

人文学・社会科学関係の データアーカイブの整備・公開や関連研究に関する取組の例

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事業」【別紙1】

人文学・社会科学のデータの分野・国境を越えた共有及び利活用を促進するオールジャパンの基盤を構築することを目的に、平成30年度から事業実施。

日本学術振興会が中核機関となり、データの公開や利用等のための共通ガイドラインの策定、分野横断的な総合的データカタログの整備等を行うとともに、データアーカイブの整備・発信等を行う拠点機関を選定し、連携してデータインフラストラクチャーを構築。

平成30年度、以下の4機関を拠点機関に選定。

- ・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
- ・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
- ・一橋大学経済研究所
- ・大阪商業大学 JGSS 研究センター

2. 各大学共同利用機関における取組及び取組支援

(1) 日本語の歴史的典籍の国際共同研究ネットワーク構築計画【別紙2】

人間文化研究機構・国文学研究資料館では、国内外の大学・研究機関と連携し、日本語の歴史的典籍約30万点を画像化したデータベースを整備し、歴史的典籍の国際共同研究ネットワークを構築する計画を実施。これにより、歴史学・社会学・哲学・医学などの諸分野の研究者が多数参画する異分野融合研究を醸成し、幅広い国際共同研究の展開を目指す。

平成30年度は、医学・理学・産業・武術・宗教分野を中心とした画像情報の作成、国際及び異分野融合の共同研究を進めるとともに、拠点大学との協働による内製化を本格化し、画像情報の作成業務を加速。

(2) データサイエンス共同利用基盤施設における取組【別紙3】

平成28年度に情報・システム研究機構に設置された「データサイエンス共同利用基盤施設」では、データ共有支援／データ解析支援／データサイエンティスト育成の3つの活動を通し、大学等のデータ駆動型研究を促進するための支援を実施。特に、人文学・社会科学関係のセンターとして「社会データ構造化センター」及び「人文学オープンデータ共同利用センター」を整備。

○社会データ構造化センター

社会調査データや公的マイクロデータ、ソーシャルビッグデータを整備するとともに、産業界等とも連携してコンソーシアムを設立し、政府と連携してオンサイト施設を全国的に開設し研究環境の整備を推進し、我が国における公的統計マイクロデータの研究利用（二次利用）を促進。

○人文学オープンデータ共同利用センター

情報学・統計学の最新技術を用いて人文学データへのアクセスを改善する研究開発を進めるとともに、オープンサイエンスの考え方に基づき多くの人々が参加できるデータプラットフォームを構築。

3. 各大学における取組及び取組支援

(1) 共同利用・共同研究拠点における取組

人文・社会科学系の共同利用・共同研究拠点として認定を受けている拠点において、それぞれの研究分野に応じた研究成果や研究データのアーカイブ整備やアクセスオープン化が行われている【別紙4】。

(2) 科学研究費助成事業による支援

研究成果公開促進費において、研究者個人又は研究者グループ等が作成するデータベースで、公開利用を目的とするものに対して助成を実施。

平成30年度は、人文科学・社会科学分野28件の取組を支援。

【基礎科学力の強化に関するタスクフォース審議のまとめ】（平成29年4月）

社会科学分野において、データプラットフォームの構築を検討する。オールジャパン体制によるデータ・サプライヤーとしての機能に特化し、中立的かつ持続的な活動を可能とする組織的な枠組みとし、社会科学系データの一元的な収集・保管・提供・活用を図る。

【学術の総合的發展をめざして～人文・社会科学からの提言～（日本学術会議）】（平成29年6月）

歴史資料・公文書・調査データなどのデータベースの構築は、人文・社会科学の新たな展開にとって不可欠であるばかりか、国際的な共同研究の基盤形成にも必要である。

<データの重要性>

- 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、その分析が当該分野の研究の国際的な潮流となり、国際的なジャーナルへの投稿に検証可能なデータ公開が条件となるのみならず、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る。こうした中、オープンサイエンスの潮流も踏まえ、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。

<我が国の現状>

- 我が国では、個人研究の度合いの強い人文学・社会科学のデータの多くが個々の研究者の管理に委ねられており、しかも、貧困や高齢化、国民意識などに関する代表的な国際調査でさえも、調査主体である研究者グループの自主的な取組に依っている。これら当該分野を支えてきた研究者が退職時期を迎えようとする今日、こうした取組が途絶え、或いはデータが散逸すれば、国際的に「**ジャパン・ミッシング**」に陥る危機にある。
※貧困や所得格差等に関し、約50か国の政府統計等を収録した最大の国際比較調査(Luxembourg Income Study)の日本の収録データは私立大学の実施した調査データである。
※約70か国の人々の価値観を時系列的に調査する世界価値観調査における日本の調査は、民間研究所の協力を得て、個人の研究者によって行われている。
- アジア諸国と比べてもデータの多言語対応は遅れており、発達する国際共同研究の中での「**ジャパン・バッシング**」の危機にある。
※2014～2016年のアメリカの人文学・社会科学系分野の共著相手国で、日本は14位と、中国3位、韓国7位の後塵を拝し、インド15位、香港17位、シンガポール18位、台湾20位と拮抗している。
- データの共有が進まないこともあり、研究者がそれぞれに同様の調査、データの編集等を行い、**研究活動の効率性を損ねている懸念**がある。

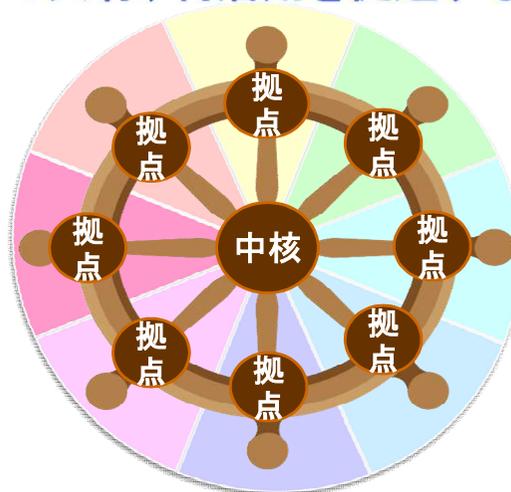
人文学・社会科学のデータの共有、利活用を促進するオールジャパンの基盤を構築

ネットワーク型データ共有基盤の構築 (拠点機関)

各分野で拠点的取組を行う組織が相互に連携しながら、それぞれの分野の研究者コミュニティの拠点となり、

- ・データ・アーカイブ機能の強化(共有化)
- ・海外発信・連携機能の強化(国際化)
- ・データ間の時系列等接続関係の整備(連結化)
- ・新たな社会基盤的調査の支援(高度化)

データ分析に基づく人文学・社会科学
研究の飛躍的發展



日本の人文学・社会科学研究の
「見える化」による国際共同研究の強化

人文学・社会科学リサーチコモンズの形成 (中核機関)

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、拠点間の強力な連携協働により、

- ・データ公開、利用、権利関係等の**共通ガイドライン**の策定
- ・分野横断的な総合的**データカタログ**を整備
- ・オンライン**オーダーメイド集計・分析システム**の開発研究
- ・公開シンポジウムの開催やニュースレターの配信等を推進

研究者が共にデータを共有・利用しあう
文化を醸成

日本語の歴史的典籍の国際共同研究ネットワーク構築計画

【事業概要】

平成30年度予算額 171,949千円
 (平成29年度予算額 180,000千円)

人文学分野の長年の課題である研究の細分化、従来型の研究手法からの脱却を図るため、日本語の歴史的典籍を対象とする大規模画像データベース整備を基に、その活用により人文学分野における組織的かつ国際的な共同研究を推進し、医学、理学分野等との異分野融合の国際共同研究ネットワークを構築する。

＜基礎データ＞

- (整備) 約39.5億円(2014～2023年度、10年計画) (※2013年度に準備経費1億円を措置)
 (運用) 2017年度から新規システムによりデータベースを公開
 (共同利用研究者数) 1,061人(うち外国人229人(14ヶ国・地域)) ※2014～2017年度実績(延べ数)
 (外部利用者数) 35,576人(うち外国人3,203人(67ヶ国・地域)) ※2017/5～2018/4のユーザー数(延べ数)
 (外部利用件数) 623,433件 ※2017/5～2018/4のページビュー数
 (成果図書・論文数) 1,036件(うち国際270件) ※2014～2017年度実績(延べ数)

【事業の実施体制】

【中心機関】大学共同利用機関法人人間文化研究機構 国文学研究資料館

【機構内連携機関】国立歴史民俗博物館、国立国語研究所、国際日本文化研究センター

【国内拠点】東京大学、京都大学、慶應義塾大学等(国内20機関)

【国外拠点】コレージュ・ド・フランス日本学高等研究所、コロンビア大学、北京外国語大学等(国外16機関)

【連携研究機関】国立情報学研究所、国立極地研究所、国立国会図書館、弘前大学、茨城大学等

【研究計画】

＜平成30年度＞

- ・医学、理学、産業、武術、宗教分野を中心に、画像情報を作成。
- ・内製化を拠点大学とともに本格的に実施し、画像情報作成を促進。
- ・海外シンポジウムの開催やHP・SNS等を通じ、研究成果発信と広報活動を実施。

＜「新日本古典籍総合データベース」の構築＞

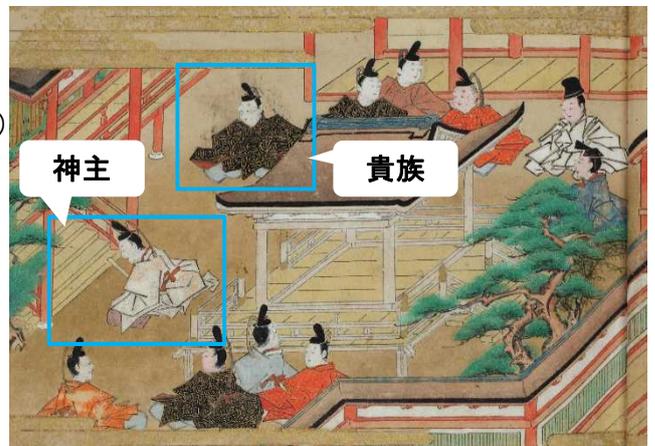
- ・歴史的典籍の大規模情報提供システム(第1期)公開、検索機能の向上化・多言語化対応(平成29年～33年)
- ・歴史的典籍の大規模情報提供システム(第2期)(平成34年～)

＜国際共同研究ネットワークの構築＞

- ・異分野連携を踏まえたネットワークの拡充(平成27年～35年)

＜国際共同研究の推進＞

- ・異分野融合研究の醸成(平成26年～平成35年)
- ・総合書物学の構築(平成26年～平成34年)
- ・文献観光資源学研究的の推進(平成29年～平成33年)



絵検索のためにタグ付けされた画像情報の例 出典:『大職冠』

【期待される成果】

- ◎人間文化研究の新たなデータベースとして、広く大学の教育・研究に寄与。
- ◎オープンデータを推進し、異分野融合研究を醸成。
- ◎文理融合による新たな研究領域の開拓。
- ◎30万点の日本語の歴史的典籍が研究者の机上で閲覧可能になり、研究の効率が飛躍的に向上。
- ◎双方向的フェロウシップによるグローバルな次世代研究者の育成。

分野	期待される成果の一例
総合書物学	文化資源学としての典籍の活用
文献観光資源学	地域観光資源の開発
典籍防災学	気候・天文等に関する文献から過去の災害の要因を知り、災害の予知予防に貢献
医学	伝統医療の樹立と開発
	成人病の新たな治療方法への貢献
薬学	日本由来の新たな生薬の開発
農学	救荒対策としての農作物の復元
生活文化学	調理方法の復元
人文地理学	統合地誌研究

社会データ構造化センター (CSDS)

構造化され、整備された社会データは、複雑な現代社会を解明し、地域環境、治安、経済を含むさまざまな課題解決に役立ちます。本センターは、日本学術会議、総務省統計センターをはじめ、国内外の調査機関、研究機関等と連携し、人間・社会データ・コンプライアンス管理プラットフォームの構築を目指しています。具体的に日本人の国民性調査などの「社会調査データ」、オンラインで利用できる「公的調査のマイクロデータ」、人の社会行動をリアルタイムで表す「ソーシャルビッグデータ」の整備・向上に取り組み、データに基づく人文・社会科学と政策立案の進捗を支援しています。(センター長: 吉野 諒三)

2017年度の主な成果:

- ▶ 日本人の意識に関する収集済みの社会調査データの属性別詳細集計表を作成・公開するとともに、アジア・太平洋価値観国際比較調査等の公開を開始しました。
- ▶ 公的マイクロデータの公開・利用について、国際会議の開催や研究コンソーシアムの事務局を担っています。昨年度は、政府・地方自治体の個票レベルのデータが利用できる「オンサイト室」をDS施設内に移転・開設しました。



アジア・太平洋価値観国際比較調査



第2回人間・社会データ構造化シンポジウム

- ▶ 話題の欧州のGDPR (一般データ保護規制制度) などを含め、ソーシャル・ビッグデータ時代の個人情報保護関連法律・倫理について広く啓蒙を図る「第2回人間・社会データ構造化シンポジウム」を開催しました。

人文学オープンデータ共同利用センター (CODH)

本センターは、人文学分野におけるデータのオープン化と共同利用に基づく研究・支援活動を推進するため、大規模に電子化されつつある日本の文化、例えばくずし字で書かれた江戸時代の古典籍や活字印刷された明治以降の書籍などを対象に、情報学・統計学の最新技術を活用した自動テキスト化や内容分析などを行っています。しかし人文学研究分野におけるデータサイエンスの発展に不可欠な、データのオープン化はまだ十分に進んでいません。そこで、世界の人文知を共有する情報基盤を構築するとともに、異分野の研究者、市民、企業等とコラボレーションしながら、オープン化の推進を図っています。(センター長: 北本 朝展)

2017年度の主な成果:

- ▶ 「日本古典籍データセット」に、『伊勢物語図録』や『武鑑』など約1,000点を追加・公開しました。
- ▶ 世界のミュージアムやライブラリで使われる国際的画像配信方式「IIIF (International Image Interoperability Framework)」を活用し、任意の画像を切り出して収集可能な「IIIF Curation Viewer」を開発・公開しました。
- ▶ 「江戸料理レシピデータセット」で、江戸時代の卵料理を現代で活用するためのレシピと写真を追加し、ウェブサイト「クックパッド」にも掲載しました。



「IIIF Curation Viewer」を用いて画像を収集した「顔貌コレクション」を公開



くずし字チャレンジ! :
くずし字解読のためのAIコンペを開催

人文学・社会科学関係の共同利用・共同研究拠点(平成30年度)
(研究データ等のアーカイブ整備、オープンアクセス化関係)

研究施設名	共同利用・共同研究拠点名	データの例
北海道大学 スラブ・ユーラシア研究センター	スラブ・ユーラシア地域研究にかかわる拠点	中東欧・旧ソ連諸国の選挙データ
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	社会調査・データアーカイブ共同利用・共同 研究拠点	各種の社会調査データ
東京大学 史料編纂所	日本史史料の研究資源化に関する研究拠 点	日本関係史料の史料情報、デジ タル画像
東京外国語大学 アジア・アフリカ言語文化研究所	アジア・アフリカの言語文化に関する国際的 研究拠点	デジタル化された辞書やデータ ベースのオンライン公開
一橋大学 経済研究所	「日本及び世界経済の高度実証分析」拠点	政府統計マイクロデータ、マーケティ ングデータ等
京都大学 人文科学研究所	人文学諸領域の複合的共同研究国際拠点	漢籍や中国書をはじめとする図 書資料等
京都大学 東南アジア地域研究研究所	地域情報資源の共有化と相関型地域研究 の推進拠点	地域研究に係る写真資料等
京都大学 東南アジア地域研究研究所	東南アジア研究の国際共同研究拠点	マレー・インドネシア語雑誌記事 横断検索システム
大阪大学 社会経済研究所	行動経済学研究拠点	「くらしと好みと満足度」調査
慶應義塾大学経済研究所附属 パネルデータ設計・解析セン ター	パネル調査共同研究拠点	日本家計パネル調査等
法政大学 野上記念法政大学能楽研究所	能楽の国際・学際的研究拠点	金春家旧伝文書デジタル アーカイブ等
早稲田大学 イスラーム地域研究機構	イスラーム地域研究拠点	カイロのイスラーム建築等
早稲田大学 坪内博士記念演劇博物館	演劇映像学連携研究拠点	演劇博物館演劇情報総合デー タベース等
神奈川大学 日本常民文化研究所	国際常民文化研究拠点	漁業制度資料等
大阪商業大学 JGSS研究センター	日本版総合的社会調査共同研究拠点	総合的社会調査に係るデータ等
関西大学 ソシオネットワーク戦略研究機構	ソシオネットワーク戦略研究拠点	RISS経済心理学データアーカイ ブ等
立命館大学 アート・リサーチセンター	日本文化資源デジタル・アーカイブ研究拠 点	ARC浮世絵ポータルデータベー ス等
文化学園大学 文化ファッション研究機構	服飾文化共同研究拠点	文化学園リポジトリ等

※本資料は文部科学省が調べたもので、あくまで一例である。