

GeoLOD: A Toponym Platform to Connect Japanese Historical Gazetteers and Historical Big Data Apps



Asanobu KITAMOTO

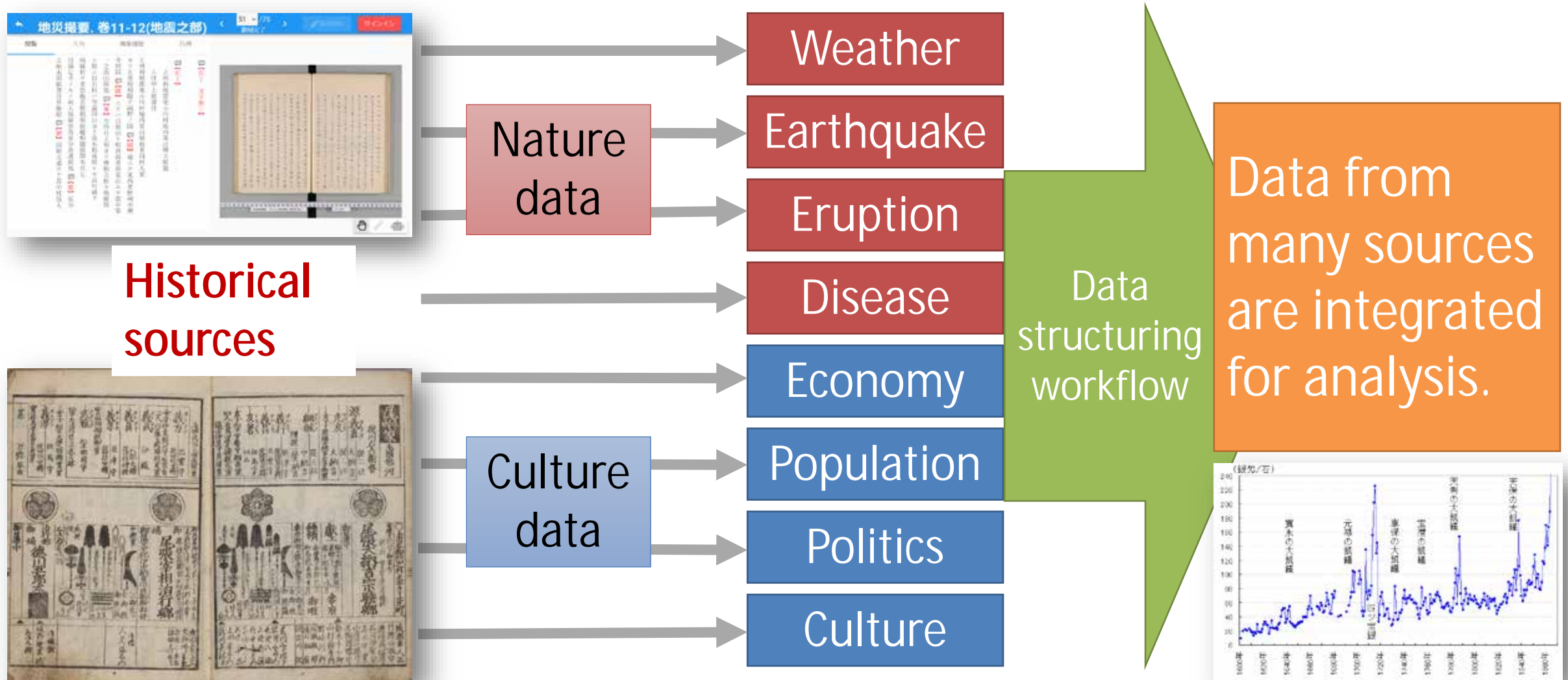
ROIS-DS Center for Open Data in the Humanities
(CODH) and National Institute of Informatics

Co-authors:

Yuta Hashimoto, Yasuyuki Kano, Junzo Ohmura

Historical Big Data (HBD) Project

<http://codh.rois.ac.jp/historical-big-data/>



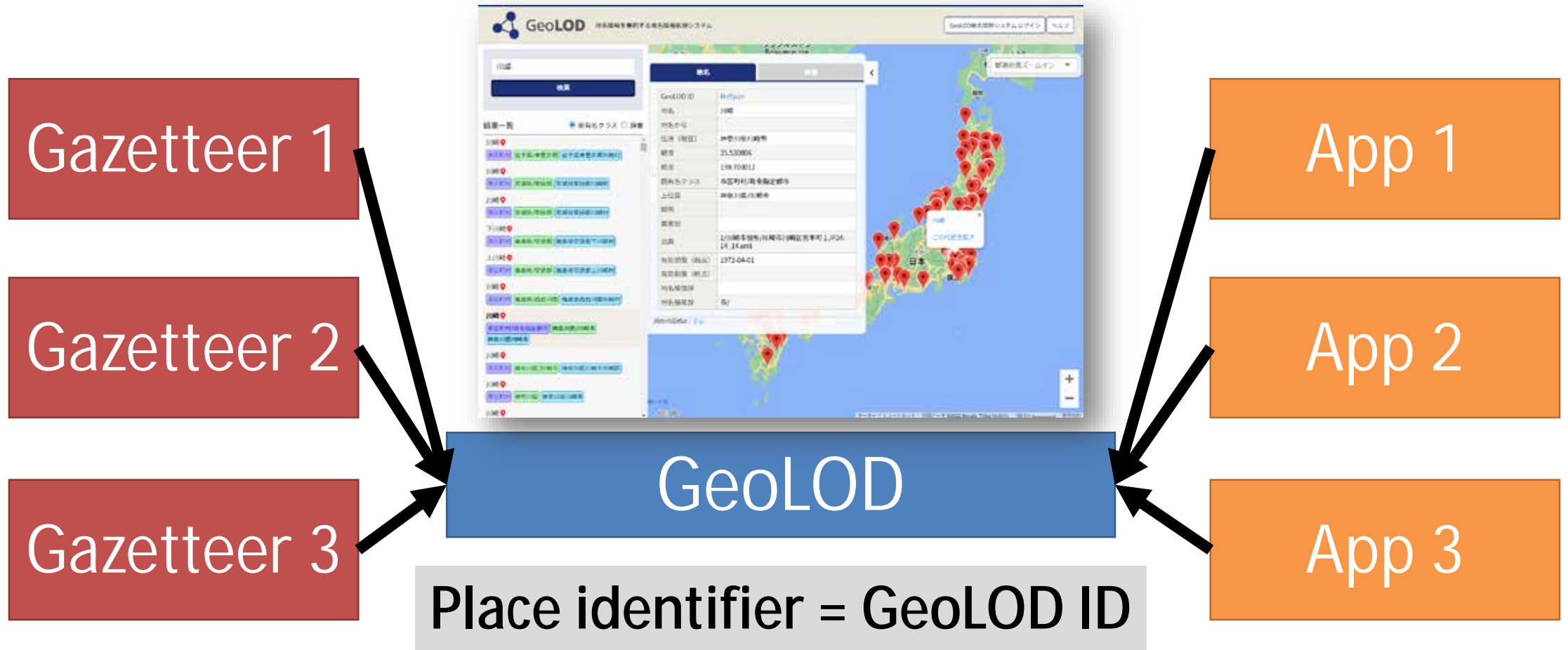
Two Aspects of Place Names



1. **The textual representation** includes variations and alternative names.
2. **The geographic representation** consists of a representative point or a polygonal boundary.
3. **Temporal dimensions** are also essential for historical big data research.

GeoLOD: Connect Gazetteers and Apps

<https://geolod.ex.nii.ac.jp/>

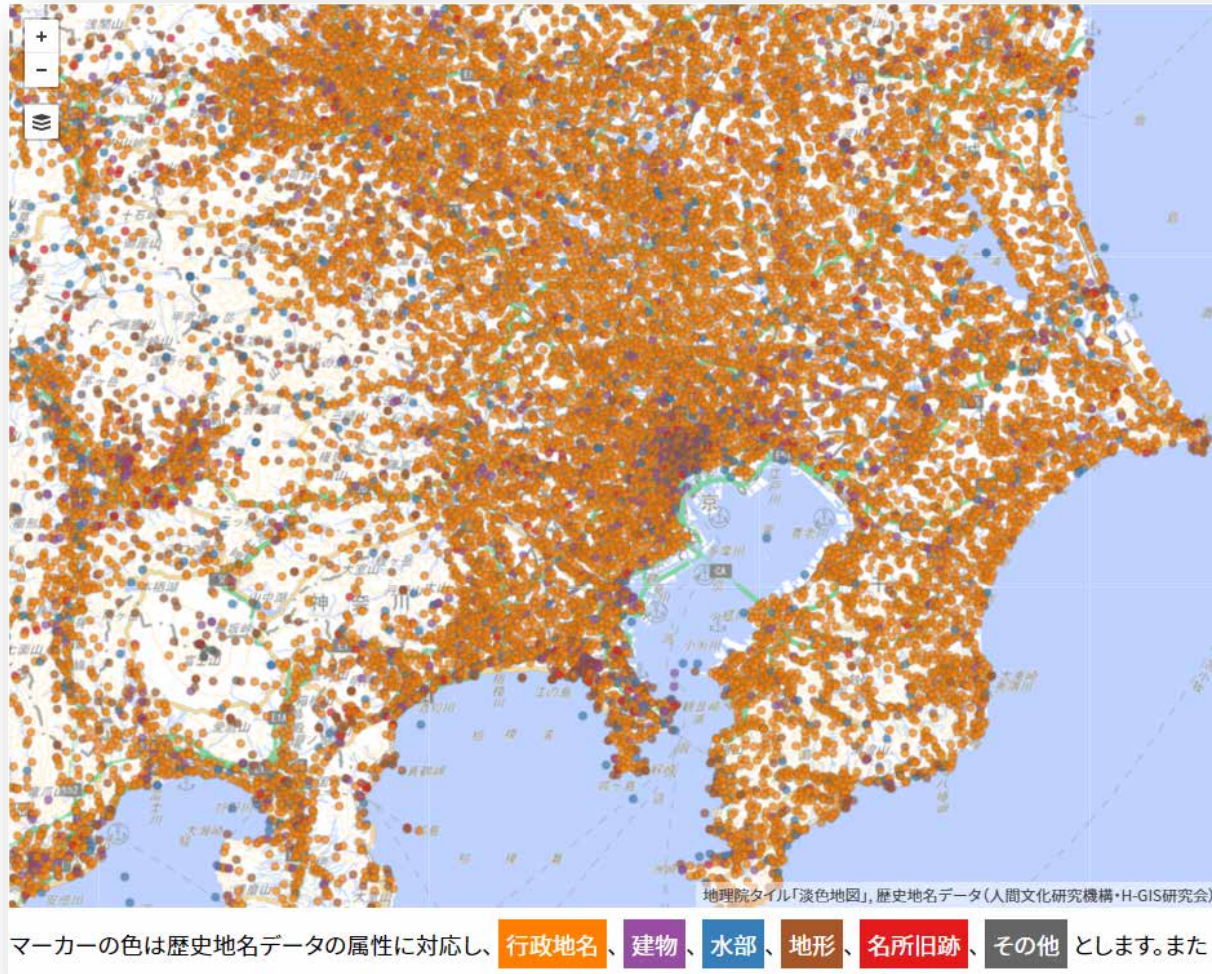


How to Create Historical Gazetteers?

1. **Convert** existing digital place name datasets into the GeoLOD format.
2. **Extract** place names from old maps and apply georeferencing when necessary.
3. **Integrate** existing datasets and assign unique identifiers for the place name.
4. **Collaborate** with a publisher to create an open dataset together.

1. Map of Historical Place names

<http://codh.rois.ac.jp/historical-gis/nihu-map/>



1. NIHU and H-GIS group released "**historical place name data**" dataset containing **298,914** place names.
2. Convert the dataset into the GeoLOD format and assigned the GeoLOD ID.

2. Edo Maps

<http://codh.rois.ac.jp/edo-maps/>

From **29** sheets, **8719** place names were extracted.

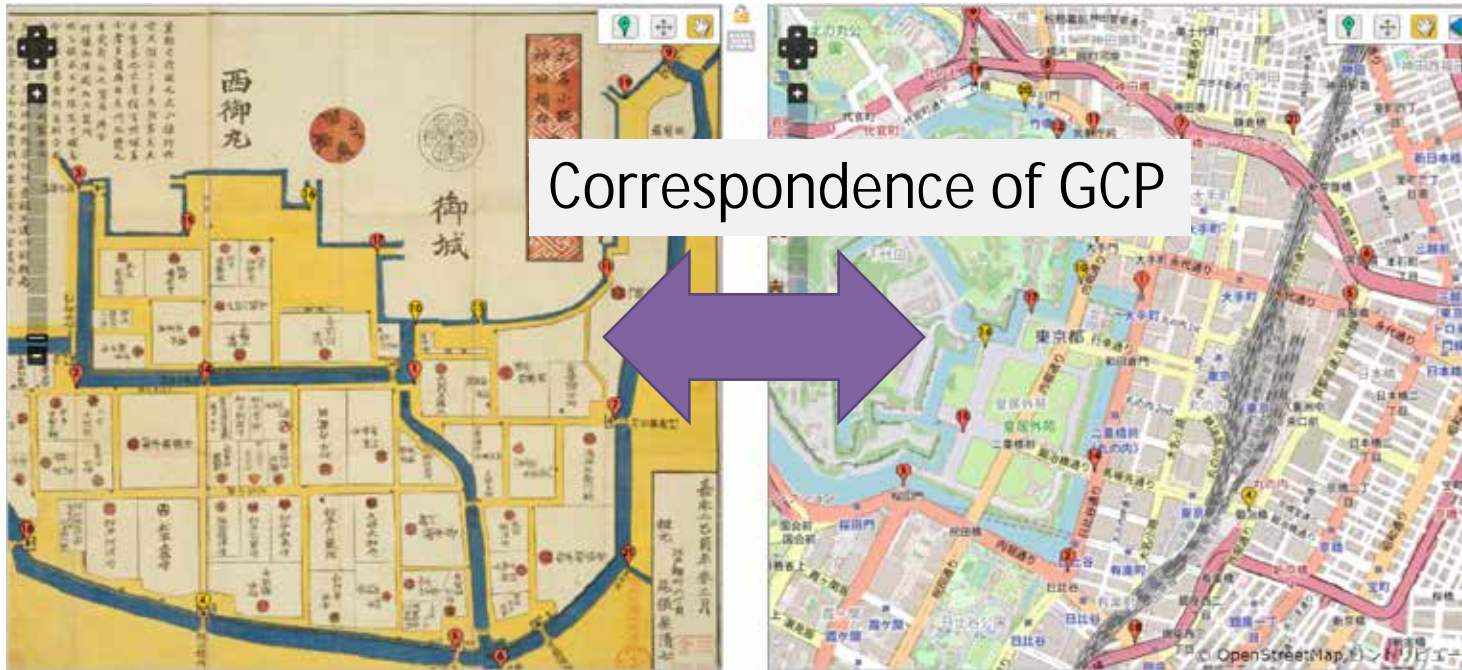
番号	分類	現代語訳	翻刻	地図
2-001	施設	幸橋御門	幸橋御門	拡大図
2-002	施設	山下御門	山下御門	拡大図
2-003	施設	数寄屋橋御門	数寄屋橋御門	拡大図
2-004	施設	鍛冶橋御門	鍛冶橋御門	拡大図
2-005	施設	呉服橋御門	呉服橋御門	拡大図
2-006	地名	一石橋	一石橋	拡大図
2-007	地名	出橋	出橋	拡大図
2-008	町名	丸屋町	丸屋丁	拡大図

[2-296]

地名：磯辺大神宮（イソベ大神宮）

分類：寺社仏閣

Georeferencing Old Maps



National Diet Library
"Edo Kiriezu"

Ritsumeikan University
Map Warper for Japanese



Edo Maps + Map Warper tile
service

3. Historical Municipal Boundaries Dataset

<https://geoshape.ex.nii.ac.jp/city/>

Municipality (city, town, village) is an important place name because it appears frequently in many types of documents and statistics.

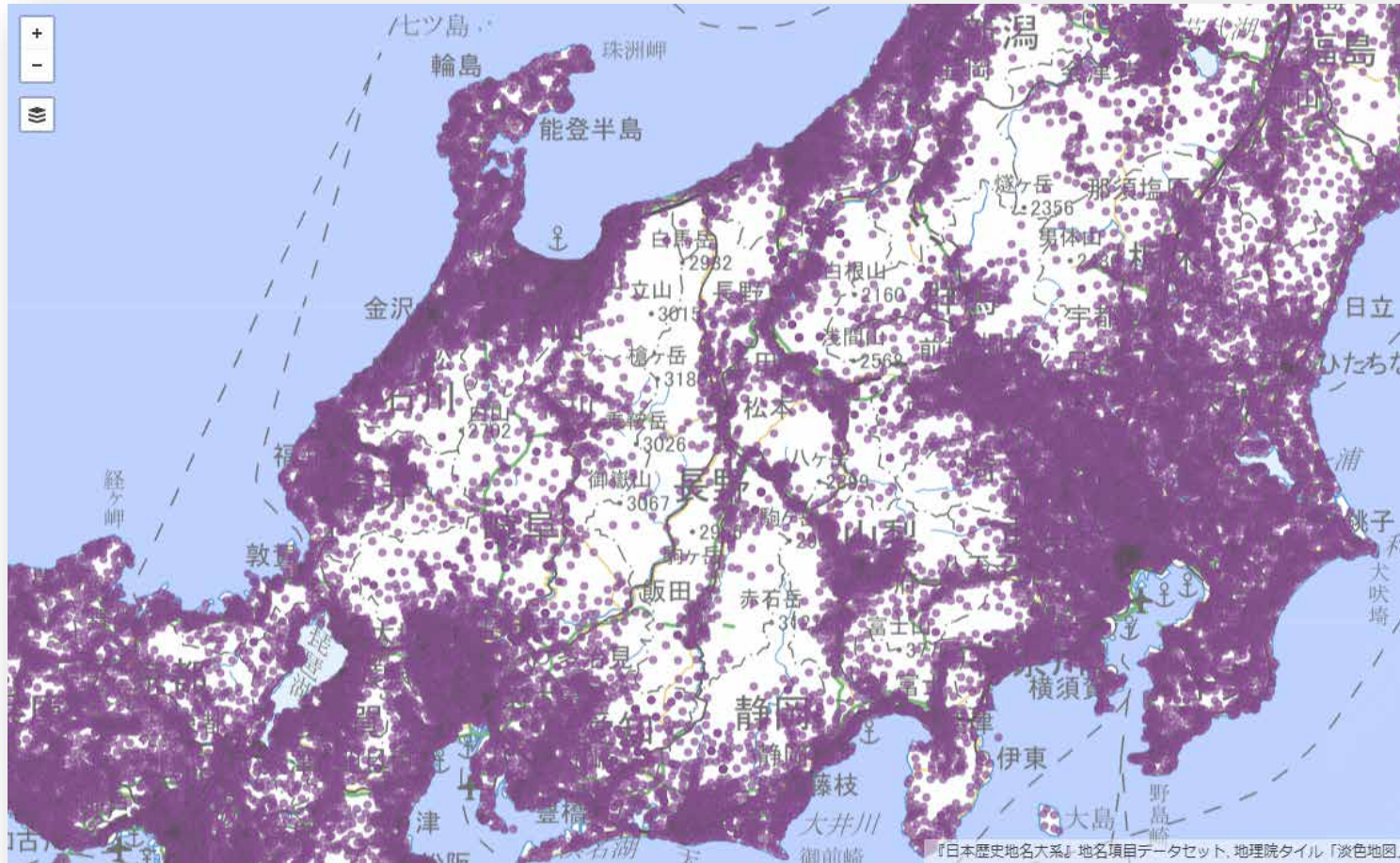


1. Official City ID **after 1968**.
2. Integrate multiple datasets to assign unique City IDs **after 1889**.
3. **16,856** cities were uniquely identified based on the identity of the spelling.
4. Each ID has attributes for a **representative location (point) and temporal boundaries (polygon)**.

4. Nihon Rekishi Chimei Taikei

<https://geoshape.ex.nii.ac.jp/nrct/>

1. Collaborate with a publisher (Heibonsha) to create an open dataset together.
2. **80,502** place names from the Edo Period (pre-1868) to present.



How to Create Historical Big Data Apps

The city is on the Potomac River, across from Virginia, and shares land borders with Maryland to its north and east.

Named Entity Recognition

The city is on the Potomac River, across from Virginia, and shares land borders with Maryland to its north and east.

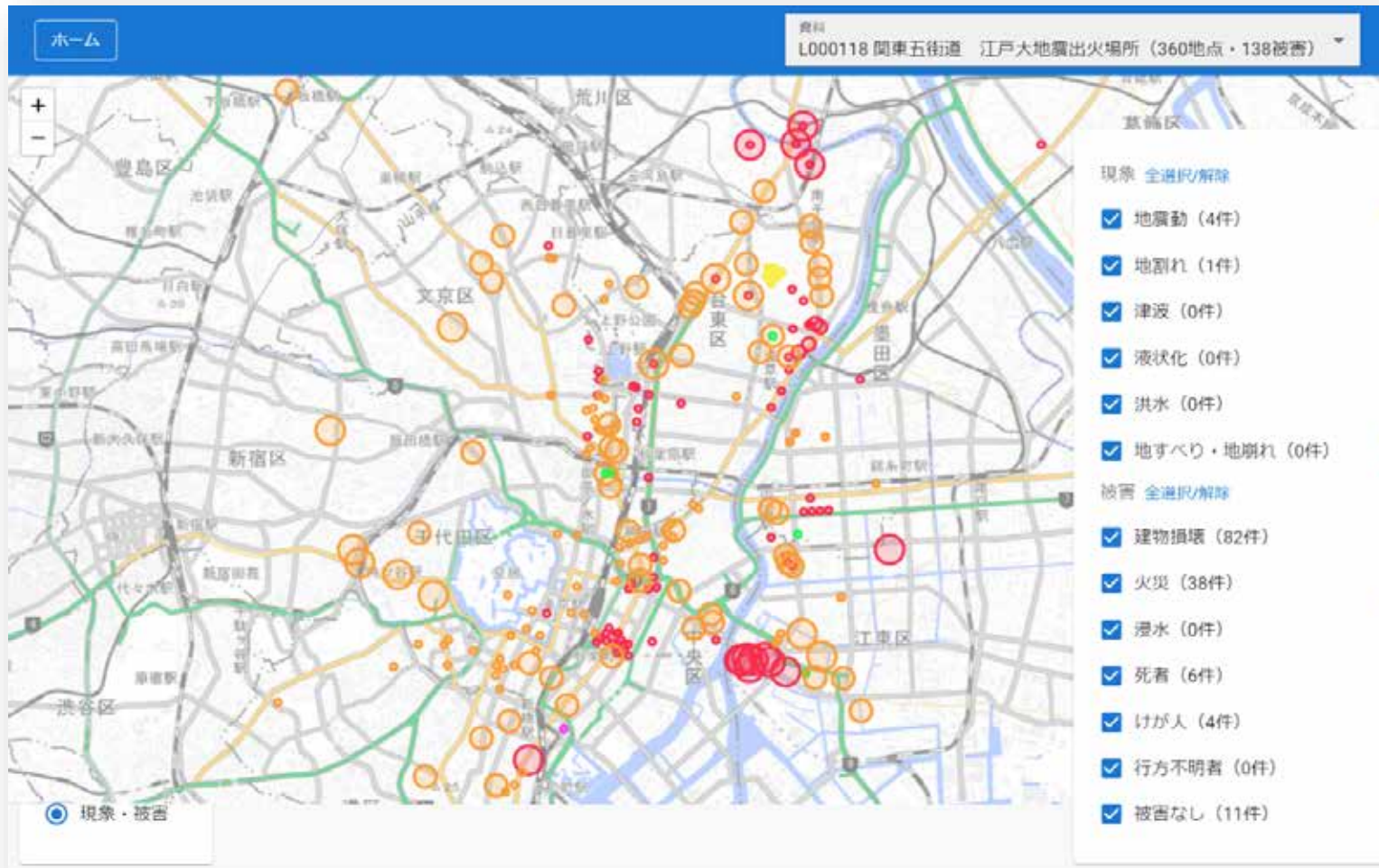
Entity linking: A mention in the text is linked with an entity in KB via the identifier (ID).

Disambiguation

Original Text	ID	Location
Potomac River	https://en.wikipedia.org/wiki/Potomac_River	geo:38,-76.335
Virginia	https://en.wikipedia.org/wiki/Virginia	geo:38,-79
Maryland	https://en.wikipedia.org/wiki/Maryland	geo:39,-77

Minna de Chushaku (Let's Annotate Together)

<https://ansei2.vercel.app/>



1. **Develop** a collaborative annotation platform for Japanese historical text.
2. **Collect** annotated data about the damage of the Ansei Edo Earthquake in 1855.
3. **Visualize** the damage on the map and analyze it with domain experts (seismologists).

Summary

<http://codh.rois.ac.jp/>

1. **GeoLOD is a toponym platform** to connect (historical) documents with the real world via place names.
2. **Gazetteers** can evolve collaboratively with scholars and citizens, especially for fine-grained place names.
3. **Apps** can diversify collaboratively with developers and domain experts, especially for domain-specific topics.
4. **Workflow to create linked data from historical documents is the basis for modern digital tools to study the past.**