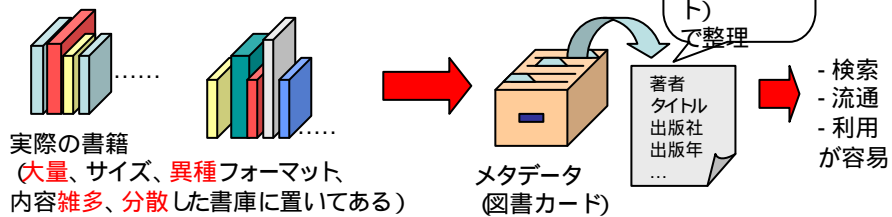


メタデータとその活用

INTAP セマンティックWeb委員会委員
 (株)富士通研究所
 津田 宏
 htsuda@jp.fujitsu.com

メタデータとは？

- データに関する情報を表すデータ
 - 属性情報
 - 他オブジェクトとの関係情報
- メタデータとは例えば、
 - 書籍に対する書誌情報
 - タイトル、作者、出版社、概要、....



メタデータ:何に使える?

INTAP

- 検索
 - 大量、異種形式、雑多、言葉で検索できない
 - 図書検索、Web情報検索、画像検索、...
- 交換 統合
 - 異種、分散したデータのとりまとめ、所在情報管理
 - クリアリングハウス、電子政府公開情報、地理情報システム (GIS)...
- 流通 配信
 - データの作成・利用法の管理
 - フィルタリング、デジタル資産管理(DAM)、...

3

(c)2002 INTAP. All rights reserved.

メタデータ規格の色々

INTAP

- 汎用
 - DC(Dublin Core), RSS,
- マルチメディア
 - MPEG7
- 電子政府
 - MIReG, e-GMS, AGLS, e-Gov
- E-learning
 - LOM(Learning Object Metadata), SCORM, LIP
- ニュース、テレビ放送
 - XMLNews, NewsML, TV Anytime, ARIB,
- 音楽
 - MusicBrainz, ...
- 地理 観光情報
 - G-XML, JMP, ...
- フィルタリング
 - PICS
- ユーザプロフィール
 - P3P
- コンテンツ管理
 - cIDF (Content ID Forum),

.....

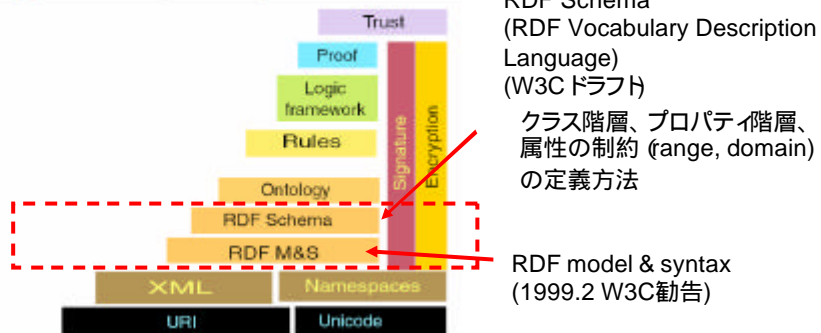
4

(c)2002 INTAP. All rights reserved.

セマンティックWebとメタデータ INTAP

– 個々のメタデータでなく メタデータの記述形式 (入れ物)を規定

Enabling Technologies - Layer Cake



(WWW2002, W3C trackより
<http://www.w3.org/>)

RDF = Resource Description Framework
 メタデータの記述モデル(3つ組)と
 流通のためのXML表現

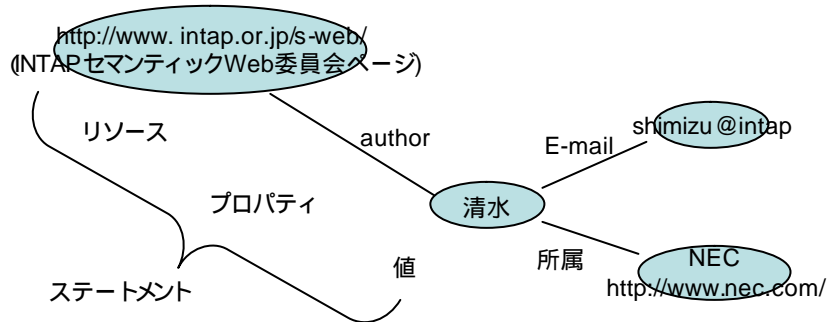
5

(c)2002 INTAP. All rights reserved.

RDF (model) INTAP

• RDF (Resource Description Framework)

– (リソース, プロパティ, 値) の三つ組みでメタデータを表そうというモデル



6

(c)2002 INTAP. All rights reserved.

RDF (syntax)

- 交換・流通のためにXML構文を持つ

```
<rdf:RDF>  
<rdf:Description about="http://www.intap.or.jp/s-web/">  
  <s:author>清水</s:author>  
</rdf:Description>  
</rdf:RDF>
```

このURLに関して、以下の属性がある

値がさらにリソースとなる場合

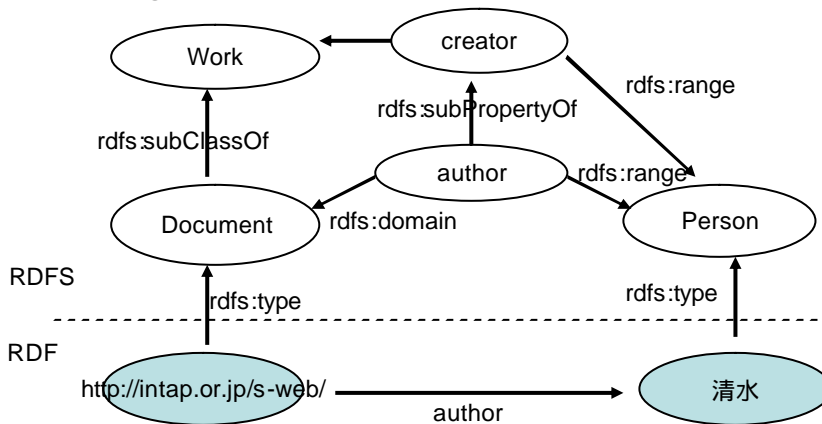
```
<s:author>  
<rdf:Description about="http://www.intap.or.jp/id/1716/">  
  <v:name>清水</v:name>  
  <v:Email>shimizu@intap</v:Email>  
</rdf:Description>  
</s:author>
```

or

```
<s:author  
  rdf:resource="http://www.intap.or.jp/id/1716/"  
  v:name="清水" v:Email="shimizu@intap/">
```

RDF スキーマ (RDFS)

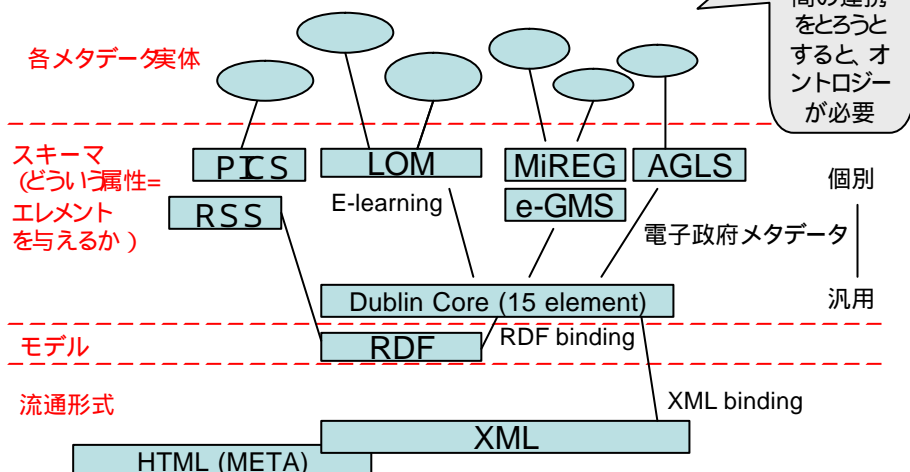
- クラス階層、プロパティ階層、値の制約 (range, domain)などを定義



SW関連メタデータ **INTAP**

- RSS (RDF Site Summary or Rich Site Summary)
 - Webサイトの要約をRDFのメタデータとして記述することで、検索・配信を容易にする
 - <http://purl.org/rss/1.0/spec/>
- PICS (Platform for Internet Content Selection)
 - 有害コンテンツのフィルタリング。ページにメタデータ(PICSラベル)を付与し、クライアントでブロック。
- P3P (Platform for Privacy Preferences Project)
 - 個人のプライバシー情報伝達のための枠組み。各サイトがプライバシーポリシーをメタデータとして公開(XMLベース)。利用者はそれを見て(半自動で)個人情報公開してよいかを判断。

メタデータの記述階層 **INTAP**



Dublin Core (ダブリン・コア)

INTAP

- ネットワーク上のものも含めた情報資源の、基本的なメタデータ要素(エレメント)。
- 1995年、米オハイオ州ダブリンで開催された国際会議の結果が元となり、このように命名
- DCMI (Dublin Core Metadata Initiative)
 - <http://dublincore.org/>
- 15の属性: (Dublin Core Metadata Element Set : DCMES)
 - Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights

11

(c)2002 INTAP. All rights reserved.

各国電子政府メタデータ

INTAP

- 行政公開文書の、国/省庁横断的管理
- 各国で、DCMESを拡張してエレメントを定義
 - (欧州)
 - 英e-GMS (e-Government Metadata Standard)
 - EU MIREG (Managing Information Resources for e-Government)
 - デンマークOIO-metadata (Offentlig Information Online)
 - アイルランドPSMS
 - (豪州)
 - オーストラリア AGLS (Australian Government Locator Service)
 - ニュージーランド NZGLS (New Zealand GLS)
 - (米)
 - GILS (Government Information Locator Service)
- さて、日本は？ cf. 電子政府の総合窓口

12

(c)2002 INTAP. All rights reserved.

メタデータ:落とし穴 1

- **メタデータデッドロック**
 - 「メタデータがあればこんな良いサービスが作れるのに...」VS 「こんなサービスがあるのだったら、ちょっと大変だけどメタデータを作っても良いのに...」
 - 一般に、メタデータを作るのは大変なうえ、みんながやらないと対費用効果も見えにくい
 - [対策] メタデータ(半)自動化技術
 - メタデータエディタ
 - メタデータジェネレータ 情報抽出、自動分類技術 (自然言語処理技術, AI技術)の応用ターゲット
 - 半自動化 → サービスを早期立ち上げ → メタデータがさらに増える → サービスさらに発展 → ... というポジティブなループの第一歩

13

(c)2002 INTAP. All rights reserved.

メタデータ:落とし穴 2

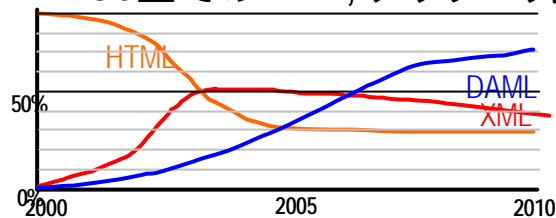
- **メタデータのトラスト**
 - HTML META タグはなぜ機能しなかったか?
 - ワードスパム攻撃
 - 良く検索される語をMETAタグに大量に入れることで、サーチエンジンの結果を騙してページ露出度を上げる攻撃
 - 結果として、今のサーチエンジンはメタデータを活用できていない
 - Internet: 性悪説なので、Tim Berners-Leeの階層最上位の「Trust」が伴わないとだめ
 - 多分、「RDFスパム」は必ず起こるが、どう対処する?
 - [対策]
 - 信頼できるコミュニティ(イントラネット、エクストラネット)からすすめる。電子政府など閉じた応用。

14

(c)2002 INTAP. All rights reserved.

これからはメタデータの時代 (か?) INTAP

- メタデータの種類、用途の広がり。特に検索以外での必要性がでてきた
- Webページ数の増加 :ますます検索は困難
 - 3.2億(98.4) 8億(99.7) 10億(00.1) 21億(00.7) 40億以上? (01 ~)
- Web上でのXML, メタデータ比率の増加予想



(by Mury Burke,
SWMU2, 2001.11)

15

<http://www.daml.org/meetings/2001/11/swmu2/>

© 2002 INTAP. All rights reserved.