

機関リポジトリの メタデータ概論

平成22年度学術ポータル担当者研修

平成22年7月29日 <会場:名古屋大学>

北海道大学附属図書館学術システム課 野中雄司

今日のお話

- **1. 基本概念**
 - 1-1 **メタデータ設計の意味**
 - 何を考えてメタデータ(入力項目)設計すればよいか
 - 1-2 **OAI-PMH**
- **2. メタデータ設計**
 - 2-1 **内部メタデータ設計**
 - 2-2 **外部へのデータ提供設計**

1. 基本概念

1-1. メタデータ設計の意味

**何を考えてメタデータ(入力項目)
設計すればよいか**

機関リポジトリとは？

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**(クリフォード・リンチ(2003))

要するに

- 学内構成員に対するサービスである。

= (イコール)

- 機関内産出の研究成果を世界中からアクセスされやすいようにする
 - 可視性(ビジビリティ)のアップ

ビジビリティアップの基本

- 自動的にページをインデクシングしてくれる Googleなどの検索エンジンから見えるように (クローリングされるように) すること。
- 文献のデータを所定の様式で用意すると、そのデータを持って行ってきて、自サービスにデータ登録し、サービス展開してくれるものがある。(JAIRO等)

→ **今日はこちらの話**

1-2 OAI-PMH

OAI-PMHとは？

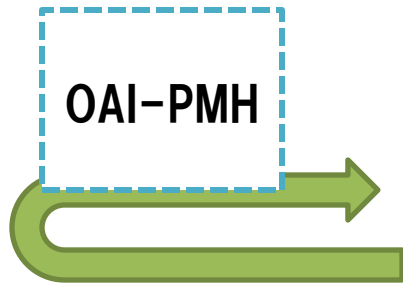
- **複数リポジトリのデータを収集し、それに基づいたサービスを提供するために開発されたデータ提供・収集用のプロトコル**
- **Open Archives Initiative によって開発**
 - 仕様書(<http://www.openarchives.org/OAI/openarchivesprotocol.html>)
 - 翻訳(<http://www.nii.ac.jp/irp/archive/translation/oai-pmh2.0/>)
- **HTTP通信を使用するため、アプリケーションに依存しない**

リポジトリから見たOAI-PMH

- 主にリポジトリ用の外部へのデータ提供を実現するためのもの
- **リポジトリを設置するということ
＝(イコール)**
「OAI-PMHで外部にデータ提供することである。」といっても過言ではない。
- 一つのサービスに提供するとその提供先が別のサービスにデータ提供し、広がっていくことも。

こんなイメージ

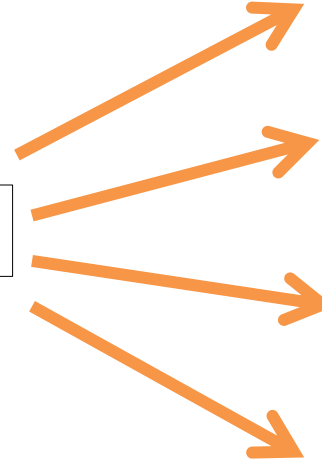
- 機関
リポジトリ
- 機関
リポジトリ
- 機関
リポジトリ



IRDBコンテンツ分析
NII Institutional Repositories DataBase Contents Analysis



SCIRUS
for scientific information only



プロトコル概要

データを提供
(機関リポジトリ)

データ
プロバイダ

①要求

`http://xx.xxxx.ac.jp/xxx?verb=xxx`

②取得(XMLで)

データを集め、それ
を使用したサービ
スを提供するもの
(JAIRO等)

サービス
プロバイダ

●HTTPで情報のやり取りを行う。

●要求

- `http://xx.xxxx.ac.jp/xxx?verb=xxx`
- Verb引数が要求する命令を表す(6種類)

●取得

- サーバ(リポジトリ)はXMLを返信

●コンテンツ本体(本文PDF)はやりとりしない。

- コンテンツの在りか(URL)や書誌情報などをやりとりする。

赤字部分を
「ベースURL」と言います

つまり

- 機関リポジトリに搭載されている文献データを他のサービスが(勝手に)持って行って(刈り取って)くれて、サービスしてくれる。

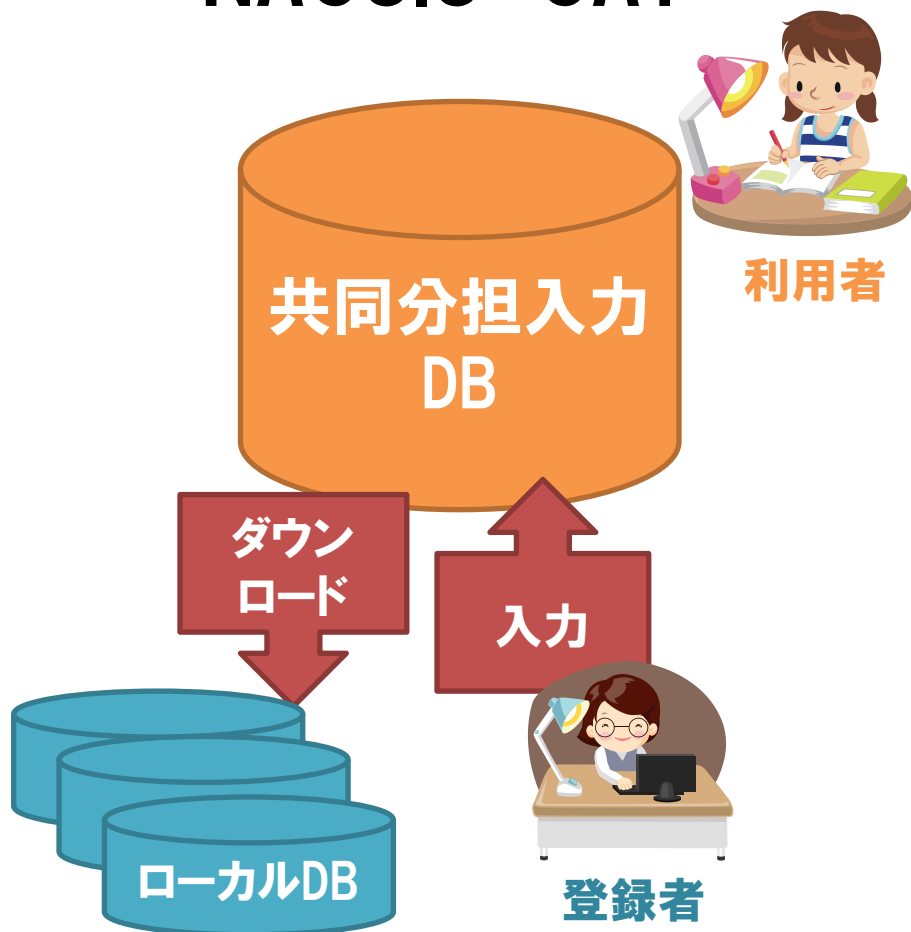
「ハーベスティング」といいます

例えば
JAIRO

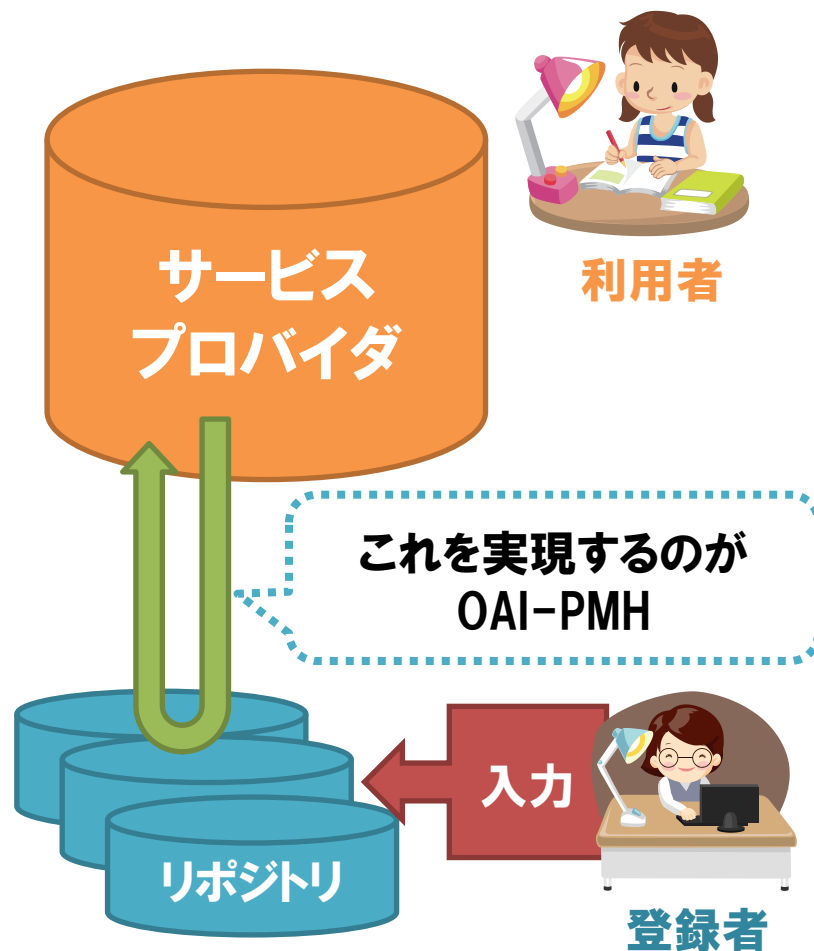
- 自分たちからデータを送付する必要がない。
- 変更, 削除記録も考慮してくれる。

モデル比較

NACSYS-CAT



OAI-PMH



コマンドとパラメタ

Verb	説明	指定可能パラメタ
Identify	どんなりポジトリですか？	
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

パラメタ	説明
m etadataPrefix	このデータ形式で
s et	この集合のデータを
f rom	この日時のデータから
u ntil	この日時のデータまで
i dentifier	このIDのデータを
r esumptionToken	次項のデータを

要求URL(コマンド)例

http://eprints.lib.hokudai.ac.jp/dspace-oai/request?

verb=ListRecords

データを全部ください

&metadataPrefix=junii2

junii2形式で

&from=2009-04-01

2009年4月1日から4月30日までに追加された(もしくは変更された)

&until=2009-04-30

&set=hdl_2115_20046

集合が [hdl_2115_20046] (内実は農学部の雑誌論文)の

実際にはどうなるのか？

北海道大学 附属図書館 | HUSCAP

Hokkaido University Collection of Scholarly and Academic Papers
Copyright(c) 2005 Hokkaido University Library, All Rights Reserved.

検索 表示言語: 日本語 一時的表示範囲: 全文献

フルテキスト
AIRway.pdf 360.69 kB Adobe PDF 見る/開く

タイトル: 機関リポジトリへのアクセス経路
その他のタイトル: Access path to institutional repositories

著者: 紙谷, 五月
野中, 雄司
杉田, 茂樹

キーワード: 機関リポジトリ
リンクリゾルバ
OAI-PMH
OpenURL
AIRway
institutional repository
link-resolver

発行日: 1-Dec-2008
出版者: 情報科学技術協会

```
<?xml version="1.0" encoding="UTF-8" ?>  
<header>  
  <metadata>  
    <junii2 xmlns:schemaLocation="http://rdnb.iiac.jp/...>  
      <title>機関リポジトリへのアクセス経路</title>  
      <alternative>Access path to institutional repositories</alternative>  
      <creator>紙谷, 五月</creator>  
      <creator>野中, 雄司</creator>  
      <creator>杉田, 茂樹</creator>  
      <subject>機関リポジトリ</subject>  
      <subject>リンクリゾルバ</subject>  
      <subject>OAI-PMH</subject>  
      <subject>OpenURL</subject>  
      <subject>AIRway</subject>  
      <subject>institutional repository</subject>  
      <subject>link-resolver</subject>  
      <NDC>007</NDC>  
      <description>特集 ファイン/ゼロリテ/向上</description>
```

リポジトリに
登録

OAI-PMH

JAIRO
Japanese Institutional Repositories Online

このアイテムのアクセス数: 22 回 (2009/07/07)
<http://jairo.nii.ac.jp/0003/00025399>

書誌情報
機関リポジトリへのアクセス経路
紙谷, 五月, 野中, 雄司, 杉田, 茂樹
情報科学技術協会 The Journal of Information Science and Technology
88 (12) , pp.610 - 614 , 2008-12-01 , 情報科学技術協会
ISSN0919-3801
NDC書誌ID:NDL-ANI0006587

本文を読む
<http://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/94975/1/AIRway.pdf>

JAIROがハー
ベストینگ
(週次, 月次等)

自動的にCiNiiに反映

CiNii
国立情報学研究所
論文情報データベース(CiNii)

機関リポジトリへのアクセス経路

機関リポジトリへのアクセス経路<特集>ファインダビリティ向上
Access path to institutional repositories<Special feature>Improving findability: Information access optimization

紙谷 五月
紙谷 五月
NONAKA Satsuki
AIRwayプロジェクト
AIRway Project, Hokkaido University Library

野中 雄司
野中 雄司
NONAKA Yuki
AIRwayプロジェクト
AIRway Project, Hokkaido University Library

杉田 茂樹
杉田 茂樹
SHIGITA Shigeki
AIRwayプロジェクト
AIRway Project, Hokkaido University Library

本文を読む/探す
CiNii PDF 国立情報学研究所 CiNii 本文PDF
Webcat Plus Webcat Plus 先行刊-所蔵情報
FullText Full-Text@Hokkaido Univ.
FullText Full-Text@Hokkaido Univ.

CiNiiからも検索可能に!
(本文(リポジトリ)へのリンク付)

2. メタデータ設計

機関リポジトリのメタデータ

デジタルコンテンツの多面的な特性を記述するもの

メタデータ構成	記述する情報
記述メタデータ	目録に相当する書誌情報
技術メタデータ	保存されているコンテンツを再生するために必要な技術情報
権利メタデータ	コンテンツの利用について規定する情報。利用者の義務や禁止・許可・制限事項等
保存メタデータ	コンテンツの作成履歴, 資料を受け入れた日付等の情報
管理メタデータ	コンテンツ収集・作成過程やコンテンツとメタデータの登録・更新情報

本表は以下を元に作成

<http://www.ndl.go.jp/jp/standards/da/index.html>

<http://www.ndl.go.jp/jp/standards/da/da.pdf>

<http://www12.ocn.ne.jp/~zuki/drf/repository.html>

機関リポジトリのメタデータ

おおきく2つの用途に分けられます

- システム内部で使用するメタデータ(内部メタデータ)
- 外部に提供するメタデータ
 - OAI-PMH(等)で外部に提供するメタデータ

内部メタデータ

- 自分たちがリポジトリに登録する(した)メタデータや、システムが自動的に付与した登録日などのメタデータ。
- 通常リポジトリ内のデータベースに登録される。

The image shows a screenshot of the HUSCAP submission interface on the left and a Microsoft Access database window on the right. The HUSCAP interface includes a header with the university logo and name, a search bar, and a series of buttons for document processing: 記述 (Description), アップロード (Upload), 確認 (Check), ライセンス (License), and 完了 (Complete). The main form is titled "投稿: アイテムの記述" (Submission: Item Description) and contains various input fields for metadata such as status, author, title, publication date, and journal name. A "Check Ro" button is visible at the bottom right of the form.

The Microsoft Access window displays a table named "クエリ1" (Query1) with the following data:

element	qualifier	text_value
creator		紙谷, 五月
creator		野中, 雄司
creator		杉田, 茂樹
date	accessioned	2008-12-05T01:41:19Z
date	available	2008-12-05T01:41:19Z
date	issued	2008-12-01
identifier	uri	http://hdl.handle.net/2115/34975
description		特集 ファインダビリティ向上
description	abstract	機関リポジトリは独自の集客力を持たず、外部の情報サービスシステムに集客を
description	abstract	nd the usage.
description	provenance	Submitted by 2班 / 理学系 図書館 (huscap2@lib.hokudai.ac.jp) on 2008-12-05
description	provenance	Approved for entry into archive by 2班 / 理学系 図書館(huscap2@lib.hokudai.ac.jp)
description	provenance	Made available in DSspace on 2008-12-05T01:41:19Z (GMT). No. of bitstreams: 1
language	iso	ja
publisher		情報科学技術協会
relation	uri	http://www.infosta.or.jp/journal/200812/jhtml#6
subject		機関リポジトリ
subject		リンク・リゾルバ
subject		OAI-PMH
subject		OpenURL
subject		AIRoway
subject		institutional repository
subject		link-resolver
subject		機関リポジトリへのアクセス経路
title	alternative	Access path to institutional repositories
type		article (author version)
creator	alternative	Kamiya, Satsuki
creator	alternative	Nonaka, Yūji
creator	alternative	Sugita, Shigeaki
subject	ndc	007
citation	title	情報の科学と技術 = The Journal of Information Science and Technology Association
citation	issn	0913-3801
citation	volume	58
citation	issue	12
citation	spage	610
citation	epage	614
citation	ncid	AN10005857

外部への提供メタデータ

- 内部メタデータから提供先に合わせた形式に変換して出力。
- OAI-PMHではXMLを出力。
- データ取得要求がくる度に変換プログラムが動いて出力する。

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。以下にドキュメントツリーを表示します。

```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2009-07-13T02:54:36Z</responseDate>
  -<request identifier="oai:eprints.lib.hokudai.ac.jp:2115/34975" metadataPrefix="oai_dc" verb="GetRecord">
    http://eprints.lib.hokudai.ac.jp/dspace-oai/request
  </request>
  -<GetRecord>
    -<record>
      -<header>
        <identifier>oai:eprints.lib.hokudai.ac.jp:2115/34975</identifier>
        <datestamp>2008-12-08T09:53:35Z</datestamp>
        <setSpec>hdl:2115_20083</setSpec>
        <setSpec>hdl:2115_1</setSpec>
      </header>
      -<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:creator>紙谷, 五月</dc:creator>
          <dc:creator>野中, 雄司</dc:creator>
          <dc:creator>杉田, 茂樹</dc:creator>
          <dc:date>2008-12-01</dc:date>
          <dc:identifier>http://hdl.handle.net/2115/34975</dc:identifier>
          <dc:description>特集 ファインダビリティ向上</dc:description>
          -<dc:description>
            機関リポジトリは独自の集客力を持たず、外部の情報サービスシステムに集客を依存する。従って、より多く利用されるためには、主たる閲覧利用者層として想定される研究者の日常的な情報探索行動の中で、収載文献が自然に発見されアクセスされるような経路の形成と誘導が必要となる。これを目的として実装され、リンク・リゾルバのための追加情報源として機能するAIRway システムを概観し、現在の被利用状況について述べ、今後の展望についてまとめる。
          </dc:description>
          -<dc:description>
            An institutional repository does not have its own power to attract large users, but commends user navigations to external service providers. Academics are assumed to be the primary users of institutional repository. Therefore, to gain their access, it is important to make the repository contents discoverable and accessible within their day-to-day research behavior. We have implemented "AIRway" system which is designed to be yet another knowledgebase for link-resolvers. This paper describes its functions and the usage.
          </dc:description>
          <dc:language>ja</dc:language>
          <dc:publisher>情報科学技術協会</dc:publisher>
          <dc:relation>http://www.infosta.or.jp/journal/200812j.html#6</dc:relation>
          <dc:subject>機関リポジトリ</dc:subject>
          <dc:subject>リンク・リゾルバ</dc:subject>
          <dc:subject>OAI-PMH</dc:subject>
          <dc:subject>OpenURL</dc:subject>
        </oai_dc:dc>
      </metadata>
    </record>
  </GetRecord>
</OAI-PMH>
```

用途による2つのメタデータを設計 する際の考え方

**世界標準は顧慮せず、
なおかつ、顧慮する**

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！**
 - 内部で文献を管理するための情報やリポジトリシステム上で見せたい項目を保持！
 - ハーベスティングしてほしいサービスに対応できる情報を保持！
- **外部へのデータ提供は標準準拠で！**
 - OAI-PMH(等)で外部へデータ提供するためのメタデータは標準準拠で！
 - 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)**の出力が求められる。
 - junii2, OALDC, MTD-MS…

2-1. 内部メタデータ設計

内部メタデータ(自由に！)

- 機関リポジトリは扱うコンテンツが多様なこともあり、メタデータ項目、記述方法について、NACSIS-CATのコーディングマニュアルのような統一的規範は**存在しない**。
- 設計は自分で(機関ごとに)しなくてはならない。
- ではどうするか？

内部メタデータ(設計考慮点)

1. どのようなメタデータフォーマットでデータを外部に提供するのかを**あらかじめ想定**し,
2. リポジトリ内部で保持する内部メタデータを**資料種別ごと**に明確に決定することが重要
3. 必要な情報を**不足なく**定義すること

資料種別ごとに必要項目を考えてみる

内部メタデータ項目は様々なタイプの資料の最小公倍数で	雑誌	紀要	会議	学位	教材
タイトル	●	●	●	●	●
著者	●	●	●	●	●
抄録	●	●		●	
著者キーワード	●	●			
学位授与機関・学位の種類				●	
掲載誌情報	●	●			
発表学会・開催地・開催年			●		
授業年度・講義科目名					●

junii2形式に対応したいので分割して保持しなくちゃ

例えば学位論文

例えば
HUSCAPでは

完全メタデータレコード

DCフィールド	値
dc.creator	加藤, 克
dc.date.accessioned	2007-08-28T06:44:36Z
dc.date.available	2007-08-28T06:44:36Z
dc.date.issued	2006-03-24
dc.identifier.uri	http://hdl.handle.net/2115/28095
dc.language.iso	ja
dc.title	ブラキストン「標本」史
dc.type	theses (doctoral)
dc.subject.ndc	900
dc.description.degreeGrantor	Hokkaido University (北海道大学)
dc.description.degreeLevel	博士
dc.description.degreeDiscipline	文学

学位授与機関・学位
の種類を内部で保持
したかったので追加

例えば雑誌論文

例えば
HUSCAPでは

完全メタデータレコード

DCフィールド

値

dc.creator Kamiya, Hiroyuki

dc.creator Fukunaga, Satoki

dc.creator Ohyama, Takashi

dc.creator Harashima, Hideyoshi

dc.date.accessioned 2009-06-17T06:07:15Z

dc.title Effects of carriers on transgene expression from plasmids containing a DNA sequence with high histone affinity

dc.type article (author version)

dc.subject.ndc 499

dc.citation.jtitle International Journal of Pharmaceutics

dc.citation.issn 0378-5173

dc.citation.volume 376

dc.citation.issue 1-2

dc.citation.spage 99

dc.citation.epage 103

dc.citation.doi 10.1016/j.ijpharm.2009.04.032

JAIROにデータ提供したいので掲載誌情報は分割して保持するように項目追加。
(junii2形式で外部にデータ提供したいので)

例えば作業管理メタ

例えば
HUSCAPでは

北海道大学学術成果コレクション
HUSCAP
Hokkaido University collection of Scholarly and Academic Papers
Copyright(c) 2005 Hokkaido University Library, All Rights Reserved.

北海道大学 | 附属図書館 | HUSCAP

検索

表示言語: 日本語

一覧時の表示範囲: 全文献

記述 記述 記述 アップロード 確認 ライセンス 完了

投稿: アイテムの記述

投稿アイテムに関する情報を下のフィールドに入力してください。ほとんどのブラウザでは、タブキーによりフィールドを移動することができます。([詳細ヘルプ...](#))

status_origin

担当者

status

作業メモ

公開画面には表示しない作業管理情報もメタデータとして保持

内部メタデータの記述

- 記述方法(書誌部分)についても、やっぱり統一の規範は**存在しない**。
- 「**学術コンテンツ登録システム**」の「**II. データ記述マニュアル**」を参考に行っている機関も。
 - <http://www.nii.ac.jp/nels/man/index.html>

「タイトルにルビが入っている場合はどうやって入力すれば？」
などなど記述方法についての疑問

2-2. 外部へのデータ提供設計

外部へのデータ提供用メタデータ (標準準拠で！)

- データ提供のためのメタデータにはフォーマットごとに記述仕様がある。(サービスプロバイダは、複数リポジトリからデータを集めてそれを元にサービスするため)

メタデータフォーマット	使用サービス等
OAI_DC(シンプルダブリンコア)	OAI-PMHで必須。OAIster等で使用
junii2	JAIRO、CiNiiで使用
ETD-MS	学位論文用。NDLTDで使用

各メタデータスキーマと参考ページ

各フォーマットの記述方法はメタデータスキーマで表現されています。

- **OAI_DC**

- http://www.openarchives.org/OAI/2.0/oai_dc.xsd
- <http://dublincore.org/documents/dces/>

- **junii2**

- <http://irdb.nii.ac.jp/oai/junii2.xsd>
- <http://www.nii.ac.jp/irp/archive/system/junii2.html>
- http://www.nii.ac.jp/irp/archive/system/junii2_guide.html
(junii2ガイドライン)

- **ETD-MS**

- <http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd/view>
- <http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>

外部へのデータ提供用メタデータ

- それぞれのフォーマットで出力するためにフォーマットごとに「内部メタデータ」→「外部に提供するメタデータ」への変換プログラムが必要

「クロスウォーク」といいます。

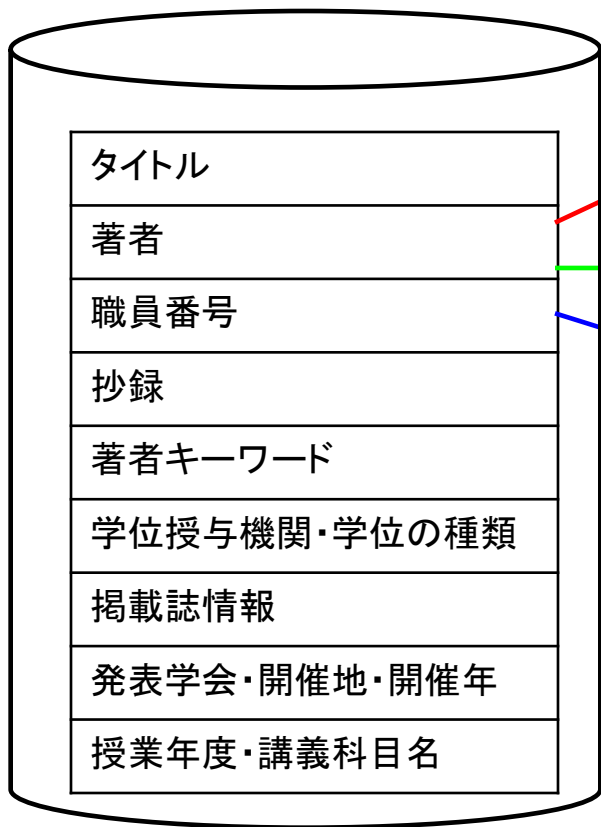
- 内部メタデータと外部に提供するメタデータのマッピングをしておく必要がある。

クロスウォーク(crosswalk)

A large orange speech bubble with a white outline, containing the text 'HUSCAPの例'.

HUSCAPの例

内部メタデータ



junii2

dc:title
dc:creator
dc:description
dc:subject

junii2:title
junii2:creator
junii2:description
junii2:subject
junii2:issn, junii2:jtitle junii2:volume, junii2:issue ほか

OAI_DC (シンプル
ダブリンコア)

ETD-MS

dc:title
dc:creator
dc:description
thesis:degree.name thesis:degree.level thesis:degree.discipline thesis:degree.grantor

まとめ

再びイメージ

外部データ提供
用メタデータは標
準仕様で！

(きとんと出力しないとそも
そも収録されないかも。)

クロスウォーク
設定

機関
リポジトリ

内部メタデータ
は自由に！
不足なく！

OAI-PMH
(junii2)

IRDBコンテンツ分析
NII Institutional Repositories DataBase Contents Analysis

OAI-PMH
(oai_dc)



OAI-PMH
(ETD-MS)

Scientific Commons

NDLTD



内部メタデータ設計とクロスウォーク設定はとっても大変そう？

- まずはいろいろリポジトリに入力，登録してみ
て，OAI-PMHの各フォーマットでどのように出
力されるか確認してみる。
 - http://XXX.ac.jp/verb?XXXXX&metadataPrefix=oai_dc&...
 - <http://XXX.ac.jp/verb?XXXXX&metadataPrefix=junii2&...>
- **必要があれば調整すればよい**

参考になるサイト等

- IRDBハーベスタ機能実装仕様 (junii2)
http://www.nii.ac.jp/irp/archive/system/irdb_harvest.html
- DRIVERガイドライン (oai_dc)
<http://drf.lib.hokudai.ac.jp/drf/index.php?tech>
- DRF-ML上での議論 (oai_dc)
DRF-ML(<http://drf.lib.hokudai.ac.jp/drfml/>)で [oai_dc] や [クロスウォーク] で検索

業者に発注する場合

- OAI-PMHの知識もあるかどうかも確認したほうがよいと思います。
- 導入時には、内部メタデータ項目のカスタマイズだけではなく、最低限oai_dc, junii2のクロスウォーク設定まで盛り込むことが重要と思います。

サービスプロバイダにはどうやってハーベスティングしてもらえばいいの？

**次の
「機関リポジトリの公開」
講義のお楽しみ！**

終わり

**世界中からアクセスされやすい
リポジトリをみんなで考えていきましょう！**