

文部科学省 共同利用・共同研究拠点事業

社会調査・データアーカイブ共同利用・共同研究拠点

社会科学系データの保管と共有に関する国際的な動向

——IASSIST 年次大会の報告から——

森いづみ・王帥・仲修平

(東京大学社会科学研究所)

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

2019年10月

社会科学系データの保管と共有に関する国際的な動向
—IASSIST 年次大会の報告から—

森いづみ・王帥・仲修平
(東京大学社会科学研究所)

SSJDA – 70
2019年10月

【目次】

1. はじめに
2. メタデータの整備
3. データ管理計画
4. 機密性の高いデータの利用
 - 4.1 秘匿処理の方法
 - 4.2 機密性の水準と運用例
 - 4.3 研究者パスポート
5. 質的データのアーカイブ
 - 5.1 質的データの保管や共有に関する情報収集
 - 5.2 質的データの公開と倫理審査との関係
 - 5.3 **Qualitative Data Repository** 担当者からの情報収集
 - 5.4 質的データのアーカイブに関する議論の進展
6. データ利用の国際連携
 - 6.1 データの多言語化
 - 6.2 機関間のデータ利用に関する国際連携
 - 6.3 データアーカイブ認証
7. SSJDA の今後の課題に向けて
 - 7.1 主な議題のまとめ
 - 7.2 データの共有に関する課題
 - 7.3 データの利用に関する課題
 - 7.4 データの管理・運用に関する課題
 - 7.5 おわりに

【要旨】

本稿の目的は、社会科学分野における調査データの管理や共有を考える上で欠かせない諸問題について、海外の事例を交えながら検討しようとするものである。筆者らが2019年5月に参加した IASSIST (International Association for Social Science Information Services and Technology) の年次大会で得た情報をもとに、海外のアーカイブ関連機関におけるいくつかの先進的な取り組みを紹介した後、それらをもとに、筆者らが所属する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 (CSRDA) の SSJ データアーカイブ (SSJDA) における今後の実務的な課題を整理する。

社会科学系データの保管と共有に関する国際的な動向
—IASSIST 年次大会の報告から—

森いづみ・王帥・仲修平（東京大学社会科学研究所）

1. はじめに

研究活動をより広く社会に開いていこうとするオープンサイエンス化の機運の中で、データを適切な方法で取得し、それを適正に管理・共有していくことの重要性はますます高まっている（三輪・佐藤 2018）。日本においても、近年は多くの学会で倫理規定が整備されるなどし、またオープンサイエンス化への対応が分野を超えて検討されるようになってい¹。こうした背景のもと、研究者個人にとっても、法令や倫理面に配慮しながらデータを収集・分析することはもちろん、データを安全かつ長期的な視野にもとづいて保管・共有することの重要性が増している。

本稿の目的は、社会科学分野における調査データの管理や共有を考える上で欠かせない諸問題について、海外の事例を交えながら検討しようとするものである。本稿では、筆者らが IASSIST (International Association for Social Science Information Services and Technology) の 2019 年の大会に参加し、そこで情報を得たいくつかの先進的な取り組みを紹介した後、それらをもとに、筆者らが所属する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 (CSRDA) の SSJ データアーカイブ (SSJDA) における今後の実務的な課題を整理したい²。

IASSIST とは、社会科学系の研究や教育におけるデータの管理・共有をサポートするための国際的な専門家のネットワークであり、データアーカイブや図書館、大学や民間研究組織、政府機関や非営利組織などから、社会調査や情報学、情報技術等の多様なバックグラウンドを持つ会員が登録している³。IASSIST の年次大会は 1977 年以後、毎年北米や欧州、あるいはオーストラリアのいずれかの地域で開催されており、日本の社会科学系研究者やアーカイブ関係者らも、近年若干名がこの会議に参加している（前田ほか 2011；前田・

¹ たとえば内閣府の国際的動向を踏まえたオープンサイエンスの推進に関する検討会

（<https://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/index.html>）では、「研究データリポジトリ整備・運用ガイドライン」（2019 年 3 月 29 日付）が公表されている。また、国内におけるオープンサイエンスのインフラとなる学術基盤を開発・運営するために、2017 年より NII オープンサイエンス基盤研究センター (RCOS) が設置されている

（<https://rcos.nii.ac.jp/>）。

² 社会調査・データアーカイブ研究センター (CSRDA) は、2010 年 4 月に文部科学省の共同利用・共同研究拠点としての活動を開始している（2016 年 4 月に認定更新）ほか、2018 年に日本学術振興会より「人文学・社会科学データインフラストラクチャー構築プログラム」の拠点機関として認定を受け、活動している。

³ 会員の背景は米国がもっとも多く、続いて欧州、カナダ、アジア・太平洋、アフリカとなっている（<https://iassistdata.org/about/index.html>）。

朝岡 2017 ; 朝岡・高橋 2019)。今回 (2019 年) の年次大会は、2019 年 5 月 27 日～30 日にオーストラリア・シドニーのニューサウスウェールズ大学にて開催された。以下、メタデータの整備、データ管理計画、機密性の高いデータの利用、質的データのアーカイブ、データアーカイブの国際連携や認証といった項目ごとに分けて、社会科学系の調査データの管理や共有について考える上で、有意義だった取り組みについて記載する⁴。

2. メタデータの整備

“Introduction to DDI: Basic Concepts and How to Develop Skills for Training Researchers”と題するワークショップでは、調査概要等の情報を意味するメタデータ (= データについてのデータ) にとって不可欠な要素とは何か、メタデータは研究を能率的に進める上でなぜ大切か、メタデータは、ライブラリアン/レポジトリ/研究者/IT 担当者等にとってどのような意味を持つかといった点について、参加者間でのグループセッションを経ながらの講義がなされた。とくに、調査を実施する当初からメタデータについて考えることの大切さが強調され⁵、DDI (Data Documentation Initiative) とはそもそも、データの合理的な利用をサポートするための一貫した枠組みなのだ、という理解が示された。

とくに印象深かった点として、“Legacy metadata may require updating” (古いメタデータは、アップデートが必要になることもある) というアドバイスがある。たとえば DDI の形式を変えると、一貫性やフォーマット変換等の観点から、既存のメタデータをアップデートすることはある程度避けられないのだというのは、SSJDA でも現在、DDI への対応や英語化で直面している問題であり、うなずける点であった。

担当講師 (ドイツのアーカイブである GESIS のアーカイビストと、カナダのカールトン大学の図書館所属のデータ・スペシャリスト) によると、メタデータ管理のソフトウェアにはいくつかの種類があるが、Nesstar Publisher については、学生の教育にはよい (ウェブ上で分析を行うのは記述統計とクロス表までがおすすめだが、それ以上の場合はダウンロードをすすめる) が、この 10 年ほどアップデートがないのが心配でもある、との懸念が示された。同じくメタデータ管理ソフトウェアである Colectica については、たとえばエクセルに対応しているソフト (スタンダード版は無料、プロフェッショナル版は有料) を使用することで、SPSS や Stata、SAS などで作成されたデータファイルを、エクセルにインポートして、その変数情報等を生かしながらメタデータ作成が可能であるとのことだった⁶。

⁴ 主に森が全体にかかわる部分を執筆し、王がデータの寄託や多言語化に関する部分、仲がデータの利用に関する部分について、それぞれ分担執筆した。なお、本稿に表明されている見解はあくまで各執筆者のものであり、社会調査・データアーカイブ研究センターの公式見解でないことをお断りしておく。また、本稿で提示する情報や URL 等のリンクは、いずれも 2019 年 10 月現在のものである。

⁵ 調査後に作成されるメタデータを“downstream”メタデータと呼ぶのに対し、調査開始当初から作成されるメタデータを“upstream”メタデータと呼ぶとされる。

⁶ Colectica に関しては、Youtube の tutorial ビデオが参考になるとのことだった。

(<https://www.youtube.com/user/Colectica>) ただし、日本語に対応しているとの情報はな

今回のワークショップの内容の大部分は、もともと DDI Alliance の公式サイト (<https://www.ddialliance.org/>) にも記載されていたものではあるが、ワークショップ形式で参加者の質問や議論を交えながら進行することで、ただウェブサイトから情報収集するよりも、各自の理解が深まるものだと感じられた。また、DDI に関して、今回のワークショップと同じような内容を自分の所属機関の同僚等に説明したい場合は、今後“Training Library”というページが追加されるとのアナウンスがあった。2019 年 10 月現在、以下の URL より、DDI に関する基本的情報を含んだいくつかの説明スライドが入手可能となっている (<https://ddialliance.org/training/training-library>)。

質疑応答の時間に、参加者からそもそも DDI は、質的データなどの異なるタイプのデータの内容を表す際にも本当に適しているのか、という質問が出され、その後の講師とのやりとりから、こうした (DDI のような) データ規格の標準化・統一化の流れに対して、一定の懐疑的な見方もありうるのだということがうかがえた。

なお、メタデータとは何かや、その日本での適用の詳細については、これまでも社会科学系のアーカイブ担当者による解説や情報共有がなされてきたので、詳細はそちらも合わせて参照されたい (前田ほか 2001 ; 前田・朝岡 2018)。

3. データ管理計画

データ管理計画 (Data Management Plan) についてのワークショップ (“DMPs: HmMMMM, I Think I Know What They Are But) は、カナダの大学のデータライブラリアン 2 名 (ワークショップやウェブサイト等を通じて、研究者らにデータ管理計画の意義や作成法、具体例等を伝えることを業務の一部としている) によって担当された。

研究者はしばしば、調査にあたり自分自身がデータ管理計画を作成するという意識が薄い、近年はそれが研究資金を得る際の必須事項となっている場合もあり、また「データをシェアできることは、よい研究実践にとって重要である」との観点からも、データ管理計画が重要であることを述べていた。データ管理計画作成のための障壁として、時間や理解不足、前例がないこと等もある中で、以下のような役立つツールについても紹介がなされた。

Portage Network (<https://portagenetwork.ca/>) とは、カナダの研究図書館連合が作成した、データ管理計画ツールである (図 1)。誰でもオンラインでアクセス可能で、右上のオレンジの箇所、「DMP Assistant」に登録すれば、データ管理計画のテンプレートが得られるとのこと。ワード形式でインポートして、その後自由に使えるとのことである。

いので、データに日本語のラベル等が含まれる場合は確認が必要かもしれない。

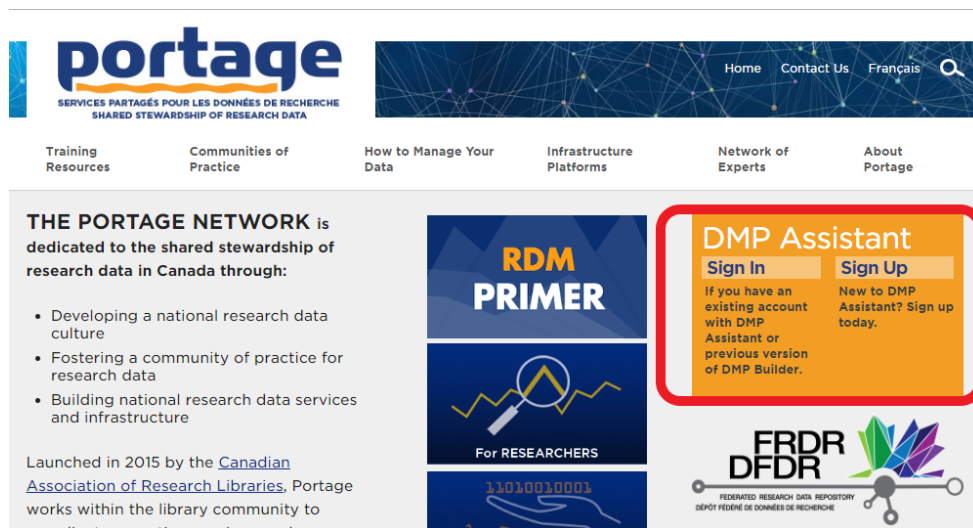


図 1 Portage Network のウェブサイトのスクリーンショット

データ管理計画について、とくに肝心なのは、研究者が研究のライフサイクルの最初から、これについて考えること（＝データをどうするのかについて、事前によく考えて決めておくこと）なのだ分かった。その際、たとえば IT 関係のスタッフと風通しの良い関係を築いておくことの重要性（例：兌換性のある保存フォーマットやストレージスペース等について尋ねられる）についての言及もあった。

そもそもなぜデータの共有が必要なのか、データシェアリングの是非についても説明がなされ、後者については、たとえばデータ捏造（data fabrication）のように注意が必要な場合も含まれることが指摘された。その後ケーススタディとして、理系分野の実際の例をもとに、「このようなデータ管理方法で、注意すべき点はどこか？」といったシミュレーションエクササイズも行われた（例：データのバックアップはきちんととれているのか、永続性の観点から、適切なソフトウェアが使われているのか等）。

結局、データ管理計画とは誰のためのものなのかについて、講師に質問したところ、研究者個人に対してだけでなく、データアーカイブを含め、誰にとっても当てはまるもの、という理解であるとの回答がなされた。（後日 Portage のウェブサイト上で確認したところ、やはりそのような記載あり）。講師らの言うように、たしかに研究者にとって、こうした管理計画を作ることに慣れるハードルが高いかもしれないが、このように全般的な計画について考えることは、調査の実施からデータの寄託・保管・共有までを視野に入れたデータライフサイクル全体を考えるという意味で、社会科学分野においても、間違いなく有意義なことであろう。

4. 機密性の高いデータ（sensitive data/ restricted data）の利用

4.1 秘匿処理の方法

“Statistical Disclosure Control”のセッションでは、秘匿処理を容易にするための R を

用いたフリーソフトウェア (sdcMicro) や、オンラインパネルデータの秘匿処理についての議論が有意義であった。sdcMicro

(<http://www.ihsn.org/software/disclosure-control-toolbox>) とは、世界銀行などの支援を得て開発し、多くの途上国の調査等を公開する際に利用されているツールである。R の操作に慣れていない人でも、Graphical User Interface (GUI) で操作できるとの紹介があった。なお、関連するリソースとして、報告者らが作成した、オンラインで参照できる秘匿処理に関するガイドライン“Statistical Disclosure Control for Microdata: A Practice Guide”が紹介された (<http://www.ihsn.org/node/685> より入手可)。秘匿処理のさまざまな方法や理論に関する説明や、リスクと利便性を秤にかけた秘匿処理の判断 (risk/utility measures) などについての記載がある。

同じ対象者に対して繰り返し調査を行うオンラインパネル調査の場合は、パネルを追うごとに個人の識別につながる指標 (identifier) が増えていくため、とくに統計的な情報開示のコントロール (statistical disclosure control) を考慮することが重要になる。オーストラリアのデータアーカイブの担当者からは、Life in Australia というパネルデータの複数のウェブを使い、実際にどのような基準で開示のリスクと利便性のバランスを取っているかについての枠組みが報告された。

4.2 機密性の水準と運用例

ICPSR では、Restricted Data に対するアクセスは「Probability of Disclosure」と「Severity of Harm」の2つの側面から4つカテゴリー (Web Download, Secure Download, Virtual Enclave, Physical Enclave) を設けている (Akmon and Jekielek 報告)。各カテゴリーは一般的な Web Download からより厳しい制約を課す Physical Enclave へととなっている。報告では、Restricted Data に対する申請からアクセスまでをいかに簡素化するかが鍵となっている点が指摘されていた。その課題に対しては、Restricted Data 専用の新しいシステムを構築しているようであった。それに際しては、Restricted Data の利用に関連する概念を統合／統一しておくことが肝要であることを強調していた。

これらの点は、SSJDA が保有する「特別データ」の扱いに対して多くの示唆を有する。たとえば、SSJDA では、それぞれの「特別データ」に対する名称や提供方法が異なっている場合がある。むろん、各データの寄託される経緯が違うために、固有の誓約書や提供方法となることは問題ではない。しかしながら、より安全でよりスムーズにデータを提供するという観点からすると、現状の仕組みは改善の余地が残されている。とりわけ、ICPSR が強調していたように「一つの仕組みの中で完結させる」ことは有効な手段となりうる可能性があるだろう。

機密性の高いデータに関しては、Birds of a Feather という少人数の対話型セッションにも参加し、色々な国や地域のデータ担当者が代わる代わる発言する中で、いくつもの

有意義な情報を得た。まず、情報共有を進める中で、“restricted data”といっても、対象者の特定につながるどのような変数がどれだけあるかにより、色々なレベルのものがある (“baby restricted”から“super restricted”まで) という指摘がなされ、参加者がみな同意した。たとえば SSJDA でも、いくつかの「特別データ」利用に関してはそれぞれ個別の取り決めがあるが、たとえば今後「病歴」等、健康関連のより機密性の高いデータを公開することになる場合、専用の分析室が必要になるのではないかといった議論が内部であり⁷、そうした検討を行う際にも、このように「特別データにもいろいろなレベルがある」ということをまず認識することは有効だと思えた。(たとえば米国の ICPSR で薬物使用やエイズに関するデータが公開されているように、扱うデータの機密性のレベルでは実際に違いもあり、それだけの設備やシステムもそろっている。ICPSR の場合、Virtual Data Enclave (VDE) と呼ばれるオンサイトの分析室がある⁸。)

以上のような議論の中で、“Five Safes” (safe people, safe project, safe settings, safe data, safe outputs) と呼ばれるデータの機密性を守るための枠組みが、とくに欧米のアーカイブ関係者の間では意識されていることがうかがえた。Five Safes の詳細は、以下のサイトで詳しく紹介されている

(<https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control/five-safes>)。つまり、機密性の高いデータの利用に際しては、各国・機関の間で対応の進度に差はあるものの、共通した努力や課題 (利用者の身分や研究計画のチェック、利用の際のトレーニング、データへの秘匿処理、オンサイト分析室やリモート集計等の環境設定、データ分析結果のチェック) がなされており、人手や労力、データの特性等の観点から、必要かつ可能な範囲での対応がなされているということである。

また、参加者の中で、自身のレポジトリでの特別データの運用経験から、restricted data といっても、本当に個票データのレベルでそれが必要な人はあまり多くない、という発言もあった。要は、研究者が特別データを使って知りたい多くの事項は、実は個票データを借りなくとも、クロス集計レベルで分かることも多いとのこと。特別データの貸し出しには、アーカイブ側の負担もそれなりにかかるにもかかわらず、である。しかし、やはり研究者にとっては、実際のデータ全体を見ないと、必要な分析だけを取り出してコードを書くのも難しいというジレンマがある。実際に借りてみた特別データを見ると、結局あまり使えなかった、ということもありうるという発言があった後、データに関するドキュメンテーション (メタデータ等) 自体は restricted である必要はないのだから、もっとそれら自体をオープンにした方がよいのではないか、といった提案もなされた。

⁷ SSJDA では現在、「社会調査室」として、特定のデータの分析を行うことのできる分析室を設置しているが、その運用の枠組み等はまだ定まっていない状況にある。

⁸ <https://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/enclave.html> ICPSR では、機密性の高い質的データについては、safe room (安全な分析室) 内で分析させ、パソコンで電子的に取ったメモのみを、アーカイブ担当者の確認後に持ち出してよい、という運用がなされているという。

日本の政府統計についても、利用のために時間や労力がかかるといった声が聞かれるので、これらの他国の現状についての議論は、実に示唆的であり、多くの国で共通の課題を抱えているのだということが見て取れた。

さらに、データの機密性に関して、質的調査データの利用についても議論が及んだ。米国シラキュース大学の **Qualitative Data Repository** では、保有するデータの利用について、個人レベルでも、人に対する **memorandum of agreement (MOU)** があれば、利用を許可しているとの話があった。質的な社会調査データの公開の際の難しい点として、面白い質問項目がたくさん含まれるのに、秘匿処理などによって“**restricted**”運用のためのデータ処理を施すと、大事な情報の多くが失われてしまうので、そもそもデータを公開することの意味が薄くなる、といったジレンマも表明されていた。

続いて、提供後の利用について若干述べておく。実際にデータを利用者へ渡さずに分析させる方法の一つとして、リモートアクセスがある。この場合、データ利用が安全に行われているかどうかを常にチェックする（参加者の言葉で“**baby sit** する”）人員が必要で、たとえば IT 担当者の役割が必要になるといったことが議論された。また、利用者に関する教育（トレーニング）を重視するという方法も挙げられた。あるアーカイブでは、機密性の高いデータ利用に関する“**secure data training**”という講習を受けて登録すれば、その研究者に対して、5 年間は認証が与えられるという、パスポートシステムのようなものを導入しているという情報もあった。

4.3 研究者パスポート

Restricted Data の管理と利用提供には莫大なコスト、及び十分な注意が必要である。それにしても、利用される過程のすべてを把握しきれず、利用に伴うリスクが **Restricted Data** を提供する上で大きな懸念となっている。**Restricted Data** のより信頼的・安全的・効率的な運用管理を実現するために、**ICPSR** は研究者パスポート (**Researcher Passport**) システムの開発と実装を試みている (Hemphill, 2019)。研究者パスポートは旅行パスポートのようなもので、所属や身分、トレーニング経験などに関する事項を証明し認証機関 (**ICSPR**) が交付するという、データ利用者に対するデジタルアイデンティティシステムである。研究者パスポートの一環として、**ICPSR** は **Restricted Data** にアクセスするために必要なトレーニングや経験などを評価し、そのステータスを表すオープンバッジ (**Open Badges**) を発行するシステムも開発している。今までは、**Restricted Data** が利用される度に、利用者の身分を証明できる書類の提出が要求され、書類審査に時間がかかった。研究者パスポートの搭載によって、利用者情報が共有されているということで、利用者にとっては身分証明の提示が一回で済み、アーカイブ機関にとっては書類審査の手間がなくなり、より簡単に迅速な申請プロセスの実現が可能になっている。また、研究者パスポートの運用基準を共有する機関が多くなるにつれ、各アーカイブ機関にとっては、より安全的な運用ができるほか、データ利活用の普及につながることを期待できる

5. 質的データのアーカイブ

5.1 質的データの保管や共有に関する情報収集

質的調査データの保管や共有に関して、近年 SSJDA でも問い合わせを受けることがある。さまざまな質的データの中でも、インタビュー形式の調査は SSJDA の対象とする利用者コミュニティにとっても比較的関連の深いものである⁹。SSJDA が現在寄託を受け付け、公開しているのは、調査票調査にもとづく回答を数値化した形でのデータのみであり、外部の寄託者からは、保管を含めてインタビューデータやその録音内容等の資料は受け付けていない。なお、附属社会調査・データアーカイブ研究センターの社会調査部門が調査を実施した社研パネル調査シリーズのインタビューデータについては、内部フォルダにて適切な管理体制や権限付与のもと、保管を行っている。

しかし近年、研究者から、インタビューデータのトランスクリプトや録音データ等を寄託し、SSJDA にて安全に保管することは可能かといった問い合わせを受けることがある。これは、昨今のデータ共有への意識の高まりや、データの安全で長期的な管理計画が一層求められるようになったという時代背景を考えれば、ある種もつともなことである。今後そうした対応がどの程度ありうるのかを含め、SSJDA 内部でも、海外の先進的な事例等も参照しながら、質的調査データの受け入れの際のルールや注意点等について、まずは情報収集から始めることにしている¹⁰。

5.2 質的データの公開と倫理審査との関係

“Working with Sensitive Data: The Qualitative Data Repository’s WSRD Project”という報告では、米国の National Science Foundation (NSF) の支援を受け、米国で質的調査の際に不可欠となる IRB (Institutional Review Board) 審査と、質的調査のデータ管理計画 (Data Management Plan) をいかに調和させるかについての取り組みがなされている様子が報告された。インタビュー調査やフィールドワーク等を行う際、調査対象者の調査への同意 (インフォームド・コンセント) を得ることは重要であり、これは IRB 審査の重要なポイントの一つともなっている。実はこの IRB で約束する事項と、データ管理計画にお

⁹ 東京大学では現在、質的調査データの保管にかかわる機関として、とくに歴史的に重要な資料等の保管・管理のために「東京大学史料編纂所」や「東京大学文書館デジタル・アーカイブ」といった組織がある。

¹⁰ たとえば、イギリスの UKDA やオーストラリアの ADA において、質的調査データの保管や秘匿処理についての詳細なポリシーが述べられており、参考になるだろう。

<https://www.ukdataservice.ac.uk/get-data/other-providers/qualitative/uk-archives.aspx>

<https://www.ukdataservice.ac.uk/get-data/other-providers/qualitative/international.aspx>

x

<https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative>

<https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation>

<https://ada.edu.au/depositing-data/>

けるデータの扱いに関する事項との関係性はあいまいであり、たとえば IRB への申請では、データの利用終了後はデータを破棄する、と記載することが通例であるものの、そのように記載した場合、アーカイブ等を通じてデータを適切に保管・管理するといったことが違反となってしまう。こうした事態を打開するため、質的調査の実施で対象者の同意を得る際に、データの共有や二次利用についても明確に同意を得ることや、複数の機関の IRB 担当者との会合を通じて、こうしたデータ共有への理解を深めようと努力している、というのが報告の内容であった。

5.3 Qualitative Data Repository 担当者からの情報収集

米国における質的調査データ専門のアーカイブであるシラキュース大学の **Qualitative Data Repository (QDR)**¹¹に関するセッションでは、たとえば先住民族に関する質的データ（インタビュー内容の他、写真や何らかの現物等）をアーカイブしたいが、完全に匿名化ができない場合に、どうしたらよいかという質問があり、データ自体は受け入れるが、ずっと外に出さず保管したままにするというオプションもある、といったやりとりもなされていた。QDR の担当者の **Sebastian Karcher** 氏は政治学の博士号を保有しており、IASSIST の学会誌 (**IASSIST Quarterly**) にも論文が掲載されているなど、質的調査データのアーカイブについて造詣が深い¹²。

SSJDA でも、最近インタビューデータの音声データについて、アーカイブで保管できないかという問い合わせを受けており、この可能性について **Karcher** 氏に尋ねたところ、たしかにトランスクリプト等と違って、インタビューの音声データは秘匿処理が難しく（対象者の名前に音声処理をするなどには多大な労力がかかるため）、今のところ QDR では、音声データの公開は行っていないとのことだった。ただし、エスノメソドロジーなどの分析で、たとえば語りや「間」の分析などでこうした音声データが研究上必要になることは十分理解できるため、今後そうした点も含めて検討することは意義がある、とのことだった。SSJDA でも、今後たとえば外部の大型プロジェクトの担当者等からインタビュー関連のデータ保管の要請があった場合などに、音声データに関しては完全な秘匿処理をすることはできないにしても、そのデータを、インタビュー対象者のリストと別々の場所に保管し、その紐づけ情報を一部の人にしか見えないように配慮するなどして、インタビューデータを保存目的で受け入れる、といった可能性については、そのメリットやデメリットを含め、様々な面から慎重に検討を進めていくべきかもしれない¹³。

¹¹ QDR の紹介は以下のリンクを参照。 <https://qdr.syr.edu/>
<https://www.rd-alliance.org/qualitative-data-repository-qdr.html>

¹² IASSIST Quarterly については以下のリンクを参照。
<https://iassistquarterly.com/index.php/iassist/issue/archive>

¹³ 質的調査データの受け入れに関しては、データの機密性や倫理面での検討事項など、とくに色々な配慮が必要になりうる。たとえば UK データアーカイブの『データの管理と共有』の「著作権」に関する事例紹介コラムでは、研究者があるインタビューデータの録音

5.4 質的データのアーカイブに関する議論の進展

IASSIST2019では、上記以外でも質的データに関して、ソーシャルメディアやゲーム関係のデータの利用やアーカイブ化についての報告もあった。その際もやはり重要視されていたのは、調査対象者への同意（インフォームド・コンセント）であり、たとえば一般公開されているウェブサイト等の情報であっても、何が公的なもの（public）で、何が私的なもの（private）なのかの識別が難しい場合もあることや、個人のプライバシーと、データの収集・公開による公共的な利益をどう秤にかけて考えたらよいかについての議論がなされていた。

また、学会中のプログラムの一つとして、質的データに関する Interest Group Discussion というものもあった。1時間のインフォーマルなセッションで、まず2名の司会担当のもと、参加者（30名程度）がそれぞれ所属と関心について述べた後、今後参加者の間でどのような情報共有が可能かについての意見交換を行うという形式を取っていた。質的データと一口に言っても、その方法論や分析、教育面に関心のある参加者（例：分析用ソフトウェアなど）と、質的データのアーカイブに関心のある参加者（後者の方が少し多め）の両方がいることが分かった。議論の中で、IASSISTはどちらかというとき量的データに関心が偏っているが、質的データについても今後もっと議論を深める必要性が提起されたり、質的調査データをアーカイブ・管理する際の基準や条件（criteria）のようなものはあるのか（それは量的データより複雑かもしれない）、というような印象深い議論がなされていた。今後、このグループ独自のウェブサイトを今回の参加者の交流拠点とし、今後質的データに関する2つ程度のトピックについてウェビナーを開催したり、来年度のIASSIST学会の前に、今回挙げた議論の要点をリマインドする、ということがアナウンスされており、継続的に情報共有がおこなわれている（<https://sites.google.com/uncg.edu/iassistqsshdig/home>）。

6. データ利用の国際連携

6.1 データの多言語化

オープンサイエンスが進むにつれ、データ共有の範囲が一国レベルを超え、国際連携レベルまで展開しつつある。欧州17か国のデータアーカイブが加盟する欧州社会科学データアーカイブコンソーシアム（CESSDA）は最も代表的な団体である。多国間のデータ検索や利用を促進するため、共有できるデータアーカイブ環境の構築がもちろん重要であるが、言葉の違いがデータ共有において大きな壁であることは間違いない。今回のIASSISTでは、CESSDAにおける持続可能な多言語コンテンツの開発にあたり、加盟機関間のメタデータ

媒体と印刷された記録の保存をデータ・センターに申し込んだ際、もし研究者が著作権の移譲または面接で得たデータの利用許諾を事前に調査対象者に対して取得していない場合、データ・センターがデータを受け入れることは問題になりうる（データ的大幅な抜粋を二次利用者が引用することができないため）といったケースが紹介されている（UK データアーカイブ 2011=2013, 日本語版 p. 30）。

管理協力の取り組みが紹介され、多言語コンテンツの翻訳品質に関する検証の結果が報告されていた。

CESSDA では、欧州における多言語コンテンツの開発やポリシーの作成、欧州の各データアーカイブから研究データを入手できるようなメタデータ管理に関する取り組みが行われている。多言語コンテンツとしての欧州言語社会科学シソーラス (ELSST) は、2000 年に開発に着手し、データ共有の効率化につながる効果が期待されている。一つまたは複数のヨーロッパ諸国の言語で概念を表す語彙網の構築や、標準化されたインデックス用語の整備により、多言語間のデータ検索が可能となっている。2019 年 9 月時点で、概念の検索に用いられる言語の種類は 14 言語 (デンマーク語、オランダ語、チェコ語、英語、フィンランド語、フランス語、ドイツ語、ギリシャ語、リトアニア語、ノルウェー語、ルーマニア語、スロベニア語、スペイン語、スウェーデン語) にも達している¹⁴。

Barbalet (2019) の報告では、ELSST の開発にあたり、CESSDA 加盟機関間のメタデータ管理協力のモデルが紹介されていた。CESSDA では、基本、すべての加盟機関が ELSST の構築にかかわり、メタデータ基準の開発や多言語コンテンツのテスト実装に協働する体制で行われている。このような協力体制によって、国境を越えて、多領域ないしは学際分野にわたる質の高い社会科学データの共有・活用を促進する効果を挙げていた。しかし、複数の国にまたがる CESSDA-ELSST プロジェクトの進展は必ずしも順調ではない。加盟機関を対象に行った多言語対応のメタデータ整備に関する中間調査では、ELSST の利用状況に違いがあることや、一部の加盟機関が ELSST を索引に導入しなかったことなど、協力体制が抱える問題点が提示されていた。

Balkan (2019) の報告では、多言語コンテンツの翻訳性能を検証するプロジェクトの取り組みが紹介されていた。多言語にまたがる作成した索引目録の一貫性や、多言語検索の有効性を確かめるために、言語レベルと実装レベルで相互に整合性を持った形で検証が進められており、多言語コンテンツ性能の検証方法として有効であると評価されていた。

日本でも今後データアーカイブの国際連携や利用者の多様化が進んだ際、こうしたデータの多言語対応が一層要請される可能性がある。CESSDA の ELSST プロジェクトは、そうした際に一つの参考になる取り組みと言えよう。

6.2 機関間のデータ利用に関する国際連携

“Transnational Access to Controlled Microdata for Research – Overcoming Barriers” と題するパネルセッションでは、国際間で個票データの共有を可能にするための 3 つの先進的な事例が紹介された。近年、データ利用に関する法的枠組みや安全なデータアクセス方法が検討される中で、詳細な政府統計や調査データ (highly detailed administrative data

¹⁴ ELSST の概要については、たとえば以下を参照。

<https://www.ukdataservice.ac.uk/about-us/our-rd/cessda-elsst.aspx>

and surveys) が次第に研究やエビデンス・ベースドの政策のために使われるようになっていく。このような認識のもと、一国内はもとより、国際間においてこうした個票データをいかに共有していくかという課題について、異なるデータセンター間での合意形成や、データ運用枠組みの統一化に関する議論がなされた。

具体的には、1) フランス、ドイツ、オランダ、イギリスの 6 つの研究データセンターによる共同のネットワークである International Data Network (IDAN)、2) 欧州の銀行関連のデータを扱う共同プロジェクトである International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA)、3) デンマーク、フィンランド、ノルウェー、スウェーデンの政府統計機関が、加盟国内で企業／事業所関連の個票データの相互利用を可能にする Nordic Microdata Access Network (NordMAN) の取り組みが紹介された。秘匿性の高いデータ利用の場合は特に、各国でバラバラである基準をまずは比較し、整備するための努力がなされていることが分かった。

International Data Access Network (IDAN)¹⁵は、2018 年に設立された欧州 6 機関のプロジェクトである。設立の経緯として、まずは 2 機関ごとにセキュリティ面や安全なデータ分析室 (Safe Rooms) 等の環境面での合意 (bilateral agreements) を形成し、その後、6 機関全体でのより統合されたデータ相互利用の枠組みが確立されつつある様子が紹介された。メタデータの英語化については、現在進行中とのこと。利用資格の付与、トレーニング (講習等)、データへのアクセス方法のそれぞれについて、各国で個別に存在している方法や方針を比較し、標準化や調和 (harmonization) に向けてプロジェクトを進めていることが紹介された。

The International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA)¹⁶は、欧州中央銀行や欧州連合統計局が中心となり運営している granular data の利用に関する国際ネットワークであり、イタリア、ポルトガル、イギリス、フランス、ドイツの各機関が参加している。国家間のデータ利用に関するワーキンググループやワークショップを定期的に関き、利用する研究者向けのトレーニングを行ったり、データアクセスに関する共通の課題や最良の事例 (例：遠隔地から安全にデータにアクセスする方法等) について議論し、技術者と協力して利用のためのオープンソフトウェアを開発しているとのこと。ミーティングに一度参加し、所属機関長の名前で公式の申請をすれば、メンバーになれるとの紹介があった。

Nordic Microdata Access Network (NordMAN)¹⁷は、北欧諸国における行政データの国際利用ネットワークであり、個人レベルと企業 (事業所) レベルの個票データ利用が可能とのこと。行政データの国際間での利用に関して、報告者の回答では、少なくとも北欧の参加機関の間では「信頼」(trust) が成り立っており、万一对象を特定するなどの不適切

¹⁵ <https://idan.network/>

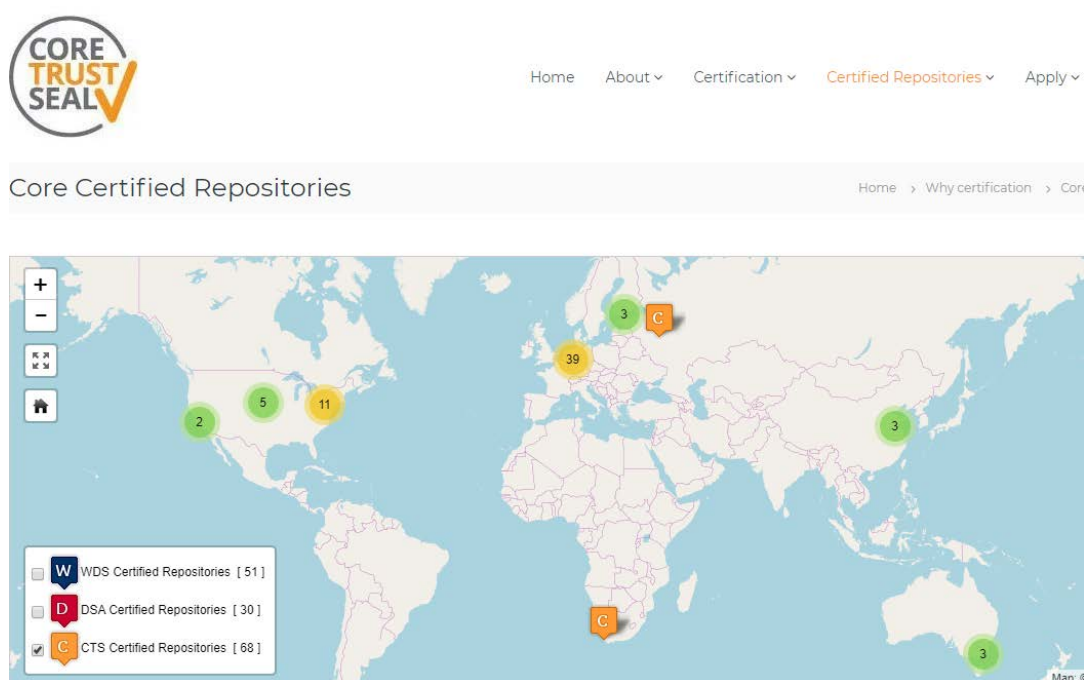
¹⁶ <https://www.inexda.org/>

¹⁷ <http://nordman.network/>

な分析が行われた場合は、その利用者や所属する利用機関を利用停止にするといった措置も辞さない（つまり、それだけ信頼が重視されている）といった説明がなされていた¹⁸。また、続いてフロアの別の参加者から、国際間でのデータへのアクセス方法はどうなっているか、との質問があり、安全なリモートアクセスを実施している（自宅から接続可能）との回答だった。（おそらくデータ自体はダウンロードさせない利用方法、ということだろう。）ウェブサイトを見ると、加えて **Output check** を行っている、というような記述もあった。

6.3 データアーカイブ認証

データアーカイブ認証とは、データリポジトリの信頼性を担保するために設けられた制度であり、デジタル資産としてのデータを、安全かつ持続的な方法で保管・利用することを保証するための国際的なベンチマークである。2017年にこの認証が開始されて以後、現在68機関がすでにこの認証を取得済みであり（図2）、SSJDAでも現在、Core Trust Seal (CTS: <https://www.coretrustseal.org/>) と呼ばれる国際認証¹⁹の取得に向けて準備を進めている。



¹⁸ SSJDA で保有する事業所データについては、たとえば設立年と、企業規模などの情報を合わせることで対象の特定が行われる可能性があるため、それらを防ぐための秘匿処理を行っている。そうした文脈から、NordMAN の報告者に秘匿処理はどのようにしているかの質問をしたところ、このような回答を得た。

¹⁹ 日本語で CTS 認証の概要を分かりやすく記載したものとして、以下が挙げられる：南山泰之，2017，「信頼できるデータリポジトリの中核的な統一要件」カレントアウェアネス・E, No.320 (<https://current.ndl.go.jp/e1888>)。

図2 CTS 認証を取得している組織 (2019年10月時点, 68機関)

IASSIST2019の大会では、“Digital Preservation, Trust, and Certification”と題するセッションに参加し、ICPSR、World Data System (WDS)、ADAの担当者によるデータアーカイブ認証に関する説明を聞いた。また、学会期間中にCTS取得済みの機関の担当者と情報交換するなどして、アーカイブ認証に関連して以下のような情報を得た：調査実施者に対しては、調査の一次利用だけでなく、アーカイブへの寄託・保存＋二次利用についても本来同意を得ることが望ましい。ただしこうしたインフォームド・コンセントや法的・権利関係の確認方法については、アーカイブによってある程度現実的な対応や解釈をしているとの回答を得た²⁰。翻ってSSJDAの寄託書を見たところ、権利関係についてはそれなりに記載があるが、まだ不足している部分もあり、今後CTSへの申請にあたって、上記のアドバイスを念頭に改めて検討を進めているところである。

その他、申請準備にあたり訳に立ちそうなアドバイスとして、以下が挙げられる。

- Core Trust Seal (CTS) 認証については、すでに別の認証を取得済みの機関の場合でも準備に数週間～1か月程度かかり、その後通常レビュアーから質問が来てやりとりがあるとのことなので、早めに進めた方がよいだろう。完全にすべての要素が「実装段階」でなくとも、そのためのタイムフレーム（いつまでに何をやる予定か）を示せば、申請自体は十分可能とのこと。またその際、コーネル大学の担当者より、一人ですべての項目には対応しきれないので、一人の取りまとめ役を中心に、関係者間で複数回打ち合わせを持つことを勧められた。
- CTS 認証にあたり、やはり関連項目に関するウェブサイトのアップデートは必要になること（ただし非英語圏のアーカイブについては、適切な説明があれば、認証のためだけに不必要な英訳しなくても大丈夫そうであること）。
- CTS のガイドラインにあるすべての要素について、逐一答えなくてはならないというわけではなく、あくまでそれらを参考に、必要な情報を含めて回答を作成すればよいとのこと。
- ポリシー関係の文書については、ウェブサイトに直接記載するのでも、PDF 文書のようなリンクを作成するのでも、やりやすい方法で問題ないとのこと。
- 内部文書をどこまで出すかについては、各アーカイブで判断が必要になること。
- Institutions と Repositories の重なりや違いについては、表記を工夫する必要があること（例：社研／CSRDA／SSJDA）。

²⁰ 各機関の担当者に口頭で尋ねたところ、たとえばオーストラリアのADAでは、初期の分析 (initial analysis) と二次分析 (secondary analysis) の両方について、同意を得ることを原則としているとのこと。ただし、寄託者 (= 調査実施者) がその同意を取り付けたかどうかに関する、物理的な証拠までは、必ずしも求めていないとのことだった。

7. SSJDA の今後の課題に向けて

7.1 主な議題のまとめ

日本でも今後、オープンサイエンス化が進むにつれ、データの管理や共有の際の倫理的配慮や注意点が一層重要になるだろう。その際、とくに本稿で述べたようないくつかの事項（たとえばメタデータやデータ管理計画、機密性の高いデータの利用等）に対する意識は高まってくるのではないだろうか。

今回の IASSIST2019 への参加を通じた全体的な所感として、データの利用に関して、機関間・国際間の情報共有や連携が進んでいることがうかがえた。たとえば今回の学会では、各アーカイブ/レポジトリの保有するメタデータや個票データ、または秘匿性の高いデータについて、複数の機関間（一国内、あるいは国際間）で、その利用基準を共通化する努力が行われていることが報告されていたし、たとえば欧州の複数の国家間で、個票データの利用を可能にするための取り組みがいくつも進行中で、今回もそのようなプロジェクトが紹介されていた。また、そうした際に重要なテーマとなるのが、メタデータの標準化、利用資格や利用基準の共通化、多言語化、法律面や倫理面での配慮などであり、今回の報告者や参加者の多くも、そうした関心を共有していたと言えるだろう²¹。

データの寄託促進や利用状況の把握、共有の促進といった、データアーカイブの中心的な機能についても、参考になる議論が多く行われており、そうした中には、SSJDA と共通する課題も多くある模様だった。一例として、利用者が常に正しくデータ出典を示すわけではないため、ICPSR では人手や労力をかけて二次利用の成果物を収集していることや、CESSDA でも、永続的な識別子 (PID: persistent identifier) の利用に関するガイドラインが現在整備されつつあることなどが挙げられる。また、参加者の間で、科学的データ利用の“FAIR”原則 (to make data findable, accessible, interoperable and reusable : データを見つけやすく、アクセスしやすく、相互運用しやすく、再利用しやすくするという公開の原則) への認識がある程度共有されていることも印象的だった²²。なお、本報告では項目として挙げていないが、データや分析の再現性 (reproducibility) というトピックも、こうした背景と関連して、しばしば議論されている項目である。

²¹ また、一国内でも、たとえばオーストラリアでは、国の機関である Australia Research Council がデータカタログを整備しており、社会科学系データアーカイブである Australian Data Archive (ADA) は、そこからのリンクで利用者を獲得することもあるとのことだった。日本でも現在、国立情報学研究所の主導のもとで統一的なデータカタログ整備についての議論が進行中であり、こうした海外の状況と共通する点も多い。

²² たとえばメタデータに関するワークショップで、メタデータの標準規格 (DDI) についての意義を議論する際、参加者の間でこのようなワードが上がってくる。欧州の GDPR (General Data Protection Regulation, EU 一般データ保護規則) へ対応するデータ管理計画に関するセッションでも、プライバシーへの配慮とオープンサイエンスをどう両立するか、といった際に、この原則が意識されていることが分かる。FAIR 原則についての日本語での説明は、たとえば以下のサイトを参照のこと (<https://current.ndl.go.jp/e2052>)。

さらに、データの国際間利用に関心のある人々の間では、日本の動向も十分に海外からの関心となっていることが印象的であった。SSJDAでは今回、データの寄託促進に関するセルフアーカイブ化のシステムについてポスター報告し、少なからずの機関や担当者が関心を寄せてくれた。また、特別データの利用に関する少人数の情報交換セッションなどでも、海外の研究者がSSJDAの特別データを利用する際の仕組みがどうなっているかについて、質問を受けたりした。

7.2 データの共有に関する課題

オープンサイエンスの実現に向けて、データポリシーの策定や情報共有基盤の整備などが重要である一方、寄託データが十分でない限りは、データアーカイブ機関の機能がうまく働かず、データ共有を実現することが難しくなる。IASSISTでは、データの共有・寄託を促進するために、各国のデータアーカイブ機関が行っている取り組みの紹介だけでなく、直面している課題の共有やその解決策に関する議論が行われていた。

オープンサイエンスの理念が多くの人に受け入れられたため、データの共有に向けた取り組みが、国主導の形で各国で進められていることは、IASSISTで確認できたことである。たとえば、JanezŠtebe (2019) の発表では、公的財源の充実と改善により、欧州諸国のデータアーカイブ機関が整備しつつある、という調査結果が紹介されていた。Lkhagva (2019) は、モンゴルにおけるデータアーカイブの現状を紹介され、2013年に開発し始めたオープンデータシステムはモンゴル国立統計局が主導して行われたことで、データ共有の範囲が拡大していると発表されていた。アフリカにおいても、研究データのオープンアクセスの動きが求められている。ただ、データポリシーや公開基準が不透明であるため、データ取り扱いの変化に伴う課題が顕在化していることが紹介された (Oblieye 2019)。

しかし、データ共有の理念が各国で進められていても、必ずしもデータ共有の行動に結びついていないことは、各国で共通の課題として議論されていた。その理由としては、研究者が自身のデータの公開に対する抵抗感や、データ共有に係る保存管理作業の負担感などが挙げられる。たとえば、Akdeniz (2019) の発表では、データ共有に対する研究者の意図を行動意思理論に基づいて分析した結果、研究者の態度によって大きく影響されることが明らかになったことが紹介された。特に発展途上国においては、データポリシーの策定やデータ共有の基準が定めておらず、データの公開に対する研究者からの抵抗感が強い。アフリカではオープンデータへの移行を強いられるため、研究者が誤って一貫性のないデータを公開してしまい、公開データの質が保証できない弊害が生じた事例も報告されていた (Oblieye 2019)。

データ共有に対する研究者の理解を高めるために、アメリカの大学で熱心に取り組んでいるのは、データアーカイブ機関のスタッフとライブラリアンが一体となり、調査データの企画から実査、公開まで研究者との協力体制を強化することである (Mullenburg, Steeves, McDougall 2019)。これまでデータ公開の経験がなかった研究者を密にサポートすること

で、データ共有に対する研究者の意識向上や、研究データの管理や公開・共有に寄与していると評価されている。また、CESSDA は加盟国間の連携強化を通じて、持続可能な情報共有基盤の構築によってデータ提供手続きの効率化・簡素化を追求しつつ、非加盟国との交流を重ね、欧州全域にわたるデータインフラ基盤の構築や、データ共有文化の拡大を牽引する役割を果たしている。

日本では、オープンサイエンスの国際潮流の中、分野を超えるデータ共有の基盤構築が推進されている。SSJDA は、日本学術振興会より「人文学・社会科学データインフラストラクチャー構築プログラム」の拠点機関として、データ寄託の効率性及びデータ利活用の利便性を図るようなシステム改善を試みている。今年度の IASSIST では SSJDA が行っている最近の取り組みについてポスター発表にて報告し、多くの関係者が日本のデータアーカイブに興味を示した。ポスターの内容は図 3 の通りである。

今回の報告では、日本における研究データ共有の政策動向や現状を示し、SSJDA の活動とデータ管理の変遷を紹介するうえで、データの寄託と提供手続きの効率化を図るために寄託者マイページの導入を試みた経緯を説明した。寄託者自身がマイページを通じて、寄託データに関する詳細情報の提供や利用情報の確認が可能になることによって、寄託手続きの利便性を高め、SSJDA へのデータ寄託・共有の増加につながることを期待できる。成果物の追跡管理を可視化することで、データライフサイクルのより良い循環に繋がることにも寄与できる。国際化及びオープンサイエンス時代に応じたデータインフラの整備が、日本の人文学・社会科学研究に貢献し、研究データの蓄積と公開を支えるものとして位置づけられている。

IASSIST は諸国間の課題共有や対話の場として機能し、日本のアーカイブや研究者に対しても示唆するものが多い。特に、国々の異なる背景を持ちながら、相互に協力し合う姿勢が大いに感じられた。日本におけるオープンサイエンスの実現に向けて、研究データの共有にはデータインフラ整備のほか、広い範囲の連携とデータ共有文化の醸成が今後さらに求められるであろう。

Promoting Data Deposit in SSJDA: Introducing the Improvements in Our New Data Management System

Shuai Wang^{*1}, Shuhei Naka^{*1}, Koichi Iriyama^{*1}, Satoshi Miwa^{*1}

^{*1}: Institute of Social Science, University of Tokyo

ssjda@iss.u-tokyo.ac.jp





Background	Introduction	Services and Activities
<ul style="list-style-type: none"> ■ The issue of “Promotion of the Humanities and Social Sciences Addressing Risk Society and Matured Intellectual Society”.(Council for Science and Technology, July 2012) ➢ Topic-setting program to advance cutting-edge humanities and social sciences research. ■ Start the project of Data Infrastructure for the Humanities and Social Sciences.(Japan Society for the Promotion of Science, April 2018) ➢ To preserve the existing social survey and social science data. ➢ To build an infrastructure to share data effectively. 	<ul style="list-style-type: none"> ■ The University of Tokyo, Institute of Social Science, Center for Social Research and Data Archives (CSRDA) created the Social Science Japan Data Archive (SSJDA) to support empirical research in the social sciences in Japan and has been disseminating raw data since April 1998. ■ The collection and storage have been continuing smoothly so far. Over 2,000 data sets have been put into public domain. 	<p>Construct Data Infrastructure for Secondary Use</p> <ul style="list-style-type: none"> ➢ Preserve, Gather, and Make Data Public ➢ Maintain and Develop the Online System <p>Educational Activities</p> <ul style="list-style-type: none"> ➢ Quantitative Analysis Seminars <p>SSJDA Awards</p> <ul style="list-style-type: none"> ➢ Award for Depositors ➢ Award for Excellent Papers <p>Coordination with Data Archives Worldwide</p> <ul style="list-style-type: none"> ➢ Network of Asian Social Science Data Archives ➢ SSJDA Seminar
Transition of DMS	Improvements	
<div style="display: flex;"> <div style="width: 20%; font-size: x-small;"> <p>1998</p><p>1999</p><p>2005</p><p>2006</p><p>2009</p><p>2012</p><p>2014</p><p>2018</p><p>2018</p> </div> <div style="width: 80%;"> <ul style="list-style-type: none"> • SSJDA established. Starts disseminating raw data. • The application and provision of data was through postal mail. • User information was managed by Excel. • Online Analysis System starts operations. • User database system starts operations. • SSJDA Direct (Online data management system) starts operations. • Starts trial operation of the meta data-publishing and online analysis system “Nesstar”. • All usage applications and usage reports are completed online and switched to the SSJDA Direct System completely. • Integrated the search system into SSJDA Direct and implemented a new feature of output search. • Starts system update. </div> </div>		
	Future Activities	
	<ul style="list-style-type: none"> ➢ Promoting data deposit and increasing researchers’ willingness to share research data ➢ Enhancing the convenience for depositors and data users ➢ Developing international collaboration with data archives worldwide 	

Our center has been authorized by the Ministry of Education, Culture, Sports, Science and Technology as a Joint Usage and Research Center for Social Research and Data Archives since April 2010, and also authorized by Japan Society for the Promotion of Science as a Core Institution for Constructing Data Infrastructure for the Humanities and Social Sciences since October 2018.

図 3 IASSIST2019 で報告したポスター

7.3 データの利用に関する課題

本小節では、“Data Usage”と“Research Data Support”のセッションで鍵概念となっていた“Data Visualization”という観点から、データ利用に関する課題を指摘しておきたい。Data Visualization とは、図表、グラフや地図などを用いて情報源を視覚的な描写へ変換

することである (<https://duke.app.box.com/v/IntroDataVisS18/>)²³。では、どのような情報源を視覚化する必要があるのだろうか。ここでは、データへのアクセス性を向上させるポータルサイトとアーカイブが保有するデータに焦点を当てたい。

Paige and Kellam の報告では、近年のアメリカにおいて“Open Data Portal”(ODP)が行政機関のあらゆる水準（ローカルからインターナショナル）において急増している背景について言及があった。ポータルサイトの増加によってデータへのアクセス性は以前より良くなっているが、各ポータルサイトが使いやすいとは限らない点が指摘されていた。多くの行政機関はデータセットのアップデートは行っているが、そのポータルがデータセットに関連する **Contextual Information**（文脈的な情報）を提供しているとは限らない。そのため、ユーザーがデータを積極的に利用するには至っていないとのことであった。その状況を解消する一つ的手段として、データの扱いに関するオンラインチュートリアルを設けることが提案されていた。

ODP の事例としては、**The U.S. Digital Analytics Program (DAP)**に関する報告がなされていた (**Marshall and Dianat** 報告)。DAP は分析ツール、分析のトレーニングや政府関係機関のサポートをすることが目的である (<https://analytics.usa.gov/?=dg>)。このウェブページでは、トップページにおいて基本的な情報を視覚的に理解できるようになっている。たとえば、何人の人が行政機関のウェブページにアクセスしているのか、あるいは人びとが使っているコンピュータの情報（デスクトップ、ラップトップやタブレットという機器、**Chrome**、**Safari**、**IE**などのブラウザ、**Windows**、**iOS**、**Android**などのOS）がある。加えて、ファイルダウンロード数の多い情報がひと目でわかる。

一方、アーカイブが保有するデータについては「データ利用の測定」が参考になるだろう (**Pienta, Hemphill and Akmon** 報告)。この報告では、データ利用／再利用を予測する要因などを **ICPSR** の利用記録データを用いて、経時的な利用の軌跡が分析されていた。多くのデータはリリース直後に活用される傾向であるが、一定年数が経過するとほとんど利用されないデータとなっている現状が明らかにされていた。この点については、**SSJDA** の利用実績とも類似する傾向であると考えられる。逆に言えば、データリリースからある程度の時間が経ったデータの利用を促進するためには、アーカイブ側が何らかの工夫を施す必要があるかもしれない。

そのためには、上述した ODP の機能の一つとして「保有データの視覚化」を組み込むことが考えられる。ここでの視覚化においては、ユーザー側とアーカイブ側の双方からの観点があると面白くなるかもしれない。前者は、「ユーザーがどのデータをどの程度利用しているのか」というアーカイブのトレンドを示すことである。そのトレンドに加えて、どのような雑誌に掲載されているか、トピックは何かという情報もあると研究テーマを絞り込むうえで参考となるだろう。

²³ Duke University では、**Data and Visualization Services Department** が創設されており、ワークショップやサービスの一部を公開している (<http://libcms.oit.duke.edu/data/>)。

アーカイブ側の観点としては、データが収集された経緯やその特徴を簡単な視覚化を通して ODP のトップページで示すことである。SSJDA では、利用実績は少ないものの、データの希少性が高いものを数多く保有している。いわば、「手つかずのお宝」が眠っているのである。もちろん、貴重なデータを保存すること自体が意味のあることだと考えているが、データは利用されることによってより輝きが増すかもしれない。そのためには、アーカイブ側は「データをよりよく見せる」工夫を考えることが必要となるだろう。

7.4 データの管理・運用に関する課題

最後に、データの管理・運用に関する課題が二点挙げられる。一点目は多言語化について、二点目はアーカイブ認証についてである。まず多言語化に関して、データアーカイブの国際連携や利用者の多様化を見込んで、現在 SSJDA でも、多言語化（まずは英語化を主とする）への対応を進めている²⁴。その内容は、メタデータ（調査概要）の翻訳、とくに利用頻度の高い調査データをピックアップしての翻訳（主に調査票と個票データについて）、ウェブサイトの英語化など、多岐にわたる。

そうした作業を進める上で、現在、調査データの翻訳作業に関するポイントや注意点が多く上がってきており、それを海外の事例に即して検討している段階にある。そこで上がってきている課題には、たとえば翻訳の質チェックの重要性、翻訳内容や手続きのルール化の必要性、予算と翻訳ペースのトレードオフ、メタデータ作成作業へのフィードバックといった点がそれにあたる。

こうした点を検討する際、今回の IASSIST2019 でも報告されていた、欧州の SERISS (Synergies for Europe's Research Infrastructures in the Social Sciences) プロジェクトから学べることは多い。たとえば SERISS の翻訳作業に関する中間まとめでは、サーベイリサーチにおける翻訳の重要さと難しさについて述べられ、よい翻訳とは何か、Survey expert with translation skill のような人材が監督者であるべき、チームで検討することの重要性、質の評価をどうするか、ガイドラインの策定、といった点が議論されており、SSJDA が直面する翻訳の課題を考える上でも非常に有用である²⁵。

また、翻訳の質チェックの際の課題として、小さな誤訳でも重大な影響を及ぼしうることや、細かなニュアンスをよくくみ取って翻訳することの重要性、翻訳者の資質や訓練等が重要であることなども述べられている²⁶。さらに、SERISS の翻訳プロジェクトでは、社会調査やコンピュータ言語学の専門家など、異なる分野のメンバーを含むチームでの検

²⁴ SSJDA の利用者が多様化する中で、海外機関在籍の利用者や、日本にいる留学生の利用者などの利用者層も増えていることがこうした背景にある（三輪・佐藤 2018: 493）。

²⁵ Diana Zavala-Rojas (ESS ERIC) – Translation and translation assessment in comparative surveys のスライドが有用である (<https://seriss.eu/resources/sen/>より入手可能)。

²⁶ 注 25 と同様のリンクより、D3.5 – First SERISS Symposium on synergies between survey translation and developments in translation sciences. Minutes of the expert meeting より、Q1: Why is survey translation assessment の部分などがとくに有用である。

討を行い、現在翻訳のための統一的なガイドラインを策定中の模様である。

二点目はアーカイブ認証の取得である。現在 SSJDA が取得を目指している Core Trust Seal (CTS) 認証の取得には、以下のようなメリットがあると言われている (L'Hours et al. 2019)。

- アーカイブ組織に関する十分な情報の把握と整理 (内部評価) により、データの長期保管を可能にすること
- 外部のレビュアーからの指摘 (やコミュニケーション) を通じて学ぶことやそのプロセス自体がアーカイブの改善にとって有意義である
- アーカイブが安全に、かつ一定の永続性をもってデータを保管できることを示すことは、アーカイブ自身に加え、研究者や利用者それぞれにとってもメリットとなる
- アーカイブのスタッフ間で、こうした点についての意識を共有するといったコミュニケーション自体が、重要で有意義なものである
- アーカイブの潜在的なリスクの特定とその対応の改善につながりうる
- アーカイブ運用のための文書化の強化と、手続きの透明性の向上につながる

実際に現在準備を進めている手ごたえからして、これらの点に共感する部分は大きい。認証取得の準備自体に時間や労力がかかるものの、国際的な基準が定まったことを契機に、その取得に向けて努力することは、意味があることだと言える。

むろん、アーカイブ認証の取得には課題も少なからずある。一般的にデジタル資産の保管には多くのコストがかかる (予算、人員、時間など) ため、取得後もその認証を継続できるかどうかは、各機関の今後の組織や運用状況にも左右される。認証取得にあたっては、アーカイブの運用が、個人情報や著作権などの様々な法的・倫理的規定に照らして適切なものになっているかどうかをチェックする必要も出てくる。こうした点は、きちんとクリアできればむしろメリットになりうる点ではあるが、その準備のために多方面の協力が必要となるだろう。また、認証の取得には、日本円で 10 万円超の支出が必要となる。2019 年 10 月現在、3 年間有効な認証取得のために、1,000 ユーロの *administrative fee* が必要とされており、認証を取得する場合は、こうした費用も予算として考える必要が出てくるだろう。

7.5 おわりに

最後に、本稿で挙げたようなデータの保管や共有に関するさまざまな問題について考えることとは、つまるところ、データアーカイブ側にとっては業務の総点検を意味しているとも言える。また、研究者側にとっては、今後社会調査を実施する際、安全かつ長期的にデータを保管 (場合によっては共有) できるかどうかまでを見越して、データの創出から分析、管理、寄託、共有を含んだ、社会調査におけるデータライフサイクルの全工程にかかわる項目について、丁寧に検討していくことを意味するだろう。本稿で述べた事例が社会科学系のアーカイブや研究者にとって役立つことを願い、結びとしたい。

参考文献

- 朝岡誠・高橋かおり, 2019, 「海外データアーカイブの動向 2——IASSIST 年次大会の報告から」『社会と統計』5, 33-41.
- L'Hours, Hervé; Mari Kleemola, and Lisa de Leeuw, 2019, “CoreTrustSeal: From Academic Collaboration to Sustainable Services,” *IASSIST Quarterly* 43(1), 1-17.
DOI: <https://doi.org/10.29173/iq936>
- 前田幸男・佐藤慶一・安藤理・米倉佑貴・朝岡誠・入山浩一・角井佑, 2011, 「Data Documentation Initiative の利用可能性」SSJ データアーカイブリサーチペーパーシリーズ No. 46. (<https://csrda.iss.u-tokyo.ac.jp/rps/RPS046.pdf>).
- 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』4: 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.
- 三輪哲・佐藤香, 2018, 「オープンサイエンス時代における社会科学データアーカイブの意義: SSJ データアーカイブのこれまでとこれから」『情報の科学と技術』68 (10), 489-494. DOI: https://doi.org/10.18919/jkg.68.10_489
- UK Data Archive, 2011, “Managing and Sharing Data: Best Practice For Researchers, 3rd Edition.” (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による翻訳, 2013, 『データの管理と共有: 研究者向け最良事例』).
- 今回参照した報告やワークショップの一覧 (すべて IASSIST2019 より)
- Akdeniz, Esra, “The Role of Subjective Norm in the Prediction of Researchers’ Intentions to Share Data: Explanations by the Theory of Planned Behavior.”
- Akmon, Dharma and Susan Jekielek, “Restricting Data’s Use: A Spectrum of Concerns in Need of Flexible Approaches.”
- Amy Pienta, Libby Hemphill and Dharma Akmon, “Data Use Trajectories Over Time in the Social Sciences.”
- Barbalet, Suzanne, “Sustainable European Multilingual Vocabularies: A Model for Cooperation in Metadata Management among European Data Archives.”
- Balkan, Lorna, “Evaluating Translation Quality in the ELSST Multilingual Thesaurus.”
- Borschewski, Kerrin, “How to deal with Persistent Identifiers? A Survey on PID usage within CESSDA ERIC.”
- Edmunds, Rorie and Heather Leasor, “Building Trust in Scientific Data: CoreTrustSeal Certification and the World Data System.”

- Fleet, Robert. "Video Game Data: The Untouched Potential of the Gamer as a Data-Generator."
- Fry, Jane and Carla Graebner, "DMPs: Hmmmmmm, I Think I Know What They Are But."
- Hemphill, Libby, "Researcher Passports and Open Badges: The latest in ICPSR's Efforts to Provide Secure, Trustworthy Access to Restricted Data."
- Hemphill, Libby. "Updates on ICPSR's Social Media Archive (SOMAR)."
- Karcher, Sebastian. "Archiving Qualitative Online Source: Obstacles, Strategies, and Lessons."
- Karcher, Sebastian. "Working with Sensitive Research Data: The Qualitative Data Repository's WSRD Project."
- Kocar, Sebastian and Nicholas Biddle. "Anonymization of Online Panel Survey Data: A Case Study."
- Kirby, Jasmine and Cameron Riopelle, "It's Not Quite Wakanda: Preserving Indigenous Cultural Heritage Augmented and Virtual Reality Environments."
- Lkhagva, Myagmarsuren, "Statistical Open Data: Comparing Mongolia with Other National Open Data Systems."
- Lyle, Jared. "ICPSR and CoreTrustSeal: Repository Certification Experiences and Opportunities."
- Marshall, Ann and Oldooz Dianat, "What the U.S. Digital Analytics Program (DAP) Data Can & Cannot Tell Us: Investigating U.S. Government Website Usage Over Time." Obileye, Olatunbosun, "Democratization of Data in an Emerging Economy."
- Moss, Elizabeth, and Jared Lyle, "Informal Data Citation: The Impact on Tracking Shared Data Reuse."
- Muilenburg, Jennifer, Vicky Steeves, Janet McDougall, "Partners in Reproducibility: Working with Researchers and Data Producers to Enhance Sharing, Reproducibility, and Long-term Access to Research Data."
- Paige, Bonnie and Lynda Kellam, "If You Build it, Will They Come? Sustaining Open Data Portals."
- Perry, Anja, Jane Fry, and Olof Olsson. "Introduction to DDI: Basic Concepts and How to Develop Skills for Training Researchers."
- Silberman, Barriers Roxanne, Kamel Gadouche, Stefan Bender, and Peter Bøegh Nielsen, "Transnational Access to Controlled Microdata for Research: Overcoming Barriers."
- Štebe, Janez, "Examining Barriers for Establishing a National Data Service."
- Wang, Shuai, Shuhei Naka, Koichi Iriyama, and Satoshi Miwa, "Promoting Data

Deposit in SSJDA: Introducing the Improvements in Our New Data Management System.”

Welch, Matthew, “Statistical Disclosure Control: A Graphic Interface for sdcMicro.”