The Expansion of Historical Municipal Boundaries Dataset and Its use in Historical Big Data Research

*Asanobu Kitamoto[1], Ken T. Murata[2]

1. National Institute of Informatics, 2. National Institute of Information and Communications Technology

In most geographic information, municipality names are central to place names. As municipality names appear in various administrative documents and statistical information, mapping these into geographical space is essential for geographical analysis. The historical change of municipalities is not developed as open data, and it has been challenging to analyze data by historical municipality names. Hence this study aims to develop unified open data on municipalities and utilize it for historical big data research (http://codh.rois.ac.jp/historical-big-data/) from the Edo period to the present.

The Historical Municipality Boundaries Dataset (https://geoshape.ex.nii.ac.jp/city/) is an open dataset of changes in the municipality from 1920 to 2022, utilizing the municipality data from the Digital National Land Information. This dataset defines criteria for judging the continuity and identity of place names. When the name does not change, even if the boundary of the place name changes, it is considered continuous and identical and assigns an identifier to the municipality. In addition, the municipality can be handled as both a point and a polygon by devising a unique algorithm for assigning representative points.

A significant problem with this dataset is the lack of continuity. For example, the oldest data in the Digital National Land Information is from 1920, but the following data is from 1950, leaving a gap of 30 years. Moreover, the gap continues for up to five years after that, so municipalities created and ceased to exist within the missing period are not included in the dataset. Therefore, we decided to integrate the Administrative Boundary Transition Database (Map Data) of the Yuji Murayama Laboratory in the Department of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, which includes the municipality data every year in succession since 1889, when the municipal system came into effect. As a result, we could assign identifiers to 4159 municipalities derived from the Japanese Local Government Code, 12260 municipalities derived from Digital National Land Information, and 399 municipalities derived from the Administrative Boundary Transition Database.

Furthermore, in cooperation with Heibonsha and the Encyclopaedia Research Center, we are now attempting to extend the identifiers in Heibonsha's Nihon Rekishi Chimei Taikei. Currently, verification is ongoing in Gunma Prefecture, and we expect to add new identifiers to the open data for towns and villages dating back to the Edo period, such as feudal villages. The release of new open data will facilitate research on extracting and mapping place names from long-term texts from the Edo period to the present, leading to a breakthrough in historical big data research.

We design the Historical Municipal Boundaries Dataset to link with diverse datasets and services. First, by calculating overlaps with the 'Census Boundary Dataset' (https://geoshape.ex.nii.ac.jp/ka/), it is possible to look up past municipal names at the town and street level. In addition, by using GeoLOD (https://geolod.ex.nii.ac.jp/), a service for assigning and sharing place name identifiers, it is possible to search for information on past municipalities via an API. It also shares its place name dictionary with GeoNLP (https://geonlp.ex.nii.ac.jp/), a service that automatically extracts place names from the text, paving the way for the extraction of place names from large amounts of text. In addition, as a service utilizing the Historical Municipal Boundaries Dataset, an application has been developed to visualize the

number of articles per region corresponding to a search keyword by linking place names assigned to past newspaper articles with the Historical Municipal Boundaries Dataset. In this way, we believe the Historical Municipal Boundaries Dataset can contribute as the foundational dataset for various historical applications.

Keywords: Historical Big Data, Municipality, Boundary data, Place name identifier, Geographic
                  information, Visualization