

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の
進捗と今後の展望（骨子案）

令和3（2021）年11月10日

独立行政法人日本学術振興会

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

（人文学・社会科学分野におけるデータインフラストラクチャーの課題等検討）

目次

1.	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の概要	2
1.1.	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の背景と基本的な考え方	2
1.1.1.	事業実施に至った背景	2
1.1.2.	事業の目標	2
1.1.3.	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	3
1.1.4.	ネットワーク型による人文学・社会科学データインフラストラクチャーの構築	3
1.2.	中核機関の取組	3
1.2.1.	手引きの策定と共通基盤の整備	3
1.2.2.	分野横断的な総合データカタログの整備	3
1.2.3.	オンライン分析システムの開発研究	3
1.2.4.	リサーチコモنزの形成	4
1.3.	拠点機関の取組	4
1.3.1.	データアーカイブ機能の強化（データ「共有化」）	4
1.3.2.	海外発信・連携機能の強化（データ「国際化」）	4
1.3.3.	データ間の時系列等接続関係の整備（データ「連結化」）	4
2.	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の活動実績	5
2.1.	体制	6
2.2.	成果	7
2.2.1.	人文学・社会科学総合データカタログ（JDCat）	7
2.2.2.	オンライン分析システム	10
2.2.3.	データ共有のための手引き	11
2.2.4.	拠点機関における取組	11
2.3.	データ利活用に関する研究者コミュニティの形成・文化の醸成	12
2.3.1.	講演・報告・執筆活動	12
2.3.2.	「JDCat サロン - データインフラの最前線」	13
3.	令和5（2023）年度以降の人文学・社会科学データインフラストラクチャーのあり方	14
3.1.	人文学・社会科学データインフラストラクチャーの拡充に向けた取組	14
3.2.	体制	15
3.3.	人材	16
3.4.	人文学・社会科学総合データカタログ（JDCat）	16
3.5.	オンライン分析システム	17
3.6.	データ共有のための手引き	17
3.7.	データ利活用に関する研究者コミュニティの形成・文化の醸成	17

1. 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の概要

概要

諸外国に比べて日本では、人文学・社会科学分野のデータインフラストラクチャーの整備が十分ではないため、日本学術振興会が平成 30 (2018) 年度より 5 年間の時限付きで、「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」を開始した。

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は、中核機関と拠点機関の連携協働によるネットワーク型で構築することとし、それぞれ以下の通り取組を行うこととした。

<中核機関の取組>

手引きの策定と共通基盤の整備／分野横断的な総合データカタログの整備／オンライン分析システムの開発研究／リサーチコモンスの形成

<拠点機関の取組>

データアーカイブ機能の強化（データ「共有化」）／海外発信・連携機能の強化（データ「国際化」）／データ間の時系列等接続関係の整備（データ「連結化」）

1.1. 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の背景と基本的な考え方

1.1.1. 事業実施に至った背景

人々の意識、状態、行動等を記す人文学・社会科学のデータは、客観的な証拠に基づく政策立案をはじめ社会の様々な意思決定などに不可欠な社会の共有資産であり、欧米のみならず、アジア諸国でも国家レベルでデータの国際的な共有・利活用の基盤が整備されている。

しかしながら、我が国では、諸外国に比べて人文学・社会科学分野におけるデータインフラストラクチャーの整備が十分ではなく、喫緊の対応が必要であった。

特に、社会科学系データに関しては、各研究機関等がそれぞれの保有するデータを管理するに留まっており、一元的に収集・保管・提供・活用する組織・機関はなかった。

このため、独立行政法人日本学術振興会（以下、「学振」という）は、人文学・社会科学のデータが組織・機関を越えて共有・利活用できるよう推進するためのキックオフとして、人文学・社会科学データインフラストラクチャー（以下、「人社データインフラ」という）を構築することとし、平成 30 (2018) 年度より 5 年間の時限付きで、「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」（以下、「本事業」という）を開始した。

1.1.2. 事業の目標

本事業は、人文学・社会科学のデータを体系的に収集・管理し、国際的な連携・発信を強化する基盤を整備することにより、分野・国境を越えて研究者がデータを広く共有し利用し合う文化を醸成していくことを目標とした。これにより、我が国の人文学・社会科学の研究水準が大きく向上するとともに、我が国との国際共同研究や交流が増え、もって人文学・社会科学全体が発展することが期待される。

1.1.3. 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

本事業の開始にあたり、平成 30（2018）年 5 月から 6 月にかけて「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¹」を実施した（1,892 の研究機関を対象に、各研究機関に所属する研究者への協力を依頼し、3,956 名の人文学・社会科学分野の研究者から回答を得た）。人文学・社会科学分野における学術的調査データの現状について情報を収集する中で、データの公開には慎重な意見が多い一方、データの利用にはきわめて積極的な姿勢が示され、公開されたデータを使いたいという研究者のニーズは大きいことが明確になった（他研究者による公開データや公開資料については 50.8%が「自分の研究に必要ななら積極的に利用したい」、43.4%が「場合によっては利用することもありうる」と回答）。

1.1.4. ネットワーク型による人文学・社会科学データインフラストラクチャーの構築

新たに人社データインフラを構築するにあたっては、中核機関と複数の拠点機関との強力な連携協働によるネットワーク型によって基盤を整備していく方向性を打ち出した。方向性を検討する段階で、人文学・社会科学分野全体を網羅するナショナルセンターはないものの、一部の分野で大学等の研究者グループを核に拠点的取組が行われており、このような拠点の中には、文部科学大臣による共同利用・共同研究拠点の認定を受ける研究機関も存在した。このため、分野毎に拠点機関の基盤を強化しデータの共有化に取り組みながら、分野横断的な体制を構築するネットワーク型の基盤整備が妥当であると判断した。

1.2. 中核機関の取組

国内外に開かれた分野横断的なデータ利活用のシステムを構築するため、拠点機関の間の強力な連携協働により、以下の取組を行うこととした。

1.2.1. 手引きの策定と共通基盤の整備

- ・ 人文学・社会科学における分野横断的なデータの共有・利活用を推進するため、メタデータやフォーマットの基準、データの利用、公開のルール、秘匿処理、権利関係等に関する手引きを策定する。

1.2.2. 分野横断的な総合データカタログの整備

- ・ 人文学・社会科学に関し、データの所在、所有者、データのタイプ等のメタデータを収録した分野横断的な総合データカタログを日英両言語に対応して整備する。

1.2.3. オンライン分析システムの開発研究

- ・ 異なるデータアーカイブ間でデータを共有するシステム、オンラインのリモート分析を可能にする

¹ 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，2019，「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」，日本学術振興会ホームページ，（2021 年 9 月 8 日取得，https://www.jsps.go.jp/j-di/data/datainfra_houkokusyo.pdf）。

【機密性 2 情報】

るシステムの開発を推進する。

1.2.4. リサーチコモンズの形成

- 分野・国境を越えて研究者がデータを広く共有し利用し合う文化の醸成を目指し、シンポジウムの開催やニュースレターの配信等による啓発を行う。

1.3. 拠点機関の取組

人文学・社会科学分野におけるデータの管理・利活用に中心的な役割を担っている研究機関を核に、分野毎に拠点を形成し、相互に連携しながら、以下の取組を行うこととした。

1.3.1. データアーカイブ機能の強化（データ「共有化」）

- 研究機関、学協会からのヒアリング、あるいは公開情報等から、データの所在、所有者、管理状況等を把握し、メタデータを整備する。
- データの散逸等を防ぎ利活用を促進するため、寄託を受けたデータも対象としつつ、優先度の高いデータについて、データクリーニング、欠測値補完、論文との紐付け可能な書誌符号（DOI）付与等を施したアーカイブ化を行う。

1.3.2. 海外発信・連携機能の強化（データ「国際化」）

- 国際的なデータの利活用に資するため、拠点ホームページ、メタデータ、調査票、ラベル等を英語化する。
- 国際共同研究として収集されたデータを、国際的な学術利用のために多言語化する。
- アーカイブの機能強化のため、海外データアーカイブ²諸機関と連携し、研究交流を推進し、シンポジウム等を開催するとともに、もって若手研究者の育成の強化に資する。

1.3.3. データ間の時系列等接続関係の整備（データ「連結化」）

データを利活用した人文学・社会科学研究及び教育利用を推進するため、大規模社会調査及び公的統計それぞれの時系列等接続をしたデータベースの整備や、連結可能なデータとするための編集・加工等を行う。

² 本稿ではデータアーカイブとは、主に研究・教育を目的として、幅広いデータを収集・保存し、提供するサービスのことをいう。

2. 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の活動実績

概要

本事業は、運営委員会、学振と国立情報学研究所からなる中核機関、公募により選定された5つの拠点機関で取り組んでいる。

中核機関においては、学振と国立情報学研究所が人文学・社会科学総合データカタログ（JDCat）及びオンライン分析システムを構築している。また、学振はデータ共有のための手引きの策定、データ利活用のコミュニティ・文化の醸成も実施している。

拠点機関においては、各拠点機関が保有するデータ及び外部から寄託されたデータを対象に、①データアーカイブ機能の強化（共有化）、②海外発信・連携機能の強化（国際化）、③〔人文学〕データ間の連携を可能にする環境の整備／〔社会科学〕データ間の時系列等接続関係の整備（連結化）に取り組んでいる。令和2（2020）年度に運営委員会によって実施された拠点機関の中間評価では、高い評価を受けている。

本事業の取組内容は図1の通りである。

これらの取組を着実に実施することで、人学データインフラの構築を推進する。

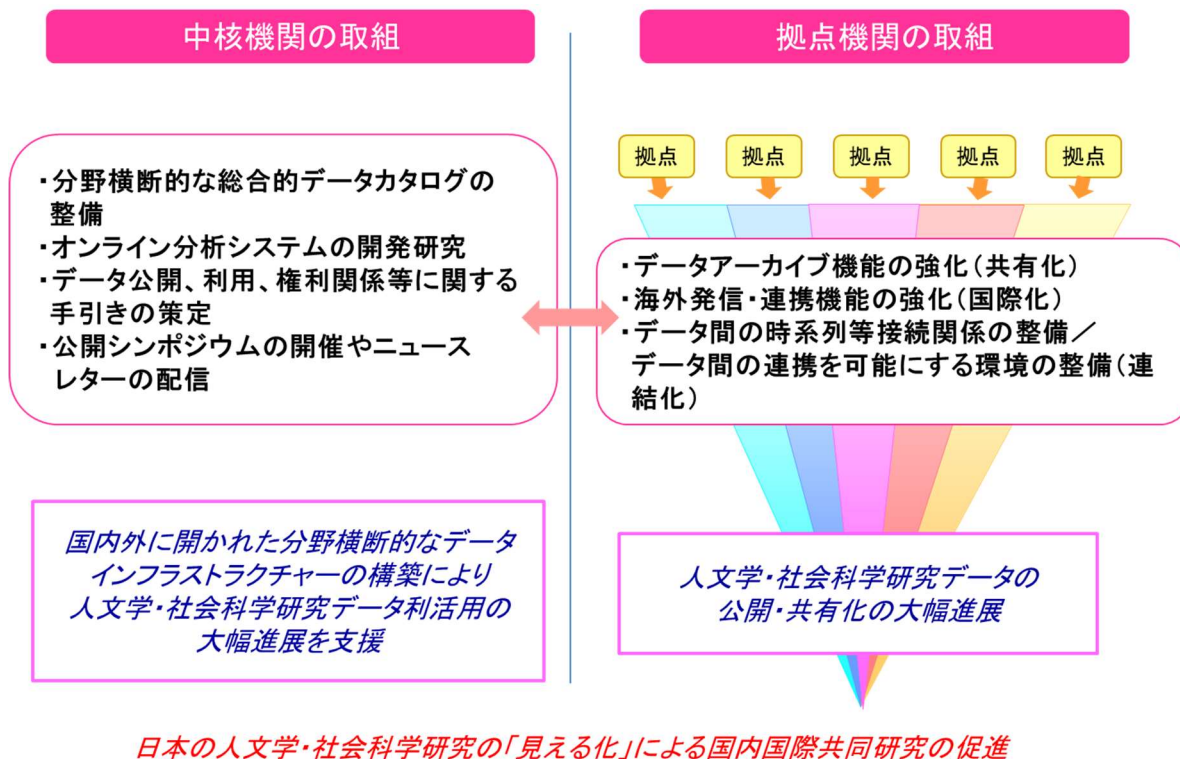


図1：本事業の取組内容

【機密性 2 情報】

2.1. 体制

本事業は、運営委員会、中核機関、拠点機関からなる（図2）。

運営委員会は、本事業の基本方針、計画の策定、拠点機関の審査・評価等を行う。また、必要に応じて作業部会を設置することとしている。これまでに設置した作業部会は、以下の通りである。

- ① 仕様等検討のための作業部会（平成 30（2018）年度）
- ② データ保有状況調査のための作業部会（平成 30（2018）年度）
- ③ 共通ガイドライン（手引き）策定のための作業部会（令和元（2019）年度～令和 2（2020）年度）
- ④ 人文学・社会科学分野におけるデータインフラストラクチャーの課題等検討のための作業部会（令和 3（2021）年度）

中核機関は、学振と国立情報学研究所（以下、「NII」という）が担う。学振は、平成 30（2018）年 11 月に人文学・社会科学データインフラストラクチャー構築推進センター（以下、「センター」という）を設置し、NII とデータインフラストラクチャーシステムの共同開発を行うこととなった。センターは、センター長（PD）1 名、研究員（PO）3～4 名で構成されるが、各員は、それぞれ所属大学等との兼務で本業務に携わっている。人文学・社会科学総合データカタログ（JDCat）等の詳細な仕様決定、拠点機関との連携調整、システム開発の進捗管理等を行っている。

拠点機関については、平成 30（2018）年度から令和元（2019）年度にかけて、公募により以下の 5 機関を選定した。

- ・ 大阪商業大学 JGSS 研究センター
- ・ 慶應義塾大学 経済学部附属経済研究所パネルデータ設計・解析センター
- ・ 東京大学 社会科学研究所附属社会調査・データアーカイブ研究センター
- ・ 一橋大学 経済研究所
- ・ 東京大学 史料編纂所

なお、拠点機関においては、必要に応じて以下の専門職員等を本事業の委託費で配置できることとしている。

- ① データ共有基盤の強化に当たり、データ・アーカイブ・マネージャー、データ・アーキビスト、データ・ライブラリアンなど、研究者のデータ利用を支援する者
- ② 研究員や専門職員など、調査実施機関においてデータに付加価値をつけることやデータを国際化、高度化するために必要な者
- ③ IT 技術者など、高度なデータ利用環境の整備・構築を支援する者

拠点機関は、データアーカイブに関する取組のほか、人文学・社会科学総合データカタログ（JDCat）（以下、「JDCat」という）との連携を中核機関と協力して行う。

また、中核機関と 5 拠点機関は定期的に開催する連絡協議会を通じて、事業の具体的な進め方等に係る協議や調整を行う。

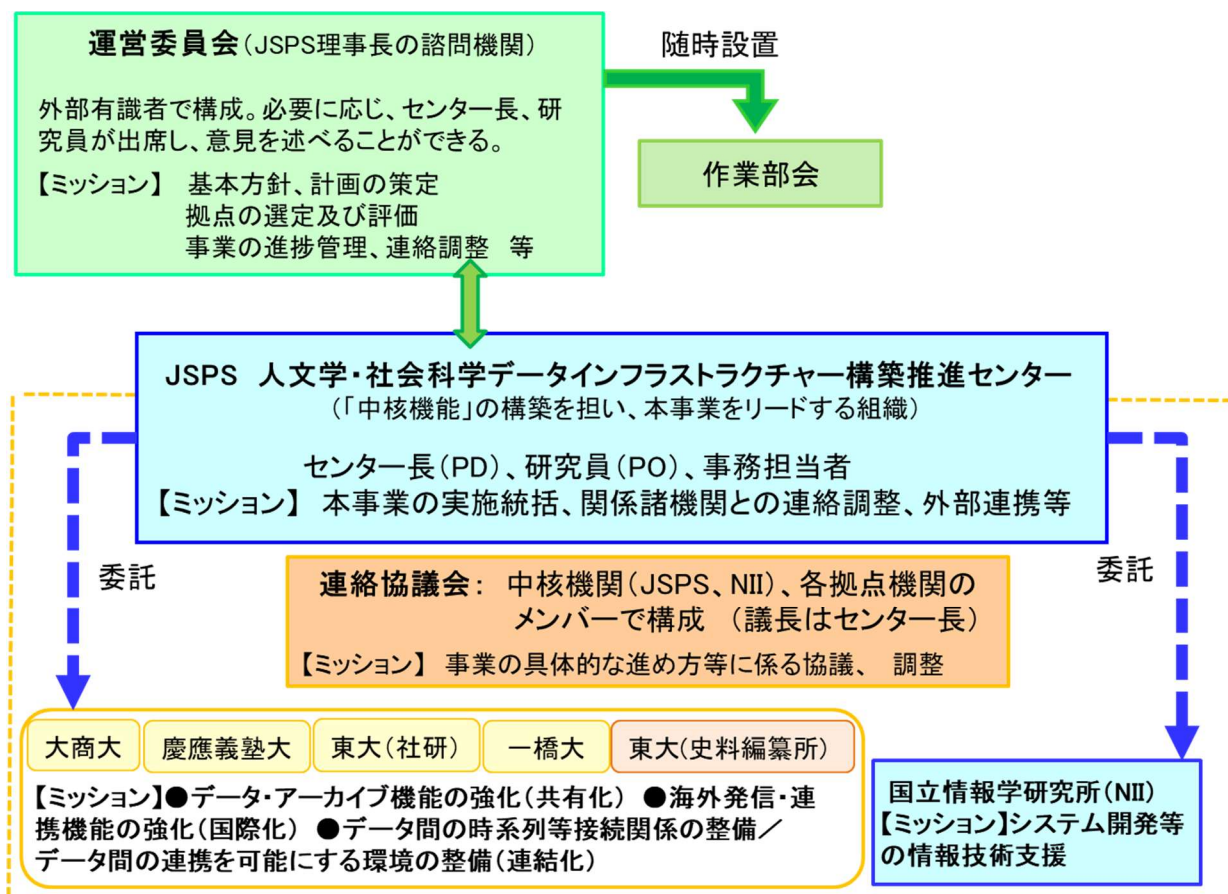


図 2：本事業の体制

2.2. 成果

2.2.1. 人文学・社会科学総合データカタログ (JDCat)

JDCat は、拠点機関が提供するデータのメタデータ³を自動収集して収録したカタログで、FAIR 原則⁴に沿って整備されており、人文学分野のデータと社会科学分野のデータを分野横断的に一括検索することが可能である (図3)。また、日本語・英語でのメタデータ整備により、海外の研究者が日本のデータを検索することも可能である。さらに、メタデータをクリエイティブ・コモンズ (CC) ライセンスの CC0⁵で公開することにより、海外からのメタデータの自動収集が容易となっている。なお、NII の提

³ データを説明する上で必要な情報 (作成者、作成時期、データの概要など) のこと。

⁴ データを共有するための基準となる国際的な原則 (FAIR Data Principles)。「FAIR」は「Findable(見つけられる)」「Accessible(アクセスできる)」「Interoperable(相互運用できる)」「Re-usable(再利用できる)」の略で、データ公開の適切な実施方法を表現しており、本原則に準拠したデータを作成する機運が国際的に高まっている。詳細は「データ共有の基準としての FAIR 原則」<https://DOI.org/10.18908/a.2018041901> を参照。

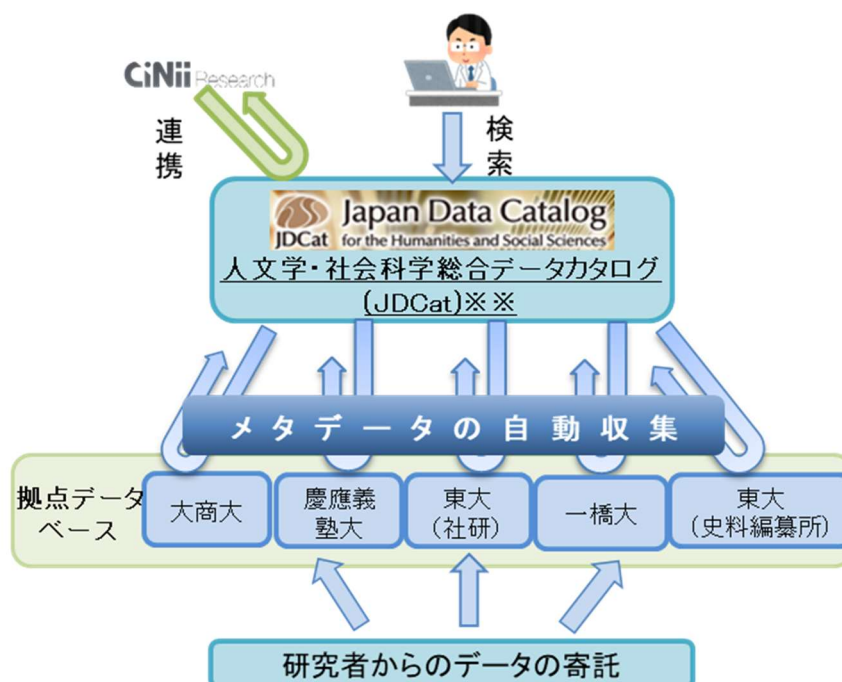
⁵ クリエイティブ・コモンズ (Creative Commons, CC) が提案し、世界的に流通している利用許諾条件 (ライセン

【機密性 2 情報】

供する検索サービス「CiNii Research⁶」との連携により、学术论文や図書、研究データ等の研究成果情報、それらの成果を生み出した研究者、研究プロジェクト等の情報と関連付けることが可能である。

利点としては以下がある：

- ・ 多様なデータアーカイブを一括で検索することにより、自身の研究に必要なデータを効率的に探すことができる。
- ・ 拠点機関にデータを寄託することで、JDCat にメタデータが掲載され、自身の研究データの共有ができる。



※※NIIが開発したリポジトリソフトウェアWEKO3を利用して構築

図 3：JDCat の構成

2.2.1.1. JDCat メタデータスキーマ

JDCat メタデータスキーマは、人文学・社会科学分野において国内外で標準的に使用されているメタ

ス) のルールに基づく表記。CC ライセンスは、著作権を保持しつつ、一定の条件下で著作物の自由な流通と再利用を可能にするルールである。なかでも CC0 (Creative Commons licenses zero) は、著作物に関するすべての権利を法的に可能な限り放棄することを示すものであり、誰もがどのような形でも使用できることになる。詳細はクリエイティブ・コモンズ・ジャパン Web サイト：<https://creativecommons.jp/licenses/>を参照。

⁶ 国立情報学研究所 (NII) が提供する、日本最大規模の学術情報検索サービス。WEKO3 に登録された研究成果や論文情報のみならず、図書、研究データ、それらの成果を生み出した研究者、そして研究プロジェクトの情報などを包括して探索することが可能である。

【機密性 2 情報】

データスキーマや、海外のデータアーカイブ等の現状の調査を経て、31 項目からなるスキーマとした。これらの項目は人文学・社会科学両分野をカバーし、かつ海外からのハーベストも視野に入れた項目となっている。

また、データへの DOI の付与を推奨することにより、データの特定が容易となっている。

さらに、社会調査メタデータの国際規格である Data Documentation Initiative (DDI) ⁷、及び JPCOAR スキーマ⁸とマッピングすることにより、国内外の相互運用性に優れている。

2.2.1.2. 統制語彙

JDCat における統制語彙は、海外で使用されている統制語彙の調査を経て、9 項目、12 種類を翻訳・作成した。「トピック」は CESSDA Topic Classification⁹を、「観察単位」「サンプリング方法」「調査方法」については、DDI Controlled Vocabularies¹⁰を翻訳した。

また、「データタイプ」等は、分類の必要上、独自の統制語彙を作成した。

2.2.1.3. NII が開発した WEKO3 を利用したメタデータのハーベストと一元的検索

JDCat は、中核機関を担う NII が開発したリポジトリソフトウェア WEKO3¹¹を利用して構築している。拠点機関からのメタデータのハーベストについては、OAI-PMH¹²または Resource Sync¹³でのハーベストとしている。これにより、各拠点機関のデータ整合性の確認が容易となっている。JAIRO Cloud¹⁴あるいは WEKO3 を用いてメタデータを登録している拠点機関においては、システム改修は不要であったが、東京大学社会科学研究所附属社会調査・データアーカイブ研究センターでは、自機関のシステムに OAI-PMH によるハーベスト機能を実装する必要があった。

⁷ 社会科学、行動科学、経済科学、健康科学の分野における調査や実験などの観察によって得られたデータを記述するための国際標準メタデータ規格。ICPSR や UKDS などの諸外国のデータアーカイブで採用されており、検索システムの運用などで活用されている。詳細は <http://www.ddialliance.org/>を参照。

⁸ オープンアクセスリポジトリ推進協会 (JPCOAR) が策定した新しいメタデータ規格。日本の機関リポジトリのメタデータの国際的な相互運用性を向上させ、日本の学術的成果の円滑な流通を図ることを目的としている。詳細は <https://schema.irdb.nii.ac.jp/ja>を参照。

⁹ 欧州の社会科学データアーカイブコンソーシアムである CESSDA (the Consortium of European Social Science Data Archives) が作成し公開している統制語彙 (CESSDA Vocabulary Service) の一種。Topic Classification はデータの主要なテーマを示すための統制語彙である。詳細は <https://vocabularies.cessda.eu/>を参照。

¹⁰ DDI Alliance (the Data Documentation Initiative Alliance) が作成し公開している統制語彙。詳細は <https://ddialliance.org/controlled-vocabularies>を参照。

¹¹ 国立情報学研究所 (NII) が提供する、研究データや関連の資料を公開するためのデータ公開基盤。

¹² OAI (Open Archives Initiative) が策定したメタデータ収集のプロトコル (Open Archives Initiative Protocol for Metadata Harvesting)。

¹³ OAI-PMH の後継規格として策定が進められているメタデータ収集のプロトコル。

¹⁴ 国立情報学研究所 (NII) とオープンアクセスリポジトリ推進協会 (JPCOAR) が共同で運営するクラウド型の機関リポジトリ環境提供サービス。

【機密性 2 情報】

2.2.1.4. 機能

JDCat の機能等の決定にあたっては、利用者の利便性の高い検索方法、表示順等を検討し、以下のような機能を実装している。

【JDCat に実装している機能】

- ① 各拠点機関から提供を受けるデータのメタデータを自動的に収集する
- ② DDI に準じたメタデータを収集する
- ③ 拠点機関のメタデータを収集するだけでなく、別途、同類のデータとメタデータを登録できるインターフェースを有する
- ④ 将来的な展開を見込み、DDI 以外のメタデータスキーマにも拡張できる
- ⑤ 利用統計情報を管理者にレポートする
- ⑥ 持続的な運用を見据え、オープンソースソフトウェアで構成する

【ユーザーが利用できる機能】

- ① メタデータを一括検索できる
- ② キーワード検索に加え、DDI で規定される各メタデータ項目による詳細検索ができる
- ③ 各拠点機関のデータカタログのみをコレクションとして表示できる
- ④ CiNii Research でも JDCat メタデータを含めた横断検索ができる

2.2.2. オンライン分析システム

オンライン分析システムは、統計ソフトをインストールしたりデータを手元にダウンロードしたりすることなしに、R や Python のプログラムを作成・実行しデータを分析できるシステムである。分析プログラムや分析結果を出力し、公開・共有を行うことも可能である。

利点としては以下がある：

- ・ 自身でデータ分析に係る環境を構築する必要なく研究を実施できる。
- ・ 大学の講義、ゼミ等において、教員は学生とデータ及びプログラムを共有し、指導を行うことができる。
- ・ 研究室において、データやプログラム等を含む分析環境を保存し、研究の再現性と継続性を担保することができる。

社会科学分野においては令和 2（2020）年 10 月から大学の講義等にて試験運用を開始している。人文学分野においては令和 3（2021）年度中に大学の講義等にて試験運用を開始予定である。なお、本事業期間終了後は NII において運用を継続予定である。

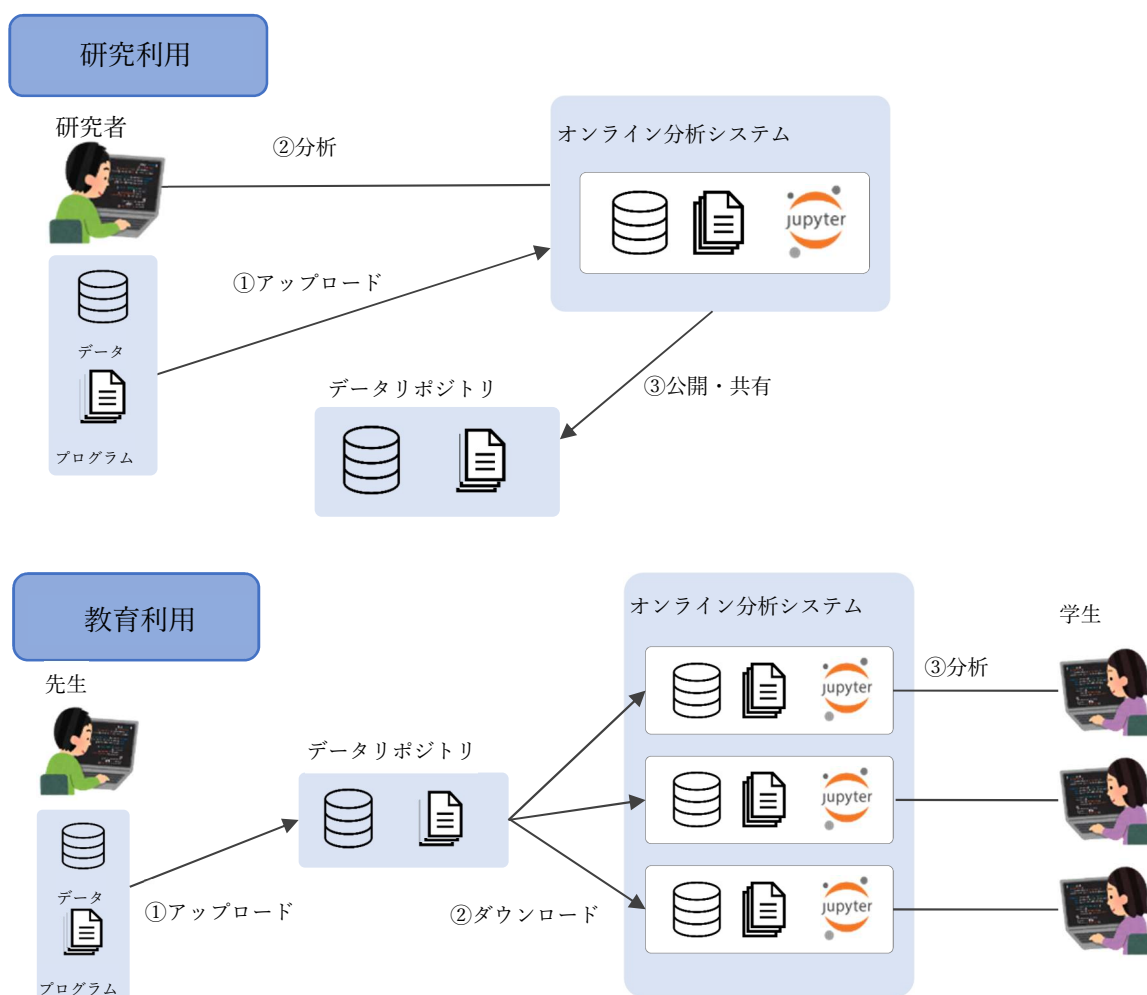


図 4：オンライン分析システムの活用例

2.2.3. データ共有のための手引き

これまで、人文学・社会科学分野におけるデータ共有のための手引きで日本において策定されたものはなく、海外機関のガイドラインを参考にしていた。そこで、令和元（2019）年度に運営委員会のもとにデータ共有のための手引き策定に係る作業部会を設置し、令和 2（2020）年度にかけて「人文学・社会科学におけるデータ共有のための手引き」（以下、「手引き」という）を策定した（現在、公開準備中）。なお、手引きの構成は、海外の複数の機関（例：UKDA、FSD）のガイドラインを参考にしている。

また、令和 2（2020）年 11 月～12 月にかけて、関連学協会等へ意見照会を実施し、寄せられた意見を検討した結果を踏まえて策定した。

2.2.4. 拠点機関における取組

社会科学分野の拠点機関については平成 30（2018）年度に 4 機関を、人文学分野の拠点機関については令和元（2019）年度に 1 機関を公募により選定した。

【機密性 2 情報】

拠点機関は、これまでデータの管理・利活用に中心的な役割を果たしてきた実績を活かして、広く人文学・社会科学系研究者の研究基盤となりえるデータインフラストラクチャーの構築の重要拠点として業務を実施できる研究機関を選定した。

5 拠点機関とその保有データ（JDCat でメタデータを順次公開予定）は以下の通りである：

【社会科学分野】

- 大阪商業大学 JGSS 研究センター
日本版総合的社会調査（Japanese General Social Surveys：JGSS）、台湾・韓国・中国との共同研究「東アジア社会調査プロジェクト（East Asian Social Survey：EASS）」
- 慶應義塾大学 経済学部附属経済研究所パネルデータ設計・解析センター
家計や企業の意識や行動を長期にわたり追跡調査できるパネルデータ
- 東京大学 社会科学研究所附属社会調査・データアーカイブ研究センター
1950年代から現在に至るまでの幅広い社会調査のマイクロデータ
- 一橋大学 経済研究所
明治初期以降の日本統計年鑑、および政府統計の調査票様式ほか、調査概要（根拠となる法令・通達等を含む）、各種分類表、長期経済統計

【人文学分野】

- 東京大学 史料編纂所
明治維新までの国内外の幅広い日本史史料

なお、5 拠点機関は、自機関の保有するデータだけでなく、外部の研究者や組織が作成したデータの寄託を受け入れている、あるいは受け入れ予定である。

また、令和 2（2020）年度に、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運営委員会において、5 拠点機関の中間評価を実施し、高い評価を受けている。

2.3. データ利活用に関する研究者コミュニティの形成・文化の醸成

当初の構想で中核機関の取組であった「リサーチ commons の形成」については、より目的に資するように「データ利活用に関する研究者コミュニティの形成・文化の醸成」とした。分野・国境を越えて研究者がデータを広く共有し利用し合う文化の醸成を目指し、講演・執筆活動等による啓発に取り組んでいる。

また、これまであまり関わりのなかった人文学・社会科学分野のドメインの専門家と、図書館情報学・情報工学の専門家の交流が盛んになるよう取り組み、データ利活用のコミュニティの形成を促している。

2.3.1. 講演・報告・執筆活動

2.3.1.1. 講演・報告

- 2019 年 1 月 第 4 回 SPARC Japan セミナー 2018 「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の取組について」

【機密性 2 情報】

- 2019 年 3 月 International Cooperation Seminar on Metadata, Data Catalogues and Tools for Findability "JSPS's Plan for a Federated Data Catalog in Humanities and Social Sciences" "Current Situation of Data Archiving for Japanese Official Statistics" (フィンランド・FSD)
- 2019 年 7 月 地域科学研究会 研究・実験データの保管・共有の推進方策 II ～研究データ管理のプラットフォーム／研究者・大学の取組と責務～「人文学・社会科学系データの共有・活用の制度基盤～データインフラストラクチャー構築推進事業の取組みと実際～」
- 2019 年 9 月 The Digital Transformation -Implications for the Social Sciences and the Humanities "Sharing Social Science Data in Japan: JSPS's Plan for a Federated Data Catalog" (DIJ in Tokyo)
- 2019 年 10 月 "The Past, Present, and Future of Social Science Data Preservation and Dissemination in Japan" (ミシガン大学図書館)
- 2019 年 11 月 三田図書館・情報学会研究大会ラウンドテーブル
- 2020 年 9 月 2020 年度統計関連学会連合大会企画セッション
- 2021 年 2 月 2020 年度経済統計学会全国研究大会報告集 (大会中止に伴い、報告集での報告)
- 2021 年 2 月 東京通信大学 情報マネジメント・フォーラム「社会調査データの保存と共有について 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の紹介を兼ねて」
- 2021 年 9 月 2021 年度統計関連学会連合大会企画セッション

2.3.1.2. 執筆

- 2019 年 6 月 中央調査報 (No.740)
- 2019 年 7 月 中央調査報 (No.741)
- 2019 年 11 月 ESTRELA (No.308)
- 2020 年 "Sharing Social Science Data in Japan: JSPS's Plan for a Federated Data Catalog." In The Digital Transformation: Implications for the Social Sciences and the Humanities, edited by Harald Kümmerle and Franz Waldenberger, 27-31. Tokyo: German Institute for Japanese Studies.
- 2021 年 1 月 日本統計学会会報 (No.186)

2.3.2. 「JDCat サロン - データインフラの最前線」

「JDCat サロン - データインフラの最前線¹⁵⁾」では、主にデータインフラストラクチャーの現場を支える若手研究者等とサロン形式で懇談し、懇談内容を学振のホームページに掲載することで、日本のデータアーカイブの現状等についての情報発信を行っている。研究者がデータを共有しあう文化を醸成するとともに、国内外の共同研究等を促進し、もって人文学・社会科学の振興を図るという事業目的に資するべく、廣松センター長の発案により令和 3 (2021) 年 3 月から実施している。

¹⁵⁾ 日本学術振興会、2021、「JDCat サロン - データインフラの最前線」、日本学術振興会ホームページ、(2021 年 9 月 8 日取得、<https://www.jsps.go.jp/j-di/jdcatsalon.html>)。

3. 令和5（2023）年度以降の人文科学・社会科学データインフラストラクチャーのあり方

概要

本事業において人文科学分野のデータと社会科学分野のデータを一括検索できるデータカタログを構築したことは有意義である。今後、人文科学分野については、取り扱う分野を拡充していくこと、また、社会科学分野についても取り扱うデータの一層の充実が望ましい。

現在の我が国の状況に鑑みれば、人事データインフラをネットワーク型で形成することは妥当であるが、構築した人事データインフラの安定的、持続的な運用を確保できるよう引き続き考えていくことが重要である。人事データインフラの拡充を、人文科学・社会科学分野においてデータの蓄積や提供に貢献している機関が連携することによって進めるには、拠点機関だけに閉じたネットワークから、他機関の積極的な参画を促す開かれたネットワークへと、段階的に移行していくことが望ましい。

また、ネットワーク型の人事データインフラでは、全体的な運用方針を定めたり拠点機関との相互調整を行ったりする機能を担う中核機関を置くことが必要である。この中核機関の果たす役割は極めて重要であるため、しかるべき専門機関において長期にわたり安定した運営ができるよう、国において制度設計を行っていく必要がある。

研究データ管理・共有・利活用にあたる人材については、第6期「科学技術・イノベーション基本計画」を踏まえ、国レベルでの検討が始まっている。人事データインフラに必要な人材については、これらの動きを参照しつつ検討することが必要ではないか。

本事業の成果である、人文科学・社会科学総合データカタログ（JDCat）、オンライン分析システム、データ共有のための手引きは、事業期間終了後も、それぞれ適切な方法等により、管理・運用されることが期待される。

データ利活用のコミュニティについては、研究機関や学会、研究者と実務者が相互に協力して形成していくべきである。また、研究においてデータを扱う専門人材が研究の推進には不可欠であるという理解を研究者コミュニティの中で広めていくことも重要である。

3.1. 人文科学・社会科学データインフラストラクチャーの拡充に向けた取組

人文科学・社会科学データインフラストラクチャーに関する諸外国の現状と比較すれば日本の整備状況は不十分で、喫緊の対応が必要であった。そうした状況を鑑みれば、本事業において人文科学分野と社会科学分野を同時に対象とし、人文科学分野のデータと社会科学分野のデータを一括検索できるデータカタログ（JDCat）を構築したことは有意義である。一方、本取組においては、人文科学と社会科学で必要なレコードの単位やメタデータの粒度にばらつきがあること等が分かってきた。

本事業の拠点機関は、人文科学分野1、社会科学分野4である。今後、人文科学分野については、取り扱う分野を拡充していくことが望ましい。また、社会科学分野についても、取り扱うデータの一層の充実が望ましい。このためには、拠点機関の数の拡充もさることながら、各拠点機関における役割の拡充・強化も求められるところである。

究極的には「全ての人文科学・社会科学」が横断的にデータを共有・利活用できることが理想である。

【機密性 2 情報】

そのために、データアーカイブに係るノウハウと実績のある機関同士の連携だけでなく、データアーカイブに興味・関心のある他機関も積極的に参画できる体制づくりが必要である。

併せて、データの寄託を促す取組を研究者コミュニティのレベルで実施していくことも重要である。具体的な取組としては、データアーカイブに係るノウハウと実績のある機関の研究者を中心に分野毎のネットワークを形成することやデータ公開に関する勉強会・研修会といったサポート活動等を通じて研究者コミュニティ全体のレベルアップを進めていくことが考えられる。そのうえで、公共財としてオープンで再現可能な形で研究データを作成・共有・利活用する文化を醸成していくことが大切である。

ただし、データアーカイブの管理運営を各機関の自助努力にまったく委ねてしまうことは避けるべきである。研究データを長期にわたって保存、管理することの困難性は各機関が共通に抱えている課題であり、国の責任において解決策を模索すべきと考える。

3.2. 体制

本事業では、学振のパイロット事業という位置づけで、学振が自ら人社データインフラ構築作業を担っており、ファンディング・エージェンシー（以下、「FA」という）が事業運営そのものに深く関与する形になっている。人社データインフラに関する諸外国の事例では、資金提供する FA はいくつかあるものの、運営自体を行っている FA はなく、本事業運営形態は、極めて特殊なケースであると言える。実際、FA である学振の組織構造、人員体制等に鑑みれば、学振が本事業を長期的に運営し続けることは難しく、今後、人社データインフラの強化を図っていくには、しかるべき専門機関において長期にわたり安定した運営ができるよう、国において制度設計を行っていく必要があるものと考えられる。

本事業では、人社データインフラを関連機関が協力するネットワーク型で形成することとしたが、現在の我が国の状況に鑑みれば、この判断は妥当であったと考えている。ネットワーク型のメリットとして、異なる専門分野のデータを統合的に扱えることが挙げられる。しかし、ネットワークに参画する各拠点機関におけるデータアーカイブの安定的、持続的な運用が確保されないとこの利点は活かされないこととなる。過去には資金面・人材面の問題で運用を継続できなくなったデータアーカイブもある。構築した人社データインフラの安定的、持続的な運用を確保できるよう、引き続き考えていくことが重要である。

本事業では、まず、人社データインフラを構築することそのものに主眼が置かれたため、公募により選定した拠点機関に閉じたネットワークで出発したが、人社データインフラの拡充段階では、人文学・社会科学分野において、データの蓄積や提供に貢献している機関が連携することによって進めていくことが良いのではないかと考える。そのためには、他機関の積極的な参画を促す開かれたネットワークへと段階的に移行していくことが望ましい。この開かれたネットワークでは、データアーカイブに関するノウハウと実績のある機関が、データを蓄積するサーバーの共有やクラウド化、あるいはデータアーカイブにおけるデータの長期保存・長期利用性の維持に関するノウハウや技術の蓄積、公開、共有について指導的な役割を果たすことが望ましい。併せて、これまでの拠点機関同士の連携等を活かしつつ、異なる分野間での相互交流を進めていくことも必要である。

ネットワーク型の人社データインフラでは、全体的な運用方針を定めたり拠点機関との相互調整を行ったりする機能を担う中核機関を置くことが必要である。この中核機関の果たす役割は極めて重要であり、長期にわたって安定的に運用できる体制が必要である。本事業では中核機関の機能を学振と NII が

【機密性 2 情報】

担うこととなったが、パイロット事業ということもあり、専任職員を 1 名しか配置することができず、他機関との兼務職員（教員・研究者）等が事業の実務における中心的な役割を担うこととなった。中核機関がその求められる機能を適切に果たすためには、各分野の研究データに関して分野専門家とコミュニケーションのとれる人材やシステム基盤の専門家など、高度な知識・技能を有する者が必要であるが、これらの業務を兼務職員に依存する体制は本事業全体を通じて大きな課題であり、本事業終了後の次のステージにおいて中核機関となる機関においては、これらの高度専門人材を採用することを含め、充実した人員体制を整えることが人社データインフラを安定運用に導くために必要であると考えられる。

このほか、人社データインフラに関する日本の状況を発信することと併せて海外の動向を把握していくことも重要である。これについては、各分野のデータアーカイブのノウハウと実績のある拠点機関がリーダーシップを発揮して海外との連携を強化していくことが望ましい。

3.3. 人材

安定的、持続的な人社データインフラの運用には、データを専門に扱うデータ・ライブラリアン等の活躍が不可欠である。しかしながら、我が国においては、研究者が研究時間を割いてアーカイブ運営に携わらざるを得ないのが現状である。諸外国の状況を踏まえると、データの利活用やその技術に関するスタッフを拡充していく必要があると考えられる。具体的には、システム基盤を管理・運営する技術スタッフやデータキュレーションなどを実施するスタッフ、データの利活用を希望するユーザーへの対応を行うスタッフ、長期に渡る安定的なデータの保管を担うスタッフ等である。

第 6 期「科学技術・イノベーション基本計画」（令和 3 年 3 月 2 6 日閣議決定）では、「社会全体のデジタル化やオープンサイエンスの潮流を捉えた研究そのものの DX¹⁶を通じて、より付加価値の高い研究成果を創出し、我が国が存在感を発揮することを目指す。」とされており、研究データ管理・共有・利活用にあたる人材についても国レベルで検討が始まっている。今後、各分野・領域等で議論が進むことが想定されるが、人社データインフラに必要な人材については、これらの動きを参照しつつ、しかるべき組織、機関において具体的な人材像の絞り込みや育成のあり方を検討することが必要でないかと考える。

3.4. 人文学・社会科学総合データカタログ（JDCat）

人文学・社会科学分野の国内外の研究者に対して、有用な研究データの発見可能性を高めるための活動は、中核機関を中心として今後も継続すべきである。JDCat には、拠点機関が精選した、クオリティが高く利活用の需要が見込まれる研究データに関する情報を集約することが望ましい。また、幅広い分野をカバーする CiNii Research など外部のシステムにもメタデータを流通させることで、データの発見可能性をさらに高めていくことが期待される。今後は、ユーザーの要望を踏まえた改良を重ね、JDCat のユーザビリティと信頼性を一層高めていくことが重要である。

拠点機関においては、利活用の需要が見込まれる貴重なデータの寄託を促進するとともに、JDCat にメタデータが掲載されているデータの長期利用・長期保存が保証されていることが望ましい。また、中核機関においてもデータの長期利用・長期保存のための枠組みを検討していくことが期待される。

¹⁶ デジタル・トランスフォーメーション。ウメオ大学（スウェーデン）のエリック・ストルターマン教授が 2004 年に提唱した概念。

3.5. オンライン分析システム

JDCat で発見したデータの利活用を目的に構築されたオンライン分析システムは、データの分析環境を教員と学生、研究グループ内で共有できる機能を備えるため、人文学・社会科学分野のデータ駆動型研究等に取り組む研究者、学生を支援する有用なツールになると考えられる。本格運用開始は令和 4 (2022) 年度となるため、その検証はこれからの課題となるが、本事業期間終了後は NII において運用を継続する予定のため、更なるデータの利活用が促進されることを期待する。

3.6. データ共有のための手引き

これまで、人文学・社会科学分野におけるデータ共有のための手引きで日本において策定されたものはなく、本事業において、海外の複数の機関（例：UKDA、FSD）のガイドラインを参考にして日本で初めて策定した成果は大きい。今後は、中核機関において適切な頻度で更新することが望ましい。共通事項と各分野あるいはデータの性質ごと（例：コンテンツの種類、ならびにデータ表現の仕方（ファイルフォーマットなど））で分けて作成することも考えられる。

3.7. データ利活用に関する研究者コミュニティの形成・文化の醸成

データ利活用に関する研究者コミュニティの形成・文化の醸成に関しては、研究機関や学会、そして、研究者と実務者が相互に協力して推進していくべきである。データを利活用した研究については、既に多くの実績がある分野もあれば、まだ経験が浅い分野もあり、その実態は学問分野、領域等によって異なる。データを利活用した研究の進んでいる分野の研究者等からデータ利活用に関する知識・技術等を積極的に学び、他の研究者と共有していくことが望ましい。そのような活動は、データ利活用に関する研究者コミュニティの形成につながり、やがては、人文学・社会科学分野全体におけるデータ利活用に関する文化が醸成されていくこととなる。研究機関や学会においては、このような研究者の活動を後押しする取組が望まれる。

併せて、研究においてデータを扱う専門人材が研究の推進には不可欠であるという理解を、研究者コミュニティの中で広めていくことも重要である。

さらに、データマネジメントプラン（DMP）や研究データの長期保存など、各種の政策の動向や社会背景を研究データのライフサイクルと関連付けるための啓発活動も、データ利活用に関する文化の醸成に貢献すると考えられる。

研究者コミュニティの形成、文化の醸成は、研究者等、当事者がメインで進めるものではあるが、国内外における情勢変化に適切に対応するためには、研究機関、学会、行政等、関係者による強力な後押し、支援も必要であると考えられる。