

デジタル歴史地名辞書の公開とその活用

関野 樹^{†1} 原 正一郎^{†2}

概要: 地名を経度緯度と対応付けるジオコーディングサービスが数多く提供されているものの、それらは現代の地名にしか対応していない。有志の研究団体である H-GIS 研究会の協力の下、2018 年 3 月に人間文化研究機構から公開されたデジタル歴史地名辞書は、歴史地名を緯度・経度と関連付けることを目的に構築されたものであり、今後の歴史研究などでの活用が期待される。

キーワード: 地名辞書, ジオコーディング, 地理情報システム, 地図, オープンデータ

Utilization of the Digital Historical Gazetteer Distributed from NIHU

TATSUKI SEKINO^{†1} SHOICHIRO HARA^{†2}

Abstract: Many geocoding services are provided and are available to associate place names with coordinate values indicated by latitude and longitude. These services, however, can be applied only for current place names. National Institutes for the Humanities (NIHU) started to distribute a dataset of digital historical gazetteer in March 2018, in cooperation with H-GIS Research Group. It is expected that the data contribute to many disciplines including history.

Keywords: Geocoding, Place Name, GIS, Map, Open Data

1. はじめに

日常の生活から政治・経済に至るさまざまな場面で、場所は地名により指定される。行政により階層的に定められた地名（いわゆる「住所」または「行政地名」）だけでなく、山・川・湖・海などの自然地形、寺・神社・港などの人工的な構造物、さらには、街道や鉄道などの地域を繋ぐ交通手段に由来するものなど、多様な地名が随所で用いられてきた。

一方で、地理的な位置を正確に指し示すには、緯度経度が用いられる。地図上にその位置を表示するだけでなく、二点間の距離や方位を計算したり、地理的な範囲同士の相対関係を検証したりするなど、解析的な処理を行うための基礎ともなる。近年、地理情報システム（GIS: Geographic Information System）が手軽に使えるようになり、さらに、Web 上で地図を表示するサービスも各社から提供されるようになると、緯度経度を必要とする場面も増えてきた。このため、地名により指定された場所を緯度経度に変換する仕組み（ジオコーディング）が Web フォーム[1]や Web API [2]として提供されている。これにより、主に行政地名を用いて、任意の場所の緯度経度を取得することが可能になった。ところが、これらは現代の地名のみを対象にしているため、歴史研究など過去の地名を扱う際に正しく適用できないという問題がある。

地名は、歴史的に変化しており、ある場所を指す地名が、消失、統廃合、分割・細分化されることがしばしばある。さらに、地名の表記が変化したり（文字の違いを含む）、地名が指し示す場所や属する上位の地名（階層構造）が変わったりすることもあり、古文書や古記録に現れる古い地名（歴史地名）を現代の地名を対象にしたジオコーディングのサービスに当てはめても、正しい緯度経度が得られるとは限らない。

このような問題に鑑み、有志の研究組織である H-GIS 研究会（Humanities-GIS Research Group）[3]の協力の下、人間文化研究機構から歴史地名辞書のデータが新たに公開された。本稿では、この歴史地名辞書データを実際の研究現場で利用する方法や注意点を中心に紹介する。

2. 歴史地名辞書の概要

この度公開された歴史地名辞書データ（以下、「本データ」）は、大阪国際大学の桶谷猪久夫氏らにより 2007 年ごろから構築が始まり、H-GIS 研究会のメンバーも加わりながら作業が続けられてきたものである。その後、歴史地名を扱うための基礎データとしての意義や重要性から、大学共同利用機関法人の一つである人間文化研究機構の研究資源共有化事業の一環としても位置付けられ、2018 年 3 月に同機構から公開された。

（歴史地名辞書データ 公開サイト）

http://www.nihu.jp/ja/publication/source_map

^{†1} 国際日本文化研究センター
International Research Center for Japanese Studies
^{†2} 京都大学東南アジア地域研究研究所
Center for Southeast Asian Studies, Kyoto University

データの元になった資料は、大日本地名辞書（吉田東伍著）[4]、式内社神名帳[5]、旧 5 万分の 1 地形図[6]の 3 つで、いずれも明治以前に編纂されたものである。このため、構築された地名辞書データは、概ね、江戸後期から明治の地名に対応できるものと考えられる。

データ構築は、これらの資料から地名を抽出し、それぞれの場所を現代または過去の地図を用いながら比定した上で、緯度経度と対応付けた。結果として、本データには全体で 298,914 件の地名データが含まれており、その内訳は、大日本地名辞書：53,528 件、式内社神名帳：2,842 件、旧 5 万分の 1 地形図：242,544 件である。

本稿では、歴史地名辞書のデータ構築にかかる詳細な手順については割愛する。これらは、それぞれの出典資料ごとに関連する文献を参照されたい（大日本地名辞書：[7][9][10][12]、式内社神名帳：[7][8][9][10][12]、旧 5 万分の 1 地形図：[10][11][12]）。なお、以下の点は、データ構築の過程で共通する問題であり、データの利用においても留意する必要がある。まず、複数の資料や地図から得られた地名が混在しているため、同じ地名に対し、緯度経度の異なる複数のレコードが存在する場合がある。また、緯度・経度の取得に際しては、作業上、相応の誤差が生じる。特に、一定の範囲を表す地名など、場所を一点に特定できない場合は、地図上の文字の位置で代表するなどしている。このため、緯度・経度は、必ずしも地図上の特定の地物の位置を指し示しているわけではない。

3. 歴史地名辞書データの構造

今回公開されたのは、歴史地名辞書のデータのみであり、検索などの機能は含んでいない。これは、本データの必要性や応用範囲の広さに鑑み、検索システムによる定まった機能のみを提供するよりも、さまざまな形でデータの利用を可能にすることを優先したためである。これにより、利用者が本データを自身のシステムに組み込んだり、データ解析の中間処理で本データを参照したりといった利用が実現する。本データ（構築途上のものを含む）を用いて実装された検索システム等は[13]などで事例が紹介されている。

本データは、保守や検索での利便性を考慮し、関係データベースのテーブルを CSV ファイルとして出力した 6 つのファイルで構成されている。これらのファイルの文字コードは UTF-8 である。図 1 は、本データに含まれるテーブル間の関係を示す。基本となるのは「地名」テーブルであり、単純なジオコーディングであれば、このテーブルのみで処理することができる。そのほかに、ヨミ、ローマ字表記、原資料での表記などを収容した「地名_別名」テーブル、地名が含まれる県や旧国名などの情報を記した「上位地名」テーブル、行政地名・自然地形・人工物などの地名が表す対象の違いを記した「地名_属性」テーブル、および、それらを相互につなぐ中間テーブルがある。

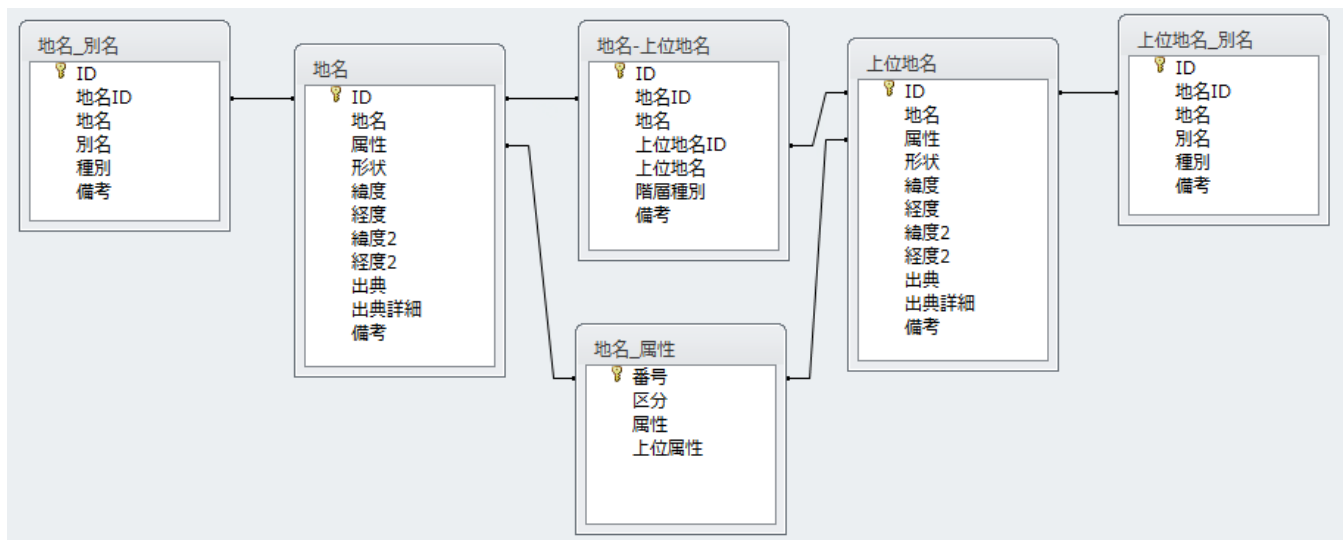


図 1 歴史地名辞書データのテーブル構成。

Figure 1 Table structure of the Digital Historical Gazetteer.

表 1 は、基本となる「地名テーブル」の各データ項目の詳細である。これ以外のテーブルについては、公開サイトよりリンクされている「データ項目」を参照されたい。

データ中の緯度・経度はすべて世界測地系 (WGS 84) に基づいている。また、データ項目[地名]に入力されている

地名は、原資料同士の表記の違いを吸収するために、原資料で使われていた旧字・異体字が新字や常用漢字との対照表[14][15]を用いて変換されている。

表 1 「地名」 テーブルのデータ項目の詳細

Table 1 Detail of data items in the "Place Name" table.

項目名	型	内容
ID	long	地名の ID.
地名	string	[ID]で表される地名の代表値. 他の地名は地名_別名テーブルに収容されている.
属性	int	地名の属性. 地名_属性テーブルの [番号]が入る.
形状	int	地名の地理的形状. 1:点, 2:線, 3:点. 3 は面の代表点を示すが, 同様の点が 1 となっている場合もあり, 実質的に 1 と 3 の区別はない.
緯度	double	地名が指し示す緯度. 形状が 2 (線) の場合は, 一方の端点の緯度.
経度	double	地名が指し示す経度. 形状が 2 (線) の場合は, 一方の端点の経度.
緯度 2	double	形状が 2 (線) の場合の, もう一方の端点の緯度. 形状が 2 (線) 以外の場合は null.
経度 2	double	形状が 2 (線) の場合の, もう一方の端点の経度. 形状が 2 (線) 以外の場合は null.
出典	string	地名の出典.
出典詳細	string	出典内の詳細情報 (巻, 頁, 項目など)
備考	string	備考. 地名の表記や場所の比定方法などを含む.

4. 利用例

本データの利用方法を理解するための一助として, 典型的な利用例を紹介する. 上述のとおり, 本データは複数のテーブルから構成されている. 単独のテーブルのみで処理する場合は, 表計算ソフトウェアなどでも容易に検索などの操作が可能であるが, 複数のテーブルを用いる場合は, 何らかの関係データベース管理ソフトウェアにデータをインポートして利用の方が効率的である.

(1) ジオコーディング (基本)

基本的なジオコーディングは, 「地名」テーブルだけで処理できる. データ項目「地名」を検索し, 当該レコードの「緯度」および「経度」を得ればよい. ただし, 他の地域 (県など) の同じ地名も同時に検索されるため, あらかじめ, データを絞り込んでおくか, 緯度経度の範囲を決めておくことが望ましい.

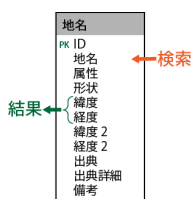


図 2 基本的なジオコーディングでのデータの利用.

Figure 2 Utilization of the data for basic geocoding.

(2) ジオコーディング (読み仮名の利用)

地名の読みは, 「地名_別名」テーブルのデータ項目「別名」に収容されている (注: 読みのデータが無い地名もある). また, データ項目「別名」には, 読み仮名以外にも, ローマ字表記や原資料での表記が入力されている. このため, 「地名_別名」テーブルのデータ項目「種別」で "ヨミ" と入力されているレコードのみを対象にする.

「地名」テーブルと「地名_別名」テーブルは, それぞれ, データ項目「ID」と「地名 ID」で結合されている. 実装では, 両テーブルを使って, あらかじめ新たな検索用のテーブルを生成しても良いし, データ検索時に双方を結合するクエリを用いても良い.

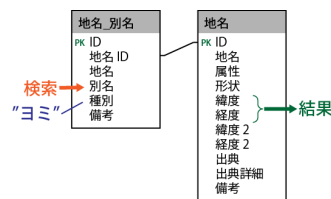


図 3 読み仮名を使ったジオコーディング.

Figure 3 Geocoding using alternate place names.

(3) 都道府県別のデータの切り出し

本データは, 収録されている地名の数も多く, 先述のとおり, 異なる地域の同じ地名が同時に検索されてしまう問題もある. このため, 何らかの形であらかじめ「地名」テーブルを絞り込んでおくことが現実的である. 例えば, 以下のように, 都道府県ごとにデータを切り出すことが想定される.

地名の階層構造は, 「地名」テーブルと「上位地名」テーブルにより表現されている. 1 つの地名に対し, 県名と旧国名が対応付けられている場合があるため, これらは, 中間テーブル(「地名_上位地名」)を介して結合されている.

特定の都道府県の地名データのみを切り出すには, 「上位地名」テーブルのデータ項目「地名」で必要な都道府県名のレコードの ID 抽出し, 中間テーブル「地名_上位地名」を介して, 該当するレコードを「地名」テーブルより抽出する.

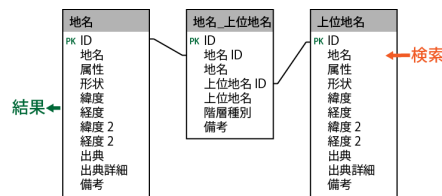


図 4 階層構造に基づく都道府県別のデータの切り出し.

Figure 4 Record extraction according to hierarchical structure.

5. 利用に際して

本データは、クリエイティブコモンズライセンスのCC-BYに近い形のライセンスで提供されている。データ構築に携わった人間文化研究機構およびH-GIS研究会の名称を明記すれば、商業利用も含め、自由に利用、加工、頒布等が可能である。なお、義務を課すものではないが、今後のデータの拡充および研究の発展に活用するため、利用事例の報告が推奨されている。また、歴史地名には、今日の目からすれば適切ではない表現が含まれていることが想定される。この点について、それぞれの出典資料が刊行された当時を表す歴史的・学術的資料としての趣旨を理解・留意の上での利用を求めている。

6. 歴史地名辞書データの今後

今回公開されたデータは、単純なCSVファイルであるため他のデータとの連携も視野に入れると、RDF形式によるLinked Dataとして公開することも考えられる。これにより、GeoNames [16] や Getty [17]といった地名に関する他の大規模なリソースと連携を図ることも可能になる。これを実現するためには、「いつごろ使われた地名か」といった、歴史地名に固有な情報を記述するための語彙の整備も必要になる。

また、今回公開された歴史地名辞書データは、階層構造に関する情報が不十分である。旧国名に対応付けられていない地名が数多くあり、旧郡名にはほとんどの地名が対応付けられていない。これらを解決するために、地名が旧国名、旧郡名ごとに記載されている旧高旧領取調帳[18]を用いることにより、地名の階層構造に関する情報を整備することも今後の課題として浮上している。

謝辞 歴史地名辞書データは、人間文化研究機構の研究資源共有化事業、および、H-GIS研究会に関連する京都大学地域研究統合情報センターの共同研究の一環として行われた。また、データ構築にあたっては、人間文化研究機構の各機関、および、H-GIS研究会に属する研究者の多大なる助言、助力をいただいた。ここに謹んで感謝の意を表す。

参考文献

- [1] 東京大学空間情報科学研究センター。“Geocoding Tools & Utilities - 位置参照技術を用いたツールとユーティリティ”。<http://newspat.csis.u-tokyo.ac.jp/geocode/>, (参照 2018-07-19).
- [2] Google LLC. “Developer Guide | Geocoding API | Google Developers”. <https://developers.google.com/maps/documentation/geocoding/intro>, (参照 2018-07-19).
- [3] “H-GIS研究会”. <http://www.h-gis.org/>, (参照 2018-07-19).
- [4] 吉田東伍. 大日本地名辞書. 1900-1907, 富山房.
- [5] 延喜式神名帳. 藤原時平ほか編, 延喜式第九・十, 927.
- [6] 日本陸軍参謀本部陸地測量部(一部:国土交通省国土地理院). 5万分の1地形図(1,444図幅). 1892-1991, 国土地理院 承認

- 番号 平成 30 情使, 第 12 号.
- [7] 桶谷 猪久夫. 人文分野における日本地名辞書の構築と地名属性の特徴分析. じんもんこん 2007 論文集, 2007, p. 79-86.
 - [8] 桶谷 猪久夫, 前川 武. 式内社データベースの構築と分布の調査. 大阪国際大学紀要「大阪国際論叢」, 2007, vol. 20, no. 2, p. 49-62.
 - [9] 桶谷 猪久夫. デジタル地名辞書構築とその利用. アジア遊学, 2008, vol. 133, p. 182-187.
 - [10] 桶谷 猪久夫. 地図・地名データベース. 研究報告人文科学とコンピュータ (CH), 2009, vol. 2009-CH-83, no. 3, p. 1-8.
 - [11] 四井 恵介, 関野 樹, 原 正一郎, 桶谷 猪久夫, 柴山 守. 明治・大正期旧 5 万分の 1 地形図をベースにした地名辞書構築. じんもんこん 2010 論文集, 2010, p. 211-216.
 - [12] 桶谷 猪久夫. 地図・地名データベースの構築. HGIS 研究協議会編, 歴史 GIS の地平, 2012, 勉誠出版, p. 79-88.
 - [13] Hara, S. and Sekino, T.. Digital Gazetteer as a Knowledgebase for Open Data Science (2nd report). Proceedings of PNC 2018 Annual Conference and Joint Meetings, in press.
 - [14] 東京大学史料編纂所. “異体字同定一覧史料編纂所データベース異体字同定一覧”. http://www.wap.hi.u-tokyo.ac.jp/ships/itaiji_list.jsp, (参照 2018-07-19).
 - [15] 日本国政府, 常用漢字表 (平成 22 年内閣告示第 2 号). 2010.
 - [16] “GeoNames”. <http://www.geonames.org/>, (参照 2018-07-19).
 - [17] The J. Paul Getty Trust. “Getty Thesaurus of Geographic Names”. <http://www.getty.edu/research/tools/vocabularies/tgn/>, (参照 2018-07-19).
 - [18] 木村礎 校訂. 旧高旧領取調帳 関東編・近畿編・中部編・中国・四国編・九州編・東北編. 竹内理三ほか監修, 日本史料選書, 1969-1979, 近藤出版社.