

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

# Web情報システム構成法

## No.12 RDF入門

萩野 達也 (hagino@sfc.keio.ac.jp)

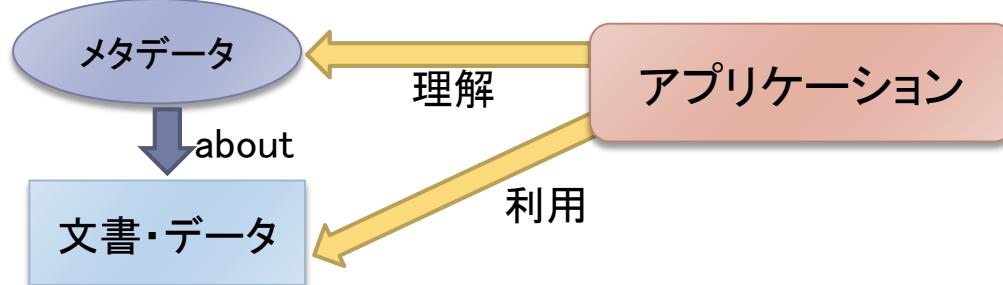
# 文書 vs データ

---

- ▶ 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



- ▶ W3CセマンティックWebアクティビティ → データアクティビティ
  - ▶ <http://www.w3.org/2001/sw/> → <https://www.w3.org/2013/data/>
  - ▶ 「セマンティックWebはアプリケーションや組織などの境界を越えてデータを共有し再利用することのできる共通のフレームワークを提供する」

# 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

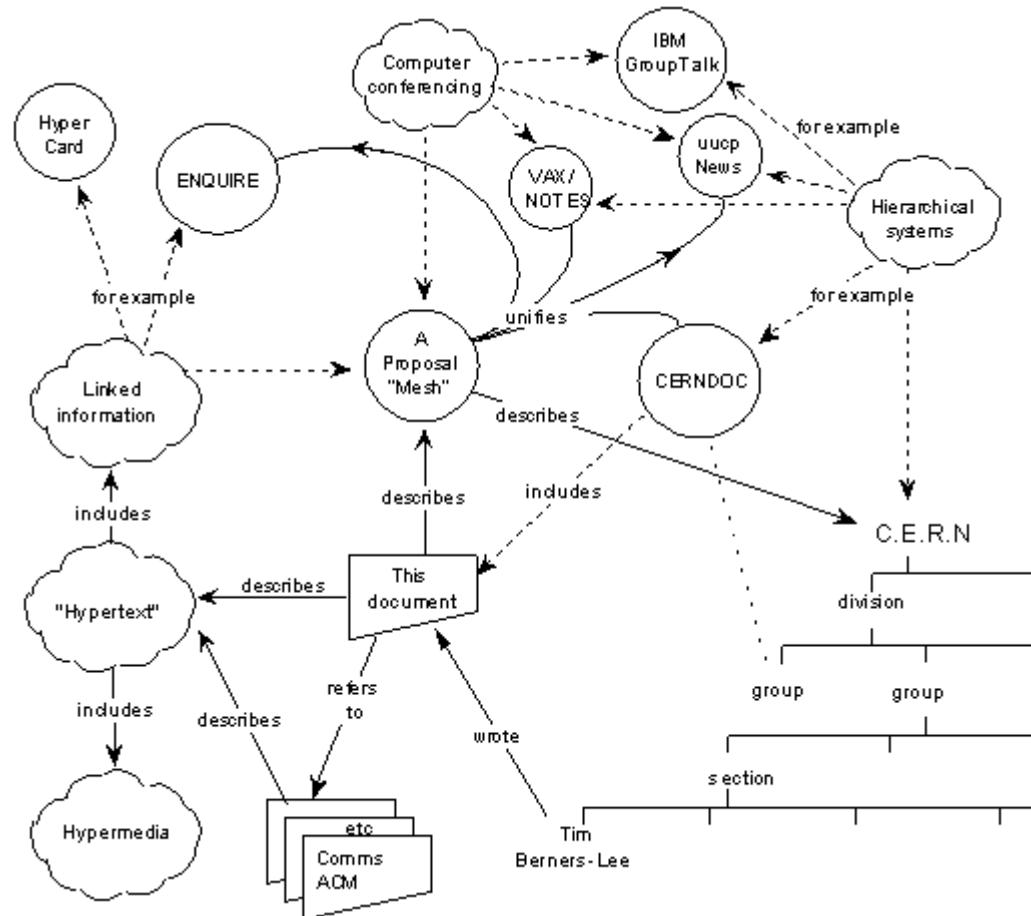


## ▶ セマンティックWeb

- ▶ 1998年

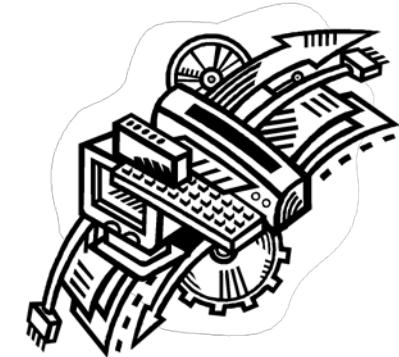


## ▶ データWeb



# 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(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ぬ



| 計営録 東海道本線 横浜・東京方面 (上り) |    |    |    |    |    |    |    |    |    |    |    | ▶ 休曜・休日の時刻表を表示 |
|------------------------|----|----|----|----|----|----|----|----|----|----|----|----------------|
| 平日                     |    |    |    |    |    |    |    |    |    |    |    |                |
| 5                      | 00 | 22 | 50 |    |    |    |    |    |    |    |    |                |
| 6                      | 01 | 休  | 休  | 休  | 休  | 休  | 休  | 休  | 休  | 休  | 休  | 休              |
| 7                      | 01 | 04 | 07 | 10 | 13 | 16 | 19 | 22 | 25 | 29 | 31 | 休              |
| 8                      | 01 | 04 | 09 | 12 | 15 | 19 | 22 | 27 | 30 | 35 | 43 | 51             |
| 9                      | 01 | 10 | 23 | 33 | 45 | 50 | 55 |    |    |    |    |                |
| 10                     | 05 | 15 | 26 | 38 | 49 | 59 |    |    |    |    |    |                |
| 11                     | 14 | 32 | 50 | 53 |    |    |    |    |    |    |    |                |
| 12                     | 02 | 14 | 35 | 44 | 50 |    |    |    |    |    |    |                |
| 13                     | 03 | 15 | 26 | 32 | 39 | 50 | 56 |    |    |    |    |                |
| 14                     | 07 | 15 | 25 | 38 | 50 | 54 |    |    |    |    |    |                |
| 15                     | 04 | 17 | 32 | 49 | 50 | 58 |    |    |    |    |    |                |
| 16                     | 02 | 16 | 28 | 39 | 42 | 50 | 55 |    |    |    |    |                |
| 17                     | 07 | 14 | 19 | 24 | 34 | 46 | 51 | 54 |    |    |    |                |
| 18                     | 05 | 11 | 22 | 26 | 39 | 45 | 50 | 57 |    |    |    |                |
| 19                     | 08 | 20 | 23 | 32 | 42 | 48 | 54 | 58 |    |    |    |                |
| 20                     | 09 | 21 | 25 | 37 | 42 | 50 |    |    |    |    |    |                |
| 21                     | 09 | 14 | 21 | 25 | 39 | 51 |    |    |    |    |    |                |
| 22                     | 04 | 19 | 25 | 33 | 39 | 45 | 50 |    |    |    |    |                |
| 23                     | 12 | 27 | 39 |    |    |    |    |    |    |    |    |                |



# RDF

- ▶ Resource Description Framework
  - ▶ データを三つ組みとして表す
- ▶ 三つ組み(Triple)
  - ▶ 資源 - プロパティ - 値
  - ▶ 主語 - 述語 - 目的語
  - ▶ 有向グラフ



この資源(この授業のWebページ)のタイトルは「Web情報システム構成法」である。

# 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

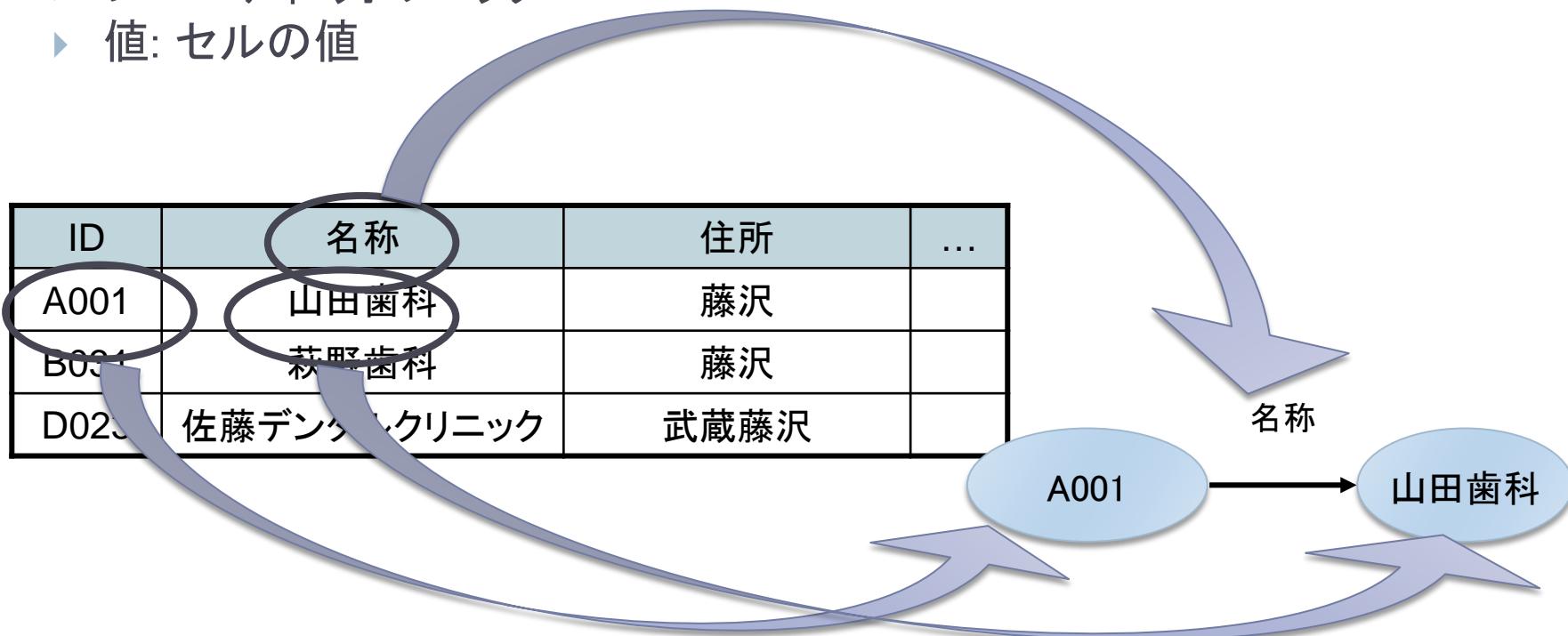
| 住所   | 県   |
|------|-----|
| 藤沢   | 神奈川 |
| 武蔵藤沢 | 埼玉  |
| ...  | ... |



| ID   | 休診日    |
|------|--------|
| B031 | 木曜, 日曜 |
| D023 | 土曜, 日曜 |

# 表をRDFに変換する

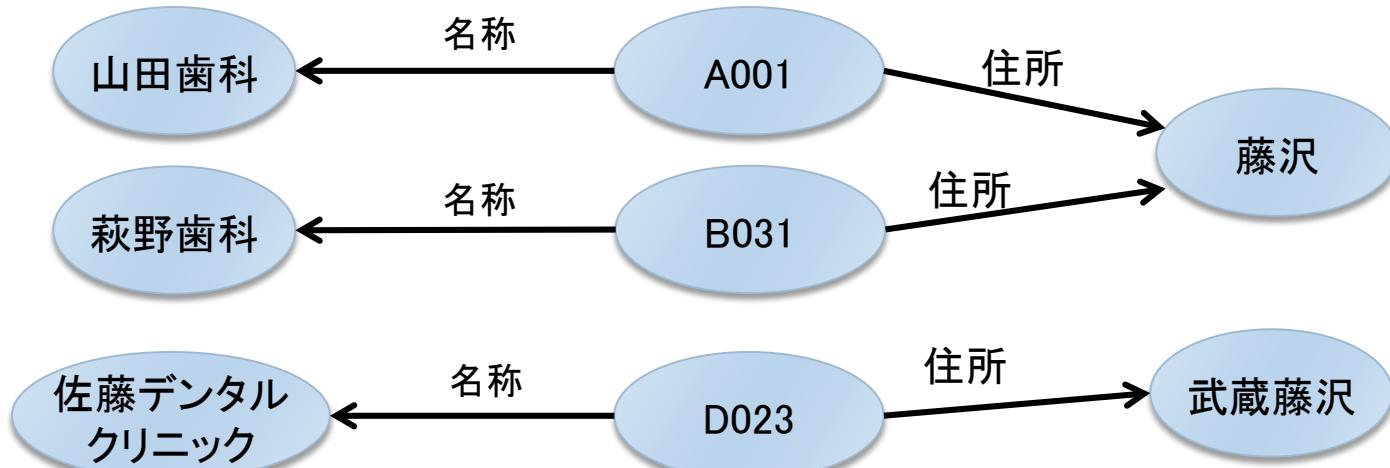
- ▶ 表の各行を三つ組みで表す
  - ▶ (資源, プロパティ, 値)
  - ▶ 資源: 主キー
  - ▶ プロパティ: 列のヘッダ
  - ▶ 値: セルの値



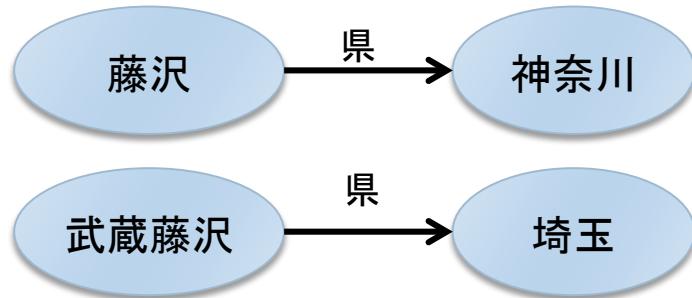
# 表をRDFグラフに変換する

- ▶ すべての資源およびプロパティをRDFに変換する
- ▶ 同じ資源は共有される

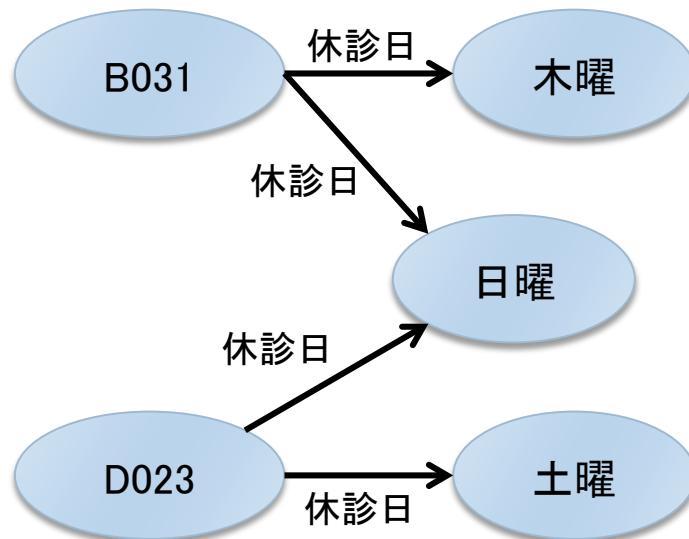
| ID   | 名称          | 住所   | ... |
|------|-------------|------|-----|
| A001 | 山田歯科        | 藤沢   | ... |
| B031 | 萩野歯科        | 藤沢   | ... |
| D023 | 佐藤デンタルクリニック | 武藏藤沢 | ... |



# 表をRDFグラフに変換する



| 住所   | 県   |
|------|-----|
| 藤沢   | 神奈川 |
| 武蔵藤沢 | 埼玉  |
| ...  | ... |

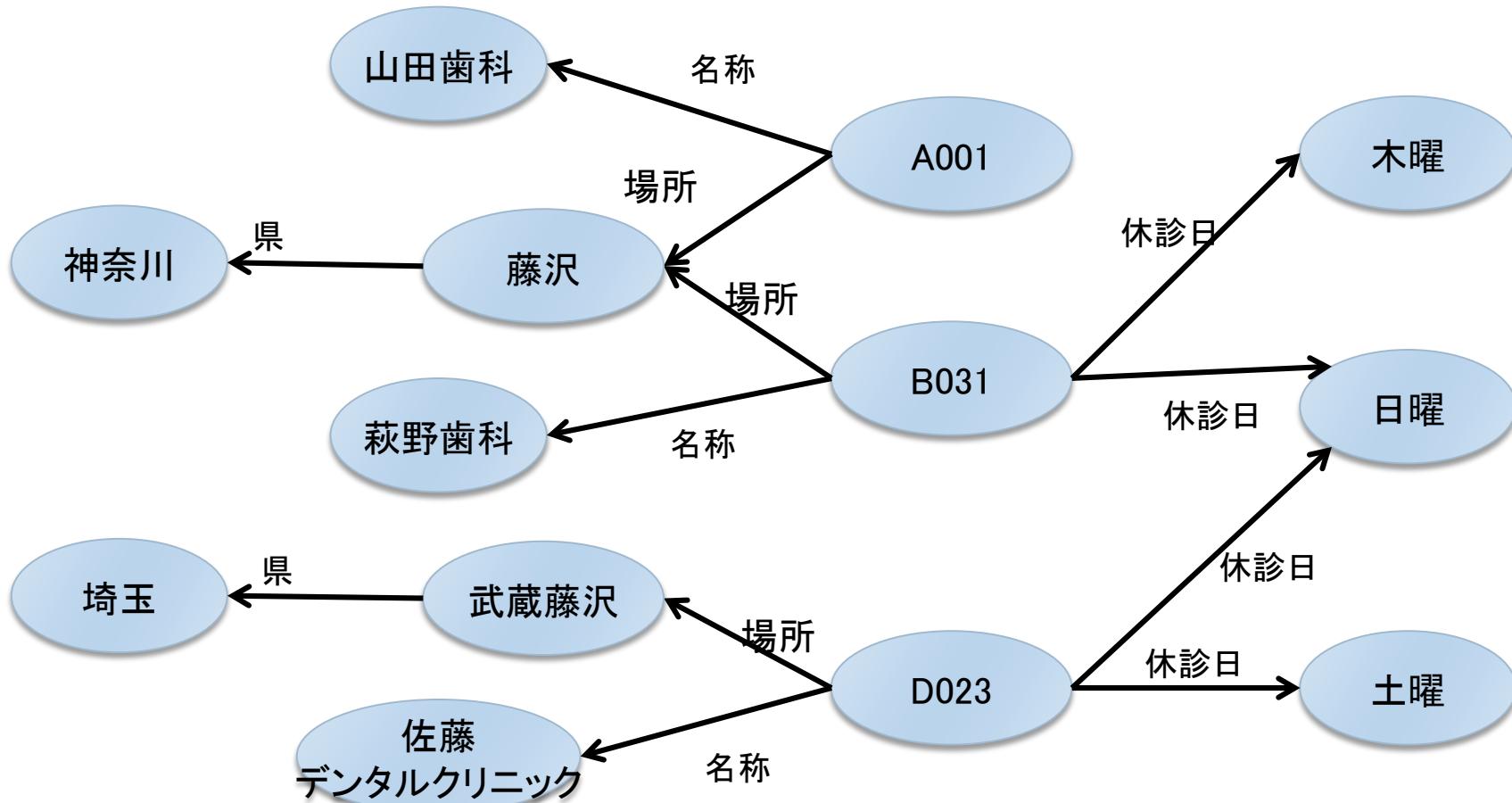


| ID   | 休診日    |
|------|--------|
| B031 | 木曜, 日曜 |
| D023 | 土曜, 日曜 |



# 複数のデータ資源を統合する

- 同じノード(資源)を共有すればよい



# 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>



# 他の表現方法

---

- ▶ グラフ表現
  - ▶ 人にはわかりやすい
  - ▶ ファイルとして保存したり、ネットワークで転送するのが難しい
  - ▶ RDFをどのように公開するのか？
  
- ▶ テキスト表現
  - ▶ シリアライゼーション
    - ▶ N-Triples
    - ▶ Notation 3 RDF (N3)
    - ▶ RDF/XML

# 意味の表現

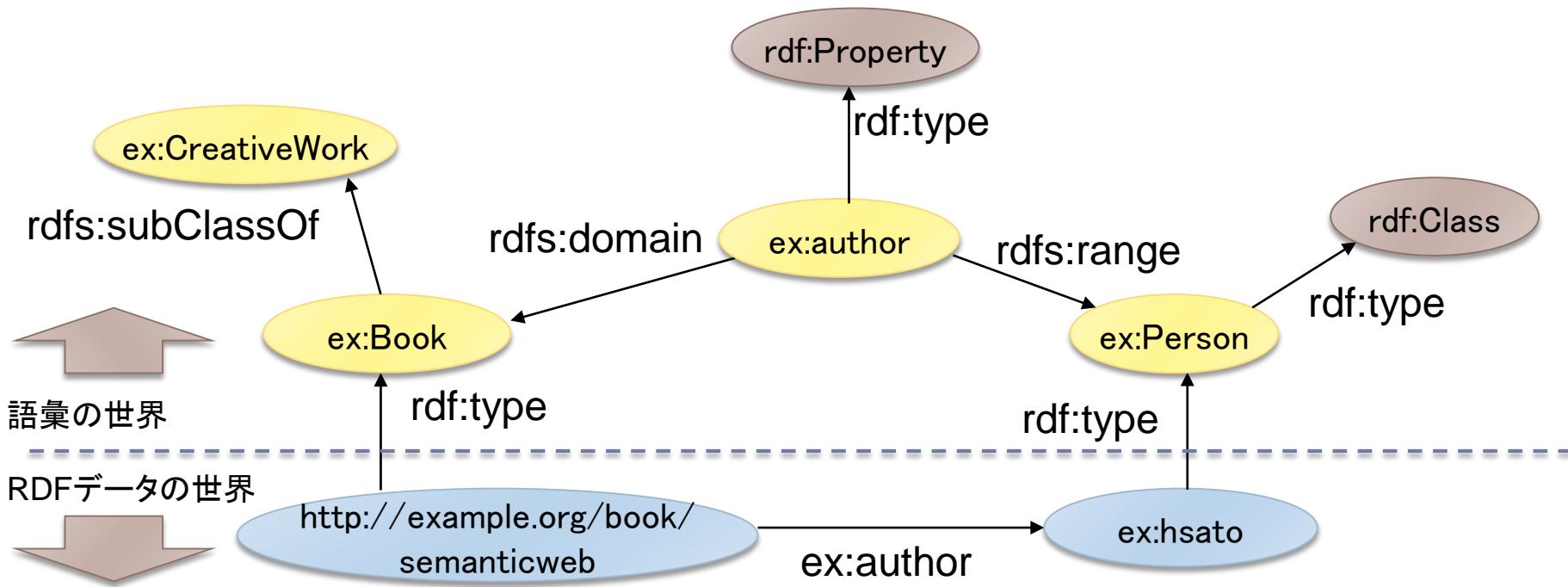
- ▶ Web資源の意味
  - ▶ RDF (Resource Description Framework)
- ▶ 語彙の表現
  - ▶ RDF Schema
- ▶ オントロジーを使って語彙の意味を定義する
  - ▶ OWL (Web Ontology Language)

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

# 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" />
    </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>
```

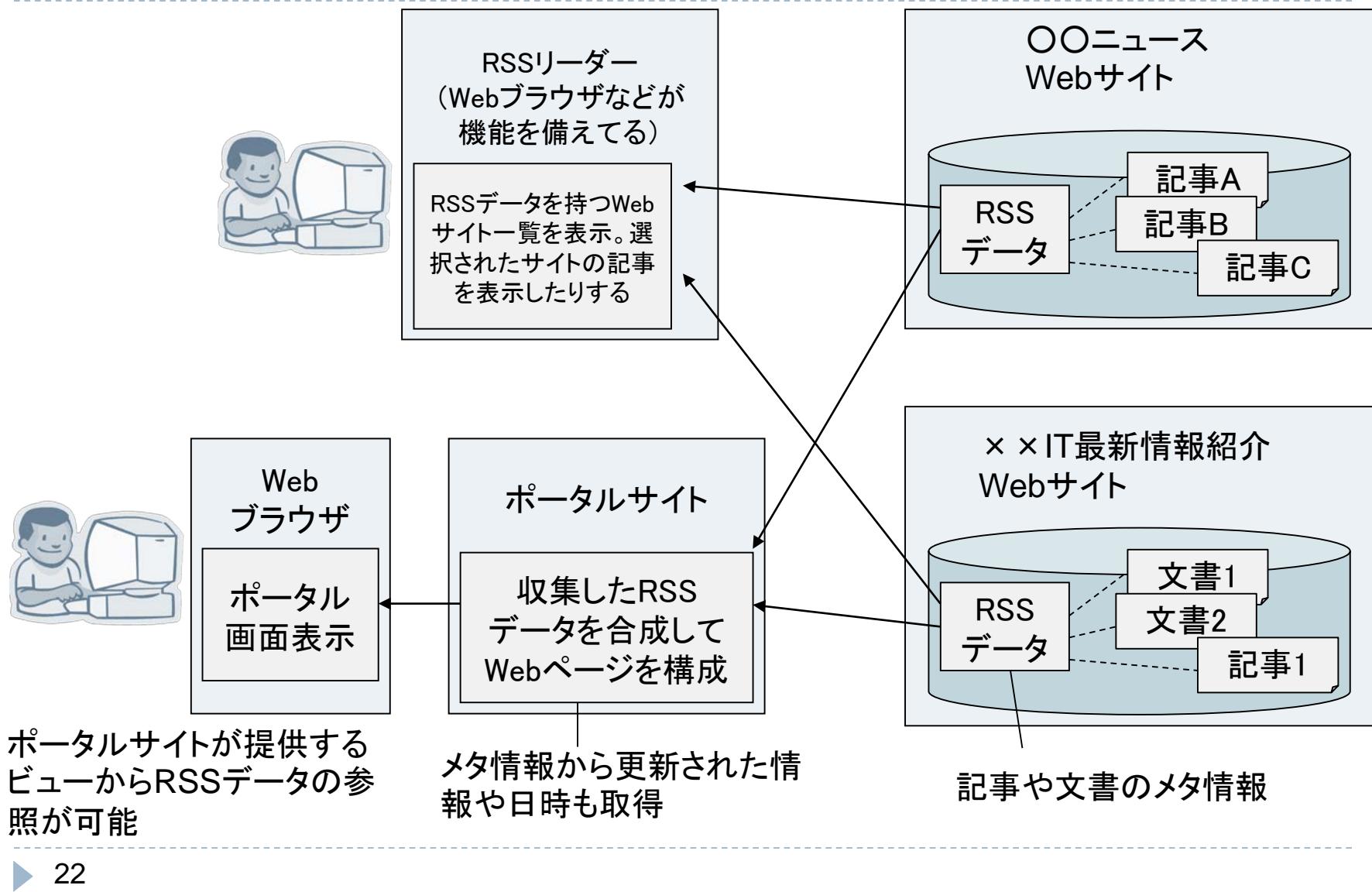
# item要素の例

```
<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



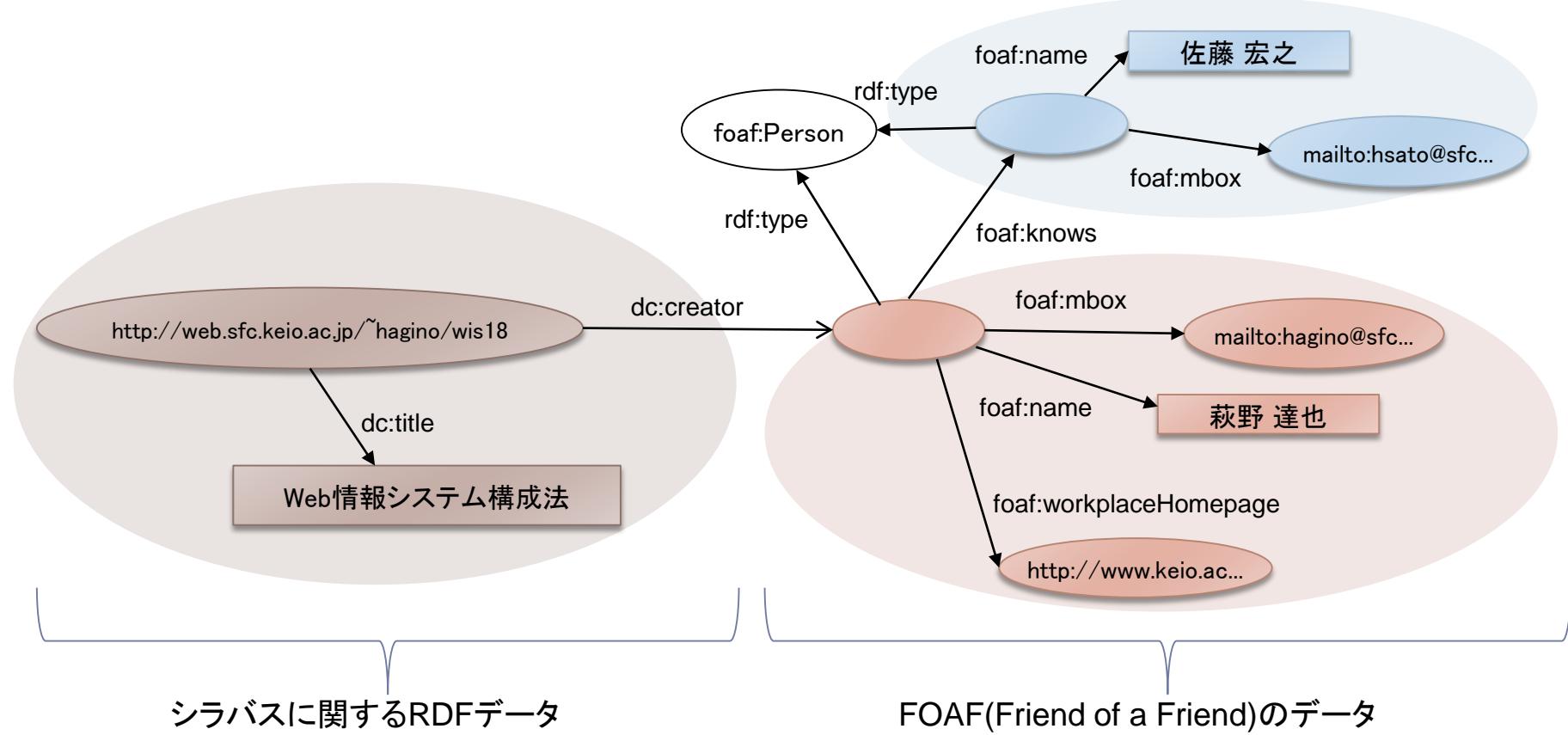
# FOAF (Friend of a Friend)



- ▶ セマンティックWebの初期からRSSの次にRDFが活用された事例
  - ▶ 人に関するメタ情報(名前、興味、どこの近くに住んでいるなど)をRDFで記述して公開するプロジェクト
    - ▶ <http://www.foaf-project.org/>
  - ▶ FOAFのメタ情報は、RSSと同様に個々のWebサイトなどどこに置いても良い
  - ▶ 知っている人を記述できるので、個々のRDFから知人のRDF記述を連鎖的に辿れるようになる
  - ▶ ソーシャルネットワークの観点から注目を集めた
    - ▶ 自分と同じ興味を持つユーザの発見
    - ▶ 人の結びつきから信頼性を測る
- ▶ 23

# 分散した情報の統合

- ▶ RDFのグラフ構造のデータモデルにより異なるデータを統合して扱うことができる

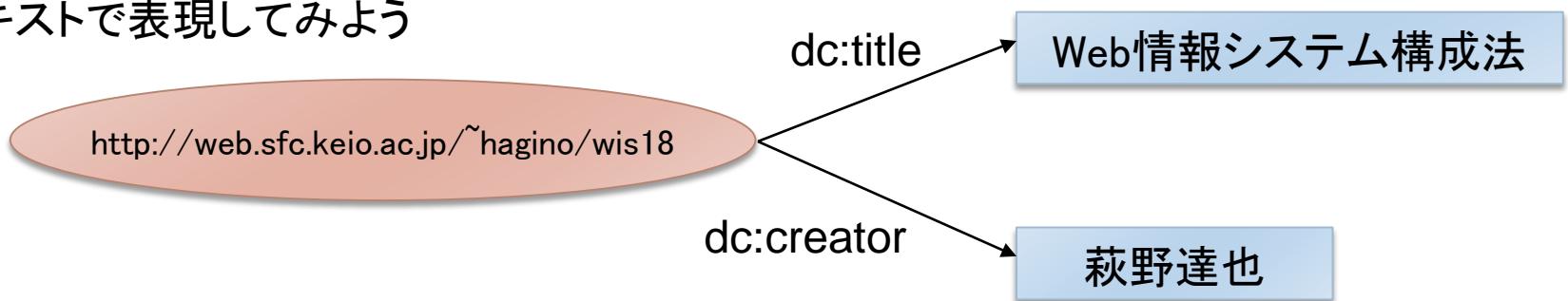


# RDFのテキスト形式は複数ある

## ▶ テキスト形式による表現(シリアル化)

- ▶ N-Triples
- ▶ Notation3 (N3)
- ▶ RDF/XML
- ▶ Turtle
- ▶ RDF/JSON
- ▶ ...

以下のグラフ構造を持つデータをテキストで表現してみよう



# N-Triplesによる表現

- ▶ 最も単純な形式
- ▶ 三つ組を直接書く
- ▶ 三つ組ごとに:リソース, プロパティ, 値の順に並べて, 「.」(ピリオド)で終わる
- ▶ URIは<>で囲み, リテラルは" "で囲む

```
<http://web.sfc.keio.ac.jp/~hagino/wis18> <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/wis18> <http://purl.org/dc/elements/1.1/creator> "萩野達也" .
```

参考: RDF Test Cases

## 3. N-Triples

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

# N3による表現

- ▶ Notation 3 RDF (略してN3)
- ▶ N-Triplesの明快さを維持しつつ簡略化できる
- ▶ TurtleはN3のサブセット

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .  
  
<http://web.sfc.keio.ac.jp/~hagino/wis18> dc:title "Web情報システム構成法" ;  
dc:creator "萩野達也" .
```

※Turtleでも同様

See: 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による表現

```
<?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/wis18">
      <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>
```

参考: RDF/XML Syntax Specification (Revised)

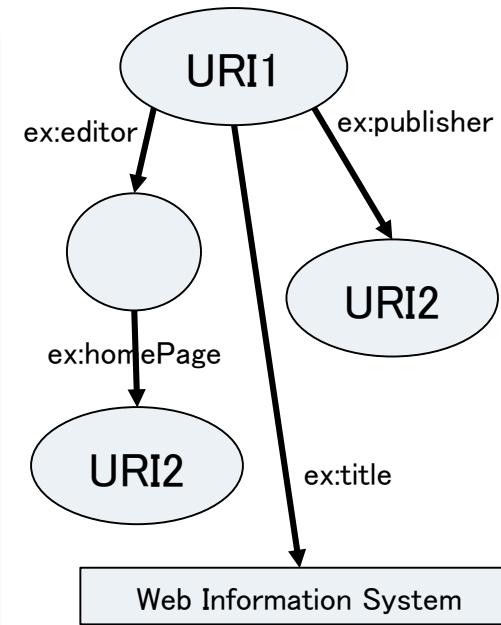
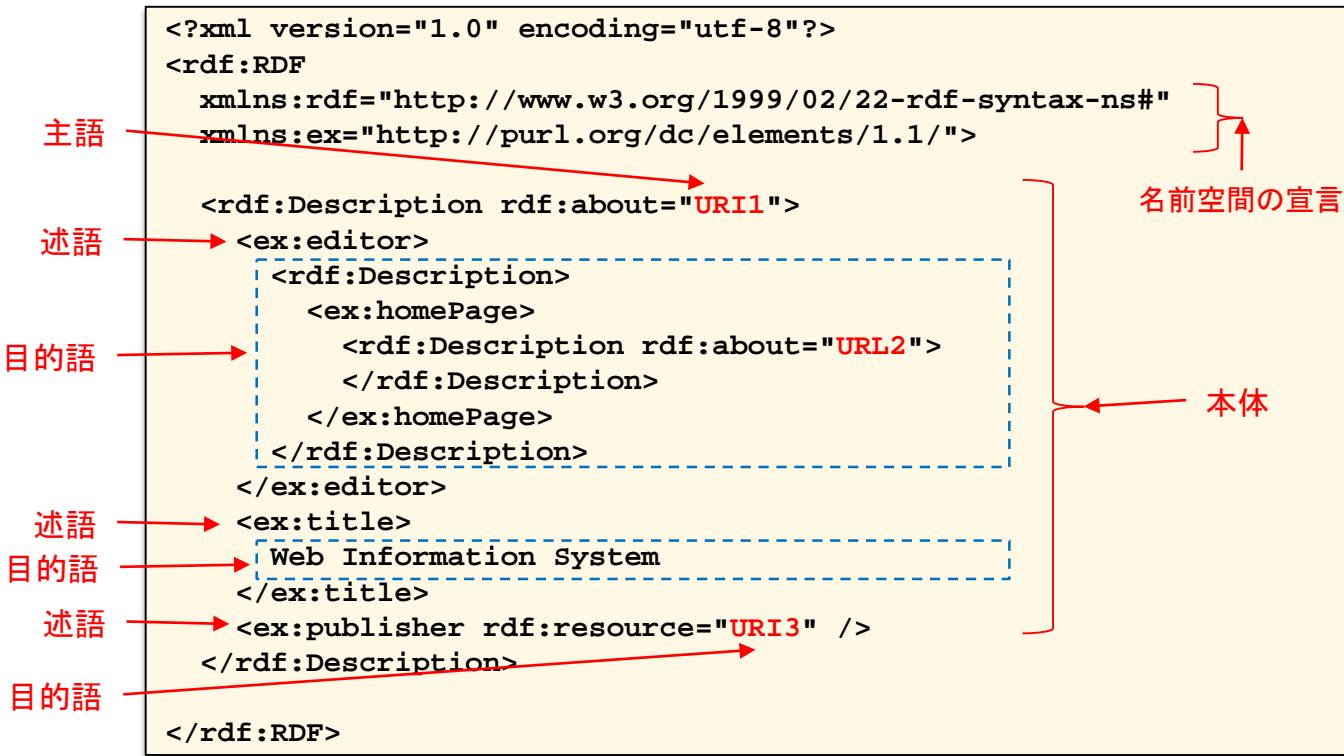
<http://www.w3.org/TR/REC-rdf-syntax/>

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>

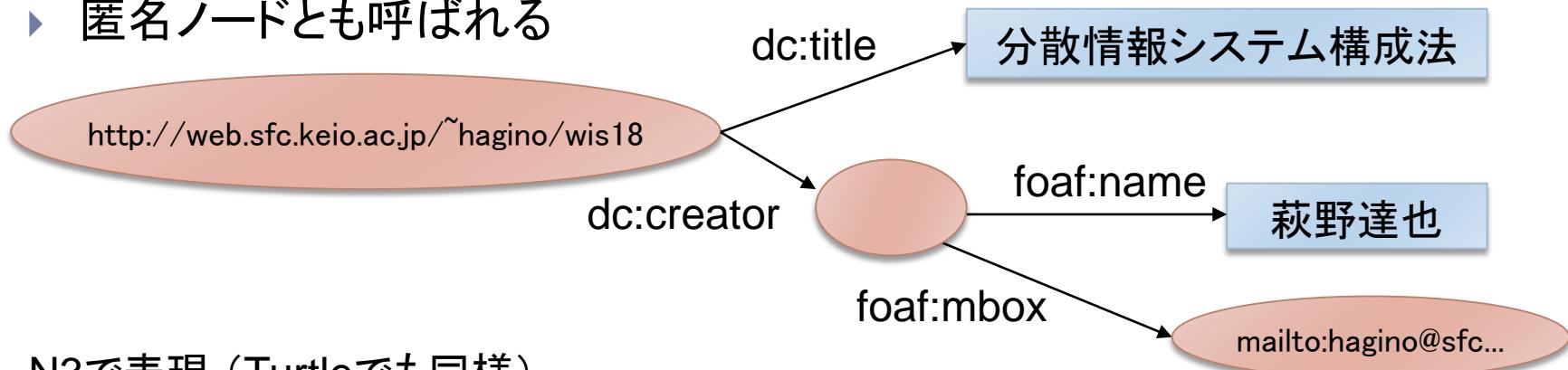
# RDF/XML形式 (1)

## ▶ 基本的構造



# ブランクノード

- ▶ グラフ表現においてURIが与えられないノード
  - ▶ 与えない方が表現の都合が良いことがある
  - ▶ システムが自動的に一意的なURIを割り当てる
- ▶ 匿名ノードとも呼ばれる



N3で表現 (Turtleでも同様)

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .  
@prefix foaf: <http://xmlns.com/foaf/0.1/> .  
  
<http://web.sfc.keio.ac.jp/~hagino/wis18>  
  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> ] .
```

※ブランクノードを主語とするトリプルは、主語を略して[述語 目的語]と表記することができる

# RDF/XML形式 (2)

## ▶ 型付ノード

```
<?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>A marvelous thing</dc:title>
  </rdf:Description>

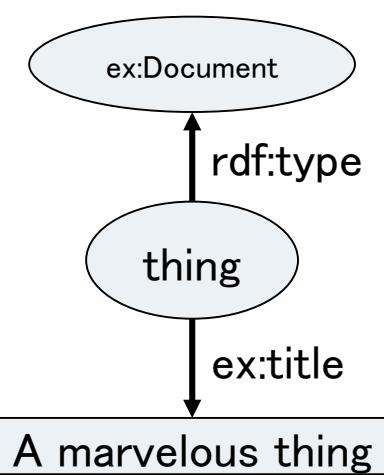
</rdf:RDF>
```

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>A marvelous thing</dc:title>
  </ex:Document>

</rdf:RDF>
```



# RDF Validation Service

- ▶ Webブラウザから利用可能

## ▶ 入力

- ▶ 直接
- ▶ ファイル
- ▶ URL

## ▶ 出力

- ▶ 解析結果
- ▶ 三つ組
- ▶ グラフ

### Validation Results

Your RDF document validated successfully.

### Triples of the Data Model

| Number | Subject   | Predicate   | Object                      |
|--------|---|---|-----------------------------|
| 1      | <a href="http://www.w3.org/">http://www.w3.org/</a> | <a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a> | "World Wide Web Consortium" |

### The original RDF/XML document

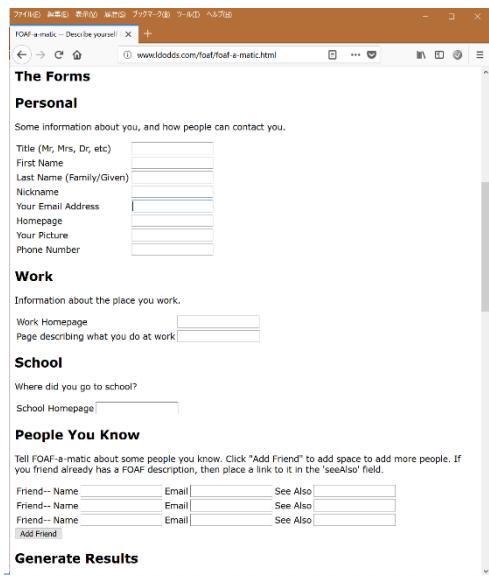
```
1: <?xml version="1.0"?>
2: <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3:   xmlns:dc="http://purl.org/dc/elements/1.1/">
4:   <rdf:Description rdf:about="http://www.w3.org/">
5:     <dc:title>World Wide Web Consortium</dc:title>
6:   </rdf:Description>
7: </rdf:RDF>
8:
```

### Graph of the data model



# 課題:FOAFを作ってみよう

- ▶ FOAF-o-Maticを使って自分のFOAFを作ってみなさい.
  - ▶ <http://www.ldodds.com/foaf/foaf-a-matic.html>
- ▶ 提出
  - ▶ <https://vu5.sfc.keio.ac.jp/kadai/>
  - ▶ 出来上がったRDF/XMLをテキストとして提出しなさい.
  - ▶ 締め切り: 7月14日正午



The screenshot shows the FOAF-o-Matic web interface. It has sections for Personal information (Title, First Name, Last Name, Email, Homepage, Picture, Phone Number), Work information (Work Homepage, Job Description), School information (School Homepage), and People You Know (Friend list). A large blue arrow points from this form to the generated XML code on the right.

```
foaf.xml
```

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:admin="http://webns.net/mvcb/"
  <foaf:PersonalProfileDocument rdf:about=""
    <foaf:maker rdf:resource="#me"/>
    <foaf:primaryTopic rdf:resource="#me"/>
    <admin:generatorAgent rdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
    <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
  </foaf:PersonalProfileDocument>
  <foaf:Person rdf:id="me">
    <foaf:name>tatsuya hagino</foaf:name>
    <foaf:title>Mr</foaf:title>
    <foaf:givenname>tatsuya</foaf:givenname>
    <foaf:family_name>hagino</foaf:family_name>
    <foaf:mbox rdf:resource="mailto:hagino@sfc.keio.ac.jp"/>
  </foaf:Person>
</rdf:RDF>
```

# まとめ

---

- ▶ Semantic Webとは
  - ▶ 機械可読なメタデータを、複数のアプリケーションをまたいで共有、統合できるようにする仕組み
- ▶ RDFのグラフ構造を持つデータモデルにより、シンプルかつ柔軟なデータ統合が実現
- ▶ 参考文献
  - ▶ 書名: Semantic Web for the Working Ontologist
  - ▶ 著者: Dean Allemang, James Hendler