

2023年10月18日

歴史的地名の「行政区画変遷」を大規模オープンデータ化
～『日本歴史地名大系』を平凡社地図出版との協働により機械可読データとして強化～

情報・システム研究機構 データサイエンス共同利用基盤施設 人文学オープンデータ共同利用センター（以下 ROIS-DS CODH、センター長 北本朝展）と株式会社平凡社地図出版（代表取締役 西田裕一）は、『日本歴史地名大系』の機械可読データ化に向けた協働を推進し、このたび歴史的地名の「行政区画変遷」に関する大規模オープンデータを公開しました。

『日本歴史地名大系』（用語解説 1）とは、全国の歴史研究者の協力を得て地名研究・地域史研究の全成果を結集し、株式会社平凡社が25年（1979年～2004年）をかけて出版した50巻51冊の地名辞典です。この地名辞典の編集にかつて関わった人々の全面的な協力を得て、地名辞典の内容の更新や位置情報（緯度経度）の追加を進め、データ駆動型研究に適した機械可読データを新たに構築しました。出版社が長年にわたって蓄積してきた膨大な学術資源をオープンデータ化することで、研究者や一般市民など社会における利用が大きく進むことを期待します。

また ROIS-DS CODH は、このオープンデータの価値をさらに高めるため、データ駆動型研究で重要な役割を果たす地名識別子（用語解説 2）として、現代から江戸時代までの市区町村の変遷を反映した市区町村 ID を整備します。市区町村などの行政地名は、さまざまな行政文書や統計データ、歴史資料などに頻出する重要な地名ですが、公的な地名識別子として代表的な「全国地方公共団体コード」は1968年以降しか整備されていません。そこで ROIS-DS CODH は、『日本歴史地名大系』のオープンデータを活用して、江戸時代の郷帳に記録された藩政村まで遡及できる地名識別子の整備に取り組みます。

このような成果に基づき、ROIS-DS CODH では大規模歴史地名データベース（用語解説 3）「れきちめ：日本歴史地名統合データベース」を構築し、歴史地名を活用した各種の歴史データ統合に基づく「歴史ビッグデータ」の研究を推進します。

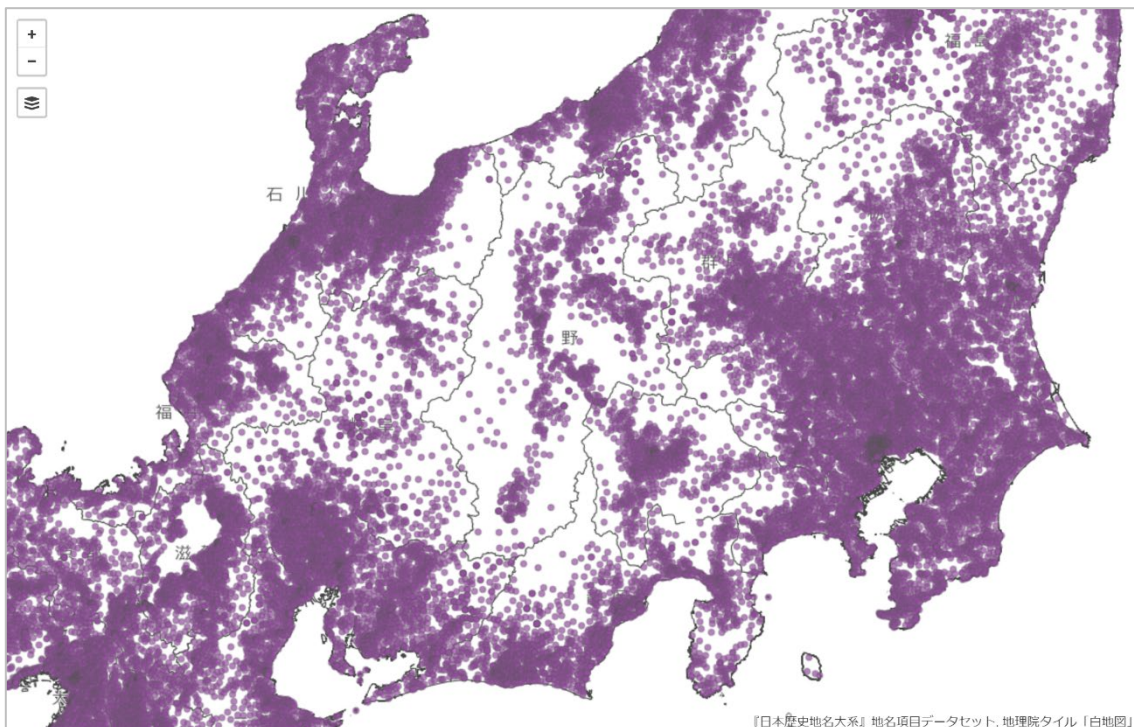


図 1：『日本歴史地名大系』地名項目データセットに含まれる地名（全国的な分布）

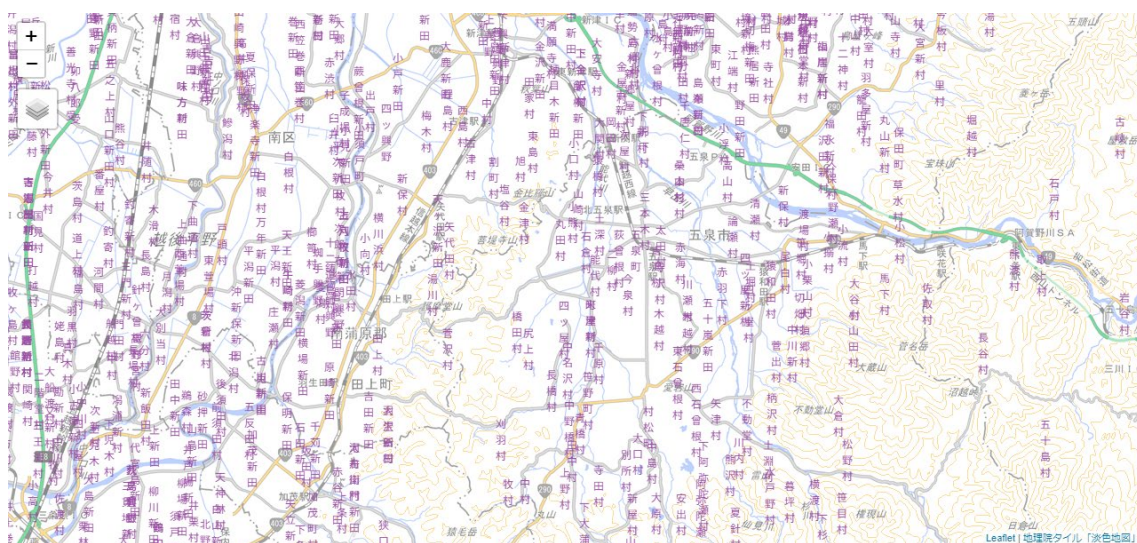


図 2：『日本歴史地名大系』地名項目データセットに含まれる地名（新潟県の拡大地図）

■新たに公開するデータセット

1. 『日本歴史地名大系』地名項目データセット（79,502 件）：

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

ジャパンナレッジ（用語解説 4）に収録された『日本歴史地名大系』の地名項目を対象に、ジャパンナレッジ項目 ID と地名、読み、緯度経度などを 1 レコードとするデータセットを公開します。特に緯度経度は、ジャパンナレッジ版『日本歴史地名大系』では一部の地名にしか付与されていませんが、今回の公

開に合わせて多くの地名の緯度経度を推定し、データセットを強化しました。ただし緯度経度の推定精度には改善の余地があるため、今後も各種資料を精査しながら精度を向上させる計画です。今回は、全国のデータセットを公開します。

2. 『日本歴史地名大系』行政地名変遷データセット (3,611件) :

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

一般社団法人百科総合リサーチ・センターによってデジタル化された『日本歴史地名大系』の付録「行政区画変遷表」をもとに作成したデータセットです。後述の「歴史的行政区域データセットβ版」は1889年以降の行政地名を対象としますが、本データセットはこれを1889年以前の明治期から17世紀の江戸期まで遡及します。1889年とは、現代の行政区域の骨格が成立した市制・町村制が施行された年で、この年を境界に行政地名のデータセットを2つに分割します。ただし現代から江戸時代までを通して使えるように、2つのデータセットを合わせて表記揺れや連続性を調査して市区町村IDを付与します。今回は、栃木県と群馬県のデータセットを先行公開しますが、今後は全国に拡大する計画です。

■連携して更新するデータセット

1. 歴史的行政区域データセットβ版 (16,857件) :

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

日本の行政区域（市区町村）に市区町村IDを付与し、境界（ポリゴン）データを紐づけるなど、行政区域の地理情報を統合したデータベースです。国土交通省「国土数値情報」と筑波大学「行政界変遷データベース」を中心に複数のデータセットを統合した成果です。2017年にオープンデータとして公開した後も更新を続けており、現在は1920年～2023年のデータセットを公開しています。また、上記『日本歴史地名大系』行政地名変遷データセットを活用し、栃木県と群馬県については1889年～1919年のデータセットも先行公開しました。今後は全国に拡大する計画です。

上記データセットの概要については、「れきちめ：日本歴史地名統合データベース」をご覧ください。

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

■今後の展開

このたび公開したデータセットは初期版です。今後、『日本歴史地名大系』地名項目データセットについては地名の緯度経度の精度を高めるとともに、『日本歴史地名大系』行政地名変遷データセットについては全国の都道府県に対象を拡大する作業を進めます。その過程では、研究コミュニティや市民のニーズを収集してデータセットの有用性を高めるとともに、日本学術会議や学会、デジタル庁、民間企業などが推進する地名のデジタル化や標準化の活動とも連携したいと考えています。

■用語解説

1. 日本歴史地名大系

『日本歴史地名大系』は、全国の歴史研究者の協力を得て編纂され、地名研究・地域史研究の全成果を結集した地名辞典です。1979年から2004年にかけて平凡社から50巻51冊の地名辞典として出版されました。日本列島47都道府県+京都市の15万におよぶ地名項目に加えて、文献解題や地図類、行政区画変遷・石高一覧などの資料も付帯しています。今回のオープンデータでは、行政区画変遷を初めて機械可読データとして公開するとともに、行政区画変遷に関連する地名項目約8万件の位置情報（緯度経度）など、元の辞典にはなかった情報も整備して公開します。

2. 地名識別子

実世界で識別可能なエンティティ（実体）に付与した固有の文字列を識別子と呼びます。英語の Identifier を省略した「ID」や、数字を用いた「コード」などの表記もよく使われます。識別子は社会のデジタルトランスフォーメーション（DX）を推進する上で鍵を握る仕組みです。例えば、マイナンバーは日本の住民に付番する一意の番号ですが、これを利用すれば複数のデータベースに登録された記録を紐づけて扱うことができ、利便性が向上すると言われています。行政区域についても、1968年以降の市区町村には「全国地方公共団体コード」が付番されていますが、それ以前の市区町村は対象となっておりません。そこで ROIS-DS CODH では過去の市区町村にも識別子を付与し、様々な文書に出現する記述を地名識別子に紐づけ、文書横断的に統合するという方法により、「歴史ビッグデータ」の研究 (<http://codh.rois.ac.jp/historical-big-data/>) を推進していきます。

3. 歴史地名データベース

現代に続く地名だけでなく、いまや使われなくなった過去の地名なども含め、歴史的に存在した地名を収集した成果をまとめたものが歴史地名データベースです。日本の歴史地名に関しては、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」(https://www.nihu.jp/ja/database/source_map) が、298,914件の地名を含む大規模なデータセットとして広く使われています。今回公開するデータセットは「歴史地名データ」とは地名の収集方法が異なりますので、両者は補完的な役割を果たすことが期待できます。

4. ジャパンナレッジ

株式会社ネットアドバンスが運営するオンライン辞書・事典検索サイトです。『日本歴史地名大系』の全項目はジャパンナレッジ版『日本歴史地名大系』(<https://japanknowledge.com/contents/rekishi/>) に登録されているため、有料契約すればすべての情報が閲覧可能です。これに加えて今回のオープン化

により、基本的な情報はすべての人々が自由に利用可能となりますので、データ駆動型研究におけるデータ統合など、多様な学術的ニーズを持つ研究者に利用が拡大することが期待できます。またオープンデータに含まれるジャパンナレッジ ID を用いることで、外部の地名データベースからジャパンナレッジの本文を調べやすくなります。このように今回のオープン化は、データ駆動型研究の時代における出版社のオープン・クローズ戦略の立案にも資する試みだと考えています。

謝辞

- 『日本歴史地名大系』の「行政区画変遷」のデジタルデータは、株式会社小学館と株式会社平凡社によって設立された一般社団法人百科総合リサーチ・センターのメンバー（森田康夫・斎藤文雄・小島晃・成田広一・大澤克行）が作成しました。
- 「地名項目データセット」の緯度経度は、株式会社平凡社地図出版が整備しました。
- 地名の緯度経度の推定には、株式会社情報試作室の相良毅氏が開発する Python 版住所ジオコーダ `jageocoder` (<https://www.info-proto.com/jageocoder/>) によるジオコーディング結果、および人間文化研究機構と H-GIS 研究会の「歴史地名データ」の情報を参照しました。「歴史地名データ」を作成された故 桶谷猪久夫先生に深くお礼申し上げます。
- 「行政地名変遷データセット」の作成には、戸田ペリカン事務所の戸田智恵氏の協力を得ました。
- データセットの整備を進めるにあたり、その公開にご理解いただいた株式会社ネットアドバンスおよび株式会社平凡社に心より感謝申し上げます。

【本件に関する問い合わせ】

大学共同利用機関法人 情報・システム研究機構
本部広報室
TEL:03-6402-6211 E-mail : koho@rois.ac.jp

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所
総務部企画課 広報チーム
TEL:03-4212-2164 E-mail : media@nii.ac.jp