SPARC Japan NewsLetter
ISSN 1883-826X
NO.36 2019 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
SPARC Japan 運営委員会
arXiv.org 会員コンソーシアム

■ 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

CC BY

SPARC Japan 運営委員会

SPARC Japan 運営委員会の会議資料をウェブサイトで公開しています。
http://www.nii.ac.jp/sparc/about/committee/

arXiv.org 会員コンソーシアム

arXiv.org（アーカイブ・ドット・オルグ）は 1991 年に創設された物理学、数学、コンピュータサイエンス及び関連分野のプレプリント・サーバです。2017 年には電気工学システム科学、経済学が新たに加わりました。2018 年末時点の論文登録件数は約 150 万件、年間ダウンロード数は約 2.2 億件以上で、総ダウンロード数は 2017 年に 10 億件を超えました。長くコーネル大学図書館（CUL）が運営してきましたが、2019 年 1 月に図書館から Cornell Computing and Information Science（CIS）に運営が移りました。また、ガバナンスのために会員諮問委員会（会員から選任）、科学諮問委員会（科学者・研究者）を設けています。世界的には、2013 年に開始した「arXiv 会員制プログラム」により、2018 年末現在で 27 カ国 231 機関が会員として支援しています。

SPARC Japan は、第 5 期事業の達成目標の一つとして「国際的な OA イニシアティブとの協調」および「学術情報流通にかかわるアドボカシー活動」を掲げ、その具体的活動として arXiv.org の日本研究図書館コンソーシアム（Consortium of Japanese Research Libraries: Coordinated by National Institute of Informatics (NII)、NII Japan Consortia）の事務局を担い、日本の参加機関の取りまとめと広報活動を行っています。当コンソーシアムを通じた

日本からの参加は現在 16 機関（2018 年新規加入 2 機関）で、その貢献に応じて会員諮問委員会に一つ席が与えられており、現在は国立情報学研究所 武田英明教授（SPARC Japan 運営委員会委員長）が代表として参加しています。

資金に関しては、主に会員からの会費のほか、コーネル大学図書館、サイモンズ財団からの支援に依っています。日本の参加機関の拠出金は毎年 NII が取りまとめて arXiv.org に支払い、各参加機関は NII に支払っています。さらに、スローン財団・ヘイシングサイモンズ財団からも次世代 arXiv イニシアティブへの資金提供があります。また、オンラインでの募金活動も 2015 年の試行以来継続して行っています。

- 会員からの資金調達額に応じて$550,000（年間、2019 年見込み）
  会費は過去 3 年間（2015~2017 年）の利用実績（ダウンロード数）の平均順位に応じて$1,000～$4,400（年間）
- コーネル大学図書館：$170,000（年間）＋間接コスト（総費用の 37%）
- サイモンズ財団：$100,000（年間）、マッチングファンド$300,000（年間）
- スローン財団：$450,000（次世代 arXiv イニシアティブ対策、総額）
- ヘイシングサイモンズ財団：$322,000（次世代 arXiv イニシアティブ対策、総額）
- オンラインでの募金活動：$56,666（年間、2017 年実績）

※最新の情報は「arXiv Update - January 2019」をご参照ください。
https://confluence.cornell.edu/display/arxivpub/arXiv+Update+-+January+2019

■ SPARC Japan セミナー報告

CC BY

第 1 回 SPARC Japan セミナー2018

「データ利活用ポリシーと研究者・ライブラリアンの役割」

2018 年 9 月 19 日（水） 国立情報学研究所 12F 会議室 参加者：70 名

2018 年 6 月、内閣府の「国際的動向を踏まえたオープンサイエンスの推進に関する検討会」により、「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」が策定されました。今後は、公的資金の配分を受ける大学や企業等においてもデータポリシーの策定が広がっていくことが見込まれることから、今回のセミナーでは、このガイドラインの位置づけや経緯を踏まえ、どのようにしてデータポリシーを策定するか、また図書館の提供サービスにおいてデータポリシーをどのように活用できるかを検討しました。

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/2018/20180919.html)

企画概要

2002 年のブタペスト・オープンアクセス・イニシアチブと 2013 年の G8 オープンデータ憲章は、科学アカデミーからのボトムアップな自主活動とトップダウンの政策提言という互いに交錯しつつも共通の方向を目指し、「オープンサイエンス」の実現を加速する転回点となった。日本におけるオープンサイエンスの推進に向け、研究データの利活用に向けた取り組みが各所で進められている。しかし、これを実現するための方針であるデータポリシーやその運用体制の整備は不十分である。この状況を踏まえ、「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」（以下、「ガイドライン」という。）が策定された（※１）。このガイドラインに基づき、国立研究開発法人のみならず、公的資金の配分を受ける大学や企業等においても、データポリシーを作成する必要に迫られている。（※２）

パネルディスカッション（左から林賢紀氏、赤池氏、林和弘氏、白井氏、丸川氏、石山氏）

ガイドラインでは、研究データ管理・利活用ポリシー策定におけるポイントや、定めるべき項目（目的、定義、制限事項、管理、メタデータ、識別子など）が示されている。研究データを日々生成し管理・公開を行う研究者はもちろん、メタデータの付与などを通じて研究データの品質管理・流通・提供に携わる図書館員も、各々の専門性を発揮してデータポリシーの策定に積極的に関わることが求められている。たとえば、ガイドラインにおいてはデータが具備すべき要件として「国際的なデータ管理原則である「FAIR 原則（※３）」に可能な限り沿うもの」とされている。ここでうたわれる「Findable（見つけられる）、Accessible（アクセスできる）、Interoperable（相互運用できる）、Reusable（再利用できる）」 という原則は、研究データのみならず、これまでの主たる対象であった文献情報を含む学術情報流通全般に適用でき、図書館におけるサービスの基底となりうる考え方といえる。

本セミナーでは、このガイドラインの位置づけや経緯を踏まえ、どのようにしてデータポリシーを策定するか、また図書館の提供サービスにおいてデータポリシーをどのように活用できるかを検討する。

※１　総合科学技術・イノベーション会議に設置された「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」による。

※２　イノベーション戦略調整会議で検討が進められている「統合イノベーション戦略（素案）」によると、国立研究開発法人のデータポリシーの作成は 2020 年度末まで、研究資金配分機関によるデータポリシー策定要請の制度化は 2021 年度予算での公募までに導入することとされている。

※３「データ共有の基準としての FAIR 原則」(2018 年 4 月 19 日、NBDC 研究チーム) DOI:10.18908/a.2018041901

参加者から

（大学/図書館関係）

・リポジトリを構築したばかりなので、オープンサイエンスに至るまではリポジトリ担当（本学の場合図書館員）がより高度な知識と取り組みを進めないと停滞したままになると危惧しました。デジタルアーカイブも現在構築中なので、国立民族学博物館の発表は大変参考になりました。

・オープンデータの有効性とリスクについて理解が深まった。

・具体的な事例や現役の研究者からの視点も垣間見ることができ、とても有意義なセミナーでした。個人的には海外の事例をもっと知りたかったです。セミナーの中で今後科研費獲得のためにはデータポリシーが必須になるかもしれないなどのお話がありましたが、現状そのような条件を課している、又は予定している国や団体があるのかなど、状況を知りたかったです。また、個人情報に関わる研究データ（医療系など）ポリシー作成についてどのような留意点があるのか、法律家からの視点もあると、より多面的なセミナーになるかと思います。

・具体例も多く、とても役立つ内容が多かったです。研究データ管理について、以前より具体的にイメージできるようになりました。〔動画中継視聴者〕

（企業/図書館関係）

・機関リポジトリとの比較が、理解の助けになりました。ポリシーの中身というより、めざすところ、策定の意味を説明頂けて、国の意図が分かった気がします。

（企業/その他）

・データの提供を研究者の方が面倒に思うことは存知あげていましたので、データをタグ付けしやすい研究支援システムや、ソフトウェアなどを、データ利活用ポリシーに沿って、国で開発されていないか？と興味を持って本日の講演にお伺いしていました。データカタロギングのツールが向上すれば、図書館の方々もRA の方々も、本件に関して積極的に参加、協力できると思います。

（その他/研究者）

・データの活用により、新しい成果が出るという話がなければ、研究者としてはモチベーションがあがらないのではないでしょうか。

企画後記

CC BY

☺研究データの公開にあたり必要となるポリシーの策定から実際の公開まで、様々な事例を伺うことができる機会となりました。研究データの公開やポリシーの作成は政策レベルでは「統合イノベーション戦略」に基づく取り組みですが、その目指すところや図書館を含む研究の現場までのつながりを実感できる議論ができました。企画担当として、ご参加いただいた皆様にお礼申し上げます。引き続き、研究者と図書館員が協働するきっかけとできれば幸いです。

林　賢紀

（国立研究開発法人 国際農林水産業研究センター）

☺今回公式アカウントでの Twitter 発信を担当しました。講演の内容をリアルタイムで発信するのですが、必ずしも追いきれなかった部分もあったかと思います。そうした中、講演を聞かれている方々が補足のツイートをしてくださることが何度かありました。本セミナーのテーマの一つは研究データをめぐる協業についてですが、まさに運営スタッフと参加者の協業がセミナーの場で実現できたと思います。

八塚　茂

（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）

☺データポリシー策定のためのガイドラインは、各機関のミッションに合わせて、オープン・アンド・クローズ戦略を定めるよう勧めています。ポリシーの策定を義務として捉えるのではなく、改めて自機関の学術情報や研究活動の在り方について再確認する良いきっかけとして考えていきたいと思いました。

石山　夕記

（一橋大学附属図書館）

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2018

「データ利活用ポリシーと研究者・ライブラリアンの役割」

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情報管理係長。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。

データポリシー策定のためのガイドライン

昨今、日本でもオープンサイエンスの推進に向けて、研究データの利活用に向けた取り組みが進められています。しかし、これを実現するための方針であるデータポリシーや運用体制の整備がまだまだです。このような動きを受けて、2018 年 6 月、内閣府で、「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」が策定されました。

私の所属する国際農林水産業研究センターも含めた国立研究開発法人のみならず、公的資金の配分を受ける大学や企業等においてもデータポリシーを策定するようにということが示されました。具体的には、「統合イノベーション戦略」を見ると、2020 年度末までにデータポリシーを策定するようにということが書かれており、皆さん大変だと思います。

このガイドラインには、研究データの管理や利活用のポリシー策定のポイントとして、目的・定義・制限事項といったポリシーに加えるべき項目、図書館員の方にはなじみのあるメタデータや識別子についても書いてあります。研究データは日々生産されていて、論文を書いている研究者はもちろんですが、図書館員もメタデータを付けることは、図書館目録などをつくっていれば日々行う作業ですし、機関リポジトリに載せる論文にメタデータを付けているものが、今度は研究データも対象になるらしいというのが今の状況だと考えています。

このような形で、研究データの品質管理を行う、メタデータや識別子を付けて流通させる、これらを用いてデータを一般のユーザーや研究者に提供する研究者や図書館員も、それぞれの専門性を持ってデータポリシーというものを考える、策定に積極的に関わるということが求められています。

セミナーの進め方

本日は 3 件のご発表で、研究データに関わるポリシーの策定の経緯やその内容、そもそもこのガイドラインがどのような意図を持ってつくられたのかという話、実際の研究データの扱いについての話を頂き、最後にディスカッションを行います。

各講演の最後の 10 分間を質疑応答に充てる予定で

す。質問ができなかった点があれば、お手元にご用意した質問票か、ウェブで質問を受け付ける Slido（スライド）というツールをお使いください。本日はどうぞよろしくお願いいたします。

第 1 回 SPARC Japan セミナー2018

「データ利活用ポリシーと研究者・ライブラリアンの役割」

統合イノベーション戦略と

研究データ管理・利活用ポリシー策定ガイドラインが目指すもの

赤池 伸一（科学技術・学術政策研究所 / 内閣府）
林 和弘（科学技術・学術政策研究所）

講演要旨

2018 年 6 月に閣議決定された統合イノベーション戦略において、オープンサイエンスは「知の源泉」を支える柱の一つとして位置づけられた。そして、その具体的な施策として、研究開発法人が研究データポリシーを策定することが掲げられ「研究データ管理・利活用ポリシー策定ガイドライン」が作成された。本講演では、統合イノベーション戦略とその戦略におけるオープンサイエンスの役割を解説し、ガイドラインについてその背景と狙いと共に紹介することで、研究データ利活用を更に具体的に進めるための議論を促す。

赤池 伸一

科学技術・学術政策研究所 上席フェロー（併）内閣府 政策統括官（科学技術・イノベーション担当）付参事官。1992年科学技術庁入庁、文部科学省、在スウェーデン日本国大使館、内閣府、科学技術振興機構、一橋大学イノベーション研究センター教授等を経て、2016年より文部科学省・学術技術学術政策研究所科学技術予測センター長、2018年より現職。内閣府政策統括官（科学技術イノベーション担当）付参事官等を併任。東京大学大学院総合文化研究科修士課程修了（学術修士）、英国サセックス大学科学技術政策部門（SPRU）修士課程修了（理学修士）、東京工業大学大学院社会理工学研究科修了（学術博士）。専門は科学技術イノベーション政策の経済社会効果、ノーベル賞受賞者の分析等。

林 和弘

科学技術・学術政策研究所 科学技術予測センター 上席研究官。1995年ごろより日本化学会の英文誌の電子ジャーナル化と事業化を大学院時代のアルバイトを端緒に行う。電子投稿査読、XML出版、J- STAGEの改善、電子ジャーナル事業の確立と宣伝活動など、幅広いフェーズで実務に基づき考察と改善を加え、当該誌を世界最速クラスで発行する電子ジャーナルに整え、2005年にはオープンアクセス対応を開始し、電子書籍（ePub）対応の技術立証も行った。その経験を生かして日本学術会議、SPARC Japanなどを通じて日本発の情報発信をより魅力的にするための活動を行い、電子ジャーナルの将来と次世代の研究者コミュニケーションのあり方についても興味を持つ。2012年より文部科学省科学技術・政策研究所において政策科学研究に取り組んでおり、科学技術予測調査に加えてオープンサイエンスのあり方と政策づくりに関する調査研究に取り組んでいる。現在、内閣府、G7科学技術大臣会合、OECDのプロジェクトにおけるオープンサイエンス専門家として活動。SPARC Japan運営委員会委員。

●**赤池**　私は科学技術・学術政策研究所と内閣府の参事官を併任しており、今日はむしろ内閣府の参事官の立場からお話をさせていただきます。

オープンサイエンスは最近盛り上がってきていて、政府全体としても大事な課題として認識しています。今日は私から、政策の背景という比較的マクロな話をさせていただいて、林からガイドラインの詳細についてお話をさせていただければと思います。

オープンサイエンスの政策的な背景

政府は 5 年に一度、科学技術基本計画をつくっており、既に第 5 期を迎えています（図 1）。科学技術基本計画では、25 兆円や 26 兆円などという財政目標があちこちでいわれますが、その中でもいろいろな重点分野や重点課題を取り上げてきています。最初のうちは研究開発の計画、研究開発を支えるシステムという側面が強かったのですが、期が進むに従って、社会の中で科学技術をどう捉えていくか、制度的な側面をどう考えていくかという色彩が強くなってきています。

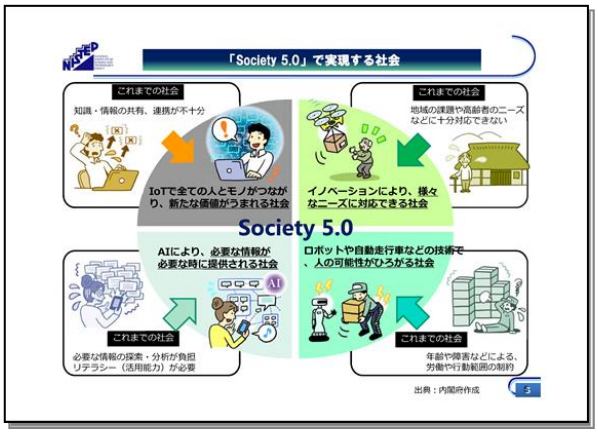
第 5 期科学技術基本計画では幾つかの柱がありますが、その中でも Society 5.0 というコンセプトを立ち上げています。なぜ 5.0 かというのは、「狩猟」が Society 1.0、「農耕」が Society 2.0、「工業」が Society 3.0、「情報」が Society 4.0、「新たな社会」が Society 5.0 ということで 5.0 です。例えばドイツなどでは Industrie 4.0 があり、アメリカでも韓国でもいろいろなイニシアチブが発表されています。Society 5.0 とは、昨今の状況の中で、サイバー空間とフィジカル（現実）空間を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の新しい社会（Society）をつくっていこうという非常に大きなコンセプトです。

具体的には、IoT で全ての人とモノがつながり、新たな価値が生まれる社会など、いわゆる情報科学だけではなく、例えば新しい農業の在り方、新しい地域の在り方、新しい介護の在り方、医療の在り方のようなものも追求していきましょうという非常に広範なコンセプトとなっています（図 2）。

これを支える一番基礎となる知の源泉を、統合イノベーション戦略では、「Society 5.0 実現に向けたデータ連携基盤の整備」「オープンサイエンスのためのデータ基盤の整備」「エビデンスに基づく政策立案／大学等法人運営の推進」として、この三つのデータ関係の柱でやっていこうとしています。

「Society 5.0 実現に向けた」というのは、例えば自動運転や農業や医療、さまざまなところで既にデータの蓄積がありますが、内閣府主導で、それらを有機的に連携させて新たな価値を生み出そうというものが実体的な側面です。「エビデンス」については、今、政府の中、大学で、さまざまな経営情報その他がありますが、そういうものをより有機的に連携させ、エビデンス・ベースト・ポリシー、大学や研究機関の経営改革をやっていきましょうという動きになっています。これらの三つの柱がそれぞれ連動しながらやっていく

（図 1）



（図 2）

という形になっています。

図 3 の政策討議で検討されたテーマは、統合イノベーション戦略をつくるために議論したテーマの一覧です。ここで注目していただきたいのは、このぐらいの粒度の中にオープンサイエンスが入っているということです。元々、オープンサイエンスというのは、もっと政策の粒度の小さいところではたくさん入ってきていたのですが、この中では産学連携や環境エネルギー戦略などと同じレベルで重点事項として入っているということで、非常に大事な柱だと認識しています。

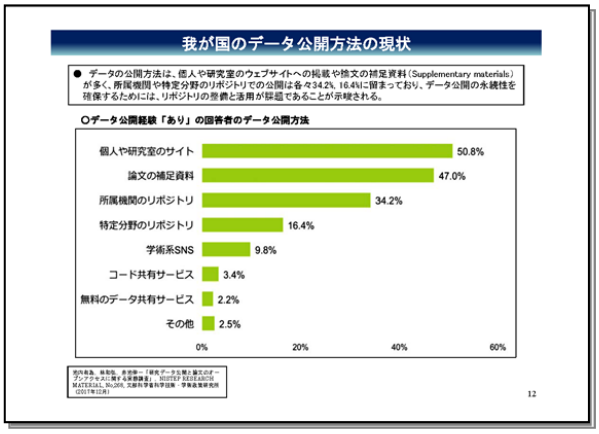
第 5 期科学技術基本計画においても、既に図 4 のような明確な記述があります。特に「公的資金による研究成果については、その利活用を可能な限り拡大することを、我が国のオープンサイエンス推進の基本姿勢とする」こととしています。もちろんその際には、適切なオープン・アンド・クローズ戦略に基づいて行うことは大事なのですが、基本的にはオープンにしていくというものを考えて、既に規定しています。

これを統合イノベーション戦略にするに当たって、少しブレークダウンしています。三つの論点から議論していて、一つ目は「研究データ利活用方針の策定」、二つ目が「競争的資金等におけるデータ管理の要請」、三つ目が「研究データ利活用のための基盤の整備」です。

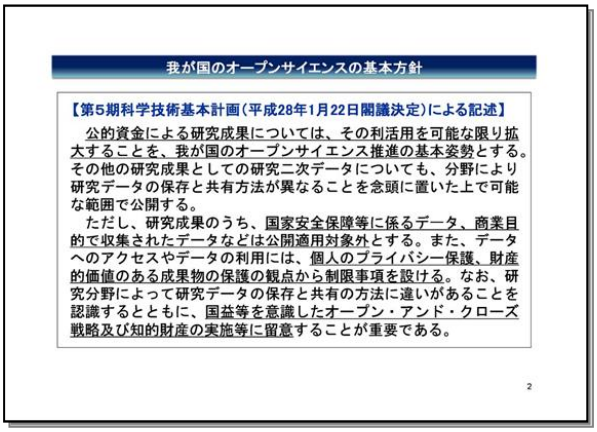
では、現状はどうなっているかというと、オープンは研究者の中でももちろん進んでいて、データ公開の経験が「あり」と答えた研究者は多いのですが、やはり個人や研究室のサイトを通じてやっていらっしゃるという方が多いです（図 5）。リポジトリの整備をきちんと制度的に担保してやっていくことが一つの課題となっています。

また、科学技術・学術政策研究所がオープンサイエンスに関する実態調査を行っており、その結果、日本は大学の数が多いため、リポジトリの数は多いのです

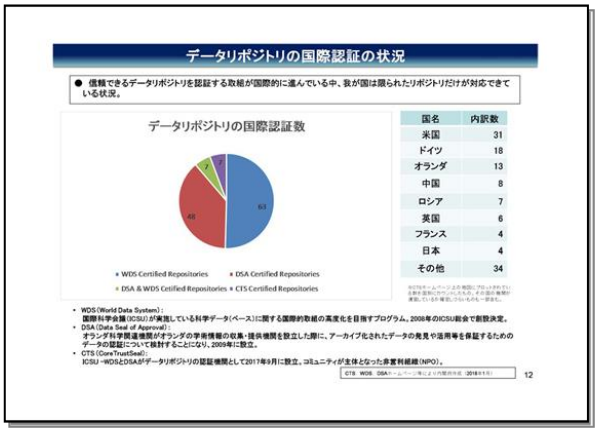
（図 3）



（図 5）



（図 4）



（図 6）

が、必ずしもそのリポジトリの中に入っているコンテンツが充実しているとは言い難い状態にあるということが分かっています。

図 6 は、データリポジトリの国際認証の状況ですが、国際認証は我が国では限られた状態であるというのが現状です。

オープンサイエンスのためのデータ基盤の整備

以上を踏まえて、先ほどお示ししたような統合イノベーション戦略の記述になっているのですが、特に力を入れているのがオープンサイエンスのためのデータ基盤の整備です。

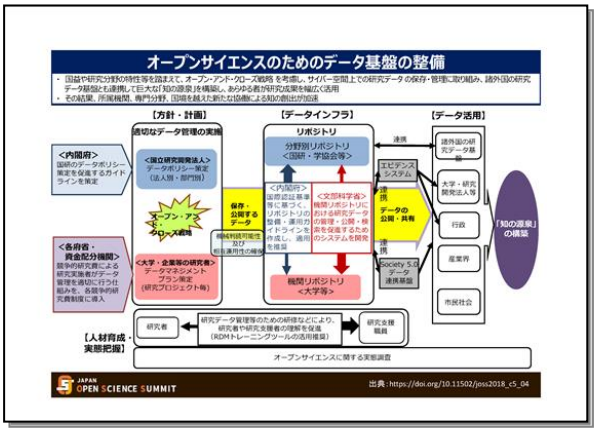
オープンサイエンスのためのデータ基盤とは、国益や研究分野の特性等を踏まえて、オープン・アンド・クローズ戦略を考慮し、サイバー空間上での研究データの保存・管理に取り組み、諸外国の研究データ基盤とも連携して巨大な「知の源泉」を構築し、あらゆる者が研究成果を幅広く活用、その結果、所属機関、専門分野、国境を越えた新たな協働による知の創出が加速する状況と考えています。図 7 は、国立の試験研究機関、大学、部門別というような組織としての取り組みを上側に書いています。また、分野別のリポジトリの整備を政府として支援していくとともに、ガイドラインを作成していくということがこちらに載っています。今回これに関するガイドラインを策定させていただいています。

組織というのはある意味、組織的ヒエラルキーの中に制度的に決まっているものですが、今度は逆に個々の研究者の立場からどう考えるかというと、もちろん組織としての一員であるとともに、外部からお金を受け取って研究活動をするという側面もあります。もう一つの側面からオープンサイエンスを押さえていこうという取り組みです。

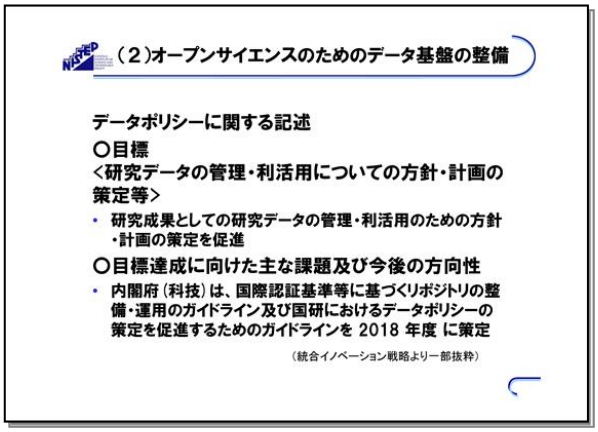
そちらの方は、むしろ資金配分機関の方から考えていて、資金配分機関、ファンディングエージェンシーはさまざまありますが、そちらでは、データ・マネジメント・プランを研究者に作成していただいて、オープン、クローズドということをきちんと研究者の方に認識していただきます。そういう制度をファンディングエージェンシーの中でしっかりつくっていっていただこうという仕組みになっています。組織としての縦の動きと、ファンディングエージェンシーからの横の動きの両方でオープンサイエンスを実現していくという考え方になっています。

もちろんこういうことをするためには、その研究者に「意識しろ」「そうしろ」と言うだけではなく、インフラを支えるということも必要で、国立情報学研究所を中心として各大学のリポジトリを結んで、研究データ基盤をつくる予算措置を文科省を中心にとっていくことも考えています。組織ごとの動き、ファンディングエージェンシー、さらにこれを支える基盤への措置で、実体的なオープンサイエンスを総合的に進めていこうという考え方です。

手前みそになりますが、このような実態把握をきち



（図 7）



（図 8）

んとして、それをまた新たな政策に反映していくということも必要なので、科学技術・学術政策研究所、私や林が中心となってオープンサイエンスの実態調査を進めていきたいということで、こういう総合的な政策の体系となっています。これが私どもの政策の概要です（図 8）。

本セミナーの参加者の中にも会員の方がいらっしゃると思いますが、オープンアクセスリポジトリ推進協会（JPCOAR）から統合イノベーション戦略に対してコメント・見解を頂き、これは大変ありがたいことだと認識しています（図 9）。私も図書館、ライブラリアンはオープンサイエンスを担う主体として非常に大事な機能、職制だと考えています。もちろん私どもがつくったガイドラインを全て「はい、そうですね」と受け入れていただけるわけではないと思いますし、批判的な側面もあるかと思いますが、政策はいろいろな方からのご意見を頂きながらアップデートしていくことが大事だと思っているので、こういう形でご見解を頂いたというのは大変ありがたいです。

詳しい内容は林から話があると思います。私のパートは以上です。

●**林和弘**　科学技術・学術政策研究所の林です。SPARC Japan の運営委員も仰せつかっています。研究データ管理・利活用ポリシー策定ガイドラインについて、書いてあることをなぞっても仕方がないので、なるべくその行間を説明するようにしたいと思います。

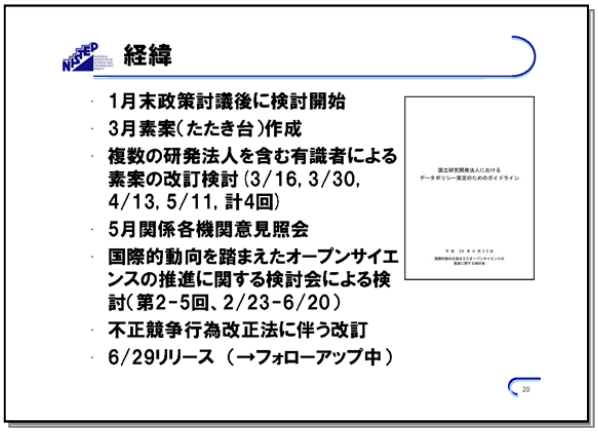
先ほど赤池からコメントがありましたが、JPCOAR の反応ということで、読み方の真意は直接関係者の方々に伺ってはいないのですが、「大枠としての方向性は JPCOAR と一致するものではありますが、具体性に乏しいので、図書館を中心とした JPCOAR としてはこういう活動を進めていきたい」という、より具体的な細かい粒度で見解が示されています。内閣府の議論としては、いろいろな議論がある中で、「内閣府は大枠を決めてくれればよくて、個別具体のことは決めてくれるな」というような議論があったということは、この場を借りてご紹介したいと思っています。

すなわち、ビジョンや大枠の方向性を決めた後は、学術情報流通や研究の将来に関する議論は現場に任せてくれというような話がある一方で、対論として、国として決めてくれなければ困るという意見もありました。そのような議論の中で、今回の統合イノベーション戦略は、大外の枠と、ベンチマーク的に数が追い掛けられるような、国立研究開発法人における研究データポリシーの策定の数のようなものを設定するなどして、達成度を測れるようにしてきたというのが現状です。

目次に沿ったガイドラインの概要

さて、では、その研究データ管理・利活用ポリシー策定ガイドラインですが、簡単に経緯をご説明させていただきたいと思います。1 月末に、赤池から紹介があった政策討議という場で、オープンサイエンスの論点が議論されました（図 10）。その後に検討を開始し、

（図 9）



（図 10）

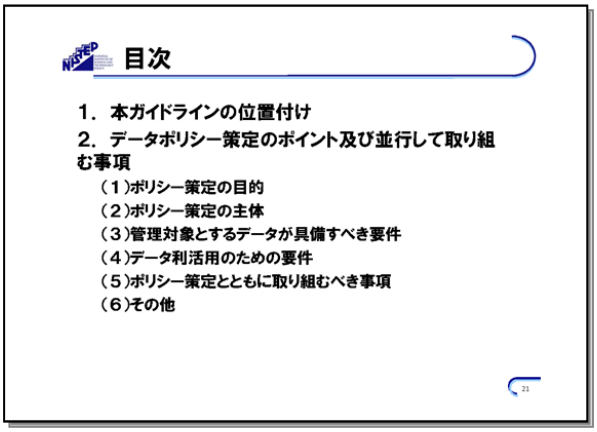
3 月に草案を作成しました。そこで、計 4 回、複数の国立研究開発法人を含む有識者による改訂を行いました。それについて 5 月に関係各機関に意見照会などをかけています。

並行して、内閣府の「国際的動向を踏まえたオープンサイエンスの推進に関する検討会」による検討も行われ、その途中で「不正競争行為改正法に伴う改訂」なども含まれて、6 月 29 日にリリースされ、それをフォローアップしているところです。

目次としては、まず「1.本ガイドラインの位置付け」があって、「2.データポリシー策定のポイント及び並行して取り組む事項」ということで、「(1) ポリシー策定の目的」、なぜこのポリシーを策定したのか、あるいは「(2) ポリシー策定の主体」「(3) 管理対象とするデータが具備すべき要件」「(4) データ利活用のための要件」「(5) ポリシー策定とともに取り組むべき事項」「(6) その他」が書いてあります（図 11）。

「本ガイドラインの位置付け」では、オープンサイエンスがそもそもどのようなものであって、政策としてはどのように考えているかということが書かれていますが、これはお題目なので、これを解説していると時間がかかってしまいます。

「データポリシーの策定のポイント及び並行して取り組む事項」は、議論の過程をお話しすると、大枠のプリンシプル（大原則）を先に決めてからガイドラインをつくった方がいいのではないかという議論がありました。その大原則に当たる部分が、この「データポリシーの策定のポイント及び並行して取り組む事項」の中に含まれているとお考えいただければと思います。そして、後半の方に、実際にガイドラインに書くべきことが書かれています。



（図 11）

当たり前なのですが、国立研究開発法人として、2 の「(4) データ利活用のための要件」のところで、「ポリシー策定にあたっては、研究分野の特性や国研のミッションを踏まえた研究データの利活用に関する考え方に基づき、公開とすべきもの、非公開とすべきもの、また制限事項を設けるか否かなどを明示する」とあり、必ずしも全部を公開、しかもフルオープンに公開することを前提にしているものではありません。

オープンサイエンスの「オープン」という言葉はいまだに誤解を持たれやすいものではありますが、研究開発法人にとって一番適切なデータの利活用の方針をまず決めましょう、それは研究開発法人であれば元々つくられたときのミッションがあるので、そのミッションに従えば自ずと導かれるであろうということになっています。

その下のポツで、図書館業界の方々には琴線に触れる「メタデータの作成」のことなどにも触れています。

その次、「(5) ポリシー策定とともに取り組むべき事項」として、研究者に過度の影響を与えないようにする、その次のポツですが、リテラシーの向上や人材の育成や創出、能力開発に努める。これはポリシーをつくるためのガイドラインとは一見関係ないようなことではありますが、先ほど申し上げた背景、プリンシプルに相当する、そのポリシーをつくることが最終的な目的ではなくて、ポリシーをつくる過程において、オープンサイエンスの推進に必要な環境整備が必要であるということをここで申し上げています。

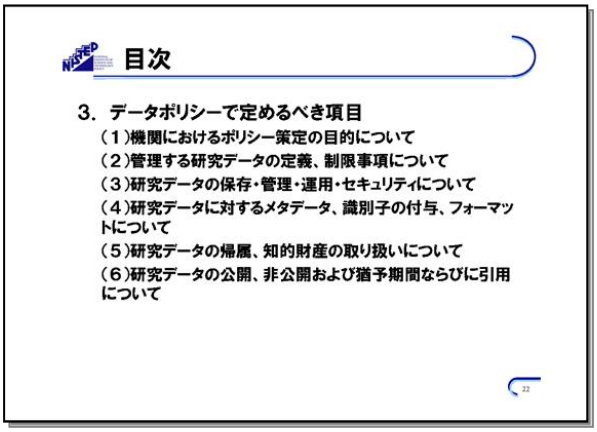
「(6) その他」では、常に必要に応じてポリシーの改訂を検討して、どんどん進めましょうということが書かれています。今時の言葉で言うと、アジャイルです。このガイドライン自体もアジャイル的につくっているもので、都度変わり得るものであるということです。あるいは、国立研究開発法人の皆さまのポリシー

もそんなに頻度高く変えるとコストがかかりますが、柔軟に対応していくべきものだろうということです。

ここまでが原則に当たるようなところです。

その次の目次は、「3.データポリシーで定めるべき項目」です（図 12）。「(1) 機関におけるポリシー策定の目的について」「(2) 管理する研究データの定義、制限事項について」「(3) 研究データの保存・管理・運用・セキュリティについて」「(4) 研究データに対するメタデータ、識別子の付与、フォーマットについて」「(5) 研究データの帰属、知的財産の取り扱いについて」「(6) 研究データの公開、非公開および猶予期間ならびに引用について」があります。ここはご覧いただくと、本当に当たり前のことしか書いていません。逆に言えば、このぐらいからまず始めていただいて、しかも、必ず全部この（1）～（6）を入れなさいというものでもありません。それも研究開発法人の特性に応じて取捨選択されてよいものではありますが、標準的にあると望ましいものということになっていると理解しています。

「(2) 管理する研究データの定義、制限事項について」が一番難しいというか、むしろ真剣に主体的に取り組むべきものであるということです。皆さんが管理すべきデータというのは一体何なのだろうか、人を含むコストが必要なので、それをかけてでも管理・保存・公開すべきデータを決めるというのは、一つ裏話をお話ししますと、これは途中のガイドラインの改訂バージョンは、「アセットデータ」を決めなさいという固有名詞化した表現を使って、しばらくのバージョンはその言葉がありました。つまり、研究者がその機関で生んだデータを資産として考えてください、そうすれば、それをどうしようかということをより真剣に考えることができるのではないかという思いを込めて入れたわけですが、議論の途中で、固有名詞化したときのディフィニションやそれが世に出た後の影響のことを考えて、最終的には取り払われたという経緯があります。



（図 12）

「(3) 研究データの保存・管理・運用・セキュリティについて」は、昨今はやはりセキュリティの問題がクローズアップされています。これも軽口をたたくわけではありませんが、各研究機関は既にセキュリティポリシーに関しては別途検討されているので、それに応じて対応してくださいということになるかと思います。

「(4) 研究データに対するメタデータ、識別子の付与、フォーマットについて」は、データポリシー策定において図書館関係者の方々が寄与する余地は十分あるのではないでしょうか。

「(5) 研究データの帰属、知的財産の取り扱いについて」も必要です。研究データには著作権は存在し得ないので、その所有や利用に関するライセンスをどのように表示するか。あるいは、CC BY であればどういうライセンスのアグリーメントを結ぶかといったことが今後必要になってくるので、そこに関する環境整備、研究機関、研究者のリテラシーを高めるようなデータポリシーが必要であるということになると思います。

「(6) 研究データの公開、非公開および猶予期間ならびに引用について」は、公開といっても未来永劫公開し続けるべきものもあれば、そうでもないというものもあるかと思います。あるいは、そもそも非公開にするけれど保存するものもあると思います。そのあたりについて、ポリシーとして定めるべきことを、繰り返しになりますが、各研究機関のミッションや研究分野の特性に応じて決めてくださいという形になってい

ます。

総じてお伝えしたいのは、これは研究者や研究機関に余計な手間をかけさせたいというものでは全くないということです。むしろ研究機関、研究者が持っているデータをアセットとして捉えて、それが不必要に流出する、その研究者・研究機関の貢献が認められないままに使われてしまう、利活用されるポテンシャルがあるにもかかわらず死蔵してしまうということがないようにするきっかけとなるガイドラインだということをお話しさせていただきたいと思います。

「事務連絡」に意外と大事なことが書いてあるので、ご紹介させていただきたいと思います（図 13）。まず、下線のところをご覧ください。ポイントを示しているものであって、個別具体的なことを本当に細かく決めるのは、各研究開発法人あるいは研究機関が自発的に行うべきものであるということです。研究分野の特性や法人としてのミッションに基づいて、自らの研究組織にふさわしい内容として主体的に定めるべきポリシーの項目もあれば、研究データの横断的連携の推進、あるいは国際的な活動の枠組みの中に入っている場合は、共通して取り組むべきものや標準的に利用しているものを採用すべきという、例えば 2 種類しかないのですが、こういうものもあるのだということを内閣府は分かった上で事務連絡を書いているという立場を取っているということになると思います。

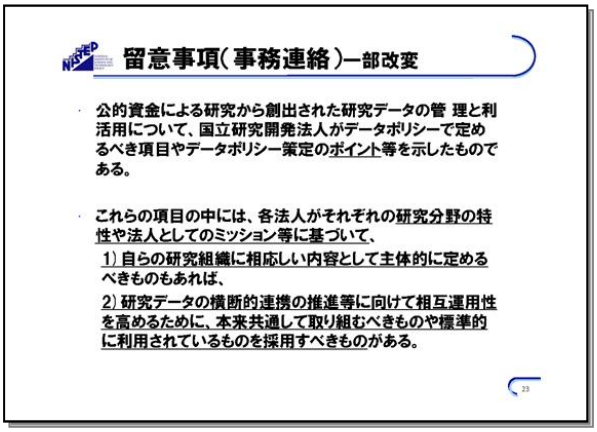
その法人においては、データポリシーに記載するそれぞれの事項の狙いや趣旨をぜひよくくみ取っていただいて、ポリシーを策定することが重要であるという話です（図 14）。

最後に、これは内閣府らしい話で、図書館業界的には直接関係ない点かもしれませんが、分野間データ連携基盤や研究開発法人等における基幹業務に関するモデルシステム等を内閣府で開発しているというのは、先ほど赤池から説明があった、「知の源泉」をつくる三本柱のことを言っています。そちらとの連携を意識しているので、それを踏まえてポリシーを制定あるいは改訂してくださいということです。

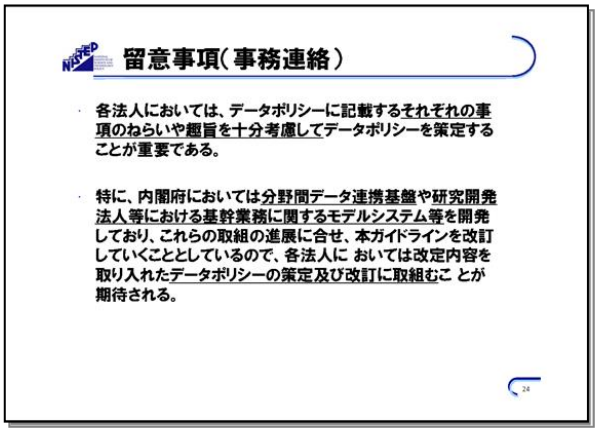
研究データ基盤として、研究者や研究者コミュニティおよびその近傍だけで使われるのではなく、産業や社会が扱うデータ基盤と相互連携をすることにより、情報の利活用をこれまで以上に飛躍的に向上させることにより社会を変えていきましょう、これが Society 5.0 を駆動する「知の源泉」の一つの構成であり、それがドライビングフォースとなっているという話になると思います。

ガイドラインが想定するもの

とはいっても、結局、ご利益は何だということで、図 15 のスライドを最後に持ってきています。研究データ管理・利活用ポリシー策定ガイドラインが想定しているのは、1 点目は、共同研究や研究費獲得で不利にならないようにすることです。ポリシーが存在していないと、共同研究が成立したときに、そのデータの帰属を相手方のいいように持っていかれる、あるいは



（図 13）



（図 14）

不利になる可能性があります。あるいは、もう欧米では動きがありますが、公的資金を獲得する際に、その所属の研究機関がデータポリシーを持っていないと申請すらできないというような流れにもなりつつあります。日本も時間の問題ではないかと個人的には思います。

2 点目は、研究データの適切な管理体制とリテラシー向上です。これは不必要な流出を避けるだけではなく、死蔵しないという目的もあると思います。

3 点目は、残すべき研究データに関する研究者・研究機関の主体的な議論と役割分担です。内閣府が「これを残してください」と言ったり、文科省が決めたりということにはならないはずです。データのことを一番分かっているのは研究者あるいは研究機関なので、そこが主体的に決めるべきであり、誰かが管理、保存コストを負担するわけですから、それをしっかり議論する必要があります。

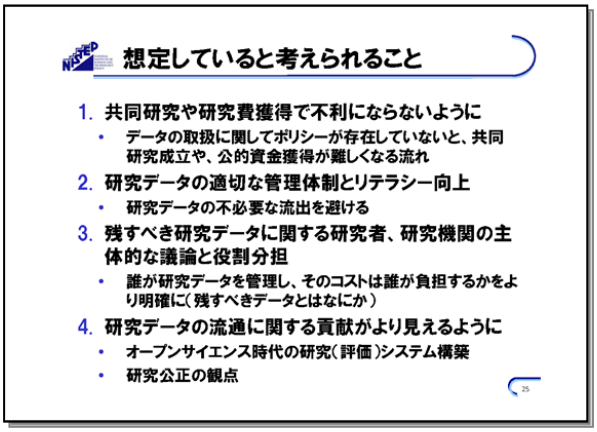
4 点目は、これはだんだん粒度が大きくなるように書いたのですが、オープンサイエンスの一番遠いビジョンである、研究データの流通に関する貢献がより見えるようにすることです。オープンサイエンス時代の研究システム、研究評価システムが出来上がる、あるいは、研究公正の観点から、論文だけに依拠し過ぎる研究評価や学術システムではない、より健全なオープンサイエンスパラダイムに向かっていく、それを狙っているということです。

どちらかというとトップダウンの施策で内閣府からガイドラインが出ています（図 16）。統合イノベーション戦略が上に出来上がって、科学研究を変容させ、社会を変革する知識基盤構築としての「知の源泉」をつくるためには、ボトムアップからの活動も必要ではないでしょうかということです。

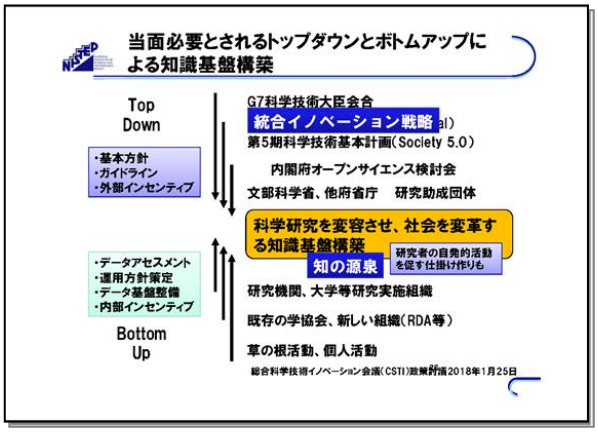
●林賢紀　ありがとうございました。それではフロアから質問を受けたいと思います。

●フロア 1　関東学院大学の職員です。赤池さんにお伺いしたいのですが、私の大学の研究者と近いところで仕事をしているのですが、教員と話したときに、オープンにしたいのだけれど忙しくてそこまで手が回らないというような話を聞いています。忙しい先生方にオープンに向けて考えていただくインセンティブについて何かお知恵があればお伺いしたいです。

●赤池　今、一番大事な側面だと思います。先ほど林が「トップダウン」と言いましたが、そういう規定だけではなく、先生方のお時間をということで、例えば、オープンサイエンスの側面ではないですが、各大学において URA、事務職員にこういうことに対するリテラシーを増やしていただくというような活動もさまざまさせていただいています。今回の統合イノベーション戦略の中では、特に事務職員、URA、図書館、その



（図 15）



（図 16）

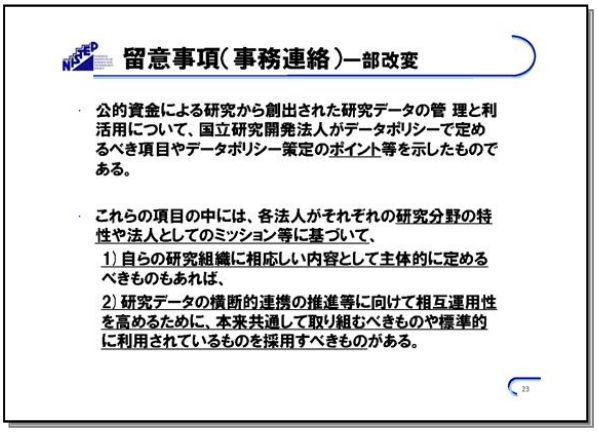
他の関連の方も含めた人材育成、それから先生方の理解という側面的なことにも力を入れてやっていきたいと考えています。

また、これはハードの側面ですが、国立情報学研究所や各大学が連携してそういう基盤をつくって、そもそもリポジトリに登録したり、それを管理をしたりという作業をできるだけ合理的に、簡潔にできるようなシステムの開発という側面も大事だと考え、そういう側面も文科省と協力してやっていっていますので、少しずつですが改善していきたいと考えています。

●**フロア 2**　国立極地研究所の図書館員です。林さんにお伺いしたいのですが、留意事項、文科省からの事務連絡の 2 番目で、「横断的連携の推進等に向けて相互運用性を高めるために」等々のポイントがあったと思います（図 13）。相互運用性を高めるために本来共通して取り組むべきものは結構いろいろな切り口であるとは思うのですが、それに関する分かりやすい事例があれば皆さんつくりやすい気がします。何かお示しできるような取り組み等があればぜひ教えてください。

●**林和弘**　そのご要望は事前に国立研究開発法人向けにガイドラインを紹介したときにも頂いており、まず、やったこととしては、既に出ている日本語のデータポリシーの実例をお付けして、国立極地研究所のデータポリシーも紹介させていただいているという状況です。

それと併せて、まだコミットはできませんが、議論



（図 13）

の方向性としては、ガイドラインを解説する資料の作成という話も出ています。その中ではまさに、今おっしゃられた、それぞれに関して具体的にこういう例があるというものをお示しする。ただ、高エネルギー物理などはかなり閉じた世界なのでその研究者の中で済むのですが、地球環境系や天文などのビッグサイエンスの場合はそもそも協調して研究が進むので、そのデータポリシーに関しては国際的な監修を踏まえたポリシーづくりが必要になるという大枠はあるかとは思います。あとは個別には解説でお示しできるようになればいいのではないかと思っています。

●**フロア 3**　農業・食品産業技術総合研究機構の研究者です。赤池さんにお伺いしたいのですが、昨今、国研も交付金が減って、だんだんお金が足りなくなってきている中で、どうしてもデータをあまりオープンにしないで、企業との共同研究契約などを結ぶことによってデータを取りに行けというような風潮にうちの法人はなっています。それを言っているのが総合科学技術・イノベーション会議の前議員だというのがちょっとどうかと思うのですが。

ただ、そういうことも必要だと思うので、民間の資金を得るという必要性と、オープンにしていく必要性のバランスをどう取ってやっていけばいいのかということをもう少し研究開発法人にインプットしていただかないと、多分、研究開発法人がばたばたしてしまって対応できないのではないかと感じています。その辺でもし何かあればお話しいただければと思います。

●**赤池**　大事なお話だと思います。直接の答えになるか分からないのですが、多分、内閣府の今回のガイドラインも含めて、「これをやってください」「このためにもっと仕事を増やしてください」という意図ではなくて、例えば民間との関係性も含めて、まず法人の経営があって、その研究所が持っているデータというリソース、それから企業が持っているリソースをどう組み合わせて法人としての経営を合理的にやっていくか

という側面があるのではないかと思います。

　そのあたりはよくご存じの理事長だと思いますので、限られたと言うとあれですが、必要なリソースを投入して、最も効果的な経営をする中で、民間企業とどういうお付き合いをしていけばいいのかなというのは逆にそういう順番で決まっていくのではないかと思います。先に企業との研究があって、そこから何でもかんでも取ってこいという話ではないのではないかという印象を持っています。当然、トップも熟知されているとは理解しています。

第 1 回 SPARC Japan セミナー2018

「データ利活用ポリシーと研究者・ライブラリアンの役割」

国立環境研究所データポリシーと データ公開に向けた試み

白井 知子
（国立環境研究所 地球環境研究センター）

講演要旨

国立環境研究所では、幅広い分野の環境研究を学際的かつ総合的に進めている。環境研究は多岐に渡るため、使われるデータも分野ごとに特徴があり、データを公開することへの関心や懸念も異なる。そのような状況の下、研究所は、平成 29 年 4 月、「国立環境研究所データの公開に関する基本方針（データポリシー）」を策定、公開した。データポリシーの紹介および策定に至る背景、研究所におけるデータ公開に向けた試みについて紹介する。

白井 知子

国立環境研究所地球環境研究センター 主任研究員。1999年東京大学理学系研究科博士課程修了。同年宇宙航空研究開発機構地球観測研究部門にて、上空大気の航空機観測や、宇宙からの地球観測センサ等を担当。2001年～2002年、米カリフォルニア大学アーバイン校に滞在し、越境大気汚染等について共同研究を行う。2004年以降、国立環境研究所にて、温室効果ガスの排出・吸収状況と、その大気中での挙動についての研究を行う。2011年からは、地球環境データのデータベース化やデータ利活用の推進にも携わっている。2018年4月～9月、企画部研究企画主幹兼務。

国立環境研究所（NIES）は、国立研究開発法人として既にデータポリシーを持っている数少ない研究所です。本日はその策定の経緯と論点、現場の例、NIES においてデータを公開していくときの課題について述べたいと思います。

自己紹介・個人的動機

私は大気化学の研究をしており、最初は真空装置を使って微量成分の分析をしていました。大気球実験で成層圏の空気を採って分析など、だんだん内容が広がってきて、データ解析の魅力に取りつかれました。その後、旧宇宙開発事業団（NASDA、現 JAXA）に就職し、地球観測部門で、航空機観測実験をしたり、衛星のプロジェクトに関わったり、アメリカの大学に滞在して共同研究をしたりしていました。

主に観測に携わっていた際、大量に得られるデータを使い尽くせないところにフラストレーションを覚え、だんだん解析や計算などを自分でやるようになってきて、両方ずっと続けることは大変なので、今は主に解析や計算をしているという状況です。

一方で、私は「地球環境データ統合解析推進室」というところに所属があり、主に観測により得られたデータのデータベースを作って発信していくという業務も行っています。また、4 月から半年間、企画部も兼務しており、今回ご紹介するデータポリシーは企画部中心に作られたものです。

個人的動機としては、観測とモデルと両方やっていると、観測の方は観測で忙しいし、論文も書かなけれ

ばいけないし、次の観測のプロポーザルも書かなければいけないということで、とにかく余裕がないです。一方、モデル研究者は、使いやすいデータはないかといつもサーチしています。どういうものが使いやすいデータかというと、欠損がなく、グリッド化されているものです。観測をしている人は、他人のためにデータ整備や発信をするのはどうも後回しになってしまいますし、時空間欠損がない観測というのは非常にまれです。かといって、モデラーがこういったデータに手を付けて整備するかというと、世界を見回すと、整備されたデータセットが見つかることが多いので、一番整理されているデータがいつも使われることになり、日が当たらないデータはずっと日が当たらないという状況になります。そのような背景から、そこをつなぐことに私は非常にやりがいを感じています。

NIES としてのデータポリシー

NIES はつくば市にあり、九つの研究実施部門と三つの管理部門があります。

（http://www.nies.go.jp/sosiki/index.html）

私がいるのは研究実施部門の一つである地球環境研究センターです。今回、ご紹介する、研究所のデータポリシーは主に管理部門が中心になって作ったというところが一つポイントです。

策定のきっかけは、ある研究成果に係る全ての情報について、総務部に開示要求が来たことでした。開示要求については、法人文書に関しては情報公開法で決められていますが、研究データに関しては何をもって「全ての情報」とするかの定義がはっきりしていません。実験ノートに至るところから全てのデータだと言いはじめるときりがないですし、担当部門で対応に困ったことがあり、同様のケースに備えるとともに、研究データの公開を促進するためにも、研究データの公開について定めたデータポリシーを策定しておくことが非常に大事であるということから、策定への流れになりました。

策定の経緯は、平成 28 年 9 月に理事から企画部にデータポリシー作成の指示が来て、10 月に企画部で素案を作成し、11 月にオープンサイエンスとの切り分けについて、オープンサイエンスを別途進めている環境情報部と相談し、整理しました。12 月に所内幹部で議論した後、平成 29 年 2 月に方針および原案を所内に周知し、意見を募りました。その際、データの帰属・利用条件に意見が集中しましたが、これについては後で詳しくご紹介します。3 月に弁護士に確認し、4 月にデータポリシー策定・公開となりました。

ポイントは、半年ぐらいで作られたということです。やる気になればできるという言い方もできますし、そうやって作られたデータポリシーなので、先ほどの内閣府のデータポリシーガイドラインのご発表を聞きながら、NIES のデータポリシーがどの項目をカバーしているかをチェックしてみたのですが、カバーしているものもあれば、全くしていないものもあり、部分的にカバーしているものもあるという状況です。

企画部、環境情報部、総務部で意見交換して決まったことは、「データポリシーはふわっとしたもので、NIES の姿勢を示す所信表明のようなものとする」「研究不正の観点でデータを公開するのとは異なり、外部で積極的にデータを利用してもらうための公開を対象とする」「データの技術的な方法ではなく、理念を示すことが主眼」「細かな規定は別に定める。最小限必要なことを記載する」「オープンサイエンス推進のために情報部で検討している内容は、論文等を公開する機関リポジトリをはじめ、扱う範囲は幅広いので、オープンサイエンスの方向性とデータポリシー策定は方向性が同じということで、矛盾していなければ、各々議論を進めてもよい」ということです。そして、オープンサイエンスについては別途ワーキンググループを設けて検討することになりました。オープンサイエンスの方が進みが遅いので、もし一緒に進めていたら、データポリシーができるのはもう少し遅れていた気がします。

データポリシーの素案策定のときに参考にした文書やポリシーは、2015 年の JST 科学技術情報委員会の

「データシェアリングのあり方に関する提言」など、図 1 に示した通りです。機関全体としてのポリシーはなくても、「○○センター」というレベルではポリシーがあるところが比較的あり、そのようなところの個別データポリシーなども参考にしながら論点を絞っていき、素案を作りました。

データポリシー素案作成の論点は、例えば「1.目的」では、「積極的にデータ公開し、国の研究開発成果の最大化に資する方針でよいか」でした。決まった方針は、「政府の基本方針はデータが自由に利用できること。NIES のデータは税金で作成したものなので、自由に使ってもらうことが大原則。ただし、状況により商用利用や改変への制限はあり得る」でした。

「2.公開するデータの範囲」についての主な論点は、「公開するデータの範囲は適当か」でした。これは、策定のきっかけとして、「公開範囲に制限を設けないと対応に困る」ということがありましたので、「公開するデータの範囲にやや制限を加える」という方針で進めています。

「3.データの品質・管理」についての論点は、「品質という曖昧な言い方でよいか。精度、真度、利用しやすさなどに言及するか」「トレーサビリティについて言及するか」で、基本方針は、言及するということにし、弁護士のアドバイスで「非保証」を入れました。

「4.データの帰属・利用条件」は、所内で意見を募った際に、一番関心が高かったところです。論点として、「知的財産権に言及するだけでよいか」「他機関との共同研究が行われたケースについて、言及する必要があるか」「データを利用する際に、引用元（NIES）を示すことを義務として記載するか。その記載例を示すか」「データの商業利用について、利用料を課すなどの条件を加えるか」について話し合いました。

データポリシー素案作成 step0

参考にした既存の文書・ポリシー等 （表の項目は一部）

機関	JST科学技術情報委員会	日本学術会議情報学委員会国際サイエンスデータ分科会	水産研究・教育機構	産業技術総合研究所地質調査総合センター	農業環境技術研究所農業環境情報センター	宇宙航空研究開発機構・国立環境研究所・環境省
表題	データシェアリングのあり方に関する提言	オープンデータに関する権利と義務	環境調査データポリシー	研究成果情報のデータポリシー	サイトおよびデータ利用ポリシー	GOSATデータポリシー
設定時期	2015年4月	2014年9月30日	2016年3月2日 2016年4月1日改正			2008年11月10日 2009年4月16日改訂 2015年3月13日改訂
データの範囲	シェアリング範囲		データの定義	用語の定義（データの範囲）	サイトへのリンク	
提供機関の権利		データに関する権利	データの帰属	保護される権利・帰属第三者に帰属する権利		
提供機関の義務	品質保証方法	データに関する義務	データの取得、保管、管理、整備			データ配布に関わる使命
データ提供期間	データ整備ルール			公開の開始・終了	格納データを予告なく変更	データ配布時期
⋮	⋮	⋮	⋮	⋮	⋮	⋮

（図 1）

まずは、知的財産権についてですが、データというのは著作物とは違って、著作権がはっきりと定められておらず、単純なデータの羅列は著作権法の保護対象ではありません。体系的に整理されたデータベースであって、創作性が認められるものは保護対象となるのですが、過去の裁判例でも著作権が認められたり認められなかったりと、曖昧なので、そこに触れはじめると記載が煩雑になるということで、著作権についてはあえて触れていません。

また、公開したデータがこちらの意図しない形で二次利用されたり、商業利用されたりすることを避ける観点から、「無断転載や二次配布は、別に定める場合を除き認めない」と、基本的にはフェイルセーフの方針を取っています。ただ実際には、公開中のデータの中には「積極的に配布していいよ」「公開していいよ」というものもたくさんあり、そのようなものについては、別途「いいですよ」と規定で定めています。基本は「認めない」となっているということが特徴かと思います。また、データ利用時には、「出典明記はお願いするが、通常は金銭的保証は求めない」という方針で進めました。

帰属については、議論が非常に分かれるところではあったのですが、「職務上取得したデータの権利は、研究所（研究所に所属する個人を含み得る）に属する」ということにしています。要するに研究者個人に付くのではなく、あくまでも研究所に帰属するという形で記述しています。

「5.データの公開期間」については、「データ公開期間に関する項目は必要か」「公開の開始時期の遅延、公開中止の理由として挙げた例文は適切か」が論点になりました。ここについても弁護士からアドバイスをもらって、「合理的な理由のある場合には」という文

言を入れました。

こうしてできた国立環境研究所のデータポリシーは以下から公開されています。

（https://www.nies.go.jp/kihon/kitei/kt_datapolicy.pdf）

このポリシーを策定したのは管理部門ですが、データの公開は実質的には各研究部門・研究者が中心となって行われています。その際、データ公開ポリシー（ライセンス含む）はデータごとに定めています。研究データ公開に関して、機関共通で実質的効力を持つポリシーを細かく定めてしまうと、多様なデータがある機関の場合は、そのポリシーを決めるだけでも非常に大変です。今の NIES のデータポリシーは、各データ提供者の意思で別途詳細を定められるようになっているので、「情報公開法との切り分けを明確にする」「データ公開を促進するベースを提供する」という目的を素早く達成することができたと思っています。

ただし、一方で、初めてデータを公開する担当者から「こういうデータを公開してもよいか」「公開の際、どこに許可を取る必要があるか」「このようなデータポリシーで公開したいが、管理部門に確認・判断・決裁が必要か」といった問い合わせが寄せられたり、「データポリシーは誰が決めるのですか」という質問があったりするほど、データポリシーに関する意識がそれほど高くない現場もあります。実際にこのデータポリシーを運用していくに当たっては、現場ごとに必要なデータポリシーを追加で定められるような、要するに、所全体の大枠のデータポリシーに基づいて、各データの細かいところを定められるような教育、支援が必要という状況です。

例として、地球環境データベースは 2016 年 9 月から DOI 付きでデータを公開しており、利用規約は、クリエイティブ・コモンズ・ライセンスまたは独自で作成した規約からの選択としています。独自ライセンスの例（図 2）では、「配布の条件」で「有料で販売してはならない」としたい人はそうしますし、「利用者は、再配布に際し、再配布により本データセットを受け取る者に対し、本データセットが本利用規約の下で提供されていることを通知しなければならない」とすることもできます。「出典の明記」では、作成者が出典の記載を削除するように要求できるようにもできます。もちろん「引用の書式」も独自で決められます。研究者がクリエイティブ・コモンズなどをあまり使わない一番の理由は、不特定多数に向けてデータの再配布や成果物の公表をしたときには報告してほしい、何かデータセットに問題があった場合は報告してほしいといった、データ利用者からの利用報告や問題報告、フィードバックが欲しいというケースがほとんどです。そのあたりをうまく拾っていけるようなポリシーが必要なのではないかと認識しています。

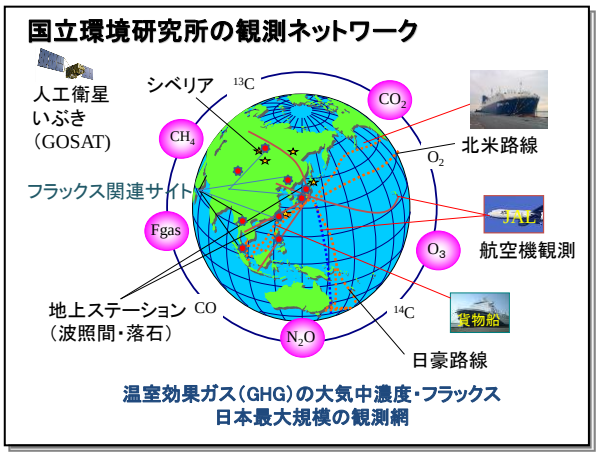
このように、大枠のデータポリシーの他に、現場で細かいことをデータごとに決めているということをご紹介しました。

利用規約の例：地球環境データベース (GED)

2016年9月からDOI付きでデータを公開。利用規約はCCライセンスまたは独自で作成。

許諾の範囲	利用者は、本データセットの内容を改変することなく第三者に再配布すること、および、本データセットをもとにして新たに知的生産物 (以下、派生物) を作成して公表することができる。
配布の条件	本データセットを第三者に再配布する場合、有料で販売してはならない。 本データセットを本利用規約に基づいて第三者に再配布した場合、再配布されたデータセットにも本利用規約が適用される。 利用者は、再配布に際し、再配布によりネデータセットを受け取る者に対し、本データセットの原配布元のURL、本データセットが本利用規約のもとで提供されていること、および、本データセットは作成者によって随時更新されていることを通知しなければならない。
最新版の利用	利用者は、特別の理由がない限り、本データセットの最新版を利用しなければならない。
出典の明記	利用者は、派生物に、その派生物が本データセットを利用して二次的に作成されたものであることを明記しなければならない。 ただし作成者は、利用者に対し、派生物から出典の記載を削除するように要求することができる。その場合、利用者は、合理的に実施可能な範囲で、派生物から出典の記載を削除しなければならない。
引用の書式	本データセットを引用する場合は、原則として、以下のように表記しなければならない。 向井ら(2014)、波照間島における大気 CO_2 濃度の連続観測データセット、Ver. x.x (*1)、国立環境研究所 地球環境研究センター、DOI:10.17595/20160901.001. (参照: YYYY/MM/DD)(*2) (*1) バージョン番号はファイル名に記載されている。 (*2) 参照年月日は、ファイルをダウンロードした日付を記載すること。 ただし、掲載誌の投稿規程等により引用の書式が定められている場合は、その掲載誌の規程に従うものとする。
利用報告	再配布、および成果物の公表が、不特定多数に向けて行なわれた場合、利用者はデータ提供機関にそのことを速やかに報告しなければならない。
問題報告	本データセットの内容について、不備や誤り、またはその疑いのある記述を発見した場合、利用者は、その情報を速やかにデータ提供機関に報告しなければならない。
無保証及び免責	作成者およびデータ提供機関は、本利用規約に基づく本データセットの利用に関して一切の保証をしない。 利用者は、派生物の内容について、作成者またはデータ提供機関が何らかの保証を与えているかのような表現を、作成者に断りなく行なってはならない。 本データセットの利用によって何らかの損害が生じた場合、いかなる損害であってもすべて利用者が責任を負うものとする。

（図 2）



（図 3）

現場の例：地球環境研究センター

現場の例として、私が所属する地球環境研究センターの取り組みについてご紹介したいと思います。地球環境研究センターは 1990 年に発足しました。人工衛星、船舶観測、航空機観測、地上ステーションなど、日本最大の温室効果ガスの観測網を持っており、そのデータの発信を主に地球環境データベースで行っています（図 3)。

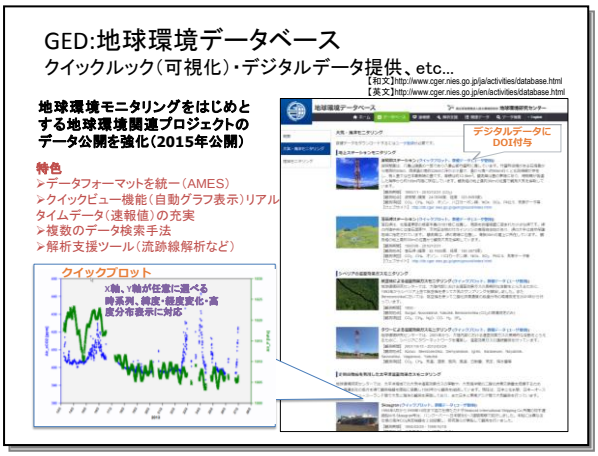
2010 年代前半までは、大気中温室効果ガス濃度や、海や陸域におけるフラックスといったデータは、アメリカなどに本拠地がある国際的なデータベースを通じての提供や、個別にリクエストをもらった研究者からの提供がほとんどでした。

2011 年からデータ統合・解析システム（DIAS）の 2 期に参加させていただき、このときにメタデータのみ DIAS に投入し、公開したのですが、実データをデータベースから提供ということはしていませんでした。ということで、データ公開用に地球環境データベースを構築し、2015 年から公開しました（図 4)。

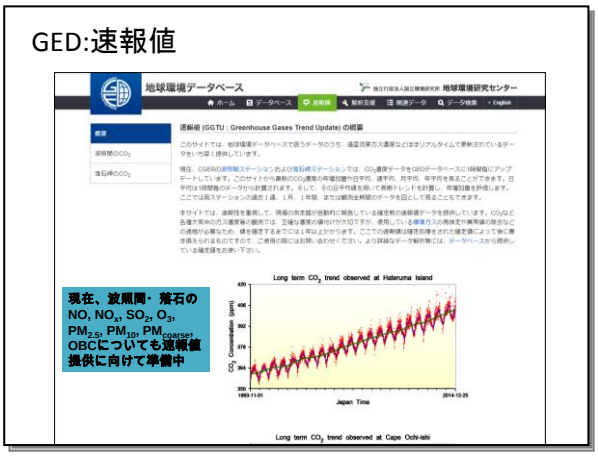
地球環境データベースでは、地上ステーションや船などで観測したデータをダウンロードできるようになっており、ユーザー登録が必要なデータもあるのですが、クイックプロットができたり、データフォーマットが統一されていたり、今まで個別に研究者が抱えていたデータを一括して提供できるようなものになっています。

例えばデータ検索なども、地点検索、メタデータ検索など、いろいろな検索ができるようになっていたりします（図 5)。検索については、現在、改良中です。

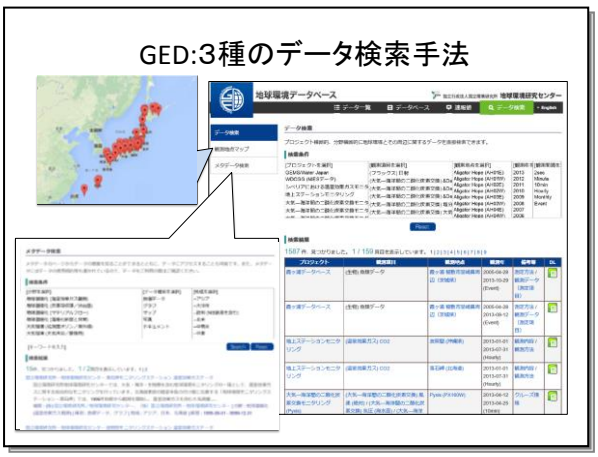
また、速報値といって、そのページに行けば今日の値が表示されているというような、沖縄の波照間島と北海道の落石岬で観測している今の大気中の CO_2 の



（図 4）



（図 6）



（図 5）

（図 7）

濃度がリアルタイムで見られるようなサイトがあります（図 6）。それらの値を別のサイトで表示したいという場合、API を提供しています（図 7）。

また、トレンド解析、トラジェクトリ、フットプリントのような、一般の方は使うかどうかは分かりませんが、学生や研究者がよく使う解析ツールも提供しています。これらは DIAS からも公開しています（図 8）。

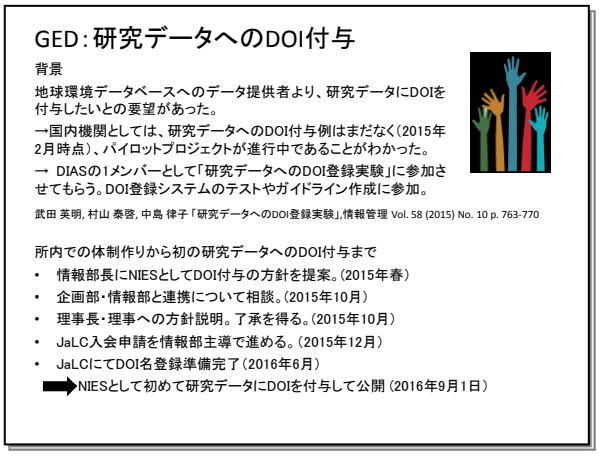
情報提供サイトも作っています。例えば、アジア太平洋の温室効果ガス関係のモニタリングサイトをまとめて、検索して一括表示できるようなものも作成しています（図 9）。

研究データに DOI を付与する流れができていることは皆さんも十分ご存じだと思うのですが、NIES でもその試みを進めています。国立情報学研究所の武田先生を中心に進めてくださった、研究データに DOI を付与するプロジェクト「研究データへの DOI 登録実験」に 2015 年から参加させていただきました。研究所の中で、NIES として研究データに DOI を付与する方針を提案し、2016 年 9 月から、NIES として研究データに DOI を付与して公開を始めました（図 10）。

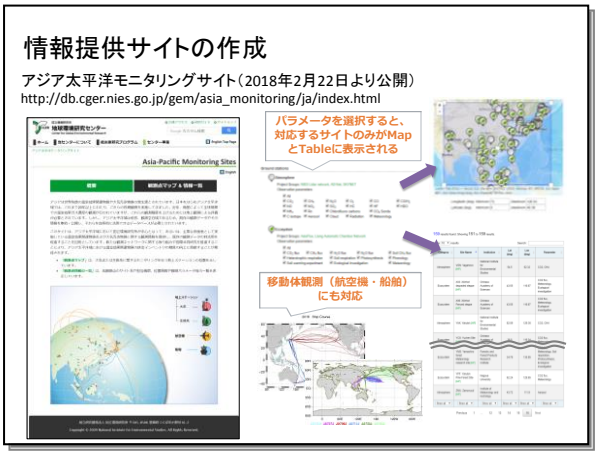
その後も順調に、地上ステーションのデータや、モデルで計算した二酸化炭素フラックスマップ、富士山頂で観測した大気中二酸化炭素濃度、民間航空機で観測している二酸化炭素濃度、オゾン層衛星観測センサのデータも、DOI 付きで公開しました（図 11）。

このようなデータセットはどんどん使ってほしいデータであり、DOI を付与した公開が、データを提供することが評価されるということにつながっていくといいなと思ってやっている面もあります。また、DOI を付けることでメタデータをきっちり作ってもらえるようになったという他のメリットもあります。ですから、率先して DOI を付与した研究データ公開を進めているところです。

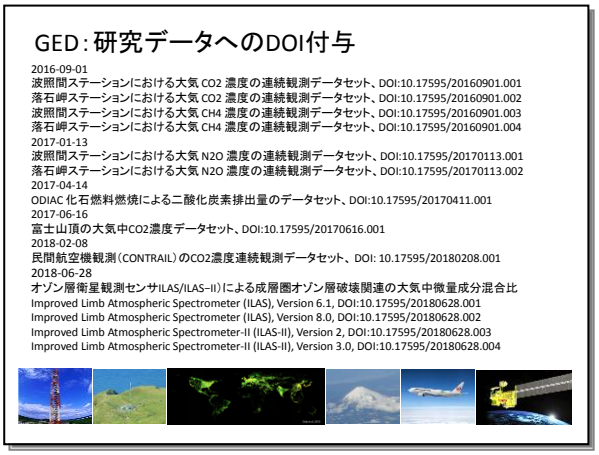
（図 8）



（図 10）



（図 9）



（図 11）

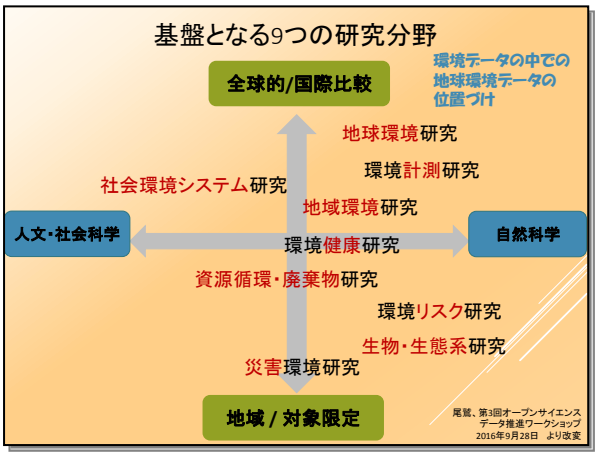
NIES におけるオープンデータの課題

NIES には非常にたくさんの分野があり、いろいろな研究をしています。基盤となる九つの研究分野として、地球環境研究、環境計測研究、地球環境研究、環境健康研究、資源循環・廃棄物研究、環境リスク研究、生物・生態系研究、災害環境研究、社会環境システム研究があります。これらは、全球的で国際的な研究か、地域的で対象が限定された研究かという縦軸と、人文・社会科学系か、自然科学系かという横軸のマトリックスで分類することができます（図 12)。その中でどこに位置するかによって、データが大きく異なります（図 13)。自然科学系はモニタリングや実験をしますが、人文・社会科学系はアンケートを取ったりするので、データそのものがどう出てくるかも違い、気になることが全く違ってきます（図 14)。

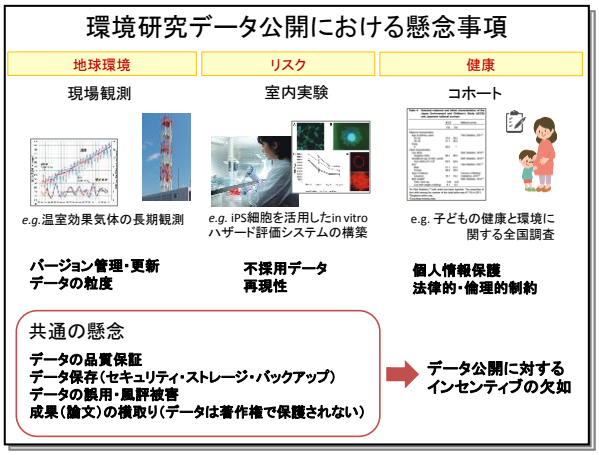
地球環境分野の現場観測であれば、バージョン管理・更新、データの粒度が問題になりますが、室内実験の場合は、失敗したデータはどうするのか、再現性をどうするのかということが問題になってきます。健康分野のコホートでは、個人情報保護や法律的・倫理的な制約が問題になります。

その一方で共通の懸念もあり、それはデータのクオリティコントロール、データ保存（セキュリティ・ストレージ・バックアップ）の問題、データの誤用や風評被害の問題、成果の横取りです。これらがデータ公開に対するインセンティブの欠如につながっています。

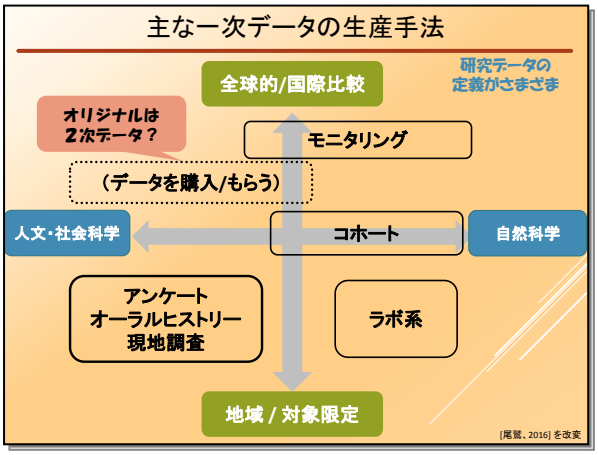
データの汎用性については、私が今携わっている地球環境データは、環境データの中では非常に全球的で、国際的にデータを使い合っていくような分野であるということ、また、自然科学系なので、ある程度データのフォーマットなどがそろっていることが多いこともあり、環境データの中ではオープンデータが進みやすい、モチベ



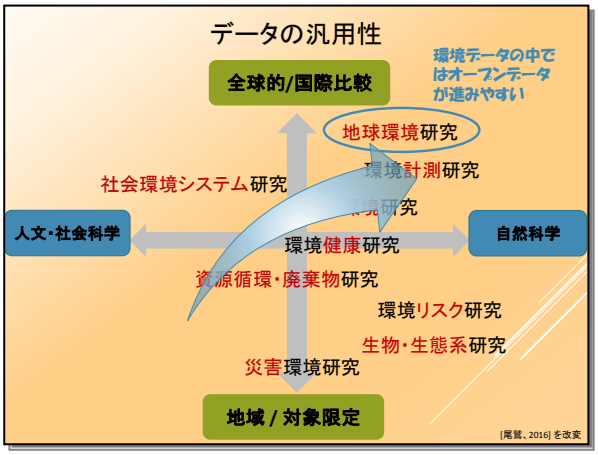
（図 12）



（図 14）



（図 13）



（図 15）

ーションが湧きやすい部類に入ります（図 15）。ところが、研究所の他の分野はもう少しハードルが高いことが多いので、研究所全体でなかなか一緒には進めないというのが現状です。

「研究現場におけるオープンデータに向けた課題」をまとめてみました。一つ目として、動機不足があります。そもそもデータをなぜ公開する必要があるのか分からない、成果を出してからデータを公開したいのに、なぜ今、データを出さなければいけないのかという疑念から、ずっとデータを抱えてしまいます。これについては、Schmidt ら（2016 年）、池内ら（2017 年）による、アンケートの結果で、データを公開する際の問題として「自分の成果を出してからデータを公開したい」という割合が非常に多いということが分かってきています。メリットがまだ分からない、未知の恐怖のようなものがあります。

二つ目は、知識不足です。学術コミュニケーションについての知識不足があり、メタデータ、著作権、DOI、研究データマネジメントなど言葉はよく聞くけれど、一体何をすればいいのかが全然分からないということもインセンティブを下げてしまいます。そのあたりは図書館員の皆さんに役割を発揮していただける部分ではないかと思っています。

三つ目は、リソース不足です。データを公開したいがリソースが足りない、これは皆さん深刻だと思うのですが、ストレージ容量や IT 人材、時間がありません。セキュリティ対策、バックアップといったデータ管理への支援不足、そのようなことを考えていくと、なかなか手が動きませんし、やりたいと思っても一人ではできないということがあります。

（図 16）

このような課題が現場では非常に大きいです。これらを全部クリアしたとしても、四つ目に、公開したデータが有効に利活用されているかという課題があります。例えば、データの見つけやすさ、アクセスの容易さ、データの品質情報・メタデータが明示されているか、データ利用条件の簡素化・無償提供をしているか、そのようなポイントが幾つかあると思うのですが、データを公開したことで満足してしまって、本当にそれが有効に活用されているか、本当はもっと使ってもらえるのではないかというところまではいっていないのが現状ではないかと思います。

今後の課題のまとめです（図 16）。現場レベルでは、一つ目は研究者の理解・意識の向上です。二つ目は、データ公開ポリシーの策定です。まだ NIES もあまりやっていないものとして、保存や公開すべきデータ、期間等についても決めた方が良いでしょう。三つ目は、運用体制で、人員・予算の継続的確保です。これが非常に難しいということがあります。四つ目は、データ管理に必要な作業のシステム化です。これは NIES でも今行っていますが、非常に地道な作業です。五つ目は、データの品質管理を行う仕組みづくりです。これは高度な話です。六つ目は、データ公開の簡易化・迅速化です。これはシステム化とも関係しています。七つ目は、データ利用の促進です。

これらの課題を考えるときに、他の民間のサービスなどがいろいろあり、スピードや手軽さでは競えないことがあるので、そのようなところも考えながら研究所の現場はやっています。

社会・コミュニティレベルの課題としては、国レベルの共通基盤整備計画がいろいろあるのですが、利用者が誰なのかがあまりはっきりしていないものもあり、大きな船に乗れるのかどうかが関心事の一つではあります。また、研究者のインセンティブ向上には、ファンディングエージェンシーや関係省庁の姿勢や方針が

非常に効いてきます。データ出版に対する評価がもう少し上がれば少し変わってくると思います。また、ビッグサイエンスや AI、ビッグデータ解析などとの連携がどんどん進んでいくと、もう少しデータを整備しようという気持ちになっていくと思うので、そのようなデータ駆動型科学の推進があるのではないかと思っています。そして、データの相互運用性の向上があります。

データポリシーは作ってあると非常に有用だと思います。それがデータを使っていくきっかけになることは確かだということをメッセージとして残したいと思います。

●**林賢紀**　ありがとうございました。それでは質問を幾つかお受けしたいと思います。

●**フロア 1**　東京大学附属図書館の図書館員です。データポリシーの策定の経緯から実際の使われ方までご説明いただきまして、どうもありがとうございました。ポリシーの策定から 1 年半経過して、その間に内閣府から、「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」が出ています。1 年半運用されてみて、あるいはガイドラインがこのように国から出た状態で、NIES のデータポリシーに関して、改訂の必要性は認識されていらっしゃるか、あるいはその運用の中で研究者からどのような意見が出ているかということを伺えればと思います。

●**白井**　とてもポイントを突いたご質問だと思います。

最初の、内閣府のガイドラインを見て改訂の必要性が感じられるかですが、それは非常にあるとは思っています。先ほど、私も最初に NISTEP の林さんのご発表を聞きながらチェックを入れたりしたという話をしたと思うのですが、全く触れていない部分などもあります。ただ、やはり内閣府としても別に全部チェックしろという意味で作られたわけではないということなので、もちろん全部を網羅するための検討をしなければいけないとは思ってはいません。一方、ガイドラインができたことで研究所での議論はしやすくなったというのはあるので、優先順位を付けて、少しずつ検討していない部分を検討して、データポリシーを改訂していくということはもちろん考えているところです。

それから、データポリシーに関する研究者からの反響ですが、あまり聞こえてきてはいません。このデータポリシーは本当にふわっとしたものなので、その先の現場で作成する細かいポリシーやライセンスについては、もちろん議論が白熱することも現場ではあるのですが、研究所ポリシーについては特に利害が反するところがないというのもありまして、議論にはなりません。平和なポリシーです（笑）。

ただ、内閣府でこれについて発表させていただいたときに、科学技術振興機構（JST）の方から共同研究のときのことをいろいろ聞かれて、そのあたりについてはもう少し書き込んでおいた方が組織としていいのではないかという認識は持っています。

●**フロア 2**　横浜国立大学（SPARC Japan 運営委員）の深貝です。社会科学系がベースで、思想研究の側なので、自然系の方の感覚を伺いたいと思いまして。白井さんのお話の最後の方で、研究者がデータをなかなか出したがらないという、当然といえば当然の話を頂きました。データポリシーを使ってなるべく公開に向けて促して、やがては赤池さんの話の Society 5.0 まで行こうという遠大な話です。実験系の方だと、善しあしはともかく、データを隠したいとなりがちだというのはそうだと思うのです。特に資金をどこから出してもらったかを含めた上で、特許や発明などいろいろなことをしていますから。それでもなおデータをオープンにしてほしいということを今やっているわけですが、観測系の方ですと、観測データが多くなることでようやく分かることがありそうで、傾向的変化という場合に、わずかな観測ではとても足りないから多く集める

必要があるわけです。そのような意味では、地球環境の分野では、データを出すことにもっと積極的であっていいのではないかと思うのですが、どうでしょうか。

それから、自分がたまたま非常に珍しいデータを観測してしまった場合はやはり隠したくなりそうですが、研究者はどう振る舞うのでしょうか。

●**白井**　1 番目の質問に関しては、地球環境データのようなものはオープンデータと親和性が高く、究極をいっているのは天文の分野で、データを使い合わないとサイエンスできないということで、分野内だけですが、ものすごくオープンデータが進んでいるようです。

地球環境の場合はその中間というか、使い合った方がいい分野とそうではない分野がまだあるのですが、やはりインセンティブとしては非常に高いですし、皆さん、研究の過程で他の人の作ったデータセットのお世話に必ずなっていることが多い分野なので、そのような意味では、データをオープンにしていくことへの認識は高い分野だと思っています。ただし、認識は高いけれども、実際に手を動かすかというのはまた話は別というのはあります。これが 1 問目の答えになります。

2 番目の珍しいデータに関しては、戦略的なオープン・クローズの判断になると思います。珍しいものはやはり自分で最初にしゃぶり尽くしたいというのが研究者の本能的な感覚だと思いますので、すごいネタになりそうなものであれば、しばらく公開しないという方針を戦略的に決めるのはある話だと思います。それは特に悪いことだとは思っていないのですが、それによって大発見が遅れてしまうとなるともったいないので、2 年などそれなりのエンバーゴ期間が決めてあるのがいいと個人的には思っています。

第 1 回 SPARC Japan セミナー2018

「データ利活用ポリシーと研究者・ライブラリアンの役割」

地域研究画像デジタルライブラリにおける データベース協働構築の実際

丸川 雄三（国立民族学博物館）
石山　俊（国立民族学博物館）

講演要旨

国立民族学博物館では、世界各地で撮影された写真や動画を集積するデータベース「地域研究画像デジタルライブラリ」の構築を進めている。当事業は科研費による学術研究支援基盤形成事業であり、公募によって選ばれた科研費プロジェクトの代表者とともに、写真のデジタル化および撮影内容のデータ化を実施する。そのためデータベースの構築においては各プロジェクトの研究者と密接に連携する必要がある。実際に支援を担当する発表者の立場から、データベース協働構築における情報システムと支援業務のあり方や課題等について報告する。

丸川 雄三

国立民族学博物館人類基礎理論研究部准教授。2003年、東京工業大学大学院博士後期課程（計算工学専攻）修了。博士（工学）。東京工業大学精密工学研究所助手、国立情報学研究所連想情報学研究開発センター特任准教授、国際日本文化研究センター文化資料研究企画室准教授を経て、2013年10月から現職。専門は連想情報学による文化情報発信手法の研究。これまで手掛けた主なサービスは、『文化遺産オンライン』、『国立美術館遊歩館』、『想 -IMAGINE 早稲田大学演劇博物館』など。

石山 俊

国立民族学博物館地域研究画像デジタルライブラリープロジェクト研究員。専門分野は、アフロ・ユーラシア乾燥地域研究および同地域における文化人類学的研究。主な調査地は、サハラ・オアシスおよびサハラ南縁乾燥地域（サーヘル・スーダン地域）。

●丸川　本発表では、研究者が調査で撮影した写真をライブラリとして活用可能な形にする支援事業「地域研究画像デジタルライブラリ」の実際について紹介します。

国立民族学博物館は 1977 年に開館し、昨年で 40 周年を迎えました。国立民族学博物館は、構想の段階から、研究施設を持つ博物館という形でデザインされています。初代館長の梅棹先生が「博物館ならぬ博情館」と語られているように、もちろん標本資料があってのことですが、情報を重視しているという点が非常に特徴的な博物館です。そのようなコンセプトを受けて、国立民族学博物館では、標本資料のデータベースの他に、研究活動の中で収集した研究資料、あるいは研究者自身が撮りためた調査写真のデジタルアーカイブズを公開しています。こういう活動を長年にわたってこつこつと進めています。

メタデータ付与

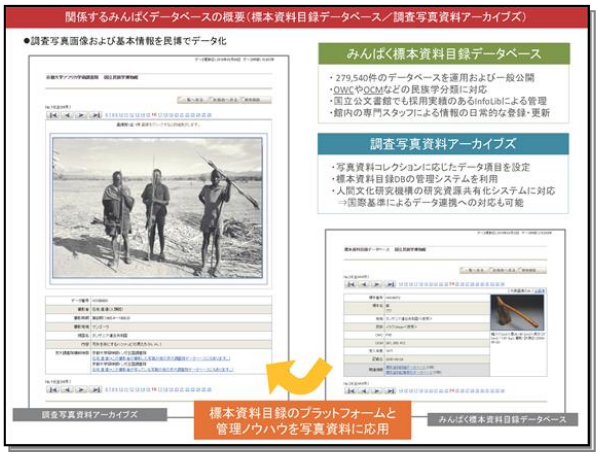
民博が地域研究画像デジタルライブラリの支援事業を実施する背景には、そのような永年の蓄積と実績があるということです。この事業は現在科研に採択されている研究者を対象に、その調査写真をデジタル化し、メタデータを付与し、データベースとしてネットワーク上で共有できるようにする、研究活動の支援をおこなうものです（図 1）。プロジェクトが採択されると、写真を研究者から提供していただきます。あまり整理されていない、ライブラリという形にはなっていない写真のまとまりとしてお預かりすることもあります。代表的なものは、ポジフィルムで撮影された調査写真です。それを整理し、デジタル化し、そこにメタデータを付与していく作業をおこないます。

国立民族学博物館では既に、支援事業という形ではないのですが、研究者の調査写真を図 2 のような形でデジタルライブラリとして公開しております。さらに右下の画面は国立民族学博物館の標本資料データベースです。35 万件ある資料をほぼ全てデータベース化して公開しています。標本資料にメタデータを付ける段階では、OWC や OCM という、イリノイ大学が開発し公開している HRAF（Human Relations Area Files、フラーフ）の分類を付与する作業をおこなっています。

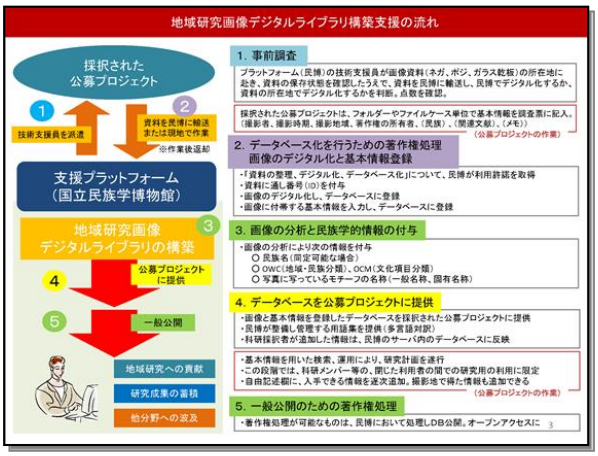
OCM は Outline of Cultural Materials の略で、日本語では「文化分類」と呼んでいますが、例えば農業で使っているものには「241（耕作）」、漁業で使っているものについては「226（漁撈）」、というように、資料の種類に応じて系統的に番号を付ける作業に 30 年来取り組んでいます。今回の地域研究画像デジタルライブラリの整理においても、この OWC や OCM を付与してきたノウハウが生かせるのではないかということです。

また、OCM、OWC という専門的な枠組みではなく、もう少し分かりやすい用語の例もあります。図 3 は展示場で公開している調査写真ですが、右側にあるような民族、カテゴリ、撮影地というような比較的狭い範囲の関連する用語を事前に整理して統一した上で、約 1,000 枚の写真にそれぞれ付与しています。この例ではさらに写真の撮影地を地図上にプロットする作業もおこなっています。

以上の実績も踏まえ、今回の支援事業では、OCM と OWC を写真に付与しています。また、ライブラリとして写真の整理をしやすくするため、コレクション



（図 2）



（図 1）

（図 3）

ごとにそれぞれ独自の用語集を整理して提供しています。当てはめる用語が決まれば、その対訳をつくることで、多言語化が比較的少ない労力で実現できることにもつながります。

フィルムのデジタル化

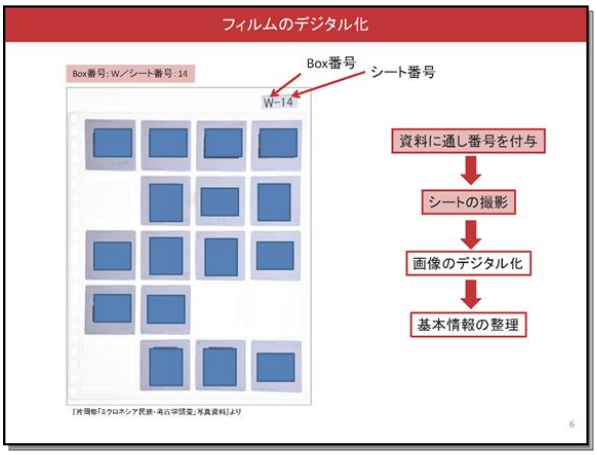
次はフィルムのデジタル化について説明します。ポジフィルムでしたら、まずコマ（スライドマウント）を特定するための番号を入れます（図 4）。写真資料の現物に対する整理をするということです。

それからスライドマウントをシートごと撮影します。さらに写真のデジタルスキャンをおこないます。現在は業務用のフィルムスキャナはほとんど製造中止で手に入れることが難しく、プロジェクトでは SlideSnap Pro を活用しています（図 5）。このスキャナはスライド上映用の装置を改造したもので、デジタルカメラを取り付けて使います。スイッチ一つでスライド送りとシャッターを同期させることができ、慣れれば比較的高速に高精細なスキャンが可能です。

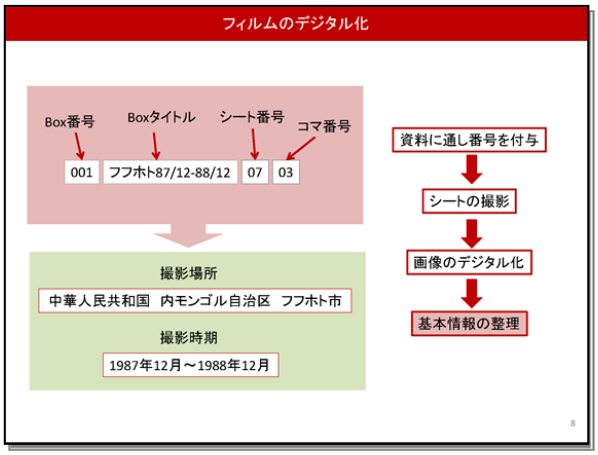
場合によっては研究者が自分のための整理として手書きなどで記入している文字情報もテキストとして取り込みます。略号なども多く読み取りはたいへんな作業なのですが、本人からのヒアリングもおこない、テキストに落とし込んでいくという作業をします（図 6）。

データベースの構築

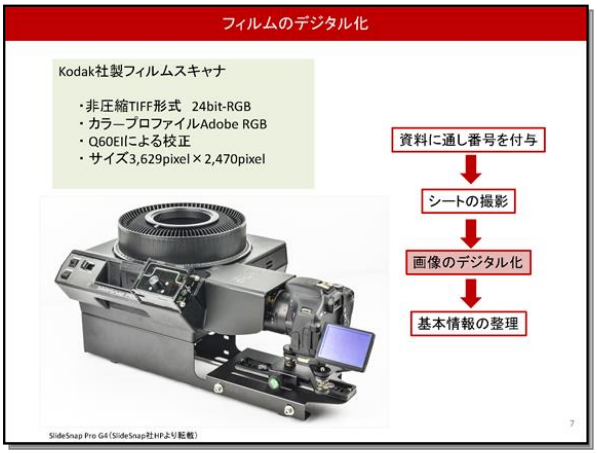
最後にメタデータの付け方です（図 7）。おおまかに五つほどのカテゴリに分けています。最初は写真の説明です。撮影者本人でなければ分からない情報として、まずは写真のキャプションを基本としています。提供時の資料などから分かる情報についてはこちらでなるべく埋めていくのですが、どのような意図で何を



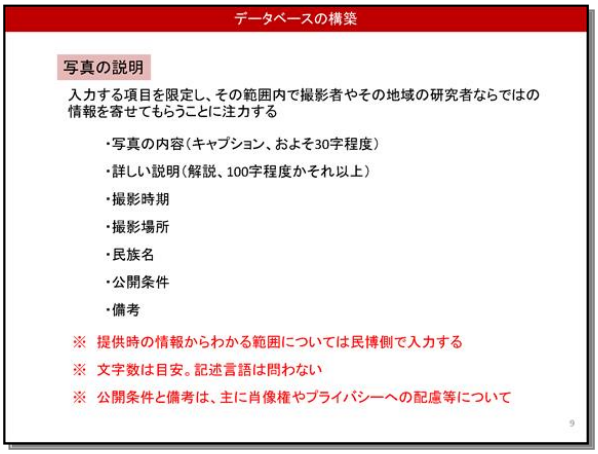
（図 4）



（図 6）



（図 5）



（図 7）

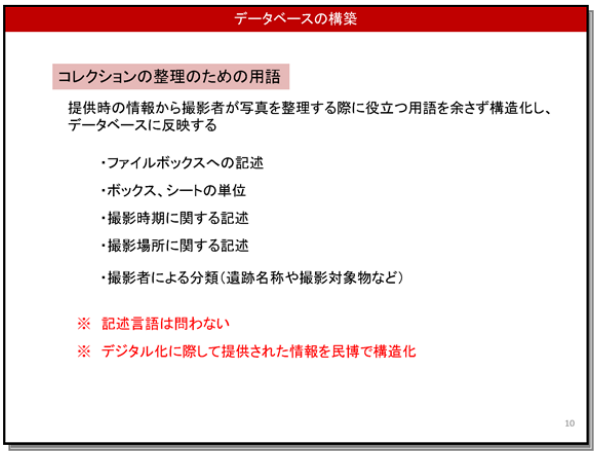
撮影したのかという部分は、やはり研究者自身に入れていただく必要があります。

二つ目は、写真整理のための用語です（図 8）。スライド写真などは、多くの場合 20 コマほどを納めたシートが一つのボックスに 20 枚程度入れられているのですが、そのようなまとまりは閲覧性が比較的良いものです。ぱっと取り出して、開いて、さっと見て、何枚目に何が写っているかということがすぐに分かります。ですがデジタル化をおこない、全体をまとめてフラットにしてしまうと、その使い勝手が失われてしまいます。そこでそのような元々の使い勝手の良さをデジタルにも落とし込んでいくためのデータを付与します。物理的な整理の情報もなるべく落とし込んで、そのまとまりで、例えばきちんとシート単位で取り出すことができるようにするということです。

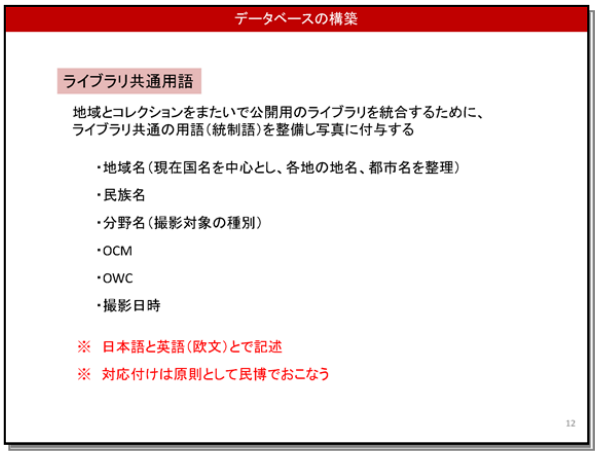
三つ目は、撮影、機材、フィルム、デジタル化に関する情報です（図 9）。撮影機器に関する情報、フィルムの種別やメーカーなどです。デジタルカメラで撮影した写真については Exif というメタデータからいろいろなことが分かります。これらの情報を適切に取り出します。

四つ目は、ライブラリ共通用語です（図 10）。地域名や民族名、あるいは研究分野によっては撮影意図や撮影対象に合わせた用語を付けます。冒頭で説明した OCM や OWC もここに含まれますが、それ以外の用語についても、コレクションをまたいで共通に利用できるよう意識しながら整理します。

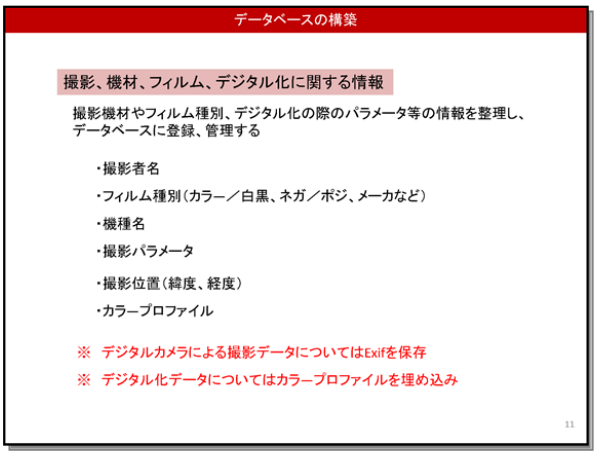
五つ目は、写真の検索時にファセットを自動的につくるためのものです（図 11）。撮影者が入力したキャプションから単語を抽出するなど、自動的なインデックスの更新を検討しています。



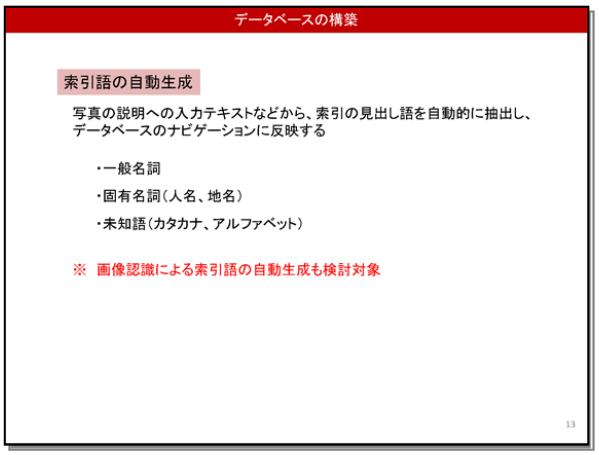
（図 8）



（図 10）



（図 9）



（図 11）

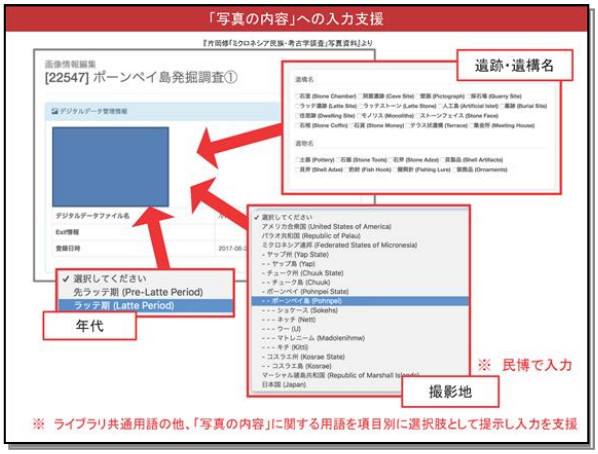
これらのデータを入力し修正するためには、編集用のインタフェースも重要です。図 12 の左が現状の画面ですが、今後はこれを右の構成図のようにブラッシュアップする予定です。先ほど物理的な使い勝手という話をしましたが、スライドを収納したシートの見栄えをある程度踏まえたような見せ方、配置ができるようなものも検討しています（図 13）。入力支援として用語を選択肢として提供することなども可能です（図 14）。

以上の要件や技術要素を組み合わせつつ、写真資料からデジタルアーカイブを構築し、さらにはライブラリとして公開し活用可能な形にしていくという作業に地道に取り組んでおります（図 15）。

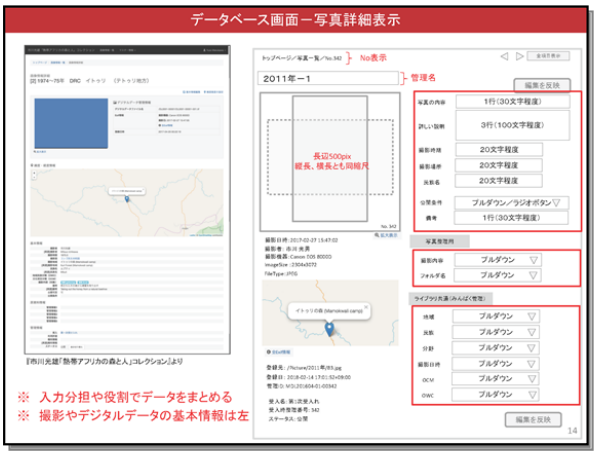
ライブラリへの展開

本事業の目的は研究者に対する支援ですので、関係者を越えた公開は必須ではないのですが、国立民族学博物館としては、差し支えのない写真についてはいずれライブラリとして公開する予定です（図 16）。一般の方にとっても非常に興味深いものになるのではないかと思います。

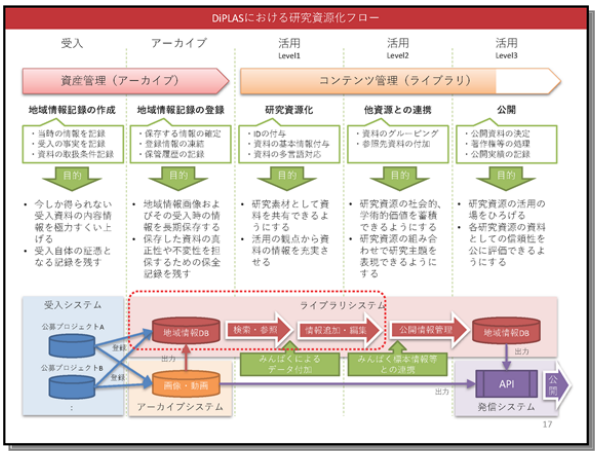
後半は石山に代わります。



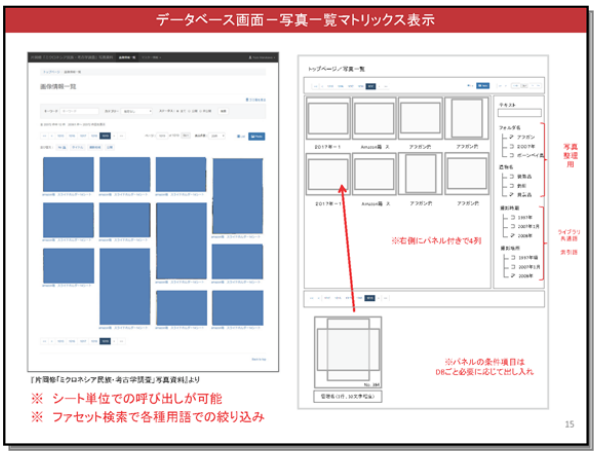
（図 14）



（図 12）



（図 15）



（図 13）

（図 16）

●**石山**　今、丸川からデータベースに関して説明がありました。当初プロジェクトの設計としては、データベース上で情報を入力していただくということをメインに考えていたのですが、実際にやりはじめてみると、Excel にまず入力して、そこからデータベースに流していこうという動きが出ています。

簡単に自己紹介をさせていただくと、本日の参加者の中で私が一番デジタルやオープンサイエンスに弱いのではないかと思っています。私の専門領域は地域研究・文化人類学で、アフリカの乾燥地を中心とした生業文化や生活文化の研究をしていて、ひょんなことから 2017 年 10 月より、この地域研究画像デジタルライブラリプロジェクトに従事しています。地域研究者は、若い方はデジタルに慣れ親しんでいる方が多いのですが、キャリアの長い方になるとアナログの方が結構いるのではないかと思います。私はどちらかというとそのような方の立場に立って支援をして、うまくデータベース化につなげていけたらいいなという気持ちで関わっています。

Excel を用いたメタデータ入力

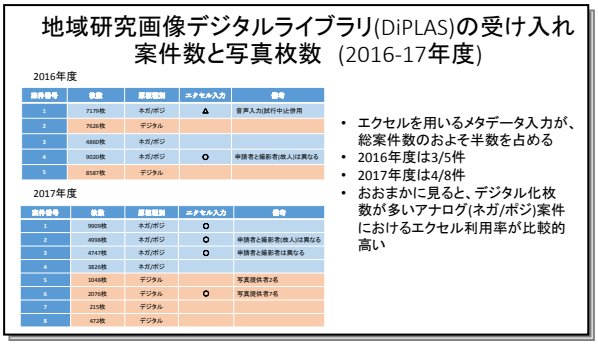
図 1 が昨年度・一昨年度の採択プロジェクトで支援を決定した案件、枚数、原板種別の一覧です。ほとんどの場合は 1 名の方が撮影者ですが、場合によっては科研プロジェクトで数名の方が撮影者になっている場合もあります。ざっと見ると、およそ半分で Excel を用いたメタデータの付与作業が進行しています。デジタル化の枚数が多くて、原板がネガ／ポジの人が、Excel を使っている傾向があるのではないかと思っています。

Excel によるメタデータ付与作業が始まった背景には、まず、文系研究者のデジタルやデータベースに対するハードルがあります。慣れ親しんでいない中で、何千枚分のデータを入れろと聞いた瞬間のプレッシャーはすごいと思います。それをいかに和らげて、みんなにやってもらうようかということがあります。また、デジタル化写真の枚数が多い、撮影者が故人である、撮影者がアナログ世代ということもあります。科研の代表者が同地域の先人の写真のデータベース化を要望するという案件もあり、その場合は、撮った方は既に亡くなって情報を付けられないという状態です。

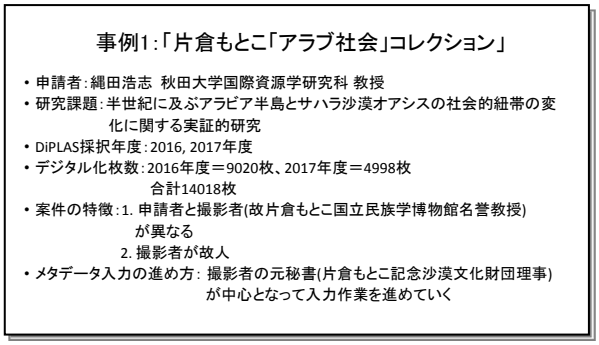
このような課題を抱えながら、Excel によるメタデータ付与作業が始まってきたのです。

事例 1：「片倉もとこ『アラブ社会』コレクション」

事例を二つほど紹介したいと思います。一つ目が「片倉もとこ『アラブ社会』コレクション」というタイトルの案件です（図 2）。申請者は、科研の代表者、秋田大学の縄田先生です。科研の研究課題は「半世紀に及ぶアラビア半島とサハラ沙漠オアシスの社会的紐帯の変化に関する実証的研究」です。この課題は、実は今年度もそうなのですが、2016 年度と 2017 年度で約 14,000 枚のデジタル化が進んでいます。案件の特徴は、申請者と撮影者（故片倉もとこ国立民族学博物館名誉教授）が異なる、撮影者が既にお亡くなりになっているということです。そのような写真に一体どうやってメタデータを付ければいいのかということで、



（図 1）



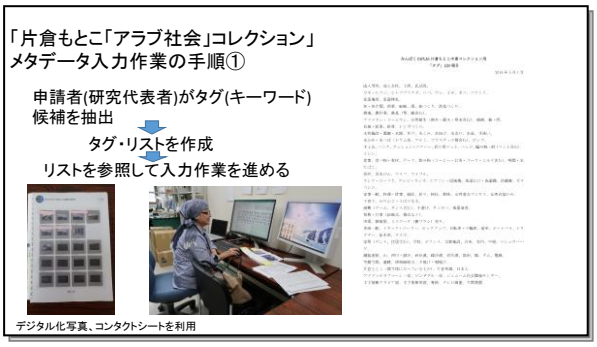
（図 2）

撮影者の元秘書の方にいろいろな情報を参考にしながら作業を進めていただいています。

その写真が図 3 です。このようにして作業を進めています。まず申請者、つまり秋田の縄田先生が写真をざっとくまなく見て、タグ（キーワード）候補を 120 ぐらい抽出します。この中から元秘書だった方がその写真に合うタグ（キーワード）を Excel に入力していきます。

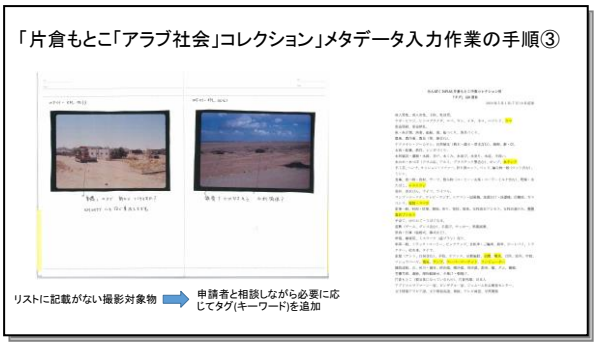
その際、片倉先生が残したフィールドノートやメモも参考にします。それから、入力者の方の記憶なども思い出しながら、どんどん情報を入れていくという作業を地道に続けています（図 4）。

ただし、その途上で最初に想定したタグ（キーワード）に当てはまらない写真が出てくるのです（図 5）。そのときは写真を印刷して、例えば「『重機』のタグを新しくつくりますか？」と、申請者（科研の代表者）とみんなでやりとりしながら、では増やそうかというので黄色マーカーをしたところは増えたり、そのような一発で決まらない往復作業が出てきています。



（図 3）

（図 4）

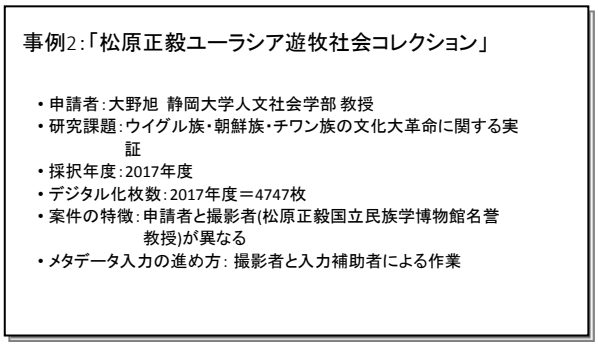


（図 5）

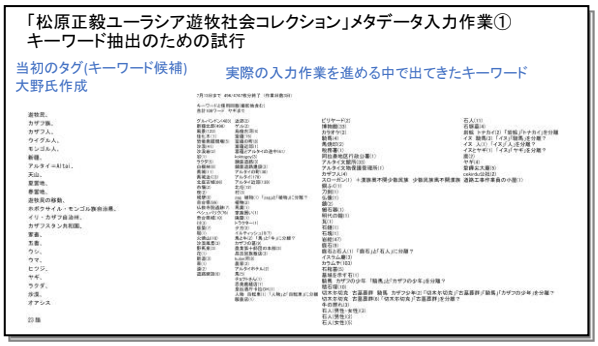
事例 2：「松原正毅ユーラシア遊牧社会コレクション」

二つ目の事例は、「松原正毅ユーラシア遊牧社会コレクション」です（図 6）。この案件も申請者と撮影者が異なる写真です。採択年度は 2017 年度で、5,000 枚弱です。この案件の場合は、撮影者である松原正毅先生と、実際に Excel に文字を打ち込んでいく入力作業者がペアになって作業を進めています。

図 7 の左は、当初のタグ（キーワード）候補です。最初に申請者の大野先生が「こんなものかな」と作成した 23 個のワードです。ただし、実際に松原先生が作業を始めると、どんどんタグ（キーワード）が増えていきました。4,747 枚中 496 枚時点で、100 いっているか、いっていないかぐらいです。まだまだタグ（キ



（図 6）



（図 7）

ーワード）は増えていくと思うのですが、これをデータベース上で検索しやすいようにするために、1 回またこれを整理する作業も必要になってくるのではないかと思われます。

実際の Excel の画面を見ると、地名がいろいろ書いてあります（図 8）。最初、タグ、タグ、タグ、キャプションと入力していくのですが、フィールドノートを見ながら作業しているので、どんどん羅列して、項目が乱れてきてしまって、これも途中で整理しなければいけないという課題も残っています。

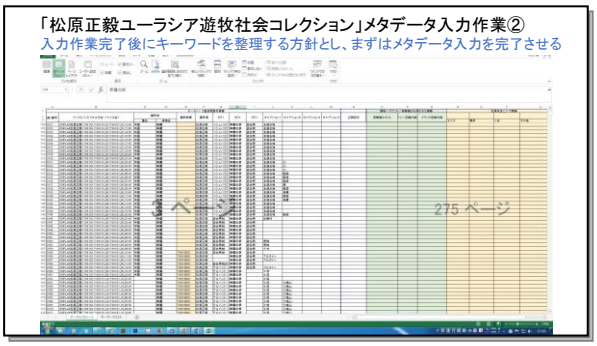
図 9 のような感じで、先生がノートを見ています。作業補助者は内モンゴル出身の総研大の研究生の方で、現地の言葉も分かり、中国語も分かり、松原先生が言った言葉をすぐ Excel に反映させていくことができています。

このプロジェクトの第一義的な目的は、科研研究への支援です。ただ、実際にいろいろな作業をする中で、今までにない新しい面白い副次的効果が見えてきました（図 10）。松原先生は国立民族学博物館の名誉教授でいらっしゃるので、実際に学生、総研大の院生の指導はしないのですが、こういう機会が生まれたことによって、研究コミュニケーションが成り立ったり、松原先生の本が中国語に翻訳されて、この楊海英先生というのが申請者の大野先生なのですが、「中国語の本が余っているから欲しい人がいたらどうぞ配ってください」と言って、留学生に配られたりしています。

事例としては一つで、そんなに大きな広がりにはなっていないのですけれど、私自身、こういうデジタル化の作業もすごく重要だと思いますが、このような副次的な効果がいろいろなところで出てくると思うのです。そのようなものも面白い結果になってくるのではないかと思います。

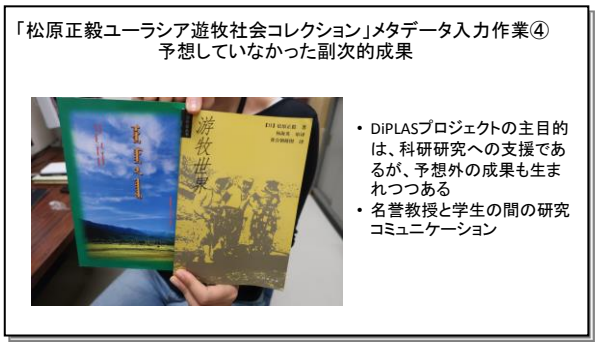
まとめと課題

やはり一番ハードルが高くなっているのはメタデータ付与作業です。何千枚、人によっては 1 万枚オーバー、それをどうやったらプレッシャーを和らげて作業を進めることができるか。ただ「データベース画面があります。やってください」でやってくれる方ももちろんいると思いますが、それではなかなかスタートできない場合は、やはり何らかの支援、その場を設ける、作業補助者に来ていただくといったことも必要だと思います。



（図 8）

（図 9）



（図 10）

（図 11）

最初にタグ（キーワード）候補を設定して、それにのっとってデータを入れていくということを想定しているのですが、実際にやってみると対応できない部分があって、増やさなければいけません。減らすということはまずありません。ですから、どうしても何往復かやりとりしながら進めていく必要があります。私にとって、デジタル化はかなり先端技術のイメージがあったのですが、そのような従来のフェース・ツー・フェースのコミュニケーションが今の段階では重要だと思っています。

このようなことが果たして、地域研究あるいは文化人類学分野の画像・デジタルデータ固有の問題なのか、あるいは他の分野の方々と共有できる何か課題・問題があるのかどうか、これは私の情報ではよく分からないのですが、考えていく必要があるだろうということです。

地域研究者あるいは文化人類学者は、一つの地域・人々・コミュニティと何年も、場合によっては何十年もかけてじっくり付き合っていく方が非常に多いです。ですから、オープンサイエンス、デジタルデータという中での地域研究の位置付け、どうやったらうまく回っていくかということは、実際にこうやっていろいろ作業をしながら考えていければと思っています。

私はこの事業に関わりながら、科研のメンバーで提供支援を受けるという側にも立っています。図 11 の左は、先ほどの片倉もとこ先生の写真です。1970 年のオアシスの灌漑農地です。右は、2018 年 5 月に行ったときの写真です。全く同じ場所ではないのですが、農業が廃れていることによって、社会、いろいろなものが変わってきている、そのような研究の一つのきっかけになる、大変有用な写真だと思います。私の個人的な研究も含めて紹介させていただきました。

●**林賢紀**　ありがとうございます。それでは質疑を受けたいと思います。

●**フロア 1**　国立国会図書館の職員です。お二人のどちらかは分からないのですが、データ公開を前提としていない研究手法ゆえの苦労かなと感じました。研究手法自体を変える方向に、支援する立場としてコミットすることはあり得るのでしょうか。

●**丸川**　写真をデジタル化して、いろいろな形で活用できるようにすることで何が起こるかというところはこれからなのですが、支援事業で、これ自体が研究ではないということで、そこまで積極的に踏み込んで情報学的な研究テーマをそこに当てはめてというアプローチはしていません。むしろライブラリが公開された後で、分野横断的に他の分野、他のサイエンスの方がこの写真のデジタルライブラリをある種のデータとして活用することでイノベーションが起こるということは期待されているのではないかと思っています。

●**フロア 2**　国立極地研究所の図書館員です。ある程度分野の知識を持っている方でなければ入力補助者は務まらないということを前提にされていると思いますが、補助者の作業時間は 1 日どれくらいなのか、あるいは 1 週間単位でこのぐらいの画像をやってというような幅を設けていらっしゃったら教えてください。

●**石山**　基本的にその地域のことを分かっている方に補助者を今のところお願いしていますが、松原先生の事例のように、「これ」と言って、それを日本語や英語で入力するだけなら、必ずしもその地域に精通した方でなくても大丈夫ではないかという気はしています。

作業時間はまちまちなのですが、午前 11 時から昼 1 時間休憩を挟んで 5 時まで、これを週 1 あるいは週 2 のペースでやっている場合が多いです。それは松原先生の案件の場合です。うまく進むときはその日のうちに 200 枚ぐらい、情報が分からないとか、時系列が狂っている場合、50 枚以下という日もあります。ですから、5 時間作業でアベレージ、大ざっぱに言って 100～200 ぐらいと考えていただければいいと思います。

●**フロア 2**　そこは正確さを重視していて、枚数などの制限を設けてやっているわけではないということでよろしいですか。

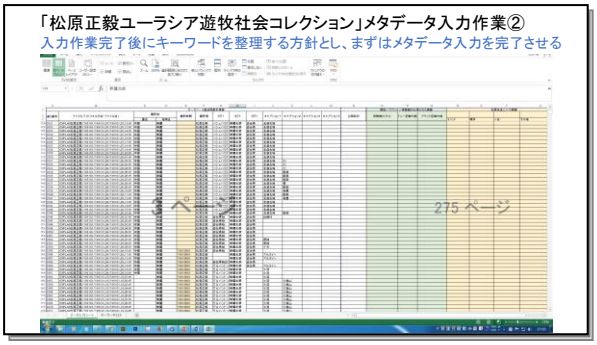
●**石山**　枚数の制限については、結構疲れる作業なので、作業をする方がやめると言うまではやっていただければいいかなと思いますし、片倉先生の案件の場合は、入力者の方が自宅でも少しやったりなど、いろいろなパターンでしています。枚数が多い場合はなかなか終わりが見えてこないということも今の課題です。

●**林賢紀**　はい、ありがとうございます。
　もう一つ質疑の時間がございますが、他に何か聞いてみたい、あるいは質問・コメント等いかがでしょうか。はい、どうぞ。

●**フロア 3**　関東学院大学の職員です。松原先生の「ユーラシア遊牧社会コレクション」メタデータ入力作業②（図 8）を見ると、松原先生のコメントを入力する欄が Excel 上で見えるのですが、独自に追加するテーブルというか、ローカルの入力はどれぐらい許容されているのでしょうか。

●**石山**　文字数などですか。

●**フロア 3**　際限がなくなってしまうので、ローカルで、例えば松原先生のコメントを取るところは何個までというような制限を設けていらっしゃるのか。



（図 8）

●**石山**　丸川さんからあった話では 30 とか 100 とか。文字数のマックスはありましたか。

●**丸川**　目安です。

●**石山**　目安ですね。今のところ松原先生の方にはそのような制約は課していません。またいろいろデータをいじっていく段階になって、長過ぎるというものが幾つか出てきたらお願いする場合はありますが、今はほとんど自由記述に近い感じでお願いしています。

●**フロア 3**　その項目は、検索のターゲットにはされているのですか。

●**石山**　それもそこから検討したいと思うところなのですが、例えば 100 字の記述の中で重要な単語があると判断した場合は、それをまたキーワードとして別に抜き出すという可能性はもちろんあります。

第1回 SPARC Japan セミナー2018
「データ利活用ポリシーと研究者・ライブラリアンの役割」

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）
赤池　伸一　（科学技術・学術政策研究所 / 内閣府）
林　和弘　（科学技術・学術政策研究所）
白井　知子　（国立環境研究所 地球環境研究センター）
丸川　雄三　（国立民族学博物館）
石山　俊　（国立民族学博物館）

●**林賢紀**　パネルディスカッションでは、モデレーターは引き続き私、国際農林水産業研究センターの林が務めます。パネリストは登壇された皆さまです。

始めに、ご発表された皆さまから先ほどの発表について補足があれば、順にお願いできますでしょうか。

●**赤池**　今日はテーマが「研究者・ライブラリアンの役割」ということなので、ライブラリアンの方、研究者の方、ぜひご意見をお聞かせいただければと思います。そこを踏まえて、また新たな施策に反映していきたいと思っていますので、よろしくお願いいたします。

●**林和弘**　1 点補足させていただくと、研究データポリシーを使って何かをしなさいという話ではないというお話をしたのですが、これまで研究データ利活用に関して、思いはあるけれども、活動がしづらかったライブラリアンの方など、関係者の方がいらっしゃったら、むしろこのデータポリシーを使って何かできることを考えていただければというのが本意でした。

●**白井**　特に補足というのはないのですが、研究の現場にいると視野が限られてきてしまうので、このような場で皆さまと一緒に情報交換できることを楽しみにしています。

●**丸川**　私も特に補足はないのですが、国立民族学博物館のデータベースは少し奥に入って分かりにくいところにあるので、それをどう活用して皆さんに見ていただけるようにするかというのが課題です。今日お越しの皆さんには、ぜひご興味がありましたら、一度眺めてみていただければと思います。国立民族学博物館はデータベースを始めたのが早いだけに、データのポリシーといったところが後手に回っているところがあるので、今日この場で勉強させていただきたいと思っています。よろしくお願いします。

●**石山**　私からも補足というわけではないのですが、国立民族学博物館は関西にはコアなファンの方がいらっしゃって、チェックしている方が多いのですが、関西を離れると、よほど興味がある方以外は「何、それ？」という反応もありますので、ぜひこれを機会にサイトを見ていただければと思います。今日は「データ利活用ポリシー」というのがテーマの頭に付いているのですが、私ができたのは現場からの報告ということで、これがどう結び付いていくかという話も考えられたらと思います。

●**林賢紀**　ありがとうございました。では、オンラインならびにフロアからご質問を頂いているので、質問から先にさせていただきたく思います。

まず、赤池さん、林さんへの質問かと思いますが、「もしポリシーをつくれなかったらどうなるのだろう」という質問が来ています。ポリシーをつくれなかった機関、あるいはつくったけれどもデータの公開が間に合っていない機関は競争的資金への応募ができなくなるのですかという質問です。

●**赤池**　法的な拘束力という意味ではないです。ただ、例えば共同研究をしたり、ファンディングエージェンシーから外部資金を獲得したりするときに必要になったり、学会から求められたり、だんだん適用せざるを得ない環境になってくるということで、事実上の、デファクトといいますか、そのようなものに徐々になっていくのではないかと思います。実態面は林さんの方がよくご存じだと思うので、よろしくお願いします。

●**林和弘**　ご説明がかぶりますが、法的拘束力がないので、できなかったからといって罰則が、例えば国から何かあるということは、構造的にはあり得ません。念のため、統合イノベーション戦略では、全研究開発法人が 2020 年までにデータポリシーを策定するということにはなっているのですが、研究開発法人傘下の全ての研究機関はすべからくデータポリシーを策定せよとは書いていないのです。これは読み方の解釈の仕方もあるのですが。ですから、できなかったからといって急に何か不利になるということもありません。

また、プレゼンのときに申し上げたとおり、余計な負担をかけるつもりは本来ない、つくった方が便利になる、もしくは不利にならないという議論の上でこれはできています。今、赤池参事官からもありましたように、世の中の流れとしては、データポリシーがあることが前提で研究費がもらえるような方向性になっているので、それに向けてご準備ください という話になっています。

でも、それで放っておかれてしまうのかということに対しては、今想定されている議論の個人的見解をお話しすると、研究開発法人の主立ったところがデータポリシーを作成すると、研究開発法人のデータポリシーの中の共通項目のようなものが出来上がるだろうということを想定しています。そうすると、それをデファクト的な標準のセットにして、テンプレート化することによって、職員数が必ずしも多くない研究開発法人向け、あるいはその下の研究機関向けにデータポリシーが策定できるようになるのではないかという想定も議論としてはあります。必ずこうするというつもりではなくて、そういう議論もした上で、こういうポリシーおよびガイドラインがつくられてきたということを補足説明させていただきます。

●**林賢紀**　これに関連して、「科研費もデータ・マネジメント・プランが義務化されるということになると、統合イノベーション戦略の期限よりももっと早い、例えば 2020 年度初めにはポリシーができていて、それが公開されているという状況の方がいいか」という質問があります。これは今の話で言うと、やはりその方がいいということになりますでしょうか。

●**赤池・林和弘**　はい。

●**林賢紀**　はい、ありがとうございます。

また、赤池さんに、「文科省と内閣府に分かれてオープンサイエンスの関連施策に取り組まれているように見えます。それぞれの所管、連携はどのようなものでしょうか」という質問が来ています。

●**赤池**　その点についてはよく批判されているところですが、このオープンサイエンスということに関しては、事実上一体的に議論をしていると言っていいと思います。もっと率直な話をすると、私自身が文科省の下部機関の科学技術・学術政策研究所（NISTEP）の職員で、本省も併任していて、縦に全部見ています。

今回、特に内閣府の検討に当たっては、相当綿密に本省、具体的には研究振興局の情報参事官付とも連携しています。もちろんそこも国立情報学研究所（NII）、科学技術振興機構（JST）などと本当に細部に至る連携をしているので、そういう面での齟齬はないと政府部内では考えています。

ただ、先ほどからご指摘があるとおり、これを実現させていく上で、例えば研究所や研究者の方々の研究活動のさまざまな場面にまで、いい仕組みになっていくかというところは、これからの課題であると考えています。そういう意味では、政府部内では齟齬はないと考えていますが、むしろ実態面をどう図っていくかというのは、私どもも皆さんの知恵を借りながらやっていきたいと考えています。

●**林賢紀**　次は林さんへの質問です。「非公開とする研究データはリポジトリには入れないという認識でいるのですが、リポジトリのダークアーカイブ的な利用は想定されているのでしょうか」。

●**林和弘**　研究機関の研究データ管理という意味においては、おっしゃるとおり、機関リポジトリとした場合は、どちらかというと公開を主眼とすると思うのですが、研究管理データ基盤という考え方の方が、データポリシーに関する管理という意味では合っているのではないかと思います。

公開というのはラストワンマイルのようなものです。まず研究機関でどういう研究が生まれて、その研究データの価値はどのようで、その価値に応じてオープン・クローズ戦略を、もしくはエンバーゴ規定を設定した上で保存すべきものは保存し、公開すべきものは公開するということになるので、リポジトリはどちらかというと「ハレ」の日の着物のようなもので、相対的にはオープンにしてどんどん使ってくださいというものを公開するイメージでいます。

それよりはもっと、「ケ」という表現は適切ではないと思いますが、普段使いのものから、時には、いい意味で泥臭い作業、私もデータ整理をよくしていたのでこれは決して悪い意味ではないのですが、そういう作業の元の生データを管理することによって、誰がいつ、どういう作業をし、研究活動に貢献したかというログが取れるようにしておくデータ管理基盤をつくっていくということになるのだと理解しています。

●**林賢紀**　オンラインで聞くと、たくさん質問が来るなと思っております。

これも林さんへの質問なのですが、「科研費を得て行った研究活動で取得・作成した研究データの公開の義務化への流れについての見解」ということで、多分これまでの流れだと、それは出す、公開することになると、今の回答とほぼ同じになるのではないかと思うのですが、こちらはいかがでしょうか。

●**林和弘**　長期的に見れば時間の問題であるのは間違いないとは思っているのですが、科研費が研究データの公開まで義務化するというのは、私個人としては相当時間が、10～20 年ぐらいかかってもおかしくないくらいだと今は思っています。

それを占う二つのシグナルがあります。まず科研費自体が論文のオープンアクセスの義務化をまだ制定していないので、その義務化が始まれば、次は研究データの公開が義務化になるという一つのシグナルになるかとは思います。もう一つは、JST のファンディング事業で、既にオープンサイエンスに関する幾つかの推奨がされていて、その中で論文のオープンアクセスはかなり強い推奨で、研究データに関してもベストエフォートで公開しましょうと書いてあります。これが義務化になったとき、また一つのシグナルになると思いますが、それなりに時間がかかるものと思っています。

●**林賢紀**　科研費を含めた義務化というのは時間がかかるというお話でした。

次は白井さんです。今日のご講演の中で、データの取得と作成が他の研究機関と共同で行われた場合の帰

属については、それらの研究機関との取り決めによって定めるというお話がありました。差し支えなければ、具体的な例を何かご紹介いただけないでしょうか。

●**白井**　具体的な例は思いつかないのですが、基本的には、私が今兼務している企画部で共同研究一切を取り扱っているので、そこで個別案件として対応しています。研究機関同士だと、財源など似たようなケースが多く、省庁が違っても同じ言語でしゃべれることが比較的あるのですが、民間企業との共同研究だと、視点やタイムスケールが違うので、こだわるポイントが伝わりにくいというところが結構センシティブだと思っています。個別案件は細かくてご紹介できないのですが、今度こういう機会があったときのために個別案件で出せるものを整理しておこうと思います。

●**林賢紀**　参考までに、私のいる研究所でも、海外との共同研究契約で知的財産権の帰属を必ず入れるようになっているのですが、例えば特許や品種などがお互いのものになるのか、現地のものになるのかというような例がよくあります。個別で先方と協議して、結果として研究契約として成立するというイメージかと思っています。

　これも白井さんです。個別の案件になるかもしれないのですが、「現場でそれぞれつくられている詳細なポリシーの作成の経緯や概要で、分野ごとの違いはあるのでしょうか」というご質問です。

●**白井**　はい、「ありまくる」というのが答えになります。本当に分野によってこだわるところが全然違います。発表の中でもご紹介したのですが、研究の中でデータを使い合うような分野、国際的な分野や自然科学系は、なるべく使ってもらえるようなポリシーにする傾向があります。一方で、限定された地域や個人の情報が入ってくるような分野では、あまりオープンではないポリシーに落ち着くという傾向があります。それぞれの分野によってデータポリシーのこだわりどころがかなり違うというのが、研究機関としての一括したデータポリシーを細かく定めるのが難しかった一番の原因かと思っています。

●**林賢紀**　また、例えば使ったらお知らせくださいというフィードバックについて、規約に明記するよりも、その手間を減らして利用者を増やした方がいいのではという視点から、「フィードバックの中身や、フィードバックがどのくらい来たかということを定量的に評価している事例はあるのでしょうか」というご質問です。

●**白井**　定量的な評価はしていませんが、DOI を付けて公開した理由の一つは、トラックバックしやすくすることです。どれだけアクセスされたかを集計しやすくするという目的を込めて、なるべく DOI を付けた公開を増やしています。

●**林賢紀**　確かにリンクや DOI があると、それだけで引用のカウントがしやすくなると個人的にも思います。

　丸川さん、石山さんへ質問というか、ご要望が来ています。図書館にとっても、メタデータの付け方など、写真の整理は悩ましい問題です。ご発表の中でも、かなり細かい、例えば保存容器や状態といったものにまできちんとメタデータが付いているということもありました。今日も一部見せていただきましたが、整理のノウハウ、メタデータの付与ルール、扱われている語彙の共有や公開のご予定はあるのでしょうか。

●**丸川**　国立民族学博物館の事業としても写真の整理は長年やっているので、そちらの方では一定のルールがあります。今回ご紹介したプロジェクトの難しいところは、資料を国立民族学博物館の所蔵にして整理するのであれば、国立民族学博物館のルールにのっとってということでいいのですが、この事業ではそれをしないで、資料はデジタル化した後に返却するという形

にしていることです。ですので、国立民族学博物館での ID を付けるという形での整理とはいろいろと違うところもあって、プロジェクト独自の整理ということでやっています。

回答としては、国立民族学博物館としての整理のルールはここで披露するだけの持ち合わせはないのですが、一定のルールの下で行っています。今回のプロジェクトでの整理法は、それとは前提とやり方が少し異なるということです。

●**林賢紀**　国立民族学博物館のルールや、今回のプロジェクトのルールを、例えば図書館の人が使えるような形で公開する予定はないのでしょうか。

●**丸川**　今回のプロジェクトの方の話ですよね。今回のプロジェクトの写真については、整理が終わったら、利用許諾あるいは著作権を国立民族学博物館側で頂くという形の著作権者との取り交わしを進めています。最終的にそのような形で国立民族学博物館側の判断で公開できるとなったものについては公開していきます。ただ、著作権あるいは利用許諾の問題はそこでクリアされて明確になるのですが、何しろ世界中の各地の人々、まちの様子が映っているので、肖像権の問題もあります。それに加えて、今、国立民族学博物館で議論しているのは、各地の文化に対する配慮、その地でこういう写真は公開するべきではないというような、それぞれの場所や文化での暗黙の了解です。それをきちんと踏まえた上で公開しなければならない、公開するに当たってはそれをきちんと踏まえなければならないということで、そこの線引きをどうするかは、館内でも議論しているところです。

どういうルールがあるかという具体例については、もしよろしければ、石山さんから話をしていただこうと思います。

●**石山**　補足すると、地域の規範・社会・文化によって、公開していい写真、できない写真に分かれてくるのです。例えば先ほど紹介した片倉もとこ先生のサウジアラビアの写真が今一番難しい事例です。

片倉先生は女性ですから、女性が家の奥の方、台所などの領域で、覆いを取って顔を出している写真を撮っているのですが、これの公開は、基本的に現在のサウジアラビアの文化・社会的には NG です。ただ、もしご本人の了解が取れれば公開できるだろうというのが、申請者の、科研プロジェクト側での今のポリシーで、現地調査に行く中で、できる限りそれを本人にいいか悪いか聞いて、もし、いいと言ってくれるものがあったら公開する方向でいきましょうということです。もしご本人がご存命でない場合、周りのご家族に聞くなどです。

女性が写っていなくても、結婚の契約は双方の父親同士でするのですが、これは男だからいいだろうと僕たちは考えたのですけれど、実際にサウジアラビアの方に聞くと、「いや、男性しか写っていない写真でも、これは公開しない方がいい」という個別の判断が求められる地域もあります。

撮影者によると、そこは割と鷹揚な方もいるそうです。例えばアフリカでは、女性が上半身裸でなければ、あとはいいのではないか、というスタンスの方もいますし、その辺は案件ごとの個別対応で、なるべく公開したいという方向で進めています。

●**丸川**　そのような状況なので、一般公開、誰でも見られるようにできるものというのはやはり限られています。ただ、それは公開しないという意味ではなくて、国立民族学博物館に来ていただいて、一定の条件を満たした上で閲覧をする、あるいは研究等に利用するというように、リポジトリという形ではなく、研究基盤としては何らか利用可能な形にしていくということになっていくと思います。

●**林賢紀**　今日の文脈でいくと、研究データのポリシーをつくって公開しないといけないのかということに対する一つの回答が、将来的には公開するかもしれな

いけれども、社会の規範、地域規範、肖像権など守らなければいけないものが当然出てくるということですね。

●**林和弘**　先ほどのダークアーカイブと今の話をつなげて、一つ言わなければいけないことを言っていなかったので申し上げます。データポリシーは外向けに公開することを推奨するような意味合いが相対的に強いと申し上げました。実は、このデータポリシーをつくろうと思うと、内規をたくさんつくらなければいけないということを十分認識しています。既に先行している事例で内規が 6〜7 個必要になっています。まず雇用契約があります。その中で、知的財産の処理に関して書いてあるものもあれば、書いていないものもある、書いていなければ決めましょうという話があります。また、そもそも内部でダークアーカイブを含めて、管理すべき研究データをどう扱い、運用するかという内部基準が必要になってきます。そういう話で、データポリシーを策定するというのは、その下に数々の内部の運用・統制のためのルールを決めることでもあるということになります。今、現場の具体的な事例などをご紹介いただいたということです。

●**林賢紀**　今の関連で、赤池さん向けの質問が来ています。「データ基盤をつくることも重要ですが、データを入れることそのものが重要です。これはデータポリシーを定めるだけでは十分ではないのではないでしょうか」という質問です。これについてコメントがあればお願いします。

●**赤池**　データの入れ場所も整理しなければいけないですし、一番大事なのは、入れていただくインセンティブです。インセンティブについては、規制と言うと強いですが、オープンとクローズドの基準のようなものと相互にリンクするところがあります。入れておくべき場所というのは結構深刻な問題と考えていて、文科省や関係当局と議論を始めています。

ただ、限られた予算、財政の中でどういう形で効率的にやっていけるか。また、大事なこととして、各機関もそうですが、まず経営方針があって、その中でどういう管理をしていくかというのが先にありきで、何でもかんでもということにはならないのではないかと思います。

ここで言っているガイドラインは、むしろ英米法か大陸法かというような議論ですが、実際に事細かに決めるということも一案として考えたのですが、それは学術の振興という側面から見てあまり望ましくないのではないかと。例えば個人情報一つとっても、石山先生からご説明があったとおり、中身によって、分野によって、個人に関する情報の意味合いが変わってきます。今のところ言えるのは、こういうことについて考えていただく必要がありますよという意味でのガイドラインであるということで、それで実際の事例を積み上げながら、そこに一般準則のようなものができてくるのではないか、そういう意味で規定そのものも進化のプロセスにあるのではないかと思います。

最初の話に戻ると、どのようにためておくか、ためていくためのルールをどう考えるかというのは関係当局と議論しているところです。

●**林賢紀**　先ほどの丸川さん、石山さんの質問に戻ります。写真そのものの公開は今の議論のとおりで、いろいろな事情で考えなければいけない、大きな言葉で言えばオープン・クローズ戦略になるのかもしれないのですが、図書館員の目で見たときに、具体的な写真の整理方法のノウハウの公開があると、現場で何か担当されている方の役に立つのではないかと思います。先ほど見せていただいたタグは、かなり民族学ベースのものになっている印象がありますが、例えばそれが海外で撮った農業研究の写真などに応用できる可能性があるとすれば、何かそういうノウハウを共有していただけると、整理をする図書館員は大変ありがたいのですが、そのあたりはいかがでしょうか。

●**石山**　これは国立民族学博物館のこのプロジェクトの議論というより、私と同僚の研究員とでよく話していることですが、先ほど丸川さんから昔のスライド装置にデジカメをくっつけて撮影する話などがありましたけれど、その辺のノウハウ、メタデータの付与のやり方など、いずれできたらいいなということは話しています。今は日々違った事例の積み重ねで、まだまだまとめられる状況ではないのですが、恐らく数年続いていけば、作業のパターン化、こういうケースはこうというのができてくるのではないかと予測しているので、その時点でそのようにできたらいいと思いますし、そういう要望を頂けるということは非常にありがたいことだと考えています。

●**林和弘**　関連した質問です。プロジェクトごとにメタデータを付けるとか、秘書さんがやるとか、個別に誰かが付いているような事例のご紹介があったのですが、いい意味でそういう人材が使い回されるようなケースがあるのでしょうか。それが一つのデータライブラリアンにつながるきっかけになるのではないかという感じで、誰か最初はバイネームで、得意な人がどこへ行っても国立民族学博物館のメタデータを付けるというようなシナリオになるような兆しはありますか。

●**石山**　今のところ、今日紹介した事例以外にはまだあのような事例はなくて、ご本人に付けていただくということが多いのですが、地域研究で問題になってくるのは、地域に対してある程度詳しい人の方が作業者として親和性があるのではないかということです。途中でパッと言われても、「あそこですね」と、非常に作業が進みやすくなるのではないかとは思っています。そういうところでも、地域というものに関心がおありの方だったら、ライブラリアンとして活躍していただくことはできると思いますし、その方もそうやって地域に関する知識をどんどん増やしていくことが喜びになれば、それはすごく面白いことだと思います。

●**林和弘**　それは少しシチズンサイエンスの様相も入ってくるような感じがしますね。

●**石山**　そうですね。来館者で、国立民族学博物館のファンの方でそういうことにお詳しい方がいると思うのです。ファンの方に限らずとも、そういう方に参加していただくということもあり得るかなと思いますし、そうなったら面白いです。

●**林賢紀**　ありがとうございます。それでは、ここまでの議論も含めて会場内から何かコメント等ございますか。

●**フロア 1**　国立極地研究所の図書館員です。今の議論を大変面白くお聞きしていました。ライブラリアンをどう広げていくかというところに特に関心があったのですが、それに関連して、前提の話をお聞きしたいのです。元の話に戻ると、好意的に、前向きに解釈すると、ガイドラインが策定されたときに、ポリシーメイキングや運用にライブラリアンが関わるということを想定されて書いていただいたということになると思うのです。

ただ、よく大学等でいわれる話として、各大学の機関には諸条件を判断する能力がないとか、あるいは、かなりきつい言い方ですが、ポリシー作成に対する発言権がライブラリーには必ずしもあるとは限らないというような話があります。でも、それでは議論が進まないというか、かみ合わないので、そもそもガイドラインが示すライブラリアンとはどういう人かということをお聞きしたいと思って来ました。ついては、国研の中でライブラリアンが機能している例があるのか、あるいはライブラリアンの機能を持つ人がどういう人なのか、お聞きできればと思っています。

今回の話を聞くと、白井さんや石山さんはライブラリアンではないかとも思えるのですが、ご本人のスタンスとしてはいかがか、そういうところもお聞きできればと思っています。

●**林賢紀**　皆さま、いかがでしょうか。

●**白井**　ちょっとびっくりしています（笑）。ライブラリアンの方にそのように言っていただけるのはうれしいというか、仲間に入れてもらったのかなという感じもあるのですけれども、研究をやっている側とライブラリアンと呼ばれるコミュニティの間にまだギャップがありまして、今ライブラリアンの方々はどういう方々で、何を日常やっていらっしゃって、関心事は何でというのが見えないのです。関心事などがマージしてきて、そういうご発言になったのかなと思うので、そこを歩み寄ることがオープンサイエンスやオープンデータを進める大きな一歩ではないかという気がしています。

国立環境研究所でも、数は少ないのですが、歩み寄り型のライブラリアンのような方がいらっしゃって、その存在は研究所全体に役立っています。研究者は、自分の研究分野についてすごく熱心なのですが、全体を俯瞰してどうこうしようという余裕のある人はそんなにいなかったりもします。関心に火が付けば走るのですが、日常生活でコンスタントに何かについて俯瞰的には仕事をしていないので、やはりライブラリアン的な方との対話を通じて、研究所としてのいろいろな知識の底上げや有機的な動きにつながっていくといいなと思っています。

●**石山**　私の方もうれしいびっくりです。基本的に研究者だから研究者という一つのラベルだけでいいはずもないし、いろいろなことができるのは、僕はすごくいいことではないかと思っています。そういう意味でそう言っていただけると、自分を見直したりします。

私がこういうメタデータ作業に関わって面白いと思うのは、人の研究手法を盗めるということです。松原先生の事例だと、先生が学生相手に、「20～30 年前はここはこうやったんや、ここでこういうことが起きた」と言うと、僕はアフリカを中心とした地域研究なのですが、「そうか」と聞き耳を立てて聞いて介入してしまったりします。先ほどのご質問ともつながるところかとも思うのですが、その地道にやっている作業のプロセスを将来的に発信していけるというのは意味があるのだと、今、改めて思ったところです。

●**赤池**　少し政策的な側面から見ると、今、大学や研究開発法人では組織改革が最大の課題になっていて、今までのいわゆる教員の方と事務の方だけではなくて、例えば科学コミュニケーションもそうですし、URA の方もそうですし、今までの職制とは違う新たな職制が求められていて、全体のリソースが増える状態ではないのですが、そのような役割の方が強く求められているのではないかと思います。だから、新しい形のライブラリアンは、新しい時代の中で、経営改革の中でどういうポジションを取っていくのか、ある意味戦略的に動ける良いチャンスでもあるのではないかと思います。

例えば、こういう横の情報連携の場もそうですし、他の職制、URA、お隣の職制との関係も含めて、どういう形で、またキャリアパスも含めて取っていけるのかというのが議論になっています。諸外国、アメリカもそうですし、ヨーロッパなども別の形でエコシステムが回っているような面もあると思うので、そういうところも参考にしつつ、もちろん政策当局としてもしっかり考えていかなければいけないですし、現場のお考えなども聞きながら考えていきたいと思います。

●**林和弘**　既にフロア 1 の方とは議論していることを話すことになるのですが、皆さんに共通する話として、データポリシーをつくる際に私も間接的に関わらせていただいているのですが、個人的な見解として、ライブラリアンとして想定している像は一つ明快なものがあります。

研究を終えて次の研究に移るときに、前の研究の管理をするのが正直、面倒くさいと思っている研究者も多いと、研究者である私自身が認識しています。つまり、研究者の内在的欲求としては新しいことをどんど

んやりたいわけです。そこで、終わった研究と言うと駄目なという意味になりそうですが、そうではなくて、終了して興味を失った、これも正しい表現ではないですね。成熟し終わった研究のデータなどについてはライブラリアンがしっかりと管理をする。この場合は、後世に残すだけでなく、知の体系化に役立てるための、教育的な汎用性のあるメタデータを付けるというような、ライブラリアンの中でもゼネラルライブラリアンのスキルが生きていく話としてあり得るのではないかと思っています。

これとの対極として、研究対象に対するパッションが研究者との対話を生み、歩み寄りのライブラリアンという話もありましたが、いい意味での感情移入をして、何か面白いことをやろう、それを具体的なサービスに実現していこうという中に、もう一つのデータライブラリアンの像があり得ると思います。日本のライブラリアンの構造上、なかなかそういうものが起きにくい状況であることも認識した上で、でも、可能性としてはそういうものもあっていいのではないかと思っています。

●**白井**　今、林さんが言葉に詰まっていらしたデータについてですが、「終わった」というよりも、研究者は自分が分かりたいことが分かって、それで論文を数本出したら、次に行くのですが、実はそのデータは、他の視点から見たり、他の部分を使ったりすると、新しいものを生み出す可能性がかなりあって、そういう意味では半生（はんなま）のデータというか、全然死んでいません。もしかしたら誰か他の人が見たら金の卵かもしれない。これからオープンデータで、今まで他の分野の人が見ることがなかったデータに日が当たって、全然違う成果が生まれることがあると思っています。

ただし、私のイントロでもご説明したのですが、次のことを考えてお金を取らなくてはいけないということがあると、自分が論文を出した後のデータを整理して、メタデータもしっかり付けて最後まで持っていくというところまでは気持ちがいかないし、手も動かないということがあります。

そういう非常に価値のあるところだということをフォローさせていただきたいと思いました。

●**林賢紀**　ちょうど今日のテーマにつながる話題ですが、この話題に関連して何か他にフロアからコメントや意見はございますか。特にライブラリアンの立ち位置というところで何かお気付きのことがあれば、フロアからお願いします。

●**フロア 2**　出版社のシュプリンガー・ネイチャーの社員です。私は図書館の担当で皆さまとお付き合いがあります。皆さまと今パネラーで座っていらっしゃる方々の一つの共通点がキュレーターという立場ではないかと思います。キュレーションしていくという面で、データをいわゆるライブラリアン語で言うと、カタロギングしていくことがかなり容易になっていれば、もっとたくさんのデータが取捨選択できたり、公開するまでのデータの用意が手早くできたりするのではないかと思うのです。

例えば研究室で先生方がデータを取ったときに、iPad などをピピピと押すと、そのカテゴリが入力できるとか、昔のデータを例えばシチズンサイエンティストの方にカタロギングしていただくときも、よく図書館にあるバーコードを読むとさっと入るというような、ツールがあればいいのではないかと思うのです。政府で今回出された「データポリシー策定のためのガイドライン」に沿ったような形の項目がうまく入力できるツールを例えば NII や JST と開発するというような企画はあるのでしょうか。

●**赤池**　大事な視点だと思います。ただ、同じような問題意識を持っている人は、JST でも NISTEP でも NII でもおりまして、それ自体が別々の形であるので、政府としてそういう知見をどのような形である意味標準化して政策に落とし込めるかというのは、私自身も

関心があるところです。国の資産としてそういうものをどう活用していくかということがすごく大事です。例えばシュプリンガー・ネイチャーさんやエルゼビアさんもそうですし、グローバルな文脈の中でデータをどう活用して、私の仕事で言えば国益をどう確保していくかというのが、私の内閣府で課せられた次の課題なのです。

だから、そういうところを実現あらしめるものとして、そのような具体的なツール、あるいはプロトコルに落とし込むのはすごく大事だと思っています。潜在的には認識していたのですが、具体的な施策にはまだ落とし込めていないので、ぜひそこのところはしっかりと考えていきたいと思います。具体的には林さんの方から。

●**林和弘**　オープンサイエンス政策の中でも科学技術政策と産業政策を分けて議論することがあって、今のは産業政策の方につながる話だと私は理解しています。オープンサイエンスの可能性を生かして、ツールというものをつくって、ビジネスチャンスを、もっと広く言えば研究プラットフォーム自体をつくって、研究データが世界で研究者のなくてはならない存在になって、そこでサステイナブルなビジネスを展開しようという大きな流れの中で、メタデータを付けるためのツールづくりは産業政策として非常に大きなテーマであるということは、私が講演するときは必ずお話しさせていただいています。

その一方で、今おっしゃったツールに限定して議論させていただくと、ツールのメタデータのフィロソフィはやはり研究者が考えなければいけません。道具のデザイン、仕様を決めるところはまだ研究開発要素が多分に含まれています。武田先生の前でこういう話をするのはあれなのですが。そういうテーマであるとは思います。今日まさに一つの事例をご紹介いただいたように理解しています。非常に大事なポイントだと思います。言いたかったのは、研究開発のチャンスもビジネスのチャンスも両方あり得るポイントだと思いますということです。国としては、支援するというのは一つの流れとしてあっていいと思います。

●**林賢紀**　確かに、ライブラリアンと研究者の共通の重なるところが、キュレーター的というあたりなのかなという印象を受けました。

この話題で他にコメントはありますか。

●**フロア 3**　広島大学の職員です。先ほどいろいろとファンディングエージェンシーの話などが出ました。大学も経営的な観点を持たなければいけないのですが、科研費など、ファンディングエージェンシーがオープンの義務化に進んでいくと、大学もやらざるを得ないという感じになっていきます。

もう一つ、データの保管場所もいろいろ議論しているとありましたが、データ管理をしていくのは経営だけではなかなか難しい問題があって、大きなデータをどうやって束ねておくかということが非常に大きくて、今、個人の研究者にそれを任せることでいいのかということもありますので、その辺をもう少し積極的な政策として進めていただけるといいなと感じました。

●**赤池**　まさにご指摘いただいたとおりでして、そういう方向と個々の研究者の研究活動を尊重するということも基本で、そこのトレードオフをどうやって取っていくかということがあります。

もう一つ重要なアクターとして、産業をビジネスとして回していくことがとても大事です。私が言うのも何ですが、パブリックセクターの世界の中だけでというのはかなり難しくなってくる時代が来ているのではないかという印象はあります。民間のリソース、お金や、大学が持っているシーズをビジネス化するなど、新しい形で次の世代をつくっていく必要があるのではないかと思っています。

私はついこの間、ドイツのメルク社の 350 周年のシンポジウムに行ったのですが、そのときに、あるアメリカのベンチャーの会社が売り込みに来ました。そこ

は研究ツールを一貫して全部見てくれます。それはどちらかというと、共同研究にするためのフォーマットに最初から研究データを入れるところからやると全部管理できますというビジネスです。世界中でビジネスとしてかなりやっているそうです。今、キュレーター、ライブラリアンとして、そのような新しい日本発のビジネスを新たに展開して、日本発の経済が成り立つようなエコシステムをつくっていけると一番望ましいのではないかと思います。ただ、これはかなり現状から飛躍していますけれども、いろいろな枠組みも広げて今ご指摘の点も含めて考えていきたいと思います。

●**林賢紀**　今日のテーマは「データ利活用ポリシーと研究者・ライブラリアンの役割」ということで、データポリシーを策定するガイドラインから、概要、実際のつくられ方、実際にデータ利活用をするに当たってこのような形でサポートするという事例を、今日、3本の講演の中で検討させていただきました。ディスカッションの中では、研究者・ライブラリアンの役割に関して、同じようなところがあるという点が得られたと考えています。

1 点だけ、オンラインで頂いたコメントで拾い忘れたものがあったので、ご紹介いたします。「機関リポジトリ以外のデータの公開の場にはどのようなものがあるのか」というご質問でした。これは他の方からもコメントを頂戴しておりまして、例えば figshare、Dryad、GitHub 等々のサービスがあるということです。他にもいろいろなサービスがあるかと思いますし、NII でも将来的に構築予定と聞いております。

それでは、パネルディスカッションはここまでにしたいと思います。パネラーの皆さまに拍手をお願いいたします。