

オープン・アーカイブ・イニシアチブ
メタデータ・ハーベスティング・プロトコル
OAI-PMHの概要

Jewel Ward <jewelw@lanl.gov>

慶應義塾大学訪問研究員

ライブラリシステム研究会セミナー

三田キャンパス 2003-06-17

謝辭

- JCDL 2001/2002: OAI-PMH Introduction
 - Hussein Suleman (then at Virginia Tech)
- JCDL 2003: Introduction to the OAI-PMH
 - Timothy W. Cole (UIUC)
 - William H. Mischo (UIUC)
 - Thomas Habing (UIUC)

謝辭

- JCDL 2003: Advanced Overview of Version 2.0 of the OAI-PMH
 - Michael L. Nelson (Old Dominion University)
 - Herbert Van de Sompel (LANL)
 - Simeon Warner (Cornell University)
- Digital Library Federation (DLF) Spring Forum 2003
 - "The OAI Static Repository: a file-based approach to exposing metadata via the OAI-PMH."
 - Herbert Van de Sompel (LANL)
 - This research was conducted by Patrick Hochstenbach (LANL), Henry Jerez (LANL) and Herbert Van de Sompel.

発表の概要

- 機関レポジトリ（ごく簡単に）
- OIA-PMHの背景と開発の経緯
- OAI-PMH の基本
- 新しい展開
- 関連サイト
- 質疑応答

機関レポジトリ

- 機関レポジトリ: 単数または複数の大学などの機関の知的成果を収集し保管するためのデジタルコレクション
- 現在の分散した方法に対して、機関の研究成果を集中できる

機関レポジトリ

- arXiv は機関レポジトリではありません！
(現在は @Cornell University)
- LANL の機関レポジトリ・プロジェクト
 - AISTI (the Alliance for Innovation in Scientific and Technical Information)
 - Within LANL

関連の動きとプロトコル

- オープン・アーカイブの動き
 - 研究成果や学術資料への一般のアクセス促進
 - 商業出版者による学術雑誌の価格高騰への対応
- OAメタデータ・ハーベスティング・プロトコル
 - Eプリント・レポジトリや電子図書館の増加
 - Eプリント/図書館のコミュニティが学術アーカイブの相互運用性を求めた

OAI-PMH 技術開発

- Gopher, FTP
- 総合目録
- Z39.50
- Kahn-Wilensky Framework
- Dienst Protocol
- Harvest
- UPS (Universal Preprint Services)
- OAI-PMH

OAI-PMHの概要

- OAI-PMHとは？
 - メタデータ・ハーベスティングを通じて、電子図書館の相互運用性のためにアプリケーションに依存しないし様を定義する規約
 - 様々なサービスや機能を可能にする、または促進する基礎を構築する規約
- OAI と OAI-PMHの違い

OAI-PMHの概要

OAI-PMHとは？ --以下のものではありません

- 検索サービス
- データベース
- オープン・アーカイブの情報システムそのもの
- メタデータの仕様定義
- ダブリンコア

データ/サービス 提供者

- データ提供者 (DPs) – レポジトリ
リソースとメタデータを所有し、OAIプロトコルによってメタデータを他のDPと共有しようとする組織や個人
- サービス提供者 (SPs) – ハーベスタ
検索など高レベルのサービスを利用者に提供するためDPからメタデータを収集する組織や個人
- DP=サーバ、SP=クライアント

OAI-PMH 要求セット

要求	機能
Identify	レポジトリに関する記述
ListMetadataFormats	サポートしているメタデータ形式
ListSets	定義されているセット
ListIdentifiers	OAI 固有識別子
ListRecords	N個のレコードのリスト
GetRecord	単独レコードのリスト

レポジトリ
に関する
メタデータ

収集要求

多くの要求は引数を持つ: dates, sets, ids, metadata formats
および resumption token (フローコントロール用)

OAI-PMH のメタデータ

- ダブリンコア・エレメントセット -- 必須
- さらに表現豊かなメタデータ形式 -- 強い推奨
- その他のメタデータ形式の例:
 - MARC
 - RFC-1807
 - Open Languages Archives Community Metadata Set
 - Electronic Theses and Dissertation Metadata Set

リソース – アイテム – レコード



← リソース

*set-membership is
item-level property*

item = identifier

ダビデ像に関する
あらゆるメタデータ

← アイテム

Dublin Core
metadata

MARC
metadata

SPECTRUM
metadata

← レコード

レコード = 識別子 + メタデータ + 日付スタンプ

固有識別子

- それぞれのアイテムは固有識別子が必須
- 識別子はURIシンタックスの後に記述
 - OAI 方式:
 - oai:<archiveID>:<recordID>
 - oai:etd.vt.edu:edt-1234567890
 - その他の方式も可：
 - http
 - handle

日付スタンプ

- 差分収集のために必須
- YYYY-MM-DDまたは YYYY-MM-DDThh:mm:ssZ (世界標準時)の方式
- メタデータ中の日付とは異なる
メタデータ収集にのみ使用
- メタデータ「レコード」の生成日付
 - 出版年月日ではない
 - アイテムの作成日付ではない

セット

- オプション (DPによる)
- setSpec & setNameが必須
それぞれのDP中のsetDescriptionは任意
- サブコレクションの収集のために階層構造を設けることも可 (“:” を使用)

OAI-PMH の要求と応答

OAI “要求”

Identify

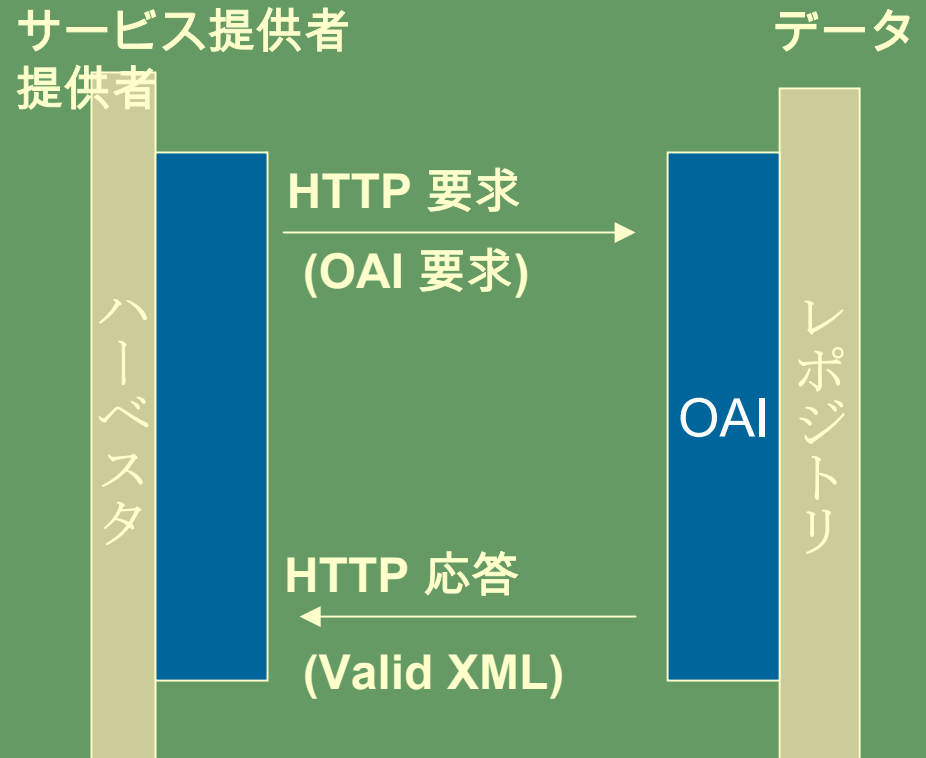
ListMetadataFormats

ListSets

ListIdentifiers

ListRecords

GetRecord



ベースURL + 要求の例

- <http://arXiv.org/oai2?verb=Identify>
- <http://arXiv.org/oai2?verb=ListSets>
- <http://arXiv.org/oai2?verb=ListMetadataFormats>
- http://arxiv.org/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc
- http://arxiv.org/oai2?verb=GetRecord&identifier=<recordID>&metadataPrefix=oai_dc
- http://arXiv.org/oai2?verb=ListRecords&metadataPrefix=oai_dc

応答の例

```
<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
  http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-06-03T20:13:50Z</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_dc"
    identifier="oai:arXiv.org:acc-
    phys/9411001">http://arXiv.org/oai2</request>
- <GetRecord>
..
</GetRecord>
</OAI-PMH>
```

レコードの例

```
- <record>
- <header>
  <identifier>oai:arXiv.org:acc-phys/9411001</identifier>
  <datestamp>2003-02-05</datestamp>
  <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>
```

コンテナエレメント (オプション)

- レポジトリレベル (set)
 - <Identify><description>
 - レポジトリに関する追加的情報
 - oai-identifier, eprints, friends, branding, other...
 - <ListSets><setDescription>
- メタデータレコードレベル
 - <about>
 - 「メタデータのメタデータ」 (レコードレベルの権利情報など)

Resumption Tokensほか

RT / フローコントロール / ロードバランス

- “resumptionToken” 不完全な応答のために使われる
- クライアント (SP) が、後で結果を受け取るため、サーバ (DP) へ示すトークンを受け取る

Resumption Tokensほか

RT / フローコントロール / ロードバランス

- RTオプション: completeListSize, cursor, and expiration date attributes
- Sequential tokenを区切るために、
from/until/metadataPrefix/set とレコード番号を使用
 - from!until!metadataPrefix!set!recordnumber
 - 2000-01-01!2001-01-01!oai_dc!All!100
- 自動終了機能とセッション管理

Resumption Tokensほか

RT / フローコントロール / ロードバランス 等冪性

- ハーベスタが応答を失ったり、大量の収集を一からやり直すのを避ける
- `resumptionToken` を利用して再度、要求を送付
- ハーベスタは最新の `resumptionToken` と以前のトークンの両方を受けつけることが必要

エラーの扱い

すべてのエラーメッセージは XML形式で送られる

- badVerb: ありえない要求
- badArgument: ありえないパラメータの値あるいは組み合わせ
- badResumptionToken, cannotDisseminateFormat, idDoesNotExist: パラメータの形式は正しいが、現在の条件に合わない
- noRecordsMatch, noMetadataFormats, noSetHierarchy: 応答すべきものがない

エラーメッセージの例

```
<?xml version="1.0" encoding="UTF-8" ?>  
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/  
  http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">  
<responseDate>2003-06-03T20:32:53Z</responseDate>  
<request>http://arXiv.org/oai2</request>  
<error code="badArgument">Verb 'ListRecords', argument  
  'metadataPrefix' required but not supplied.</error>  
</OAI-PMH>
```

OAI-PMH 静的レポジトリ

動機づけ

- OAI-PMH は障壁の低いプロトコルである
- OAI-PMH はデータ提供者に簡便さをもたらす

OAI-PMHの起源となった Santa Fe Convention
におけるバイアス

OAI-PMH 静的レポジトリ

動機づけ

実装はいつも簡単というわけではない

- 技術の不足
- コレクションが小さすぎて投資に足らない
- データベースのセキュリティへの配慮
- ISP がサード・パーティのソフトウェアを許可しない
- Cf. OLAC, union catalogue, LoCKSS

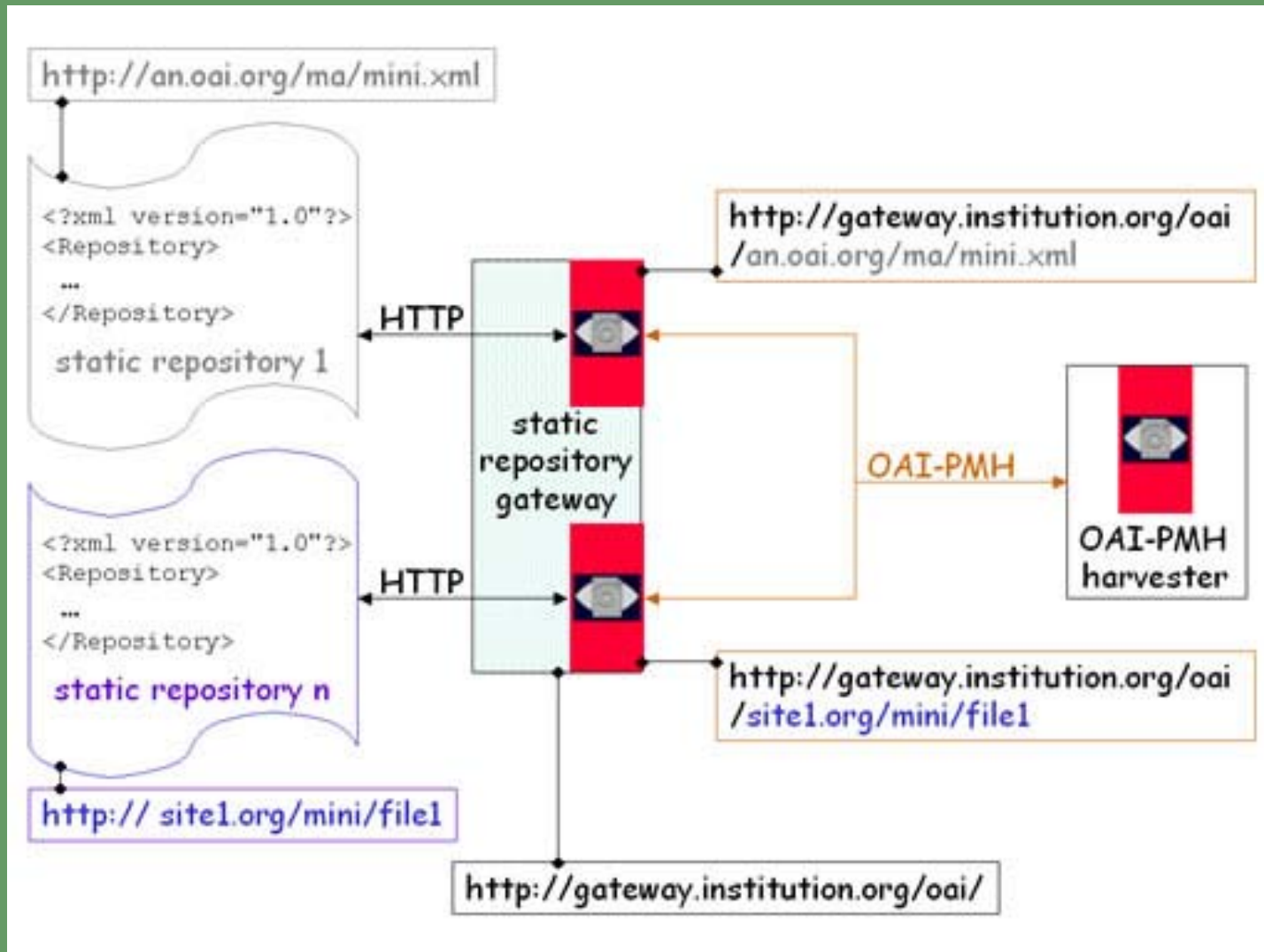
OAI-PMH 静的レポジトリ

そこで、、、

OAI-PMHを通じて、メタデータコレクションを

共有するための障壁を、さらに下げるための研究を開始！

OAI-PMH 静的レポジトリ



権利に関する工夫

- 権利に関する検討:
 - リソース
 - メタデータ
- Creative Commons (CC)におけるフレームワーク
- JISC/OAI/CCの共同研究
JISC: “Joint Information Systems Committee”
(RoMEOを含む)

関連サイト

- Institutional Repositories Partnering with Faculty to Enhance Scholarly Communication
 - <http://www.dlib.org/dlib/november02/johnson/11johnson.html>
- SPARC Institutional Repository Checklist & Resource Guide
 - http://www.arl.org/sparc/IR/IR_Guide.html

関連サイト

- Open Archives Initiative
 - <http://www.openarchives.org>
- OAI Metadata Harvesting Protocol
 - <http://www.openarchives.org/OAI/openarchivesprotocol.htm>
- OAI-PMH Tools Index
 - <http://www.openarchives.org/tools/index.html>
- Virginia Tech DLRL OAI Projects
 - <http://www.dlib.vt.edu/projects/OAI/>
- Repository Explorer
 - http://purl.org/net/oai_explorer
- ARC Cross-Archive Search Service
 - <http://arc.cs.odu.edu/>

関連サイト

- ARC Cross-Archive Search Service
 - <http://arc.cs.odu.edu/>
- OAI-PMH Static Repository
 - Registration
 - <http://libtest.lanl.gov/registry.html>
 - Example Repository
 - http://libtest.lanl.gov/cgi-bin/gateway.cgi/lib-www.lanl.gov/%7Ehochsten/desktop.xml?verb=ListRecords&metadataPrefix=oai_dc
 - Specification
 - <http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>

関連サイト

- Creative Commons
 - <http://www.creativecommons.org/>
- JISC
 - <http://www.jisc.ac.uk/>
- Dspace
 - <http://dspace.org/news/dspace-news.html>
- E-Prints DL-in-a-box
 - <http://www.eprints.org>
- Greenstone Digital Library
 - <http://www.greenstone.org/english/home.html>

関連サイト

- NDLTD
 - <http://www.ndltd.org>
- XML Schema Validator
 - <http://www.w3.org/2001/03/webdata/xsv>
- Dublin Core Metadata Initiative
 - <http://www.dublincore.org>
- XML Tools at W3C
 - <http://www.w3.org/XML/#software>

ご質問は？