

《資料》

## 海外データアーカイブの動向 2

### —IASSIST 年次大会の報告から—

朝岡 誠

高橋 かおり

#### 【要旨】

社会調査データは今後の社会の発展に寄与する公共財であり、広くデータが利活用される環境を整備する必要がある。本稿では 2018 年 5 月に開催された IASSIST 年次大会で報告された内容をもとに海外のデータアーカイブで行われている先進的な取り組みを紹介する。また国立情報学研究所オープンサイエンス基盤研究センター（NII RCOS）と行った共同プロジェクトについてポスター発表を行ったので、その様子について報告する。

キーワード：データアーカイブ，データ利用，データ共有

#### I はじめに

社会調査は複雑な現在社会の状況を見通すための重要なツールであり、今後の社会の発展に寄与する公共財である。現在、日本の学术界で進展しているオープンサイエンスの理念はこうした社会調査データのありかたを的確にとらえたものであり、現在それに向けた環境整備が急ピッチで進められている。その中でデータアーカイブは広く調査データの利活用を促進するための役割を果たす必要がある。

それでは、調査データの利活用を促進するためにデータアーカイブはどのような体制を整備し、そしてどのようなサービスを提供すべきなのだろうか。これらの問題関心のもと、著者らは海外の先進的なデータアーカイブの取り組みを確認することを目的に、2018 年 5 月にモントリオール（カナダ）で開催された IASSIST（International Association for Social Science Information Services and Technology）年次大会に参加した<sup>1)</sup>。

社会科学に関わる情報サービス・技術をテーマとする IASSIST は、“Foster and promote a network of excellence for data service delivery”，“Advance infrastructure in the social sciences”，“Provide opportunities for collegial exchange of sound professional practices”の 3 つを活動目的に掲げ、データアーカイブ業務に従事／関係する専門家が現在行っている先進的な取り組み、および関連した動向を発表する場として機能している。報告されるコンテンツも多岐にわたり、多くの成果発表が行われた。

本稿では、前田・朝岡(2017)に引き続き、IASSIST にて開催されたワークショップや報告

された内容をもとに海外のデータアーカイブで行われている先進的な取り組みを紹介し、今回 RUDA が行ったポスター発表について報告する。また、今回の IASSIST では RUDA の紹介がてら国立情報学研究所オープンサイエンス基盤研究センター (NII RCOS) との共同研究の成果報告を行った。現在 RUDA では前田・朝岡(2018)で紹介したとおり、横断的検索システムの構築に向けて NII RCOS とデータ公開環境基盤の整備に向けた研究プロジェクトを行っており、RUDA と SSJDA のメタデータのいくつかを DDI2.5 ベースのメタデータに変換し、横断的検索システムにアップロードし、試験的に運用した。この共同研究についても日本のアーカイブの動向として紹介する。

## II IASSIST における研究動向

### 1. データ・ライブラリアンの教育

筆者のうちの1人(高橋)は実務面で RUDA の運用管理に従事しているが、このような立場は欧米においてはデータ・ライブラリアン (data librarian) として職務の確立がされている。すなわち、図書館の司書(ライブラリアン)が書籍貸出や運用管理、レファレンスの相談に乗るように、データについての貸出や運用管理、活用相談に乗るのがデータ・ライブラリアンなのである。

このようなデータ・ライブラリアン向けのワークショップの紹介が「Putting the CURATE Model in Practice: Training for Data Curation」のパネルセッションでなされた。ここでは、2017年12月に行われたデータ・キュレーション・ワークショップの様子とそのフィードバックを取りあげる。ここでいうデータ・キュレーションのゴールは「本来の目的を越えてデータを活用するための方法を用いて調査データを準備・維持し、必要なデータを確実にそろえ、長期間にわたり再利用と引用可能性を促進する」ことである。そのためには、メタデータの扱いや文書管理、アクセスや保管方法など様々な工夫や共通ルーツの遵守が必要である。

この調査グループが示すデータによれば、全米3分の2の大学図書館で、データ・キュレーション・サービスがすでに設置されている。そしてこの動きを全米に広げるため、Data Curation Network (DNC) が2017年に設立された<sup>1)</sup>。DNCは独自の「CURATEモデル」を提唱し、データの保存と利活用を専門的に担う人材の育成や、それに興味を持つ人々への知識の共有を行っている。

社会調査をめぐるあり方として、調査設計や実査が得意な研究者と、分析が得意な研究者は必ずしも一致しない。また、院生や若手研究者などは資金やネットワークの面から大規模調査に関われないことは往々にしてある。日本においても今後、図書館に付随する形でデータ利用について専門的に助言や指導を行える人材が一層求められるだろう。私的に手に入るデータを利用するだけではない、より目的に合致したデータを得るためのゲートキーパーとして、データ・キュレーションのできるデータ・ライブラリアンの存在は今後

日本で必要とされる役割である。

## 2. データ活用のための学生教育

データサイエンス教育というと分析手法が強調されがちである。しかし海外、とりわけ欧米では書誌情報や実験データ、地理情報などについては、データ・ライフサイクル (data lifecycle) に基づいた運用が進みつつある。そのなかで個人でのデータ活用実践として、データのアーカイブ化とその活用を学生に教える試みがなされている。それはいわゆる質問紙調査のデータセットだけではなく、実験データの情報なども含まれる。単なる情報をいかにして研究データにするのか目的にある。「Instruction: Data Management Skills」のセッションではこれがテーマとなっていた。

アメリカ合衆国の大学のシラバスを分析した Ashley Doonan と Evan Cosby の発表 (“A Follow-up on Data Management and Data Sharing Training in Graduate Education in the Social Science”)によれば、データの分析 (analysis) と解釈 (interpretation) に触れたシラバスは4分の3にのぼるという。他方、管理 (management) と共有 (sharing) についても扱う大学はわずかながらある。歴史学と政治科学、地理学においてはマネジメントへの注目が高いとはいえ、全体では正規授業でデータ管理や共有について学ぶ機会はアメリカにおいても多くはない。

実践例としては、オンタリオ農業大学における取組が紹介された (Michelle Edward and Carol Perry, “Preserving the agricultural data story at the Ontario Agricultural College”)。農学に関する実験データや情報を、いかにリサーチ・データ・ライフサイクルに載せるのかということを講義とワークショップを通じて学生に伝授していた。ファイル名をつける法則や、管理するための表の作り方を通じて、情報を「今の自分」が理解できるものではなく、将来の自分や他の学生も活用できるデータとする手法が紹介された。これは学生向けのプログラムであったが、教員にとっても革命的な取り組みだったという。

初歩的な取り組みであるが、このようなやり方は意識しなければ身に付かない。データアーカイブへの寄託や二次利用といった取り組みのみならず、個人がいかにデータを管理し、利活用できるようにするのもまた、データサイエンス教育において重要な論点である。

## 3. データアーカイブにおけるプライバシーの問題

データ公開やシェアが進むにつれて、取扱注意のデータ (sensitive data) をどのように扱うのかもまた共通課題となりつつある。これは後述する質的データにおいてより問題となる。「Sharing Sensitive Data: Interpreting Legal and Policy Framework」のセッションでは、南アフリカ (Council for Scientific and Industrial Research)、アメリカ (Quantitative Data Repository, Syracuse University: QDR)、フランス (GENES) の取り組みがそれぞれ紹介された。

この議論の大前提として、取扱注意のデータについては完全に匿名化することが最適解

ではないという共通認識を持つ必要がある。そのうえでそれぞれの取り組みから、各機関、あるいは公開形態（報告書、学術論文、書籍出版など）による基準のずれが見えてきた。さらに、公的統計と社会調査のデータの扱いの違いなども問題になる。

例えば QDR の発表では、学内倫理委員会（IRB）は参加者の安全や守秘義務を守ることが前提にあるのに対し、データアーカイブ（リポジトリ）はデータの安全性を守ることに重きを置くという各主体のミッションの違いが示された。その際、学内倫理委員会とデータアーカイブの協働が鍵になる。ただし、QDR がアメリカで公的助成を一定程度受けている 50 機関を対象に学内倫理委員会のガイドラインを分析したところ、現状ではデータ・シェアリングの基準まで明記しているところはほとんどなかった。今後は、協力者のプライバシーや人権に配慮しながら、データとしても共有可能にするための基準、すなわち将来的にデータアーカイブに収録される可能性について明記することが推奨される。

日本においても日本社会学会による「日本社会学会倫理綱領（2005 年初版発行）」<sup>2)</sup>とそれに伴う「日本社会学会倫理綱領にもとづく研究指針（2006 年初版発行，2016 年改訂）」<sup>3)</sup>、あるいは社会調査協会による「倫理規定（2009 年施行）」<sup>4)</sup>が公布され、社会学者個人による倫理基準の順守意識の高まりや、学内倫理委員会での審議の必要性が説かれるようになっていく。例えば社会調査協会の「倫理規定」においては二次利用についての言及があるが、例外としての目的外使用の記述に留まる。今後オープンデータ化がより進む中で二次利用を前提にした倫理規定や綱領への書き換えも必要であろう。また、質的データについても厳重なデータ管理が前提とされる中で、研究者個人のみの財産とせず学術界としての財産にするための取り組みはあつてしかるべきである。加えて学内倫理委員会が医療系、心理学系の実験や臨床を前提に基準や判断が行われている現在、社会調査における基準をどのように確立していくのかもさらなる問題である。

#### 4. 質的データアーカイブの活用に向けたネットワークの試み

量的調査や量的データへの関心が高い IASSIST においても質的データの問題に特化したネットワークがある。IASSIST Qualitative Social Science & Humanities Data Interest Group (QSSHDIG)の短時間のミーティング（Birds-of-a-Feather Discussions）では、2016 年にできたこのネットワークの趣旨説明がなされた<sup>5)</sup>。会場には 20 名弱が集まり、一言ずつの自己紹介の時間も取られた。当日参加者の多くはライブラリアンの方々であった。

主催者側からは過去のワークショップの取り組みとネットワークへの参加の紹介があった。他のセッションにおいても質的データへの配慮や関心があった。取り組みとしてはまずは既存の研究発表や論文情報の集約とリストアップが主な活動であったが、今後の展開を注視していく必要があろう。

#### 5. DDI4.0 の動向

DDI とは社会調査データの国際標準として広く使われているメタデータ基準である。現

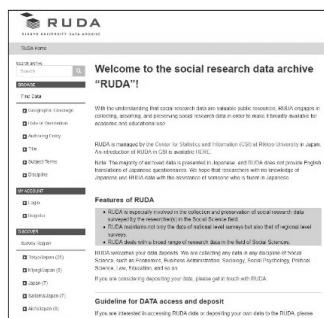
行では、DDI には大きく分けて 2 つのバージョンが存在する。一つは調査データを収集し、公開するデータアーカイブのために設計された DDI Codebook であり、もう一つは調査を実施している研究者が調査記録を残すために設計された DDI Lifecycle である。DDI4.0 はこの 2 つのバージョンと互換性を持ち<sup>6)</sup>、RDF(Resource Description Framework)や XKOS というオンライン集計指向のシステムに対応している。昨年 DDI4.0 のプロトタイプが公開され、今後の国際標準メタデータ規格となるかと思われるが、実際に運用されるのはまだ先のことになりそうである。

### III 日本のアーカイブの動向

今年度の IASSIST では RUDA と NII RCOS の共同研究の成果をポスターと口頭発表にて報告した。ポスターの内容は次頁のとおりである。今回の報告では日本の社会調査データアーカイブの現状を示すために RUDA の現状について触れ、日本の小規模アーカイブの現状について説明し、アーカイブ利用者の広範のニーズに応えるために NII RCOS が開発した研究データ基盤を用いた国内アーカイブ間相互検索システムの導入を試みた経緯を説明した。そして NII RCOS が開発した研究データ基盤が 3 つに分かれ、この 3 つの中の公開基盤にそれぞれのアーカイブのメタデータをアップロードすることで各アーカイブのポータルサイトとして活用することと、そのためには各アーカイブのメタデータを RUDA で考案したメタデータ基盤<sup>7)</sup>に変換して運用することについて説明した。

# A Pilot Study towards Cross-Searchable Social Science Data Archive in Japan

Makoto Asaoka (Rikkyo University)  
 \*corresponding author: asaoka@rikkyo.ac.jp  
 Yutaka Maeda (Kwansei Gakuin University)  
 Miho Funamori (National Institute of Informatics)  
 Masaharu Hayashi (National Institute of Informatics)  
 Kazutsuna Yamaji (National Institute of Informatics)



RUDA (<https://ruda.rikkyo.ac.jp/>)

## Rikkyo University Data Archive(RUDA)

In March 2010, Rikkyo University established Center for Statistics and Information (CSI) to support statistical education and survey research. Rikkyo University Data Archive(RUDA) a data archive of quantitative social survey data maintained by CSI in April 2011.

### Data Collection

- 55 data collections available.
- Collects mainly social survey data conducted by researchers of social sciences.
- 3~7 data collections added per year.

### Users

- 484 registered user.
- Used mainly for secondary analysis, social survey, and statistical training.
- 10~20 data downloaded per year.

### Management

- Ruda uses DSpace 5.5 as repository system.
- metadata is described under (qualified) Dublin Core scheme.
- 2 assistant professors(fixed-term) are engaged staff of RUDA.

## Pilot study for Cross-searchable Social Science Data Archives in Japan

RUDA and National Institute of Informatics Research for Open science and Data Platform (NII RCOS) collaborate to establish a cross-searchable infrastructure between social science data archive in Japan using Research Data Infrastructure for Open Science developed by NII RCOS.

### Research Data Infrastructure for Open Science

The Infrastructure consists of three platforms

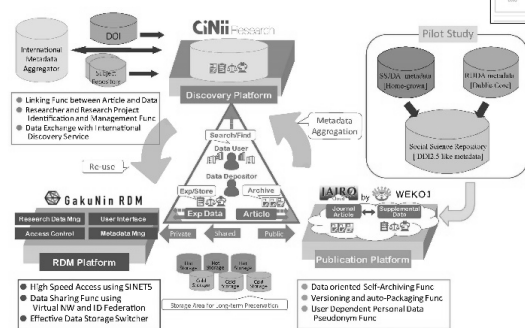
- RDM Platform(GakuNin RDM) : Manage and share research data within the research project member. so-called "Japanese version Open Science Framework."
- Publication Platform(WKOC3) : Publish and store research data and related materials that researchers can decide to make openly accessible.
- Discovery Platform(Cinii Research) : Discovery academic resources from institutional repositories (such as WEKO3), and other open databases.

### Pilot study

- Construct "Social Science Repository" on WEKO.
- Unify metadata used in each data archive into a common format (DOI-like).
- 2 data archives have posted metadata into Social Science Repository  
 SSJDA(@ The University of Tokyo): 3datasets metadata  
 RUDA: 10 datasets metadata

### Issues

- How should we construct thesaurus for effectively search?
- In Japan, a thesaurus that efficiently describes survey data is not shared yet.
- How do we construct the regulations regarding rights of (meta)data?
- It is not clear on who hold rights of metadata ( depositor or data archive)?



Research Data Infrastructure for Open Science (NII RCOS)



Social Science Repository in WEKO (Pilot study)

パイロットスタディとして SSJDA に協力していただき、Social Science Repository というアカウントに RUDA のメタデータを 10 件、SSJDA のメタデータを 3 件アップロードして試験運用したが、問題なく運用することができたので、今後は他のアーカイブにも協力を呼びかけていく予定であるが、各アーカイブ間で研究ジャンルやキーワードが統一されていないために、いくらか調整が必要である点や NII RCOS にアップロードされたメタデータの権利についてどのように扱うのが今後の問題であると説明した。

今回のポスター発表は日本のアーカイブに興味を示す、特に NII RCOS の研究データ基盤に対して関心を持つ研究者やアーカイビストが多くおり、盛況であった。海外からもアクセスできるのか、英語対応するのかという言語対応についての質問が主であり、日本語調査メタデータの英語化は今後行っていく必要があるだろう。

#### IV さいごに

本稿では IASSIST での発表内容に基づき、海外データアーカイブによるデータ利用者およびデータアーカイビストへの教育やデータ共有に関わる動向について議論してきた。データサイエンスといえば、まだ日本では統計教育に限定されているが、少ないながらも DDI Lifecycle に根差したデータ作成、保存を念頭においたカリキュラムが組まれており、調査をサポートするデータ・ライブラリアンという新しい職業を見越した教育がなされつつある。調査データが計量研究の一分野だけではなく社会全体に浸透しつつあるといえる。したがって、今後のデータアーカイブはデータ寄託者とデータ利用者の視点に立ち、寄せられる期待や現状に柔軟に応える形でデータアーカイブの整備、およびサービスを提供する必要がある。そのためには、各アーカイブが独自のポリシーを遵守するだけでなく、アーカイブ間で相互運用性を満たすためのガイドラインを設け、ネットワークをつなげる努力が必要になると考えられる。

今回 RUDA は IASSIST で得た知見をもとに国内アーカイブ間相互検索システムの試験運用を試みた。一通りの成果は得たが、海外の先進的な状況を日本の状況と見なすことはできないので、国内データアーカイブの現状、およびデータ寄託・利用者の期待を正しく把握し、それに対応する中で事業を進めていくことが必要である。

#### 注

- 1) 今年度は Association des bibliothèques et archives cartographiques du Canada との共同開催となった。
- 2) <http://www.gakkai.ne.jp/jss/about/ethicalcodes.php> (2018 年 12 月 29 日取得)
- 3) <http://www.gakkai.ne.jp/jss/about/researchpolicy.php> (2018 年 12 月 29 日取得)
- 4) <http://jasr.or.jp/jasr/documents/rinrikitei.pdf> (2018 年 12 月 29 日取得)
- 5) QSSH DIG の web サイトは以下のとおりである。(2018 年 12 月 29 日取得)  
<https://sites.google.com/uncg.edu/iassistqsshdig/home?authuser=0>

- 6) DDI Lifecycle は DDI codebook と互換性を持つように設計されているが、フレームが大きく異なるため、対応させるのは難しい。DDI4.0 は DDI codebook の情報と DDI Lifecycle の情報を別々に持つように設計されていると考えられる。
- 7) RUDA で考案したメタデータは前田・朝岡(2018)に記載されている。

## 参考文献

- 前田豊・朝岡誠，2017，「海外データアーカイブの動向—IASSIST 年次大会の報告から—」『社会と統計』3:27-35.
- 前田豊・朝岡誠，2018，「RUDA の DDI 対応に向けた取り組み」『社会と統計』4:25-35.



## Summary

# The Trend in Foreign Data Archives 2 : From the Presentations in IASSIST Annual Conferences

Makoto Asaoka and Kaori Takahashi

### [Abstract]

Social research data is a public good that contributes to the development of future societies. Therefore, we need to facilitate an environment in which people can use this data appropriately. This paper reports on some cutting-edge researches and projects presented at the International Association for Social Science Information Service and Technology & Association of Canadian Map Libraries and Archives (IASSIST & CARTO) annual conference held in May 2018, focusing especially on the role of data librarians, and how to deal with sensitive data. In addition to that, we share the results and audience responses to the paper presentation about the collaborative research project with the National Institute of Informatics (NII) Research Center for Open Science and Data Platform (RCOS).

Keywords : data archive, data usage, data sharing