

第 5 回 SPARC Japan セミナー2012

「Open Access Week — 日本におけるオープンアクセス, この10年これからの10年」

eScience への広がり: Open Repository

大園 隼彦

(DRF/岡山大学附属図書館)

講演要旨

欧米ではオープンアクセスに加えてオープンデータの動きが始まっており、Open Repositories 2012 では機関リポジトリでの研究データの保存、管理、再利用に関する発表が数多くあった。これらの事例や最近の関連する動向を紹介し、国内の大学図書館がデータ管理に係わる可能性について考える。



大園 隼彦

2002年より図書系職員として勤務、岡山大学の機関リポジトリの立ち上げに携わる。

2009年よりデジタルリポジトリ連合 (DRF) 技術サポートワーキング・グループに参加。

Open Repositories 2012 概要と所感

Open Repositories とは、デジタルリポジトリの開発や運用に携わる関係者が世界中から集まり、情報の共有や議論を行う国際会議です。7回目に当たる

Open Repositories 2012 は、7月9日から13日にエジンバラ大学で開催され、30カ国超から450人以上の関係者が出席しました。

私が参加したセミナーのうち、特に話題となったものを二つ紹介します。一つがオープンアクセスを促進するための共同リポジトリや共同インフラについての話でした。特に、イギリスの助成研究の成果を公開する共用インフラを用意している UK RepositoryNet+ について、また、EU の助成研究の成果を公開するためのインフラ整備と、リポジトリの連携促進によるア

クセス性の改善を目指す OpenAIRE が話題に上りました。日本においても、大学図書館で共同リポジトリが進んでいます。こちらは参加機関が増加して、コンテンツの量の増加からさらに一歩進んで、サービス連携等を進めて、コンテンツのアクセス性の改善を進めようという話を中心だったかと思います。

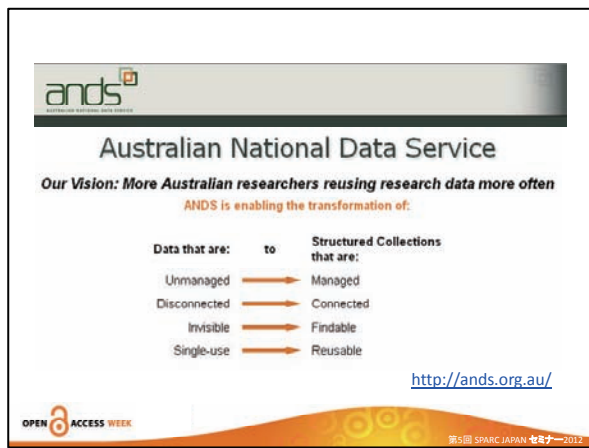
二つ目に、コンテンツのアクセス性や連携に関連して、研究データ管理についての話も大きく取り上げられていました。データの長期保存、データの永続アクセス、データの再利用、データの引用等のために、研究データを管理する必要があるのではないかとということです。特徴的だったのが、研究データを保存して統合検索するためのインフラを整備しているという発表で、グリフィス大学の Research Hub やオックスフ

オード大学の DataFinder といったシステムが紹介されていました。

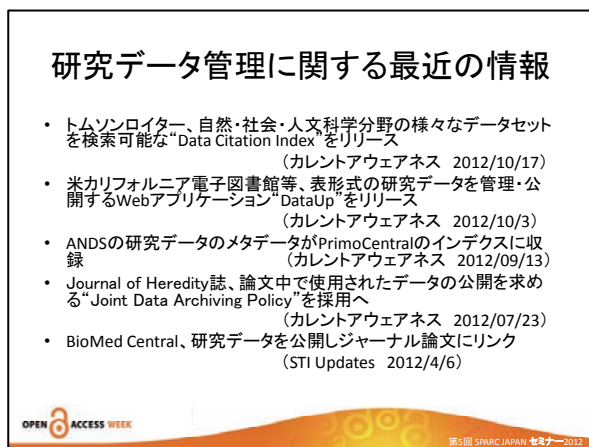
研究データ管理の分かりやすいコンセプトとして、ANDS (Australian National Data Service) をご紹介いたします (図 1)。これは、より多くのオーストラリアの研究者が、さらに研究データを利用できるようにすることを目的としています。データを結び付けて発見できるようにして、再利用できるようにするということをうたっていました。

日本における研究データ管理を考える

研究データ管理に関する最近の情報を調べてみると、研究データの公開、データの再利用についての話題がたくさん挙がってきています (図 2)。こうした状況を鑑み、日本でも研究データ管理を始めるべきではないかというのが、Open Repositories に参加しての率



(図 1)



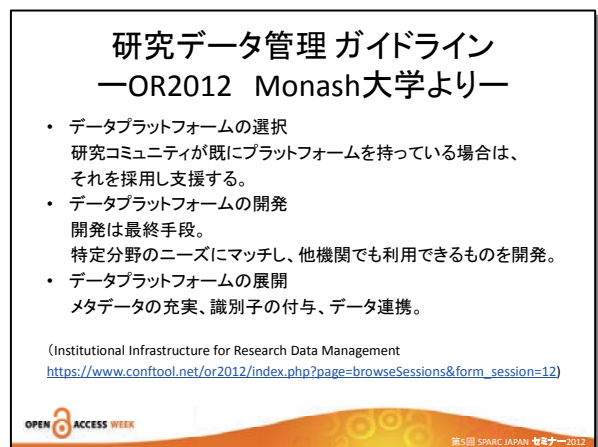
(図 2)

直な感想です。特に先ほどご紹介したデータの長期保存、アクセス、再利用に対しては、図書館の機関リポジトリのノウハウが生かせるのではないかと考えています。まずは図書館でできることを進めるべきではないかということで、まず保存すること、次に引用できる環境を構築していくべきではないかと思いました。そこで、何を保存するのか、どこに保存するのか、どのように保存するのか、サービスの展開はどうしたらいいのかということを考えてみました。

これは、Open Repositories に参加していたモナシュ大学のガイドラインです (図 3)。まず、データプラットフォームの選択について、研究コミュニティが既にプラットフォームを持っている場合は、それを採用して支援するのがいいのではないかと述べられていました。また、データプラットフォームの開発について、開発は最終手段であり、特定の分野のニーズにマッチして、さらに他機関でも利用できるものを開発した方がよいと発表されていました。もう一つ、データプラットフォームの展開に関しては、メタデータの充実や識別子の付与、データ連携が求められると発表されていました。

何を保存するのか

これは、Open Repositories でも発表していたオックスフォードの DaMaRO プロジェクトのスライドを参考にしました (図 4)。研究データとは、研究で利用する全ての情報を指し、そのうち構造化データとし



(図 3)

データベースや表、非構造化データとしてはテキスト、イメージ、音声、記録、メール等を保存した方がいいのではないかとのことでした。

最近出た「ビッグデータ時代におけるアカデミアの挑戦」という報告書では、観測データやソーシャルデータといった規模や発生頻度の大きいものについて触れられていましたが、図書館がすぐにそこまで行くのは難しいので、今回はそこについては考慮せず、できる範囲のところで話を進めていきたいと考えています。

どこに保存するのか

次に、どこに保存するのかということです。モナシュ大学の事例でもあったように、図書館がするのならば、当然、既存のプラットフォームを活用した方がよいでしょう。既に多くの大学図書館が機関リポジトリを持っているので、それを活用して相互運用性での貢献ができるのではないかと考えました。その他の可能性としては、これもモナシュ大学の報告にありましたが、研究者はコミュニティのシステムをより利用する傾向にあります。例えば Dryad といった生物科学分野のデータリポジトリがあります。また、新設のプラットフォームとしては、Open Repositories で紹介されていた Research Hub や DataFinder や Open Exeter などが当てはまると思います。今日は図書館の取り組みということで、既存のプラットフォームの活用についてお話ししたいと思います。

国内の機関リポジトリに現在どれぐらいのデータが

(図 4)

入っているのか、JAIRO で調べたところ、「データ・データベース」というところに 5 万 2,659 件のデータが既に入っていました (図 5)。そのうち 5 万 2,223 件、ほぼ全てが千葉大学のデータであることが分かります (図 6)。先ほど、主に研究の過程に出てくるデータを管理・保存すべきだと言いましたが、この内容を見てみると、過程というより成果としてのデータ、あるいはデータベースそのものへのリンクや図書館の貴重書のデータが入っていたので、研究データ管理という意味ではまだまだ日本の機関リポジトリは機能していないと思いました。

どのように保存するのか

今後、リポジトリをデータ管理に使うのであれば、どのように保存するのかということが大切になってきます。まず、オープンアクセスジャーナルのようにフ

(図 5)

(図 6)

ルテキストの検索は期待できないので、メタデータを充実させることがアクセス性の改善につながります。さらに、データセットに永続 ID を付与することで、引用やアクセス性が改善します。従って、メタデータと識別子という二つにリポジトリが対応できるようにする必要があるのではないかと考えました。

メタデータの充実という点に関しては、DataCite という研究データのアクセスを推進しているコンソーシアムが Metadata Schema というものを公開しています。ここでは、識別子、作成者、タイトル、公開者、公開年を必須項目として挙げています。分類・キーワード、寄与者、日付、言語、資源タイプ、その他の ID、サイズ、フォーマット、バージョン、権利、その他の詳細情報は任意項目となっています。ざっと見た限り、これらは今、機関リポジトリの一つ標準となっている junii2 でほぼカバーできるのではないかと思います。また、今後は相互運用性をどのように考えるかが重要になってくるでしょう。

識別子の付与に関しては、まず考えられるのがデータコンテンツに識別子を付与することです。その一つが DOI です。先ほど紹介した Research Hub の場合は、Australian National Data Service が DOI を付与しており、DataFinder については DataCite が DOI を付与する予定となっています。国内においては、今、JaLC がそういうことをしようとしているので、今後データに対しても保存した場合、DOI を付与することができるのかどうか、展開に期待しています。

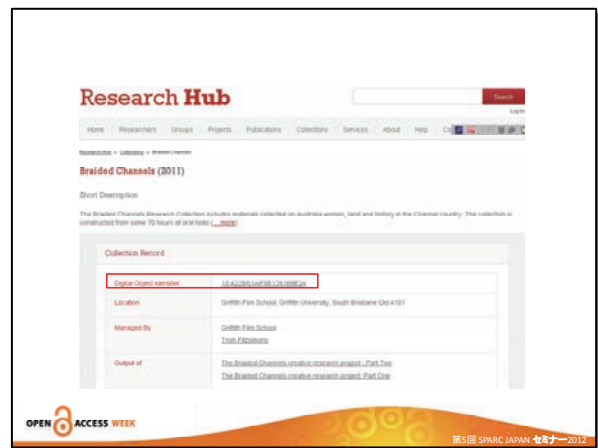
もう一つが研究者 ID で、Research Hub、ANDS と NLA (National Library of Australia) では、研究者 ID で相互にリンクするような仕組みができています。国内においては金沢大学の CSI プロジェクトで、科研費番号や研究者リゾルバー ID で名寄せをしているので、こうしたことを進めると、リポジトリの方でも名寄せができるのではないかと考えています。同時に、ORCID の動向も視野に入れる必要があると思います。

これは、Research Hub の検索画面です (図 7)。赤で囲んだ部分で分かる通り、DOI が付与されています。

これが Research Hub と ANDS が提供している Research Data Australia のサービス、それと National Library of Australia のサービスの三つの画面です (図 8)。それぞれがお互いにデータを保有していて、研究者で Research Data Australia から Research Hub に飛んだり、National Library of Australia の検索結果から Research Hub に飛んだりという仕組みが既にできているようです。

サービスの展開をどうするのか

これまでご紹介したとおり、まずは研究データを保存し、次に利用の環境を提供する必要があります。より多くのデータが保存され、さらに再利用されるため



(図 7)



(図 8)

には、ポリシーの作成や、ほかのシステムとの相互運用性、データ再利用の仕組みを整える必要があるのではないかと思います。

例えば、リポジトリ内での論文とデータのリンクや、学内システム、研究者データベースとのつながり、あるいは機関リポジトリで進めるのであれば、JAIROにデータをハーベストされているので、例えばナショナルデータという形でまとめることもできます。さらに、メタデータのところで分野という項目もありましたが、そこでサブジェクトでくくることもできるのではないかと思います。

研究データ管理支援を開始する図書館への

10の提言（欧州研究図書館協会）

今年7月、欧州研究図書館協会が「研究データ管理支援を開始する図書館への10の提言」という文書

研究データ管理支援を開始する図書館への10の提言
— LIBER (欧州研究図書館協会) —

- ① 研究データ管理の支援活動(助成金申請、知的財産の助言等)を提供する。
- ② メタデータ及びデータの基準の発展に貢献し、メタデータサービスを提供する。
- ③ データライブラリアン職を作り、専門職員を育成する。
- ④ 組織のデータポリシー開発に積極的に参加する。データのライフサイクルに適したオープンデータポリシーの採用を促す。
- ⑤ 研究者やデータセンターと連携し、データの相互運用基盤(共有、アクセス性)を促進する。

第5回 SPARC JAPAN セミナー2012

(図9)

を公開しました(図9、図10)。これは私が訳した文章なので、興味のある方は原文を当たっていただければと思います。今日触れた部分でいえば、「メタデータの基準の発展に貢献し、メタデータサービスを提供する」「研究者やデータセンターと連携し、データの相互運用基盤を促進する」「保存、発見、永続アクセスを提供することで、研究データライフサイクルを支援する」「データに永続IDを付与し、データの引用を促進する」といったことが提案されています。

また、この提言ではデータ管理ポリシーについても触れられています。参考までに、Open Repositoriesが開催されたエジンバラ大学のポリシーをご紹介します(図11、図12)。今日触れたところでは、「全ての新しい研究企画は、データの保存、管理、完全性、信頼性、共有、公開に取り組む、研究データ管理計画を含まなければならない」「大学は今後の研究データの

研究データ管理ポリシー
University of Edinburgh

1. 研究データは研究データライフサイクルを通して最高基準で管理される。
2. プロジェクトにおける健全な研究データ管理の責任は研究責任者にある。
3. 全ての新しい研究企画は、データの保存、管理、完全性、信頼性、共有、公開に取り組む、研究データ管理計画を含まなければならない。
4. 大学は研究データ管理計画のガイドラインやトレーニング、サポート、アドバイスを提供する。
5. 大学は、今後の研究データのアクセスのため、その保存、バックアップ、登録のサービスとシステムを提供する。

第5回 SPARC JAPAN セミナー2012

(図11)

研究データ管理支援を開始する図書館への10の提言
— LIBER (欧州研究図書館協会) —

- ⑥ 保存、発見、永続アクセスを提供することで、研究データライフサイクル支援をする。
- ⑦ データに永続IDを付与し、データの引用を促進する。
- ⑧ 既存のインフラを活用して保存、検索サービスを提供する。
- ⑨ 特定分野のデータ管理の実践に参加する
- ⑩ IT部門との協力、クラウドサービスを活用することで、動的・静的研究データの安全な保存を実現する。

(<http://current.ndl.go.jp/node/21693>)

第5回 SPARC JAPAN セミナー2012

(図10)

研究データ管理ポリシー
University of Edinburgh

6. 他の場所で管理されているデータ、例えば、国際的なデータサービスやサブジェクトリポジトリ、は大学にも登録する。
7. 研究データ管理計画は、適切な安全対策の下でデータがアクセス、再利用されることを保証しなければならない。
8. 分野研究データの合法的な所有は守らなければならない
9. 今後歴史的に重要になる研究データ、研究成果を検証するデータを含む全ての研究データは、適切な国または国際的データサービス、分野リポジトリ、大学リポジトリでの登録、保存を提供され、評価される。
10. ファンドの条件に関わらず、データをオープンに再利用する権利を保持せずに、データを公開・再利用する排他的な権利を商業出版社に渡すべきではない。

(<http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy>)

第5回 SPARC JAPAN セミナー2012

(図12)

アクセスのため、その保存、バックアップ、登録のサービスとシステムを提供する」。さらに、ほかの場所で管理されているデータ、例えば国際的なデータサービスやサブジェクトリポジトリも大学に登録しなければならないといったことがポリシーとしてうたわれています。

まとめ

データ管理に既存のシステムを生かすという観点では、図書館で管理している機関リポジトリを利用できるのではないかと思います。また、新しくデータ管理システムを構築する場合には、機関リポジトリ構築のノウハウを生かして、図書館がシステム連携等に貢献できるのではないかと思います。

今後、全てのデータを保存する必要があるということになるのならば、機関リポジトリがその保存先の選択肢の一つとして考えられてもいいのではないかと。そのために機関リポジトリ、図書館の側でも効率よくデータの管理ができるように、例えばメタデータの整備、識別子の付与といったことにも対応していく必要があるのではないかと思います。

データの収集については、論文については既に雑誌に公開していますが、論文がなかなか機関リポジトリに集まってこないという話もあり、これは図書館側の悩みでもあります。データについては、保存する場所がまだ整っていない状況なので、図書館の側もデータの収集から論文の収集へとアプローチしていくという方法もあるのではないかと考えています。