

機関リポジトリの メタデータ概論

平成21年度学術ポータル担当者研修

平成21年9月10日 <会場:国立情報学研究所>

一橋大学学術システム課 高橋菜奈子

Special Thanks to 野中雄司(北海道大学)

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。**
 - **構築後の設定でも十分対応可能**
- **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サービスプロバイダー (JAIRO等) がある。**
 - **最初の設定が重要!**

→ **今日はこちらの話**

1-2 OAI-PMH

OAI-PMHとは？

- The Open Archives Initiative Protocol for Metadata Harvestingの略
- 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

OAI-PMHに対応すると...

- 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プロバイダーが(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。

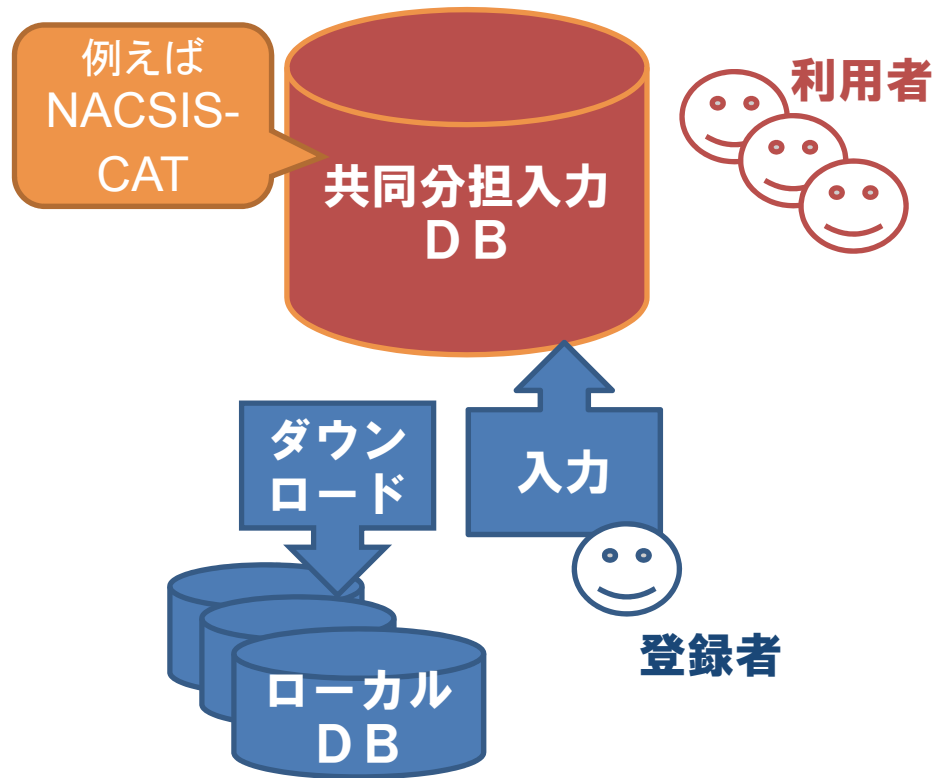
「ハーベスティング」といいます

例えばJAIRO

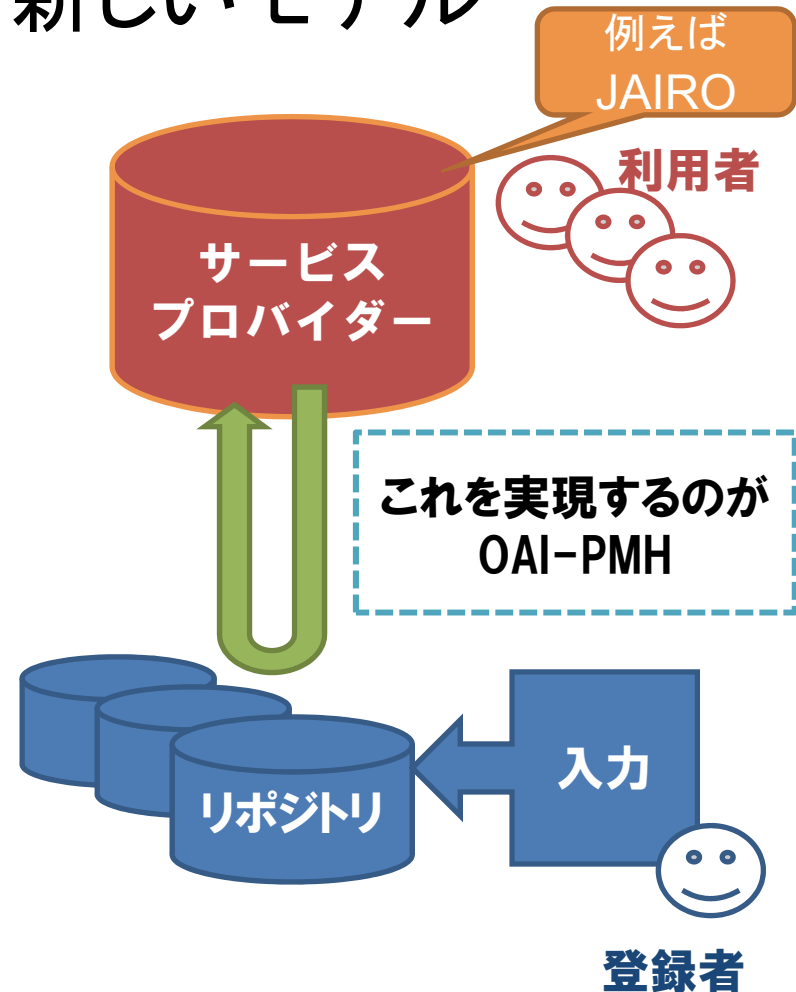
- 自分たちからデータを送付する必要がない。
 - リポジトリはデータを蓄積するのみ！
- 変更，削除記録も考慮してくれる。

OAI-PMHで実現できたこと

- 今までのモデル



- 新しいモデル



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)**

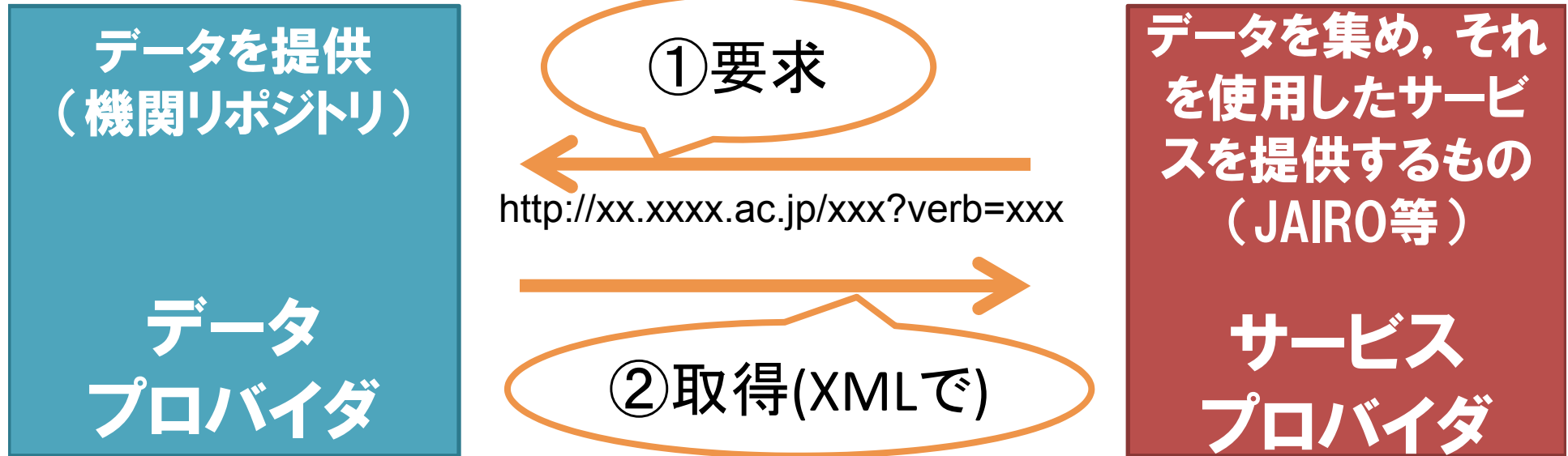
「OAI-PMHで外部にデータ提供することである。」といっても過言ではない。

- 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



プロトコル概要



●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

コマンドとパラメタ

Verb	説明	指定可能パラメタ
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

パラメタ	説明
metadataPrefix	このデータ形式で
set	この集合のデータを
from	この日時のデータから
until	この日時のデータまで
identifier	このIDのデータを
resumptionToken	次項のデータを

本表はOAI-PMHの仕様書に加え以下を参考に作成
<http://www12.ocn.ne.jp/~zuki/drf/repository.html>

要求URL(コマンド)例

ベースURL

<http://hermes-ir.lib.hit-u.ac.jp/rs-oai/request?>

verb=ListRecords

レコードの一覧をください

&metadataPrefix=junii2

junii2形式で

&from=2009-08-01

&until=2009-08-30

2009年8月1日から8月30日までに追加された(もしくは変更された)

実際にはどうなるのか？

一橋大学機関リポジトリ
HERMES-IR
Hitotsubashi University

このアイテムの引用には次の識別子を使用してください: <http://hdl.handle.net/10086/16457>

Title: NACSIS-CATにおける韓国・朝鮮人著者名典拠の
Other Titles: Identification of Korean Author name authority records in NACSIS-CAT
Authors: 高橋, 菜奈子
Issue Date: Aug-2007
Publisher: 学術文献普及会
Citation: 大学図書館研究
Volume: 80
Start Page: 65
End Page: 73
Language: jpn
Text Version: publisher
Appears in Collections: 010a Journal Articles = 雑誌掲載論文

Files in This Item:

File	Description	Size	Format
0100805701.pdf		351Kb	Adobe PDF

アイテムの詳細リンクを表示する

機関リポジトリに登録

JAIRO
Japanese Institutional Reports

このアイテムのアクセス数: 88 (2009/08/19)
<http://jairo.niac.jp/0033/00014300>

書誌情報

NACSIS-CATにおける韓国・朝鮮人著者名典拠の同定
Identification of Korean Author name authority records in NACSIS-CAT
NACSIS-CAT ニオケル カンコク チョウセンジン チョシヤメイ デンキョ ノドウテイ
高橋, 菜奈子, Takahashi, Nanako
大学図書館研究
80(65) - 73, 2007-08, 学術文献普及会
ISSN0098-6507
NACSIS(DC-ID): AN00136087

本文を読む
<http://hdl.handle.net/10086/16457>

JAIROがハーベスティング(月1回)

OAI-PMH

自動的にCiNiiに反映

```

<alternative>
  Identification of Korean Author name authority records in NACSIS-CAT
</alternative>
<alternative>NACSIS-CAT ニオケル カンコク チョウセンジン チョシヤメイ デンキョ ノドウテイ</alternative>
<subject>著者名典拠</subject>
<subject>同定</subject>
<subject>韓国</subject>
<subject>朝鮮人</subject>
<subject>NACSIS-CAT</subject>
<subject>著者名</subject>
<subject>同定</subject>
<subject>同音異名</subject>
<subject>別名</subject>
<subject>name authority records</subject>
<subject>identification</subject>
<subject>Korean author names</subject>
<subject>personal author heading</subject>
<subject>name name</subject>
<subject>different names</subject>
<subject>variant names</subject>
</description>
  一橋大学附属図書館で行なったハンゲル図書への追加入力作業として、NACSIS-CAT における韓国・朝鮮人著者名典拠の同定作業について報告する。著者名典拠とのリンク率が低いこと、その原因として同姓同名や同音異名、別名表記によるリンク形成の困難さが指摘できる。リンクを形成するにはその著者の著作、漢字表記、学位・学歴情報が同定根拠として有効であることも明らかになった。各機関が著者名リンクを放棄しないこと、手元の図書から判明する情報を追加していくことの重要性を指摘した。
</description>
<fullTextURL>http://hdl.handle.net/10086/16457/fullTextURL</fullTextURL>
<textVersion>publisher</textVersion>
<publisher>学術文献普及会</publisher>
  
```

CiNii
国立情報学研究所
論文情報ナビゲータ(サイエンス)

検索: takahashi nanako

登録済

NACSIS-CATにおける韓国・朝鮮人著者名典拠の同定
Identification of Korean Author name authority records in NACSIS-CAT

高橋, 菜奈子 Takahashi, Nanako

本文を読む/探す

Webcat Plus Webcat Plus 所蔵情報
Hermes 一橋大学附属図書館OPAC

キーワード

著者名典拠
同定
韓国・朝鮮人名
NACSIS-CAT
著者名
同姓同名
同音異名
別名表記
name authority records
identification
Korean author names
personal author heading
same name
different names
variant names

収録刊行物

大学図書館研究
Journal of college and university libraries 80 pp.65-73 2007-08
学術文献普及会

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

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。

スキーマ	要素	限定子	値	言語	
dc	contributor	author	高橋, 菜奈子		抹消
dc	date	accessioned	2009-01-20T06:22:41Z		抹消
dc	date	available	2009-01-20T06:22:41Z		抹消
dc	date	issued	2007-08		抹消
dc	identifier	uri	http://hdl.handle.net/10086/16457		抹消
dc	description	abstract	一橋大学附属図書館で行なったハンゲル図書の迎及入力を素材として、NACSIS-CATにおける韓国・朝鮮人名の著者名典拠の同定作業について報告する。著者	en	抹消
dc	description	provenance	Submitted by work1 hermes-ir (ir-work1@www.lib.hit-u.ac.jp) on 2009-01-20T06:22:41Z	en	抹消
dc	description	provenance	Made available in DSpace on 2009-01-20T06:22:41Z (GMT). No. of bitstreams: 1 0100805701.pdf: 359447 bytes, checksum:	en	抹消
dc	format	extent	359447 bytes		抹消
dc	format	mimetype	application/pdf		抹消
dc	language	iso	ja	en	抹消
dc	publisher		学術文献普及会	en	抹消
dc	subject		著者名典拠	en	抹消
dc	subject		同定	en	抹消

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交換プログラムが動いて出力する。

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns:xs="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.openarchives.org/OAI/2.0/"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2009-08-20T07:40:03Z</responseDate>
  <request metadataPrefix="oai_dc" identifier="oai:hermes-ir.lib.hit-u.ac.jp:10086/16457" verb="GetRecord">http://hermes-ir.lib.hit-u.ac.jp/rs-oai/request</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:hermes-ir.lib.hit-u.ac.jp:10086/16457</identifier>
        <timestamp>2009-07-27T13:35:35Z</timestamp>
        <setSpec>hdl_10086_9</setSpec>
      </header>
      <metadata>
        <?xml version="1.0" encoding="UTF-8" ?>
        <oai_dc:dc xmlns:xs="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:creator>高橋 菜奈子</dc:creator>
          <dc:date>2007-08</dc:date>
          <dc:identifier>http://hdl.handle.net/10086/16457</dc:identifier>
          <dc:description>
            一橋大学附属図書館で行なったハングル図書の選及人力を素材として、NACSIS-CATにおける韓国・朝鮮人名の著者名典拠の同定作業について報告する。著者名典拠とのリンク率が低いこと、その原因として同姓同名や同音異名、別名表記によるリンク形成の困難さが指摘できる。リンクを形成するにはその著者の著作、漢字表記、学位・学歴情報が同定根拠として有効であることも明らかになった。各機関が著者名リンクを放棄しないこと、手元の図書から判明する情報を追加していくことの重要性を指摘した。
          </dc:description>
          <dc:language>ja</dc:language>
          <dc:publisher>学術文献普及会</dc:publisher>
          <dc:subject>著者名典拠</dc:subject>
          <dc:subject>同定</dc:subject>
          <dc:subject>韓国・朝鮮人名</dc:subject>
          <dc:subject>NACSIS-CAT</dc:subject>
          <dc:subject>著者標目</dc:subject>
          <dc:subject>同姓同名</dc:subject>
          <dc:subject>同音異名</dc:subject>
          <dc:subject>別名表記</dc:subject>
          <dc:subject>name authority records</dc:subject>
          <dc:subject>identification</dc:subject>
          <dc:subject>Korean author names</dc:subject>
          <dc:subject>personal author heading</dc:subject>
          <dc:subject>same name</dc:subject>
          <dc:subject>different names</dc:subject>
          <dc:subject>variant names</dc:subject>
          <dc:title>NACSIS-CATにおける韓国・朝鮮人著者名典拠の同定</dc:title>
        </oai_dc:dc>
      </metadata>
    </record>
  </GetRecord>
</OAI-PMH>
```

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？

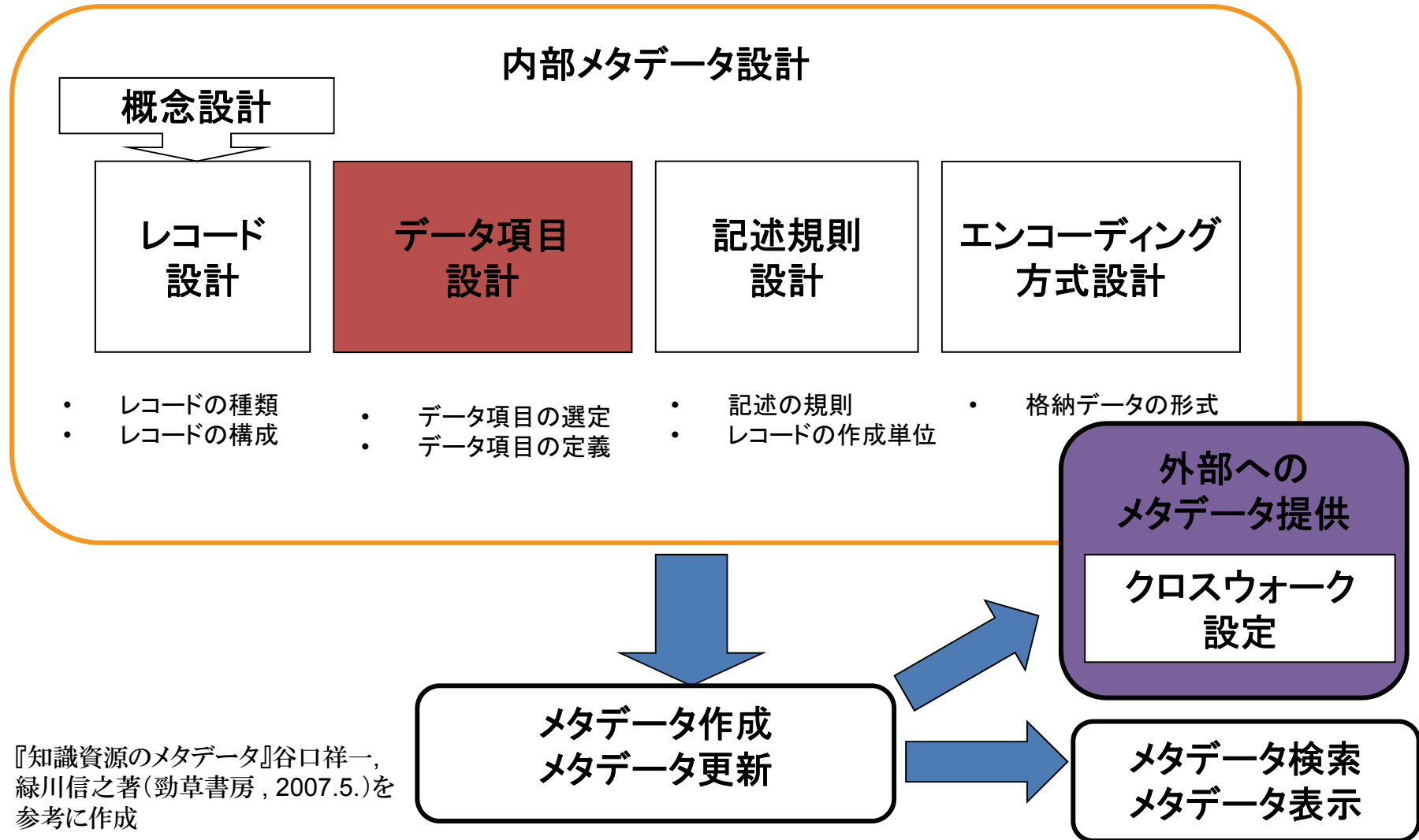
- **内部メタデータ設計は自由に！**
 - 内部で文献を管理するための情報やリポジトリシステム上で見せたい項目を保持！
- **外部へのデータ提供は標準準拠で！**
 - OAI-PMH(等)で外部へデータ提供するためのメタデータは標準準拠で！
 - サービスプロバイダによっていろいろな形式(メタデータフォーマット)の出力が求められる。

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、記述規則について、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**。
- 設計は自分でしなくてはならない。
- ではどうするか？

内部メタデータの設計



『知識資源のメタデータ』谷口祥一，
緑川信之著(勁草書房，2007.5.)を
参考に作成

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

以下を参考にした

http://drf.lib.hokudai.ac.jp/drf/index.php?plugin=attach&refer=DRF-Ookayama&openfile=020909_05_handout.pdf

内部メタデータ項目設計

資料種別ごとに
必要なデータ項目を
考えてみる

	雑誌	紀要	学会	学位	WP DP	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			●			
バージョン情報					●	
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
HERMES-IRでは

DC Field	Value
dc.contributor.author	坂内, 徳明
dc.date.accessioned	2008-03-26T03:29:42Z
dc.date.available	2008-03-26T03:29:42Z
dc.date.issued	1995-03-08
dc.date.submitted	1994-12
dc.identifier.uri	http://hdl.handle.net/10086/15407
dc.language.iso	ja
dc.title	ロシア民俗学史再考
dc.title.transcription	ロシア ミンゾクガクシ サイコウ
dc.contributor.alternative	Bannai, Tokuaki
dc.type.NII	Thesis or Dissertation
dc.type.DCMIType	Text
dc.language.iso639-2	jpn
dc.affiliation	一橋大学大学院言語社会研究科
dc.description.degree	博士(社会学)
dc.description.degreeNo	乙第139号
dc.description.physdesc	2冊
dc.description.degreeGrantor	一橋大学

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
HERMES-IRでは

DC Field	Value
dc.contributor.author	小島, 清
dc.date.accessioned	2009-04-15T09:22:26Z
dc.date.available	2009-04-15T09:22:26Z
dc.date.issued	2009-04
dc.identifier.uri	http://hdl.handle.net/10086/17283
dc.language.iso	ja
dc.publisher	世界経済研究協会
dc.title	合意的国際分業の原理(上) : 世界経済の異質化と同質化
dc.title.transcription	ゴウイテキ コクサイ ブンギョウ ノ ゲンリ ジョウ : セカイ ケイザイ ノ イシツカト ドウシツカ
dc.contributor.alternative	Kojima, Kiyoshi
dc.type.NII	Journal Article
dc.type.DCMIType	Text
dc.language.iso639-2	jpn
dc.citation.issn	0488-132X
dc.citation.NCID	AN0039049X
dc.citation.jtitle	世界経済評論
dc.citation.volume	53
dc.citation.issue	4
dc.citation.spage	33
dc.citation.epage	44
dc.textversion	publisher
dc.affiliation	一橋大学

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)

例えば管理メタデータ

例えば
HERMES-IRでは

スキーマ	要素	限定子	値
dc	contributor	author	渡辺, 雅男
dc	contributor	editor	一橋大学一橋学会
dc	Local	hdarid	AZ00011334 移行データの旧番号
dc	Local	kknrid	70175123 科学研究費研究者番号 (著者の名寄せに使用)
dc	Local	sortkey	130-004-01 目次順表示用のソートキー

公開の画面には表示しない
管理情報も保持

内部メタデータ記述規則設計

- 記述方法(書誌部分)についても、やっぱり統一規範は**存在しない**。

これが最大の悩み！

「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？」
「連載記事のタイトルはどうしよう？」
などなど記述方法についての疑問

- 「**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>

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

- **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>

- **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>

クロスウォーク(crosswalk)

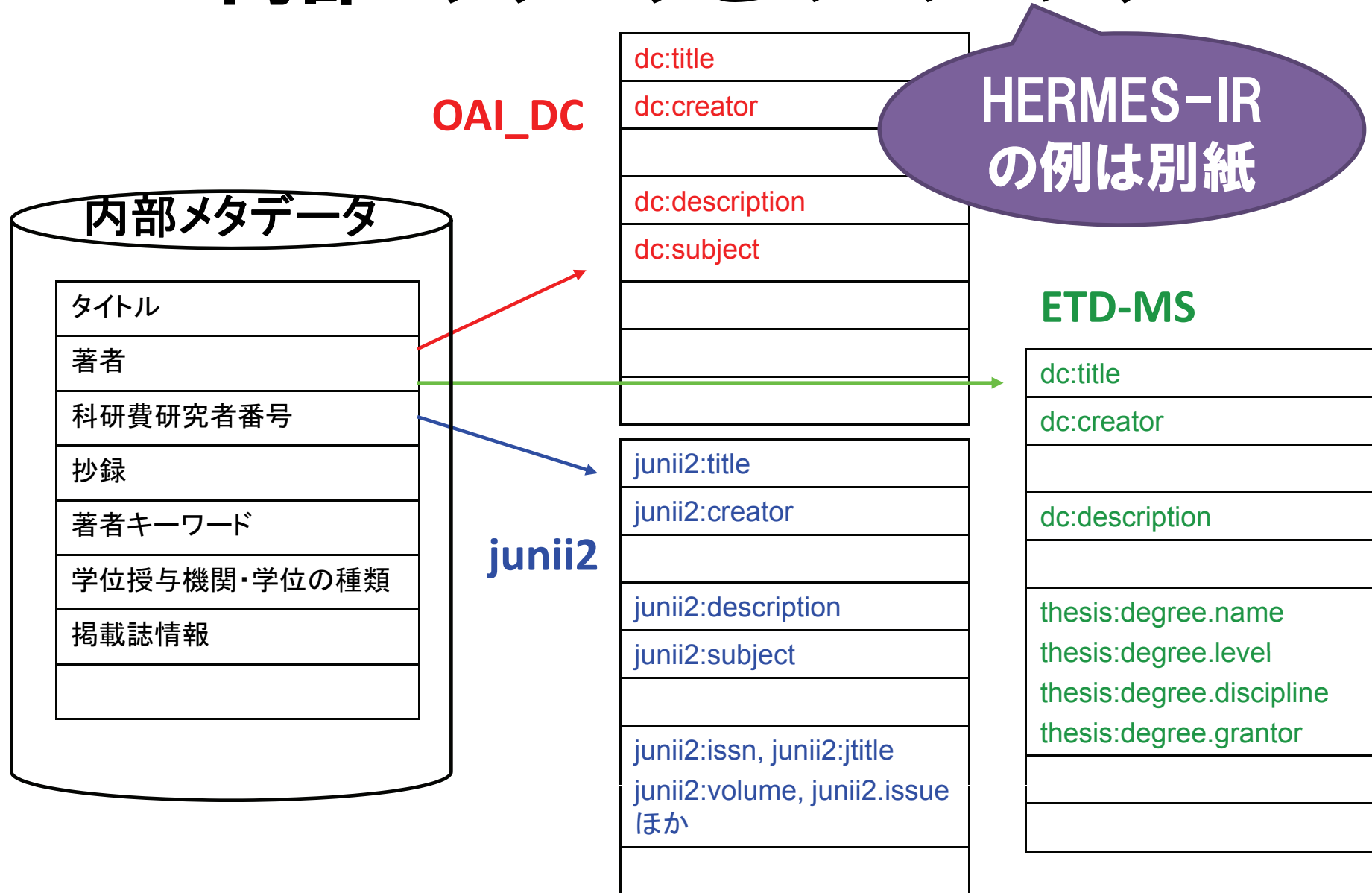
- それぞれのフォーマットで出力するためにフォーマットごとに「内部メタデータ」→「外部に提供するメタデータ」への変換プログラムが必要

「クロスウォーク」といいます。

HERMES-IR(DSpace)の場合
javaで書かれている。

- 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。

内部メタデータとのマッピング



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)



OAI-PMH
(oai_dc)



OAI-PMH
(ETD-MS)



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？

- **まずは業者に発注できる知識があればよい。**
 - 内部メタデータの項目の設定
 - 外部に提供するメタデータとのマッピング
- **必要があれば調整することも想定しておけばよい。**
 - (実際、調整せざるをえない。)

業者が発注する場合

- **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。**

自力でやってみてみたい方は...

- 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)
http://drf.lib.hokudai.ac.jp/cgi-bin/namazu.cgi?query=oai_dc&submit=Search

**サービスプロバイダにはどうやってハーベストイン
グしてもらえばいいの？**

**明日の「機関リポジトリの公
開」講義のお楽しみ！**