

# J2Cat サロン

## データインフラの最前線

### 多様な研究データと再利用を 踏まえた支援サービスに期待

酒井 由紀子（さかい・ゆきこ）

帝京大学 共通教育センター 准教授

人社データインフラ事業に運営委員会委員として携わられている酒井由紀子さんに、図書館情報学の観点からお話をお聞かせいただきます。

#### —ご自身の研究についてお聞かせください。

学術コミュニケーションの変容を探る研究の一環として、慶應義塾大学の倉田敬子先生、宮田洋輔先生らと一緒に全分野の研究データの再利用の実態調査を行っています。自身の研究分野である図書館情報学の調査研究でも、社会科学系の質問紙調査をたびたび採用するなど、研究データの収集、利用、再利用を実践しています。

#### —人社データインフラ事業とは、どのように関わられるようになったのですか？

研究データ再利用に関する調査の一環として、2017年に米国および日本の社会科学系データリポジトリの活動について現地調査を行いました。さらに、2018年には米国の大学図書館における人文学・社会科学を中心とした研究データ支援サービスについての現地調査を行いました。このような経験もあり、ご縁があって現在では人社データインフラ事業運営委員会の委員を務めています。

#### —ではまず、人社データインフラ事業で扱うデータについて、どのような視点で捉えられていらっしゃるかお聞かせください。

研究データ再利用の文献調査を通じて、人文学と社会科学では、収集される研究データの性質が異なっているという印象を持っています。一般に、多くの社会調査データは一定のモデルをもとにデータが取得されているため、すでに構造化されています。一方で、人文



慶應義塾大学メディアセンター（図書館）勤務（1983年～2013年）。慶應義塾大学より博士（図書館・情報学）取得（2012年）ののち、慶應義塾大学文学部准教授（2013年～2018年）。東京財団政策研究所政策データラボ シニア・マネジャー、研究員（2018年～2020年）を経て、2020年より現職。

系で収集される研究データは、もともと研究を想定して構造化されたデータというよりも、新聞記事などのテキストや、数値であっても異なる角度から分析を行うなど、研究目的に応じた構造を与える必要のある「素材」であることが多いのではないのでしょうか。

#### —「素材」という表現は初めて聞きました。

例えば、医学分野のデータであるカルテは、どちらかと言えば素材であると考えています。また、研究グループ内では政府統計データなども素材に近いだろう、といった議論をしています。海外の文献では business resource という表現が使われたりもするようですが、いずれにせよ分析に適した構造化がなされていない状態を指す言葉として使っています。

#### —「素材」と「研究データ」はどのような使い分けになるのでしょうか。

研究データの定義は研究者によって異なりますので、構造化されているかどうかを使い分けの基準として考えるのは難しいと思われま。素材を研究データとしない立場もありますが、研究の再現性や透明性を確保するためだけでなく、再利用というメリットを想定すると、個人的には研究データを広範囲に捉え、素材も研究データの一部として扱う方が適切であると考えています。一方で、データを保存・共有する範囲、研究目的にかかわる文脈を表すメタデータの記述のありかたは、両者で大きく異なるものと推測されます。研究データの下位カテゴリをどう捉えるか、まだ議論が必要

だと感じています。

—続いて、各分野における研究データの共有について、どのように捉えていらっしゃるかお聞かせください。

各国の調査研究の結果を見ている限りでは、現状のところデータ共有は研究の公正性を確保するために必須の行為として捉えられることが多いようです。このような観点からは、迅速に手間をかけずにデータを共有できることが多くの研究者によって何より重要となります。したがって、使い慣れているデータリポジトリやデータアーカイブがあれば、そちらが最優先されることになるでしょう。

—研究者がデータアーカイブを選択する際、ほかに重視するポイントはどこになるのでしょうか。

本来、研究データの共有は検索や再利用を目的に行われます。検索や再利用の場面では、データを収集した研究の文脈を理解し、新たな研究にそのデータが使えるかどうかを判断する必要があるため、こういった情報を適切にメタデータとして記述できることは重要な点になるはずです。実際、再利用が習慣化している分野では標準化された構造やメタデータをもつデータアーカイブが優先されています。

もう一つのポイントとして、研究データを再利用する際にはリポジトリに対する信頼が重要、という研究結果もあります。通常、リポジトリの信頼性は各分野の研究者ネットワークを通じて担保されています。機関リポジトリのように身近な組織が運営しているシステムの場合、データ共有のために手厚い支援が受けられ共有の場として広く使われるようになれば、おのずと信頼も得て、さらに検索や再利用の対象として認識され活用される可能性はありそうです。

—機関リポジトリは、どのような場合に活用が考えられそうでしょうか。

特定の論文に付随する研究データを探す場合は、論文中のデータ利用可能性ステートメント (DAS)<sup>1</sup>を参照して登録先のリポジトリを検索することになります。

この場合、機関リポジトリが第一の選択肢になる訳ではありませんが、機関リポジトリでの公開が一般的であるコンテンツ、例えば大学で出版されている紀要論文に付随する研究データなどは活用の対象となりやすいかもしれません。

また、自分の研究に合った研究データを広く検索したい場合に機関リポジトリを検索することも考えられます。一般論として、第一には自分の分野データアーカイブを検索するものと考えられますが、適切なデータが見当たらなければ汎用的なデータアーカイブへも検索対象を広げていく人が多いと考えられます。その際、信頼できるデータアーカイブとして機関リポジトリが候補に挙がる可能性はあると思われます。

—関連して、人文学・社会科学データの再利用についてより詳しくお聞かせください。

研究データの再利用は、特定の研究で収集された「データセット」だけではなく、前述しました「素材」も対象になり得ます。ここで言う再利用は、いわゆる二次分析に限定されず、論文本体に含まれる図表やサプリメンタルデータの閲覧、及びこれらの経験から得られるちょっとした気づきなどを含んでいます。研究者は、自分自身でもデータを収集していますが、こういった small fact を確認する際に研究データの再利用を行っています。

—研究データの再利用は、日常的に行われているのですね。

はい。また、このような研究データの再利用を前提とする場合、先ほど挙げた論文のサプリメンタルデータや、本文に埋もれているデータ、分野に限定されない「素材」となるような統計、テキストなどを効率よく検索できる仕組みがあると、より広い分野で研究データの再利用が進むのではないかと考えています。

—例えば、どのような仕組みが考えられるのでしょうか。

図書館情報学の分野では、研究データの検索パター

---

<sup>1</sup> データ利用可能性ステートメント (Data Availability Statement) とは、研究論文を作成する際に分析・生成されたデ

ータにどこからアクセスできるか、またどのような条件でアクセスできるのかを記載するもの。

ンに焦点を当てた研究もいくつか存在しますが、正直なところまだ良く分かっていないというのが現状と思われま。自分自身の経験に照らしてみると、実際にデータを探す際には、結局論文などを手掛かりに芋づる式に探している印象があります。データの場合、調査方法が極めて重要な観点になりますので、例えばどのようなプロトコルでデータが作成されたのか、JDCat<sup>2</sup>の中で繋がってくると効率の良い検索が実現できるかもしれません。

—実現に向けた課題はどういった点にありそうでしょうか。

前提として、研究データを再利用するイメージがまだ曖昧であり、学術コミュニケーションのエコシステムの中に組み込まれていない印象があります。例えば、学術誌のポリシーとして、サブリメンタルデータ<sup>3</sup>はPDFでの登録しか受け付けられないケースが散見されます。また、各リポジトリのデータ登録フローもまだ洗練されているとは言えません。当面は、特定分野の研究者の再利用パターンを理解し、研究データを適切に公開できる人材、検索してくれる人材の育成が課題になりそうです。

—では、引き続きデータアーカイブの人材についてお聞かせください。

データアーカイブを支える人材には、データキュレーター<sup>4</sup>、データライブラリアン<sup>5</sup>と呼ばれる職種が位置づけられます。これらの養成・育成プログラムと雇用については、北米の医学図書館界で2000年に提唱された話題になった「インフォメーションリスト」、いわば医療や医学研究に深くコミットするスーパーライブラリアンの雇用／育成に関する議論や実践が思い起こされま

す。当時の動向として、臨床部門では、医学図書館員が臨床医学の基礎を学んだ上で、カンファレンスや回診に同行するなどして医療職のニーズを把握し、必要なエビデンスの収集・提供をするようになりました。一方、基礎研究では、例えばバイオインフォマティクスの教育・訓練を受けたスペシャリストが、新たに医学図書館に雇用されることが多かったようです。

—北米では、どのような契機でこのような活動が広まったのでしょうか。

歴史的には、1970年代に病院図書館員が密接に医療に関わった経験があるようです。これらの実践が、「インフォメーションリスト」提唱に伴いいわば医学版のエンベデッドライブラリアン<sup>6</sup>として北米で想起され、この知見を拡大させるための助成金がつきました。この潮流に沿い、戦略的に医療の現場へ図書館職員を送り込んで実績を積む大学や、バイオインフォマティクスの専門家を図書館員として雇用して教育及び研究支援を行う大学も現れ始めました。すなわち、図書館情報学を出自に持つデータライブラリアン、生命科学等の主題領域を出自に持つライブラリアンの両方が起点となり、データライブラリアンの職種が確立されていったと言えます。

—現在では、誰がどのような研究データ支援サービスを展開しているのでしょうか。

2018年に現地調査をした北米では、研究データ支援が大学において必須のサービスとして注目されていました。サービスの実施主体は様々ですが、大学図書館、研究支援部門、情報技術部門等の間で、ある意味取り合いになっている状況とも聞き及んでいます。また、視察で会ったデータキュレーターやデータライブラリア

<sup>2</sup> JDCat とは、Japan Data Catalog for the Humanities and Social Sciences の略。人文学・社会科学総合データカタログ。学振が実施する「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。

<sup>3</sup> フォーマットや紙面の都合により論文本体には掲載できないものの、当該論文にとって有用な図、写真、生データの表、動画などの補足資料全般を指す。

<sup>4</sup> 研究データの利活用のために、コラボレーションなども利用して潜在的な価値を顕在化するデータ専門職。

<sup>5</sup> 研究データの整理のために、メタデータ付与やデータ連携などを安定的に行うデータ専門職。

<sup>6</sup> 日常の業務において図書館を離れ、利用者が活動している場から、利用者との活動をともにしつつ情報サービスを提供している図書館司書を指す。

ンのキャリア形成は様々で、背景の異なるスペシャリストがそれぞれの既得の知識やスキルを活かし、あるいはさらに必要なスキルや経験を追加してサービスにあたっていました。例えば、生物学の学位を持つサブジェクトライブラリアン<sup>7</sup>がサービスをデータ支援にまで拡大していたり、大学図書館員が夏季集中の情報技術を中心とした科目から成る Data Matters を受講して研究データ支援サービスに備えていたりしています。また、大学図書館内にオフィスを移転した社会科学分野のデータアーカイブでは、iSchool<sup>8</sup>の修士課程でアーカイブとレコードマネジメントの教育を受け、かつ社会科学分野の複数の研究プロジェクトで支援の実践を積んだスタッフが、主にコンサルティングを中心としたサービスで活躍していました。さらにほかの大学では、新たに立ち上げた人文学・社会科学分野の研究データ支援サービス実践のために、情報技術者を新たに雇用したり、経済学研究科の大学院生をアシスタントとして雇用したりしていました。

—日本で同様のサービスを展開するためには、どこが課題になるでしょうか。

日本では、既に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研究データ管理人材育成のためのオンライン教材<sup>9</sup>を開発し、広く利用されています。次の課題としては、図書館あるいは各分野のスペシャリストが、それぞれのキャリアや経験に基づいて必要な追加的なスキルを養成、育成できるプログラムを、どこがどの程度用意できるかが課題でしょう。研究支援サービスをするのだから自身にも「研

究」の経験は必須で、その意味では大学院はふさわしいですが、社会人課程が増えたとはいえ、正規のフルタイムの大学院生しか受講できないプログラムではいかにも柔軟性に欠けます。北米のライブラリアン養成課程で最近最も多く新設された科目は「データサイエンス」ですが、日本でも同様の科目が図書館情報学の専門課程として導入されるのであれば、聴講生等も参加出来たり、図書館関係団体との連携で継続教育の一環として受講できたりする仕組みがあってもよいと思います。ほかには北米の Data Matters のように、ビジュアライゼーションや統計解析等の基本的な技術支援のための科目も日本にも必要でしょう。このような情報技術系の育成科目には、最近増えている学部や大学院のデータサイエンスの専門課程の協力を仰いで、支援サービスの担当者が学べるプログラムを開発するといった動きがあってもよいのではないのでしょうか。

—研究データ支援サービスを始めるためには、どこから手をつけると良いでしょうか。

令和3年度学術情報基盤実態調査大学図書館編<sup>10</sup>によると、国立大学に限定すれば約半数が「研究データ管理やオープンサイエンスの対応」を大学図書館の機能面での課題として認識しているものの、小中規模も多い公私立大学を合わせると、全体では2割程度と関心が必ずしも高くないようです。また、日本では研究データ管理計画が研究助成申請の際の必須事項となる<sup>11</sup><sup>12</sup><sup>13</sup>など、研究を実施するためにクリアすべき手続きに目が向きがちですが、利用者が必要とする研究データ支援サービスはもっと幅広いはずであり、自機関に

<sup>7</sup> ある特定主題分野の図書館情報資源と利用者に特化したサービスに従事するライブラリアンのこと。

<sup>8</sup> 図書館情報学をある程度の基盤としつつも、「図書館司書養成」の枠を超え、幅広い情報領域に関する研究・教育を志向する大学院課程のこと。

<sup>9</sup> 研究データ管理サービスの設計と実践。

<https://lms.nii.ac.jp/enrol/index.php?id=52>

<sup>10</sup> 令和3年度「学術情報基盤実態調査」の結果報告について—大学における大学図書館及びコンピュータ・ネットワーク環境の現状について—。

[https://www.mext.go.jp/b\\_menu/houdou/2021/1418398\\_00005](https://www.mext.go.jp/b_menu/houdou/2021/1418398_00005)

.html

<sup>11</sup> オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針。

[https://www.jst.go.jp/pr/intro/openscience/policy\\_openscience\\_r4.pdf](https://www.jst.go.jp/pr/intro/openscience/policy_openscience_r4.pdf)

<sup>12</sup> AMED における研究開発データの取扱いに関する基本方針、AMED 研究データ利活用に係るガイドライン、データマネジメントプラン。

<https://www.amed.go.jp/koubo/datamanagement.html>

<sup>13</sup> NEDO プロジェクトにおけるデータマネジメントについて。

[https://www.nedo.go.jp/jyuhoukoukai/other\\_CA\\_00003.html](https://www.nedo.go.jp/jyuhoukoukai/other_CA_00003.html)

おける研究者のニーズを汲み取るところから始める必要がありそうです。

もっとも、実際のところ養成・育成プログラムの開発には時間がかかります。研究支援の適性や既存の知識や技能が十分であるか、あるいは基本的な関連研修を受けた意欲のある人材がいれば、研究プロジェクトの一員に送り込み、研究者のニーズを汲んで支援を行う経験を積んでもらうというのも一案かと思います。先ほど紹介した社会科学分野のデータアーカイブでコンサルティングサービスに従事するスタッフも、iSchool で受けた教育はごく基礎的なもので、何より役立ったのは研究プロジェクトでの実践だったと語っていました。

—日本の場合、どのような人材がサービスを担う可能性があるのでしょうか。

私自身も大学図書館員でしたので、心情的には図書館員に頑張ってほしいところですが、日本の大学図書館がどこまで研究データサービスを推進するかは、リーダーシップに依存するのではないかと考えています。また、大学図書館として組織的な取り組みがなければ、適性のあるライブラリアンが、リサーチ・アドミニストレーター（URA）のように図書館外で活躍する可能性も高いのではないかと推測しています。

—最後に、人文学・社会科学のデータリポジトリ／データアーカイブが果たしていく役割について、お考えをお聞かせください。

まず、人社データインフラ事業で構築した複数のデータリポジトリ／データアーカイブの横断検索システムは長年の懸案だったと聞き及んでおり、関係の皆様のご尽力に敬意を表したいと思います。そのうえで、やはり必要な人材の養成・育成が人文学・社会科学データインフラの最大の課題です。現状では、支援されるべき研究者が研究データ支援サービスを実施している状況であり、早急な解決が望まれます。

課題解決の方向性としては、既存のデータリポジトリ／データアーカイブが、大学や機関の中で研究データの拠点になっていくイメージを持っています。例えば、学内における研修の機会提供、統計相談、人材育

成などの活動を通じて、学内に活動成果を還元していく仕組みがあると、より支援を集めやすくなるのではないかと考えています。特に拠点機関の皆さまには、データ保存・再利用のためのシステムを提供するだけでなく、その業務を通じて養った知識やスキルを活かし、人材養成・育成の場としても、研究者の研究データにかかわる支援サービスの間としても一定の役割を果たしてくださることを期待します。

（座談会開催：令和4年9月1日／聞き手：南山泰之）