

人文学・社会科学のデータインフラストラクチャーの 今後の方向性について (中間まとめ)

2022年8月29日

文部科学省

研究振興局 振興企画課 学術企画室

【背景】

○分野特性による程度の差はあるものの、人文学・社会科学においても広範なデータ収集と分析という研究手法は一般的なものとなっており、研究成果の質に直結する極めて重要な要素の一つである。

○他方、デジタル人文学などの取組が欧米で拡大しつつあるように、多様な資料や研究データのデジタル化とともに、データサイエンスを応用したデータ駆動型の新しい研究手法を取り入れることにより、これまで得られなかった学術的・社会的成果を人文学・社会科学研究が生み出すことに大きな期待が寄せられている。また、オープンサイエンスという視点から、研究の基盤となるデータの公開を通じた共同利用の促進に向けた動きも並行して進められている。

○しかしながら、日本の人文学・社会科学においては、一部の機関において資料やデータの整備・利活用の促進に取り組んでいるものの、全体的には研究データの利用環境整備は諸外国の後塵を拝している。特に、社会科学の諸分野においては共同利用可能なデータ拠点と言える体制整備が大幅に遅れており、国際共同研究の相手先としての日本の魅力が相対的に低下している要因にもなっているとの指摘がある。

【これまでの取組状況と課題】

○平成30年度より、日本学術振興会（JSPS）では人文学・社会科学の研究データを共有・利活用する基盤を構築することを目的に「人文学・社会科学データインフラストラクチャー構築推進事業」を実施している。

○この事業において、JSPSが国立情報学研究所（NII）の協力を得て事業全体を取りまとめ遂行する中核機関となり、データアーカイブ機能の整備及びデータの管理・利活用に関し中心的な役割を果たしてきた5つの拠点機関（人文学1、社会科学4を公募・選定）との運営体制を構築した。

○そして、拠点機関が提供するデータを一括検索できる人文学・社会科学総合データカタログ（JDCat）の構築、オンライン分析ツールの研究開発、データ共有のための手引きの策定、拠点機関における寄託受入れ体制の整備、拠点ホームページやメタデータの英語化などに取り組んできた。

○これまでの取組により、JDCatで公開されているメタデータは7,199件（2022/6/20現在）、JDCat新規ユーザー数は約6か月（2021/12/20～2022/6/17）で32か国3,874件に増加、各データへのアクセス数は1か月（2022/3）で36,771件、となっている。

○一方で、人文学分野は1機関の日本史史料のデータであり、JDCatがカバーできる範囲は限定的である。社会科学分野においても、重要度は高いものの家計パネル調査や若年パネル調査、総合的社会調査といった社会調査データと明治初期以降の日本統計年鑑及び公的統計の調査票様式・調査概要等に留まっている。

【今後の方向性（案）】

（人文学・社会科学総合データカタログ（JDCat）をはじめとした基盤の充実・強化）

○JDCatが国内外の研究者にとって有用な研究データの発見可能性を高めるツールとして研究者コミュニティ等に貢献するためには、取り扱うデータの一層の充実が求められる。このため、現在の5拠点以外の大学・機関に、どのようなデータがあり、どのようなニーズがあるのか、JSPSが行った令和元年（2019年）の調査（※）の分析等を踏まえ、次期（2023年～2027年）において必要な分野に関する、新たな拠点機関を検討することが必要である。加えて、データ間の連携や情報の発見・活用を促進するツール提供など、ユーザーの利便性を向上させる取り組みが期待される。

なお、より多くの機関に開かれたデータインフラへと拡充していく際には、クオリティの確保等、JDCatにデータを登録する際のデータの在り方やその方法について、さらに検討する必要がある。

（※）「人文学・社会科学系学術的調査データの現状に関する調査報告書」（日本学術振興会人文学・社会科学データインフラストラクチャー構築推進事業運営委員会作業部会）

○JDCatは現在、NIIが提供する日本最大規模の学術情報検索サービス「CiNii Research」と連携し、CiNii Researchで検索した情報からJDCatに移動して詳細なメタデータを見ることが可能となっている。このような仕組みを、例えば海外のデータアーカイブのカタログと連携することでデータ発見可能性が更に高まることが期待される。

（研究者がデータを共有し利用し合う文化の醸成とコミュニティの形成等）

○これまで、データの利用・公開のルール、権利関係等に関する「データ共有のための手引き」の策定、公開シンポジウムやニュースレターの配信等を通じて、研究者をはじめ学生等に対するデータを共有・利用し合う文化の醸成・コミュニティの形成に資する啓発活動を進めてきた。

今後は、データ共有のための手引きの更新に加え、実際のデータ構築のモデル例の作成やデータ共有に当たり留意すべき点について勉強会や研修会等による知識・技術の共有や普及、これまであまり関わりのなかった人文学・社会科学分野の専門家と図書館情報学や情報工学の専門家の交流が盛んになる取り組み、データ利活用に関するコミュニティの形成を促進する取組などを通じて、コミュニティ全体のレベルアップが図られることが期待される。

（運営体制）

○現在の人文学・社会科学のデータインフラストラクチャーは、各拠点機関が保有するデータのメタデータを作成し、中核機関が統合的に検索できるシステムであるJDCatを公開・運用するとともに、拠点機関との連絡・調整機能を担う運営体制となっている。

今後、本データインフラストラクチャーをより開かれた基盤としてデータの共有・利活用を促進していくためには、今まで以上に研究者コミュニティと連携した取り組みが望まれることから、大学・研究機関が連携し、ガイドラインの更新やシステムの改修などを相互に協力して実施するコンソーシアム形式による自律的な運営に移行することが望ましい。しかし、現時点では各大学・研究機関において十分な人的体制やシステム等が備わっていないことから、将来的なコンソーシアム形式の実現に向け、当面はデータアーカイブに係るノウハウや経験のある機関が中核機関となり新たな拠点機関を巻き込みながら、ネットワーク型として運営していくことが望ましい。

○その際には、中核機関と拠点機関の連携はもとより、拠点機関間でも積極的に情報共有やデータ間連携を行うことなど、密接な連携に努めることが必要である。加えて、基盤の定期的な点検とそれに伴うシステム改良を通じて、ユーザーに対する信頼性の確保に努めることも必要である。このためユーザーのニーズ等を確認できるような運営体制が望まれる。

○また、開かれたデータインフラストラクチャーの実現を目指し、拠点機関以外のデータのJDCatへのメタデータの登録に対応できる仕組みを検討することが望まれる。例えば、中核機関がメタデータの受入れ方針及び受入れ体制を整えるとともに、拠点機関がJDCatへのメタデータ登録等に必要な情報提供や技術支援ができる体制を構築することが考えられる。なお、受入れ方針等を検討するにあたってはユーザーのニーズやオープンサイエンスの動向などにも留意することが必要である。

○加えて、本データインフラストラクチャーのみならず、我が国の人文学・社会科学分野におけるデータの共有・利活用を更に発展させていくためには、データの長期保存や利活用に関する専門性を備えた人材の育成・活用が不可欠である。人文学・社会科学分野におけるデータ共有・利活用に必要な人材については、他分野等の動きを参照しつつ、在り方を検討することが必要である。