地名にかかる情報技術に関する研究会

（2014 年 7 月 19 日開催）

報告書

京都大学地域研究統合情報センター・地域情報学プロジェクト
「地域の知」の情報学―時間・空間・語彙に注目した地域情報学の展開―

国立情報学研究所公募型共同研究「GeoNLP プロジェクト：
自然言語文を対象とした地名情報処理とコミュニティの展開」

大学共同利用機関法人人間文化研究機構・研究資源共有化事業委員会

はじめに

空間情報は研究資源の探索や解析に有効な糸口や手段を与えてくれる。この中で地名は、場所を特定するために最も古くから用いられてきた要素であり、歴史的な史資料も含めて空間情報を時間的に掘り下げてゆける可能性ももっている。しかしながら、地名を地理情報システム（GIS）や検索などの情報技術の中で扱おうとすると、緯度経度などの座標との対応付け、地名データの抽出、地名同士の時空間的な関係の表現など、さまざまな技術的な課題が浮かびあがってくる。

本報告書は、こうした地名にかかる情報技術に関する課題に対して取り組みを進めている以下の３つの研究組織が共催で 2014 年 7 月 19 日に国立情報学研究所において開催した研究会「地名にかかる情報技術に関する研究会」の概要をまとめたものである。

・京都大学地域研究統合情報センター・地域情報学プロジェクト「地域の知」の情報学―時間・空間・語彙に注目した地域情報学の展開―（H-GIS 研究会）
http://www.h-gis.org/

・国立情報学研究所公募型共同研究「GeoNLP プロジェクト：自然言語文を対象とした地名情報処理とコミュニティの展開」
http://agora.ex.nii.ac.jp/GeoNLP/

・大学共同利用機関法人人間文化研究機構・研究資源共有化事業委員会
http://int.nihu.jp/

総合地球環境学研究所
関野 樹

目次

報告１ …… 1
歴史資料としての地名の活用
榎原雅治（東京大学史料編纂所／人間文化研究機構）

報告２ …… 11
歴史地名辞書をいかに作るか？
桶谷猪久夫（大阪国際大学）

報告３ …… 33
資源共有化における地名の役割と時空間への展開
関野 樹（総合地球環境学研究所）

報告４ …… 43
地名情報処理環境 GeoNLP の紹介と歴史的な地名に関する課題
北本朝展（国立情報学研究所）

報告５ …… 61
地名データのセマンティック Web への応用
原 正一郎（京都大学地域研究統合情報センター）

歴史資料としての地名の活用

榎原雅治

地名は名づけられたその土地の成り立ちや空間として利用されてきた歴史を語るものであり、貴重な資料として歴史研究でしばしば活用されてきた。

歴史学において最も早くから地名が活用されてきたのは、古代の条里制研究においてである。戦前以来、各地の平野部に残る「一ノ坪」「二ノ坪」などの小字に注目することによって、条里制のプランが空間的に推定復元されてきた。また高度成長による都市化や、それに伴う人口移動によって伝統的農村の景観が急速に失われ始めた1970年代末ごろからは、おもに中世史の分野において伝統的農村の空間的復元やその形成過程の研究が盛行するようになったが、そこでは現在に残る小地名が重要な手掛かりとなっている。そして在地領主の支配拠点の具体的な姿や、集落の形態、耕地開発の過程などの解明などの成果がそこから生み出されてきた。

こうしたミクロな地域観察の場合とともに、マクロに列島の歴史を考えるうえでも、地名はきわめて重要な資料となる。黒坂周平『東山道の実証的研究』（吉川弘文館）は「センドウ」という発音される地名に注目することによって、畿内から奥羽に至るまでの中世の東山道（とうせんどう）のルートを浮かび上がらせた労作である。筆者もまた「宿」という地名に注目することによって、中世の東海道や山陽道のルートを明らかにした経験がある（『中世東海道を行く』中公新書）。また九州各地の海辺部には「京泊」という地名が存在するが、これらの立地状況を詳細に分析することによって、戦国期の九州から京都へ向かう航路が浮かび上がってくる。交通史研究においては空間を広い面でとらえることが必要であるが、地名はどこにでも存在する資料である。文献史料の残存状況に粗密のある中世の交通状況をとらえるにあたって、地名は大きな力を発揮すると考える。

このように有効な歴史資料である地名であるが、現存する地名を活用できる体制は十分であるとはいえない。日本における地名は、都道府県―市・町・村―大字―小字という構造となっている。どのレベルの地名が大字／小字として位置づけられているかは多様であるが、巨視的に見れば、小字が近世、あるいはそれ以前から用いられていた地名を継承している場合が多いのは確かである。歴史研究において用いられてきた地名もおもには小字に相当するものである。現在、小字を収集した刊本としては角川書店から刊行されている都道府県別『歴史地名体系』の巻末付録が最も整っている。全部または一部が収録されていない府県もあるが、市町村レベルの自治体史、戦前の郡誌なども併用すれば、かなりの割合で全国の小字を網羅することも不可能ではない。しかし現状では大部分は手作業によるほかない。これらをＤＢ化、地図への自動表示できるならば、地名の収集や相互比較は容易となり、歴史資料としての有効性は飛躍的に高まるだろう。小字の一元的なＤＢを作るのが困難なのであれば、研究者がそれぞれに集めた小字を登録すると、自動的に地図表示されるような仕組みができないか、というのが地名を歴史研究の手掛かりとしている私の願望である。

第9回人間文化件空情報資源共有化研究会
2014年度第2回H－GIS研究会
「自然言語文を対象とした地名情報処理とコミュニティの展開」第1回研究会
地名にかかる情報技術に関する研究会

歴史資料としての地名の活用

榎原　雅治

2014年7月19日
於　国立情報学研究所

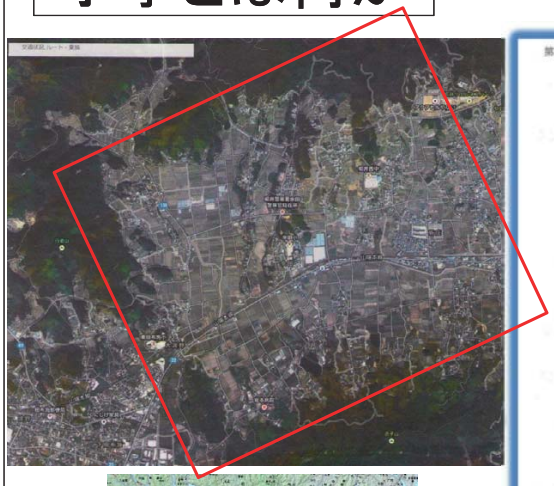
近代地名の階層性

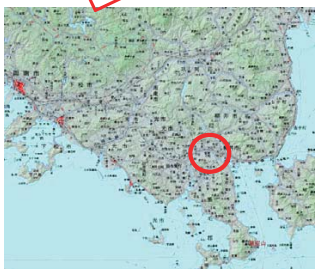
中世の字
↓
郡 ― 近世村（1874年　80,572）― 近世の字
↓
（1871年　大区小区制）
1873年　地租改正令公布
旧高旧領取調帳
近代の字
（地番設定）
1878年　郡区町村編制法　郡　―　町・村（74,325）
地方行政便覧
1889年　市町村制　郡 ― 町・村 ― 大　字 ―――― 小 字
（15,820）
法務局地籍図

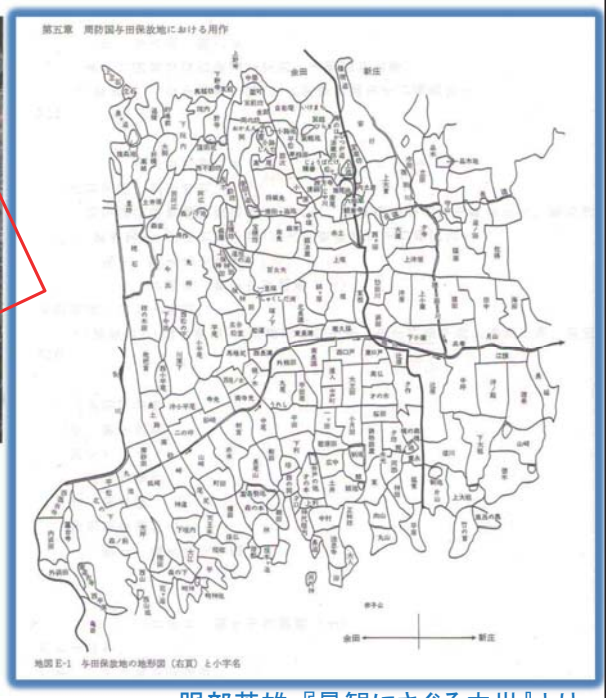
歴史的な字と小字の関係は多様
近世末の字がほぼそのまま踏襲された場合
近世末の字を地租改正時に整理統合した場合
郡区町村編制で統合される以前の近世村を「小字」と呼ぶ場合
（現代、地図に採用される「小字」はこの場合が多い）

小字とは何か

居住者のいない場所にもつけられた地名







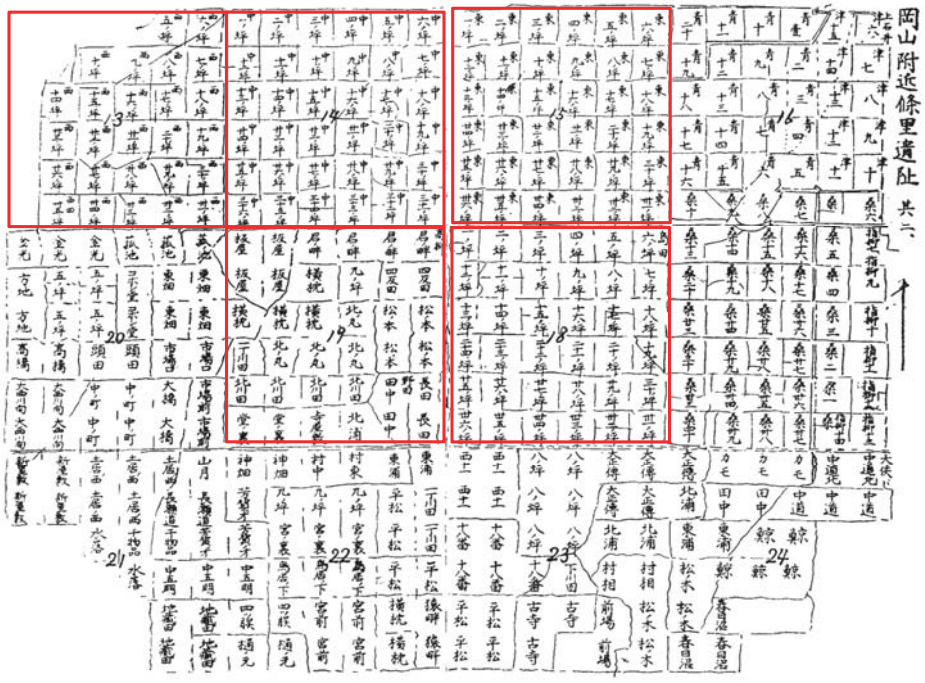
服部英雄：『景観にさぐる中世』より

日本史研究における地名の活用

- 近世村は現在までつながる ⇒
  その名称の登場は現代までつながる集落（居住形態）の成立を意味するので、文献史料上におけるその初出年代は重要
- 文献史料の欠落を補う史料としては、小字が圧倒的に重要（地名自体の史料的価値）
  文献史料のない地域にも小字地名は存在
  小字にはその土地の開発事情を示すものが多い
  小字には近世以前にさかのぼるものが多い

◆どんな研究に使われてきたか？
ミクロな地域観察
　古代の条里制の復元
　中世の荘園・村落研究
マクロな地域観察
　交通研究

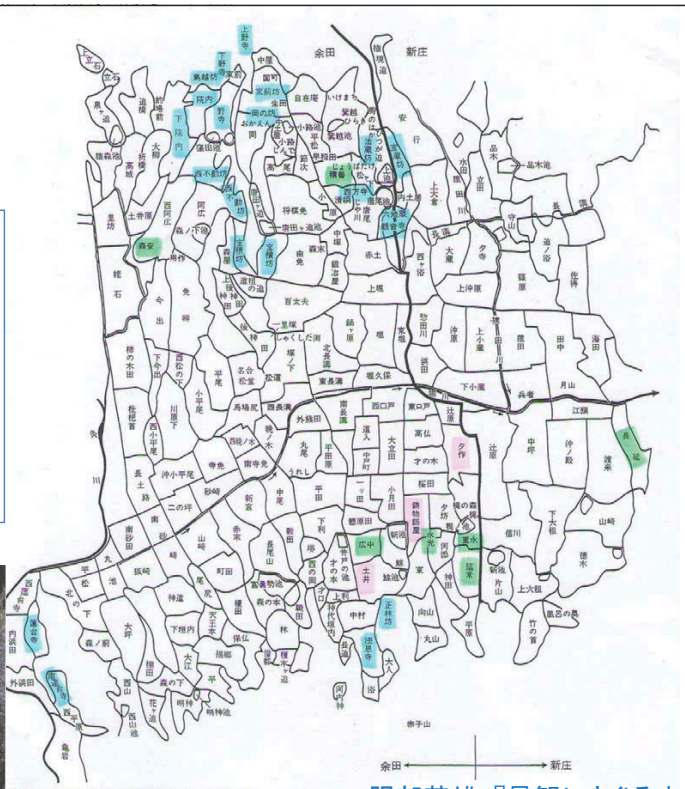
条里の復元（『岡山市史8』より）



中世村落景観の復元

周防国与田保
現・山口県田布施町

領主の拠点関係
土井・・・領主館
夕作・・・用作（直営地）
宗教施設
〇〇寺、〇〇院、〇〇坊
有力農民の居住地名（みょう）名

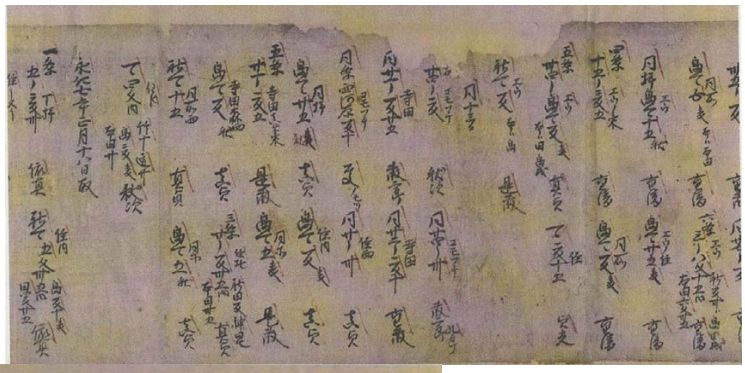


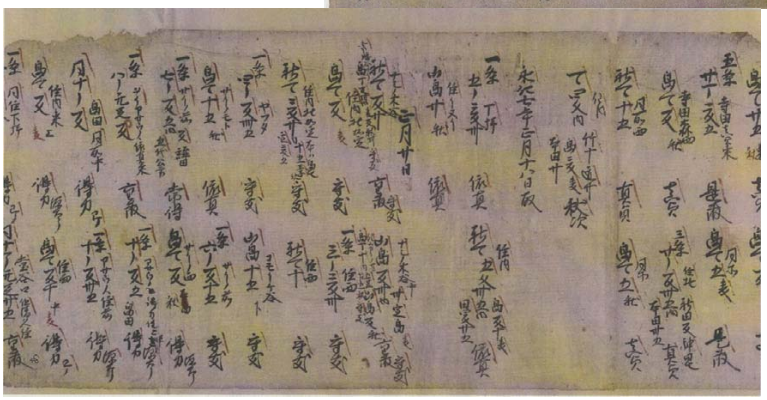
服部英雄：『景観にさぐる中世』より

播磨国矢野庄
（現・兵庫県相生市）

13世紀末の実検取帳（土地台帳、歩いた順に記載）

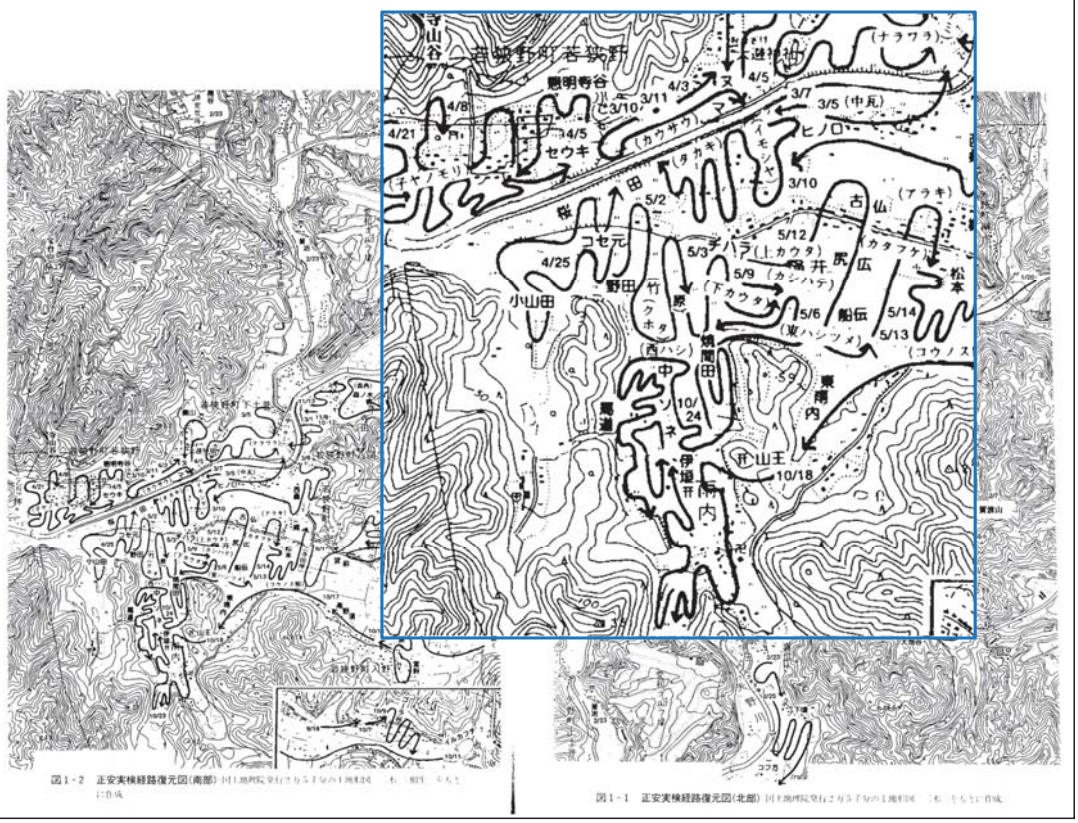
記載地名と現在の小字を照合

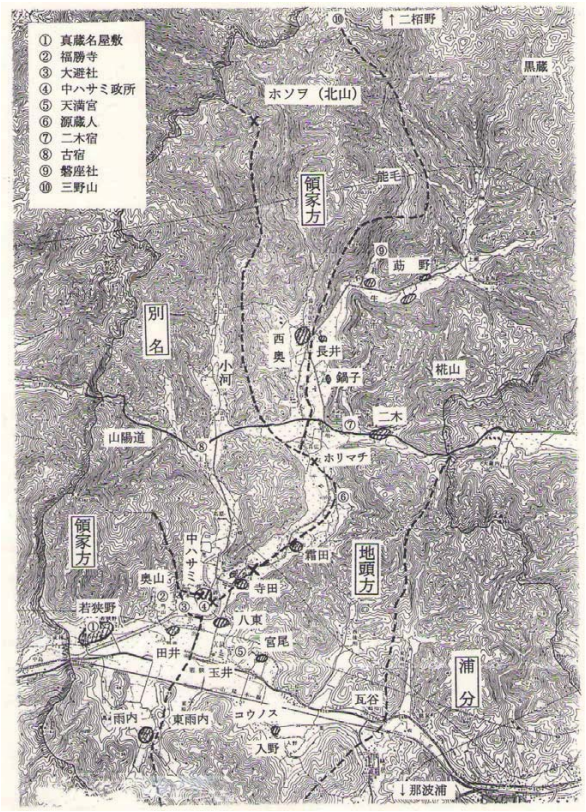




行政地名としては失われたもの
⇒ 聞き取り調査

転訛 （源蔵人→ 下頃）





実検順以外に何がわかったか？

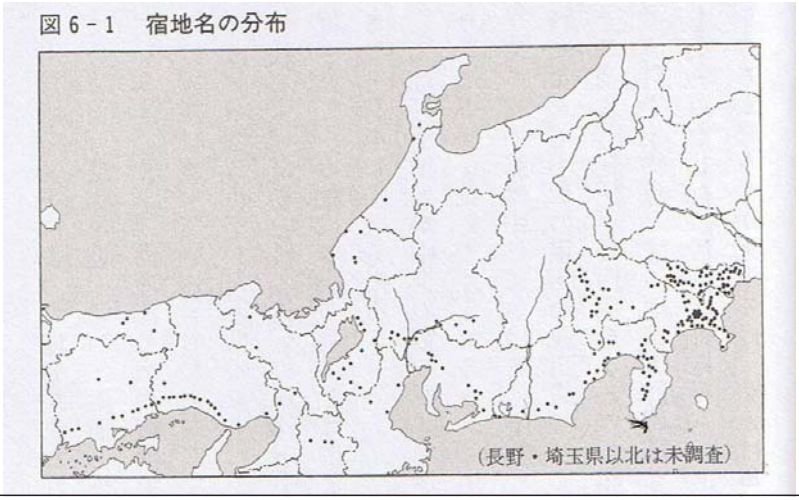
係争地、住居などの記載に注目すれば、その所在地点を特定できる

矢野庄固有の問題
・所領分割の実態
（領主の違い）

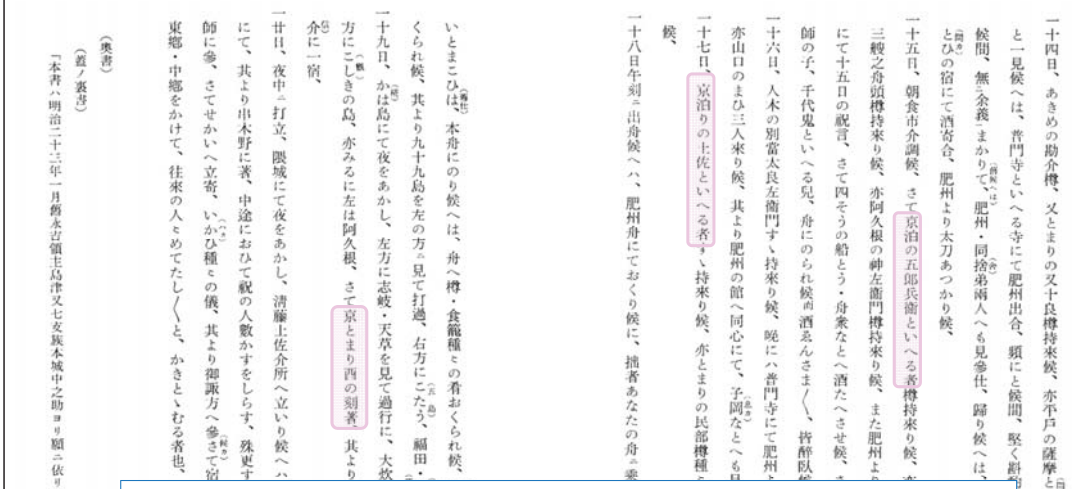
中世村落一般の問題
・住居、集落の位置
（13世紀末には現在の集落の原型が成立）

マクロな地域観察　交通路の探索

宿地名の収集（小字レベルでの収集）
　駿河以西と甲斐以東での「宿」地名の性格の違い
　駿河以東・・・「宿」は限定的　⇒
　　文献・考古資料と照合すると、中世の国家的幹道の痕跡と考えられる（地名が中世の交通路を浮き彫りにする）

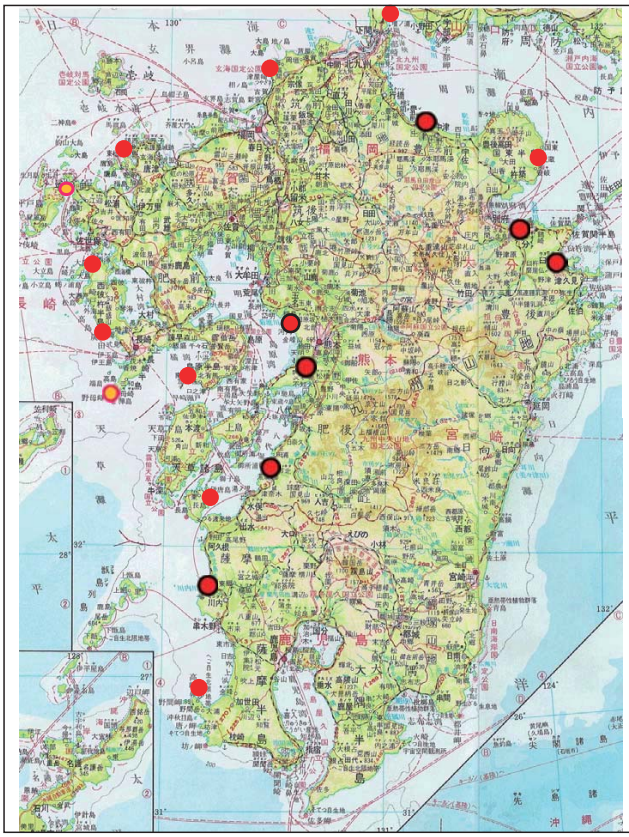


天正三年島津家久上京日記（島津家文書）



この日記に登場する「京泊」は薩摩川内の港町
島津家久が京都から薩摩に到着した時の湊

「京泊」という地名に意味はあるか？
⇒ 「京泊」地名の探索



九州各地の京泊

京泊の二類型

●河川でつながる後背地をもち、政治的権力の拠点の近傍
都市の外港

●岬や半島の突端で、陸上交通的には僻遠の地
畿内に向かう海路上の停泊地

● 京崎

ベースマップは帝国書院『新詳高等社会科地図』1984

九州各地の唐坊

- 西海岸のみ
- 入江に立地
- 河川との関係はさまざま

⇒

大型船の入港地

積出基地としての性格は希薄

遺跡調査も参照

⇒

日宋貿易時代の中国人居留地

とうぼう
唐房
当房
当方
東方

柳原敏昭「九州の唐坊・唐房の由来を探る」（『別冊歴史読本　地名を歩く』）による

ベースマップは帝国書院『新詳高等社会科地図』1984

地名の情報化への期待

- 小字の把握方法
  - 各法務局・市町村役場税務課の公図（字切図）
  - 角川書店『日本歴史地名体系』の附録
  - 各自治体史（戦前の郡誌、近年の市町村史）
  - 県別小字集成（愛知、兵庫）

  統一的な把握方法はまだ存在しない

  全国的な把握・比較が可能になれば、地名の地域性、意味についての理解を深めることができる

- 小字の詳細な位置の把握方法・・・当面は情報化は困難
  - 大半は法務局・市町村役場税務課の公図のみ
  - 各自治体史の附録、地域図書館（ごく一部）

小字データベースの試みー滋賀県での事例ー

角川地名体系の小字一覧をエクセルでDB作成（約89000件）

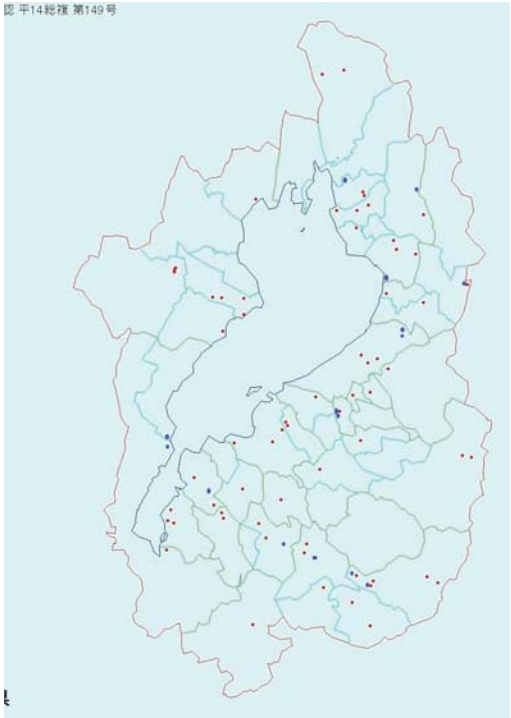
活用の一例：宿地名と市地名

☆宿と市の相違

国家的交通拠点としての宿⇔地域流通の拠点としての市

☆内陸から湖岸に至る地域内の交通路

小字の地図表示



歴史地名辞書をいかに作るか？

桶谷猪久夫（大阪国際大学）

１．はじめに

近年の地理情報システム(GIS:Geographic Information System）の急速な普及により、考古学、歴史学、民俗学、国文学、地理学、地域研究などといった人文科学の各分野においても、デジタル化された地図などの地理情報との連携が共同研究支援に有効になってきている。しかし、海外における時空間情報を利用した人文科学研究の事例は増えつつあるが、国内では、人文科学分野の研究に有効なデジタル地名辞書などの研究基盤となるツールが皆無に近いのが現状である。このような状況下で、我々は歴史的なデジタル地名辞書等の作成、有効なデータベース化と共同開発した地理情報共有システム（HuMAP :Geographic Information Sharing System for Humanities）とのインターフェースを開発した。

デジタル地名辞書の基礎となる文献は、下記に述べるが「大日本地名辞書」（52,824 件）、延喜式巻９と巻 10（延喜式神名帳）に記載された神社（式内社）2,861 社、日本寺院総監に記載された 78,588 寺院のデータベース化、旧５万分の１地形図から抽出した地名（242,735 件）、旧高旧領取調帳に記載された地名（97,358 件）、六十六部廻国の寺社（1,261 件）、廻国供養塔データベース（8,738 件）などであり、それらの位置情報等との統合を実現した。最後に、我々が本デジタル地名辞書に付与した地名情報から実際に検索した例を示す。

２．デジタル地名辞書構築に直接対象とした各文献の特徴

（１）大日本地名辞書（The Dictionary of Place Names in Greater Japan）：地名数　52,824 件

吉田東伍（1864 年：元治元年～1918 年：大正 7 年）が 1895 年（明治 28 年）から日本には統一した地誌がないことに着目し，地名辞書の編纂を決意、独力で 13 年かけて編纂した日本で最初の本格的な地名辞書である。国土の国名、郡名、町村名、郷名、神社、寺院、山川、湖沼、港湾などの地名に関する考証・変遷などが詳しく解説されている。大日本地名辞書は、明治 33 年に初版が発行され、刊行から百年以上も経た現在でも地理学、歴史、古典文学や民俗学の研究に必要不可欠な基礎資料として活用されている。大日本地名辞書の収録範囲は、第１巻から第８巻まであり、各地域別に分けて掲載されている。初版は全 11 巻からなり、総ページ数は 5,580 ページ、1,200 万文字で構成されている。その後増補され、現在、日本全土（北海道から沖縄（琉球））までが収録されている。

（２）式内社　　：2,861 社

延喜式全 50 巻のうち第１巻から第 10 巻が神祇関係（神祇式）で、第９巻と第 10 巻が延喜式神名帳といわれ、そこに記載された神社を式内社と言い社格の一つとされ、当時朝廷から「官社」として重要視された神社であることを示している。延喜式神名帳に記載された神社（式内社）は全国で 2,861 社、そこに鎮座する神の数は 3,132 座である。「神名帳」とは、古代律令制における神祇官が作成していた官社の一覧表のことあり、そこには国名、郡名別に神社が記載されており、また祭神、社格なども記載されている。

（3）日本寺院総監　　：78,558 寺院

日本寺院総監データを元に、旧国名、旧郡名と片仮名読みの付与、現在の市町村との対応付け、さらに位置情報、緯度・経度を付与しデジタル化を行った。

（4）旧５万分の１地形図　　：地形図枚数：1,435、地名数：242,735 件

明治初期から 2 万分の 1 縮尺をベースとして関東では迅速測図として測量された地図が作成され、関西では第四師管で准正式地形図の測図、仮製地形図が作成された。明治 18（1885）年からは全国整備を目的として正式 2 万分の 1 地形図が作成されていくが、明治 23（1890）年からは早期に全国をカバーするために縮尺を 5 万分の 1 に改めて整備がすすめられた。大正 5（1916）年に全国整備が完了する。日本で最初に精密測量された大縮尺かつ全国をカバーする旧 5 万分の 1 地形図をデジタル地名辞書の収録対象とし、それらに掲載されている全ての地名を採取した。

（5）旧高旧領取調帳

旧高旧領取調帳に掲載された維新期の 65,210 ヵ村を追加登録した。（重複があり、97,358 カ所）明治時代初期に政府が各府県に作成させた、江戸時代における日本全国の村落の実情を把握するための台帳である。取調帳には幕末時点での村名と当時の領主、および明治初期の各村の石高が記載されており、おおむね慶応年間から明治 4 年（1871 年）頃までの実情を示す史料である。

活字化と電子化は、明治大学教授の木村礎が、蘆田本を底本として東大本および『天保郷帳』『地方行政区画便覧』などを用い、欠落部分および筆写本ゆえの誤写と思われる箇所に校訂を加えたものが、近藤出版社から日本史料選書の一部として刊行された（全 6 冊）。さらに、現在では国立歴史民俗博物館が木村本にさらに校訂を加えてデータベースを作成し、同館のホームページ上でこれを公開している。

（6）六十六部廻国　　：1,261 寺社

六十六部廻国は、陸奥・出羽から薩摩・大隅までの六十六の国の寺社に法華経を奉納して回るという巡礼である。長い場合は、数年をかけて成就された。観音信仰をもとに畿内とその周辺を巡る西国三十三カ所巡礼や弘法大師空海の修行の寺社、開創の寺社を巡る四国八十八カ所遍路と違い、それらの寺社も含んでいるが、全国六十六カ国にまたがって、一宮、国分寺、八幡社を中心に周りの寺社を手当たり次第に巡り法華経を奉納して回るという巡礼であった。

対象とした文献（納経帳）

① 越後金益の六十六部廻国　　納経先：558 か所
② 善統不軽の廻国　　納経先：　78 か所
③ 六十六部権右衛門聖地選択　　納経先：548 か所
④ 下総国香取郡松子村宗心の廻国　　納経先：　77 か所

（7）廻国供養塔データベース　　：8,738 件

廻国供養塔データベースは、六十六部の廻国巡礼に関連して造立された石造物すべてを対象とし、具体的な採録基準は、「六十六部」または「廻国」「回国」の刻字を碑面にもつものは原則的に採録対象とした。ただし、「六十六部」の刻字があっても六十六部読誦や六十六部書写の供養塔であるこ

とが明白なものは除外している。「六十六部」や「廻国」「回国」の刻字はもたないが「奉納大乗妙典」とある供養塔については、 次の条件の一にかなう場合に採録する。 ①現状では判読困難ながら、「六十六部」「廻国」「回国」などの文字があったと推測される。 ②六十六部廻国にかかわることを示す出土物・文献・伝承などがある。③造立の中心的役割を果たしたとみなされる人物が「行者」などの肩書きをもつか他国の者で、 廻国巡礼の実態がともなっていると判断されたものを採録した。

我々がデジタル地名辞書として構築した文献毎の地名数を表１．に示す。

表１．デジタル地名辞書に格納された各文献毎の地名数とその合計

文献名	地名件数
大日本地名辞書	52,824
式内社	2,861
寺院総鑑	78,558
旧５万分の１抽出の地名	242,735
旧高旧領取調帳	97,358
六十六部廻国	1,261
六十六部廻国供養塔データベース	8,738
総合計	484,335

３．デジタル地名辞書の検索機能

具体的な検索例として、式内社である「杵築大社（現、出雲大社）」を検索文字列（キーワード）として指定し、検索した結果を示す。検索方法は、プロジェクトの HP、http://pnc-ecai.oiu.ac.jp/chimei/ を立ち上げ、検索メニュから検索対象の地名属性（旧５万分の１地形図）を指定し、該当する地名（出雲大社）を入力したデジタル地名辞書検索システムの検索画面を図１に示す。図２は検索結果表示画面であり、検索後「旧５万分の１地形図」と「出雲大社」でヒットした一覧が表示される。この場合は，ヒット数（：１）が表示され、その内容が項目毎（地名、図名、現在地名、地名属性：建物　神社）が表示されている。左側の「詳細」をクリックすると該当する「出雲大社」の地名、緯度、経度、現在の住所、地名属性（建物　神社）、図名、測量年

図１．地名辞書検索システムの検索画面

等が詳細に図３に表示される。「図名画像」の「この地名の位置を表示」をクリックすると、その所在情報が国土地理院の旧５万分の１地形図で表示される。図４は、「出雲大社」の所在地を表示した旧５万分の１地形図の GeoTIFF 画像の表示である。GeoTIFF は位置情報を持っているため該当「出雲大社」の位置情報が旧５万分の１地形図上に赤色（丸印）でブリンクして表示される。

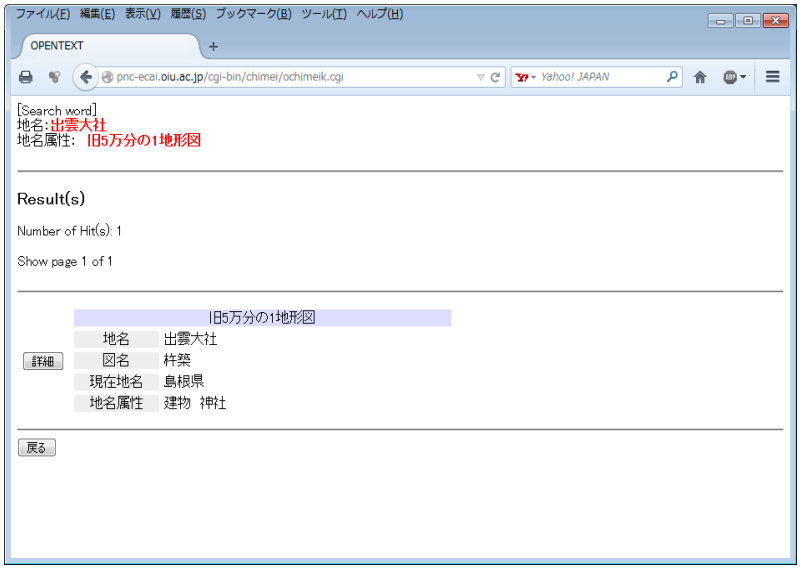


図２．地名辞書検索システムの検索結果表示画面

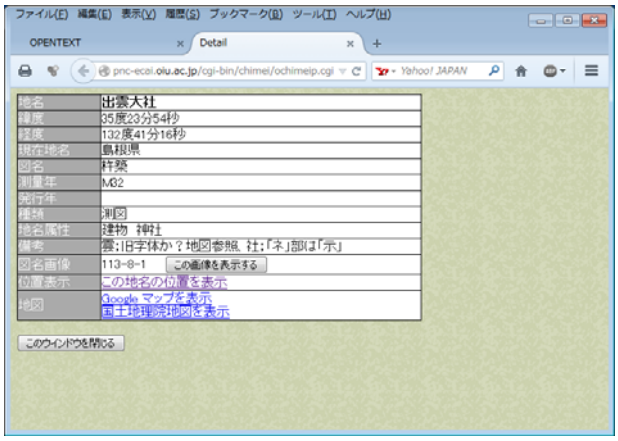


図３．地名辞書検索システムの検索結果詳細表示画面

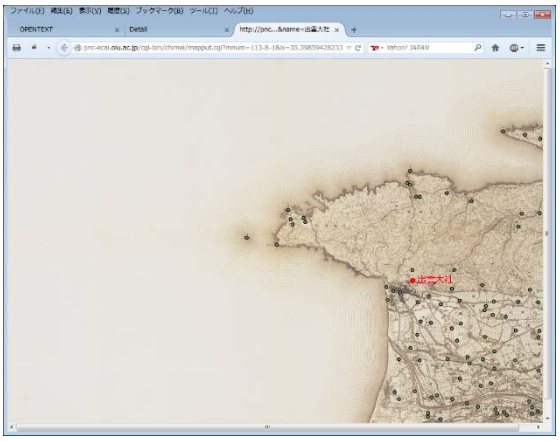


図４．検索結果の該当所在地（出雲大社）の旧５万分の１地形図表示画面

参考文献

(1) Ikuo OKETANI, “The Construction of the Digital Gazetteer and the Topographical Maps Database based on Humanities GIS and the Coordination of the Full Text Retrieval System of Japanese Historical Resources”, PNC 2012 Annual Conference and Joint Meetings, UC Berkeley School of Information, December 7-9, 2012, ppt(29 slides)

(2) 桶谷猪久夫、「六十六部資料の可視化からみえること」、CIAS 複合ユニット第 1 回研究会、京都大学地域研究統合情報センター、2014 年 6 月 13 日～14 日、ppt(38 slides)

歴史地名辞書をいかに作るか？

地名にかかる情報技術に関する研究会
（国立情報学研究所（学術総合センター）

2014年7月19日

大阪国際大学
桶谷猪久夫　oketani@oiu.ac.jp

1

- 近年の地理情報システム(GIS) の普及
- 人文科学分野：考古学、歴史学、民俗学、国文学、地理学、地域研究など

  ＊地理情報との連携が研究支援に有効
- 海外における時空間情報を利用した人文科学研究の事例は増えつつある
- 国内では、デジタル地名辞書などの研究基盤となるツールが皆無に近い
- 人文科学のための地理情報共有システム (Geographic Information Sharing System for Humanities)と共同で開発

デジタル地名辞書構築に直接対象とした各文献の特徴（1）

（1）大日本地名辞書(*The Dictionary of Place Names in Greater Japan*)　：地名数　**52,824**

- 吉田東伍（よしだ　とうご）編纂　　：元治元年(1864)～大正7年(1918)
- 明治28年（1895）日本には統一した地誌がないことに着目し、地名辞書の編纂を決意
- 明治33年(1900)に初版が発行
- 明治40年（1907）に完成　：13年間かけて編纂
- 日本で最初の本格的な地名辞書
- 刊行から百年以上も経た現在でも歴史地理学の基礎資料として活用

デジタル地名辞書構築に直接対象とした各文献の特徴（2）

（２）式内社

- 延喜式全50巻のうち第１巻から第10巻が神祇関係（神祇式）で，第９巻と第10巻が延喜式神名帳といわれ，そこに記載された神社を式内社と言い社格の一つとされ，当時朝廷から「官社」として重要視された神社であることを示している。
- 延喜式神名帳に記載された神社（式内社）は全国で2,861社で，そこに鎮座する神の数は3,132座である．「神名帳」とは，古代律令制における神祇官が作成していた官社の一覧表のことあり，そこには国名，郡名別に神社が記載されており，また祭神，社格なども記載されている．

（３）日本寺院総監

- 日本寺院総監データを元に，旧国名，旧郡名と片仮名読みの付与，現在の市町村との対応付け，さらに位置情報，緯度・経度を付与しデジタル化を行った（78,558寺院）．

4

デジタル地名辞書構築に直接対象とした各文献の特徴(3)

（４） 旧５万分の１地形図

- 明治初期から2万分の1縮尺をベースとして関東では迅速測図として測量された地図が作成され，関西では第四師管で准正式地形図の測図，仮製地形図が作成された。
- 明治18(1885)年からは全国整備を目的として正式2万分の1地形図が作成されていくが，明治23(1890)年からは早期に全国をカバーするために縮尺を5万分の1に改めて整備がすすめられた。
- 大正5(1916)年に全国整備が完了する。
- 日本で最初に精密測量された大縮尺かつ全国をカバーする旧5万分の1地形図をデジタル地名辞書の収録対象とし，それらに掲載されている全ての地名を採取した。
- 地形図枚数：1, 435、地名数：242, 735

5

デジタル地名辞書構築に直接対象とした各文献の特徴(4)

（５） 旧高旧領取調帳

- 旧高旧領取調帳に掲載された維新期の65,210ヵ村を追加登録した。(重複があり、97,358カ所)
- 明治時代初期に政府が各府県に作成させた、江戸時代における日本全国の村落の実情を把握するための台帳である。
- 取調帳には幕末時点での村名と当時の領主、および明治初期の各村の石高が記載されており、おおむね慶応年間から明治4年(1871年)頃までの実情を示す史料である。
- 活字化と電子化は、明治大学教授の木村礎が、蘆田本を底本として東大本および『天保郷帳』『地方行政区画便覧』などを用い、欠落部分および筆写本ゆえの誤写と思われる箇所に校訂を加えたものが、近藤出版社から日本史料選書の一部として刊行された(全6冊)。さらに、現在では国立歴史民俗博物館が木村本にさらに校訂を加えてデータベースを作成し、同館のホームページ上でこれを公開している。

6

デジタル地名辞書構築に直接対象とした各文献の特徴(5)

（６）六十六部廻国

- 六十六部廻国は、陸奥・出羽から薩摩・大隅までの六十六の国の寺社に法華経を奉納して回るという巡礼
- 長い場合は、数年をかけて成就された。観音信仰をもとに畿内とその周辺を巡る西国三十三カ所巡礼や弘法大師空海の修行の寺社、開創の寺社を巡る四国八十八カ所遍路と違い、それらの寺社も含んでいるが、全国六十六カ国にまたがって、一宮、国分寺、八幡社を中心に周りの寺社を手当たり次第に巡り法華経を奉納して回るという巡礼であった。

- **対象とした文献（納経帳）**

（１）越後金益の六十六部廻国　納経先：558か所
（２）善統不軽の廻国　納経先：78か所
（３）六十六部権右衛門聖地選択　納経先：548か所
（４）下総国香取郡松子村宗心の廻国　納経先：77か所

7

デジタル地名辞書構築に直接対象とした各文献の特徴(6)

（７）廻国供養塔データベース（8, 738件）

＊採録の対象
・廻国供養塔データベースは、六十六部の廻国巡礼に関連して造立された石造物すべてを対象
・廻国供養塔の語は、本来、廻国成就の供養の目的（および、それに準ずる目的）で造立される石塔を指すべきもの。

＊具体的な採録基準：
・六十六部」または「廻国」「回国」の刻字を碑面にもつものは原則的に採録対象
・ただし、「六十六部」の刻字があっても六十六部読誦や六十六部書写の供養塔であることが明白 なものは除外する。
・「六十六部」や「廻国」「回国」の刻字はもたないが「奉納大乗妙典」とある供養塔については採録

8

廻国供養塔データベース（8,738件）のマッピング

形態	置換え	個数
自然石	自然石	2142
山状角柱	角柱	1003
櫛型	櫛型	827
丸彫型	丸彫型	823
角柱	角柱	807
笠付型	笠付型	439
光背型	光背型	301
駒型	駒型	245
宝篋印塔	宝篋印塔	191
板碑	板碑型	56
頭丸角柱	角柱	50

無縫塔	無縫塔	30
板碑型	板碑型	27
宝塔型宝篋印塔	宝篋印塔	27
香箱型	箱型	17
自然石板碑	自然石	17
灯籠	灯籠	17
宝塔	宝塔	15
「板碑型」	板碑型	12
光背型（舟型文字塔）	光背型	12
「石碑」	石碑	11
石幢	石幢	11
丸彫型×6基	丸彫型	10

9

デジタル地名辞書に格納された各文献毎の地名数とその合計

大日本地名辞書	52,824
式内社	2,861
寺院総鑑	78,558
旧5万分の1抽出の地名	242,735
旧高旧領取調帳	97,358
六十六部廻国	1,261
六十六部廻国供養塔データベース	8,738
合計	484,335

10

設計したデータ項目（１）

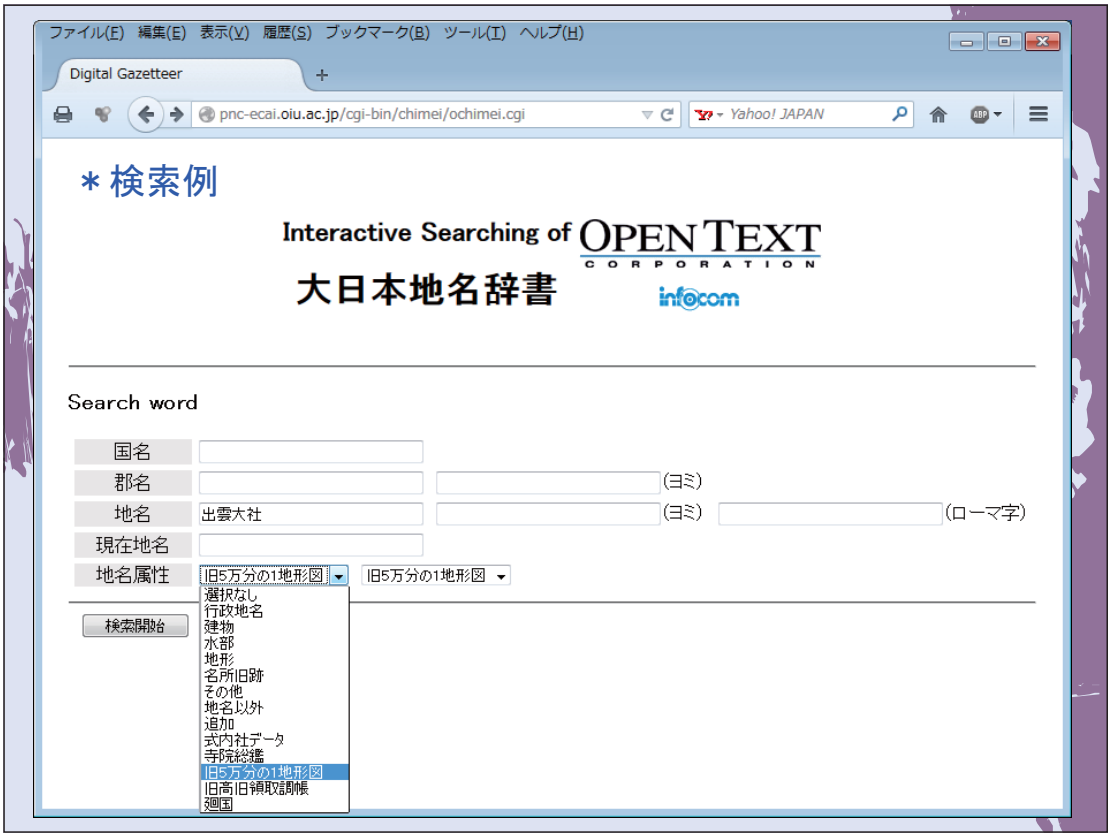
- ID：全レコードについてユニークな数字
- 地名：漢字、仮名表示、ローマ字表記（ヘボン式・英米規格・外務省・道路標識・駅名表・連母音を繋げずに表記、訓令式）
- 郡名：地名が包含される旧郡名、漢字、仮名表示、ローマ字表記（上に同じ）
- 国名：地名が包含される旧国名、漢字、仮名表示、ローマ字表記（上に同じ）
- 比定地名：地名に対応する現在の地名、漢字、仮名表示、ローマ字表記（上に同じ）
- 平成大改訂の地名：漢字、仮名表示、ローマ字表記
- 形状：地名が表す場所が、点、線、面であるかの区別

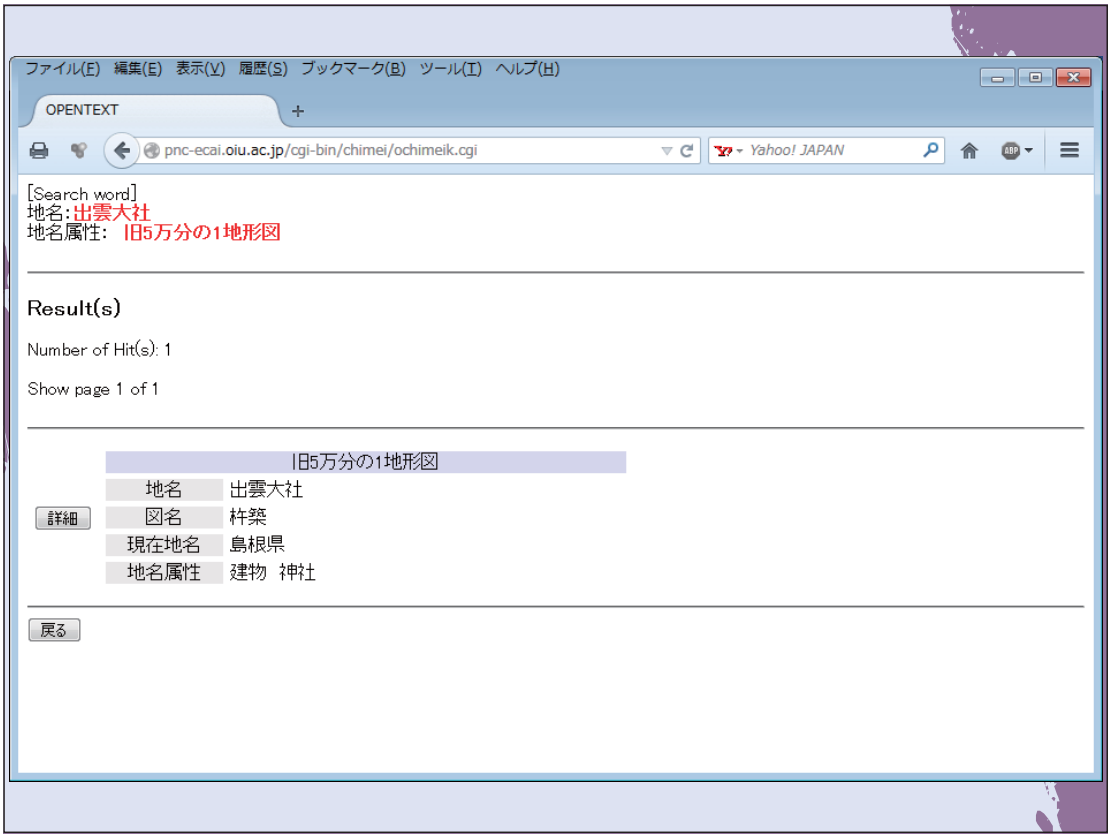
11

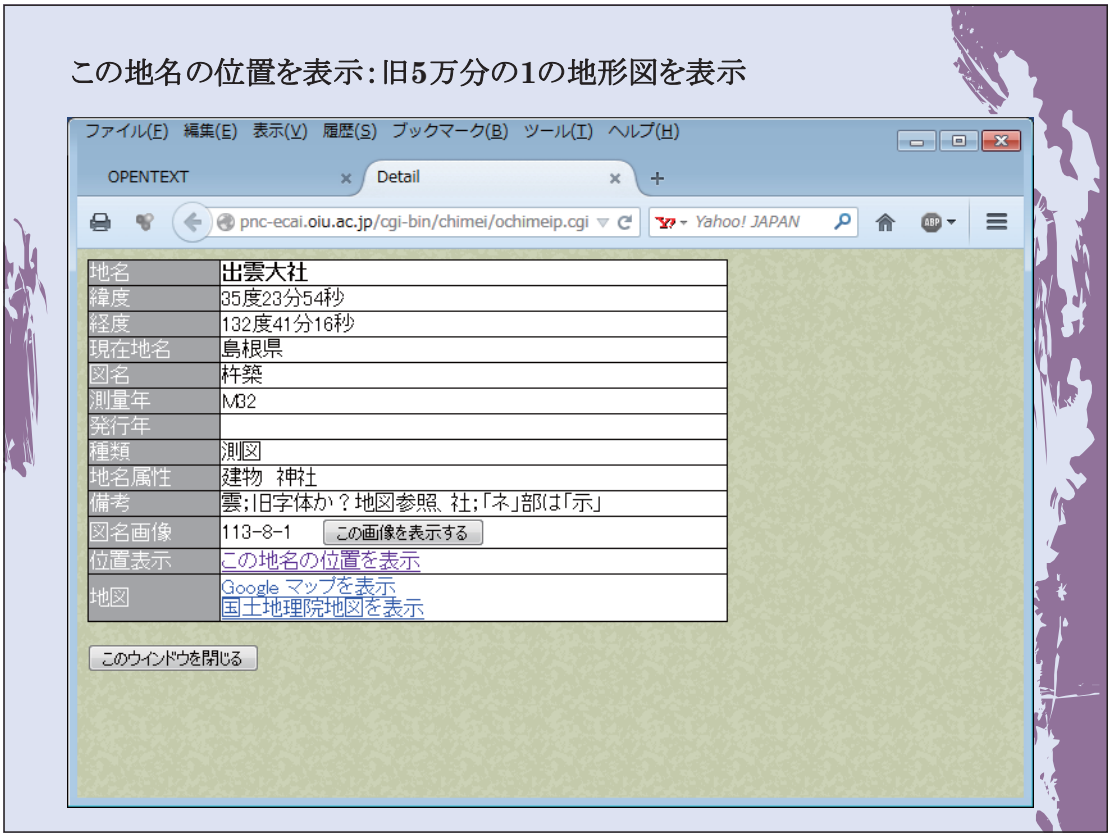
設計したデータ項目（２）

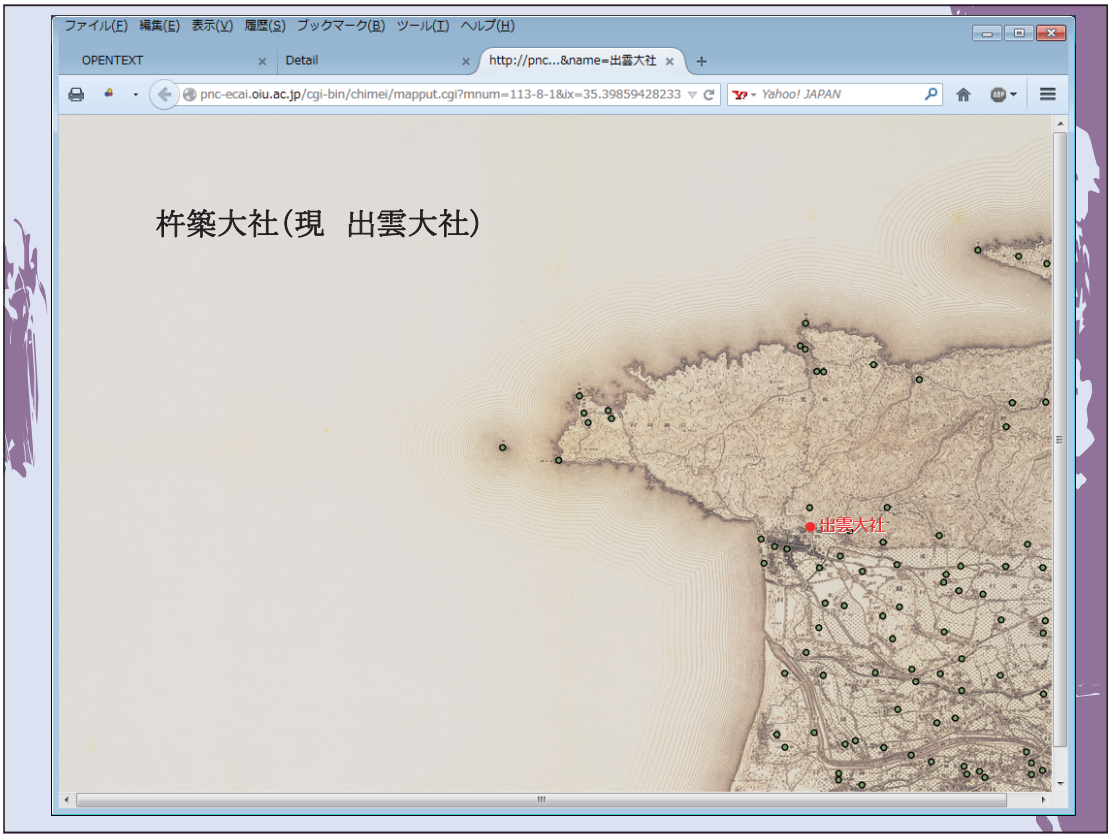
- 位置記述法：点、線、矩形、多角形
- 緯度：地名を代表する地点の緯度
- 経度：地名を代表する地点の経度、記述法は緯度に準ずるものとする。
- 地名属性：行政地名、山、岳、河川、湖沼、湾、建物（神社、寺院）、橋など
- 上限時代：その地名が有効であった上限年代（yymmdd）
- 下限時代：その地名が有効であった下限年代（yymmdd）
- 大日本地名辞書の巻号、該当ページ（拡張を考慮）
- 備考：緯度・経度・地名の比定に関わる情報

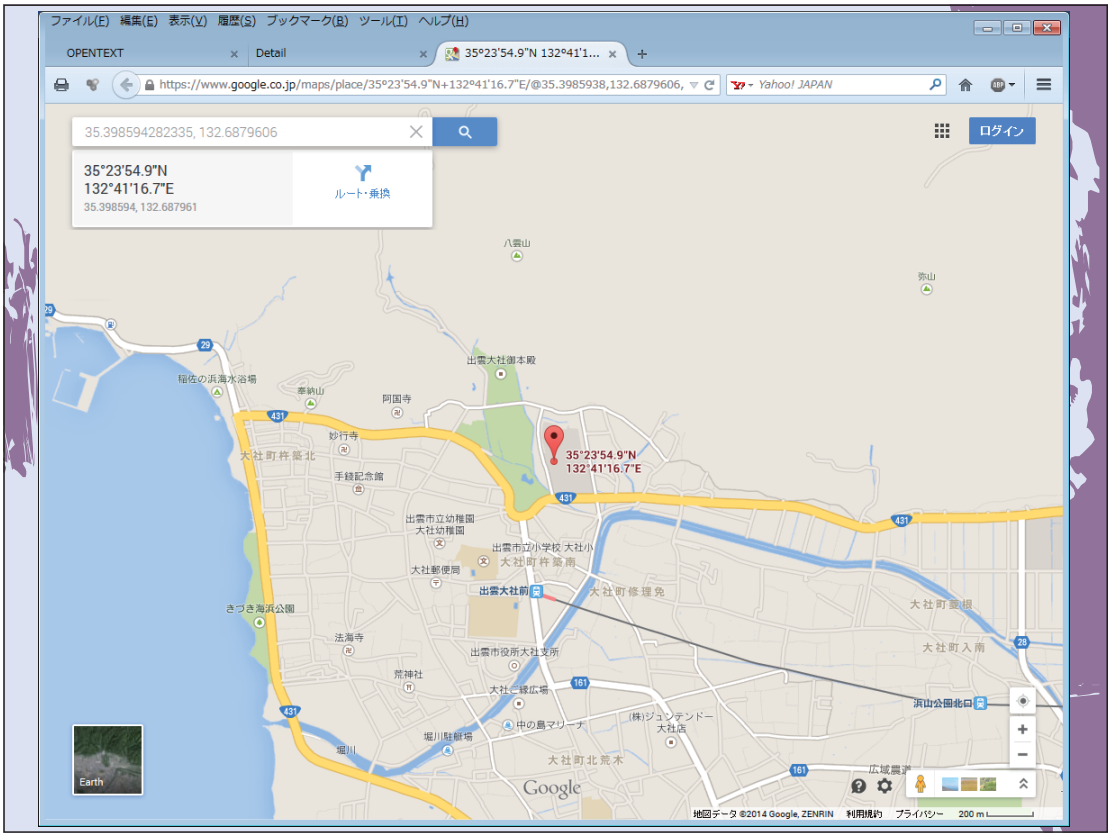
12

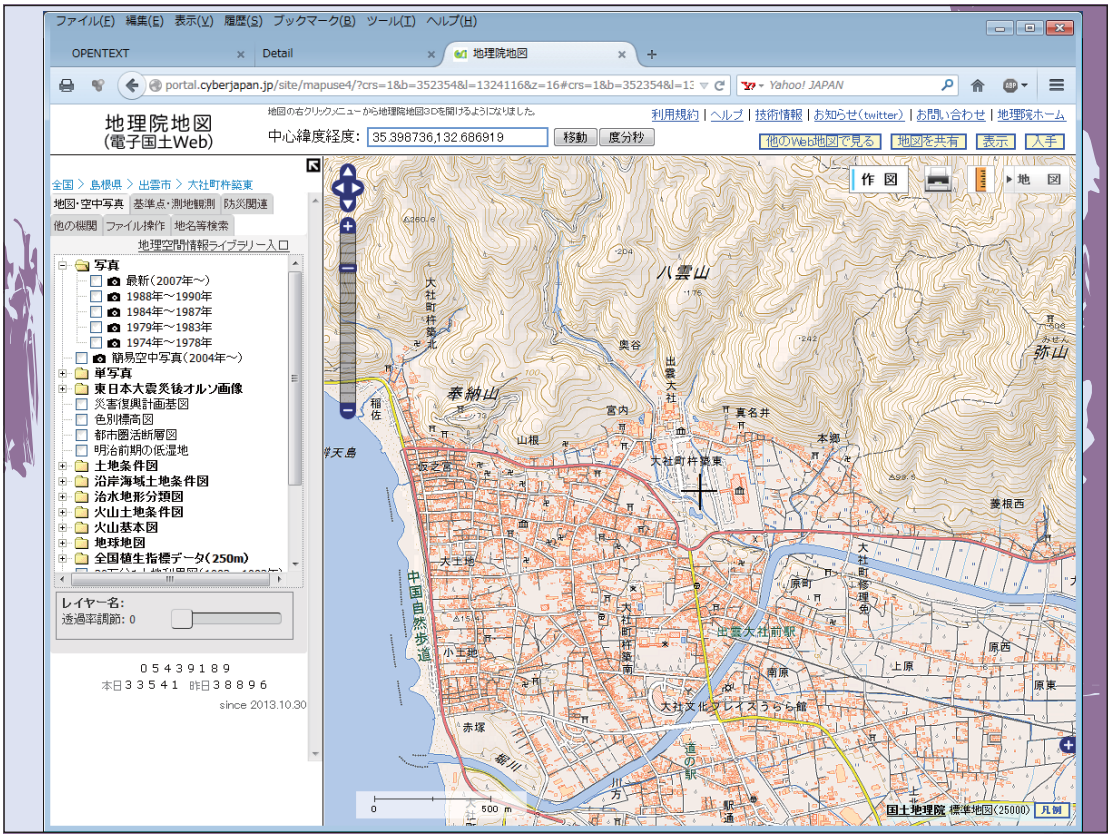

ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
Digital Gazetteer
pnc-ecai.oiu.ac.jp/cgi-bin/chimei/ochimei.cgi
Yahoo! JAPAN
＊検索例
Interactive Searching of OPEN TEXT CORPORATION
大日本地名辞書
infocom
Search word
国名
郡名
(ヨミ)
地名
出雲大社
(ヨミ)
(ローマ字)
現在地名
地名属性
旧5万分の1地形図
旧5万分の1地形図
検索開始
選択なし
行政地名
建物
水部
地形
名所旧跡
その他
地名以外
追加
式内社データ
寺院総鑑
旧5万分の1地形図
旧高旧領取調帳
廻国


ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
OPENTEXT
pnc-ecai.oiu.ac.jp/cgi-bin/chimei/ochimeik.cgi
Yahoo! JAPAN
[Search word]
地名:出雲大社
地名属性: 旧5万分の1地形図
Result(s)
Number of Hit(s): 1
Show page 1 of 1
旧5万分の1地形図
詳細
地名 出雲大社
図名 杵築
現在地名 島根県
地名属性 建物 神社
戻る


この地名の位置を表示：旧5万分の1の地形図を表示
ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
OPENTEXT
Detail
pnc-ecai.oiu.ac.jp/cgi-bin/chimei/ochimeip.cgi
Yahoo! JAPAN
地名 出雲大社
緯度 35度23分54秒
経度 132度41分16秒
現在地名 島根県
図名 杵築
測量年 M32
発行年
種類 測図
地名属性 建物 神社
備考 雲;旧字体か？地図参照、社;「ネ」部は「示」
図名画像 113-8-1 この画像を表示する
位置表示 この地名の位置を表示
地図 Google マップを表示
国土地理院地図を表示
このウインドウを閉じる


ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
OPENTEXT
Detail
http://pnc...&name=出雲大社
pnc-ecai.oiu.ac.jp/cgi-bin/chimei/mapput.cgi?mnum=113-8-1&ix=35.39859428233
Yahoo! JAPAN
杵築大社（現　出雲大社）
出雲大社


ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
OPENTEXT
Detail
35°23'54.9"N 132°41'1...
https://www.google.co.jp/maps/place/35°23'54.9"N+132°41'16.7"E/@35.3985938,132.6879606,
Yahoo! JAPAN
35.398594282335, 132.6879606
ログイン
35°23'54.9"N
132°41'16.7"E
35.398594, 132.687961
ルート・乗換
八雲山
弥山
出雲大社御本殿
稲佐の浜海水浴場
奉納山
阿国寺
妙行寺
大社町杵築北
手錫記念館
35°23'54.9"N
132°41'16.7"E
出雲市立幼稚園
大社幼稚園
出雲市立小学校 大社小
大社町杵築南
大社郵便局
出雲大社前
大社町修理免
大社町鳶根
大社町入南
きづき海浜公園
法海寺
出雲市役所大社支所
荒神社
大社一級広場
中の島マリーナ
(株)ジュンテンドー
大社店
浜山公園北口
堀川駐車場
堀川
大社町北荒木
広域農道
Earth
Google
地図データ ©2014 Google, ZENRIN 利用規約 プライバシー 200 m


ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
OPENTEXT
Detail
地理院地図
portal.cyberjapan.jp/site/mapuse4/?crs=1&b=3523548l=1324116&z=16#crs=1&b=3523548l=1
Yahoo! JAPAN
地理院地図
(電子国土Web)
地図の右クリックメニューから地理院地図3Dを開けるようになりました。
中心緯度経度: 35.398736,132.686919
移動
度分秒
利用規約 | ヘルプ | 技術情報 | お知らせ(twitter) | お問い合わせ | 地理院ホーム
他のWeb地図で見る
地図を共有
表示
入手
全国 > 島根県 > 出雲市 > 大社町杵築東
地図・空中写真
基準点・測地観測
防災関連
他の機関
ファイル操作
地名等検索
地理空間情報ライブラリー入口
写真
最新(2007年～)
1988年～1990年
1984年～1987年
1979年～1983年
1974年～1978年
簡易空中写真(2004年～)
単写真
東日本大震災後オルソ画像
災害復興計画基図
色別標高図
都市圏活断層図
明治前期の低湿地
土地条件図
沿岸海域土地条件図
治水地形分類図
火山土地条件図
火山基本図
地球地図
全国植生指標データ(250m)
レイヤー名:
透過率調節: 0
0 5 4 3 9 1 8 9
本日 3 3 5 4 1 昨日 3 8 8 9 6
since 2013.10.30
作 図
地 図
八雲山
奉納山
弥山
大社町杵築北
大社町杵築東
大社町杵築南
出雲大社
出雲大社前駅
大社文化プレイスうらら館
中国自然歩道
国土地理院 標準地図(25000) 凡例

六十六部日本廻国のマッピング

（１）越後金益の六十六部廻国[2)]

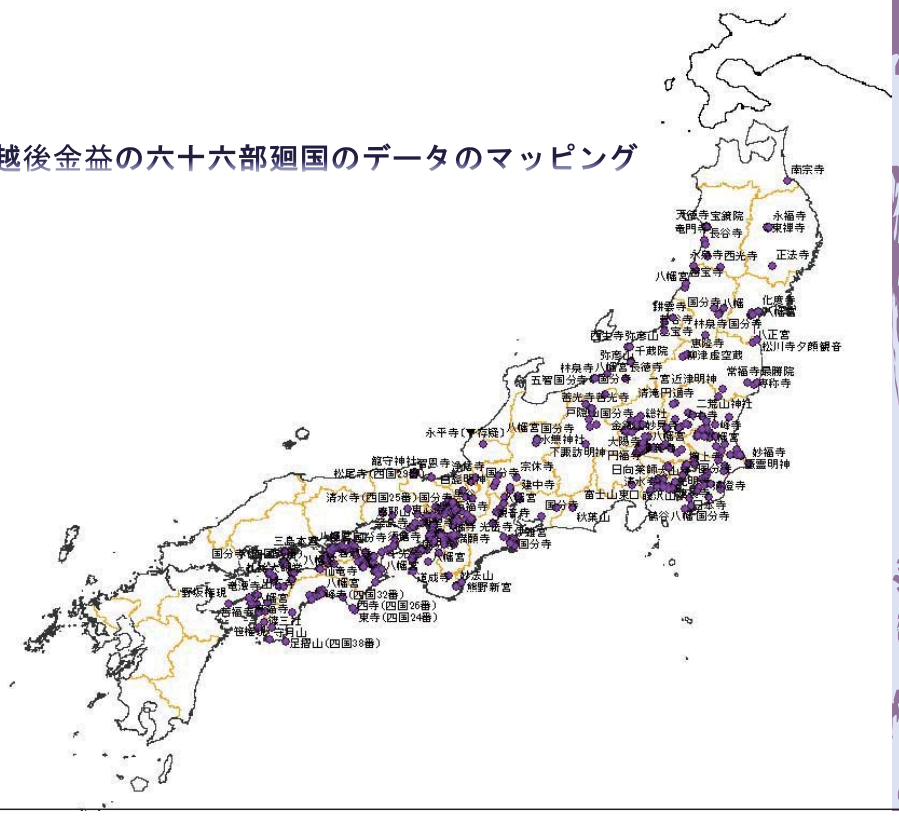
・史料：越後金益納経帳全17冊（徳島県つるぎ町・東福寺蔵）、納経先：558か所

・金益：越後国古志郡中村（現長岡市（旧　栃尾市）出身、俗名　金六

・廻国は、天保5年（1834）6月～嘉永元年（1848）6月：あしかけ15年にわたる納経帳

19

越後金益の六十六部廻国のデータのマッピング



六十六部日本廻国のマッピング

（２）善統不軽の廻国

・史料：善統不軽納経帳全１冊（福島県相馬市・横山重久氏蔵）、納経先：78か所

・善統不軽：史料所蔵者・横山家の10代前の先祖、相馬藩の刀鍛冶

・俗名 弥五左衛門吉成、享保３年（1718）没

・廻国は、宝永３年（1706）２月～同４年７月の１年５か月かけて巡礼

・納経帳は、1か所（高野山）を除き、廻国順に請取が並ぶ

21

善統不軽の廻国のデータのマッピング

六十六部日本廻国のマッピング

(3) 六十六部権右衛門聖地選択

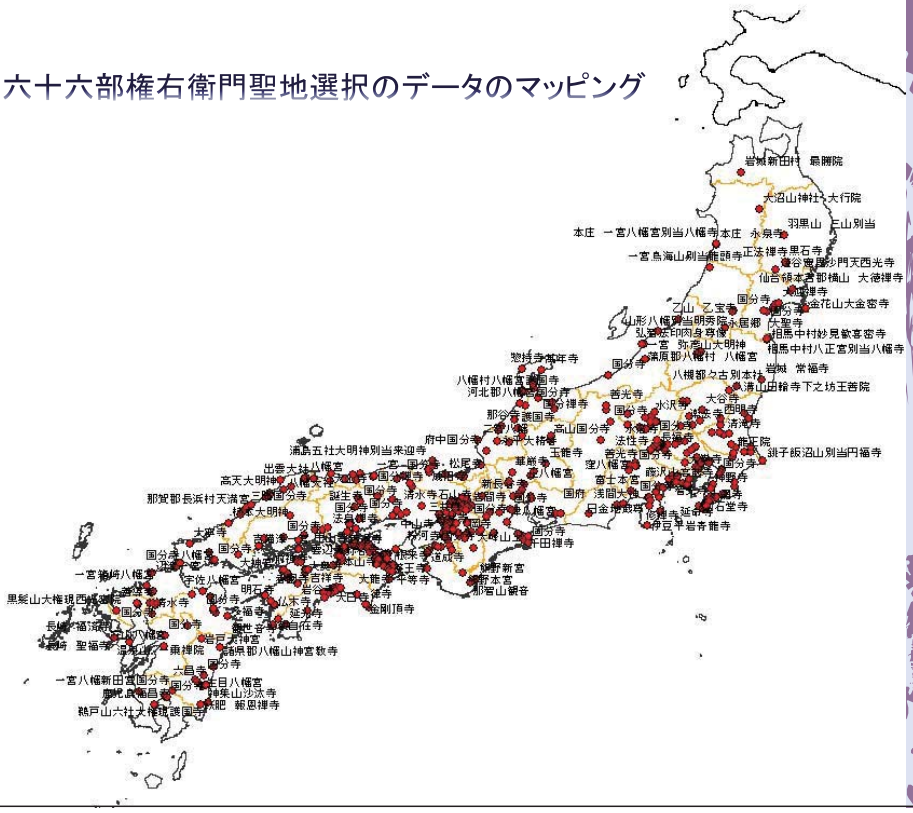
・史料：千葉県館山市立博物館、納経帳全3冊、納経先：548か所

・権右衛門：安房国山本村出身

・廻国は、安永4年3月1日（1775年3月31日）から安永8年10月18日（1779年11月25日）まで約4年8か月かけて巡礼

23

六十六部権右衛門聖地選択のデータのマッピング



六十六部日本廻国のマッピング

（4）下総国香取郡松子村（香取市）宗心の廻国

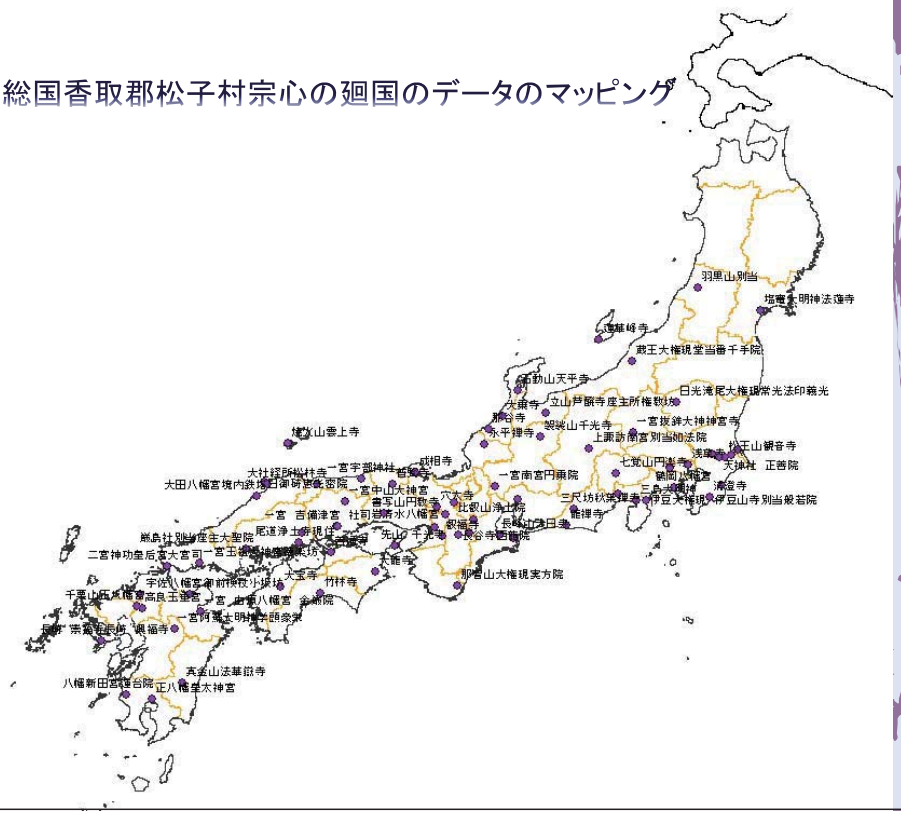
・史料：千葉県香取市松子宮野長太郎氏所蔵、
　　納経先：77か所

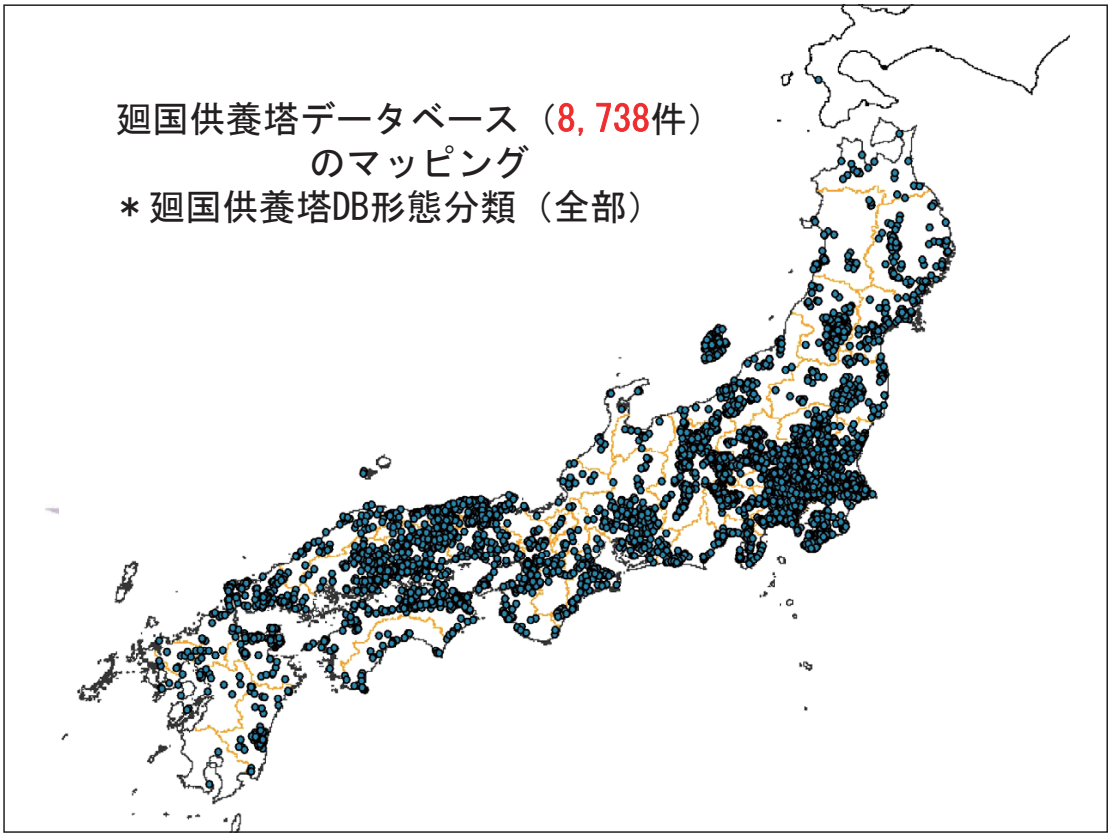
・宗心：下総国香取郡松子村出身

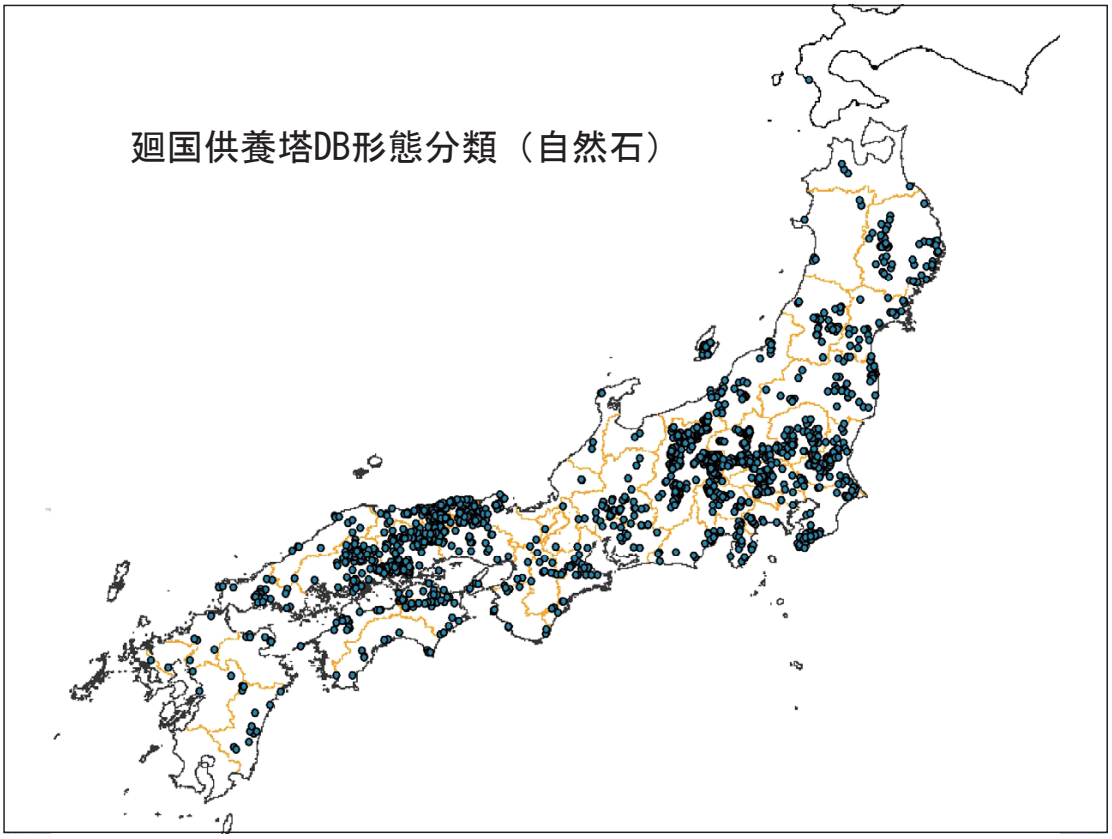
・廻国は、宝永4年（1707年）1月から宝永5年（1708年）10月まで約1年9か月かけて巡礼

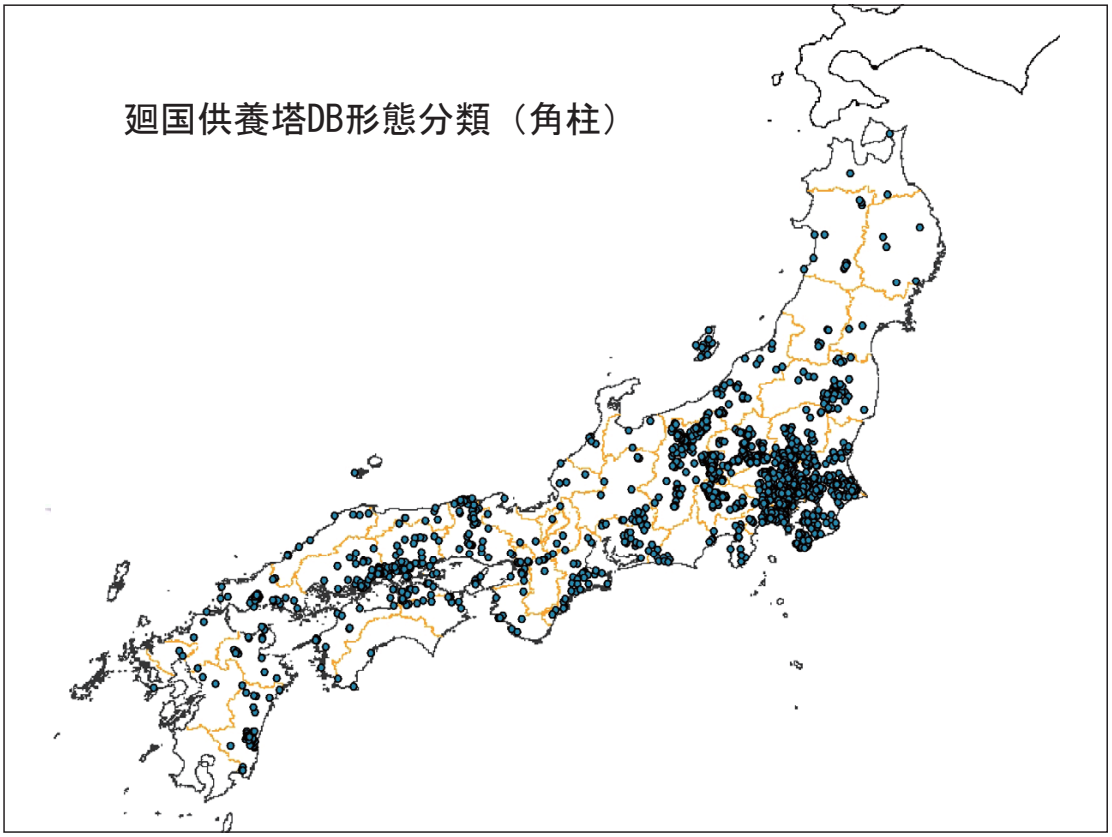
25

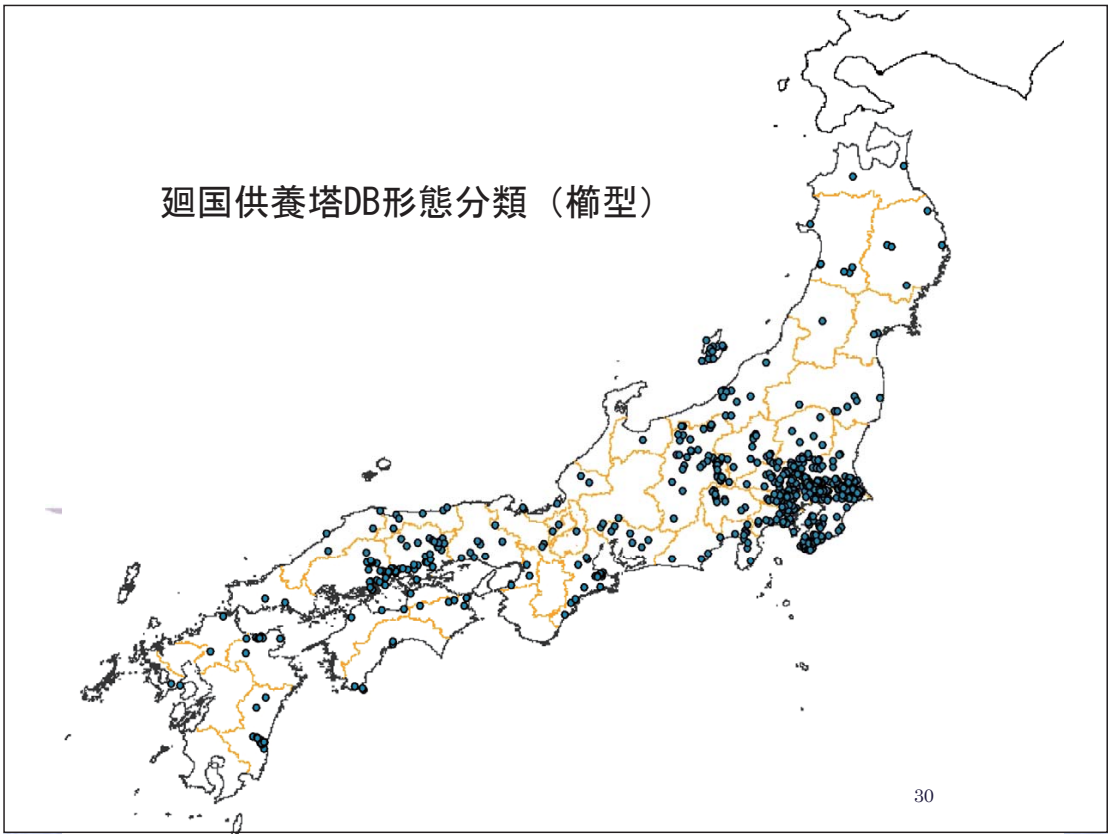
下総国香取郡松子村宗心の廻国のデータのマッピング




廻国供養塔データベース（8,738件）
のマッピング
＊廻国供養塔DB形態分類（全部）


廻国供養塔DB形態分類（自然石）


廻国供養塔DB形態分類（角柱）


廻国供養塔DB形態分類（櫛型）
30

六十六部廻国のトレッキング

＊六十六部廻国は、陸奥・出羽から薩摩・大隅までの六十六の国の寺社に法華経を奉納して回るという巡礼のため、奉納した日付（年月日）が分かる。そのため、マッピングとともにトラッキングも可能である。
＊今回は、六十六部廻国の４つのデータでトラッキングも完成した。

31

下総国香取郡松子村宗心の廻国のトラッキング

宗心の廻国.mxd

32

六十六部資料の可視化から見えること

１．宗心の廻国のトレッキングでの問題点
納経場所の空白が109日、45日、44日、35日の4回ある。
＊時間スケールバーを強制的に進ませるか、トレッキングの間隔を延ばして対応
＊トラッキングは巡礼が動的・視覚的に表示されるため巡礼の実態を見るために有効

２．規範リストと実際の巡礼寺社との関係
＊規範リストに掲載された寺社に巡礼しているとは限らない。
＊「善統不軽の廻国」と「下総国香取郡松子村宗心の廻国」はだいたい１か国１か所の寺社への巡礼であったが、あとの２つ「越後金益の六十六部廻国」と「六十六部権右衛門聖地選択」は、全国六十六カ国にまたがって、一宮、国分寺、八幡社を中心に周りの寺社を手当たり次第に巡り法華経を奉納して回るという巡礼であった。

33

資源共有化における地名の役割と時空間への展開

関野 樹（総合地球環境学研究所）

○はじめに

Google Maps などを利用して空間情報が手軽に扱えるようになり、その活用の幅も広がっている。その中で地名は空間を特定するための重要な要素であり、その利用環境が整えられてきた。たとえば、住所（地番）や施設の名称などを入力すると、それが緯度経度に変換され、地図上に表示されるサービスが Web 上で手軽に利用できる。しかしながら、こういったサービスの多くは現代の地名に限られており、過去の地名（歴史地名）などについては充分な利用環境が整っていない。一方、人間文化研究機構で進められている研究資源共有化事業のように、過去の文書や資料を扱ったデータベースを相互に連携させる仕組みにおいては、空間情報はそれらの重要な接点であり、当然、歴史地名を適切に取り扱う仕組みが必要である。このため、本稿では、資源共有化に関連させながら、情報技術における地名の取り扱いの現状と課題の整理を試みる。

ところで、議論に先立って、ここでの「地名」の範囲を明確にする必要があろう。地理学では、地名は、土地、山や河などの地形、文化圏などに付けられた名称を指し、施設や家屋などの名称含めないのが一般的である。ただ、資源共有化などの情報技術の中では、空間を特定するための何らかの識別子として「地名」の用語が用いられることも少なくない。したがって、ここでは施設や家屋なども含む、空間と結び付けられるすべての名称を「地名」として扱うものとする。

○地名と空間

情報技術の中で地名と空間はどのような関係で扱われているであろうか。空間情報に関する国際規格（ISO 19100 シリーズ）では、空間を特定する方法として、座標による方法（ISO 19111）と地理識別子による方法（ISO 19112）の 2 通りを定めている。前者は緯度経度などによるもので、空間上の位置を正確に特定することができるため、地図上に情報を表現する場合や GIS を用いた解析をしようとする場合は、この座標による方法が適当である。しかしながら、複雑な形状の空間範囲や複数の位置を一緒に扱う場合などは、処理が複雑になってくる。

もう一方の地理識別子は地名や郵便番号など、空間に結び付けられる何らかの名称や記号を指す。地名だけでは空間上の位置を特定することはできないので、多くの場合、地名と座標を結び付けることで地図上に情報を表現している。地名の場合は、階層構造（例：国 > 県 > 市町村 > 大字 > 小字）に基づいて対象範囲を拡げたり絞ったりすることも容易である。また、空間上の位置よりも施設や土地自体に主眼が置かれている場合、施設の位置が移動していても同じ「場所」として扱える利点もある。たとえば、「御所」での出来事を異なる複数の史資料から探す場合、時代と共に御所の空間上の位置が移動していても、

御所という「場所」はそのまま扱うことができる。さらに、地名の場合は、セマンティック Web 技術を用いながら他の種類の情報と連携させることも容易である。

地名に関して、空間上の位置を特定するための座標（緯度経度）、上下の階層の地名、地名の変遷等の情報をまとめた地名辞書データベース（gazetteer）が構築されている。各時代の地名についてもこれらが基盤情報として整備されてゆくことにより、地名を使った研究資源の連携や GIS を使った解析の促進されてゆくことが期待される。

○地名を使った検索・解析

地名に関する解析やデータ構築を行おうとした場合、まず、史資料から地名を抽出する作業が必要になる。ここでまず、どの単語が地名であるかを判断することが作業を進める上での壁となる。たとえば、「吉田」であれば、地名としての吉田（京都市左京区吉田や吉田神社など）があるが、他にも、人名、団体名、モノの名称など様々な「吉田」が存在する。これらを自動的に判別することはなかなか容易ではない。次に、地名を使った場所の表現である。「吉田のあたり」、「吉田の北 1km」なども場所を指しているので地名とみなせるが、実際の空間とどのように結びつけるのが難しい。

次に、地図上での表現や GIS を使った解析を行う場合は、地名と空間上の位置を結び付ける作業が行われる。この作業でよく問題になるのが「どこを代表点にするのか」である。たとえば、「京都市」の場合、その空間範囲を座標で指定することも可能ではあるが、近隣の町村との合併でその範囲は時と共に変化するし、また、データが複雑かつ大きくなるために扱いにくくなる。結果として、どこかを代表点とし、その座標を「京都市」に結びつける場合が多い。一般に行政地名などでは役所の場所（京都市であれば京都市役所の場所）を代表点とすることが慣例となっている（市役所のどの場所か、という問題は依然として残るが）。そのほかの地名も同様に代表点を与えることが行われるが、その点をどこに置くかが必ずしも決まっているわけではなく、２つの地名が同じものかを空間上の位置をたよりに検討する場合などに支障が生じてくる。

地名に基づいて研究資源を検索しようとする場合は、その地名が、史資料の内容で使われている地名なのか、史資料の作成や保管などの場所なのかといった違いを、どのように分けて、どのようにメタデータに記述するのかが課題となる。メタデータの仕様が研究資源ごとに異なると、複数の研究資源を連携させようとする資源共有化が難しくなってくる。

最後に、地名データを公開する際に問題となるのが、人権問題への配慮である。特に歴史地名においては被差別部落などの地名を含むことがあり、場合によってはその場所を容易に特定できる情報が含まれる。こうした地名を含むデータの取り扱いについては、利用目的や公開方法に一定の配慮が必要となる。

資源共有化における地名の役割と時空間への展開

関野 樹（総合地球環境学研究所）

なぜ「地名」か？

ここでの「地名」の範囲

・土地につけられた名称
・山・河・・・などの地形名に対して用いられる傾向
・文化圏・工業地域・都市圏なども地名の一種
・駅名・線路名・家屋名・・・などは含めないのが普通
（『地理学事典（改訂版）』より）

→ 空間と結び付けられる名称
（土地、地物すべてを含める）

空間情報としての地名－時間との対比

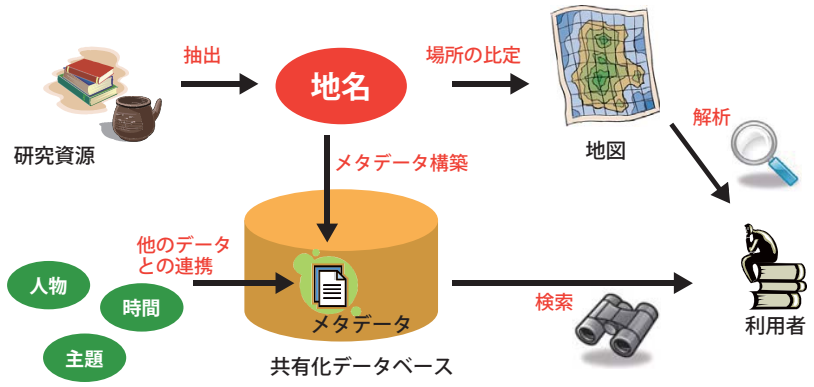
空間	時間
・GPSを携帯しているケースは稀 （少なくとも最近までは） ・史資料に現れるのは地名 ・地名はそのままでは空間座標にならない （位置関係が分からない）	・ほとんどの人が時計を持っている ・資料中でも日時で表現される ・日時はそのまま時間軸に落とせる

空間情報の利用の拡がり

・地理情報システム（GIS）の発達と普及
・Google Mapsなどによる手軽な利用環境

地名の処理にかかる多くの課題

地名の使われ方の一例



まずは、それぞれの過程での課題を概観

地名の抽出

どれが地名か

吉田

地名
・京都市左京区吉田・長野市吉田・富士吉田市
・吉田駅（近鉄けいはんな線）・吉田駅（JR弥彦線）
・吉田神社（京都市左京区ほか）・吉田城（豊橋市ほか）
・吉田病院（旭川市）・吉田食堂（京大内・会津若松市）

人名
団体名
・吉田東伍（地理学者）・吉田茂（政治家）ほか多数
・吉田カバン・吉田酒造

モノの名前 ・吉田うどん

吉田食堂写真：http://gekikaracurry.blog123.fc2.com/blog-entry-219.html

どこまでが地名か？

あいまいな地名、相対的な地名
・××のあたり
・××の北〇〇km
・△△のとなり

旧字・異体字 吉吉 田冈田田

「グリフウィキ」より

現状の多くは・・・

・手作業での抽出
・地名辞書とのつきあわせ
・それぞれの基準

メタデータの中での地名

何の地名か？

- 内容で扱われている場所
- 作成場所
- 保管場所

メタデータ構築

- どのように記録するのか

Coverage
- 資源のコンテンツが対象としている時間的・空間的な範囲。
- 時間情報と空間情報は分けて対応付ける。

＜時間情報＞
- 資源に記述されている時間的情報はここに対応させる。
- 例：文書DBの日付
- 資源の作成や公開に関する日付はDateに対応させる。
- 資源に記述されている時間的情報が、その資源の作成等の時間情報でもある場合は、Dateにも対応させる。

＜空間情報＞
- 例：論文DBの対象地域、錦絵DBでの描かれている地名
- 資源が製作あるいは使用された地理的情報もここに記述する。
- 例：製作地、使用地、所在地、収集地

Where
- 場所に関する情報。
- Coverageの空間情報にマッピングされるデータ項目は、これに該当する。
- 実在しない架空の地名か記載されたデータ項目でもよい。
- DCのSourceやRelationにマッピングされる空間情報もここに対応付ける

「NIHUメタデータマッピング規則」より

何に使われるか？

研究資源の検索
- 同じ地名
- より上位の階層の地名
- 同じ場所
- 近隣の場所

研究資源の利用
- 地図上に表示展開
- GISを使った解析

地名と空間の関係を考える

- 地名だけでできること
- 空間を考えないとできないこと

地名と空間

地理識別子の一つとしての地名

地理識別子
- 地名
- 郵便番号
- 電話番号（市外局番）
- 市町村コード

座標
- 緯度・経度
- UTM
- 地心直交座標系
- 平面直角座標系

地理情報
ー地理識別子による空間参照
ISO 19112 (JIS X 7112)

地理情報
ー座標による空間参照
ISO 19111 (JIS X 7111)

空間、地物（の位置）

地名と空間

地理識別子

地名　日本経緯度原点

（東京都港区麻布台二丁目18番1）

座標

地心直交座標　X軸 −3,959,340.203m
Y軸　3,352,854.274m
Z軸　3,697,471.413m

緯度経度　北緯　35度39分29.1572秒
東経　139度44分28.8869秒

空間、地物（の位置）

地名が持ち得る情報

地名：吉田
階層
京都市 ＞ 左京区 ＞ 吉田
種類
地域名

座標
測地系
WGS84

別名？

地名

地名の変遷
場所の変遷

空間、地物
形状
点、線、面
種類　（特に地物の場合）
地域

地名辞書
データベース

検索・解析における地名と空間

地名による検索・解析		空間による検索・解析
吉田 検索		
対応しない	別の場所に同じ地名がある場合	対応する
対応する（種類の情報があれば）	違う種類で同じ地名がある場合	対応する（種類の情報があれば）
階層情報による（下の階層の・・・）	より広域や限定的な検索をする場合	位置関係、距離や面積による（近隣の・・・）
対応しやすい	場所が変わる場合（「御所」など）	対応しない
対応しない	名称が変わる場合	対応する
対応する	場所が比定できない場合（「邪馬台国」など）	対応しない
対応しない	地図に表示する	対応する
対応しない（相対的な位置関係は除く）	位置関係の解析	対応する

場所の比定にかかる課題

地名と実際の空間との関連付け

対応する場所（空間）を探す
- いつの時代の？
- 何に基づくのか（緯度経度だけとは限らない－絵図など？）

仮に、対応する場所（空間）が見つかったとしても・・・

地点
- どこを代表点にするのか？
- あいまいな場所の扱い

範囲や境界
- 境界をどこまで定めるのか（そもそも定める必要があるのか？）
- 範囲や境界が変わる場合
- 隣接関係

他の情報との関連

時間

- 地名の変遷
- 地名が使われた期間　など

その他

- 別名（他言語も含む）
- 郵便番号
- 関連する情報資源（Wikipediaなど）
- 人物
- 歴史上のイベント
- ほか


```
−<rdf:RDF>
 −<gn:Feature rdf:about="http://sws.geonames.org/1853051/">
    <rdfs:isDefinedBy rdf:resource="http://sws.geonames.org/1853051/about.rdf"/>
    <gn:name>Sakyō-ku</gn:name>
    <gn:alternateName xml:lang="ja">さきょうく</gn:alternateName>
    <gn:alternateName xml:lang="ja">Sa-kyō-ku</gn:alternateName>
    <gn:alternateName xml:lang="ja">Sakyō-ku</gn:alternateName>
    <gn:officialName xml:lang="ja">左京区</gn:officialName>
    <gn:featureClass rdf:resource="http://www.geonames.org/ontology#A"/>
    <gn:featureCode rdf:resource="http://www.geonames.org/ontology#ADM3"/>
    <gn:countryCode>JP</gn:countryCode>
    <wgs84_pos:lat>35.16563</wgs84_pos:lat>
    <wgs84_pos:long>135.79625</wgs84_pos:long>
    <gn:parentFeature rdf:resource="http://sws.geonames.org/1857906/"/>
    <gn:parentCountry rdf:resource="http://sws.geonames.org/1861060/"/>
    <gn:parentADM1 rdf:resource="http://sws.geonames.org/1857907/"/>
    <gn:parentADM2 rdf:resource="http://sws.geonames.org/1857906/"/>
    <gn:childrenFeatures rdf:resource="http://sws.geonames.org/1853051/contains.rdf"/>
    <gn:locationMap rdf:resource="http://www.geonames.org/1853051/sakyo-ku.html"/>
    <gn:wikipediaArticle rdf:resource="http://en.wikipedia.org/wiki/Sakyō-ku"/>
    <rdfs:seeAlso rdf:resource="http://dbpedia.org/resource/Sakyō-ku"/>
  </gn:Feature>
 −<foaf:Document rdf:about="http://sws.geonames.org/1853051/about.rdf">
    <foaf:primaryTopic rdf:resource="http://sws.geonames.org/1853051/"/>
    <cc:license rdf:resource="http://creativecommons.org/licenses/by/3.0/"/>
    <cc:attributionURL rdf:resource="http://sws.geonames.org/1853051/"/>
    <cc:attributionName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GeoNames</cc:attributionName>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2006-01-15</dcterms:created>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2014-03-08</dcterms:modified>
  </foaf:Document>
</rdf:RDF>
```


GeoNamesの例

共通の課題

文字に関する課題

- 旧字、異体字の扱い

データ公開に関する課題

- データの公開形式
- インタフェース
- 検索方法
- 人権問題に関連する地名

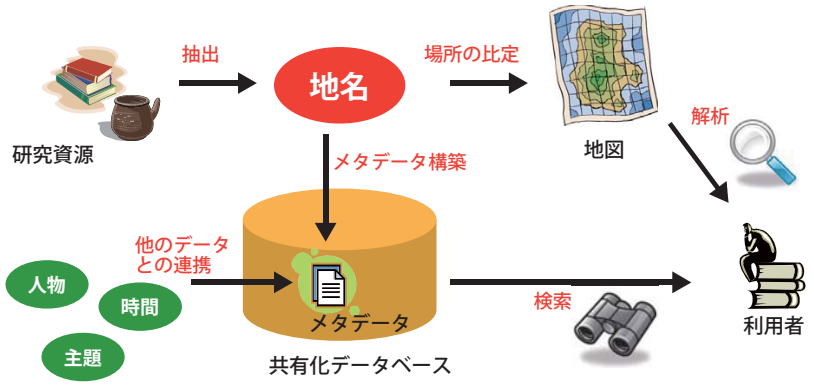
経度	緯度	地名	属性	備考	図幅	測量年	発行年月日	リスト番号	白黒	種類	発行年月日	都道府県名
140.8593	35.79932	鹿嶋灘	21		銚子	明36	M39/06/30	67-2-1	*	測図	19060630	茨城県
140.7875	35.78457	鹿嶋	3		銚子	明36	M39/06/30	67-2-1	*	測図	19060630	茨城県
140.5598	35.96613	行方	3		鹿島	明36	M39/06/30	67-5-1	*	測図	19060630	茨城県
140.5965	35.9757	北浦	23		鹿島	明36	M39/06/30	67-5-1	*	測図	19060630	茨城県
140.5912	35.97763	水原洲吠崎	46		鹿島	明36	M39/06/30	67-5-1	*	測図	19060630	茨城県
140.5993	35.96631	爪木ノ鼻	46		鹿島	明36	M39/06/30	67-5-1	*	測図	19060630	茨城県
140.6266	35.93267	鰐川	22		鹿島	明36	M39/06/30	67-5-1	*	測図	19060630	茨城県

人間文化研究機構資源共有化事業で提供している地名辞書データの例
（CSV, Excel, KML, シェープファイルで県ごと）

最後に

総合討論へむけて

- 各研究グループや研究者の取り組み、要望などの共有
- 技術面、運用面などでの課題の整理
- 今後の方向性（何ができるか・できそうか）



地名情報処理環境 GeoNLP の紹介と歴史的な地名に関する課題

北本 朝展（国立情報学研究所）

1．はじめに

地名は地理情報における重要な一要素であるにもかかわらず、従来の地理情報システム（Geographic Information Systems / GIS）では扱いづらい存在であった。地理情報システムは、座標系という幾何学的な空間で、多様な地物を一様に扱えることが大きな特徴である。しかし地名は概念であって、そこで問題となるのは位相学的な空間である。例えば「○○の一部」といった関係性も重要であるし、特に歴史地名の場合はその概念が指し示す中心点や境界も不明確である。こうした状況は GIS のモデルと整合しないにもかかわらず、従来の歴史地理情報処理では GIS を無理やり使うという方法論が主流であった。そこで、我々が推進するプロジェクト GeoNLP は、地名を対象とするオープンな地名情報システム（Toponym Information Systems / TIS）を構築し、GIS とは異なる体系で地理情報を扱える情報基盤を創生することを目指す。

２．地名の曖昧性

そもそも「地名とは何か？」という問題に統一された定義はないが、GeoNLP プロジェクトでは「位置に関する属性をもつ固有名（named entity）」と定義する。狭義の地名は、土地に付与された名前のみを扱うが、GeoNLP では広義の地名として、住所などの行政的な位置表現だけでなく施設などの人工物に付与された名前も対象とし、こうした名前を固有名と位置（緯度・経度）との対応表である地名辞書にあらかじめ登録する。この問題設定において地名情報処理の課題とは、1) 自然言語文中に出現する地名を見つけ、2) これに対応する地名辞書のエントリを検索し、3) その位置を結果として示す、と整理できる。

ここで技術的に困難な挑戦課題となるのが、地名という概念の「曖昧さ」の処理である。上記に示した 3 段階の処理では、それぞれ対処すべき曖昧さの課題が異なっている。まず 1)では、自然言語文の中から地名文字列を切り出さねばならないが、日本語の場合は単語が空白文字等で区切られていないため、形態素解析等を用いて単語に分割せねばならず、ここに曖昧さが生じる。次に 2)では、ある地名文字列に対して地名辞書に複数のエントリが存在する場合が多々あり、そのどれを選ぶかという点に曖昧さが生じる。この曖昧さを多義性（ambiguity）と呼ぶ。最後に 3)では、地名に対する位置をどう定義するかという点に曖昧さが生じる。代表点および境界を明示的に定義できる場合は問題ないが、日常的

に使う地名や歴史的な地名の場合はどちらも曖昧な場合が多々ある。この曖昧さを漠然性（vagueness）と呼ぶ。こうした曖昧さを解決できる地名情報基盤として GeoNLP を誰でも簡単に使えるようにすることが、プロジェクトの最終的な目標である。

現在のところ GeoNLP は多義性の問題を中心的に扱っており、曖昧性解消のために自然言語文に地名が出現する際の文脈情報を利用する方式を研究している。特に地名の共起に注目し、同じ種類の地名や距離的に近い地名が出現しやすいとの仮定のもと、共起の起こりやすさを評価する関数を用いて共起しやすい候補を選んでいる。ただし評価関数は改善の余地が大きく、文脈情報が不足している場合に適切な候補が選ばれにくいという問題がある。さらに根本的な問題として、precision と recall のジレンマという問題は今後の重要な研究課題である。地名辞書のエントリを増やしていけばいくほど、確かに抽出可能な地名は増える（recall は上昇する）ものの、地名ではない単語を誤って地名として抽出してしまう場合も増える（precision が低下する）ことが避けられない。これはこの種の処理において原理的に避けられない問題であり、特に歴史地名のように一般名詞的な地名が多い地名辞書を使えば、より深刻な問題となることも予想される。実用的な観点から precision と recall のバランスをコントロール可能な手法が必要である。

３．歴史地名に関する課題

GeoNLP を歴史地名に適用する際に生じる重要な問題として、1) 古文用形態素解析器の導入、2) 地名の変遷のデータベース化、3) 地名の有効期間の活用、を論じておきたい。1)は GeoNLP が用いる形態素解析器（MeCab）が現代文に最適化されており、古文には必ずしも適用できないという問題を指すが、これは形態素解析器の差し替えを可能にすることで部分的には対処可能である。2)は地名の変遷をどう辞書に登録するかという問題を指すが、これは地名辞書の本体に登録するのではなく、我々が開発する GeoLOD に地名の関係を Linked Data として登録する方法で解決したい。最後に 3)は、地名の変遷を登録するだけでなく、地名が主に出現する期間を属性として定義することで、期間を指定すると特定の地名が選ばれやすくなる評価関数を定義する方法を考えている。いずれも、実際の歴史的なテキストで試しながら改善していく必要があり、共同研究などを進めていきたい。

GeoNLP はオープンソースとオープンデータという、オープンな方針のもとに地名情報処理環境を実現することを目標とする。プロジェクトで開発したソースコードやデータは、ウェブサイト https://geonlp.ex.nii.ac.jp/ から入手できる。しかしこの問題は我々だけで解決できるものではなく、ソフトウェアとデータの改善はコミュニティベースで進めていきたいと考えている。関心のある方々は、ぜひご協力いただきたい。

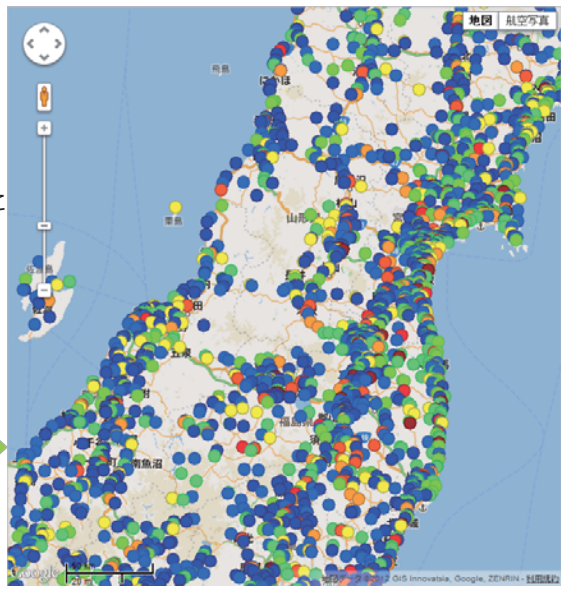
地名情報処理環境 GeoNLPの紹介と 歴史的な地名に関する課題

国立情報学研究所
北本 朝展
https://geonlp.ex.nii.ac.jp/

テキストのジオタギング

宮城県などによると、気仙沼市や多賀城市などで大規模な火災が発生。気仙沼市は津波で市街地の3分の1が水没し、気仙沼湾に浮かぶ大島の4集落が壊滅状態となった。女川町もほぼ壊滅という。岩手県では陸前高田市がほぼ壊滅し、山田町や宮古市の一部が水没。福島県では南相馬市の約1800世帯が壊滅状態という。宮城県警は東松島市のJR野蒜駅付近で、この脱線した列車から乗客ら9人、別の列車からも11人を救出した。

- テキストから地名の候補を抽出。
- 同綴異義地名をさらに解析（例えば岩手県「宮古」市と沖縄県「宮古」島）。
- テキストの周囲に現れる他の地名を手掛かりに、複数候補の中から最適候補を決定。
- 地名辞書と照合して地名の緯度経度を検索。
- 地図上にマッピング。



地名解析の問題

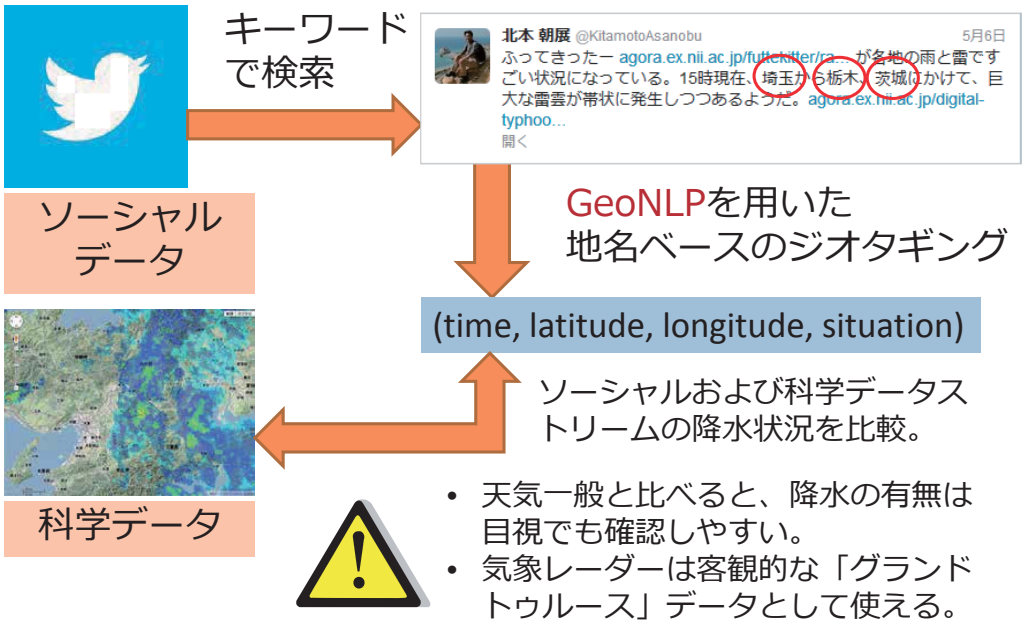
横浜は雨だよ。川崎は雪？

横浜は好きだよ。川崎は好き？

- 神奈川県横浜市と神奈川県川崎市？
- 青森県横浜町と福岡県川崎町？
- 横浜さんと川崎さん？
- **各種の曖昧さを解消し実世界にGrounding。**

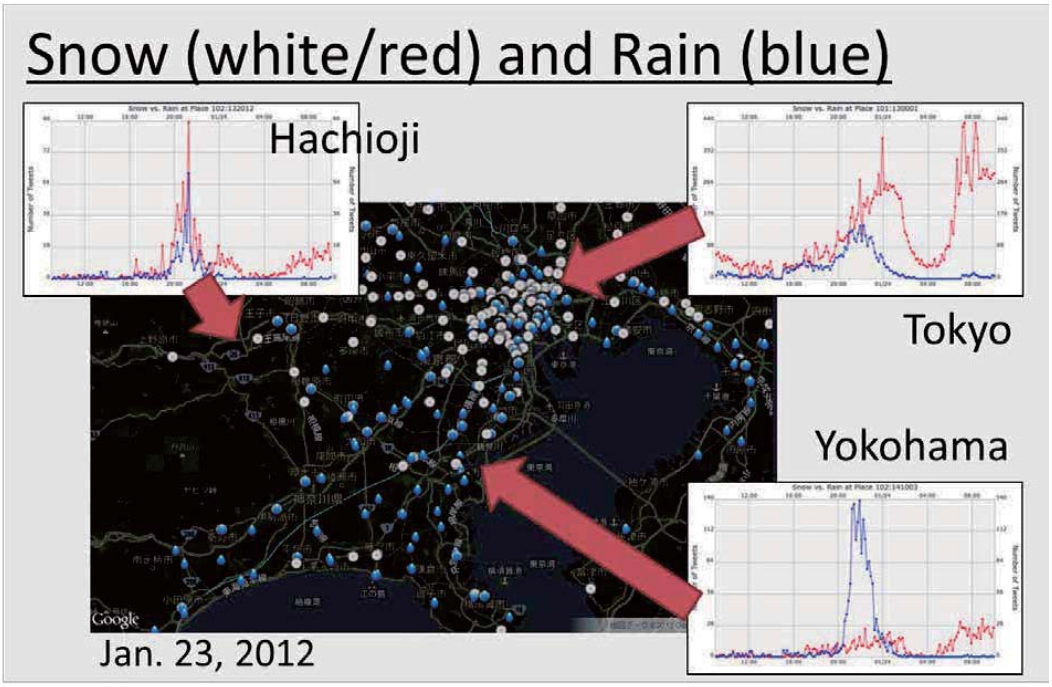
2014/7/19 地名情報技術研究会 3

ソーシャル気象観測



- 天気一般と比べると、降水の有無は目視でも確認しやすい。
- 気象レーダーは客観的な「グランドトゥルース」データとして使える。

2014/7/19 地名情報技術研究会 4



ふってきたー: http://agora.ex.nii.ac.jp/futtekitter/

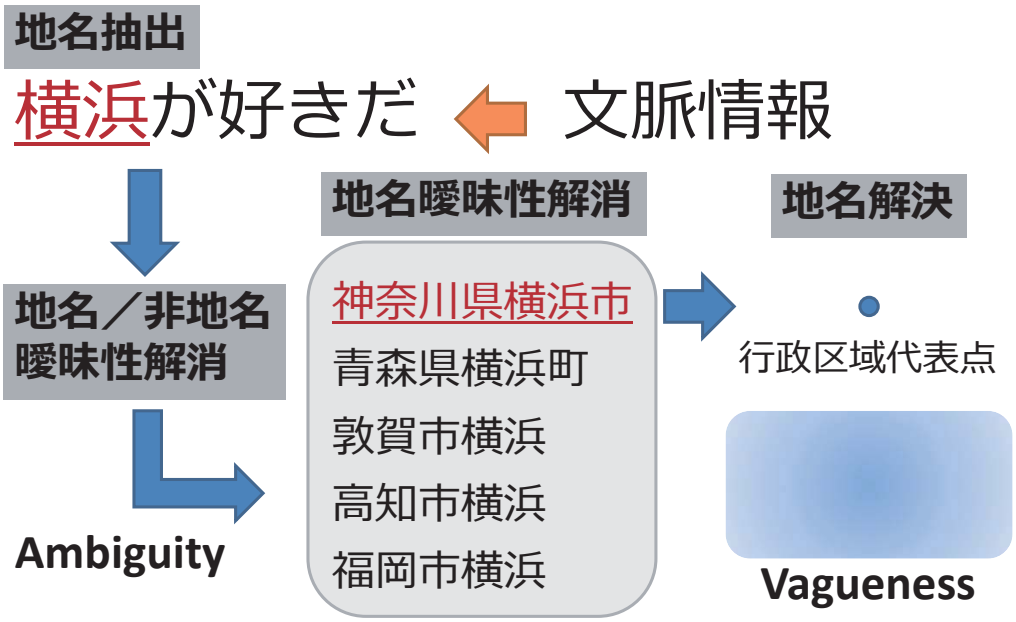
地名とは何か？

- GeoNLPプロジェクトでは「位置に関する属性をもつ固有名」と定義する。
  1. 行政区域などの公的な地名。
  2. 階層構造にしたがう住所。
  3. 山などの地物に付与された自然地名。
  4. 人工物に付与された施設名。
  5. 日常的によく使われる地域名。
  6. （将来）郵便番号等のコード系地名。
  7. （将来）道路、鉄道など線状の地物。
  8. （未定）屋内の区画名、店舗名等。

地名情報システムGeoNLP

- 地理情報システム（GIS）は、座標系という幾何学的なモデルの上で、多様な種類の**地物**を扱う。
- 地名情報システム（TIS）は、関係性という位相学的なモデルの上で、不明確な境界をもつ**概念**を扱う。
- GeoNLPは地名情報システムとして、GISとは異なる体系で空間情報を扱う。

地名に関する曖昧さ



地名／非地名の曖昧さ

1. 地名と人名のどちらか？
   – 例：「横浜が好き」横浜市？横浜さん？
2. 地名と一般名詞のどちらか？
   – 例：「風呂に行く」地名の可能性ほぼゼロ。
3. 歴史的経緯を考えても、地名・人名・一般名詞の語彙の重複は大きい。
4. 一般名詞を地名とする誤りは、結果のサプライズが大きく信頼度を低下させる。

2014/7/19 地名情報技術研究会 9

地名に関する2種類の曖昧さ

概念定義（Vagueness）	指示対象（Ambiguity）
• 「東京に行く」の「東京」は、どこの範囲を指す？ • 「ディズニーランド」は東京？「奥多摩」は東京？ • 行政区域なら曖昧さはないが、日常語の意図する範囲は公式区域と不一致。	• 「横浜に行く」の「横浜」は、全国の横浜のうちどれ？ • 神奈川県横浜市だけでなく、青森県横浜町など、複数の候補がある。 • 「第一小学校」など、施設名でも多くの同綴異義語が存在。

2014/7/19 地名情報技術研究会 10

地名のVagueness

Physical Grounding

- 公式の行政区域と、住民等がもつ感覚的な境界が、一致しない場合がある。
- 地形などの物理的な特徴を基準として境界を推定する。
- 川や盆地、高原など、地形に特徴がある場所なら有用である。

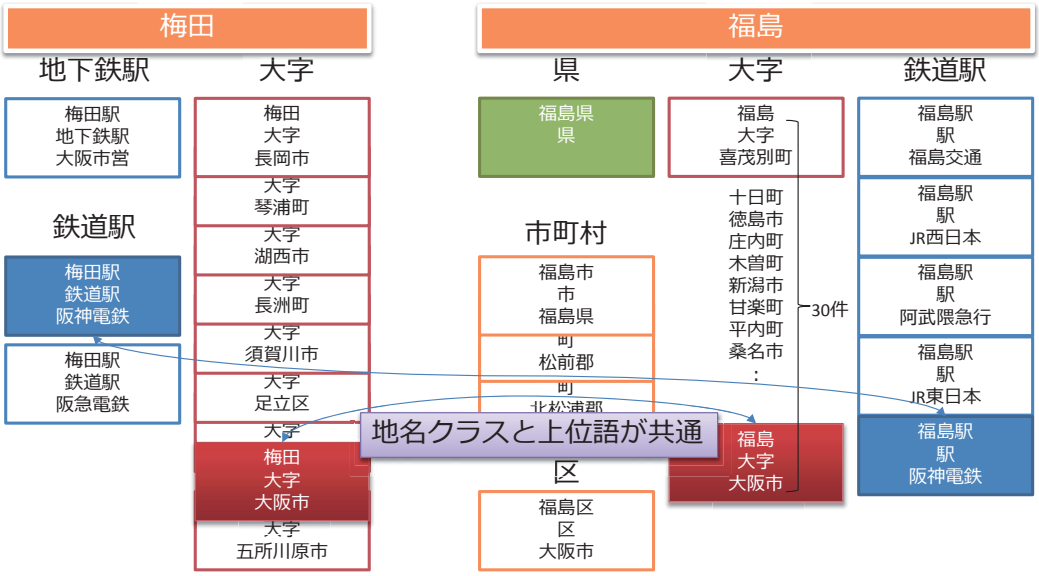
Social Grounding

- 人々（コミュニティ）の行動から境界を推定する。
- 写真とPOIの共起から、どこがひとまとまりの地域かを推定。
- 旅行者等のヒューマンプローブデータから、場所のモデルを構築。
- SNSでの地名言及から、人々の地理的感覚を推定。

2014/7/19 地名情報技術研究会 11

地名のAmbiguity

「今日は梅田から福島に行きます。」の「梅田」と「福島」の組み合わせは12×40通り。



2014/7/19 地名情報技術研究会 12

スコア加算方式

$Score = 1500\varsigma_1(n_{fsb})+500\varsigma_1(n_{psb})+1500\varsigma_1(n_{chd})+2000\varsigma_1(n_p)+200\varsigma_1(n_{cls})+100\varsigma_1(n_{dic})+\beta_{spa},$

ただしς_1はゲイン１のシグモイド関数

梅田	梅田駅 地下鉄駅 大阪市，御堂筋線	梅田駅 鉄道駅 阪神電鉄，本線	梅田駅 鉄道駅 阪急電鉄，神戸線	梅田 大字 大阪市，北区	梅田 大字 長岡市	梅田 大字 五所川原市
n_{fsb}:上位語完全一致数	0	1 阪神本線に福島駅がある	0	0	0	0
n_{psb}:上位語一部一致数	1	2 阪神電鉄，本線の2つが一致	0	1 大阪市に福島という字があるが北区ではない	0	0
n_{chd}:子地名数	0	0	0	0	0	0
n_p:親地名数	0	0	0	0	0	0
n_{cls}:同クラス数	0 地下鉄駅と鉄道駅は別クラス	1	1	1	1	1
n_{dic}:同辞書数	1	1	1	1	1	1
B_{spa}:空間ボーナス	0	0	0	0	0	0
スコア計	277	1211	138	369	138	138

現行手法の問題点

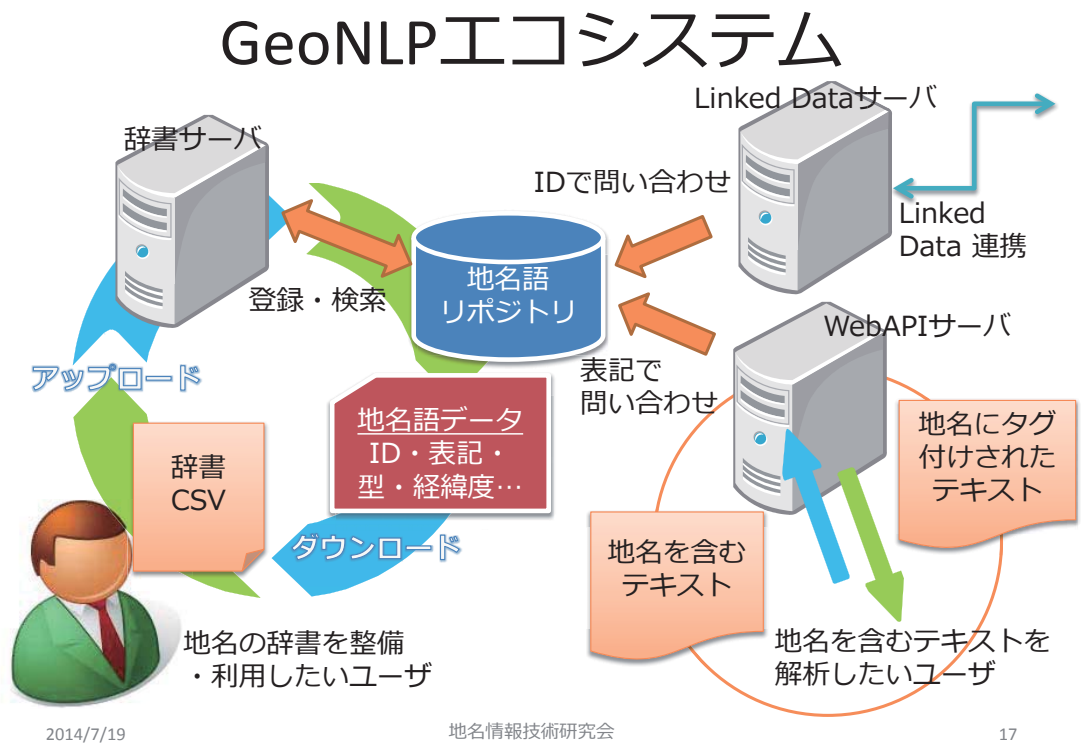
1. 上位語やクラスといった属性に依存
   - 属性の定義は文脈非依存、評価関数は文脈依存。
2. 評価関数の調整が困難
   - コーパスがあれば学習できるが、コーパスがない。
3. テキスト中の出現位置を考慮せず
   - テキスト中の距離が近い語ほど影響が大きい？
4. 地名の事前知識（知名度等）を考慮せず
   - 「福島」→多くの人は「福島県」を想起。
5. 地名語以外の語との共起を考慮せず
   - 「今日は**梅田**にいますが、週末は**震災ボランティアで福島**に行きます。」→福島は「福島県」か「福島市」、大阪ではない。

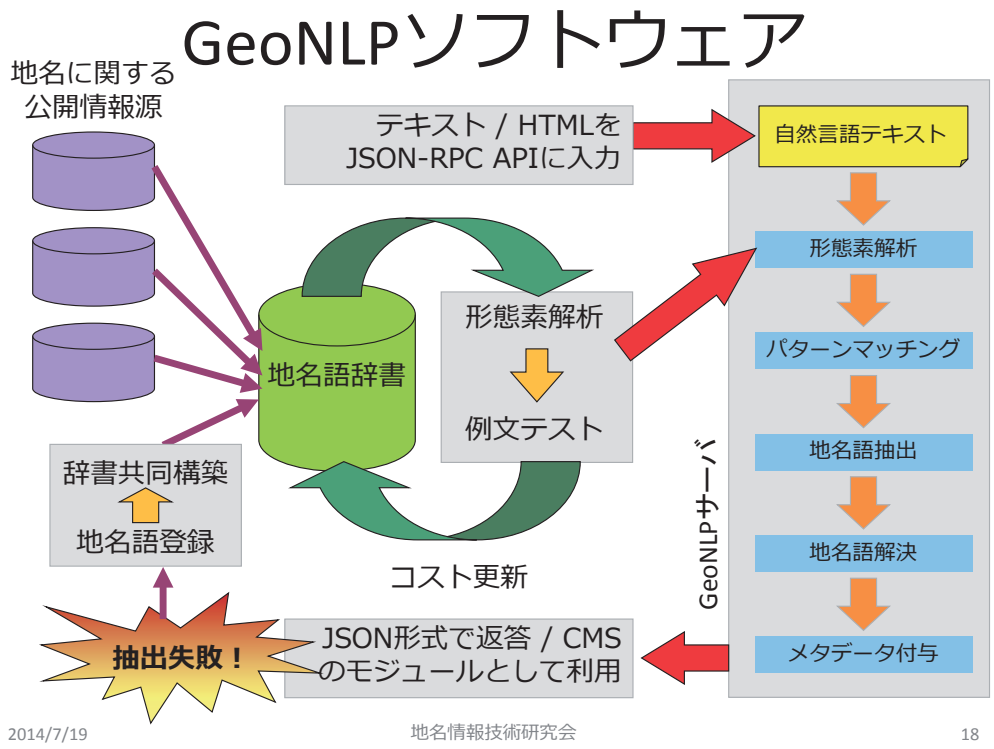
PrecisionとRecallのジレンマ

- より多くの地名を抽出する性能（Recall）を向上させると、地名だけを抽出する性能（Precision）が低下する「ジレンマ」。
- 地名語辞書の拡張は、Recallを向上させる裏で、Precisionを犠牲となる（特に小字は一般名詞的な地名が多く困る）。
- 技術によって、Precision-Recall曲線自体を向上させるしかない。

GeoNLPの構成

- **GeoNLP Software**
  - オープンソースソフトウェアとして配布。
  - ユーザが自分のサーバにインストールして利用。
- **GeoNLP Data**
  - 地名語辞書を作成してアップロード。
  - サイトからダウンロードして多目的に利用。
- **GeoNLP Service**
  - ソフトウェア機能（の一部）をAPI経由で利用。
  - Linked Open Dataとして他データと統合。


GeoNLPエコシステム
Linked Dataサーバ
辞書サーバ
IDで問い合わせ
Linked
Data 連携
地名語
リポジトリ
登録・検索
WebAPIサーバ
アップロード
表記で
問い合わせ
地名語データ
ID・表記・
型・経緯度…
地名にタグ
付けされた
テキスト
辞書
CSV
ダウンロード
地名を含む
テキスト
地名の辞書を整備
・利用したいユーザ
地名を含むテキストを
解析したいユーザ
2014/7/19
地名情報技術研究会
17


GeoNLPソフトウェア
地名に関する
公開情報源
テキスト / HTMLを
JSON-RPC APIに入力
自然言語テキスト
形態素解析
パターンマッチング
地名語辞書
形態素解析
例文テスト
地名語抽出
辞書共同構築
地名語登録
GeoNLPサーバ
地名語解決
コスト更新
抽出失敗！
JSON形式で返答 / CMS
のモジュールとして利用
メタデータ付与
2014/7/19
地名情報技術研究会
18

API応答はGeoJSONに準拠


```
{ "surface" : "神保町",
  "geonlp_id" : "STc3x1",
  "geo" : {
    "type" : "Feature",
    "geometry": {
      "type": "Point",
      "coordinates" : [139.757606, 35.695966]
    },
    "properties" : {
      "name" : "神保町 駅",
      "kana" : "じんぼうちょうえき",
      "icon" : "http://maps.google.co.jp/mapfiles/ms/icons/rail.png",
      "entry_id" : "9939",
      "dictionary_id" : 4,
      "body" : "神保町",
      "prefix" : [""],
      "suffix" : ["駅", ""],
      "body_kana" : "じんぼうちょう",
      "prefix_kana" : [""],
      "suffix_kana" : ["えき", ""],
      "ne_class" : "Station",
      "hypernym" : ["東京メトロ", "半蔵門線"],
      "priority_score" : 0,
      "latitude" : 35.695966,
      "longitude" : 139.757606,
      "address" : "東京都千代田区神田神保町二丁目2",
      "code" : {"ekidata.jp":"2800807", "TokyoMetro":"Z-07"},
      "CityCode" : "13101"
    }
  }
}
```

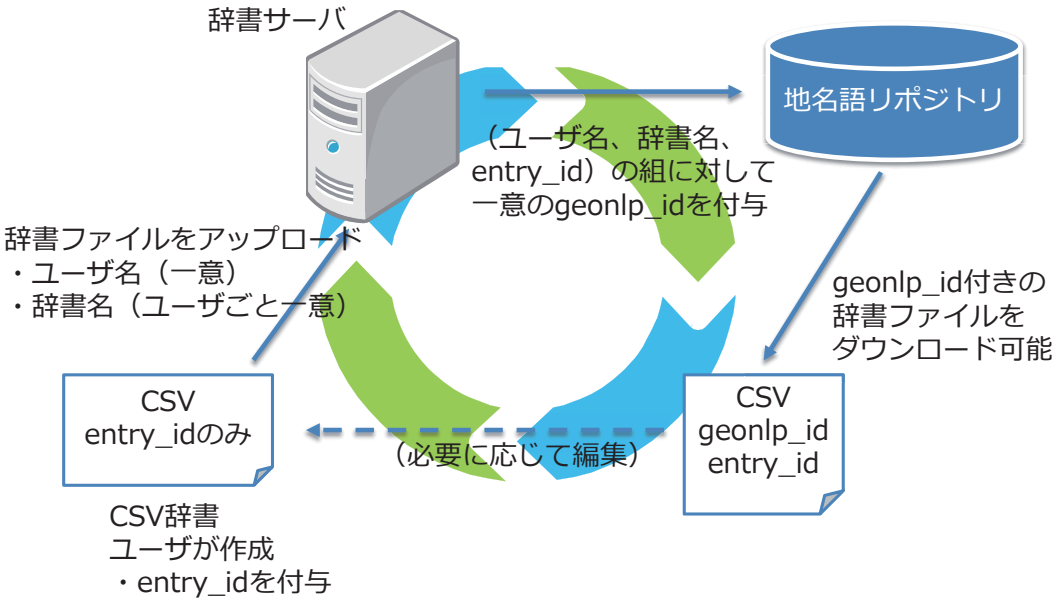

- WebAPI応答はGeoJSON準拠
- 地図へのオーバレイが容易に。
- 既存のコンバータ（SVG変換等）も活用できる。

GeoNLPサービス

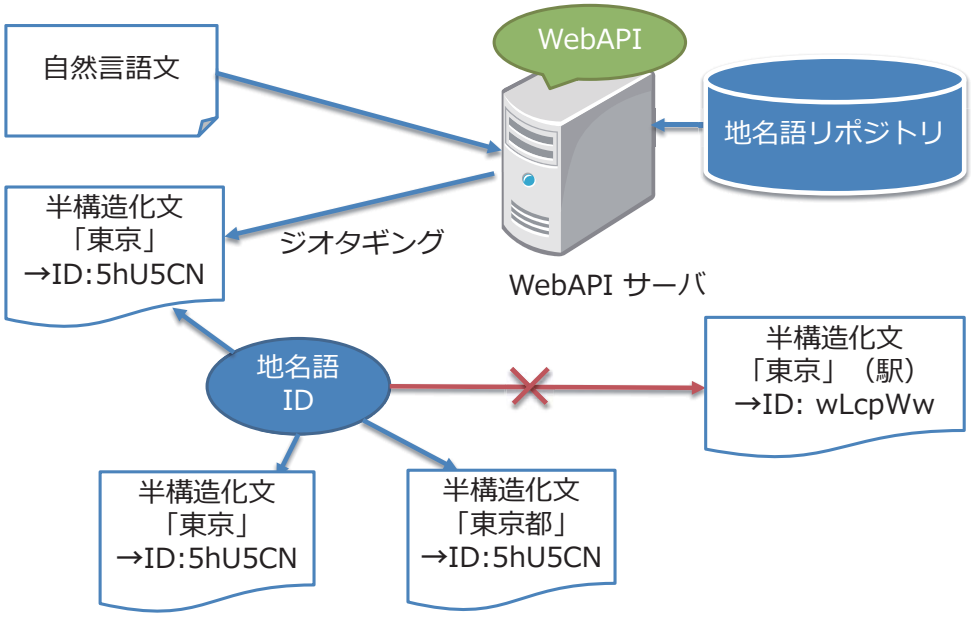
https://dias.ex.nii.ac.jp/geonlp/

任意の自然言語文を入力すると、自動的に解析して地図化

GeoNLP IDの付与

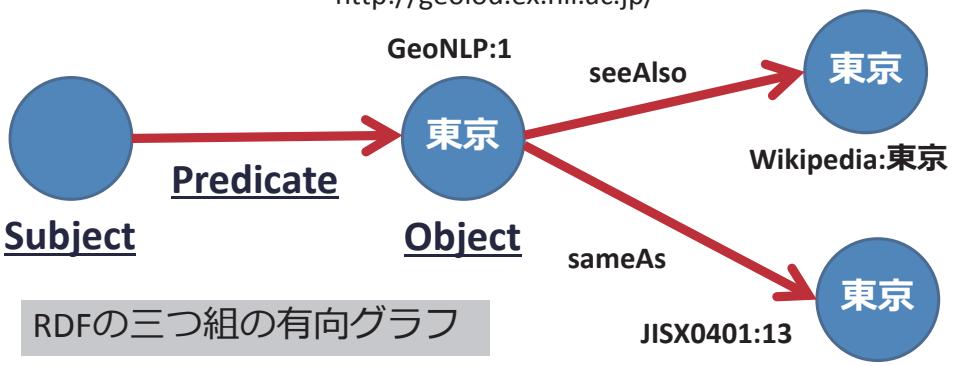


GeoNLP IDを用いたタグ付け



GeoLOD

http://geolod.ex.nii.ac.jp/



RDFの三つ組の有向グラフ

- Semantic Webの技術などを用いて、データを（URIで）リンクする仕組みを開発し、オープンなデータのネットワークを構築。

GeoNLPデータ

https://geonlp.ex.nii.ac.jp/

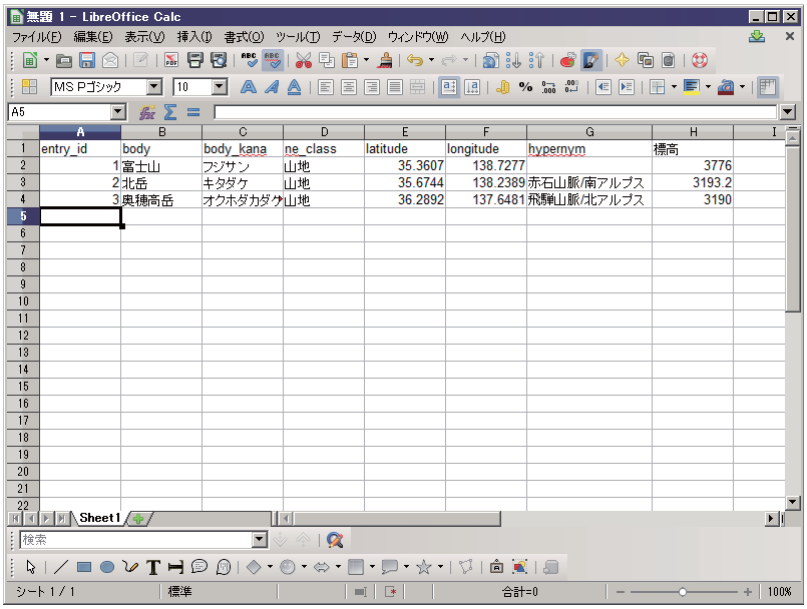
- 地名辞書を共有して、GeoNLPソフトウェアで活用。
- 行政・企業のオープンデータを地名辞書の形式に加工。
- 個人・グループの草の根的な地名辞書をオープン化。

利用ライセンス

ライセンスの種類	Creative Commons (CC)	Open Data Commons (ODC)
Public domain	CC0	PDDL (public domain dedication and license)
Attribution	CC-BY	ODC-BY
Attribution-ShareAlike	CC-BY-SA	ODbL (Open Database License)

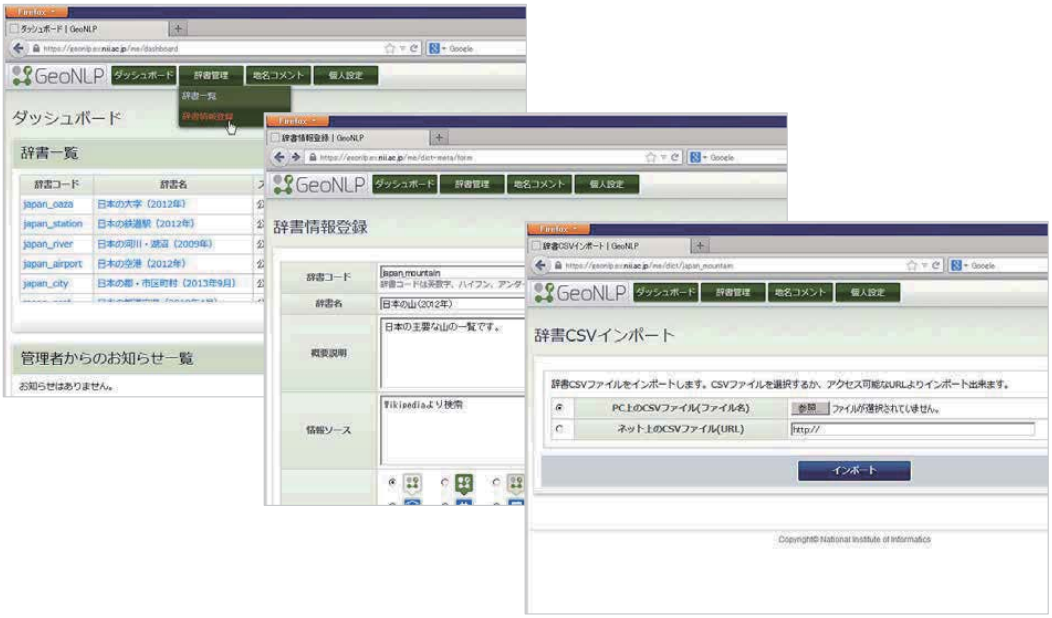
- 独自設定も可能だが、利活用を促進するため、制限が少ないライセンスを推奨。

地名辞書を作成してみよう



	A	B	C	D	E	F	G	H
1	entry_id	body	body_kana	ne_class	latitude	longitude	hypernym	標高
2	1	富士山	フジサン	山地	35.3607	138.7277		3776
3	2	北岳	キタダケ	山地	35.6744	138.2389	赤石山脈/南アルプス	3193.2
4	3	奥穂高岳	オクホダカダケ	山地	36.2892	137.6481	飛騨山脈/北アルプス	3190

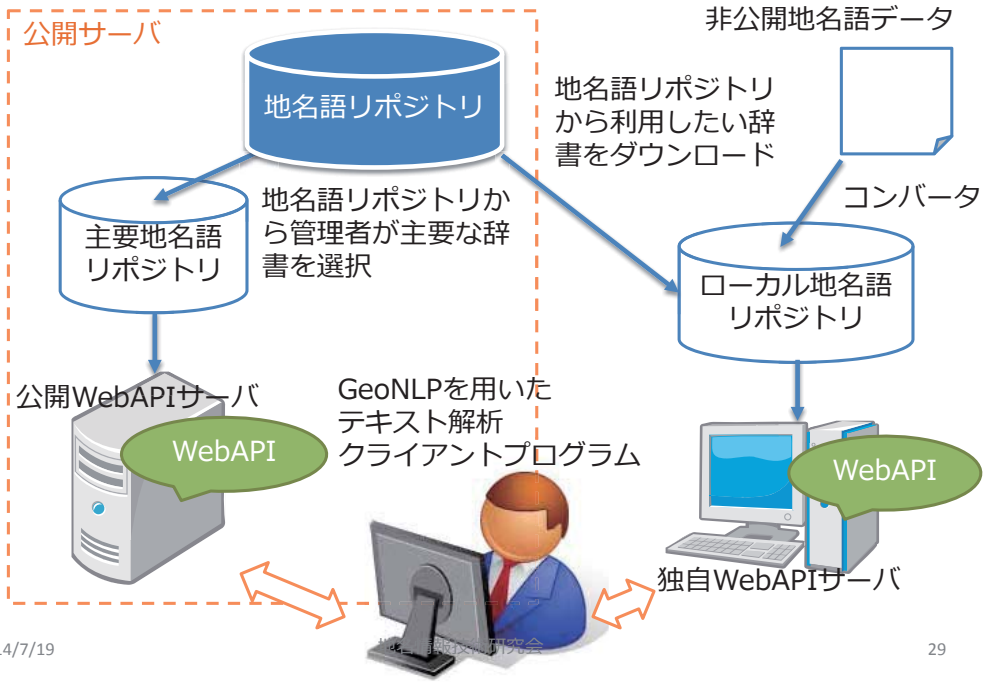
地名辞書を作成してみよう



地名語辞書のスキーマ

フィールド名	フィールド英語名	必須種別・複数の別	型・制約	例
■ ID（Identifier）				
GeoNLP ID	geonlp_id	必須	char(6), primary key, unique	0sUwdt
エントリID	entry_id	必須（空欄不可）	varchar(255), not null	1
辞書ID	dictionary_id	必須	int	41
■ 表記情報(Notation)				
原型	body	必須（空欄不可）	varchar(255), not null	札幌
接頭辞	prefix	推奨・複数可	varchar(255)	“”
接尾辞	suffix	推奨・複数可	varchar(255)	飛行場/空港
読み	body_kana	推奨	varchar(255)	さっぽろ
接頭辞読み	prefix_kana	拡張・複数可	varchar(255)	“”
接尾辞読み	suffix_kana	拡張・複数可	varchar(255)	ひこうじょう/くうこう
■ 関係情報(Relation)				
固有名クラス	ne_class	必須	varchar(255), not null	
上位語	hypernym	推奨・複数可	varchar(255)	日本/北海道
優先スコア	priority_score	拡張	int, default=0	0
■ 属性情報(Attribute)				
代表点緯度	latitude	必須	varchar(32)	43.1175
代表点経度	longitude	必須	varchar(32)	141.381389
住所	address	推奨	varchar(255)	札幌市東区丘珠町
地名コード	code	拡張・複数可	varchar(255)	IATA:OKD/ICAO:RJCO
有効期間（開始）	valid_from	拡張	date	“1942-09”
有効期間（終了）	valid_to	拡張	date	“”

公開・非公開辞書の統合解析



歴史的な地名への適用

1. GeoNLPは形態素解析に基づく。
   - 古文用形態素解析器はあるか？
2. GeoNLPは地名辞書を必要とする。
   - GeoNLP形式に変換する必要あり。
3. 地名の有効期間をどう活用するか？
   - 仕様には入っているが、未実装。
   - 地名の変遷をモデル化したLinked Data。
4. 地名の境界を返す機能は今後の予定。

歴史的な地名への課題

1. 歴史地名の地名語辞書をアップロード、共有し、GeoNLPで分析することは可能。
2. 有効期間の扱いについては、実際のニーズをお持ちの方と議論したい。
3. 地名の変遷、地名の境界はGeoNLPの外側。GeoLODの拡張として扱いたい。
4. 古文用形態素解析器に差し替えるための機能が新たに必要？

謝辞

- GeoNLPの研究・開発は、相良毅氏（株式会社情報試作室）との共同研究によるものです。
- GeoNLPウェブサイトの開発には、（株）トライアックスの協力を得ました。
- JSTさきがけ、地球環境情報統融合プログラム、国立情報学研究所共同研究費等の支援を受けました。

https://geonlp.ex.nii.ac.jp/

GeoNLP

地名データのセマンティック Web への応用

京都大学地域研究統合情報センター
原　正一郎

１．地名辞書データベース構築の経緯

PNC2001 に出席した際に、米国の歴史研究者から日本の歴史地名に関するデータベースの情報提供を求められた。本人は日中交易史を研究していたそうで、史料に現れる日本の地名を地図上で同定することに難儀しているとのことであった。その当時、すでに米国では Alexandria Digital Library Gazetteer などのフリーの地名データベースがオンライン公開されていたが、日本において該当するような地名データベースは存在していなかった。そこで H-GIS 研究グループにおいて、掲載されている地名数が多く、地名辞書として一定の評価が得られており、かつ著作権の切れている史料を検討した結果、吉田東伍編纂による「大日本地名辞書」のデジタル化を試みることとなった。

2005 年 11 月 17 日に、本稿執筆者が在職していた国文学研究資料館と吉田文庫との間で、電子化された「大日本地名辞書（索引）」の構築に関する覚書を締結し、「人間文化研究機構連携研究　日本とユーラシアの交流に関する総合的研究　『湿潤アジアにおける「人と水」の統合的研究』代表：秋道 智彌（総合地球環境学研究所）」の経費により、大阪国際大学桶谷教授の監修のもとで、データベース化を開始した。

大日本地名辞書を手始めに、これまでに延喜式、寺院名鑑、旧５万分の 1 地図から地名を抽出し、地名データの集積を進めてきた。現時点において、大日本地名辞書：49,557 件、式内社：2,842 件、寺院：78,588、旧 5 万分の 1：192,122 の計 323,109 地名を収録している。

地名辞書のデータ構造は図 1 のようになっている。

1	2	3	4	5	6	7	8	9a	9b	9c	10a	11a	10b	11b	12	13	14	15	16	17
番号	地名	ヨミ	国	郡	郡ヨミ	形状	位置記述法	現在地名a	現在地名b	現在地名c	緯度	経度	緯度	経度	現在名称	巻	ページ	大項目名	地名属性	備考
11997	笠木山	カサギ	美濃																	
38325	飛州	ヒ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	509	飛騨国	2	
38385	飛騨山脈	ヒダ	飛騨			1	1	岐阜県			36.17.21	137.38.52				5	509	飛騨国	33	
38399	飛騨国	ヒダ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	509		2	
38400	斐陀国	ヒダ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	509	飛騨国	2	
38356	飛騨	ヒダ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	510	飛騨国	2	
38377	飛騨川	ヒダ	飛騨			2	2	高山市	美濃加茂市		36.5.8	137.32.34	35.26.43	137.3.8		5	510		22	
38384	飛騨高原	ヒダ	飛騨			1	1				36.8.34	137.15.14				5	510	飛騨川	34	
38388	飛騨谷	ヒダ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市	加茂郡	中津川市	35.39.41	137.9.34				5	510		9	郡と同類
38401	飛騨工	ヒダ	飛騨	飛騨谷	ヒダダニ											5	510	飛騨国		役職名？
9459	大船戸	オオフナト	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.39.46	137.9.46			大船渡	5	511	金山	8	
13047	金山	カナヤマ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.39.41	137.9.34				5	511		4	
13459	上麻生	カミアソウ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡七宗町			35.32.38	137.7.10				5	511		5	
14992	神田郷	カンダ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡東白川村			35.38.33	137.19.29				5	511		6	
15017	神土	カンド	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡東白川村			35.38.33	137.19.29				5	511		8	
15140	神淵	カンブチ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡七宗町			35.34.39	137.5.22			神渕	5	511		5	
16232	桐洞	キリボラ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.38.43	137.7.36			菅田桐洞	5	511		8	
17607	黒川	クロカワ	飛騨	飛騨谷	ヒダダニ	2	2	加茂郡白川町	加茂郡白川町		35.35.47	137.19.12	35.35.9	137.11.55		5	511	白川	22	
21287	笹洞	ササボラ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.38.42	137.6.23			菅田笹洞	5	511	桐洞	8	
21651	佐見	サミ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡白川町			35.40.34	137.16.15				5	511	白川	5	
22734	七宗山	シチソウ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡七宗町			35.35.53	137.8.22				5	511		31	

図１．地名辞書データベースの構造

データは XML で記述されている。以下に例を示す。

```
<item pid='26682'><kuni>山城</kuni><gun>上京</gun><gunyomi>カミキョウ</gunyomi>
        <placename reading='ソウコクジ>相国寺</placename>
        <pname1>京都市上京区</pname1><pname2></pname2><pname3></pname3>
        <shp>1</shp><loc>1</loc><lat>35.2.0</lat><long>135.45.46</long>
        <lat2></lat2><long2></long2><altname></altname><parent></parent>
        <attribute>13</attribute><page>23</page><remarks/>
</item>
```

データベースは XML サーチエンジンと Web API の組み合わせである。初期の検索画面例を図 2 に示す。

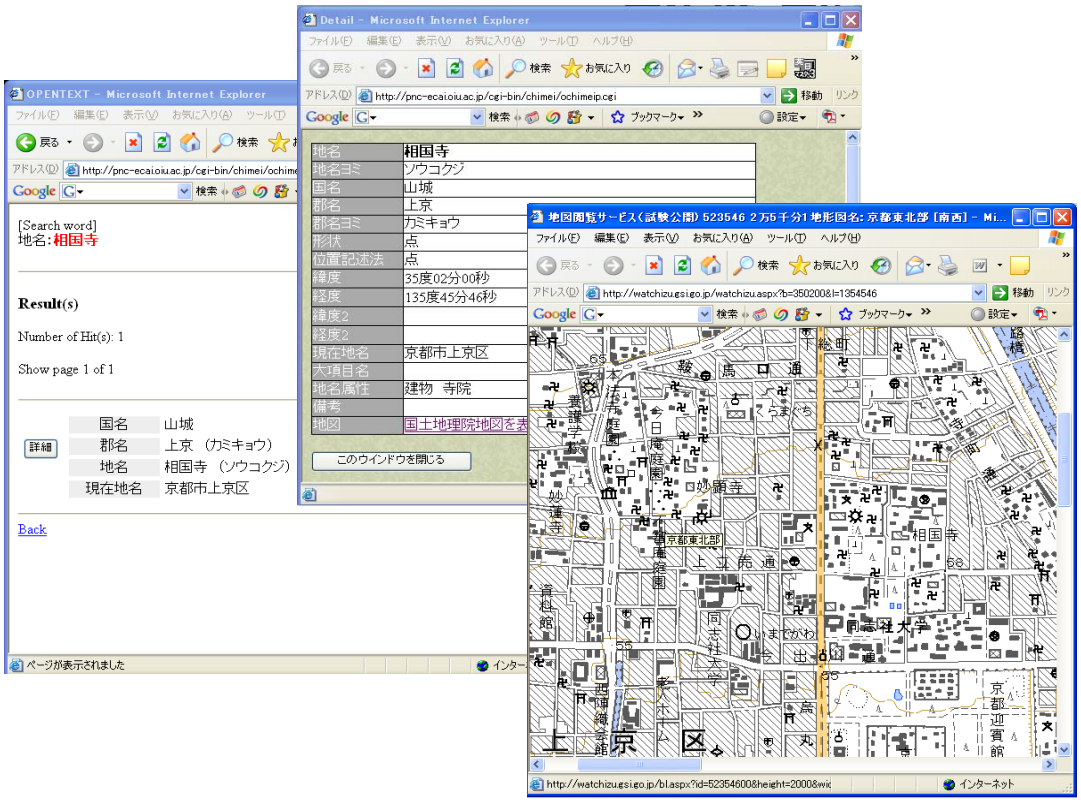


図２．地名辞書データベースの検索例

２．地名辞書データベースの問題点と API の導入

地名辞書データベースは簡単な検索画面を持っており、特定の地名を検索するだけなら十分な機能を提供している。しかし、他のデータベースなどと組み合わせた処理を行うとすると、地名辞書の検索結果を別のデータベースに手入力あるいはコピー＆ペーストするしか方法がない。CGI を使うことも可能であるが、返礼が HTML であるため、後処理が面倒である。つまり、地名辞書をプログラムで利用することは困難である。

この問題を解決するために、地名辞書に REST（Representational State Transfer）的な Web サービスとしての API を追加した（ここで REST 的とは、RPC のスタイルに合わせた簡易

な XML+HTTP インターフェイスという意味合いで有り、Fielding の REST アーキテクチャスタイルの原則に合わせた Web サービスシステムではない)。この API の利用法は、CGI や Servlet と似ているので、ユーザ側プログラミングが容易であること、返戻が XML/JSON であるため，プログラム実装の自由度が高く，既存のツールを利用できる点に特徴がある。図 3 に API 仕様の概要を示す。

- api_base_url'/'Content_Type?'/'Database_ID'?'Parameters+
- Content_Type:none | /*XML text/xml*/
- json | /*JSON application/json*/
- simplejson | /*Simple JSON application/json*/
- jsonp | /*JSONP text/javascript*/
- simplejsonp   /*Simple JSONP text/javascript*/
- Parameters: operation=('searchRetrieval' | 'explain')
- '&'version='1.2'
- ('&'Query)?
- '&'recordSchema=('mods' | 'dc' | 'original')
- ('&'etc.)*
- Query: query='not'? (Term Op Keyword) (('and'|'or') 'not'? (Term Op keyword))*
- Term: Original_Field_Name | Mods_Terms | DC_Terms | 'cql.anywhere' | etc.
- Op: '=' | '==' | '<>' | etc.

図３．**REST 型 API の仕様の概要**

この API において、例えば「寺院名が相国寺であるレコード」を検索し、結果を Simple JSON 形式で返戻する query は、

http://*ServerURI*?operation=searchRetrieve&version=1.2&query=(c3="%E7%9B%B8%E5%9B%BD%E5%AF%BA") &recordSchema=original

となる。ここで *Server URI* は地名辞書の URI、C3 は地名辞書において地名が記載されているフィールド名、"%E7%9B%B8%E5%9B%BD%E5%AF%BA"は相国寺の URI エンコードである。返礼は

{"c1":"10026682",   "c2":"10026682",
"c3":"相国寺",   "c4":"ソウコクジ",
"c5":"35.03333333",   "c6":"135.7627778",
・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
"c16":"山城",   ・・・・・・・・・・・・・・・・・・・・・・
"c18":"上京区",   "c19":"カミキョウ",   ・・・・・・
"c22":"京都市上京区////",   "c23":"京都市上京区////",   ・・・
"c25":"大日本地名辞書",   "c27":"30048501：相国寺と重複"・・
・・・・・・・}
・・・・・・・・・・

のようになる。この検索結果から得られた相国寺のある国名（c16）の値「山城」を利用し、「相国寺と同じ国にある地名」を検索し、結果を Simple JSON 形式で返戻する query は、

http://serverURI?operation=searchRetrieve&version=1.2&query=(c16="%E5%B1%B1%E5%9F%8E")&recordSchema=original

となる。結果は、「会初川、相之京、青谷、赤井、赤井河原、不明門、県神社、県井戸、県宮、赤日埼、秋山、安居院、旦椋神社、朝日山、飛鳥路、愛宕神社、愛宕山、愛宕念仏寺、阿刀神社」などとなる（ここでは地名フィールド（c3）のみを書き出している）。

REST 的 API を利用することにより、地名辞書の検索結果を利用して別のデータベースを検索する、Google Maps と連携させて位置を地図上に可視化する、統計パッケージなどと連携さてデータマイニングを行う、などの一連の操作をアプリケーションプログラムで実行できるようになった。

３．セマンティック Web を応用した新たな展開

REST 型 API の導入により、地名辞書データベースは Mashup などで利用できるようになり、研究に資するデータベースとして一定の進展を実現できた。ただし、以下のような課題が残っている。

- **柔軟なデータベース連携が困難である：**他のデータベースと連携検索する際に必要な join 操作ができないので、同様の操作を手続き的に記述する必要がある。
- **人文学で使われる手続きを素直に記述できない：**系譜を辿ったり（祖先や子孫を辿る）関係者を探し出したりする（仲間の仲間の・・・・）ことは人文学分野における典型的な情報処理操作である。しかし現状の手続的な API でデータベースを結合しながら関連を辿ることは容易でない

このような課題を解決するためには、従来の関係データモデルではなく、データ要素間の関連記述を重視したネットワーク型のデータモデルが有効であると考えた。ネットワーク型のデータモデルは多く提案されているが、Web 上にある情報資源の記述に対応し、データモデルや記述法が国際標準となっており、ツールやアプリケーションが豊富で、応用実績があるものとして、RDF（Resource Description Framework）がある。

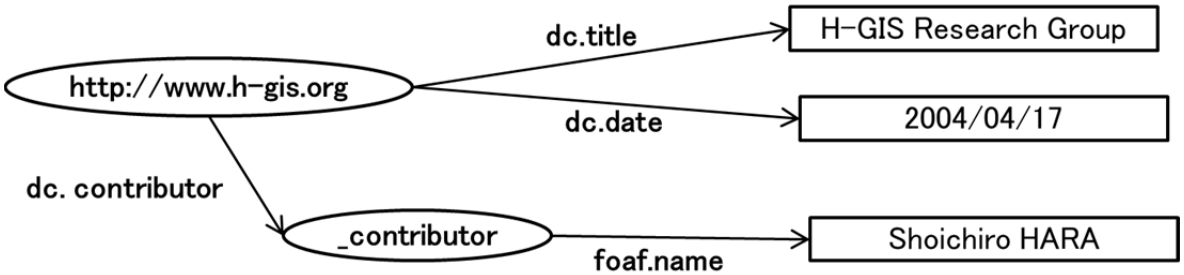


図４．RDF によるネットワークモデルの例

RDF のデータモデルでは、2 つの情報資源の関連を、主語（subject）、述語（predicate）、目的語（object）の 3 つの要素で表現し、これをトリプル（triple）と呼ぶ。主語は情報資源、目的語は主語に関連した別の情報資源あるいは名前などの値、述語が両者の関連を示す。RDF において Triple はグラフで表現され、通常の情報資源（URI で識別できるもの）は丸いノードで表し、テキストなどの値は四角のノードで表す。各ノードの間は、これらの関係を表すエッジで表現する。RDF が定義しているのはこのトリプルに基づく抽象構文であり、具象構文としては XML を利用した RDF/XML が別に定義されている。例えば図 4 上部分のグラフは、「http://www.h-gis.org とは Humanities' GIS research group のことである」という情報を表現している。ここで、「http://www.h-gis.org 」が主語、「のことである（dc.title）」が述語、「Humanities' GIS research group」が目的語となる。これを RDF/XML で記述すると次のようになる。

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
    <rdf:Description rdf:about="http://www.h-gis.org">
        <dc:title>Humanities' GIS research grouop</dc:title>
    </rdf:Description>
</rdf:RDF>
```

RDF において、情報資源は URI で一意に識別される。異なるデータベースに存在するデータであっても、同じ情報資源に同じ URI が付与されていれば RDF のデータそのものが自動的に対応づけられる。つまり、ある情報の目的語が URI で示され、さらに同じ URI を主語に持つ情報が別のデータベースにあると、それらの目的語と主語が連結されて、新しいネットワークが構成される。このメカニズムにより、単純なトリプルから巨大なネットワークが構築される。さらに RDF には、語彙の意味やカテゴリ・分類を定義する RDFS（RDF Schema）、語彙の厳密な意味や相互関係（オントロジー）を記述する OWL（Web ontology language）、RDF で記述されたデータを検索・操作する SPARQL（SPARQL Protocol and RDF Query Language）などの整備が進んでおり、これらを駆使することにより、多様な情報記述と情報処理を実現している。

４．RDF を利用した地名辞書データベースの再構築

地名辞書データベースを Web 上におけるより使い勝手の良い研究資源とすることを目指して、①データモデル設計、②語彙定義、③RDF データへの変換と RDF データベース（RDF Triple Store）構築、④検索用アクセスポイント（SPARQL Endpoint）公開を試みた。

４．１ データモデルの設計

地名辞書には、①地名を行政地名・建物などの属性で分類している、②地名は、それが属する町、その町が属する郡、その郡が属する県のように階層化されている、③緯度・経度が付与されている、④地名の出典情報があるなどの特徴がある。この特徴を新しい地名辞書のデータモデルにおいても踏襲した（図 5）。

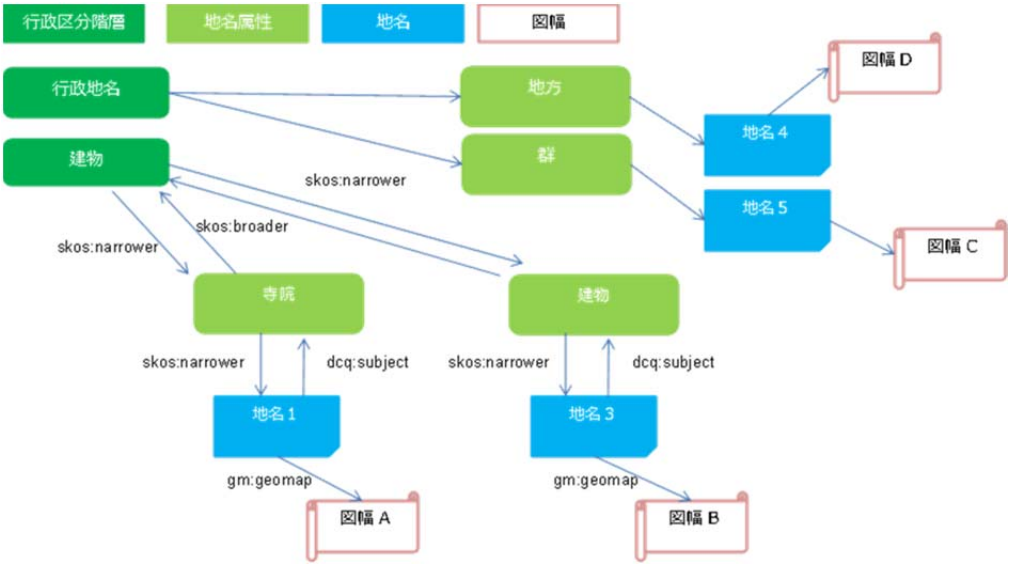


図５．地名辞書のデータモデル

４．２ 語彙と名前空間の定義

RDF ではデータを主語・述語・目的語の Triple で表現し、各 Triple は名前（語彙）で識別されている。地名辞書のフィールド名のうち、「国名」のように地名辞書に特有な語彙は Triple の語彙としてそのまま利用するとしても、地名の階層関係や同義語関係などはシソーラスやオントロジーなどで使われている既存の語彙を使った方がデータ連係の点から都合がよい。しかし「title」が雑誌目録では「題名」の意味で、人名録では「肩書き」の意味で用いられるように、同じ語彙が異なる分野において異なる意味で使われていることがある。そのため、各語彙は名前空間で識別する。名前空間と語彙を組み合わせ URI とする。これにより、地名辞書の語彙は一意に識別される。地名辞書で利用している名前空間の一部を以下に示す。

地名辞書	gzt http://gazetteer.chikyu.ac.jp/gzt/elements/1.0/
図幅	gm http://gazetteer.chikyu.ac.jp/gm/elements/1.0/
地名属性	http://gazetteer.chikyu.ac.jp/id/placeattribute/番号
行政区分	http://gazetteer.chikyu.ac.jp/id/category/番号
地名	http://gazetteer.chikyu.ac.jp/id/placename/番号
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
dc	http://purl.org/dc/elements/1.1/
dcq	http://purl.org/dc/terms/
owl	http://www.w3.org/2002/07/owl#
skos	http://www.w3.org/2004/02/skos/core#
skosxl	http://www.w3.org/2008/05/skos-xl#
geo	http://www.w3.org/2003/01/geo/wgs84_pos#
vcard	http://www.w3.org/2006/vcard/ns#
gn	http://www.geonames.org/ontology

・・・・・・・・・・

以下の語彙の一部を示す。

ID	主語
Original ID	dc:identifier
地名	rdfs:label
地名よみ	skosxl:altLabel
緯度経度情報	geo:Point
緯度	geo:lat
経度	geo:long
緯度 tky	gzt:lat
経度 tky	gzt:long
x 座標（UTM）	gzt:x
y 座標（UTM）	gzt:y
ゾーン（UTM）	gzt:zone
形状	dc:format
位置記述法	gzt:projection
地名属性	dc:subject rdfs:label dcq:subject
旧国名	gzt:country
旧国名元表記	gzt:countryOriginalNotation
旧群名	gzt:county
旧郡名よみ	gzt:countyAltLabel
旧郡名元表記	gzt:countyOriginalNotation
都道府県名	vcard:region
合併前所在地	gzt: localityOriginalNotation
現所在地	gzt: locality
現在地名	skosxl:prefLabel
出典	dc:source
備考	rdfs:comment
図幅情報	gm:geomap
図幅名	rdfs:label owl:sameAsI gm:north gm:west gm:east gm:south

４．３ RDF Triple の変換と RDF Triple Store の構築

上記で定義した語彙を用いて、地名辞書の各レコードを RDF Triple に書き換える。以下は RDF で記述された「阿位郷」の例である。

```
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:     <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc:       <http://purl.org/dc/elements/1.1/> .
@prefix dcq:      <http://purl.org/dc/terms/> .
@prefix owl:      <http://www.w3.org/2002/07/owl#> .
@prefix skos:     <http://www.w3.org/2004/02/skos/core#> .
@prefix skosxl:   <http://www.w3.org/2008/05/skos-xl#> .
@prefix geo:      <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix vcard:    <http://www.w3.org/2006/vcard/ns#> .
@prefix gzt:      <http://gazetteer.chikyu.ac.jp/gzt/elements/1.0/> .
@prefix gm:       <http://gazetteer.chikyu.ac.jp/gm/elements/1.0/> .

<http://gazetteer.chikyu.ac.jp/id/placename/10000001>
        a           geo:SpatialThing ;
        rdfs:label "阿位郷" ;
        gzt:country "出雲" ;
        gzt:county "仁多郡" ;
        gzt:countyAltLabel "ニタ" ;
        gzt:locality "仁多郡奥出雲町////" ;
        gzt:localityOriginalNotation "仁多郡奥出雲町////" ;
        gzt:projection
                [ rdfs:label "1" ;
                  owl:sameAs <http://gazetteer.chikyu.ac.jp/pn/pntype/Point>
                ] ;
```

```
dc:format
        [ rdfs:label "1" ;
          owl:sameAs <http://gazetteer.chikyu.ac.jp/pn/pntype/Point>
        ] ;
dc:identifier "10000001" ;
dc:source "大日本地名辞書" ;
dc:subject
        [ rdfs:label "郷・里" ;
          dcq:subject <http://gazetteer.chikyu.ac.jp/id/placeattribute/6>
        ] ;
geo:Point
        [ gzt:lat "35.13" ;
          gzt:long "132.94" ;
          gzt:x      "3889169.23606705" ;
          gzt:y      "312534.868129257" ;
          gzt:zone "53 S" ;
          geo:lat "35.1279467847798" ;
          geo:long "132.942567549788"
        ] ;
skosxl:altLabel "アイ" .
```

４．４ **SPARQL Endpoint** の構築

RDF Triple Store に格納された地名データを検索する一般的な方法は、RDF Triple 用検索言語の SPARQL を利用することである。SPARQL クエリを受け付ける API の URI を SPARQL Endpoint という。以下は、地名辞書用の SPARQL Endpoint において「相国寺のある国に属する地名一覧」の表示するための SPQRCL クエリの例であり、

```
PREFIX rdfs:        <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gzt:         <http://gazetteer.chikyu.ac.jp/gzt/elements/1.0/>

select distinct ?place ?name
  from <http://gazetteer.chikyu.ac.jp/rdf/placename>
where     {
          {?s rdfs:label "相国寺"}
          {?s gzt:country ?country}
          {?place gzt:country ?country}
          {?place rdfs:label ?name}
  }
```

その結果の一部は以下のようになる

100 / 8353 (表示件数/ヒット件数)　ダウンロード: RDF/XML 形式 N3/Turtle 形式 CSV 形式

Place	Name
http://gazetteer.chikyu.ac.jp/id/placename/10023952	朱雀野
http://gazetteer.chikyu.ac.jp/id/placename/10042178	真葛原
http://gazetteer.chikyu.ac.jp/id/placename/10003404	池坊

・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・

SPARQL クエリの記述は一般利用者には困難であるため、RDF Triple Store に集積されたデータを検索したり可視化したりするためのインターフェイスや GUI を開発する必要があると思われた。そこで今回の試行では、これまでの地名辞書データベースの API と同じ動作をするインターフェイスを SPARQL 検索エンジンの上に作成した。

５．まとめ

データベースの高度利用を目指す試みとして、地名辞書データベースを RDF Triple Store として再構築した。有効性の検証はこれからであるが、以下の点に留意する必要があると考えている。

- 利用のシナリオ・Good Practice・事例の収集
- 多くのデータベースを RDF Triple Store 化して連携の対象を充実させる
- API・可視化などユーザインタフェースの機能充実
- データマイニングツール（統計パッケージ、GIS 等）との連携

地名データの
セマンティック Webへの応用

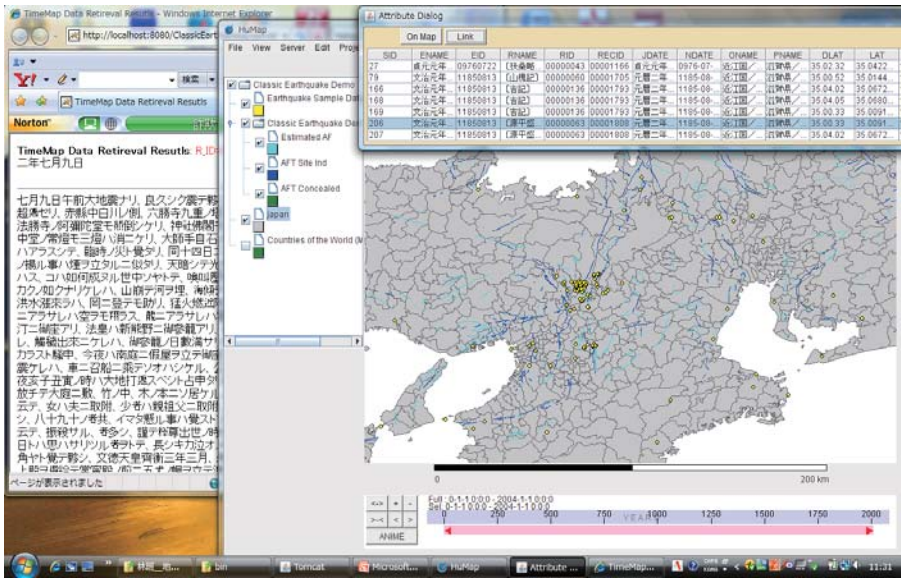
京都大学地域研究統合情報センター
原　正一郎

地名にかかる情報技術に関する研究会@国立情報学研究所20140719
京都大学地域研究統合情報センター地域情報学プロジェクト2014年度第2回H-GIS研究会
国立情報学研究所公募型共同研究「GeoNLPプロジェクト：自然言語文を対象とした地名情報処理とコミュニティの展開」第1回研究会
第9回人間文化研究情報資源共有化研究会

そもそも・・・・・

- 2001（?）年頃のPNCにおいて、地名データのデータベース公開に関してクレームを受けた（Free Rider）
  - ✓実際にWebで利用可能なフリーのデジタル地名辞書は見つからなかった
- 2003年に基盤研究(A)「古代・中世の全地震史料の校訂・電子化と国際標準震度データベース構築に関する研究」開始
  - ✓デジタル地名辞書の必要性（震度データベース構築等における位置同定）
- 2004年にデジタル地名辞書の簡単な調査を行っている
  - ✓データソースとして大日本地名辞書があげられている（H-GIS研究会）
- 2005年11月17日に、吉田文庫と国文学研究資料館の間で、電子化された「大日本地名辞書（索引）」の構築に関する覚書を締結
  - ✓歴史地名データベース化を開始
    - 主要経費：人間文化研究機構連携　湿潤アジアにおける「人と水」
    - 作業監修：大阪国際大学桶谷教授
    - 2008年（？）以降は旧5万分の1地図のデジタル化と地名入力を開始
    - 大日本(49,557)+式内社(2,842) +寺院(78,588)+旧5万分の1(192,122) = 323,109

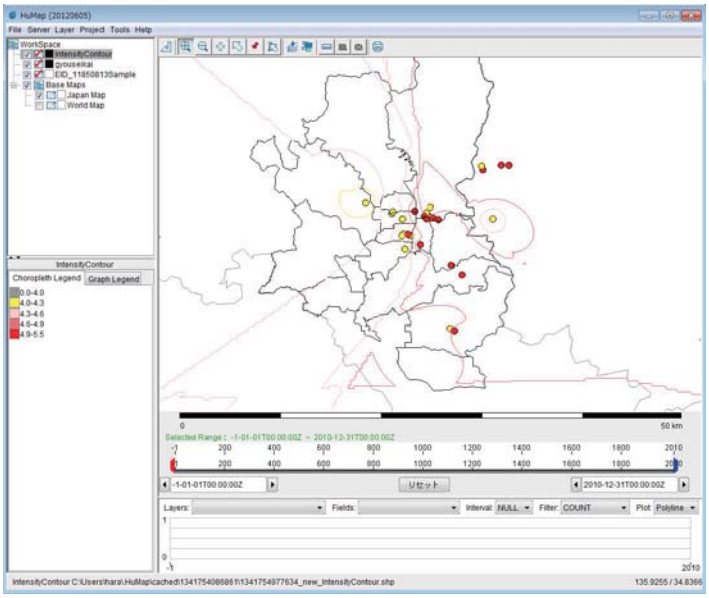
こんな使い方を想定1



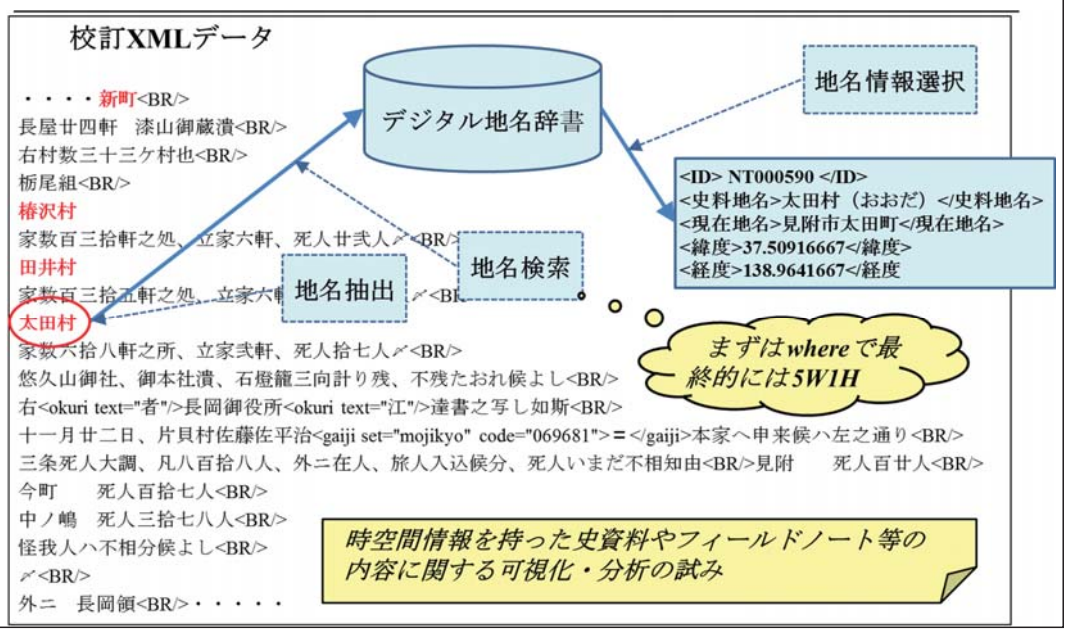
こんな使い方を想定1 *(cont.)*

地震名	地震ID	史料名	史料ID	レコード	和暦	西暦	古代地名	現在地名	Y	X	備考	震度	震度判定概拠
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	山城国／京都／法勝寺	京都府／京都市／左京区／岡崎法勝寺町	35.01028	135.7902778	現法勝寺町バス停付近に「法勝寺跡」の石碑あり	5	法勝寺九重塔顛落、重々垂木以上皆落也、毎層柱扉連子被相殘、露盤八殘其上折落、阿彌陀堂并金堂之東西廻廊、礎楼、常行堂之廻廊、南大門西門三宇(ママ)、北門一宇、皆顛倒(ママ)、無一宇全、門築垣皆壊、南北面少々相殘云々、【『玉葉』によれば阿弥陀堂?倒は七月十三日かもしれない】
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	山城国／京都／得長寿院	京都府／京都市／左京区／岡崎徳成町	35.01194	135.7811111	左の場所に得長寿院跡の石碑あり	5	闍得長壽院(千體正觀音｜鳥羽院御願)顛倒云々
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	山城国／京都／御所(閑院)	京都府／京都市／中京区／古城町	35.00889	135.7569444	当時の御所(里内裏)は閑院。現中京区押小路小川西入ル西福寺付近。東は西洞院通、西は油小路通、南は御池通、北は二条通に囲まれた地域にほぼ相	4.5	東中門廊顛危、東子午透廊北車寄顛倒、四面賀垣(スガキ？)大略破壊
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	山城国／京都／法成寺	京都府／京都市／上京区／松蔭町	35.0175	135.7705556	現府立鴨沂高校付近に「法成寺跡」の石碑あり	5	法成寺内廻廊皆?倒、東塔(西塔未造畢、只有組物許不傾、)北傾、東面築垣皆壊、諸門無事、南門築垣近日皆修理、十本許壊、
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	山城国／京都／五条摂政亭(近衛亭)	京都府／京都市／下京区／上平野町	34.99222	135.7630556	五条摂政亭(里内裏)は五条東洞院。緯度・経度は現在の五条東洞院郵便局。	5	五條攝政亭(當時坐近衛亭)寢殿悉平伏、西子午透廊顛倒、四面築垣皆壊、
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	山城国／宇治橋	京都府／宇治市／宇治	34.88944	135.8091667	左は現在の宇治橋の位置	5	宇治橋皆以顛倒、干時渡之人十餘人乘橋入水、其中一人溺死云々、
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	美濃国	岐阜県／不破郡／垂井町／府中	35.3775	136.5277778	仮に、現在の垂井府中簡易郵便局の緯度・経度を入力	3	美濃・伯耆等の国より来たるの輩いわく、はなはだの大動にあらず
文治元年七月九日	11850813	〔山槐記〕	60	1705	元暦二年〇文治元年七月九日庚寅	1185-08-13	伯耆国	鳥取県／倉吉市／国府	35.43056	133.7925	緯度・経度は、伯耆国庁跡附法華寺畑遺跡	3	美濃・伯耆等の国より来たるの輩いわく、はなはだの大動にあらず
文治元年七月九日	11850813	〔吾妻鏡〕	140	1785	文治元年七月十九日庚子	1185-08-23	相模国／鎌倉	神奈川県／鎌倉市			【石橋追加】	3	地震良久
文治元年七月九日	11850813	〔吾妻鏡〕	140	1785	文治元年七月十九日庚子	1185-08-23	山城国／京都／源廷尉六條室町亭	京都府／京都市／下京区／西若松町	34.99083	135.7563889	緯度・経度は六条通と油小路通の交差点付近のもの	4	源廷尉六條室町亭、云門垣、云家屋、無聊顛傾云々、可謂不思議歟、
文治元年七月九日	11850813	〔百錬抄〕	76	1787	文治元年七月九日庚寅	1185-08-13	山城国／京都(京中)／宮城／内裏(大内)	京都府／京都市／上京区／小山町	35.01667	135.7463889	緯度・経度は内裏の中心付近のもの	4.5	宮城瓦垣并京中民屋、或破損、或顛倒、一所而不全、

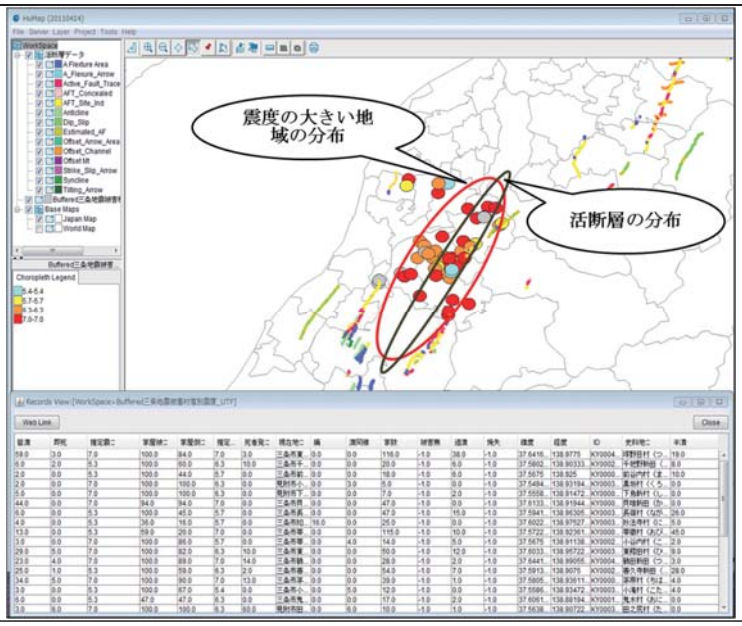
こんな使い方を想定１*(cont.)*



こんな使い方を想定２

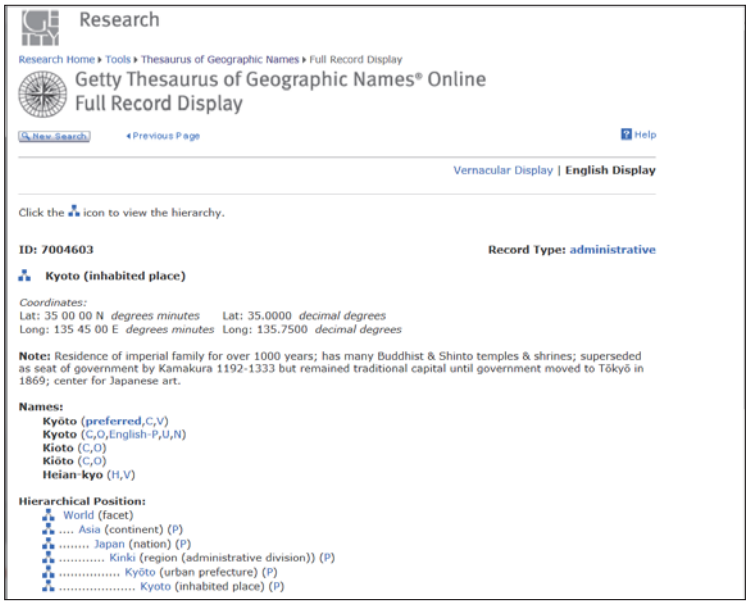


こんな使い方を想定2*(cont.)*



地名辞書の構造例1

-Getty Thesaurus of Geographic Names-



Research

Research Home ▸ Tools ▸ Thesaurus of Geographic Names ▸ Full Record Display

Getty Thesaurus of Geographic Names® Online
Full Record Display

New Search | ◂Previous Page | Help

Vernacular Display | **English Display**

Click the icon to view the hierarchy.

ID: 7004603 **Record Type: administrative**

Kyoto (inhabited place)

Coordinates:
Lat: 35 00 00 N *degrees minutes* Lat: 35.0000 *decimal degrees*
Long: 135 45 00 E *degrees minutes* Long: 135.7500 *decimal degrees*

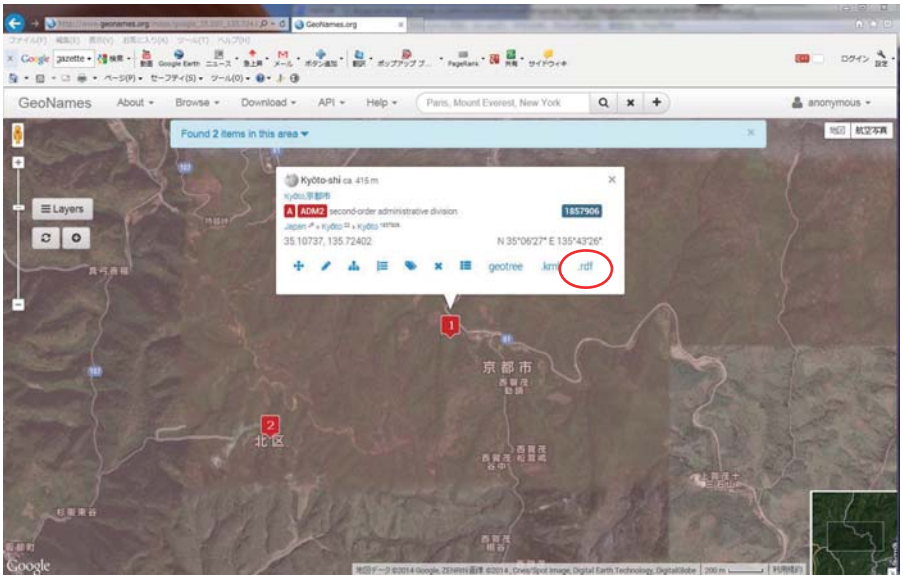
Note: Residence of imperial family for over 1000 years; has many Buddhist & Shinto temples & shrines; superseded as seat of government by Kamakura 1192-1333 but remained traditional capital until government moved to Tōkyō in 1869; center for Japanese art.

Names:
Kyōto (preferred,C,V)
Kyoto (C,O,English-P,U,N)
Kioto (C,O)
Kiōto (C,O)
Heian-kyo (H,V)

Hierarchical Position:
World (facet)
.... Asia (continent) (P)
..... Japan (nation) (P)
..... Kinki (region (administrative division)) (P)
..... Kyōto (urban prefecture) (P)
..... Kyoto (inhabited place) (P)

地名辞書の構造例2

−geonames−



歴史地名データベースの構造

Past Names

Inclusion in the Past

Present Names

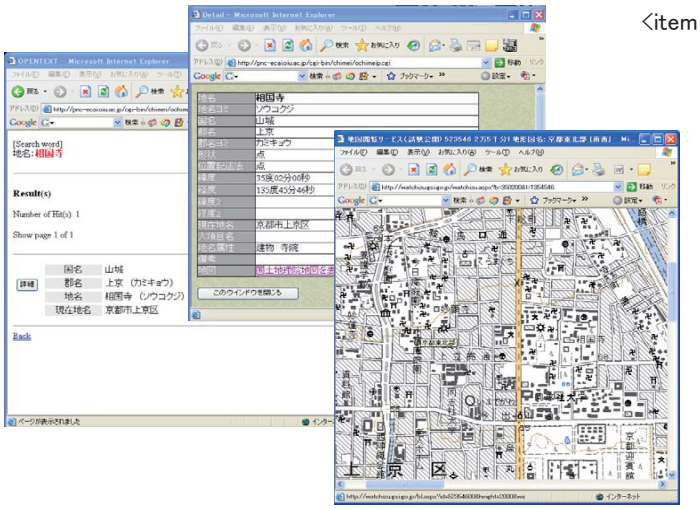
Present Inclusion

Coordinates

Feature Types

1	2	3	4	5	6	7	8	9a	9b	9c	10a	11a	10b	11b	12	13	14	15	16	17
番号	地名	ヨミ	国	郡	郡ヨミ	形状	位置記述法	現在地名a	現在地名b	現在地名c	緯度	経度	緯度	経度	現在名称	巻	ページ	大項目名	地名属性	備考
11997	空木山	カサギ	美濃																	
38325	飛州	ヒ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	509	飛騨国	2	
38387	飛騨山脈	ヒダ	飛騨			1	1	岐阜県			36.17.21	137.38.52				5	509	飛騨国	33	
38390	飛騨国	ヒダ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	509		2	
38400	斐陀国	ヒダ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	509	飛騨国	2	
38356	飛騨	ヒダ	飛騨			3	4	岐阜県			36.8.34	137.15.14				5	510	飛騨国	2	
38377	飛騨川	ヒダ	飛騨			2	2	高山市	美濃加茂市		36.5.8	137.32.34	35.26.43	137.3.8		5	510		22	
38384	飛騨高原	ヒダ	飛騨			1	1				36.8.34	137.15.14				5	510	飛騨川	34	
38388	飛騨谷	ヒダ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市	加茂郡	中津川市	35.39.41	137.9.34				5	510		9	郡と同類
38401	飛騨工	ヒダ	飛騨	飛騨谷	ヒダダニ											5	510	飛騨国		役職名？
9459	大船戸	オオフナト	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.39.46	137.9.46			大船渡	5	511	金山	8	
13047	金山	カナヤマ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.39.41	137.9.34				5	511		4	
13159	上麻生	カミアソウ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡七宗町			35.32.38	137.7.10				5	511		5	
14992	神田郷	カンダ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡東白川村			35.38.33	137.19.29				5	511		6	
15007	神土	カンド	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡東白川村			35.38.33	137.19.29				5	511		8	
15140	神淵	カンブチ	飛騨	飛騨谷	ヒダダニ	1		加茂郡七宗町			35.34.39	137.5.22			神渕	5	511		5	
16233	桐洞	キリボラ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.38.43	137.7.36			菅田桐洞	5	511		8	
17607	黒川	クロカワ	飛騨	飛騨谷	ヒダダニ	2	2	加茂郡白川町	加茂郡白川町		35.35.47	137.19.12	35.35.9	137.11.55		5	511	白川	22	
21287	笹洞	ササボラ	飛騨	飛騨谷	ヒダダニ	1	1	下呂市			35.38.42	137.6.23			菅田笹洞	5	511	桐洞	8	
21651	佐見	サミ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡白川町			35.40.34	137.16.15				5	511	白川	5	
22734	七宗山	シチソウ	飛騨	飛騨谷	ヒダダニ	1	1	加茂郡七宗町			35.35.53	137.8.22				5	511		33	

歴史地名データベースの例



```
<item pid='26682'>
    <kuni>山城</kuni>
    <gun>上京</gun>
    <gunyomi>カミキョウ</gunyomi>
    <placename reading='ソウコクジ'>
     相国寺
    </placename>
    <pname1>京都市上京区</pname1>
    <pname2></pname2>
    <pname3></pname3>
    <shp>1</shp>
    <loc>1</loc>
    <lat>35.2.0</lat>
    <long>135.45.46</long>
    <lat2></lat2>
    <long2></long2>
    <altname></altname>
    <parent></parent>
    <attribute>13</attribute>
    <page>23</page>
    <remarks/>
</item>
```

REST型APIの付与
－Webサービス（Webプログラム環境）－

- ユーザサイドとサーバサイドの機能分担
  - ✓サーバサイド：レポジトリ機能と検索機能を分担．外部に対して検索サーバの所在を示すURIとWeb ServiceのAPIを公開
  - ✓ユーザサイド：分析や表示を分担．プログラムは公開されているMyデータベースのURIとAPIを利用してユーザが作成
- ユーザサイドとサーバサイドの連携
  - ✓REST（Representational State Transfer）的なWeb Service
    - CGIあるいはServletのクライアントプログラムと似ているのでユーザ側プログラミングは容易である
    - 返戻がXML/JSONであるため，プログラム実装の自由度が高く，既存のツールを利用できる
    - JavaScriptやXSLT等を使ったマッシュアップを目指している
- 逐次型検索機能（縦型検索）等の実現
  - ✓前段の検索結果に応じて後段の検索法が変化する

REST型APIの仕様(概要)

- api_base_url '/' Content_Type? '/' Database_ID '?' Parameters+
- Content_Type:none | /*XML text/xml*/
  json | /*JSON application/json*/
  simplejson | /*Simple JSON application/json*/
  jsonp | /*JSONP text/javascript*/
  simplejsonp /*Simple JSONP text/javascript*/
- Parameters: operation=('searchRetrieval' | 'explain')
  '&' version= '1.2'
  ('&' Query)?
  '&' recordSchema=('mods' | 'dc' | 'original')
  ('&' etc.)*
- Query: query= 'not' ? (Term Op Keyword) (('and' |' or') 'not' ? (Term Op keyword))*
- Term: Original_Field_Name | Mods_Terms | DC_Terms | 'cql.anywhere' | etc.
- Op: '=' | '==' | '<>' | etc.

REST型APIによる検索例1

寺院名(c3)が「相国寺(%E7%9B%B8%E5%9B%BD%E5%AF%BA)」であるレコードを検索しSimple JSON形式で返戻

http://gazetteer.chikyu.ac.jp/api/sru/simplejson/G0000005geoname?operation=searchRetrieve&version=1.2&query=(c3="%E7%9B%B8%E5%9B%BD%E5%AF%BA")&recordSchema=original

{"c1":"10026682", "c2":"10026682",
"c3":"相国寺", "c4":"ソウコクジ",
"c5":"35.03333333", "c6":"135.7627778",
・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
"c16":"山城", ・・・・・・・・・・
"c18":"上京区", "c19":"カミキョウ", ・・・・・
"c22":"京都市上京区////", "c23":"京都市上京区////", ・・・・
"c25":"大日本地名辞書", "c27":"30048501:相国寺と重複" ・・・・
・・・・・・}
・・・・・・・・・

REST型APIによる検索例

「相国寺」と同じ「国（c16）」にある寺院を検索する

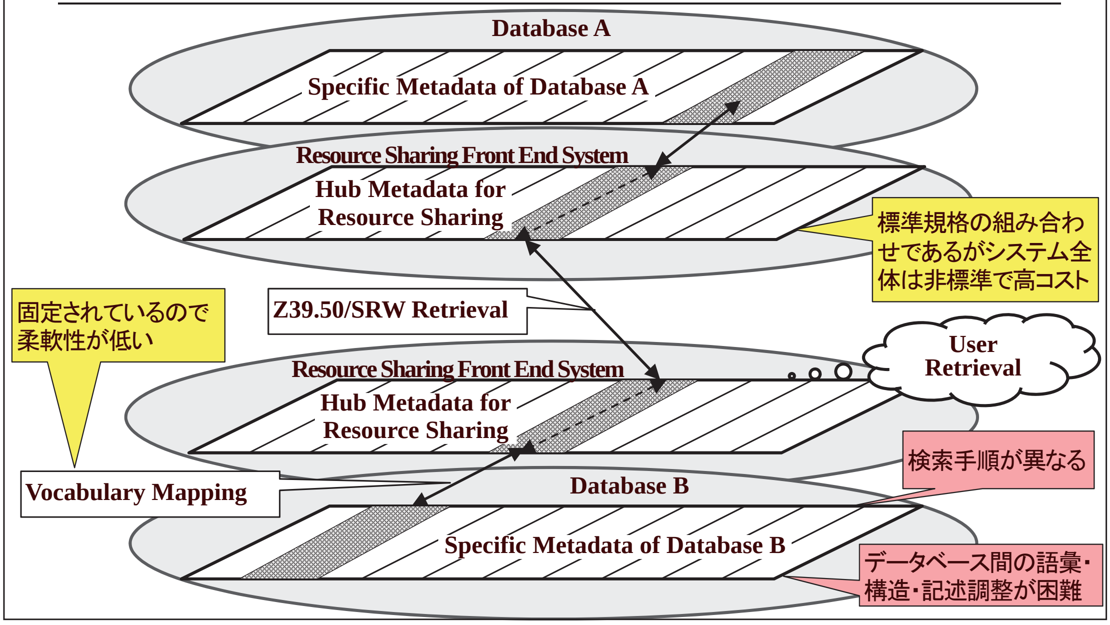
http://gazetteer.chikyu.ac.jp/api/sru/simplejson/G0000005geoname?operation=searchRetrieve&version=1.2&query=(c3="%E7%9B%B8%E5%9B%BD%E5%AF%BA")&recordSchema=original

{"c1":"10026682", "c2":"10026682",
"c3":"相国寺", "c4":"ソウコクジ",
"c5":"35.03333333", "c6":"135.7627778",
・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
"c16":"山城", ・・・・・・・・・・
"c18":"上京区", "c19":"カミキョウ", ・・・・・
"c22":"京都市上京区////", "c23":"京都市上京区////", ・・・・
"c25":"大日本地名辞書", "c27":"30048501：相国寺と重複 ・・・・
・・・・・・}
・・・・・・・・・

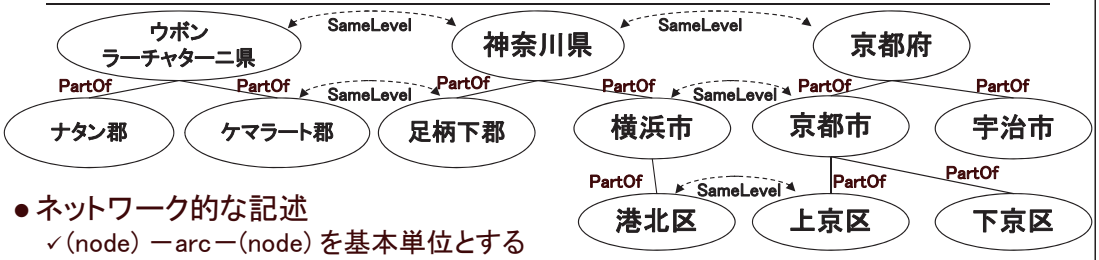
http://gazetteer.chikyu.ac.jp/api/sru/simplejson/G0000005geoname?operation=searchRetrieve&version=1.2&query=(**c16="%E5%B1%B1%E5%9F%8E"**)&recordSchema=original

- とりあえず動いている
- 他の辞書との連携をどうするか（資源共有化フレームワーク？）
- 手続き型になってしまう（手順を記述する必要がある）。

ところで資源共有化システムとは・・・



ネットワーク型のメタデータ



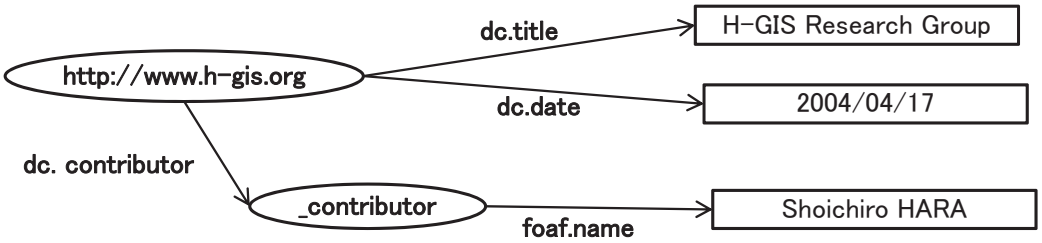
- ネットワーク的な記述
  - ✓(node) ーarcー(node) を基本単位とする
    - (京都市) ーPartOfー (京都府)なら PartOf (京都市, 京都府) のように書ける
  - ✓表やツリーやネットワークは基本単位を順次接続して構築する
    - PartOf(上京区,京都市)とPartOf(京都市,京都府) から PartOf(上京区, 京都府) を導く
  - ✓ 意味の階層(たとえばシソーラス)を簡単に利用できるかも
- 検索に述語論理(のようなもの)が利用できたら便利かも
  - ✓手続き型よりも宣言型の方が検索には便利な場合が多い
    - 例:祖先とは私の親あるは親の親である
  - ✓理論的な枠組みが整ってると便利(ホーン節等)
    - 例: PartOf(Narrow, Wide).
      PartOf(Narrow, Wide):-PartOf(Narrow, Next), PartOf(Next, Wide).
- NodeやarcをWeb上で一意に同定できたらとても便利

Resource Description Framework (RDF)

ー基本的な要求を満たしているー

Resource Description Framework:Web上における柔軟なデータ連係を実現するための情報記述

- ✓Triple: Subject + Predicate + Object or Predicate(Subject, Object)
- ✓URIによるリソースの一意識別
- ✓URIによるWebリソースのリンク
- ✓語彙拡張性(RDFS, OWL)
- ✓コンピュータによる推論の実現を目指す(SPARQL等)

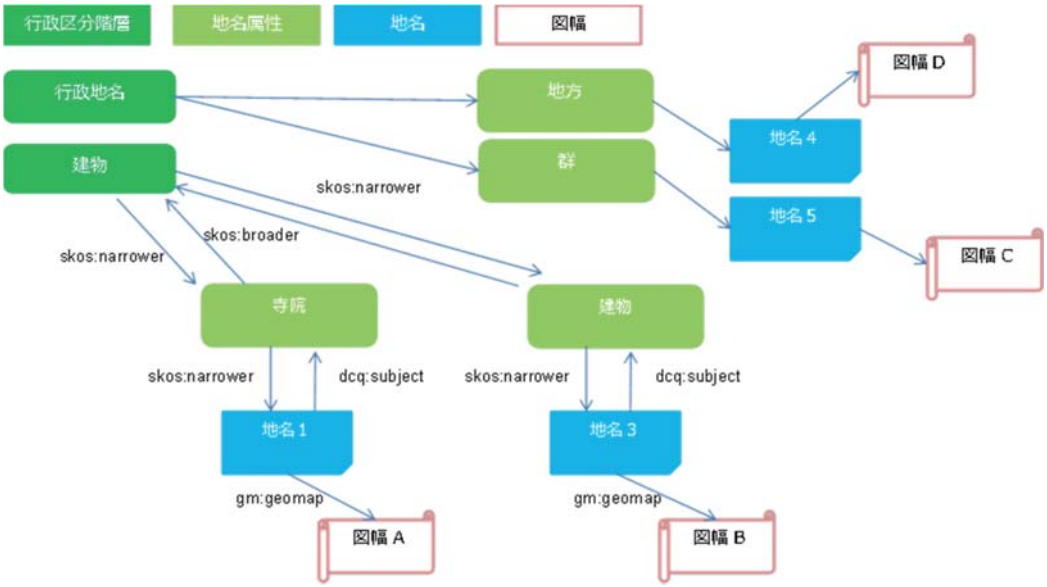


目標:RDFをベースにした共有化プラットフォームの構築と利用可能性の検討

RDFの利用に関する試み

- 次世代資源共有化システムの基礎研究として、とにかく手を動かしてみよう
- 地球環境学リポジトリを「知のデータベース」の情報基盤とするためのセマンティックWebに関する基礎的な研究
- 地域研究統合情報センターおよび総合地球環境学研究所が構築中の地名データベースを対象として、RDFデータの設計・変換・構築に関して以下の研究を実施
  - ✓ 地名辞書に関するデータモデルの設計
  - ✓ 名前空間定義
  - ✓ 語彙の定義
  - ✓ スキーマ定義RDFデータ変換とRDFストアの構築
  - ✓ 簡易検索ツールの構築

地名辞書のデータモデル



名前空間の定義（一部）

- 地名辞書　gzt　http://gazetteer.chikyu.ac.jp/gzt/elements/1.0/
- 図幅　gm http://gazetteer.chikyu.ac.jp/gm/elements/1.0/
- 地名属性　http://gazetteer.chikyu.ac.jp/id/placeattribute/番号
- 行政区分　http://gazetteer.chikyu.ac.jp/id/category/番号
- 地名　http://gazetteer.chikyu.ac.jp/id/placename/番号
- rdf　http://www.w3.org/1999/02/22-rdf-syntax-ns#
- rdfs　http://www.w3.org/2000/01/rdf-schema#
- dc　http://purl.org/dc/elements/1.1/
- dcq　http://purl.org/dc/terms/
- owl　http://www.w3.org/2002/07/owl#
- skos　http://www.w3.org/2004/02/skos/core#
- skosxl　http://www.w3.org/2008/05/skos-xl#
- geo　http://www.w3.org/2003/01/geo/wgs84_pos#
- vcard　http://www.w3.org/2006/vcard/ns#
- gn　http://www.geonames.org/ontology#

・・・・・・・・・・・・・・・・・・・・

行政区分のRDFデータ例(一部)

```
@prefix rdf:        <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos:       <http://www.w3.org/2004/02/skos/core#> .
@prefix rdfs:       <http://www.w3.org/2000/01/rdf-schema#> .

<http://gazetteer.chikyu.ac.jp/id/category/01>      /*区分　1:行政地名、2:建物等*/
        rdf:type    skos:Concept ;                  /*タイプ*/
        rdfs:label  "行政地名" ;
        skos:narrower                               /*下位階層*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/1> ,  /*地方*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/2> ,  /*国*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/3> ,  /*郡*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/4> ,  /*町*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/5> ,  /*村*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/6> ,  /*郷・里*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/7> ,  /*市*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/8> ,  /*字*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/9> ,  /*その他*/
                    <http://gazetteer.chikyu.ac.jp/id/placeattribute/83> , /*県*/
                    ・・・・・・・・・・・・・・・・・・・・・・・・・・・・
<http://gazetteer.chikyu.ac.jp/id/placeattribute/1>
        rdf:type    skos:Concept ;
        rdfs:label  "地方" ;
        skos:broader        <http://gazetteer.chikyu.ac.jp/id/category/01> .
```

地名メタデータの語彙定義（一部）

ID	主語
Original ID	dc:identifier
地名	rdfs:label
地名よみ	skosxl:altLabel
緯度経度情報	geo:Point
緯度	geo:lat
経度	geo:long
緯度tky	gzt:lat
経度tky	gzt:long
x座標（UTM）	gzt:x
y座標（UTM）	gzt:y
ゾーン（UTM）	gzt:zone
形状	dc:format
位置記述法	gzt:projection
地名属性	dc:subject
	rdfs:label
	dcq:subject
旧国名	gzt:country
旧国名元表記	gzt:countryOriginalNotation
旧群名	gzt:county
旧郡名よみ	gzt:countyAltLabel
旧郡名元表記	gzt:countyOriginalNotation
都道府県名	vcard:region
合併前所在地	gzt: localityOriginalNotation
現所在地	gzt: locality
現在地名	skosxl:prefLabel
出典	dc:source
備考	rdfs:comment
図幅情報	gm:geomap
図幅名	rdfs:label
	owl:sameAsI
	gm:north
	gm:west
	gm:east
	gm:south

地名辞書のRDFデータ例

```
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc:     <http://purl.org/dc/elements/1.1/> .
@prefix dcq:    <http://purl.org/dc/terms/> .
@prefix owl:    <http://www.w3.org/2002/07/owl#> .
@prefix skos:   <http://www.w3.org/2004/02/skos/core#> .
@prefix skosxl: <http://www.w3.org/2008/05/skos-xl#> .
@prefix geo:    <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix vcard:  <http://www.w3.org/2006/vcard/ns#> .
@prefix gzt:    <http://gazetteer.chikyu.ac.jp/gzt/elements/1.0/> .
@prefix gm:     <http://gazetteer.chikyu.ac.jp/gm/elements/1.0/> .

<http://gazetteer.chikyu.ac.jp/id/placename/10000001>
    a       geo:SpatialThing ;
    rdfs:label "阿位郷" ;
    gzt:country "出雲" ;
    gzt:county "仁多郡" ;
    gzt:countyAltLabel "ニタ" ;
    gzt:locality "仁多郡奥出雲町////" ;
    gzt:localityOriginalNotation
        "仁多郡奥出雲町////" ;
    gzt:projection
        [ rdfs:label "1" ;
          owl:sameAs <http://gazetteer.chikyu.ac.jp/pn/pntype/Point>
        ] ;
    dc:format
        [ rdfs:label "1" ;
          owl:sameAs <http://gazetteer.chikyu.ac.jp/pn/pntype/Point>
        ] ;
    dc:identifier "10000001" ;
    dc:source "大日本地名辞書" ;
    dc:subject
        [ rdfs:label "郷・里" ;
          dcq:subject
<http://gazetteer.chikyu.ac.jp/id/placeattribute/6>
        ] ;
    geo:Point
        [ gzt:lat "35.13" ;
          gzt:long "132.94" ;
          gzt:x   "3889169.23606705" ;
          gzt:y   "312534.868129257" ;
          gzt:zone "53 S" ;
          geo:lat "35.1279467847798" ;
          geo:long "132.942567549788"
        ] ;
    skosxl:altLabel "アイ" .
```

図幅のRDFデータ例

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dcq: <http://purl.org/dc/terms/> .
@prefix gn: <http://www.geonames.org/ontology#> .
@prefix gm: <http://gazetteer.chikyu.ac.jp/gm/elements/1.0/> .

<http://gazetteer.chikyu.ac.jp/id/map/76-4-1>        /*リスト番号（主語）*/
        a gn:Map ;                                   /*タイプ*/
        rdfs:comment “測図” , “*” ;                   /*地形図の種別等*/
        rdfs:label ″木更津″ ;                         /*図幅名*/
        gm:no “76-4-1” ;                             /*リスト番号*/
        dcq:created “1903” ;                         /*測量年*/
        dcq:issued “1911/03/08” .                    /*発行年月日*/
```

SPARQL検索例

「相国寺」という地名のある国に属する地名一覧

```
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gzt:     <http://gazetteer.chikyu.ac.jp/gzt/elements/1.0/>

select distinct ?place ?name
        from <http://gazetteer.chikyu.ac.jp/rdf/placename>
where   {
                {?s rdfs:label ″相国寺″}
                {?s gzt:country ?country}
                {?place gzt:country ?country}
                {?place rdfs:label ?name}
        }
```

100 / 8353 (表示件数/ヒット件数)　　ダウンロード： RDF/XML形式N3/Turtle形式CSV形式

Place	Name
http://gazetteer.chikyu.ac.jp/id/placename/10023952	朱雀野
http://gazetteer.chikyu.ac.jp/id/placename/10042178	真葛原
http://gazetteer.chikyu.ac.jp/id/placename/10003404	池坊
http://gazetteer.chikyu.ac.jp/id/placename/10026298	施薬院
http://gazetteer.chikyu.ac.jp/id/placename/10040014	富家殿
http://gazetteer.chikyu.ac.jp/id/placename/10032728	富小路

・・・・・・・・・・・・・・・・・・・・・・・・・・・・・

最近の作業(1)
ータイ語地名辞書ー

#	Unit	Code	Name English	Name Local	Lat	Lon	国	州	郡	区	村	寺
5	Temple	TH340301001		วัดถ้ำเหวสินธุชัย	15.30678	105.4844	TH	34	3	1	0	1
5	Temple	TH340301002		สำนักสงฆ์คำผักกูด	15.32455	105.4836	TH	34	3	1	0	2
5	Temple	TH340301011		วัดโขงเจียม	15.3174	105.5017	TH	34	3	1	1	1
5	Temple	TH340301012		สถานปฏิบัติธรรมอัศวเทพ	15.31657	105.4895	TH	34	3	1	1	2
5	Temple	TH340301021		วัดถ้ำคูหาสวรรค์	15.30836	105.4854	TH	34	3	1	2	1
5	Temple	TH340301022		วัดบุปผาวัน	15.32072	105.4933	TH	34	3	1	2	2
5	Temple	TH340301023		ที่พักสงฆ์ห้วยค้อ	15.32657	105.4852	TH	34	3	1	2	3
5	Temple	TH340301031		สำนักสงฆ์วัดห้วยสะคาม	15.33203	105.4751	TH	34	3	1	3	1
5	Temple	TH340301032		ที่พักสงฆ์ถ้ำฝ้ายเวินหีบ	15.33216	105.4773	TH	34	3	1	3	2
5	Temple	TH340301041		สำนักสงฆ์หัวเห่ว	15.29192	105.465	TH	34	3	1	4	1
5	Temple	TH340301051		วัดสว่างวารี	15.29769	105.4173	TH	34	3	1	5	1
5	Temple	TH340301052		สำนักสงฆ์วัดป่าถ้ำชัยเจริ	15.31543	105.4455	TH	34	3	1	5	2
5	Temple	TH340301071		สำนักสงฆ์ห้วยหมากใต้	15.31877	105.5222	TH	34	3	1	7	1
5	Temple	TH340301072		ที่พักสงฆ์สภาบุญดงขี้ยา	15.28714	105.5354	TH	34	3	1	7	2
5	Temple	TH340301073		ที่พักสงฆ์ฮ่องคำเม็ก	15.31593	105.521	TH	34	3	1	7	3
5	Temple	TH340301081		ที่พักสงฆ์ศิลาดะ	15.31561	105.5538	TH	34	3	1	8	1
5	Temple	TH340301082		ที่พักสงฆ์สถานปฏิบัติธร	15.31001	105.5566	TH	34	3	1	8	2
5	Temple	TH340301091		สำนักสงฆ์ผาประตูทอง	15.30764	105.4145	TH	34	3	1	9	1
5	Temple	TH340301101		ที่พักสงฆ์บ้านท่าแพ	15.29994	105.4948	TH	34	3	1	10	1

- 異なる構造の地名データを連係させる方法の検討
- オントロジーの利用

最近の作業(2)
ー簡易インタフェースの構築ー

- Myデータベースと互換性を有するAPIの構築
  - ✓CIASの情報基盤はLOD (Linked Open Data)を基盤とするシステムへ移行する予定
  - ✓多くの研究データベースは表形式である
  - ✓RDFストアに蓄積されたデータを表データとして扱う場合、一般ユーザにとってSPARQLでは複雑すぎるかも
  - ✓SPARQL形式に加えてCQLによる検索APIおよび，XML，HTML，JSON形式の返戻を実現する

課題

1. 一般公開の可能性を探る（アイデアを！）
2. 使い方事例の収集
   - ✓GISとの連携（特定地名の分布等）
3. APIとテキストマイニング技術の応用事例
   - ✓全文データ中から自動抽出した地名への緯度・経度データと包含関係・地名属性へのリンクデータを埋め込むアプリケーションを試作し、オントロジーデータベースを利用したアプリケーション連携の可能性を探る
4. 属性の拡充
   - ✓特に地名の経時変化等（使い方事例からのフィードバック）
5. コンテンツの拡充
   - ✓ここまで集めても不十分
   - ✓参加型のシステムが必要
   - ✓大きな資料がそろそろ枯渇（小字？）
   - ✓特定地域に限定すれば、地域研究者や郷土史家の協力を得やすいか

地名にかかる情報技術に関する研究会（2014 年 7 月 19 日開催）
報告書

編集　京都大学地域研究統合情報センター・地域情報学プロジェクト「地域の知」の情報学―時間・空間・語彙に注目した地域情報学の展開―

国立情報学研究所公募型共同研究「GeoNLP プロジェクト：自然言語文を対象とした地名情報処理とコミュニティの展開」

大学共同利用機関法人人間文化研究機構・研究資源共有化事業委員会

発行　2014 年 10 月 31 日
京都大学地域研究統合情報センター　地域情報学プロジェクト
〒606-8501 京都市左京区吉田下阿達町 46

印刷　大光社
〒604-0086 京都市中京区小川通丸太町下ル中之町 76

Printed in Japan
※図表の転用・転載を禁止します