

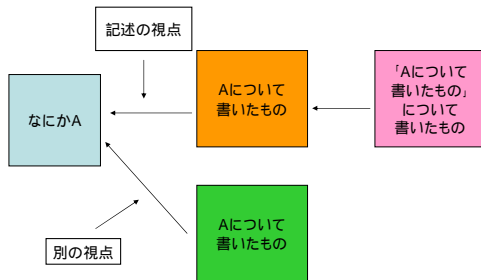
メタデータについて議論してみよう

杉本重雄
知的コミュニティ基盤研究センター
図書館情報メディア研究科
筑波大学

目的と進め方

- メタデータに関する自由な議論がしてみたい
- メタデータそのものの概念はとても広いので、はじめは、メタデータについての基本概念を述べ、つづいて、本研究室で扱っているメタデータについて述べる。
- そのあと、筆者にとっての最近の研究的に関心のあるトピックをいくつか挙げて、議論したい

メタデータ



メタデータ

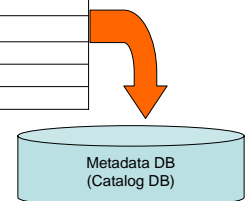
- メタデータの例
 - レストランのメニュー、おもちゃに書いてある想定利用者の年齢、ソフトウェアの利用条件、食品の内容、お店の看板など、いろいろ
- 筆者の研究室で扱ってきたメタデータ
 - 情報資源の記述と発見のためのメタデータ
 - 目録、書誌情報のようなもの
 - 利用者、利用者環境の特性のメタデータ
 - 資源の保存のためのメタデータ
- 主にメタデータの記述規則に関する研究

メタデータとメタデータスキーマ

- メタデータを書くためには記述規則が必要
- メタデータの記述規則
 - 形式を決める
 - 実際に記述の指針を決める
- メタデータスキーマ
 - メタデータの記述形式を決めるもの

メタデータスキーマについて: 具体表現の例

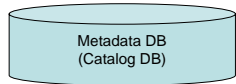
Title (en)	Digital Library Researches in Tsukuba:
Creator (en)	Sugimoto, Shigeo
Creator (ja)	杉本, 重雄 (pronunciation: すぎもと, しげお)
Subject (en)	Metadata Schema
Subject (en)	Interoperability
Publisher (en)	Shanghai Library
Publishing Location (en)	Shanghai, China
Date (ISO-8601)	2004-10-13
Language (ISO-639)	en



メタデータスキーマについて: 具体表現の例

```

<dc:title> Digital Library Researches in
  Tsukuba: </dc:title>
<dc:creator> Sugimoto, Shigeo</dc:creator>
<dc:creator xml:lang="ja">
  <rdf:value>杉本, 重雄 </rdf:value>
  <uliscore:pronunciation>すぎもと, しげお
  </uliscore:pronunciation>
</dc:creator>
...
  
```



メタデータスキーマを考える: 構造的な制約

実現形式には依存しない制約

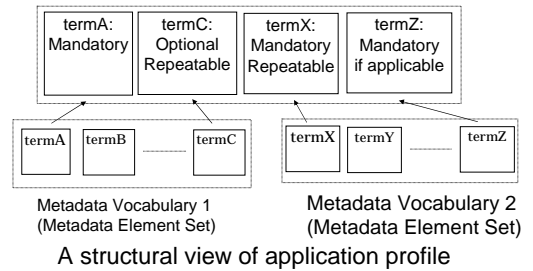
- Title:** Mandatory
- Creator:** Mandatory if Applicable, Repeatable
- Subject:** Optional, Repeatable
- etc.

メタデータスキーマを考える: エlement 基本的要素 (記述項目と値の語彙)

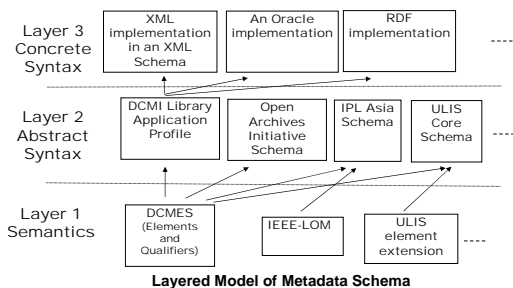
Title
Element Name: Title
Label: Title
Definition: A name given to the resource.
Comment: Typically, Title will be a name by which the resource is formally known.

Creator
Element Name: Creator
Label: Creator
Definition: An entity primarily responsible for making the content of the resource.
Comment: Examples of Creator include a person, ...

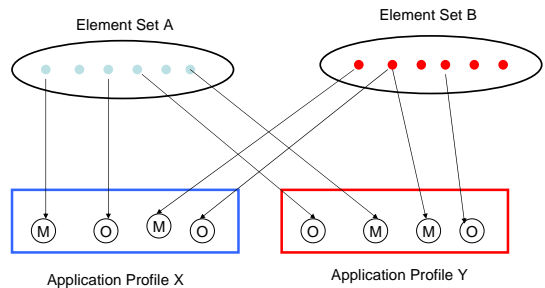
Application Profile



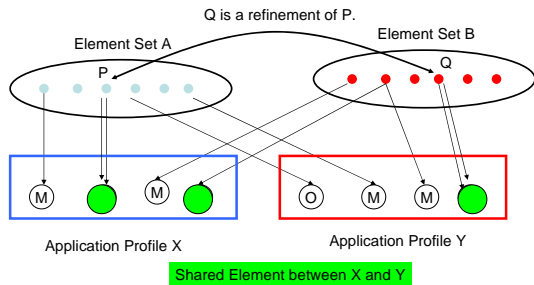
メタデータスキーマを階層的に見てみると



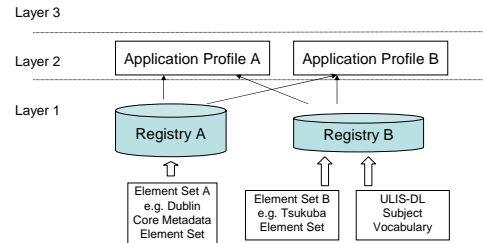
Application Profileと相互運用性



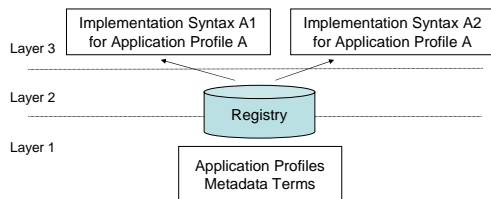
Application Profileと相互運用性



階層モデルとメタデータスキーマレジストリ



階層モデルとメタデータスキーマレジストリ



主題語彙に関して

- ULIS-DLの経験から
 - 中味を一覧できるようにしたい
 - 適切な見出し語を用意したディレクトリが作りたい
 - 収集したリソースに対し、既存の語彙(分類、件名標目表)は適用しやすくない

ULIS-DLでの主題語彙の評価

- 2003年までに作られたメタデータのSubjectエレメントから取り出したキーワード(生データで約16000語)を正規化後、何語でどれだけのメタデータ件数をカバーできるかを評価してみた。

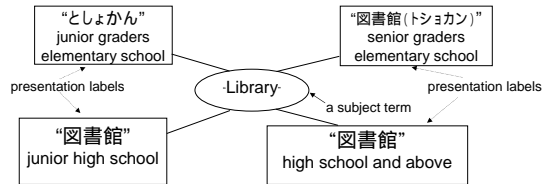
	キーワード	カバーできなかったメタデータ件数	非カバー率
CTS-2	3979	1519	6%
CTS-3	2045	2083	8%
CTS-4	1366	2590	10%
CTS-5	1025	2801	11%

岡山県立図書館のデジタル岡山大百科

- デジタル岡山大百科
 - 県立図書館による地域ポータル
 - 図書館員のほか教員や地域のボランティアも参加
 - Simple Dublin Coreを基礎
 - 主題分類の語彙
 - NDC
 - 子ども分類
 - 夢作り分類

子ども分類

- ひとつの概念に対して複数のラベル



主題分類の語彙の比較

- NDCと子ども分類、夢作り分類の比較
- 主題分類はディレクトリを作るのに必要
 - Web上で発信される資料に対してNDCは必要か
 - 独自に分類を作ると後のメンテナンスが大変

分類の比較 (子ども分類 = KV、夢作り分類 = PV)

NDC	000	100	200	300	400	500	600	700	800	900	total
#Kids terms	17	8	8	196	58	54	28	62	6	6	443
#NDC in KV	7	7	3	44	27	27	20	26	4	4	169
#Pref terms	15	2	12	171	30	56	44	17	1	1	349
#NDC in PV	4	2	5	34	11	18	25	15	1	1	116

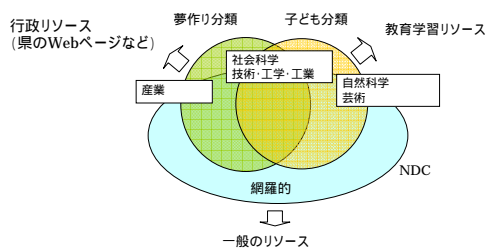
NDC: 000=総記, 100=哲学, 200=歴史, 300=社会科学, 400=自然科学, 500=技術・工学・工業, 600=産業, 700=芸術, 800=言語, 900=文学

Distribution of Terms in the NDC term space - Okayama's Case -

NDC	000	100	200	300	400	500	600	700	800	900	total
#Kids terms	17	8	8	196	58	54	28	62	6	6	443
#NDC in KV	7	7	3	44	27	27	20	26	4	4	169
#Pref terms	15	2	12	171	30	56	44	17	1	1	349
#NDC in PV	4	2	5	34	11	18	25	15	1	1	116

NDC: 000=総記, 100=哲学, 200=歴史, 300=社会科学, 400=自然科学, 500=技術・工学・工業, 600=産業, 700=芸術, 800=言語, 900=文学

主題分類の比較



自治体の分類の語彙の比較

Web上に出ている分類表から

	茨城県	香川県	神奈川県	岡山県
トップレベル	23	11	14	16
深さ	2	2	2	3
語数	115+	82	88	280+

主題語彙について

- 対象領域にあった主題語彙がほしい
- 作ること、メンテすること、再利用することのできるだけ簡単におこないたい

Identifierは何を指す？

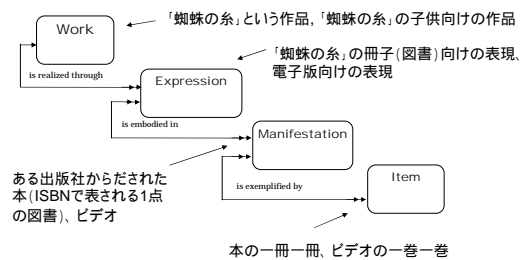
- ISBNの場合
 - 本につけられた番号
 - 書棚に並べられた一冊一冊の個々の本ではなく、出版物としての1点に対して与えられる。
- 電話番号の場合
 - 回線につけられた番号
 - 電話機につけられた番号ではない
 - 転送される場合もあり、実際にどこにつながるかはわからない。

Identifierは何を指す？

- URLの場合
 - <プロトコル> <ドメイン> <ドメイン内の場所・パラメータ>
 - ドメイン名はIPアドレスに対応付けられている
 - <ドメイン内の場所・パラメータ>の部分の解釈はサーバに任される。
- Identifierは、それを解釈するもの(resolver)があってはじめて意味を持つ。

FRBRによる実体間関係のモデル

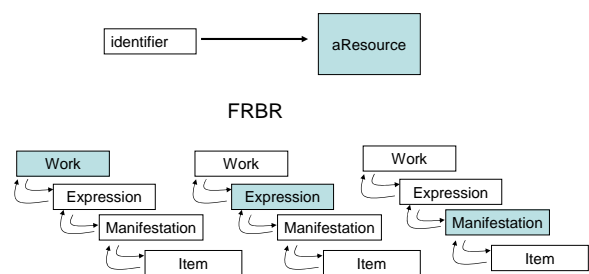
メタデータの記述対象は何か？

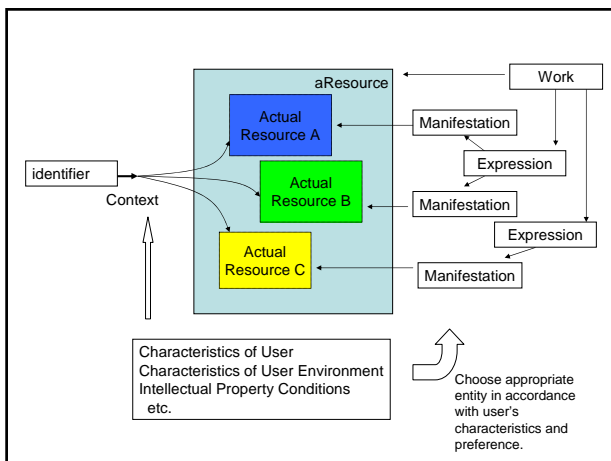
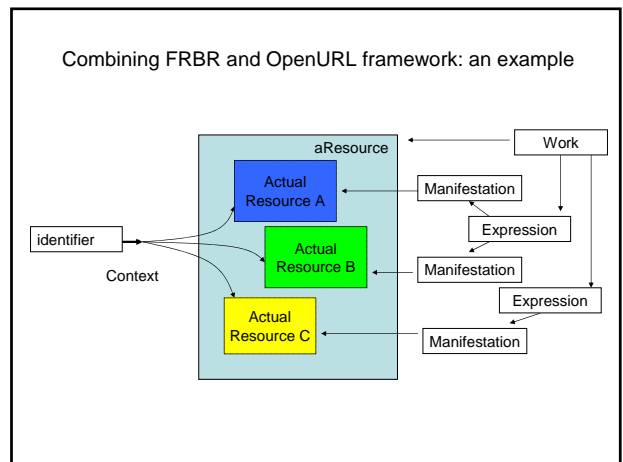
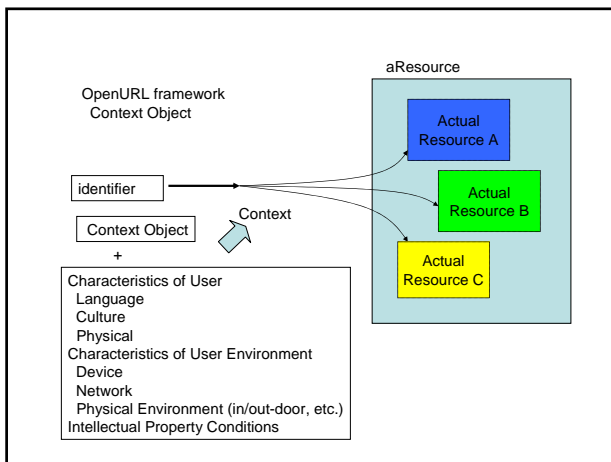
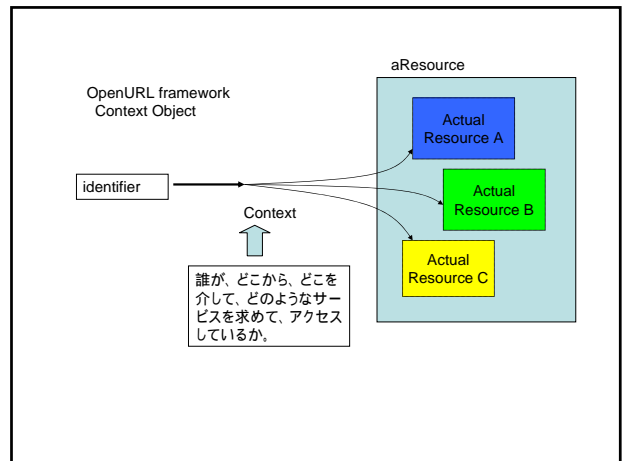
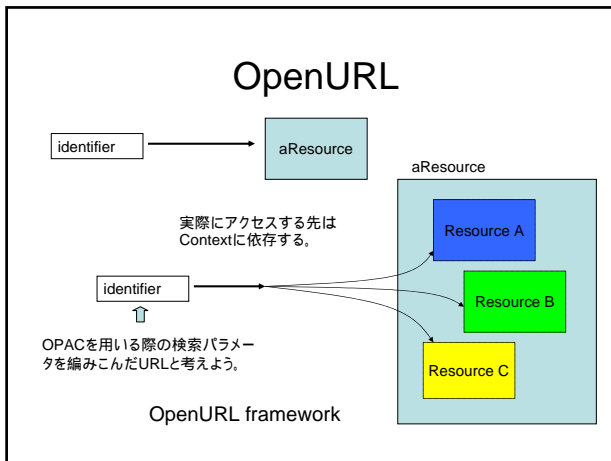


「何」に対してメタデータを書くか

- IFLAのFRBRモデル
 - 書誌記述のためのモデル
 - 4段階でとらえる
 - Work: 表現を限定しないで識別でされる作品
 - Expression: Workをなんらかの方法で表現したもの
 - Manifestation: Expressionを何らかの媒体上に実現したもの
 - Item: Manifestationの個別の実体

Identifierは何を指す？





Accessibility, FRBR, OpenURL

- Accessibilityを良くするには、利用者、利用環境に応じて、最も適切なリソースを探し出し、適切に提供することが必要である。
- OpenURLはコンテキストにあったアクセスのための基本モデルを与えてくれる
- FRBRは「同じ内容」のリソースの間の関係性を与えてくれる
- Accessibility Metadataは利用者、利用環境の特性の記述の枠組みを与えてくれる
- これらをくみあわせてみればどうなるか？

メタデータについて議論してみよう

- メタデータはどのように作る
- メタデータの語彙
 - エlementセット
 - 主題語彙、分類
- メタデータスキーマのモデルとレジストリ
 - Global Interoperability vs. Domain Specificity
 - 相互運用性とSemantic Webの技術
- そのほか

メタデータはどのように作る？

- メタデータの作り方
 - 専門家による (Librarian/Cataloger model)
 - 機械による (Software-generated Metadata model)
 - リソース作成者による (Submission model)
- 専門家による作成は高くつき、大量のものを作ることが難しい
- ソフトウェアによる作成は品質に問題
- リソース作成者によるものは、品質の問題が残る

主題語彙の問題

- 大きな語彙vs小さな語彙
 - Dublin Coreは小さな語彙
 - 大きな語彙(MPEG7, MARC, etc.)
- 主題語彙に関する素朴な疑問
 - 図書館でよく使われる語彙はそのままWebリソースに使えるか
 - 素人利用者に使いやすいか
 - 見せ方をうまくできないか

メタデータスキーマのモデルとレジストリ

- 相互運用性はメタデータのキーワード
 - 相互運用のために共通化を進めたい
 - 応用毎にできるだけ適した記述をしたい
- Application Profileの役割
 - いろいろなElementセットから必要な項目を取り出せる
 - 異なるスキーマを組み合わせる際のキー

メタデータスキーマのモデルとレジストリ

- Semantic Webの技術
 - メタデータスキーマのベースは語(語彙)の定義
 - 定義した語を共有したい
 - 定義した語を再利用したい
 - 定義した語をメンテナンスしやすくしたい
 - Dumb-downとCrosswalkのための仕組み

その他

- Dublin Coreのはたした役割
 - Dublin Coreはシンプルなメタデータとして有名になった
 - 「15項目だけ!」
 - 本当にシンプルか?
 - Application Profileは重要な概念
 - RDFと組み合わせることで、メタデータの相互運用性を支える基盤概念
 - 日本国内で求められるものは?
 - (頼りになる)翻訳?
 - 組織?
- そのほか