

Slide URL

<https://vu5.sfc.keio.ac.jp/slide/>

Web情報システム構成法

No.14 RDFとLOD

萩野 達也 (hagino@sfc.keio.ac.jp)

文書 vs データ

▶ Web上の文書

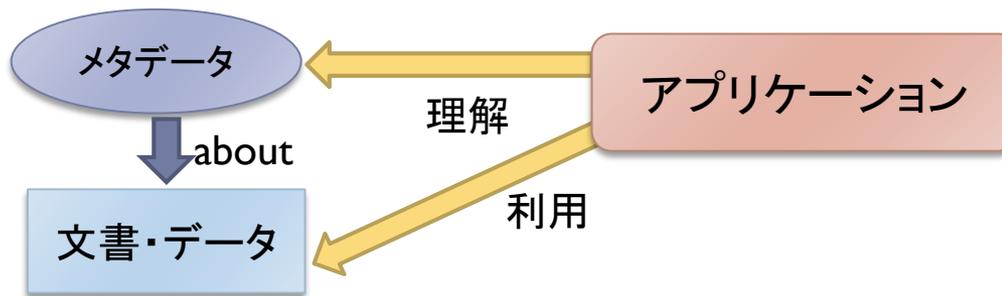
- ▶ インターネット上のハイパーテキストシステムとして成功
- ▶ HTMLは広く使われるようになった
- ▶ 人が読む文書が大量にある(ありすぎ?)
- ▶ HTMLをインターフェイスとする便利なアプリケーションもある
- ▶ 文書の意味を理解せずに検索エンジンが処理をして関連するページを見つける

▶ Web上のデータ

- ▶ 計算機可読なデータ
- ▶ Web上のデータはまだ活用されていない

セマンティックWebとは

- ▶ データのWeb
 - ▶ 「文書のWeb」から「データのWeb」へ
- ▶ メタデータ
 - ▶ メタデータ = 文書やデータに関するデータ
 - ▶ 計算機可読なメタデータをアプリケーションで共有する
 - ▶ データの共有や統合を可能にする



Information Management: A Proposal

▶ Webの提案

- ▶ Tim Berners-Lee
- ▶ 1989年
- ▶ リンクには意味がある
- ▶ リンクされたデータを処理することで新たな知識を得る

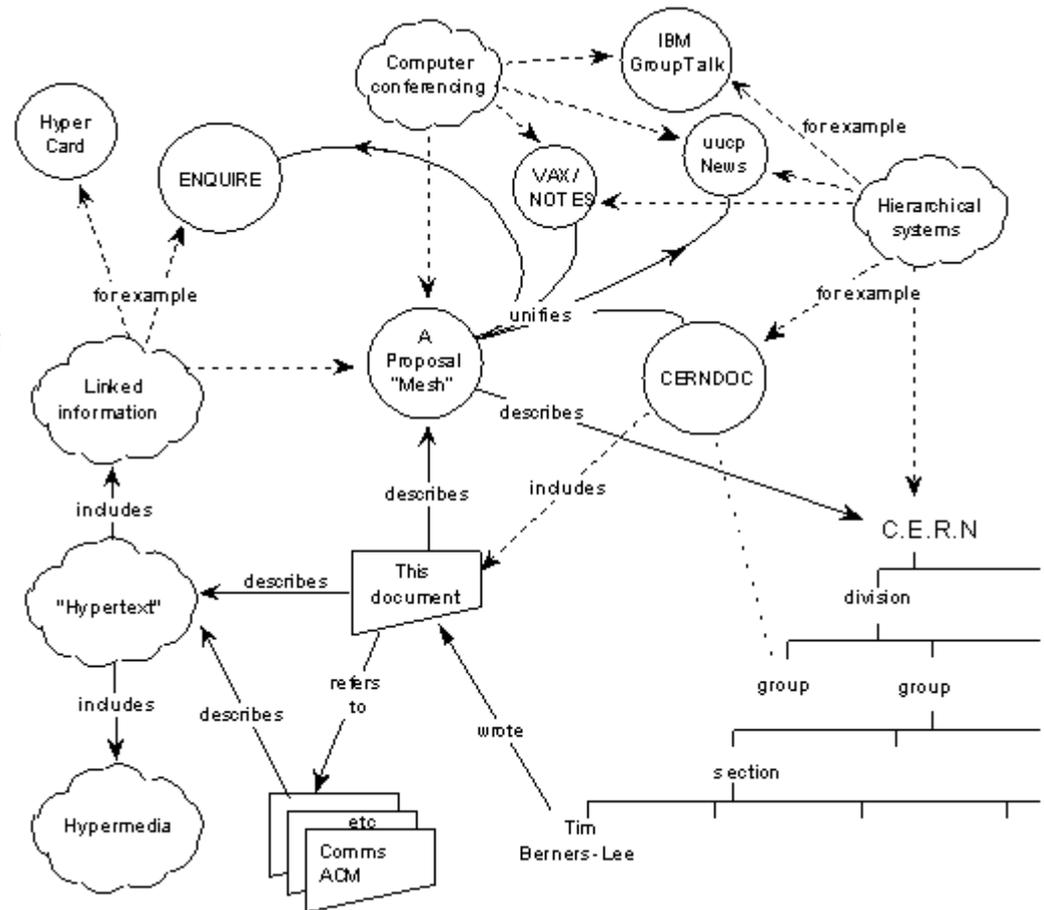


▶ セマンティックWeb

- ▶ 1998年

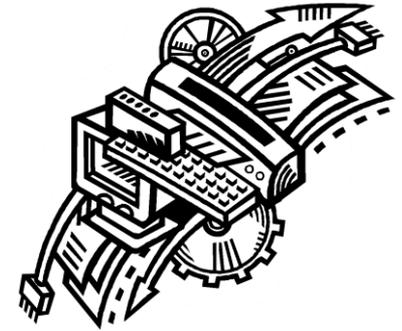


▶ データWeb



計算機可読データとは？

- ▶ 計算機可読データ
 - ▶ データが標準形式で書かれている
 - ▶ データを機械が自動的に(人手を介さずに)に処理できる
 - ▶ データに意味が付いている
- ▶ PDF文書の中の表は計算機可読か？
 - ▶ データ形式的に計算機可読ではない
 - ▶ 図と表の区別がつかない
- ▶ HTMLのテーブルのデータは計算機可読か？
 - ▶ データ形式的には計算機可読であるが，
 - ▶ 単純に表として表示する目的において
 - ▶ 意味的には計算機可読ではない
 - ▶ データ形式的にも複雑すぎる
- ▶ CSVファイルは計算機可読か？
 - ▶ データ形式的には計算機可読であるが，
 - ▶ 意味的には計算機可読ではない
 - ▶ ヘッダ行では十分ではない
 - ▶ 列の間の意味が分からない



計算機 東海道本線 横浜-東京方面(上り)		平日		土曜・休日の時刻表を表示	
5	00 27 50				
6	快高 02 11 19 27 34 40 44 47 52 55 58	快高	高		
7	快高 01 04 07 10 13 16 19 22 25 29 31 34 37 40 46 49 51 54 58	快高	快高	快高	快高
8	01 04 09 12 15 19 22 27 30 35 43 51 54	快高	高	高	
9	01 10 23 33 45 50 55				
10	05 15 26 38 49 59	快高			
11	14 32 50 53	快高			
12	02 14 35 44 50				
13	03 15 26 32 39 50 59	快高			
14	07 15 25 38 50 54	快高			
15	04 17 32 40 50 58	快高			
16	02 16 26 29 42 50 55	快高			
17	07 14 19 24 34 46 51 54	快高	快高		
18	05 11 22 26 30 35 45 50 57	高	高	高	
19	08 20 23 32 42 48 54 58	快高	快高		
20	09 21 25 37 42 50	快高	快高		
21	00 14 21 25 39 51	快高			
22	04 19 25 33 39 45 58	快高			
23	高 12 27 39	高	高		

Web上の分散情報



ID	名称	住所	...
A001	山田歯科	藤沢	...
B031	萩野歯科	藤沢	...
D023	佐藤デンタルクリニック	武蔵藤沢	...

Web

住所	県
藤沢	神奈川県
武蔵藤沢	埼玉県
...	...

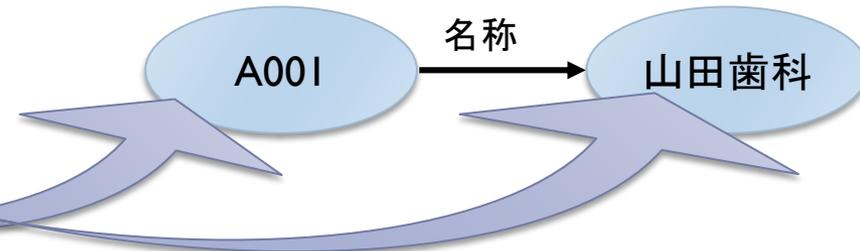


ID	休診日
B031	木曜, 日曜
D023	土曜, 日曜

表をRDFに変換する

- ▶ 表の各行を三つ組みで表す
 - ▶ (資源, プロパティ, 値)
 - ▶ 資源: 主キー
 - ▶ プロパティ: 列のヘッダ
 - ▶ 値: セルの値

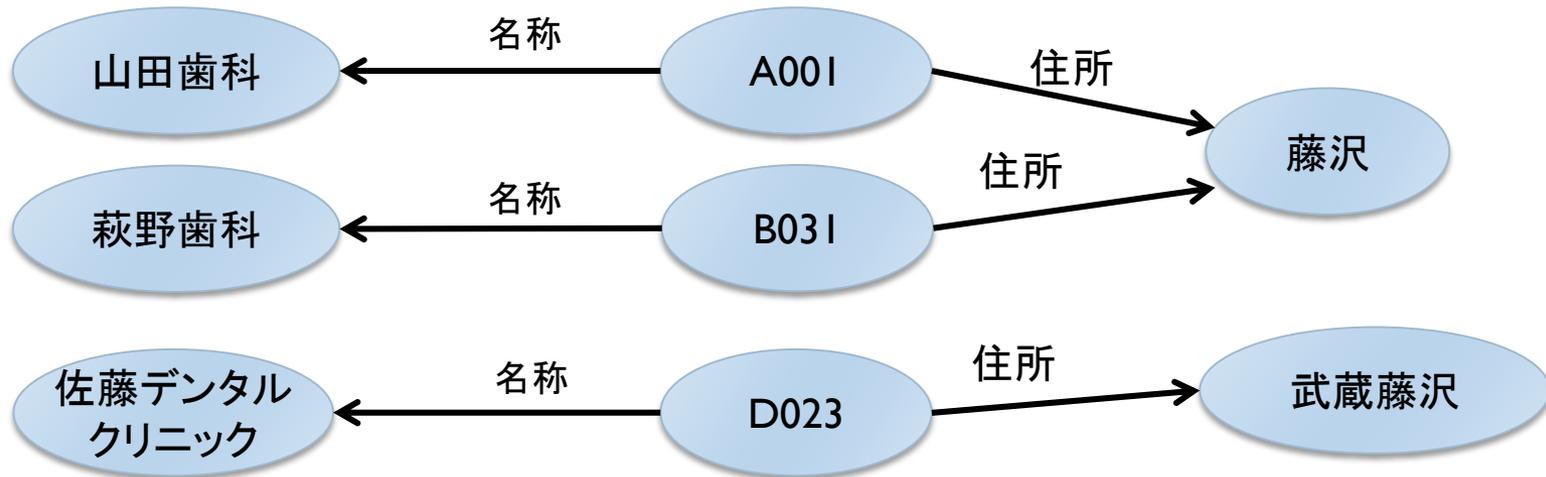
ID	名称	住所	...
A001	山田歯科	藤沢	
B001	森野歯科	藤沢	
D023	佐藤デンタルクリニック	武蔵藤沢	



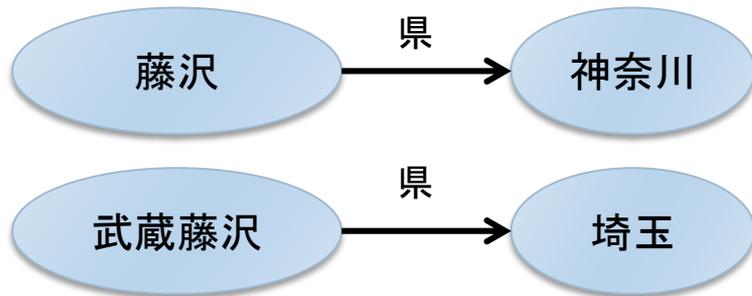
Convert Table to RDF Graph

▶ 同じ資源は共有される

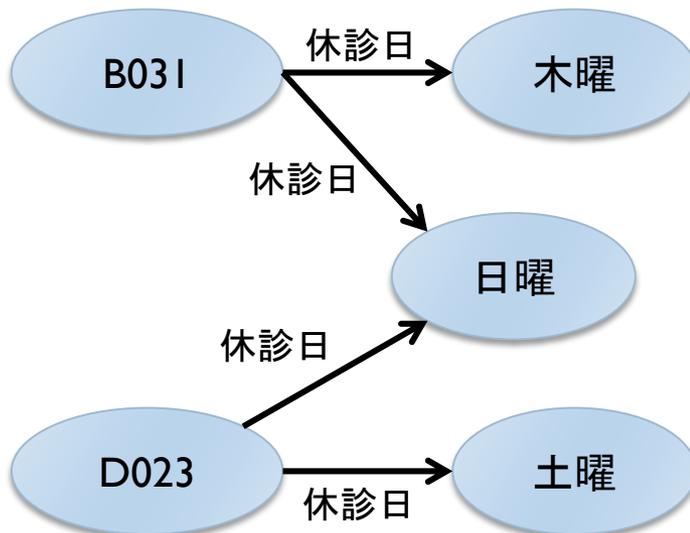
ID	名称	住所	...
A001	山田歯科	藤沢	...
B031	萩野歯科	藤沢	...
D023	佐藤デンタルクリニック	武蔵藤沢	...



Convert Table to RDF Graph (cont.)



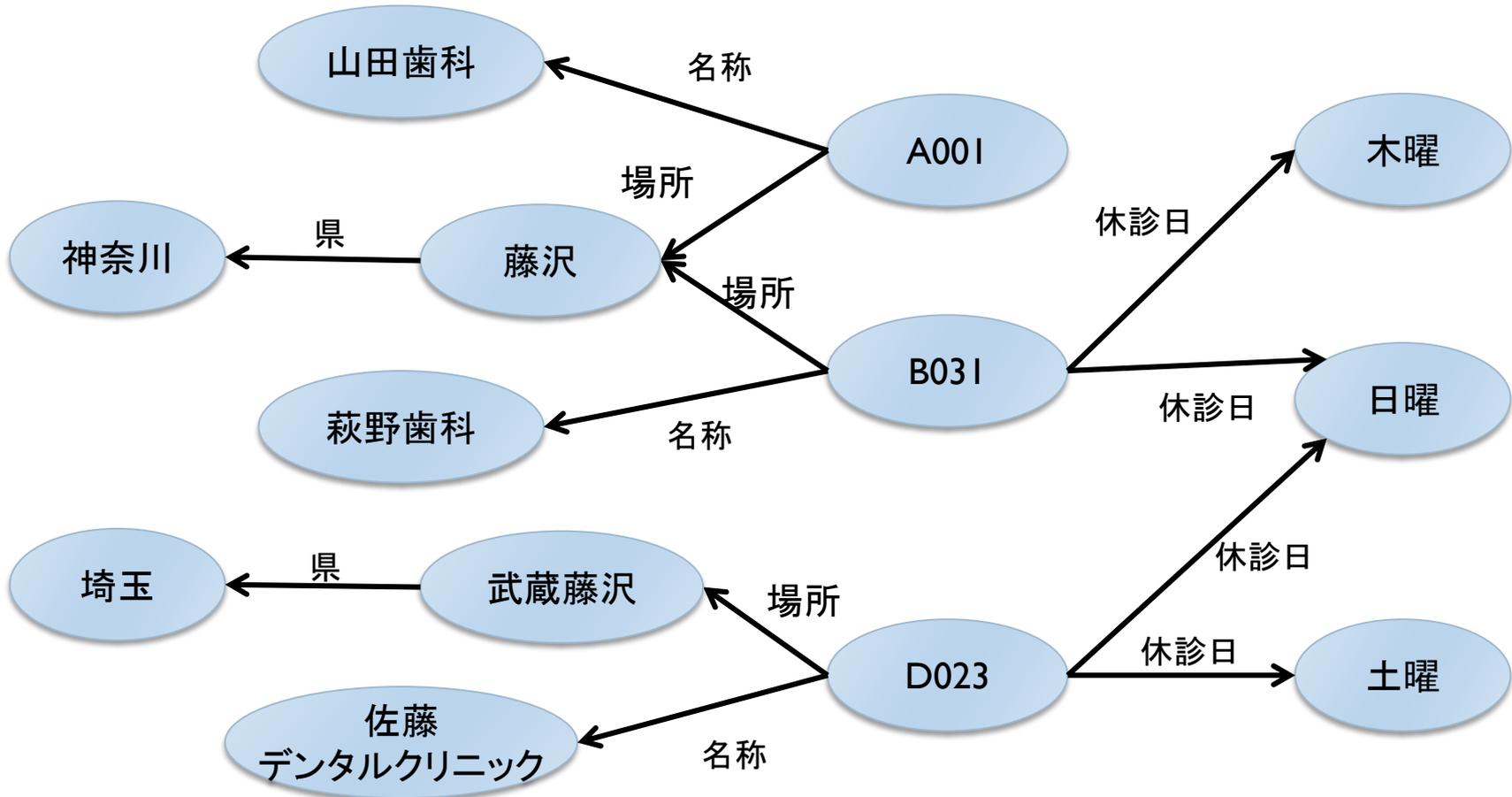
住所	県
藤沢	神奈川
武蔵藤沢	埼玉
...	...



ID	休診日
B031	木曜, 日曜
D023	土曜, 日曜

複数のデータ資源を統合する

- ▶ 同じノード(資源)は自動的に共有される
- ▶ 表の結合(join)は必要ない



URIを使う

- ▶ 同じノード(資源)を共有することでデータを統合する
 - ▶ Web資源はURIを持つ
 - ▶ <http://example.org/dentist.rdf#B031>
 - ▶ 名前空間を用いると簡潔
 - ▶ `ex:B031`
 - ここでexは<http://example.org/dentist.rdf#>にバインド
- ▶ プロパティは語彙である
 - ▶ AAAの原則: Anybody can say anything about any topic.
 - ▶ だれでもが自由に語彙を使うことができる(制限がない)
 - ▶ プロパティもURIを用いる
 - ▶ 例: titleプロパティはダブリンコアで定義されている
 - ▶ <http://purl.org/dc/terms/title>

<http://web.sfc.keio.ac.jp/~hagino/wwb23>

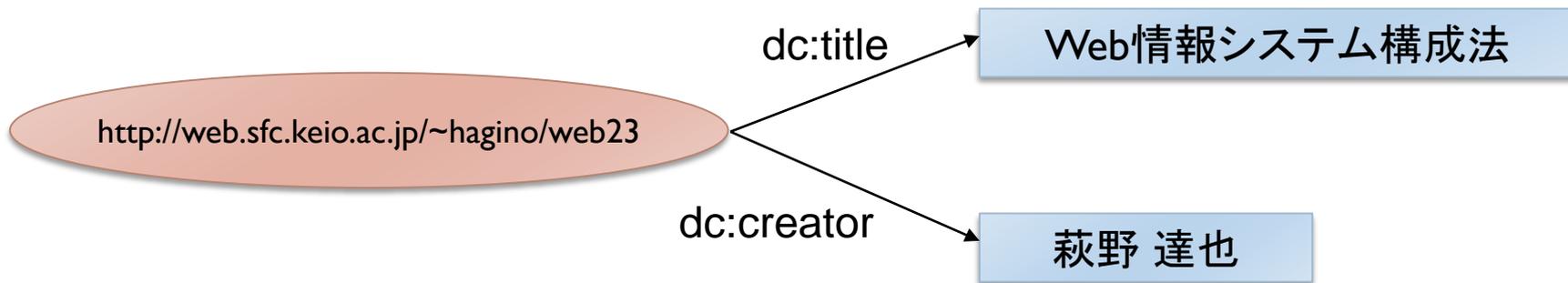
<http://purl.org/dc/terms/title>

Web情報システム構成法

RDFのテキスト形式

- ▶ テキスト表現 (シリアライゼーション)
 - ▶ N-Triples
 - ▶ Notation3 (N3)
 - ▶ RDF/XML
 - ▶ Turtle
 - ▶ RDF/JSON
 - ▶ ...

次のRDFのテキスト表現



N-Triples

- ▶ 単純な形式
- ▶ 3つ組を書くだけ
- ▶ それぞれの3つ組は, 資源, プロパティ, 値
- ▶ 3つ組は「.」で終わる
- ▶ URIは<>で囲み, 値は"\"で囲む

```
<http://web.sfc.keio.ac.jp/~hagino/web23> <http://purl.org/dc/elements/1.1/title> "Web情報システム構成法" .  
<http://web.sfc.keio.ac.jp/~hagino/web23> <http://purl.org/dc/elements/1.1/creator> "萩野 達也" .
```

参照: RDF Test Cases

3. N-Triples

<http://www.w3.org/TR/rdf-testcases/#ntriples>

N3

- ▶ Notation 3 RDF (またはN3)
- ▶ N-Triplesを単純化
- ▶ TurtleはN3の一部

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .  
<http://web.sfc.keio.ac.jp/~hagino/web23> dc:title "Web情報システム構成法" ;  
                                           dc:creator "萩野 達也" .
```

※Turtleでも同じ

参照: Tim Berners-Lee, Primer: Getting into RDF & Semantic Web using N3

<http://www.w3.org/2000/10/swap/Primer.html>

神崎正英, Notation3: RDFの簡易表記から論理表現まで

<http://www.kanzaki.com/docs/sw/n3.html>



RDF/XML

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description
    rdf:about="http://web.sfc.keio.ac.jp/~hagino/web23">
    <dc:title>Web情報システム構成法</dc:title>
    <dc:creator>萩野 達也</dc:creator>
  </rdf:Description>
</rdf:RDF>
```

参照: RDF/XML Syntax Specification (Revised)

<http://www.w3.org/TR/REC-rdf-syntax/>

神崎正英, RDF -- リソース表現のフレームワーク

<http://www.kanzaki.com/docs/sw/rdf-model.html>



RDF/XML (1)

▶ 基本構造

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="URI1">
    <ex:title>
      Web情報システム構成法
    </ex:title>
    <ex:publisher rdf:resource="URI2" />
    <ex:editor>
      <rdf:Description>
        <ex:homePage>
          <rdf:Description rdf:about="URI3">
          </rdf:Description>
        </ex:homePage>
      </rdf:Description>
    </ex:editor>
  </rdf:Description>
</rdf:RDF>
```

名前空間の宣言

主語

述語

目的語

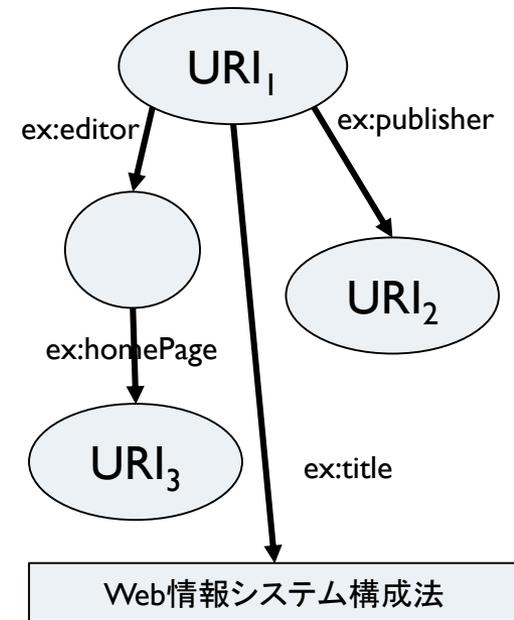
述語

目的語

述語

目的語

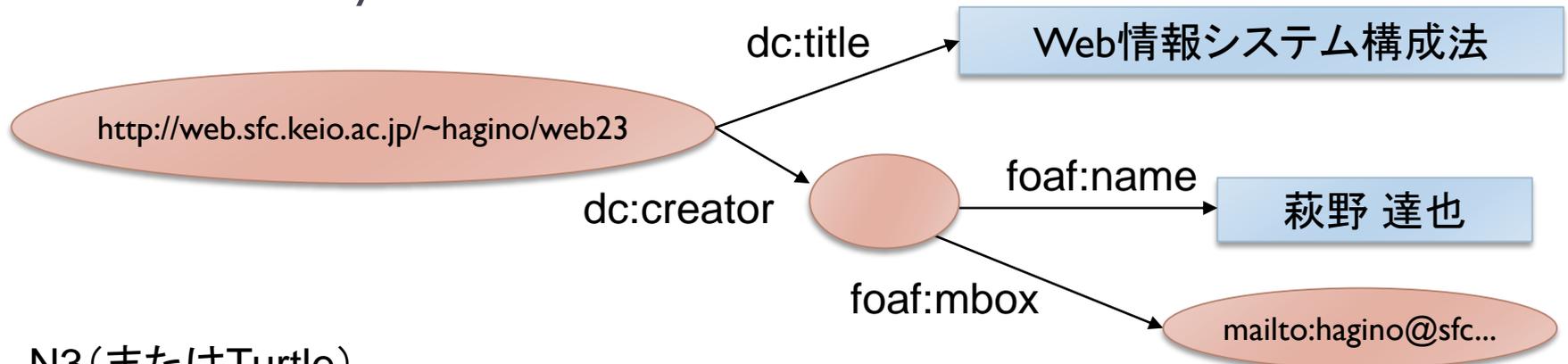
本体



空白ノード

▶ URIの無いノード

- ▶ システムが自動的に一意的なURIを割り当てる
- ▶ 匿名 (anonymous) ノード



N3(またはTurtle)

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<http://web.sfc.keio.ac.jp/~hagino/web23>
  dc:title "Web情報システム構成法" ;
  dc:creator [ foaf:name "萩野 達也" ;
              foaf:mbox <mailto:hagino@sfc.keio.ac.jp> ] .
```

※空白ノードには [述語 目的語; 述語 目的語; ...] を使う.

RDF/XML (2)

▶ 型付きノード

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/"
         xmlns:ex="http://example.org/stuff/1.0/">

  <rdf:Description rdf:about="http://example.org/thing">
    <rdf:type rdf:resource="http://example.org/stuff/1.0/Document"/>
    <dc:title>なにか素晴らしいもの</dc:title>
  </rdf:Description>

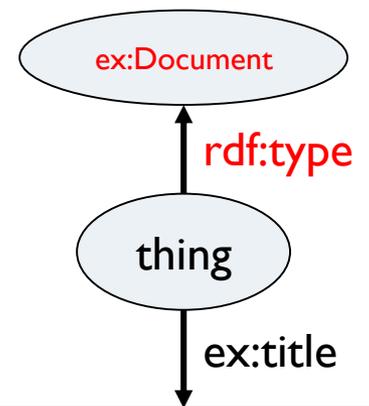
</rdf:RDF>
```

同じ

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/"
         xmlns:ex="http://example.org/stuff/1.0/">

  <ex:Document rdf:about="http://example.org/thing">
    <dc:title>なにか素晴らしいもの</dc:title>
  </ex:Document>

</rdf:RDF>
```



なにか素晴らしいもの

RDF Validation Service

- ▶ 入力
 - ▶ 直接
 - ▶ ファイル
 - ▶ URL
- ▶ 出力
 - ▶ 3つ組
 - ▶ グラフ表現

Validation Results

Your RDF document validated successfully.

Triples of the Data Model

Number	Subject	Predicate	Object
1	http://www.w3.org/	http://purl.org/dc/elements/1.1/title	"World Wide Web Consortium"

The original RDF/XML document

```
1: <?xml version="1.0"?>
2: <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3:   xmlns:dc="http://purl.org/dc/elements/1.1/">
4:   <rdf:Description rdf:about="http://www.w3.org/">
5:     <dc:title>World Wide Web Consortium</dc:title>
6:   </rdf:Description>
7: </rdf:RDF>
8:
```

Graph of the data model



RSS (RDF Site Summary)



- ▶ ニュースの更新情報などを伝える
 - ▶ チャンネル情報を送るために作られた
 - ▶ ブログではトラックバックを送るために利用
- ▶ RSSは有益
 - ▶ Webサイトに行かなくとも更新情報が分かる
 - ▶ 複数のサイトから似たような情報を集めることができる
 - ▶ ニュースを比較できる
- ▶ RSSはRDFの良い例
 - ▶ RDFの簡単なアプリケーションの一つ
 - ▶ 広く用いられている

RSS1.0の構造

▶ 昔のセマンティックWeb委員会のRSSを例に説明

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
<rdf:RDF xmlns="http://purl.org/rss/1.0/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:sy="http://purl.org/rss/modules/syndication/" xml:lang="ja">
```

```
<channel rdf:about="http://www.net.intap.or.jp/INTAP/s-web/">
```

...

```
</channel>
```

channel要素：サマリー記述の対象となるサイトの基本情報を記述

```
<item rdf:about="http://www.net.intap.or.jp/INTAP/s-web/data/conference2003/index.html">
```

...

```
</item>
```

item要素：サイト内の個々の記事や文書に関する情報を記述（複数個記述可能）

```
</rdf:RDF>
```

channel要素の例

```
<channel rdf:about="http://www.net.intap.or.jp/INTAP/s-web/">
```

```
<title>INTAP セマンティックWeb委員会</title>  
<link>http://www.net.intap.or.jp/INTAP/s-web/</link>  
<description>INTAPセマンティックWeb委員会ではセマンティックWeb技術の調査研究ならびに普及活動を行っています。</description>
```

タイトル、対象とするサイトのURL、サイトの内容を記述(必須項目)

```
<dc:language>ja</dc:language>  
<dc:rights>(c)2002 Interoperability Technology Association for Information Processing, Japan</dc:rights>  
<dc:date>2003-06-05T00:00:00+09:00</dc:date>
```

Dublin Coreのボキャブラリを用いて記述された言語や権利情報や日付に関するメタ情報

```
<items>  
  <rdf:Seq>  
    <rdf:li rdf:resource="http://www.net.intap.or.jp/INTAP/s-web/data/conference2003/index.html" />  
    <rdf:li rdf:resource="http://www.net.intap.or.jp/INTAP/s-web/publicdata.html#gijutsu" />  
  </rdf:Seq>  
</items>
```

item要素でサマリーを記述する個々の記事や文書の情報リソースとその順番を示す(必須項目)

```
</channel>
```

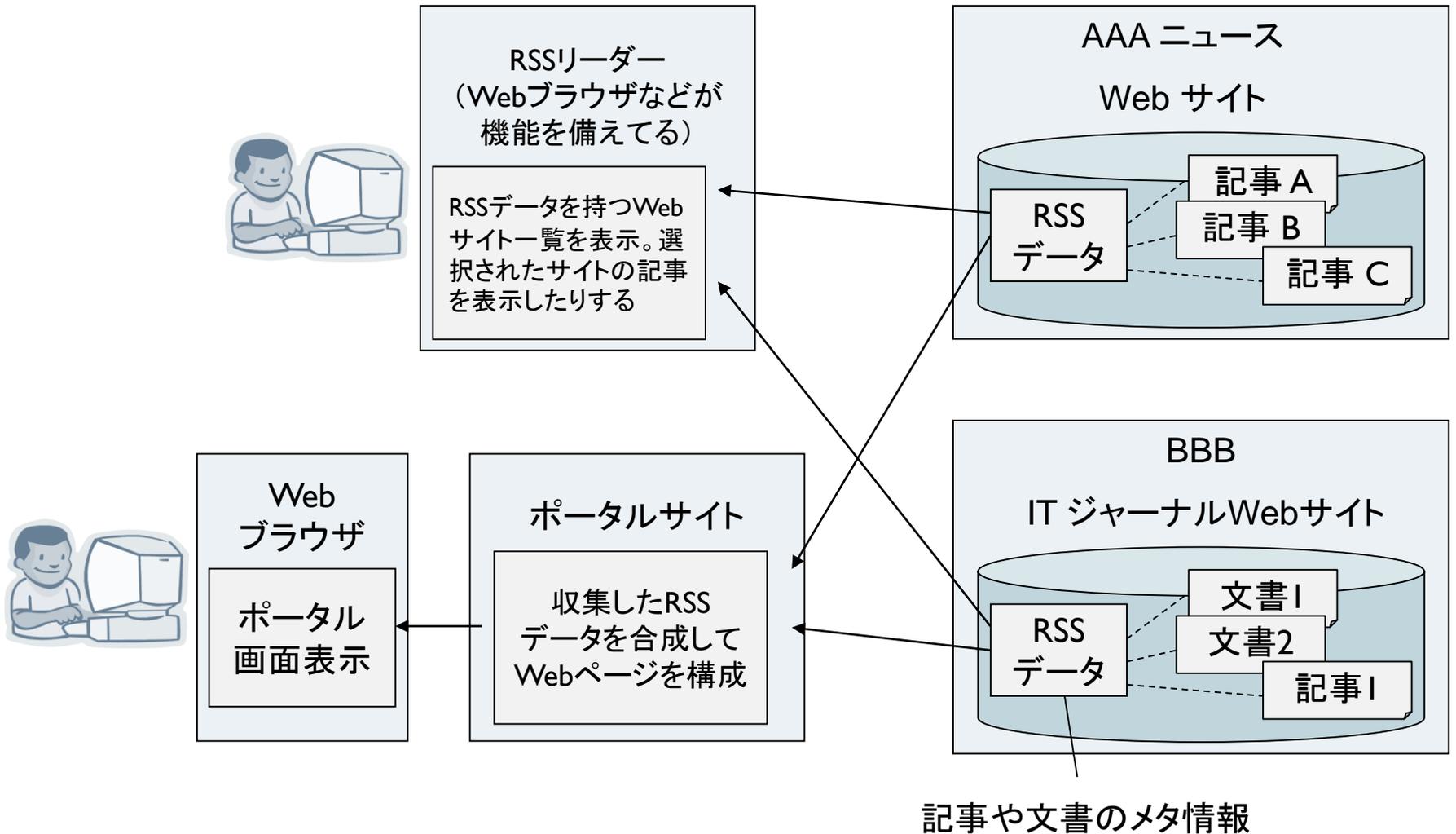
item要素の例

```
<item rdf:about="http://www.net.intap.or.jp/INTAP/s-web/data/conference2003/index.html">
  <title>セマンティックWebコンファレンス2003 開催案内 (11/17開催)</title>
  <link>http://www.net.intap.or.jp/INTAP/s-web/data/conference2003/index.html</link>
  <description>INTAPと慶應 ... セマンティックWebコンファレンス2003を開催致します。</description>
  <dc:subject>セマンティックWebに関するお知らせ</dc:subject>
  <dc:creator>INTAP事務局</dc:creator>
  <dc:date>2003-10-14T00:00:00+09:00</dc:date>
</item>
```

```
<item rdf:about="http://www.net.intap.or.jp/INTAP/s-web/data/14-semanticweb-report.pdf">
  <title>平成14年度セマンティックWeb技術の調査研究報告書</title>
  <link>http://www.net.intap.or.jp/INTAP/s-web/data/14-semanticweb-report.pdf</link>
  <description>本報告書はセマンティックWeb委員会の平成14年度の活動...</description>
  <dc:subject>セマンティックWebに関する文献</dc:subject>
  <dc:creator>INTAPセマンティックWeb委員会</dc:creator>
  <dc:date>2003-05-01T00:00:00+09:00</dc:date>
</item>
```

- ▶ タイトル、対象とする情報リソース(記事や文書)のURLは必須項目
- ▶ それ以外にDublin Coreのボキャブラリを用いてトピック(dc:subject)、作者(dc:creator)、日付(dc:date)に関するメタ情報を記述している。

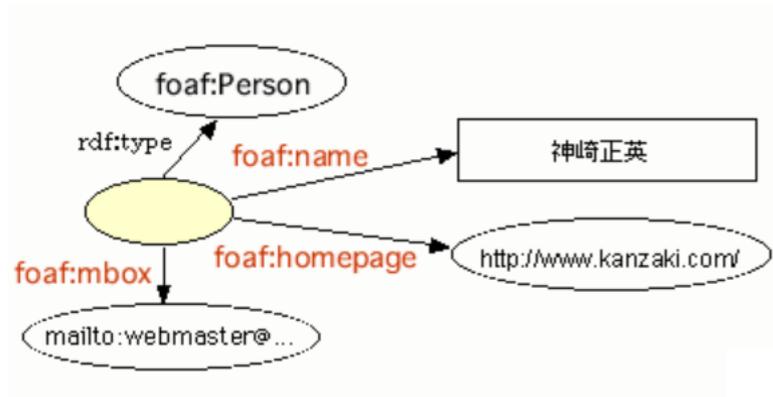
RSSデータのユーザ利用



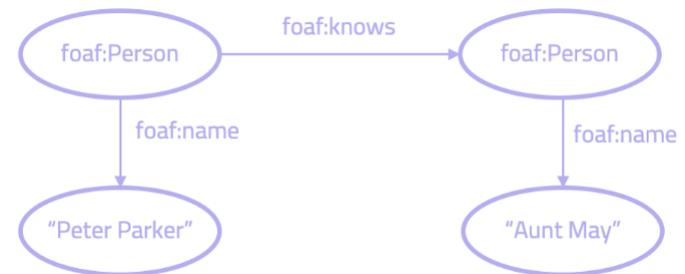
FOAF (Friend of a Friend)



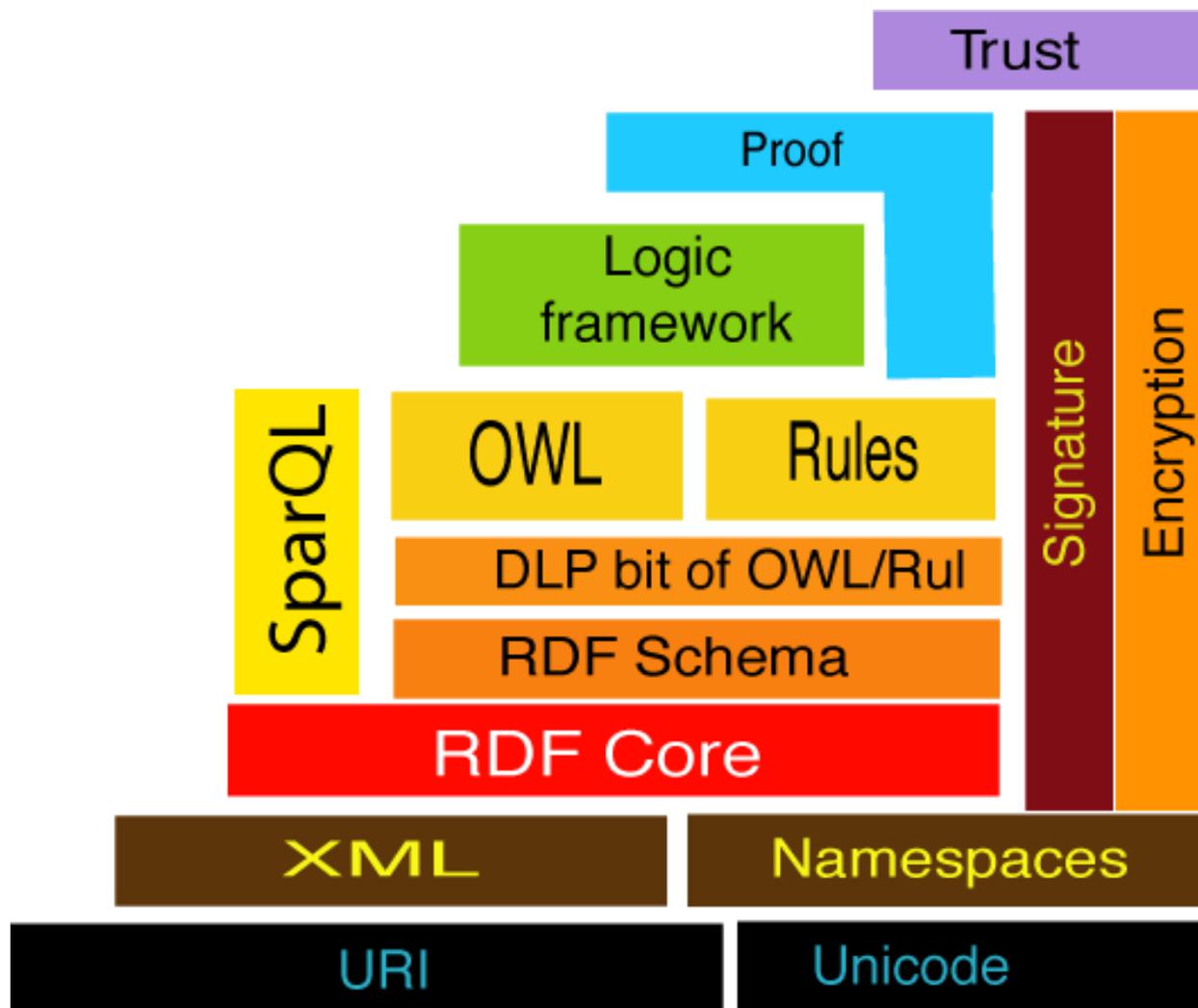
- ▶ 人に関するメタ情報をRDFで記述
 - ▶ 名前, メールアドレス, 住所, 趣味など
 - ▶ <http://www.foaf-project.org/>



- ▶ 友達をつなぐ
 - ▶ foaf:knowsプロパティを使う



セマンティックWebのレイヤー



意味の表現

- ▶ Web資源の意味

- ▶ RDF (Resource Description Framework)

- ▶ 語彙の表現

- ▶ RDF Schema

- ▶ オントロジーを使って語彙の意味を定義する

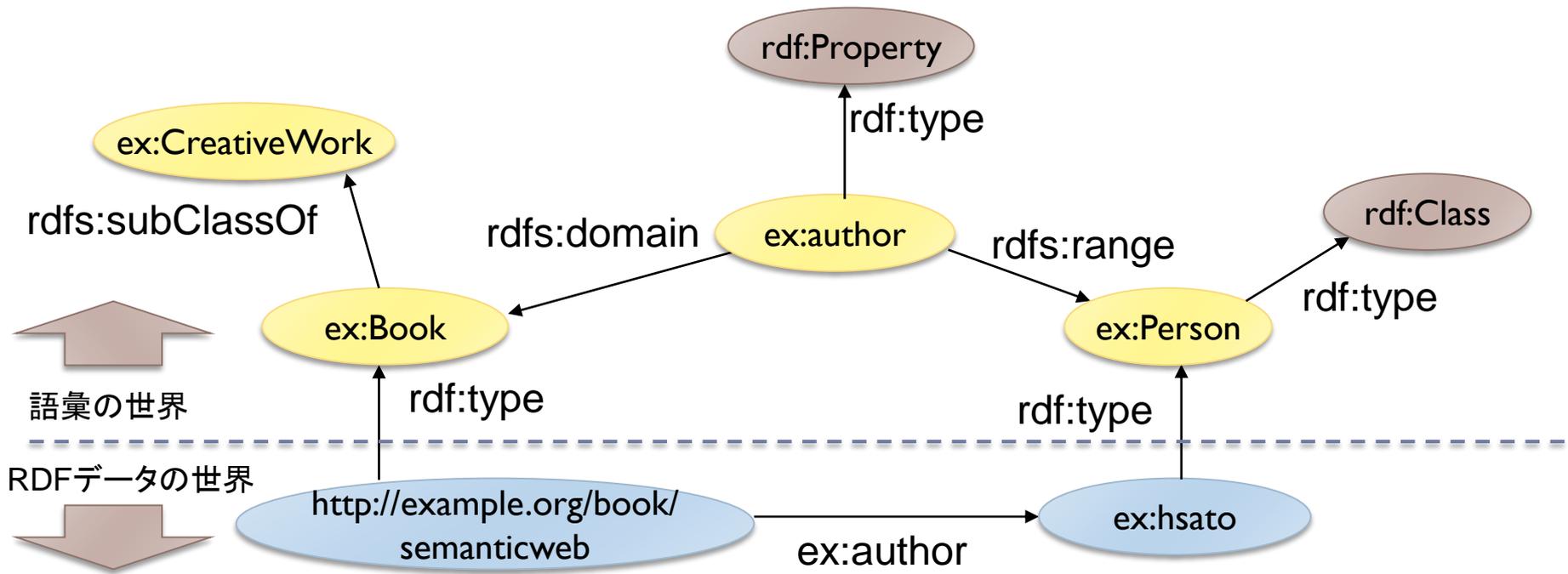
- ▶ OWL (Web Ontology Language)

より高度な
知識の表現



RDF Vocabulary Description Language: (RDF Schema)

- ▶ 資源が所属するクラスを定義
 - ▶ オブジェクト指向プログラミング言語と同じようにクラス階層がある
 - ▶ スーパー(親)クラスとサブ(子)クラス
- ▶ プロパティの語彙を定義
 - ▶ プロパティにも階層がある
 - ▶ プロパティの定義域, 値域を定義



Web Ontology Language (OWL)

- ▶ 概念や資源の間の深い関係を記述する
 - ▶ 同値 : `equivalentClass`
 - ▶ 例 : 「Member」と「Club」は同じインスタンスを持つ
 - ▶ 互いに素 : `disjointWith`
 - ▶ 例 : 「Student Member」であり「Full Member」であるインスタンスは存在しない
 - ▶ 集合 : `unionOf`
 - ▶ 例 : 「Member」は「Student Member」と「Full Member」の共有集合である
 - ▶ その他 : クラスの共通部分, 補集合, プロパティの制約

セマンティックWebもWeb

▶ AAAの原則

- ▶ Anyone can say anything about any topic.
- ▶ 信用できないデータや矛盾するデータが含まれる
- ▶ 複数の名前やIDがあるかもしれない(名前の非同一性)
- ▶ 開世界(閉世界仮説は使えない)

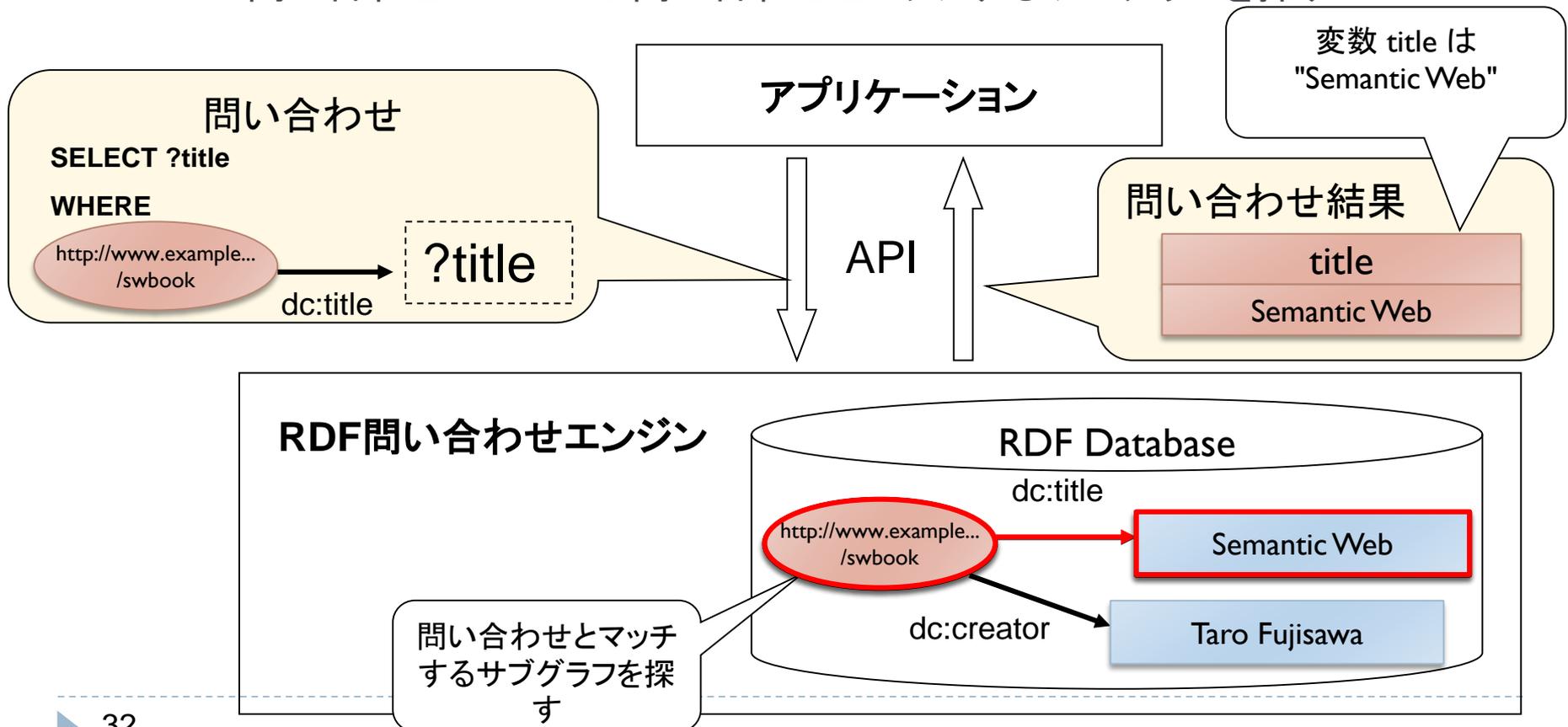
▶ ネットワーク効果

- ▶ 参加者が増えれば価値が高くなる
- ▶ 価値が高くなれば, 参加者も増える

- ▶ データはつながっている
 - ▶ Web文書がハイパーリンクでつながっているように

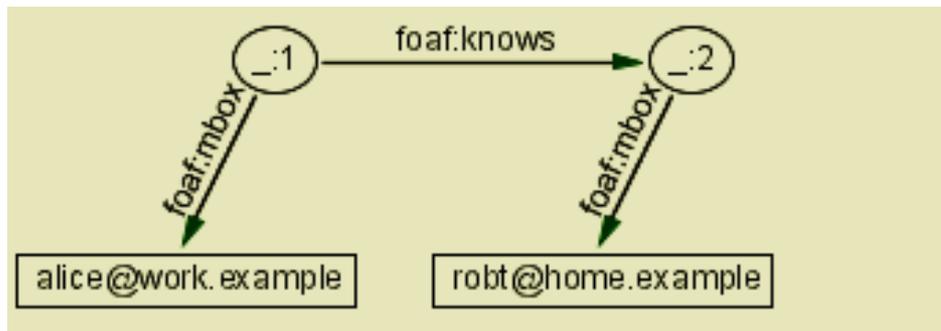
RDF問い合わせ言語

- ▶ RDFデータについて問い合わせる
 - ▶ RDFデータベースはRDFをグラフとして持っている
 - ▶ 問い合わせは変数付きグラフ
 - ▶ RDF問い合わせエンジンは問い合わせとマッチするサブグラフを探す

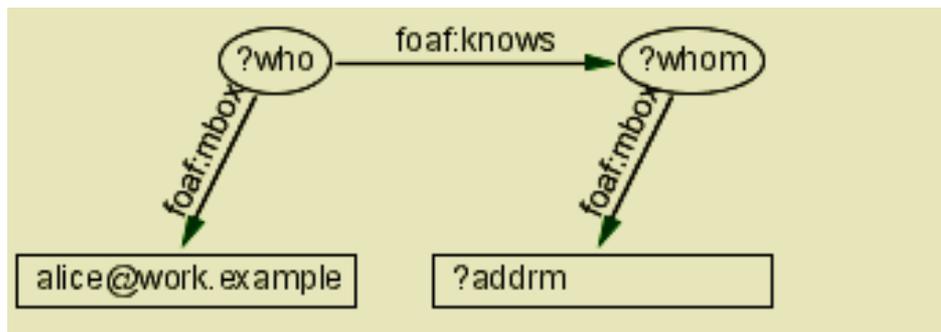


SPARQL

- ▶ SPARQLプロトコルとRDF問い合わせ言語
- ▶ グラフベースの問い合わせ言語(3つ組のパターンマッチ)



グラフ
(RDFデータベース)



グラフパターン
(問い合わせ)

who	whom	addrm
_:1	_:2	"robt@home.example"

問い合わせ結果
(list matched values for variables in the query)

SPARQL問い合わせ (例1)

- ▶ SELECT clause
 - ▶ 欲しい変数を並べる
- ▶ WHERE clause
 - ▶ 変数を含んだグラフ(3つ組)

データ

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
_:a foaf:name "Johnny Lee Outlaw" .
_:a foaf:mbox <mailto:jlow@example.com> .
_:b foaf:name "Peter Goodguy" .
_:b foaf:mbox <mailto:peter@example.org> .
```

問い合わせ

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name, ?mbox
WHERE
{ ?x foaf:name ?name .
  ?x foaf:mbox ?mbox }
```

変数

3つ組

パターン

結果

name	mbox
"Johnny Lee Outlaw"	<mailto:jlow@example.com>
"Peter Goodguy"	<mailto:peter@example.org>

SPARQL問い合わせ (例2: 制約)

データ

```
@prefix dc: <http://purl.org/dc/terms/> .
@prefix : <http://example.org/book/> .
@prefix ns: <http://example.org/ns#> .

:book1 dc:title "SPARQL Tutorial" .
:book1 ns:price 42 .
:book2 dc:title "The Semantic Web" .
:book2 ns:price 23 .
```

問い合わせ

```
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX ns: <http://example.org/ns#>
SELECT ?title ?price
WHERE { ?x ns:price ?price .
  FILTER (?price < 30.5)
  ?x dc:title ?title . }
```

変数

制約 (priceが
50.5より小さい)

結果

title	price
"The Semantic Web"	23

SPARQL問い合わせ（例3：オプションナル）

```
@prefix foaf:      <http://xmlns.com/foaf/0.1/> .
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

```
_:a  rdf:type      foaf:Person .
_:a  foaf:name     "Alice" .
_:a  foaf:mbox     <mailto:alice@work.example> .

_:b  rdf:type      foaf:Person .
_:b  foaf:name     "Bob" .
```

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE { ?x foaf:name ?name .
       OPTIONAL { ?x foaf:mbox ?mbox }
}
```

オプションナルパターン
(満たす必要はない)

name	mbox
"Alice"	<mailto:alice@example.com>
"Bob"	

DBpedia SPARQLエンドポイント

- ▶ DBpedia

- ▶ Wikipediaのinfoboxを抽出
- ▶ <https://www.dbpedia.org/>



- ▶ DBpediaデータのSPARQLエンドポイント

- ▶ <https://dbpedia.org/sparql>

- ▶ 日本語DBpediaのエンドポイントも存在

- ▶ <https://ja.dbpedia.org/sparql>



Linked Data と Linked Open Data (LOD)

▶ オープンデータ

- ▶ 誰でもが利用して良いようにデータを公開する
 - ▶ 商用利用も許す
- ▶ 再配布も可能
- ▶ 政府のデータ, 製品情報, など

▶ Linked Data

- ▶ 事物(資源)の識別子としてのURIを使う
- ▶ データ記述の標準フォーマットであるRDFを使う
- ▶ 共通のURIを使うことでデータ同士をリンクして活用する

▶ Linked Open Data

- ▶ オープンなライセンスの下で提供するLinked Data



オープンデータの5つ星

▶ <http://5stardata.info/>

★ オープンライセンスでWebでデータ(形式はなんでもよい)を公開する

★★ 構造化データ(スキャナの画像ではなくExcelなどを使う)を公開する

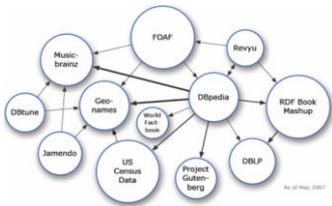
★★★ 公開された形式でデータを公開する(ExcelではなくCSVにする)

★★★★ URIを使って表現し, 他のデータから参照可能にする

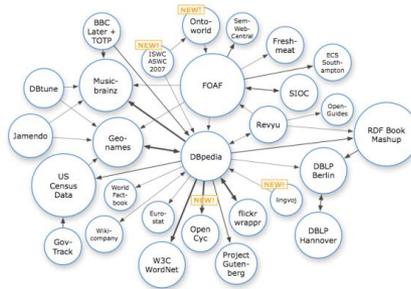
★★★★★ 他のデータにリンクする



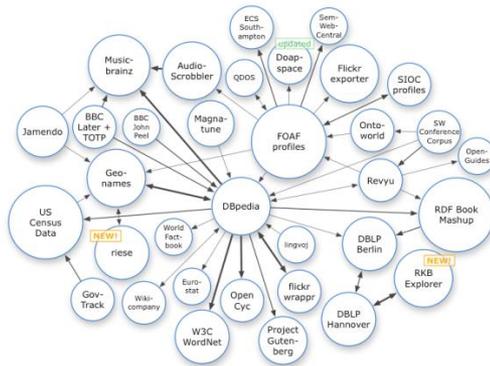
Linked Open Data



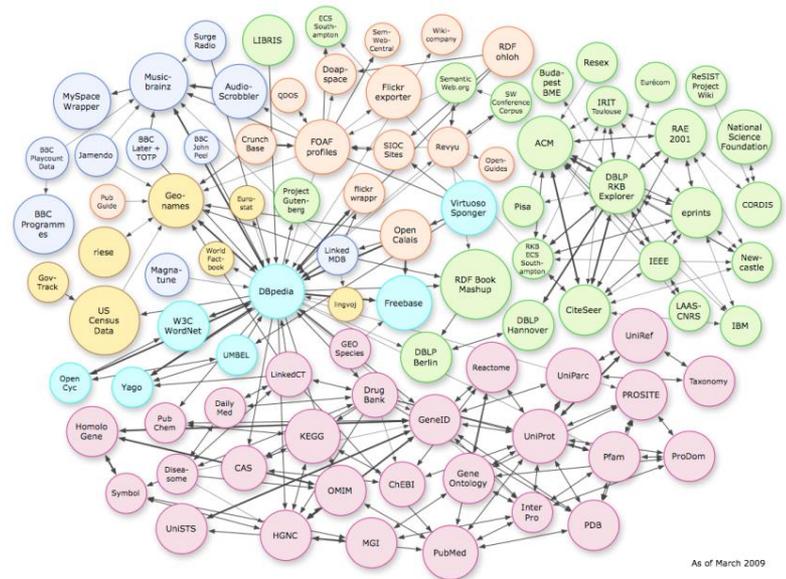
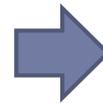
2007/5/1



2007/11/7



2008/3/31

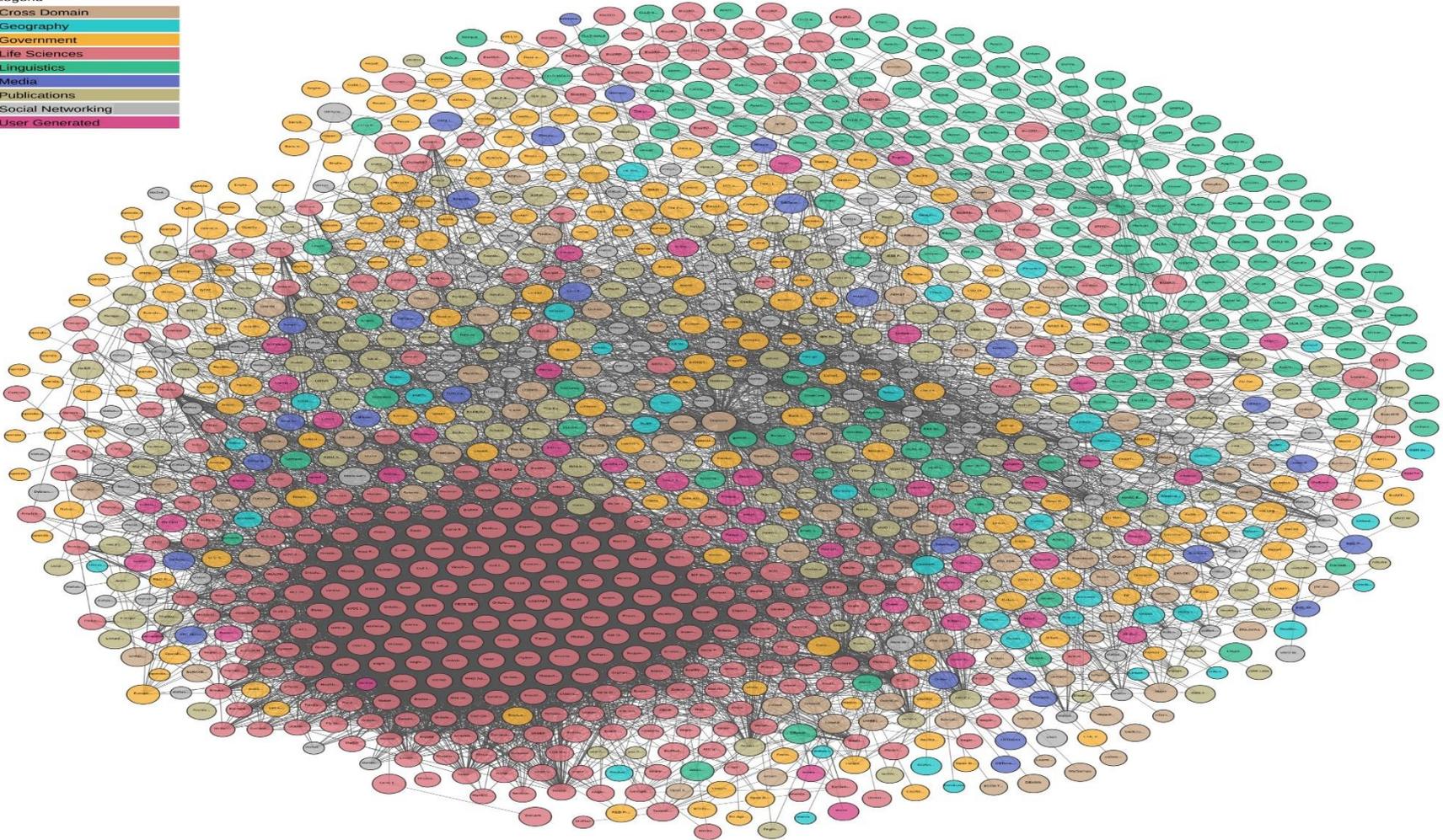


2009/3/27

As of March 2009



Linked Open Data (LOD)

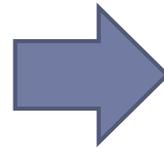


The Linked Open Data Cloud from lod-cloud.net



データをつなぐことの効果

2008年に弁護士がオハイオ州のゼーンズビルの水道の整備地図を作成



白人の居住状況とマッシュアップ



黒人多い居住地に差別的に水道が整備されていないことが分かり、市が10.9億円の賠償を命じられる

<https://www.dispatch.com/story/news/2008/07/11/racism-ruled-jury-finds/23536573007/>

まとめ

- ▶ セマンティックWeb
 - ▶ 計算機可読メタデータ
- ▶ RDF
 - ▶ 3つ組: 資源, プロパティ, 値
 - ▶ 資源とプロパティにURIを使う
- ▶ SPARQL
 - ▶ RDFデータの問い合わせ言語
- ▶ Linked Open Data