

第3回 SPARC Japan セミナー2017

「オープンサイエンスを超えて」

デジタル時代の研究プロセスと 大学、大学図書館における支援のあり方

倉田 敬子

(慶應義塾大学文学部)

講演要旨



多様な情報源の探索、研究データの収集、分析、成果の発表まで、研究者たちの研究プロセスはデジタル化されつつある。大学図書館は従来、研究に必要とされる資料を提供することでその研究活動を支えてきた。学術コミュニケーションにおけるデジタル化とオープンアクセスの進展は、大学図書館のこの役割を縮小させているといえる。大学は大学図書館の予算および研究者たちの経済的基盤を提供することにより、研究活動を支えてきたが、グローバルな競争とオープンサイエンスを志向する国の政策や社会状況の中でより直接的な関与がもたらされてきている。大学図書館、さらには大学というコミュニティは研究者の研究活動をいかに支援できるのか、研究プロセスのデジタル化という観点から考えてみたい。



倉田 敬子

1987年慶應義塾大学文学研究科博士課程修了。1988年慶應義塾大学文学部図書館・情報学科助手、1993年同助教授、2001年から文学部図書館・情報学専攻教授。専門は学術コミュニケーション、特に情報メディアのデジタル化、オープンアクセス、研究データに関心を持っている。著者に『学術情報流通とオープンアクセス』（勁草書房）など。2008年度日本図書館情報学会賞受賞、2010年科学技術への顕著な貢献2010（ナイスステップな研究者）選定。SPARC Japan 運営委員会委員。

オープンサイエンスとは何なのか

今回のセミナー全体としては、オープンサイエンスとは何か大きなテーマです。特に午前中は、科学そのものとは何なのか、データドリブンとはどういうことなのかということが話の中心でした。でも、私はやはりオープンサイエンスとは何かはよく分かりません。データが重要だということは、研究者なら誰だって分かっている話ですが、データドリブンだと言われてしまうと、今までの科学とは何が違うのかということになります。いわゆる第4の革命というような、データがまずあって研究が始まるというところまで本当にいくのかは、社会科学、人文科学の立場から言うと、ま

だよく分からないし、制度化という点ではまだまだ未知数のところがあると考えています。

オープンサイエンスのもう一つの側面は、デジタルであるということです。非常に広い文脈で社会全体がデジタル化している中で、研究も完全にデジタルの世界へ行くだろうといわれています。

もう一つ、個人での研究がないわけではもちろんありませんし、個々の研究自体はもちろん研究者個人がするのですが、プロジェクトとしては、共同・共有・協働が科学の世界においても非常に重要視されています。これまで個人が中心だったといわれていた人文科学も含めて、共同・共有・協働が研究の一つの特徴と

されるようになってきていると考えています。

このデジタルと共同（協働）という環境の中で、今、研究者たちはどのような研究をしているのか、こちらの方に今日は焦点を当てて事例を紹介いたします。このようなことを「デジタルスカラーシップ」と言う方もいると思います。ただ、今回はオープンサイエンスがテーマなので、これもオープンサイエンスの一つの側面と考えます。

デジタル時代の研究者

いろいろな焦点の当て方があると思うのですが、私は「研究ライフサイクル」という表現に関心を持っています。最近よく目に付くのでググってみたところ、図1のような図がたくさん出てきました。Pinterestでもたくさんピンされており、面白かったので幾つかここに示しています。後ろに隠れている、文章だけで表しているようなものもありますが、やはり写真映え、インスタ映えするような図にしている方が非常に多くて、図書館のサイトにもこのような図を使って図書館でのサービスを説明しており、すごいと思った次第です。

研究ライフサイクルの各要素を見ると、共通している要素がかなりあります。図2は、KramerとBosmanが2015年のカンファレンスで発表したものです。学術コミュニケーションの101の新機軸、イノベーションと言っています。内側の円に2005、真ん中の円に2010と書いてあります。一番外側が2015年という意

味なのだと思います。ポスターなので細かい発表は分からないのですが、これはデジタルに焦点を当てたもので、このような形で学術コミュニケーションを支えるさまざまなサービスが出てきているということです。

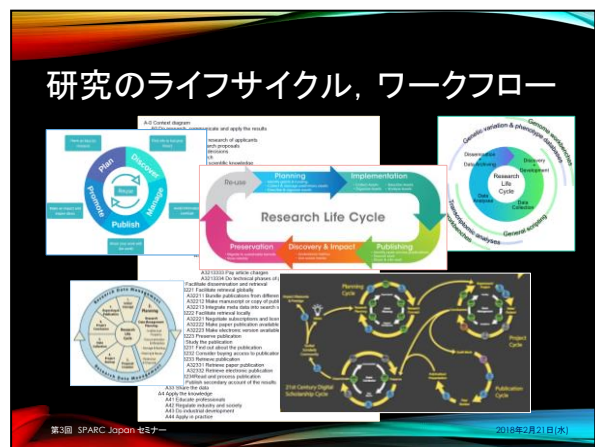
それぞれのワークフロー、研究のライフサイクルに応じてこれら新機軸のサービスを利用していることを示しています。最初の段階がDiscovery（発見）です。Web of Scienceのようなデータベースを使っていた人たちが、今やGoogleやGoogle Scholar、あるいはarXivのようなプレプリント・サーバで論文・文献を探すようになっていきます。

Analysis（分析）に関しては、SPSSなど個々のソフトウェアを使っていた人たちが、今やインターネット上で盛んに共有できるようになっているRやGitHubを使いはじめています。

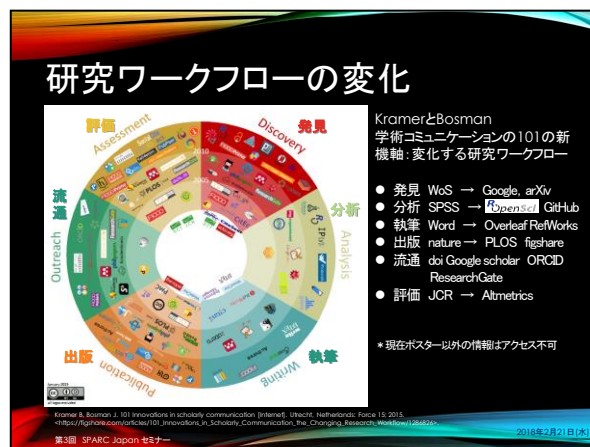
Writing（執筆）はWordになって長年たちますが、OverleafやRefWorksをみんな使いはじめているのではないかと思います。

Publication（出版）は申し上げるまでもなく、伝統的な「Nature」のような雑誌から、オープンアクセス・ジャーナル、さらにはfigshareで学会発表のものを共有するというものも出てきています。

Outreach（流通）は、もちろんDOI、ORCIDといったIDを付けることから、Google Scholarでも業績の一覧を作れますし、ResearchGateで論文を共有することもあります。このOutreach（流通）とDiscovery



(図1)



(図2)

(発見) のところはある意味では表裏だと思うのですが、いろいろな形で情報が手に入るようになっていきます。

最後は Assessment (評価) です。以前は JCR やインパクトファクターぐらいしかなかったものが、altmetrics などさまざまなものが出はじめています。このように、現在、研究のワークフローにいろいろな変化が出てきているということが言われています。

図 3 はさらに 2017 年に FORCE2017 という学会で発表されたもので、今、研究者は、それぞれの段階でさまざまな新機軸、サービスの中から好きなものを選ぶ段階に来ているということです。これは逆に、選ばなければいけなくなっているということで、何をどういう形で選んでどうやっていくのかを決めるのは、これはこれで結構大変です。

一人の研究者として

ここまでが全体としての概略で、では私はどうしているのかを簡単にお話しします。本当であれば、研究者は今このように変わっていますと言いたいのですが、これはまださまざまで、それについての調査などの成果もそれほど出ていないので、社会科学者としての私の一例をご紹介します。

大学の研究者の場合、複数の目的・役割を同時進行で進めていかなければいけません。もちろん教育にも非常に時間を割かれます。教育は自分の研究が土台になりますし、大学の研究者には研究者の育成という役

割もあるので、これをないがしろにはできません。でも、自分の研究と教育はやはり少し違います。

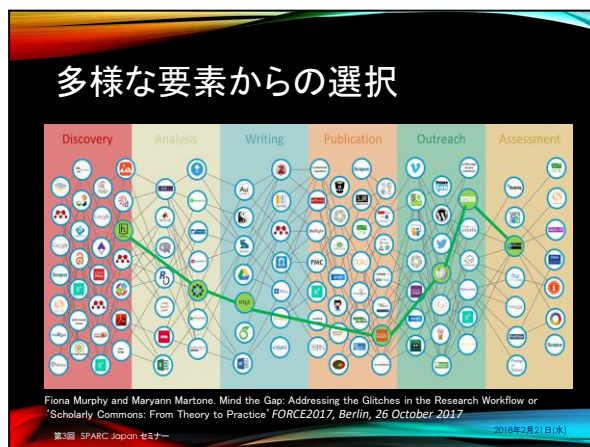
次に、大学や学会において何らかの役割を果たさなければならず、デジタル化の影響はここにも来ています。研究を応用できる側面ももちろんなくはないのですが、それとは全く違うことが求められます。

さらに、研究プロジェクトが一つという方もそうはいらっしゃらないと思います。やはり幾つも同時進行していて、目的もステップも異なります。例えば、このような講演を頼まれるということも入ってくるわけです。そうなったときに、どういう形で、自分で自分の行動をコントロールしていくのかということが非常に難しくなってきます。

発見の側面もデジタル化されて、情報探索の側面はほぼ完全にデジタル化されていると思いますが、どれをどう使うかは組み合わせていると思います。私もいまだに図書館の OPAC はよく使いますし、その一方でやはり Google Scholar は手離せませんし、ResearchGate からメールはよく来るし、何かあれば figshare をちょっと覗いてみる、それを当たり前のようにもうやっているわけです。

しかし、その中で、雑誌そのものを見に行くということもやはり今も行っています。重要な雑誌のモニタリングをやめるということはしません。そういう最新の情報に触れるというカレントアウェアネスと、研究プロジェクトのそれぞれの段階に合ったところで網羅的に探す、この二つが似ているようであり、やはり全く目的が違うので、文献、情報の探し方も違ってくるということになります。

さらに、これは当たり前なのですが、今、入手する段階において、デジタルで入手できるという点では非常に便利になりましたが、残念ながらまだ図書や日本語の論文に関しては紙でしか手に入らないものがあります。読むときには、私はまだ紙が優位で、デジタルも増加はしているのですが、紙で読む場合がまだまだあります。何かを記録しようとするときのアイデアのメモや文献の注釈はデジタルで随分できるようにはな



(図 3)

ってきていますが、やはり紙も使っています。読んで、情報を整理して、メモを作ってしまう段階を、私はもっとデジタルで統合化したいと常々思っているのですが、これは結構難しいのです。さらに、共同研究になると、ここが非常に複雑になるので、この辺でももう少し効率的なサービスができないのだろうかということは常々思っています。

次の段階で、調査を計画しようとした場合、もちろん先行研究を集めてきて参考にします。先行研究でのデータも最近では入手可能ですが、これまでも例えば社会調査の質問票などは情報として十分参考にしていました。でも、これからは個票データが入手可能になります。現にもう幾つかの社会調査ではデータまで公開しています。では、それで調査計画のやり方が変わるのかというのは、私にはまだよく分かりません。理論があって目的があって、自分なりの考えで調査を計画しています。そのときに人のものは参考にはなるけれど、そこでデータが手に入ったということで、本当に研究のやり方まで変えるのかというのはよく分かりません。

さらに大変なのは助成金を取ることで、今は、学内でも調査は倫理委員会にかけなくてはなりません。また、日本ではまだ大きくなってはいませんが、欧米でいわれているデータ・マネジメント・プラン（DMP）を今後は出していくことになると思います。煩雑な書類や手続きが必要で、これはできれば、研究者ではなく、もっと違う人たちに助けてほしいとは思っています。ただ、それもやはり研究の一部なわけで、どこまでをどのように研究者としてうまく効率化しながら、やらなければいけないことをきちんとやっていけるかというのはなかなか大変であるというところに来ていると思います。

私自身は調査の分析で、分析ツールまで共有することはまだそれほど多くはないのですが、やはり統計や、例えば AI とまでいかななくても、そのような最新の研究をされている方の中では、分析ツール自体を共有する、分析の方法自体を共有することが始まっています。

このような傾向をどのように自分の研究の中でうまく管理していくのかというのは、非常に大変になってきていると思います。

そんなに大きな調査ではなくても、少し調査をすればデータは山のように出てきて、解析はそのたびに、段階ごとに無数のファイルを作ってしまうので、これをきちんと管理するのはとても大変なことです。基本的に今、私は Dropbox や Wiki、Dropbox Paper などを使っていて、最後のデータは仕方がないので Zenodo や figshare で公開するという方法を取っています。

執筆・発表の段階になると、これはある意味ではこれまでの伝統的なやり方に従うのですが、基本的には共同研究などでは完全に共同執筆になります。部分ごとに分担して書いて、それを上げて、みんながもう一度全部読み直して、人が勝手に手を入れていってということは何回もやることで論文が出来上がっていきます。

最後は、やはり英文校閲にかけなければいけないので、ここは業者をお願いすることになります。今は引用の形も、場合によっては投稿まで業者が引き受けてくれます。これは研究者にとってはとても便利で、頼みたくなります。というのは、投稿はオンラインシステムなのですが、雑誌によってまちまちで、そのたびにやり方が少し違うため、すごく面倒なのです。それを例えば 1 万円でやってくれると聞くと、ついふらふらと頼みたくなるという気持ちは大変よく分かります。

Outreach（流通）のところは、一番変わっているのではないかと思います。今まで研究者自身が outreach を意識することはあまりなかったと思います。でも、例えば今、私は Google Scholar にプロフィールを登録していて、そうすると、自分で論文を刊行すると、1 週間ぐらいたつと、Google から「これはあなたの発表ですか」とメールが来ます。それにただ「Yes」とするだけでいいのです。「引用されましたよ」ということもすぐに知らせてくれます。

もちろん ResearchGate にも登録していて、そうすると ResearchGate で「この人があなたの論文を読みまし

たよ」ということを教えてください。今まで、そのようにリアルタイムで反応が分かるということはあまりなかったもので、そういう意味では、研究者も outreach を直に感じられるようになってきているのではないかと思います。また、ORCID 番号を取っているのですが、これは Scopus と連動してくれているので、ここでも自分の業績がある程度自動的に積み重なっていきます。

それに比べて、これは慶應義塾大学の例ですが、大学の研究成果データベースは非常に入力に面倒で、多くの研究者ができていないと思います。何とか最低限やっているのですが、ここ 1~2 年はそれすら怠りがちです。Google Scholar や ResearchGate のサービスに慣れてしまうと、今の大学のシステムはとても遅れていると感じざるを得ません。ただ、慶應のシステムは新しくなるそうなので期待しています。

大学、大学図書館による支援

では、このような研究者を大学や大学図書館はどう支援できるのかということが問題なのだと思います。前半で、村山先生が学会には学会のやるべきことがあるとおっしゃったのはまさにそのとおりで、学会こそ社会と研究をつないでくれるのです。では大学はとなると、ここはとても難しいところだと思います。学会はアカデミックなコミュニティとして確実にあります。でも、大学はインスティテューションでしかないので、同じインスティテューションに所属しているのですが、それぞれ違う学会に所属している研究者たちを取りまとめているわけです。その大学の中の図書館に何ができるのかということになると、これはなかなか難しいと思います。

日本ではまだなかなかできていませんが、欧米の大学図書館での研究支援の例は随分たくさん報告されています。ただ、全体として、最後の結論だけ言ってしまうと、その研究支援はインストラクションやコンサルタントなのです。それが悪いとはもちろん言いません。とても重要な点です。例えば、助成の条件がどう

なっているのか、データ・マネジメント・プランをどう作ればいいのか、そういうことに関して図書館員がアドバイスしてくれるというのは大変重要だと思います。ただ、これはインストラクションです。要するに、全体の動向をきちんと把握して、それを研究者に仲介する役割だと思います。これで止まっているのかということが私の疑問なのです。

そうではなくて、学術コミュニケーションシステムを改革するところに大学図書館の役割があるのではないかという話です。例えば、機関リポジトリはそのためになられて、今、それは一つの役割を果たしているとは思いますが、オープンアクセスの中でものすごく重要なステークホルダーになり得ているかという、残念ながらそこには至っていないと言わざるを得ないと思います。

例えばデータ・マネジメント・プランの作成というところでは一定の役割を果たしていますが、データそのものの管理を本当に図書館ができるのかということころは難しいですし、図書館でなくても、大学ができるのかということすら難しいのです。figshare や Zenodo と争って勝てるのかという話です。既存のデータアーカイブとどういう関係を作っていくのかということですね。競わなくてもいいのですが、役割分担は必要だと思います。そういう意味で、学術コミュニケーションの生態系（エコシステム）の中で、大学図書館は一体何ができるのかを考える必要があるのではないのでしょうか。

前半の David のお話からすると、実現は困難ということになりますが、全く違うシステムを構築するという運動があります。今、面白いと思っているのは、Scholarly Commons というムーブメントです（図 4）。これは、研究のさまざまな素材やツールなどを全てオープンにして共有するという原則に賛同、協力してくれる研究者、組織、出版社などだけのコミュニティを作れないかというものです。完全にオープンでなくても、ミクスチャーのレベルをできるだけオープンの方に近づけた形で、しかも研究成果だけでなく研究プロ

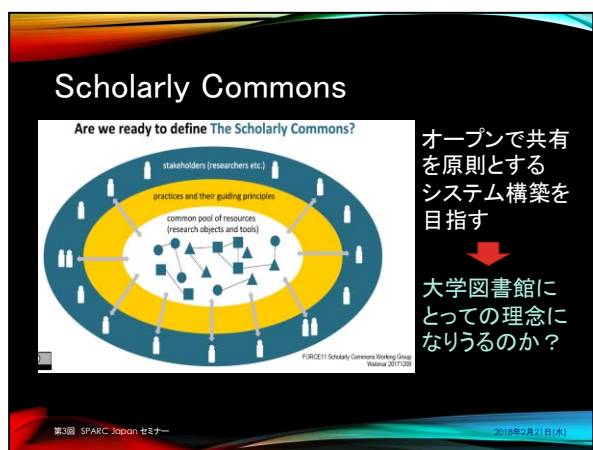
セスのところまでも含んだ形でのコモンズを作れないかという動きです。これはあくまでも一つの理念ですが、そういうところに一体、大学図書館は本当にどこまで関わり得るのかという話だと思います。

そこまでは無理なら、大学でせめてこれくらいできないかなというのが、研究情報管理です（図 5）。大学というインスティテューションの中で、さまざまな研究に関わる情報が生まれ、管理する必要が生じているわけですが、これらが全く統合化されていません。せめて、研究成果、研究者の情報、教育歴、広報、研究助成といった情報の統合的なシステムを作るということぐらいは、大学がやるべきではないかと思えますし、十分やれることではないかと考えています。

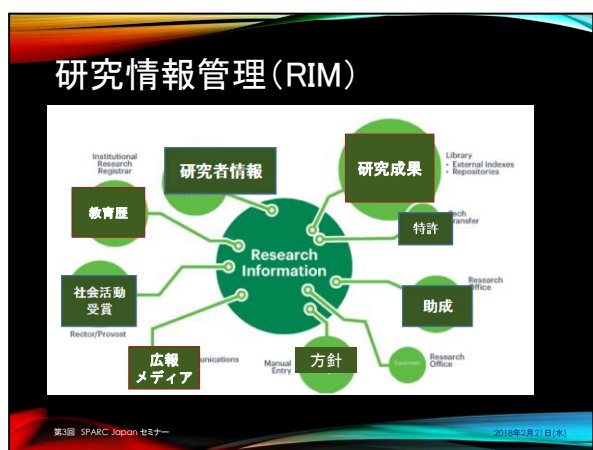
図 6 は、Horizon 2020 の一つとして行われている Common European Research Information Format というものです。これは図書館が今まで取り組んできた文献に

対するメタデータを、研究情報の標準化にも生かしていこうという考え方だと私は取っています。図書館が扱ってきたメタデータを、もう少し違う形で生かしていけるというのは、図書館にとって親和性が高い活動として、今後このような形での貢献も考えられるのではないのでしょうか。インストラクションやコンサルタントだけではなく、このような形で学術コミュニケーションのインフラのところに关わるような活動を、図書館が行えるなら大きな貢献と言えるのではないかと考えています。

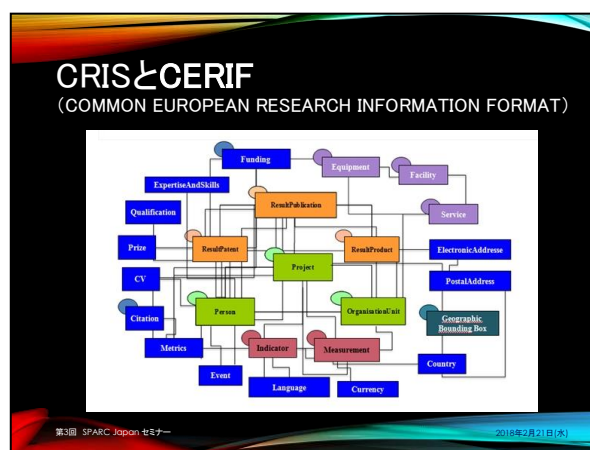
● Joseph 大変素晴らしかったです。Force11 の Scholarly Commons モデルを取り上げてくださってとてもうれしいです。図書館コミュニティでは、中央部分に位置する共有するツールと、運営に関する共通原則について、取り組むべきことがたくさんあります（図 4）。私たちは、こういった原則を定義し、ここでの成果、ワークフロー、ワークスペースをよりしっかりと管理できるようにならなければなりません。



(図 4)



(図 5)



(図 6)