

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ます。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。

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

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



(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。

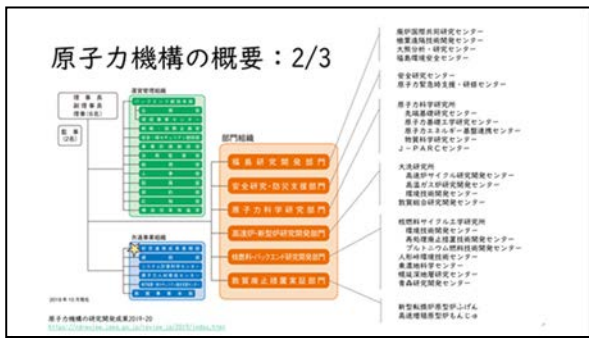
一つの大きなプロジェクトの実施のために研究をしている組織と、それぞれの大きなテーマの中で自由度高く研究開発を行っている組織があり、多様な活動が行われています。それぞれの部門では、研究職の方は研究開発を行い、技術職の方が設備や機器の維持、研究所全体の放射線管理などを行っています。また、各部門には管理系の業務を行う事務職も所属しています。

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。

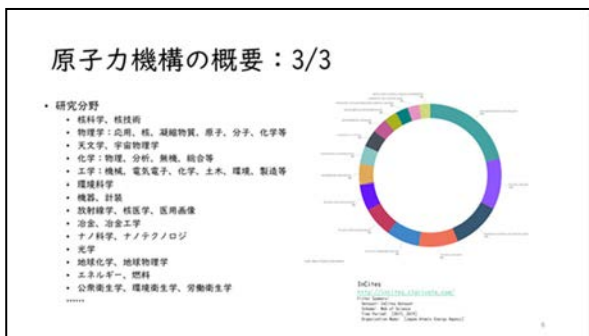
公開の例を挙げます。図 4 は技術レポートですが、DOI、本文の PDF ファイルのリンク、付録へのリンクがあります。利用できる形で公開していますが、ライセンスは明示していません。他にも、福島研究開発部門が放射性物質のモニタリングデータを公開しており、利用者の便利のためにいろいろなフォーマットを掲載しています。このように、統一的な方針の下ではありませんが、利用可能な形で、ある種のデータを公開している例があります。



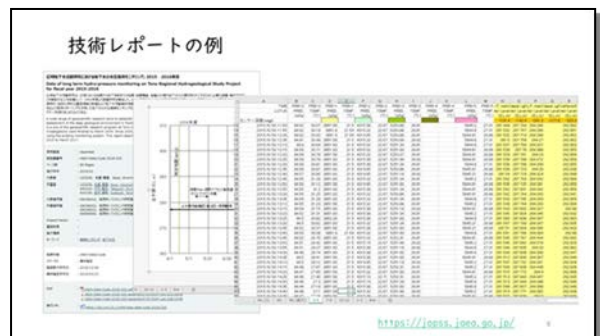
(図 2)

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なく、実践するためにはルール化が必要だと



(図 3)

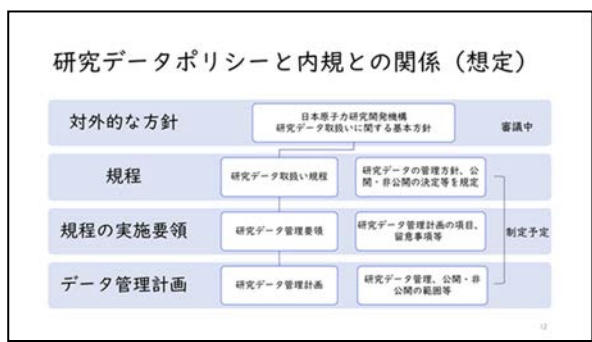


(図 4)

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)

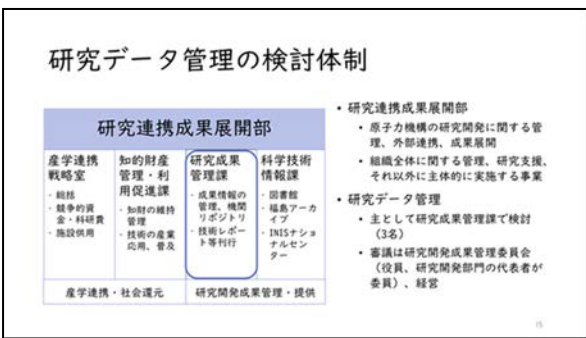
検討のプロセス

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。

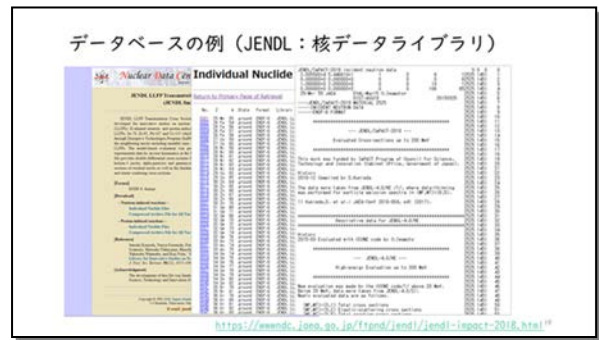
その例がJENDL（核データライブラリ）です（図9）。これはデータベースで、輸出管理の手続きを経て、核データのライブラリとして公開しているものです。

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をお伝えしました。

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



(図 9)



(図 8)



(図 10)

の研究分野で慣習があるということも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、整理して固定化して保存・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。

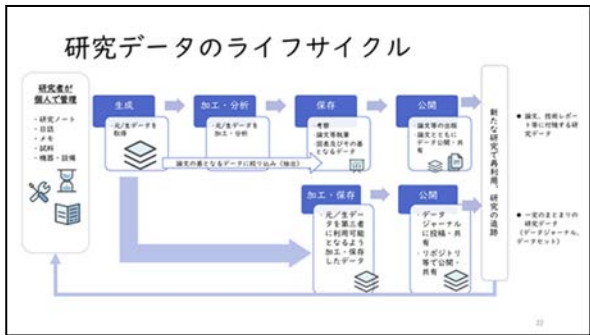
整理をしながら研究データポリシーの本文も検討し、ドラフト版について理事の確認を経て、パブコメを実施しました。パブコメは原子力機構の職員の方なら誰

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 11)

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、

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ならないのか」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ます、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。

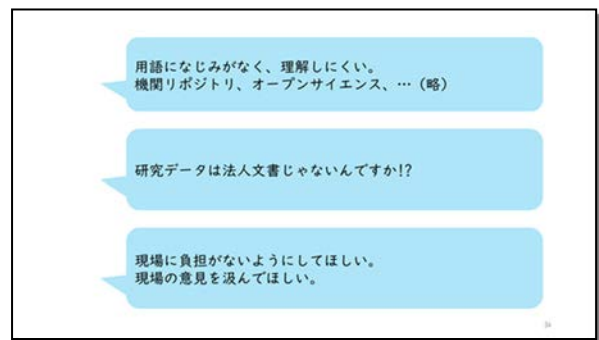
今後の取り組みと課題

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

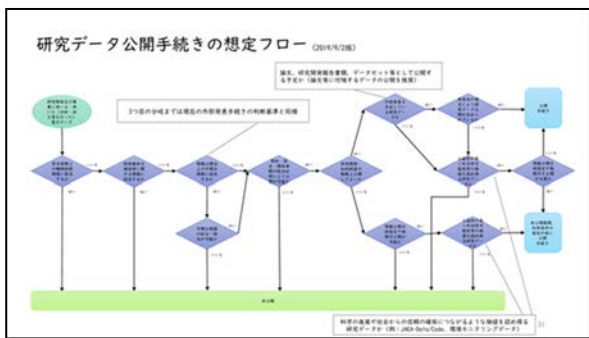
管理を考えるのですが、原子力機構に当てはめることができるものとできないものがあります。また、実際に実践するのは組織の中の個人です。原子力機構や研究分野のことをより学び、研究データ管理を考えなければいけないのだということを感じています。

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。

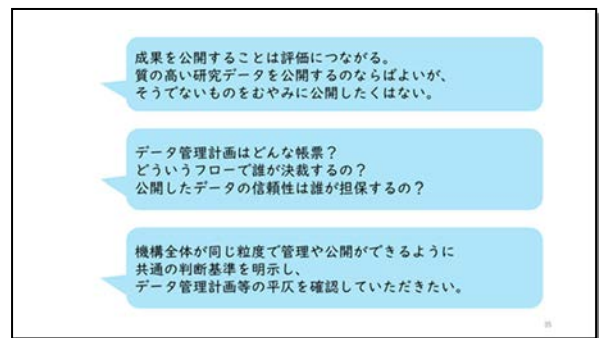
よって、取りまとめる側に要望される事柄としては、ルールや判断基準のガイドラインなどをつくること、情報管理や責任の所在を明確にしていくことがあります。プロジェクト型の研究を行っている部門も、個人の自由度高く研究を行っている部門もありますし、研究分野や手法によっても異なる状況があります。個人的には、そんな画一的なルール化ができるのだろうか、そして、そのような内容で実際にルール化した場合、



(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う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。

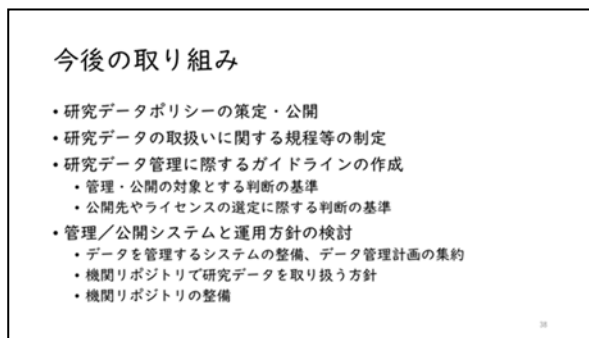
最後に、全体的な課題を申し上げます。図 20 の上の「生データ」と下の「生データ」は、同じものを指しているでしょうか。データの標準化を図る非営利コンソーシアム (casrai) の用語集では、「生データ (raw data)」とは「処理されていないデータ」だと定義されています。「生データ」とは、これらの用例や定義が間違っているわけではなくて、それぞれの人のバックグラウンドや文脈によって、使う言葉の語義が

異なるのだらうと思います。

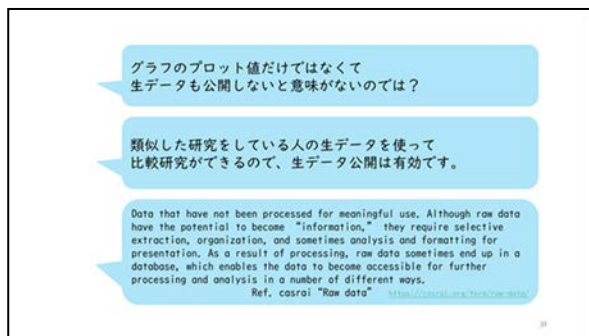
同じことは研究者とのインタビューでも、研究データ管理における組織の中や図書館員同士の議論でも、この SPARC Japan セミナーの場でも起こり得るのではないかと思います。研究データ管理について同じ認識で議論していくことがとても難しく、私たちが使っている言葉の一つずつを、意味や意図を丁寧に確認しながら議論していかなければならないと感じています。

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。



(図 19)



(図 20)

●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ましょ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、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、当面必要ない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、いろいろ手を取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。