



■ 第4回 SPARC Japan セミナー 2014

「グリーンコンテンツの拡大のために我々はなにをすべきか？」

2015年3月9日(月) 国立情報学研究所 12F 会議室 参加者:68名

学術研究活動の中で生成された多様な学術資源のオープン化が、学術的観点からも、社会的要請の観点からも求められてきている。しかしながら、現在の我が国の学術資源の現状を考えると、研究分野によって、また機関・組織によって、オープン化に対する認知や動機・普及状況に大きな意識の差が生じている。また、多くの機関・組織においては、機関・組織内において生成された学術資源が複数のシステムに散在しており、組織的なオープン化推進のためには、横断的な管理体制モデルの構築も急務である。

現在、機関リポジトリに登録されているコンテンツの大半は論文であるが、本来の研究成果は研究データや標本等も含む多様なものである。本セミナーでは、まず学術研究機関が機関リポジトリ等において公開・発信する学術コンテンツを「グリーンコンテンツ」と再定義する。その上で、論文以外の研究データ、博物館資料のメタデータ及び資料画像データ等もオープン化のターゲットとして捉え、コレクション構築や利用促進など、学術資源管理の具体的な道筋を探りたい。本セミナーが、将来における学術資源のオープン化のイメージを共有するための第一歩になれば幸いである。

セミナー概要は以下のとおりです。当日の配布資料等含め詳細は SPARC Japan の web サイトをご覧ください。

(<http://www.nii.ac.jp/sparc/event/2014/20150309.html>)

講演

図書館によるデータ管理への道筋

南山 泰之(国立極地研究所)

2002年に Budapest Open Access Initiative で定義された「オープンアクセス」活動は、その後大きな広がりを見せ、2015年現在において、学術出版に関わる者で知らない人はいない、と言えるまでになっている。さらに、近年においてはより大きな概念である「オープンサイエンス」をキーワードに、従来の論文のオープン化に留まらず、その根拠となるデータや研究プロセスの公開が期待されている。本セミナーでは、データや研究プロセスの公開を通じて、知識の循環・再利用によるイノベーションが期待でき、また他業界との連携が促進される、というオープンサイエンスの理念に呼応した形で、データをキーワードにして研究データ、博物館資料などを繋ぐことを試みたい。

では、そもそもなぜデータが重要なのか。近年、研究の根拠であったり成果であったりするデータをもっと効果的に活用したい、という試みを全面に押し出すものとして、「データ中心科学」というあり方が提唱されている。「データ中心科学」では、論文もデータとして扱われ、収集された大規模で複雑なデータに基づき新しいデータが作られる

(データ駆動型の研究)という循環が期待される。一方では、研究データガバナンスの問題として、研究不正への対応・研究の透明性という切り口からも、データの組織的な管理の重要性が再認識されつつある。

上記のようなデータ保存・公開の意義を早くから認識し、海外では DCC (Digital Curation Centre)、RDA (Research Data Alliance) などの助成機関がデータの公開支援を行っている。これらを受けて大学図書館による取り組みも始まっており、英エジンバラ大学、パデュー大学などでは研究データリポジトリを立ち上げ、積極的なデータ公開支援を行っている。また、国内に目をやれば、研究者や助成機関、博物館等のコミュニティによるデータ公開の取り組みが既にあり、2014年12月頃からは内閣府や国立国会図書館、日本学術会議などにおいても大変活発な検討が行われている。このような現状を踏まえ、国内における大学図書館が果たすべき役割や、他コミュニティとの連携の具体的な手法を検討したい。



大学博物館における学術資料情報のオープン化に関する取組み

山下 俊介(京都大学宇宙総合学研究ユニット)

オープンアクセスやリンクドオープンデータなどの取り組みでは、情報の効率的な共有と活用が目指され、付加的な情報資源の蓄積も期待できる。しかし、完成物としての研究論文やジャーナル、あるいは公共サービスなどの実際業務と一体となって生み出される情報に比べ、博物館資料や学術アーカイブ資料といった手が届かない学術資料情報を作り出す活動への検討は進んでいない。山下氏は、これまで京都大学総合博物館において、学術アーカイブの構築に携わってきた。大学博物館では、研究プロセスと標本の生成・公開プロセスが密接に関係していること(大学博物館は object-based research の中心)、京都大学総合博物館での約 260 万点に上る学術標本資料の現状を紹介した。学術資料情報のオープン化については、“論文→標本、標本→論文”の関係を維持・確保することが重要であり、例えば分類学では、学名を記載(論文を公開)する際には、証拠標本の情報を論文に記載するという決まり(命名規約)が存在する。

京都大学で現在進めている「京都大学研究資源アーカイブ」は、京都大学における教育研究の過程において収集・作成された様々な資料類を体系的に収集・保存し、新たな教育研究の資源(研究資源)として運用することを目的としている。このプロジェクトでは、既存の学内資料保存施設では対象外となるフィルムや日誌などの資料を収集し、PEEK(研究資源アーカイブデジタルアーカイブシステム)において公開を行っている。氏は、堀田満映像資料(映画フィルム、1960-ca.1982)を例に、植物標本と映像資料を紐付けることの重要性を述べたが、アーカイブ公開における課題としては、固有 ID の付与やアーカイブ資料の階層性(コレクション>シリーズ>アイテムといった階層的記述を含む)と併せて、情報公開に至るまでのコスト負担についても将来的に解決する必要がある。

将来的な展望として、CCR(Connection between Collection and Research)構想が挙げられる。CCR では、タイプ標本(生物の新種記載を行う際に根拠となる標本)以外にも、多様な標本コレクションと研究(成果・データ)を繋



ぐインフラを構築し、研究サイクルとアーカイブ(コレクション生成)のサイクルを一体的に連携させることを目的としている。これにより、Web 上

のオープン化されている学術資料情報と論文・出版物、博物館での収蔵資料をリンクし、それらの関係性を明確化することが可能となる。学術資料のオープン化に際しては、拡散した学術情報をどのように再集約し実際の学術活動、つまりおおとの標本などの資料と結びつけていくのかが、大きなカギとなるであろう。コストを要する物資料などのアーカイブ化を促進・支援する枠組み(スキーム)こそが必要であり、本分野の今後の進展が望まれる。

機関リポジトリと DOI

- JaLC における DOI 付与について -

武田 英明(国立情報学研究所)

デジタル化以前は、研究者の最終的な研究成果は論文を意味し、データは論文執筆のための情報にすぎなかった。しかしながら、急激



にデジタル化が進みデータが巨大化している現在、データそのものが研究成果となり、論文とデータは一体化してきている。理論科学、実験科学、シミュレーション科学に続いて、近年、データ中心科学が出ている。シミュレーション科学やデータ中心科学の場合は、論文は研究の成果ではなく研究の紹介でしかない。データこそが研究成果である。ではなぜ研究データをオープンにするべきか?を考えると、1) 社会的な成果の共有、2) 公的資金による成果の公共性、3) 研究成果の継承と発展 4) 再現性の担保等が理由として挙げられる。

研究データ流通を支える情報基盤のレイヤーとして、メタデータの上位の識別子が重要となりつつある。データを記述するためのメタデータスキーマは多様で、メタデータのみでデータを同定し流通をコントロールすることは困難になりつつある。識別子には DOI の他、ORCID(研究者の識別子)、FundRef(助成団体の識別子)などがある。DOI は識別子からデジタルオブジェクトが存在する URI に変換するサービスである。もともとは出版社が論文識別子を共有するために作った制度であるが、現在は論文にとどまらず様々なデジタルオブジェクトの識別子に成長した。DOI サービスのメリットは、コンテンツへの確実なアクセス手段の提供であり、著者、読者、出版社、助成団体、すべてのステークホルダーにとって大きなメリットがある。

DOI 運営は、全体を統括する IDF(International DOI foundation)、登録機関(Registration Agency)、DOI 付与組織の三層構造になっており、論文の DOI 付与をミッション

としている CrossRef も RA である。その他、データセットへの DOI 付与を目的としている RA に DataCite がある。

ジャパン・リンク・センター (JaLC) も RA で、2012 年に発足し、第 1 段階としては主にジャーナル論文に DOI を付与してきた。2014 年 12 月に新システムに移行すると共に、JaLC は国内の様々な DOI 付与のニーズに対応できるようにポリシーを整備した。特に機関リポジトリのコンテンツに DOI が付与できるように拡張されている。JaLC DOI の目指す方向は、研究者の業績の全てをカバーできる DOI である。たとえば科研費等の成果のすべてに DOI を振ることができれば、研究者本人から研究機関、助成団体に有用であろう。

JaLC では、2014 年 10 月より研究データに対する DOI 付与実験プロジェクトも始めている。目標は研究データへの DOI 登録ポリシーの策定と運用フローの確立であり、国内ではおそらく初の分野をこえた研究データ関連機関の連携プロジェクトである。研究データには、メタデータスキーマや、データ粒度、データのライフサイクルと担当者の関係など、論文とは異なる多くの課題があがってきており、その解決にむけてプロジェクトを進めているところである。

研究成果はいずれは「データ」になっていくであろうし、研究データ流通基盤は必須となるであろう。DOI はその流通を指させる重要な要素となっていくと考えている。

パネルディスカッション

<グリーンコンテンツの拡大に向けて>

モデレーター: 堀井 洋 (一般社団法人 学術資源リポジトリ協議会)

パネリスト: 林 和弘 (科学技術・学術政策研究所) / 南山 泰之 (国立極地研究所) / 山下 俊介 (京都大学 宇宙総合学研究所) / 武田 英明 (国立情報学研究所)

パネルディスカッションに先立ち、林和弘氏 (科学技術・学術政策研究所) による「オープンアクセスからオープンサイエンスへ: 俯瞰と政策的要点」の紹介、今回のモデレーターでもある堀井洋氏 (一般社団法人学術資源リポジトリ協議会) による「学術資料に関連した活動のご紹介」があった。以下に概要を示す。

林氏: 国際社会は 2010 年代に入りオープンアクセスからオープンサイエンスに移行しつつあり、市民がサイエンスに意識・無意識的にも近づく時代になってきている。近年はサイエンスコミュニケーションなど積極的に市民にコミットする試みもあり、オープンアクセスの時代よりも様々なステークホルダーが関与してきている。

第 4 期科学技術基本計画においてもオープンアクセスの推進に言及はあったが、これを受けた活動は機関リポジ



トリの構築と学会誌の電子化支援にとどまっていた。大きく動きがあったのは 2013 年、G8 のオープンデータ憲章への合意と並行して、G8 科学技術大臣会合においてリサーチデータのオープン化につき日本も合意している。

これらの潮流を踏まえ、今後どのようなことが求められているのか。政策的には、学術面に加え、経済的効果、産業振興、教育効果などその波及が求められてくる。もっとも、全ての分野の研究者がオープンアクセスによって有利になる訳ではないため、国益という観点からは、今後何をオープンにして何をクローズにするのか、を検討する必要がある。これからは、クローズにすることに理由が求められる時代になるかもしれない。図書館業界としてはデータジャーナルの動向に注目すべきであり、研究データの質、データ作成者の貢献度などといった新しい要素が現れてくるだろう。

堀井氏: 学術資源リポジトリ協議会 (愛称: Re*poN) は 2014 年 10 月に設立され、大学・企業など学術に関わる人間でデータ化に取り組んでいる。Re*poN ではこれまでに明治～昭和期における科学実験機器資料・教育掛図資料の電子化、金沢大学 virtual museum project の構築などを行ってきており、その他資料の調査分析、メタデータ・デジタルデータの作成、組織間の交渉などを通じて情報基盤の構築・整備を担うことを目指している。研究データ生成の事例としては「加賀藩先祖由緒并一類附帳」があり、科研費で情報の抽出・電子化を行ったものを研究コミュニティ「加賀藩研究ネットワーク」で限定公開している。一般公開への課題としては、1) 汎用的な公開研究データとしての精度・完成度、2) 資料所蔵者・関係者との、公開に関する承諾、3) 公開データを作成する労力・コスト負担があり、データの整形・監修者の存在が重要と考えている。

パネルディスカッションでは、今回のテーマ「グリーンコンテンツの拡大」に即して、様々な立場のパネリストから意見が述べられた。以下、テーマごとに列挙する。

【博物資料や研究データにまで学術資源のオープン化の範囲を広げる意義や目的】

(武田氏) オープン化の流れは国際社会の状況変化によ

るところが大きい、個人的な興味としては図書館や機関リポジトリがどう対応していくか、どのような意義を見出すのか、というところにある。この問題は、研究プロセスの中に図書館がどこまで関与できるのか、という問題にも関係してくる。また、今後学術分野の独自性は強まるだろうが、共通部分を探ることが重要になるだろう。

(山下氏) 博物館においては、以前から資料のデータベース化や公開はなされてきた。オープン化によって、これまでのそうした成果が他とどう繋がるのか、を考えるタイミングにきたのではないかと。

(南山氏) 図書館にとって公開・整理という行為は本質的なものであり、存在意義のようなもの。電子媒体を取り扱う上でも、その役割は変わらない。

(林氏) オープン化による学術分野や関連分野の産業の発展なくしては、研究予算の増加もあり得ない。図書館がどのようにデータに関わるか、という点に関して言えば、データジャーナルで問題になるクオリティコントロールは、データの内容そのものに関するチェックではなく、あくまで様式に関わるもの。様式のチェックに関する知見は、図書館に蓄積されている。

【オープン化への筋道】

(南山氏) オープン化の対象を確定させるためには、オープンとクローズの線引きを急ぐ必要がある。図書館においては、その前提としてまずリポジトリでどの範囲のデータを取り扱うのか、といったポリシー策定を、研究者と連携しながら行う必要がある。実務においては、(そもそも図書館員が主題を理解できれば早いのだが) まずは院生や URA と協力して取り組みたい。

(林氏) データに関しては、データマネジメントプランが必

-----参加者から-----

(大学/図書館関係)

- ・大変充実したセミナーで大いに勉強になりました。

Re*poN についても初耳でしたし、オープンサイエンスの動きなども教えて頂き、わずか4時間で最近の動きや課題、ニュースまで手に入れることが出来るありがたいセミナーでした。

・直接すぐに業務に反映するような内容ではありませんでしたが、世界のトレンドや、今後の日本の動きが分かり刺激を受けた。

・思っていたよりも概念的な内容が多いと思った。全体像をつかむのには良かった。

(大学/大学・教育関係)

要。研究を計画する時点から図書館が関わる必要があり、これに関するトレーニングは一つの提案になる。ポリシー策定は、アクションプランとして URA と協力すべき。

【博物資料の孤児化対応】

・作成者と連絡が取れず、ライセンス処理ができない資料をどう取り扱うか。

(山下氏) 難しい問題。現在はアーカイブ化対象資料の公募時に公開可能、権利処理可能なものを優先的に選ぶことで対応している。

(林氏) 筋としては文化庁裁定になるが、過去に学会論文で同様の事例があった際には、ネット上での告知で対応した。論文であれば訴訟リスクは少ないと考えての対応だが、博物館資料ではわからないのであまりお勧めできない。

(フロア) 学術流通に関する内容なので、公開に問題があるとも思えない。まず公開してしまい、クレームがあった際に誠意を持って対応すれば良いのではないかと。

【博物資料公開のプライオリティ】

(山下氏) 研究論文等に使用されるタイプ標本・資料に関する情報を優先的に公開することも必要ではないかと。

(堀井氏) 学術情報の公開に際しては、その資料の概要的な情報の公開からはじめて、状況に応じて段階的に公開情報の内容を詳細にしていく方法もあると考える。

議論も尽きないところではあったが、最後に各パネリストからデータ管理に関する決意表明がなされ、ディスカッションは終了となった。

- ・データ公開に関する様々な課題がわかった。

(大学/その他)

・政策としてサイエンスデータのオープン化への流れが既にあることに気付かせてもらい有益だった。

(大学/学術誌編集関係)

・研究データのシェアに関する世界の潮流を理解することが出来ました。雑誌の編集の立場からは、この流れに対しどのようなことが出来るのか考えさせられました。(データジャーナルの刊行?リポジトリへのデータ提供を著者に呼びかける?)ただ研究者の間ではデータのシェアについてはあまり話題になってないように感じます。

(企業/大学・教育関係)

・政策レベルでどのような動きになっているかについて
の話を興味深く聞かせていただいた。大きな潮流の中で
図書館がどう動くべきかという問いはこのような俯瞰的な観
点が必要。

(企業/その他)

・グリーンコンテンツの現状や今後についてよくわかった。
(その他/図書館関係)
・機関リポジトリを進める上で、必要な考え方、研究デー
タの扱いを考える必要性に気付かされました。

-----企画後記-----

😊 実は SPARC Japan に参加するのは、これが初めてで
「なるほど、こういう場なんだ」というのが感想です。

それはさておき、最近、いろいろな方向から研究デー
タの話がやってくるのですが、考えれば考えるほどデータと
論文のとりあつかいの違いを強く感じます。でも、ものすご
く大事な話だということも日々感じています。そもそも論文
がいつまでも学術情報の中心でありつづけるのかどうか、
わからなくなってきています。

三角 太郎 (千葉大学附属図書館)

😊 データのオープン化がテーマ、ということで始まった
本セミナーの企画でしたが、並行して国内の関係機関が
急速な動きを見せ、日ごと情報を追いかけての準備と
なりました。将来のデータ管理を様々な立場から議論する、
という楽しさを皆様と共有できていましたら、大変嬉しく思
います。

南山 泰之 (国立極地研究所)

😊 企画 WG はもとより、SPARC Japan セミナーへの参加
自体も初めてでしたが、色々と刺激的な体験をすることが
できました。今回のセミナーでは、オープン化の対象範囲
に博物資料も(!?)ということで、一般社団法人学術資
源リポジトリ協議会のメンバーでもある京都大学山下先生
とともに、博物資料に関する現状と課題を少しでも多くの
参加者の方々にお伝えしたいと考えました。古文書や標
本などのアナログな資料をネットワーク上で扱えるようにす
るには、多くの人々の努力と協力が必要です。パネルディ
スカッションでの議論が纏まらなかったことは、大変申し訳
なく思っておりますが、今後も様々な立場の人間が学術情
報のオープン化について、ざっくばらんに楽しく議論でき
る場が形成されることを切望します。

堀井 洋 (一般社団法人 学術資源リポジトリ協議会)