JAPAN OPEN SCIENCE SUMMIT 2019
5月28日10：00～11：30
E1 人文・社会科学におけるオープンサイエンス

人文学・社会科学のデータ共有基盤について

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
ymaeda@iss.u-tokyo.ac.jp

前置き

- 社会科学データと言ってもその内容は多様である。
  - 無作為抽出に基づく標本調査データ
  - 経済時系列データ
  - 国家や紛争を単位とした国際政治データ
  - 政治家の声（音声）や選挙ポスター（画像）
  - 実験データ
- ここでは、主に標本調査データの共有基盤について議論する。
- JSPSが推進している活動の具体的内容ではなく、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論ずる。
- 本日の話はあくまで個人的見解です。

集合行為問題

- 社会全体の最適解と個々人の利己的行動との関係
- 市場メカニズムに委ねると社会全体としては十分な公共財が供給されない。
  - フリーライダーが発生する
- 解決策は二つ
  - 政府・独占的権力による公共財提供
  - 利己的行動と公共財提供が連動する制度設計

社会調査データを保存・共有する理由（1）

- 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。
- 大規模標本調査には費用がかかる
  - 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。
- データの収集・整理には相当の時間がかかる。希少な資源を効率的に利用するためには、データを共有する方が賢明。

社会調査データを保存・共有する理由（２）

- 多くの調査データは異なる観点からも分析可能
  - 費用および回答者への負担を考えると、既存データで分析可能ならば、新しくデータを収集する必要はない（収集しないほうが良い）。
- 社会科学における再現性を保証する必要がある。
  - Michael LaCour事件（データのねつ造）
  - 社会科学の主要ジャーナルは再現に必要な資料（データとコードの）提出を査読の条件として要求
    - DA-RT (Data Access and Research Transparency)

データの保存・共有が進まない理由（1）

- 文化の問題
  - 所有権意識が強く、外部の研究者とデータを共有したがらない（長期的にはデータが散逸する危険性）。
  - 他の研究者が自分のデータに興味を持つと思わない。
- 誘因の不足
  - 研究者は研究業績で評価が定まる。
  - データの共有は潜在的競争相手を利する行為
  - データを共有するには手間と暇がかかる。

データの保存・共有が進まない理由（２）

- 技術的問題１（情報技術）
  - データ共有に必要なメタデータの内容がわからない
  - 適切な機械可読ファイルの形式がわからない。
  - データ・ファイルが劣化して読めなくなった。
- 技術的問題２（法的・倫理的側面）
  - データを共有することを事前に検討せずに調査に着手
  - 他者のデータを加工した場合、共有可能なのか不明
  - データの所有権について理解がまちまち
  - 秘匿処理の手順や基準がわからない。

データの保存・共有が進まない理由（３）

- 制度・組織の問題
  - 誰がデータ共有のための時間・費用を負担するのか。
  - データの共有を橋渡しする組織・制度の不在
  - データの共有について指導・助言する組織の不在

- 10年ほど前にフィンランドとスウェーデンについてそれぞれまとめられた報告書の内容は、現在の日本の状況と共通している。

ヒトに関するデータの難しさ　プライバシー

- 調査に際して、次のような説明を行うことが多い。
  「回答はすべて統計的に処理されますので、回答の内容から個人が特定されたりすることはありません。またもちろん、回答内容が第三者に伝わるようなことはありません。」
- 住民基本台帳の閲覧について
  - 閲覧者、調査名そして閲覧範囲（丁目）が公表される。
- 学術利用に限るという慣行が根強い
  - 民間企業の研究員や一般市民の方に対する利用の道は閉ざしていることが一般的。完全にオープンではない。

政策的に何が必要とされるか

- 文化の問題（研究者コミュニティー）
  - 長い時間をかけて変えていくしかない
- 誘因の不足（研究助成団体・研究者コミュニティー）
  - データ・メタデータの準備を研究費に積算することを認める。
  - データの引用を制度化する。
  - データの共有・DMPを研究助成の条件にする
- 技術的問題（公的機関・研究助成団体）
  - 基準やガイドラインを信頼できる機関が提示することで緩和
- 制度・組織の問題
  - 国・公的機関が組織的努力を支援することで対応可能

諸外国の政策的対応

- OECD, "OECD Principles and Guidelines for Access to Research Data from Public Funding" 2007

- ESRC “ESRC Research Data Policy” 2018
- NSF Social, Behavioral and Economic Sciences Directorate “Data Management for NSF SBE Directorate Proposals and Awards” 2018

  他国も当然 Data Policy はあるが、社会科学に特化したものは多くないかもしれない。

欧米の社会調査・統計調査データアーカイブ（例）

- アメリカ（1962）　Inter-University Consortium of Political and Social Research (ICPSR) https://www.icpsr.umich.edu
- ドイツ（1960/1986） Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) https://www.gesis.org/en/home/
- オランダ （1964/2005） Data Archiving and Networked Services (DANS) https://dans.knaw.nl/en
- イギリス (1967) UK Data Archive (UKDA) http://www.data-archive.ac.uk/ (2011) UK Data Service https://www.ukdataservice.ac.uk/
- ノルウェー（1971/2015）NSD - Norwegian Centre for Research Data http://www.nsd.uib.no/nsd/english/
- フィンランド（1999) Finnish Social Science Data Archive - FSD http://www.fsd.uta.fi/en/
- スウェーデン（？/2008）Swedish National Data Service – SND
- https://snd.gu.se/en

欧米主要アーカイブが実施している活動
（Funding Agency や National Academy が求める活動）

- 質の高いデータ（とメタデータ）の保存と共有
- データの二次利用の促進
- データ利用に関する学習機会の提供
- 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活動
  - 方法論、プライバシーの保護、研究倫理等も含まれる。

例えば、UKDS, SND, NSD

JSPSの活動

- 昨年度、公募で４機関を選定。
  - 大阪商業大学、慶應義塾大学、東京大学、一橋大学
- 各機関は所蔵データと強みを生かして、データの共有化・国際化を推進。
- JSPSは、各機関と協力しつつ活動。
- 本年度から来年度にかけて、データ共有に関するガイドライン／手引きを検討。

（一般論として）政策的効果

- 研究の発展
  - データの利用可能性が高まると、研究の数は増える。質も高まる。
  - ただし、効果が明示的に出るまでは時間がかかるだろう。
- 研究者養成
  - データが利用できる、できないは、研究内容だけではなく、研究者の進路にも大きな影響がある。

データの保存と共有を支える組織・人材の必要

- データの共有活動を支える組織とその組織を担う専門的能力を持つ職員が必要。
- 諸外国ではデータ・アーカイブの常勤職員が、手引きやガイドラインについて調査・研究
- 諸外国ではデータの保存・共有に関する専門職のキャリアパスがある程度成立（政府機関データ部門・学術調査機関・レポジトリ・図書館等の間を異動）。
- 研究者が片手間でやるものではない。
- 大学の一部局がやるものでもない。

参考文献リスト

- Carlhed, C., & Alfredsson, I. (2009). Swedish National Data Service's Strategy for Sharing and Mediating Data : Practices of Open Access to and Reuse of Research Data – The State of the Art in Sweden 2009. Presented at the IASSIST/IFDO 2009.
- Kuula, A., & Borg, S. (2008). Open Access to and Reuse of Research Data - The State of the Art in Finland.
- Corti, L., Van den Eynden, V., Bishop L., & Woollard M. (2014) Managing and Sharing Research Data: a Guide to Good Practice, Sage Publication.
- Lupia, A., & Elman, C. (2014). Openness in Political Science: Data Access and Research Transparency: Introduction. *PS: Political Science & Politics, 47*(1), 19-42.