

ROIS 戦略的研究プロジェクト 成果報告会
(第3期名称:機構間連携・文理融合プロジェクト)

研究課題名:
人文学研究データのオープンサイエンスにおける活用を最大化する情報基盤の構築

2024年3月28日

研究代表者

所属 国立情報学研究所

氏名 大波 純一

◆背景と研究目的

本プロジェクトでは、人文学分野における研究データや論文などの学術情報の利活用機会を、情報基盤を利用して最大化させるために、実際の基盤の状況を調査し必要なシステムを構築する。この中で、大学共同利用法人情報・システム研究機構（以下、ROISと記述）の国立情報学研究所（以下、NIIと記述）が有する学術情報基盤（主に検索基盤 CiNiiResearch、公開基盤WEKO3、及び人文学・社会科学総合データカタログJDCat）と、人間文化研究機構（以下、NIHUと記述）の有する人文学分野の学術情報を対象とする。最終的に**内部メタデータ、利用者アクセス情報、自然言語処理による解析結果などを機構間で交換し、相互に情報高度化の恩恵が得られる体制を構築**することを目的とする。

◆国内外の類似・競合する研究との関係

これまで日本国内では、NIIの電子リソースレポジトリREOでも「人文社会科学系データコレクション」だけが分離しているなど、分野ごとのデータが水と油のように分かれて情報流通が為されていた。しかし近年、ROISとNIHUが機構合同シンポジウム「人文知による情報と知の体系化」（2017年開催）で発表した数々の研究のように、**文理の部分ごとの融合と情報の交流が開始されている状況**である。

海外では、例えばヨーロッパではCO-OPERASと呼ばれるプロジェクトで「人社系データのFAIR原則の適用を進める」活動を推進しており、日本と同様に文理情報の流通強化に着手しつつある。欧州研究図書館協会（LIBER）でも「デジタル人文学」と呼称する人文学事業のデジタル化を推進するワーキンググループを2019年に発足させて活動を続けている。米国でもデジタル人文学に関する情報をまとめたサイト（The Digital Humanities Literacy Guidebook）をカーネギーメロン大学が構築するなど、**デジタル人文学の潮流はこの数年で活発な動き**を見せている。

また、新型コロナウイルスへの国際的な対応を見ても、ウイルスの生物学的な特徴を掴むだけでなく、情報工学によるデータを利用したシミュレーションや統計学的な分析、社会学に基づく人間の動態制御などが同時並行で実施されており、「様々な学術分野の情報が融合しやすい流通状態を準備する」ことは有効である。

◆ 本研究の位置付け

論文のオープン化や、オープンサイエンスの広まりによって、学術情報の流通が研究評価や社会への還元のための手段として周知されつつある。例えば研究で創出された学術データは、適切な分類やメタ情報が付与されて公開され、他の研究で利用されることで、データ駆動型の研究成果が次々と見出されている。オープンな環境は、他分野との学際研究での活用や国際的なコラボレーション、市民科学での利用などにも秀でており、研究活動の健全で発展的な循環をもたらすことができる。

しかし人文学では、研究規模が小さく属人的な研究となることが多いこと、研究対象となる資源が文化や言語など概念的である場合もあること、デジタル化するための技術情報が人文学から離れた情報学の知識を必要とすることから、オープン化をスムーズに進めるのが難しい。今後もこの状況が続くのであれば、日本の人文学研究は分野間や国際間、市民コミュニティとの繋がりを作り出すことが難しくなり、他の分野に比べて小規模で視野の狭いものだとは認識されてしまう恐れさえ生じる。それを回避するためには、**人文学の多様なデータがデジタル化され情報基盤へ適切な状態で格納されることが肝要である。**

本提案では、以下の項番で示す手段で人文学分野の研究データの高度化を実施し、人文学研究の価値を最大化する。この成果が人文学研究者のデジタル化やオープン化へのインセンティブとなり、研究活動の好ましい循環が開始することも期待できる。

1) 研究の概要	本プロジェクトでは、人文学の学術情報流通プロセスの現況調査を実施し、改善箇所や要求される技術の洗い出しを実施する。それに基づいて人文系研究データの処理方法のベストプラクティスを提案し、さらに実際の情報基盤における利用を提案する。 最終年度となる2023年度は、2022年度までの結果を元に、全学術分野への研究データ流通プロセス方法の見直しを実施する。見直しの状況に基づきNIIの基盤を修正する。また国立歴史民俗博物館、国際日本文化研究センター経由での流通データ拡張を模索し、ネットワークをより強固なものとする。			
2) 実施計画・実績	2021年度		2022年度	2023年度
	FS (Feasibility Study) ★ 6/19 FS採択審査会 ★ 3/18 FS評価審査会(本研究採択)		本研究 ★ 1年目実績評価	★ 10/17 2年目成果報告会 ★ 最終成果報告
	費用 (千円)	予算	1,000	3,000
	費用 (千円)	執行	980	2,980
実施者 (所属機関)	代表者: 大波 純一(国立情報学研究所・オープンサイエンス基盤研究センター) 共同研究者: 山地 一禎(情報学研究所・オープンサイエンス基盤研究センター) 林 正治(情報学研究所・オープンサイエンス基盤研究センター) 朝岡 誠(情報学研究所・オープンサイエンス基盤研究センター) 後藤 真(国立歴史民俗博物館) 関野 樹(国際日本文化研究センター) 南山 泰之(情報学研究所・オープンサイエンス基盤研究センター) 南山 泰之(情報学研究所・オープンサイエンス基盤研究センター・2022年度より参加)			
3) 研究成果の概要	FSの結果、海外でも明確なパスやリテラシが定まっていない、「既存の人文学研究データをどのように流通させるか」について、日本国内の既存基盤を利活用及び拡張させて解決することを目標とし以下の①～⑤の課題(① NIHUの統合検索システムとNII RDCの統合試験、② 汎用リポジトリ設置の構想、③ CiNii Researchにおける本文検索の実装テスト、④ 分野別メタデータスキーマの増強、⑤ 人文学研究の成果の最小単位での公開に関する検討)を実施した。さらに2022年度、基盤間相互運用におけるデータ品質向上のため新たに課題⑥(⑥ NII RDCの人文学基盤還元するデータの要件検討)を追加して実施した。			