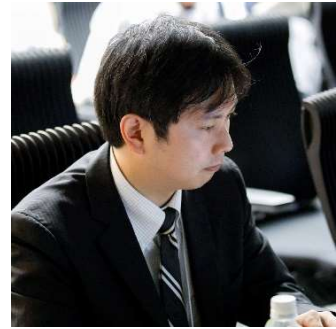


# J2Cat サロン

## データインフラの最前線

### 機関リポジトリとデータアーカイブの接点を探る



南山泰之（みなみやま・やすゆき）

情報・システム研究機構国立情報学研究所

オープンサイエンス基盤研究センター 特任技術専門員

図書館・図書室勤務のご経験を持ち、データキュレーションを研究テーマにされている南山泰之さんが考える日本の機関リポジトリとデータアーカイブの接点とは？

#### データキュレーションの魅力と日本の現状

——研究テーマであるデータキュレーションの魅力とは、

データは単独で解釈ができないので、データを再利用するためにはデータがどのように作成されたかという情報を整理する必要があります。加えて、データがどのように使われたか、といった文脈をデータに付与していくことで、データを理解できる人が増え、より多様な使われ方に繋がっていくところに魅力を感じています。また、このようなデータキュレーションを情報システムで扱うことで、データが持つ知識や関連性の表現を支援したいと考えています。

——現在の大学図書館のあり方と機関リポジトリの関係についてお考えをお聞かせください。

大学図書館はこれまで、紙というか、実物があるものをどう整理して研究者に提供するかを中心に考えてきました。20～30年前から目録はオンラインで提供されるようになりましたが、目録に留まらず本文ファイル自体の提供や、オンライン上でのデータ解析といった動きが出てきた中で、多くの大学図書館はまだウェブを前提とした研究支援までたどり着けていない

2005年、中央大学法学部法律学科卒業。2005年より国立極地研究所情報図書室に勤務。2007～08年、第49次日本南極地域観測隊に参加。東京大学駒場図書館（2011～14年）、国立極地研究所情報図書室（2014～18年）、東京財団政策研究所（2018～19年）を経て現職。

という感触です。機関リポジトリは紀要や論文のデータを中心にコンテンツを掲載し、一步進んだ研究支援を実施してきましたが、それは研究の中心部分ではなくて出口だけです。まだ扱えていない研究の本質の部分にどう関わっていくかが大きなチャレンジだと捉えています。

——日本のデータアーカイブの現状をどう見ますか。

人文学・社会科学の分野では、研究対象となる事象がチームによって大きく異なるため、データ共有の必然性がなかなか見出されず、一部の実践に留まってきた印象があります。この点、地球科学や宇宙科学、生命科学など、国際的なデータ収集、共有を前提に研究が進む分野とは様相が異なっています。一方で、最近ではデジタルアーカイブをキーワードに、人文学分野では自分たちの手元にあるものをデジタル化して活用しようという潮流が広がっています。さらに、この潮流に国立国会図書館のような大規模な機関が参画したことによって、小規模・中規模の機関でもデータアーカイブを作る機運が高まり始めています。今後、より多くの小規模機関が参画していくことが、分野の裾野を広げていくことになると思います。

<sup>1</sup> データキュレーションとは、データの整理・統合・編集な

どを通して利活用しやすい形に加工するデータ処理のこと。

## 学振の人文・社会科学データインフラ構築推進事業への示唆

——学振の人文・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取組について、忌憚のないご意見をいただければと思います。

国内のデータインフラ整備は極めて重要です。公的資金で得られたデータの長期保存だけではなく、保存されるデータの質の向上や新しい比較研究にも繋がります。キュレーションの観点からも、データの繋がりを見えるようにするにはデータアーカイブがないと始まりませんから。

ただ、それを支えるには、コミュニティ、システム、そして資金が必要です。特にコミュニティは、データを利用する人だけでは駄目で、データアーカイブを発展させるために積極的に改善を提案する人が必要です。JDCat<sup>2</sup>の構築は、データアーカイブ発展のためのコミュニティ作りのきっかけになっていると思います。

——人文・社会科学の分野横断コミュニティ形成の可能性とは。

分野横断コミュニティ形成は大きな目標になり得ると思いますが、一方でコミュニティが大きすぎると密な議論がしづらいので、小規模に共通の関心事を議論できる場が複数立ち上がっていくことが良いと思います。例えば、人文分野・社会科学分野に共通する話題としては、データを保存するときのフォーマットなど実務に密着した話題が考えられそうです。また、JDCatの情報システムは JAIRO Cloud<sup>3</sup>のベースとなる WEKO3 で構築されているため、図書館の機関リポジトリ担当者も交えたりリポジトリのシステム運用に関する話題もあり得るかもしれません。

——海外では、大学とデータアーカイブの連携が強いという印象です。

例えば、イギリスの UK Data Archive (UKDA) では、大学のリポジトリ関係者が一定期間 UKDA に出向して実務を経験することでコミュニティを形成しているようです。国内の機関リポジトリ担当者はまだデータアーカイブとは何か、といった明確なイメージを持っていないと推察されますので、例えばデータアーカイブと機関リポジトリで共同の実務研修や人材交流をすると、連携に向けた道筋が見えるかもしれないと思っています。

——理想のデータアーカイブに必要な人材と体制とは。

分野の知識もあってデータを解釈できる人、さらに持続的な情報収集能力を兼ね備えた人材が必要です。ただ、データアーカイブの実践は研究成果になりづらいので、研究者が全てを担うことは難しいという認識です。一方で、データアーカイブの周辺にはリポジトリを運営している図書館員や、リポジトリ運営自体が機関の評価になるという認識を持っている方々も一定数おり、またデータ整理の経験は教育面でも重要と考えられます。

今後のあり方として、全体をハンドリングするベテラン研究者、中堅として運用を支える図書館員や技術者、さらに教育を兼ねてデータキュレーションなど実務の一端を担う大学院生といった形で、様々な職種の方々との連携を前提にした体制構築も考慮に値するのではないのでしょうか。

(座談会開催：令和3年3月23日／編集：深澤はるか)

<sup>2</sup> JDCat とは、Japan Data Catalog for the Humanities and Social Sciences の略。人文・社会科学総合データカタログ。学振が実施する「人文・社会科学データインフラストラクチャー構築推進事業」の成果の一部で、複数のデータアーカイブのメタデータが一括検索できる。本事業は、学振が平成30年度から実施する事業で、人文・社会科学研究に

係るデータを分野や国を超えて共有・利活用する総合的な基盤の構築により、研究者がデータを共有しあい、国内外の共同研究等の促進を目指している。

<sup>3</sup> JAIRO Cloud とは、オープンアクセスリポジトリ推進協会 (JPCOAR) と国立情報学研究所が共同運用する、クラウド型の機関リポジトリ環境提供サービス。