

SPARC*Japan NewsLetter

ISSN 1883-826X

NO.44 2021 年 6 月

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。

※所属、肩書はすべて開催当時のものです。

CONTENTS

■ SPARC Japan 活動報告

学術情報流通推進委員会

arXiv.org 活動支援について

arXiv.org の 2022 年度以降の年会費
モデルについて

CLOCKSS 活動支援について

SCOAP³ 活動支援について

■ SPARC Japan セミナー報告

企画概要

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

学術情報流通推進委員会

学術情報流通推進委員会の会議資料をウェブサイトで公開しています。

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



arXiv.org 活動支援について

arXiv.org< <https://arxiv.org/> >は、物理学、数学、コンピュータサイエンス等のプレプリントサーバとして、同分野を中心とした研究成果のオープンアクセスとアーカイブを進めています。2014 年 12 月に 100 万論文を突破し、2019 年には約 155 万件、年間ダウンロード数は約 2.6 億件以上で、総ダウンロード数は 2020 年 3 月に 15 億件を超えました。

利用件数上位の機関による財政支援があり、2013 年に開始した「arXiv.org 会員制プログラム」により 2019 年末現在で 30 カ国 243 機関が参加しています。

日本においては、arXiv.org の利用実績が多い研究機関がコンソーシアムを形成してこの取組みを支援しており、2009 年から国立情報学研究所にて、参加意向確認および年会費の取りまとめを行っています。

arXiv.org の利用頻度が高い国内機関に対して 2020 年の参加意向調査を行った結果、2020 年 12 月末現在の参加機関は 16 機関です。

関連会議は、以下のとおり開催され、日本の参加機関を代表して国立情報学研究所武田英明教授が参加しました。

- Annual Meeting 2020 年 10 月 5 日～7 日（オンライン）
- Member Advisory Board 会議 2020 年 12 月 14 日（オンライン）

arXiv.org の 2022 年度以降の年会費モデルについて

「arXiv.org 会員制プログラム」では、長期的に十分な運営予算を確保できる持続可能性モデルを確立すること等を目的に、論文ダウンロードモデルから、論文投稿モデルへの変換が 2022 年度から行われます。

従来 Tier は 3 年間の平均ダウンロード数の機関別ランキングに基づいていましたが、機関の IP アドレスからのダウンロード数は arXiv.org 全体の 15% しか占めておらず、このランキングでは、年による大きな変動や集計コストの発生という課題がありました。

新しい会費モデルでは、ダウンロード数ではなく、3 年間の平均投稿数に基づいて会費のランクが決まり、Tier も 1-6 までであったのが、ゴールド、シルバー、ブロンズの 3 つの主要な Tier となります。

2021 年 4 月 15 日に arXiv.org 事務局と日本の arXiv.org 参加機関のミーティングを行い、これらの情報が共有されました。

国立情報学研究所では、引き続き 2022 年度以降の会費モデルについて情報収集を行い、参加機関に対して情報を共有してまいります。

CLOCKSS 活動支援について

CLOCKSS (Controlled Lots of Copies Keep Stuff Safe) <<https://clockss.org/>>は、全世界の研究者のために電子的な学術コンテンツの長期保存を実現することを目的として活動しています。具体的にはアーカイブとそれを運営するコミュニティを構築して、コンテンツが出版社から提供されなくなった場合にアーカイブ上のコンテンツを広く利用できるようにするなど取り組みを行っています。また、コンテンツを保持する世界 12 のノード機関の一つとして、国立情報学研究所が参画しています。

日本においては、2013 年から国立情報学研究所にて、参加意向確認および年会費の取りまとめを行っています。

大学図書館コンソーシアム連合 (JUSTICE) 会員館を中心に 2020 年の参加意向調査を行った結果、2020 年 12 月末現在の参加機関は 101 機関です。

関連会議は、以下のとおり開催され、日本の参加機関を代表して国立情報学研究所武田英明教授が参加しました。

- CLOCKSS Board of Directors 会議 2020 年 6 月 23 日、9 月 16～17 日、2021 年 2 月 23 日（オンライン）

SCOAP³ 活動支援について

SCOAP³<<https://scoap3.org/>>は、欧州原子核研究機構 (CERN) が主導する、高エネルギー物理学 (HEP) 分野の査読付き学術雑誌論文のオープンアクセスを実現する国際連携プロ

ジェクトです。日本においては、SCOAP³対象誌を購読している研究機関を中心に、従来図書館が出版社に支払ってきた購読料を、論文出版加工料（APC）に振り替えることで、HEP分野の主要な学術雑誌のオープンアクセスを支援しています。

日本においては、2014年から国立情報学研究所にて、参加意向確認および年会費の取りまとめを行っています。

フェーズ3（2020-2022年）を支援する国内機関を中心に2020年の参加意向調査を行った結果、2020年12月末現在の参加機関は81機関です。

関連会議は、以下のとおり開催され、日本の参加機関を代表して高エネルギー加速器研究機構（KEK）野崎光昭主任 URA および国立情報学研究所山地一禎教授が参加しました。

- ・ GOVERNING COUNCIL 会議 2020年10月20日～21日（オンライン）

■ SPARC Japan セミナー報告

第1回 SPARC Japan セミナー2020



「研究データ公開：フルオープンと制限公開の境界線」

2020年10月2日（金） オンライン開催 参加者（Webex Events）：247名

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。

<https://www.nii.ac.jp/sparc/event/2020/20201002.html>

企画概要



研究データの公開が広く推奨されている一方で、条件を満たした特定の利用者によりのみデータを提供する制限公開を行っている学術分野がある。昨今、研究データのライセンスや機密保持の観点からこの制限公開に対する関心が高まっているが、どのようなルールに基づいて研究データを公開、制限公開、非公開に区別しているのか、データ所有者、データ利用者等のステークホルダーとどのような取り決めを行っているのかについてはよく知られていない。また、特定の研究者、機関内のみ限定して研究データを提供する制限共有との違いは何であろうか。

現在、制限公開や制限共有を行う機関から制限公開の実態（データ提供者との関係性、制限公開の提供フローとその労力、利用者情報の管理）を共有することで、機関が制限公開を行う際の課題を明らかにする。



参加者から

(大学/図書館関係)

・図書館の業務として今すぐ「役に立つ」わけではないが、特に前半のデータの定義や公開するデータの線引きの話などは、かなり実践的で研究データのオープン化の雰囲気がかなり現実的に把握できた。

(その他/図書館関係)

・人文系の研究図書館として、研究員の研究データの整理・公開・活用を考えている。大いに参考になる事例であると同時に、課題となる点も共有できた。

(その他/研究者)

・今回のセミナーを拝見して、これまでの研究データをオープンにしようという論調から、データ活動を維持していくための費用をどのように捻出するかという論調へ、現実路線へと焦点が移り変わった段階に来たことが見て取れた。

(その他/その他)

・データとは何か、公開・非公開の考え方、各機関の運用など大変勉強になった。講演者が非常にわかりやすく説明してくれて、大変おもしろかった。

・どの機関も実際のケースでの経験を積みながら、着実にデータについて取り組んでおられることが分かり、自身の業務の参考と励みになりました。データ管理・共有・公開について、絵にかいたようなフローでは収まらないというのが実情であり、データの品質やセキュリティ保持には一定のコストがかかるということは、具体的な事例がないと見えづらいところです。今回はそれが可視化されたという点でも非常によい企画だったと思います。

企画後記



😊 各分野で「制限」をどのように考えているのか、汎用的な制限公開を見いだせるのかに関心があり、今回、初めて SPARC Japan の企画に携わらせていただきました。どの分野でも、フルオープンにしたいが法的な問題があって今はできない、JAXA の海老沢先生の「制限公開データは、いずれすべて公開データになる」というコメントが印象的でした。

朝岡 誠
(国立情報学研究所)

😊 フルオープンと制限公開を線引きする「基準」だけではなく、そこに至るまでにどのような議論が行われたのか、残された課題は何で、どのようなアプローチが考えられるのかを伺える大変貴重な機会となりました。分野による違いや共通点は、今後の議論や実務に役立つものと思います。企画 WG としては、参加者の皆さんの満足度を上げられるようなオンライン開催の方法を模索していきたいです。

池内 有為
(文教大学)

😞 パネリストからの「事例を積み上げながら練習問題を解いている」というご発言が印象に残っていま

す。税金により研究成果を生み出している国立研究開発法人としては、その成果は何らかの形で公開し社会に還元する必要があります。しかし、研究データの公開に関しては、それをいつ、どのようなタイミングで行うか、研究分野によっても違いやそれぞれの課題があることが今回のセミナーでは明らかになりました。

研究データにまつわる話題を前回に続き企画しましたが、この「練習問題」をどのように解いていったか、引き続き皆様と共有することで課題の解決につながればと考えています。

林 賢紀
(国際農林水産業研究センター)

😞 今回パネルディスカッションのモデレーターを担当しました。データをオープンにすることが当たり前になったからこそ、その「制限」も確かな根拠に基づいて真剣に議論されなければなりません。各パネリストのお話には、いずれも真剣な議論を経た深みがあり、私自身大変勉強になりました。

八塚 茂
(バイオサイエンスデータベースセンター)

■ SPARC Japan セミナー報告

第2回 SPARC Japan セミナー2020



「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

2020年12月18日（金） オンライン開催 参加者（Webex Events）：251名

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。

<https://www.nii.ac.jp/sparc/event/2020/20201218.html>

企画概要



プレプリントによる研究成果の迅速な共有は複数の分野で拡がりつつあり、過去の SPARC Japan セミナーでも数回にわたって取り上げてきた。現在、COVID-19 を契機としてプレプリントの活用はますます加速している。さらに、プレプリントを公開するためのプラットフォームとして、arXiv など特定の分野を中心としたサーバのみならず、オープンピアレビューから出版までのフローを含む F1000 Research、Research Gate などのアカデミック SNS、学術出版社が提供するサービスなどが登場している。

本セミナーでは、多様なプラットフォームによるプレプリント公開の最新動向や目的を共有することによって、プレプリントの方向性を展望する。その際、オープンアクセスリポジトリ連合（COAR）による『Bibliodiversity（書誌多様性）の形成に向けた行動の呼びかけ』で学術情報流通における多様性の障壁として挙げられた4項目を論点としたい。すなわち、「(1) 共通言語としての英語の優位性」、「(2) 基盤とサービスの集中」、「(3) 限定的資金モデル」、「(4) 学術雑誌ベースの評価という偏狭な視点」である。

パネルディスカッションでは、さまざまなプレプリントのあり方を4項目に沿って検討することによって、学術情報流通の多様性をどこまで実現できるのかを議論したい。



参加者から

(大学/学術誌編集関係)

・オンライン開催で地方にいても気軽に参加できました。コロナが終息してもオンラインでも視聴できる環境を整えていただければありがたいです。

(大学/図書館関係)

・プレプリントサーバの位置付けについていろいろな立場からの意見を聞くことができて大変参考になりました。特に坊農先生と引原先生のご講演は、研究活動と学術メディアの関わりという学術コミュニケーションの原点に立ち戻ることができる内容で、大変興味深く拝聴しました。

(大学/その他)

・プレプリントの利活用といっても、それにとどまら

ず、研究評価やSNSでの結びつき、インタラクティブに反応や評価を得ることにまでつながっていることが学べたのがよかった。また、学術コミュニケーションの改革のお話も興味深く拝見しました。

(企業/図書館関係)

・実際に投稿している先生のお話やURA担当の方のお話が聞けたのは貴重な機会で大変ありがたく思いました。

(企業/研究者)

・F1000 Research 販売代理店として広く論文投稿を巡る環境や研究者コミュニティの問題意識の側面を垣間見ることができたように思います。

企画後記



😊 2020年はCOVID-19の世界的な流行を契機としてプレプリントの活用が進み、新たなプラットフォームも登場しました。今回は最新の動向を共有しつつ、書誌多様性の4つの課題、すなわち学術情報流通における言語、プラットフォーム、資金、評価という本質的な問題について、さまざまな立場から語っていただきました。学術の発展にとって何が重要で、どこに問題が残されているのか。この熱い議論を5年後、10年後、30年後に振り返るのも楽しみです。

池内 有為
(文教大学)

😊 初めてSPARC Japanの企画に携わらせていただきました。従来とは異なるオンラインでの開催による難しさもある一方、直接の参加が難しい方々にもオンラインだからこそ参加いただけたと思います。プレプリント公開の最新動向を皆様と共有する機会を得ることができ、私自身も大変多くのことを学ばせていただきました。

安原 通代
(国立情報学研究所)

😊 このセミナーで議論されたプレプリントと書誌多様性の話題は、研究成果発表を巡る最新の議論であり、サイエンスのあり方を巡る議論の最前線だったのではないかと思います。オンライン開催故の難しさや不便さもあったかと思いますが、ご参加いただいた皆様方にとって、このセミナーでの議論がプレプリントの現状とこれから、書誌多様性の観点、そして今後のサイエンスのあり様を考える一つのきっかけになっていれば幸いです。

矢吹 命大
(横浜国立大学)

😊 研究のライフサイクルにおける立ち位置は不変なれど、発信のあり方、活用のされ方は日々多様化の一途を辿るプレプリント。書誌多様性という切り口から議論いただきましたが、聴衆の皆さまにはどう響いたでしょうか。オンライン開催2回目にて、議論の取り回しや質疑応答の受付方法など、企画ワーキングメンバーとしては試行錯誤の回となりましたが、失敗も成功もきっと今後に活きるもの！と割り切って楽しませていただきました。

山形 知実
(北海道大学)

■ SPARC Japan セミナー報告

第3回 SPARC Japan セミナー2020



「初めての研究データ」

2021年2月18日（木） オンライン開催 参加者（Webex Events）：282名

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。

<https://www.nii.ac.jp/sparc/event/2020/20210218.html>

企画概要



大学や研究機関で研究データの管理や公開を検討しているところは多いと思われるが、これら始めるにあたっては検討しなければならない課題がいくつもある。

本セミナーでは、国内外における研究データ管理・公開の実例やベストプラクティス、その評価方法、GakuNinRDM や J-Stage Data などのプラットフォームの活用法について話を伺う。

本セミナーが、研究データの管理や公開に踏み出す一助になれば幸いである。



参加者から

(大学/図書館関係)

・「初めての研究データ」ということで参加いたしました。これまで参加した「研究データ」関係の研修やフォーラムよりは、今回の各ご講演は大変参考になりました。ただ、できましたら「研究データ」の概念や事例についてもご紹介いただきたかったと思います。本学は人文系の大学のため、「研究データ」の例でよく取り上げられる理系・医療系の事例では、なかなかうまく「研究データ」自体がイメージできずにあります。

・短期大学という環境であり、人材や予算確保の関係で、何もしていないと情報取得の機会を逸することもあるので、オンラインでの参加は、とてもありがたい。

(企業/学術誌編集関係)

・海外の機関での取り組みを現場の方から聞けてためになった。

(その他/図書館関係)

・J-Stage Data と GakuNin RDM という今後日本のスタンダードとなりそうな2つのインフラの基本的な仕組みを聞いたことと、北大の機関リポジトリでの実践事例が聞いたことがとても良かったです。初心者に親切な内容だったと思います。ただ、「国内大学・研究機関における RDM 取り組み状況調査」の結果報告も聞けなかったので少し残念でした。

(その他/研究者)

・情報管理・発信者側の視点に加えて、情報を生み出す研究者側の視点の発表があればさらに良い。

(その他/その他)

・データの定義、公開・非公開の各機関の違いなどを知ることができた。

企画後記

😊 研究データ管理・公開のインフラの開発に携わっている者として、聴衆の皆様からいただいたご質問は大変参考になりました。研究データを誰がどのように管理・公開し、どのようにベストプラクティスを集めていくのかいろいろと課題はありますが、まずは走ってみるしかないを実感しました。このようなセミナーを通じて走ってみた結果を共有していけると良いなと思いました。

朝岡 誠

(国立情報学研究所)

😊 研究データ管理やその運用の実例が少しずつ積み上がっています。図書館員としては、これまでの機関リポジトリの運用等が、研究データ管理という新たな領域に生かせるのか不安なところもありますが、今回のご講演では、これまでの経験が生かせることを示唆頂きました。やはり研究者とのコミュニケーションで得られた経験や知見が重要になると感じています。引き続き、この SPARC Japan セミナーが、グッド・バッドなプラクティスを共有する場になれば幸いです。

林 賢紀

(国際農林水産業研究センター)



😊 多くの方にご参加いただき、研究データ管理・公開への関心の高さを改めて感じました。国内での事例はまだあまり多くありませんが、今後更に事例が増え、今回のセミナーのように情報共有できる機会が増えていけばと思います。

安原 通代

(国立情報学研究所)

😊 初めて司会を担当しました。今回のセミナーでは、とても多くのご質問をいただきました。オンラインツールの使用により質問しやすい環境になったおかげでもあります、何よりも皆様が研究データ管理・公開を自分事としてとらえ、真剣に参加してくださったからだと思います。今後も、研究データ管理・公開の実践知を共有する場が SPARC Japan セミナーで引き続き提供されるといいですね。

八塚 茂

(バイオサイエンスデータベースセンター)

本誌についてのお問い合わせ

国立情報学研究所 SPARC 担当

E-mail co_sparc_all@nii.ac.jp FAX 03-4212-2375

<https://www.nii.ac.jp/sparc/>

第1回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

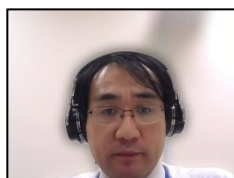
開会挨拶 / 概要説明

朝岡 誠

(国立情報学研究所 / オープンサイエンス基盤研究センター)



朝岡 誠



国立情報学研究所 / オープンサイエンス基盤研究センター特任助教。2020年度SPARC JAPANセミナー企画ワーキングメンバー。東北大学大学院文学研究科修士課程修了、博士課程単位取得満期退学。東京大学社会科学研究所、立教大学社会情報教育研究センターを経て、2019年より現職。現職では人文学・社会科学データインフラストラクチャー構築推進事業に携わり、制限公開によるコンテンツ提供について研究している。

セミナーの企画概要

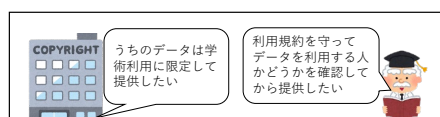
現在、オープンサイエンスの潮流により、研究成果の公開が広く推奨されています。研究データも誰でもアクセスできる形で公開することが望ましいと考えられる一方で、機密保持の観点から公開が望ましくない研究データもあります。医学や社会科学の分野では、調査対象者のプライバシー保護の観点から、研究データにアクセス制限を設け、条件を満たした利用者のみデータを提供することが慣習になっています。このようなデータ提供方法のことを制限公開と呼びます

(図1)。現在、各分野、各機関で研究データ公開ポリシーが整備される中、研究データのオープン・アンド・クローズ戦略の観点から、オープンとクローズの境界にある制限公開に対する関心が高まっています。

しかし、一言で制限公開と言っても、メールアドレスを登録するだけのものから、所属機関長のはんこが必要なものまで、多様な方法があり、データを公開する機関や公開される研究データの属性によって、その方法は千差万別です(図2)。どのような条件を付けるのが相応なのかを、法律の観点、その分野の慣習、

制限公開とは？

- 制限公開：条件を満たした利用者のみデータを提供
 - データ公開が促進される一方で、機密保持等の観点から公開になじまないデータもある
 - 研究データ公開ポリシーが制定されていく中で、制限公開によるデータ提供が増えることが予想される



制限公開の例

(図1)

様々な制限公開

- 制限公開の方法は一つではない
 - メールアドレス登録後、ダウンロード可
 - 所属機関長による許可が必要 など
 - ケースに応じて、どのような条件を設定すればいいのか？
- 制限共有とは？
 - 特定の機関、研究者グループにのみ研究データを提供
 - 制限公開との違いは？使い分ける必要があるのか？



公開したら、問題になる可能性はあるんじゃないの？

制限共有ってどういう方法と何が違うんだろう

(図2)

データ提供者の要望などから判断しなければいけないので、むしろ制限公開というグレーゾーンが増えた分、研究データの公開判断が一層難しくなったと言えます。

また、制限公開とは別に、制限共有という研究データの共有方法もあります。研究データにアクセス制限を設けて、条件を満たした利用者にデータを提供するという点では制限公開と同じですが、利用の対象を特定の機関や研究グループに絞っている点で、制限公開よりもクローズドな方法として知られています。しかし、制限公開との具体的な違いについては明らかになっていません。

本日のセミナーでは、研究データ公開ポリシーを基に、公開・非公開と、その間の制限公開・制限共有を区別している機関様、そしてそのデータを運用している機関様にご講演いただくことで、オープンとクローズドの境界にある制限公開・制限共有はどのようなデータに適用するべきなのかを明らかにしていきたいと考えています。そして、制限公開・制限共有データを運用するには、データ所持者やデータ利用者などのステークホルダーとどのような取り決めが必要であり、どのようなプロセスを経てデータが提供されるのかについて明らかにしていきたいと考えています。

第 1 回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

「研究データの公開・利用条件指定ガイドライン」と 制限公開

池内 有為

(文教大学文学部／

研究データ利活用協議会 研究データライセンス小委員会)

講演要旨



2019年に研究データ利活用協議会(RDUF)研究データライセンス小委員会は、研究者やステークホルダーが研究データのライセンスに関する共通理解を深め、適切な公開方法を決定するためのツールとして、「研究データの公開・利用条件指定ガイドライン」を公開した。ガイドラインは、文献調査、インタビュー調査、質問紙調査、イベントでの公開ディスカッションなどに基づいて策定された。講演では、ガイドラインの概要、およびフルオープンと制限公開を判断するプロセスについて紹介することによって、本日の議論の枠組みを示したい。

池内 有為

文教大学文学部専任講師。博士(図書館情報学)。慶應義塾大学法学部卒業(1995年)、同文学研究科図書館・情報学修士課程修了(1997年)の後、1997年から2005年までフェリス学院大学附属図書館勤務。主婦、筑波大学大学院博士後期課程を経て、2019年より現職。専門領域は研究データ共有とオープンサイエンス。2017年に研究データ利活用協議会にて研究データのライセンス検討プロジェクト小委員会を立ち上げ、2019年から研究データライセンス小委員会としてガイドラインを策定。2020年度SPARC Japanセミナー企画ワーキングメンバー。



本日は研究データ利活用協議会の研究データライセンス小委員会として参りました。小委員会では「研究データの公開・利用条件指定ガイドライン」を作成したのですが、その中で、本日のテーマであるフルオープンと制限公開の境界線に関する課題がいろいろと見えてきましたので、お話しさせていただきます。

研究データ利活用協議会と研究データライセンス小委員会の概要

研究データ利活用協議会(RDUF)は2016年6月に設立されました。研究データを公開して利活用しようとしたときに、ライセンスの問題や、データの引用が

進まない、あるいはリポジトリをどうするかなど、さまざまな問題が出てきます。それらについて、それぞれ関心を持つ産官学の人たちが集まって小委員会を作り、課題解決に向けてボランティアベースの活動を行っています。

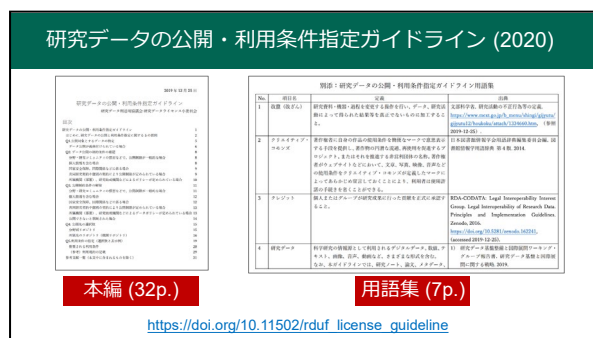
研究データライセンス小委員会は2017年10月から活動しています。研究データを公開する際に、制限公開したいとき、何か条件を付けたいときに、どのようにライセンスを付ければいいのか、あるいは付けられているライセンスを正しく理解するにはどうしたらいいのかということがなかなか難しいので、分かりやすいガイドラインを作ろうという小委員会です。産官学

から多種多様な愉快的なメンバーが集まって活動していますので、ご関心のある方はぜひご参加ください。本日この後ご講演される海老沢先生や上島さんにもメンバーに加わっていただいています。

「研究データの公開・利用条件指定ガイドライン」は2019年12月にできました(図1)。当初は1年半くらいでできるかと思っていましたが、いろいろと難しく、討論に討論を重ねて、最終的には時間をかけて作りました。今年の頭から公開していますので、ご関心のある方は、ぜひご覧ください。シンプルに作ったつもりですが、いろいろ盛り込んで、本編が32ページとなりました。それに加えて7ページの用語集を付けています。そもそもデータの公開とはどういうことか、というところから非常に難しいのですが、こうした「データ公開」などの基本的な用語についても定義を行っています。

ガイドライン作成の経緯と得られた知見

そもそもの前提として、データを公開するなら、理想はOpen by defaultでフルオープンになっていると、使う側としては非常に安心して使えます。しかし、私たちが調べた中でも、分野によって慣習があったり、研究機関や共同研究者の事情でどうしても制限せざるを得ない、機密情報やプライバシー情報、個人情報やDNA情報が含まれているからフルオープンにはできない、特許や商業的利益に関わる、あるいはフルオープンにして誤解を生んだり剽窃されてしまったらどうしようなど、なかなかフルオープンにはできない、制限で公開したいというニーズがあります。



(図1)

では、制限を付けて公開すればいいというように話は簡単に進まず、制限公開する際にもいろいろな問題があることが分かりました。まず、利用条件が複雑です。多くの研究データは著作物として保護されません。公開者としては、自分の望む利用条件で使ってほしいのですが、その契約なり条件付けに非常にコストがかかります。再利用しようとする利用者としても、ライセンスや契約が付いているけれども何をすればいいのか確信が持てない、利用条件や契約内容がよく分からないとなると、やはりデータの公開や利活用が進みません。

では、標準的な、みんなが共通理解を持っているようなライセンスがあれば、それに乗っかっていけば進みそうなのですが、なかなかそうもなっていません。既によく普及していると予想されるクリエイティブ・コモンズ・ライセンスなども、まだ認知度はそれほど高くありません。また、クリエイティブ・コモンズ・ライセンスの場合は、意思表示に使うことはできるのですが、データそのものに著作物性が認められない場合などは、それを侵された場合も法的な保護を受けられない可能性があるという問題があります。

国際的には、2016年にResearch Data Alliance (RDA)とCODATAが「研究データの法的相互運用性：指針と実施のガイドライン」を公開しています。これをそのまま翻訳して公開することも考えましたが、大部であり、また日本の状況にどれだけ当てはまるか検討が必要でした。

日本版ガイドラインの策定

そこで作成したのが日本語版のガイドラインです。データを公開したい側も、データを利用したい側も、すぐにライセンスを理解し、分かりやすくライセンスを付けられるようなものを目指しました。

このガイドラインを作る上でのプロセスは図2のとおりです。まず、基礎文献の収集・翻訳から始めて、インタビュー調査をあちこちの機関に対して行いました。2、3年前になりますが、まさに本日講演をして

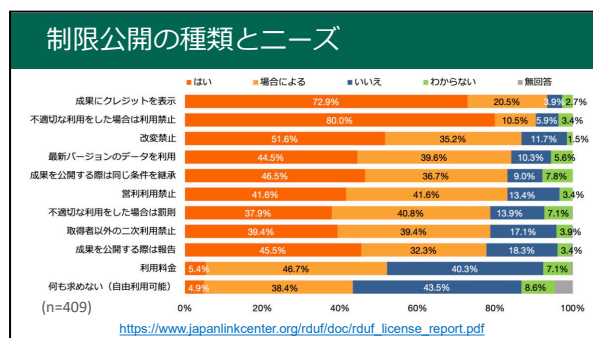
くださっているような機関にメンバーが足を運び、靴底を減らしてお話を伺ってきました。アンケート調査は、研究者の方や、研究の支援をされる方、図書館員の方などに対して行い、その結果をもって Japan Open Science Summit (JOSS) などのイベントでセッションを開催し、ステークホルダーの方や会場の方とも話しながらいろいろな事例を集めました。そしてドラフト版を作り、関係者の方からレビューを頂き、ガイドラインが昨年末に完成しました。このあたりの話は南山委員長が記事で紹介していますので、もしよろしかったらご覧ください。

制限公開のニーズ

図3はアンケートの結果です。データを公開する際に何か条件を求めたいと思われませんかということで、インタビュー調査で聞いてきた項目を挙げてお伺いしたところ、やはり全体的に制限公開したいというニーズがあることが分かりました。「成果にクレジットを表示したい」「不適切な利用をした場合は利用を禁止したい」「改変禁止」といったことを研究者の方は求



(図2)



(図3)

めています。「何も求めない」はやはり少なく、公開に当たっては何らかの制限を求めたいというニーズがあることが分かりました。

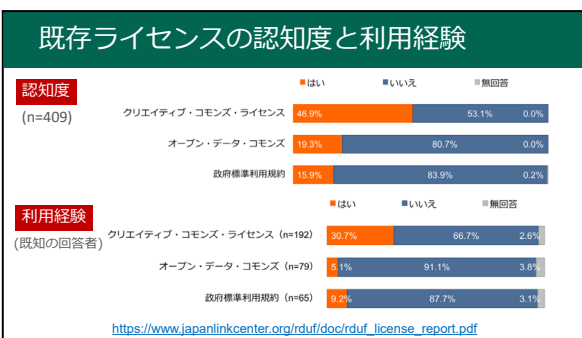
そして、既存ライセンスの認知度についても伺いました(図4)。クリエイティブ・コモンズ・ライセンス、オープン・データ・コモンズ、政府標準利用規約についてそれぞれ聞きましたが、認知度の「はい」の比率はあまり高くなく、ご存じの方であっても、実際に使っている方はごく限られていることが分かりました。

また、求める条件が遵守されるならば、自身のデータを公開してもよいと思われませんかという問いに対しては、「公開してもよいと思う」「やや公開してもよいと思う」は合わせて 88.5%でした。ライセンスあるいは制限公開の条件がきちんと整理され、かつそれがうまく遵守されることが担保されれば、データの公開や利活用が進むのではないかということが示唆されました。

本日の論点：制限公開とフルオープンの境界線

本ガイドラインでは、研究データの公開・利用条件を指定するときのワークフローを用意しています(図5)。まずは対象のデータを選び、公開データと非公開データの区別、条件の解除を一定期間後にするかどうか、公開先、そして利用条件を考え、データの公開に至るというものです。

その中で、聞き取り調査やディスカッションをしていて、やはりここが難しいという話になったのが、まさに本日のテーマである公開と制限公開の部分です。

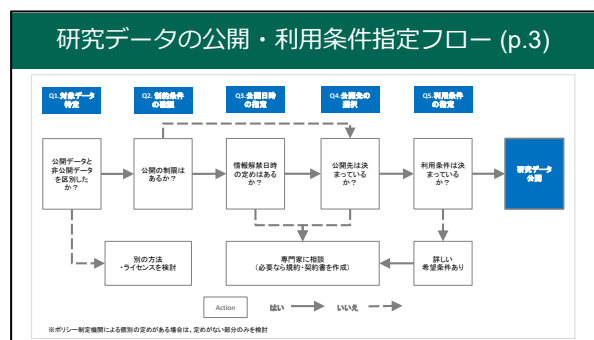


(図4)

公開・非公開をどう線引きするか、その上で、フルオープンにするのか、それとも一定の条件を設けるのか、機関などでのコンセンサスを得ることが非常に難しいというのが、これまでの調査の感触です。では、具体的にどう制限を設けるのかということも、また一つ判断が困難な部分です。一つの機関や個人の研究者が考えていくことは非常に難しいですが、本日のようにいろいろな機関の間で話し合う機会があると、相互に変役立つのではないかと考えます。

さまざまな区分と制限

この話をするときに、四つのレイヤーがあるというところが少しややこしいので、改めてご説明します（図 6）。公開に関しては、フルオープンと制限公開があります。そして制限を付けつつ、条件を満たしたとしても外部の人（機関外、グループ外の第三者）には見せないというのが制限共有です。こういう線引きが割と多くなされているので、そこで線を引き、基本的にはこの四つに分けて考えていただければと思います。



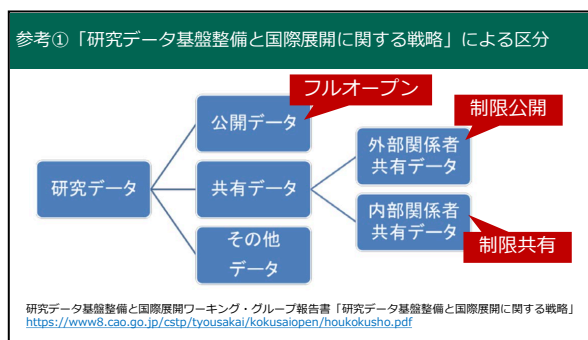
(図 5)

例えば、内閣府の「研究データ基盤整備と国際展開に関する戦略」では、公開データ・共有データ・その他と分けています（図 7）。公開についてはフルオープンです。共有データの中に外部関係者にも共有するものがあり、これは本セッションでいう制限公開に当たります。共有データの中でも内部の者にしか共有しないものが制限共有です。

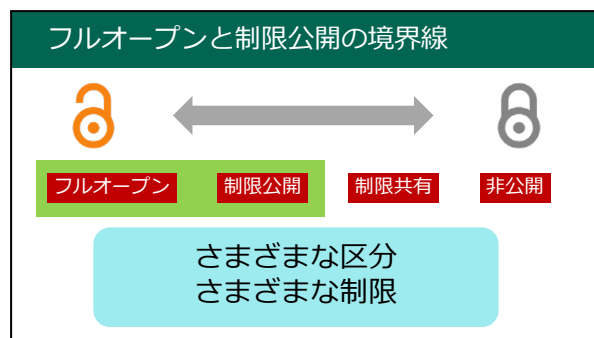
図 8 は国立研究開発法人日本医療研究開発機構（AMED）による区分です。公開と共有は、公的データベースに載せるか、一機関のデータベースに載せるかで線を引いています。このように、いろいろな線の引き方があるのが現状です。

フルオープンと制限公開の課題

まとめとして、フルオープンと制限公開の課題について3点挙げさせていただきます。一つは、フルオープンと制限公開を分けるときに、条約や法律などのハードローによる制約もあれば、分野ごとのソフトロー（規範や慣習、文化）もあります。これが十分に醸成されている分野もあるものの、そうでない分野も非常



(図 7)



(図 6)

参考② AMEDによる区分		
公開レベル	手続き	リポジトリ
公開	不要	公的データベース
制限公開	アクセス申請→承認	
制限共有	アクセス申請→承認	自機関データベース等
非公開	(なし)	

ゲノム医療実現のためのデータシェアリングポリシー (AMED)
<https://www.amed.go.jp/content/000060867.pdf>

(図 8)

に多く、まだ混沌としています。ここの線引きをどうするかが大きな課題です。

二つ目は、制限公開に対するアクセス申請をして、OK であれば公開というときに、その妥当性をどう判断するかです。この線引きの難しさもあります。

三つ目は、制限公開の今後の可能性です。どのような方法があるのかというところまで、本日の議論で踏み込めればと考えています。

第 1 回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

宇宙科学研究所における データ公開・制限公開に関する議論の紹介

海老沢 研

(宇宙航空研究開発機構 宇宙科学研究所)

講演要旨



数年前まで、宇宙科学研究所（ISAS）では、何を「データ」と呼ぶのか、その「データ」を公開してよいのかどうか、はっきりしないような状態が続いていた。2017 年度より所内に「科学データ専門委員会」が設立され、正式なデータポリシーを策定した。そこでは対象となる「データ」を定義するとともに、原則としてすべてのデータは公開とし、非公開とせざるを得ない理由を限定した。また、非公開データについては、担当者と非公開期限を定めることにした。

データポリシーが策定されたおかげで、データに関わる議論は進めやすくなったが、具体的に、各データのどの部分を公開するか非公開とするか、非公開の場合はその期限をどれだけにするか、など判断が難しい場合も多い。それらについて、委員会で議論を進めながら、個別に判断している。

本講演では、そのような議論を紹介するとともに、同じ課題を抱えている他機関との情報交換も期待している。

海老沢 研

1986年京都大学理学部卒業、1991年東京大学大学院理学系研究科修了。宇宙科学研究所にて日本学術振興会特別研究員を経て、1992年よりNASAゴダード宇宙飛行センターにてX線天文学の研究とX線衛星データアーカイブ開発に従事。2001年から2004年までNASAから派遣され、ジュネーブ天文台インテグラル衛星データセンターに勤務。2005年より宇宙科学研究所教授。X線天文学の研究と大学院教育（東京大学天文学教室）に従事しつつ、科学衛星運用・データアーカイブユニット（C-SODA）にて科学データアーカイブDARTSの開発・運用を行う。宇宙科学研究所科学データ専門委員会委員長。



自己紹介

私が所属する宇宙航空研究開発機構（JAXA）の宇宙科学研究所（ISAS）は、1964 年設立の古い研究所です。私は科学衛星運用・データ利用ユニット（C-SODA）に属しており、衛星データ処理、データアーカイブ基盤の開発・運用に携わっています。それから、科学データ専門委員会の委員長も務めています。私自身は X 線天文学を専攻しており、2005 年に JAXA に入所する前はアメリカ航空宇宙局（NASA）と欧州宇宙機関（ESA）で X 線天文学の研究と衛星アーカイブ開発に従事していました。本日は専門の研究ではなく、

業務として行っているデータ整備・データ公開の話をさせていただきます。

ISAS におけるデータ整備・公開の状況

ISAS には、宇宙理工学委員会という諮問委員会があります。これが所長に対していろいろな提言をします。2016 年度に「宇宙科学研究所が保有するデータの取り扱いに関する提言」を提出しました（図 1）。それ以前からいろいろな活動をしていたのですが、このあたりから物事が整理されていったので、そこから話を始めたいと思います。その資料は宇宙理工学メン

バーに限定公開していますが、そのエッセンスを紹介しします。これらの提言を実行に移す形で、ISAS を中心として宇宙科学データの整備・公開が進められています。

まず、データの公開・非公開についてです。データの種類によらず汎用かつ有用なデータは公開し、長期的に保管すること。非公開データは保管期限を定め、期限後に扱いを見直すこと。データに関するさまざまな事項を審議し、所に進言するための有識者委員会を設けること。また、無条件に商業利用や宣伝利用しても良い科学データを置く公開領域を設けてはどうかといった提言をしています。この提言に従って、2017年に ISAS 内に科学データ専門委員会が作られ、私が今まで委員長を務めてきました。

「宇宙科学研究所のデータポリシー」の紹介

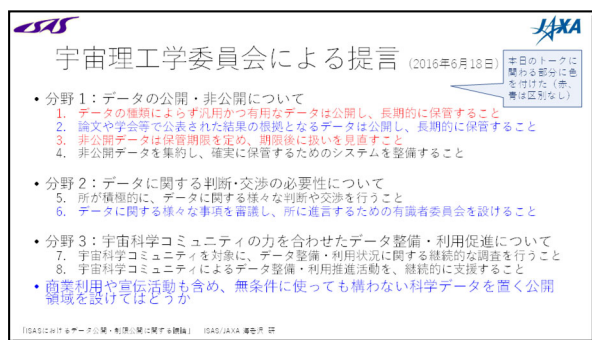
この科学データ専門委員会が、「宇宙科学研究所のデータポリシー」を策定しました(図2)。諮問委員会が所に対して行った提言に留まらず、所としてフォーマルなデータポリシーを作ろうということで、議論

を重ねました。日本語・英語で ISAS のホームページより公開しています。ただ、ホームページで公開するだけでは正式な文書になりませんので、文書番号を付けて JAXA 内で公式な社内文書にしています。このように、フォーマルなデータポリシーを定めました。そのうち、本日の話に関係する部分を紹介させていただきます。

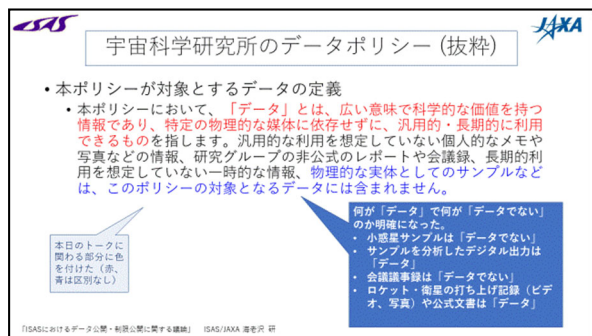
まず、データについて定義しています。『データ』とは、広い意味で科学的な価値を持つ情報であり、特定の物理的な媒体に依存せず、汎用的・長期的に利用できるものを指します、「物理的な実体としてのサンプルなどは、このポリシーの対象となるデータには含まれていません」としています。例えば私たちの研究所で、「はやぶさ」がイトカワのサンプルを採ってきました。そのサンプルはここではデータとはいいません。しかし、そのサンプルを分析したデジタル出力はデータです。また、いろいろな会議の議事録も保管されているのですが、それはデータとは呼びません。その一方で、ロケットや衛星の打ち上げ記録などの公式文書はデータということになります。

このように、まずは、何がデータで何がデータでないかをはっきりさせました。以前はそれがはっきりしていなかったもので、そもそも何がデータで何がデータではないかという議論もできませんでしたが、それがはっきりしてきたということです。

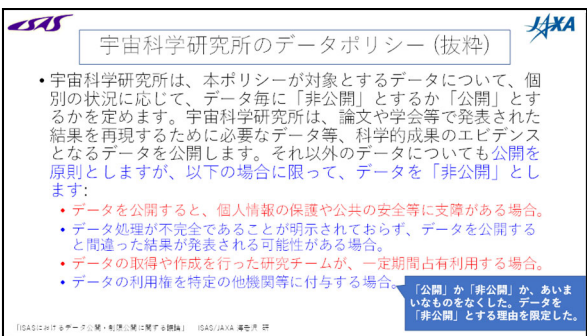
データポリシーとしては、まず公開を原則とします(図3)。しかし、次の場合に限ってデータを非公開とします。データを公開すると個人情報の保護や公共の安全等に支障が出る場合。データ処理が不完全である場合。データの公開すると間違った結果が発表される可能性がある場合。データの取得や作成を行った研究チームが、一定期間占有利用する場合。データの利用権を特定の他機関等に付与する場合。



(図1)



(図2)



(図3)

ることが明示されておらず、データを公開すると間違った結果が発表される可能性がある場合。データの取得や作成を行った研究チームが、一定期間、占有利用する場合。データの利用権を特定の他機関等に付与する場合です。このように、公開か非公開かが曖昧なものをなくし、データを非公開とする理由を限定しました。

非公開データについては、利用範囲と非公開期限を定め、その後、公開に移行するか、非公開を続けるか、あるいは廃棄するかを判断します。これは非常に大事で、後々、いろいろな場面で効いてきます（図4）。

非公開といっても、誰も使わない非公開はあり得ません。また、永久に非公開のデータも原理的にあり得ませんので、どこかで廃棄する、あるいは公開に移行することになります。非公開を長く続けるという手もありますが、極端な話、100年間非公開というのは考えにくいのです。私は天文学者ですが、新しい星が発見されたとき、100年前の写真を見て、100年前の空に、実はその星が受かっていた、ということはよくあることです。ですから、100年後に宇宙科学研究所のデータを使うことは十分あると考えています。

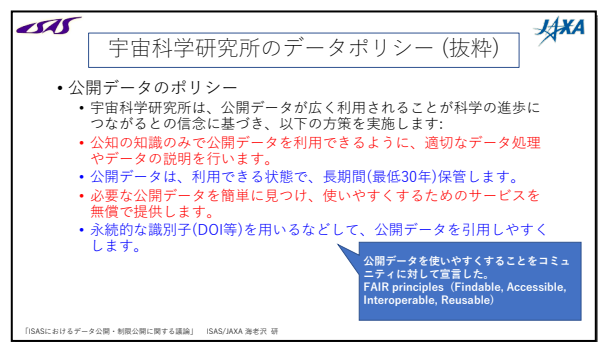
公開データのポリシーは図5のようになっています。これは後で気が付いたのですが、世界的に公開データを Findable, Accessible, Interoperable, Reusable にしようと言う FAIR principles というものがあるそうです。私たちが自分たちの頭で考えたものが、こういうものと合致しています。

公開データ利用の際のルールは図6のとおりです。公開データは、原則として、営利目的、非営利目的を

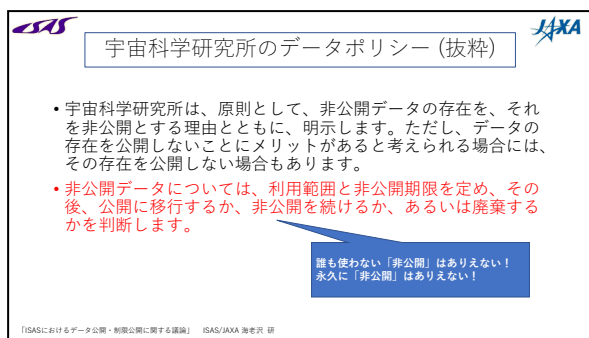
問わず、複製、送信、加工も含め無償で利用できる。また、必ず出典を明示すること。宇宙科学研究所は、利用者がデータを用いて行う一切の行為について何ら責任を負うものではない。このように非常に寛大なルールで、政府標準利用規約（第2.0版）や CC BY 4.0 など実績のある外部のルールを引用しています。公開データのルールを定めることにより、データ利用が進み、安心してデータを公開することができます。私たちは ISAS のデータポリシーに則っている、このデータポリシーは JAXA の正式文書としてフォーマルなものであるということで、物事が着々と進んでいます。

公開データ・データ利用の例

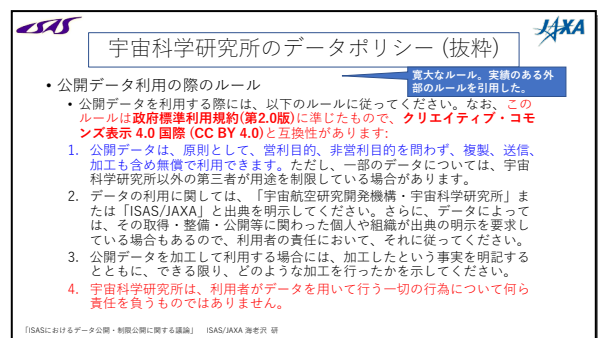
ISAS のデータポリシーに沿って、着実にデータ公開が進んでいます。例えば「かぐや（SELENE）」は月を周回する衛星です。「かぐや」については、NHK が作ってきれいに整備されたハイビジョンカメラのデータがあったのですが、それが公開か非公開かが長い間、曖昧な状態でした。それも、データポリシーに従って、公開するか、非公開だったら何年間非公開かと



(図5)



(図4)



(図6)

いうことを決めなければいけません。そこで所長の命で、JAXA の広報部の方なども含めて NHK と交渉し、公開することになりました。公開か非公開か曖昧なものを作ってはいけない、というポリシーが効いた例です。

また、「あかり」衛星の場合には、未処理の生データがたくさんありました。このデータは貴重なものなので公開しよう、そのためにはデータポリシーに従って使える状態にして公開しなければいけないということで、予算を付けて、高次処理をして公開しました。

それから、データへの DOI 付与も進んでいます。

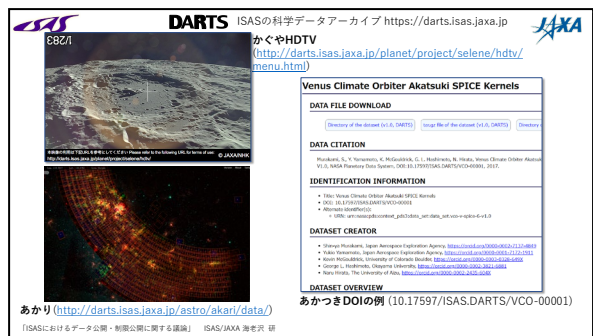
現在稼働中の「あらせ」衛星のデータ処理は名古屋大学が主に行っていますが、過去の ISAS の衛星データがいろいろな大学に散らばっています。これについてもデータ整備・公開を進めなければいけないということで、ISAS が予算を付けて支援しています。このようにしてデータ整備・公開が着々と進んできています。

いくつか例を示します(図7)。「かぐや」のハイビジョンのデータです。非常に美しい映像で、JAXA と NHK のロゴが入っています。公開できなかったら本当にもったいないことでしたが、きちんと公開できています。「あかり」のデータも素晴らしいもので、きちんと整備して公開しました。それから、DOI が付いた例としては金星を回る「あかつき」のデータがあります。DOI をたたくと「あかつき」のデータが出てきます。これらのデータもよく利用されています。

また、データの商業利用も進んでいます(図8)。これはグリー株式会社と一緒にやった仕事です。やは

りプロのプログラマーはすごいと思ったのは、データの使い方を知らせると、すぐに VR のコンテンツにするのです。宇宙を題材にした子ども向け VR 体感サイエンスツアー「ありえな LAB」というものです。もちろんデータ利用にお金は取りませんが、必ず私たち JAXA、ISAS の DARTS (データベース) に acknowledge していただき、CC BY 4.0 の規定に従って使用していただいています。

一般利用の例としては、最近気付いたのですが、YouTube にきれいな「あかつき」のデータが加工されてアップされています(図9)。かなりよくできるアマチュアの方ようで、「DARTS は JAXA のいろいろな探査機の資料を見ることのできる宝箱でした」と書いていただいて、とても嬉しかったです。一般の方も ISAS のデータを使って、何か科学的発見をされるかもしれません。公開して使ってもらわなければ何が出てくるか分からないということで、データを公開しています。



(図7)



(図8)



(図9)

「制限公開」に関わる議論の紹介

ここからが本題になるかと思いますが。制限公開に関わる議論は非常に難しく、私たちも悩んでいることがありますので、本日の研究会でいろいろな専門家の意見を聞きたいと思っています。

まず私たちのデータポリシーですが、公共の安全等に支障がある情報は非公開にします（図 10）。例えば周波数情報は非公開です。人工衛星と地上で通信しますが、その周波数で何かを送って人工衛星を止めるなど、原理的にはできるかもしれません。それから、人工衛星と地上との間の通信を邪魔することができるかもしれません。過去の衛星なら問題ないだろうと思っても、同じ周波数を現在の衛星にも利用していることがあるそうです。工学の先生方の話を聞くと、やはり周波数は出したくない、出す必要もないということです。

JAXA リポジトリが国立情報学研究所（NII）にあります。ISAS 報告で、M-V 型ロケットについての科学論文があります。これは公開されていますが、「〇〇のレーダ・トランスポンダ」と白塗りになっています。隠されている部分は周波数です。周波数情報は非公開という具体的な例で、その判断にはあまり困っていません。

次が、困っている例です（図 11）。テレメトリデータと工学データです。1970 年代からのテレメトリデータと、それを分かりやすくした、探査機の位置・姿勢制御、温度、電圧などの工学データが保管されています。これらをどうするべきか。さすがに直ちに破棄はないでしょう。しかし、先ほどのロジックで永久非

公開もないでしょう。ただ、公開すると、公共の安全や研究チームの一定期間占有利用に抵触しないか。

「はやぶさ」で有名な先生にも意見を伺ったのですが、そういうものは公開しなくていいのではないかという意見でした。

「はやぶさ」の地球再突入直前には、機微なオペレーションがあったそうです。そこでどんなことをしたかを人に知らせると、衛星や探査機メーカーのノウハウも知られてしまいます。そこで非公開にしようとなったのですが、永久非公開はあり得ない。では何年待って公開するか。これはまだ決まっています。私たちの委員会でも散々議論しているのですが、困っているところです。

もう一つ困っているのが、「宇宙科学資料データベース」です（図 12）。ISAS とその前身の研究所の 1950 年代からの数十万枚の画像、映像が保管されています。かつては ISAS に映像記録班が存在していました。昔は国家公務員として写真を撮ったり、映像を撮ったりする専門的な方がいて、プロフェッショナルなクオリティで、ほぼ全てのロケット・衛星開発記録

テレメトリデータ・工学データ

- 1970年代からの科学衛星の「テレメトリデータ」（生データ）が保管されている
- テレメトリデータを解読した、探査機の位置・姿勢制御、温度、電圧などの「工学データ」が存在する。
- 終了した衛星のテレメトリデータ・工学データをどうするか？
 - 破棄はないだろう
 - 永久非公開はありえない
 - 公開すると「公共の安全」や「研究チームの一定期間占有利用」に抵触しないか？
 - たとえば、「はやぶさ」の地球再突入直前のオペレーションがわかってしまう
 - 衛星や探査機メーカーのノウハウも知られてしまう
 - 一定期間は非公開としたら、その期間は？（10年～30年？）

「ISASにおけるデータ公開・制限公開に関する議論」 ISAS/JAXA 海老沢 研

(図 11)

「公共の安全等に支障がある」情報

- 「周波数情報」は非公開
- <http://id.nii.ac.jp/1696/00033361/> など
- 周波数情報を白塗りにして公開

「ISASにおけるデータ公開・制限公開に関する議論」 ISAS/JAXA 海老沢 研

(図 10)

「宇宙科学資料データベース」

- ISASとその前身の研究所の1950年代からの数十万枚の画像、映像が保管されている
- かつてISASに「映像記録班」が存在した
 - プロフェッショナルなクオリティ
 - ほぼすべてのロケット・衛星開発記録がデジタル化されている
- 現在、JAXA内でアカウントを申請して閲覧・利用できる
 - Dspaceというシステムに載せていて、検索も可能

「ISASにおけるデータ公開・制限公開に関する議論」 ISAS/JAXA 海老沢 研

(図 12)

を取っていました。もちろん昔は写真、フィルムでしたが、膨大なお金をかけてこれを全てデジタル化しました。現在、宇宙科学資料データベースは JAXA 内でアカウントを申請して閲覧・利用できます。DSpace というシステムに載せているので検索も可能です。これは一部公開していますが、まだ全部は公開していません。

他にも非公開資料がたくさんあります。例えば 1950 年代のカップロケットや、日本初の人工衛星「おおすみ」打ち上げ後の街の様子など、写真が何十万枚もあるわけです。これらの画像や映像の公開には、著作権、肖像権、公共の安全、メーカーのノウハウなどの観点から問題がないか、今、議論しているところです。

私たちの委員会では、宇宙科学資料データベースを 30 年後には公開することを決めました。世の中には 30 年ルールというものがあるそうで、いろいろな公文書が 30 年で公開されています。ただ、その多くはまだ公開できていないというのが現状です。担当の方によると、事なかれ主義という言葉は悪いですが、絶対に問題を起こしたくないという意識が高いように思われます。一方で、数十年にわたり、多額の税金をかけて収集・整備した資料を公開しないことは許されるのかというと、やはり公開しなければいけないのだと思います。責任がある誰かに、びしっと「公開すべきだ」と言って貰えれば良いのですが。

今こういう問題があり、困っているということをご紹介しました。ぜひ多くの皆様のご意見を頂ければと思います。

第 1 回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

データ流通市場におけるデータ公開の事例・論点

上島 邦彦

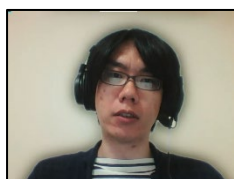
(株式会社日本データ取引所)

講演要旨



民間企業のデータ公開にはしばしば多くの障壁が生じます。その背景には法的課題や技術的問題だけでなく、ブランド毀損やビジネス機会の損失リスク、IT リテラシー格差といった諸問題もつきものです。利用規約の作成、業務フローの標準化、提供内容・方式の棲み分け、UI/UX デザインによる事故防止など、さまざまな工夫が行われています。

本報告では、スタートアップ企業の視点から、日本企業のデータ取引によくある論点と、データ流通市場サービスを介してデータを限定公開する事例を紹介します。



上島 邦彦

株式会社日本データ取引所に創業メンバーとして参加。事業企画・プロダクト開発全般に従事。企業間のデータ調達・販売支援を始めとして、データ製品開発や社内データ戦略企画などにも携わる。研究データライセンス小委員会メンバー、データ流通推進協議会技術基準検討委員会副主査。

日本データ取引所 (J-dex) 会社概要

弊社はその名のとおり、データの取引を専門に行う会社として創業し、今年で5年目に入ります。企業と企業の間を取り持つことが多いです。データ活用戦略の立案やデータ調達の仲介——本日の論点でもある、どこまで公開・共有できるかも含めた企画・設計に携わっています。

利用条件やプライバシーポリシー、セキュリティポリシーの整備など、組織間でデータを流通させるときの法的相互運用性の向上についても、弁護士と連携しながら実務支援をさせていただいています。

研究データ利活用協議会のライセンス小委員会にも参加している他、データカタログやデータ品質に関するガイドラインの策定、国際標準化活動にも、微力ながら貢献しています。

本日は、企業の一従業員の目線でお話をさせていただけます。企業がデータを公開・共有するときの肌感覚——ソフトローがどのようなものであるかを実感していただければと思います。キーワードは、「鍵のかかった部屋」をいかに持ち運ぶかです。

データとは何か

データとは、ざっくりと言うと記録された事実であり、価値ある情報の素材です (図 1)。日本工業規格が 1994 年に、情報とは何か、データとは何かを定義していて、これがよく参照されます。情報がデータとして存在を認められ、利用可能となるまでに、複数の工程が必要であることは皆さんも実感的に理解されていることでしょう。多くの工程を経てデータが流通しています (図 2)。ビジネスや学問、行政を支えるデ

ータが、世界中で動き回っています。他の製品やコンテンツと同じで、何かしら対象物に関する記録を行い、それを蓄積、転送して統合する、いわゆる原料生産に当たるプロセスがあり、外に出してもいい製品に加工するプロセスを経て、使う側が評価したり、それを基に判断・行動するという一連の流れを回す。これを広く「データ流通」といっています。

「データ製品」という用語もあります。先ほど申し上げた、概念としてのデータではなく、外に出せる、売ってもいい、他の方にお使いいただける、生の情報を加工したデジタルコンテンツの総称です。その製品情報を要約・整理したものを、「データ製品カタログ」と呼んでいます。この言葉が指す対象は、学術研究におけるデータ出版の対象データや、研究データリポジトリとほぼ同じかと思います。

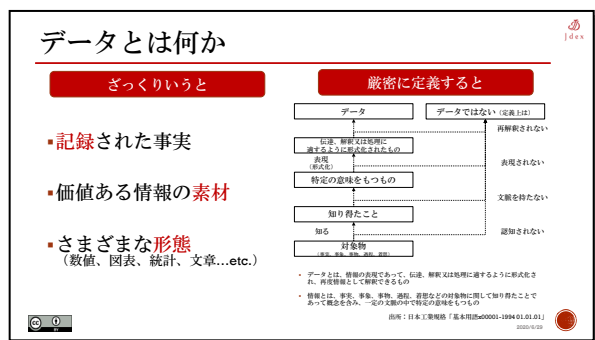
データ取引についても簡単に説明します。ごく抽象的に考えると、誰かが持つデータを他の相手に見せたり渡したりすることは、広く「データ取引」だと言えます。取引交渉によってお金とデータを交換するだけでなく、お互いにデータを開示したり、共同研究のよ

うに、データ共有をして共同で論文を書くことも含みます。オープンサイエンスのために世の中にデータを公開することも、対価のない取引だと考えられます。

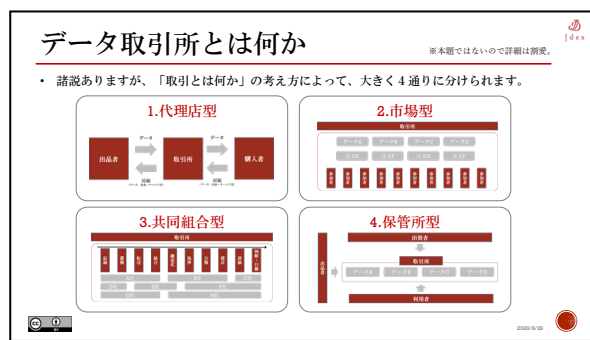
弊社名にもある「データ取引所」には、いろいろなパターンがあります（図3）。単純に仲介する「代理店型」。注文と商品をマッチングする「市場型」。データ流通プロセスの一個一個を皆さんで担う「共同組合理型」——これは魚の市場に近くて、同じ場所に皆で集まってやる方が楽だし早いというもの。そして、研究データリポジトリはこちらに近いと思うのですが、取引所が出品者からデータをお預かりし、そこに第三者が出資し、また第三者の方が利用する「保管所型」。これらをひっくるめて「データ取引所」と呼んでいます。

データ公開の方法整理

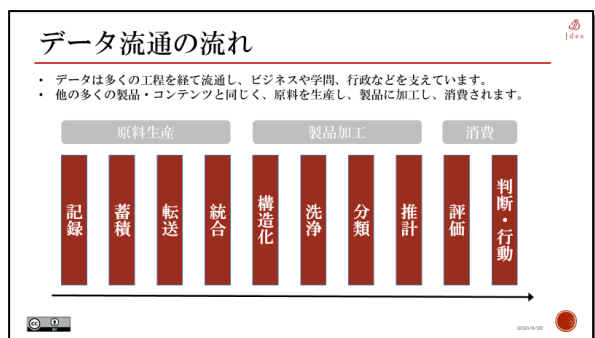
データの公開・共有にグラデーションがあることは、企業のデータ取扱担当者も日々感じながら仕事をされています。例えば、図4の左下——企業のデータ取扱担当者が何かデータを持っているとして、それを誰に



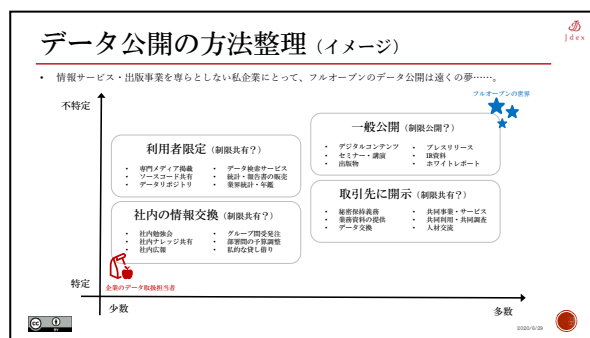
(図1)



(図3)



(図2)



(図4)

も公開しないことはよほどあり得ず、少なくとも社内の誰かとは共有するでしょう。そこから出発して、右上にあるように、条件指定や利用制約なしで、フルオープンで公開するという大きな目標に向かっていく。

2つの軸で考えてみます。横軸は「どれくらい多くの方にご覧いただくか」（多数・少数）。縦軸は「どれくらい顔の見える方にご覧いただくか」（特定・不特定）。この間にグラデーションがあり、どこまで多数の人に、どこまで不特定の人に情報を出すかによって、情報を出す形式・方法が変わると考えてみます。

例えば、社内の情報交換のために制限共有する（左下）。利用者を限定して、ある程度専門知識や業界固有の事情が分かる方に公開する（左上）。あるいは取引先と契約を結んだり、共同の事業・サービス・プロジェクトを組んで開示する（右下）。私ども（民間企業）が「一般公開」と呼ぶものは、実質的には「制限公開」ですが、デジタルコンテンツとして世に問うたり、セミナー・講演でグラフを見せたり、出版物を出したりするときには、著作権をはじめ各種の制約が生じます。もろもろの法的観点の整理なども面倒になってきます。

企業のデータ公開にありがちな障壁

企業のデータ公開に当たって、ありがちな障壁が5つあります（図5-6）。大きいのはやはり法的課題です。一番大きい国際法のレベルから、国内の法令レベル、企業間での契約レベル、そしてここが今一番やっかいなところかと思いますが、実務レベル。それぞれで問題が生じます。データ作成者が——研究データでいう

と研究者が——これら全ての知識を持つことは、かなり難しいのではないかと感じています。

技術的問題もあります。単純に言えば、知識です。データに関する知識と業界に関する知識の両方が分かる人は、昔も今も将来も少数派でい続けるでしょう。また、データ製品開発のためにいろいろな技術が考案され、提供されていますが、慎重を期すれば期するほど手間や労力がかかります。「そこまでして外に出したいデータなのか」と、皆さん悩みながら作業されています。それから、先週から今週にかけて各地で事故が起きたとも報じられていますが、安全にデータを移転するためのセキュリティ、プライバシー、証跡管理などもよく問題になります。これらはハードな制約です。

ソフトな制約もあります。よく聞くのは、企業・部署としてのブランドが毀損されないかといった何となくの不安感や、ステークホルダーの意向や考えの板挟みになってなかなか身動きが取れないということです。

ビジネス機会を逸するのではないかという不安も、現場の方や経営者の方が感じることは多いようです。そもそも「データを持っています」と公に言わないこと自体が競争力になりますし、データの公開・共有が自社のブランディングにつながる、そういうことが好きな人材が来てくれる、その作業自体で知識習得できると分かっているにしても、投資対効果といえますか、どれくらい労力をかけていいのかよく分からないという声はよく聞きます。個人の感覚ですが、家に本がどんどん積み上がって減っていかないことに似ていますね。ITリテラシー格差などもよく問題視されます。

企業のデータ公開にありがちな障壁 (1/2)

- ・ **法的課題**
 - ・ 国際レベル：通商協定、越境データ移転、ローカリゼーション規制
 - ・ 法令レベル：営業秘密、知的財産権、個人情報、パブリシティ権
 - ・ 企業レベル：秘密保持契約、データ利用規約、情報管理指針、社内決裁
 - ・ 実務レベル：利用条件の整理、契約書の起草、ライセンスの設定
- ・ **技術的問題**
 - ・ 統計調査・データ分析と専門分野の双方に精通した人材は少数派（昔も今も将来も）
 - ・ 「欲しいデータ」「外に出していいデータ」の判断も意外にむずかしい
 - ・ データ製品開発には多様な技術が介在する（慎重を期すほど高コスト体質に）
 - ・ デバイス性能理解、データ品質評価、データ洗浄手順、匿名化・曖昧化
 - ・ 安全なデータ移転のためのセキュリティ確保、プライバシー保護、証跡管理
 - ・ 業務フロー自体に根ざすリスクを減らすには、働き方から変える必要がある

(図 5)

企業のデータ公開にありがちな障壁 (2/2)

- ・ **ブランド毀損**
 - ・ 業界慣習、社内外の視線、敏感な専門家、無関心な一般生活者の板挟み
 - ・ 売上リスクに当意即妙で応じられる広報スキルと法務リソースを持つ担当者は少数
- ・ **ビジネス機会の損失リスク**
 - ・ 「データの保持」を促すこと自体が競争力になる
 - ・ サンプルデータや項目定義だけでも技術流出リスクが懸念
 - ・ サプライチェーンの上流・下流で、交渉材料として使われる情報開示
 - ・ 得られる利益（ブランディング、人材獲得、知識習得）の不確実性
 - ・ 経営層・現場ともに「すべきた」と分かっているにもかかわらず、設備・人員の投資判断を行いつらい
 - ・ 自宅の競合品がどういっわけか一向に減らないことに似ている
- ・ **ITリテラシー格差**
 - ・ 組織内のデータ活用スキルのばらつき、データ活用部門の大いなる忙殺
 - ・ 取引相手との知識ギャップ（スーツとギークの、理論と応用の、大企業とベンチャーの...）

(図 6)

企業のデータ≡「鍵のかかった部屋」

つまり、企業のデータは「鍵のかかった部屋」のようなもので、中に何かがあり、価値もあるのだろうけれども、その価値を知るには鍵を開ける必要があって、なかなか入りづらく、中にいる人、外にいる人、双方にとって苦しい状況に陥りがちです。

そんな中で、誰もが安心して自由にデータを使える——先ほどの図の一番右上までたどり着くにはどうすればいいか。一般論でいうと、「鍵のかかった部屋」のドアをノックして中に招き入れてもらい、その部屋の住人になればいい。鍵を手に入れて、長居して、ゆくゆくはゴールを目指すといったことを考える。

少し常識から外れて、概念的に考えると、そもそも鍵があるからいけないのだ、出入り自由にすればいい。部屋そのものを解体して、壁をなくせばいいといった、空間デザインのような発想もできるでしょう。そして、私が最近考えていて、本日まで説明したいのは、「部屋ごと持ち運ぶ」こともできるのではないかな。

「鍵のかかった部屋」をどう扱うか。企業データの取扱法になぞらえると、20 世紀的方法として、組織のメンバーシップを入手する、データのオーナーシップを持つ、データの利用権限を付与してもらうなどがあります。これらはデータポリシーの整備など、制度・法務面の対応でクリアできます。加えて昨今増えているのは、コミュニティを運営して、そこで公開・共有していくことです。いわば 21 世紀的方法です。一定のライセンスを設定して、より多くの方が使えるようにする。さらには、弊社がまさにそうですが、機能的なサポートを受ける、マーケットプレイスに出品するといった挑戦も出てきています。大きく 2 つ、説明します。

まず、データの権限を付与してもらう事例です。先ほど申し上げた 4 つの課題に対して、リーガル面だけではなく、サービス面、システム面、コンテンツ面でも工夫することで、総体としてリスク・コストを下げる手法が取られます。利用規約を作成したり、業務フローを標準化して技術的な問題が起こりにくくしたり、

システム側で事故が起こりにくくしたり、出せるものと出せないものを切り分けたりします。リテラシーに関しては、もう頑張るしかないという感じですが。

弊社が関わった事例では、しょうゆや豆乳のメーカーとして知られるキッコーマンと、PR 会社のオズマピーアール、東京大学大澤幸生研究室の 4 者による共同研究があります。約 20 社間でデータ流通の個別の利用許諾を得て、データ活用の枠組みを作りました。これは、「やればできるけれども大変だ」という事例です。

2 つ目は、データ製品の出品登録・購入リクエストの事例です。（利用許諾の手続きを）システム化してしまうことです。私どもの挑戦として、データの出品と購入ができる会員制のマーケットプレイス「JDEX™」を 11 月ごろから始めようとしています（図 7）。データの制限公開と制限共有を線引きするための微妙な取り決めなど、細かいところをサポートする機能を備えて、実際にお使いいただこうと思っています。データのカatalog を作れるようにして、データ自体を出品しなくていいようにしたり、審査制にしたり、特定の相手だけに注文・出品できるようにしたり。匿名加工やライセンス整理など、実務面のサポートもシステム側で行おうとしています。使い慣れれば簡単にできると期待していますが、もちろん日本ではまだ普及していません。「まだまだこれから」の事例です。

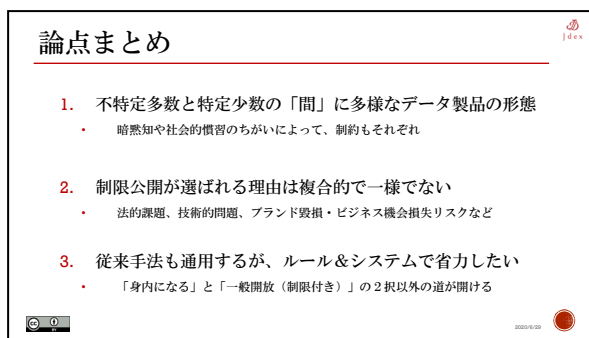


(図 7)

論点のまとめ：「鍵のかかった部屋」をいかに 持ち運ぶか

企業のデータは、なかなか鍵を開けて外に持ち出せない。その状況で、「鍵のかかった部屋」をどう持ち運ぶか。境界線をずらす、動かすという発想をした方が、物事が進みやすいのではないかと思います。

まとめると大きく 3 つの論点があります（図 8）。まず、不特定多数と特定少数の間に多様なデータ公開の仕方があり、ソフトローによって制約もそれぞれです。また、制限公開あるいは制限共有が選ばれる理由は、複合的で一様ではなく、トータルでリスクとコストを下げる必要があります。1 つの解決策として、実務的に頑張って労力を割く従来の方法も有効ですが、今後は、ルールやシステムで省力できるといいのではないかと考えています。



(図 8)

第 1 回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

ヒトデータの共有のための取り組み

三橋 信孝

(科学技術振興機構 バイオサイエンスデータベースセンター)

講演要旨



2013年に運用を開始したNBDCヒトデータベースは、生命科学や医療の分野で生み出されたヒトに関する様々なデータの共有を推進しています。個人単位で決定されたゲノム配列が主な格納データです。ヒトデータの共有では、データの使いやすさと研究参加者のプライバシー保護のバランスが非常に重要です。

NBDCヒトデータベースでは、データ提供者は、研究参加者との間のインフォームドコンセント、データのプライバシーレベル、研究の進捗状況などによって、適切な共有方法を選択することができます。個人情報の保護や倫理面にも配慮しながら、データの利用を促進するためのこれまでの取り組みを紹介したいと思います。

三橋 信孝

国立研究開発法人 科学技術振興機構（JST）バイオサイエンスデータベースセンター（NBDC）研究員。東京大学大学院理学系研究科修士課程修了後、民間企業にてライフサイエンス分野のソフトウェア開発に従事。ライフサイエンスデータベースセンター（DBCLS）研究員を経て、2011年より現職。NBDCヒトデータベースの情報セキュリティポリシー策定・情報システムを担当。NBDC、DBCLSの研究者と共同で、日本人ゲノム変異の統合データベース「TogoVar」の開発・運用に取り組んでいる。



機関紹介

バイオサイエンスデータベースセンター（NBDC）は科学技術振興機構（JST）の一組織です。ライフサイエンス研究の成果が最大限活用されるように、オープンサイエンスという方針の下で、ライフサイエンス分野のデータを統合して利活用を促進するセンターとして2011年に設立されました。私どもは以下の三つのテーマで活動を行っています。一つは「ひろく」ということで、データを共有するためのルールやガイドラインを作成してデータを集めることです。二つ目は「つなげて」で、RDFというフォーマットで集めたデータの統合を行っています。その他、辞書を作ったり、さまざまな統合をするための技術開発を行ったりしています。三つ目は「つかう」ということで、統合

したデータを使いやすくするためのアプリの開発、データ駆動型研究に向けての研究などを行っています。

NBDCのWebサイトでは、統合データベースとそれを使うためのツールが誰でも無料で利用できるようになっています（図1）。共同研究先のライフサイエンス統合データベースやファンディングプロジェクトで作られたサービス、データベースなども利用できます。TogoTVという動画による解説がありますので、ご覧いただければと思います。

NBDCのサービスとヒトデータベース

サービスの一覧のところに、今日紹介させていただくNBDCヒトデータベースやバリエーションのデータベースがあります（図2）。なぜヒト由来のデータを集

めて共有する必要があるのか。人間のゲノムにはごくわずかな個人差があります。このゲノムの個人差が、病気のかかりやすさ、薬の効きやすさ、治療率などの個人差につながっていきます。このようなゲノムの個人差を調べることができるようになったのは、ヒトのゲノムの塩基配列、DNA の配列を決定するコストが非常に下がったからです。今、1 人当たり約数百ドルで配列決定ができるといわれています。

こうしてゲノムの個人差と疾患の関係等が分かったと、ゲノムの違いによって治療方針を決めることができる個別化医療（プレジジョンメディスン）が実現するといわれています。そのためには、より多くの個人の医療情報とセットになったゲノム情報を蓄積したデータベースの構築が必要になってきます。

欧米では、米国国立衛生研究所（NIH）の The database of Genotypes and Phenotypes (dbGaP) や、ヨー

ロッパの European Genome-phenome Archive (EGA) という公共データベースが構築されました。そのため、ヨーロッパのデータが圧倒的に多いです。ゲノムにももちろん人種差がありますので、日本人のデータベースを作ることは非常に重要です。日本でも疾患に関する知見を得るためのコホート研究を、バイオバンク・ジャパンや東北メディカル・メガバンクが数千人から十万人を集めて行っています。そういった研究から出たゲノムのデータや医療データなどを共有するための受け皿として、ヒトデータベースが作られました。

NBDC のヒトデータベースは、国立遺伝学研究所の DDBJ センターが運営する Japanese Genotype-phenotype Archive (JGA) と共同運営しています。NBDC が、制限公開をするためのガイドラインの作成や、制限公開時にはデータ提供や利用するための審査を実施しています。一方、DDBJ は、データの管理やユーザーへの払い出しなどの業務を行っています。

公開からちょうど 7 年たちますが、300 件程度の申請があり、追加更新なども行われています。約 32 万人のデータが格納されていて、そのうち約 150 件が公開されており、約 140 件の利用申請があります。

公開、制限公開、非公開の区別

ここからは、セミナー概要で問い掛けられていた四つのテーマに沿ってヒトデータベースの内容を紹介させていただきます。

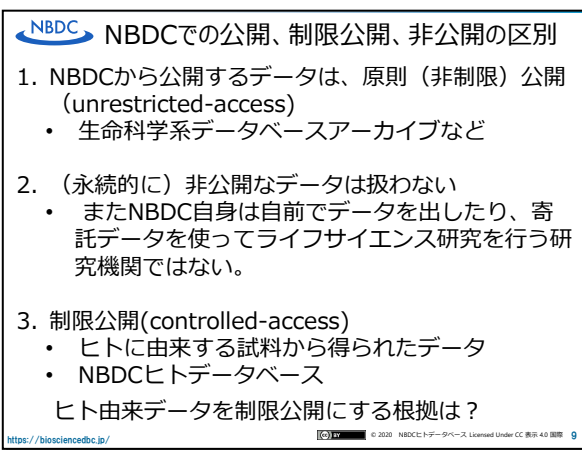
まず、データの区分です（図 3）。NBDC では、ヒ



(図 1)



(図 2)



(図 3)

ト以外のデータも扱っていますが、基本的には非制限公開（フルオープン）です。生命科学系データベースアーカイブなどを、クリエイティブコモンズなどのライセンスで公開しています。NBDC 自体は、論文発表前のデータなど一時的に非公開のデータを預かることはありますが、永続的に非公開なデータを扱うことはありません。また、NBDC 自体が NBDC データベースのデータを使ってライフサイエンス研究そのものをするということもありません。中立的な立場でデータベースを運営しています。

その中で、やはりヒト由来のデータだけではどうしても制限公開（controlled-access）で扱わなければいけません。その根拠が、ヒト由来試料を対象にした医学研究を実施する上での医学指針やゲノム指針など、各種の倫理指針です（図 4）。これに沿って研究なりデータ共有なりをしなければなりません。また、これらの指針は個人情報保護法に裏付けられています。

ゲノム指針は 2001 年に策定されました（図 5）。基本方針の 5 番目、「個人の権利の保障の科学的又は社会的利益に対する優先」というところが一番重要ではないかと思います。そういった精神に沿って制度設計がされています。

また、個人情報保護法も、完全施行されたのが 2005 年ですが、2017 年に改正されました。個人情報の定義が明確化され、ヒト由来データに関しても、その枠組みで個人情報として扱われることになりました。

では、具体的に何が個人情報になるのかというと、

NBDC ヒト由来試料を対象とした医学研究を実施する上で遵守すべき指針

- ・人を対象とする医学系研究に関する倫理指針
(適用範囲：人を対象とする医学系研究全般)
- ・ヒトゲノム・遺伝子解析研究に関する倫理指針
(適用範囲：ヒトゲノム・遺伝子解析研究)
- ・遺伝子治療等臨床研究に関する指針
(適用範囲：遺伝子治療等臨床研究)

これらの指針の根拠法：

- ・個人情報の保護に関する法律
- ・行政機関の保有する個人情報の保護に関する法律
- ・独立行政法人等の保有する個人情報の保護に関する法律
- ・個人法保護条例

<https://biosciencebc.jp/> © 2020 NBDCヒトデータベース Licensed Under CC BY 4.0 図4 10

(図 4)

まず、ゲノム研究で扱うレベルの DNA の配列情報は、ほぼ全て個人識別符号という個人情報に位置付けられました。また、病歴や投薬情報といった医療関係データは、さらに厳格な管理が求められる要配慮個人情報に該当することになりました。これらの定義は法律や政令にきちんと定義されています。要配慮個人情報になると何が大変かという、データを第三者に提供するためにあらかじめ同意を得なければいけない、いわゆるオプトインの手続きを経ないといけないことです。

ステークホルダーとの取り決め

2 番目に、データ所持者、データ利用者等のステークホルダーとどのような取り決めを行っているか。ヒトのデータですので、データ提供者、データ利用者の他に、研究参加者（試料提供者）もステークホルダーになります。データ提供者が研究のために研究参加者から試料をもらうときには、どういう研究に使うかという目的を説明した同意説明文書が必ずあり、研究参加者はその文書にサインをしているはずであるという取り決めがあります。また、データベースセンターである NBDC とデータ提供者や利用者との間の取り決めとしては、NBDC が作ったヒトデータ共有ガイドラインがあり、これに沿って運営しています。

ガイドラインの中には、データ提供時に提供者が守らなければいけない必要事項が書かれています。まず、データ提供者が実施した研究がその組織の倫理審査委員会で承認されているということを、研究計画書、承

NBDC ゲノム指針（2001年）の基本方針

1 基本方針

本指針は、遺伝情報が得られる等のヒトゲノム・遺伝子解析の特色を踏まえ、全てのヒトゲノム・遺伝子解析研究に適用され、研究現場で遵守されるべき倫理指針として策定されたものである。本指針は、人間の尊厳及び人権が尊重され、社会の理解と協力を得て、研究の適正な推進が図られることを目的とし、次に掲げる事項を基本方針としている。

- (1) 人間の尊厳の尊重
- (2) 事前の十分な説明と自由意思による同意（インフォームド・コンセント）
- (3) 個人情報の保護の徹底
- (4) 人類の知的基盤、健康及び福祉に貢献する社会的に有益な研究の実施
- (5) 個人の権利の保障の科学的又は社会的利益に対する優先
- (6) 本指針に基づく研究計画の作成及び遵守並びに独立の立場に立った倫理審査委員会による事前の審査及び承認による研究の適正の確保
- (7) 研究の実施状況の第三者による実地調査及び研究結果の公表を通じた研究の透明性の確保
- (8) ヒトゲノム・遺伝子解析研究に関する啓発活動等による国民及び社会の理解の増進並びに研究内容を踏まえて行う国民との対話

<https://www.mhlw.go.jp/files/06-Seisaku/jouhou-10600000-Dajinkankouseikagaku/0000153405.pdf> © 2020 NBDCヒトデータベース Licensed Under CC BY 4.0 図5 11

(図 5)

認書とともに示していただく必要があります。そして、もちろん研究参加者からインフォームドコンセントを取っていて、なおかつ、そのインフォームドコンセントの中にデータベースでデータを共有するということが明記されていることを必ず確認しています。これに同意がないものはヒトデータベースで扱うことができません。最後に、例えば特定の疾患に関する研究にしかこのデータを利用できないなど、インフォームドコンセントに基づいた制限事項があれば、それも申請していただくことになっています。

同じように、データ利用時の必要条項もガイドラインで決めています。データを利用する研究も倫理審査委員会の承認を得なければいけません。また、研究を行う方はそのデータを使った分野での研究経験があるということで、論文等の書誌情報を求めています。あとは、利用するインフラ、IT 環境できちんとセキュリティ対策が行われていることをガイドラインのチェックリストを使って確認しています。

図6は、インフォームドコンセントに基づいた個別の制限事項の表示例です。左にあるのが各データの内容を表示する概要ページです。どのようなデータに関してもこういった概要データは公開しています。その中にポリシーのリンクがあります。NBDC ポリシーはガイドライン標準のポリシーです。隣に hum0101policy というリンクがあるのは、このデータセットに特有の制限事項を付けることができるという例です。

(図6)

制限公開と制限共有の違い

では、制限公開と、特定の研究者・研究機関内のみ限定して研究データを提供する制限共有との違いは何でしょうか。NBDC ヒトデータベースにも制限共有という仕組みがあります。これはデータの公開に先駆けてグループの中で共有する仕組みです。将来的には制限公開系のデータベースに入れる、もしくは非制限公開することが条件になったデータを、一時的に制限共有する枠組みになっています。現在、NBDC グループ共有データベースを運用しており、国立研究開発法人日本医療研究開発機構（AMED）が実施する三つのプロジェクトから出てきたデータが制限共有として格納されています。もちろん、AMED 以外のグループもグループ共有データベースを作る枠組みがあります。

図7の表は、制限公開と制限共有の違いをまとめたものです。まずガイドラインは、制限共有の場合もNBDC ヒトデータ共有ガイドラインに準拠していただきますが、AMED ゲノム制限共有データベース（AGD）の場合は AMED の方針でガイドラインを多少変更することができ、データ提供者が合意した場合のみデータが利用できるというように変更されています。ですので、AGD のデータを利用したいユーザーは、あらかじめデータ提供者の合意を得て NBDC に利用申請していただかないといけません。その代わり、制限共有の場合は、国際 DNA データベースに登録する際に発行されるアクセッション番号が発行されませ

	制限公開	制限共有
ガイドライン	NBDCヒトデータ共有ガイドラインに準拠	左記共有ガイドラインに準拠。助成機関やプロジェクト等の方針をNBDCと協議の上、反映できる
利用審査時のデータ提供者の合意	不要	原則不要。AMEDゲノム制限共有DBでは必要
アクセッション番号の発行	可能	不可能（制限公開データベースへの移行が必要）
データ利用者が利用可能になる時期	遅くともデータ提供者の成果論文発表時	登録直後。合理的な一定期間後に制限公開DBに移行
維持費用	データ提供者、利用者の負担はなし	データ提供者、助成機関やプロジェクト等が負担する

アクセッション番号：DDBJ/EMBL/GenBank国際DNAデータベースに登録された塩基配列の認識番号
<https://biosciencedb.jp/>

(図7)

ん。論文発表するためには、データを使った場合には基本的にこのアクセス番号を書かないといけませんので、きちんとした論文発表ができないことがあります。その場合は制限公開に移行していただく必要があります。これはデータ提供者だけではなく利用者も同じ条件になりますので、利用者にとってもやや制限の多い共有となっています。

制限公開や制限共有を行う機関の実態

次に、データ提供者との関係性についてです。NBDC 自体は、データを作り出す研究の助成はしていません。米国の dbGaP というデータベースを運営する NIH は巨大なファンディング機関ですので、制度上、データ公開を義務化することができますが、われわれはそういうことができません。

その代わり、データ公開を促進する工夫として、まず日本の助成機関である AMED と連携しています。具体的には、AMED 研究データ利活用に係るガイドラインの中に、AMED の研究では必ずデータマネジメントプランを作るようにして、その中にデータ公開予定を記載することを義務化しました。その受け入れ先として AGD やヒトデータベースが含まれています。

2 番目に、アクセス番号を取ることが論文受理の条件となっています。これが制限公開を促進する最も大きなドライビングフォースだと思っています。

3 番目に、インフォームドコンセントに基づいた個別の制限事項を受け付けます。これによって、再同意を取らなければデータベースに寄託できないものに関しても、なるべく制限事項も一緒に取り込むことで公開のハードルを下げる工夫をしています。

図 8 の図が制限公開共有までの流れです。データ審査に 2 週間ほどかかり、専用サイトを作り、JGA にデータを登録するとアクセス番号が発行されるという仕組みになっています。

一方、データ利用者に関しては、ひたすらヒトデータを使いやすい環境を整備してデータ利用を促進しています。これによって制限公開等の必要性も高まるの

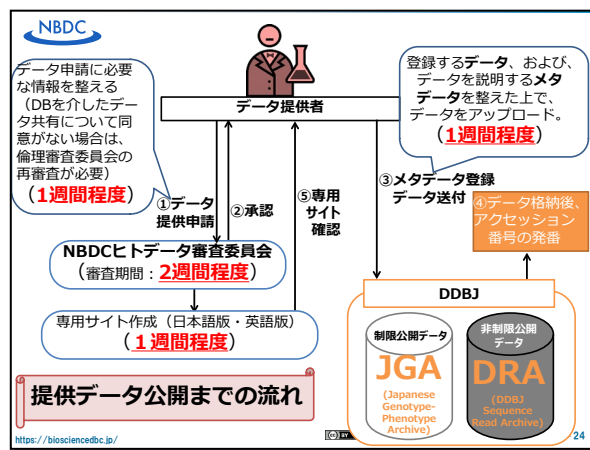
ではないかと考えています。方法としては、まず申請手続きの効率化のために Web システム化を行っており、その運用が始まっています。

2 番目に、ダウンロードしたデータを解析する必要がありますが、そのためにセキュアな計算環境を整えるにはかなりの手間がかかります。そこで利用者が所属する機関内に自前で計算機を用意することを求めているガイドラインを改訂し、自組織以外に利用可能な解析用の計算環境を提供することにしました。現時点では JGA を運用している DDBJ と東北メディカル・メガバンクという、どちらもスパコンを持っているところで、非常に潤沢な解析環境が利用可能です。これを機関外サーバと呼んでいます。

3 番目としては、制限公開データの概要を把握できるように、ゲノムの個人差をあらかじめ検出して、ヒトデータベース内での個人差の頻度を非制限公開として出すことによって、データの中身を少しでも把握してもらおうとしています。

図 9 が利用の流れです。申請していただいて、2 週間ほどの審査の後、JGA からダウンロードできるようになります。図 10 は機関外サーバの仕組みです。以前は自組織のサーバ以外は使えませんでした。現在は自組織のサーバも、DDBJ や東北メディカル・スパコンも使えるようになっています。

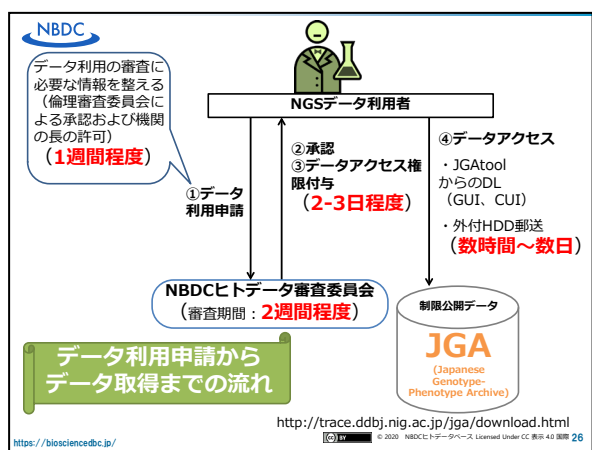
図 11 はヒトデータベースのデータセットの一覧ですが、データの中身がこれだけだと把握しにくいということで、重要なバリエーションの部分だけを取り出し、



(図 8)

頻度という統計情報に加工して非制限公開で閲覧できるようにしたものが TogoVar というデータベースです (図 12)。左側にあるのが、ヒトデータベースに格納されている、それぞれのプロジェクトから出てきたゲノムのデータです。これを同じ手法で再解析して、日本人という大きな集団でのバリエーション頻度を計算します。これは非制限公開できますので、ユーザーが見て、ヒトデータベースに利用申請するかどうかを決めることができます。

同じく、研究に必要なさまざまな他の集団での頻度データや文献データなども、ワンストップで解析できるようにして利便性を上げています。



(図 9)

NBDC NBDCヒトデータベースのデータ一覧

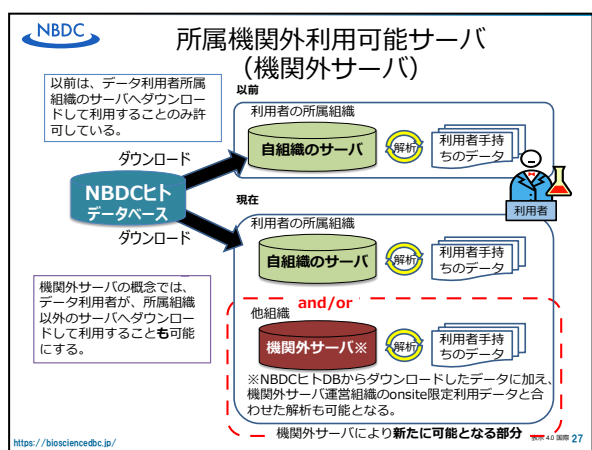
Research ID	研究題目	公開日	データの種類	解析方法	解析者	提供先 (対象施設)	提供先 (対象施設)	アクセス制限
homo0136.v1	広島県立広島医療センターから提供されたゲノムデータ (広島県立広島医療センター)	V1: 2018/05/22	SNP-chip	ブロードバンド解析 (Affymetrix (Axiom Axiom))	広島県立広島医療センター	広島県立広島医療センター	広島県立広島医療センター	非制限公開
homo0134.v1	岡山県立岡山医療センターから提供されたゲノムデータ (岡山県立岡山医療センター)	V1: 2018/04/06	NGS (Exome)	ブロードバンド解析 (Illumina (HiSeq 2500))	岡山県立岡山医療センター	岡山県立岡山医療センター	岡山県立岡山医療センター	制限 (Type 1)
homo0133.v1	愛媛県立愛媛医療センターから提供されたゲノムデータ (愛媛県立愛媛医療センター)	V1: 2018/04/06	NGS (Exome)	ブロードバンド解析 (Illumina (HiSeq 2500))	愛媛県立愛媛医療センター	愛媛県立愛媛医療センター	愛媛県立愛媛医療センター	制限 (Type 1)
homo0129.v1	徳島県立徳島医療センターから提供されたゲノムデータ (徳島県立徳島医療センター)	V1: 2018/04/06	NGS (Exome)	ブロードバンド解析 (Illumina (HiSeq 2500))	徳島県立徳島医療センター	徳島県立徳島医療センター	徳島県立徳島医療センター	制限 (Type 1)
homo0128.v1	小笠原町立小笠原医療センターから提供されたゲノムデータ (小笠原町立小笠原医療センター)	V1: 2018/04/27	SNP-chip	ブロードバンド解析 (Affymetrix (Axiom Axiom))	小笠原町立小笠原医療センター	小笠原町立小笠原医療センター	小笠原町立小笠原医療センター	制限 (Type 1)
homo0124.v1	ヒト1号染色体から提供されたゲノムデータ (ヒト1号染色体から提供されたゲノムデータ)	V1: 2018/04/06	NGS (Exome)	ブロードバンド解析 (Illumina (HiSeq 2500))	ヒト1号染色体から提供されたゲノムデータ	ヒト1号染色体から提供されたゲノムデータ	ヒト1号染色体から提供されたゲノムデータ	制限 (Type 1)

課題: データの中身を把握しにくい

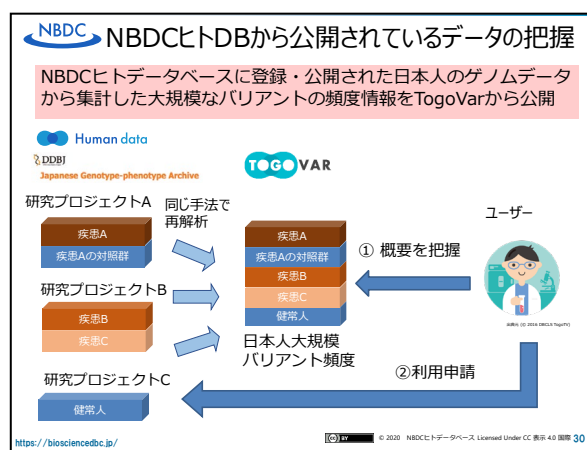
<https://biosciencedbc.jp/>

© 2020 NBDCヒトデータベース. Licensed Under CC BY 4.0 図11 28

(図 11)



(図 10)



(図 12)

第 1 回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

農研機構統合 DB の構築と データ共有の取り組みについて

桂樹 哲雄

(農業・食品産業技術総合研究機構 (農研機構))

講演要旨



近年、農業分野における研究環境の ICT 化にともなって電子データが急増している。多大な労力をもって生み出された研究データは、それ自体が貴重な財産であることから、これらを適切に収集・管理し、利活用することが研究活動を加速化し、従来の分野・領域を超えた学際的な研究を創出できると期待される。そこで、農研機構では研究データの適切な保存・管理・整理を目的としたデータ基盤として、「農研機構統合 DB」を構築し、2020 年度から試験運用を開始した。データ再利用の観点から、機構内のデータは基本的に機構職員が自由に参照できる仕組みを構築した一方、ライセンスや機密保持の観点から、参照制限をかけるものも存在する。

本講演では、農研機構統合 DB の概要を示したのち、オープン・クローズド戦略に基づいて実際にどのようなルールと仕組みでデータの共有を行っているかを紹介する。まだ手探りのところもあるため、本講演を通じて皆様のご意見を伺いたい。

桂樹 哲雄



2002年大阪府立大学工学部海洋システム工学科卒業、2005年同大学院工学研究科機械系専攻博士前期課程修了、2011年同航空宇宙海洋系博士後期課程単位取得退学、2014年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。日本学術振興会特別研究員、豊橋技術科学大学情報・知能工学系助教を経て、農業・食品産業技術総合研究機構農業情報研究センター データ戦略推進室主任研究員に着任。流体計算の分野で研究キャリアをスタートし、バイオインフォマティクスの分野で植物代謝の計算手法を開発、ケモインフォマティクスの分野で化合物の構造活性相関の研究に従事した。現在は農業データ利用促進のためのデータベースに関する研究・開発を行う。

農研機構統合 DB 開発経緯

私の所属するデータ戦略推進室は 2019 年 4 月に設置されました。そこから農研機構統合 DB を作ろうということで、2020 年 3 月に DB 運用ガイドラインを整備し、6 月から機構内限定で試験公開を始めました。農研機構統合 DB のデータを外部ユーザーと一部共有開始したのが今年 9 月です。2021 年 4 月から本格運用を開始予定です。

開発の背景としては、データがたくさんたまってきたことと、経年によってデータがどんどん散逸しているということがありました。職員が退職するとデータ

がどこにあるか誰も知らないということがよく起きてきたので、分野横断的かつ統一的なデータ基盤の構築を急務として、そのデータ基盤として農研機構統合 DB を構築しました(図 1)。コンセプトは「人をつなぐ、データをつなぐ」で、農研機構全体での研究データの共有・活用を目指し、研究データを通じて人をつないだり、また、われわれはグラフデータベースを導入していますので、それでデータ自体をつないだりしています。

DB の構成としては、オンプレミスとクラウドに分散しており、DB を 1 次 DB と 2 次 DB に分類してい

まず（図 2）。1 次 DB は、メタデータを整理して研究データをカタログとして塊で置いておくものです。2 次 DB はグラフデータベースで、データを連携して関連データをつなげて解析できるようにするものです。1 次 DB にためたものから選抜して 2 次 DB に持っていくという構成にしています。農研機構内に 1 次 DB、クラウド環境に 2 次 DB を置いて、両方に機構内からも機構外部からも適切にアクセスできるように、ファイアウォールや認証システムなどを導入してアクセス制御をしています。

1 次 DB (カタログ型 DB)

1 次 DB は、カタログ型のメタデータを入れて管理するメタデータ DB と、そのデータ本体を置くオブジェクトストレージの二つで構成されています。データにメタデータを付けてカタログ型 DB で管理し、その実体を Amazon S3 互換オブジェクトストレージに格納するという設計になっています。

特長としては、機構内のデータを集約して安全に保管することで、研究者のデータ保存の負担を軽減し、

自分たちでハードディスクを買わずに済むということと、メタデータを整備することで機構内の有用な研究データの存在を可視化できるということがあります。メタ情報を基に検索可能で、他の人がどのような研究をしているかといったことから新たなテーマを発見したり、展開したりできます。揭示板機能によって研究者間のコミュニケーションを促進することも目指しています。また、柔軟なアクセス制御によってデータを安全に共有できます。

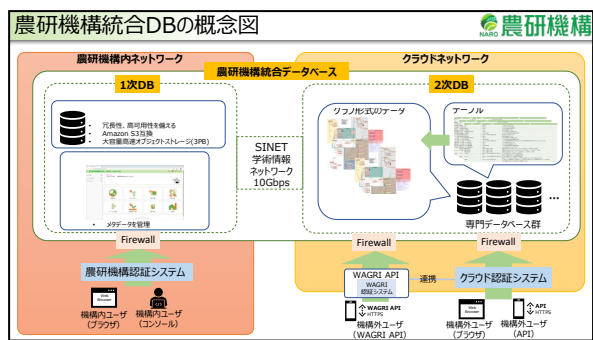
それから、完全に外部に出してしまうデータに関しては、別システムを立ち上げました。もし公開用のシステムに何らかの不具合で外部から不正に入られてしまっても、機密情報は出ないようにするという仕組みを持たせています。

この1次DBには、農研機構に集まってくるあらゆるデータが入ります(図3)。ドローンから撮影した画像セットや、定点カメラあるいはセンサーなどで取得した観測データなどがリアルタイムで入ってくることも想定して設計しています。もちろん Excel データやゲノム配列データ、それから文献データや設計図書も置いてもらおうと考えています。

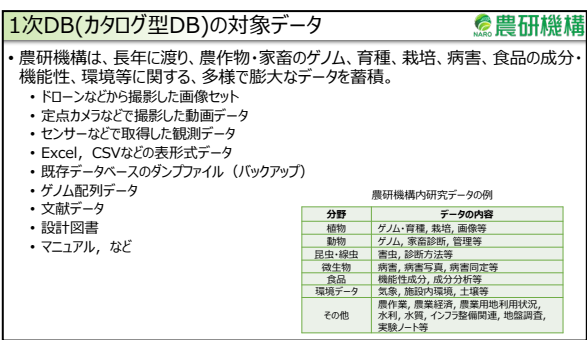
データに付けるメタデータの項目も、昨年（2019年）、一生懸命考えました。もちろんデータをただ置いておくだけでは他の人にはどのようなデータが分からないし、どんなライセンスかということも分からないので、そういう情報を整理しようということになりました。ただ、あまりたくさん付け過ぎると、実際に入力する人に、入力したくない、面倒くさいと言われるので、どれぐらいのデータであればみんな入れてく



(图 1)



(图 2)



(圖 3)

れるか、そして検索にきちんと引っ掛かるかということを考えてみました。

設計方法としては、Dublin Core をベースにして、農研機構で必要と思われる項目を追加し、入力には煩雑にならないように語彙統制をしつつ、しかし柔軟に、項目は欲張らずというということでメタデータ項目を考えています。結果的には Dublin Core や JPCOAR と似た項目となりました。このメタデータは、これだけでは足りないということで、今後カスタムメタデータを付けられるように設計しようと、データベースの改修を行っているところです。

2次DB

2次DBについては少しだけ説明します。グラフデータベースやテーブルデータベースとして Oracle のクラウドのデータベースを使いましょうということでデータを入れていっています。Linked Open Data などと連携できます（図4）。

2019年に私どもの部署ができてからデータを集め出し、今、その入れ物を農研機構統合DBとして構築したところで、これからどんどんデータが入ってくることを期待しているのですが、データの種類としてはあらゆるデータがあります。紙データ、リアルタイム計測データ、マルチメディアデータ、実験データなどがあるのですが、結局のところ、カテゴライズするとしたら、電子ファイルといってもテキストデータが入っているものと入っていないものに分かれます。入っていないものは1次DBにメタデータを付けて放り込み、テキストデータがあるものに関しては全文検索が

できるようにしようと考えています。また、テーブルデータやグラフデータは、2次DBで専門のテーブルデータなどを扱えるデータベースにしていますので、そこに入れていこうと考えています。

農研機構統合DBのオープン・クローズド戦略

これは皆さんのご意見も頂けたらと思っているのですが、農研機構でも、オープンデータとシェアードデータ・クローズドデータを分けて考える必要があるという議論をしてきています（図5）。

オープンデータに関しては、選抜したデータを公開することと、公開用DBを別途用意して安全性を考慮することを考えています。メタデータに関しては一部だけ検索できるようにしようとしています。

シェアードデータ・クローズドデータに関しては、農研機構は国の機関ですし、機構のお金で作ったデータは機構全体で共有することを基本に考えていこうということになっています。これまで農研機構ではあまりそういう考え方が一般的ではなかったのですが、もうそういう時代ではないので、どんどん共有していきましょうということで動いています。メタデータも共有して、農研機構内のユーザーであればメタデータからデータを検索できるようにしようと考えています。

ただし、もちろん農研機構のデータを何でも共有していいわけではありませので、それぞれのデータセットで個別に共有範囲を指定することもできるようにしています。1次DBは柔軟なアクセス権を設定して、ユーザー権限とグループ権限を設定することが可能ですので、便利なNASのように利用できるということ

2次DBの概要	農研機構
1次DB内のデータセットからデータを選抜し、整理・格納 - 有用性の高いもの、他のデータとの関連性の高いものを選抜 - プロパティグラフ、RDFなどの形で整理 - 統計的解析・機械学習による知識発見をサポート - RDFデータサーバを構築 - クラウド型グラフDBを導入 - FusekiによるSPARQLエンドポイントの提供 - テーブルデータも格納可能 - 解析ツールを提供（順次追加） - 育種データビューワ(NARO Pedigree Viewer)を開発	
1次DB → データにメタ情報を付けて、塊で格納 2次DB → 1次DBの中から解析用に選抜されたテーブルデータ、グラフデータを格納	

(図4)

農研機構統合DBのオープン・クローズド戦略	農研機構
- オープンデータ - 選抜したデータを全世界に公開 - 公開用データベースを別途用意し、安全性を考慮（2021年4月運用開始予定） - メタデータ(一部抽出)による検索も可能 - シェアードデータ・クローズドデータ - データは機構内全体で共有するのが基本 - メタデータも共有 → 農研機構内ユーザーはメタデータから検索可能 - ただし、それぞれのデータセットで個別に共有範囲を指定することも可能 1次DBは柔軟なアクセス権設定が可能 → 便利なNASのように利用できる - ユーザー権限、グループ権限を設定可能 - 外部利用者を登録して共有することも可能（ただし、メタデータの利用は限定的） - メタデータを秘匿することも可能 - メタデータはインデックス化されており、通常は機構内ユーザーの検索結果に表示される（機構内限定共有）。 - 検索結果にも表示されないように設定することも可能（要申請）	
データ作成者の権利は、研究データ利用規約によって担保（データが勝手に他人に流用されることはない）	

(図5)

を宣伝しています。また、外部利用者（県の試験場の方、大学の先生など）を登録して共有することも可能です。ただし、その場合にはメタデータの利用を限定的にしようとしています。というのも、メタデータで検索できてしまうので、データが見えてしまうのです。メタデータが見えてしまうのはまずいということで利用を限定しています。

それに関連して、メタデータ自体を機構内のユーザーに対しても秘匿することを可能としています。というのは、メタデータはインデックス化されているので検索でき、検索結果に一部のデータが表示されてしまいます。そこにデータが存在することが分かってしまうのも困るということがありますので、検索結果にも表示されないように設定することが可能です。これは申請ベースで受け付けています。

先ほど、農研機構の文化のようなことを少し言いましたが、やはりデータ作成者には、これからはデータを機関内で共有するのだといっても、勝手に使われてしまうのではないかと心配される方が非常に多いです。従って、データ作成者の権利は研究データ利用規約で担保されているのでデータが勝手に他人に利用されることはないということを強調しています。

公開用メタデータとしては、一部だけ公開するという事で項目を選択しています。例えば、機構内のIDなどは機構外の方には関係ないものなので出しません。問い合わせ先（連絡窓口）に関しては、データセットの責任者に直接連絡できるようにしようと考えていますが、これについてはまだ少し変わるかもしれません。今、内部で相談しているところです。

利用促進のための3要素

データベースは今やっと試験運用が始まったところなので、これからどんどん皆さんに使ってもらいたいと思い、利用促進のための3要素を考えました。まず、運用ガイドラインを策定しました。それから、データ利用規約も策定しました。また、AI スパコンを構築しましたので、それと併せてDBに関しても教育を行

って広報活動をして、皆さんに使ってもらおうと考えています。

まず、運用ガイドラインでは、研究データの定義を、研究の過程でできたデータの全部としています。研究で発生したデータは機構全体にとって貴重な研究資産なので、個人やチームだけで保存するのではなく、機構が定めた安全な場所に永続的にバックアップし、分野横断的に使っていきたいということを定めています。DBの運用開始後は、研究成果となるデータは原則機構内での再利用を認めるということを明記しています。ただし、秘匿性が高い場合など、別途認められた場合は利用範囲を限定できます。

そしてデータ利用規約では、「データを利用した研究について、学会発表や論文投稿等の外部発表を行う場合や特許提案等を行おうとする場合は、事前にデータ提供者に連絡し、著者、引用方法、謝辞、出願等について協議すること」「データ提供者は、提供したデータによってデータ利用者が被ったいかなる損害にも責任を負わないこと」というように、データを出した人が全ての責任を負うわけではないことを明記しているのがポイントです。

クローズドデータへのアクセス制御

統合DB全体をファイアウォールやWebアプリケーションファイアウォールできちんと保護しています。あるいは、IdPでアクセス管理しているのですが、それでユーザーごとやグループごとに権限を制御し、トークンを用いて認証・認可ができるようにしています。

実際の手順

実際にDBにデータが入ってきたときに、われわれが農研機構内でどうしているかということを紹介しましょう。農研機構の研究は、大課題、中課題、小課題に分類されています。その中課題ごと、あるいは部署を横断するプロジェクトがある場合にはその責任者をオーナーとするフォルダを作ります。責任者がそのフォルダ内のデータに責任を持って必要なメンバーをユーザ

ーID やグループ単位で登録し、それぞれに権限を付与する形でデータを管理しています。

それから、メタデータのアクセス制御ももちろん行っていますし、先ほどから言っているように、秘置したい場合には秘置することも可能です。

異動・退職時のデータ保存・権限移譲

職員の異動・退職時は、データがどこかへ消えてしまうタイミングでもあるので、統合 DB への登録をもってデータ移管を完了することにしていきます（図 6）。異動・退職者はアクセス権を上長または後任者などに変更し、変更後は退職者のアカウントは速やかに削除するという運用方法を採ろうとしています。そうすると、退職者が持っている全てのデータのアクセス権を手で変更するというのは非常に大変ですので、その移動の仕組みを現在開発中です。退職後については、研究成果管理規定に従って、業務上、認められれば外部からのデータアクセスを一部許可することにしていきます。

まとめ

以上、農研機構統合 DB と、そのオープン・クローズド戦略の実際について紹介しました。取得されたさまざまなデータをより高度に利活用できるよう、今後も農研機構統合 DB を改良していきたいと思っています。

異動・退職時のデータの扱い	農研機構
・異動・退職時のデータ保存・権限移譲 ・異動・退職時は統合DBへの登録を以ってデータ移管を完了する。 ・異動・退職者は、アクセス権を上長または後任者などに変更する。 ・退職後、退職者のアカウントは速やかに削除する。 ・1次DBには権限移譲の仕組みを導入する（今年度末から利用可能）。 ・退職後のデータアクセスについて ・研究成果管理規定に従い、業務上、認められれば外部からの一部データアクセスを許可。	

（図 6）

第 1 回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

材料分野の研究データ公開における理想と課題 ～NIMS の事例から～

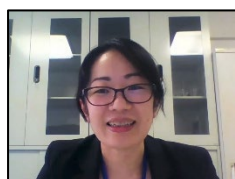
篠田 陽子

(物質・材料研究機構 統合型材料開発・情報基盤部門)

講演要旨



物質・材料研究機構では2017年に材料データプラットフォームセンター（DPFC）を設置し、材料分野におけるデータ駆動型研究の促進に貢献する材料データプラットフォーム DICE の構築に取り組んでいる。DICE は材料データの入り口（収集）から出口（利活用）までを一貫して扱うシステムであり、取り扱う研究データの来歴や性質は多岐にわたるため、DPFC では各サービスの在り方や運用について検討し議論を重ねてきた。本発表では DICE での事例を紹介しつつ、研究データ公開における理想と課題を考える機会としたい。

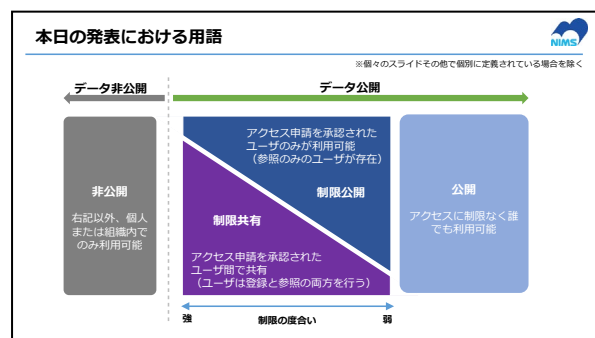


篠田 陽子

国立研究開発法人物質・材料研究機構 統合型材料開発・情報基盤部門 材料データプラットフォームセンター データサービスチーム所属。研究機関、研究費配分機関、大学研究室で研究支援業務に従事後、2016年に同機構に転職し、2019年6月より現職。研究データポリシーやサービス利用規約の策定に携わっている。

本日の発表における用語

本編に入る前に、本日の発表における用語を整理させていただきます（図 1）。「公開」はアクセスに制限なく誰でも利用可能なこと、「制限公開」は承認ユーザーのみが利用可能なこと、「制限共有」は承認ユーザーの間で共有（ユーザーは登録と参照の両方を行う）



(図 1)

ザラ間でデータを共有すること、そして、これらを合わせて「データ公開」、それ以外を「非公開」「データ非公開」とします。

物質・材料研究機構（NIMS）について

最初に、NIMS について簡単にご紹介します（図 2）。NIMS は、その名が示すとおり、国内で唯一、物質・材料科学の研究を専門とする公的研究機関です。茨城県つくば市内の 3 地区と、兵庫県西播磨地区の SPring-8 内に事業所があります。役職員は約 1,600 人で、このうち 50～60%が研究職員、25%がエンジニア職員として働いています。研究組織としては 8 つの研究拠点・部門があり、各組織で特色ある物質・材料

の研究開発を行っています。

私が所属する統合型材料開発・情報基盤部門（MaDIS）は、データとの融合で物質・材料の研究開発の在り方を革新することを目指して 2017 年 4 月に設置された新しい部門です。材料データプラットフォームセンターでは、データ駆動型の研究開発を支える基盤として、本日紹介する材料データプラットフォームの構築・運用に取り組んでいます。私はここで、主にデータの利用に関する枠組みやルールづくりに携わっています。

材料分野の研究データ

材料分野の研究データにはさまざまなものがあります。それらのデータをどのように取り扱うかを考えると、図3のようなイメージになります。左側の制限公開や非公開のところにあるものが多めになっています。実際には個々のデータで状況が異なるため、必ずしもこの限りではないのですが、材料分野の研究データは、例えば生物や化学、物理など他の基礎科学の分野に比べると、相対的にオープンにされにくい傾向にあると

言えます。

では、なぜオープンにされにくいのか。大きな理由の一つとして、材料は産業界や知財との関連が強いということが挙げられます。使われてこそその材料といわれるように、材料は社会や産業に役立つ物質であり、使える、そして使われる材料を作り出すための化合物組成や構造、成形プロセスなどは、研究戦略やノウハウに直結しやすい状況があります。

その一方で、近年のオープンサイエンスの流れや、マテリアルズ・インフォマティクス、マテリアルズインテグレーションといったデータ駆動による材料設計・開発への取り組みも加速度的に進んでおり、データの利活用に対する期待や要望も高まっています。

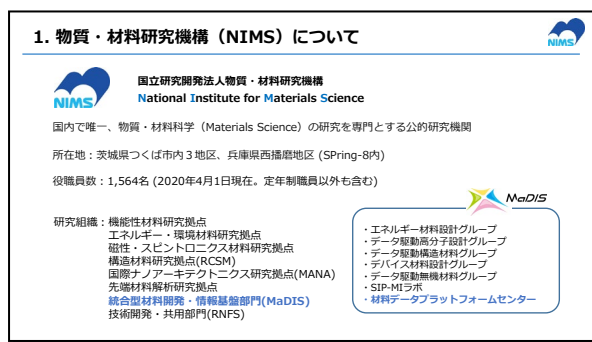
こういった背景からも、材料分野の研究データにおいては、研究の新たな発展、そしてイノベーション促進のためにも、保護と利活用のバランスを取りながら戦略的にオープンにしていくことが重要になってきます。

NIMS の研究データポリシー

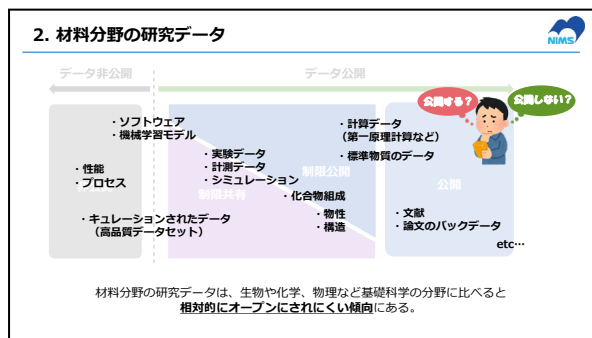
NIMS は国研として日本全体のイノベーションシステムをけん引する役割があり、また、NIMS の強みであり資産である材料データの利活用を図っていくという使命があります。このため NIMS では、研究データに対する基本的な考え方や利用ルールの整備と、データ利活用のためのデータインフラの構築・運用を両輪で進めてきました。

2018年8月に研究データポリシーを制定しました。このポリシーの最初でも述べているのですが、研究活動を通じて得られる研究データは重要な知的資産であり、適切な管理と広く有効な利活用を図っていくために、研究データの収集、管理、利活用に関する基本的な方針を、研究データポリシーとして定めています（図4）。基本的な考え方としては、NIMSの研究戦略や分野特性を考慮するとともに、研究データの性質や来歴等に留意しつつ取り扱うこととしています。

一日に研究データといっても、形式や種類はさまざま



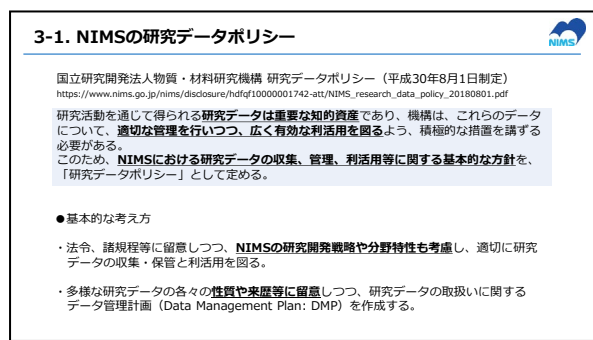
(圖 2)



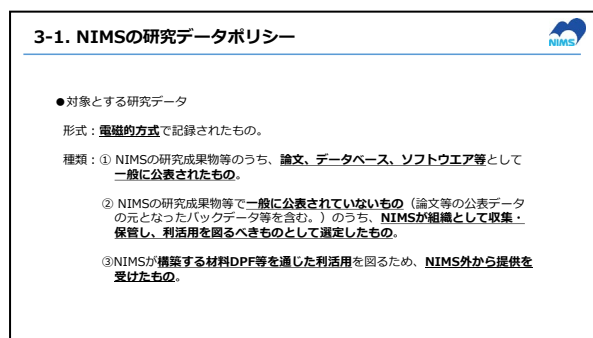
(圖 3)

まです。NIMS の研究データポリシーでは、対象とする研究データを電子データとしています（図 5）。種類としては大きく 3 つ、①NIMS の研究成果物等のうち、一般に公表されたもの、②NIMS の研究成果物等で一般に公表されていないもののうち、NIMS が組織として収集・保管し、利活用を図るべきものとして選定したもの、③NIMS が構築する材料データプラットフォーム等を通じた利活用を図るため、NIMS 外から提供を受けたものとしています。

これらの研究データを提供・公開するときの考え方として、まず研究データの提供・供用については、NIMS 内外のデータの利活用によって発展的な成果を生み出すことを目指すとともに、必要に応じてアクセス制限や利用の限定などの措置も取り得ることとしています。研究データの公開（ポリシーでは、インターネット上で公表しアクセスできるようにすることを指す）については、公開することを原則としつつ、権利保護や情報保護の観点から配慮が必要なものは公開対象外としています。また、公開に際しては必要に応じて条件を付けることとしています。



(図 4)



(図 5)

材料データプラットフォーム DICE

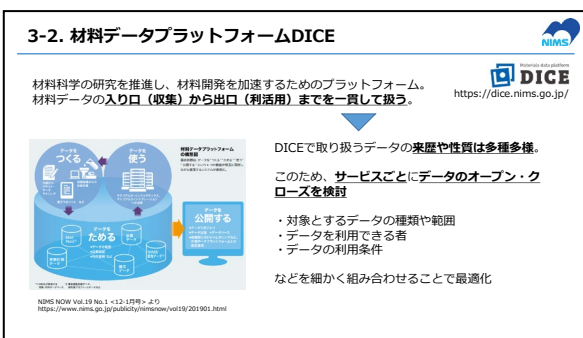
この研究データポリシーと両輪となっていくデータインフラが、材料データプラットフォーム（DICE）です（図 6）。材料科学の研究を推進し、材料開発を加速するためのプラットフォームであり、材料データの入り口から出口までを一貫して取り扱うシステムです。

DICE で取り扱うデータは、研究活動における日々の実験データや計測データ、データベース、論文や報告書、さらには利活用のために NIMS 外から提供されるデータなど、来歴や性質がさまざまです。このためプラットフォームでは、提供サービスごとにデータのオープン・クローズを検討する方法を取っています。具体的には、何をどこまで公開するのか、誰に使わせるのか、どう使わせるのかを細かく組み合わせることによって最適化を図っています。今回は DICE のデータ公開の事例として、MDR、MatNavi、共有データリポジトリの 3 つのサービスについて説明します。

Materials Data Repository (MDR)

MDR は、「公開」の事例です（図 7）。MDR は材料研究データや文献等を収集・保存し、インターネットを介して広く利用できるようにするための公開データリポジトリで、2020 年 6 月 15 日から公開を開始しています。名前に「材料データ」とあるように、材料に関するメタデータ項目をサポートしており、登録データに DOI を付けることによって引用を容易にしているのが特徴です。

MDR で公開するデータは主に NIMS の公知データ



(図 6)

です。公開されたデータは誰でも利用することができ、検索・閲覧・ダウンロードが可能です。データ登録に関しては、現在は NIMS ユーザーに限定しています。

データの利用条件としては、各データに示されているライセンス条件に従うこと、それから、公表時に出典を明示することとしています。

NIMS 物質・材料データベース (MatNavi)

MatNavi は、「制限公開」の事例です（図 8）。高分子、無機、金属などの十数種類の材料データベース等で構成された統合データベースシステムで、予測システムなどのアプリケーションも提供しています。これまでたくさんのユーザーにご利用、ご活用いただいています。

MatNavi では、NIMS が組織として収集したデータ（データベース）や、NIMS が利用および管理の権限を保有しているデータを公開しています。データベースの利用にはユーザー登録をお願いしています。利用者はデータの検索・閲覧ができますが、Web スクレーピング等によるデータの大量取得は禁止しています。



(図 7)



(図 8)

また、MatNavi のデータは、教育または研究開発目的での自己利用に限定して利用を許諾しています。

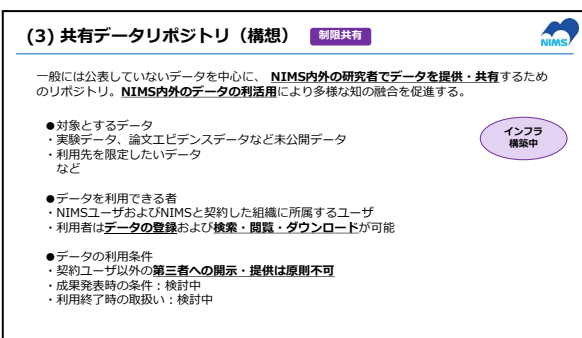
共有データリポジトリ (構想)

共有データリポジトリは、「制限共有」の事例です（図 9）。こちらはまだ構想段階ですが、NIMS 内外の研究者でデータを提供・共有し、これらの NIMS 内外のデータの利活用によって多様な知の融合を促進することを目指すものです。

共有データリポジトリでは、実験データや論文エビデンスデータなどの未公開データ、あるいは利用先や利用者、利用する方法などを限定したいデータなど、一般には公表していないデータを扱うため、利用者は NIMS ユーザーおよび NIMS と契約した組織に所属するユーザーに限定されます。そして、「共有」という名が示しているように、利用者は検索・閲覧・ダウンロードの他にデータの登録も行えることになります。また、このリポジトリで共有されるデータは第三者への開示・提供は原則不可となります。

DICE の事例のまとめ

紹介した 3 つのサービスを比較すると、図 10 の表のようになります。現状、NIMS では、研究データの公開の仕方を基盤あるいはサービスごとに検討していますが、各サービスで扱えるデータの範囲が限定されており、これを広げていくことや、サービス間の連携をうまく図っていくことなど、まだ課題が残されている状況です。



(図 9)

研究データ公開における理想と課題

最後に、研究データ公開における理想と課題について、5つの項目を挙げて話題提供したいと思います。なお、これは主にプラットフォームのサービス担当者の視点から見たもので、発表者の私見も含まれていますので、組織としての考え方とは異なっている場合もあることをご承知おきいただければと思います。

1つ目は、本日のセミナータイトルにもなっている公開（フルオープン）と制限公開／制限共有の境界線です（図11）。明確な基準があり、一律に線引き・判断できるのが理想ですが、実際はデータの内容やデータ提供者の考え方によって公開可否の判断はさまざまであり、さらに時間軸が加わって状況が変化していくこともあります。ですので、結局はデータごとに細やかに判断していく必要があります。また、国研としての役割の中でのバランスという観点もあります。知の最大価値化の実現に向けて、どのポイントに比重を置くかが重要になってきますが、安全や保護を意識すると非公開に偏りがちになります。

2つ目は、データ（提供者）の権利の保護という点

3-2. 材料データプラットフォームDICEの事例（まとめ）			
	公開	制限公開	制限共有
サービス名称	MDR	MatNavi	共有データリポジトリ
対象データ	公知データ	利活用を図るため、NIMSが組織として収集したデータ	一般には公表していないデータ（未公開データ、利用先を限定するデータ、など）
データの提供者	主にNIMS	NIMS	NIMS内外の研究者（組織契約をしたユーザ）
データの利用者	制限なし	利用登録をした者	NIMS内外の研究者（組織契約をしたユーザ）
データ利用者ができること	検索・閲覧・ダウンロード	検索・閲覧	検索・閲覧・ダウンロード
データの利用条件*	- 各データ等のライセンス条件に従う - 出典の記載	- データの大量取得は禁止 - 教育又は研究開発目的での自己利用に限定 - 出典の記載	- 第三者への開示・提供は原則不可 - 成果の取扱いを検討中

*サービス利用に際しましては、規約、約款等で詳細をご確認ください。

（図10）

4. 研究データ公開における理想と課題

●公開（フルオープン）と制限公開／制限共有の境界線

理想：明確な基準があり、一律に線引き・判断ができる。

課題：データの内容やデータ提供者の考え方によって公開可否の判断は様々。時間経過とともに状況が変化することもある（例：論文発表や出版など）すばった線引きや一律に白黒つけることは難しい。

課題：国研としての役割の中での（バランス）。知の最大価値化の実現に向けて、どこに比重を置くか？「安全」「保護」を意識しすぎると非公開に偏りがち。



（図11）

です（図12）。理想としては、法律ですべてきれいに保護されて安心できると思いますが、実際にはデータは無体物であり、民法上も権利の対象とはなりません。また、著作権、特許権、営業秘密なども、ある一定の条件を満たしていないと保護されないということがあり限定的です。このためデータの保護においては、原則として契約を通じて保護を図る必要があります。この手続きで難しい部分が出てきます。

これと関連するのが、3つ目の利活用のシナリオです。データ利用者とデータ提供者の双方にとってうれしい使われ方がされるのが理想ですが、課題としては、例えば成果の取り扱いについて、データを使って得られた生成物・成果物は誰のものかということがあります。特に昨今のマテリアルズ・インフォマティクスやマテリアルズインテグレーション、AIで利用したときの成果の取り扱いなどは難しいところがあります。約款等であらかじめ扱いを定めておくことが必要になる場合もあります。また、データ（提供者）の貢献をどう評価していくか。データの引用・利活用状況の可視化や、その可視化された貢献を組織内で適正に評価に反映していく仕組みづくりも必要だと感じています。

4つ目はデータの提供方法です（図13）。いろいろなデータ公開のステータスがありますが、非公開から公開までワンストップサービスで提供できるのがやはり理想です。課題としては、データの形式や性質、権利関係などが一様ではなく、ステークホルダーも非常に多い状況になっているため、1つのシステムで多様なデータに対応することは、技術面あるいはコスト面の問題もあり、実際のところ困難ということがありま

4. 研究データ公開における理想と課題

●データ（提供者）の権利の保護

理想：法律で保護される。

課題：データは無体物であり、民法上、所有権や占有権、用益物権、担保物権の対象とはならない。著作権、特許権、営業秘密などによる保護は限定的。データの保護は原則として契約を通じて図る必要がある

●利活用シナリオ

理想：データ利用者とデータ提供者にとって双方よし

課題：成果の取扱い。データを使って得られた生成物・成果物は誰のもの？（MIやAIで利用したとき）約款等で予め扱いを定めておくことが必要になる場合もあり。

課題：データ（提供者）の貢献の評価。データの引用・利活用状況の可視化（DOI付与や出典記載その他）や組織内で評価に反映する仕組みづくりが必要。

（図12）

す。

また、ルールを整備していくところでも、ステークホルダーが多くなればなるほど調整が難しくなります。現状としては、サービスごとに対応しながらそれらをシームレスにつなぐ仕組みをつくることで、ワンストップに近いサービスをつくっていくことが必要ではないかと思っています。

5 つ目は費用です。ユーザー視点でいえば費用負担は低い方がいいですが、運用側の視点でいうと、インフラの維持・運用、セキュリティ確保などにおいては多大な費用がかかるという現実があります。持続可能なデータ公開のためには有料化を視野に入れることも必要であり、データを集めるにも公開するにも費用が発生する時代が来るのではないかと思っています。こういった費用が大きくなればなるほど、研究活動そのものにも負担が生じ、影響が出てくる恐れがあります。こういうところも視野に入れながら、うまいデータ公開の仕組みを考える必要があると思っています。

まとめ

材料分野は産業や知財との関連が強く、研究データはオープンにされにくい状況がありますが、研究の新たな発展やイノベーションを促進するためには、研究データを戦略的にオープンにしていくことが求められます。

また、研究データを保護しつつ利活用を図るためには、データ公開に際し、適切なアクセス制限や利用条件を付すことも重要です。研究データの使われ方や利活用シナリオを念頭に置いた上で、ハード（インフラ）

とソフト（契約・規則など）を両輪としてデータ公開の枠組みをつくっていくことが大切だと感じています。

それから、安心・安全なデータ公開を実現するためには、やはりお金がかかります。持続可能なデータ公開のためには、インフラの維持・運用、セキュリティ確保にかかる費用も考えていく必要があると思います。

最後に、当発表に当たり多くのアドバイスとお力添えを頂きました当センターの谷藤センター長、データサービスチームの田沼チームリーダー、桑島さんに御礼を申し上げます。

4. 研究データ公開における理想と課題

●データの提供方法

理想：非公開～公開までワンストップサービスで提供

課題：データの形式や性質は様々でなく、ステークホルダーも多い。
1つのシステムで多様なデータに対応することは、技術とコストの問題もあり困難。
ルールの面でも、ステークホルダーが多くなるほど調整は難しくなる。
サービスごとに対応しつつ、これらをシームレスにつなぐ仕組みが必要。

●費用

理想：ユーザー視点では、費用負担は低いほうが良い。

課題：運用側の視点では、インフラの維持・運用、セキュリティ確保などには多大な費用がかかる。
持続可能なデータ公開のためには、有料化を視野に入れることも必要か。
データを集めるにも公開するにも 費用が発生する時代が来る？
データ公開・利用にかかる費用の増大は、研究活動そのものにも影響がでる恐れあり。

(図 13)

第1回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

社会科学分野におけるデータ提供の実態 ーデータアーカイブ機関と利用者の最前線ー

仲 修平

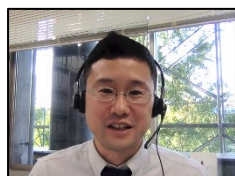
(東京大学社会科学研究所)

講演要旨



本報告の目的は、データアーカイブ機関と利用者の関係に着目して社会科学分野におけるデータ提供の実態を共有することである。具体的には、(1) どのようなフローでデータを利用者へ提供しているのか、(2) データの提供に際していかなる検討を実施しているのかという点である。両者について、東京大学社会科学研究所附属社会調査・データアーカイブ研究センターが構築している SSJ データアーカイブ (Social Science Japan Data Archive) を事例としてみていきたい。そのことを通して、学術目的で利用するデータの公開をより一層進める一方で、利用者へ提供する際に残されている課題を考察する。

仲 修平



東京大学社会科学研究所附属社会調査・データアーカイブ研究センター助教。日本学術振興会特別研究員を経て2018年4月より現職。専門分野は、社会階層論、計量社会学。主な研究関心は、雇われない働き方の実態を量的／質的調査に基づいて検討すること。主著は、『岐路に立つ自営業—専門職の拡大と行方』(勁草書房、2018年)。データアーカイブ研究センターでは、調査基盤研究分野(公開データの準備や提供等に関わる業務)で活動している。

本報告の目的

本報告の目的は、社会調査・データアーカイブ機関の役割、個票データを提供するフローの実態、利用者の申請に対する検討の内実という3点を通じて、社会科学分野におけるデータの公開と制限の境界線を共有することです。

先に結論を申し上げておきますと、原則として利用者へ個票データをより広く公開していこうというスタンスで運用しています。ただし、一定の条件を満たさない場合に限り制限することがあります。重い制限をかけるのはあくまで例外的な運用であることをご理解いただきたいと思います。

所属機関の役割

私が所属している東京大学社会科学研究所附属社会調査・データアーカイブ研究センターは、四つの研究分野から成り立っています(図1)。調査基盤研究分野は、データの保存・収集・公開のプロセスに携わっています。その中でも私は利用者へデータを提供するプロセスに関わっています。社会調査のデータを作っていく社会調査研究分野や、データの利用を促進するためにセミナーや研究会を開いたりする計量社会研究分野もあります。また、国際調査研究分野は、世界のデータアーカイブ機関と連携していくために、学会に行ったり、実際に人と会ってネットワークを構築したりすることを始めており、特に東アジアのデータアー

カイクの連携が進みつつあります。これが全体の状況です。

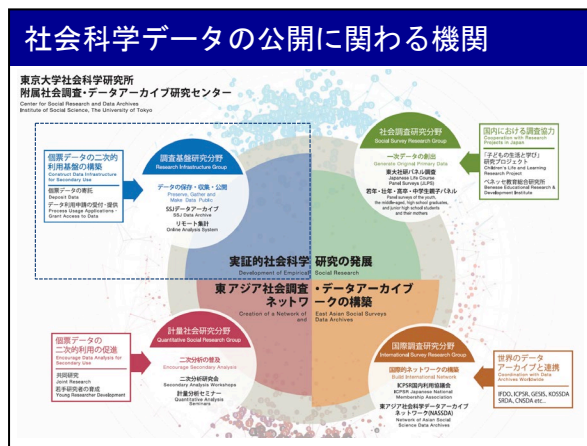
社会科学分野で扱う個票データ

まず、社会科学分野で扱う個票データとは何なのかということを中心に説明します。図2は東大社研パネル調査の調査票の一部です。パネル調査とは、同じ個人を何年も追跡していくタイプの調査です。例えば、現在の仕事について伺いますという形で、働き方や仕事内容について質問項目を立てています。それに対して、回答者は数字を選んでいきます。例えば正職員で専門職であれば、「(1) 働き方」では2番に、「(2) お仕事の内容」では1番に丸を付けます。そういったタイプのデータを作っていきます。

ここで強調しておきたいのは、全ての対象者に対して調査票を郵送配布し、訪問によって回収しているデ

ータであるということです。東大社研パネル調査のサンプルサイズは、おおよそ 5000 人です。調査対象者の自宅に調査員が行って回収しますので、1 票のデータを得ることにかなりの労力がかかっています。ですので、データを公開するときにも、こうしていろいろな人の苦労が積み重なって集まったデータを保存していますので、データがとてもかわいく見えてきます。そういったデータを扱っています。

実際に個票データを配布するときには、図3のような電子ファイルで提供しています。表の見方としては、例えば 1 行目であれば、性別に関しては 1 (男性)、出生年に関しては 1981 (1981 年生まれ) だということが分かります。社会科学分野で主に扱うデータの形は、基本的にはこのように 1 行に一つの単位 (個人や組織) が入っていて、列方向にそれぞれの質問項目が入っているというものです。こういった数値情報の集まりを社会調査データとして公開しています。



(図1)

Social Science Japan Data Archive

(SSJDA) の仕組みと利用状況

では、実際にどういうデータプロセスによって公開していくのかという模式図が図4です。まず各主体が調査を実施して、それを SSJDA に寄託して、公開できるようにデータの処理をして利用者に使っていただくというプロセスで進めています。

図5は、実際に集められたデータと利用している方のおおよその数です。SSJDA は 1998 年からスタート

(図2)

社会科学分野で主に扱う個票データ (例)

個票データ

調査票によって得られた情報を
1行に調査対象1つの情報となるように置き換えたもの

調査対象者の番号 性別 出生年 出生月

	PanelID	CN	sex	ybirth	mbirth
1	1	1	男性	1981	2
2	2	1	1	1975	1
3	3	1	1	1967	10
4	4	1	女性	1985	5
5	5	1	2	1985	10
6	6	1	2	1977	10
7	7	1	2	1978	5
8	8	1	2	1976	9
9	9	1	2	1974	6

表の見方) 1行目: 男性, 1981年2月生まれ

(図3)

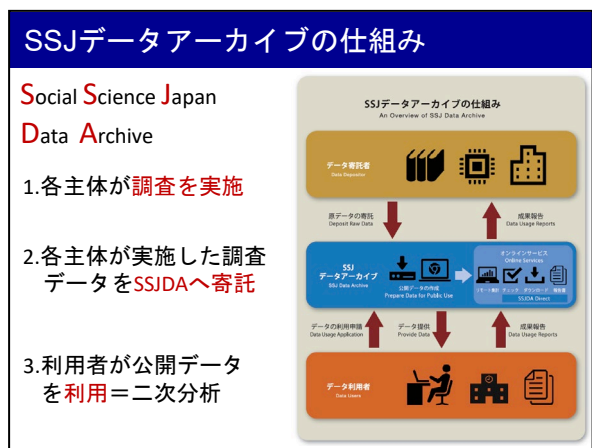
して、現在では公開データセット数や利用者が年々非常に増えていることがご覧いただけると思います。直近3年では約70件ずつ新規のデータセットを公開しています。研究者はもちろん、大学院生や、教育目的で学部生にも利用していただいています。また、日本に関心のある海外の研究者の利用も最近になって増えてきています。

図6はSSJDAを利用した研究業績です。主に学術著書や論文、学位論文です。2次分析による研究成果は増加傾向にあり、若手研究者による学術書の出版、卒論、修論、博論等でご活用いただいています。図書館職員の方からも問い合わせがあります。学部生の教育目的で使っていただく機会が最近になって増えてきているという特徴があります。調査や研究に困った学生がいれば、SSJDAを紹介すると何か研究ができそうだと思っていただけると幸いです。

保有データ

SSJDAで保有しているデータは、経済学と社会学のデータが多くなっています。それ以外の分野でも、経営学や教育学、政治学など、幅広い分野を扱っているという特徴があります（図7）。

現在保有しているデータは、多くは公開中ですが、全てを公開しているわけではありません。公開に至るまでには、データのクリーニングや、ローデータに対して後からコードを付与するアフターコーディングと呼ばれるプロセスもありますので、そのような公開準備をしているもの、あるいは寄託の段階から保存目的で寄託していただいているデータもあります。保有している多くのデータは公開を前提としていますが、公開に向けていろいろなプロセスの途上にあるデータがあります。



(図4)

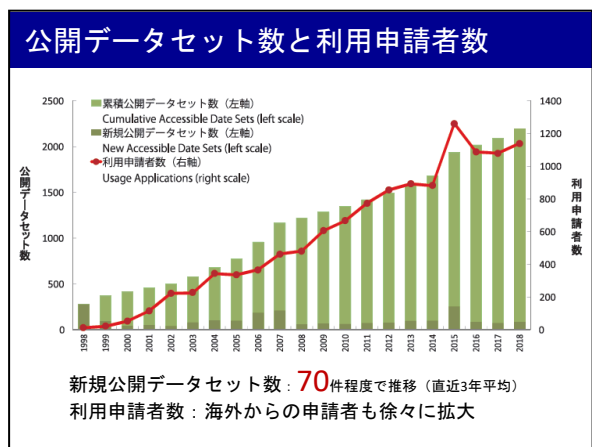
SSJDAのデータを利用した研究業績

	2016年	2017年	2018年
著書	12	19	15
論文	41	52	53
学位論文	124	165	161
合計	177	236	229

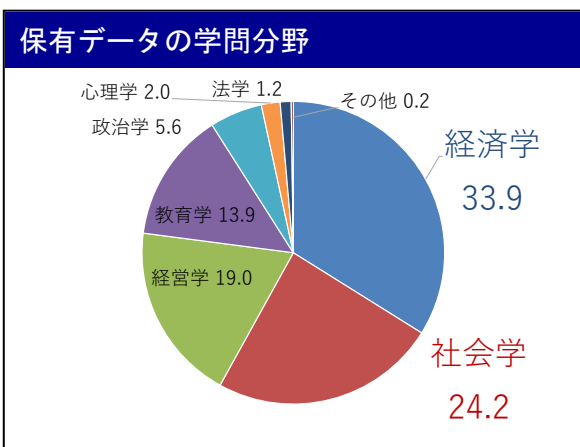
注：成果物は数年経ってから届く場合があるため、公表値とは異なる可能性がある。

二次分析による研究成果は増加傾向
若手研究者による学術書の出版
卒論・修論・博論で活用実績

(図6)



(図5)



(図7)

データ提供フロー

先ほどお示したデータ提供の三つのプロセスの中で、特に本セミナーに関わるのが、SSJDA と利用者との間でどのようなやりとりがなされているかという点です。このプロセスにおいて、セミナーのテーマである公開と制限が行われています。データ提供フローは、図8のような三つのステップになっています。

一つ目は、利用者による利用申請書の提出です。実際の申請画面では、利用の目的や、どのデータを使うかということを選択していただいた上で、研究計画という形で、なぜこのデータが必要なのか、どのように使うのかということを書いていただきます。

二つ目に、書かれた研究計画を見て、SSJDA の中で承認判断をします。今のところ、全ての研究計画を目視によって判断して、場合によっては申請者に補足的な対応を問い合わせる形で進めています。その結果、承認という判断になれば、三つ目としてデータを提供していきます。

提供／制限の内実

提供と制限に対する基本的なスタンスとしては、原則は、利用目的に合致した場合、データを利用者へ提供します。制限は例外的で、あくまで利用条件に満たない場合を想定しています。

利用の承認については、①利用条件に合致しているか、②申請内容が妥当であるか、③利用状況が適切であるかという三つの観点から判断しています。それぞ

れについて、具体的にどのような点を見ているかということをお示したいと思います。

まず、①利用条件に合致しているかという点については、二つのポイントがあります。一つは利用資格、もう一つは利用状況です。SSJDA は、大学または公的研究機関の研究者、あるいは教員の指導を受けた大学院生に利用資格があります。NG になるケースで多いのは、民間企業や学部学生、高校教員などの場合です。ただ、学部学生に関しては、指導教員からの申請で受け付けています。利用状況というのは、過去の申請に対してきちんと対応しているかどうかということです。利用期間は基本的に1年間ですが、それに対して利用報告をしているか、また、教育目的で使った場合には受講者リストをきちんと提出しているかという点から判断しています。通常、図書館でも、借りた本を先に返してから次の本を借りると思いますが、そのように、過去の申請にきちんと対応しているかということもポイントになっています。

②申請内容の妥当性に関しては、研究計画のボックス欄に書かれた内容・目的が妥当であるかどうかという点から判断しています。研究に関しては、研究計画の内容が十分に示されているか、示された内容と調査に含まれている変数が合致しているか、また、教育に関しては、卒業論文の中身や使われる授業名などの記載が必須ですので、その有無を見ています。

③利用状況の適切性については、一度の申請で大量のデータを申請する方がごくまれにいらっしゃいます。データによってはシリーズで提供しているものもあり、そういった場合にはたくさんのデータが必要になります。しかし、そうではなくて、いろいろなデータをとにかく使ってみたいということで、同じ研究計画の目的で大量のデータを申請された場合には、それほど同時に必要ないことを確認した上で、データを絞って提供するケースもあります。また、学生が教員としてアカウントを作成する、複数名で一つのアカウントを共有する、教員が自らのアカウントを学生へ貸し出すなど、アカウントの不正利用が明らかに見られた場合に

データ提供フローの3ステップ

- 1.利用者による利用申請書の提出
- 2.SSJDAによる承認の判断
- 3.データの提供

<申請画面>

利用目的
☒ 研究 ☐ 教育

すべての研究計画を目視によって確認したうえで、承認可否を判断

研究番号	研究名	データ形式	利用期間	更新頻度	備考
PD010	東大総研・東大バカ丸調査 (LPS-VI Wave) 基本データ、2007	東京大学社会科学研究所バカ丸調査プロジェクト	ダウンロード	*期間なし	※利用料

※研究データの提供

研究計画 (200文字以上 400文字以内)

※研究計画の記入欄をここに記入

(図8)

は制限をかけます。具体的に制限を加えたケースを次に紹介したいと思います。

制限のケース

これまでに最も重い制限を加えたケースとしては、利用者のアカウントを停止する措置をとったことがあります。そのケースは次のようなものです。まず、教員が学生にアカウントを貸し出して、その学生がデータを申請してくる、場合によってはその学生からSSJDAの担当者に問い合わせが来るという状況になっていました。この不正利用が発覚した時点で該当の教員に対して改善を促しましたが、十分な対応がなされませんでした。そこで、私たちの機関の中で検討を重ね、合意をもってアカウントを停止することになりました。基本的にはデータをどんどん使っていただきたいのですが、使い方が良くない場合には制限が加わります。ただし、その場合でも、何度も検討を重ねて、結果としてやむなく停止しているという形です。

もう一つ制限を加えた例として、学部ゼミメンバーでアカウントを共有していたケースがあります。同じゼミに所属する人たちが一つのアカウントで利用申請を繰り返したというケースです。これについては先方の先生に改善を依頼して対処されたと伺っていますが、その対応と同時に、SSJDAの中でもシステムを再検討し、2016年3月からは、学部学生の申請は指導教員の先生による申請という形に限定しています。

まとめ

提供と制限の境界線については、データ提供の原則は「常にオープン」という形になります。ただ、制限の可否は利用条件の合致を個別に確認した上で検討しています。特に重い制限は、データに対して不誠実な利用者に対してのみ行っています。

最後に、一担当者から見た提供フローの課題を示したいと思います。個々の申請全てを目視で対応しているので、人員と時間のコストが非常に大きくなっています。ですので、今後データの公開数や利用者が増え

ていった場合に、申請件数に対する処理にも制約が出てくるのではないかと感じています。それに対しては、申請フローを自動化あるいは半自動化にする、あるいは、現在は申請ベースで検討しているものを利用者ベースにする検討を加えるなどの形が可能ではないかと考えています。

第1回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

パネルディスカッション

八塚 茂	(バイオサイエンスデータベースセンター)
林 賢紀	(国際農林水産業研究センター)
池内 有為	(文教大学／研究データ利活用協議会 研究データライセンス小委員会)
海老沢 研	(宇宙航空研究開発機構)
上島 邦彦	(株式会社日本データ取引所)
三橋 信孝	(バイオサイエンスデータベースセンター)
桂樹 哲雄	(農業・食品産業技術総合研究機構 (農研機構))
篠田 陽子	(物質・材料研究機構)
仲 修平	(東京大学社会科学研究所)



●八塚 このディスカッションでは、池内さんをご講演で示された三つの論点を中心に議論していきます。それから、皆さまから頂いているご質問に回答したいと思います。

まず、皆さまのご講演の振り返りを兼ねて、二つの論点についてお一人ずつコメントを頂きます。一つ目は、フルオープンではなく制限公開をする場合の根拠と、制限公開に区分を設ける場合、どのような基準で区別しているかということです。二つ目は、制限公開をしている場合はアクセス申請を受け付けることとなりますが、その妥当性の判断をどのようにされているかということです。

●仲 一つ目の制限するときの根拠は、利用資格に合致しているか、過去の利用状況はどうか、不正な使用方法をしていないか、寄託していただいている大切なデータを扱っていますので、それに対して誠実に扱ってもらっているかという点で検討しています。

二つ目の妥当性の判断に関しては、個別に一つ一つ検討しているのが実際のところですが、基本的には多くのデータを使って新たな学術的価値をつくっていただきたいと思っているので、どんどん公開したいのですが、制限をかけるときには利用条件や使い方の点で判断しています。その判断が割れるときには、上長に確認したり、あるいは内部の委員会にかけたりして、制

限をかけるときには慎重に対応しています。

●篠田 一つ目の制限公開と公開区分の基準の根拠については、材料分野はどちらかというとオープンしにくい状況があるのですが、その中でも論文発表や外部発表をして一般に公表したデータは公開リポジトリに入れやすく、また、外部発表をするときには機構の中での手続きもありますので、そういうところを公開の一つの根拠としています。

利用者のアクセス申請の妥当性については、例えばMatNaviですと、ユーザー登録をしていただいた方は承認されるという形になっています。申請は必要ですが、非常に制限が緩やかな制限公開という形です。あるいは、もう一つの例としてご紹介した共有データリポジトリなどは組織契約を基にしますので、相手方に内部で共有利用の責任を持っていただけるかなどの審査フローを経た上で、妥当性を確認することになる予定です。

●桂樹 一つ目の根拠については、個人に属するデータ、例えば腸内細菌の細菌叢を調査する場合には、これだけで個人のデータになってしまいますので、そういったもの。あるいは国として、または県などでもあるのですが、他の県には見せたくないといったデータは見せないという形で考えています。

二つ目のアクセス申請に対する判断は、一般的な回答になってしまいますが、活用目的によって個別に判断しています。ビジネス目的よりはアカデミックの方が緩くできるのかなという感じです。

●三橋 制限公開の根拠は、ヒトのゲノムデータ、病歴データ、投薬データなどは個人情報保護法や政令で基本的には個人情報となっていますので、制限公開にせざるを得ないという判断です。

申請の妥当性については、ヒトに関する研究に関しては各研究機関で倫理審査委員会が確立しており、そこで了解を頂かないといけないので、われわれとして

は基本的にはそれを確認するという仕組みでやっています。

●上島 企業のデータの扱いに関して言いますと、各組織共通の事情として、逆に公開することに根拠が求められます。今回ご登壇された皆さまは、皆さんが指摘されているような法的根拠や、コミュニティにおける貢献が重視されるところ、また各分野の研究動向や学術的な潮流を踏まえて、機関として公開していかないといけないという意味決定をされているかと思いますが、（企業でも）そのところの理由付けが必要です。取引先がやっているから、社会的要請として必要だから、あるいはパートナーシップを増やすために、例えば研究者に使っていただくためのデータを公開していくといった根拠づけが求められます。

区分の基準として、大きいところでは、やはり秘密保持契約のように、相手方との取引関係において情報が漏れない約束事があるかが焦点になりがちです。利用規約に同意されているか、実際の利用実態はどこまで調べていくのかということを、秘密保持契約にさらに追加していろいろ課していくことが一般的に多いと感じます。

●海老沢 JAXA 宇宙科学研究所の場合、データセンターが利用者の可否判断をすることはありません。公開されていないデータがたくさんありますが、衛星や探査機計画に参加した人たちが、1年程度、優先的に使うための非公開です。プロジェクトごとに、装置をつくった人、運用計画に参加した人、科学的アドバイスをした人などを集めてチームをつくっていて、チームの中で、こういう人はデータを使える、使えないという判断、1年間はデータを公開しない、などの判断をします。各プロジェクトやチームごとの判断であり、所として、あるいはデータセンターとして、誰が使っている、使ってはいけないという判断はしていません。誰がデータの優先権を得るかは、衛星ごとの研究チームによって決まります。

●池内 今、改めてお話をお伺いして、それぞれの事情があることがよく分かりました。データを公開するとなると、分野を越えて企業のデータを研究者などいろいろな人が使っていくときに、その分野の暗黙知のようなものを理解していない、あるいは知らない人が使うことになります。そのときに、それをどう伝えるかということも、その先の議論としては重要だと感じました。

●林 池内さんからコメントを頂いたのですが、例えば桂樹さんのご発言で、品種の情報はライバルである他県にはあまり見せたくないというのは農業分野では分かりやすい話ですが、他分野の方から見れば、別に県同士だからいいのではないかと思われるかもしれません。文化が違うという言葉でいいかどうか分かりませんが、分野によっても事情や意識は違うのではないかと思います。このあたりは Slido でも質問が来ていましたので、後で広げられればと思います。

●八塚 ありがとうございます。ここからは、Slido で頂いた質問を交えながら話をしていきたいと思います。

桂樹さんに対するご質問として、メタデータの一覧や、公開・非公開の表がスライドであったと思いますが、そのメタデータの定義にたどり着くまでにはどのようなディスカッションがあったのか。それから、カスタムメタデータのお話がありましたが、統一しにくいメタデータもあると思います。それをどのような方法で記録したり、管理したりすることを想定されているか。桂樹さん、いかがでしょうか。

●桂樹 メタデータの考え方ということだと思うのですが、最初、ゼロからどうやってメタデータを付けていくかという議論をした中で、世の中に存在している標準的なものをいろいろ勉強して、この辺かなということで、一番標準的な Dublin Core の項目を見たりしながら、農研機構ではどのようなデータを必要として

いるか、メタデータとして記録しているかを考えながら設計していきました。その中で、基本的に農研機構内でデータを整理することなので、どの予算を使ったか、どんなプロジェクトでやったかということを入れていくなど、そういうことを基準に考えています。

カスタムデータについては、分野横断的に、例えばドローンデータ、画像データを扱うプロジェクトが幾つか立ち上がって、その方たちと相談して、どんなメタデータを付けていけばいいか、検索したいとしたらどんなデータか、どういう項目が付いていたら便利かということをディスカッションしながら、それぞれの分野で必要なものを付けていこうという形です。そのデータを使っている人々、幾つかのグループに聞いて回って、では、この辺を付けましょうかということをしました。

●八塚 なるほど。では、カスタムと言いつつ、実はある程度は調停したというか、ここに使うということでは考えられていたわけですか。

●桂樹 データを後で検索するときに、なるべく統制をかけたいということがありまして、それでかけています。

●八塚 分かりました。ありがとうございます。

では、他のご質問も紹介したいと思います。他から取ってきた 1 次データを加工して、2 次データ、3 次データを作った場合、その共有化を義務付けることについてはどう思われるか。各領域では、そもそもどのようなルールがあるのか、ないのか。あるいは皆さん自身、規範的にどう考えられるかというご質問が来ています。これはお一人ずつ伺いたいと思います。

●仲 2 次データや 3 次データの共有化を義務とするかというのは判断に迷うところかと思います。ただ、学術目的で集められたデータは、いろいろな人に使わ

れて初めて価値が出てくると個人的に考えています。ですので、可能な限りデータの公開と共有をできるように、調査の段階で合意を取って、将来的には公開されるという流れを想定しています。例えば科研費などの公的な資金を得て調査をしている場合には、データそのものにはより公共的な価値があると考えられます。そういった場合には、2次データとして使えるようにしていく形がいいのではないかと考えています。

●篠田 これは大変難しい問題だと思っています。特に材料分野では、今、マテリアルズ・インフォマティクスや機械学習による材料設計のようなものがあり、そういうところに使うためにデータセットを作ったり加工したりすることがあります。基本的には元のデータがどういうライセンスをしているかに関わってくると思いますが、そういうデータセットを作るときに、ライセンス条件が違うものが混ざったときにはどうしたらいいのかということがあります。ですので、一律に共有化や公開がなかなかできないところだと思っています。

●桂樹 皆さんがおっしゃっているとおりだと思います。われわれの場合には、まだデータが集まってきていない部分もありますので、これから考えていくのですが、そんな中でも、取ったときのライセンスが同じプロジェクトの中でも違う、あるいは継続して次のプロジェクトに移って、今回のプロジェクトはこのように皆さんでデータを共有しましょうといったときに、うっかり前のものを一緒に枠組みの中で公開してしまいそうになったりします。時間軸も考えなければいけないところが難しいと思っています。

●三橋 ヒトのデータの場合は、研究参加者から同意をベースに試料を提供していただきますが、その場合には同意撤回という制度があります。同意撤回されてしまうと、データベースからそのデータを削除しないといけません。そうすると、あまり2次加工、3次加

工してトレースできないものが増えてしまうと困ります。基本的にはデータを使って論文発表ができないと意味がないので、論文発表で必要なデータは加工して載せていただくということまではある程度認めているのですが、それ以上の流通は実はあまり認めていないというのが実情です。

●上島 どこまでご参考になるか分かりませんが、コンテンツビジネスでは企業が2次利用やファンフィクションのためのガイドラインを公表することがあります。オープンデータに近いライセンス表示と、著作権表記（all rights reserved）による権利保持を併用する運用も見たことがあります。フェアユースの範囲をどう設定していくかという、著作物の扱いと近い枠組みで考えられると思っています。

実務的には、弊社の例でいうと、あるデータ作者が作ったデータを別の方が加工して公開したいといった場合に、出どころの表示をどう書くか、その方がどのような加工をしたかどこまで明記すればいいのかということが、細かいところですが論点になっています。そのあたりは、産業界、学術会、行政も含め、何となくの慣習ができてくると、心理的な抵抗感がなくなってくるのかなと感じています。

●海老沢 これは大事な課題だと思います。私たち宇宙科学研究所のデータポリシーは、論文を出したときのエビデンスデータは公開すべき、というものですが、実はそこまで至っていません。宇宙科学の分野では、幾つかのジャーナルは、必ずエビデンスデータを公開してURLを示さないと受け付けません。「Nature」「Science」も厳しいですし、American Geophysical Union (AGU) もそうになっています。しかし、JAXAも含め、国内の大学等にそういうデータリポジトリがないのが現状です。

一方で、海外では、誰でもデータを置いて公開できるリポジトリを運用しているところがあります。国内の研究者が海外の著名なジャーナルに論文を出すため

にはデータを公開しなければいけないので、そういう優れたデータがどんどん海外に置かれるという状況が見られ始めており、これはよろしくない状況だと言われています。これは国内で私たちが真剣に考えないといけない課題だと思います。それと関連して、今、NII が主導して JAIRO Cloud を始めています。そういうところが国内の研究者のデータ公開のリポジトリになっていけるのではないかと注目して見守っているところです。

●池内 今お話を伺っていて、2 次利用、3 次利用されていった場合に、トラッキングの問題はデータを出す側にもあるし、やはり利用する側にもあると感じました。孫データやひ孫データになると追いつけなくなってしまうという状況も出てくるでしょうし、もちろんコストも莫大になっていくと思います。使う側としても、そのデータが 1 次公開されたデータなのか、2 次データなのか、3 次データなのかという判断が非常に難しく、そこを気にするとデータの再利用がしづらい状況になりそうだと思います。上島さんがおっしゃったように、やはり文化の醸成が必要であることと、出典や加工の詳細をしっかりと書くという方向に行くのか、それとも、ある程度もういいだろうということにだんだんになっていくのかは、データの種類や分野の慣習によって異なると思いますが、今後の展開がどうなるのかなというコメントをさせていただきます。

●八塚 ありがとうございます。私自身はリポジトリを運営していて、もちろん 2 次データ、3 次データを作ってもらうためにリポジトリを公開していますので、そこは非常に意識しています。逆に、1 次データこそまさにオープンでないと、それを追跡するという話になると大変なコストがかかってしまいます。そういう意味では、1 次データにどのようなライセンスを付けるのか、どのような制限をかけるのかということは非常に大きな問題になってくると思います。林さん、何かありますでしょうか。

●林 そもそも他のデータを使ったときに、どのようにそのことを表現するのか。論文を書いて、引用文献として書く、あるいは謝辞という形で何かしらの貢献を示すという方法もあるかと思います。しかし、こんなデータを使ってすごく助かった、参考になった、あるいはデータの誤りがあった、ここのデータがもっとこうだったらいいという要望など、データを使ったときの表現や貢献を示すに当たっての慣習が皆さまの分野であればコメントを頂けますか。

●海老沢 他の分野でもそうかもしれませんが、私たち天文学・宇宙科学に関わる分野で、データアーカイブを作ると、みんなに使って喜んでもらえるのだけでも、それが自分の研究評価として評価されないという、若手の不満を聞くことがあります。そこで、データを作って DOI を付けて、データサイテーションが自分の業績になる仕組みができればいいという話があります。最近、宇宙科学の論文では、DOI があるときは必ず DOI を示しています。その DOI に行くと、そのデータを作ったのは誰かということが分かって、データを作ることが評価される。これが、データサイテーションが進んでいくということになるのではないかと思います。

●八塚 ありがとうございます。まさに DOI を使ったデータサイテーションは、RDF でも専門の小委員会がありますけれども、本当にこれから重要になってくると思います。

●仲 海老沢先生がおっしゃったように、成果物に対して、引用や謝辞など、DOI をフォローするようにしています。特に論文の中で、データの利用に対する謝辞の記載があるかどうかを確認しています。基本的には利用が終わった段階で成果物を登録していただくのですが、その際に論文を現物あるいは PDF で送っていただいて、一つずつ謝辞を確認するという形を取っています。

もう一つは、一連のデータを寄託していただいたり、寄託いただいたデータの利用件数が多い場合には、年に一度、該当の寄託者様を表彰する仕組みをつくっています。SSJDA を利用する側と寄託する側の双方がよりよく結びつくように取り組んでいます。

●**桂樹** われわれはデータがまだ全然集まっていない段階ではあるのですが、これからデータを作った人をきちんと評価できるように、まさにデータごとに DOI が取れる仕組みをつくっていかうとしてるところです。それでデータを作る人のモチベーションも上げていきたいと考えています。

●**三橋** 塩基配列のデータベースに関しては、公的 DB に登録するとアクセッション番号という番号が発行され、それを論文に書くと、そのデータを出した人が評価されるという仕組みになっています。あとは、ヒトデータベースの利用者のガイドラインに必ず年に 1 回、利用報告を出すように決めており、その中で、論文発表をしたらそれを報告するようという決まりをつくっています。

●**八塚** ありがとうございます。大体この話題については出尽くしたかと思うので、次のご質問にいききたいと思います。

これも組織の中の話になるかと思いますが、データ公開において、例えば知財の部署との相談はされていると思いますが、それをどのように進めているのでしょうか。個別に相談すると件数が多くなったり、時間がかかったりする場合があります。かといって、知財部署との調整なしでやるわけにもいかないと思います。この辺で、こういう苦労をしたとか、こういう方法がありますという方はいらっしゃいますか。

●**海老沢** 私たち宇宙科学研究所のデータポリシーの中で、商業利用も OK というポリシーにするときに、JAXA としてそれが許されるのかどうか気がになりま

した。そこで JAXA の広報部や新事業促進部などの意見を伺ったのですが、特に問題はありませんでした。科学データに関しては、知財などの問題はなく、商業利用目的ということで公開できました。意外とみんな知財や商業利用に関して臆病というか、びくびくしているような感じがして、もっとオープンにデータを公開しても良いのでは、という印象を持ちました。一方、地球観測などに関しては、宇宙観測と違って、知財などの難しい問題も出てくるのかなという印象も持っています。

●**上島** どこまで参考になるか分かりませんが、私の関わった事例ですと、組織内の関連する部門——広報、法務、知財あるいは情報システムからご担当にお集まりいただいてタスクフォースを結成して、そこで定期的に会合を持つ形で、寄せられてきた案件をさばいたり、集中審議をかけたりする。組織としてそういう枠組みをつくっているところはございました。

●**八塚** ありがとうございます。やはり個別に処理するというよりも、ポリシーや仕組みをつくるのがとても大事だと思います。

それでは、次のご質問にいききたいと思います。実際的な話ですが、どのくらいのスタッフで運用しているのか、それから、どんな分野のスタッフで構成されているのかというご質問です。これは実際にデータを提供されているところということで、海老沢さんからお話しいただけますか。

●**海老沢** 私たちの研究所では、データを作る立場と公開する立場を分けていて、基本的にデータを作るのは衛星などの「プロジェクト」です。プロジェクトは、どこかでスタートして、どこかで終了します。しかしデータセンターは定常組織なので終わりがなく、ずっと継続するわけです。私たちのデータセンターは非常に小さくて、私たちの部署で DARTS というデータアーカイブに関わっているのは数人の正職員のほとんど

が兼任、専任の派遣職員が数人、トータル 10 人足らずです。一方で、データを作る方は集中的に短期間で、例えば 1 年間に 10 人で、10 年の間に延べ 100 人ぐらいという形になります。データを作る側とデータを維持する側が違うというところがポイントだと思います。

●三橋 われわれの場合、DDBJ と NBDC が共同で運営しているのですが、それぞれに少なくとも 1 人ずつ、バイオロジーで Ph.D.を取って、なおかつ倫理等のことをきちんと勉強している、ほぼ専任のコアなメンバーがいます。それにプラスして、他をサポートする私のような IT 系の人間や、外注などのスタッフで何とか動かしているという形です。

●桂樹 私はデータ戦略推進室という部署なのですが、上長が 1 人いて、Ph.D.を持つ研究員が 3 人、システムをやる人が 2 人、SE が 1 人です。すごくこじんまりとした状況で、あらゆるデータを集めてこいと言われて四苦八苦しているところです。倍ぐらいの人がいたらいいのにと思うことがよくあります。

●篠田 私の所属する材料データプラットフォームセンターは、材料プラットフォームを構築するというミッションの下で動いていますので、非常に大所帯で、正確には把握していないのですが、あるときには 60 人ぐらいいると聞いていました。その中でも、基盤の構築とサービス提供を両方していますので、構築の方でシステムエンジニアなどがかなり多くいます。ただ、要素技術の開発も行っていますので、研究者も参加しており、実際に材料分野の研究者でセンターでサービス構築を進めている方もいます。

あとは、サービスを提供する側として、私どもは併任も多いのですが、サービスチームというものがあります。こちらは 14~15 人のメンバーがいて、構築されて運用が始まったものをユーザーに提供していくところを担当しています。

●仲 東京大学社会科学研究所附属社会調査・データアーカイブ研究センターでは、任期付きの教職員を含めて 15 人程度で運用しています。その中でも、今日ご紹介したデータ提供のプロセスでは、実質的には 2 人で対応しています。もちろんデータを作る部門にはデータアーカイブのマネジャーがいて、短時間で働いてくださっている方も数人います。思ったよりも小さい組織で運用しています。

●林 先ほどの桂樹さんのお話の中で、例えばメタデータを作るのに苦労されているということがありました。また、篠田さんのご発表の最後にも関連しますが、スタッフの中にライブラリアンというか、メタデータの扱いが得意な図書館の人が入っていると、お手伝いになるのではないかと思います。

●八塚 ありがとうございます。では、他の質問にきたいと思います。

先ほど、2 次データや 3 次データも含めたデータの利用について話題が出ましたが、それを実際どれくらいチェックされているのか。例えば、利用受託を守っているか、引用と謝辞を行っているかということは先ほどのコメントで少しあったと思うのですが、他にこういう調査をしているということがありましたら、どなたかコメントを頂けますか。

●篠田 直接の利用については、やはりデータに DOI を振って引用状況を見たり、謝辞は確認が難しいので、皆さんから「こういうものを載せたよ」という報告を頂いて情報収集する形を取っています。2 次利用がされて、それが成果につながったかは別にして、直接的に言えるデータのダウンロード件数やページへの訪問者数を一つの指標として、データの利活用状況などをモニターしている状況です。

●海老沢 私たちは衛星プロジェクトごとに、どれだけその衛星データが使われたかということでカウント

しています。たくさんの論文が出れば、良い衛星であり、良い成果ということです。一方で、データセンターの貢献と衛星の貢献の区別は難しいのです。私たちデータセンターは、いろいろなデータを集めて使いやすくしてデータ利用を促進しています。DARTS というデータアーカイブをしているのですが、どの部分で貢献したかを判断するのが難しく、データセンターとしての貢献を測るのは難しいです。衛星のデータは、宇宙科学研究所からダウンロードされることもあるし、例えば NASA から同じデータがダウンロードされることもあるのです。ですから、データセンターの貢献を評価してもらう指標は、なかなか難しいです。

●八塚 ありがとうございます。それでは、ここでいったん質疑から離れまして、ディスカッションの論点に戻ります。

制限公開の今後がどうなっていくのかということについて、皆さんのご意見を伺いたいと思います。例えば、分野を越えた共通のフレームワークが必要になるのか。あるいは、既に商用もある程度あると思います。そういった学術以外の利用が進んでいくのかどうか。それから、完全公開のデータに比べると、当然、制限公開の場合はコストがかかるのですが、そのコスト負担は一体誰がすべきなのか。これは池内さんからお願いできますか。

●池内 まずコストの話からすると、やはり現時点でもトラッキングのコストや判断のコストはすごく大きいと思いました。一方で、DOI で追っていくのは、技術的に解決できる部分が今後出てきそうだと思いますので、まずは DOI を振るという慣習を根付かせることが、すごく遠回りみたいですけれども、やはり必要だという気持ちを新たにしたところです。各データアーカイブの機関においても、登壇されている皆さまは十分されているかと思うのですが、そういうことが大事だということを改めて感じました。

制限公開に関しては、できるだけ安全側に倒して制

限しているという部分もあるかと思うのです。例えば公開するという意味では、論文を公開するということについては、私たち研究者は抵抗感が少ないと思います。論文であっても剽窃されてしまう可能性はあります。つまり、データも論文もそんなに変わらない部分もあるはずですが、やはりデータに関しては、まだまだ論文と比べると公開している歴史が短い分、恐れてしまう部分はあるのかと思います。

私も個人的にもどきどきしながら公開するので、そういった気持ちも十分理解できますが、実はそんなに恐れることはないかもしれないという文化ができていたり、公開したことでこんなに良いことがあったというベストプラクティスが共有されたりしていくと、だんだん制限公開に対する考え方も変わってくるのかもしれないと思いました。分野特化ではなくて全般的な話なので、ふわっとしていますが、そのように考えました。

●海老沢 私が専攻している天文学、宇宙科学は、基礎科学の中でも特にオープンな分野だと思います。宇宙科学のデータを公開しない、地上天文台のデータを取って公開しないということは、ほとんどなくなってきました。逆に公開しなくて悪かった例はたくさんあり、過去に天文学の分野で間違った論文がたくさん出ていますが、それらはほぼ 100%非公開データから出ています。データを公開しないと信頼されないのです。

これは実は恥ずかしいのですが、私の研究所で、以前、あるロケットを打ち上げて大きな成果を上げて、大きな物理学の賞を取った人がいます。この研究所の所長にまでなりました。一方では、他の大学で、地上の望遠鏡である発見をして同じ賞を取った人がいます。どちらの発見も、今では否定されています。それらのデータは公開されていませんでした。これも名前を出すすゝと血が流れるかもしれませんが、現在でも、別の地上の望遠鏡でずっと何十年も観測を続け、大発見をしたと言っている先生がいます。しかし、データを公開

しないのです。データを公開しないと信頼されないという流れは、はっきりしています。

私たちの分野で課題になっているのは、人工衛星をつくった人たちのデータの優先権をどのくらいにするかです。大体1年が目安ですが、それをもっと短くした方がいいのではないかと、あるいは優先権は全くなくてもいいのではないかとという議論もあります。

いずれにしろ、宇宙科学の分野に限って言えば、データ公開の流れは止まりません。過去にはデータを公開せずに間違った結果が報告されたという科学的な損失もありました。現在も残念ながらあります。

●上島 産業界から見て、アカデミズムの引用のやり方や、DOI がきちんと付されていてそれを業績としてカウントしているということは、まだまだ知られていないところがあります。世の中的にもう少し知られていくといいのだろうと感じます。

また、コストをどう負担するかについては恐らく二つしかなくて、一つは、コストに見合う収益をどう稼ぐか、もう一つは、コストをいかにスリム化するかです。コストのスリム化に関して言うと、ある程度の利用手数料を頂いたり、何らかの有償化する方策が、(研究活動の) 持続可能性を考えると必要だろうと、皆さんのお話を伺っていて感じました。収益化に関して言うと、企業のデータであれば、製品化して商売として売りましょう、値段を付けましょうということで解決するのですが、オープンなデータを研究成果として公開するとなると、なかなか難しいと重ねて感じます。

今後の安心感をどう醸成していくかについては、リスクをどう考えるかに尽きると思います。リスクにもコントロール可能なリスクとコントロール不可能なリスクがあります。コントロール可能なリスクは、審査制など、マニュアル(人的) 対応の中で皆さんコントロールしようとされていて、企業の取り組みともかなり重なるところがあります。コントロール不可能なリスクは、恐らくコミュニティとして負うことは難しいので、抽象的な言い方になるのですが、何らかの「保

険」のような仕組みが学界でも必要ではないかと感じました。

●三橋 提供者が獲得した研究費でデータを集めて研究論文を書くまではコスト負担者と受益者が一致しています。そうすると、論文を書くまでは研究者がある程度、独占的に使ってもいいけれども、論文を書いたら、やはり公開してもらおう。その公開のコストに関しては、今もそうですが、提供者ではなく、われわれのデータベースセンターで負担するのがいいのではないかと考えています。

●桂樹 農研機構では、農業データ連携基盤(WAGRI)に参加しています。データのアグリゲーションとマーケット、集めてくると売っていくことができる仕組みを目指してそういうものを構築しているのですが、それが、データを出すに当たってコストをどうすればいいかという一つの解決策になればと思っています。仕組みとしてはB to B to Cと呼ばれているもので、一つ目のBの人がデータを提供して、そのデータを真ん中のBの人がビジネスにできるような仕組みにします。そしてCの方々に提供する。例えば農薬を買ってもらうことでお金を集めて、データを提供することに対するコストを負担してもらおう。そういう仕組みをつくっていかうとしています。

●篠田 材料分野の方は公開するよりも非公開、制限公開の方に偏りがちで、オープンにしにくいということがあるので、どのようにオープンの方に寄せていくか、戦略的にオープンにしていくかが課題になるのですが、やはり懸念事項として、デジタルデータは一度公開して拡散してしまうと、もうそれを非公開にするのは原理的に不可能ということがあり、オープンにするところはかなり心理的ハードルがあり、慎重にならざるを得ません。こういったところは少しずつ事例を積み上げていく必要があります。一方で、材料分野の中でも、データによってはオープンにして活用しても

raitaiというニーズもあり、そのような研究者やデータの種類もあります。そういう事例を積み上げながら、制限公開と公開をうまく使い分けていければよいと思っています。

コスト負担については、NIMS は基盤なども提供していますが、国研のミッションとしてこういうことをしていますので、お金を頂かずに皆さんの税金でちゃんと還元できるところに関しては、まずは組織として運用を進めていく。ただ、一方でそれが商用利用などにつながったときには、また別の考え方も必要になってくるかと思っています。繰り返しになってしまうのですが、こういうところは事例を積み上げながらいろいろと練習問題を解いていっている状況です。

●仲 私からは、一人の研究者としての立場と、組織に所属する立場から簡単にリプライしたいと思います。

まず、研究者の立場としては、やはり制限公開であったとしても、利用できる多様なデータがあり、それにいろいろな人が相互に乗り入れるプラットフォームがあるというのは、研究においては本当に大事なことでと思っています。特に1次調査は費用負担が非常にかかってきますので、大型の研究プロジェクトに関わっていない方や学部学生、院生を含めていろいろな方が無償でデータを使えるのは、研究のアイデアで勝負できるという点で可能性が広がると思っています。ですので、どんどん公開できればと思っています。

その一方で、組織としてはデータアーカイブセンターが永続的に続いていかないと、データが散逸してしまうという問題が生じてしまいます。そういう意味では、何らかの形で収益を一定程度稼げる状態をどこかでつくっていくことを視野に入れる必要があると考えています。現時点では完全にフリーで、お金を頂かずにやっています。セミナーに関しては一部費用を頂いていますが、それは講師料などで、データの提供は無償で行っています。

●八塚 ありがとうございます。それでは、最後に林

さんにまとめていただけないでしょうか。

●林 練習問題を懸命に解いているフェーズにある分野がすごく多いのだろーということ、皆さまのご発表や今のディスカッションを通じて感じました。特に、使ったことを例えば DOI で表現するなどで示す、あるいは何か違う方法があるのかもしれませんが、まずはデータに関しても貢献を示していく。巨人の肩に乗るというのが研究だと思いますので、論文と同じように、データについてもきちんと貢献を示すことができるようにする。その上で、公開に当たって一定期間制限があるのか、あるいは最初からオープンになるのか、その最終的なインセンティブや必要性、どうしても公開できないけれどもいずれ公開になるかもしれない、あるいは、これまでは公開に制限なり限定がされていたけれども、もう少ししたら公開が当たり前になるかもしれない。そういう発展もあるのかなということを皆さまとの議論を通じて考えました。

気の利いたまとめになっていないかもしれませんが、こんな感じでどうでしょうか。

●八塚 ありがとうございます。これでパネルディスカッションを終わりたいと思います。パネリストの皆さま、本当にありがとうございました。

第2回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

開会挨拶 / 概要説明

矢吹 命大

(横浜国立大学大学戦略情報分析室)

矢吹 命大



2020年度SPARC Japanセミナー企画ワーキングメンバー。横浜国立大学大学戦略情報分析室准教授/研究推進機構リサーチアドミニストレーター。筑波大学大学院人文社会科学部国際政治経済学専攻単位取得退学、修士（国際政治経済学）。筑波大学大学院人文社会系特任研究員として巨大大学を巡る国家間関係の研究に従事した後、2014年横浜国立大学特任教員（講師）・URA。2020年4月より現職。大学戦略情報分析室において大学経営判断を支援する各種情報収集、提言を行うと共に、URAとしては研究力分析、オープンサイエンスの推進に取り組んでいる。



セミナーの企画概要

昨今の COVID-19 を契機としてプレプリントの活用がますます活発化していることについては、皆さまお聞き及びのことかと思えます。それに加えて、プレプリントが多様なプラットフォームにおいて公開されるという状況にあり、もはや学術出版社においてもプレプリントの公開を後押しするといった状況にあります。本セミナーでは、多様なプレプリントの公開について、最新動向やその目的を皆さまと共有し、今後のプレプリントの方向性を展望します。そのため、講演では多様な立場からのプレプリントに関する話題と事例を紹介します。

冒頭では、国立情報学研究所の河合先生から「機関リポジトリによるプレプリント公開」と題して、機関リポジトリとプレプリントの関係について、また、パネルディスカッションでも中心となる Bibliodiversity（書誌多様性）に関してお話しいたします。

続いて、筑波大学の森本様から「研究成果公開のグローバルスタンダードに向けた筑波大学の取り組み」と題して、人文社会系評価指標のお話と、最近始まっ

た、オープンピアレビューから出版までのフローが含まれる F1000Research を利用する筑波大学ゲートウェイのご紹介を頂きます。

続いて、シュプリング・ネイチャーのアントワーン・ブーケ様より、「コロナ時代における研究情報発信—プレプリントに関する出版社の取り組み」と題して、学術出版社におけるプレプリントについて話題提供を頂きます。

後半では、広島大学の坊農先生から「生命科学研究におけるプレプリントやSNS活用の現状と課題」と題して、研究現場におけるプレプリントの利用と課題、アカデミック SNS の活用についてご紹介いただきます。

最後に京都大学の引原先生から、「Preprint が誘導する研究サイクルの力学考」と題して、大きな観点から研究におけるプレプリントについてお話を頂きます。

パネルディスカッションにおきましては、最初の講演者の河合先生が監訳されたオープンアクセスリポジトリ連合による論文「Fostering Bibliodiversity in Scholarly Communications—A Call for Action!」（邦訳：「学術情報流通における『書誌多様性』の形成に向けて—行動の

呼びかけ」)に示されている学術情報流通における多様性の障壁とされた4項目、すなわち、①共通言語としての英語が優位であること、②基盤とサービスが一部に集中していること、③資金モデルが限定的であること、④学術雑誌ベースの評価という偏狭な視点があることに沿って議論します。

第2回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

機関リポジトリによるプレプリント公開

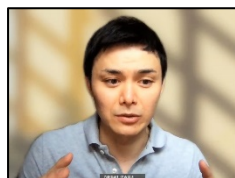
河合 将志

(国立情報学研究所 / オープンサイエンス基盤研究センター)

講演要旨



本報告では、機関リポジトリによるプレプリント公開の可能性について検討する。具体的には、プレプリントサーバーとしての機能が求められるようになった背景や、実装にあたって見込まれる問題などについて論じる。加えて、機関リポジトリのこうした機能変化が、学術情報流通における書誌多様性に与える影響についても言及する。



河合 将志

大阪大学大学院国際公共政策研究科博士後期課程修了。オープンサイエンス基盤研究センターではデータ解析を担当しており、システムの利用統計処理などに携わっている。

報告内容

本報告では、近年、プレプリントサーバーへの関心が高まっていることを受けて、機関リポジトリでプレプリントの公開が可能なのかという点について検討します。まず、機関リポジトリとプレプリントサーバーの違いについて確認した上で、機関リポジトリによるプレプリント公開の可能性について検討します。また、プレプリントと書誌多様性の関係が本日のパネルディスカッションのテーマですので、機関リポジトリによるプレプリント公開が可能になった場合における書誌多様性への影響についても最後に言及します。なお、書誌多様性という概念は、単純に学術情報流通の各側面における多様性ということかと思えます。

機関リポジトリとプレプリントサーバーの違い

まず、機関リポジトリとプレプリントサーバーの位置付けを確認します。

公開基盤であるオープンアクセス (OA) リポジトリの中の大きいカテゴリーの一つに、機関単位で運営

される機関リポジトリがあります (図 1)。国内の機関リポジトリとしては JAIRO Cloud や紅が挙げられます。もう一つの大きいカテゴリーとしては、個別の研究分野ごとに運用される分野リポジトリがあります。例えば PubMed Central (PMC) などです。この分野リポジトリのうち、プレプリントに特化したものがプレプリントサーバーです。arXiv や Social Science Research Network (SSRN) が例として挙げられます。

次に、機関リポジトリとプレプリントサーバーのコンテンツの違いを、学術雑誌論文を例に見ていきます (図 2)。まず、機関リポジトリの場合は、出版社に



(図 1)

よる体裁調整がなされた出版社版がコンテンツになる場合もありますが、出版社版の一つ前のバージョンで、査読はなされているものの体裁調整がなされていないポストプリントが主なコンテンツです。

一方のプレプリントサーバーの場合は、ポストプリントのさらに一つ前のバージョンで、査読も体裁調整もなされていないプレプリントがコンテンツです。

次に、プレプリントサーバーが着目される背景を把握するためにも、機関リポジトリのデメリットについて確認します(図3)。機関リポジトリのコンテンツは査読がなされたポストプリントなので、機関リポジトリが公開の迅速性や更新性に関して優れているというわけではありません。従って、研究者から見た場合には、査読が済んで掲載が決定しているのに機関リポジトリにわざわざ登録するメリットは何なのだということになるかと思います。

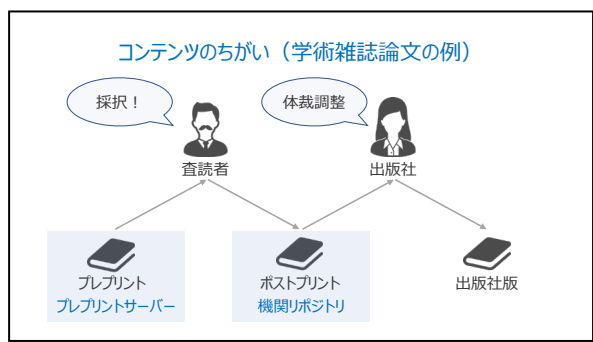
研究者が機関リポジトリにメリットを見いだせていないということは、私が行っている解析の結果にも表れています。図4は機関リポジトリの学術雑誌論文コンテンツを増やす上での各活動の効果の推定値を示し

たものです。縦軸は各活動、横軸はコンテンツを増やす上での活動の効果を表しています。図書館が研究者にポストプリントの提供を依頼する学術雑誌論文提供依頼や、登録されたコンテンツのダウンロード件数を通知するダウンロード件数通知の効果が高い一方で、研究者自身がコンテンツを機関リポジトリに登録するセルフアーカイブの効果は非常に低いことが示されています。その理由としては、研究者からのセルフアーカイブの要請がないためにそもそも実施していないことが多いようです。つまり、研究者は図書館から頼まれたらポストプリントを提供するけれども、セルフアーカイブをするほどの魅力を機関リポジトリには見いだせていないということです。

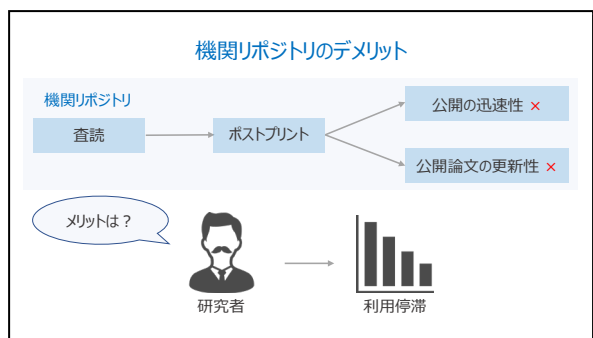
一方で、プレプリントサーバーのコンテンツは査読がなされていないプレプリントなので、プレプリントサーバーは公開の迅速性や更新性に優れていると言えます。従って、研究者からプレプリントサーバーを見た場合には、先取権を確立するためのタイムスタンプを気軽に押せる魅力的な基盤に見え、そのことが利用評価につながっていると思います。

以上のように、プレプリントサーバーのメリットは、プレプリントサーバーそのものの性能というよりは、査読がなされていないというプレプリントの性質に由来していると言えます。

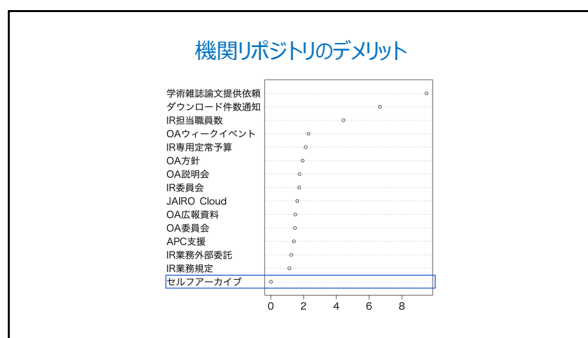
そう考えたときに、プレプリントサーバー以外のOAリポジトリによってプレプリントを公開する可能性はないのかという疑問が当然ながら生じてきます。わざわざプレプリントサーバーによって公開するのではなく、日本において充実した整備状況にある機関リ



(図2)



(図3)



(図4)

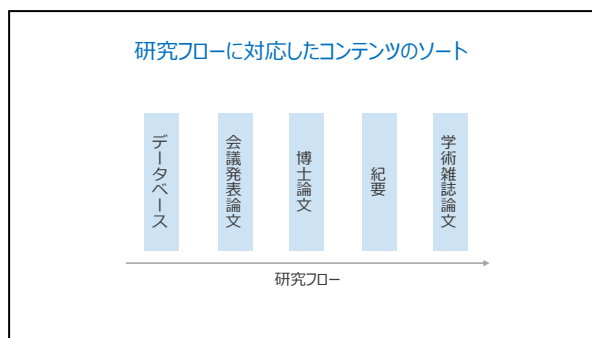
ポジトリによってプレプリントを公開することが可能かどうか、検討する必要があると思っています。より端的に言えば、既にある機関リポジトリを活用すれば済むのではないかという疑問が生じてくるわけです。

機関リポジトリによるプレプリント公開の可能性

疑問点を確認したところで、機関リポジトリによるプレプリント公開の可能性を検討します。

日本において機関リポジトリの数はリニアに増えており、現在 800 以上の機関リポジトリがあります。この数は世界的に見ても非常に多いものとなっています。そのコンテンツとしては、紀要論文が最も多く 54%で、次いで学術雑誌論文が 14%となっています。また、博士論文や会議発表論文、データベースなどもあり、多様なコンテンツが登録されています。

これらのコンテンツのうち、主要な幾つかを研究フローに対応した形でソートしてみます（図 5）。当然、分野によって順序は変わってきますが、大体このようにソートできるかと思います。これを眺めると、研究フローをおおよそカバーするコンテンツが登録されていることが分かります。このソートしたコンテンツの並びにプレプリントを落とすと、データベースと会議発表論文の間あたりになるかと思います。ここで言いたいのは、機関リポジトリの既存コンテンツの中にプレプリントを位置付けることに不自然さが無いということです。機関リポジトリは単に整備状況がいいだけではなく、プレプリントとの相性もいいので、機関リポジトリによるプレプリントの公開は可能なのではな



(図 5)

いかと考えています。

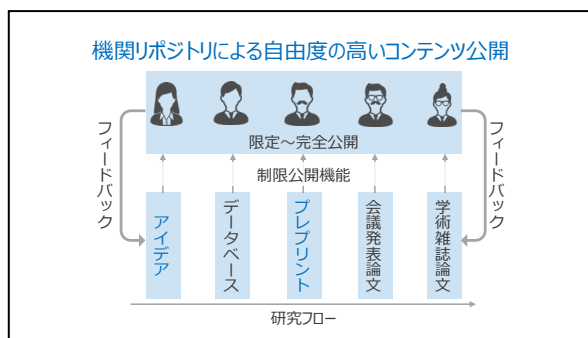
少し話がそれるかもしれませんが、より将来的な話としては、プレプリントだけではなく、研究フローをさかのぼって研究のアイデアなどもコンテンツ化し、なおかつ、公開後に SNS のようにフィードバックも得られるようにすることが考えられます（図 6）。もちろんフィードバックは欲しいけれども、コンテンツによっては、まだアイデア段階なので完全公開まではしたくないというニーズもあると思うので、公開の範囲を限定できる制限公開機能なども併せて使うことが考えられます。この制限公開機能については、国立情報学研究所（NII）の朝岡先生が現在開発中です。

研究者は機関リポジトリにメリットを感じていない現状にありますが、プレプリントに限らず、望むコンテンツを望む範囲で公開でき、なおかつフィードバックも得られるとなれば、プレプリントサーバー以上のメリットを機関リポジトリに感じられるようになると思います。

書誌多様性への影響

最後に、機関リポジトリによるプレプリント公開や、その発展版である、より自由度の高いコンテンツ公開が可能になった場合の書誌多様性への影響にも言及したいと思います。

先ほどのような自由度の高いコンテンツ公開を可能とする基盤があると、各研究者は公開コンテンツと公開範囲を自由に選択できるようになります。その選択が堆積していくと、結果として書誌多様性が生まれる確率が高くなります。ただ、多様性がないということ



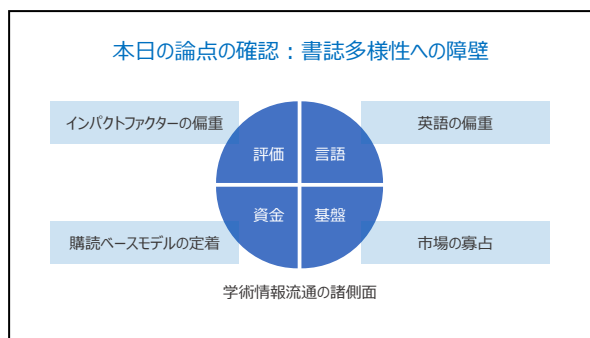
(図 6)

は、例えばルールが明確であるとも言えますし、多様性がある方が良いということが必ずしも自明なわけではありません。従って、そこが不明確なまま、多様性の構築そのものが目的化しないように注意する必要があります。

本日の論点の確認：書誌多様性への障壁

本日のパネルディスカッションの論点である書誌多様性への障壁について簡単に確認しておきたいと思います（図 7）。学術情報流通の諸側面として「言語」「基盤」「資金」「評価」があるとして、言語面では英語の偏重、基盤面では市場の寡占、資金面では購読ベースモデルの定着、評価面ではインパクトファクターの偏重という問題があり、各側面において多様性が失われていることがレポートで指摘されています。パネルディスカッションでは、それが良いことなのか悪いことなのかという根本的な点も含めて議論していくことになると思っています。

●矢吹 プレプリントがリポジトリに入ってくるとなった場合には、その後のポストプリントもある、掲載される論文もあるということになります。多様なバージョンが全て1カ所に入ってくる場合、見ている側にとっては、これはどうするのだという感じにならないかと思いました。限定公開のような仕組みに走ればうまく切り分けられるかと思いつつも、そういうことがあるのではないかと思います。



(図 7)

●河合 例えば、プレプリントを載せた後にポストプリントを載せたら上書きされるという形になれば、紛らわしくないかと思います。

●矢吹 同じ趣旨の質問が来ています。「各機関の方針やシステムによると思いますが、機関リポジトリで、プレプリント版や、査読を経て学術雑誌等で公開された版といった複数のバージョンが公開される場合、バージョン管理にはどのような方法が考えられるでしょうか」。

●河合 最新のもののみを公開したいのか、それまでの変遷を公開していきたいのかも、ユーザーである研究者が選べるようにしていくべきだと考えています。

●矢吹 もう一つ質問が来ています。「内容の質保証についてはどのようにお考えでしょうか」。プレプリントは査読前のものということもあり、そういったものがリポジトリに入ってくるときにどう考えるかという意図でしょうか。

●河合 プレプリントである以上、その公開には研究者のための公開という側面があると思います。フィードバックのための公開といった側面です。

●矢吹 次は大きな質問です。「プレプリントは主にはどの学術分野で行われているのでしょうか」。ご存じの範囲でお願いします。

●河合 どこで特に盛んかという点については、把握できていません。

●矢吹 この後の講演でも幾つかの分野についてのお話が出てきますので、そこでもこの質問の答えが出てくるかもしれません。

もう一つ、これもまた難しい質問です。「機関リポジトリにおけるアイデアの公開は著作権で保護されま

せんが、研究活性化とアイデアの保護のバランスはどのように取っていくことができるでしょうか。

●河合 フィードバックが欲しい一方で、アイデアが取られてしまう可能性は当然あると思いますが、そこはあくまでもユーザーの責任、ユーザーの判断で行ってもらわなければならないと思っています。

●矢吹 次は NII 船守先生からです。「研究者がプレプリントを機関リポジトリに登録するインセンティブが必要ですが、例えば Plan S の権利保持戦略を日本の研究助成機関に採択してもらい、かつプレプリントを公開することを義務化するというのはどうでしょうか。日本の研究のビジビリティの向上にもつながると思います」。

●河合 そういった規定があった場合には登録が加速すると思います。また、ユーザーから見ると、例えば国のリポジトリのような感じで一つのブランドに見えるような仕組みをつくらないと、モチベーションが上がらないと思います。

●矢吹 実は私も、リポジトリは機関別に整備されていて、一方でプレプリントのサーバーは分野別にあるので、その辺の関係はどうなるのかと思っていたのですが、今のお話がそれに当たると思いました。

質問がもう一つ来ました。「機関リポジトリは機関ごとにつくられたものですが、研究者が分野内でコミュニケーションを取れることの方が大事ではないでしょうか。分野ごとに機関リポジトリをグルーピングするような構想はありますか」。

●河合 先ほどの話ともつながると思いますが、ユーザーから見た場合に、そういったくくりがあるように見えるようにしなければいけないと思っています。

●矢吹 深貝先生からコメントです。「休憩時間の前

にでも、フォスタリング文書について皆さんの翻訳をご紹介いただければと思います。パネルに入る前に、関心のある方は休憩時間などにこれにアクセスできるでしょうから」ということです。先ほど翻訳していた文書と元の文書のリンクは、後でチャット欄等を通じてシェアさせていただければと思います。

もう一つ質問が来ました。「リポジトリの登録へのインセンティブとして、登録すると世界的なデータベースへの取り込みも自動的にされるということがあるように思いますが、どうでしょうか」。

●河合 今のところ具体的なイメージは持っていませんが、今後は考えていかなければならないと思っています。

第 2 回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

研究成果公開のグローバルスタンダードに向けた 筑波大学の取り組み

森本 行人

(筑波大学 URA 研究戦略推進室)

講演要旨



人社会系研究の成果発信、そして社会との関連づけがより強く求められている中で、筑波大学では国内の人社系の研究成果やそのクオリティを、日本のみならず世界に示すことを目指し、新たな評価指標や研究成果の発信について議論を続けてきました。その成果として学術誌の多様性をはかる iMD (index for Measuring Diversity) を開発し、さらに新たな学術情報の発信手法として 11 月末にオープンを予定している筑波大学ゲートウェイを用いて、より多くの方に人文知を届けられる仕組みを構築しています。



森本 行人

京都生まれ。関西大学大学院経済学研究科にて博士（経済学）の学位を取得。関西大学URAを経て、2013年度より筑波大学本部URA。2017年に人文社会分野から特許出願（特願2017-138751）、2018年にはURA業務の一環として、科研費の奨励研究獲得。

背景

私は筑波大学ゲートウェイの広報を主に担当している立場から、F1000 筑波大学ゲートウェイの話というお題を頂いたのですが、まず筑波大学ゲートウェイ導入に至る経緯を説明します。

The Times Higher Education (THE)、Quacquarelli Symonds (QS) などの世界大学ランキングの評価項目の中に、論文の被引用数に基づく評価項目というものがあります。これは Scopus 等の論文・引用データベースが元となっていますが、収録誌の多くは英語論文であり、それ以外の言語で書かれた論文はほとんど収録されていません。

さらに、現在用いられている評価指標として、Impact Factor (IF) があります。1 論文当たりの平均被引用数に相当するもので、Web of Science に収録され

ている収録雑誌の 3 年分のデータを用いて計算されています。また、Scimago Journal Rank (SJR) や Eigenfactor は引用という側面から出発した指標です。ただ、引用という行為にポジティブな評価を認めるという前提には、分野によっては若干の温度差があると思います。

では、人文社会系分野の学術誌はどうかというと、2 バイトの文字という特性があるので、論文・引用データベースに収録されにくいのです。論文が日本語で書かれており、稿末に引用文献一覧を付けない論文も少なくありません。また、引用数の集計が大変しづらく、従来の評価指標では見えにくいということで、評価の対象となってきました。

そこで筑波大学は、新たな研究評価指標の開発に着手しました。それが 2013 年、今からおおよそ 7 年前の

ことです。本学の教員と URA でスペシャルチームを組んで取り組み、さらに京都大学学術研究支援室（KURA）や、大阪大学の経営企画オフィス／URA 部門の方々とも調査研究を行いました。また、矢吹先生も含めて全国の人文社会系の URA の方や、国立情報学研究所の先生方、科学技術振興機構（JST）の方にもヒアリングをさせていただくなど試行錯誤を繰り返し、多くの方からご意見を頂きました。そして本学で index for Measuring Diversity（iMD）を開発し、特許を出願しました。さらに評価も考えた上で、人文社会系の研究成果をどのように公開しようかと考えていたところ、筑波大学ゲートウェイを公開するに至りました。

人文社会系分野の研究成果指標としての iMD

まず、学内紀要よりは全国学会誌、全国学会誌よりは国際学会誌という評価には、一般論として多くの研究者の同意が得られるのではないかと思います（図1）。これは、より多くの人がその雑誌を読んでいるであろうという考えに立った評価です。では、これを定量化するにはどうすればいいのかということで、学術誌の著者の所属機関の多様性を定量化しました。これは、質は定量化できないということが大前提にあります。IF も質ではなく話題性を定量化したものですし、iMD も質ではなく多様性を定量化したものです。繰り返になりますが、より多くの国や研究機関の研究者から論文が発表されているジャーナルほど、多様な著者に評価されているジャーナルであることが言えるの

ではないかと思います。特定のジャーナルに1年間で発表された全論文の著者の、所属機関とその立地国の数を基にスコアを算出したものが iMD です。

IF には、長期間引用される分野と、短期間しか引用されない分野がありますが、スコア算出の根拠となる被引用数のデータの透明性が低いなど、さまざまな問題があります。しかし、iMD は著者の所属機関と立地国をカウントしているだけで、データベースを全く必要としないので、このような問題を解決することができました。

図2は iMD の計算式です。C（country：所属機関の国数）と A（affiliation：所属機関数）を足しています。 α と β は、この後、重み付けができるようにあて置いているのですが、今は1です。これは対数を取って計算しています。理由としては、例えば学内紀要と「Nature」を比べると、「Nature」では著者数が年間3,000人などですが、紀要はそれほど多くなく、比べると時には3,000対2ということもあります。それだとあまり比較にならないので、スムージングをしています。

図3の左上のように、従来の評価Aでは、A誌、B誌、C誌、D誌と雑誌を並べたときに、IFのある雑誌だけが見えていました。雑誌はたくさんあるのですが、IFの付いたものか、あるいはゼロでしか評価されてこなかったということです。

一方、従来の評価Bでは、とにかく本数で評価されていました。いい雑誌に載っても、紀要であっても、

【iMD®の考え方】



- ・「学内紀要<全国学会誌<国際学会誌」という評価には、一般論として、多くの研究者の同意が得られる
- ・これを定量化するには？
 - ・学術誌の著者所属機関の多様性を定量化
- ・質は定量化できない。
 - ・IFも質ではなく話題性を定量化
 - ・iMDも質ではなく多様性を定量化

（図1）

【iMD®】



- ・学術誌を著者所属の多様性の観点から定量化する指標
- ・計算式

$$\log_n (\alpha \times C + \beta \times A)$$

- ・C:所属機関の国数
- ・A:所属機関数
- ・ α と β はCとAの重み付け係数

138751 「評価システム、評価方法及びプロ

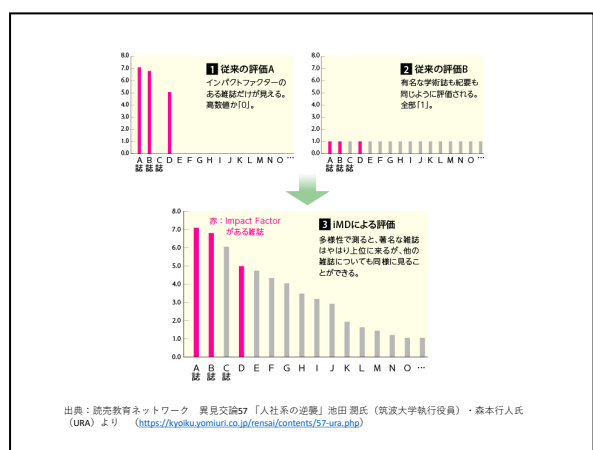
（図2）

1本は1本とカウントされていました。

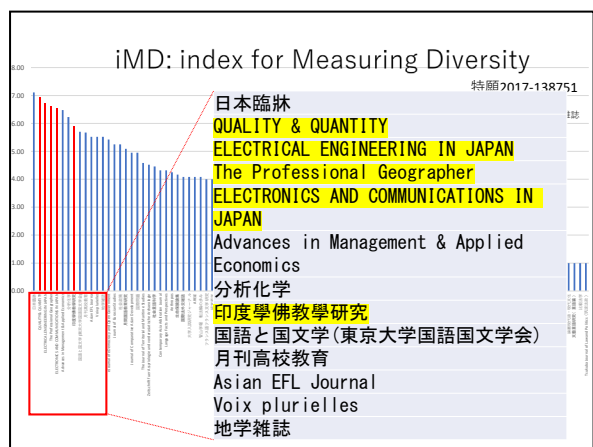
しかし、iMD による評価では、その雑誌にどのような所属の人が論文を書いているのかという多様性を見るので、何かしらの数字が付きます。iMD による評価で並べると、IF のある雑誌の方が圧倒的に少なく、著名な雑誌はやはり上位に来ますが、今までゼロも付かなかったような他の雑誌についても同様に見ることができました。

図4のグラフは、筑波大学の人文社会系の教員がある年に書いた雑誌を横に並べたものです。iMD が高い順番に示していますが、黄色でハイライトした雑誌には IF が付いており、それ以外は付いていません。ここで、従来の被引用数に基づく評価では見えてなかった学術誌の多様性を見ることができました。

図5のグラフは横軸に IF、縦軸に iMD を取って相関を見たものですが、やはり IF が高い雑誌は iMD も



(図3)



(図4)

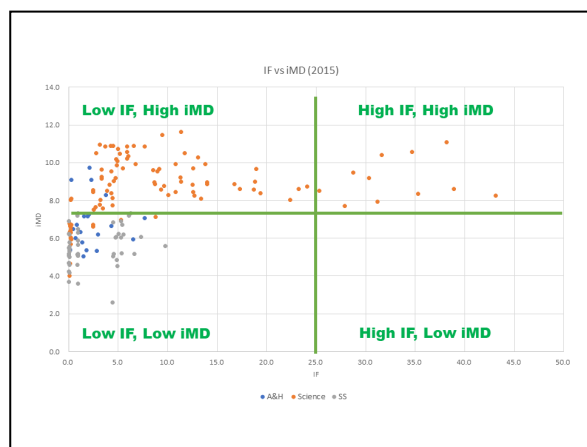
高くなるという結果が出ました。一方、IF がゼロ線上にある雑誌は、従来見えてこなかったものですが、iMD によって数値化することができ、見える化に成功したと考えています。ドットの数 Web of Science の Science Citation Index (SCI) から 200 誌、Social Sciences Citation Index (SSCI) から 50 誌取ったものです。Arts & Humanities もあります。

筑波大学ゲートウェイ

ここからは、筑波大学ゲートウェイについてご紹介します。

まず、導入の経緯です。2019 年 2 月 15 日に虎ノ門ヒルズフォーラムで「人文社会系分野における研究評価シンポジウム」（筑波大学主催）を行いました。そこには本学の池田潤先生、国立歴史民俗博物館の後藤真先生、Elsevier 社の Anders Karlsson さん、F1000 Research 社の Rebecca Lawrence さんにご参加いただき、また Research England の David Sweeney さんからは日程が合わなかったということでビデオメッセージを寄せていただき、人文社会系分野における研究評価について考えました。

ここでご縁があり、Lawrence さんから筑波大学ゲートウェイのご提案がありました。そのときに、F1000 Research 社の理念として、「研究と学問、そして言語には壁があってはならない」とおっしゃったことに強く共感を覚え、話を進めていくことにしました。日本のオープンサイエンスは着実に進んでいますが、諸外



(図5)

国と比較すると、やや出遅れている感があることは否めません。そういった意味で、コントリビューションは学术界に対してそれほど多くないと思っています。また、英語に過度に依存した学術情報コミュニケーションにも問題がないと言い切れません。

そこで、筑波大学は、日本で初めて F1000Research 出版モデルを大学として活用し、世界で初めて英語以外の言語の出版モデルを可能とすることにしました。日本と世界の学術情報コミュニケーションに一石を投じたいという考えから導入に至りました。最終的にはこれが波及して、どの国の言葉でも迅速に、オープンに、制約なしに研究成果を発信できることがグローバルスタンダードになることを願っています。

Lawrence さんは 2 度来日してくださり、本学の研究担当副学長と人文社会系長、学長補佐室長、URA が直接説明を受けました。それを受けて、本学の研究担当副学長が学内で筑波大学ゲートウェイを導入したときに投稿するかという希望調査を行ったところ、一定のニーズがあることが分かりました。最終的には学長にご判断いただき導入することになり、今年に入って学内説明会を夏に 1 回、秋に 1 回、そして 12 月にも実施して、各研究者に投稿のご検討を頂いているところです。

困難だった点をお話しします。文系にとっては APC が 1,350 ドルと高額だったので、人文社会系長と話し合い、独自の投稿支援プログラムを準備していただきました。これは徐々に全学的に広がっているところです。また、大変困難だったのが、国内に代理店がなかったので契約手続きは直接やりとりをしていたのですが、契約の文化が違って手間と時間がかかった点です。ただ、最近代理店ができたので、今後、自分の機関でも導入しようと考えている方は、そちらの代理店を通じてゲートウェイあるいはプラットフォームを導入することが可能です。

図 6 は F1000Research の特徴です。従来の学術誌では、論文を投稿すると匿名で査読が行われ、査読を通った論文だけが出版され、論文の著作権は発行する学

会や出版社が持ちます。そして、オープンアクセスジャーナル以外は論文を読むために何らかの料金が発生します。それに対し、F1000Research では、投稿すると最低限のチェックとトリアージを行い、それを突破した論文を XML で組み直して Web で公開、そこでパブリッシュということになります。公開時は査読なし論文の扱いになりますが、出版されているのでこの時点から引用が可能になります。公開後にオープンピアレビューが始まり、査読で 2 名以上にアプルーブされるとアクセプトされます。すると査読付き論文となり、Google Scholar、Scopus、PubMed、PubMed Central などのデータベースに収録されることになります。

他にも、伝統モデルに依存するたくさんのモデルがありますが、説明は省略します。

このスライドでお示したように、F1000Research の出版モデルは、オープンアクセスジャーナルとも、プレプリントとも、リポジトリとも異なる、新しい出版であるということです。

図 7 は投稿の流れを表したものです。順を追ってご説明します。

まず投稿です。筑波大学ゲートウェイは、単著であれば筑波大学の教員でなければ投稿できませんが、共著の場合は共著者の中に筑波大学の教員が 1 人以上入っていれば投稿可としています。また、英語であれば全ての分野で投稿が可能です。人文社会系分野については英語ないし日本語で投稿していただくことができます。図 8 は投稿の画面です。まず原稿のタイプを選

【F1000 Research】

現在の主流な査読モデル

投稿→非公開査読→出版(著作権は出版社、引用可)→有料購読

F1000 Researchモデル:

投稿→出版(著作権は著者、引用可)→無料閲覧→公開査読

ほかのモデル(いずれも伝統的モデルに依存):

オープンアクセスジャーナルモデル: 伝統的モデル+無料閲覧
ハイブリッドモデル: 伝統的モデル+出版時に有料購読か無料閲覧かを選択
プレプリントモデル: 出版前自主公開 → 無料閲覧(+伝統的モデル)
リポジトリモデル: 伝統的モデル+出版後自主公開(別DOIで引用)→無料閲覧

(図 6)

び、タイトルを入力します。それからアブストラクトとキーワードを書きます。この辺は他の雑誌と一緒にと思いますが、興味深いのは、その下の Authors Contributions に全ての共著者の役割を入力してもらうことです。

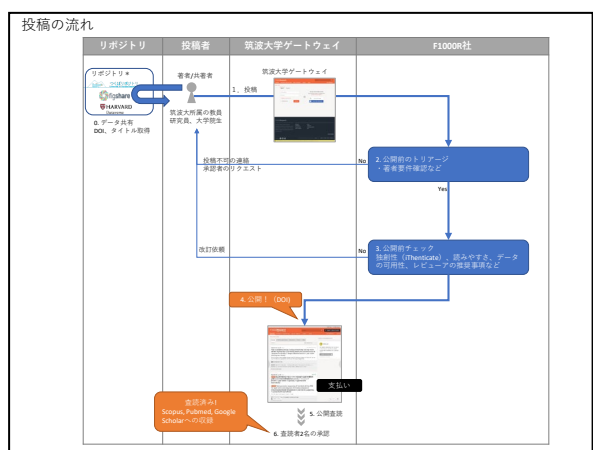
それから、筑波大学ゲートウェイは、原稿だけではなく、データガイドラインに沿ってデータもアップロードしてオープンにする必要があります(図9)。これが一番大きい特徴だと思います。そして APC の支払い方法を選びます。論文は Creative Commons の CC-BY、データは CC0 のライセンスで出版されることになり、その条件に同意する必要があります。最後に、投稿に当たり共著者から同意を得ているか、二重投稿に当たらないかなどの誓約チェックがあります。

これらを全て書き込んでサブミットのボタンを押す

と、公開前のトリアージということで、F1000 Research 社側のチェックが入ります。次に、剽窃がないかが iThenticate に掛けられて確認され、あとは読みやすさやデータの可用性などが確認されます。この段階で査読者の推薦も求められます。剽窃については非常にまれなケースであると F1000 Research 社の方から伺っています。

これらのチェックが終わると、論文が XML に組み直されて、図10のような画面で公開されます。

それから公開査読が開始します。「？」あるいはチェックマークをクリックすると査読者のコメントを読むことができます。2人以上の査読者から承認を得ると査読終了となり、Scopus、PubMed、Google Scholar に収録されます。All Metrics が付いていて、views や downloads、citations、それから Google Scholar に飛んでいくということも見るできるようになっています。



(図7)

(図9)

(図8)

(図10)

す。また、PDF でダウンロードすることもできます。

以上、筑波大学ゲートウェイの説明をしてまいりましたが、これを研究成果出版の一つの新たな選択肢として、筑波大学の研究者に活用していただきたいと思っています。特に、論文をすぐに出版したい、論文の認知度を高めたい、論文を Scopus や PubMed、PMC に収録したい、論文の citation を上げたい、論文の著作権を保有したい、リーズナブルにオープンアクセスにしたいという方々にとってはメリットが大きいのではないかと考えています。

料金は図11のようになっています。カテゴリーAだと 800 ドル、B だと 1,000 ドルです。多くの人文社会系の先生方の研究成果は、リサーチアーティクルに該当すると思いますので、カテゴリーC の 1,350 ドルに該当すると思います。

おわりに

最後に筑波大学ゲートウェイの可能性をお示しして、私の話を終わりにしたいと思います。

これまでは見えにくいといわれてきた人文社会分野の研究成果を国際的に見える化するために、iMD や筑波大学ゲートウェイを導入してまいりました。研究者は研究費などを自身で獲得し、実験や調査を経て、アウトプットとして論文を出しているわけですが、そこに購読料や掲載料が発生しています。APCは必要経費として仕方がないにしても、研究者が無償で提供している論文を過度に商品化する原稿の出版モデルに対

し、多くの研究者が違和感を覚えているのではないのでしょうか。筑波大学ゲートウェイが波及し、どの国の言葉でも迅速かつオープンに、制約なしに研究成果を発信できることがグローバルスタンダードになることを願っています。

●矢吹 「APC補助のための支援プログラムでは、どの程度の予算を用意されたのでしょうか。総長裁量経費、学長裁量経費などから割いているのでしょうか」という質問が来ています。

●森本 まだそこには至っておらず、人文社会系の中でいくつかの論文に対して支援すると伺っています。

●矢吹 続いての質問です。「見える化推進の成果として、F1000Research 導入前後における citation 数への影響等の調査等は計画されていないのでしょうか」。

●森本 大変的確な質問です。これは計画しています。筑波大学の中に筑波大学ゲートウェイ運用委員会があり、そこで毎年、見直しを行い、次年度の目標を上げるのですが、そこで計画について話し合っています。

●矢吹 次に、「F1000Research はプラットフォームなのだと思いますが、ここに投稿された論文は、どの学術雑誌に掲載されることになるのでしょうか。あくまで論文単位で収録されるのであって、○○誌というラベルは付かないのでしょうか。仕組みが飲み込み切れず申し訳ありません」というご質問です。

●森本 イメージが湧かないところではあると思いますが、ご指摘通り F1000Research はプラットフォームです。そのプラットフォームにぶら下がっているのが筑波大学ゲートウェイです。プラットフォームが1本あり、そこに複数のゲートウェイがぶら下がっている



(図 11)

形になりますので、投稿して引用する場合は F1000Research ということになります。DOI でそのように明記されています。従いまして、ご質問への回答としましては、F1000Research となります。

●矢吹 次は iMD についての質問です。「雑誌単位の評価と論文単位の評価は必ずしもイコールではないと思います。ご検討の過程でもし論文単位の評価も試みられたようでしたら、ご紹介いただけないでしょうか」。

●森本 論文単位の評価はしないという約束で、雑誌単位の評価について全学の先生に協力を頂いています。

●矢吹 ありがとうございます。思わず私にもこっとしてしまいました。

次の質問は、「人文社会科学分野の学術誌は引用等で評価できないとのことで、そういう分野があることを初めて知りました。これまではどういう形で評価してきたのでしょうか」。

●森本 本数がほとんどだと思います。本数で組織評価や教授への昇進が行われていたと聞いています。もちろん内容を精査して教授に昇進ということもあると思いますが、定量的な部分では本数ということです。

●矢吹 次の質問です。「F1000Research で、公開前のステップにおいて、大学としては質保証などに関与しているのでしょうか」。

●森本 補助金を渡す先生についてはアブストラクトを送っていただき、それでジャッジして配分するという支援プログラムだと聞いています。

●矢吹 まだ始まったばかりですからね。

●森本 そうなのです。

●矢吹 次の質問は、「F1000Research のオープンピアレビューにおいて、例えば競争相手や悪意のある第三者から意図的に悪い評価がなされたとき、受理への道が遠くなってしまうことはないのでしょうか。そのような問題を避けるために編集担当の見識や判断が必要ではないかと思うのですが」ということです。

●森本 第三者から悪い評価がなされるのは、恐らく公開査読のときだと思います。基本的には投稿したら受理される、どのような論文もオープンにするというのが F1000 Research 社のスタンスです。F1000Research が面白いのは、編集者がいないところです。つまり、エディターバイアスを受けないことが特徴の一つになります。恐らくおっしゃりたいのは、自分が査読している論文が自分の論文によく似ていて、自分の論文がまだ出ていないから、自分の論文が出てからこの査読中の若手の論文をアクセプトしようということが起こるのではないかという質問だと思いますが、そのようなエディターバイアスもレビューバイアスも受けなことが筑波大学ゲートウェイの特徴となっています。

●矢吹 次に、「世界への研究成果発信力を高めるという目的は達成されつつあると感じますか。今のところの効果について教えてください」という質問です。

●森本 実は 11 月 26 日に完成したばかりで、まだ 1 カ月もたっていないのです。投稿が始まっているので、今まさに私たちも効果を静観している状況です。

第2回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

コロナ時代における研究情報発信 - プレプリントに関する出版社の取り組み

アントワーン・ブーケ

(シュプリングー・ネイチャー (日本))

講演要旨



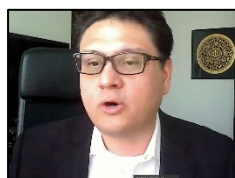
研究成果への早期かつ迅速なアクセスは、科学のペースを加速させ、発見を進展させるために重要である。特にコロナ時代において、その重要性は顕著である。そのため、出版社は、研究をできるだけ早くアクセスできるように取り組んでいる。シュプリングー・ネイチャーのジャーナルは、長きにわたって、プレプリントの共有をサポートしてきた。また、Nature のジャーナルおよび BMC ジャーナルは、20 年以上も前から、出版前の原稿を共有することへの価値を認識してきた。

シュプリングー・ネイチャーでは、ジャーナルへの投稿と査読（ピアレビュー）プロセスに、早期共有と透明性の高い査読を統合した革新的なサービス、In Review（イン・レビュー）を導入している。

今回の発表では、コロナ時代における研究出版の変化、プレプリントの長所と短所、および In Review を通じてどのように研究の早期共有を推進しているかを紹介する。また、どのようにして研究をオープンにしているのか、さまざまなオプションについても解説する。

アントワーン・ブーケ

アントワーン・ブーケは、シュプリングー・ネイチャーのヴァイス・プレジデントとして、東京を拠点に日本、東南アジア、オセアニアのインスティテューショナル・セールスを担当している。アジア太平洋地域の学術出版界で20年以上の経験を持ち、シュプリングー・ネイチャー（日本）の代表取締役社長も務める。これまでアジア地域の出版プログラムを指揮したほか、日本で医学情報事業を立ち上げた。また、書籍のコミッショニング・エディターとしての経験もある。オーストラリアで生まれ、ブリスベンのグリフィス大学卒業。東京大学で博士号（物理学）を取得。1994年から日本在住。



COVID-19 に関連する研究出版物

本日紹介するデータは、ほとんどが Digital Science の Dimensions のデータベースから 2020 年 10 月ごろに取り込んだものですが、Digital Science の COVID-19 のデータベースはリアルタイムで毎日アップデートされているので、そのデータレポートも紹介します。また、2020 年 6 月に Digital Science が COVID-19 の研究成果に関する大変面白いレポートを発表したので、その詳細もぜひご覧ください。

Dimensions のデータを分析すると、2020 年 1 月から、

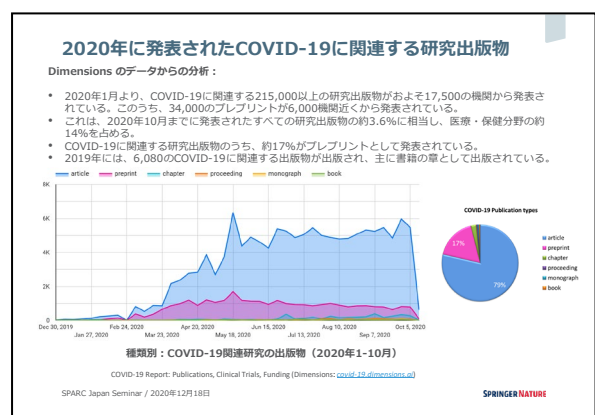
COVID-19 に関連する 21 万 5,000 報以上の研究出版物が 1 万 7,500 機関から発表されています（図 1）。そのうち 3 万 4,000 件（17%）はプレプリントで、6,000 機関近くから発表されています。特に各国のロックダウンなどが始まると、4 月、5 月ごろから研究論文とプレプリントの数が大幅に増加しました。研究論文は 10 月まで毎月ほとんど同じくらいの程度で出版が続く、プレプリントはだんだん減っています。COVID-19 に関する研究論文のトータル数は、Dimensions データベースに入っている研究論文の約 3.6%で、医

療・保健分野の約 14%を占めます。毎年、世界の研究論文は 2～3%成長するといわれているので、COVID-19 に関する研究はその成長を 2 倍ぐらい加速させたと言えるかもしれませんが、このコロナ時代の間に、COVID-19 に関する研究だけでなく、全体的に研究のアウトプットが大幅に増えました。

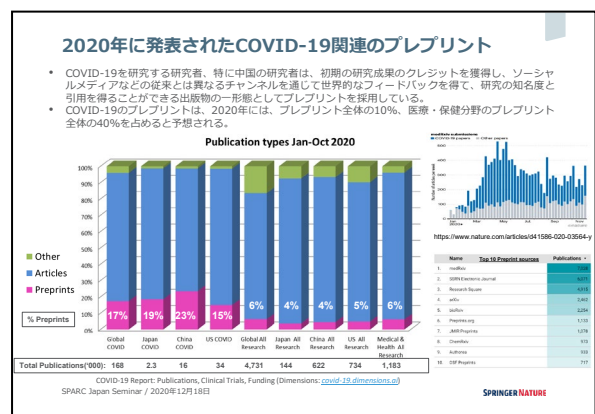
COVID-19 に関連する研究出版物のうち、17%はプレプリントです。同じ検索カラムで 2019 年のデータを見るとたった 6,000 件だったので、ほとんどは今年に出版されたものです。

COVID-19 に関するプレプリント

世界の COVID-19 に関連するプレプリントは、世界の COVID-19 に関連する出版物のうちの 17%です (図 2)。同じ時期に出版されている研究出版物全体に占めるプレプリントの割合はたったの 6%なので、COVID-19 に関する研究のプレプリントはその 3 倍で



(図 1)



(図 2)

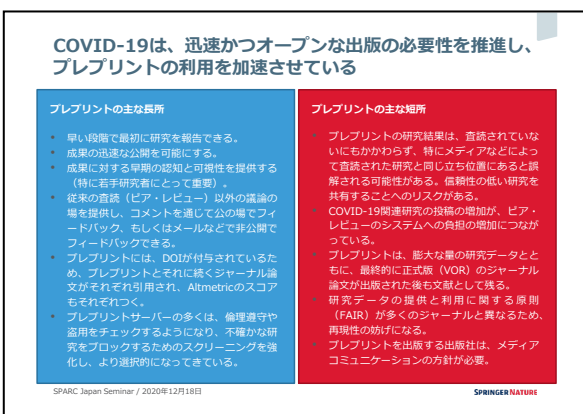
す。特に中国の研究者は約 4 分の 1 がプレプリントの記事で出版しています。日本の場合は 19%です。そして、医療・保健分野のプレプリントのうち 40%が COVID-19 に関するものでした。

面白いのは、プレプリントがどこで発表されているかということです。Dimensions の中で一番多いのは medRxiv で、あとは SSRN Electronic Journal や Research Square、arXiv です。medRxiv は 2019 年 6 月にスタートしたばかりで、今のところプレプリントの 3 分の 2 が COVID-19 に関するものなので、ほとんど COVID-19 のプレプリントのリポジトリになっています。

プレプリントの長所と短所

プレプリントの主な長所は、早い段階で「最初に私がこの研究をしました」という報告ができること、成果の迅速な公開が可能であること、その成果に対する早期の認知と可視性が提供できることです (図 3)。

この 3 点は特に若手研究者にとってとても大事なことです。また、従来は、査読のプロセス以外で、投稿する前にフィードバックはほとんど得られませんでした。が、プレプリントは出せば公開もしくは非公開のフィードバックが得られます。また、プレプリントは DOI が付与されているので、プレプリントとそれに続くジャーナル論文の両方が引用され、それぞれに Altmetric のスコアも付きます。また、F1000Research がいい例ですが、プレプリントサーバーの多くが倫理遵守や盗用をチェックするようになっており、不確かな



(図 3)

研究をブロックするためのスクリーニングが強化され、より選択的になっています。

プレプリントの主な弱点は、まず、プレプリントの研究結果は査読されていないにもかかわらず、特にメディアなどによって査読された研究と同じ立場にあると誤解される可能性が高く、信頼性の低い研究を共有することへのリスクが非常に大きいことです。アメリカで結構有名なケースも出ています。また、COVID-19に関する研究の投稿が大変増加しており、ピアレビューのシステムの負担も増加しています。先にプレプリントが存在するので、科学雑誌は、より早く査読のプロセスを進めて、できるだけ早く査読済みの記事として出したいと考えます。また、プレプリントは膨大な量の研究データとともに、最終的に正式版（Version of Record）のジャーナル論文と全く同じように文献になるので、文献が重複して文存在することになります。それから、プレプリントに掲載されている研究データの提供と利用に関する原則（FAIR 原則）が必要です。F1000Researchのようなプレプリントサーバーはきちんとチェックされていますが、全体的にプレプリントサーバーはデータの提供についてはまだ少し弱いのです。それから、プレプリントを出版する出版社はメディアコミュニケーションの方針が必要です。リポジトリも含めてメディアはいろいろなところでプレプリントが見えているので、きちんとしたメディアコミュニケーションの方針が必要ということです。

プレプリント論文のインパクト

プレプリントの数がどんどん増えて、単純に言えることは、インパクトのあるプレプリント論文はインパクトのあるジャーナル論文として出版されるということです。一例を挙げます。10月のDimensionsデータで一番引用されているプレプリントは武漢のグループが2月9日にmedRxivで出版したのですが、2カ月半後に同じものが「The New England Journal of Medicine」で論文として出版されています。その引用数は、DimensionsのCOVID-19に関する論文で2番目に高い

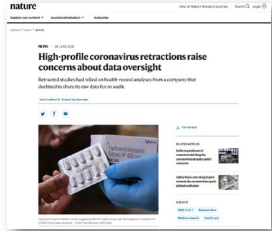
です。

研究者の目から見て、何がプレプリントの信頼性を高めているかという調査が行われました。「Royal Society Open Science」に掲載された論文によると、欧米の研究者約4,000人を調査したところ、プレプリントの信頼性を評価するために大事なものは、基礎となる研究内容の透明さ、例えば研究データへのリンクが透明であること、そしてオープンサイエンスのコンテンツを支持するデータがプレプリントに載せてあることだという結果が出ました。それほど大事ではないのは、著者の所属、著書の名前、グループ、それから、どこのジャーナルに投稿するつもりかということでした。主なプレプリントサーバーには、研究データのリンクなど、透明性を見せるリンクがまだ存在していないこともあるので、プレプリントの信頼性を高めるためにどんどんこのようなスタンダードをinsistすべきだと思います。

出版物の撤回

プレプリントについては、さまざまなスキャンダルなどの問題もありました（図4）。例えばアメリカで二つの有名なケースがあり、結局「Lancet」で出版された論文とそれに関するプレプリントは両方撤回されています。アメリカの大統領がよく言っていたヒドロキシクロロキンに関する有名なプレプリントです。撤回の主な理由は、第三者が研究データにアクセスできないということでした。

データの検証不足によるCOVID-19論文の撤回



- 米国のヘルスケア分析会社Surgisphereが提供したデータに基づくCOVID-19の論文2篇が、データの信頼性が低いために撤回された。
- Lancetに掲載された論文では、抗マラリア薬であるヒドロキシクロロキンがCOVID-19の患者にとって危険である可能性が報告された。
- New England Journal of Medicineに発表された論文では、特定の心臓の治療薬がCOVID-19に感染した人に与える影響が報告されており、安全性への懸念がないことが発表された。
- さまざまなところから2つの研究について懸念が示されたものの、Surgisphereは、第三者の監査人が検証できるように生データを提供しなかったため、著者は論文の撤回を求めた。
- この一連の件は、特にこのパンデミック下において、データ評価の厳格性、査読プロセス、医療への影響について深刻な疑問を投げた。

Sources: <https://www.nature.com/articles/d41586-020-01695-w>
<https://www.thescientist.com/news-opinion/lancet-retracts-surgispheres-study-on-hydroxychloroquine-67613>

SPARC Japan Seminar / 2020年12月18日

SPRINGER NATURE

(図4)

しかし、本当に COVID-19 に関する論文の撤回は大きな問題なのでしょうか。今まで 39 報が撤回され、その中の 15 報はプレプリントでした。実は、それは撤回された論文のトータルのたった 2.3%です。ですから、特に大きな問題にならないと思います。撤回は、例えば結果のエラー、データのエラーなど、よく予想できる理由によるもので、プレプリントでも論文の記事でも、COVID-19 に関する論文の撤回は大きな問題ではないようです。

シュプリング・ネイチャーのプレプリントに関する方針

研究データを FAIR の原則でオープンにすることの必要性について先ほど言いましたが、特に COVID-19 の影響で、オープンデータを利用したい人がさらに増えています。シュプリング・ネイチャーの調査によると、他の人のデータを利用したい研究者が大幅に増えているので、プレプリントでもデータのリンクが大変必要になってきています。

2019 年 5 月から、シュプリング・ネイチャーのジャーナルは全タイトルのプレプリントをエンカレッジしています。なぜかという、早い段階で著者のプライオリティや人名が認められるからです。プレプリントはグリーン OA の重要な形態だと思います。査読のプロセスに入る前の段階でできるだけ早く結果が出せるので、出版社にとってはとてもいいことです。

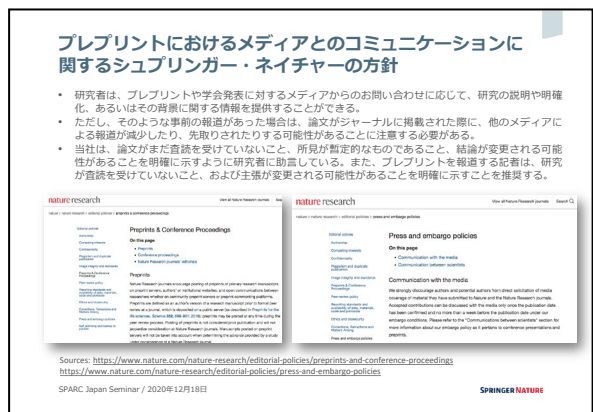
プレプリントにおけるメディアとのコミュニケーション

ョンに関するシュプリング・ネイチャーの方針は図 5 のとおりです。やはり研究者には、プレプリントや学会発表に対してメディアから問い合わせがあれば、必ずそれがまだ査読されていない論文であることを説明していただきたいです。それから、プレプリントの段階でメディアからの注目が大きい場合は、その後、査読された論文がジャーナルに出るところでマスコミの興味が減少するので、もし編集や訂正があっても、正しいデータがそれほど注目されない可能性があります。プレプリントサーバーでも出版社でも、メディアの方針が必要であり、それはとても大事です。

In Review

In Review は、COVID-19 に関する 3 番目のプレプリントサーバーです（図 6）。シュプリング・ネイチャーが出資する Research Square という会社のプレプリントサーバーなので、シュプリング・ネイチャーのジャーナルと In Review はかなり深い関係があります。In Review のプレプリントサーバーの主な特徴は、F1000 と似ていて、プレプリントが存在しながらジャーナルの査読のプロセスを透明にし、査読されている間に早い段階で成果も分かるサービスであることです。

投稿するときには、著者が In Review のサービスにオプトインできます。そして品質管理チェックが行われます。このプレプリントは、著者のバージョンをジャーナルの記事の形に編集して、その段階で公開します。In Review には公開されているプレプリントのバー



(図 5)



(図 6)

ジョンが表示され、ここでは Under Review のバッジが付いています。これは Nature Research の In Review のプレプリントサーバーに載せている記事です。このプレプリントは、雑誌「BMC Developmental Biology」で出版されます。すると雑誌のサイトに記事として移動して、Accepted のマークが付きます。BMC の場合はオープンアクセスの CC-BY のライセンスがあるので、それが In Review のサーバーにも残ります。

最近 Research Square は、In Review の特に COVID-19 の研究によって大変人気になり、2020 年 11 月 16 日で 5 万報のプレプリントを発表しています。過去半年間に発表したプレプリントの 60% は COVID-19 に関するものでした。

データで見る In Review

In Review のプレプリントのデータを見てみると、In Review で出されたプレプリントがそのままジャーナルに受理される割合は 43% です（図 7）。これが高いかどうかは分かりませんが、1 回リジェクトされて違うジャーナルに投稿されるケースも含めると、その割合は大きくなると思います。プレプリントは結構見られており、平均 8 回のダウンロードがあります。最も多くダウンロードされたプレプリントでは約 5,000 回ですが、やはり雑誌の投稿論文よりはダウンロードされていません。

引用については、今までに Research Square で約 2,000 件の総被引用数があります。引用されるプレプリント

はたくさん引用されますが、ほとんどは引用されていない状態で、全プリプレントのうちたった 3% 未満しか引用されていません。

また、Altmetric のスコアで注目性を見ると、あるプレプリントはたくさん注目されていますが、多くのものには Altmetric スコアがなく、注目されていません。これはジャーナルの記事と似たパターンだと思います。

コロナ時代における査読への負担

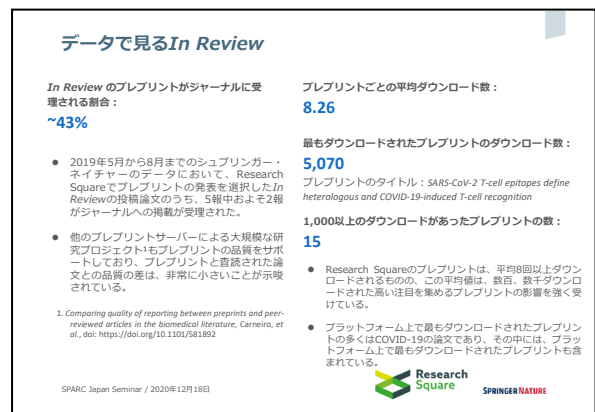
先ほど、ピアレビューのシステムのプレッシャーについて少しお話ししました。COVID-19 の時代で、COVID-19 に関連する論文を出版社はより迅速に優先的に出版したいのですが、質の高い査読は維持したいのです。従って、大変限られた数の査読者に、かなり大きな負担が掛かっていると思います。シュプリング・ネイチャーと他の出版社も入っている C19 Rapid Review Initiative では、ボランティアの査読者のレビューと身元が出版社やジャーナル間で共有できるようになっています。

ピアレビューに関する大変面白いデータとして、COVID-19 に関する論文は大変速く査読のプロセスが行われているのですが、COVID-19 以外の論文はその負担を受けてだいぶ時間がかかったということがあります。

OA ステータス別：2020 年の COVID-19 関連論文

最後に、プレプリント以外のジャーナル論文の OA のステータスを見てみます（図 8）。現在、COVID-19 に関する世界の論文の約 80% は完全に無料アクセスできるようになっています。そのうち純粋なゴールド OA は全体の 20% を占めており、これは世界的全論文のオープンアクセス率である 17% に近い数字です。

なぜ 80% になったのかというと、出版社がクロズドの論文をブロンズ OA の形で無料提供しているからです。それがウィズ COVID-19 の時代で 40% くらいになっていますが、平均的にブロンズ OA は 17% ぐら



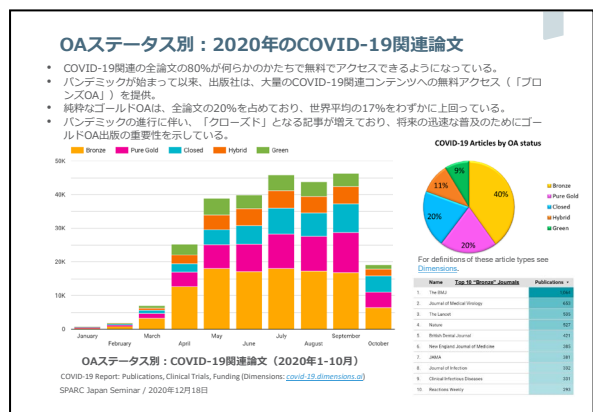
(図 7)

いです（図 9）。来年、再来年になってアフター COVID-19 の時代になると、ブロンズ OA の影響は減っていくと思います。

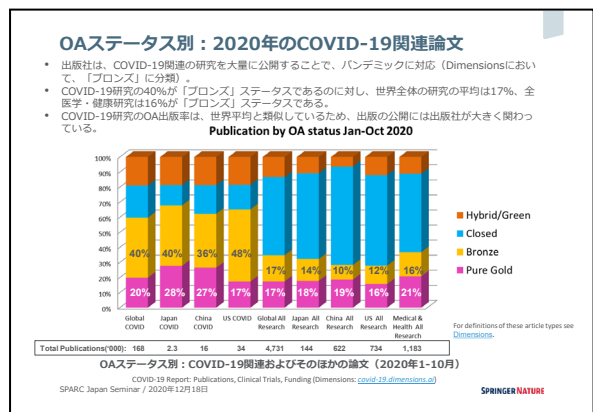
迅速な研究をオープンに公開するためのベストな方法は、ゴールド OA かもしれません（図 10）。COVID-19 に関する研究論文をこれからより迅速に出版し、高いインパクトを達成するためには、ゴールド OA が最適な形態だと思います。

●矢吹 一つ目のご質問は、「図 3 でプレプリントの主な短所として挙げられている、ジャーナル論文が出版された後も文献として残る点について、もう少し詳しくお願いできますか」というものです。

●ブーケ プレプリントサーバーには、ジャーナルへ



(図 8)



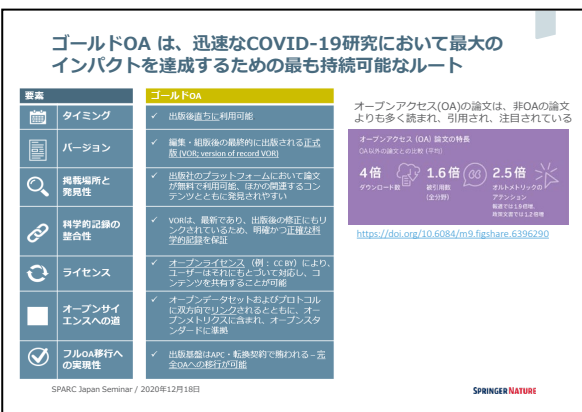
(図 9)

の投稿と査読のプロセスが終わった後もプレプリントとジャーナル論文の両方が文献として残ります。査読のプロセスで記事が改善されたり、新しい意見が出たりするので、若干異なるけれども同じようなものが重複して残ります。たくさん論文の発見（discovery）は必要なのですが、数が増えるので検索する側には大変な負担になるのではないかと考えています。

●矢吹 次の質問は、「シュプリング・ネイチャーではプレプリント論文の引用を支持するということがありますが、査読前で信頼性が担保されていない状態の内容を引用することへの懸念はないのでしょうか」というものです。

●ブーケ ほとんどのプレプリントは投稿してそのまま雑誌に出版されることになりましたが、メディアの方針でよく言われているように、引用する側は「これはプレプリントを引用しています」とはっきり書いていただきたいです。査読のプロセスでこの結論が変わる可能性があることをみんなが理解できるようにするためです。プレプリントの良さは、より早く研究成果などが出ることでありますが、やはり科学のプロセスでさまざまな改善があるので、読者は引用されているプレプリントだけではなく、最新データもチェックするべきだと思います。

●矢吹 坊農先生から、「図 3 の短所の下から二つ目、



(図 10)

プレプリントに掲載されたうんぬんのところで、FAIR が再現性の妨げになるという書きぶりになっていますが、誤訳ではないですか」というご指摘があります。「FAIR は Findable・Accessible・Interoperable・Reusable にするための原則なので違うのではないのでしょうか。もし元の文章があれば、そちらも示していただけるとありがたいです」とのことです。

●**ブーケ** これは、プレプリントサーバーは FAIR の原則を利用しないとその再現性が問題になるという意味です。つまり、プレプリントサーバー、リポジトリは FAIR の原則をフォローしなければならないという意味です。

●**矢吹** もう一つご質問です。「査読を通らなかった論文、リジェクトされたもののプレプリントは残るのでしょうか。また、不正などによって取り下げられた論文のプレプリントはどのようなのでしょうか」。

●**ブーケ** その二つのケースはだいぶ違います。リジェクトされたプレプリントは、そのままプレプリントとして残るケースが多いと思います。プレプリントはそのまま文献として存在感があるので、ジャーナルでそのままアクセプトされても、違う形でアクセプトされても、リジェクトされても、そのまま残ります。しかし、不正 (plagiarism, fraud) に関するプレプリントは、撤回のプロセスがあります。また、Retraction Watch はジャーナルの論文だけではなくプレプリントのサーバーも調査しているので、プレプリントも撤回になることがあります。今までは割と少なくても 17 報のみですが、その中に結構有名なケースがありました。フランスのプレプリントで、COVID-19 に関するものでした。政府のポリシーに大きな影響があったのですが、それはやはり撤回されました。

第2回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

生命科学研究における プレプリントや SNS 活用の現状と課題

坊農 秀雅

(広島大学 大学院統合生命科学研究科 生命医科学プログラム)

講演要旨



bioRxiv (バイオアーカイブ) をはじめとするプレプリントの活用が生命科学分野においても急速に広まり、演者も海外での学会発表の前にはその発表のプレプリントをアップロードするようにしている。また、ResearchGate に代表される研究者向けの SNS も普及し、それを介した (電子的な) 別刷り請求やポストドク申し込みなど、従来の学術コミュニケーションに変革が訪れている。しかしながら、それぞれその情報の質や有償コンテンツの受け渡しなど問題点も多く指摘されている。そこで本講演では演者の実体験を交えながら、その利点と欠点に関して生命科学分野におけるプレプリントと SNS の使われ方の現状を紹介する。

坊農 秀雅

理化学研究所、埼玉医科大学を経て、2007年より情報・システム研究機構ライフサイエンス統合データベースセンター (DBCLS) にて統合TVや統合データベース講習会AJACSを立ち上げ、日本DNAデータベース (DDBJ) と連携して第3塩基配列データ解析技術開発とその利活用を行う。2020年4月より広島大学大学院統合生命科学研究科 特任教授として卓越大学院 (ゲノム編集先端人材育成プログラム) でのバイオインフォマティクス教育に取り組みつつ、ゲノム編集データ解析基盤技術の開発とバイオインフォマティクスによる遺伝子機能解析を行う研究室 (bonohulab) を立ち上げている。京都大学博士 (理学)。



私は以前、この SPARC セミナーの運営委員をしており、ちょうど2年前 (2018年10月) にプレプリントの話をしたので、この2年でどう変わったかというところに注目して聞いていただければと思います。今日はプレプリントと研究者版 SNS のお話を、私の事例を使ってご紹介します。

私は2020年4月に広島大学に移りました。ゲノム編集 (Genome editing) を使うセンターにも所属しています。ゲノム編集の実験を行うには、その前段階でさまざまなデータを集めなければいけません。例えば、どの遺伝子を狙ってゲノム編集をするかを決めたり、ヒトではなく新しい生物の場合は、ゲノム配列を読ん

で遺伝子を決めたりします。そのためのゲノム編集データ解析基盤技術の開発と、バイオインフォマティクスの遺伝子機能解析研究を、新しい研究所を立ち上げながら進めています。

bioRxiv

図1は2年前に出したスライドです。私は公共のデータベースの再解析 (メタ解析) を行うための手法を論文を作って出したのですが、2年前にその話をしました。当時、私は静岡に住んでいて、2018年2月23日の富士山の日とその論文のプレプリントを bioRxiv に出して、8カ月後の10月にプレゼンをしました。そ

の8カ月の間にアブストラクトの閲覧が907回、PDFのダウンロードが約300回あり、たくさん見られるようになってうれしいという話をしました(図2)。その後どうなったかという、2年後の2020年11月末現在でアブストラクトの閲覧数は3,000を超え、PDFのダウンロード数も1,000を超えています(図3)。また、プレプリントでアップするときはPDFを論文の体裁にしてアップするのですが、bioRxivのシステムが良くなり、そこからテキストマイニングをして、普段ジャーナルで見るようなフルテキストへの組み替えも自動的にできる仕組みになりました。これはmedRxivもそうです。

その他の違いとしては、「これはプレプリントなので注意して見てください」という赤枠の注意書きが付きまして。それから、左の真ん中あたりのだいたい色で囲んでいるところですが、査読論文が通ると自動的にリンクが張られるようになりました。プレプリント

で放ったらかしていても、古いDOIのものとトレースできる仕組みになっています。ですので、これはもう古いのではないかと懸念は少ないのではないかと思います。ただ、もちろん漏れがあるので、プレプリントのままで、査読の方に出ているのにつけられないということはあるかもしれません。

そういう状況で、bioRxivのシステムも社会も2年で変わりましたが、研究者社会の方はあまり変わっていません。プレプリントを出すメリット・デメリットは全く変わってなくて、残念ながら、研究者でも一部の人間だけが先取制を担保するために出すということが多いです。多くの保守的な日本の研究者は、「プレプリントはよく分からないからいいや」という状況です。

変化したのは、medRxivがCOVID-19の影響で広く知られるようになったことです。これはもちろんいいのですが、査読論文との違いを無視したり、プレプリントに関する無理解が露呈したりする状況があります。また、一部の共同研究者は、査読前にbioRxivに出し、より早く「自分たちがこの研究をしたのだ」ということを言い、エディターの人たちが悪いことをして別の論文を先に通すということがないようにしています。早くbioRxivに出せば、自分たちが先にやったということが分かるので、そのようにする人が増えてきています。

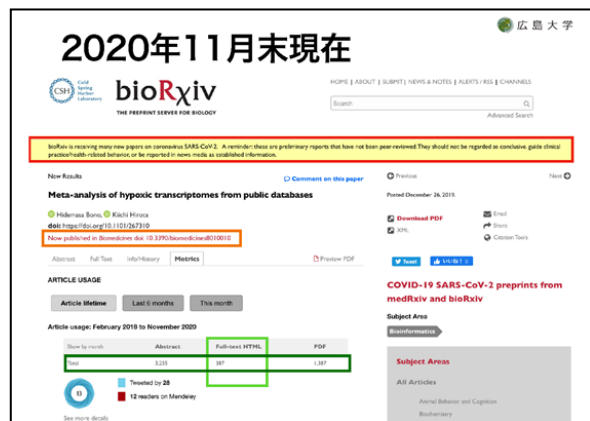
しかし、プレプリントが増えている中で、bioRxivで受け入れてくれないタイプの論文もあります。例え



(図1)



(図2)



(図3)

ば、私は生命科学の中でも特にバイオインフォマティクスというプログラムを書いたりするタイプの研究をしているのですが、みんなで集まって Hackathon (hack と marathon をかけた言葉) というミーティングのようなものをしたときのレポートを私たちは査読論文誌にずっと出してきたのですが、そのプレプリントを bioRxiv に出したら「ミーティングレポートは受け付けません」と断られました。

それ以外の変化として、論文誌を出しているシュブリンガー・ネイチャーや MDPI において、論文投稿と同時に自分たちのところがやっているプレプリントに投稿されるような仕組みができ、プレプリントに誘導する仕組みがだいぶ出てきています。有名なのは、今月に入ってから出た eLIFE というものです。割といい論文誌を出すということで研究者の間では「ここに出したい」という人が増えています。その eLIFE が、"publish, then review" ということで、まず論文をパブリッシュして、その後レビューを行うということを始めるとアナウンスしています。これも結構話題になっており、このような F1000Research モデルが広まっていくといいなと思って見ています。

BioHackXiv

先ほどの Hackathon のレポートは bioRxiv で受け入れられなかったのが、BioHackXiv に出しました (図 4)。これもプレプリントで、DOI は付くし、ORCID などにもレコードが入りますが、Europe PMC には載って

いません。こういう形で、自分たちが BioHackathon のミーティングでまとめたレポートをプレプリントとして論文に出すことができます。これは 2019 年の秋に私どもがヨーロッパの BioHackathon で議論したことをレポートにして出したものですが、こういう形で残していくことができるのは、それはそれでいいなという例の一つです。

F1000Research

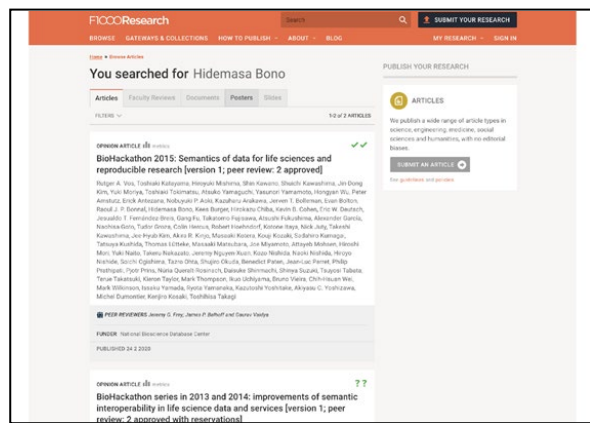
図 5 は F1000Research で、これも私が出したものです。実は BioHackathon のレポートとして 2 回、F1000Research に出して、上と下で別々の論文です。2015 というものと、2013 and 2014 というものですが、上の 2015 はピアレビューで OK となっていて、2 人分のチェックが付いており、PubMed で検索できるようになっています。下の 2013 and 2014 は「??」と付いています。これは、レビューされたがアプルーブされず、PubMed などにも載らず、このままの状態でのプレプリントのように F1000Research のサイトにあるという状態です。

ResearchGate (RG) : 研究者向けの SNS

ResearchGate (RG) は研究者向けの SNS です。割と古くて、2008 年にボストンで設立され、2013 年からはドイツのベルリンに本部を移しています。少し古いデータですが、2016 年の公称利用者数は 900 万人です。いつから自分が登録しているかという詳しいレコード



(図 4)



(図 5)

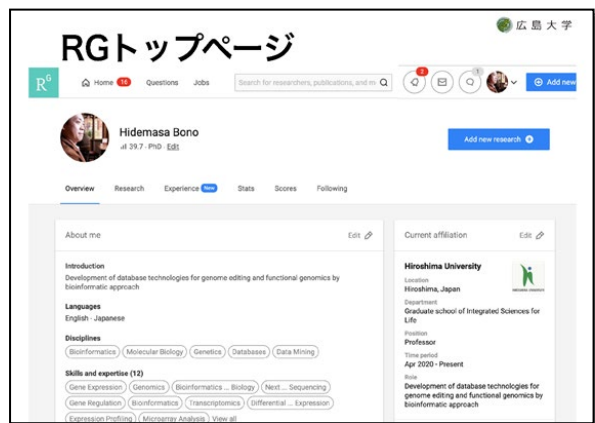
は残っていませんでしたが、私はどうも 2015 年ごろから登録して使っているということが分かりました。

基本的には researchmap や ORCID のように、研究者の論文業績リストのような感じかと思っていましたが、ORCID と同様、この ResearchGate もどんどん自動で登録されていくようです。誰かが登録しないといけないようですが、共著者のうちの誰かが登録すれば他の人にも行くという仕組みが古くからできていたようで、自動で入るような印象があります。researchmap も最近バージョンが上がってそういう仕組みになってきているので、大して変わらないとは思いますが、そういうことを早く取り入れたというイメージです。

それから、following や followers という概念もありますが、Twitter のようにフォローしていないと見られないということではなく、どちらかという研究に関係があるキーワードや誰かの名前を検索して見に行くような感じです。全然知らない、フォローしていない人ともやりとりができ、実際のところ followers や following というのはあまり関係ない仕組みになっています。

私がいいと思っているのは、業績に対する統計値の可視化ができることです。自分たちの論文が引用された・されていないということが 1 週間ごとの解像度で出るので、割とリアルタイムにそういうものが見えて面白いです。

図 6 は私自身のトップページです。名前や PhD などが書いてあります。右上にあるアラートは、Facebook のアラートと同じですが、鐘マークのところ



(図 6)

に 2 と書いてあります。クリックすると、自分の論文を誰かが新しく引用したというお知らせが来ています。放っておくとすぐたまってしまうので、自分のトップページを見るよという誘導がうまくできていると思います。

右上の会話マークはチャットのようなものです。グレーで 1 と書いてあるのは、見たけれどもずっと残してあるということです。具体的には「論文の PDF を下さい」というリクエストが来ています。そういうものも全部ここに出ます。

SNS の機能もちろんあります。私自身は 4 年間で 10 個以下のメッセージしかありませんが、ほとんど国外からのメッセージです。一番多いのが別刷りの PDF の請求です。私はオープンアクセスの論文にばかり出しているので、よく論文を読んでいないのだけれど思いながら URL を教えてあげて解決しているのですが、中には「オープンアクセスではない論文誌の PDF が欲しい」というメッセージも来て、これはどうなのかと思っています。そこは個人によって判断が分かれるところで、私は、自分がファーストオーサーやコレスポンディングオーサーでないものは断っています。あとは雑誌特集号への寄稿依頼や、最近増えてきたのはポスドク申し込みです。そういうものがメッセージとして送られてくることが多いです。

また、researchmap 的な機能として、研究業績が出ます。特徴的なのは、その論文のページを何回見たかということが集計されることです。論文を登録する際に、普通の査読論文だと DOI を入れれば勝手に登録されますが、プレプリントの場合は論文本体の PDF をアップしないと登録されません。CC-BY などになっている普通の査読論文だと、勝手に収集されてここに張られていくようです。

そして、スコアが研究者ごとに付いています。このスコアは、論文を出した数、誰かからの質問に答えた数、followers の数などで出のですが、私の場合はほとんどが論文で数えられています。

それから、h-index も同じように計算されています。

これに関しては、ほとんど Google Scholar と一致しており、ある程度信頼できるのではないかと思います。

図7は1週間ごとの統計値です。これは11月と10月ですが、誰かが見に来た回数、フルテキストを読んだ回数、Citations、あとはRecommendationsということで、誰かがその論文をRecommendするとその数も出るので、私の研究だとそれはあまり出ていません。Research Interest など、そういうものを計算して ResearchGate 的にスコア化しているものだと思います。これらは weekly 以外に monthly でも見られます。そういう可視化の機能がいいと思っています。

その他の機能としては、自分が出した論文を研究のプロジェクトとしてまとめることができるので、頭の整理になっていいというところもあると思います。

それから、海外のものばかりですが、宣伝情報や job 情報がメールで来たり、本の広告が JREC-IN 的な感じで来たりします。

まとめ

以上、プレプリントと ResearchGate (研究者版 SNS) のお話をしました。両方とも日本の研究者がほぼ全員使っているということではなく、まだ一部の人が便利だと言って使っている感じです。しかし、フルテキストは ResearchGate に、プレプリントは Google にインデックスされていますから、ググって当たり着き、知らないうちに使っていたという例は結構あるようです。

しかし、研究者は Twitter を匿名 (のつもり) で使

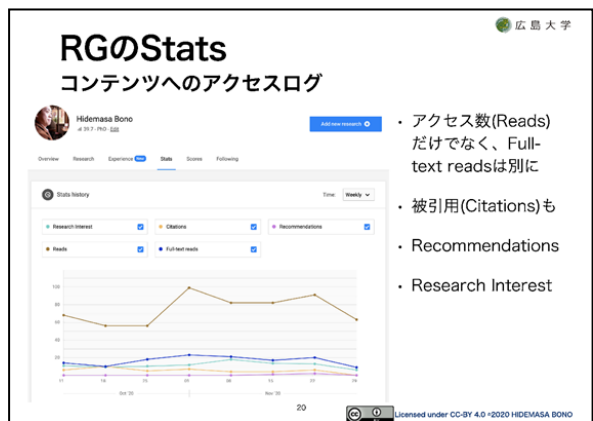
っている人が多いです。実名にしていないのは、今後の評価などに影響すると考えているからのようです。

そういう状況なので、やはりプレプリントは研究費を出している側で「出してください」と義務付けるべきだと思います。研究費を出す海外の機関、例えばウェルカム・トラストは、独自にプレプリントのサーバーを用意し、「そこに入れてください」というようにしていますし、National Institutes of Health (NIH) の PubMed Central (PMC) も、NIH グラントでやったものは自分たちのところに入れることになっていたと思うので、日本もトップダウンでそのようにしていった方がいいと個人的には思っています。

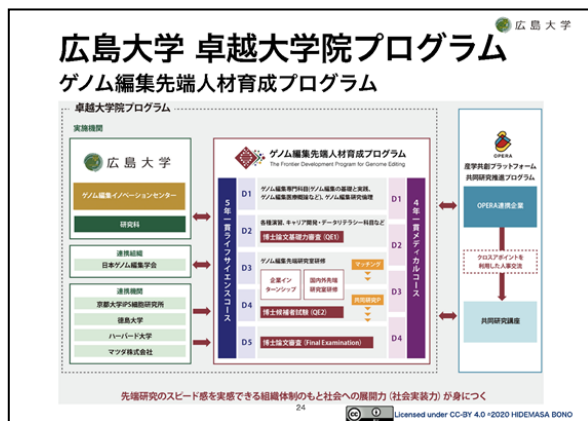
一番問題なのは、査読論文しか評価しないという現在の基準だと思います。研究者は割と皆さんそういう頭があり、「査読論文しか出さない。他はやっても労力の無駄だ」と思っていることが多いので、そういう意識を変えていくような活動を若い世代に対して行っていく必要があると思います。そして、やはりトップダウンの強制力が必要です。これも前から言っているのですが、なかなか日本では実現しません。

広島大学 卓越大学院プログラム

そのようなことのための教育を、現在、広島大学の卓越大学院プログラムでさせていただいています (図8)。ゲノム編集先端人材育成プログラムの中でバイオインフォマティクスの講義を担当し、コンピューターを使ったデータ解析や、データサイエンス関係で生物



(図7)



(図8)

の研究を進められる人材の育成を行っています。

●矢吹 先生方とお話をしていると、ResearchGate を使っている方がいらっしゃる一方で、全く使わない方もいらっしゃいます。使っていると、もうこれだけでいいという感じになってくる先生もいらっしゃって、人によっては相当使い勝手がいいのだらうなという印象を受けていますが、プレプリントの話を含めて、ResearchGate が研究活動自体にうまく活用できている部分は何かありますか。

●坊農 やはり自分から情報発信しないと集まってこないという部分はありますが、私が自分の研究テーマで論文を出していると、特集号などのお知らせが来て「それに出してくださいね」という依頼が来ることもあるので、それはプラスになっているかなと思います。

興味がある人が質問してくるということがあればいいのですが、私に関しては今のところそれほどなくて、「ポストドクに来たいです」という人が来るくらいです。しかし、人によってはここでそういう研究交流が生まれて、そこから共同研究が始まることもあるのではないかと思います。私は ResearchGate ではなくて、Twitter の日本語でのやりとりで共同研究が始まったことは割とありまして、むしろ Twitter 世代です。まだ利用者が多くないときの Twitter でそういうやりとりがあり、プラスになっています。

●矢吹 ありがとうございます。ちょうど私も大学の研究力分析を行っており、URA という立場もあって、研究活動と SNS に関心があったので伺いました。

コメントが一つ来ています。「政府によるプレプリントの義務化あるいは著者最終稿の公開義務化には大賛成です。ぜひパネル討論で議論してください」ということなので、この後、パネルで言及していただければと思います。

もう一つ、「プレプリント論文におけるデータ公開の実例をご紹介いただけますか」という質問が来ています。

●坊農 プレプリントの場合は、プレプリントを出したときにデータも全部出さないとはいけません。査読論文の場合は査読論文の出版時に出していたのが、先になるということです。私自身は塩基配列、DNA の配列などを決めてそれを出すのですが、国際塩基配列データベースに登録し、その登録番号を論文に含めないといけません。査読中はそういうものは伏せておく、あるいは reviewer だけが見られるリンクを与えるのですが、arXiv などのプレプリントの場合は、アップロードした瞬間にそれが公開されていないといけないことになっています。それで、論文が出たときに、既にデータも利用可能になっているという状況が大事なのです。

今言ったのは塩基配列のデータベースの話ですが、定型の塩基配列のデータだけではなく、論文に付随したさまざまな種類のデータを、例えば Figshare や datadryad などにアップして、その DOI を論文に書いておき、そういうものは無制限に誰でもアクセスできるようにしておくことがプレプリントでは必須とされています。しかし、なかなかそれも広まっておらず、データはまだ公開していないけれども取りあえず自分たちの先取性のためだけに bioRxiv などのプレプリントにアップしているという例がたまにあるようです。ぜひ皆さんにプレプリントのルールを知っていただき、データがより公開・再利用されるような研究社会になればよいと考えています。

第2回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

Preprint が誘導する研究サイクルの力学考

引原 隆士

(京都大学図書館機構長)

講演要旨



arXiv.org が醸成してきた学術コミュニティにおける内的手段であった preprint が、既にこの手段の準備があったバイオ系を含む分野において、コロナウイルス感染下の情報共有及び日々変化する状況への議論の手法として脚光を浴びた。この preprint を含む動きが学術情報流通の新しい手段、もしくは別の手段を生み出すかどうかは、研究のサイクルを生み出す動力学を理解することで、議論することができる。多くの OA ムーブメントの位置づけも合わせて、成熟したからこそ考えるべき、日本の今後の研究の多様性と展開を生む必要条件を論じる。

引原 隆士



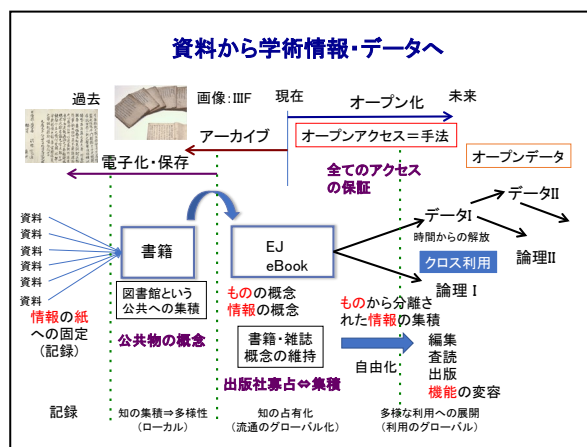
1997年以来京都大学大学院工学研究科に所属し、現在同電気工学専攻教授。研究分野は非線形力学の工学的応用、計測とシステム制御、パワープロセッシング。2012年以来京都大学図書館機構長として長きに亘り、オープンサイエンスに関わる活動に従事し、2015年にオープンアクセスポリシー、2020年にオープンデータポリシーを機関として取りまとめた。これらの活動を通じて、オープンサイエンスと同時に図書館だけでなく科学技術活動全般のデジタルトランスフォーメーションを指向している。2016-2018年 arXiv.org MAB。

プレプリントは、ステークホルダー間の単純なつばぜり合いではありません。研究活動の活力になるといえる本質的な面を捉える必要があります。そのことを図書館関係者あるいは研究者の方々に理解していただくことが本日の一番大きな目的かもしれません。私は専門分野が数理物理なので arXiv.org をよく使いますが、それに関する話は2017年10月のSPARCの講演で既にさせていただきましたので、今日お話しするストーリーから外させていただきます。

研究のライフサイクルにおけるプレプリント

モデルなくお話しすることは非常に難しいので、まず議論の前提として、学術情報とは何かということからお話しします(図1)。従来から、情報と呼ばれるものが書籍や論文などに固定化され、ものとして流通

してきました。ものとしてあるということはエビデンスであったわけですが、現在では、大きく見てエビデンスとして論理とデータを考えないといけないというのが学術情報における重要なポイントだと私は認識しています。そのうち、論理とデータに関しては、今は



(図1)

情報として扱われます。ところが、論理とデータは、ものがなくても成立するので、ものの中に固定化されなくなっているというのが現状です。

プレプリントを研究のライフサイクルで見えます（図 2）。モチベーションから始まり、研究、下書き、論文と進んでいく中にプレプリントがあり、研究のライフサイクルにおいてプレプリントはあまり大きな意味を持たないとされていると思います。特に図書館関係の方々は、数年前からプレプリントを気にしているかもしれませんが、arXiv.org に関してはもう 20 年以上運用されています。

このライフサイクルの個々のポイントで研究者のどのような能力や技術が必要かということについては、後でゆっくり見ていただきます。研究者にとって、研究を進めてキュレーションをする能力、一般化する能力、さらに論文への落とし込み能力は必須条件です。それに加えてさまざまな技術がなければ、このライフサイクルは回せません。従って、議論になっている論文の部分、サイクルでいう下の方のコミュニティの部分だけを取り上げてプレプリントを考えるのは非常に拙速だと思っています。

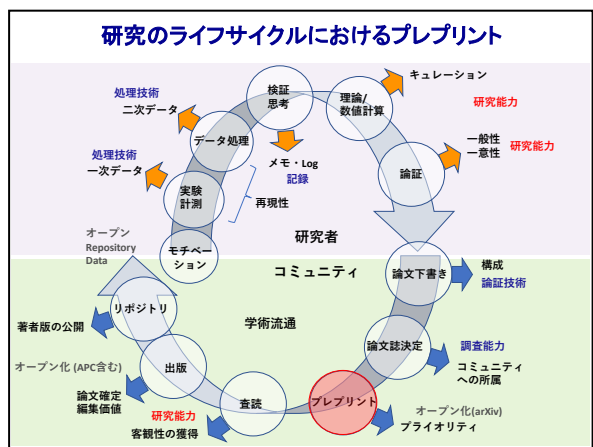
現状は、図 3 のように、プレプリントの範囲が広がっています。本来、プレプリントは小さなコミュニティの中で回す情報流通でした。それが、下書きの段階から論文誌決定、査読に至るところまでもコミュニティで回すようになってきています。先ほど、リポジトリをプレプリントサーバーとする話もありましたが、

そうなるとこの研究サイクルは非常に混乱を来します。ResearchGate も同じで、リポジトリやプレプリントが両方載っている状態の中でこのサイクルをどう回したいのかということが、研究者から見れば非常に重要なポイントです。従って、これをきちんとひもとかないといけません。

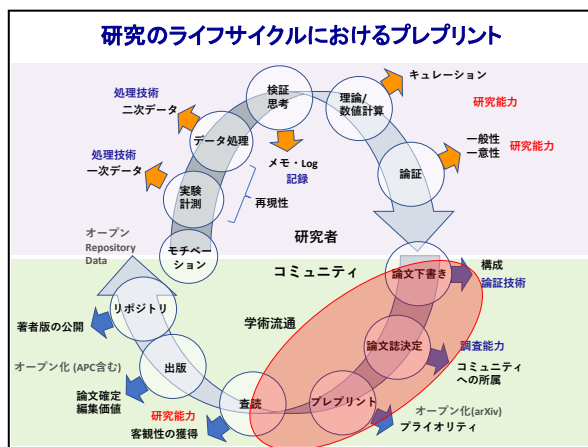
研究・研究者の動向分析

研究者の認識からすれば、研究のオリジナリティは非常に重要なポイントです。オリジナリティを高めるときにコミュニティがあるかないかということは別として、研究を創成し、それを展開して成熟させていく過程の中で、インパクトファクターが礼賛されるフェーズは、人が行ってきた研究をさらに進めていくフェーズです。しかし、オリジナリティを尊重する研究者から見れば、これは終わったフェーズであるわけです。ここで議論していても、研究は広がりはしますが、プラス α 、プラス β と数を増やしていくだけの世界なので、数が増えたからといって、その人を集めてきて研究者として充実しているかという何もできません。どこかでコピーを探してくる人たちのフェーズなわけです。はっきり言って、インパクトファクター礼賛フェーズは、研究者あるいは研究機関から見ると何の意味もありません。ですから、「今までの人事は失敗だった」と多くの大学が言っています。

プレプリントは、創成から展開のフェーズに関与しているということを理解していただかないといけない



(図 2)



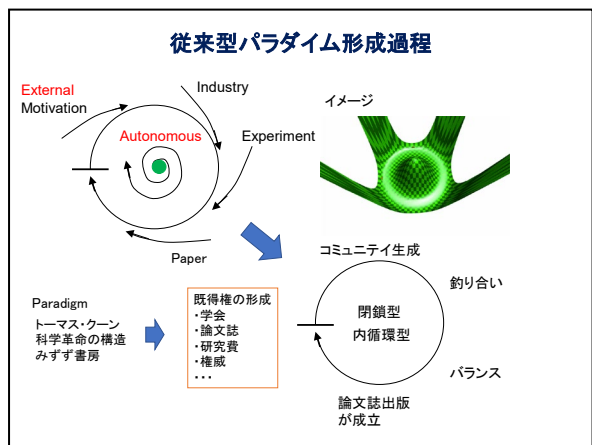
(図 3)

と思います。成熟期間ですぐ出てくるような論文をプレプリントで出しても意味がありません。出版社が「もうすぐ載りますよ。プレプリントも置きましょう」と言うのは成熟期間です。展開期間や創成期間は、その分野に人がいなくて、評価者もいないので、今までの雑誌であれば排除されていた世界です。これをきちんと残していく世界がプレプリントの中にあるのだということを理解していただかないと、何の意味もないのではないかという意見も出てきています。

また、論文がたくさん出てきたからといって、パラダイムシフトが起こるわけではありません。国の政策も大きく間違っていて、はやっている分野に人を集めればいつの間にかパラダイムシフトが起きて新しい製品が出てきて、今まで見えなかったものが見えるようになるという議論がいろいろなところでされていますが、そのようなことは一切ありません。チャージアップしたからといって分野は生まれませんということです。商業的には、雑誌の方々は、創成・展開の部分は人がいなくて買ってもらえない以上、いくらやっても仕方がないのです。ですからこの分野は、プレプリントが見るべき分野であると言えます。

従来型パラダイムと閉鎖型コミュニティ

少し別の観点から、従来型の学会やパラダイムの形成過程を見てみましょう（図4）。自発的に生み出した部分が広がっていく流れと、外から入ってくるさまざまな産業的要求、実験結果、他の論文、モチベーシ

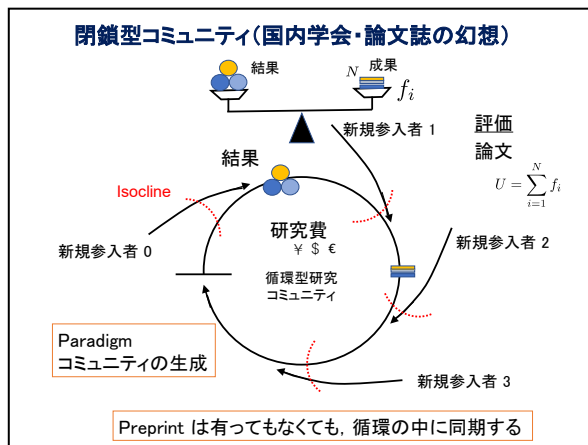


(図4)

ョンから、研究のサイクルがバランスを取ってぐるぐる回り出し、収束していく過程が一つのパラダイムになっています。イメージとしては、真ん中が持ち上がっていて、あるいは少し下がっていて、外から入り込むようなポテンシャルのたわみがあり、そこでぐるぐる回って自己充足型になっています。Thomas Kuhn のパラダイムの話から言えば、ここで出来上がるものは既得権です。要するに、学会であり、論文誌であり、研究費であり、権威です。これを守ることがパラダイムなのです。従って、パラダイムを守っているのが論文誌であることも事実です。閉鎖型・内循環型コミュニティの形成過程で生まれたものですから、今ある雑誌や学会は、それを維持するためのものであるということを知っていただく必要があります。

たくさんの新規参入者も全て循環型に入ってくるだろうというのが、閉鎖型コミュニティの幻想です（図5）。研究費はここにあるからみんな寄ってくるだろう、それだけの数の成果をバランスを取って出せますよという世界をつくり上げているわけです。外から入ってくる人も、このサイクルの中でフェーズを合わせながら生活すると研究費にありつけるというのが今までのパターンです。これは私見かもしれませんが、周りで起きているのはこういうことです。

この中では、プレプリントもあってもなくても一緒です。要するに、このフェーズに合わせて一緒に回っていれば、プレプリントがなくても周りに認められていくという状況があります。プレプリントには意味が



(図5)

ないと発言される方々は、結局はこの循環の中でぐるぐる回っている人だと言えると思います。

arXiv のサブミッションから見る変化

さて、arXiv のサブミッションのデータから何を見るかということです（図 6）。arXiv の創立の過程から言えば、高エネルギー物理ですから、一つの高エネルギー物理の分野でフェーズが回っていて、その中でコミュニティが形成されます。それがその後、いろいろな分野に分かれてくるわけです。もちろん数が変わっているのですが、比率だけで見るとあまり正確ではないのですが、この分野の中でそれぞれ別のコミュニティの中で回り出すものを、真ん中の辺りはサポートしていたと言えると思います。

しかし、今、この内循環型のパラダイムが変わりつつあります（図 7）。変わりつつあるというか、既に変わっているわけですが、要するに世界中で学会等がタコつぼ化する、あるいは学会が学会誌を出版社に買い取られて、そのコミュニティが小さくなっている、コミュニティが存在しなくなっているということが起きています。結果として、その中に若い人が入ってこなくて研究者が老化する、あるいは予算が減って既得権が小さくなり人が小さくなっていく。そうすると当然、中の引力は小さくなりますから不安定化します。そして外からいろいろな情報が入ってくると、中にいる人たちは外のものを求めて出ていくという状況が起きてきます。コミュニティが維持できなくなって、外

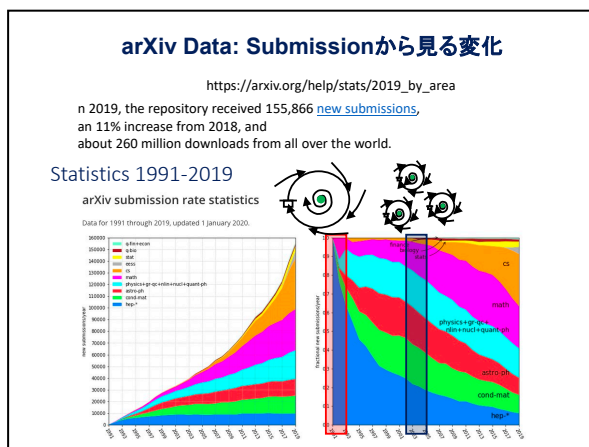
に人を発散し、予算も発散していく、あるいは論文誌も総合的なものも変わっていく。その中では、境界はできるかもしれませんが、フローが起きてくる。それが代替的に力学的に見た流れになるかと思います。同じようなことが arXiv でも起きています。

統計を見ると、サブミッションはどんどん増えているのですが、それ以上にダウンロードがものすごい数になっています。これは新しいものを求めてきていることを示しているのですが、一つの分野の人だけではなく、何かいい解析方法はないか、いいデータはないか、いいアルゴリズムがないかということを求められるのが arXiv のデータベースとしての在り方です。

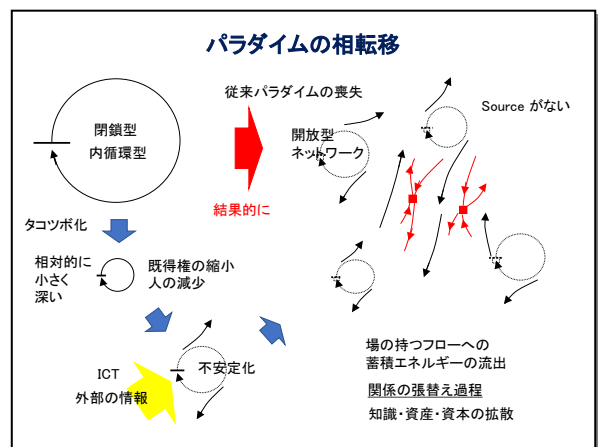
要するにここで起きているのは、コンピューターサイエンス、math など、横串的なものが広がってきて、高エネルギー物理や凝縮系の部分の比率が減り、どちらかという全体としてフローを起こすような研究内容が広がっているということです。それが今求められている arXiv のプレプリントとしての在り方です。

開放型コミュニティの出現

それを絵で描き直すと、図 8 のようになります。arXiv は一つのパラダイムを作っていましたが、それが開放型のコミュニティになってきています。そういうところは元々あったのですが、オリジナルな方々もプレプリントを持って外に出ていき、それを他の研究と合わせていくようなことが起きています。新規参入の方々も、一時期ここで in phase するのですが、プレ



(図 6)



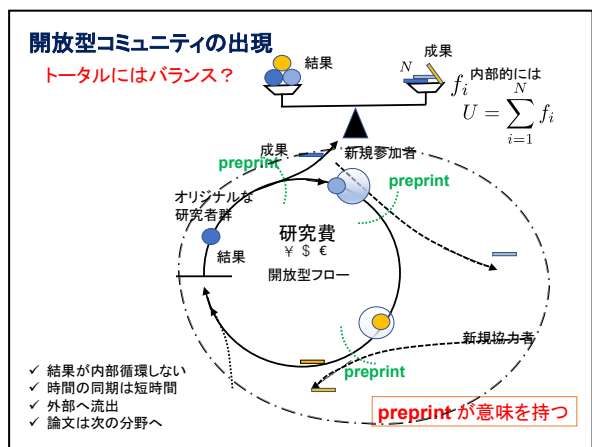
(図 7)

プリントという成果を持って出ていくという形です。そして協力者も同じような形で、あるときだけプレプリントを書いて、外に出ていく。つまりプレプリントは、ここにいる一瞬の間だけの in phase を保つための手段になっているということです。

固定的な流れの中ではプレプリントとジャーナルはほとんど変わらなかったのですが、分野的には、理論物理系などは時間もかかりますし、ジャーナルを待っている暇はありません。そのコミュニティの中だけでなく皆さん一緒に回るので何も問題ないのですが、少し寄ってまた外に出ていく人には待てません。だからプレプリントで成果を残して外へ出ていく。プレプリントはそういう大きな意味を持っているということになります。結果として、内部循環がもう存在していませんから、ある一瞬の時間、その切り出しが大変重要になります。論文は次の分野へ行ってまた書くということにもなり得るわけです。

受け皿の乱立

アーカイブ、プレプリントサーバーが山のようにあります。開放型になった結果として受け皿が乱立しているのですが、この中で権威があるかないかという話も出てきます。これは「乱立」でも、「受け皿がたくさんある」でも、どちらの言い方でもいいかもしれませんが、いずれもが閉鎖型や内部循環型のシステムにはなっていません。bioRxiv でも数理的なものがたくさんありますし、最近の COVID-19 でも数理的な感染



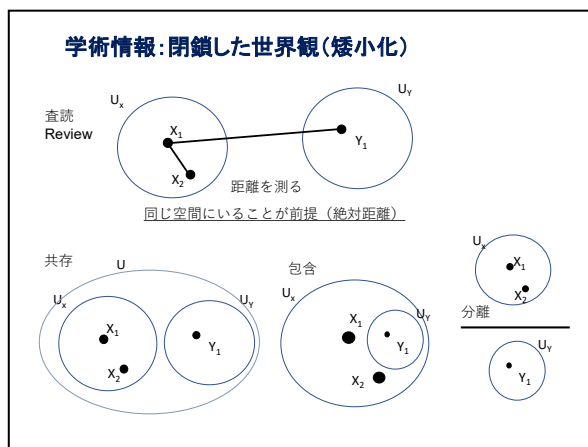
(图 8)

拡大のモデルなどの話は bioRxiv にも出ますし、arXiv にも出ますし、他のところにも出ています。それらを どうやってリンクするかという話は重要ですが、どこ でもいいわけです。要するに自分たちのメインのところ に寄って、そこで「これはこんな効果がありますよ」と示すことが重要なので、そこから出て論文誌だけ別のところで出すことは当然可能です。ですから、この世界自体がもうフローの中にあると思います。

言いたいことは、それぞれの受け皿が固定的に権威を持って残るかどうかなどというのは別のことだと思えますが、今この情報流通が非常に必要な時代になっているということです。きつい言い方をすれば、出版社の方々はこれを何とか取り込もうと網を張っています。しかし、そうではなくて、このフローを維持しなければ、科学の世界はまた既得権の収奪の世界になってしまいます。フローを維持することを考えなければ研究は進まないと思います。

学術情報：閉鎖した世界観・開放された世界観

学術情報が閉鎖した世界観を描くと、図9のようになります。論文を出した人たち (X_1 、 X_2 など) があって、別の人 (Y_1) も出したとします。そのときに、同じ空間にいるかどうかの距離を測るのが査読です。同じものである必要はなく、自分たちと同じパラダイムの中で動いているかどうかを見るのが査読だと思います。その結果、「同じ世界にあるから共存しましょう」「一緒だから包含しましょう」となる、あるいは



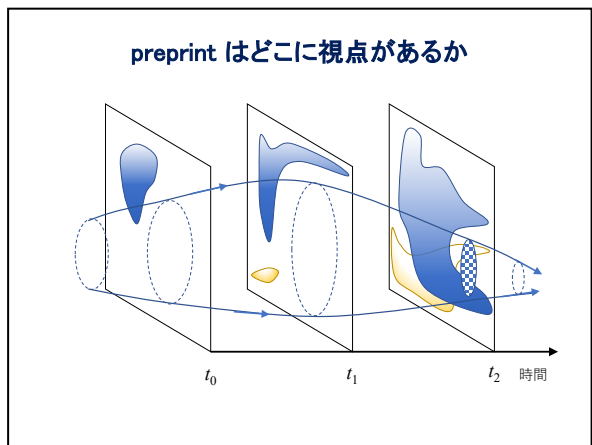
(圖 9)

「違うから別の世界に行きなさい」と切り出してしまいうわけです。これが今までの閉鎖した学術情報流通の世界観です。論文誌というものが、そのレッテルになるわけです。

ところが、グローバル化されたら、そうではなくて、ある一定の距離を保ちながら相互の関係を満たさないといけなくなります。そこで相互の関係を測るメジャーが必要になります。このメジャーがプレプリントです。あるいは論文でもいいのかもしれませんが、それぞれ同じ世界にいるわけではないので、共通で見えるメジャーが必要になります。それが今起きていることで、別の世界へ行っても自分が測ってもらえるようにメジャーのポイントとなるものを残していく。その行為がプレプリントになっていると私は研究者として見えています。

プレプリントはどこに視点があるか

話をもう少し進めて、プレプリントはどこに視点があるかということを考えてみます（図 10）。研究が始まり、創始的なものから最後に結論が出ていくという時間経過の中で、オリジナリティがある過程と、広がっていく過程があります。その中で、研究のロングスパンをまず描きます。オリジンからターゲットがだんだん広がり、包括領域が広がって、最後はその応用領域として収束していきます。この中で必ずあるのが、この視点の中に入っていない外側の成果です。それを、自分たちがこの中だけのロングタームを肯定するため



(図 10)

に否定してはいけません。否定しないために、外側にあるものをきちんと残していかなければいけないというのが科学の世界、研究者の世界です。

青で描いたものが広がっているのに、自分たちの世界観だけにいると、それが収束していったときに全くずれてしまうということが起きます。プレプリントは外側にある世界を見いだすものだということも意識しないといけません。中のものでプレプリントを置いていくということは、時間としてどちら側にあるかだけの話で、あまり意味がないのです。

研究者的な研究促進とファンドの研究ターン促進

研究者が研究を促進したいときは、研究開始から結論が出るまでの時間を短くします。ですから、人が集まっているところで書き出して、人が去っていく前に終わらせてしまうのが、一番賢い、世の中をうまく回る研究者のパターンだと思います。そして目標を達成して予算をもらって出ていってしまいます。

しかも、この研究者の性（さが）はファンドにうまく利用されています。選択と集中ですが、ほとんど結果が見えてからファンドを与えて、みんなが頑張ってくれる間に成果を評価します。別にファンドがこの分野を育てているわけでも何でもなく、既に育てているのです。育てるのだったら、誰も見てくれないところや、始まる手前のものを本当は育てないといけませんが、それにはファンドは何の効果もありません。ですから、ファンドがプレプリントを要求したところで、そんなものは意味がないと私は思っています。プレプリントの外側にあるものが重要であり、プレプリントは、目の目を見ていないところにきちんと残していくことが重要です。なぜかということを次にお話します。

図書館的視点

図書館が現在プレプリントとして意識しているものは、既存パラダイムの外にあると思います（図 11）。

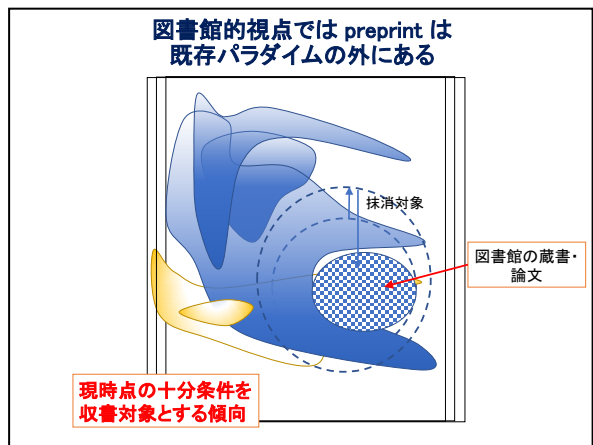
図書館の蔵書や論文は、非常に狭いもので、最終形として残ったものです。それはパラダイムとしてぐるぐる回って成熟したのですが、これしか残りません。ところが、科学の世界観、グローバリズムからいえば、その外側に重層的にさまざまな成果が上がっているわけですが、それを否定してこれを残すのかというのが、今までの論文誌の在り方の問題です。

プレプリントとして、自分の時間を提示しながら、こちらが手前で向こうが後なのということもきちんと残しながら置いていくことが、科学の透明性として非常に重要です。成果が出たものに対してものを言うことはいくらでもできますが、それは後付けの理屈です。本当は、この過程が広がって狭まる、外側ではないということを示した点線の部分を残していくことが科学の成立なわけです。ところが、現時点での十分条件だけを求めているのは、論文誌も出版社も変わりません。それが現実で、皆さんがされているのは営利事業だと思います。

arXiv が示す原点

arXiv が示した原点は、要するに研究をエンカレッジするのだということです。ですから、ステークホルダーが取り合いをしているという話ではなく、各研究者が研究できるようにするということなのです。アーカイブやプレプリントの議論の中では、その観点が抜けていると思います。

もう一つ重要なことは、クリアでシンプルなインタ

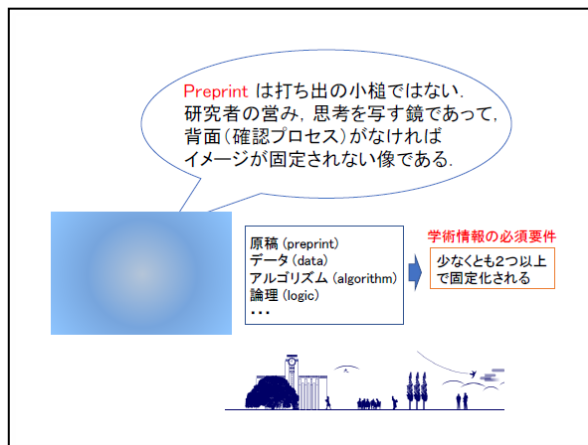


(図 11)

ーフェースであるべきだということです。いろいろなものが乱立している中で、出版社自身が「プレプリントを置いたら載りますよ」と言うこと自体が大きな間違いだと思います。それは研究者の活動の取り込みにすぎません。営利活動だからそうなるのかもしれませんが、それは研究活動を支援するものではないと思います。研究者は、研究分野のアーカイバーではありません。研究者は過去のものをずっと考えていく者ではなく、まだ見ぬ世界を探検する者ですから、それをエンカレッジするようなものでなければならないと思います。そして、プレプリントがその可能性を持っているものであり、その外側の世界が本当は探求する世界だということを皆さんはあまり意識されていないということを、あえて申し上げました。

プレプリントは研究者の営みや思考を映す鏡

当然ながら、プレプリントは打ち出の小槌ではありません。研究者の営みや思考を映す鏡であって、背面（プロセス）がなければイメージが固定されないものだと思います（図 12）。つまり、今見えるプレプリントの評価はジャーナルに射影したものであって、ジャーナルの観点からプレプリントが見られています。これは間違っていると思っています。ジャーナル自体が研究の一断面にすぎず、プレプリントはもっと広い世界を持っているのだということを見たら、むしろプレプリントの射影がジャーナルにすぎないのではないかというのが私の意見です。



(図 12)

また、学術情報の必須条件は、原稿やデータ、アルゴリズム、論理などのうち、少なくとも二つ以上で固定・検証されるものです。一つだけで検証されるものは科学としては成立しません。それはプレプリントでも変わらないということを最後に申し上げたいと思います。

ので、ご質問はすぐには出づらいのかもしれませんが、大変勉強になりました。

●引原 プレプリントを現象論で見るのか、手法論で見るのか、観念で見るのかで大きく違うと思うので、今役に立たないというような判断をするのは、関係者はやめた方がいいだろうというのが私の基本的な姿勢です。

●矢吹 深貝先生から、「引原先生の創成段階に焦点を当てた強調は、社会科学の側からも見て分かりやすいものです。力学的観点でご説明いただきましたが、知の進化論といった設定も可能です。さて、これは後のパネルに向けてですが、コロナ状況という新規事象の中で、既存の細分化を越えた領域横断を迫るという観点でプレプリントを位置付けることができれば」という、この後のパネルに向けてのコメントが来ていますが、もし応答があればお願いします。

●引原 おっしゃるとおりだと思っています。プレプリントは、大きな変化をするときに非常に強力な武器になると思います。COVID-19 がそれを示しているように、以前ですと、高温超伝導が発見されたとき、ほとんど一晩でいろいろな成果が変わっていくときに、社会も大きく変わる意見もたくさん含みながらプレプリントが発散的に広がった時代があります。同じように、大きく変わるときであればこそプレプリントが意味を持ち、論文誌が役割を失うという現実があるということを見たときに、社会が定常で安定した状態であれば、多分ほとんどなくなると思います。しかし、社会が発展するためには安定を続けてはいけなくて、安定が一回不安定になって次の安定を求めていくという発展形を作らないといけないとすると、プレプリントのある世界を広げて、そこで見ない限りは伸びないだろうというのが私の個人的な意見です。

●矢吹 ダイナミズムを整理していただいたところな

第2回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

パネルディスカッション



- 池内 有為** (文教大学 文学部)
山形 知実 (北海道大学附属図書館)
河合 将志 (国立情報学研究所 / オープンサイエンス基盤研究センター)
森本 行人 (筑波大学 URA 研究戦略推進室)
アントワーン・ブーケ (シュプリンガー・ネイチャー (日本))
坊農 秀雅 (広島大学 大学院統合生命科学研究科 生命医科学プログラム)
引原 隆士 (京都大学図書館機構長)

●池内 パネルディスカッションは、書誌多様性に関する四つの障壁に沿って、関連する質問を取り上げながら、パネルの方お一人ずつにコメントを頂きながら進めていきます。四つの論点以外の質問に関しては、時間が残りましたら個別にお答えいただきたいと思います。

共通言語としての英語の優位性

●池内 まずは「共通言語としての英語の優位性」についてです。これは常々、日本の私たち、日本語を母語とする者たちは感じていることかと思えます。F1000Research による日本語論文・成果の公開という話が森本さんからありました。また、坊農先生への質問として、「ResearchGate は researchmap よりも機能が豊富なようですが、日本国内の研究者同士で情報交換するのに ResearchGate をそのまま日本語でのやりとりで使うこともありますか」という質問が来ています。このあたりを絡めつつ、皆さんに一言ずつお話を伺っていききたいと思います。まず河合先生、お願いします。

●河合 英語で書かれていることによるメリットは当

然大きいと思いますが、アカデミアだけではなく社会への普及ということも考えると、現地語で書かれていることのメリットも当然あると思います。ですから、例えばオリジナルの論文自体は英語であったとしても、近年、自然言語処理が大変発達しているので、自動で現地語に翻訳するような機能を実装していくことなどがあるのではないかと考えています。

●池内 ありがとうございます。機関リポジトリでの公開ということになると、やはり日本語のものの公開も期待されるところだと思います。

森本さん、ご講演の中でも研究と学問、そして言語には壁があってはならないという理念が紹介され、ここは本当に多くの方の共感を呼んだ部分ではないかと考えます。

●森本 ありがとうございます。まさにおっしゃる通りで、その理念が共有できたので、本学学長の承認を受けて筑波大学ゲートウェイというものが日本で初めて導入できたのではないかと考えています。講演の中でも言いましたように、日本と世界の学術コミュニ

ケーションに一石投じたいと考えており、その視点から英語の優位性ということについて一言で言うなら、キーワード検索ではないかと思います。つまり、筑波大学ゲートウェイに投稿されると Google 検索でも検索可能になるので、日本語での検索が世界からできます。今のところ、Scopus に収録されている雑誌でいうと、Google に反映された分ですけれども、英語しかヒットしません。仮に将来 F1000 が自動翻訳などを取り入れて、どの言語でも検索できるようになれば、どの言語でもいいと思うのですが、今のところは元の言語でしか検索できないと思います。日本語の論文を検索したい研究者は、母国語が何であれ日本語で検索することが予想されますので、F1000 と筑波大学ゲートウェイで日本語論文を公開することは、日本と世界の学術コミュニケーションに大いに貢献するのではないかと考えています。

●池内 ありがとうございます。母国語で高等教育を受けたり、専門書を日本語で読んだりできるという恵まれた環境もある一方で、やはりそこを越えて英語で発信するということにどうしても障壁がある中で、こういった日本語のプラットフォームができることは大変期待が大きい取り組みではないかと思っています。

続いてブーケさんにお話を伺います。もちろんシュプリンガー・ネイチャーのジャーナルは英語のものが多いと思いますが、こういった日本語での発信が出てくることに関して何かコメントがありましたらお願いします。

●ブーケ 今までは研究論文とプレプリントサーバーは大体、その雑誌やプレプリントサーバーを作ったコミュニティによって作られているので、それが例えば日本語かフランス語のコミュニティだったら、当然その言葉を利用します。今のところは世界的な科学の情報の流れはほとんど英語ですが、出版社としては、価値がある情報があればどの言葉でもアクセスさせたいと思います。

森本さんがおっしゃったように、技術にその役割があります。どのような言葉で出版しても世界中で誰でもアクセスできるような技術が利用できれば、この問題はなくなると思います。最近、AI を利用した自動機械翻訳のツールはだいぶ進んでいます。プレプリントサーバーでも、科学雑誌などを読みながら自分の好みの言葉で情報が出てくるツールが発展すれば、そういう問題もだんだんなくなると思います。もちろん、社会学の方は言葉によってニュアンスが違うなど、言葉は非常に大きな影響があるので、技術ではこの問題は解決できないと思います。

確かに今までは英語ばかりで、日本人は disadvantage を感じていたかもしれませんし、世界中のコミュニティに join するためには英語でないと難しいと思われていますが、だんだん AI などのツールが出てきて、バリアは低くなると思います。

●池内 どうもありがとうございました。力強いお言葉を頂いたと思います。自動翻訳ではテクニカルタームの問題や新語などへの対応が難しいという点はあるかと思いますが、技術的解決がその障壁をだいぶ減らしていくのではないかというメッセージだったと思います。

続きましては、坊農先生にお願いしたいと思います。質問への答えも併せてお願いできればと思います。

●坊農 まず質問への回答です。ResearchGate で日本人から何か来るということは 1 回しかなかったのですが、それはもちろん日本語でした。ただ、日本人だと、多分、電子メールで来てしまうので、ResearchGate で来るかと言われると、むしろ ResearchGate を使うようになるという方がバリアが高いのかなと思います。

また、共通言語としての英語の優位性ということは、私のいる生命科学はやはり英語で論文を出さないと業績にならないという文化がすごく根付いている分野です。これはもちろん感じているところです。おっしゃるように、最近 DeepL などのようなツールが盛ん

に使えるようになってきていて、だいぶバリアが下がっているのですが、どの言葉でも発信できるようになると思います。ただ、そうなってくると、やはり誰が発信しているのか、誰が言ったかということをきちんと分かるようにするべきだと思います。日本人は匿名で語りたがるので、もう少し日本人の研究者も自分の言ったことに責任を持ってやれるようになっていくと思うのですが、なかなかそういうところが厳しいのかなという気はしています。

●池内 ありがとうございます。ResearchGate は、意外と日本の研究者からは今のところは来ていないということですが、ResearchGate であれば匿名ではないやりとりになるので、また違ったコミュニケーションなのかと思います。ResearchGate も ResearchGate である種、閉じているというか、ある程度コミュニティの中だと思うのですが、そこを越えて実名で発信していくことも大事ではないかというコメントだったと思います。

それでは、引原先生からもお願いします。

●引原 英語に関しては、素養として勉強することは重要だと思います。それはどの言語でもいいと思いますが、人としての成長、英語を学んだ結果として研究者が論文を書くというのが一つのステータスだと思うのです。読む側にとっての敷居を下げるためにはいろいろな翻訳ツールを使っていくことは重要だと思いますが、書く側においては、自分の言葉として日本語で書くのもいいですし、得意な言語できちんと論理を構築して出していくことが重要であろうと私は思っています。それは日本語でも、英語でも、フランス語でも、ドイツ語でもいいのではないかと思います。また、ヨーロッパ系の言語でラテン語系と英語系なら、何となく似た言語は読めますよね。そういう感覚は重要だと思います。ですから、英語は否定しませんが、外国語として出すということは、研究者はやはりしていった方がいいというのが正直なところですよ。

●池内 研究者としてのコメントを、私も大変背筋が伸びる思いで聞いておりました。言語は何であれ、論理の部分をしっかり書けるようにすることが研究者としてはすごく大事だというお話だったと思います。ありがとうございました。

基盤とサービスの集中

●池内 二つ目の論点は「基盤とサービスの集中」です。今まではある程度限られたプラットフォームだったのが、いろいろなものが出てきたということで、本日の講演でもさまざまなお話を頂きました。こちらに関しても質問が幾つか来ています。

まずは河合先生への質問です。「各機関で整備されたりポジトリを一体として見せることができるような基盤が必要というお話でしたが、IRDB のブランディングや機能追加で目的を達成できる可能性はありますでしょうか」というご質問が来ています。河合先生、お願いします。

●河合 ポータルのようにする形で、例えば 800 ぐらいある機関リポジトリを一つにまとめることは、当然、技術的には可能だと思います。そうすると、見る側としても非常に便利です。アップロードする側も、ばらばらのものよりも日本で一つという感じのものに上げる方がモチベーションだと思いますので、これは技術的にも話としてもいいのではないかと思います。

●池内 ありがとうございます。続いて森本さんにお話を伺いたと思います。「F1000 をオールジャパンのものとして日本のゲートウェイとするのはいかがでしょうか」というコメントが来ています。いかがでしょうか。

●森本 よくぞ言ってくださいました。まさに私どもの学長が目指しているところはここです。どうしていけばそうなるのかという議論は既に始めています。

一方で、F1000 Research 社としてはプラットフォームの数も増やしていきたいですし、F1000Research のページを見ていただくと、プラットフォームとゲートウェイは数が非常に増えてきているので、数を増やすことが彼らの仕事となっている部分もあります。その辺の折り合いをどの辺に持っていくのが問題になってくるとは思いますが、ぜひとも拡大していくということになってきたら、また意見交換などをさせていただければと思います。

●池内 日本の代理店もできたということで、障壁もだいぶ下がったのではないかと想像します。ありがとうございました。

それでは、ブーケさん、基盤についていかがでしょうか。プレプリントについては、グリーン OA の一形態として推奨をというお話でしたが、コメントがありましたらお願いします。

●ブーケ 講演でお話したように、プレプリントはグリーン OA の非常に大事な形態だと思います。プレプリントは、古くはロスアラモス国立研究所のアーカイブサーバーの時代から無料で公開しているもので、典型的なグリーン OA の形です。査読されているジャーナルの論文の前の段階で、できるだけ早く情報をオープンにして公開のコメントや参照を集めるサービスは、やはり緑の形でないと無理だと思います。

また、プレプリントサーバーの経済のモデルは、ほとんどは非営利のグループや大きな大学が維持しています。しかし、エルゼビアやシュプリンガー・ネイチャーなどの出版社も、Research Square のような組織と協力してプレプリントサーバーを維持しています。それは、これからリサーチサイクルで利用するサービスのプラットフォームにもなり得るからです。プレプリントは出版する前の段階のサービスであると考えてそういうサービスを利用すれば、研究論文はだんだん改善され、より良いものになるので、出版社はそれをサポートしています。質問の答えになるかどうか分かり

ませんが。

●池内 実はそれは三つ目の論点でお伺いしようと思っていたのですが、先にお答えいただいたということで、ありがとうございました。

続きまして坊農先生、お願いします。

●坊農 私からすると、F1000 に出したのは、「出してみよう」ということでみんなで一致して出したのです。ですから、森本先生の流れとは全く別だったのですが、とにかく、いろいろな形態の論文があるのがジャーナル側の都合で出ないという状態になってきていて、私の研究なども割とメジャーではないので、そのようになっています。しかし、いろいろな形態でプレプリントサーバーができて、そういうものがきちんとインターネットで検索できるようになっていくと、これからすごくいい時代になっていくのではないかと思います。

●池内 研究者としての視点からのコメントをありがとうございました。

それでは引原先生、ご講演では、プラットフォームが多いのは決していいことではない、研究者はアーカイブではないのだというお話がありましたが、改めてコメントをお聞かせいただければと思います。

●引原 そもそも緑のオープンアクセスの概念は、そのコミュニティを再生するものだとは私は理解しています。今まで各大学や学会など、いろいろな小さなコミュニティがあったと思います。それは日本ではまだ残っているかもしれませんが、世界的にはほとんど崩壊しています。アジアの国などは、もうアメリカあるいはヨーロッパのコミュニティに属しないと研究者として存在できないという状況にあります。それはそれでいいかもしれませんが、文化として破壊している部分があるということは認めるわけです。それを再生するためには、オープンアクセスという手段を使っ

て、グリーンに自分たちのコミュニティを維持するプラットフォームが必要だろう、それは言語だけではなく、コミュニケーションとして必要だろうと思います。

そのときに重要なのが論文という話になるかと思うのですが、プレプリントは情報交換の手段であって、論文との違いはキュレーションがあるかないかだけです。そのキュレーションがまた特殊で、言語、グループに属している、あるいはインパクトファクターがどうだとか、何か利害が絡む部分なわけです。コミュニティは利害を位置付けるものかという、そうではないと思います。ですから、グリーンなプレプリントがコミュニティを再生できるのであれば、そういうプラットフォームがあってもいいだろうと思っています。

もう一つは、研究者自身が自分の論文の公開をマネージしていくというプラットフォームのサービスがあり得ますから、政治や経済に左右されないというのはいいことだと思います。機関に「ここの分野でやりなさい」と言われることも越えられると思いますし、あるいは大学がサブスクリプションしているかどうかとも関係ありません。研究者の自発的行動を促すものだと思いますから、プレプリントを中心に見れば、一つの手段を研究者が手に入れているのに、それをみすみす手放す必要はないのではないかと思います。例えばカンファレンスがあるのではないかとっても、今までそういうものも残ってきていません。あるいは、論文誌だっていつ消えるか分かりません。消えたものが見られないということは往々にしてあります。従って、今のコミュニティをどう維持するかということがアーカイブにも必要なかもしれないと思います。少し矛盾したことを言いますが、要するに、現状としてコミュニティをきちんと維持しなければ、科学の一つの分野が消えてしまうだろうと思っています。そのサポートという意味です。

●池内 ありがとうございます。例えばプレプリントサーバーの運営が立ち行かなくなって閉鎖されてしまうような事態もあります。これは完全にそのコミュニ

ティ側ではなく外のプラットフォームの論理で、今まで重ねてきた知が失われてしまう、コミュニティが解体されてしまうという状態だと思います。やはり研究者なり研究コミュニティの方がしっかりとハンドリングできるプラットフォームを持っていることがすごく重要なのかなと、お話を伺っていて強く思いました。

坊農先生が紹介してくださった ResearchGate など、そこに出すものや、やりとりをどうするか、発信をどうするかということが、坊農先生の意味でかなりハンドリングできているようでした。そういったものが、もう少しコミュニティレベルで、グリーン OA のプラットフォームのようなところでしっかりと持てるようになると、コミュニティの再生などにつながるという理解でよろしいでしょうか。

●引原 はい。ResearchGate もこちらでコントロールした方がいいと思います。そもそも私がこういう考え方に至ったのは、ジャーナル問題を何とかしなさいということで学術研究懇談会（RU11）で海外出張をさせていただいたのですけれども、そのときにアメリカの大学の幾つかと議論させていただいて、「やはり今は危機的な状況で、学内でもコミュニティが崩壊しつつあり、分野を越えて同じことをやっているのになかなかコミュニケーションが取れない。この状況は研究者個人個人がジャーナルの方を向いていることによる。データあるいはプレプリントの段階でお互いに議論ができないのは大きな課題だ」ということを何度か聞いたからです。それはそのとおりだと私も思います。

●池内 ありがとうございます。今後それがどういった基盤になるかは未知数ですが、例えば機関リポジトリなども期待が持てる場所ではないかと思います。

もう1人のモデレーターの山形さんは大学図書館に所属されています。大学図書館のスタッフとしてのコメントを一言頂けますか。

●山形 引原先生が、研究者の皆さんは研究をするこ

とが本分であって、アーカイバーではないとおっしゃっていましたが、まさにそのアーカイバーというのは図書館の役割の一つでもあると思います。また、機関リポジトリのこれからの可能性のところで河合先生に言及していただいたと思いますので、そういったところでは今後も協力できるのではないかと、まだ役割はあるのではないかと感じています。

●池内 ありがとうございます。本日、大学図書館からも多数ご出席いただいていると思いますが、考えるきっかけにいただければ大変うれしく思います。

限定的資金モデル

●池内 三つ目の論点は「限定的な資金モデル」です。私たちにとって頭の痛い APC の問題などがずっとありますが、これに対する出版社からの見解や、機関リポジトリによるプレプリント公開の可能性のお話もありました。森本さんには、「F1000 に関して、論文投稿の費用についてはお話しいただいたと思いますが、プラットフォームの費用などについてもお伺いしたい」という質問が来ています。

それでは、まずは河合先生からお願いします。

●河合 例えば契約価格が高騰しているということであれば、間接的なアプローチではありますが、コミュニティとして OA を進めていくということもあるかと思えますし、他の策でもいいと思うのですが、何らかの策を持ち合わせていることによって、出版社との交渉も少しは有利に進めていくことができる可能性が出てくると思います。まずはコミュニティとしてどうしていくべきかという大きな方針を明確化することが重要だと思います。

●池内 ありがとうございます。それでは、森本先生、お願いします。

●森本 F1000Research の論文投稿の費用については、

Google で「F1000Research APC」と打っていただくとすぐに出てくと思いますが、実は 2020 年 9 月 1 日に料金の改定がありました。文字数に応じて料金が決まっていたのですが、特にプラン S の新しい指針が出され、その影響を受けて、現在は、三つの料金形態となっています。

もう 1 点はプラットフォームに関する費用についてのご質問ですが、F1000Research の場合はプラットフォームとゲートウェイがあり、ここで大きくインシャルコストが変わってきます。プラットフォームを作る場合は、自分たちでルールを作るので、かなり自由度が高くなり、どういったプラットフォームにしたいのかということを反映することができます。その分、少しお高いというわけです。ゲートウェイについては、筑波大学ゲートウェイの場合は F1000Research のプラットフォームにぶら下がっている状況なので、F1000Research のプラットフォームのルールに従います。そこにゲートウェイというものを付けて、今回、初めての試みで、日本語も投稿可能なゲートウェイに構築しました。ランニングコストとしては、保守・メンテナンス費用が 1 年ごとにかかってきます。

●池内 詳細なご説明をありがとうございました。

次に、ブーケさんへの質問です。これはすごくストレートな質問で、「プレプリントは結局 APC から出ているんじゃないの？」というのは皆さん考えられるところかと思いますが、率直なところをお聞かせいただければと思います。

●ブーケ 大変面白い質問です。つぶれたプレプリントサーバーはまだ存在していませんが、大体コミュニティのサービスとして維持されているので、大手の出版社も少しずつプレプリントサーバーに興味を持ち始めています。F1000 は大変面白いモデルだと思います。科学ジャーナルとプレプリントサーバーのハイブリッドで、プレプリントのみをホスティングするのは無料ですが、ピアレビューの査読サービスと論文を改善す

るサービスがあれば APC がかかってしまいます。このモデルは増えると思います。それである程度のコストを支持できますし、プレプリントサーバーがずっと継続できる保証になります。

科学出版社は、過去に出版された論文のレコードのアーカイブを保存する約束を守ることが必要です。例えば学会と契約するとき、学会誌を出版するときは、必ず Portico のようなサービスで、もし出版社がつぶれた場合、Portico の Web サイトでそのコンテンツを永遠に保存するので、そのコストがあります。科学論文の数は毎年どんどん増えているので、確かにそれを保存するメンテナンスのコストは毎年増えています。

これからプレプリントサーバーでも同じような問題が発生すると思います。プレプリントを出してもジャーナルの記事が出てくるのだから、プレプリントがなくなってもいいのではないかと思われる人もいますが、プレゼンで言ったように、リジェクトされてもアクセプトされてもプレプリントは残ります。それは、自分が最初にこの結果を出したという証拠としてずっと残るべきだと思います。ですから、これからプレプリントサーバーもずっと保存するコストは発生するので、何か経済的なモデルを作らないといけません。慈善で維持するか、政府が維持するか、コミュニティが維持するか、APC の形で投稿するときに少額のお金を払うか、どれがいいかわかりませんが、確かに価値があるサービスを継続的に続けるようなサポートが必要になると思います。

●池内 プレプリントの長期保存のコストに関する議論はまだ十分に行われていないように思います。新たな論点が出てきたということで、また将来の SPARC のネタが一つできたかなという気がします。そのコストを誰が負担するのか、誰が持っていくのか。Portico の話が出ましたが、ああいったバックアップも必要なのかということも含めて、SPARC のようなさまざまなステークホルダーがいる場で長期的に考えていく必要がある課題が出されたと思います。ありがとうございます。

ました。

続きまして、坊農先生、お願いします。

●坊農 私ども生命科学の研究者は、やはり自分たちで取ってきた競争的資金で APC を賄わないと論文は出せないと思っています。プレプリントは、今のところ bioRxiv は無料なので、そこに出そうという流れになっていて、もっと利用が進むかと思ったらそうでもないというのが今の私の印象です。やはりそこは評価軸が変わっていかないとなかなか変わらず、そういうものは急には日本では変わらないのだなと痛感している段階です。

●池内 意外にもお金の面やコストの面だけではドライブしないというか、プレプリントへの投稿があまり促されていないというところかと思っています。一方で、坊農先生の講演にもあったように、論文として出版されないものであってもプレプリントとしてずっと残っていくものも確かにあり、それがプレプリントサーバーに割とみんなに見える形、検索もされる形で残るといことは、もう少し広めていきたいとも思いました。ありがとうございました。

それでは、引原先生、お願いします。

●引原 難しい問題です。arXiv.org 自身が最近では大手ユーザーの国や地域でシェアしてお金を払うというモデルに切り替わっていますが、やはり National Science Foundation (NSF) やゲイツ財団などのサポート部隊があって何とか慈善で成り立っている部分と、ファンドで成り立っている部分があります。それを利用者が一部負担するというのは正しいと思いますが、その価値を皆さんが認める方向の試みが少ないとは思いますが。ですから、受益者負担にすると、多分、尻すぼみになってしまいます。その価値を認めて、それによって他のプレプリントが生きてくるということもあると思うので、そういう動きをしないといけないのではないかと思います。

●池内 これも長期的な課題ということで受け止めました。今後 SPARC Japan セミナーで2年後、3年後、10年後に議論するときに、どのような展開になっているのだろうと考えました。ありがとうございました。

学術雑誌ベースの評価という偏狭な視点

●池内 四つ目の論点は「学術雑誌ベースの評価という偏狭な視点」です。これは評価の話で、皆さんの講演の中にもいろいろと出てきました。従来型のインパクトファクターや学術雑誌ベースの評価だけではない新しい評価が必要だということはずっといわれていたのですが、ここに対して、プレプリントあるいはiMDのような評価指標がどのような役割を果たせるかについてご意見を伺いたいと思います。河合先生からお願いします。

●河合 まず出版の形態などに関してですが、発表で示したように、自由度の高い公開が可能であれば、研究者の意思に反して出版の形態が制約されるということは減ってくると思います。また、インパクトファクターは当然説得力のある指標だと思いますが、先ほど坊農先生が紹介してくださったように、閲覧回数、ダウンロード回数、SNS 上での言及回数など、指標に関しては現時点で既にかなり多様化してきているという認識があります。

●池内 ありがとうございました。

続いて、「そうはいっても、やはり研究者は有名なジャーナルに掲載されることに価値を置かれているかと思うのですが、この辺について、筑波の研究者の方々からは何か反響はあるでしょうか」という質問が来ています。森本さん、お願いします。

●森本 若手の先生、人社系の先生、理工系の先生と、大きく三つに分けられると思います。理工系の先生、特にベテランの先生は、もう静観者で、「いいのができたね」「そうなんだ」という意見を頂いています。

人社系の先生については、ドイツ語で書いている先生やフランス語で書いている先生などいろいろいらっしゃって、中には『源氏物語』の研究をしている先生が古語を交えながらというものもあります。要するに、どの言語で自分の研究を発信すべきかということは、その研究者が一番よく分かっているのです。つまり、どのコミュニティに一番読んでほしいのかということです。その言語で書くことが、自分の研究をどんどん尖らせていく。そのコミュニティでどんどん議論していきたいということでその言語を選択されているので、そういう先生たちに「先生、やはり英語で書いてください」というのはなかなか難しいところがあります。

筑波大学ゲートウェイはまだ日本語ですが、とはいえ世界大学ランキングでは Scopus が採用されています。そういった世界的なコミュニティが使っているデータベースに自分の論文が収録されること、英語で書かなくても収録されることについては大変興味を持っています。幸い、その投稿料を補助するプログラムもあります。人社系の先生からはかなりご質問も頂き、反響も大きかったので、人社系の教員を対象に説明会を実施したりもしました。

最後に、非常に面白い意見が若手の先生から出ています。任期付きの助教や、テニュアトラックに乗っている助教の先生、助教はほとんどそういうことなのですけれども、やはり従来の雑誌ですと、査読期間が半年、長いもので2年ぐらいかかるときがあるのですが、その間に自分の研究室を移動することは多々あります。そのときに、レビューから再実験するように言われて、もう研究室を移動してしまったので、とてもではないけれども再現実験できない。では、その論文はお蔵入りになるのかということになる。そういう論文が既に2、3本ありますということでした。そういう意味では、筑波大学のゲートウェイは、査読は、出版されてから始まるのです。最初は awaiting peer review という状況で公開され、そこで査読が始まるので、早く出版 (printed、published) されているという状況になり、2年近くも公開されないということは避けられま

す。ですので、若手の先生は結構興味があるのではないかという話を伺っています。

●池内 ありがとうございます。もう1件、質問が来ています。「iMD を筑波大学のゲートウェイに適用した場合、どうしても筑波大学の先生方に偏るのではないか。そうすると比較的数値が低くなってしまいますが、研究者にとってアドバンテージになるのですか」というご質問です。

●森本 筑波大学ゲートウェイは、親分がF1000Research プラットフォームなので、多様性はかなり高くなります。従って、筑波大学ゲートウェイに出したからといって、いつも自分が投稿している雑誌よりも iMD が低く出るような雑誌に投稿するということにはなりにくいと思います。

●池内 ここは何か誤解を生みそうな点でもあると思ったので、質問していただけてとても良かったです。どうもありがとうございました。もし他にも一言ありましたらお願いします。

●森本 学内では、従来のジャーナルとF1000Research、つまり筑波大学ゲートウェイを、ブティックとコンテンツに例えています。ブティックに行くと、いろいろなメーカーの服や靴が置いてあります。つまり、その店にどの商品を並べるかという目利きの方がいらっしゃるのです。どの論文を掲載すべきかという目利きの人がいるのが従来のジャーナルです。一方、筑波大学ゲートウェイや F1000Research は、どの論文でも掲載するので、楽天のような感じで、あまたある商品の中から自分でいいものをチョイスするようになっているのではないかと考えていました。しかし、よくよく考えると、最近 Amazon's Choice みたいなものが出てきて、新しい指標が出てきたのは確かなので、筑波大学ゲートウェイも盛り上がってくると、そういうチョイスが出てくるかもしれないと思っています。

ます。

●池内 大変面白いお話をありがとうございました。それでは、ブーケさん、お願いします。

●ブーケ この質問の基本的な意味は、プレプリントとピアレビュー・ジャーナルがどのように違うかということだと思いますが、論文を書く人の立場と読者の立場では、意味が若干違うと思います。論文を書く研究者にとってピアレビューのジャーナルの役割は、一つ目は「最初にこの研究をしました」という報告をすることです。それはとても大事です。二つ目は研究結果の公開です。査読のプロセスが終わった後で初めてジャーナルに公開するので、レビューのスピードはとても大事です。三つ目は認知度とビジビリティです。ジャーナルで出版すれば、自分の研究が認知されて、他の研究者に認められるということです。これらの三つの役割があります。

プレプリントの特徴を見ますと、この三つの役割は、ある程度は達成できると思います。例えば、最初に研究を報告するということはジャーナルより早くできます。そして、成果をジャーナルより迅速に公開できます。また、成果に対して早い段階で早期に認知度と可視性が提供できます。それから、より早くフィードバックできます。従って、論文を書く人にはある程度、プレプリントがその三つの大事な役割を達成できると思います。

では、読者に対してのピアレビューのジャーナルの役割は何かというと、クオリティとインパクトの決定です。クオリティについては、実験をきちんと行った、メソッドは再現性がある、データはみんな見られるなど、基本的な正しさを査読のプロセスでチェックします。私も昔は研究者でしたが、たくさんリジェクトされた経験があります。査読のプロセスは論文とサイエンスの結果を改善するのにとても重要で、査読者のコメントをもらわないと、やはり研究は良くなりません。自分が査読者になったときには、そのプロセスで自

分の経験と知識を著者に伝えられます。正しいかどうかは大事ですが、いろいろなスタディを見ると、やはり7割の論文はある程度正しいでしょう。「PLOS ONE」や「Scientific Reports」のレベルでしたら、インパクトがあってもなくても出版します。しかし、インパクトを決めるのはとても大事な役割です。確かにジャーナルはたくさんありますが、インパクトによって分けられます。例えば「Nature」「Science」「Cell」などインパクトの高いジャーナルは、正しくてもインパクトがなければそのジャーナルに出版しません。ジャーナルは、コミュニティに対してインパクトがあるかどうかで出版するかどうかを決めます。

インパクトで決めるので、読者は信頼できる研究をすぐ早く見つけることができます。読者の一番貴重な指標は時間です。時間は限られていて、全部読むことは不可能なので、ジャーナルのインパクトを見ます。インパクトはインパクトファクターとよく間違えられますが、インパクトファクターはただの popularity、人気度をメジャーするものです。本当のインパクトは、社会や研究コミュニティにインパクトがあるかどうかということを表すもので、それはジャーナルが決めます。そして、そのジャーナルにインパクトがあるものが出るので、読者は貴重な時間を使ってそのジャーナルを利用するわけです。

今、この二つの役割はプレプリントにはありません。もしコミュニティがオープンピアレビューなどのコメントの正しさとインパクトをある程度メジャーできる時代になったら、だんだんジャーナルに近くなると思います。しかし、今のところはそうでもありません。なぜ APC のお金がかかるかというと、やはりインパクトと正しさをチェックするプロセスがあるからです。ジャーナルはそのプロセスを管理しています。

もう一つ言いたいことは、プレプリントは論文の卵だということです。プレなので、プリントの前の段階です。プレプリントサーバーのようなところで、ジャーナルで絶対に出版しないものが出版できるかというのが大事な質問です。例えば、negative result や研究の

ノートブックには大事なレコードがありますが、ジャーナルはそういう情報はなかなか出版できません。そういうものは、研究論文のような査読の必要はありません。ただ記録として残したいようなものはプレプリントサーバーの形が非常に適していると思います。

●池内 ジャーナルの評価がある一方で、プレプリントやノートブック、失敗事例の共有のようなところも出てきているということですね。ありがとうございます。

それでは、坊農先生、お願いします。

●坊農 もしここに公的予算などの審査の基準を作っている方がいたら、ぜひ査読論文だけを評価しているということを変えていくようなことをしていただきたいと思います。いつも言っているのですが全然変わらないので、そのように変えていってほしいという、それだけです。

●池内 ありがとうございます。

それでは、引原先生、お願いします。

●引原 学術雑誌ベースの評価というのは言い方が微妙なのですが、評価をする側がフェアであり、透明性があって、誰から見ても一点の曇りもないということが認められないと、その評価は正しくないはずですが。しかし、今、例えば雑誌のインパクトファクターには多くの不正があるのも事実です。それは、エディトリアルフェアネスが確保できていないから、レビュアーのフェアネスが確保できていないからです。ScopusかWeb of Scienceか分からないですけども、そういうプラットフォームがそのまま放置しています。

極端な例ですと、論文誌を収奪して新しい論文誌のような顔をして載せてしまったものを放置しています。これは日本では有名な話で、大学の紀要論文を、名前だけそのまま使ってサイトを書き換えたというものがあります。そういうものは、フェアという観点からは

すぐに処置すべきだと思うのですが、1 年たっても放置しています。それで本当に評価する側の立場を保てるのかと思ってしまいます。ですから、インパクトがある、あるいは評価があるということを主張するのであれば、その評価をきちんと管理しないと、それは維持できないのではないかというのが私の個人的な意見です。

●池内 ありがとうございます。

いろいろな論点が出てきて、その他の質問までは取り上げられず、失礼しました。プレプリント、結果の早期共有、査読プロセスの透明化といったことも出てまいりました。今後のプレプリントのテーマとして、引き続き SPARC で取り上げていければと思います。

最後に山形さんから一言、お願いできますか。

●山形 皆さま、長い時間ありがとうございました。今回のお話はプレプリントが中心ということで、大学図書館にはあまり関係ないと感じる方もいらしたかと思いますが、最後の学術雑誌ベースの評価という部分にも関係するかもしれませんが、プレプリントが研究者同士の間でどう評価されるかということだけではなく、世間からどう評価されるか、あるいは学生や一般の方がどのように見るのかというのは、図書館員の力が生かせるところではないかと思います。例えば教育・啓蒙活動などで関係していくことができるのではないかと感じました。皆さま、本当にありがとうございました。

●池内 それでは、これにてパネルディスカッションを終了します。どうもありがとうございました。

第3回 SPARC Japan セミナー2020

「初めての研究データ」

開会挨拶 / 概要説明

八塚 茂

(バイオサイエンスデータベースセンター)



八塚 茂

国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター研究員。システムエンジニア等を経て、2015年10月より現職。生命科学データのリポジトリ運営に従事。研究データ利用協議会（RDUF）傘下のジャパン・データリポジトリ・ネットワーク（JDARN）代表、2020年度SPARC Japanセミナー企画ワーキングメンバー。



セミナーの企画概要

ちょうど1年前のSPARC Japan 2019の中で、「実践研究データ管理」というセミナーを開催しました。このセミナーは、単なる理論ではなく、参加者が明日からできる研究データ管理をイメージすることを目的として開催したものです。その中では、データリポジトリや図書館の実例、研究データ管理を支援するためのツール、研究データ管理を担う人材育成の取り組みについて発表が行われました。パネルディスカッションでは、研究データの管理における研究者、研究支援職員、図書館職員の役割や、役割を超えた全ての関係者が共有すべき理念やリテラシーについて議論しました。

今年はその第2弾として、研究データの管理・公開のための実践知を共有することを目的としてこのセミナーを開催しています。特に今回は、研究データの管理や公開をこれから始めようとしている方々を応援したいと考えています。そして、これを始めるに当たって検討しなければならないことがたくさんあるので、課題やその対応方法について皆さんと共有したいと思います。

本日は、まず研究データの管理・公開の実例や課題を知るということで、日本の大学における実例を三上

様から発表していただきます。それから、海外（ヨーロッパ）の事例については神谷様からご講演いただきます。次に、研究データのリポジトリの評価方法や日本の現状についての調査実績を安原様からご講演いただきます。そして、実際に研究データ管理を行うための便利なインフラが既にできています。その代表格として、J-STAGE Data について加藤様から、GakuNin RDM について込山様からご説明を頂きます。そして最後にディスカッションという順番で進めたいと思います。

第 3 回 SPARC Japan セミナー2020

「初めての研究データ」

データリポジトリ J-STAGE Data でのデータ公開

加藤 斉史

(国立研究開発法人科学技術振興機構)

講演要旨

国立研究開発法人科学技術振興機構（JST）は J-STAGE 利用者を対象とするデータリポジトリ J-STAGE Data のパイロット運用を令和 2 年 3 月より開始し、令和 3 年 3 月に本格運用に移行する予定である。J-STAGE Data サービス概要およびパイロット運用状況について紹介する。



加藤 斉史

教育サービス会社でeラーニング教材の開発業務に従事した後、2000年に科学技術振興事業団に入職。現在、科学技術振興機構情報基盤事業部にて、J-STAGEおよびJ-STAGE Dataの運営業務に従事。

はじめに

本日の話の内容は三つに分かれています。まず、J-STAGE Data 導入の背景として、データ共有が広まってきた状況についてお話しします。続いて、J-STAGE Data の概要として、データの登載から公開までの流れ等について紹介します。最後に、現在までパイロット運用を行っていますので、その取り組みと結果について紹介します。

J-STAGE Data 導入の背景

研究データ公開が必要とされる背景には、オープンサイエンスやビッグデータを活用したデータ駆動型研究の潮流があると考えています。データの共有、利活用による新たな価値を創造する取り組みが期待されています。

その一例として、欧州におけるオープンサイエンスの対応状況について 2019 年に調査を行いました（図 1）。32 カ国のポリシーを確認したところ、オープンサイエンスに関するポリシーがある国が 2017 年には

12 カ国だったのが、2019 年には 18 カ国に増えていきます。このうち、データに関するポリシーがある国は、2017 年の 12 カ国に対し、2019 年には 15 カ国に増えています。

また、研究の不正防止や透明性担保のために、多くの大学や研究機関、研究資金助成機関等が、データの管理・公開に係る方針を策定しています。図 2 は、アメリカやヨーロッパ、アジアの代表的な助成機関の対応状況についてまとめた表（2019 年時点）です。10

OS・OAに対する欧州諸国の対応状況

オープンサイエンス・オープンデータについてのポリシーの有無

	2017年3月	2019年8月
ポリシーの有無	12/32	18/32

(欧州の32カ国：うち非EU 4カ国)

うち、データ、サンプル、ソフトウェアに対するポリシーの有無

	2017年3月	2019年8月
ポリシーの有無	12/32	15/32

参考文献：SPARC Europe's Annual Report (2019)

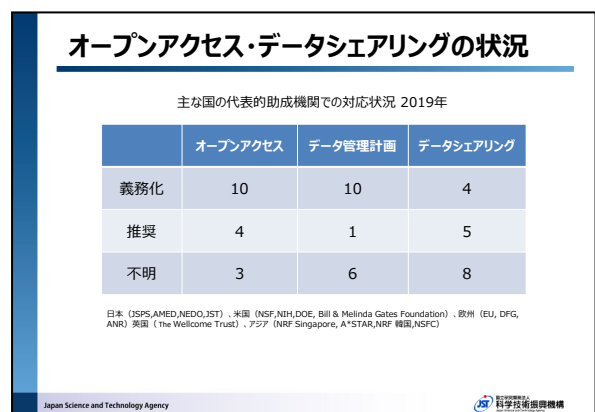
Japan Science and Technology Agency

国立研究開発法人科学技術振興機構

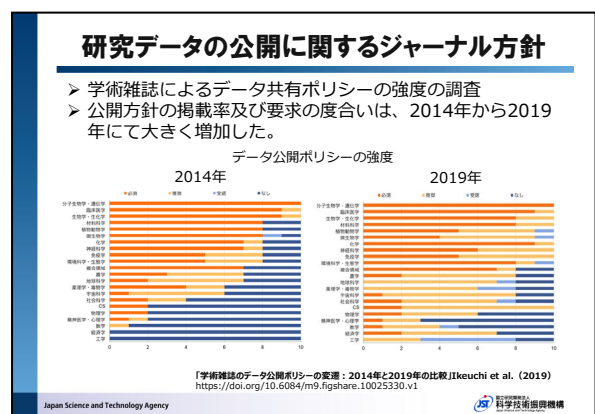
(図 1)

機関がオープンアクセスやデータ管理計画について義務化しており、進んできていることが見て取れます。データシェアリングについては、義務化4機関、推奨5機関と、少し強度がばらついています。これは、データの特性が分野によって異なることなどから、ポリシーの強度がばらついたのではないかと考えています。このように、データ共有についてもポリシーを定める機関が増えてきています。

次に、データの公開や共有に関するポリシーの整備がジャーナルでも進展していることについて紹介します。図3は、学術雑誌220誌に対してデータ共有ポリシーの強度の調査を行ったものです。左が2014年、右が2019年で、公開方針について、必須・推奨・受託・なしを色分けして表示しています。青色がデータ公開ポリシーなしですが、2014年から2019年になると、ほとんどの分野においてデータポリシーなしが減ってきていることが見て取れます。



(図2)



(図3)

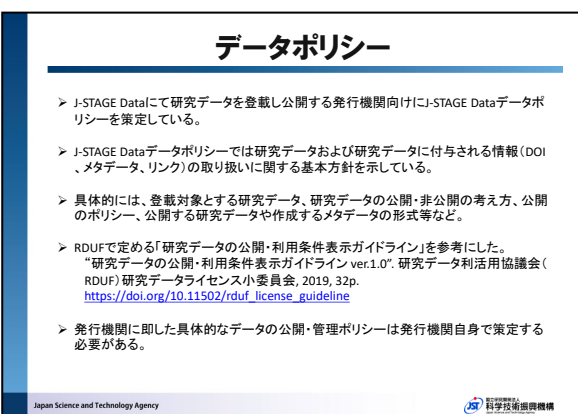
以上のように、研究者が研究成果論文を発表する際、その根拠となるデータ公開を求められる場面が多くなっています。

このような状況を受けて、J-STAGE Data を立ち上げ、J-STAGE の掲載記事に関連するデータを登録・公開するデータリポジトリを提供することにしました。その方針としては、全てオープンアクセスで公開し、登録データには DOI を付与するという規格にしました。

J-STAGE Data の概要

J-STAGE Data は2020年3月にリリースし、現在はパイロットジャーナルによる試行運用を行っています。登録できるデータは、研究成果論文の根拠となった研究データや、その記事を補足する図表等のデータです。機能としては、J-STAGE 掲載の記事との相互リンク、他のデータベースからのアクセス、Google Dataset Search や Dimensions からの検索が可能です。

また、J-STAGE Data を立ち上げるに当たり、推奨するデータポリシーを策定しました(図4)。簡単に言うと、研究データや研究データに付与される情報(DOI、メタデータ等)の取り扱いに関する基本方針です。もう少し具体的に言うと、J-STAGE Data への登録対象とする研究データや、その研究データの公開・非公開の考え方、公開のポリシー、公開する研究データや作成するメタデータの形式などをポリシーの中で定めています。このポリシーを定めるに当たって



(図4)

は、研究データ利活用協議会（RDUF）の定める「研究データの公開・利用条件表示ガイドライン」を参考にしています。なお、発行機関に即した具体的なデータの公開・管理ポリシーは発行機関自身が策定する必要があります。

今、発行機関という用語を使いましたが、少し分かりづらいと思うので、データポリシーの中で書かれている状況を整理した図をお示しします（図 5）。データポリシーの適用範囲は、発行機関が J-STAGE Data に掲載し、公開した研究データとなっています。この定義は詳しく紹介しませんが、発行機関が研究データを J-STAGE Data に掲載することは少し後のフローにも出てくるので、頭に入れていただければと思います。

J-STAGE Data への掲載対象となる研究データは、J-STAGE で公開されている、または公開予定の記事・論文にひもづくデータ、すなわち論文記事の根拠となったデータ、あるいは研究の実施過程で得られた情報および関連情報（表、図、音声・動画、ソフトウェア、プロトコル、その他ポスターセッションやプレゼンテーションの資料など）です。

公開・非公開については、著者および発行機関が判断して決定することとしています。その判断の際に留意すべきデータ公開への制約があります。研究データに含まれる内容（機密性、プライバシー等）や、研究当事者の要望等によって、データ公開に制約が生じる場合があります。例えば分野・研究コミュニティなどの慣習で公開制限が一般的な場合や、個人情報を含

む場合、国家安全保障等に係る場合、共同研究契約等の契約による制限がある場合、所属機関や研究助成機関などによりポリシーが定められている場合などが関わることがあるので、このようなところに留意しながらデータ公開をしていただくことになります。

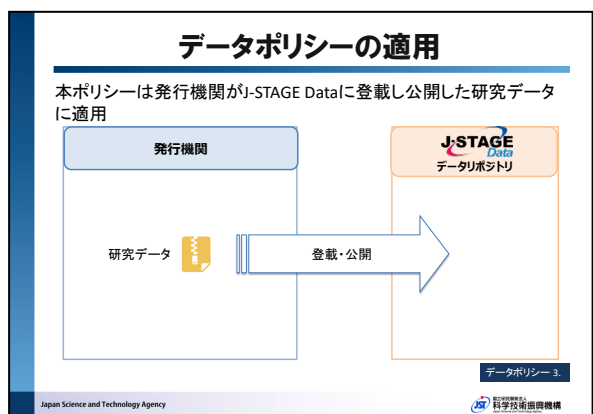
なお、これらの制約の中には、データに適切な処理（個人情報の匿名化処理）を施すことや、一定期間が過ぎるとデータを公開することができる場合もあります。

公開のポリシーとしては、原則としてエンバゴ期間を設定することは推奨していませんが、何かの事情でやむを得ない場合は、研究データに一定期間非表示にするためのエンバゴ期間を設定することができます。エンバゴ期間中はメタデータのみが表示されます。推奨では、エンバゴ期間は 12 カ月以内としています。

さらに、研究データを公開しない場合でも、研究データの内容の詳細（データ収集方法やデータの性質、精度等）や問い合わせ先をメタデータに記述してメタデータのみ公開することができます。これは、データが存在したことを証跡として残すことで、将来の研究活動を支えることができると考えているからです。

公開する研究データについては、できるだけ再利用可能なように、特定のアプリケーションに依存しない形式（例：PDF より xls 形式、xls 形式より csv 形式）、あるいは研究コミュニティの中で標準化されたものがあればそれにのっとることで、相互運用性が高まると考えていただければと思います。作成するメタデータについては、データの収集方法やデータの性質等、研究データの内容をできるだけ詳細に記述していただければと思います。これも、語彙等について標準化されたものがあれば、それに従うのがよいと考えています。

続いて、研究データの公開のワークフローについて説明します。図 6 の一番左は、J-STAGE に論文を投稿してから公開に至るまでの線です。ジャーナルを発行する機関が原稿を受け付けて、査読・編集した上で J-STAGE への掲載作業を行います。図の右側は、研究



(図 5)

データを登録していく流れです。矢印が2本あるのは、2パターンあると考えているからです。左側の青い線は、著者が記事とセットで投稿し、発行機関が登録するという流れです。こちらは著者が発行機関に研究データを渡し、発行機関が査読・編集を行った上で、J-STAGE Data への登録から公開まで行うものです。右側の青い線は、著者が J-STAGE Data へ投稿し、その上で発行機関が査読して公開作業を行うという流れです。このように、研究データの公開には大きく分けて二つのパターンがあると考えています。

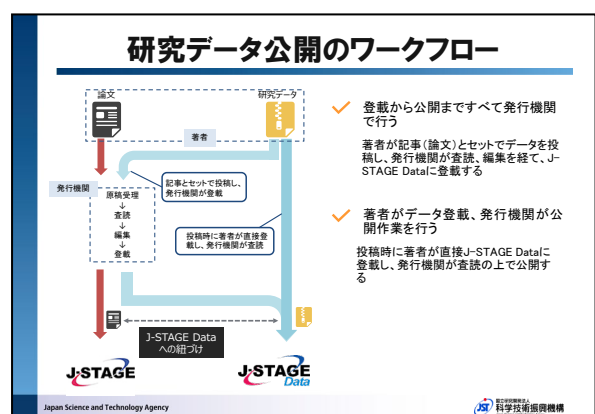
実際に研究データを登録するときの画面を紹介します。図7は登録するデータファイルをアップロードするための選択画面です。メタデータ（タイトル、アイテムタイプ、著者名等）についてはここで記載します。ライセンスについては、CC ライセンスを基本として、記載のものから選択するようになっています。

ここまでがデータを登録する機能についての紹介で

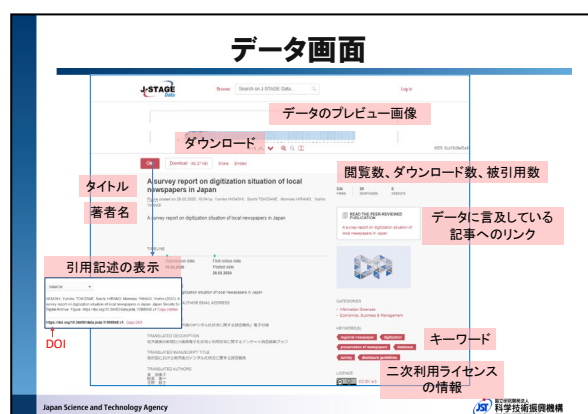
す。ここから先は、データを閲覧する画面等の紹介をしていきたいと思います。

まず、J-STAGE Data のトップ画面を開くと、幾つかの最新の登録情報等がサムネイルと共に表示されます。サムネイルをクリックすると、図8のようなデータ画面にいきます。ダウンロードと書いてあるところからデータのファイルをダウンロードします。ダウンロードの左にある「Cite」と書いてある赤いボタンをクリックすると、このデータを引用する際の引用記述を見ることができます。この中に DOI も記載されています。その下にタイトルや著者名、その他メタデータ等が表示されています。右側には閲覧数やダウンロード数、被引用数、データに言及している記事へのリンク、キーワード、二次利用ライセンスの情報等が表示されます。

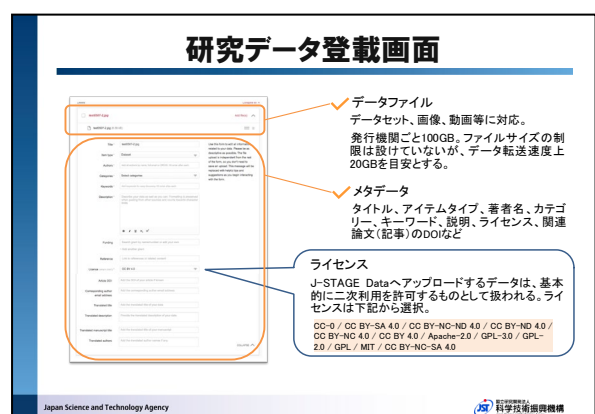
このようなデータ画面と J-STAGE の相互リンクの状況について説明します。図9の左側が J-STAGE の



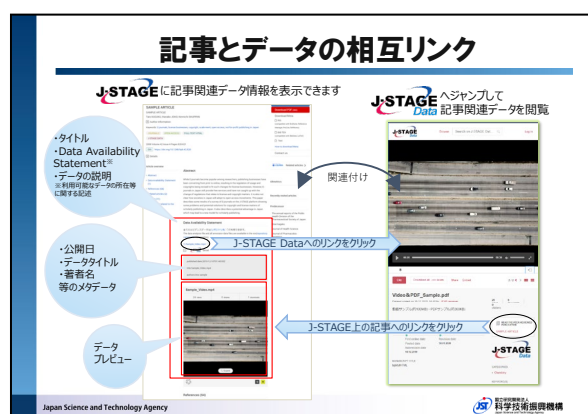
(図6)



(図8)



(図7)



(図9)

画面です。J-STAGE の記事情報の真ん中あたりにアブストラクトがあり、その下に J-STAGE Data のデータ情報が掲載されます。タイトルや、Data Availability Statement といって、利用可能なデータの所在等に関する記述を掲載することができます。また、データの説明等についてもここに記載可能です。公開日、データタイトル、著者名等のメタデータが表示された後に、データプレビューも表示されます。

J-STAGE Data へのリンクをクリックすると、該当する J-STAGE Data のデータ画面に飛ぶことができます。J-STAGE 上の記事へのリンクはデータ画面の右側に表示されます。

さらに、Google Data Search や Dimensions から J-STAGE Data にリンクができますが、これらは全てのデータではなく、データ種別によります。

パイロット運用の結果

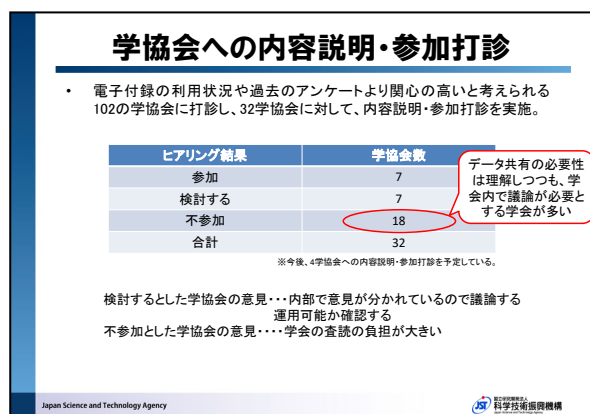
続いて、パイロット運用の結果を紹介します。パイロット運用期間は 2020 年 3 月から始まっており、2021 年 3 月までを考えています。本格運用に先立ち、必要となる運用フローや登載ガイドラインの策定、事業運営等の検討を行うためにパイロット運用を行い、10 誌程度のパイロット参加を受け付けています。

図 10 は現在のパイロット参加状況です。8 誌が参加しており、既にダウンロード数の増加や、他のデータベースでのインデックス等、研究成果発信力の向上が見られています。ダウンロード数の増加等について

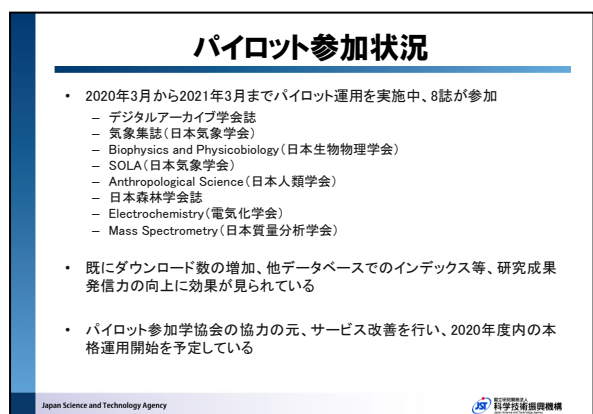
はこの後紹介します。パイロット参加学協会の下、サービス改善を行い、2020 年度内の本格運用開始を予定しています。

パイロット運用期間中に学協会に J-STAGE Data の説明や参加打診を行ったので、その状況について紹介します（図 11）。電子付録の利用状況や過去のアンケートから、関心の高いと思われる 102 の学協会に打診し、話を聞いてもいいという回答を頂いた 32 学協会に対し、内容説明、参加打診を実施しました。その結果、「参加」が 7 学協会、「検討する」が 7 学協会、「不参加」が 18 学協会となりました。「検討する」という学協会については、データ共有の必要性は理解しているが、学会内で議論が分かれており、検討に時間がかかりそうだということです。「不参加」という学協会からは、学会の査読の負担が大きいという意見が寄せられました。

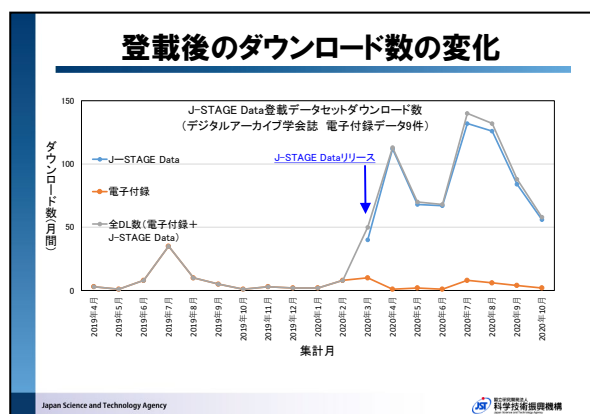
図 12 のグラフは、登載後のダウンロード数の変化



(図 11)



(図 10)



(図 12)

を表したものです。電子付録のときと比べて、J-STAGE Data に登載するとダウンロード数が増加していることが分かります。

参考：第3回 J-STAGE セミナー

2021年3月1日にJ-STAGE セミナーを行い、「ジャーナルから見た研究データ」と称して、参加学協会の事例や、国立研究開発法人における研究データの公開の取り組みなどを紹介する予定です。J-STAGE のイベント情報から申し込みができます。今年度は通年で研究データについて取り上げています。本日の内容の補足にもなると思うので、ご覧いただければと思います。

●八塚 早速、二つ質問を頂いています。一つ目は、「データを公開する際に、個人情報や機密が含まれたデータについては制約を課すというご説明がありましたが、こういった機密等が含まれていて完全公開できないようなものは、研究者が判断するのか、それともJ-STAGE 側でチェックしていただけるのか、どちらになるのでしょうか」という質問です。

●加藤 研究者側で判断していただきます。J-STAGE の登載時に、J-STAGE Data 側でのチェックはかけていません。

●八塚 あくまでも研究者側で判断していただく。従って、その責任も研究者側に課せられるということですね。

●加藤 そのとおりです。

●八塚 次の質問です。「メタデータは、あらかじめ用意されている項目のみ利用できるのでしょうか。あるいは、雑誌や学協会別にオリジナルのデータ項目も

用意することが可能なのでしょうか」。

●加藤 設定項目については、こちら側で用意した標準的なメタデータ項目を使っていただきます。項目を加えることはできません。

●八塚 あらかじめ用意したメタデータ項目のみということですか。

●加藤 はい。自由に項目を増やせるということではありません。

●八塚 続いて、「先ほどのご説明の中で、著者、研究者自身が J-STAGE Data にアップロードできるようなフローがありましたが、その中で、例えば論文がリジェクトになったのにデータが載ってしまって具合が悪いという場合には削除できるのでしょうか」という質問です。

●加藤 基本的に、著者が直接アップロードした場合についても、最終的な公開承認はジャーナル側、発行機関が行うことになっているので、論文がリジェクトされることが想定されていれば公開しないのではないかと意味で、起こりにくいケースだと思います。

（後日注：当日パネルディスカッションで補足しましたように、リジェクトの場合、発行機関の中で査読審査をしている中でリジェクトになっていますので、データ自体、まだ公開されていません。また、メタデータも公開されていない状態ですので、その状況の中で削除するかどうかは発行機関が決めることができます。）

なお、公開されたデータについては、できるだけ保持することが望ましいですが、削除になったときについてもポリシーの中で取り決めがあり、削除したときに削除した旨を設けることで、研究データ自体を非公開扱いにすることはできると記載しています。

●八塚 次は、「データファイルのサイズは発行機関ごとに 100 ギガとありましたが、それを越えたファイルサイズのデータは扱えないのでしょうか」という質問です。

●加藤 パイロット運用期間中は 100 ギガバイトを目安として進めています。100 ギガバイトでは少ないという声も学協会から頂いているので、対応を考えていきたいと思いますが、現状は 100 ギガバイトとしています。

●八塚 続いての質問です。「ダウンロード数が増えたのは、学協会の参加の呼び掛けで皆さんがアクセスした、あるいは関係者の皆さんがアクセスしたといったものが含まれていませんか」。最後のグラフのご説明のところだと思います。

●加藤 それは多分にあると思います。詳しく分析していませんが、初速があったことは否定できません。

●八塚 次に、「既に J-STAGE に掲載されている論文について、後付けでデータを登録することはできますか」という質問を頂いています。

●加藤 はい。できます。実績もあるそうです。

●八塚 次は、「論文自体がもし J-STAGE に登録されない場合は、データも削除されてしまうのでしょうか」という質問です。つまり、データと論文を一緒に登録したけれども、後で論文がリジェクトされてしまった場合に、データも一緒に消されてしまうのですかという質問です。

●加藤 リジェクトされたら、消されることは消されると思いますが、DOI を付与されているので、メタデータ自体の公開は継続されることになると思います。

●八塚 メタデータは継続するけれども、データ自体はもうないということですか。

●加藤 はい。非公開と同じような扱いです。データ自体はないけれども、こういうデータが過去に存在したという記載と共にメタデータは残ることになります。

●八塚 次の質問です。「同じ論文の中で、公開可能なデータと非公開のデータが混在する場合がありますと思いますが、非公開データも含めて管理できますか」。

●加藤 はい。できます。それぞれに別個の DOI を付けられるかどうかという質問でしょうか。

●八塚 そこははっきりしませんが、非公開であっても DOI は付けられるということですか。

●加藤 はい。そのとおりです。

●八塚 分かりました。

次の質問は、「例えば学会や研究会の中で J-STAGE Data を使って制限共有することは可能でしょうか」というものです。

●加藤 制限共有はできません。

●八塚 一般に公開することが前提ということですね。

●加藤 はい。公開となった場合は一般公開です。公開前の登載の中で、学協会の事務局での確認ということは想定しているので、編集委員会等が見るという使い方は想定していますが、もう少し大きな集団、例えば学会内だけの限定公開はできません。

第3回 SPARC Japan セミナー2020

「初めての研究データ」

全国規模な研究データ管理サービスの 提供者視点でのベストプラクティス

込山 悠介

(国立情報学研究所)

講演要旨



研究データを研究者が独自に管理すると、貴重な研究データが研究室で紛失・消失、また気密性の高い研究データが外部に流失してしまう恐れがあります。また、研究者の所属機関が組織的な研究データ管理を行っていない場合は、研究不正の温床になる可能性があります。また、最近では競争的資金による研究プロジェクトでは、データ管理計画に基づく研究データ管理が要求されていることも増えています。国立情報学研究所では2021年から国内の学術機関に向けて、研究データ管理サービス GakuNin RDM の本運用を開始しました。本発表では、全国規模の研究データ管理サービスの提供者の目線から、実際の研究データ管理の事例を踏まえて、現実的なベストプラクティスについて述べます。

込山 悠介

国立情報学研究所コンテンツ科学研究系助教。博士（農学）。2014年3月、東京大学大学院農学生命科学研究科博士課程を修了。日本学術振興会特別研究員（DC2）として生物情報工学の研究に従事。2014年4月、東京大学医科学研究所ヒトゲノム解析センター特任研究員としてスーパーコンピュータ運営や生命科学データベース開発のプロジェクトに従事。2016年4月より現職。2017年4月より国立情報学研究所オープンサイエンス基盤研究センターの発足に伴い、研究データ管理サービス GakuNin RDM の研究開発、運営および国際連携に従事。



研究データ管理（RDM）サービスとは？

国立情報学研究所（NII）の研究データ管理サービスの背景についてお話しする前に、まず研究データ管理についてお話しします。研究データ管理サービスは、英語では Research Data Management の頭文字を取って RDM サービスといいます。RDM は、ある研究プロジェクトにおいて使用された、または生成された情報をどのように組織化・構造化していくかということを指す言葉です。

そして、「RDM サービス」にも広義と狭義とがあります。広義の RDM サービスに当たるのは、本日のメインテーマにもなっているような、研究データのライフ

サイクル（研究データの計画・管理・共有・解析・公開・検索）の全てのプロセスに関わる業務をサポートしていくための IT インフラ、あるいはその研究支援体制（意思決定、予算、人的リソースの配置、教育プログラム）の提供です。本講演内においては狭義の RDM サービスを、研究者が発表前の研究データを管理・共有することに特化した IT インフラとして紹介させていただきます。

国立情報学研究所（NII）の学術情報基盤の全体像

今回は、国立情報学研究所をご存じない方もいらっ

しゃいますので、弊所の学術基盤サービスを簡単に紹介します（図 1）。参加者の皆様は図書館に所属している方が多いと思いますので、日本語論文検索の CiNii や機関リポジトリのサービスである JAIRO Cloud 等にご存じだと思いますが、その他にも NII の事業サービスとしては、全国の都道府県に張り巡らされている学術情報用ネットワークの SINET や、学術認証フェデレーションといて、所属機関の全学共通のログインシステムで、さまざまな Web サービスヘユーザがログインして使っていただける認証フェデレーションを提供しています。他にも無線 LAN のサービスである eduroam 等も提供しています。また、RDM サービスとも関連しますが、クラウドサービスに関する導入支援等を行うサービスも行っています。

RDM サービス提供の背景

NII が RDM サービスを作ってきた背景として、2015 年ごろから政府のオープンサイエンス政策の流れがあり、その中で研究データ基盤を構築していくことになりました（図 2）。

直近では、「統合イノベーション戦略 2019」の中に、「NII を中心とした研究データ基盤・リポジトリの整備、研究データの管理・利活用方針」といったことが盛り込まれています。また、「統合イノベーション戦略 2020」の中には、NII の研究データ基盤システムを用いて先進的な研究データ、研究マネジメントを推進するということが書かれています。それがさらに進み、

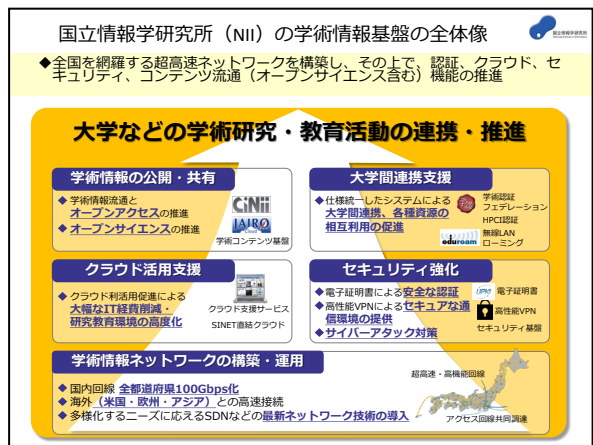
内閣府ムーンショット型研究開発制度の運用・評価指針の中にも、NII の研究データ基盤 NII Research Data Cloud (NII RDC) 等を使って先進的な研究データをマネジメントしていく事を推奨するということが盛り込まれています。NII の GakuNin RDM は、ムーンショット型研究開発制度で先行利用の形で、お使いいただいていることもあり、一部、研究代表者の先生方への説明会等も始まっています。

GakuNin RDM は、少し広い視点で見ると検索・公開・管理の 3 つの基盤からなる NII RDC のうちの一部となっています（図 3）。データの公開基盤である JAIRO Cloud と、データを検索するための基盤である CiNii Research と連動する形で、公開前の研究過程のデータを取り扱う管理基盤 GakuNin RDM を新たに提供していく位置づけになります。

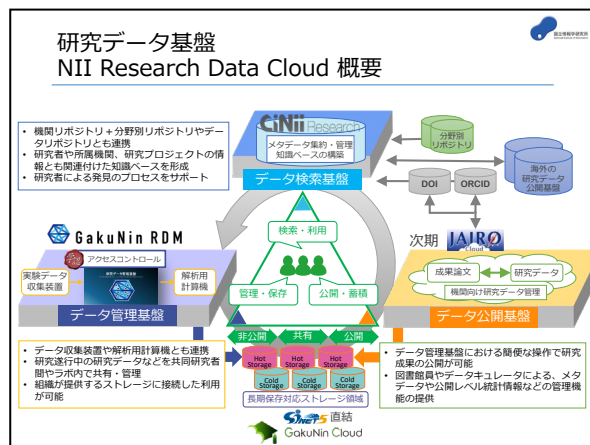
オープンサイエンス時代の 研究データ基盤構築に向けた国内の政策的経緯

- ・ 2015年3月：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
- ・ 2016年1月：政府「第5期科学技術基本計画」
- ・ 2016年2月：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5月：G7茨木・つくば科学技術大臣会合つくばコミュニケ（共同声明）
- ・ 2016年5月：政府「科学技術イノベーション総合戦略2016」
- ・ 2016年7月：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」（提言）
- ・ 2017年6月：政府「科学技術イノベーション総合戦略2017」
- ・ 2018年6月：政府「統合イノベーション戦略」
- ・ 2019年6月：政府「統合イノベーション戦略2019」
- ・ 2020年6月：日本学術会議「オープンサイエンスの深化と推進に向けて」（提言）
- ・ 2020年7月：政府「統合イノベーション戦略2020」

（図 2）



（図 1）



（図 3）

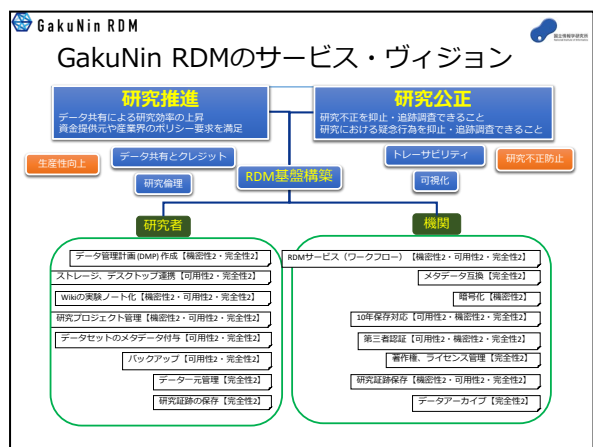
研究データ管理サービス GakuNin RDM

既に GakuNin RDM を実証実験でご利用いただいている機関の方もいらっしゃいますが、改めて説明させていただきます（図 4）。研究データライフサイクルにおける研究計画の段階から、データを収集し、データ分析等を行い、論文を書いてパブリッシュする研究者の日常を支える基盤で、Web ブラウザ上で研究データを保存・管理することができるようになっています。2021 年 4 月からは 24 時間 365 日の運用体制で全国の学術機関に提供していく予定になっています。2021 年 2 月現在、夜間帯はありませんが 365 日の有人の運用体制を取っています。

GakuNin RDM については、研究を推進していくための機能と、研究公正（研究不正の防止）という二つの大きな柱で研究データ基盤を構築してきました（図 5）。ターゲットはもちろん研究者が主役ですが、それ



(図 4)

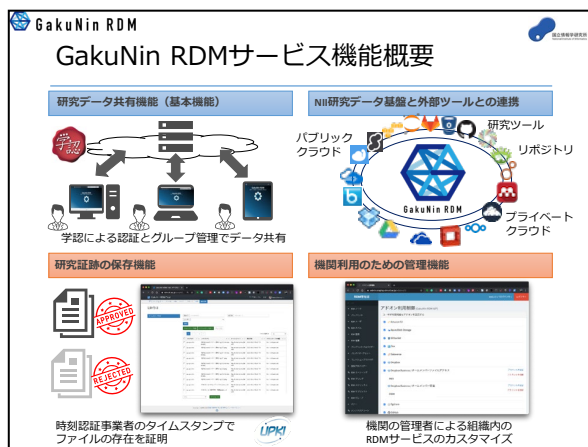


(図 5)

と同じくらい研究機関にとっても重要な機能を盛り込んできました。開発黎明期は、米国の Center for Open Science が開発した Open Science Framework (OSF) をベースに国内の需要を盛り込んで機能拡張を行っていました。初期は本当に国内では私のコンピューターの中だけで動いていたようなときもありましたが、2017 年 4 月にはオープンサイエンス基盤研究センターが発足し、本格的な開発がスタートしました。2019 年 4 月から約 1 年半の実証実験を全国で実施し、この 2021 年 2 月 15 日から本運用サービスの提供を開始しました。

GakuNin RDM の機能

GakuNin RDM では、所属機関が学術認証フェデレーション（学認）に参加していれば、いつも利用されているアカウントでログインすると、利用機関同士であれば、すぐに機関を超えてデータの共有（制限共有、限定共有）を開始することができます（図 6）。もちろん研究者 1 人のワークスペースとしてもご利用いただけます。そういった基本的なデータ保存という機能があります。しかし、それだけでは研究データ管理のサービスとは言えないので、研究者が普段使うソースコードを保存するリポジトリや、引用文献を入れるような追加機能、さらに、研究プロジェクト等でより大容量のクラウドストレージを連携できる機能を持っています。他にも基本機能として、システム中に保存したファイルに対して研究証跡の保存機能が働き、NII



(図 6)

以外の時刻認証事業者（第三者機関）のタイムスタンプによって特定時刻のファイル存在を証明するという機能が動作しています。また、機関のシステム管理者に向けて、機関内での RDM サービス提供を実施しやすくするための機能を充実させるような開発もおこなってきました。

GakuNin RDM のログイン画面に行くとプルダウンメニューに参加機関のリストが並んでいます。自機関の名称を選択すると、自機関で普段使っている認証システムを経由し、ダッシュボードにログインできるという仕組みになっています。また、研究室内や共同研究者間での階層的な研究プロジェクトの研究データ管理も行えるようになっていきます。階層自体は無制限に作ることができるので、うまく活用すると、より大規模な研究プロジェクトでも使っていただけたと考えています。

さらに、図 7 のように、さまざまな研究ツールや外部のクラウドストレージにも接続することができます。右側の図は極端な例ですが、一つの研究プロジェクトで十数種のクラウドを連携させて使うといったことも可能です。

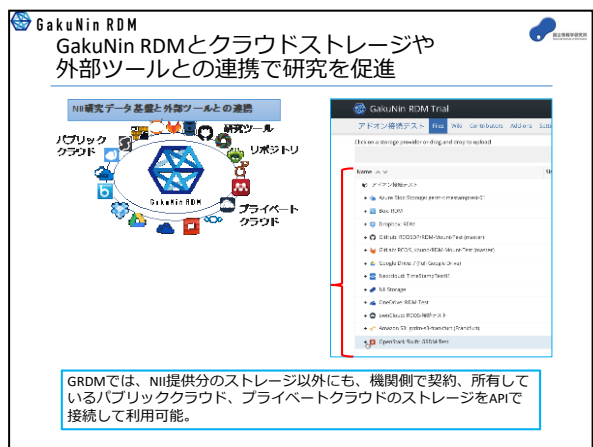
それから、研究データの証跡管理を行うことができます。一度システム中にファイルをアップロードして、もしシステムの裏側でファイルが削除されたり、書き換えられたりすることがあった場合は、それを検出する機構が入っています。それにより、プロジェクトを管理する研究主宰者は、プロジェクトに所属する研究

者の研究データへの不正操作を検知できることで、研究不正を抑止することができます。

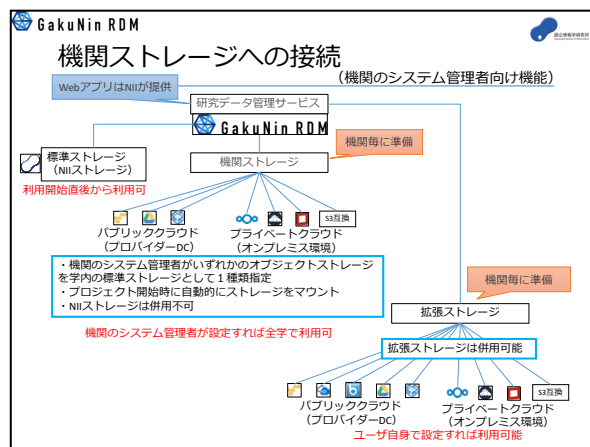
機関のシステム管理者へ向けて開発した機能として、GakuNin RDM には機関がお持ちのストレージを接続して利用していただける機能もあります（図 8）。機関がエンタープライズ契約している商業用のクラウドサービスや、機関の中にストレージサーバーをお持ちであれば、GakuNin RDM へ接続して使っていただくことができます。もちろん、GakuNin RDM をご契約いただくと、標準ストレージ（NII ストレージ）として、利用開始直後から使用可能な領域も一定容量あります。加えて、ストレージを準備していただく必要がありますが、機関ごとまたは、プロジェクトあるいは部局単位で切り分けてストレージを設定したいという要望にもお応えできる形になっています。

ここで簡単に、RDM サービスのおすすめの構成を大学の規模別に紹介します（図 9）。実証実験の段階では大規模大学と中規模大学の事例が多かったのですが、大規模大学は、想定される取り扱い研究データが大きい分野・部局があり、あるいは部局ごとの自治があり、全学的なシステム統一が困難、あるいは外部との大規模な共同研究のプロジェクトが多いという場合には、標準的に準備されているストレージと拡張ストレージの併用を検討されているところが多かったです。

中規模な大学では、全部局で統一した RDM サービスが導入可能、あるいは地域の研究所や企業との共同研究プロジェクトがあるということで、学内にサーバ



(図 7)

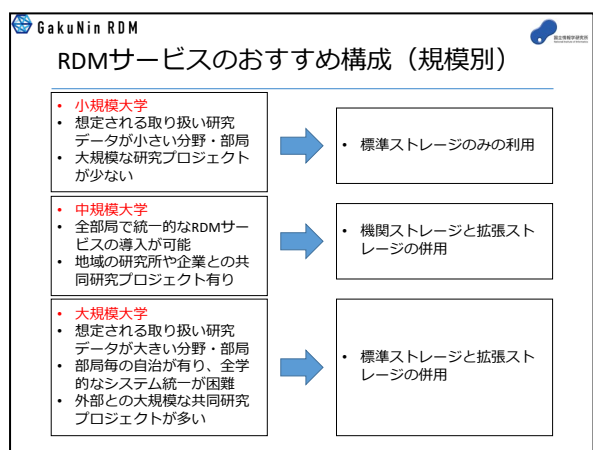


(図 8)

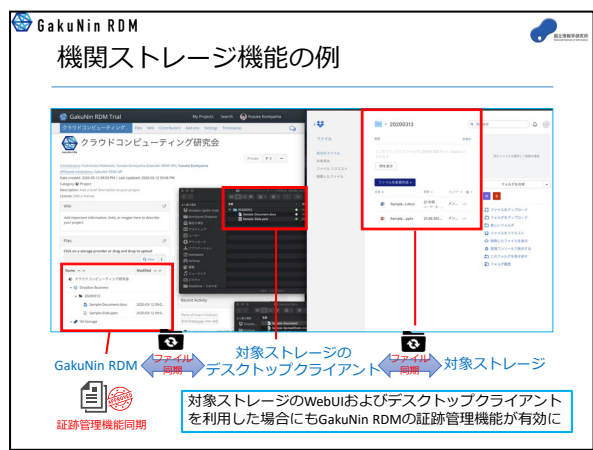
ーがあるストレージを接続して、標準ストレージと置換えて使っている事例がみられました。それだけでは不十分ということで、さらに外部機関とも連携するためのストレージを拡張ストレージへ割り当てているという事例もあります。

小規模な大学では、情報科の教員や情報システム部門の職員が学内 IT システムの面倒を見ている場合があります。そのような場合に、標準ストレージだけを使うのであれば、ご契約いただいた直後から全学的な RDM サービスをご利用いただけます。

機関ストレージの具体的な例をお示しします。図 10 は、GakuNin RDM の裏側に、機関で利用する契約をしている Dropbox Business のストレージを割り当てた図ですが、例えば既に学内で使われているデータ共有のストレージがある場合は、そのデスクトップクライアントと RDM をうまく連動させて、なおかつ研究



(図 9)



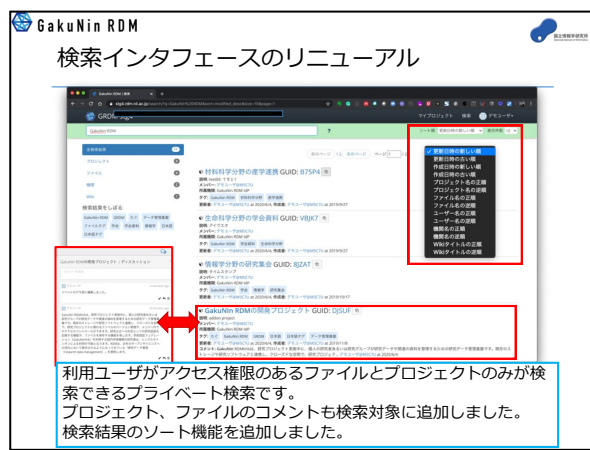
(図 10)

証跡機能も働くということを一部のクラウドストレージでは実現しています。これを使うと、エンドユーザーにとっては、何も考えずに普段の Windows や Mac でファイル管理を行えば、それが自動的にクラウドに同期されると同時に研究データ管理が行われます。

また、システム中に保存されたデータで、かつ自分にアクセス権があるデータのみを検索も実現しています（図 11）。システム中に入ったファイルに関して、ファイル名やアノテーション、ディスカッションのコメント、ID を検索できる機能も拡張しています。

また、留学生や海外の研究者も含めて利用したいということで、ユーザインタフェースやメッセージは日本語と英語に対応しています。Web ブラウザの言語設定を切り替えれば、それだけで言語が切り替わるという仕様になっています。

GakuNin RDM は、2021 年 2 月 15 日から本運用でサービスを提供しています（図 12）。プレスリリースと同時に利用申請の Web サイトもオープンしました。機関のシステム導入の担当者様にお申し込みいただくことになっています。現在既にお申し込みの受付が始まっています。これまで先行利用していただいていた機関につきましては、データ移行することなくそのままご利用いただけます。冒頭でも申し上げましたが、4 月から 24 時間 365 日のフルスペックでの有人オペレーションを始めていくということで、2020 年度内はその準備として、夜間を除く 365 日の有人オペレーション体制を取っています。



(図 11)

当面は、アプリケーションの利用料は無料としています。また、大学からご要望が多い NII クラウド導入支援サービスのクラウドチェックリストにも、サービス開始と同時に掲載させていただきました。ポリシー面の詳細につきましてはクラウドチェックリストを参照していただけます。2021 年度はサービス運営として ISO27001 (ISMS: 情報セキュリティマネジメントシステム) の認証取得を進めていく予定です。

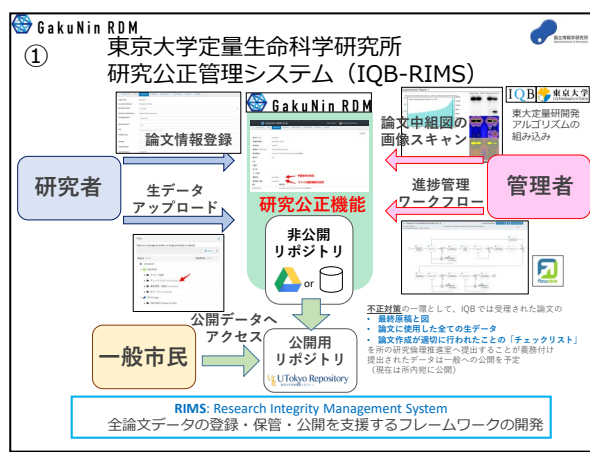
GakuNin RDM の利用事例

ここから GakuNin RDM のユースケースの説明に移ります。昨年度、この SPARC Japan セミナーに登壇させていただいた際、GakuNin RDM の実証実験では利用機関数は 18 機関でしたが、現在では 22 機関に増えています。最近ではムーンショット型研究開発制度の採択大学のプロジェクトマネージャー（領域代表）からもお問い合わせが来ています。また、これまでは国立大学が中心でしたが、私立大学や NPO も参加していただいております、学術機関の規模や種別も多様化しています。

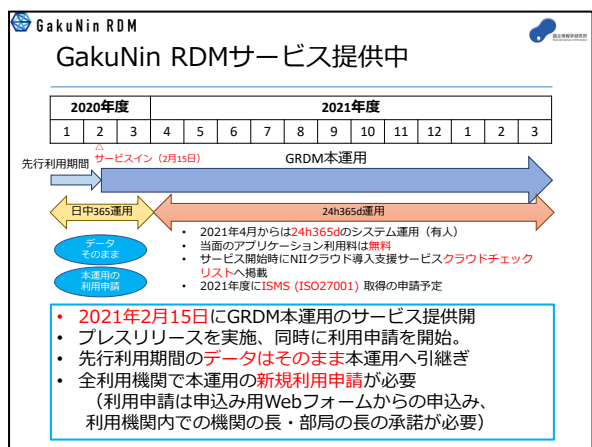
利用事例を紹介します。一つ目は、東京大学定量生命科学研究所 (IQB: Institute for Quantitative Biosciences) と共同で開発している研究公正管理システム (IQB-RIMS) です (図 13)。これは、生命科学分野の論文に含まれている画像を、部局に所属する全ての研究者の論文を対象として出版前に倫理部門でチェックしたいというユーザーシナリオがありました。これまでは、

メールベースのやりとりをされていて、スプレッドシートでデータを管理しリマインドもメールで行い、論文中の画像データの解析もかなりの部分が手動で行われていました。その半自動化という目的を中心に GakuNin RDM を入れていただき、機能拡張することで研究公正機能を実現されています。これには一部、業務プロセスを管理するワークフローエンジンを導入・連携しています。2019 年度の紹介時は開発段階でしたが、2020 年度は実証実験に進んでおり、今後、大学の公開用リポジトリに、全ての論文にひも付くデータを公開する機構を入れようとされています。

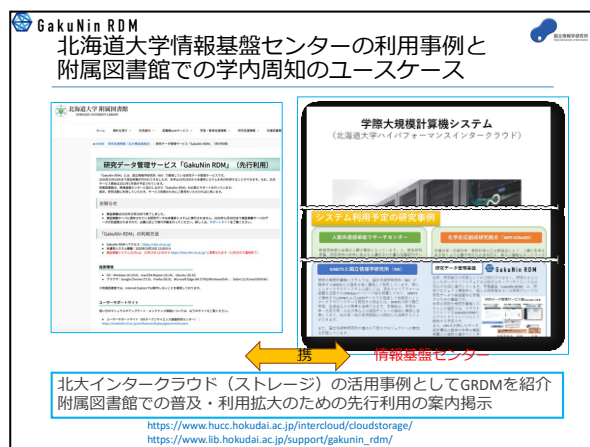
二つ目は北海道大学の事例です (図 14)。情報基盤センターのスーパーコンピュータのユーザーが利用できるストレージを GakuNin RDM に割り当てることができるといことで、情報基盤センターのパフレット等に掲載していただいています。一方で、附属図書館



(図 13)



(図 12)



(図 14)

の方では、利用方法や必要な情報等を掲載していただき、学内研究者への普及をしていただいております。こういった図書館と情報基盤センターの連携が RDM サービスの利用者を実際に広げていく上で非常に大切なことだと思っています。そして、URA・研究推進部門から出されている北大コアファシリティ構想の中で、学内に点在している実験装置や遠隔地の実験装置をリモートで操作するプロジェクトのデータの受け皿として GakuNin RDM を検討していただいております。

三つ目は名古屋大学の事例です（図 15）。学内で提供されているクラウドストレージを GakuNin RDM に連携できるということで、情報戦略部門でかなり丁寧なマニュアルを作って掲示していただいております。これも情報基盤センターと情報戦略部門との連携の事例と考えています。

また、その他の事例として、2019 年度はテキストのみの紹介でしたが、2020 年度は大型研究プロジェクトのデモムービーも作成しました（図 16）。京都大学の医師と北海道大学の数理学者が連携してデータを取得し、分析を行い、最後は医師がアノテーションするという用途で利用していただいております。

最後の事例は、大学経営統合での IT インフラ統一のユースケースです（図 17）。こちらは北海道国立大学機構が三つの大学を統合するときに、IT インフラを統合したいということで GakuNin RDM の導入をご検討いただき、学内に設置したクラウドストレージ、学内にあるサーバーのストレージと GakuNin RDM を

連携させるということで既に動き出しています。

まとめ

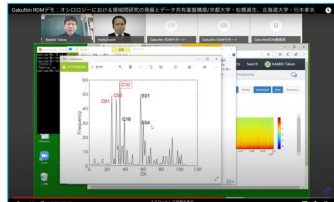
以上、全国規模で提供している研究データマネジメントサービス GakuNin RDM の紹介と、その機能概要について話題提供させていただきました。2021 年 2 月 15 日から GakuNin RDM の本運用サービスが動いており、2021 年 2 月現在の利用機関は 22 機関となっています。今後も受付を行い、さらに関係各所と連携しながら、各大学で使っていただけるように進めていきたいと考えています。

●八塚 既に幾つかのご質問を頂いています。まず、「個人情報を含む研究データのように秘匿性の高いも

GakuNin RDM

⑤ 大型研究プロジェクトでのユースケース

医師と数学者による神経科学・数理学・医学分野の共同研究
(新学術領域オシロロジー)



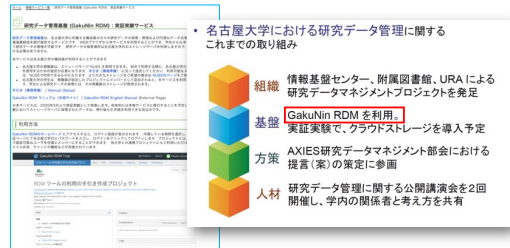
GakuNin RDM デモ：オシロロジーにおける領域間研究の発展とデータ共有基盤構築/京都大学・松橋真生先生、北海道大学・行木孝夫先生

GakuNin RDM 利用デモ動画を YouTube で公開中
<https://youtu.be/5zS8-o5B3vw>

(図 16)

GakuNin RDM

④ 名古屋大学での実証実験のユースケース



名古屋大学における研究データ管理に関するこれまでの取り組み

- 情報基盤センター、附属図書館、URA による研究データマネジメントプロジェクトを発足
- GakuNin RDM を利用、実証実験で、クラウドストレージを導入予定
- AXIES 研究データマネジメント部会における提言(案)の策定に参画
- 研究データ管理に関する公開講演会を2回開催し、学内の関係者と考え方を共有

名古屋大学情報推進本部
<http://www.icts.nagoya-u.ac.jp/ja/services/rdm/>

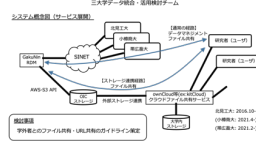
国立大学経営改革促進事業「研究データの大学間相互利用に向けたアカデミッククラウドの構築」の一環として GakuNin RDM を活用

(図 15)

GakuNin RDM

⑥ 大学経営統合での IT インフラ統一のユースケース

工学・商学・農学の異分野の大学経営統合によるオープンイノベーション
(北海道国立大学機構：北見工业大学、小樽商科大学、帯広畜産大学)




GakuNin RDM デモ：GakuNin RDM を軸としたオープンイノベーション/北見工业大学・井澤洋志先生、小樽商科大学・三浦克寛先生

<https://youtu.be/316KX83b-q>

(図 17)

のに関しては、GakuNin RDM で管理できますか」という質問です。

●**込山** 現状、システム側では暗号化等を行っていないので、秘匿性の高いデータを入れる場合は、ファイルをユーザー自身が暗号化して入れる必要があります。

●**八塚** コミュニティのポリシーやユーザー側の対策が必要ということですね。

●**込山** ユーザー側の機関ポリシーに従って使ってよいかということ判断していただく必要があると思います。ただ、ファイル単位の暗号化がユーザーからご要望の多い機能だということは認識しています。

●**八塚** 次に、「証跡管理機能として、リポジトリに関与する全ユーザーの編集履歴などをログファイルで取得することは可能でしょうか」という質問です。

●**込山** GakuNin RDM の証跡管理機能では、プロジェクト（ご質問ではリポジトリ）の管理者はプロジェクトに関する全ユーザーの編集履歴などを、ログファイルで取得することが可能です。また、学内の調査委員会などから指示を受けた機関のシステム管理者が、プロジェクトごとに全ユーザーの編集履歴をログファイルとして取得することも可能です。

（注：当日の回答が質問内容に沿ったものではございませんでしたので、正しい回答に修正いたしました）

●**八塚** 次の質問です。「利用している研究者が他の大学等に移った場合はどうなるのでしょうか」。

●**込山** 現状では、所属機関ごとにアカウントが作成されるので、他大学に移った場合は別人という形になります。初期の頃はこれを統合する仕組みを考えていましたが、機関を超えて研究者が移籍する場合に、研究者が勝手にデータを持って異動してしまうことはい

かがなものであるという意見もあり、現在は別のユーザーとして新たにアカウントを作ってログインしていただく形になっています。

第3回 SPARC Japan セミナー2020

「初めての研究データ」

研究データ管理サービスってどうやってはじめるの？ (欧州での例を参考に)

神谷 信武

(チューリッヒ大学 アジア・オリエン特研究所図書館)

講演要旨



研究データ管理サービスをはじめるにあたり、他の機関の様子を見るのは大変参考になる。というわけで、少し古い資料だが、DCC の“RDM strategy: moving from plans to action”から、特に University of Edinburgh を対象として取り上げる。また、研究データ管理サービスを発展させるにあたり、SparcEurope が提供している RISE をもとにした RDM サービス自己査定ツールも役立つと思うので紹介する。



神谷 信武

司書 (チューリッヒ大学 アジア・オリエン特研究所図書館)。今回はJDARNの一員としてお話しします。

プレゼンの目的

この発表では、研究データ管理サービス (RDM) をどうやって始めるのかということを中心に考えていきたいと思います。①RDM をどうやって始めるのか、②初めから完璧な RDM サービスなどあるのか、③現時点で完璧な RDM があったとして、将来も継続してそうあり続けるだろうかという三つの疑問を考えながら、主にイギリス・スコットランドのエディンバラ大学の RDM サービスの取り組みを見ていきます。まずエディンバラ大学の現状をお話しします。次に、エディンバラ大学は 2011 年ごろから本格的に RDM サービスの提供を開始しましたが、その初めの段階を見ていきます。それから、エディンバラ大学は 2013 年以降、RDM サービスをさらに発展させていくためのロードマップを幾つか書いているので、それを見ながら、どのように段階的に発展していったか見てい

ます。そして、発展段階において必要な自己査定についても見ていきます。

エディンバラ大学とその RDM

エディンバラ大学は古い大学で、デイヴィッド・ヒュームやコナン・ドイルなどを輩出した大学でもあります。研究大学として RDM にも積極的です (図 1)。例えば MANTRA はセルフラーニングのオンラインコ

エディンバラ大学とそのRDM I

- イギリスのスコットランドにある1583年に設立された大学
- 哲学者ヒュームや医者・作家コナン・ドイルなどを輩出
- RDMに関しても積極的である
 - [MANTRA: Research Data Management Training](#)
 - [OCLC: The Realities of Research Data Management](#)
 - [Delivering Research Data Management Services \(FutureLearn\)](#)
 - RDM Roadmaps

(図 1)

ースです。RDM に関するセルフラーニングの資料も出しています。また Online Computer Library Center (OCLC) に対しても、自機関で行ってきた研究データ管理のインフラに関する報告書を書いています。FutureLearn という e ラーニング、セルフラーニングのプラットフォームで研究データマネジメントサービスのモジュールも提供するなど、大変積極的な大学です。RDM ロードマップに関しては後ほど説明します。

具体的にエディンバラ大学がどのようなサービスを提供しているかは、ホームページで見ることができず (図 2)。RDM サービスに関して必要な事項が分かりやすく述べられていて、大学関係者に対してもしっかりと情報が行き届くフローができており、大変素晴らしいと思います。学内の研究者や、研究データを管理する者に対してもトレーニングを提供していますし、さまざまな研究段階に対応してツールの情報も提供しています。

RDM Strategy: moving from plans to action (2013)

ただ、最初からそのような状況だったわけではありません。最初がどうだったかということは、イギリスの The Digital Curation Centre (DCC) が 2013 年に発表した「RDM strategy: moving from plans to action」というドキュメントで見ることができます (図 3)。イギリスのエディンバラ大学、サウサンプトン大学、サリー大学で RDM サービスを導入したときに、三つの大学で異なる状況がある中で、それぞれがどうやって始めて、発展させていったのかということを変に短くまと

めたものです。

これによると、エディンバラ大学は、2011 年 5 月に RDM ポリシーを策定しました。策定のために、研究者にあらかじめアンケート調査を行っていました。その前の段階で、既にエディンバラ大学はデータシェアリングやリポジトリのツールを持っていましたが、それだけでは足りない状況がありました。そして 2011 年 5 月にポリシーを策定し、それを基に予算を算出しました。具体的には 100 万ポンドをハードソフトなどの予算として算出し、年間予算としては 25 万～40 万ポンドという概算を出しました。その後、データマネジメントの設計、研究データのストレージ、データシェアのリポジトリの三つの分野に関してパイロットプロジェクトを施行し、さらに、具体的にどういったサービスをどのように提供していくかということを考えながら発展させていったようです。

RDM strategy には、三つの大学が RDM サービスを発展させていく過程で共通して課題となった箇所もまとめられています (図 4)。一つ目は、コストの算出です。RDM サービスを提供するときにかかるお金を

RDM strategy: moving from plans to action I

2013年に発表された報告で、エディンバラ大学・サウサンプトン大学・サリー大学の3つの異なる大学でRDMサービスを導入しその概況を比較し、共通する問題点や学びを短くまとめたもの。

エディンバラ大学は...

- 2011年にRDMポリシーを策定
- それをもとに予算を算出
- データ・マネジメントの設計、研究データのストレージ、データ・シェアのためのリポジトリにパイロット・プロジェクトを施行

(図 3)

エディンバラ大学とそのRDM II

エディンバラ大学のResearch Data Service

ホームページ : <https://www.ed.ac.uk/information-services/research-support/research-data-service>

- 様々なトレーニングの提供
- 様々な研究の段階に対応してツールの情報を提供

(図 2)

RDM strategy: moving from plans to action II

共通して課題となった箇所

- コストの算出
- 予算割当の確保
- サービスの提供の仕方

(図 4)

どう計算するのかということです。二つ目は、仮に予算を算出できたとして、それをどうやって割り当てて確保できるのかということです。三つ目はサービスの提供の仕方です。完璧な RDM サービス全般を一気に出せなかったら、その場合は順番に出していくのかということです。

では、得られた学びは何だったのでしょうか（図5）。まず、やれることからやるということです。例えば、ポリシーがあるに越したことはないけれども、ポリシーをつくるのに時間がかかり過ぎて何もできないのであれば、別のことから始めた方がいいという意見がありました。次に、目的となるサービスが自機関においてどれくらいのスケールを要するか把握することです。それから、シニアの学者からのコミットメントを得られるようにすることです。学内でも学者からのコミットメントは必要なので、できるだけ有力な学内の方からも助力を得られる状況をつくっておくことが大事です。そして、サービスに関する情報はオープンに、なるべくトランスペアレントな形で提供していくことです。さらに、RDM サービスの発展には、なるべく柔軟だけれども実現可能なタイムスケールの設定をすべきであることです。あまりにも理想に走ってはいけないし、がちがちな期間設定も難しいというのが学びということでした。

エディンバラ大学の RDM Roadmaps

エディンバラ大学は RDM ロードマップのバージョン 2（2012～2016 年）とバージョン 3.2（2017～2020 年）を公開しています。3、4 年のペースでロードマ

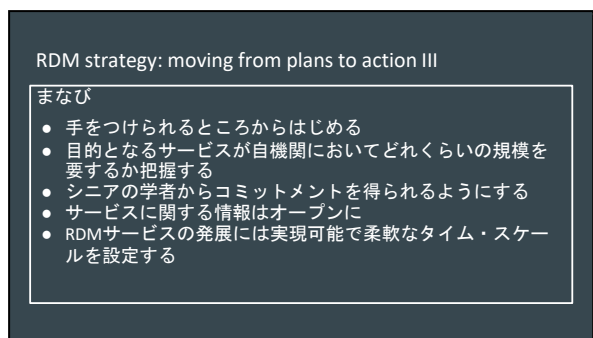
ップが書かれています。

ロードマップでは、RDM プランニング、アクティブ・データ・インフラストラクチャー、データスチュワードシップ、データ・マネジメント・サポートの四つのサービスカテゴリーにまず大きく区切り、その中から細目を取って、さらにその細目のプライオリティを付けて、誰が、どれぐらいで達成するのかという大まかな時間の設定をしています。

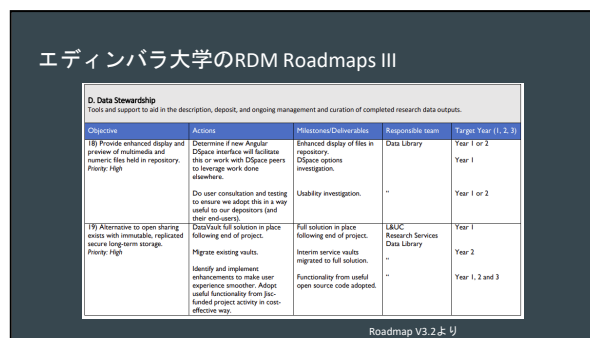
分りにくいと思うので具体的にお見せします（図6）。データスチュワードシップというカテゴリーの中で、例えばリポジトリの中のプレビューやディスプレイを見やすくするという細目を立てます。そのためには何を必要があるのか、それを誰が実現するのか、そして目標時間はどれくらいかという細かい項目を立てて、なるべく実現できるような努力をしています。これがロードマップです。

エディンバラ大学のロードマップのホームページには、バージョン 3.2 がほぼ達成できたと書いてあり、素晴らしいのですが、最初から全て素晴らしいものを提供する必要はありません。今こういうものが提供できていて、次のステップとして何をいつまでに、どのように良くしていったらいいのかということを考え、それをある程度、表に出しながら話を進めている感じ

です。このようなエディンバラ大学の素敵な RDM サービスですが、ロードマップを見ると、彼らもいろいろと模索している印象があります。では、これから RDM サービスを提供したい人はどうしたらいいのかということをおなりに考えてみました。取りあえず獲得でき



(図 5)



(図 6)

るリソース、予算を考えつつ、できることから始める。そして、余裕がある場合は現状提供している RDM サービスを自己査定し、実現可能な枠でその改善を試みる。この2点を念頭に置くといいのではないかと思います。

RDMの自己査定とRISE

自己査定はどのようにするかというと 2018 年に SPARC Europe が公開した自己査定ツールがあります (図 7)。項目に沿って自分の大学の RDM サービスの現状を入力していくと、最終的にレーダーチャートが作成できるというものです。私も 1 回使ってみました。登録しようとする SPARC Europe にデータが蓄積されるので、嫌な人はやめておいた方がいいと思います。ただ、レーダーチャートで見ると、良いところ、良くないところが一目で分かるので便利です。

ちなみに、このレーダーチャートは Research Infrastructure Self-Evaluation (RISE) というフレームワークを基に作られています。このフレームワークは基本的に、RDM ポリシー&戦略、ビジネスプランと持

続性、アドバイスサービス、トレーニング、DMP、アクティブ・データ・マネジメント、査定とリスクマネジメント、保存、アクセスと出版、ディスカバリーという 10 個の大きな項目から成っていて、図 8 が実際のレーダーチャートです。

RISE に関して細かい説明はしませんが、同じくデータキュレーションセンターが発表した Using RISE, the Research Infrastructure Self-Evaluation Framework というドキュメントの PDF 版の 9 ページ以降に細かい説明があります (図 9)。先ほど挙げた 10 項目の下に、さらに三つの細目があります。例えば、ポリシー・ディベロップメントがその一つです。その細目をさらにレベル 1~3 に分けて、最終的にリストでポイントを加算して自分の査定ができるものになっています。

ヨーロッパの場合は自己査定を公開していない大学があり、公開の例としては二つしか見つけれませんでした (図 10)。一つは、オランダのデルフト工科大学が提供している自己査定です。これはリスト形式なのでレーダーチャートではありませんが、自分の大学の中でどういうレベルのものを提供できているか把握

(図 7)

(図 9)



(図 8)

(図 10)

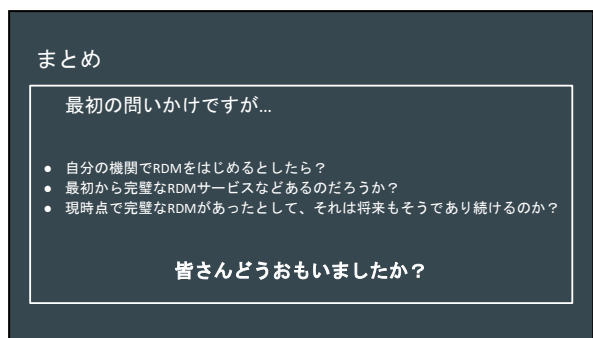
できるようになっています。もう一つはライブニッツ大学のものです。これもリストで、現状は1レベルで最終的には2レベルにまで持っていきたいということをも自分たちで判断し、次の段階に持っていくために何をしたらいいかということをも自分で決めることができるツールになっています。

これを公開することによって、他の大学の参考にもなります。大学ごとに基準や出発点が違うので、一概に一般化はできませんが、自分の大体の立ち位置が分かるようなツールです。

まとめ

最初の問いかけをもう一度見てみましょう（図 11）。どうやって RDM を始めるか。最初から完璧なサービスがあるのか。完璧な RDM があったとしたら、それは将来もそうであり続けるのか。

私は、初めから完璧な RDM サービスはないと思っています。また、現時点で完璧な RDM があったとしても、将来はそうではないと思います。ウォーターフォール型の RDM サービスが1回で提供できたとして、それは恐らく将来継続して優秀なサービスとしてファンクションしないのではないかと思います。ただ、完璧な RDM サービスがいきなりできないから何もやるべきではないということではなく、できることから始めたらいいのではないのでしょうか。みんなで一緒に悩みながら進んでいきましょう。



(図 11)

●八塚 質問を頂いています。まず、「エディンバラ大学では RDM ポリシーがなくともやれることからやるというサービスを始めたということでした。RDM を始めるときに、RDM があってポリシーがないというのはいいけれども、そもそも RDM がない場合、どのようなことから始めたらいいと思いますか」という質問です。

●神谷 そもそも、RDM がなくとも、実際にリポジトリなどはある気がするのですが。

●八塚 私が推測するしかないのですが、要するに、マネジメントがまだできていない、これから始めなければならないというところもあるのではないかと思います。今だと機関リポジトリも結構ありますが、そういったものもまだない状態で、まず何から始めたらいいかということだと思います。

●神谷 まずは持っているものから確認することですが、そういえば Japan Data Repository Network (JDARN) の小委員会で、RDM サービスを提供するためにどういったタスクが必要なのかというリストを作られていましたね。

●八塚 はい。ガイドラインは作りました。

●神谷 そういうものを見ながら、どういうタスクが発生して、われわれが所属している機関にはこういうものがあって、これを使えばある程度のことは実現できるのではないかということは、意外とすぐできるのではないかと思います。その辺から始めてはいいかがでしょうか。

●八塚 なるほど。ありがとうございます。

次の質問は、大学図書館の方から頂いています。

「ヨーロッパの大学や研究機関における RDM は、どこが主体となって実施しているのでしょうか。図書館員が何らかの役割を担っているパターンがあるのでしょうか。あるいは、ご自身の所属機関や近隣等で RDM に関する取り組みがあるようでしたら教えてください」。

●神谷 私の所属する大学は、図書館が研究者に対して DMP などをサポートすると言っているので、一応、図書館が中心になって RDM サービスを提供しようと頑張っているところです。ただ、私の大学もそんなに完璧ではありません。担当の人に話を聞くと、「私もよく分からない」という返事が返って来たりしますが、目標としては、研究者の一番近くにいた図書館員が、いわゆるサブジェクトライブラリアン的なファンクションを担いつつ、データ管理なども並行することを目指しています。

スイスの場合は、国で SWISSUbase というリポジトリを立ち上げようとしていて、公開用のデータをスイス国内から吸い上げられるようなインフラをつくらうとしています。それが今はプロジェクト段階にあり、主に図書館員の方たちに説明しているところです。

●八塚 ありがとうございます。次の質問は、「EU においては、データ共有にまつわる法的な背景が変化していますが、研究者以外には、例えばこういったバックグラウンドを持った方が Evaluate your RDM Offering などのメンテナンスに関わっているのでしょうか。ご存じであれば教えていただきたいです」というものです。

●神谷 EU 全体で言うのは難しい気がします。図書館からの動きはあり、欧州研究図書館協会 (LIBER) や、ヨーロッパの図書館連合も積極的に関わっていますが、具体的にどこの誰がということは分かりません。

●八塚 私から一つ質問です。神谷さんは図書館にお

勤めですが、日本だと、図書館員だけではなく、図書館の研究者も含めて、どうしても誰がやるかということで押し付け合いめいたことがあります。どのようにリーダーシップを取ったり、協調したりして進めていくべきでしょうか。ヨーロッパから日本を見て、日本はこういうところを良くしたらいいのではないかと思いますか。

●神谷 日本は GakuNin RDM もあるし、インフラがたくさんあるので素晴らしいと常々思っています。私は、例えば JDARN などと一緒に活動させていただいて、皆さん同じ悩みを抱えているのは分かっているので、それならみんなでフランクに悩みを話し合えばいいのではないかと思います。ただ、日本に関しては、専門リポジトリと大学が教員全般に提供するリポジトリが違っているのも、その辺をどのように共有したいのか分からないのですが、例えばデータキュレーターと一緒に話ができる土台があるといいのかなと思います。また、デジタルアーカイブ学会も、デジタルアーカイブに関するクオリティの査定などのツールを作っているようなので、その辺とも一緒に話ができたら楽しいのではないかと思います。

第3回 SPARC Japan セミナー2020

「初めての研究データ」

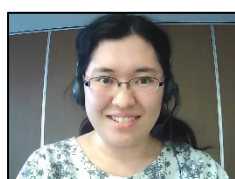
研究データ公開実践のための課題を探る： 北海道大学での事例を通じて

三上 絢子

(北海道大学附属図書館)

講演要旨

国内の大学において、研究データ公開を支援するための体制は未だ十分に整備されていない。研究者の意向を反映した形での支援体制を構築する上では研究データ公開の実践事例の蓄積と分析が必要と考えられる反面、研究データ公開希望に対する国内大学での実践事例はこれまでほとんど共有されてこなかった。本講演では、2017年から2020年まで北海道大学の機関リポジトリ（HUSCAP）上でデータを公開した14件の事例について、研究者がデータ公開を希望した背景とデータ公開時に機関リポジトリ担当部署側が行った対応を紹介する。また、これまでの研究データ公開に関する議論を踏まえてこれらの事例を分析することで、現在の機関リポジトリの運用体制では対応が不十分な点やこれから改善すべき点を発見する。



三上 絢子

2016年より北海道大学附属図書館勤務、現在は研究支援課研究支援企画担当係員。2020年度オープンアクセスリポジトリ推進協会（JPCOAR）研究データ作業部会員。2019年度国立情報学研究所実務研修において研究データ管理・公開業務に携わる。2019年より現在まで北海道大学の機関リポジトリ業務を担当。

自己紹介

私は北海道大学附属図書館の職員です。研究支援企画担当ですが、3年ほど前まではシステム担当という名前でした。その名のとおり、主な業務は図書館の各種オンラインサービスの管理です。このサービスの中には、図書館業務に必要なオンラインカタログや、電子ジャーナルに学外からアクセスするためのシステム、その他、機関リポジトリも含まれており、実質的に機関リポジトリ担当ということになります。今回はこの機関リポジトリ上での実践を中心にお話しします。

北海道大学学術成果コレクション（HUSCAP）の概要

北海道大学学術成果コレクション（HUSCAP）は、

学術成果公開を目的とした機関リポジトリです（図1）。運用は2005年からで、当時から機関リポジトリ構築用のオープンソースソフトウェア DSpace をベースに、機能を継ぎ足して運営しています。主な公開資料は北海道大学所属の研究者の論文や学会発表資料、北海道大学発行の紀要ですが、ここ数年はごく少数な

北海道大学学術成果コレクション（HUSCAP）の概要

66,910件の学術資料を公開する機関リポジトリ(2021/02/04現在)
オープンソースソフトウェア「DSpace」をベースに構築している(運用は2005年から)
主な公開資料:
・北大所属の研究者から公開依頼のあった資料(論文、教育資料、学会発表資料など)
・北大発行の紀要
・研究データ(少数)

北海道大学

(図1)

がらも研究データを公開する事例が出てきています。

研究データ公開事例の概要

2017年から2020年までにHUSCAP上で行われた研究データの公開事例を紹介します。今回の事例紹介で扱う研究データの範囲は、HUSCAP上で公開された資料のうち、メタデータの資料種別がdatasetのものに限定しています。実際のデータの形式としては、数値データを集めたExcelファイルや画像ファイルが大半でした。対象の範囲は2017～2020年で、この間に14件の依頼がありました。これはファイル数ではなく依頼件数ベースでカウントしています。そのうち北海道大学発行の紀要が1件で、残りの13件は雑誌投稿論文に関連したデータです。ちなみに、2016年以前には公開したデータセットはありませんでした。

データ公開は北海道大学所属の研究者の依頼により行っています。図2の左側に依頼者の所属部局、右側に、部局ごとにデータに関連する論文が投稿された雑誌名をまとめました。今回の場合は、やや分野に偏りが見られ、大まかに言えば地球科学分野や医学分野が多かったことが分かります。

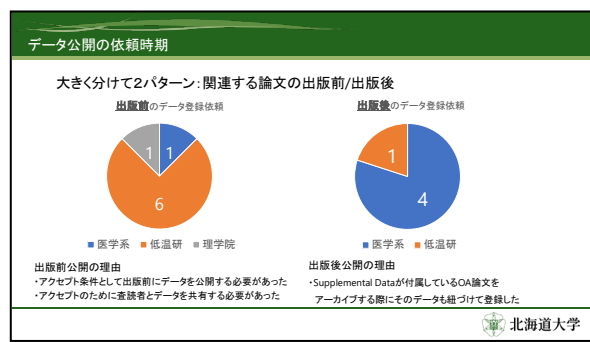
また、今回の実例では、研究データ公開が依頼される時期が、データに関連する論文が出版される前と出版された後の二つに大きく分けられました(図3)。件数としては、出版前に依頼された件数が若干多く、特に低温科学研究所所属の研究者の依頼は出版前、医学系所属の研究者の依頼は出版後が多い傾向にありました。出版前に公開する理由としては、論文のアクセプト条件として事前に研究データを公開するという要

件が課せられていたからというものがほとんどです。出版後に公開したものは、既に他の場所で公開されているオープンアクセスの論文をアーカイブする際に、関連して公開されていた研究データも一緒にアーカイブしたものです。

次に、これらの研究データの公開依頼に対して実際に行った作業を紹介します。ここでは公開事例をデータ公開のタイミングごとに分けて順に解説していきます。

パターン1：投稿前の論文の根拠データを公開

パターン1-1は、校正時に研究データのURLを確定させる必要があったケースです(図4)。実際に研究データをオープンにしたのは、論文投稿の流れ図の右側、最後の論文公開時ですが、論文誌面上にURLを掲載する必要上、校正の段階で事前にURLだけ確定させておきたいというケースでした。これが今回の事例の中では最多で、5誌7件の依頼が該当しました。ポイントとしては、論文の公開前にURLを確定させておいて、論文が公開されるタイミングでデータを公

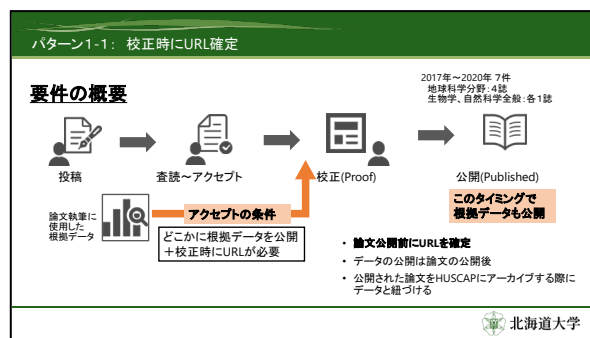


(図3)

依頼元の研究者が所属する部局		
部局	件数	投稿雑誌
低温科学研究所	7	Journal of Geophysical Research: Atmospheres Mammal study
医学研究院	3	低温科学 Earth and Planetary Science Letters
その他医系部局	2	PNAS Polar Science
理学院	1	Behavioural brain research Scientific data
水産学部紀要	1	Molecular Cell The Journal of Clinical Endocrinology & Metabolism
		理学院 Tectonics
		水産学部紀要 海洋調査漁業試験要報

北海道大学

(図2)



(図4)

開することと、論文の公開後にデータと論文とのひも付けを行うことです。機関リポジトリ側の作業としては、まず非公開状態でデータをアップロードし、この時点で URL を確定させました。その URL を校正用にエディターに送り、論文が公開された後に、研究者から公開された旨の連絡をもらってデータを一般に公開しました。公開の際に、研究データのメタデータに関連する論文の情報を入力しました。特に、データの公開は論文が公開された後に行いたいという要望が多かったため、データの公開前に URL を確定できる機能があったことは作業していて助かりました。

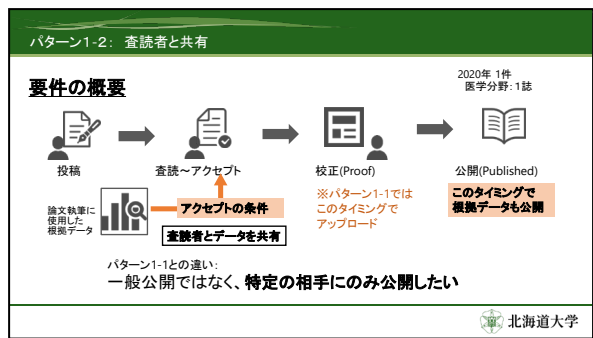
パターン 1-2 は、査読者とデータを共有したい場合です（図 5）。先ほどのパターンでは、流れ図の最後の公開までは完全にデータ非公開状態で構いませんでしたが、このパターンでは、一般公開よりも前の査読の段階で特定の相手にだけ公開したいという要件が入ってきます。これも、雑誌の規定でこれがアクセプト条件だったというものです。まずは先ほどのパターンと同じく、研究データを非公開状態でアップロードしましたが、HUSCAP とその大本になった DSpace の場合、特定の非公開アイテムだけが読み取り可能なアカウントを元々の機能として作成することができるので、査読者にこの研究データのアクセス権限付きのアカウント ID を渡しました。これは査読者と共有するところ以外は最初のパターンと同じで、論文が公開されたタイミングで依頼元の研究者から連絡してもらってデータを一般公開しました。

ここまでの 2 パターンは、論文が公開された後に研究データも一般公開しましたが、次のパターン 1-3 は、

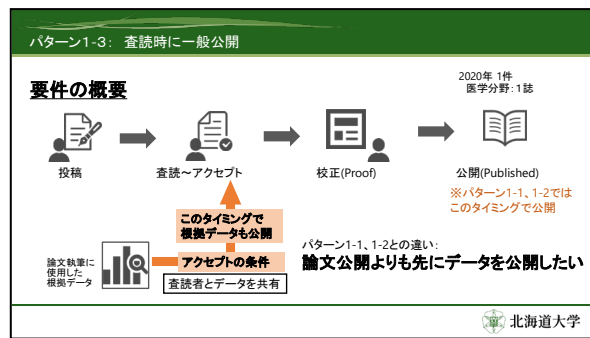
査読時にデータを一般公開したいケースです（図 6）。パターン 1-2 では、査読のタイミングでは査読者のみとの共有でしたが、今回は雑誌投稿規定により、データだけ先行して一般に公開する必要があります。これは、システム的には非公開を使わずにそのまま公開すればいいので逆にシンプルなのですが、HUSCAP で公開する場合、システム以外の問題で一つ手続きが必要になりました。現在の HUSCAP の運用規定では、原則として公表済みの資料を公開することを前提としています。これは論文の剽窃の疑惑を避けるためや、機関として公式に公表できる資料かどうかについて、既に学術資料として一般に公表済みであるということが一つの判断基準として有用であるためという理由が挙げられます。従来の登録資料は公開済みの論文が大半なので、この基準で特段問題はありませんでした。しかし、今回のパターンでは、論文がまだ公表済みではないため、付随するデータも未公表資料と考えられます。これは規定を明文化した北海道大学学術成果コレクション資料提供・公開細則に資料の種別として明示されている「公表済資料」に当てはまらないため、このパターンの場合は「その他、特に館長の認めた資料」として都度、館長が決裁して掲載しています。ただ、このポリシーについては今後改定して、もっとスムーズに研究データを公開できるようにする予定です。

パターン 2：投稿後の論文に付随するデータを公開

パターン 2 は、既に公開されている論文のアーカイブに伴って付随データを公開する場合です（図 7）。



(図 5)



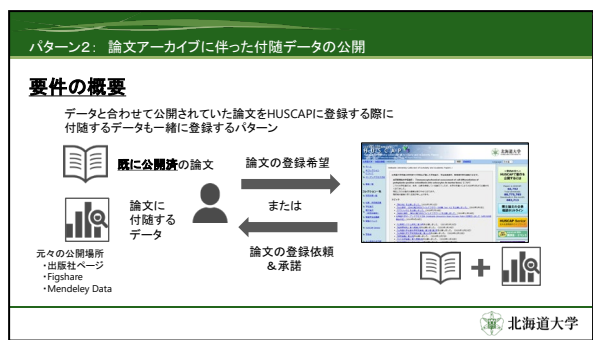
(図 6)

こちらは、データ公開の依頼時期は必然的に論文出版後になります。公表済みの資料をアーカイブするという点では、これまで機関リポジトリが行ってきた論文アーカイブ業務とほぼ同じです。従って、論文同様に研究データにもメタデータを付与してファイルを登録し、公開します。こちらは特段、非公開処理や運用規定などの判断が必要ないので、シンプルなパターンと言えます。

ちなみに、北海道大学の場合、論文アーカイブの経緯は主に二つあります。一つは、研究者から登録の依頼がある場合です。もう一つは、図書館が北海道大学の研究者が主要な著者となっている論文を定期的に検索し、依頼を出し、研究者から承諾を得てアーカイブをする場合です。

パターン1とパターン2に共通する対応

論文公開後の作業として、これまでのパターンの全てにおいて、メタデータを使って論文と研究データの関連付けを行っています。こちらのメタデータはJPCOARスキーマに沿っていて、その中で規定されているリレーション要素の属性 isRequiredBy で、元論文の DOI と研究データの関連付けを示すことができます。ここに入力することによって、出版社や学会のサイトで掲載されている論文へのリンクを貼ることができます。加えて、HUSCAP 内にも論文をアーカイブしている場合には、HUSCAP 内へのリンクを別に貼って関連付けています。



(図 7)

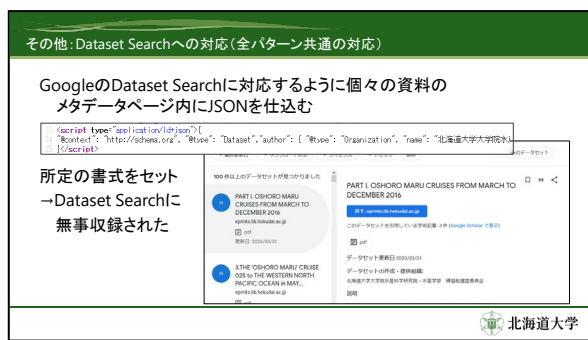
その他

その他に水産学部紀要「海洋調査漁業試験要報」を1冊データセットとして公開しています。元々冊子体で発行していたものをPDF形式でアーカイブしているもので、内容としては、北海道大学水産学部が所有する練習船「おしよろ丸」から得られたデータを記録しています。

また、研究データを公開する際の作業として、GoogleのDataset Searchという検索エンジンに収録されるよう、メタデータページ内にJSON形式でデータを入れています。所定の形式でメタデータを入れておけば、Googleによってクロールされる際に、その研究データがDataset Searchに拾われて収録されます。これは、Google Scholarの場合にクロールしてもらったためのデータとも多少形式が違い、Dataset Searchに対応したものを新しくコードを書いて出力したものと聞いています。図8のとおり、北海道大学水産学部附属練習船「おしよろ丸」のデータを掲載した紀要がDataset Searchに登録されていることが分かります。

研究データ公開例のまとめ

ここまでのまとめです(図9)。2017年からの4年間で14例のデータを公開しました。論文の根拠データの公開手順は、いつ登録・公開するか、誰と共有するかによって幾つかのパターンに分けられました。公開手順に関係する事柄としては、非公開状態やIDによる制限公開が思いのほか役に立ったこと、従来からの運用要綱に当てはまらないパターンもあること、デ



(図 8)

ータと論文を関連付けたり、検索エンジンに登録したりといった発展性を高める手法をいろいろと試してみたことなどが挙げられます。

研究データ公開事例から考える今後の課題

ここからは、今回の事例を踏まえて、今後の研究データ公開の課題を考えていきます（図 10）。今回紹介した事例は、割と限られた範囲の実例であるとも言えます。一方で、望ましい研究データ公開支援の先行研究や議論の中では、もっと幅広い事例でのデータ公開を想定しています。今回の事例でカバーできていなかった範囲についての対応が今後の課題と言えます。例えば今回の事例における公開理由は、雑誌のアクセプト要件だからというものがほとんどでしたが、今後の研究データ活用推進においては、その他の公開理由も出てくると想定されます。多彩な公開理由を想定した公開基準の策定、また、今回は特に考えていなかった再利用を促進するドキュメントの整備や、公開場所としての機関リポジトリの位置付けも、今後検討すべき課題と言えます。これらの課題について、一つずつ詳

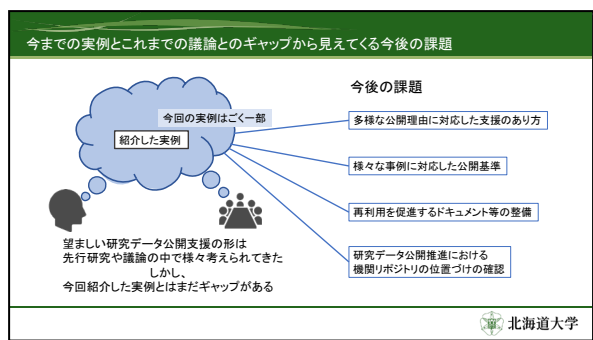
しく見ていきます。

まずは公開理由についてです（図 11）。先にも述べたとおり、今回の事例では、ほぼ全てのデータの公開理由が、雑誌の投稿規定にあるからというものでした。一方で、2017 年に科学技術・学術政策研究所（NISTEP）科学技術予測センターが行った実態調査の結果によると、データ公開を行った経験のある研究者の公開理由としては、「雑誌のポリシーだから」が全体の中で 2 番目に多かったです。一番多かった理由は「研究結果を認知してもらいたいから」で、その他に「科学研究や成果実装を推進したいから」「他の研究者のリクエストに応じて」などが続きます。このように、雑誌の投稿規定として義務化されているからというだけでなく、研究者側のモチベーションによってデータ公開を希望する事例も想定されます。この場合の疑問としては、そういった理由やモチベーションに対応したリポジトリの運用方針や公開基準がまだ明確ではないのではないか、それから、今回の事例では雑誌の投稿要件を満たしていればそれでよかったのですが、その他の理由で公開を希望する研究者については、また違った要件があるのではないかとあります。

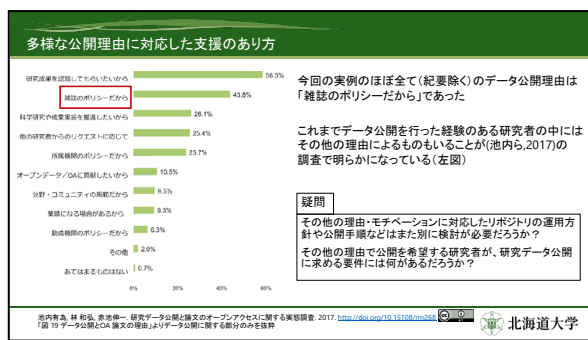
続いて、さまざまな事例に対応した公開基準についてです（図 12）。今回の事例では、全て論文に関連するデータで早急に公開する必要があり、なおかつ、個人情報や知的財産権などの問題がないと考えられたため、データの公開可否を特段問わず登録・公開しました。しかし、研究データの中には、公開に対して何かしらの問題が出てくるものも予想されます。例えば、

研究データ公開のまとめ	
2017年～2020年に登録した14例（論文の根拠データ13例、紀要1例） 論文の根拠データの公開手順はいくつかのパターンに分けられた	
公開手順を判断する要素	公開手順に関する事柄
いつ、登録/公開するか？ - 論文アクセプト前（査読時） - 論文アクセプト後（校正時） - 論文公開後	制限公開の機能 - URLのみ確定させ非公開の状態 - 特定の相手とのみ共有
誰と共有するか？ - 非公開状態 - 査読者とのみ共有 - 一般公開	研究データ公開を想定した運用要綱 データと論文の関連付け 検索エンジン（Dataset Search等）への登録

（図 9）



（図 10）



（図 11）

著者名が明記される論文と違って、データの場合、特にデータが単体の場合は、権利表記において誰がクレジットされるべきか明確に答えが出ません。また、共同研究者との間で契約上の問題がある場合や、分野内で別途ガイドラインが作成されている場合、個人情報や特許権などの法律的な判断が必要な場合も想定されます。

以前、私はNIIの実務研修において、公開可否判断ワークフローというものをつくったことがあり、このような問題についてどういった順番で判断していけばよいかということをフローチャート図にまとめました。しかし、実務において、このフローチャート図は誰か1人、あるいは単独部署でたどれるものではないことに気がきました。権利表記や共同研究者との合意は研究の関係者しか知り得ない情報ですし、例えばデータ公開が特許申請の何かしらの妨げになり得るかどうかの判断には、知的財産権の専門知識が必要になります。このように、公開に際しての問題を誰が判断し、解消するかについては、図書館などの主な支援部署単独では難しく、研究者や関連部署との連携が必要と言えます。

また、そのときに、研究データ公開に取り組む意義や問題意識をどうやって共有し、協働するかについては、現状集まっている事例だけ考えることは難しいと言えます。これは、複数の利害関係者のそれぞれの意見を聞きながら進めていくべき事項だと思います。

また、今回の公開事例では研究データ公開に際して特別な手順は定めておらず、通常、論文を公開するときと同様にメタデータを付与し、ファイルをリポジ

リで公開していました（図13）。一方で、これまでの研究データ公開の議論の中では、公開時に考慮すべき点がいろいろと挙げられています。例えば、国内のJPCOAR 研究データタスクフォース作成の「RDM トレーニングツール」によれば、ファイルの命名規則、ドキュメントの整備、再利用時のライセンスの明示なども考慮すべき点ですが、これらは今回の事例では特に行っていませんでした。理由としては、従来の論文の公開手順をほぼ準用したからということと、単純に出版社の定めた要件、すなわちファイルの公開以上のことは必要なかったからという2点が挙げられます。

ここで考えるべきことは、従来の公開手順では考慮できていない朱色で示した点について、これからどのように対応していくかということです。また、出版社の定めるファイル公開以上の要件、例えば他の研究者に再利用してほしいといった研究データ公開の要望が今後起き得るかどうか、起き得るとすれば、そこにどのようなモチベーションがあり、それに対してどういう公開手順で対応することが望ましいかを考える必要があると思います。

最後に、今回は北海道大学の機関リポジトリHUSCAP 上での公開事例のみを取り上げましたが、実際には、その他にも、例えば分野別のリポジトリや、Figshare、Zenodo などの汎用リポジトリが既に存在します。今回の事例で、HUSCAP でのデータ公開を選んだ事情が分かるケースについて図14に幾つか示しています。理由としては、既存のリポジトリが第1候補だったが制約を満たすことが難しく HUSCAP を選んだケース、既に他の公開場所で公開されているデー

様々な事例に対応した公開基準

今回の公開事例の場合：
特設、データの公開可否の判断をしていない
(明確な判断基準も現時点では存在しない)

公開可否判断ワークフロー例：
三上 倫子 研究データ公開ワークフロー、
(2019年度NII実務研修成果集、一部抜粋)
https://nii.ac.jp/2019/07/25/20190725_01/

データによっては公開の際に
確認すべき問題があることが予想される

- 権利表記の合意(誰を作成者とするか?等)
- 共同研究者との契約上、守秘義務が生じる場合
- ゲノム情報など分野ごとのガイドラインが存在する場合
- 個人情報を含む場合
- 知財関連(特許など)が絡む場合

判断に専門知識を必要とする問題もある

疑問

誰が問題を判断・解消すべきか?
支援部署だけでは対応は難しい?
研究チーム・あるいは関係機関へ事前に
専門部署(知財等)へ事前に対応可能性を協議?
複数の利害関係者間で問題意識を共有し、協働
できるか?

北海道大学

(図 12)

再利用を促進するドキュメント等の整備

今回の公開事例では特別な公開手順は定めていない(論文同様、提供ファイルをリポジトリで公開)

- ファイルの命名規則は特になし(論文公開同様)
- データの内容を説明するドキュメントは特になし
- 論文のAcceptanceに必要→出版社の定めた要件を満たしていればよい
- データの中身については研究者に任せている
- 再利用については考慮していない

研究データ公開で考慮すべき点(JPCOAR研究データタスクフォース作成「RDMトレーニングツール」(2017)より)

ファイルのフォーマット

- ファイルの組織化(命名規則)
- 再利用のためのドキュメントの整備
- メタデータの付与
- 発見性やアクセス性の向上
- 再利用時のライセンスの明示

※現時点で考慮できていないものを朱色で表記

疑問

従来の公開手順では考慮できない点に対し、
どのように対応していくか?
出版社等の要件以上のこと、例えば「再利用の
利便性を高めるため、ドキュメント整備を行うなど
の場合の研究者/支援者のモチベーションは何か?
どういった手順でそのモチベーションに対応できるか?

JPCOAR研究データタスクフォース RDMトレーニングツール 2017 <http://rdm.nii.ac.jp/1458/000000023/>

北海道大学

(図 13)

タをアーカイブしたケース、図書館で研究データ公開の相談を受けた流れで機関リポジトリでの公開を選んだケースがありました。

ここで一つ疑問となるのが、研究データ公開における機関リポジトリの位置付けです。論文公開においては、機関リポジトリはオープンアクセスの一つのGreen OA、すなわち許諾の得られた投稿済み論文をアーカイブするところで、長期保存が見込まれ、比較的出版社から許諾の出ることが多い非商用リポジトリという位置付けがありました。しかし、研究データ公開においては、そういった論文のオープンアクセスの事情とは異なるので、機関リポジトリに求められるものが何なのか、あるいは、目的に応じてその他の分野別のリポジトリや汎用リポジトリの活用も視野に入れて支援すべきかということがあります。

まとめ

今回の講演のまとめです（図 15）。前半で事例紹介を行い、後半で今後の課題を挙げました。ここで挙げた課題は支援部署（図書館や URA）が単独で解消できることではなく、研究データ公開の主体である研究者の意見や、各研究機関の体制や方針等も併せて考えていくべき難しい問題だと思います。今回の実践から見えた今後考慮すべき事柄はたくさんありますが、ひとまずは、今立っている場所を確認することが、望ましい研究データ公開の実践の足掛かりになればと思います。

緑色の丸い小鳥は「はすかつぷちゃん」といって、北海道大学機関リポジトリ HUSCAP のマスコットキ

ャラクターです。今後もはすかつぷちゃんと共に、オープンサイエンスの取り組みに注力していきたいと思っています。

●八塚 質問が幾つか寄せられています。一つ目は、「北海道大学で公開された研究データを使って2次分析を行った場合に出てくる研究成果を、リポジトリ側ではどのように扱う予定でしょうか」というものです。

●三上 今回については、特段ライセンス等は付与していませんが、例えば研究者がクリエイティブ・コモンズ・ライセンスで公開したいという場合はそれに従います。ご質問は、明示していない場合の対応でしょうか。

●八塚 恐らくそうだと思います。

●三上 その場合は、論文の転載もそうですが、いったん機関リポジトリ担当に問い合わせをいただいて、そこからデータの権利を持っている先生に問い合わせ、それにお答えするという形になると思います。

●八塚 続いて、「今回ご紹介いただいた事例には見当たりませんでしたが、例えば外部資金（科研費など）の募集要項においては、公開を推奨しているものが多いと思います。そういった理由が見当たらないというのは、もしかすると、研究者側の公開意識が低いので

研究データ公開推進における機関リポジトリの位置づけの確認

HUSCAPでのデータ公開を選んだ事情

第一候補だった分野別リポジトリはデータフォーマットが厳密に決められていたため、制約のないHUSCAPで公開することを希望した

既に他の場所で公開されているデータのアーカイブ
公開場所例: FigShare, MendeleyData, 出版社のリポジトリ

図書館で研究データを公開したいという問い合わせを受け、FigShare等の公開リポジトリと併せてHUSCAPでも公開できる旨を伝えた

初めからHUSCAPで公開希望との連絡があった

その他のリポジトリでの公開状況は現時点では把握していない

疑問

論文公開における機関リポジトリの位置づけ
→GreenOAに対応したアーカイブ公開場所

データ公開における機関リポジトリの位置づけとは何か？
(同じくデータ公開を担う分野別リポジトリ、公開リポジトリとの違いは何か)

北海道大学

(図 14)

まとめ

模索は始まったばかり

今後の課題

- 多様な公開理由に対応した支援のあり方
- 様々な事例に対応した公開基準
- 再利用を促進するドキュメント等の整備
- 研究データ公開推進における機関リポジトリの位置づけの確認
- とそれぞれについて現時点の疑問

事例紹介 +

新しい形を目指した雪だるまを作るはすかつぷちゃん
(HUSCAPのマスコットキャラクター)

北海道大学

(図 15)

はないでしょうか」という質問です。

●三上 確かに今回は、論文の投稿段階で公開したいという事例がほとんどでした。ただ、既に研究データが公開されていたもののアーカイブを行った事例もあり、機関全体で本当に外部資金で公開していないかどうかは分からないところです。機関リポジトリ担当も、相談を受けた分の研究データ公開の事情は把握していますが、それ以外のことは把握していません。実際に研究者がどう考えているかというのは、図書館にとってもなかなか把握が難しいところだと考えています。

●八塚 次の質問です。「研究者にとってはリポジトリが各種存在していて、どれを選択するのが適切なのか悩む場面もあると思います。北海道大学では、図書館を主体とした、研究データに対する相談窓口があるのでしょうか」。

●三上 まだ明示的に研究データに関して図書館にご相談くださいとアナウンスしているわけではありませんが、事実上、機関リポジトリに関連して、図書館の機関リポジトリ担当に相談が来るケースが多くなっています。その中では、機関リポジトリに公開したいと初めから決まっているわけではなく、どこかしらのリポジトリに公開したいという要望もあります。分野別のリポジトリまでは図書館の担当はあまり把握していませんが、一般的な Figshare、あるいは HUSCAP での公開も可能だという案内はしており、結局 HUSCAP で公開したケースはありました。

●八塚 次に、「三上さんのご経験から、データ公開を支援する部署は何人くらいの人員であることが理想的とお考えでしょうか」という質問を頂いています。

●三上 今、機関リポジトリ担当は4人で行っています。今後、研究データの取り扱いが増えてくるとして、今の機関リポジトリ担当と仕事が被る点もあるので、

図書館内であれば恐らく現状か、研究データの判断の専門人員が要るでしょうか。

●八塚 現状は4人で足りているのですか。それとも人手不足の状態なのですか。

●三上 正確に言うと、4人と、他に部局ごとの図書担当が対応しています。業務内容にもよるので、今の時点で考えるのはなかなか難しいところです。

●八塚 ありがとうございます。次の質問です。「先ほどご説明いただいた中で、パターン 1-3 の査読時に一般公開する場合において、もし論文がリジェクトされてしまった場合は、データもリポジトリから取り下げるとことはあり得るのでしょうか」。

●三上 リジェクトの場合だったか分かりませんが、過去に1度、実際にリポジトリから公開請求を取り下げたものがあります。北海道大学のリポジトリでは、研究データは1度公開したら取り下げ不能という規定は特に設けていないので、先生から取り下げしてほしいという要望があれば、そのように対応すると思います。

第 3 回 SPARC Japan セミナー2020

「初めての研究データ」

「リポジトリのグッドプラクティスのための COAR コミュニティフレームワーク」と 「国内機関における研究データ管理の取り組み状況調査」

安原 通代

(国立情報学研究所)

講演要旨



2020 年 10 月にオープンアクセスリポジトリ連合 (COAR) が "COAR Community Framework for Good Practices in Repositories" を公開しました。JPCOAR では、このフレームワークの日本語訳を作成し、2020 年 12 月に JPCOAR サイトで公開しました。本報告ではこのフレームワークの概要説明及び、2020 年に JPCOAR 研究データ作業部会 RDM 事例形成プロジェクトと AXIES-RDM 部会が協力し行った「国内大学・研究機関における RDM 取り組み状況調査」について紹介します。



安原 通代

2020年度SPARC Japanセミナー企画ワーキングメンバー。京都大学図書系職員。2020年4月から国立情報学研究所学術基盤推進部図書館連携・協力室オープンアクセスリポジトリ推進協会 (JPCOAR) 事務局担当。JPCOAR研究データ作業部会員。

本日の内容

本日は、前半では昨年 10 月にオープンアクセスリポジトリ連合 (COAR) が公開した「リポジトリのグッドプラクティスのための COAR コミュニティフレームワーク」についてお話しします。セミナーのテーマは「初めての研究データ」ですが、この COAR コミュニティフレームワークは、データに関する項目もありますが、研究データリポジトリのためだけのフレームワークではないので、その点をご了承ください。後半では、2020 年 11 月から 12 月にかけて AXIES-RDM 部会と JPCOAR 研究データ作業部会が連携して実施した「国内機関における研究データ管理の取り組み状況調査」について説明します。

JPCOAR とは

オープンアクセスリポジトリ推進協会 (JPCOAR)

は、2016 年 7 月に、リポジトリを通じた知の発信システムの構築を推進し、リポジトリコミュニティの強化と日本のオープンアクセスならびにオープンサイエンスに資することを目的として設立されました (図 1)。国公立大学図書館協力委員会と国立情報学研究所 (NII) との間の協定書に基づき設置された「大学

JPCOAR とは

- ・ オープンアクセスリポジトリ推進協会
 - JPCOAR : Japan Consortium for Open Access Repository
 - <https://jpcoar.repo.nii.ac.jp/>
- ・ 2016年7月 リポジトリを通じた知の発信システムの構築を推進し、リポジトリコミュニティの強化と、我が国のオープンアクセス並びにオープンサイエンスに資することを目的として設立
- ・ 国公立大学図書館協力委員会と国立情報学研究所 (NII) との間の協定書に基づき設置された「大学図書館と国立情報学研究所との連携・協力推進会議」と連携し、活動を行っている
- ・ JAIRO Cloud (共用リポジトリサービス) を NII と共同運営
- ・ 会員機関数 : 662機関 (2021年1月8日現在)

(図 1)

図書館と国立情報学研究所との連携・協力推進会議」と連携して活動しています。共用リポジトリサービスである JAIRO Cloud を NII と共同運営しており、2021 年 1 月 8 日現在、JPCOAR の会員機関数は 662 機関です。

JPCOAR の運営組織体制は図 2 のようになっています。参加機関の意思を直接反映させる場が総会です。運営委員会は、運営に関する基本方針を策定する役割を担っています。その運営委員会の下にあり、研修、広報など各種事業の具体的な活動の中心となる組織が作業部会です。作業部会は、JPCOAR の参加機関に所属する職員で構成されています。「国内機関における研究データ管理の取り組み状況調査」は、このうちの一つである研究データ作業部会が行ったプロジェクトの一つです。

事務局は、JPCOAR の活動に必要な業務を遂行する部署です。事務局の専任職員は参加機関からの出向職員となっています。私は京都大学からの出向職員としてここに所属しています。

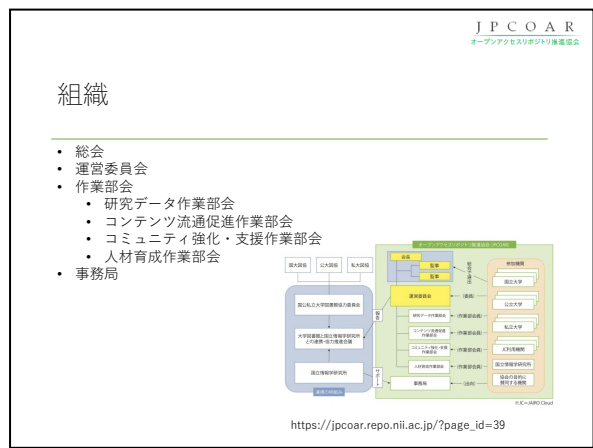
COAR とは

次に、オープンアクセスリポジトリ連合 (COAR) という組織について簡単に説明します。COAR は、図書館、大学、研究機関など世界中の 157 のメンバーとパートナーから成る国際協会で、JPCOAR もメンバーとして加入しています。国際的なリポジトリの協力体制による学術研究成果の可視化と利活用の促進を目的

としていて、世界中のリポジトリコミュニティの情報共有を中心に、コミュニティ間をつなぐ役割を果たす国際組織です。COAR がリポジトリの運用の評価、改善を支援するために策定したのが「リポジトリのグッドプラクティスのための COAR コミュニティフレームワーク」です (図 3)。これは 2020 年 10 月 8 日に COAR の Web サイトで公開されました。日本語訳は JPCOAR が作成し、同年 12 月 4 日に JPCOAR の Web サイトで公開しています。この日本語訳は JPCOAR の機関リポジトリからダウンロードでき、原文と同様に CC BY で利用可能になっています。COAR の Web サイト上では現在、原文の英語を含めて 8 カ国語で公開されています。

このフレームワークは、リポジトリの運用の評価、改善を支援することを目的として策定されています。他にも多くのリポジトリの評価フレームワークが存在していますが、それらの評価基準は異なる組織に散在しており、特定の地域や特定のタイプのリポジトリしか利用できないことが多いです。COAR フレームの目的は、そういった既存のフレームワークの基準をまとめて、さまざまな地域、さまざまなタイプのリポジトリで利用できるようにすることだそうです。

COAR の現事務局長である Kathleen Shearer 氏が回答しているインタビューの記事に、フレームワークの目的や今後の作業についてももう少し詳細に書かれていたので紹介させていただきます (図 4)。既存のフレームワークの多くは、求める基準、リポジトリ要件を



(図 2)



(図 3)

非常に高く設定していて、世界中の多くのリポジトリ、特にあまりリソースのないリポジトリはその基準に到達できないので、そういったフレームワークは無関係なものになってしまっている。そのため、COAR のフレームワークは、多くのリポジトリが採用できるように、基準を高く設定し過ぎないようにしたとのことです。また、COAR のフレームワークは、正式な認証を実施するために作成されたものではなく、リポジトリの改善を支援する自己評価のためのツールとされています。一方で、このフレームワーク作成の次のステップとして、より詳細なガイダンスをリポジトリの管理者に提供できるための作業が既に開始されているとのことです。

この記事は 2020 年 10 月 30 日に公開されたもので、インタビューの時点では 11 月から COAR のワーキンググループがそのような活動を開始すると事務局長が答えているのですが、現時点で COAR から次のステップに関する情報はまだ出ていません。

作成プロセス

次に、COAR のフレームワークがどのようなプロセスに基づいて作成されたかを紹介します (図 5)。まず、既に存在する多くのリポジトリの評価フレームワークを COAR のワーキンググループがレビューしてギャップを特定し、その重要性、妥当性、実施可能性のレベルを評価して、各特性を「必須」または「望ましい」に分類して草案を作成しました。2020 年 6～8

月にこの草案が COAR メンバーに配布され、実施が困難なものはないか、重要な基準が欠けていないか、解釈が難しいものはないか、不適切なものはないかという観点からさらにフィードバックを受け、それを反映して公開に至ったということです。

今後は、コミュニティがこれらのベストプラクティスを採用する際に役立つ事例やガイド、インストラクションへのリンクや引用を COAR のホームページで提供することを予定しているそうです。また、このフレームワークは完成形ではなく、毎年 1 回、7 月から 8 月にかけてフレームワークを見直し、最新のベストプラクティスを反映して更新することを予定しているそうです。

フレームワークで取り上げる特性の項目

フレームワークの内容について、ごく簡単に説明します。図 6 にあるような九つの項目に分類され、各項目に必須の特性と望ましい特性が挙げられています。実際に読んでいただくと、明確なチェック項目を定めているわけではないことがお分かりいただけると思います。

一つ目の項目では、リポジトリのリソースの発見可能性を高めるために必要とされる特性が取り上げられています。リポジトリ内のリソースが利用できなくなった場合でもメタデータが利用可能か、検索機能を提供しているかなど、必須とされる特性と、望ましいとされる特性に分けられています。

J P C O A R
オープンアクセスリポジトリ推進協会

目的

- 基準を高く設定しすぎずに、リポジトリがグッドプラクティスを採用できるようにするためのもの
- 正式な認証を実施するために開発されたものではなく、リポジトリの改善を支援する自己評価のためのツールであり、リポジトリのリソースの割り当ての正当性を示すために役立つ
- フレームワークの次のステップとして、ベストプラクティスに準拠できるように支援する、より詳細なガイダンスをリポジトリ管理者に提供するための作業が開始されている

Kathleen Shearer on COAR's newest community framework
<https://www.openaire.eu/kathleen-shearer-on-coar-s-newest-community-framework>

(図 4)

J P C O A R
オープンアクセスリポジトリ推進協会

作成プロセス

- COARワーキンググループが既存のフレームワークをレビューし、ギャップを特定し、その重要性、妥当性、実施可能性のレベルを評価し、各特性を「必須」または「望ましい」に分類
- 2020年6月～8月にCOARメンバーに配布され、コミュニティからの更なるフィードバックを受け付け、反映
 - 採用のしやすさ - 本質的な特性の中に、実施が困難なものはないか?
 - ギャップ - 重要な基準が欠けていないか?
 - 明確性 - 解釈が難しいものはないか?
 - 妥当性 - 不適切なものはないか?
- 今後、COAR はコミュニティがこれらのベストプラクティスを採用する際に役立つ事例やガイド、インストラクションへのリンクや引用を提供していく
- COAR リポジトリ評価作業部会は、毎年 7 月から 8 月にかけてフレームワークを見直し、最新のベストプラクティスを反映し、更新する

(図 5)

二つ目は、リソースへのアクセスを保証するための項目です。必須とされる特性として、ユーザーがリポジトリ内のリソースにアクセスするためのコストは発生しないこと。リポジトリは、公示された期間におけるリソースへの継続的なアクセスを保証すること。リポジトリは、障害者のために、文書やメタデータへのアクセスをサポートすること。デバイスの中立性として、ユーザーがリポジトリにアクセスするために特定のデバイスを必要としないこと。こういった四つの特性が挙げられています。具体的な期間や、サポートのための具体的な整備内容については規定されていません。期間が定められて公示されているか、何らかのサポートがされているかといった確認内容になっています。アクセスに関する望ましい特性としては、大容量ファイルの取り扱いや、アクセス制限のあるリソースへの間接的アクセスのサポートがあるかなどが挙げら

れています。

三つ目は、リソースの再利用についての条件や、フォーマットに関する項目です。必須の特性が四つと、望ましい特性が二つ挙げられています（図7、8）。

四つ目は完全性と真正性です。私は最初、この項目名をただけでは内容が分かりませんでした。原文では integrity and authenticity となっています。リポジトリのリソースが不正に変更されていないことを保証するための文書やポリシー、記録の保持に関する内容となっています。バージョンの管理については必須事項とされています。

五つ目は品質保証に関する項目です。メタデータやリソースの品質を保証する項目が図9のように挙げられています。

六つ目はセンシティブデータのプライバシーに関する項目です。特に研究データに関わる項目になるかと

J P C O A R
オープンアクセスリポジトリ推進協会

フレームワークで取り上げる特性の項目

- ・ 発見可能性
- ・ アクセス
- ・ 再利用
- ・ 完全性と真正性
- ・ 品質保証
- ・ センシティブデータのプライバシー（例：被験者情報など）
- ・ 保存
- ・ 持続可能性とガバナンス
- ・ その他

(図6)

J P C O A R
オープンアクセスリポジトリ推進協会

再利用（望ましい特性）

3.5 リポジトリは、フルテキストハーベ스팅やテキストマイニング・データマイニングをサポートするためのオープンなAPIを持つ

3.6 リソースは、機械可読なコミュニティ標準フォーマットで保存される

(図8)

J P C O A R
オープンアクセスリポジトリ推進協会

再利用（必須の特性）

3.1 リポジトリは、再利用条件を規定するライセンス情報をメタデータレコードに含む

3.2 リポジトリは、引用可能なPIDを提供する（1.4参照）

3.3 リポジトリは、推奨される、独自仕様でないフォーマットのリストを提供する

3.4 ランディングページには、引用に必要な情報を含むアイテムに関するメタデータが機械及び人間が読める形式で掲載される

(図7)

J P C O A R
オープンアクセスリポジトリ推進協会

品質保証

- ・ 必須の特性

5.1 リポジトリはメタデータ（及び該当する場合はリソース）の基本的なキュレーションを行う

5.2 リポジトリは、どのようなキュレーションプロセスがリソースとメタデータに適用されているかを概説した文書またはポリシーを提供する

- ・ 望ましい特性

5.3 リポジトリは、リソースとメタデータの外部注釈、コメント、レビューをサポートする

(図9)

と思いますが、リポジトリがセンシティブデータを収集している場合には、データ所有者が許可されたユーザーのみにアクセスを制限できる仕組みを提供していることが必須の特性とされています。望ましい特性になると、さらにデータのセキュリティ要件のレベルの違いに基づいて、段階的なアクセスを提供できることが加わります。

七つ目は保存に関する項目です。

八つ目に挙げられているのが、リポジトリの管理責任を持つ組織の明確化や、管理組織における運用体制の明確化についての項目です。この項目のみ、望ましいとされる特性はなく、全て必須とされる特性になっています。

最後の項目はその他です。支援窓口やヘルプデスクがあるかや、利用統計についてなど、他の八つの項目には含まれていない特性が挙げられています。

フレームワークで取り上げられている項目は以上です。必須となっている特性でも、ポリシー作成が必要だったり、機能面ではシステム構築から必要になったりして、すぐには対応が難しい内容も多く含まれています。最初にも申し上げましたが、COAR のフレームワークは、正式な認証を実施するために作成されたものではなく、リポジトリの改善を支援する自己評価のためのツールとされています。グッドプラクティスに基づいて自機関のリポジトリを自己評価して改善点を見つけ、さらに将来に向けて機能やポリシー、運用について検討するために活用するものと考えています。

JPCOAR の関係者から、このフレームワークを基にして日本国内に向けたチェックリストを作ってはどうかという案が出されており、現在、ミーティングや意見交換を行っています。また、COAR のフレームワークも毎年更新が予定されているので、その進展についても今後報告していきたいと思っています。

「国内機関における研究データ管理の取り組み状況調査」

次に、2020 年末に実施した「国内機関における研

究データ管理の取り組み状況調査」について紹介します（図 10）。この調査は、JPCOAR の作業部会の一つである研究データ作業部会が AXIES-RDM 部会と連携して行っている RDM 事例形成プロジェクトの一環になります。RDM 事例形成プロジェクトは、参加機関による RDM の取り組みを支援し、データポリシー策定の足掛かりを作成することや、ディスカッションやワークショップを通じて事例集を作成・公開することなどを目的としています。プロジェクト全体については、2020 年の NII オープンフォーラムで、NII オープンサイエンス基盤研究センター（RCOS）の南山泰之先生が報告されていますので、ご参照ください。

今回行った「国内機関における研究データ管理の取り組み状況調査」は、国内の大学・研究機関を対象に、機関内での研究データ管理の取り組み状況を把握することを目的として、JPCOAR 会員機関、AXIES 会員機関、その他の国内機関にご協力いただいたアンケート調査です（図 11）。質問は全部で 46 問あり、七つのセクションに分かれています。それぞれのセクションについて、内容に応じて適切と想定される回答部署を記述していますが、想定回答先は各大学・研究機関の体制によって異なると思いますので、あくまでも想定です（図 12）。

この調査を行うに当たっての AXIES-RDM 部会と連携体制は図 13 のような形になっています。こちらの図も 2020 年末の AXIES 年次大会で RCOS の南山先生が講演で使われたものを使用させていただいています。

J P C O A R
オープンアクセスリポジトリ推進協会

RDM事例形成プロジェクト (AXIES-RDM部会との連携事業)

- ・参加機関によるRDMの取り組みを支援し、データポリシー策定の足掛かりを作成する
- ・ディスカッションやワークショップを通じて事例集を作成・公開する
- ・AXIES-RDM部会とJPCOAR研究データ作業部会が連携して実施する

2020 NIIオープンフォーラム
AXIES研究データマネジメント部会合同トラック「学術機関における研究データ管理フレームワーク」
https://www.nii.ac.jp/openforum/2020/day2_os4.html
研究データ管理のステークホルダーと事例収集 南山泰之（国立情報学研究所）
<https://www.nii.ac.jp/openforum/upload/cb0bfc6d76ac96dd6c1340e2f86e6e68b4c155.pdf>

(図 10)

調査開始時（11月）の JPCOAR の会員機関数は 657 機関、AXIES の会員機関数は 133 機関、両方に重複して加盟している機関が 112 機関でした。また、どちらにも加盟していない国内機関にもご連絡して、最終的には合計 658 機関にご協力をお願いしました。

Google Form 経由での回答をメールで依頼して、Google Form にアクセスできない機関については、Word ファイルに回答を入力の上、メール添付での送信をお願いし、最終的に 352 の機関から回答を頂くことができました（図 14）。内訳としては、研究開発機関が 62%、国立大学が約 19%、公立大学が 12%、私立大学が 2%、大学共同利用機関等が 2%です。もう少し内容を報告できたらよかったのですが、現在、報告書は取りまとめ中になっています。

プロジェクト自体は 2019 年度から 3 年を予定しており、本年度が 2 年目です。今年度はアンケート調査

自体に対する報告書をまとめて、来年度は、調査結果やプロジェクトの成果報告などをまとめて公開します。

その他の研究データに関する取り組み

最後に、JPCOAR の作業部会による活動の中から、研究データに関するその他の取り組みを幾つか紹介します。

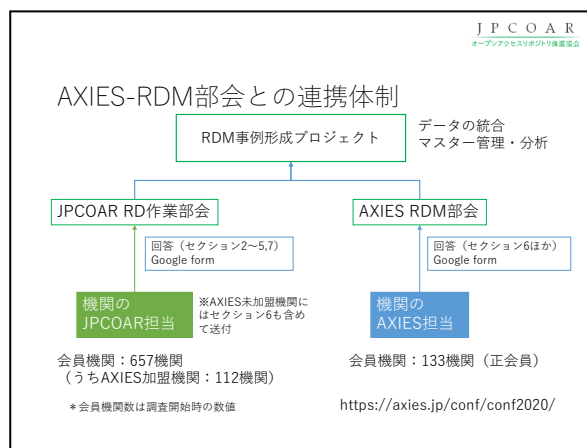
2020 年 10 月に教材「研究者のための研究データマネジメント」を公開しています。研究支援者の目線から、大学や研究機関等に所属する研究者の方に向けて作成した教材集です。研究データ管理の場面に応じた 12 のテーマに分かれていて、研究者自身がこの教材によって必要な知識を得る他に、研究を支援する側が各教育機関の研究環境やニーズに応じた形で、この教材を加工して研究データ管理サービスを提供することを想定しています。また、これは 2017 年公開の研究

J P C O A R
オープンアクセスリポジトリ推進協会

国内機関における研究データ管理の取り組み状況調査

- 目的：国内の大学・研究機関を対象に、機関内での研究データ管理の取り組み状況を把握する
- 調査期間：2020年11月27日～12月28日
- 質問数／所要時間：全46問／約50分
- 対象：JPCOAR会員機関、AXIES会員機関、その他国内機関（計685機関）

(図 11)



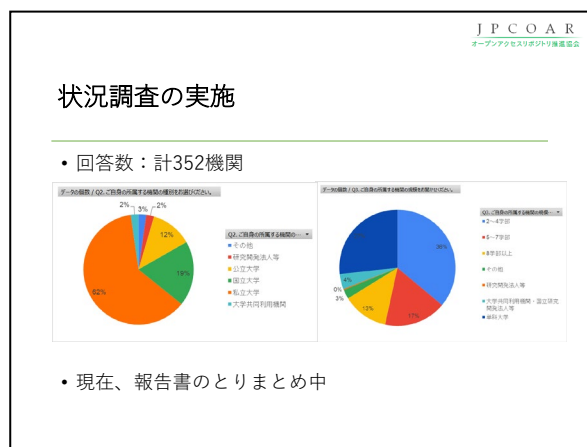
(図 13)

J P C O A R
オープンアクセスリポジトリ推進協会

セクション別回答想定先

- セクション1（概要説明）：（全1問）
- セクション2（基礎情報）：JPCOAR参加機関 窓口担当部署または機関リポジトリ担当部署（全3問）
- セクション3（ニーズの把握）：機関内の研究データ管理（RDM）担当者／対応部署（全5問）
- セクション4（データ管理体制の構築状況）：研究推進部署／URA等（全6問）
- セクション5（研究データ管理サービスの実施状況）：大学図書館等（全11問）
- セクション6（情報インフラの整備状況）：情報基盤センター等（全18問）
- セクション7（オープンアクセスリポジトリ推進協会（JPCOAR）について）：JPCOAR参加機関 窓口担当部署または機関リポジトリ担当部署（全2問）

(図 12)



(図 14)

支援者向けの教材「RDM トレーニングツール」、2018 年公開の教材「研究データ管理サービスの設計と実践」に続く第 3 弾の教材となっています。2018 年公開の教材「研究データ管理サービスの設計と実践」については、2 月 10 日に第 2 版が公開されています。こちらは大学図書館員や URA、技術スタッフなど、研究支援職員が研究者の研究プロセスに沿って研究データ管理サービスの設計と実践の方法を身に付けることを目的としたものです。

そして、3 月にはデータベースレスキューマニュアルを公開する予定となっています。

●八塚 質問を頂いています。「COAR のフレームワークの JPCOAR 版という話がありましたが、研究データ利活用協議会（RDUF）が昨年策定したガイドラインとの違いや位置付けなどについて教えていただくと幸いです」。私が推測するには、これは RDUF の中のジャパン・データリポジトリ・ネットワーク（JDARN）で作成したリポジトリガイドラインのことだと思いますが、その辺の違いで、もしご存じのことがあれば教えていただけますか。

●安原 すみません、私の方で明確な違いは分かりません。

●八塚 違いについてはこれから検討しようとしているところだと思いますので、なかなか見えてこないと思いますが、私の感覚では、COAR のフレームワークは CoreTrustSeal などと違って、誰かに認証されるような類いではなく、あくまでもリポジトリ側で自己改善していくためのツールという位置付けだと思います。そういうことでよろしいでしょうか。

●安原 はい。そのとおりです。

●八塚 ありがとうございます。では、次の質問にしたいと思います。「調査状況については報告書を取りまとめ中ということでしたが、もし興味深い結果があれば、何か情報を頂けませんか」。

●安原 「研究データの管理サービスを展開するに当たって、自機関で、研究者から要望が高いと思われるサービスをお聞かせください」という質問があり、複数選択可として、こちらからデータ管理計画（DMP）の作成支援や研究データのストレージ提供など幾つかの選択肢を挙げたのですが、最も回答数が多かったのは「分からない」でした。その上で、「ここで選択した研究データ管理に関する支援を実施する際に必要と思われる知識やスキルはどういったものでしょうか」という質問をして、データキュレーションスキルや ICT スキルといったスキルを幾つか挙げて、こちらも複数回答可としたのですが、ほぼ全てのスキルを「はい」とした方が多かったです。ですので、どういった要望が高いかは分からないけれども、あらゆるスキルや知識が必要とされているという認識があるのだらうと思いました。

このアンケートを行った際に、セクションに分けて、取りまとめを依頼している部署から各関連部署に質問していただいたのですが、その際に非常に思ったのが、研究データ管理における関連部署の連携の重要性です。必要とされるものは分からないけれども、必要とされる能力は多岐にわたっているので、これは本当に一つの部署では対応できるものではありません。そうすると、学内・機関内のさまざまな部署で連携していくことが必要だと感じました。

●八塚 今の質問に関係するのですが、「結果については取りまとめ中ということですが、どういった設問だったのかをご紹介いただけないでしょうか」という要望が来ています。

●安原 図 12 のように、セクションに分かれており、

ニーズの把握やデータ管理体制の構築状況、研究データ管理サービスの実施状況、情報インフラの整備状況、最後は JPCOAR についての質問もさせていただきました。何らかの研究データ管理サービスを実施しているか、その中で図書館が実施しているものはあるかといった質問や、現在はしていないけれども図書館で実施を計画・検討しているものはあるか、また、過去に研究データ管理に関する研修やワークショップ、シンポジウム等のイベントを実施したことはあるかといった質問で、全部で 46 問あります。

●八塚 設問についても、まだ公開されていないというのでしょうか。

●安原 設問自体は機関の皆さまにお送りしただけで、まだ公開していません。

●八塚 私から一つ質問させてください。今回、COAR のフレームワークを翻訳されて、これが現状の日本の機関リポジトリや分野別リポジトリにどれだけマッチしていそうかという感触は何かありますか。

●安原 やはり日本のリポジトリに当てはめるには少し難しいものがあると感じまして、JPCOAR 内で、これを基にして日本に向けたチェックリストを作ろうということになっています。

●八塚 例えばどのあたりが食い違っていそうですか。

●安原 特定のことになってしまいますが、私は、障害者が文書やメタデータにアクセスするのをサポートするあたりで、少し具体的なリソース例やサポート対応を挙げていく必要があると思いました。

第3回 SPARC Japan セミナー2020

「初めての研究データ」

パネルディスカッション



朝岡 誠	(国立情報学研究所)
林 賢紀	(国際農林水産業研究センター)
加藤 斉史	(国立研究開発法人科学技術振興機構)
込山 悠介	(国立情報学研究所)
神谷 信武	(チューリッヒ大学)
三上 絢子	(北海道大学附属図書館)
安原 通代	(国立情報学研究所)

アンケート

●朝岡 ディスカッションの前に、Slido のライブ投票機能を利用した簡単なアンケートを行います。質問は二つあります。

一つ目は、加藤様から紹介があった J-STAGE Data を利用したいかどうかという質問です。研究者の皆さまは、自分あるいは自分のプロジェクトで使ってみようかどうか、図書館員や URA、学術出版関係の皆さまは、自分の業務に関係する研究者に薦めたいかどうか、「ぜひ利用したい」「利用したいが課題がある」「他のリポジトリを使いたい」「そもそも Web を介したこのようなサービスを利用して研究データの管理または公開を行いたくない」という4択でお答えください。特に「利用したいが課題がある」という方は、Q&A でその課題または要望について書いていただけると幸いです。

二つ目は、込山様から紹介があった GakuNin RDM の利用についての質問です。ご回答をお願いします。

—アンケート入力—

それでは、回答を締め切らせていただきます。ご回答いただきありがとうございました。

皆さまの回答結果を林さんから報告していただきます。

●林 まず、加藤様の講演で紹介していただいた J-STAGE Data については、「ぜひ利用したい」「利用したいが課題がある」が多かったです。課題としては、今までご説明いただいたような中身だけなのか、あるいはもう少し何か情報があるのか、また、論文がリジェクトされたときにデータはどうなるのかなどが挙げられていますが、このあたりはこの後のディスカッションで広げてゆきたいと思います。

●朝岡 ありがとうございます。加藤様、この結果についてコメントを頂けますでしょうか。

●加藤 今年度から本格運用に移りたいと思っていますので、「利用したい」とご回答いただいた方、ありがとうございます。ぜひ問い合わせを頂ければと思います。一番多かったのが「課題がある」ということで、本格運用には入れると思っているのですが、まだ課題

解決に至らないところは、またお伺いしながら改めていきたいと思います。本日伺えれば伺いますし、後日でもお問い合わせいただければと思います。

●朝岡 ありがとうございます。

次に、GakuNin RDM についての回答結果を林さんから紹介していただけますか。

●林 こちらも「利用したい」「利用したいが課題がある」が合計 87%で、ほぼ皆さん使ってみたいということですが、「課題がある」が半分くらいで、「他のサービスを使いたい」も若干あります。ただ、質問でもそれほど大きな課題はなかったような気がしますが、そのあたりはどうでしょうか。

●朝岡 そうですね。GakuNin RDM 自体に何か課題があるというより、もしかしたら自分たちの機関内でポリシー的な意味で課題があるということかもしれません。込山様、今のこの結果についてコメントをお願いしますでしょうか。

●込山 GakuNin RDM はサービス開始直後なので、これだけ「利用したい」という声を頂けたのは非常にありがたいと思っています。同時に、課題がたくさん残っているということで、もう少し機関の皆さまの声を聞きながら、開発に反映できることは反映し、本日のお話にあったような共通のフレームワークや何らかのルーブリックに落とし込んでいけるのであれば、それに合わせたシステムの改修を進めていくことで共通化を図っていきたいと考えます。

●朝岡 ありがとうございます。

本日の講演内容の振り返りとディスカッション

●朝岡 それでは、パネリストの皆さまの報告を振り

返りながら、本パネルディスカッションの論点について説明します。

本日の講演では、研究データ管理と研究データ公開のためのインフラとその実例、そして研究データ管理・公開のグッドプラクティスを構築していく取り組みについて講演していただきました。研究データ管理のインフラについては込山様に、エディンバラ大学の実例とグッドプラクティスを構築していく取り組みについては神谷様に、研究データ公開、とりわけ論文の根拠データ公開のインフラについては加藤様に、北海道大学の実例については三上様に、グッドプラクティスを構築していく取り組みについては安原様に紹介していただきました。

研究データの管理も公開も、神谷様や三上様が説明されたように、手を付けられるところ、要望があったところから始めていくことが重要です。自分たちに何ができるかというのは、神谷様が紹介された Evaluate your RDM Offering や、安原様が紹介された COAR コミュニティフレームワークの下で、サービスの長期的目標を立てて展開していくものであるということを、本日の講演を聞いて実感しました。取りあえず手の届くところから始めてみよう、そして他の人の取り組みをまねしてみよう、それで長期計画を立てていこうということです。

その取り組みを行っていく中で、込山様が紹介された GakuNin RDM や、加藤様が紹介された J-STAGE Data や、今の機関リポジトリが活用されることになると思います。今回、「初めての研究データ」というタイトルを見て、すぐにでも研究データ管理・公開を始めたい聴衆の皆さまからすると、自分の組織はどうやって始めていけばいいのか、誰がイニシアティブを取っていけばいいのか、研究者側なのか、それとも機関側なのかということについて疑問を持たれたのではないかと思います。GakuNin RDM も J-STAGE Data も、どちらも研究者や学会がイニシアティブを取って活用するインフラではありますが、実際、北海道大学やエディンバラ大学の事例を見ると、研究教育機関がデー

タポリシーにのっとってイニシアティブを取って行っており、研究者と、研究教育機関のスタッフである図書館員の皆さま、URA の皆さま、場合によってはインフラの関係者の皆さまとの関係性が重要だと感じました。

また、今回、海外におけるグッドプラクティスに向けた取り組みが紹介されましたが、海外の Evaluate your RDM Offering や COAR コミュニティフレームワークでは少し敷居が高いのではないかと、日本版のこういったものができるのかどうか、また、できた場合、自分たちも実践として使えるのかどうかについて疑問を持たれたのではないかと思います。

そこで、本日は次の三つの論点を中心にディスカッションを進めていき、その中で皆さまから Slido で頂いている質問に回答していきたいと思います。

第1の論点として、研究データ管理における機関と研究者の関係についてディスカッションしていきます。データ管理の話題ということで、まずは神谷様から、エディンバラ大学等、海外の機関では機関と研究者はどのような関係になっているのかコメントを頂けますでしょうか。また、エディンバラ大学の事例を紹介していただいた中で、RDM のロードマップが刻々と変化したと聞きましたが、機関と研究者の関係性はロードマップの進行とともに変化が見られたのかどうかについて、お聞かせいただけますか。

●神谷 私はエディンバラ大学で働いているわけではないので、確実な情報は出しづらいのですが、発表でエディンバラ大学の該当ページを紹介したように、研究データ管理サービスのところは、研究者に対して研究データ管理のやり方などのサポートをしていますし、研究データを取り扱うツールに関しても扱いやすいように改善しています。そういうところを含めて、機関がサポートしているところは確実にあるはずです。

ロードマップによって研究データに関するサービス内容が変化しているので、もちろん研究者と機関の役割が変わってきています。また、研究データ管理を、

ある機関の中でインフラとして定着させるときに、研究者からの理解とコミットメントが確実に必要になります。ロードマップを一緒に作成しながら、研究者がこれまで研究データ管理のサービスを受けてきたけれども、これから改善したいのはここですよねという話は確実にしていく必要があると思います。ですので、役割分担は、そのコミットメントの具合やステージによって変わってくるのではないのでしょうか。

●朝岡 ありがとうございます。

次に込山様にお伺いします。GakuNin RDM の設計者として、機関と研究者の関係をどのようにお考えになって GakuNin RDM を設計したのでしょうか。実証実験を通じて、その中で分かったことも含めてコメントを頂けると幸いです。

●込山 GakuNin RDM の設計の基軸という意味では、機関と研究者の軸、もう一つ、研究の推進と公正という軸があります。一見すると、機関はガバナンスの強化で研究データ管理を研究者にやらせたいというように捉えられてしまうのですが、決してそれだけではありません。機関もちろん研究を推進しようとしていますし、研究者は研究者で、自分たちの研究の不正防止だけではなく、真面目に取り組んでいる研究者からすると、何か外から訴えられたりしたときの証明手段にもなるなど、幾つかの軸で切って考えていきました。

実際の設計の進め方としては、本日も幾つか紹介がありましたが、国内外の RDM サービスのフレームワークからシステムに関する項目やガイドラインを切り出して、分解してみ、機能要件を考えて、それを軸に、先ほど言った二つの観点から分類して、これは RDM の必須機能だろうというところから優先的に埋めていったというのが設計段階で行ったことです。

実証実験の段階になると、大学側からの要望なども出てくるので、それをいかに束ねるか。全ての要望に応えることはできませんが、より多くの大学から要望が出ている機能については取り組んでいこうとしまし

た。そこでは、研究者が望むことと組織が望むことの違いや、どの時期にどの機能を充実させていくかということに悩みました。国内の場合、当初はどうしても研究不正防止や研究データの10年保存の影響を受けてデータ管理をしたいという、どちらかという機関の研究担当理事などに理解していただくための実装や説明を進めてきて、最近になってようやく研究者に近い機能を充実させはじめたという経緯があります。

●朝岡 ありがとうございます。

次に、データ公開の見地から、加藤様、三上様、安原様にお伺いしたいと思います。研究データ公開は研究データ管理に密接に関連していて、研究データ管理の時点で必要なメタデータ情報がないと、十分な形で研究データを公開することができません。研究データを公開する立場からすると、研究データ管理における機関と研究者の関係はどのようなものが望ましいと思われるでしょうか。まずは加藤様からコメントをお願いします。

●加藤 私たちからは、研究者の所属機関はあまりよく見えていないところがあります。所属機関に関しては、所属機関がデータに関するポリシーを決めていくと研究者が共有を進めやすくなるということぐらいしか見えていません。発行機関についても同様ですが、ポリシーにてデータ共有について定めるジャーナルが増えてきています。こういったことでポリシーが決まってくると、研究者にもデータ共有をするということが広まっていくので、世の中的にもデータ共有という考え方が浸透していくのではないかと考えています。

●朝岡 ありがとうございます。次に三上様、図書館員としての見地も含めてコメントを頂けると幸いです。

●三上 研究データ管理については、機関と研究者との間でメリットを共有できるのが望ましいと思っています。

ます。支援側で研究データ公開を推進しようと思っても、管理に必要なメタデータと公開に必要なメタデータは違うと思うのです。公開に必要なメタデータには研究者の余分な労力を使う以上、研究者側にメリットがないものを作ってほしいと言ってもあまりうまくいきません。ただ、研究者側でメリットを判断する上で、助成や出版社など研究データ公開が必要になっていく側面について、URAや図書館はいろいろと情報を持っています。そういった情報を幅広く持つ支援部署が研究者に向けたサポートや情報を出していった研究者をサポートする形がいいと思います。

あとは、機関側から研究データの公開をするための管理や公開の意思決定のための情報提供、支援の案内をして、研究者の方でメリットを感じてくれるような状態が望ましいと考えています。

●朝岡 ありがとうございます。

最後に安原様、2020年11月から12月に行われた「国内機関における研究データ管理の取り組み状況調査」の結果も踏まえてコメントを頂けると幸いです。

●安原 今回のアンケート調査を行って、研究データ管理における、サポートする側の関連部署の連携の重要性を改めて感じました。それとともに、研究データライフサイクルの中のどの段階でどのような課題があるのか、研究者の方はそういう場合にどこに相談するかを分かりやすくすることで相互にメリットを感じていただけるのではないかと思います。

●朝岡 ありがとうございます。

●林 Slidoに頂いている質問にもありましたが、研究データを公開していかないといけないという話が出てきたとき、機関の中で図書館員ができることは何かというと、例えば「メタデータを付けることは得意です。目録でずっとやってきた業務だし、できます。しかし、研究の中身はさっぱり分かりません」となりま

す。しかし、いろいろな政策文書などを見ていると、図書館職員がうまくコミットしてサポートしていきとされています。今までは論文とそのメタデータを扱うことで学術情報の流通を担ってきたわけですが、それが研究データにもどんどん幅が広がってきていて、だから私たちはこういったセミナーを開催するし、サポートもするという事で、先ほど三上さんから、大学という現場でいろいろな取り組みをされている事例をご紹介いただきました。

では、そもそも図書館や図書館職員というのは、こういう研究データを出す、公開する、管理するという課題を持っている機関や研究者にとって相談しやすい人たちなのでしょうか。相談したいのだけれども、何か忙しそうだし、最近は電子ジャーナルもあるから電子もいけそうだけれども、紙しか分かっていないのではないかと思われていたら嫌だなと思ったりするので。もちろん図書館職員はサービスをするのが仕事ですから、いろいろなところに出て行って、こんなこともできます、あんなこともできます、サポートをやりますよと言って歩くのが大好きで、私も好きです。しかし、研究者から見たときに、相談しやすい、話を持っていきやすい存在なのか。神谷様や三上様、何かご経験や知見があれば教えてください。

●神谷 もちろんそれは各図書館員が置かれている状況によって違うと思うので、私からの視点が正しいかというとは何も言いがたいのですが、私は日本学の図書館で働いていて、先生方とは、コロナ禍ではない状況だと会う機会が結構頻繁にあり、インフォメーションリテラシーなどの授業も入門コースの中で受け持たせてもらっているという経緯があります。ですので、話しかけやすい方なのではないかとは思いますが。ただ、他のところでもそうなのかというと、それは少しまた別の話なので、やはり各図書館の置かれている状況を見ながらということではないでしょうか。

●林 なるほど。確かに「いや、うちの図書館は何か

話しにくいんだよね」みたいな状態だと、それで全体として回るのかというのはちょっと分かりませんね。

●神谷 私がなぜ研究データ管理を大事だと思っているかというと、私はもちろん公開された研究データを使いたいと思っていますし、それをオープンサイエンスとして再度利用するのに役立てていきたいと思っているからです。そういう意図があるときに、私が研究者に対して、「図書館でもそういうサービスを行おうと思っています」という話を振っていきけるような状況をつくっていければとは思っています。それは図書館員が思っただけで話し掛けられない限り、起こりません。やはり誰かがイニシアティブを取って話し掛けられない限り、何も起こらないのではないのでしょうか。

●林 確かにそうですね。

●三上 それでいえば、今回の公開事例の中には、HUSCAP は論文を公開する場所だと初めから知っていて研究データを公開したいという先生がいらした記憶があります。HUSCAP は定期的に北海道大学で論文を出した研究者の方に「登録しませんか」とお誘いのメールを送ったり、私の上の世代の話なのであまり詳しくはないのですが、立ち上げのときに少し営業で回ったりしたという経緯があって、一応、機関リポジトリという論文を公開できる場所があるという認知があったのが相談につながったのかなと思っています。

もう一つ、GakuNin RDM を図書館で宣伝していると、最初に GakuNin RDM で研究データを公開できるかという問い合わせが来て、GakuNin RDM ではできないけれども、Figshare や HUSCAP のようなリポジトリであれば可能ですとご紹介することがあります。そのように、何かしら研究データや公開が図書館と結び付くヒントがあれば、あるいは、という気はします。

●林 確かに北海道大学の HUSCAP は、割と機関リ

ポジトリのはしりのところからずっと運用されているというイメージがあります。それだけの蓄積と認識があるからこそ、そこに行けば自分たちのものを出してもらえる、情報が分かるだけではなくて発信もできるというイメージの切り替え、あるいは新しい機能のイメージを付けることがうまくいったのかなと、話を聞いていて思いました。

●朝岡 おっしゃるように、研究者と機関のどちらがするかというよりは、まずは研究データの管理にしろ公開にしろ、やらなければいけないということよりも、管理や公開をすることで自分の意見が発信できるのだというように、ポジティブな研究者なり図書館員をうまく具合に結び付ける環境自体が必要だと思いました。

それでは、次の論点に移りたいと思います。第2の論点は、研究データ、特に論文の根拠データの公開についてです。加藤様が発表されたように、海外のジャーナルに投稿する場合、根拠データの公開が義務付けられていることもあり、研究者からのニーズが高まっていますが、リポジトリ側はその際にどこまでサポートすべきか、または GakuNin RDM などの RDM で管理した研究データを公開するに当たってどのような点に連結していくのか、特に研究データ管理と研究データ公開はそれこそ連結しているものなので、その連結についてどのように考えていくべきなのかということを議論したいと思います。

まずは研究データ公開ということで、加藤様と三上様にお伺いします。J-STAGE Data では投稿後の論文の根拠データを対象に、論文著者、研究者側が根拠データをメタデータも含めてアップロードするという研究者ファーストのスタイルを取っていて、ジャーナル発行機関は公開にあまりタッチしません。一方で北海道大学では、機関に所属する研究者の要請に応じて図書館員がメタデータを準備して、根拠データをアップロードする、そして投稿前の論文の根拠データの公開に当たっては査読者との限定的な共有対応なども行っています。ジャーナルと機関では幾らか違いがあるか

もしれませんが、研究者とそれをサポートする機関の関係は、研究データの公開においてはどうあるべきだと思いますでしょうか。研究者が上げたいデータを自分で上げていくというセルフアーカイブみたいなやり方もあると思いますが、機関リポジトリの方で幾らか研究者をサポートし、機関として研究データを公開するパターンもあると思います。あとは、根拠データの公開において必ずしておかないといけないこと、例えばライセンスの確認や必須メタデータの確認なども含めてコメントを頂けると幸いです。まずは加藤様、J-STAGE Data での査読前の論文の取り扱いに対する展望も含めてコメントを頂けると幸いです。

●加藤 少し緊張して発表でうまく伝わらなかったのですが、まず、J-STAGE Data はジャーナル発行機関のためのサイトですので、ジャーナル発行機関がタッチしないということではなくて、ジャーナル発行機関に投稿されるものが最終的には発行機関の判断で公開に結び付くという流れになっています。その前に、ジャーナル発行機関の中で J-STAGE Data と別に投稿を受け付けて、審査の上で OK であったものを J-STAGE Data に登載します。また、著者が直接 J-STAGE Data に登載していくことも可能です。それをジャーナル発行機関が確認した上で公開するかどうかというルートの違いがございしますが、基本的にはジャーナル発行機関が確認して公開に結び付きます。

査読前の原稿の根拠データについても、先ほどの流れのように公開前に載せることができますけれども、同じようにジャーナル側で判断して公開することになりますので、ジャーナル側のポリシーによりませんが、査読前に公開してよいというポリシーを取らなければ公開されないということになります。

そういう意味ではジャーナルと結び付いた形になっていますので、投稿先がまだ決まらない、査読前の原稿の段階で根拠データを載せる先としては J-STAGE Data はふさわしくないという状況になっています。

併せて、先ほど説明の中でリジェクトのご質問を頂

いたときの回答で間違っただけをしてしまいましたので訂正します。リジェクトの場合、発行機関の中で査読審査をしている中でリジェクトになっていますので、データ自体、まだ公開されていません。また、メタデータも公開されていない状態ですので、その状況の中で削除するかどうかは発行機関が決めることができます。

●朝岡 ありがとうございます。関連して、J-STAGE Data について、パイロット運用参加中の学会の方から質問が来ていますので紹介します。「学会として、ジャーナル側ではなくて投稿者に、論文の受理後のデータのアップロードの公開などのオプションを付けたいと考えている機関があります。そのようなことを、あくまで学会として投稿者に提供するサービスの一つとして位置付けることについて、J-STAGE としてはいかがお考えでしょうか」ということですが、いかがですか。

●加藤 論文にひも付かなくてもデータを載せたいということですね。J-STAGE Data のポリシーとして、J-STAGE 掲載の論文と関連するデータとなっていますので、関連度にもよると思いますが、ひも付ける論文がない場合には対象外となります。

●朝岡 上げられるデータはかなり限られている感じですね。

●加藤 そのとおりです。論文とひも付いているということになりますので、その論文を書くに当たって研究の中で発生したデータなどは関連度が多少低くても上げることができますが、全く論文にひも付かないままデータを置く場所としては使えません。

●朝岡 ありがとうございます。

次に三上様へ、「学内で研究データの取り扱いについて検討したり、実装したりする組織が他に存在する

のか」という質問を頂いています。例えば研究支援担当の中で情報システム担当、研究者と図書館だけではなく、他に組織で連携しているかどうかについてコメントを頂けると幸いです。

●三上 GakuNin RDM の導入のときには情報基盤センターと連携して機関のストレージを入れたりしています。ただ、公開については今のところ他部署とはあまり連携していません。一応、URA が割と科研費などの申請の面倒を見ているので、国内で研究データの公開が助成に必須の場合が多くなると、そこでの連携の芽はあるのではないかと個人的には思いますが、現時点では特段、どこと連携しているということはありません。

●朝岡 ありがとうございます。次に安原様にお伺いします。JPCOAR では、根拠データの公開においてリポジトリを運用する機関はどのような役割を担うべきだと議論されているのでしょうか。

●安原 正直なところ、JPCOAR 側ではそういった議論はあまり深まっていないというのが実状です。少し話がずれるのですが、私個人としては、以前、ドイツのハノーファー大学でインタビューした際に、ハノーファー大学では研究者の意見を取り入れて、ドイツ語で利用でき、ドイツ国内で保存される安全で信頼できる研究データの公開場所として、研究データのためだけの機関リポジトリが作られたという話を聞きました。リポジトリを運用する機関としては、そういう安全で信頼できる研究データの公開場所の提供という役割を担うものは重要だと感じました。

●朝岡 ありがとうございます。

次は、データ管理の見地から神谷様と込山様にお伺いします。研究データ管理は、先ほど言ったとおりに研究データ公開に密接に関連していて、幾らか研究データ管理と公開の中で連携が必要だと思います。まず

は神谷様、海外の事情を踏まえて、管理と公開の関係についてコメントを頂けると幸いです。

●神谷 先ほどからずっと話に上がっているように、公開するときにはメタデータをきちんと付けて、FAIR 原則にのっとるなどして、みんなで後で使えるようにしようというコンセンサスがあり、管理は、もちろんプロジェクトが走っている間はその研究者なり、研究者をサポートする人なりが行うことになると思います。公開のときにメタデータを付けて公開したデータを長期間公開できるようにリポジトリに上げておくところの連携をどうするかは、図書館なり他の機関なりが考えるべきことだと思っています。日本の場合だと GakuNin RDM を使って管理した後で、ではどうするのかというところが大変気になるのですが、その辺はどうなのですか。

●込山 NII の三つの基盤は発表の中でも紹介させていただきましたが、管理の基盤と公開の基盤は安全のために完全に切り分けています。もちろん公開の基盤、機関リポジトリで JAIRO Cloud には大勢の機関がいらっしやいますので、まずは WEKO3 を目標に GakuNin RDM からデータを送出できる仕組みを開発していく予定です。その際、国際的に使われている標準的なリポジトリの通信プロトコルを使い、国内外のリポジトリで使われているメタデータスキームに沿って、きちんとデータを送り出せるように設計しておけば、それ以外のリポジトリ等とも連携していけるのではないかと思います。

●朝岡 ありがとうございます。私も NII で公開の基盤をサポートしている立場からすると、管理と公開の関係は連携していかなければいけないものだと思います。特に今、私は機関リポジトリの開発に携わっていて、制限共有や、メタデータでもいろいろなタイプのものを自分でエディットできるような形で作成しているので、こちらも皆さまに使っていただけると幸い

です。

最後の論点に移ります。日本版 Evaluate your RDM Offering や、日本版 COAR コミュニティフレームワークが実現できるかどうかについてディスカッションしたいと思います。研究データ管理をする機関にとって、自分のサービスの長期的目標を立てていく上で重要になると思いますが、パネリストの皆さまはどのようにお考えでしょうか。まずは管理の方の日本版 Evaluate your RDM Offering の実現可能性について、まずは神谷様、お願いいたします。

●神谷 日本の方々の需要があればもちろんできますけれども、需要がなければ要らないのではないのかと思います。その辺は本当に皆さんと一緒に話し合って決めるべきことではないでしょうか。

●朝岡 海外では需要がやはりそれだけ高くて、管理する大学機関などがきちんと使うことが浸透しているという感じなのでしょう。

●神谷 これは私の主観的な印象ですが、みんな手探りで分からない状況なのだと思います。その中でも、エディンバラ大学など特に盛んなところが例を出していて、それを見て参考にしているところもあったりします。ヨーロッパの中で、例えばいろいろなフレームワークが共有されていて、ある程度共有されているベースがあるから、ではうちもこういう感じにしていこうというような、共有感覚のようなものがあるのではないのでしょうか。こういうところから、ある程度経験や価値観を共有していくようなものが欲しいという感覚が生まれているような気はします。

●朝岡 確かにそうですね。まずはやる気がある機関などで、お互いに高め合うというよりも、お互いにそのように目標を持っていこうとすることでそのようなものができていると思います。共通するフレームワークという意味であれば、GakuNin RDM は使う機関が

幾らかあるので、そういった共有はしやすいと思うのですが、込山様、この辺についてどのように思いますか。

●**込山** GakuNin RDM が適用できるのはどうしても IT インフラなどの項目になってくると思いますので、そこは一つ埋まるとして、その周りをいかに埋めていくか。他の項目等も埋めていかなければいけないので、そういった意味で実際に動作している RDM サービスとして使いながら、組織の中での RDM サービスのあり方を検討していただくことはできるのではないかと思います。それを題材にして、牽引力のある学術機関、あるいはワーキンググループやタスクフォースの中でリーダーシップを発揮していただく機関やキーパーソンがいた方が、日本の場合はうまく、機関への RDM サービスの導入が進んでいくのかなと考えていました。

完成した万能な RDM サービスのモデルとなるユースケースが一機関から出てくることは、恐らくないのだろうとは思っています。それぞれの機関で最適な RDM サービスのベストプラクティスが生まれてくるのだと考えます。

●**朝岡** ありがとうございます。

次に、三上様と加藤様に、日本版の COAR コミュニティフレームワークの実現可能性についてお伺いします。研究データの管理・公開を行っている図書館員の立場から意見を頂けると幸いです。

●**三上** 大本の COAR フレームワークを眺めた個人的な感想ですが、結構、機関リポジトリの使っているシステムに依存するところが多いと思いました。恐らく日本の大学は JAIRO Cloud みたいに図書館内で管理していない場合もあるので、システムと人員を分けて考えるのはありかなと思いました。

インフラの JAIRO Cloud は恐らく共通のフレームワークだと思うので、判断基準や、大学が違っててもシステムが同じであれば、とここに注目して確認すればいい

というベストプラクティス的なものを共有できるかもしれません。一方で人員の話については、なかなか機関内での組織体制や相談窓口を明らかにするのは難しいかもしれません。ただ、国内で機関リポジトリを持っている大学が結構あります。北海道大学も機関リポジトリの論文の業務から伸ばす形でいろいろと実践していったので、国内は、今あるものから伸ばしていく形で考えるのが向いているのではないかと、個人的な感想ですが思いました。

●**加藤** われわれだけでベストプラクティスを探って広げるのには限界がございます。重要なことではあるけれども十分にはできていないところですので、日本版 COAR コミュニティフレームワークに期待し、関わっていただけたらと思っています。

●**朝岡** ありがとうございます。最後に、COAR コミュニティフレームワークを紹介していただいた安原様、JPCOAR では COAR コミュニティフレームワークをこれからどのように広げていくのか、国内機関とどのように関わっていくのかということについてコメントを頂けると幸いです。

●**安原** 先ほども申し上げましたけれども、まだ検討段階ではありますが、チェックリストを作成しているという意見は JPCOAR 内でも出ています。この COAR コミュニティフレームワーク自体も、今後毎年グッドプラクティスを採用してどんどん更新していく予定になっています。ですので、JPCOAR でチェックリストを作ったとしたら、それが完成形というよりも、実際に皆さまに使っていただいて、さらにより良いものにしていく必要があると考えています。また、先ほど紹介した国内機関の取り組み状況調査についてのアンケートで「研究データ管理の取り組みを行っている」と回答した機関に、可能であれば取り組み状況をより詳しく伺って共有させていただくことも考えています。そういう形で、実際に行っているグッドプラクティス

を共有していける形にしていきたいと考えています。

●朝岡 ありがとうございます。

林さん、皆さまの意見を聞いて、いかがですか。

●林 まさにこの SPARC Japan セミナーが、グッドな、またバッドなプラクティスを共有する場だと思いますので、ぜひ今回ご紹介いただいたコミュニティフレームワーク等々を実際に回してみて、そのご感想などを発表していただけると嬉しいです。

●朝岡 ありがとうございます。

今回、聞いてみて思ったのは、やはりまずは意欲のある人が始めてみるということです。始める中で、機関を超えて、ステークホルダーを超えて、団結してお互いの価値観を合わせられるようなフレームワークなどがつくれると、公開や管理が進んでいくのではないかと思います。

これでパネルディスカッションを終了します。どうもありがとうございました。