インタビュー

研究データの共有と活用の基盤整備をめざして

DataCite Executive Officerに聞く

Aiming at building infrastructure of research data sharing and reuse

An interview with Executive Officer of DataCite

Brase, Jan¹

インタビュー：恒松直幸²

.....

¹ DataCite

² 独立行政法人科学技術振興機構 情報企画部（〒102-8666 東京都千代田区四番町5-3）

¹ DataCite

² Department of Information Planning, Japan Science and Technology Agency (5-3 Yonbancho Chiyoda-ku, Tokyo 102-8666)

情報管理 57(12), 871-881, doi: 10.1241/johokanri.57.871 (<http://dx.doi.org/10.1241/johokanri.57.871>)

1. DataCiteの会員，歴史，組織，活動状況

——DataCiteが属するドイツ国立科学技術図書館 (Technische Informationsbibliothek: TIB)^{注1} の役割や重要性についてご説明いただけますか。

Brase氏：1957年のソ連での人類初の人工衛星スプートニク打ち上げ成功を知り，西洋諸国は「ソ連の科学技術に追いつき追い越すにはどうしたらよいのだろう」と考えたのです。ドイツでは科学技術に特化した図書館を作ろうという動きが起こり，専門の国立図書館を設立することになりました。その分野でドイツ最大の蔵書数を誇っていたハノーバー大学の技術図書館が1959年に国立図書館となりました。第二次大戦時にドイツの大半の図書館が空爆で破壊さ

れましたが，同大学では図書館職員がすべての蔵書を馬や荷車に載せて坑道に隠したため，無傷で残っていたのです。

ドイツの他の国立図書館は，ドイツ語またはドイツ人による出版物のすべてが収集対象です。一方，TIBは言語や国に関係なく世界中の科学技術関係の出版物を収集し，ドイツ国内だけでなく国外の人々にも提供するという国の設置法に基づいています。TIBは，科学技術図書館としては世界最大です。

——博士はどのような経緯でこの分野にお入りになったのですか。

Brase氏：ハノーバー大学で純粋数学を専攻し，副専攻はコンピューターサイエンスでした。卒業時に同

本稿は，2014年10月17日に行ったインタビューを弊社編集事務局にて編集・再構成したものである。

Dr.Braseの肩書はインタビュー当時のもの。同氏は2015年1月からドイツ国立科学技術図書館 (TIB) EU リサーチリエゾンオフィサー。

大学からコンピューターサイエンスの博士号を取らないかという誘いがあり、そこでメタデータや情報検索に取り組み図書館分野に足を踏み入れました。図書館と協働でプロジェクトを行うことになり、私たちが研究してきたことが突然実際のユースケースとなったのです。2年間検討してきたメタデータについて図書館の人たちに話したところ、彼らは目録データ (catalog data) と呼んでいましたが、私たちと同じ問題を150年間も抱えていたのです。これは面白いと思い図書館分野にかかわるようになったのです。

——世界的なネットワークのためにどのような組織がDOIとDataCiteをサポートしているのでしょうか。またDataCite設立と発展の経緯もお話してください。

Brase氏：DataCiteはデジタルオブジェクト識別子 (Digital Object Identifier: DOI) の登録機関の1つとして出版業界も含め各界から多くのサポートを受けています。現在会員は31機関です。ジャパンリンクセンター (JaLC) 注2)、大英図書館、TIBなど22機関が正会員で、DataCiteのローカルノードとして機能し、DataCiteのサービスを自国で提供しています。この正会員が世界中で350のデータセンターと協働しています。

残りの9機関は準会員です。マイクロソフトリサーチ、Digital Curation Centre、ハーバード大学などの組織、さらにDataCiteと共通する利害はあってもDOI登録は積極的に行っていないものもあり、IEEE (Institute of Electrical and Electronics Engineers) などが含まれます。さらにre3data、DatabibのようにDataCiteを支援してくれる組織とは、関連分野での協働のための覚書を交わしています。つまり会員は、科学者のデータ公開を手助けし、データセットを引用可能にするという明確なミッションを共有する組織の集まりで、世界に広がっています。

歴史をちょっとお話する¹⁾と、2003年にドイツ研究振興協会 (Deutsche Forschungsgemeinschaft: DFG) 注3) の資金援助によりデータ公開や引用の可能性を探るプロジェクト (Citability of primary data) が

Jan Brase氏 略歴

ドイツ国立科学技術図書館 (TIB) EU リサーチリエゾンオフィサー (Research Liaison Officer)

1999年ハノーバー大学数学専攻学士、2005年同大学コンピューターサイエンス専攻博士取得。専門分野はメタデータ、オントロジー、デジタルライブラリー。1999～2003年ハノーバー大学情報システム部門、2004年同大学L3Sセンターを経て2005年1月～2006年8月TIBでデジタルライブラリーの研究に携わる。2006年9月～2009年12月TIBのDOI登録機関としての活動を統括する。

2010年1月～2014年12月DataCiteのExecutive OfficerおよびInternational DOI Foundation (IDF) の委員長 (chair)。2015年1月から現職。また現在、International Council for Scientific and Technical Information (ICSTI) の会長 (president) ならびにCODATA Task group on Data Citationの副委員長 (Co-Chair)。2011年にドイツの「図書館ハイテク賞」を受賞。

発足し、データに識別子を登録する図書館が必要だという結論に至りました。TIBがこのプロジェクトに加わることになり、私もちょうど博士課程を終えた

図1 DataCiteをとり巻く構造

ところで、参加することになりました。

2004年にTIBでデータへのDOI登録を開始し、その後ドイツ国内のデータセンターのデータへのDOI登録が順調に進んでいくと、海外のデータセンターが接触してきました。スイス、イギリス、そして米国のデータセンターまでもが、データへのDOI登録に関心をもっていたのです。ところが、海外のデータセンターの研究助成機関は、自国だけでも提供できるかもしれないサービスのためにドイツの機関に資金を出すことに難色を示しました。科学をグローバルによりよくできるアイデアが見つかったというのに、国境という問題があったのです。データ引用を実現したければ、国ごとのレベルではなくグローバルに活動しなければならないことがわかりました。

そこでスイスやイギリスの図書館と会合を開き、協働でプロジェクトを行えないか話し合いました。科学はグローバルなもので科学者はグローバルに研究を行っていますが、自国のネットワークの一部でもあります。科学者は、自国の大学や研究所の一員として国の助成を受けているため、自国を代表する窓口が必要だと考えていました。そこで2008～2009年にコンソーシアム設立の動きが広がりました。2008年にはヨーロッパの図書館6館が私たちとの協業に関心を寄せ、2009年には世界中の機関と会合をもちました。同12月、ロンドンにグローバルなコンソーシアムであるDataCiteを設立したのです。

DataCiteのネットワーク体制は階層構造をもつという点ではインターネットとよく似ていますが、詳細はもっと複雑です。たとえば、ドイツや米国のようにDataCiteの会員が複数の国もあります。歴史的または構造的な理由からです。ドイツは連邦国家なので、1会員に絞るのは難しいのです。TIBのミッションの1つはデータにDOIを登録することですが、対象は科学技術関係のデータに限られています。そのため、ドイツ国内の他の会員であるドイツ国立医学中央図書館(ZB Med)、ドイツ経済学中央図書館(ZBW)、ドイツ社会科学センター(GESIS)が各専門分野のデー

タにDOI登録を行っています。米国も全米をカバーする1つの中央図書館や機関がないため、3機関が会員となっています。もちろん、まだ会員のいない国もあります。

——グローバルな組織が話題になる場合、グローバルな大義に自国の資金を提供する必要があるのかが問題となります。なぜ税金を使うのか、その支出からどのような利益が得られるのかというものです。

Brase氏：DataCiteのよいところは、DataCite自体が非常に小さな組織だということです。現在スタッフはたった2名で、比較的少ない費用で運営しています。

費用の内訳は、国際DOI財団(the International DOI Foundation: IDF)へ支払うライセンス費用、私自身と他の職員の給与、それに旅費です。各国の会員が作業の大部分を行うので、DataCiteの予算はわずかで済みます。また会員が支払う費用の大部分は自国の人件費とインフラに使われるので、その国の利益となります。会費と引き換えにライセンスを得て、自国のシステムとグローバルなシステムの一部になれるのです。

——現在DataCiteにはどれくらいのデータセットが登録されていますか。

Brase氏：370万件のDOIがあると思います。データセットの定義にもよりますが、DataCiteではデータセットを非常に広く定義しています。大部分は、誰もがデータと理解できる数字が並んだ表のデータです。さらに従来の学術出版以外の科学情報、すなわち研究や研究者に影響や刺激を与えるものすべてを科学情報と考えています。

この中には、測定データ、写真、ビデオ、プレゼン資料などが含まれます。1つの定義として、研究の動機づけとなるものはすべて研究データとなるということです。DataCiteには、DOIを登録し目録に収録しているオブジェクトが370万件あるといえます。2014年末までにオブジェクトを400万件まで増やした

いと考えています^{注4)}。

2004年に最初のDOIを登録し、その5年後、まだTIBしか始動していなかったのですが、70～80万件に増えました。DataCite設立の2009年には、100万件から200万件に増えました。そして2年前、250万件まで増加しました。2013年9月には300万件だったのが、この1年で100万件以上増加し、急激に伸びています。

——では、DOIの解決数はどれくらいなのでしょう。つまり、どれくらいの人々がデータへのアクセスのためにDOIを利用しているのでしょうか。

Brase氏：おおまかな統計データによれば、解決数は通常月100万件です。2013年には、1年で1,000万件でした。解決数は増え続け、2014年夏にはたしか月100万件をちょっと超えるくらいでした。

——ところで、DataCiteの名前が記載された論文がどのくらい発行されているか数えたことはありますか。

Brase氏：常に参考文献リストにアクセスできるとは限らないので簡単ではありません。そこで他の機関との提携を始めたのです。データ引用索引にDataCiteのデータセットを多数使用しているトムソン・ロイター社のData Citation Index^{注5)}と契約したばかりです。同社が参照数をカウントし、DataCiteのデータセットが実際に引用されると、同社がチェックしてくれます。またCrossRefとも協働のための契約に至り、

インフラの構築を始めました。新しいDOIがCrossRefに登録されると、CrossRefは参考文献リストを確認し、DataCiteのDOIがないかどうか調べます。同様に、DataCiteはDOIの登録や更新があると、CrossRef用のメタデータを確認するといった具合に情報交換を行います。

そのためのインフラを2015年の上半期に確立したいと考えています。DataCiteのメタデータが閲覧されると、他のオブジェクトに関係するDataCiteのDOI数を調べます。現在、DataCiteの約50万件のDOIが他のオブジェクトに関係しています。関係とは引用を意味しています。補足的な関係は引用の一部となります。このことからDataCiteの数十万件のデータセットが出版に結びついていることがわかりますが、これはDataCiteが把握している情報の一部にすぎず、実際はこれよりはるかに大きな数値かもしれません。

2. データ公開の重要性と技術

——データ引用の必要性をどのようにお考えですか。

Brase氏：データを引用可能にするという考えはかなり古くからありました。科学技術分野の論文引用に関しては300年もの歴史をもつ確立したシステムがあります。科学者は論文執筆以外の仕事も行うにもかかわらず、それを取り扱えないこのシステムには欠陥もあります。データ収集は科学者にとって重要な

図2 DataCiteと連携する研究データ公開機関のグローバルな構造

科学研究の一環ですが、論文のように認められず、注目されないのは残念なことです。

国際極年から国際地球観測年への拡張（1957～1958年）を機にWorld Data Centerが設立されましたが、当時はデータ公開をサポートする技術がありませんでした。データ公開の考えは50年前からありましたが、この15～20年間でインターネットや新しいテクノロジーが開発され、データ公開をサポートする技術が突然手に入ったのです。

——データ公開の技術についてお聞かせください。

Brase氏：2つの技術があります。1つ目はHTMLの技術です。データセットを引用可能にするには、データセットを保管できるシステムを用意し、利用者に公開し提供しなければなりません。これは通常インターネットのHTMLページで行われます。HTMLを利用してWorld Wide Web（WWW）からデータへアクセスできる機能は重要な技術です。2つ目はDOIです。DOIは、データへのリンクが安定していて、「404 not found」のエラーが起こらないよう保証する永続的な識別子です。DOIシステムは1998年に創設され、2004年にTIBはデータへのDOI登録を始めました。

——デジタル技術は急速に発展し、デジタル方式で記録したものは長期保存できないという理由から、識別子について多くの人が懐疑的になっています。DOIをより長く利用するために、どのような仕組みになっているのでしょうか。

Brase氏：DOIシステムは識別子という技術だけでなく、技術にもとづく信頼性のある関係機関が集まり運用する基盤です。たとえば、DataCiteでDOIを登録するとDataCiteの会員組織がDOIに付随しているので、リンクの安定が保証されます。会員同士、会員とデータセンターとの契約により、協働してコンテンツを永続性のあるものにします。たとえば、現在DVDに保存されているコンテンツは、将来技術が変わった場合でも他のプラットフォームへのコンテン

ツの移行が保証されます。

DOIは目で見える部分は小さな文字列ですが、その背後には多くの組織が参加している大きなネットワークがあります。DOIの魅力は、技術だけでなくこの組織のネットワークが信頼や契約も提供することです。DataCiteはこの協働イニシアチブの枠組みを提供しています。

原則的にはDOIはHTTPと、インターネットに依存したものに深く埋め込まれています。DOIはインターネットベースの技術なので、現時点ではインターネットなしではまったく意味がありませんが、インターネットベースである必要はないのです。仮に数十年後にインターネットがなくなり、想像もつかないような技術が開発されたとしても、DOIがそれに適応できる可能性は十分にあります。文字列をHTTPと切り離し、別の何かに取り付けて、新たな技術を利用することも可能です。

3. データ引用をめぐるステークホルダー

——研究分野によってはデータ引用の受け入れに時間がかかるなどということはあるですか。

Brase氏：それは大きな問題の1つです。すべての研究分野に共通の類似点はなく、研究分野ごとにデータの定義やその取り扱い方について独自の方法があります。たとえば、地球科学では幸運なことに、データに関して非常に優れた定義があり、データの取り扱い方も素晴らしいのです。地球科学者は、今日の天気は明日計測できないので、今日、データを作成しておかなければなりません。将来的に不要かもしれませんが、データを収集・保存しておきます。10年後にこのデータをもとにしてその日の天気を予測するかもしれません。つまりこの分野にはデータを収集し、保管し、利用するという伝統があり、この基本こそ、データ公開や引用の優れた事例なのです。

化学の場合、古典的な実験論文のデータを公開しないかと科学者にもちかけても、誰も興味を示しませ

んでした。なぜなら実験の手法は論文に書いてあるので、データは実験から得られると考えるからです。別の科学者は特許の問題や巨額の価値がある新しい実験だから、という理由でデータを共有しようとしないうのです。また医学研究者は、患者に対する守秘義務が守れないという倫理的問題を抱えています。

分野ごとにそれぞれ問題があり、すべての問題を解決できる共通の方法はなく、各分野に独自のワークフローが必要です。エンバーゴ（公開禁止）期間の設定も必要です。たとえば、天文学者にとって一番の恐怖は、自分たちの天体の観測値を公開すると、一般市民が新しい惑星を発見して名前を付けてしまうかもしれないことです。天文学者は、「データを収集したのは私なのに、データを分析する時間がなかったために、別の人が惑星に名前を付けてしまった」と考えます。このような理由から天文学分野では半年間のデータのエンバーゴ期間が設けられています。

データ引用の利点は、誰もが理解できるモデルだということです。科学者は論文には必ず引用を記載するので、論文が他の科学者に引用され評価が得られることを知っています。科学者に「データはどうしますか。あなたのもう1つの成果ですよ」と伝えることが大事です。科学者なら誰でも、データも公開すれば、データ引用により評価が得られるということを理解できます。このわかりやすさによりデータ引用があらゆる分野で世界的に機能しているのです。

——データセンター、図書館、DOIは出版社に広く利用されているというお話でしたし、もちろん研究コミュニティもそうです。今後のデータ引用のステークホルダー（利害関係者）としては、どのような人たちが考えられますか。

Brase氏：忘れがちなのですが、学術雑誌の編集委員会が興味深いステークホルダーの1つになるでしょう。編集委員会は科学者にデータの公開やデータを引用した論文の執筆を促すべきです。編集委員会は、データ引用を利用した論文の提出を科学者に要求するという重要な役割を担っています。また、科学関連団体も同様の活動を行い、会員に優れた手本を示すことができるでしょう。

研究助成機関も重要な利害関係者です。2003年からのプロジェクトも、研究助成機関であるドイツ研究振興協会の利益となりました。同協会は、極地調査のために北極探査船ポーラスターのような船舶に何百万ユーロという資金を投入していました。北極海に出航し多くの観測を行い、船の建造費を含め航行中は1日約100万ユーロかかる計算になるでしょう。何らかの理由でデータが利用できなくなったり紛失したりしたら、資金がすべて無駄になってしまいます。プロジェクト報告書という形でプロジェクトの結果だけは見られたとしても、実際の結果であるデータを失ってしまったら多額の投資が無駄になってしまいます。

研究助成機関からデータとプロジェクトの結果すべてを公開するようと言われたときに、DataCiteはそのためのインフラを提供します。結果を公開する場はありますし、どこに公開すればよいのか提案し、DOIを登録します。これでデータは引用可能となり一挙両得です。研究助成機関は「プロジェクトに100万ユーロ出資したのだけれど、これがその結果だ。これらのデータセットがプロジェクトの成果なのだ」と言えるのです。一方、研究者は「プロジェクト終了時に時間がなくて論文は書けなかったが、こんな実験をした。実験データを公開して引用可能にすれ

ば、データ引用によって認められるようになるのだ」と言えるのです。

話は変わりますが、研究者の間に労働分業が生まれています。たとえば、先ほどのポーラスターの例では、1日中データを収集している科学者は詳細を分析する時間も論文を書く時間もないわけです。これまで論文を書いて名声を得るのは必ずしもデータを収集した科学者ではなかったのですが、研究データの引用によって、このような科学者が突如名声を得ることができるようになるのです。

4. メタデータに対するDataCiteの取り組み

——メタデータに対するDataCiteの取り組みについてもお話しください。私は社会学者ですが、1980年代、社会学者間のデータ共有など、当時はまだ方法が確立していなかったので、メタデータに何を記述すればよいのか、何を記述してはいけないのか、わからないことだらけでした。現在は状況が違っていると思いますが、いかがでしょうか。

Brase氏：さまざまな分野が存在し、データ公開のワークフローすべてに適用できる共通の解決方法はありません。メタデータもしかりです。2003年のプロジェクト開始当初、地球科学分野の3つのデータセンター（Alfred-Wegener-Institut für Polar-und Meeresforschung, Deutsches Klimarechenzentrum

[DKRZ], Deutsches GeoForschungsZentrum [GFZ]) は、同じ分野であるにもかかわらずメタデータについて合意に達することができませんでした。DataCiteのメタデータには引用に必要なデータが含まれているので、最終的には、階層化したアプローチを取り、情報を追加することができましたが。

従来の項目は、標題、著者名、出版者、日付、識別子（DOI）で、さらに推奨される項目は、抄録やコンテンツタイプ（テキスト、写真、番号付データセットなど）です。重要なメタデータ項目の1つに関係要素があります。関係要素はオブジェクト間の関係を記述するもので引用も含まれます。ほかの項目も含めると、全部で25項目になります。

メタデータ・スキーマの最新版3.1^{注6)}が2014年10月17日（インタビュー当日）にリリースされたばかりです。この中には基本的な核となるメタデータがうまくまとめられています。DataCiteに登録する各オブジェクトにはこれらの要素が必要で、すべてに関して概要を知ることができます。詳細を知りたい場合は、特に社会科学に関しては、研究分野のレベルまで掘り下げなければなりません。この分野にはDOIを利用した非常に優れたメタデータ構造があり、分野内でうまく機能している独自のメタデータ・スキーマがあります。もし社会科学分野に関するすべてのデータセットのメタデータを目録に収録すれば、一部のユーザーにのみ公開している追加のDOI項目があったとしても、DOIを利用してデータセンターのページにたどり着けば、完全なDOIの情報を入手できるというわけです。

問題は「メタデータを何に利用するのか」ですが、メタデータはユースケース次第で非常に役立つものになります。プロジェクト開始当初は、すべてのデータセットのメタデータを収録するデータベースを構築すれば、データを検索し比較することができると考えていました。しかし、実際はさまざまな異種のデータが含まれているため無理でした。その代わりDataCiteのメタデータ・スコアを利用して、基本的な書誌メタ

データ（著者名、標題、抄録など）の検索ができ、1つの事柄に関して概要を得ることができます。

データセットを科学的に比較したい場合は、データセンターレベルで検索しなければなりません。このためデータセットのDOIを解決する際に必ずデータセンターへと導かれるようになっています。データセンターは、科学技術分野の検索に対応できる完全かつ幅広いメタデータを提供します。たとえば、「30歳の日本人男性と30歳のドイツ人男性を比較したすべての研究を知りたい」などの検索は、データセンターレベルで解決できます。DataCiteではデータの範囲が広すぎて、集中して行うことはできないのです。

5. 図書館の変化、そしてライブラリアンの役割と情報教育

——現在は科学知識の中心的存在が図書館からデータセンターへと移行している転換期なのでしょうか。

Brase氏：おそらく完全には移行しないでしょう。先ほど述べたように、複雑な検索にはデータセンターを利用しなければなりませんが、ほとんどのデータセットは図書館の目録を通じて入手可能です。図書館には幅広くさまざまなコンテンツタイプがあるので、今なお重要な場所です。たとえば、「光合成について知りたい」といった基本的な情報要求に対して、出版物、論文、ビデオ、細胞中の光合成の測定を説

明したデータセットなどを提供できます。その中に興味のあるデータがあれば、目録から該当するデータをクリックし、データセンターへと導かれ、そのデータセンターで調査を続けることができます。しかしそれでも、図書館が情報への入り口だと考えます。なぜならば、図書館は先頭アドレスであり、幅広い種類の情報を提供するからです。もっと詳しく知りたいと興味が湧くものが見つかったら、さらに詳しい情報を所有するデータセンター、ピクチャーギャラリー、大学のリポジトリなどを図書館が教えてくれます。

——図書館は、データセンターが所有するデータのメタデータをサービスの一部として組み込まなければならぬのですね。

Brase氏：そうです。DataCiteは図書館とデータセンターが協働すべきと考え、積極的な活動を行っています。DataCiteは図書館の目録情報にマッピング可能な、メタデータに関する書誌情報を提供しています。コンテンツ自体は別の場所にあったとしても、永続性のある識別子によって図書館目録で一括して検索できるのです。

従来の図書館では目録は所蔵資料への窓口でしたが、技術の進歩によって、所蔵していないたくさんのコンテンツも案内できるようになりました。もはやデータを保存する必要はなく、データを記述するメタデータとデータセンターへの永続的なリンクが必要だけです。データセンターと信頼できる契約関係を構築しなければなりません。DataCiteも協働し、図書館にコンテンツのDOIを提供するので、図書館はDOIを解決しコンテンツを提供可能な状態にする必要があります。コンテンツを変更した場合はDataCiteに連絡し、DataCiteの目録にメタデータが収録されるようにします。あるユーザーからの検索に関連するコンテンツがデータセンターにあるとします。図書館がDOIを使ってそのデータセンターにリンクすると、ユーザーはデータセンターとして機能す

る図書館に誘導され、そこで検索ができます。しかしユーザーにとっては図書館が検索を開始した場所なのです。

——大学のデータセンターに保管しているかなりの量のデータを公開し、他の科学者に引用や利用をしてもらうには、信頼できる図書館にデータセンターをリンクしなければならないということでしょうか。

Brase氏：それはインフラの1つです。データを保管する物理的な場所は変わりません。理想は大学にデータを保管する永続的なリポジトリがあり、質のよいメタデータを所有していることが望まれます。リポジトリ自体が信頼できるデータセンターに相当するのです。必要なのは若干の仕組みの追加と契約関係だけです。さらに理想を言えば、大学と図書館が協働し合意に達することです。大学側がリポジトリでデータを保管し、引用を可能にし、一般に公開することです。

——多くの人が論文を検索するとき、最初にGoogle Scholarを使用しています。探しているデータをキーワード入力するだけで結果を表示するような大きな検索エンジンが出現すると思いますか。

Brase氏：可能性はあります。個人的な見解ですが、Google Scholarのようなものは素早くおおざっぱな検索には適していて、ある程度の概略は得られるでしょう。しかし特定の情報を深く掘り下げて探す場合は、図書館に軍配があがります。Google Scholarはあらゆるものに関して幅広い情報を提供していますが、図書館が提供する結果は精選した一握りの情報だからです。熟練したライブラリアンがメタデータを記述し、コンテンツを選択し、アドバイスをを行っているお陰です。

——研究データ同盟 (Research Data Alliance: RDA)^{注7)}の総会によく出席しますが、研究データの共有に積極的なライブラリアンたちの話題はトレーニングプログ

ラムです。その1つはデータサイエンティスト向けの基本コースで、データサイエンティストのトレーニング教材に図書館学の基本が必ず含まれるよう望んでいます。データ引用の新しい時代に対応するために、ライブラリアンや科学者への教育をどのように行えばよいのか、考えを聞かせてください。

Brase氏：図書館は転換期にありライブラリアンの役割は進化を続けているので、ドイツでも図書館教育がそのペースに追いついていくのは大変です。現在図書館で働く人の多くが私のような部外者です。私はコンピューターサイエンス出身ですし、異なる分野出身の人が図書館で働くことは珍しくありません。

図書館情報学はライブラリアン向けに、より高度なコースを用意できると思います。また、いまはどの学科でも多かれ少なかれデータを扱いますが、データを扱う学科のある大学の科学者や科学技術を専攻している学生は全員、何らかのデータ管理コースを取るべきです。大学はすべての科学者、あるいは少なくとも博士号をもつすべての科学者にデータ管理コースを提供して、データ保管の技術、データのさまざまな形式、データ保管の多様な方法、データ公開、メタデータなどについて教えるべきです。このような基本は、ライブラリアンだけでなく全科学者向けのカリキュラムに組み込まれるべきだと思います。

——ライブラリアンと研究者の関係が変わっていくのですね。ライブラリアンの役割についてどのようにお考えですか。

Brase氏：ライブラリアンは情報の入手方法や検索方法にたけているので、情報のゲートキーパーとして、学生にとって重要な存在です。ポケットに収まる小さな機器で世界中のあらゆる情報にアクセスできる時代となり、学生はあふれる情報の流れの中で途方に暮れています。情報とは何か、どのように情報を保管するのか、情報にどう対処するのか、情報にどのように注釈を付けるのかといった技術は、ライブラリアンが従来もっている資質なのです。図書館は

基本的な情報トレーニングを学生に提供し、もっと積極的に大学における役割を果たすべきです。ライブラリアンは国内外のネットワークの要であり情報の大海での水先案内人なのです。

ドイツの国立図書館は膨大な所蔵資料をもち、コンテンツの大部分はインターネットからアクセス可能になっています。そこで地域の小さな大学図書館がいまだに必要かという疑問が湧きますが、私は必要だと考えます。科学者はローカルなインフラの一部です。たとえば、日本の科学者がグローバルに活動しても、彼らは国の助成を受けた日本の一員であり、研究を行っている地域の大学の一員でもあるからです。

仲間の科学者と会うことができ、疑問について手助けしてくれるライブラリアンがいる地域の場所、図書館があれば、そこへ行って「昨年北米で会議があって、こんな話が出たと聞いたのですが、抄録を探すのを手伝ってくれませんか」と尋ねることができるのです。Google Scholarで探すよりよほど手取り早い方法です。

6. データ公開と出版社の動向

——出版社の動向はどうでしょうか。たとえば、出版社が論文の著作権を所有しているかぎり、論文の購読料を払わなければなりません。また出版社が収集を始めたデータについても懸念があります。大手出版社の論文にデータを掲載すれば、出版社のサーバーやリポジトリにデータを保管するよう勧められるでしょう。この場合、たとえば数値データには著作権は発生しませんが、ビデオや写真には原則として著作権が発生します。所属大学のデータセンター、図書館、機関リポジトリなどにデータを渡せば、著作権問題の観点からもデータによりアクセスしやすくなります。このことに関して見解をお聞かせください。

Brase氏：そのとおりです。それはある意味、当初からDataCite設立の目的の一部でもありました。科学技

術分野の伝統的な学術出版システムにはバックサイドがあることに誰も異論はないと思います。科学者の立場から納得できない点があるからです。自分の論文を提供して、その論文に対して購読料を支払わなければならない、同僚や著者や出版社と共有することもできないのです。

データ公開を始めた当初から、科学者の要望は、論文出版で起こったことをデータで繰り返さないように、データの公開を前提とし、出版社ではなく大学や学術コミュニティが主体となって進めようというものでした。これはとてもうまくいきました。最初からデータを完全公開とし、データセンターと協働してコンテンツを自由に入手できるようにし、データの引用やデータへのアクセスを可能にしたのです。

もちろん、倫理的な理由からデータを完全には公開していないデータセンターもありますが、ほとんどすべてのデータは公開されています。DataCiteは出版社とも協力しているので、たとえばエルゼビアの提供する論文の根拠データにもアクセスすることができます。ScienceDirectを契約していなければ、論文の抄録のみ閲覧可能で、論文自体にアクセスするには20ドル支払わなければなりませんが、データへのリンクが表示されており、データには無料でアクセスできます。つまり論文に対する権利はなくても、データを閲覧する権利はあるのです。このようにデータへのアクセスが非常にうまく機能していることを誇りに思っています。出版社さえもこのシステムに敬意を払っています。

2007年のブリュッセル宣言^{注8)}で、STM出版社は、根拠データへのアクセスは自由にできるようにすべきであると宣言しました。出版社の定義による「自由にできる」や「根拠データ」の意味においてですが。現在では出版社とDataCiteとの契約を通して、データに登録されたDOIを表示し、論文からデータをクリックでき、それが生データの場合、著作権や購読料も発生せず、無料となりました。DataCiteが存在したか

からこそ実現したのです。

もし10年前に出版社が10億ドルで1つのデータアーカイブを作成し、すべてのデータをそこに保管し、データのアクセス料に20ドルを課金することにしよう」と決めていたら、私たちがここに座っていることもないし、状況も違ったでしょう。実際にはDataCiteが創設されインフラが構築されました。DataCiteが多くの先例を作り、多数のデータセンターが稼働し、データが無料であることに科学者が慣れてしまったため、もはや出版社はこの状況を変えることはできません。

ただし商業出版社が有料の追加サービスを提供することは容認しています。たとえば、トムソン・ロイター社のData Citation Indexのほとんどのデータは無料で入手可能ですが、追加サービスは有料です。データを見やすく視覚化したり、あちこち閲覧できるモデルを作成したりといったサービスは有料となるわけです。データに加えて追加のサービスを提供するのであれば課金してもよいと思います。しかし、根拠データは無料であるべきで、科学者が望んでいることであり、DataCiteが確立したことなのです。

本文の注

注1) TIB <http://www.tib.uni-hannover.de/en/>

注2) JaLC https://japanlinkcenter.org/top/doc/141031_1_kobayashi.pdf

注3) Deutsche Forschungsgemeinschaft (DFG) 日本代表部 <http://www.dfg.de/jp/index.jsp>

注4) 2015年1月末現在、430万件に達した。

注5) サイテーション・インデックス（引用索引）について http://ip-science.thomsonreuters.jp/ssr/citation_indexing/

注6) DataCite, メタデータスキーマ バージョン3.1を公開 <http://johokanri.jp/stiupdates/education/2014/10/010428.html>

注7) Research Data Alliance <https://rd-alliance.org/>

注8) Brussels Declaration on STM Publishing http://www.stm-assoc.org/2007_11_01_Brussels_Declaration.pdf

参考文献

- 1) Brase, Jan; Sens, Irina; Lautenschlager, Michael. "The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five Year History of DataCite". D-Lib Magazine, 2015, vol. 21, no.1/2. doi: 10.1045/january2015-brase, (accessed 2015-1-20).