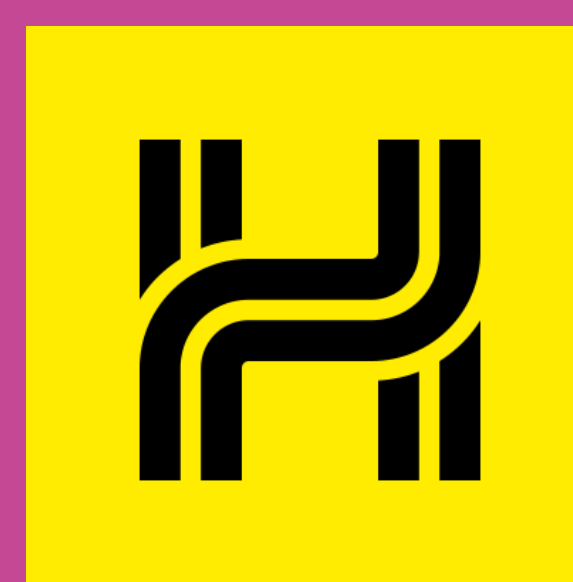


# 日本の文化と歴史をデジタル技術で探る



## - 「データ駆動型人文学」と「人文学ビッグデータ」による人文学研究のデジタル変革(DX) -



ROIS-DS人文学オープンデータ共同利用センター 北本 朝展・山本 智子

### どんな研究？

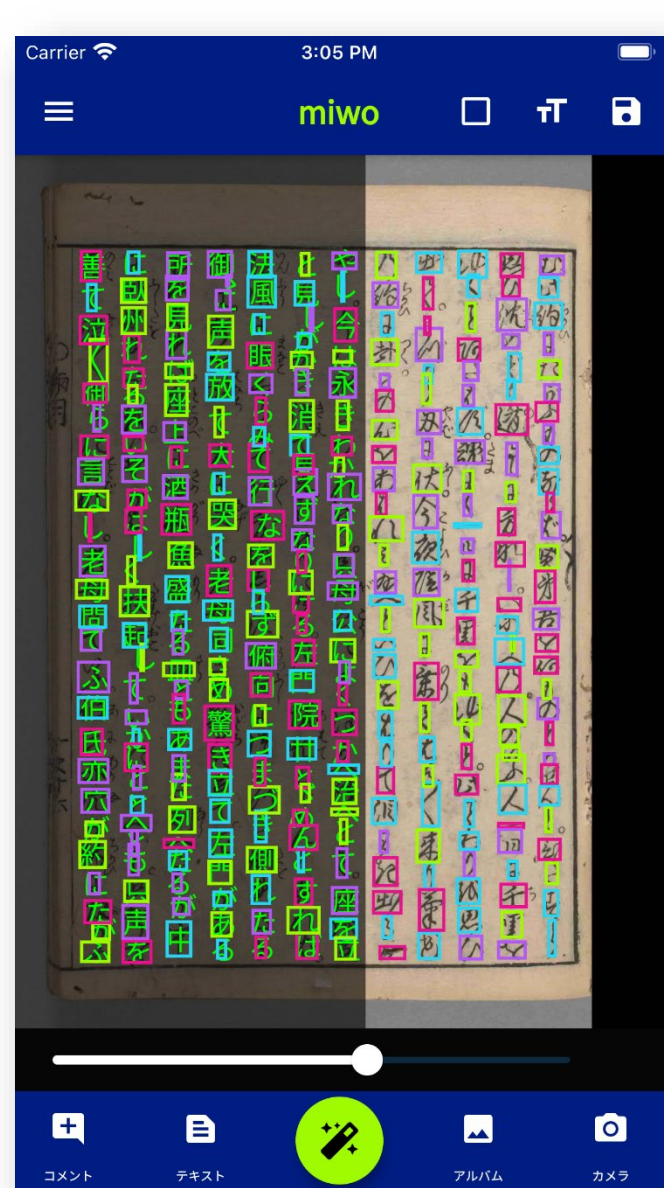
- **データ駆動型人文学**：情報学・統計学の最新技術を用いて人文学資料（史料）を分析
- **人文学ビッグデータ**：人文学研究の成果に基づき構築したデータセットを超学際的に活用
- **人文学のデジタル変革**：オープンサイエンスなど新しい潮流を取りこんだ人文学研究へ

### 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し、人文知に基づく新たな活動を開拓

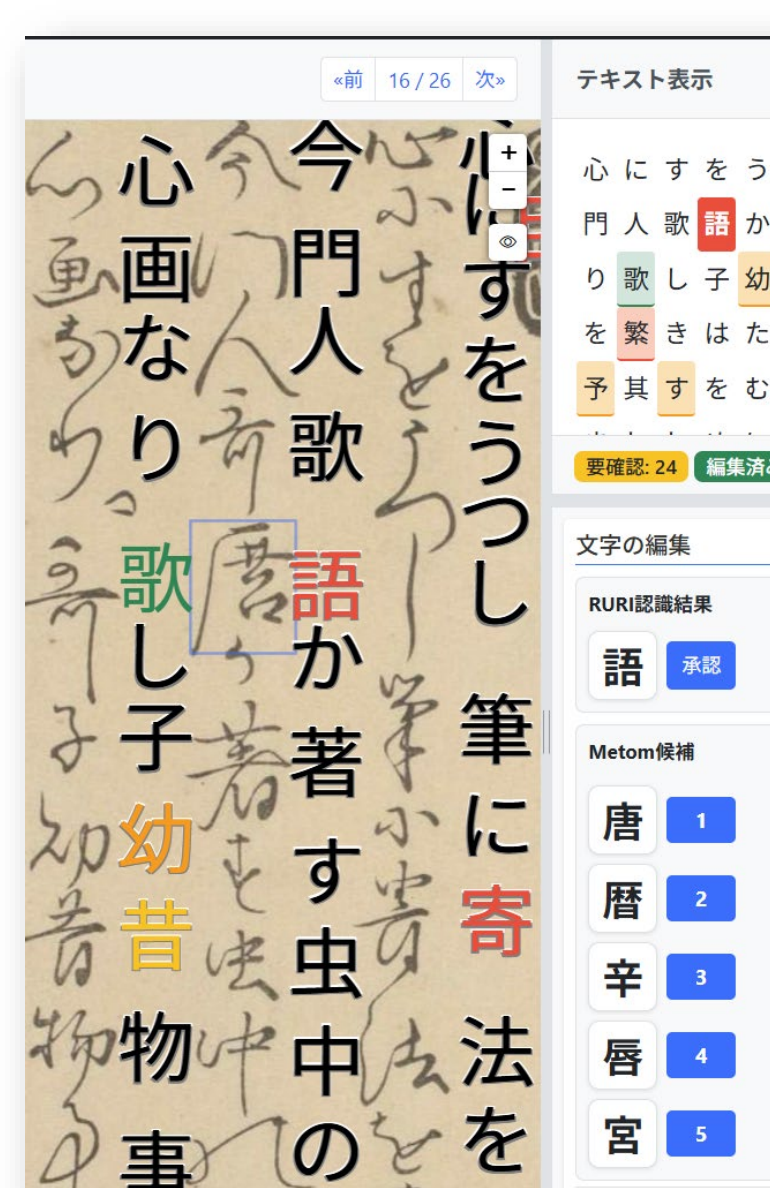
### 研究内容（方法・結果・結論）

#### みを：AIくずし字認識アプリ



- **物体検出アルゴリズム**を用いて、画像からくずし字を認識し、現代日本語文字に変換
- **iOS/Androidアプリを無料公開**、ダウンロードは20万回以上
- これまでの**認識画像枚数は約350万枚**、教育や調査に幅広い利用実績

#### つくし翻刻エディタ：AI強化UI



- **画像単位・文字単位くずし字OCR**の出力結果を統合し、誤りパターンも収集する翻刻ツール
- 他のAIツールも将来的には統合し、**AI支援UIの活用により、翻刻作業コストの低減**を目指す
- **現在開発中**、近日公開を目指す

#### つくしビューア：IIF×LLM×古典籍



- **IIF画像ビューア**にLLMとの**チャット機能**を追加
- 表示された古文を対象に、**現代日本語で内容を質問**できる
- **サーチ (RAG) 機能**とも統合。ページや本全体について質問できる

#### つくし堂：AI生成書店



- **古典籍の要約やキャッチコピー**を、くずし字OCRデータからAIが生成
- **サイト構成もAIが検討**し、コードもAIが生成
- **人間では不可能な分量の作業も、AIであれば実現可能**

#### 歴史ビッグデータ



- **過去の資料から構造化データを抽出**し、現代のビッグデータ技術を活用して過去を復元
- **地名識別子 (GeoLOD)**を開発し、地名と地図を結合
- **データ構造化ツール、データ基盤アプリ**などを多数公開 (edomi、武鑑全集など)
- 古地震 (安政江戸地震) を対象に、**実世界データ**を検証

#### メモリーグラフ



- **同一構図撮影を支援するカメラアプリ** (iOS/Androidアプリを無料公開)
- **複数の写真を重ねて撮影**し、他者と共有することが簡単に可能
- **地域文化資源、観光、防災、人文学研究**などの様々な分野における**フィールドワーク**で利用
- **デジタルアーカイブ**を外に持ち出し、データで人々の活動を支援
- **日本や世界**で複数の活動が進行中



連絡先：ROIS-DS人文学オープンデータ共同利用センター

URL：https://codh.rois.ac.jp Email：北本朝展 kitamoto@nii.ac.jp, 山本智子 t\_yamamoto@nii.ac.jp

# 日本の文化と歴史をデジタル技術で探る

## - 「データ駆動型人文学」と「人文学ビッグデータ」による 人文学研究のデジタル変革(DX) -

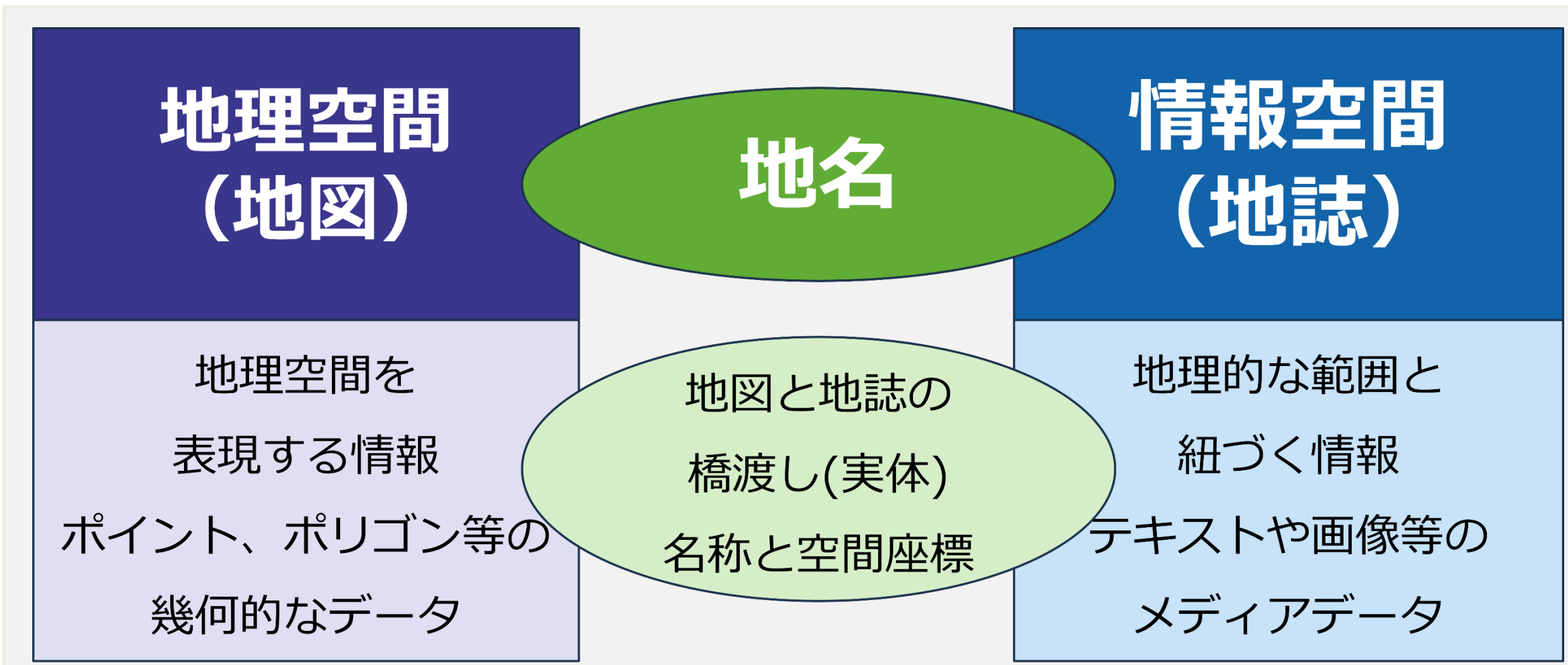
ROIS-DS人文学オープンデータ共同利用センター 北本 朝展・山本 智子

### どんな研究？

DHコンソーシアムプロジェクト (DiHuCo) は、文部科学省が推進する「人文学・社会科学のDX化に向けた研究開発推進事業」の中で構成されるコンソーシアムです。CODHは連携機関として「研究実践ハブ」を運営し、「地図・地誌類領域」を中心とした人文学・社会科学のDX化に取り組みます。

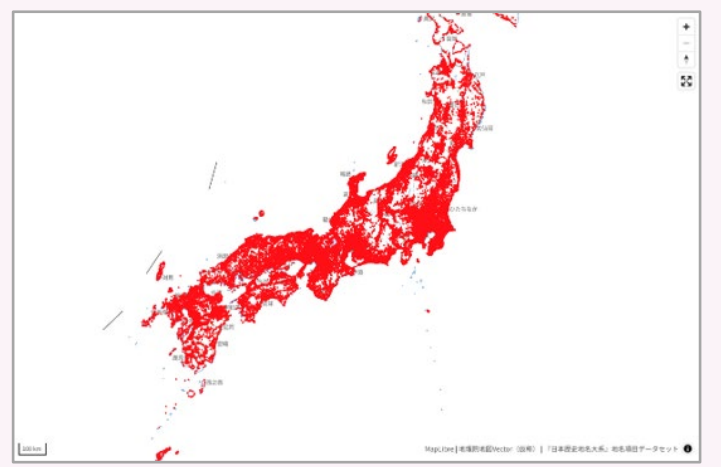
### 背景・目的

以下の3つのテーマに関するデータセット、ソフトウェア、サービス等を公開します。



### 何がわかる？

『日本歴史地名大系』地名項目データセット  
ジャパンナレッジで公開する、平凡社刊行の『日本歴史地名大系』の地名項目と関連情報 (80,502件)

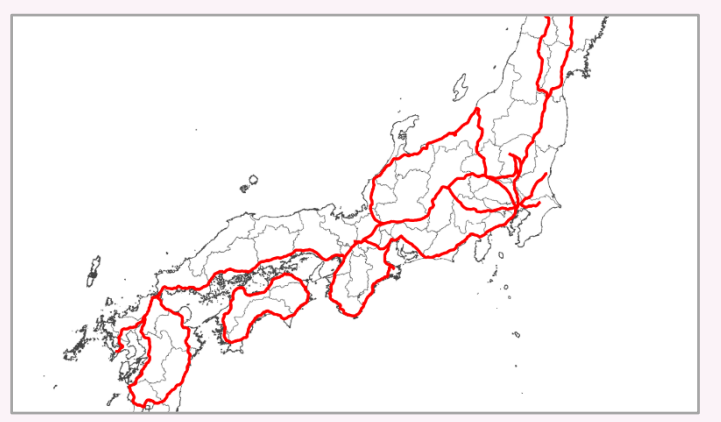


『日本歴史地名大系』施設・地点項目データセット  
ジャパンナレッジで公開する、平凡社刊行の『日本歴史地名大系』の施設・地点項目と関連情報 (32,038件)



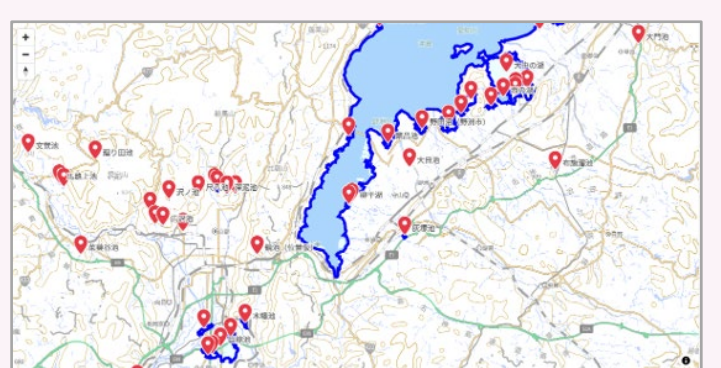
江戸主要街道データセット

江戸時代の主要街道(五街道および主要脇街道)の地理データ (約7700km)



江戸末期海岸線/水域/湖沼データセット

江戸末期の海岸線、川、湖、沼などの地理データ



データセットは全てオープンデータとして公開中！  
(CC BY 4.0ライセンスの元に提供)

<https://codh.rois.ac.jp/>



## 研究内容

### ①江戸マップ

古地図「江戸切絵図」から地名を抽出して地名データベース化し、  
歴史ビッグデータや歴史GIS(Geographic Information Systems)、  
江戸都市空間の地理情報基盤として活用

#### 課題：過去の地図の活用が困難

- 地図が歪んでいるため、現代の地図と重ねられない
- 膨大な地名の管理・検索・活用には地名IDが必要

#### 方法：29枚の地図（地名）の段階的な位置合わせ

- ジオレファレンスによる位置推定
- 「れきちず」を基盤地図とした手動修正
- 現代地図と比較した寺社等の精密修正
- 複数地図の相互比較による整合性確認

#### 成果：江戸全体の地名データベース構築

- 8,788件の地名をデジタル化、地名ID(識別子)を付与
- 江戸時代末期(1850年頃)の地名が検索・閲覧可能に

#### 展望：

- 他アプリとの連携  
例) edomiポータルの  
商店情報、旅行記とリンク

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



### ②DiHuCoガイドライン

講演の音声記録と講演資料を素材に、生成AIで記事を生成し、  
研究の実践的な知見を検索可能なテキストとして蓄積・共有する

#### 背景：

人文学研究のDX化の活用事例を集約し、人文データ規格の  
モデルガイドラインを策定する

#### 課題：研究実践の知見をどう集め、共有するか

- 論文には書かれにくい知見が、講演ではよく語られている
- 知見を共有する新しいガイドラインの仕組みが求められる

#### 方法：講演を素材に生成AI(Claude)で記事を生成

- プロジェクト機能で生成・レビューのワークフローを構築
- 記事の構成は固定せず、講演内容に応じて部品を取捨選択
- 個別事例から、共通パターン・論点へと段階的に展開

Use Case : 個別の実践を具体的な文脈ごとに記録

Best Practice : 複数のUCから共通パターンを帰納

Issues : パネル議論の論点を複数立場で整理

#### 展望：

- 今年度中にDiHuCo ガイドラインとして公開

<https://codh.rois.ac.jp/dihuco/>

