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）

本論文は，歴史地名の構造化と統合のための地名リソースの構築を進めるとともに，地理的史料と地名リソースの間でのエンティティリンキングの可能性を複数のケーススタディの結果に基づき検証する．またこうして構築した地名情報基盤の上で「歴史的行動記録」を構造化しつつ，江戸という都市または時代に関するビッグデータ分析を進めていく上での現状と研究課題について議論する．

Construction of Edo Big Data Based on Structuring and Integrating Historical Placenames on Spatial Historical Sources

Asanobu Kitamoto / Chikahiko Suzuki (ROIS-DS Center for Open Data in the Humanities / National Institute of Informatics)
Shoko Terao / Misato Horii / Hiroshi Horii (AMANE LLC)

This paper focuses on the construction of the placename resource for structuring and integrating historical placenames, and validates the possibility of entity linking between spatial historical sources and the placename resource from the result of several case studies. We then work on structuring “historical activity record” on the toponym information platform, and discuss the current status and research challenges for promoting big data analysis for the city and the era of Edo.

１．まえがき

歴史ビッグデータとは，過去の史料に含まれる世界の記述を構造化し統合することで，データに基づき過去の世界を分析する研究である．そのためには，過去の史料の記述をどのように構造化して機械可読データに変換するかが技術的課題となる．そこで本論文は，データ構造化の中でも特に空間情報の問題に着目し，史料中に出現する地名を地名リソースの識別子とリンクし，現在の地図上で扱うという，エンティティリンキングのための手法に焦点を合わせる．

本論文は特に「江戸」を対象として研究を進める．まず地名リソースとして，江戸時代の古地図である「江戸切絵図」の地名を構造化するとともに，人間文化研究機構が公開する「歴史地名データ」も活用可能とする．次に地理的資料として「武鑑」「買物案内」「名所案内」を取り上げ，それらの史料中に出現する地名を地名リソースとリンクした結果を報告する．最後に江戸ビッグデータ（江戸に関する歴史ビッグデータ）の構築に向けて，歴史的行動記録を構造化する例を示す．

２．歴史ビッグデータと空間情報

歴史ビッグデータのデータ構造化ワークフローの概念図を図 1 に示す．これはデジタル化した史料をテキスト化し，そこから地名を抽出し，さらにこれをエンティティリンキングすることで，Linked Data として解析に用いるという流れを示したものである．ここで本研究が対象とするエンティティリンキングは，史料中に出現する文字列を，知識基盤に定義された地名識別子とリンクする作業である．これを実現するための研究テーマには，地名リソースの構築と地名アノテーション基盤の構築の 2 つがあり，さらに後者に関してはテキスト中の地名文字列を抽出する固有表現抽出のステップと，その文字列を実際の地名に対応付ける曖昧性解消のステップが必要となる．

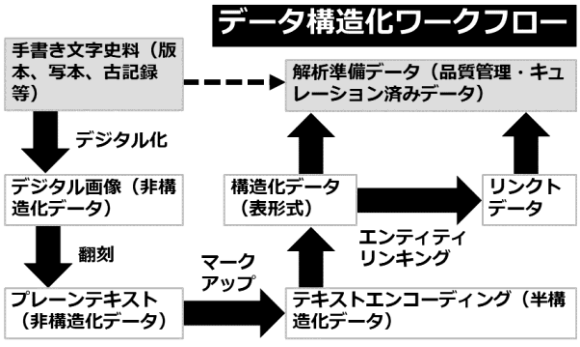

図 1 データ構造化ワークフローの概念図.

前者に関しては，テキストから地名文字列をマークアップすることが課題であり，TEI（Text Encoding Initiative）やその他のマークアップ技術

を用いた取り組みが進んでいる．一方後者に関しては，歴史地名の場合に以下の 2 つの問題を考慮する必要がある．第一に，知識基盤に定義された公式の地名と，史料中に出現する実際の地名との間に，表記揺れなどの原因で不一致が生じるという問題がある．第二に，同一あるいは類似する地名が複数存在する場合に，その中のどれに対応するかを決定する必要がある．こうした作業を効率的に行うためのアノテーション基盤の構築は歴史ビッグデータの重要な課題である．

歴史ビッグデータのアプローチと従来の歴史研究で用いられてきた歴史 GIS [1]のアプローチの大きな違いは地名もしくは識別子の扱いにある．歴史 GIS では，史料に出現する地名を地理情報システム（GIS）に登録し，それを地図上に可視化することで空間的な分布の分析を可能とした．このアプローチでは緯度経度が必須情報となるため，緯度経度を持たないまたは確定させることが難しい地名を扱うことが困難という問題があった．そこで本論文は，地名を軸に各種の情報を統合する地名情報基盤[2]のアプローチを用いる．ここでの地名とは，名称や緯度経度などを属性としてもつ識別子のことを指すが，緯度経度は必須項目でないため，歴史的地名を扱いやすいという利点がある．このように識別子を軸とした情報統合は，Linked Data が目指す世界でもある．

このようなアプローチの背後には，データ構造化とデータ分析・可視化を疎結合にするという考え方がある．従来のデータセット構築では，構造化と分析・可視化が密結合していたため，特定の目的には使えるが再利用性が低いデータセットが生み出される傾向があった．例えば歴史地名を緯度経度に変換する場合，その方法が目的によって様々に異なる場合がある．したがっていきなり緯度経度に変換してしまうよりも，まずは地名の識別子で情報を統合し，目的に合わせて変換方法を使い分けた方が，データの再利用性は高くなる．このように，地名を軸としてデータ構造化と分析・可視化を分離し，特定の目的に合わせたデータを基本データの派生データとして構築することが重要である．歴史ビッグデータでは，同様の疎結合の考え方を，空間だけでなく時間や人物，イベントなど，様々なデータに一般化していくことを長期的な目標としている．

こうした識別子に関する著者らのこれまでの取り組みを紹介する．まず「歴史的行政区域データセットβ版」[2]は，1920 年以降の多くの市区町村に識別子を付与して活用可能とした．また GeoLOD [3]はさらに分野横断的に地名の識別子を付与する取り組みであり，地名辞書の構築や地名の登録を通して，個別の地名に識別子を付与することができる．さらに GeoNLP は，GeoLOD で識別子を付与した地名辞書を用いて，自然言語文から地名を自動的に抽出して曖昧性を解消するエンティティリンキングのソフトウェアである．

図 2 江戸切絵図の「御江戸大名小路絵図」出典：国立国会諸図書館デジタルコレクション．

然言語処理の分野においても様々な研究がある．著者らも，現代文を対象とした地名固有表現認識のためのソフトウェア GeoNLP の研究を進めている．しかし史料の場合はそもそもテキスト化されていない場合が多く，テキスト化されていても形態素解析やその他の言語処理ツールにおいて十分な対応が進んでいない現状がある．そこで本論文では GeoNLP のような自動的なアプローチではなく，手動を中心に進めていくこととする．

こうした考え方のもと，本論文では江戸を対象とした歴史ビッグデータにおける空間情報の現状と様々な試みを紹介する．まず第 3 章では江戸に関する地名リソースの収集を紹介し，第 4 章では江戸に関する地理的資料を紹介する．続いて第 5 章では古地図のジオレファレンスとして，古地図上の地名リソースに対して現代の緯度経度を付与する作業を議論する．そして第 6 章では地理的資料を地名リソースにエンティティリンキングした結果を紹介し，最後に第 7 章ではこれらの技術に基づく歴史的記録の活用をまとめる．

3．江戸に関する地名リソース

3.1 江戸切絵図

「江戸切絵図」とは江戸時代に作成された江戸の分割地図である[4]. 江戸切絵図には多くの種類があるが, 本論文の対象は「尾張屋版」(1849-1862) と呼ばれる江戸切絵図である．これは全部で 32 枚から構成されるが，うち 3 枚は他の地図が対象とする領域の拡大版であるため，本論文は残り 29 枚を対象とする．これらはすべて国立国会図書館デジタルコレクションで IIIF により公開されているため，これを本研究では利用する．このうちすでに 28 枚分の作業を完了しており，残る「小石川絵図」も間もなく作業を完了する予定である．そして 28 枚分の作業成果として，8354 か所の地名を構造化したデータベースを公開した．

ここで構造化の対象とする地名は，住所としての町名等だけでなく，「広義の地名」としての施設名[2]も含むものとする．このような施設名は現

代では Point of Interest (POI)とも呼ばれる．本論文は POI も含め，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 カテゴリに地名を分類して収集した．なお屋敷地には大名屋敷に加えて旗本屋敷などもあり，後者は将来的には後述の「武鑑」と接続できる可能性があるものの，件数が多いことからその構造化は将来課題に回すこととした．

地名の構造化には IIIF Curation Platform [5,6]を利用した．まず江戸切絵図を IIIF Curation Viewer で開き，地名文字列の領域を四角で囲んでキュレーションに登録し，そこに翻字などのメタデータを追加した．この作業を地図全体にわたって行うことで，IIIF のキャンバス座標を含む構造化データを作成した．なおメタデータについては，異体字・旧字の常用漢字への変換，「◯◯丁」→「◯◯町」等の表記の統一，「◯◯イナリ」→「◯◯稲荷」等の平仮名・片仮名の漢字への変換など，基本的なクリーニング作業によってデータを整理し，整理後の地名を主に活用した．

またデータ公開にも IIIF Curation Platform を活用した．江戸切絵図の地図表示を想定し，IIIF Curation Viewer にアノテーションビューモードを追加し，IIIF 画像上に地図マーカーなどを表示可能とした．このような準備を経て，「江戸マップβ版」ウェブサイトを 2019 年 11 月に公開した[7]. 江戸切絵図の 1 枚である「御江戸大名小路絵図」を表示した例を図 2 に示す．IIIF Curation Viewer 上に地名のマーカーが表示できており，マーカーをクリックすると個別の地名ページへのリンクが表示される．

3.2 歴史地名データ

人間文化研究機構と H-GIS 研究会が公開する歴史地名データは，大日本地名辞書（明治 33 年初版）から 53,528 件，延喜式神名帳（延長 5 年（927 年）編纂）から 2,842 社，旧 5 万分の 1 地形図（明治 29 年から昭和 10 年に測量）から 242,544 件の地名を収集し，それぞれの地名に識別子と緯度経度を付与した地名辞書である[8]. その意味でこのデータは，先述した地名リソースの条件を満たしており，全国をカバーしている点も優れている．一方，上記の年代からもわかるように明治以降の時代に収集した地名が中心になっているため，江戸時代の地名を扱う際にどのくらい使えるかは明らかではない．また地名の粒度としても，都市域における細かい町名はあまり含んでいないという特徴がある．

この歴史地名データを活用するために，まずウェブ地図として使いやすい「歴史地名マップ」を公開した[9]. これは図 3 に示すように，バイナリベクトルタイル技術を活用することで，全国 298,914 件の地名を同時に表示しながらズームイン・アウトができるようになっている．次に歴史地名データを地名辞書形式に変換し，GeoLOD に登録することで，GeoLOD ID 識別子も活用できるようにした．こうして，地名リソースとしての活用と，アノテーションを支援する可視化の両方を用意して，活用のための準備を整えた．

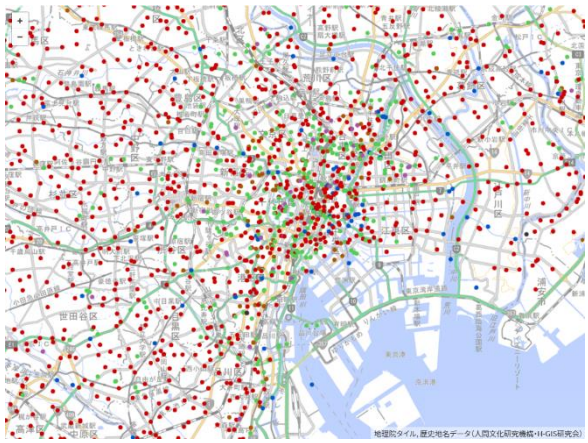

図 3 歴史地名マップの東京付近の拡大図.

4． 江戸に関する地理的史料

本章では，以下の 3 種類の地理的史料を対象とし，史料中に出現する地名のリストを整備する．

4.1 武鑑

武鑑とは，江戸時代に出版された大名家および幕府役人の名鑑であり，大名家や幕府役人に関するデータが項目別に詳細に記録されている[10]. 本論文では『寛政武鑑』(1789)のデータを用いた．武鑑には，居城地，上屋敷，菩提寺などの地理情報が地名として記されている．そこでこれらを江戸切絵図の地名 ID や，現代地図の緯度経度に手動でマッピングした．また IIIF Curation Viewer を用いて，江戸切絵図上で上屋敷または菩提寺に対応する領域を囲んだキュレーションも作成，江戸切絵図上のキャンバス座標も取得した．さらに菩提寺については，駒沢学園が公開する「寺院資料データベース」[11]を寺院に関する典拠として活用し，地理情報はそちらも参照できるようにした．その成果は「武鑑全集」で公開している[12].

4.2 観光案内

江戸時代の各世紀から 2 種類ずつ「名所記」「名所案内」をピックアップし，『江戸名所記』（1662）『江戸名所百人一首』（1663）『絵本江都の見図』（1795）『絵本江戸桜』（1795）『江戸名所図会』（1834-1836）『絵本江戸土産』（1850-1867）の計 6 種類 21 冊を対象に，IIIF Curation Viewer を用いて挿絵を切り抜いた．そして，これらの挿絵に書かれた地名を翻刻し，その他のメタデータとともに整備した[13]. さらに地名と歴史地名データの ID や江戸切絵図の ID との紐づけも手動で行った．その成果は「江戸観光案内」で公開している[14].

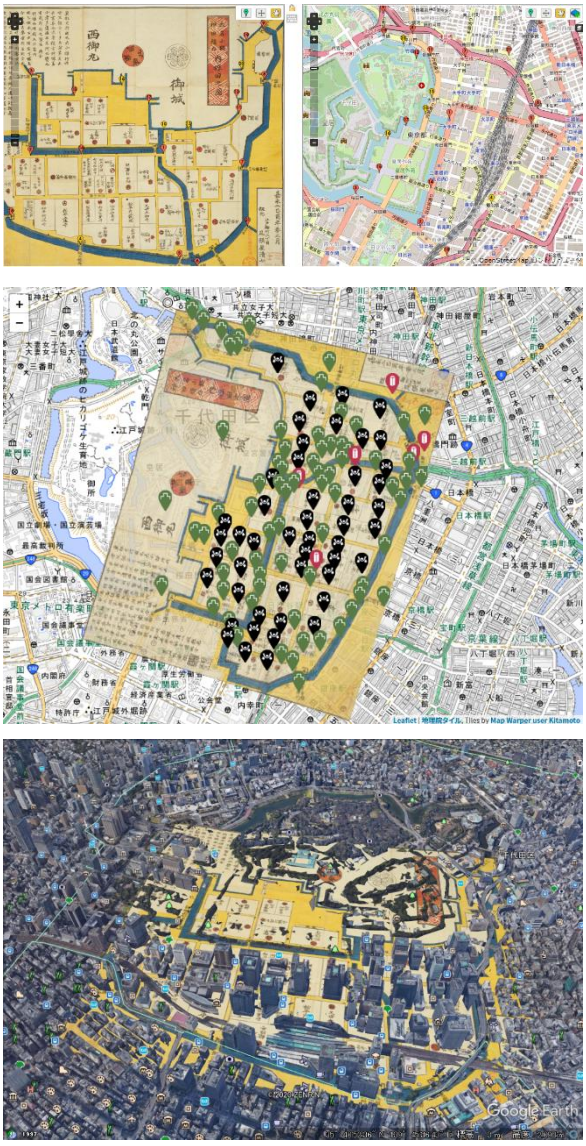

図 4 江戸切絵図の「御江戸大名小路絵図」のジオレファレンス．上から，日本版 Map Warper を用いたジオレファレンスの作業画面，ジオレファレンス結果の地理院地図上のタイル表示，Google Earth 上のタイル表示を示す．

4.3 買物案内

『江戸買物独案内』(1824)は，江戸（一部は他地方を含む）の商店・職人を業種別に一覧化した出版物である[15]．商人名や居所だけでなく業種や紋などのグラフィカルな要素も含むが，うち文字情報については国立歴史民俗博物館の「江戸商人・職人データベース」がすでに整備されているが，居所の現在地は現在の区レベルでしか把握できないという問題がある．そこで，より詳細な情報を含む江戸切絵図とリンクする計画であるが，本論文ではその予備的な調査として居所と江戸切絵図との自動マッチングの例を紹介する．その成果は「江戸買物案内」で公開している[16]．

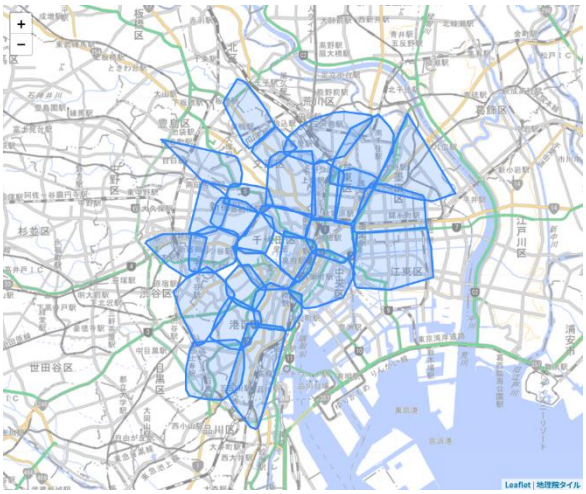

図 5 江戸切絵図 28 枚の地図ごとの制御点の外接多角形を可視化した地図．それぞれの地図の大まかな範囲を示す．

4．古地図からの地名リソース構築

4.1 古地図と地名の関係

地名リソースを構築する場合，一般的に地名は古地図から収集する．本研究でも江戸の地名は江戸切絵図から収集する．このとき地名の位置として，古地図上の画像座標は容易に取得できるが，これを現代の座標系としての緯度経度にどう変換するかが次の課題となる．最も直接的なのは，歴史地名を現代のジオコーディング技術によって緯度経度に直接変換する方法である．例えば，Google マップに地名を入力してマーカーが立てば，その緯度経度を読み取ればよい．しかし現代の位置情報サービスは現代地名が主な対象のため，歴史地名の検索には難がある．そこで本論文ではこの問題を以下の方法で解決する．

第一が，すでに別の方法で緯度経度が推定された地名リソースとリンクする方法である．例えば「歴史地名データ」はすでに緯度経度が付与されているため，古地図上の地名を歴史地名データの識別子に紐づければ緯度経度を付与できる．ただし，地名の表記が文字列レベルで完全一致するとは限らず，かなと漢字の変換や，表記揺れなどの修正も必要となる．また歴史地名データに複数の同名地名が存在する場合，地図で緯度経度を確認しないと，誤った位置に対応付けする危険がある．

第二が，古地図全体を現代地図上に重ねるジオレファレンス処理を行い，地名の緯度経度を丸ごと推定する方法である．これは古地図上の画像座標と緯度経度座標との対応関係を近似する数学的な変換式を推定することが課題となる．ただし江戸切絵図のように測量成果に基づかない地図は，地図の歪みが不均等であるため，通常のジオレファレンス処理では精度に限界がある．さらに精度を高めるには，推定された緯度経度を個別に修正していく必要も生じる．

4.2 江戸切絵図のジオレファレンス

古地図のジオレファレンスについては様々な手法が提案されている．例えば北本らは，北京古地図のジオレファレンスのために，基準点と基準線を用いた新手法を提案し，203 ページ 290 億画素に及ぶ巨大な地図をジオレファレンスした[17]. これは古地図を変形させて現代地図に重ねる方法であり，地図全体の位置合わせが可能という利点があるが，地図中の文字や地物形状も歪んでしまうため，史料としての活用が困難になるという欠点がある．一方，北本らはシルクロードの古地図を対象に，古地図を変形させずに一点のみを位置合わせするインタラクティブな位置合わせツール「マッピニング」も開発した[18]．これは史料の可読性が高いという利点はあるが，位置合わせした以外の点の緯度経度を推定することはできない.

こうした様々な手法がある中で，本研究では簡便な位置合わせを行うツールとして，立命館大学が公開する「日本版 Map Warper」[19]を活用した．このウェブサービスは，内部的には GDAL（Geospatial Data Abstraction Library）の gdalwarp プログラムを活用し，古地図と現代地図の制御点から地図の変形を計算する．制御点は全体にわたってまんべんなく指定する必要があり，我々の例ではおおむね 20 点前後を指定した．このように指定すると，地図の切り抜きと変形をブラウザ上で行うだけでなく，変形後の地図データをタイル配信することも可能なため，古地図の公開基盤としても活用することができる.

このツールを活用して江戸切絵図をジオレファレンスした結果，またその結果得られる地図タイルを地理院地図および Google Earth 上に表示した結果を図 4 に示す．現代地図と重ね合わせることにより，その地図がどこを示しているのかを理解しやすくなることがわかる．さらに，地図ごとにカバーしている範囲を示すため，ジオレファレンスに活用した制御点の外接多角形（convex hull）を計算して表示したのが図 5 である（作業中の小石川絵図は除く）．これを見ると，江戸切絵図が，東は現代の山手線の内側，西は江東区から墨田区の付近までをカバーしていることがわかる．ただし，これらは制御点に用いた範囲であって，地図自体はそれよりも外側に広がっている．江戸切絵図は初期に出版された江戸の中心付近は比較的距離と方位を守って描かれているものの，後期になるにしたがって周辺部に移り，そこでは距離や方位を犠牲にしてでもその地区にある観光名所の情報を入れることに重点を置いていることがよくわかる．したがって後期になるほど位置合わせは困難となり，その地図は道案内のための位相的地図と考えた方がよいことになる.

こうした古地図の位置合わせに関する問題点は，人権問題などへの十分な配慮なしに被差別部落などの地名が現代の地図上に重ね合わされることで，偏見や差別を肯定・容認する方向に活用されてしまう点にある．江戸切絵図についても，過去にそうした問題が発生している[20]. ただし，問題の存在自体を隠ぺいすべきではなく，むしろ歴史に関する正しい理解を促すための十分な解説とセットで公開すべきであるとの意見がある．我々もその方向で公開の準備を進めている.

５．歴史地名のエンティティリンキング

5.1 エンティティリンキング

地名リソースの構築に続いて，エンティティリンキングの課題として，本論文では地名抽出・アノテーションを以下の 2 つの方法を併用して進めていく．第一が，史料から地名を人手で抽出した上で，地名リソースの地名と一つ一つ照合して識別子を付与する方法である．これは地名の文字列が一致しない場合でも，人間が補助資料を参照して考証すればリンクできるという利点があるが，コストは大きくなる．第二が，史料から地名を抜き出し，クリーニングして史料ごとの地名集を構築し，地名集と地名リソースとをプログラムで一度に照合する方法である．史料ごとの地名集を構築する際には，史料内での表記ゆれを統一する校訂は必要だが，全体の照合はプログラムで行うためコストが低いという利点がある．そこで本論文では，武鑑と観光案内には第一の方法，買物案内には第二の方法を用いることとした.

5.2 武鑑の結果

武鑑の地名情報である上屋敷と菩提寺が，地名リソースとどのくらいリンクできるかを検証した．ただしこれらは一般的に施設名に含まれるものであり，施設名は表記揺れが比較的激しいことから，人間が多種多様な資料を参照しながら慎重にリンクを進めていった.

まず上屋敷については，大名家 264 件中現在地にマッピングできたのは 261 件，またこのすべてを江戸切絵図の地名とリンクできた．上屋敷は江戸の重要施設であるため，江戸切絵図でも重点的に記載されている．ただし江戸切絵図の作成は，『寛政武鑑』よりも 60 年以上後であるため，その期間にはいくつかの上屋敷が移転していることもわかった．武鑑中に記載された住所と，江戸切絵図から推定される場所とが異なるものは，264 件中 24 件と 1 割弱である．歴史的史料は時間方向に密に存在するわけではないため，時間のずれをどう吸収するかは常に問題になる.

次に菩提寺については，典拠となる駒沢学園の「寺院資料データベース」へのリンクは 246 件，一方江戸切絵図への接続は 242 件となった．菩提寺の中には江戸切絵図の外側に位置するものがあること，また移転等によって追跡しづらい場合があることなどを考えると，上屋敷より件数が少ないのは妥当な結果である．ただし 9 割以上

図 6 江戸観光案内に含まれる地名を現代の地図上に可視化した結果.

の菩提寺が江戸切絵図とリンクできたということは，寺院という施設も江戸切絵図において重要であることを伺わせる結果となった.

5.3 江戸観光案内の結果

江戸観光案内は 1255 件の挿絵を収集しており，その中で地名が明確に記されている挿絵 1219 件を対象に，地名リソースにどのくらいリンクできるかを検証した.

まず「歴史地名データ」とリンクできたのは 765 件と，全体の 6 割強に達した．先述のように「歴史地名データ」は明治時代に収集された地名が多く含まれるにも関わらず，多くの地名がリンクできたということは，観光名所のような POI については，江戸時代と明治時代との間に大きな変化がなかった可能性がある．一方，江戸切絵図とのリンクについてはまだ検証が終わっていない．ただし江戸切絵図の用途として観光は重要であり，特に後期においては距離と方位を歪めてでも観光名所を地図内に取り入れる編集方針が明確であるため，今後は「江戸切絵図」とのリンクを通して江戸観光ビッグデータを構築したいと考えている.

なお図 6 には，江戸観光案内を現代の地図にマッピングした結果を示す．江戸を中心に関東一円に観光名所が散在していることがわかる.

5.4 江戸買物案内の結果

江戸観光案内とは異なり，江戸買物案内の居所は○○町などの行政地名や，○○通りや○○橋などの交通に関する地名で示されることが多い．この種の地名は歴史地名データにあまり含まれていないため，江戸切絵図とのリンクを試みた．また買物案内の居所が 2454 件と多いため，人手による精度の高いリンクの前に，文字列マッチングアルゴリズムによる地名の相互比較を行うこととした.

文字列マッチングを適用する前に，データのクリーニングが必要である．具体的には，買物案内に出現する地名の表記をある程度統一した地名集を構築した．ただし買物案内は広告であることから，居所の表現も広告主にゆだねられていると考えられる．客を引き寄せるためによりわかりやすい表現として，地名の前後にランドマーク的な名称を合わせて書いている場合がある．こうした状況を踏まえると，文字列マッチングにおいては完全一致だけでなく，前方や後方部分のみが一致する前方一致や後方一致，さらには居所の文字列の一部に地名リソースの地名を含む部分一致などのパターンを調べていく必要がある．そこで本論文では以下の文字列マッチングアルゴリズムをこの順番で適用することにより，どの程度の居所が地名リソースとリンクしうるかを検証することとした．以下に各種の文字列一致の例を示す.

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**）

表 1 買物案内と江戸切絵図の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	-	-

表 1 は文字列マッチングの結果を示す．まず完全一致は 956 件と全体の 4 割弱を占めた．ただし，同名の地名が複数存在する曖昧性の問題が検出された地名も 168+36=204 件あり，これらをエンティティリンキングするには曖昧性を解消してどれか一つの地名を選ぶ必要がある．ここで重要となるのが地図内か地図間の曖昧性かである．地図内の曖昧性は，同一町名が道の両側に存在する「両側町」などで発生する場合が多いため，リンク先としてどれを選んでも実用上の大きな違いは生じない．ゆえに実用上の問題が生じるのは地図間の曖昧性（36 件）である．これは一つ一つ確認する必要があるが，完全一致の中では 4%弱を占めるだけであり，大きな問題にはならない．これに対し，前方一致や後方一致，部分一致では，地図間の曖昧性の割合が 4 割程度と高く，この場合は曖昧性解消において人間の判断がより重要性を増す．とはいえ，部分一致を含めて江戸切絵図の地名と一致した居所は 2062 件と全体の 84%に達しており，買物案内と江戸切絵図は

多くの地点でリンクできる可能性が高い．

一方，江戸切絵図と一致しない地名は 392 件であり，その原因は今後調査する必要がある．両者の発行年が 25 年以上ずれていることによる町名変更などの影響も考えられるものの，異体字への対応が不十分だったり，地図には書かれていない有名施設だったりする場合も多いため，人間による作業でリンクできる可能性が高い．

5.5 考察

以上の結果により，歴史地名データと江戸切絵図が対象とする地名は大きく異なることがわかった．両者の地名が完全に一致するのは 687 件と，江戸切絵図の地名 8354 件のうち 8%程度に過ぎない．逆に言えば，両者は補完的な地名リソースとして使えるとも考えらえる．観光案内は歴史地名データで 6 割ほどリンクできたことから，観光名所のような POI には歴史地名データが比較的強いと考えられる．一方，買物案内は江戸切絵図が部分一致を含めて 8 割ほどカバーできたことから，生活に密着した地名は江戸切絵図の方が適している．

このように各種の地名リソースがそれぞれの特色を持つとすると，それらをすべて統合した一つの地名リソースを構築する方が便利である．その役割を果たすのが GeoLOD である．GeoLOD が付与する統一的な識別子を史料中の地名表現に付与することで，分野や目的を超えて，地理的なデータを共有することが可能になる．これが歴史ビッグデータの考える疎結合性と Linked Data の実現につながる．

6．江戸ビッグデータに向けて

6.1 歴史的行動記録

江戸に関する地理的史料は，この他にもまだまだたくさん残されている．こうした史料から過去の世界に関する知見を得るためには，史料に残された歴史的記録を構造化し，収集し，統合的に分析する必要がある．歴史的記録には様々な種類がありうるが，我々は大別して「歴史的状況記録」「歴史的行動記録」「歴史的状態記録」の 3 種を想定している[21]．これらを歴史的な記述の最小単位としつつ，目的に応じて組み替えることで，歴史の新たなストーリーが見出せるようになる．

本論文ではこのうち「歴史的行動記録」に焦点を合わせる．これは現代のビッグデータ分析における OD（origin-destination）データに相当するものである．OD データでは，(1)どこからどこへ，(2)いつ，(3)誰が，(4)どのような手段で，(5)どのような目的で移動したかなどが最小単位の記録となる．そして，多数の人々の動きの記録を収集して統合的に分析することで，人々の移動パターンの分析などを行う．以下，この記録が持つべき属性について検討する．

まず「どこからどこへ」については，現代では GPS などを用いて緯度経度単位で精密に測定することが可能である．ただしビッグデータ分析においては，精密なデータはプライバシーにも関係するため，意図的に精度を落とすことがある．すなわち地球表面をある程度の広さを持つゾーンに分割し，人々の移動をゾーン ID 間の移動としてモデル化する．このことを踏まえると，歴史地名の粒度が荒いとしても，それ自体が問題というわけではないことがわかる．次に「いつ」については，現代では UTC に基づく精密な記録が可能であるが，史料では和暦に基づく大雑把な記録となる場合が多い．しかし HuTime[22]を用いて和暦を西暦に変換することで，日単位では問題なく扱うことができる．「誰が」については，理想的には人物識別子を用いるべきであるが，通常は文字列で人物を表現することになる．「どのような手段で」については，通常は徒歩や馬，船などが使われるだろう．最後に「どのような目的」については類型化が難しいが，研究の問いに最も密接にかかわる属性でもあり，ここは目的に応じて付加していく仕組みが必要となる．

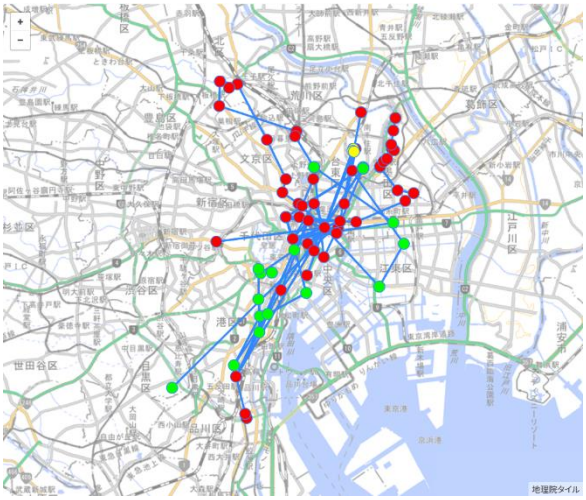

図 7 『西遊草』の歴史的行動記録を現代の地図上に可視化した結果．

6.2 『西遊草』の分析例

こうした歴史的行動記録の有効性を確かめるため，本論文では清河八郎著『西遊草』を用いる．『西遊草』11 巻 8 冊は，清河八郎が安政 2 年（1855）母を奉じて西遊したときの日記である[23]．八郎の郷里である山形県を出発し，母とともに伊勢詣でをした後，多くの土地を巡って半年後に郷里に戻った．江戸には安政 2 年 7 月 26 日に入り，8 月 29 日に出立するまで約 1 か月滞在している．清河とその母の訪問した地名を抽出し，地名リソースや地理的資料から取り出した情報とリンクすることで，歴史的行動記録を分析した例を示す．

今回利用した地名リソースは江戸切絵図，地理的資料は江戸観光案内である．後者は歴史地名データとの紐づけも行っているため，マッチした

地名には信頼性の高い緯度経度情報を与えることもできる．ただし抽出した地名 164 件にヒットしたのが，江戸切絵図が 132 件に対し，観光案内は 119 件であり，江戸切絵図の方が使える地名が多かった．観光に特化すれば観光案内・歴史地名データの方が強いものの，清河八郎の行動パターンでは家などを訪問することも多く，その場合は江戸切絵図の町名の方が有効なためである．また藩邸などの同定には武鑑全集も有効に活用できた．一方，リンクできない地名には，地図作成時点には存在しない地名（「御台場」等）や，通称等で地図には出てこない地名（「於玉池」等）などがあった．

その結果を示すのが図 7 である．赤点は八郎，緑点は母，そして黄点は八郎と母が訪れた地点である．二人は江戸を縦横に移動し，様々な場所を訪問していることがわかる．また両者は共に吉原を訪ねており，それ以外にも母は目黒や深川などかなり遠くまで足を延ばしていることがわかる．一方，八郎は観光の動きだけでなく，上屋敷などの仕事に関係する動きも多く，母とは異なる行動パターンを示している．このような歴史的行動記録を様々な史料から構造化すれば，江戸観光の典型的なルートや人々の買物行動などが分析できる可能性があると考えられる．

7．おわりに

本論文は，歴史地名に関する基盤の構築，および地理的資料とのリンクに基づく，江戸ビッグデータの構築に向けた試みを紹介した．今後はさらに多くの史料を構造化しつつ，データ駆動型のアプローチで過去の世界を探る研究をさらに広げるとともに，過去の世界を現在に引き寄せるために過去のデータを現代のプラットフォームに載せる研究も展開していく計画である．

参考文献

[1] HGIS 研究協議会編，歴史 GIS の地平，勉誠出版, 2012.
[2] 北本 朝展，村田 健史, “歴史的行政区域データセットβ版をはじめとする地名情報基盤の構築と歴史ビッグデータへの活用”，情報処理学会技術報告, Vol. 2020-CH-124, No. 1, pp. 1-8, 2020.
[3] GeoLOD, https://geolod.ex.nii.ac.jp/, (参照 2020-11-08).
[4] 俵 元昭, “江戸の地図屋さん 販売競争の舞台裏”，吉川弘文館, 2003.
[5] 北本 朝展，本間 淳, Tarek SAIER, “IIIF Curation Platform：利用者主導の画像共有を支援するオープンな次世代 IIIF 基盤”，人文科学とコンピュータシンポジウム じんもんこん 2018 論文集, pp. 327-334, 2018.
[6] 北本 朝展, “オープンな画像の利活用を開拓する IIIF Curation Platform”，カレントアウェアネス, No. E2301, 2020.
[7] 江戸マップβ版, http://codh.rois.ac.jp/edo-maps/ (参照 2020-11-08).
[8] 関野 樹，原 正一郎, “デジタル歴史地名辞書の公開とその活用”，研究報告人文科学とコンピュータ（CH）, Vol. 2018-CH-118, p. 1-4. 2018.
[9] 歴史地名マップ, http://codh.rois.ac.jp/historical-gis/nihu-map/, (参照 2020-11-08).
[10] 藤實 久美子, “江戸の武家名鑑 武鑑と出版競争”,吉川弘文館, 2008.
[11] 駒沢学園寺院資料研究センター, “寺院資料データベース”, http://jiin-shiryo.komajo.ac.jp/, (参照 2020-11-08).
[12] 北本 朝展, “人物データの分析――江戸時代のデータブック「武鑑」の構造化と歴史ビッグデータ解析――”，電子情報通信学会誌, Vol. 102, No. 6, pp. 569-571, doi:10.20676/00000350, 2019.
[13] Chikahiko Suzuki, Asanobu KITAMOTO, “Pre-modern Japanese Books as Data of Humanities: Finding Image of Edo Famous Place from Meisho-Ki 名所記 and Meisho-Zue 名所図会 using IIIF Curation Platform”, JADH2019, pp.42-45, 2019.
[14] 江戸観光案内. http://codh.rois.ac.jp/edo-spots/, 公開準備中.
[15] 鈴木 親彦，北本 朝展, “IIIF Curation Platform による『江戸買物独案内』のマイクロコンテンツ化:非文字情報を軸に”，人文科学とコンピュータシンポジウム じんもんこん 2019 論文集, pp. 11-18, 2019.
[16] 江戸買物案内. http://codh.rois.ac.jp/edo-shops/, 公開準備中.
[17] 西村 陽子，北本 朝展, “Google Earth と『乾隆京城全図』を用いた北京歴史空間の情報基盤”，人文科学とコンピュータシンポジウム じんもんこん 2008 論文集, pp. 81-88, 2008.
[18] 西村 陽子，北本 朝展, “ディジタル史料批判と歴史学における新発見”，人工知能学会誌, Vol. 31, No. 6, pp. 769-774, 2016.
[19] 矢野桂司，鎌田遼, “日本版 Map Warper の構築と活用”，地理情報システム学会講演論文集, 2017.
[20] 石松 久幸, “カリフォルニア大学バークリー校における日本古地図のデジタル化プロジェクトについて”，情報の科学と技術, Vol. 59, No. 11, pp. 557-562, 2009.
[21] 市野 美夏，橋本 幸恵，平野 淳平，増田 耕一，北本 朝展, “目撃情報の収集による歴史的状況記録パスファインダーの構築”，人文科学とコンピュータシンポジウム じんもんこん 2018 論文集, pp. 343-350, 2018.
[22] 関野 樹, “時間名による時間参照基盤の構築－Linked Data を用いた期間の記述とリソース化”，じんもんこん 2019 論文集, pp. 267-272, 2019.
[23] 小山松勝一郎 編訳, “西遊草 清河八郎旅中記”，平凡社, 1969.