

SPARC Japan セミナー2021

「研究データポリシーが目指すものとは」

学術情報インフラが実現する 研究データの管理と循環

大波 純一

(国立情報学研究所 オープンサイエンス基盤研究センター)

講演要旨



研究データポリシー実現のためには、研究データを適切に管理する学術情報インフラの環境が必要である。一方これを実現するため、大学を含む研究機関は新たな仕組みの構築や基準を策定する必要に迫られた。そこで国立情報学研究所では公的な情報基盤を構築し、利用者の研究データ管理の負担軽減となるように対応を開始している。NII Research Data Cloud (NII RDC) と呼ばれるこの基盤は、多様なアプリケーションとしての機能と高い処理能力を有し、国内の機関や研究者の研究データ管理をサポートし、研究活動における知の循環を促進する。このNII RDC を利用した研究データ管理や、研究データの公開・検索を含めたプロセスについて紹介する。

大波 純一



国立情報学研究所オープンサイエンス基盤研究センター特任准教授。博士（理学）。2009年3月、東京工業大学大学院生命理工学研究科博士課程を修了。製薬企業向けのシステムエンジニアとして民間で働いた後、国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンターで研究員として、生命科学系データベースや検索基盤の研究に当たる。2020年4月より現職で学術検索基盤CiNii Researchの研究開発、運営および国際連携に従事。

はじめに

本日は国立情報学研究所（NII）が提供しているインフラと、研究データポリシーや研究データの管理に関する状況についてご説明します。

最初に少し自己紹介をさせていただきます。私は元々、生物学の方で学問を修めており、動物の遺伝子解析のようなことをしていました。卒業後はIT企業で製薬企業向けのシステム開発などを行い、生物系や情報工学の分野ではとても多くのオープンデータを使ってきたことから、オープンサイエンスの重要性を認識しました。その後、公共データベースの現場へ移り、2012年から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で横断検索などのシステムを開発し

ました。2020年に国立情報学研究所のオープンサイエンス研究センターに移り、現在は新しい検索基盤のCiNii Researchを担当しています。本日は、この基盤の説明からさせていただきます。

学術コンテンツ流通の広がり

元々、学術コンテンツといえば論文や図書、学位論文などが主流だったと思いますが、最近はオープンアクセス推進の流れがあります。国際社会のさまざまな要請もあり、実際にそれをどうしていくかという段階になっています。さらに、研究データを研究データ管理（RDM）の流れで出したり、歴史的・文化的資料についてはデジタルアーカイブで公開したりといった、

幅広い提供の仕方をしていくという世界が学術コンテンツにおいて広がってきています。

このような中で国の政策としては、機関リポジトリを有する全ての大学、大学共同利用機関法人、国立研究開発法人において、2025 年までにデータポリシーの策定率が 100%になるようにしてくださいということが提言されました。さらに、公募型の研究資金の新規公募分において、2023 年度までにデータマネジメントプラン（DMP）およびこれと連動したメタデータの付与を行う仕組みの導入率が 100%になるようにしてくださいということもいわれるようになりました。このことは、国の示している第 6 期科学技術・イノベーション基本計画の方針で指示され、それに従って研究・管理をしていくという動きが必要になってきています。

実際にどのような動きをするかということも国で定めています。例えばデータキュレーター、図書館職員、URA、研究の第一線から退いた方など、研究の周りに位置しているさまざまな人たちの協力を頂き、図書館のデジタル転換などの取り組みについて、2022 年度までに方向性を定めていくということです。

そこで、ここからは、「公的資金による研究データの管理・利活用に関する考え方」、研究データ基盤システム NII RDC（NII Research Data Cloud）、およびメタデータ共通項目と呼ばれるデータのテンプレートについてご説明します。

「公的資金による研究データの管理・利活用に関する基本的な考え方」

「公的資金による研究データの管理・利活用に関する基本的な考え方」は、2021 年 4 月に統合イノベーション戦略推進会議にて決定されました。この内容は内閣府の研究デジタルトランスフォーメーションの Web ページで公開されていますが、主に図 1 に示す四つの記述のとおりです。

一つ目は、研究データ基盤システムである NII Research Data Cloud を中核的なプラットフォームと位置

付けて、産学官における幅広い利活用を図るため、メタデータ（データを説明するための情報から構成されるデータ）を検索可能な体制を 2023 年度までに構築することです。

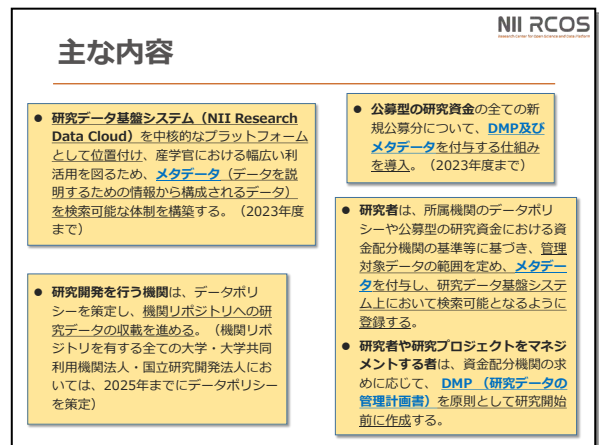
二つ目は、研究開発を行う機関は、データポリシーを策定し、機関リポジトリへの研究データの収載を進めるということです。

三つ目は、公募型の研究資金の全ての新規公募分について、DMP（研究データの管理計画書）やメタデータを付与する仕組みを 2023 年度までに導入するということです。

四つ目は、研究者は、所属機関のデータポリシーや公募型の研究資金における資金配分機関の基準などに基づき、管理対象データの範囲を定めてメタデータを付与し、研究データ基盤システム上において検索可能となるように登録する。そして、研究者や研究プロジェクトをマネジメントする者は、資金配分機関の求めに応じて、DMP を原則として研究開始前に作成し提出するということです。

しかし、これらの項目には、研究データのメタデータや DMP の作成、機関リポジトリへの収載、2023 年度までに行うなど、さまざまな新しい内容が含まれており、これらを完璧に準備できている機関はまだ少ないと思います。

図 2 は「公的資金による研究データ管理・利活用に関する基本的な考え方」についてのスライドです。特に注目していただきたいのは、真ん中に赤で示した研



（図 1）

究データ基盤システム（NII Research Data Cloud）です。ここに、大学・研究機関や分野ごとのデータプラットフォームの中から、メタデータやデータそのものを登録していただき、それを検索・利活用できるようにしてアカデミアや産業界に出していくことを推進しようとしています。この結果、先ほど電気通信大学での Society5.0 につながるさまざまなアウトプットの話があったように、さまざまな方向から社会が良くなることを実現していこうという流れになっています。

このデータの流れを実現するために、資金配分機関や NII ではメタデータを収載する際の基準となる共通項目を制定しています（図 3）。こちら内閣府の Web ページで公開している内容ですが、「ムーンショット型研究開発制度におけるメタデータ説明書」でパイロット的にデータ登録している分野において、メタデータのテンプレートを設定しました。これを基準と

して、どのような項目のメタデータを研究データから登録して流通させるかを考えています。

研究データ基盤システム（NII Research Data Cloud）

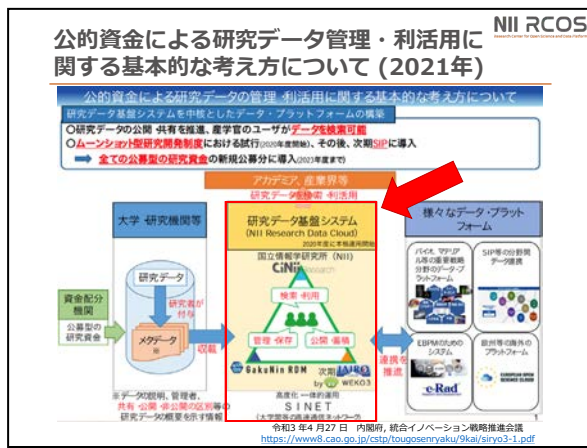
さらに、NII では近年、NII Research Data Cloud（NII RDC）という基盤を提供しています（図 4）。これは大学や研究機関で創出された学術情報を NII のクラウドにまとめていただき、活用・保存していくための場所です。大まかには、データ検索基盤、データ管理基盤、データ公開基盤の三つで成り立っており、それぞれが連動しています。

データ管理基盤では GakuNin RDM と呼ばれるサービスが動いており、さまざまなデータの登録やプロジェクト内のデータ管理ができるようになっています。

データ公開基盤では、JAIR Cloud というクラウド環境、もしくは WEKO3 と呼ばれるサービスにデータを登録して外に公開することができます。

そして、データ検索基盤では CiNii Research というサービスが動いています。流通しているメタデータをここに集約し、利用者は検索して情報を発見することができます。

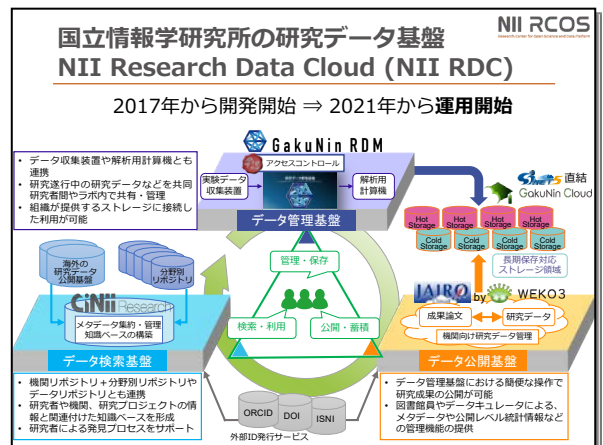
NII RDC を実際に使うときの流れは図 5 のようになっています。基本的には研究者などが GakuNin RDM に日々の研究データを管理・共有するためのスペースを作って登録します。メタデータや研究データ本体、そして DMP の情報も登録できます。さらに公開基盤



(図2)

項目	必須/任意	備考
1 資金配分機関情報	必須	公費型の研究資金を配分した資金配分機関（所有者含む）の英語略称
2 e-Radの課題番号	必須	公費型の研究資金の場合は、e-Radに登録した課題番号
3 プロジェクト名	必須	公費型の研究資金以外の場合は記入不要
4 データNo.	必須	管理対象データを一意に特定するための番号
5 データの名称	必須	公費型の研究資金を配分した資金配分機関が付与
6 掲載日・掲載更新日	必須	予言資料、報告資料、測定結果などの半導体の分らない場合は避ける
7 データの説明	必須	データの詳細な説明、掲載更新日
8 データの分野	必須	e-Radの研究分野（主分野）、e-Radとの連携により、自動入力される予定。
9 データ種別	必須	研究データ登録システムでは、通常は「データセット」を標準とするが、データの特性に応じて「データセット」以外の種別を選択可能
10 掲載データ量	任意	1GB未満、1GB以上10GB未満、10GB以上100GB未満、100GB以上の区分により記載
11 管理対象データの活用・提供方針	必須	無償、有償、ライセンス情報、その他の条件（利用の仕方等）等を記載
12 データの公開予定日	必須	公開/共有/非共有、非公開/公開開始期日から選択
13 データの公開日	必須	公開開始期日を選択した場合は、公開予定日を記載
14 データの公開日	必須	公開開始期日を選択した場合は、公開予定日を記載
15 データの公開日	必須	公開開始期日を選択した場合は、公開予定日を記載

(図3)



(図4)

では、研究データを実際に外に出すための設定をします。大学では図書館員などがその処理に当たるとこともあったと考えています。最後に、そのようにして公開されたデータを検索基盤の CiNii Research で外に出すと、組織管理者や外部研究者が確認したり、一般市民の方が情報を見たりすることができるようになります。そういった使い方を想定しています。

管理基盤 GakuNin RDM

管理基盤 GakuNin RDM は、登録すると図 6 のような画面になり、プロジェクトの中でデータ共有やデータ流通などをすることができます。まだ構築中の部分もありますが、実際に登録するときはフォームに入力していただきます。さまざまな互換機能も内部に備えています。メタデータの共通テンプレートに入力するのは大変だと思った方もいると思いますが、できるだ



(図 5)



(図 6)

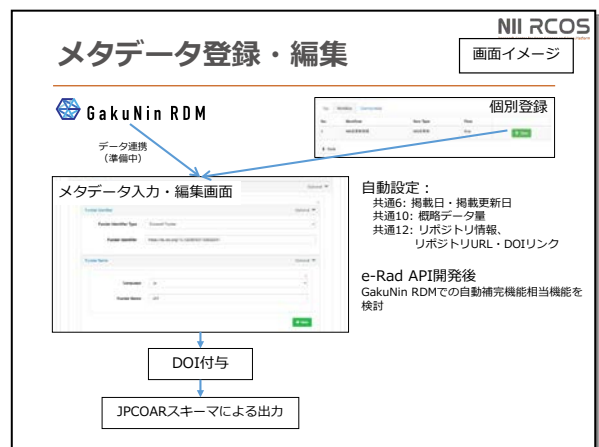
け Web のインターフェイスで入力しやすいように設定しています。日本語、英語の対応や、さまざまなスキーマへのフィッティングなども自動で行うことを検討しているため、お使いいただければと思っています。最終的に登録した機能をレポートニングすることもできる予定です。DMP の登録なども予定しています。

公開基盤 JAIRO Cloud

図 7 は公開基盤 JAIRO Cloud の画面のイメージです。登録されたデータ一つ一つを論文に出すとき、あるいは最終的に世の中に広く見てもらいたいものがあるときはこちらに登録し、DOI を付与して、スキーマなどを入力して見せることができるようになっています。

検索基盤 CiNii Research

検索基盤 CiNii Research は 2021 年 4 月から正式公開された国立情報学研究所の新サービスで、一つの検索画面から多様な学術情報へ一挙アクセスできる基盤です (図 8)。3,000 万件以上のデータが入っており、論文だけではなく、さまざまな研究データ、本、研究のアウトプットを検索することができます。元々 NII が提供していたさまざまなサービスもあるのですが、そのまま登載したわけではありません。例えば、Google で「CiNii Research」と検索すると CiNii Research の現在のトップ画面にアクセスできますが、その中に「研究データ」や「プロジェクト」というタブがあり、必要に応じてフィルターをかけることができるように



(図 7)

なっています。また、検索結果の画面では、データソースとしてどこから資金配分がなされたかでフィルターをかけたり、その研究データがどういったプロジェクトでつくられたかを記述したりと、高度な使い方ができるように工夫しています。

このような研究データおよびメタデータの登録については、ムーンショット型研究開発制度の情報をパイロット情報として、内閣府と登録の方法を構築しつつあります。この中では、特にメタデータの登録の推奨案として、研究プロジェクトの管理者や研究者がまずは管理基盤 GakuNin RDM にデータを登録し、そのデータを公開するときは JAIRO Cloud にそのまま連携させて、そのデータが検索基盤の CiNii Research に移り、自動的に外から検索できるようにして公知にしていこうというサポートの仕方を構築中です。

サービス統合のお知らせ

最後に1点、お知らせです(図9)。元々NIIの基盤で論文を検索するためのインフラとして提供していた CiNii Articles を多くの方にお使いいただいていたと思いますが、こちらを2022年4月1日から新しい論文検索基盤 CiNii Research に統合する予定です。実際は、サービスのドメインなどが主に変わって、元々の CiNii Articles の URL はそのままリダイレクトされて CiNii Research の中のページに移るので、ユーザー側で大きな対応をしなければいけないということは少ないと考えています。ただ、API や高度な利用をしてい

た方については少し変更点があります。大変お手数ですが、CiNii Research もしくは現在の CiNii Articles のページに、各ページの右上にお知らせリンクが出ていますので、そちらの内容を見ていただければ幸いです。

一例として、CiNii Articles と CiNii Research のページを、ある同じ論文について示してみます(図10)。CiNii Articles では左のようなメタデータが見えていますが、新しい CiNii Research では、メタデータだけでなく、参考文献のリンクや関連プロジェクト、ものによっては関係する研究データなどにつながって、そのままシームレスにさまざまなデータを見つけたり使ったりすることができるようになっています。ご利用いただければ幸いです。

まとめ

発表のまとめです(図11)。まず、「公的資金によ

(図9)

(図8)

(図10)

る研究データの管理・利活用に関する基本的な考え方」(2021 年)に沿う研究データの管理が必要な状況になってきました。それに対応して、メタデータ共通項目にマッピングされた研究データを、研究データ基盤システム(NII RDC)の三つの基盤である管理基盤、公開基盤、検索基盤を利用して適切に管理し、新たな知の循環に持っていくことができると考えています。最後に、2022 年 4 月に CiNii Articles は CiNii Research に統合します。より発展的に使うことができるものになっていますので、よろしくお願いいたします。

●矢吹 二つ質問を頂いています。一つ目は、「GakuNin RDM と JAIRO Cloud の連携では、メタデータだけでなく、研究データ本体の連携が可能な見込みでしょうか」という質問です。いかがでしょうか。

●大波 研究データ本体の連携が可能な見込みです。研究データ本体は、管理基盤、公開基盤まで移って、正確にはメタデータも JAIRO Cloud に移るのですけれども、メタデータだけが検索基盤にも移る、連携するというものです。分かりづらくて申し訳ありません。

●矢吹 ありがとうございます。もう一つ質問があります。「国際ジャーナルなどに掲載された論文などの出版物と、それらに関連する研究データはどのよう

にリンクされるのでしょうか。ご存じでしたらご教示いただきたく存じます」。

●大波 現在、国際ジャーナルに載ったデータを流通させる基盤として、海外のデータ基盤である CrossRef や DataCite といった海外データ流通リポジトリと連携を行い、情報を CiNii Research などに取り込もうとしています。その海外由来のデータを使って国際ジャーナルのものや英語のものをつなげていき、そちらで成果物を確認できるようにしています。若干途上の部分もありますが、そういったポリシーで進めています。

●矢吹 ありがとうございます。さらに追加の質問です。「JAIRO Cloud 以外のリポジトリと GakuNin RDM との連携予定について伺えますでしょうか」という質問です。

●大波 JAIRO Cloud 以外のリポジトリに、分野別リポジトリや、さまざまな大学固有のリポジトリがあるパターンもあります。したがって、画面では国立情報学研究所の基盤を使うことをお勧めしましたが、必ずしも使う必要はなく、別途これを迂回するようなルートを経て CiNii Research で検索できるようにする方法でも大丈夫です。都度対応ということにはなりますが、大学や研究所とのご相談の上でこの流れに載せるということは十分可能ですので、ご理解いただければと思います。

●矢吹 ありがとうございます。もう一つ質問があります。「ご紹介されたメタデータのスキーマは、『2023 年度までに、DMP と連動したメタデータの付与の仕組み』で述べられているメタデータの標準スキーマとになっていく予定なのでしょうか」というご質問です。

●大波 メタデータと連動していくかというところは内部でかなり議論があり、「そうだ」と言いづらいところがあります。ただ、そうしていくことが、諸外国

NII RCOS

まとめ

- 『公的資金による研究データの管理・利活用に関する基本的な考え方』(2021年)に沿う、**研究データ管理が必要な状況**
- メタデータ共通項目にマッピングされた研究データは、**研究データ基盤システムNII RDC**(NII Research Data Cloud)
 - GakuNin RDM (管理基盤)
 - JAIRO Cloud (公開基盤)
 - CiNii Research (検索基盤)
 を利用して適切に管理し、**新たな知の循環へ**
- 2022年4月にCiNii ArticlesはCiNii Researchに統合

(図 11)

でも進められている機械可読の DMP (machine-actionable DMP) の利用の仕方ともつながっていき、単純にデータを登録するだけでなく、それがまた社会を良くするアウトプットが出せるものになっていくと考えて取り組んでいるので、大変良いご質問だと思うのですけれども、引き続き検討中という答えにさせていただきます。

●矢吹 ありがとうございます。最後にもう一つ、「GakuNin RDM 上で DOI を取得した場合、リポジトリにデータを連携した場合にその DOI を引き継ぐのでしょうか。あるいは、GakuNin RDM とリポジトリのそれぞれで DOI を取得することになるのでしょうか」という質問です。

●大波 まず、DOI は GakuNin RDM で取得できるのではなく、公開基盤の JAIRO Cloud で登録できるようになっています。一方で、DOI はそもそもジャパンリンクセンター (JaLC) や他の分野別データベース、あるいは海外のリポジトリで並行して取られてしまう場合もあり、複数の DOI を持った場合にどちらを優先させるかということは DOI の中で大きな議論になっていると思います。一つの答えとしては、最終的に CiNii Research で検索するときに名寄せをして複数の DOI を示すようにしています。そのような形で、DOI を複数取ってしまった場合でも、それを織り込む形で基盤の方で対応できるようにしています。