

特集2

セマンティックWebにおける メタデータとその活用

松井 くにお^{†1} (株) 富士通研究所

津田 宏^{†2} (株) 富士通研究所

上田 健次^{†3} 九州日本電気ソフトウェア (株)

小泉 雄介^{†4} (株) NEC総研

豊内 順一^{†5} (株) 日立製作所

布目 光生^{†6} (株) 東芝

■メタデータとは

セマンティックWebで重要な役割を果たすメタデータあるいはメタ情報とは「データ(情報)についてのデータ」のことである。本稿では、メタデータの歴史をひもときつつ、本稿で説明するセマンティックWebに関連する主なメタデータを紹介しよう。

メタデータの歴史

メタデータとして最も身近なのは、書籍に対する書誌情報であろう。個々の書籍について、著者やタイトルなどの属性と値とをまとめ、カードやDBなどで検索できるようになっているのが普通である。その歴史は、図書館学における書誌コントロール(bibliographic control)として16世紀まで遡ることができる。その後、情報技術の発展によりMARC(Machine Readable Cataloging)¹⁾のように、計算機上で取り扱うことが求められるようになった。人文系テキストのSGML(Standard Generalized Markup Language)化プロジェクトTEI(Text Encoding Initiative)²⁾におけるTEI Headerも有名である。

インターネットの普及とともに、ネットワーク上の情報リソース(資源)についてもメタデータの必要性が指摘されるようになった。中でも、1994年にIETF(Internet Engineering Task Force)でドラフトがまとめられたIAFA Template(Internet Anonymous FTP

Archives)³⁾は先駆的なものである。FTPサーバ内のソフトウェアなどのリソースに対して、そのメタデータを公開し、分散リソースの検索システムを効率よく運営しようというものである。RFC822準拠による、属性/値対によるテンプレートとして、文書やソフトウェア用のものが提案されたが、拡張性の少なさやエンコードの問題から実用には至らなかった。

1995年、欧米の図書館学の団体により、書誌情報およびネットワークリソースについてのメタデータ要素(エレメント)の最小セットが合意された。これが、Dublin Core⁴⁾である。わずか15要素と思われるかもしれないが、メタデータの要素集合で初めて合意をみたという点で重要である。Dublin Coreを拡張するかたちでMIReG(EU Government Metadata Framework)、GILS(Global Information Locator Service)、AGLS(Australian Government Locator Service)など欧米豪の電子政府で公開情報のメタデータが記述されつつある。

Webにおけるリソースのメタデータ

HTML内にメタデータを記述するのは、METAタグをはじめ、それを拡張したSHOE(Simple HTML Ontology Extensions)⁵⁾や、会社や商品情報のタグづけIDML(International Development Markup Language)⁶⁾といった多くの試みが行われてきた。ただし、現在HTML内のMETAタグ自体、活用するには問題がある。とい

^{†1} E-mail:matsui.kunio@jp.fujitsu.com

^{†4} E-mail:y-koizumi@pd.jp.nec.com

^{†2} E-mail:htsuda@jp.fujitsu.com

^{†5} E-mail:toyouchi@sdl.hitachi.co.jp

^{†3} E-mail:k-ueda@qb.jp.nec.com

^{†6} E-mail:kosei.fume@toshiba.co.jp

うのは、一時期検索エンジンで検索結果のランキングにMETAタグ情報を利用したところ、METAタグ内に本文と無関係の語を大量に入れることで自分のページのランキングを操作する、いわゆるワードスパム攻撃が行われたためである。

Webにおいては、プッシュなどのコンテンツ流通を効率的に行うメタデータも有用である。CDF (Channel Definition Format)⁷⁾は、チャンネルと呼ばれる文書集合に対して、その更新間隔などの情報をXMLにより記述する。それに従い収集することで、効率的にプッシュ型のインタフェースが実現できる。

Webにおける人／サービスのメタデータ

Webでは、人やサービスに関するメタデータ（あるいは「プロフィール」）も重要である。

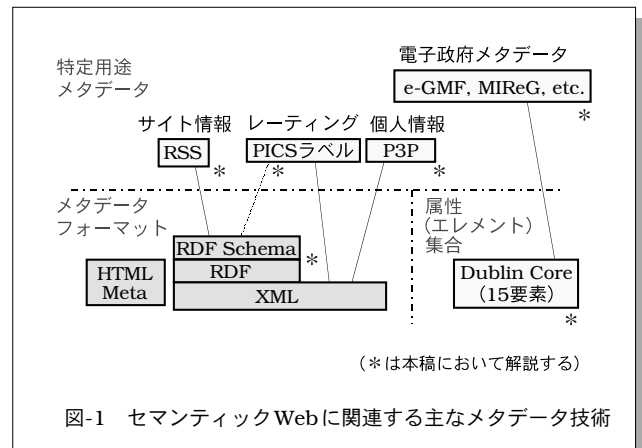
PICS (Platform for Internet Content Selection) は、閲覧者に対して暴力などの有害コンテンツのアクセスコントロールを行うためのメタデータの利用方法である。情報発信者や第三者の格付け機関がコンテンツのメタデータを付与し、クライアント（ブラウザ）でブロックするという仕組みである。

P3P (Platform for Privacy Preferences Project) は、W3C (World Wide Web Consortium) により定められた個人のプライバシー情報伝達のための枠組みである。これは、Webサイトがプライバシーポリシーをメタデータで伝えることができ、さらに利用者がブラウザ等に設定した個人情報を、どのようなサイトにどれだけ公開できるかを定めることができる。

次世代Webとして有望視されているWebサービスについてのメタデータは、WSDL, UDDI (Universal Description, Discovery and Integration), SOAP (Simple Object Access Protocol) などサービス統合のための一連の形式で定められている。詳細は本特集「セマンティックWebとは」の章にゆずる。

セマンティックWebにおけるメタデータ

セマンティックWebでは、Webリソースに対してマシンリーダブルなメタデータを付与することで、高度な検索サービスなどを可能にしようとしている。そのためには、メタデータを記述／交換するための書式がまず必要であり、それがRDF (Resource Description Framework)⁸⁾である。RDFはW3Cで1999年に勧告されている。RDFでは（リソース、プロパティ、値）の3つ組による有向グラフで情報を表現しXML (extensible Markup Language) による表現形式も持つ。RDFによりDublin Coreをはじめとする実際の多くのメタデー



タを表現することが可能である。

ただし、誰もが自由にRDFを書くだけでは、前述のMETAタグにおけるワードスパムのような攻撃に対して脆弱である。セマンティックWebを現実のものとするには、今後はメタデータの収集・流通などの運用や信頼性の検討が重要となるだろう。そして、HTMLやXMLなどのコンテンツに加えて、信頼できるメタデータも大量に存在し、人、もの、サービスが柔軟に連携する世界が形づくられていくことになるだろう。

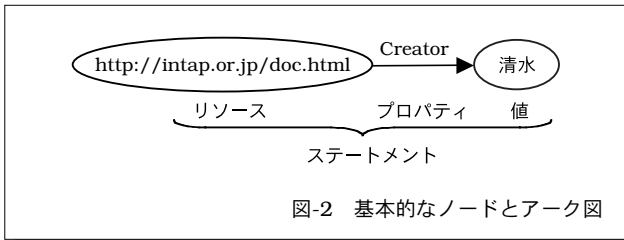
以下、こうしたメタデータに関する個々の技術を紹介していこう。図-1は、以降で紹介する個々の技術の関連を示したものである。まず、フォーマットおよび要素集合の基盤技術としてのRDFとDublin Coreを紹介する。次に、これらの技術の上に作られたメタデータの実例として、PICS, P3P, RSS, および各国電子政府メタデータを紹介する。

■ RDF

RDFはデータとその意味を記述するフレームワークで、W3Cが標準化を行っている⁸⁾。Web上の資源にメタデータをRDFで付与することにより、マシンによる意味理解と自動処理を容易にするものである。セマンティックWebにおいてはメタデータ処理基盤としてRDFが用いられる。

RDFのモデルは、項目と項目の関係を表現するだけの単純なものである。否定的な表現や曖昧な表現が含まれておらず、したがって単純明快に情報の意味を記述できることから、マシンは定義されたかたちで的確に動作することが可能となる。

RDFを記述する言語としてはXMLが用いられる。RDFのモデルを用いない単純なXMLだけでもDTDやスキーマを使うことによりWeb情報に意味を持たせる記述は可能だが、その記述方法は多岐に渡り、また使用



```
<rdf:RDF>
<rdf:Description about=" http://intap.or.jp/doc.html ">
  <s:Creator>清水</s:Creator>
</rdf:Description>
</rdf:RDF>
```

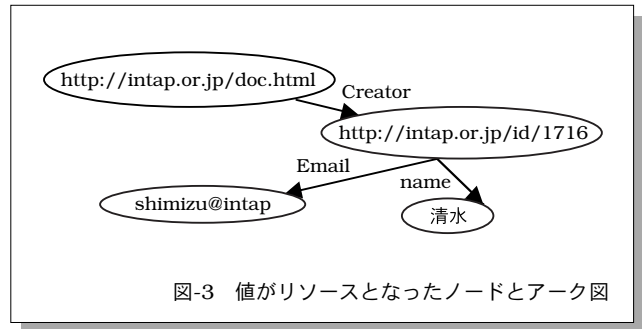
例-1 基本的な RDF 記述

```
<rdf:RDF>
<rdf:Description about="http://intap.or.jp/doc.html"
  s:Creator="清水"/>
</rdf:RDF>
```

例-2 基本的な RDF 記述

```
<rdf:RDF>
<rdf:Description about=" http://intap.or.jp/doc.html ">
  <s:Creator rdf:resource="http://intap.or.jp/id/1716"
  v:Name="清水"
  v:Email="shimizu@intap" />
</rdf:Description>
</rdf:RDF>
```

例-3 値がリソースとなった RDF 記述



```
<rdf:RDF>
<rdf:Description about="http://intap.org/doc.html">
  <dc:Creator>
    <rdf:Seq ID="CreatorsAlphabeticalBySurname">
      <rdf:li>清水</rdf:li>
      <rdf:li>上田</rdf:li>
    </rdf:Seq>
  </dc:Creator>
  <dc:Identifier>
    <rdf:Bag ID="MirroredSites">
      <rdf:li rdf:resource="http://intap.or.it/doc.html"/>
      <rdf:li rdf:resource="http://intap.or.jp/doc.html"/>
    </rdf:Bag>
  </dc:Identifier>
  <dc:Title>
    <rdf:Alt>
      <rdf:li xml:lang="it">Il Pagio di Web Fuba</rdf:li>
      <rdf:li xml:lang="ja">いけてるWebページ</rdf:li>
    </rdf:Alt>
  </dc:Title>
</rdf:Description>
</rdf:RDF>
```

例-4 コンテナを使った RDF 記述

するアプリケーションごとに異なるなど、事実上、自動処理が難しくなる。これらの問題を改善するためにセマンティック Web でのメタデータ処理基盤を RDF と定め、記述を統一している。

RDFモデルと構文

RDFモデルと構文は、W3C勧告の「RDF Model and Syntax Specification」⁹⁾で決められている。RDFモデルは基本的に「リソース」、「プロパティ」、および「値」というステートメントから構成されており、リソースにおける値の意味をプロパティで示している。たとえば、ある Web ページ <http://intap.or.jp/doc.html> の作成者 (Creator) が清水氏であることをノード (楕円の部分) とアーク (矢印の部分) を使った図解で示すと図-2 のようになる。なお、アークの方向が重要であることに注意。

また、この図解と同じ記述を、RDF/XML 記述で行うと例-1 のような構文になる。ここで、Description 要素の about 属性がリソースを示し、Creator 要素は作成者を示す。したがって、マシンが処理する場合、「<http://intap.or.jp/doc.html> の作成者 (Creator) は清水である」と理解できる。なお、完全な XML ドキュメン

トには `xmlns:rdf` や `xmlns:s` といった名前空間宣言が必要であるが、ここでは割愛する。また、RDF 文中の 2 バイトコードは UTF-8 である。

また、属性を Description 要素の XML 属性とした省略形文法を用いると例-2 のように記述することができる。

これらの RDF は、いずれも同じ意味を示す。ただし、HTML ドキュメントに埋めて利用する場合、最初のケースは Web ブラウザに値 (清水) が表示される可能性がある。

1 つのステートメントにおける値は、さらなるプロパティに対するリソースになり得る。たとえば図-3 においては、「社員番号 ID1716 番の名前は清水で、電子メールアドレスは shimizu@intap であり、<http://intap.or.jp/doc.html> の作成者である」ことを示す。

また、そのときの RDF 記述は例-3 のようになる。

さらに、RDF はコンテナを使って複数の値を記述することもできる。例-4 は、ある 2 人が作成した Web ページの日本とイタリアのミラーサイトを示す例である。ここで、Seq、Bag、Alt はコンテナとよばれ、複数の値を記述している。Seq (Sequence) は列挙された値の順序を重視するコンテナであり、この例では作成者の順

番を示している。同様に Bag は値が同等であり、例ではミラーサイトに用い、Alt (Alternative) はいずれか1つの選択を表す値のリストであり、例ではタイトルに用いている。

RDF Schema

RDF Schema は、W3C の「RDF Schema Specification 1.0」で定められ、RDF のステートメントを解釈するためのものである。RDF Schema は、今後、RDF Vocabulary Description Language となるが、ここでは RDF Schema として記述する。通常、RDF の先頭に名前空間として定義されるものである。

RDF Schema は、Java のようなオブジェクト指向プログラミング言語に似ており、リソースのプロパティ定義やプロパティとリソース間の定義、リソースのクラスとクラス間の関係と制約条件などを定義することができる。人名や年齢などのボキャブラリ自体を定義したものではなく、あくまで人名は文字列、年齢は数字であるといったカテゴリ分類やその記述方法を示すものである。

RDF Schema は、他の RDF データの流用、一部修正した再利用や拡張利用などが可能である。対象とする情報やサービスに応じて規定する必要があり、各種サービスに共通する汎用的なスキーマは標準化対象になり得る。

RDF Schema には、中心となるクラスであるリソース定義 (Resource)、プロパティ定義 (Property)、クラス定義 (Class) がある。また、Property クラスの事例でありながら他クラスとの関係を示すものとして、タイプ定義 (Type)、クラス間の関係 (SubClassOf)、プロパティ間の関係 (subPropertyOf)、などがある。さらにプロパティとクラスの制約条件を示すものとして、プロパティ値の範囲の定義 (range)、プロパティを持つクラスを示す定義 (domain) などもある。

例-5 の RDF Schema 例は、クラス階層を表現する単純な例として、親クラス「自動車」から派生させた「乗用車」、「トラック」、「バン」を定義し、さらに「乗用車」と「バン」両方の特徴を備えたサブクラスである「ミニバン」を定義したものである。

```
<rdf:RDF xml:lang="ja"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdf:Description ID="自動車">
    <rdf:type resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
  </rdf:Description>

  <rdf:Description ID="乗用車">
    <rdf:type resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#自動車"/>
  </rdf:Description>

  <rdf:Description ID="トラック">
    <rdf:type resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#自動車"/>
  </rdf:Description>

  <rdf:Description ID="バン">
    <rdf:type resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#自動車"/>
  </rdf:Description>

  <rdf:Description ID="ミニバン">
    <rdf:type resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#乗用車"/>
    <rdfs:subClassOf rdf:resource="#バン"/>
  </rdf:Description>
</rdf:RDF>
```

例-5 RDF Schema

■ Dublin Core

Dublin Core の概要

Dublin Core は、情報資源の作者、タイトル、作成日といった書誌情報の特徴を統一的に記述するメタデータの要素集合 (ボキャブラリ) を定める⁴⁾。Dublin Core が記述の対象とする情報資源は主にネットワーク上の文書オブジェクトであり、これらのオブジェクトを効率的に発見するには、どのような表現が適しているかという議論を通じて、表-1 に示す15項目の属性が定義された。

Dublin Core 制定の経緯

これらのメタデータ・ボキャブラリの定義と、相互運用可能なメタデータ標準の普及を進めているのはメタデータ研究会 DCMI (Dublin Core Metadata Initiative)⁴⁾ である。ちなみに、DCMI の活動において中心的な役割を担う OCLC (Online Computer Library Center) の予算の約9割は米国の図書館コンソーシアムのメンバから得られている。DCMI は1994年に Chicago で開催された WWW 会議での非公式会合に端を発し、翌1995年3月の第1回メタデータ研究会で正式にその活動を開始した。米 Ohio 州 Dublin に、多くのコンピュータ科学者、情報学者および図書館司書を集めて開催されたこの研究会の成果が、Dublin Core Metadata Set あるいは単に Dublin Core と呼ばれる Web 上のさまざまな情報資源を記述するための13個のメタデータ要素である。なお、

要素名	説明
Title	資源に与えられた名前
Creator	資源の内容に主たる責任を持つ者
Subject	資源の内容の主題
Description	資源の内容の明細
Publisher	資源を提供している主体
Contributor	資源の内容に貢献している者
Date	資源のライフサイクルにおけるイベントに関連する日
Type	資源の内容の種類またはジャンル
Format	資源の物理的または電子的形式
Identifier	特定の文脈での曖昧さのない資源への参照
Source	ある派生した資源から、元の資源への参照
Language	資源の内容を記述する言語
Relation	関連するリソースへの参照
Coverage	資源の範囲または領域
Rights	資源内または全体の権利の情報

表-1 Dublin Coreのメタデータ要素

Dublin Coreは1996年にはテキストとイメージの記述のため、その要素数が13から15に拡張された。

Dublin Coreの拡張と記述

Dublin Core Metadata Element Set (DCMES)¹⁰⁾という仕様で定義された15個の要素は、30カ国(25言語)間で共通要素としてのコンセンサスが得られている。Coreという言葉通り、このメタデータ要素は多くのユーザが共同に利用するための最小の基本部分であり、用途ごとに拡張が可能である。もちろん、拡張する場合には拡張内容の徹底がユーザ間で必要となり、また拡張することによって相互の互換性はなくなるの言うまでもない。また、DCMESはボキャブラリ自体を定義するが、HTMLでの記述形式の推奨形式を決めた以外には記述法や構文は定義しておらず、実際にどのような構文規則で実装するかはユーザに任されている。ただし、上述のRDFはDublin Coreを記述するのに適したフレームワークと考えられており、文献9)、12)では、Dublin Coreの記述例が示されている。

Dublin Coreの詳細化

Dublin Coreは、図書館司書のような書誌情報管理の専門家でない、一般的な文書オブジェクトの作者や出版者でも記述可能という目標のもと決定された。したがって、そのボキャブラリには、非常に単純化された

定義がなされている。しかしボキャブラリに、より詳細な意味を持たせたい場合には、この単純さがデメリットとなり得る。一例としてDateという要素は、「リソースのライフサイクルにおけるイベントに関連する日」と定義されているが、これではドキュメントの作成日と最終更新日のいずれを示すのかも、明確でない。

このような問題に対しDCMIは2000年に基本要素タイプの意味を深化する修飾子(DC Qualifier)^{11)、12)}の勧告を行った。修飾子には、要素タイプの意味の詳細化をするもの(Element Refinement)と、単位や記述ルールを明確にするためのもの(Encoding Scheme)との2種類がある。これにより、たとえば、Dateの基本タイプを作成日、有効期日または期間、利用可能日または期間、発行日、最終更新日の5種類に詳細化して利用することが可能となった。

■ PICS

PICSとはPlatform for Internet Content Selection(インターネットコンテンツ選択のための技術基盤)の略であり、W3Cが1995年夏から開発を進め、1996年から1997年にかけて仕様策定した、インターネット上のコンテンツの自主規制を目的とした技術基盤である¹³⁾。

PICSの仕様書は2002年4月現在、Ver1.1(PICS-1.1)が公開されている¹⁴⁾。PICSラベル記述方法と実装方法に関する“PICS Label Distribution Label Syntax and Communication Protocols”，フィルタリング規則の記述方法に関する“PICS Rules 1.1”，およびレイティングサービスの記述方法に関する“Rating Services and Rating Systems (and Their Machine Readable Descriptions)”からなっている。

PICSの特徴は、インターネットにおける情報発信を制限することなく、受信者が設定するレベルに合わせて、選択的に情報を受信(フィルタリング)できるようにするところにある。すなわち、あるWebコンテンツに対してPICSラベルを付加する(これをレイティングという)ことによって、受信者側でフィルタリングソフトを通じて当該ラベルを参照し、好ましくない性質のコンテンツについてはブロックすることを可能とする(図-4参照)。

PICSラベルとは、Webコンテンツを、Webページ単位、Webサイトを構成するディレクトリ単位、またはWebサイト単位で、ヌード・暴力等の一定のカテゴリ別に評価した結果を記述したものである。PICSラベルはメタデータの種類であるが、HTML言語で記述される。PICSラベルはセルフレイティング(情報発信者自身

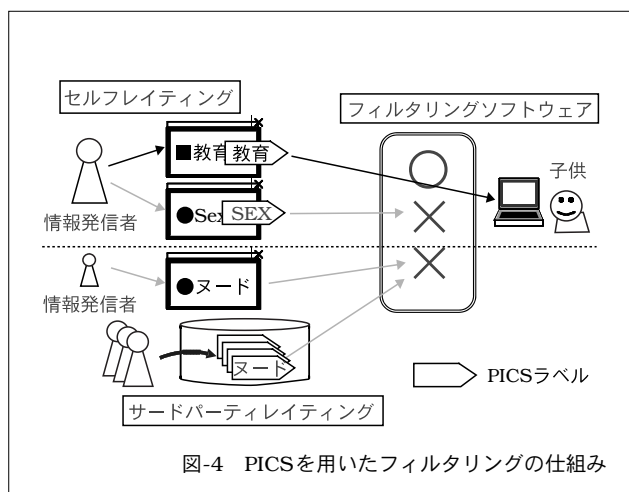


図-4 PICSを用いたフィルタリングの仕組み

が行うレイティング)の場合は、WebページにMETAタグとして付加されるか、またはHTTPのヘッダに付加され、サードパーティレイティング(発信者以外の第三者が行うレイティング)の場合は、第三者機関の保有するラベルデータベースに保存される。

以下に、PICSラベルの一例を挙げる。http://intap.or.jp/shimizuというWebのディレクトリに対して、後述のICRAが定めたレイティング基準でレイティングしたラベルデータを示している(例-6)。

PICSを用いるメリットとしては、以下の3つが挙げられる。標準データ形式と標準プロトコル方式を用いているためフィルタリングソフトウェア間やレイティングサービス間でPICSラベルの交換が可能であること、PICSラベルをWeb上のラベルデータベースで公開することによってPICS対応のフィルタリングソフトウェアを用いて世界中からそれらのラベルを参照できること、および受信者はいろいろなレイティング基準に基づくレイティングサービスの中から自分に合ったものを選択できることである。

PICS準拠のシステム・ツールとしては、英国のICRA(Internet Content Rating Association)が、サイト運営者がICRA基準のPICSラベルをWeb上で作成することができるセルフレイティングツールを2000年12月から公開している¹⁵⁾。以来、2002年3月末までに、59,000件弱のラベル登録がある。また、2002年3月にはICRAラベルをフィルタリングでき、またブラックリスト/ホワイトリスト機能のあるICRAfilterをリリースしている。

インターネット協会では、サードパーティレイティングに基づくラベルデータベースおよびサーバ型フィルタリングソフト(SFS)を無償公開している¹⁶⁾。ラベルビューロは日本語サイトを中心に約45万URLのデータを保有している(2002年3月現在)。

```
<meta http-equiv="PICS-Label" content="(PICS-1.1
"http://www.icra.org/ratingsv02.html" | gen true for
"http://intap.or.jp/shimizu"r (nr 1 ni 1 vu 1 vk 1 lz 1
ca 1 od 1))">
```

例-6 PICSラベルの例

また、MicrosoftのInternet Explorer3.0以降、またNetscapeのコミュニケーター4.5以降(6.0未満)にはPICS準拠のフィルタリング機能が組み込まれている。

PICSで表現できるメタデータは、仕様上、カテゴリ名と数値の組のみであり、著者に関する情報などをメタデータとして表すことはできない。そのため、PICSを拡張して、より多くの表現を可能にしようとする取り組みがなされてきた。W3Cの計画では、XMLとRDF以前に策定されたPICS-1.1を、RDFの規定に従って定義されたPICS-2.0に移行することになっていたが¹⁷⁾、2002年4月現在はペンディング中のようなのである。

■ P3P

P3PとはPlatform for Privacy Preferences Project(プライバシー情報取扱いに対する個人の選好を支持する技術基盤)の略で、W3Cが1997年夏から開発を始め、仕様策定を進めてきたインターネット上のプライバシー保護を目的とした技術標準である¹⁸⁾。2002年4月にP3P1.0の勧告版がリリースされた¹⁹⁾。

現在のところ、Web上のプライバシーポリシーは会社により記述形式がまちまちであり、また、その都度各会社ごとのプライバシーポリシーを読んで確かめることは消費者にとり大きな負担となる。P3Pの標準では、プライバシーポリシーの掲載項目等を標準化し、かつマシンリーダブルな形式(XML形式)で記述することによって、ポリシーを自動処理できるようにしてある。XMLで記述されたP3P準拠のプライバシーポリシーをP3Pポリシーという。

P3Pポリシーの掲載項目は、P3Pポリシーが適用されるURL、当該企業・団体に関する情報、本人情報へのアクセス、紛争解決手段、個人情報利用目的、個人情報の提供範囲、個人情報の保存期間、収集する個人情報の種類などである。

例-7に、P3Pポリシーの一例を挙げる。CatalogExampleというサイトのプライバシーポリシー(収集する情報やその利用目的等の情報を含む)を示している。

消費者側では、P3P対応のクライアントツールによって、個人情報収集画面においてP3Pポリシーを参照したり、あらかじめ登録しておいた消費者のプリファレン

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
  <POLICY name="forBrowsers"
    discuri="http://www.catalog.example.com/PrivacyPracticeBrows
    ing.html" xml:lang="en">
    <ACCESS><nonident/></ACCESS>
    <STATEMENT>
      <PURPOSE><admin/><develop/></PURPOSE>
      <RECIPIENT><ours/></RECIPIENT>
      <DATA-GROUP>
        <DATA ref="#dynamic.clickstream"/>
      </DATA-GROUP>
    </STATEMENT>
  </POLICY> T
```

例-7 P3Pポリシーの例

ス (Webサイトに提供してもよい個人情報およびその利用目的と提供先について規定した個人情報利用原則) と P3P ポリシーとを照合して、個人情報を提供するか否かの判断を半自動で行ったりすることもできる。

このように P3P とは企業が Web 上で公開するプライバシーポリシーに基づいて、消費者がその企業の信頼性を自動的に判断するための技術標準である。消費者のプライバシー意識が高まる中、電子商取引の効率化を促進する技術として、今後の普及が期待されている。

P3P を実際に運用するためには、Web サイト側 (サーバ側) と Web 利用者側 (クライアント側) の両方で P3P を導入する必要がある。

サーバ側のツールとしては、IBM、AT&T、NEC 等が、Q&A 方式で比較的容易に P3P ポリシーを作成し Web サイトに導入することができる P3P 対応 Web サイト構築ツール (P3P ポリシージェネレータ) を試験的に開発している。また、インターネット協会が 2001 年 3 月に日本語サイト向けの P3P ポリシージェネレータを開発し、Web 公開した²⁰⁾。インターネット協会の P3P ポリシージェネレータは P3P1.0 の勧告候補版に対応しており、タグ画面上で質問に回答することにより容易に P3P ポリシーを作成できる。

クライアント側のソフトウェアとしては、Microsoft が 2001 年にリリースした Internet Explorer 6.0 に P3P 対応のクッキー管理機能を標準装備されている。

2002 年 4 月現在、Microsoft、AOL、AT&T、IBM、HP、米商務省等、米国のサイトを中心に、W3C サイトに登録されているだけでも約 380 サイトが P3P 対応となっている²¹⁾。日本のサイトでは、BIGLOBE、ニフティ、インターネット協会等が対応している。

■ RSS

RSS (RDF Site Summary) は、Web サイトの要約を RDF のメタデータとして簡単に記述する方法で、Syndication フォーマットである²²⁾。RSS における Syndica-

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns
  xmlns="http://purl.org/rss/1.0/">

  <channel rdf:about="http://www.intap.or.jp/xml/news.rss">
    <title>INTAP</title>
    <link>http://intap.or.jp/tech</link>
    <description>
      INTAPの技術情報サイトです。
    </description>
    <items>
      <rdf:Seq>
        <rdf:li resource="http://intap.or.jp/tech/rdf/rdf.html" />
        <rdf:li resource="http://intap.or.jp/tech/rss/index.html" />
      </rdf:Seq>
    </items>
  </channel>

  <item rdf:about="http://intap.or.jp/tech/rdf/rdf.html">
    <title>RDF技術情報</title>
    <link>http://intap.or.jp/tech/rdf/rdf.html</link>
    <description>
      INTAPで調査したRDFに関する技術情報を掲載しています。
    </description>
  </item>

  <item rdf:about="http://intap.or.jp/tech/rss/index.html">
    <title>RSS技術情報</title>
    <link>http://intap.or.jp/tech/rss/index.html</link>
    <description>
      INTAPで調査したRSSに関する技術情報を掲載しています。
    </description>
  </item>

</rdf:RDF>
```

例-8 RSSの構文

tion とは、検索や収集のためにデータをオンラインで使用可能にすることやオフラインで発行可能とすることである。RSS は、各サイトの要約情報を集約、管理するためのものから発展し、軽量で簡潔なため、ニュース配信や製品紹介、フォーラム、Web ページのカスタマイズなど、実にさまざまな分野で利用されている。セマンティック Web ではさまざまなコンテンツや情報を RDF で意味を持たせるという観点から、RSS の応用はセマンティック Web におけるある特定領域の一事例といえる。

構文についての概要を説明すると、RSS1.0 においては、基本部分であるコア構文と拡張性を持たせたモジュールがある。

コア構文は、RSS 文書ごとに channel 要素と 1 つ以上の item 要素からなる。channel 要素にはタイトル (title)、RSS の対象 URI (link)、概要説明 (description)、item 要素で記述するリソースの目次 (items) などの必須項目のほか、ロゴ (image) や文字列入力 (textinput) がある。Item 要素には、item のタイトル (title)、リソースの URI (link)、リソースの要約 (description) がある。

例-8 は、コア構文を使って書いた RSS 例である。

RSS のモジュールでは、XML の名前空間を用いて、

独自の要素型を追加定義することが可能である。現在、一般的には、Dublin CoreとSyndicationモジュールがある。たとえば、コア構文に含まれないポキャプラリーとして日付や更新日などを表すものとして、Dublin Coreの日付(Date)要素などが有効である。

■各国電子政府にみるメタデータの標準化と適用事例

情報資源記述手段としてのメタデータを、利用分野に応じて標準化・実装する試みがあるが、その1つとして各国電子政府における行政公開文書へのメタデータ付与の事例がある。

行政公開文書はその性質上、表題や作成日といった情報に加え、廃棄時期や保存場所、管理担当部署、関連条項といった管理情報が必要である。従来のような管理情報に加え、今後の利用が想定される情報資源記述を含み、さらに横断的に構造化されたメタデータ定義セットの標準化が行われれば、これに準拠した文書管理を行うことで省庁間や国別に散在する情報資源を最大限に有効活用できる基盤が整うことになる。

現状では、メタデータ標準の要素集合として最小限の管理情報を記述するDCMES(前述)が存在するが、この基本15要素そのものは定義内容が明確で一般に普及しやすいという利点がある一方、精度の高い情報を記述する場合やコミュニティ独自の情報記述のためには要素の拡張や精密化が必要となる。

Dublin Coreの制約に準拠し、各国ではその実情に合わせて要素・限定子の拡張が行われている。英国のe-GMS(e-Government Metadata Standard)²³⁾、デンマークのOIO-metadata(Offentlig Information Online Metadata)、アイルランドのIPSMS(Irish Public Service Metadata Standard)、フィンランドのFinnish Dublin Core extension、オーストラリアのAGLS(Australian Government Locator Service)²⁴⁾、ニュージーランドのNZGLS(New Zealand Government Locator Service)、米国のGILS(Government Information Locator Service)等がその例である。

EU標準メタデータセット

行政公開情報のメタデータ標準化の1つとして、EU標準のメタデータセット(MIReG Metadata Element Set)「(2002年1月Draft)²⁵⁾がある。

これは英国のe-GMSとほぼ同様であるが、その基本方針は利用方法が単純であること、国際標準や他のEC標準に準拠すること、拡張可能であることを目標とし

エレメント名	参照	要素の概要	(参考) 国内行政文書項目との対応
Contributor	DC	(Creatorに次いで) 貢献した人・組織	(編集者・翻訳者)
Coverage	DC	情報資源内容の適用を受ける期間、場所	*
Creator	DC	情報資源の作成にあたり最も責任を有するもの	作成者
Data	DC	情報資源そのものに関する日時情報	作成(取得)時期
Description	DC	情報資源に含まれる情報の説明	タイトル/(備考)
Format	DC	情報資源の記述形式	*
Identifier	DC	内容に対する一意の識別番号	*
Language	DC	記述されている言語	(ja)
Publisher	DC	情報資源を発行した責任者	管理担当課・係
Relation	DC	関連する情報資源への参照	*
Rights	DC	情報資源全般の権利情報	*
Source	DC	提示された情報資源への参照	*
Subject	DC	主題およびキーワード	大・中・小分類
Title	DC	タイトル	タイトル
Type	DC	情報資源の種類	(分類カテゴリ)
Audience	MIReG	利用者のカテゴリ	*
Disposal	MIReG	情報資源の保持と処理	保存期間・廃棄時期 保存期間満了時の措置結果
Location	MIReG	情報資源の物理的な場所	保存場所
Preservation	MIReG	永久保存資源のための備考データ	(保存期間・保存場所)
Function	AGLS	資源内容が関連する業務種別	*
Availability	AGLS	(主としてオフライン資源に接触するための) 資源作成者/管理者情報	*
Mandate	AGLS	資源作成のための法的根拠、条項の記述	*
Keywords	Canada	資源のキーワード(検索システムに対する単語の重み付けに影響)	*

表-2 Dublin Coreと各国要素の拡張例

ており、柔軟性のある定義となっている。例-9は、メタデータ記述例で、適切なMIReG名前空間がDublin Coreの精密化として定義されているものと仮定する。

文書に対する精密化記述は適切に限定される必要がある。たとえば要素名“coverage”(リソース内容の表す期間と場所の定義)の利用であれば、データ内容を広く包含する情報を付加するのではなく、情報資源を特定できるような空間的・時間的な制約となることが望ましい。メタデータの記述例として、たとえば例-10のような表現も可能である。


```
<rdf:RDF xmlns:rdf="http://www.w3.org/RDF/RDF/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcq="http://purl.org/dc/terms/"
:
<rdf:Description
about="http://sample.go.jp/doc/sample.html">
<dc:title>委員会名簿リスト</dc:title>
<dc:creator>creator@sample.go.jp</dc:creator>
<dc:description>地方議会、常務委員会のメンバリスト。
連絡先の詳細や委員会での役割を含む。
</dc:description>
</dc:coverage.temporal>
<dc:data.create>2002-02-21</dc:data.create>
<dc:format>text/html</dc:format>
:
```

例-9 メタデータ記述例

```
例：2003/04の期間に徴収される2002/03期の納税
申告（文書）に対する場合
COVERAGE.TEMPORAL.BEGINNING
DATE:2002-04-01 END DATE:2003-03-31
START DATE OF CAPTURE:2003-08-01 END
DATE OF CAPTURE:2004-04-01
```

例-10 情報資源の記述例

国で活用され始めた事例について述べた。メタデータはその効用を示し、かつ標準化動向を常に見据えながら、コンテンツの充実と並行して作成していくことが肝要となる。周辺の関連ツールなどの研究開発も重要な課題である。

標準化の動き

電子化政府を対象としたメタデータエレメントセットは、DCMI²⁵⁾やDC-Gov (Government Working Group)での議論を中心に欧州(北欧)、豪州、米国といった国々での実際の利用を経てフィードバックされ、確実に広まっている。

我が国でも近年、電子政府の総合窓口(e-Gov)をはじめとして電子化政府への取組みが進んでいるが²⁶⁾、情報資源の記述(メタデータ)の利用に関しては、こうした欧米の動きにやや遅れをとっている印象を受ける。

単純な例として、現在の「電子政府の総合窓口」(政府ポータルサイト)における『リサイクル』の検索結果(一部)を例-11に示す。この結果を見る限りは、カテゴリごとの意味構造を見出すことは難しい。ここに、一定の基準に沿ったカテゴリ定義、すなわちメタデータ付与の統一的枠組みがあれば、その枠組みを活用し、たとえば時系列・地域別といった、より有意な検索機能を提供できると思われる。また、メタデータ利用によるこうした検索機能は、各国の電子政府窓口において、すでに実現されている例も多い。

■おわりに

本稿では、メタデータの意味するところと、世界各

参考文献 (URL)

- 1) <http://www.loc.gov/marc/>
- 2) <http://www.tei-c.org/>
- 3) IAFA Template: Deutsch, P., Emtage, A., Koster, M. and Stumpf, M.: Publishing Information on the Internet with Anonymous FTP (1994).
<http://www.roads.lut.ac.uk/Reports/iafa-draft/iafa-draft.html>
- 4) Dublin Core Metadata Initiative, <http://dublincore.org/>
- 5) <http://www.cs.umd.edu/projects/plus/SHOE/>
- 6) <http://www.idmlinitiative.org/>
- 7) <http://www.w3.org/TR/NOTE-CDFsubmit.html>
- 8) <http://www.w3.org/RDF/>
- 9) <http://www.w3.org/TR/REC-rdf-syntax/>
- 10) <http://dublincore.org/documents/dces/>
- 11) <http://dublincore.org/documents/dcmes-qualifiers/>
- 12) Expressing Qualified Dublin Core in RDF/XML,
<http://dublincore.org/document/2001/11/30/dcq-rdf-xml>
- 13) W3CのPICSに関するWebページ: <http://www.w3.org/PICS/>
- 14) PICS仕様書は次のURLにある。
<http://www.w3.org/TR/REC-PICS-labels>,
<http://www.w3.org/TR/REC-PICSRules>,
<http://www.w3.org/TR/REC-PICS-services>
- 15) ICRAのWebサイト: <http://www.icra.org/>
- 16) (財)インターネット協会のレーティング/フィルタリングに関するWebページ: <http://www.iajapan.org/rating/>
- 17) Berners-Lee, T.: W3C Data Formats W3C NOTE (29 Oct. 1997).
<http://www.w3.org/TR/NOTE-rdfarch>
- 18) W3CのP3Pに関するWebページ: <http://www.w3.org/P3P/>
- 19) P3P1.0仕様書は次のURLにある。
<http://www.w3.org/TR/2002/REC-P3P-20020416/>
- 20) (財)インターネット協会のプライバシー情報管理システムに関するWebページ: <http://www.nmda.or.jp/enc/privacy/index.html>
- 21) http://www.w3.org/P3P/compliant_sites
- 22) RSS 1.0, <http://purl.org/rss/1.0/spec>
- 23) The UK e-Government Metadata Standard v2.
- 24) The AGLS pages on the National Archives Website:
<http://www.naa.gov.au/agls>
- 25) The MIReG Metadata element Set, Draft v0.1
- 26) 電子政府の総合窓口: <http://www.e-gov.go.jp>

(平成14年5月21日受付)

大分類名	中分類名	小分類名	タイトル1行目	タイトル2行目	タイトル3行目
廃棄物・リサイクル	循環企画	リサイクル	第5, 6回ごみ減量化国民大会	—	—
廃棄物・リサイクル	循環企画	リサイクル	分別収集促進計画, 市町村分別収集計画	愛知県	—
廃棄物・リサイクル	循環企画	リサイクル	平成12年度	保管施設一覧	岩手県, 宮城県, 秋田県, 山形県, 福島県
廃棄物・リサイクル	循環企画	広域臨海環境整備センター	フェニックス法 大綱・要綱等作成過程	大綱, 要綱, 五六予算, 件名登録, 関係資料	昭和55年8月~ 56年1月

例-11 検索結果

