

SPARC*Japan NewsLetter

ISSN 1883-826X

NO.43 2021年3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
SCOAP³ for Books パイロットプロジェクト
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/>



SCOAP³ for Books パイロットプロジェクトについて

SCOAP³は、高エネルギー物理学分野の査読付きジャーナル論文のオープンアクセス化を実現することを目的とした、国際連携プロジェクトです。2014年以来、同分野で世界をリードする11のジャーナルで35,000以上の研究論文をオープンアクセスで公開することに貢献してきました。

この実績を受けて、2019年5月のSCOAP³ Governing Council（以下、GC）において、SCOAP³イニシアチブの範囲を従来のジャーナルから書籍に拡大することが提案され、2019年10月のGCにおいて大学レベルの教科書等の関連書籍78タイトルについて、SCOAP³ for Books パイロットプロジェクトとして進めることが承認されました。

国内のSCOAP³事務局を務める国立情報学研究所は、2020年10月にGCからの依頼を受けて、2020年度SCOAP³プロジェクト参加機関に向けて本プロジェクトへの寄付のお願いについて情報共有をいたしました。参加意向に係る調査へのご回答は、2021年3月26日をもって締め切らせていただきましたが、本プロジェクトの今後の動向を注視し、SCOAP³プロジェクト参加機関に向けて引き続き情報を共有してまいります。

<https://www.nii.ac.jp/sparc/about/international/scoap3/survey2021b.html>

<https://scoap3.org/scoap3-for-books-pilot-project-started/>

■ SPARC Japan セミナー報告



第3回 SPARC Japan セミナー2019

「実践 研究データ管理」

2020年2月7日（金） 国立情報学研究所 12F 会議室 参加者：66名

今回は、「研究データ管理」をテーマとし、研究データ管理を日々実践している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ツールの紹介、研究データ管理を担う人材育成の取り組みに関する発表を行いました。また、パネルディスカッションでは研究データ管理における研究者・研究支援職員・図書館職員の役割、役割を超えてすべての関係者が共有すべき理念やリテラシーとは何かについて議論しました。

次ページ以降に、当日参加者のコメント（抜粋）、企画後記およびドキュメント全文（再掲）を掲載しています。その他の情報は SPARC Japan の Web サイトをご覧ください。

<https://www.nii.ac.jp/sparc/event/2019/20200207.html>

企画概要



オープンサイエンスを実現するための基盤として、保管や公開を中心とする研究データの適切な管理は必要不可欠になっています。しかしながら、研究データ管理の必要性が強調される一方で、研究データ管理を「どのように」行うべきなのかに関する情報は極めて乏しいと言わざるを得ません。

こうした現状においては、データの主たる提供者としての研究者だけでなく、データ管理の実務を担うことを期待されている大学等の研究支援職員や図書館職員も「研究データを管理しろと言われても、明日から何をすればよいかわからない」と大いに戸惑っているのではないのでしょうか。

本セミナーでは、研究データ管理を日々実践している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ツールの紹介、研究データ管理を担う人材育成の取り組みに関する発表を行います。また、パネルディスカッションでは研究データ管理における研究者・研究支援職員・図書館職員の役割、役割を超えてすべての関係者が共有すべき理念やリテラシーとは何かについて議論したいと思います。

参加者の皆様が、明日からできることをそれぞれイメージできるようになっていたら幸いです。



パネルディスカッション（左から林氏、山田氏、熊崎氏、込山氏、結城氏、八塚氏）

参加者から

(大学/図書館関係)

・今まで中々複数の視点からの情報提供をいただく機会がなかったため、研究データ管理の知識を総合的に学び返す良い機会になった。既にポリシー策定に向けて当大学でも動いているが、より具体的な道筋が見えたように思う。

・研究データの現状と今後の展開について知ることができて良かった。

・組織へ関連する準備や業務をどう割り振っていかうと構想し、かつ、それらを構築するための準備を進めたのか、今回、原子力機構の熊崎様、NIIの込山先生からの発表はとかくタテ割り部局になりがちな大学や研究所での進め方に有効なモデルロールに感じました。

(大学/その他)

・本テーマに関する現状を短時間で知ることができました。本日の資料をもとに学んでいきたいと思います。

・研究データ管理の必要性は理解しているが、その手法が明瞭ではない大学が多いので、大学の取り組み事例がもう少し充実していれば良かったと思う。

(その他/図書館関係)

・他機関での実践例や、GakuNin RDMの解説が役に立った。理系の研究データ管理に対する考え方を、人文系の図書館でどう応用・実践していくかが課題であると感じた。

企画後記



😊 今回のセミナーは、単なる理想論を語る場にしたくないという思いから企画しました。研究機関・図書館・データリポジトリそれぞれにおける研究データ管理の取り組みはまだ不十分かもしれません。しかし、そうした現状を受け止め、「自分ごと」として一歩でも前進することの重要性を多くの方にご理解いただけたように思います。

八塚 茂

(バイオサイエンスデータベースセンター)

😊 「研究データを管理しろと言われても、明日から何をすればよいのかわからない」という現状からちょっと前に進むため、すでに取り組まれている例を共有すべく今回のセミナーを企画いたしました。研究データを適切に扱うためには、研究者・研究支援職員・図書館職員といったそれぞれの役割でなすべきことがあり、(研究データに限りませんが)互いに信頼できるパートナーとして連携する必要があることが再確認できたように思います。

林 賢紀

(国際農林水産業研究センター)

😊 今回初めて企画立案から関わらせていただきました。印象的だったのは「研究データ管理についてステークホルダーで共通認識をもって議論することが難しい」という点です。組織・立場・領域によって研究データ管理を語る文脈が異なるため、合意に時間がかかるかもしれません。今回のセミナーで紹介された事例等が、そうした合意を築くような「実践」の一助になれば幸いです。

西村 恭佑

(富山大学)

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3回 SPARC Japan セミナー2019

「実践 研究データ管理」

開会挨拶 / 概要説明

林 賢紀

(国際農林水産業研究センター)



林 賢紀

国立研究開発法人国際農林水産業研究センター企画連携部情報広報室情報管理科。2014年4月より現職。博士（情報学）。図書館や公式Webサイトの運営管理、データベースの構築支援など、研究情報の収集及び発信に携わっている。



当セミナーの企画概要

今日のテーマは「実践 研究データ管理」です。オープンサイエンスを実現するための基盤として、研究データの保管や公開を中心に適切な管理が必要不可欠です。しかし、いろいろなところから研究データを保管しなさいといわれ、その必要性は分かっても、具体的にどのように行っていくのかを知る機会は少なく、情報もあまりないところかと思います。研究者が研究を通じて生み出したデータを提供するわけですが、一方で、その管理の実務を誰がするのか、例えば大学であれば研究支援職員や、またこれからの図書館員の役割の中に研究データ管理が入っていくといわれています。しかし、私もそうですが、正直、「明日から研究データの管理をしよう」といわれても、そもそも研究データには何があって、何を管理して、どうやって目録を取ればいいのかも分からないというところです。

今日のセミナーでは、研究データの管理を日々実践しているリポジトリや図書館の実例、あるいは、これからポリシーを考えなくてはいけないので、どうやって扱うかを考えてみたという事例を発表していただき

ます。また、研究データ管理を支援するためのツールは、国立情報学研究所（NII）でも開発が進んでいます。こういうツールを使うことで、より簡便に、より手軽にできるという話題や、われわれは何を学習して、何を進めていけばいいのか、どういう人がいれば研究データの管理が回っていくのかという人材育成の取り組みに関しても発表いただく予定です。

その後、パネルディスカッションにおいて、研究データ管理に係る研究者、研究支援の職員、図書館の役割、あるいはその役割を超えたところにある共有すべきものは一体何かというような理念やリテラシーを議論できればと考えています。

今日は研究者や大学の職員、あるいは図書館員などいろいろな方がお集まりです。皆さまが、今日のセミナーが終わったときに、「じゃあ、こんなことができるのではないか」「自分はこれができるのではないかとイメージできるようなセミナーを予定しています。

第 3 回 SPARC Japan セミナー2019

「実践 研究データ管理」

名古屋大学における 研究データ基盤整備推進組織の整備について

竹谷 喜美江

(名古屋大学附属図書館事務部 / 情報連携統括本部情報推進部)

講演要旨



2019 年 12 月、名古屋大学における研究データに関する取り組みを推進することを目的として、名古屋大学研究戦略・社会連携推進分科会のもとに研究データ基盤整備部会が設置されました。同部会は、学内の情報施策を担う情報連携統括本部の部長を主査としつつ、情報連携統括本部、情報基盤センター、附属図書館、研究協力部、学術研究・産学官連携推進本部、IR 本部の、各担当理事、各部署の責任者および担当者を構成員としています。当部会の設置までの経緯や運営について報告するとともに、事務局としてどのような活動を行ったか紹介します。



竹谷 喜美江

平成元年東京大学附属図書館中央図書館に就任後、医学図書館、農学部図書館、国立教育研究所教育図書館を経て、筑波大学附属図書館、一橋大学附属図書館、文部科学省官房政策課資料係、国立国会図書館支部図書館協力課システム係を経験。その後、新潟大学附属図書館情報サービス課長、金沢大学附属図書館情報企画課長、名古屋大学附属図書館情報管理課長、同附属図書館事務部次長を経て、現職に至る。

自己紹介

名古屋大学は 2020 年 4 月から岐阜大学と法人統合することになっており、現在、ラストスパートの 2 カ月で事務的にも大変な状況です。しかも、先日次期スーパーコンピュータ導入が決定しまして、今週初めの記者発表等に追われました。「富岳」と同等の CPU を持ち、今の 10 倍のスピードになります。名古屋大学の所在地である「不老町」とデータの「フロー」を掛け合わせて「不老」という名前が付けられました。そういう事情で、本日登壇するはずだった情報推進部推進課の古島唯が来られなくなり、ピンチヒッターとして部長の私が、名古屋大学の取り組みはどういうものかということ、古島とは少し違う視点になるかと思

いますが、報告したいと思います。

登壇予定だった古島は、元々、東京大学の図書館職員として東京大学総合図書館を 3 年と医科学研究所の図書室を 2 年経験した後、名古屋大学に出向し、経験を積んでいます。最初の 2 年間は、附属図書館でサービス関連業務を担当し、2019 年 4 月からは情報推進部に異動して、事務系職員として筆頭である総務系の業務を担当しています。

私も古島と同様、東京大学に 1989 年に入職しました。その後の職歴は資料の通りですが、主に東京大学と筑波大学と名古屋大学に 6~7 年お世話になっています。大学図書館以外の経験としては、国立教育研究所の図書館や、文部科学省の官房政策課では、図書館

業務以外に、いろいろな評価の仕組みなど政府の仕事も知ることができました。国会図書館でも支部図書館・協力課という省庁全体の図書室の面倒を見る課をはじめ、いろいろな部署を経験させていただいたことはどれも非常に大切な経験になっています。現在は名古屋大学で図書館事務部と情報推進部の部長として励んでいます。

研究データ基盤整備部会設立までの背景（１）

研究データマネジメント・プロジェクト

今日は名古屋大学の研究データの基盤整備について、その中身よりもプロジェクトをどのように進めているかということをご紹介しますと思います。名古屋大学では、情報連携統括本部が研究データに関しては古くから積極的に動いています。2018 年 5 月に情報連携統括本部のプロジェクト専門委員会の中で、研究データをマネジメントするプロジェクトが組織され、その担当教員を中心に、技術職員、図書館職員、URA が加わって研究データに関する検討を進めていました。

主な活動としては、GakuNin RDM の実証実験への参加、補助金によるクラウドストレージの試験的導入、そして、2019 年 4～5 月に実施した、研究データ管理に関するアンケート実施などになります。現在、大学 ICT 推進協議会 (AXIES) で、全国にアンケートを行う予定ですが、名古屋大学のアンケートを土台にしていただく予定です。また、全学への周知目的の公開講演会もこの 2 年間で 4 回企画しています。

大学 ICT 推進協議会 (AXIES) は、図書館の方々に
は馴染みが薄いと思いますが、2011 年に設立され、

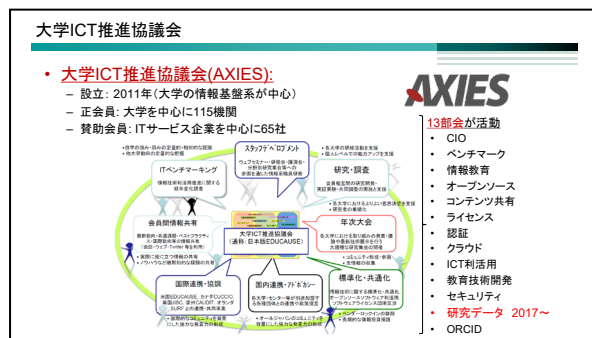
国立大学、私立大学を含めて 115 機関が入っています（図 1）。大学の研究情報をどのように活用していくかという取り組みで、13 部会が立ち上がっており、それぞれ 115 機関の教員や事務職員が集まって検討しています。研究データマネジメント部会は 2017 年に発足しています。

また、AXIES では、「学術機関における研究データ管理に関する提言」（以下、「提言」という）を出しています。冊子体でも出ていますし、ホームページからダウンロードすることもできるので、ぜひ図書館の人や、情報の関係ではない方も見ていただくと、データ管理に関して何が問題なのかということが分かってくるかと思います。

図2は部会の会員一覧です。京都大学の青木学聡先生が主査で、副査であるNIIの船守美穂先生の下に名古屋大学の松原茂樹教授が、そして部会員として外山勝彦教授がいます。この2人の先生が主に名古屋大学の研究データに対する中心になっています。その他のメンバーにも、名古屋大学と関連の深い方が何人もいます。

提言には、学術機関における目的・効果がⅠ、Ⅱ、Ⅲ章に書かれ、方法・機能・維持管理がⅣ、Ⅴ、Ⅵ、Ⅶ章に書かれています。今日は、Ⅴの「学術機関における研究データ管理のための仕組み」を名古屋大学がどうつくったかというところをご紹介します。

図3が提言の中のV章です。研究データ管理のための仕組みとして「これだけはやらなくてはならない」ことがバイブルのように書いてあります。例えばaの「研究データ管理の機関内関係部署の明確化と仕組み



(图 1)

大学 ICT 推進協議会 研究データマネジメント部会		
部会員一覧		
部会員33名、2017設置		
部員理事 松原 玲二 (広島大学) 部会主席 青木 孝樹 (京都大学) 部会幹事 松平 美穂 (国立情報学研究所) 松原 茂樹 (名古屋大学)	部員 (フューチャースタレーション) 藤田 雅夫 (国立京大府)	
部会員 (順不同) 前田 司司 (京都大学) 野田 隆司 (京都大学) 金ヶ野 雅樹 (岡山大学) 尾崎 孝一 (国立情報学研究所) 長野 昌弘 (国立情報学研究所) 山田 昌孝 (京都府立大学) 山崎 一誠 (国立情報学研究所) 松本 正浩 (国立情報学研究所) 桐田 洋 (筑波大学) 土井 雅彦 (豊橋技術科学大学) 辻井 直浩 (奈良先端科学技術大学院大学)	西村 浩二 (広島大学) 尾方 淳 (筑波大学) 田中 孝 (筑波大学) 松平 元也 (金沢大学) 南田 昌孝 (金沢大学) 水本 隆夫 (京都府立大学) <u>高田 直幸 (名古屋大学)</u> 宮本 幸雄 (日立製作所) 荒木 孝 (伊藤忠テクノソリューションズ)	

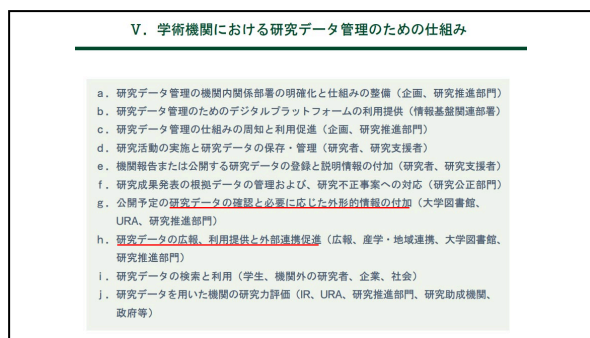
(图 2)

の整備」は、企画部や研究推進部門が、b の「研究データ管理のためのデジタルプラットフォームの利用提供」は情報基盤関連部署が、d の「研究活動の実施と研究データの保存・管理」は、研究者と研究支援者が担当しなければなりません。大学図書館は g と h で、研究データをしっかり管理して外形的に情報を深めたデータを振ることと、研究データを公開して広報し、その利用提供をしなくてはなりません。そして、これらの研究データの利用をするのは i の学生や研究者、社会であり、研究データを用いた機関の研究力を評価するのは、j の IR、URA、研究推進部門や企画部門です。このように研究データ管理には複数の担当部門がそれぞれの役割分担をする必要があると書かれています。この提言に基づき、名古屋大学でも各部署に役割を落としていくことになりました。

ところが、複数の部署（ステークホルダー）が共通の方向性を確認し、各々が主導的な役割を果たすような体制づくりは、簡単ではありません。一部署である情報連携統括本部が、隣の部署に声をかけて会合を開いても、なかなかうまく連携に至ることはできません。どうやってうまく機能して進めていくかということが大きな課題でした（図 4）。

研究データ基盤整備部会設立までの背景（2） 研究データに関する関連部署による打ち合わせ

研究データに関する関連部署による打ち合わせが開始されるきっかけは、2019 年 7 月 26 日のことでした。内閣府総合科学技術・イノベーション会議（CSTI）の委員である本学の松尾総長から、当該会議での研究



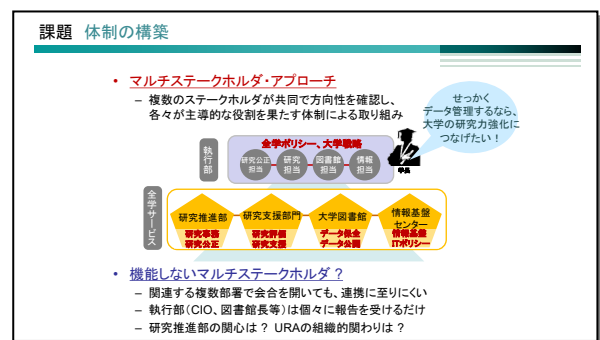
(図 3)

データに係る国の動向を受けて、名古屋大学における研究基盤の整備にかかる課題の抽出をするようにとのミッションが出されたのです。

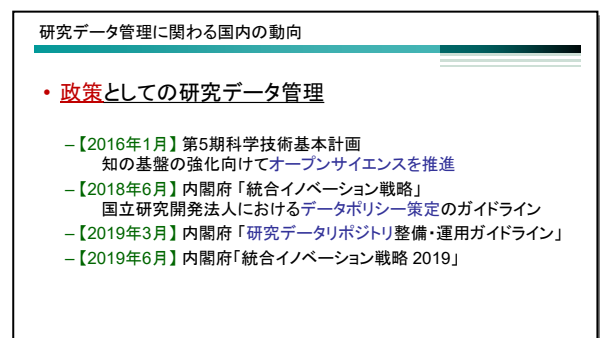
総長から集められたのは、企画部、研究支援部、附属図書館と情報連携統括本部のそれぞれの担当理事、部局長と、部課長でした。意見交換をした結果、とりまとめは、これまでの研究データの整備に関する検討を進めていた、松原先生などを中心とする情報連携統括本部が担当することになりました。

研究データ管理に関わる国内の動向としては、図 5 の通りです。2016 年 1 月に「第 5 期科学技術基本計画」で既にオープンサイエンスを推進するという流れができています。それから 2 年たって、内閣府の CSTI でデータポリシー策定のガイドラインが出され、まず国研から始めようということになりました。

「第 5 期科学技術基本計画」では、具体的に次の二つが大事であると思います。まず、「公的資金による研究成果は、その利活用を可能な限り拡大することをオープンサイエンスの基本姿勢とすること」です。公的資金で研究したものは基本的に全て公開しましょう、



(図 4)



(図 5)

ということです。

それから、「研究分野によって研究データの保存と共有の方法に違いがあることを認識するとともに、国益を意識したオープン・アンド・クローズ戦略の実施に留意」することです。これは、受益者に相応の負担を求めることができそうなデータは全部オープンにしないでいいということです。どれをオープンにして、どれをクローズにするかという見極めはポリシーの中に盛り込まなくてはなりません。

研究データ基盤整備部会設立までの背景（3）

研究データ基盤整備 WG（時限的設置）

総長からのミッションを受けて、研究データ基盤整備 WG を時限的に設置しようということになりました。情報連携統括本部長である担当理事と、センター長である副本部長が中心となり、誰を集めるかを決めました。情報連携統括本部からは教授 2 名（松原、外山）と部課長に、附属図書館からは館長他教授 1 名と部課長に、研究協力部からは課長に、産学官連携推進本部からは URA に入ってもらいました。そして、情報連携統括本部の研究データプロジェクトの方々には、コアメンバーとして入ってもらい、最終的に図 6 のようなメンバーになりました。

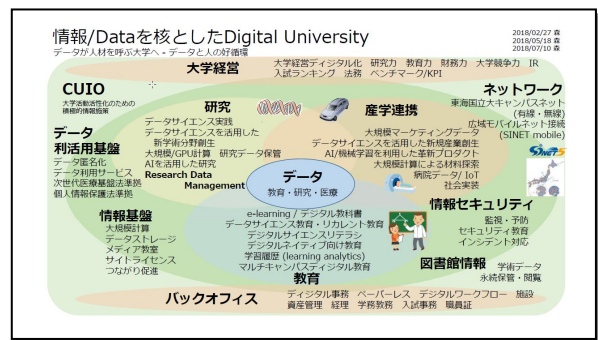
総長からは夏休み明けには報告をと指示されており、9 月の 1 カ月間で 2 回の会合を開催して、「研究データ基盤の整備に関する課題整理とスケジュール案」を提出しました。

図 7 は名古屋大学の森センター長が作成した図です。大学ではデータが研究と産連と教育の三つに使われて

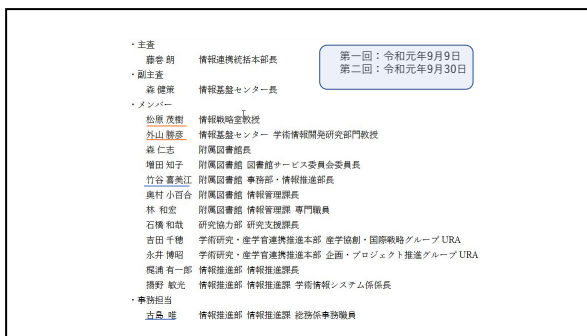
います。データ利活用基盤や情報基盤、バックオフィス、図書館情報、ネットワーク、大学経営まで、データは大学の根幹としていろいろなことで使われているということを整理した図です。

名古屋大学のデータ利活用に関する将来構想としては、図 8 のように、たくさんの大きなデータに対応する研究データ基盤としての認証・ストレージをしっかりと整え、そこに記録・保存をして、それらのデータをオープンにすること、そして、それらのデータを再利用することです。また、個々の研究者は今後、研究費の申請の際に「研究データの管理計画」の策定が求められることになります。その作成支援も必要です。これらのことを円滑に進めるために、研究データ管理ポリシーを規定化したり、権利・インセンティブを整理したりする必要があります。

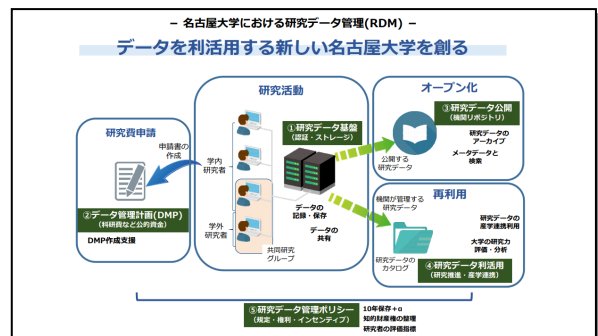
では、誰が何をやるのか。ここをきちんと明確にすることが肝心です。図 9 の表は、名古屋大学の研究データの要素とその担当を整理したものです。①研究データ管理基盤の導入と運用、②データ管理計画（DMP）を作成する研究者の支援、③研究データの



（図 7）



（図 6）



（図 8）

アーカイブと公開・発信、④研究データの利活用促進、⑤研究データ管理のポリシー策定に、先ほどの提言の中にあった、やらなくてはならないことを全て落とし込んで、担当部署を割り振りました。

図 10 はスケジュールです。本格運用は 2021 年を目指します。あとわずかです。2019 年度は終わりますが、国の動きに合わせて、2020 年から試験運用に入るという結構タイトなスケジュールです。

このように、情報連携統括本部が人を集めて WG を作り課題の洗い出しとスケジュールを作成しましたが、今後は、大学の基幹会議である教育研究評議会の下での研究戦略・社会連携推進分科会で全学的に課題を整理していくべきであると、総長に報告をしました。

その結果、教育研究評議会の下に、研究データ基盤整備のための分科会を組織することになりました。分科会には、各部署から承認された構成員が参画して、定期的に会議が開催されることになりました。

研究データ基盤整備部会設立までの背景（4）

研究データ基盤整備部会は、当初の目論見通り、研究戦略・社会連携推進分科会の下に、2019 年 11 月 26 日に設置されました。その狙いは、全学に関わる基幹的な会議の下で組織を作ることと、関連するマルチステークホルダーのトップの参画です。参画いただいた部署の方々には、それぞれの理事の下でイニシアチブを持って積極的に課題に取り組んでいただくようお願いしました。

情報連携統括本部の松原先生と外山先生には、引き続きこの分科会をリードしていただくことになりまし

た。構成員は、理事が 5 人、URA が 4 人、教員が 4、5 人と、部課長を中心とした事務職員になりました。

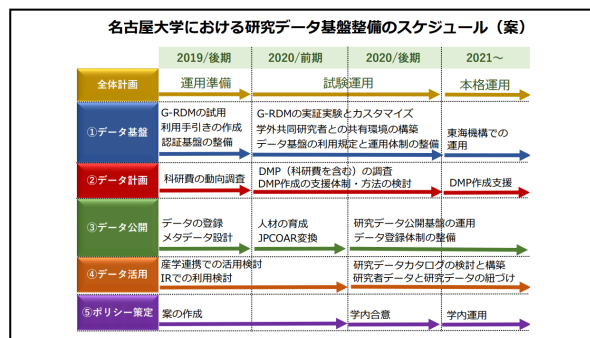
多忙な理事が 5 名もいるため、この整備部会は日程調整が難しく、第 1 回目の会議のセットができたのが約 1 カ月後の 12 月 13 日でした。そこで、会議を円滑に進めるために、その間の 1 カ月を利用して、各部署を 1 カ所ずつ巡って事前に意見交換をして課題等の洗い出しをしておこうということになりました。

まず附属図書館を訪ねました。図書館は、リポジトリを使って研究データの公開に関する支援を進めていきたいと考えていました。ただ、研究データへのメタデータ付与は未知の領域であることが課題とされました。図書館職員は、図書のメタデータ付与についてはプロだけれどもデータのそれについては全く初めてで、人材育成が必要であるということでした。また、デジタルアーカイブの管理についてもやっていきたいとのことでした。

続いて研究協力部を訪ねました。研究協力部は、データ管理計画の作成を周知したり、説明会を開いたりすることはできるので、協力したい、また日本学術振興会（JSPS）をはじめとしたファンドの動向を見て予算獲得が可能な情報があったら提供したいということでした。課題は何かあるかと尋ねると、公開した研究データを勝手に利用されてしまうことに抵抗がある研究者がいるだろうということ、また、登録の手間がかかるとそれが障壁になり研究者はやろうとしないだろうということ、そしてそのような課題を払拭するためには、まずは共通のポリシーのような枠組みをしっかりとつくっていくということが大切であるという結論が

研究データ管理の要素	情報連携統括本部	附属図書館	研究協力部	理工学部	医学部	農学部	経済学部	法学部	文学部	教育学部	国際学部	グローバル教育センター	研究推進部	研究戦略・社会連携推進分科会	研究データ基盤整備部会	研究データ管理の要素
① 研究データ管理の導入・運用	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	研究データの管理・運用
② データ基盤	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	研究データの基盤整備
③ データ計画	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	研究データの計画・設計
④ データ公開	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	研究データの公開・発信
⑤ データ活用	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	研究データの活用・利活用
⑥ ポリシー策定	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	研究データのポリシー策定

(図 9)



(図 10)

共通理解されました。

次に訪れたのは、学術研究・産学官連携本部です。知財・技術移転グループからは、対象が著作物ならば担当として検討するけれども、研究データはライセンスの対象とはしない、とクールにかわされました。産連本部の担当者からは、海外研究機関では、研究データのデータベース化やビジネスにつなげる展開も多く存在しているとのコメントをいただき、方向性に賛同いただきました。今後、論文、特許を含めて価値があるものについては、学外に対して積極的に売っていく部隊が必要だとのコメントもいただきました。

最後は、IR 本部を訪ねました。こちらは、これまでは各部署がそれぞれの目的に応じてシステム設計をしてきたため、システム間の連携がとれていないので、全体を統括した設計が必要であるということで、ぜひとも連携していきたいということでした。今後教員が作成した研究データのデータベースを保存公開していくなら、それを IR のためにもぜひ使いたい。そのためには、IR システムに連携するためのキーが必要である。例えば、ORCID の取得を義務付けするなど、何か共通となるキーを使って IR が持っているデータベースとのひも付けをすることが必要不可欠だということ、かなり具体的な意見が出されました。

第1回の部会は、これらの各部署にリサーチした内容をもとにして意見交換を行い、データポリシーを策定することと、研究データを保存公開する上での課題をまとめるという二つのことが決定されました。

今日は附属図書館の方が多いということで、図書館職員の方に向けて附属図書館が関係する検討課題を少しリサーチして参りました（図 11）。この資料は附属図書館情報管理課の林専門職員が提供してくれたものです。

図書館のリポジトリ担当者は、一言で研究データと言っても手書きのメモからデジタルデータ、果ては標本まで様々な種類が想定されるため、どこまで受け付けてどう整理したら良いのかということに非常に悩んでいます。しかし、既に研究者からは、論文を投稿しようとしたら、根拠となる研究データをきちんと公開していないと受け付けないといわれたので、登録を頼みたいという相談が何件か来ており、対応に苦慮しつつも何件かはリポジトリにデータを登録して公開しているとのことでした。公開されているのは、大気海洋科学、大気科学、宇宙科学などの分野のものです。

図書館で考えられる課題としては、ストレージ、人員、ランニングコスト、大容量のファイルの処理方法、ファイルの受け渡し手段、などがあります（図 12、図 13）。

★ 附属図書館が関係する検討課題（4/5）

どうやってリポジトリに載せるか？
研究者、投稿先、助成機関が望む形で載せられるか？

<具体例から考える検討事項>

- ・公開希望が増えた場合、ストレージや人員がどこまで対応可能か
- ・ストレージが拡張可能な場合でも、ランニングコストは確実に膨らむ。
- ・公開対象が広い場合、どこまで対応できるか
- ・1件のアイテムに、数GBのファイルや数百個のファイルを登録してよいものか

（図 12）

附属図書館が関係する検討課題

★ 附属図書館が関係する検討課題（1/5）

● 研究データ公開基盤の仕様

- ・公開システム → 次期JAIROCloudを公開基盤として想定
- ・メタデータ → JPCOARスキーマによるデータ連携 } DataCite対応
- ・識別子付与 → DOI
- ・著者名典拠 → 現在は、研究成果を登録した者のみ、内部IDで名寄せ

● 研究データ公開のための体制整備

- ・ポリシー策定 → 先ずは目的・対象。図書館は特に公開についてオープンアクセスポリシー、リポジトリ登録要項とのすり合わせも必要
- ・登録の作業フロー
- ・作業量に応じた体制
- ・ランニングコスト

（図 11）

★ 附属図書館が関係する検討課題（5/5）

どうやってリポジトリに載せるか？
研究者、投稿先、助成機関が望む形で載せられるか？

<具体例から考える検討事項>

- ・ファイルの受け渡し手段
- ・公開時、全てまとめて圧縮してもよいものか
- ・利用のためのプログラムコード、マニュアルも併せて提供が必要
- ・ドメイン・メタデータは、研究者の協力が必須
- ・DOI付与やDataCite連携以外に求められる、条件はあるか
- ・手続き申請やデータ受取のワークフローはどうするか
- ・論文投稿や助成で求められるデータ公開の場合、公開のタイミングは？

（図 13）

第2回部会

第2回部会は1月20日に行いました。第1回目の部会で研究データポリシーを策定することは決まっていたので、第2回部会では、策定スケジュールと今年度末までにまずデータポリシー案をつくることを目標としました。時間もあまりないので、データポリシーのたたき台を、情報連携統括本部で、国立研究所が策定し公開しているポリシーを参考にして作成することになりました。

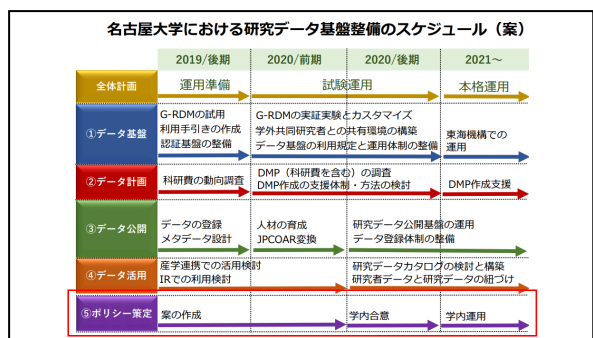
今後の予定

今後は、データポリシーの基本方針案が形になるまで詰めていきます。今はお見せできる状況ではありませんが、3月には案を完成させる予定です。

今月（2月）の部会では国立情報学研究所のオープンサイエンス基盤研究センター長である山地教授にレクチャーをしていただきます。研究データがどういうもので、どのように扱うべきか、ということ具体的にレクチャーしていただき、名古屋大学全体の研究基盤を整理する機運を盛り上げたいと考えています。

最後に再掲しますが、本学のポリシー策定のスケジュールは、図14のようになっています。2020年度の終わりには「案」が取れてポリシーを策定したいと考えています。

今日ご紹介したパワーポイントのうちの何枚か、特に整った美しい資料ですが、情報連携統括本部の松原教授からご提供いただいたものです。その他、情報推進部情報推進課の揚野係長、附属図書館情報管理課の林専門職員から、本資料作成に当たってご自身の資料



（図14）

や貴重な情報をご提供いただきました。この場を借りて御礼を申し上げます。

●フロア1 NIIの木下です。名古屋大学は岐阜大学と一つの法人になる予定だということですが、研究データ管理については岐阜大学と何か話し合いは進めているのですか。

●竹谷 岐阜大学では、研究データ管理をしっかり検討する基盤がまだ育っていないため、名古屋大学が先行して検討を進め、それに岐阜大学も合わせていく予定です。

●フロア2 高エネルギー加速器研究機構の職員です。最近の研究はどの分野も共同研究だったり、国際共著論文を出すことが推奨されていたりして、研究データも他機関に所属する複数の研究者が関わるものが多いかと思います。名古屋大学で作ったガイドラインやルールと他機関のルールが必ずしも一致するわけではないかと思いますが、その辺の調整はできるような仕組みになっているのでしょうか。

●竹谷 難しい問題だとは思いますが、共著者である方々の所属する機関ごとにルールが違えば、どこのリポジトリにデータを置くかは、著者が決めればよいのではないかと思います。名古屋大学に置くなら、名古屋大学のポリシーに従わなければいけないですが、名古屋大学のポリシーと合わなければ、共著者の所属機関にストレージを置いて公開することになるかと思っています。今のご意見は貴重なものでした。ポリシー策定の検討の際に参考にさせていただきたいと思います。

●フロア3 京都大学の職員です。研究データにはデジタル化されているもの以外にもありますが、それに

については今回は対象外ということなのでしょう。

すか。

●竹谷 これも難しい問題です。情報連携統括本部としてはデジタルデータだけと考えたいのですが、大学全体の目線から研究データを考えると、デジタルデータだけに限定することはできないでしょう。紙データや電子媒体ではないデータ、昆虫の標本なども含めて検討していく必要があると思います。

●フロア 4 オープンにするデータと、ある程度クローズにして大学のために役に立てるデータを取り分ける必要があるという話がありました。オープンにするかしないかの判断は重要であると思うのですが、誰がそれを判断するのか。大学が判断すればいいのか、公的な資金であれば政府がそこに関与できるのか、元を正せば国民の税金なので一般市民が入るべきなのか。ほとんどの基礎研究はオープンにして構わないのでしょうけれども、微妙なデータもきっと出てくると思います。その辺はどのように仕分けをしようとされているのか、突っ込んでいくとどんどん難しい例が思い付くのですが、いかがでしょうか。

●竹谷 名古屋大学のデータをどう管理するかです。で、基本的に名古屋大学が決めることではないかと思っています。個人的には、利用ニーズの高いものについては、公開のための原資を捻出するために、今まで投入してきた資金の一部を回収すると考えてはいかがだろうかと思っています。どのように仕分けをするかといったことについては今後、研究者とともにしっかり検討していく必要があると思います。

●フロア 5 琉球大学の職員です。上層部がどのように研究データ管理を進めようとしているのかは分かったのですが、現場の研究者の方々に理解していただけないといけないと思うのです。そのような取り組みは、アンケートをされたということは伺いましたが、その他にはどのぐらいの頻度でどういうことをされていま

●竹谷 現場の研究者の方々への周知はとても大事であると思います。私どもは今年度に入って、研究データに関する勉強会や研修会を3回学内で行っています。正確に申し上げると、学内だけではなく、広く学外にもPRして参加者を募りました。本学の教員は結構興味を持って参加してくれています。今後も、引き続きこのような企画は続けていかなければいけないと考えています。ありがとうございました。

第 3 回 SPARC Japan セミナー2019

「実践 研究データ管理」

糖鎖科学における研究データ管理

山田 一作

(野口研究所)

講演要旨



糖鎖科学コミュニティにおいて塩基配列やタンパク質のリポジトリは利用できたが糖鎖構造のリポジトリはなかった。2013年に中国・大連において開催された国際会議において、糖鎖科学研究において糖鎖構造の明確化について議論された。この会議の結果、糖鎖構造のリポジトリを構築し糖鎖構造に固有のアクセッション番号を付与することが合意された。そして2014年8月に国際糖鎖構造リポジトリ GlyTouCan が公開された。また、糖鎖科学では質量分析法により糖鎖構造を解析するが、その際に生成するデータも重要であり、これらのリポジトリとして UniCarb-DR および GlycoPOST が開発された。UniCarb-DR は質量分析データを解析したピークリストなどのリポジトリであり、GlycoPOST は、質量分析の実験法や生データのためのリポジトリである。本発表ではこれらのリポジトリの開発の経緯なども含めて紹介したい。



山田 一作

公益財団法人野口研究所インフォマティクスプロジェクトリーダー。1997年に東京都立大大学院工学研究科にて博士号取得。2002年から野口研究所研究員、2006年から同所にて糖鎖科学研究を開始し、糖鎖構造表記法、オントロジー、データベースなど糖鎖インフォマティクス研究に従事している。

自己紹介+α

今日は、糖鎖科学の領域の研究データをどうやって扱っているのかという糖鎖の領域のお話で、図書館の方たちと少し違うのですが、図書館の方や他の領域の方の意見を聞きながら、より良いものにしていきたいと考えています。

私は学生のときから有機化学を専攻していて、野口研究所に入って糖鎖科学を始めました。有機化学の中にも糖鎖構造はありますが、糖鎖科学は少し特殊というか、私としては取っ付きにくかった領域で、すごく難しいものだと考えていました。しかし、有機化学を学んでいたので化学構造には馴染みがあり、それを中心に糖鎖科学を見ていくと、難しいのですが実は面白

い、やりがいがあるものだと思います。結局今も何とかこの分野を良くしていきたいという思いで続いています。

難しいという一つには、化学が決して簡単なわけではないということもあるのですが、例えば化学だと炭素を表すのに元素記号で C と書けば間違いなく炭素だという認識があります(図1)。肩に 12 と書くと同位体表記があるという共通認識があります。それは世界共通だと思います。そういう世界でずっとやってきたのですが、では糖鎖科学ではどうなっているかというと、糖鎖の化学の人たちは元素記号を使って化学構造式を書いたりします。ただ、サイエンスとなると、化学以外にも生物や医学などさまざまな人が絡んでき

ます。そういう人たちが集まってグルコースといったときに、国際純正・応用化学連合（IUPAC）では Glc と書くと決まっていますが、これでいいの。駄目かといわれると「いいんじゃない？」という答えが大体返ってくるのですが、化学から見ると Glc の化学構造式はどのようなものなのかが気になります。

Glc はグルコースだと考えると、図 2 は欧州バイオインフォマティクス研究所（EBI）の ChEBI というデータベースですが、丸で囲ったところにグルコースがあります。化学構造式で書けるものに関しては四角の左側に化学構造が記載してあります。しかし、化学構造式で書けないもの、非常に書きにくいものに関しては構造の記載がありません。

グルコースといった場合に、非常に曖昧な部分が含まれていて、その部分を明確に表すことができません。グルコースには D 体と L 体という鏡像異性体があり

ます。グルコースといったときに、生物の方などは生体内に一般的にあるような D-glucose を思い浮かべますが、D-glucose も化学構造式が載っていません。それは、グルコースにも、環状になっていたり、直鎖状になっていたり、さまざまなパターンがあるからです。そういうところが明確に記載されていないと、さらにグルコースの細かいデータを付加していかないと情報が正確ではないことになります。そんなところが私としては非常に面白いです。

今日の内容は、最初に、糖鎖とはどのようなものなのか。その後に、糖鎖科学における研究データとはどのようなものか。そして、私自身が構造に興味を持っているということもあり、糖鎖の構造と、糖鎖科学における標準化・共通化の取り組み、そして、それらを使って糖鎖科学でデータを保管するリポジトリを作っているのも、その利用を含めてご紹介したいと思います。

糖鎖の機能

図 3 は糖鎖の機能の一例です。糖鎖は単体で機能することもあります。蛋白質や脂質と複合体を形成して機能することもあります。蛋白質や脂質は水溶性ではないものが多いのですが、そこに糖鎖という水溶性が高いものが付加すると溶解性が増して、その付き方によって機能が変わっていきます。また、蛋白質に糖鎖が付くのですが、その糖鎖の構造が変わることによって、蛋白質の品質管理を行っています。

糖鎖構造は面白く難しい

化学には元素記号がある

炭素 → C → ^{12}C

糖鎖科学では

グルコース → Glc ← これでいい？

Glc の化学構造式は？

(図 1)

Glc の化学構造は？

Glc は、グルコース？

“Glc” とは何？

(図 2)

糖鎖の機能

- 蛋白質や脂質の機能向上
 - 溶解性、細胞局在、蛋白質品質管理
- 細胞間情報伝達
 - 細胞の認識や接着、病原体の認識、がん増殖・転移などに関与
- 個の識別
 - ABO型の血液型

(図 3)

二つ目の機能は、細胞間の情報伝達に使われているということです。糖鎖は細胞の表面に非常に多く存在し、細胞間の認識や接着に関わっています。病原体の認識や、がんによって糖鎖の構造が分かるといったことがよく知られています。

三つ目の機能は個の識別です。糖の話をするときによく出てくるのが ABO の血液型の話です。この血液型は糖鎖の構造が異なることによって型が決まっているので、私から見ると糖鎖の構造が非常に重要だということになります。

糖鎖の分野における研究の手法を全て網羅することはなかなか難しいですが、例えば図4のようなものがあります。合成で糖鎖を作るとなると、化学合成や酵素合成という方法が行われています。あとは分析です。例えば糖鎖の構造を解析したり、各種の相互作用や機能を見ていく、そして糖鎖の局在などをイメージングするということが行われています。

糖鎖の研究では、どのような糖がどのような機能になっているのか、どういう相互作用があるのかということが重要になってきて、一つの鍵として糖鎖構造というものが考えられます。

糖鎖の構造

合成や分析などいろいろな研究があるのですが、糖鎖の構造とその構造を決めるための分析データが、糖鎖研究の中でも重要な位置にあります。それ以外にも非常に重要なのですが、今回は構造にフォーカスしたお

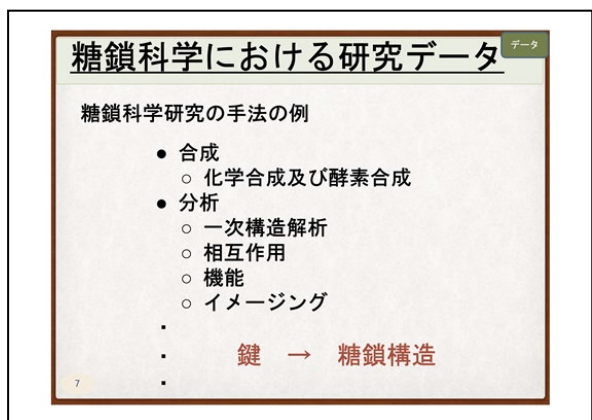
話をしたいと思います。

先ほど糖鎖の構造データということで、Glc（グルコース）の話をしました。その文字列の表記方法としては、図5のIUPACの下に長い文字が書いてありますが、このような表記法がよく使われます。それ以外に、この構造の糖鎖はこういう名前ですよという名称がいろいろ付いています。

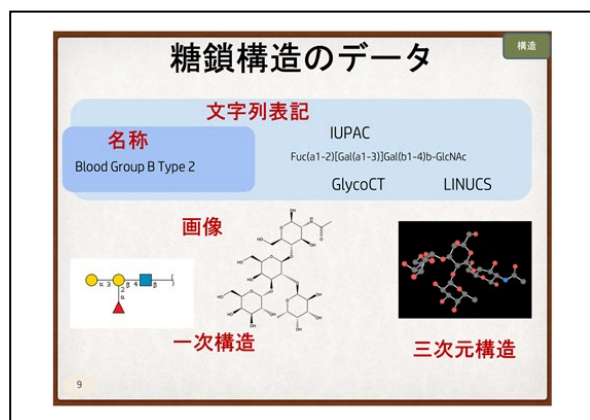
左下のカラーのものは単糖のシンボルが複数つながって糖鎖になっていますし、真ん中の画像のような元素記号を使った化学構造式や、シミュレーションで使われる三次元構造など、さまざまな構造データがあります。こういうものが論文などいろいろなところにばらばらに存在しています。全て一つの糖のことを表しているのですが、分野によって、自分たちに一番わかりやすいように書き表し方が違ってきます。しかし、データを見る上では、どれかを見たときに他のデータも見えるようにならないといけないということがあり、これらのデータをうまく扱っていただけるような仕組みが必要です。

糖鎖科学における標準化

そういった中で、糖鎖科学における標準化、共通の言葉やルールをつくっていこうという動きがあります（図6）。JIS Z 8002:2006（標準化及び関連活動—一般的な用語）における標準化の定義には「最適な秩序を得ることを目的として」と書いてあります。標準化は、相互理解や互換性の確保、品質、正確な情報という点



(図4)



(図5)

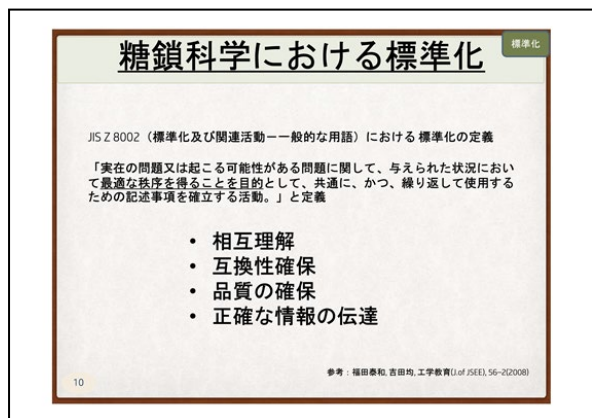
で非常に重要であり、糖鎖の分野でもそういうところを意識した取り組みを行っています。その取り組みの中で、糖鎖構造の記号について、それから Glc などの文字列表記について、そして、糖鎖のガイドラインづくりや、糖鎖構造に ID を付けてひも付けることで全てが同じであることを示す仕組みづくりに取り組んでいます。

図 7 は糖鎖構造で使う記号で、糖鎖と単糖です。これを糖鎖の領域でみんなで使っていきたいという流れになっています。これは日本だけではなくアメリカの国立生物工学情報センター (NCBI) や国立衛生研究所 (NIH) などの人たちが主導しているのですが、それにいろいろな形で世界の研究者が関わり、いろいろなルールを決めていき、そしてワーキンググループで、この記号についてどうするかという議論が行われています。その中で、最初に 2015 年に、このように糖鎖の記号を表しましょうという論文が出ました。その後もアップデートされるごとに論文を出すと同時に、研究者たちがインターネットでフリーで見られるような形で情報を提供しています。

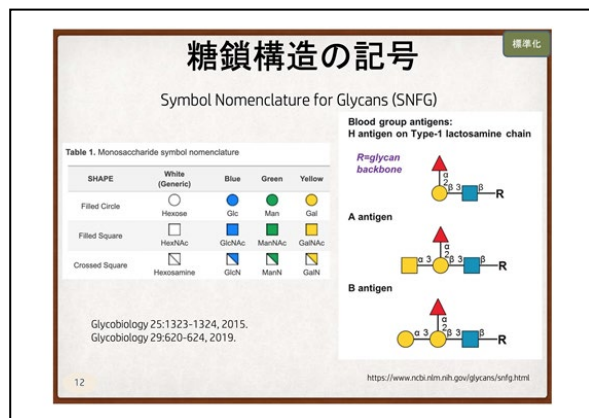
これだけ見ると、糖の領域はこのように書いてあればみんな分かるではないかというイメージを持つかもしれませんが、私が 14、15 年前に糖の領域の学会に行ったときには、この記号がポスターやスライドによってみんなばらばらだったのです。このグルコースは青で書いてありますが、赤の丸がグルコースという脚注があるポスターなども何割かありました。私

自身、その構造に興味を持っているので、一時期、学会のポスターや講演で糖鎖の研究者の人たちがどんな記号を使っているのか、SNFG の前には CFG、オックスフォード、IUPAC などいろいろな記号があるので、それぞれ統計を出していたのですが、非常にばらばらでした。それがだんだん統一化されてきています。普及啓蒙活動が功を奏していると思うのですが、昨年参加した学会では、かなりの人たちがこれに従っている、またはこれに近い形で書くようになっていました。

次は文字列表記です。IUPAC など糖鎖を表現するときがあるのですが、糖鎖の構造のデータ管理をするときに、やはりコンピューターを使って糖鎖の構造を表せないとなかなか管理が難しいです。そういうときに、国際糖鎖構造リポジトリを作ることになったのですが、そのときに使える糖鎖構造を表すユニークな文字列が必要だということで開発が始まったのが Web3 Unique Representation of Carbohydrate Structures (WURCS) です。ユニーク性があるので、同一の糖鎖構造かどうか、この文字列を使うことによって判断できるようになってきています。糖鎖の構造は、分岐構造、繰り返し構造、環状構造、分岐とリピートがあるなど非常に複雑です (図 8)。化学構造と化合物は、ある意味似ているものがあります。既に DDBJ という塩基配列のリポジトリなどでは配列で書けるので、そういうものは比較的やりやすかったのですが、そういう方法が糖鎖では取れなかったので開発したのが WURCS です。



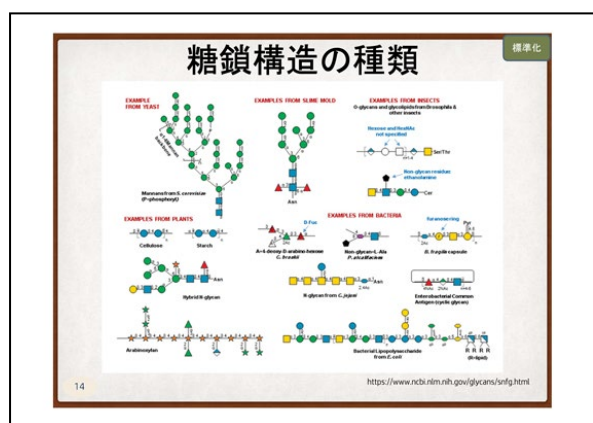
(図 6)



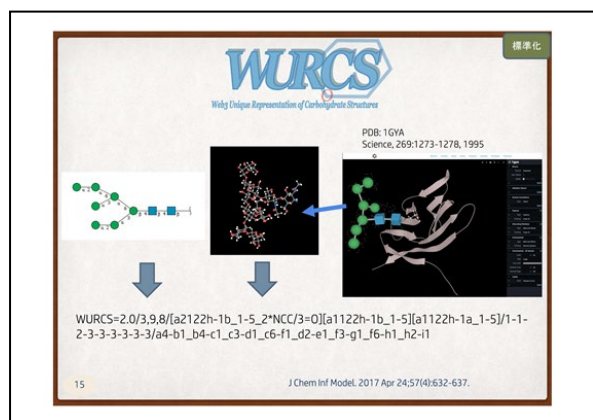
(図 7)

一例をお示しします。図 9 は Protein Data Bank (PDB) です。蛋白質の立体構造のリポジトリですが、蛋白質と三次元の SNFG という表記方法で書いたものがあると、この蛋白質に糖鎖が付いていて、この糖鎖が何なのだろうと取り出したときに、糖鎖の SNFG 記号で表すことができ、原子レベルで三次元で表すことができます。WURCS という表記法に変換することで同じものだと確かめることができ、他のエントリーなどで同じ糖鎖の構造があれば、同じ文字列になることが保証できるというものです。

これまで標準化・共通化の話をしてきました。少し毛色が違いますが、ガイドラインが糖鎖科学の領域で作られてきました。Minimum Information Required for A Glycomics Experiment の頭文字を取って MIRAGE というプロジェクトなのですが、目的としては、文献の糖鎖関連データの品質を向上させようというもので



(図 8)



(図 9)

す。実験をどのようにするかというものではなくて、実験結果を正しく解釈して再現できるようにする。論文文化されたデータを誰でもきちんと再現できるようにする。そのためにはどんな情報が必要かということを、MIRAGE のグループに世界中の研究者たちが参加して議論し、いろいろなガイドラインが作られています。そのガイドラインに従って論文やデータベースを利用していきましょうという動きがあります。

MIRAGE のガイドラインは糖鎖の実験のガイドラインで、サンプルの調整方法、質量分析の実験、マイクロアレイの実験、液体クロマトグラフィの実験などについて定めています。来月にもまた MIRAGE の会議が開かれて、新しいガイドラインについて検討されます。このように必要に応じてガイドラインを作って、それを領域で利用していこうという動きがあります。

図 10 は、糖鎖構造に対して ID を付加してあげようという話です。糖鎖構造にはたくさん書き方がありますが、それに対して ID を付けてあげると、全部同じものだということが分かります。そのために WURCS という文字列を作ってきましたが、その基になったのは 2013 年に中国の大連で開催された ACGG-DB の会議でした。日本をはじめアメリカ、オーストラリア、ドイツ、ロシア、中国、韓国、台湾の人たちが参加し、糖鎖のデータベースやリポジトリ、アクセス番号が必要だろうという議論がなされました。それ以前にもこういう議論はされていましたが、この会議において、具体的に国際的なリポジトリを 1 個作

(図 10)

ってみんなで利用していきましょうという合意が得られました。データの範囲としては、糖鎖の構造と登録者、登録日ぐらいのシンプルなリポジトリを作り、それを利用しようということになりました。

糖鎖科学のリポジトリ

研究者が実験をして糖鎖の構造などのデータを登録すると、リポジトリはその構造を見て適切なアクセッション番号を研究者に発行します。研究者は、このアクセッション番号を論文に記載することにより、どんな糖鎖を用いたかを明確にします。一方で、それ以外の研究者たちは、論文を読んで、そこにアクセッション番号が記載されていると、このアクセッション番号を基にリポジトリのデータを見ます。論文には自分に馴染みのない構造で書いてあったとしても、リポジトリを見ると、自分にとって分かりやすい表記がされています。そうすることで、どんなものを扱っているかがよく分かるという仕組みです。

このようなリポジトリとして、GlyTouCan という糖鎖構造のリポジトリを作りました。それから、UniCarb-DR と GlycoPOST という二つのリポジトリがあるのですが、これは最初の頃にお話した糖鎖構造を決めるための質量分析に関するデータを保存するためのリポジトリです。これらのものを、標準化を利用して作っています。

まず GlyTouCan に関してです (図 11)。これは国際糖鎖構造リポジトリといって、世界で唯一の糖鎖構造

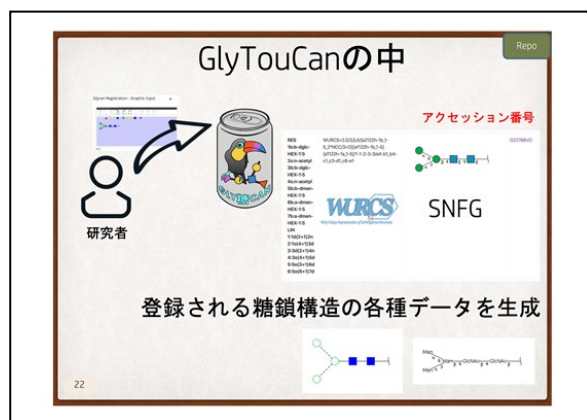
に対するリポジトリです。世界中の研究者が自由に登録・利用できるようにしています。研究者には糖鎖構造を書くためのウェブツールを提供しています。それを使ってここでお絵描きをして登録ボタンを押すと、GlyTouCan のシステムが働いて、アクセッション番号を研究者に返すという形になっています。先ほどの WURCS という表記法や GlycoCT、IUPAC など複数の構造でデータを入力しても、同じようにアクセッション番号を返すというシステムです。

図 12 は GlyTouCan の中身です。先ほど言ったように研究者がお絵描きをして、その構造を GlyTouCan の中に投げ込みます。例えば GlycoCT という表記方法のものを GlyTouCan のシステムで WURCS という表記方法に変換したり、SNFG 記号のイメージを作ったり、オックスフォード、IUPAC というさまざまな表記方法で糖鎖構造を表現します。あとは WURCS という文字列が GlyTouCan の中にあるかどうかを検索して、もしあれば、その構造は既に登録されているので、そのアクセッション番号を研究者の方に返します。なければ、新しいアクセッション番号を発行して研究者に返すという形になっています。

図 13 は、UniCarb-DR というアノテーションした質量分析データのリポジトリです。MIRAGE のガイドラインに、サンプル調整の方法、どういう条件で測定したか、スペクトルのピークデータなど、アノテーション済みのデータを登録することができます。この MIRAGE のガイドラインに沿ったデータをウェブの



(図 11)



(図 12)

インタフェースで書き、それをエクセル形式でエクスポートすることができます。

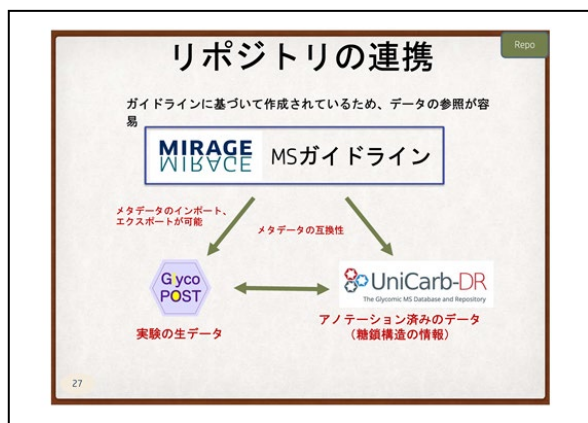
図 14 は GycoPOST というリポジトリです。こちらでも質量分析データのリポジトリで、MIRAGE のガイドラインに従ってメタデータ等を作っていく、UniCarb-DR で作られたエクセルデータをインポートして利用することもできます。違いは、測定したときに機械から出てきた生データです (図 15)。UniCarb-DR は、人がアノテーションしたデータを入れるリポジトリですが、GycoPOST はアノテーションされていない、生で機械から出てきたデータをそのまま入れているリポジトリです。これらは、MIRAGE のガイドラインに従ってそれぞれアノテーションされたデータと生のデータが作られているので、そのメタデータには互換性があり、それぞれのデータを参照することができます。



(図 13)

次に、分析データのリポジトリです (図 16)。アノテーションが付いているデータを、UniCarb-DR というリポジトリから UniCarb-DB というデータベースにデータをインポートするときに、糖鎖構造に GlyTouCan のアクセッション番号を付けて UniCarb-DB の方に持ってきてデータベースを整理します。UniCarb-DB の方では GlyTouCan のアクセッション番号が付いているので、糖鎖の研究者たちはどんな糖鎖構造なのかがよく分かるようになっています。

図 17 のように、GlyTouCan という糖鎖構造のリポジトリを作って、それを GlyCosmos、Glycomics@ExPASy、GlyGen という、日本、スイス、アメリカのプロジェクトが共通して利用していきましょうということで、糖鎖領域の三つのプロジェクトで GlySpace Alliance をつくり、データの共通利用やライセンスについてみんなで協力していく形になっていま



(図 15)



(図 14)



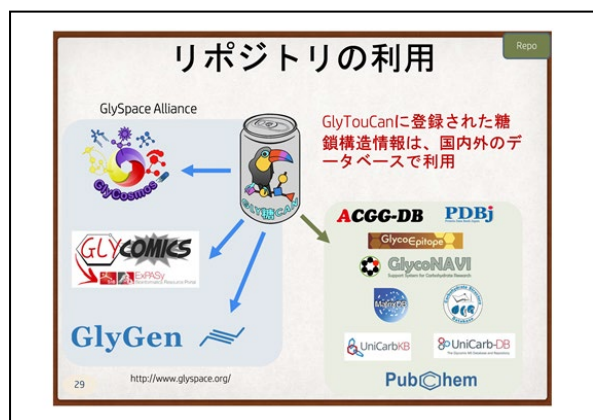
(図 16)

す。それ以外にも、糖のデータベースや、先ほどの PDB、PubChem Compound のデータベースなどに対しても GlyTouCan のアクセッション番号を付けることによって、いろいろな領域において糖鎖の構造がどのようなものかが分かるような取り組みをしています。

まとめ

糖鎖科学におけるデータ管理のために、標準化・共通化をして、糖鎖の記号と文字列表記を決めてアクセッション番号を付け、品質向上のためにガイドラインづくりを行ってきました。これらの標準を使って、GlyTouCan、UniCarb-DR、GlycoPOST の各リポジトリを運用・連携しています。これらの整備によって、リポジトリとデータベースを利用した正確な研究データが利用可能になっています。

これらのプロジェクトは、GlyTouCan に関しても現在進行形で研究開発しており、まだまだいろいろな問題があります。ユーザのサポートをしなければいけないなど、運用でも大変なところはありますが、糖鎖の研究者に役立つような環境を提供していきたいということで、私だけではなく、日本だけではなく、世界の研究者が協力して糖鎖領域のデータを良くする取り組みをしています。



(図 17)

●フロア 1 高エネルギー加速器研究機構の職員です。物理屋からすると、とてつもなくたくさんのバリエーションがあって気が遠くなるような話ですが、糖鎖以外にもさまざまな物質が世の中にはあると思います。それらの標準化を世界的に統括するような仕組みはあるのでしょうか。例えば物理だと、国際純粋・応用物理学連合 (IUPAP) という国際的な委員会のような組織が大きなことを決めていくのですが、化学の世界でも、例えば幾つか流儀があったときに統括したり、足りないところを補ったりするような、仕切る組織があるのでしょうか。

●山田 まず化学の方では、IUPAC で元素記号のルールが明確に決められており、糖鎖の方では国際生化学・分子生物学連合 (IUBMB) と IUPAC が連携して命名法のガイドライン的なものを出していますが、それだけでは対応できないものが結構あります。あまり悪口を言っては良くないですが、お医者さんで、化学構造式で書いても分からない人、名前でも書いても分からない人、分野ごとに自分たちが分かればいい、それで研究が済むという領域の人たちは、それ以外のことは必要ないけれども、論文を読んだときにそれが何なのかが分からなければいけないので、それをどうするか。糖鎖の中でそれができていないということで、大連で開催されたミーティングでそれを明確にしようという動きがありました。

糖鎖の領域には幾つか学会がありますが、学会主導でガイドラインを作るというよりも、ある程度コミュニティが主導になって、「これが必要だ」という動きになる場合が多いです。

第 3 回 SPARC Japan セミナー2019

「実践 研究データ管理」

原子力機構における 研究データポリシー策定に向けた検討

熊崎 由衣

(日本原子力研究開発機構)

講演要旨



日本原子力研究開発機構（原子力機構）では、研究データの管理と公開に関する方針を定めた研究データポリシーの策定に向けた準備を行っている。策定にあたり、現在の研究データの取扱いや公開状況に関する調査や、研究者からのヒアリングを行った。これまでの検討において、原子力機構の研究開発の特徴のため公開には慎重な判断が必要であること、研究開発分野が多岐にわたることから統一的な方針を策定する難しさ、公開データの質や信頼性を確保する必要性などについて議論している。今回は、これらの調査や検討の状況を紹介し、組織における研究データ管理について考える機会としたい。



熊崎 由衣

日本原子力研究開発機構 研究連携成果展開部 研究成果管理課に勤務。高校教員、コンサルティング企業（図書館、産学連携担当）の後原子力機構図書館に転職した。2018年7月より現職にて、テクニカルレポート等の編集・刊行と研究データポリシー策定を担当。前職の経験を生かして、研究開発法人でのオープンサイエンスの実践を手探りしている。研究データ活用協議会企画委員、研究データライセンス小委員会メンバー。

国立研究開発法人日本原子力研究開発機構（原子力機構）は、研究データの管理や利用に関する組織としての方針、研究データポリシーを策定しようとしています。本日はその現時点の検討内容について、検討のプロセスを中心にご紹介します。

クル技術（使用済み燃料を発電に再利用する技術）、施設の廃止措置が主要なテーマです。

これらの研究開発を行う拠点が国内に 11、施設は 90 ほどあり、約 3,100 名が勤務しています。研究職はそのうち 600～700 名です。私は茨城県の原子力科学

原子力機構の概要

原子力機構は総合的な原子力の研究開発機関で、研究と技術開発を行っている国立研究開発法人です（図 1）。研究テーマとしては、2011 年の福島第一原子力発電所事故以降は、その対処が大きなミッションとなっています。他に、原子力の安全や基礎基盤研究、バックエンド技術（放射性廃棄物の処理）、高速炉サイ

原子力機構の概要：1/3



- ・我が国唯一の総合的な原子力の研究開発機関
- ・主要テーマ
 - ・ 福島の再生・復興に向けた技術の確立
 - ・ 原子力安全の継続的改善
 - ・ 原子力を支える基礎基盤研究
 - ・ バックエンド技術の確立
 - ・ 高速炉サイクル技術の確立
 - ・ 「もんじゅ」、「ふげん」の廃止措置の実施
- ・ 職員数：
 - ・ 約3,100名（研究職／技術職／事務職）

(図 1)

研究所の図書館に勤務しています。

組織は大きく 3 種類に分かれています (図 2)。管理系の業務を行う組織、機構全体に関する管理と研究開発のそれぞれの事業を行う組織、研究開発を行う部門です。その下に研究所やセンターがあります。

一つの大きなプロジェクトの実施のために研究をしている組織と、それぞれの大きなテーマの中で自由度高く研究開発を行っている組織があり、多様な活動が行われています。それぞれの部門では、研究職の方は研究開発を行い、技術職の方が設備や機器の維持、研究所全体の放射線管理などを行っています。また、各部門には管理系の業務を行う事務職も所属しています。

多様だと申し上げましたが、図 3 のパイグラフは過去 3 年分の Web of Science のドキュメント数と研究分野を調べたもので、例えば緑が核科学、核技術、紫が物理学の応用です。皆さまの原子力機構のイメージと合っていますでしょうか。私も転職する前は「エネルギーなどかな？」と単純に思っていたのですが、今は環境や地球科学、それから、グラフに表れていないのですが地質や環境学、コンピュータサイエンスなどの

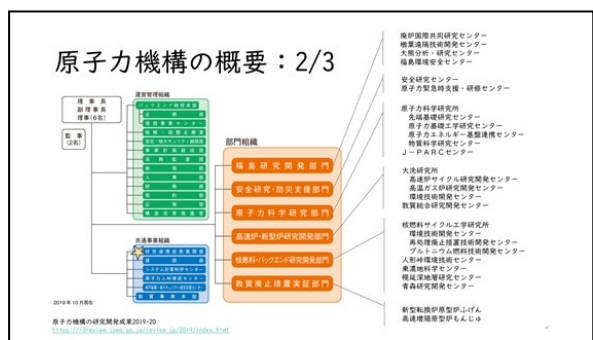
研究をしている人もいます。アプローチもそれぞれで、理論の方も実践の方もいますし、フィールドワークに出掛けていく方もいます。原子力と一言で言っても、研究分野も手法もさまざまです。

原子力というテーマであるため、職員が論文投稿や口頭発表を行う際には、事前にその内容の確認と決裁を受けなければならないというルールがあります。その情報を図書館で集約・管理して機関リポジトリなどの形で発信しています。ですが、研究データ管理はそれぞれの部門に委ねられていて、公開も、業務における考え方や必要性、研究分野の文化、慣習によっており、統一的な手続き、方針などは現時点ではありません。

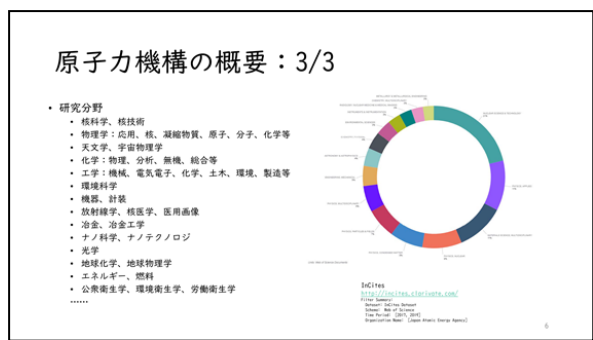
公開の例を挙げます。図 4 は技術レポートですが、DOI、本文の PDF ファイルのリンク、付録へのリンクがあります。利用できる形で公開していますが、ライセンスは明示していません。他にも、福島研究開発部門が放射性物質のモニタリングデータを公開しており、利用者の便利のためにいろいろなフォーマットを掲載しています。このように、統一的な方針の下ではありませんが、利用可能な形で、ある種のデータを公開している例があります。

研究データポリシーの概要

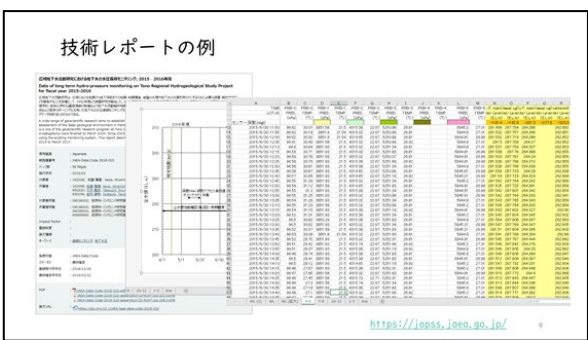
このような状況がある中で、組織としての研究データの管理の方針を考えています (図 5)。現時点での研究データポリシー (想定) は、ポリシーと内部ルールを階層で考えています。つまり、ポリシーをつくるだけではなく、実践するためにはルール化が必要だと



(図 2)



(図 3)

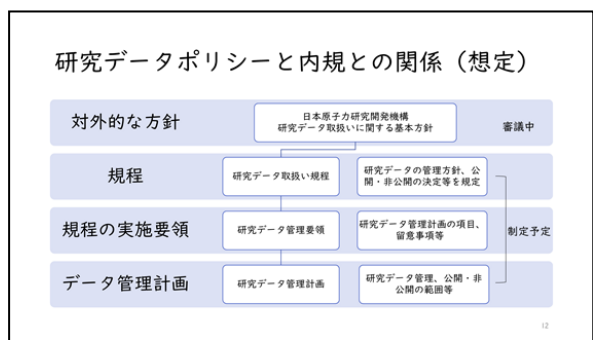


(図 4)

いう解釈です。ポリシーの本文を検討する際にも、運用をある程度想定しています。

ここでは要点だけご紹介します（図 6）。本文は A4 で 2 ページ程度、対象は、「統合イノベーション戦略」に「サイバー空間上での研究データの保存管理」とあるので、電子データを対象としています。その管理に当たっては、データ管理計画をつくりまします。データ管理計画は研究開発部門の決裁の他に、取りまとめ部署が確認するという方針になる見込みです。

公開する対象は、論文などとして公開する成果に付随するもの（エビデンスデータ、バックデータ）、データセットとしては先ほど例を挙げたような形で公開するものを想定していて、これらは検索や二次利用を可能にしてライセンスや利用条件を明示して公開したいと考えています。公開しないデータについては、必要な情報管理と制限を行います。右側の同心円は内閣府の資料を参考にして少し表現を変えたものです。以降、本発表では、管理するけれども公開しないデータを緑色、公開するデータを水色と色分けして表します。



（図 5）



（図 6）

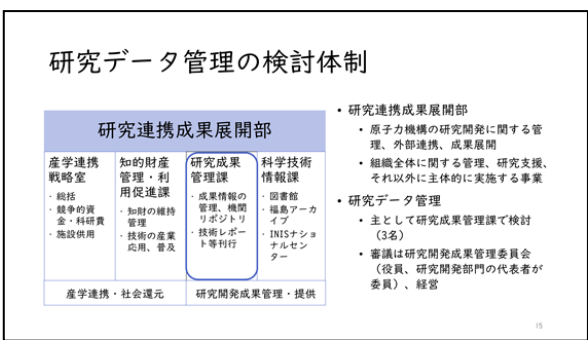
検討のプロセス

研究データ管理は、共通事業組織の研究連携成果展開部の中の研究成果管理課で検討しています（図 7）。研究連携成果展開部は図書館の業務と産学連携業務が 2 本柱で、原子力機構の成果情報の管理、発信から展開までの研究支援の業務を行っています。研究成果管理課は、成果情報の管理、リポジトリの運用、技術レポートの編集・刊行などを行っています。私の主な業務はレポートの編集や校閲です。

当課の課長と上司と私で研究データ管理を考えているのですが、課長は基幹システムが専門で、上司と私は修士を出て研究の経験があるとはいえ人文社会科学系です。原子力機構では、研究者や情報システム担当部署に意見を聞いてはいますが、実際の策定や検討に研究者が直接関わっているわけではありません。ただ、審議などは研究系の方も含めて行っています。

取り組みのきっかけは、「統合イノベーション戦略」への対応です。研究データポリシーだけでなく、競争的資金の獲得の際にもデータ管理計画が必要とされたため組織的に対応しようとしています。以前から情報収集などをしていたのですが、2018 年度から本格的に着手して、2019 年度にポリシーや素案の検討をしています（図 8）。

まず、組織や私の所属する研究連携成果展開部内外の方にデータポリシーのことを知ってもらうためのアドボカシー活動をしました。キックオフとして、主に研究開発部門の幹部の方を対象に説明会を開催しました。これは「オープンサイエンスとは」「なぜデータポリシーをつくるのか」という説明と、今後の協力を



（図 7）

依頼するものです。原子力科学研究所の会議室と全国16カ所をテレビ会議でつないで60名程度の方に説明しました。

続いて、科学技術・学術政策研究所（NISTEP）科学技術予測センターの林和弘上席研究官を招いて、「なぜ研究データポリシーが必要なのか」というテーマで意識を高めるための講演会を開催しました。これには70名程度の研究者が参加されました。まずこのような周知と理解促進、依頼をしておき、その後状況把握のための調査依頼をしました。

それと並行してポリシーを検討しました。書きはじめたのは2018年の年末ごろで、年度明けから本格的に練りはじめ、5〜6カ月でドラフトを作って内部パブコメを実施しました。以降、本資料で研究者という言葉が出てきますが、特に職種を指定せず大きく研究開発を行う人という意図で使います。この状況把握とポリシーの検討について、もう少し詳しく申し上げます。

ポリシーを作るといっても一体どうしたらいいのか分からない状況ですので、現状を把握するため、研究開発部門を対象に調査を行い、既に何かルールや実践がないかを問い掛けてみました。研究データの管理ルールがあれば参考になるかと思ったのですが、統一的な方針はない、ただし公開している例はあるということをお伝えいただきました。

その例がJENDL（核データライブラリ）です（図9）。これはデータベースで、輸出管理の手続きを経て、核データのライブラリとして公開しているものです。

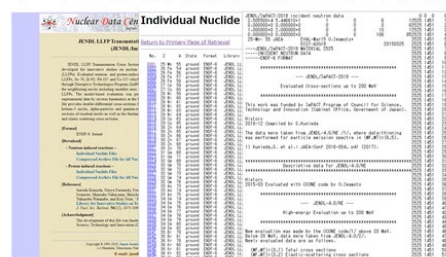
また、原子力科学研究所で観測している環境放射線

のモニタリングの例では、地図に合わせて、逐次、データを公開しています（図10）。このように業務で取得したデータは技術レポートとして刊行することもあります。ただ、研究開発だけが主目的ではなく、業務のために取得していて価値を持ち得るデータがあるということ、これで把握しました。

続いて、研究者の方に、どんな研究活動をしているかを教えてもらうためのインタビューを20名程度の方を対象に実施しました。これは随時継続しており、実際に幾つかの実験施設や装置も見学させていただきました。管理はそれぞれの組織の中で必要なものを調達して、それぞれがルールをつくって行っていました。公開は論文と共にデータを公開したことがある、ただ、それは supplemental information、appendix、supporting information などとして、他者が利用可能なデータを公開したということでした。投稿前に何らかのリポジトリに登録した事例も知りました。

また、例えば物理学の核反応データについては、論文を発表したら、そのデータをデータベースに登録するまでが論文を書くということであるなど、それぞれ

データベースの例（JENDL：核データライブラリ）



（図9）

実施した取り組み

- ・ アドボカシー
 - ・ 2019/4：キックオフ・協力依頼の説明会（全職員対象）
 - ・ 2019/6：理解促進・意識醸成のための講演会（研究職対象）
- ・ 状況把握
 - ・ 2019/5-6：研究開発部門対象の調査
 - ・ 2019/5-9：研究者へのインタビュー
 - ・ 随時：組織内外、出版社等の状況調査
- ・ 研究データポリシーそのものの検討
 - ・ 2019/9：ポリシーへの内部パブコメ（全職員対象）
 - ・ 2019/10：研究開発成果管理委員会（研究開発部門の幹部）による審議
 - ・ 2019/11：役員による審議



（図8）

環境放射線監視情報の例



（図10）

の研究分野で慣習があるということを教えてもらいました。個別にデータをリポジトリに登録するという方法ではないだけで、他者が利用できる形でのデータの公開は、研究分野によっては特別ではなく行われてきたのだと思います。

また、管理や公開についての考え方を尋ねると、一定の理解はあるものの、負担を懸念する声や、管理システムが研究にとって必要不可欠である、あるいはデメリットを上回るメリットがあるならば使う、また研究者として評価されるのであれば公開してもよいという意見を頂きました。データを公開した例があると聞くと、その研究者の論文を検索してみたり、外部発表手続きの書誌データを確認したり、データジャーナルで出版している方を検索するなど、並行して状況の調査、整理をしていきました。

図 11 の研究データのライフサイクルは、主にデータを取得して成果を公開するまでの過程です。上段は、論文と共に公開する場合です。大量に取得したデータ、取得したそのまま加工していない生データ、例えば条件を変えて複数回実験を行うかもしれません。それがいろいろな分析や取捨選択によって小さくなり、保存（固定化）されて、論文などを執筆し、その中のデータを公開する。下段は、データのまとまりとして他者も利用可能な形に加工し、整えて固定化して保存・公開する。このような流れになるのではないかと整理してみました。

整理をしながら研究データポリシーの本文も検討し、ドラフト版について理事の確認を経て、パブコメを実施しました。パブコメは原子力機構の職員の方なら誰

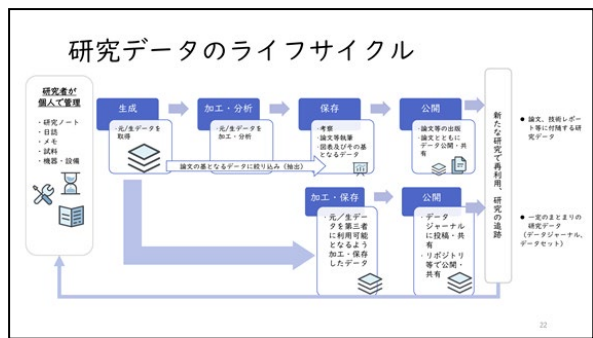
でも回答できるものです。ポリシー本文の他に、解説資料として左側にポリシーの本文、右側に解説や用語説明という構成の資料を5ページほど添え、かつ、概要説明のために「オープンサイエンスから今後の予定まで」という資料を添えてパブコメを実施しました。

すると、研究データ管理はどの範囲で行うべきか、例えば組織なのかという問題や、その先の運用についてさまざまな意見がありました。ただ、ポリシーの本文についてはあまり多くの意見が出ませんでした。これら全てのコメントに回答を作成し、原子力機構内に発信しました。

パブコメの後に再度検討し、次は研究開発成果管理委員会で審議をしていただきました。これは、研究連携成果展開部の図書館部分が研究支援のために行っている事業について審議する委員会で、当部の担当理事と部長、そして各研究開発部門の推薦された委員から成るものです。審議の結果、素案は承認されたのですが、規程との整合性を取ってほしい、公開・非公開の判断のポイントやガイドラインを示してほしいという意見を頂いたため、データ管理計画を規程に定め、今後、要領やマニュアルを整備するという対応を予定しています。

これらを経て担当としては、ポリシーそのものというよりも、研究者はその実現に伴う影響や日々の業務への影響に関心があるのではないかと考えました。また、多様な研究分野がある中で統一したルールを作れるのだろうかというのが個人的には疑問ですが、頂いたコメントに対応すべく、さらに検討、整理を深めています。

図 12 は研究データポリシーの対象についての検討です。パブコメで、データ管理計画は研究系の部署だけ行えばよいのではないかという意見があったのですが、管理の対象とする研究データは部署ではなくてデータの内容で考えようとしています。原子力機構の事業の中で取得するデータは何種類かに分けられます。一つ目は、もちろん研究開発のデータです。二つ目は、例えば実験を行った機械の精度や校正などのデータで、



(図 11)

研究公正の観点からは必要かもしれませんが、今回の対象ではないだろうと区別しました。そこで、研究開発そのもののデータと、事業活動を実施するためのデータ、さらにその中で、科学的価値などの何らかの価値を認めるものとし、これらを各部門で判断して管理する、と現時点では考えています。

それを、少し表現を変えると、図 13 のピラミッドの絵になります。緑は管理するけれども公開しないデータの範囲、水色が公開するデータの範囲です。いろいろな議論の中で「データは全部公開しなければいけないのか」という話題があり、論文のエビデンスデータとして求められる範囲を調査するため、出版社のポリシーや、各ジャーナルの guide for authors、FAQ を見てみました。結果、研究分野によって異なるのでしようけれども、大量のデータを求めているわけではなく、必要な最小限のデータセットを何らかの整理をして出すことが求められているのだろうと考えました。例えば、PLOS は「ミニマルデータセットの提供」、Springer Nature は「公表された研究の主要な所見を支持する最小のデータセット」という書き方をしています。

す。よって、ピラミッドの水色と緑の範囲は伸縮して、その度合いは研究者や研究分野、場合によってはレビューアやエディターによって変わり得るのではないかと現段階では考えています。

ただし、水色の部分の公開については留意が必要です。安全保障、法令や組織としての戦略に基づいて必要な情報管理を行っていく必要があり、その判断のガイドラインやポイントは今後整理していかなければなりません（図 14）。

ここまでの整理を、先ほどのライフサイクルの図に当てはめてみます（図 15）。緑は、ある程度伸縮するけれども管理するデータ、そして水色の重なっているところは、それぞれ研究分野や研究者によって変動するけれども公開する可能性のあるデータと想定しています。管理するけれども公開しない範囲は、個人で管理するもの、内部の関係者で管理するもの、外部を含めた共同研究などのプロジェクトの中で管理するものの三つに分かれると今は整理しています。この整理は、今後のデータのアクセス制限の区分やデータの管理システムの運用の中でもう少し詰めていきます。

研究データポリシーの対象

- 原子力機構で取得しているデータ
 - 研究開発そのもののデータ
 - 研究開発を実施するための周辺データ（施設・機器の校正、環境等）
 - 事業活動を実施するためのデータ
 - 事業活動を実施するなかで外部に提供・公開するデータ
- 研究データポリシーの対象とする範囲
 - 研究開発・事業活動の過程、あるいは結果として取得・作成したもの
 - 科学的価値、公益性や利活用可能性等の価値を認め得るもの

(図 12)

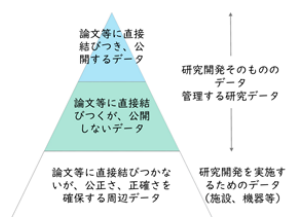
研究データ公開の留意点

- 研究データ公開に際する留意点、必要な情報管理
- 安全保障、法令等
 - 安全保障：核不拡散や軍事転用規制
 - 法令等：情報公開法上の不開示情報（個人情報、法人情報、国家安全情報、公共安全情報、審議検討等情報、事務事業情報）
- 組織としての考慮
 - 契約、協定、関係者間の取り決め
 - オープン・クローズド戦略（研究開発戦略、知的財産戦略上の考慮）

(図 14)

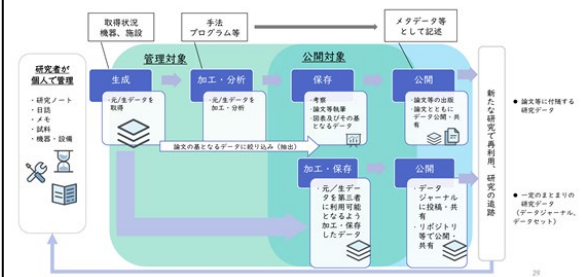
公開する研究データ

- 公開の対象
 - 論文、技術レポート、データベース等として公表するもの
 - 公益性が高く、利活用可能性等の価値を認め得るもの
- 論文等に付随するデータ
 - 出版社は全てのデータセットの公開を求めているわけではなく、論文の結果を検証可能な最小限のデータセット



(図 13)

研究データのライフサイクルとデータポリシーの対象



(図 15)

実際の運用を想定するために、データを公開する際の手続きに至る判断のポイントをフローにして考えてみました（図 16）。三つ目の分岐までは安全保障や法令で、その先がオープン・アンド・クローズ戦略です。このような運用に関する検討は、例えば管理システムだと GakNin RDM を Open IdP で試行させていただいたり、実は今週もまだ研究者にインタビューを続けていたり、繰り返し行っています。

検討の過程で行った事柄を大きく分けると、組織の研究開発について知る、外部の動向・方針を知る、そして組織のルールを調べるということです。整理する、形にするということを積み重ねて、当たり前のことをひたすらにこつこつと行っています。明日からできるという本セミナーのテーマからすると、きっと皆さん何かしら「何だ、当たり前だな」と思われることやできることがあるのではないかと思います。

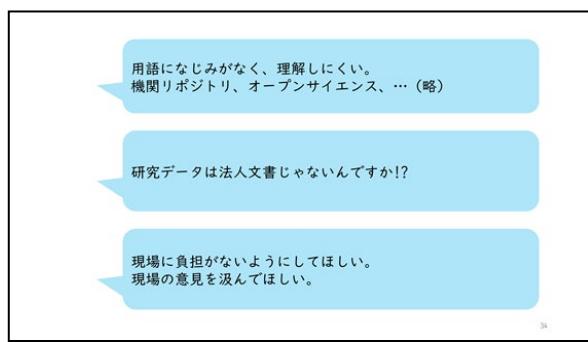
今後の取り組みと課題

審議の過程で頂いた印象的なコメントを紹介します。図 17 はパブコメや打ち合わせでのコメントで、研究データ管理について丁寧に説明する必要があることや、コミュニケーションが重要だということを感じています。図 18 はもう少し広い視点でのコメントで、組織として成果を公開することに伴う評価や責任をどう担保するのか。担保するとすれば、統一的なルールによる必要があるのではないかと、原子力機構では研究データ管理を組織の中で組織の文化を含めて考えざるを得ない状況で、それが難しい点だと思っています。いろいろなレポートや海外の事例を参考に研究データ

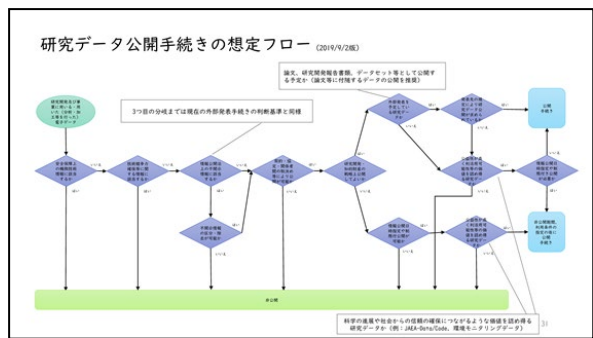
管理を考えるのですが、原子力機構に当てはめることができるものとできないものがあります。また、実際に実践するのは組織の中の個人です。原子力機構や研究分野のことをより学び、研究データ管理を考えなければいけないのだということを感じています。

原子力機構は、多様な研究分野があり、原子力故の情報管理や業務管理の適正さ、厳格さが求められるという特徴があります。研究データ管理の方針としては、全体で統一して実施する、判断という人の面とセキュリティというシステムの面の両方で、既存の情報管理も研究データ管理についても統一した方針やルールで厳格に行うという方針になる見込みです。

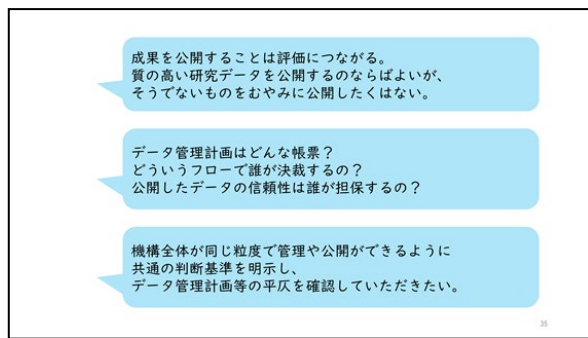
よって、取りまとめる側に要望される事柄としては、ルールや判断基準のガイドラインなどをつくること、情報管理や責任の所在を明確にしていくことがあります。プロジェクト型の研究を行っている部門も、個人の自由度高く研究を行っている部門もありますし、研究分野や手法によっても異なる状況があります。個人的には、そんな画一的なルール化ができるのだろうか、そして、そのような内容で実際にルール化した場合、



(図 17)



(図 16)



(図 18)

実際に研究開発に携わる方々が、より良い技術開発を行うことができるのだろうかということを懸念しています。

図 19 は今後の取り組みです。まず、研究データポリシーを策定・公開する。本来は研究データ管理に関するルールやサービスを設計するという思想に立つて行すべきかもしれないのですが、ポリシーに沿って規程やそれらの判断のガイドラインをつくるということを進めていこうと考えています。管理や公開のシステムについては、国の基盤である NII Research Data Cloud に期待していて、これらを 2020 年度中にできる限り進めたいと思っています。

最後に、全体的な課題を申し上げます。図 20 の上の「生データ」と下の「生データ」は、同じものを指しているでしょうか。データの標準化を図る非営利コンソーシアム (casrai) の用語集では、「生データ (raw data)」とは「処理されていないデータ」と定義されています。「生データ」とは、これらの用例や定義が間違っているわけではなくて、それぞれの人のバックグラウンドや文脈によって、使う言葉の語義が

今後の取り組み

- ・研究データポリシーの策定・公開
- ・研究データの取扱いに関する規程等の制定
- ・研究データ管理に際するガイドラインの作成
 - ・管理・公開の対象とする判断の基準
 - ・公開先やライセンスの選定に際する判断の基準
- ・管理/公開システムと運用方針の検討
 - ・データを管理するシステムの整備、データ管理計画の集約
 - ・機関リポジトリで研究データを取り扱う方針
 - ・機関リポジトリの整備

(図 19)

グラフのプロット値だけではなくて
生データも公開しないと意味がないのでは？

類似した研究をしている人の生データを使って
比較研究ができるので、生データ公開は有効です。

Data that have not been processed for meaningful use. Although raw data have the potential to become "information," they require selective extraction, organization, and sometimes analysis and formatting for presentation. As a result of processing, raw data sometimes end up in a database, which enables the data to become accessible for further processing and analysis in a number of different ways.
Ref. casrai "Raw data" <https://casrai.org/text/raw-data/>

(図 20)

異なるのだらうと思います。

同じことは研究者とのインタビューでも、研究データ管理における組織の中や図書館員同士の議論でも、この SPARC Japan セミナーの場でも起こり得るのではないかと思います。研究データ管理について同じ認識で議論していくことがとても難しく、私たちが使っている言葉の一つずつを、意味や意図を丁寧に確認しながら議論していかなければならないと感じています。

また、パブコメで「用語に馴染みがなく、理解ににくい」と書かれてしまったのですが、図書館としての日頃の業務や研究データ管理についてのアドボカシー活動が必要だと強く感じています。これらは研究データ管理について考える上での基礎的な課題です。

続いて、ポリシーを実践する際の研究者への負担と組織のルール化のバランスを組織の課題として感じています。研究者が研究しにくくなってしまうと本末転倒ですし、それはひいては組織のパフォーマンスに影響を与えると思います。その一方、質の保証という点もあり、組織として研究データを世に送り出すときに、どう責任を持つのか。データを公開したら思わぬ使われ方をして新たな展開が生まれ、イノベーションにつながることを期待するという語られ方をするオープンサイエンスの文脈の中で、あらかじめ質を見極める、保証することは意味があることなのだろうか、そして可能なのだろうか。そのための技術や人を私たちは組織の中でどう確保していったらいいのだろうか、ということを考えています。まとめると、組織として何のために管理して公開していくのだろうか、という議論です。世界や国の方針と組織の考え方のギャップの中で、手探りをしながら今後も進めていかなければならない、と考えているところです。

●フロア 1 軍事問題研究会の職員です。公開できないデータの取り扱いについてのお話がありました。これについて、例えばデータそのものがないのか、元々

研究がされていないのか、それとも研究がされているけれども出せないのかということは外部からは分かりません。従って、いずれ公開されるときが来るのかという問題も出てくると思います。例えばそれが公開できないにしても、その公開データに何か統一的な番号のようなものが振られて、「こういったものが何月何日に出ました」と明らかにされるのか。明らかにされなくても、表題が出て、例えばこういう理由でもって何年何月何日まで非公開だというようなことが出されるのか。出されない場合は、当然、請求する側とすれば、どうしてもこういう理由で欲しいということもあると思うのです。その辺の手続きはどのようになるのでしょうか。

●熊崎 正直なところ、まだそこまで詰めきれていないのですが、オープン・アンド・クローズ戦略の考え方の中で、例えばメタデータだけ公開して、データも必要であれば共同研究を実施しようという方法もあると思います。どういう研究が行われているか、どんな資産を持っているかということは、原子力機構だけでなく他の大学や研究機関でも選択的に出していくものだと思いますので、それぞれの組織の中での考え方に基づいて選び抜いて出していくという流れになるのではないかと想像しています。

●フロア 2 原子力関係のデータは大変重要な意味を持ち続けます。福島の実験からしても、例えば放射線医学のデータは広島・長崎のデータにかなり頼っていると聞いています。福島に関しても、チェルノブイリなど過去の何十年も前のデータが役に立つということです。これはあまり想像したくないですが、数十年後に世界のどこかで同じようなことがあったときに、「福島はどうなっていたのだろう」ということが残念ながらあり得ると思うのです。そうすると、当面必要と思われるデータが将来的に大変役に立つ可能性を秘めていると想像します。数十年という長いスパンにわたって、今あるデータを、今すぐ公開するかどうか

かは別にしても、後々使えるようにきちんと取っておくことが非常に重要ではないかと思います。

そう考えると、そのための技術的な基盤は、先ほどNIIを頼りにしているとおっしゃいましたが、NIIという組織が50年後にあるかどうか分かりませんし、さらにいえばMicrosoftという会社が50年後にあるかも分かりません。そこまで考えると、かなり長期的な視点でセーフティネットというか、いろいろ手打っておかないといけないと思います。

似たような問題で、われわれは素粒子の研究で大量のデータを取ると、そう簡単に大きな装置を世界中で作れませんので、とにかく次世代に残すことを考えます。次世代の人が新しい知見を基に昔のデータを解析すると、何か新しいものが得られるかもしれないので、どうやってそれを残すかということが非常に悩ましい問題なのです。似たような問題を抱えているかと思うので、何か知恵があったらぜひ教えてください。

●熊崎 おっしゃることはそうだと思いますし、福島のときはTMIの事故（スリーマイル島原子力発電所事故）などのデータを調査するようにと図書館へのレファレンス依頼があったと聞いています。お答えにはなっていないかもしれませんが、原子力機構では図書館で「福島原子力事故関連情報アーカイブ（FNAA）」を作っています。これはNDL（国立国会図書館）のWARPというウェブアーカイブの仕組みを使って、福島事故への対応として、環境回復や廃止措置というように選別したウェブアーカイブのポータルサイトです。ウェブサイトでの情報発信が福島事故のときは非常にたくさん行われたのですが、それが散逸してしまわないように選りすぐって保存する目的です。例えばウェブアーカイブという手段があるかもしれませんが、おっしゃるように、そもそも電子データの長期保存をどう考えるか、どう取り組んでいくかという世界的な図書館が抱える課題にも通じるのではないかと思います。

第3回 SPARC Japan セミナー2019

「実践 研究データ管理」

研究データ管理サービスの概要と 利用事例の紹介

込山 悠介

(国立情報学研究所)

講演要旨



学術機関における研究データ管理の需要が高まりつつある中で、国立情報学研究所（NII）は2020年度後半から総合的な研究データ基盤 NII Research Data Cloud の一部として、研究データ管理サービス GakuNin RDM を新たな学術情報基盤サービスとして提供していく。実際に機関で研究データ管理を行うために必要になるツールを提案するとともに、全国の学術機関で実施中の実証実験におけるユースケースを紹介する。



込山 悠介

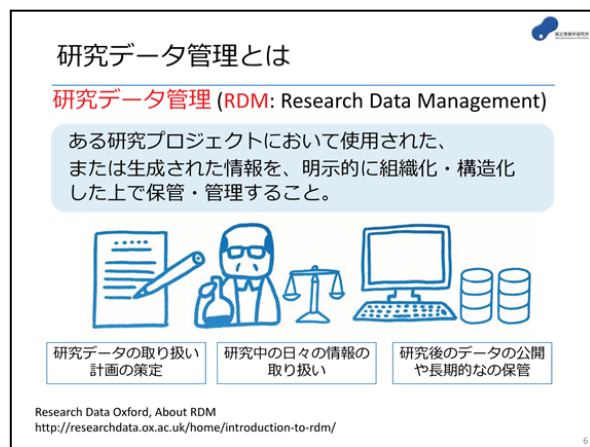
国立情報学研究所コンテンツ科学研究系助教。博士（農学）。2014年3月、東京大学大学院農学生命科学研究科博士課程を修了。日本学術振興会特別研究員（DC2）として生物情報工学の研究に従事。2014年4月、東京大学医科学研究所ヒトゲノム解析センター特任研究員としてスーパーコンピュータ運営や生命科学データベース開発のプロジェクトに従事。2016年4月より現職。2017年4月より国立情報学研究所オープンサイエンス基盤研究センターの発足に伴い、研究データ管理サービス GakuNin RDM の研究開発、運営および国際連携に従事。

NII が開発している研究データ基盤 NII Research Data Cloud の中で、研究データ管理を取り扱っている GakuNin RDM というサービスがあります。まずはこれでこの先 5 年、10 年を頑張っていくことで 50 年後につながっていくのだらうと、われわれのグループは一丸となって頑張っています。本日はこの研究データ管理サービスを紹介します。

本日は、研究データ管理の簡単な背景を少し述べて、NII Research Data Cloud、GakuNin RDM を紹介します。図書館研究者、URA、学術出版職の方々が、自分の組織内でコストミニマムに研究データ管理を実践するためのサービス GakuNin RDM の概要を理解することを本日のゴールとして設定しています。

研究データ管理の背景

釈迦に説法かもしれませんが、研究データ管理とはそもそも何かということを簡単に述べます（図 1）。英語では Research Data Management の頭文字を取って



(図 1)

RDMです。情報を探すときは、RDMという言葉で検索すると海外の先進的な文献がヒットするかと思います。これは、ある研究プロジェクトにおいて使用された、あるいは、そのプロジェクト中で生成されたリサーチデータを明示的に組織化・構造化する、つまりはメタデータ化したり、共通の語彙を作って、スキーマや名称を統一した上でデータベース間の言葉のずれ等をなくしたりすることも含めて行った上で、研究データを保管・管理していくことと定義されています。

具体的に現場の研究者は何をしたらいいかというと、研究に着手する、あるいは研究費申請書を作成する段階で研究者が、いつ、誰が、どこで研究データを取得して、どのくらいの保存するための容量、公開するための領域が必要になるか、そして、いつまで誰が保管して管理するのかを含めた研究データ管理計画を書く必要があります。さらに、最初に紙で研究データ管理計画を出すだけでなく、走りながら中間評価的に研究データ管理計画をアップデートしていくというアドバンストな考え方もあります。

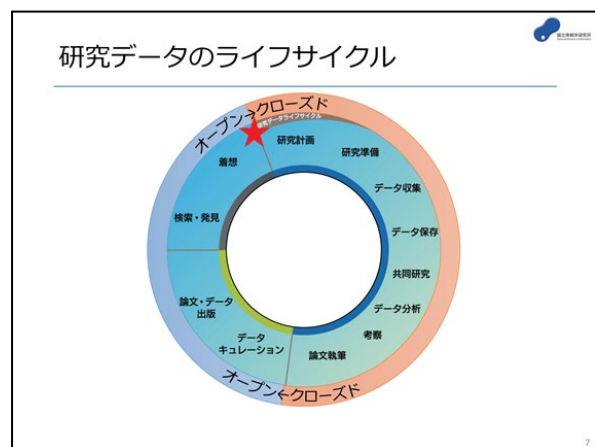
その研究データ管理計画に基づいて、研究者は日常出てくるデータのある程度構造化して、まずは自分が分かるようにします。それを共著者に渡したときには、共著者が何をやっているか分かるような情報は最低限付けます。データに“Read Me”を付けてという言葉はよく言われますが、そうした形でデータを人に渡す形で残していきます。その上で、論文がパブリッシュされた後、研究後のデータ公開、長期的にどこで保存

していくかといったこともあらかじめ決めておく必要があります。これが研究データ管理といわれている行為です。

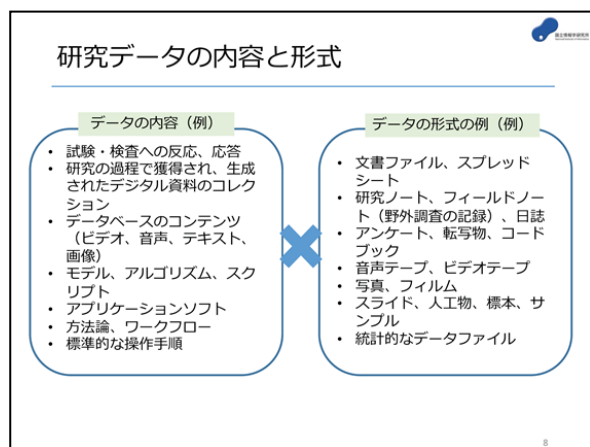
図2は研究データのライフサイクルです。研究の着想段階、研究を計画する段階、データを集めて分析し、考察し、論文を書く。あるいはその中間では共同研究者とデータを共有するといったプロセスもありますが、それは論文が出る前のクローズドなプロセスとされています。

そして、論文が出た後は、データをどこかのリポジトリで公開します。サブジェクトリポジトリと呼ばれる分野リポジトリにデータを預けて、そのIDを取得して論文を書いて出すというパターン、あるいは、昨今では機関ごとのリポジトリに論文や紀要と併せてリサーチデータも公開していくという流れをつくろうという動きがあります。ここはリポジトリ図書館が管轄しているところです。そこに入ったデータがハーベスティング（収集）され、検索、発見できるようにしていく。これが研究データのライフサイクルといわれているものです。

図3は、実際に研究データの中身にはどんなものが含まれているかということを一リストアップしたものです。やはり分野によって全く違うことは想像されます。汎用的には、データベース内のコンテンツとしてビデオや音声、テキスト、画像、あるいは理論系の話だと、微分方程式のモデル、アルゴリズムそのもの、プログラミングソースコードといったものを研究データとし



(図2)



(図3)

て保存する、あるいはデータだけではなくて、操作手順や正確に実験を行うための手順なども研究データです。秘伝のたれではないですが、研究室ごとに一子相伝的に伝えられるので、それを知るために留学するなど、そのラボに行かないと手順が分からないからいくら同じ装置があってもできないという場合もあり、こういったものも研究データに含まれています。それを表現しているのは、データの形式としてはワードファイル、エクセルファイル、あるいは音声テープやスライド資料などです。

一つの論文ができるまでには、10 の下支えになるスライド資料、さらにはその下に大量の生データがあるという文献もあります。こういったものを全て含めて保存していくのが研究データ管理の流れだと考えています。

研究データ管理が始まる前に、大学の皆さまは研究データの保存について、2014～2015 年ごろに 10 年保存というガイドラインの整備を進められました（図 4）。データの保存については各大学の部局ごとにガイドラインが既に整備されているかと思いますが、そこに書かれているのは、部局ごと、あるいは研究室ごとに DVD にデータを入れて保存しておきなさいということです。しかし、それではモダンな研究データ管理としては不十分です。

次に、研究データ管理の第二の流れといいますか、最近では資金配分機関から研究者が競争的資金を獲得する際には、研究データ計画を書きなさいと言われてま

す（図 5）。これは国立研究開発法人日本医療研究開発機構（AMED）、国立研究開発法人科学技術振興機構（JST）が先行的に行っていて、今年度は科研費の「学術変革領域研究」でも、研究者は申請して採択された際はデータ管理計画を改めて書かないといけません。そういう時代になっています。

そこで、やはり研究データ管理は機関単位、大学あるいは研究所の中で行った方がいいということで、学術機関で整備するのがよいだろうということになっています。なぜ研究データ管理を学術機関で整備すべきかということを二つの観点で整理します（図 6）。

一つは研究公正です。研究不正を起さないという流れから来ている考え方です。もう一つは研究推進です。これは、研究をどんどん促進するためにオープンサイエンス、オープンイノベーションで全国・全世界の研究者とデータを共有していくという前向きな観点

国内の研究助成機関のDMPの動向

機関名	方針策定状況
国立研究開発法人 日本医療研究開発機構 (AMED)	「疾病克服に向けたゲノム医療実現化プロジェクト」及び「難病克服プロジェクト」ゲノム医療実現のためのデータシェアリングポリシー（2016.4） 2018.5.1以降にAMEDが新規公募する事業について、<u>データマネジメントプラン</u>の提出を義務化（2018.3）
国立研究開発法人 科学技術振興機構 (JST)	「オープンサイエンス促進に向けた研究成果の取扱いに関するJSTの基本方針」策定（2017.4） 「戦略的創造研究推進事業におけるデータマネジメント実施方針」（2016.2）に基づき、<u>データマネジメントプラン</u>の提出を要求
経済産業省 産業技術環境局	「委託研究開発におけるデータマネジメントに関する運用ガイドライン」策定（2017.12） <u>データマネジメントプラン</u>の作成を要求
独立行政法人 日本学術振興会（JSPS）	「論文のオープンアクセス化に関する実施方針」策定（2017.3） - データに関する方針は未定

*データマネジメントプラン（データ管理計画書）の提出を要求

（図 5）

研究データ保存のガイドラインと規程

- 文部科学省**
 - 『研究活動における不正行為への対応等に関するガイドライン』（2014.8.26）
 - 研究機関は、研究者に対して一定期間研究データを保存し、必要な場合に開示することを義務付ける旨の規程を設けるべし
- 日本学術会議**
 - 『（回答）科学研究における健全性の向上について』（2015.3.6）
 - 論文等の形で発表された研究成果のもととなった実験データ等の研究資料は、当該論文等の発表から10年間の保存を原則とする
- 京都大学**
 - 『京都大学における公正な研究活動の推進等に関する規程』（2015.2.24）
 - 『京都大学の研究データ保存（理事裁定）』（2015.7.30）
 - 保存期間は、当該論文等の発表後少なくとも10年とし、これを下回って定めではない

（図 4）

研究データ管理を学術機関で整備

研究データ管理を学術機関で整備

- 研究公正**
 - 研究者はこれまで研究室や専門分野ごとの慣習で、研究データ管理をしてきた、その重要性は変わらない
 - しかし、「説明責任」の観点では、組織的なRDMを実施する必要がある、研究者が所属する機関が研究公正の責任を取る必要がある
- 研究推進**
 - 個々の研究者がRDMを実施するために「インフラ」を調達・運用することは、負担が大きく非効率的であり、持続性も担保されない
 - 学術機関にとってRDM環境を整備することは、オープンサイエンス時代における研究推進につながる

（図 6）

です。個別に詳細を見ていきます。

まず、研究公正です。よくわれわれは研究所を訪問して、「われわれの研究データ管理サービスを使ってください」とお願いすると、「われわれの研究所はもう何年も前から研究データ管理をきちんとしている。そんな不正が起こることはない」ということを時々言われるのですが、決して今まで専門分野や研究室ごとで長年データ管理されているものを否定するわけではありません。その重要性は変わりませんが、大学あるいは学術機関の組織として、もし何か事が起こったら調査委員会を立ち上げないといけません。そして、不正が起こったラボのどこにデータがあるのかということを出してもらわないといけません。組織としては説明責任の観点で、大学が決めた場所にデータを保存しておく必要があるだろうというのが、研究公正の文脈での研究データ管理です。

最初、日本国内では、研究公正のためにデータマネジメントのインフラを整備しないとイケないということでキャンペーンをしていました。しかし、それだと研究者受けが悪い、つまらない、もっと楽しく研究を便利にするためのインフラとして使っていった方がいいというアドバイスを受けて、それ以降は、研究推進のための基盤ということもかなり強めに言っています。

われわれのサービスを導入していただくと、研究者が研究室で個別に予算を取ってサーバーを買うことになります。最初はいいのですが、研究者は物を買うことしか考えていないので、その後のメンテナンスは大概ラボの学生やポスドクの人に任せてしまいます。私も言われた経験があるのですが、それでメンテナンスに入ると、スキルを持った若手の研究者が異動してしまうと、その後は有名なデータベースでも維持管理されなくなるということがよくあります。論文がたくさん引用されている、教科書に載っているようなデータベースだけれどもソースコードがないとか、データベースをアーカイブしようにも後任の先生が着任してサーバーを捨ててしまったなどという話は、私のバックグラウンドの分野でも聞いたことがあります。

図7は大学ICT推進協議会（AXIES）の提言書の中の記事を表で分かりやすく整理したものです。研究データ管理を大学研究機関で行っていくための学内のステークホルダーをリスト化してあります。第一には、大学等の執行部、研究担当理事、情報担当理事の先生方にご理解いただいて、導入をトップダウンで進めていただきます。

ただ、それだけでは不十分で、ストレージやネットワーク、認証サービスを準備していただくためには情報基盤センターの先生方に入っていただく、あるいは情報基盤センターと連携してシステム運用をされているベンダーの企業の方も一緒にやっていく必要があります。特に、研究者にいかにもマネジメントしてもらいかということと、データ管理から公開へ移る部分でのサポートという意味では、図書館が大きな役割を持っています。いかに研究データが管理されているかということが徐々に見える化されてくると、研究者にとっては嫌なのですが、大学の理事クラス、あるいはURAや研究推進部門の方からすると、自分たちの大学の研究力を分析するために研究データ管理の情報をを使って分析すること（IR研究）にも将来使っていくことになると考えられます。

NII 研究データ基盤 NII Research Data Cloud

NII の統合的な研究データ基盤、NII Research Data Cloud をご説明します（図8）。先ほどの研究データ

研究データ管理とは何をすれば良いのか		
学内での研究データ管理の位置付・方針決定	研究データ管理規定（ポリシー） 機関リポジトリ運用規則（利用規程、データのエンバゴ/解禁期間の設定等）	大学執行部 研究推進部 大学図書館 研究推進部
情報基盤の整備	ストレージ、機関リポジトリ、認証、セキュリティ 各手続きのためのシステム構築・運用（10年保存（研究者転出等への対応含む）、公開・長期保存、研究評価等向け可視化）	情報基盤センター 情報基盤センター
研究データ管理体制	研究データ管理の事務体制（DMP、研究データ10年保存） 研究過程における研究データ管理（データ構造化、管理、説明資料等付加） 研究データ登録・申請手続き（公開データ、公開データ、長期保存データ） 研究データ保存・公開手続き（保存・公開前の内容確認、データ管理情報等の付与）	研究推進部 URA 研究推進部 研究推進部 大学図書館
啓蒙啓発、研修体制	データ管理計画（DMP）作成支援 研究データ管理研修（院生・研究者向け、研究支援者向け）	研究推進部 大学図書館 大学図書館 研究推進部
研究データ利用体制	研究評価・研究戦略策定のための利用体制と手続 共同研究、産学連携等のための連携体制と利用	URA URA 研究推進部

（図7）

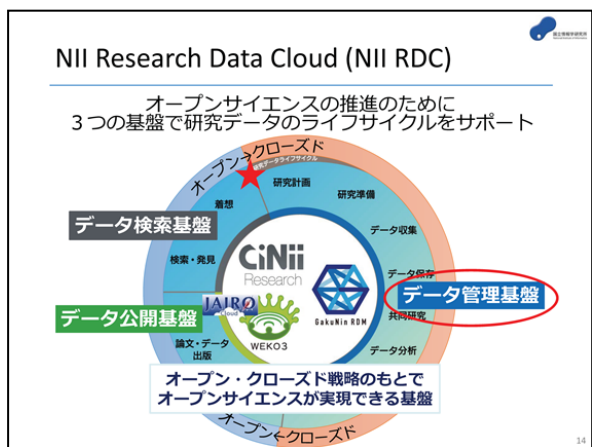
ライフサイクルの図をサポートするために、研究者の日常的なアクティビティをフルサポートするという意味で、NIIでは三つの基盤を開発しています。

一つは、データ公開基盤（WEKO3）です。これはJAIRO Cloudと呼ばれる、NIIが全国の570以上の機関に提供している機関リポジトリのクラウドサービスです。今までは文献だけでしたが、現在はデータも保存できるような改修が進んでいます。そこに登録されたデータが、次世代 CiNii と呼んでいるサービスで検索できるようになるという仕組みになっています。検索基盤（CiNii Research）では、論文と研究データだけではなく、科研データベースのようなものも併せて検索できるようになっていくだろうという見通しが立っています。もう一つ、新しくつくろうとしているのが、GakuNin RDM と呼ばれている研究データ管理基盤です。

この三つの基盤を使い分けることで、あるデータはクローズドで温めておく、あるいはどんどん公開して共同研究者等に見せていく、また、世の中のために出していった方がよいデータはオープンな領域の基盤に入れていただくということになります。本日はデータ管理基盤についてご紹介します。

研究データ管理基盤 GakuNin RDM の概要

図9はGakuNin RDMのサービスビジョンを示しています。研究推進と研究公正の二つの観点から、さらに研究者と機関にとってうれしい機能を分析しました。

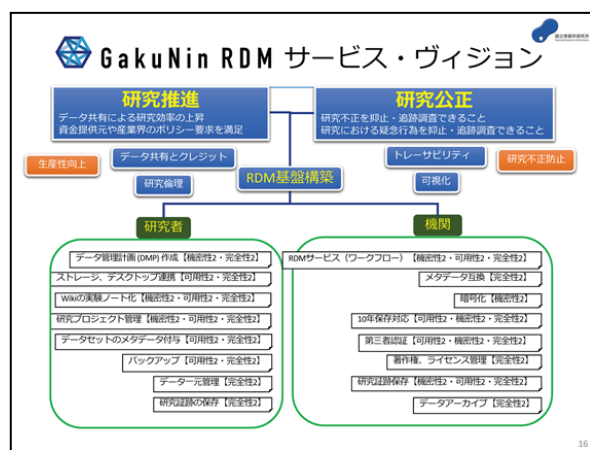


(図8)

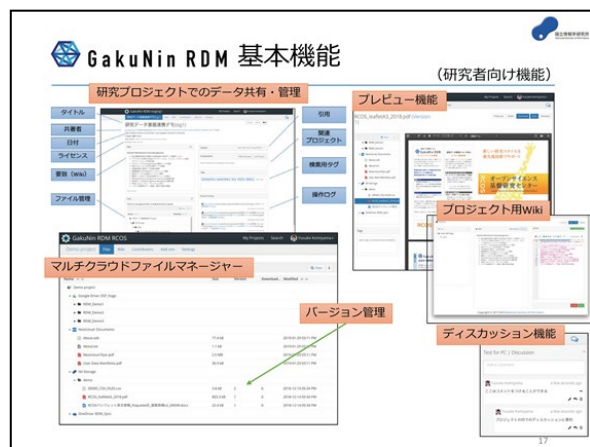
これは勝手にわれわれがリストアップしたのではなく、大学 ICT 推進協議会（AXIES）の RDM の提言書や政府文書等を読み解き、RDM にとってそれぞれ最低限必要な機能はこういった機能だというものをリストアップして、それらに基づいて RDM 基盤の開発を進めています。

図10は実際の画面です。GakuNin RDM はウェブアプリケーションです。ブラウザで、パソコンやスマートフォン、タブレットからもアクセスして利用することができます。ログインして、研究プロジェクトが始まると、研究者はシステム中で研究データを管理するポータルのようなものを簡単に立ち上げることができます。

そこに研究の概要やライセンスの情報、プロジェクト開始の日付や共著者の情報等を記入することができます。特徴的なのはマルチクラウドファイルマネージ



(図9)



(図10)

ャーで、ファイルを自分のデスクトップ上のシステムからドラッグ・アンド・ドロップでウェブ上に自由に保存できることです。保存するとファイルには自動的に内部に ID が振られます。ID 管理がされ、なおかつ日付も管理されて、誰がいつ保存したデータかが分かります。同じファイル名称であれば、ファイルのバージョン管理が可能です。研究者の「ファイルのバージョン管理はできるのですか」という要望にはお応えできていると思っています。

便利な機能としては、約 30 種類のファイルのプレビュー機能を持っています。おおよそオーソドックスな Office ドキュメント等はプレビューができるようになっています。他にも XML ファイルや、PDB ファイルを 3 次元可視化する機能も持っています。そして、研究プロジェクトを進める上で便利なのは、Wiki の機能も研究ポータルの数だけ立ち上げることができて共同研究で使っていただくことができます。あるいは、研究を進める上でディスカッションを行うコメント機能も作っています。

研究者が新しい研究ポータルを立ち上げ、研究データを管理していきます。デスクトップからファイルを GakuNin RDM 上にアップロードするのは非常に簡単な操作です。決して理工系だけが使うことを目指しているわけではなく、人文の先生でもすぐに使っていたらと思うと信じてこういったインターフェイスを用意しています。

ファイルのプレビュー機能では、ワードファイル、パワーポイント、エクセルシートを開くことができます。また、CSV だとこの場で編集することも可能です。ZIP ファイルも中を見られます。

Wiki は、複数人が同時編集できるような機能を持っています。入力すると、プロジェクト概要としてどういう研究データなのかという説明書きを書いたことができます。

バージョンコントロールの機能は、先ほどのパワーポイントを編集・保存して再アップロードするという流れですが、同じパワーポイントファイルをアップロ

ードするとファイルバージョンが変わり、個別のファイルごとに履歴として残ります。過去のバージョンにさかのぼってもう一度見ることもできますし、誰がそのファイルを変更したかということも克明に記録されます。こういったことも、研究データ管理を自動化していくという意味では非常に楽なのではないかと考えています。

プロジェクトが増えたときはタグ等を付けて、一部メタデータを付けて、プロジェクト単位で検索できるような仕組みになっています。

あとは来歴管理です。ファイルがどのようにアップロードされた後に変更・削除されたかを管理者がトレースできるので、研究代表者はそれをトレースして情報をダウンロードすることができます。

GakuNin RDM は最初から標準ストレージ（NII ストレージ）で 100GB 使えるのですが、ここにさらにパブリックなクラウド、あるいはプライベートクラウドをマウントして、学内で契約されている、あるいはオンプレミスで学内で立てているクラウドを接続して、同じクラウド上で利用することができます。やはり学内のシステムにデータを置いておきたい、しかしメタデータは研究データポータルとして GakuNin RDM の中で管理したいといった要望には応えていくことができていると考えています。

例えば Google ドライブのような代表的な商用クラウドをマウントして、外部のクラウドにも同じようにデータをデスクトップから保存できますし、クラウド間でデータを渡すことでバックアップなどの用途にも利用できるのではないかと考えています。

GakuNin RDM 中にアップロードしたファイルには、研究証跡保存機能が働きます。これは NII のサービスだけではなく、その時間に確かにそのファイルがあったということを、第三者機関の時刻認証局の発行するタイムスタンプによって証明するという仕組みです。実際の画面では、何か不正があったということがリストアップされ、管理者はそれを見て、〇〇さんのデータが疑わしいということを調べられる仕組みになって

います。

GakuNin RDM のストレージタイプとしては、GakuNin RDM の利用直後から、標準ストレージとして研究者 1 人当たり 100GB まで利用可能なストレージが準備されています。それだけでは不十分な場合は、研究プロジェクトや研究室ごとにご契約いただいたクラウドを接続する拡張ストレージがあります。十数個の外部クラウドサービスを接続して利用していただけます。クラウドストレージだけではなく、GitHub や Bitbucket といったソースコードリポジトリも接続できますし、一部、Figshare などの図表サービスとも接続することができます。

そして、今年度新たに作った機関ストレージという機能があります。これは NII 標準ストレージと置き換える形で、起動直後から使える自機関が持っているストレージをあらかじめ接続しておいて、利用者がログインしたら、自機関のクラウドがマウントされている状態で研究データ管理システムが使えるという仕組みです。直近では、代表的なクラウドプロバイダーのエンタープライズ級のクラウドサービス（実際には Dropbox Business を指しており、2020 年 4 月時点ではリリースされ公知になっている）に向けて開発を進めており、間もなくクラウドプロバイダーのサービスを GakuNin RDM のバックエンドとしてお使いいただけるようになります。そうしますと、学内で既に導入されている場合は、それに GakuNin RDM を被せる形で、クライアントソフトも全て流用しながら使っていただくことができるので、あまり無理がない形で研究データ管理を始めていただけるのではないかと考えています。システム管理者の立場では、あまり研究クラウドを増やされると困るということがあると思うので、その制御も機関ごとにコントロールすることができます。それを反映させて、自機関だけのストレージと RDM だけを作ることも可能です。

ここから少し話題が変わります。管理リポジトリは、当然、データ公開のために渡す機能の開発を進めています。リポジトリの標準的なプロトコルの次世代パー

ジョンに合わせていくということで、今年度、次年度と開発を進めていきますが、特に次年度、力を入れていくということになっています。先ほど紹介した GakuNin RDM のデスクトップは、主にクラウドストレージが標準的に持っているようなデスクトップクライアントを流用できる形を考えています。一つが、これも将来の形にはなりますが、JupyterHub という研究分析ツールの統合環境との連携もプロトタイプというレベルでは進めています。GakuNin RDM 中のデータを JupyterHub のデータ解析環境に送り、実験が終わって、プログラミングの実行が終わったら GakuNin RDM に戻ってきて、また自動的にデータ管理していきます。ここが研究者に向けた便利機能になっていくと考えています。

国内の学術機関でのユースケース

GakuNin RDM の実証実験を 1 年半ほど行っており、2020 年度の後半から本格運用に入っていこうとしています。現在、GakuNin RDM の実証実験参加機関は 18 機関です（図 11）。国内 16 機関と、海外ではマレーシアの 2 大学です。北海道大学では、情報基盤センターのパンフレットに RDM と共同利用のストレージを接続するといったことを書いてくださったり、附属図書館で学内に向けて利用や実証実験参加の呼び掛け等をしてくださっています。名古屋大学では、研究データ管理の体制づくりのベースとして利用していただいています。



(図 11)

経営統合の話では、北海道国立大学機構の3大学連携の中で、北見工業大学を中心に小樽商科大学、帯広畜産大学にも試していただける計画になっています。

京都大学では、2019年8月夏に20名程度の教員の方に集まっていただき、多人数で使った場合にどうなるかというワークショップを開催していただきました。

東京大学定量生命科学研究所では、もう二度と研究不正を起こさないという目的で、研究所内で出される全ての論文を研究倫理推進部門がチェックしています。その中心に GakuNin RDM を置いて利用していただいています。

他に大型研究プロジェクトとしては、文科省新学術領域で一部使っていただいています。

実証実験の参加方法

実証実験に関心を持たれた場合は、学術認証フェデレーションに参加していただく必要があるのですが、情報システム部門とご相談いただき、NII に連絡していただければ、すぐに使える状況になります(図12)。一番大変なのは、やはり学内調整だと聞いています。われわれが開発を進めているオープンサイエンス基盤研究センター(rcos-office@nii.ac.jp)までご連絡いただければと思います。

まとめ

NII では研究データ管理のためのサービスを2020年後半から提供していこうとしています。名前は

GakuNin RDM の導入方法

- ・学認フェデレーション参加のIdPでの接続 (必須)
 - ・ IdP接続前に関係者がログインして試用する方法も準備
- ・ストレージの準備 【技術面】
 - ・ (1) 標準ストレージ: NIIストレージ (研究者一人当り100GB)
 - ・ (2) 拡張ストレージ (オプション): 学内へ提供中の貴学ストレージを接続して実験
 - ・ (3) 機関ストレージ (オプション): 標準ストレージと学内ストレージを切替えて利用
- ・学内調整 【ポリシー面】
 - ・ 実証実験参加のための学内での合意形成
 - ・ 実験にご協力いただける学内の部局・研究プロジェクトへのご案内・ご紹介

(図 12)

GakuNin RDM で、研究推進と研究不正防止に非常に有効なものを簡単に学内や研究所内に導入していけると思います。実証実験参加機関に手を挙げてくだされば、われわれの担当者が行って使えるようにしますので、ぜひお声掛けいただければと思います。

●フロア 1 北里大学白金図書館の職員です。図書館としては、研究者は書誌データを researchmap に登録していますが、その生データを GakuNin RDM で管理するというイメージでしょうか。例えば researchmap の書誌を個人登録に委ねているため、研究者間や大学間で統一が図れていない部分がとても多いと思うのです。GakuNin RDM で生データの書誌を researchmap に提供することで書誌データの統一が図れると思うのですが、この先、GakuNin RDM と researchmap の連携はあるのでしょうか。

●込山 researchmap も有用なサービスだと思っています。彼らも API のバージョンアップを進めているので、逆に向こうから情報を一部頂いて私たちの方の書誌情報に取り込むなど、相補的にできることもあるかと思っています。現在、機能としては持っていませんが、将来的に何か researchmap 側から情報を取り込むようなことはあるのではないかと思います。

●フロア 2 国際共同研究などで例えば大きな加速器のデータを入れようというのではなく、これをもう少し運営サイドに使えないかと思ったのですが、外国人の研究者が自由に使えるようなものになるのでしょうか。

●込山 国際共同研究における運営や教育にお使いいただいても研究推進の目的に合致しますので、お使いいただけます。システム的には既に海外からも同じプラットフォームに入っていていただいてデータ共有の空間

を作れるようになっています。ただ、安全保障貿易に係る機微技術管理の規定、あるいはヨーロッパ国籍の方やヨーロッパからアクセスする方は EU 一般データ保護規則（GDPR）に従うなど、法的な問題は残っています。システム的には今すぐにでも可能ですが、大切なのは研究データ管理に関するルールをどうつくるかということです。

●フロア 2 そのルールはどこかで議論されているのですか。

●込山 研究所あるいは大学ごとにそういうルールがつくられて、共同研究の中でどうデータ共有をすればよいかというポリシーが組織ごとにできてくるのではないかと考えています。もちろん、大きなコンソーシアム等で議論していく必要はあると思います。

実験レベルの話ですが、一つこぼれ話があります。学内のストレージで共同研究をしてもよいと認めている、一番ポリシーが緩い大学のストレージを使い、みんなでそこに預けるということが過去の RDM の実験では起こっていました。

第3回 SPARC Japan セミナー2019

「実践 研究データ管理」

JPCOAR による研究データマネジメント人材の育成と 研究データに関する取組について

結城 憲司

(北海道大学附属図書館/JPCOAR 研究データ作業部会)

講演要旨



本報告では、オープンアクセスリポジトリ推進協会（JPCOAR）が取り組んできた研究データマネジメント人材の育成（RDM 教材、研修）の他、研究データに関する取組（データベースレスキュープロジェクト、RDM 事例形成プロジェクト、JPCOAR スキーマ等）について紹介する。その他、大学 ICT 推進協議会研究データマネジメント部会（AXIES-RDM 部会）や国立大学図書館協会オープンアクセス委員会における研究データに関する諸活動についても触れる。



結城 憲司

北海道大学附属図書館研究支援課長。2018年4月から現職。2018年4月からJPCOAR運営委員会委員、2019年4月からJPCOAR研究データ作業部会主査。その他、大学ICT推進協議会研究データマネジメント部会（AXIES-RDM部会）に参加。国立大学図書館協会（JANUL）オープンアクセス委員会には事務局として関与。

今回は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主査という立場で、JPCOAR が行ってきた人材育成と研究データに関する幾つかの取り組みについてご説明します。また、大学 ICT 推進協議会（AXIES）の RDM 部会の活動にも関与しており、国立大学図書館協会の OA 委員会でも研究データの取り組みを行っていますので、それについてもお話しします。

JPCOAR の取り組み

JPCOAR は、国公私立の図書館、主に大学図書館あるいはそれ以外の機関の図書館等が加盟している団体です。設立は2016年7月で、2020年1月14日時点で

は629機関が参加しています。主に大学等が使っている JAIRO Cloud という機関リポジトリのシステムを国立情報学研究所（NII）と共同運営しています。

2019年から2021年にかけて、「オープンアクセスリポジトリ戦略」として五つの戦略を立てています。その1番目の「オープンサイエンスの推進に寄与するため、研究データの公開、流通に関する先導的な取組を行う」が今回に関わるものです。こちらは私が担当している研究データ作業部会が行っています。

内閣府から2018年6月に出された「統合イノベーション戦略」は、オープン・アンド・クローズ戦略に基づき幾つかの提案がされており、それについて JPCOAR は図書館としてどのようなことができるかと

ということで、三つの見解を述べています。まずは、JAIRO Cloud により研究データの公開に取り組めるのではないかと、また、これまで整備してきた JPCOAR スキーマ（次世代のメタデータのスキーマ）により研究データ等に対応していくことができること、それから、RDM トレーニングツールの開発・普及・利用促進についてもやっていけるのではないかとということです。

人材育成の取り組みとしては、教材開発を行い、三つの教材を作ってきました。最初に作った教材が「RDM トレーニングツール」です。こちらは研究データ管理に関する基礎的な知識を習得するためのツールとなっています。図1がその内容ですが、結構詰め込んだ形です。それまで日本語の教材がなかったので、初めて作ったツールということで、いろいろ弱点はあったのではないかと思います。

それらのツールをもう少し理解しやすくしようと作ったのが、「オープンサイエンス時代の研究データ管理」です。MOOC 用に公開された動画付きのトレーニングツールで、4週にわたって受講する形で内容の再編を行っています（図2）。

そして、「研究データ管理サービスの設計と実践」という研究支援者をターゲットとした教材も公開しました。それまでは、誰が対象者なのかよく分からない教材だというコメントを頂いていたので、そのあたりに力点を置いて作っています。特徴は、研究データのライフサイクルに基づいて教材を構成していることです。例えば図書館の職員なのか、URA の方なのか、研究推進の方なのか、情報系の方なのか、それぞれに

立場によってどこで研究者に関与できるのかという視点で教材が理解できるようになったのではないかと思います。図3がその中身です。研究前の支援、研究中の支援、研究後の支援、日常的な支援ということで整理しました。

これまでに作った RDM 教材は、土木学会の e ラーニングに収録されたり、「統合イノベーション戦略」や「研究データリポジトリ整備・運用ガイドライン」の中で取り上げられたりしています。

ここからは、今年度行ってきたことになります。まずは学認 LMS という NII が開発しているラーニング・マネジメント・システムで、二つの RDM 教材を提供し、昨年9月から12月にかけて22機関が参加して試験運用を行いました。その報告会を12月に開催し、いろいろなご意見を頂いたので、今後反映させていきたいと思っています。ただ、この学認 LMS のシステムは開発に若干時間がかかりそうということで、実際の運用はもう少し先になりそうです。また、参加機関による研究データ管理に関する取り組みの情報共有を行い、他館の状況を把握できたという別のメリッ

「オープンサイエンス時代の研究データ管理」の構成	
章	内容
第1週：研究データ管理とは	- 本講座の背景とねらい - 研究データ管理の重要性が増している背景 - 研究データ管理の定義 - 効果的なデータ管理に必要なデータ管理計画 - データ管理計画をめぐって国内外の動向 - データ管理計画の実践
第2週：保存・共有・文書化	- 研究データの保存や長期保存・共有・再利用 - 研究データの組織的な管理 - データを文書にまとめる
第3週：メタデータ・法倫理的問題	- データにメタデータを付与する - データの引用 - 研究データをめぐる著作権・再利用促進のためのライセンス - 研究不正・研究倫理 - センシティブデータ
第4週：運用に向けて	- ポリシーとは - 研究データ管理サービスとは？ - サービスの設計～学習者が自機関での研究データ管理サービスを構築するためのステップ

(図2)

「RDMトレーニングツール」の構成	
章	内容
第1章 導入	研究データ管理の重要性が増している背景や、研究データ、研究データ管理の定義について学ぶ。
第2章 データ管理計画 (DMP)	効果的なデータ管理に欠かせないデータ管理計画に関し、作成義務化の動向や、その構成要素について学ぶ。
第3章 保存と共有	研究データの研究期間中の保存や長期保存に関する留意点について学ぶ。また、研究データの共有に関して、その意義や検討すべき点、共有方法について学ぶ。
第4章 組織化、文書化、メタデータ作成	研究データを長期的に管理・活用するために欠かせない、一定のルールに則ったデータの組織化や、データについて説明する文章やメタデータの作成について学ぶ。
第5章 法・倫理的問題	研究データをめぐる著作権や、再利用を促進するためのライセンスの仕組みについて学ぶ。あわせて、センシティブデータを取り扱う上での留意点や、研究倫理についても学ぶ。
第6章 ポリシー	国や助成団体、機関、学会等が、研究データの保存や共有を求めるポリシーを策定する例が増えており、これらの動向及びポリシー策定の要件について学ぶ。
第7章 研究データ管理サービスの検討	学習者が自機関での研究データ管理サービスを構築していくためのステップを学ぶ。

(図1)

「研究データ管理サービスの設計と実践」の構成	
章	内容
第1章：序論	研究データ管理に関する基礎的な知識を学ぶ
第2章：サービス設計	自機関の実情に合ったサービスを設計するために必要なことを学ぶ
第3章：研究前の支援	研究者が研究を開始する前段階で、どのような支援が求められているかに学ぶ
第4章：研究中の支援	研究者が研究を実施している最中に求められる支援について学ぶ
第5章：研究後の支援	研究終了後に、得られたデータを公開することについて学ぶ
第6章：日常的な支援	日常的な研究データ管理サービスについて学ぶ

(図3)

トもあったと思っています。

もう一つ、RDM 教材の取り組みとして、若手研究者向けの教材作成を行っています（図 4）。受講した方は分かると思いますが、既存の教材は受講にかなり長い時間がかかる教材で、あれを研究者の皆さんに受けてくれというのは少し無理があると思います。また、研究者の皆さんであれば、ピンポイントで知りたいことが学べればいいわけです。例えば研究データの引用をどうすればいいのか、データのマネジメント計画をどのように作ればいいのかということをピンポイントで教えてくれる教材がないと先生方に受けてくれとは言えないので、もう少し分割して若手研究者向けに作ってみようという取り組みをしています。千葉大学の方で先行して開発中の教材があるということで、そちらを参考に作っています。今年 3 月には試行版を作成し、来年度中にブラッシュアップして公開できるようにしていきたいと思っています。

また、JPCOAR では研修事業を行っています（図 5）。こちらは JPCOAR の参加機関の方が対象です。来年度の計画としては、基礎研修としてオープンアクセス新任担当者研修（仮称）を行う予定です。これまでも JPCOAR では機関リポジトリ新任担当者研修を行っていましたが、もう少し幅広に内容を盛り込み、年に 2 回程度開催したいと思っています。それから専門研修としては、今年 10 月から新 JAIRO Cloud が稼動することになっているので、操作説明会を秋以降、全国 5 会場程度で開催する予定です。JPCOAR セミナーは、来年開催できるかどうか不確かですので、参考程度にご覧ください。

若手研究者向けの教材作成

- 今回の教材のターゲットは若手研究者
- 既存の教材は受講に時間がかかりすぎ、研究者は利用しにくい
- 既存のRDM教材を分割して（マイクロコンテンツ化）、若手研究者向けに再構成
- 先行している千葉大学で開発中の教材等も参考に
- 2019年度の目標は
 - 現在の支援者向け教材をベースに、研究者が行うタスクと支援者が行うタスクを仕分ける
 - 他の「研究者向け教材」から、現在の支援者向け教材に不足しているコンテンツを洗い出す
- 2020年3月末までに試行版を作成予定

(図 4)

JPCOAR の研究データに関する取り組みを幾つか紹介します。図 6 の「失われていく過去の研究データ」は、研究データ作業部会のメンバーである京都大学の天野さんが調査したものです。この調査によると、科研費等の助成を受けて作成されるデータベースのうち、半分ぐらいが適切に継続管理されていないということです。既に失われてしまっているようなデータもあり、そういったデータを救済する必要があるのではないかとこのコンセプトの下に考えているのが、データベースレスキュープロジェクトです。管理していた研究者がいなくなると、どうしても継続が困難になるということがあります。大学によってはデジタルアーカイブが受け皿になる可能性もありますが、機関リポジトリがその受け皿になる可能性を模索するプロジェクトです。2019 年度の活動として、作業部会員の所属する機関においてこのプロジェクトを進めました。ただ、実際には公開まで持っていくには少々時間がかかるというのが正直なところ。こちらからアプローチしても、そのデータがどこかのハードディスクにはあるのだけれどもどこかよく分からず、なかなかデータの

2020年度のJPCOARの研修（計画中）

- JPCOAR参加機関対象の研修
- 基礎研修
 - オープンアクセス新任担当者研修（仮称）
 - これまで開催してきた「機関リポジトリ新任担当者研修」をリニューアル
 - 学術情報流通の概論、OA出版、RDM、オープンサイエンスの動向等の内容を盛り込む
 - 年2回開催予定
- 専門研修（他にも検討中）
 - JPCOARセミナー
 - JPCOARスキーマ、OAI-PMH、IIIF、DOI、ORCID等の学術コミュニケーション要素技術の解説、夏頃開催予定
 - 新JAIRO Cloud操作説明会
 - 秋以降、全国5会場程度開催予定

(図 5)

失われゆく過去の研究データ

URLが生きている／移動している

メンテナンス状態

安定 不安 不可 未オープン (CDのみ、有料、要登録)

「KAKEN: 科学研究費助成事業データベース」で平成21～23年度採択分（文系）を調査した結果

・2016年及び2019年に調査（京都大学・天野氏による調査）

科研費の助成を受けて作成されたデータベースの半分が適切に継続管理されていない

(図 6)

救済には至らないケースもあります。

図 7 は過去に取り組んだ事例です。「矢内原忠雄文庫植民地関係資料データベース」は琉球大学の事例です。これは 2017 年度に JPCOAR の研究データタスクフォース（当時）として取り組んで公開されたものです。「北の息吹：日本の魅力的なワイルドフラワー500種」は本学の事例です。部局のホームページで公開していた研究データが、その先生がいらっしゃらなくなる予定のため公開の維持が難しいということで図書館に相談に來られ、機関リポジトリが受け皿として公開した事例です。「ピラネージ画像データベース」は機関リポジトリというよりはデジタルアーカイブでレスキューした、東京大学の事例です。

次に、RDM 事例形成プロジェクトの取り組みをご紹介します。こちらは JPCOAR と、後で述べる AXIES-RDM 部会との連携プロジェクトになります。目的は RDM に関する事例を実際に積み重ねていくことで、2019 年度に参加したい機関を募集し、2020 年、2021 年と 2 年間かけて蓄積していくプロジェクトです。前に述べた教材でも、日本国内の事例が少なく海外の事例ばかりではないかということをおっしゃる方がいらっしゃいました。まだ大学の事例が多くないので、これを積み上げていき、次に続く大学が参考にできるようなものを公開していきたいと考えています。JPCOAR としては、研究データの公開、発信等の部分を中心に担当する予定です。

もう一つの JPCOAR の取り組みとして、「JPCOAR スキーマ」があります（図 8）。現在、日本の機関リポジトリでは「junii2」という、どういう基準でデー

タを記述するのかということを決めているメタデータスキーマがあるのですが、その次世代のものを開発して、2017 年 10 月に公表しています。目的はオープンサイエンスおよびオープンアクセス方針への対応で、研究データに対応したメタデータスキーマとなっています。これまでの「junii2」では、論文に必要なメタデータはあったのですが、研究データとなるともう少し幅広い項目が必要になってきますので、それらを追加したり、国際的な相互運用の向上やメタデータ構造の修正を行うなどしています。JPCOAR スキーマへの対応を始めており、ジャパンリンクセンター、CiNii Articles、医中誌、新 IRDB、次に今年 10 月に公開する新 JAIRO Cloud という形で公開が順次進んでいます。今後は JPCOAR スキーマが主流に変わっていくのではないかと考えています。

その構造を表したのが図 9 の図です。各国内の機関リポジトリから IRDB の方に JPCOAR スキーマでデータが送られ、各種サービスに提供されて、各利用者が利用できるようになります。

JPCOAR
オープンアクセスリポジトリ推進協会

レスキュー事例

- 「矢内原忠雄文庫植民地関係資料画像データベース」（琉球大学附属図書館）※2017年度JPCOARレスキュー事例
<http://manwe.lib.u-ryukyu.ac.jp/yanaihara/>
- 「北の息吹：日本の魅力的なワイルドフラワー500種」（北海道大学附属図書館）
<https://eprints.lib.hokudai.ac.jp/dspace/html/2115/68332/Blessings-from-the-Northern-Island.html>
- 「ピラネージ画像データベース」（東京大学学術資産アーカイブ化推進室）※デジタルアーカイブでレスキューした事例
<https://iif.dl.itc.u-tokyo.ac.jp/repo/s/piranesi/page/home>

23

(図 7)

JPCOAR
オープンアクセスリポジトリ推進協会

JPCOARスキーマ (1)

「JPCOARスキーマ」は、日本の機関リポジトリで標準的な「junii2」に代わる次世代のメタデータ規格としてJPCOARが策定し、2017年10月に公表

- 目的：日本の機関リポジトリのメタデータの国際的な相互運用性を向上させ、日本の学術的成果の円滑な流通を図ること
- JPCOARスキーマ策定の基本方針
 - オープンサイエンスおよびオープンアクセス方針への対応 ⇒ 研究データのメタデータスキーマ策定
 - 識別子の拡充に伴うメタデータ構造の修正
 - 国際的な相互運用性の向上
- JPCOARスキーマへの対応状況
 - 2019年3月 ジャパンリンクセンター (JaLC) 、 CiNii Articles、医中誌
 - 2019年4月 新IRDB (学術機関リポジトリデータベース)
 - 2020年10月 新JAIRO Cloud (予定)
 - (対応中) CiNii Research

(参考) 「JPCOARスキーマガイドライン」 <https://schema.irdb.nii.ac.jp/>
(参考) 「JPCOARスキーマの策定：日本の学術成果の円滑な流通を目指して」 <https://doi.org/10.1211/johokanri.49.719>

25

(図 8)

JPCOAR
オープンアクセスリポジトリ推進協会

JPCOARスキーマ (2)

①各機関リポジトリからIRDBにメタデータをハーベスティング ②各種のサービスにデータを流通

国内の機関リポジトリ → IRDB → JPCOAR スキーマ → 各種のサービスにデータを流通

図は、以下の資料からの引用
片岡朋子「JPCOARスキーマが支えるデータ引用」、Japan Open Science Summit 2019
https://openlinkcenter.org/rub/dsc/pas2019_rdc_05.pdf

26

(図 9)

AXIES-RDM 部会の取り組み

次に、AXIES の RDM 部会の取り組みについてお話しします。この部会のメンバーである NII の船守先生から、こちらの団体は情報系のセンターの教員が中心の組織なので図書館の人も入ってもらいたいという話があり、私が入らせていただいて取り組みに協力しています。

AXIES は情報系のセンターの教員を中心に運営されている組織で、その中に研究データマネジメント (RDM) 部会があります。主な取り組みとしては、2019 年 5 月に「学術機関における研究データ管理に関する提言」を公開しており、2020 年 1 月に「大学における研究データ管理に関するアンケート (雛形)」の公開をしています。

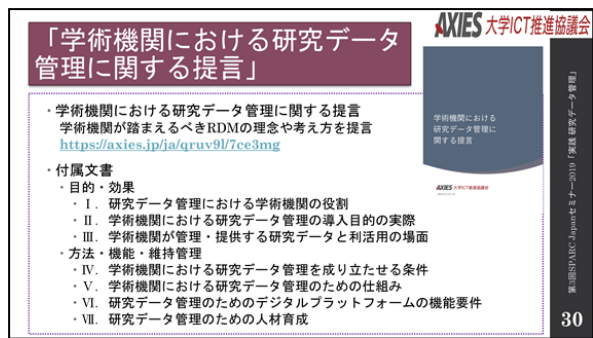
「学術機関における研究データ管理に関する提言」については、図 10 にあるような資料が公開されています。研究データの管理・公開について非常に具体的に書かれているので、いろいろ疑問に思っていた部分が分かりやすくなっていると思います。例えば、研究データを誰が管理するのかという話については、それは研究者だということが明確に書かれており、それを管理するための環境等を機関としてきちんと提供しないといけないということが書かれています。管理という言葉にすると、教員は上から全部管理されるのかと拒否反応を示す場合が多いので、その辺は丁寧に説明する必要があるのではないかと思います。また、公開に関してもいろいろ誤解がありますので、その辺でも参考になる資料ではないかと思います。

もう一つの「大学における研究データ管理に関する

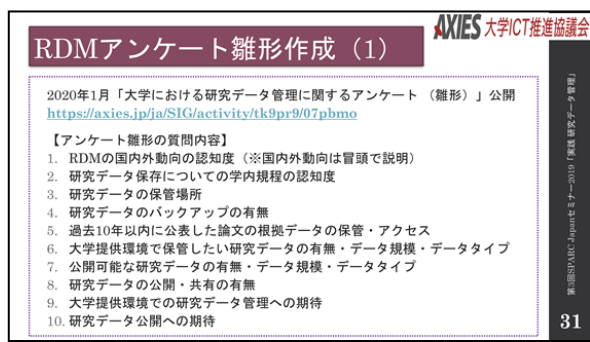
アンケート (雛形)」の作成が、私がメインで協力している取り組みです (図 11)。つい先日、1 月下旬に雛形が公開されました。質問は 10 問程度と、割とシンプルにまとめています。これ以上質問数が増えると回答率が下がるという意見があり、必要最小限のものとして 10 問に絞っています。こちらは元々、名古屋大学で実施されたものを多少修正して、幾つかの大学でアンケートを実施させていただきました。

2019 年 8 月に雛形案を作成、2019 年 10~12 月に京都大学、広島大学、大阪府立大学、本学でアンケートを実施、2019 年 12 月に福岡で開催された AXIES 年次大会でアンケートの実施結果を報告、2020 年 1 月に雛形公開、2020 年 3 月に全体的な結果を情報処理学会 IOT 研究会で報告というスケジュールで取り組んでいます。

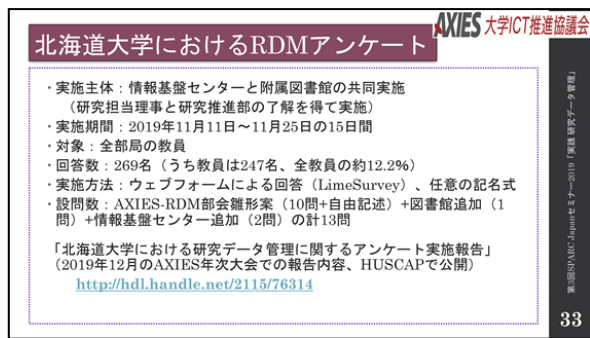
北海道大学で実施した事例が図 12 です。情報基盤センターと附属図書館が共同で、研究担当理事や研究推進部の方にも話を通した上で実施しています。実施期間は昨年 11 月 11 日~11 月 25 日の 15 日間です。全部局の教員を対象としており、回答数は 269 名 (うち



(図 10)



(図 11)



(図 12)

教員は 247 名、回収率は全教員の 12.2%) でした。ウェブフォームで実施したのですが、AXIES-RDM 部会の雛形の 10 問の他に、図書館で 1 問、情報基盤センターで 2 問追加しました。

特徴的なのは、任意の記名式にしたことです。なぜかというと、これを入り口としたいと思っていたからです。名前を挙げてくれる方には個別にお問い合わせすることもありますということを明記した上で、来年度インタビュー調査を行いたいと思っています。その中で研究データに関心をお持ちの先生方に話をし、実際に研究データの公開等につなげていければと思っています。アンケートの集計結果については本学の機関リポジトリ HUSCAP で公開していますので、そちらを見ていただければと思います。

簡単に説明しますと、研究データの 10 年保存ルールについては、65%の方が知っていると回答しており、割と認知度が高かったです。科研費の申請をするときに研究倫理教材を受けさせられるので、その中で認知度が上がっているのではないかと思います。

研究データの保管場所については複数回答を認めているのですが、一番多いのはやはり個人のパソコンで 88%、研究室のサーバーやストレージが 49%でした。

機関による RDM 環境が必要だと回答した方は 62%でした。それに対して、研究データの公開ニーズについては 32%の方しか必要だと回答しておらず、やはり公開に関しては少し抵抗感が強いということが分かりました。これに基づいて、個別の先生にいろいろ聞いていきたいと思っています。

国大図協 OA 委員会の取り組み

続いて、国立大学図書館協会（国大図協）のオープンアクセス委員会の取り組みを説明します。国大図協は、全国 86 の国立大学、放送大学、大学共同利用機関等の図書館を会員とする団体です。ここでもオープンサイエンスや研究データ管理に関する取り組みを行っており、2019 年 3 月に取り組みの文書を公表していますし、6 月の岡山の総会研究集会において、「大

学図書館のオープンサイエンスに向けた取り組み」というテーマで報告および議論を行っています。そして 2020 年 1 月に「オープンサイエンスの推進に向けた協会の行動計画」を公表しました。この中で、まず取り組むこととして、アドボカシー活動の推進、人材の育成、先導的事業の推進について、短期と中期に分けて計画しています。

国大図協の中にオープンアクセス委員会というものがああります。こちらはオープンアクセスだけではなく、オープンサイエンスの研究データのことを扱っているので、活動を少し紹介します。

昨年度、機関リポジトリの再定義を行いました。「統合イノベーション戦略」の中で機関リポジトリという言葉が急に出てきたわけですが、当初、大学図書館の皆さんにとっては、機関リポジトリはそれぞれの大学の研究者の皆さんの論文、あるいはそれぞれの大学で発行している紀要、あるいは博士論文を公開する場所だという認識でした。しかし、「統合イノベーション戦略」で求められているのはもっと幅広いもので、通常、商業流通されている学術文献、あるいは大学の活動で生み出される学術文献以外に、研究データやデジタル化された研究資料があります。これまでデジタルアーカイブとして図書館がいろいろ公開してきた画像や資料も機関リポジトリの中にも含まれてもいいのではないかと考えられます。その辺の線引きはいろいろシステム的な問題があるのですぐには統一できないと思いますが、将来的には分け隔てなく収録していく必要があるのではないかと考えるの下に、機関リポジトリの再定義を行いました。

もう一つ、昨年取り組んだ事項として、「オープンサイエンスに向けての国立大学図書館が担う具体的な役割」というものがあります。内容としては、①各会員館の役割、②会員館共通の取り組み、③協会としての取り組みという三つのレベルを設けて、それぞれどんなことに取り組んでいくのかということを列記しました。特に①会員館の役割については、各大学の事情や規模が違うので、こういうことは取り組めるのでは

ないだろうかということを一覧した上で、それを取捨選択し、それぞれの大学で取り組んでいただく形にしております。

その背景を説明したのが図 13 の図です。大学の図書館の皆さんに、もっとオープンサイエンスに向けて取り組んでほしい、こう変わっていくのだということを図示しています。旧来は大学図書館という資料の収集や保存、提供など紙ベースのものを中心に行っていて、その後、電子ジャーナルが出てきたら電子ジャーナルに対応していたと思います。ただ、2005 年ごろから、それだけでは駄目で、オープンアクセスの推進や機関リポジトリの取り組みを行う必要があるだろうということで、それぞれの大学の研究成果を公開する取り組みを行ってきました。そして今度は、研究データのことも段階を踏んで行っていく流れとなっています。

今年度の取り組みについて簡単に説明します。一つは、「研究者に対する研究データ調査の項目リスト（案）」の作成です。大学図書館がこのオープンサイエンスや研究データ管理に取り組む上で何ができるかということは、このオープンアクセス委員会の中で議論になりました。研究者の皆さんで、オープンサイエンスにそこまで取り組まれている方はまだ多くはないと思いますが、まずはその実態を知ろうということです。方法としては、①アンケート調査、②個別インタビュー調査という2段階を想定しています。以前、オープンアクセスをやっている時代にもこのような取り組みをやってきたので、それを参考に組み立てています。



(图 13)

①アンケート調査については、AXIES の作ったアンケートの雛形を利用推奨しています。こちらを使って次の個別インタビュー調査につなげるような形で名前をお聞きした上で、研究データに関心を持っているような研究者を捕捉します。②個別インタビュー調査については、AXIES の雛形のようなものがないので、オープンアクセス委員会の方で雛形を用意して公開する予定です。

もう一つ必要ではないかと議論になっているのが、教員向けに研究データのオープン化のメリットを示すような資料です。インタビュー調査に行ったときに必要ではないかということです。そこで、「研究データのオープン化の具体例とそのメリット（仮称）（案）」を作っています。ただ、論文を公開することのメリットは割とスムーズに説明できるのですが、研究データを公開することのメリットとしては、引用されることや、科学のさらなる発展につながることも言えると思いますけれども、実感としてメリットがあるということとをどれだけ研究者の皆さんに伝えられるかというのは難しいところです。今たたき台を作っているところですが、難航しています。また、もう一つこれでやりたいと思っているのは、誤解を解くことです。先ほど言いましたように、研究データ管理というと、管理されてしまうのかとか、オープンというと、全部オープンにしなければいけないのかなどと思っている先生もいらっしゃるので、そのあたりの誤解も解きたいと思っています。こちらについては3月末までにまとめる予定です。

今年度のこの二つの取り組みについては、公開までにはもう少し時間がかかるかと思います。公開しましたら国大図協でお知らせしたいと思っています。

●林 JPCOAR の取り組みをいろいろご紹介いただきました。私は国の研究所にいますのでこれに入っているのですが、JPCOAR の会員機関として国の研究機関は

どのくらい入っているのでしょうか。

●**結城** 機関数は把握していませんが、作業部会にも研究所の方に何人か参加していただいています。どうしても大学を主体として話をしてしまうので、大学とは違った視点でいろいろお話を頂けるのは非常にありがたいです。特に研究と近いところにいらっしゃる人が多いので、大変参考になる意見を頂いています。

第3回 SPARC Japan セミナー2019

「実践 研究データ管理」

パネルディスカッション

- 林 賢紀 (国際農林水産業研究センター)
- 八塚 茂 (バイオサイエンスデータベースセンター)
- 山田 一作 (野口研究所)
- 熊崎 由衣 (日本原子力研究開発機構)
- 込山 悠介 (国立情報学研究所)
- 結城 憲司 (北海道大学附属図書館/JPCOAR 研究データ作業部会)



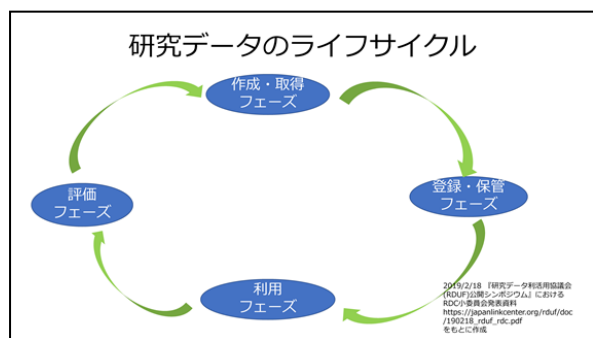
●八塚 初めに私の方から代表的に質問をさせていただきます。その後に議論をしていただき、その他に皆さまから頂いた質問についてパネリストの方に答えていただければと思います。

まず自己紹介を簡単にします。私は元々IT のエンジニアをしていましたが、現在はバイオサイエンスデータベースセンターにいます。担当しているのは生命科学のリポジトリの運営で、システムの運用からデータのキュレーションまで何でもやっています。いろいろなご縁があり、SPARC Japan セミナーのワーキンググループに参加しています。

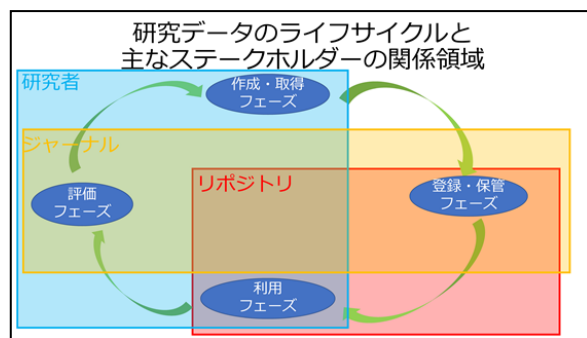
登壇者の皆さまにいろいろなライフサイクルの図をお出しいただいたと思いますが、それを非常に単純化し、データ流通に重きを置いた形で四つのライフサイクルの図を描いてみました (図 1)。一番上に研究の

現場でデータを作成したり、そこから取得して作ったりするフェーズがあります。そして、そのデータを登録したり、保管したりするフェーズがあります。それから、その保管されているデータを利用するフェーズがあります。そして評価のフェーズは、例えば代表的なものとしては、論文にこのデータを使うという引用のシーンを想定しています。

このフェーズに対して、主なステークホルダーがどう関わっているかを示したのが図 2 です。研究者に關していうと、データの作成、利用、それから論文等で評価するということに関わってきます。私どもリポジトリは、データの登録・保管、それから、利用者の方に提供するというところで関わっています。ジャーナルは二つにまたがって書いています。当然データを評価する場がジャーナルであったりしますし、最近で



(図 1)



(図 2)

はデータの公開を要求するジャーナルは増えています。しかも、こういうリポジトリで公開しなければならないという条件を決めているところが結構増えています。そういう意味でいうと、ジャーナルも登録・保管に大きく関わってきているのではないかと考えています。

では、このデータのライフサイクルにどういった課題があるのでしょうか（図 3）。作成・取得フェーズに関していうと、リポジトリに登録するまでは、やはり自分でデータを管理する必要があります。GakuNin RDM を使うとその辺がうまくできるのではないかと思います。いずれにしても、研究者自身がそこを管理しておく必要があります。

それから、DMP もだいぶ広まってきたとはいえ、「何ですか？ それ」という反応はまだあるような気がします。そして、リポジトリに預ける段階になると、メタデータの作成は一体誰がするのかという問題が出てきます。そもそも、信頼できるリポジトリとは何かが分からないということがあります。リポジトリとしては、ストレージやインフラを持続して持つていく必要があります。

それから、きれいなデータばかり出てくればうれしいのですが、フォーマットが整っていないものがあり、そういったものの整備も必要になります。

そして今度は利用者から見ると、欲しいデータがなかなか見つからない、どうやって探すのかという問題が出てきます。

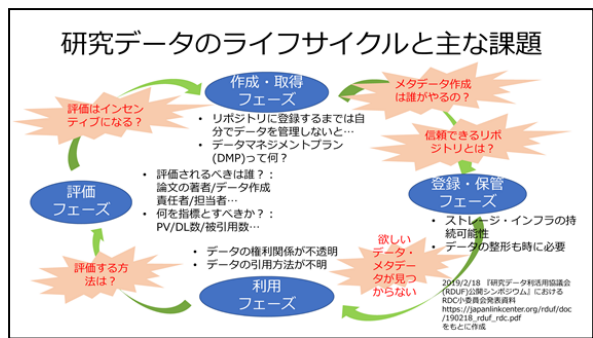
また、いざ利用してみると、データは手に入ったのだけれども使っていないかどうかの権利関係がよく分からないというものがたくさんあります。それから、デ

ータを引用したいけれども、どうやって引用すればいいか、URL を書けばいいのか、タイムスタンプを書けばいいのかなど、いろいろ分からないことがあります。

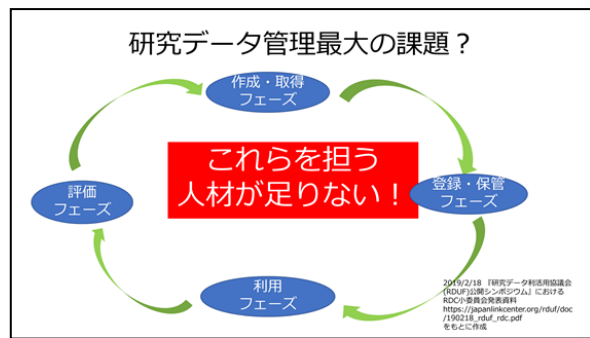
そして、評価です。論文を引用すればいいのですが、それは論文の著者が評価されるべきなのか、作成した人が評価されるべきなのか。また、論文に書かれる以外にも、リポジトリの方から見ると、PV なのか、ダウンロードなのか、何を評価指標として出せばいいのかということが問題になります。そして、高い評価を受けたとして、それが次の研究予算のアップにつながるのかというインセンティブの問題も出てくるかと思っています。

これは後で皆さんに伺いたいのですが、いろいろな課題がある中で私が一番の問題だと思っているのは、そもそもこの辺に関わる人が圧倒的に少ないということです（図 4）。

そういう問題意識を出した上で、パネリストの皆さんに質問させていただきたいことが 3 点あります。1 点目は、皆さまの立場から見て、研究データの管理における最大の課題とは一体何でしょうか。2 点目は、この会場にいらっしゃる方、あるいは中継を見ていただいている方を含めて、誰がどのような役割を果たせるとお考えでしょうか。そして 3 点目は、先ほども幾つかのステークホルダーを図示しましたが、他のステークホルダーの方に理解していただきたい、これは知っていただきたいということがあれば、せっかくでするのでここで述べていただければと思います。



(図 3)



(図 4)

●林 頂いている質問としては、これは特に山田様宛てだと思われそうですが、「自分が大学所属の研究者だしたら、冒頭の名古屋大学や北海道大学の発表を踏まえて、研究データ管理共有・公開において、どこが楽だったり、期待できたり、大変そうか、あるいは自分たちの取り組みと連携できるかについてコメントしてほしい」というリクエストが来ています。異なるステークホルダーというところで何かあればコメントを頂ければ幸いです。

●山田 自分が研究者だったとして、先ほど八塚さんも人が足りないとおっしゃっていたように、データを公開する上でサポートしてくれる人がいると楽になりそうというところはあると思います。今は分野のリポジトリの立場ですが、機関リポジトリでも大変そうなのに、維持・管理に加えてサポートが不意にやってくることが多くて、それが続くと他の仕事が回らなくなってしまうことがあります。

連携という意味では、分野のリポジトリ間の連携は私の方ではいろいろ進めているのですが、機関リポジトリをまたいだ連携なども、データとしてはつながることに関連性が出てくるので、いろいろな意味で有益だと思います。ただ、その機会があまりないので、そういう機会をつくることも有益ではないかと思っています。研究者からすると、そういう形でいろいろなデータがつながっていくと、自分自身の将来的にも有用になるのではないかと思います。

●熊崎 運用はこれからなので、連携とまではいかないかもしれませんが、例えば保管は GakuNin RDM を使わせていただくという方法があると思います。データが順調に進んで公開できたとして、使われないとあまりその先の発展につながらないと思います。きちんとメタデータを付けて、NII Research Data Cloud なり、他の分野別のデータリポジトリなり、原子力機構から公開するとしても、何らかの別の枠組みで流れていくようなデータの連携の仕組みをうまくつくことで、

研究者の方々にデータを公開していただいたその先につながっていくのではないかと、そうなることを期待します。

●込山 分野間の連携のための連携の仕組みをつくるために、われわれ NII はシステムの連携をきちんとしてお届けし、データ管理のレベルから公開に渡していくところをつくっていく。そうすれば、既存のリポジトリの仕組みでもうまく連携して、国際的な学術情報の検索の仕組みに乗ることができ、他の分野の研究データがきちんと探せるようになる。そういう連携を生むために、まず私たちはその下のシステムの連携をきちんとするということかと思っています。

●結城 順番にお答えします。1 点目は、研究データにおける最大の課題は何かということですが、大学において研究データ管理にはいろいろな部署が関わってきます。その間の協調、仕事の分担が非常に難しく、どこが音頭を取るのかが押し付け合いになりそうとあります。具体的に言いますと、私は図書館にいて、これまで研究成果の公開をしてきたので、研究データの公開については、それが使われるのであれば、もしかしたら図書館の資料の一環なのかもしれないということで理解できます。しかし、それ以外に研究データを収める基盤となる情報系のセンターがありますし、科研費や外部資金等の申請を行う研究推進の部門等があります。それから、先生方のお手伝いをしている URA という職種の皆さんもいらっしゃいます。そのあたりがどのように連携して、この研究データ管理を進めていくのか。最初にお話があった名古屋大学ではトップダウンで話が進められていて、そういう大学は取り組みやすいのですが、上の方が動いても、下が動くかというのはまた別の話です。現場レベルでどこまで仕事につなげていけるかということがあると思っています。

もう一つの課題は、研究者の皆さんをどう動かすかです。実際に研究をされるのは研究者の皆さんなので、

その方々にとってこれが非常にやりやすい形にならない限りは、実質的なものとはみなされないのではないかと考えています。

2 点目は、セミナーに参加して、私は図書館の職員なので図書館の立場でどういうことに取り組めるかということです。図書館としては研究データの公開、あるいは、もしかしたら利用の部分に関わってくるかと思っていますが、まずは研究者の皆さんが研究データをどのように扱っているかを知ることから始まるのではないかと考えています。その中で研究データの公開に関わるお手伝いができるのであれば、具体的に資源として持っている機関リポジトリ等を使ってどのように公開を進めていくお手伝いをしたら先生方が助かるのかを知って進めていけたらと思っています。その後、リポジトリを研究データに対応させていくためにどのようにすればいいのかということをやっているのではないかと考えています。

大学で研究データポリシーをつくるのは、なかなか難しいです。いろいろな学部があって、学生や院生がいるところも国研とかなり違うところかもしれません。そのあたりのことに配慮しながら全学的な研究データポリシーができるのかということ、なかなか思い付かないというのが正直な印象です。

●八塚 ありがとうございます。ディスカッションなのであえて調和を崩すようなことを申し上げますと、皆さんのお話で非常に印象的だったのが連携という言葉が出てきたことですが、私は正直、連携というだけではこの分野は進まないと思っています。なぜかというと、日本の場合は結構、「おまえ、やれよ」というふうに互いに押し付け合ってしまうと思うからです。もちろん連携は最終的に大事ですが、最初に熱量を持った人がどれだけいるかということが気になっています。皆さんは現場をいろいろ見てこられたと思いますが、そういった熱量を持った人がどのくらいいるのか、多少摩擦を起こしてでも進めていこうという感じの人がどれだけいるのか、どなたかコメントがあり

ましたらお願いします。

●込山 そういった方は、リサーチデータや IT など、違う業界の違うイベントに行ってもいらっしゃるし、図書館のイベントにも毎回来られています。そういう先生方や熱意のある図書館職員など、顔見知りばかりになるのは良くて、どんどん新しい方が増えてきて、熱量を持って「私だったらここだったらできる」ということを言うてくださるメンバーをどんどん増やしていかないといけない段階だと感じています。自分の周りだと、そういう方は 30~40 人はいらっしゃるのではないかと印象です。

●林 もちろん熱量というものも当然あるでしょうし、頂いたご質問で、「トップダウンの動きがまだあまり機関内でもありません。逆に、誰か熱い人がいてボトムアップ的に何かするということはないのですか」という質問が 1 点来ています。

それから、これは熊崎様への質問で、「原子力機構で上層部向けにアドボカシー活動を始めたということですが、理事長からの指示がない状態で、まず上層部に必要性を認知させたということなのでしょうか」という質問が来ています。

●熊崎 まず質問にお答えしますと、私が着任したときには既に、それはやるべきこと、なぜなら「統合イノベーション戦略」に示されており、私たちは国立研究開発法人なので国の位置付けの下で取り組まなければならない、という状況が始まっていました。検討の体制や、どのように進めるかという上層部の判断は、私はつまびらかには存じ上げていません。

ただ、このような状況の中で、研究連携成果展開部の研究成果管理課が担当するのだと決まったときに、ボトムアップで働き掛ける必要を強く感じました。組織の中の他の部署では、これは原子力機構でも行うことであるとあまり認識されていない方もいますし、研究者の中には「海外では聞くけれど、とうとう原子力

機構でもするのですね」とおっしゃる方もいます。ボトムアップの強い旗振りや先導ではなく、個人の対話であっても、何か動き出して進めていくことで、大きなポリシーをつくることや、少しの理解を得ることにつながっていくのではないかと思います。先ほどの八塚様の熱量ということであれば、原子力機構で研究者の方に話をする中で、とても力になってくれる方や、これは知財ポリシーなどと同様に非常に重要だから一緒に頑張りましょうと言ってくれる方もいますが、それはこのような働き掛けの中から生まれてきたものだと感じます。まず声を上げる、何かやってみることから始まると思います。

●林 アドボカシー活動を通じて、ボトムの声拾い上げたというイメージでしょうか。

●熊崎 そう思います。

●林 私も国研の人間なので、「統合イノベーション戦略」から降ってきたら、本当にものすごいトップダウンで、やらなければいけないからやるのだという空気があります。しかし、原子力機構の場合は地道なアドボカシー活動を通じて、結果としては研究者の皆さまの声を拾うことができ、上からも来ているし、下からも挙がるという、ある意味、いい感じでできているような印象を受けました。

他のパネラーの方は何かありますでしょうか。実は自分が熱量を持っているといったことでもいいです。

●山田 国研でもない、分野の方でやっている糖鎖のリポジトリは元々、研究者の方から、こういうものが必要だという形で始めたのです。それを受けて、私たちシステムをつくっていくメンバーも重要性を認識してつくっていききました。つくり終わって、本当に使ってくれるのかと不安なところもありましたが、1~2年たったころに国際会議で実際に研究者の方から、これをもっとみんなに宣伝して使っていこうという声を

上げていただきました。それで糖鎖の関連ジャーナルの先生や会場にいた人たちに、連名で使いましょうとメーリングリストで賛同者を募り、80名ぐらいの賛同を得ました。糖鎖の領域はそんなに大きな団体ではないのでそのぐらいなのですが、一部の研究者の熱意が私たちのやる気につながっているというところは分野的にはありました。

●林 ありがとうございます。八塚さん、何かコメントはありますか。

●八塚 今、山田さんからもコメントがありましたように、現場の方が利用者のメインだと思います。研究データ管理は必ずしもイコールオープンではないと思いますが、逆に、管理を全くせず、今までのままでも困らないですかという問いを立てたら、困るという方が出てくると思います。そういうところに熱量とやる気はあるのではないかと思います。もちろん、やらなくてはいけないというのは、国や組織からのトップダウンでもあると思いますが、やらないと自分たちが困るという意識を皆さんと共有できたらとはすごく思います。

●林 誰かしら熱量がある人が1人いれば、旗を振ってくれるし、付いていこうという感じになるのでしょうか。けれども、やはり言い出しっぺの法則が働いて、言った人がやらなければいけなくなるところもあるので、なかなかやりづらいのではないかと思います。

いろいろなステークホルダーの人がいますし、例えば研究倫理の話では、「データの管理を行う上では、研究倫理教育の推進が関係するのではないか。こういう人たちとは連携されているのか」という質問が来ています。あるいは、研究公正という視点から、実験ノート（ラボノート）を厳密に管理しなさいと言われていて、研究推進のためにオープンアクセスもすごく大事です。しかし負担は減らしてほしい。そういうことを言われると、やはり研究公正を担当する部署と

も話をしないといけません。質問の3点目にも関係するのですが、違うステークホルダーに、何かこれを図ってほしいということはありますか。

●**込山** 研究公正という話が出ましたので、GakuNin RDM のユースケースの事例を一つ紹介します。

東京大学の事例です。研究者と事務セクションの研究倫理推進部門で、研究所内で出版される論文の全てについて画像をチェックするという業務フローがあり、メールで「先生、データを出してください」「このフォーマットではチェックできません」というやりとりを何往復もして、専門員が付いてプログラムを動かして目で見てチェックしているところがあります。その真ん中に GakuNin RDM を置いていただくことで、研究者は日常的に RDM からデータを入れていくのですが、事務局部門に「投稿されたのでチェックしてください」というアラートやメールが飛んできて、それによって業務ワークフローの処理が楽になるというふうに、うまく合わせて使ってもらう仕組みになっています。研究者と倫理部門のやりとりを楽にするのも、研究データマネジメントサービスの一つの役割だと考えています。

●**林** 研究公正をチェックするのに GakuNin RDM があれば、今まで苦労していたものが解消できるというイメージですか。

●**込山** そうですね。今までは、PDF 等、クラウドストレージなどを使うのですが、管理も全てメールでされていたということです。全ての論文のチェックのやりとりをするのは、担当者も1名ぐらいしかいらっしやらないので非常に大変で、記録もメールをさかのぼるのは意外と大変なので、記録が残ることは良いことだとは言われています。

●**林** そういった動きは、今後ポリシーにどうやって反映させるのでしょうか。ポリシーに書いてあること

を実装するようなことでしょうか。

●**込山** この場合は、RDM ができる前から研究所内ポリシーのようなものをお持ちで、そのガイドラインに従ったチェックをしていました。ある程度、「この形でデータを提出しなさい」「ファイル形式はこうです」「どういうデータでどこに論文を投稿したかということを書きなさい」というガイドラインがあった上で、第2期として RDM と連携させていくということになっています。

●**林** ガイドラインをつくった人たちとシステム側でうまく話し合って、GakuNin RDM だったらガイドラインに書かれているタスクをこなせるということになったのですね。

●**込山** そうですね。GakuNin RDM を追加拡張する形ではやっているのですが、システム拡張する場合でも、先にポリシーとガイドラインと登場人物と業務ワークフローが紙になっていると設計も非常にしやすくなり、うまくはまったという事例です。

●**林** なるほど。研究公正の部署や研究者側のニーズをうまくシステム側ですくえたという感じですね。

では、これも込山先生にお聞きますが、逆に、これはシステム側では吸収しきれないから、ポリシーなりガイドラインなりの運用側をもう一度見直した方がいいのではないかという事例はありましたか。

●**込山** それは割と多くあるような気がします。私たちのフレームワークだと、学術認証フェデレーションで認証などをしているので、アカデミックのサポートを NII は得意としています。やはり、産業界との産学連携はどうするのかというのが私たちの課題です。あとは、他省庁の管轄の研究所と一緒にどうしていけばいいのか。もちろん同じ仕組みでつくって二重開発にならないように技術提供をした方がいいと思います。

本当は中で共同研究ができた方がいいのですが、ポリシーや政策の話ではそういった難しい面もあるかと思っています。あとはやはり、大学と国研が抱えているポリシーづくりはだいぶ違うということがあります。

●林 次に、例えば今のシステムの実装や、ポリシーをつくる、あるいはメタデータを付けていくという運用周り全て含めてだと思のですが、「実際にデータペーパー、研究データに関する論文の事前サーベイをしてからインタビューに来てくれるとうれしい」というコメントを頂いています。

あとは、「研究の現場が忙しくて、聞きに来られてもなかなかお答えできないし。アンケートと言われてもなかなか対応できません。その結果、取りこぼしてしまうのですが、具体的な研究データに関する論文を書かれていたりする事例があります。そういった先行論文のサーベイはされているのですか」という質問が来ています。要は、今までも出したことがあるから、それを見てくれた方が早いのだけれどもということですが、先行事例やデータを使ったサーベイは、皆さんそれぞれの立場でされていますか。

●熊崎 原子力機構は外部発表のデータが集積されているので、発表前のもも発表した後のメタデータも把握しています。ですから、「データジャーナルに投稿した人がいるらしい」「では見てみよう」という調査はできますし、実施しています。インタビューの前に、その方がどういう分野の研究をされているかは少なくとも確認しますが、「私たちは文系で、詳しく理解できないので教えてください」となってしまう。伺って、インタビュー後にその方の論文を再度検索する、などをしています。ただ、「何も分からないのに来るな。勝手にやればいいではないか」とおっしゃる方ももちろんいるので、対話が必要だと思います。

●林 そのあたりの視点で、「研究データ管理の導入について、研究者サイドとして率先してできることは

何かありますか」という、非常に熱量を感じるような質問も頂いています。先に論文を見てくれるとうれしいという人もいれば、逆に率先してできることはないかと言ってくれる人もいて、ありがたいことですが、研究者の方にお伝えしたいことは何かありますか。

●込山 GakuNin RDM を普及させようとしているわれわれNIIの立場からすると、まだGakuNin RDMが導入されていない大学や機関の場合、研究者から情報システム部門、情報推進部門に「RDMをやりたいのになぜないのか」「RDMを早く使えるようにしてくれ」というリクエストがあり、情報システム部門からNIIに連絡が来るということが一番多いです。それは私たちも直ちに使えるようにしようという動きになります。一番早いのは、情報システム部門に「NIIのRDMを試してみたい」と言っていただくことです。いきなり全所的に使うのは難しいと思うので、「試してみたい」というリクエストを情報システム部門に連絡していただくことを、私たちとしては一番期待しています。

●山田 糖鎖の方では、いろいろ開発していく上で、研究者に役立つものをつくりたいという気持ちが強いので、研究者から直接こちらに何でもいいのでもっと意見を言ってほしいです。全てに対応できるかどうかは分かりませんが、さまざまな意見を多くの人から頂いて、優先順位を付けてやっていきたいと常々思っています。言いにくい方もいるかもしれませんが、積極的に情報発信していただくことは研究者コミュニティの中でも有益になると思うので、そうしていただけるといいかと思います。

●林 今の山田さんのコメントは、もちろん組織として研究公正でやらなければならないし、上からいろいろ言われたり、もっとすごい上の方から降ってきたりして、やらなければいけないこともあるけれども、発信すれば自分たちの研究コミュニティにフィードバックなりメリットがあるということでしょうか。それは

すごく大きなことです。

次に、これは一番「いいね」が付いている質問で、「オープン化のメリットは、利活用が考えられますが、やはりアーカイブするだけではなかなか難しいのではないかと思います。単にアーカイブだけではないのではないかという部分について、もし何かいい事例やお考えがあれば頂きたいです」ということです。

それから、「結城さんの発表の中で、AXIES-RDMで船守先生から図書館にも関わってほしいという要請があったと聞いたのですが、事務職の業務、作業の一端としてやってほしいという話なのか、それとも、対等の立場で議論したいということなのかを教えてください」という質問が来ています。

●結城 そこまでの意図はよく分からず引き受けてしまったのですが、こういう研究データのいろいろな活動に関わっている理由としては、私がその位置にたまたまいたということが大きいのです。しかし、実際にAXIESに参加してみると、割と対等な立場で扱っていただいていると思います。当然、研究者ではないので、そのスタンスでの話はなかなかできないのですが、なるべく図書館員としての見解を述べるようにはしています。それが研究者と少し違った視点でAXIESの活動に役立っているのであれば、よかったと思います。

●林 私も事務職という立場なので、対等なのか、事務として淡々とやるべきなのかというのは悩ましいところです。今のお話、あるいは発表全体を聞いて、みんなで対等にやっていくということはすごく感じられました。それは熊崎様の発表の中の原子力機構の事例でも、相手をつなげる過程でこれだけ対等にやられているのだなと個人的にはすごく感じました。確かに違うステークホルダーというだけであって、上下だとか、事務だから事務作業を受けますという感じではないのだらうという思いがあるのですが、やはりパネラーの皆さまもそういうイメージでしょうか。そうでないと、三つ目の「理解していただきたい」が、「理解しろよ」

となってしまうとよろしくないのでは、お互いに対等に議論し合うということだと思いますが、いかがですか。

●八塚 もちろん私も組織の人間なので、ポリシーを決めたりすることが非常に大事だということはよく分かります。ただ、まずはデータを出していくということを1回でもやらないと、課題が見えてこないのです。私も一通りやってみてよく分かりました。よくある例では、公開してほしいというデータが来るのだけれども、このデータがどういう意味なのかがそもそもよく分からないし、生データを見ても全然分からないし、メタデータを研究者は書いているつもりなのだけれども私はさっぱり理解できないので、どういうことですかと研究者に尋ねます。そこで初めて会話が生まれて、お互いに理解できて、データを公開することになります。

また、ライセンスはこれでいいということで最初は大枠は合意していたはずなのに、いざ出そうとすると、ちょっと待ってくれという話になって、そこでまた議論しだすということもよくあります。実際にデータを生み出していくことによって課題が見えてくるし、どうすればいいかが見えてくるし、お互いの理解が進むと思うのです。もちろんポリシー決めは大事なのですが、むしろポリシーは最初は大ざっぱでもいいので、まずはデータをそのスキームで生み出して、その中でお互いに対等の立場で話をしてフィードバックをして、どんどん洗練させていくのがいいのではないかと思います。

●林 八塚さんが最初に示したスライドの管理のフェーズで、パネラーのそれぞれがどこに重点を置いて携わるかが違って行く中でも、対等にみんなうまく回して、被っているところもあるし、それぞれが得意なところもあるので、うまく分担して進めていくのがいいのではないかと、今、皆さまの意見を聞いていて思いました。

引き続き、頂いた質問です。これは皆さまにと書かれています、「研究データの取り扱いがオープン・アンド・クローズ戦略に基づいていると聞いています。ただ、線引きが分野や時代で変わるのではないのでしょうか。今のこの情勢下だとクローズだけれども、昔はオープンだった、あるいは 10 年後の情勢だとオープンにすべきなのではないか、もしかするともっと間口が狭くなってクローズに向かっているかもしれないというふうに、いろいろなオープンアクセスの方針やデータポリシーがありますが、時代に合わせて将来的に改定していくことは議論されているのですか」という質問です。

●熊崎 あまり時代までは想定していなかったのですが、データポリシーの内部パブコメの際に、これは将来改定するかもしれないのだから、「改定するかもしれない」と書いたらどうですかという意見があり、結局外部に宣言することではなく内部で控えておく事項として整理しました。内閣府のガイドライン（国立研究開発法人におけるデータポリシー策定のためのガイドライン）にも改訂するように書いてあります。ですので、時とともに、あるいは組織の方針、法改正によって変わっていくものだと認識しています。

●林 そうなると、例えば受けるシステム側も、中長期的には開発がある程度続いていくようなイメージになるのでしょうか。

●込山 そうですね。基本機能を兼ね備えて 2020 年度からスタートすると思いますが、当然ポリシーが変わっていくので、裏側のネットワーク、ストレージをそれに準拠してどう整備するかという、運用面の開発はずっと続いていくことになると思っています。質問の中で、どういった体制でスタートするかという話もありましたが、今まさに NII の事業部門で運用していくために、CiNii や JAIRO Cloud と同じように事業部門として提供していけるように、単に R&D だけやっ

て、はい終わりというわけではなくて、ずっと続いていくサービスとしての体制を敷いている最中です。

●林 今回の関連で、今日の議論の中で 50 年後に NII はあるのかという話がありました。50 年後ではなく、当面で構わないのですが、NII での GakuNin RDM の運用体制や料金体系について、もし今の時点でお話しできることがあればお願いします。

●込山 NII はようやく今年度 20 周年だと思いますが、2022 年度以降にアプリケーションを含めて受益者負担でお金を頂いていく方向にあるということは聞いています。われわれは初期の導入という意味では標準的なストレージは準備しますが、徐々に大学側で、クラウドでもオンプレミスでもいいのですけれども、ストレージも準備していただいて接続して、そこにデータを残していけるように持っていこうという戦略です。ただ、一定容量はしばらくの間はお使いいただけるようなストレージを準備するという方針になっています。

●林 逆にいうと、機関側では自分たちのストレージ分は当然負担しなければいけないし、GakuNin RDM にも何かしらの利用料金というか、一定の負担はあるのですか。

●込山 サービスの運用というか、アプリケーション代は頂かないと維持できなくなるのはもう見えています。

●林 確かに、データを預けたはいいけれども、お金が切れてしまつてというのは本当に避けたいところです。

●込山 研究大学で巨大なストレージが要するところは、やはり学内に大きなストレージを置いていただいて、それを GakuNin RDM につなげていただく。そうでは

ない単科大学や小さな規模の大学だと、GakuNin RDM のデフォルトの領域で受けられる可能性もあるかもしれないので、大学の規模やプロジェクトの内容によって使い分けていくことになると考えています。

●林 機関側の金銭的な負担やシステム的な負担を考えたときには、RDM の入力インターフェースの部分よりも、むしろストレージが一つのポイントだという印象ですが、そんな感じですか。

●込山 実際にお金がかかってくるのはストレージの料金です。そこは使いたい人が使いたい部分を確保するという受益者負担になっていくことは間違いありません。

●林 先ほどの 50 年の話ではないですけども、今、ストレージは安くなっているといわれていますが、安くなる代わりに使う分も当然増えるでしょうし、先ほどの研究公正の議論からすれば、1 回登録したら 10 年ぐらいで消えてなくなるという想定はあまりないだろうと思います。ですから、負担はだんだん大きくなっていくかもしれないという印象を受けました。

●八塚 今の永続性の部分に関して、研究データリポジトリの認証の CoreTrustSeal の中で、長期計画があるということが一つの要件になっています。もちろん長期計画を示す必要がありますが、その中には終了計画も含まれています。いつ終わるかは決まっていなくても、終わったときにどうするのかということまできちんと書くということです。終わったときに、例えば、より大きなリポジトリに全部を預ける、あるいは、これはあまり現実的ではないと思いますが、寄託者にデータを全部返すなど、そういうことも考えなさいということが明記されています。

●林 確かに、分野別のリポジトリであれ、今日の話題の一つであった GakuNin RDM であれ、本当にイン

フラだと思います。なくなってしまうというのは避けたいので、そういった認証制度がきちんと用意されていて、その中で考えていく事項は当然あるということだと思います。

あとは、これも少し要望が高めなので、皆さまに伺います。「国研など国の機関で GitHub、AWS のような海外の資本のサービスを使うことについて、パネラーの方からのご意見をお聞かせください」という質問が来ています。国研だけではなくて大学なども含まれると思いますが、いかがですか。私の知っている範囲では、全く駄目ではなくて、掲載する情報によって海外のクラウドでもいいかどうかを考えた方がいいという認識でいるのですが、もしコメントがあればお願いします。

●熊崎 国研などの政府機関の統一基準としては、「政府機関等の情報セキュリティ対策のための統一基準群」という基準があります。何年度版というふうに改訂されていくもので、この中に例えば、クラウドサービスで取り扱われる情報に対して、国内法以外の法令が適用される場合の対策や、その事業の実施場所や契約に定める準拠法、裁判管轄が指定できることについての項目があります。海外資本だからというよりも、それによって適用される法が変わってしまっていて、結果、不利益を被ることがないようにするということだと理解しています。それは国研だからではなく、重要なデータを扱う組織として考えておくべきことなのではないかと思います。

●林 ちなみに、今 GitHub を使っているチームが GakuNin RDM を入れたときに、今までどおりの GitHub の使い方をしていれば、GakuNin RDM がいい感じで連携してくれるのでしょうか。

●込山 GakuNin RDM の研究ポータルとひもづける必要はありますが、それ以降は GitHub を通常どおり利用すれば、その記録が RDM 側にログインすると履

歴として残っていきますし、RDM から GitHub へのアクセスもできます。GitHub は変わらずそのまま使ってください。

少しだけクラウド利用の補足をしたいと思います。外資のクラウドベンダーは全て駄目かという、一概にそういうわけではありません。大手のクラウドベンダーは必ず日本国内にもデータセンターを持っていて、国内法を準拠しているか、国内法人を置いているか、あるいはサービスのシステム中でクラウドのオプションとして日本の法律で係争するということをきちんと選択できるような会社もあります。そういったチェックリストを見る必要があると思います。

●林 チェックが済んでいれば連携はできるということですね。AWS、具体的には S3 のようなストレージでもそういう連携はできるのですか。

●込山 はい。そこの連携は間違いなくできます。プロバイダー各社のストレージと接続できる仕組みを GakuNin RDM は持っています。

●林 ありがとうございます。

最後に、パネラーの皆さまからコメントを一言ずつお願いします。

●山田 やはり図書館の方たちと分野別のところでは感覚が違うところはあるのだと思います。私たちがつくってきているものは、トップダウンでという認識は全くなく、分野を良くしたいという思いによるものです。図書館の機関リポジトリの方も、良い環境を提供したい、良いものをつくりたいという意味では全く同じだと思います。連携するだけでは駄目だという話もありましたが、では実際に連携を前提にして、どうすればそれをより良くできるのか、意味があるものにできるのかということを議論できる場を今後もつくっていく必要があると感じました。

●熊崎 私は図書館だからという意識で研究データ管理の業務をしておらず、原子力機構という組織の中の仕事をしていると思っています。理想論ですが、図書館だからやるとか、誰々がやるとかではなくて、みんなで組織の研究の環境を良くして、みんなが気持ち良く仕事ができるように動いていけたらいいと思っています。八塚様の「登壇者の皆様への質問」にありましたが、異なるステークホルダーを理解していただきたいというよりは、対話をして一緒に考えていただきたいです。面倒くさいかもしれませんが、アンケートをするのであれば、おざなりではなくて一緒に考えていただき、意見を言っていただく。また、データポリシー策定後はポリシーに基づく事例をつくることが大事で、そのための相互の協力が必要だと原子力機構でも思っています。みんなでお互いに一歩歩み寄ることができたら、もう少し研究しやすくなるのではないかと思います。

●込山 まずは研究者に GakuNin RDM にデータを入れていただき、そこで研究を進めていただいて、論文を書くなど何らかの結果が出たというユースケースが集まれば、次につながる原動力になっていきます。ですから、まずは研究者に便利に使っていただくような活動を継続していきたいと考えています。便利に使っていると、裏側でシステムがアシストして、勝手に研究データマネジメントをしていたという状況を早くつくりたいと考えていますので、皆さまのご支援をよろしくお願いします。

●結城 トップダウンの話で気になっていることがあります。最初の事例の名古屋大学が割とトップダウンで進めていますが、そういう大学は多くないのが現状ではないかと思います。本学もそうで、いろいろな事情はあるのですが、そういうことが全体的に進んでいるわけではありません。そういう状況で、私は図書館職員なので図書館職員の立場で言うのですが、だから何もしなくてもいいという話でもないと思っています。

図書館として取り組みやすい研究データの公開の部分や、リポジトリの整理の部分は頑張りたいですし、アンケートやインタビュー調査も今後実施したいです。ステークホルダーの皆さんと仲良く一緒にできればさらに良いのですが、そうできない状況もあると思うので、できる範囲で一生懸命やって、皆さんが乗り気になってきたときにうまく協力することができれば、研究者の皆さんにとっての良い環境づくりのお手伝いができるのではないかと考えています。

●八塚 この SPARC Japan セミナーに私も何年か関わっていますが、普段はなかなか会えない異なるステークホルダーの方が集まって対話できることはとてもいいと思います。これをきっかけに、もっと研究データについて皆さんと熱く対話できればと思っています。

●林 今日の議論をまとめると、自分たちにできることはどこかに必ずあり、研究データ管理という大きな目的の中のどこかのニーズでその役割が果たせるのではないかということかと思います。

フロアから質問をお送りいただいた方、議論に参加いただいた方、中継をご覧いただいた方、中継先から質問を寄せていただいた方、全ての方に感謝します。これでパネルディスカッションを終了します。