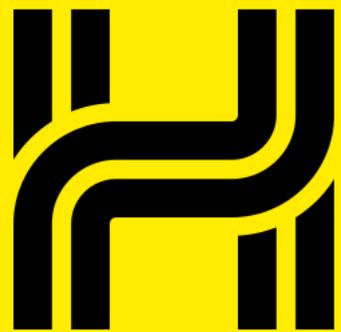


# 地理的史料を対象とした歴史 地名の構造化と統合に基づく 江戸ビッグデータの構築



**北本 朝展**・鈴木 親彦（ROIS-DS人文学オープン  
データ共同利用センター／国立情報学研究所）  
寺尾 承子・堀井 美里・堀井 洋（合同会社AMANE）  
<http://codh.rois.ac.jp/>

# 歴史ビッグデータとは？

1. 現代のデータを想定して開発されたビッグデータの方法論を、過去の記録に延長すること。
2. **生データの構造化**：機械可読なデータを生成し、様々な分野における解析などに利用する。
3. **多様な（Variety）データの統合解析**：単体のデータだけでは見えてこない新しい知見を得る。
4. **人文学研究のデジタルトランスフォーメーション**：研究基盤の構築が必須。技術要素だけではなく、研究体制や制度の変革も必要。

# 歴史ビッグデータの統合解析



歴史的資料  
(史料)

自然  
科学的  
データ

人文  
社会的  
データ

天気

天候

地震

噴火

経済

人口

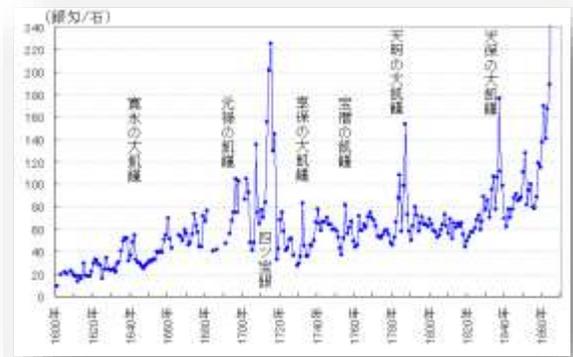
政治

文化

データ  
構造化  
ワーク  
フロー

過去のビッグデータを統合解析するための基盤技術の研究

歴史ビッグデータ基盤  
(機械可読)





Time Machine Europe

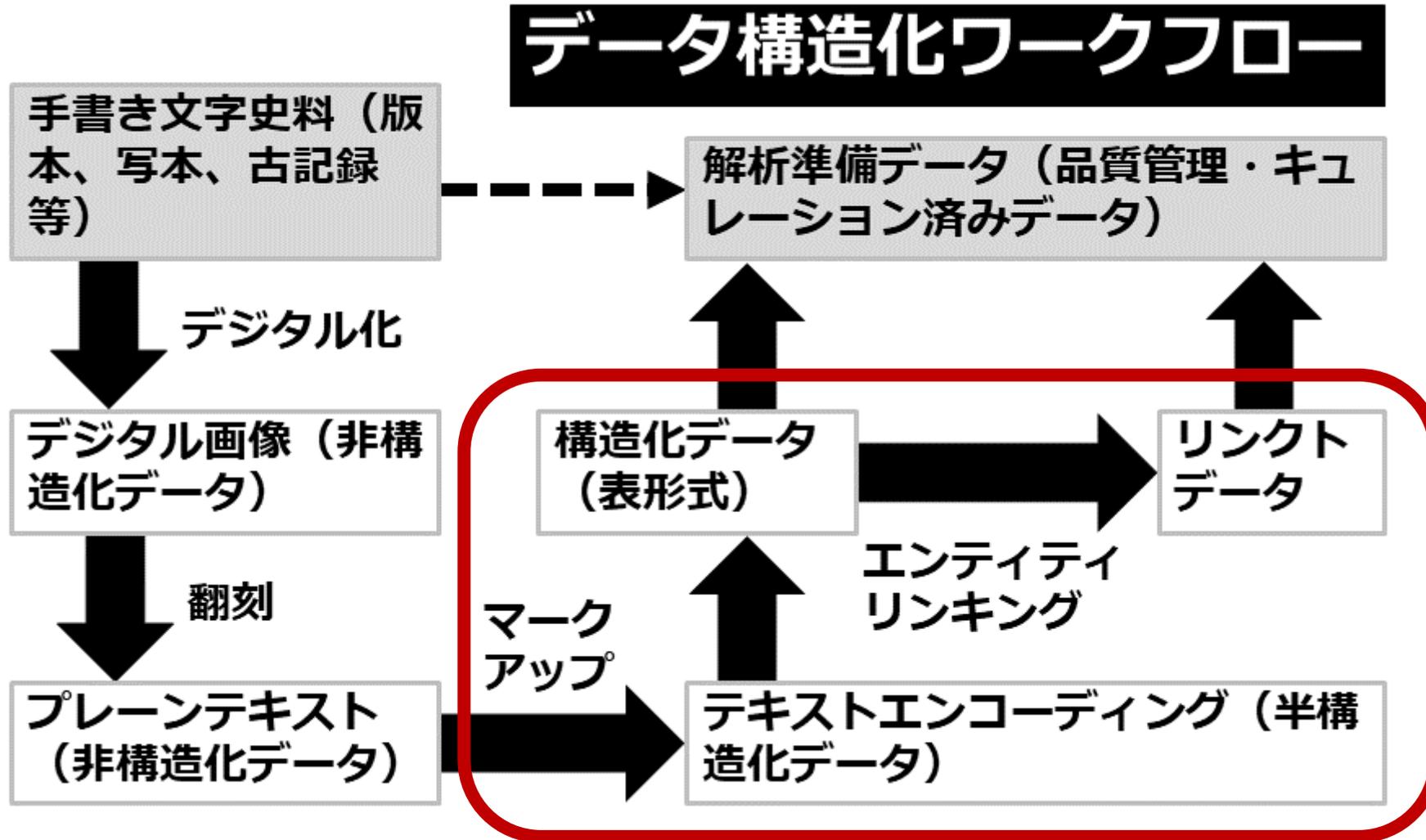
<https://www.timemachine.eu/>

カレントアウェアネス-E  
「過去のビッグデータ」を探る  
欧州タイムマシン研究計画

<https://current.ndl.go.jp/e2248>

ADDING A NEW DIMENSION TO THE  
PAST

# データ構造化ワークフロー



# 史料と地名

1. **地名**：空間の一部に対する呼称。住所等の「行政地名」と、Point of Interest (POI) 等の「施設名」を含む。
2. **地名リソース**：地名に識別子と属性を与えたデータセット。地名情報基盤で統合・共有する。
3. **歴史地名への対応**：時代による変化、表記揺れ（文字種の揺れを含む）、別名、履歴などを考慮。
4. **史料と実世界のリンク**：史料中の地名を経由して実世界と紐づける、一般的な枠組み（基盤）が必要。

# 史料とエンティティのリンク



宇田川丁、三島丁・  
神明丁、此分潰家多、  
土蔵残所なし

固有表現認識

宇田川丁、三島丁・  
神明丁、此分潰家多、  
土蔵残所なし

曖昧性解消

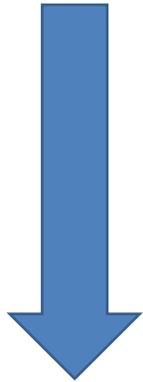
御江戸大地震大破并出火類焼場等書上之写（みんなで翻刻）

1. 江戸マップに出現する地名を、地名リソース（エンティティのデータベース）として整備する。
2. 史料の文字列から固有表現（地名）を抽出する。
3. 特定のエンティティとリンクすることで、実世界と紐づける。

原資料表記	江戸マップID	江戸マップ表記
宇田川丁	4-358	宇田川町
三島丁	4-290	三島丁
神明丁	4-294	神明町

# 地名識別子と疎結合性

私は**東京**に行く。



35.676666, 139.762222

地名を直接的に緯度経度に変換する方法は後から修正がしづらく、データ構造化と分析・可視化が一体化している（**密結合**）。

私は**東京**に行く。



地名識別子：UoGwci



35.676666, 139.762222

地名を識別子にリンクし、識別子の属性の緯度経度を用いる方法は、分析・可視化の目的に応じて属性を変更できる（**疎結合**）。

# 本発表の構成

- **地理的史料からの地名抽出**
  1. 武鑑全集
  2. 江戸観光案内
  3. 江戸買物案内
- **地名リソースの構築**
  1. 歴史地名マップ
  2. 江戸マップβ版
- **地名リソースと地理的史料のエンティティリンクング**
- **歴史的行動記録での活用**



# 武鑑と地名（地理情報）



- **上屋敷**や**菩提寺**などの施設は、居所に関する情報を持つ。
- 様々な情報を参考に**現代地図**とリンクする。
- **江戸マップ**中の**屋敷**や**寺院**ともリンクする。

上屋敷地図： <http://codh.rois.ac.jp/bukan/book/200018823/kamiyashiki/>

# 江戸観光案内 – 観光ビッグデータ

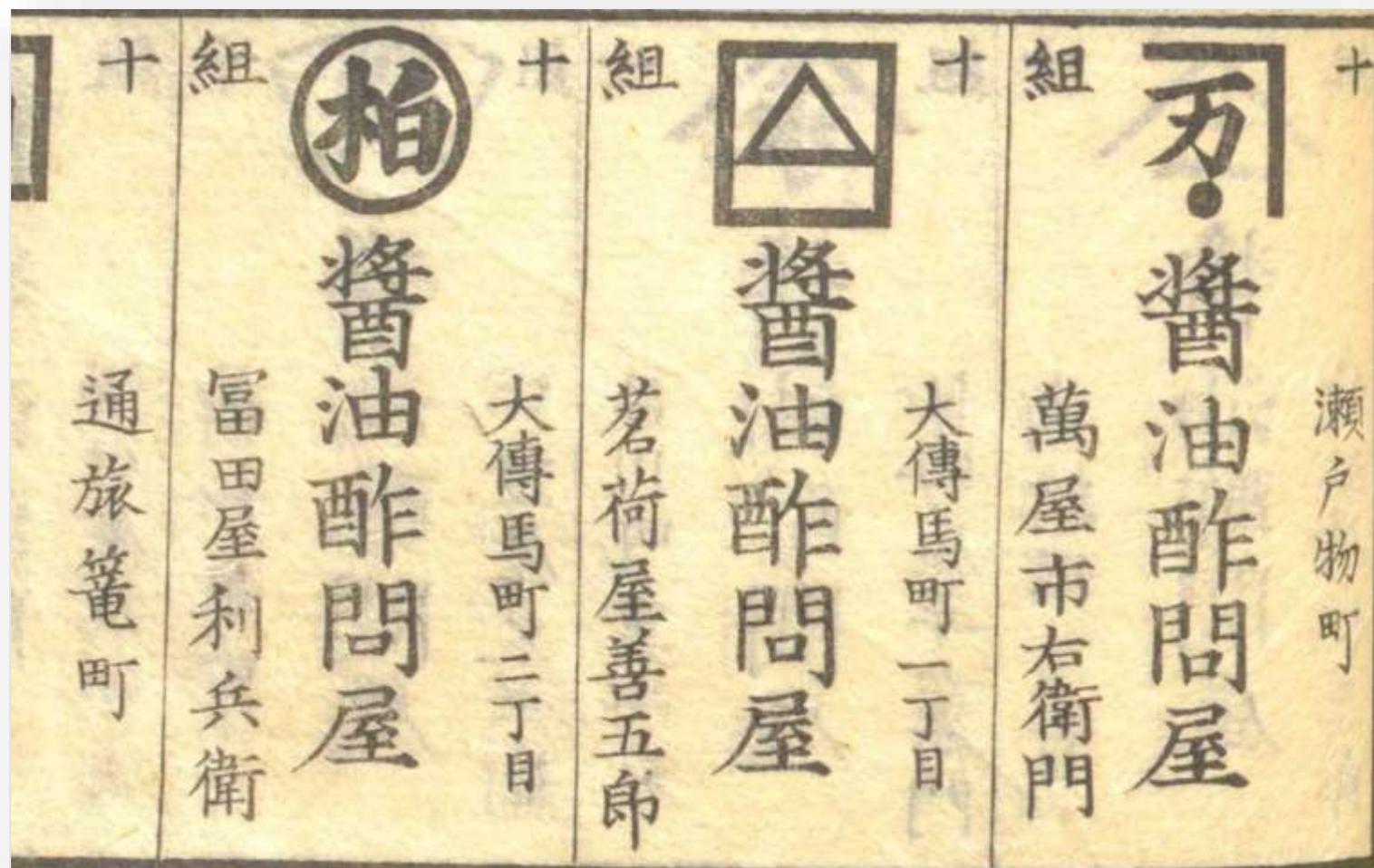
<http://codh.rois.ac.jp/edo-spots/>



- 江戸時代の名所記や名所案内を各世紀2点ずつ選択。
- IIF Curation Viewerを用いて挿絵を切り抜き、地名を翻刻し、メタデータを付与。
- 江戸マップ、歴史地名データとリンク。

# 江戸買物案内 – 商業ビッグデータ

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



- 『江戸買物独案内』（1824）から、商人名や居所などを抽出しデータベース化。
- 歴博の「江戸商人・職人データベース」は現在の区レベル。
- 江戸買物案内は、江戸マップの町レベルでリンク。

# 歴史地名マップ

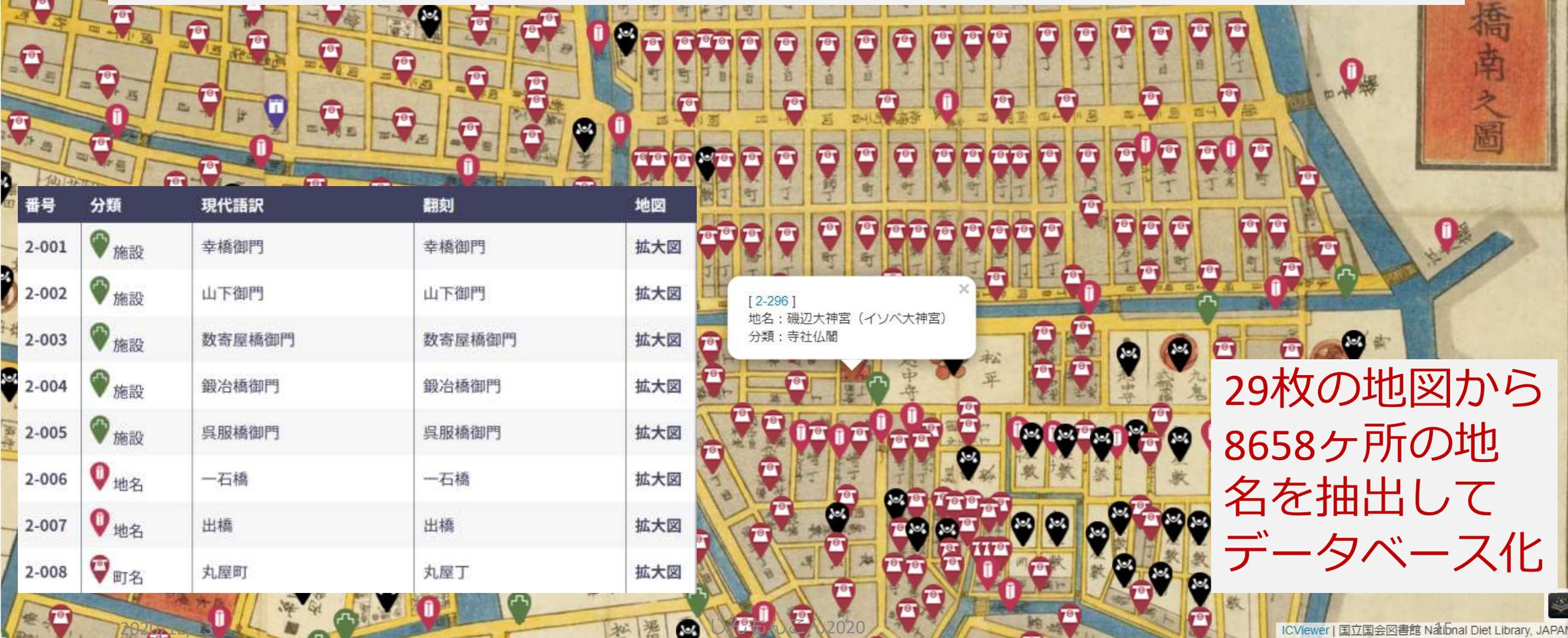
<http://codh.rois.ac.jp/historical-gis/nihu-map/>



1. 人間文化研究機構・H-GIS研究会が公開する「**歴史地名データ**」298,914件を地図にまとめて表示。
2. **バイナリベクトルタイル**の活用により、多数地点をウェブ地図に表示。
3. **歴史地名データの識別子と属性**を、地名リソースの一つとして活用。

# 江戸マップ - 江戸切絵図の地名リソース化

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



The image shows a historical map of Edo with numerous red location pins overlaid. A table in the bottom-left corner lists specific locations, and a tooltip in the center provides details for a selected pin. A large red text box on the right summarizes the project's scope.

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

29枚の地図から  
8658ヶ所の地  
名を抽出して  
データベース化

# 江戸切絵図の構造化

1. 尾張屋版（1849-1862） = 江戸切絵図の中で最も大量に売れて普及した地図。うち29枚を対象とする。
2. 国立国会図書館「デジタルコレクション」で公開されるIIF画像を、IIF Curation Viewerで読み込み。
3. 地名を矩形で囲んで画像座標を記録し、矩形中の文字を翻刻・現代語訳し、データベース化。
4. 「施設」「屋敷地」「寺社仏閣」「店名」「地名」「町名」「海川池」「観光地」「その他」の9分類を付与。ただし旗本屋敷などは数が多いため除く。

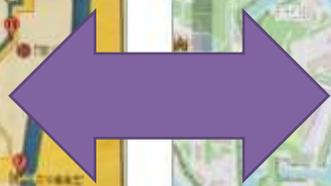
# 古地図を地名リソース化

1. 過去の地名識別子を現代に接続するには、緯度経度を属性として付与する必要がある。
2. Google Maps等のジオコーディング：緯度経度が直接得られるが、歴史地名への対応に難がある。
3. 歴史地名データ（NIHU等）：一点ずつ緯度経度を確認すれば、精度は高いがコストも大きい。
4. 古地図のジオレファレンス：古地図全体を一度に変換して緯度経度を付与できるが、精度に限界がある。

# 古地図のジオレファレンス



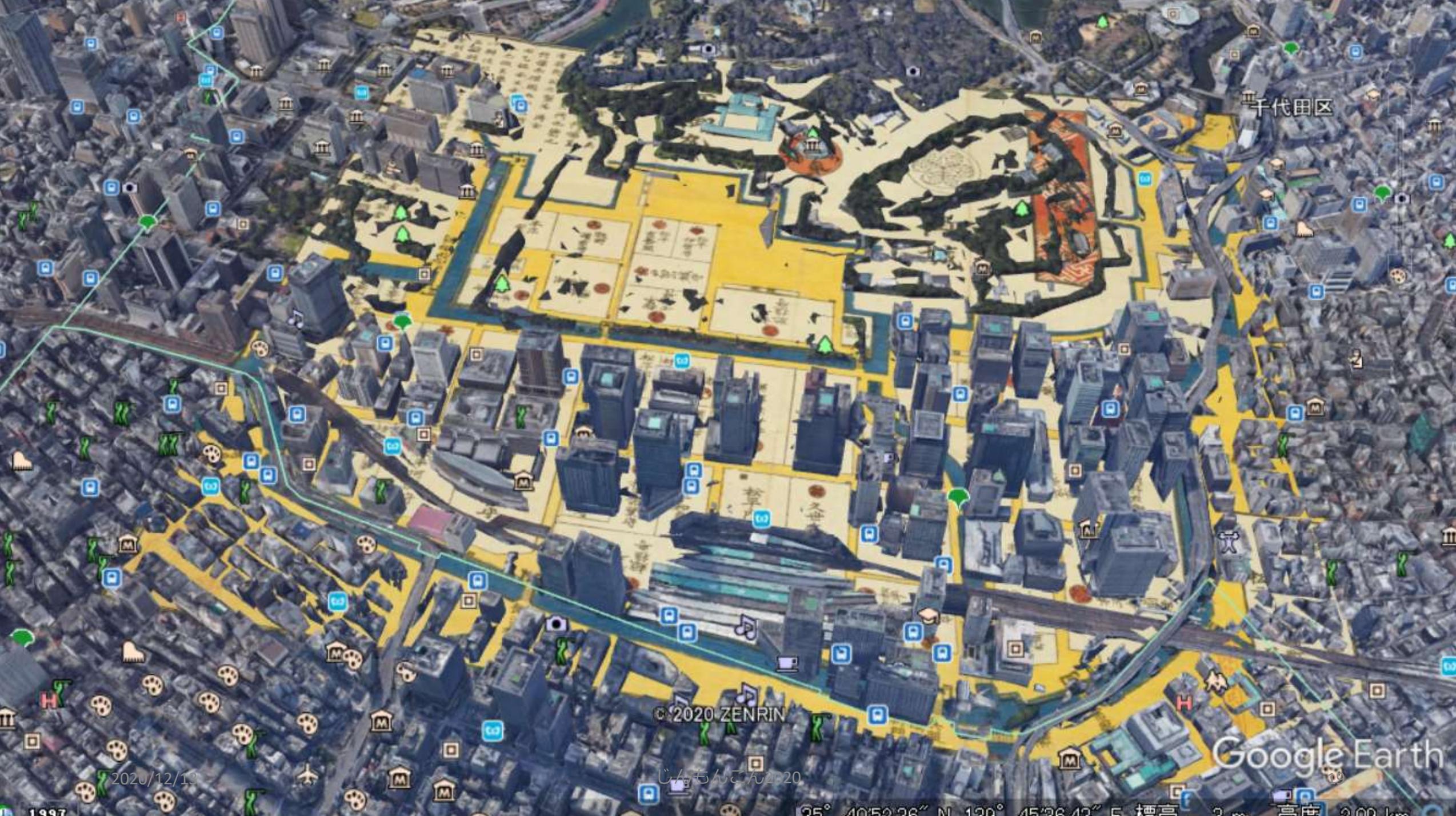
GCPの対応付け



国立国会図書館  
『江戸切絵図』

立命館大学  
日本版Map Warper

江戸マップβ版 + 日本版Map  
Warperタイル配信サービス



千代田区

Google Earth

© 2020 ZENRIN

じゅんちんちん2020

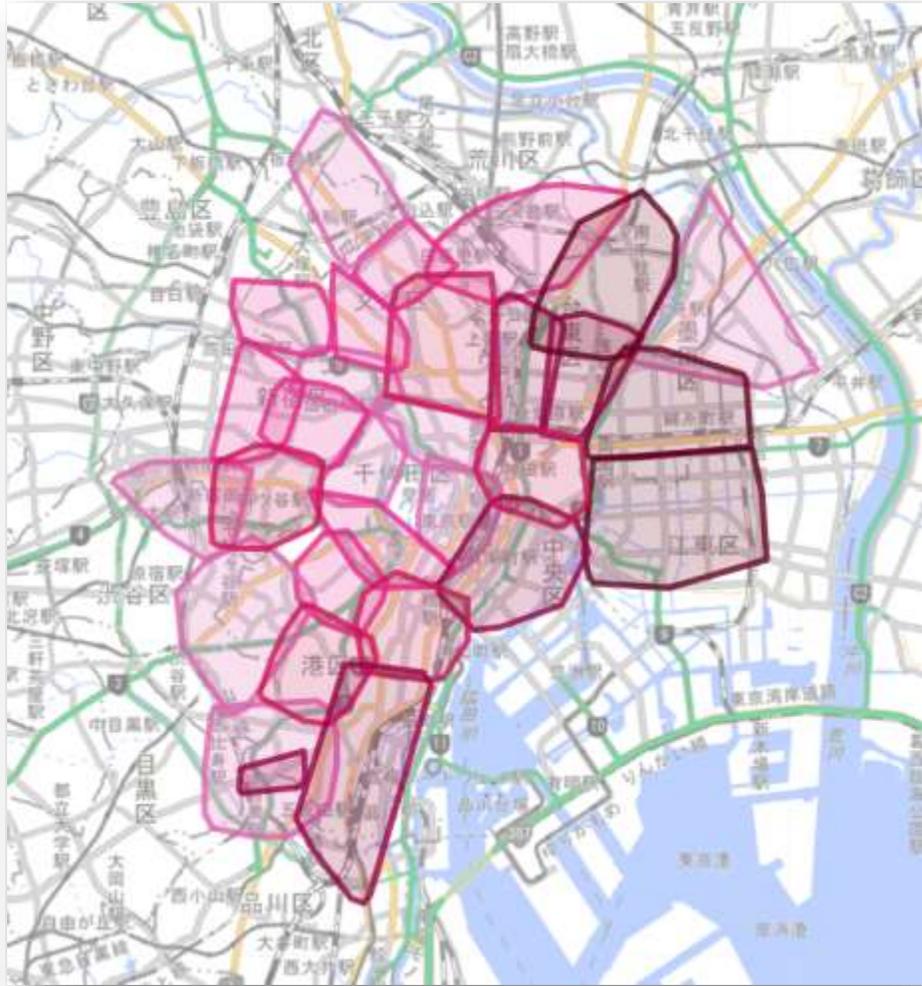
2020/12/13

1997

35° 40'52.26" N 139° 45'26.42" E 標高 2 高度 2.00 km

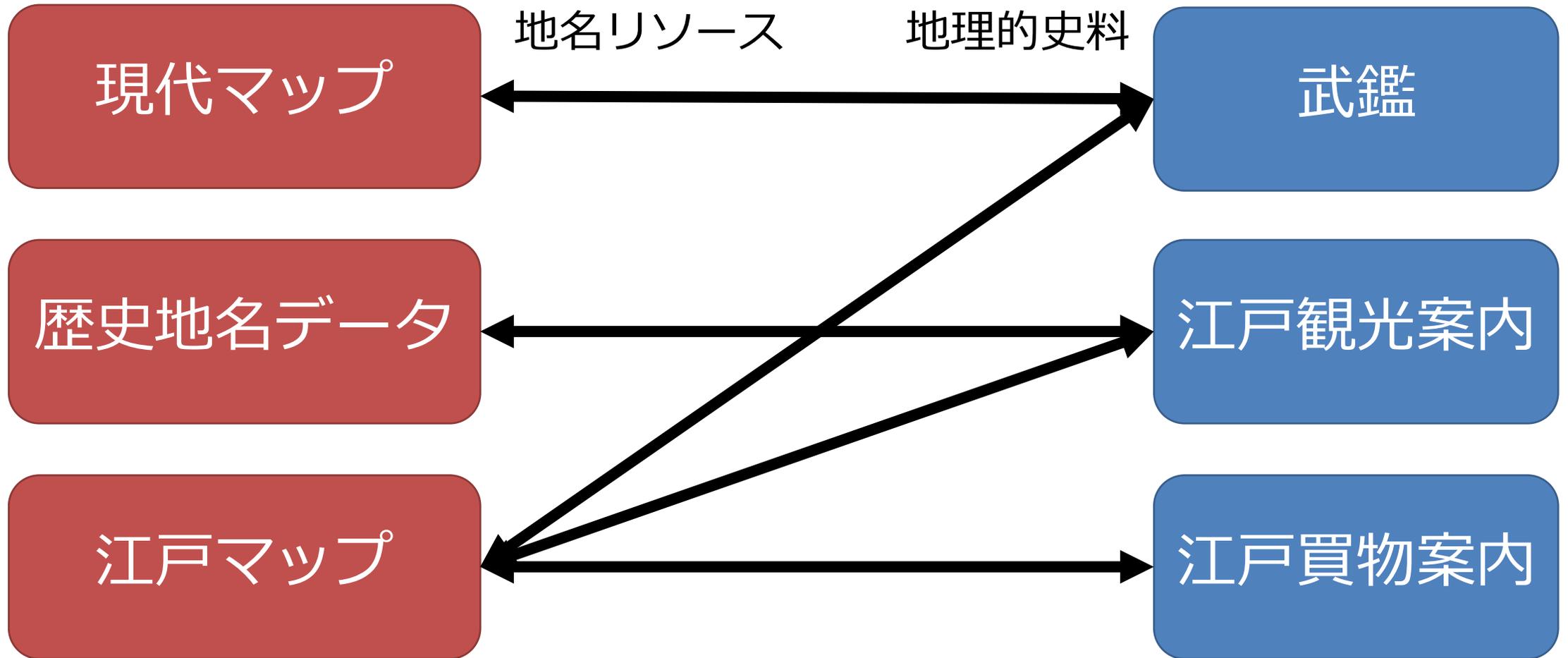
# 江戸切絵図の範囲

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



1. 位置合わせ可能な地名を対象とした外接多角形。**地名が多いほど濃い赤**で着色。
2. **東側の方が西側より地名が多い**傾向がある。
3. 地名が少ないのは、**武家屋敷**または**田畑**が多いため。
4. 観光地を入れるため、**意図的に歪めた地域**もある。

# 地名リソースと地理的史料のエンティティリンキング



# 武鑑の結果

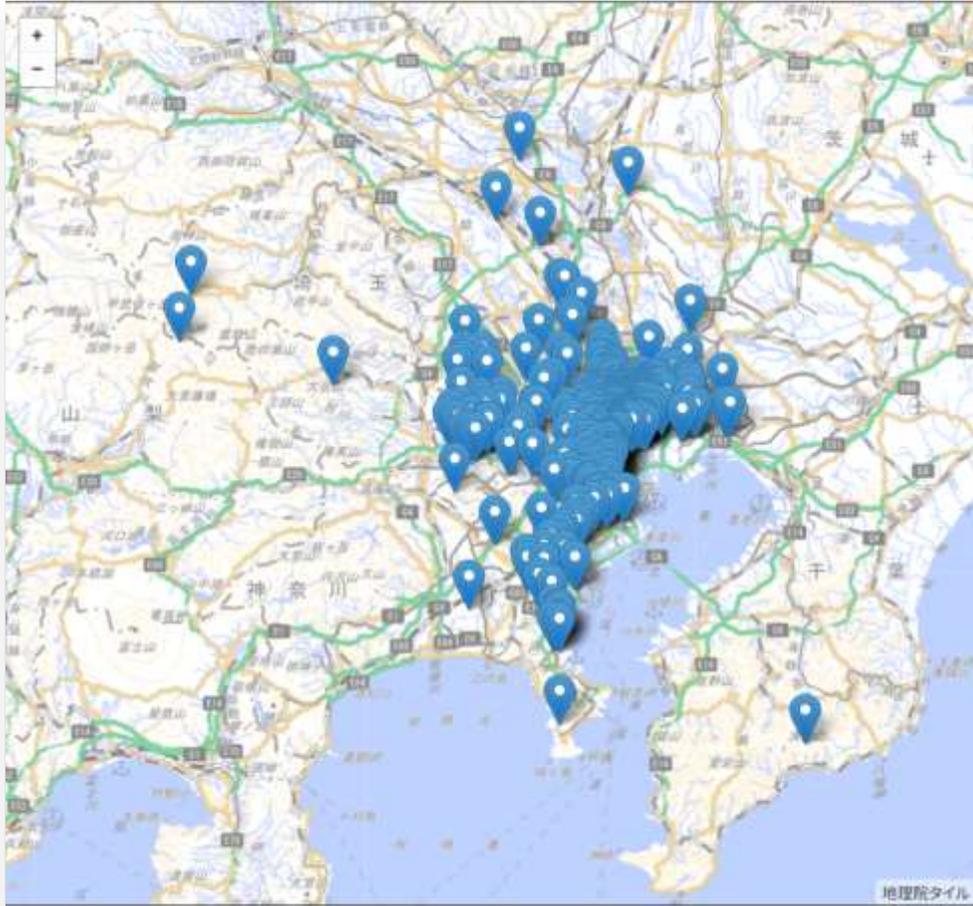
## • 上屋敷（264家）

- 現代マップへのリンクは261家で可能。
- 江戸マップへのリンクは264家すべて可能。
- ただし、江戸切絵図は『寛政武鑑』より60年以上後のため、その間に264家中24家が移転している。

## • 菩提寺（264家）

- 駒沢学園「寺院資料データベース」へのリンクは246家可能。
- 江戸マップへのリンクは242家可能。
- 江戸外の菩提寺や寺院の移転などがあるため、上屋敷よりもリンクできる件数は少ない。

# 江戸観光案内の結果



- **1255件**の挿絵を収集。その中で地名が明確な挿絵は**1219件**。
- **歴史地名データ**とのリンクは**765件（全体の6割強）**で可能。
- 歴史地名データは明治時代の地名が多いが、**観光名所の地名は江戸と明治で大きな変化はなかった可能性**がある。
- **江戸マップ**とのリンクは作業中。

# 江戸買物案内の結果

居所 (原本表記)

本町三丁目 (73)	通油町 (44)	南新堀一丁目 (42)	南茅場町 (35)	小網町三丁目 (34)
日本橋通一丁目 (33)	本町四丁目 (32)	小舟町一丁目 (28)	大橋馬町二丁目 (28)	
大橋馬町一丁目 (26)	伊勢町 (24)	小網町二丁目 (23)	日本橋通二丁目 (22)	通徳町 (21)
通旗屋町 (20)	南新堀二丁目 (20)	本船町 (20)	橋山町一丁目 (19)	橋山町二丁目 (19)
本石町四丁目 (19)	室町一丁目 (18)	下谷池之端仲町 (17)	堀江町一丁目 (17)	
尾張町二丁目 (16)	日本橋通四丁目 (15)	北新堀町 (15)	室町二丁目 (14)	
浅草諏訪町 (14)	小舟町三丁目 (14)	堀留町二丁目 (14)	本石町二丁目 (13)	
本町二丁目 (13)	大伝馬町二丁目 (12)	瀬戸物町 (12)	本石町三丁目 (12)	
堀留町一丁目 (12)	室町三丁目 (12)	日本橋通三丁目 (11)	本材木町一丁目 (11)	
本町一丁目 (11)	馬喰町二丁目 (10)	小網町一丁目 (10)	両国吉川町 (9)	芝神明前 (9)
深川佐賀町 (9)	浅草南馬道町 (9)	本郷一丁目 (9)	品川町墨河岸 (9)	堀江町四丁目 (9)
小舟町二丁目 (9)	橋山町三丁目 (8)	浅草駒形町 (8)	馬喰町一丁目 (8)	
箱崎町一丁目 (8)	本湊町 (8)	てりふり町 (8)	鎌倉町 (8)	本郷二丁目 (8)
浅草並木町 (8)	本石町十軒店 (7)	小伝馬町二丁目 (7)	大伝馬町一丁目 (7)	
本郷三丁目 (7)	尾張町一丁目 (7)	新南橋町四丁目 (7)	南傳馬町二丁目 (7)	富澤町 (7)
豊岸崎銀町二丁目 (7)	本銀町四丁目 (7)	青物町 (6)	麹町五丁目 (6)	日本橋方町 (6)
江戸橋四日市 (6)	新南橋町 (6)	新和泉町 (6)	長谷川町 (6)	本郷四丁目 (6)
馬喰町四丁目 (6)	本石町一丁目 (6)	元飯田町 (6)	池之端仲町 (6)	大橋馬場町 (6)
京橋銀座一丁目 (5)	浅草聖天町 (5)	通油町南新道 (5)	新大坂町 (5)	深川仲町 (5)
田所町 (5)	芝口一丁目 (5)	芝田町九丁目 (5)	三河町一丁目 (5)	橋町三丁目 (5)
上橋原町 (5)	本材木町四丁目 (5)	裏町 (5)	豊岸崎四日市町 (5)	豊岸崎四日市町 (5)
元飯田町中坂 (5)	南傳馬町三丁目 (5)	小伝馬町三丁目 (4)	神田須田町二丁目 (4)	
通新石町 (4)	日本橋青物町 (4)	日本橋瀬戸物町 (4)	神田鍋町 (4)	京橋金六町 (4)
小伝馬町一丁目 (4)	深川上大崎町 (4)	堀江町三丁目 (4)	日本橋室町一丁目 (4)	
船町四丁目 (4)	小日向水道町 (4)	浅草北馬道町 (4)	濱町二丁目 (4)	両国吉川町 (4)
元濱町 (4)	南傳馬町一丁目 (4)	豊岸崎町 (4)	湯嶋横町 (4)	京橋弓町 (4)
浅草茅町一丁目 (4)	小傳馬町三丁目 (4)	箱崎町二丁目 (4)	両国樹橋 (4)	
赤坂表伝馬町一丁目 (3)	下谷広徳寺前 (3)	本銀町二丁目 (3)	大門通油町 (3)	

- 商人の居所は、〇〇町などの行政地名や、〇〇通り／〇〇橋といった交通に関する地名で示されることが多い。
- 江戸の細粒度地名は歴史地名データに少ないため、江戸マップとのリンクを試みた。
- 異なり地名**1046件**中、江戸切絵図外が**15件**ある。不明**7件**。

# 文字列マッチング

- 江戸買物案内の居所と、江戸マップの翻刻の文字列を比較し、**自動的なマッチングが可能か**試してみる。
  1. **完全一致** (芝口一丁目⇔芝口一丁目)
  2. **前方一致** (駿河町北側⇔駿河町)
  3. **後方一致** (日本橋新右衛門町⇔新右衛門町)
  4. **部分一致** (神田紺屋町薬師新道⇔紺屋町)
- **江戸買物案内の地名は統制が取れていない**。広告主が客にわかりやすい地名表現を選んでいるため？

# 文字列マッチングの結果

	成功	地図内の曖昧性	地図間の曖昧性
完全一致	956	168	36
前方一致	642	310	277
後方一致	384	219	148
部分一致	80	55	37
それ以外	392	-	-

部分一致も含め、江戸マップと対応した居所は**2062件**と**2454件中の84%**。一方**人手では99%**マッチング完了。

# 考察

1. 江戸観光案内のような**観光名所（POI）**には、**歴史地名データ**の方が適している。
2. 江戸買物案内のような**生活に密着した地名**には、**江戸マップ**の方が適している。
3. **歴史地名データと江戸マップは、補完的な地名リソースとして使える。**
4. 歴史地名データと江戸マップの間で完全一致する地名は**703件**。江戸マップ8658件の**8%程度**に過ぎない。

# GeoLOD : 地名リソースの統合基盤

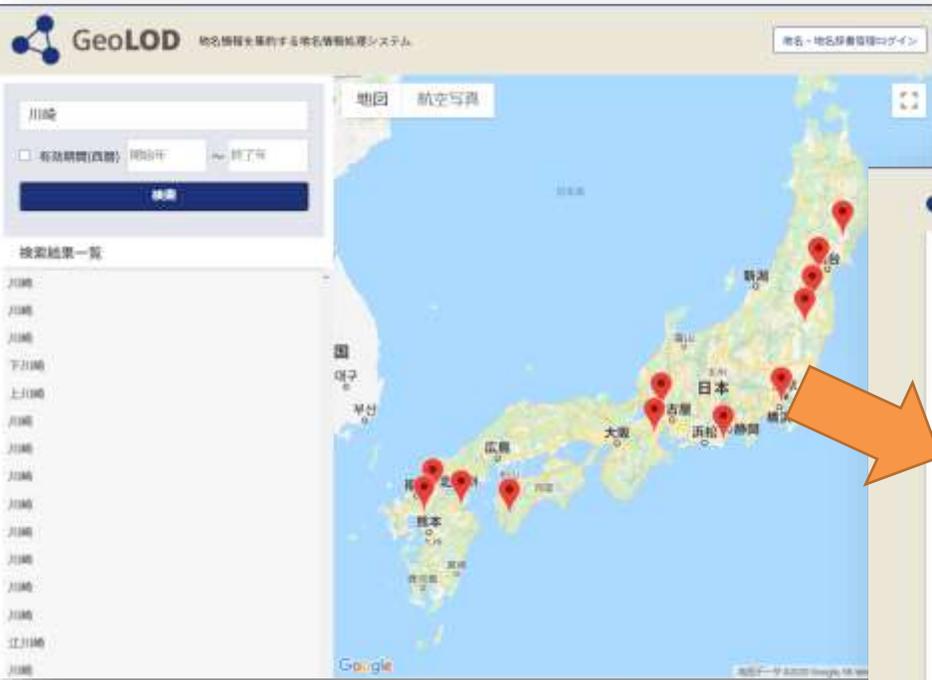
<https://geolod.ex.nii.ac.jp/>

地名に固有の識別子 (GeoLOD ID) を付与する。

地名辞書ごとに得意な分野は異なる。

統一的な識別子を付与する地名情報基盤を構築。

名称や緯度経度等の属性情報も管理。



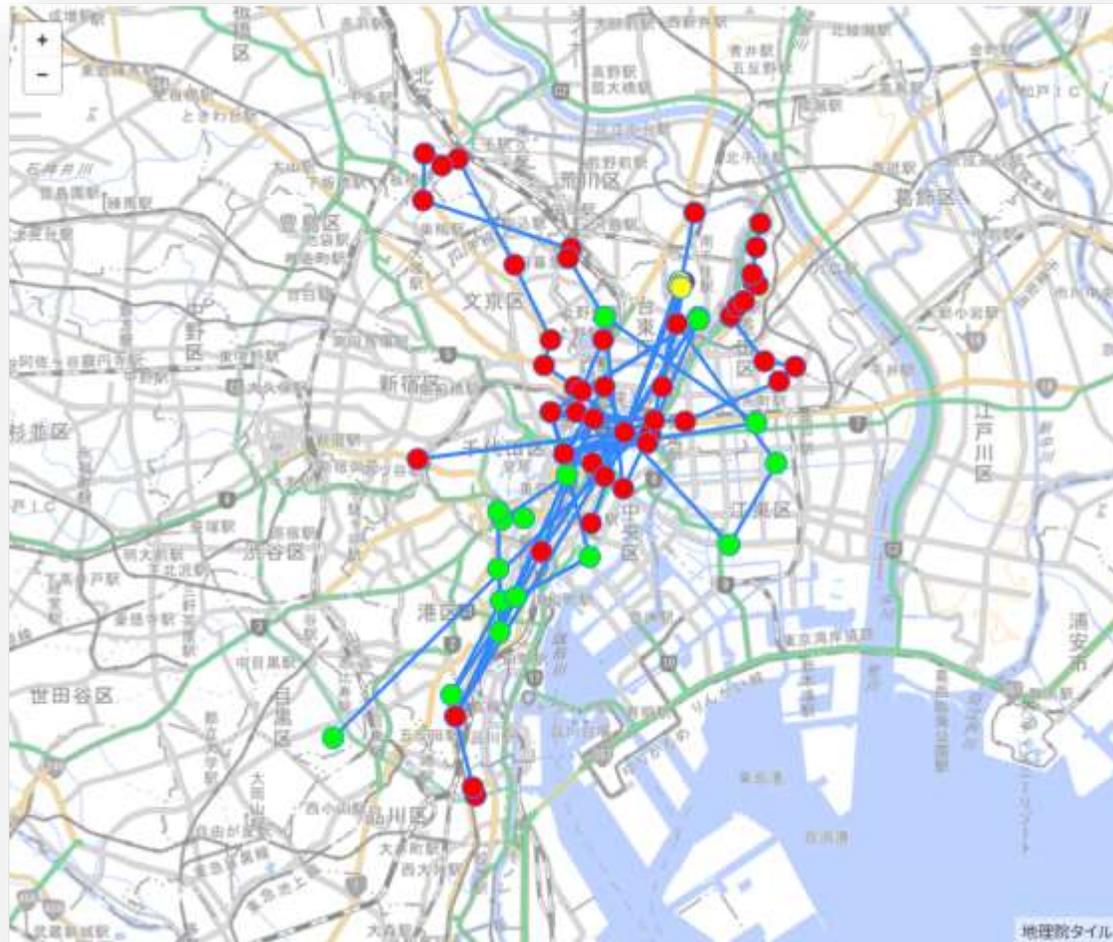
地名辞書の形式は、地名のエンティティリンキングシステム GeoNLPと相互運用可能。



# 江戸ビッグデータに向けて

- **歴史的状況記録**：どこで、いつ、だれが、何に対し、どのような観察を行ったか（天候や地震の分析に利用）。
- **歴史的行動記録**：どこからどこへ、いつ、だれが、どのような手段で、どのような目的で移動したか（Origin-Destination (OD) データ）。
- 「どこ」を地名リソースとのリンクとして管理することで、地理的なビッグデータ分析を可能とする。

# 歴史的行動記録での活用



- 清河八郎が安政2年（1855）に江戸を訪問した日記『西遊草』を分析。
- **日記から地名を抽出し、江戸マップや江戸観光案内とリンク。**
- **164件の地名中、江戸マップと132件、江戸観光案内と119件がリンク可能。**

# まとめ

1. 歴史ビッグデータのデータ構造化の課題の一つとして、**地名情報基盤の構築**を議論した。
2. **江戸に関する地理的資料**をIIF Curation Platformにより構造化し、**地名リソースとのリンク**を試みた。
3. **江戸マップ**や**歴史地名データ**の特徴を分析し、**複数の地名リソースの統合**への方向性を明らかにした。
4. 今後の課題：歴史ビッグデータのための**地名データ構造化を支援するシステム**の構想を進める。

# 謝辞

- 本研究は以下の支援を受けました。
  1. 歴史ビッグデータ研究基盤による過去世界のデータ駆動型復元と統合解析, 科学研究費補助金 基盤研究(A), 19H01141, 研究代表者: 北本 朝展, 2019-2021
  2. 人文学ビッグデータにおける構造化ギャップの克服と分野横断的利用の検証, 機構間連携・文理融合プロジェクト, 情報・システム研究機構, 研究代表者: 北本 朝展, 2018-2020
  3. 人文学資料マイクロコンテンツ化の実践研究 -江戸の都市空間再構築を通して-, 科学研究費補助金 若手研究, 20K20141, 研究代表者: 鈴木親彦, 2020-2023



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