第 2 回 SPARC Japan セミナー2015

「科学的研究プロセスと研究環境の新たなパラダイムに向けて
－e-サイエンス，研究データ共有，そして研究データ基盤 -」

事務職員のためのオープンデータ入門

星子　奈美
（九州大学附属図書館）

講演要旨

この発表では、事務職員という立場からオープンデータへの貢献について考えていきたい。オープンサイエンスの潮流が急速に高まる中、研究データ共有の重要性は学内事務組織においても認識されている。しかしながら、「研究データをオープンにすべき」という漠然とした理解を超え、事務部門として何をすべきか具体的に検討を始めている日本国内の大学は、現状ではまだ少ないと思われる。大学の事務職員がオープンデータに対応する新たな業務の枠組みを今後構築するにあたり、そのプロセスに対するイメージが明確になるよう、オープンデータの基本的情報を共有する機会となれば幸いである。

星子　奈美

九州大学附属図書館eリソースサービス室リポジトリ係長。平成27年度 SPARC Japanセミナー企画ワーキンググループメンバー。2006年9月より6か月間、実務研修のためQueensland University of Technology（オーストラリア）において機関リポジトリ業務に携わる。2007年から2010年、および、2013年から現在まで、九州大学の機関リポジトリ業務を担当。

私は九州大学附属図書館で機関リポジトリの業務を担当しています。今年度の SPARC Japan セミナーの企画ワーキングのメンバーを務めており、今回お話しさせていただく機会を頂きました。今回のセミナー講師の方は研究者が中心でしたので、私はそれとは異なる事務職員という立場から話をしたいと思っています。

研究機関におけるオープンデータの現状

まず、研究機関におけるオープンデータの状況について説明します。機関リポジトリは、初期に導入した機関で運用開始から 10 年ほど経過し、安定的に運用されている状態かと思います。そんな中、今年 3 月に内閣府から「我が国におけるオープンサイエンス推進のあり方について」という報告書が報告され、さらに 4 月に京都大学でオープンアクセス方針が採択されたこともあり、2015 年はオープンサイエンスの推進において象徴的な年だと言えます。

内閣府の報告書では、オープンサイエンスをオープンアクセスとオープンデータとに分けて定義づけていますが、各大学において、現時点ではオープンアクセスを優先的に進めている状況ではないでしょうか。九州大学においても、オープンアクセス方針策定のため検討を行っている段階です。

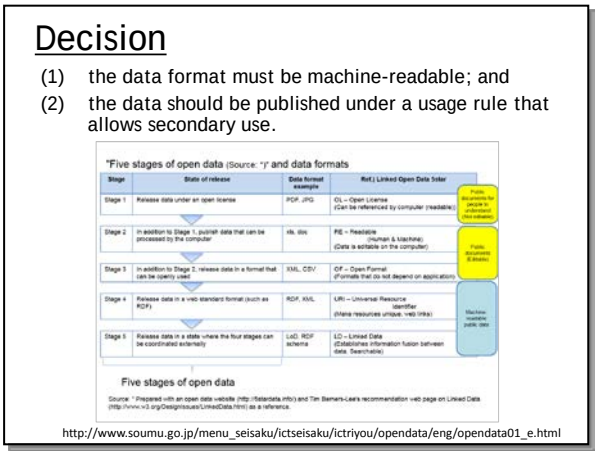
また、これまで各大学の事務組織においては、「研究データ」というものを業務の中で扱った経験がないので、データをオープンにするのだろうという曖昧な理解はあっても、業務フローの中でオープンデータをどうイメージしたらいいのか認識するのが難しい状況

です。

そこで、本日のプレゼンテーションは、非研究者による、非研究者にとっての、オープンデータ入門となればと思っています。お話しする内容は非常に基本的なことになるかと思いますが、少しでも具体的なイメージを共有できればと思っています。また、研究者として参加していただいている聴衆の方にも、非研究者が理解できていること、できていないことを知っていただくことにより、オープンデータについて今後協力していく上での対話がスムーズになればと思います。

オープンデータとは何か

次に、オープンデータの定義について確認します。総務省の「電子行政オープンデータ戦略」の中では、オープンデータと言えるための条件として、2 点が挙げられています（図 1）。1 点目は機械判読に適したデータ形式であること、2 点目は二次利用が可能なルールで公開されたデータであることです。さらに、オープンデータの五つの段階とデータ形式も提示されています。オープンライセンスの下、データを公開するという 1 段階目から、ウェブ標準のフォーマットで、外部連携可能な状態でデータを公開するという 5 段階目までに分かれています。これを見ると、オープンデータは、データをネットワーク上のどこかの場所に置いて終わり、というものではなく、データを共有しやすいよう、公開方法に関する具体的な検討が必要になってくるということが分かります。



（図 1）

この定義を見ると、オープンデータにおいては、共有できる状態であること、shareability が大きな重要性を持っているという印象を受けました。機関リポジトリは「研究成果のショーケース」と呼ばれることがあります。例えば店頭にケーキが美しく並んでいるように、研究成果を公開しているという意味合いですが、オープンデータの場合は少し状況が異なり、データが様々な人にとって道具として使える状態であることが重要になります。複数の人が共有する物品を棚の中で保管する時、取り出しやすいように片付いていて、どこに何が置いてあるのか分かるように整理してあれば、使う人全員にとって便利です。オープンデータはそんなイメージで考えると分かりやすいのではないかと思います。

行政データのオープン化の動き

ここからは、オープンデータを事務職員の業務の中で扱うことについて、より具体的なイメージを持つために、行政データのオープン化の動きについて見てみたいと思います。内閣府の報告書「我が国におけるオープンサイエンス推進のあり方について」の中でも、「これまでオープンデータに関する活動は、政府が保有する行政データ等の公開活動として捉えられることが多かった」と述べられているように、行政データの公開においては実運用が進んでいるため、研究データを扱う際に参考にできる部分があるのではないかと考えました。

例えば、図 2 は「オープンデータをはじめよう～地方公共団体のための最初の手引書～」という資料で、地方公共団体の事務担当者のために簡易な表現で書かれた手引書です。「担当チームを決めよう」から始まるオープンデータの手順が分かりやすく説明されています。

また、地方自治体ごとにデータ公開用のプラットフォームもできています。福岡市の Open Data of FUKUOKA CITY では、消防局が発信している火災や救急などのお知らせメールや、大気環境の測定データを

公開しており、研究目的で使うことも、一般市民の方が使うことも可能です。また、Open Data of FUKUOKA CITY では、オープンソース・ソフトウエアの CKAN（シーカン）が用いられていますが、大学等の研究機関でも、データの公開に当たり既存の機関リポジトリを用いるのか、あるいは、CKAN のようなデータ用のソフトウェアを別途利用すべきか、今後検討していく必要があると考えられます。

さらに、取り扱うデータの整備に関し、行政データにおいては、共通語彙基盤整備事業という事業があり、情報処理推進機構が情報連携用語彙データベースの構築やツールの整備を進めています（図 3）。これにより、異なる機関の間でデータ項目やデータ型、語彙が統一され、データを利用したサービス構築が容易になります。

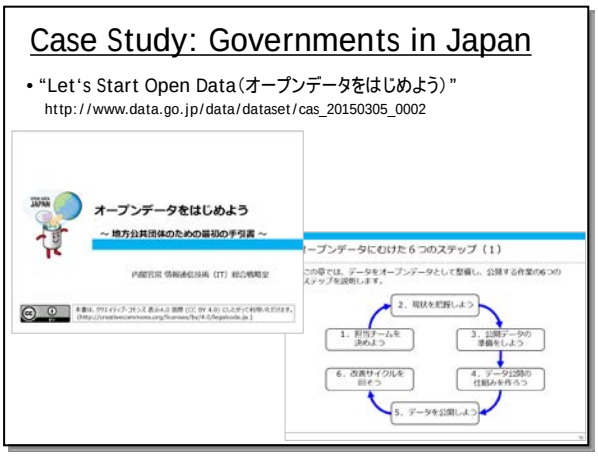
このように行政データで既に実施されているオープンデータの取り組みを、研究機関の事務職員の立場に置き換えることによって、今後の業務が想像しやすくなるのではないかと思います。

その一方で、「我が国におけるオープンサイエンス推進のあり方について」の中では、「国としての方針策定に当たっても、行政データのオープン化との違いを理解した上で検討すべきである」と明示されており、行政データとは異なる研究データ固有の課題についても認識しておくべきだと思われます。

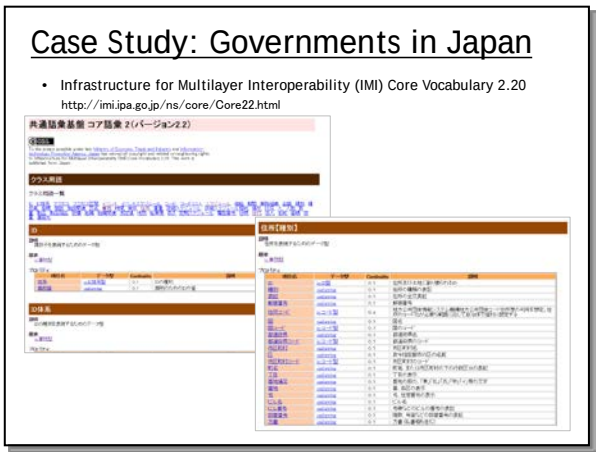
例えば、研究データの取り扱いに関してよく言及されるのが「データ管理計画（DMP：Data Management Plan）」です。DMP は、ある研究においてどのような種類のデータを扱うのか、あるいはどこに保存し、どういう取り決めで公開をするのかといった、データの取り扱いについての計画書です。研究助成団体に対し申請する際に、各団体の定める書式に従って DMP を作成し提出する必要があります。オープンサイエンス推進の潮流の中で、DMP を提出する研究助成団体が今後増えていくことも予想され、計画書を効率的に作成するための支援が事務職員の役割として重視される可能性があります。カリフォルニア大学が提供している DMP Tool は、テンプレートを選択して必要な事項を入力することにより、DMP が生成されるもので、カリフォルニア大学以外にも複数機関がパートナーとなって使用しています。

あるいは、研究データ固有の課題としてデータ量の大きさや多様さも挙げられます。オーストラリアのクィーンズランド工科大学が今年 5 月付で公開した「Guidelines for the Management of Research Data at QUT」という手引では、扱うデータの種類をイエス・ノーで分類し、それに応じて適切なストレージの選択肢を示しています。つまり、ストレージはデータの性質や用途に応じて使い分ける必要があり、その点に対応した公開プラットフォームを構築すべき、ということになります。

また、10 月 6 日に文部科学省から「学術情報のオープン化の推進について（中間まとめ）」が公表されましたが、この中では、大学等に期待される役割とし



（図 2）



（図 3）

て、データ管理計画の作成に関する支援、研究データの保管に係る基盤の整備（特にアカデミッククラウドの活用）、DOI 付与によるデータの永続的な管理、データ作成者の貢献の評価、データキュレーター等の人材育成が明記されています。

こうした課題への対応が必要になるにつれ、各研究機関においては、機関全体としての大きな動きが求められるようになると思われます。データ保管の基盤整備に関してはシステム部門、特許や法的な問題に関しては法務部門、研究戦略についてはリサーチアドミニストレーターというように、研究者だけでも、図書館だけでもなく、様々な学内関連部署の協力が重要になると思われます。また、オープンにしたデータをどうアピールするかという点では、大学のレピュテーションマネジメントという側面もあるかと思います。

図書館員が果たすべき役割

本日は図書館員の方も多く参加されていますので、特に図書館員が果たすべき役割についても考えてみます。

一つ目は、コーディネーターとしての役割です。あるいはファシリテーターという表現がより適切かもしれません。図書館員は日常の業務やイベント等を通じ伝統的に緊密なネットワークが存在し、大学内の他部署と比較しても全国的な情報共有が進んでいると言えるのではないでしょうか。オープンデータに関しても情報をいち早く入手し、学内・学外の様々な関係者間において調整役として貢献することが考えられます。

二つ目は、データキュレーターとしての役割です。これは「学術情報のオープン化の推進について（中間まとめ）」にも明記されており、今後具体的取り組みが推奨されていくのではないかと思われます。

以上、事務職員としてオープンデータに携わる際にポイントになると思われる点を列挙いたしました。関心を持たれた事項がございましたら、詳細については各情報源をご覧いただければ幸いです。