



SPARC Japan ニュースレターでは、各回セミナーの報告に講演やパネルディスカッションを書き起こしたドキュメントを加え、さらにそのほかの SPARC Japan の活動をご紹介します。今回は国立情報学研究所実務研修（CERN）についてご報告します。

CONTENTS

■ SPARC Japan 活動報告
 SPARC Japan 運営委員会
 国立情報学研究所実務研修(CERN)

■ SPARC Japan セミナー報告
 概要
 参加者から
 企画後記
 ドキュメント
 (講演・パネルディスカッション)

■ SPARC Japan 活動報告



SPARC Japan 運営委員会

SPARC Japan 運営委員会の会議資料をウェブサイトで公開しています。

<http://www.nii.ac.jp/sparc/about/committee/>

国立情報学研究所実務研修(CERN)

第4期及び第5期 SPARC 事業の達成目標の一つとして「オープンサイエンスへの活動スコープの拡大」を掲げており、2016年に高エネルギー物理学分野の情報サービスに係る国際連携協定を締結しました。その具体的活動として国立情報学研究所実務研修生を CERN に派遣しましたので報告します。

- 物理学分野のデータベースに関するデータキュレーション（国立情報学研究所実務研修）
 NII、KEK（高エネ研）、CERN の3機関の国際連携協定に基づき、国立情報学研究所実務研修の一環として、以下を実施しました。
 研修期間：平成28年3月11日（火）～平成29年3月2日（木）
 所属機関・氏名：京都大学・大村 明美
 研修場所：CERN（欧州原子核研究機構）
<http://www.nii.ac.jp/hrd/ja/jitsumu/h28/index2.html>

■ SPARC Japan セミナー報告



第 3 回 SPARC Japan セミナー 2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

2017 年 2 月 14 日（火） 国立情報学研究所 12F 会議室 参加者：118 名

本セミナーでは、今年度のセミナーの総括として、これまでのセミナーの内容を受けつつ、オープンサイエンスのあり方とその方向性について様々な面から検討しました。研究データをテーマとした研修の報告、トレーニングツールの紹介もあり、オープンサイエンスに係る企画や実践の参考となりました。

次ページ以降に、当日参加者のコメント（抜粋）、企画後記およびドキュメント全文（再掲）を掲載しています。その他の情報は SPARC Japan の Web サイトをご覧ください。(<http://www.nii.ac.jp/sparc/event/2016/20170214.html>)

概要



今年度の SPARC Japan セミナーでは、年間テーマの「オープンサイエンス時代の文献とデータの流通：科学的知識創成の新たな標準基盤へ向けて」に対し、第 1 回セミナー「オープンアクセスへの道」では、グリーンオープンアクセスとゴールドオープンアクセスの関係性や担うべき役割について議論がなされ、国内研究者の APC 支払額の把握と SCOAP³ モデル発展の必要性などが確認された。

第 2 回セミナー「研究データオープン化推進に向けて：インセンティブとデータマネジメント」では、データ・サイテーション、データ・ジャーナル、データ・オーナーシップなど研究者へのインセンティブのあり方や、それを支える研究データ管理の具体的方策（オープン化にかかるコストとのバランス、研究データ管理に対する報酬等）、などが議論された。

第 3 回セミナーでは、上記を受けて、一年間の SPARC Japan セミナーを総括し、オープンサイエンスを「しなければならないこと」「すべきこと」「したほうが利益があること」「せざるを得ないこと」など多角的に再考することで、オープンサイエンスの先にある科学的知識創成の新たな標準基盤について考えてみたい。

* 標準基盤：学術研究（＝科学的知識創成）を支援する基盤は、それを構築するプロセスにおいて、多分にボトムアップな活動によって構築され、いずれは利用者に意識されることなく学術研究の環境として位置付けられていく。このようなプロセスが、インターネットスケールで起こっているのが現在であり、地域や学術分野を越えて、学術環境のスタンダードを確立する活動が芽吹いている。本セミナーではこのような活動がもたらすスタンダードを標準基盤と定義する。

参加者から

(大学/図書館関係)

・東大の小野先生の発表でのアンケート結果がとても興味深かったです。

(大学/研究者)

・北本先生の講演はディープラーニングの最新動向がきけて大変興味深かった。Dekker 氏の発表はヨーロッパの現状がよくまとまっていたよかった。尾城氏の発表は実践的で役立ちそうである。ぜひ使ってみたい。

(企業/図書館関係)

・情報がオープンになるスピード、びっくりしました。



(その他/図書館関係)

・今までのオープンサイエンス、オープンデータのセミナーやフォーラムは概念や理想の話あるいは実現の道程は遠く、どうしたものかと思う話ばかりでしたが、今回は実際の現場に沿った話がきけて前向きな気持ちになれました。

・従来のオープンデータいけいけ！の勢いから制限共有、制限公開のほうへ流れが変わったのかと思わせる講演がいくつかあり、パネルディスカッションで小賀坂氏や小野氏のお話を聞くまでは訝しく感じていました。けれども、最後にお二人が語っていたようにオープンサイエンス、オープンデータを否定するものでも制限共有ありきでもなく、研究者にオープンデータが浸透し納得のいく形で進めていくことが大切で、やるからにはオープン化の結果に責任を持つ、見通しをつけておかなければいけない、との話に、今まで以上に重大さが身にしました。また勢いに流されずに落ち着いていかなければと思いました。

(その他/その他)

・著作権とデータの関係について明確になった。DMPの具体例が紹介され、イメージがつかめた。

企画後記

😊 1年間を総括する今年度第3回の SPARC Japan セミナーの主査という大役を、無事果たすことが出来ました。これも、登壇者、運営委員、事務局、企画 WG、ご来場の皆さまのお力添えがあつてのことだと思います。この場をお借りして皆さまに御礼申し上げます。SPARC Japan セミナーは研究者と図書館職員が協働し、企画・開催しています。今後も様々な立場から学術情報流通に関わるホットな話題をご提供できるセミナーであり続けることを期待します。

梶原 茂寿
(北海道大学附属図書館)

😊 今回は twitter 係として、セミナーの内容の情報発信を担当しました。「再考」という重いテーマでしたが、数歩先を進んでいる欧州の実際、データ共有、DMP やライセンスの現実の姿など、地に足を着けて「再考」できたのではないかと思います。企画担当でしたが、お話を伺いたかった方のお話を伺えて役得でした。

小野 亘
(東京学芸大学)



😊 深層学習、計算機科学の分野では、arXiv が最新成果を出す主戦場になっており、プログラムソースや結果の公開・引用が1日単位で行われていることは驚きでした。スピードがオープンサイエンス推進の原動力になっていることは、現在ではある特殊な分野での出来事かもしれませんが、一つの要素として念頭に置いておいて損はなさそうです。また、オープンサイエンスの中でファンディング機関が果たす役割は大きい、データマネジメントプランに関する現実的な実施方



針・評価軸策定は、日本だけでなく欧米でも手探り状態であることも分かりました。今後は、データの提供者である研究者を大きく巻き込んでいく方策を考えていかないといけないですね。

能勢 正仁
(京都大学大学院理学研究科)

😊 本年度の最後を締めくくるセミナーは第1回の論文のオープンアクセス、第2回の研究データ共有の議論を踏まえつつ、総括として新しい研究プラットフォームの様子をさぐり、具体的に何ができるかを考えるというものでした。欧州を中心としたオープンサイエンスのビジョンから DMP 義務化や作成支援まで、幅広く採り上げられ、研究者の本音も飛び交うディスカッションコントロールが中々に大変でしたが、様々なステークホルダーが集まるようになった SPARC Japan セミナーの醍醐味として今後も対話の場が提供できればと思いました。

林 和弘
(科学技術・学術政策研究所)

😊 本セミナーでは twitter での発信を担当しました。第1回、第2回を受けての総まとめとなる今回、議論が収斂する方向に向かうかと思いきや最先端の話題が次々と飛び出し、改めてこのテーマの奥深さと関係者の本気度を実感しました。国内でも着々と事例が積み重なっている今、来年度の展開が一層楽しみです。

南山 泰之
(国立極地研究所)



第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

開会挨拶／概要説明

梶原 茂寿

(北海道大学附属図書館)



梶原 茂寿

2014年より、北海道大学附属図書館本館に勤務。学術システム課システム管理担当で、図書館情報システム及びネットワークの管理と機関リポジトリを担当。2015年、2016年SPARC Japanセミナー企画WGメンバー。

第3回 SPARC Japan セミナー2016「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」にご参加いただき、誠にありがとうございます。

第3回セミナーで目指すもの

今年度の SPARC Japan セミナーの年間テーマは、「オープンサイエンス時代の文献とデータの流通：科学的知識創成の新たな標準基盤へ向けて」でした。

第1回セミナー「オープンアクセスへの道」では、文献に焦点を当て、グリーンオープンアクセスとワールドオープンアクセスの関係性や担うべき役割について議論がなされ、国内研究者の APC 支払いの捕捉、SCOAP³ モデルの発展の必要性などが確認されました。

第2回セミナー「研究データオープン化推進に向けて：インセンティブとデータマネジメント」では、データの流通に焦点を当て、データサイテーション、データジャーナル、データオーナーシップなどの研究者へのインセンティブの在り方、それを支える研究データの管理の具体的な方策などについて議論がなされました。

そして、この第3回セミナーでは第1回・第2回のセミナーを受けて、1年間の SPARC Japan セミナーを総括し、オープンサイエンスについてさまざまな方面から再考してみます。オープンサイエンスの先にある、科学的知識創成のための新たな標準基盤について、皆さまと共に考えたいと思います。

セミナーの流れ

本セミナーでは、午前に二つの基調講演を行い、午後は五つの発表を行います。その後、パネルディスカッションということで、長丁場になりますが、皆さまにも活発に発表、質問などしていただけるようになっていますので、ぜひご参加いただければと思います。

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

欧州から見たオープンサイエンス

Ron Dekker

(欧州委員会研究・イノベーション総局

(European Commission (DG Research & Innovation)))

講演要旨



オープンサイエンスは政治的な課題として重要になっている。2016年に欧州理事会でオランダが議長国となっている間に、28加盟国はオープンサイエンスに関するEU競争力担当相理事会の結論を採択し、オープンサイエンスに関するアムステルダム行動要請(仮訳、Amsterdam Call for Action on Open Science)を表明したオープンサイエンスの会議があった。欧州委員会はヨーロッパオープンサイエンスアジェンダを策定し、オープンサイエンスクラウドやオルトメトリクスといったテーマでいくつかの専門家グループを設置した。またEU加盟の数カ国では国家的にオープンサイエンスポリシーや戦略を制定している。その他、大学や出版者、資金配分機関等のステークホルダーもまた市民科学を含むオープンサイエンスの活動に関わっている。同時に、例えば学術情報流通の方法といったことへの重要な変更は先行者不利益によって阻害され、大きな財政的な再分配を必要とする。

そしていま、われわれはどのような状況にいるのか。どのようにオープンサイエンスエコシステムの変化を誘導できるか。どのようにして出版への配慮をしつつ、オープンアクセスへの移行を触媒し変化を起こすことができるか。さらに、研究データシェアリングを促進するためには何が必要とされるか。

本講演では、ヨーロッパの状況と上記の課題について述べるとともに、可能な解決策についての議論を呼び起こしたい。また出版や研究データについて主に語る一方で、オープンサイエンスの他の側面やイノベーションを含めた一般的な科学と社会の関係についても触れたい。

Ron Dekker

マーストリヒト大学で経済学を専攻し、労働市場研究のキャリアをスタートした。1995年にティルブルフ大学に入り、研究の軸足をデータマネジメントに移した。これはオランダ科学研究機構(Netherlands Organisation for Scientific Research (NOW))での業務のきっかけとなり、1997年にはデータ部門の責任者となった。その後、社会科学部と中央計画・研究所部のチームコーディネーターとして勤務した。2007年NOW研究所ディレクター。2013年にはオランダ高等教育研究ITイノベーション機構(仮訳、Dutch IT-innovation organisation for Higher Education & Research (SURF))でディレクターを務め、2014年にオランダのEU議長国としての準備のためにオープンサイエンスのプロジェクトリーダーとして教育・文化・科学省(Ministry of Education, Culture and Science)に出向した。2016年、欧州委員会研究・イノベーション総局オープンサイエンス担当加盟国出向専門家(仮訳、Seconded National Expert on Open Science at the European Commission, Directorate-General Research & Innovation)となった。2017年3月から、Consortium of European Social Science Data Archives (CESSDA)ディレクターとなる予定。CESSDAはヨーロッパ最大のインフラ(いわゆるESFRI(研究インフラ欧州戦略フォーラム)Landmark)のひとつで、本拠地はノルウェーのベルゲンである。



私は欧州委員会に出向している立場なので、欧州委員会を代表して発言するわけではありません。今回は欧州委員会のオープンサイエンスに関する動向を踏まえ、オープンサイエンスとは何であるか、欧州のオープンサイエンスポリシーについて、出版物とデータのオープン化に焦点を絞ってお話しし、それからオープンサイエンスの展望について述べたいと思います。

オープンサイエンスとは何か

オープンサイエンスとは、科学を変化させるものです。アイデア・概念、分析、データ収集、出版、レビューという研究サイクル全体が変わります。この状況は、デジタル化によって生じています。生産されるデータの量は指数関数的に増加しており、透明性が求められるようになっていきます。例えば、心理学では公表された結果の約半分に再現性がありません。従って、説明責任を満たした研究を実現する新しい方法を見つけなければなりません。また、科学は、企業を含む社会全般とより良いつながりを持ち、特に貧困と疾病という大きな社会的課題に取り組む必要があります。

オープンサイエンスとは、データ、出版、ソフトウェア、オープンノート（研究公正）、市民が積極的に科学に参加する市民科学から構成された、包括的用語と見ることができます（図 1）。オープンサイエンスを一つに定義することは難しいですが、私は「オープンサイエンスとは、あらゆる種類の科学的知識は、発見と同時ぐらいに、早いタイミングで公開されなけれ

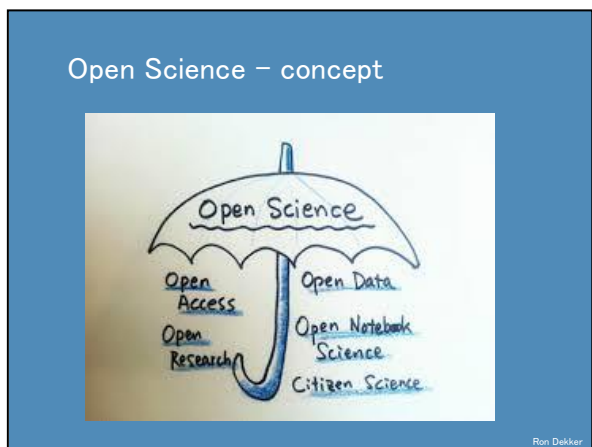
ばならないという考えである」というマイケル・ニールセン（Michael Nielsen）の定義が好きです。直ちに共有しなかったり、壁を維持したりする理由はいろいろあるでしょうが、核心はできるだけ早く公開することです。

現在、既に利用可能なオープンサイエンスのツールがたくさんあります（図 2）。この内側のサイクルは、データ収集、分析、出版、レビューといった伝統的な科学です。外側のサイクルはオープンサイエンスで、内側のサイクルに比べてより外界との相互作用があります。オープンサイエンスとは、このように科学をつないでいくものなのです。

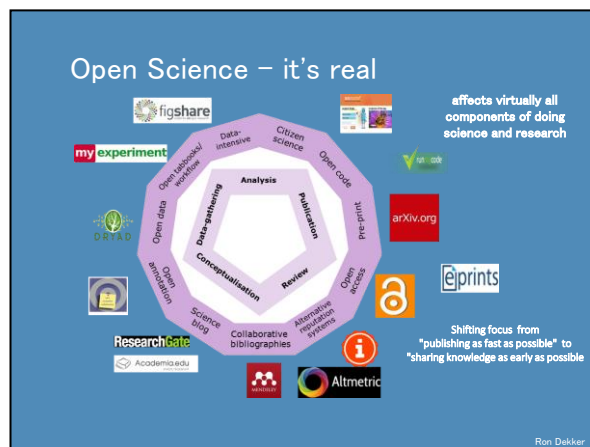
欧州のオープンサイエンス政策

2014 年に欧州委員会は、研究者、出版社、資金提供者といったオープンサイエンスのステークホルダーとのコンサルテーションを開始しました。主なトピックは、出版、データおよび研究インフラでした。

図 3 は、コンサルテーションで出てきたオープンサイエンスの主な促進要因です。同時に促進の障壁となるものも挙げられました。その一つが品質保証です。それは、直ちに公開された場合に、どのように品質を保証するのか。データを共有したいとき、いかにクレジットを提供するのか。データと知識を共有するのに十分なインフラはあるか。研究者とそれ以外の人々は利点を認識しているか、ということでした。そこで私たちは、データと出版物について多くの優先事項を設



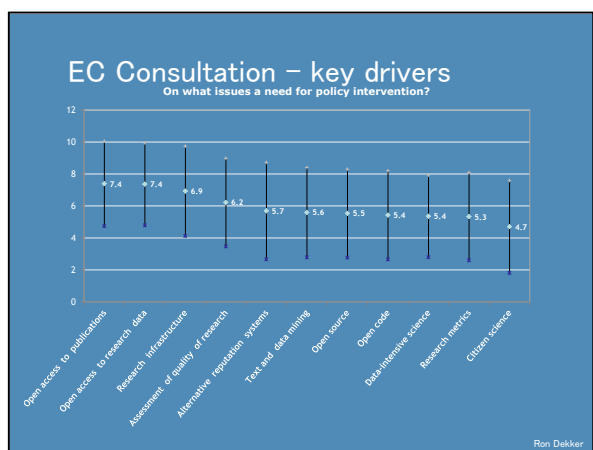
(図 1)



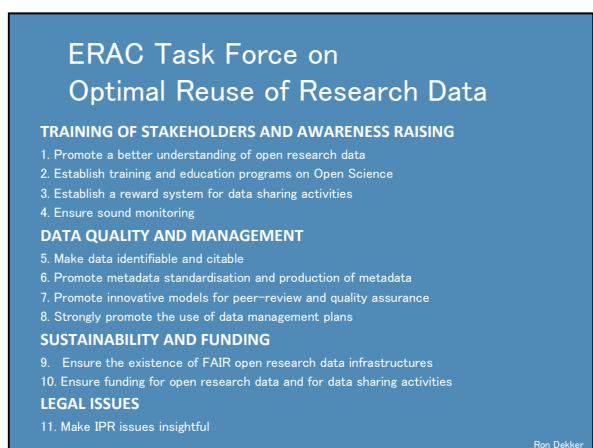
(図 2)

定しました。これは欧州委員会委員のカルロス・モエダス（Carlos Moedas）だけでなく、オランダ教育・文化・科学省副大臣のサンダー・デッカー（Sander Dekker）にも取り上げられました。

昨年、オランダは欧州理事会の議長を務め、オープンサイエンスを優先させることを決定しました。まず簡単に議長国オランダの成果を紹介し、欧州委員会の優先する政策について話します。私たちは、議長国時代にデータと政策に関する成果物をつくることに決めました。ERAC Task Force on Optimal Reuse of Research Data（研究データの最適な再利用に関するタスクフォース）は、データトレーニング、データ管理、持続可能性、IP 問題などに関する数多くの活動と勧告に取り組みました（図 4）。データの最適な再利用を追跡することはとても複雑です。私たちはここで、あえて「オープンデータ」という言葉は使いませんでした。



(図 3)



(図 4)

「オープンデータ」という言葉には、データは直ちに公開されるべきだということを示唆する面がありますが、データの中には機密性があるものや、そのデータを最初に使用する権利があるものもあるからです。そうした場合には、自分に再利用の権利があることを明らかにしなければなりません。

これらを実現するため、アムステルダムで会議を開催し、オープンサイエンスに関するアムステルダム行動要請（仮訳、Amsterdam Call for Action on Open Science）を表明しました（図 5）。そこで決めた目標は、一つ目は、2020 年までに科学出版物の完全なオープンアクセス化を達成することです。二つ目は、データ共有のための新しいアプローチを準備することです。三つ目は、データ共有の際に人を認識し、クレジットを与える新しい報酬と評価のシステムをつくることです。東ヨーロッパの現在のシステムでは、データの共有に対する報酬はありません。権威ある雑誌に出版すれば報酬が与えられますが、教育、データの共有、知識の共有などに対しては報酬がありません。この報酬制度を変えていかなければ、共有するインセンティブがありません。四つ目は、少し秘密めいていますが、つまり知識の共有と監視システムです。率直に言うと、国や大学には出版の費用について何らアイデアがありません。購読の契約費用は知っていますが、例えば、ゴールドオープンアクセスの論文掲載料（APC）については知らないのです。従って、これらのデータを収集し、出版コストに関する情報を共有する必要がある



(図 5)

ます。他にも、オープンサイエンスに関するアムステルダム行動要請には解決の方向性を提示する 12 の推奨事項があり、各ステークホルダーは自分がやるべきことに取り組みます（図 6）。

2016 年 5 月、欧州理事会の全 28 加盟国は、EU 競争力担当相理事会の結論を採択しました（図 7）。全加盟国が「オープンサイエンスを望んでいる」と表明したことは、強力な政治声明でした。この結論の一つ目は、オープンサイエンスの重要性です。二つ目は、ステークホルダーの代表で構成される Open Science Policy Platform を設立することです。このプラットフォームは、出版社、資金提供者、大学、応用研究の従事者、若手研究者からなる 25 人のメンバーで構成されています。欧州委員会は、オープンサイエンスを進める方法についてアドバイスを求めました。三つ目は、オープンサイエンスが進むべき方向として、できるだ



(図 6)



(図 7)

け早く結果をオープンにし、テキスト・データマイニング（TDM）を行い、著作権を保持するということです。これは、有料アクセスかまたは全くアクセスがなくなってしまう、という現状から抜け出す一つの解決策になり得ます。四つ目は、出版物のオープンアクセスです。全ての加盟国は、2020 年までに出版物をオープンアクセス化することに合意しました。2020 年までのオープンアクセス化ということは非常に重要でした。なぜなら、政治声明で期限を示すことは重要なことからです。五つ目は、研究データを、より一般的に最適に再利用できるようにすることです。データは公共財でなければなりません、この意見はデータを扱う全ての研究者の間で共有されているわけではありません。

欧州では、現在、強力な政治声明を出しています。これは全ての加盟国間で合意しています。私たちはオープンアクセスとオープンサイエンスに関する国家政策を実行に移すために、ヨーロッパオープンサイエンスアジェンダ、ハイレベル専門家グループ、オープンサイエンスポリシープラットフォーム、Horizon 2020 をつくりました。

アムステルダム行動要請と同様に、ヨーロッパオープンサイエンスアジェンダは、障壁の除去、インフラを提供するインセンティブをいかに与えるか、オープンサイエンスをノーマルサイエンスにすることを目的としています。オープンサイエンスは、できるだけ早く、知識を生産し共有するための一般的な方法とならなければなりません。

このアジェンダでは、報酬システム、品質および影響の測定、将来の出版モデル、探索可能・アクセス可能・相互運用可能・再利用可能（FAIR）なオープンデータ、欧州オープンサイエンスクラウド、研究公正、市民科学、オープンエデュケーションおよびスキルの八つの重要な問題に焦点を当てることが決められました。品質に関しては、現在のジャーナルのインパクトファクターは、研究や論文ではなく、ジャーナルの品質を測定するためのプロキシです。欧州オープンサイ

エンスクラウドは、データと全ての研究成果に対するインフラとしての役割を果たしています。研究公正に関しては、結果を再現可能なものにすることが重要です。欧州委員会は、これらの項目に関するハイレベル専門家グループを既に設置し、または今後設置する予定です。

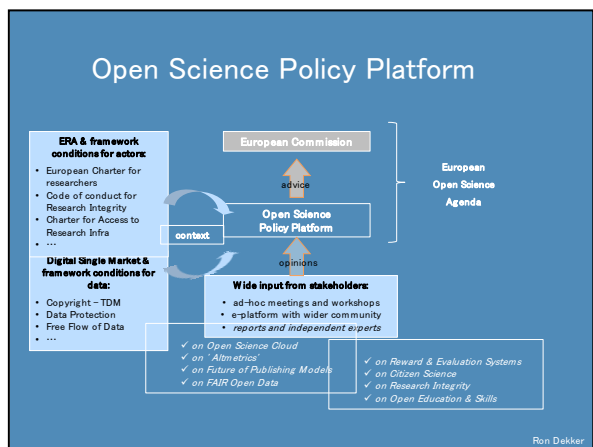
また、欧州委員会は、オープンサイエンスポリシープラットフォームを導入しました。このポリシープラットフォームは、委員会に助言する 25 人のステークホルダーから構成され、ヨーロッパオープンサイエンスアジェンダの八つの項目を、欧州委員会がオープンサイエンスへの移行を促進し、触媒することを推進できるように、包括的に把握することが期待されています。

チャートを使って説明します。まず左は入力側です（図 8）。オープンサイエンスポリシープラットフォームはワーキンググループを持ち、ステークホルダーから意見をもらうことができます。そしてプラットフォームは、最終的に欧州委員会に助言するだけでなく、結果を機関に還元します。ワーキンググループのステークホルダーがプラットフォームの方策に合意すれば、それは全てのステークホルダーによって容易に実施されることになります。

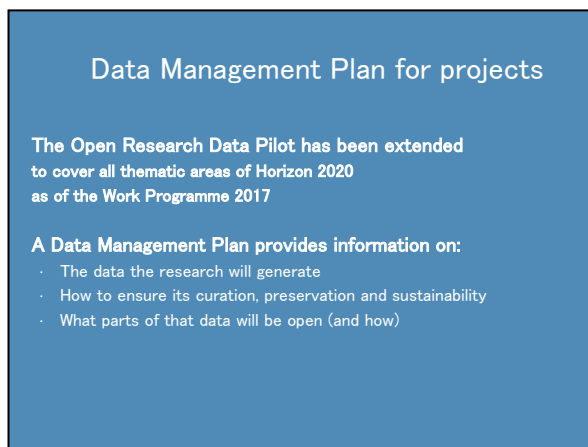
また、欧州委員会は、リサーチプログラムの中でオープンサイエンスに取り組んでおり、プログラムの中でも EU 第 8 次研究・イノベーション枠組み計画「Horizon 2020」が最も重要なものです。その中では、出版物をグリーンまたはゴールドでオープンアクセス

化することを義務付けています。できるだけ早くデポジットし、領域ごとに決められた 6～12 カ月間のうちに利用可能にする必要があります。ゴールドオープンアクセスの費用は、助成金から支払うことができます。同じことがデータにも当てはまります。助成金受領者は、データをリポジトリに保管し、アクセスを提供する必要があります。私たちは FAIR 原則にのっとり、出版物だけでなく、データも共有したいと考えています。

データ管理計画（DMP）のパイロットも行いました（図 9）。これは若手研究者に、他人に有益なものを生み出すことへの意識づけをするためです。パイロットへの参加は任意であり、本パイロットに関わる人の約 3 分の 2 が参加しています。本パイロットの適用範囲外のプロジェクトから、さらに 10% が自発的に参加しました。このパイロットについては、Horizon 2020 の全てのプログラムをカバーするように拡張することにしました。研究によって生成されるデータ、どのようにしてデータを提供するかといった手順は、むしろ簡単でシンプルであり、研究者はデータキュレーションとセットアップに関わる情報を求められます。これによって、研究者の、他の人にとって価値のあるものをつくりだしているという意識を高め、データを研究者のラップトップから引き出し、機関リポジトリや国家・各専門分野のデータベースに取り込むのです。この DMP は非常に重要で、データを共有するための必要条件です。



(図 8)



(図 9)

データと出版物のオープンアクセスに関する欧州委員会の政策活動をご紹介します。私たちは政治声明、ポリシー、ハイレベル専門家グループ、ステークホルダーから成るプラットフォームを有しており、これらを補助金交付規程に反映しています。

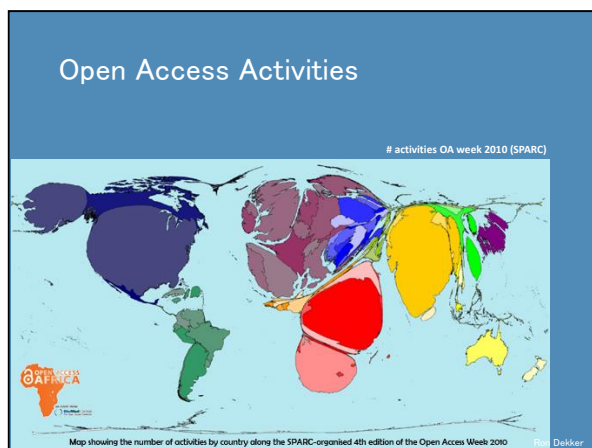
現在、ほとんどの国がオープンアクセスに関する政策を導入しています。実際に G7 では昨年、声明を発表しました。それは同時に G20 の目的でもあります。オープンサイエンスは、今度欧州で開催される G7 会議の議題になる予定で、報酬制度とクラウドインフラに焦点が当てられます。この議題は、日本と欧州委員会が共同で作成したものです。

学術出版物のオープンアクセス

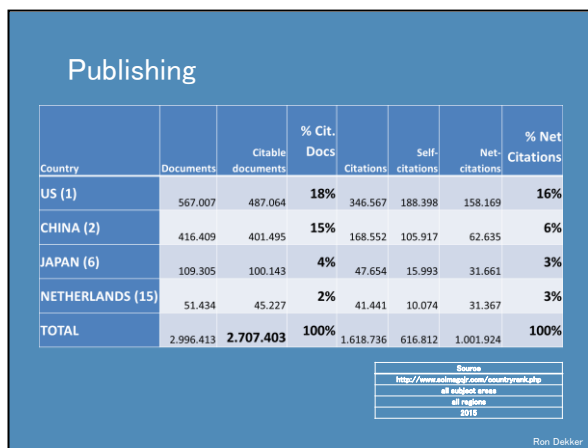
出版物とデータについてに移ります。誰がオープンアクセス政策に関与しているかを考えるため、まず私

は SPARC による OA Week 2010 の図 10 を使って、オープンアクセス政策が活発な国を紹介します。オープンアクセスは多くの国々で議題になっています。しかし、2050 年に予測される人口動向を見ると、アジア諸国への大きな移動があります（図 11）。人々が住んでいるところは、すなわち、科学、イノベーション、経済成長の可能性があるところです。アジアで、科学を社会にもっとうまくつなげることができれば、科学の枠を超えて大きな可能性があることは明らかです。

出版に話を戻すと、毎年約 300 万件の出版物があります（図 12・13）。90%が引用可能で、その約半数が引用されており、自己引用を差し引くと、270 万件の論文につき約 100 万件が引用されていることになります。つまり、3 件のうち 1 件のみ引用されているので、インパクトファクターは 0.3 です。ところが、論文によってはインパクトファクターが 30 も 40 も付与



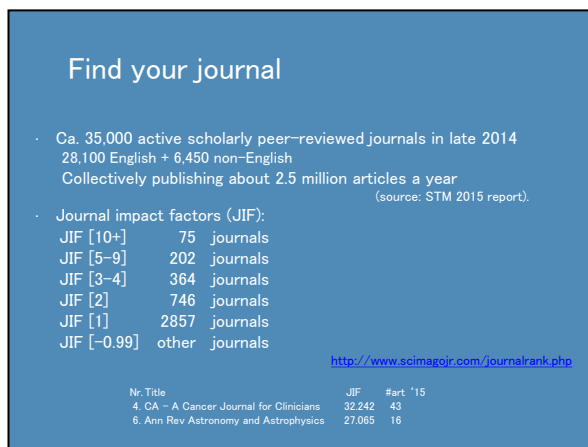
(図 10)



(図 12)



(図 11)



(図 13)

されます。これはつまり、多くの人々が引用を全くされていないということを意味します。これらの論文は誰のために出版されているのでしょうか。誰とこの知識を共有しているのでしょうか。データを見ると、引用率がオランダと日本で同じぐらいであることが分かります。

3万5千種のジャーナルがあるので、ジャーナルを見つけることは難しいでしょう。さらに、これら3万5千種のうち1%だけが5以上のインパクトファクターを持ち、3~4のインパクトファクターを持つものも1%です。オランダでは、研究者は2以上のインパクトファクターを持つジャーナルに掲載することが推奨されていますが、これは全研究ジャーナルのわずか4%にすぎません。

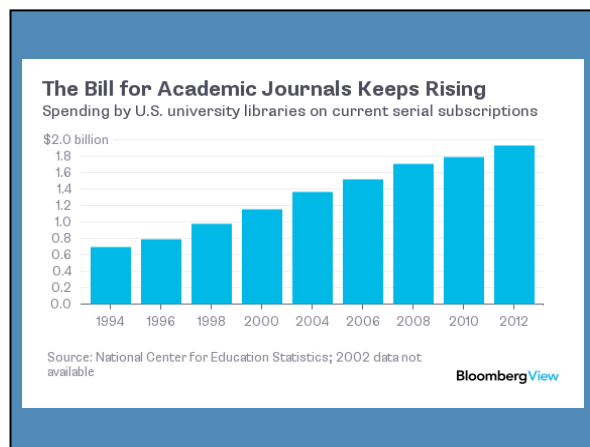
その関連性を述べると、インパクトの高いジャーナルに掲載したい場合はレビューを書かなければなりません。これが最も多く引用されるからです。しかし、天文学と物理学を例に取ってみると、2015年にアクセプトされた論文は16件しかないのです。その16件の中の一つでなければならないということです。この中に入れば論文は引用されるのですが、これが科学でしょうか。私たちは論文をこのようなジャーナルの一つに掲載するために、公的資金を費やしているのでしょうか。私には、これらのジャーナルに掲載されること自体が目標になってしまっているように思えてなりません。

出版社は、IPアドレスの助けを借りてマーケットを細分化し、インターネットを最大限に活用しています。多くのジャーナルを購読できることが魅力のパッケージ購読がありますが、これにより、一つの購読をやめることが、全てのジャーナルの購読をやめることになるため、難しくなっています。ジャーナルのインパクトファクターはまた、外に出ることが非常に難しい経済的ロックインをもたらします。ですから、他のステークホルダーらはジャーナルのインパクトファクターを品質の代わりとして使用するのです。これは、リサーチカウンスルで助成金を決定する際に関係しま

す。大学のランキングを決めるときにも関係します。ここには先行者不利益があります。そのシステムから出ることを決めたら、もはや大学のランキングに入ることはできないでしょう。それは政治的または金銭的不利益をもたらすかもしれません。

従って、有意で肯定的な結果を得なければならないために、リスクの高い研究を行わず、権威あるジャーナルに掲載することにインセンティブが働きます。有意ではなく否定的な結果しか出なければ、トップレベルのジャーナルには掲載されません。これは出版パイアスにつながります。一例として、臨床試験では、公表されやすいのはポジティブな結果が出た試験ですが、インターネットで検索すれば、ネガティブな結果が出た試験や、有意でなかった結果が多く見られます。これでは私たちは情報のほんの一部を取得するだけになってしまうので、この報酬制度を変える必要があります。一方、現在のシステムは研究者と出版社には利益をもたらしています。研究者は出版社に著作権を与え、出版社はそれによってビジネスを行う機会を得て、優れた論文を提供してジャーナルの価値を高めています。出版社は研究者の評判を高め、その評判は研究者が大学内で助成金や威信を得るのに役立ちます。

これは win-win であると言えるかもしれませんが、誰かがお金を支払う必要があります（図 14）。各誌の価格上昇やジャーナルの数の増加によって、上昇し続けるジャーナルのコストを払わなければならないのは、図書館や大学です。このことは他のモデルを考えてみ



(図 14)

る理由の一つです。

私たちはなぜオープンアクセスに移行したいのでしょうか。第一に、インターネット技術を利用すれば、コピーや配布は簡単で安価です。また、資金提供者はより多くの利益を望んでいます。それは、出版においてのみ言っているのではなく、知識や成果を共有するということも意味しています。そのインパクトは、科学的なインパクトだけでなく、社会やイノベーションに対するインパクトまで拡大しなければならないのです。

重要な原則の一つが、公的資金を受けた研究の結果は公的に利用可能であるべきだということです。資金提供に値する研究は分かち合う価値があるはずです。科学を社会とつなげなければ、出版物や成果物へのアクセスを改善する必要があります。人々は今、専門分野内の成果ばかりに焦点を当てているので、アクセスの改善は科学に役立つかもしれません。エルゼビアのSTM Digestは、各専門分野で何が起きているかを説明するためのものです。これは一般の人が使用するものではなく、研究者が他の分野について学ぶために使用されています。従って、これは科学が前進する助けとなっています。私たちは、科学の公正性と信頼の向上に取り組むことができるのです。

ゴールド、グリーン、ハイブリッドなどさまざまなバリエーションのオープンアクセスがあります。しかし、オープンサイエンスには多くの誤解があります。公表する義務はありません。それは特許と相反するものではありません。特許を取りたいのであれば、伝統的ジャーナルであろうとオープンジャーナルであろうと、まず特許を取ってから出版する必要があります。また、ピアレビューのプロセスにも相違ありません。どちらにもピアレビューがあります。オープンアクセスには悪質なジャーナルもありますが、それは伝統的なシステムでも同じです。一部のジャーナルでは既存の論文をコピーして購読モデルにすることさえあります。

つまり、われわれがオープンアクセスを望むなら、

政策が必要です。お伝えしたとおり、ほとんどの国は既にオープンアクセスに関する政策を持っており、多くはグリーンに関してですが、一部はゴールドに関するものです。恐らく最も見事な政策はゲイツ財団のものだと思います。それは1ページで書かれ、五つの項目からなります。オープンアクセスで直ちに公開し、データを共有し、ゲイツ財団がコストをカバーして、公正な価格を支払うというものです。この方針には例外がなく、2年間浸透に費やした後、発効しています。従って、出版物のオープンアクセスについては、私たちはどの方向性に進むべきかもう分かっているのです。

研究データの最適な再利用

研究データの最適な再利用については、さらに少し複雑になります。少なくとも欧州では、出版は一つの出版モデルから新しい出版モデルへの移行期間にあります。

研究データは項目の組み合わせです。欧州委員会には三つの柱があります。一つ目はインフラを管理し、コンテンツと連結する欧州オープンサイエンスクラウド（EOSC）です。研究インフラ欧州戦略フォーラム（ESFRI）は大量のデータをつくる大規模な研究インフラであり、コンテンツの一部です。二つ目の柱は、高性能コンピューティング（HPC）ネットワークングソフトウェアを提供する欧州データインフラストラクチャー（EDI）です。これは、大規模な欧州HPC、ネットワークング、ソフトウェアを持ち、これらのサービスと活動を組み合わせて一つのサービスに集約するものです。つまり、サービスとしてのインフラです。研究者として、私はストレージと計算機能が必要です。同僚とつながる必要もあるので、これらはサービスとして提供されて然るべきです。三つ目の柱はアクセスを拡大することで、それには中小企業、産業界、政府が含まれます。

EOSCに焦点を当てると、委員会の前議長はかつて、これは欧州のものではなく国際的なものであると言いました。一部のデータはクローズドなので完全なオー

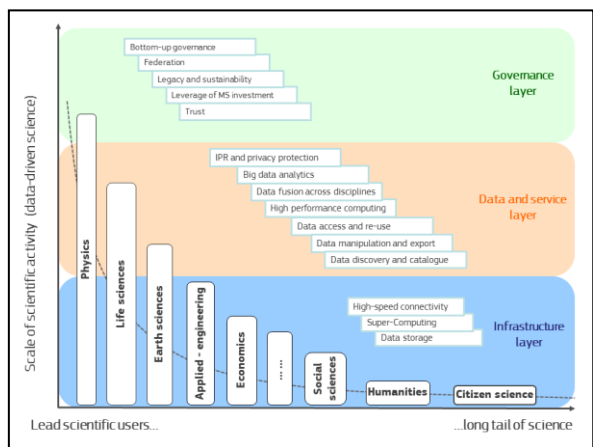
ブンではありません。また、科学のみならず、イノベーション、公共的なものでもあります。これはクラウドでありながら、現実のものであり、設備を提供する地上のデータサービスです。研究者がデータを保存、管理、再利用するための環境をつくり出しています。私たちは、大学や研究所に、そして国家的または欧州レベルでスーパーコンピュータとともに多くのインフラが既に存在していることを認識しています。課題は、既存および新しいインフラといかに連携するかということです。ガバナンスの問題とこれらの施設との連携については、昨年 10 月に発表された EOSC レポートに記載されています。

項目の一つを見ると、図 15 はガバナンスの設定、データまたはコンテンツと新しいサービスの提供、またはインフラの稼働についてのものです。これはこのクラウドの挑戦です。オランダの DANS 研究所の試みは、データに対する一種のミシュランガイドを提供することです。私たちは、FAIR 原則の最初の三つの部分（探索可能・アクセス可能・相互運用可能）のそれぞれに対して五つの質問をし、成績をつけることができます。平均を取ることで、一般的な成績が分かります。これはデータを知らせる最初の方法です。私は、課題の一つはデータを発見可能にすることだと思います。たくさんのデータの中から、どうすれば見つけることができるのでしょうか。

私はクラウドは既に存在していると強調したいと思います。例えば、アメリカ国立衛生研究所（NIH）の

NIH Commons においてです。アメリカ国立科学財団にもクラウドがあり、マイクロソフトとアマゾンには商業クラウドがあります。私が印象的に思うのは、ほとんどがアメリカの例であるということです。彼らはまず、NIH が運営している 600 万ドルのクラウドのようなパイロットから始めています。欧州ではもっとトップダウンで行っています。私たちはガバナンスについて考え、始める前に全てを整えたいと思っています。しかし、このアプローチを取れば、日本には既にバックボーンと設備があるため、有利なスタートをすることができます。アメリカがクラウドに投資し、アメリカの研究がクラウドに投資すれば、国家レベルの投資になるでしょう。欧州がクラウドに投資したいと思ってプロバイダに進めば、国際的な投資になるでしょう。また、これらのサービスを提供するために国内または欧州のプロバイダを調整し準備する必要があります。このために、このパイロットが必要なのです。

先ほど述べたように、データは見つけることができなければなりません、再利用するためにはデータは信頼できるものでなければなりません。DANS 研究所と日本の World Data System（WDS）による「Data Seal of Approval」は、データが良好なフォーマットで見つけ出され、信頼できるものであるという保証を与えるツールを提供します（図 16）。また、データセットの識別子も提供します。データを作成した研究者にクレジットを与えたいならデータを参照できるようになっていることが大切です。



(図 15)



(図 16)

しかし、データを共有し最適に再利用するには、共有を推進するための信頼を勝ち取る文化的な変化が必要です。このためにはデータを参照できる必要があるため、識別子を利用すべきです。しかし、共有するデータを生産する者に対して報酬やインセンティブも与えなければなりません。また、データ生産者とユーザーの認証も必要です。理想としては、データを提供する、使用する際に自分が何者かを知らせるために、データにアクセスするためのある種のシングルサインオンがあったらいいでしょう。特に生命科学や社会科学では、機密データをどう扱うかに注意を払わなければなりません。一部のデータはインターネット上で公開することはできませんが、それでも研究には関係してきます。われわれはそれにどのように取り組んでいけばいいのでしょうか。データインフラについて結論付けると、その方向性はあまり明確ではありません。私たちは、共有のためのインセンティブ、報酬の仕組み、識別子が必要であることは分かっていますが、これらをどう実現するかはまだ明確になっていません。

オープンサイエンスの展望

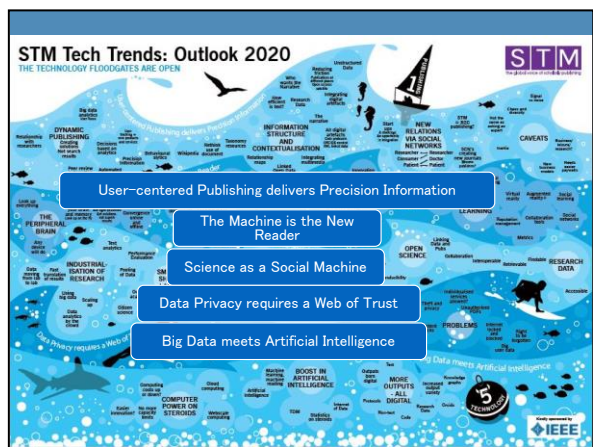
オープンサイエンスに関するビジョンについては、まず出版物から話を始めて、次にデータに移りたいと思います。国際 STM 出版社協会が出した「STM Tech Trends 2015」では、「The Article in a Hub and Spoke Model (ハブ&スポークモデルにおける論文)」と「Data as first class Research Object (第一級研究対象と

してのデータ)」とあります。データは研究成果であって、共有する必要があるということになったのです。

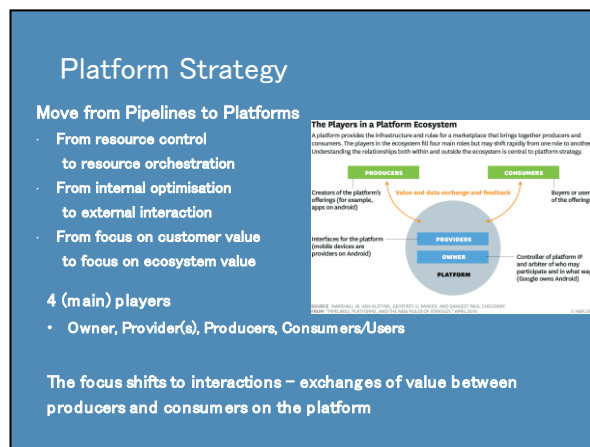
「STM Tech Trends: Outlook2020」は、ビッグデータと、データと人工知能の組み合わせに関するものです(図 17)。機械は、論文やデータを読む新しい読者になるかもしれないし、研究者を助けることになるかもしれません。そのためには、正確な情報を提供することが非常に重要です。私は 1 年間に発行される 300 万の論文の中から 300~400 件の論文しか読むことができないので、選択の際に何らかの助けが必要です。以前は 1~2 誌だけしか読めなかったものが、たくさんの雑誌が出てきました。関連論文が他誌に出たりすることもあり、フィルタリングが必要です。

出版物に追加された価値はフィルタリング機能です。例えば、私の 15 歳の息子はコンテンツに対してお金を支払おうとはしませんが、インターネットで音楽や映画を見つけるときには助けが欲しいかもしれません。そのために彼はソーシャルネットワークを利用します。一方、研究では、関連情報を見つけるのに助けが必要です。出版物において、読者は関連する論文を探したいと思っていますが、著者はできるだけ多くの読者にアプローチしたいと考えるのは明かです。出版物の供給と需要という市場があるのです。

新しい方向は、プラットフォーム戦略を持つことです。図 18 は、戦略を持ちたければ異なった考えをしなければならぬと言う「ハーバード・ビジネスレビュー」の論文に基づいています。私たちは焦点をシフ



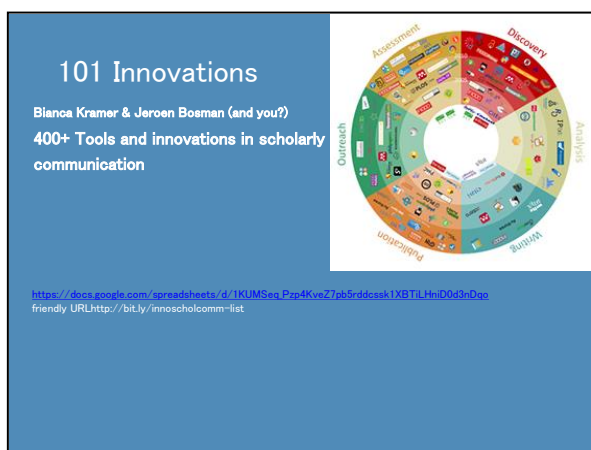
(図 17)



(図 18)

トしています。それは限界コストに関するものではなく、プラットフォーム上の生産者とユーザー間の相互作用と価値交換に関するものです。これらのプラットフォームには、所有者、プロバイダ、生産者、ユーザーがいます。一つの大きな例は Apple です。彼らは iPhone と iPad というプラットフォームを所有していて、これは電話のためでなく、そこでしか使えないアプリケーションを用意しています。ユーザーを見つける App Store と iTunes のことです。Apple の成功は、このつながりをつくったことにあります。

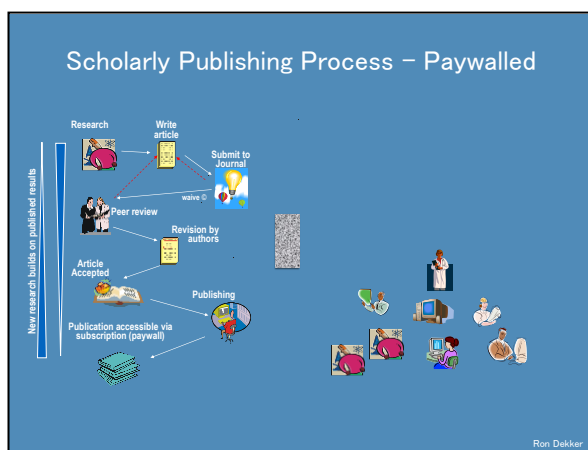
もう一つの例はソニーの PlayStation です。PlayStation 4 では、月会費を払えば、オンラインで世界中の他のプレイヤーとプレイできます。このように、プラットフォーム上の付加価値は重要です。私は、ユーザーと生産者が良好なつながりを持つことが、出版物の方向であるべきだと思います。



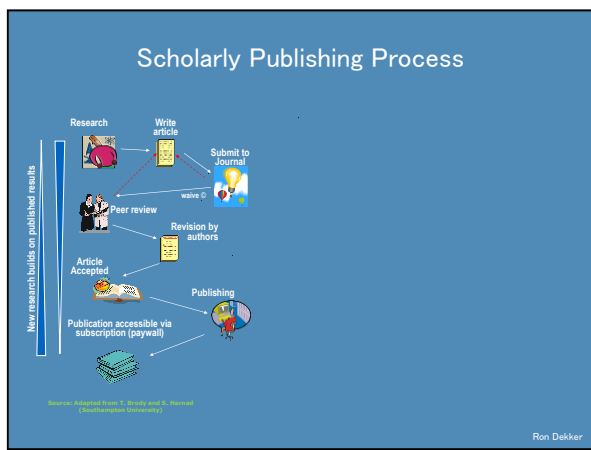
(図 19)

第 3 回 SPARC Japan セミナー2015 で、図 19 を見たことがあると思います。「101 Innovations」は、利用可能な全てのツールを記録したのですが、現在はオープンサイエンスを行うときに利用可能なツールの数であるため、「400+ Innovations」と呼ばれるべきです。

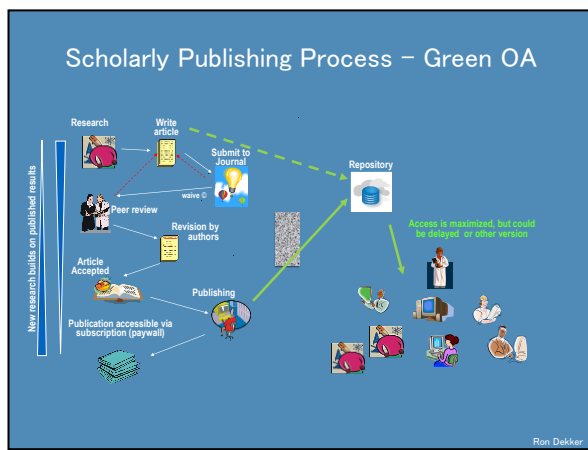
図 20 は伝統的な出版プロセスを示しています。論文を書く、提出する、却下される、別のジャーナルに再度提出する、査読される、拒否されるか改善する、論文が受け入れられる、発表する、これがジャーナルに掲載されるということです。ペイウォールモデルでは、科学界の外の人々は見ることができず、購読していなければ論文にアクセスするために 30 ドル支払わなければなりません (図 21)。グリーンオープンアクセスでは、エンバゴ方式とプレプリント方式によって利用可能です (図 22)。従って、私たちはプロセスの最後を変えるゴールドオープンアクセスを考え出し



(図 21)



(図 20)



(図 22)

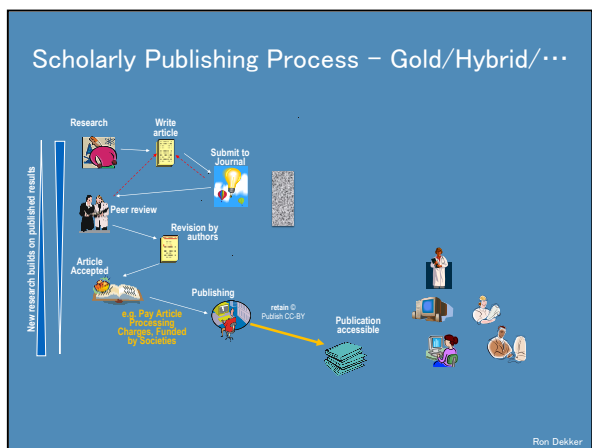
ました（図 23）。私たちが支払いをし、誰でも利用できるようにします。このゴールドモデルは標準になるのでしょうか、それとも他のモデルが出るのでしょうか。

出版ではウェルカムトラストによる新しいアイデアがあります（図 24）。これは「Open Research」と呼ばれ、論文執筆後 1 週間以内に論文が発表されるものです。その後、論文が公開の場で査読され、結果を利用できるようになります。また、これは出版物に加えて検査結果やデータにも適用されます。どのようにしてデータが調査されるかは分かりませんが、これによって、データセットの品質を保証する新しい方法が見つけれられると思います。最初に出版し、次に品質評価を行うというように、物事の順番が変わってきています。さらに進むと、研究を再び途中の段階に戻すことができます（図 25）。研究して、それを公開するかどうかを決定します。公開する場合、伝統的なジャーナル、

ゴールドオープンアクセス、またはオープンリサーチに送ることができます。私の場合は査読を選ばないで、プレプリントで投稿します。そうして、アクセスを得ることができます。これは研究者にとって非常に簡単な方法です。

この方法は、オープンアクセスをさらに容易にするため、資金提供者にとってもとても簡単です（図 26）。このモデルは、約 750 ドルという FAIR な価格で査読を受けられます。これは新しい方法かもしれませんが、いまだ既存のジャーナルシステムの中にあります。

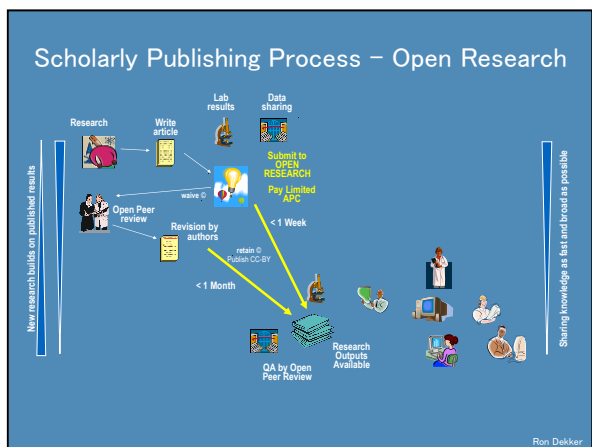
別の動きはプレプリントの中に見いだすことができます。資金提供者が識別子を取得するプレプリントを喜んで認め、プレプリントに関する助成金を容認すると言えば、若手研究者にとって助けとなります。2 年間の博士研究員の助成金を受けている場合、発表されるまでに 1 年かかる論文を提出する時間がありません。



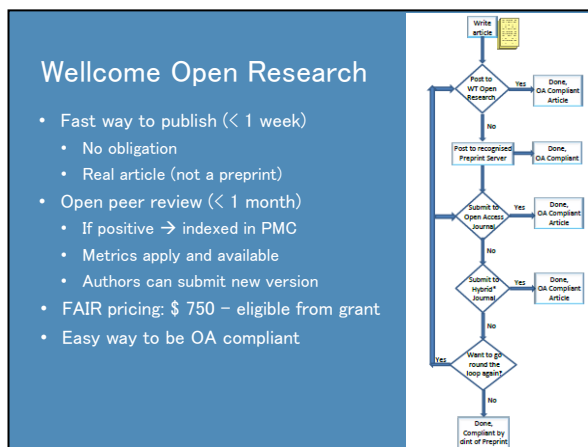
(図 23)



(図 25)



(図 24)



(図 26)

このプレプリントでタイムスタンプを取得した場合、助成金に対する新しい申請でそれを使用することができます。これによって出版や知識の共有もスピードアップします。私たちはプレプリント用の信頼できるリポジトリのシステムを持つことになるでしょう。このように、出版における知識共有の動きを追うことで、将来の方向性を見いだすことができます。

データに関しては、皆さんの力が必要です（図 27）。データを再構成するために必要な労力を考慮しても、データには莫大な価値があります。研究者がデータを準備し共有することに労力を割けば、他の誰かが、論文や新しい知識を生産するのに多くの時間を節約して収入を得ることになります。従って、データ共有に関する市場をどのように設定したらいいかというのが大きな疑問の一つになります。私たちにはインフラがあるので、パイロットプログラムも行うことができます。私たちは、全ての学問分野の国家政策を立てるべきか、国際的にそれぞれの学問分野ごとに政策を立てるべきか、あるいはその両方を行うべきかということについて議論することもできます。

市場を創出するに当たって重要な win-win の一つの例は、Structural Genomics Consortium (SGC) です（図 28）。SGC は、創薬研究に関する大規模な官民連携による研究団体であり、創薬ターゲットとなるヒトのタンパク質の 3 次元構造を同定・解析する研究を行っています。製薬企業は、競合同士であっても、対立するそのときまではこのコンソーシアムで協働しています。



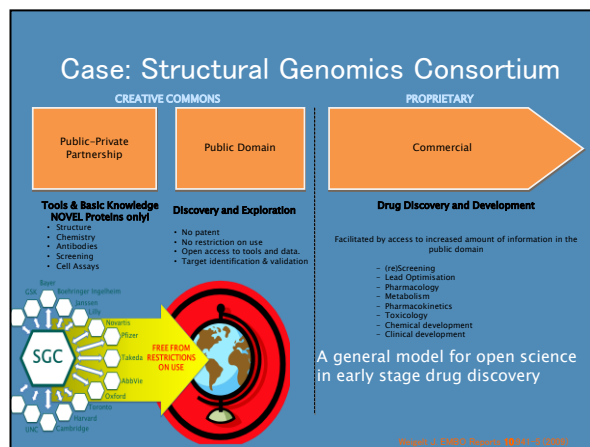
(図 27)

このコンソーシアムでは、信頼できる組織を設立し、10 の研究グループを持っています。組織はそこに調査したいと考えている上位 3 種のタンパク質を伝えます。その中から、興味を持っている参加団体が多かったトップ 30 のタンパク質について、どの団体がそれに興味を持っているか、それについてどのような取り組みをしているかは明らかにせずに研究しています。参加団体はこれに同意しています。これは win-win です。多くのデータと研究能力を持っているのですから、需要と供給、生産者とユーザーをマッチさせることが、データに対する課題です。

膨大なデータの中から、価値あるデータを探すことができるようにしなければなりません。これが基本です。図書館は、（データと研究者の）仲介役を果たすことができるかもしれません。図書館同士はつながっているため、他の図書館で何が起きているかを知っています。データに関する質問がある場合、図書館員は手助けすることができます。

結論として、私たちはオープンサイエンスの新しい時代の一部とならなければなりません。私たちはより多くの人々にアプローチし、科学だけでなく社会にも大きな影響を与えたいと考えています。それは二度手間や出版バイアスを防ぐことができます。科学、社会、イノベーションに影響を与えるためには、将来使うためにデータを保存する必要があります。

課題は、いかに協力し、イノベーションを生み出し、共有するかということです。私は日本に來られてとて



(図 28)

もううれしいです。私が日本を知るようになったきっかけは、伝統的な映画「檜山節考」、伝統的な企業でどのようにイノベーションを起こすかについて説明した『知識創造企業』という本です。私たちをオープンサイエンス、データの再利用、出版物のオープンアクセスに導いてくれるのは、伝統の融合、過去を理解、イノベーションと変革への勇気だと考えています。最後に、南アフリカ共和国のことわざをご紹介します。終わります。「早く行きたいなら、一人で行きなさい。遠くに行きたいなら、一緒に行きなさい」。

● **深貝** 横浜国立大学の深貝と申します。SPARC Japan の運営委員をしています。オープンサイエンスについて、このところ日本でも非常に活発に議論しています。歴史を振り返れば、近代の入り口の科学革命のころ、グーテンベルクの活版印刷を通じて、こもっていた局所的な知識がオープンになって、そして科学者が振る舞い方を変えていきました。それと似たようなことが今起きはじめています。ネットワークを使って、今までの印刷物ではない情報が飛び交うようになった中で、人々の振る舞い方が変わってきているのです。

電子ジャーナルが出たというのは、単に印刷物が電子版で出ただけですから、ある意味で伝統的です。しかし、最近では、人々がネットワークを使って、知識を交換し合っています。知識を交換することによって、多くの人々が知識の生産に関わり、アイデアがひらめくということになりつつあるのです。それを一層加速しようというのがオープンサイエンスだと思います。

ただ、その場合のバリアは、学術の成果が誰かに帰属し、その人がインパクトファクターによって評価を受けるという方向に加速していることです。そうすると、科学者・研究者は、高い評価を受けるために、知識を狭めた上で成果を挙げることに走りがちです。しかし、広い目で見ると人間の知識を多く豊かなものにしていくためには、オープンにしていっていい方がいい。そ

のような揺らぎの中にあるわけです。

その中で、公的資金を使って得られた成果をオープンにするということを今やっていますが、それだけで十分なのかどうか、お考えをお示しいただけるとありがたいです。

● **Dekker** 現在は、印刷物をインターネットで利用できるようにするのではなく、そもそも成果を最初からデジタルで出すようにしなければならない、そういうパラダイム転換の最中にあると言えます。私はウェルカムトラストの「Wellcome Open Research」が好きなのですが、その理由は伝統的な論文と中間結果の両方に焦点を当てているからです。そこでは重要なデータを公開し共有したいと思えば、することができます。ジャーナル「Science Matters」では、研究コミュニティにとって重要であると思われる場合は、中間結果を公表することができます。私たちは、伝統的な出版を越えて、成果の新たな発信方法を見つける必要があるのです。

これらの成果を測定する際は、ジャーナルのインパクトファクターを超えた新たな指標が必要になります。資金提供者は、代替物や新しい成果の発信方法に報酬を与えるべきです。イノベーションまたは社会的議論に重要な貢献をした場合も、資金提供者や大学によって報酬が支払われるべきです。研究者に対する独自の報酬システムを持つ大学や医療機関もあります。資金提供者は、出版物の数で評価することをやめるよう、所属する委員会のメンバーに求めるべきです。代替案としては、意味のないインパクトファクターを積んでいくのではなく、査読者がトップ3またはトップ5の出版物を精査することが挙げられます。異なるものを比較しているのですから、分野を超えてインパクトファクターを比較することはできません。DORA イニシアチブは、研究を評価するために定性的な情報を使用すると述べています。それが進むべき道であり、また、私たちは、プレプリントや他の公開方法の価値を認めるべきです。

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

ディープラーニングとオープンサイエンス ～研究の爆速化が引き起こす摩擦なき情報流通へのシフト～

北本 朝展

(国立情報学研究所)

講演要旨



講演者はこれまでオープンサイエンスの意義を利便性、透明性、参加という3軸から考えてきたが、これだけでは科学研究にオープン化がどう貢献するかを理解しづらいという問題があった。そこで「スピード」という新たな軸を導入することで、オープンサイエンスの意義を再考してみたい。その背景にある仮説は、研究のスピードが極限まで高速化すると、情報流通もそれに追従して高速化せねばならないため、情報流通の妨げとなる「摩擦」を取り除く方向に進化して、結果的に科学研究がオープン化するというものである。こうした仮説を検証するのに最適なフィールドが、人工知能および機械学習の一分野である「ディープラーニング（深層学習）」である。この分野ではいったい何が起きているのか？この例外的な研究分野から得られる教訓がどれほど一般化できるかは未知数であるが、これまでの動向を分析することで、オープンサイエンスの一つの可能性を先取りしてみたい。



北本 朝展

1997年東京大学工学系研究科電子工学専攻修了。博士（工学）。現在、国立情報学研究所コンテンツ科学研究系准教授、総合研究大学院大学情報学専攻准教授、情報・システム研究機構人文科学オープンデータ共同利用センター準備室長。画像データの分析を中心に、データ駆動型サイエンスを人文科学や地球科学、防災等の分野で幅広く展開する。文化庁メディア芸術祭アート部門審査委員会推薦作品、山下記念研究賞などを受賞。オープンサイエンスの展開に向けた、オープン化や超学際的研究コラボレーションにも興味を持つ。

私の研究分野は情報学で、画像処理や画像データベースの研究をしていたのですが、その後、データ駆動型サイエンスに発展し、気象情報、地球環境情報、人文科学情報などの分野で研究を行うようになりました。この流れで、最近はオープンサイエンスの問題にも関わっています。

1. オープンサイエンスの背景

1-1. オープンサイエンスへの収束

オープンサイエンスとは、「オープン」という言葉をてこにして、サイエンス（研究）の方向を変えるも

のだと考えています。共通する部分は、今よりもよりオープンにしようという方向性です。いろいろな方向に、いろいろな運動があるのですが、それを「よりオープンに」という一語で束ねるとどういう世界が見えるかという話になると思います。

ただ、個々の活動で「オープンサイエンス」の意味は異なるので、単一の定義は困難です。よく私が言っているのは、オープンサイエンスは同床異夢であるということです。オープンにしたいという思いは共通しているのだけれど、どんなオープンを夢見ているかという認識を共有しないと、実は違う夢を見ていたとい

うことがよくあります。私は、図1にある、オープンアクセス、オープンデータ、データマネジメント、シチズンサイエンス（市民科学）、研究の再現性など、透明性・参加・協働・共有の視点で動いている多様な活動は、全て「オープンサイエンス」というキーワードでまとめられると考えています。

「オープン」には三つの側面があると考えています。一つ目は、他者が使えるということです（再利用）。外部の人が研究成果を自分の目的に再利用できます。これにはオープンデータやオープンアクセスなどが当てはまります。

二つ目は、他者が検証できるということです（透明性）。外部の人がエビデンスを検証し、正当性を判断できます。政府の政策を検証するオープンガバメントも同じような目的を持っていますが、他者の行った研究を再現・検証することも非常に重要なテーマになってきています。

三つ目は、他者を受け入れることです（参加）。研究のプロセスに他者を巻き込むことで、今まで得られなかったような成果を生みます。これにはオープンイノベーションや市民科学が当てはまります。

1-2.制度を分析する四つの視点

オープンサイエンスを分析するときには、法、規範、市場、アーキテクチャの四つの視点があると考えています。この四つの視点は、Lawrence Lessig が書いた“Code: And Other Laws of Cyberspace”（邦題『CODE

ーインターネットの合法・違法・プライバシー』）という本で挙げられていて、Lessig は制度をどうやって変えていくかをこの四つの視点から考えることができると主張しています。

日本語で言うと、法は「しなければならない」、規範は「すべきである」、市場は「した方が利益がある」、アーキテクチャは「せざるを得ない」と言い換えられます。このような道具を使って制度を変えていく方法は、オープンサイエンスでも有効ではないかと考えます。

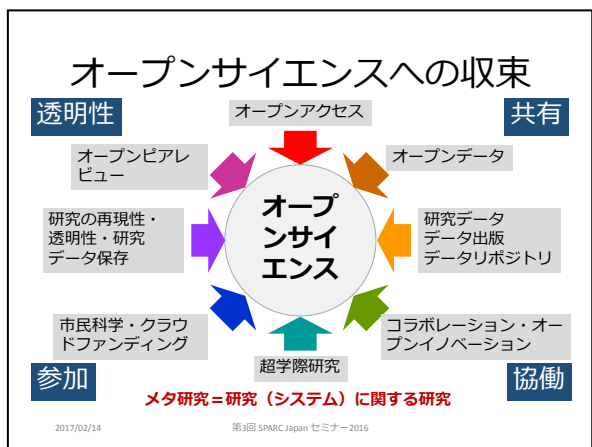
例えば、法によるオープン化は、資金提供機関が決めるルールを変えることによってオープン化を進めていくというものです。データ管理計画なども基本的にはそうで、ルール変更が重要な役割を果たします。例えば、資金提供機関が、どれだけオープン化したかで研究評価するというルールになれば、研究のやり方が変わっていきます。ただ、これは上からの押し付けで副作用も大きいので、本当にそれを適用していいのかという見極めは非常に重要です。ゆっくり慎重に進めていくことが求められます。

規範によるオープン化は、そもそもサイエンスはオープンであるべきであるという考え方です。これは分野によっては一般的な考え方です。例えば、地球科学では地球全体のデータがないと研究できないので、シェアしなければ研究が進まないという状況があります。こういう分野は、データ共有が不可欠なので、オープンな文化が発達したと言えます。

それだけではなく、世代の差もあります。上の世代より若い世代の方がもともとオープンな文化に生まれているので、共有に抵抗がありません。

ただ、こういう規範によるオープン化は、文化圏が違うと同じ文化を共有できないので、全分野で共有することはなかなか難しいです。

市場によるオープン化は、「した方が得だ」というものです。オープン化によって報酬が得られるのであれば、みんなオープン化するだろうということです。研究成果をオープン化すると例えば引用が増加するな



(図1)

らば、オープン化した方が得だからオープン化しようという考えを持つ人も出てくるでしょう。ただ、これは報酬と裏腹に損失への不安があります。他者に成果を横取りされるのではない、報酬は労力に見合うのか、こういう損失への不安があると、よほど報酬が確実でないとなかなかオープン化できなくなります。

アーキテクチャによるオープン化は、Lessig の本でも情報社会特有の方法であると述べられているのですが、何らかのプラットフォームの上で作業すると、必然的にオープンに誘導されてしまうというものです。例えば、何かツールを使っている、ワンクリックでオープン化できる機能があれば、オープン化へと誘導されてしまうかもしれません。プラットフォーム上の見えないルールで、いつの間にか誘導されるというのがこの方法の特徴です。

また、苦痛の軽減という視点も重要です。オープン化は大変だけれど、有償のサービスを使えば簡単にできるとなると、そういうプラットフォームの上に乗ってオープン化を進める状況も当然起こり得ます。こういうものは良くも悪くも企業にとってはベンダーロックインのチャンスであり、そうした利益がオープン化を進めていくという未来もありえます。

このように四つのオープン化は方向が全く異なりますので、オープン化の話題ではどのオープン化に焦点を合わせているのかを明確にすることが有益ではないかと思っています。

1.3. 研究データのオープン化

研究データの問題についても触れたいと思います。私は研究データを三つの種類に分けています。研究資源データ、論文付属データ、研究過程データです。

研究資源データは、研究の入力となるデータです。観測データや、評価用データセットなど、これを使って研究するというものです。これは再利用が目的ですから、再利用しやすいオープンデータになっているかという点が評価基準になります。

論文付属データは、研究の出力となるデータです。

典型的なのは論文のエビデンスとなるデータです。これは、再利用も可能ではありますが、それよりも透明性、すなわち研究が再現できるかの方が重要ですので、再利用とは違った視点でのオープン化が求められます。

研究過程データは、研究の入力と出力の間で生み出されるデータです。例えば、日々の研究活動のエビデンスとなるようなデータです。このデータは透明性のために重要で、研究不正防止のためのデータ保存の義務化などもこれに関連しますが、基本的にオープン化が目的ではなく、クローズドに長期保存し、本当に必要があれば限られた人に対してオープンにするものです。

私が関心を抱いているのは、研究資源データをどうやって再利用のためにオープン化するかということです。このニーズは、データ駆動型研究が広まるにつれてますます高まっています。図 2 は、「Preparing for the Future of Artificial Intelligence」という、2016 年 10 月にホワイトハウスから出たものです。ホワイトハウスのウェブサイトは最近大幅に変わりましたので、今アクセスできるかは分かりません。ここでは、マシンラーニングや AI を発達させるには、データをオープンにすることで研究を活性化させることが重要であると強調されています。この AI とマシンラーニングが今後のオープンデータの一つの焦点になります。

データについて、考えるべきはライフサイクルです。ライフサイクルとは、データが誕生してからたどるプロセスを段階の遷移として表すモデルです。研究者は

データ駆動型研究からの要請

PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

<https://www.whitehouse.gov/blog/2016/10/12/administrations-report-future-artificial-intelligence>

- Recommendation 1: **Private and public institutions are encouraged to examine whether and how they can responsibly leverage AI and machine learning in ways that will benefit society.**
- Recommendation 2: **Federal agencies should prioritize open training data and open data standards in AI.**

2017/02/14

第3回 SPARC Japan セミナー 2016

(図 2)

全体に関わりますが、そのうちの一部にはライブラリアンも関わります。例えば、データ管理計画などはライブラリアンが関わるのが期待される段階です。

例として、Plan、Collect、Assure、Describe、Preserve、Discover、Integrate、Analyze という段階があって、これがぐるぐる回るのがデータライフサイクルです（図3）。個々の段階で何をするかということが重要になるのですが、左のサイクルと右のサイクルをご覧になって分かるように、データライフサイクルはそれぞれ違います。ここには示していませんがもっとたくさんの図がありますので、唯一の正解があるというよりは、プロジェクトに適したライフサイクルは何かを考え、その中で誰がどの部分を担当するかを考えることが重要だと思います。

1-4.ライブラリアンの役割変化

こういう状況の中で、ライブラリアンも役割が変化していくことが考えられます。従来のライブラリアンは本を扱っていました。本はこれ以上加工されない最終の生成物であり、これを利用者（読者）のために整理していればよかったのです。しかし、データ管理計画など、データを扱わなければいけないとなると、データライブラリアンが必要だという話になります。

データは、本のような最終生成物ではなく、変わっていくものです。また、データにどんなメタデータを付けるのかという話になったとき、メタデータに決まったフォーマットがあればいいですが、ない場合は作者

のところに行って、このデータをどのような目的でつくったのかインタビューしながらフォーマットを決めるプロセスが入ります。本で言えば、作者に作品の意図を聞いて、それを記述するというような、編集者や出版者に近い役割を果たさなければなりません。ですから、従来の最終生成物だけ扱ってればよかった時代と、役割もかなり変わってくるのです。

ライブラリアンは、データに標準的な価値を与えて使えるようにすることが目的ですが、一方でキュレーターは、データに付加価値を付けることが仕事です。データライブラリアンが整理した段階で、標準的なメタデータが付いて誰でも使えるようにはなりますが、そこにさらにデータキュレーターが関わり、このデータにはどんな価値があって、どういう目的に使えるかを売り込む。そんな人がいると、データの利活用も進むでしょう。このようにライブラリアンの役割変化が今後進んでいくと考えています。

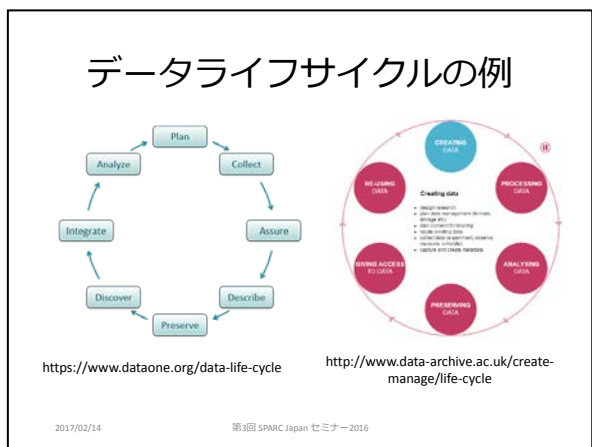
現段階でデータライブラリアンに求められるスキルの一つ目は、データライフサイクルに関する知識と、DOIなどの識別子を与えてデータをアクセス可能にすることです。

二つ目は、メタデータを付けないとデータにアクセスできないので、メタデータ規格や語彙に関する国際標準との相互運用性などに関する知識です。

三つ目は、データは置いておくだけでは使われないので、リポジトリに入れるなど、データの共有や利活用を促進する方法に関する基本的な動向の把握です。

四つ目に、上級編として、研究者コミュニティとの議論やエスノグラフィ的観察に基づき、データ管理システムを設計し構築する能力です。これができるデータライブラリアンは大変ありがたいと思います。

ただ、10年後を考えると、こういう仕事のかなりの部分はAIに代替される可能性があります。今からデータライブラリアンになろうという方は、10年後のキャリアを目指しますから、そのときまでに代替される知識を学んでしまったら、人生戦略上、非常にまずいです。ですから、10年後どうなるかを予想しな



(図3)

がらデータライブラリアンに必要なスキルを考えなければいけません。

メタデータの付与についても、例えば、アクセスが目的であれば、項目や品質の考え方も変わりますし、AI 向けメタデータであれば、人間が付与するかどうかも分かりません。ですから、AI との関わり方は必ず考慮すべきポイントになります。この場合、データライブラリアンには二つの道があると思います。一つは、AI を使いこなす側に立って、データライブラリアン兼データサイエンティストというスーパーマンになる道です。もう一つは、AI の下で働く道です。現実世界のダークデータが AI が使いやすいクリーンデータに整理して提供する。スーパーマンになるのもいいですが、下で働くというのも必ずしも悪いことではありません。今の世の中にも実際にはそうなっていて、Google のためにデータをきれいにしてあげるとするのは、みんながやっていることです。ですから、データライブラリアンになるのであれば、10 年後も維持できる価値を考えなければいけないというのが前半部分のお話です。

2.オープンサイエンス再考

AI が出てきましたが、後半は AI の研究の世界で起きていることの話です。

2.1.オープン化の第四の軸「スピード」

これまでオープンサイエンスの意義は、「再利用できる」「透明性がある」「参加できる」といった三つを考えてきました。ただ、これだけだと研究者にはあまり実感がなく、これだけではオープン化が進まないという感じがありました。そこで、オープン化の意義として、「スピード」があるのではないかと今日をお話したいと思っています。研究のスピードが極限まで高速化すると、情報流通もそれに追従して高速化しなければいけないので、情報流通の妨げとなるような摩擦を取り除く方向に情報流通プラットフォームが進化し、結果的に科学研究がオープン化するという

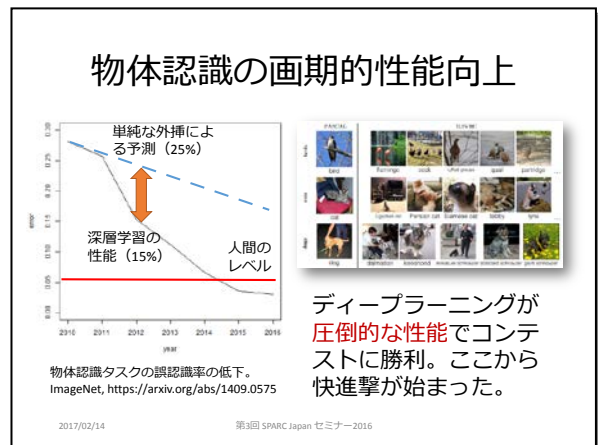
ような仮説です。その例をディープラーニング（深層学習）という分野の現状から検証したいと思います。

2.2.ディープラーニング登場

皆さん、ディープラーニングはお聞きになったことがありますか。AI という言葉はニュースで日々お聞きになっていると思います。今、第3次人工知能ブームと人工知能学会会長も言っていますが、第3次人工知能ブームの中心的存在がこのディープラーニングで、技術的には「ニューラルネットワーク」と呼ばれるものです。ニューラルネットワークの中で特に層が多いものが「ディープ」と呼ばれます。原理は昔から知られていたのですが、ビッグデータとアルゴリズムの改良で画期的な性能向上があり、それが衝撃を与えて人工知能ブームにつながっています。

画期的な性能向上は、まず画像認識の分野から始まりました（図4）。画像だけを与えて、それが何かを認識する物体認識タスクのコンペティション「ImageNet Large Scale Visual Recognition Challenge (ILSVRC)」が2010年から始まりました。緩やかなスピードで誤認識率が改善していくと多くの人が思っていたところ、2012年には誤認識率がいきなり25%から15%まで下がり、ディープラーニングが圧倒的な性能で優勝しました。みんながその結果に驚き、その後は一斉にディープラーニングに走り、今は人間よりも性能が高くなっています。

ここからいろいろな分野にディープラーニングの適



(図4)

用が始まったのですが、私が一番好きなのは囲碁のソフトウェアの AlphaGo です（図 5）。人間のチャンピオンも AlphaGo に負けました。過去のデータはもちろん学んでいますが、自己対戦によって戦略を深化させ、人間とは異なる戦略を用いて勝利したことが非常に衝撃的でした。

私も感動して、AlphaGo 観戦ツイートをまとめて Togetter ページをつくりました。私は囲碁というゲームがプレイできますので、この勝利がどう衝撃的かということも具体的に分かりましたし、この勝利は今まではちょっと違うと感じました。この AlphaGo をつくっている会社が DeepMind です。この会社、今は Google の傘下にありますが、何のプロダクトも公表されていない時点で数百億円相当の金額で買収されたとして話題になりました。この会社が今、ディープラーニングでは重要なキープレイヤーになっています。

まずディープラーニングで面白いのは、オープンソース競争が始まっているという点です。ディープラーニングの最先端のライブラリを各社が競ってオープンソース化しています。Google の TensorFlow、Microsoft の CNTK、Facebook の Torch、他には日本でも Preferred Networks という会社が Chainer というライブラリを公開して頑張っています。

本当に最先端のライブラリは知的財産の塊なのに、それがどんどんオープンソース化されているのはなぜでしょうか。やはりオープンソースに人がおびき寄せられることを狙っているのでしょうか。オープンソース

のコミュニティができ、協力者が増えれば、そこから創出される価値も増えます。例えば、それを使ったライブラリや可視化ソフトなどいろいろなものが出てくれば、全部自社で開発しなくてもよくなります。まさにオープンイノベーションですが、そこでは競争すべき領域と協調すべき領域があります。差別化できる部分は守りつつ、外部の力を使えるところは使うというのが今の状況です。

では、Google がオープンソース化したからといって、みんなが Google と同じことができるかというと、それはありません。ソフトウェアは確かにオープンですが、例えば AlphaGo を実行するのに、数十億円のコンピューター資源が必要になるなら、誰でもできることはありません。また画像認識でも、数億枚の画像データを使わなければいけないとなると、そんなデータを持っているところは限られます。だから実際は同じことはできないのですが、少なくともソフトウェアについてはオープン化がどんどん進んでいます。

このようにオープンソース化が急速に進み、論文の実験コードなども再利用や再現性の観点からオープンソース化され、誰でも試せるようになりました。共通基盤データも、例えば先ほどの物体認識タスクの実験データがオープン化されているので、誰でもソフトウェアをつくれれば性能が比較できます。さらに応用分野がどんどん広がっているのも、他の分野の研究者や一般の人々も大挙参入しています。ですから、一刻も早く成果を公表しないと、自分の成果が古くなってしまうのではないかと、という恐怖を皆さん持っています。スピードへの強烈な圧力がかかるという、他の分野にはない独自の状況が生まれているのです。

AlphaGoの衝撃



<https://deepmind.com/research/alphago/>
 アルファ碁観戦ウェブサイト
<https://togetter.com/li/983741>

- ディープラーニングは、人間とは異なる戦略を用いて、人間のチャンピオンに勝利した。
- 過去データを学ぶだけでなく、自己対戦で戦略を深化させた。
- 開発：DeepMind社（Googleが買収）

2017/02/14
第3回 SPARC Japan セミナー 2016

(図 5)

2-3.arXiv の利用

そこで登場するのが arXiv です (図 6)。なぜ今さら arXiv なのだと疑問を持たれる方もいらっしゃると思います。これは 1991 年登場の元祖プレプリントサーバーで、登場してからもう 25 年以上たっているからです。現在はコーネル大学が運営しています。実は今、この arXiv が大きく変わっているのです。

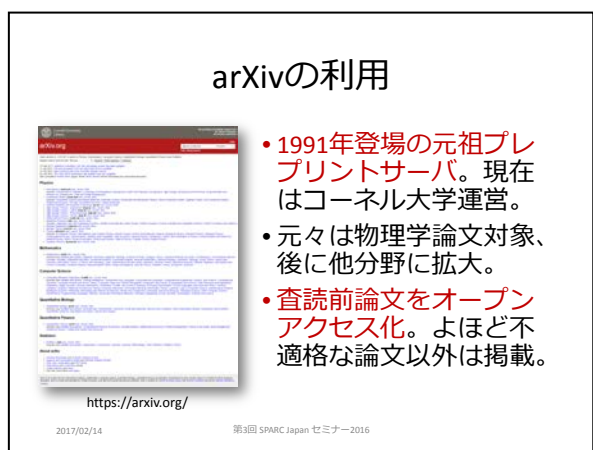
arXiv への投稿数はずっと上昇しています (図 7)。2010 年ぐらいから加速しているように見えます。分野別の投稿状況を見ると、もともと high energy physics が多かったのですが、割合としては減ってきました。math も多いですが、cs (コンピューターサイエンス) が近年になって割合を増やしてきています。これはディープラーニングの盛り上がりと同期した変化です。

ディープラーニングの分野では、この arXiv が主戦場になってきています。AlphaGo をつくった

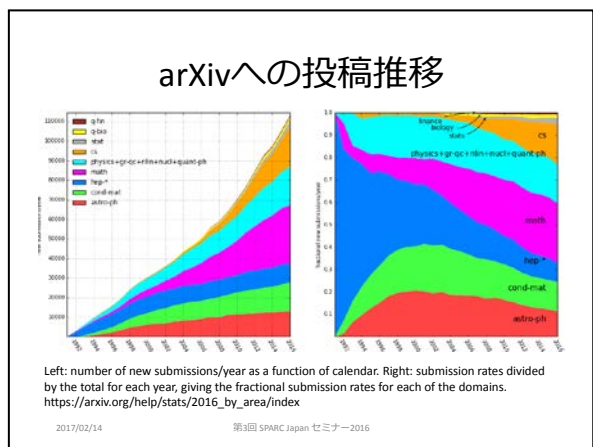
DeepMind のサイトには Publications というページがあり、134 件 (雑誌 10 件、arXiv 系 34 件、著名国際会議 90 件) の出版がリストになっています。雑誌は「Nature」などです。著名国際会議のリストもすごいのですが、そのような国際会議や「Nature」に混ざって、arXiv が同格でリストに並んでいます。別に彼らがそれらを区別している感じはありません。実際、DeepMind が出している論文は、arXiv で終わってしまうものもあります。つまりレビューは受けていない。arXiv にまず最新の成果をどんどん流す文化があり、その中で必要なものだけ査読を受けているような印象を受けます。

その arXiv がどう主戦場かを見てみましょう。例えば図 8 は 2016 年 10 月 10 日に arXiv に投稿された、東京大学の松尾豊先生らの論文です。この論文を、10 月 11 日に arXiv に投稿された論文が引用しています。この論文を書いたのが DeepMind の人なのです。DeepMind の人は arXiv を毎日毎日チェックし、1 日前の論文でも引用して投稿しています。まさに引用の爆速化が起こっています。ネットの世界から見れば、Twitter なら 1 分前でも引用できますので、1 日前の情報を引用すること自体は別に珍しくないと思います。ただ、従来の学術情報出版の基準から見ると、考えられないようなスピードで引用が進んでいると言えます。

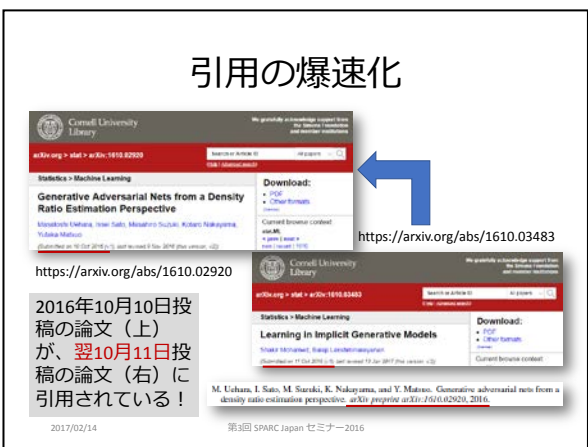
さらに、最近進展している分野なので、2016 年や 2015 年の論文が必然的に多くなります。そうすると、論文のほとんどが arXiv にあるので、arXiv を読んでい



(図 6)



(図 7)



(図 8)

れば基本的な動向は追えることになります。よほど古いものでない限り arXiv を見ていればいいということになれば、過去論文データベースの重要性が低下してきます。

皆さんにとって心配なのは、arXiv に投稿することが二重投稿にならないかという点でしょう（図 9）。二重投稿と arXiv がどう整理されているか。ディープラーニングの中心的な国際会議である Neural Information Processing Systems (NIPS) は、“Prior submissions on arXiv.org are permitted.” と明言しています。他の会議では、International Conference on Machine Learning (ICML) も “e.g., in arXiv” と書いています。一方、Computer Vision and Pattern Recognition (CVPR) という会議は、arXiv.org への投稿は出版ではないからオーケーと言っています。この「出版ではない」というのは微妙な線引きですが、レビューされてないから出版ではないという解釈を取っています。ただ NIPS では数語で済むところ、CVPR は数パラグラフを使って arXiv がなぜ出版でないかを説明しており、この長さの違いが文化の違いを反映していると考えられます。

NIPS は遅くとも 2010 年にはこう書いているので、昔からそうであったと言えます。一方 CVPR は 2012 年からこのような注意書きが入りました。ですから、2012 年前後が普及への臨界点だったのではないかと思います。

さらに、図 10 は 2016 年 12 月 9 日に出たプレスリリースで、「なお、本研究成果は 2016 年 12 月 5 日に

『Computing Research Repository』に公開されました」と書いてあります。この「Computing Research Repository」というのは arXiv です。つまり、arXiv に投稿したのでプレスリリースしますというものです。

このプレスリリースを行った大阪府立大学の黄瀬浩一教授および岩村雅一准教授は知り合いだったため、個人的にメールを出してプレスリリースの背景について幅広く意見交換させて頂きました。特に、通常のプレスリリースは査読済み論文が出版されるタイミングが多いのに対し、査読がない arXiv への投稿の段階でプレスリリースを行うことへの考え方をお聞きしました。すると、「慎重にすべきという意見もあったが、速報性を重視してこのタイミングでリリースすべきと個人的に判断した」とのことでした。

このような成果の迅速な公開は、実は Ingelfinger Rule と呼ばれるルールに反します。これまで査読論文に関連する内容の公表を遅らせてきたのは、プレプリントやプレスリリースを論文の公開前に行ってはいけないというルールのためでした。しかし速報性を重視する圧力が高まっており、そうも言っていない状況になってきています。

ディープラーニングの会議で、International Conference on Learning Representations (ICLR) というものがあります。ここは面白くて、著者が arXiv に論文を投稿し、そのリンクをプログラム委員会に連絡することで、査読が始まるというプロセスを採用しています。つまり論文を最初に arXiv に投稿しなくてはならない

二重投稿とarXiv

[NIPS] Prior submissions on arXiv.org are permitted.
<https://nips.cc/Conferences/2016/CallForPapers>

[ICML] Submission is permitted for papers that are available as a technical report (or similar, e.g., in arXiv).
<https://2017.icml.cc/Conferences/2017/CallForPapers>

[CVPR] Note that such a definition does not consider an arXiv.org paper as a publication because it cannot be rejected. It also excludes university technical reports which are typically not peer reviewed.
http://cvpr2017.thecvf.com/submission/main_conference/author_guidelines

NIPSは遅くとも2010年に、CVPRは2012年からarXivに言及しており、2012年前後が普及への臨界点？

2017/02/14

第3回 SPARC Japan セミナー 2016

(図 9)

プレスリリースとarXiv



なお、本研究成果は2016年12月5日に『Computing Research Repository』に公開されました。
 論文タイトル: Deep Pyramidal Residual Networks with Separated Stochastic Depth
 掲載論文 (Cornell University Library「Computing Research Repository」)
<https://www.osakafu-u.ac.jp/news/publicity-release/pr20161209/>

2017/02/14

第3回 SPARC Japan セミナー 2016

(図 10)

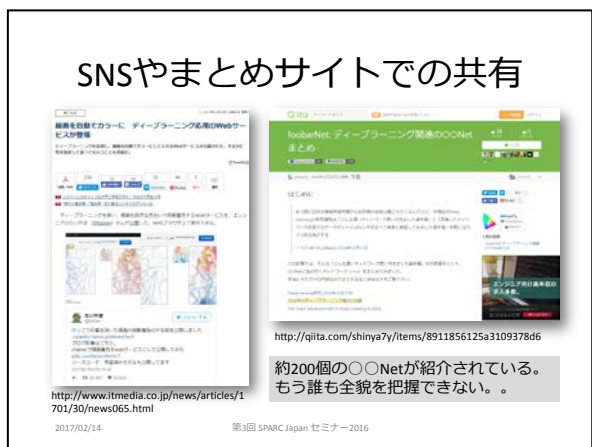
のです（図 11）。

最近はさらにオープンピアレビューに移行しています。OpenReview というシステムを使い、投稿した瞬間に論文が公開されるだけでなく、その後のレビューも公開されることになります。ICLR のサイトには、“The friction in our publication system is slowing the progress of our field.”、つまり friction（摩擦）があるようなシステムは問題であり、私たちは新しいやり方に変えるのだという宣言が書いてあります。私は本報告のテーマ「摩擦なき」に決めたとき、実はこの言葉を知らなかったのですが、やはり摩擦というアナロジーに行き着くのだと感慨を覚えました。

このような急速な進歩は、SNS やまとめサイトでどんどん共有され、どんどん広まっています（図 12）。この「ディープラーニング関連の〇〇Net まとめ」には、〇〇Net というものが約 200 個紹介されていて、



(図 11)



(図 12)

もう誰も全貌を把握できないという状況です。このようなソーシャルな方法で、情報共有が加速的に進んでいます。

これはある意味、市場によるオープン化であって、オープンソース、オープンデータにする方がお得だという動機からオープン化が進んでいきます。オープンアクセスリポジトリも、最初に「出版」して、その後には査読に行く世界になってきています。ソーシャル化で研究成果は SNS で迅速に共有され、公開サービスもいろいろな人が立ち上げ、日々どんどん進化している状況です。さらに、その公開サービスは、多くのユーザーがすぐに試すので、迅速性や頑健性も日々ネット上でチェックされます。そういう時代になってきました。

Michael Nielsen は、邦題『オープンサイエンス革命』という本を書いた人です（図 13）。実はこの英語のタイトルは“Reinventing Discovery: The New Era of Networked Science”です。「オープンサイエンス」とはタイトルでは言っていないのですが、オープンサイエンスに関することも論じています。

この本で描かれているオープンサイエンスの世界が、ディープラーニングでかなり実現してしまっているのではないかというのが私の感想です。いろいろな形でオープン化が進んで、知識が共有されています。面白いのは、Nielsen が 3 冊目に書いた本がディープラーニングの本なのです。両方ともネットワークという点で、2 冊の本には多くの共通する面があるようです。



(図 13)

法によるオープン化も着々と進んでいます（図 14）。ゲイツ財団は、財団の助成を受けた論文の、財団の OA 方針に対応していないトップジャーナルへの掲載は認めないこと、ウェルカム財団はプレプリントを助成金申請に記載してもよいことを表明しています。

3. 摩擦なき学術情報流通へのシフト

このような動きの中で起こっていることは、摩擦なき学術情報流通へのシフトだと思います。そこで考えなければいけないのは、我々が本当に欲しいものは何かということです。「人々が欲しいのは、ドリルではなく穴である」という有名な言葉があります。ドリルを買いに来た人がいて、ドリルがないときに、別のドリルを勧めるのではなく、そもそもなぜドリルが必要なのかを聞き、穴が必要なのだったら穴を開ける別の機械をお勧めすればよい。目的と手段は何かと考えることを忘れてはいけないという言葉です。

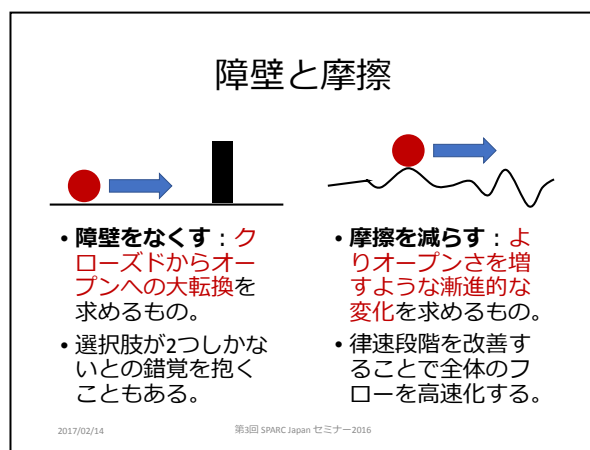
私は昨年の SPARC Japan セミナーで障壁について話しました。障壁をなくすのは壁を取るという考え方で、時としてクローズドからオープンへの大転換を求めることにもなりますが、摩擦を減らすというのは動きをスムーズにするということで、そうした漸進的な変化も必要ではないかと考えています（図 15）。

摩擦の代表例として査読があります。arXiv への投稿も、要は査読の問題をどうするかに集約されます。査読があることで公表が遅れる、不当な査読がある、査読者を見つけるのが大変で持続的でないといった問

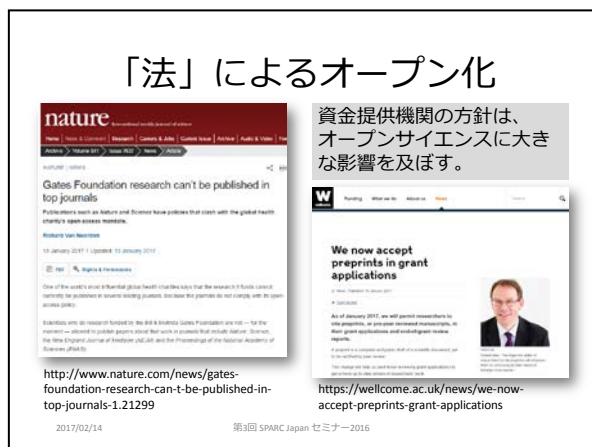
題があります。

これまでの査読はゲートキーパー、つまり優れた論文だけを選別する門番の役割を果たしていました。これは確かにスペースが有限であれば有効な方法です。紙面や時間枠の制約があるので、門番があなたは通す、あなたは通さないと選ぶ方法が有効でした。しかし今はもうネットとなってスペースは無限なので、むしろ利用者の注意力が有限であるという点が制約になってきます。利用者にとって重要なものを、既に出ているものからどうやって拾ってくるか、ということが大きな課題となります。

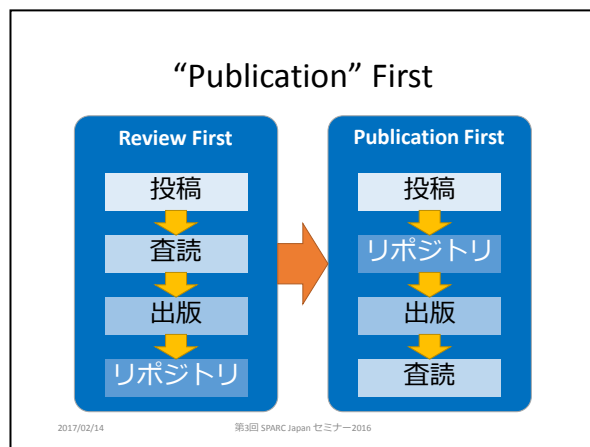
そのとき、一つの予測として、選ぶという行為が付加価値サービスになる可能性があります。誰にとっても、自分が読むべき論文を選んでくれるとありがたい。これはある意味では、出版物のオープンアクセスが、著者支払いモデルから購読モデルに戻るとも言えます。



（図 15）



（図 14）



（図 16）

つまり、あなたのために論文を厳選してくれる付加価値サービスを購読するのです。これを私は Subscription model 2.0 と呼びたいと思います。ただこのような状況では、どのリポジトリにデポジットするかが流量勝負となるかもしれません。そうすると arXiv のようなリポジトリに比べ、機関リポジトリなどは難しくなるかもしれません。いずれにしろ Publication が最初、後で査読というモデルに移行する可能性が高いと思います（図 16）。

出版は人類の知の共有が目的で、評価は人類の知の継承が目的だと思います。ですから、オープンな出版と、多様な評価を組み合わせることで、出版と評価を分離できればいいのではないかと。最終的に欲しいのは門番ではなく、この人の言うことは信用できるというブランドではないかと考えます。

4.まとめ

オープンサイエンスの重要な駆動力としてスピードがあるというのが本報告の主張です。ディープラーニングはスピードへの圧力が強烈な特殊事例かもしれませんが、決して孤立事例ではないと考えています。物事は臨界点を過ぎると急速な変化が起きるものであり、そういう時期を既に迎えた分野としてディープラーニングを捉えることもできます。本セミナーは図書館の参加者が多いと思いますが、この未来に図書館が果たせる役割は何か、次のアクションをぜひ考えていきましょう。

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

JST におけるオープンサイエンスへの対応 (DMP 導入試行をはじめとして)

小賀坂 康志

(科学技術振興機構)

講演要旨



JST は 2013 年に研究成果論文へのオープンアクセス方針を公表して以来、国内外動向をふまえて研究成果へのアクセス・流通の促進（いわゆるオープンサイエンス対応）について検討・対応を行ってきた。論文については引き続き環境整備を進めると共に、研究データについても、2016 年には RDA 総会を誘致開催（東京）し、また戦略的創造研究推進事業の一部の研究課題について、データ管理計画（DMP）を導入する等、対応を進めている。本講演では、JST におけるオープンサイエンス対応の状況について報告する。



小賀坂 康志

学習院大学大学院自然科学研究科博士後期課程修了。博士（理学）。JSPS海外特別研究員（NASA/GSFC）等を経て、2000年名古屋大学大学院素粒子宇宙物理学専攻助手。X線天文衛星を用いた高エネルギー天体物理学研究及びX線望遠鏡開発に従事。2009年JST入職、科学技術振興調整費、A-STEP、再生医療、CREST・さがけ等のファンディング事業を経て、2015年10月より現職で情報事業（J-STAGE、JaLC、researchmap等）を担当。

科学技術振興機構（JST）の戦略的創造研究推進事業では、2016 年からデータマネジメントプラン（DMP）の大規模な導入を始めています。国内では大規模導入は恐らく初めての事例なので、その経緯も含めて皆さんに共有できたらと思います。

あるいは科学コミュニケーション、理数学習支援です。

この JST の立場でオープンサイエンスの推進をするというときに、踏まえていただかなければならないのは、各種の政策文書あるいは提言です（図 1）。上から順に内閣

1.概況（JST におけるオープンサイエンス対応）

JST には三つの機能があります。一つ目は、研究開発戦略の立案や提言といったシンクタンク機能です。二つ目は、科学技術イノベーションを興すファンディング事業です。具体的には、基礎研究、シーズの実用化研究、産学連携、国際共同研究をやっています。三つ目は、基盤事業です。具体的には私がいる情報事業、

日本におけるオープンサイエンス推進の動き

- ・内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」報告書（2015年3月）
→公的研究資金による研究成果の利活用促進の拡大
- ・第5期科学技術基本計画（2016年1月閣議決定）
→第4章(2)③ オープンサイエンスの推進
- ・文部科学省 科学技術・学術審議会 学術分科会 学術情報委員会「学術情報のオープン化の推進について」（2016年2月）
- ・G7茨城・つくば科学技術大臣会合（2016年5月）
つくばコミュニケ（共同声明）
→オープンサイエンスを効率的に推進し、適切に活用
- ・科学技術イノベーション総合戦略2016（2016年5月）
→オープンサイエンスの推進の基本姿勢の下、研究成果・データを共有するプラットフォームの構築
- ・日本学術会議「オープンイノベーションに資するオープンサイエンスのあり方に関する提言」（2016年7月）

Japan Science and Technology Agency

第3回SPARC Japanセミナー2016（2017/2/14 JST小賀坂）

科学技術振興機構

(図 1)

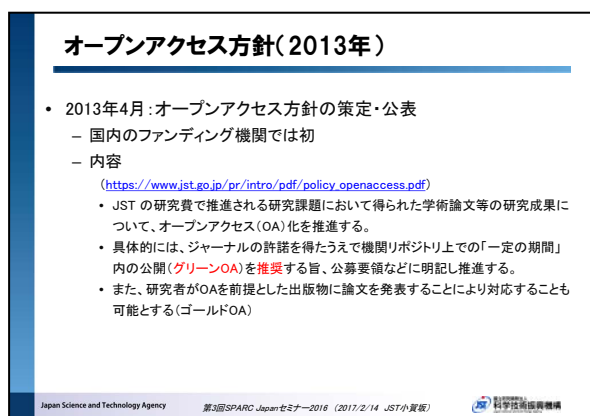
府、今年度から始まっている第5期科学技術基本計画、文部科学省、果ては、昨年日本で開催されたG7科学技術大臣会合、科学技術イノベーション総合戦略、日本学術会議の提言、こういうところでオープンサイエンスということが盛んに言われるようになってきています。研究資金配分機関であるJSTとしても、この対応を考えていかなければなりません。

最初にこれまでどんなことをやってきたのか簡単にご説明します。振り返ると、2013年、JSTは国内のファンディング機関としては初めてオープンアクセス方針を策定・公表しました(図2)。JSTの研究費で推進された研究課題において得られた学術論文等の研究成果については、オープンアクセス化を推進することを宣言しています。具体的な内容としては、グリーンオープンアクセスを推奨し、ゴールドオープンアクセスについても可とします。可とするというのは、それに必要な経費を研究費から支出してもよいという意味です。このような方針を打ち出したところです。

続いて、ジャパンリンクセンター(JaLC)としての活動ですが、2015年に研究データにDOIを付与するという取り組みを始めています(図3)。ここでは研究データについて関心のある機関やプロジェクトが集まってDOI登録をする実証プロジェクトを行い、この結果について得られた知見を基に、「研究データへのDOI登録ガイドライン」を公表しました。

昨年はリサーチデータ・アライアンス(RDA)のplenary meetingを東京に誘致・開催しました(図4)。おかげさまで340人という多数のご参加を頂いたという結果でした。

また、これも昨年ですが、研究データ利活用協議会を設置して、研究データのさらなる利活用に向け、国内主要機関や多くの参加者を得て、種々、議論を行っています(図5)。



オープンアクセス方針(2013年)

- 2013年4月: オープンアクセス方針の策定・公表
 - 国内のファンディング機関では初
 - 内容
 - JSTの研究費で推進される研究課題において得られた学術論文等の研究成果について、オープンアクセス(OA)化を推進する。
 - 具体的には、ジャーナルの許諾を得たうえで機関リポジトリ上で「一定の期間」内の公開(グリーンOA)を推奨する旨、公募要領などに明記し推進する。
 - また、研究者がOAを前提とした出版物に論文を発表することにより対応することも可能とする(ゴールドOA)

https://www.jst.go.jp/pr/intro/pdf/policy_openaccess.pdf

(図2)

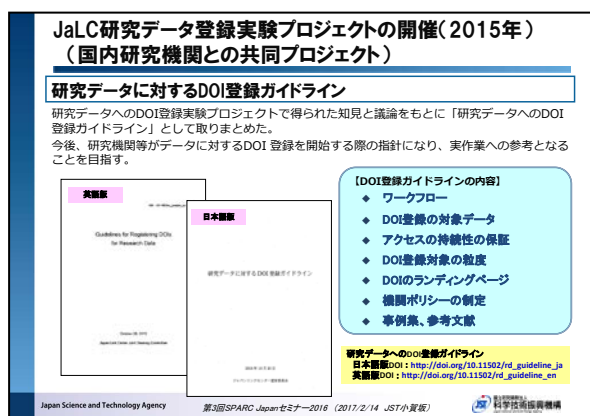


RDA総会誘致(2016年)

- 2016年3月: RDA Plenary Meeting(東京)を誘致・開催
 - テーマ: Making data sharing work in the era of Open Science
 - 日時: 平成28年3月1日(火)~3日(木)
 - 会場: 一橋講堂(東京都千代田区)
 - 主催: RDA及びJSTによる共同開催
 - 参加者: 342人

Japan Science and Technology Agency 第3回SPARC Japanセミナー2016 (2017/2/14 JST小夏版)

(図4)



JaLC研究データ登録実験プロジェクトの開催(2015年) (国内研究機関との共同プロジェクト)

研究データに対するDOI登録ガイドライン

研究データへのDOI登録実験プロジェクトで得られた知見と議論をもとに「研究データへのDOI登録ガイドライン」として取りまとめた。
今後、研究機関等がデータに対するDOI登録を開始する際の指針になり、実作業への参考となることを目指す。

【DOI登録ガイドラインの内容】

- ◆ ワークフロー
- ◆ DOI登録の対象データ
- ◆ アクセスの持続性の保証
- ◆ DOI登録対象の粒度
- ◆ DOIのランディングページ
- ◆ 機関ポリシーの制定
- ◆ 事例集、参考文献

研究データへのDOI登録ガイドライン
日本語版DOI: http://doi.org/10.11501/rd_guideline_ja
英語版DOI: http://doi.org/10.11501/rd_guideline_en

(図3)



研究データ利活用協議会の設置(2016年) (ジャパンリンクセンターの活動)

- 設立背景
 - 第7回RDA総会等を通じた研究データ利活用の議論活性化
 - 「研究データへのDOI登録実験プロジェクト」によって醸成された分野横断型の実務担当者コミュニティの形成
- 目的
 - 研究データに関する多様なセクターが集い、研究データの共有と公開に関する課題解決と技術の共有
 - 海外等の関連組織・活動との情報共有とコラボレーション
- 参加機関

6機関(NDL、NII、NIMS、NICT、JST、千葉大)39名、および個人参加25名
- 活動状況

開催	開催	開催日	開催
1	設立準備会議(第1回)	5月26日(木)	JST
2	設立	6月3日(金)	NDL
3	公開キックオフミーティング	7月25日(月)	JST
4	研究会(第1回) RDA Eighth Plenary Meeting 等参加報告	10月3日(月)	NII
5	研究会(第2回) SPARC Japanとの共催	10月3日(月)	NII
6	公開シンポジウム(サイエンスアゴラ内)「研究データの利活用の未来-オープンサイエンスの実現手段-」	11月4日(金)	JST
7	研究会(第3回) 第1回COOHセミナー-Big Data and Digital Humanities(後援)	1月23日(月)	NII
8	研究会(第4回) WDS国内シンポジウムとの共催(来年度の活動計画の検討)	3月9日(木)~10日(金)	—

(図5)

2.戦略的創造研究推進事業における DMP の導入

このような経緯を踏まえて、本題である、戦略的創造研究推進事業における DMP の導入についてご説明したいと思います。

まず、戦略的創造研究推進事業についてごく簡単に申し上げます。趣旨は、科研費のような純粋基礎研究とは異なり、国が定める戦略的な目標の達成に向けて基礎研究を推進するというものです。研究体制は、複数の研究者から成るネットワーク型の研究所を構築し、そこにプログラムオフィサーたる研究総括を配置し、研究領域の運営をするというスタイルです。

DMP の導入に向けての経緯として、オープンデータの動向の高まりがあり、これに対して JST 内でも議論しなければならないという機運が高まりました（図 6）。これが 2014 年ごろの話です。

このような背景を受けて、戦略事業においても DMP 導入を検討してきましたが、まずは対象となる研究者に対して意識調査を行うところから始めました。これが 2015 年のことです。

これに基づいて、2016 年 2 月に事業としてのデータマネジメントポリシーを策定しました。その中で、DMP に書き込むべき事柄などを定めています。

それで、いよいよ 2016 年の公募から、一部の研究領域について、DMP の提出を義務化しました。それでは、この経緯を少し順を追ってご説明したいと思います。

2-1.導入準備：研究者の意見照会

まず、DMP の導入に先立って、研究者に対する意見照会を行いました（図 7）。対象研究者に対してデータを提供する側、あるいはデータを利用する側のそれぞれの立場から、データを共有した経験の実態を調査し、それを基に、研究者にとって本当に意義のあるデータ共有を促進するための方針を策定したいと考えたのです。期間は 2015 年 10～11 月、ウェブアンケートで行い、対象となった研究者は 769 名、これに対して 4 割弱の 300 名弱の回答を得ています。主な項目三つについてアンケート結果をご紹介します。

最初は、データの利用の経験について聞いたところ、「他者のデータを利用した経験がある」という割合は 6 割弱でした（図 8）。その経験者に、「最も効果的に利用できたデータは何であったか」と聞いたところ、その 7 割が、公的研究機関や学会が作成した公開データでした。また、利用者の 8 割以上は「満足」であったという回答でしたが、その「満足」に至った要因と

導入準備：研究者に対する意見照会

- 目的：DMP導入を検討するに当たって、**研究データの提供者・利用者それぞれの立場から、データ共有経験の実態を調査し、データ共有の成功要因を明らかにするとともに、研究者にとって意義のあるデータ共有を促進するための方針策定に活用する。**

- 対象：CREST・さがりけ・ERATO・ACT-Cの研究代表者等
- 期間・手法：2015年10～11月、Webアンケート
- 回答率：37% (288名/769名)

Japan Science and Technology Agency
第3回SPARC Japanセミナー2016 (2017/2/14 JST小室版)
科学技術振興機構

(図 7)

DMP導入の経緯・経過

- 「オープンデータ」動向の高まりを受けて、JST内でもDMP導入について議論を開始(2014年～)
- 戦略的創造研究推進事業において導入を検討
 - 研究者に対して意識調査を実施(2015年)
 - データマネジメントポリシーを策定(2016年2月)
 - 一部の研究領域についてDMP提出を義務化(2016年公募より)

Japan Science and Technology Agency
第3回SPARC Japanセミナー2016 (2017/2/14 JST小室版)
科学技術振興機構

(図 6)

研究者アンケート結果①

データの利用について

- データの利用経験率は約57%
- 最も効果的に利用できたデータの7割が、公的研究機関や学会作成による公開データ
- データ利用の総合的な満足度は8割以上
- 満足に至った要因としてはデータの信頼性、次いで利用のしやすさへの満足が重要

【データマネジメント実施方針への反映】

→公的データベースへの登載を推奨するのが良い

Japan Science and Technology Agency
第3回SPARC Japanセミナー2016 (2017/2/14 JST小室版)
科学技術振興機構

(図 8)

して一番大きなものは「データの信頼性が高かったこと」という意見がありました。

いろいろ論点はあるのですが、データの利用者がこういう感想を持っているということなので、これを踏まえて、事業として、データの登録先は公的データベースをできるだけ使うように推奨することがよからうと考えるに至りました。これもデータマネジメントポリシーに書き込んであります。

他方、今度はデータを提供する側としての経験を聞いたところ、約半数の研究者が「データを提供した経験がある」と答えました（図 9）。提供の結果、「コラボレーションがうまくいった」という回答が半数強でした。何をもって成功と思ったかについては、「共同研究に進展した」「相手側の研究が進展した」が多かったです。

その他、いろいろありますが、こうした回答を踏まえて、特に私どもが重要だと思ったのは、利用者からのフィードバックが得られるような仕掛けを組み込むことです。そのため、出典を明記すること、データと論文情報のひも付けをしっかりとすることも方針として書き入れました。

また、データを提供した経験を有する研究者の過半数が「個人的な要望に応じて研究データを提供した」と回答しました。これは逆に、個人的なコネクションがないと研究データの共有に至っていないという現状があると解釈し、パブリックなデータベースにできるだけ登録を促すことが重要であろうと考えました。

「データを提供したことがない」という場合、その

理由を聞いたところ、「相手側からの要望がない」「義務負担」「自身へのメリットが不在」「管理責任が重い

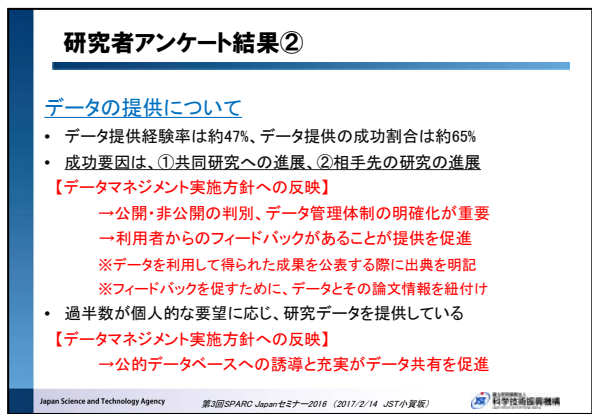
「論文発表」「特許出願」「守秘義務契約に抵触するという懸念」などがありました（図 10）。これはわれわれも想像がつくところではあります。特に、若手層のデータ共有の機運を高めていくことも大切であろうということで、まずは将来のある若手研究者に DMP の作成をしっかり促すことが重要であると考えました。

2.2. データマネジメントポリシーの策定

こうした調査を踏まえてデータマネジメント実施方針を立て、その中で DMP の記載内容を図 11 のとおりに決めました。これが出たのが 2016 年 2 月、ちょうど 1 年ぐらい前だと思います。

内容としては 6 項目あり、1 点目は、管理対象となる研究データの保存・管理方針を定めることです。

JST としては、国の方針である、アカデミッククラウドの活用を推奨することを明記しています。



研究者アンケート結果②

データの提供について

- データ提供経験率は約47%、データ提供の成功割合は約65%
- 成功要因は、①共同研究への進展、②相手先の研究の進展

【データマネジメント実施方針への反映】

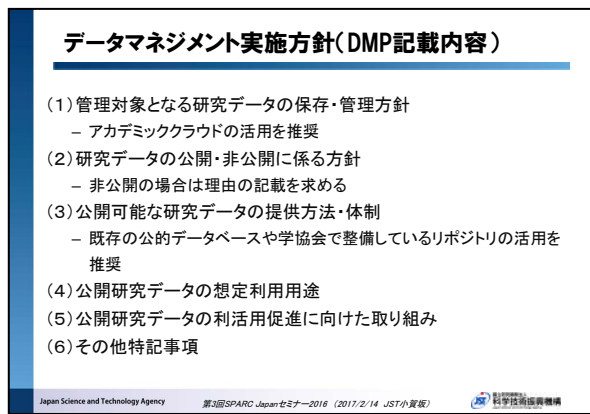
- 公開・非公開の判別、データ管理体制の明確化が重要
- 利用者からのフィードバックがあることが提供を促進
- ※データを利用して得られた成果を公表する際に出典を明記
- ※フィードバックを促すために、データとその論文情報を紐付け

- 過半数が個人的な要望に応じて、研究データを提供している

【データマネジメント実施方針への反映】

- 公的データベースへの誘導と充実がデータ共有を促進

(図 9)



データマネジメント実施方針(DMP記載内容)

- 管理対象となる研究データの保存・管理方針
 - アカデミッククラウドの活用を推奨
- 研究データの公開・非公開に係る方針
 - 非公開の場合は理由の記載を求める
- 公開可能な研究データの提供方法・体制
 - 既存の公的データベースや学協会で整備しているリポジトリの活用を推奨
- 公開研究データの想定利用用途
- 公開研究データの利活用促進に向けた取り組み
- その他特記事項

(図 11)

2 点目は、研究データの公開・非公開に係る方針を定めることです。非公開の場合には理由の記載を求めていることにしています。ここで一言付け加えると、そもそも DMP は、JST がそれ以外にデータ取り扱い方針を定めていない場合は、研究者が自分の意思でもって決めることです。公開せずともよいわけです。しかし非公開の場合は、なぜか、ということは書いていただきたいという思いが込められています。

3 点目は、公開可能な研究データがあった場合、その提供方法や体制を定めていただくことです。JST としては、既存の公的データベースや学協会で整備しているリポジトリの活用を推奨するとしています。

その他は、公開研究データの想定利用用途、公開研究データの利活用促進に向けた取り組み、その他特記事項の記載を求めています。

2-3.公募要領への記載

公募要領には、例えばこのように記載されています（図 12）。「平成 28 年度以降に新たに設定された研究領域で採択された研究代表者は（中略）『データマネジメントプラン』を研究計画書と併せて JST に提出していただきます。また、上記方針に基づいてデータの保存・管理・公開を実施していただきます」。このように求めることにしました。

ここでお気付きだと思いますが、欧米の方針と違い、まず、採択された研究者に求めることにしています。これには大きな理由があります。欧米では応募時に求めることが一般的であろうと思いますが、それは採択

審査のときの評価基準にすることを意味しているのです。しかし、実際にはそれほど審査に重く使われているわけではないと聞いています。私も NSF や DOE のオフィサーと話をしましたが、その辺は悩ましいということです。

われわれも DMP が出たところで、それをどのように評価軸として設定し、どれぐらいウエートを置けばよいのかということについては、知見がありません。また、評価を行う研究者も、そのことに必ずしもリテラシーがあるとは言えないと思います。従って、現時点では採択審査の基準にはせずに、採択後に研究計画書の一部として出していただくことにしました。

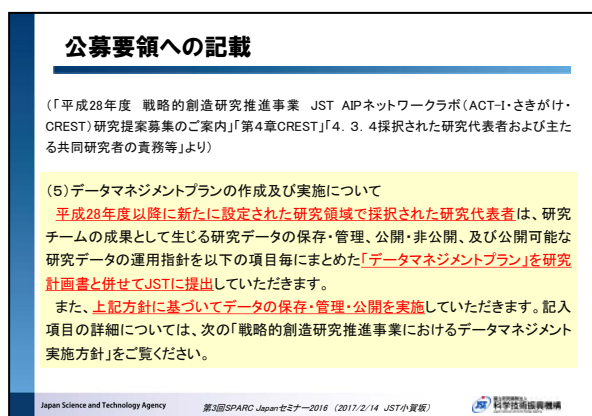
2-4.DMP の導入・提出状況及び研究者からの反応

公募の結果はもう出ており、採択課題も決定しています（図 13）。DMP 提出の対象となった研究課題は 79 課題あります。一つのプロジェクトが複数のデータ項目についての管理方針を定めてくるので、提出されたデータ項目は全部で 230 件ぐらいありました。

その内訳を少しここにブレイクダウンしてあります。データの内容は、実験データ、計算データ、あるいは計算モジュールそのものであったり、多様でした。

公開・非公開に係る方針については、「非公開」「限定的な公開」「公開」の 3 段階で記載を求めているのですが、約半数が「限定的な公開」という回答でした。残りは「公開」と「非公開」という回答が半分ずつぐらいでした。

保存場所については約半数が「研究室等の自前のサ

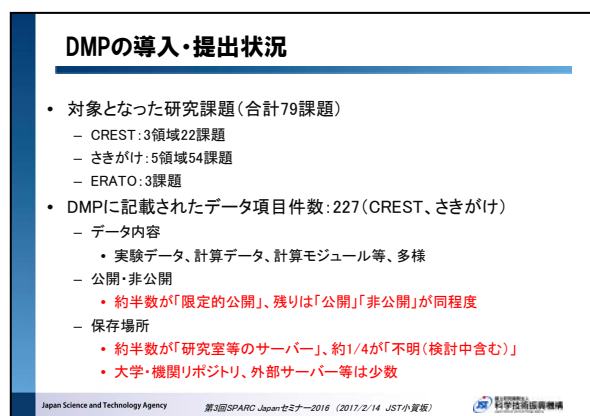


公募要領への記載

（「平成 28 年度 戦略的創造研究推進事業 JST AIP ネットワークラボ（ACT-I）さがけ・CREST」研究提案募集のご案内「第 4 章 CREST」）4. 3. 4 採択された研究代表者および主たる共同研究者の責務等より）

（5）データマネジメントプランの作成及び実施について
平成 28 年度以降に新たに設定された研究領域で採択された研究代表者は、研究チームの成果として生じる研究データの保存・管理、公開・非公開、及び公開可能な研究データの運用指針を以下の項目毎にまとめた「データマネジメントプラン」を研究計画書と併せて JST に提出していただきます。
また、上記方針に基づいてデータの保存・管理・公開を実施していただきます。記入項目の詳細については、次の「戦略的創造研究推進事業におけるデータマネジメント実施方針」をご覧ください。

(図 12)



DMP の導入・提出状況

- 対象となった研究課題（合計 79 課題）
 - CREST: 3 領域 22 課題
 - さがけ: 5 領域 54 課題
 - ERATO: 3 課題
- DMP に記載されたデータ項目件数: 227 (CREST、さがけ)
 - データ内容
 - 実験データ、計算データ、計算モジュール等、多様
 - 公開・非公開
 - 約半数が「限定的な公開」、残りは「公開」「非公開」が同程度
 - 保存場所
 - 約半数が「研究室等のサーバー」、約 1/4 が「不明（検討中含む）」
 - 大学・機関リポジトリ、外部サーバー等は少数

(図 13)

「サーバー」と回答しています。残る半数のうちのさらに半分、全体の4分の1が「検討中」も含む「不明」という回答でした。翻って、「大学」「機関リポジトリ」「外部サーバー」等は、ごく少数の、1〜2割程度という回答でした。確かにそうであろうと思う反面、少々期待外れな面はあったと言わざるを得ません。

もう少し所感的なことを申し述べると、全体的な傾向として、公開の範囲を限定的にするという回答が半分ぐらいあったのですが、その公開の範囲は、おおむね研究チーム内、研究領域内、CRESTならCRESTの事業の中の研究者ないしは共同研究者内でした（図14）。

限定的な公開には2種類あるといわれています。一つは「制限共有」といって、共有グループをかなり明示的に限定した者の中での共有です。これに対して、「制限公開」という考え方があります。これは基本的に公開なのですが、メンバーシップを限定するということです。例えば、利用申し込みに対して要件を満たしていれば利用を許すという形です。そういう観点からすると、回答はほぼ「制限共有」にとどまるというレベルであったということです。

しかしながら、公開すべきかどうかについては、それほどかつちりとした答えを持っていないのではないかという印象を持っています。というのは、要請があれば「非公開」から「限定的公開」へ、あるいはそこから「公開」する用意もあるという回答もあったからです。

一つ面白かった意見は、厳しいといえは厳しいので

すが、「利活用を想定して（安全を見て）多くのデータ保管をすることはしない。保管にもデータ解析にもコストが掛かるので、他の目的のためにデータを再解析するよりも、むしろ、その目的に合ったデータを再測定する方がよい」というものでした。これも容易に想像がつく意見なのですが、こういうご指摘に対して、なぜデータ保管をするのかという説明をわれわれはしていかなければいけないのだらうと思った次第です。

そういうわけで、DMPの作成はおおむね順調に進んでいます（図15）。うまくいった理由は、過剰なボリュームを求めなかったこと、事業側であらかじめ記載例を作成したこと、場合によってはJST側の担当者が作成を支援したことだと考えられます。今後は、データの保管場所についての問題、共有や公開する場合の支援など必要になるでしょう。この方針は平成29年度も続くので、次第に導入する研究課題の数は増えていくと思っています。

3.JSTのその他の取り組み

最後に、今後、JSTのオープンサイエンス対応として何を考えているか、簡単にご紹介します。

一つはオープンアクセスに関するもので、CHOR Pilot Projectです（図16）。CHORとは米国にベースを置く出版者の連合体ですが、CHORUSという取り組みをしています。CHORUSの詳細は申しませんが、JSTと千葉大学はCHORと組み、この取り組みが日本でワークするかどうかについてのパイロットプログラムを走らせていて、近々、結果が出ようとしています。

DMP記載内容・研究者からの反応

- 全体的な傾向
 - －「限定的公開」の共有範囲は、概ね研究チーム内、研究領域内、ないし共同研究者（いわゆる「制限共有」であって「制限公開」ではない）
 - －要請があれば「非公開」→「限定的公開」、「限定的公開」→「公開」とする用意があるケースが多い
- 研究者からの反応
 - －DMP作成についての特段の問い合わせはなし
- 研究者からの意見（例）
 - －「利活用を想定して（安全を見て）多くのデータを保管することはしない。むしろ、必要が生じたら再測定する。保管にもデータ再解析にもコストがかかることから、他の目的のためにデータを再解析するよりも、その目的に合ったデータを再測定する方がよい。」

Japan Science and Technology Agency 第3回SPARC Japanセミナー2016（2017/2/14 JST小夏版）

（図14）

今後の課題

- DMP作成は概ね順調
 - －適切なボリューム（1ページ程度）
 - －あらかじめ記載例を作成
 - －JST課題担当者が作成を支援した例もあり
 - 研究現場への導入は概ね問題なく進捗
- DMPに沿ったデータ管理の実践
 - －データ保管場所の確保は課題
 - －共有・公開の支援も必要
 - policy into practiceの難しさはこれから

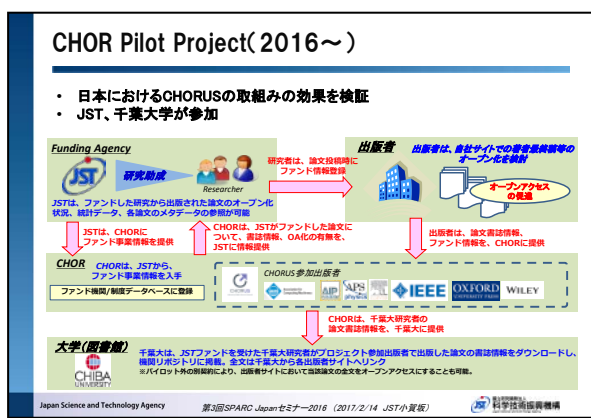
Japan Science and Technology Agency 第3回SPARC Japanセミナー2016（2017/2/14 JST小夏版）

（図15）

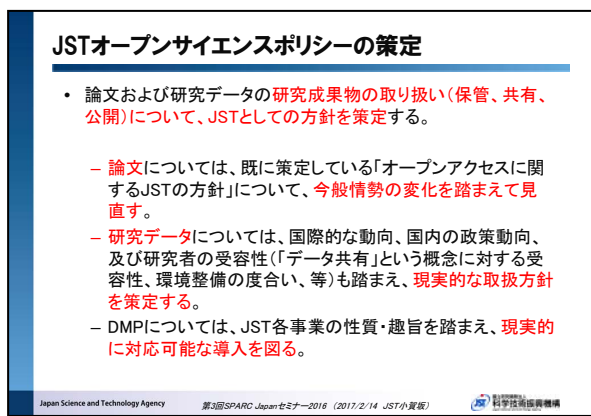
うまくいけば、ファンディングエージェンシーとしてこの取り組みを維持していくことも考えています。

また、少しテクニカルな話になりますが、オープンサイエンスの動向が高まると、インターオペラビリティの確保の観点から、早晚、persistent identifier (PID) をどう管理するのかということが課題になってくるかと思っています。端的に言うと、グローバルな PID のシステムとドメスティックなシステムとの相互乗り入れをする機能が早晚必要になるのではないかと考えていて、JST というよりは JaLC として、こういうシステムが必要ではないかと検討しています。

もう一つ、JST としてのオープンサイエンスポリシーの策定を進めています（図 17）。論文については、既に策定しているポリシーの見直しを行って再策定しようと思っています。研究データについても、国内外の動向を踏まえ、研究者の受容性についても十分考えた上で、現実的な取り扱い方針を策定したいと考えています。



(図 16)



(図 17)

DMP については、今、JST の一部の事業に導入していますが、全事業に対して導入を図ることを検討しています。その際には JST は多様な事業があるので、各事業の性質・趣旨を踏まえ、現実に対応可能な導入を図っていくことを考えています。そう遠くない将来に、JST としての統一方針をお示しできるものと思います。

●古賀 天理大学の古賀と申します。ご紹介いただきましたポリシーあるいはプランの中で、データ保存の永続性を保証するための手段は今後どこまで考えられるのか、お聞きしたいと思います。その手段としては、システムリプレースやお金を継続的に提供することも考えられるかと思いますが、データ永続保存の取り組みに関して、考えをお聞かせいただければと思います。

●小賀坂 データそのものが失われないように保管するという観点と、公開の場合はデータを検索可能な状態に置くという観点の2点があると思います。前者については、国の方針としてNIIのアカデミッククラウドの整備も進むので、この中で国としてセキュアな環境が整っていくことになっています。

一方、検索可能性については、今のところは論文に振られているDOIを研究データにも適用するのがよいのではないかと、各国で取り組みが進んでいて、日本でもJaLCがそういう取り組みをしました。こういうものが整備されていくのではないかと考えています。もし考えに抜けがあれば、どなたかコメントいただければと思います。

●林 NISTEP の林です。DMP がうまくいった理由として、記載例を用意したという話と、担当者が作成を手伝ったという話がさらっとありましたが、もう少し詳しく具体的に、どういう人材がどういうふう何を手伝ったか教えていただけますか。それがつまり、

DMP を支援する人材の一例であるように見えたのです。その観点からもう少し踏み込んで、SPARC Japan セミナーという場でもあり、ライブラリアンがそういうことを支援しようとした場合に役立つようなことを知りたいので、実際に何が起きたかお聞かせ願えますでしょうか。

●小賀坂 JST のプロジェクト担当者が研究計画書を見て、この研究であれば、例えば NMR をやると言っているの、このデータが出てくるのではないかと推し量って、これは厳しそうだから限定公開かなということを担当者の考えでもって書いて、「先生、こんなものでしょうか」とお示しします。先生はそれでよければ「それでいいよ」とおっしゃるし、そうでなければ「ここは違う」とおっしゃいます。行為としてはそういうことをやった担当者もいたということです。

なぜそういうことをする、ないしはできるかという、基本的に JST の課題担当者は、研究領域の設定の段階からコミットして、研究分野の専門的内容に踏み込んだ検討をしてプログラムの設計をしますし、出てきた研究計画書について、きちんとそれで研究が実施されるかというところまで踏まえて中身を見て、研究者とやりとりをするので、その中で研究内容についても学んでいきます。特段、何かのレクチャーを受けたということではなく、OJT ができるようになっていくというのが本当のところではないかと思います。

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

材料科学分野における データ利用のライセンスの考え方 －図書館からの視点

小野寺 千栄

(物質・材料研究機構)

講演要旨



データには著作権がないため、利用に際してはライセンスが重要となる。本発表では、物質・材料研究機構で2016年に開催した3回の連続セミナーをもとに、データ利用におけるライセンスの考え方について整理し、データ利用の実例について紹介する。また、図書館はそこでどのような役割を担うことができるのかを考える。



小野寺 千栄

物質・材料研究機構 図書館員。大学図書館等での勤務を経て、2015年より現職。研究情報の受発信を含めた図書館業務全般に携わっている。

今日は材料科学分野におけるデータ利用のライセンスの考え方について、図書館からの視点でご紹介します。まず、物質・材料研究機構（NIMS）についてご紹介した後、NIMS で開催した「オープンサイエンスと著作権」のセミナーについて、次に NIMS においてデータがどのように扱われているのか具体的な事例をご紹介します、最後にデータ利用について図書館がどのように関わることができるのかを皆さんと一緒に考えられればと思っています。

1.NIMS の概要

NIMS は国内で唯一の物質・材料科学の専門機関で

す。2016 年 10 月には理化学研究所、産業技術総合研究所と共に、特定国立研究開発法人に選定されました。組織のミッションとして、物質・材料科学技術に関する基礎研究および基盤的研究開発、研究開発成果の普及とその活用の促進、NIMS の施設および設備の共用、研究者・技術者の育成およびその資質の向上の四つを掲げています。職員数は約 1,500 人で、そのうちの約半数が研究に関わる職員です。私はそこで図書館司書として働いています。

図書館では、研究に関わる職員へのアドボカシー活動として、セミナーを開催しています。2016 年は「オープンサイエンスと著作権」について3回の連続

セミナーを開催しました。NIMS のデータ利用についてご紹介する前に、データの著作権についての考え方を整理するため、まず、このセミナーの概要をご紹介します。

第1回は2016年1月に開催しました（図1）。講師は、内閣府の真子博さん、文科省の著作権分科会の委員でもある末吉互弁護士です。約100名の参加者のうち、NIMS 内部からは研究系職員の約6%に当たる48名の参加がありました。真子さんにはオープンサイエンス政策の背景と今後の展開についてお話しいただき、末吉弁護士にはオープンサイエンスに必要な著作権、特に論文と研究データの著作権についてお話しいただきました。

講演では、(1) デジタル著作物を利用するには現行の著作権法は不十分であり、足りないところは契約で補うことができる、(2) 著作物とは、思想または感情を創作的に表現したものであり、データは著作物には該当せず、著作権法の対象にはならない、(3) そのため、第三者がデータを利用する際の条件は、契約で規定することになる、とお話しいただきました。この第1回の反響を受けて、第2回のセミナーを開催しました。

第2回はデータ共有を切り口に開催しました（図2）。講師は出版社、研究者、弁護士の3名でした。出版社からは、ネイチャー・パブリッシング・グループが発行しているデータジャーナル「Scientific Data」のエディターVarsha Khodiyar さんにお越しいただきました。研究者としては、この後、NIMS の事例でもご紹介する高分子データベースのデータ収集について桑島

功さんに、弁護士として第1回に引き続き、末吉弁護士にお話しいただきました。

Varsha さんからは、データ共有のプラットフォームとしてのデータジャーナルについてお話しいただきました。末吉弁護士からはあらためて、データは著作物に当たらないこと、そのためデータを収集して提供するデータベースも著作物として認められることは過去の判例を見ても非常に稀であり、データの利用は、データ提供元と利用者との契約によって規定されるということをお話しいただきました。

最後となる第3回のセミナーでは、材料・科学分野においてもデータ駆動型研究が推進されている背景から、テキスト・データマイニングや人工知能といった次世代の知的財産権を取り上げ、弁護士と出版社から講師をお迎えしました（図3）。

まず、内閣府の次世代知財システム検討委員会の委員でもある福井健策弁護士からは、人工知能をめぐる知的財産の論点として、三つのポイントをお話しいただきました。一つ目は、人工知能の学習用データの権利を著作権で保護するかどうか、保護する場合はどこ

図書セミナー「オープンサイエンスと著作権」第1回

- 2016年1月8日開催
- テーマ：オープンサイエンスと著作権
- 講師：内閣府、弁護士
- 参加者：101名
 - ・ NIMS（研究系職員）：48名
 - ・ 外部：53名（研究機関、図書館、政府機関、出版社・学会、大学）
- 概要：
 - ・ オープンサイエンス政策への取り組みと今後の展開。
 - ・ 論文と研究データの著作権。現行の著作権法はデジタル開示された著作物の利用には不十分であり、契約で補う。データは著作物ではなく、著作権法の対象にならないため、利用者が利用する条件は契約で規定する。

(図1)

図書セミナー「オープンサイエンスと著作権」第2回

- 2016年3月10日開催
- テーマ：(続) オープンサイエンスと著作権－データ共有
- 講師：出版社、研究者、弁護士
- 参加者：83名
 - ・ NIMS（研究系職員）：51名
 - ・ 外部：32名（企業、研究機関、政府機関、大学、図書館、出版社）
- 概要：
 - ・ データキュレーターから見るデータ共有（データジャーナル）。
 - ・ NIMSにおけるデータ収集と共有。
 - ・ データベースが著作物として認められることは、過去の判例を見ても非常に稀。著作物ではないデータは自由に利用できるが、データ出典元との契約があれば、その契約に従う。

(図2)

図書セミナー「オープンサイエンスと著作権」第3回

- 2016年11月24日開催
- テーマ：(続々)オープンサイエンスと著作権－データマイニングからAIへ
- 講師：弁護士、出版社
- 参加者：82名
 - ・ NIMS（研究系職員）：61名
 - ・ 外部：21名（研究機関、大学、企業、出版社、政府機関、図書館）
- 概要：
 - ・ AIをめぐる知的財産権の論点について：AIの学習用データの権利をどこまで著作権で保護するか。アルゴリズム・学習モデルの保護範囲（あるいは特許、プログラムの著作権として保護）、AIによる自動生成物は著作物か否か。
 - ・ 出版社におけるテキストデータマイニングの実例紹介。

(図3)

までを保護するのか。二つ目は、アルゴリズム・学習モデルの保護範囲をどうするか、特許やプログラムの著作物として保護するのか。三つ目は、人工知能による自動生成物を著作物とするかどうかということです。

出版社からは、エルゼビア社のヴァイスプレジデントの Anders Karlsson 氏にお越しいただき、エルゼビア社が提供している研究者向けのテキスト・データマイニングについて紹介していただきました。

3 回のセミナーを通じて、(1) 論文は著作物として著作権法で保護されるが、データは著作物ではないということ、(2) だからといって、何をやってもいいのではなく、契約によって条件を設定して公開できるということ、(3) 人工知能が絡んできて、知的財産権は今まさに議論が進んでいるところだ、と整理できました。

2.NIMS におけるデータ

データの著作権について整理できたところで、NIMS におけるデータ利用の実例を、作成者、利用者、提供者の三つの立場からご紹介します。

2-1.Case1

Case1 はデータの作成者としての関わりです (図 4)。研究データのメタデータ付与については以前から取り組まれていましたが、データ駆動型研究ではデータ共有がますます重要になるため、メタデータ作成の効率化・共有化が進められようとしています。

その一つが、計測データのメタデータ記録システムの開発です。NIMS 内の計測系研究グループで検討ワ

NIMSにおけるデータ：case 1 作成者として

- 研究データのメタデータ付与の効率化、共有化の取り組み
- 計測データのメタデータの記録システム

課題	対応方法
手書きのラボノート	データ採取時にデジタル化して記録
装置の操作で手がふさがる	音声入力
全てを入力するのは面倒	可能な箇所はシステムが自動計算
実験手順と入力項目の順番の相違	入力したい項目を入力したいタイミングで入力
装置が異なると、語句が異なる	語句の違いを吸収するためのテンプレート
メタデータと実験データの対応付け	メタデータ記録システムに実験データも保存

- 学会でも議論を共有して、課題解決に取り組む

(図 4)

ーキングが組織され、その議論から開発されたシステムです。研究者の立場からさまざまな工夫が凝らされています。

まず、計測データのメタデータを記録するときの課題として、実験時に手書きのラボノートしか記録していない場合、別途デジタル化作業が必要となります。そこで、計測装置を制御する PC から利用できるよう、ブラウザベースの入力インターフェースによって、データ採取時にデジタル化して記録できるようになっています。

また、装置の操作で手がふさがるため、音声入力にも対応していたり、全ての項目を入力しなくても済むように、可能な箇所はシステムが自動計算してくれるようになっています。

実験手順と入力項目の順番が違っていると画面を行ったり来たりしなければならないので、入力したい項目を入力したいタイミングで入力できるようになっています。また、装置が異なると語句が異なるため、その違いを吸収するためのメタデータテンプレートが用意されています。そして、メタデータと実験データ (生データ) を確実に対応付けるため、メタデータ記録システムの中に実験データも保存できるようになっています。メタデータ付与については、組織内だけでなく、学会でも議論が共有され、課題解決のため分野全員での取り組みが進められています。

2-2.Case2

Case2 は、データの利用者としての関わりです (図 5)。NIMS では MatNavi というサイトで、材料開発に

NIMSにおけるデータ：case 2 利用者として

～ データサイエンティストによるデータ収集 ～

- 物質・材料データベースサイト (MatNavi) > 高分子データベース
- 高分子材料設計に必要とされるデータの提供
- データ収集元：出版論文 (購読誌、OA誌) からデータのみを収集
- 作業工程

文献管理

データ収集・論文データ作成はデータサイエンティストとして研究OBが担当

データ収集

物質データ収集

データベース作成

その他ファイル作成

インターネット公開



<http://mits.nims.go.jp/>

(図 5)

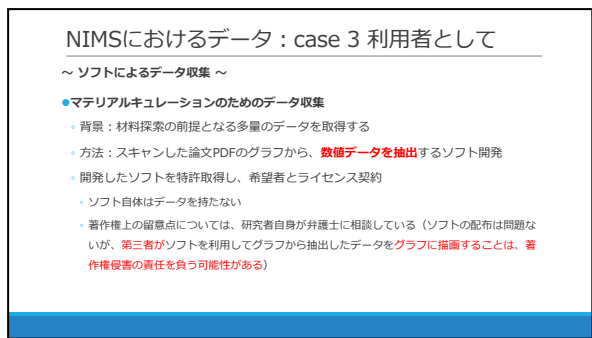
役立つさまざまなデータベースを公開しています。MatNavi はユーザー登録すれば無料で利用できます。今日は MatNavi で提供しているデータベースの中から、高分子データベース「PoLyInfo」のデータ収集についてご紹介します。

PoLyInfo は高分子設計に必要なデータを提供するデータベースです。データの収集元は、購読誌を含む出版論文で、その中からデータのみを収集しています。論文を収集元としている理由は、データの所在が明らかなこと、膨大な数が確実に出版されて公開されていること、必要とする情報が含まれていること等が挙げられます。

作業工程は、まず収集対象の文献を選定します。高分子分野は年に約 3,000 論文が出版されているようですが、スクリーニングによって 600～700 文献まで絞り込みます。次に、その文献から物性データを収集して、その物性データから構造検索などを可能にするための辞書データを作成します。その他の必要なファイルを準備して、高分子データベースとして公開しています。このうち、データ収集と辞書データ作成については、データサイエンティストとして、高分子研究 OB や有機化学研究 OB が担当しています。Case2 では、データサイエンティストによるデータ収集についてご紹介しました。

2-3.Case3

Case3 も利用者としての関わりです（図 6）。マテリアルキュレーションという手法によって、新しい材料探索を研究している方の例をご紹介します。それには



NIMSにおけるデータ：case 3 利用者として

～ ソフトによるデータ収集 ～

- マテリアルキュレーションのためのデータ収集
 - 背景：材料探索の前提となる多量のデータを取得する
 - 方法：スキャンした論文PDFのグラフから、**数値データを抽出**するソフト開発
 - 開発したソフトを特許取得し、希望者とライセンス契約
 - ソフト自体はデータを持たない
 - 著作権上の留意点については、研究者自身が弁護士に相談している（ソフトの配布は問題ないが、**第三者がソフトを利用してグラフから抽出したデータをグラフに描画することは、著作権侵害の責任を負う可能性がある**）

（図 6）

多くのデータを取得する必要がありますが、ボーナデジタルではない論文のデータも必要だったため、スキャンした論文 PDF のグラフから数値データを抽出するソフトを開発されました。同様の無料ソフトもあったそうですが、使い勝手が良くなく、効率的にデータ取得できなかったため、自分の研究用途に合うソフトを開発したところ、企業からも自社の古いデータをデジタル化したいというニーズがあり、特許を取得してライセンス契約することになったそうです。

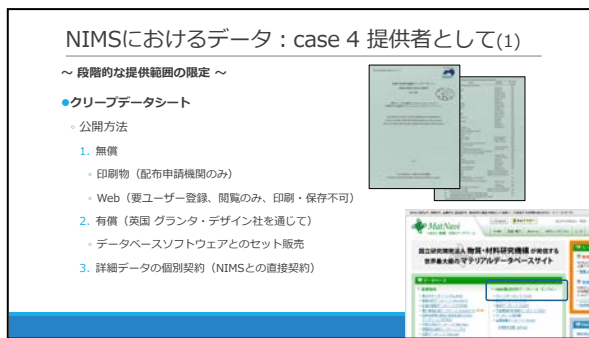
ソフトを提供する際の著作権上の留意点については、研究者自身が弁護士に相談しており、グラフからデータを抽出することは著作権法上の問題はないが、抽出したデータをグラフに描画することは、著作権侵害の責任を負う可能性があるため、ソフト利用者にもそのように伝えているそうです。Case3 ではソフトによるデータ収集についてご紹介しました。

2-4.Case4

Case4 は提供者としての関わりです（図 7）。MatNavi で提供しているデータシートのうち、クリープデータシートがどのように提供されているかをご紹介します。

その前にクリープとは何か、ごく簡単に説明します。クリープとは、物体に一体の荷重を加え続けることでゆがみが生じる現象のことをいいます。NIMS では高温下で金属を引っ張り続けたゆがみを測定していて、長い試験では約 40 年引っ張り続けて、世界最長記録を更新しました。

このクリープデータを「無償」「有償」「詳細データ



NIMSにおけるデータ：case 4 提供者として(1)

～ 段階的な提供範囲の限定 ～

- クリープデータシート
 - 公開方法
 1. 無償
 - 印刷物（配布申請機関のみ）
 - Web（要ユーザー登録、閲覧のみ、印刷・保存不可）
 2. 有償（英国 グランタ・デザイン社を通じて）
 3. データベースソフトウェアとのセット販売
 3. 詳細データの個別契約（NIMSとの直接契約）

（図 7）

の個別契約」という三つの方法で提供しています。無償提供は、印刷物とウェブの MatNavi です。印刷物は、配布希望の申請をした機関のみに配布していて、申請の際には利用目的も明記するようになっています。これはデータを営利目的で利用されないように管理するため、過去には配布を拒否したケースもあったそうです。ウェブの場合はユーザー登録により、利用を管理しています。ウェブ利用においては、閲覧のみで印刷や保存ができないようになっているため、無償提供の場合、デジタルデータは一切提供していません。

有償提供のデータは、無償提供しているものと基本的には同じですが、イギリスのグランタ・デザイン（Granta Design）社が開発しているデータベースソフトウェアとセットで有償販売されています。さらに詳細なデータが欲しいというユーザーとは、提供内容を個別に契約しますが、これは NIMS との直接契約になります。

提供方法によって、利用を認める範囲は異なります（図 8）。まず、無償配布の場合、データをプロットしたり、データ解析結果を公表することは可能ですが、データを営利目的で利用することは認めていません。二次利用をコントロールする目的もあって、利用は登録制となっています。

有償販売の場合、その範囲はグランタ・デザイン社との契約に基づくことになります。詳細データを個別に契約する場合も、その範囲はケース・バイ・ケースですが、データそのものを開示しないこと、第三者へ提供しないこと、データを基に解析した結果は公開してもいいという 3 点は、原則として共通しています。

NIMSにおけるデータ：case 4 提供者として(2)

●利用を認める範囲

1. 無償配布
 - OK：データをグラフにプロットする、データ解析結果を公表する
 - NG：データの営利目的利用
⇒ 配布/ユーザー申請時の理由によっては、拒否することもある
2. 有償販売
 - 利用範囲は契約に基づく
3. 詳細データの個別契約
 - データ開示禁止
 - 第三者への提供禁止
 - 解析結果の公表は可能

(図 8)

Case4 では、データ提供の範囲を段階的に限定している事例についてご紹介しました。

2-5. Case5

最後も提供者としてのケースです（図 9）。NIMS では情報統合型物質・材料開発イニシアティブ（MI²I）を進めています。これは JST のイノベーションハブ構築支援事業の採択課題の一つで、データ科学をこれまでの物質・材料科学に融合させることによって、新しい物質・材料科学の研究を加速する取り組みの場となることを目指しています。

活動の根幹を成すものの一つに、コンソーシアム活動があります。活動の概念図が右下の図です。データベースやツールを会員が使用することで、会員が持つコアな技術が拡大し、コンソーシアムで提供されるスクールや研究者のネットワークが技術の開発を促進します。技術が成長することで、雨粒のようにイニシアティブの中に落ちてきて、個別の共同研究を開始する契機になることを目指しています。

このため、研究コミュニティの発展に向けて、会員間でできる限りのデータ共有を目指しており、データ共有のためのプラットフォームを NIMS が提供しています。Case5 では、会員の利用規則という契約による、限定的なデータ提供の事例についてご紹介しました。

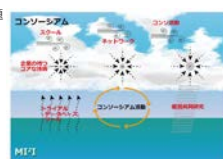
いずれのケースでも著作権法の範囲内とそうではないところを明確に区別して、著作権法の及ばない範囲については、契約でその利用範囲をコントロールしていることがお分かりいただけたかと思います。

NIMSにおけるデータ：case 5 提供者として

～ 限定的な提供範囲 ～

●情報統合型物質・材料開発イニシアティブ（MI²I）

- JST のイノベーションハブ構築支援事業の採択課題
- データ駆動型の物質・材料研究への取り組み
- コンソーシアム会員を募って活動
- 会員の利用規則に沿ったデータ利用
- データプラットフォームは NIMS が提供



コンソーシアム活動の概念図
<http://www.nims.go.jp/MI1-1/h6g4r00000002oc-att/h6g4r00000002pe.pdf>

(図 9)

3. 図書館業務とデータ

ここまで NIMS の研究者とデータとの関わりをご紹介しましたが、この中には図書館は関わっていません。データを扱うにはその分野の専門知識も必要ですが、図書館員がその必要な知識を習得するのを待っている時間的な余裕はありません。そこで、図書館員が今あるスキルでできることは何か、三つ挙げられるのではないかと思います（図 10）。

まず、図書館の伝統的業務である資料の保存と管理のスキルを生かして、研究者にデータの保存と管理に必要な情報を伝え、それを通じて図書館員もデータの扱いを学ぶということです。これについては、この後、尾城さんから詳しくご紹介いただけるかと思います。

次に、機関リポジトリのノウハウを生かして、研究者に保存と提供のプラットフォームを用意し、その対話の中で図書館員も必要なプラットフォームについて理解を深めるということです。NIMS では、文献をセルフアーカイブするリポジトリとは別に、マルチメディアデータをセルフアーカイブする imeji というリポジトリを研究者に提供しています。imeji ではメタデータ項目を研究者自身が設定できるので、拡張的なメタデータ管理ができるようになっていますが、先にご紹介した研究者自身が開発したメタデータ記録システムと共同するという考えられると思います。

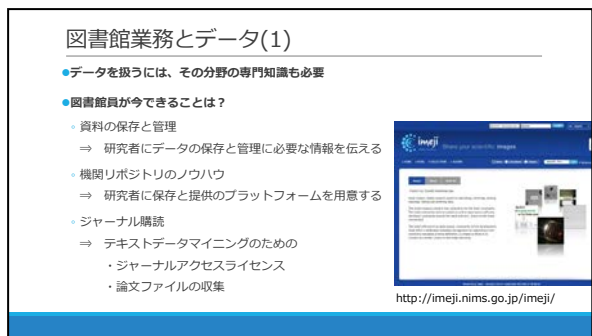
そして、ジャーナル購読のノウハウの利用です。データ駆動型研究では、どれだけのデータを集められるかがポイントとなるので、テキスト・データマイニング用に購読とは異なるジャーナルアクセスのライセンスや論文ファイルの収集が必要となります。NIMS で

つい最近、そうした事例があったのでご紹介したいと思います。

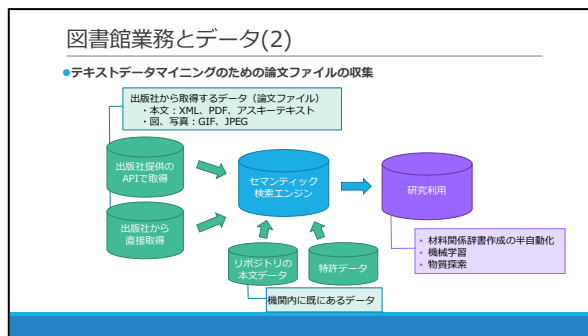
出版社とのコネクションを持っているのが図書館なので、データマイニングのための出版社とのライセンス契約を図書館を通じて行いましたが、出版社によって提供可能な範囲はさまざまでした（図 11）。まず、購読ジャーナルしか認めないというところ、研究目的で NIMS 内だけで使うなら提供できるというところ、研究用途として実験的に無償提供するところ（学会）など、さまざまでした。マイニング用のデータなので、論文 PDF のように美しく成形されている必要はなく、こちらの希望としては本文データと図や写真のデータだけを入手できれば足りるのですが、出版社によって、API 経由だったり、出版社サーバーから直接取得だったり、それぞれ可能な方法でデータを収集しました。

取得したデータを知識ベースとして、セマンティックな検索エンジンに投入しますが、データをよりリッチにするために、機関内で既にあるデータ（リポジトリの本文データや特許データ）についても、このセマンティック検索エンジンに投入します。それを材料関係辞書作成の半自動化や、機械学習、物質探索に利用することを予定しています。

契約交渉を通じてあらためて感じたのは、ライセンス契約の範囲設定の難しさです（図 12）。出版社から論文データを購入する場合、テキスト・データマイニングの成果を機関内にとどめておくのか、それとも第三者に提供するのかで、料金が大きく変わってきます。また、ソフトウェア会社との契約でもセマンティック検索エンジンにより生成した辞書データをソフトウェア



(図 10)



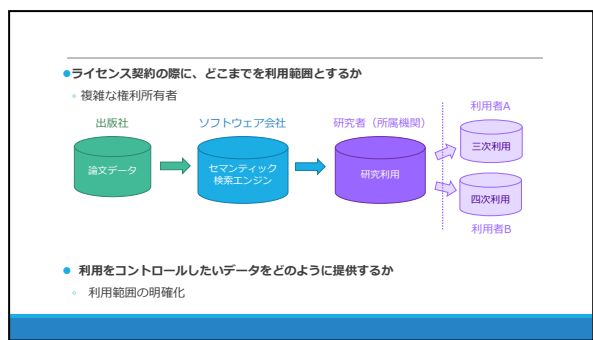
(図 11)

ア会社に提供するかどうかで、やはり料金が変わってきます。ソフトウェア会社に辞書データを提供することは、出版社のデータを第三者に提供することになるので、出版社が第三者への提供を認めていない場合はできません。

NIMS では今回、研究で利用するために NIMS 内だけで限定利用することを前提として、細部については、その都度確認しながら進めるということで、出版社からデータを購入しました。NIMS の中だけでの利用、つまり、二次利用までのライセンスですが、研究成果を公開して、三次利用あるいは四次利用というのを考えると、権利の所有者はどんどん複雑になって、ライセンスの内容も、都度、協議するというわけにはいかなることが想定されます。

今回のまとめとして、材料科学分野ではオープンにできるデータとそうでないデータがあり、オープンにできない、つまり、利用をある程度コントロールしたいデータの提供方法の検討ポイントは、利用範囲を明確にすることではないかと思っています。

ライセンスについてはまだまだ勉強中で試行錯誤しているところなので、もし、そういうセミナーを開催される際には、ぜひ教えていただけるとうれしいです。



(図 12)

●角田 JAMSTEC の角田と申します。著作権法をよくよく調べられていて、よく分かりました。ありがとうございます。ただ、そもそも著作権法をいくらじったところで、多分、データそのものをどうこうするのは全く別の話になってしまうのではないかというのが、私がまた別途、弁護士さんに聞いたときの感じでした。

著作権法というのは、中身が事実であろうとシミュレーションであろうと全く関知しないし、中身が面白かろうがつまらなかろうが、陳腐だろうが独創的だろうが、それも全く関知しません。中身をどう表現するかという表現法が独創的かどうかで保護するというのが著作権法の本来の精神なので、データそのものを著作権法でどうこうしようとしても、それは本質的に無理で、データをどう表現するかに関して、びっくりするような表現方法、独創的な表現方法があるのであれば、それは著作権法の保護の範囲内になるかもしれません。

では、論文はどうかと皆さん思われるのですが、もし論文が本当に定型化されて書かれていて、しかも、どの AI でも、体裁、表現が同じような論文が書けるということであれば、論文の中身がいくら独創的であっても著作権法では保護されないというすごい状況が起きるので、データそのものをどう保護するかということになった場合には、もう著作権法とは全く関係なしにライセンス契約なり何なりを考えるのがいいのではないかと思った次第です。

●小野寺 おっしゃるとおりで、データは著作権で保護される範囲ではないということは、NIMS で開催したセミナーでも弁護士の先生が繰り返しおっしゃっていたことです。データというのは客観的な事実を示したもののなので、そこに独創性や創造性は関わらないということを学んだところです。

●**角田** 全くそのとおりなのですが、データにはシミュレーションデータというものもあります。人間が本を書いたり、絵を描いたりするときも、事実をそのまま写すというものと、シミュレーションしたり、頭の中にあることを事実に基づかずに表現するということもありますよね。そういうこともありますのでということを一言付け加えさせてください。

●**小野寺** ありがとうございます。

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

研究データ共有の理想と現実、そして実践可能性 ～地球環境分野の研究基盤に関する意識調査から～

小野 雅史

(東京大学地球観測データ統融合連携研究機構)

講演要旨



「研究データの共有は科学の発展に貢献する良い考えだ」という理念については、研究者のみならず、政策立案者、資金提供者、市民等、多くのステークホルダーが基本的には同意するはずである。しかし、実際には、多くの研究データが、共有または公開される状況に至っていないのが現状である。このオープンサイエンスの理想と現実の間には、どういった課題が存在するのだろうか？これを明らかにするために、我々は地球環境情報分野の現場の研究者を対象として意識調査を実施した。そこで、本講演では、この意識調査の結果をもとに、我々がオープンサイエンスを前に進めるために実現可能な取り組みについて、考察した内容を紹介する。



小野 雅史

東京大学地球観測データ統融合連携研究機構 (EDITORIA) の特任研究員。DIAS (データ統合・解析システム) という研究データ基盤を中心とする事業「地球環境情報プラットフォーム構築推進プログラム」に参加している。過去に、地理情報標準ISO/TC211の仕様検討委員、GEO (Group on Earth Observations) のオントロジータスクチーム、Belmont Forum E-Infrastructures and Data Management Collaborative Research ActionのData sharingグループのメンバーとして活動。

本日のセミナーのテーマは「オープンサイエンス再考」ですが、私の報告のタイトルにある「研究データ共有」は、オープンサイエンスの要素の一つです。研究データ共有の理想、広義にはオープンサイエンスの理想といったときに、皆さんは恐らくこういうものをイメージされると思います。「研究者がみんなデータを提供して、それをみんなでシェアして、それが科学の発展を促進し、ひいては社会への貢献に資する」。これは非常にいい考えです。みんないいことだと思っています。でも、その一方で現実はどうなっているのでしょうか。

実はこれをとてとても簡単に知る方法があります。例え

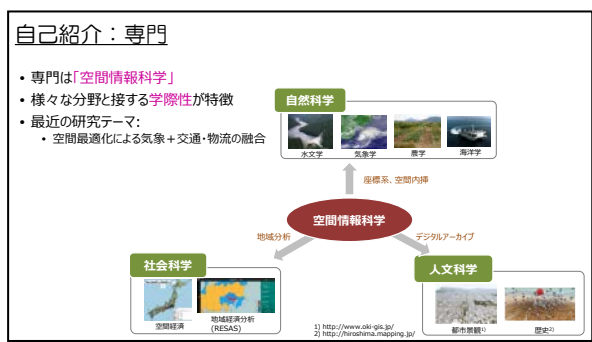
ば、誰でもいいので研究者を捕まえて、こういう質問をしてみてください。「先生、あなたが持っている研究データのうち、何パーセントぐらいを他の人とシェアしていますか」。これは言うまでもなく、非常に歯切れの悪い答えが返ってくると思います。これはオープンサイエンス推進派といわれている先生でも、そんなに変わりはないと思います。つまり、こういうことなのです。みんないいことだと思っているのに現実にはそうっていない。ですから、この理想と現実の間にどういった空白があって、その隙間を埋めるために何をしたらいいのか、そのためのヒントになるような話を今日はできたらと思っています。小賀坂さんや小野寺

さんのお話とも非常に符合する部分があるので、思い出しながら聞いていただけたらと思います。

自己紹介

私の専門は空間情報科学です（図 1）。この分野の特徴は、自然科学だけでなく、社会科学、人文科学などいろいろな分野と接することが多い学際性です。ある分野とある分野を融合させるということも比較的やりやすくて、最近の私の研究テーマは気象と交通・物流の融合です。研究とは全く関係ないのですが、その他、色、調味料といった民間資格を持っています。私は料理が結構好きなので、家に調味料が 80 種類以上あって、それを組み合わせて毎回新しいカレー粉をつくる、そんなことをしていたりします。一見これは全然関係ないように聞こえるかもしれませんが、分野の融合のように、異質なものを組み合わせて何か新しいものをつくるのが好きな人間、それが私です。

こうした私自身の特性を生かしながら、これまでさまざまな異分野の融合関係のプロジェクトを実施してきました（図 2）。



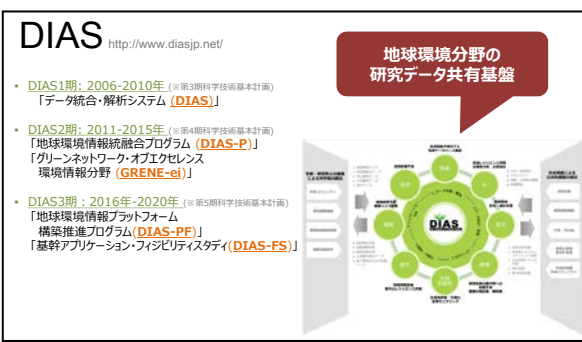
(図 1)

一例として、ISO/TC211 の geographic information standards に関する仕事、国内の地球環境データ統合の Data Integration & Analysis System (DIAS: データ統合・解析システム) というプロジェクト、DIAS の国際版のような Group on Earth Observation (GEO: 地球観測に関する政府間会合) のオントロジー等のタスクグループでの仕事、ベルモント・フォーラムのデータシェアリンググループのメンバーとしての活動などを行ってきました。

Data Integration & Analysis System (DIAS)

今でも私のメインのプロジェクトは DIAS です（図 3）。DIAS とは、地球環境分野の研究データ共有基盤のことです。DIAS という名前が使われたのは、第 3 期科学技術基本計画の 2006 年からで、ちょうど今年から DIAS 第 3 期に入りました。

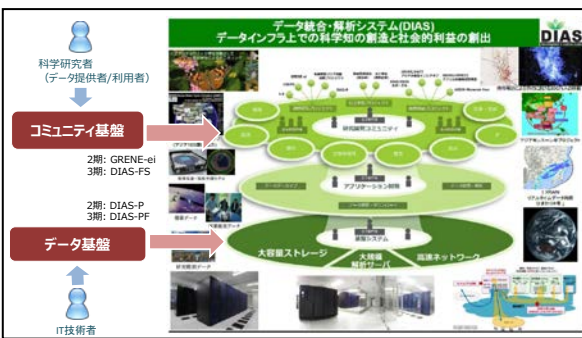
この DIAS の特徴を一つ挙げると、過去の SPARC Japan のイベントで、北本先生からお話があったと思いますが、コミュニティ基盤とデータ基盤の二つのレイヤーを持っていることです（図 4）。このデータ



(図 3)



(図 2)



(図 4)

基盤を支えるのは主に IT 技術者が中心で、コミュニティ基盤の中核となるのが、いろいろな分野のサイエンスの研究者です。このコミュニティ基盤の上で地球環境に関する課題を取り上げ、その課題であれば私はこういうデータを持っている、そのデータを使わせてくれれば私はこういう分析ができるというような議論をしながら、研究を進めているという枠組みになっています。

このような形で研究を進めていくことに力を入れているのは言うまでもなく、その他にも力を入れている領域があり、その一つが教育です（図 5）。東京大学での講義、サマースクールなどの企画、アジア地域でフィールドスタディなどを行っています。

他に産学連携もしています。図 6 は河川のダム操作支援システムの事例です。例えば、東京大学がモデルや手法を考案して、電力会社は電力データを出して、日本工営などの建築コンサルはシステムのオペレーションをするという枠組みで共同作業を行います。

このように民間企業との仕事経験を蓄積して、今期の DIAS の第 3 期で力を入れている領域がビジネ

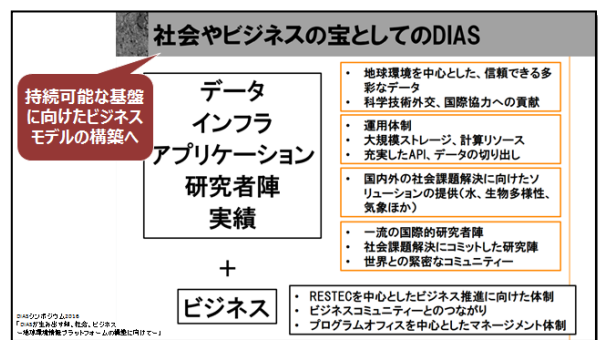
スモデルの構築です（図 7）。なぜ研究基盤がビジネスなのかと思うかもしれませんが、キーワードは「サステナビリティ」です。国家財政もこれから緊縮していく中で、やはりデータ基盤は自分で独自に継続していくことを考えなければいけないのです。European Open Science Cloud でもサステナビリティの議論はいろいろ出ていて、コマーシャルサービスの利用も挙げられていると思うのですが、今、DIAS でもそちらの検討に力を入れています。非常にチャレンジングな課題ですが、いろいろ考えているところです。

ただ、ビジネスモデルをつくるといっても、別に DIAS のメンバーが MBA を取りに行くというような話ではなく、ポイントになるのは人だと思っています（図 8）。つまり、DIAS の考えにどれだけの人が乗ってきてくれるか、それがポイントになると思うので、潜在ユーザーがどういうところにいるかを調査しながら、リーチする活動をしています。

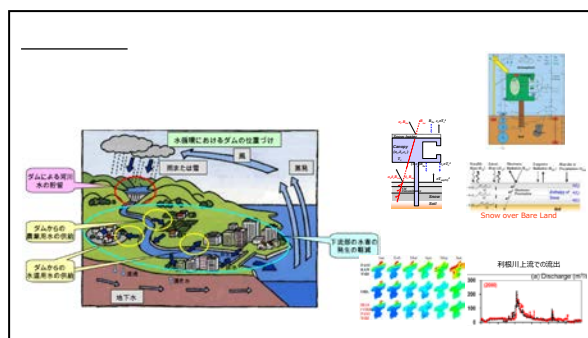
このように、いろいろな活動を続けているのですが、当然、全てが順調にいくわけではありません。中にはデータ利用者側のニーズとデータ提供者側のシーズが



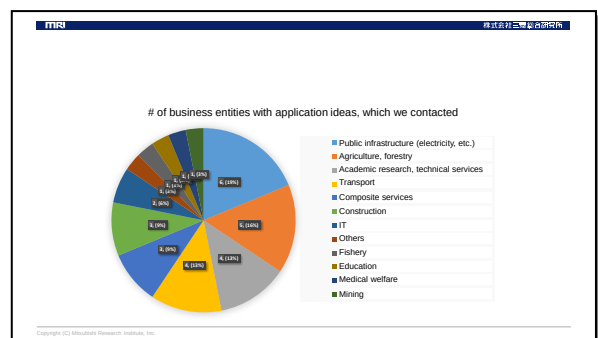
(図 5)



(図 7)



(図 6)



(図 8)

うまく折り合わないといったことが出てくるという問題があります。

データ共有とデータ公開

問題の詳細に入る前に、データ共有とデータ公開の二つのコンセプトについて説明します（図9）。

データ共有（data sharing）の定義は、特定の個人・組織間でルールを決めてデータを提供することです。一方、データ公開（open data）とは、特別な制限なしで任意の他者にデータを提供することをいいます。

これはオープン・データ・イニシアティブや、ベルモントでも同じような意味で使っていたので、国際的にも通用する定義だと思います。オープンサイエンスというのは、広い意味でこの両者を包含する概念であると考えます。

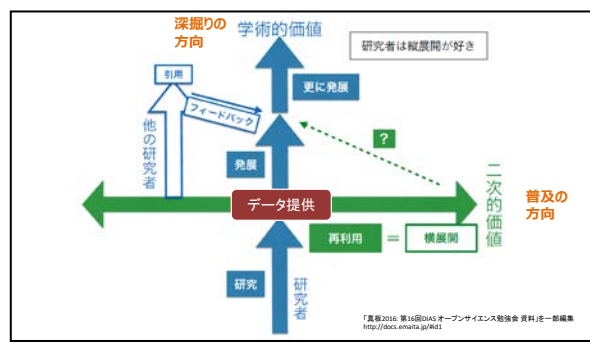
こうして比較してみたときに、ひょっとすると皆さん、データ公開の方が制限がない分、いいことなのではないかと思えるかもしれませんが、一概にそうとは言えないのです。なぜなら、データ提供者の視点から見ると、利用者からのフィードバックを得やすい枠組みは、どちらかというとデータ共有の方だからです。

これは少し考えれば分かると思うのですが、特定の範囲でデータだけではなくルールもシェアしてデータを提供し合うという枠組みなので、相手の姿もイメージしやすいですし、コミュニケーションが取りやすいのです。一方で、データ公開の方は、任意の第三者なので要は相手が誰かも見えにくい、あるいはどこで使われるのかも分からない、そもそも連絡が取れるのかも分からないというような問題があって、フィードバ

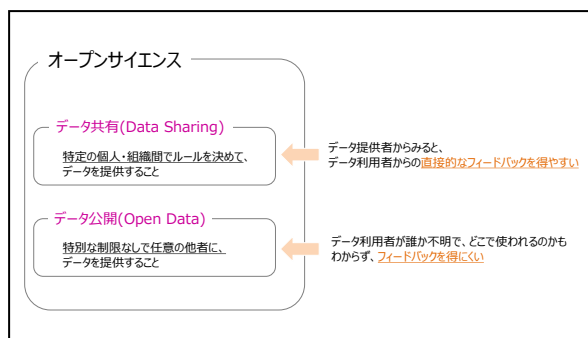
ックを得にくい枠組みです。

このフィードバックを得られるかどうかは、研究にとって非常に重要な要素です。図10は、DIASの勉強会で国立環境研究所の真板英一さんがベースをつくられた図です。平たく言うと、研究には普及の方向と深掘りの方向の二つの軸があるということです。尖った研究者で、一つの研究テーマをどんどん深掘りしていくタイプは、縦軸の方向に進みます。でも、オープンサイエンスの基本戦略は、どちらかというと横軸の普及の方向性です。

つまり、オープンサイエンスは、オープンにしてみれば誰かが使ってくれて、何か面白いことが起こるかもしれないというような割に楽観的な推論に基づいているのです。確かに誰かが面白い使い方をしてくれるかもしれないのですが、一つの研究テーマを深掘りするタイプにとっては、それは直接的にはあまり関係ないのです。ですから、そういうタイプの人にはどちらもフィードバックが非常に重要で、これが不可欠になります。こうした観点で見ると、DOIやデータサイテーションはインセンティブの本丸といわれていますが、



（図10）



（図9）

フィードバックという観点から考えると

- DOIやデータサイテーションは、データ提供のインセンティブになるか？
- 被引用回数の増加による効果：
 - 業績が主目的で、研究成果は手段というタイプに対しては、強いインセンティブになる
 - 自分の研究の深化が目的で、業績を気にしない（楽しく研究したい）タイプの場合、実はあまり関係がない
 - もちろん、（数は少なくても）自分と同等以上の能力を持つ研究者からの、直接的なフィードバックのほうがインセンティブになりえる

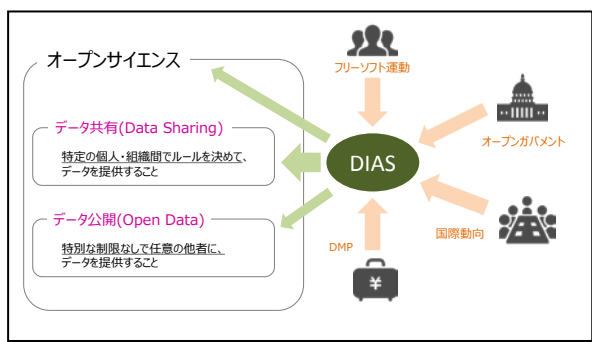
（図11）

これがインセンティブとして効く相手と効かない相手がいるということも分かってきます（図 11）。

例えば、業績が主目的で、研究成果は手段というタイプに対しては、強いインセンティブになります。その一方で、自分の研究を深掘りするのが目的だったり、他の人がどう思っているかは関係ない、私は楽しい研究がしたいのだというタイプに関しては、実はあまり関係ありません。そういうタイプにとっては、自分の分野で、かつ、自分と同等以上の能力を持つ研究者からの一言という方が、強いインセンティブになるのです。このような議論をわれわれ DIAS でしてきたので、DIAS はデータ公開よりはデータ共有の方をアイデンティティとしてやってきたという経緯があります（図 12）。

意識調査の実施

このような中、オープン化に対するプレッシャーがだんだん強くなってきて、その方向性も考えなければいけない、対応していかなければいけないという状況になったので、今後の指針を得るために調査を企画し



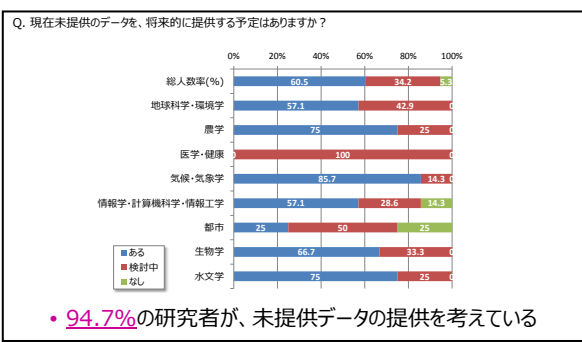
(図 12)

ました（図 13）。この詳細については「情報管理」11月号に掲載されているので、そちらをご覧くださいと思います。今日はその調査の一部と、そこから分かったことを中心にご紹介します。

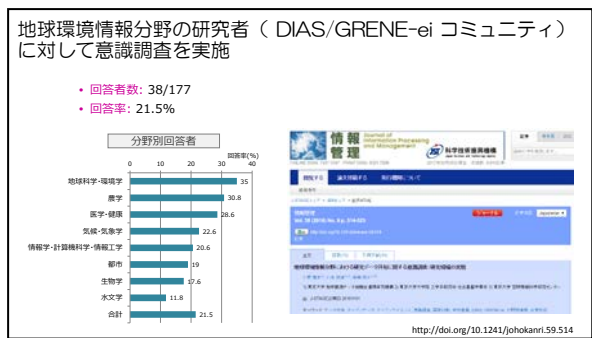
まず、研究者はそもそもデータの提供に反対しているのかという素朴な疑問があります。それに対する答えが図 14 で、要はそんなことはないという結果でした。実に 94.7%の研究者から、まだ出していないけれど、提供の予定がある、あるいは検討しているという回答が得られています。

次に、そんなに考えてくれているのなら、なぜそれを出してくれないのかという疑問が沸き起こると思います。

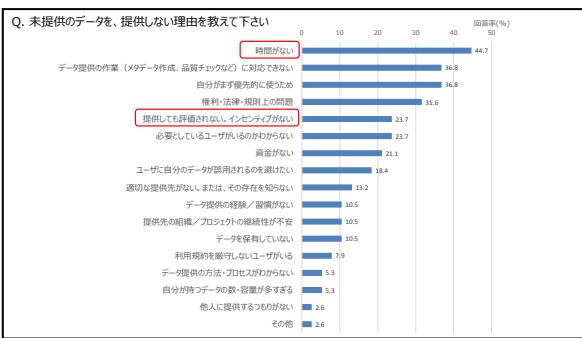
それに対する答えは、「時間がない」ということがやはり一番強い理由になるかと思えます（図 15）。この理由が「インセンティブがない」という理由の倍近くを占めています。ここから、仮にインセンティブとして、DOI、データサイテーションが進んだとしても、何らかのサポートがないと思うとおりにデータの提供は進まない可能性があるということが見えてきます。



(図 14)



(図 13)



(図 15)

誰にだったら提供してもいいのかという疑問に対する結果は、図 16 です。青い方がメタデータに対する答えで、赤い方がデータに対する答えです。上の方に行くほど公開の範囲が広めで、下の方に行くほど狭めという整理の仕方をしています。やはりメタデータに関しては別に任意の人に見せてもいいけれども、データに関しては、「関係する分野だったらいいけれど、ちょっと公開は」という結果が見えてくると思います。

続いて、いざ提供するとき、データ共有あるいはデータ公開するときの条件はどういうものがふさわしいかという質問をしました。

図 17 が、データ共有の結果です。左側が私の調査の結果、要は国内の結果で、右側が私が調査の参考にした DataONE という国外のプロジェクトの結果です。青いバーが、データ提供側としてこういう条件を入れてほしいという結果です。赤いバーが、データ利用者として、その条件であればのんでもいいという結果です。

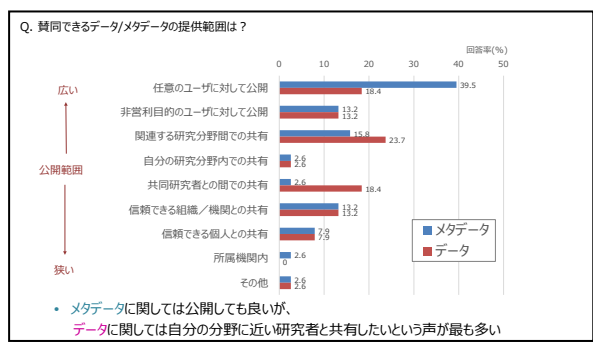
この結果から面白いことがいろいろ分かりました。三つだけポイントを言うと、一つ目に、国内の結果で

も国外の結果でも同様に、データ共有の条件として、そのデータを使ったということを論文で引用すること、データ利用規約を守ることがトップの方に来ています。

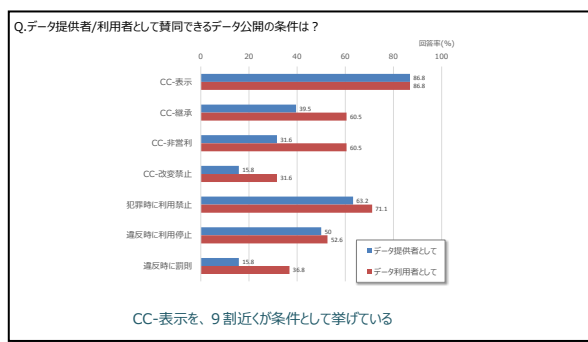
二つ目は、国内の結果では、データの提供者よりも利用者として受け入れ可能な条件が多い、つまり、データを使わせてくれるのなら、その条件は少々厳しくてもむという結果が出ているということです。

三つ目に、プロジェクトへの参加機会を提供するという条件で、国外の結果と国内の結果で顕著に差が出たことです。つまり、あなたのプロジェクトに参加させてくださいというデータ共有の条件を、海外では重視している、国内では重視していないという結果です。これはどういうことかという、データ共有をコラボレーションのチャンスとして見る向きが海外では強く、国内では弱いという結果なのではないかと解釈しています。

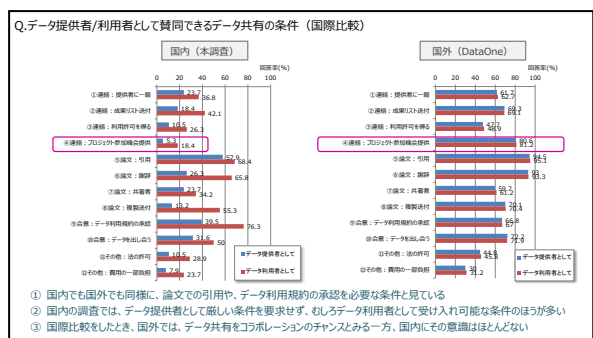
続いて、図 18 はデータ公開の条件の結果です。クリエイティブ・コモンズ・ライセンス (CC ライセンス) は表示がマストだという回答が 9 割近くあることがまず分かりました。



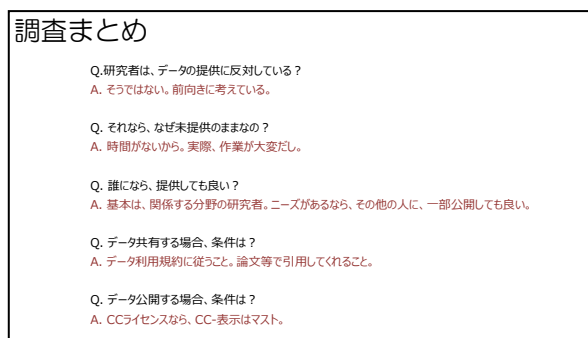
(図 16)



(図 18)



(図 17)



(図 19)

CC ライセンスは、それを守らなかったときも特にペナルティがなく、フリーライディングが起きやすいという問題が指摘されていました。ではペナルティとしてどういうものが適切なのか、私が個人的に興味を持ったので、そのような質問を入れてみました。それによると、罰則まではいかなければ、利用停止処分はしたいという結果が見えました。

私の調査を FAQ 形式で簡単に整理すると、図 19 のような結果になりました。この調査をやってみて、出てきた結果の全てをまだ消化しきれていないところがあって、それに対する具体的なソリューションがすぐには出てこない状況です。しかし、これが研究者の現在の考えであるということはきちんと受け止めて、今後何をするかを考えていく必要があると思っています。

私たちはこれから何をすればいいのか

これから何をすればいいのか、考えたことを、簡単にご紹介したいと思います。

まず大前提として、焦って間違った方向に行ったり、強制力のあるルールをすぐ採用したりしない方がいいのではないかと思います。最近、特に気になっているのは、オープンサイエンス系の議論だと、とにかく海外は進んでいて日本は遅れているから早くやらなければいけないという論法を使う人が結構いることで、これに私は違和感があります。

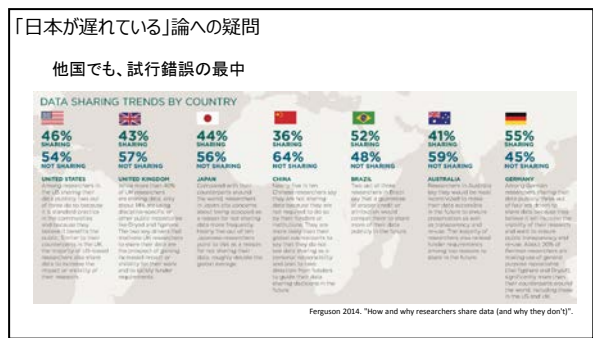
というのは、私の経験からも、確かに海外の方がいろいろなチャレンジ、トライアル、新しいコンセプトが出てはいるのですが、実際に全ての研究者の実態を調査すると、全分野に進んでいる人も遅れている人も

いて、その統計結果を出すと、そんなに変わらないのではないかという印象があります。

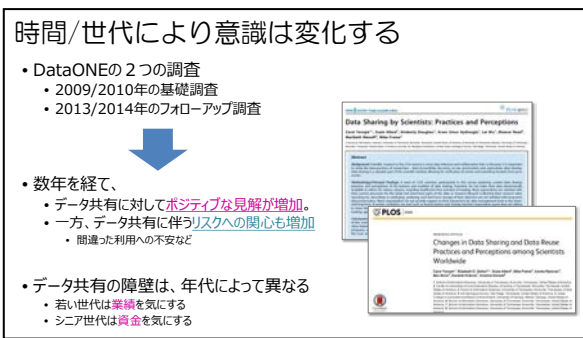
それを裏付けるかのような調査結果も出ていて、例えばイギリスはオープンデータですごく進んでいるというイメージがあるかもしれませんが、実際には日本と、sharing と not sharing のレートはほとんど変わらないのです（図 20）。ですから恐らく、海外でも進んでいる人は進んでいるけれど、全体としてはそんなに変わらないのです。日本でも海外でも同様に試行錯誤の最中であるというのが実態に近くて、国別に進んでいる、進んでいないというよりは、ライフサイエンスのように進んでいる分野もあれば、全然まだ進んでいない分野もあると言った方が、私にとっては実感として腑に落ちる感じがします。

また、焦って何かをやらない方がいいということをや裏付けるかのように、時間がたてば、やはり意識は変化するということが分かってきています。DataONE が基礎調査とフォローアップ調査を行い、数年たつて研究者の意識がどう変化したのかを調べた有名な論文では、やはり研究データ共有に対してポジティブな見解がだんだん増えていったという結果が出ています（図 21）。ただ、その一方でリスクへの関心も増加しました。これは私なりに解釈すると、最初にいいことばかりを言い過ぎて、現実の方があまり追い付いていないということがあったので、この落差を見て、本当に大丈夫なのかという不安が増加したのではないかと思います。

ですから、何か結果を出そうと焦ってやるよりは、北本さんからスピードの話がありましたが、（物事



(図 20)



(図 21)

を進めるときには) 適切な速度というものがあるので、そこを見極めながら進めていく必要があるのではないかと思います。

この論文は他にも面白いことが書いてあります。やはり年代によって考え方が違って、若い世代は業績を気にする、シニア世代は例えば教授だともう地位は安定しているので業績はあまり気にせずにプロジェクトで予算が取れるかどうかを気にするというので、インセンティブは年代によってもだいぶ違うのではないかと思います。

ということで、焦ってやるよりは継続の方が重要です。DIAS もこの名前で活動しているのは 2006 年からですが、30 年以上かけて今のところまで来ています (図 22)。

オープンガバメントとオープンサイエンス

その他に、オープンガバメントの文脈をオープンサイエンスに持ち込んで、例えば政府標準利用規約に従って研究データをオープンにしろと言うポリシーメーカーの方がたまにいますが、この二つは全く違う概念だということを説明しておこうと思います (図 23)。



(図 22)

オープンガバメントの文脈をオープンサイエンスに転用するのは良くない

- **オープンガバメント: 公共データのオープン化**
 - Open by default
 - 公共データは、一般知識があれば理解できる。
 - ユーザの数が多。
 - 政府は、市民にデータ公開を要求されたら従う立場 (Public servant)
- **オープンサイエンス: 研究データのオープン化**
 - Ideally open
 - 研究データを扱うには専門知識が必要。
 - ユーザが少ない。
 - 公的資金で作成したデータがオープン? ⇒ 研究者は、Public servantなんですか?

(図 23)

例えば、公共データの人口統計データを考えたときに、これは別に特別な知識がなくても使えます。その一方で、研究データは誰が使うのか。先ほど河川の例を出しましたが、水循環の分布型流出モデルのチューニング化された初期値で使うパラメーターセットを誰が使うのだということです。こういうものを第三者にオープンにすることは、あまり意味がないのではないかと、その辺のバランスも考えながら進めなければいけないのではないかと思います。

もう一つ重要な論点として、立場の違いもあると思います。例えば、民主主義の基本原則から言って、市民がデータを公開しろと言えば、政府はパブリックサーバントなので出さなければいけないのです。でも、公的資金で作成したデータがオープンという論法は、国に言うのなら分かるのですが、研究者に言うのは適切なのか、私の方でも違和感が残っています。研究者はパブリックサーバントなのかということです。

これは突き詰めて考えると非常に深いテーマだと思います。やはり市民は間違えることがあるのです。そういうときに、こちらの道が正しいのだと誰が言うのか。その役割を負っているのが研究者ではないか、だから研究者はパブリックサーバントというよりは、ファクトに対するサーバントであるべきだと思っています。いくら国や市民が言ったとしても、それを全て鵜呑みにするのは、民主主義を正常に回すために良くないのではないかと考えています。

データ・マネジメント・プラン (DMP)

私も海外で実際につくられているデータ・マネジメント・プラン (DMP) はどんなものか見てみたのですが、文章で説明して書いてあるものから、データの細かいスペシフィケーションを書いてあるものまで、まちまちでした (図 24)。これをつくることにどのくらい実効性があるのかという疑問が出てきて、きちんと検証した方がいいのではないかと思います。しかし、小賀坂さんの報告を聞くと、その辺は慎重に丁寧にされているということなので、安心しました。逆に、

日本のアドバンテージは、いったんルールが決まると丁寧なやることだと思うので、国内で丁寧な作業、仕事をして、日本はうまくいっているという事例を見せられたらいいのではないかと考えています。

私たちはこれから何をすればいいのかをまとめると、
すぐに結果を出そうと焦らない、やるべきことは継続
する、やること（やったこと）は有効だったかどうか
検証することではないかというのが私の考えです。

文献とデータの比較

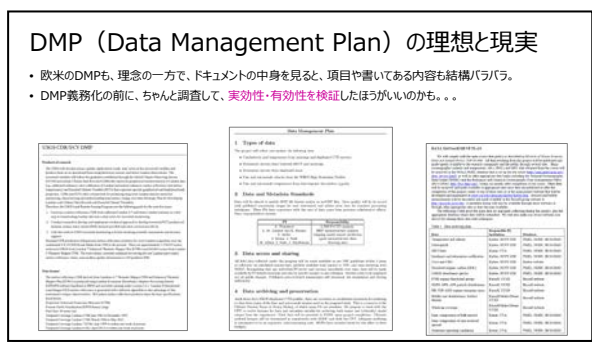
最後に、SPARC Japan セミナーは図書館の参加者が多いイベントで、図書館は、文献管理で得たノウハウが研究データに転用できるのか悩んでいる方が多いのではないかと予想したので、文献の世界とデータの世界を少し比較してみました。

まず、文献と研究データのボリューム感は、国立国会図書館の今の蔵書数が約 4,100 万であるのに対して、DIAS の公開データが昨年度時点で約 5,800 万でした（図 25）。DIAS の本体は非公開領域なので、それを入れると数倍以上に膨らむかもしれませんが、このぐ

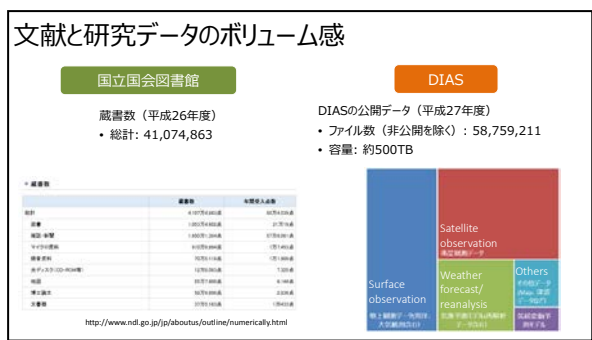
らのボリューム感です。これだけのボリュームがあると、メタデータをどうやって付ければいいのか悩ましいところで、適切なメタデータの粒度に悩んでいるところです。皆さんも、国会図書館のメタデータをもう一回ゼロから付け直すと言われたら、かなりうんざりで、それは無理だと思いますよね。そういうデータの世界で悩んでいるのがわれわれだということです。メタデータの粒度問題について本当に悩んでいるので、もしいいアイデアがあれば教えてほしいです。

図 26 は大きさの比較です。世界一大きい本は通称『オバママニア』という本で、厚さが約 34cm だそうです。世界一大きい画像データはアルプス山脈の図で、約 46TB あるらしいです。

言いたいことは、データの方が大きいとかそういうことではなく、大きい本は子どもでも何とか持ち運べるサイズである、でもデータは移動させるだけでも特別な環境がないとできない、こういう違いがあるということです。この違いは結構大きいのではないかと思います。



(圖 24)



(圖 25)



(圖 26)

まとめ

最近のオープンサイエンスの議論は抽象論、技術論に引きずられている印象があります（図 27）。そこにはデータ提供者とデータ利用者という重要なプレイヤーに対する視点が抜け落ちていると感じています。データ提供者と利用者は、文献の世界で言う、著者と読者の関係です。その著者と読者を歴史を越えてつなげていくという役割が図書館にあるのではないかと思います。ですから、データ提供者とデータ利用者をどうつなげていくか、そのアーキテクチャをどうつくっていくかという視点で今後を考えていただければうれしいです。データキュレーター、データライブラリアンの役割もそこにどう関われるかがポイントになるのではないかと考えています。

文献の著者・読者⇒データ提供者・利用者

- 最近の「オープンサイエンス」の議論は、「オープンアクセス」、「オープンイノベーション」などの抽象論や、「DOI」や「メタデータ」などの技術論が目立つ。
- そこには、「データ提供者」と「データ利用者」という、重要なプレイヤーに対する視点が抜け落ちている

図書館は、時代を超越しようとする著者と読者との出会いを支えてゆく重大な役割と責任を果たしていかなければならない

飯島 利雄（早稲田大学図書館長）「時代を越えて生きるために——著者、読者および図書館の責任」

文献の世界の著者・読者の関係が、研究データの世界のデータ提供者・データ利用者の関係。

データ提供者をリスペクトし、データ利用者のニーズを汲みながら、データ提供者とデータ利用者をつなぐ場をどう作っていくか、そこから考えてみましょう。

（図 27）

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

研究データを用いたサービスの調査・企画

A 班：機関内研究情報統合システム(IRIIS)の企画構想

B 班：オープンデータ利活用総合基盤システム「ODEN」の企画

大向 一輝 (国立情報学研究所)

田村 峻一 (滋賀医科大学附属図書館)

梶原 茂寿 (北海道大学附属図書館)

講演要旨

国立情報学研究所では教育研修事業の一環として、様々な機関に所属する図書館員が協力し、学術情報流通に関する諸課題の解決を目指す学術情報システム総合ワークショップを毎年開催している。今年度は8名の受講者が「研究データを用いたサービスの調査・企画」をテーマとして約6ヶ月に渡る検討を行った。本講演ではワークショップの概要と目的、その効果について述べる。参照：平成28年度カリキュラム及び講義資料/成果物(<http://www.nii.ac.jp/hrd/ja/ciws/h28/curritxt.html>)



大向 一輝

1977年京都生まれ。2005年総合研究大学院大学博士課程修了。博士(情報学)。2005年国立情報学研究所助手、2007年同助教、2009年同准教授。セマンティックウェブやソーシャルメディア、オープンデータの研究とともに、学術情報サービスCiNiiの開発に携わる。株式会社グルコース取締役。著書に『ウェブがわかる本』(岩波書店)、『ウェブらしさを考える本』(丸善出版)がある。

A 班：機関内研究情報統合システム(IRIIS)の企画構想

当班では、研究データを含めた研究資源や研究に関する情報の管理を、機関単位で集約的に行うことを想定し、「機関内研究情報統合システム(略称 IRIIS)」を企画構想した。企画にあたっては、機関内の情報を網羅的に収集できること、および簡便に業務が行えることを重視した。今後このような仕組みが標準化し普及することで、より多くの機関が研究データ管理に取り組みやすくなることが望ましい。



田村 峻一

2013年、滋賀医科大学に図書系職員として採用。以来、図書受入業務及び機関リポジトリ業務を担当している。日本医学図書館協会認定資格ヘルスサイエンス情報専門員(基礎)。平成28年度学術情報システム総合ワークショップ受講者。

B 班：オープンデータ利活用総合基盤システム「ODEN」の企画

B 班では、散在するオープンデータのメタデータを一元的に集約し、ワンストップでの検索を可能にするオープンデータのメタデータデータベース「オープンデータ利活用総合基盤システム - ODEN」の構築を企画した。本講演では、そのコンセプトと概要について紹介する。



梶原 茂寿

2014年より、北海道大学附属図書館本館に勤務。学術システム課システム管理担当で、図書館情報システム及びネットワークの管理と機関リポジトリを担当。2015年、2016年SPARC Japanセミナー企画WGメンバー。

研究データを用いたサービスの調査・企画



大向 一輝 (国立情報学研究所)

私は、日本の論文を検索する CiNii や、大学図書館の蔵書データを検索するサービス、博士論文を検索するサービスといった、国立情報学研究所 (NII) の情報サービスの企画設計を担当しています。

研究者としては、オープンデータの話でも時々出てきますが、リンクト・オープン・データと呼ばれるようなデータフォーマット、データの表現方法に関する研究をしています。でも今日はその話は一切出てきません。

NII の教育研修事業

NII は多面的な組織で、情報学に関する研究の場所であり、大学共同利用機関として、大学のインフラを提供しています。そのインフラの中にネットワークサービスと学術情報サービスがあり、さらに学術情報サービスの中にはオンラインサービスの提供と人材育成もミッションとして入っているため、今日はその部分についてのお話になります。

NII の教育研修事業では、大学図書館の方々を対象として提供しているプログラムがたくさんあります。これまでの成果を図 1 のようにサイトに載せています。

現在行っている研修のメニューが図 2 に並んでいます。一番上は講習会で、システムの使い方をレクチャーしたりするのですが、下に行くほど専門性が高まっ

て、専門研修では、例えば目録の取り方の一番難しいところについて研修したり、リポジトリ事業についてもシステムの使い方を覚えていただかなければいけないので、その研修を行います。

さらにレベルが上がっていくと、最終的には三つの研修 (NII 実務研修、学術情報システム総合ワークショップ、大学図書館職員短期研修) があります。NII 実務研修は、NII に半年や 1 年来て、机を並べて普通に仕事をしていただくものです。完全に OJT で、何が研修なのかよく分からないのですが、そこで全国の大学図書館に共通する課題について取り組んでいただきます。大学図書館職員短期研修は、数日間にわたってレクチャーを受けて学ぶものです。

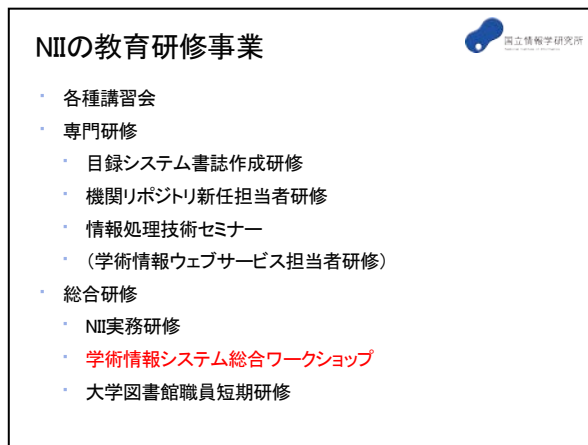
この後、学術情報システム総合ワークショップの成果発表が 2 件あるので、私はその前座として、この後の発表はどういうコンテキストの中でつくられてきたものかについてお話しします。

学術情報システム総合 WS とは

大学図書館、学術情報流通をめぐる問題は日々刻々と変わっており、その一つがオープンサイエンスになると思います (図 3)。もはや特定の部署で頑張ったら解決する、組織で頑張ったら解決するという問題は、皆さんの能力が高いのでとっくに終わっていて、残っ



(図 1)



(図 2)

ているのは横断的に取り組まないといけないものばかり、だからこそ非常に難しく感じるのです。そういう問題をみんなで連携・協力しながら解ける力を身に付けてもらうことがこの研修の目的です。この研修では、全くレクチャーは行われません。約6カ月間、ただ放っておかれて、いつまでに報告書をつくって持つてくるように、と指示される、極悪な研修スタイルを取っているのです。3回ぐらいは実際に集まるのですが、その間は基本的にはオンラインでやりとりします。

この研修は4年目を迎え、今年も8名の方に、どういう研修をするかを知りながら来ていただいているはずなので、非常にありがたいことです。

6月に1回初顔合わせをして、次に2カ月後に会いましょうと解散して、また2カ月ぐらいたって集まって、最後に報告書を書くというスケジュールです(図4)。日本中の図書館員の方なので、その間は基本的に

学術情報システム総合WS

- ・ 目的
 - ・ 大学図書館等と学術情報流通をとりまく環境は常に変化しており、それらをめぐる課題は**個別の機関で解決不能な問題**が増えている。本ワークショップにおいては、様々な立場から、共通する課題に**連携・協力**して取り組み、問題解決を図ることのできる総合力を身につけることを目的とする。**3回の短期集合研修(前期・中期・後期)**および**自館における作業**を通じ、自律的に課題に取り組む姿勢を涵養する。
- ・ 期間: 約6ヶ月
- ・ 参加者: 8名

(図3)

総合WSの進め方

- ・ 第1回: 6/30~7/1
- ・ (オンラインでの共同作業)
- ・ 第2回: 8/25~8/26
- ・ (オンラインでの共同作業)
- ・ 第3回: 11/1~11/2
- ・ (オンラインでの共同作業)
- ・ 最終報告書の提出: 12/16
- ・ ツールによる支援
 - ・ ビデオチャット
 - ・ タスク管理システム



(図4)

はじかに会うことはほとんどなく、オンラインのツールを使ってやりとりをします。ビデオチャットを使う場合もあれば、われわれが提供しているタスク管理システムの場合もあります。メールで連絡を取ると、議論発散というか、日常の仕事と混ざってしまうので、ツールを使ってきちんと管理しながらやってもらえるようにすることは支援します。こういうスタイルで、今年も半年間無事に最終成果まで持ってきていただきました。

テーマは、毎年ころころ変えています(図5)。初回は、デジタルアーカイブ的な、どこにどんなデジタルリソースがあって、それをどうやって活用するべきかを調査していただきました。

次の年は、CiNiiを中心にログを実際に見て、そのユーザーがどんな振る舞いをしていて、その振る舞いに基づいてどんなサービスをさらに提供するべきかを考えていただきました。

去年は、今もホットな話題ではあるのですが、総合目録システムをどのように新しく変えていくべきかということ、データを探りながら考えていただきました。

今年は、オープンサイエンスに係る「研究データを用いたサービスの調査・企画」というテーマにしました。研究データのオープン化、オープンサイエンスは、理念が分かると非常に大事なような気がするし、制度がどんどんつくられていて、それに沿って人々が行動する、そういうところの議論は非常に先行していると

総合WSのテーマ

- ・ 2013「デジタル化された資料の活用」
- ・ 2014「学術情報サービスにおけるユーザー理解」
- ・ 2015「NACSIS-CATの運用モデル再考」
- ・ **2016「研究データを用いたサービスの調査・企画」**
 - ・ 研究データを用いた、研究支援を目的としたウェブサービスを現状やニーズの調査を踏まえて企画する。本ワークショップでは**蓄積**した研究データをオープンにし、**利活用**することを想定してサービスの企画を立てる。また、どのような蓄積・管理の在り方が望ましいかも検討する。

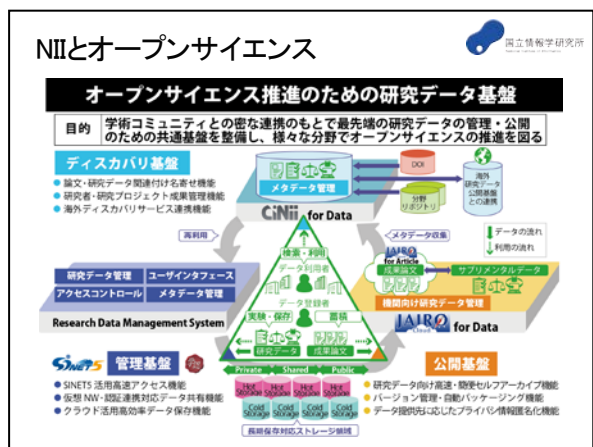
(図5)

感じます。では実際に、研究者、あるいは図書館員など学術情報を支える人たちはどうなるかという、情報サービスを普段使用中で、その理念や制度に従うことになっていくはずですが、北本先生の報告では Lessig の四つの力に触れられましたが、規範、法、市場などがあるけれど、やはりアーキテクチャ、何かをせざるを得なくなる、何かを仕向けるものが、実際の生活の中で一番関わるところであって、そのシステムをもう少し深掘りして考えないと、実際にオープンサイエンスをするということが何なのか理解できないのではないかという個人的な疑問もありました。

それで、そこにフォーカスして、別に制度を決めてもらうのではなく、情報サービスを企画すること、サービスの姿を描き表すことをテーマにしました。

これは架空の設定ではなく、NII の中でもこれからオープンサイエンスはかなり強く打ち出していくことになっています。図6は古いもので、単語はアップデートされたものがこれから出てくると思うのですが、NII の中で「三角の図」と呼ばれているものです。

NII が提供する情報サービスには三つの要素が必要であると考えられています。一つ目は情報管理で、研究者が普段、データをシェアリングしながら研究することをサポートするためのものです。二つ目は、何かの成果を公開して世の人に使ってもらうための基盤です。三つ目は、そういうものをもっと集めて検索可能にし、発見可能にする基盤です。この三つがそろって情報サービスによる支援だということでやろうとして



(図6)

いて、これに来年度からかなり力を入れていくので、これをやるための研究員やスタッフを今まさに募集しています。興味がある方はアプライしていただけると大変うれしいです。

学術情報システム総合 WS のより大きなテーマ

こういうこともあって、情報サービスの中身をつくらせていただくということを研修プログラムにしました。そこで、ある考え方を受講者の方をお願いしています。それは、現状について調査・分析することは大切ですが、現状に引きずられてはいけない、オープンサイエンスは試行錯誤の真っ最中であって、積み上げベースで次のサービスを考えるというよりは、図書館としてはこういうふうに関わりたいということを明確にしながらサービスを描き出していくべきであろうということです。図7は、研修にも参加していただいた岡本真さんが描いた図を引きました。まずビジョンを立てた上で、その中でわれわれがやるべき仕事、それを実現する手段を書き下していくことで、システムの仕様にしていって下さいということを言いました。

また、単に理念ベースで書き下すだけではなく、仕様書のスタイルで書き起こしてください、画面の設計はきちんとやってくださいということを言いました。やりたいこと、願望だけを書くのではなく、要る機能、要らない機能をきちんと選別してものにしていくこと、願望をたくさん書いて、それに合わせて絵を描くと大変な量になって、そもそもマンパワーとして無理なの



(図7)

で、ある種の制限をかけながら考えることを指示しました。

これから二つのチームの発表がありますが、それぞれ全く出自、バックグラウンドの異なる人たちで考えたことです。コラボレーションに慣れていくことが、今後の図書館による問題解決にとって非常に重要な基盤になるのではないかと考えて実施しました。

まだ来年度もこのスタイルは続けるのですが、テーマは検討中で、もう一回オープンサイエンスを扱うのか、違うテーマを扱うのか分かりませんが、ぜひ興味がありましたらご参加ください。資料・報告書等は全て図8に載っています。

成果発表


国立情報学研究所

- A班
 - 機関内研究情報統合システム(IRIS)の企画構想
- B班
 - オープンデータ利活用総合基盤システム「ODEN」の企画
- 資料・報告書
 - <http://www.nii.ac.jp/hrd/ja/ciws/h28/curritxt.html>
- 2017年度の総合WS
 - テーマ検討中…

(図 8)

A 班：機関内研究情報統合システム（IRIIS）の企画構想



田村 峻一（滋賀医科大学附属図書館）

ワークショップの A 班のディスカッションの内容について、発表させていただきます。

1. 班員紹介とワークショップの流れ

まず、私も含めて A 班の班員を紹介します（図 1）。田村、野間口、堀、柏倉という 4 名が A 班に分かれて作業を行いました。私と野間口さんは、滋賀医科大学と京都大学の図書館員として勤務しています。滋賀医科大学は医科系の単科大学で、京都大学は総合大学ということで、少し毛色の違う大学になっています。残りのお二人は、堀さんは自然科学研究機構国立天文台研究力強化戦略室で、URA の研究評価の仕事をしています。柏倉さんは海洋研究開発機構（JAMSTEC）の地球情報基盤センターで、地球シミュレータ研究成果リポジトリというリポジトリの管理業務に当たっています。このように、A 班はバックグラウンドの異なる 4 人が集まったので、大変だったところもあるのですが、お互いに自分とは違う世界のことを知れて、大いに勉強になりました。

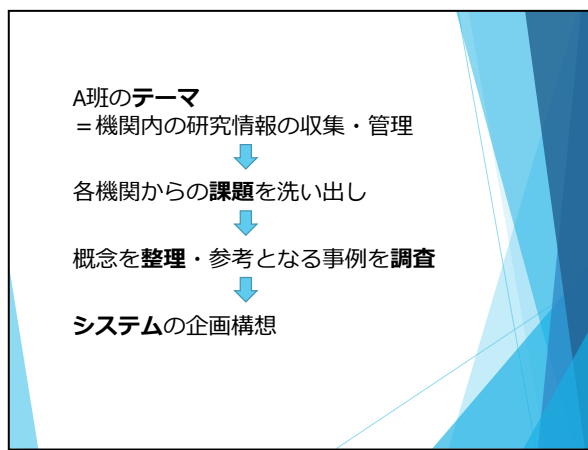
なぜこの 4 人で集まったかということ、自分の機関内の研究情報を収集し、管理する仕事をしていることが共通していたからです（図 2）。よって、この共通テーマからディスカッションをスタートさせました。各機関で感じている課題を洗い出し、その後、概念を整

理して、参考となる事例を調査、最終的にシステムの企画構想という流れでした。各段階について説明していきます。

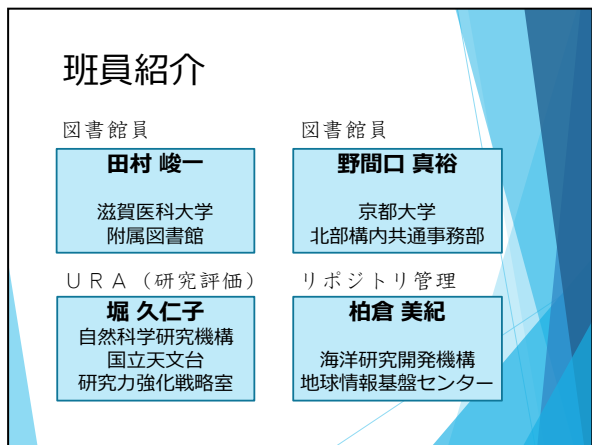
課題の洗い出しでは、各機関で感じている課題を持ち寄りました（図 3）。いろいろありましたが、大きく集約化と簡便化という二つのベクトルに整理しました。

集約化については、現状だと、研究グループごとに研究成果の出どころがばらばらになっていて非常に探しづらい、そのため、これを機関でまとめれば、研究評価や分析に当たって便利であり、外部に対する広報としても良いということです。

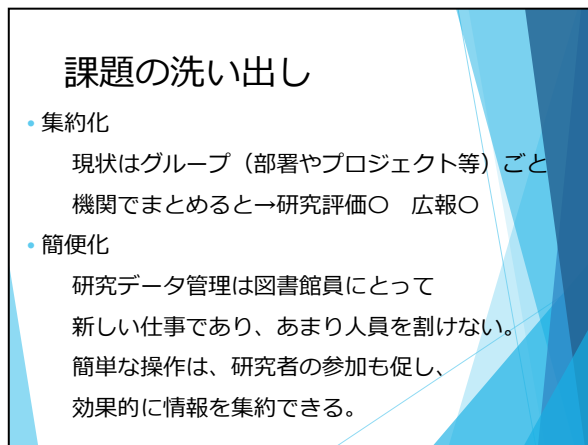
簡便化については、研究データ管理は図書館員にと



（図 2）



（図 1）



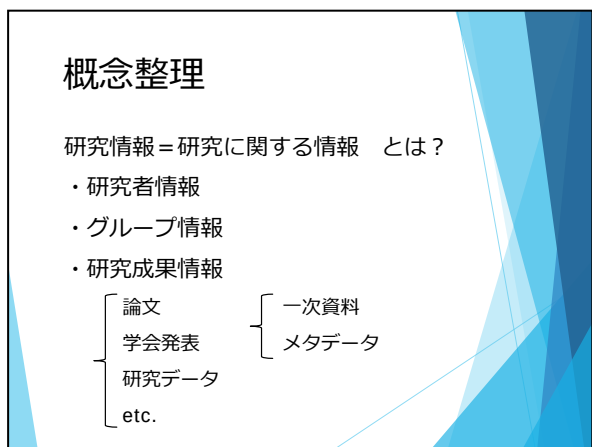
（図 3）

って新しい仕事で、それを導入するに当たっては、簡単な操作で行えることを重視したいと考えました。職員も入力・作業をしますが、研究者にも作業に参加してもらって、みんなが使いやすい、みんなで使ってもらえるようなシステムをつかった方がいいという話になりました。

「研究情報」という言葉を先ほどから使っていますが、これは研究に関する情報全般を指しています(図4)。これを細かく分析すると、研究者に関する情報、研究のグループ(研究者の集まり)に関する情報、研究成果に関する情報に分かれます。研究成果情報はさらに、論文や学会発表や研究データなどに細かく分かります。研究成果情報は、一次資料(論文のPDFファイルそのもの、研究データそのもの等)とメタデータから成るという別の観点からの分類もあります。

「データ」という言葉、「情報」という言葉を安易に使うと、かなり議論が錯綜してしまったので、このようにまとめて考えました。

図5は参考事例調査の一つの大きな例です。OAIS参照モデルという、研究成果が生産者によって生産され、それを保存計画に基づいて運用し、消費、アクセスに至るまで管理して計画していくというモデルを見て、この四角形に当たるようなシステムをつくらうかということになりました。

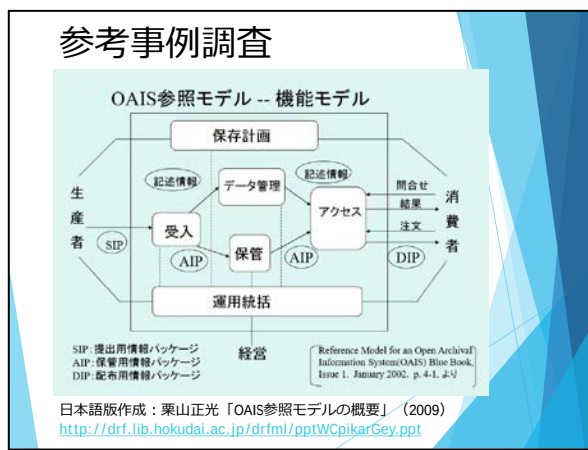


(図4)

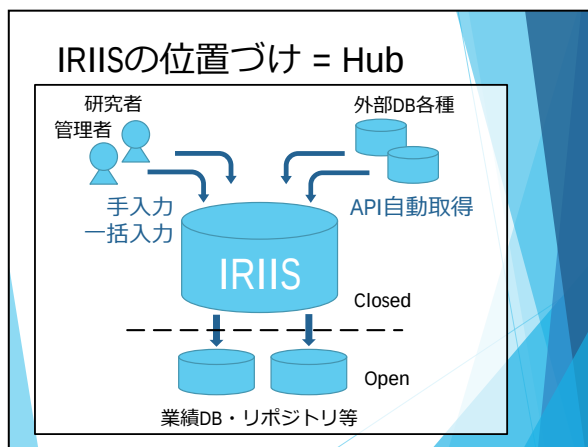
2. IRIIS の概要

そしてできたシステムが IRIIS (Institutional Research Information Integrated System : 機関内研究情報統合システム) です。最初は I がもう一つ少なくて IRIS と呼んでいたのですが、某図書館システムと同じ発音になってしまうので、Integrated の I を 1 個足して、統合するシステムだとしました。われわれが実務を担当している大学や研究機関など、一つの機関内で完結して使用するシステムを想定しています。

IRIIS がどういう構成になっているか紹介していきます。IRIIS はハブとしての位置付けを持っており、複数のソースから情報を取ってきて、いったん蓄積します(図6)。ユーザーは研究者と管理者の2種類を設定しています。研究者は研究者で、管理者というのは、図書館員や URA のような事務職員を設定しています。



(図5)



(図6)

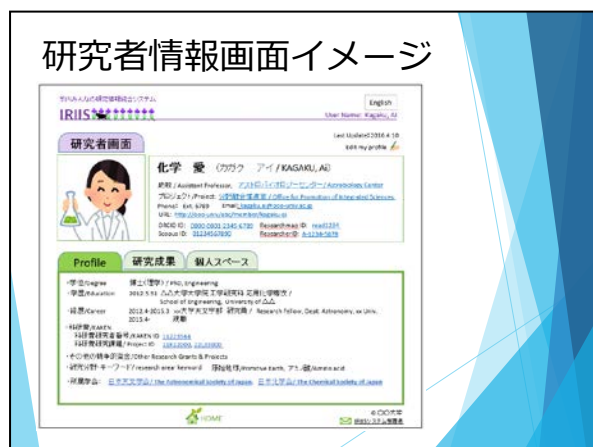
外部のデータベースからも自分の機関の研究に関する情報と思われるものを自動的に引っ張ってきて、それをプールします。プールされた情報は、クローズドのまま保持するものもありますが、オープンにできるものはオープンに切り替えて、それをまた別のサーバー、リポジトリにはき出していくというシステムです。

3.IRIIS の四つのテーブル

IRIIS の中身は主に四つのテーブルから成っています(図7)。先のスライド(図4)で研究情報を分類した名称を使っていますが、それぞれ研究者情報テーブル、グループ情報テーブル、研究成果情報テーブル、さらにデータ保存スペースという機能を付けて、大きく四つのエンティティから成るシステムを構築しました。以下、それぞれについて画面イメージと業務フローの概念図を示して説明していきます。



(図7)

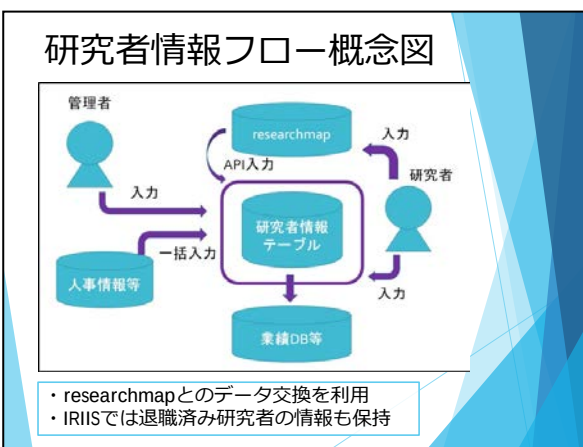


(図8)

3-1.研究者情報テーブル

図8のような画面のイメージ図をつくりました。研究者に関する情報が表示されています。タブで研究成果の画面にも行きますし、後で述べる個人スペースの画面にも行くことができます。

研究者情報は researchmap とのデータ交換が自動的に行われます(図9)。これはもう滋賀医科大学等でも既にやっていることなのでイメージはできると思います。人事情報など、機関内の他のデータベースで持っているような情報も、TSV や CSV ファイルなどを使って一括で取り込めるようにすると楽になるのではないかと考えました。人事の情報はカレントで存在している研究者の情報ですが、退職済みの研究者についても学術情報としては保持しなければならないので、退職済みフラグのようなものを使って情報を保持するというイメージです。



(図9)

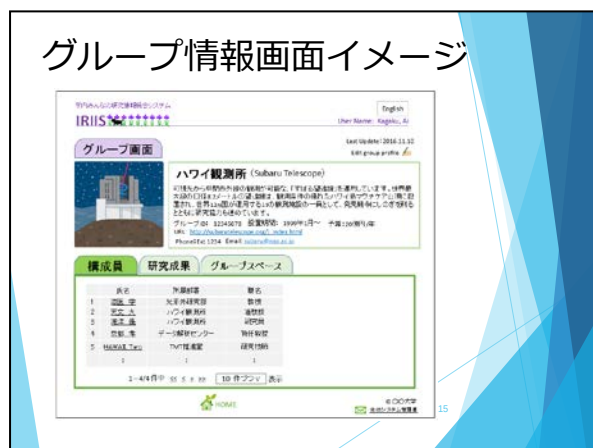
3-2.グループ情報テーブル

グループ情報画面も同じような画面になっています（図 10）。グループ独自の情報として、誰が所属しているのかということと、設置期間があります。グループの情報を独自に研究者とは別に持つことで、グループ単位での研究成果の集約に有効活用できるのではないかと考えました（図 11）。

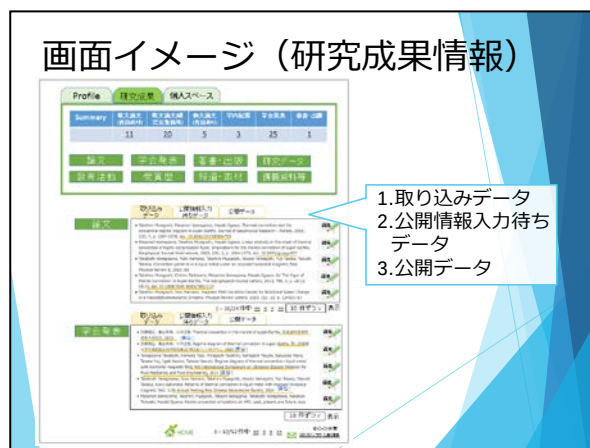
3-3.研究成果情報テーブル

研究成果情報については、API 等で取り込んできたデータは、そのまま取り込みデータとして持って、これは確かに私のデータである、この機関のデータであると承認されたものが次の段階に行きます（図 12）。そのうち公開できるデータには公開フラグのようなものを立てて、という3段階のレイヤーにデータを分けています。

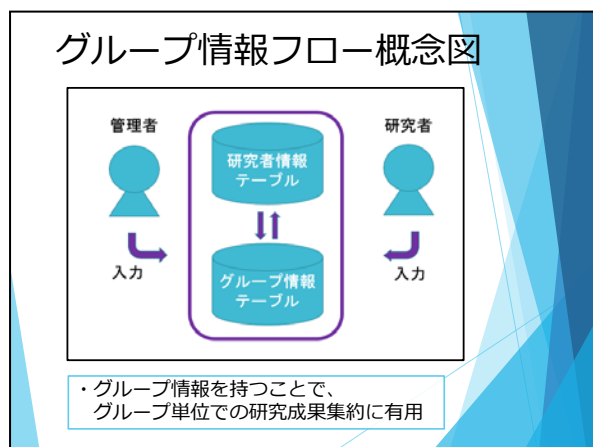
図 13 は研究成果情報のフロー図です。外部から取ってきて、管理者や研究者が内容を成形して、アタッチファイルを付けられるようなものは付けて、それをためて、機関リポジトリ等のサーバーに移せるようにしておきます。研究データについても文献と似たような感じで、DataCite を例として考えていますが、きちんとこれができるかどうかは、今後見ていかなければいけない課題だと思っています。



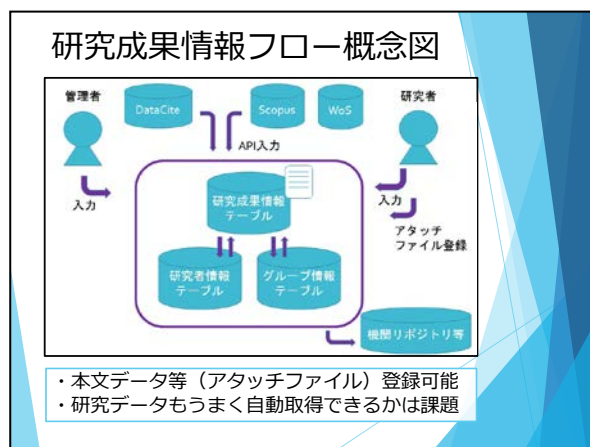
(図 10)



(図 12)



(図 11)



(図 13)

3-4.データ保存スペース

研究者が研究に使うファイルをアップデートして保存しておけるデータ保存スペースもつくりました。それに付随的に To Do リストやカレンダーの機能も付けて、Dropbox、Backlog のようなプロジェクト管理ツールを参考にしたような設計になっています (図 14)。

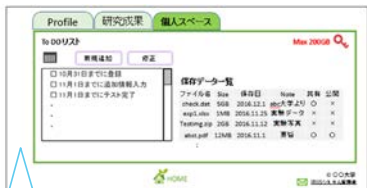
研究中のデータは保存スペースに保存されますが、研究期間が終わったら、これは公開できるようになったという通知が管理者に行って、管理者がメタデータを入力して、同じようにデータリポジトリのような別のサーバーに移動できるシステムを考えました (図 15)。

図 16 は本発表の振り返りです。今後の展開としては、本システムの IRIIS は仮想の大学に導入するという設定で構想したので、抽象的なモデルのようになっていますが、実際に各機関で導入するとすると、機関

ごとの環境や予算状況に合わせて、具体的に協議することが必要だと考えられます (図 17)。連携する外部のシステムも、これから変化していくものが多いと思うので、それに合わせてつくり替えていくことが必要だと思います。

今後、この IRIIS のようなモデルが標準化し、広く普及することで、多くの機関が研究データ管理に取り組みやすくなる環境が整うことが望ましいと考えています。

画面イメージ (データ保存スペース)



- ・ To Doリスト、カレンダー機能が付属
- ・ 各研究者およびグループに200GB
→ 申請により増設可能
- ・ 管理者からも見られない
→ 必要に応じ他の研究者と共有

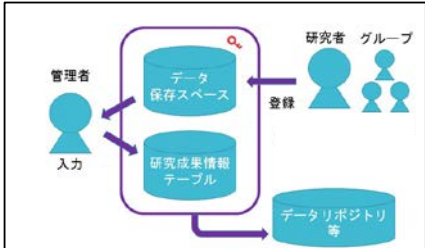
(図 14)

本発表の振り返り

- ▶ 機関内の研究情報を収集・管理する実務者としての立場から、統合システム“IRIIS”を提案
- ▶ 集約化・簡便化を指向
- ▶ 4つのテーブルを軸としたシステム構成

(図 16)

保存スペースからの公開



- ・ 研究終了後の研究データは管理者がメタデータ入力公開

(図 15)

今後の展開

- ▶ 本システムは仮想の大学向け
- ▶ 外部システムの状況も変化の只中
(researchmap2.0, CiNii for Data, ORCID, JaLC2 ...)
- ▶ このようなモデルが標準化し、広く普及することが望ましい。

(図 17)

B 班：オープンデータ利活用総合基盤システム「ODEN」の企画



梶原 茂寿（北海道大学附属図書館）

B 班は、オープンデータ利活用総合基盤システム「ODEN」の企画について説明します。

ODEN の企画・開発

B 班のメンバーは、北海道大学の梶原と、九州大学の法常さん、国立国会図書館の町屋さん、東京大学駒場図書館の松原さんの 4 人です（図 1）。われわれの班は A 班とは違って、研究データのユーザー側からの視点で、どのようにオープンデータを使っていくかを考えた企画となりました。

この企画は、データはオープンに公開されつつあるということが前提です（図 2）。データは既にいろいろなドメインやコミュニティで持っていて、オープン

にされているものもあるけれども散在していて、その利活用を支援する仕組みがまだ不十分な状況があることから、オープンデータ利活用総合基盤システム（Open Data Enhancing Network）、頭文字を取って、愛称「ODEN」の開発を企画することにしました。

基本コンセプトは、「散在するオープンデータのメタデータを一元的に集約し、研究データのワンストップで検索を可能にするためのメタデータデータベース」です。利用を促進するための機能として、Amazon の「このデータを使った人はこんなデータも使っています」というレコメンド機能や、ライトユーザーに向けたゲーミフィケーション要素を導入しています。また、どのようなレベルの利用者であっても利活用しやすいように、ユーザーが交流・情報交換できるフォーラムや、ユーザーによる自発的な情報の蓄積を可能とするツール&マニュアル機能も準備しています。

ワークショップの最初の集まりで示された説明資料の中に、「サービスの企画をする際には、VMSO（Vision、Mission、Strategy、Objectives）を定めよう」ということがありました（図 3）。われわれもそれに沿って今回の企画をしていきました。われわれが考える ODEN のビジョンは、「DATA for Your HAPPINESS 研究データを解放し人々を自由にする！！」です。

そして、そのビジョンを実現するために、「あらゆ

B班メンバー	
梶原 茂寿	北海道大学附属図書館
法常 知子	九州大学附属図書館
町屋 大地	国立国会図書館
松原 恵	東京大学駒場図書館

↓

研究データのユーザー側の視点から企画

2017/2/14 第3回 SPARC Japan セミナー2016

(図 1)

ODEN の企画趣旨

- データはオープンに公開されつつある
(という前提で企画)
- 十分な利活用がされていない。
← データが各所に散在している
← データの利活用支援の仕組みが不十分
- なので、……

2017/2/14 第3回 SPARC Japan セミナー2016

(図 2)

VMSO

VMSOを定めよう

Vision	・実現すべき未来
Mission	・達成すべき使命
Strategy	・実施すべき戦略
Objectives	・到達すべき目標

Copyright © 2016 NII. All Rights Reserved. 学問をまかす社会へ

平成28年度学術情報システム総合ワークショップ
岡本真「サービスの企画」資料より (CC BY 2.1 JP)
<http://www.nii.ac.jp/hrd/ja/ciws/h28/1-02.pdf>

2017/2/14 第3回 SPARC Japan セミナー2016

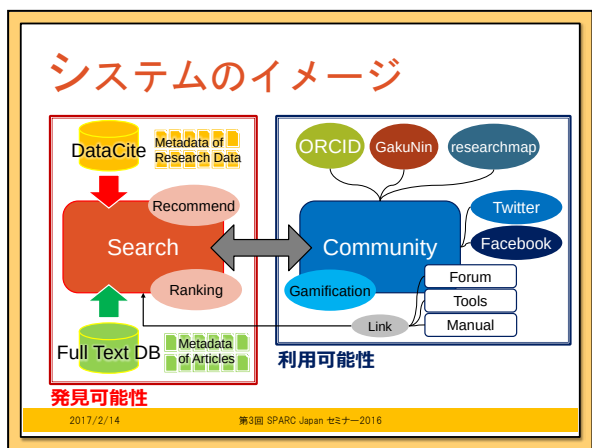
(図 3)

るオープンデータを見つけやすく、使いやすい、場
「コミュニティ」を実現する。そのことにより、全社
会の研究・イノベーションを加速させる」というミッ
ションを掲げています。

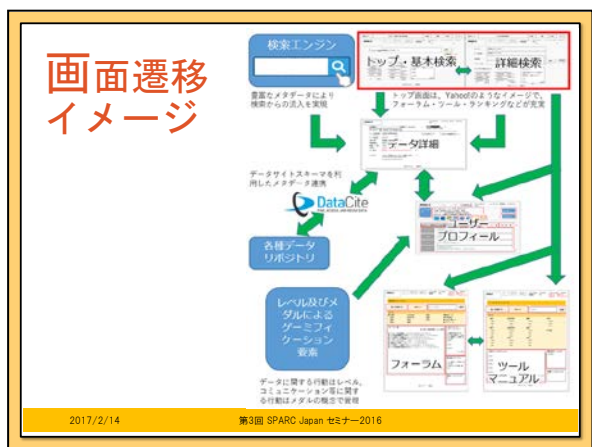
システムのイメージ

図 4 は ODEN のシステムのイメージです。単純に
発見の可能性を上げる検索のための部分と、見つけた
オープンデータを使っていただくための部分の大きく
二つに分かれます。

図 5 は画面遷移イメージです。検索エンジン、
ODEN のトップ画面、基本検索画面、詳細検索画面
があります。そこからこちらで収束しているメタデー
タにたどり着いて、そのメタデータは基本的に
DataCite のスキーマを利用して連携しているという形
です。そこからデータのユーザーのところへ飛んだり、



(図 4)



(図 5)

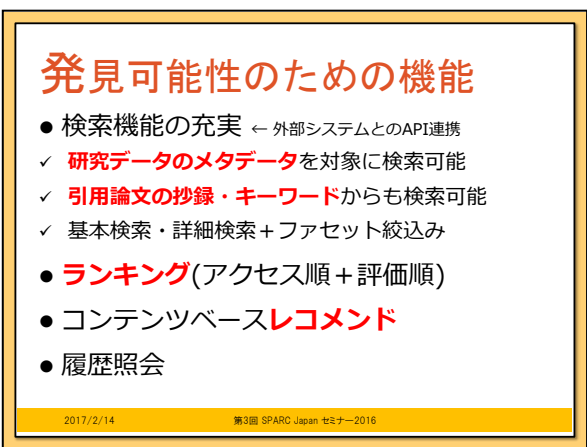
ユーザープロフィールなどもあるのですが、そこでレ
ベルやメダルといったゲーミフィケーション要素の管
理もできます。また、フォーラムやツールマニュアル
なども用意しています。

次に、発見可能性のための機能について詳しくご説
明します (図 6)。検索機能を充実させており、研究
データのメタデータを対象とした検索が可能、引用論
文の抄録・キーワードからの検索も可能であることが
コアな機能です。

また、メタデータがどれだけアクセスされたかとい
うアクセス回数順、ユーザーによるメタデータの評価
順のランキングを示す機能があります。さらに、「こ
のデータを使った人はこちらのデータも使っています」
というレコメンド機能や、マイページで今までに使っ
たデータの照会もできる機能を考えました。

利用可能性を高めるための機能としては、フォーラ
ム機能、ツール&マニュアル機能、ゲーミフィケーシ
ョンがあります (図 7)。フォーラム機能では、ユー
ザーが相互に「こんなデータを探しているのだけれど、
何かいいデータはないですか」という情報交換や交流
をすることで、自助的な能力育成につながると考えて
います。DIAS などが既にこういうものを実現してい
ると思います。

ツール&マニュアル機能では、ユーザーがナレッジ
を蓄積し、ある一定以上のレベルにあるユーザーや研
究者が自分でツールやマニュアルをつくって、ここに
提供することができるようになっています。



(図 6)

ゲーミフィケーションについては二つの機能があって、一つはレベルです。例えばデータにアクセスした回数など、データに関する行動を起こすとレベルが少しずつ上がっていくようなことを考えています。もう一つはメダルです。メダルは ODEN の具になっていて、フォーラムに3回参加すると大根がもらえる、といった具合になっています。

まとめ・今後の展望

データを集めるだけがデータベースの作成ではありません（図 8）。研究者やデータを扱っている人だけに対応しているようなシステムでは、オープンアクセスの促進につながらないのではないかということから、さまざまなユーザーに対応できるようなものを考えていきました。そして、データベースのリピーターをつくることも大事なので、ライトな遊びの部分も必要だと考えて、ゲーミフィケーションを取り入れました。

あらゆるオープンデータがを見つけやすく、使いやすい場「コミュニティ」を実現する。そのことにより、全社会の研究・イノベーションを加速させるということを考えながら今回の企画を行ったということです。

ワークショップ成果物に仕様書と報告書が載っているので、A 班のものも、B 班のものも、ぜひそちらを見ていただければと思います。

利用可能性のための機能

- フォーラム機能
 - ✓ ユーザー相互の**交流・情報提供**による**自助的な能力育成**
- ツール&マニュアル機能
 - ✓ ユーザーによる**ナレッジの蓄積**による活発なデータ利用
- ゲーミフィケーション
 - ✓ **ライトユーザー**が継続的に利用し続ける**動機づけ**
 - ✓ 心理的インセンティブによる**コミュニティの形成**

2017/2/14

第3回 SPARC Japan セミナー2016

(図 7)

まとめ・今後の展望

- ・ データを集めるだけがデータベースの作成ではない。
- ・ 様々なユーザーに対応できるシステムを構築することによって、データベースの利活用をより促進する。
- ・ データベースのリピーターを作るための遊びも必要。

2017/2/14

第3回 SPARC Japan セミナー2016

(図 8)



●**大向** 半年かけて、それぞれ4名の班で、仕様書をつくって提案をしていただきました。私は中身にコントロールはかけないようにやってきました。サービス全体の姿もそうなのですが、その所々に表れる細かいディテールが重要だと考えています。

例えば、そもそも研究データはメタデータを付けて検索できたらそれでいいのかということに対しては、どちらの提案からも鋭く突っ込みが入っています。私自身は、例えば研究の途中でデータが欲しくなったら、データ検索エンジンにアクセスして、そこにキーワードを入れて、リストを見て、クリックするというような振る舞いが本当にあるのか、非常に疑問に思うわけです。欲しいデータがあれば、欲しいと分かっているものであれば、そうやって探すかもしれませんが、そうではないものをデータ検索エンジンで探すということは、今のところ自分の研究プロセスの中にはありません。それよりも、例えば論文のデータ、プロジェクトの何か、日本であれば科研費のデータベース、プロジェクトベースのデータベースを見て、そこで使われたデータということで初めてアクセスする可能性があります。

そういうことを考えると、これまでの図書館的に、データに対してきちんとリッチなメタデータを付けて、それをデータベースに入れれば、オープンサイエンスの支援ができたとは到底言えないというのがこの二つの班の提案から見ても非常に思うところです。

どちらの報告にも、かなり人ベースということが出てきていました。自分のソーシャルメディアのプロフィール画面のようなものがあって、それにひも付いた形でデータが載っている、そういうことが割と自然に出てきていることを考えてみても、やはりメタデータでは済まない、図書館なら図書館で、人のネットワークのファシリテーションまで手を出さないと、研究データアクセスに対する支援になっていないという可能性が見えてくると思います。

これからわれわれとしても、何かしらの形でそのシ

ステムのプロトタイピングを進めていかなければいけません。直接的なメタデータアプローチは、ゼロでは全くサービスにはならないですが、そういうエッセンスをどう取り込んでいくか。そして、まだ顕在化していないニーズ、情報要求がどこにあるのかということ、架空の設定ではあるけれども考えていただくことによって、少し浮き上がった部分があるのではないかと考えました。

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

RDMトレーニングツールの紹介

尾城 孝一

(東京大学附属図書館)

講演要旨



世界各国で公的研究資金による研究成果について、論文だけではなくエビデンスとなる研究データの保存や共有を求めるポリシーを制定する例が相次いでいる。適切に研究データを管理するには、研究者自身に加えて研究支援者が研究データ管理に係る知識やスキルを習得する必要がある。本発表では、機関リポジトリ推進委員会の研究データタスクフォースで開発を進めている、研究データ管理に関する研究支援者向け学習教材について紹介する。



尾城 孝一

東京大学附属図書館事務部長。1983年1月、名古屋大学附属図書館に採用され図書館職員としてのキャリアを開始。その後、東京工業大学附属図書館、国立国会図書館、千葉大学附属図書館、国立情報学研究所、東京大学附属図書館、大学図書館コンソーシアム連合事務局を歴任。2015年4月より、現職。

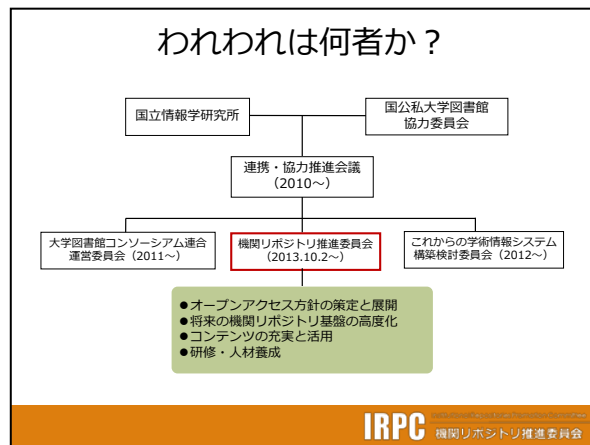
今日は、機関リポジトリ推進委員会の研究データタスクフォースが開発した、RDM (Research Data Management) トレーニングツールについて紹介させていただきます。

機関リポジトリ推進委員会の取り組み

国立情報学研究所 (NII) と大学図書館の連携・協力の枠組みの中に、機関リポジトリ推進委員会という委員会が置かれており、そこで機関リポジトリ、オープンアクセス、オープンサイエンスに関するさまざまな活動をしています (図 1)。

機関リポジトリ推進委員会には、基幹的な業務のサポートをするための三つの常置の作業部会があります (図 2)。その他、先端的なプロジェクトに取り組むためのアドホックなタスクフォースが六つ設置されて

おり、われわれは其中で研究データタスクフォースとして活動しています。これから紹介する RDM トレーニングツールの開発や、ケーススタディを通じて研究データ管理のノウハウを蓄積していくような活動をしています。



(図 1)

図3がそのタスクフォースのメンバーです。私の他に、NIIの山地先生、大学や研究所の図書館員の皆さん、URAの皆さんにも入ってもらい、一緒に活動しています。数は少ないのですが、これは今考えられる最強の精鋭部隊だと自負しています。

さて、オープンサイエンスなのですが（図4）、研究データに対して研究者の皆さんはどういう意識を持っているのでしょうか。DataONEというプロジェクトに関連する調査があり、図5はその最初の方の調査結果なのですが、1,329名の研究者のうち、78%が、少なくともデータの一部はアクセス制限のないデータリポジトリに登録したいと考えている、しかし、自分自身のデータに他の研究者が容易にアクセスできていると回答した回答者は36.2%にとどまっています。データを電子的に利用できるようにできない理由としては、やはり時間がない、お金がない、データの登録先

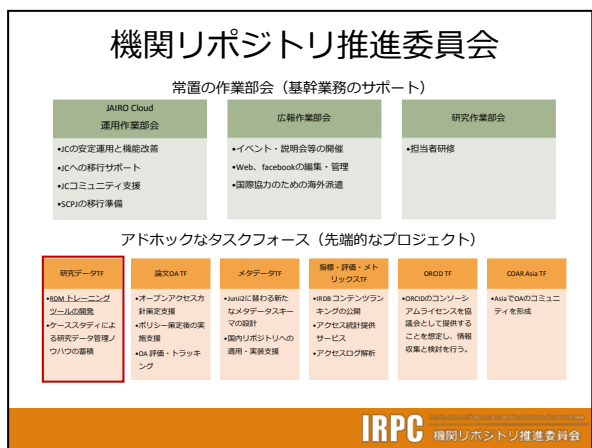
がないといったことが挙げられています。

この調査から、データを公開したい、あるいは研究資金配分機関等のポリシーによって公開しなければならない、けれども実現できないでいる研究者が結構たくさんいるということがうかがえるかと思います。

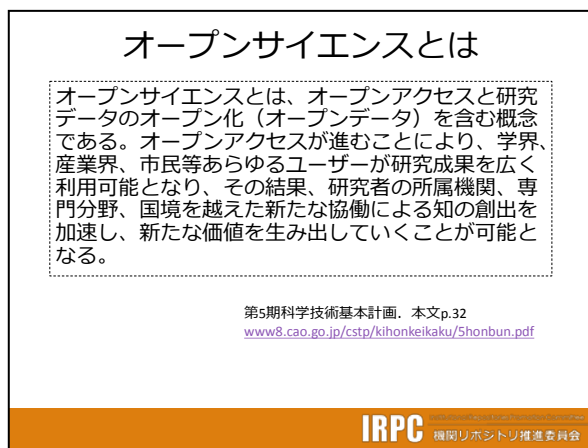
こういう研究者を支援し、オープンサイエンスを進めていくためには三つの基盤が必要だと考えています。一つ目は研究データ基盤（共通システム）、二つ目は研究者を支援する人の基盤（支援スタッフの確保・育成）、三つ目は組織の基盤（コミュニティ）です。今日はこのうち二つ目の支援スタッフの確保・育成につながるトレーニングツールの紹介をします。

RDM トレーニングツールとは

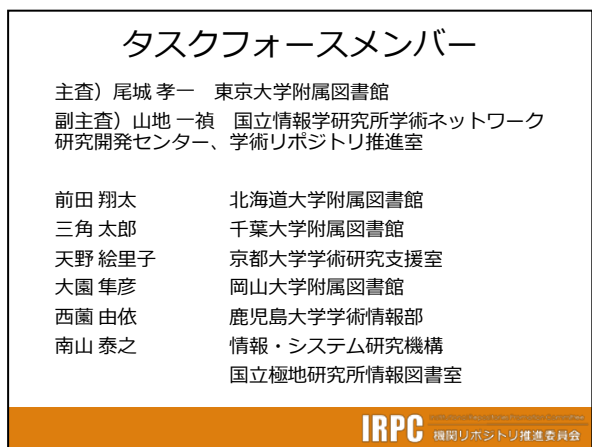
文部科学省が2016年2月に公開した審議のまとめの中でも、研究データのオープン化に関して、大学等



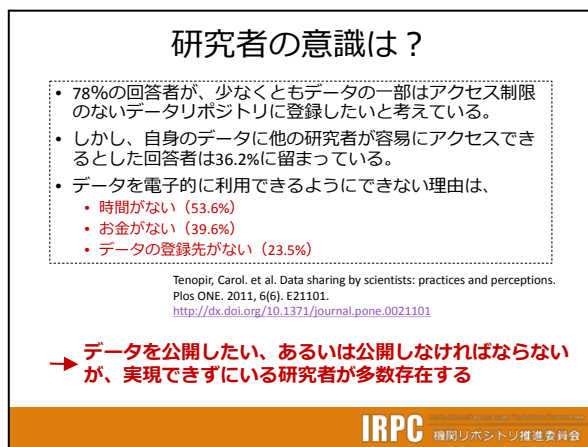
(図2)



(図4)



(図3)



(図5)

に期待されている役割が幾つか挙げられているのですが、その一つとして、技術職員、URA 及び大学図書館職員等を中心としたデータ管理体制を構築し、研究者への支援に資するということが書かれています（図 6）。

しかし、研究データ管理の支援をやれと言われても、一体何をやっていいのかよく分かりません。これは海外でも事情はどうも同じようで、どういうことをやればいいのか、それが学べるオンラインのツールがたくさんできています（図 7）。その日本版の教材をつくらうというのが、われわれが RDM トレーニングツールの開発に取り掛かったきっかけです。

ツールの目的は、学習者が RDM に関する基礎的な知識を得ること、研究者支援サービスの足掛かりを得ることです。

教材の形式は、e ラーニング教材にして、MOOC で

提供できるようにします（図 8）。講義、確認テスト、ワークなども適宜挟みこんでいきます。受講者は最初は図書館員を想定していたのですが、せっかくならもっと幅広く、IT 部門の人や、URA といった研究支援部門の人たち、それからオープンサイエンスや研究データに関心を持っている人たちに使ってもらえるようなツールにできればということで、教材の開発に取り掛かって、約 1 年かかってほぼ完成に近づきつつあります。

全体の教材の構成は図 9 のようになっており、7 章から成っています。データの生成、加工、分析、保存、公開、再利用と、データのライフサイクル全般についてひとつとおり学べるようになっていきます。

北本先生から、2017 年のスキルと 2027 年のスキルというお話がありましたが、取りあえずこのツールは 2017 年のスキルについて学べるようになっていきます。

研究者支援への期待

『学術情報のオープン化の推進について（審議まとめ）』
平成28年2月 科学技術・学術審議会 学術分科会 学術情報委員会
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm

（大学等に期待される役割のひとつ）

技術職員、URA 及び大学図書館職員等を中心としたデータ管理体制を構築し、研究者への支援に資する

IRPC 機関リポジトリ推進委員会

(図 6)

教材の概要

- eラーニング教材（MOOCにて提供予定）
 - 講義、確認テスト、ワーク
- 受講者は幅広く想定
 - 大学や研究機関等において研究者の支援に携わる方（図書館・IT部門・研究支援部門の職員（URA含む）、研究室で研究補助業務に携わる方等）
 - オープンサイエンスや研究データ管理に関心のある方

IRPC 機関リポジトリ推進委員会

(図 8)

研究データ管理の教材

（海外の教材）

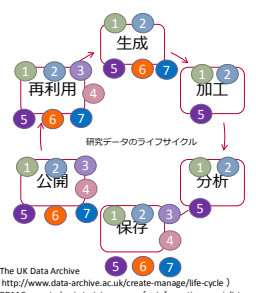


→ 日本版の教材を作ろう！

IRPC 機関リポジトリ推進委員会

(図 7)

データのライフサイクルと教材の構成



- ① 1章 導入
- ② 2章 データ管理計画
- ③ 3章 保存と共有
- ④ 4章 組織化・メタデータ
- ⑤ 5章 法倫理的問題
- ⑥ 6章 ポリシー
- ⑦ 7章 サポートサービスの検討

* The UK Data Archive
 (<http://www.data-archive.ac.uk/create-manage/life-cycle>)
 * RDM Support - basic training course for information specialists
 (<http://dx.doi.org/10.6084/m9.figshare.1285113>) より

IRPC 機関リポジトリ推進委員会

(図 9)

若い人にとっては、これは人生戦略上の損失につながってしまうようなツールかもしれないのですが、まずは基礎的なところから学んでいくのも大事ではないかということです。

図 10・図 11 に、各章のシラバスを簡単に書いてあります。導入、データ管理計画、保存と共有、組織化・メタデータ、法倫理的問題、ポリシー、サポートサービスの検討、このような内容になっています。

例えば第 1 章では、研究データ管理とは何かという説明があり、それに関するワークなども挟みこまれています（図 12・図 13）。

また、JST の小賀坂さんからデータ管理計画（DMP）のお話がありましたが、DMP のつくり方なども学べるようになっています（図 14・図 15）。

今後の展開

この教材は、まず 3 月末までに完成させて、MOOC 化も併せて完了させて公開したいと考えています（図 16）。その後、4 月から、JPCOAR（オープンアクセスリポジトリ推進協会）というコミュニティが本格的に活動を始めます。今の機関リポジトリ推進委員会の活動も、この協会の活動に引き継がれていきます。われわれ研究データタスクフォースもこの新しい協会の中で、教材をより広く活用してもらって、学んだ人からフィードバックを得て改定して、より良いサービスにしていきたいと考えています。

2027 年のスキルについて一緒に考えて、こういう活動を共にやっていきたいという方がいらっしゃいましたら、ぜひ声を掛けていただければと思います。

シラバス

第1章 導入
研究データ管理の重要性が増している背景や、研究データ、研究データ管理の定義について

第2章 データ管理計画
効果的なデータ管理に欠かせないデータ管理計画に関し、作成義務化の動向や、その構成要素について

第3章 保存と共有
研究データの、研究期間中の保管や長期保存に関する留意点について。研究データの共有に関し、その意義や検討すべき点、共有方法について。

第4章 組織化・メタデータ
研究データを長期的に管理・活用するために欠かせない、一定のルールに則ったデータの組織化や、データについて説明する文書やメタデータの作成について。


機関リポジトリ推進委員会

(図 10)

サンプル（未定稿）

1.3.1 研究データ管理
(Research data management) とは

ここでいう「研究データ管理」という言葉は、ある研究プロジェクトにおいて使用された、あるいは生成された情報やどのように組織化し、保存し、管理していくのかといったことにまで至る。

研究データ管理には次のようなことが含まれる：

- どのようにデータを取り扱っていくかの計画（データ管理計画/DMP）— 多くの助成機関においてデータ管理計画が申請書類の一部として求められている
- 研究プロジェクトが行われている間の、日々の情報の取り扱い方
- 長期的にデータを何に使うのか（研究プロジェクトが終わった後はそのデータを何に使用するか） <http://researchdata.ox.ac.uk/home/introduction-to-rdm/> より


機関リポジトリ推進委員会

(図 12)

シラバス（続）

第5章 法倫理的問題
研究データをめぐる著作権や、再利用を促進するためのライセンスの仕組みについて。センシティブデータを取り扱う上での留意点や研究倫理について。

第6章 ポリシー
国や助成団体、機関、出版者等による、研究データの保存や共有を求めるポリシー策定の動向や要件について

第7章 サポートサービスの検討
前章までの内容を踏まえ、学習者が自機関での研究データ管理サービスを構築していくためのステップについて


機関リポジトリ推進委員会

(図 11)

サンプル（未定稿）

1.4.1 ワーク（基礎）

次の文章の空欄に相応しい言葉を、選択肢a～fの中から選択してください。

オープンサイエンスとは論文へのオープンアクセスと「」のオープン化を含む概念であり、我が国では内閣府の出した第五期科学技術基本計画でその取り組みの姿勢について言及された。

研究データはオープンサイエンスや「」の観点からも適切な保存や公開が求められている。

研究資金の助成機関では「」の提出が義務付けられていることもある。

研究データ管理のことを英語で「」という。そのメリットとして、「」や「」などが挙げられる。

選択肢：
a データ消失リスクの減少 b Research Data Management (RDM)
c 研究データ d 研究効率の向上 e Data Management Plan (DMP)
f 不正対策


機関リポジトリ推進委員会

(図 13)

サンプル（未定稿）

WEKO JARQ IRP

2.1.1 データ管理計画（DMP）とは

「研究プロジェクト等における研究データの取り扱いを定めるものであり、具体的にはデータの種類、フォーマット、アクセス及び共有のための方針、研究成果の保管に関する計画などについて記載されるもの」（文部科学省科学技術・学術審議会学術分科学術情報委員会（2016）「学術情報のオープン化の推進について（審議のまとめ）」用語解説より）

- 研究の実施段階から終了後に至るまでの期間において、研究データがどのように生成、管理、共有、保存される予定か、を文書化する。
- この計画に従った管理を行い、研究データが利活用可能な状態で適切に管理されるようにする。
- 研究データの適切な保管・管理は、研究データの公開を進めるための前提である。

IRPC 機関リポジトリ推進委員会

(図 14)

サンプル（未定稿）

WEKO JARQ IRP

2.4.1 データ管理計画理解度チェック

1. 次の文章の空欄に相応しい言葉を、選択肢a～fの中から選択してください。

データ管理計画は以下に挙げるようなことで役立つ：

- 自分の研究データや記録が、[] であること、完全であること、真正性があること、信頼できること、これらのことを保証する（ラベリングが意味のあるもので完全性があるということを保証するのに役立つ）...
- 長期的に見て、[] と資源の節約になる。
- データの[] を高め、そのことにより、データ損失の[] を最小化する。
- 研究の公正性と、他者による[] を保証する。
- 他者に自分のデータを利用させることにより、手続の[] を防ぐ。

選択肢：

a. セキュリティ b. 再現性 c. 正確 d. 重複 e. 時間 f. リスク

IRPC 機関リポジトリ推進委員会

(図 15)

今後の展開

• 教材の公開（3月末予定）

↓



2017.4
本格始動

- 教材の活用促進
- 持続可能な教材開発体制の確立
- 教材の段階的拡充

IRPC 機関リポジトリ推進委員会

(図 16)

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

ディスカッション 「科学的知識創成の新たな標準基盤へ向けて」



林 和弘	(科学技術・学術政策研究所)
Ron Dekker	(欧州委員会研究・イノベーション総局 (European Commission (DG Research & Innovation)))
北本 朝展	(国立情報学研究所)
小賀坂康志	(科学技術振興機構)
小野寺千栄	(物質・材料研究機構)
小野 雅史	(東京大学地球観測データ統融合連携研究機構)
大向 一輝	(国立情報学研究所)

●**林** 本年度の SPARC Japan のセミナーを締めるこのディスカッションでは、今日の話と、年間を通して掲げたオープンサイエンスの再考という非常に大きなテーマを採りあげます。これをどうまとめたらいいか、まだ私の頭の中には何もないのですが、皆さまと一緒に何かつくり出せればと思っています。

今日は Ron さんに、日本もここまで来ているのだということをお見せしてきたので、あらかじめ Ron さんから、ご紹介した日本の取り組みについてコメントを頂きたいと思います。

●**Dekker** お聞きした内容は非常に有用でしたし、多くのことを学びました。急いで決断しないことが重要です。RDA の報告書「Riding the Wave」は、何が決定されているかを理解せずに決定を下さないよう警告しています。私は今日、オープンサイエンスに関する日本の状況を概観することができました。日本には既に有用な情報がたくさんありますし、インフラや政府など、いろいろな要素もあります。皆さんは恐らく、パイロットやディスカッションを通してさらに理解を深める必要があるでしょう。

同時に、野心的な目標を設定することが課題です。

オランダが EU 理事会の議長国になったとき、私たちは野心的な目標を掲げたと思っていましたが、その95%を達成しました。そのため私はもっと野心的であってもよかったかもしれないと思っています。最初に立てた目標を達成できなかったとしても、それは大きな失敗ではありません。野心を持って、一つ一つの要素をつなぎ合わせていってください。

「オープン」の目指すものはさまざまです。ですから、自身が求めるオープンとは何かを明確にしてください。ユーザーは誰か。誰が論文を読むのか。誰がデータを使用するのか。皆さんはコミュニティを拡大するためにデータを再利用するに当たって、初期採用者やアンバサダー（代表者）を見つけることができます。ピーター・スーバー（Peter Suber）が述べたように、オープンサイエンスの最大の障壁は誤解です。アウトリーチは非常に重要です。

最後に、ODEN やオープンアクセスリポジトリ推進協会（JPCOAR）のような素晴らしい例が幾つか見られました。これらは、前進し、野心的な目標を設定し、構成条件を設定するための優れた取り組みです。皆さんはどんな条件の下で何をしたいですか。それを明確にすることが所有権を管理することにもつながり

ます。特にテキストとデータマイニングを考慮すると、これは重要な問題です。

市場を開放し続けることは重要なことです。少なくともヨーロッパでは、出版はもはや自由市場ではありません。一部の出版社がルールを決めて、図書館がそれに従ってお金を支払うという構造になっています。私たちはデータをオープンにするための市場を維持したいと思っています。そのためには、多くのユーザーと生産者が必要です。生産者とユーザーのつながりを促進するためには、図書館員のような仲介者や、ODEN が必要です。

●林 ありがとうございます。ODEN チームには非常に力強いご講評を頂きました。

小野さんから日本もそんなに遅れているわけではないという話がありましたが、日本もいろいろな取り組みを、しかも地に足が着いた形でしているということを Ron さんに示せたと思います。ですから、ここからは国際的にも問題になっていることについて議論していければと思います。

会場から頂いた質問が多様で非常にまとめにくいで、出だしからハードな話題に行きたいのですが、「小野さんの講演にあった『データのオープン化を強要、義務化するのはいくく悪い』という意見について、パネリストの皆さんの本音を聞きたい」というご質問があります。拙速過ぎるのは良くないということかと思いますが、小野さんは既に態度を表明されているのですが、他の皆さんはいかがでしょう。小賀坂さんなどは実はこの件は前から一言お持ちの方だったので、どうですか。

●小賀坂 これはイエスでもありノーでもあります。ご存じのように、オープンサイエンス推進については、欧州はかなりトップダウンで進んでいます。対して、アメリカはかなりコミュニティドリブンで進んでおり、これは方針の問題だと思います。大事なことは、政策として進めたいと思うなら進めれば良い、強制もすれ

ば良い、ただ、その結果に対してきちんと国や行政が責任を持つべきだ、それに尽きると思うのです。強制的に研究者の振る舞いを変えさせた方がひいては国の産業競争力の強化に役立つと思うならば、そうやれば良いけれど、その帰結についてはきちんと責任を取るべきです。

と申し上げた上で私の個人的な考え方は、研究者は自身のやりたいようにやるので、彼らに受容されないポリシーは意味がないということです。JST はずっとその方針でやってきています。今回もその方針を取っているつもりです。

●林 この点、他のパネリストで何か申し上げたい方はいらっしゃると思います。大向さん、どうですか。

●大向 今日は図書館活動を技術的に支える立場として参画しておりますが、私も技術的なソリューションのないポリシーは勘弁してほしいと思います。もちろん、法律や何かの制度によって固めることは必要だと思うのですが、それを裏付けるテクノロジーがなくて、研究者がメインの課題ではないところに時間を使うことになってしまい、最終的なアウトプットのクオリティを下げることになってしまうのであれば、それは非常に問題です。ポリシーとそれを裏付けるソリューションを一緒に考えたいということが一番思うところで

●林 ということで日本は、どうしても、という枕が正しいかどうか分かりませんが、研究者側からのボトムアップ型になりがちです。でも、ある OSTP (Office of Science and Technology Policy : アメリカ合衆国科学技術政策局) の方の言説を借りると、ボトムアップだけに頼り過ぎるといつまでたっても時間がたつばかりで進まない、だから、carrot and stick の stick が必要なのだとおっしゃっています。今日は既に幾つかヒントはあったと思うのですが、ボトムアップのアプローチをより加速させる手法について、何かパネリス

トの方、あらためて申し上げたいことはありますでしょうか。北本さんなどはたくさんあるのではないのでしょうか。

●北本 私も、基本的にはオープンにした方が得だという状況で進めたいと思っています。ボトムアップなのですが、ボトムアップの単位がどこかだと思うのです。ある経済学では、個人として最適な行動をすると全体最適にならないという合成の誤謬のような考え方があります。ボトムアップで本当に個人として最適な方向に行くと、全体最適にならないというのは、恐らくオープンサイエンスでも同じだと思います。

ですから、「個人としての最適性をどうぞ追求してください」だと、いつまでたっても進まないかもしれませんが、全体の、例えばコミュニティ単位で最適化するというのであれば、要は、個人としての最適性から少し違う方向に進んでくださいということであれば、全体としては良くなる可能性があるのです、そういう単位で偉い人が考えないと進まない面があるのではないかと思います。

●林 そうですね。リサーチコミュニティの見識のようなものが研究者の集団をリードすると。Center for Open Science の Brian Nosek さんは、オープンサイエンス・フレームワークをつくるときに、あまり先を走り過ぎると人が付いてこない、寄り添い過ぎるとそこに停滞してしまう。だから、マラソンランナーを教育するときに、10m 先にペースが速い人を走らせるように、少し前を走って付いてきてもらうのがよいのだとおっしゃっています。

そういうアプローチが日本でもできるかという話があるかと思うのですが、小野さんは実際に DIAS に関わっていて、今のようなアプローチは絵に描いた餅だといった話など、もう少し具体的なエピソードをご披露いただけますでしょうか。

●小野 確かに、おっしゃるとおり、研究者は言わな

いとなかなか動かない、周りに言われないと自分の好きなことだけをやり続けるという面があると思います。強要するのが良くないと言ったのは、今すぐデータを出せというような言い方は良くないのではないかと思います。

小賀坂さんが DMP でされていたように、考えるところから始めろという強制の仕方はありだと思います。取りあえず、公開を前提としてまず考えてくれ、できないならできない理由を考えろという強制の仕方はありかなと思っています。いきなり出せ、とにかくいつまでに出せ、明日出せというのは良くないでしょう。

それでもなかなかやらないようであれば、これも小賀坂さんがおっしゃったことに私も全く同感なのですが、国が責任を持って強制させる、ただ、強制させた結果、うまくいったかどうかを検証する必要があると思います。間違っていたらやめなければいけないし、それでオープンサイエンスが進んだという結果が出たのであれば、それは研究者の側が間違っていたことになります。国が責任を取る、きちんと検証するという前提で強制するならば、それはそれで一つのやり方ではないかと思います。

●林 Ron さんに戻してみたいのですが、今のうちに、日本ではボトムアップ型である中で、政府がトップダウン型で進めるなら、きちんとポリシーに対する責任を担保しなければならないという議論になりました。一方で、ヨーロッパは Horizon 2020 に従って、例えば 2020 年までにオープンアクセスは 100% 実現するということを明快にうたっています。既に宣言されたポリシーに対する責任、チェック体制についてはどのようなになっているか、お聞かせ願えますでしょうか。

●Dekker オランダが欧州理事会の議長国のとき成功した要因の一つは、政治的バックアップを得た包括的アプローチでした。大学と資金提供者は、自分たちはオープンアクセスを望んでいると述べました。一部の出版社でさえ、私たちは変化が欲しいと言っていま

した。共通の目標を持った包括的な提携があれば、前進することができます。私は、何が求められているかを明確にして、官僚的形式主義を最小限に抑えることに気を使いながら、トップダウンのアプローチを組み合わせたいと思っています。オランダのリサーチカウンスルが直ちにオープンアクセス化したいと言えば、それは誰にとっても非常に明確です。

先週、ある若手研究者が、「全員に適用される限り、研究者は義務的であるかどうかは気にしない」と言っていました。それが義務でなければ、研究者は余分な努力をしたいとは思わないでしょう。誰もが変わるよう誘導されるべきです。科学者も（他の人間と同じように）変化を好みませんが、時には特定の方向へのインセンティブが必要な場合もあります。だからこそ、私たちは政治的声明と、それをしたいと主張する資金提供者の声明を提供しようとしたのです。慣れるには時間がかかります。私たちにはパイロットがあり、再びゲイツ財団の例を挙げると、結果がオープンアクセスで公開されるというのが資金調達の条件です。

●林 という状況なので、やはりトップダウンは慎重でありつつ、ある程度は必要になってくると思います。そこを日本のファンディングエージェンシーを含めて、今後それなりにポリシーを定めていく必要があるのではないかと思います。私もどちらかというのを後押しする役割で今 NISTEP にいるとも言えます。

これは大枠の話でしたが、個別に見ると小野寺さんへの質問が一番多いです。「既存の研究データ管理ツールではなく、自前でデータ管理をすることのメリット・デメリットを知りたいです」というご質問があるのですが。

●小野寺 自前で管理するというのは、既存のもので足りない部分があったから、それを補うために自分たちで開発するというのが背景だと思います。ですから、デメリットについてはないというか、あってもそれを解決するために、コミュニティの中で議論して、その

デメリットも解決する方向に持っていこうとしていると理解しています。

●林 強いて言うと、課題はモチベーションの維持だったりしませんか。やりたい人がいるうちはいいのだけれども、そういう人が異動してしまったり、モチベーションが下がったりしたときにどうするか。自前でつくるときは、結構ボランティアなリソースを使うことが多いと推察されるのですが。

●小野寺 人がいなくなったときという問題は確かにあると思うのですが、ご紹介した NIMS の事例では、小さいグループではなく、同じ計測系の研究グループ、組織の中のコミュニティとして活動しているので、ある人が抜けたからといってそこでストップしてしまうことにはならないと思います。

●林 コミュニティとして持続性を確保するという感じですか。

●小野寺 はい。

●林 先ほどの北本先生の、ボトムアップはコミュニティベースでという話と通ずる議論になってきて、狙っていなかったのですがきれいにつながりました。

次に、ライセンス、著作権、ビジネス回りのところを気になさっている方が多く、「NIMS のデータで無償と有償と制限公開がある。その使い分け、違いの基準を機関としてどういう方針でつくられているか」。これは説明があったような気もするのですが、あらためてこの部分についてご紹介いただけますか。でも、これは研究者側が選択しているという感じで、機関として何かそれを推奨する、ガイドをするということはないのでしょうか。

●小野寺 そうですね。

●林 それは DIAS もそうですね。出すデータの公開レベルを研究者側が選べると伺っているのですが。

●小野 はい、基本的には、データ提供者がどうしたいかというのを中心に決めていただいているという状況です。

●林 今度は制限公開からビジネスサステナビリティの方に話を変えたいと思います。有償でやる場合に、まず小野寺さんに質問をしたいのが、イギリスの会社と組んでお金を取っているという話だったのですが。

●小野寺 お金を取っているのはイギリスの会社です。

●林 NIMS にもお金は流れてくるのですか。

●小野寺 はい。

●林 それは差し支えない範囲で、どのぐらいをカバーするのでしょうか。

●小野寺 この発表でご紹介するためにインタビューをしたのですが、そこまでは聞いていないので、お答えできません、すみません。

●林 その辺が結構クリティカルになってきて、ある質問では、「私の分野ではデータベースを研究者に対しても有償で売る研究者がいて、データ統合を進める上で大きな障害となっています」というご指摘があります。一方で、データベース事業を回すためには、お金をもうける目的ではない範囲で収入を得ることも重要な場合がある、これがデータ研究管理基盤の大きなトピックになっています。

これについて、小野さん、DIAS 側からとして、確かソフトウェアを販売するなどいろいろなケースがあったと思うのですが、見通しについてはどのようにお考えでしょう。

●小野 DIAS 自体もビジネスや自律的な運用に向けて検討は進めていますが、実際にそこまでいいないので、いろいろ考えながらやっていっている状況です。さまざまな課金モデルが考えられるのですが、先ほどのお話にあったような、データに対してお金を取るというやり方は、ちょっとうまくいかないのではないかと今のところ思っています。

では、どういうところでお金を取るかとなるのですが、やはりデータをどのように使いたいかによるのです。われわれがやっているのは地球環境関係のデータ基盤なので、これがあつたらユーザーも満足だし、世の中の地球環境の課題に役に立つというアプリケーションをつくるために DIAS にあるデータを使って開発するという場合、アプリケーションの開発部分にユーザーにお金を出してもらおうようなモデルがあるのではないかと考えているところです。

●林 その一方で、事業の持続性が見えない中で、やみくもに研究データを登録することに対する懸念を示す質問もありました。要は、体のいい研究者のバックアップ保存先として使われてしまうのではないかと懸念ですが、それについて何かコメントができる方はいらっしゃいますか。

まだそういう話よりも、いかにデータを登録するかというフェーズではあるかと思うのですが、早晩、逆転するはずですよね。みんなが登録するようになったら、全部は登録できないとなって、何かを捨てる方の議論になるのですが。今このテーマを扱うのはちょっと難しいでしょうか。

●北本 私が DMP で一つ恐れているのはそれなのです。義務化と言って、「では出せばいいのか」と形式的にデータを出すことが広まってしまうと逆効果になりかねません。あまり規制を厳しくするとそういうことが起こり得るので。

●林 トップダウンで早くやり過ぎるとそうなってし

もう。

●北本 やり過ぎると、では出せばいいのかということになってしまいます。例えば、DIAS だとセクションはしているのですが、そのセクションをどこまですべきかという問題があります。figshare などは何でも受け付けていますが、それでよいのかという疑問はあります。そこをどうするのかは結構大きな問題です。ただ研究データの墓場のようにすべきでないとは思います。

●小賀坂 似たようなことがかつて知財でも起こりました。昔は基礎研究の研究者は知財取得に対しては無関心でしたが、国がそれではいけないということで、研究者に知財を出しなさいと言う時代がありました。

ところが、その維持経費が莫大になって、支えきれないまで膨れ上がり、逆に何でもかんでも出せばいいというものではないという時代が来ました。大学も、最近ではライセンス可能な特許しか支援しない傾向にあるそうです。今度は逆に、将来化けるかもしれない基本特許すら出せないという寄り戻しが来ています。ただ、多分バランスが取られていいところに落ち着くのではないのかと思っています。

データについては、明らかに保管インフラに掛かる経費とのバランスですよね。多分、あまりアプリアリな答えはなくて、現実的なところで落ち着くのではないかという気がしています。

●林 今の議論に関して、Ron さん、ヨーロッパではデータをどこまで持つか、あるいはデータをどうやって捨てたらいいかという議論はどのぐらい進んでいるのでしょうか。それとも、まだ研究者にデータを登録してもらるのが先というフェーズにあるのでしょうか。

●Dekker それは分野によります。物理学や生命科学のように多くのデータを生み出す科学では、データを保管するためのプロトコルがあります。大学には

10 年間データを保管するという行動規範があります。プロセスの一部として作業が終了した後、その作業成果を単純に捨てることのできるセクターは他にありません。民間部門でそうしたら解雇されるでしょう。したことについて責任を負うことは、良い研究をすることの一部です。

オランダでは、心理学の分野で大きな研究不正が起きました。それまで高く評価されていた教授がデータを改ざんしていたのです。データを作成するように求められたとき、彼はそれを渋りました。引っ越しのときに調査結果を紛失したというわけです。彼が最終的にデータを提出したとき、統計モーメントが全て 3 次および 4 次モーメントと同じだったので、これは不正だとみなされました。この事件は、再現性の確保のためにデータを保管しなければならないという議論に拍車をかけました。

再利用のためにデータを保管することの必要性に関する議論はあまりありません。保管はコストのほんの一部です。コストは、研究者とそのアシスタントがメタデータを追跡し記述するために費やされる作業時間に掛かるものです。さらに、IPR（知的財産権）は非常に複雑です。少なくともオランダでは、データの権利所有者は、（一部においては）データを得るための資金を提供した者、つまり資金提供者または大学になります。彼らはデータを追跡し、記録する責任があります。他の研究者がデータを再利用できるように、データを見つけられる状態にすることが課題です。議論は、データを保管すべきかという問題から、再利用に適したデータの品質をどう保証するかという問題に移行していくでしょう。

●林 日本は小さなところの議論をしてしまったのではないかという感じになっていますが、話を変えます。今日の話題提供の中で一つ大きなインパクトのある話は、ウェルカムトラストがつくったようなオープンリサーチ・プラットフォーム、ないしはプレプリントサーバーという意味での arXiv の広がりだったと思いま

す。それに関するご質問が何点か来ています。まずは小野さん宛てに、「プレプリントサーバーのような、雑誌を経由しない論文の公開の場があれば、研究者にとって本当にいい世界になると思いますか」という質問が来ています。研究者を代表するわけではありませんが、今日の話を伺っていかがでしょうか。

●小野 何とも言えないところがありますが、プレプリントサーバーなりに論文をどんどんアップロードしていくという活動が評価されるようであれば、そういうものに乗る研究者の方もいるのではないかと思います。でも、大抵の研究者は、それなりに権威があるところをファーストプライオリティとして投稿するのではないかと思います。

最近聞いた話で、PLOS ONE はオープンで比較的投稿しやすいということで、一時期とても盛り上がり、投稿数が増えていたのですが、だんだんそれが減って、一時の盛り上がり下がってきているようです。だから、今プレプリントサーバーのようなものを用意しても、そんなに集まらないような気がします。

●林 生々しい本音をありがとうございます。一方で、北本さんのご紹介のように、ディープラーニングの世界ではスピードがすごく大事ということで、極端な表現をすれば、査読なんかしてられない感じですね。だから、そういうのは新しいディシプリンの中で機能するだけなのか、既存のエスタブリッシュされた研究、ディシプリンにも影響を与えるのか、これがこれから注目されていくことになるのだろーと思います。

その辺の見通しについて、パネリストの皆さんは何か思うことはありませんか。あるいはそろそろ会場の方から何かありましたらお願いします。この話はみんなで考えないといけない、新しいパラダイム、研究パラダイム、評価パラダイムを生み出すためのプラットフォームづくりにつながると思うのです。端的に言うと、旧来のピアレビューをいい意味でどのように捨てていく、変えていくかというような話にもなると思う

のですが。Ron さん、どうぞ。

●Dekker それは品質評価を破棄するものではありません。ウェルカムトラストは二つのことを行っています。Wellcome Open Research にはピアレビューされた出版物が含まれますが、大きな違いは、まず発行して、その後、レビューが行われることです。

ウェルカムトラストは、プレプリントを認める多数の資金提供者の連合に加わっています。プレプリントを掲載するためには、プレプリントを保管するためのアーカイブや施設が必要です。幾つかの分野、特に生物学、医学、保健においては、成果物にタイムスタンプを入れることが非常に重要です。それは、導入、分析、データ、結果、結論を含む本格的な論文である必要はないということにもつながります。論文が幾つかの重要な結果に分解されることになります。プレプリントでも、研究コミュニティにとって重要で関連性の高い結果を公表することができるのです。それはまさにエボラ危機の間に起こったことです。他の研究者がこの伝染病を抑える方法について結論を下すことができるように、データと結果ができるだけ早く公開されました。

●林 そうですね。結局、とにかく出して、その成果が公開されるタイムスタンプをまず付ける。その後、今までどおりのピアレビューもやって、今までの評価のスタイルも担保する、そういう世界が始まろうとしています。これもまた、ディシプリンによると思うのですが、そういう流れの中で、研究機関、大学職員、研究者そのものをどう育成するかという話になると思います。そういう質問も来ています。

「プレプリントに出すのが当たり前の世界、ないしはオープンリサーチ・プラットフォームのような、とにかく出して後からピアレビューをするという世界になるとき、大学や研究機関の図書館員の育成、採用方法を大幅に変える必要があるように思いますが、そのような可能性は日本にあるのでしょうか」という、悲

観的なご指摘も頂いています。

意識すると、プレプリントサーバーが当たり前の世界、オープンリサーチ・プラットフォームが当たり前の世界、公開までいなくても研究データのシェアリング、コラボレーションが進むような世界になったときに、大学の職員やその他はこの激動の時代にどう対応していったらいいのかということだと思います。

●大向 何が起こるかさっぱり分からない、明日になればポリシーが外的要因によってがらっと変わってしまう、そして最適な情報システムの姿がまるで変わってしまうということは、まだ学術情報分野ではあまり経験しないことかもしれませんが、ウェブの世界で私たちは散々経験しています。Facebook が来た、Google が来た、Amazon が来た。いつの間にか、フィットしなければいけないというわけではないけれど、フィットした方がより生活のクオリティが上がるという状態に置かれる中で、近年、図書館が幾つかある選択肢のうち、どの情報システムを選ぶということぐらいしかやってこなかったことはとても危ないと思っています。だから、自分でつくらないといけなるときにはつくらなければいけないのだ、制度もアーキテクチャーも含めてつくる側に回らなければいけないということを、何とかメッセージとして伝えたいと思って、NII の研修プログラムを数年前から始めているのです。

研修プログラムで目指していることは、知識を増やすことではなく、動じないというか、新しいシチュエーションに対して、クオリティが高いかどうかは別として、ソリューションをひねり出して次に行こうと思えるようになることです。それを今鍛えておかないと怖いと思ってやっています。

●林 10 年前、機関リポジトリを始めたころに比べると、格段に図書館の方々が研究に対してどうコミットできるかを一生懸命考えています。10 年前だとまだ、「出版した論文の著者原稿を下さい」と言うのが研究者とのコミュニケーションで大事なことだと真面

目に議論していました。それはそれでその当時は大事だったのですが、10 年たって、また違うコンテキストで研究データにどう向き合うかを考えています。それを繰り返して、本日も北本さんが 2027 年のデータライブラリアンのスキル要件についておっしゃっていましたが、ライブラリアンが何に対応できるかきちんと考えれば、多少ポジティブな方向に向かうのではないかと思います。

北本先生、何か追加のコメントはありますか。

●北本 大向先生がおっしゃったように、ウェブの世界はずっと変わってきたのに、学術情報が変わらなかったのもう限界というところではないかと思えます。ですので、今後大きく変わっても別に不思議ではありません。

先ほど Ron さんもおっしゃいましたが、プレプリントサーバーができたから査読が不要になるということではないのです。何らかの品質管理は依然として必要です。ただ、品質管理の方法がピアレビューだけではなくる可能性が、プレプリントによって開けます。

自分でチェックすべきものを他人にアウトソースするというか、他のレビュアーのチェック結果を使わせてもらうことによって品質管理が簡単になるというのがピアレビューの仕組みですが、今後はそれを AI が行うかもしれません。あくまで目的は品質管理であって、ピアレビューという手段が欲しいわけではないというのが変化だと思います。

●林 そうですね。ピアレビューという仕組みも、手紙で研究情報をやりとりする時代に生まれたレガシーなフレームワークなので、必ずしも様式にこだわる必要はありません。けれども、これも分野によると伺っています。例えば high energy physics の研究でプレプリントがはやったのは、書く人数と読む人数が同じくらいで小規模なため、査読をしなくてもその中身が分からない人はもぐりだという文化があるから進んだからだという話を物理の研究者に聞いたことがあります。

一方、私が専攻した、製薬につながる有機合成化学は、書く人に対して読む人の方が圧倒的に多いので、信頼できる情報が欲しい、誰かによって信頼が担保された情報が欲しいということで、ピアレビューが重要になります。ただ、今の ICT のインフラ、ハードウェア、ソフトウェアがあり、これからデジタルネイティブの人材も増える中で、新しいクオリティコントロール像をどうつくっていくかという話になってくると思います。

さて、これはどこで出しているかわからなかったのどこまで引っ張ったのですが、Ron さんに質問があります。「アメリカではアメリカファーストの政策の下に保護主義的な動きが見えて」、ここからは私が付け加えます。「環境系の研究などは情報を出すことを制限するような動きがあります。さらに、イギリスの EU 離脱もあり、ヨーロッパに所属している国々が保護的に動く恐れがあるのではないのでしょうか。そういう中で、Horizon 2020 や European Open Science Cloud を進めていく上で、どのように見通しを立てていますか」というご質問です。

●Dekker 私はアメリカが保護主義であると言ったのではなくて、保護主義的世界が起こりつつあると言いました。しかし、新しい知識を深めたい場合、早い段階で可能な限りオープンにしなければなりません。お金を得たい場合は、ある段階まで知識を保持して、特許またはライセンスを取れば良いと思います。

Structural Genomics Consortium を例にとると、非常に競争力のある製薬会社ですら、独自で事業を行うには複雑過ぎるし、あまりにも高価になってしまうので、共同で行うことを決めました。資金は大きな要素になります。自分にとって何が良いのか、他にとって何が良いのか、両者にとって良いものは何かを考えて、win-win の状況をつくり出す必要があります。リサーチを行う際、データの再利用は推進力となります。リサーチをすることによって、選挙で生き残ることができます。ヨーロッパでは、カトリック教会の次に大学

が最古の組織です。1500 年以上存続しており、恐らく将来も存続するでしょう。

しかし、変化が速くなっています。追求すべき方向が分からない場合、まず 10~15 の花を咲かせて始めてみましょう。それが私がアメリカで見いだしたものです。試してみて、それが駄目なら、進んでその取り組みを中止するのです。そうすることによって、専門知識を構築していくのです。これが不確実な新しい未来に備えるための最良の方法です。

●林 頼もしいコメントをありがとうございます。これで全ての質問でもないのですが、大体お答えしたことになります。ここからフロアからのご指摘やご質問なども受け付けたいと思います。

●深貝 今までオープンサイエンスという話になると、図書館側はどうやったらいいかという戸惑いがあり、研究者側は成果は自分のものとして上げないと競争に勝てないという思い込みにはまり込んでいました。しかし、今日のお話を伺っていると、図書館あるいは情報を学術的に管理する方々が新しい試みをしつつあって、多いに励みになりました。今の学術研究のスタイルにどっぷり漬かった、「自分の研究データは自分のものである」という人々が変化していく可能性を開くためのプラットフォームをいろいろつくっていただいていると思います。

そういう意味で、今後は実際にデータを集める仕組みを試してみたり、ユーザーとしてデータを試してみたりすると思います。具体的には、この面についてもの分かりがよい研究者を頼りにしながら、特に学生と接している教員に頼んで、実際にデータを使ってみます。また、柔軟な発想の院生たちにデータを使ってもらいつつ、データを集めて登録することに参加してもらいます。それから、狭いディシプリンの考え方に染まっていない学部生に、報告にあった ODEN を使ってもらえば、分野の垣根を越えた柔軟な発想がどう生きるかを試してみることができると思うのです。

ただし、図書館の職員だけでそれは無理で、教員に頼まなければいけないので、実験的にそういうデータベースの新しい試みをなされている方々は、ぜひ図書館長を巻き込んでいただくと進んでいく可能性はあります。逆に、弊害が見えるかもしれないけれども、それはそれで貴重なことだと思います。

●林 ご指摘ありがとうございます。深貝さんはまだ図書館長でいらっしゃいますか。去年まででしたか。「現図書館長から頼もしいお言葉を」と返そうと思ったのですが。そうやって事例に取り組んでいく、進めていくこと自体が大事であるということですね。今のお話は、FOSTER という Horizon 2020 の取り組みを想起させます。若い人にオープンデータだけを使って研究をさせるという研修プログラムがあると伺っています。

最後に、今日のテーマは「オープンサイエンス再考」なので、今日のディスカッションないしはこの1年のSPARC Japan を振り返って、あなたにとってのオープンサイエンスとは何か、何がキーになるかをパネリストの皆さんから一言ずつ頂いて終わりにしたいと思います。

●小賀坂 JST は、決まった額のお金で成果を最大化していただくためにやれることをやるという立場ですから、その文脈でオープンサイエンスを推進すべきなのですが、この議論をするときに、ステークホルダーが偏っていると思うのです。今足りていないのは、一つはドメインサイエンティストだと思います。この話にリテラシーのない人たちにきちんと入ってもらわなければいけないのです。もう一つは、リワードシステムに関してなのですが、評価担当者です。評価担当者をどうやって評価に組み込んでいくかをやっていただきたいと私の立場では思います。

●小野寺 材料科学の分野では共有できるものとそうでないものがありますが、共有できるものについても、

そのデータの持ち主がコントロールできるような、どう使われているかが把握できるような仕組みができればと思っています。学術情報流通は研究者がコントロールできない状況になってしまいましたが、オープンサイエンスを進めるにあたっては、研究者が意図しない方向に行かないようにコントロールできればいいと思います。

●小野 私は最初に間違ったメッセージを発してしまったかもしれないと反省しています。「焦ってはいけない」とは言ったのですが、それは出さないという意味ではなくて、やはり出せるデータと出せないデータがあるのです。出せるデータも、まだすぐには出せないから、ちょっと時間を下さいという部分があるということです。

●林 エンバーゴの議論になっていなかったですね。

●小野 そうですね。だから、それを今すぐ出せと言うのはやめてほしいということです。われわれも徐々にそういうことを考えていきながら、出せるタイミングができたら出していきますということなのです。それを言いたかったのです。

まだ出せないデータはどうしても出せないのも、それはちゃんと理由を聞いていただいて、「それだったらしょうがないね」と納得していただきたいということがまずあります。今出ていないデータも、今後実際は出ていくと思います。そこに関してはすごくポジティブに見ています。

年代によって考え方が違うという話をしましたが、先ほど、まだ偏見がない学生と一緒にやったらいいのではないかというお話がありましたが、全くそれに同感で、下の世代に行くほどオープンマインドなのです。データもシェアするし、何だったら個人情報もシェアしているような、そこはやり過ぎではないかという気がするのですが。時間がたてば世代が変わるので、今後どんどんそういうことが進んでいくのではないかと

思っています。時間がたてばいずれ解決する問題があるので、焦らずじっくり構えていけばいいのではないかと思います。図書館の方も、気難しい研究者と一緒にやるよりは、なるべくオープンマインドの人を探して、一緒に活動が続けていった方がいいのではないかと思います。

もう1点、小賀坂さんが、ドメインサイエンティストのオープンサイエンスの議論への参加が足りないとおっしゃっていたのですが、私はその他にも、経済や法律の専門家の参加が特に日本は足りてない気がします。私はベルモント・フォーラムのデータシェアリングに参加していましたが、そこには経済と法律の専門家がいました。経済は、サステナビリティを考えていく上で必要になります。法律は、これも結構重要で、データの利用規約のところで結構研究者は悩んでいるのです。そこがどうしたらいいか決められないので、出せないという人が多いのです。逆にこういうところを国に支援していただいて、どういうデータの利用規約が研究データにとって必要かを考えていただきたいと思います。

データの法的なライセンスの話は、われわれも弁護士に相談したのですが、法律も、知財に詳しい人、民事に詳しい人など、分野が分かれています。その中でデータの専門家が全然いないので、相談する相手がいないという状況です。それは研究者のレベルではなかなかできないところなので、その辺の支援を考えていただきたいと思っています。

●林 大事なポイントですね。

●大向 私にはあまり深い見識はないのですが、オープンデータ、オープンガバメント、オープンサイエンスなど、「オープン」の付くムーブメントと長く付き合ってきて思うことは、「オープンは笑ってやること」です。しかめっ面しながら、「オープンにしなければ」と言いつつやることは決してうまくいきません。新しいものを取り入れるときは、取り入れたらこんなこと

ができるようになった、こんな機能があることが分かった、今まで見えていなかったものが見えるようになったという理由で、みんながアダプトしていくというプロセスがあります。オープンかクローズドかは、普通にコスト計算をすると、クローズドの方が良かったとしても、オープンの方が面白いから、楽しいから、便利だからということで乗り越えてきたこれまでの歴史があるのです。特にウェブを見ていると、それが成功要因だと強く感じる場所があります。

具体的にどうというわけではないのですが、オープンサイエンスとはどのようなものか体験してもらうために、オープンサイエンスをすることはどのような感覚なのか知ってもらうために、体験にフォーカスした形で仕組みをつくっていく、動かしていくことにも目を向けてみてもいいのではないかと思います。この議論に限らず、普段思っているのはそういうことです。

●北本 私は、本来的に、オープンがいいところもあるし、クローズドがいいところもあると思うのです。それが歴史的な経緯によって逆転しているようなところがあって、必ずしも全体的に一番いい形になっているとは思えない部分があります。ですから、オープンが全てでもないし、クローズドが全てでもないのですが、どちらがいいのかも一度考えるといいと思います。本当はオープンがいいのに、いろいろな経緯でクローズドになってしまっている部分はオープンにしませんか、ということではないかと思っています。

「研究とは shoulder of giants（巨人の肩）の上に乗って行うものである」というニュートンの有名な言葉がありますが、そういう「巨人の肩」をどうやってつくっていくかが、科学の非常に大きな課題です。それは個人でつくった方がいい場合もあるし、集団でつくった方がいい場合もあります。集団でつくるといのはまさにオープンなサイエンスだと思います。何となく人工的なバリアをつくって、何となく個人でやって、無駄な競争をして消耗するのはあまり賢くないので、集団でつくった方がいい場合に、オープンサイエンス

を取り入れることが非常に重要だと思います。

●Dekker あるときは個人で、あるときは集団でという考えに同意します。私にとってオープンサイエンスとは、科学が変わる一つの道です。好むと好まざるにかかわらず、科学は変わっていきます。列車に飛び乗ってそれをフル活用するのは大学や国次第です。私は、日本はオープンサイエンスを実現するための要素の全てを持っていると思います。皆さんは優れたインフラを持っていますし、列車に飛び乗ってそれをフル活用できる明晰な人たちがいます。今までも革新的な新製品を開発してきたと思いますが、今や革新すべきは科学であり、その列車に飛び乗るのです。私ならば政府を待たず、若い人たちに投資します。

●林 ありがとうございます。第1回セミナーでは、グリーンやゴールドといった研究論文のオープンアクセスの話をして、第2回セミナーでは研究データの話をしました。論文のオープンアクセスについては大体固まったという議論がある中で、今日衝撃的だったのが、グリーンでもゴールドでもないプレプリントサーバーないしはオープンリサーチ・プラットフォームに載せた後に、ピアレビューを行うという仕組みが生まれたことです。このような形で、まだまだ学術情報流通は変わりつつあります。論文の世界ですら、変わりつつあることが今回分かりました。

そして、研究データの方も、インセンティブやビジネスモデルに問題はあるのですが、今日は日本でも地に足の着いたさまざまな試みが繰り返されているという報告がありました。NII では研究データ基盤も来年度から進むという時代になりました。

ということで、いよいよ policy into practice、オープンサイエンスを実践する段階に入ってきました。できることが山ほどあって困るぐらいになっていると思います。お越しいただいているステークホルダーの方々に、科学研究そのものが変わるというビジョンの下、今後5年間、10年間の間にやれることは何かという

課題をお持ち帰りいただければ、本セミナーないしは今年度の SPARC Japan セミナーはうまくいったということになるのではないかと思います。

これにてパネルディスカッションを終了させていただきます。どうもありがとうございました。