

# The Construction of Geometric Data Sharing Site 'Geoshape' Including Historical Administrative Region Dataset $\beta$

\*Asanobu Kitamoto<sup>1,2</sup>, Ken T. Murata<sup>3</sup>

1. ROIS-DS Center for Open Data in the Humanities, 2. National Institute of Informatics, 3. National Institute of Information and Communications Technology

## 1. Introduction

There are two approaches for integrating geographic information. The first approach is to use a coordinate system such as latitude and longitude and mathematically define an addressing system. The second approach is to use the database of entities and enumerate all addresses. The entity is a unit expressed as the combination of a unique ID and multiple attribute information. Geometric data such as a point or a polygon is included in the attribute information, but the unit of integration is the unique ID. We aim at building entity-based geographic information systems, starting from GeoNLP (<https://geonlp.ex.nii.ac.jp/>) to extract entities, GeoLOD (<http://geolod.ex.nii.ac.jp/>) to create linked data for entities, and Geoshape (<http://geoshape.ex.nii.ac.jp/>) to share entity-based geometric data.

## 2. Changes in the Format of Geometry Data

For the sharing of geometric data, the Shapefile format and the GML (Geography Markup Language) format have been used. These formats, however, are not best-fit to the trend of Web technology, and we employed alternative formats, namely GeoJSON and TopoJSON, variants of the JSON format. GeoJSON can describe the shapes and attributes of entities such as points, lines, and polygons. TopoJSON simplifies the representation of coordinates in comparison to GeoJSON, and is also adapted to color-coded maps (choropleth maps).

In addition, vector tile formats for efficient data access are gaining popularity in recent years. A vector tile format incorporates the idea of tiling, which is already standard in raster maps, and accesses only tiles that are required for the real-time rendering of maps on the browser. It also has advantages of quicker rendering and on-the-fly configuration of the map styling. We selected the Mapbox Vector Tile format, which uses Protocol Buffers for the binary encoding of the vector data.

## 3. The Construction of Historical Administrative Region Dataset $\beta$

The most demanding geometrical data is the administrative region dataset of municipalities. Municipalities are not only used as a unit of geographic information, but also linked to the daily life of residents and the history of memories. Although the number of municipalities has decreased with each major merger from the Meiji to the Heisei era, memories of the former municipalities still remain in multiple layers. To integrate information from the present to the past, we therefore need to construct a dataset that can keep track with historical changes in administrative regions.

In January 2017, we released a website “Historical Administrative Region Dataset  $\beta$ ,” which structures and visualizes the transition of administrative regions with reference to the “Kokudo Suuchi Joho from Ministry of Land, Infrastructure, Transport and Tourism.” This dataset not only provides the administrative regions as polygon data at 27 time points from 1920 to 2019, but also summarizes the proximity and overlap of administrative regions based on geometric computation. In addition, by referring to the municipal offices and public facility dataset, a public facility within a polygon is selected as a representative point when possible so that a meaningful place name can be selected as a representative

point of a polygon.

In addition, we defined unique IDs that uniquely identify administrative regions from the past to the present. Prefectural and municipal codes nationwide set by the Ministry of Internal Affairs and Communications (at the time of the Ministry of Home Affairs) can be used for administrative regions after 1968, but cannot be used for administrative regions that have disappeared before then. We therefore developed an algorithm to assign new unique IDs to all administrative regions by dividing the name space of unique IDs before and after 1968.

#### 4. Geoshape –A Site for Sharing Geometry Data

Starting with the Historical Administrative Region Dataset, we plan to share various types of geometric data on the Geoshape site. For example, “Town and Village Boundary Data (2015)” from e-stat offers administrative regions in a finer scale. Moreover, in February 2020, we released a website that allows users to seamlessly zoom in from a city-level to a town and village-level nation-wide. We also computed the overlap between those polygons so that users can easily check which cities in history a present town or village belonged to.

We are now adding geometric data such as river data and meteorological data to Geoshape and evolve it as the hub of geographic information with entity-based search, visualization and integration functionalities.

#### Acknowledgment

This work was supported by ROIS-DS-JOINT 020RP2018 and 035RP2019 to Takeshi Murata.

Keywords: Geometric Data, Administrative Region, Dataset, GeoJSON, Vector Tile, Entity-Based Geographic Information Systems