

「みんなで翻刻」にみる 歴史地震研究への 非専門家の参加

加納 靖之

東京大学 地震研究所

東京大学 地震火山史料連携研究機構

みんなで翻刻

橋本雄太（国立歴史民俗博物館）・中西一郎（京都大学名誉教授）

京都大学古地震研究会メンバー・みんなで翻刻の参加者

第1回 SPARC Japan セミナー2019

「人文社会系分野におけるオープンサイエンス ～実践に向けて～」

きょうのお話

- 私の身の回りのオープンサイエンス的な取り組み
 - 『京都の災害をめぐる』 / 古典オーロラハンター
 - 地震学：満点計画 / 歴史的記録 / オープンデータ
 - みんなで翻刻
- 非専門家が参画することの意義
 - マンパワー / 学び / 他には？
 - 専門家側の目論見と実際
 - 何をもって成功とするか？

古典オーロラハンター

[illegible]

満点計画

近況報告

鳥取の地震計設置に阿武山観測所のボランティアサポーターが参加しました。

2015年3月22日～26日にかけて、鳥取県での地震計設置に阿武山観測所のボランティアサポーター6人が参加しました。

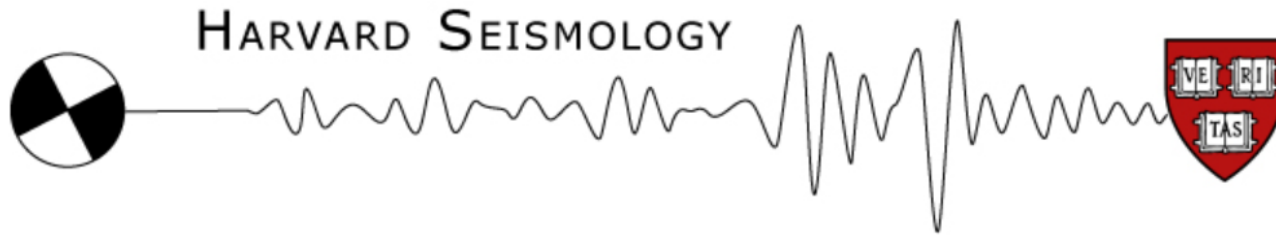
これは、以前に鳥取県で行った地震計の設置点調査にもとづき、満点地震計を50点設置するという作業でした。この作業を阿武山観測所のサポーターたちも担い、実際に何点かの地震計を設置しました。

- 専門家とともに地震観測に参画
- 計器もプロ仕様
- 解析にも関わるように



サポーターによる地震計設置の様子

地震計記録のデジタル化プロジェクト (ハーバード大学)

[HOME](#)[PEOPLE](#)[RESEARCH PROJECTS](#)[HRV](#)[RESOURCES](#)[CONTACT US](#)

地震計記録のデジタル化プロジェクト

[English Version\(英語\)](#)

[石井水晶](#)

[協力校専用ページへ \(パスワード認証\)](#)

地震計で記録された最初の地震は1889年4月17日にドイツで観測された日本の地震です。それ以降、1980年後半まで地震計の記録は紙などのアナログ媒体で残されています。この100年以上におよぶ記録は地震、火山、地球温暖化、等様々な研究にとって非常に貴重なデータであるにも関わらず、アナログのままでは解析できません。その上、紙などの劣化も進んでいます。保管場所の問題などもあり、このままでは100年以上の記録が失われてしまう可能性があります。

膨大なデータのデジタル化に向けて2018年7月より[学校改革フォーラム](#)（代表：盛永俊弘）との共同プロジェクトを立ち上げました。日本の高校生にハーバード大学で開発されたDigitSeisソフトウェアを使ってデジタル化を手伝ってもらう試みです。このプロジェクトを通して、地震学や自然科学をより身近に感じてもらえるよう解説や交流なども行なっています。

<http://www.seismology.harvard.edu/research/DigitSeisJapan/index.html>

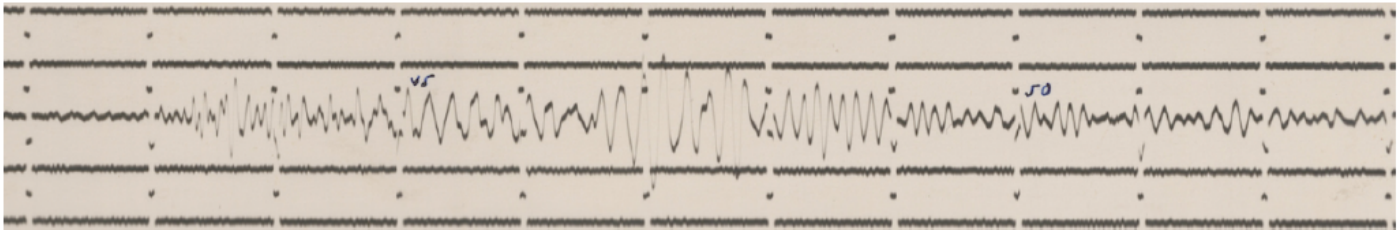
地震計記録のデジタル化プロジェクト (ハーバード大学)



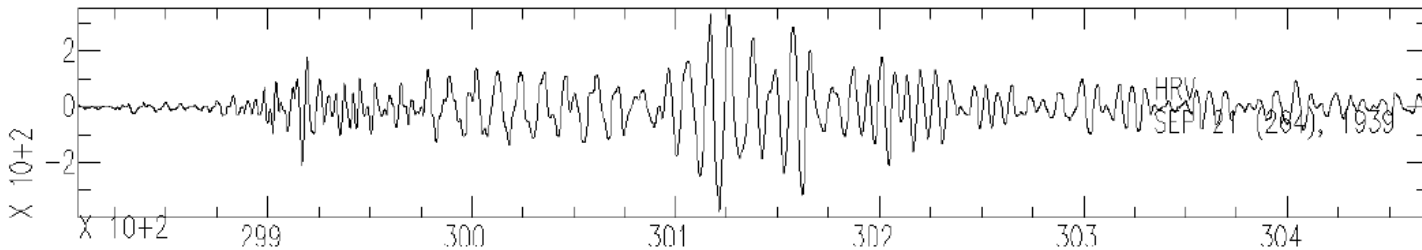
DigitSeis

- 画像を時系列に変換するソフトウェア

画像



研究材料となる時系列



ソフトはインターネットで公開中

<http://seismology.harvard.edu/research/DigitSeis.html>

地震計記録のデジタル化プロジェクト (ハーバード大学)

February

専門家がデータとして使える (SACファイル)

Date	Instrument	Components
----------------------	----------------------------	----------------------------

March

Date	Instrument	Components		
03/02	SP	Z (1.41 GB) Info JPG (35 MB)	NS (1.4 GB) Info JPG (34 MB)	EW (1.4 GB) Info JPG (33 MB)
	LP	Z&EW (1.4 GB) Info JPG (35 MB)	NS (1.39 GB) Info JPG (33 MB)	Z&EW (1.4 GB) Info JPG (35 MB)
03/08 - 03/09	SP		NS (1.43 GB) Info JPG (45 MB)	EW (1.43 GB) Info JPG (44 MB)
	LP	Z&EW (1.46 GB) Info JPG (53 MB)	NS (1.42 GB) Info JPG (43 MB)	Z&EW (1.46 GB) Info JPG (53 MB)
03/09 - 03/10	SP	Z (1.48 GB) Info JPG (55 MB)	NS (1.45 GB) Info JPG (51 MB)	EW (1.47 GB) Info JPG (50 MB)
	LP	Z&EW (1.48 GB) Info JPG (55 MB)		Z&EW (1.48 GB) Info JPG (55 MB)
03/09 - 03/11	LP		NS (1.46 GB) Info JPG (51 MB)	
03/13 - 03/14	SP	Z (1.45 GB) Info JPG (49 MB)	NS (1.44 GB) Info JPG (48 MB)	EW (1.44 GB) Info JPG (47 MB)
	LP	Z&EW (1.43 GB) Info JPG (47 MB) SAC(Z) (34 MB) analysis by Manato Sano of Ritsumeikan Keisho Senior High School (March 2019)		Z&EW (1.43 GB) Info JPG (47 MB) SAC(EW) (34 MB) analysis by Manato Sano of Ritsumeikan Keisho Senior High School (April 2019)
03/13 - 03/15	LP		NS (1.44 GB) Info JPG (45 MB)	
03/23 - 03/25	LP		NS (1.46 GB) Info JPG (48 MB)	
03/24 - 03/25	SP	Z (1.45 GB) Info	NS (1.45 GB) Info	EW (1.44 GB) Info

オープンデータと地震学

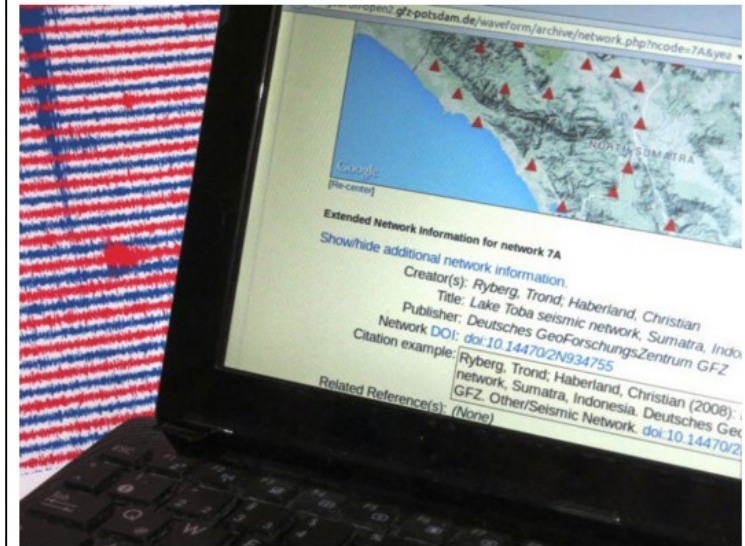
(S23) オープンデータと地震学 [コンビーナ: 加納靖之・汐見勝彦]

研究データの公開と流通が学術の発展に不可欠な要素であるという考えが広がりつつある。地震学の周辺では、地震波形データをはじめとして公開・流通など先端的な事例がある一方で、データの生産に関わる研究者・技術者の評価などの課題も多い。研究データ共有の手段である永続的識別子付与やデータ出版などの取り組みについて、他の研究分野や国内外の情勢、各大学・研究機関での具体的な取り組みについて情報交換し、地震学の発展に寄与するデータ共有のあり方について考えたい。

口頭発表17件(招待3件)
ポスター発表4件
事例, 課題, 将来展望 etc.
データDOI, データジャーナル etc.

Why Seismic Networks Need Digital Object Identifiers

In a move to give credit where it's due, the International Federation of Digital Seismograph Networks will link digital object identifiers to data from seismic networks and project deployments.



A new system of digital object identifiers will bring long-overdue recognition to scientists providing data through seismic networks.
Credit: Peter Evans

FDSN Evans, et al. (2015) EOS
doi:10.1029/2015EO036971. P

歴史的記録（京大防災研紀州観測点）

どのようなメタデータが適当？必要？

The screenshot displays a Windows File Explorer window with the address bar showing the path: PC > DVD RW ドライブ (D:) 201511241730 > JPEG > 紀州 1969-1974. The main area shows a grid of 539 JPEG files, each with a thumbnail and a filename. The files are organized in a grid with 5 columns and 11 rows, with the last row containing only 4 files. The filenames follow a pattern: 00768_b.jpg, 00768_f.jpg, 00769_b.jpg, 00769_f.jpg, 00770_b.jpg, 00770_f.jpg, 00771_b.jpg, 00771_f.jpg, 00772_b.jpg, 00772_f.jpg, 00773_b.jpg, 00773_f.jpg, 00774_b.jpg, 00774_f.jpg, 00775_b.jpg, 00775_f.jpg, 00776_b.jpg, 00776_f.jpg, 00777_b.jpg, 00777_f.jpg, 00778_b.jpg, 00778_f.jpg, 00779_b.jpg, 00779_f.jpg, 00780_b.jpg, 00780_f.jpg, 00781_b.jpg, 00781_f.jpg, 00782_b.jpg, 00782_f.jpg, 00783_b.jpg, 00783_f.jpg, 00784_b.jpg, 00784_f.jpg, 00785_b.jpg, 00785_f.jpg. The left sidebar shows the 'クイック アクセス' (Quick Access) section with links to 'デスクトップ' (Desktop), 'Dropbox', 'Creative Cloud', 'ドキュメント' (Documents), 'ダウンロード' (Downloads), 'ピクチャ' (Pictures), '1603', 'gmt', '紀州デジタル', and '立替私H27年度'. Below this, there are links to 'Dropbox', 'OneDrive', 'notes', 'work', and 'ドキュメント'. The 'OneDrive - 京都大学' (OneDrive - Kyoto University) section is also visible. The bottom status bar shows '539 の項目' (539 items), '1 個の項目を選択' (Select 1 item), and '0.98 MB'.

京都大学古地震研究会

- 史料解読の勉強会
- 研究に関する情報交換 / 情報技術の活用（情報科）

図書館員の皆さんとの勉強会も



市民参加の取り組み



<https://honkoku.org/>

みんなで翻刻

仲間と、AIと協力して、
古文書解読に参加しよう！

参加する

※現在新バージョンの試験運用中です。安定稼働するまで
旧バージョンも併せてご利用ください。

いいね！ ツイート B!ブックマーク LINEで送る

- 市民参加で(地震)史料を翻刻
- やりがい搾取？ プロの仕事をとる？ ⇄ 学習・楽しみの提供

「みんなで翻刻 地震史料」(バージョン1)

- 概要

- URL: <https://v1.honkoku.org>
- 歴史災害史料の市民参加型翻刻プロジェクト
- 2017年1月10日に Web サイトを正式公開
- 京都大学古地震研究会が開発・運営

- プロジェクトの目的

1. 災害史料の大規模テキストデータベース構築
2. 翻刻作業を通じた市民の防災意識向上

バージョン1の成果(1015日目)

- 総入力文字数…618万文字
- 翻刻が完了した画像…8274/8925枚(91%)
- 翻刻が完了した史料…498/508点
- 参加登録者数…5320名

東大地震研の所蔵和古書はすでに全点翻刻完了

設計の特徴

- 古文書解読の学習サービスとして設計されている
 1. くずし字学習支援アプリKuLAとの連携
 2. 参加者相互の「学び合い」の支援
- くずし字解読を学習しながら、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



12万回DL(2019/2)

参加者のツイッターより

- 「みんなで翻刻 honkoku.org、ちょっと読める人があらかた翻刻すれば、後で達人が誤刻箇所や判読不能箇所を修正してくれる最高の環境なのでとても勉強になる #みんなで翻刻」 (2017-01-09)
- 「みんなで翻刻、前の人を読んだの見ながら、自分にも新たに読めるところがあって面白いね。国文学やってた学生時代にこういうのあったらなあ。時代は変わったね。」 (2017-01-12)

成果物の品質について

- 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

- 学術出版される史料集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た

バージョン1の課題

1. 対象史料の拡大

- 地震研のデジタルアーカイブに特化した構成になっている

2. 初学者のサポート

- 参加登録者5,000人のうち、実際に翻刻に参加したのは350人程度

バージョン2の新機能

1. IIIF対応
2. くずし字認識AIの導入
3. 翻刻文マークアップ言語の導入
4. 翻刻文の画像埋め込み
5. 翻刻文の全文検索

1. IIIF 対応

- IIIF (International Image Interoperability Framework)
 - デジタルアーカイブが公開する画像データの相互運用のための国際標準規格
 - 機関ごとに乱立していた画像データの提供方式に統一的方法を与えた
 - 日本国内のIIIF対応機関
 - NDL, 国文研, 東京大学, 京都大学, 九州大学, 慶応大学, 関西大学, 近畿大学, 歴博,...

現在進行中の翻刻プロジェクト

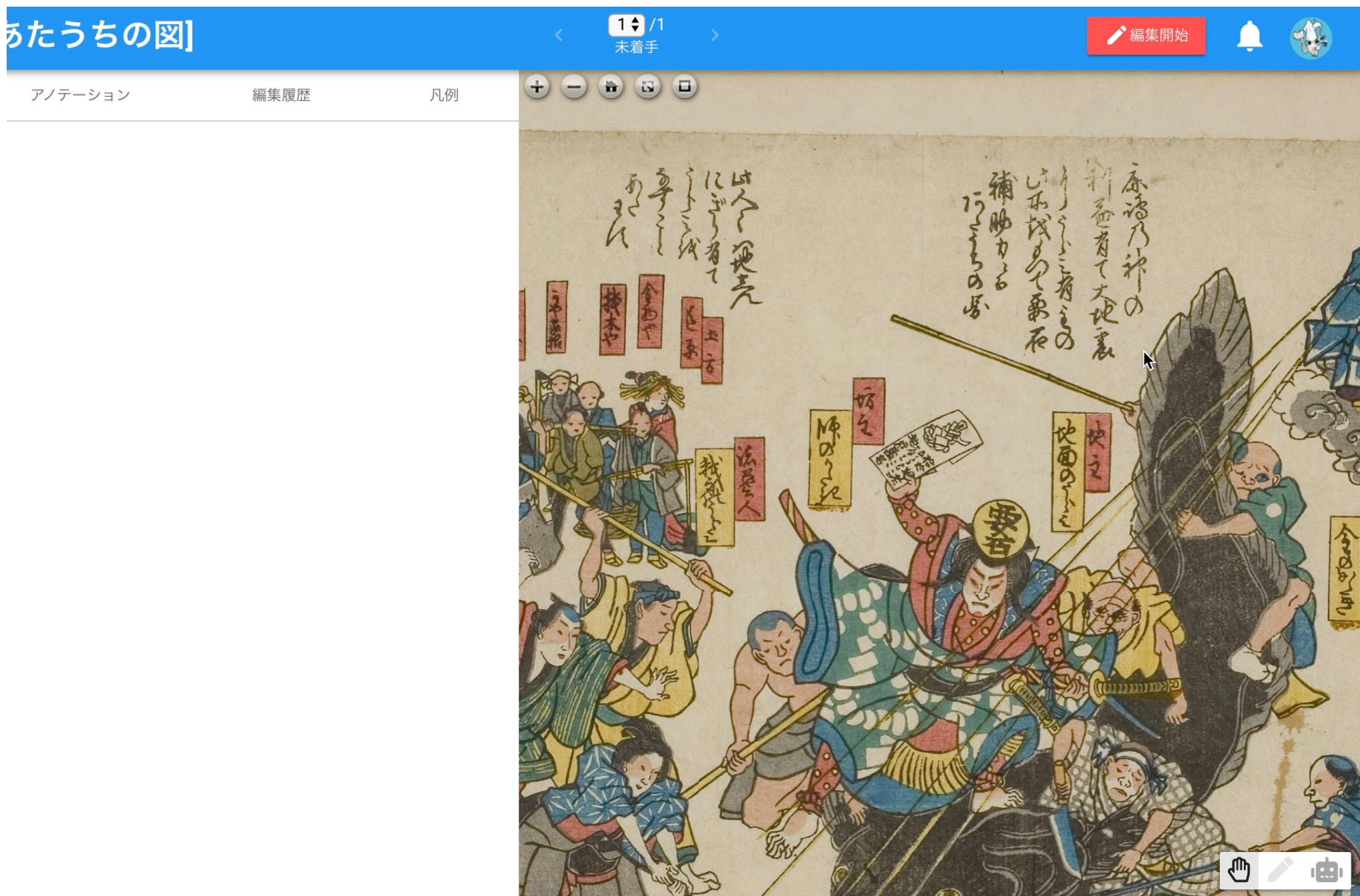
1. 翻刻！ 東寺百合文書

- ユネスコ世界記憶遺産にも指定されている東寺百合文書の翻刻プロジェクト
- 百箱のうち「セ函」「ヤ函」の210点を公開中

2. 翻刻！ 石本コレクション

- 東大総合図書館が所蔵する石本巳四雄の災害史料コレクション578点
- 近世のかわら版や鯰絵を多数含む

2. くずし字認識AIの搭載



くずし字認識AI

2種類の認識プログラムを選択可能

1. CODH koguma-net

- CODHの研究者タリン・カヌーワットさんが開発
- JavaScriptで動作し、クライアント側で実行可能
- 翻刻文をAIの教師データとして活用する研究も

2. 凸版印刷くずし字認識AI

- 凸版印刷のサーバーにあるAIをネットワーク経由で利用

バージョン2の経過(95日目)

- 総入力文字数…48.5万文字
- 翻刻が完了した史料…299/788点
- 参加登録者数…400名

ニコニコ生放送 『みんなで翻刻してみた』

ドワンゴとの共同企画. 2017年3月から月1回放映



「ニコニコ超会議2017」への出展

2017年4月末に開催された「ニコニコ超会議2017」にブース出展



今年のニコニコ超会議2018(4/28-29)@幕張メッセにも出展予定

みんなで翻刻の工夫

- システムと素材
 - 学習ベース（添削希望 etc.）
 - くずし字学習支援アプリKuLAとの連携
 - 参加の楽しみ（ゲーム性 / SNSでの自慢）
 - 東大地震研石本文庫（図書室 / 既にオープン）
- 広報
 - リリース時期
 - (株)ドワンゴとのコラボ
 - 学術研究支援室(KURA)の支援

目論見と実際

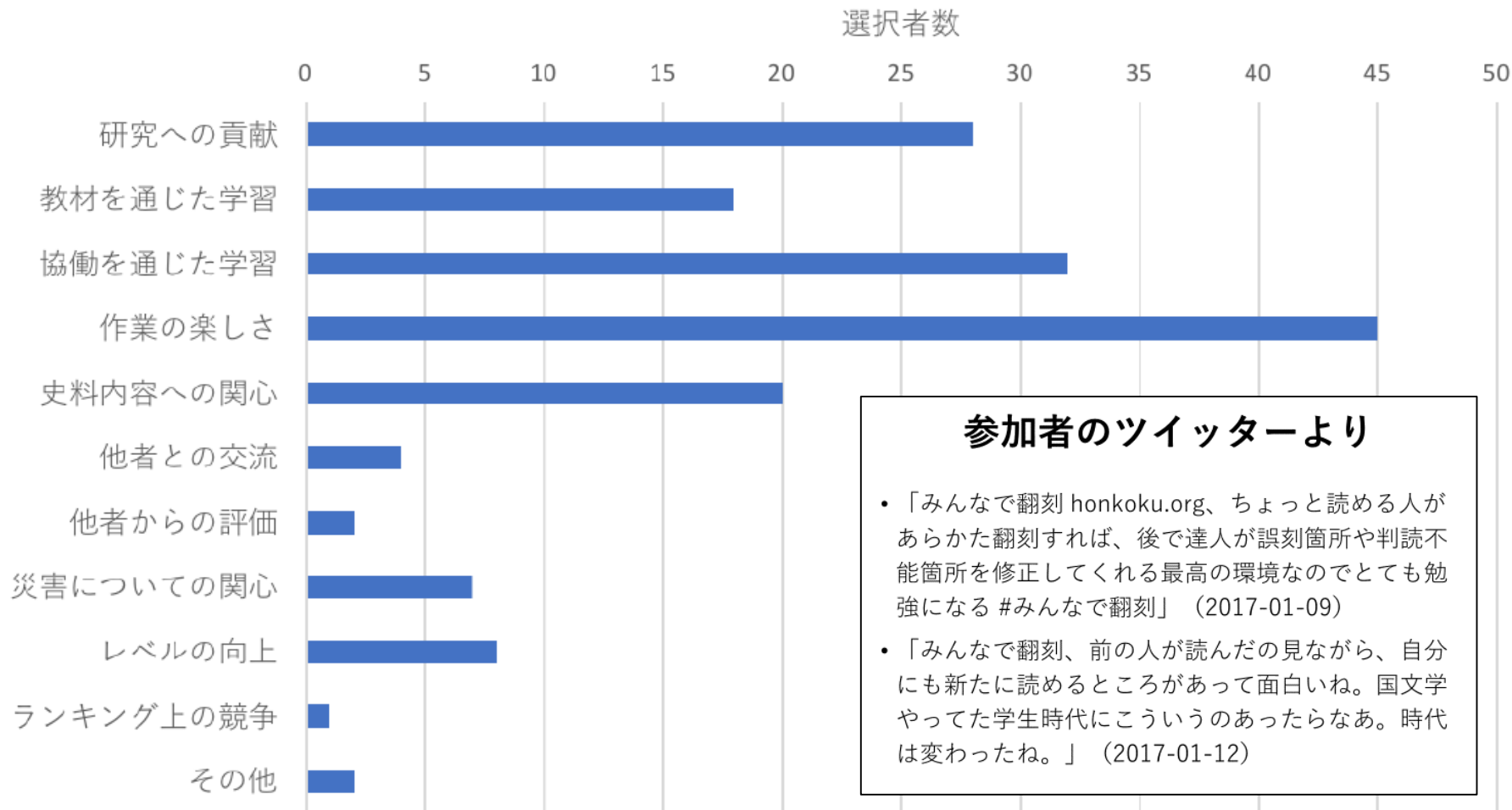
波及効果

- 過去の地震に関してこれまでに知られていない事実が発見される可能性があります。
- 過去の解析の前提となったデータの再検討が可能になります。
- くずし字の知識習得者を増加させます。
- 地震史料のテキスト化を加速することで、史料を後世に残します。
- SNS 機能を利用し、日本各地で眠る災害史料の発掘に寄与します。
- 地震史料以外にも、日本各地、各組織で進む史料のデジタルアーカイブ化やデジタル公開に新たな戦略をもたらします。
- 過去の災害について学ぶことで、市民の防災意識の向上をめざします。

京都大学プレスリリース（2017年01月10日）

http://www.kyoto-u.ac.jp/ja/research/research_results/2016/170110_1.html

目論見と実際



グラフは橋本雄太(2019)京都大学博士論文より

<https://repository.kulib.kyoto-u.ac.jp/dspace/handle/2433/233817>

目論見と実際

- 「ほんとはね、特定の地域の史料はその地域の人
がやるのが一番いい。だって、小字だとか今はない郡
名だとかを自然な形で触れているから、処理のス
ピードがそのほかの地域にいるよりずっと早いはず
だし、なにより愛着を持って取り組めるんじゃない
かなって。（2017-01-13）
- 「諸国海辺地震津波書」から軽井沢関係の部
分・・・「かるい沢宿浅間山へん大ひにあえるまた
あさま山けむり此日一向にいてず其外山々大しんど
う」（1854年安政東海地震・安政南海地震のこ
と）」（2017-01-17）

こんな意見も

- 「みんなで翻刻って若い院生のバイトと勉強の場を奪うことにならんかな、という一抹の不安が。」
- 「みんなで翻刻、報酬がないのでやらない」

自慢したいこと

- たくさんの参加者を得た → 集合知
- 安定的に動作・運用できている（改善点はある）
- 学習・知識の伝達
 - くずし字解読・歴史史料への関心
 - 自然災害への気づき
- オープンなデジタルアーカイブの利用例を提示
 - 東寺百合文書
 - 東京大学総合図書館蔵石本コレクション

がんばらなければならないこと

- 専門家による研究への活用
 - 歴史地震：既にある程度研究されている
 - 歴史地震史料コーパス？
- 非専門家による活用
 - 読みもの
- 開発・運用体制
 - 大学等の研究としては限界がある？

等身大の地震学

日本地震学会の改革に向けて：行動計画2012

- 1) 提言：重要な課題について議論の機会を増やすための会員間の議論の場や機会を設けるべき
対応：「特別シンポジウム」を継続的に実施し、議論の内容を「モノグラフ」の刊行によって伝えていく。
- 2) 提言：地震・津波防災に関連する他学会との連携を強化していくべき
対応：講演会における他学会との共催による「特別セッション」の企画を推進すると共に、特に日本地震工学会との連携を強化していく。
- 3) 提言：国家によるプロジェクトに関する学会における議論の場を提供すべき
対応：国家プロジェクトに関連する「特別セッション」を継続的に企画し、議論の内容を、「ニュースレター」や「モノグラフ」を活用して出版する。
- 4) 提言：“地震予知”への取り組みを見直すべき
対応：用語の整理に基づいて、“地震予知”に関する誤解や無用な議論を避けると共に、地震発生予測の研究の現状をアウトリーチ活動を通じて社会に伝えていく。
- 5) 提言：社会に対して、“等身大”の地震学の現状を伝えていくべき。
対応：関連委員会によって「地震学を社会に伝えるワーキンググループ」（仮称）を創設し、地震学の現状を一般市民の目線に立って社会に伝えていくとともに、地域防災への貢献及び社会からの要請を受け止める場となることをめざす。また、委員会構成について見直しを行う。
- 6) 提言：学術情報発信のあり方を検討すべき。
対応：「地震」（和文）と「Earth, Planets and Space」（英文）今後のあり方について、臨時委員会（またはワーキンググループ）を構成して半年～1年程度の期間をかけて検討する。

人文社会系分野における オープンサイエンスの在り方

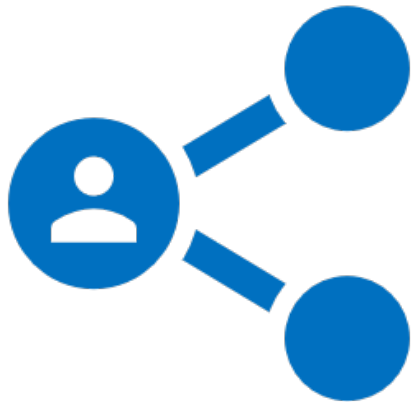
- 各種の悩みは分野を問わない
 - やってみたい / やるべきか / 意味があるのか
 - どうしたらいいか
- 自然科学分野のほうが親和性があると見られがち？
 - 数値データ vs 非数値データ（テキスト/画像 etc.）
 - データ公開に対する考えかた
 - 非専門家と接するという意味では同じ
- 人文情報学の隆盛が追い風に

今後の展開

- みんなで翻刻
 - 各地のデジタルアーカイブとの協力
 - 開発・運用体制
- 地震史料搜索隊 ← 古典オーロラハンター
 - IIFを利用して、とにかくパラパラと史料を閲覧
 - 「地震」などキーワードを抽出 → フラグ
- 先輩方がもともとでやってきたこと
 - デジタルも活用して、より広くつながりたい

研究データやプロセスの共有

- 同床異夢だとしても，興味・関心を共有できる対象（プロジェクト）の形成が肝心なのでは？
 - 大義だけでは動かない 「研究」はそんなに大事？
- 専門家・非専門家の相互のコミュニケーションをうまくデザイン/プロデュースできないか



第3回東京大学学術資産アーカイブ化推進室主催セミナー

東京大学デジタル万華鏡

～デジタルアーカイブズ構築事業の成果紹介と
これからの活用を考える～

2017年度にスタートした東京大学デジタルアーカイブズ構築事業では、各部局の資料デジタル化と公開の支援を行っています。三年目となる現在、各事業の個性的な資料公開が続き、また2019年6月にはデジタル化資料を横断的に検索できる「東京大学学術資産等アーカイブズポータル」を公開しました。

今回のセミナーでは、2018年度までに行われた事業の紹介と、デジタル画像を用いた新たな教育・研究活動の実践に関する講演を行います。

東京大学が持つ資料の多種多様さを知っていただくとともに、デジタルアーカイブ構築、デジタルコンテンツ活用等についての情報を分かち合うことで、学内・学外を問わず、デジタルアーカイブを育てていくヒントを共有できればと考えています。是非ご参加ください。

2019年11月29日（金）13:30→17:15 受付 13:00～

東京大学 弥生講堂一条ホール（本郷地区弥生キャンパス）

【プログラム】

■ デジタルアーカイブズ構築事業の成果から

・東京大学学術資産等アーカイブズポータルの紹介

・ライトニングトーク

デジタルアーカイブズ構築事業の各担当者が、公開データの魅力を語ります！

登壇予定：総合研究博物館、文書館、法学部法制史資料室、明治新聞雑誌文庫、文・部呂実習施設、経済学部、教養学部（駒場図書館）、情報学環、史料編纂所、地震火山史料連携研究機構、空間情報科学研究センター、総合図書館 など

・デジタルアーカイブ活用のために 最新技術の紹介

中村 寛（情報基盤センター助教／学術資産アーカイブ化推進室）

■ これからの活用に向けて

・講演1 加納靖之（東京大学地震研究所）
総合図書館蔵石本コレクションと「みんなで翻刻」



・講演2 田村隆（東京大学大学院総合文化研究科）
東大本『源氏物語』と
新たな本文研究プラットフォーム



参加対象者：デジタルアーカイブ、デジタルコンテンツ
活用などに関心のある方

学内者、学外者を問いません

以下のサイトから事前にお申し込みください

<https://www.lib.u-tokyo.ac.jp/ja/library/contents/event/20191015>

東京大学附属図書館ウェブサイト > イベント



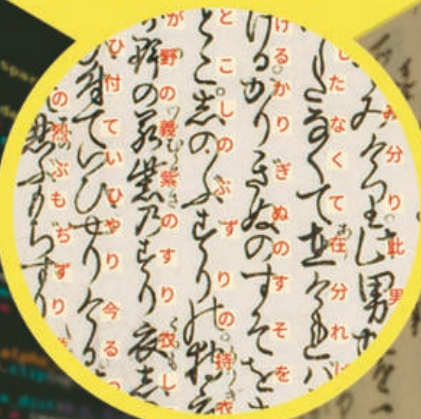
お問い合わせ先：

東京大学学術資産アーカイブ化推進室（附属図書館総務課）
digital-archive@lib.u-tokyo.ac.jp / 03-5841-2613

※障害等のため、会場設備等に配慮が必要な場合はご連絡ください

AI がくずし字を 読む時代がやってきた

AI（人工知能）を活用してくずし字を読み解く研究の、
過去・現在から未来まで



基調講演

内田 誠一

佐々木 孝浩

（九州大学大学院システム情報科学研究科）

（慶應義塾大学附属研究所系連文庫）

Session1

馬場 上

青池 橋本

（奈良文化財研究所）

（東京大学史料編纂所）

基

聡

亨

雄太

（国立国会図書館電子情報部電子情報企画課
次世代システム開発研究室）

（国立歴史民俗博物館）

Session2

山本 和明

北本 朝展

（国文学研究資料館）

（ROIS-DS 人文系オープンデータ
基盤技術センター／国立情報学研究所）

Session3

カラーヌワット タリン

Alex Lamb

（ROIS-DS 人文系オープンデータ
基盤技術センター／国立情報学研究所）

（セントリオール大学 NELA）

Session4

ロバート キャンベル

喜連川 優

（国文学研究資料館）

（国立情報学研究所）

kaggle コンペ入賞者

日本文化とAI
シンポジウム

2019年11月11日（月）10:00-17:45

一橋講堂 東京都千代田区一ツ橋2丁目1-2
学術総合センター2F 参加無料 要事前申込

お申し込みは WEB で

<http://codh.rois.ac.jp/symposium/japanese-culture-ai-2019/>



大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics



大学共同利用機関法人
情報・システム研究機構
Research Organization of Information and Systems



大学共同利用機関法人 国文学研究資料館
National Institute of Japanese Literature