

資料と公共性 : 2018年度研究成果年次報告書

岡崎, 敦

九州大学大学院人文科学研究院 | 九州大学大学院統合新領域学府 : 教授

市澤, 哲

神戸大学大学院人文科学研究科 : 教授

石田, 栄美

九州大学附属図書館 | 九州大学大学院統合新領域学府 : 准教授

後小路, 雅弘

九州大学大学院人文科学研究院 : 教授

他

<https://doi.org/10.15017/2230688>

出版情報 : 2019-03-14. 九州大学大学院人文科学研究院

バージョン :

権利関係 :

シンポジウム「オープンデータと大学」へのコメント
—アーカイブズ学の視点から—

清原 和之

2019年1月30日に九州大学大学院統合新領域学府ライブラリーサイエンス専攻と九州大学附属図書館の共催で、大学における多様な情報資源のオープン化の問題を議論する「オープンデータと大学」と題するシンポジウムが行われた。まず、趣旨説明を行った岡崎敦氏（九州大学）から、大学で行われた活動に関する情報の管理は誰の責任と負担のもとでなされるのか、その責任体制のあり方やそこでの専門職の役割について、問いが投げかけられた。そして、中村覚氏（東京大学）から、東京大学デジタルアーカイブズ構築事業における学術資産のオープンデータ化の取り組みが報告された。次に、南山泰之氏（東京財団政策研究所）から、研究データの管理に関する動向整理がなされ、研究データの利活用に向けた課題として研究者の意識や情報インフラの問題、専門職の認知度などの問題が指摘された。続いて、畑埜晃平氏（九州大学）から、「くずし字」の文字認識問題への取り組みとそのデータのオープン化について事例報告がなされた。その後、報告を行った4名をパネリストとして、フロアの参加者との間で活発な議論が行われた。以下では、主に議論となった論点をまとめ、アーカイブズ学の視点から、若干のコメントを加えたい。

パネルディスカッションでは、まず、データをオープン化していく際のインセンティブについて、質問がなされた。これに対して、中村氏は部局等にはオープン化によって問い合わせが減る、といった実務面でのメリットを紹介しているが、一研究者に対しては何らかの業績となるようなアプローチが必要であると回答した。また、南山氏は、トップダウンの方策として成立させるための理念の共有と、草の根的な研究者への働きかけの二つのアプローチをとることが重要であると指摘した。

次に、大学評価との関係でオープンデータやオープンサイエンスの研究／教育への取り組みを数値化していくことはできないか、という質問がなされた。これに対し、中村氏からはオープンデータを使った活用事例を集め、それを数値化して提示する可能性を示唆した。南山氏からは、研究データに関してはデータの産業的価値も考慮する必要があり、単純に数値目標に落とし込むことは難しいこと、畑埜氏も現状ではオープンデータ化が進んでいる分野とそうでない分野があるため難しいとの見解であった。他方、岡崎氏は、オープンデータやオープンサイエンスはただ公開すればよいというのではなく、その資源が集合知として創造的な利活用ができるというところで本来評価されるべき、との意見を述べた。

続けて、デジタル・ヒューマニティーズやシチズン・サイエンスとの関連で、大学の学部教育への取り組みはあるかという質問が出され、この点に関して畑埜氏は、くずし字のデータセットとディープラーニングを組み合わせたデータサイエンス教育の有効性を提示

された。他方、岡崎氏はデジタル・ヒューマニティーズやオープンサイエンスと教育の問題は人材養成やキャリア形成の問題とセットで考える必要があると指摘された。これに対し、中村氏はデータ管理に関しては専門職が担うべきであるが、デジタル・ヒューマニティーズについては、現在の研究のトレンドとして研究者個人が身に付けるべき研究手法の一つなのではないか、と述べられた。

パネリストの大方の見解として、デジタル・ヒューマニティーズやオープンサイエンスは何か特別なものではなく、将来はそれが当たり前となってくることが示唆されたが、その一方で、情報管理の人材養成は必要か、また、そうした人材が必要であれば情報管理専門職は研究者のデータ管理にどこまで関与すべきか、という点が議論となった。この点に関しては情報管理を専門的に行う人材は必要であるという意見が大半であったが、南山氏は、データ管理の分野は非常に動きが速くその時点で最も有効なものを提案していくことが専門家の責任であるといえるが、長期的に次世代でどう評価するかという観点も見ていく必要があると指摘された。また、畑埜氏は、研究データの公開や管理が義務になった時にそれをサポートしてくれる専門職の存在は不可欠であるとし、専門分野ごとにメタデータなどの取り決めが違ったとしても他分野からの利用という点でメタデータやキュレーションに関する一般的な知識を持った人材の意見も取り入れていく必要があるのではないかと述べられた。

このあと、研究者が自身の研究データをオープンにする際、どういったサポートないしサービスが必要か、という質問が出された。これについて南山氏は、データ公開を支援するレファレンス業務の必要性や、より踏み込んだ研究関連データの収集、分析の補助をサービスとして行うことも考えていると回答した。中村氏はオープンライセンス等の知識を得るための機会の提供をあげ、畑埜氏、岡崎氏は、データ公開のための簡単なマニュアルを作って、何か問題があれば担当機関がすべて責任を持つ、というサービスのあり方が望ましいとした。

以上のように、パネルディスカッションでは、データのオープン化を促すためのインセンティブの問題、オープンデータやオープンサイエンスの評価や教育をめぐる問題、データ管理と人材養成／キャリア形成、データ公開のための支援の問題などが議論された。ここでは、シンポジウムを踏まえて、改めて、大学はオープンデータ化に対していかなる役割を果たしうるのか、そこでの専門職の役割とは何かについて、アーカイブズ学の近年の議論に拠りながら、考えてみたい。

まず、研究データのオープン化とは何かについて、簡単に整理しておきたい。今回のシンポジウムでは、研究資源のオープンデータ化、研究データの管理／オープン化、くずし字データのオープン化など、デジタル・ヒューマニティーズ、オープンサイエンス、シチズン・サイエンス等とも関連する多様な話題が扱われ、議論が多少錯綜した面があったように思われる。この点に関して、国立情報学研究所の北本朝展氏は、個々の活動は「よりオープンに」する方向性を共有しつつ、「オープンサイエンス」として収斂しつつあるとし、

その上で、研究データを3種類に区分している。すなわち、①研究の入力となるデータで、再利用のためのオープン化が求められる「研究資源データ」、②研究の出力となるデータで、透明性のためのオープン化が求められる「論文付属データ」、③研究の入力と出力の間に生み出されるデータで、日々の研究活動のエビデンスとなり、研究不正防止等のために長期保存が求められる「研究過程データ」、である [北本、2017]。今回のシンポジウムでは、オープンデータという点では共通しつつも、中村氏と畑埜氏は主に「研究資源データ」のオープン化、南山氏は「研究過程データ」のオープン化について取り上げていたといえる。ここで注目したいのは、デジタル・キュレーションの概念である。デジタル・キュレーションという、データの解釈や利活用に焦点が当てられる傾向にあるが、イギリスのデジタル・キュレーション・センター (DCC) では、デジタル・キュレーションとは、「その作成時点から、将来のための発見や再利用のための利用可能性を確保するために、データの使用を管理する活動」と定義され、データ・キュレーションの管理プロセスには、「データの保管、セキュリティ、品質管理、データに関する重要情報の記録、データに対する方法と変化の分析、将来におけるアクセスと再利用を可能とするためのデータの保存」が含まれる¹⁾。ここで重要なのは、データの利活用のためにキュレーションの活動が「作成時点から」行われるものとして捉えられている点で、アーカイブズ学では、研究者によるデータの作成から公共空間における利活用までを包括的に捉える「データ・キュレーション・コンティニューム」という概念が提示されている。データ・キュレーション・コンティニュームは、オーストラリアのフランク・アップワードによって提示された記録管理のためのレコード・コンティニューム・モデルから派生したもので [Upward, 2000]、データ・キュレーション・コンティニューム・モデルでは、ある調査者ないし調査チームによって調査データが捕捉され、キュレーションされるプライベート・ドメイン、コア・チームの外部ないし組織外の調査者との間で調査データが共有されるコラボレーション・ドメイン、データが公共空間でアクセス可能となるパブリケーション・ドメインの3つの領域が設定されている。そして、プライベート・ドメインからコラボレーション・ドメインへ、コラボレーション・ドメインからパブリック・ドメインへの移行措置 (マイグレーション) の重要性が指摘されている [ANDS Guide, 2016]。

モナッシュ大学では、このキュレーション・コンティニューム概念を適用して、大学としての情報管理戦略の実施やポリシー、ガイドライン、ハウツー・ガイドの作成などが行われているが、特に興味深いのは、データ管理を支援するためのリポジトリの役割である。モナッシュ大学では当初、長期保存に適したオープンアクセスを提供する研究データのリポジトリ²⁾を構築していたが、この一元的なアプローチでは、キュレーション・プロセス全体にわたって研究者によって生み出されるデータを適切に管理することに失敗し、また、研究者からのデータ提供も芳しくなかったという。そこで、パブリケーション・ドメインのリポジトリに加えて、コラボレーション・ドメインにおけるリポジトリを構築するプロジェクトが立ち上げられた³⁾。このコラボレーション・リポジトリによって、アクセス可

能な範囲を制限したうえで調査チーム同士の研究データの共有、分析やアノテーションを容易にし、データ管理のためのベストプラクティスに沿った形での調査研究が行われることとなった。このような2つのタイプのリポジトリを設定することで、データのオープン化に対する研究者の抵抗感をなくし、コラボレーション・ドメインからパブリック・ドメインへのデータ移行が促進されていったという [Treloar, et. al., 2007; Treloar et. al., 2008]。この事例のように、パブリケーション・ドメインとコラボレーション・ドメインの領域それぞれにリポジトリを構築することは資金面や技術面、また人的資源の問題で難しいところもあると思われるが、研究データのオープン化を促していく上で、公開の面だけでなく、オープン化を促進させるための、研究者個人の研究活動や共同研究を活性化させる研究データ管理支援のあり方も考える必要があるのではないだろうか。

また、データ管理のための専門職に関しても、データ・キュレーションの専門教育が盛んなノースカロライナ大学チャペル・ヒル校のアレックス・H. プールは、データ・アーキビストはキュレーション・ライフサイクルの初期段階から関与すべきである、と主張する [Poole, 2015]。そして、研究データのキュレーションには、出所、評価・選別、真正性、メタデータ、リスク管理、信頼といったアーカイブズ的実践が肝要な役割を果たすとし、アーキビストは、「データの信頼性と真正性の確保、データの証拠的価値の決定、データのコンテキストの記述、ステークホルダーの役割と責任の実体化のために [データ] 作成者と協同すべきである」、という [Poole, p. 113]。特に、今回のシンポジウムとの関連では、研究データの評価・選別とメタデータ付与が重要である。情報学研究者のクリスティン・ボグマンも指摘するように、「データ・キュレーション問題の核心は、どんなデータが保存に値するか、なぜ、誰が、誰に対して、どのくらいの期間保存するかという疑問である」 [ボグマン、2017年、15頁、324-345頁]。この点で、データの評価選別に関わるアーカイブズ学の知見は有用であろう。プールはデータを評価する際に鍵となるのは、将来の利用可能性とともに、データ (materials) の保存コスト、その型式、状態、量、アクセス可能性、唯一性であるという。そして、データの評価選別プロセスにおいては、「①データを保存する義務はあるか? ②作成分野において、あるいは、他の学問分野において、使用しないし再利用のためにデータを保持しておく理由はあるか? ③データは検証や文化的、財政的目的のために保持しておくべきか?」の三つの問いが考慮されることが重要である、と指摘する [ibid, p. 115]。また、メタデータについても、それぞれの学問の専門領域でメタデータ・スキームや標準が発展しているが、研究者たちは相互運用性やアクセス、長期保存のためのメタデータはほとんど考慮していない、とプールはいう [ibid, pp. 116-7]。相互運用性を確保し、データを長期にわたってオープン化していくためのメタデータ付与を行っていくうえで、情報管理専門職の関与は不可欠であるといえよう。

シンポジウムでも指摘されたように、研究者の流動性が極めて高い大学という場でオープンデータを制度化していくことには、ステークホルダー間でのその必要性も含めたより深い議論と合意形成が必要である。そして、何のためのオープン化かを明確にしたうえで、

情報管理専門職のデータの公開、利活用のためのキュレーション・ライフサイクルのより初期段階への関与や、研究者のインセンティブを高めるための、研究データの管理や分析、保存を容易にしうる、包括的な支援の充実が求められている。

註

¹) Digital Curation Centre, Curation lifecycle model, available at <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

²) Australian Research Repositories Online to the World(ARROW) project, available at <http://arrow.edu.au/>

³) Dataset Acquisition, Accessibility and Annotation e-Research Technologies(DART) project, available at <http://dart.edu.au/>

Australian ResearCH Enabling enviRonment(ARCHER) project, available at <http://archer.edu.au/>

参考文献

ANDS Guide, “Curation Continuum”, 2016, available at <http://www.ands.org.au/guides/curation-nontinuum>

Poole, Alex H., “How has your science data grown? Digital curation and the human factor: a critical literature review”, *Archival Science*, 15, 2015, pp. 101-139.

Treloar, Andrew, David Groenewegen and Cathrine Harboe-Ree, “The Data Curation Continuum: Managing Data Objects in Institutional Repositories”, *D-Lib Magazine*, 13(9/10), 2007, available at

<http://www.dlib.org/dlib/september07/treloar/09treloar.html>

Treloar, Andrew, and Cathrine Harboe-Ree, “Data management and the curation continuum: how the Monash experience is informing repository relationships”, *Proceedings, VALA Conference 2008*, available at

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.572.4602&rep=rep1&type=pdf>

Upward, Frank, *Modeling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond – a personal reflection*, 2000, available at https://www.ltu.se/cms_fs/1.83844!/file/Modellning_the_continuum.pdf

北本朝展「データキュレーションへの期待と課題：自然科学から人文科学まで」、九州大学ライブラリーサイエンス専攻シンポジウム「オープンデータとデジタル・ヒューマニティーズ」、2017年1月18日

https://catalog.lib.kyushu-u.ac.jp/opac_download_md/1794498/kyushu17-ppt.pdf

クリスティン・L. ボーグマン (佐藤義則/小山憲司訳) 『ビッグデータ・リトルデータ・ノーデーター研究データと知識インフラ』 勁草書房、2017年。