ポータルの最新技術応用

OAI-PMHによるメタデータ活用

名古屋大学情報連携基盤センター
山本哲也

話の順番

- *OAI-PMHの概要*
- プロトコルの詳細
  - HTTP
  - XML
  - verb
- *使い道の例*
  - *名古屋大学では*

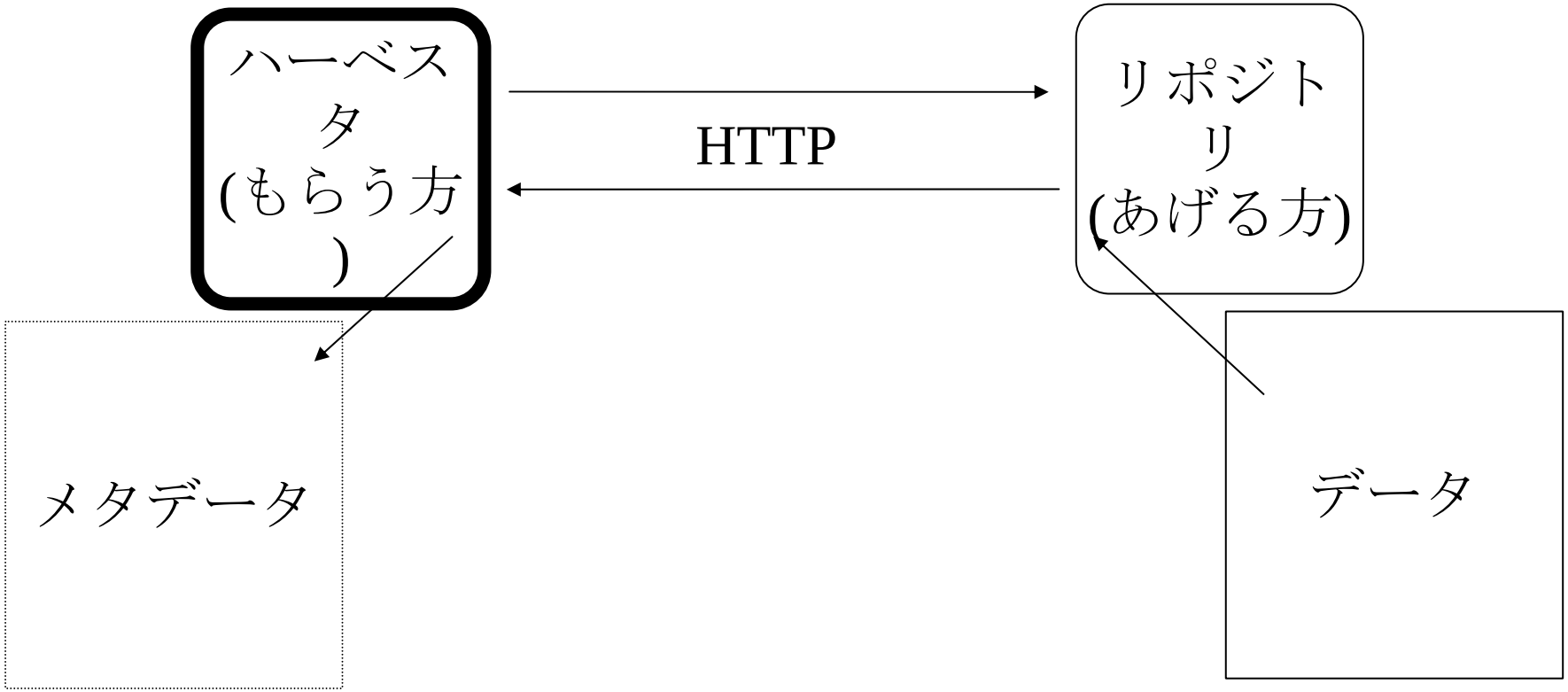
OAI-PMH概要

- 「メタデータ」の交換プロトコル
- Open Archives Initiative によって開発が進められている
- 仕様書 http://www.....
- 翻訳 http://www....
- アプリケーションに依存しない・相互使用可能
- 世界中にこれをサポートした公開リポジトリが存在する[TODO]

【repository】
[名][C]
1 容器(container)。
2 保管場所、たんす、押入れ；倉庫；博物館；商店；
　地下納骨所；（知識の）宝庫；もの知り、事情通。
3 秘密を打ち明けられる人。

（ジーニアス英和辞典　第3版）

イメージ

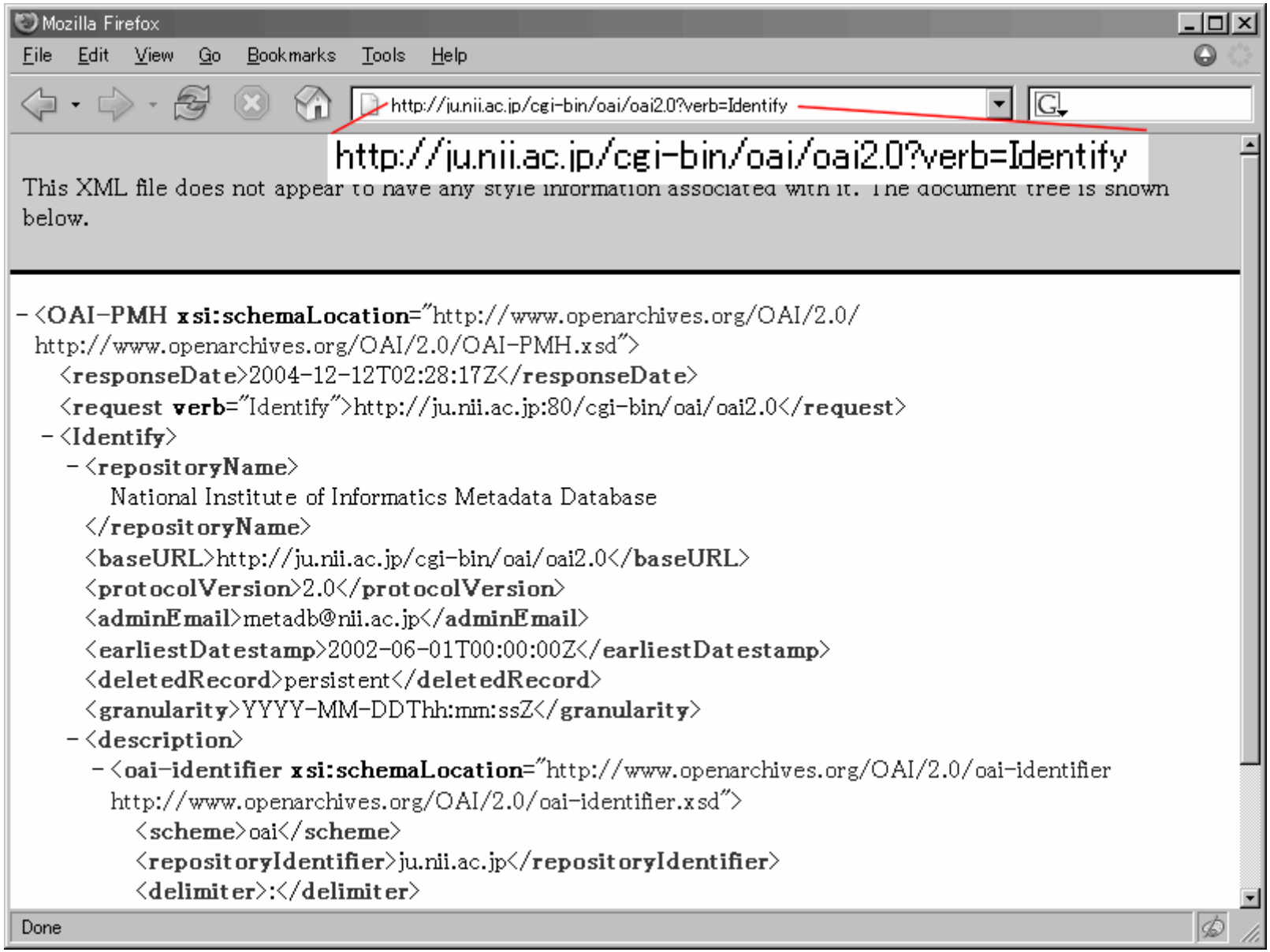


プロトコル概要

- HTTPに乗って情報を受け渡す
  (http://...../.....?verb=Identify)
- サーバーは、HTMLでなくてXMLを送信する
  (Content-Type: text/xml)
- verb引数が、要求する機能をあらわす
  (6種類)

Identify(あなたは何者)

- サーバーの基本情報を得る
  - リポジトリの名前
  - OAI-PMHのバージョン
  - 管理者情報


Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://ju.nii.ac.jp/cgi-bin/oai/oai2.0?verb=Identify
http://ju.nii.ac.jp/cgi-bin/oai/oai2.0?verb=Identify
This XML file does not appear to have any style information associated with it. The document tree is shown below.
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2004-12-12T02:28:17Z</responseDate>
<request verb="Identify">http://ju.nii.ac.jp:80/cgi-bin/oai/oai2.0</request>
-<Identify>
-<repositoryName>
National Institute of Informatics Metadata Database
</repositoryName>
<baseURL>http://ju.nii.ac.jp/cgi-bin/oai/oai2.0</baseURL>
<protocolVersion>2.0</protocolVersion>
<adminEmail>metadb@nii.ac.jp</adminEmail>
<earliestDatestamp>2002-06-01T00:00:00Z</earliestDatestamp>
<deletedRecord>persistent</deletedRecord>
<granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
-<description>
-<oai-identifier xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
<scheme>oai</scheme>
<repositoryIdentifier>ju.nii.ac.jp</repositoryIdentifier>
<delimiter>:</delimiter>
Done

verb=Identify

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
 http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2004-12-12T02:28:17Z</responseDate>
  <request verb="Identify">http://ju.nii.ac.jp:80/cgi-bin/oai/oai2.0</request>
  <Identify>
    <repositoryName>
      National Institute of Informatics Metadata Database
    </repositoryName>
    <baseURL>http://ju.nii.ac.jp/cgi-bin/oai/oai2.0</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>metadb@nii.ac.jp</adminEmail>
    <earliestDatestamp>2002-06-01T00:00:00Z</earliestDatestamp>
    <deletedRecord>persistent</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
    <description>
      <oai-identifier xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-
identifier http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
        <scheme>oai</scheme>
        <repositoryIdentifier>ju.nii.ac.jp</repositoryIdentifier>
        <delimiter>:</delimiter>
        <sampleIdentifier>oai:ju.nii.ac.jp:00000001</sampleIdentifier>
      </oai-identifier>
    </description>
  </Identify>
</OAI-PMH>
```

GetRecord(1件取得)

- 指定したIDのメタデータを1件取得する
- IDは、あらかじめ分かっていること
  (後述のListIdentifiresで知ることができる)
- metadataPrefixを指定して、「見え方」を指定できる(後述)

verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ju.nii.ac.jp:00017296

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
 http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2004-12-12T03:49:13Z</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_dc"
identifier="oai:ju.nii.ac.jp:00017296">http://ju.nii.ac.jp:80/cgi-
bin/oai/oai2.0</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:ju.nii.ac.jp:00017296</identifier>
        <datestamp>2002-12-02T07:14:30Z</datestamp>
        <setSpec>0097</setSpec>
        <setSpec>natural_science</setSpec>
      </header>
      <metadata>
         (省略)
      </metadata>
    </record>
  </GetRecord>
</OAI-PMH>
```

<matadata>の中身が、oai_dc形式で表現されている

```
<metadata>
  <oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title>伊藤圭介文庫</dc:title>
    <dc:title>イトウ ケイスケ ブンコ</dc:title>
    <dc:title>Itō keisuke bunko</dc:title>
    <dc:title>Ito Keisuke Collection</dc:title>
    <dc:creator>名古屋大学附属図書館</dc:creator>
    <dc:creator>ナゴヤ ダイガク フゾク トショカン</dc:creator>
    <dc:creator>Nagoya Daigaku Fuzoku Toshokan</dc:creator>
    <dc:subject>NDC:470</dc:subject>
    <dc:subject>伊藤圭介</dc:subject>
    <dc:subject>植物学</dc:subject>
    <dc:description>日本における近代植物学の祖といわれる伊藤圭介の稿本(手書き本)188冊を
電子化したもの。</dc:description>
    <dc:publisher>名古屋大学附属図書館</dc:publisher>
    <dc:publisher>ナゴヤ ダイガク フゾク トショカン</dc:publisher>
    <dc:publisher>Nagoya Daigaku Fuzoku Toshokan</dc:publisher>
    <dc:type>デジタルミュージアム</dc:type>
    <dc:format>text/html</dc:format>
    <dc:identifier>
      http://www.nul.nagoya-u.ac.jp/db/keisuke/index.html
    </dc:identifier>
    <dc:language>jpn</dc:language>
  </oai_dc:dc>
</metadata>
```

<matadata>の中身が、junii形式で表現されている

```
<metadata>
  <meta xsi:schemaLocation="http://ju.nii.ac.jp/oai http://ju.nii.ac.jp/oai/junii.xsd">
    <code>00017296</code>
    <userid>A07070</userid>
    <fano>FA002407</fano>
    <adate>20021129</adate>
    <udate>20021202</udate>
    <institution>名古屋大学</institution>
    <title>伊藤圭介文庫</title>
    <title.transcription>イトウ ケイスケ ブンコ</title.transcription>
    <title.alternative>Ito Keisuke Collection</title.alternative>
    <creator xsi:type="NC">
      <ahdng>名古屋大学附属図書館</ahdng>
      <aid>DA02287075</aid>
    </creator>
    <creator.transcription xsi:type="NC">
      <ahdng>ナゴヤ ダイガク フゾク トショカン</ahdng>
      <aid>DA02287075</aid>
    </creator.transcription>
    <subject xsi:type="NDC">470</subject>
    <subject>伊藤圭介</subject>
    <subject>植物学</subject>
    <description>日本における近代植物学の祖といわれる伊藤圭介の稿本(手書き本)188冊を電
子化したもの。</description>
    <publisher xsi:type="NC">
      <ahdng>名古屋大学附属図書館</ahdng>
      <aid>DA02287075</aid>
    </publisher>
    <publisher.transcription xsi:type="NC">
      <ahdng>ナゴヤ ダイガク フゾク トショカン</ahdng>
      <aid>DA02287075</aid>
```

ListIdentifires(ID一覧を見る)

- GetRecord等で指定できるIDの一覧を取得できる
- メタデータそのものの一覧ではない
  (これは、後述のListRecordsで行う)
- 日付の範囲指定ができる
- Set(部分集合)の指定ができる
- リストが大きすぎるときには、ページ分けのような仕組みで数回に分けて取得する

verb=ListIdentifiers&metadataPrefix=oai_dc&set=natural_science

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2004-12-12T04:51:19Z</responseDate>
  <request verb="ListIdentifiers" metadataPrefix="oai_dc"
set="natural_science">http://ju.nii.ac.jp:80/cgi-bin/oai/oai2.0</request>
  <ListIdentifiers>
    <header>
      <identifier>oai:ju.nii.ac.jp:00003718</identifier>
      <datestamp>2002-06-14T10:51:47Z</datestamp>
      <setSpec>natural_science</setSpec>
    </header>
    <header> ... </header>
    <header> ... </header>
    <header> ... </header>
    <header> ... </header>
    ...
    <resumptionToken completeListSize="4150" cursor="100" expirationDate="2004-12-
13T04:51:23Z">1102827083296521100</resumptionToken>
  </ListIdentifiers>
</OAI-PMH>
```

※次のページは、
verb=ListIdentifiers&resumptionToken=1102827083296521100
で表示できる

ListMetadataFormats(見せ方一覧)

- リポジトリの中のあるアイテムがどのフォーマットに対応しているかを一覧で取得する
- oai_dcというフォーマットは、全リポジトリ・全アイテムで必須
- アイテムによって、対応フォーマット(metadataPrefix)の集合は違うときがある
- いろいろな独自フォーマットを用いるコミュニティがある

verb=ListMetadataFormats&identifier=oai:ju.nii.ac.jp:00017296

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2004-12-12T04:55:38Z</responseDate>
  <request verb="ListMetadataFormats"
identifier="oai:ju.nii.ac.jp:00017296">http://ju.nii.ac.jp:80/cgi-
bin/oai/oai2.0</request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>

<metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>junii</metadataPrefix>
      <schema>http://ju.nii.ac.jp/oai/junii.xsd</schema>
      <metadataNamespace>http://ju.nii.ac.jp/oai</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

ListRecords(全件取得)

- まさにハーベスティング(刈り取り)
- 対象リポジトリの全アイテムの各メタデータをすべて取得
  (日付範囲指定、Set指定も可)
- 巨大なリストは数回にわけて取得

verb=ListRecords&metadataPrefix=oai_dc&set=natural_science

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2004-12-12T04:48:04Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc"
set="natural_science">http://ju.nii.ac.jp:80/cgi-bin/oai/oai2.0</request>
  <ListRecords>
    <record>
      <header> ... </header>
      <metadata> ... </metadata>
    </record>
    <record>
      <header> ... </header>
      <metadata> ... </metadata>
    </record>
    <record>
      <header> ... </header>
      <metadata> ... </metadata>
    </record>
      ...
    <resumptionToken completeListSize="4150" cursor="100" expirationDate="2004-12-
13T04:48:08Z">1102826888296181100</resumptionToken>
  </ListRecords>
</OAI-PMH>
```

ListSets(セット一覧)

- どういう部分集合の指定ができるかを一覧表示
  - 十進分類?
  - 機関?
  - その他?
- 子セット、孫セットの表現も可能(必要かな?)

verb=ListSets

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2004-12-12T04:56:24Z</responseDate>
  <request verb="ListSets">http://ju.nii.ac.jp:80/cgi-bin/oai/oai2.0</request>
  <ListSets>
    <set>
      <setSpec>0001</setSpec>
      <setName>北海道大学</setName>
    </set>
    <set>
      <setSpec>0002</setSpec>
      <setName>小樽商科大学</setName>
    </set>
    ...
    <set>
      <setSpec>general_works</setSpec>
      <setName>General_Works</setName>
    </set>
    <set>
      <setSpec>philosophy</setSpec>
      <setName>Philosophy</setName>
    </set>
  </ListSets>
</OAI-PMH>
```

OAI-PMHの機能で「ない」もの

- データの「実体」のやりとり
  - あくまで「メタデータ」
- 検索
  - 日付範囲や部分集合の指定程度。文字列の検索はない
  - ハーベストした後の使い道次第
- データ操作
  - データの挿入、更新、削除の機能はない
  - リポジトリのデータ管理は、OAI-PMHの仕事ではない

必要なスキル

- HTTPの扱い
- XMLの扱い
- プログラミング？パッケージ利用？
  - フレームワーク利用

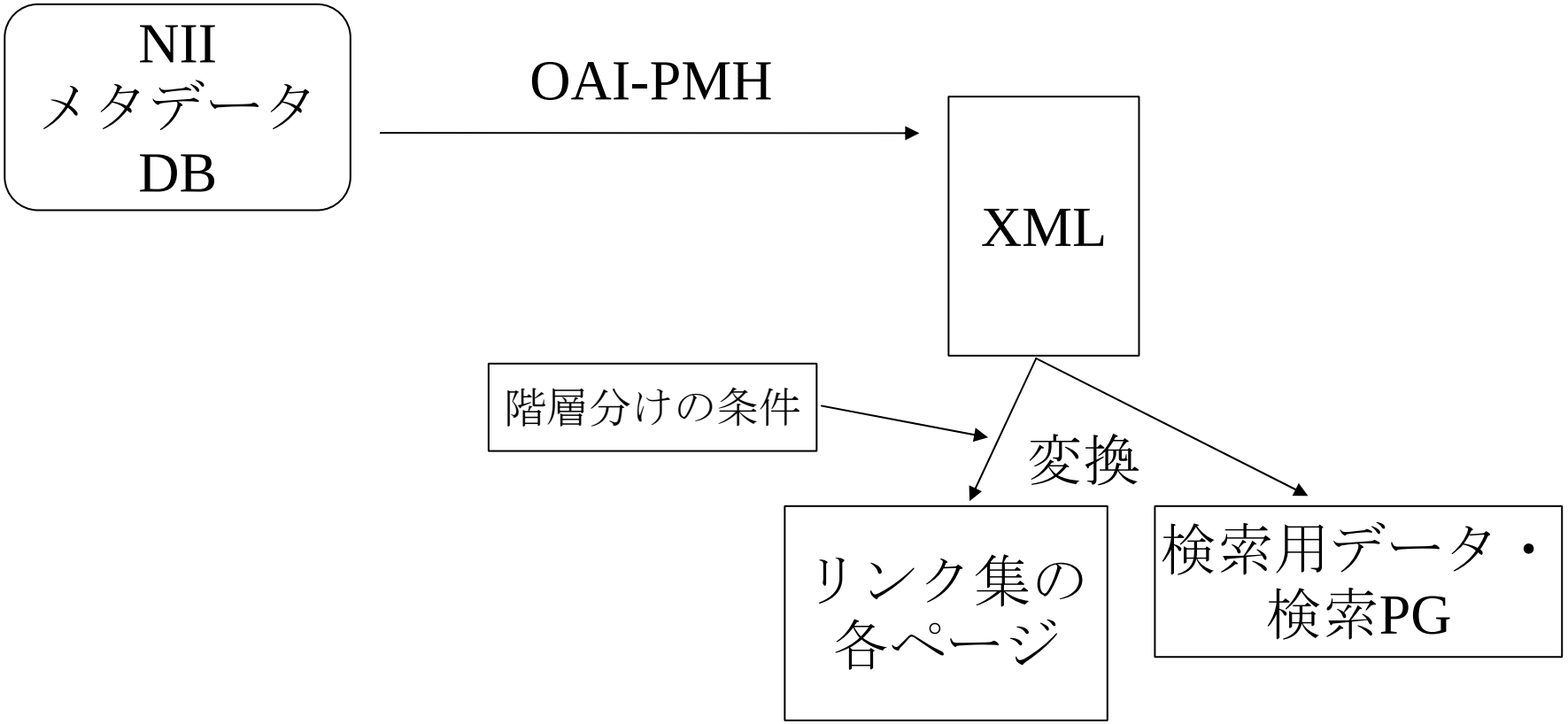
使い道の例（前提）

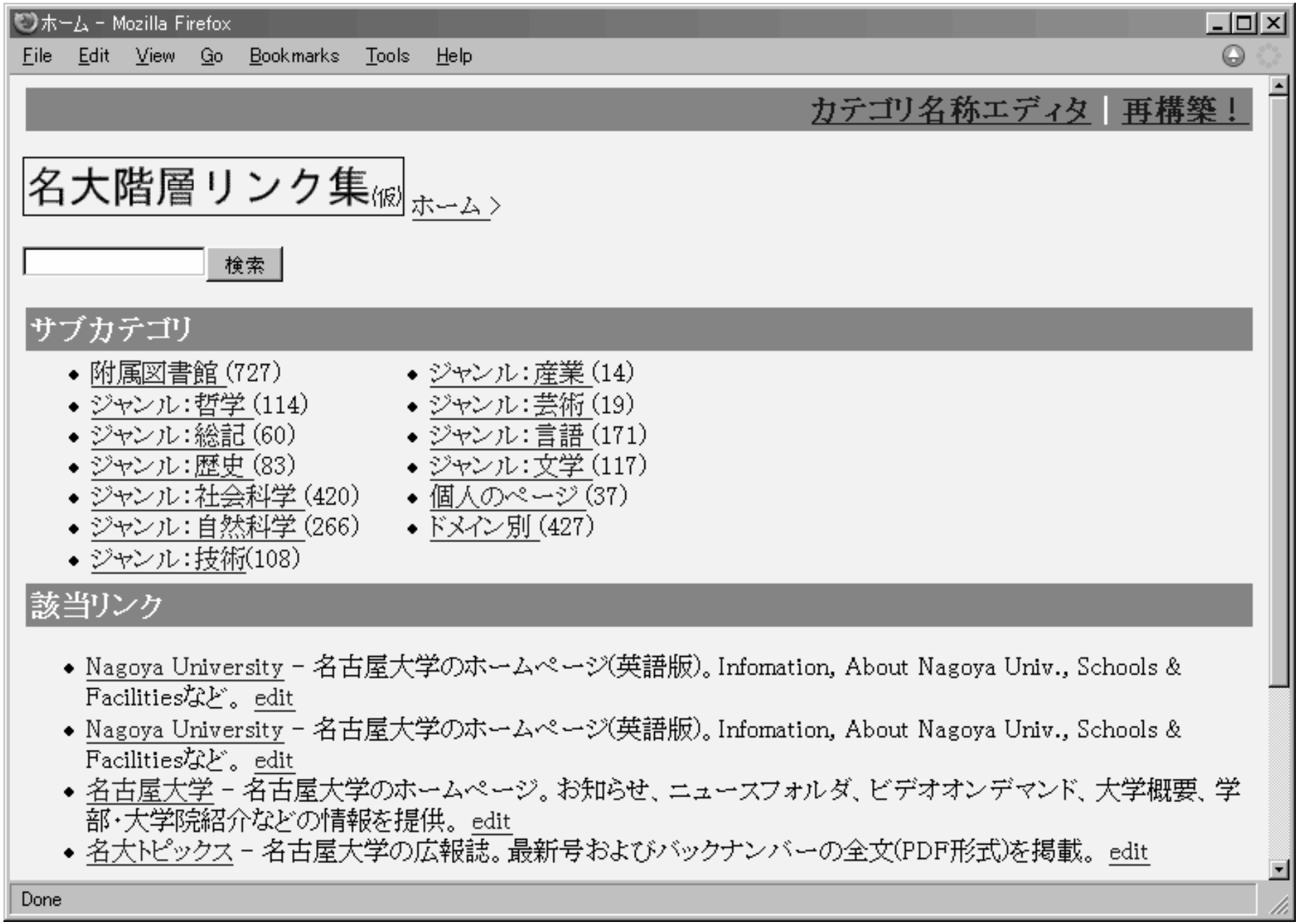
- 名古屋大学付属図書館で、メタデータ・データベース共同構築事業への参加という形でデータを登録していた(1,000件程度)
- このデータを、自前でも活用するような方法はないか？
- ここに登録したメタデータを、OAI-OMHでやりとりするサービスが始まったらしい

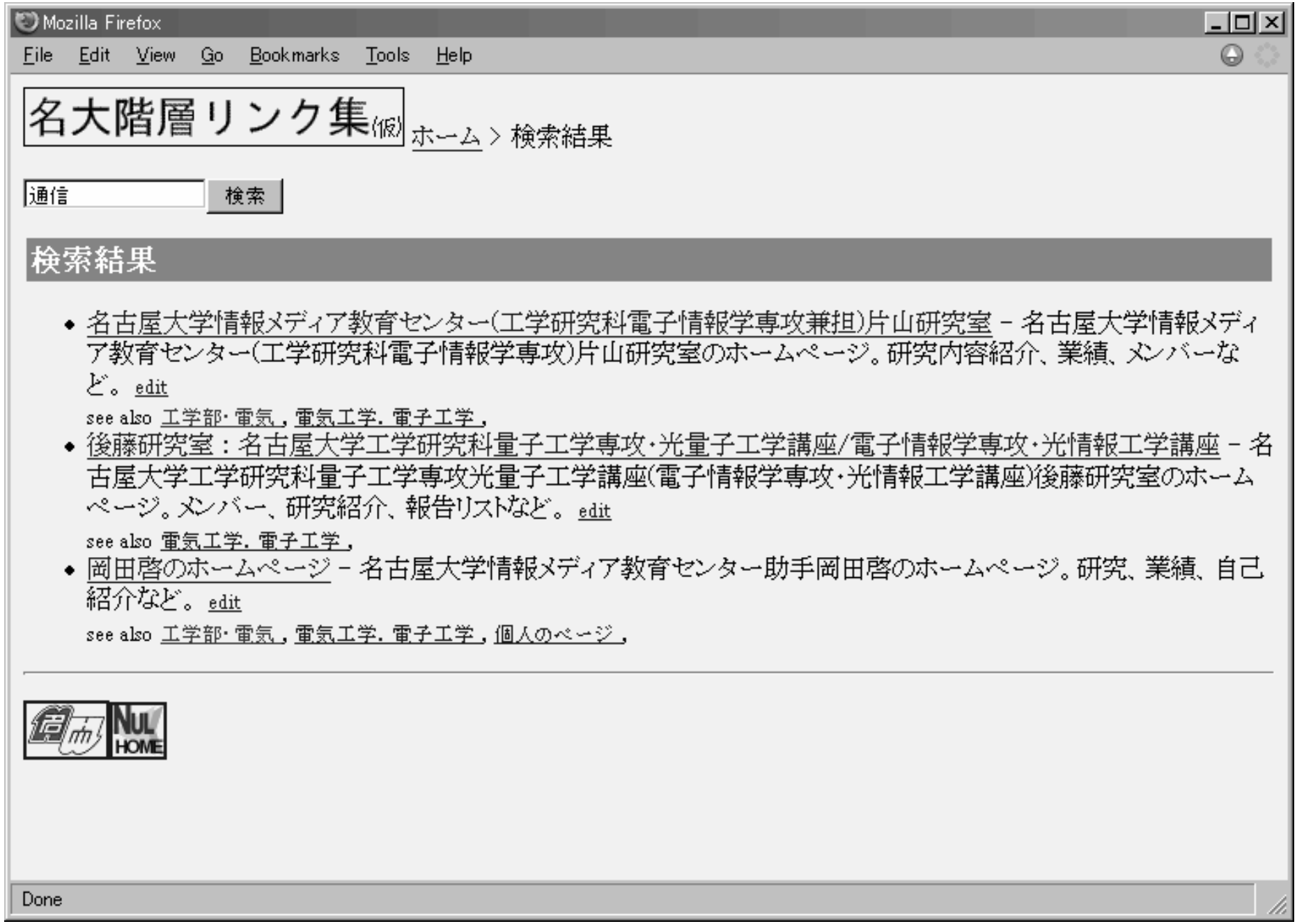
使い方の例（ある実装）

- 「階層リンク集」を作ってみる
  - 必要なデータを、OAI-PMHで「刈り取って」くる
  - メタデータを手がかりにカテゴリーの階層を作って、各データを投げ込む
  - ワード検索機能をつけてみる

データの流れ




ホーム - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
カテゴリ名称エディタ | 再構築！
名大階層リンク集(仮)
ホーム >
検索
サブカテゴリ
附属図書館 (727)
ジャンル：哲学 (114)
ジャンル：総記 (60)
ジャンル：歴史 (83)
ジャンル：社会科学 (420)
ジャンル：自然科学 (266)
ジャンル：技術(108)
ジャンル：産業 (14)
ジャンル：芸術 (19)
ジャンル：言語 (171)
ジャンル：文学 (117)
個人のページ (37)
ドメイン別 (427)
該当リンク
Nagoya University - 名古屋大学のホームページ(英語版)。Infomation, About Nagoya Univ., Schools & Facilitiesなど。 edit
Nagoya University - 名古屋大学のホームページ(英語版)。Infomation, About Nagoya Univ., Schools & Facilitiesなど。 edit
名古屋大学 - 名古屋大学のホームページ。お知らせ、ニュースフォルダ、ビデオオンデマンド、大学概要、学部・大学院紹介などの情報を提供。 edit
名大トピックス - 名古屋大学の広報誌。最新号およびバックナンバーの全文(PDF形式)を掲載。 edit
Done


Mozilla Firefox
File Edit View Go Bookmarks Tools Help
名大階層リンク集(仮)
ホーム > 検索結果
通信
検索
検索結果
名古屋大学情報メディア教育センター(工学研究科電子情報学専攻兼担)片山研究室 - 名古屋大学情報メディア教育センター(工学研究科電子情報学専攻)片山研究室のホームページ。研究内容紹介、業績、メンバーなど。 edit
see also 工学部・電気 , 電気工学．電子工学 ,
後藤研究室：名古屋大学工学研究科量子工学専攻・光量子工学講座/電子情報学専攻・光情報工学講座 - 名古屋大学工学研究科量子工学専攻光量子工学講座(電子情報学専攻・光情報工学講座)後藤研究室のホームページ。メンバー、研究紹介、報告リストなど。 edit
see also 電気工学．電子工学 ,
岡田啓のホームページ - 名古屋大学情報メディア教育センター助手岡田啓のホームページ。研究、業績、自己紹介など。 edit
see also 工学部・電気 , 電気工学．電子工学 , 個人のページ ,
NUL HOME
Done