donnée, il y aura donc des projections cartographiques plus appropriées que d'autres. Outre les caractéristiques des projections cartographiques, il faut prendre en compte les dimensions de la région dont on veut établir la carte, son orientation principale (par exemple nord-sud contre est-ouest) et l'emplacement de la région sur le globe (par exemple zones polaires, latitudes moyennes ou zone équatoriale).

A2.14. Les manuels de cartographie et de nombreux manuels de SIG sont assortis de listes détaillées indiquant quelle projection cartographique convient le mieux à telle ou telle application. Dans certains cas, le meilleur choix peut être une projection qui ne décrit parfaitement aucune caractéristique. La projection Robinson qui est appréciée pour les cartes mondiales, par exemple, est une projection de compromis conçue surtout à des fins esthétiques notamment pour établir des cartes d'atlas. Dans d'autres cas, par exemple si la carte à établir ne concerne qu'une zone relativement restreinte, les distorsions apportées par n'importe quelle projection pourront être négligeables pour une application particulière.

A2.15. La figure A2.6 montre quelques projections cartographiques couramment utilisées. À la partie supérieure, la Terre apparaît sous la forme d'une sphère, ainsi qu'en coordonnées sans projection de latitude et de longitude, positionnées comme s'il s'agissait de coordonnées planaires. Il faut noter au passage que de nombreux distributeurs de données de SIG diffusent des données cartographiques numériques se

Figure A2.6

Projections cartographiques courantes



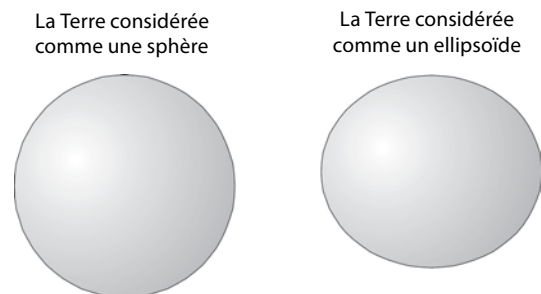
référant à des coordonnées « géographiques » sans projection parce qu'il est en général plus simple pour un utilisateur de convertir des données de latitude et de longitude dans n'importe quel système de projection cartographique et qu'il est parfois plus difficile de passer d'un type de projection cartographique à un autre.

D. Établissement de cartes plus précises : systèmes de référence géodésique

A2.16. Un autre facteur vient compliquer la conversion des coordonnées sphériques de latitude et de longitude : le fait que la Terre n'est pas une sphère parfaite avec un rayon constant. Des mesures précises montrent que la surface de la Terre est très variable et qu'elle se modifie constamment. Autre facteur encore plus important : la Terre est aplatie aux pôles de telle sorte que la distance entre le centre de la Terre et le pôle nord (demi petit axe) est plus courte que celle entre le centre de la Terre et l'équateur (demi grand axe). Pour établir des cartes précises, il est plus exact de considérer le globe comme un ellipsoïde ou un sphéroïde avec une relation déterminée entre le rayon polaire et le rayon équatorial (voir figure A2.7). Les paramètres décrivant l'ellipsoïde ainsi que l'origine et l'orientation du système de coordonnées utilisé pour référencer des objets cartographiques sont qualifiés de système de référence géodésique (du nom de la science qui a pour objet la mesure de la Terre : la géodésie).

Figure A2.7

Sphère et ellipsoïde



A2.17. Les paramètres les plus appropriés utilisés pour se rapprocher de l'ellipsoïde varient suivant les régions du monde. C'est pourquoi des centaines de systèmes de référence géodésique ont été mis au point. Heureusement, chaque service cartographique national n'utilise qu'un seul système de référence géodésique standard pour toutes ses activités cartographiques et seuls quelques-uns de ces systèmes sont utilisés pour la cartographie régionale, continentale ou mondiale. Des complications apparaissent lorsque le service cartographique modifie le système de référence géodésique standard. Ces systèmes ont constamment été perfectionnés au cours des deux derniers siècles. En conséquence, les anciennes cartes concernant une région peuvent être basées sur un système de référence géodésique tandis que les nouvelles ont été établies avec un nouveau système plus exact.

A2.18. Pour les cartes à petite échelle couvrant une grande région ou pour établir des croquis cartographiques dans des applications qui n'exigent pas une grande exactitude, les problèmes posés par des systèmes différents demeurent négligeables. Mais pour établir des cartes plus précises à de plus grandes échelles, l'écart peut être assez important. Le tableau A2.1 montre les coordonnées du bâtiment du Siège de

l'ONU dans le système de coordonnées de la projection transversale universelle de Mercator (UTM), qui est examiné plus en détail ci-dessous. Les coordonnées de latitude et de longitude du bâtiment des Nations Unies ont été projetées selon la même projection en utilisant différents systèmes de référence géodésique. L'écart nord-sud entre les anciens sphéroïdes de Clarke, qui étaient la norme aux États-Unis jusqu'à une date récente et le système géodésique mondial plus récent est d'environ 300 m sur le sol ou plus d'1 cm sur une carte à l'échelle du 1/25 000. Si l'on considérait la Terre comme une sphère parfaite plutôt qu'un ellipsoïde, on aurait un écart de plus de 18 km.

Tableau A2.1

Coordonnées projetées du bâtiment du Secrétariat de l'ONU à New York utilisant différents ellipsoïdes de référence

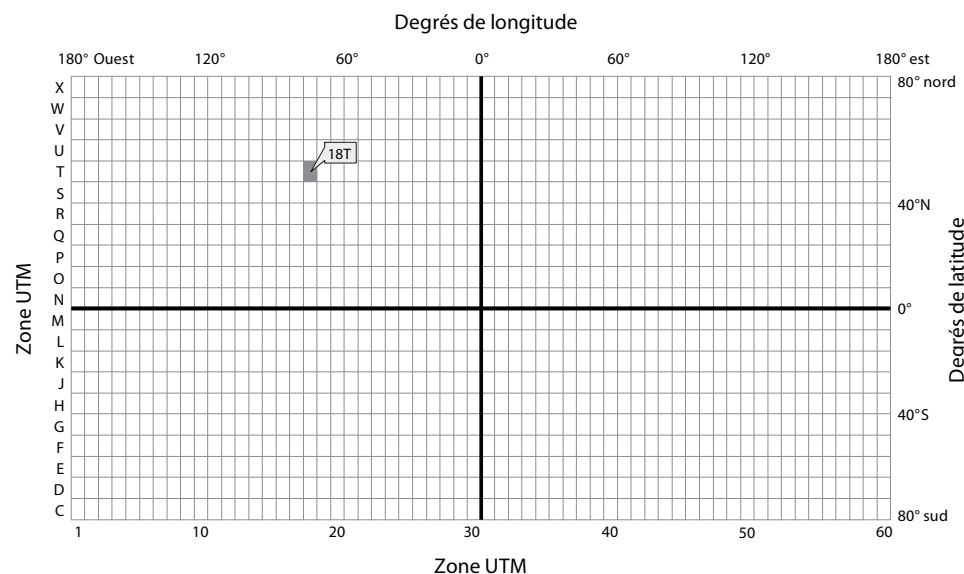
Ellipsoïde de référence	Coordonnées UTM (mètres)	
	Abscisse (x)	Ordonnée (y)
Clarke 1866	587 141,3	4 511 337,1
Clarke 1880	587 142,6	4 511 245,1
WGS84	587 139,0	4 511 549,7
Bessel	587 128,5	4 511 095,4
Sphère	586 917,2	4 529 920,6

Système de projection transversale universelle de Mercator (UTM)

A2.19. Un système de référence cartographique mérite d'être analysé plus en détail : le système UTM. C'est l'un des systèmes les plus couramment utilisés pour la cartographie à grande échelle dans le monde entier. Il est fondé sur une projection cylindrique transversale (projection transversale de Mercator) dans laquelle le cylindre est en contact avec le globe le long d'un méridien. Un méridien « local » différent est choisi pour différentes parties du monde. Les distorsions d'échelle, de forme et de distance le long de cette tangente sont très faibles. Le système UTM mondial comporte soixante zones de longitude (voir figure A2.8).

Figure A2.8

Le système UTM



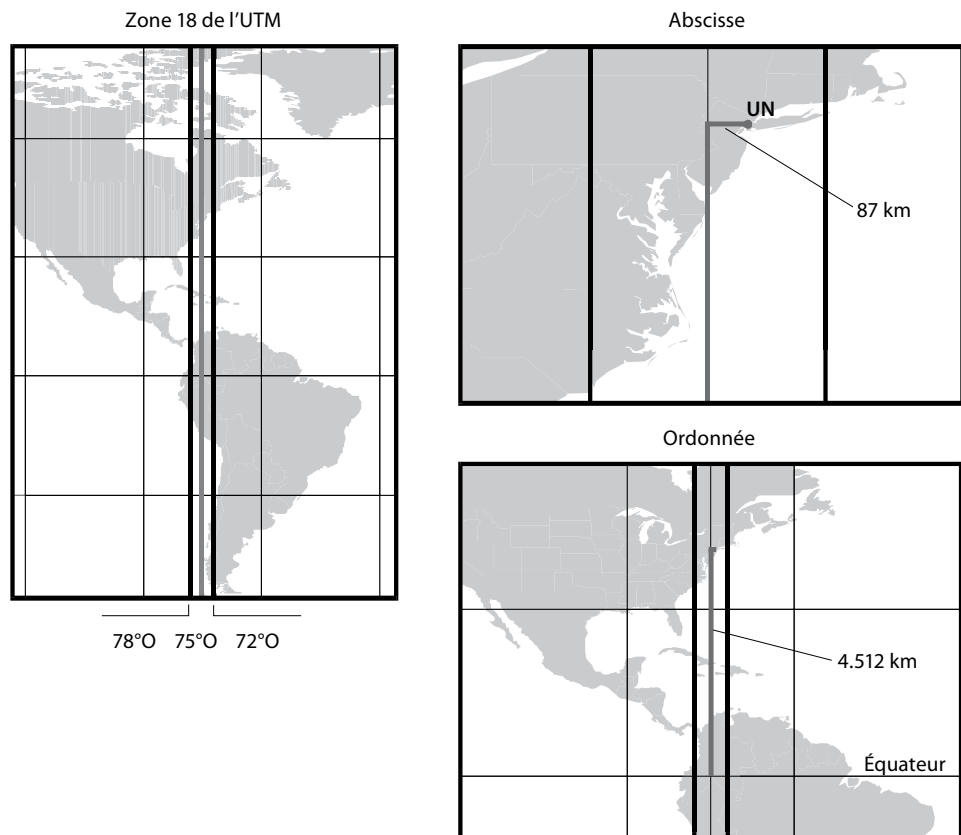
A2.20. Chaque zone a une largeur de six degrés de longitude, soit trois degrés dans chaque direction à partir du méridien tangent. Les zones du système UTM sont numérotées séquentiellement d'ouest en est à compter de 1 pour la zone comprise entre 180° O et 174° O, ayant pour méridien central 177° O. Les zones sont de plus divisées en rangées d'une hauteur de 8° . Des lettres leur sont affectées du sud au nord à partir de 80° sud avec la lettre C. La distorsion aux pôles étant très importante, aucune zone UTM n'est définie pour les régions situées au-delà de ces limites.

A2.21. Les coordonnées sont mesurées en mètres (ou en pieds) à partir du méridien central en abscisse dans la direction est-ouest et en ordonnée dans la direction nord-sud. Pour éviter les nombres négatifs, on ajoute 500 000 à l'abscisse. Pour la même raison, on ajoute 10 millions à l'ordonnée, mais seulement pour les coordonnées dans l'hémisphère sud. Ces décalages sont qualifiés de « fausse ordonnée ».

A2.22. Pour illustrer l'utilisation du système UTM, un exemple est présenté sur la figure A2.9. Le bâtiment du Siège de l'ONU à New York est situé à $40^{\circ}45'01''$ de latitude nord et $73^{\circ}58'04''$ de longitude ouest. Cet emplacement se situe dans la zone 18T du système UTM, qui s'étend du 72° au 78° ouest et du 40° au 48° nord. Les coordonnées x et y du système UTM exprimées en mètres sont 587139,0 et 4511549,7. Cela signifie que le bâtiment des Nations Unies se trouve approximativement à 87 km à l'est du méridien central de la zone 18 de l'UTM (75° O) et à environ 4.512 km au nord de l'équateur.

Figure A2.9

La situation du bâtiment du Siège de l'ONU dans le système de référence UTM



E. Échelle cartographique

A2.23. Il existe de grandes différences entre les cartes du point de vue de la surface au sol qu'elles couvrent. Les cartes nationales ou régionales ne montrent que les objets les plus importants, tandis que les cartes locales montrent de nombreux détails tels que des maisons individuelles ou de petits ruisseaux. La dimension ou la surface couverte sur une feuille de carte standard ou sur un affichage numérique est déterminée par l'échelle cartographique choisie pour tracer la carte. Cette échelle est représentée par une fraction reliant la distance sur la carte à la distance du monde réel sur le sol. Par exemple, sur une carte topographique au 1/25 000, 1 cm sur la carte représente 25 000 cm ou 250 m dans le monde réel.

A2.24. L'échelle de la carte étant une fraction ou un rapport, plus grande est la distance au sol représentée, plus faible est l'échelle de la carte. Par exemple, une carte à l'échelle du 1/1 000 000 est une carte à petite échelle car 1 divisé par 1 million est un nombre très petit (0,000001). Une carte au 1/5000 a une échelle relativement grande car 1 divisé par 5000 est un nombre relativement plus grand (0,0002). Ainsi, les cartes à petite échelle montrent de grandes superficies, tandis que les cartes à grande échelle se concentrent sur des zones peu étendues. En pratique, petite et grande échelle sont souvent confondues parce que dans la langue courante, « grand » et « petit » se réfèrent à la surface couverte ou aux dimensions du phénomène plutôt qu'à la fraction. Les modèles du climat mondial, par exemple, sont souvent appelés modèles à grande échelle. Il est donc utile pour éviter les malentendus de mentionner de façon explicite l'« échelle cartographique ».

A2.25. Quelques échelles courantes de cartes :

Échelle cartographique	1 cm sur la carte représente	
1/5 000	50 m	grande échelle
1/25 000	250 m	
1/50 000	500 m	
1/100 000	1 km	
1/500 000	5 km	
1/1 000 000	10 km	petite échelle

A2.26. Compte tenu de la transition de cartes analogiques à des cartes géographiques numériques, il est important de souligner que les données géographiques numériques sont essentiellement dépourvues d'échelle. Une fois que les coordonnées qui définissent les objets géographiques sont entrées dans un SIG, elles peuvent être affichées à n'importe quelle échelle spécifiée. L'utilisateur peut agrandir la carte ou faire un zoom arrière en explorant les données, et pour cela il change d'échelle rapidement et progressivement. Néanmoins, il importe d'avoir présent à l'esprit que les données ont été probablement tirées d'un document source (cartes, images, etc.) à une échelle initiale donnée. Des cartes imprimées à différentes échelles, par exemple, présenteront des degrés de détail différents. On pourra distinguer sur une carte au 1/25 000 les bâtiments individuels constituant un village. Par contre, sur une carte au 1/500 000, le village tout entier sera représenté par un point, si toutefois il figure sur la carte.

A2.27. Le processus par lequel des objets cartographiques sont simplifiés ou agrégés s'appelle la généralisation. Il joue un rôle important en cartographie. Du fait de la généralisation des objets — les routes rurales sinueuses deviennent des lignes droites, des détails des limites de districts disparaissent — il ne sert à rien d'imprimer au 1/5 000 une carte numérisée à partir d'une carte au 1/250 000 ou de combiner

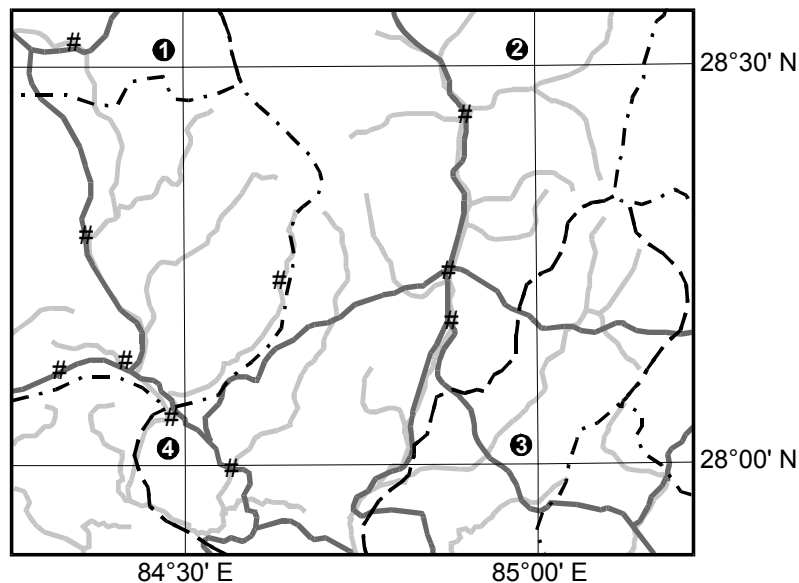
des ensembles de données numériques tirées de cartes à des échelles très différentes. Cela montre qu'il est très important d'indiquer l'échelle de la carte source dans les documents qui accompagnent un ensemble de données géographiques numériques. D'autre part, compte tenu de ces questions d'échelle cartographique, il importe dans le cas d'un projet de grande ampleur prévoyant l'élaboration de cartes numériques de déterminer suffisamment tôt les exigences d'échelle du produit à obtenir, pour que l'on puisse établir la base de données sur la base de documents sources adéquats.

F. Exemple de géoréférencement

A2.28. Le problème de géoréférencement d'une carte numérisée ou scannée dans des coordonnées unitaires appropriées en vue de son enregistrement dans un SIG a été examiné au chapitre II du présent *Manuel* dans la section consacrée à l'intégration des cartes numériques. On trouvera aux paragraphes ci-après un exemple réaliste destiné à illustrer l'opération du géoréférencement. La figure A2.10 ci-dessous montre une carte qui a été numérisée en plusieurs couches. Après numérisation, les coordonnées se trouvent référencées en unités de table à numériser, dans le cas présent des pouces. Afin d'utiliser la carte numérisée avec d'autres données numériques concernant cette région géographique, nous devons convertir les coordonnées du numériseur en coordonnées du monde réel qui correspondent à la projection initiale de la carte. Les lecteurs connaissant mal les systèmes de coordonnées et les projections cartographiques souhaiteront peut-être revoir les éléments figurant dans les sections A2.1 à A2.7 ci-dessus.

Figure A2.10

Points de contrôle sur une feuille de carte



A2.29. La première opération consiste à établir des points de contrôle bien définis. Elle fait en général partie du processus de numérisation. Les points de contrôle doivent être bien répartis à travers la zone présentant de l'intérêt pour améliorer l'évaluation des paramètres de transformation. Cela signifie qu'ils ne doivent pas être tous dans une

même zone ou au centre de la carte. Outre les routes, les cours d'eau, les circonscriptions administratives et les villes, la carte montre aussi un réseau régulier de lignes de latitude et de longitude à des intervalles d'un demi-degré. Les intersections de ce « graticule » permettent de définir commodément des points de contrôle, car leurs coordonnées sont faciles à déterminer. Sur la carte, les quatre points de contrôle choisis sont numérotés de un à quatre. Leurs paires de coordonnées sont respectivement (84,5–28,5), (85,0–28,5), (85,0–28,0) et (84,5–28,0). Il faut noter que parce que les programmes de SIG utilisent des coordonnées planaires, nous devons spécifier des paires de longitude/latitude (c'est-à-dire x/y) plutôt que la latitude/longitude. Pour la même raison, nous devons préciser les coordonnées en degrés décimaux plutôt qu'en degrés, minutes et secondes comme cela est courant pour les cartes sur papier ou dans les répertoires toponymiques.

A2.30. Malheureusement, nous ne pouvons pas utiliser les coordonnées de longitude/latitude directement pour cette transformation, parce que la carte originale sur papier n'a pas été enregistrée dans des coordonnées géographiques de latitude/longitude; très peu de cartes le sont et souvent cela est indiqué par le fait que le réseau de latitude et longitude n'est pas constitué de lignes droites. La projection originale de la carte dans cet exemple est la projection conique équivalente d'Albers, avec les paramètres suivants :

- Parallèles standard : 27° et 30° nord
- Méridien central : 84°
- Latitude d'origine : 28°

A2.31. Ces paramètres cartographiques sont en général mentionnés sur la feuille de carte. Avant de procéder à la transformation des coordonnées, nous devons convertir les coordonnées de longitude/latitude du point de contrôle en coordonnées correctes du monde réel dans la projection d'Albers. Dans la plupart des logiciels, cela peut être fait en répertoriant les paires de longitude/latitude (la longitude étant la coordonnée x et la latitude la coordonnée y) dans un fichier texte ou par l'intermédiaire d'une interface menus et en spécifiant les paramètres de projection pertinents dans le module de changement de projection du système.

A2.32. Naturellement, si nous pouvons lire les coordonnées du monde réel des points de contrôle directement sur la carte, cette opération supplémentaire est inutile. Cela est possible, par exemple, sur des cartes topographiques référencées dans la projection UTM. Il en va de même si les points de contrôle ont été déterminés sur le terrain au moyen d'un récepteur GPS qui convertit automatiquement les coordonnées dans une projection géographique particulière.

A2.33. Nous avons maintenant les quatre paires de coordonnées des points de contrôle exprimées en coordonnées de la table à numériser, ainsi qu'en coordonnées de projection du monde réel — dans ce cas mesurées en mètres. Les deux ensembles de coordonnées sont énumérés sur le tableau A2.2. Le premier point de contrôle, par exemple, est situé à environ 49 km à l'est du méridien central (84°E) et à 55,5 km au nord de la latitude d'origine (28°N).

A2.34. La troisième opération consiste à calculer les paramètres de transformation sur la base des deux ensembles de paires de coordonnées. La plupart des logiciels de SIG offrent cette option. Techniquement, les paramètres sont évalués en utilisant les équations de régression suivantes :

$$x' = a + bx + cy$$

$$y' = d + ex + fy$$

dans lesquelles x' et y' sont les coordonnées du monde réel et x et y sont les coordonnées au numériseur des points de contrôle ; a , b , c , d , e et f sont les paramètres à évaluer. Les erreurs d'évaluation dans cette transformation sont les résidus de la régression.

A2.35. Le tableau A2.2 indique pour chaque point de contrôle la paire de coordonnées dans le système de coordonnées d'entrée (unités de numériseur) et dans le système de sortie (projection d'Albers en mètres). En outre, le tableau indique les erreurs de transformation (résidus) que le système a calculées en unités de sortie (mètres). Nous constatons que l'erreur de transformation est d'environ 7,8 mètres dans le sens des x et d'environ 14,6 mètres dans le sens des y . Ces erreurs sont rarement égales à zéro. Les sources d'erreur peuvent être des déformations des cartes sur papier dues au rétrécissement et au pliage, ainsi qu'une erreur de mesure commise lors de la numérisation des coordonnées du point de contrôle. Une erreur très importante portant sur un ou plusieurs points de contrôle résulte en général d'une grave méprise, par exemple, la permutation des coordonnées x et y ou des identificateurs du point de contrôle. En général, cette opération doit être effectuée avec soin, car elle influe beaucoup sur l'exactitude et par conséquent l'utilité de la base de données de SIG.

A2.36. Le tableau fournit aussi une indication de l'erreur globale intervenue dans la transformation. Il s'agit de l'erreur quadratique moyenne indiquée en unités de coordonnées d'entrée et de sortie (respectivement en pouces et en mètres). Une erreur quadratique moyenne élevée indique que les emplacements des points de contrôle dans les unités des cartes d'entrée et de sortie ne correspondent pas aux mêmes emplacements relatifs. Pour un projet de conversion de données à grande échelle, il conviendra de définir et de ne pas dépasser une erreur quadratique moyenne maximale acceptable. Ce qui est considéré comme acceptable dépend de l'échelle cartographique des cartes imprimées originales et des exigences de l'application en ce qui concerne l'exactitude. La cartographie pour le recensement n'exige pas forcément un degré très élevé d'exactitude, par contre, les applications cadastrales, par exemple, doivent respecter des normes beaucoup plus élevées.

Tableau A2.2

Paramètres de transformation

Point de contrôle	Coordonnées en unités de numérisation (pouces)		Coordonnées en unités projetées du monde réel (mètres)		Erreurs calculées en unités du monde réel (mètres)	
	x	y	x	y	x	y
1	11,777	19,660	48936,2	55529,6	-14,59	7,80
2	26,670	20,661	97871,5	55835,2	14,60	-7,81
3	27,696	3,824	98333,0	409,3	-14,55	7,78
4	12,751	2,810	49166,9	102,3	14,54	-7,77
Erreur quadratique moyenne (entrée, sortie)			0,005034	16,524		

A2.37. Le système convertit en une seule opération toutes les coordonnées figurant dans la base de données cartographiques dans le système des coordonnées de sortie. La base de données de sortie est ensuite correctement référencée dans les coordonnées de la carte originale sur papier. Cette carte peut être ensuite projetée dans une projection cartographique différente, par exemple pour l'intégrer à une base de données détaillée dans une autre projection standard. Cette description avait pour objet de souligner les principes généraux de la transformation. Même si la mise en œuvre effective est liée à un logiciel particulier, le fait de comprendre les étapes du géoréférencement aide à apprécier l'importance de cette opération.

G. Considérations pratiques

A2.38. Tout projet de cartographie numérique important (par exemple établissement de cartes de recensement) exige l'intégration d'informations cartographiques provenant de nombreuses sources différentes. C'est pourquoi un système standard de projection et de coordonnées doit être choisi. Idéalement, le système de référence sélectionné devrait être le même que celui utilisé dans d'autres activités cartographiques dans le pays. La plupart des pays utilisent un système standard de projection et de coordonnées qui est optimal pour leur territoire national et pour la série de cartes nationales à différentes échelles.

A2.39. La quasi-totalité des logiciels de SIG comportent des fonctions de transformation de coordonnées d'un système de référence dans un autre (par exemple de mètres en pieds ou d'unités de numérisation en unités cartographiques), afin de convertir des cartes numériques du système latitude/longitude à la projection cartographique ou pour passer d'une projection à une autre. Ils permettent à l'utilisateur de choisir un système de référence géodésique et d'autres paramètres pertinents, quels qu'ils soient. Dans des cas rares, il arrive qu'ils ne puissent prendre en charge une projection particulière et un logiciel de projection particulier doit alors être utilisé. Les systèmes de positionnement universel (GPS), examinés en détail au chapitre IV, prennent aussi en charge certaines projections cartographiques et les systèmes de référence géodésiques les plus courants. Les coordonnées collectées pendant les travaux sur le terrain peuvent ainsi être saisies sous la forme de paires de coordonnées de latitude et de longitude ou dans un système de projection.

A2.40. Les informations relatives à la projection et au système de référence géodésique figurent généralement sur les cartes topographiques. Mais avec les ensembles de données cartographiques numériques un problème peut se poser du fait que les formats de SIG courants n'enregistrent pas nécessairement de façon explicite les informations relatives au système de projection. Par exemple, une agence de recensement peut se procurer un ensemble de données de SIG relatives aux routes ou à l'hydrologie sans informations complémentaires sur leur projection cartographique. Si ces données sont combinées aux cartes de recensement numériques, on risque de ne pas avoir de correspondance parfaite entre les unes et les autres. Ainsi, l'intégration verticale est impossible à moins que les deux ensembles de données soient convertis dans le même système de projection. Si l'on ne peut déterminer la projection cartographique en recherchant l'origine de l'ensemble de données dans les cartes sources, la seule possibilité consiste à harmoniser les deux cartes numériques en fonction des besoins, ce qui peut occasionner de sérieuses erreurs. Il importe donc que tous les ensembles de données soient accompagnés de documents appropriés et que les métadonnées — informations relatives aux données — soient conservées avec le lot de données de la carte numérique.

A2.41. Enfin, on examinera ici un aspect pratique concernant la conversion de différents formats d'enregistrement des coordonnées de latitude et de longitude. Celles-ci sont généralement exprimées en degrés, minutes et secondes. Celles du bâtiment du Siège des Nations Unies à New York sont par exemple 40°45'01" de latitude nord et 73°58'04" de longitude ouest. Pour entrer ces coordonnées dans un SIG ou un système de projection cartographique, nous devons d'abord les convertir en degrés décimaux. En pratique, cela les fait ressembler à des coordonnées normales cartésiennes x et y . Pour convertir des degrés, minutes et secondes en degrés décimaux, nous calculons par exemple de la façon suivante la latitude et la longitude du Siège des Nations Unies :

$$40 + \frac{\left(45 + \frac{1}{60}\right)}{60} = 40,7502778$$

$$73 + \frac{\left(58 + \frac{4}{60}\right)}{60} = 73,9677778$$

A2.42. Étant donné que la longitude du Siège des Nations Unies se situe à l'ouest du méridien de Greenwich, elle est indiquée sous la forme d'un nombre négatif en degrés décimaux (par exemple -73,97). De même, les valeurs de la latitude dans l'hémisphère sud sont également exprimées sous la forme de nombres négatifs.

A2.43. Pour convertir par exemple en sens inverse la latitude en degrés, minutes et secondes :

Degrés : 40

Minutes : $0,7502778 \times 60 = 45,016668 = 45$

Secondes : $0,016668 \times 60 = 1$

Annexe III

Modélisation de données

A. Introduction

A3.1. La présente annexe analyse les questions de modélisation de données géographiques et présente un exemple de contenu d'un dictionnaire de données détaillé qui peut être utilisé par un bureau de recensement pour fournir des informations sur les bases de données géographiques établies à des fins de recensement. On trouvera à l'annexe IV un dictionnaire de données plus simple destiné à accompagner les produits géographiques de recensement diffusés dans le public.

B. Définition des termes essentiels

A3.2. Un modèle de données spatiales décrit des entités géographiques telles que des habitations, des unités administratives ou des cours d'eau et les relations entre ces entités. Dans les modèles de données orientées objet, la définition inclut aussi habituellement les opérations effectuées sur les entités. Un modèle de données est indépendant de tout logiciel particulier. L'utilisateur peut donc mettre en œuvre le modèle de données dans n'importe quel logiciel de SIG complet.

A3.3. La structure de données spatiales met en œuvre un modèle de données particulier. Elle consiste en structures particulières de fichier qui sont utilisées pour représenter différents types d'entités. Par exemple, des unités administratives ou des étendues d'eau seront représentées sous la forme de polygones — c'est-à-dire une série de coordonnées dans laquelle la première est la même que la dernière. Une structure de données autorise des opérations logicielles qui définissent les relations entre les entités géographiques. Par exemple, une route peut coïncider avec une partie du périmètre d'un polygone qui définit une unité administrative.

A3.4. Le *format de données* est un terme plus général généralement appliqué à un ensemble spécifique de données dans le cadre d'un logiciel. Certains formats de données commerciaux ont été utilisés si couramment qu'ils sont devenus une norme de fait. Le format DXF (Drawing eXchange Format) par exemple a été élaboré initialement pour le logiciel AutoCAD. Il est maintenant pris en charge par la quasi-totalité des logiciels commerciaux de SIG.

A3.5. Un *dictionnaire de données* est un document maître qui décrit le modèle de données en détail, ainsi que les codes éventuellement utilisés pour identifier les entités et leurs attributs.

A3.6. Enfin, un *schéma de base de données* est une description des relations logiques entre les entités spatiales, les tables d'attributs et les règles d'intégrité qui définissent une base de données spatiales complète et détaillée.

C. Exemple de modèle

A3.7. L'exemple suivant est tiré de la définition très complète des définitions de l'entité géographique dans la *Base de données topographiques nationale – Dictionnaire de données* (Géomatique Canada, 1994).

Tableau A3.1

Informations réunies pour définir un modèle de données spatiales

Nom d'entité	Nom concis de l'objet géographique.
Définition	Description détaillée de l'entité géographique.
Attributs de domaine permanent	Attributs qui ne peuvent avoir qu'un nombre limité de valeurs prédéfinies. Par exemple, le type d'unité administrative (district, province, etc.), ou le type de surface d'une route. Ces codes prédéterminés constituent le domaine des valeurs possibles.
Attributs de domaine variable	Attributs qui ont un nombre potentiellement infini de valeurs possibles. Leur domaine peut donc ne pas être défini. On peut citer l'identificateur unique de l'unité administrative, la population de l'unité ou le nom d'un cours d'eau. Chaque attribut est décrit par les informations suivantes : • Nom; • Type par exemple alphanumérique (A), entier (I) ou réel (R); • Nombre de caractères ou de chiffres autorisés; • Le <i>domaine</i> des valeurs — c'est-à-dire une liste de toutes les valeurs possibles et de leurs définitions, pour les attributs de domaine permanent, ou la <i>définition</i> de l'attribut, pour les attributs de domaine variable.
Combinaisons autorisées de valeurs d'attributs	Pour les attributs de domaine permanent, toutes les combinaisons autorisées d'attributs sont énumérées. Par exemple, dans le cas d'unités administratives, seuls les districts et les provinces peuvent avoir un chef-lieu administratif officiel. Par conséquent, si le type d'unité administrative n'est pas le district ou la province, un autre attribut qui énumère les noms de chefs-lieux doit être vide. Les renseignements relatifs aux combinaisons autorisées de valeurs d'attributs sont utiles pour la vérification automatique de la cohérence. Si l'entité n'a pas de domaine permanent, on inscrira « aucun ». S'il n'existe qu'un seul attribut de domaine permanent, toutes les valeurs autorisées seront énumérées. S'il existe plusieurs attributs de domaine permanent, toutes les combinaisons de valeurs autorisées seront énumérées.
Relations	Il s'agit ici de décrire les relations que l'entité géographique peut avoir avec d'autres objets spatiaux. Cela est utile, par exemple, pour définir comment des cours d'eau ou des routes peuvent coïncider avec des limites d'unités administratives ou de districts de recensement. Les relations sont définies par les caractéristiques suivantes : • <i>Nom d'entité et géométrie de l'entité</i> — par exemple point (P), ligne (L) ou surface (A); • <i>Relation</i> — c'est-à-dire <i>raccordement</i>, dans le cas d'une ligne qui se raccorde à un point; ou <i>partage</i>, dans le cas d'une surface qui partage une limite avec une ligne; • <i>Cardinalité</i>, définie par une paire de valeurs définissant le nombre minimum et maximum de fois qu'une entité peut intervenir dans une relation. Par exemple, une intersection de routes est liée à des caractéristiques de routes. L'intersection doit comporter au moins une route qui s'y raccorde et elle peut en théorie être reliée à un nombre infini de routes. Si le nombre maximal ne peut être déterminé, il est représenté par N. La relation entre l'intersection routière et la route est donc (1, N); • <i>Nom et géométrie de l'entité connexe</i>. <i>Remarque</i> : Il faut noter que cela ne se réfère qu'aux relations entre des objets géographiques. Les relations entre des champs de la table d'attributs géographiques et des tableaux extérieurs doivent être définies séparément.

Représentation géométrique et dimensions minimales (mètres)	Il s'agit de la figure géométrique utilisée pour représenter l'entité. Dans le cas d'unités administratives, il s'agira presque toujours de surface (polygones). Cependant, pour d'autres objets, la représentation géométrique d'une entité spatiale pourra dépendre de l'échelle cartographique. Par exemple, un village pourra être représenté comme une surface présentant un périmètre aux grandes échelles cartographiques (par exemple 1/25 000), tandis qu'il prendra la forme d'un point aux petites échelles cartographiques (par exemple 1/250 000). À la même échelle cartographique, un village étendu ou une ville pourra être représentée comme une surface, mais un petit village prendra la forme d'un point. En fonction du type d'objet, la dimension minimale des entités pourra se référer à leur « superficie », « largeur », « longueur » ou « hauteur ».
Notes	Toute information complémentaire requise pour définir l'entité, ainsi que les notes de bas de page concernant l'un quelconque des autres champs descriptifs.
Diagramme	Pour indiquer comment une entité est modélisée, un graphique illustre les relations de cette entité avec diverses autres entités.

A3.8. L'information la plus importante dans le modèle de base de données est la définition de chaque entité et la description détaillée de tous les attributs enregistrés pour les objets géographiques. Pour de nombreux projets cartographiques de recensement, ces descripteurs essentiels de base de données peuvent suffire. Cependant, surtout si la base de données de recensement doit être incorporée dans une base de données nationale de SIG, il sera souhaitable de consacrer plus de temps et de travail à l'élaboration d'un modèle de base de données qui assure la compatibilité avec les informations fournies par d'autres organismes. Dans ce cas, les relations entre les circonscriptions administratives ou les unités de recensement et d'autres objets géographiques devraient être clairement définies.

A3.9. Pour préciser le contenu d'un dictionnaire de données, le tableau A3.2 donne un exemple décrivant la définition d'une couche de données d'une unité administrative. Ce tableau est présenté uniquement à titre d'exemple. La spécification exacte variera en fonction de la mise en œuvre dans chaque pays. Il convient de noter que les « attributs de domaine permanent » sont distincts et discontinus.

Tableau A3.2

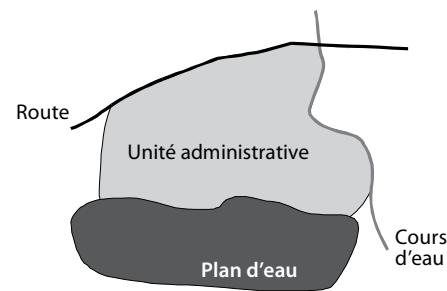
Exemple : unités administratives pour un pays comportant trois niveaux subnationaux

Unité administrative	
Une zone géographique aux limites définies par la loi créée en vue de l'exercice de fonctions publiques administratives et autres.	
Attributs de domaine permanent	
Unité administrative de type I (1) :	
1 — Province	Unité administrative de premier niveau
2 — District	Unité administrative subnationale de second niveau
3 — Localité	Unité administrative de troisième niveau
Indicateur rural/urbain I (1) :	
1 — Sans objet	Seules les localités sont classées dans la catégorie rurale ou urbaine
2 — Rural	Unité administrative aux caractéristiques surtout rurales
3 — Urbain	Unité administrative constituée d'une ville grande ou moyenne
Attributs de domaine variable	
Identificateur d'unité administrative I (14)	
<i>Note :</i> Dans cet exemple de base de données, tous les renseignements relatifs aux attributs (par exemple nom, autre nom, nombre de ménages, population, etc.) sont stockés dans des tableaux de données liés à la table des attributs géographiques par l'identificateur d'unité administrative.	

Combinaison autorisée de valeurs d'attributs			
Province : sans objet			
District : sans objet			
Localité : urbaine			
Localité : rurale			
Note : Seules ces combinaisons sont possibles. Par exemple, il n'y a pas de provinces urbaines ni de districts ruraux.			
Relations			
Unité administrative (P)	Part	(0, N)	Route (L)
Unité administrative (P)	Part	(0, N)	Cours d'eau (L)
Unité administrative (P)	Part	(0, N)	Étendue d'eau (P)
Note : Les routes et les cours d'eau sont représentés sous la forme de lignes (L) et ils peuvent coïncider avec certaines parties de la limite d'une unité administrative, qui est représentée par un polygone (P). De même, une limite administrative peut coïncider avec le rivage d'une étendue d'eau telle qu'un lac, qui est représenté par un polygone. (0,N) se réfèrent à la cardinalité de la relation. Cela signifie que par exemple, au minimum zéro route peut coïncider avec une limite d'unité administrative et que le maximum n'est pas déterminable (indiqué par N, qui signifie n'importe quel nombre).			
Représentation géométrique et dimension minimale			
L'unité administrative est représentée par un objet ponctuel si sa superficie est inférieure ou égale à 1 kilomètre carré et par un objet superficiel si elle dépasse 1 kilomètre carré.			

NOTE : Les unités administratives doivent coïncider avec les limites de districts de recensement. Les unités administratives doivent couvrir complètement le territoire national. En d'autres termes, il ne peut exister aucune partie du territoire du pays qui ne soit pas attribuée à une unité administrative.

Diagramme



Annexe IV

Exemple de dictionnaire de données établi à des fins de diffusion

A4.1. On trouvera ci-après un exemple de dictionnaire de base de données visant à permettre la diffusion d'une base de données géographique censitaire concernant des localités dans le pays hypothétique de Poplandia. L'expression « dictionnaire de données » est parfois remplacée par « métadonnées » alors qu'en pratique ces deux concepts sont différents. Les dictionnaires de données sont antérieurs aux métadonnées et font généralement référence aux informations distribuées avec les données qu'une agence diffuse. Les normes ISO en matière de métadonnées fournissent des directives spécifiques. Chaque pays peut utiliser le « profilage » pour adapter un format de métadonnées commun à ses propres applications nationales. L'exemple ci-dessous n'est présenté qu'à titre indicatif. Le contenu effectif du dictionnaire de données doit être établi avec soin par l'office national du recensement pour prendre en compte des questions spécifiques au pays considéré.

Dictionnaire de données : base de données censitaires de SIG concernant des localités

Titre de la base de données	Base de données géographique de recensement concernant des localités de Poplandia
Source	Organisme national de statistique, Service du recensement, section cartographique (1996), recensement national de la population et de l'habitat de Poplandia, 1995.
Contenu de la base de données	La base de données est constituée d'une couche de données de SIG concernant les localités du pays tout entier. Cette base de données est diffusée dans le format « shape file » d'ArcView (Environmental Systems Research Institute, Inc.), le format Interchange de MapInfo (MapInfo, Inc.) ou sous la forme d'un simple fichier texte de coordonnées. La présente documentation se réfère à la version shape file d'ArcView. La table de données des attributs géographiques de la couche de données SIG des localités (LOC.DBF) contient seulement des informations de base y compris le code de la localité (LOC_CODE) et les noms des unités administratives dans lesquelles elle se trouve. Deux tables de données extérieures sont diffusées avec la base de données de SIG, l'une contenant les caractéristiques démographiques tirées du recensement (POP.DBF) et l'autre les attributs des ménages (HH.DBF). Ces tables de données peuvent être liées à la base de données de localité du SIG en utilisant le champ commun LOC_CODE. Sauf indication contraire, toutes les données concernent le recensement du 1^{er} juillet 2005.
Unité administratives et déclarantes	La base de données contient des informations concernant 1291 localités dans 9 provinces et 123 districts.
Besoins en logiciel et matériel	La base de données peut être visionnée avec n'importe quel logiciel cartographique de SIG ou de bureau qui peut importer des shape files d'ArcView ou des fichiers en format Interchange de MapInfo. La configuration minimale du système dépend du logiciel utilisé pour accéder aux données. En général, un ordinateur personnel compatible IBM sous 486Mhz ou plus avec au moins 8 Mo de mémoire vive suffira. On peut accéder à la base de données à partir du CD-ROM ou l'installer sur le disque dur de l'ordinateur. Elle exige 16 Mo d'espace disque.

Titre de la base de données	Base de données géographique de recensement concernant des localités de Poplandia
Format de distribution de la base de données	La base de données est distribuée sous forme non comprimée sur le CD-ROM et on peut y accéder directement.
Projection	Conique équidistante
Parallèles standard	20° nord et 60° nord
Méridien central	140° ouest
Unités des coordonnées	Mètres
Décalage des coordonnées	Aucun
Échelle de la carte source	Variable. La plupart des localités ont été délimitées sur des cartes au 1/25 000 ou plus, les localités rurales sur des cartes au 1/50 000 ou à une échelle inférieure.
Renseignements sur l'exactitude moyenne	D'après les renseignements fournis par l'agence cartographique nationale, l'exactitude moyenne estimative est de +/-100 mètres dans les zones urbaines et +/-30 mètres dans les zones urbaines.
Unités déclarantes non consécutives	Certaines localités sont constituées de plus d'un polygone. La table des attributs contient un champ (FLAG) qui a une valeur de 1 pour le plus grand polygone (le seul pour des localités constituées d'un seul polygone) et de 0 pour de petits polygones. Pour éviter les doubles comptages lorsque les données de recensement sont agrégées l'agrégation ne devra être opérée qu'après avoir sélectionné les localités dont la valeur de FLAG est égale à 1.
Produits connexes	L'organisme national de statistique publié des bases de données numériques de SIG comparables pour des districts de recensement. Le nombre de districts de recensement étant très élevé, il existe des bases de données de SIG séparées pour chaque province. On prendra contact avec l'office national de la statistique pour obtenir plus d'informations.
Références	L'organisme national de statistique (2005). <i>Rapport technique sur les activités de cartographie pour le Recensement national de la population et de l'habitat de Poplandia, 2005. Service du recensement. Section cartographique.</i> L'organisme national de statistique (2005). <i>Rapport méthodologique et administratif pour le Recensement national de la population et de l'habitat de Poplandia 2005. Service du recensement.</i> L'organisme national de statistique (1996). <i>Résultats du Recensement national de la population et de l'habitat de Poplandia 2005. Service du recensement. Section cartographique.</i>
Adresse	Organisme national de statistique, Service du recensement Section cartographique, service utilisateurs P.O. Box 9999 Tarota, Sambas Province Tél : 99-99-99999 Fax : 99-99-99998 Adresse électronique : geog@census.gov.xx Site Web : www.census.gov.xx

Fichiers de données géographiques

LOC.SHP — Base de données SIG des limites de localités

Nom du fichier :	LOC.SHP
Type de fichier :	ESRI Arc Shape File
Types d'objets :	Polygones

Fichiers associés :	LOC.DBF	Table d'attributs polygonaux (partie du shape file)
	POP.DBF	Indicateurs démographiques de recensement
	HH.DBF	Indicateurs de ménages pour le recensement
	LOC.SHX	Fichier géographique interne utilisé par ArcView

Fichiers de données attributs

LOC.DBF : caractéristiques de la localité

Nom de champ	Description	Définitions de champ	Intervalle	Codes	Valeurs manquantes
LOC_CODE	Code officiel de localité. Lien avec les tableaux de données externes pop.dbf et hh.dbf. Le géocode est établi en concaténant des identificateurs administratifs : province 2 chiffres + district 3 chiffres + localité 3 chiffres	Ent.8	Valeur positive	aucun	-999
AREA	Superficie de la localité en km ²	Réel, 6.1	Valeur positive	aucun	-999
FLAG	Indique si le polygone est le plus grand pour la localité. Pour les localités constituées de deux polygones ou plus, seul le plus grand ou le plus important aura une valeur égale à 1.	Ent.1	0-1	0-petit 1-grand	
URBAN	Indique si la localité est classée dans la catégorie urbaine ou rurale.	Ent.1	0-1	0-rurale 1-urbaine	-1
LOC_NAME	Nom de la localité	Char.25	Aucun	Aucun	s.o.
DIST_NAME	Nom du district	Char.25	Aucun	Aucun	s.o.
PROV_NAME	Nom de province	Char.25	Aucun	Aucun	s.o.
AREA_TOTAL	Superficie totale de la localité en km ²	Réel 10.3	Valeur positive	Aucun	-999
AREA_LAND	Superficie totale de la localité couverte par les sols en km ²	Réel 10.3	Valeur positive	Aucun	-999
AREA_WATER	Superficie totale de la localité couverte par de l'eau en km ²	Réel 10.3	Valeur positive	Aucun	-999

POP.DBF — Indicateurs démographiques de recensement

Nom du champ	Description	Définitions de champ	Intervalle	Codes	Valeurs manquantes
LOC_CODE	Code officiel de localité. Lien avec les tables de données attributs loc.dbf et hh.dbf.	Ent.8	Valeur positive	Aucun	-999
POP_TOT	Population recensée totale	Ent.7	Valeur positive	Aucun	-999
POP_DENS	Densité démographique en habitants par km ² (POP_TOTAL / AREA)	Réel 5.1	Valeur positive	Aucun	-999
...	...	...	...	...	...

HH.DBF — Indicateurs de ménages pour le recensement

Nom du champ	Description	Définitions de champ	Intervalle	Codes	Valeurs manquantes
LOC_CODE	Code officiel de localité.	Ent.8	Valeur positive	Aucun	-999
HH_NUM	Nombre de ménages.	Ent.7	Valeur positive	Aucun	-999
HH_HEAD	Sexe du chef de famille.	Ent.1	0-1	0-homme 1-femme	-1
...	...	...	...	...	...

Annexe V

Élaboration d'une carte thématique

A. Introduction

A5.1. La présente annexe propose un bref aperçu des éléments liés à la conception à considérer pour créer des cartes thématiques. Il est impossible de traiter de toutes les questions relatives au contenu informatif des cartes dans le cadre de cet aperçu; il convient dès lors de consulter un manuel au besoin. Les cartographes distinguent différents types de cartes. Les cartes générales servent de cadre de référence à des fins d'orientation. Elles présentent surtout les objets géographiques réels que l'on peut observer sur le terrain. Ces objets sont soit naturels — cours d'eau, montagnes, littoraux — soit artificiels, comme les routes ou les localités. Les cartes de référence montrent aussi des objets qui ne sont pas visibles sur le sol. Le meilleur exemple en est les frontières politiques et le réseau de référence montrant les latitudes et les longitudes. Les cartes topographiques entrent dans la catégorie des cartes générales ou des cartes de référence. Elles jouent un rôle important dans la cartographie des secteurs de recensement car elles fournissent des informations sur les objets que l'agent recenseur utilise pour s'orienter dans le secteur de travail qui lui est affecté.

A5.2. Les *cartes thématiques* présentent plus d'intérêt lorsqu'il s'agit d'exposer des résultats de recensement. Elles montrent la répartition géographique des phénomènes physiques et culturels qu'il n'est pas facile d'observer directement sur le sol. Elles peuvent être fondées sur des informations qualitatives ou quantitatives. On peut citer à titre d'exemple de cartes du premier type une carte montrant la répartition de la population par langue maternelle ou par religion. Les cartes thématiques quantitatives, appelées parfois cartes statistiques, fournissent par contre des informations sur les dimensions relatives des objets cartographiés. On peut citer à titre d'exemple une carte sur laquelle les symboles représentant les villes d'un pays sont gradués en fonction des dimensions de chaque ville. Un autre exemple est celui d'une carte sur laquelle des unités déclarantes telles que des districts sont ombrées en fonction de leur densité démographique. La plupart des cartes établies pour un atlas de recensement sont de ce type.

B. Principes de la conception des cartes

A5.3. Bien que fréquemment utilisées à des fins d'analyse, les cartes se prêtent mal à la présentation de données exactes. Sur une carte, les valeurs de données sont exprimées par des symboles. Le cartographe doit affecter les valeurs de données à des intervalles de classes pour obtenir un nombre raisonnable de catégories représentées sous forme de couleurs ou de symboles. Cela signifie qu'une partie des informations est perdue dans la présentation de la carte. Si les cartes permettent de présenter correctement des tendances, des grandeurs relatives et la répartition des valeurs d'indicateurs, on obtient de meilleurs résultats avec les tableaux de données ou les cartes numériques dont la base de données peut être interrogée si l'on est intéressé par des valeurs exactes.

A5.4. L'établissement de cartes de présentation est l'opération par laquelle le cartographe communique une idée ou un concept au lecteur (Monmonier, 1993). Elle est comparable à d'autres modes de communication d'informations qualitatives ou quantitatives sous forme graphique au moyen de diagrammes, d'images ou d'autres illustrations. En conséquence, les mêmes principes de conception qui président à la conception graphique s'appliquent aussi à la cartographie.

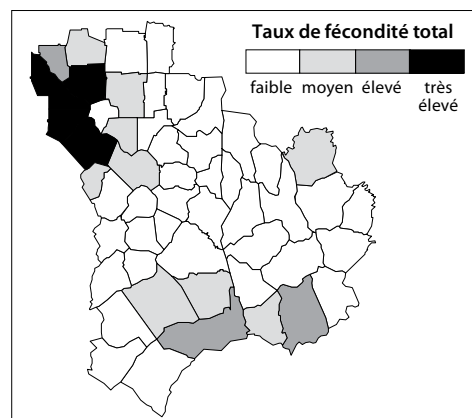
A5.5. La simplicité et la clarté sont les principes essentiels en matière de conception. De nombreuses cartes sont encombrées de détails parce que le cartographe a voulu présenter trop de choses sur un petit espace. Tufte (1983) offre un concept utile avec son rapport « maximisation des données/consommation d'encre » : appliqué à la cartographie, il signifie que l'encre utilisée doit servir à représenter des données géographiques plutôt que des informations non essentielles. Il faut donc renoncer à fournir des informations superflues. Les titres commençant par « carte de ... » ou « légende » sont inutiles, de même que de nombreux encadrés, lignes et souvent mais pas toujours, les flèches indiquant le nord et l'échelle. Comme la plupart des principes, celui-ci a naturellement ses limites. Certains éléments de la carte, comme la légende elle-même, un titre concis et des informations sur la source sont évidemment requis pour comprendre la carte.

A5.6. La simplicité signifie aussi qu'aucun espace ne doit être perdu. Aujourd'hui, on dispose presque partout d'imprimantes laser à haute résolution et par conséquent les cartes n'ont pas besoin d'être imprimées en très grand format pour montrer tous les détails. La carte peut être imprimée dans un format d'autant plus réduit qu'elle est mieux dessinée. Mais économiser l'espace signifie aussi éviter des polices, des symboles de légendes ou des encarts disproportionnés.

A5.7. La notion de hiérarchie visuelle est un autre concept important. Elle s'applique à tous les éléments figurant sur la carte elle-même ainsi qu'à la disposition de toutes ses composantes. Sur la carte elle-même, le choix des couleurs ou des symboles reflète l'ordre de valeurs des données. Sur une carte de mortalité infantile, par exemple, les unités déclarantes présentant les valeurs les plus élevées pourront être ombrées avec la couleur la plus brillante ou la nuance de gris la plus sombre. Ce sont les points chauds qui doivent retenir immédiatement l'attention du lecteur. Sur la figure A5.1, par exemple, les catégories de taux faibles à élevés sont volontairement ombrées en nuances de gris clair pour faire ressortir la catégorie des « taux très élevés ». C'est le contraste entre les couleurs sombres et les nuances plus délicates qui crée la hiérarchie visuelle. Une zone relativement claire entourée de teintes sombres ressortira tout autant. Le choix des couleurs est examiné ci-après plus en détail.

Figure A5.1

Établissement d'une hiérarchie visuelle par le choix des couleurs ou des nuances de gris



A5.8. Le cartographe peut aussi recourir à d'autres stratagèmes pour diriger l'attention du lecteur vers une zone particulière de la carte. Il peut entourer d'un contour précis les objets les plus importants sur la carte, par exemple, en les faisant ressortir par rapport à l'arrière-plan. On utilise parfois des annotations ou des flèches orientées vers certains objets, mais cela a souvent pour effet d'encombrer la carte.

A5.9. Les mêmes principes s'appliquent à la composition d'ensembles de la carte. Les éléments les plus importants de la carte sont l'information cartographique elle-même, le titre et la légende expliquant les symboles utilisés. Ils doivent être les éléments les plus apparents sur la page de la carte. On n'introduira d'autres éléments qu'avec prudence.

A5.10. L'utilisation de la couleur dans les cartes thématiques statistiques implique plusieurs considérations sous-jacentes (voir également par. A5.4 ci-dessus). Les cartographes doivent avoir conscience des corrélations négatives entre une présentation visuelle recourant à la couleur et les coûts supplémentaires liés à l'encre et à d'autres fournitures. Le daltonisme ne doit pas être perdu de vue. Les cartes devraient éviter l'utilisation du rouge et du vert pour des valeurs de données, puisque les daltoniens ne peuvent pas les distinguer normalement. Plus généralement, une carte couleur doit être reproductible sur une imprimante en noir et blanc sans perte du contenu informatif. Cela signifie qu'il faut respecter la densité de l'ombrage et remplacer les configurations de remplissage au besoin. Enfin, les cartographes doivent veiller à ne pas heurter une partie quelconque de la population par leurs choix de conception. Ils doivent être au fait des sensibilités des diverses régions ou groupes de la population. Certains symboles ou couleurs peuvent s'accompagner de connotations positives ou négatives pour différents groupes ethniques ou raciaux du pays. Lors de la conception d'une carte, les symboles associés à des stéréotypes concernant un sous-groupe quelconque de la population doivent être évités. Et surtout, les cartographes doivent être conscients que les couleurs doivent être utilisées judicieusement et non gratuitement.

1. Éléments d'une carte thématique

A5.11. Une carte thématique est constituée de plusieurs éléments. La carte elle-même consiste en une carte de base indiquant les limites des secteurs présentant de l'intérêt, par exemple, les frontières du pays et éventuellement des objets de référence tels que les principaux cours d'eau ou les grandes villes. Ils orientent le lecteur qui souhaite comparer la valeur d'une variable dans une partie du pays à cette même valeur dans une autre partie. Le deuxième élément principal est le recouvrement de la carte thématique qui présente la répartition géographique de la variable.

A5.12. Outre les données cartographiques réelles, une carte susceptible d'être publiée comprend d'autres éléments. Il peut s'agir des éléments suivants :

- Les **titres et les sous-titres** doivent être courts et très descriptifs.
- Les **sources de données, remerciements et date d'établissement de la carte** renseignent l'utilisateur sur la fiabilité et la crédibilité de la carte. Il convient par ailleurs d'ajouter les informations relatives aux droits d'auteur. Certaines institutions qui établissent régulièrement des cartes ajoutent aussi des numéros de référence et de version à usage interne. Il convient d'y ajouter des explications importantes pour permettre de comprendre le contenu de la carte. Dans le cas des cartes imprimées en grand format, les paramètres de la projection cartographique doivent aussi être indiqués.
- La **légende de la carte** décrit la façon dont les valeurs de la variable cartographiée ont été traduites en symboles cartographiques, par exemple, quelles

couleurs sont utilisées pour représenter un intervalle donné de valeurs de densité démographique. Il importe de toujours inclure les unités de mesure dans la légende, par exemple le « nombre d'habitants au kilomètre carré ».

- **L'échelle de la carte** permet au lecteur de mesurer les distances sur la carte. Pour une série de cartes thématiques comme celles d'un atlas de recensement, où toutes les cartes sont tracées à la même échelle, cette information n'a pas besoin de figurer sur chaque page. Il en va de même pour les cartes de dimensions relativement réduites de régions connues où il est peu probable que le lecteur veuille effectuer des mesures de distances. Il est en général préférable d'ajouter une échelle graphique plutôt que de spécifier une échelle numérique (par exemple 1/1 000 000). Si la carte est réduite ou agrandie pendant la photocopie, l'échelle graphique reste utilisable. Par contre, l'échelle cartographique nominale utilisée pour tracer la carte initiale sera erronée.
- Une **flèche indiquant le nord** n'est pas absolument nécessaire sur une carte de référence étant donné que toutes les cartes sont orientées vers le nord. C'est particulièrement vrai si la carte montre une zone géographique connue telle qu'un pays tout entier. Si l'on fait tourner la carte pour mieux l'adapter à la page, on devra dans tous les cas inclure une flèche indiquant le nord.
- Les **limites et bords intérieurs de cartes** servent à séparer différents éléments de la carte; l'utilisation de ces éléments graphiques est surtout une question de conception. On a une impression de fouillis si la carte comporte un trop grand nombre de lignes et d'encadrés. On ne devra donc utiliser des limites supplémentaires que si les éléments de la carte ne sont pas suffisamment séparés.
- Des **noms de lieux et des étiquettes** servant à identifier des objets géographiques ou de zones statistiques.
- Un **graticule**, c'est-à-dire le réseau des latitudes et des longitudes (parallèles et méridiens) qui facilite l'orientation sur la carte. Ces éléments doivent être inclus sur les cartes à petite échelle.
- Des **cartes de positionnement** sont utilisées pour indiquer l'emplacement de la zone couverte par la carte principale. Par exemple, une carte de densité démographique au niveau du district peut être accompagnée d'une petite carte montrant l'emplacement de ce district dans le pays ou la province.
- Les **cartons intérieurs** sont semblables aux cartes de positionnement. Mais au lieu de montrer l'emplacement de la zone couverte par la carte principale, ils montrent une portion réduite de la carte à une échelle cartographique supérieure. Par exemple, une carte au niveau d'une province peut être accompagnée d'un petit carton intérieur montrant la région de la capitale ou fournissant des informations plus détaillées sur un district moins étendu.
- Du **texte et des annotations** fournissent des informations ou des explications de base qui doivent être courtes et appropriées.
- Des **éléments graphiques supplémentaires** peuvent inclure un histogramme montrant la répartition géographique de la variable ou le symbole du bureau qui a établi la carte.

A5.13. Les figures A5.2 et A5.3 montrent deux exemples de cartes comprenant beaucoup d'éléments de la carte thématique. La figure A5.2. est une carte d'unités administratives de premier niveau dans le pays hypothétique de Poplandia. La carte est couverte d'un réseau de latitudes et de longitudes définissant les références géodésiques. La capitale nationale, les chefs-lieux administratifs de provinces et les principaux

cours d'eau sont ajoutés à titre de référence. Tous les objets sont étiquetés correctement en utilisant différentes polices pour divers types d'objets. La marge au-dessous de la carte comporte une échelle graphique, la légende indiquant les types d'objets thématiques représentés et l'origine de la carte. Si le bureau statistique dispose d'un logo, celui-ci peut également être ajouté sur chaque carte. Une flèche indiquant le nord a été omise pour deux raisons. La carte ne présente pas d'orientation particulière et les lignes représentant la longitude indiquent clairement que le nord se situe au haut de la carte. L'autre raison, moins évidente, est que la projection géodésique utilisée pour la carte fait converger les latitudes vers le nord. Cela sous-entend que le nord se trouve dans une direction légèrement différente aux différentes longitudes.

Figure A5.2

Échantillon de carte des unités administratives et des principaux centres urbains

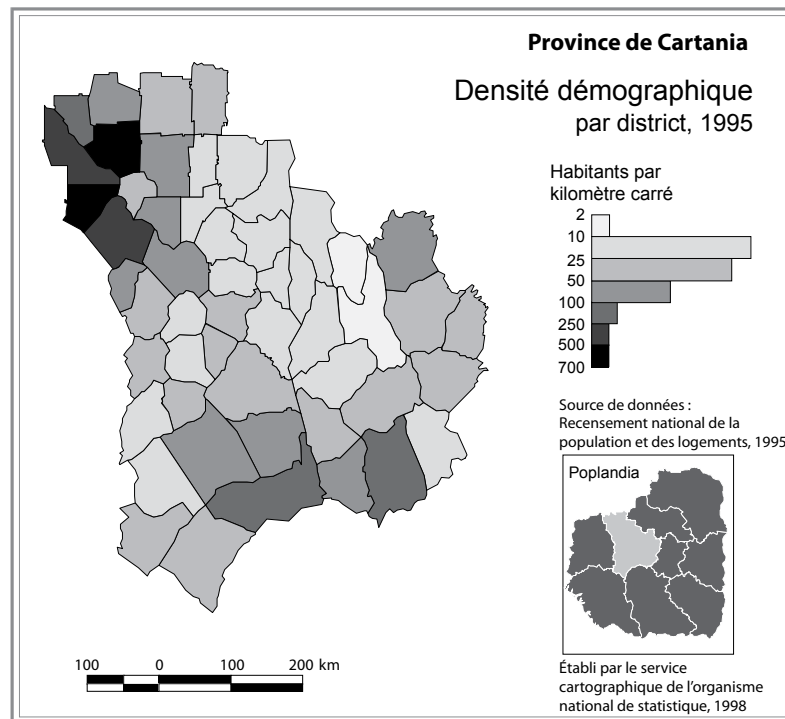


A5.14. La carte thématique représentée sur la figure A5.3. montre la densité démographique dans une des provinces de Poplandia. Une carte de ce type pourrait par exemple accompagner des tableaux montrant les caractéristiques démographiques par province dans une publication censitaire. Une telle carte pourrait par ailleurs comporter des graphiques ou d'autres représentations graphiques. Le tracé de la carte est très simple. Le titre décrit le thème de la carte et le sous-titre indique la zone géographique. Au lieu de la légende traditionnelle présentant des couleurs dans des cases de dimensions égales, la légende de cette carte montre les catégories de densité démographique sous la forme d'un histogramme. Celui-ci joue le rôle d'une légende traditionnelle — associant les valeurs à des nuances de couleurs — et il indique en outre la répartition

de fréquence des valeurs du district. Pour des cartes plus complexes, comportant plusieurs zones, on peut ajouter le nombre réel de districts relevant de chaque catégorie. Par souci de simplicité et de clarté, cela n'a pas été fait dans le cas de la présente carte. Au-dessous de la légende et de l'indication de la source, une petite carte de localisation montre l'emplacement de la province de Cartania dans le pays. Il est souvent inutile d'ajouter des annotations à une carte de localisation représentant le pays, parce que la forme du pays est facilement reconnaissable pour les lecteurs.

Figure A5.3

Exemple de carte thématique représentant la densité démographique



2. Niveaux de mesure et variables graphiques

a) Dimensions spatiales

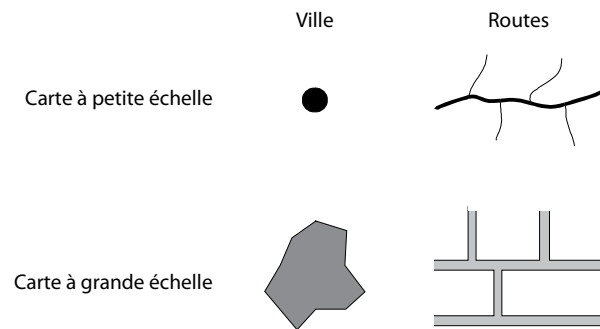
A5.15. Les cartes thématiques n'indiquent pas seulement l'emplacement d'un objet, elles fournissent aussi des informations concernant cet objet — la valeur de la variable en chaque point géographique. Une carte thématique est donc composée d'éléments géographiques et de certains attributs correspondant à ces éléments. Cela signifie qu'en dessinant une carte thématique, nous devons envisager la dimension spatiale des objets géographiques et nous devons avoir présent à l'esprit le niveau de mesure de chaque variable. Ces deux facteurs déterminent les options cartographiques disponibles pour produire une carte attrayante sur le plan visuel, précise et facile à interpréter.

A5.16. Les objets géographiques sont représentés dans une base de données géographique par des primitives géométriques : des points, des lignes et des surfaces. D'autres catégories, moins souvent utilisées en cartographie, ajoutent une troisième et une quatrième dimensions : le volume et l'espace-temps. La forme géométrique utilisée pour un objet du monde réel dépend parfois de l'échelle spatiale de la carte ou

de l'ensemble de données. Par exemple, on pourra représenter un village ou une ville comme des surfaces sur des cartes à grande échelle mais ils seront représentés sous la forme d'un point sur des cartes à plus petite échelle au niveau de la province ou du pays (voir figure A5.4). Une route pourra être représentée par une ligne sur une carte de province, mais elle prendra la forme d'une double ligne — c'est-à-dire un objet surfacique — sur une carte de ville.

Figure A5.4

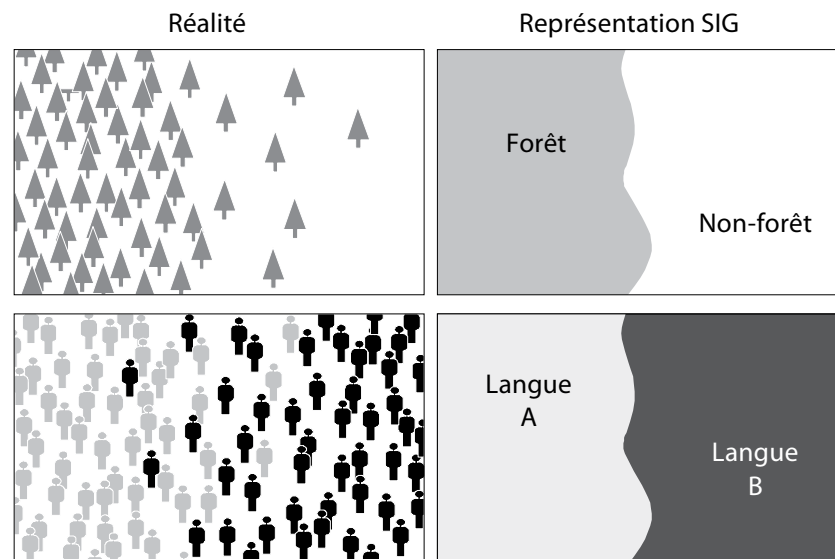
Effets de la généralisation sur l'affichage d'objets spatiaux



A5.17. Il importe de se souvenir que les limites et les emplacements ne sont pas toujours aussi clairement définis qu'ils le paraissent dans la représentation discontinue d'une carte ou d'une base de données géographique. Les objets complexes du monde réel ont souvent besoin d'être généralisés, simplifiés ou abstraits pour pouvoir être représentés dans la base de données d'un ordinateur. Par exemple, de nombreux objets du monde réel n'ont pas de limites précises. Il existe souvent une zone de transition entre la forêt et la non-forêt. Si la forêt est représentée sous la forme d'un objet surfacique (et non sous forme de points correspondant à chaque arbre), il y aura nécessairement une perte d'informations (voir figure A5.5).

Figure A5.5

La complexité du monde réel doit parfois être simplifiée pour faciliter sa représentation dans un SIG



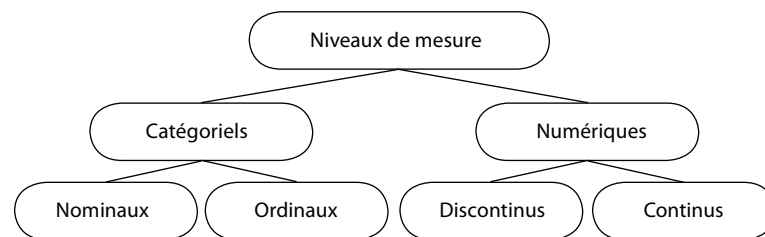
A5.18. Dans le domaine socioéconomique on a des exemples de limites mal définies avec la répartition des groupes ethniques ou linguistiques. Bien que le mode de répartition de ces groupes soit parfois très bien défini, il existera probablement à la périphérie de chaque région des zones où se mêlent les divers groupes ethniques ou linguistiques. Les cartographes utilisent parfois des lignes pointillées pour représenter ces limites mal définies, bien que cela ne résolve pas le problème de l'emplacement de la limite sur la carte.

b) Niveaux de mesure

A5.19. Il faut mentionner un autre facteur important : la façon dont est mesurée la variable dont on souhaite établir la carte. Il faut surtout distinguer les informations numériques de celles relatives à des catégories (voir figure A5.6). Les données relatives à des catégories se subdivisent à leur tour en données nominales ou ordinales. Les données nominales ou qualitatives décrivent simplement un type d'objet, sans qu'il y ait un ordre naturel entre ces catégories. On peut citer par exemple les différents types de maisons, par exemple les maisons en pierre ou en bois. Par contre, les données ordinales supposent une hiérarchisation des catégories, même si l'on ne connaît pas l'intervalle qui les sépare. Par exemple, à partir des réponses obtenues suite à une enquête, nous pourrions classer les ménages en fonction de leur niveau de bien-être, faible, moyen ou élevé. Mais nous ignorons si la différence entre un niveau faible et un niveau moyen est la même qu'entre un niveau moyen et un niveau élevé.

Figure A5.6

Mesure des variables



A5.20. Si nous pouvons quantifier la différence entre catégories, nous avons affaire à des données numériques. Les données discontinues correspondent à des dénombrements, par exemple le nombre de chambres par ménage, ainsi que la population totale. Les variations continues ou variables proportionnelles peuvent prendre toute valeur souhaitée. Elles peuvent être mesurées avec une grande précision. En ce qui concerne les données de recensement, les variables continues sont en général des indicateurs calculés pour des unités de recensement agrégées telles que la densité démographique, la proportion de population ayant accès à l'eau potable ou le taux de fécondité total.

c) Variables graphiques

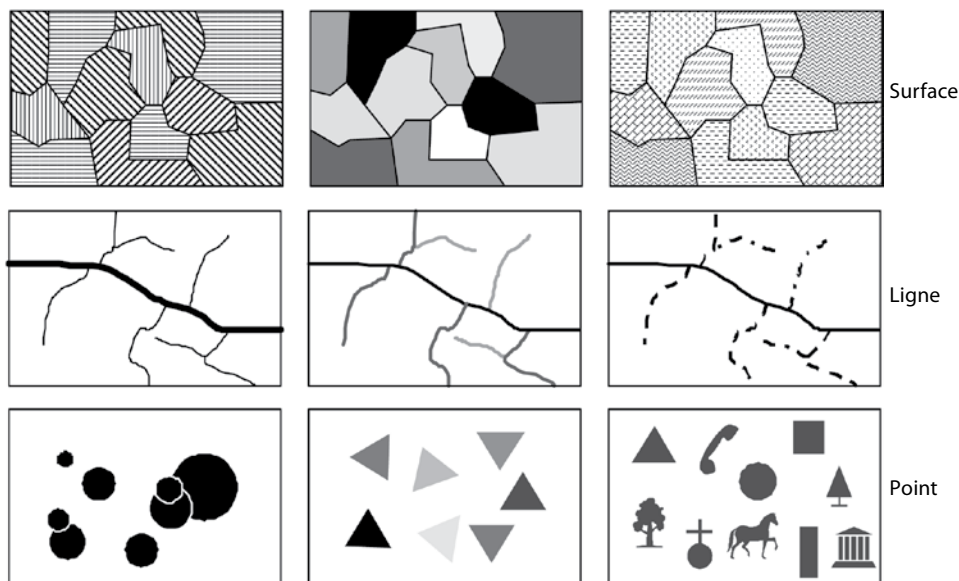
A5.21. Sur une carte thématique, les symboles graphiques indiquent au lecteur des différences de valeur ou de catégorie entre des objets géographiques. Les concepts de symbolisation appliqués à la cartographie sont semblables à ceux élaborés pour les applications infographiques par Bertin (1983; voir aussi MacEachren, 1995). Bertin distingue les variables graphiques suivantes :

- La **taille**, indicateur de différences ordinales ou numériques. La taille est très importante pour les objets ponctuels ou linéaires, pour indiquer par exemple la dimension des villes à l'aide de cercles gradués ou l'importance de la migration entre régions à l'aide de lignes ou de flèches d'épaisseur variable.
- L'**orientation** est utilisée, par exemple, pour les hachures d'objets surfaciques. De même, des objets ponctuels géométriques peuvent être présentés dans différentes orientations. L'orientation n'implique aucune différence concernant la grandeur d'une variable et elle est donc utile pour présenter des données nominales.
- La **texture** correspond à la densité d'un motif régulier qui varie suivant les zones. Elle peut servir à représenter des différences ordinales ou numériques. Ce peut être un raccourci utile lorsque les périphériques de sortie n'ont que des capacités limitées pour imprimer des couleurs ou des nuances de gris. La texture est aussi très utile pour représenter une information stratifiée dans laquelle deux variables sont figurées l'une au-dessus de l'autre. Mais il n'est pas facile de préserver la clarté de ces cartes qui sont donc utiles surtout pour des applications d'analyse exploratoire.
- La **forme** est très importante pour les objets ponctuels. Les ensembles de symboles et de polices présents dans les logiciels commerciaux de SIG et les logiciels de cartographie informatique comportent un grand nombre de symboles différents. Les symboles les plus connus en cartographie sont ceux qui représentent des bâtiments tels que des lieux de culte ou des hôpitaux.
- La **couleur** est tout à fait appropriée pour illustrer des différences numériques et dans une certaine mesure géographiques. Le choix de la couleur est l'une des questions les plus importantes en ce qui concerne la conception d'une carte et c'est pourquoi il est examiné plus loin de façon plus détaillée.

A5.22. En principe, chacune de ces dimensions graphiques peut servir à caractériser les divers types d'objets géographiques — à savoir les objets ponctuels, linéaires et surfaciques. Mais dans la plupart des cas, on utilise seulement un sous-ensemble de variables graphiques pour les différents types d'objets. La figure A5.7 en fournit

Figure A5.7

Variables graphiques correspondant à des polygones, des lignes et des points



quelques exemples. Les variables graphiques d'une carte thématique sont choisies de façon à correspondre au type de mesure utilisé pour l'indicateur faisant l'objet de la carte. Par exemple, les dimensions et la couleur sont très importantes pour représenter des valeurs numériques. Les formes des symboles ponctuels ou la texture des objets surfaciques représentent des valeurs nominales différentes.

3. Types de cartes thématiques

a) Représentation cartographique d'objets discontinus

A5.23. Les données de recensement compilées en vue de leur publication sont constituées de nombres agrégés par unité déclarante telle que le district ou le secteur de recensement. Le meilleur moyen de représenter ces données consiste dans l'utilisation de cartes choroplèthes. Le terme choroplèthe vient des mots grecs « choros » (lieu) et « plèthe » (valeur). Les cartes choroplèthes montrent des données correspondant à des unités déclarantes discontinues souvent définies indépendamment de la répartition spatiale effective des données (par exemple des limites administratives). Le symbole — c'est-à-dire la couleur ou la texture — utilisé pour ombrer chaque unité déclarante, est déterminé par la valeur. Les cartes choroplèthes diffèrent des cartes de catégories de zones pour lesquelles les unités déclarantes sont déterminées par les données. Par exemple, sur une carte représentant un couvert forestier, les unités déclarantes seront définies à partir des limites séparant les zones forestières et les zones non forestières.

A5.24. Un exemple de carte choroplèthe a déjà été présenté à la figure A5.3. Les cartes choroplèthes sont construites en subdivisant d'abord en un ensemble de catégories toute la série de valeurs de données correspondant aux unités déclarantes. On assigne ensuite une couleur ou un type d'ombrage à chaque catégorie. Étant donné que les données d'une énumération suivent un ordre naturel, le choix des couleurs ou des tons répond à une certaine logique, par exemple une progression des nuances de clair à sombre ou des motifs d'irrégulier à régulier. L'objectif consiste à donner au lecteur une idée intuitive de la grandeur de la valeur dans chaque unité déclarante. Il existe de nombreuses façons différentes de déterminer les symboles utilisés pour ombrer les cartes choroplèthes. Le choix dépend du type de variable, de l'intervalle des valeurs de données et aussi du périphérique de sortie utilisé pour présenter la carte. Le choix des symboles étant très important, il sera examiné en détail dans la prochaine section.

A5.25. Les cartes choroplèthes donnent de bons résultats lorsqu'il s'agit de représenter la répartition générale des valeurs de données sur une carte ou de comparer les répartitions entre différentes cartes. Il n'est en général pas possible d'obtenir la valeur exacte de chaque unité déclarante parce que les couleurs ou nuances représentent seulement des intervalles de valeurs semblables. Pour présenter les valeurs exactes, il est préférable d'utiliser des tableaux de données ou de procéder à une requête interactive dans un SIG.

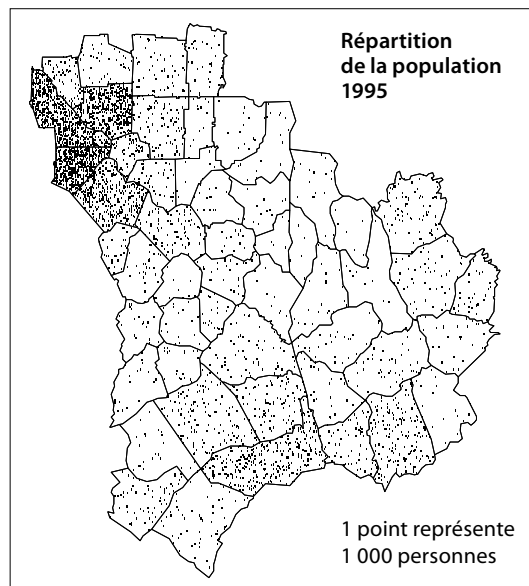
A5.26. Les valeurs utilisées pour produire des cartes choroplèthes sont presque toujours des coefficients, des proportions ou des densités. Il peut s'agir de coefficients géographiques, où une valeur de données telle que la population est divisée par la surface pour calculer la densité démographique. Ou il peut s'agir de coefficients généraux dans lesquels le dénominateur est une valeur autre qu'une surface, par exemple le taux de natalité brute exprimé par le nombre de naissances par 1 000 habitants. Le plus souvent, lorsqu'on établit une carte de variables socioéconomiques, la dimension des unités déclarantes n'est pas constante. Par exemple, il existe souvent de très grandes différences de dimensions et de population entre divers districts et provinces. Si nous voulions établir la carte d'une variable de dénombrement telle que la population totale

plutôt qu'un coefficient, les districts les plus grands seraient probablement ombrés avec les couleurs les plus sombres, même si leur population était peu nombreuse par rapport à leur superficie. Le recours aux cartes choroplèthes est donc inapproprié si l'on veut représenter des valeurs absolues.

A5.27. Une autre méthode utilisable pour afficher des données de dénombrement est celle des cartes de points. Ce type de carte a été utilisé pour la première fois en France en 1830 pour établir une carte montrant la répartition de la population. Sur des cartes à points, le symbole d'un point sert à représenter une ou plusieurs unités d'une variable cartographiée. Par exemple, chaque point peut représenter 10 000 personnes ou 1 000 ménages. La grandeur de la variable est alors représentée par la densité plus ou moins élevée des points dans l'unité déclarante. La figure A5.8 montre une carte à points typique, présentant une répartition de la population.

Figure A5.8

Carte de densité par points



A5.28. Pour placer les points sur la carte, deux méthodes sont applicables. Le cartographe, connaissant la répartition effective de la population dans chaque district, peut choisir l'emplacement des points à partir de celle-ci. Par exemple, il placera davantage de points dans les zones urbaines et autour de celles-ci que dans les régions rurales moins peuplées. Pour certaines applications, on a fait appel à des cartes d'utilisation des sols ou de couverture du sol pour aider à déterminer les densités de points dans chaque unité déclarante. On peut aussi recourir à des masques virtuels pour qu'aucun point ne soit placé dans des zones dont on sait qu'elles sont inhabitées, comme les nappes d'eau, les forêts très denses ou les réserves naturelles protégées.

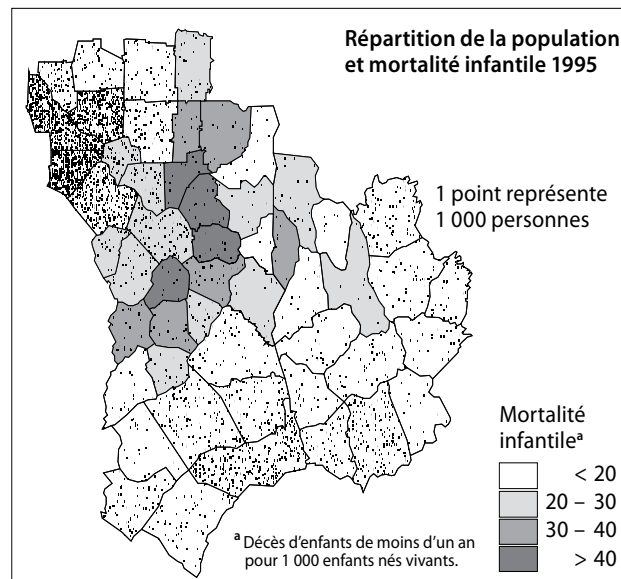
A5.29. La seconde possibilité consiste à placer les points aléatoirement dans chaque district. Dans ce cas, la densité des points reflète la densité d'ensemble de la valeur. Les logiciels de SIG et de cartographie sur ordinateur offrant des fonctions de représentation de la densité placent généralement les points de façon aléatoire. L'utilisateur ne peut choisir que la dimension de chaque point ou le symbole utilisé pour les points. Ce symbole peut être choisi de façon à représenter la variable faisant l'objet de la carte, mais un simple point offre généralement la représentation la plus claire.

A5.30. Certains programmes spécialisés écrits permettent de placer les points en recourant à d'autres couches de la carte. Le placement manuel des points compte tenu de la connaissance que le cartographe peut avoir de la variable est naturellement très fastidieux.

A5.31. Les cartes de points constituent un moyen efficace pour représenter les informations relatives à la densité, à condition que le placement des points soit guidé par la répartition géographique réelle de la variable représentée ou que la répartition dans chaque unité déclarante soit dans l'ensemble homogène. Cette méthode présente le grand avantage que les cartes de densité par points se prêtent bien à la reproduction, qu'elles soient photocopiées ou imprimées, car il s'agit essentiellement de cartes monochromes (en noir et blanc). Les cartes de densité par points peuvent être utilisées en combinaison avec des cartes choroplèthes pour illustrer simultanément deux variables — par exemple la carte de la figure A5.9 montre qu'il n'existe pas de relation entre les densités démographiques élevées et les taux élevés de mortalité infantile. Dans ce cas, la densité des points ne devra pas être très élevée pour que l'on puisse facilement déterminer les couleurs ou les nuances correspondant aux districts sous-jacents.

Figure A5.9

Combinaison de cartes de points et de cartes choroplèthes



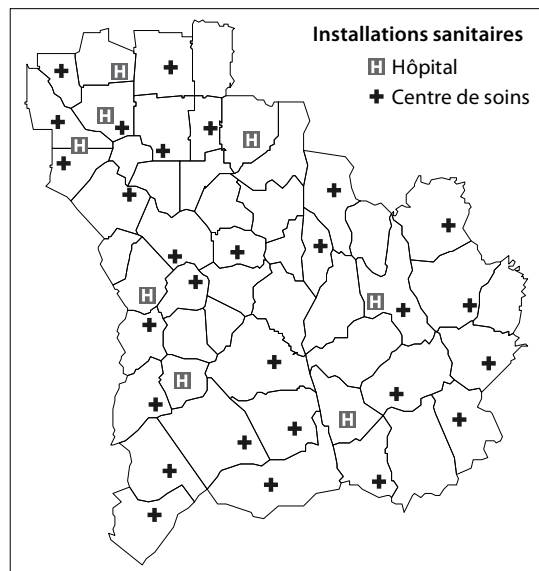
b) Données ponctuelles nominales

A5.32. Le type le plus simple de carte de points est celui où chaque point représente un élément distinct comme une exploitation agricole ou un hôpital. Ces données ponctuelles nominales représentent des catégories d'objets plutôt qu'un attribut compte ou dimension. Dans les cartes de symboles ponctuels simples, l'emplacement du point représente correctement l'emplacement de l'élément. La dimension, la couleur ou le symbole utilisé peuvent refléter divers types d'objets, tels que des centres de services de santé distincts des hôpitaux, comme sur la figure A5.10. On pourra utiliser des figures géométriques simples comme des cercles, des carrés et des triangles pour représenter divers types d'objets ponctuels. Inversement, les logiciels de cartographie numérique ou de SIG permettent à l'utilisateur de spécifier un symbole qui corresponde au type d'objet représenté. Par exemple, la carte figurant sur la figure A5.10 montre la réparti-

tion de deux types d'installations sanitaires au moyen de symboles faciles à interpréter. Les symboles utilisés sont généralement des caractères d'une police de caractères ou des dessins en mode point. La plupart des logiciels ont leurs propres polices qui offrent un grand nombre de symboles cartographiques classés par thème, comme les transports, les services collectifs ou établissements ouverts au public. Certains systèmes permettent aussi à l'utilisateur d'importer des symboles en mode point conçus par lui. Le mieux consiste toutefois à utiliser un jeu de symboles standard chaque fois que cela s'avère possible afin de réduire le besoin de spécialisation nécessaire pour la reproduction.

Figure A5.10

Représentation d'objets ponctuels distincts



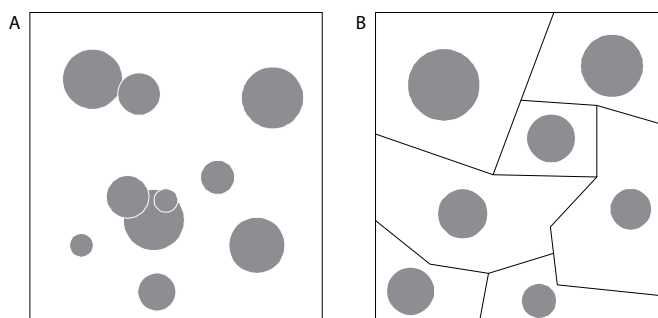
c) Symboles ponctuels proportionnels

A5.33 Les symboles ponctuels peuvent aussi servir à représenter une quantité en un certain lieu. Un type apprécié de carte de recensement, par exemple, montre l'emplacement et la dimension des grandes villes au moyen de cercles et de carrés gradués en fonction des valeurs numériques de chaque objet. Ces cartes sont qualifiées de proportionnelles ou de cartes de symboles gradués. Les cartes de symboles gradués sont utiles pour représenter la valeur absolue d'une variable. Elles le sont moins pour représenter une valeur relative telle qu'une densité ou un coefficient.

A5.34. Il existe deux types de cartes de symboles gradués. Dans un cas, les données se réfèrent à un objet ponctuel tel qu'une ville ou un ménage. Dans ce cas, l'emplacement du symbole correspond à l'emplacement de l'objet (voir figure A5.11A). Dans le second cas, les symboles sont utilisés pour représenter les valeurs d'objets surfaciques tels que des districts. Dans ce cas, on doit choisir un emplacement représentatif dans chaque unité déclarante (voir figure A5.11B). Il faut noter que la plupart des systèmes entourent chaque cercle d'un halo afin que l'on puisse distinguer des cercles très rapprochés les uns des autres. Le système trace les plus grands cercles en premier lieu pour éviter que les plus petits soient recouverts.

A5.35. Comme précédemment, un logiciel de cartographie numérique nous permettra de choisir un symbole correspondant au thème de la carte. Ces symboles figuratifs peuvent rendre la carte plus attrayante. Cependant, si les symboles sont trop complexes, ils risquent de détourner l'attention du lecteur de la principale information

Figure A5.11

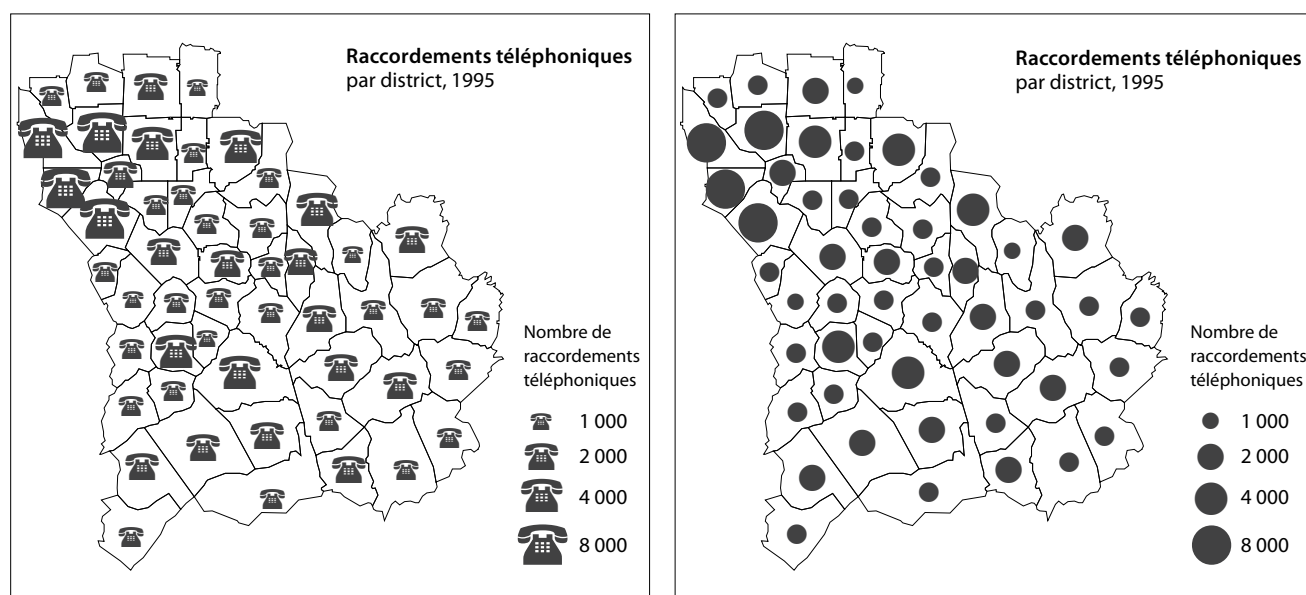
Symboles proportionnels correspondant à des objets ponctuels et surfaciques

qu'il s'agit de communiquer : la grandeur relative de la variable dans différentes régions. Comparons les deux versions d'une carte montrant le nombre de raccordements téléphoniques sur la figure A5.12. Bien que le symbole représentant l'appareil téléphonique soit très simple ; il est plus difficile d'apprécier la dimension de la variable sur la carte de gauche que sur la carte de droite plus simple. Le cartographe doit présenter l'information de façon claire et intelligible tout en s'efforçant de rendre la carte attrayante. Dans la plupart des cas, on obtiendra de meilleurs résultats avec des symboles simples qui ne détournent pas l'attention du lecteur de la grandeur relative de la variable étudiée.

A5.36. Des symboles proportionnels peuvent aussi être utilisés pour préserver deux variables en même temps. Par exemple, la dimension des cercles peut représenter le nombre de ménages dans une unité déclarante, tandis que la couleur ou la nuance de gris de chaque cercle indique le pourcentage de ménages ayant un raccordement au téléphone. Ici encore, le cartographe doit éviter de surcharger la carte d'informations. Si le nombre d'unités déclarantes est très élevé ou si les unités sont très petites, il peut être préférable de présenter les deux variables sur des cartes séparées.

A5.37. Outre les cercles, on utilise couramment des symboles géométriques tels que des carrés ou des triangles. En faisant varier l'orientation des triangles, on peut

Figure A5.12

Pictogrammes ou symboles graphiques simples

montrer la valeur de variables divergentes telles que l'immigration et l'émigration de chaque unité déclarante (voir figure A5.13). Des nuances de gris ou des couleurs différentes facilitent encore l'interprétation.

A5.38. D'autres cartes apparentées aux cartes de symboles gradués représentent les différences de valeurs à l'aide d'un certain nombre de symboles standardisés

Figure A5.13

Présentation de l'importance et de la direction des flux au moyen de symboles graphiques simples

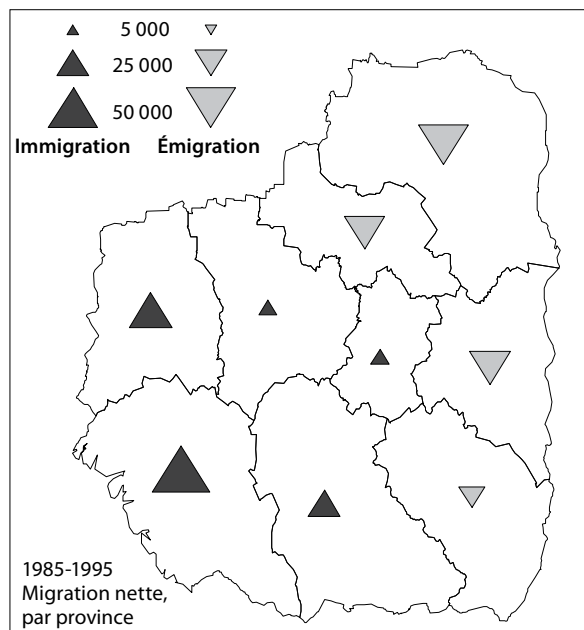
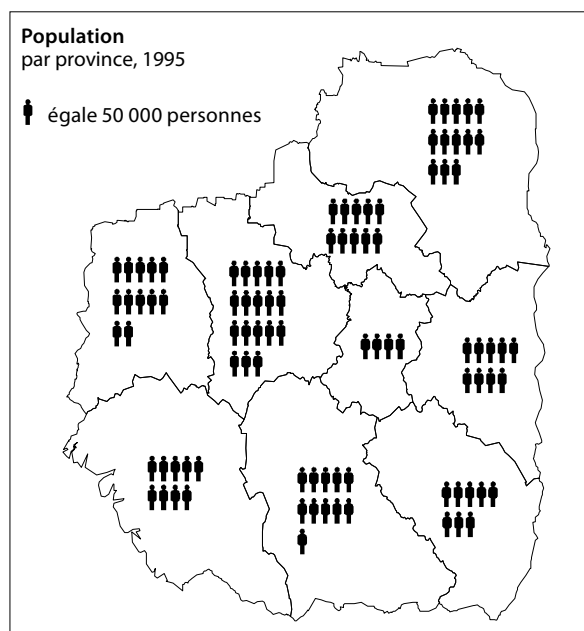


Figure A5.14

Représentation de valeurs de données faisant varier le nombre de symboles cartographiques pour chaque objet



tracés pour chaque unité géographique. Par exemple, la population totale peut être représentée comme sur la figure A5.14. Ce type de carte était très apprécié en cartographie thématique. Mais, de même que pour les symboles figuratifs, ces cartes sont facilement surchargées de détails et difficiles à interpréter. Il est préférable de représenter la grandeur des différentes valeurs à l'aide de symboles proportionnels.

d) Cartes-diagrammes

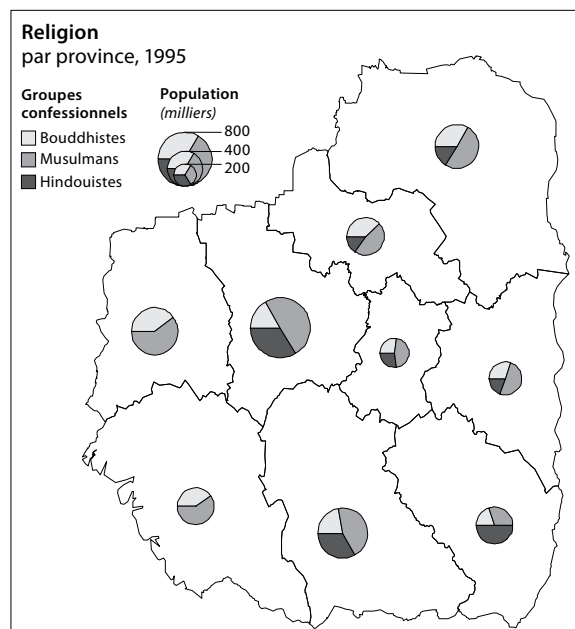
A5.39. Les cartes présentant des informations statistiques sous forme de diagramme pour chaque observation géographique rencontrent un grand succès du fait de leur disponibilité dans les logiciels commerciaux de SIG et de cartographie numérique. Comme plusieurs types de cartes examinés précédemment, les cartes-diagrammes sont facilement surchargées d'un trop grand nombre d'informations. Il existe malheureusement beaucoup d'exemples de cartes de ce type dont il est difficile ou impossible d'extraire des informations.

A5.40. Les types les plus courants de cartes-diagrammes utilisent des graphiques à secteurs, à barres ou à colonnes. Les diagrammes sont en général échelonnés de telle sorte que la dimension de chaque graphique à secteurs, par exemple, reflète l'importance du dénominateur. Par exemple, la figure A5.15 montre la répartition géographique de la proportion des principaux groupes confessionnels. Les diagrammes sont échelonnés en fonction de la population totale. Nous devons donc présenter deux types d'information dans la légende : la couleur qui se réfère à chaque groupe confessionnel et les totaux de population qui correspondent à une dimension donnée du graphique à secteurs.

A5.41. Les cartes-diagrammes donnent les meilleurs résultats lorsqu'il y a relativement peu d'observations géographiques et très peu de groupes à représenter. Par exemple, une carte-diagramme à secteurs qui comporte seulement deux catégories peut être très utile en combinaison avec une carte choroplèthe simple pour montrer plusieurs variables simultanément (voir figure A5.16) : La répartition spatiale des divers niveaux

Figure A5.15

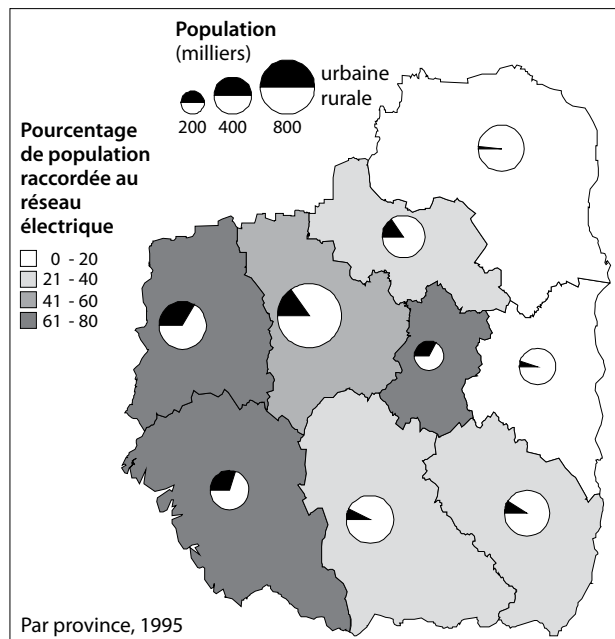
Carte-diagramme à secteurs



d'accès au réseau électrique, la population totale dans chaque province et la proportion de la population rurale ou urbaine. Sur cette carte, il semblerait que les provinces ayant une proportion élevée de population urbaine ont un coefficient plus élevé de raccordements au réseau électrique. Une carte bien conçue qui n'est pas surchargée de symboles, de couleurs et de nuances peut faciliter l'analyse multivariable de différentes variables. Cependant, la carte-diagramme à secteurs et les cartes semblables peuvent facilement devenir difficiles à interpréter et leur utilisation devrait être limitée aux cas dans lesquels le message cartographique n'est pas encombré de trop de symboles et de catégories.

Figure A5.16

Combinaison de cartes choroplèthes et de cartes-diagrammes à secteurs

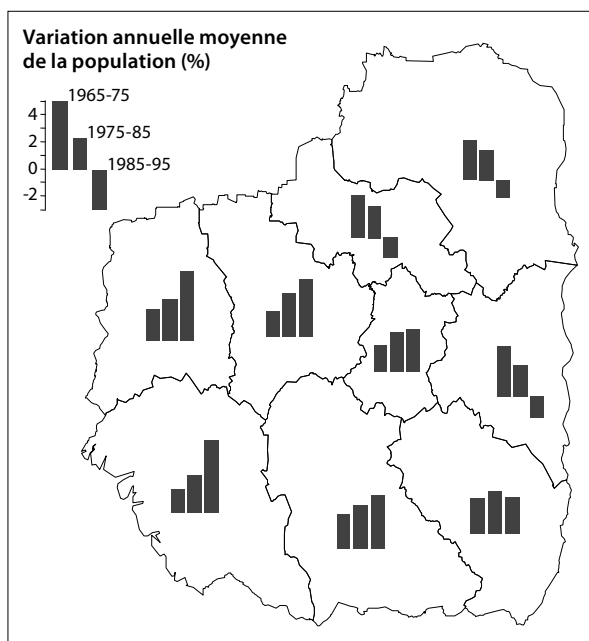


A5.42. Les cartes-diagrammes servent aussi à illustrer les tendances au cours du temps. La carte de la figure A5.17, par exemple, montre la variation annuelle moyenne en pourcentage de la population dans chaque province entre les trois derniers recensements. Les diagrammes à barres sont très simples, sans bordure et sans ligne de base, car pour ces données, il ne peut y avoir de doute quant aux barres correspondant à un accroissement ou à une diminution de la population. Comme précédemment, il s'agit en fait de montrer des changements relatifs dans le temps et non des valeurs exactes, qu'il est préférable de présenter sous forme de tableau.

A5.43. Un type de diagramme qui présente une grande importance en ce qui concerne les données de recensement de la population est naturellement la pyramide des âges. Les pyramides peuvent être combinées avec une carte de base des unités déclarantes pour montrer comment la répartition par âge et par sexe varie par région (voir figure A5.18). Les pyramides de population sont des diagrammes très complexes. Cela sous-entend qu'elles ne peuvent être représentées de façon satisfaisante que si le nombre de régions sur la carte est relativement faible. En général, cela signifie qu'elles ne seront présentées sur un atlas de recensement qu'au premier niveau subnational. Cependant, en pratique les logiciels commerciaux de SIG et de cartographie numérique ne produisent pas automatiquement des diagrammes-pyramides. Il faut donc les créer extérieurement, par exemple dans un tableur et les ajouter à une carte de base dans un logiciel graphique ou dans le module de mise en page d'un programme de cartographie numérique.

Figure A5.17

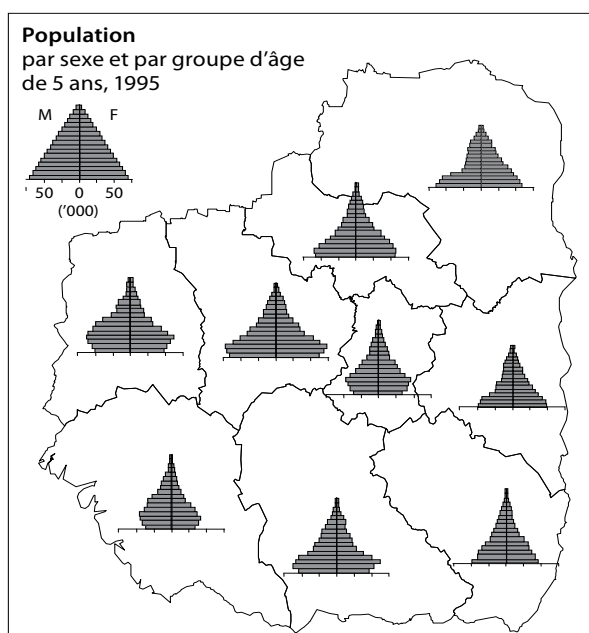
Carte montrant l'évolution dans le temps au moyen d'histogrammes



A5.44. Les pyramides de population présentées pour plusieurs régions sont significatives s'il existe des différences dans la forme des pyramides. Si les répartitions par âge et par sexe sont plutôt constantes dans tout le pays, les cartes résultantes ne fourniront guère d'indications intéressantes. Sur la figure A5.18, il semblerait que les provinces du sud-est aient connu une baisse de la fécondité au cours des 15 dernières années, tandis que cela n'a pas été le cas pour les provinces du nord. En outre, on

Figure A5.18

Combinaison de cartes et de pyramides de population

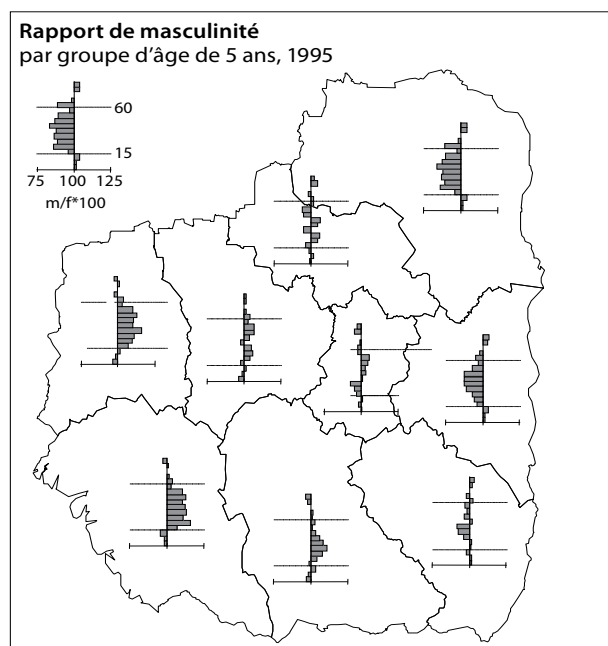


constate que les provinces du nord-est présentent une répartition dissymétrique du rapport de masculinité. Il semble qu'il y ait plus de femmes que d'hommes dans la population économiquement active. Au sud-ouest, la situation paraît être inverse.

A5.45. Les variations du rapport de masculinité peuvent être illustrées en utilisant un type différent de diagramme à barres, comme le montre la figure A5.19. Ces diagrammes montrent l'excédent ou le déficit d'hommes ou de femmes dans chaque province. La tendance qui apparaissait sur la carte des pyramides de population est ici beaucoup plus nette. Cependant, la carte est assez complexe et visuellement peu attrayante. On examinera un autre moyen de représenter les rapports de masculinité à la fin de la présente annexe.

Figure A5.19

Représentation des rapports de masculinité sur une carte



e) Cartes de flux

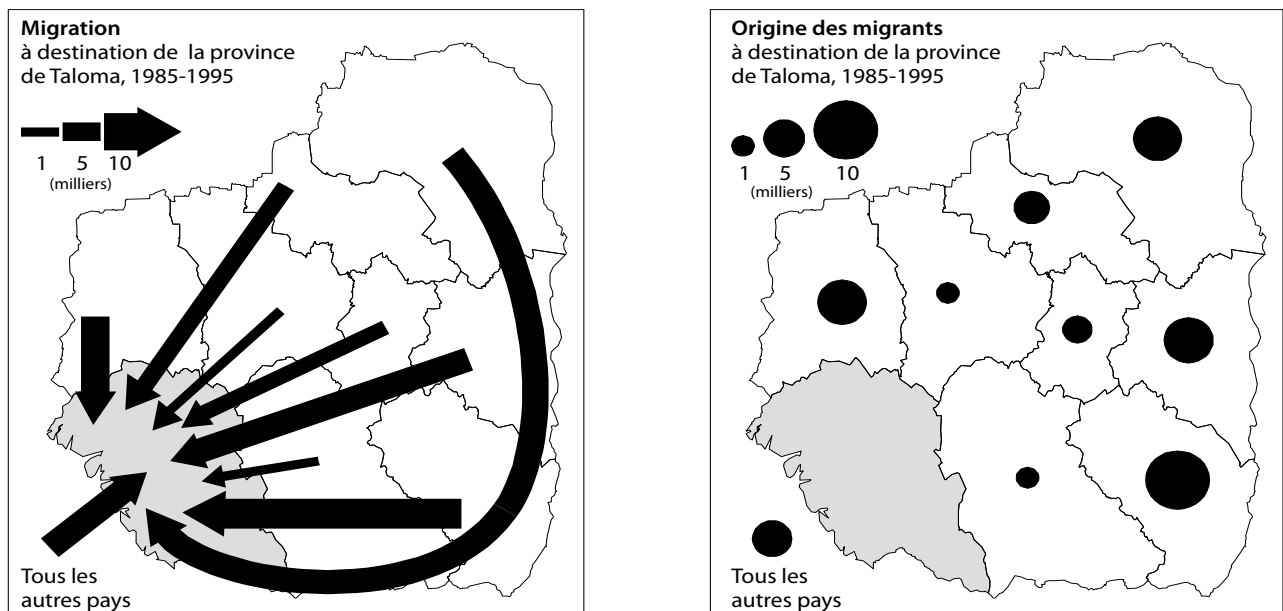
A5.46. La migration est une variable démographique qui représente les mouvements de population d'une partie du pays à une autre (migration interne) ou entre le pays et le reste du monde (migration internationale). La migration peut être représentée sur les cartes de diverses façons. Les taux de migration sont représentés au moyen de cartes choroplèthes des taux d'immigration, d'émigration ou de migration nette. Le volume de l'immigration ou de l'émigration peut être représenté en utilisant des cartes de symboles gradués (voir plus haut figure A5.13). Inversement, si l'on dispose d'informations complètes sur la migration, des cartes de flux — appelées également cartes de lignes flux — peuvent être utilisées. Ces cartes montrent divers aspects de la migration : l'itinéraire du flux de migration et la direction (de ... à), en utilisant une flèche comme symbole et l'importance du flux, en faisant varier l'épaisseur de la ligne.

A5.47. Les cartes de migration peuvent devenir rapidement très complexes. Même avec notre carte échantillon de provinces comportant seulement neuf unités déclarantes, il existe 72 flux éventuels — sans compter la migration internationale ou la migration à l'intérieur des provinces. C'est pourquoi on établit rarement des cartes complètes des flux montrant tous les itinéraires de migration possibles à l'intérieur

de la région du pays. Il existe plusieurs options possibles. L'une consiste à ne pas tenir compte des flux migratoires les moins importants et de représenter les plus importants et les plus significatifs. Une autre possibilité consiste à établir des cartes séparées pour chaque province montrant seulement l'immigration ou l'émigration de la province (voir figure A5.20). Pour nos provinces échantillons, cela donnerait une série de neuf paires de cartes. Même ces cartes plus simples peuvent facilement devenir surchargées. Le cartographe doit souvent dessiner des flèches courbes autour de la carte si les régions d'origine et de destination sont très éloignées.

Figure A5.20

Autres présentations des flux entre régions



A5.48. Dans les cartes de flux utilisant des flèches comme symboles, l'impression visuelle dépend de la longueur et de l'épaisseur de la flèche. Une flèche étroite et assez longue peut avoir un impact visuel plus grand qu'une flèche plus courte et plus large du fait de sa surface plus grande. Dans certains cas, le cartographe voudra peut-être utiliser ce fait pour attirer l'attention sur un flux migratoire intéressant en provenance d'une région éloignée, mais souvent le lecteur aura des difficultés pour évaluer l'importance relative des flux représentés par des flèches de longueur différente. Si l'on veut mettre l'accent sur le niveau absolu de migration en provenance de chaque région d'origine, d'autres présentations seront plus appropriées. Par exemple, au lieu de flèches, on pourra utiliser des symboles gradués pour montrer l'importance des flux migratoires par origine ou par destination (voir figure A5.20).

A5.49. En utilisant des types spéciaux de symboles gradués, chaque carte peut montrer à la fois la migration à destination de la province ou en provenant, comme le montre la figure A5.21. Ici, on utilise des demi-cercles de différentes couleurs ou nuances de gris pour distinguer l'immigration de l'émigration.

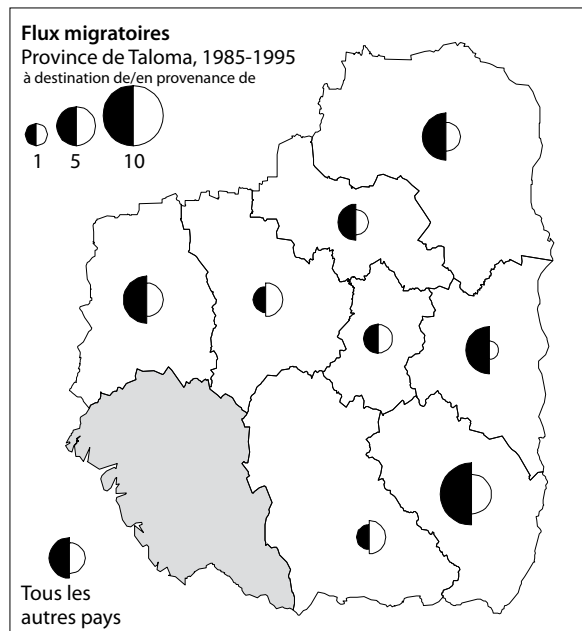
f) Représentation de phénomènes continus

A5.50. Les types de cartes présentés dans les précédentes sections conviennent pour des données concernant des objets géographiques distincts tels que des emplacements ponctuels ou des surfaces. Cependant, certains phénomènes géographiques

sont continus. La température ou l'altitude, par exemple, varient de façon continue dans l'espace. Mais il est aussi possible de considérer la répartition de la population comme une variable qui varie de façon plus ou moins continue. Les unités déclarantes sont définies assez arbitrairement et les valeurs agrégées présentées sous forme de tableau pour ces unités ne permettent pas de voir les variations spatiales à l'intérieur de chaque unité. Les atlas de population et, de plus en plus, les ensembles de données de SIG, présentent donc parfois la densité démographique et la répartition de la population comme des variables continues.

Figure A5.21

Représentation de l'immigration et de l'émigration

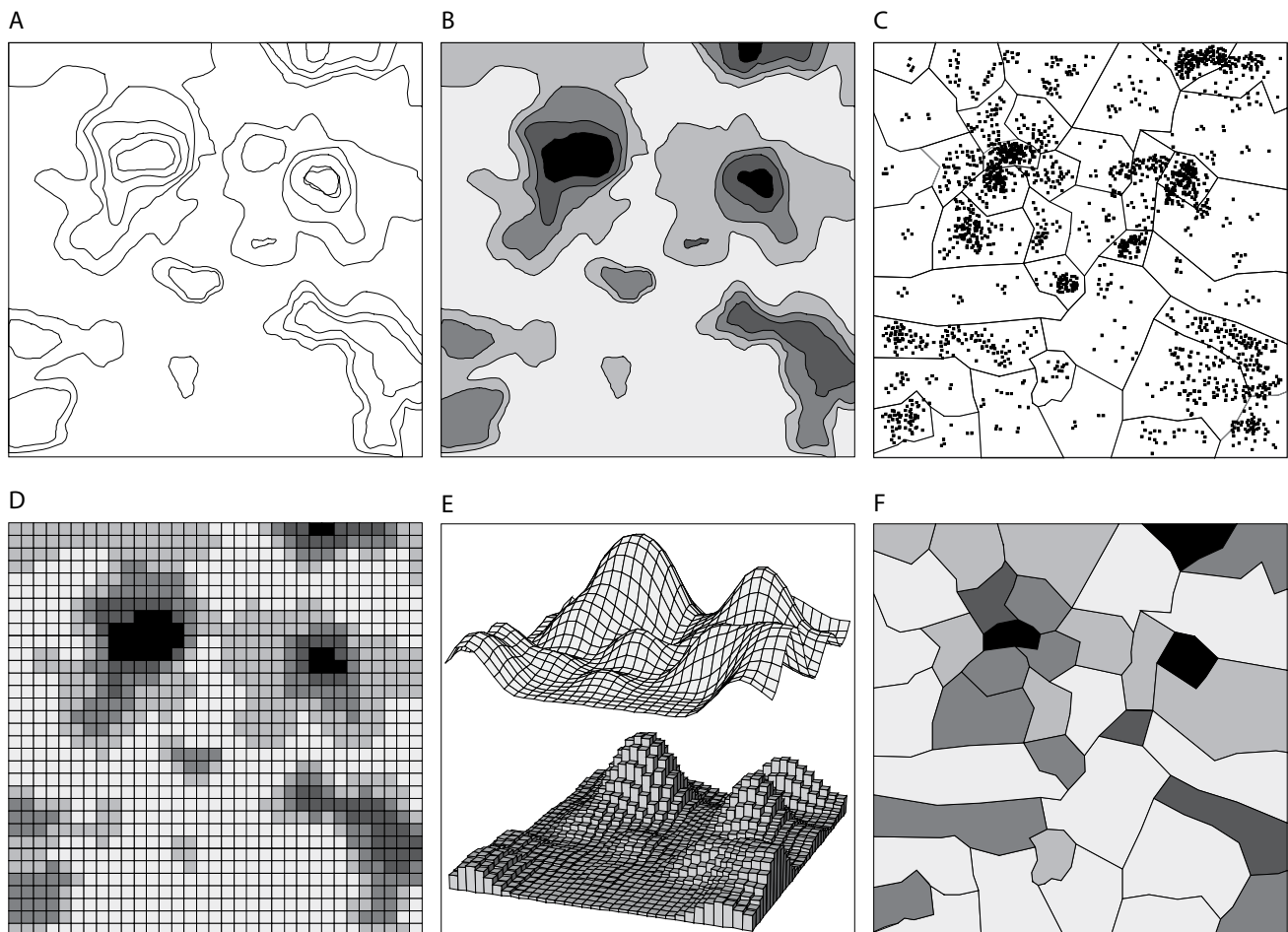


A5.51. La continuité effective ne peut pas être représentée facilement sur une carte imprimée ou dans une base de données d'ordinateur. Même si nous pouvions théoriquement établir une valeur différente pour chaque point exact du pays, il nous faudrait discrétiser les données d'une façon ou d'une autre aux fins de la cartographie. Plusieurs méthodes sont présentées sur la figure A5.22.

A5.52. Le moyen le plus courant pour représenter des données continues consiste à utiliser des isolignes ou des maillages réguliers. Les isolignes — le mot grec « iso » signifie égal — sont des lignes de valeur constante et on les appelle aussi lignes de niveau (figure A5.22A). Elles sont utilisées sur les cartes topographiques qui indiquent l'altitude. Les cartes de niveaux peuvent aussi être ombrées et elles ressemblent alors davantage à des cartes choroplèthes (voir figure A5.22B). Les couleurs représentent des valeurs dans la série de données comprise entre deux intervalles de niveau. Les cartes à points peuvent aussi être utilisées pour fournir une description plus continue de la répartition de la population ou d'une variable semblable. Comme on l'a indiqué plus haut, la plupart des logiciels de SIG produisent des cartes de points en traçant aléatoirement des points dans chaque unité déclarante. Dans le cas présent, nous n'obtenons aucune information supplémentaire par rapport à une carte choroplèthe. Mais si les points sont placés en fonction de données supplémentaires sur la couverture terrestre ou des emplacements de villages, par exemple, on pourra obtenir une image plus continue de la répartition de la variable (voir figure A5.22C).

A5.53. Aux fins de modélisation et d'analyse dans un SIG, les données continues sont en général enregistrées dans des maillages réguliers (figure A5.22D). La dimension des mailles est choisie de façon à préserver la variabilité dans l'ensemble de données, mais un maillage serré donne des dimensions de fichiers très importantes. Enfin, les logiciels de cartographie numérique, de même que les logiciels d'infographie en général, offrent différents moyens pour présenter sous forme de surfaces des ensembles de données variant de façon continue. La figure A5.22E montre deux exemples : un modèle fil de fer et un diagramme à barres à deux dimensions. Ces techniques sont très utiles pour présenter des informations sur les terrains fondées sur un modèle numérique d'altitude. Parfois, ces cartes peuvent aussi très bien illustrer la répartition de la population. Sur ces cartes, des collines et des pics représentent des groupements de densité démographique très élevées, tandis que les vallées symbolisent des zones faiblement peuplées. Cependant, en ce qui concerne la population et d'autres informations socioéconomiques semblables, il est souvent difficile d'évaluer la véritable répartition spatiale sur des surfaces. Intuitivement, nous ne pouvons interpréter des altitudes, mais il est beaucoup plus difficile d'associer rapidement les hauteurs de surface d'autres variables à leurs valeurs respectives. Il est donc en général plus indiqué de recourir à des techniques cartographiques plus classiques. Par comparaison, la figure A5.22F montre une carte choroplèthe où les unités déclarantes ne sont pas déterminées par la répartition des données.

Figure A5.22

Différentes méthodes cartographiques utilisables pour présenter des données continues

C. Classification des données

A5.54. On a examiné dans les sections précédentes les outils dont dispose le cartographe pour présenter des informations thématiques sur les cartes. Le concepteur de la carte doit choisir les variables graphiques et les types de cartes thématiques les plus appropriés pour la variable à représenter. Dans certains cas, les types de symbole et les valeurs de variables seront mis en correspondance, un à un. Ce sera le cas si l'on représente un petit nombre de catégories nominales, par exemple, avec des symboles ponctuels de taille semblable, mais de forme différente. Toutefois, même avec des données catégoriques, il est souvent nécessaire de représenter plusieurs objets ayant des valeurs similaires par le même symbole graphique. Par exemple, les ménages unis ou plurifamiliaux peuvent être représentés par le même symbole ponctuel. Les données numériques doivent presque toujours être catégorisées avant d'être appariés à des dimensions ou des couleurs de symboles.

A5.55. Le processus par lequel des observations présentant des valeurs semblables sont regroupées pour être représentées par le même symbole graphique est appelé classification. Il est semblable aux méthodes de classification dans le domaine statistique, qui groupent les valeurs en catégories de façon à réduire au minimum la variance des observations dans la même catégorie et à maximiser la variance entre différentes catégories. Les logiciels de cartographie numérique proposent des méthodes par défaut pour affecter des symboles à des valeurs ou à des intervalles de valeurs. Même si les interfaces graphiques utilisateur se sont améliorées, ces valeurs par défaut peuvent ou non être appropriées pour la variable représentée — le plus souvent elles ne le sont pas. Des outils de classification inappropriés donnent souvent des tracés de cartes inappropriés ou même susceptibles d'induire en erreur. On examinera donc plus en détail les possibilités de classification aux paragraphes suivants.

A5.56. Les classes de données numériques sont en général des intervalles de valeurs contigus. Le nombre de classes est déterminé par plusieurs facteurs : la répartition des données (c'est-à-dire la variation des valeurs dans l'ensemble de données), l'exactitude recherchée de la représentation des données et — facteur non moins important — la capacité du périphérique de sortie de présenter de petites différences de couleurs et de texture. On n'améliore pas nécessairement une carte thématique en accroissant le nombre de classes, car il est de plus en plus difficile pour le lecteur de distinguer différentes classes. Il importe plutôt de déterminer les intervalles de classes de façon à refléter exactement les variations de l'ensemble de données.

A5.57. Le choix d'une technique de classification appropriée dépend de la répartition des données de la variable. Une méthode permettant d'obtenir une carte exacte et visuellement attrayante pour un ensemble de données uniformément réparti (c'est-à-dire qui comporte un nombre à peu près égal de valeurs élevées, moyennes et faibles) ne donnera pas nécessairement de bons résultats pour une répartition de données dissymétrique — c'est-à-dire une répartition comportant beaucoup de valeurs faibles et peu de valeurs élevées.

A5.58. Pour établir des cartes de qualité appropriée pour la publication, les données devront donc toujours être évaluées en utilisant des diagrammes statistiques. Malheureusement, les logiciels de SIG et de cartographie numérique n'ont que des capacités limitées en matière de tracé de diagrammes. Mais ils permettent d'exporter les données vers des tableurs ou des programmes statistiques qui possèdent des fonctions développées pour tracer des diagrammes.

A5.59. Le type le plus utile de diagramme pour déterminer les intervalles de classe est un classement par ordre de grandeur. Tous les points correspondant aux

données sont classés en fonction de leur valeur de bas en haut. Ils sont ensuite tracés l'un à côté de l'autre — l'axe des x indique le rang de chaque observation et l'axe des y les valeurs de données. Les intervalles verticaux ou *discontinuités naturelles* entre des points de données voisins sont de bons candidats pour tracer des limites de classes, mais souvent on aura plus ou moins d'intervalles que le nombre de classes souhaité.

A5.60. Les pages qui suivent présentent des exemples de méthodes de classification courantes pour trois variables comportant des répartitions différentes de données statistiques. La variable densité démographique présente une répartition dissymétrique. Il existe de nombreuses valeurs dans l'intervalle de 21 à environ 110 personnes au kilomètre carré, et seulement quelques valeurs très élevées. La valeur la plus élevée (791) est égale à environ deux fois et demi la deuxième valeur par ordre d'importance (320). Cela n'est pas rare en ce qui concerne la densité démographique. Le district présentant les valeurs les plus élevées, par exemple, peut inclure la capitale d'une province par ailleurs rurale. La seconde variable est le taux d'alphabétisation pour les districts. Les valeurs sont assez uniformément réparties, ce qu'indique la ligne presque droite correspondant aux observations dans le classement par ordre de grandeur. Il n'y a pas de valeurs extrêmes.

A5.61. La troisième variable citée est l'indice synthétique de fécondité. Le classement par ordre de grandeur montre un accroissement rapide des valeurs pour les observations inférieures, une section moyenne étendue avec une augmentation beaucoup moins rapide et de nouveau un accroissement plus rapide des valeurs pour les observations très élevées vers la droite. Cela correspond à une répartition dite normale, qui est caractérisée par un plus faible nombre de valeurs extrêmement faibles et élevées et de nombreuses observations dans les intervalles intermédiaires. Naturellement, les exemples cités ne sont là qu'à titre d'illustration. Les mêmes variables pour d'autres zones géographiques pourront présenter des répartitions très différentes.

A5.62. Les exemples montrent que l'apparence d'une carte dépend au plus haut point du choix de la méthode de classification, qui peut ou non être appropriée pour la répartition des données. Cela confirme que les méthodes de classification fournies dans les logiciels de SIG doivent être utilisées avec une certaine prudence.

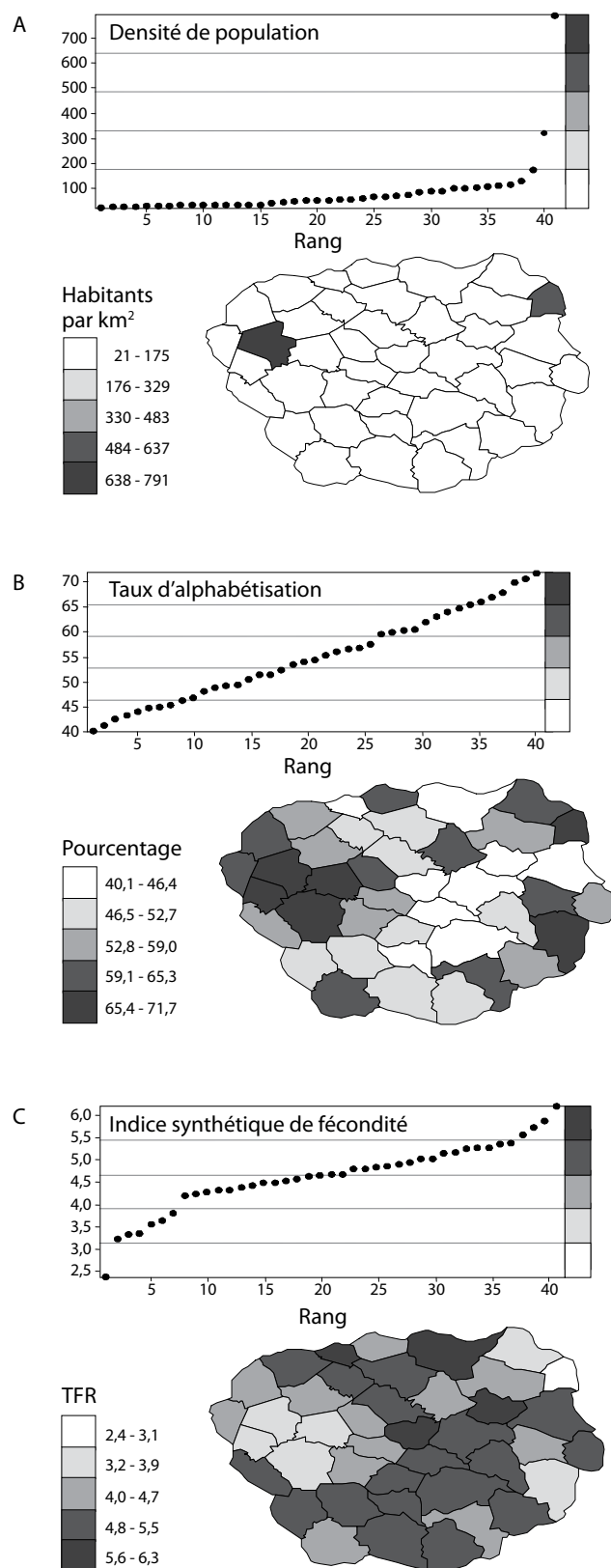
1. Classification de données en série

A5.63. L'une des méthodes de classification les plus simples consiste à diviser la gamme des valeurs de données en *intervalles égaux* (figure A5.23). Le cartographe détermine d'abord combien il utilisera de classes. L'intervalle de valeurs de données — la valeur la plus élevée moins la plus faible — est ensuite divisé par le nombre de classes pour obtenir l'incrément, appelée également raison de la progression. La première classe s'étend alors de la valeur la plus faible à cette même valeur augmentée de l'accroissement et les classes suivantes sont définies en ajoutant l'accroissement à la valeur supérieure de la classe précédente. Il peut être nécessaire d'arrondir si les nombres sont indiqués dans la légende avec un faible degré de précision.

A5.64. Pour la variable densité démographique, la valeur inférieure est égale à 21 et la plus élevée à 791. L'intervalle est donc de 770. Puisque nous voulons utiliser cinq catégories, la raison de la progression est donc égale à $770/5$ soit 154. Ainsi la première catégorie va de 21 à 175, la suivante de 176 à 329 et ainsi de suite.

A5.65. La carte de densité démographique montre pourquoi cela peut occasionner des difficultés. L'intervalle de valeurs est influencé par une valeur très élevée. La raison de la progression est donc si élevée que l'intervalle de la première catégorie comprend toutes les observations sauf deux. Il est évident que la carte obtenue n'est pas très instructive.

Figure A5.23

Intervalles égaux

A5.66. Cette méthode donne de biens meilleurs résultats avec le taux d'alphabétisation, qui est réparti de façon plus uniforme. L'ensemble de données est divisé en nombres à peu près égaux d'observations dans chaque catégorie et la carte résultante donne une bonne impression de la répartition de l'alphabétisation entre les districts.

A5.67. Enfin, la carte de l'indice synthétique de fécondité donne lieu à des difficultés comparables, mais moins marquées que celles rencontrées pour la carte de la densité démographique. On ne compte qu'une seule observation dans la catégorie inférieure et la carte est plutôt dominée par les valeurs des intervalles de la catégorie moyenne. Cependant, par une coïncidence, les intervalles de classes entre la deuxième et la troisième et entre la quatrième et la cinquième catégories correspondent assez bien aux discontinuités dans la répartition des données.

A5.68. Outre ce procédé des intervalles égaux, il existe d'autres solutions pour la classification des données en série. L'une d'entre elles consiste à utiliser une progression géométrique régulière telle que 0-2, 2-4, 4-8, 8-16, etc. Cela peut donner de bons résultats pour des répartitions dissymétriques de données comme dans le cas de la variable densité démographique.

2. Classification statistique

A5.69. Une méthode de classification consiste à avoir un nombre approximativement égal d'observations géographiques dans chaque catégorie. Ce résultat peut être obtenu en utilisant le concept statistique des quantiles, qui divisent l'ensemble de données en classes comportant le même nombre d'observations. Si l'on a quatre classes, on parle de quartiles, s'il y en a cinq, de quintiles, etc.

A5.70. Pour déterminer les quantiles, le nombre d'observations est divisé par le nombre de catégories souhaité et, si nécessaire, arrondi à l'entier le plus proche. Dans le classement par ordre de grandeur, les premières observations n sont affectées à la première catégorie, les n suivantes à la seconde, etc. Tout nombre impair est affecté à la première ou à la dernière catégorie.

A5.71. L'établissement de cartes par quantiles est assuré par de nombreux logiciels de cartographie numérique et c'est pourquoi cette méthode est devenue très courante pour l'établissement des cartes.

A5.72. Les trois échantillons de cartes ont tous un aspect attrayant. On a par définition une bonne répartition des observations entre les classes de telle sorte que toutes les cartes utilisent pleinement toute la gamme des gris.

A5.73. Si l'on considère la répartition des données, la classification concernant la variable taux d'alphabétisation paraît très appropriée. En fait, la carte ne paraît pas très différente de celle qui utilise des intervalles égaux.

A5.74. Cependant, pour les cartes de la densité démographique et de l'indice synthétique de fécondité, nous constatons que la méthode regroupe des valeurs semblables dans des catégories différentes. Pour l'indice synthétique de fécondité, par exemple, les deux observations présentant les valeurs les plus élevées dans l'intervalle de données inférieur (2,4-4,2) ressemblent beaucoup plus aux observations de la seconde catégorie qu'aux observations de la première catégorie. Bien plus, on a trois observations avec une valeur de 5,3, une qui est affectée à la quatrième classe et deux à la cinquième (certains logiciels de cartographie numérique assouplissent le critère du nombre égal d'observations pour éviter de tels cas).

A5.75. Les cartes par quantiles doivent donc être utilisées avec prudence. Très souvent, des valeurs semblables se retrouvent affectées à des catégories différentes et

Figure A5.24

Établissement de cartes par quantiles (fréquence égale)

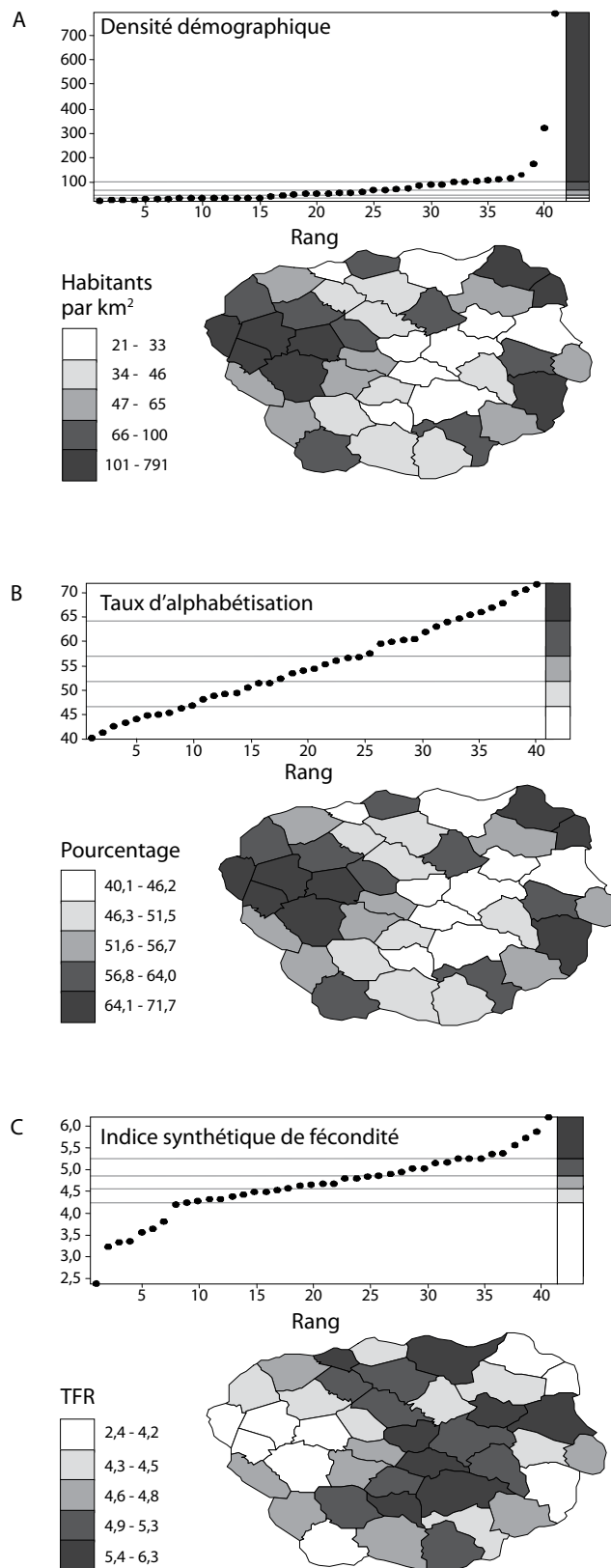
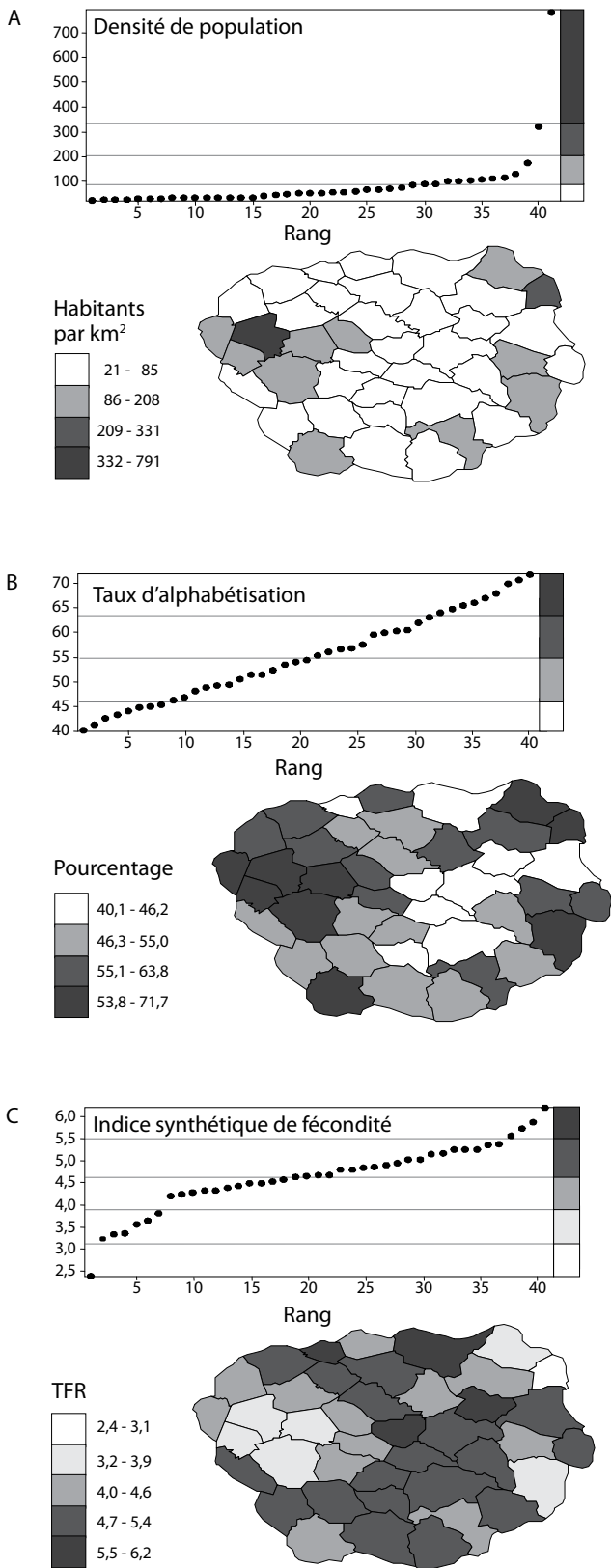


Figure A5.25
Écart-type



des valeurs différentes sont affectées à la même classe. Bien que les cartes obtenues soient visuellement attrayantes, cette impression risque d'induire en erreur.

A5.76. Il existe une autre technique de classification statistique fondée sur des évaluations sommaires de la répartition des données. Une méthode consiste à déterminer des intervalles de classes en utilisant l'écart-type de la répartition de la variable. L'écart-type est égal à la racine carrée de la variance. La variance est calculée en tant que moyenne des carrés des écarts entre les valeurs de données et la valeur moyenne globale. Par exemple, pour le taux d'alphabétisation, l'écart-type est égal à 8,9.

A5.77. Les catégories de cartes fondées sur les écarts-types montrent ainsi comment des observations individuelles — concernant par exemple des districts — se comparent à la valeur moyenne d'une province ou du pays tout entier.

A5.78. Les catégories sont obtenues en soustrayant l'écart-type de la moyenne (55 pour le taux d'alphabétisation) ou en le lui ajoutant. Les intervalles des catégories sont donc constants, comme avec la méthode des intervalles égaux.

A5.79. En ce qui concerne le taux d'alphabétisation, le premier intervalle de données (40,1-46,2) correspond à des valeurs supérieures à un écart-type, mais inférieures ou égales à deux écarts-types au-dessous de la moyenne. La répartition des données étant assez dense, toutes les valeurs sont comprises entre +/- deux écarts-types et quatre catégories seulement sont nécessaires. Comme nous le voyons sur la figure A5.25B, cette méthode divise les valeurs du taux d'alphabétisation en nombres d'observations à peu près égaux dans chaque catégorie, ce qui donne une carte avec un bon contraste visuel.

A5.80. Cependant, cette approche convient beaucoup moins bien pour la variable densité démographique. Parce qu'il y a beaucoup de valeurs peu élevées, la densité démographique moyenne est plutôt faible (85,4) et l'écart-type plutôt élevé (124,8). La première catégorie — correspondant à des valeurs situées à moins d'un écart-type de la moyenne — devrait donc être en fait comprise entre -39,5 et 85,4. D'autre part, la valeur la plus élevée (791) est à plus de cinq écarts-types de la moyenne. Nous devrions donc utiliser beaucoup plus de catégories, dont la plupart ne contiendraient aucune observation. Au contraire, la catégorie la plus importante sur la carte présentée ici inclut toutes les valeurs à plus d'un écart-type de la moyenne. Il est évident que les écarts-types ne constituent pas un bon choix pour cette variable.

A5.81. Les écarts-types donnent des résultats un peu meilleurs pour l'indice synthétique de fécondité, avec une moyenne de 4,6 et un écart-type de 0,8. Cependant, seule la valeur très faible de 2,4 se situe dans la catégorie la plus basse, soit à plus de deux écarts-types au-dessous de la moyenne.

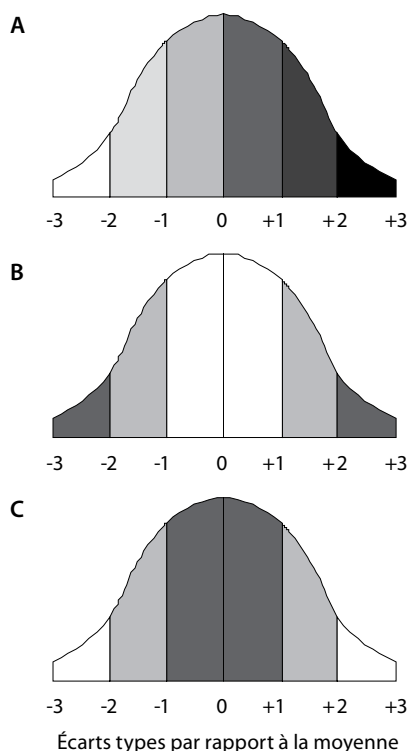
A5.82. La méthode de classification par les écarts-types est intéressante parce qu'intuitive, étant étroitement associée aux techniques statistiques descriptives. Elle donne de bons résultats si les données sont normalement réparties, avec une variance relativement faible, de telle sorte qu'au plus six catégories suffisent pour inclure toutes les valeurs.

A5.83. Les écarts-types peuvent être utilisés pour représenter différents types de tendances dans un ensemble de données [voir figure A5.26; voir également Dent (1999)]. Dans les exemples fournis sur la figure A5.25, on utilise une échelle de gris allant du clair au sombre. Les cartes illustrent la progression des valeurs faibles aux valeurs élevées de densité démographique, de taux d'alphabétisation et d'indice synthétique de fécondité correspondant au classement par catégories présenté sur la figure A5.26A. C'est en fait l'application la moins commune de la classification fondée sur les écarts-types.

A5.84. Cette méthode est plus couramment utilisée pour illustrer des tendances divergentes. Par exemple, pour montrer des niveaux de revenu, nous souhaiterions peut-être faire ressortir quels sont les districts les plus pauvres et le plus riches. Dans ce cas, nous affecterons des couleurs vives ou une texture marquée aux districts dont les valeurs se situent à plus d'un et deux écarts-types de la moyenne et des couleurs relativement pâles aux districts situés au centre de la répartition de données (figure A5.26B).

Figure A5.26

Affectation de couleurs aux catégories déterminées par les écarts-types



A5.85. Si l'on s'intéresse seulement à la distance par rapport à la moyenne — sans se préoccuper de savoir si les valeurs sont supérieures ou inférieures à la moyenne — on pourra utiliser les mêmes couleurs de part et d'autre. S'il s'agit aussi de savoir si les valeurs sont supérieures ou inférieures à la moyenne, on devra utiliser différentes couleurs ou textures pour chaque côté. Par exemple, sur une carte imprimée en couleurs, on pourra affecter des teintes rouges allant du clair au sombre aux catégories inférieure à la moyenne et des teintes bleues correspondantes aux catégories supérieures à la moyenne.

A5.86. Dans d'autres cas, nous souhaiterions peut-être faire ressortir les intervalles moyens (figure A5.26C). MacEachren (1994), par exemple, examine une carte de l'Irlande du Nord (Royaume-Uni) publiée par Fothergill et Vincent (1985) qui montre la proportion de Protestants et de Catholiques. Sur cette carte, les valeurs de l'ordre de 50 %, indiquant un nombre à peu près égal de Protestants et de Catholiques, sont mises en lumière en affectant aux catégories moyennes une couleur vive (jaune). Les zones où les Catholiques ou les Protestants sont en nette majorité sont présentées avec des couleurs plus pâles (vert et orange, respectivement).

3. Points d'inflexion naturels

A5.87. Comme nous l'avons vu avec les précédents exemples, la plupart des méthodes donnent des cartes de nature à induire en erreur pour des variables qui ne présentent pas une répartition très uniforme. Il arrive souvent que des valeurs semblables soient affectées à des catégories différentes ou que des valeurs très différentes soient regroupées ensemble. Une méthode logique pour classer les données en cartographie consiste à trouver un mode de regroupement des données qui optimise l'opération en minimisant les différences entre valeurs dans chaque catégorie et en maximisant la variation entre groupes.

A5.88. Cet objectif peut être atteint en procédant à une inspection visuelle de la répartition des données et ensuite au choix des inflexions entre catégories. Des exemples de cette méthode sont présentés sur la figure A5.27. Pour la variable indice synthétique de fécondité, cela est très simple, car elle présente plusieurs inflexions distinctes dans la répartition.

A5.89. Cela s'avère plus difficile pour les deux autres variables. Pour la densité démographique, une stricte application de la méthode pourrait consister à affecter toutes les valeurs inférieures à la même catégorie et les valeurs supérieures à un certain nombre de catégories séparées. Mais cela doit être mis en balance avec le souci de préserver une variation peu marquée dans les intervalles de valeurs inférieures.

A5.90. De même, pour la variable taux d'alphabétisation uniformément répartie, les points d'inflexion ne sont pas très nets car la différence de valeur entre les observations ne varie pas beaucoup.

A5.91. Malgré ces difficultés, la classification à partir des inflexions naturelles, prenant explicitement en considération la répartition des données, permet en général d'obtenir des représentations cartographiques précises des données avec un bon contraste visuel.

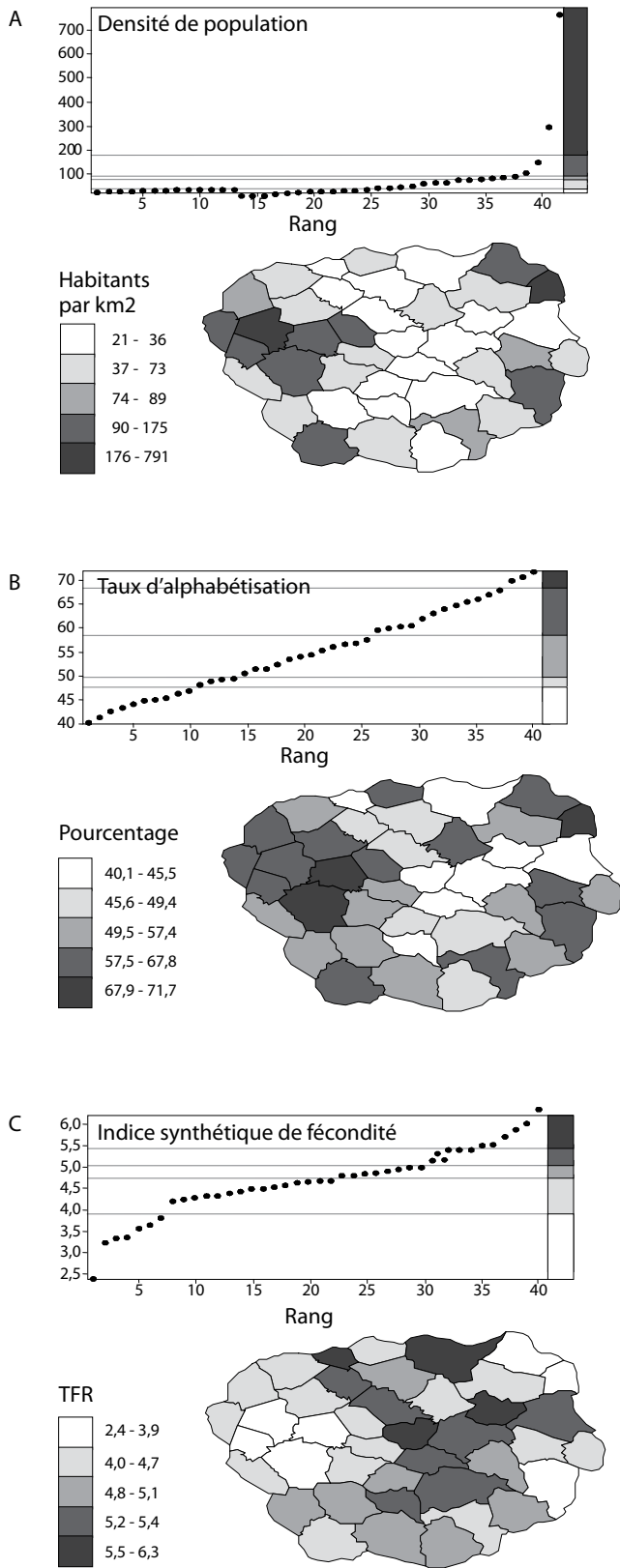
A5.92. Plutôt que de se fier à une appréciation quelque peu subjective, on peut aussi laisser l'ordinateur déterminer les inflexions naturelles ou optimales. De nombreux logiciels de SIG et de cartographie numérique possèdent des fonctions permettant de définir des points d'inflexion naturels fondés sur une évaluation automatisée de la répartition des données (méthode de classification optimale de Jenk). On peut aussi utiliser des fonctions de classification ou de regroupement dans des logiciels statistiques.

4. Cartes choroplèthes sans intervalles de catégories

A5.93. Avec les cartes choroplèthes dites sans catégories, le cartographe n'a pas à choisir de méthode de classification. Grâce à une technologie améliorée d'affichage et d'impression, les écrans d'ordinateurs et les imprimantes peuvent obtenir une large gamme de couleurs différentes ou de nuances de gris. Pour une carte sans catégorie ou de catégorie n , les valeurs de données détermineront directement, par exemple, le pourcentage de gris. Par exemple, pour une variable exprimée en pourcentage, nous pouvons choisir le niveau de gris sur une échelle allant de 0 % de gris (blanc) à 100 % de gris (noir) qui correspond à chaque valeur de l'observation. Mais si la méthode de reproduction peut donner un nombre suffisant de nuances distinctes, il est recommandé d'éviter le blanc comme couleur, car c'est aussi généralement la couleur du fond de page.

A5.94. En pratique, cela ne donnera peut-être pas des résultats optimaux. Cela tient notamment au fait que de nombreuses variables ne vont pas de 0 à 100, mais ont des valeurs concentrées sur un intervalle plus restreint. La carte peut ainsi ne présenter en définitive que des nuances de gris très claires ou très sombres. Nous pouvons éviter

Figure A5.27
Points d'inflexion naturels



ce problème en « étirant » la répartition des données : l'utilisation de la couleur la plus claire pour la valeur la plus basse et de la plus sombre pour la valeur la plus élevée donnera des cartes plus faciles à interpréter par le lecteur.

A5.95. Mais en général, il y a une limite au nombre de nuances de gris et de couleurs qui peuvent être facilement distinguées. Si un système d'ombrage continu est utile à des fins analytiques, le classement des valeurs de données entre un nombre limité de catégories est généralement préférable pour des cartes établies à des fins de présentation.

5. Classification des données extérieures

A5.96. Dans certains cas, le système de classification est fourni par un organisme extérieur. Par exemple, pour établir une carte de la pauvreté par district, le cartographe utilise une valeur seuil de donnée du revenu moyen — appelée seuil de pauvreté — au-dessous de laquelle un district est considéré comme pauvre. Autre exemple de système de classification fourni extérieurement, celui dans lequel on procède à des comparaisons avec les cartes imprimées existantes pour lesquelles les données initiales ne sont pas disponibles. Pour permettre une comparaison exacte entre les cartes, par exemple pour le taux de fécondité dans les différentes provinces d'un pays, la classification doit être identique.

6. Remarques générales

A5.97. L'aperçu qui précède montre qu'il existe de nombreuses méthodes pour affecter des valeurs de données à des catégories. La plupart des logiciels de SIG et de cartographie numérique prennent en charge les intervalles égaux, les quantiles, les écarts-types et les points d'inflexion naturels. De plus, tous les logiciels permettent à l'utilisateur de définir une classification de données personnalisée.

A5.98. Chaque méthode a ses avantages et ses inconvénients, qui sont présentés sur le tableau A5.1. Le choix de la méthode la plus appropriée dépend de la répartition des données et des objectifs de la carte. En général, la répartition des données devrait toujours être évaluée en utilisant un diagramme statistique tel que le tableau d'ordre de grandeurs présenté plus haut. De tels outils sont proposés dans certains logiciels de SIG. Le nombre optimal de catégories et les meilleurs points d'inflexion paraîtront alors souvent tout à fait évidents.

Tableau A5.1

Évaluation des différentes techniques de classification

Méthode de classification	Avantages	Inconvénients
Intervalles égaux	Facile à mettre en œuvre. Méthode appropriée pour des données uniformément réparties.	Aucun rapport entre le système de classification et la répartition des données. Les intervalles de catégories étant fixes, des valeurs semblables peuvent être affectées à différentes catégories et des valeurs différentes à une même catégorie. Ne convient pas pour des répartitions dissymétriques de données ou des ensembles de données avec des valeurs hors norme.
Progression géométrique	Facile à mettre en œuvre. Appropriée pour des données avec une répartition très dissymétrique (beaucoup de valeurs faibles et peu de valeurs très élevées).	La progression géométrique appropriée doit être déterminée par l'utilisateur. Les intervalles de catégories étant fixes, des valeurs semblables peuvent être affectées à différentes catégories et des valeurs différentes à une même catégorie.

Méthode de classification	Avantages	Inconvénients
Quantiles (fréquence égale)	Un bon contraste visuel est garanti. Appropriée pour des données très uniformément réparties.	Des valeurs semblables ou identiques peuvent se retrouver dans des catégories différentes.
Écarts-types	Convient pour représenter des tendances divergentes centrées autour de la valeur moyenne. Établit un rapport entre des catégories individuelles et la valeur moyenne globale. Convient pour des données normalement réparties.	Des répartitions dissymétriques ou des ensembles de données contenant des valeurs hors norme (peu de valeurs très élevées ou très faibles) nécessiteront un grand nombre de catégories (plusieurs écarts-types au-dessus ou au-dessous de la moyenne)
Inflexions naturelles	Des valeurs semblables seront affectées à la même catégorie. Le nombre de catégories est souvent suggéré par le nombre d'inflexions.	Les intervalles de catégories obtenus peuvent être très inégaux. Exige une appréciation subjective (détermination visuelle). Ne permet pas la comparaison des cartes dans le temps.
Cartes choroplèthes sans classification	Il n'est pas nécessaire de définir des interruptions entre catégories. La nuance de gris ou la couleur est déterminée directement par la valeur de la donnée. Fait ressortir une répartition continue des valeurs dans l'ensemble de données.	La plupart des périphériques de sortie n'acceptent qu'un nombre limité de nuances de gris ou de couleurs distinctes. Les cartes présentant de faibles différences de gris ou de couleur sont difficiles à reproduire (par exemple à photocopier). Difficile à mettre en œuvre dans la plupart des logiciels de SIG et de cartographie.

A5.99. Il convient de noter que les points d'inflexion naturels ne sont pas appropriés si plusieurs cartes sont présentées ensemble à des fins de comparaison, par exemple, une série chronologique des rapports de masculinité par district ou des cartes d'accès à l'eau potable pour deux provinces du pays. Dans ces cas, les limites de classes doivent être choisies constantes. Un système de classification défini par l'utilisateur et fondé sur une évaluation de toutes les séries chronologiques doit être choisi à cette fin. Les cartes par quantiles peuvent aussi être utilisées, si l'objectif consiste seulement à comparer le classement des différentes observations dans le temps ou l'espace plutôt que les valeurs de données effectives. Deux cartes par quartiles pourraient par exemple présenter les 25 % des districts possédant les taux d'alphabétisation les plus élevés tels qu'ils apparaissent à la suite du dernier recensement et du recensement précédent.

D. Choix de la couleur

A5.100. Les exemples de cartes examinées dans la présente annexe utilisent tous des échelles de gris pour la symbolisation. Les publications en noir et blanc sont moins coûteuses à produire et les cartes utilisant des échelles de gris demeurent lisibles lorsqu'on les reproduit sur une photocopieuse en noir et blanc. Par contre, l'utilisation de la couleur offre au cartographe beaucoup plus de possibilités dans la conception de la carte. Le prix des imprimantes couleurs continue de baisser. D'autre part, dans un proche avenir, beaucoup plus de cartes seront présentées électroniquement sur des sites Web ou des publications électroniques. Dans ces cas, la couleur pourra être largement utilisée pour établir les cartes, même s'il conviendra de l'utiliser judicieusement.

A5.101. Il est utile de savoir comment un ordinateur interprète les couleurs lorsqu'on doit définir un système de couleurs pour une carte choroplèthe. Les couleurs sont définies dans un ordinateur en utilisant un modèle parmi plusieurs modèles de

couleurs. Deux des plus courants sont le modèle teinte-saturation-intensité (TSI) et le modèle rouge, vert et bleu (RVB). Le terme « teinte » désigne ce que l'on comprend généralement par la couleur, comme le « rouge » ou le « bleu ». Physiquement, la teinte se réfère à l'intervalle spectral de la lumière réfléchi et va du violet, avec une faible longueur d'onde, au bleu, vert, jaune, orange et rouge qui a la longueur d'onde la plus élevée dans le spectre visible. L'intensité est appelée parfois luminosité [teinte-saturation-luminosité (TSL)]. Elle détermine la différence entre, par exemple, un rose pâle et un rouge sombre, qui ont tous deux la même teinte. Enfin, la « saturation » est la mesure de la brillance ou de l'intensité. Une couleur ayant une saturation faible apparaît plus pâle ou grise tandis qu'une couleur à haute saturation paraît plus pure.

A5.102. Le modèle RVB est un modèle dans lequel les nouvelles couleurs sont définies par addition en combinant différents niveaux de rouge, de vert et de bleu. Les écrans d'ordinateur ou de téléviseur utilisent la méthode RVB. Des niveaux égaux des mêmes trois couleurs donnent des nuances de gris. Les niveaux inférieurs de rouge, vert et bleu combinés donnent du noir et les valeurs les plus élevées du blanc.

A5.103. Le choix de la couleur dépend du niveau de mesure de la variable, du type de carte utilisé et du message que le cartographe souhaite communiquer. L'œil humain distingue facilement différentes teintes, ce qui est tout à fait approprié pour distinguer des catégories distinctes. Par exemple, des cercles bleus peuvent être opposés à des cercles rouges pour montrer différents types d'écoles. Cependant, il faut tenir compte du daltonisme lorsqu'on choisit une couleur uniquement pour distinguer différents symboles cartographiques. Les daltoniens risquent de ne pas pouvoir distinguer le rouge et le vert — forme la plus courante du daltonisme touchant 1 % des hommes — ou le bleu et le jaune. Certains ne peuvent pas voir la partie verte du spectre des couleurs. En général, il est préférable de ne pas s'en remettre aux différences entre le rouge et le vert dans la composition des cartes.

A5.104. Pour présenter des variables mesurées en continu comme la population, le revenu ou des coefficients et des pourcentages, on utilise des variables graphiques qui comportent un classement distinct. Les différences de niveau de couleur (c'est-à-dire les nuances claires à sombres d'une même couleur) sont facilement associées aux valeurs de la variable, tandis que les nuances sombres sont en général associées aux valeurs les plus élevées de la donnée. Par exemple, les niveaux de densité démographique sont souvent représentés par des teintes rouges, allant des rouges très pâles pour une faible densité démographique aux rouges sombres pour les zones de forte densité démographique. Pour les répartitions de données dissymétriques, les niveaux de couleur ne sont naturellement pas liés en proportion directe aux valeurs des catégories de données. Sinon, dans l'exemple de la variable densité démographique utilisé dans la précédente section, par exemple, nous utiliserions de nombreuses nuances très pâles, difficiles à distinguer, de rouge clair pour les nombreuses valeurs peu élevées et une nuance très sombre pour un petit nombre de valeurs élevées. On utilise au contraire des degrés égaux de niveau de couleur pour représenter les catégories d'une progression géométrique ou similaire.

A5.105. Si une classification comporte de nombreuses catégories, nous pouvons nous retrouver avec un plus grand nombre de catégories que ce que l'on peut distinguer clairement sur une page imprimée. Dans ce cas, les nuances de couleurs proches les unes des autres peuvent être combinées — ce que l'on appelle une gamme de couleurs d'un spectre partiel. Pour utiliser à nouveau l'exemple de la densité démographique, nous pouvons commencer par des nuances de jaune clair et continuer jusqu'aux rouges sombres en passant par les oranges. Il importe de considérer qu'il doit y avoir une nette progression des couleurs moins dominantes aux couleurs plus

dominantes. Les cartes utilisant plusieurs teintes brillantes dominantes pour des valeurs faibles ou levées d'un intervalle continu ou ordinal de catégories ne communiquent pas un message clair et sont une source de confusion pour le lecteur.

A5.106. Il paraît approprié d'utiliser des couleurs différentes pour représenter un intervalle continu de données dans une application telle qu'une échelle de données divergente. Par exemple, une carte de l'immigration nette par unité administrative comportera un éventail de catégories partant de nombres négatifs élevés pour une forte émigration et atteignant des valeurs positives élevées correspondant à une forte immigration en passant par zéro. Pour faire ressortir les valeurs négatives et positives élevées — où la migration influe le plus sur la dynamique de la population — nous pouvons utiliser un système de couleurs allant par exemple du rouge intense au rouge pâle ou au rose, puis au blanc pour les taux de migration nette proches de zéro et du bleu pâle au bleu intense pour les taux d'immigration les plus élevés.

A5.107. Une dernière observation concernera la catégorie multivariable, avec laquelle deux variables sont présentées en combinaison. Par exemple, une carte pourra associer différents niveaux d'alphabétisation et de fécondité en utilisant une légende constituée essentiellement d'une matrice de combinaisons possibles des catégories alphabétisation et fécondité. Le cartographe doit trouver un système de couleurs approprié dans lequel, par exemple, les différences de niveaux d'alphabétisation sont représentées par des teintes voisines dans un système spectral partiel et les différences de taux de fécondité sont représentées par des variations de l'intensité de la couleur. Malheureusement, ces cartes sont difficiles à interpréter. Le lecteur doit constamment consulter la légende pour assortir les couleurs aux valeurs de données concernant les deux variables. En général, on devra éviter de recourir à des cartes multivariables. La section A5.6 présentera quelques approches alternatives à la présentation géographique d'informations multivariables.

A5.108. Revenant aux types de niveaux de mesure examinés plus haut, le tableau A5.2 présente, sous forme résumée, les directives concernant l'utilisation des nuances de gris et des couleurs [voir également Brewer (1994)].

Tableau A5.2
Choix des nuances de gris et des couleurs

Niveaux de mesure		Exemple	Cartes en noir et blanc	Cartes en couleurs
Nominaux	Binaires	Accès à l'eau potable	Blanc ou noir, gris pâle ou gris sombre.	Nuances fortement contrastées de différentes couleurs comme le bleu et le rouge ou le jaune et le vert.
	Catégoriels	Langue dominante (anglais, français, espagnol, etc.)	Variations de structure avec une dominance visuelle semblable.	Différentes couleurs avec des niveaux semblables de valeur et de saturation qui ne supposent aucun classement, par exemple bleu, vert, jaune, violet.
Ordinaux		Niveau d'éducation (école primaire, école secondaire, etc.)	Nuances de gris échelonnées avec des différences relativement importantes entre niveaux de gris. Les différences de texture font ressortir encore mieux le caractère ordinal des données.	Même nuance ou intervalle spectral partiel de couleurs, avec des différences relativement élevées entre catégories. Par exemple, jaune pâle, orange, rouge moyen, rouge sombre.
Discrets		Dimension des ménages (1, 2, 3... personnes) et non-dimension moyenne des ménages	Même situation que pour les données ordinales, mais des différences plus faibles entre nuances de gris sont acceptables.	Même situation que pour les données ordinales, mais des différences plus faibles entre nuances de gris sont acceptables.

Niveaux de mesure		Exemple	Cartes en noir et blanc	Cartes en couleurs
Continus	Séquentiels	Taux d'alphabétisation (toute valeur comprise entre 0 et 100 %).	Gamme continue de nuances de gris. Le niveau de gris peut ou non être proportionnel aux valeurs de données. Des différences faibles mais observables des niveaux de gris sont acceptables.	Gamme continue de la même couleur ou dans un intervalle spectral partiel. Des variations faibles du niveau de couleur sont acceptables.
	Divergents	Taux de masculinité (inférieur à un = plus de femmes que d'hommes ; supérieur à un = plus d'hommes que de femmes)	On doit recourir à des différences de texture/motif. Des ombrages pleins d'une part et des différences de texture d'autre part peuvent donner de bons résultats.	Couleur neutre (blanc ou gris) au centre avec une gamme continue de deux couleurs différentes de part et d'autre. Par exemple de l'orange clair à l'orange foncé pour les valeurs inférieures à un et du vert clair au vert foncé pour les valeurs supérieures à un.

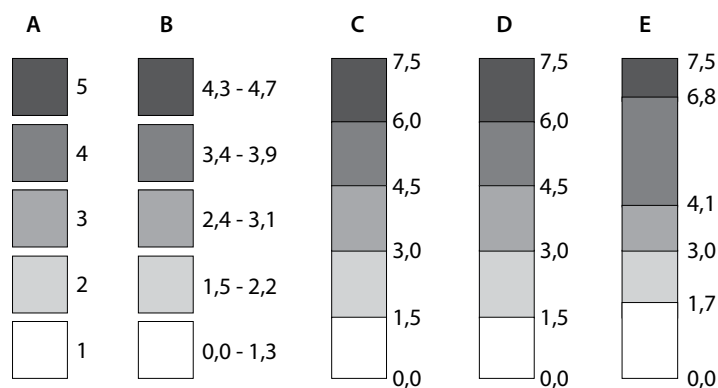
E. Conception de la légende cartographique

A5.109. Le niveau de la mesure peut être reflété dans la conception de la légende, qui établit la relation entre les valeurs de données ou les intervalles de valeurs et les symboles graphiques utilisés. Les logiciels de SIG et les logiciels de cartographie numérique comportent une fonction intégrée de conception des légendes qui convient pour la plupart des applications. Pour une conception cartographique plus soignée, cependant, nous pouvons modifier la légende par défaut, soit dans le module de tracé des logiciels cartographiques, soit dans un logiciel graphique extérieur.

A5.110. La figure A5.28 en fournit quelques exemples. Pour les données catégorielles, les divers cadres de légendes doivent être séparés les uns des autres (voir figure A5.28A). De même, on peut faire ressortir ainsi des séries de catégories non contigües — par exemple, il existe un intervalle entre la limite supérieure d'une catégorie et la limite inférieure de la suivante (voir figure A5.28B). En général, cependant, on devra éviter d'utiliser des légendes de ce type. Les cadres de légendes contigus font ressortir le caractère continu de variables telles que des coefficients ou des densités (voir figure A5.28C). La continuité des valeurs de données ressort encore plus si les cadres individuels de catégories ne sont pas entourés d'un contour (voir figure A5.28D). Enfin, la figure A5.28E montre l'exemple d'une légende pour la classification d'une variable continue avec des séries de catégories irrégulières.

Figure A5.28

Différents types de légendes pour des cartes ombrées

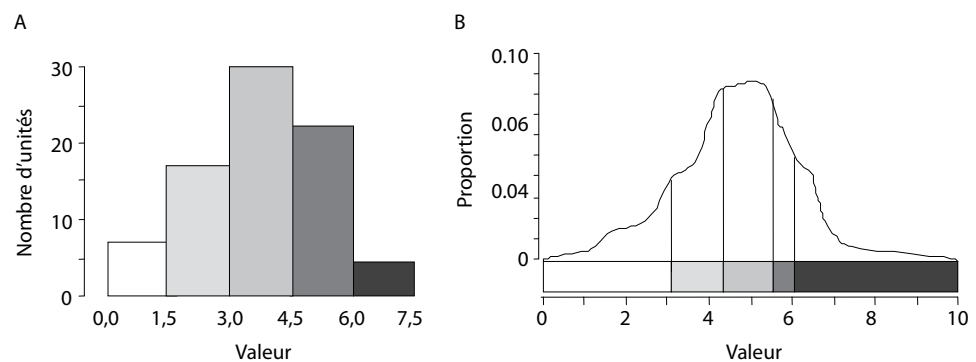


A5.111. Les légendes des figures A5.28C, D et E comportent des points d'inflexion plutôt que des séries de données distinctes. Lorsqu'on utilise des séries de données pour avoir une répartition continue, on se heurte à un problème qui consiste à présenter une même valeur de donnée dans deux catégories, par exemple, 0-10, 10-20, 30-40. Ce problème peut être résolu en utilisant le symbole inférieur à pour affecter chaque valeur à une seule catégorie : par exemple 0-<10, 10-< 20, 20-30. Avec des catégories ouvertes, on peut utiliser le symbole plus grand que ou égal : <10, 10-< 20, >20.

A5.112. La légende peut aussi être intégrée à un diagramme statistique qui résume la répartition des données de la variable. Des histogrammes dans lesquels les couleurs des barres correspondent à la nuance de couleur sont souvent utilisés à cette fin (voir figure A5.29A). Si les intervalles de catégories ne sont pas constants, les barres de l'histogramme peuvent être tracées avec une largeur variable. Si le logiciel cartographique ne prend pas en charge les histogrammes, il est possible d'en tracer dans un programme graphique ou de les importer d'un tableur ou d'un logiciel statistique. Il existe deux possibilités pour déterminer la hauteur de chaque barre. L'approche la plus conventionnelle consiste à utiliser le nombre d'unités géographiques dont les valeurs entrent dans chaque catégorie. Certains logiciels de cartographie numérique affichent le nombre d'unités entrant dans chaque catégorie de la légende. Cela comporte cependant une difficulté : les unités — par exemple des districts — peuvent avoir des tailles de population très différentes. Au lieu d'utiliser le nombre d'unités, on pourra donc déterminer la hauteur des barres de l'histogramme à partir de la taille de la population sous-jacente. Pour une carte de la densité démographique, par exemple, on utilisera le nombre d'habitants présents dans chaque intervalle de densité. Naturellement, la forme de l'histogramme sera très différente et le procédé utilisé devra être indiqué clairement sur la carte ou dans le texte qui l'accompagne.

Figure A5.29

Légendes montrant la répartition des données statistiques



A5.113. Les logiciels statistiques permettent aussi de calculer des schémas de densité montrant la répartition des données sous une forme plus continue qu'avec un histogramme (voir figure A5.29B). La surface située au-dessous de la courbe de densité est égale ou total à un, de telle sorte que la fréquence approchée d'une valeur de donnée individuelle quelconque peut être lue sur le graphique. Des légendes de ce type ont été utilisées, par exemple, dans l'*Atlas of United States Mortality* (*Atlas de la mortalité aux États-Unis*) [National Center for Health Statistics (NCHS), 1997].

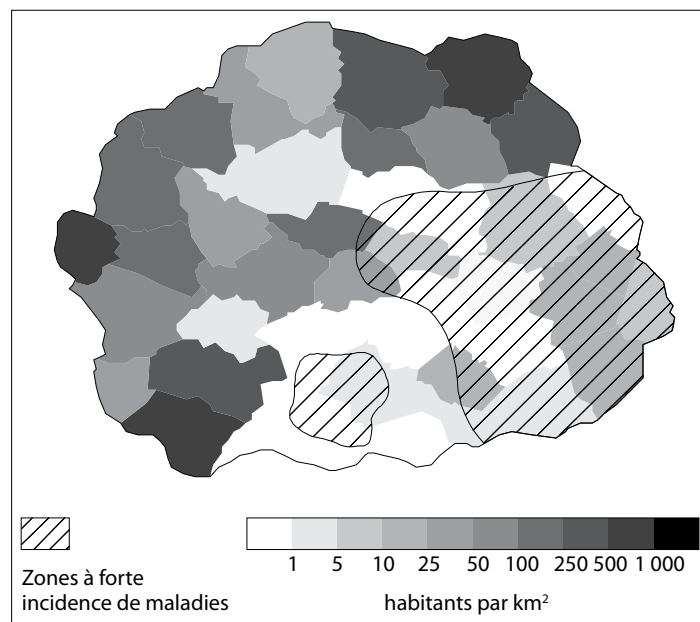
F. Cartes qui racontent des histoires

1. Cartes multivariées

A5.114. À quelques exceptions près, les exemples précédents n'ont montré qu'une variable à chaque fois. C'est le type le plus courant d'affichage utilisé dans un atlas de recensement. À des fins analytiques et pour illustrer les rapports entre variables, nous souhaiterions parfois afficher plus d'une variable à la fois. Dans la section consacrée au choix des couleurs, nous avons fait observer que les cartes multivariées qui utilisent un système de couleurs complexe pour montrer deux variables sur la même carte sont en général difficiles à comprendre. Une autre solution, comme nous l'avons mentionné plus haut, consiste à utiliser un tracé comportant une couleur de fond transparente sur une carte choroplèthe ombrée. Cela donne de bons résultats si la variable superposée ne comporte que quelques catégories ou s'il s'agit d'une variable binaire (présence ou absence) (voir figure A5.30).

Figure A5.30

Combinaison de symboles ombrés pleins et hachurés pour présenter deux variables sur la même carte



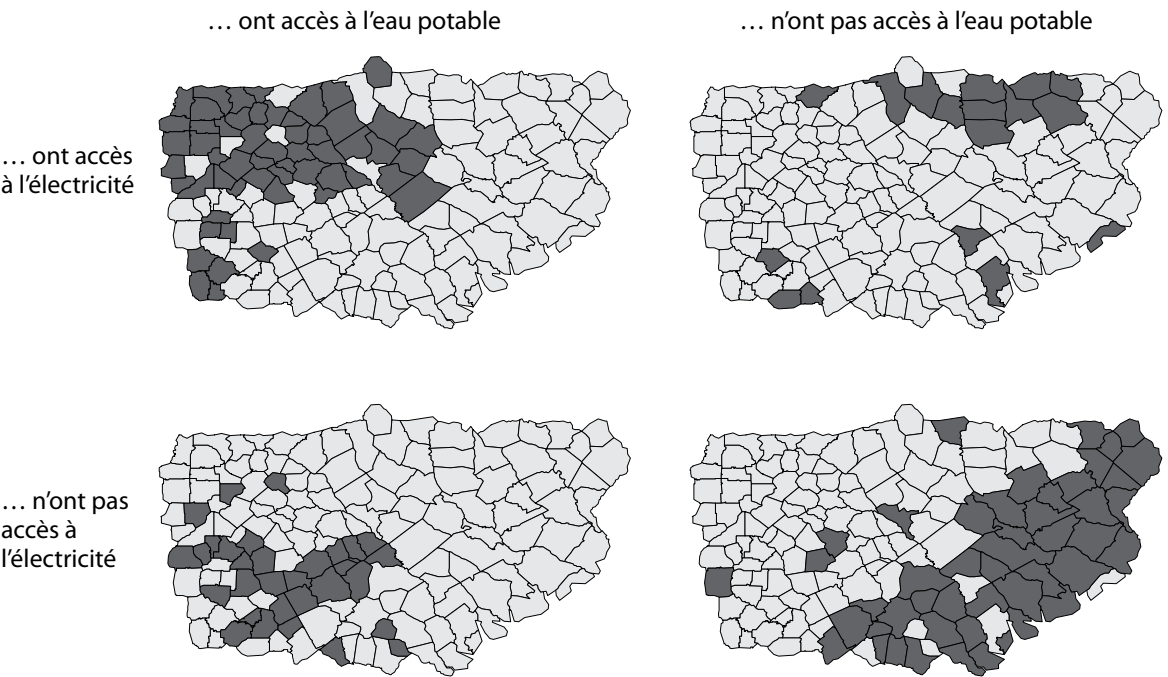
A5.115. Dans l'analyse de données statistiques, deux variables catégorielles qui ne prennent qu'un petit nombre de valeurs sont analysées en utilisant des tabulations croisées. Ces tableaux sont aussi appelés tableaux de contingence. Les lignes et les colonnes d'un tableau à double entrée montrent les catégories des deux variables et les cases montrent le nombre d'observations qui prennent les valeurs correspondantes pour chaque variable. Cet arrangement permet d'évaluer rapidement leurs relations. Par exemple, nous pourrions avoir converti deux variables d'un recensement d'habitations — le pourcentage de ménages ayant accès à l'eau potable et le pourcentage de ménages ayant accès à l'électricité — en deux variables binaires indiquant si la majorité des ménages du district a ou non accès à ces services publics. La tabulation croisée peut ressembler à ce qui suit :

Majorité des ménages ayant...			
	... accès à l'eau potable	... n'ayant pas accès à l'eau potable	Total
...accès à l'électricité	55	17	72
...n'ayant pas accès à l'électricité	31	48	79
TOTAL	86	65	151

A5.116. Si nous voulons présenter ces informations sous forme géographique, nous pouvons établir une carte comportant quatre catégories, une pour chaque case de la tabulation croisée. Cependant, étant donné que ces quatre catégories ne comportent aucun classement naturel, il sera difficile pour le lecteur de découvrir des tendances sur une telle carte. Une meilleure méthode consistera à traduire le concept d'un tableau à double entrée directement en langage cartographique. La figure A5.31. montre une carte équivalant à un tableau à double entrée. Chaque carte indique les districts corrélatifs à la case correspondante dans le tableau à double entrée. Les cartes n'ont pas besoin d'être assorties d'une légende détaillée, car la nuance sombre fait nettement ressortir les districts présentant de l'intérêt.

Figure A5.31
Carte équivalant à un tableau à double entrée

La majorité des habitants ...



A5.117. Des tendances apparaissent immédiatement même sur une petite carte qui ne couvre qu'environ un tiers de page. La plupart des districts du nord-ouest ont accès à la fois à l'eau potable et à l'électricité, tandis que la majorité des ménages dans les districts du sud-est n'ont accès ni à l'un ni à l'autre. Dans les tabulations croisées, les cases hors diagonale sont souvent les plus intéressantes. Dans certains districts du nord-est la plupart des ménages n'ont pas accès à l'eau potable, mais ont accès à l'électricité. Dans un groupe de districts au sud-ouest, la situation est inversée.

A5.118. Cette approche pourrait être étendue à des tableaux plus complexes, par exemple lorsqu'une variable présente trois valeurs (à savoir faible, moyenne et éle-

vée) et une autre comporte deux catégories. Il n'est pas nécessaire de tracer de grandes cartes. Même avec de nombreuses unités géographiques — dans le présent exemple 151 districts — de petites cartes suffisent, car seules deux couleurs ou nuances de gris contrastées sont requises.

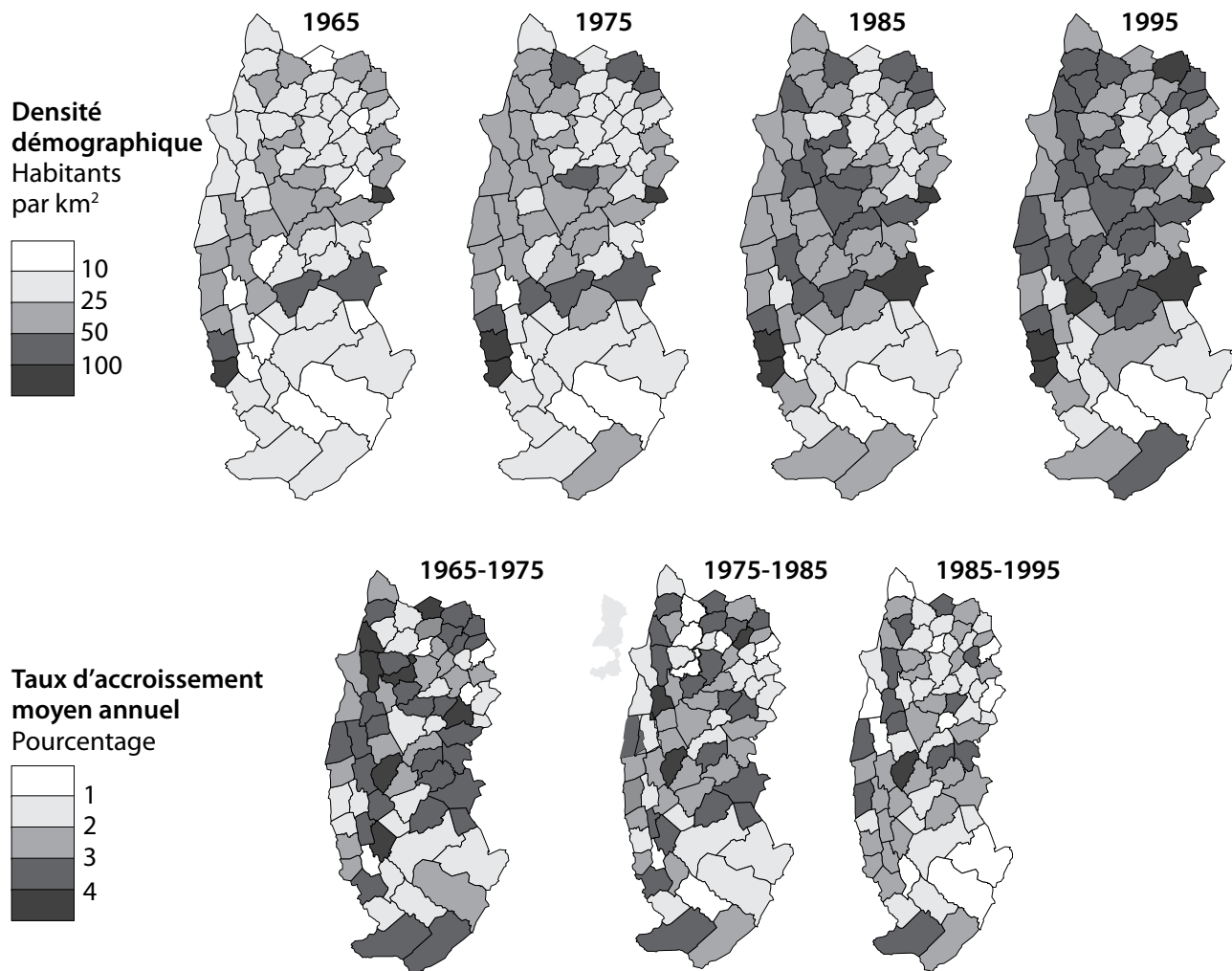
2. Petites cartes multiples

A5.119. On peut aussi présenter efficacement des informations évolutives en disposant les données sur plusieurs cartes. La figure A5.32 montre l'accroissement démographique dans le temps à partir des chiffres provenant de quatre recensements successifs. Les cartes de densité démographique montrent où l'accroissement démographique a été le plus élevé. Pour permettre des comparaisons dans le temps, les limites de catégories doivent être les mêmes sur toutes les cartes. Cela signifie que les systèmes de classification fondés sur la répartition des données (par exemple les inflexions naturelles) ne sont pas appropriés. Les cartes de densité sont complétées

Figure A5.32

Petites cartes multiples — représentation de l'évolution dans le temps

Dynamique de la population, 1965-1995



par trois cartes plus petites montrant les taux d'accroissement démographique moyen annuel entre les recensements.

A5.120. Des présentations du type présenté sur la figure A5.32 sont appelées petites cartes multiples [Bertin (1983) et Tufte (1983)]. Le même tracé de carte est répété pour chaque année ou chaque sous-groupe de population. Le tracé étant le même pour toutes les cartes, elles s'avèrent faciles à comprendre pour le lecteur. Cela permet au concepteur de la carte de présenter une plus grande densité d'informations que ce qui serait autrement possible. Les relations multivariées sont souvent plus évidentes avec des présentations comportant plusieurs cartes qu'avec des cartes composites accompagnées d'une légende éventuellement compliquée.

A5.121. Un document des Nations Unies, *Geographical Information Systems for Population Statistics* [Nations Unies (1997); voir figure 4.8] offre un autre exemple de carte utilisant le concept des petites cartes multiples. Il montre les taux de masculinité concernant des tranches d'âges de cinq ans dans 75 districts au Népal. Le graphique montre 17 petites cartes, avec une classification divergente centrée autour d'un taux de masculinité équilibré. La version en couleurs de cette carte montre un excédent de femmes dans différentes nuances de rouge et un excédent d'hommes sous forme de bleus. La version en noir et blanc utilise des nuances de gris pleines pour un excédent de femmes et une combinaison de points de densité variable pour un excédent d'hommes. À n'en pas douter, la couleur renforce le message communiqué par ces cartes. Bien qu'elles contiennent une grande quantité d'informations, les cartes peuvent être interprétées très facilement, car les groupements de valeurs semblables sont très apparents. Il est évident qu'un tableau de 1 275 valeurs (17 fois 75) serait beaucoup plus difficile à interpréter que les mêmes informations présentées sous forme de cartes géographiques. En fait, le graphique correspondant au Népal fait ressortir clairement certaines tendances qui peuvent être attribuées à la migration pendant le cycle de vie des hommes dans un certain nombre de districts.

Annexe VI

Glossaire

Accès sélectif — Dégradation de l'exactitude des signaux de satellite GPS conservée jusqu'en 2000 par le Département de la défense des États-Unis. Elle pourrait toutefois être réinstaurée en cas de guerre.

Adresse — Nombre ou désignation semblable affectée à une unité d'habitation, une entreprise ou tout autre bâtiment. Les adresses sont surtout utilisées pour la distribution du courrier, mais elles sont aussi importantes à des fins administratives, par exemple dans les systèmes d'enregistrement des déclarations d'état civil et pour les recensements.

American Standard Code for Information Interchange (ASCII) [Code standard américain pour l'échange d'informations] — Code informatique mis au point pour faciliter l'échange de données alphanumériques et de caractères spéciaux entre ordinateurs et entre systèmes d'exploitation. Un code au niveau de l'octet, de valeur comprise entre 0 et 255, est affecté à chaque caractère.

Analyse de réseau — Procédures d'analyse des relations entre des points ou des adresses sur un ensemble de lignes dans une base de données qui peuvent représenter par exemple un réseau de voirie. On utilise l'analyse de réseau pour les choix d'implantations ou de cheminement par exemple pour la gestion des situations d'urgence.

Analyse spatiale — Ensemble de techniques utilisées pour extraire des informations utiles de données géoréférencées. L'analyse spatiale inclut l'intégration d'ensembles de données géographiques, des méthodes qualitatives et quantitatives d'évaluation de données, de modélisation, d'interprétation et de prévision. Dans un SIG, l'analyse spatiale se réfère souvent à des méthodes d'intégration de données de SIG telles que la superposition de polygones ou l'analyse de voisinage. Dans un sens plus large, elle inclut, par exemple, des modèles de traitement spatial (par exemple dynamique de la migration) et des statistiques spatiales (par exemple modèles de régression rendant compte des relations et arrangements spatiaux entre observations).

Annotation — Texte utilisé pour affecter un libellé à des objets sur la carte. L'annotation peut être enregistrée dans un SIG et tracée sur les cartes en vue de l'affichage ou de l'impression. Contrairement aux informations texte sur un tableau d'attributs, l'annotation n'est utilisée que pour l'affichage cartographique et non à des fins d'analyse.

Appariement d'adresses — Processus par lequel on apparie à des emplacements géographiques sur un réseau de voirie des informations générales relatives aux attributs en utilisant une adresse réelle. Par exemple, on peut mettre en correspondance un registre d'adresses sous forme de tableau et une carte numérique détaillée du réseau de voirie pour établir une couche de points de SIG montrant

l'emplacement de chaque ménage. Cette opération est parfois appelée également **géocodage**.

Arc — Voir **Ligne**.

Attribut — Caractéristique d'un objet géographique. Par exemple, zone de texte ou champ numérique enregistré dans un tableau de base de données relationnelle qui peut être associé aux objets géographiques d'un SIG. Les attributs d'un district de recensement, par exemple, peuvent être son identificateur unique, sa surface en kilomètres carrés, sa population totale et le nombre de ménages. On distingue parfois les attributs géographiques et les attributs généraux. Les premiers sont enregistrés dans un tableau de données qui est étroitement associé aux fichiers de coordonnées géographiques et qui contient des champs tels que des identificateurs internes, des codes d'objets et des surfaces. Les attributs généraux sont normalement enregistrés dans des tableaux de données séparés qui peuvent être associés au tableau des attributs géographiques.

Bande — Couche d'une image multispectrale de télédétection montrant les signaux mesurés dans un intervalle défini du spectre électromagnétique. Voir aussi **Image multispectrale**.

Base de données — Collection logique d'informations interdépendantes, gérées et enregistrées sous la forme d'un ensemble, par exemple dans le même fichier informatique. Les expressions « base de données » et « ensemble de données » sont souvent utilisées l'une pour l'autre. Une base de données de SIG contient des informations sur l'emplacement des objets du monde réel et les caractéristiques de ces objets.

Base de données géographique — Rassemblement logique de données concernant des objets associés à des lieux situés sur la surface terrestre.

Beidou — Alternative au système GPS proposé par la Chine, constitué de 35 satellites et offrant une exactitude pouvant atteindre 10 mètres.

Binaire — Constitué de deux ou se référant à deux, comme les variables binaires (oui /non). Également forme de codage informatique fondé sur des éléments individuels d'information appelés bits qui peuvent prendre deux valeurs, 0 et 1.

Bit — Élément binaire qui peut prendre une valeur de 0 ou de 1.

Bits par seconde (BPS) — Mesure de la vitesse de transfert dans les réseaux numériques de télécommunications.

Cadre géographique de recensement — Unités de collecte et de présentation utilisées par un bureau de recensement pour le dénombrement censitaire et la tabulation des données. Ce cadre englobe la structure hiérarchique des unités de recensement et des unités administratives, leurs désignations et leurs codes et les relations entre les différentes unités. Voir **Hiérarchie géographique**.

Canal — Partie de l'électronique d'un récepteur GPS qui capture le signal du satellite. Les récepteurs multicanaux captent et traitent simultanément des signaux provenant de plusieurs satellites.

Canevas géodésique — Réseau de repères ou points fondamentaux mesurés avec précision et exactitude qui servent de base pour obtenir de nouvelles mesures d'emplacements. Appelé également réseau de points de repère.

Carte — Représentation d'une partie de la surface terrestre tracée sur une surface plane (par exemple sur papier ou écran).

Carte choroplèthe — Carte statistique sur laquelle les valeurs enregistrées pour les unités déclarantes sont affectées en premier lieu à un certain nombre de caté-

gories ou d'intervalles de classes discontinus. Les unités déclarantes sont alors ombrées en utilisant des symboles (couleurs ou motifs) choisis pour chaque catégorie.

Carte de base — Carte montrant des objets géographiques essentiels qui peuvent être utilisés comme références de localisation. On peut citer des échantillons d'objets comme les routes, les limites administratives et les localités. Les cartes de base sont utilisées pour rassembler des données géographiques nouvelles ou comme référence pour l'affichage des données d'une carte thématique.

Carte de flux — Carte montrant les mouvements, par exemple de marchandises ou de personnes, suivant un cheminement linéaire.

Cartes de points — Une carte dans laquelle les quantités ou les densités sont représentées par des points. En général, chaque point représente un nombre défini d'objets discontinus tels que du bétail ou des habitants. Les points peuvent être disposés aléatoirement dans les unités déclarantes ou ils peuvent être placés de façon à refléter la véritable répartition sous-jacente de la variable.

Carte de référence — Dans le contexte de la cartographie censitaire, produit cartographique (sur papier ou numérique) montrant une partie du cadre géographique de recensement, c'est-à-dire un ensemble de données ou une unité de diffusion statistique.

Carte marine — Carte conçue en priorité pour la navigation maritime et aérienne, par exemple carte nautique ou aéronautique.

Carte planimétrique — Carte qui, contrairement à la carte topographique, ne montre que les emplacements d'objets, mais pas leur altitude. La carte planimétrique peut inclure les mêmes objets qu'une carte topographique, à l'exception des courbes de niveau, mais elle ne montre en général que des objets sélectionnés dans un but particulier.

Carte schématique — Voir **Croquis cartographique**.

Carte thématique — Carte qui illustre un concept, un sujet ou un thème spécifique. Une carte thématique peut présenter des informations quantitatives ou qualitatives.

Carte topographique — Carte présentant essentiellement des objets du monde réel, notamment des courbes de niveau, des cours d'eau, des routes, des localités et des points de repère. Les feuilles de carte standard créées par les services cartographiques nationaux à différentes échelles sont le plus souvent des cartes topographiques.

Cartogramme — Carte construite en mettant à l'échelle les unités déclarantes en fonction de la valeur d'une variable enregistrée pour elles. Appelé aussi représentation de la valeur par unité de surface.

Cartographie — Art et science consistant à créer une représentation bidimensionnelle d'une partie de la surface terrestre. Les objets représentés peuvent être des objets réels (cartographie topographique) ou peuvent représenter des concepts et caractéristiques plus abstraits (cartographie thématique).

Cartographie automatisée/gestion d'installations — Applications de SIG dans le secteur des services collectifs et des travaux publics intéressant particulièrement les questions d'ingénierie et de maintenance.

Centroïde — Centre mathématique d'un polygone. Pour des polygones de forme irrégulière, on peut le considérer comme un « centre de gravité ».

Chaîne — Voir **Ligne**.

Champ — Colonne dans un tableau de base de données.

Classification — Affectation d'objets à des groupes partageant les mêmes caractéristiques ou des caractéristiques semblables. En cartographie, processus consistant à affecter des symboles à des objets cartographiques semblables ou ayant des valeurs semblables. La classification est utilisée pour simplifier la carte de façon à mieux faire passer le message du cartographe.

Clé étrangère — Dans des systèmes de gestion de base de données relationnelle, champ ou rubrique d'un tableau qui contient une valeur identifiant des lignes dans un autre tableau. Elle est utilisée pour réunir deux tableaux en définissant la relation entre deux éléments d'une base de données relationnelle. La clé étrangère est la clé primaire dans l'autre tableau.

Clé primaire — Un ou plusieurs champs dans un tableau d'attributs qui assurent l'identification unique d'une instance, d'une ligne ou d'un enregistrement spécifique.

Client — Ordinateur qui utilise des données ou un logiciel enregistrés sur un autre ordinateur (serveur) souvent éloigné.

Codage par longueur de ligne — Technique de compression pour des données matricielles, rectangulaires ou des données image. Au lieu d'enregistrer chaque valeur de cellules adjacentes présentant la même valeur, le système enregistre la valeur et le nombre de répétitions de celle-ci. La compression est importante quand des objets discontinus sont enregistrés dans un SIG matriciel.

Code — Caractères alphanumériques utilisés pour identifier des objets géographiques. Les codes sont aussi utilisés pour identifier des catégories d'attributs tels que des intervalles de densité démographique, des classes d'utilisation des sols ou des industries. Voir aussi **Code géographique**.

Code géographique — Identificateur alphanumérique unique affecté à une unité juridique, administrative, statistique ou à une unité déclarante.

Coïncidence géographique — Lorsque deux objets géographiques ou plus partagent le même emplacement ou les mêmes limites. Par exemple, certaines unités déclarantes ou statistiques peuvent aussi être des circonscriptions administratives.

Colonne — Dans un SIG, groupe de cellules ou de pixels alignés verticalement dans un réseau ou une base de données maillée de SIG. Dans les systèmes de gestion de base de données, champ ou rubrique d'un tableau d'attributs.

Compilation d'une carte — Processus consistant à assembler, évaluer et interpréter des mesures et des documents cartographiques en vue d'établir une nouvelle carte.

Composition d'une carte — Arrangement des éléments de la carte en vue de créer un produit cartographique qui soit visuellement attrayant et reproduise correctement les phénomènes présentés.

Computer Graphics Metafile (CGM) — Format de fichier normalisé pour échanger des images ou des données vectorielles.

Conception assistée par ordinateur (CAO)/conception et dessin assistés par ordinateur — Système logiciel qui comporte des outils de dessin et de conception destinés spécialement à des applications d'ingénierie et d'architecture. Les systèmes de CAO utilisent un système de coordonnées graphiques et sont donc semblables aux systèmes d'information géographique.

Connectivité — Dans un SIG topologique, cas où deux lignes ou plus sont réunies en un seul point ou nœud.

Contiguïté — Cas dans lequel deux objets géographiques ou plus sont voisins ou adjacents.

Contrôle de la qualité — Étapes et procédures d'un projet de création de base de données ou d'un système de production cartographique qui assurent que les données ou produits obtenus soient conformes à des normes spécifiées d'exactitude et d'exploitabilité.

Conversion de données — Transfert de données d'un format à un autre. En général, la conversion de données se réfère au passage à la forme numérique de l'information cartographique sur papier. Dans un sens plus large, la conversion de données géographiques englobe aussi le transfert d'informations numériques d'un format de fichier SIG à un autre.

Coordonnée — Deux ou trois nombres qui décrivent la position d'un point en deux ou trois dimensions (par exemple x/y ou x/y/z, où z désigne l'altitude). Une coordonnée à deux dimensions est parfois appelée une paire de coordonnées, une coordonnée à trois dimensions, un triplet de coordonnées. Dans les bases de données de SIG, les coordonnées représentent les emplacements correspondants à la surface terrestre par rapport aux autres emplacements.

Couche — Ensemble individuel de données de SIG contenant des objets qui intéressent le même thème, comme des routes ou des maisons. Le terme « couche » se réfère à la capacité d'un SIG de superposer et de combiner différentes couches thématiques référencées dans le même système de coordonnées. Appelée aussi **couverture**.

Couche thématique — Voir **Couche**.

Courbe de niveau — Ligne sur une carte qui relie les points de même altitude. Voir aussi **Isoligne**.

Couverture — Dans un SIG, la couverture désigne parfois un ensemble de données vectorielles de SIG qui contient des objets géographiques concernant un seul thème, comme des unités censitaires ou des routes.

Croisement — Fonction de SIG utilisée pour intégrer topologiquement ou combiner deux couches de données spatiales de telle sorte que seuls les objets situés dans la zone commune aux deux soient préservés.

Croquis cartographique — Carte (souvent tracée à la main) montrant les principaux objets d'une zone donnée, mais qui peut ne pas présenter un degré élevé de précision de localisation et par conséquent risque de ne pas représenter correctement les distances et les dimensions des objets. Un croquis cartographique peut cependant posséder un degré élevé d'exactitude logique, ce qui signifie que les relations entre objets sont correctement représentées. Appelé également **carte schématique**.

Dépassement de trait — En numérisation, ligne tracée au-delà du point où elle devait se raccorder à une autre ligne. Le segment de ligne erroné qui en résulte est parfois appelé segment suspendu (*dangle*).

Dictionnaire de données — Catalogue de données qui décrit le contenu d'une base de données. Il regroupe sous la forme d'une liste les informations concernant chaque champ dans les tableaux d'attributs ainsi que le format, les définitions et les structures des tableaux d'attributs. Un dictionnaire de données est un élément essentiel de l'information sous forme de métadonnées.

Dispositif centralisateur — Dans le contexte des infrastructures nationales de données spatiales, dépôt permettant de rassembler et de diffuser les données et métadonnées de SIG.

District de recensement — En général, la plus petite unité géographique pour laquelle des données censitaires sont agrégées, compilées et diffusées. Un district de re-

censement est défini par des limites décrites sur un croquis cartographique ou dans une base de données de SIG. Ces limites peuvent ou non être visibles sur le sol. Appelé également îlot ou secteur de recensement.

Documents de référence — Données et informations d'un type quelconque utilisées pour compiler une carte ou une base de données de SIG. Ils peuvent inclure des observations sur le terrain, des photographies aériennes et terrestres, des images satellitaires, des croquis, des cartes thématiques, topographiques, hydrographiques ou hypsographiques, des croquis et dessins cartographiques, des informations sous forme de tableaux et des rapports écrits concernant des objets géographiques naturels ou artificiels.

Données-cadres — Dans le contexte d'activités nationales de SIG, ensemble de thèmes géographiques à usage général ou de données de base, telles que limites administratives, altitude ou infrastructure de transport. Les initiatives relatives à une infrastructure de données-cadres ou de données spatiales nationales visent à coordonner l'élaboration et la normalisation d'ensembles de données de SIG intéressant les données-cadres dans un pays.

Données de base — Voir **Données-cadres**.

Données spatiales — Informations concernant la localisation, les dimensions, la forme d'objets géographiques et leurs relations réciproques. Dans un SIG, les données spatiales sont classées techniquement sous forme de points, de lignes, de surfaces et de matrice.

Données vectorielles — Modèle de données de SIG dans lequel l'emplacement et la forme des objets sont représentés par des points, lignes et surfaces essentiellement constitués de coordonnées x,y.

Échelle — En cartographie, relation entre la distance sur la carte et la distance correspondante sur la surface terrestre. L'échelle est indiquée sous la forme d'un rapport, par exemple, 1/100 000 qui signifie qu'un centimètre sur la carte équivaut à 100 000 cm sur la surface terrestre. L'échelle étant un rapport, une carte à petite échelle montre une surface relativement étendue, tandis qu'une carte à grande échelle montre une petite surface. Plus généralement, l'échelle se réfère au niveau de l'observation ou de l'enquête, qui peut aller des phénomènes à petite échelle aux phénomènes à grande échelle.

Éléments d'une carte — Éléments d'une carte thématique ou topographique tels que le titre, la légende, l'échelle, la flèche indiquant le nord, le graticule, les limites et bordures de finition.

Ellipsoïde — En cartographie, forme tridimensionnelle utilisée pour représenter la Terre. L'ellipsoïde terrestre est caractérisé par une distance plus courte du centre aux pôles (demi petit axe) que du centre à l'équateur (demi grand axe). Appelé également sphéroïde.

Ensembles de données — Collection logique de valeurs ou d'objets d'une base de données concernant un sujet unique.

Entité — Phénomène du monde réel d'un type donné. Dans les systèmes de gestion de base de données, collection d'objets (par exemple personnes ou lieux) qui partagent les mêmes attributs. Les entités sont définies lors de l'élaboration conceptuelle de la base de données.

Enveloppe (buffer) — Zone ou surface à une distance spécifiée autour d'un objet géographique (points, lignes ou polygones). L'utilisation d'enveloppes est l'une des fonctions géospatiales fondamentales.

Équateur — En cartographie, le parallèle de référence, autrement dit la latitude 0° nord et sud.

Étendue de la carte — Coordonnées en unité cartographiques définissant le rectangle qui comprend tous les objets figurant dans une présentation spécifique de la carte ou dans une base de données de SIG, c'est-à-dire les coordonnées x et y minimales et maximales d'une base de données numériques ou la partie d'une base de données reproduite dans une présentation de carte.

Étirement élastique — Procédure consistant à modifier de façon non uniforme la forme et l'emplacement d'objets dans une base de données de SIG. L'étirement élastique est souvent utilisé pour introduire dans un système de coordonnées connu un ensemble de données de SIG provenant d'un système de coordonnées inconnu. Les adaptations sont définies en spécifiant de nombreux liens entre des emplacements figurant dans l'ensemble de données d'entrée et leurs points de référence ou de contrôle correspondants dans le système de coordonnées de sortie.

Exactitude — Absence d'erreur. Mesure dans laquelle une mesure ou une représentation est conforme aux valeurs véritables du monde réel. La détermination du niveau d'exactitude acceptable et l'élaboration d'une norme d'exactitude figurent parmi les premières phases d'un projet de SIG. L'exactitude ne doit pas être confondue avec la précision, qui se réfère à la capacité de distinguer de petites quantités lors d'une mesure. Par exemple, l'emplacement d'un point peut être mesuré avec précision (c'est-à-dire avec quatre décimales significatives) mais de façon inexacte (à plusieurs mètres de distance de son emplacement dans le monde réel).

Exactitude logique — Terme indiquant dans quelle mesure les relations entre objets géographiques sur une carte ou dans une base de données de SIG sont représentées correctement (par exemple adjacent à, relié à). Une base de données de SIG peut être logiquement exacte, même si l'exactitude de la localisation est limitée.

Fichier d'attributs géographiques — Tableau de base de données étroitement lié aux objets spatiaux enregistrés dans un fichier de coordonnées de SIG. Le fichier ou tableau d'attributs géographiques contient des informations spécifiques sur chaque objet tels que l'identificateur, le nom et la superficie. Dans certains systèmes, ce fichier est également appelé tableau d'attributs ponctuels, linéaires ou polygonaux. Les données enregistrées dans des tableaux externes peuvent lui être rattachées en utilisant une base de données relationnelle.

Fichier de référence géographique — Fichier de base tabulaire et numérique contenant la liste des noms, codes géographiques et éventuellement des attributs de toutes les entités géographiques intéressant le recensement et la collecte de données topographiques.

Format de données — Désigne en général un ensemble spécifique de structures de données, éventuellement propre au fournisseur, dans un système logiciel.

Format DXF (Drawing Exchange Format) — Format ASCII utilisé pour décrire un graphique ou un dessin élaboré par Autodesk, Inc. (Sausalito, Californie). Élaboré initialement pour des applications de CAO, il est aussi devenu une norme pour l'échange de données de SIG.

Fusion — Fonction de SIG qui supprime les limites entre des polygones adjacents présentant la même valeur pour un attribut particulier. Par exemple, des polygones de districts de recensement peuvent être fusionnés à partir du code de leurs unités de supervision pour créer des cartes de supervision.

Galileo — Alternative au système GPS développée par l'Union européenne. Composé de 30 satellites et deux stations au sol, Galileo sera compatible avec le système GPS des États-Unis au niveau de l'utilisateur.

Généralisation — Voir **Généralisation cartographique**.

Généralisation cartographique — Processus permettant d'isoler des objets du monde réel en réduisant le degré de détail pour permettre leur représentation sur une carte. Nécessite des opérations de sélection, de classification, de simplification et de symbolisation.

Géocodage — a) Processus consistant à affecter des codes géographiques à des objets dans une base de données numériques; b) Fonction de SIG déterminant la localisation d'un point à partir d'une adresse. Voir aussi **Appariement d'adresses**.

Géométrie coordonnée (COGO) — Terme utilisé par les arpenteurs-géomètres pour désigner des mesures précises d'emplacements.

Géoréférencement — Processus consistant à déterminer la relation entre les coordonnées de pages et les coordonnées du monde réel. Le géoréférencement est nécessaire après la numérisation, par exemple pour convertir les coordonnées de page mesurées en unités de numérisation (c'est-à-dire en centimètres ou en pouces) en coordonnées du monde réel utilisées pour tracer la carte de référence. Voir aussi **Transformation**.

Géospatial — Terme parfois utilisé pour décrire des informations de nature géographique ou spatiale.

Geo-TIFF — Voir **Tagged image file format**.

Global Navigation Satellite System (GLONASS) — Système exploité par le Ministère de la défense de la Fédération de Russie et correspondant au système GPS des États-Unis. Il est très semblable au GPS, mais ne fait pas l'objet d'une disponibilité sélective. Certains récepteurs combinent les signaux de GPS et GLONASS pour améliorer l'exactitude des coordonnées.

GPS différentiel (GPSD) — Ensemble de techniques utilisées pour améliorer la précision de coordonnées saisies par un récepteur GPS en calculant l'erreur de signal (décalage) pour un second récepteur GPS (la station de base) situé dans un lieu déterminé avec précision et exactitude. Le facteur de correction est appliqué aux coordonnées saisies par l'unité mobile, soit en temps réel, soit en mode post-processeur (c'est-à-dire en utilisant une base de données d'informations de correction). Dans certaines parties du monde, des informations de correction différentielle sont diffusées en permanence par un ensemble de stations de base. Voir aussi **Station de base**.

Grand cercle — Cercle formé par un plan passant par le centre d'une sphère. Par exemple, tous les méridiens et l'équateur sont des grands cercles. Sur la sphère, le plus court chemin entre deux points se trouve sur le grand cercle passant par ces deux points.

Graphic Interchange File (GIF) — Format de fichier image graphique élaboré initialement pour transmettre des images par l'intermédiaire de serveurs électroniques de messages et d'informations. Le format GIF, qui permet une compression efficace de la dimension des fichiers, est utilisé pour la plupart des graphismes sur les pages Web.

Graticule — En cartographie, réseau des longitudes et des latitudes tracé sur une carte.

Hierarchie géographique — Dans le contexte de la cartographie censitaire, système d'unités de surface généralement emboîtées créées à des fins administratives ou

pour la collecte de données. Par exemple, un pays est subdivisé en provinces, divisées en districts, etc., jusqu'au niveau le plus bas qui peut être le district de recensement. Voir aussi **Cadre géographique de recensement**.

Hydrographie — Éléments appartenant aux eaux superficielles, tels que lacs, cours d'eau et canaux.

Hypsographie — Éléments faisant partie du relief ou concernant l'altitude du terrain.

Image — Représentation d'une partie de la surface terrestre. En général, cependant, l'image est produite au moyen d'un dispositif optique ou de détection électronique. Par exemple, les photographies aériennes scannées ou les données de télé-détection sont généralement qualifiées d'images. En ce qui concerne l'enregistrement et le traitement des données, une image est très semblable à un maillage ou à une matrice.

Image multispectrale — Ensemble de données de télédétection constitué d'un certain nombre de bandes ou de couches. Il s'agit essentiellement d'images séparées d'une même zone prises simultanément et dont chacune présente le signal d'un intervalle différent du spectre électromagnétique.

Image panchromatique — Image de télédétection qui enregistre le signal dans un large intervalle du spectre électromagnétique, de la même façon qu'une photographie en noir et blanc.

Image satellitale — Ensemble de données numériques enregistrées à partir d'un satellite sur orbite terrestre soit photographiquement, soit par un détecteur à balayage à bord du satellite. Une image satellitale dans un SIG est semblable à un ensemble de données matricielles ou rectangulaires.

Index spatial — Tableau ou structure de recherche de renseignements dans une base de données géographique qui est utilisé par un SIG ou un système de gestion de base de données pour accélérer les requêtes, les opérations analytiques et l'affichage d'objets spatiaux.

Information cadastrale — Enregistrement qui décrit les droits et intérêts passés, présents et futurs concernant la propriété des terres à des fins juridiques. Les cartes cadastrales indiquent l'emplacement géographique et l'étendue des parcelles de terrain. Dans de nombreux pays, les levés cadastraux utilisent aujourd'hui des SIG pour enregistrer ces informations. Appelée également information relative aux titres fonciers.

Infrastructure — Système de travaux publics dans un pays, un état ou une région, comprenant les routes, les lignes des services d'utilité publique et les bâtiments publics.

Infrastructure de données spatiales — Voir **Données cadres**.

Inscription d'un point dans un polygone (Point-in polygon) — Opération d'un SIG par laquelle des objets ponctuels sont combinés à des polygones pour déterminer quel point doit être attribué à un polygone. Cette opération permet d'ajouter des attributs de polygones à chaque enregistrement correspondant dans le tableau des attributs de points (par exemple de informations relatives à une circonscription de service sanitaire pour un point d'un échantillon d'enquête) ou bien des attributs de points peuvent être résumés pour chaque polygone correspondant (par exemple le nombre d'hôpitaux dans chaque district).

Inscription d'une ligne dans un polygone — Opération d'un SIG par laquelle des objets linéaires sont combinés à des objets polygonaux pour déterminer quelles lignes sont attribuées à tel ou tel polygone. Par cette opération, des attributs

polygonaux peuvent être ajoutés à chaque enregistrement correspondant dans le tableau des attributs linéaires (par exemple le district dans lequel se trouve la route) ou des attributs linéaires peuvent être résumés pour chaque polygone correspondant (par exemple la longueur totale d'une route dans un district).

Insuffisance de trait — Lors de la numérisation, ligne qui n'a pas été prolongée jusqu'au point où elle aurait dû se raccorder à une autre ligne.

Intégration — Dans un SIG, processus consistant à compiler un ensemble cohérent de données spatiales à partir de sources hétérogènes. L'intégration verticale se réfère à la capacité du SIG de combiner différentes couches de données référencées dans le même système de coordonnées

Intégration horizontale — Technique d'édition manuelle ou automatisée dans un SIG appariant des objets partagés numérisés à partir de feuilles de carte adjacentes. L'intégration horizontale peut être nécessaire, par exemple, pour raccorder des routes ou des limites d'unités administratives après avoir réuni des cartes numérisées séparément.

Intégration verticale — Voir **Intégration**.

Interaction spatiale — Interdépendance entre des entités géographiques. Elle se réfère souvent au flux de marchandises, services, informations ou personnes entre divers lieux géographiques. L'analyse d'interaction spatiale est importante pour l'étude de la migration humaine.

Internet — Système mondial de réseaux informatiques interconnectés qui offre des services de communication de données tels que l'ouverture de session à distance, le transfert de fichier, le courrier électronique, les serveurs de messages et d'informations et les groupes de discussion. Internet est aussi la base du World Wide Web (WWW).

Internet Protocol (IP) — Le plus important ensemble de codes et de conventions permettant le transfert de données numériques sur Internet.

Interpolation — Processus consistant à évaluer une valeur de variable en un lieu à partir de valeurs mesurées en des lieux voisins. Utilisé pour produire un ensemble de données complet aux points de grille à partir d'informations d'échantillonnage ponctuel, par exemple une surface de précipitations à partir de stations pluviométriques.

Interpolation aérolaire — Transfert d'un attribut d'un ensemble de zones déclarantes à un autre ensemble de zones incompatible; par exemple, estimation des totaux de population pour des régions écologiques sur la base d'un ensemble de données de SIG concernant la population par district.

Isoligne — Ligne sur une carte « isarithmique » reliant des points de valeur constante. L'exemple le mieux connu est l'isohypse, qui montre des lignes d'égale altitude. Voir aussi **Courbe de niveau**.

Java — Langage de programmation qui permet de créer des logiciels utilisable sur des plateformes (systèmes d'exploitation) multiples. Des programmes Java appelés miniprogrammes peuvent être envoyés ou recherchés sur Internet pour être exécutés sur un ordinateur distant.

Joint Photographic Experts Group (JPEG) — Format de fichier graphique utilisé surtout pour des images photographiques et qui autorise une compression importante des dimensions de fichiers.

Jointure — Dans les systèmes de gestion de bases de données relationnelles, processus consistant à joindre des valeurs d'un tableau de base de données à un autre tableau en reliant une clef étrangère à la clef primaire dans la table extérieur.

Langage de requêtes structuré (SQL) — Dans les systèmes de gestion de données relationnelles, syntaxe standard utilisée pour définir, manipuler et extraire des données.

Largeur de bande — Quantité ou volume de données numériques pouvant être transféré(e) par une connexion de télécommunications.

Latitude — « Coordonnée y » dans un système de coordonnées polaires sur une sphère. Mesurée comme la distance angulaire en degrés au nord et au sud de l'équateur. Désignée également par le mot parallèle.

Légende — En cartographie, informations sur une carte qui expliquent les symboles utilisés pour les objets et variables représentés sur la carte. La légende inclut l'explication des symboles requise pour interpréter la carte, par exemple les nuances de couleurs et intervalles de valeurs correspondants d'une carte de densité démographique.

Levé topographique aérien — Levé cartographique réalisé au moyen d'une photographie aérienne ou d'une autre technique de télédétection.

Ligne — Objet unidimensionnel. Type de donnée géographique constituée d'une série de coordonnées x, y, où la première et la dernière coordonnées sont appelées nœuds et les coordonnées intermédiaires sont des sommets. Qualifiée aussi parfois d'arc ou de chaîne. La partie d'une ligne comprise entre deux intersections avec d'autres lignes est appelée segment linéaire ou subdivision d'arc.

Limite — Ligne qui définit l'étendue d'une unité surfacique ou les lieux où deux surfaces se rencontrent. Une limite est représentée dans un SIG sous la forme d'un objet linéaire pouvant définir le côté d'un polygone. La limite peut ou non être visible sur le sol, c'est-à-dire qu'elle peut suivre des objets du monde réel tels que des routes ou des cours d'eau, ou elle peut être définie uniquement par des coordonnées géographiques.

Longitude — La « coordonnée x » dans un système de coordonnées polaires sur une sphère. Mesurée en tant que distance angulaire en degrés à l'est ou à l'ouest du méridien de Greenwich.

Maillage — Modèle de données géographiques représentant des informations sous la forme d'un tableau de cellules carrées uniformes. Chaque maille de la grille a une valeur numérique qui se réfère à la valeur réelle d'un phénomène géographique en ce point (par exemple densité démographique ou température) ou qui indique une classe ou une catégorie (par exemple l'identificateur de district de recensement ou le type de sol). Voir aussi **Matrice**.

Matrice — Modèle de données géographiques représentant des informations sous la forme d'un tableau régulier de lignes et de colonnes, semblable à une grille ou une image. Les cellules d'une matrice sont généralement, mais pas toujours, carrées. Les objets surfaciques ou linéaires sont représentés par des groupes de cellules matricielles adjacentes de même valeur.

Méridien — Ligne de référence définie par la longitude correspondante. Par exemple, le méridien de Greenwich.

Méridien central — Longitude qui définit l'origine de la coordonnée x d'une projection cartographique.

Méridien de Greenwich — Longitude de référence, c'est-à-dire 0° Est ou Ouest. Il passe par la ville britannique de Greenwich, dans la banlieue de Londres.

Métadonnées — Données relatives à des données. Ensemble d'informations décrivant le contenu, la qualité, les conditions, le format, la lignée et toute autre caractéristique pertinente d'un ensemble de données.

Modèle — En cartographie, dessin normalisé d'éléments cartographiques périphériques (bordures, limites, flèches indiquant le nord) qui peuvent être utilisés pour une série de cartes normalisées. Dans les systèmes de gestion de base de données, tableau vide créé à des fins multiples, pour lequel seuls les champs ou rubriques ont été définis.

Modèle altimétrique numérique (MAN) — Représentation numérique d'informations relatives à l'altitude d'une partie de la surface terrestre. Un MAN est généralement un ensemble de données ligne par ligne dans lequel les valeurs d'altitude sont enregistrées pour des cellules dans une grille à mailles fines, mais on peut aussi recourir aux formats vectoriels pour enregistrer l'altitude. Un MAN est quelquefois qualifié de modèle numérique de terrain (MNT).

Modèle de données — Concept théorique propre à l'utilisateur d'un ensemble de données décrivant les entités de la base de données et leurs relations réciproques.

Modèle entités-relations — Modèle de données qui définit les entités et leurs relations réciproques, par exemple les relations entre les districts de recensement et les régions de supervision.

Modèle numérique de terrain — Voir **Modèle altimétrique numérique (MAN)**.

Modélisation surfacique — Terme utilisé pour décrire des données matricielles ou image de SIG qui décrit un phénomène continu, variant régulièrement comme une altitude ou une température.. Même la densité démographique est parfois représentée sous la forme d'une surface maillée.

Nadir — En photographie aérienne et en télédétection, point de la surface terrestre situé directement au-dessous du capteur ou de l'appareil de prise de vues.

Nœud — Point initial ou point final d'un objet linéaire ou point de rencontre de deux lignes ou davantage.

Normalisation — Procédure conceptuelle d'élaboration d'une base de données qui élimine les redondances dans une base de données complexe en définissant des rapports de dépendance et des relations entre entités de la base de données. La normalisation réduit les besoins en capacité de stockage et évite des incohérences à l'intérieur de la base de données.

Normalisation des données — Élaboration d'un accord sur des définitions communes de données, formats, représentations et structures de toutes les couches et éléments de données.

Normes — En matière de calcul, ensemble de règles et de spécifications adopté par une autorité qui définit par exemple les besoins en matière de précision, les formats d'échange de données, les matériels ou les logiciels.

Numérisation — Processus consistant à transformer en coordonnées numériques des informations relatives à des objets géographiques figurant sur des cartes papier. La numérisation se réfère couramment au processus manuel consistant à tracer des lignes sur une carte papier fixée sur une table à numériser à l'aide d'un curseur de type souris qui saisit les coordonnées et les enregistre dans une base de données de SIG.

Numérisation frontale — Technique de numérisation n'exigeant pas l'utilisation d'une table à numériser. Les éléments sont tracés sur l'écran avec la souris, soit à partir d'une image scannée affichée à l'arrière-plan, soit en suivant des objets tracés sur une feuille transparente (par exemple du mylar) fixée sur l'écran de l'ordinateur.

Objet — Objet géographique représenté sur une carte ou enregistré dans une base de données de SIG. Les objets peuvent être des objets naturels ou des objets du monde réel, créés ou aménagés par l'homme (cours d'eau ou localité) ou des objets conceptuels ou déterminés (par exemple des limites administratives).

Objet géographique — Élément ou phénomène géographique défini par l'utilisateur et qui peut être représenté dans une base de données géographique. On peut citer comme exemple les rues, les parcelles de terrain, les puits ou les lacs.

Objets géographiques discontinus — Entités individuelles faciles à distinguer, comme des maisons ou des routes — contrairement à des phénomènes géographiques continus.

Octet — Groupe de huit chiffres binaires ou bits pouvant être traités comme une unité par les programmes informatiques. Un kilooctet comporte environ mille octets, un mégaoctet un million d'octets et un gigaoctet un milliard d'octets.

Orthophotographie — Voir **Orthophotographie numérique**.

Orthophotographie numérique — Image numérique ou photographie aérienne, généralement de très haute résolution, qui a été géométriquement corrigée. Une orthophotographie numérique, appelée également ortho-image, associe le détail d'une photographie aérienne à l'exactitude d'une carte topographique.

Parallèle de référence — Latitude qui définit l'origine de la coordonnée y d'une projection cartographique.

Parcelle — Unité cadastrale ou propriété foncière isolée.

Phénomènes géographiques continus — Variables géographiques qui varient sans ruptures ou interruptions clairement apparentes, par exemple, la température ou la pression atmosphérique — par opposition aux phénomènes géographiques discontinus.

Photogrammétrie — Art et science permettant d'obtenir à partir de photographies des mesures et d'autres informations. Dans le contexte cartographique, procédures appliquées pour collecter des informations sur des objets du monde réel à partir de photographies aériennes ou d'images satellitaires.

Photographie aérienne — Techniques de prise de vues photographiques à partir d'une plate-forme aérienne, en général un avion volant à basse altitude. Appelée aussi parfois photographie à la verticale ou orthophotographie. Les photographies aériennes sont utilisées par la cartographie photogrammétrique qui permet d'obtenir un degré élevé d'exactitude.

Pixel — De « picture element » (élément d'image). Semblable à une cellule dans une image, un réseau ou une carte matricielle.

Point — Objet de dimensions zéro. Coordonnée x, y utilisée dans une base de données géographique numérique afin de représenter des objets trop petits pour pouvoir être figurés par des lignes ou des polygones. Par exemple, les ménages, les puits ou les bâtiments sont souvent représentés par des points.

Point de contrôle — Point sur une carte ou une photographie aérienne ou dans une base de données numériques pour lequel les coordonnées x, y et éventuellement l'altitude sont connues. Utilisé pour l'enregistrement géodésique d'objets de la carte.

Polygone — Objet bidimensionnel. Objet surfacique représenté dans un SIG vectoriel sous la forme d'une série séquentielle de coordonnées x/y. Ces dernières définissent les lignes entourant la surface; autrement dit la première et la dernière coordonnée du polygone sont identiques.

PostScript — Langage de description de page, souple, à résolution élevée, utilisé le plus souvent pour envoyer aux imprimantes des informations graphiques telles que des cartes produits à partir de SIG. Le format PostScript encapsulé (EPS) comprend une représentation réduite du graphisme en mode point à des fins de visualisation préalable.

Précision — Capacité de distinguer de petites différences dans une mesure. Dans un SIG, la précision des coordonnées est déterminée par le type de donnée utilisé pour enregistrer les coordonnées x et y (en général double précision, soit 16 octets pour chaque nombre).

Précision d'alignement — Terme utilisé pour indiquer dans quelle mesure les emplacements sur une carte ou dans une base de données de SIG sont correctement enregistrés par rapport à leur localisation exacte sur la surface terrestre. L'exactitude logique ne concerne par contre que la représentation correcte des relations entre objets géographiques.

Projection cartographique — Procédé mathématique permettant de convertir des points situés à la surface de la Terre dans un système de coordonnées planes. En fonction de la formule mathématique appliquée, les projections cartographiques ont différentes propriétés. Certaines conservent la forme des régions sur le globe, d'autres la surface, les angles ou les distances relatives.

Projection conforme — Projection cartographique dans laquelle tous les angles sont conservés correctement en tous points.

Projection cylindrique transverse [Universal Transverse Mercator (UTM)] — Projection cartographique cylindrique souvent utilisée pour les cartes à grande échelle (c'est-à-dire locales).

Projection équidistante — Projection cartographique qui conserve l'échelle le long d'une ou de plusieurs lignes, ou entre un ou deux points et tous les autres points de la carte.

Projection équivalente — Projection cartographique dans laquelle toutes les régions sont représentées dans une proportion correcte par rapport à leur surface dans le monde réel.

Protocole — Ensemble de conventions qui déterminent le traitement, l'échange et le formatage de données dans un système de communications électroniques. Semblable à une norme de données, mais appliqué aux procédures.

Protocole de transfert de fichiers [File transfer protocol (FTP)] — Ensemble standard de conventions pour l'échange de fichiers informatiques dans des systèmes numériques de télécommunications comme Internet.

Quadrangle — surface rectangulaire limitée par des paires de méridiens et de parallèles.

Quantile — Méthode de classification statistique ou cartographique qui affecte un nombre égal d'objets à un nombre déterminé de catégories. Les systèmes de quatre catégories sont appelés quartiles, de cinq catégories quintiles et de dix catégories percentiles. Par exemple, le premier des quatre quartiles d'une répartition de données contient 25 % des observations comportant les valeurs les plus basses.

Rangée — Dans un SIG, groupe de cellules ou de pixels alignés horizontalement dans un réseau ou une base de données matricielle de SIG. Dans les systèmes de gestion de base de données, enregistrement ou instance dans un tableau d'attributs.

Rayon — Distance du centre d'un cercle à sa circonférence.

Réalité de terrain — Informations collectées lors d'un levé de terrain pour vérifier ou calibrer des informations extraites de données de télédétection.

Rectification — Processus par lequel les coordonnées d'une image ou matrice sont converties en coordonnées réelles, nécessitant généralement une rotation et une mise à l'échelle de mailles et exigeant donc un nouvel échantillonnage ou une interpolation des valeurs de mailles. Similaire à la transformation de données vectorielles.

Repérage — Processus consistant à apparier des objets sur deux cartes ou deux couches de données de SIG de telle sorte que les objets correspondants coïncident. Le repérage se réfère à une série de points de contrôle terrestres et est lié à la transformation et à l'étirement élastique.

Répertoire toponymique — Liste de noms de lieux avec leur localisation géographique (en général latitude et la longitude).

Réseau grande distance [Wide Area Network (WAN)] — Réseau informatique reliant des ordinateurs à grande distance au moyen de liaisons de télécommunications à grande vitesse ou de satellites.

Réseau local [Local Area Network (LAN)] — Réseau informatique reliant des ordinateurs relativement peu éloignés les uns des autres, par exemple dans le même immeuble de bureaux.

Résolution — Mesure du plus petit détail pouvant être distingué sur une carte ou dans une base de données numérique. La résolution détermine l'exactitude avec laquelle l'emplacement et la forme d'un objet cartographique peuvent être représentés à l'échelle considérée de la carte. En ce qui concerne les données matricielles de SIG et les données image, la résolution sert parfois à indiquer la dimension des cellules ou des pixels.

Saisie des données — Conversion de données concernant les coordonnées géographiques à partir de sources sur papier ou à l'aide de mesure sur le terrain dans un format lisible par l'ordinateur. La saisie des données nécessite généralement la numérisation ou le scannage de cartes papier ou de photographies aériennes.

Satellite géostationnaire — Satellite terrestre qui reste dans une position fixe au-dessus d'un point de la surface terrestre. Désigné également par son orbite géosynchrone.

Scannage — Technique de saisie de données utilisée pour collecter des informations sur des documents imprimés (par exemple sur papier ou mylar) et pour les convertir en image numérique au moyen d'un dispositif optique photosensible. Pour les données cartographiques, le scannage peut remplacer l'entrée de données par numérisation. Après avoir scanné une carte, les données image sont en général converties en format vectoriel au moyen d'un logiciel de conversion mode maillé/mode vectoriel ou par traçage sur écran d'objets linéaires et ponctuels.

Seconde d'arc — Une seconde de latitude ou de longitude, soit 1/3600° de degré.

Segment de contrôle — Réseau mondial de station de surveillance et de contrôle de GPS qui assure l'exactitude des signaux de satellites.

Segment spatial — Partie du système GPS située dans l'espace, c'est-à-dire les 24 satellites GPS.

Segment utilisateur — Partie du système GPS qui inclut tous les types de récepteurs de signaux GPS.

Séparation des couleurs — Procédé consistant à diviser un document graphique en pages ou dossiers séparés pour chacune des quatre couleurs (cyan, magenta, jaune et noir). La séparation des couleurs est à la base de la plupart des procédés professionnels d'imprimerie.

Serveur — Ordinateur mis en place pour fournir certains services à d'autres ordinateurs (clients), par exemple un serveur Web est un service central d'archivage de données, de logiciels ou de contenu pour le World Wide Web.

Sommet — Élément d'une série de coordonnées x, y définissant une ligne. Le premier et le dernier sommet de la ligne sont généralement appelés nœuds.

Spatial Data Transfer Standard (SDTS) — Norme de données et métadonnées pour l'échange d'ensembles de données de SIG entre producteurs et utilisateurs de données et entre systèmes logiciels et formats de fichiers. De nombreuses normes nationales et internationales ont été réalisées ou suggérées.

Sphère — Corps sphérique semblable à une balle. La Terre est de façon très approchée une sphère, mais en réalité, elle est plus exactement représentée par un sphéroïde. Voir **Ellipsoïde**.

Station de base — Récepteur GPS dont l'emplacement a été déterminé avec exactitude et précision, qui diffuse et/ou reçoit des informations différentielles de correction pour les récepteurs GPS mobiles. Voir aussi **GPS différentiel**.

Structure de données — Mise en application d'un modèle de données constitué de structures de fichiers utilisées pour représenter divers objets.

Superposition — Combinaison de deux couches de données qui se trouvent dans le même système de référence géographique. La superposition peut être opérée à des fins de présentation cartographique ou les deux couches peuvent être physiquement combinées pour créer un nouvel ensemble de données de SIG (par exemple superposition de polygones, inscription d'un point dans un polygone, d'une ligne dans un polygone).

Superposition de polygones — Opération de SIG par laquelle deux couches de données de polygones sont combinées pour créer une nouvelle couche de données. La couche obtenue est constituée des zones d'intersection des deux ensembles de polygones introduits. Le tableau des attributs de la nouvelle couche de données contient les attributs des deux ensembles de données entrés. La superposition de polygones est l'une des opérations fondamentales d'un SIG et elle est souvent utilisée pour intégrer des informations provenant de sources hétérogènes telles que des données démographiques et environnementales.

Surface — Étendue limitée, bidimensionnelle de la surface terrestre représentée par un polygone dans un SIG.

Symboles — En cartographie, éléments de conception graphique utilisés pour représenter des objets cartographiques. Les types de symboles sont des points, des lignes et des polygones d'une certaine forme. La symbolisation suppose que l'on choisisse des variables graphiques telles que la forme, la dimension, la couleur, la structure et la texture.

Symboles gradués — En cartographie thématique, utilisation de symboles (c'est-à-dire des cercles et des carrés) pour représenter l'importance d'une variable en un point ou dans une unité déclarante. La dimension du symbole est proportionnelle à la valeur de la variable.

Système de coordonnées — Système de référence utilisé pour spécifier des emplacements sur une carte ou dans une base de données de SIG. Un système de coordonnées cartographiques est défini par la projection de la carte, un ellipsoïde de référence, un méridien central, un ou plusieurs parallèles standards et des déplacements éventuels des coordonnées x et y .

Système de coordonnées cartésiennes — Système de lignes qui se coupent en formant des angles perpendiculaires dans un espace à deux dimensions. Ce système fournit un cadre pour des localisations de référence précises appelées coordonnées x/y .

Système de coordonnées planes — Système utilisé pour localiser un point dans lequel deux groupes de lignes droits se coupent à angle droit et ont comme point d'origine une intersection perpendiculaire donnée. Voir **Système de coordonnées cartésiennes**.

Système de couleurs — Moyen permettant de représenter les couleurs numériquement dans un ordinateur. Par exemple, dans le système de couleurs RVB, les couleurs sont représentées sous forme de niveaux numériques de rouge, vert et bleu. Le rouge pur, par exemple, est défini par les chiffres 255, 0, 0. Il existe d'autres systèmes de couleurs tels que le système TSL (teinte, saturation, luminosité) et le système CMJ (cyan, magenta, jaune).

Système de gestion de base de données (SGBD) — Logiciel conçu pour gérer et manipuler les données de tableaux. Un SGBD est utilisé pour saisir, enregistrer, manipuler, chercher et interroger des données. La plupart des SIG utilisent un SGBD relationnel pour gérer des données d'attributs.

Système de gestion de base de données relationnelle (SGBDR) — Système de gestion de base de données qui permet de réunir temporairement ou en permanence des tableaux de données appartenant à une zone commune (clé primaire et clé étrangère). Chaque ligne, enregistrement ou instance dans une base de données présente un ensemble déterminé d'attributs ou de champs. Chaque tableau a une clé primaire qui identifie chaque enregistrement par une référence unique. Le tableau peut aussi contenir une clé étrangère, identique à une clé primaire dans un tableau extérieur. Un lien relationnel est établi en appariant les valeurs de la clé étrangère aux valeurs correspondantes de la clé primaire du tableau extérieur.

Système de référence géodésique — En cartographie, ensemble de paramètres définissant un système de coordonnées. Plus précisément, un système de référence géodésique sert de référence ou de base à des mesures ou des calculs. Par exemple, un système de référence géodésique constitue le cadre de référence pour les activités cartographiques du pays.

Système d'information géographique (SIG) — Ensemble d'ordinateurs, logiciels, données, géographiques et personnel réunis en vue de saisir, enregistrer, rechercher, mettre à jour, manipuler, analyser et afficher des informations géographiquement référencées.

Système d'information sur les terres (LIS) — Terme parfois utilisé pour une application de SIG qui contient des informations sur une région particulière, y compris des renseignements cadastraux, des informations sur l'utilisation des sols, la couverture des terres, etc.

Système mondial de localisation (GPS) — Système de 24 satellites sur orbite terrestre qui émettent des signaux utilisables pour déterminer la position géographique exacte d'un point à la surface de la terre. Le GPS est largement utilisé pour les levés sur le terrain, l'arpentage et la navigation. Le système GPS est exploité par le

Département de la défense des États-Unis. Voir aussi **GPS différentiel (GPSD)**, **Beidou**, **Galileo** et **GLONASS**.

Table à numériser — Périphérique informatique utilisé pour saisir des données de coordonnées extraites de cartes sur papier ou de matériaux cartographiques semblables. Appelé également numériseur.

Tableau — Dans les systèmes de gestion de base de données, ensemble d'éléments de données arrangées en lignes (enregistrements et instances) et en colonnes (champs ou rubriques). Le nombre de colonnes est généralement fixé par la définition de la structure du tableau, tandis que le nombre de rangées demeure variable.

Tag Image File Format (TIFF) [Format de fichier image étiqueté] — Format standard de fichier image ou matriciel qui peut enregistrer des images en noir et blanc, en gris ou en couleurs sous forme comprimée ou non. Les scanners et autres dispositifs utilisés pour créer des données image comportent souvent une sortie en format TIFF. Dans les SIG, Geo-TIFF est un fichier image standard en format TIFF qui décrit une image de télédétection, une orthophotographie numérique ou un ensemble de données matricielles de SIG. Il comporte un fichier associé avec une extension .tfw qui contient des informations concernant les données de référence géographique de l'image, la dimension des cellules en unités du monde réel et d'autres informations pertinentes.

Télédétection — Processus consistant à acquérir des informations sur un objet à distance, c'est-à-dire sans contact physique. La télédétection se réfère généralement à l'acquisition d'images à partir de capteurs satellitaires ou de photographies aériennes.

Thème — Dans un SIG, ensemble d'objets géographiques appartenant généralement au même groupe de sujets (par exemple routes ou localités) et qui sont stockés dans la même base de données de SIG.

Topologically Integrated Geographic Encoding and Referencing (TIGER) [Codage et référencement géographiques topologiquement intégrés] — Format de données élaboré par le Bureau of the Census des États-Unis pour prendre en charge les programmes de recensement et les dénombrements. Les fichiers TIGER sont des ensembles de données de SIG dans un format interne qui contient des intervalles d'adresses le long du réseau de voirie et des secteurs de recensement ainsi que des limites d'îlots de recensement. Le système TIGER figure parmi les premières initiatives prises en vue de créer une base de données numérique complète de SIG pour l'ensemble d'un pays.

Topologie — Dans un SIG, terme désignant les relations spatiales entre objets géographiques (par exemple points, lignes, nœuds et polygones). Une base de données topologiquement structurée enregistre non seulement des objets individuels mais aussi la façon dont ces objets sont en rapport avec d'autres appartenant à la même classe ou à une classe différente d'objets. Par exemple, le système enregistre non seulement l'ensemble de lignes représentant un réseau routier, mais aussi les nœuds définissant les intersections des routes, ce qui lui permet de déterminer des itinéraires empruntant différents segments de routes. D'autre part, au lieu de stocker des polygones sous la forme de boucles fermées enregistrant deux fois les limites entre polygones adjacents, un SIG structuré topologiquement n'enregistre chaque ligne qu'une seule fois, avec des informations indiquant quel polygone se trouve à gauche ou à droite de la ligne. Cela évite des redondances et facilite la mise en œuvre de nombreuses fonctions de SIG et d'analyse spatiale.

Traceur (ou table traçante) — Périphérique informatique destiné à reproduire des fichiers graphiques, semblable à une imprimante, mais utilisé en général pour obtenir des produits en grand format.

Trajets multiples — Erreur dans les lectures de GPS due à la réflexion et à la dispersion des signaux de GPS sur les structures voisines telles que des maisons ou des arbres. L'erreur due aux trajets multiples constitue un problème surtout pour les levés topographiques de grande précision.

Transformation — Conversion de données spatiales numériques d'un système de coordonnées à un autre par translation, rotation et mise à l'échelle. La transformation est utilisée pour convertir des données cartographiques numériques obtenues par numérisation d'unités de numériseur (par exemple centimètres ou pouces) en unités du monde réel correspondant à la projection cartographique et au système de coordonnées de la carte de référence (par exemple mètres ou pieds). Voir aussi **Géoréférencement**.

Transmission Control Protocol (TCP) [Protocole de contrôle de transmission] — un des protocoles de base d'Internet.

Tuile — Dans un SIG, terme parfois utilisé pour désigner des feuilles adjacentes de carte numérique enregistrées dans des fichiers séparés. Les tuiles peuvent être de forme régulière (carrées ou rectangulaires) ou elles peuvent suivre des tracés irréguliers tels que les limites de districts ou de provinces. L'enregistrement de toutes les tuiles dans le même système de référence géographique permet de réunir des tuiles adjacentes à titre temporaire ou permanent.

Type de données — Caractéristique de champ des colonnes dans un tableau d'attributs. Par exemple : caractère, virgule flottante et nombre entier.

Unité administrative — Zone géographique qui sert à des fins administratives et gouvernementales. Elle est généralement définie et créée par une disposition légale.

Unité cartographique minimale — En général, dimension du plus petit objet qui sera inclus sur la carte. À une échelle cartographique donnée, c'est aussi la dimension à partir de laquelle un objet polygonal compact peu étendu est représenté par un point ou un objet polygonal long et étroit est représenté par une ligne. Par exemple, une ville est représentée par un polygone si son diamètre dépasse 3 mm sur une page et par un point si elle est plus petite.

Unité gouvernementale — Voir **Unité administrative**.

Unité surfacique — Surface naturelle ou artificielle souvent utilisée pour compiler et présenter des données agrégées. On peut citer à titre d'exemple des zones de couverture terrestre ou des **districts de recensement**.

Unités cartographiques — Unités de mesure dans lesquelles les coordonnées d'une base de données de SIG sont enregistrées; à savoir : centimètres, mètres ou degrés, minutes et secondes.

Vector product format (VPF) [Format de produit vectoriel] — Format vectoriel de SIG élaboré par la National Map and Imagery Agency des États-Unis (précédemment Defense Mapping Agency) et destiné à devenir un format d'échange de données vectorielles universellement accepté.

World Wide Web (WWW) [Web ou Toile] — Il a été mis au point initialement par l'Organisation européenne pour la recherche nucléaire (CERN), en Suisse, en tant que système permettant de distribuer des documents électroniques constitués de nombreux fichiers différents de divers formats, n'importe où dans le monde, ou se rapportant à ces fichiers. Les documents sont créés dans un lan-

gage normalisé hypertexte avec liens (Hypertext markup language ou HTML) qui peut être interprété par des navigateurs Web sur l'ordinateur de l'utilisateur. Les emplacements des documents HTML sont définis par des liens ou des adresses appelées localisateurs universels de ressources (universal resource locators ou URL). Le Web s'est développé rapidement et il devient de plus en plus un moyen important permettant de diffuser des documents et des données. Des logiciels spécialisés de SIG permettent à un organisme de présenter des cartes sur le Web. Par exemple, un utilisateur éloigné peut concevoir et afficher une carte thématique en utilisant des bases de données de SIG situées sur le serveur Web de l'organisme assurant la distribution.

D'autres glossaires et dictionnaires sont mentionnés chez Padmanabhan et autres (1992), ASCE (1994), McDonnell et Kemp (1995) et Dent (1998). Des ressources peuvent également être consultées en ligne :

Canada Centre for Remote Sensing	http://www.ccrs.nrcan.gc.ca/
Geographer's Craft Project (University of Texas)	http://www.colourado.edu/geography/gcraft/contents.html
GPS World magazine	http://www.gpsworld.com/gpsworld/static/staticHtml.jsp?id=8000&searchString=glossary
Perry-Castañeda Library, University of Texas	www.lib.utexas.edu/Libs/PCL/Map_collection/glossary.html
United States Census Bureau	www.census.gov/dmd/www/glossary.html
United States Geological Survey	http://interactive2.usgs.gov/learningweb/explorer/geoglossary.htm

Annexe VII

Adresses utiles, URL

Logiciels de SIG

Autodesk Inc.	San Rafael, CA	AutoCAD	www.autodesk.com
Bentley Systems Inc.	Huntsville, AL	MicroStation	www.bentley.com
ESRI, Inc.	Redlands, CA	ArcGIS, ArcInfo, ArcView, ArcExplorer, Atlas GIS	www.esri.com
Intergraph	Huntsville, AL	GeoMedia	www.intergraph.com
MapInfo Corp.	Troy, NY	MapInfo GIS	
Microsoft Corp.	Redmond, WA	MapPoint	www.microsoft.com
Oracle Corp.	Redwood Shores, CA	Oracle Spatial	www.oracle.com
UNSD Software Project	New York, NY	PopMap	www.un.org/Depts/unsd/softproj/index.htm
Siemens	Munich, Allemagne	SICAD Spatial Desktop	www.siemens.com
Smallworld Systems Inc.	Englewood, CO		
PCI Geomatics Group	Richmond Hill, Ontario, Canada	SPANS and PAMAP	www.pci.on.ca
ThinkSpace Inc.	London, Ontario, Canada	MFWorks	www.thinkspace.com
Vision* Solutions	Ottawa, Ontario, Canada	Vision*	

Logiciels spécialisés

Blue Marble Geographics	Gardiner, ME	Coordinate management and GIS development tools	www.bluemarblegeo.com
Caliper Corp.	Newton MA	Maptitude, GIS+, TransCAD	www.caliper.com
Core Software Technology	Pasadena, CA	TerraSoar (bases de données géospatiales – diffusion), ImageNet (diffusion en ligne de données géospatiales)	www.coresw.com
Quantum GIS		Logiciel libre	http://qgis.org
Thuban		Logiciel libre	http://thuban.intevation.org
Open EV		Logiciel libre	http://openev.sourceforge.net

Systèmes de traitement d'images de télédétection

Leica GeoSystems/Erda	Atlanta, GA	ERDAS Imagine	www.erdas.com
Earth Resource Mapping	San Diego, CA	ER Mapper	www.ermapper.com

Clark Labs	Worcester, MA	Idrisi GIS	www.clarklabs.org
MicrolImages Inc.	Lincoln, NE	TNTmips	www.microimages.com
PCI Geomatics Group	Richmond Hill, Ontario, Canada	EASI/PACE, OrthoEngine	www.pci.on.ca
Research Systems Inc	Boulder, CO	ENVI logiciel de visualisation	www.rsinc.com

Images satellitaires à haute résolution et orthophotographie numérique

GeoEye	Thornton, CO	Satellites Carterra et Ikonos	www.spaceimaging.com
Digital Globe	Longmont, CO	Satellites QuickBird et EarlyBird	www.digitalglobe.com
Orbital Imaging Corp.	Dulles, VA	Satellites Orbimage	www.orbimage.com
EROS Data Center	Sioux Falls, SD		
Spot Image		Satellites Spot	www.spot.com
Maps Geosystems	Munich, Allemagne	Photographies aériennes (Afrique, Moyen-Orient)	www.maps-geosystems.com
EarthSat	Rockville, MD	Services de satellites et de cartographie	www.earthsat.com

Systèmes de positionnement universel

Magellan Corp.	Santa Clara, CA	www.magellangps.com
Ashtech	Santa Clara, CA	www.ashtech.com
NovAtel Inc.	Calgary, Alberta, Canada	www.novatel.ca
Sokkia Corp.	Overland Park, KA	www.sokkia.com
Trimble Navigation Ltd.	Sunnyvale, CA	www.trimble.com
Garmin		

Revues

GeoWorld, GeoAsia, GeoEurope, GeoInformation Africa, Mapping Awareness, Business Geographics	GeoWorld, Fort Collins, CO	www.geoplance.com
GPS World		www.gpsworld.com
International Journal of Geographical Information Science	Taylor & Francis, Londres, Royaume-Uni	
GeoInfosystems	Advanstar Pub., Eugene, OR	
Journal of the Urban and Regional Information Systems Association	URISA, Park Ridge, IL	http://www.urisa.org/
ISPRS Journal of Photogrammetry & Remote Sensing		www.itc.nl/isprsjournal

Divers

National Center for Geographic Information and Analysis	Santa Barbara, CA	GIS research center	www.ncgia.ucsb.edu
International Institute for Aerospace Survey and Earth Sciences (ITC)	Enschede, Pays-Bas	Cours de formation sur les SIG	http://www.itc.nl/
European Umbrella Organization for Geographic Information (EUROGI)	Pays-Bas		www.eurogi.org
U.S. Federal Geographic Data Committee	Reston, VA		www.fgdc.gov
Permanent Committee on GIS Infrastructure for Asia & the Pacific			www.permcom.apgis.gov.au/
Odyssey		Communications sur les SIG	
ESRI GIS jump station		Liens avec des applications de SIG dans le monde entier	
GeoWorld business links			

