

OAI-PMH をめぐる動向

はじめに

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) は, OAI (Open Archives Initiative)⁽¹⁾が策定したメタデータ収集(メタデータ・ハーベスティング)のためのプロトコル(規約)である。2002年6月に第2版⁽²⁾⁽³⁾が公表されて以来, OAI-PMHを採用する機関, プロジェクトの数は上昇の一途をたどっている⁽⁴⁾。未だに米国情報標準化機構(National Information Standards Organization: NISO), 国際標準化機構(International Organization for Standardization: ISO)等の公式の規格とはなっていないものの, 事実上の標準としての地位を確立している。

本稿では, まず OAI および OAI-PMH の概要, プロトコルとしての特質について述べる。次いで, 国内外における適用事例を紹介し, 最後に, OAI-PMH をめぐる今後の課題と展望について俯瞰することとしたい。

1. OAI とは

OAI は, メタデータ収集を通じて多様なリポジトリ(電子情報庫)間の相互運用を促進することを使命とした国際的な運動である。

OAI の歴史は, 1999年10月に米国のサンタフェで開催された会議にまで遡ることができる⁽⁵⁾。当時, その数を増しつつあったeプリントアーカイブ(電子論文保管庫)の相互運用性の確立を目的として開催されたこの会議において, メタデータの収集を通じて複数アーカイブの相互運用を図るという基本的な枠組みが合意された。

その後, 2000年8月には, 電子図書館連合(Digital Library Federation: DLF)とネットワーク情報連合(Coalition of Networked Information: CNI)が支援を表明し, その活動範囲もeプリントアーカイブから各種電子情報コンテンツのリポジトリへと拡大されることとなる。そして2001年1月には OAI-PMH と呼ばれるメタデータ収集のためのプロトコル第1版が制定され, さらに翌2002年の6月には第2版が発表されている。

2. OAI-PMH とは

2.1 プロトコルの概要

OAI-PMH の基本的な枠組みは「データプロバイダ」と「サービスプロバイダ」と呼ばれる2種類の参加者によって形作られる。データプロバイダは, 各種電子情報を蓄積したサーバを維持し, OAI-PMH によりメタデータを開示する。一方, サービスプロバイダは,

OAI-PMH を使用してデータプロバイダが提供するリポジトリからメタデータを収集し, それに基づき各種の付加価値サービスを提供する。

データプロバイダが維持する「リポジトリ」は, OAI-PMH に定義された要求(リクエスト)に対して応答することのできるサーバである。一方, サービスプロバイダがメタデータ収集に使用する「ハーベスタ」は, OAI-PMH の要求を発行するクライアント側の応用ソフトウェアということになる。

続いて, プロトコルに従ってやりとりされるメタデータに関する概念と定義について見ていくことにする。

まず「アイテム」であるが, これはリポジトリの構成要素のひとつであり, あるひとつのリソースに関するメタデータを複数のフォーマットで蓄積する概念的な入れ物(コンテナ)であると定義されている。アイテムにはリポジトリのなかで一意になる識別子が付与される。

アイテムと対になる概念として「レコード」がある。レコードは, あるひとつのフォーマットで表現されたメタデータのことである。レコードは OAI-PMH の要求に対して, XML でコード化されて返戻される。

OAI-PMH では, 複数のメタデータ・フォーマットによるレコードの送信が認められているが, 最低限の相互運用性を保証するために, 限定子の付かないダブリン・コア(シンプル・ダブリン・コア)での送信が必須の要件となっている。

リポジトリには, 複数のアイテムをグループ化するために「セット」という概念を導入することができる。例えば, 主題, リソース種別(図書, 論文, 教材等)あるいは作成機関などのセットが考えられる。ただし, セットは必須の要素ではなく, あくまでオプションとして設定することができる。ハーベスタは, こうしたセットや日付を利用して, リポジトリの中から選択的にメタデータを収集することが可能になるのである。

OAI-PMH で使用される要求は以下の6つのみである。

- Identify (リポジトリに関する情報を取得する)
- ListMetadataFormats (リポジトリにおける利用可能なメタデータ・フォーマットの一覧を取得する)
- ListSets (リポジトリのセット構造を取得する)
- ListIdentifiers (リポジトリからレコード中のヘッダー情報のみを取得する)
- ListRecords (リポジトリから条件に合致するレコード全てを取得する)
- GetRecord (リポジトリから個々のレコードを取得)

する)

このうち, Identify, ListMetadataFormats, ListSets の 3 つは, リポジトリに関する情報を入手するための要求である。残りの 3 つの要求が, 実際にリポジトリからメタデータを収集する際に利用される。ほとんどの要求には引数を指定することができる。

2.2 プロトコルの特徴

OAI-PMH のプロトコルとしての特徴は, その簡潔性につける。可能な限り多くのデータプロバイダが OAI-PMH に基づく相互運用性の枠組みに参加できるように, 意識的に限定された機能しか組み込まれていない。

また, 既存のウェブ技術標準をそのまま活かし, できるだけ実装者の負荷を軽減させようという意図が伺える。例えば, 要求は全て HTTP の GET もしくは POST を使って送信する。それに対する応答には XML を使用し, 文字コードには UTF-8 を採用している。Z39.50 (ISO23950) のような規格と比較すると, まさに「敷居の低い」プロトコルと言えよう。

さらに 2003 年 10 月には, データプロバイダの負担を一層軽減するための仕組みである静的リポジトリ・ゲートウェイ (Static Repository Gateway) の仕様が公開されている⁽⁶⁾。

2.3 OAI-PMH の適用事例

2.3.1 関連プロジェクト

(1) FAIR (Focus on Access to Institutional Resources): FAIR は英国の JISC (CA1501 参照) の助成金によるプログラムであり, 2002 年 1 月に開始された。OAI-PMH を通じて学術機関の知的資産を配信し, 新たなサービスを創出することをめざしている。合計 14 のプロジェクト (参加機関数 50) により構成される⁽⁷⁾。

(2) NSDL (National Science Digital Library): NSDL は全米科学財団 (National Science Foundation: NSF) が助成するプロジェクトであり, 科学に関連する多様なデジタル・コンテンツを提供する電子図書館の構築をめざしている。今後 5 年間で, 数百万の利用者を対象として, 数千万の電子リソースを提供する予定である。NSDL のシステム・アーキテクチャにおいて OAI-PMH は不可欠の役割を演じている⁽⁸⁾。

(3) メロン財団のイニシアチブ

2001 年 8 月, メロン財団から OAI-PMH に基づく多様なサービスを開発するための研究費が 7 機関に配分された。助成金の総額は, 150 万ドルに達する。各プロジェクトは以下の 3 点の課題に取り組んでいる⁽⁹⁾。

- ・複数機関, 複数分野にまたがる広範囲なメタデータに基づくポータル・サービスの設計
- ・アーカイブや特殊コレクションからのメタデータ収集
- ・特定のトピックに基づく, 様々なフォーマットの資料に関するメタデータ収集

2.3.2 データプロバイダ

OAI-PMH に準拠したデータプロバイダの例については, OAI のホームページを参照されたい。2003 年 10 月 27 日現在, 121 のリポジトリが登録されている⁽¹⁰⁾。

2.3.3 サービスプロバイダ

一方, サービスプロバイダについても OAI のホームページにリストが掲載されている⁽¹¹⁾。以下に代表的な事例を挙げる。

(1) ARC: オールド・ドミニオン大学が開発した複数リポジトリの横断検索システム⁽¹²⁾。

(2) my.OAI: OAI-PMH 準拠のデータベースを統合検索するための多機能サーチエンジン⁽¹³⁾。

(3) NDLTD OAI Union Catalog: 電子学位論文ネットワーク (Networked Digital Library of Theses and Dissertations: NDLTD) の OAI 版総合目録⁽¹⁴⁾。

(4) OAIster: 上記メロン財団のイニシアチブのプロジェクトのひとつ。イリノイ大学が開発したハーベスタを使用して, 203 の機関からメタデータの収集を行い, 170 万件を超えるレコードの検索サービスを提供している (2003 年 10 月 1 日現在)⁽¹⁵⁾。

(5) SCIRUS: エルゼビア・サイエンス社が開発した, 学術文献に特化したサーチエンジン。OAI-PMH によって収集されたメタデータを含む⁽¹⁶⁾。

2.3.4 国内の事例

千葉大学附属図書館では, 学内で生産されるさまざまな電子的な研究成果を蓄積し, 保存し, 発信するための「千葉大学学術情報リポジトリ」の構築が進められている。このリポジトリに蓄えられたコンテンツの視認性をさらに高めるために, OAI-PMH のリポジトリ機能が実装された。

一方, 国立情報学研究所では, 大学等からの情報発信を支援することを主たる目的として, 平成 14 年度からメタデータ・データベース共同構築事業を開始している。国立情報学研究所は, OAI-PMH に準拠したハーベスタを開発し, 「千葉大学学術情報リポジトリ」から定期的にメタデータを収集し, メタデータ・データベースに格納している。

OAI の枠組みの中で両機関の役割をとらえるなら,

千葉大学がデータプロバイダ，国立情報学研究所がサービスプロバイダとして位置づけられよう。

3 今後の展望と課題

3.1 OAI-PMH をめぐる展望

以上，OAI-PMH を適用した国内外の事例を紹介してきたが，このプロトコルが持つ潜在力は，これらの適用事例以外にも効力を発揮するに違いない。例えば，図書館が提供する OPAC に OAI-PMH を実装すれば，分散統合型の総合目録を形成することも夢ではない。また，各機関で構築が進んでいる貴重書データベースや古文書データベースなどの電子図書館に適用することによって，さまざまな付加価値を持ったポータル（統合的情報窓口サービス）を開発することも可能となる。

3.2 今後の課題

今のところ，日本において OAI-PMH のリポジトリとしての機能を備えたデータベースは，千葉大学附属図書館の学術情報リポジトリと国立情報学研究所のメタデータ・データベースの2つのみである。OAI-PMH が持つ潜在的な可能性を最大限に発揮させ，さまざまな付加価値を備えたサービスを開発するには，このプロトコルを世の中に広め，その実装を促すための広報・普及活動が求められる。

また，新規に OAI 準拠のリポジトリを構築したり，既存のデータベース等を OAI 準拠にするための支援ソフトウェアの開発も期待される。海外では既にオープン・ソース化された無償のソフトウェアが多数存在する。日本においても，こうしたツール類の開発および無償頒布が望まれる。

おわりに

昨今，さまざまな「電子図書館」プロジェクトの旗の下に，多種多様な電子情報資源が日々生産され，インターネット上に蓄積されている。しかしながら，これらの情報資源は現状では孤立した存在であると言ってよい。利用者は個々の情報資源の提供場所を探し出し，そこを訪問し，それぞれ異なるインターフェイスを介して情報にアクセスすることを強いられている。

貴重な電子情報資源をインターネットの大海のなかの「孤島」に終わらせないためにも，「孤島」間の相互利用性を確立する枠組みの構築が求められている。本稿で紹介した OAI-PMH はそのための鍵となる要素技術である。多様な可能性を秘めたこのプロトコルのなお一層の浸透が望まれる。

（千葉大学附属図書館：尾城孝一^{おじろこういち}）

- (1) Open Archives Initiative. (online), available from < <http://www.openarchives.org/>>, (accessed 2003-10-27).
- (2) Open Archives Initiative Protocol for Metadata Harvesting [Protocol Version 2.0]. (online), available from <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>, (accessed 2003-10-27).
- (3) OAI-PMH2.0 日本語訳. (online), available from < <http://www.nii.ac.jp/metadata/oai-pmh2.0/>>, (accessed 2003-10-27).
- (4) Lagoze, Carl et al. The making of the Open Archives Initiative Protocol for Metadata Harvesting. Library Hi Tech. 21(2), 2003, 118-128.
- (5) Van de Sompel, Herbert et al. The Santa Fe Convention of the Open Archives Initiative. D-Lib Magazine. 6(2), 2000. (online), available from <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>>, (accessed 2003-10-27).
- (6) Specification for an OAI Static Repository and an OAI Static Repository Gateway. (online), available from <<http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>>, (accessed 2003-10-27).
- (7) FAIR. (online), available from <http://www.jisc.ac.uk/index.cfm?name=programme_fair>, (accessed 2003-10-27).
- (8) NSDL. (online), available from <<http://nsdl.org/>>, (accessed 2003-10-27).
- (9) Waters, Donald J. The Metadata Harvesting Initiative of the Mellon Foundation. ARL Bimonthly Report. 217, 2001. (online), available from <<http://www.arl.org/newsltr/217/waters.html>>, (accessed 2003-10-27).
- (10) Registered Data Providers. (online), available from <<http://www.openarchives.org/Register/BrowseSites.pl>>, (accessed 2003-10-27).
- (11) Registered Service Providers. (online), available from

- <<http://www.openarchives.org/service/listproviders.html>>, (accessed 2003-10-27)
- (12) ARC. (online), available from <<http://arc.cs.odu.edu/>>, (accessed 2003-10-27).
- (13) my.OAI. (online), available from <<http://www.myoai.com/>>, (accessed 2003-10-27).
- (14) NDLTD OAI Union Catalog. (online), available from <<http://rocky.dlib.vt.edu/~etdunion/cgi-bin/index.pl>>, (accessed 2003-10-27).
- (15) OAIster. (online), available from <<http://oaister.umd.umich.edu/o/oaister/>>, (accessed 2003-10-27).
- (16) SCIRUS. (online), available from <<http://www.scirus.com/>>, (accessed 2003-10-27).

尾城孝一. OAI-PMH をめぐる動向. カレントアウェアネス. 278, 2003.12, 12-14
本稿の著作権は国立国会図書館に帰属する。