

機関リポジトリの可視性の向上と メタデータ, OAI-PMH

平成24年度DRF新任担当者研修「メタデータ概論」「機関リポジトリの公開」の資料を元に再編集しています

三大学連携機関リポジトリ研修会 2014.8.5@お茶の水女子大学
千葉大学附属図書館亥鼻分館 武内 八重子

この講義の目的

- 研究成果の公開に、機関リポジトリを使うメリットを説明するための基礎知識を得ましょう
 - メタデータ（入力項目）設計の意味
 - OAI-PMHというデータ提供方法について
 - リポジトリ内部のメタデータと外部へ提供するメタデータ
 - oai_dcやjunii2のクロスウォーク設計について

この講義の内容

1 基本理念

- 1.1 何を考えてメタデータは設計されているか
- 1.2 OAI-PMH

2 メタデータの設計方法

- 2.1 機関リポジトリのメタデータ
- 2.2 内部のメタデータ設計
- 2.3 外部へのメタデータ提供設定

3 可視性を高めるための登録

3

メタデータとは

- **資料の内容表示や検索などを容易にするために付けられる、データのこと。**電子データの書誌データとっていれば間違いない。PDFや画像などの電子データは、ファイルを開くまでは中身がどのようなものであるか不明であるし、検索などもできない。そこで、タイトルや作成者、作成年などのメタデータを付けることで、検索などができるようにする必要がある。こうして付けられるのがメタデータである。リポジトリへの登録は、すなわちメタデータの付与であるといっても過言ではない。どのような項目が必要で、どのように記述するかは各リポジトリで自由であるが、ダブリン・コアに準拠した項目は持っておくこと。

DRFPedia より「メタデータ meta-data」

<http://drf.lib.hokudai.ac.jp/drf/index.php?Keywords>

4

1.1 何を考えてメタデータは設計されているか

5

1 基本理念

機関リポジトリとは？（復習です）

クリフォード・リンチ（Clifford Lynch, 2003）

大学がその構成員に提供する、大学とその構成員が創造したデジタル資料の管理や発信を行うために、大学がそのコミュニティの構成員に提供する一連のサービス

||

機関で産出された研究成果を世界中からアクセスされやすくする

||

可視性（visibility:ビジビリティ）の向上

6

可視性を向上させるには

- Googleなどの検索エンジンが、すべてのコンテンツをクロール（読取→索引化）してくれるようリポジトリを設定する
- CiNiiなどの外部サービスが、機関リポジトリのデータをもとにサービス提供してくれる。このようなサービスをしてもらえるように、
 1. リポジトリを設定する
 2. 外部サービスに登録する

7

1.2 OAI-PMH

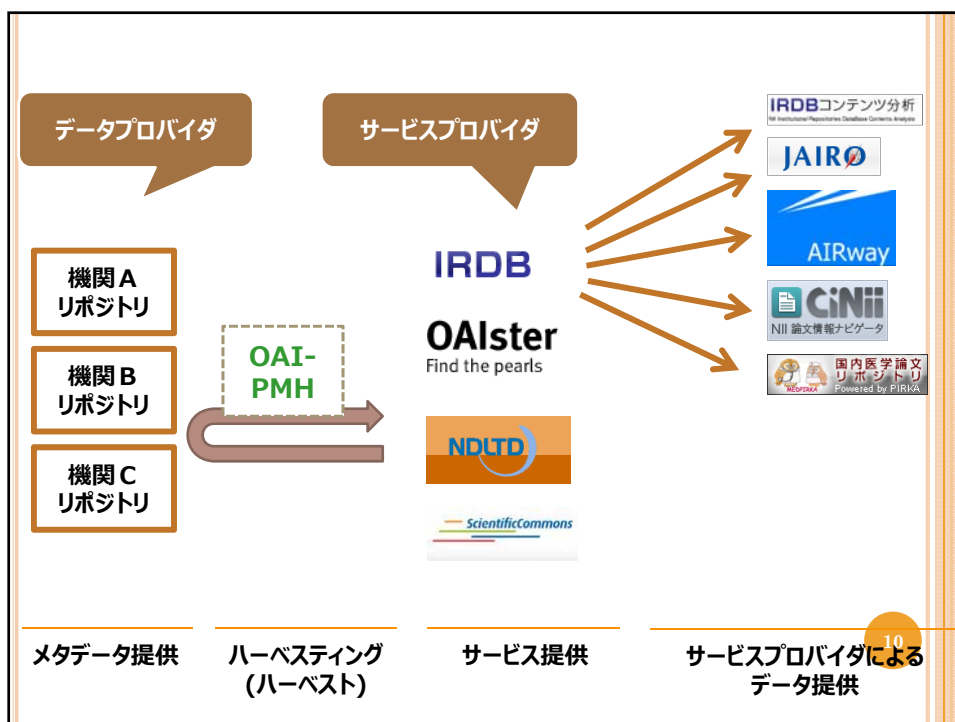
8

1 基本理念

OAI-PMHとは？

- 複数のリポジトリのメタデータを収集し、それに基づいたサービスを提供するために開発された、データ提供・収集用のプロトコル（通信手段）
- HTTP通信を用いていて、各リポジトリソフトウェアでサポートされている
- **O**pen **A**rchives **I**nitiative **P**rotocol for **M**etadata **H**arvesting
- <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- 日本語訳 <http://www.nii.ac.jp/irp/archive/translation/oai-pmh2.0/>

9



10

- 「リポジトリを公開する」ということ

||

OAI-PMHで外部にメタデータを提供すること

- 機関リポジトリに収録されたメタデータを、外部サービスがOAI-PMHで自動的に収集して（刈り取って）くれて、サービス提供してくれる

ハーベスティング

11

プロトコル概要

- OAI-PMH要求
 - サービスプロバイダは、URLの形で要求を送る
`http://xx.xxxx.ac.jp/xxx?verb=xxx`
 - Verb引数が要求（6種類）を表す
- メタデータ提供
 - データプロバイダは、XML形式でメタデータを返信する
- HTTPで情報のやり取りを行う
- コンテンツ本体（本文PDF）はやり取りに含まれない
 - やり取りするのは、コンテンツの入手先（=URL）や書誌情報など（=メタデータ）だけ

赤字部分が「ベースURL」

12

要求と値

http://xxx.xxxx.ac.jp/xxxx?verb=[要求]
&[引数]=[引数の値]
の形で記述

要求	説明	指定できる値
Identify	どんなリポジトリですか？	
ListMetadataFormats	どんな形式でデータを出力できますか？	i
ListSets	どんな集合がありますか？	r
GetRecord	データを1つください	m* , i*
ListIdentifiers	IDリストをください	m* , s, f, u, r
ListRecords	データを全部ください	m* , s, f, u, r

引数	説明
metadataPrefix	このデータ形式で
set	この集合のデータを
from	この日付のデータから
until	この日付のデータまで
identifier	このIDのデータを（データごとのユニークな識別子）
resumptionToken	次項のデータを

*のついた引数は
必須

13

この XML ファイルにはスタイル情報が関連づけられていないようです。以下にこのキ... を表示します。

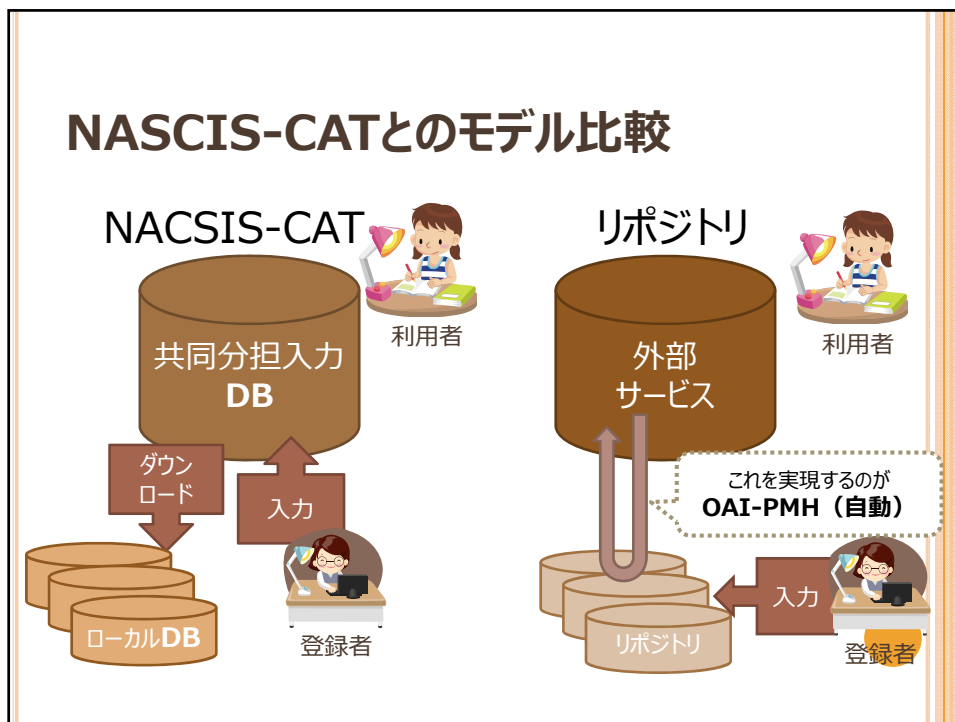
```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?xml:lang="en" ?xml:base="http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2014-08-01T07:28:06Z</responseDate>
<request verb="ListRecords" metadataPrefix="junii2" from="2014-04-01" until="2014-05-31" ?>
<ListRecords>
<record>
<header>
<identifier>oai:mitizane.ll.chiba-u.jp:00020144</identifier>
<timestamp>2014-04-25T06:07:31Z</timestamp>
<setSpec>setA</setSpec>
</header>
<metadata>
<junii2 xsi:schemaLocation="http://ndb.nii.ac.jp/oai/http://itd.nii.ac.jp/junii2" ?>
<title>大学改革と大学図書館</title>
<alternative>ダイガクカイカクト ダイガクトショカン</altern...>
<creator>土屋 俊</creator>
<creator>ツチャ、シュ</creator>
<NDC>0177</NDC>
<NDC>010</NDC>
<NDC>377</NDC>
<NDC>0073</NDC>
<subject>大学図書館</subject>
<subject>管理・運営</subject>
<description>
現在の大学は、一方では高等教育と先端科学技術研究開発という社会的役割の履行を強める社会からの改革の要請に答えると同時に、デジタルネットワークの社会インフラ化によって加速される社会の情報化という動向にも対応することを迫られている。(中略)大学図書館は、この大学改革と社会インフラの情報化というふたつの圧力によって示される方向性を自覚して、その展開を図らなければならないと考えられる。...文部省図書館職員短期研修講演(2000年11月1日東京大学)テキスト(pp.19-23)
</description>
<publisher>千葉大学文学部行動科学科認知情報科学講座</publisher>
<publisher>チバ ダイガクランガクブ コウドウ カガクカ ニンシ ヨウホウ カガク コウザ</publisher>
<date>2000-11-01</date>
<NIItype>Others</NIItype>
<type>その他</type>
<format>text</format>
</metadata>
</record>
</ListRecords>
</request>

```

http://mitizane.ll.chiba-u.jp/cgi-bin/oai/oai2.0
ベースURL
?verb=ListRecords
&metadataPrefix=junii2
junii2形式でデータをがばってください
&from=2014-04-01&until=2014-05-31
更新日が2014/4/1~5/31まで

NASCIS-CATとのモデル比較



2.1 機関リポジトリのメタデータ

16

2 メタデータ的设计方法

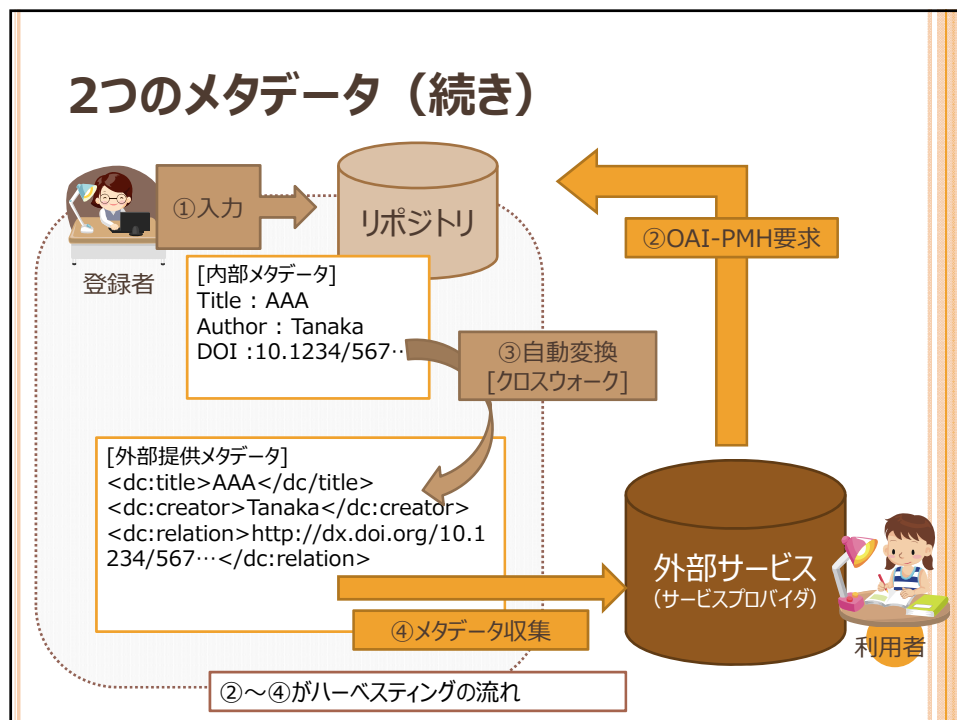
2つのメタデータ

- **内部メタデータ** (リポジトリシステムで使用するメタデータ)
 - 目録に相当する書誌情報や, システムがコンテンツを管理するためのデータなど
 - データの持ち方は使用するリポジトリシステムにより異なる

- **外部への提供メタデータ**
 - OAI-PMHなどで外部に提供するデータ
 - 要求の都度、内部メタデータを自動変換して出力
 - 提供先に応じていくつかの種類を設定できる

17

2つのメタデータ (続き)



2.2 内部メタデータ設計

19

2 メタデータの設計方法

内部メタデータ設計で考慮すべきこと

- 機関リポジトリはメタデータ項目、記述方法について、
目録のような **標準化された規則はない**

必要になるのは…

- **資料種別ごと**に明確に決定する
- 必要な情報を不足なく定義する
- 外部への**提供メタデータの形式**、**項目をあらかじめ想定**する

20

資料種別ごとに必要項目を考える

	雑誌論文	紀要論文	学位論文	会議発表	科研報告	授業教材
タイトル	●	●	●	●	●	●
著者	●	●	●	●	●	●
著者キーワード	●	●	●		●	
抄録・概要	●	●	●		●	
掲載誌情報	●	●				
学位授与機関・学位の種別 etc.			●			
会議名・開催地・日時				●		
課題番号・研究期間・種目etc.					●	
授業年度・授業科目etc.						●

内部メタデータの記述

- 書誌の記述方法について、NACSIS-CATコーディングマニュアルのような **統一規範はない**
- 外部メタデータ用のjunii2ガイドラインも参考に
 - <http://www.nii.ac.jp/irp/archive/system/junii2.html>
- NII「学術コンテンツ登録システム」の「システムデータ記述手引」>「データ項目」を参考にしている機関もある
 - <http://www.nii.ac.jp/nels/man/>

内部メタデータの例

YNU-R

dc.contributor.author 高橋, 知也
dc.contributor.author 小池, 高史
dc.contributor.author 安藤, 孝敏
dc.contributor.alternative Takahashi, Tomoya
dc.contributor.alternative Koike, Takashi
dc.contributor.alternative Ando, Takatoshi
dc.contributor.transcription タカハシ, トモヤ
dc.contributor.transcription コイケ, タカシ
dc.contributor.transcription アンドウ, タカトシ

TeaPot

creator 館取, 直子
creator 茂出木, 理子
creator.alternative ETORI, Naoko
creator.alternative MODEKI, Riko
creator.alternative えとり, なおこ
creator.alternative もでき, りこ

CURATOR

Creator 中村, 愛
Creator.Transcription ナカムラ, アイ
Creator.Alternative NAKAMURA, Ai
Creator 永田, 亜希子
Creator.Transcription ナガタ, アキコ
Creator.Alternative NAGATA, Akiko
Creator 山本, 利江
Creator.Transcription ヤマモト, トシエ
Creator.Alternative YAMAMOTO, Toshie

23

2.3

外部へのメタデータ提供設定

24

2 メタデータの設計方法

外部提供メタデータは標準準拠

- 外部提供メタデータには**標準化されたメタデータフォーマット仕様がある**

- サービスプロバイダは複数リポジトリのデータを収集するため

25

メタデータフォーマット

- OAI_DC

- OAI-PMHで必須, OAIsterなどで使用
- DRIVERガイドライン
<http://driver-support.eu/managers.html>
- http://www.openarchives.org/OAI/2.0/oai_dc.xsd
- <http://dublincore.org/documents/dces/>

Dublin Core :
インターネット上の情報資源の
メタデータを記述する世界標準。
OAI_DCはDCのデータ項目を使っている。

26

○ junii2

- JAIRO, CiNiiなどNIIで使用
- junii2ガイドライン
<http://www.nii.ac.jp/irp/archive/system/junii2.html>
- <http://irdb.nii.ac.jp/oai/junii2.xsd>

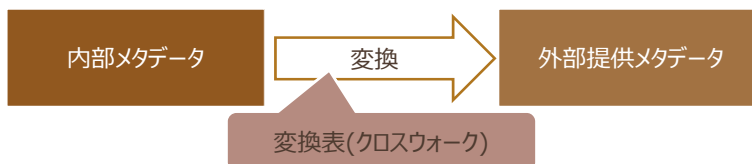
○ ETD-MS : an Interoperability Metadata Standard for Electronic Theses and Dissertations

- 学位論文用, NDLTDで使用
- <http://www.ndltd.org/standards/metadata/etdms/1-0/etdms.xsd>
- <http://www.ndltd.org/standards/metadata/>

27

外部提供メタデータへの変換

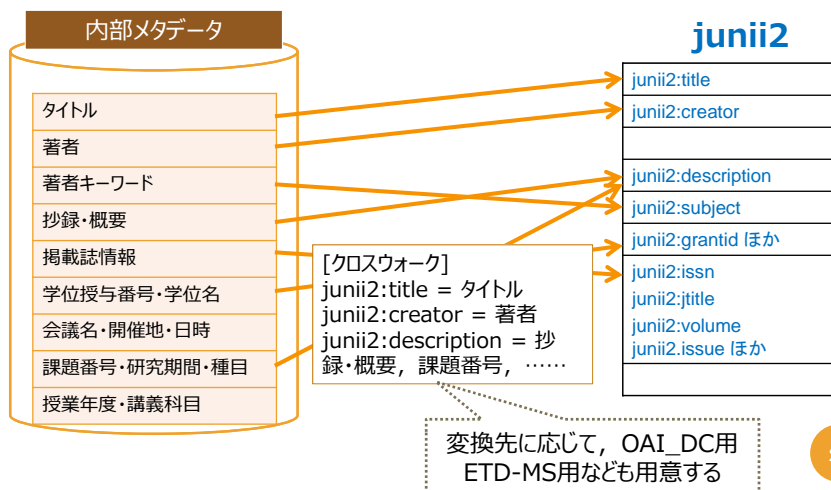
- OAI_DC, junii2, ETD-MSなどでメタデータを提供するために、変換表に基づいた変換が必要。



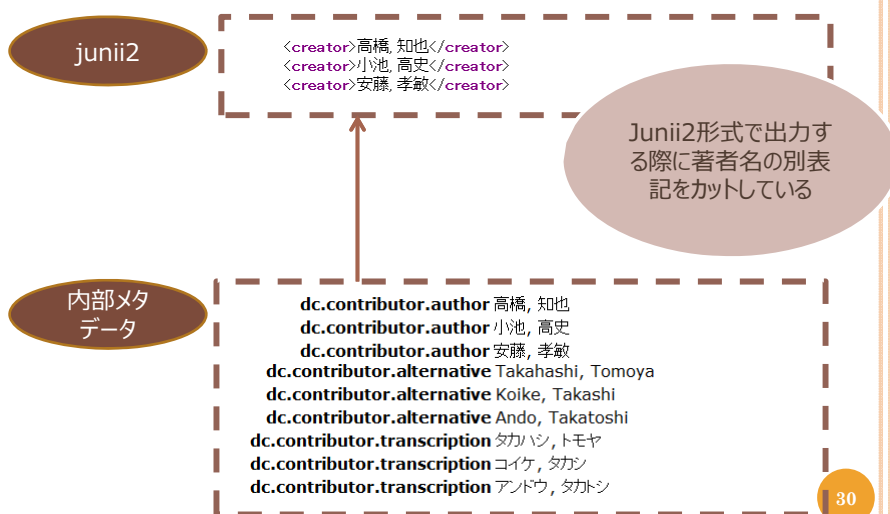
- 内部メタデータから、外部提供メタデータへの変換ができるように、両者のデータ項目間でマッピングしておく必要がある。
- マッピングを示した変換表がクロスウォーク。

28

メタデータ項目のマッピング



外部メタデータの例



メタデータ設計の考え方

○世界標準は顧慮せず，なおかつ顧慮する

○内部メタデータ設計は自由に！

- リポジトリ上で見せたい情報や，コンテンツを管理するためのデータを保持
- ハーベスティングしてほしいサービスに対応できるように，情報を保持

○外部への提供メタデータは標準準拠で！

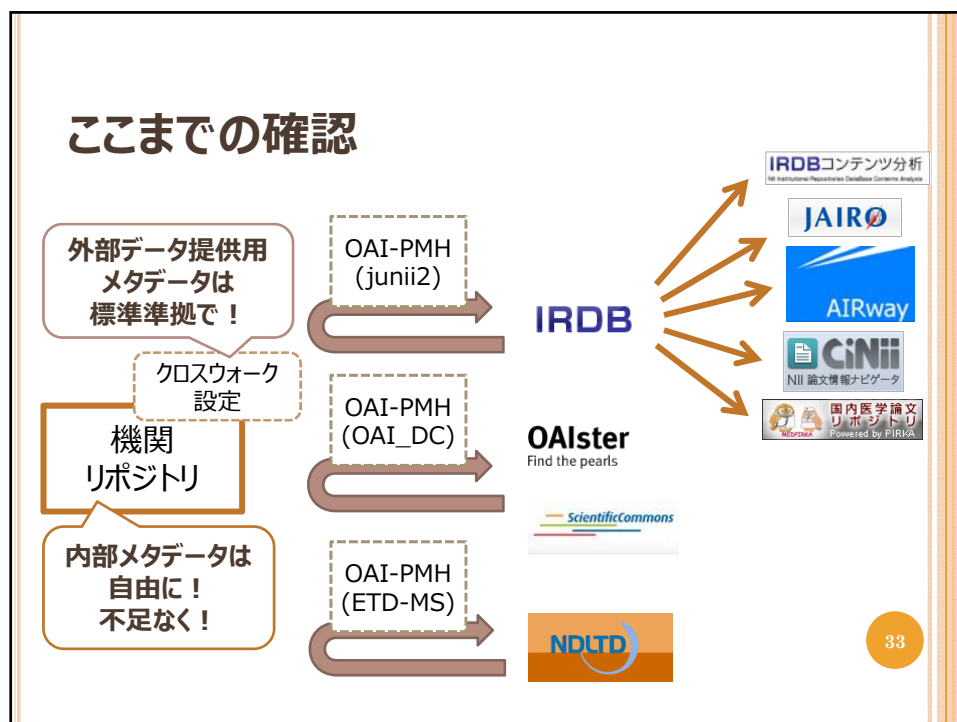
- OAI-PMHなどで外部へデータ提供するために標準準拠で！
- サービスプロバイダによっていろいろな形式（メタデータフォーマット）での出力が求められる
 - junii2, OAI_DC, ETD-MS...

31

3

可視性を高めるための登録


32



データプロバイダの一覧リストに登録する

国内外で維持管理されている機関リポジトリの一覧リストに登録する

- OAI Open Archives Initiative
 - メタデータ収集を通じて多様なリポジトリ間の相互運用の促進を目的とした国際的な取組み
- NIIの機関リポジトリ一覧
- OpenDOAR Directory of Open Access Repositories
 - SHERPAが運営
- ROAR Resitory of Open Access Repositories
 - サウサンプトン大学が運営



Registered Data Providers

This page lists registered OAI conforming repositories, registered through our [registration and validation page](#). Currently there are 2370 such repository table may be sorted either by the [Repository Name](#), the [base URL](#), or the [oai-identifier namespace](#) (if defined; used in the [oai-identifier](#) scheme). For repository you may view the registration record from the database, or alternatively, if your browser can render XML, you may issue an [identify request](#) to selected repository and receive the current XML response.

Service providers can get an XML formatted list of base URLs of registered data providers from <http://www.openarchives.org/Register/ListFriends>.

View	Repository Name	base URL (sort)	oai-identifier namespace-identifier used) (sort)
Record Identify	æ*æµá¼c «ãñšã- jëæá ±ããã, äã*	http://kamome.lib.ynu.ac.jp/dspace-oai/request	
Record Identify	Ochanomizu University Web Library - Institutional Repository (TeaPot)	http://teapot.lib.ocha.ac.jp/ocha-oai/request	
Record Identify	Chiba University Scholarly Repository	http://mitizane.ll.chiba-u.jp/cgi-bin/oai/oai2.0	mitizane.ll.chiba-u.jp
Record Identify	Yamaguchi University Navigator for Open access Collection and Archives	http://petit.lib.yamaguchi-u.ac.jp/info/lib/oai_repository/repository	
Record Identify	Tsukuba Repository (Tulips-R)	http://dspace.tulips.tsukuba.ac.jp/dspace-oai/request	
Record Identify	HamaMed-repository	http://hikumano.hama-med.ac.jp/dspace-oai/request	
Record Identify	Ryukoku University Institutional Repository	http://repo.lib.ryukoku.ac.jp/oai/request	

2014.7.31現在

サービスプロバイダの検索分析対象になる

国内外で機関リポジトリ（データプロバイダ）のメタデータ等を収集して、新たなサービスを展開しているサービスプロバイダがいる
 検索対象になれば、アクセス数が劇的に増加し、登録コンテンツが多くの人に利活用される

- IRDB/JAIRO/CiNii
- Google/Google Scholar
- OAIster
- NDLTD
- Westlaw Japan
- RePEc
- 各大学の業績データベース

→ 詳細は月刊DRF第19号などを参照

サービスプロバイダの検索対象であるなら…

- サービスプロバイダで使用するメタデータフォーマットのためのクロスワークを設定しておく必要がある
- 削除履歴を残す設定にする
 - 削除したレコードの情報が相手先に伝わって削除される
- リポジトリのURIが変わった時などには設定変更が必要
 - データプロバイダの一覧リストについても同様に登録変更が必要

37

まとめ

- 機関リポジトリは、OAI-PMHによって、登録したコンテンツのメタデータを外部サービスに拡散しやすいシステム
 ||
 研究成果の可視性を向上させる
- そのためには………をきちんとしましょう
 - 適切なメタデータ、クロスワークの設定
 - データプロバイダ一覧やサービスプロバイダへの登録

38