

メタデータ配信による画像資料活用の可能性

黒崎浩行 (國學院大學神道文化学部専任講師)

はじめに：インターネット公開された画像資料を研究活用に結びつけるには

國學院大學学術フロンティア構想「劣化画像の再生活用と資料化に関する基礎的研究」では、大場磐雄・折口信夫・桜井満・柴田常恵・梶山林継写真資料のデジタル化が進められてきた。その成果の一部は平成13(2001)年10月から順次ウェブサイト上で公開されている(<http://www2.kokugakuin.ac.jp/frontier/>)。

事業の成果を広く公開し、国内外の研究目的利用に応え、さらには新たな研究の方向を創発していくために、この画像資料を含む成果のインターネット公開は、大きな意味を持っている。

だが、公開される画像資料の総体、あるいは個々の画像資料は、どのようにすれば、それを必要とする利用者の目に留まり、またその研究活用を促すことになるのだろうか。

このようなある種当たり前の問いを発すれば、各コレクションの概説や、画像データとともに入力された資料目録データを公開し、それに対する検索インターフェイスを提供することができれば、この要求に応えることができるのではないかと、即座に答えられるかもしれない。

実際はそう単純なことではない。インターネット上には今や膨大な量の情報が公開されている。利用者は、たまたま知っているサイトや、サーチエンジンで上位にヒットするサイトを訪れることで精一杯というのが現状だろう。

WWW (World Wide Web) の最大の特徴は、ハイパーリンクによって、情報間の関係の網の目を張りめぐらせることができる、ということにある。ウェブサイトの URL (Uniform Resource Locator: インターネット上の情報源のアドレスを一意に表す表記方法) をあらかじめ知っていれば、そのサイトに直接アクセスすることができるが、たいていはそうではなく、別のウェブサイト (サーチエンジンを含む) からのリンクの網の目を探索して訪れる。サイトを設置して発信する側は、このリンクの網の目に加わっていき、さらにはその形成に働きかけることによって、そのサイトが利用者にとって「たまたま知っているサイト」や、「サーチエンジンで上位にヒットするサイト」になるように、さまざまな工夫を凝らしてきた。その努力なしでは広範な利用が望めないからである。

それは従来、たとえばサーチエンジン対策として、適切なキーワードやタイトルをページに埋め込んだり、関係するサイトにリンク登録を依頼したりする、といった形で経験則的に取り組まれてきた。その一方で、情報間の意味のあるリンクづけをコンピュータによって自動化できるよう、データの「意味」についての記述・交換方式を定め、インターネット上のサーバ間での相互運用性を高めようという構想が提唱されている。これはセマンティック・ウェブ (Semantic Web) と呼ばれ、さまざまな分野でその実現が期待されてきた。本稿では、その動向を押さえつつ、とりわけ「メタデータ配信」という側面から、この事業の成果を有効ならしめる情報公開の方向性を考察してみたい。

1. セマンティック・ウェブとその技術動向

WWW の技術標準を協議する団体、W3C (World Wide Web Consortium) のウェブサイトには、WWW の提唱者であるティム・バーナーズ=リー (Tim Berners-Lee) が記した1999年の文書、「Web Architecture: Describing and Exchanging data」が公開されている。この文書のなかで、セマンティック・ウェブ

は「明示的に事物間の関係を記述し、機械によって自動的に処理できるように意図された意味論的情報を含んでいる文書群、またはその一部からなるウェブ」と定義されている (Berners-Lee 1999)。

機械可読な意味情報の記述は「メタデータ (metadata)」と呼ばれ、これに関連して RDF (Resource Description Framework) Model and Syntax という規格が、1999 年に提唱されている (W3C 1999, 図 1)。「メタデータ」とは、「データについてのデータ」ということだが、より具体的には、あるウェブ上のリソース (資源) について、そのプロパティ (特性、性質) を記述したデータを指す用語である。RDF Model and Syntax は、その記述方法を次のように規定している。

まず、リソースを指す主語 (Subject)。これは RDF においては、URI (Uniform Resource Indicator) という、それを一意に表す記法によって示される。次に、そのリソースがもつ性質を述べる述語 (Predicate)。さらに、述語が指し示す目的語 (Object)。RDF はこの基本 3 要素からなる。例えば、「<http://www2.kokugakuin.ac.jp/frontier/ooba/>の名称は大場磐雄資料目録である」や、「<http://www2.kokugakuin.ac.jp/frontier/ooba/>の制作者は國學院大學學術フロンティア事業実行委員会である。」という文は、RDF の要件を満たしている。<http://www2.kokugakuin.ac.jp/frontier/ooba/> が主語、「大場磐雄資料目録」、「國學院大學學術フロンティア事業実行委員会」が目的語、「~の名称は~である」、「~の制作者は~である」が述語にあたる。上記の例では目的語が字句 (literal) そのものだが、別のリソースによって表現することもできる。すなわち、先の 2 番目の例の「國學院大學學術フロンティア事業実行委員会」を <http://www2.kokugakuin.ac.jp/frontier/> という URI に置き換え、「その名称は國學院大學學術フロンティア事業実行委員会である」といった文をそこに加えてもよい。このように、RDF 文は入れ子構造にすることが可能になっている。

また、RDF 文は通常、上記のような自然言語ではなく、文書マークアップ言語である XML (Extensible Markup Language) を用いて記述する (図 2)。それによって、文をコンピュータが解釈し、処理を行うことが可能になる。

RDF は意味を記述するための「構文 (syntax)」を提供する規格であり、具体的な対象について、その「意味 (meaning)」をどのような語彙を使って表現するか、どのような事柄を「意味」として拾い上げるのか、ということについては定めていない。

これはオントロジー (ontology) と呼ばれ、しばしばクラス階層によって、語彙の体系をまとめあげていく試みとして研究されている (Maedche 2002)。

書誌情報に関するメタデータ語彙を定めた規格としては Dublin Core Metadata Element Set (IS015836) が有名である。ここには、書誌情報を記述するための 15 の要素が定められている (title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights)。

2. 人文科学での利用における問題

セマンティック・ウェブを構築し、データを広汎に流通させるには、意味記述に用いる共通語彙を確立し、それを定着させるオントロジーの普及がポイントとなる。そうしなければ、相互に互換性のない意味記述が交錯するにとどまると予想されるからだ。しかし、そのようなことは可能なのだろうか、という疑問が頭をもたげる。「...Semantic Web が提供する情報の共有は、人の思考資産というレベルの話でなく、企業間取引の様な定型の情報のみが対象となり、これに応える ontology は、データ交換用プロトコルとしての意味が表現できれば十分になる」(大矢 2001 P.15)。大矢は、それぞれの専門領域を背景にもつ人文科学研究者が望んでいるのは、ontology による定型的な共通語彙の提供で

はない、とする。そうではなく、むしろその基礎となっている SGML/XML によって「自然言語で示された情報を記述する一般的な手法を確立すること」(同) が重要なのだと主張する。

大矢の議論はおおむね、研究者が手持ちの文字資料に対してマークアップ (タグづけ) を行って、その構造と解釈を提示するような研究を視野に置いていると思われる。たしかにそうした視点からの研究では、記述のための一般的な「構文 (syntax)」の機械可読な形式・手法を第一に目指すことが妥当であろう。

だが、セマンティック・ウェブのコンセプトを人文科学研究に適用するさいには、機械可読な意味記述方法の追求という理論的な視点と同時に、未知のデータへのアクセスを容易にし、それによって知見を豊かにするという実践的な視点も考慮に入れなければならないだろう。

その有効な打開策にヒントを与えるような取り組みとしては、国立国文学研究資料館が開発を進めている「メタデータベース・システム」が挙げられよう。さまざまな研究機関が独自に構築・管理するデータベースを包括的に検索できるようなサービスを提供しようとするさい、一つのサービス提供組織にデータベース実体を寄託し、そこにおいて共通のデータ属性項目への統一的な変換を行い、そのように加工された包括的なデータベースに対して検索するインターフェイスを提供する、という方法が考えられる。この方法の最大の問題点は、データベース実体が、その構築に責任と権利をもつ研究機関の手をいったん離れ、サービス提供組織による変換・加工がなされることにある。また、実際の運用上の負担も大きい。これに対して「メタデータベース・システム」では、「ゲートウェイ・システム」の手法をとる。すなわち、インターネット上でデータベースを提供・公開している複数の研究機関のサーバに対し、それぞれのデータベースで用いられている多様な属性項目を共通の属性項目にマッピングし、これに対する検索インターフェイスを提供することで、複数のデータベースを結合した横断的な検索を可能にしているのである。共通の属性項目には Dublin Core メタデータが採用されている (原・安永 2001)。

つまり、人文科学分野においてオントロジーの共通規格を定めるさい、それを各資料の記述に直接利用することを想定する必要はなく、また望ましくもない。そうではなく、まずは資料の性質に即した語彙による機械可読な記述を行い、それらに対する包括的な検索が必要な場合に限り採用するものと考えべきだろう。ただし、データベース構築のさいに 5W1H のような基礎知識として Dublin Core などを習得すべき必要性はいうまでもない。

3. 一般ユーザーレベルの新たな兆候

実はこの一年の間に、以上のような専門研究レベルの動向とは全く別の流れから、RSS (RDF Site Summary) (Libby 1999) (Brickley et al 2000) という、RDF のひとつの応用形態が一般ユーザーレベルで注目され、新たな広がりを見せはじめている (図 3)。それは、ウェブログ、Web 日記、Wiki、CMS (Contents Management System) などといった、ユーザーがオンラインでコンテンツを更新管理できる Web アプリケーションの急速な普及と関わりがある。

これらのアプリケーションでは、更新手続きと同時に、更新内容を RSS 形式で自動的に出力、公開する機能が備わっていることが多い。RSS には、そのサイトの概要と、更新された各記事の要約、および記事本文へのリンクが記述される。RSS を自動的に作成する機能が Web アプリケーションに組み込まれているため、サイト管理者は特に意識することなくコンテンツそのものの更新に集中することができる。

いっぽう、公開された RSS データは、インターネット上の別のサイトが定期的に収集し、「ヘッド

ライン」などとして紹介することができる。また、個人ユーザーが、更新された情報のみを巡回したいときに、クライアントマシン上で RSS リーダーと呼ばれるソフトを用いて閲覧することができる。

サイトの最新情報が効率的に配信、閲覧できることにより、ユーザーが関心をもつ複数のサイトをまとめて訪れることが容易になっている。さらには、サイト同士の言及による相互リンクの活性化をも加速する結果を生んでいる。

ウェブログではさらに、「トラックバック」(trackback) という機能が相互リンクを促進している。代表的なウェブログ管理ソフトウェア、Movable Type (<http://www.movabletype.org/>) に実装されているもので、あるウェブログ運営者が他人のウェブログ記事に対してコメントしたいときに、記事を自分のサイトに掲載すると同時に、その URL と要約を含むデータ (XML 形式で記述されたもの) を、元記事のサイトに送信することができる。そのデータを受信した元記事のサイトには、コメント記事へのリンクが自動的に追加される。それにより、元記事の読者は、この記事に対する他のサイトでのコメントの存在を知り、リンクをたどってそれを読むことができる。これは従来、リファラー (リンク元) ログを公開するという手法でも行われてきたことだが、こちらがあくまで機械的にリンク元をリストアップしていく仕組みであるのに対し、トラックバックはサイト間の意味のあるインタラクション (相互行為) を可視化する効果を意図的に狙っているといえよう。また、掲示板などは同一サイト内に記事とコメントの双方を掲載していくが、それらとコミュニティ形成のうえで違いを生んでいく可能性にも関心が持たれている。

RSS の配信やトラックバックにより、そのサイトへの単なるリンクではなく、そのサイトの最新情報の要約を含んだリンクが可能となり、さらに共通の関心によるハイパーリンクの形成とそれを通じたアクセスが加速しているのである。

4 . セマンティック・ウェブの展開が画像資料のインターネット公開に与える示唆

このように、一般ユーザーレベルでセマンティック・ウェブのインフラが急速に拡大しつつあるが、それらは「テキスト系サイト」と総称されることがあるように、文字情報が主体である。画像情報をセマンティック・ウェブのなかに生かすには、さらにもう一段階の仕掛けが必要となるだろう。

これまでのデジタル化された画像資料の公開は総じて、サムネイルギャラリー、スライドショーによる順次閲覧型のインターフェイスや、キーワード入力による検索型のインターフェイスによって提供されてきた。検索結果表示が画像資料のメタデータ提供になっているということもできるかもしれないが、それとは分離して画像データのみがリンクされる可能性を否定することはできない。そうなった場合、元の検索結果に示されたメタデータ記述は、その検索インターフェイスを利用せずに直接リンクをたどってきたユーザーにとっては失われてしまう。

このことは従来、知的財産権保護の文脈から問題視されてきた。そのために、画像データの直接リンク禁止を主張し、また実際に直接リンクできないように技術的な対処を行うケースがしばしば見られてきた。しかし、画像資料のインターネット公開による研究活用促進という視点からすれば、むしろ発想を逆にすべきであろう。

すなわち、その画像データに直接リンクすることを許可し、それと同時に、その画像データにリンクしているサイトへの逆方向のリンクを、画像データ提供サイトにおいて行うのである。具体的な方法は、先に挙げたトラックバックやリファラーログの公開が考えられる。このようにすれば、画像データの提供のみでなく、その画像資料をめぐるインタラクションをすべてのユーザーにとって可視化できるようになる。リンクのなかには不適切なものが含まれるかもしれないが、そうした判断は可視

化によってはじめて可能になる。

また、画像データとともに付加情報として、その画像資料についてのメタデータを記述した RSS を提供する（図4）。これは、画像データベースから自動的に生成することが可能である。ユーザーはこの RSS にリンクを張るか、あるいは自サイト内に取り込むことで、画像資料への適切な参照を行うことが可能となる。

これはデータ提供者側にとってもメリットとなる。なぜならば、画像データのみが直にリンクされることにより不正確、不適切な文脈に置かれることを未然に防ぎ、また著作権表記などの配布条件に関わる情報をも付加することができるからである。

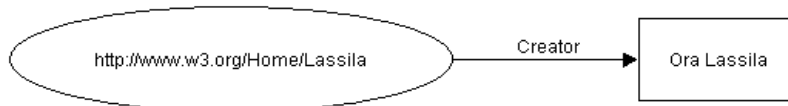
書誌情報をベースとした Dublin Core にもとづく RSS の記述語彙が、画像資料にとってどこまで妥当かという問題や、サイトの更新情報収集以外の場面でメタデータを自動的に集約し関連づけるためのサービスやツールが未成熟であることなど、いくつかの問題を抱えてはいるものの、本格的な公開データベースの構築時にはぜひ検討すべき課題として提案しておきたい。

参考文献

- Berners-Lee, Tim. 1999 *Web Architecture: Describing and Exchanging Data*. W3C Note 7, June, 1999.
<http://www.w3.org/1999/06/07-WebData>
- Brickley, Dan et al. 2000 *RDF Site Summary (RSS) 1.0*. RSS-DEV Working Group. <http://purl.org/rss/1.0/spec>
- 原正一郎・安永尚志 2001 「メタデータによるマルチメディアデータ統合の試み」『情報処理学会研究報告』2001-CH-51 (2001): 47-54。
- ISO TC 46/SC 4. 2003 ISO 15836:2003(E) *Information and documentation -- The Dublin Core metadata element set*. 2003-02-26, International Standard. <http://www.niso.org/international/SC4/n515.pdf>
- Libby, Dan. 1997 *RSS 0.91 Spec, revision 3*. Netscape Communications.
<http://my.netscape.com/publish/formats/rss-spec-0.91.html>
- Maedche, Alexander. 2002 *Ontology Learning for the Semantic Web*. Boston: Kluwer Academic Publishers.
- 大矢一志 2001 「SGML/XML データと記述に関する規則: RDF と人文科学研究との接点」『情報処理学会研究報告』2001-CH-51: 11-16。
- World Wide Web Consortium. 1999 *Resource Description Framework (RDF) Model and Syntax Specification*. W3C Recommendation 22 February 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>

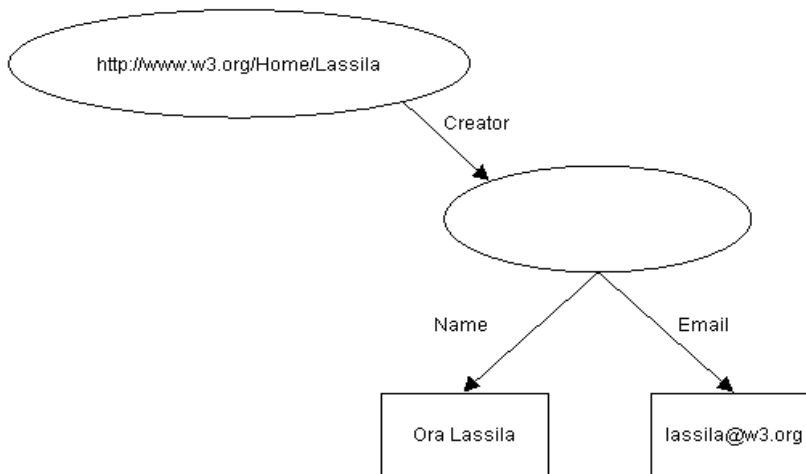
Subject (Resource)	http://www.w3.org/Home/Lassila
Predicate (Property)	Creator
Object (literal)	"Ora Lassila"

In this document we will diagram an RDF statement pictorially using directed labeled graphs (also called "nodes and arcs diagrams"). In these diagrams, the nodes (drawn as ovals) represent resources and arcs represent named properties. Nodes that represent string literals will be drawn as rectangles. The sentence above would thus be diagrammed as:



[D](#)

Figure 1: Simple node and arc diagram



[D](#)

Figure 2: Property with structured value

図 1 RDF の概念図 (World Wide Web Consortium1999)

```
<rdf:RDF>
<rdf:Description
about="http://www.w3.org/Home/Lassila">
<s:Creator>Ora Lassila</s:Creator>
</rdf:Description>
</rdf:RDF>
```

図 2 XML による記述例 (World Wide Web Consortium1999)

```

<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <channel>
    <title>神道と日本文化の国学的研究発信の拠点形成</title>
    <link>http://21coe.kokugakuin.ac.jp</link>
    <description>國學院大學21世紀COEプログラム</description>
    <lastBuildDate>Sat, 1 Nov 2003 3:11:51 JST</lastBuildDate>
    <docs>http://backend.userland.com/rss/</docs>
    <generator>XOOPS 2.0.2</generator>
    <category>News</category>
    <managingEditor>21coe-info@kokugakuin.ac.jp</managingEditor>
    <webMaster>21coe-info@kokugakuin.ac.jp</webMaster>
    <language>ja</language>
    <image>
      <title>神道と日本文化の国学的研究発信の拠点形成</title>
      <url>http://21coe.kokugakuin.ac.jp/themes/21coepurple/images/fujisanicon.gif</url>
      <link>http://21coe.kokugakuin.ac.jp</link>
      <width>88</width>
      <height>31</height>
    </image>
    <item>
      <title>ミニ・国際シンポジウム「神道 はどう翻訳されているか(2) 近現代の神道を中
        心に」</title>
      <link>http://21coe.kokugakuin.ac.jp/modules/news/article.php?storyid=53</link>
      <description>[中略] </description>
      <pubDate>Thu, 30 Oct 2003 5:20:26 JST</pubDate>
      <dc:date>2003-10-30T17:20:26+09:00</dc:date>
      <guid>http://21coe.kokugakuin.ac.jp/modules/news/article.php?storyid=53</guid>
    </item>
  </channel>

```

図3 RSS の例 <http://21coe.kokugakuin.ac.jp/backend.php>



2023 伊号住居址炉
昭和25年4月9日
82mm×107mm



RDF

```

<?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:kf=" http://www2.kokugakuin.ac.jp/frontier/ " >
  <rdf:Description about=" ... " >
    <dc:title>伊号住居址炉</dc:title>
    <dc:date>1950-04-09</dc:date>
    <dc:format>82mm×107mm</dc:format>
    <dc:creator>大場磐雄</dc:creator>
    <kf:image>
      http://www2.kokugakuin.ac.jp/frontier/ooba-photo/ob2023.jpg</kf:image>
    <dc:rights>Copyright © 2003 Kokugakuin University</dc:rights>
  </rdf:Description>

```

図4 メタデータ配信のもたらす可能性

画像資料公開のさい、説明・ナビゲーション・検索手段だけでなく、ユーザーが画像資料に対して「意味づけを行うリンク」をするためのデータを提供