

10 March 2025

English

**United Nations Group of Experts on
Geographical Names**

2025 session

New York, 28 April – 2 May 2025

Item 5(b) of the provisional agenda *

Technical expertise: Geographical names data management

**Standardization of *kanji* characters used in the geographical
names database**

Submitted by Japan**

* GEGN.2/2025/1

** The full report was prepared by Geospatial Information Authority of Japan.

1. Introduction

In Japan, geographical names are written by combining multiple characters such as *kanji*, which are adapted from the logographic Chinese characters for Japanese writing, and *hiragana*, which are phonetic characters. As a national mapping agency, the Geospatial Information Authority of Japan (GSI) has developed and manages a database of approximately 420,000 domestic geographical names that appear on the national base maps. The geographical names database holds the *kanji* characters of the geographical names as well as attribute information such as *hiragana* that indicates the pronunciation, and the coordinate value obtained at a representative location within the area indicated by the geographical name. It is utilized as the basic data for geographical names to be displayed on the maps produced by the GSI. In the past, some *kanji* characters used in the geographical names database could not be displayed correctly in some system environments, so they were replaced with *kanji* characters or *hiragana* that can be handled under a system environment such as a general computers and smartphones. Recently, the Japanese government has been preparing character information commonly handled by administrative agencies. In line with this initiative, the GSI has also been working on standardizing and unifying the *kanji* characters used in the geographical names database by introducing the character specifications recommended by the government. That is what we will be discussing in this report.

2. The current state of *kanji* characters in Japan

In Japan, 2,136 common *kanji* characters have been established as a guideline for *kanji* characters used in everyday life, but the total number of *kanji* characters used by administrative agencies in their duties exceeds 60,000, including those with slight differences in character shape (Figure 1). When handling characters on a system environment, it is necessary to assign a character code to each character for processing by the system environment, based on its shape to be recognized as a character by humans. In this paper, the set of characters to which each character code has been assigned is called a character set. Since there is no character set that can correspond to all *kanji* characters including those with slight differences in their shapes, each administrative agency uses a character set that adds its own created characters to a general character set. As a result, inconsistencies occur in data exchange between administrative agencies, unique character information cannot be obtained, and data is not displayed correctly on general computers and smartphones. For this reason, the government is promoting the identification of *kanji* characters that can be commonly used by administrative agencies and the assignment of character codes, as well as their use in the core business systems of local governments.

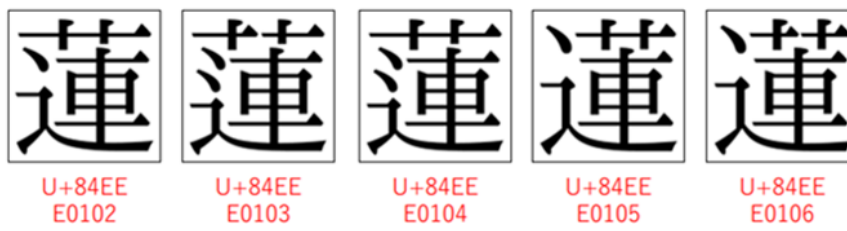


Figure 1: Example of *kanji* characters of only slightly different forms being used as different *kanjis*

3. Issues associated with *kanji* characters in geographical names database

The GSI continuously collects residential geographical names established by local governments through resolutions in their assemblies, and common geographical names (natural geographical names, etc.) that have been confirmed by local governments as being used uniformly in the local area and updates and manages the geographical names database. It is necessary to use *kanji* characters as

accurately as possible, but at the time the specifications for the geographical names database were established, the number of *kanji* characters that could be displayed correctly with the character code used in the system environment was limited. For this reason, the GSI limited the range of *kanji* characters to be used to match the character code used at the time, applied substitute characters for *kanji* characters outside the range and made public on the Internet a correspondence table of original *kanji* characters and substitute characters. However, there was a possibility that map users were unaware that the substitute characters were used and assumed that these substitute characters were the original *kanji* characters.

4. Initiatives and effects of standardizing *kanji* characters

In recent years, character codes that are widely used in system environments are capable of correctly displaying a wider range of *kanji* characters, and environmental constraints are decreasing. And *kanji* characters in geographical names are an important element in establishing an identity. But as mentioned above, creating and using original characters will result in a loss of data compatibility. Therefore, in light of the problems with displaying *kanji* characters in geographical names and the data compatibility, the GSI considered changing the specifications for the *kanji* characters used in the geographical names database and it was decided that the *kanji* characters used in the geographical names database would conform to the character specifications recommended by the Japanese government.

As a result, it has become possible for almost all *kanji* characters to be displayed correctly in general system environments. In addition, for *kanji* characters with slight differences in shape, the government has recommended a correspondence table, which eliminates the need for the GSI to develop its own correspondence table, thereby achieving consistency with the standardization and unification of *kanji* characters in Japan.

5. Concluding remarks

The standardization of *kanji* characters is being promoted by the Japanese government, and machine-readable data that ensures interoperability will enable correct information to be conveyed in electronic texts and speech recognition. This will lead to the avoidance of *kanji* characters that are not displayed properly depending on the system environment. The Geospatial Information Authority of Japan will continue to work on the efficient and proper development of the geographical names database, and on providing geographical name information that is useful and easy to understand for users.

Points for discussion

The Group of Experts is invited to:

- (1) Express its views on the report and sharing whether other countries have the same experience(s) towards the standardization, including the identification of various characters and the assignment of the character codes.