

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」について

廣松 毅 | Hiromatsu Takeshi

(独)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センター
センター長



■ 1972年東京大学大学院経済学研究科修士課程修了(経済学修士)、同大学助手・助教授・教授を経て2009年同大学定年退職、情報セキュリティ大学院大学教授・特任教授を経て、現在同客員教授。18年11月より現職。

はじめに

独立行政法人日本学術振興会(以下、「学振」と略)は平成30(2018)年度から「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」と略)に着手している。本事業の目的は、人文学・社会科学研究に係るデータを分野や国をこえて共有・利活用する総合的なシステム(データインフラストラクチャー)を構築することにより、研究者がデータを共有・利用しあう文化を醸成するとともに、国内外の当該分野の共同研究を促進し、それによって人文学・社会科学の振興を図ることである。

本稿では、公表資料を基に本事業の概要について説明を行う^{注1)}こととして、人文学・社会科学インフラストラクチャー構築推進センターの活動そのものや本事業を推進するために必要な技術的支援については本号の他の二つの論文

に譲ることとする。

なお、本稿の内容は原稿執筆時点(令和元年9月)のものであり、実施体制、取り組み等については、随時変更されることがある。

背景

現在、科学研究全体に大きな変革が生じつつある。なかでも、人文学・社会科学研究においてはこれまで個々の研究者ないしは少人数のグループが資料・データを収集し、その分析・解釈を通じて意味づけを行うという手法が主流であった。しかし、近年のICTの発展によって多数の研究者が共同でデータを収集・分析する実証的なアプローチが国際的に急速に発展し、国際的に評価の高い学術雑誌への投稿などでは、データを用いた実証研究に分析の再現可能性が求められるようになってきた。こうした傾

注1) 本稿の記述はあくまで筆者個人のものであり、学振の公式見解を示すものではない。

向と同時に、世界的なオープンサイエンス^{注2)}の潮流を踏まえて、多くの国々においてナショナルセンターなど国家レベルでデータの集積と管理、利活用を推進するデータインフラストラクチャーが整備され、これらの機関が国境をこえて相互に連携して、データの共同利用を促進するとともに大規模で継続性を重視した国際比較調査を協働して実施している。

日本の現状

日本においても一部の分野で大学等の研究者グループを核に拠点的取り組みが行われている。しかし、個人研究の度合いが強い人文学・社会科学では、多くの場合、データの管理が個々の研究者に委ねられており、研究者の異動や退職、研究室の改廃などによって散逸することも少なくない。また、特に社会科学の学術研究に不可欠な基盤である国勢調査や国民経済計算などの大規模な政府統計に関しては、分類方法の変更や市町村合併などの環境変化によって過去の調査結果との時系列的な接続関係が不明なものが少なからずあり、この解明のためにマイクロデータを分析する膨大な作業が個々の研究者ないしはグループの献身的な努力に依存しているのが現実である。一方、例えば人文学における「日本

語の歴史的典籍のデータベースの構築」^{注3)}などのように学術研究の大型プロジェクトとして進められており、大きな成果を上げている例もある。

とはいえ、人文学・社会科学の国際的ネットワークの急速な発達に伴って、諸外国の機関等と連携した大規模な国際共同比較調査などが増大している中で、例えば貧困や高齢化^{注4)}、国民意識^{注5)}などに関する国際共同比較調査において日本の調査は個々の研究者グループの献身的な取り組みに依存しており、国際的に求められる役割を十分に果たせているとは言い難い。

そして、これら研究者ないしはグループの努力によって続けられている拠点的な取り組みは、そのほとんどが所属大学の運営費か科研費等の外部資金によって賄われている。しかし、昨今の大学の財政状況は厳しく、また、継続性が求められる調査や国際共同比較調査については新規性や独創性を求める外部資金の審査で不採択となることもあり、その運営基盤は脆弱でかつ限界を迎えつつある。さらに、これらの取り組みを牽引してきた研究者層が高齢化し退職の時期を迎えており、こうした取り組みが途絶えることによりこれまで蓄積されてきたデータが散逸すれば、国際的に重要なデータで「ジャパン・ミッシング」^{注6)}に陥る危機にある。

注2) オープンサイエンスに関しては、第5期科学技術基本計画（平成28年1月閣議決定）、科学技術・学術審議会学術分科会学術情報委員会「学術情報のオープン化の推進について」（平成28年2月）、日本学術会議オープンサイエンスの取組に関する検討委員会「オープンイノベーションに資するオープンサイエンスのあり方に関する提言」（平成28年7月）など、多くの検討・提言がなされている。全体を通じて、オープンサイエンス＝オープンアクセス＋オープンデータと捉えられているようである。

注3) <https://www.nijl.ac.jp/pages/cijproject/> を参照。

注4) 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世界最大の国際比較調査である 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。

注5) 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、個人の実験者によって行われている。

注6) 世界の標準的な統計調査等において、日本のデータが収集ないしは公表されていないため「日本が存在していない」として、無視される状況をいう。研究者自らが調査収集したデータであっても、それが他の研究者誰もが利用できる公表データとなっていない場合、それを使った論文は、第三者による検証が不可能であるとして、国際的に評価の高い学術雑誌の審査基準を満たさず、応募さえ認められない場合が増えつつある。

加えて、欧米のみならずアジア諸国のデータアーカイブ機関も多言語化によって相互の連携を深めているのに対して、日本の多言語対応は遅れている。データ分析が人文・社会科学研究の国際的な潮流となりつつある中、分析可能なデータの存在は、国際共同研究や研究者交流において研究成果を残せるかを決定づける鍵であり、国際共同研究が急速に進む中で「ジャパン・パッシング」^{注7)}の危機にもある。

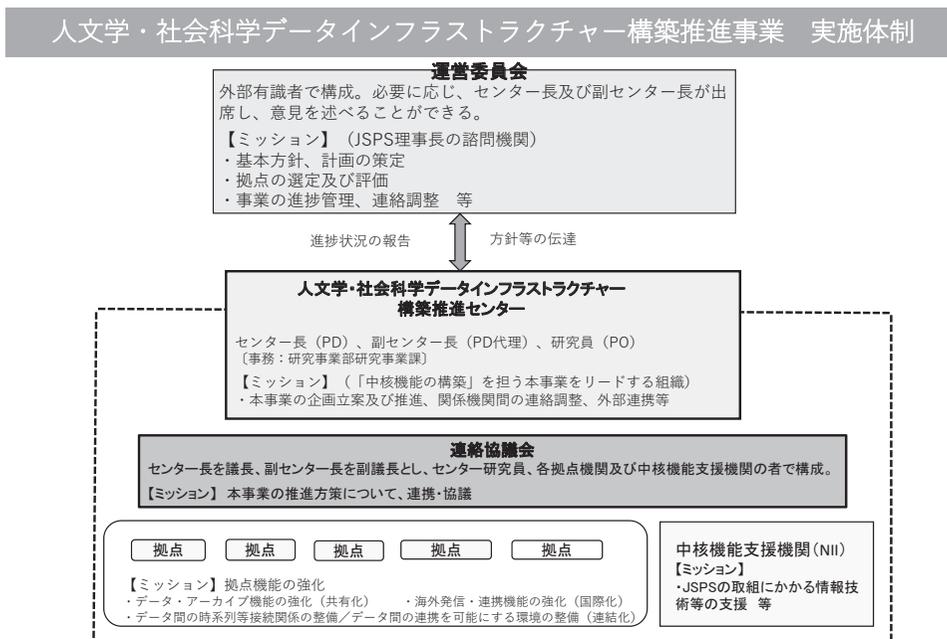
本事業の取り組み^{注8)}

上記のような背景と日本の現状を踏まえて、これまででも多くの関係者・機関によって、人文・社会科学分野におけるデータインフラストラク

チャーの整備がたびたび提起されてきた。しかしながら、それは人文・社会科学分野においてナショナルセンターとなり得る機関が存在しなかったこともあり、実現されずに残された課題となっている。本事業は、国、学術振興機関、大学共同利用機関、共同利用・共同研究拠点、大学等、そして個々の研究者という各主体の協力を通じてこの課題に取り組むために、学振の中にセンターを設けて平成30(2018)年度に着手したものである。実施体制は下図のとおりである。

前述のように、現状、日本には人文・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、一部の分野では大学等の研究者グループを核に拠点的取り組みが行われ、文部科学大

図1 実施体制



注7) 日本が素通りされてしまうこと。

注8) 学振は科学研究費助成事業、いわゆる科研費の申請受付・審査・配分・管理の事業もやっている。これは学振の公的な研究資金の配分機関としての機能である。現時点で、本事業はそれと直接関係はしていない。

特集 ■ 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構築の推進

臣による共同利用・共同研究拠点の認定を受けている拠点も存在する。このため、本事業では学振を中核とするとともに人文学・社会科学における分野毎に選定した拠点（5 拠点）の基盤を強化し、それぞれの専門分野のデータの共有化に向

けて取り組みつつ、拠点間の強力な連携協働によって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を目指して、取り組んでいる^{注9)}。

以下の二つの図は、この取り組みを図示したものである。

図2 取り組み①

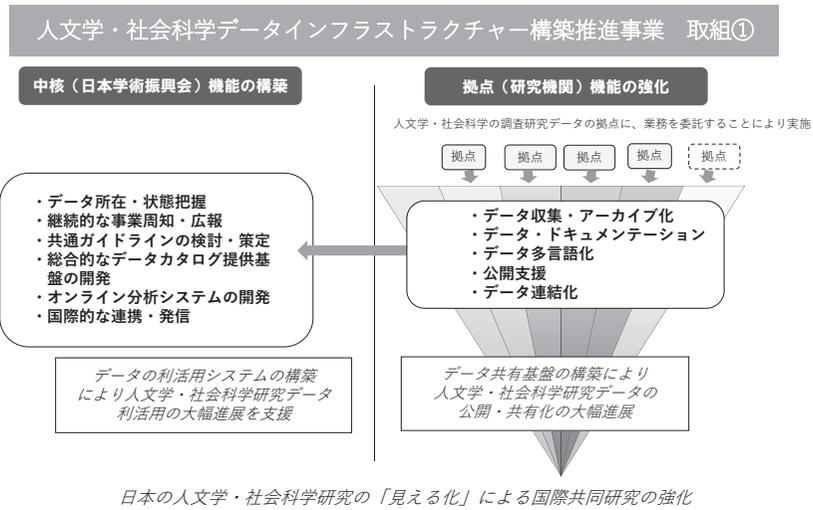
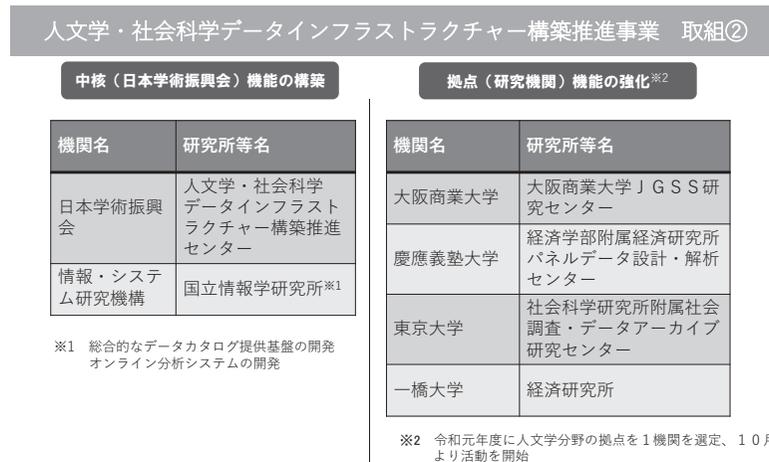


図3 取り組み②



注9) 本事業の平成30(2018)年10月16日時点での取り組み体制については <https://www.jsps.go.jp/j-di/torikumi.html> を参照。この時点での4拠点はいずれも社会科学分野の拠点であり、平成30(2018)年10月1日より活動を始めている。人文学系の拠点に関しては令和元年度に選定され、10月1日より活動を開始することになっている。

取り組みの具体的な内容は、以下のとおりである。

I. 各拠点の取り組み

— ネットワーク型データ共有基盤の構築 —

大学等の研究所等が、関連分野において、広く人文学・社会科学系研究の基盤となり得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次の取り組みを実施する。また、取り組みの実施に当たっては中核である学振が行う取り組みにも連携協力する。

- ア. データアーカイブ機能の強化（共有化）
- イ. 海外発信・連携機能の強化（国際化）
- ウ. データ間の時系列等接続関係の整備／データ間の連携を可能にする環境の整備（連結化）

II. 中核の取り組み

— 人文学・社会科学分野のリサーチコモンスの形成 —

学振は、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、以下の取り組みを行う。

- ・ データ共有、利用、権利関係等の共通ガイドラインの策定
- ・ 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
- ・ 分野横断的、総合的データカタログの整備
- ・ リサーチコモンスの形成（公開シンポジウム開催やニュースレター配信等による啓発）

将来の展望

本事業の最終目的である人文学・社会科学データインフラストラクチャーの構築は、国、学術振興機関、大学共同利用機関、共同利用・共同研究拠点、大学等、そして個々の研究者が、主体的に取り組むことによつてのみ実現できるものである。現時点では、着手してまだ2年目ということもあり、対象としているデータの種類もそれほど多くはない。現時点で各拠点が保有しているデータは図4のとおりである。

特に人文学系のデータに関しては、本年10月1日から活動開始であるため、本稿執筆時点では未知数である。われわれとしては、今後、各拠点に多くのデータが寄託されデータインフラストラクチャーとして充実していくことを目指す覚悟である。

しかしながら、将来直面すると考えられるいくつかの困難がある^{注10)}。一つ目は本事業のような基盤整備には財源措置を伴い、国や学術振興機関の予算措置や枠組み整備等を必要とすることである。本事業は、一部の分野で研究者グループの努力によつて続けられている拠点的な取り組みを核としているものの、すでに指摘したように、現状ではそのほとんどが所属大学の運営費か外部資金によつて賄われてきている。しかし、こうした大学のミッションをこえる取り組みを独自財源で賄うには限界を迎えており、基盤構築のために行うべき取り組みについては、財政支援が必要となる。

二つ目に、現在中核・拠点となっている機関は現状把握やガイドラインの整備等に同時に着手し努力をするとともに、その努力を今後も継

注10) 本事業は、平成30(2018)年度から5年間の時限付きである。したがって、期限が過ぎた後の本事業の継続に関しては、現在のところ未知数である。

図4 各拠点が保有するデータ

各拠点が保有するデータ

- ・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2,000セットを超える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
 - 長期経済統計
- ・慶應義塾大学 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
- ・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)

続していく覚悟であるが、構想の実現に向けて関係主体が共に踏み出すことが肝要である。本事業のようなネットワーク型の基盤構築には、財源とともに時間がかかる（あるいは、最終終着点のない）事業であり、継続していくためには人材の育成も必要である。さらに三つ目の点として、本事業の関係者には参加意識、主体性を高める効果がある反面、関係組織の意思決定等の複雑化による弊害が現れないとは言い切れない。その面からも、一定の時期にこのような

事業のあり方について検証する必要があるであろう。

これらの将来考慮すべき問題点があるにしても、われわれとしてとにかく人文学・社会科学データインフラストラクチャーのプロトタイプを構築して、その成果を国、学術振興機関、大学共同利用機関、共同利用・共同研究拠点、大学等、そして個々の研究者などの関係者に問うべきであると考え、本事業に取り組んでいく所存である。