

Open Repositories 2012 参加報告

西薊由依 (鹿児島大学附属図書館)

大園隼彦 (岡山大学附属図書館)

Open Repositories は、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ある。ヨーロッパとアメリカを交互に開催地としながら2006年より毎年開催されており、7回目を迎えた Open Repositories 2012 (OR2012) は、2012年7月9日-13日にイギリスのエジンバラ大学で開催された。参加者リストによると、30ヶ国超から450人超の関係者が出席している。OR2012のコンセプトは、"Open Services for Open Content : Local In for Global Out"であり、オープンコンテンツ、拡張コンテンツ、分散型システム、情報発信基盤といった現在の動向を反映したものであった。

この度デジタルリポジトリ連合 (DRF) から、ポスター発表と、リポジトリの国際的動向の調査を目的に参加した。以下、ポスター発表、参加したワークショップ、テーマ別セッションの主なものについて報告する。

1. ポスター発表

今回 DRF^{(*)1}はイギリスの Repositories Support Project (RSP)^{(*)2}とともにポスター発表^{(*)3}を行った。

ポスター出展に至った経緯を簡単に紹介すると、2012年1月に、RSPプロジェクト・マネージャ (当時) の Jacqueline Wickham 氏を招いて行われた DRF RSP Meeting^{(*)4}に端を発する。この会議では、イギリス・日本両国におけるリポジトリ・マネージャ向け研修活動に関する報告・意見交換が行われたが、その際にお互いの目的、活動に多くの共通点を見だし、協力関係を深く確実なものにすることが両国の機関リポジトリ推進に非常に有効であるとの認識を持つこととなった。これを受けて、同年3月、DRF、RSP、The United Kingdom Council of Research Repositories (UKCoRR)^{(*)5}との三者間において国際連携に関する了解覚書 (Memorandum of Understanding (MoU)) が締結された^{(*)6}。この覚書は、機関リポジトリの推進のための人材養成が両国において重要性を増してきていることを踏まえ、機関リポジトリ構築に係る人材養成における知見と経験の共有、各組織が開催する機関リポジトリ構築をテーマとした会合やワークショップへの各組織代表者の招聘及び参加、ウェブサイト等における協力関係の報知、各国内及び国際的な助成受入への協力を通じた各組織の活動の継続性確保への助力を目的としたものである。この MoU 締結を受け、今後の関係構築の第一歩として、OR2012 へのポスター共同出展を行った。

ポスター発表は "Sharing experiences and expertise in the professional development of promoting OA and IRs between repository communities in Japan and the UK" と題し、DRF と RSP それぞれの活動内容や、DRF、RSP、UKCoRR の MoU 締結について紹介するとともに、日本とイギリス両国のリポジトリ・コミュニティが人材育成に関して相互協

力していくことが両国のオープンアクセス（OA）、リポジトリ促進に繋がるという主旨の内容で行った。

全体では各国から 60 超のポスター発表が寄せられ^(*)、研究データ、リポジトリと CRIS^(*)との連携を扱ったものが複数見受けられた。OR2012 総会に続いて行われた"Minute Madness"では、会議参加者の前で各ポスター発表者が 1 分間のプレゼンテーションを行い、その後ドリンクレセプション会場やセッション会場でポスター展示が行われ、参加者は自分が関心を持った発表について発表者と意見交換を行うことができるように運営されていた。

=====

Ref. (URL はすべて最終アクセス: 2012-09-27)

- *1 <http://drf.lib.hokudai.ac.jp/drf/>
- *2 <http://www.rsp.ac.uk/>
- *3 ポスター発表関連資料は以下で公開: <http://drf.lib.hokudai.ac.jp/drf/index.php?Papers%20and%20Talks>
- *4 DRF RSP Meeting については以下を参照: http://drf.lib.hokudai.ac.jp/drf/index.php?DRF_RSP_meeting, http://drf.lib.hokudai.ac.jp/drf/index.php?plugin=attach&refer=%E6%9C%88%E5%88%8ADRF&openfile=DRFmonthly_25.pdf
- *5 <http://ukcorr.org/>
- *6 MoU については以下を参照: http://drf.lib.hokudai.ac.jp/drf/index.php?Press%20release%20%2F%20DRF_RSP_UKCoRR_MoU_201203, http://drf.lib.hokudai.ac.jp/drf/index.php?plugin=attach&refer=%E6%9C%88%E5%88%8ADRF&openfile=DRFmonthly_27.pdf
- *7 全発表の一覧・内容は以下で公開: https://www.conftool.net/or2012/index.php?page=browseSessions&form_session=6
- *8 Current Research Information Systems

2. ワークショップ

会議初日と 2 日目には、総会に先立ち、リポジトリのソフトウェアやサービス、研究データ管理、デジタル保存などをテーマとした計 14 のワークショップが開催された。本章では、筆者らが参加した RSP と COAR のワークショップについて報告する。

2. 1. RSP: Building a national network

【概要】

(1)Baiviar Notay 氏による JISC の活動についての発表("The Role of JISC in supporting repository development")、(2) Jackie Wickham 氏による SHERPA、RSP の紹介 ("The

SHERPA Project"、"The Repositories Support Project")、(3)Marie Cairney 氏による事例発表 ("The University of Glasgow institutional repository: a case study"、(4)Gareth Johnson 氏による UKCoRR の紹介 ("UKCoRR: an independent organisation for UK repository staff") が行われ、英国のリポジトリ事情を中心とした内容であった。その後、参加者を 4 グループに分けリポジトリに関連する話題のディスカッションも行われた。

【内容紹介】

(1)現在イギリスのリポジトリの数は 200 を超えるまでに発展しており、2002 年以來の JISC^(*)の多数のプロジェクトがこの発展に寄与してきたことが紹介された。OAI-PMH を用いて学術機関が持つ様々なタイプの知的資産を検索できるようにした FAIR (Focus on Access to Institutional Resources) プログラムに始まり、Repositories and Preservation プログラムでは各高等教育機関がゼロからリポジトリを構築するのを促進。インフラの土台を築くと、その拡張 (REF 対応プラグインなど) と革新に取り組んできた。近年の取り組みの 1 つがメタデータの自動生成とテキストマイニングであり (例えば MERLIN プロジェクトでは、リポジトリコンテンツのテキストマイニングによるキーワード抽出と重み付けを行い、主題特化した検索インターフェースを提供している)、2012 年 3 月に発表された報告書"Value and benefits of text mining"の紹介がされた。

2002 年以來約 300 のプロジェクトが実行され、OA をサポートするインフラに約 2,000 万ポンドが投資されてきたが、これらは国の戦略・政策に沿って行われてきたものである (Houghton レポート (2009)^(*)、Department for Business, Innovation and Skills (DBIS) "Innovation and Research Strategy for Growth" (2011)^(*)、Universities UK "Efficiency and effectiveness in higher education" (2011)^(*)、Finch レポート (2012)^(*))。最近では効率性・効果性に重点が置かれ、リポジトリと学術成果管理のインフラを共有化するプログラムが進行中だが、この理由としては、効率的・効果的なインフラにより成果の確実な管理・共有・再利用を図る、大学がより効果的・効率的な学術コミュニケーション形成に確実に貢献しオープンな学問に取り組む、JISC の方針や共有インフラを国内外の関係機関と提携させ、より効率的・効果的ワークフローを可能にする、先行投資に見合う価値を供すること等が挙げられる。このプログラムの事例の 1 つが、UK RepositoryNet+プロジェクト (EDINA)^(*)で、プログラムが目指すインフラサービスの環境を整える製品を供するとともに、この製品を通してサービス内容の検討や市場調査を行うことも目的としており、製品のアイデアを生み出す場である Innovation Zone (UKOLN) と連携している。実現の必要がある機能として、どのコンテンツが登録可能なかの明示、ふさわしい場へのコンテンツ登録、登録済データの質向上 (知的所有権・著作権情報、メタデータ生成、識別子サービス、典拠ファイル)、登録済データの分析 (引用分析、統計の集約、アクティビティ・データ)、登録済データの保護等が挙げられ、これらに対応するサービスとして、SHERPA RoMEO、JULIET、IRUS (機関リポジトリ利用統計)、1 回の登録で複数のリポジトリへ登録可能にする Repository Junction Broker (RJB)、組織やリポジトリの同定ツ

ール Organisation and Repository Identification (ORI)、Open Access Repository Registry (OARR) が挙げられた。

Finch レポートはイギリス国内外で様々な議論を引き起こしたが、UK Open Access Implementation Group (UKOAIIG. 英国 OA 実行グループ) (*7)でも、この転換期にリポジトリがどのような役割を果たせるかについて検討予定であるとの紹介があった (*8)。

JISC は 2012 年 8 月より保証有限責任会社となる組織改革が行われたところだが、このように JISC 自身の展望が政治的・組織的にも複雑である上、各機関・JISC とともに財政事情は厳しく、また学術情報流通におけるビジネスモデルは予測が難しい、といった様々な課題に面していることが最後に述べられた。

(2)現在 Nottingham 大学の The Centre for Research Communications (CRC) (*9)が管理している SHERPA (*10)、RSP について、その概略や意義等が説明された。

SHERPA は、2003-2006 年に実行されたプロジェクトを発端とし、以来雑誌掲載論文のセルフアーカイブに関する出版者のポリシーを集約した RoMEO (*11)、リポジトリのレジストリ OpenDOAR (*12)、助成団体のポリシーを集約した JULIET (*13)など、その事業は国内外の OA 推進・サポートに大きな役割を果たしてきている。

CRC のもう 1 つのプロジェクト RSP は 2006 年のプロジェクト開始以来、リポジトリ・マネージャのネットワーク構築や研修開催等で、英国内の各リポジトリがその目的により合致し、標準化し、持続可能となるためのサポート活動を行っている。イギリスのリポジトリが抱える課題としては、メタデータのみでなく本文の登録率の向上、研究データのような論文以外のタイプのコンテンツへの対応、Research Excellence Framework (REF) に対応するためどのような研究が行われているかの報告、助成団体の要件を順守しているかどうかの報告、リポジトリと大学内他システムとの統合等が挙げられた。

なお、2011 年に行ったイギリスにおける全国調査の結果が以下のとおり紹介された：

- ・使用ソフトウェア：EPrints 58%、DSpace 27%
- ・登録主体者：代行 21%、部分的に代行 51%、場合による 24%、代行なし 4%
- ・保存ポリシーの制定：有 55%、無 45%
- ・本文登録率：本文有のみ登録 24%、メタデータのみ登録 76% (このうち本文登録率は 29.85%)
- ・機関の登録義務化：41%
- ・(学内の) リポジトリ業務の責任が図書館に有：86%
- ・研究管理システムとの統合：済 16%、実装中 30%、計画中 29%、予定なし 25%

(3)Glasgow 大学の機関リポジトリ "Enlighten" (*14)について、その計画から実際のサービス実施までの報告が行われた。

最初の JISC 助成プロジェクト "DAEDALUS" (2002-2005) では、大学は研究成果の OA 化だけでなく、学内集中型の出版物データベースを開発することも目指したが、2004 年の

リポジトリサービス開始時点では教員のボランティアな登録サービスだった。2006年にサービスを"Enlighten"と改名。2007年開始の Glasgow Theses Service では、数年かけて学位論文の登録義務化(卒業判定に影響)を果たした。そして2008年の Research Assessment Exercise (RAE. 研究成果の査定)で教員の出版物を規則的・継続的に集約する必要に迫られ、同年"Enlighten"へ出版物の書誌情報の登録が義務化された。

教員の負担減のためシンプルな運用に努めるほか、大学内の各種システムの核となる研究システムと統合(例えばプロジェクト終了時にはその関連出版物をリポジトリに登録するようシステムが研究者にリマインダを送る)するなど、大学内でのシステム間連携にリポジトリシステムも含まれている。

2009年からは著者名データの集約や助成情報とのリンク付け、スタッフプロフィールページの作成、現在は2014年の REF (RAEの後継)への対応準備を行うなど、改善を続けている。

10年かけて、"Enlighten"は大学全体のリポジトリかつ出版物の中心的データベースとなった。本文登録率の向上が課題の1つだが、近年のOA誌出版の流行がその解決に役立っている。

(4) UKCoRR は2006年に SHERPA Plus プログラムから派生したもので、イギリス国内のリポジトリ・スタッフのための独立組織である。国外のリポジトリ・コミュニティとの協力関係構築も進めており、日本の DRF とも2012年に MoU を結んでいる。

機関ではなく個人ベースの参加組織という点が特徴の一つで、現在リポジトリ業務に携わっている者のみが参加できる。参加費は無料で、現在は121機関から265人が参加している。参加費が無料のため資金がない。助成も受けていないため利害関係の対立は起こらないが、それゆえに活動内容は限られてくる(RSPの代替とは成り得ない)。活動内容は優良事例の情報交換や学術出版者へのロビー活動などで、満足度は保証する、なお最近の話題は Finch レポートとそれへの反響、リポジトリと CRIS との関係などであるとの紹介があった。

=====

Ref. (URLはすべて最終アクセス: 2012-09-27)

*1 <http://www.jisc.ac.uk/>

*2 <http://www.jisc.ac.uk/media/documents/publications/rptheconomicapublishing.pdf>

*3 <http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf>

*4 <http://www.universitiesuk.ac.uk/Publications/Documents/2011/EfficiencyinHigherEducation.pdf>

*5 <http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf>

- *6 <http://www.repositorynet.ac.uk/>
- *7 <http://open-access.org.uk/>
- *8 その後、7月25日に開催された会議の記録"OAIG July meeting summary" (8月6日付) が UKOAIG の Web サイトに掲載されている: <http://open-access.org.uk/2012/08/oaig-july-meeting-summary/>
- *9 <http://crc.nottingham.ac.uk/>
- *10 <http://www.sherpa.ac.uk/>
- *11 <http://www.sherpa.ac.uk/romeo/>
- *12 <http://www.openoar.org/>
- *13 <http://www.sherpa.ac.uk/juliet/index.php>
- *14 <http://eprints.gla.ac.uk/>

2. 2. COAR: Workshop: COAR Repository Interoperability Roadmap

【内容紹介】

OA の実現を支援するための国際連携組織「オープンアクセスリポジトリ連合」(Confederation of Open Access Repositories (COAR))^(*1)の Eloy Rodrigues 氏が"COAR Repository Interoperability Roadmap"と題した発表を行った。

リポジトリの数は増加を続け、最近では世界規模で e-Research インフラの主要な構成要素となってきたが、各リポジトリの価値を最大化するにはこれらのネットワーク化、すなわち相互運用性が肝要となる。COAR はワーキング・グループ 2 における 2010-2011 年の主要目的を「OA リポジトリ間の相互運用性、そして広範な e-Research インフラの一部としての相互運用性について議論を促進すること」と定め、段階を踏みながらロードマップの作成に取り組んできた。本ワークショップの目的は、相互運用性の展望、どの基準やプロトコルがどのサービスをサポートしているのか、各機関でどのように相互運用性の戦略を立てるべきか、といったことを参加者に理解させることで、同時にロードマップ最終版発表前の意見聴取も意図していた。なお、DRIVER プロジェクト^(*2)との関わりについて、COAR のガイドラインは DRIVER ガイドラインを参照はしているがそれに置き換わるものではない、との説明があった。

ロードマップは 2 部構成で、第 1 部では OA リポジトリの相互運用性の状況 (2012 年 7 月末に発表予定との説明)、第 2 部では将来の方向性 (2012 年末に発表予定との説明) を扱っており、本ワークショップでは前者の第 1 部について内容の説明がされた。前述のように OA リポジトリは世界規模で e-Research インフラの主要な構成要素となってきたが、各サービスで使用されている様々な規格類をその目的により整理すると例えば以下のようなになる。

- ・メタデータ・ハーベスティング : OAI-PMH

- ・リポジトリのネットワーク：DRIVER、OpenAIRE
- ・利用統計：COUNTER、KE Usage Statistics Guidelines、OA-Statistik、PIRUS、SURE
- ・自動登録：SWORD、OA-RJ (Open Access Repository Junction)
- ・著者同定：Author Claim、ORCID、DAI (Digital Author Identification)
- ・永続識別子：PersID、DOI (Digital Object Identifier)、ハンドルシステム、DataCite
- ・複合オブジェクト管理：OAI-ORE (Open Archives Initiative Object Reuse and Exchange)
- ・研究管理：CRIS-OAR^(*3)

ロードマップ第1部ではこれらの規格類について1つずつその名称やURL、分類、対象地域、現在の更新状況、目的、実装例等を説明する(11月のベルリン会議^(*4)でもプレゼンテーション予定)。

2013年には、第1部の改訂版とデータ・リポジトリとの相互運用性も含めたロードマップ2.0を発表予定とのことである。

=====

Ref. (URLはすべて最終アクセス: 2012-09-27)

*1 <http://www.coar-repositories.org/>

*2 <http://www.driver-repository.eu/>

*3 CRIS: Current Research Information Systems, OAR: Open Access Repositories

*4 <http://www.berlin10.org/>

3. テーマ別セッション

会議では多数のテーマのセッションが開催されたが、ここでは関心の高かった共同リポジトリ・共同インフラをテーマにしたものと、研究データ管理をテーマにしたものを中心に報告する。

3. 1. 共同リポジトリ・共同インフラ

リポジトリサービスやインフラの共同構築については、このテーマを冠したセッション^(*1)が2コマ開催されたほか、ショートプレゼンテーションの集まりである **Repository Fringe**^(*2)でもこのテーマの発表が行われるなど、世界的な関心度の高さが窺えた。

日本でも共同リポジトリによるリポジトリ構築が特に中小規模機関で拡まっており、同じく関心度は高い。本項では、会議で筆者が拝聴した中からいくつか関連の発表を内容紹介し、その後、国内の状況について少し述べたい。

(1)"The Development of a Socio-technical infrastructure to support Open Access Publishing through Institutional Repositories" / Andrew David Dorward, Peter Burnhill,

Terry Sloan (EDINA, United Kingdom) [S]

JISC 助成プロジェクト"UK RepositoryNet+ (RepNet) "について紹介された^{(*)3}。これは、学術出版の新しい OA 出版モデルをサポートするために、機関リポジトリをベースにしたインフラを構築することを目的としており、持続的かつよく使われる一連のサービスを提供することでリポジトリのコスト効率性をより高めることができると強調されていた。具体的には、出版料は著者か研究責任者 (PI) により研究助成金 (官民間問わず) から支払われ (=透明性の担保)、助成者の公開義務化方針により、論文ファイルが DOI と助成情報のメタデータを付されて誰もが利用できるリポジトリ (助成者・主題・機関) にアップロードされる、というものである。

この RepNet の背景には、OA を促進する政府の方針 (公的資金を受けた研究の OA 化。ゴールド OA 推進を提言した Finch レポート)、資金提供者の公開義務化方針、出版者の実践 (グリーン OA が STM 出版者にどのような影響を与えるか試算した PEER プロジェクト^{(*)4})。グリーン OA は出版者サイトでのダウンロードを促進するが、ビジネスモデルとして好ましいのはゴールド OA) が挙げられる。そのスコープには、ゴールド OA のコスト負担の仕組み (APC を雑誌の自由市場下に置くには透明でなければならない)、研究データセット (NERC^{(*)5}や ESPRC^{(*)6}が登録義務化。将来データセット・レジストリを作る第一歩に)、OpenAIRE^{(*)7}との連携 (まず MoU を締結。DRIVER との統合やゴールド OA、データセット・レジストリとの連携にも関心) が入っている。

本会議中にサービスの第 1 段階 (Wave 1) の構成要素が公式披露されたが、それらの最適化、続いて第 2 段階 (Wave 2) の構成要素の詰めが今後の焦点となる。2013 年 3 月に JISC のサービスポートフォリオレビューを受けることになっている。

(2)"Repository communities in OpenAIRE: Experiences in building up an Open Access Infrastructure for European research" / Najla Rettberg, Birgit Schmidt (State and National Library, University of Gottingen, Germany) [S]

EU が助成した研究成果を OA で提供するリポジトリ連携プロジェクト OpenAIRE について、そのビジョンや概要の紹介があった。ヨーロッパでは、欧州委員会 (EC) が、オープンサイエンスのためのオープンインフラストラクチャの構築を政策として進め、助成プログラム"Horizon 2020" (2014-2020 年実施) や EU 加盟国の助成を受けた研究成果についても OA 化の方針を発表したところである。

OpenAIRE は 2009 年に 27 カ国の参加で始まったプロジェクトで、EC の助成を受けた査読済論文の OA 義務化 (Open Access Pilot の OA に関する特別条項 39 (SC39))、第 7 次研究フレームプログラム (FP7) により助成を受けた研究成果の OA 化促進といった EC の方針を背景に、研究者がこれらの方針に従うためのインフラを構築するとともに、それらの研究成果を公開しているリポジトリを連携させアクセス性を高めた。活動を通してヨーロッパ内の人々が経験やベストプラクティスを共有するネットワークの形成も促進され

た。約 5,000 のプロジェクトから生産された約 27,000 の FP7 出版物が登録されており、うち約 10,000 が OA で（その他はエンバーゴ等でクローズド）、ソースはリポジトリのほか雑誌からのハーベスト、文献データベース等がある。

2011 年 12 月よりフェイズ 2 として OpenAIREplus が 33 カ国の参加で始まった。これは OpenAIRE の上に構築され、出版物からデータセットや FP7 以外の助成プログラムへのリンクを形成することで、ヨーロッパにおける OA の研究生産物全体へのアクセスを促進するプロジェクトであり、OA のデータセットをハーベスト・改善・保管する e インフラが構築されることとなる。データリポジトリ向けのガイドラインは既存の DRIVER と OpenAIRE のガイドラインの拡張で対応する。

(3)"Enhancing repositories and their value: RCAAP repository services" / Clara Parente Boavida, Eloy Rodrigues, Jose Carvalho (University of Minho, Portugal) [S]

RCAAP^(*)は、FCCN^(*)が全体調整を、Minho 大学が技術面の調整をそれぞれ担当する態勢で 2008 年に始まったプロジェクトである。ポルトガルの研究成果の可視性・アクセシビリティ・普及度の向上、成果物情報へのアクセス改善、OA に関する国際的イニシアティブへのポルトガルの参加等を目的とし、年々その提供サービスを拡大してきている。

2008 年当初より提供しているのが、リポジトリホスティングサービス (SARI) と検索ポータルサイトである。前者は DSpace1.8.2 を基にした無料のサービスで、各ユーザごとにインターフェースが異なり、機関ごとに URL も異なる。現在 26 リポジトリをホスティングしている。後者では、現在 42 のリソース（機関リポジトリ 35、雑誌 5、データリポジトリ 1、ブラジルのポータル 1）からハーベストされた約 45 万のコンテンツをまとめて検索でき、機能面での特徴としては、コンテンツの全文検索、文献管理ツールや研究者情報サイトとの連携、RSS 提供、等が挙げられる。

2009 年より開始したのが RCAAP Validator と共同リポジトリのサービスである。前者は、DRIVER のガイドラインに基づいたもので、リポジトリの OAI-PMH 情報を入力すると統計やエラー情報のレポートが提供される。後者では、1 つのリポジトリに現在 13 機関がまとめられている。2010 年よりデータリポジトリ、ブラジルとの連携を開始した。

2011 年より雑誌ホスティングサービス (SARC)、利用統計ツール (SCEUR) の提供を開始した。前者では、各雑誌は自律的に運用され、個々の URL を付与されるほか、SWORD を用いた機関リポジトリとの連携も行われる。後者は RCAAP が多数提供している DSpace のアドオンの一つで、リポジトリ利用統計の中央集約サービスとして用いられており、統計のグラフィック表示が可能である。OAI-PMH データプロバイダでもある。なお RCAAP が提供している DSpace アドオンはこのほか、OAI 拡張アドオン（ヴァーチャル・セット作成）、共有バーアドオン（ソーシャルネットワークや文献管理ツールとの共有、研究者情報サイトへの送信）、DeGóis アドオン（研究者情報サイトとの連携）等がある。

最後に、ポルトガルのリポジトリ現況について、リポジトリ数は 30 超（2008 年の RCAAP

開始後増加)、コンテンツ数は 10 万弱、登録義務化した機関数は計 14 との報告があり、その発展の要因は、相互運用性、コミュニティの存在、研究システムへの統合であるとの分析が示された。

(4)"Shared and not shared: Providing repository services on a national level" / Jyrki Ilva (National Library of Finland, Finland) [S]

フィンランドでは現在 48 機関が機関リポジトリを有するが、うち 36 機関は国立図書館が運用するインフラを用いたリポジトリ (DSpace) で、比較的大規模な 6 つの研究大学は自機関リポジトリ (DSpace/Fedora) をローカル構築している。フルテキストコンテンツ数は、国のインフラを利用したリポジトリが計 62,900、その他のリポジトリが計 57,700 とほぼ同程度である。

当初機関リポジトリは、各機関でローカル構築され、検索エンジンに OAI-PMH ハーベストされるという見通しだったが、リポジトリによっては、資金に乏しい、ほとんどコンテンツがない、という例も見られる。共同サービスやホスティングサービスの利用は、コストの節約や、より資金をかけたサービスといった重要な利点を得られる可能性がある。フィンランドでは他国より共同化が進んでいるが、この理由としては大学図書館が伝統的に共同サービスを構築してきたこと、大学の資金源である官庁が中央集約型の国のインフラ構築を強く支援してきたことが挙げられた。リポジトリは国立図書館が提供する中央集約型サービスの一つだが、国からは少額の一時的なプロジェクト資金が付いたのみで恒久的な資金提供はないため、現在は参加機関がその費用を拠出している。

フィンランドのリポジトリは、1990 年代後半から 2000 年代前半における学位論文や雑誌の出版を発端とし、リポジトリの概念はその数年後に導入された、という経緯がある。国が 2007 年に始めた Doria^(*)10)は、全機関の全コンテンツを一つの大きな DSpace リポジトリに収めようというものだったが、当初約 10 機関の参加しか得られず、多くの大学は既にローカル構築していたリポジトリをアップグレードすることを選択し、共同サービスやホスティングサービスの利点はリポジトリ・マネージャにはそれほど受け入れられなかった。その後、応用科学分野でのサービスプロバイダに選ばれたことや法律が制定されたことを背景に国立図書館のリポジトリサービスの基盤化が進んだ。

現在のサービスモデルでは、国立図書館は技術面でのメンテナンスに専従しており、参加機関は、複数機関によるリポジトリ (Theseus、Doria) への参加か、独自の DSpace リポジトリ構築 (ホスティングは国立図書館) かを選択できる。デジタル出版物の蓄積と普及のための手頃なシステムへのニーズは高く、リポジトリサービスの長期的アクセスや永続的地址、各機関の他システムとの統合は大きなメリットだが、さらにリポジトリのプラットフォームは多様なデジタルコンテンツ管理に適しており、グリーン OA にとどまらない多様な利用例が存在する、と今後の展望が示された。

フルテキストコンテンツをタイプ別に見ると、セルフアーカイヴは 4 機関で登録義務化

されているにも関わらずその登録コンテンツ数は少ない。圧倒的に比率が高いのが学位論文で、この OA 出版がフィンランドの成功でありこの国の学術出版の在り方を変えてきたことは明らかである、と締め括られた。

(5)"The Library Consortium of New Zealand's Shared IRR Infrastructure" / Allison Brown (The University of Otago, New Zealand), Andrew Lockett (Unitec Institute of Technology, New Zealand), Craig Murdoch (AUT University, New Zealand), Kate Nixon, Andrea Schweer (The University of Waikato, New Zealand) [R]

The Library Consortium of New Zealand ^(*11)が進めてきた"Institutional Research Repository"プロジェクトの紹介が行われた。コンソーシアム構成員のうち Victoria 大学は自機関リポジトリを持つが、AUT、Otago、Waikato の 3 大学と Unitec の 1 機関がこのプロジェクトに参加しており、Waikato 大学の情報技術サービスが各機関リポジトリのホスティングを行うとともに、専任の技術専門職を置き、プロジェクト管理も行っている。各リポジトリの運用方針や優先順位はそれぞれ異なっているが、このコミュニティで知識やスキル、リスク、改善、コスト等を共有し、自機関構築よりも共同構築の方がカスタマイズやアップグレードも効果的に行える等のメリットが説明された。

【国内状況】

日本国内におけるリポジトリサービスやインフラの共同構築の現状を考えたとき、一つに共同リポジトリの構築が進んでいることが挙げられるだろう。共同リポジトリの形態の一つが、主に県域単位で複数機関が集まりリポジトリを構築する地域共同リポジトリである。2007 年 7 月の山形 ^(*12)を端緒に、2008 年に広島 ^(*13)、埼玉 ^(*14)、岡山 ^(*15)、2009 年に、新潟 ^(*16)、福井 ^(*17)、山口 ^(*18)、2010 年に沖縄 ^(*19)、2011 年に愛媛 ^(*20)、2012 年に鹿児島 ^(*21)、長野 ^(*22) (後述の JAIRO Cloud 利用)、群馬 ^(*23)で、それぞれ地域共同リポジトリが構築され、試験公開中の青森 ^(*24)を含めると、2012 年 9 月現在で 13 の地域共同リポジトリが運用され、150 近くの機関等が参加している。この地域共同リポジトリに続いて、中央型の共同リポジトリも生まれている。国立情報学研究所 (NII) は平成 23(2011)年度から JAIRO Cloud 事業を開始した ^(*25)。これは、個別の機関リポジトリ構築が困難な大学等を対象に、NII が共用リポジトリシステムを提供することにより機関リポジトリの構築をサポートするというものである。平成 27(2015)年度までに 200 機関の新規構築を目指しており、この数字は国内の博士後期課程を持つ大学のすべてにリポジトリが構築されることを想定している。2012 年 4 月より正式サービスを開始し、2012 年 9 月 27 日現在で 27 の機関リポジトリが運用中である ^(*26)。文部科学省が 2012 年 7 月に発表した「学術情報の国際発信・流通力強化に向けた基盤整備の充実について」 ^(*27)では、「機関リポジトリの活用による情報発信機能の強化について」と題された第 4 章で、機関リポジトリの役割・意義、現状、今後の課題が述べられている。この中で、大学等による個別の機関リポジトリ構築に加え、地域等における機関間連携による共同リポジトリの整備が進んでいることを

評価するとともに、NIIによる共用リポジトリサービス提供について、今後機関リポジトリのより一層の整備・普及につながる、とその役割に大きな期待を寄せている。

OR2012における発表でも、共同リポジトリについて、コミュニティの形成やその中での知識やスキル等の共有、コスト削減、効果的なシステム改善等、国内同様そのメリットが報告されていたが、同時に興味を引くのが、フィンランドにおけるコスト負担の事例である。フィンランドでは、リポジトリを自機関構築している大学もあるが、国立図書館が中央集約型サービスの一つとしてリポジトリサービスを提供している。国のインフラとして提供されているものだが、国からの資金提供は当初の一時的なもので恒久的なものではなく、現在は参加機関がその費用を拠出している。国内の共同リポジトリの多くは、現在コスト面で持続可能な運用モデルに基づいているとは言い難い。共同リポジトリプロジェクト Share（平成20-21年度）が2010年3月に発表した「共同リポジトリプロジェクト報告書：国内の地域共同リポジトリの分析」^(*28)では、第6章を「共同リポジトリのコスト分析」と題し、共同リポジトリの多くがその運営経費をCSI委託事業経費やホスト機関の学内経費でまかなっていることが多いこと、CSI委託事業終了後の持続可能な運用モデルの構築が必要であることを指摘している。一方、NIIによるJAIRO Cloudでは、その利用料金について当面は無料とされているが、将来的には有料化する可能性も示唆している。そのコストモデルの分析は、持続可能な運用体制を構築するために各参加機関に相応の負担が求められる時点まで待つ必要があるが、日本における共同リポジトリに関するコストモデルの構築を検討する際、フィンランドの事例は一つの参考材料となるかもしれない。

また、OR2012における発表では、リポジトリと、OA出版、助成情報、文献データベース、研究者情報等との連携の取り組みが紹介された。これらは、ヨーロッパにおけるOA化促進政策、資金提供者による公開義務化方針等を背景に、より広範な研究成果物へのアクセスを促進する共同のインフラ構築を目指したものであり、かつ持続可能性、コスト効率性が重視されたものである。日本では政策面での後押しに強制力がないという違いはあるが、より広範な研究成果物へのアクセスを促進するという目的に沿った様々な取り組みには参考になることもあるだろう。共同リポジトリ（共用リポジトリ）がリポジトリ構築機関数の増加という量の成果を上げることは間違いないが、OAを推進していくには、例えば、リポジトリと既存のサービスとのさらなる連携で研究成果関連情報をスムーズに繋げることで、サービスやコンテンツの価値を向上させるといった質の成果も求められる。

国内の学術情報基盤や研究環境の整備に関しては、NIIや科学技術振興機構（JST）、国立国会図書館（NDL）、日本学術振興会等複数の機関が関係しており、様々なサービス・事業を展開している：NIIによる、機関リポジトリ掲載学術情報のポータルサイトJAIRO^(*29)、国内の各機関リポジトリの詳細情報、コンテンツ内容の統計分析情報を提供するIRDB（学術機関リポジトリデータベース）コンテンツ分析システム^(*30)、国内の学術論文情報データベースCiNii^(*31)、博士論文書誌データベースを含む様々な専門分野のデータベースを集めたNII-DBR^(*32)、科学研究費助成事業により行われた研究の採択課題、研究成果を収録

した、科学研究費助成事業データベース KAKEN (*33)、JST による、研究者、文献、特許などの情報をつなぐ科学技術総合リンクセンターJ-GLOBAL (*34)、研究者向け研究基盤サービス ReaD&Researchmap (*35)、NDL による博士論文デジタル化 (*36)等。リポジトリとの連携が既に行われているものもあるが、連携が一層促進されることで、研究成果関連情報のアクセシビリティやファインダビリティの向上、評価分析への活用、インフラとしてのリポジトリの価値向上といったことが期待できるだろう。

なお国内に限らず様々なサービスの連携のためには相互運用性が必要であり、識別子を介した連携もその一つの実現例である。そして、世界で用いられている様々な規格類を整理し、相互運用性についての展望を立てる取り組みが、前述の COAR によるロードマップ作成であることは説明するまでもなく理解頂けるだろう。

=====

Ref. (URL はすべて最終アクセス: 2012-09-27)

- *1 "Shared Repository Services and Infrastructure". 後述の紹介では発表タイトルの末尾に[S]を付している。
- *2 "Repository Fringe : Pecha Kucha - National Infrastructure". 後述の紹介では発表タイトルの末尾に[R]を付している。
- *3 RepNet については本会議中複数の機会でも取り上げられ、そのうちワークショップでの議論の様相については報告書が以下のページで公開されている: http://www.repositorynet.ac.uk/wp-content/uploads/2012/08/RepNet_WS_OR2012_release.pdf
- *4 <http://www.peerproject.eu/>
- *5 Natural Environment Research Council <http://www.nerc.ac.uk/>
- *6 Engineering and Physical Sciences Research Council <http://www.epsrc.ac.uk/>
- *7 Open Access Infrastructure for Research in Europe <http://www.openaire.eu/>
- *8 Repositório Científico de Acesso Aberto de Portugal <http://www.rcaap.pt/>
- *9 Foundation for National Scientific Computing <http://www.fcn.pt/>
- *10 <http://www.doria.fi/>
- *11 <http://www.lconz.ac.nz/>
- *12 ゆうキャンパスリポジトリ <http://repo.lib.yamagata-u.ac.jp/>
- *13 広島県大学共同リポジトリ : HARP <http://harp.lib.hiroshima-u.ac.jp/>
- *14 埼玉県地域共同リポジトリ : SUCRA <http://sucra.saitama-u.ac.jp/>
- *15 岡山共同リポジトリ : O-AIR <http://www.lib.okayama-u.ac.jp/o-air/>
- *16 新潟県地域共同リポジトリ : NiRR <http://nirr.lib.niigata-u.ac.jp/>
- *17 福井県地域共同リポジトリ : CRFukui <http://crf.flib.u-fukui.ac.jp/dspace/>
- *18 山口県大学共同リポジトリ : 維新 <http://ypir.lib.yamaguchi-u.ac.jp/>
- *19 沖縄地域学リポジトリ <http://okinawa-repo.lib.u-ryukyu.ac.jp/>
- *20 愛媛地区共同リポジトリ : IYOKAN <http://iyokan.lib.ehime-u.ac.jp/dspace/>

- *21 鹿児島県学術共同リポジトリ : KARN <http://karn.lib.kagoshima-u.ac.jp/>
- *22 信州共同リポジトリ <https://shinshu.repo.nii.ac.jp/>
- *23 群馬県地域共同リポジトリ : AKAGI <https://gair.media.gunma-u.ac.jp/>
- *24 ひろさき地域共同リポジトリ <http://hrr.ul.hirosaki-u.ac.jp/dspace/>
- *25 <http://www.nii.ac.jp/irp/repo/>
- *26 <https://community.repo.nii.ac.jp/link/>
- *27 http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/1323857.htm
- *28 <http://www.lib.hiroshima-u.ac.jp/share/seika/ShareReport.pdf>
- *29 <http://jairo.nii.ac.jp/>
- *30 <http://irdb.nii.ac.jp/analysis/index.php>
- *31 <http://ci.nii.ac.jp/>
- *32 <http://dbr.nii.ac.jp/>
- *33 <http://kaken.nii.ac.jp/>
- *34 <http://jglobal.jst.go.jp/>
- *35 <http://researchmap.jp/>
- *36 <http://dl.ndl.go.jp/#etd>

3. 2. 研究データ管理

研究データ管理をテーマとしたセッションも 2 度組まれており、またポスター発表でもデータ管理に関するものが多く、現在のトレンドであるように感じた。

研究データ管理はデータ保存に関すること、データ管理に関すること、データ再利用に関することの 3 つに大別される。特徴的なものとして、保存については分散ファイルシステムでセキュリティ強化、管理については DOI の付与や相互利用可能なメタデータのスキーマを整備、再利用については統合システムを整備、研究データと研究成果の相互運用性の実現等が挙げられる。以下、主な発表の概要を紹介する。

(1) Institutional Infrastructure for Research Data Management / Anthony Errol Beitz, Paul Bonnington, Steve Androulakis, Simon Yu, David Saint(Monash University, Australia)

【内容紹介】

研究の検証や再生産等を目的として、研究データ管理のためのプラットフォームの構築は、研究機関にとってますます重要になってきている。プラットフォームの機能としては、

- ・データの自動保存、キュレーション
- ・収集データの組織化
- ・データ再利用のライセンスの整備
- ・データの発見性の促進、再利用、長期保存

- ・データの引用の実現

等が挙げられ、様々な分野の研究者の環境（ツール、ワークフロー、方法論）に合致する必要がある。

Monash 大学では、オーストラリアの研究者が研究データを再利用できるようデータの管理化、可視化、再利用促進等を進めている ANDS (Australian National Data Service *1) による研究データ管理のサポートを活用し、タンパク質結晶学、歴史学、生態系研究、気象学等の様々な分野で研究データ管理の基盤を開発し、その経験から以下のガイドラインを作成している。

- ・研究データプラットフォームの選択

研究者は組織よりも研究コミュニティへの従属意識が強く、研究コミュニティの研究データ管理プラットフォームをより利用する傾向がある。研究コミュニティが既に優れたプラットフォームを持っており研究者のニーズに合致しているならば、組織はそのプラットフォームの採用をサポートするべきであり、そのためには必要に応じてローカル環境用のカスタマイズを施すこともあり得る。

- ・研究データ管理プラットフォームの開発

開発コスト、大きいリスク、ソフトウェア維持の問題を考慮すると新しいプラットフォームの開発は最終的な手段とすべきである。しかし特定要件の対応を考える必要がある場合は、独自仕様のプラットフォームの開発もあり得る。研究者指向のプラットフォーム開発の原則と方法は、

- ・研究者も開発に関わり、研究コミュニティがプラットフォームの管理をする。
- ・アジャイルソフトウェア開発の手法を採用する。
- ・時間を節約しリスクを軽減するため、できる限り既存の研究データ管理プラットフォームのカスタマイズをする。
- ・他機関でも簡単に採用できるソフトウェアにする。
- ・結局は誰も利用することのない、汎用的なシステムを開発しない。
- ・特定の分野の要件に合致するプラットフォームを選び、研究分野における共通点・相違点を見つけることでソフトウェアの改善をする。
等が挙げられる。

- ・研究データ管理プラットフォームの展開

研究データ管理プラットフォームの一部は、組織レベル、州レベル、国レベルのクラウド環境で提供されているが、組織は研究者のデータ保存の場所として適しており、その寿命も州・国レベルのものよりも長いので、組織レベルでの展開が望ましい。

データを公開し、再利用するにはメタデータを充実（研究者、寄付団体、出版物に関する情報）させることが望ましい。研究データ管理プラットフォームは以下の仕組みを保持するようになる。

- ・データコレクションと研究管理データとの関連付け。

- ・研究データコレクションと永続 ID の関連付け。
- ・同期、ハーベストによるデータの公開。

研究データ管理に取り組む際には、研究管理システムと研究データ管理プラットフォームの連携を考える必要がある。

(2) Big Data Challenges in Repository Development / Leslie Johnston(Library of Congress, United States of America)

【内容紹介】

ビッグデータはすでに存在しており、それは常に変化している。LC がデジタルコレクションを構築・管理した 15 年間で得た最大の教訓は、コレクションの利用方法は決して推測することができないということである。多くの研究者は新しい方法で情報をマイニングし、組織化することでコレクション全体を利用することを望んでおり、情報を知識へと変換するツールやアルゴリズムをますます利用している。大規模データコレクションと研究活動をサポートする LC の 4 つの実例を紹介する。

①Web アーカイブ

2000 年に開始し、360TB の容量がある。Web の閲覧のために開始したが、現在は分析のために利用される。Web アーカイブの管理・組織化とこれらのアクセス可能性、分析可能性の提供が重要となる。

②新聞アーカイブ

Chronicling American Collection^{(*)2} には 500 万ページもの OCR イメージがある。1 日当たり 400 万回の閲覧があり、データ分析の依頼も多い。これらのデータは API 経由で利用することが可能となっている。

③Twitter

Twitter のアーカイブは約 210 億ファイルがあり、ニュースの地理的な展開や新しい言語の研究等多くの研究の依頼がある。Twitter の発展の規模は早く、インフラの規模や研究者への新たなサービスの提供等考えるべき課題がある。

④Viewshare^{(*)3}

Viewshare はデータをアップロードして共有する Web ツール。Viewshare でデジタルコレクションを分析・可視化し、そのデータや結果、視点を共有することができ、リポジトリの分析ツールとしての利用可能性が考えられる。

ビッグデータを扱うには、データ取り込みプロセスの規模の検討、事前に利用方法を考える、フルテキスト索引のインフラの規模、分析ツールの提供等、考慮すべきことがある。

(3) Moving from a scientific data collection system to an open data repository / Michael David Wilson, Tom Griffin, Brian Matthews, Alistair Mills, Sri Nagella, Arif Shaon, Erica Yang (STFC, United Kingdom)

【概要】

多くのオープンリポジトリの運営者がファンドの要件を満たすために研究データの保存をそのサービスに組み込む中、研究データ管理システムからオープンリポジトリへと移行する逆のプロセスを紹介。この方法から、正確なデータの保存、データ引用のサポート、成果物へのリンク、データ発見サービスのサポート、セキュリティポリシーのサポート、データ再利用のためのソフトウェアやデータについての詳細情報の保存、リポジトリへの信頼性の向上等、研究データ管理における様々な教訓を得ている。

【内容紹介】

STFC (The Science and Technology Facilities Council *4) は多くの研究施設を持ち、多くの研究データを生産している。例えば、施設の一つである、ISIS (*5)では 25 年間で 800 万ファイル、25 万データセット、8TB のデータを生産してきた。STFC では、これらのデータコレクションをオープンデータリポジトリへと移行する取組みを行っている。

データコレクションサービスからオープンリポジトリへの移行の経験は、将来における発見性やアクセス性、データの再利用と解釈をサポートする仕組みの必要性を示している。データを保存する主な目的は以下のとおり。

1. 研究者が自分のデータにアクセス。
2. 他の研究者が公開した結果を検証。
3. データを再利用し、新たなトレンドを発見するメタ研究 (meta-study)。
4. 新しい理論やモデルを実験に適用。
5. 再実験することなく新しい研究を行う。

実験データを自動保存し、研究課程で作成される様々なデータ (生データ、検証データ、分析データ等) を検索可能にしており、研究とデータについて高レベルの情報を保存する CSMD (Core Scientific Metadata Model *6) をメタデータの記述に採用することで研究成果に実験データや分析データを追加し、リンク形成している。

【システム構成】

- ①実験機器から実験データを収集することでワークフローが開始。
- ②データは定期的に分散ファイルシステムに保存。保存したデータには BL (British Library) DOI Server から DOI が付与され、データの発見とアクセスサービスを提供している DataCite (*7)で検索ができるようになっている。
- ③研究データのメタデータは ICAT Sever に収集して保存し、ユーザはそのインターフェースである TopCAT Server から利用する。
- ④STFC 保存システムにデータファイルを保存する。SDB (Tessella Safety Deposit Box *8) ベースの STFC 保存システムは研究データの不変性チェックを行い、その情報を STFC e-Science data archives に情報を登録する。

システム構成の際には様々なユーザグループを考慮し、

- ・複数のセキュリティとデータポリシーをサポート
 - ・結果の検証をサポートする分析レコード履歴の提供
 - ・データの詳細情報とデータの再利用のための分析ソフトウェアの情報を提供
- 等を考慮する必要がある。

(4) Towards a Scalable Long-term Preservation Repository for Scientific Research Datasets / Arif Shaon, Simon Lambert, Erica Yang, Catherine Jones, Brian Matthews, Tom Griffin (STFC, United Kingdom)

【概要】

STFC は ISIS のような様々な施設から生成された大量の科学データを保存、展開する義務がある。大規模科学データは長期的な再利用を保証する拡張可能な保存方法を必要としている。STFC は多くの保存に関する EU のプロジェクト、特に Apache Hadoop^(*9)を利用したクラウドベースの保存方法を進める SCAPE プロジェクト^(*10)に参加している。このプロジェクトにおける STFC の目的は、既存の STFC データアーカイブを拡張可能な長期保存基盤へ改良することである。

【内容紹介】

(3)のデータの再利用を主な目的とした既存システムに加えて、新たな拡張可能な保存リポジトリの開発計画があり、将来的に SCAPE システムと現行のシステムの統合を検討している。新システムでは STFC 科学データの保存ワークフローを進展させることを目的としている。新システムは以下のサービスで構成される。

・ Staging Service

研究に関連するすべてのデータセットとその情報を収集する「中継サービス」として機能する。保存ワークフローの開始となるサービスで、ICAT のメタデータ、分散ファイルシステムの実験データ、ISIS のような関連施設の Web サイトの情報、データ整理や分析過程に関わるソフトウェアからコピー等を収集している。コレクションレベルのメタデータの記述については STFC 保存メタデータポリシーによって決定。

・ Pre-Ingest Service

研究レベル、コレクションレベルでメタデータの整合性の検証 (XML、チェックサム、メッセージダイジェスト等) を行い、実験データセットの提出用情報パッケージ (SIP : Submission Information Package) を生成し、SCAPE Ingest サービスに渡すことを目的とする。ワークフローは大規模データの分散処理技術であるハドゥープを基にした STFC Hadoop-based Workflow Execution Platform により実行される。

・ Ingest Service

Pre-Ingest サービスから実験データセットの SIP を受け取り不変性の検証（チェックサムの計算、メッセージダイジェスト、デジタルサイン、Pre-Ingest のデータとのマッピング）を行う。認証された SIP は Action サービスに渡される。

Pre-Ingest と共に、STFC Workflow Execution Platform でワークフローが実行される。

・ Action service

SIP であらかじめ定義されている、保存に関するアクションを実行することを主な目的としている。アクションは、Preservation Network Model (PNM ^{*11}) にまとめられた保存目的（データだけではなく分析に利用したソフトウェアも保存等）と依存関係の分析により定義されている。

保存アクションは事前に発行されたワークフローに統合される。ワークフロー及びアクションは「Policy Store」の保存ルールとポリシーを考慮する必要があり、最終成果の OAIS 準拠保管情報パッケージ (Archival Information Package : AIP) は STFC e-Science Preservation Archive に保存される。

今後は PNM の分析によるリンクデータセットの充実した文脈や関連表現情報の保存を計画しており、最終的には保存基盤計画の適応性の理解を進めるモデルを開発する。

(4) DTC Archive: using data repositories to fight against diffuse pollution / Mark Hedges (King's College London), Richard Gartner, Mike Haft, Hardy Schwamm (Freshwater Biological Association, United Kingdom)

【概要】

DEFRA^(*12)による貯水池実証テスト (Demonstration Test Catchment : DTC) は長期間イギリスの貯水池の水質調査を行っている。プロジェクトは長期にわたる水質の拡散汚染の解決を目的とおり、汚染の原因を理解するために詳細な環境要因のモニタリングを行っている。

プロジェクトの主な構成要素は、DTC Archive プロジェクトであり、貯水池で生成・保存されたデータを展開し、様々なステークホルダーへ適切な再利用サービス提供する義務を負っている。対処すべき主な課題は、センサーから取得される様々なデータやメタデータ及び専門用語の多様性と複雑さであり、プロジェクトでは SKOS モデルを定義し、データ記述の正式な語彙を開発している。データセットが保存されると、SPARQL 等を元にした可視化とデータセット間の意味的な検索を可能とする、RDF モデルにマッピングされる。

【内容紹介】

一般的データモデルは地理情報の観測・測定のためのスキーマを定義する ISO19156^(*13) を基にして作成している。基本情報は、「実世界の調査」、「現象の測定」、「時空間情報」、「結果」、「プロセス情報」等で、これに従うことで、様々なコミュニティ間で調査方法や結果の交換・共有を進めている。

ISO19156 は外部インターフェースを提供しており、同じ基準に従う他システムとの相互運用性を進めている。ISO19156 の下層の実装は自由であり、図書館やアーカイブズの基準に準じた長期保存とデータの再利用を進める表現を採用している。

保存表現 (archive representation) では、各データファイル (CSV のような形式) はメタデータと一緒にデジタルオブジェクトとして METS ドキュメントにパッケージされ、Fedora/Islandora ベースのリポジトリに保存される。各オブジェクトの記述メタデータは一般的な MODS (Metadata Object Description Schema) を利用し、様々な分野の特殊なメタデータ要件に対応する簡易な方法として活用している。

データを再利用する (異なるデータセット間で検索や分析をする) には、個々の列・セルレベルでオブジェクトを記述する必要がある。このためには各データ要素が同じ概念上の実体に一致している必要がある。そのため、データ要素が参照する概念のレジストリを開発しているが、これはデータ要素をリンク付けする MADS (Metadata Authority Description Schema: 各典拠レコードに URI を付与できる統制語彙をエンコードする XML アプローチ *14) を利用している。リンクは自体は METS の構造マップの機能を利用し、内部ではデジタルオブジェクトの要素は <area> エlement を使って記述される。より効率的にデータセット間でアクセスするには、様々な関係者のための高度な分析と可視化に加えて、SPARQL を基にしたデータセット間の意味検索を可能にする、METS のアーカイブ表現でパッケージしたリンク機能を基にした RDF モデルにマッピングされる。

(5) Postgraduate Research Data: a New Type of Challenge for Repositories? / Jill Evans, Gareth Cole, Hannah Lloyd-Jones (University of Exeter, United Kingdom)

【概要】

Open Exeter (*15) は生化学、医学分野で大規模データセットの安全な長期保存をするための DSpace リポジトリを構築する、2010-2011 年の JISC ファンドによるプロジェクトである。現在 Open Exeter は試験的なデータリポジトリサービスである Exeter Data Archive (EDA) を学内の全分野に拡大することに取組んでおり、公開した研究と基盤データを同レコードからアクセスできるように、研究成果リポジトリである ERIC (Exeter Research and Institute Content Archive) と EDA を結合することを目的としている。

【内容紹介】

研究データの定義について、かなりオープンなアプローチをとっており、Word ドキュメント、実験のプリントアウト、Excel スプレッドシート、インタビュー記録、音声ファイル、PDF、CSV 形式の生データ、グラフ、写真、解析ファイル、統計ファイル等を含めている。

Exeter のリポジトリは、ERIC、EDA、DCO (Digital Collection Online) があり、それぞれを統合することを計画している。現システムは DSpace1.8.2 で、様々なコンテンツのアップロードをする際に SWORD2 を利用している。

現状は学位論文登録の義務化は行っているが、そのデータについては、戦略、ガイダン

ス、データの保存場所等を用意せず、登録の対象外となっている。大学院生の研究データを保存する理由としては、発見性の改善、研究者プロフィールの充実、ファンドのコンプライアンス、データの開発等が挙げられる。

現状の課題としては、データの再利用の際に問題となる大学院生のデータの紛失、可視性・利用可能性の欠如等を認識しており、学位論文とまとめた簡易的な登録方法の用意、長期のキュレーション（ファイル形式を問わない、永続 ID の付与）、研究コミュニティへの啓発等の解決方法を用意している。ポリシーを 2013 年までに作成する予定（データ登録の義務化、データチェック、学生と指導教員の責任の所在の明確化等）。将来的な課題としては、データを登録する時期、データ内容のチェック、管理コストと持続性が考えられる。

(6) Building an institutional research data management infrastructure / Sally Rumsey (University of Oxford, United Kingdom)

【概要】

オックスフォード大学は研究データの管理、保存、発見のためのインフラを開発している。研究データの目録である DataFinder のようなシステムの実装は急務であると考えている。

【内容紹介】

データリポジトリの必要性は、ファンド機関の意向、データの安全な管理、他の研究者のデータへアクセス、Data 引用の必要性、機関の知的財産としてのデータなど様々あげられる。

JISC のファンドによる DaMaRo (Data Management Rollout at Oxford ^{*16}) は大学における研究データ管理のポリシーと基盤を作成している。DaMaRo は、

- ・研究データ管理のポリシー
- ・研究者やスタッフへのトレーニング、ガイダンス
- ・技術開発・保守
- ・持続可能性へのビジネスプラン

の 4 つの要素で構成されている。

DataFinder は大学の全インフラの Hub として機能する。データを蓄積している DataBank ^(^{*17}) 上に構築されており、それらは互換性があり両システムの入出力はシームレスに行われている。DataBank へのデータ登録は DataStage から SWORD を介して実施されている。DataFinder は、DataBank 以外に Regional DataFinder と Colwiz (Collective Wisdom : 研究者、学生、大学の研究管理と生産のプラットフォーム ^{*18}) からもデータを取得している。連携システムからメタデータを取得後は、コンテンツを関連付け、ID の付与を行う。DataFinder は非デジタルオブジェクトも含めてほぼ何でも登録でき、オブジェクトによってメタデータセットが用意されており、研究者の省力化と分野間の連携性を考慮して必要最小限の構成となっているが、分野やファンドによってはフィールドの追加を

することができる。

(7) Griffith's Research Data Evolution Journey: Enabling data capture, management, aggregation, discovery and reuse. / Natasha Simons, Joanne Morris (Griffith University, Australia)

【概要】

大量に増え続けている研究データ管理の取組みは国際的規模に直面している。グリフィス大学は 2009 年後半に ANDS の支援を受けて、研究データの保存、管理、統合、発見、再利用を進める計画に着手した。計画は研究データリポジトリ、セマンティック web に従うメタデータ保存方法、データの発見に関わる Research Hub^(*19)といった基盤の開発を含んでいる。

【内容紹介】

グリフィス大学は Research Hub で研究データサービスを行っている。Research Hub は、ANDS の資金提供の下、大規模研究データを管理する世界的傾向やファンド主体の要望、グリフィス大学の研究成果のアクセス性の改善等を考慮して開発された。

Research Hub に組み込まれた Meta Data Exchange Hub が学内の研究成果(既存の DB、機関リポジトリ、研究管理データベース等)のメタデータを収集する。また、ANDS や NLA (National Library of Australia) から個人の永続識別子を取得することで、Hub 内部のデータの関連性を構築している。Research Data Australia^(*20)や大学図書館のディスカバリーサービス等の他機関にメタデータを提供することで研究成果のアクセス性を向上させている。

Research Hub は VIVO^(*21)をベースにして構築されている。VIVO はコーネル大学が開発したオープンソースのセマンティック Web アプリケーションで、様々なデータベースからデータを取得し、研究者の研究活動や成果を機関内で分野を超えて発見することができる。データ交換には ISO2146^(*22)に準拠した Registry Interface Format-Collections and Services(RIF-CS^{*23})スキーマを利用している。

ISO2146 は、ネットワーク環境でアプリケーションやドメインの範囲を超えてコンテンツを管理し情報サービスを提供するために図書館等の機関が必要とする collection (データの集合)、party (個人及びグループ)、activity (成果を作り出す活動)、service (ユーザへのインターフェース)に関する情報を提供するデータベースのルールを構築している。

Hub プロジェクトのメリットとしては

- Research Hub はグリフィス大学の研究成果のワンストップサービス。
- 一度データを登録すると何度でも再利用できる。
- 複数のソースを自動収集。
- セマンティック Web を利用し、発見性の改善、豊富なデータマイニングが可能。
- 研究者プロフィールはグリフィス大学のショーケースに。

・データを管理し、つなげ、発見でき、再利用できる環境を構築する。
等が考えられる。

【国内の状況】

文部科学省は2012年7月4日に「ビッグデータ時代におけるアカデミアの挑戦：アカデミッククラウドに関する検討会 提言」^(*)24)を公開した。この提言では、大学等の教育研究機関で発生する膨大なデータを共有する場がなく、有効活用されていない現状に対して、「ビッグデータに関し、我が国が研究開発を進めるべき事項を早急に整理し、分野間連携、国際連携、人材育成を図りつつ、戦略的に研究開発や環境構築に取り組む必要がある」としている。

一方、欧米に目を向けると「米国 NSF や NIH (米国国立衛生研究所) がファンドする研究では、研究により生み出されたデータの提出を義務付けられており」「欧米諸国でも国を挙げての取り組みが進められている」と報告されている。欧米における研究データ管理については上述のとおり、保存的側面、管理的側面、再利用的側面で様々な試みがなされている。

国内においては、機関リポジトリと研究者 DB の連携、研究者 ID プロジェクトや JaLC による DOI の付与等、機関リポジトリのコンテンツの管理、再利用を進める動きはあるが、そのコンテンツの中にデータが含まれているとは言えず、その保存、管理、再利用については今後の課題になると思われる。3つの側面のうち管理的側面と再利用の側面が特に重要になるだろう。これまでの機関リポジトリのコンテンツは論文ファイルがメインであり、それは単体でもコンテンツとしての価値を持つものであったが、データについてはこれまで以上に保存後の展開を考えなければ、価値が向上しないと思われる。したがって、ただ保存するのではなく、保存後のサービスを考えてメタデータを充実させること、データを利用した研究成果と相互につながることで、統合システムを整備しデータにまつわる様々な情報が相互につながることを考慮することで、利用者はデータの存在に気付き、その性質や意味が理解でき、その再利用につながると思われる。

今後に向けてはインフラの整備（データの規模はどれくらいか、どこに（国、組織）保存するのか）、ポリシーの整備（どのようなデータを対象にするのか、データの管理主体はどこになるのか）、サービスの検討（データの統合、API や分析機能等の付加サービスはどこまで提供するのか）等、考えるべきことは多々あると思われる。欧米の事例を参考に国内においても早急な研究データ管理システムの構築が求められる。

以下、研究データ管理について重要と思われるトピックを箇条書きで列挙する。

- ・研究データ管理プラットフォームは、データの自動保存、キュレーション、収集データの組織化、データ再利用のライセンスの整備、データの発見性の促進、再利用、長期保存、データの引用の実現といった要件を満たすことが求められる。
- ・利用性や管理コストの面を考慮すると、研究データ管理のプラットフォームは既存の

研究コミュニティのものが存在するならば活用する方が望ましい。

- ・汎用的なシステムではなく研究コミュニティの要件に合致したシステムを構築するほうが良い。
- ・研究データの管理・再利用のためには、研究者、寄付団体、出版物に関する情報メタデータを充実といったメタデータの充実が求められる。特に研究者 ID による研究者の識別、DOI 等によるコンテンツの識別は重要となる。
- ・研究データを正しく評価するには分析に利用したソフトウェアの情報を保存することも求められる。
- ・研究データは、研究成果物等の関連情報との相互運用性を持つことが特に重要となる。互いに交換できるためのメタデータの整備や、統合システム環境を構築し、まとめて利用できる環境を構築することが望ましい。

=====

Ref. (URL はすべて最終アクセス: 2012-09-27)

*1 Australian National Data Service

<http://www.andis.org.au/>

*2 Chronicling American Collection

<http://chroniclingamerica.loc.gov/>

*3 Viewshare

<http://viewshare.org/>

*4 The Science and Technology Facilities Council

<http://www.stfc.ac.uk/>

*5 ISIS

<http://www.isis.stfc.ac.uk/>

*6 Core Scientific Metadata Model

<http://code.google.com/p/icatproject/wiki/CSMD>

*7 DataCite

<http://www.datacite.org/>

*8 Tessella Safety Deposit Box

<http://www.digital-preservation.com/solution/safety-deposit-box/>

*9 Apache Hadoop

<http://hadoop.apache.org/>

*10 SCALable Preservation Environment

機関の保存戦略の立案、実行について、拡張可能なサービスを開発している。SCAPE 保存要素は、特性評価や傾向分析をとおしてリポジトリの保存の必要性を特定すること、保存計画やポリシーの形式的記述の必要性への対応を定義すること、自由度が高く拡張可能なプロセス、保存プロセスの質の監視を可能にする。

- <http://www.scape-project.eu/>
- *11 Preservation Network Model
<http://epubs.cclrc.ac.uk/work-details?w=51437>
 - *12 Department for Environment, Food and Rural Affairs
<http://www.defra.gov.uk/>
 - *13 ISO19156
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32574
 - *14 Metadata Authority Description Schema
<http://www.loc.gov/standards/mads/>
 - *15 Open Exeter
http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure/openexeter.aspx
 - *16 Data Management Rollout at Oxford
<http://damaro.oucs.ox.ac.uk/>
 - *17 DataBank
<https://databank.ora.ox.ac.uk/>
 - *18 Colwiz
<http://www.colwiz.com/>
 - *19 Research Hub
<http://research-hub.griffith.edu.au/>
 - *20 Research Data Australia
<http://researchdata.andis.org.au/>
 - *21 VIVO
<http://vivoweb.org/>
 - *22 ISO2146
http://www.iso.org/iso/catalogue_detail.htm?csnumber=44936
 - *23 Registry Interface Format-Collections and Services
<http://www.andis.org.au/guides/rif-cs-awareness.html>
 - *24 http://www.mext.go.jp/b_menu/shingi/chousa/shinkou/027/gaiyou/_icsFiles/afieldfile/2012/07/10/1323376_1_2.pdf

4. その他

テーマ別セッションのうち、今後大きく利用が普及する気配のある Altmetrics に関するセッションと、国内のリポジトリの主要な位置を占めている DSpace のセッションについて簡単に報告する。

- (1) Open Metrics for Open Repositories / Brian Kelly (University of Bath, UK), Nick Sheppard (Leeds Metropolitan University), Jenny Delasalle (University of Warwick), Mark Dewey (University of Bath, UK), Owen Stephens, Gareth Johnson (University of Leicester), Stephanie Taylor (University of Bath, UK)

【内容紹介】

リポジトリのメトリクスは将来における政策決定をサポートするリポジトリの利用傾向の理解を進めることに利用できる。

組織的な側面として、リポジトリの重要な目的が出版物へのアクセスの最大化と長期保存であるならば、リポジトリ管理者が興味のあるメトリクスは、総レコード数、フルテキストのみのレコード数、メタデータへのアクセス数、フルテキストへのアクセス数、レコードへのアクセス経路等と思われる。

フルテキストとメタデータをもつリポジトリは研究管理システムの側面を持っており、その利用統計は組織やコミュニティにとって重要なものとなる。一方、メトリクスの可用性においては、実装方法の点で技術的な問題があり、スタッフの専門知識が求められることである。そのため Google Analytics はその簡易的な方法として代用できるかもしれない。

論文レベルのメトリクスはリポジトリへの投稿に影響を及ぼす可能性がある。PLOS^(*) や arXiv^(**) は既に論文と一緒に論文レベルのメトリクスを公開しており、機関リポジトリでも同様でできると思われるが、閲覧人数といった数値を気にする著者もいる。

PLOS は PubMed Central からのアクセス数をメトリクスとして表示しており、JISC の PIRUS プロジェクト^(***) は、論文を共有するメトリクスの課題を調査している。著者は様々なソースのメトリクスを扱える方が良いと考えており、論文レベルのメトリクスに取り組むことは、リポジトリにとって著者に向き合う機会となる。このようなインパクトの測定は、全ての研究成果に関連するオンライン活動を含んでおり、Blog や Tweet のブックマークや retweet、facebook の「いいね」等の測定は、学術コミュニティを超えて数値以上の価値の指標となるかもしれない。Altmetrics マニフェスト^(****) は Web メトリクスの重要性和情報の質の評価と保証についてのメトリクスの役割を述べている。また、Altmetrics は Total-Impact のような複数のソースからメトリクスを収集するツールへリンクしている。

研究者の立場を考えると、リポジトリ管理者は論文に関連するソーシャルな活動を測定し、還元する必要がある。リポジトリのメトリクスは様々な目的と利害関係者のために必要なものである。リポジトリ管理者はリポジトリのメトリクスを積極的にオープンにする意思を示すべきである。

=====

Ref. (URL はすべて最終アクセス: 2012-09-27)

*1 <http://www.plos.org/>

*2 <http://arxiv.org/>

*3 <http://www.jisc.ac.uk/whatwedo/programmes/pals3/pirus.aspx>

*4 <http://altmetrics.org/manifesto/>

(2) UGDS: DuraSpace Plenary and Roadmap

【内容紹介】

DuraSpace の総会では、DSpace と Fedora それぞれのロードマップの説明が行われた。DuraSpace は、デジタルコンテンツの保存に関わるコミュニティの教育や、ソフトウェアの改善等の役割を担っているが、2011 年の成果としては DSpace1.8 と Fedora3.5 のリリース、初のクラウドサービス DuraCloud（保存・リポジトリ・研究データ管理）リリース等が挙げられた。以下、日本国内のリポジトリの多くが採用している DSpace に関連した事項を簡単に紹介する。

DSpace に関する 2011 年以降の達成事項としては、前述の DSpace1.8 リリース、コードの GitHub への移行、バックアップや復旧がたやすくなる “dspace-replicate” アドオンの提供等が挙げられた。DSpace1.8.2 の主な新機能としては、設定変更可能なレビューワークフロー、SWORD クライアント提供、SWORDv2 対応等が挙げられた。また今後の DSpace リリースにあたってのバージョンナンバリングについて、大きなバージョンアップと、バグの修正等の小さなバージョンアップを 2 階層で表現する（例えば「3.0」だと「3」が大、「0」が小）との説明があり、毎年 10 月か 11 月に大きなバージョンアップを行うとのことである。2012 年 10 月か 11 月に DSpace3.0 リリースを予定しているが、主な新機能として、アイテムのバージョン管理、様々なタイプのエンバーゴへの対応、コミュニティやコレクションについてのメタデータサポート、ハンドルだけでなく DOI のような外部の識別子も利用可能にすること、といった見通しが示された。なおこれらのリリースは、2009 年に発表されたプロトタイプ DSpace2.0 や、まだ議論段階にある「DSpace with Fedora Inside」とは無関係である。

(3) UGD1: DSpace User Group Session 1

【内容紹介】

DSpace ユーザグループセッションからは、DuraSpace サービスプロバイダ@mire の活動報告を簡単に紹介する。DSpace バージョンアップに関しては、1.6 の統計機能、1.7 のファセット検索のようなディスカバリ機能、1.8 の設定変更可能なワークフロー等に取り組んできた。3.0 では、検索結果におけるスニペット表示やハイライト表示、階層的ファセット、関連アイテム表示、アクセス権設定（利用対象者の限定）、ユーザビリティ改善に取り組んでおり、統計機能についてもパフォーマンスの最適化、検索履歴、ワークフローの統計に取り組んでいるとのことである。

5. 最後に

OR2012の会場となったエジンバラ大学では、リポジトリ周辺領域についてインフォーマルなスタイルで議論を交わすイベント **Repository Fringe** ^(*1)が 2008 年より毎年開催されているが、今回の OR2012 のプログラムにはこの **Repository Fringe** も組み込まれ、リポジトリをめぐってより幅広い議論が行われた。このほか、学生を含め開発者がリポジトリに関連したサービス等のアイデアを発表し競い合う **Developer Challenge** ^(*2)、総会における **Cameron Neylon** 氏による講演等、テーマ別セッションも含め、本報告に含めなかったものも多くある。これらのプログラムおよびほとんどの発表資料は OR2012 のサイト ^(*3) で公開されており、動画や報告ブログなどの情報も同サイトで確認できるので、ぜひそちらも参照頂きたい。

今回、この会議に参加する機会を得たことに感謝するとともに、会議から受けた多くの刺激を今後の活動に活かしていきたい。

=====

Ref. (URL はすべて最終アクセス: 2012-09-27)

*1 Edinburgh Digital Library, Edina, Digital Curation Centre の三者で開催している。

*2 <http://devcsi.ukoln.ac.uk/developer-challenges/developer-challenge-or-12/>

*3 <http://or2012.ed.ac.uk/>

6. 写真集

	
<p>会場となった Appleton Tower</p>	<p>Appleton Tower のホール (Coffee breaks)</p>
	
<p>ポスター発表の様子</p>	<p>ドリンクレセプション会場の様子</p>
	
<p>テーマ別セッションの様子</p>	<p>ディナー会場の様子 National Museum of Scotland</p>