

# 歴史資料の市民参加型 翻刻プロジェクト「みんなで翻刻」

---

国立歴史民俗博物館  
研究部 橋本 雄太

2024年9月7日  
日本学術会議 近畿地区会議 学術講演会

# 目次

1. 「みんなで翻刻」の概要
2. プロジェクトの成果
3. オンラインシチズンサイエンスの課題

# 「みんなで翻刻」の概要

---

# みんなで翻刻

- 概要

- 歴史資料の市民参加型翻刻プラットフォーム
- 翻刻・・・歴史学用語で史料を活字に起こすこと
- 京都大学古地震研究会を中心に開発・運営
- <https://honkoku.org/>
- デジタルアーカイブアーカイブジャパン（DAJ）アワード2024受賞🏆

- バージョン

- Ver. 1・・・2017年公開. 前近代の地震史料を対象.
- Ver. 2・・・2019年公開. 歴史資料一般を対象.

# 京都大学古地震研究会



- 地震研究や防災研究への応用を目的として古文書を解読
- 地震学，気象学，日本史，地理学，科学史，情報学，図書館司書など多分野の参加

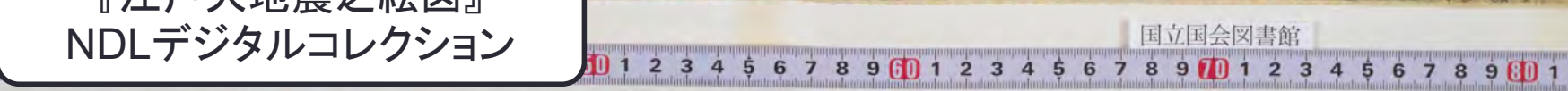


- 2012年の活動開始当初から活発な情報技術活用の試み
- 橋本はD2の2014年から参加



[illegible][illegible][illegible][illegible]

一、土產穀類、穀類、計十萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 二、土產、油、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 三、土產、鹽、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 四、土產、糖、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 五、土產、茶、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 六、土產、布、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 七、土產、紙、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 八、土產、藥、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 九、土產、其他、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）  
 十、土產、其他、計一萬三千九百九十五石、  
（此數、係以舊曆、爲標準、而計算、之數、也、）



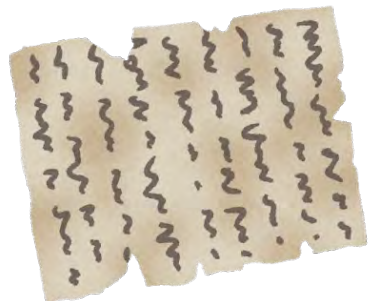
國立國會圖書館



# 「みんなで翻刻」がめざすもの

江戸時代以前の地震・災害史料  
(東京大学地震研究所所蔵)

テキストデータ  
(CC BY-SAで公開)



翻刻



活用

活用



地震・防災研究



オンラインで参加するユーザー




機械学習ベースの  
文字認識研究

# さまざまな翻刻プロジェクト


[ホーム](#) [プロジェクト](#) [リンク集](#) [ご案内](#) [使い方](#)

🔍 翻刻文を検索... [A](#) [文](#) [Twitter](#) [サインイン](#)




**日本の仏典を翻刻**  
SAT大蔵経テキストデータベース研究会による、日本の仏典の翻刻プロジェクト  
所蔵機関：SAT大蔵経テキストデータベース研究会等

56/96点の翻刻が完了




**Code4Lib JAPAN × みんなで翻刻**  
Code4Lib JAPAN 2020カンファレンスで登録した資料です。  
所蔵機関：国立国会図書館など

26/37点の翻刻が完了




**デジタルアーカイブ福井の資料を翻刻**  
デジタルアーカイブ福井で公開している地域資料を翻刻するプロジェクトです。  
所蔵機関：福井県文書館

22/37点の翻刻が完了




**疫病関連資料を翻刻！**  
近世の疫病に関連する資料を翻刻するプロジェクトです。  
所蔵機関：京都大学、国立歴史民俗博物館、他、各地のデジタルアーカイブ

25/52点の翻刻が完了




**東京学芸大学「学びと遊びの歴史」を翻刻！**  
東京学芸大学が所蔵する近世以降の学びや遊びの資料翻刻プロジェクトです。  
所蔵機関：東京学芸大学附属図書館

6/10点の翻刻が完了




**Gallicaの日本資料を翻刻！**  
フランス国立図書館（Bibliothèque nationale de France, BnF）のデジタルライブラリーGallicaで公開されている日本資料の翻刻プロジェクトです。  
所蔵機関：フランス国立図書館（BnF）

76/296点の翻刻が完了




**あいおいニッセイ同和損保所蔵災害資料を翻刻する**  
あいおいニッセイ同和損保所蔵災害資料を翻刻するプロジェクトです。京都文化博物館での「伝える-災害の記憶展」での出品資料を中心に翻刻します。  
所蔵機関：あいおいニッセイ同和損害保険株式会社

53/117点の翻刻が完了




**翻刻！東寺百合文書**  
京都学・歴史館が所蔵するユネスコ世界記憶遺産「東寺百合文書」の翻刻プロジェクトです。  
所蔵機関：京都府京都学・歴史館

73/210点の翻刻が完了



**茨城大学図書館所蔵資料を翻刻**  
茨城大学図書館が所蔵する近世資料の翻刻プロジェクト  
所蔵機関：茨城大学図書館

15/15点の翻刻が完了



**翻刻！江戸の医療と養生**  
江戸時代、一般の人たちの間に普及していた医療の知識、セルフケア、健康法などの関連資料を翻刻するプロジェクトです。  
所蔵機関：京都大学、国会図書館、他

8/17点の翻刻が完了

- ・ 地震史料の翻刻プロジェクトを含む23件の翻刻プロジェクトが現在進行中
- ・ 連携先…福井県文書館、東京学芸大、茨城大、京都府歴史館、琉球大、関西大、歴博など
- ・ IIIF (International Image Interoperability Framework) 経由で資料画像をインポート





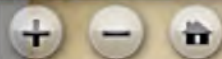
閲覧

入力

ノート

編集履歴

天明二壬寅の今年以前三十餘年の



天明二壬寅の今年以前三十餘年の  
あゝ自由を思ふ事なかりて  
徳川の世より今に至るまで  
里原村火災の御供書焼失  
御供書焼失せりい時後  
去年の壬寅の年我生年  
りて山崎郡の山村に  
い道ありて山崎郡  
巡



# 歴史資料を対象とする シチズンサイエンスの課題

## 1. 規模の問題

- 数百～数千人規模の市民の参加が必要

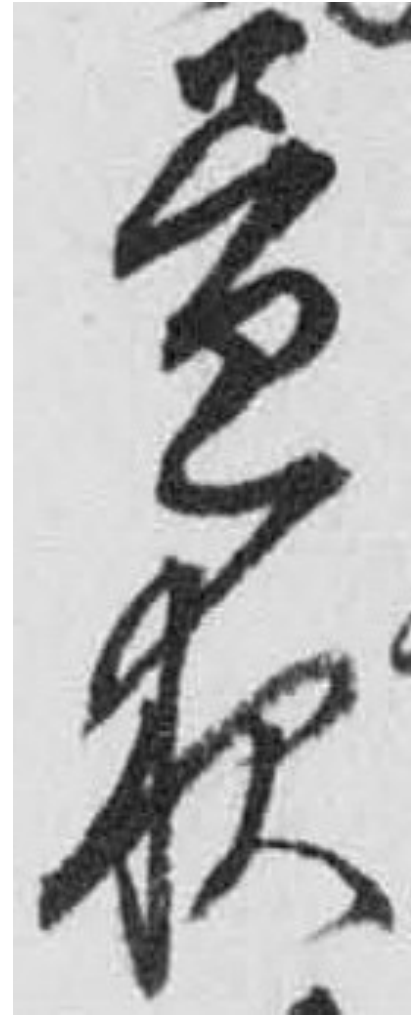
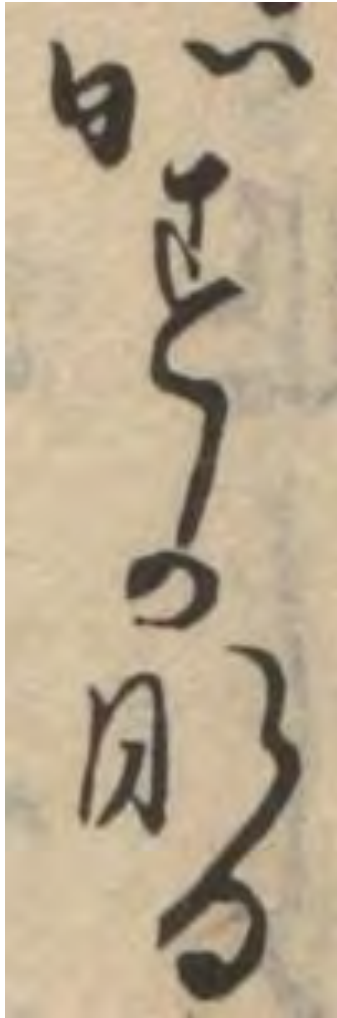
## 2. 動機付けの問題

- 善意に頼るだけでは長期的参加は見込めない

## 3. 能力の問題

- くずし字の正確な翻刻には相当の訓練が必要

# なんと読むでしょう



# 『みんなで翻刻』のアプローチ

- **くずし字解読の学習サービス**としてデザイン
  1. くずし字解読の学習コンテンツ提供
  2. 参加者間のコラボレーションによる「学び合い」
  3. AIによる初学者の解読支援
- **くずし字学習の延長**として自然に翻刻作業に参加してもらうことが目的
  - ・「やりがい搾取」にならないように
  - ・参加者が明確なメリットを得る形で参加してもらう



# ①くずし字学習アプリとの連携

- くずし字学習支援アプリKuLA
  - 阪大文学研究科を中心に開発されたくずし字の学習支援アプリケーション
  - 2016年2月の公開後25万回DL
- 「みんなで翻刻」との連携
  - KuLA の全コンテンツをサイトに転載
  - アプリで基礎学習→みんなで翻刻で発展学習を行うフローを整備した



## ②翻刻文の相互添削

差分表示画面



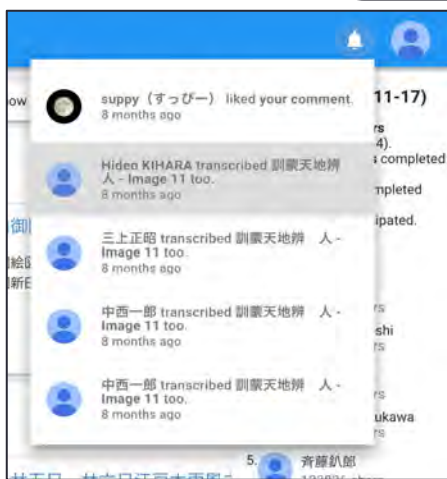
翻刻文エディタ



翻刻

フィード  
バック

通知パネル



共有

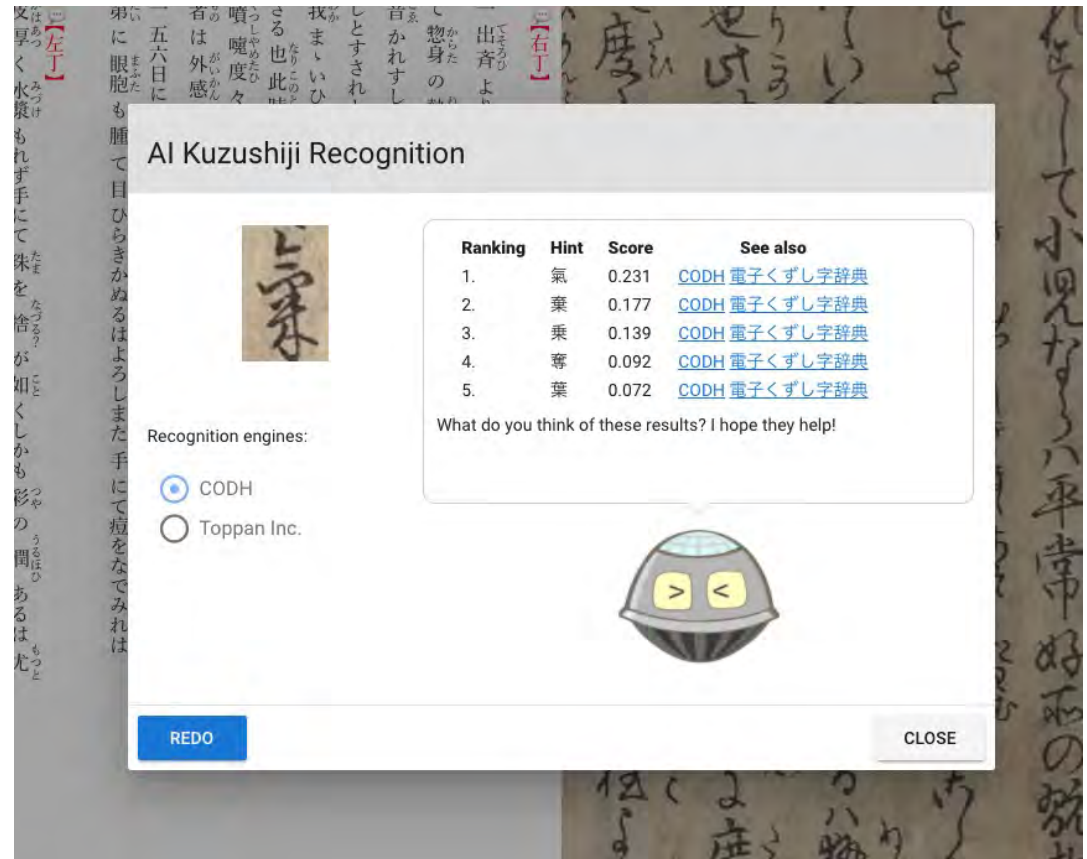
タイムライン



添削

# ③くずし字認識AIによるサポート

- くずし字の読み方の候補をスコア付きで提示する文字認識 AI を提供
- 2種類のAIが利用可能
  - Center of the Open Data for the Humanities (CODH)
  - 凸版印刷
- AIは補助輪として参加者をサポートする



# プロジェクトの成果

---



# バージョン1の成果 (2017年1月～2019年3月)

- 翻刻が完了した史料 500/508点
- 翻刻が完了した画像 8,411/8,925枚 (94%)
- 参加登録者数 5,346人
- 総入力文字数 625万字

東京大学地震研図書室の  
所蔵和古書499点の全文翻刻を達成

# バージョン2の成果 (2019年7月～)

- 翻刻が完了した史料 1,800/3,469点
- 翻刻が完了した画像 66,512/138,995枚
- 参加登録者数 3,426人
- 総入力文字数 3,343万字

翻刻対象を歴史資料一般に拡大  
現在も毎日2万字のペースで進行中

# 翻刻文の正確性

- 翻刻文10万文字を検証した結果：

表 3.2 検証作業を通じて見つかった要修正箇所の数（資料種別・タイプ別）

資料種別	合計文字数	誤読箇所		未読箇所		表記ゆれ箇所		合計	
		実数	比率 (%)	実数	比率 (%)	実数	比率 (%)	実数	比率 (%)
木版本	78,774	400	0.5	285	0.4	395	0.5	1,080	1.3
筆写本	21,901	210	1.0	105	0.5	132	0.6	447	2.0
合計	100,675	610	0.6	390	0.4	527	0.5	1,527	1.5

- 木版本で98.7%，筆写本で98.0%の正確性
- 学術出版される史料集の品質には及ばないものの、  
内容把握や全文検索には十分な品質が得られている

# オンライン シチズンサイエンスの課題

---



# システム開発の問題

- シチズンサイエンスに用いるシステムの開発を外注することは非常に困難
- 理由
  - コモデティ化した業務システムと大きく異なる仕様
  - 良質なユーザー体験を提供する必要性
  - 仮説検証を繰り返す機動的・継続的开发の必要性
- 解決策：内製開発
  - しかしシステム内製が可能な日本の学術機関は稀…

# Zooniverse のスタッフ

## ADLER PLANETARIUM



### Cliff Johnson, Zooniverse Co-Director & Science Lead

Cliff Johnson is a jointly-appointed astronomer at Northwestern University and the Adler Planetarium. As a research, he is a longtime Zooniverse collaborator as science lead for the Andromeda Project and the Local Group Cluster Search. Cliff joined the Zooniverse team at the Adler in May 2019.



### Cory Chambers, Mobile Developer

Cory is a Mobile App Developer. In his free time he likes to run at the park, work in the yard, read books, travel to new and exciting places, and spend time with his family.



### Delilah Clement, Senior Developer & Frontend Lead

Delilah is a web developer for the Zooniverse team at Adler Planetarium. She joined the team in 2021 with a background in aquatic biology. You can find her playing ice hockey or longboarding on Chicago's lakefront.



### Laura Trouille, Zooniverse PI

Laura is VP of Science Engagement at the Adler Planetarium, was co-PI for Zooniverse from 2015-2023, and since September 2023 serves as PI of the Zooniverse. While earning her Ph.D. in 2010 studying galaxy evolution, she embodied cosmic collisions as a roller derby queen aptly named "The Big Bang".



### Mark Bouslog, Developer

Mark is a front-end web developer for the Zooniverse team at the Adler Planetarium. Joining in November 2015 from a career in accounting, he's thrilled to focus his number crunching to 1's and 0's and is constantly inspired and humbled by the power of programming, citizen science and coffee.



### Meredith Stepien, Manager of Creative Content and Production

Meredith is based at the Adler Planetarium, and works with the Adler and Zooniverse on creating silly songs and videos that make difficult science concepts fun! She loves her dog and being cozy and falling asleep by 8pm.



### Michael Zevin, Astronomer

Mike is an astronomer at the Adler Planetarium specializing in gravitational waves, compact objects, and stellar evolution. He is one of the lead researchers of the Zooniverse project Gravity Spy. When space isn't on his mind, he can often be found playing music around the Windy City.



### Michelle Yuen, Developer

Michelle is a software developer for the Zooniverse team at Adler Planetarium. Born and raised in Chicago, the Adler was one of her favorite places in the city (still is!). In her free time, you can find her on the tennis court, eating tacos, kickboxing, cuddling with her cat, or sharing her passion for math, logic problems and coding to anyone willing to listen.



### Samantha Blickhan, Zooniverse Co-Director & Humanities Lead

Samantha Blickhan is based at the Adler Planetarium, leading humanities research and development efforts for the Zooniverse. Outside of work she's most often making music or fiber art projects, or hanging out with her dog.



### Sean Miller, Designer

Sean is a UI and UX designer for the Zooniverse. He also assists the Adler with design projects having to do with technical aspects such as 3d printing. His background as a maker contributes to unique perspectives and solutions to problems. In his free time he likes playing baseball or jamming with friends.



### Zach Wolfenbarger, Developer

Zach is a software developer. He was a molecular biologist, but then the lab needed some code to be written and the die was cast. He's also in a couple of bands and can be found playing shows at bars and comic conventions all over the midwest.

## OXFORD



### Alisa Apreleva, Project Liaison and Researcher

Alisa had been a forensic linguist, a journalist, a homeschooling mom and a music therapist before joining Zooniverse in 2024. During her doctoral studies Alisa helped people with MND / ALS preserve their voice through singing and playing melodica. After all these years, Alisa is still fascinated by humans.



### Chris Lintott, Senior Scientist

Astronomer and co-founder of both Galaxy Zoo and the Zooniverse that grew from it, Chris is interested in how galaxies form and evolve, how citizen science can change the world, and whether the Chicago Fire can get their act together.



### Oluwatoyosi Oyegoke, Software Developer

Oluwatoyosi is a software developer with four years of experience and a BSc(ed) in Economics. Outside of work, he enjoys gaming and is dedicated to teaching kids how to code.



### Shaun A. Noordn, Web Developer

Shaun was raised by Nintendo consoles and somehow transformed his love for video games into a love for creating interactive experiences. When not at his PC playing games or reading comics, he's at his PC studying web design and coding experimental apps.

32名の関係者のうち11名がソフトウェア開発者/デザイナー

# おわりに

- みんなで翻刻
  - 前近代の歴史資料を対象としたシチズンサイエンスプロジェクト
  - 7年間で約8,000人の参加者により2,300点（4,000万字）の資料を全文翻刻
- シチズンサイエンスの普及に向けた課題
  - 外注によるシステム開発の難しさ