

第 2 回 SPARC Japan セミナー2015

「科学的研究プロセスと研究環境の新たなパラダイムに向けて
- e-サイエンス, 研究データ共有, そして研究データ基盤 -」

日本の研究データ基盤

加藤 斉史

(科学技術振興機構)

講演要旨

研究データの共有、再利用を図るために、日本における永続的な識別子 DOI の登録機関であるジャパンリンクセンターは、研究データへの DOI 登録ができるよう 2014 年 12 月にシステムの拡張を行った。一方で、研究データへの DOI 登録に関しては、世界的にみても様々な課題の検討が進められている段階であり、日本においても課題抽出、検討を進める必要がある。そのため、JaLC において研究データへの DOI 登録を開始するのに先立ち、DOI 登録実験プロジェクトを立ち上げ、研究データへの DOI 登録に特有の課題抽出とその解決、運用方法の確立、DOI の活用方法などの検討を行ない、報告書およびガイドラインとしてとりまとめた。



加藤 斉史

教育サービス会社で e ラーニング教材の開発業務に従事した後、2000 年に科学技術振興事業団に入職。現在、科学技術振興機構知識基盤情報部にて、ジャパンリンクセンターの運営業務に従事。

本日は DOI を登録する日本の基盤である Japan Link Center (JaLC) と、JaLC が研究データへの DOI 登録を始めるまでの取り組みについてご紹介したいと思います。

Japan Link Center と DOI

初めに、JaLC について簡単に説明します。JaLC は DOI を登録する機関です。DOI とは標準的なパーマネントな識別子の一つで、国際 DOI 財団

(International DOI Foundation, IDF) が取りまとめており、JaLC はその下で活動しています。日本には、世界で 9 機関ある DOI の登録機関の一つとして JaLC があります。JaLC では、学術論文や書籍、研究デー

タを対象として DOI を登録しています。DOI を登録するに当たってはコンテンツの所在情報も管理しています。国内の学術機関が共同で運営をしており、事務局は科学技術振興機構、現在の会員数は正会員が 24 機関、準会員が 950 機関です。

続いて、研究データに対する DOI の登録状況について説明します。研究データへの DOI 登録は DataCite が先行して実施しました。DataCite は世界で九つある DOI 登録機関のうちの一つです。2004 年にドイツ科学技術図書館で DOI 登録を開始し、DataCite の活動もそこから始まっており、現在の DataCite 正会員は 23 機関、準会員 8 機関です。正会員の 23 機関はさらにその下に機関を持っているところも多いので、

23 機関というよりは、その下にさらに広がりがある、世界中の研究データを扱うような機関が参加している、割と大きな組織だと思っていいです。

JaLC は自身が DOI を登録する機関であると同時に、DataCite の会員にもなっています。ですから、JaLC は DataCite を経由するパターンと JaLC 自身で DOI を登録するパターンを持っています。さらに、学術論文、書籍を中心として DOI 登録を行っている CrossRef の会員として、CrossRef を経由して DOI を登録するパターンもあります。最終的に同じ DOI になるのですが、DataCite、CrossRef を経由することにより、DataCite、CrossRef のデータベースにもデータを登録することができます。JaLC 自身の登録だと DataCite、CrossRef のデータベースには入りません。

研究データに対する DOI の登録を始めるに当たって、研究データへの DOI 登録実験プロジェクトを立ち上げました。これを始めるに至った経緯は、平成 26 年 12 月に JaLC における DOI 登録対象コンテンツを、研究データや書籍、報告書、e ラーニング等といったこれまでの論文以外へも拡大したことです。そこで、研究データも DOI 登録がシステムとしては可能になったことに加え、研究データの公表、保全、再利用への関心も高まっている中で、DOI 登録に関して課題の検討の必要がありました。参考にするようなノウハウや日本の事例にうまく適用できるようなノウハウが少なかったのです。

運用方法の検討や DOI 登録のテストを実施して今

後の安定運用につなげることを目的とし、研究データへの DOI 登録のポリシーを策定すること、適切な運用フローを確立することをゴールとしてプロジェクトを進めていくことにしました。

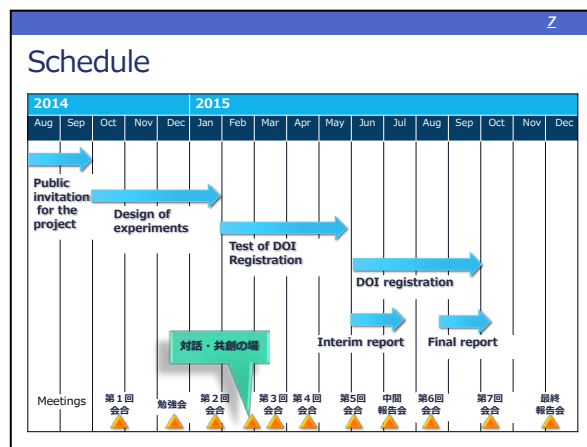
プロジェクトの内容は、研究データに対する DOI 登録の方式検討、DOI テスト登録（JaLC へのデータ登録、更新、削除等のテスト）を行った上で、DOI 本番登録をするというものでした。その他、DOI 登録の方式検討の結果なども含めて、プロジェクトの実施報告書、登録ガイドライン、登録手順書を成果物として取りまとめる予定でスタートしました。

参加に当たっては、研究データへの DOI 登録に関わろうとしている機関には広く参加してほしかったので、1 カ月弱という短い期間ですが、JaLC のサイトで公募しました。図 1 にあるように、図書館や DIAS も参加しています。また、リーダーとして国立情報学研究所（NII）の武田英明先生、サブリーダーとして情報通信研究機構（NICT）の村山泰啓先生に入っていました。

スケジュールとしては、1 年間のプロジェクトで進めました（図 2）。公募を行い、参加機関を募った上でプロジェクトをスタートして、全 7 回の会合と、「対話・共創の場」や中間報告会なども設けています。大体 2014 年 10 月～2015 年 1 月に方式検討を行い、2015 年 2～5 月にテスト登録をして、2015 年 6～9 月に実際の登録を行いました。途中で中間報告会を設けて、ガイドライン等の状況について中間報告をして意

Participants	
機関	DOI登録を検討するデータの概要
科学技術振興機構	生命科学系データベースアーカイブの収録データセット
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	データリポジトリの収録データ（歴史的資料物写真データ、科学衛星観測データ）
産業技術総合研究所	材料系データベース等の収録データ
情報通信研究機構	超高層物理学分野（地磁気データ、電離層データ、オーロラ画像データ、衛星データ）
千葉大学附属図書館	機関リポジトリの収録データのうち、主に植物標本
物質・材料研究機構	材料科学分野（電子顕微鏡画像データなど）
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・神経科学分野データベースの収録データ

(図 1)



(図 2)

見をもらいつつ、ただ今、最終報告書が取りまとまったという段階です。

JaLC を使った DOI 登録の流れについて一度ご説明しておきます。まず、先にできていた論文の方から説明します。図 3 の真ん中にあるのが JaLC で、会員が登録したい DOI 名とそれに対するメタデータを JaLC に送ります。JaLC 自身が RA として IDF に送るという流れと、論文の場合は CrossRef という RA のメンバーになっており、CrossRef を経由して IDF に送るという流れがあります。最終的には同じ IDF のサーバーにある DOI のレジストリに登録されるのですが、その途中、CrossRef を使う場合には CrossRef のデータベースにも登録されるというところが違いになっています。

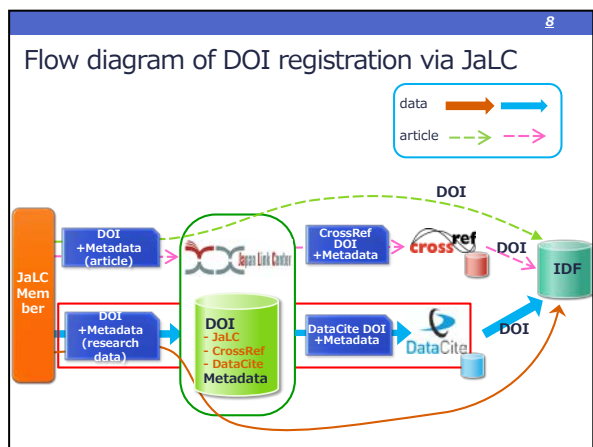
続いて、研究データの流れを説明します。同じように、JaLC に会員から DOI とメタデータが登録されます。JaLC が直接 IDF のデータベースに登録するパターンと、DataCite を経由して登録するパターンがあります。

メタデータの定義

JaLC で研究データの対応を進めるに当たって、DOI と一緒に JaLC で登録するメタデータについては項目を定義しました。原則として DataCite の定義に沿い、DataCite で Metadata Schema 3.1 として公開されているものを参考にしました。ただし、JaLC では先行して論文のメタデータ定義が決まっていたので、その

他のメタデータと横並びということを意識し、タグ名の付け方、複数言語の持ち方、ファンド情報やマルチプルレゾリューションの項目追加について可能な範囲で横並びでそろえています。このあたりが DataCite と少し違うところです。また、変更という点では、研究データにおいて、データの品質などを保証することを記載する記述が必要であろうという理由で、JaLC で独自に signature というタグを入れています。

具体的なメタデータ項目は図 4・図 5 のとおりです。DOI 名は JaLC が発番するのではなく、登録したい人の方で発番して JaLC に送ってくることになります。また、DOI の実際の飛び先となる URL もメタデータとして登録のときに必要になります。その他に、タイトル、キーワード、作成者、機関、研究者 ID、寄与者なども入っています。版を書くこともできます。版はその下にバリエーションとバージョンとフォーマット



(図 3)

10 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-Rad、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、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 4)

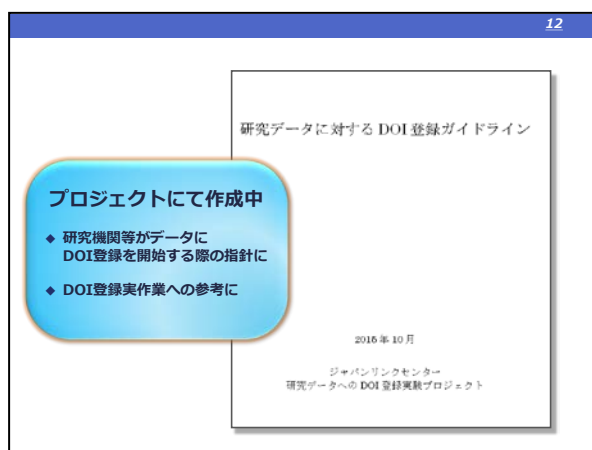
11		
項目	参考：項目（英語）	説明
関連コンテンツ	Related content	DOI、URL、relation type（引用、補足、別バージョンの存在、等）
代替識別子	Alternate ID	ローカルDBのIDなど
言語	Content language	コンテンツの言語
日付	Date	受理日、収集日、更新日、等
リソース種別	Resource type	データセット、ソフトウェア、等
権利	Rights	Creative Commons 等
追加情報	Description	抄録、研究方法、シリーズ情報、等
署名	Signature	データの品質保証等（JaLC独自）
位置情報	Geolocation	地域、場所の名称、緯度経度
研究資金情報	Fund	研究資金提供機関、課題番号、等
マルチプル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 5)

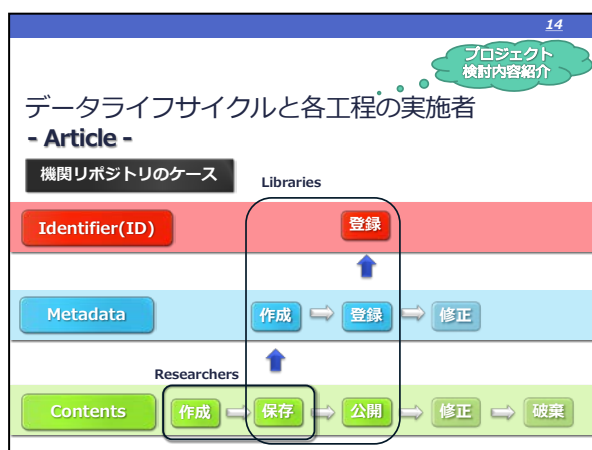
トの三つを設けています。

また、関連するコンテンツとして、DOI を付けたものと別のコンテンツとの関係性も記載できるようになっています。例えば、派生してできたものや、親コンテンツなどを書くことができるようになっています。その他、signature はデータ品質保証用として JaLC 独自で入れています。

プロジェクトでは、DOI を登録する際の研究データ独自の指針が必要であろうということで、ガイドラインを定め、DOI の登録作業のときに参考にさせていただくことを検討しました（図 6）。DOI 登録のワークフロー、DOI 登録の対象データ、DOI としての持続的な識別子として求められているアクセスの持続性をどう保証するか、DOI を登録する対象の粒度をどうするか、DOI のランディングページにどのような情報を記載しておくべきか、研究データに登録した



(図 6)



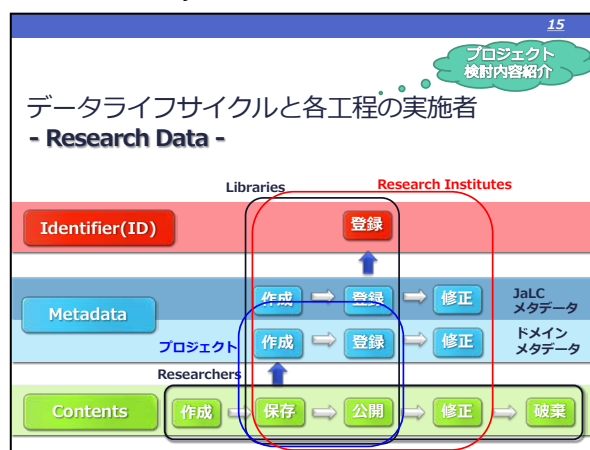
(図 7)

DOI をどう活用していくのかということをもとめて、参考になるようにしています。

考えたワークフローが図 7 です。まずは論文の機関リポジトリのケースです。一番下のところが実際に論文が生まれていくところです。研究者が論文をつくって、図書館の機関リポジトリに保存して、図書館側の作業として公開します。その際にメタデータを図書館で取って、それを登録していきます。その登録したメタデータを基に、DOI 登録用のメタデータが作られていくという流れになっています。

次は研究データのケースを見ていきます（図 8）。データが生まれてメタデータがつくられて DOI として登録するという大きな流れは論文のケースと変わりませんが、メタデータの層の役割が増えています。また、分野ごとのメタデータと、DOI 登録するために必要なメタデータの層が違うのではないかとということで、二つに分けています。

下からもう一度説明していくと、研究データが生まれて、それが保存されて公開されていきます。個人個人の研究者がプロジェクトの中で研究データを取りまとめていくことになります。そのプロジェクトの中で、うまくデータを共用するというので、そのドメインの中のメタデータが作成されて登録されるということになります。このメタデータの登録作成については、図書館の機関リポジトリと同じようなパターンもあれば、研究機関内で行われている場合もあると考えています。さらに、JaLC に DOI を登録するためのメタデ



(図 8)

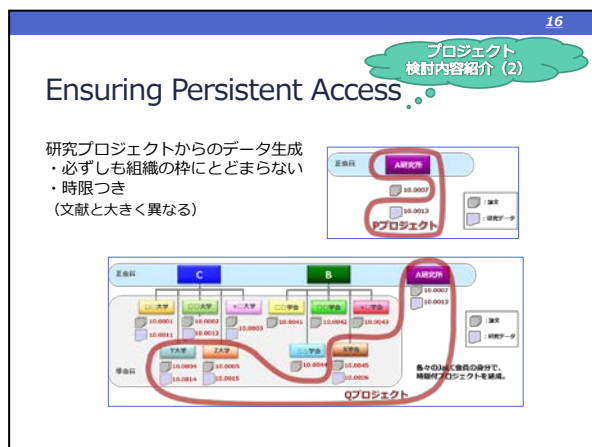
ータを作成して登録していくという流れになっています。メタデータについてはどのように使うかによって、ドメイン中のメタデータと、DOIとして付けるところが違ってくるといことになります。

アクセスの持続性の保証については、データを付けるに当たって、DOIのメタデータやコンテンツとしてのデータ自体をどう管理するかを検討しました。図9にあるように、いろいろなパターンがあるのですが、代表する機関がそのプロジェクトの中のDOI登録したデータを一元化して管理するというパターンもあるのではないかとということで、必ずしもJaLCの会員という組織の枠だけで考えるのではない、プロジェクトとしての単位があり得るといことも考えました。

今後の取り組み

今後、これまでできていた論文のDOI登録の他に、研究データについても実験プロジェクトが終わり、このようなことを踏まえてDOI本登録が進んでいくことになります。DOIは、論文にも研究データにも付いているということで、関連情報がだんだん充実してくるので、DOIを使った論文間の引用や、論文からの研究データの引用ができるようになり、評価にも使えるようになると考えています。実験プロジェクトは10月で終了して、最終報告会は来る12月3日に予定しています(図10)。こちらについても、ぜひご関心ある方は参加いただければと思っています。実験プロジェクトに集まっていた方、あるいはその周辺

の方と意見交換するような機会も今後も設けていけたらと考えています。



(図9)

Figure 10 is a poster for the "Research Data DOI Registration Experiment Project Final Report Meeting". The poster is titled "研究データへのDOI登録実験プロジェクト 最終報告会". It includes the following information:

- 日時:** 2015年12月3日(木) IDF Outreach meeting*の中で開催
- 会場:** 国立研究開発法人科学技術振興機構 東京本部 (詳細は後日決定)
- 主催:** ジャパンリンクセンター運営委員会
- 参加費:** 無料

* IDF Outreach meeting について

- IDF (The International DOI Foundation: 国際DOI財団) は、毎年冬にOutreach meetingを各国で開催。
- 今年は、日本およびアジア地域におけるDOIの普及のため、JaLCのホストにより、東京で開催。

昨年度のDOIアウトリーチ会議については集会報告として <http://doi.org/10.1241/johokanni.57.936>に掲載中。

Outreach meeting 2014の様子 (The Grand Visconti Palace Hotel, Milan, Italy)

詳細が決まり次第、ご案内をJaLC webサイトへ掲載予定。

(図10)