

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 1
- 告知板..... 6

社会科学データを共有する制度基盤

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)

日本学術振興会では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を開始した。その背景には様々な要因があるが、研究者の世代交代とともに貴重な調査データが失われつつあるのではないかと懸念、諸外国で一般的に利用可能なデータが日本については利用できない、あるいはデータが日本語のみで提供されることで、日本を事例とした、あるいは日本が含まれる国際比較研究が減少するのではないかと懸念があった。

上述の危機感や懸念は、データの保存と共有を促進する制度基盤を整備する契機の説明としては十分である。しかし、何故、データを保存・共有するための制度基盤を構築する事業が必要なのかについては、より構造的な説明が必要であるように思われる。本稿では、日本学術振興会が推進している事業の具体的な内容ではなく、このようなプログラムが必要とされる学術的・政策的背景について、諸外国の事例も参照しつつ論ずる。なお、政策的背景を論じた本稿は、あくまで前田個人の見解であることをお断りしておく。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の具体的な内容については、改めて次号で解説する予定である。

公共財としてのデータ共有基盤

データ共有基盤は学術的コミュニティにとって公共財である。ここで公共財という用語は、単に公共の利益に資する財という意味ではなく、社会科学の理論における私的財と公共財

との対比で使っている。公共財は一旦提供されると、費用を負担しない者による消費を排除できないが故に、フリーライダーが発生し、十分な量の財が供給されないというのが一般的説明である。この考え方を社会科学データの共有に

関わる仕組みに適用するならば、コミュニティー全体の最適解と個々の研究者の(利己的)行動に基づく集合的結果とが乖離し、極めて不十分なレベルでしかデータの保存と共有が行われない状況に陥ることを意味する。本稿が関心を寄せるのは、この状況を打開するための政策的対応策である。

社会科学データと言ってもその範囲は広いが、ここでは、標本調査に基づくマイクロデータを中心に議論を進めたい。以下では、まず、社会調査の標本データを保存・共有する理由について簡単に説明する。次に、社会調査データの保存と共有とが進まない理由を検討し、最後に、対応策について筆者なりの考えを述べる。

社会調査データを保存・共有する理由

社会調査データを広く共有するべきという命題自体を否定することは今日では難しいであろう。経済協力開発機構(OECD)の方針にもあるように、とりわけ公的資金で得られたデータは、広く共有されるべきだとされている(Organisation for Economic Co-operation and Development 2007)。では、なぜ共有するべきなのであろうか。最初に確認されるべきは、社会調査データの資料的価値である。公文書や書簡が重要な歴史的な資料であるように、確率抽出された標本調査個票データは、その時代の社会の姿を記録する重要な資料である。政治学ならば無党派層の増加、人口学であれば晩婚化など、歴史的趨勢を検討する場合は、長期的なデータの蓄積があって初めて分析が可能になる。

次に、大規模標本調査には費用がかかることも重要な点である。筆者自身の経験では、計画標本3000人、回収率50%、回答者数1500人

程度の選挙調査で最低2000万円程度の予算が必要となる。これだけの規模になると、ある程度の実績を持つ複数の研究者を集めて研究プロジェクトを組織しなければ調査の遂行は難しい。調査に必要な資源が容易に得られる訳ではない以上、希少な資源を効率的に利用するために、1つの調査から得られたデータを研究者間で共有する方が賢明である。

また、学術調査の場合1回の訪問面接の聞き取りに要する時間は、30分から60分程度だと思われるが、調査票の分量によっては90分近くになることもある。回答者の時間的負担を考えると、既存データで分析可能ならば、新しくデータを収集することは控えるのが良策であろう。

聞き取り調査に要する時間が長いことは、調査票が相当数の設問を搭載していることを意味する。特に継続調査の場合は、過去に利用された設問を繰り返すと同時に、時宜に応じた設問を加えるために、どうしても質問の数が多くなる。搭載された質問には様々な組み合わせがあり得るので、データを収集した研究者が想定しなかった観点からの分析は十分に可能である。その意味では、調査を実施した研究者以外がデータを分析する意義は大きい。

以上は、従前から言われてきたことの確認である(例えば、佐藤・石田・池田 2000)。その一方、近年では研究の透明性を高めるという観点からもデータの保存と共有の必要性が強調されるようになっている。筆者が専門とする政治学では、2015年に当時カリフォルニア大学ロサンゼルス校の大学院生と著名な政治学者による共著論文が、ねつ造されたデータに基づいたものであったことが大きな問題となった(Noorden 2015)。それ以前から、政治学分野における統計分析の再現性を高める必要があることは指摘

されていたが(King 1995)、近年では論文に利用した研究データの共有と研究の透明性を一体として理解することが強調されるようになった(Lupia and Elmanr 2014; Lupia and Alter 2014)。査読誌の掲載要件として、再現性を担保するデータとプログラム・ファイルの提出と公開を義務づけるべきであるという議論もなされている(Dafoe 2014)。

データの保存・共有が進まない理由

理屈の上では、多くの研究者がデータを共有することの意義は理解しているはずである。では、なぜデータの保存と共有が理想よりも進まないのでしょうか。ここでは、10年程前にデータインフラストラクチャーの問題を検討したフィンランド(Kuula and Borg 2008)とスウェーデン(Carlhed and Alfredsson 2009)の報告書を参照しつつ、筆者なりに問題点を整理したい。筆者が理解する限り、この二つの報告書で検討されている内容は今日の日本の状況との共通するところが多い。

1つ目は、研究者文化の問題である。データの共有が進まない原因の一つには、研究者によっては、努力の結晶とも言うべきデータへの所有権意識(あるいは愛着)が強く、外部の研究者とデータを共有したがないことがあるように思われる。外部の研究者とデータの共有が進まないことは、長期的にはデータが散逸する危険性を高めるように思われる。

2つ目は、データの所有権・著作権についての考え方である。データの所有権や著作権についての考え方を議論することは筆者の能力を超えた問題であるが、権利関係についての理解が確定しないことは、どのような手続きを経ることでデータを当初の研究チームの範囲を超えて

共有できるのかを不明瞭にする。結果としてデータを共有することのハードルを高くしているように思われる。また、著作権の考え方を援用すると、一次データ収集者が他界している場合、遺族に対して許可を求める必要が生ずる。複数の著作権者が他界している場合、許諾を得るために必要な労力は決して小さなものではない。

3つ目は誘因の欠如である。研究者の評価は刊行された研究業績で定まると考えることが従来の一般的考え方である。ならば、データの共有に要する手間と暇を研究者が自発的にかけると考えることは、あまりにも楽観的に過ぎるであろう。第三者がデータを正しく利用するために必要な記録を適切に残す作業には相当の手間と暇がかかる。無論、本人にとっても、長期的には記録を残すことは重要なはずであるが、少なくとも短期的には不要な作業である。記録をきちんと残す誘因を提供するか、あるいは、記録を残さざるを得ない制度的な工夫が必要になるであろう。

4つ目は技術的な問題である。長期的なデータ共有のためにはデータが適切に保存されていることが重要である。研究者個人がデータを保存している場合、長期的には、磁気媒体の劣化や、機械的な事故でデータファイルが失われる可能性がある。また、データが特定の商用ソフトウェアに依存しないことも重要である。広く使われているソフトウェアが10年後にも利用可能であるという保証はない。また、ロングセラーのソフトウェアでも、30年以上前のバージョンで作成されたファイルを、現在のバージョンで読めるかは相当疑問である。

5つ目は、プライバシーを保護するための匿名化の問題である。調査に協力して下さった方々のプライバシーを守るのは当然である

が、具体的にどのようにデータを加工すれば十分な匿名化と言えるのかは必ずしも自明ではなく、専門家でも判断に迷うことがある。例えば、2000年代前半の段階では、社会調査データに都道府県程度の地理的範囲の情報を残しておくことは一般的であった。しかしながら、情報技術の発達、および、様々な種類のデータの入手が容易になったことから、複数のデータの組み合わせによって個人が特定される危険性が高まったことが指摘されている。今日では、二次分析用のデータファイルには都道府県単位の情報をデータファイルに残さずに共有することが増えている。技術的变化に応じて、匿名化に必要な処理方法が変化するのである。都道府県変数の削除はかなり簡単な例であるが、秘匿処理に必要な技術的变化を個々の研究者が理解し、作業を行うことは必ずしも容易ではない。

政策的に何が必要とされるか

以上、データの保存と共有が進まない理由について整理して論じてきか、裏を返すと、これらの理由に対する対応がある程度できれば、状況は改善するのではないと思われる。

上述した1つ目の研究者文化については、長い時間をかけて変えていくしかない。しかし、筆者の個人的印象では、共有されている大規模データを分析して学位論文を執筆した若い研究者はデータの保存と共有に前向きであるように思われる。その意味では、適切な政策的対応がなされれば、研究者の意識も少しずつ変化するであろう。

2つ目のデータの所有権・著作権については、少なくとも事前にデータを共有する方針及びその方法について研究チーム内で合意形成がなされていれば、問題を回避できるはずである。また、順

番は前後するが4番目の技術的問題、5番目の秘匿処理の問題についても、事前に十分な計画が練られていれば、費用や人員の問題を除くと、対応はそれほど難しいものではないと考えられる。

以上述べてきたことはどちらかと言えば、技術的問題であるが、三つ目の研究者の誘因については、いかなる対応が可能なのであろうか。その方法の一つとして、社会科学分野においてもⁱ、各国と同様にデータ管理計画(Data Management Plan)を研究助成申請段階で義務づけ、データの適切な保存・管理を促すことが挙げられる。これは研究者側から見ると誘因というよりは規制であるが、データのライフサイクルの各段階における作業について周到に計画を立てておけば、研究データの保存と共有をより効率的に実現できるはずである(Corti et al. 2014)。また、費用面では、データの加工や、メタデータの準備に関わる費用を研究費から支出できるようにすることで、研究者をデータの保存と共有へと促すこと出来ると思われる。

一方、研究コミュニティにおける評価に関連して、研究成果だけではなく、データを他の研究者と共有すること自体を評価する仕組みの整備も必要であろう。最近では、データの適切な引用を促すと同時に、データの引用を論文の引用と同様に評価するべきと言う議論もなされている(Altman and Merce 2014)。また、データの内容そのものを紹介する論文を掲載するResearch Data Journal for the Humanities and Social Sciencesのような専門査読誌も発行されるようになっていく。

以上、研究者にデータの保存と共有を促すための政策的対応について簡単に論じた。データを共有するための予算を準備することで、あるいはデータの共有を評価する誘因を提供するこ

とで、一定程度研究者個人個人の行動に影響を与えることができるものと思われる。ただし、データの長期的保存や、広範な提供は、個々の研究者や研究室が良くなし得るものではない。データの保存と共有は、研究者側の誘因の問題だけではなく、広くデータの保存と共有を支えるインフラストラクチャーの整備と表裏一体に考える必要があるだろう。

欧米諸国においては、政府機関あるいは助成機関が、データの保存と共有を担う組織を設置している、あるいは財政的に支援していること

が珍しくない。例えば、イギリスのUKDSⁱⁱ、スウェーデンのSNDⁱⁱⁱ、そしてノルウェーのNSD^{iv}のウェブサイトを見ると、学術データの保存と共有を支援するだけではなく、データの保存・共有の実践への助言、ガイドラインの提示、データの共有を促進するための研究活動などが組織の目的として例示してある。日本においても、データの保存と共有を橋渡しする組織、そしてデータの保存と共有について適切な助言を与えることのできる専門的人材の育成が必要であるように思われる。

参考文献リスト

- Altman, Micah, and Crosas Merce. 2014. "The Evolution of Data Citation: From Principles to Implementation." IASSIST QUARTERLY 37:62-70.
- Carlhed, Carina, and Iris Alfredsson. 2009. "Swedish National Data Service's Strategy for Sharing and Mediating Data." IASSIST QUARTERLY 32:30-39.
- Corti, Louise, Veerle van den Eynden, Libby Bishop, and Matthew Woollard. 2014. Managing and Sharing Research Data: A Guide to Good Practice. SAGE.
- Dafoe, Allan. 2014. "Science Deserves Better: The Imperative to Share Complete Replication Files." PS: Political Science and Politics 47 (1) :60-66.
- King, Gary. 1995. "Replication, Replication." PS: Political Science and Politics 28 (3) :444-52.
- Kuula, Arja, and Sami Borg. 2008. Open Access to and Reuse of Research Data - The State of the Art in Finland. Tampere: Finnish Social Science Data Archive (FSD) .
- Lupia, Arthur, and George Alter. 2014. "Data Access and Research Transparency in the Quantitative Tradition." PS: Political Science and Politics 47 (1) :54-59.
- Lupia, Arthur, and Colin Elman. 2014. "Introduction." PS: Political Science and Politics 47 (1) :19-42.
- Noorden, Richard Van. 2015. "Retracted Gay-Marriage Study Debated at Misconduct Meet-up." Nature 522 (June) :14-15.
- Organisation for Economic Co-operation and Development. 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD Publishing.
- 佐藤博樹・石田浩・池田謙一. 2000.『社会調査の公開データ：2次分析への招待』東京大学出版会.

i 日本でも科学技術振興機構や日本医療研究開発機構はデータ管理計画を義務づけている。

ii <https://www.ukdataservice.ac.uk/about-us.aspx>

iii <https://snd.gu.se/en/about-us>

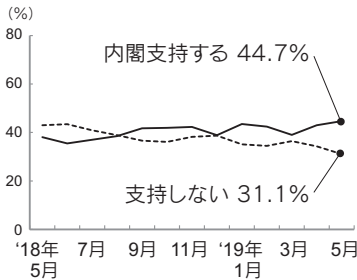
iv <https://nsd.no/nsd/english/index.html>

◇ 告知 板

5月の時事世論調査

5月の時事世論調査によると、安倍内閣の支持率は前月比1.7ポイント増の44.7%、不支持率は3.2ポイント減の31.1%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5月10日から13日に実施。有効回収(率)は1,256(62.8%)。



この時期の国内の動きは、

2024年度に紙幣刷新：麻生財務相は、1万円札、5千円札、千円札のデザインを一新すると発表した。2024年度上期から流通させる方針(4月9日)。

ブラックホールの撮影成功：国立天文台などが参加する日米欧などの国際共同研究プロジェクトによる、世界各地の電波望遠鏡による観測で史上初めてブラックホールの撮影に成功した(4月10日)。

五輪担当大臣が辞任：桜田五輪担当相が失言の責任を取って辞任。2012年の安倍首相の政権復帰以降、閣僚の辞任は8人目(4月10日)。

WTO最終審で日本が逆転敗訴：2011年の東京電力福島第1原発事故後、韓国が福島など8県産の水産物輸入を禁止している問題で、世界貿易機関の上級委員会は、禁輸を不当とみなした紛争処理小委員会の判断を取り消し、事実上、日本の逆転敗訴とした(4月11日)。

総人口、8年連続減：総務省の2018年10月1日現在の日本推計人口によると、総人口は前年比26万3000人(0.21%)減の1億2644万3000人で8年連続の減少。減少数と減少率は過去最大。生産年齢人口の割合も59.7%と、

1950年以降過去最低(4月12日)。

統一地方選後半戦：平均投票率は、59市長選47.50%、283市議選45.57%など最低を更新。また、同日に実施された衆院大阪12区と沖縄3区の両補欠選挙で自民党候補が敗れた(4月21日)。

日米貿易交渉を加速：安倍首相は、トランプ大統領とホワイトハウスで会談し、新たな日米貿易協定交渉の早期合意に向けた交渉の加速で一致(4月26日)。

「令和」へ改元：天皇陛下が202年ぶりに退位の儀式を行い、4月30日をもって平成が終わる。皇太子徳仁親王殿下が翌5月1日第126代天皇に即位し、元号が令和に改まった(5月1日)。

民間ロケット、初の打ち上げ成功：北海道大樹町のベンチャー企業が小型観測ロケットMOMO3号機を同町の発射場から打ち上げ、宇宙空間に到達させた(5月4日)。

国外では、

イスラエル総選挙：右派のネタニヤフ首相率いるリクードを中心とする右派勢力の過半数獲得がほぼ確定し、ネタニヤフ氏の5期目続投が確実に(4月10日)。

英EU離脱、10月末まで再延期：欧州連合の臨時首脳会議で、英国のEU離脱期限を最大で10月末まで再延期することで合意した(4月11日)。

ノートルダム大聖堂で大火災：パリ中心部にある観光名所、ノートルダム大聖堂で大規模な火災が発生。8時間以上燃え続け、屋根や尖塔が焼け落ちた(4月15日)。

スリランカ連続爆発：スリランカで発生した教会やホテルなどを狙った8件の爆発により、死者数が200人以上、負傷者数も450人超にのぼった(4月21日)。

口朝首脳会談：プーチン大統領と金正恩朝鮮労働党委員長は、ウラジオストクで初の首脳会談を行い、首脳間の関係構築を図った(4月25日)。

北朝鮮が飛翔体発射：北朝鮮の東岸から北東に向けて短距離の飛翔体数発が発射され、日本

海方向に70～200km飛行した(5月5日)。

モスクワ空港でロシア機炎上：モスクワのシェレメチエボ空港でエアロフロート・ロシア航空の旅客機が緊急着陸して炎上。落雷が原因とみられ、乗客・乗員78人のうち41人が死亡した(5月5日)。

米、対中関税引き上げ：米中両国は、ワシントンで閣僚級の貿易協議を開催したが、物別れに。約22兆円相当の中国製品への追加関税を10%から25%に引き上げ、譲歩を迫ったが中国は拒否(5月10日)。

政党支持率 自民党は前月比0.2ポイント減の26.5%だった。立憲民主党は0.6ポイント減の3.7%、公明党は0.1ポイント減の3.8%、共産党は0.2ポイント増の2.1%、日本維新の会は0.3ポイント増の1.6%。支持政党なしは59.6%で前月より1.1ポイント増。

政党支持率 (上段：5月、下段：4月)

自民党	立憲民主党	公明党	共産党	日本維新の会	希望の党	その他の政党	支持政党なし
26.5	3.7	0.6	3.8	2.1	1.6	0.1	0.2
26.7	4.3	0.6	3.9	1.9	1.3	0.1	0.1

国民の景気感 「良くなった」は前月より0.2ポイント増の5.4%、「悪くなった」は1.4ポイント減で25.0%。時事世論景気指数は、前月比3ポイント増の117。

時事世論景気指数

2011年	12年	13年	14年	15年	16年	17年	18年
84.1	97.2	143.8	129.1	126.0	120.0	131.9	130.0
18年 (4月) (5月) (6月) (7月) (8月) (9月) (10月)							
131 139 132 122 131 125 129							
19年 (11月) (12月) (1月) (2月) (3月) (4月) (5月)							
127 130 126 115 114 114 117							

暮らし向き 今年の今頃と比べて「楽になった」は前月比0.1ポイント増の6.0%、「苦しくなった」は1.7ポイント減の19.4%だった。