

中央調査報

(主な内容)

- 社会科学データを共有する制度基盤 (2) 1
- 告知板..... 6

社会科学データを共有する制度基盤 (2)

前田幸男

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・研究員
(東京大学大学院情報学環・教授)

前号では、人文学・社会科学領域においてデータを保存・共有するための制度基盤が必要な理由について述べた。本号においては、すでに公表された資料類を前提に¹⁾、日本学術振興会が2018年度から実施している人文学・社会科学データインフラストラクチャー構築推進事業（以下データインフラ事業）について説明する。執筆している前田の専門・担当の関係で、基本的に社会科学に議論が集中し、例示も政治学の例が多くなることをお許し頂きたい。

背景

データインフラ事業が構想された段階では、主に3つの関心・懸念があったように思われる。第一は、現状を放置すると貴重な調査データが散逸するのではないかと懸念である。第二は、データが日本語のみで提供される、あるいは、国外の研究者にとって入手困難な状況にとどまることで、日本を対象とする国際的あるいは国際比較研究が減少するのではないかと懸念である。第三は、研究データの共有が進まないために、研究活動の効率性が損なわれ、国際競争力のある研究発信が難しくなることであった。以下、順を追って掘り下げて説明したい。

最初の論点であるデータの散逸についてだが、古い調査データは今日では入手不可能なものが少なくない。日本においても1950年代には学術的調査が開始されている。1955年に日本社会学会が実施した社会階層と社会移動全国調査（日本社会学会調査委員会編1958）は、今日でもその電子データは、東京大学のSSJDAを通じて利用可能である（SSJDA 0759）。また、政治

学分野における最初の全国学術調査は1966年に政治参加の7カ国国際比較の一環として実施された市民意識調査（池内 1974）だと思われるが、その磁気データは今日においてもICPSRあるいはSSJDAを通じて利用可能である（ICPSR 7768, SSJDA M001）。ただし、これらの例は、どちらかと言えば例外であるように思われる。

筆者の専門である政治学分野からの例を挙げれば、戦後直後に蠟山政道らが行った選挙区における意識調査のデータは今日残っていないと思われる（その報告は、蠟山他1955）。1970年代から1980年代にかけて北海道大学の研究チームが行った札幌市の選挙調査データ（荒木1994）も、筆者が知る限り制度的に保存・共有された形跡はない。日本においては、1990年代半ばまでは調査データを共有する仕組みも乏しく、データの保存と共有は、主に個人研究者あるいはチームの活動に委ねられてきた。データインフラ事業が着手された背景には、研究者の世代交代が進み、また、情報端末や電子媒体の技術革新が続くことで、貴重な調査データが散

逸するのではないかという危機意識があったことは間違いない。

二点目の国際的なデータの提供であるが、他の非英語圏諸国と比べても日本のデータの多言語対応は遅れている。日本国外の研究者にデータを提供するためには、日本国内においてデータを保存・共有するための作業とは異なる準備が必要となる。数値データそのものは言語に依存しないが、データを理解するためには、数値が何を意味するのかを示すメタデータが必要である。質問票だけではなく、調査概要などの文書を含めて英訳をする必要があるが、専門的な内容を含むため、単純に翻訳業者に依頼して済む作業ではない。非英語圏のデータ・アーカイブにおけるデータの英語化の状況について網羅的に調べることは難しいが、例えば、フィンランドのFinnish Social Science Data Archive (FSD)では、ほとんどのデータについてメタデータの英訳を行い、国外の分析需要に対応している。

また、国際比較調査データを作成するためには、一定のルールに従ったデータの加工・統一作業が必要である（それらの作業はData Harmonizationと言われることが多い）。例えば、経済学分野の国際比較家計データとして著名なLISについては、国際比較用データを作成するためには、変数名の変更、基準に従った値の再割当、乗率の付与、データの補定等、多くの作業が必要となる(LIS 2019)。これらの作業の中には相当の技術的習熟が必要なものも多い。データインフラ事業を開始した背景には、データの多言語化や国際比較データの作成を通じて、日本のデータを国外の研究者に対して可視化することも念頭にあった。

最後は、研究の効率性の問題である。データの共有が進んでいない状況においては、それぞれの研究者あるいは研究チームが同様の調査、あるいは同様のデータの編集作業等を独立に行うことになる。データを共有できれば、直ちに分析を開始することが出来るにもかかわらず、各研究者がデータ収集や編集に時間と労力を割くために、研究活動の効率性が損なわれる。

政治学分野の例で言えば、選挙結果データがその典型的な例である。選挙結果のデータは極めて重要であるが、公式の選挙結果を簡便に利用可能な形で整理・保存・提供している機関は

存在しない。総務省が提供している選挙データ²⁾は『選挙結果調』を作成した際のエクセル・データを掲載しているに過ぎず、そのままデータ分析に利用することはできない。現在は、まだエクセル形式で提供されているだけ改善されたが、以前は、PDF形式であり、おおよそデータの二次的利用を考えた代物ではなかった。これらのデータを学術的な分析に耐えるように編集・加工、さらには、異なる時期の選挙を超えて接合するためには相当の時間とデータの内容に関する習熟が必要である。無論、その作業自体は研究者自身が担うべきものかもしれないが、データを共有する制度基盤が欠如しているが故に、同種の努力が相互に連絡がないまま繰り返されているのが現状であるように思われる。

また、政府統計の分野で言えば、平成の大合併のように自治体数が大幅に減少すると同時にその区域が変更されると、一貫した分析を行うためのファイルを作成するためにはデータの編集作業が必要となる。経済の時系列ファイルで言えば、統計作成の基準が変更されると、その時点を挟んで、同じ基準で比較することは難しくなる。そのためには、同一の基準を適用して、データを編集・加工する必要になるが、そのデータが共有しなければ、同じような努力が繰り返されることになる。データを共有する基盤を作成することで、個々の研究者がデータの分析に集中できる環境を作ることが、研究成果を生み出す上で重要なのである。

より大きな政策的文脈においては、文部科学省の基礎科学力の強化に関するタスクフォースの議論においては、「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」ことが必要であることが指摘されていた(2017年4月)³⁾。また、日本学術会議は「学術の総合的発展をめざして一人文・社会科学からの提言一」において、「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。また、インターネット時代の到来に応じて、データベースは希望する者のだれもがアクセス可能なもの

にしなければならない」と主張していた(2017年6月)⁴⁾。このような政策的背景のもと、人文学・社会科学データの共有・利活用を促進するために2018年度より日本学術振興会の「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」が開始された⁵⁾。

日本学術振興会の事業

データインフラ事業の活動内容は、大きくは2つに分けられる。1つは、振興会が主体となつて行う活動であり、主には、①分野横断的な総合的なデータカタログの整備、②データの共有、利用、権利関係等の共通ガイドラインの策定、③オンライン分析システムの開発、④その他、必要に応じてデータの所在状況調査や広報活動等である。

その一方、振興会が公募を通じて拠点となる大学を募り実施している活動がある。具体的には、(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)データ間の接続作業あるいは連携可能にする環境整備(連結化)である。ここでは、まず後者の活動について、先に概要を説明したい。

(1) 拠点機関の活動

最初に、共有化が意味するのは、様々な学術データを受け入れ、整理・保存し、そして、研究者に提供する作業及びその基盤の整備を意味する。具体的な活動としては、データを長期的に保存・共有するためのシステム構築・環境整備や、データと研究成果・論文等との関連づけやDOIの付与、個体識別の可能性が高いデータを分析する環境の整備を想定している。

次に国際化であるが、既に実施された重要な調査データについてメタデータを英語に(場合によっては他の言語にも)翻訳すること、あるいは、国際比較に必要な加工・編集をデータに施すことを想定している。また、海外の同種の活動を行っている機関との連携、海外に対する情報発信も重要である。

最後に、連結化で意味しているのは、複数のデータを一定の基準に編集して、統合し、様々な分析を容易にする作業のことである。英語で言えば先述のData Harmonizationに近い言葉であろう。具体的には、繰り返し行われる同じ名称の社会調査であっても、社会状況の変化によって同じ質問の選択肢が変更されることは、

珍しいことではない。その場合、データを提供する側で複数のデータ・ファイルを統合する作業をしておけば、格段に利用者の手間は省ける。実際、諸外国の著名な調査については、統合ファイルが準備・提供されていることが稀ではない⁶⁾。また、同様の作業は経済時系列データや様々な政府統計の接続でも起こりうる。

上記の記述は少し抽象的であるが、公募を通じて、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、そして、一橋大学経済研究所が採択されている。これらの機関はそれぞれの経験と強みを生かして、上記の(ア)(イ)(ウ)について具体的な活動を行っている。

事業の序盤では、自らの機関が実施した調査データについて主に活動を実施している拠点もあるが、徐々に自研究所等以外が保有する既存データについても他機関からデータを取得し、各取組(アからウ)について実施することになっている。なお、データ共有基盤の強化に当たり、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を配置することを促している。詳細については、振興会のウェブサイトを確認頂きたい。

(2) 日本学術振興会の活動

振興会が行う活動の1つがデータカタログの準備である。今日においては、インターネット上の検索を通じて、データを探し出し、そして取得(ダウンロード)することが一般的である。本事業ではデータの提供自体は拠点機関に委ねているが、データの所在を確認するために、それぞれの機関のウェブサイトを閲覧する必要があるのは不便である。また、研究状況について習熟している日本国内の研究者であればともかく、大学院生や国外の研究者にとっては、データを見つける窓口となるカタログの存在は重要である。そこで、振興会は、拠点機関が収集・整理・公開するデータを横断的に一括検索できるデータカタログを構築するために準備を進めている。

実際の作業は国立情報学研究所の協力を得て行っている。現在、各拠点機関が提供するデータのメタデータを機械的に収集し、検索・閲覧に供するための仕組みを検討しているところ

である。その際のメタデータの企画は、社会調査データを記述するメタデータの規格として諸外国のデータ・アーカイブあるいはレポジトリの多くが用いているData Documentation Initiative(DDI)に準拠ることにしている⁸⁾。また、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されることを条件としている。

カタログはメタデータを搭載するのみでデータを直接入手出来るわけではない。ただし、カタログから直接データを所蔵する機関に連動させることで、データの検索と入手をできるだけ円滑な流れにおいて実施できることを目指している。長期的には、拠点機関のメタデータを収集するだけでなく、当該サイトにも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ことを計画している。また、拠点機関以外からのメタデータも登録できるように設計する予定である。

欧州における主要データ・アーカイブを網羅したCESSDAのデータカタログも⁹⁾、メタデータを横断検索し、データの提供自体は各国のデータ所蔵機関が行っているのも、振興会が準備しているデータカタログも発想としては類似している。また、ドイツのGESISは、58ものデータ所蔵機関のデータを横断検索するカタログのベータ版を公開している¹⁰⁾。

国を跨ぐ大規模なカタログだけではなく、一国内で様々な研究資源のカタログをつなぐ努力は存在する。例えば、フィンランドのEstinという統合型データカタログはフィンランド国内のデータについて35分野を横断して検索できるようになっている¹¹⁾。また、スウェーデンのSwedish National Data Service(SND)のデータカタログはSND自体が提供しているデータだけでなく、他機関が提供している医学・疫学データについても検索できるようになっている。様々な研究データについてメタデータを流通させ、多くの研究者から見つけやすくすることは世界的潮流であるように思われる。

また、もう一つの重要な取組は、共通ガイドラインの検討・策定である。共通ガイドラインという名称はいささか仰々しいが、研究データの保存と共有を円滑に進めることを促すための手引きである。研究者は研究の専門家であるが、データの保存と共有のためには様々なルールを事前に理解することは必ずしも容易なこと

ではない。諸外国では、人文・社会科学分野におけるデータの保存・共有を促進するためにデータ・アーカイブが、データの効率的な保存と共有のためのガイドラインを提供する例は珍しくない。例えば、ICPSRのGuide to Social Science Data Preparation and Archivingは1997年に初版が発行されたのち、現在は第5版が発行されている(ICPSR 2012)。UKDAは冊子体の手引きが取得可能であるだけでなく(Eynden et al. 2011)、教科書も刊行している(Corti et al. 2014)。フィンランドのFSD¹²⁾、あるいは欧州全体を包含するCESSDA¹³⁾のようにウェブ上での閲覧できるようにガイドや手引きを載せている場合もある。

これらのガイドについて、例えばデータ・フォーマットやドキュメンテーションについては、翻訳をするだけで十分な部分もある¹⁴⁾。しかしながら、インフォームド・コンセントは社会文化的な問題でもあるので、日本の状況に即した検討が必要である。データの著作権を巡る問題も、国内法の問題であるが故に、独自の検討が不可避である。データの保存と共有のために対処すべき課題を明確にし、そのガイドラインを示すことで、研究データの円滑な利活用を促すことを目指している。

このほか、オンライン分析システムの開発準備や、研究者を対象としたデータの保存と共有に関する意識調査も実施しているが、紙幅の関係で割愛する。

終わりに

最後に、筆者の個人的感想を述べることをお許し頂きたい。筆者自身がデータの保存・共有活動に周道的に関わるようになって20年近くになるが、この間、データの保存と共有に関わる環境は劇的に変化した。以前は社会科学のデータ・アーカイブは、社会科学の閉じた世界で完結していた印象があるが、ここ10年ほどで、急激に図書館情報学や人文学、さらには工学・自然科学諸分野との垣根が下がってきているように思われる。磁気媒体の研究資源を長期的に保存する、あるいは、効果的に共有するという意味では、自然科学・社会科学・人文学の各分野の課題は類似しているからである。そのことは、自然科学のデータ・レポジトリの認証を行っていたWorld Data Systemの基準と社会科学のデー

タ・レポジトリの認証基準であるData Seal of Approval とが統合されてCore Trust Seal という一つの認証基準になったことに如実に表れている¹⁵⁾。社会科学のデータ・アーカイブも、その中核的機能は維持しつつも、より広いデータ・サービスに事業を展開している。2015年にノルウェーのNorwegian Social Science Data Services が Norwegian Center for Research Data と名称を変更したことは、その変化を象徴していると思われる。人文・社会科学のデータを共有するために要

求される専門的知識・技能の水準も高いものになりつつある。筆者のような政治学者が、政治学の研究をしつつ、片手間にできるものではないというのが正直な感想である。その意味では、データ保存・共有を成功させるために一番必要なのは、それを専門的に担う人材の育成である。周辺的な形ではあるが関わっているものとして、本事業が研究データの保存・共有を担う専門家の育成について考える契機になって欲しいと考えている。

参考文献リスト

- Corti, Louise, Veerle van den Eynden, Libby Bishop, and Matthew Woollard. 2014. *Managing and sharing research data: a guide to good practice*. Los Angeles: SAGE.
- Eynden, Veerle Van den, Louise Corti, Matthew Woollard, Libby Bishop, and Laurence Horton. 2011. "Managing and sharing data: Best practice for researchers." Colchester, Essex: UK Data Archive <https://ukdataservice.ac.uk/media/622417/managingsharing.pdf>
- Inter-university Consortium for Political and Social Research (ICPSR) . 2012. "Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (5th ed.) ." Ann Arbor, MI. <https://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- LIS. 2019. "The LIS user guide: 2019 Template." <https://www.lisdatacenter.org/wp-content/uploads/files/data-lis-guide.pdf>
- 荒木俊夫 . 1994. 『投票行動の政治学：保守化と革新政党』. 北海道大学図書刊行会.
- 池内一. 1974. 『市民意識の研究』. 東京大学出版会.
- 日本社会学会調査委員会編. 1958. 『日本社会の階層的構造』. 有斐閣.
- 蟬山政道・鶴飼信成・辻清明・川原次吉郎・中村菊男編. 1955. 『総選挙の実態』. 岩波書店.

1) 日本学術振興会が公表している資料を越えた例示や説明については、前田の個人的見解として理解して頂きたい。

2) http://www.soumu.go.jp/senkyo/senkyo_s/data/index.html

3) http://www.mext.go.jp/a_menu/kagaku/kihon/1384933.htm

4) <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t242-2.pdf>

5) 本事業は、データ共有基盤の構築のために実施するものであり、実際に標本調査を行うこと自体は、事業の範囲外である。

6) 米国の全米選挙調査 American National Election Studies (ANES) や総合的社会調査 General Social Survey (GSS) は、統合ファイルが存在する。ANES については 1948 - 2012 の統合ファイルが提供されている (ICPSR 8475)。GSS の統合ファイルは複数存在するが、1972 - 2016 の統合ファイルが最新版のようである (ICPSR 36797)。

7) <https://www.jsps.go.jp/j-di/torikumi.html>

8) <http://www.ddialliance.org/>

9) <https://datacatalogue.cessda.eu/>

10) <https://datasearch.gesis.org/start>

11) <https://etsin.fairdata.fi>

12) <https://www.fsd.uta.fi/aineistonhallinta/en/>

13) <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>

14) UKDA のガイドは日本語訳が東京大学社会科学研究所から発行されている。

<https://csrda.iss.u-tokyo.ac.jp/UKDAguide.pdf>

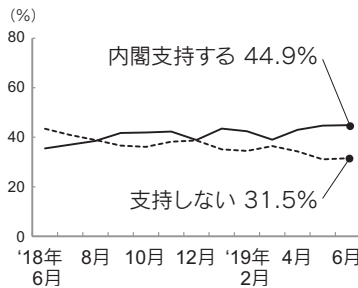
15) <https://www.coretrustseal.org/>

◇ 告 知 板

6月の時事世論調査

6月の時事世論調査によると、安倍内閣の支持率は前月比0.2ポイント増の44.9%、不支持率は0.4ポイント増の31.5%だった。

調査は全国の18歳以上の男女2,000人を対象に、個別面接聴取法で6月7日から10日に実施。有効回収(率)は1,255(62.8%)。



この時期の国内の動きは、

仁徳天皇陵、世界遺産へ：ユネスコの諮問機関である国際記念物遺跡会議(イコモス)は、仁徳天皇陵を含む「百舌鳥・古市古墳群」(大阪府)を世界文化遺産に登録するよう勧告した(5月14日)。

韓国に仲裁委設置を要請：日本企業が賠償を命じた韓国人元徴用工訴訟をめぐる、日本政府は、日韓請求権協定に基づいて仲裁委員会の設置を韓国政府に要請した(5月20日)。

食品ロス削減法が成立：食べられるのに捨ててしまう「食品ロス」の削減を目指す食品ロス削減推進法が、参院本会議において全会一致で可決、成立。同法は、政府や自治体、企業の責務のほか消費者の役割を定め、国民運動として問題解決に取り組むよう求めている(5月24日)。

トランプ大統領が国賓来日：トランプ米大統領が令和初の国賓として来日。2017年11月に続き2回目。米大統領の国賓来日は14年4月のオバマ氏以来(5月25日)。

小学生ら19人死傷：川崎市多摩区の登戸駅周辺で、私立小学校の児童ら19人が男に刃物で次々と刺され、小学6年の児童と30代男性が死亡した。犯人は、自分の首を刺して死亡(5月28日)。

旧優生保護法は違憲：旧優生保護法に基づき不妊手術を強制された女性2人が国を相手取り損害賠償を求めた訴訟で、仙台地裁は旧法の規定について憲法に違反すると判断した。一方、被害者救済の立法措置が必要不可欠であることが明白だったとは言えないとし、賠償支払いは認めなかった(5月28日)。

NHKのネット同時配信開始へ：NHKのテレビ番組を放送と同時にインターネットに流すことを認める改正放送法が成立。NHKは2020年3月にも同時配信を開始する予定で、スマートフォンやパソコンなどで番組を視聴できるようになる(5月29日)。

出生率1.42、3年連続減：厚生労働省の2018年人口動態統計によると、1人の女性が生涯に産む子どもの推計人数を示す「合計特殊出生率」は1.42で、前年を0.01ポイント下回り、3年連続で低下。出生数は、91万8397人で3年連続過去最少を更新。一方、死亡数は戦後最多の136万2482人で、人口減少はさらに加速している(6月7日)。

G20財務相会議：福岡市で開かれた20カ国・地域の財務相・中央銀行総裁会議が、2日間の討議を終えて閉幕。同会議が日本で開催されたのは初めてだったが、米中の貿易摩擦緩和の具体的な成果はなかった(6月9日)。

国外では、

部品調達で強まる包囲網：米企業との取引を禁じた米国の制裁措置を受け、中国通信機器最大手の華為技術(ファーウェイ)への部品調達や技術提供などの取引停止が広がる。日本でも同社製スマートフォンの販売延期が相次ぎ、英国の半導体設計大手が取引停止を発表した(5月22日)。

インド総選挙、与党圧勝：5年の任期満了に伴うインド下院選挙は、与党・インド人民党が単独で過半数議席を確保して圧勝。総選挙での勝利で、ナレンドラ・モディ首相の続投が確実に(5月23日)。

ルノー、ゴーン被告を告訴へ：フランスの経済・財務相は、仏自動車大手ルノーが、約13億円の不正支出疑惑をめぐる、前会長兼最高経営責任者のカルロス・ゴーン被告を告訴する見込みだと語った(6月5日)。

メイ英首相、党首辞任：イギリスのメイ首相が欧州連合離脱をめぐる混乱の責任を取り、与党・保守党の党首を辞任した。首相としても7月中に退く(6月7日)。

政党支持率 自民党は前月比1.2ポイント増の27.7%だった。立憲民主党は0.4ポイント減の3.3%、公明党は1.2ポイント減の2.6%、共産党は0.3ポイント減の1.8%、日本維新の会は0.4ポイント減の1.2%。支持政党なしは59.4%で前月より0.2ポイント減。

政党支持率 (上段：6月、下段：5月)

政党	6月 (%)	5月 (%)
自民党	27.7	26.5
立憲民主党	3.3	3.7
公明党	0.6	0.6
共産党	2.6	3.8
日本維新の会	1.8	2.1
希望の党	1.2	1.6
その他の政党	0.1	0.1
支持政党なし	0.3	0.2
支持政党なし	59.4	59.6

国民の景気感 「良くなった」は前月より1.0ポイント減の4.4%、「悪くなった」は6.0ポイント増の31.0%。時事世論景気指数は前月比12ポイント減の105と低下した。6年半ぶりに指数が110を割った。

時事世論景気指数

年	12月	13年	14年	15年	16年	17年	18年
2011年	84.1	97.2	143.8	129.1	126.0	120.0	131.9
18年	(5月)	(6月)	(7月)	(8月)	(9月)	(10月)	(11月)
	139	132	122	131	125	129	127
(12月)	19年	(1月)	(2月)	(3月)	(4月)	(5月)	(6月)
	130	126	115	114	114	117	105

暮らし向き 昨年の今頃と比べて「楽になった」は前月比2.0ポイント減の4.0%、「苦しくなった」は2.2ポイント増の21.6%だった。