【基盤研究（S）】

史料データセンシングに基づく日本列島記憶継承モデルの確立

研究代表者	東京大学・史料編纂所・准教授 山田　太造（やまだ　たいぞう）　研究者番号：70413937
研究課題情報	課題番号：24H00011　研究期間：2024年度～2028年度 キーワード：日本史、史料、データインフラストラクチャ、データ駆動型、異分野融合

なぜこの研究を行おうと思ったのか（研究の背景・目的）

●研究の全体像

日本列島に起こったイベントを解明していく上で史料は重要な資源である。時代横断・地域横断といった観点で分析可能なデータとして史料を蓄積・管理・共有することで、政治・経済だけではなく、災害・気候・文化・土地利用など多様な分野に関わる過去のイベントを探ることができるが、これを実際に利用可能とするため、

（１）“史料データセンシング”手法の確立
（２）データ駆動型データ分析基盤の構築
（３）異分野融合研究環境の形成

について取り組む。最終的に、過去から現在を、多々の分野をシームレスにつなぐデータインフラストラクチャとして成長させ、日本列島記憶継承モデルの確立を目指す。

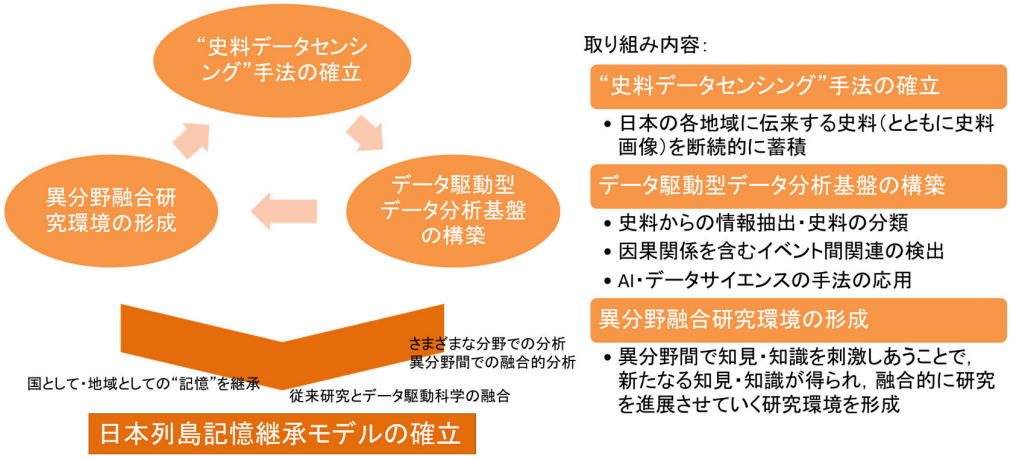

図1　本研究の概要

●本研究の特色

時代横断・地域横断といった観点で分析可能かつシームレスなデータとして史料を蓄積・管理していくことで、史料から、政治・経済だけではなく、災害・気候・文化・景観・土地利用など多様な分野に関わる過去の状況を、すくなからず復元し、分析しうると考えている。これまでの日本列島で起こったイベントを、日本史や人文学のみならず、自然科学の知見・知識をも踏まえた、異分野融合研究の実践が為せるならば、国として・地域としての“記憶” を継承していくことに繋がり、史料収集・蓄積に基づく日本列島記憶継承モデルを確立していくことができると確信している。本研究手法は、研究資源を史料画像として置き、そこから分析した結果もデジタルな環境にて利用していくことから、日本史研究過程のデジタル環境での実践、つまりは日本史研究DXの実現に他ならない。非常にクローズに思われがちな日本史研究を、他分野の研究者にも参入可能なオープンな研究環境として整備されていく可能性を秘めており、活発な議論・研究展開が可能になる。データ駆動、さらには異分野融合による史料分析を為すことで、史料画像のみではない、多角的視点に基づくデータへと成長していることから、史料そのものの高付加価値化へ繋がり、さらに、異分野融合研究に基づく研究結果をもとに次なる新たな学際研究へと繋がると考えている。

この研究によって何をどこまで明らかにしようとしているのか

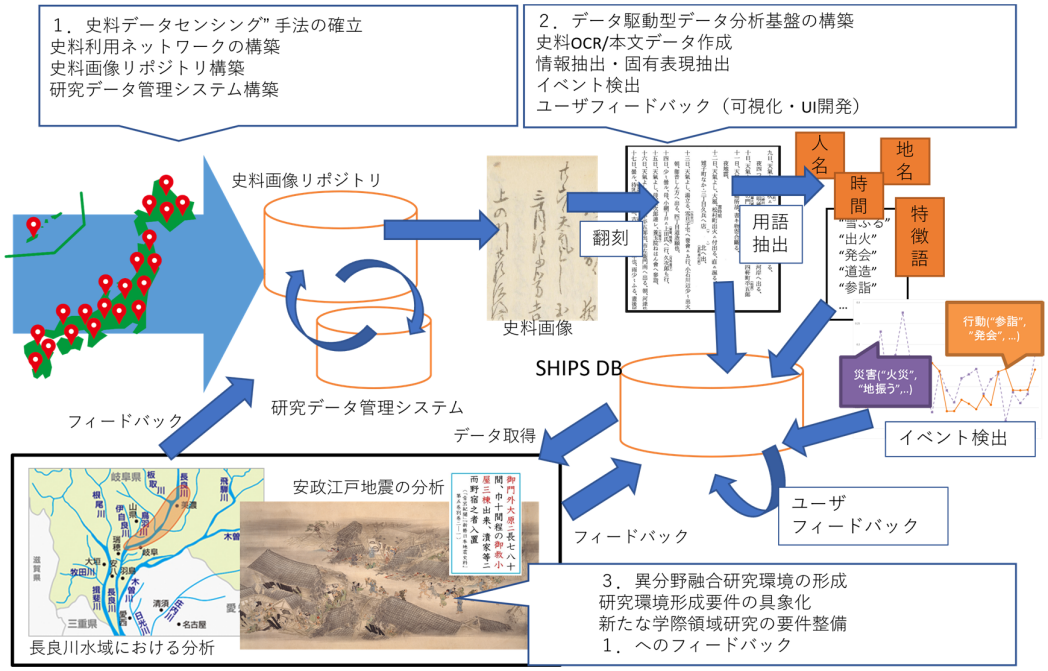

図2　研究計画

●計画（１）：“史料データセンシング”手法の確立

日本各地からデータセンシングのように史料画像データを次々と収集することができる“史料データセンシング”手法を確立する。

・史料利用ネットワークの構築：日本史史料の収集・拡充
・史料画像リポジトリの構築：日本列島の記憶となるスナップショットを集積していくデータリポジトリへ
・研究データ管理システムの構築：OAIS参照モデル（長期利用・長期保存の国際標準規格）に準拠、史料画像：由来・生成手法・利用条件・データの公開までの過程なども管理

●計画（２）：データ駆動型データ分析基盤の構築

2000万点を超える史料画像に対して、本文・人名・地名・時間といった史料を読解・理解する上で不可欠なデータを付与し、イベントとの関連付けを行うことで、 史料の性格を表現していく

・本文作成：史料OCRの実現：既存のデータセットやAI-OCRを応用しながら精度を高めていく。本文データの作成（TEI P5の応用と日本史史料用のタグの提案）
・情報抽出・固有表現抽出：深層学習による固有表現抽出手法を検討し実施。予備実験では近世後期の日記への適用でF値が0.7746
・イベント検出：トピックモデルを応用する（LDAなど教師なし学習やBERTを用いたトピック検出手法）。検出したイベント・トピックの知識表現を行う。可視化・UI開発とユーザフィードバックも行う。

●計画（３）：異分野融合研究環境の形成

・異分野融合を可能にする研究環境を形成：史料から取得しうるデータを多角的に分析し、異分野間の知見・知識が相互に刺激し合うデータ環境を明らかにしていく。
・ユースケース：地域横断・時間横断に係るデータを利用（歴史地震や環境動態解析など）。活用される/求められるデータやデータ間関連，研究過程や成果により新たに生成されるデータを格納・分析。
・新たな学際領域において必要な要件：検証を進め、新たなユースケースへ。異分野融合研究形成に不可欠となりうる要件を具象化。
・“異分野融合データファブリック”の実現へ：史料データ高付加価値化の実現

ホームページ等	https://lab.hi.u-tokyo.ac.jp/