

CA1690

動向レビュー

## デジタルリポジトリにおける メタデータ交換の動向

### 1. はじめに

近年、大学図書館を中心に、デジタル化された学術論文や研究資料を蓄積・保存し、インターネットで公開を行う、いわゆるデジタルリポジトリの構築が盛んである。こうしたリポジトリは、収録コンテンツを自機関の成果物に限定するものを機関リポジトリ、特定主題分野に収集対象を定めるものをサブジェクトリポジトリなどと呼びならわしている。制度的にもシステムのにもいまだ発展途上の段階にあり、より安全でより便利なものにするための工夫が続けられている。

リポジトリが機能を高めるための方法の一つとして、保持するコンテンツ（データ）に関するさまざまな情報を記録したデータ、すなわちメタデータを有効活用することが考えられる。リポジトリにおけるメタデータは、一般に、従来の書誌情報に相当する記述メタデータと管理的な情報を記録する管理メタデータとに大別され、前者は情報資源の発見、後者はデジタル資料の管理・保全、さらには長期保存などの目的に利用することが想定されている。これらのメタデータはリポジトリ間で交換・共有されてこそ意味があり、そのための規定作りや標準化が盛んに行われている。

本稿では、そうしたメタデータ交換の動向を、OAI-OREとPREMIS/METSという二つの活動を中心に探っていきたい。

### 2. OAI-PMH から OAI-ORE へ

2000年代の初頭から現在に至るまで、いくつものリポジトリ用ソフトウェアが開発され、改訂を重ねてきた。現在、多くのリポジトリが採用している代表的なソフトウェアとして、DSpace<sup>(1)</sup>、EPrints<sup>(2)</sup>、Fedora<sup>(3)</sup>などがある。

これらのソフトウェアに共通している基本的な機能として、OAI-PMH対応がある。OAI-PMHは各リポジトリが蓄積しているメタデータを共有するためのプロトコルである（CA1513参照）。欧米では早い時期からOAI-PMHにより収集したメタデータを利用した検索サービスが立ち上がっており、ミシガン大学が運営するOAIster<sup>(4)</sup>はその代表例として有名である。

日本の大学図書館が運営するリポジトリも多くはDSpaceやEPrintsを採用しており、それ以外の

ソフトウェアもOAI-PMH準拠をうたっているのだが、日本でOAI-PMHを利用したメタデータ提供が本格的に行われるようになったのは比較的最近である。国立情報学研究所（NII）が「junii2」<sup>(5)</sup>というメタデータ・フォーマットおよび学術機関リポジトリデータベース収集方針<sup>(6)</sup>を定め、国内の機関リポジトリで生成されるメタデータをOAI-PMHで収集し、統合的に検索できる学術機関リポジトリポータル（JAIRO）<sup>(7)</sup>を公開している。このJAIROは2008年10月に試験公開され、2009年4月に正式公開された。また、国立国会図書館（NDL）のデジタルアーカイブポータル（PORTA）の外部提供インターフェースにOAI-PMHが追加されたのは2008年12月である（CA1677参照）。なお、NIIはJAIROとあわせて、収集したメタデータを利用して、参加機関の統計分析情報を提供するIRDBコンテンツ分析システムも構築している<sup>(8)</sup>。

一方、OAI-PMHを策定したオープンアーカイブズイニシアティブ（OAI）では、2008年10月、オープンアーカイブのオブジェクトの再利用・交換に関するプロトコル“Open Archives Initiative - Object Reuse and Exchange（OAI-ORE）”の正式版<sup>(9)</sup>を発表した。これはウェブ上の情報資源の集合体の記述および交換の標準を定めたものである。

現在、ウェブ上では多くの情報資源がひとまとまりで利用されるものの、その構成要素や境界が曖昧なため、コンピュータで処理できないという問題がある。たとえば物理学等の学術論文のプレプリントサーバ“arXiv”における論文詳細表示画面では、論文タイトル、著者名、版、抄録、本文へのリンク、同じコレクションの別の論文へのリンクなどが表示されており、このページのURI（Uniform Resource Identifier）が論文を示す識別子としてよく使われる。しかし、この画面の諸要素は人間の目には一目瞭然だが、コンピュータがarXiv専用の特別な処理をすることなしに識別することは困難である。

OAI-OREはこうした問題を解決するために、情報資源の集合体をコンピュータが解釈できる形で記述するリソースマップというものを規定し、これをもとにコンピュータがデータ交換を行う<sup>(10)</sup>。

つまり、OAI-OREではメタデータのみならずリポジトリ内の個々の情報資源（コンテンツ）そのものを（必要に応じてメタデータと共に）柔軟にやり取りできるようになるわけで、アルファ版、ベータ版の段階から関係者の注目が集まった。日本においてもOAI-OREに対する関心は高く、たとえばNDLはカレントアウェアネス-Rで逐一その動向を伝えており、NIIはアルファ0.2版と正式版（1.0版）の仕

様書およびユーザガイドを翻訳公開している<sup>(11)</sup>。

### 3. OAI-ORE の活用

OAI-ORE は正式版の発表があって間もないため、リポジトリの現場で活用されるのはこれからだが、試験的プロジェクトはいくつか実施されている。ここでは英国情報システム合同委員会 (JISC) が助成する二つのプロジェクトを取り上げる。

カレントアウェアネス-Rでも紹介されているが、FORESITE (Functional Object Re-use and Exchange: Supporting Information Topology Experiments) は OAI-ORE を利用して、学術雑誌のバックナンバーデータベース JSTOR に収録されている雑誌のデータをリソースマップ化し、SWORD (Simple Web-service Offering Repository Deposit) インターフェースを介して、Atom 出版プロトコル (Atom Publishing Protocol)<sup>(12)</sup> に準拠した文書として DSpace に取り込むというものである。この SWORD も JISC の助成を受けたプロジェクトで開発されたプロトコルで、Atom 出版プロトコルの上に、リポジトリへの標準的なコンテンツ納入方式を規定したものである<sup>(13)</sup>。これによりコンテンツを複数のリポジトリに同時に納入したり、リポジトリ間でコンテンツをやり取りしたりするといったことが可能になる。OAI-ORE ではリソースマップの記述方法の一つとして Atom が利用可能であるため、FORESITE のように両者を組み合わせて使う応用例が今後出てくると思われる。

PRESERV2 プロジェクトは、デジタル資料の長期保存という観点から OAI-ORE の活用を目指している。ラムジー (Sally Rumsey) とオステーン (Ben O' Steen) は、OAI-ORE の利点として、差分アップデートができること、すべてのメタデータおよびデジタル・オブジェクト間の関係や履歴も含めて情報の喪失がない複製が可能なこと、リソースマップが機械可読であるためコンピュータが定められた方針に従って最適の保存手続きを確定してくれることなどをあげている。これらにより、あるリポジトリのコンテンツを別のリポジトリに容易にコピーができるため、数多くの複製を持って安全性を高めることができるとしている<sup>(14)</sup>。このプロジェクトでは、OAI-ORE 活用の具体例として、EPrints と Fedora という二つの異なったソフトウェアで構築されたリポジトリ間でデータを交換する仕組みを作り、2008 年 4 月、英国サウサンプトンで開催された「オープン・リポジトリ (Open Repositories) 2008」会議でデモを行った<sup>(15)</sup>。

このように、OAI-PMH がもたら資源発見のため

のメタデータ共有の仕組みであったのに対し、OAI-ORE はリポジトリにおけるデジタル資料の長期保存体制作りにも活用できると期待されている。

### 4. PREMIS 2.0

デジタル資料の長期保存に資するメタデータに関しては 2000 年代初頭からさまざまな研究が行われてきた (CA1489、CA1561 参照) が、現在、それらの頂点に立つのが PREMIS (Preservation Metadata: Implementation Strategies) の『保存メタデータのためのデータ辞書』第 2.0 版 (通称 PREMIS2.0)<sup>(16)</sup> であろう。これは保存活動に必要と思われるさまざまな情報を整理・体系化し、その意味範囲や用法を規定した、まさにメタデータ要素の辞書なのだが、具体的な実現方法とは独立したものにするため、あえて要素 (element) とは言わず、意味単位 (semantic unit) という言葉を使っている。

このデータ辞書の第 1 版は 2005 年に発表され、英国のデジタル保存賞、米国アーキビスト協会保存出版賞を受賞するなど高い評価を受けた。日本にも 2007 年の時点での保存メタデータの到達点として紹介がなされている<sup>(17)</sup>。

第 1 版の発表後、PREMIS 作業グループは解散し、このデータ辞書の維持活動の場は米国議会図書館 (LC) に移り、改訂のための編集委員会が設置された。同時に、実際に運用されているリポジトリへの適用の検討も始まった。たとえば上述の PRESERV2 プロジェクトの前身である PRESERV では、PREMIS のメタデータを機関リポジトリのシステム (EPrints) にマッピングする研究を行っている<sup>(18)</sup>。

2007 年、PREMIS 編集委員会は十分なフィードバックが得られたとして改訂作業に入り、2008 年 4 月、現在の第 2.0 版が発表された。改訂の経緯および主な改訂内容については、ラヴォワ (Brian F. Lavoie) の記事に詳しい<sup>(19)</sup>。

### 5. PREMIS と METS

PREMIS が規定するのはメタデータの意味単位であり、実際にそれを記録するための方式は別に定めることになるが、PREMIS ではデータ辞書に沿った XML スキーマを作成して公開している<sup>(20)</sup>。一方、XML ベースでメタデータを記録し、交換するための標準に METS (Metadata Encoding & Transmission Standard; CA1489、CA1552 参照) があり、PREMIS 同様、LC が維持管理を行っている。当然、PREMIS の XML スキーマで記録したメタデータを METS に入れ込もうという動きが出てくるわけだが、現実にはすんなり行かないことが明らかになってきている。

一つはセクションの分け方の違いで、PREMIS がエンティティのタイプ（オブジェクト、イベント、エージェント）で分けているのに対し、METS はメタデータのタイプ（記述、技術、来歴、構造など）で分けているため、PREMIS の各要素を METS の該当するセクションに振り分けるマッピングを行わなくてはならない。もう一つは、両方に定義されているため重複する要素があるということで、これはそのまま重複して記録するのか（その際には修正があった場合、齟齬がないよう両方ともきちんとして修正しなくてはならない）、それともどちらか一方の側に記録するのかを決めなくてはならない<sup>(21)</sup>。

PREMIS を METS で利用する代表的な事例として、オーストラリア METS プロファイル (Australian METS Profile) がある。これは 2007 年、LC に登録されたもので、前年の PRESTA (PREMIS Requirements Statement) プロジェクトを受けて完成されたものである<sup>(22)</sup>。また、イリノイ大学を中心とした ECHO DEPOSITORY プロジェクトでも、やはり PREMIS や、記述用メタデータ MODS (Metadata Object Description Schema ; CA1552 参照) を拡張スキーマとした METS のプロファイル (ECHO DEP プロファイル) を作成している<sup>(23)</sup>。PREMIS の Web サイトでは、PREMIS を METS 内で扱う際のガイドラインを作成して公開するとともに、そうした METS プロファイルを掲載するなど情報提供を行っている<sup>(24)</sup>。

しかし、マクドナフ (Jerome McDonough) も指摘するように、こうしたプロファイルは狭い範囲でしか通用せず、それぞれ独立して策定された標準をすり合わせるのには容易ではないようである<sup>(25)</sup>。PREMIS データ辞書第 1 版の作成に携わったフロリダ図書館自動化センターのキャプラン (Priscilla Caplan) は、コーネル大学、ニューヨーク大学などをパートナーに、TIPR (Towards Interoperable Preservation Repositories) プロジェクトに取り組んでいる。TIPR では、さまざまなマッピングの試みを検討した上で、上記 ECHO DEP プロファイルを起点にして、汎用性のある交換プロファイルを定義することを目標としており、PREMIS や METS および両者の使用ガイドラインの改訂に貢献したいとしている<sup>(26)</sup>。

(常磐大学：栗山正光)

(1) DSpace. <http://www.dspace.org/>, (accessed 2009-04-12).  
 (2) Eprints. <http://www.eprints.org/>, (accessed 2009-04-12).  
 (3) Fedora Commons. <http://www.fedora-commons.org/>, (accessed 2009-04-12).  
 (4) OAIster. <http://www.oaister.org/>, (accessed 2009-04-12).  
 (5) “メタデータ・フォーマット junii2”. 国立情報学研究所.

<http://www.nii.ac.jp/irp/archive/system/junii2.html>, (参照 2009-04-12).  
 (6) “国立情報学研究所 学術機関リポジトリデータベース収集方針”. 国立情報学研究所. [http://www.nii.ac.jp/irp/archive/system/pdf/1\\_nii\\_irdb\\_shushu.pdf](http://www.nii.ac.jp/irp/archive/system/pdf/1_nii_irdb_shushu.pdf), (参照 2009-04-12).  
 (7) JAIRO : Japanese Institutional Repositories Online. <http://jairo.nii.ac.jp/>, (参照 2009-04-12).  
 (8) IRDB コンテンツ分析システム. <http://irdb.nii.ac.jp/analysis/index.php>, (参照 2009-04-12).  
 (9) Open Archives Initiative Object Reuse and Exchange. <http://www.openarchives.org/ore/>, (accessed 2009-04-12).  
 (10) “ORE User Guide - Primer”. <http://www.openarchives.org/ore/1.0/primer.html>, (accessed 2009-04-12).  
 (11) “ORE 仕様書およびユーザガイド - 目次”. <http://www.nii.ac.jp/irp/archive/translation/oai-ore/1.0/toc.htm>, (参照 2009-04-12).  
 (12) “RFC 5023: The Atom Publishing Protocol”. Internet Engineering Task Force. <http://www.ietf.org/rfc/rfc5023.txt>, (accessed 2009-04-12). 和訳版; [http://www.ricoh.co.jp/src/rd/webtech/rfc5023\\_ja.html](http://www.ricoh.co.jp/src/rd/webtech/rfc5023_ja.html), (accessed 2009-04-12).  
 (13) “Welcome to SWORD APP”. SWORD Site. <http://www.swordapp.org/>, (accessed 2009-04-12).  
 (14) Rumsey, Sally ; O’Steen, Ben. OAI-ORE, PRESERV2 and Digital Preservation. *Ariadne*. 2008, (57). <http://www.ariadne.ac.uk/issue57/rumsey-osteen/>, (accessed 2009-04-12).  
 (15) Tarrant, David et al. Using OAI-ORE to Transform Digital Repositories into Interoperable Storage and Services Applications. *The Code4Lib Journal*. 2009, (6). <http://journal.code4lib.org/articles/1062>, (accessed 2009-04-12).  
 (16) PREMIS Editorial Committee. “PREMIS Data Dictionary for Preservation Metadata. version 2.0”. March 2008. <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>, (accessed 2009-04-12). 和訳版; 栗山正光訳. PREMIS 保存メタデータのためのデータ辞書. 第 2.0 版. (日本図書館協会にて翻訳権取得交渉中)  
 (17) 後藤敏行. デジタル情報保存のためのメタデータ: 現状と課題. *情報管理*. 2007, 50 (2), p. 74-86. <http://joi.jlc.jst.go.jp/JST.JSTAGE/johokanri/50.74>, (参照 2009-04-12).  
 (18) Hitchcock, Steve et al. “Preservation Metadata for Institutional Repositories: applying PREMIS”. *Preserv.* <http://preserv.eprints.org/papers/presmeta/presmeta-paper.html>, (accessed 2009-04-12).  
 (19) Lavoie, Brian F. PREMIS With a Fresh Coat of Paint: Highlights from the Revision of the PREMIS Data Dictionary for Preservation Metadata. *D-Lib Magazine*. 2008, 14 (5/6). <http://dx.doi.org/10.1045/may2008-lavoie>, (accessed 2009-04-12).  
 (20) “Schemas for PREMIS”. PREMIS: Preservation Metadata Maintenance Activity (Library of Congress). <http://www.loc.gov/standards/premis/schemas.html>, (accessed 2009-04-12).  
 (21) Guenther, Rebecca S. Battle of the Buzzwords: Flexibility vs. Interoperability When Implementing PREMIS in METS. *D-Lib Magazine*. 2008, 14 (7/8). <http://dx.doi.org/10.1045/july2008-guenther>, (accessed 2009-04-12).  
 (22) Pearce, Judith et al. The Australian METS Profile: A Journey about Metadata. *D-Lib Magazine*. 2008, 14 (3/4). <http://dx.doi.org/10.1045/march2008-pearce>, (accessed 2009-04-12).  
 (23) “ECHO Dep Generic METS Profile for Preservation and Digital Repository Interoperability”. Metadata Encoding and Transmission Standard (METS) Official Web Site . <http://www.loc.gov/standards/mets/profiles/00000015.html>, (accessed 2009-04-12).  
 (24) “Using PREMIS with METS”. PREMIS: Preservation Metadata Maintenance Activity (Library of Congress). <http://www.loc.gov/standards/premis/premis-mets.html>, (accessed 2009-04-12).  
 (25) McDonough, Jerome. “Structural Metadata and the Social Limitation of Interoperability: A Sociotechnical View of XML and Digital Library Standards Development”. *Proceedings of Balisage: The Markup Conference 2008*. Montréal, Canada,

2008-08-12/15.

<http://balisage.net/Proceedings/print/2008/McDonough01/Balisage2008-McDonough01.html>,  
(accessed 2009-04-12).

- (26) Caplan, Priscilla. Repository to Repository Transfer of Enriched Archival Information Packages. D-Lib Magazine. 2008, 14 (11/12).  
<http://dx.doi.org/10.1045/november2008-caplan>,  
(accessed 2009-04-12).