

Toward Sharing Historical Place Names Using Toponym Information Platform GeoLOD

*Asanobu Kitamoto¹

1. ROIS-DS Center for Open Data in the Humanities

GeoLOD is a toponym information platform that assigns identifiers related to place names and has functions to manage and share various attributes related to place names. A place name here refers to a unique name with a geographical concept as an attribute. In addition to addresses, which form a hierarchical structure of place names, other typical place names include natural entities such as mountains and rivers and entities related to facilities and points of interest (POIs). GeoLOD is a system mainly designed for data integration using place names and targets not only the use of public open data on place names but also the collaborative editing of place name dictionaries by communities.

GeoLOD consists of a place name management system and a place name publication system. First, the place name management system consists of a place name dictionary and place names. The place name dictionary has identifiers and attributes, and it is assumed to manage a set of place names that belong to a specific theme. On the other hand, individual place names also have identifiers and attributes, and latitude and longitude attributes play a vital role in linking place names to the real world (but is not a required field). There are two types of place name dictionaries: upload and cloud dictionaries. First, an upload dictionary is a dictionary of place names generated from some information source and managed in the CSV format of the GeoLOD schema. On the other hand, a cloud dictionary is a dictionary of place names built on the GeoLOD system and can be downloaded in the CSV format of the GeoLOD schema. Only the creator of the cloud dictionary can edit the place names, but it also offers a collaborative editing environment by inviting another user.

Next, the place name publication system searches the place names stored in the GeoLOD and visualizes them on the map. GeoLOD also provides an API that allows users to narrow down the search by latitude and longitude and by partial matching of place names, thus enabling cross-application data integration using GeoLOD place name identifiers. This approach has been tried on the "Rekiske" system to create a place name card using the GeoLOD identifiers.

GeoLOD collaborates with other systems to provide an integrated processing environment for place names. First of all, while GeoLOD represents the location of a place name as a representative point, Geoshape represents the range of a place name as an area (polygon) so that a place name can be treated as both a point and an area by linking GeoLOD and Geoshape. In addition, by registering the dictionary of place names, derived from the "Historical Administrative Area Dataset Beta" published on Geoshape to GeoLOD, we can use the information on municipalities in the past. Furthermore, using the place name dictionary constructed by GeoLOD with the text geotagging system GeoNLP, place names can be automatically extracted from text, disambiguated, and visualized on a map. In this way, GeoLOD is at the core of a place name processing platform by accumulating information on place names and utilizing it for data integration.

The integration of place names is also an essential issue in the historical big data project. If we can integrate a place name that appears in one source with a place name that appears in another source, and decide that they are the same, we will be able to link and interpret multiple events that occurred in one

place. A loosely coupled design such as GeoLOD, where immutable identifiers are used for data integration and variable data is centrally managed in GeoLOD, has the advantage of being resistant to various changes. In research fields that deal with growing and uncertain data, such as historical big data, an approach that gradually increases the reliability of the data is essential, and we believe that the GeoLOD approach is suitable for this purpose.

In order to share geographic information on historical big data across applications, we are designing an integrated management system for historical records called "Rekiroku". The system combines document space identifiers with entity space identifiers and enables quantitative interpretation of the contents of historical records, so that the accumulated data can be used for various scientific applications. In order to promote this concept, it is crucial to collect more historical place names and grow them as a reliable source of information in the future. We want to promote the collaborative accumulation of historical place names combining upload and cloud dictionaries of GeoLOD.

Keywords: Historical Big Data, Place Name, Toponym Information Platform, GeoLOD, Historical Place Name, Data Integration