

機関リポジトリと NIIコンテンツサービス メタデータを中心に

平成22年度国立情報学研究所実務研修
成果発表
静岡大学附属図書館 杉山智章

職場からの眺望



- ▶ 静岡平野、大崩海岸、駿河湾、御前崎・・・

実務研修以前



機関リポジトリ(IR)との関わり

▶ 2007年度

- 静岡大がCSI学術機関リポジトリ構築連携支援事業 追加委託機関になる
 - 館内WGでシステム担当(当時雑誌情報担当)

▶ 2008年度

- 平成19年度委託事業報告交流会で事例報告

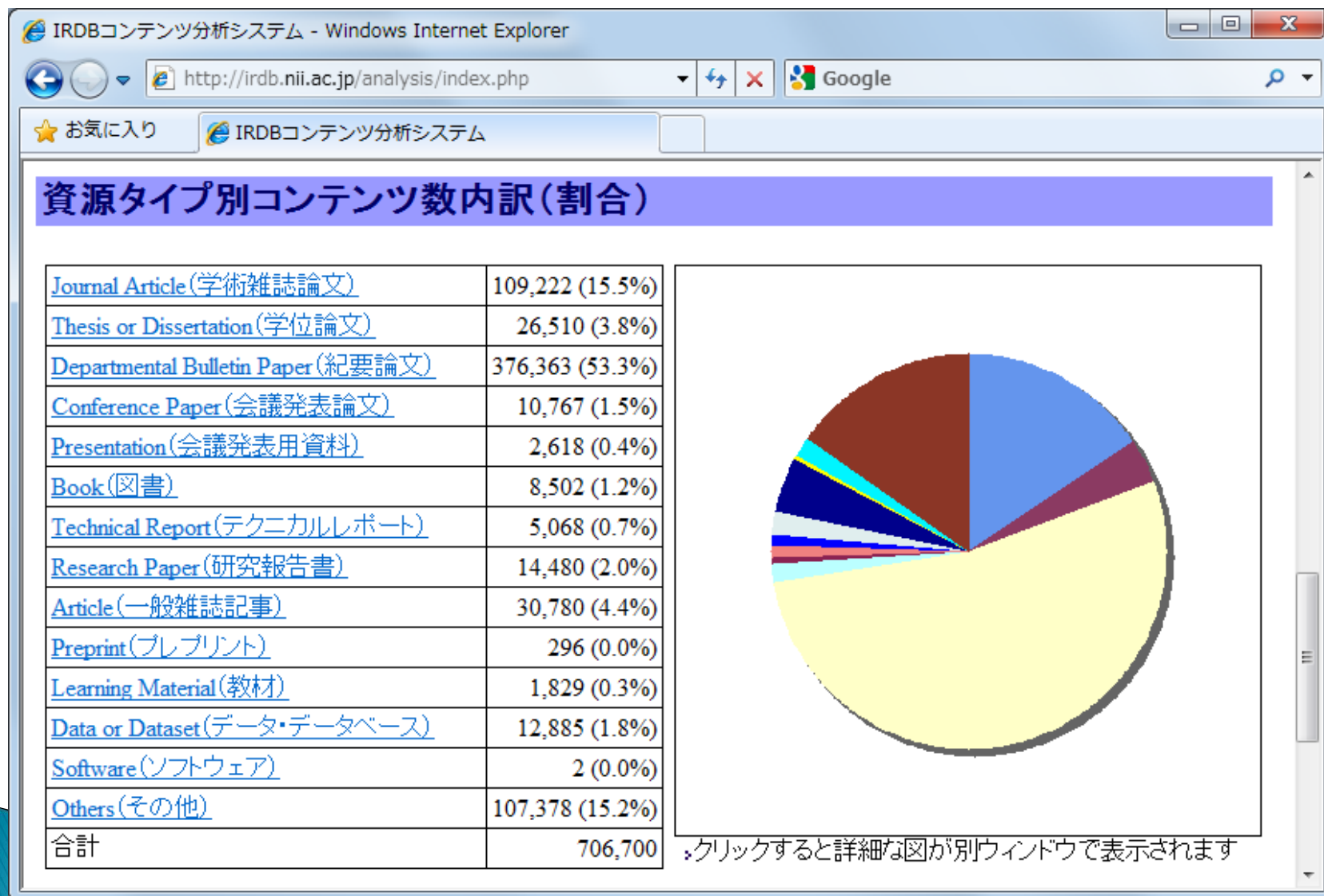
▶ 2009年度

- 図書館連携作業部会 IRシステム検討WG
 - メタデータ班「junii2ガイドライン」検討メンバー

IRメタデータを考えたきっかけ

- ▶ IRのメタデータとは、コンテンツの特性を記録したデータで、サービスに重要なのは「書誌情報」として扱われる部分
- ▶ 図書館がIRを運営する意義の一つに、この「書誌情報」を正しくメンテナンスすることがあるのではないか？
- ▶ 目録業務経験者として、IRのメタデータが“自由すぎて”危機感をもっていた。
→ 書誌情報を正しく伝えられない！

IRコンテンツ ～ 多様な資料種別



2009年度WGでの活動

- ▶ さまざまな資料種別のメタデータをjunii2でどのように表現するのか??
 - 決めるべき！！
 - しかし、その段階では無理でした。
- ▶ 雑誌論文メタデータを前提とした指針づくりをスタート
 - 「[junii2ガイドライン](#)」(IRとNIIとのメタデータ交換のガイドラインで、入カマニュアルではありません)

junii2ガイドラインの私のイメージ

書誌情報として
正しいこと
(SISTなど)

NIIサービスにとって
リーズナブルなこと

バランス

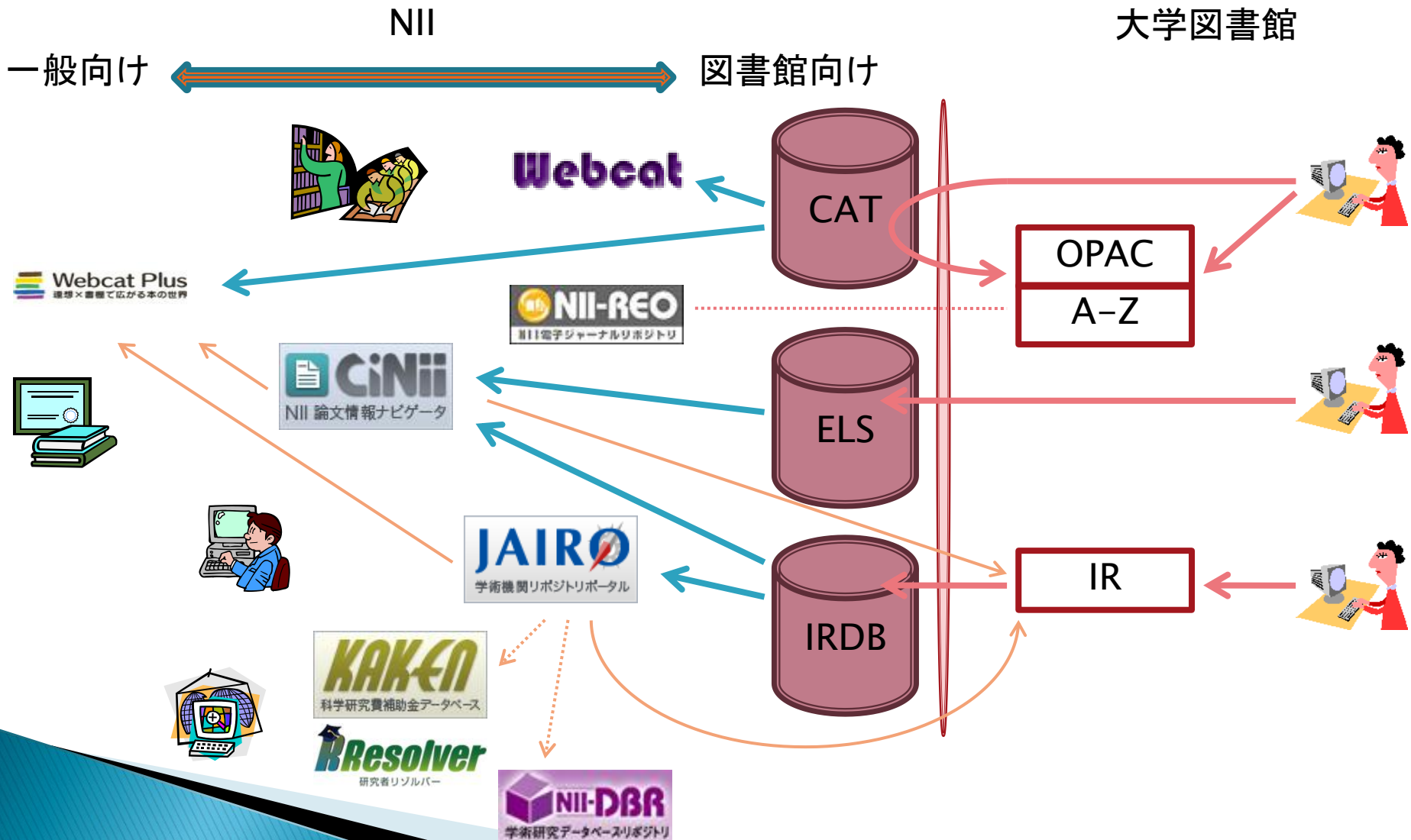
OAI_DC(DRIVER2.0)
にダムダウン可能
なこと

OpenURLで
解決できること

そして、実務研修・・・



大学図書館とコンテンツ系サービス



実務研修のスタートライン

- ▶ まずは、学術コンテンツ課の中の仕事を知らなかった。
- ▶ そして、各サービスがどのような関係にあるのかを理解しなかった。
- ▶ その上で、自分に最も適したテーマを探したかった。

→ 結果

自分が今、実際に大学で業務を担当している
機関リポジトリをテーマとした。

研修概要 -振り返って-

月	内容
7月	学術コンテンツ課の業務を通して、コンテンツ系サービスの概要を理解し、研修目標を設定。
8月	機関リポジトリのメタデータについて、今後のあり方の検討をするとともに、必要技術を習得。
9月	検討結果をもとに、実環境と同じ条件下での実装と検証。今後に向けた課題の整理。研修のまとめ。

NIIコンテンツ系サービスの展開

- ▶ 新CiNii
- ▶ 新KAKEN
- ▶ 研究者リゾルバー
- ▶ Webcat後継
- ▶ 博士論文台帳DB(?) NII-DBR(?)
- ▶ クラウド型リポジトリ(?)

機関リポジトリとの新たな連携はあるのか？

webの動向

▶ セマンティック・ウェブ

- 機械がweb上の資源のセマンティックス(意味)を扱えるように

rdf, rss, foaf, owl etc.

IRの具体的仕様としては「OAI-ORE」

▶ 「DC-NDL」(国立国会図書館) etc.

<http://www.ndl.go.jp/jp/library/data/meta.html>

方向性は2つ

1. junii2を拡張する(これまでのxml)
 - OAI-PMH
 - フラットな構造を維持するか？
2. junii3を策定する(セマンティックなステージへ)
 - OAI-OREも含む RDF/XMLによる実装
 - リテラルからリソースへ
 - OAI-PMHに限らない、アイテムページがRDFクローラ対応(?)

現行フォーマットの拡張(案)



junii2拡張

- ▶ NIIのサービス展開に応じて、minimumな拡張を行っていく。

NII側の現在の課題は、
「JAIROによる著者名寄せ」

“Authority Control”

- ▶ 著者にAuthority Control(典拠コントロール)を実施する
 - 同姓同名の著者を正しく識別するには典拠コントロールが必要
 - 研究者リゾルバーによるリゾルブ
 - JAIRO、CiNii、KAKENなどサービスを越えた典拠の可能性が広がる

NACISIS-CATの著者名典拠

▶ コーディングマニュアル

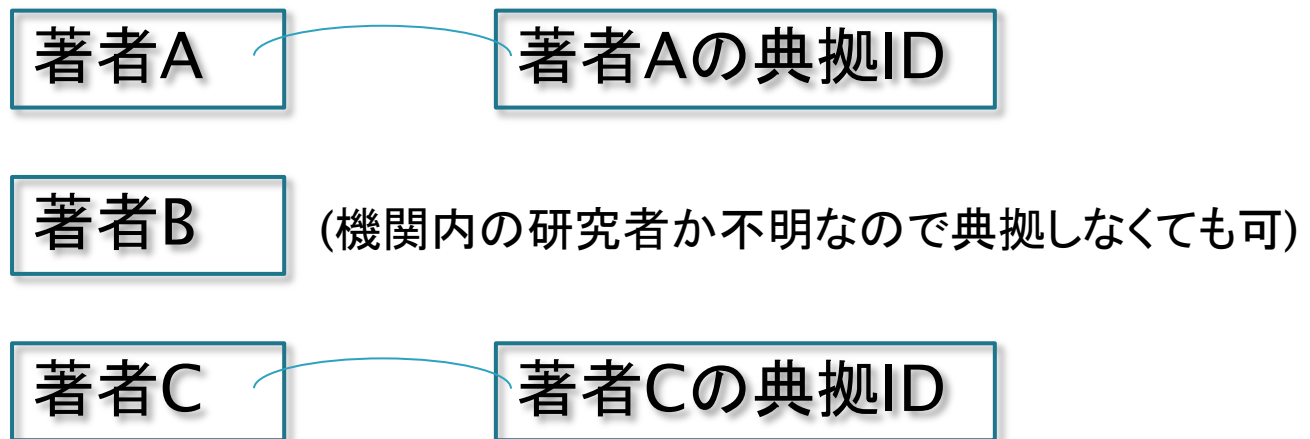
- 「著者名リンクは可能な限り形成する。」
- しかし、典拠リンクされていない著者が多い
 - どの誰か分からない著者の典拠作成は大変

▶ IRで同様なことを行い成功するのか？

- 典拠コントロールするのは、原則として機関に関係(所属あるいは所属した)する著者(研究者)のみである。
 - 守備範囲の明確化

システム要件

- ▶ IRがjunii2で<creator>の研究者名と、その典拠IDを「ひもづけ」して出力できるようにする。



世界は同じことを考えている

DSpace1.6 metadata value テーブル

Column	Type	説明	例
metadata_value_id	integer	キー(通番)	356227
Item_id	Integer	アイテムID	3676
metadata_field_id	Integer	フィールドID	3 (contributor.author)
text_value	text	値	Kato, Kenji
text_lang	character (24)	言語コード	en
place	Integer	フィールド繰り返し	1
authority	character (100)	典拠ID	70169499
confidence	Integer	典拠信頼性	600

SURE: Shizuoka University REpository: アイテムの編集 - Windows Internet Explorer

http://ir.lib.shizuoka.ac.jp/tools/edit-item?item_id=433&submit=%E7%B7%A8%E9%9B%86...

お気に入り SURE: Shizuoka University REpository: アイ...

スキーマ	要素	限定子	値	言語	
dc	title		柿田川を中心とした富士山湧水の細菌群集		抹消
dc	title	alternative	Microbial Abundance in Kakitagawa springs and springs in Mt. Fuji		抹消
dc	contributor	author	長岡, 篤子 🌱 <input type="text"/> 🔒 🔍		抹消
dc	contributor	author	木村, 浩之 🌱 30377717 🔒 🔍		抹消
dc	contributor	author	加藤, 憲二 🌱 70169499 🔒 🔍		抹消
dc	contributor	transcription	キムラ, ヒロユキ 🌱 30377717 🔒 🔍		抹消
dc	contributor	transcription	カトウ, ケンジ 🌱 70169499 🔒 🔍		抹消
dc	contributor	alternative	Nagaoka, Atsuko 🌱 <input type="text"/> 🔒 🔍		抹消
dc	contributor	alternative	Kimura, Hiroyuki 🌱 30377717 🔒 🔍		抹消
dc	contributor	alternative	Kato, Kenji 🌱 70169499 🔒 🔍		抹消
dc	subject	ndc	465		抹消

著者IDのOAI-PMH応答

junii2Crosswalk.javaの改修

```
DCValue[] dcvalues = item.getDC(element, qualifier, Item.ANY);
```

(略)

```
for (int j = 0; j < dcvalues.length; j++) {  
    element    = dcvalues[j].element;  
    qualifier  = dcvalues[j].qualifier;  
    String value = dcvalues[j].value;  
    String authority = dcvalues[j].authority;
```

値

典拠ID

拡張例1

- ▶ 属性を使用する。

```
<creator id="70169499">Kato, Kenji</creator>
```

※「id」の属性名は未定

- ▶ junii2の仕様にはすでに「オプションの属性」がある。
(lang, version)

拡張例2

- ▶ 要素を使用する。

<creator>

<著者名>kato, Kenji</著者名>

<著者ID><http://.../70169499></著者ID>

</creator>

※「著者名」「著者ID」の要素名は未定

※IDはURIとする可能性あり

- ▶ 拡張性は高いが、これまでの「junii2はフラットである」というポリシーを変更することになる。

テスト環境での実装

- ▶ 「拡張例1」「拡張例2」とも、SUREのテスト環境に実装済みである。

「拡張例1」(クリック)

「拡張例2」(クリック)

```
http://lsv9.lib.shizuoka.ac.jp/dspace-oai/request?verb=GetRecord&metadataPrefix=junii2a2&identifier=0 - Windows Internet Explorer
http://lsv9.lib.shizuoka.ac.jp/dspace-oai/request?verb=GetRecord&metadataPrefix=junii2a2&identifier=0
お気に入り http://lsv9.lib.shizuoka.ac.jp/dspace-oai/re...
<identifier>oai:lsv9.lib.shizuoka.ac.jp:10297/514</identifier>
<timestamp>2010-07-12T01:00:00Z</timestamp>
<setSpec>hdl_10297_11</setSpec>
</header>
- <metadata>
- <junii2 xmlns="http://ju.nii.ac.jp/junii2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://ju.nii.ac.jp/junii2 http://ju.nii.ac.jp/oai/junii2.xsd">
  <title>柿田川を中心とした富士山湧水の細菌群集</title>
  <alternative>Microbial Abundance in Kakitagawa springs and springs in Mt.
    Fuji</alternative>
  <creator>長岡, 篤子</creator>
  <creator id="30377717">木村, 浩之</creator>
  <creator id="70169499">加藤, 憲二</creator>
  <NDC>465</NDC>
  <publisher>日本微生物生態学会</publisher>
  <NIIttype>Journal Article</NIIttype>
  <format>application/pdf</format>
  <URI>http://hdl.handle.net/10297/514</URI>
```

```
http://lsv9.lib.shizuoka.ac.jp/dspace-oai/request?verb=GetRecord&metadataPrefix=junii2a1&identi - Windows Internet Explorer
http://lsv9.lib.shizuoka.ac.jp/dspace-oai/request?verb=GetRecord&metadataPrefix=junii2a1&identifier=o
お気に入り http://lsv9.lib.shizuoka.ac.jp/dspace-oai/re...
</header>
- <metadata>
- <junii2 xmlns="http://ju.nii.ac.jp/junii2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://ju.nii.ac.jp/junii2 http://ju.nii.ac.jp/oai/junii2.xsd">
  <title>柿田川を中心とした富士山湧水の細菌群集</title>
  <alternative>Microbial Abundance in Kakitagawa springs and springs in Mt.
    Fuji</alternative>
  <creator>長岡, 篤子</creator>
- <creator>
  <value>木村, 浩之</value>
  <authority>http://lsv9.lib.shizuoka.ac.jp/browse?
    type=author&authority=30377717</authority>
</creator>
- <creator>
  <value>加藤, 憲二</value>
  <authority>http://lsv9.lib.shizuoka.ac.jp/browse?
    type=author&authority=70169499</authority>
</creator>
<NDC>465</NDC>
```


典拠の管理

- ▶ 誰が、どんなIDを、どう管理するのか？
 - 研究者リゾルバーID, 科研研究者ID, ReaDのID?
 - 全研究者をカバーしているわけではない
 - 機関ごとのローカルID
 - 1人1IDを保証できるのか
 - IR内で異なる複数典拠ファイルの使用は許されるのか
 - ORCID, VIAF等の国際的な動向
- 「研究者リゾルバー」で解決(IDマッピング)

入力支援機能の必要性

▶ DSpace

「Populate Metadata From PubMed」

[https://wiki.duraspace.org/display/DSPACE/Populate
MetadataFromPubMed](https://wiki.duraspace.org/display/DSPACE/Populate+MetadataFromPubMed)

+ LC Name Authority

これもテスト環境に実装済み。
「ログイン」

- 論文IDを入力すると、メタデータと著者IDが自動入力される
→ 例えば、CiNiiと研究者リゾルバーの組み合わせ

普及のための課題

- ▶ DSpace旧バージョン用パッチ
 - パッチあり
 - 実装作業は誰が？費用はどう賄うか
 - 入力支援機能の開発が求められる
- ▶ 他システム(Xoonips, E-Repository etc.)
 - 開発費用が発生するであろう
 - NII製WEKOに実装して乗り換えを促す!?
- ▶ 現コンテンツに対する遡及的ID入力
 - 技術や費用的支援は考えられるか？

–htmlメタタグ–

- ▶ Google Scholarで使用
 - citationの詳細情報や本文URLを出力することは、「junii2」を実装している日本のIRがすでにやってきたこと。

Google Scholar

[Inclusion Guidelines for Webmasters](#)

SURE

実装例

高次なサービスを提供するには、
“それなり”のメタデータが求められる。

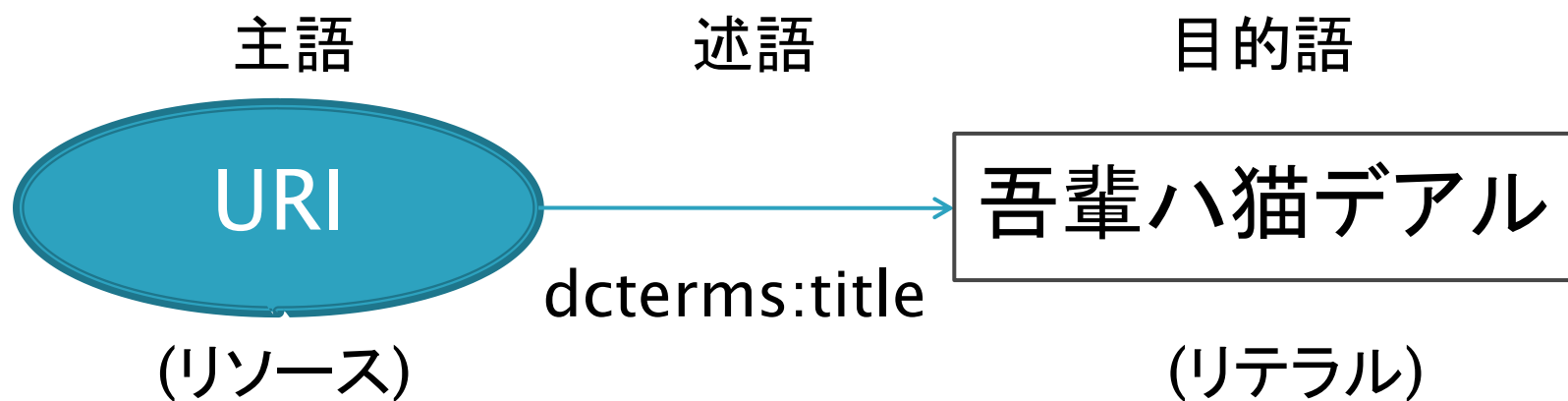
```
37
38 <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
39 <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
40 <meta name="DC.title" content="柿田川を中心とした富士山湧水の細菌群集" />
41 <meta name="DC.creator" content="長岡, 篤子" />
42 <meta name="DC.creator" content="木村, 浩之" />
43 <meta name="DC.creator" content="加藤, 憲二" />
44 <meta name="DC.publisher" content="日本微生物生態学会" />
45 <meta name="DC.issued" content="2005-10-30" scheme="DCTERMS.W3CDTF" />
46 <meta name="DC.relation.ispartof" content="日本微生物生態学会講演要旨集" />
47 <meta name="DC.citation.volume" content="21" />
48 <meta name="DC.citation.spage" content="69" />
49 <meta name="DC.citation.epage" content="69" />
50 <meta name="DC.rights" content="日本微生物生態学会: 本文データは学協会の許諾に基づ
51 きCiNiiから複製したものである" />
52 <meta name="DC.identifier"
53 content="http://ir.lib.shizuoka.ac.jp/bitstream/10297/514/1/080130014.pdf" />
54 <meta name="citation_title" content="柿田川を中心とした富士山湧水の細菌群集" />
55 <meta name="citation_author" content="長岡, 篤子" />
56 <meta name="citation_author" content="木村, 浩之" />
57 <meta name="citation_author" content="加藤, 憲二" />
58 <meta name="citation_publisher" content="日本微生物生態学会" />
59 <meta name="citation_date" content="2005-10-30" />
60 <meta name="citation_journal_title" content="日本微生物生態学会講演要旨集" />
61 <meta name="citation_volume" content="21" />
62 <meta name="citation_firstpage" content="69" />
63 <meta name="citation_lastpage" content="69" />
64 <meta name="citation_pdf_url"
65 content="http://ir.lib.shizuoka.ac.jp/bitstream/10297/514/1/080130014.pdf" />
```

新メタデータフォーマットの課題



junii3に向けて

RDF –セマンティックwebを実現する技術–



OAI-PMH用 RDF Crosswalk(リンク)

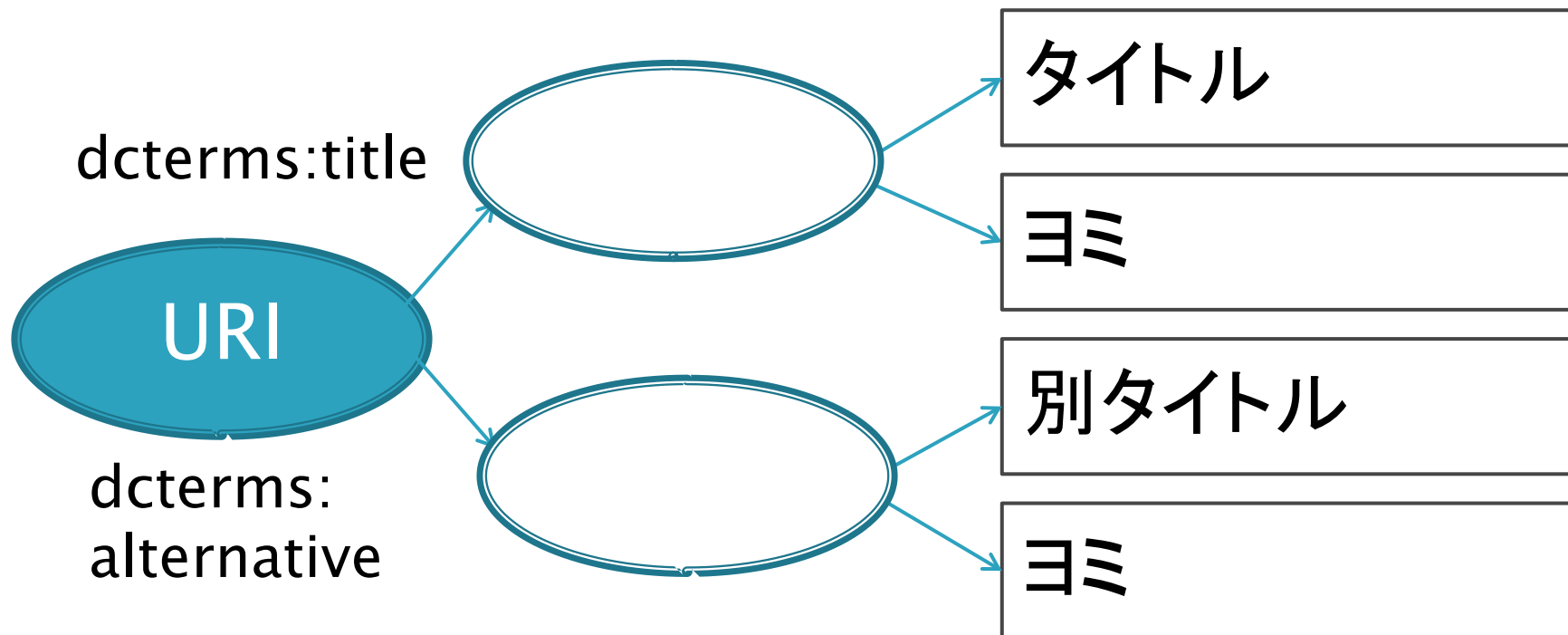
http://lsv9.lib.shizuoka.ac.jp/dspace-oai/request?verb=GetRecord&metadataPrefix=rdf&identifier= - Windows Internet Explorer

http://lsv9.lib.shizuoka.ac.jp/dspace-oai/request?verb=GetRecord&metadataPrefix=rdf

お気に入り http://lsv9.lib.shizuoka.ac.jp/dspace-oai/re...

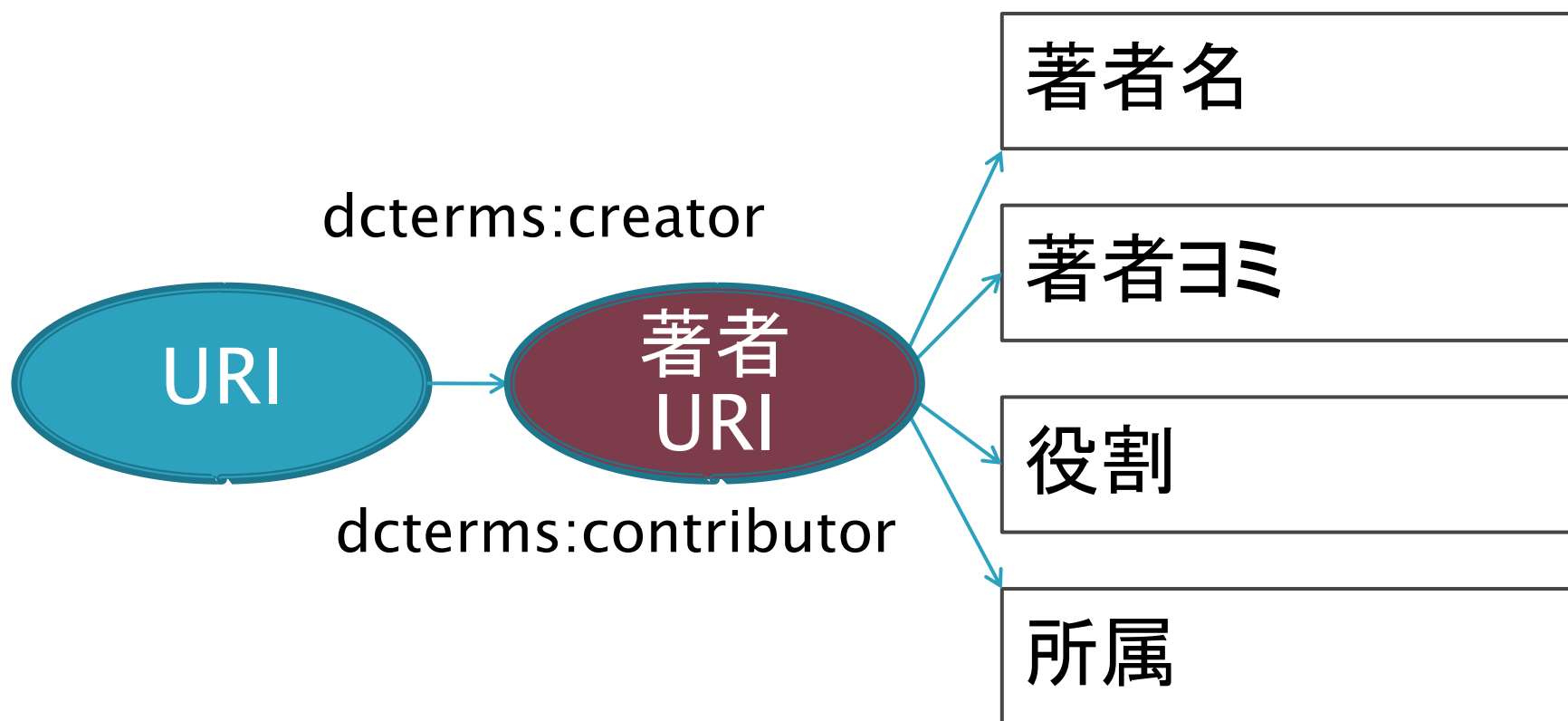
```
</header>
- <metadata>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ow="http://www.ontoweb.org/ontology/1#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ds="http://dspace.org/ds/elements/1.1/"
  xsi:schemaLocation="http://www.w3.org/1999/02/22-rdf-syntax-ns#
  http://www.openarchives.org/OAI/2.0/rdf.xsd">
- <ow:Publication rdf:about="oai:lsv9.lib.shizuoka.ac.jp:10297/514">
  <dc:title>柿田川を中心とした富士山湧水の細菌群集</dc:title>
  <dc:title>Microbial Abundance in Kakitagawa springs and springs
  in Mt. Fuji</dc:title>
  <dc:creator>長岡, 篤子</dc:creator>
  <dc:creator>木村, 浩之</dc:creator>
  <dc:creator>加藤, 憲二</dc:creator>
  <dc:subject>NDC:465</dc:subject>
  <dc:description>textversion:publisher</dc:description>
  <dc:publisher>日本微生物生態学会</dc:publisher>
  <dc:date>2005-10-30</dc:date>
```

フラットから構造化へ



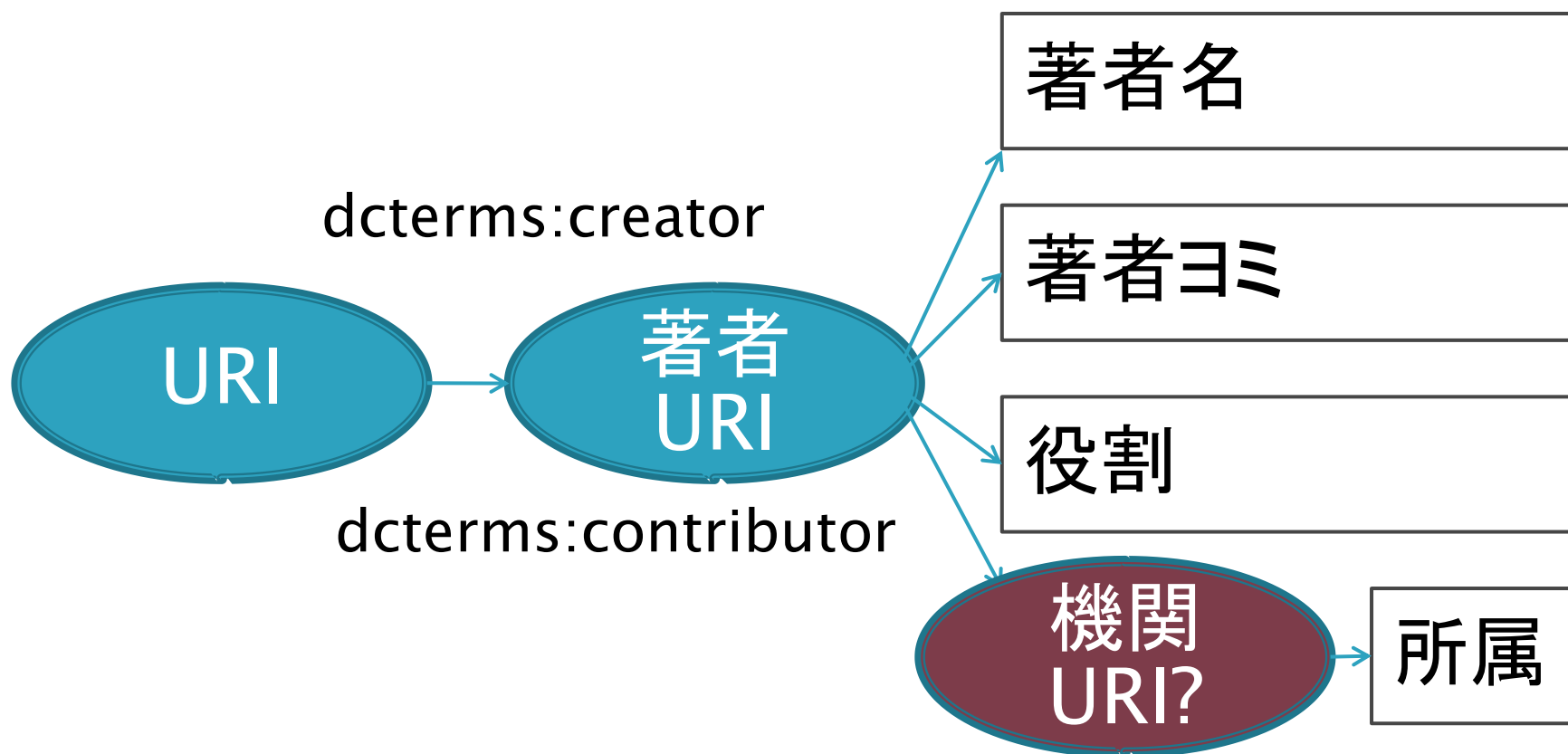
「ヨミ」のtermが悩ましい、そもそも「ヨミ」を扱うのか？

著者



「人」の情報を表す一般的なtermはFOAF。

著者



リソースは増えていくが、手間も増えていく。

例えばCiNii連携



NCID [雑誌書誌ID]	AN00123278
YEAR [年月次]	20010131
VOLN [巻号]	51(2)
ATTR [ページ属性]	P
TITL [論文名(日)]	被服イメージ判断におよぼす被服志向性の影響
TITY [論文名よみ]	
TITE [論文名(英)]	Effects of clothing orientation on the dress image judgment
AUTH [著者名(日)]	諸井,克英 / 鈴木,弥生 / 染谷,知雅 / 平田,幸恵
AUTY [著者名よみ]	モロイ,カツヒデ / スズキ,ヤヨイ / ソメヤ,チカ / ヒラタ,ユキエ
AUTE [著者名(英)]	MOROI,Katsuhide / SUZUKI,Yayoi / SOMEYA,Chika / HIRATA,Yukie
AFFN [著者所属(日)]	静岡大学人文学部 / 静岡大学人文学部 / 静岡大学人文学部 / 静岡大学人文学部
AFFE [著者所属(英)]	THE FACULTY OF HUMANITIES AND SOCIAL SCIENCES SHIZUOKA UNIVERSITY / THE FACULTY OF HUMANITIES AND SOCIAL SCIENCES SHIZUOKA UNIVERSITY / THE FACULTY OF HUMANITIES AND SOCIAL SCIENCES SHIZUOKA UNIVERSITY / THE FACULTY OF HUMANITIES AND SOCIAL SCIENCES SHIZUOKA UNIVERSITY
PAGE [ページ]	A1-A31
RTYP [記事種別(日)]	
RTYE [記事種別(英)]	
ABST [抄録(日)]	
	The purpose of this study is to investigate the effects of the subject's clothing orientation on the colored dress image judgment. In the 1st session, Clothing Orientation Scale (developed by the authors; COS) and the Self-Consciousness Scale (Fenigstein et al. 1975: public and private self-consciousness) were administered to

NII-ELS(学術コンテンツ登録システム)では、「日、ヨミ、英」のメタデータをCiNiiに提供できる。

→現在のjunii2では無理。

機関リポジトリとNII-ELSは、相互補完的??

KAKEN

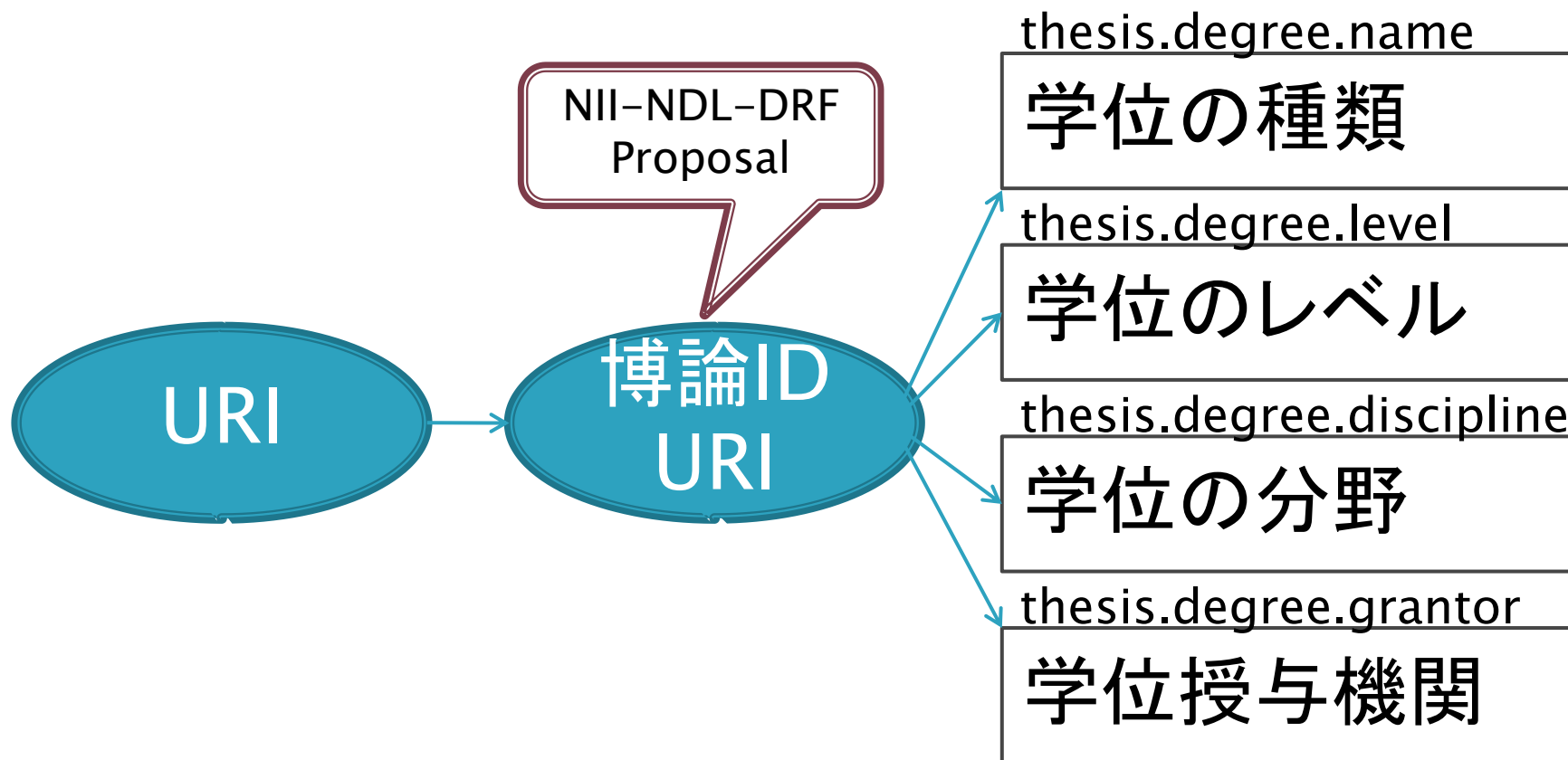
```
Mozilla Firefox
http://kaken.nii.ac.jp/0003/2008/8/ja

-<rdf:RDF xml:base="http://kaken.nii.ac.jp/ns">
  -<kaken:Project rdf:about="http://kaken.nii.ac.jp/p/17200003">
    -<dcterms:abstract>
      -<kaken:Document rdf:about="http://kaken.nii.ac.jp/d/p/17200003">
        -<dcterms:hasPart>
          -<kaken:Document rdf:about="http://kaken.nii.ac.jp/d/p/17200003/2008/8/ja">
            <dcterms:hasFormat rdf:resource="http://kaken.nii.ac.jp/pdf/2009/seika/jsps%2d1/13801/17200003seika.pdf"/>
            -<dcterms:title xml:lang="ja">
              KAKEN - アンテナの指向性を利用するユビキタスインフラストラクチャに関する実証的研究(17200003) - 2008年度
            </dcterms:title>
            <dcterms:date rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">2008</dcterms:date>
          </kaken:Document>
        </dcterms:hasPart>
      </kaken:Document>
    </dcterms:abstract>
    <kaken:projectNumber>17200003</kaken:projectNumber>
    <kaken:projectTitle xml:lang="ja">アンテナの指向性を利用するユビキタスインフラストラクチャに関する実証的研究</kaken:projectTitle>
  -<kaken:projectTitle xml:lang="en">
    Empirical Study on Ubiquitous Infrastructure using Antenna Directionality
  </kaken:projectTitle>
  <kaken:yearFrom rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">2008</kaken:yearFrom>
  <kaken:yearTo rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">2008</kaken:yearTo>
  -<kaken:member>
    -<kaken:Researcher rdf:about="http://kaken.nii.ac.jp/r/90201201">
      <foaf:lastName xml:lang="ja">渡辺</foaf:lastName>
      <foaf:firstName xml:lang="ja">尚</foaf:firstName>
      <foaf:lastName xml:lang="en">WATANABE</foaf:lastName>
      <foaf:firstName xml:lang="en">Takashi</foaf:firstName>
      <rns:affiliationTitle xml:lang="ja">静岡大学・創造科学技術大学院・教授</rns:affiliationTitle>
    -<kaken:role>
      -<kaken:Role>
        <kaken:yearIn rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">2008</kaken:yearIn>
        <kaken:roleCode>1</kaken:roleCode>
        <kaken:roleName xml:lang="ja">研究代表者</kaken:roleName>
      </kaken:Role>
    </kaken:role>
    <kaken:researcherNumber>90201201</kaken:researcherNumber>
  </kaken:Researcher>
</kaken:member>
-<kaken:member>
  -<kaken:Researcher rdf:about="http://kaken.nii.ac.jp/r/60395043">
```

研究報告書

- ▶ 少なくともIRに登録された「科学研究費補助金成果報告書」については、メタデータに研究課題番号(URI)を定義することによりIRとKAKENの相互Linkの可能性が高まる。
- ▶ 助成金を得て書かれた論文等についても、助成金情報を入力するか？
 - 例えばKAKEN「発表文献」との関係

学位論文



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

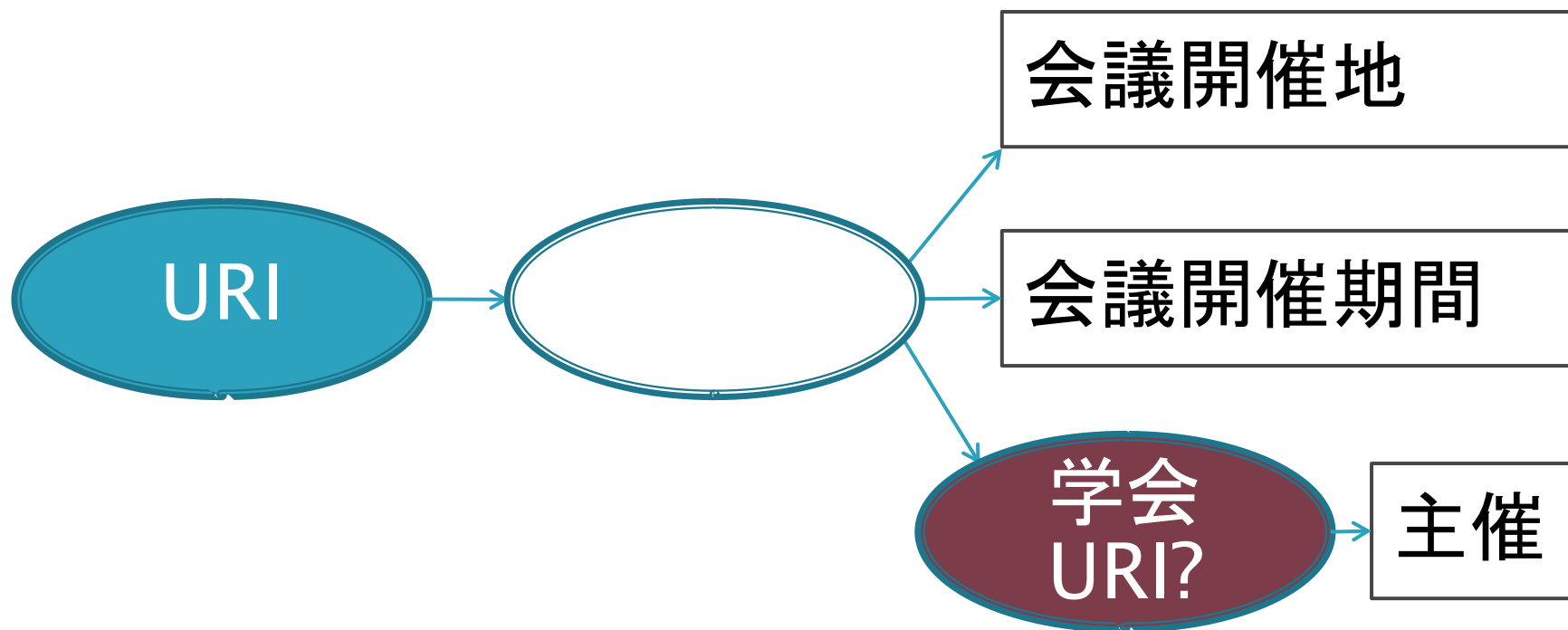
NII-NDL-DRF

Metadata Elements for Dissertaion:Proposal

- ▶ 識別子 機関番号(JST学校機関コード) +
区分番号(甲>A,乙>B,旧制>Z) +
報告番号(6桁)
- ▶ 国立国会図書館・・・デジタル化に係る複製物についてメタデータをDC-NDLで大学に提供予定
- ▶ NII・・・学位論文のIR横断サービスは
DBR? それともJAIRO上か。

当提案の実効性は？

会議情報



会議情報はセマンティックな表現がしやすい。

リテラルからリソースへ

- ▶ NIIが定義したもの
 - NIItype (NII資料種別)
 - NCID (NII書誌ID)
 - CAT-APIの開発
- ▶ 外部で定義されたスキーム
 - ISSN
 - 一般的にはURN
 - NDC
 - どのような名前空間を使うか？作るか？

書誌情報(SIST02)に準拠するならば

- ▶ 「図書」と「図書の章」(粒度異なる)
 - ▶ 図書の総ページ数
 - ▶ ISBN
 - ▶ テクニカルレポート番号
 - ▶ edition, version
 - ▶ 掲載図書名
 - ▶ 雑誌特集標題
 - ▶ 記事番号
- ... etc. IRのメタデータをよりリッチにする必要がある

課題

- ▶ 構造化したメタデータへの変換・実装はカスタマイズで可能 → 技術をどのように確保し、コストをどのように抑えるか？
 - × 現在の大学図書館に開発能力を期待するには無理がある。(高コスト)
- ▶ これまでIRに登録されたメタデータの置換や追加入力する作業が生じる
 - × それまで体力のある図書館は希少

課題

- ▶ メタデータをバージョンアップすることによる大学側の具体的利益は？説明できる論拠が必要。

→ 図書館連携作業部会WGやCSI委託機関等と連携しながら、NIIを中心とする数機関で開発実装のプロジェクトを進め、webの動向を見極めつつ、改善を重ねながら普及のタイミングを計ってはどうか。

おわりに

- ▶ 研修の感想
- ▶ 今後に向けて
- ▶ 謝辞