

【2003年度愛知淑徳大学大学院図書館情報学専攻研究討論会：Ms. Jewel Wardによる講演の要旨】

デジタル図書館におけるメタデータ Metadata in digital libraries

講演者：Jewel Ward*

講演記録：櫻木貴子**
Takako SAKURAGI

概要

本講演は平成15年度愛知淑徳大学図書館情報学専攻研究討論会として2003年6月13日に愛知淑徳大学において行われたものである。

講演は次の3つのパートに分かれている。

- (1) 講演者の所属するLos Alamos National Laboratory (LANL) およびLANL Research Libraryの紹介
- (2) 電子図書館におけるメタデータ
- (3) 講演者の修士論文『A quantitative analysis of unqualified Dublin Core metadata element sets usage within data providers registered with the Open Archives Initiative』の概要

* Assistant Researcher, Los Alamos National Laboratory Research Library

** 愛知淑徳大学文学部図書館情報学科非常勤講師

Part time lecturer, Department of Library and Information Science, Aichi Shukutoku University

1. ロスアラモス国立研究所 [1] および同研究所研究図書館 [2] の活動

米国国立の研究所であるLos Alamos National Laboratory (LANL) の使命は、科学技術を利用した世界の安全化や大量殺戮兵器の脅威の排除、エネルギーや環境、健康に関する安全の向上である。

LANL Research LibraryはLANLの研究員に対する科学技術情報の提供を主とした研究図書館であり、54名の図書館員が1万人超の研究員に対してサービスを行っている。時間や場所を問わず研究に必要な情報へのアクセスを確保し、研究員の活動を支援することにより、研究員の科学コミュニケーションや共同研究を促進することを目的としている。

利用者は次のような3種類に区分されている。

- (1) LANL内の研究者
- (2) LANLの電子プロダクト購入者
- (3) LANL外の研究者 (LANL研究者との共同研究者等)

LANL Research Libraryの主なセクションは次の5つである。

(1) Business Support

LANL Research Libraryに関わるあらゆる研究者に対してサービスを行う。LANL内の研究開発部門によって作成された製品等の管理も行う。

(2) Customer service

レファレンスサービス、コピーサービス、図書館相互貸借、貸出等のパブリックサービスを行う。

(3) Library Integrated Products

図書や雑誌、データベースの購入、目録作業、排架等のテクニカルサービスを行う。

(4) Library without Walls

研究開発部門に当たる部署。LANLの研究員およびその共同研究者が、インターネットを通じて時間場所を問わず、あらゆる壁を外

した図書館として、情報資源にアクセス可能なシステムを指向)。

(5) Reports

The vault (地下室) において保存・管理する130万件の技術レポート、データベース、フルテキストレコード、書誌レコード等へのアクセスを提供する。印刷体の情報資源を電子化して提供するデジタル化プロジェクトを行う。

【質疑応答】

Q : Classified reports (機密文書) へのアクセスは制限されているのか

A : 中央集中管理がされており、機密のレベルが高くなるにつれ、アクセスしようとする人間に対する調査の詳細さが増す。最高機密「Q」は利用目的を明確にしないとLANL内部の人間でもアクセスできない。

Q : 3種類の利用者の優先順位や提供されるサービスに違いはあるのか

A : LANL内の研究員を最優先にサービスを行っているが、協定を結んでいるSandia National Laboratoriesの研究員に対しても一定のサービスを提供している。

2. 電子図書館におけるメタデータ

2. 1 メタデータ

メタデータは「データのデータ」と呼ばれ、情報源に関する記述により情報源の検索や管理を簡単にするための構造的な情報である。米国議会図書館はこのメタデータを3種類挙げている。

(1) 記述的メタデータ (Descriptive metadata)

情報源の発見を主目的とする。MARC formatやEncoded Archival Description (EAD), Dublin Core等。Open Archives Initiative (OAI) プロジェクトでは、相互

互換性を重要視しているため、Dublin Core (DC) を最低限必要なメタデータとして採用するよう求めている。

(2) 構造的メタデータ (Structural metadata)

特定のデジタルオブジェクトを利用者に閲覧させるためにシステムが使用する。

(3) 管理的メタデータ (Administrative metadata)

「Policy」とも呼ばれ、デジタルコレクションの管理に使用する。創成や質的管理のための技術的情報やアクセス管理のための必須要件、保存のための情報等が含まれる。保存に関してはDay [4] の論文やOAISに関する論文 [5] を参照のこと。

管理的メタデータを利用した主なプロジェクト例として、「METS」というコード化用管理メタデータを含む標準を生み出す基となった米国のMaking of America II (MOA2) [6] や、保存に重点を置いたプロジェクトを展開しているNational Library of Australia [7]、デジタル資料の保存を重視すべきであると世に知らしめた英国のCedars [8] がある。

2. 2 Open Archives Initiative Protocol for Metadata Harvesting

Metadata harvestingとはメタデータを収穫するという意味である。Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) ではDublin Core Metadata Element Set (DCMES) の付与を最低限必要としているが、他のより表現豊かなメタデータ (例：MARC, RFC-1807, Open Languages Archives (OLA) Community Metadata Set [9], Electronic Theses and Dissertation Metadata Set (ETD-ms) [10]) を作成利用しての記述を強く求めている。OAIにおけるDCを利用して記述されたレコードの例は図1参照のこと。

2. 3 メタデータの問題点

Cornell大学でNational Science Digital Library (NSDL) プロジェクトに携わっているDushay [11] は、メタデータの主な問題点として次の4点を挙げている。

- (1) データのないエレメントセットの存在
- (2) 不正確なデータの存在
 - a. エレメントセットに対して内容が不正確
 - b. 必要のないデータによる不正確さ (例：空白にすべき項目に「unknown」の記述)
- (3) 紛らわしいデータの存在
- (4) 不十分なデータの存在

また、講演者が82のデータ・プロバイダ (DP) から収集した大量のメタデータの分析の結果を以下に挙げる。

〈DCMESの利用率〉

最も利用されている上位5つのエレメント (creator, identifier, title, date, type) だけで全体の71%を占め、最も利用されていない5つのエレメント (language, format, relation, contributor, source) は全体の6%に留まった。

〈データに付与されるメタデータ〉

54%のDPは、全エレメントの約半数に対して2つのエレメント (creatorとidentifier) のみを付与していた。したがって、DCMESは全て利用されていないことと、Dushayが明らかにしたデータの一貫性の不備を合わせて考えると、OAIにおけるより複雑なサービス (情報源横断検索サービス等) の提供には困難を伴うことが予想される。

【質疑応答】

Q：NSDLプロジェクトの詳細な説明を聞きたい。また、収録されているデータの作成者は誰か。挙げられた問題点はメタデータ作

成中に明らかになったものか

A : 教育者のための科学サイトを集めてメタデータを作成するという大規模プロジェクトである。基本的にオリジナルのメタデータを投稿するものであり、データ作成者はWebサイト作成者自身の可能性がある。例えば、基のメタデータがある図書館の目録のもので、不明部分に「unknown」と記述する規則があった場合、NSDLのDCを基礎とするメタデータに変換される際に特定のエレメントに必要な「unknown」が入ることが起こりうる。つまり、データのクリーニングは必須条件ではないだろう。

Q : 挙げられた問題点はNSDLのメタデータに変換時に起こるものと考えて良いか

A : NSDLプロジェクトは新たなメタデータを作成するわけではなく、教育者のための科学サイトという主旨に沿うデータを収集・統合することが目的である。したがって、投稿されたオリジナルのメタデータの質がNSDLのメタデータの質に関わるということだと考えられる。

Q : OAIについて詳細な説明を聞きたい

A : OAI-PMHは、複数のDPで作成されるメタデータの相互互換性向上を目的としている。メタデータの仕様を決めるものでもなく、DCそのものでもない。また、検索サービスやデータベースでもなく、OAIのサービスとも異なる。

次に、DPとサービスプロバイダ (SP) について。前者はレポジトリと呼ばれ、メタデータ自体を保持するものであり、後者はハーバスタと呼ばれ、DPが提供するメタデータを集めて検索サービス等を行うものである。つまり、OAI-PMHはSPがDPからデータを集める際の規約である。これらの関係は図2を参照のこと。SPがOAI-PMHを使ってDP (図2ではMetadata

Provider) にHTTPでリクエストを送る。DPは、受け取ったリクエストに則して保持するメタデータをDCに変換し、SPにHTTPで答えを返すという仕組みである。しかし、仕組みが良くてもメタデータ自体が貧弱である状況から、サービスの質も高くないのではないかと考えられている。OAIについては [16] を参照のこと。OAI-PMHは国立情報学研究所が日本語版を提供している [17]。

3. 修士論文『Open Archives InitiativeにおけるDublin Core Metadata Element Setの利用に関する分析』の概要

3. 1 研究背景・目的

OAIに登録されたDPは、横断検索の促進や電子図書館の相互互換性の向上を目的として、保持するメタデータをSimple Dublin Core (DCS) に対応させることを求められる。本研究の目的を下に挙げる。

- (1) 最も利用され、また利用されていないDCSのエレメントの把握
- (2) DPの違いとエレメントの利用率の関連性を明らかにすること
- (3) DCSの全エレメントが利用されているのかを明らかにすること

3. 2 調査分析方法

- データ収集期間：2002年5月～11月
- OAIに登録されている82のDPからデータを収集
- OAI-PMH Ver.1.1を使用
- データ集計にはPerlで作成したプログラムを使用

3. 3 結果

約91万件のデータのデータを収集した。DPは3種類が存在し、各DPのレコード数の

割合は下記の通りであった。

- 科学技術 (27) : 21%
- 人文科学 (33) : 43%
- コンボ (科学技術+人文科学) (22) : 26%

OAIは科学技術分野から始まったものであるにも関わらず、人文科学に分類されるレコード数が最も多いことに驚いた。また、1レコード当たり平均8つ、1DP当たり平均約91,000のエレメントが使用されていた。最も使用されている5つのエレメントの合計は、全エレメント使用数の71%を占めた。最も使用されていない5つのエレメントは合計しても6%であった。

DP別に見た結果においても、最も使用されている5つのエレメントは同じだが、使用されていないエレメントに変化があった (languageとformatの代わりにrightsとcoverageが入った)。さらに、DPの種類別に見た場合、最も使われているエレメントに変化が見られた。各分野1つずつ入れ替わっており、科学技術とコンボではdate、人文科学ではrightsが入った。50%以上を占めるエレメントは、科学技術ではcreatorとidentifier (50%以上)、人文科学ではtitle, identifier, creator, type (48%)、コンボではcreator, identifier, date (60%)であった。また、54%のDPにおいてcreatorとidentifierの2つのエレメントだけで50%以上となっていた。

3. 4 結論

DCMESの全てが使用されているわけではないことが明らかになった。したがって、OAIにおける情報源横断検索サービス等の提供は難しいと考えられる。本研究に関する詳細は講演者のサイト [20] を参照のこと。

【質疑応答】

Q : DPは公的なものと商業ベースのものどちらが多いのか。エレメントが全て使われ

ていない理由は何か

A : DPは誰でもなれるが、作成者による分類はしていないためわからない。DCが使われていない理由は「不明」。考えられるのは、基になるメタデータをDCにマッピングしていないということである (DCが好きではない、もしくはDCがシンプルすぎてマッピングできない)。さらに、メタデータ作成者により質が違う可能性があるため、作成者別の比較を考えている。また、追加研究として、DCのみを使用したメタデータと、独自に詳細に作り使用しているNative metadataからDCへ変換している場合の比較を考えている。

Q : OAIの人間はあなたの修士論文をどのように受け取ったか

A : 具体的にどうしろとは言われていない。より詳細なメタデータをDCにマッピングすることはそれほど難しくはない。しかし、DCはシンプルすぎるとの考えが一般的で、あまり使われない方向に来ていると感じ、メタデータの側に問題があるのではないかと考えている。システムの作りは良いが、メタデータの生で上手く働かないのではという指摘を受けたことがある。

Q : LANLのLibrary without Wallsの活動を具体的に紹介して欲しい

A : 現在、OA Static Repository (OASR) というOAIのアグリゲータのようなものを作ろうという大規模プロジェクトがある。OASRは「データを持っているだけ」のレポジトリのデータをOAIが認識可能な形に変換する仕組みである。例えば、言語学者は研究対象である言語に関する内容をメタデータとして記述することは可能だろうが、DPになるための知識に関しては難しいと考えられる。OASRはそれを支援し、複数の情報資源やメタデータを自由に使い、障

壁を低くするものと言える。

Q : メタデータ付与者は基データ作成者とメタデータ付与の専門家のどちらが良いのか

A : 図書館員等の専門家が作成した方が質は良いだろう。しかし、現在デジタルデータが非常に多くなっており、自動作成されたメタデータが増え、質が落ちる可能性がある。このような場面で情報専門家の対応にかかっている。また、メタデータの作成を自分の仕事として認識しない基データの作成者が付与者である場合には、適当に処理をして質を悪くする可能性もある。

【講演者略歴】

(e-mail : jewelw@lanl.gov)

North Carolina大学Chapel Hill校School of Information and Library ScienceにてMaster of Science in Information Science取得。現在、米国Los Alamos National Laboratory Research LibraryにてAssistant Researcherとして勤務。

【引用／参考文献】

- [1] Los Alamos National Laboratory. Los Alamos National Laboratory Homepage [online]. 2003 [cited 2003-10-02]. Available from World Wide Web : <<http://www.lanl.gov/worldview/>>.
- [2] LANL Research Library. LANL Research Library [online]. 2003 [cited 2003-10-02]. Available from World Wide Web : <<http://lib-www.lanl.gov>>.
- [3] Cornell University Research Library/ Research Department. Table 5.1 Metadata Types [online]. 2003 [cited 2003-10-02]. Available from World Wide Web : <<http://www.library.cornell.edu/preservation/tutorial/metadata/table5-1.html>>.
- [4] Day, Michael. Joint RLG and NPO Conference : Guidelines for Digital imaging [online]. 1999 [cited 2003-10-02]. Available from World Wide Web : <<http://www.rlg.org/preserv/joint/day.html>>.
- [5] OCLC / RLG Working Group on Preservation Metadata. Preservation metadata and the OAIS (Open Archival Information System) information model : A metadata framework to support the preservation of digital objects [online]. 2002 [cited 2003-10-02]. Available from World Wide Web : <http://www.oclc.org/research/projects/pmwg/pm_framework.pdf>.
- [6] Hurley, Bernard J. et al. The Making of America II Testbed Project : A digital library service model. [online]. 2001 [cited 2003-10-02]. Available from World Wide Web : <<http://www.clir.org/pubs/abstract/pub87abst.html>>.
- [7] National Library of Australia. National Library of Australia [online]. 出版日付不明 [cited 2003-10-02]. Available from World Wide Web : <<http://www.nla.gov.au/>>.
- [8] Sergeant, D. M. The Cedars (CURL (Consortium of University Research Libraries) exemplars in digital archives) Project : Metadata for Digital Preservation [online]. 2001 [cited 2003-10-02]. Available from World Wide Web : <<http://www.leeds.ac.uk/cedars/metadata.html>>.
- [9] Simons, Gary ; Bird, Steven. OLAC Metadata Set [online]. 2001 [cited 2003-10-02]. Available from World

- Wide Web : <<http://www.language-archives.org/OLAC/olacms.html>>.
- [10] Atkins, Anthony ; France, Robert. NDLTD (Networked Digital Library of Theses and Dissertations) - ETD-ms : an Interoperability Metadata Standard for Electronic Theses and Dissertations version1.01 [online]. 2001 [cited 2003-10-02]. Available from World Wide Web : <<http://www.ndltd.org/standards/metadata/etd-ms-v1.01.html>>.
- [11] Dushay, Naomi ; Hillmann, Diane I. Analyzing metadata for effective use and re-use [online]. 2003 [cited 2003-10-02]. Available from World Wide Web : <http://www.cs.cornell.edu/naomi/DC2003/dushay_hillmann_draft.pdf>.
- [12] Hodge, Gail. Metadata made simpler [online]. 2001 [cited 2003-10-02]. Available from World Wide Web : <http://www.niso.org/news/Metadata_simpler.pdf>.
- [13] Liu, Xiaoming et al. "Technical report interchange through synchronized OAI caches". *ECDL 2002 : Research and Advanced Technology for Digital Libraries 6th European Conference*. Agosti, Maristella ; Thanos, Costantino ed. Rome, 2002-09, Berlin, Springer, 2002, p.176-189.
- [14] Nelson, Michael L. Better interoperability through the Open Archive Initiative. *The New Review of Information Networking*. vol.7, 2001, p.133-146.
- [15] Ward, Jewel. "A quantitative analysis of unqualified Dublin Core Metadata Element Set usage within data providers registered with the Open Archives Initiative". *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries*. Delcambre, Lois et al. ed. 2003-05, Houston, Texas, ACM / IEEE. Washington, DC, IEEE Computer Society, 2003, p.315-317.
- [16] Open Archives Initiative. Open Archives Initiative [online]. 2003 [cited 2003-10-02]. Available from World Wide Web : <<http://www.openarchives.org>>.
- [17] 国立情報学研究所開発・事業部コンテンツ課文字情報係. OAI-PMH2.0日本語訳 [online]. 2003 [cited 2003-10-02]. Available from World Wide Web : <<http://www.nii.ac.jp/metadata/oai-pmh2.0/>>.
- [18] Day, Michael ; Powell, Andy. UKO LN Metadata [online]. 2003 [cited 2003-10-02]. Available from World Wide Web : <<http://www.ukoln.ac.uk/metadata/>>.
- [19] International Federation of Library Associations and Institutions. Digital Libraries : Metadata Resources [online]. 2002 [cited 2003-10-02]. Available from World Wide Web : <<http://www.ifla.org/II/metadata.htm>>.
- [20] Ward, Jewel. foar.net [online]. 2002 [cited 2003-10-02]. Available from World Wide Web : <<http://foar.net/research/>>.

```

- <record>
- <header>
  <identifier>oai:arXiv.org:acc-phys/9411001</identifier>
  <timestamp>2003-02-05</timestamp>
  <setSpec>physics:acc-phys</setSpec>
  <setSpec>physics:physics</setSpec>
</header>
- <metadata>
- <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc"
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
  instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
  http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:title>Symplectic Computation of Lyapunov Exponents</dc:title>
  <dc:creator>Habib, Salman</dc:creator>
  <dc:creator>Ryne, Robert D.</dc:creator>
  <dc:subject>Accelerator Physics</dc:subject>
  <dc:description>..</dc:description>
  <dc:description>Comment: 12 pages, uuencoded PostScript (figures included)</dc:description>
  <dc:date>1994-10-31</dc:date>
  <dc:type>text</dc:type>
  <dc:identifier>http://arXiv.org/abs/acc-phys/9411001</dc:identifier>
</oai_dc:dc>
</metadata>
</record>
    
```

図1 OAI_DC Record 例

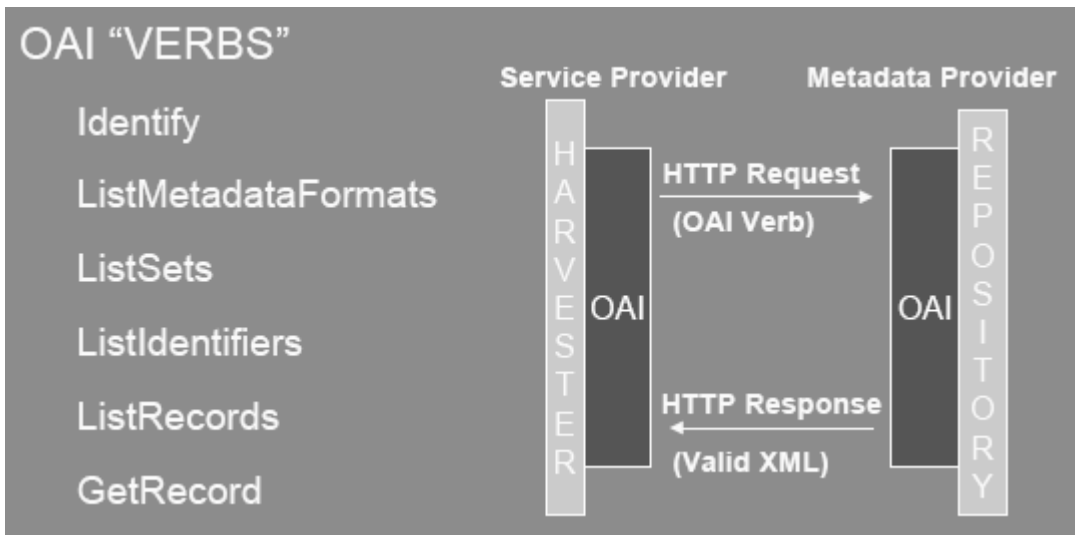


図2 OAI_PMH の仕組み