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/

どんな研究？

- **データ駆動型人文学**：情報学・統計学の最新技術を用いて人文学資料（史料）を分析
- **人文学ビッグデータ**：人文学研究の成果に基づき構築したデータセットを超学際的に活用
- **人文学のデジタル変革**：オープンサイエンスなど新しい潮流を取りこんだ人文学研究へ

何がわかる？

- データ駆動型研究を進めるための、**機械可読データセット**を構築・公開
- **オープンソースソフトウェア**を公開し、各種のサービスを外部からも活用
- **共同研究**を通して知識や資源を提供

研究背景

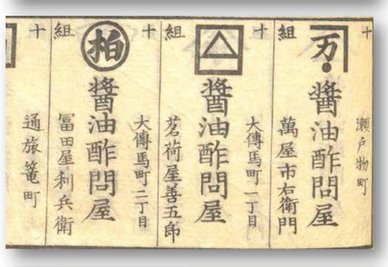
データサイエンス共同利用基盤施設（DS）は、情報・システム研究機構（ROIS）内に設置された研究組織。生命科学・地球科学から人文学・社会科学まで、データサイエンスを幅広い分野で推進する6つのセンターがある。

人文学者や情報学者などが分野横断的に協働し、人文学的な問いを情報学的な手法で解く、人文学資料から作る過去のビッグデータを分析する、などの研究に取り組む。人文学的な視点は、AIなどのテクノロジーを社会に取り入れるためのガイドとしても重要になりつつある。

研究内容

江戸地名データ「江戸切絵図」

商業ビッグデータ「江戸買物案内」



観光ビッグデータ「江戸観光案内」

文書空間 ⟷ **実体空間**

文書を読み解きながらデータを構造化したい

実世界の構造に合わせてデータを構造化したい

地名識別子

GeoLOD

文書空間と実体空間を、識別子を介して双方向に結合する

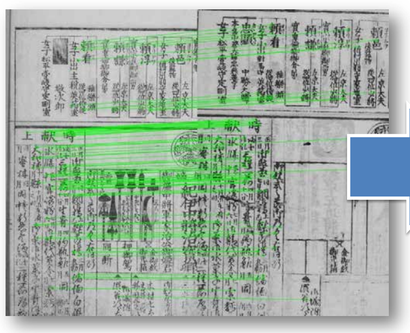
古地図のジオレファレンス

データポータル「edomi」

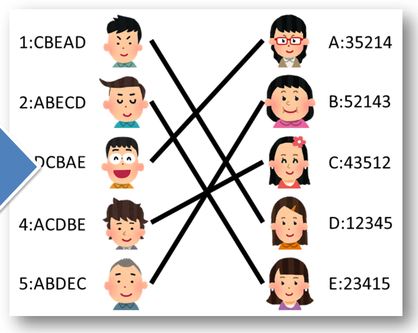
江戸ビッグデータ

武鑑全集

同一板木だが一**部だけ修正**されているかを確認

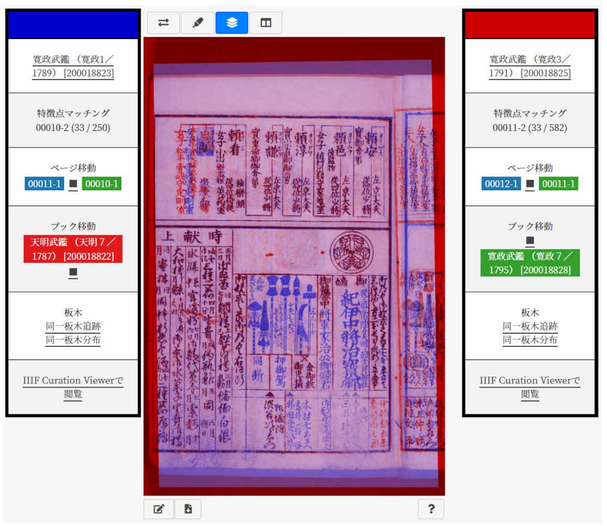


1. ページ照合：特徴点ベースの画像マッチングアルゴリズム



2. ブック照合：画像マッチング結果を利用した安定結婚アルゴリズム

3. 板木追跡：ブック照合結果に基づき、同一板木を書籍横断的に接続



vdiff.js：画像比較ツールにより同一板木の更新を強調

連絡先：北本朝展／ ROIS-DS人文学オープンデータ共同利用センター センター長

Email : kitamoto@nii.ac.jp

国際的な画像共有に基づくビジュアルな人文情報学研究

IIIFを用いたキュレーションの展開

どんな研究？

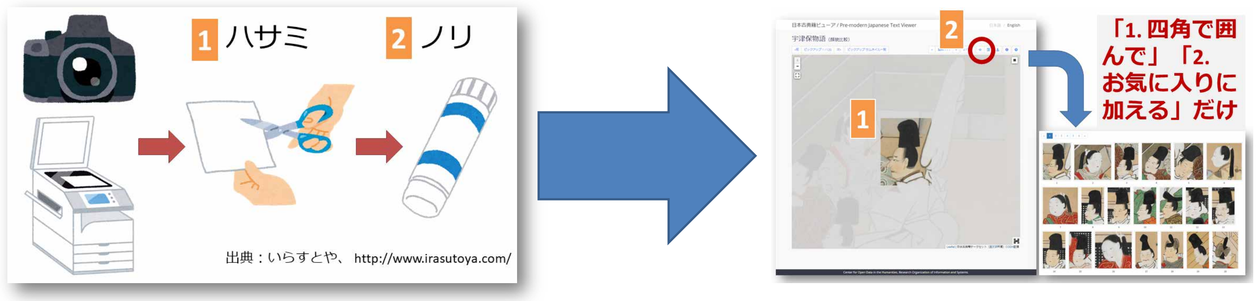
- **IIIF（International Image Interoperability Framework）**を用いたオープンな画像共有が、全球規模で急速に進展。
- **IIIF Curation Platform（ICP）**をオープンソースで公開。キュレーション等の独自機能が評価され、**国内外のいくつかのプロジェクトで採用**。

何がわかる？

- **IIIFキュレーション**：IIIF画像から一部を切り取ってアノテーションし、テーマごとに集め、自分のコレクションとして公開する機能。
- 「キュレーション」の**網羅性と検証可能性**を高めることにより、**ビジュアルな資料を用いる人文学研究**に幅広く適用可能な方法論を提案。

研究背景

提供者視点から利用者視点へ：IIIFはミュージアムやライブラリ等の**画像提供者側の標準化**に大きく貢献した一方、**画像利用者のニーズ**はあまり考慮されていない。**IIIF Curation Platform**は、研究者が**画像を収集、注釈、処理**するというニーズに応える機能を実現した。

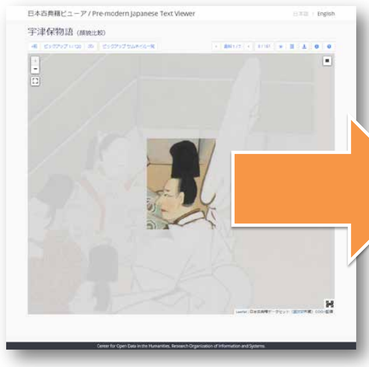


ワークフローのDX：コピー・はさみ・ノリを用いて資料の部分をテーマごとに収集し比較するアナログな手法に対し、ICPはワークフローの全体をデジタル化。

研究内容

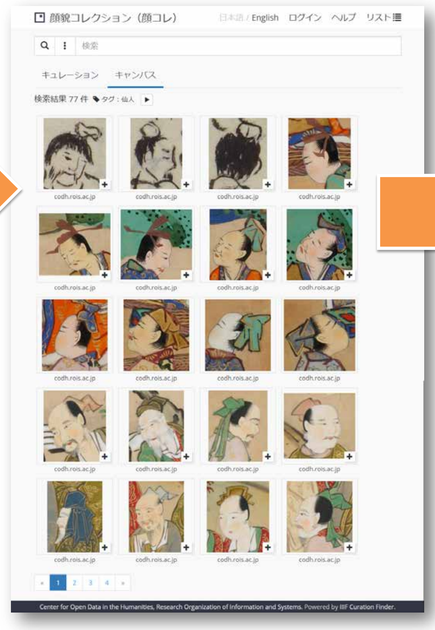
ICPを用いた絵巻作品の顔貌分析：画像の一部を収集する**マイクロコンテンツ**の技法＋オブジェクトとその文脈を切り替えながら画像を比較する**GM法＝網羅的な様式分析**を実現。

IIIF Curation Viewer

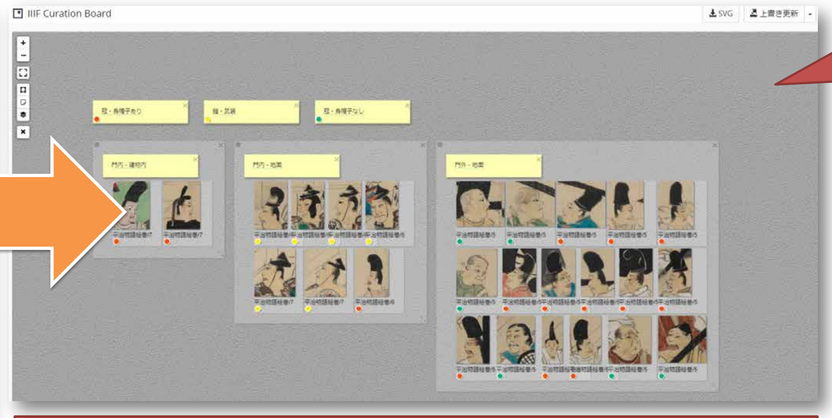


顔貌コレクション（顔コレ）：日本の絵巻物・絵本から顔貌だけを切り出した顔貌データセット

IIIF Curation Finder



IIIF Curation Board



GM法：キュレーションした画像を二次元平面上に配置し、メタデータなどの類似性を用いてグルーピング。美術作品の網羅的な分析が初めて可能となり、美術史研究を大規模化。

国宝の絵巻分析にも活用

浮世絵を対象とした「顔コレ」を公開。今後も**様々な作品やオブジェクト**に拡大し、データセットを多様化する計画。

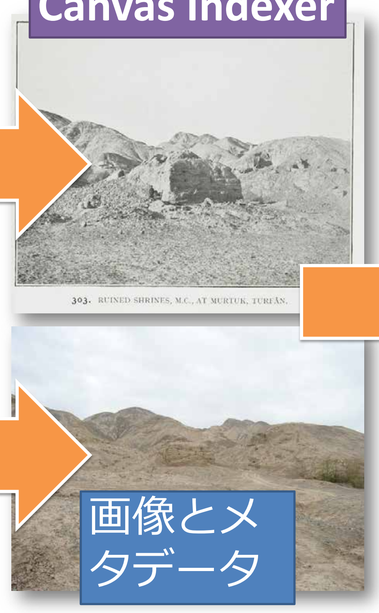
ICP と OmekaS の連携に基づくシルクロード遺跡データベースの構築：ICPを用いて**キュレーションを資料横断的に集約**し、画像を根拠とするデータベースを構築。

IIIF Curation Viewer

IIIF書籍画像から切り取り

IIIF写真画像から切り取り

Canvas Indexer



画像とメタデータ

Omeka S

2枚の画像をvdiff.jsで比較

物体検出技術による作業効率化：Faster R-CNNを用いることで70%程度の顔貌を自動認識。残りは手動で追加。

- 鈴木 親彦, 髙岸 輝, 本間 淳, Alexis Mermet, 北本 朝展, “日本中世絵巻における性差の描き分け－IIIF Curation Platform を活用した GM 法による『遊行上人縁起絵巻』の様式分析”, 人文科学とコンピュータシンポジウム じんもんこん2020論文集, pp. 67-74, 2020年12月（**情報処理学会 山下記念研究賞 受賞論文**）
- 西村 陽子, 北本 朝展, “カード単位の照合エビデンスを共有するシルクロード考古遺跡情報の統合データベース”, 人文科学とコンピュータシンポジウム じんもんこん2021論文集, pp. 146-153, 2021年12月（**人文科学とコンピュータシンポジウム じんもんこん2021 最優秀論文賞 受賞**）

連絡先：北本朝展／ ROIS-DS人文学オープンデータ共同利用センター センター長
Email : kitamoto@nii.ac.jp

誰でも簡単に使える！

AIくずし字認識アプリ「みを」

どんな研究？

- ディープラーニングによるAIくずし字認識をスマホアプリから利用。
- iOS版とAndroid版の両方をリリース。
- AIくずし字認識だけでなく、認識結果修正など、さまざまな機能も搭載！

何がわかる？

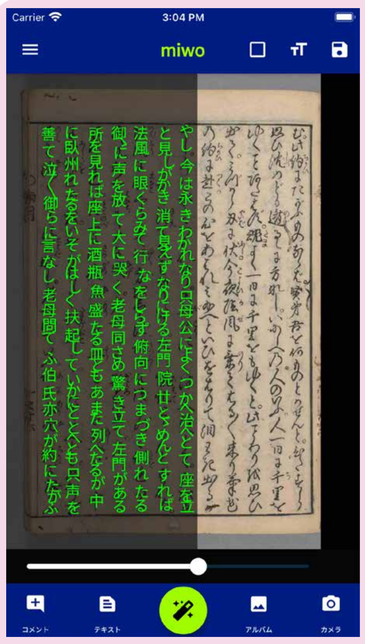
- 手持ちのくずし字資料を、スマホカメラで撮影し、数秒でくずし字認識を行って、現代日本語文字を表示する。
- くずし字が読めない人にも、くずし字資料を利用する道を開く。

研究背景

『国書総目録』に古代から1867年まで170万点の古典籍が登録されている。全国の古典籍の数は300万点ほど。古文書を含めると1億点以上。くずし字が読めないと解読できない問題。

2019年にAIによるくずし字認識モデルKuroNetを開発し、IIIF画像に対するくずし字認識サービスを実現。続いて2021年、「手持ちの資料を認識したい！」というユーザーのニーズに応えるためのアプリを開発した。

研究内容

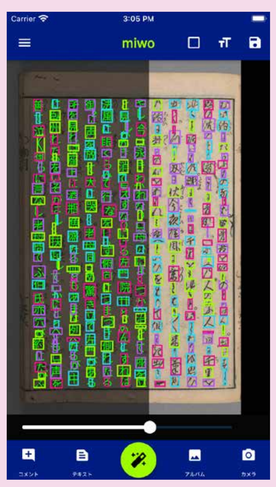


- 物体検出アルゴリズムにより、画像からくずし字を認識し、現代日本語文字に変換する。
- Flutterを用いることで、単一のコードベースでiOS、Androidの両方に対応できた。
- 2021 年 8 月 30 日にリリース。これまで約5.6万回ダウンロード。
- 現在までに認識した画像の枚数は53万枚以上。

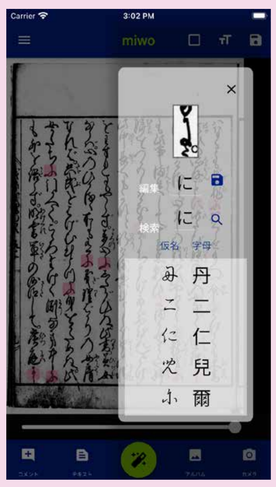
- 「みを」とは、『源氏物語』第14巻、澪標（みをつくし）に由来。
- 「みを（船の水路）を示すために立ててある杭」の意。
- 「みをつくし」が人々の水先案内となるように、「みを」アプリがくずし字資料を読むための道案内となることを目指している。
- 開発：カラーヌワット・タリン氏

「みを」のさまざまな機能

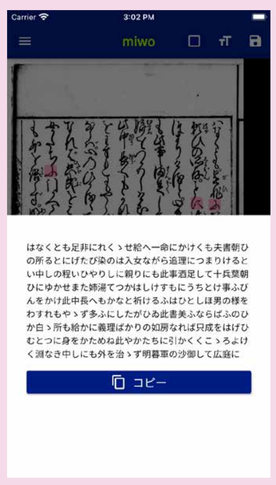
スマホカメラで資料を撮影し、認識ボタンを押すと、数秒で認識結果を表示します。



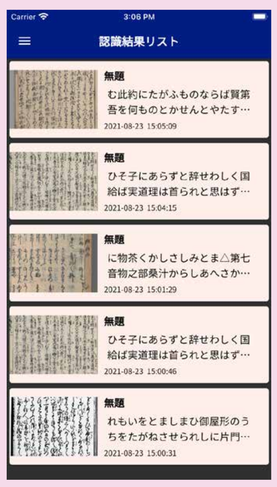
認識結果の文字に加えて、文字のバウンディングボックスも表示できます。



変体仮名の字母確認、認識結果の修正、文字検索による字形確認などの機能を用意しました。



認識したテキストを出力します。またコピーボタンを押すと、他のアプリに貼り付けできます。



認識した資料、修正した結果を保存できます。また、資料に関するコメントも入力できます。

アプリダウンロード

Download on the App Store

GET IT ON Google play

連絡先：北本朝展／ ROIS-DS人文学オープンデータ共同利用センター センター長
Email : kitamoto@nii.ac.jp

過去の記録を統合解析し、超学際的な課題に挑む

歴史ビッグデータ

http://codh.rois.ac.jp/historical-big-data/

どんな研究？

人文学、理学、工学にまたがる連携により、構造化された歴史情報を蓄積し、現代のビッグデータと同様に分析することで、過去の環境や社会の状況を復元するための基盤を構築する。

何がわかる？

過去のさまざまな事象（たとえば、飢饉など）の複合的な課題を、分野横断的な枠組みで歴史的な分析をし、再考することで、新しい発見や新しい歴史の研究につながる。

研究内容

歴史ビッグデータ

歴史資料（史料）を利用するさまざまな分野の人々が協働し、情報基盤を構築し、活用し、研究を促進する枠組み。

- 分野横断的な協働：歴史ビッグデータ研究会・各共同研究
- 多分野の協働による情報基盤：データ構造化・れきすけ

出典：＊1早稲田大学図書館古典籍総合データベース・＊2国立国会図書館デジタルコレクション

れきすけ　史料に関する知識・経験を多分野で共有

利用してほしい・教えてあげたい	探している
史料に〇〇の記録を見つけた。もっと史料を利用してほしい。	〇〇に関する記録がある史料を探す。研究などに史料を利用したい。

利用実績＋新発見

資料の情報を登録

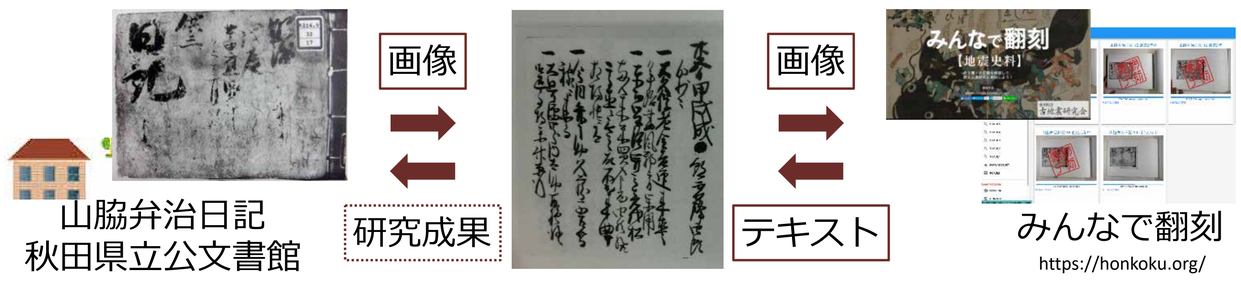
https://rksk.ex.nii.ac.jp/

資料の情報を検索・取得

詳細検索

複数の登録者対応

一つの史料の情報をカードに分割しそれらのカードを相互にリンク。

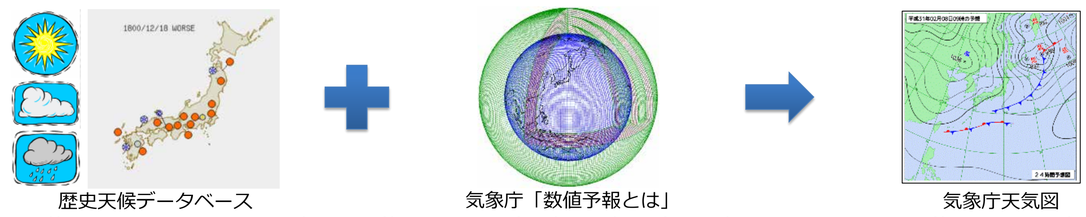


Mika Ichino & Kooiti Masuda (2022) GEOSCIENCE DATA JOURNAL, WILEY Open Access. DOI: 10.1002/gdj3.148
市野 美夏, 増田 耕一, 北本 朝展 (2020) じんもんこん2020論文集

ミレニアム再解析

古天気記録で気象モデルを拘束するデータ同化手法を駆使し、過去の気候・気象（気圧、気温、降水、風など）を再現する。

観測データ	気象モデル	再解析
高密度な古日記などの天気記録	エネルギー保存則運動方程式	古天気情報データ同化19世紀以前の大気場



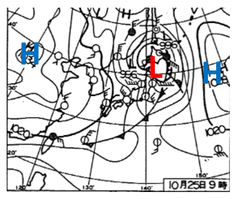
歴史天候データベース
http://tk2-202-10627.vs.sakura.ne.jp/

気象庁「数値予報とは」
https://www.jma.go.jp/jma/kishou/know/whitep/1-3-1.html

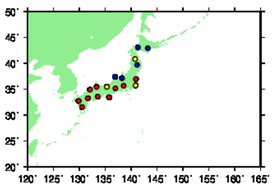
気象庁天気図
https://www.jma.go.jp/jp/g3/

分類した天気を日射量に変換しデータ同化

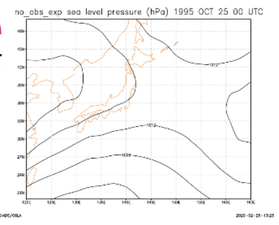
実況
現代の観測にもとづく天気図
(1995-10-25)



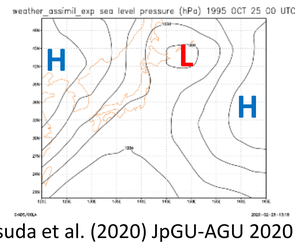
天気情報
気象庁による18地点の天気概況を、晴、曇、雨の3段階に分類



天気情報を入れない
同化実験の気圧分布
→北海道近辺の低気圧（L）が再現されない。

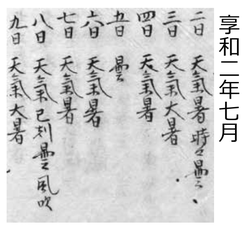


天気情報を入れた
同化実験の気圧分布
→北海道近辺の低気圧（L）が再現された。



Kooiti Masuda et al. (2020) JpGU-AGU 2020

定性的な古天気記録を定量化



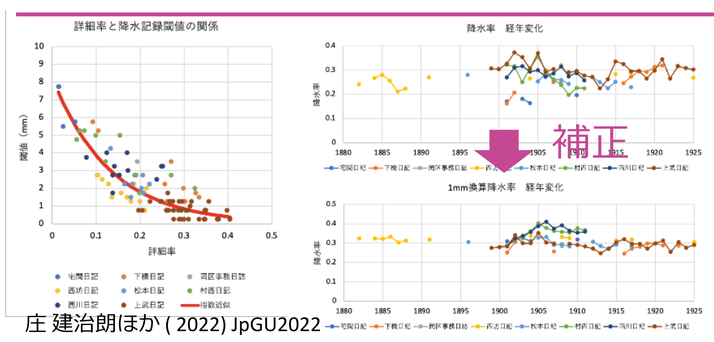
享和二年七月

二日	1802/07/30	天気暑、時々曇
三日	1802/07/31	天気大暑
四日	1802/08/01	天気暑
五日	1802/08/02	曇
六日	1802/08/03	天気暑
七日	1802/08/04	天気暑
八日	1802/08/05	天気、巳刻曇風吹
九日	1802/08/06	天気大暑

- 古天気記録から気象要素を抽出
- 翻刻、和暦を西暦に
- 記録地を緯度経度に
- 分類、コード化し、数値化

府中市郷土の森博物館研究資料集　第1集　県居井蛙録

古天気記録の詳細率で降水日を補正



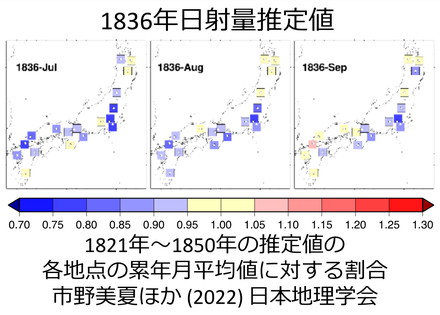
庄 建治朗ほか (2022) JpGU2022

古日記の天気記録の信憑性を古い気象観測データと比較し、記録が詳細であるほど、気象庁の降水日数に近づくことを明らかにし、日記による降水日数を補正した。

庄 建治朗, 鎌谷 かおる, 冨永 晃宏 (2017) 水文・水資源学会. DOI: https://doi.org/10.3178/jjshwr.30.294

古天気記録と米価でみる天保飢饉

1836年日射量推定値



1821年～1850年の推定値の各地点の累年月平均値に対する割合
市野美夏ほか (2022) 日本地理学会

大阪堂島米価1833～1840年

古天気記録から天保飢饉の日射量を推定し、1836年の大阪米価の異常高騰との関係を示した。

神戸大学高槻准教授提供
Mika Ichino et al. (2020) EGUに加筆

市野美夏, 三上岳彦, 増田耕一 (2018) 地学雑誌. DOI: 10.5026/jgeography.127.543

連絡先：市野 美夏　ROIS-DS人文学オープンデータ共同利用センター ／ 国立情報学研究所
Email : m_ichino@nii.ac.jp

デジタル・ヒストリー実践のための
歴史マイクロナレッジ
歴史3Dデータ

小川　潤（ROID-DS人文学オープンデータ共同利用センター/国立情報学研究所）

どんな研究？

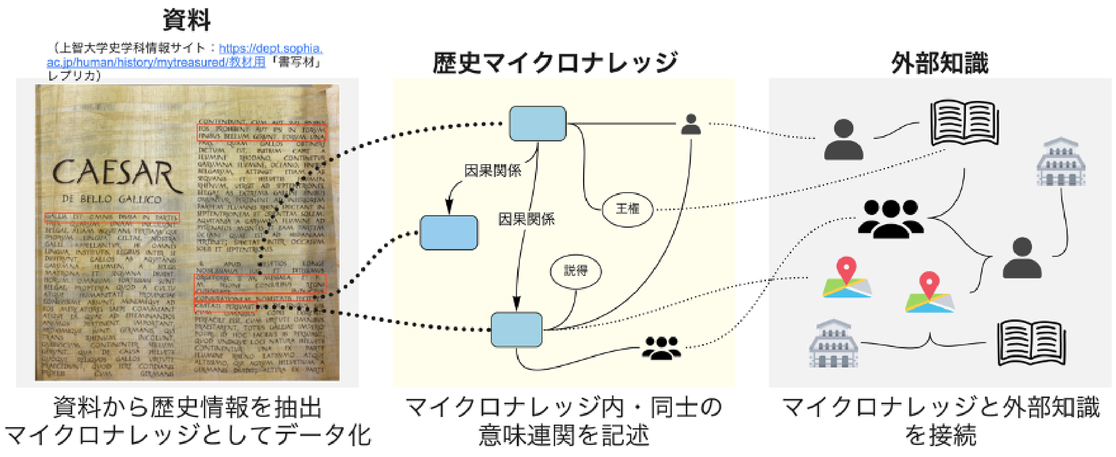
- ヒトやモノの行動や状況、交流に関する資料記述を、知識グラフを用いて外部情報とも接続された「歴史マイクロナレッジ」として構造化
- 歴史的空間・事物を3D空間で再構築・注釈するためのデータ構築、そうしたデータの蓄積・共有手法の確立

何がわかる？

- 歴史マイクロナレッジの蓄積・接続により、資料の内容解析を含むデータ駆動型歴史分析を実現
- 3Dで歴史空間を再構築することで、仮説の検証、データの収集、学術コミュニケーションを実現

研究内容

歴史マイクロナレッジの定義と構築



資料から歴史情報を抽出
マイクロナレッジとしてデータ化

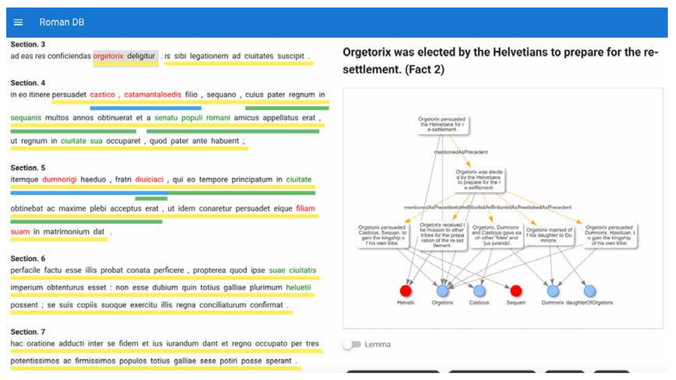
マイクロナレッジ内・同士の意味連関を記述

マイクロナレッジと外部知識を接続

マイクロナレッジとは？
資料においてヒト・モノの行動や状況、交流について言及する個々の記述箇所をデータの一単位（リソース）として記述したもの

特徴
- 知識グラフの活用
- 資料内知識の構造化
- 資料外知識との接続

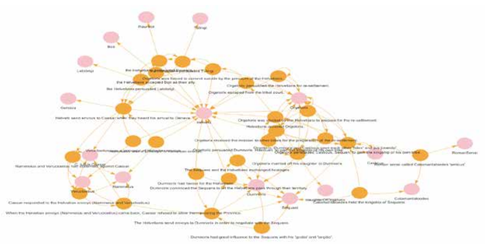
文字資料マイクロナレッジの構築と可視化例



下線は行動・状況・交流に関する資料記述箇所を表示/
フォント色はヒト・モノ・場所を表示

資料記述内容の可視化
行動・状況・交流と、それに関わるヒト・モノ・場所をネットワークとして可視化

資料全体の意味内容構造の可視化も可能に



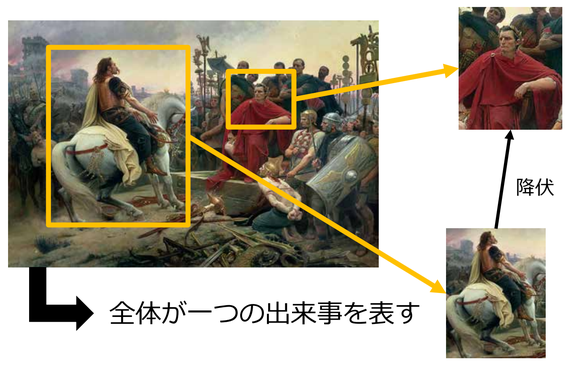
人文学資料マイクロコンテンツとの関係

文字資料マイクロナレッジとの接続
- 文字資料の中に現れるヒト・モノ・場所
- 碑文・写本、古文書等の文字画像


顔貌コレクション（顔コレ）


日本古典籍くずし字データセット

非文字資料マイクロナレッジの構築



全体が一つの出来事を表す

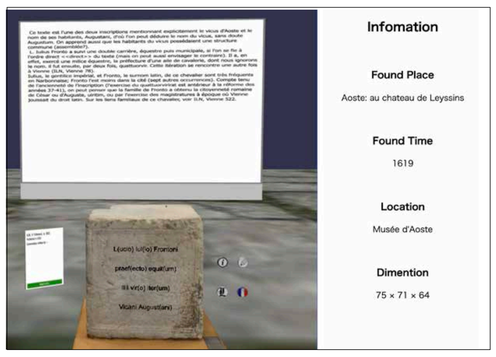
内容解析に資する「深い」歴史知識ベースの構築

歴史3Dデータ

3D Scholarly Edition
＊人文学資料のテクストデータについて提唱されてきたDigital Scholarly Editionの概念を3D・多次元に拡張
＊3Dでの再構築の過程に関わる種々の情報を明示・保存

3Dモデルデータの作成
・フォトグラメトリ・レーザースキャン
・現物が存在しない場合には、資料の記述に基づいてモデリング

文字・図像資料→3Dモデル
正確性をどこまで担保できるか
（自動化は可能？）



学術的アノテーション
・メタデータ & パラデータ
・テキストデータ
・文献情報
・資料の解釈に関わる情報
など

こうしたデータは人文学の専門家が構築するほかないが、どのようにデータを蓄積していくか

3D対応の人文学データ形式の整備

モノについての資料

https://helvia.uco.es/xmlui/handle/10396/15518?locale-attribute=en

資料構造化
3Dレンダリングの可能性を踏まえた構造化の必要

構造化された資料データ
TEI　RDF　iiif

構造化データの3D空間へのレンダリング
既存の人文学データを3D表現に利用可能な形に変換

「モノ」を歴史空間に配置
コンテキストの再構築

空間についての資料


Image created by Jean-Claude Golvin
(https://jeanclaudegolvin.com/lugdunum-lyon/)

ストラボン4巻3章2節

ガリアの全部族によって共同でカエサル・アウグストゥスに捧げられた聖域が、町に向き合う形で、両川の合流地点に築かれた。そこには60を数える部族の碑文を刻んだ特筆すべき祭壇と、それら部族一つ一つにつき彫像が…

文字・図像資料→VR空間
再構築に際しての学術的意思決定を可視化する必要

種々の歴史資料・解釈を統合した“場”としての歴史空間の再構築

これまでの取り組み・今後の構想

>文化資源多次元アノテーションについての東京大学の研究プロジェクトに参加（継続）
>東京国立博物館ポンペイ展での3D計測

>石造物3Dアーカイブとの連携（展望）

連絡先：小川 潤／ ROIS-DS人文学オープンデータ共同利用センター／国立情報学研究所
Email：jun_ogawa@nii.ac.jp