

みんなで注釈：近世災害史料の クラウドアノテーションの試み

国立歴史民俗博物館 研究部

橋本 雄太

2023年11月22日（水）14:00-

第1回ヒストリーテック勉強会

自己紹介

- 橋本雄太, 博士 (文学)
- 所属: 国立歴史民俗博物館
- 専門: 人文情報学 (デジタル人文学), 近代西洋科学史
- 京都大学大学院文学研究科 情報・史料学専修出身
- 修士課程修了後, 民間で3年間システムエンジニアとして勤務
- これまでに開発したシステム/ソフトウェア
 - SMART-GS, くずし字学習支援アプリKuLA, みんなで古写真【渋沢栄一伝記資料】, みんなで翻刻
- 言語: Ruby, TypeScript/JavaScript, Python, Java, etc...

みんなで注釈

- <https://ansei2.vercel.app/>
- 歴史災害史料を対象としたクラウドアノテーションシステム（開発中）
- 旧名：みんなでマークアップ
- 主要機能
 1. 災害史料のテキストアノテーション
 2. 外部知識ベースとの紐付け（エンティティリンクング）
 3. 災害の現象・被害記述の可視化, データ出力

みんなで注釈【安政江戸地震史料】

ホーム

資料一覧

- コレクション1
- コレクション2
- コレクション3

可視化

- 被害記録マップ

リンク集

- みんなで翻刻
- 京都市古地震研究会
- 江戸マップβ版
- 地名情報処理基盤 GeoLOD
- 古地図 with MapFan - 江戸
- お問い合わせ

岩田久美加さんが史料テキストにリンク作業を実施しました
6日

江戸御屋鋪方町方共 地震類焼場所巨細書之寫 潰家潰土蔵共 取調附街道筋近在迄 - コマ2
御公儀様右御救小屋場所 幸橋 御門外久保町原 浅草広小路 深川海辺新田 深川八幡宮境内 上野御山下火除地 東叡山御門主様ヨリ御小屋...

岩田久美加さんが史料テキストにリンク作業を実施しました
6日

江戸御屋鋪方町方共 地震類焼場所巨細書之寫 潰家潰土蔵共 取調附街道筋近在迄 - コマ2
御公儀様右御救小屋場所 幸橋 御門外久保町原 浅草広小路 深川海辺新田 深川八幡宮境内 上野御山下火除地 東叡山御門主様ヨリ御小屋...

岩田久美加さんが史料テキストにマーク作業を実施しました
6日

目次

1. 開発の背景・目的・アプローチ
2. テキストアノテーション
3. エンティティリンキング
4. 可視化・データ出力

開発の背景・目的・アプローチ

開発の背景と目的

背景

- 災害史料のテキストDB化の進行
 - 「みんなで翻刻」では近世地震史料500点(600万文字)の翻刻を達成
- 一方で翻刻テキストは自然言語で記述された**非構造化データ**

目的

- 機械処理や定量的分析が容易な**構造化データ**を史料テキストから抽出する方法を開発し、
- クラウドソーシング**を通じて大規模なデータ構造化を実現する
- ただしまずは小規模データから

「古代・中世」地震・噴火史料データベース (β版)
Since Jun.15.2009 034300

このデータベースはβ版であり、まだ修正作業中で、不定期に更新されます。論文等でご利用可な場合は、本文中や文献欄などに右記の最終更新日を明記されることをお願いします。

最終更新日 2020年3月30日(月) 11:33:11*
著者 古代中世地震史料研究会

構成と使い方 閲覧検索画面 更新履歴

この古代・中世の地震と噴火および関連事象に関する記録・典籍に記された記事のすべてについて、検討を加えて、検索の便に供するものです。

期間は、復元上、近世初期の1607年2月(慶長十二年一月)までとしています。このデータベースの表示は、すべて横書きです。

研究費補助金による以下の研究課題の成果に修正を加えたものです。
番号15201040
文化庁・国際標準化データベース構築に関する研究
番号18310124
データベース化と共有型拡張・活用システムの開発

みんなで翻刻【地震史料】
古文書・古記録を解読して、歴史災害研究に参加しよう!

地震史料集テキストデータベース

地震史料集テキストデータベース

検索要引を入力してください (例: 寛政二年大地震)

検索・要引 ヘルプ

データベースの概要

このデータベースは、新潟県・山形県・長野県とその周辺で江戸時代に発生した歴史地震に関する史料および未刊の文献史料について、史料学的検討を加えて、検索の便に供するものです。とした歴史地震は以下のとおりです。

- 1714年信濃小谷地震 (正徳四年三月十五日)
- 1751年越後高田地震 (寛延四年四月二十六日)
- 1762年佐渡地震 (宝暦十二年九月十五日)
- 1802年佐渡小水地震 (享和二年十月十五日)
- 1828年越後三条地震 (文政十一年十一月十二日)
- 1833年出羽庄内沖地震 (天保四年十月二十六日)
- 1858年信濃大町地震 (安政五年三月十日)

※1751年越後高田地震・1802年佐渡小水地震・1828年越後三条地震については、校訂済みの文献史料を掲載しています。
※1751年越後高田地震・1828年越後三条地震については、家屋倒壊率に基づく震度分布図を提示しています。

※このデータベースは、文部科学省の科学技術試験研究委託事業による委託業務「ひずみ集中華の重点的調査観測・研究」(中核機関:独立行政法人防災科学技術研究所 平成20~24年度)のうち、東京大学地震研究所が実施した「古地震・津波等の史資料の収集と解読」(6-1)の成果の一部を公表するものです。

※動作環境について
Internet Explorer (Ver.9) で動作確認を行っています。

データ構造化のフロー

翻刻済みテキスト



重要情報のアノテーション
(日時, 場所, 人物, 被害・現象)

アノテーション済みテキスト



外部知識ベースとの接続 (entity linking)
アノテーション間のリレーション記述



時空間的分析が可能な構造化データ

翻刻テキストに対するアノテーション

去月十五日、十九日之御状相達、致拜見候

(六月)

殿様倍御機嫌能被成御座候、然は去月十五日之暁八時前ち四時過迄地震烈敷、同暁七時頃御構之内御見廻り之處、御蔵脇尺廻り程穴明キ泥砂吹出、当り五六間之内地破れ、追々右同様二も可被成御考、猶又六時過二御雜蔵倒れ候由、尤天保元年ち此度八軽キ風間之旨、猶其後十六、十七日、十八日之極晩迄、昼夜共時々震候処、十九日朝ち平穩二相成、弥以御屋敷向外御別条無之由、委細兩度御注進之趣、夫々申上候、右御報為可申入如此御座候 恐惶謹言

七月三日 塙儀左衛門様

御書方

御側向 連名

去月十六日之御状相達致拜見候、然は同月十五日暁八時頃其表稀成強キ地震有之、翌日迄昼夜数十ケ度小震いたし鎮り兼候よし、尤中之嶋、堂嶋、天満三ヶ所御屋敷差而破損所無之、先御別条無御座候旨、為御注進御紙面之趣致承知候、右御報為可申入如此御座候、以上

七月三日 安井七兵衛様

御側向

1854年伊賀上野地震後に書かれた書状

(加納靖之「1854年伊賀上野地震の際に伏見で発生した局所的な液状化被害地点の検討」, 2018)

翻刻テキストに対するアノテーション

去月十五日、十九日之御状相達、致拜見候

(六月)

殿様倍御機嫌能被成御座候、然は去月十五日之曉
八時前ち四時過迄地震烈敷、同曉七時頃御構之内
御見廻り之處、御蔵脇尺廻り程穴明キ泥砂吹出、
当り五六間之内地破れ、追々右同様二も可被成御
考、猶又六時過二御雑蔵倒れ候由、尤天保元年ち
此度八軽キ風間之旨、猶其後十六、十七日、十八
日之極晩迄、昼夜共時々震候処、十九日朝ち平穩
二相成、弥以御屋敷向外御別条無之由、委細兩度
御注進之趣、夫々申上候、右御報為可申入如此御
座候 恐惶謹言

七月三日

塙儀左衛門様

御書方

御側向

連名

去月十六日之御状相達致拜見候、然は同月十五日
曉八時頃其表稀成強キ地震有之、翌日迄昼夜数十
ケ度小震いたし鎮り兼候よし、尤中之嶋、堂嶋、
天満三ヶ所御屋敷差而破損所無之、先御別条無御
座候旨、為御注進御紙面之趣致承知候、右御報為
可申入如此御座候、以上

七月三日

安井七兵衛様

御側向

1854年伊賀上野地震後に書かれた書状

(加納靖之「1854年伊賀上野地震の際に伏見で発生した局所的な液状化被害地点の検討」, 2018)

地名と日時表現のエンティティリンクング

七月三日 安井七兵衛様 御側向
御座候、以上
御紙面之趣致承知候、右御報為可申入如此
損所無之、先御別条無御座候旨、為御注進
尤中之嶋、堂嶋、天満三ヶ所御屋敷差而破
日迄昼夜数十ヶ度小震いたし鎮り兼候よし、
十五日暁八時頃其表稀成強キ地震有之、翌
去月十六日之御状相達致拝見候、然は同月



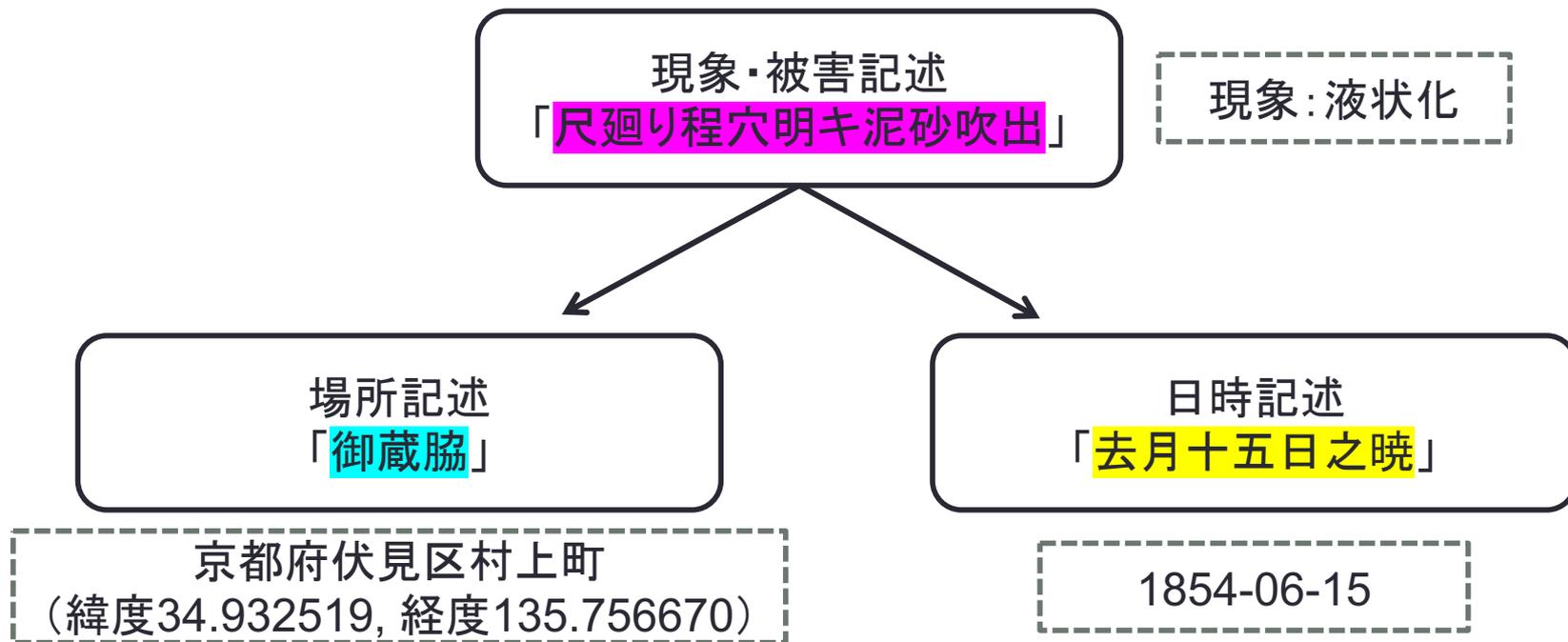
<https://datetime.hutime.org/date/1854-07-10>



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

```
"entryId": "10031585",  
"geolodId": "SqksUA",  
"hypernym": "摂津/大坂",  
"latitude": 34.697777,  
"longitude": 135.509444,  
"neClass": "行政地名/字",
```

アノテーション間のリレーション設定



アノテーション間にリレーションを設定することで、5W1H情報を機械可読形式で記述することができる

作業対象と進捗管理

作業対象

- 安政江戸地震（1855年）に関する記述のある史料約100点
- いずれも「みんなで翻刻」上でテキスト化されたもの

進捗管理の方法

- コマ（画像）単位で進捗を管理
- 全コマについてアノテーションとエンティティリンクングが完了すれば作業完了

The screenshot shows a web interface for managing digitization progress. The header is blue with the text 'みんなで注釈【安政江戸地震史料】'. On the left is a navigation menu with items like 'ホーム', '資料一覧', 'コレクション1-3', '可視化', '被書記録マップ', 'リンク集', and 'みんなで翻刻'. The main content area displays four document cards, each with a thumbnail image of a document and a color calibration chart. Below each thumbnail is the document title, the organization '東京大学地震研究所 Earthquake Research Institute, the University of Tokyo', and two progress indicators: 'マーク完了' (Marked complete) and 'リンク完了' (Linking complete). The progress is shown as a fraction of 'コマ' (pages).

Document Title	Organization	マーク完了 (コマ)	リンク完了 (コマ)
地震年代記	東京大学地震研究所 Earthquake Research Institute, the University of Tokyo	2/32	1/32
諸國地震年代記	東京大学地震研究所 Earthquake Research Institute, the University of Tokyo	1/17	0/17
大江戸 地震後教 人事改正 平かな附	東京大学地震研究所 Earthquake Research Institute, the University of Tokyo	1/8	1/8
江戸 大地震未代断の種 全	東京大学地震研究所 Earthquake Research Institute, the University of Tokyo	0/17	0/17

テキストアノテーション

アノテーション作業のUI

- スタンドオフアノテーションを採用

- アノテーション対象とアノテーション本体を分離する方式
- 低い作業負荷
- アノテーション間のオーバーラップが可能

- アノテーションの方法

- 画面右でアノテーション種別を選択
- テキストをマウスで選択し、アノテーションを付与

江戸御屋鋪方町共 地震類焼場所巨... 編集作業 作業履歴 注釈データ一覧 現象・被害データ一覧 被害状況マップ

ホーム / コレクション2

2 / 16コマ

あなたが編集集中

編集 閲覧 画像 履歴 マップ

保存 破棄

場所 日時 人物 現象・被害

マークを削除

十五 本郷ゆしま谷中王子辺
十六 小石川白山駒込溝邊
十七 小石川御門外込高田辺
十八 神田橋内和田倉門矢八代洲川岸辺
十九 外桜田霞ヶ関永田町辺
二十 芝本芝一田品川迄
廿一 日本橋南新橋まで
廿二 八丁堀つぎち 豊岸橋辺
廿三 南北本所のうち
廿四 深川一田にておぼる

一 日本橋北之方八千五百六十六号
小塚原町両側不残焼る新高原八幡候上
江戸町京町二ヶ所出火五丁不残焼る大門外
五十間高札の側のある田中川谷新馬越町八
家土蔵共町家寺院不残焼る事夥し
久田町二丁目同巻丁自山川丁北馬道金龍山
北谷中谷之寺院不残焼失北新丁(割書・聖天一横丁)より
芝居町三丁やける東か八少ツのころ南馬道より

4種類のアノテーション

1. **日時**…特定の日時や時間の範囲を指定する記述.
2. **場所**…地理的な場所を指定する記述.
3. **現象・被害**…地震とその随伴現象を含む自然現象についての記述. あるいはそれによる人的・物的被害についての記述.
4. **人物**…特定の人物を指し示す記述.

アノテーション作業マニュアル

• <https://wiki.honkoku.org/doku.php?id=annotation>

• Dokuwiki で作成した「みんな
で翻刻ウィキ」上に作成

• 複雑化してきたため、単純な
ルールに更新予定

みんなて翻刻ウィキ

検索

最近の変更 メディアマネージャー サイトマップ

ログイン

annotation

マークアップ作業マニュアル

マークアップ作業では、資料中に現れる「場所」「日時」「人物」「現象・被害」の各表現をマークアップします。それぞれどのような表現をマークアップすべきか、以下で解説します。

場所

特定の地理的場所を指し示す語句をマークアップしてください。これには地名だけでなく、「此辺」「近辺」などの指示的表現や、「品川より千住まで」といった空間的範囲を指す表現も含まれます。

地名の場合は、次のような表現がマークアップの対象になります。

- 地域・・・国、街道筋、郡、街、丁、村、郷、字など
- 地形・・・山、島、川、湖、沼、海、湾など
- 施設・・・城、堀、橋、寺、神社、屋敷、温泉など

注意が必要なケース

- 指示語を伴う場所表現・・・「同所」「近辺」「同町」など、場所を指示表現で示すものについても、場所表現としてマークアップしてください。
- 人名が含まれる場所表現・・・大名屋敷など、場所名に人物名・役職名が含まれる場合、その人名・役職名は「人物」としてマークアップしてください。例：「松平美濃守様御やしき」「黒田豊前守様御やしき」のような表現はそれぞれ、「松平美濃守様」「黒田豊前守様」を人物としてマークアップし、「松平美濃守様御やしき」「黒田豊前守様御やしき」全体を「場所」としてマークアップしてください。「～やしき」「～宅」と明示せず、人物の邸宅の意味で個人名を列挙する場合もあります。例：「江戸芝居町焼失の中にて残り候分役者宅は坂東しうか伴[吉弥][中村福助][市村羽左衛門][森田助弥][尾上菊次郎]其外茶屋二三げん残る」。このような場合は、各人名を「人物」と「場所」両方でマークアップしてください。
- 階層を伴う場所表現・・・「〇〇郡△△村□□家」のように階層順に地名が表記されている場合、地名を分割せず、一括してマークアップしてください。
- 範囲を示す場所表現について・・・「本所より深川まで」「品川から千住宿まで」のように、範囲によって場所を指定している場合は、個別の地名ではなく、表現全体をマークアップして

@CloudHonkokuさんのツイート

みんなて翻刻さんがリポストしました

OK @. 11月21日

「常陸国郷帳」を647文

エンティティリンキング

日時アノテーションのエンティティリンクング

- 元号/年/月/日/時刻をフォームに入力
 - 自動入力にも対応
- HuTime 暦変換サービスを利用し，グレゴリオ暦で表示
- リンクを作成すると，アノテーションに「」マークが表示される



The screenshot shows a web interface for linking date and time information. At the top, there is a blue header with navigation links: 編集作業, 作業履歴, 注釈データ一覧, 現象・被害データ一覧, 被害状況マップ. Below the header, there is a breadcrumb trail: マップ > 1 / 16コマ > あなたが編集集中. On the right side of the header, there are two buttons: 保存 (Save) and 破棄 (Discard).

The main content area is titled "日時情報をリンク" (Link date and time information). Below the title, there is a text box with the instruction: "マーク箇所に記載されている日時を和暦で入力してください。年/月/日は半角数字で入力する必要があります。" (Please enter the date and time recorded in the marked area in Japanese calendar. Year/month/day must be entered in half-width numbers.)

Below the instruction, there are two buttons: リンク作成/更新 (Link creation/update) and リンク解除 (Link removal). The form consists of several input fields:

- 元号 (Era name): 安政 (Ansei)
- 年 (年角数字) (Year, half-width number): 2
- 月 (月角数字) (Month, half-width number): 10
- 日 (日角数字) (Day, half-width number): 2
- 時刻 (時刻表現を原文の通り入力してください。) (Time, enter the original expression as is.): 夜四ツ時 (Night 4 o'clock)

At the bottom of the form, there is a button labeled 自動入力 (Automatic input). Below the form, there is a text box showing the result: HuTime暦変換サービスによる変換結果：1855-11-01.

On the left side of the form, there is a vertical sidebar with text: 地震類焼場所巨細書之写 (Copy of the Great Earthquake Burned Places Book) and 潰家潰土蔵共取調附御道筋近在迄 (Including the collapsed house and collapsed earthenware storehouse, including the investigation of the main street and vicinity).

場所アノテーションのエンティティリンクング

- GeoLOD API 経由で地名を検索し，対応する地名を選択する
 - 地名は複数選択可能
- 検索結果は[国土地理院タイル](#)上で確認可能
 - マップ表示には Leaflet を使用



The screenshot displays the GeoLOD API interface. At the top, there's a navigation bar with '履歴' (History), 'マップ' (Map), and 'あなたが編集集中' (You are editing). Below this, a sidebar lists various locations like '御公儀様方御救小屋場所' and '浅草広小路'. The main area is titled '場所情報をリンク' (Link location information) and contains a search box with the keyword '千住宿' (Senju-juku). Below the search box, there are buttons for 'リンク作成/更新' (Create/Update link) and 'リンク解除' (Remove link). A checkbox option is present: '検索結果を東京（島嶼部除く）近郊に限定' (Limit search results to Tokyo (excluding islands) suburbs). The map shows the Senju area with a green pin at '千住' and another at '東京藝術大' (Tokyo University of Arts). A table at the bottom lists search results with columns for GeoLOD ID, Name, Source, and Delete.

GeoLOD ID	地名	ソース	削除
Z6UWcE	千住町	歴史地名データ辞書	
aowS0a	千住	歴史地名データ辞書	

現象・被害アノテーションのエンティティリンキング

- 外部知識ベースの代わりに、現象・被害情報を作業者が構造化して記述する
 - 現象・被害の種別
 - 現象・被害の規模
 - 被害の対象
- リレーション設定
 - 近傍の日時・場所アノテーションを指定することで、現象・被害の発生場所日時を表現

画面 履歴 マップ 2 / 16コマ あなたが編集集中

保存 破棄

リンク作成/更新 リンク解除

現象の種類

地震動 地割れ 津波 液化化 洪水

地すべり・地崩れ

被害の種類

建物損壊 火災 浸水 死者 けが人

行方不明者 被害なし

現象・被害の規模

記述なし 少々 半分 大いに 残らず

上記以外の記述

被害の対象

一般家屋 寺社 屋敷・大店 土蔵 石垣

城郭 田畑

日時・場所

場所・日時を関連付け

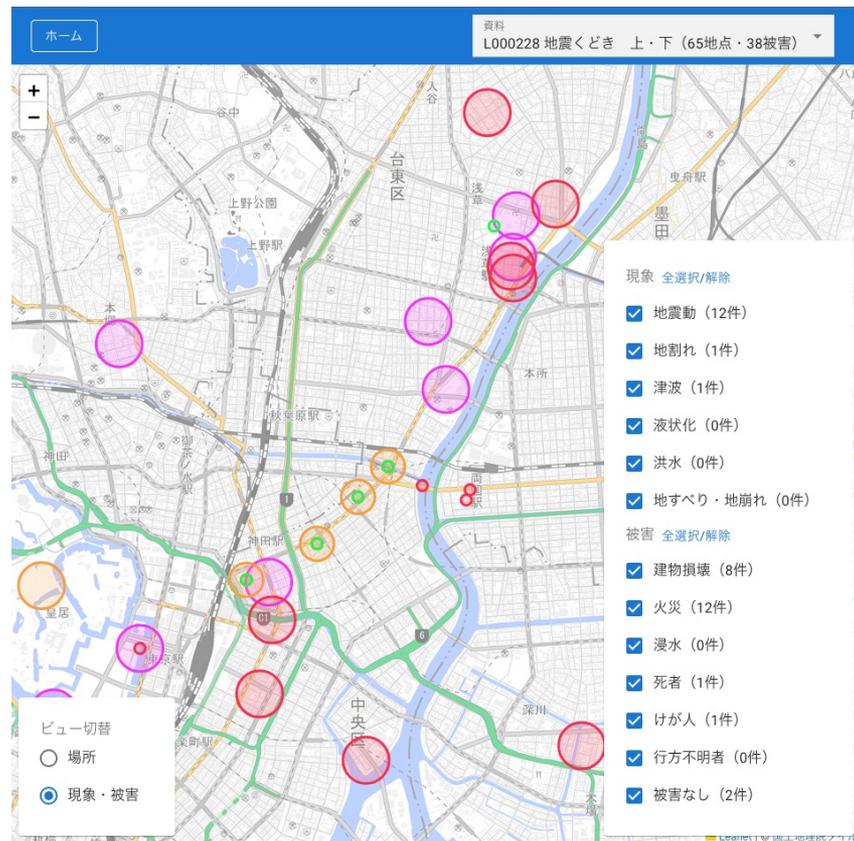
この現象・被害が発生した日時・場所を示すマークを選択します。

一 日本橋と北の方八千住宿大半くつれ
小塚原町両側不残焼る新吉原八幡候上
江戸町京町二ヶ所出火五丁不残焼る大門外
五十間高札之倒のころ 田中山谷新鳥越町八
家土蔵共町家寺院不残焼る事夥し
久田町二丁目同巻丁目山川北馬道金龍山
北谷中谷之寺院不残焼失北新丁(割書・聖天(橋丁)ち
芝居町三丁やける東か八少ツくのころ南馬道より

可視化・データ出力

現象・被害記述のマップ表示

- 現象・被害が発生した地点・種別・規模をマップ上に表示
- リレーションを設定された場所アノテーション経由で緯度経度を取得している
- 複数地点が登録されている場合は重心を表示



現象・被害データのテーブル表示

- 現象・被害マップを
テーブル形式で表示した
もの
- JSONやTSV/XLSX形
式での出力にも対応予
定

地震出火 男女死人怪我人書上之寫						
ホーム / コレクション1						
編集作業 作業履歴 注釈データ一覧 現象・被害データ一覧 被害状況マップ						
場所テキスト	現象・被害テキスト	現象・被害種別	規模	推定緯度	推定経度	関連付けられたGeoLodアイテム
本所	変死	死者	0	35.702204266666...	139.80286713333...	本所 (GOoNc0),北本所荒井町 (3lcIGP),北本所表町 (6q...
吉原町	右同断	けが人	0	35.723559333333...	139.79444033333...	惣名新吉原町 (3aWQam),新吉原 (AoYG2V),吉原 (HCo2...
吉原町	右同断	けが人	0	35.723559333333...	139.79444033333...	惣名新吉原町 (3aWQam),新吉原 (AoYG2V),吉原 (HCo2...
吉原町	右同断	けが人	0	35.723559333333...	139.79444033333...	惣名新吉原町 (3aWQam),新吉原 (AoYG2V),吉原 (HCo2...
吉原町	右同断	けが人	0	35.723559333333...	139.79444033333...	惣名新吉原町 (3aWQam),新吉原 (AoYG2V),吉原 (HCo2...
渋谷	変死	死者	0	35.654624777777...	139.70821755555...	渋谷広尾町 (py2OCH),渋谷町 (P86eQ9),渋谷 (O6QOEC...
目黒	変死	死者	0	35.635247933333...	139.70593546666...	目黒村 (gQaah),下目黒町 (g24sU3),中目黒村 (X4swiJ)
久保町	怪我	けが人	0	35.668002666666...	139.75256466666...	久保町 (UwUsWi),○久保町原ト云 (gooOoc),久保町 (z...
浅草	変死	死者	0	35.707302	139.789017	浅草 (Gu2SK3),
浅草	変死	死者	0	35.707302	139.789017	浅草 (Gu2SK3),
浅草	変死	死者	0	35.707302	139.789017	浅草 (Gu2SK3),
浅草	変死	死者	0	35.707302	139.789017	浅草 (Gu2SK3),
久保町	怪我	けが人	0	35.668002666666...	139.75256466666...	久保町 (UwUsWi),○久保町原ト云 (gooOoc),久保町 (z...
浅草	変死	死者	0	35.707302	139.789017	浅草 (Gu2SK3),
目黒	変死	死者	0	35.635247933333...	139.70593546666...	目黒村 (gQaah),下目黒町 (g24sU3),中目黒村 (X4swiJ)

1 row selected

Rows per page: 100 1-53 of 53

構築技術

- TypeScript
 - JavaScript を拡張した静的型付言語
- Material UI
 - Material Design に準拠した UI フレームワーク
- Next.js
 - React ベースの Web アプリケーションフレームワーク
 - サーバーサイドレンダリングに対応
- Firebase
 - Google が運営する Backend as a Service
 - 認証やDB, ホスティング, Cloud Function 機能などを提供
- Vercel
 - Next.js 開発企業によるホスティングサービス

課題

- クラウドソーシングの難しさ
 - 翻刻と異なりアノテーションの「正解」を決めづらい
 - 作業には歴史資料の読解や歴史地名特定の能力が要求される
 - 特に安政江戸地震の場合は江戸の地理についての知識が重要
- 当面は史学分野のトレーニングを受けた専門家に作業を依頼

- 機械学習を適用する可能性
 - 現時点では作業を大きく省力化するほどの精度は難しいとの認識
- プロジェクトを通じてAI訓練用の教師データを提供してゆく

おわりに：今後の展望

- 災害史料の構造化プロジェクトとして
 - 安政江戸地震史料の大規模構造化データ化
 - 地震史料集テキストデータベース等，他DBとの連携
 - 任意のテキストのアノテーションへの対応
- 史料テキストのアノテーションプラットフォームとして
 - 数値などアノテーション種別の拡張
 - カスタムラベルによるアノテーションのサポート
 - 横書きUI のサポート

宣伝：ベータ版テスター募集中

• じせデジリーダー

- 国立国会図書館デジタルコレクションのインターネット公開資料を対象とした参加型モバイルリーダー

• 主要機能

- モバイルに最適化された資料リーダー
- 全文検索, OCRテキスト表示
- ユーザー独自の資料リスト, 読書ノ-資料レビューの作成・共有
- コメント, いいね等のユーザー間交流
- プッシュ通知

