

歴史地理データから考える 人文学DXの未来像



北本朝展 (KITAMOTO Asanobu)

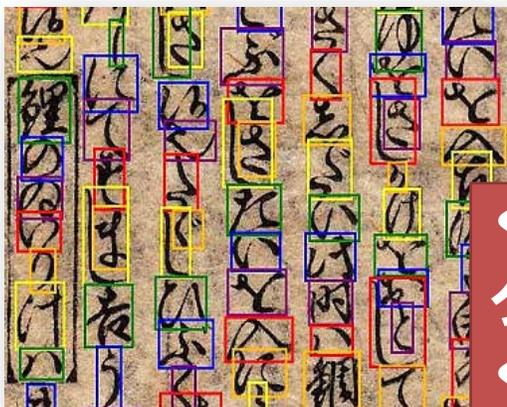
ROIS-DS人文学オープンデータ共同利用センター (CODH)

国立情報学研究所

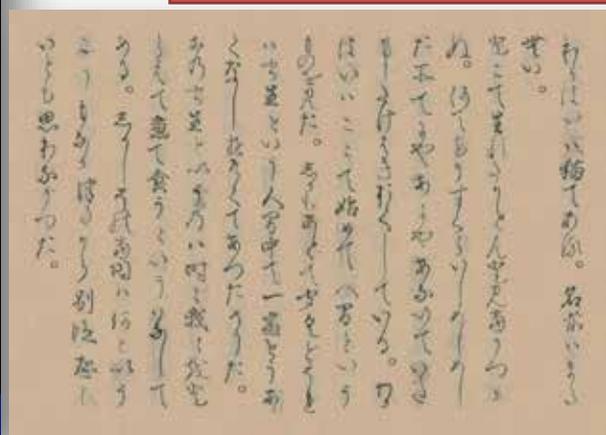
<http://codh.rois.ac.jp/>

<https://researchmap.jp/Kitamoto/>

ROIS-DS 人文学オープンデータ共同利用センター (CODH) <http://codh.rois.ac.jp/>



くずし字データセット
くずし字認識
くずし字生成



iiif
キュレーション



歴史ビッグデータ



歴史地理データ

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

1. 平凡社『日本歴史地名大系』のデータを一部オープン化
2. 平凡社地図出版と協働し、位置情報などの新たな価値を加え、機械可読データ化
3. ジャパンナレッジIDを用いて、オンラインDBとリンク
4. 出版社のオープン・クローズ戦略の一つの事例



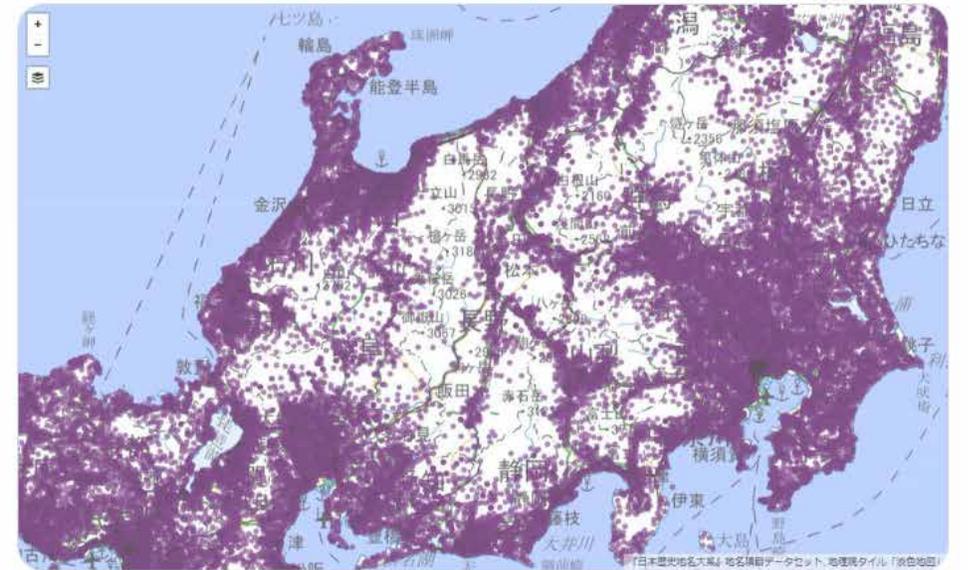
Center for Open Data in the Humanities (CODH) ✓

@rois_codh

歴史的地名の大規模オープンデータを公開しました。

codh.rois.ac.jp/news/#20231018

江戸時代の村までさかのぼる、79,502件の行政地名のデータです。平凡社地図出版と協働し、『日本歴史地名大系』の一部をオープン化しました。自分の先祖がどこに住んでいたかなど、地名から探っていくのも面白いですね。



午後5:21 · 2023年10月18日 · 79.3万 件の表示

📌 ポストのエンゲージメントを表示



5



3,422



7,491

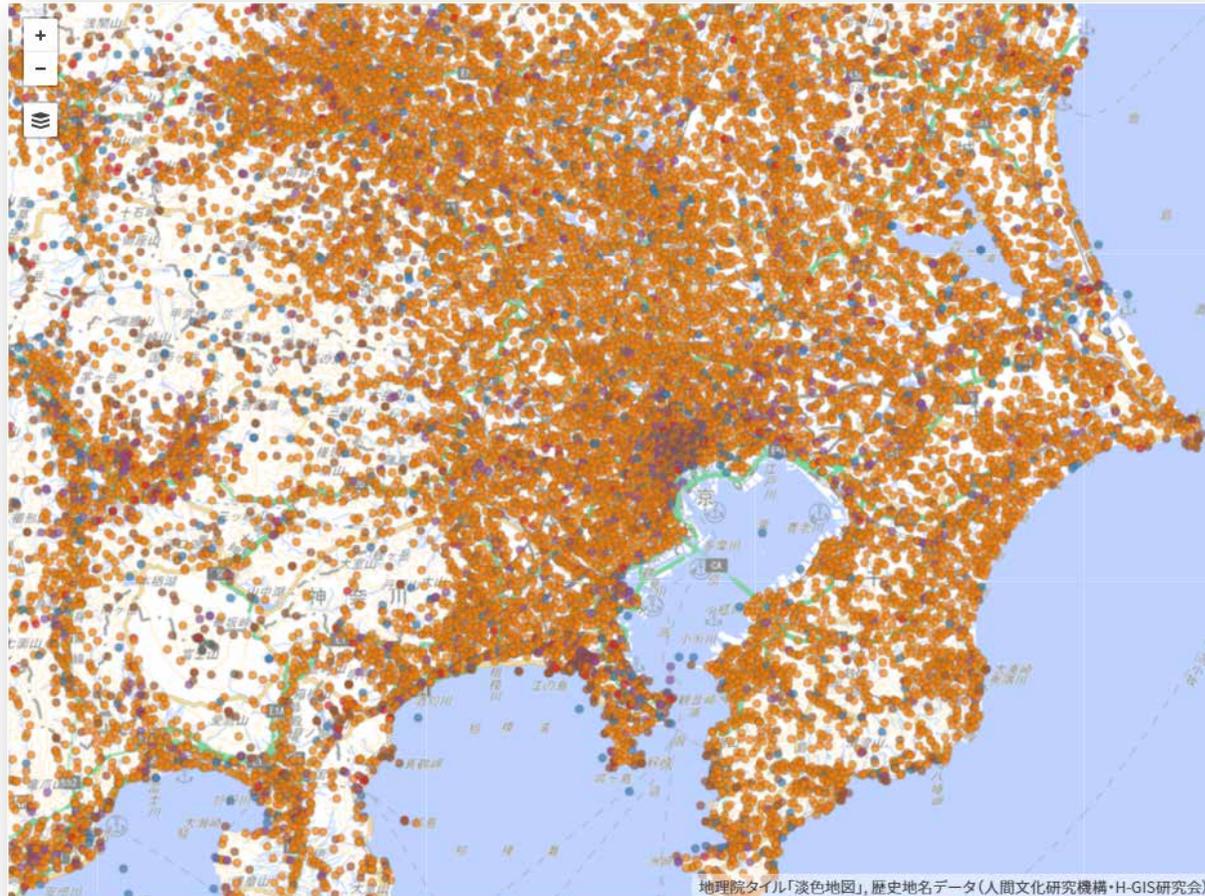


2,691



歴史地名マップ

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



マーカーの色は歴史地名データの属性に対応し、**行政地名**、**建物**、**水部**、**地形**、**名所旧跡**、**その他** とします。また

1. 人間文化研究機構・H-GIS研究会が公開する「**歴史地名データ**」298,914件を活用
2. 明治～昭和期の地図から抽出した地名が、全体の8割以上
3. **バイナリベクトルタイル**の活用により、多数地点をウェブ地図に表示

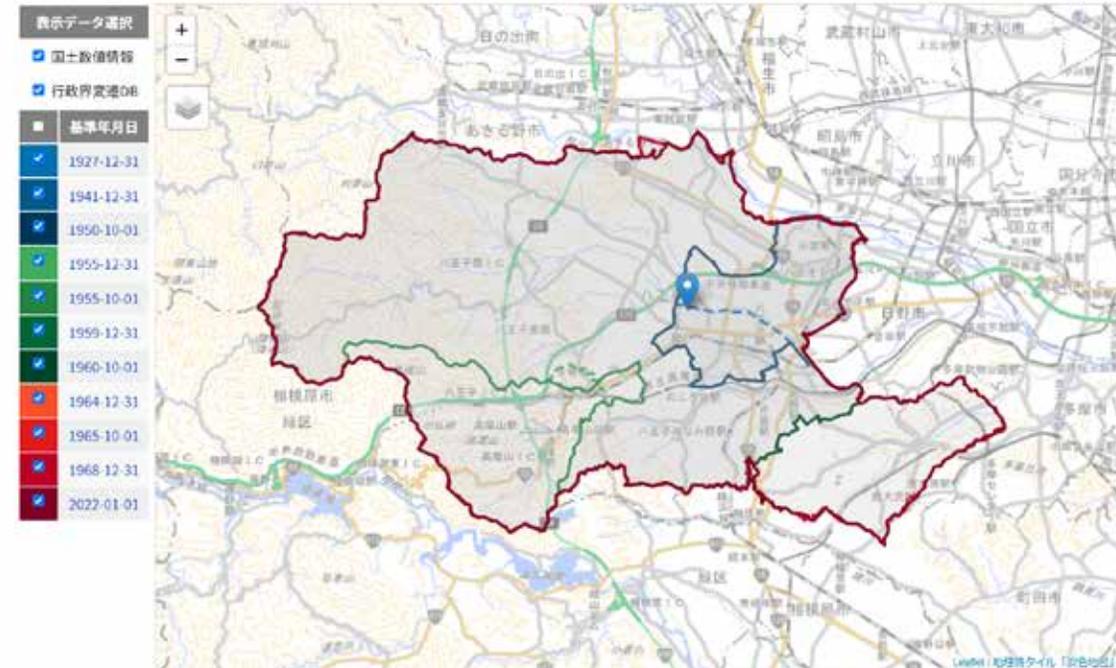
歴史的行政区域データセットβ版

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

行政区域境界の歴史の変遷

地図表示

市区町村の行政区域境界は出典のデータを変換し、国土交通省「国土数値情報」(実線)と筑波大学「行政界変遷データベース」(点線)に分けて表示します。ただしデータ作成方法の限界により、必ずしも当時の行政区域を正確に表しているわけではなく、両者のデータに不一致が生じている場合もあります。今後データの修正を進める計画ですが、正確な情報が必要な場合は、より信頼できる文献(古地図等)の情報を必ずご確認ください。



地図表示では、地図左上のレイヤ切り替えボタンを利用して、背景地図を変更できます。初期設定の「淡色地図」から「標準地図」「色別標高図」「陰影起伏図」「赤色立体地図」「傾斜層図」に変更すると、地形との位置関係を推察的に把握しやすくなります。

この市区町村との関係の歴史の変遷

赤は主要な市区町村、黄は重なりが大きい市区町村、灰は重なりが小さい(1%以下)市区町村を示します。なおポリゴンデータの注意事項も参考にしてください。

現在の市区町村行政区域と重なる過去の市区町村一覧

現在の市区町村行政区域を基準とし、それと過去の市区町村との重なり割合(「現在と過去の市区町村の共通部分の面積」/「現在の市区町村の面積」)を計算して一覧表示します。現在の市区町村が過去のどの市区町村で成り立っているかを調べるのに便利です。

表示データ選択: 国土数値情報 行政界変遷DB

市区町村選択: 重なりが小さい市区町村を隠す

市区町村	1920-01-01	1950-10-01	1955-10-01	1960-10-01	1965-10-01	2021-01-01
東京都八王子市/東京府八王子市 (13201A1568)	4%	17%	19%	40%	100%	100%
東京都多摩郡旗山町/東京府多摩郡旗山村 (1380070002)	4%	4%				
東京都多摩郡旗山町/東京府多摩郡旗山村 (1380070003)	2%	2%				
東京都多摩郡旗山町/東京府多摩郡旗山村 (1380070004)	4%	4%				
東京都多摩郡元八王子村/東京府多摩郡元八王子村 (1380030005)	7%	7%				
東京都多摩郡小宮村 (1380070006)	7%					
東京都多摩郡山口村/東京府多摩郡山口村 (1380070009)	1%	1%				
東京都多摩郡山口村 (1380070010)	1%					
東京都多摩郡由井村/東京府多摩郡由井村 (1380070018)	7%	7%				
東京都多摩郡由木村/東京府多摩郡由木村 (1380070019)	12%	12%	13%	12%		
東京都多摩郡小宮村/東京府多摩郡小宮村 (1380030009)						
東京都多摩郡旗山町 (1380070011)		1%	1%			
東京都多摩郡小宮町 (1320050001)						
市区町村	1920-01-01	1950-10-01	1955-10-01	1960-10-01	1965-10-01	2021-01-01

過去の市区町村行政区域と重なる現在の市区町村一覧

過去の市区町村行政区域を基準とし、それと現在の市区町村との重なり割合(「過去と現在の市区町村の共通部分の面積」/「過去の市区町村の面積」)を計算して一覧表示します。過去の市区町村が現在のどの市区町村に属しているかを調べるのに便利です。

表示データ選択: 国土数値情報 行政界変遷DB

市区町村選択: 重なりが小さい市区町村を隠す

市区町村	1920-01-01	2021-01-01
東京都八王子市/東京府八王子市 (13201A1568)	100%	100%
東京都昭島市 (13207A1964)		0%
市区町村	1920-01-01	2021-01-01

市区町村IDの付与



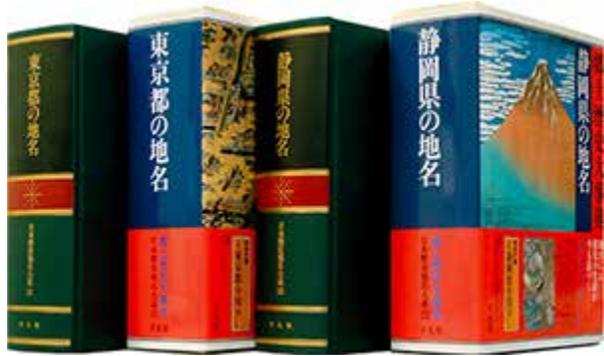
複数のオープンデータを統合し、1920年以降の市区町村に網羅的なIDを付与

- 市区町村コード (XXXXXAYYYY) : **4159**件
- 国土数値情報 (PPBQQQRRRR) : **12234**件
- 行政界変遷DB (PPCQQQRRRR) : **336**件
- 合計 : **16729**件

- 1919年から1889年 (市制及町村制) への遡及
- 明治時代 (1888年以前) から江戸時代への遡及

出版社との協働

<https://japanknowledge.com/contents/rekishi/>

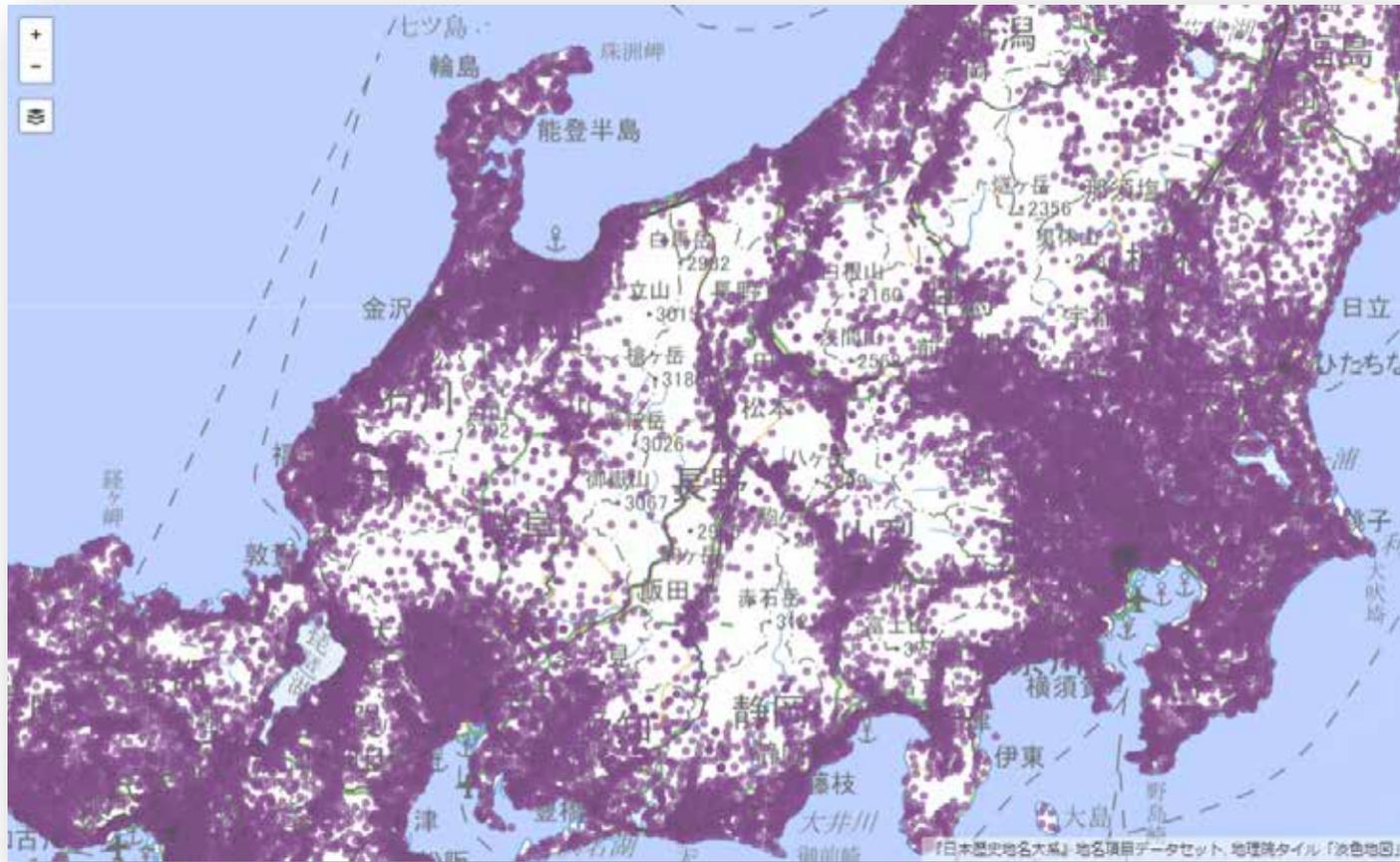


日本歴史地名大系
平凡社
巻冊数：50巻51冊
書籍版：1979-2004
現在、ジャパンナレッジで電子版を公開中

1. 出版社が有する大規模・高品質データをどのように活用するか？
2. 日本歴史地名大系（平凡社）の「行政地名変遷表」を活用し、市区町村IDを江戸時代に接続できないか？
3. 一般社団法人「百科総合リサーチ・センター」が進める、行政地名変遷表のデジタルデータ公開に向けた作業に協力
4. 平凡社地図出版とも連携し、約8万の行政地名項目をオープンデータ化

『日本歴史地名大系』地名項目データセット

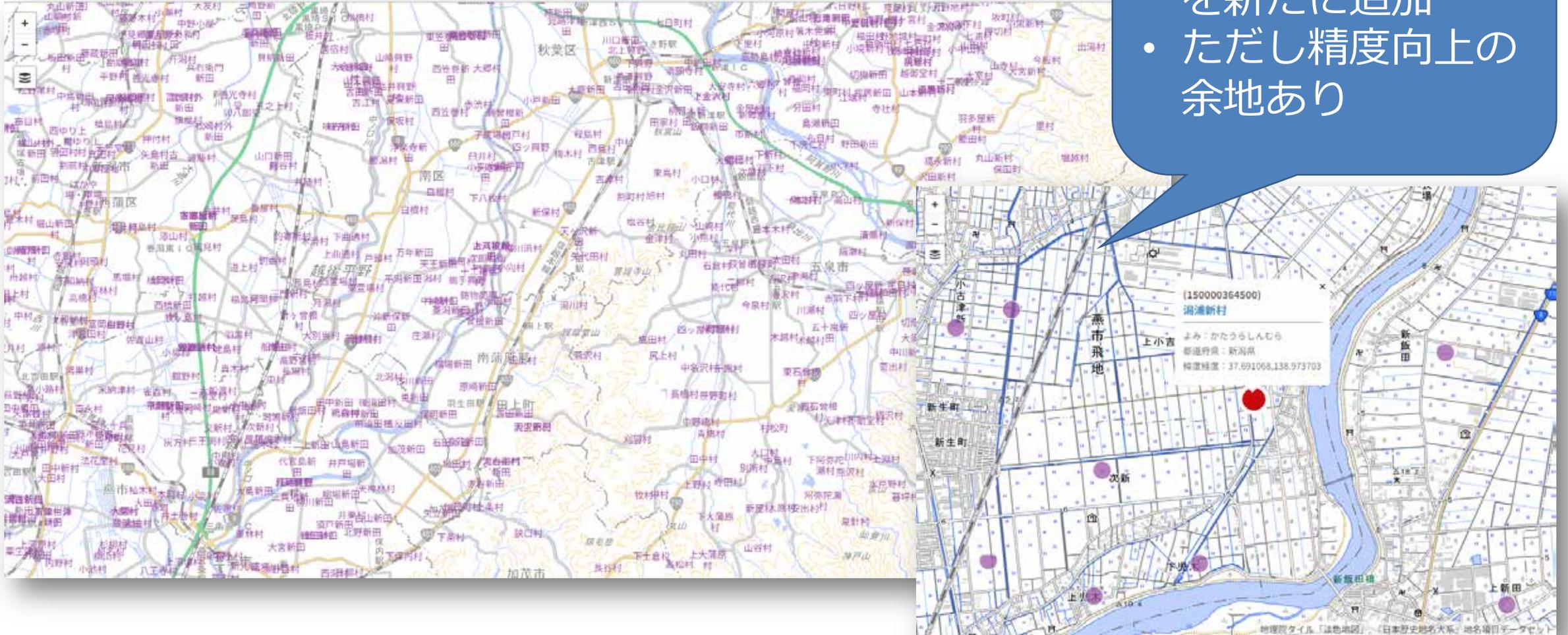
<https://geoshape.ex.nii.ac.jp/nrct/>



1. 日本全国約8万件の地名項目をデータセット化
2. jageocoder / 歴史地名データを利用し緯度経度を推定
3. 読みのデータも統合し利便性を向上

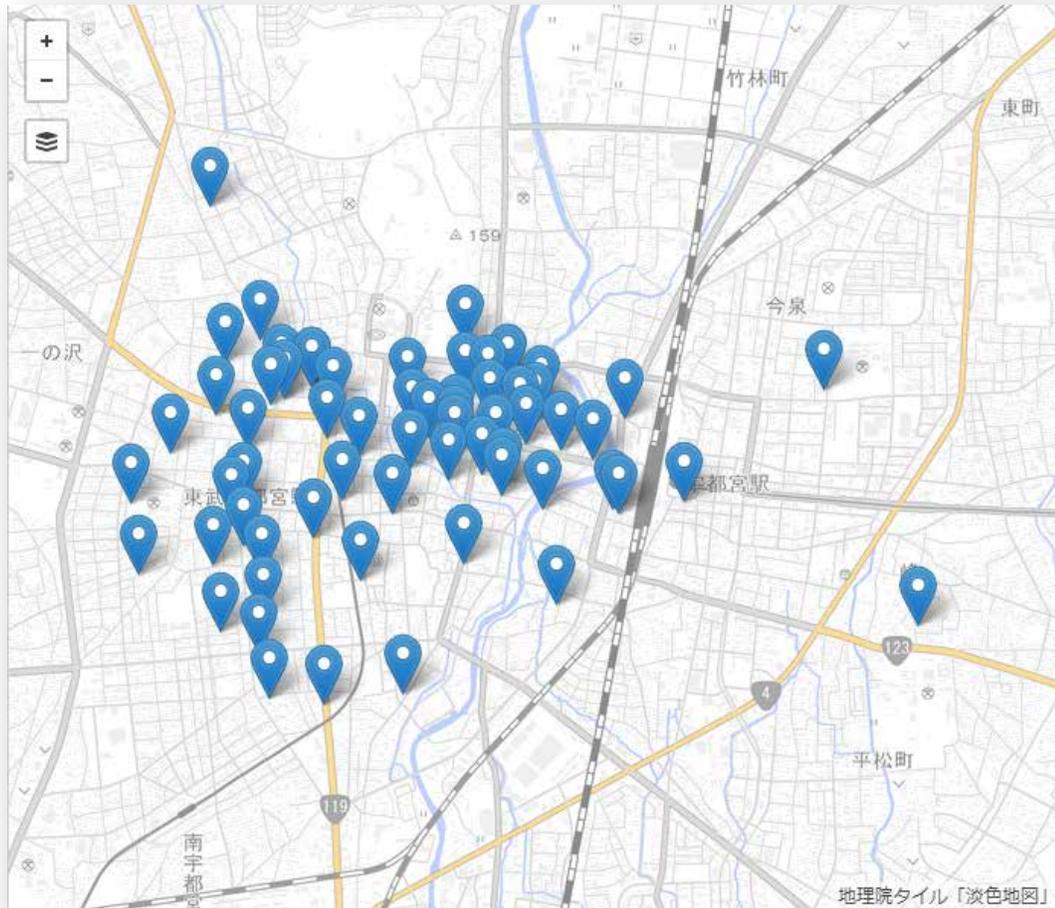
地名項目データセット拡大図

- 『日本歴史地名大系』に、位置情報を新たに追加
- ただし精度向上の余地あり



『日本歴史地名大系』 行政地名変遷データ セット

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



1. 歴史的行政区域データセットと連続する市区町村ID
2. 市制・町村制以後の市区町村と『日本歴史地名大系』地名項目との関係を整理
3. 市区町村→藩政村→ジャパ
ンナレッジのリンクが接続
4. 完了：2県／47都道府県

Japan Knowledge

日本歴史地名大系

クローズ

オープン

ジャパナ
レッジID

日本歴史地名統
合データベース

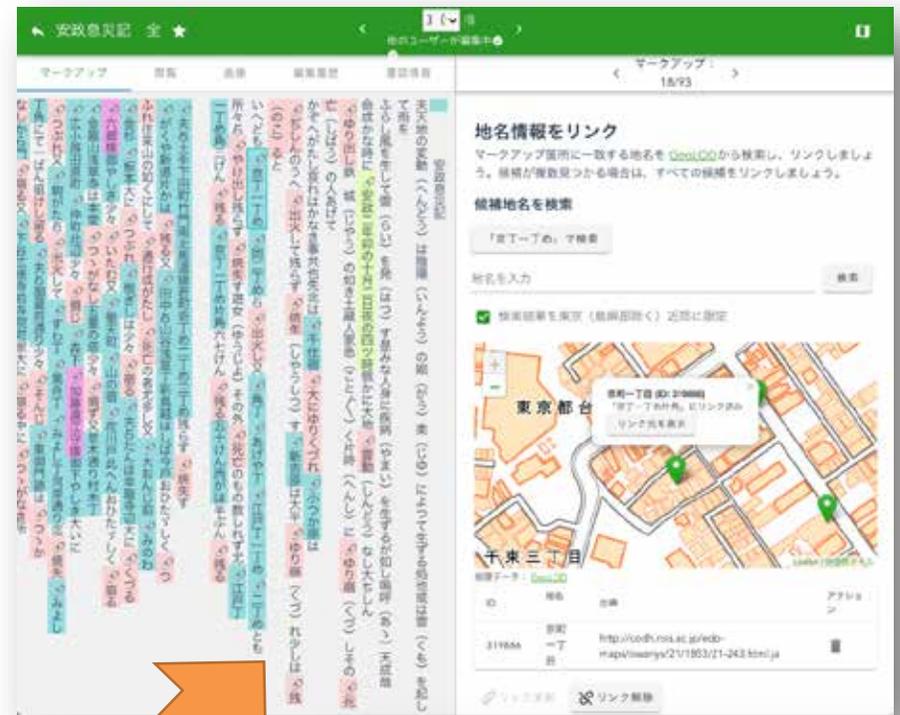
市区町村ID

歴史的行政区域
データセット

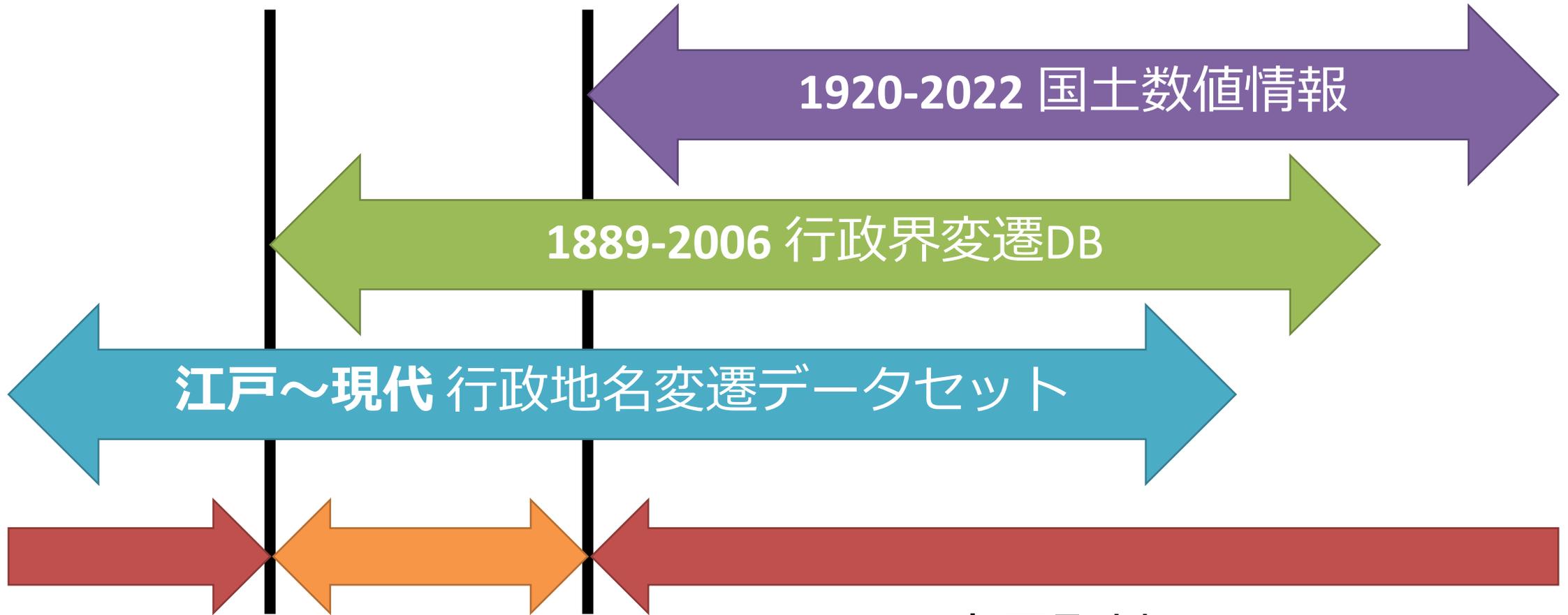
GeoLOD ID

GeoLOD

識別子サービス



市区町村IDの付与（進行中）



過去と現代のギャップをつなぐ

市区町村ID

地名識別子と疎結合性

私は**東京**に行く



35.676666, 139.762222

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

私は**東京**に行く



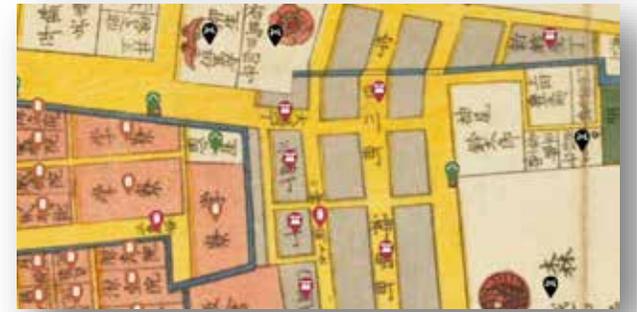
地名識別子 : UoGwci



35.676666, 139.762222

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

史料と地名のリンク



宇田川丁、三島丁・
神明丁、此分潰家多、
土蔵残所なし

固有表現認識

宇田川丁、三島丁・
神明丁、此分潰家多、
土蔵残所なし

曖昧性解消

御江戸大地震大破并出火類焼場等書上之写（みんなで翻刻）

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

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

すべての論文に地名識別子を埋め込む

大垣新田藩領渥美郡畠村に進出した知多大野商人 (3)				
傳助かし家 沓軒 綿諸色 尾州大野北浦町	次郎右衛門	一借宅荷イ商人 内	三軒	NRCT:250000093200 GeoLOD:Q4O6Wc
与右衛門かし家 沓軒 綿諸色 尾州名古屋杉町	与八	惣次郎かし屋 沓軒 鍋売	近江国栗太郡辻村 清吉	
五助かし家 沓軒 綿諸色 尾州大野高須賀町	九三郎	喜右衛門かし屋 沓軒 茶売	尾州大野柴崎村 藤八	NRCT:230000328800 GeoLOD:4YKcwm
五平次かし家 沓軒 綿諸色 尾州大野北浦町	十兵衛	善八かし屋 沓軒 茶売	尾州大野新町 四郎三郎	
甚太郎かし家 沓軒 綿諸色 尾州大野高須賀町	吉田田町 九左衛門	善八かし屋 沓軒 茶売	幡豆郡谷村 弥三郎	NRCT:230000521300 GeoLOD:yNkkCa
勘四かし家 沓軒 綿諸色 尾州大野高須賀町	平七	源三郎方ニ罷有候 一鍋売出入荷イ商人 源八	近江国栗太郡辻村 源八	
惣休かし家 沓軒 尾州大野大草村 彦八		一出入茶売商人 内	若見村 多吉	NRCT:230000353200 GeoLOD:8oCIQP
4年前の宝永4年と比較すると、地元商人は、12軒から8軒に減っている。善八と新兵衛の2人だけが残っているだけで他の6人は新たに参入している。彼らは、「味噌塩酢酒穀類諸色少々宛商耕作致両かせぎ仕候」であ		半七郎方罷有候 沓軒 茶売	額田郡檉山村 清九郎	

地名に識別子を埋め込むと

同名地名の区別が可能

外部情報源 (JK等) へのリンクも可能

石川洋一. "大垣新田藩領渥美郡畠村に進出した知多大野商人." 愛知大学総合郷土研究所紀要 68 (2023): 51-59

天文学における識別子の活用

<https://simbad.u-strasbg.fr/simbad/>

Query : M31

Basic data :
M 31 -- Galaxy
Other object types: LIN (), G (2006AJ,LEDA,...), * (AG,BD,...), GSO (2010A&A,[VV2006],...), AGN ([VV2000c],[VV2003c],...), *gam* (2FGL,3FGL,...), Rad (2C,DA,...), IR (IRAS,IRC,...), X (2MAXI,XSS), GiC (GIN), GiG (K79)

ICRS coord. (*ep=J2000*) : 00 42 44.330 +41 16 07.50 (Infrared) [] C 2006AJ....131.1163S
FK4 coord. (*ep=B1950 eq=1950*) : 00 40 00.095 +40 59 41.73 []
Gal coord. (*ep=J2000*) : 121.174329 -21.573309 []
Radial velocity / Redshift / cz : V(km/s) -300.0 [4.0] / z(spectroscopic) -0.001000 [0.000013] / cz -299.85 [4.00] C 2012AJ....144....4M
Parallaxes (*mas*): 6.0 [14.1] E 1995GCTP..C.....0V
Morphological type: SA(s)b D 2013AJ....146...67B
Angular size (*arcmin*): 199.53 70.79 35 (Obj) D 2003A&A...412...45P
Fluxes (6) :
U 4.86 [0.03] D 2007ApJS..173..185G
B 4.36 [0.02] D 2007ApJS..173..185G
V 3.44 [0.03] D 2007ApJS..173..185G
J 2.094 [0.016] C 2006AJ....131.1163S
H 1.283 [0.017] C 2006AJ....131.1163S
K 0.984 [0.017] C 2006AJ....131.1163S

References (12047 between 1850 and 2023) (Total 12047)
Simbad bibliographic survey began in 1850 for stars (at least bright stars) and in 1983 for all other objects (outside the solar system).
[Follow new references on this object](#)

Reference summaries :
from: 1850 to: \$currentYear
Display or select by : (not exhaustive, [explanation here](#))

2022A&A...657A..150 [X ,2]
Astronomy and Astrophysics, volume 657A, 15-15 (2022/1-1)
A first estimate of the Milky Way dark matter halo spin.
OBREJA A., BUCK T. and MACCIO A.V.
Simbad objects: 6

2022A&A...657A..26M [X ,1]
Astronomy and Astrophysics, volume 657A, 26-26 (2022/1-1)
Discovery of four super-soft X-ray sources in XMM-Newton observations of the Large Magellanic Cloud.
MAITRA C. and HABERL F.
Simbad objects: 23

2022A&A...657A..41R [X ,1]
Astronomy and Astrophysics, volume 657A, 41-41 (2022/1-1)
The outermost stellar halo of NGC 5128 (Centaurus A): Radial structure.
REJKUBA M., HARRIS W.E., GREGGIO L., CRNOJEVIC D. and HARRIS G.L.H.
<Available at CDS ([J/A+A/657/A41](#)): fields.dat catalog.dat>
Simbad objects: 7
Status at CDS: *Tables of objects will be appraised for possible ingestion in SIMBAD.*

すべての天体に識別子が付与され、
各種の情報が統合されているため、
例えば「ある天体に関する論文」な
どが簡単に検索できる

日本古典籍にも識別子が付与されている

<https://www.nijl.ac.jp/pages/cijproject/>

DOI = Digital Object Identifier (文書の識別子)

Point2 引用しやすい

電子データに付与される国際的識別子、DOI (デジタルオブジェクト識別子) の採用により、永続的なアクセスを保証。引用したいときに便利!

 <https://www.doi.org/>



データベースの URL が代わっても DOI は変わらず、リンク切れが発生しない

国書DBパンフレット

https://kokusho.nijl.ac.jp/page/kokusho-pamphlet_202309_web_JP.pdf

1. 国文研の古典籍にはDOIが既に付与されている
2. すべての論文がDOIを使えば、「この本に関する論文を検索」することは可能
3. あとはやるだけ

大規模言語モデル（LLM）の活用

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

LLMを人文学研究でどのように活用するかは、今後の重要な課題である

1. LLM自体に人文学データを入れる（くずし字翻刻→大規模コーパス構築→LLM事前学習）
2. 各種検索エンジンのフロントエンドとしてLLMを活用する（データベース構築→自然言語検索・要約 by LLM）

例：ある識別子に関する論文を自然言語で検索→その識別子に関する研究成果をLLMで要約→自分の研究に活用

人文学DXに向けて

1. 人文学DXとは、機械の支援を受けることで、研究を深化、大規模化、多様化することである
2. 機械の支援を受けやすくするには、機械可読データを強化し、AI用学習データも増やす必要がある
3. 出版社が有する大規模・良質のデータを機械可読・オープン化し、共存共栄を目指すことが重要である
4. 研究者が識別子を使いこなすことで、知の世界はより構造化・整理され、研究にも活用しやすくなる