

長期・歴史的な基盤データの構築と分析・可視化 ～デジタル台風・歴史的行政区域 データセット・CODH等の事例紹介～

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 (NII)

北本 朝展 (Kitamoto Asanobu)

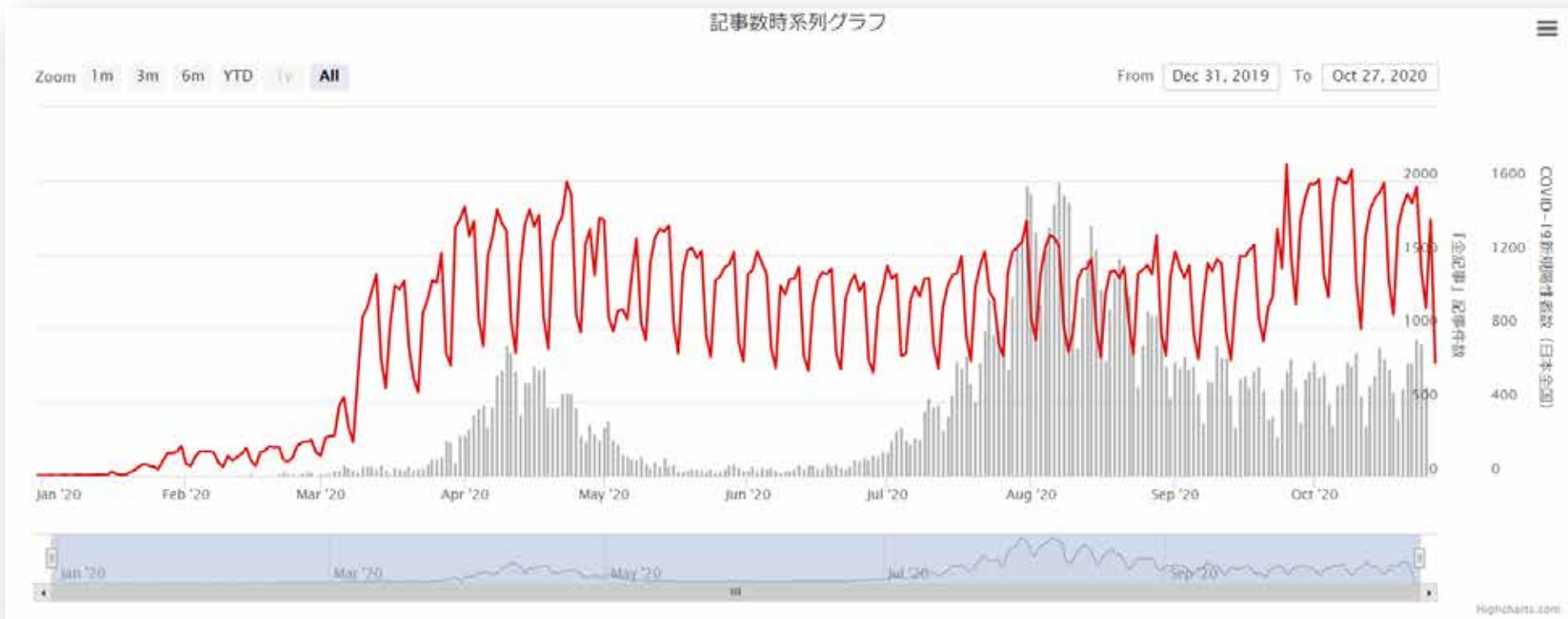
<https://researchmap.jp/kitamoto/>

長期・歴史的な基盤データ

1. **長期** = (例えば) 10年以上にわたって、連続的に構築されたデータ。
2. **歴史的** = (例えば) 現在から数十年前のデータを含む過去のデータ。
3. **基盤** = 他の研究から参照される、または他の研究で再利用されるデータ。
4. **検索技術の開発、統一的な識別子の付与、データ多様性に対応できる運用。**

新型コロナウイルス感染症 (COVID-19) ニュース分析

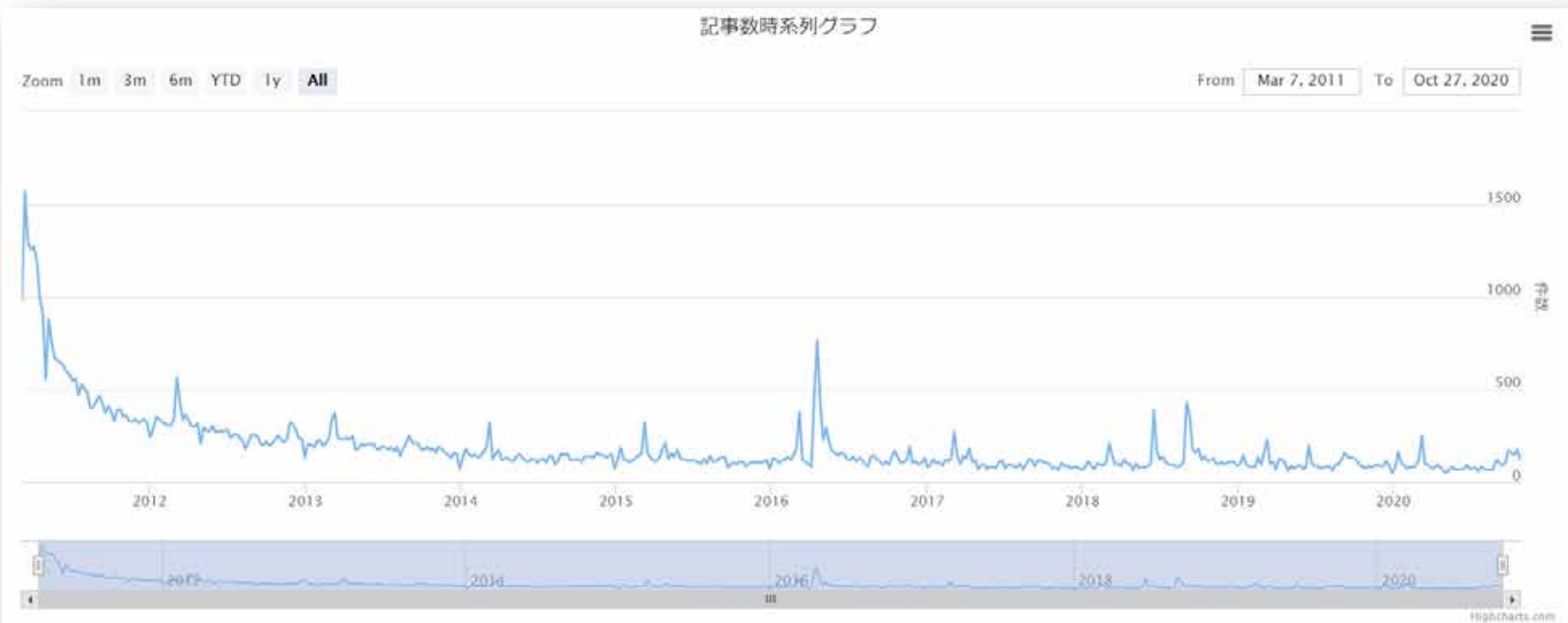
<http://agora.ex.nii.ac.jp/crisis/covid-19/mass-media/>



データを収集し続け、整理し続けることで、長期データが生まれる

東日本大震災ニュース分析

<http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/mass-media/>



データを収集し続け、整理し続けることで、長期データが生まれる

デジタル台風

<http://agora.ex.nii.ac.jp/digital-typhoon/>

The screenshot displays the 'Digital Typhoon: Typhoon Images and Information' website. The header includes navigation links like Home, Earth, and Digital Typhoon, along with a search bar and site map. A section titled 'Real-time Typhoon Information' (Updated: 2010-09-01 12:42 (UTC)) features three columns for typhoons T 201006, T 201007, and T 201008, each with a satellite image and a table of data including name, date, location, and pressure. To the right, a 'Number of Typhoons' section shows statistics for the current year and an average, followed by a 'Latest Typhoon Information' section with links to news, topics, and other resources. At the bottom, there is a 'Typhoon Database' with various search filters and a 'Meteorological Satellite Images' section showing a satellite image of the Earth from 2010-09-01 11:00 (UTC).

1. 「台風」をテーマに、あらゆるデータを統合したウェブサイト。
2. 現在の状況をキーとして過去の類似事例を検索できる。
3. 専門家から市民まで、多くの人々が調査研究や趣味などに活用している。

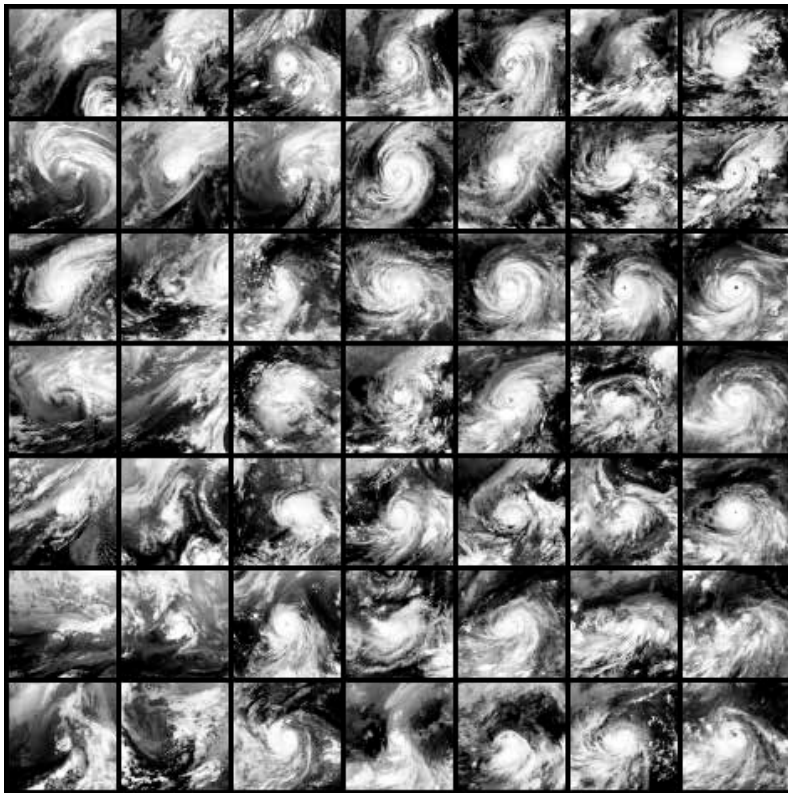
年間約2000万PVに達する、日本で最も著名な台風情報サイトの一つ。

多種データアーカイブ

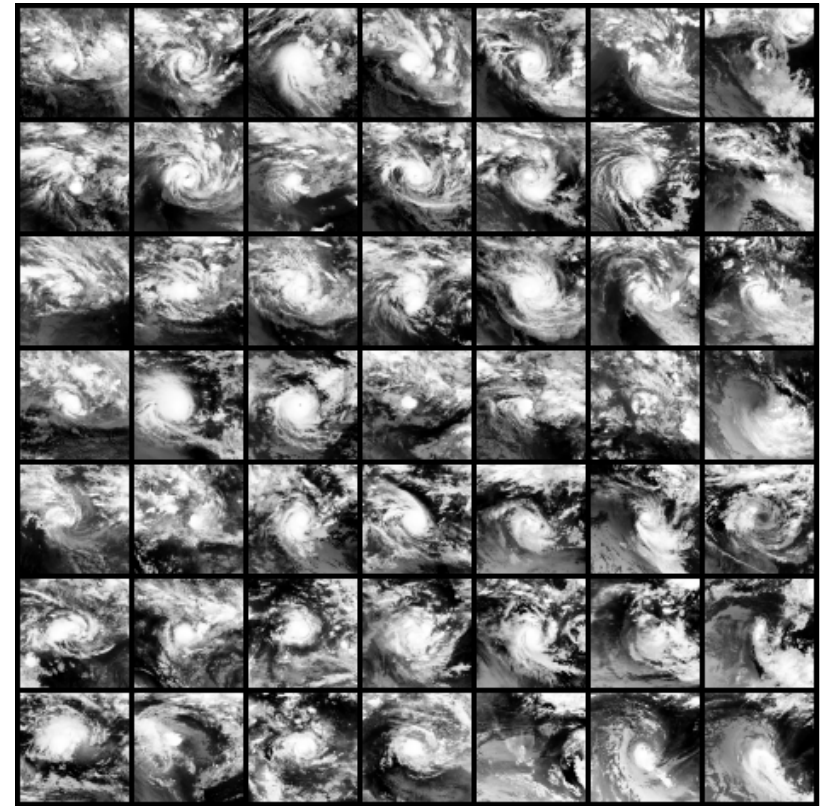
静止気象衛星画像（全球）	1978-
静止気象衛星画像（台風）	1978-
台風ベストトラック（軌跡）	1951-, 1907-
アメダスデータ（センサ）	1976-
災害レポート（構造化データ）	1971-
気象レーダー（2次元×時間）	2004-, 1998-
気象庁数値予測データ（3次元×時間×変数）	2002-
オンラインニュース（テキスト）	2003-
ソーシャルメディア（テキスト）	2004-, 2009-
天気図（線画）	1883-
気象庁防災情報XML（構造化データ）	2012-

台風画像コレクション

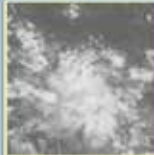
1978年以来、178,000画像（北）、58,500画像（南）



北半球

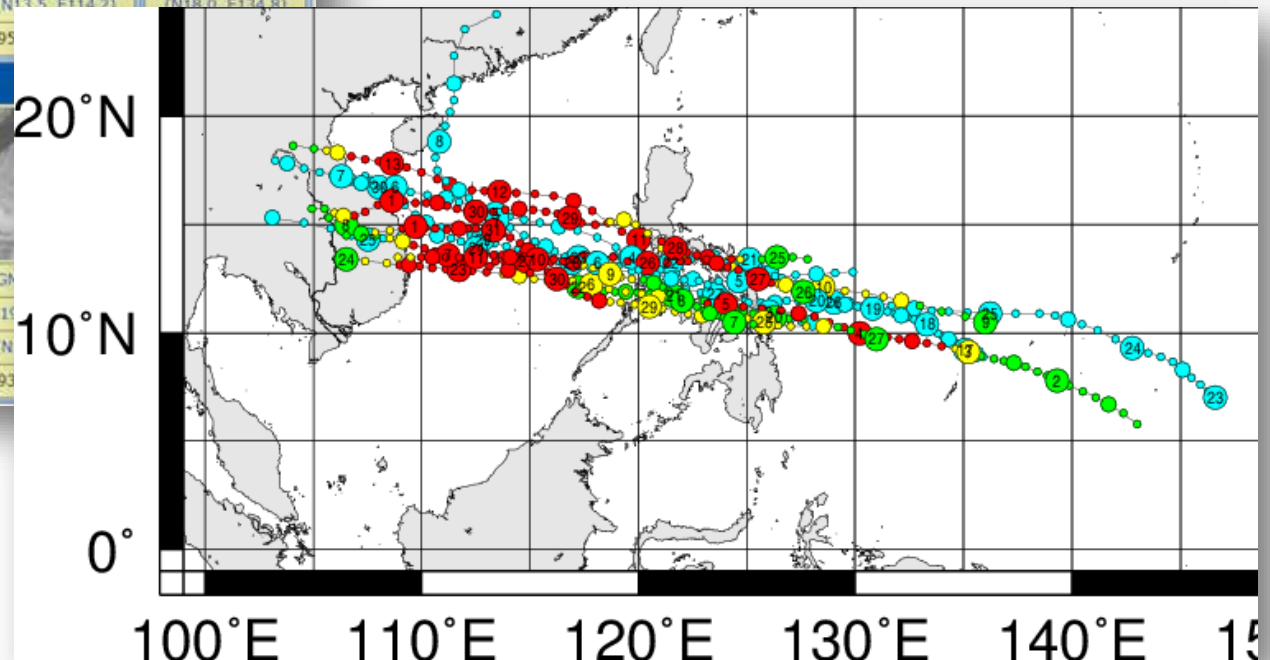


南半球

Query 1	1	2	3	4
				
HMW820102705	GMS596072313	HMW819100622	GMS490082410	MTS113110712
202018 (WNP)	199608 (WNP)	201919 (WNP)	199015 (WNP)	201330 (WNP)
(N13.3, E113.5)	(N16.4, E125.4)	(N15.0, E150.1)	(N12.7, E143.8)	(N10.2, E129.1)
950 hPa / 85 kt	980 hPa / 50 kt	955 hPa / 80 kt	1002 hPa / 0 kt	895 hPa / 125 kt
5	6	7	8	9
				
GMS500102713	GOE903112209	GMS492112205	GMS501111011	GMS492083120
200020 (WNP)	200321 (WNP)	199230 (WNP)	200123 (WNP)	199215 (WNP)
(N13.4, E124.8)	(N8.8, E151.3)	(N12.6, E149.2)	(N13.5, E114.2)	(N18.0, E134.8)
975 hPa / 60 kt	985 hPa / 50 kt	915 hPa / 100 kt	950 hPa / 0 kt	950 hPa / 0 kt
10	11	12	13	14
				
GOE904112801	GOE903071816	GMS388061718	GMS388061718	GMS388061718
200427 (WNP)	200307 (WNP)	198804 (WNP)	198804 (WNP)	198804 (WNP)
(N5.7, E153.5)	(N10.5, E136.8)	(N5.0, E141.0)	(N5.0, E141.0)	(N5.0, E141.0)
1006 hPa / 0 kt	990 hPa / 45 kt	1008 hPa / 0 kt	1008 hPa / 0 kt	1008 hPa / 0 kt

過去の検索

1. 様々な種類のデータに対して類似度を定義して検索。
2. 過去の事例を未来の対策への参考とする。



100年天気図データベース

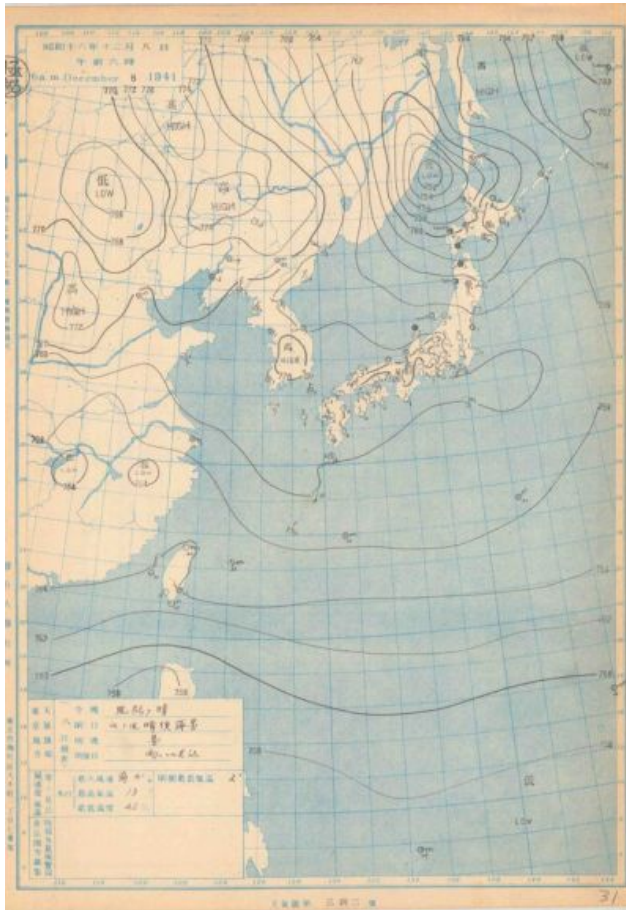
<http://agora.ex.nii.ac.jp/digital-typhoon/weather-chart/>



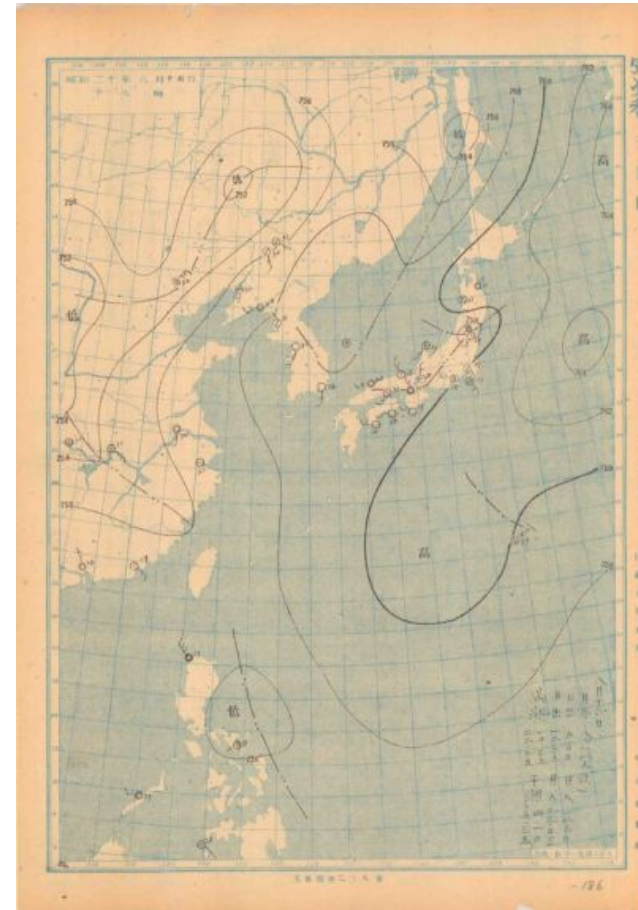
- 1883年3月1日以来の天気図アーカイブ。日時で検索し、過去の天気図にアクセスできる。
- 1883年当時の最新通信技術「電信」が迅速な天気図作成を実現した。
- 気象のみならず社会の動きも反映している。

注：気象庁「天気図」をもとに国立情報学研究所「デジタル台風」が作成

史料としての読み解き



太平洋戦争開戦（1941年12月8日）



太平洋戦争敗戦（1945年8月15日）

歴史的行政区画データセットβ版

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



- 1920年から2020年まで、28時点。
- 行政区画件数 = 16,447件、行政区画境界データ件数 = 99,296件
- 2017年1月に公開。毎年データ更新と機能追加を継続中。

データソース

単なる「行政区域データ」の形式変換ではない！データのクリーニングも必要。

1. **国土数値情報「行政区域データ」**（N03）
 - 1920年～2020年
2. **国土数値情報「市区町村役場データ」**（P34）
 - 2014年
3. **国土数値情報「市町村役場等及び公的集会施設データ」**（P05）
 - 2010年
4. **国土数値情報「郵便局データ」**（P30）
 - 2013年
5. **政府統計の総合窓口（e-Stat）「廃置分合等情報」**
 - 1970年4月1日以降
6. **政府統計の総合窓口（e-Stat）「標準地域コード」**
7. **政府統計の総合窓口（e-Stat）「統計LOD」**

地名の変遷

江戸

東京府東京市
(15区)

東京府東京市
(35区)

東京都特別区部
(23区)

1. 行政地名とその境界はいつどう変遷してきたか？
2. 公式データ：全国地方公共団体コードが定められた1968年以降のみ。
3. 行政界変遷データベース（筑波大学）：所属自治体の変遷を採録。1889年から2006年まで。
4. 『全訂 全国市町村名変遷総覧』（加除出版）。

市区町村IDの付与

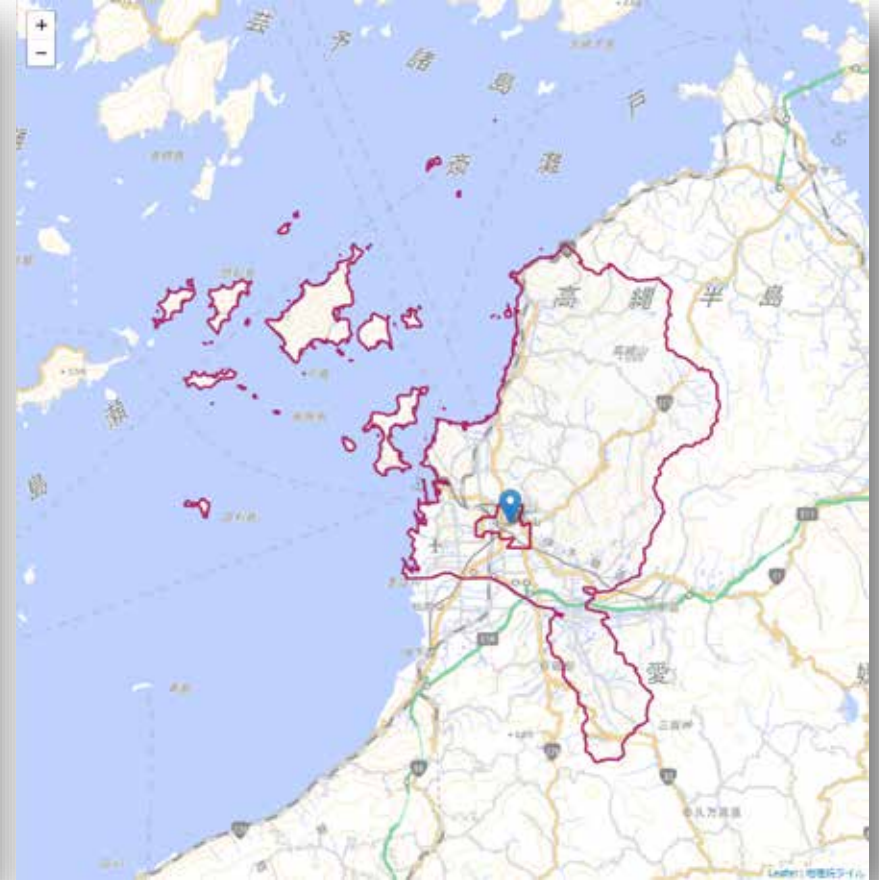
13B0010022	東京都野増村/東京府野増村
13B0030001	東京都西多摩郡霞村/東京府西多摩郡霞村
13B0030002	東京都西多摩郡吉野村/東京府西多摩郡吉野村
13B0030005	東京都西多摩郡古里村/東京府西多摩郡古里村
13B0030006	東京都西多摩郡戸倉村/東京府西多摩郡戸倉村
13B0030007	東京都西多摩郡三田村/東京府西多摩郡三田村
13B0030008	東京都西多摩郡小河内村/東京府西多摩郡小河内村
13B0030009	東京都西多摩郡小宮村/東京府西多摩郡小宮村
13B0030010	東京都西多摩郡小曾木村/東京府西多摩郡小曾木村
13B0030013	東京都西多摩郡成木村/東京府西多摩郡成木村
13B0030014	東京都西多摩郡西秋留村/東京府西多摩郡西秋留村
13B0030015	東京都西多摩郡西多摩村/東京府西多摩郡西多摩村
13B0030016	東京都西多摩郡青梅町/東京府西多摩郡青梅町
13B0030019	東京都西多摩郡増戸村/東京府西多摩郡増戸村
13B0030020	東京都西多摩郡多西村
13B0030021	東京都西多摩郡大久野村/東京府西多摩郡大久野村
13B0030022	東京都西多摩郡調布村/東京府西多摩郡調布村
13B0030025	東京都西多摩郡東秋留村/東京府西多摩郡東秋留村
13B0030028	東京都西多摩郡氷川町
13B0030030	東京都西多摩郡平井村/東京府西多摩郡平井村
13B0070001	東京都南多摩郡稲城村/東京府南多摩郡稲城村
13B0070002	東京都南多摩郡横山村/東京府南多摩郡横山村

1. 1968年以降 (A) : 公式の市区町村コードが存在。
2. 1968年以前 (B) : 名称の連続性などを根拠に固有のIDを付与 (網羅性無保証)。
3. 2種のID形式に分割。
 1. XXXXXAYYYY
 2. PPBQQQRRRR

合併で拡大した市区町村の例



山形県鶴岡市



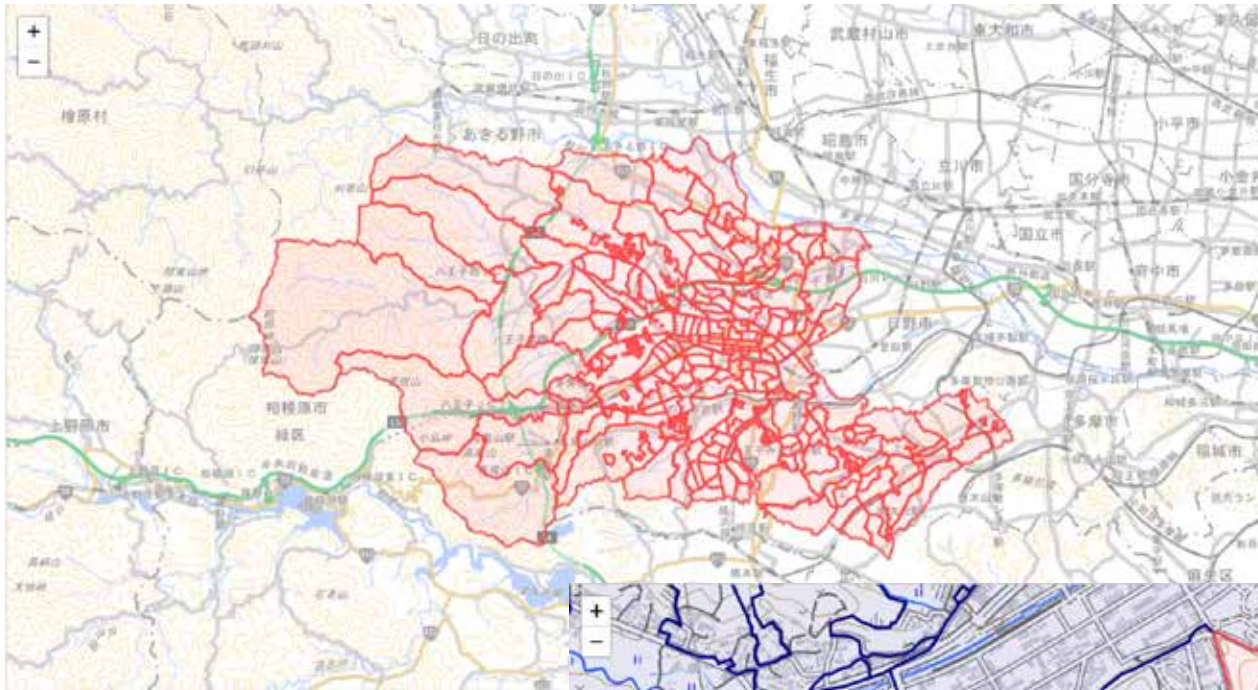
愛媛県松山市

市町村大合併と地域感覚

1. **明治の大合併**（約300～500戸規模）：71,314
（M21）→ 15,859（M22）
 2. **昭和の大合併**（約8,000人規模）：9,868
（S28）→ 3,472（S36）
 3. **平成の大合併**：3,218（H14）→ 1,727（H22）
- 幾度もの大合併を経て大きくなり過ぎた市区町村では、**行政区域内部での地域差が拡大**。
 - **旧市区町村名は住民の日常生活や歴史の記憶とリンク。災害情報等も感覚的に理解できる。**

参考：<https://www.soumu.go.jp/gapei/gapei2.html>

町丁・字 の表示



<https://geoshape.ex.nii.ac.jp/ka/resource/13201.html>



<https://geoshape.ex.nii.ac.jp/ka/resource/13/13201050001.html>

国勢調査町丁・字等別境界データセット

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

1. 国勢調査実施毎（5年毎）に設定した調査区の境界を中心としたデータセット。
2. 2015年版のデータとして、町丁・字等情報 = **219,271**件を公開。
3. 政府統計の総合窓口（e-Stat）のデータをNICTが加工し、NIIがさらに加工。
4. **時間は1点のみだが、空間解像度は市区町村行政区域よりも細かい。**

ベクトルタイル地図

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

歴史的行政区画データセットβ版 ベクトルタイル地図 | Geoshapeリポジトリ

GeoNLPプロジェクト | GeoNLP | GeoNLP地名ウェブサービス | GeoCOD | Geoshapeリポジトリ | 歴史的行政区画データセットβ版 | ベクトルタイル地図

歴史的行政区画データセットβ版のコロナレス地図（塗り分け地図）をベクトルタイル地図として提供します。これにより、コロナレス地図よりも軽量に高精細データを表示することが可能となります。

データ作成時点および表示方法（政令指定都市の統合）を選択してください。

データ作成時点：2020-01-01 政令指定都市の統合：政令指定都市を統合しない

緯度経度を指定すると、市区町村の変遷を一覧できる。



利用例

1. デジタル台風：歴史災害データベース

GeoNLPプロジェクト | GeoNLP | GeoNLP地名ウェブサービス | GeoCOD | Geoshapeリポジトリ | 歴史的行政区画データセットβ版 | ベクトルタイル地図

Aianoba KITAMOTO, ROIS-DS Center for Open Data in the Humanities, CC BY-SA 4.0

歴史的行政区画データセットβ版 ベクトルタイル地図 | Geoshapeリポジトリ

GeoNLPプロジェクト | GeoNLP | GeoNLP地名ウェブサービス | GeoCOD | Geoshapeリポジトリ | 歴史的行政区画データセットβ版 | ベクトルタイル地図

歴史的行政区画データセットβ版のコロナレス地図（塗り分け地図）をベクトルタイル地図として提供します。これにより、コロナレス地図よりも軽量に高精細データを表示することが可能となります。

データ作成時点および表示方法（政令指定都市の統合）を選択してください。

データ作成時点：2020-01-01 政令指定都市の統合：政令指定都市を統合しない

緯度：35 経度：135 更新

データ作成時点を変更



利用例

1. デジタル台風：歴史災害データベース

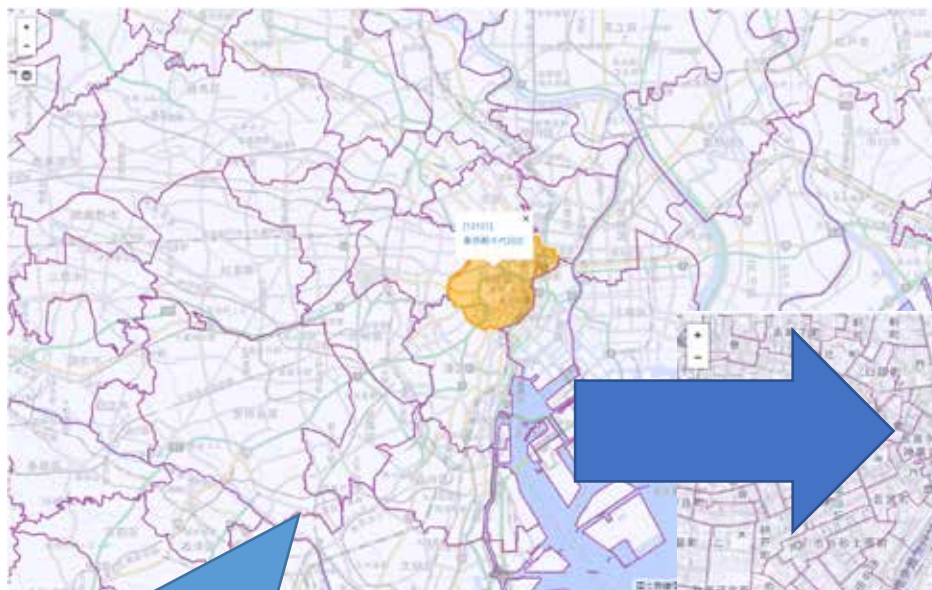
GeoNLPプロジェクト | GeoNLP | GeoNLP地名ウェブサービス | GeoCOD | Geoshapeリポジトリ | 歴史的行政区画データセットβ版 | ベクトルタイル地図

Aianoba KITAMOTO, ROIS-DS Center for Open Data in the Humanities, CC BY-SA 4.0

北緯35度、東経135度を入力

市区町村と町丁・字の統合

<https://geoshape.ex.nii.ac.jp/vector-adm/>



ズームレベル12までは、市区町村行政区域データセットを表示。

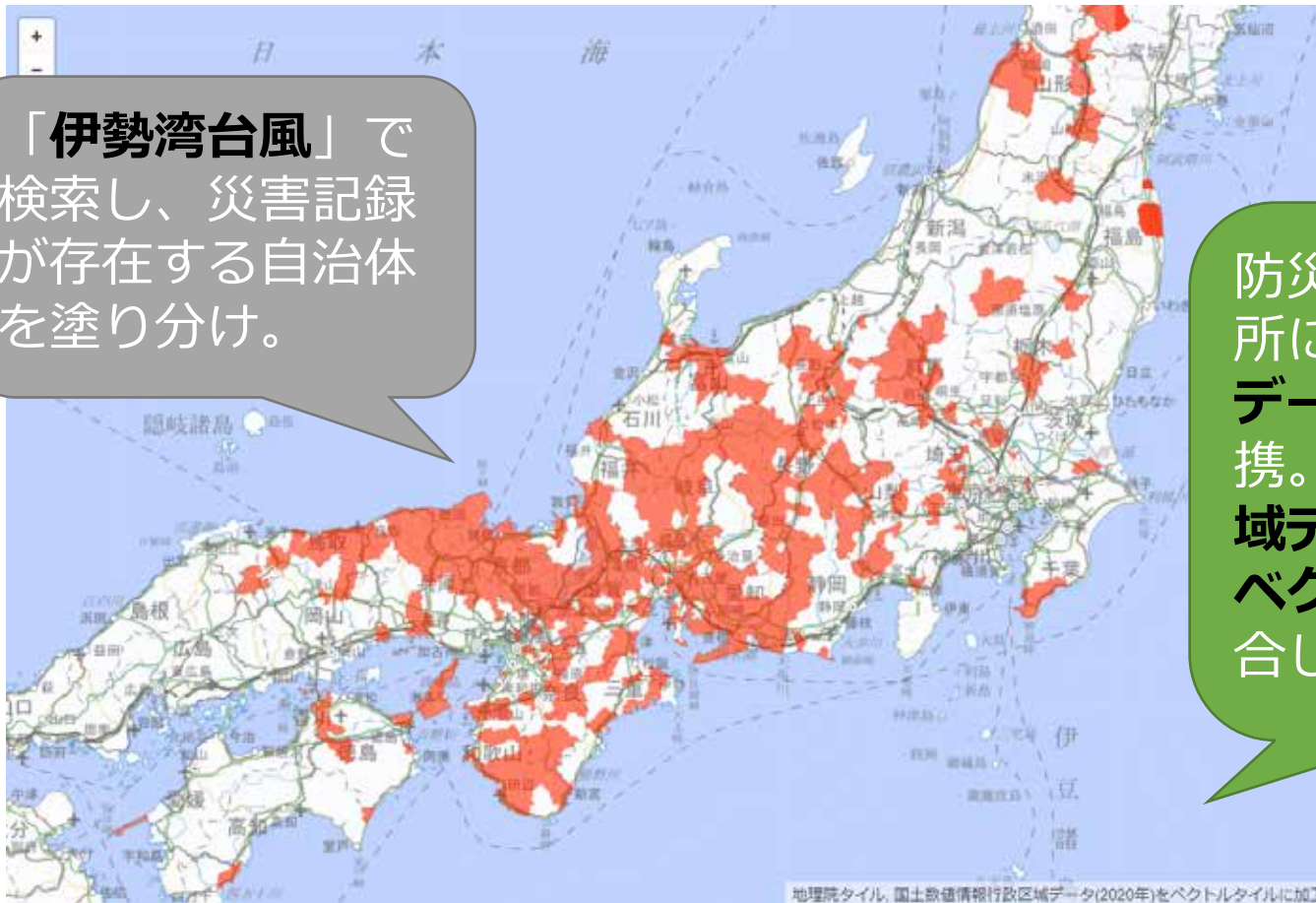


ズームレベル13以降は、町丁・字データセットに切り替え。

デジタル台風：歴史災害データベース

<http://agora.ex.nii.ac.jp/digital-typhoon/disaster/history/>

「伊勢湾台風」で
検索し、災害記録
が存在する自治体
を塗り分け。



防災科学技術研究所による災害事例
データベースと連携。歴史的行政区
域データセット
ベクトル地図を統
合した可視化。

歴史ビッグデータの課題

地災撮要. 巻11-12(地震之部) 51 / 75 翻刻完了

サインイン

閲覧 入力 編集履歴 凡例

【右丁 文字無し】

【左丁】

上州利根郡東小川村 荒山 立候儀
ニ付申上候書付

上州利根郡東小川村 荒山 候者同村人家
ヨリ五里程相隔ニ 両野ノ国 境ニテ東西 野州中禪
寺同国 白根山ト 唱西面者則荒山ニテ郡中第
一之高山候処 当四月上旬ヨリ煙相立折々鳴動致
シ候ニ付右村ハ勿論同山ヨリ流水相成候トモ品川通り
両縁村々者恐怖苦情相唱候趣相聞候間本月七
日捕亡手ノモノ兩人取締旁為見分差遣候処 見分
之始末別紙書付并宛絵 其中村役人

上州利根郡
東小川村

荒山

白根山

史料の記述を
実世界の実体
と連結する

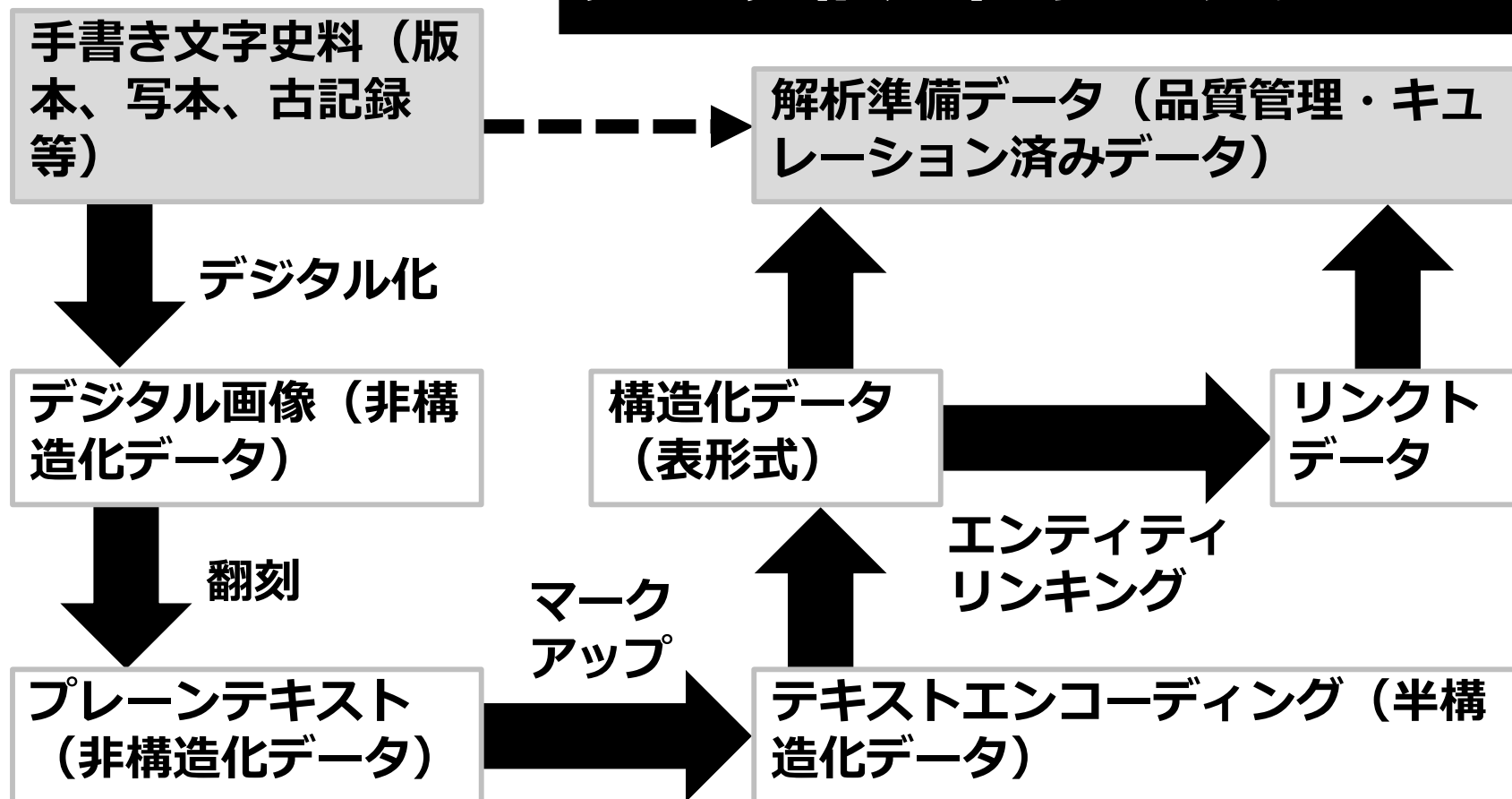


みんなで翻刻：地災撮要. 巻11-12（地震之部）（国立国会図書館蔵）

<https://honkoku.org/app/#/transcription/1B4A9AF6F6E2A2F0D4311060F1CD14F2/51/>

データ構造化ワークフロー

データ構造化ワークフロー



江戸マップβ版

<http://codh.rois.ac.jp/edo-maps/>

国立国会図書館が公開する「江戸切絵図」から地名を抽出してデータベース化。



立命館大学が公開する「日本版 Map Warper」を活用し、古地図を現代の地図に重ね合わせ。

歴史地名マップ

<http://codh.rois.ac.jp/pn-nihu/>

人間文化研究機構
が公開する「歴史
地名データ」の大
部分をマッピング。



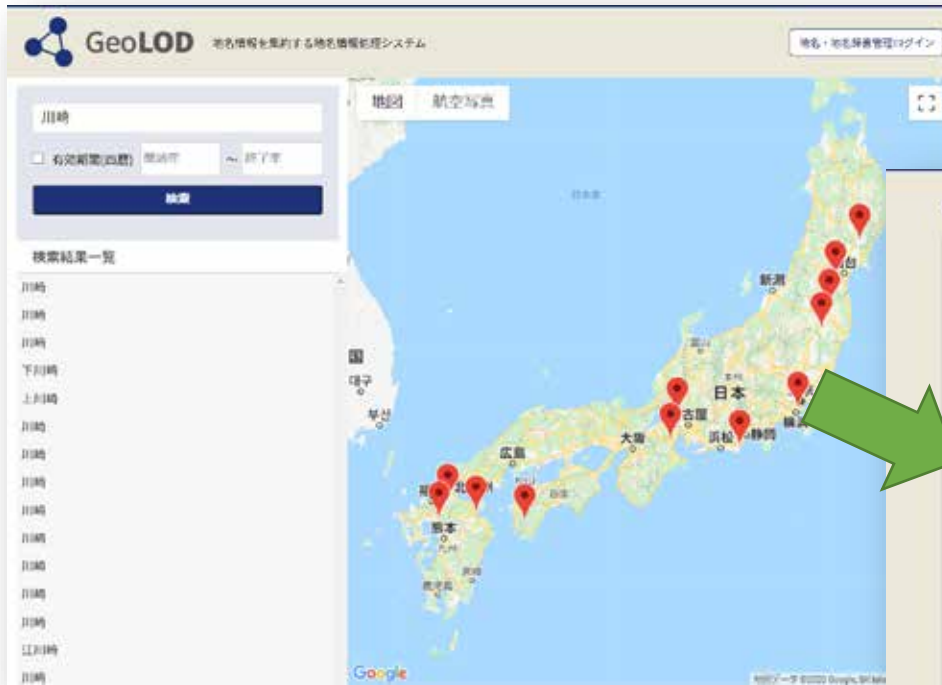
凡例:歴史地名データの属性にしたがい、以下の区分で色分けしている。行政地名、建物、水部、地形、名所旧跡、その他。またマーカーについては、線状の地名については一方の端点のみを表示している。

GeoJSONをベクトル
タイル化すること
で、全国20万件
以上のポイント
データを可視化。

GeoLODによる地名識別子管理

<https://geolod.ex.nii.ac.jp/>

地名に固有の識別子（GeoLOD ID）を付与する。



地名辞書の形式は、地名のエンティティリンキングシステム GeoNLPと相互運用可能。

Name	Property	Value
body		川崎
body_kana		カサキ
address		神奈川県川崎市
geolod_id		Geo5eFM
latitude		35.530806
longitude		139.703012
ne_class		市区町村
prefix		()
suffix		町 ()
valid_from		1889-04-01
valid_to		1924-07-01

辞書	
name	歴史的行政区域データセット[地名辞書]
description	歴史的行政区域データセット[地名辞書]は、1920年から2020年までの国土数値情報「行政区域データ」を対象に、そこに出現する行政地名をリスト化し、独自の識別子を付与して公開しています。
entry_id	1480340012
dictionary_url	http://agoraa.nii.ac.jp/GeoNLP/entry/geoshape-city
landing_url	http://agoraa.nii.ac.jp/GeoNLP/entry/geoshape-city/
code	geoshape-city
created_at	2020-07-28T15:27:26.278338+09:00

長期・歴史的データの課題

1. データフォーマットの変更
2. データ項目・定義の変更
3. データの欠損・紛失
4. 表記ゆれや誤りの訂正
5. 価値＝網羅性＆安定性＆高品質
6. 絶え間ないメンテナンス→表に見えにくい
7. 長期的な運用体制の維持
8. 競争領域ではなく協調領域。運用モデルは？

おわりに

1. 現在を「最新の過去」断面と捉え、歴史的な文脈にデータを位置付ける。
2. 分析・可視化では、データの定義の変化やセンサの経年変化なども考慮する。
3. 網羅性＆安定性＆高品質に向けた、地道な作業が不可欠である。
4. 基盤データのオープン化に対する正当な評価の仕組みが必要である。

• 参考文献

1. 北本 朝展, 村田 健史, "歴史的行政区画データセットβ版をはじめとする地名情報基盤の構築と歴史ビッグデータへの活用", 情報処理学会技術報告, Vol. 2020-CH-124, No. 1, pp. 1-8, 2020年9月

• 研究費

1. 「歴史的行政区画データセットへの現代町丁目地理データ接続とCODHからのWeb公開」, ROIS-DS公募型一般共同研究, 020RP2018
2. 「町丁目地理データ（行政境界データ）のスケラブル可視化スキーム作成と外部利用Web」, ROIS-DS公募型一般共同研究, 035RP2019
3. 「バイナリベクトルタイル化された歴史的境界データを活用した自治体防災WebGISの試み」, ROIS-DS公募型一般共同研究, 008RP2020