

第 4 回 SPARC Japan セミナー2014

「グリーンコンテンツの拡大のために我々はなにをすべきか？」

機関リポジトリと DOI - JaLC における DOI 付与について -

武田 英明

(国立情報学研究所 情報学プリンシプル研究系)

講演要旨

ジャパン・リンク・センター (JaLC) は 2012 年に発足し、第 1 段階としては主にジャーナル論文に DOI を付与してきた。2014 年 12 月に新システムに移行すると共に、JaLC は国内の様々な DOI 付与ニーズに対応できるように、多様なコンテンツに DOI 付与ができるようにポリシーを整備した。ことに機関リポジトリのコンテンツに積極的に DOI が振れるように拡張されている。また、2014 年 10 月より研究データに対する DOI 付与実験プロジェクトも始めている。本講演では JaLC の DOI 付与のポリシーに関して、機関リポジトリの関連を中心に紹介する。



武田 英明

国立情報学研究所教授。総合研究大学院大学 (総研大) 教授兼任、ジャパンリンクセンター運営委員会委員長、ORCID理事。東京大学大学院工学系研究課博士課程修了。工学博士。研究の関心はセマンティックWeb、知識共有システム、設計学。

「論文」の未来—人間可読から機械可読へ—

今、研究の世界で何が起きているのでしょうか。10~20 年前のデジタル化以前の研究者であれば、データは自分で作ることもありましたが、あくまで紙の論文を読んで、紙の論文を出していました (図 1)。しかし、論文がデジタル化して、研究者はデジタル論文を読んで、デジタル論文を発表するようになりました。一方、対象もデジタルになってきて、自分がその対象から直接データを取るよりも、誰かが作ったデータをまた使うというスタイルの研究が増えてきました (図 2)。もうじき、論文もデータも、どちらも研究成果となるような時代が来るでしょう (図 3)。

研究成果データとの一体化というのはどういうこと

かという、論文もアウトプットであるデジタルデータの一つになるということです。すなわち、データがデジタル論文の一部として付随していたり、データが論文自身 (データ論文) になったりします。もう一つ、違う視点で言うと、論文自体がデータとして扱われます。これはバイオサイエンスの世界では既に構築されつつあります。バイオサイエンスでは 1 日何千件と論文が出てくるので、人間が全部を読んでいられません。そこで、賢い検索をしようと頑張っています。要するに、半分、機械が論文を読んでいて、自分に関係することが必ず引っ掛かるようにしているのです。従って、人間のための論文からコンピュータのための論文になっています。これが既に現在起きている事態です。

さて、視点を変えて、そもそもなぜ研究データをオープンにするのでしょうか。その理由は、基本的には4点に集約されると思っています。1点目は社会での成果の共有、知識の共有のため、2点目は公的資金の公共性のためです。これらは社会的要請に分類されます。特に最近の研究資金助成機関のオープンデータポリシーなどは、2点目のコンテキストから来るものです。要するに、タックスペイヤーのアカウントビリティのために成果を公開するのです。

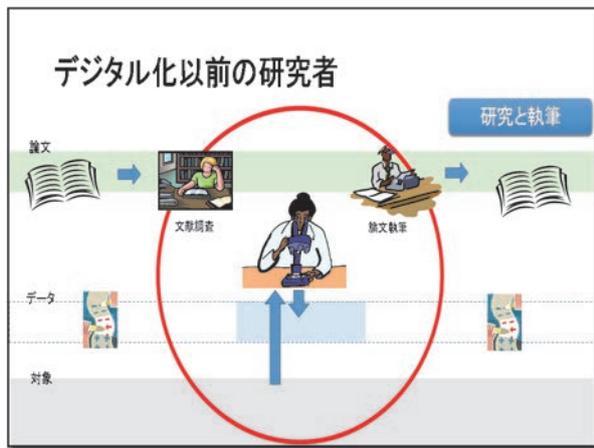
3点目は研究の発展性のためです。これは、巨人の肩に乗るためには、先人の研究やデータが手に入らなければいけない、研究の自律的発展のために必要であるということです。4点目は再現性担保のためです。これは不正や捏造がなく、正しく研究が行われたことの証拠のためということです。これらは研究のオープン性に分類されます。

現在は、これら4点全てが相まって研究のオープン性になっていることが状況をややこしくしています。ある視点から見るとオープンにしたら困る、別の視点から見るとオープンにしなければいけないという状況があります。ただ、これは全てオープンにすべきだという議論ではありません。こういうことになったときに、この四つの視点から見て、どこまでをオープンにするのが正しいかを決定するのが、その研究コミュニティや機関、ファンディングエージェンシーの役割です。

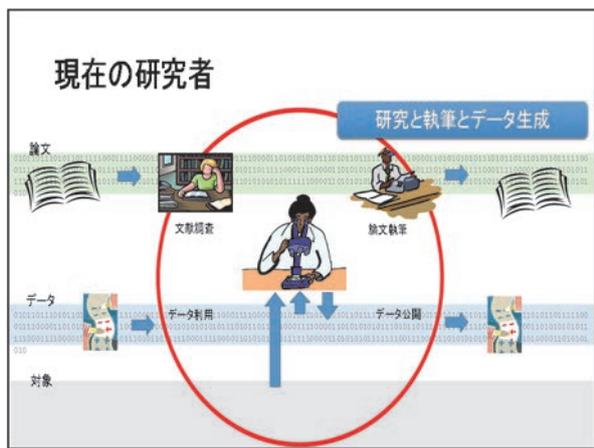
科学のタイプと再現性

再現性の確保は、第1の科学（実験科学）であれば再実験による仮説検証、第2の科学（理論科学）であれば演繹、第3の科学（シミュレーション科学）であれば、再シミュレーションによる検証によってなされます。しかし、第3の科学になると、実際に再現するのは非常に難しくなります。本当に同じ環境や同じデータがないとできないことがあるからです。第4の科学（データ中心科学）になるともっと難しくなります。つまり、第3、第4の科学になると、実際に検証可能かどうかを論文から察することができないので、もう論文は研究の成果ではなく、研究の紹介でしかなくなってしまいます。これはある研究者が述べていることで、大体私も同意します。

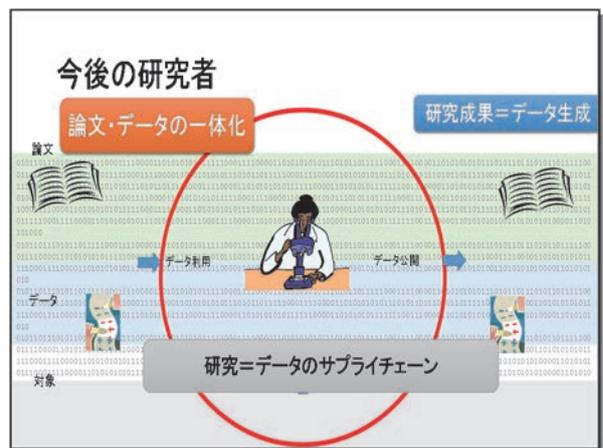
では、研究の成果はどうやって公開するかというと、仕方がないので、研究のプロセスを全て公開するしか



(図1)



(図2)



(図3)

ありません。プログラムやプロセス（ワークフロー）、どのデータをどのように処理して何をやったかを公開せざるを得ません。特に第4の科学（データ中心科学）になると、もう論文の役割が変わりつつあります。これが私の理解する、論文、研究の現状です。

研究データ流通を支える情報基盤とは

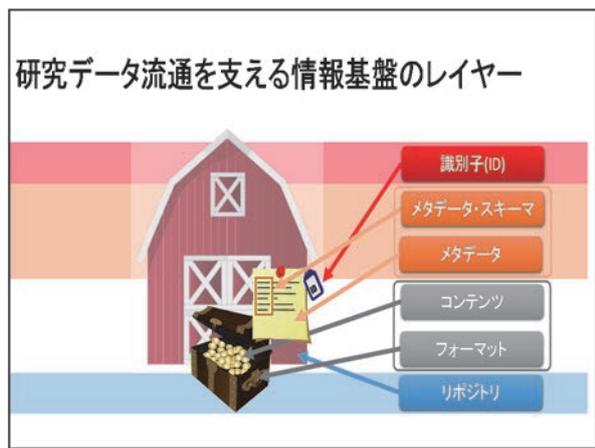
さて、そうした現状の中で、研究データ流通を支える情報基盤をどう考えればいいのかということが今日の本題です。ここでは大まかに研究データと言っていますが、研究データは、いろいろな研究分野で作られる、研究者にとってのお宝です。その分野の研究者は、それが何であり、どういうものであるか、どんなフォーマットで書かれているかも分かっています。同じ研究コミュニティの中では、あのデータは価値がある、例えば京都大学のあの標本は価値があるということは分かるわけです。ですから、その分野ではあまり困っていないのです。今までもやってきたのだから、何も問題は変わらないではないかと思われる人もいるかもしれませんが、オープン性を考えるときには変わってきます。分野を超えた連携を行う段階、あるいはオープンにする段階で、自分たちの分野以外の人たちもデータを使うことがあるということを考えたときに、現状に問題があるのは明らかです。

そのために、普通はメタデータを付けなければいけません。しかし、メタデータも勝手気ままに使われると困るので、メタデータ・スキーマが必要になります。

研究の現場の人なら、メタデータがなくても分かるので、あまり気にならない問題ですが、オープン性の観点では、それでは困ります。さらに、メタデータを付けるだけではなく、それが一意に同定できるように ID も付けなければいけません。ここまで来てはじめて、データが分野を超えて流通できる仕組みになるのです。

一方、これは研究者の現場だけではなく、機関が保持する仕組みも必要で、それはリポジトリです。大まかに分けても、識別子（ID）、メタデータ・スキーマ、メタデータ、コンテンツ、フォーマット、リポジトリというレイヤーが必要です（図4）。これらのレイヤーは、共有という意味では、個別の機関や個別の分野で解決しては駄目で、みんなで解決しなければいけない問題になっています。コンテンツ、フォーマットは、各研究コミュニティの中で考えても解ける問題です。しかし、識別子（ID）やメタデータ・スキーマ、メタデータ、リポジトリは、そのコミュニティだけではなかなか解けない問題です。リポジトリの話は重要なのですが、私の今日の話提供はむしろそれ以外、すなわち識別子（ID）、メタデータ・スキーマ、メタデータの方です。

それぞれに抱える問題があります（図5）。例えばメタデータやメタデータ・スキーマであれば、どんな言語で書くのか、どのように収集、検索、共有、変換するのかという問題が関わってきます。識別子（ID）であれば、どんなところが管理するのか、どんな体制



(図4)

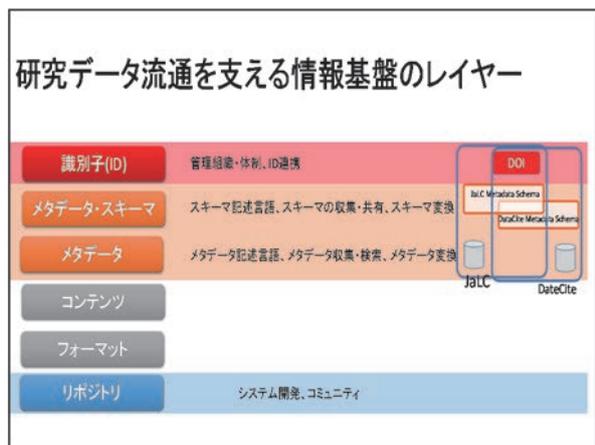


(図5)

にするのか、ID 同士の連携はどうするのかといった問題が関わってきます。リポジトリでは、システム開発、あるいは、ニーズに合わせたものをコミュニティでどうやって作っていくかが課題になると思います。メタデータ・スキーマでは、関係する組織や関係するスキーマ、システム、技術などの課題があります。このような課題が、たくさん同時並行で動いているのが現状です。

ちなみに、この後にお話しする DOI の仕組みは図 6 のように位置付けられます。DOI そのものは識別子でしかありませんが、例えば JaLC や DataCite が、DOI に付随するメタデータのスキーマを決めています。それぞれがメタデータのデータベースを持っています。そういう意味で、識別子 (ID)、メタデータ・スキーマ、メタデータという三つのレイヤーに関わることになっています。

JaLC や DataCite は、分野を超えた、共通性の高いメタデータしか集めないし、提供しません。ですから、本当にミニマムなメタデータということになります。それぞれの分野では、それではもちろん足りないわけです。そのため、研究コミュニティごとにメタデータを決めて、そのメタデータと DOI のメタデータが連携するという仕組みを作ることになると思います。従って、DataCite や JaLC は、分野ごとのメタデータに直接はコミットしないということです。その連携をするためにどうすればいいかということを考えていくことになります。



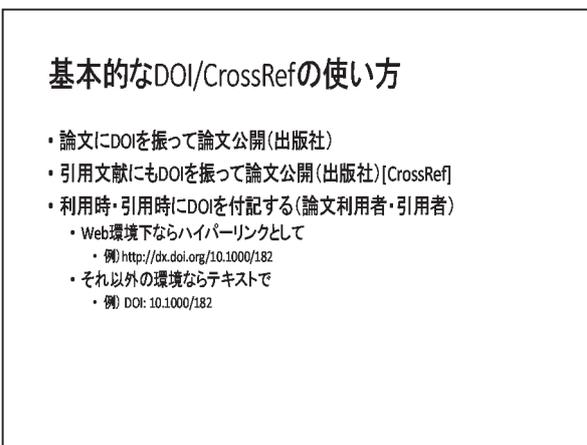
(図 6)

DOI の使い方

DOI の役割は、解決可能 (resolvable)、持続可能 (persistent)、相互運用可能な (interoperable) リンクを提供することです。1998 年に、もともとは出版社がジャーナル論文の識別子を共有するために作った制度ですが、今はさまざまなオブジェクトに使われるようになりました。ここが一つのポイントです。もともとは明らかに論文の識別子でしたが、この仕組み自体は非常に一般的な仕組みを提供していて、実はインターネット上にあるどんなオブジェクトでも構わないのです。そこで、まさに DataCite のように、研究データにも使おうという話が出てきたということです。基本的な使い方は、論文の場合、論文に DOI を振って、引用文献にも DOI を振った状態で論文を提供するというものです (図 7)。

DOI のメリット

それによって得られるメリットは、論文著者にとっては、自分の論文が読まれやすく、引用されやすくなることです。これは、短い記号列で書くことができ、簡単に参照されやすくなるからです。論文の読者にとっても、その裏返しで、確実に論文の本文テキストにたどり着けるようになります。もし DOI がなかったら、仕方がないから Google に聞いてみたりするけれど、Google に聞いても、必ずしも当たるとは限りません。うまく書誌のカット・アンド・ペーストをしないと出てこない上に、そもそも Google に引っ掛から



(図 7)

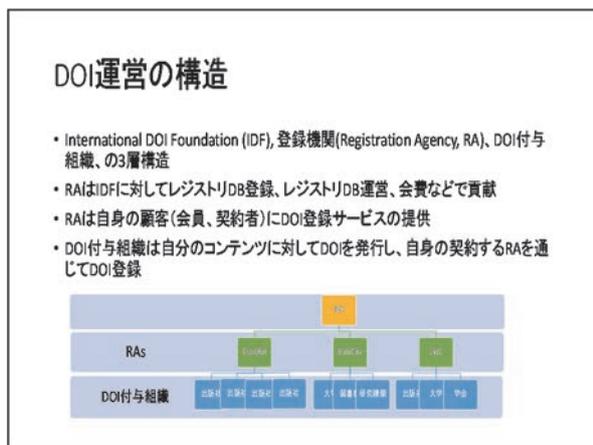
ないものもあつたりするので、なかなかうまくいかないのです。あるいは、URL を直接論文に書かれてしまうと、その URL が変わってしまうかもしれません。また、引用文献の同定も楽になります。

出版社にとってのメリットは、確実に自社サイトに誘導できることです。これは議論があるところで、特に機関リポジトリにとっては、著者版に誘導されないという問題があるのですが、それは少し置いておきます。また、引用文献の同定が楽になります。これは大きなところですが、要するに、自分の出版する論文の引用文献に DOI を付与すると、引用が楽になる。そうすると、結果的に、資金提供者や大学管理者は引用文献を簡単に同定できるようになる。従って、その評価がしやすくなるということです。

こういうことがメリットとして挙げられます。これは論文の話なのですが、まさに DOI が現在データに対してできることも、ほぼこれをデータと読み替えたものになります。これと同じようなメリットは、データに DOI を付けたときも期待できるということです。逆に言うと、これ以上のことを求めるのは難しいです。

DOI 運営の構造

DOI の運用構造については、International DOI Foundation (IDF) の配下に登録機関 (Registration Agency : RA) があり、その下に DOI 付与組織があるという 3 層構造になっています (図 8)。NPO である IDF が、DOI の運営をしています。その運用資金は



(図 8)

RA が払う会費から得ています。RA は RA で、また自分の会員からお金を集めて運営しています。持続可能で、解決可能で、相互運用可能なシステムを運用するためにはお金が掛かるのです。そのお金は基本的にはこの仕組みの中で賄っています。

DOI の RA は、図 9 のように 9 個ありますが、我々に関係するのは、CrossRef、DataCite、Japan Link Center (JaLC) の三つぐらいです。

CrossRef は、1999 年に始まった DOI の RA です。論文に付いている DOI はかなりの確率で CrossRef の DOI です。主な会員は出版社です。

DataCite は今から 6 年前に、ドイツの技術情報図書館が中心となって作られた、データに対して DOI を付与する仕組みです。このドイツの図書館はこれよりさらに前、今から 11 年前ぐらいから実際にデータに DOI を付与する作業をしていて、それをコンソーシアム型に拡大したのが DataCite です。この DataCite も IDF の RA です。

よく例として出されるのが PANGAEA で、図 10 がデータセットのランディングページです。DOI が書いてあり、このようにサイテーションしなさい、この論文を引くときにはこう引きなさいと書いてあります。これはジオデータなので、Abstract や Coverage など、どの辺を指しているということも書いてあります。

ちなみに、DataCite のメタデータは、実はほとんどがダブリン・コア準拠で、ダブリン・コアプラス、緯度経度を書くことができます (図 11)。少し自由度の

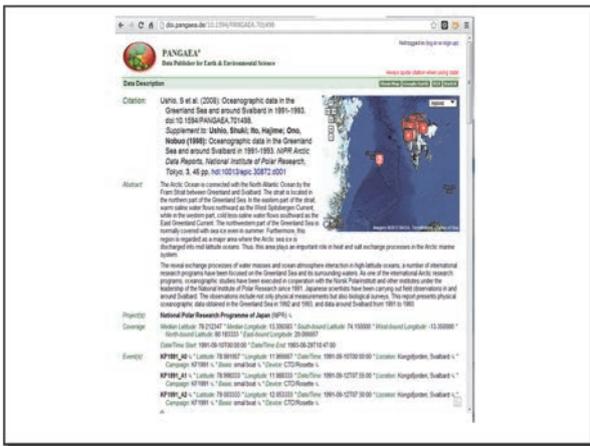


(図 9)

ある項目があって、「この〇〇というリレーションは値が〇〇」という関係を書くことができるような仕組みが入っています。そこをどう活用するかは、DataCite 自体は規定していません。DataCite のメタデータは昨年 10 月に 3.1 に変わっていて、私も最新まで付いていないので、DataCite の話は、詳しくはどこかで確認してください。

JaLC について

次に、JaLC の話に移ります。JaLC は、2012 年 3 月に世界で 9 番目に IDF から RA の認定を受けました。JST (科学技術振興機構)、NII (国立情報学研究所)、NDL (国立国会図書館)、NIMS (物質・材料研究機構) の共同の運営ですが、事務局は JST で、実際には JST の多大な人材、費用負担で動いているという状況です。私はこの中で委員長を務めています。



(図 10)

JaLC の設立趣旨は、日本初の学術コンテンツに DOI を付けたい、特に今まで CrossRef しか選択の余地がなかったところから、その選択肢を増やし、CrossRef が付けないようなものにも付けるということです (図 12)。日本語文献の適切なメタデータを取り扱い、日本にある多様な電子コンテンツを入れ、大学の紀要や日本語の学会誌、電子化由来の J-STAGE や機関リポジトリなどのものも取り込みたいという趣旨になります。

実際に関係するのは、出版社や学会、大学、研究機関、国立国会図書館、アグリゲータなどです (図 13)。CrossRef は出版社や学会しか対象ではなかったけれど、JaLC は他のものもターゲットにしています。現在、例えば博士論文は既に DOI が実際に付けられています。これは NDL と大学がうまく役割分担をしながら、DOI の付与を実際に始めているということです。NII

ジャパン・リンク・センター(JaLC)設立趣旨

- JaLCはRAの一つとして、日本発の学術コンテンツ情報を収集し、普及、利用を促進する目的で2012年に設立された。
- 日本におけるRAとしてJaLCを設立した目的は、日本発の学術コンテンツの書誌情報を網羅的に収集することによって日本国内の利活用を促すと共に、世界から日本の研究成果へのアクセス環境を向上することにある。DOIは学術コミュニケーションの重要なツールとなり、国内の研究者がDOIを利用することは日常的に普及しつつも、国内でのDOI登録は浸透していない。主たる理由として、運営母体が海外の組織であるために、日本における学術コミュニケーションの事情に合わせた対応が困難であったことが挙げられる。日本の事情とは、具体的には日本語文献の管理・流通および、日本に存在する電子化された学術コンテンツの多様性、電子化由来の多様性などである。

(図 12)

DataCite Metadata Properties

• Mandatory Properties

ID	Property
1	Identifier (with type attribute)
2	Creator (with type and name identifier attributes)
3	Title (with optional type attribute)
4	Publisher
5	PublicationYear

• Optional Properties

ID	Property
6	Subject (with scheme attribute)
7	Contributor (with type and name identifier attributes)
8	Date (with type attribute)
9	Language
10	ResourceType (with description attribute)
11	AlternateIdentifier (with type attribute)
12	RelatedIdentifier (with type and relation type attributes)
13	Size
14	Format
15	Version

(図 11)

JaLC会員と対象コンテンツ

- 出版社
 - 学術出版 (雑誌論文、書籍)
 - 学会
 - 学術出版 (雑誌論文、書籍)
 - 大学、研究機関
 - 博士論文
 - 大学出版物 (紀要、テクニカルレポートなど)
 - 機関リポジトリコンテンツ (教員出版物、OA公開物など)
 - 研究データ
 - 国立国会図書館
 - 自館出版物、パブリックドメイン出版物、博士論文
 - アグリゲータ: プラットフォーム提供者やハーベスターなど (JST, NII など)
 - 準会員の支援
- } CrossRefの主たる対象

(図 13)

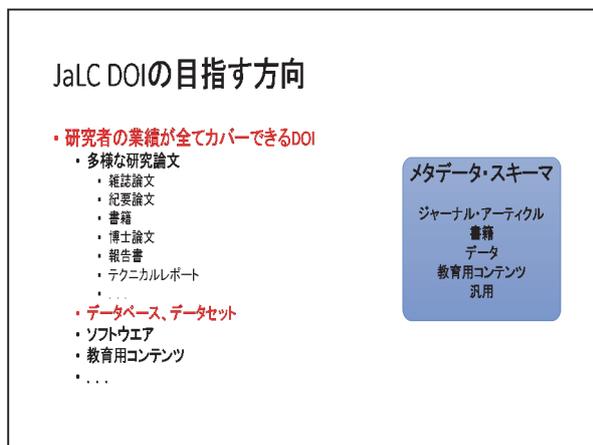
も仲介に立っています。これはコンテンツの多様性を示す一例です。

JaLC が目指しているのは、研究者の業績が全てカバーできる DOI です (図 14)。私の意見では、科研費の成果報告などは全て DOI を付けたいです。そうすれば、研究者も大学も楽になります。そういう世界をつくりたいと思っています。その中で、当然、データベースやデータセットが入ってきます。JaLC が発足して 2 年は、最初に運用したシステムが CrossRef にほぼ準拠したシステムだったため、論文以外のメタデータを入れる仕組みがうまくできませんでした。今はジャーナルとアーティクル、書籍、データ、教育用コンテンツ、汎用という形で、6 個のタイプのメタデータ・スキーマを提供するようになりました。データを入れられるようになったということがポイントになります。

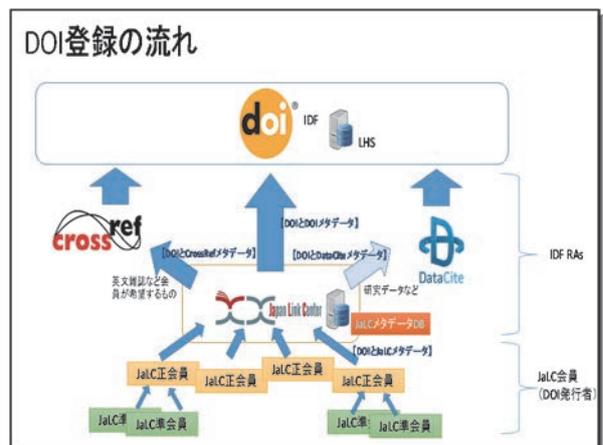
また、少し面白いのは、JaLC は IDF の会員、RA であると同時に、CrossRef の会員でもあり、DataCite の会員でもあるという二重構造を作っていること (図 15)。これは、もし CrossRef にインデックスされるのが適しているものであれば、CrossRef の DOI を取れば良いということで、その仲介をしています。ただし、CrossRef は 1 件当たり 1 ドル掛かりますが、JaLC は年会費制の会員制度を取っていて、DOI 1 件当たりの費用は取っていません。また、DataCite に登録していて、インデックスすると思えば、その DataCite との仲介をします。DataCite も会員制度なの

で、それぞれが DataCite の会員、あるいは準会員にならない限り、デポジットできないのです。それも JaLC が仲介します。しかし、JaLC は、DOI と CrossRef メタデータ、DOI と DataCite メタデータ、DOI と DOI メタデータという全てのメタデータのデータベースに登録して、メタデータとしては検索可能になるということが、直接投入するときとの違いになります。これは他の RA 組織にはない面白い仕組みなのですが、少し分かりづらい仕組みではあります。

2014 年 10 月～2015 年 9 月の 1 年間、JaLC の中で、研究データへの DOI 登録実験プロジェクトを始めました。目的は、日本における研究データへの DOI 登録の仕組みを本プロジェクトへの参加機関と共に新たに構築すること、研究データへの DOI の持続性、対象の粒度、DOI の活用などを検討すること、運用方法の検討や DOI 登録のテストを実施し、今後の安定運用につなげることでした。いざ DOI 登録を始めようとしても、ジャーナルアーティクルであれば図書館等の機関リポジトリ担当の方々もジャーナルアーティクルがどういうものであるかは大体分かっているのでもそれほど困ることはないですが、研究データは図書館の方も不慣れな上、そもそも図書館以外で、どのようところにニーズがあるかも分かりませんでした。誰が興味を持っているのか、やりたいのか分からない。どんなルールがあればスムーズに流れるのかも分からないということで、まず実験として 1 年やってみることにしました。ゴールは、研究データへの DOI 登録が



(図 14)



(図 15)

リシーの策定、新 JaLC システムにおける適切な運用フローの確立です。

現状では、図 16 のグループに参加してもらっています。左は代表機関です。プロジェクトの場合は、その配下にまたいろいろな研究所が入っているので、これは全てではありません。面白いのは、日本でメジャーなデータを作りそうな機関がほとんど入っていることです。しかも、実際にデータに DOI を振るという話なので、理念ではなくて実務として集まった初めての分野横断型のグループではないかと思っています。まさに、初めて横につながることを発見するという段階だと思います。

JaLC の研究データ用のメタデータが図 17・18 です。アールと違うのは、2 枚目で「位置情報」が入っているところ。また、「関連コンテンツ」というのは、関連も自分で定義して、その関連先の値を定義するという、かなり自由度の高いメタデータです。ここに、もし必要であれば、ここにない項目を入れてもらうという仕組みになっています。既に実験 DOI のミーティングを 2 回開いていますが、最初にしたバージョンでは、やはり不都合があって、今も改定しています。何とか 3 月中に改定しようとしています。ここはもっとこうあるべきだというご指摘を受けたので、それを直して、メタデータの更新をかけているところ。やはりやってみないと分からないところがあるので、今、やっている最中です。

検討中の課題はたくさんあります。一つ目は、一体

参加機関	
機関	DOI登録対象とするデータの概要
科学技術振興機構	生命科学分野
国立極地研究所	地球科学・環境分野 (北極域、南極域における実験及び観測データ)
国立情報学研究所 (地球環境情報統合プログラム (DIAS-P))	地球科学データ (地上観測データ、衛星観測データ、気象予測モデル、気候変動予測モデル、その他社会データ)
国立情報学研究所	機関リポジトリの保有する研究データ等 (テキスト、画像、音声)
産業技術総合研究所	地質および材料系
情報通信研究機構	超高層物理学分野 (地磁気データ、電離圏データ、オーロラ画像データ、衛星データ)
千葉大学附属図書館	一般的な研究データ、外部資金の助成による研究成果、観測データ等のアーカイブ
物質・材料研究機構	<計画変更により再検討中>
理化学研究所 脳科学総合研究センター神経情報基盤センター	脳・神経科学分野

(図 16)

どういう運用フローにするかです。例えば、プロジェクトなどがあった場合、どの機関がどこにどう取りまとめるのか。データのライフサイクルと担当者・担当機関はどうするか。どのようなデータを DOI 登録の対象にするのか。Prefix をどういう単位で与えるのか。二つ目は、アクセスの持続性の保証です。アクセスの持続性は、もちろんどの大学であっても、どの機関であっても、永遠に存在するわけではないですが、それにしても、例えば時限プロジェクトなどがあったときに、どういう扱いをすればいいのかという問題があります。

三つ目は、DOI 登録対象の粒度です。これはジャーナルやアールであればほぼ自明ですが、研究データの場合、自明ではありません。それをどうするのか。四つ目は、DOI のランディングページの要件の検討です。五つ目は、特殊なデータの特性に応じた

主なメタデータ項目		
項目	参考項目 (英語)	説明
DOI	DOI	DOI
URL	URL	URL
タイトル	Title	データのタイトル
主題	Subject	主題、キーワード、分類、等
作成者	Creator	データ作成者の名前
所属機関	Affiliation	データ作成者の所属機関
研究者ID	Researcher ID	ORCID、e-Res、科研費、等の番号
出版年月日	Publication date	研究データの出版・発行年
出版者	Publisher	研究データの出版者(研究機関等)
寄与者	Contributor	データ管理者、プロジェクトマネージャー、研究資金配分機関、等
版	Edition	Variation (出版版、著者版、等) Version (1.0、2.1等) Format (ファイルフォーマット)

(図 17)

項目	参考項目 (英語)	説明
関連コンテンツ	Related content	DOI、URL、relation type (引用、補足、別バージョンの存在、等)
代替識別子	Alternate ID	ローカルDBのIDなど
言語	Content language	コンテンツの言語
日付	Date	受理日、収集日、更新日、等
リソース種別	Resource type	データセット、ソフトウェア、等
権利	Rights	Creative Commons 等
追加情報	Description	抄録、研究方法、シリーズ情報、等
署名	Signature	データの品質保証等 (JaLC 独自)
位置情報	Geolocation	地域、場所の名前、緯度経度
研究資金情報	Fund	研究資金提供機関、課題番号、等
マルチレゾリューション優先度	Multiple resolution priority	ミラーサイトやアーカイブサイトがある場合のランディングページ優先度

(図 18)

取り扱いです。例えば DOI 登録後にデータ追加がある場合など、動的なデータはどうすればいいか、データが多い場合はどうするかなどです。六つ目は、研究データに登録した DOI の活用です。このような問題が未解決で、これから DOI のポリシーミーティングを開いて考えることになっています。

データのライフサイクルと担当者・担当機関

例えば、データのライフサイクルと担当者はどういう関係なのかを考えると、頭がとても痛いです。例えばコンテンツは作られて、保存されて、公開されて、修正されて、消えます。メタデータも作成されて、登録されて、修正されます。そして DOI に登録されます。今までの機関リポジトリであれば、コンテンツは研究者が作って保存し、図書館がそれからメタデータを作成して登録して、DOI も登録していました(図 19)。このように、割と分担の仕方が分かりやすかったのです。

しかし、データになるとどうなるでしょうか。例えば、研究者が保存、公開までずっとやるのか。修正が多いときにはどうなるのか。そのときに、例えば図書館は作成、登録、修正までも対応してやらなければならないのか。どういうワークフローで動くのか。あるいは、研究者ではなくてプロジェクトが中間にある場合でプロジェクトがメタデータ作成をするのであれば、図書館とどう連携するのかということも出てくるわけです。一方でプロジェクトもいろいろな性格がありま



(図 19)

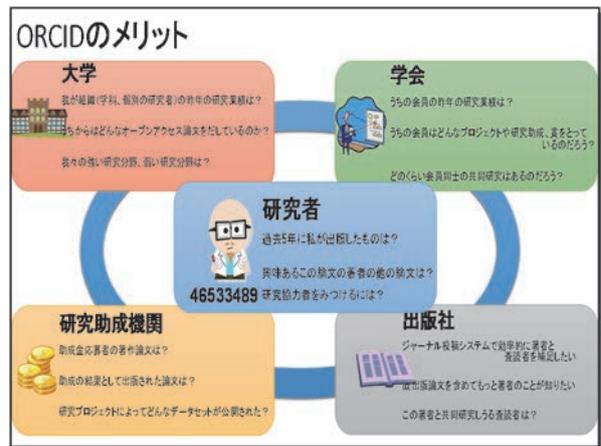
す。プロジェクト、あるいは研究センターのようなものがあつたときに、そこがメタデータ作成をやってくれるかもしれないし、やってくれないかもしれない。この辺をどのように分担するのかということが大きく問題になります。もちろんこれに唯一の答えがあるとは言いませんが、こういう問題をある程度明らかにしなければ、ワークフローが作れません。

研究者の識別子

もう一つ、重要な識別子として、研究者の識別子があります。研究者の識別子は DOI では書けません、書くべきではありません。生きているオブジェクトなので、扱いが違うからです。実体としての研究者の多面性があるので、それは別の ID を振るのが正しいだろうということです。ユニーク性と多面性を両方、担保しなければいけないのです。

それに対して、ORCID (Open Researcher and Contributor Identifier) という NPO が 2011 年から活動しており、著者や研究寄与者を一意に同定するための識別子を付与しています。一つ一つの研究者 ID は、研究者本人や大学、学会、出版社、研究助成機関など、いろいろな側面から使われます。このような多面的な視点があり、それに対してそれぞれメリットがあります(図 20)。ORCID は、ID というのはハブだと言って、これらをつなぐところにあるのが ORCID の ID になります(図 21)。

ORCID は順調に伸びていて、ちょうど今年の時点



(図 20)

で 100 万 ID になりました (図 22)。濃い青の「Direct via orcid.org」は、ORCID サイトに直接登録した件数で、これは自主的に入れたことが多いことを示します。薄い青の「Via integration」は、他のシステムから入れた件数です。多くは出版社の manuscript management system あたりから入ってくるものです。投稿しようと思うと、「ORCID ID がありますか」と聞かれるので、その時点で ORCID ID を取って入れるということが結構増えています。従って、ORCID は論文投稿時には割とよく使われるようになりつつあるということになります。ちなみに、DataCite のメタデータも、JaLC の新しいメタデータも、研究者 ID を入れる欄があって、そこに ORCID を選んで入れることができるようになってきています。特に研究データの場合、純粋な研究者ではなく、コントリビューターというデータを作成した人などが入ったりすることがあります。これはどうしても目に見えない貢献なのですが、DataCite でコントリビューターのところに、そういう人の名前や ID を入れることによって、業績が可視化されることが期待されます。登録研究機関は、欧米系が多く、アジアはまだ少ないです (図 23)。

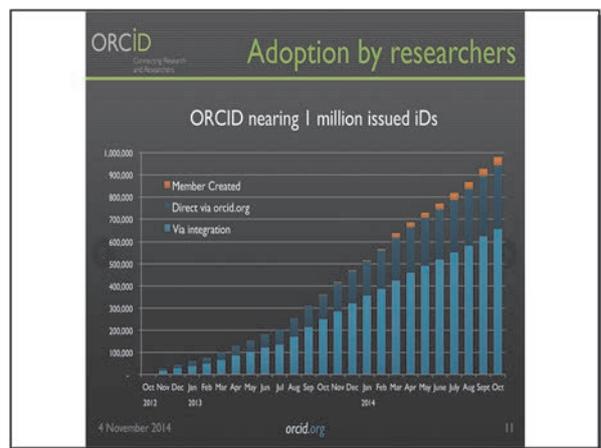
まとめ

研究成果はいずれ「データ」になります。そのため、研究データ流通基盤は必須の仕組みで、研究データ流通には、識別子、メタデータ・スキーマ、メタデータ、コンテンツ、フォーマット、リポジトリという幾つか

のレイヤーがあります。

DOI は研究データ流通の重要な識別子です。これまでは論文に対する識別子でしたが、これからはデータに対する識別子でもあります。現在、JaLC ではデータ DOI 実験プロジェクトを実施中です。国内では初めて分野を超えた研究データの実務者が集まっている会議になっていて、まさに理念ではなくて、実務として、どう問題を解いていくかということについて、やっと議論が始まったところです。

たくさんの課題があります。研究データ流通については、例えば DataCite の前身は 10 年前からやっていて、我々はその 10 年遅れで、やっと今の段階に立っています。そのような意味では大変遅れているのですが、これは情報流通なので、協調しなければいけません。ですから、別にオリジナルを作らなくていいところは全部利用すればいいのです。



(図 22)



(図 21)



(図 23)

ただ、一方でやはり競争もあります。例えば ORCID には似たような ISNI (International Standard Name Identifier) というものがあって、ISNI とある種の競争だったりします。このように、今、協調と競争が世界中で起こっています。早く日本人たちもそれに交じらなければいけないということを最後に訴えたのです。ぜひ皆さんにご参加いただけたらと思っています。