

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.13 No.3 (Jul. 2003)

~~~~~ 目次 ~~~~~

## 特集「科学技術データの活用」

|      |                                                                                                               |                                                               |     |
|------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|-----|
| 巻頭言  | 会長に就任して .....                                                                                                 | 細野 公男                                                         | 1   |
| 論文   | Discovery Regularities by Materials Databases<br>Ying CHEN, Shuichi IWATA, Yasunori KANETA and Pierre VILLARS |                                                               | 3   |
| 論文   | データマイニング技法を用いた診断アンケート支援システム<br>.....                                                                          | 田中 猛彦, 田中 康幸, 中川 優, 小倉 光博, 板倉 徹                               | 12  |
| 論文   | Frequency Distribution of the Number of Amino Acid Triplets in the<br>Non-Redundant Protein Database<br>..... | Joji M. OTAKI, Tomonori GOTOH and Haruhiko YAMAMOTO           | 25  |
| 論文   | 材料ファクトデータベースから得られるXML記述の知見ノート<br>藤田 充苗, 徐 一斌, 加治 芳行, 塚田 隆, 小野瀬 庄二, 益子 真一, 芦野 俊宏                               |                                                               | 39  |
| 論文   | 相互エントロピーを用いたアライメントの改良<br>.....                                                                                | 池 正人, 佐藤 圭子, 谷田貝 甲児, 大矢 雅則                                    | 49  |
| 寄稿   | 化学データベース .....                                                                                                | 田隅 三生, 朽津 耕三, 細矢 治夫, 早水 紀久子, 田辺 和俊, 森岡 義幸, 坂本 章, 佐藤 寿邦, 廣田 勇二 | 58  |
| 調査報告 | 用語は名詞でなければいけないか?—用語規格における用語の品詞的考察—<br>.....                                                                   | 太田 泰弘                                                         | 73  |
| 研究速報 | 「動き」を採り入れた電子マニュアルの製作とCGアニメーション画面のユーザビリティ評価 .....                                                              | 能城 正志, 野須 潔                                                   | 84  |
| 活動報告 | 第11回(2003年度)研究報告会雑感 .....                                                                                     | 平田 周                                                          | 91  |
|      | 第11回(2003年度)研究報告会抄録 .....                                                                                     | 国沢 隆                                                          | 93  |
|      | 平成15年度総会議事録 .....                                                                                             |                                                               | 99  |
| お知らせ | 情報知識学会主催・第8回SGML/XML研修フォーラム「XMLとデータベースを活用した新しいビジネスモデル」 .....                                                  |                                                               | 106 |
|      | 情報知識関連新刊図書一覧 .....                                                                                            | 平田 周                                                          | 108 |
|      | 第16回専門用語シンポジウム .....                                                                                          |                                                               | 109 |
|      | 投稿規定と執筆要領 .....                                                                                               |                                                               | 110 |



情報知識学会

## は し が き

情報知識学会主催の「SGML/XML 研修フォーラム」は、会員、参加者各位の支援をえて、今年その第8回を開催する運びとなった。各位もすでにご案内のとおり、SGML/XMLの応用分野は、今日ますますひろがりを見せている。そして、その多くがデータベースに関連した各種の事業・活動と密接なかかわりをもって展開されている。この点に鑑みて、今回の「SGML/XML 研修フォーラム」は、「XML」とデータベースを活用した新しいビジネスモデル」をテーマに掲げて企画した。また、この趣旨から、参加者に便利なよう、データベース総合展「DATABASE 2003 TOKYO」に併設の研究会として開催することにした。

今回のフォーラムでは、XMLとデータベースの関連に特に注目して、メタデータ、Webサービス、ビデオ配信サービス、電子出版、地理情報、財務情報といった幅広い応用技術、適用分野の動向を見わたすことができるように配慮し、また、すでに市場化されているXMLデータベースのためのソフトウェア製品数種について、その活用事例を紹介して、具体的な適用イメージが描けるようにしている。一方、特許や電子自治体等、電子政府に関する動きにも目配りして、全体として、データベースをXML化してゆくことにより、今後どのような新たなビジネス展開が可能であるかを検討するものとした。

ここに、本フォーラム開催にご協力賜った「DATABASE 2003 TOKYO」の主催者である(財)データベース振興センター、データベース協会をはじめとする関係各位に感謝するとともに、これらの講演を踏まえて、参加者・講演者間での活発な意見が交わされて議論が深まることにより、「ITで創る21世紀」にいささかなりとも貢献しうよう、本フォーラム企画者一同として、大いに期待するところである。

2003年10月

SGML/XML 研修フォーラム実行委員長

根岸正光

プログラム

- <<第1日>> 2003年10月30日(木)
- 9:10～ 受付開始
- 9:40～ 9:45 開会挨拶 実行委員長：根岸正光／国立情報学研究所教授
- 【 午前の部 】
- 9:45～10:35 ライフサイエンス分野における Web サービス  
浦本直彦／国立情報学研究所客員助教授・IBM東京基礎研究所
- 10:35～11:25 G-XML の活用事例 奥住啓介／(財)データベース振興センター事務局長
- 11:25～13:00 {昼食 95分}
- 【 午後の部 】
- 13:00～13:50 XML とメタデータ - メタデータの基本概念 杉本重雄／筑波大学教授
- 13:50～14:40 オーディオビジュアルコンテンツのためのメタデータ国際標準  
亀山渉／早稲田大学教授
- 14:40～14:50 {休憩 10分}
- 14:50～15:25 電子カタログにおけるネイティブ XML データベース『Sonic XIS』の活用  
武末徹也／プログレスソフトウェア(株)
- 15:25～16:00 EsTerra による汎用システムとの連携 (活用事例)  
中村真二／メディアフュージョン(株)
- 16:00～16:35 Oracle XML DB の活用事例 作田淳子・有賀洋平／日本オラクル(株)
- <<第2日>> 2003年10月31日(金)
- 【 午前の部 】
- 9:45～10:35 土木施工における XML の活用 古屋弘／(株)大林組 土木技術本部
- 10:35～11:25 財務情報の XML 化事例 水谷学／ピー・シー・イー(株)常務取締役
- 11:25～13:00 {昼食 95分}
- 【 午後の部 】
- 13:00～13:50 特許事務処理システムの XML 化 安久司郎／特許庁情報システム課課長補佐
- 13:50～14:40 これからの電子自治体システムと人材 溝江言彦／福岡県情報政策課情報企画監
- 14:40～14:50 {休憩 10分}
- 14:50～15:35 電子書籍制作における XML の活用 前川真二／大日本印刷(株)C&I 事業部
- 15:35～16:20 XML パブリッシングのための出版標準フォーマットへの取り組み  
田原恭二／凸版印刷(株)E ビジネス推進本部係長
- 16:20～16:25 閉会挨拶

SGML/XML 研修フォーラム実行委員会

- |      |      |                            |
|------|------|----------------------------|
| 委員長  | 根岸正光 | 国立情報学研究所教授                 |
| 副委員長 | 長塚 隆 | (株)ジー・サーチ主席部長、日本データベース協会会長 |
| 副委員長 | 石塚英弘 | 筑波大学教授                     |
| 委員   | 細野公男 | 慶應義塾大学教授                   |
| 委員   | 岩淵幸雄 | 日本創造学会評議員                  |
| 委員   | 菊田昌弘 | (株)シナジー・インキュベート代表取締役       |
| 委員   | 川上能徳 | 大日本印刷(株)C&I 事業部            |
| 委員   | 小川恵司 | 凸版印刷(株)E ビジネス推進本部          |

## ライフサイエンス分野における Web サービス

国立情報学研究所および日本IBM東京基礎研究所

浦本直彦 (uramoto@jp.ibm.com)

### 1. はじめに

次世代の Web の基盤として登場した Web サービスは、現在、関連仕様の標準化やビジネスアプリケーションを中心とした普及が進んでいる。SCM、ERP、CRM といったビジネス分野における Web サービスの利点は、コンポーネント化したサービスインターフェイスを、簡単に結合し呼び出すことができることである。一方で、大規模なデータや知識を統合するための枠組みとしても Web サービスは使い勝手のよい基盤であり、ライフサイエンスなど、大量の情報が蓄積され、情報の統合が重要視されている分野においても、有望な手法として注目を集めている。本論文では、主に、情報統合の立場から、ライフサイエンス分野における Web サービスへの期待や適用例について述べる。

### 2. 情報統合基盤としての Web サービス

Web サービスは、Web 環境におけるコンポーネントおよびアプリケーションの統合のための基盤技術である。広義には、HTML、Java Script、JSP、サーブレットなどの従来技術に基づくものも含むが、特に、Simple Object Accessing Protocol (SOAP)、Web Services Description Language (WSDL)、Universal Description, Discovery and Integration of Web Services (UDDI) に代表され

る、XML に基づく標準技術を用いて構築されたものを指す。Web サービスでは、ネットワーク上のコンポーネントを、標準化された入出力を持つ「サービス」として定義し、それらを組みあわせることで、より複雑なサービスを構築することができる。

Web サービスの利点については、すでに様々な場所で言及されている。ここでは、以下の2点を挙げる。

- 1) サービスインターフェイスの統合。WSDL を用いて入出力の型を定義することで、ユーザは遠隔手続き呼び出し(RPC)や、XML メッセージを使って、簡単にサービスを利用することができる。また、これらのサービスを合成することで、ビジネスアプリにおけるワークフローを構築することが可能となる。
- 2) 情報・データの統合。分散された環境にある異種のデータを結合し、ひとつの仮想的なデータベースとして表現したり、異なる情報を関連つけたりすることができる。ユーザは、複数のデータベースに、それぞれのやり方でアクセスするのではなく、横断的に情報にアクセスし、統合された情報を入手することができる。



従来の Web サービスに関する議論では、主に(1)、つまり様々な環境にあるビジネスプロセスを容易に記述し、結合することの利点が強調されてきた。しかし、サービスの背後にある情報や知識を統合するための枠組みを与えることができるという(2)の利点も Web サービスの特徴のひとつであると考えられる。

そもそも、情報統合(データ統合)は、古くから人工知能やデータベースの分野で研究が進められてきた。Ullman は、情報統合システムの共通アーキテクチャとして、ソース(source)、ラップ(wrapper)、メディエータ(mediator)によるモデルを紹介している[1]。これに、関連する構成要素を加えたモデルを図1に示す。

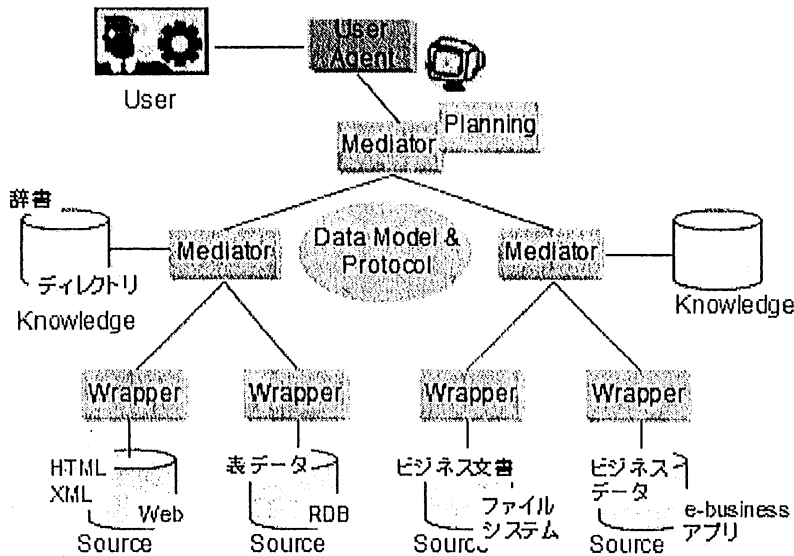


図1：古典的な情報統合のモデル

ソースは、それぞれ固有の(ローカルな)スキーマで記述された情報源を示す。XML 文書における Document Type Definition (DTD)や XML スキーマ、関係データベースにおけるデータベース(表)スキーマ、アプリケーションが用いるデータの型などがローカルスキーマの例である。HTML 文書のように意味的なローカルスキーマが定義されていない場合もある。

ラップは、個々のソースを定義するローカルスキーマを共通のグローバルスキーマへ「翻訳」するために用いられる。グローバルスキーマは、各ソースの違いを吸収し、ソースの利用者に対して、共通のビューを提供するものである。Web サービスでは、サービス記述である WSDL 文書がグローバルスキーマに相当する。

メディエータは、グローバルスキーマへ翻訳されたソースを統合し、利用者からの要求に必要な情報を提供する。メディエータは、別のメディエータと連携することができる。また、メディエータを統合し複雑な処理を行うための戦略を受け持つプランニングエンジンが別に用意されることがある。

ラッパやメディエータが参照するグローバルスキーマに関する情報を供給するのが辞書、オントロジ、ディレクトリからなる知識源である。辞書はグローバルスキーマで用いられる単語あるいは概念の定義、オントロジは概念間の関係や制約(概念定義がオントロジに含まれる場合もある)、ディレクトリは、メディエータにアクセスするための情報(例えば所在情報やアクセス形式)を提供するために用いられる。単にサービスの型を統一するのではなく、サービスが取り扱う情報や知識をうまく統合しようと思うと、このような構成要素が基盤として必要になってくる。

### 3. ライフサイエンス分野からみた Web サービスの利点

ここで、ライフサイエンス分野が Web サービスに注目する理由について考えてみよう。

ライフサイエンス分野においては、初期の段階から文献情報や遺伝子、タンパク質に関する配列情報や数値情報が精力的に収集され、公開されてきた。例えば、米国生物工学情報センター (NCBI) では、ヒトゲノム計画の成果である遺伝子配列情報や、1960年代から1200万件に及ぶ医用文献抄録データベース(PubMed、<http://www.ncbi.nlm.nih.gov/entrez/>)など、様々な形式の膨大な情報やツールを無料で公開している。NCBIを含め、多くの機関が有用な情報を整理して公共データベースとして公開しているのは、この分野の大きな特徴である。

このような大量でかつ様々な形式のデータを関連付けすることが非常に重要である。言い換えると、情報統合が本質的に必要不可欠であること[2][3][4][5]が、Web サービスを用いることの大きな要求となっている。例えば、ある遺伝子配列(A、T、G、Cの4種の塩基の並び)があるとしよう。この配列に対して、類似する遺伝子配列を検索する、あるいは、この遺伝子がコードしているタンパク質やそれに構造的あるいは機能的に類似するものを探す、そのタンパク質について言及されている文献を検索する、等々、ユーザは様々な情報源を渡り歩きながら、必要な情報を探し出す。データは文献、特許などのテキストデータだけでなく、配列データ、数値データ、画像データなど多岐にわたる。

さらにこれらの情報は、分散された環境に配置されていることが多い。よって情報源をいかに検索し、関連付けるかが大きな意味を持つ。個々のデータベースは、それぞれの研究機関で管理されていることが多いし、頻繁に更新されるので、人手による関連付けが非常に難しいからである。これらのサービスは従来 HTML ベースで提供されていたが、情報の関連付けを行うには、XML を使ってデータを構造化し、Web サービスを用いてサービスを統合するのは自然な流れである。データの XML 化についても、前述の PubMed では、全件の抄録データがすでに XML 形式で閲覧可能である。また、京都大学化学研究所バイオインフォマティクスセンターが提供する DBGET データベース([www.genome.ad.jp/dbget/dbget.links.html](http://www.genome.ad.jp/dbget/dbget.links.html))は、複数のデータソースに対して、DNA、タンパク質、リガンドなどの仮想的なデータベースを定義し、単一の検索で、複数のデータベースを横断的に検索するための機能を提供している。個々のデータベースに対してはラッパに相当する機能を用いて、複数データベースへの透過的なアクセスを実現している。

最後に、ライフサイエンス分野では、配列間の相同性(類似性に相当するもの)の計算や、タンパク質の構造および機能予測に大きな計算機パワーを必要とする点が挙げられる。これらの処理を

単一の計算機で実行するのは難しい。そこで、処理の並列化やグリッド化が必要となってくる。グリッド環境におけるミドルウェアとして、現在、Globus が脚光を浴びているが、最新のバージョンである Globus Toolkit 3.0 (<http://www.globus.org/toolkit/>)では、グリッド上のプログラムを、Web サービスとして定義し、ユーザが実際の計算機リソースがどのように使われるかを気にせずに用いることを可能にする。大阪大学下條らによる BioGrid グループ([www.biogrid.jp/](http://www.biogrid.jp/))では、ライフサイエンス分野でよく使われるプログラムをグリッド上の Web サービスとして定義し、巨大な BioGrid 上で動かすことを提案している[5]。

#### 4. Web サービスからみたライフサイエンスの利点

それでは、逆に Web サービスからみたこの分野の利点は何だろうか？

前述したように、情報統合では、セマンティクスを持ったデータを考慮することが必要である。これは、データのセマンティクスをできるだけ考慮することなくサービスの統合を実現しようとする Web サービスの理念とは相容れない場合がある。

しかし、前述の DBGET の仮想データベースでも示されているように、ライフサイエンス分野においては、遺伝子、タンパク質、アミノ酸、疾病、生物種、など、コミュニティの中で共通に使われる主要な概念が比較的少数である。うまく、共通のセマンティクスとデータモデルを設計することができれば、有効な統合システムを構築することができる(図2)。もちろん、同じタンパク質が異なる名前と呼ばれるなど、オントロジ(用語体系や同義語など)を整備していく必要があるが、PubMed における MeSH ターム(数万語レベルの概念体系)など、すでに利用可能な知識もいくつかある。

また、ビジネスアプリケーションにおいては、Web サービスのセキュリティが大きな問題となるが、ライフサイエンス分野においては、多くの公開データベースが存在するため、比較的統合システムを構築しやすいことも Web サービスが根付く土壌となりえる(もちろん、社内のデータと統合するような場合は、セキュリティの確保が重要であろう)。

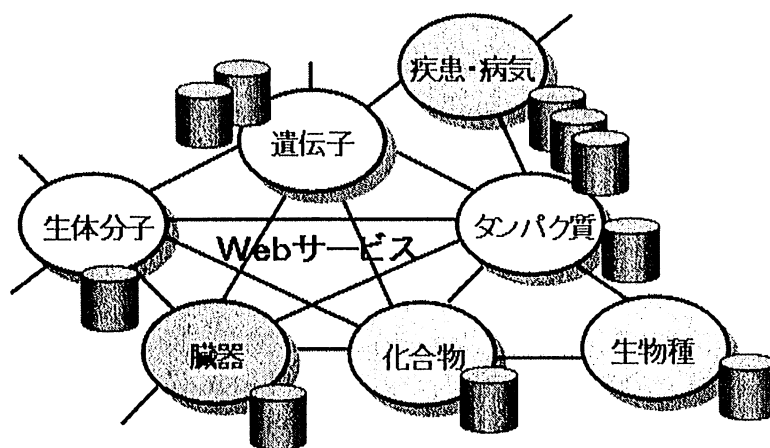


図2 ライフサイエンスにおける概念とサービス

## 5. 事例

ここまで、ライフサイエンス分野における Web サービスの重要性と適用のための枠組みについて述べた。本節では、実際の適用例をいくつかあげる。

人間の遺伝子配列を同定する国際ヒトゲノムプロジェクトのヨーロッパの拠点でもある European Bioinformatics Institute (EBI、<http://www.ebi.ac.uk/>)では、遺伝子配列データを表現するためのXML構文である XEMBL を提案している。配列データは、CGI や Web サービスによって外部に公開されている。図3に XEMBL サービス記述(WSDL 文書)を示す。

ライフサイエンス分野でのソフトウェアの相互運用性を確保するために必要な技術を統一するために発足したコンソーシアムが、Interoperable Informatics Infrastructure Consortium (I3C、[www.i3c.org](http://www.i3c.org)) である。現在、UDDI レジストリや Semantic Web[7][8]の利用などが検討されている。

日本遺伝学研究所生命情報・DDBJ 研究センターでは、ライフサイエンス分野でよく使われるツールを Web サービス化し公開している(<http://www.xml.nig.ac.jp/index.html>)。現在登録されているサービスを以下に示す。

- Blast
- ClustalW
- DDBJ
- ExClustalW
- Fasta
- GetEntry
- Gtop
- SRS
- TxSearch

これらのサービスの多くは、HTML ベースで公開されているが、Web サービスとして提供されることで、コンポーネントの一部として、より大きなシステム中に組み込むことができる。

IBMの先進ソフトウェア提供サイトである Alphaworks からは、Web Services for Life Sciences というツールが公開されている(<http://www.alphaworks.ibm.com/tech/ws4LS>)。提供されているのは、PubMed、GenBank、Blast、Phylogenic Tree、Clustal W に対する Web サービスである。

## 6. おわりに

本論文では、ライフサイエンス分野における Web サービスの重要性と適用事例について述べた。この分野における Web サービスの普及は始まったばかりであるが、これからの発展を期待したい。

```

<definitions name="XEMBL" targetNamespace="http://www.ebi.ac.uk/XEMBL"
              xmlns:tns="http://www.ebi.ac.uk/XEMBL" .... 一部略
<!-- 入力の型定義 -->
<message name="getNucSeqRequest" xmlns:tns="http://www.ebi.ac.uk/XEMBL">
  <part name="format" type="xsd:string">
    <documentation>出力形式を指定するパラメータ </documentation></part>
  <part name="ids" type="xsd:string">
    <documentation>配列の識別子(アセッション番号) </documentation></part>
</message>
<!-- 出力の型定義 -->
<message name="getNucSeqResponse">
  <part name="result" type="xsd:string">
    <documentation>結果のXML メッセージ </documentation></part>
</message>
<!-- サービスオペレーションの定義 -->
<portType name="XEMBLPortType">
  <operation name="getNucSeq">
    <input message="tns:getNucSeqRequest" name="getNucSeqRequest"/>
    <output message="tns:getNucSeqResponse" name="getNucSeqResponse" />
  </operation>
</portType>
<!-- 結合情報の定義 -->
<binding name="XEMBLServiceBinding" type="tns:XEMBLPortType">
  <soap:binding style="rpc" transport="http://schemas.xmlsoap.org/soap/http" />
  <operation name="getNucSeq">
    <soap:operation soapAction="http://www.ebi.ac.uk/XEMBL#getNucSeq" />
  <input>
    <soap:body use="encoded" namespace="http://www.ebi.ac.uk/XEMBL"
              encodingStyle="http://schemas.xmlsoap.org/soap/encoding" /></input>
  <output>
    <soap:body use="encoded" namespace="http://www.ebi.ac.uk/XEMBL"
              encodingStyle="http://schemas.xmlsoap.org/soap/encoding" /></output>
  </operation>
</binding>
<!-- サービスポイントの定義 -->
<service name="XEMBLService">
  <port name="XEMBLPort" binding="tns:XEMBLServiceBinding">
    <soap:address location="http://www.ebi.ac.uk:80/cgi-bin/xembl/XEMBL-SOAP.pl" />
  </port>
</service>
</definitions>

```

図 3: XEMBL のサービス記述 (WSDL 文書)

## 参考文献

- [1] Ullman, J., "Information integration using logical views", Theoretical Computer Science, pp.189-210, Vol.239, No.2, 2000.
- [2] Stein, L., INTEGRATING BIOLOGICAL DATABASES, Nature Reviews Genetics, pp. 337-345, Vol.4, May 2003.
- [3] Davidson, S. B., et al, K2Kleisli and GUS: Experiments in Integrated Access to Genomic Data Sources, pp. 512-531, IBM Systems Journal, 40(2), 2001.
- [4] Wheeler, D., et al, Database resources of the National Center for Biotechnology Information, Nucleic Acids Research, Vol.29, No.1, 2001.
- [5] 上田ら, メタデータを用いたバイオ情報データベース連携検索手法の提案, 科学技術フォーラム(FIT), 2003.
- [6] 浦本, Web における情報統合—セマンティック Web と Web サービス—, 情報処理学会誌, Vol. 44, No. 7, 2003.
- [7] 浦本, "Semantic Web - 機械のための Web -". 人工知能学会誌 16 卷 3 号, 2001.
- [8] Semantic Web, World Wide Web Consortium, <http://www.w3.org/2001/sw/>

## G-XML の活用事例 －G-XML を用いた電子地質図高度利用化の研究開発プロジェクト－

財団法人データベース振興センター

奥住 啓介

### 1. はじめに

G-XML は、インターネット上の情報流通を促進する有効かつ新たな方法として、利用・普及が著しい XML を GIS (Geographic Information Systems : 地理情報システム) 分野に応用し、電子地図自身やその構成要素、GIS コンテンツを記述することを可能にしたプロトコルである。これにより、従来、異なった GIS エンジン間で共有することが困難であった地図情報や位置関連情報をインターネット上で容易に記述、通信、交換することを狙いとしている。

従来、G-XML では主として地上系の、市街地図、地形図、道路地図等とその上のコンテンツを対象として捉えられてきたが、本プロジェクトにおいては、国際的にも関心が高まりつつある地質図へ G-XML を適用し、デジタル地質図及びその関連データであるボーリング情報等の提供、流通、高度利用の促進を図ることにより、知的基盤としての地質情報の高度化を図ることを目的としたものである。

### 2. G-XML

#### 2.1 GIS の課題

GIS は、「地理的な位置を手がかりに位置に関する情報を持った電子データまたは地理情報を総合的に管理・加工し、視覚的に表示し、高度な分析や迅速な判断を可能にするシステム」であり、自治体を中心に、施設管理、防災・災害対策、財産管理、固定資産管理、事業計画シミュレーション等の分野で利用されている。しかしながら、既存 GIS は、他社製システム間での地図データの相互利用ができないことやシステムを入れ替えるとデータも変換し直す必要があり、情報の相互利用や流通のニーズへの対応が困難なことや、技術革新に伴うシステムのリリース時の経費増大化を招いている。

#### 2.2 G-XML とは

このような課題を解決するため、異なる電子地図・GIS パッケージ間での GIS コンテンツの相互流通を可能にするため、財団法人データベース振興センターは経済産業省及び情報処理振興事業協会の協力のもと平成 11 年度から XML を GIS 分野に特化したプロトコル G-XML を開発してきた。この G-XML は平成 13 年 8 月には G-XML 2.0 に基づく J I S 規格 (J I S X 7199 地理情報－地理空間データ交換用 XML 符号化法) に制定されている。この J I S 化により地方公共団体を中心に G-XML の導入が加速され、岐阜県、三重県、大阪府豊中市、北海道石狩市などでシステム構築・運用中である。

### 3. 地質図の位置づけ

地質図は、地表下の地盤や構造、成り立ち等を明らかにした知的基盤として、資源開発、防災、都市開発、地域開発、土木建設、土壌環境対策等を講ずるため、関係自治体やデベロッパーをはじめとする産業界によって幅広く活用されている。従来、アナログ図で提供、利用されてきたが、近年の情報化の進展、インターネットの普及や電子政府化の動きの中で、WEB を通じて、デジタル地質図を、それも加工編集が可能なベクトルデータの形態で、提供することが求められている。このことは、政府の定めた GIS アクションプラン 2002・2005 (平成 14 年 2 月閣議決定) や、日本工

業標準調査会知的基盤整備特別委員会の策定した知的基盤整備計画（平成14年7月）においても明確にされている。また、地質図は、単独で用いられるよりは、ボーリングデータをはじめとする他の地盤関連情報や地形情報、資源情報、気象情報などとともに利用される場合が少なくない。

特に事業者においては、独立行政法人産業技術総合研究所（以下 産総研）や地方自治体から提供される地質図、ボーリングデータなどの公的な情報に加えて、事業者独自に有する情報を統合的に利用することによりその付加価値が高められる。こうした統合利用は、従来、異なった GIS エンジン間では、ファイル構成の相違等により困難であったが、G-XML を活用することにより各種の位置情報関連の共通基盤としての整備に資することが期待されている。

#### 4. 研究開発の概要

##### 4.1 研究開発の目標

本研究開発は以上の背景を踏まえ、産総研が整備する一般的な4種類の縮尺（100万分の1、20万分の1、5万分の1、2.5万分の1）の電子地質図を G-XML プロトコルに対応させる手法を検討・開発するとともに、地質図の世界で広く利用されている DLG (Digital Line Graph) 仕様から G-XML 仕様に変換/逆変換するソフトウェアを開発すること、地質情報の重要な要素であるボーリングデータを G-XML 化し、2.5 万分の1地質図とそれに対応する地域のボーリングデータを組み合わせ合わせた利用を可能とする地質情報提供システムの研究開発及び構築を行い、その有効性を実証したものである。G-XML についての機能拡張は、地質図に対応するカバレッジ（被覆）プロトコルの検討・開発、ボーリングデータ対応のオブザベーション（観測値）モデルの導入を行った。概要をまとめると以下ようになる。

- ①産総研の保有する地質図の利用促進
  - ・地質図、ボーリングデータ、断面図
- ②地質図のインターネットによる提供、二次利用可能な形式での提供 (G-XML)
  - ・DLG ↔ G-XML 変換/逆変換
  - ・Web での閲覧
  - ・データのダウンロード 商用 GIS エンジンでの利用
- ③G-XML の適用分野の拡大 (G-XML2.0 の機能拡張)
  - ・カバレッジ（被覆）用プロトコルの開発
  - ・オブザベーション（観測値）モデルの導入
- ④オープンソースソフトウェア指向による利用拡大
  - ・開発成果（ソフトウェア）の公開

##### 4.2 研究開発の詳細

###### 4.2.1 地質図データの G-XML 化

###### (1) 地質図データの整備

本研究開発（以下本プロジェクト）では、産総研で発行する100万分の1、20万分の1、5万分の1、2.5万分の1の地質図を数値化したDLGデータを使用した。100万分の1の地質図は、刊行済みのDLGデータの修正版を使用し、20万分の1、5万分の1地質図は、未刊行ではあるが刊行準備が進められているDLGデータを、2.5万分の1地質図については「筑波研究学園都市およびその周辺地域の環境地質図」をDLGデータ化したものを使用した。本プロジェクトに使用した100万分の1及び2.5万分の1の数値化データについては、一般の利用に供するため、G-XML化データを産総研から公開することを予定している。

###### (2) 地質図の分析

上記各縮尺の地質図について、以下の分析を行った。



①地質図の記載内容の調査

100 万分の 1、20 万分の 1、5 万分の 1、2.5 万分の 1 地質図の、記載内容及びそのデータ仕様(DLG フォーマット)の調査。

②ユーザニーズの検討 (サービス定義)

産総研で把握するユーザニーズ、電子地質図に関連する標準化状況、国土交通省の土木・地質調査業電子納品における土木地質図の基準等から想定されるユーザとその利用例を検討。

③サービス実現のための地質情報データモデル (応用スキーマ) の検討

デジタル化された地質図を中心に、地質情報のデータモデル (応用スキーマ) を検討。地質図情報の描画方式については、現状技術で実現可能な範囲とし、一般的な商用 GIS ソフトウェアで検証。

④データモデルに基づくタグの決定

G-XML プロトコル 2.0 でサポートしていない新規タグおよび地質図固有タグの仕様作成。

⑤メタデータの作成

地理情報標準をベースにした地質図のメタデータ項目の検討。

(3) G-XML プロトコル対応手法の開発 (モデル、スキーマ、DTD 等)

上記 (2) で実施した地質図の分析を踏まえ、カバレッジプロトコルを開発し、モデル、スキーマ、DTD 等のプロトタイプを作成した。

4. 2. 2 地質図フォーマット変換ソフトウェアの開発

DLG フォーマットの地質図と、G-XML の相互変換ツールを開発するため、DLG→G-XML 変換ソフトウェア (DLG ファイルの幾何形状およびトポロジー情報の変換プログラム)、G-XML→DLG 変換ソフトウェアの開発を行った。

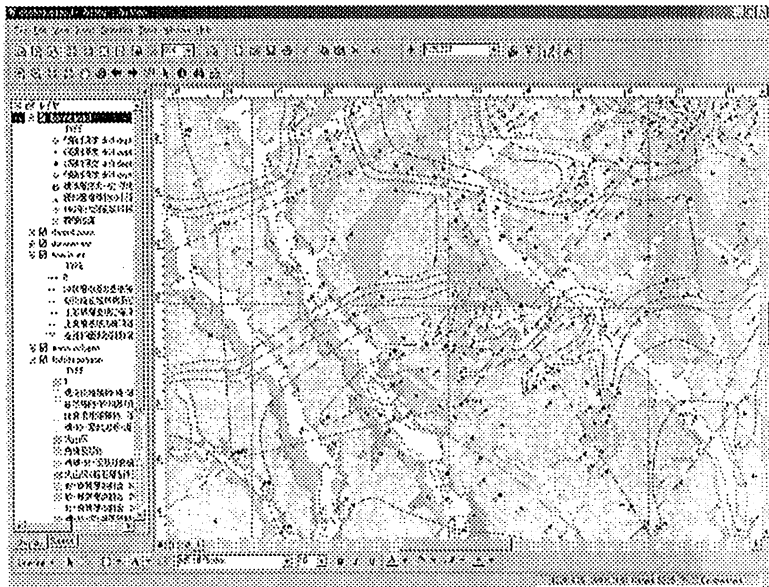


図1 1/2.5万地質図 ボーリングデータ 断面線

#### 4. 2. 3 ボーリングデータの G-XML 化

産総研が保有するボーリングデータについて、一般に提供するために必要なデータ整備を行い、また第三者がこれらデータの加工・編集を容易に行いうる環境を整備するため、以下の研究開発を行った。

##### (1) ボーリングデータの整備

産総研の刊行した「筑波研究学園都市及び周辺地域の環境地質図」の作成に利用した約5,000本のボーリングデータのうち、公開可能なもの約1,000本のデータを数値化した。このG-XML化したボーリングデータは、個人情報等を除き公開可能なデータにして、産総研より公開される予定である。

##### (2) ボーリングデータ標準仕様フォーマットへの変換

提供するボーリングデータの保持すべき情報の内容を分析・整理し、ボーリングデータの標準仕様を検討し、これに沿って G-XML 仕様のボーリングデータを流通させるためのプロトコルを開発した。

本研究開発に関する詳細項目は以下のとおりである。

##### ①ボーリングデータの記載内容の調査

以下のボーリングデータおよびボーリング柱状図作成要領について、記載する情報の項目、内容等を分析・整理。

- ・産総研が提供するサンプルボーリングデータ
- ・ボーリング柱状図作成要領（財団法人日本建設情報総合センター）
- ・土質調査要領（日本道路公団）
- ・地質調査標準示方書（東日本旅客鉄道株式会社）

##### ②ユーザニーズの検討（サービス定義）

整理したボーリングデータの記載項目・記載内容について、ユーザニーズの視点からの必要性を分析・整理し、各項目・内容の用途、利用場面、利用方法の検討。また、地質図、断面図とボーリングデータの組み合わせ利用を想定した利用範囲を設定。

##### ③サービス実現のためのボーリングデータモデル（応用スキーマ）の検討

上記②で設定したボーリングデータの利用範囲に対し、利用対象となる情報の項目・内容、データ間の関連性等を精査し UML（統一モデリング言語）により記述・検討。

##### ④データモデルに基づくタグの決定と G-XML への変換方法の検討

G-XML プロトコル 2.0 でサポートしていない新規タグおよびボーリングデータ固有のタグ仕様作成。

##### ⑤メタデータの作成

「地理情報標準第 2 版 IXメタデータ」に基づき、コアメタデータとして挙げられている項目のほか、ボーリングデータを説明するのに必要な項目を整理し、ボーリングデータのメタデータ項目を設定。

##### (3) G-XML での三次元データ処理プロトコルの研究開発

G-XML でのボーリングデータの三次元データ処理プロトコル研究開発のための、開発要件の検討・抽出を行った。

なお、G-XML での三次元処理のための要件の検討・抽出にあたっては、日本及び諸外国での関連 XML 技術等の適用事例などについて調査分析を行い、その結果を基に適切な要件抽出を行った。

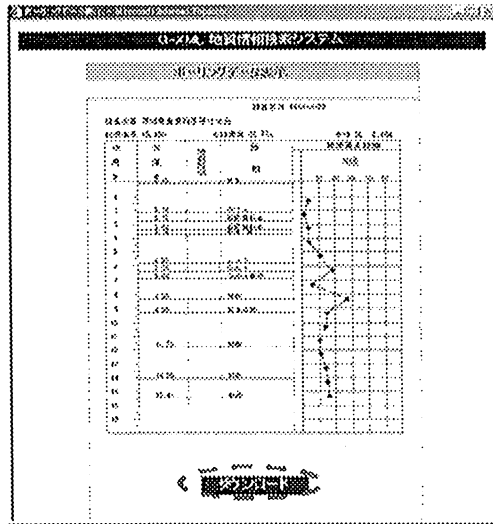


図2 ボーリングデータ表示画面

#### 4. 2. 4 地質情報提供システムの構築

上記で作成された各縮尺の地質図及びボーリングデータの情報提供システムの開発・構築を行った。本研究開発に関する詳細項目は以下のとおりである。

①100 万分の 1，20 万分の 1，5 万分の 1，2.5 万分の 1 地質図の G-XML プロトコルでの提供システム（二次元データ）

（財）データベース振興センターが一般公開している e-G View をカスタマイズし、G-XML 化された 100 万分の 1，20 万分の 1，5 万分の 1，2.5 万分の 1 地質図をブラウザで表示できるシステムを開発・構築する。開発する機能は以下のものとした。

- ・データ蓄積機能
- ・データ検索機能
- ・データ提供機能（Web サーバによる）

②2.5 万分の 1 地質図とボーリングデータ（メタデータ及びファクトデータ）の組み合わせ提供システム（二次元データ）

ボーリングデータの位置情報やカテゴリ等のメタデータを含んだ 2.5 万分の 1 地質図の G-XML ドキュメントを表示し、表示された 2.5 万分の 1 地質図内の G-XML 化されたボーリング情報を検索・参照・追加するため上記基本システムの拡張機能の開発を行った。

- ・データ蓄積機能
- ・データ検索機能
- ・データ提供機能（Web サーバによる）

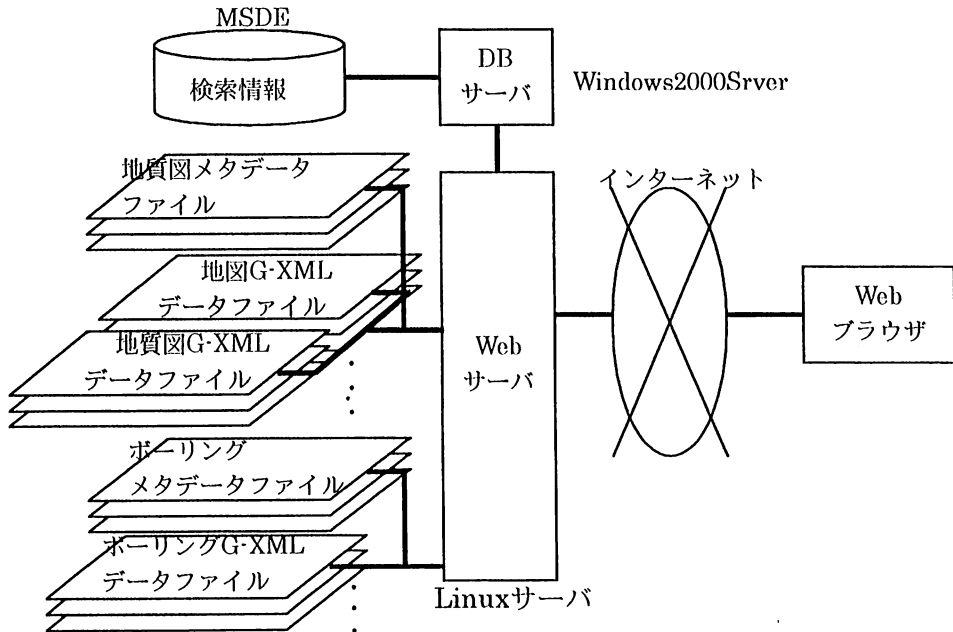


図3 提供システム構成図

## 5. 今後の課題

### 5.1 地質図系データの流通促進のための標準化

100万分の1地質図は、統一した凡例区分と色情報がなされており、ユーザにとっては理解しやすいものである。しかし、20万、5万、2.5万分の1地質図は、標準化がなされていないことから地質図データの標準化は必要不可欠である。また、同様にボーリングデータも作成したが、ボーリングデータのみでの情報では全体が把握しにくく、地質平面図・断面図の各データとの二次元・三次元的な相互利用を図っていく必要があり、それに伴うボーリング柱状図作成要領の統一、地層区分名称の統一が望まれる。

また、さらに情報公開が進展するにつれて、これらのデータ利用者が第三者であることが多いと考えられることから、提供データにはメタデータの付与が不可欠であり、このための総合的な標準化活動が必要である。

### 5.2 利用可能な地質図データの整備

地質図G-XMLデータをデータ提供サービスとしてダウンロード後に使用する場合、システムのレンダリングサービスにおいては地質図データと重ね合わせた形で表現されるが、ダウンロードされた地質図には地質図データは含まれない。このため、利用者が地質図G-XMLデータと同じ範囲をカバーする地質図データを用意する必要があり、ユーザが容易に入手でき、また幅広い利用が可能な地質図データの整備が求められる。

### 5.3 情報提供システムの高速度化

本システムではG-XML化データをSVGに変換して配信し、クライアントPCのプラグインソフトで描画を行っている。既にSVGに変換されている場合は、そのSVGデータファイルを配信する。

従って、初めて描画される G-XML のデータは変換処理が発生するため、表示完了までに非常に時間がかかる。そのため、システム起動時に全ての G-XML ファイルを SVG に変換することで、配信時の変換時間削減が考えられるが、データ量が非常に多い場合にシステム起動時の負荷が大きくなり、運用負荷の増大に繋がり、これらへの対処を含んだ運用ルール等が今後の課題となる。

## 6. 今後の取り組み

### 6.1 本プロジェクトの今後の計画

今年度は、産業技術総合研究所から電子的な提供が始まりつつある地質図及びボーリングデータの G-XML による提供を可能とする技術レベル、即ち、二次元表記された地質図、ボーリング情報を提供するための技術開発を行った。しかし、地質の世界は、本来三次元の世界であり、15年度及び16年度においては、こうした三次元情報（場合によっては時間軸もあわせた四次元情報）の取扱いを行うためのプロトコル開発を含む技術開発を実施する予定である。三次元地質モデルについては、我が国のみならず世界的にも様々なモデルの提案がなされ、研究が進められている段階にある。従って、三次元モデルの取扱いを可能にするということは、14年度事業がこれまで紙ベースで提供されてきた地質図という知的基盤を電子化、G-XML 化を通して高度化するというものであったのに対して、三次元地質モデルの構築という地質図自体の高度化を図るためのツールを提供するものとも位置づけられる。即ち、本プロジェクトにおいて G-XML による三次元情報の処理を可能にすることが進められることと平行して、三次元地質モデル自体の研究開発が進められ、双方の進展が相まって、知的基盤としての地質図の高度化に資することが期待される。

### 6.2 G-XML の動向

2001年8月に JIS 規格「JIS X 7199 地理情報—地理空間データ交換用 XML 符号化法」が制定されたが、その後、国際的な民間団体 OGC との協力により、OGC-GML2.0 と G-XML2.0 を統合化した、次の版といえる国際統合版が ISO/TC211 (地理情報) に提案され、新作業項目 (ISO19136) として、2002年5月に正式承認された。この国際統合版は、先ず 2003年2月に OGC において GML3.0 として実装仕様となったが、現在、この OGC-GML3.0 を基本にして ISO においても GML を国際規格とするための作業が進んで来ている。

一方 G-XML3.0 は、より一層実用的な適用範囲を対象に、OGC-GML3.0 の応用層として開発が進められ、2003年3月に完成した。

今後、OGC-GML から変更して、ISO/TC211 の国際規格群と整合のとれた ISO-GML をコアとして引用しながら、G-XML プロトコル仕様を更に拡張する予定である。国際統合版である GML 又は ISO-GML には、G-XML (JIS X 7199) のすべての機能が反映されているわけではないので、これらを反映すべく、新 G-XML を ISO-GML の応用層として再構築することにした。新 G-XML の開発に当たっては、現 G-XML (JIS X 7199) からの互換性を最大限に考慮して、当該標準化の作業を進めて行く必要があると考える。

応用層に関しては、LBS (位置情報サービス) やコンテンツ流通サービスのための諸機能を集約しており、既存 JIS との互換性確保と国際規格群との整合を両立し得る構成となっている。従って、LBS やコンテンツ流通サービスの面においては、日本が技術的にも、多様なサービスを実現できるという点でも、国際的に先行している分野であり、G-XML が可能とするデータの相互交換・流通の促進により、ネットワークワイドの分散協調型サービスや消費者参加型サービスの面で、日本企業の一層の競争力強化を図ることが可能になると期待できる。さらに、国際規格群とも整合・即応する形で、サービスレベルで G-XML を国際社会に対して提案して行くことの影響力も大きいと考える。

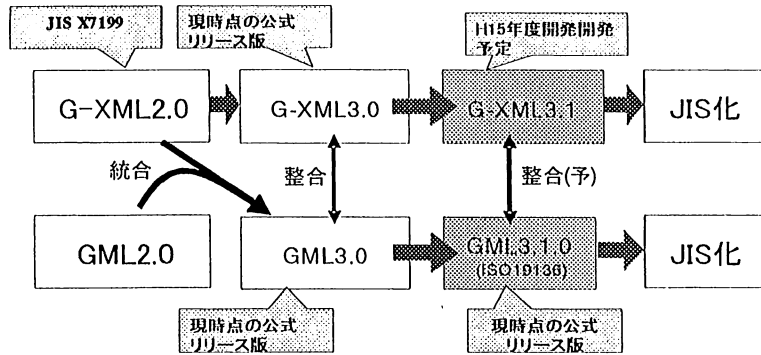


図4 G-XMLの標準化計画

### 7. むすび

本プロジェクトは、地質図、ボーリングデータ等が第三者に利用されることを主眼として、実用目的に研究開発を行ったものであり、G-XML が情報の流通・利用促進に役立つ事を実証したものと考えている。今後、このプロジェクトを契機にさらなる G-XML の応用分野の出現を期待している。

なお、本稿は新エネルギー・産業技術総合開発機構（NEDO）から財団法人データベース振興センターへの委託事業（正式名称：知的基盤創生・利用技術研究開発 G-XML 技術を用いた電子地質図の高度利用化の研究開発）成果を使用した。本プロジェクトを支援して頂いた NEDO、及びプロジェクトに参画した独立行政法人産業技術総合研究所、アジア航測（株）、国際航業（株）、（株）パスコ、（株）NTTデータの関係者に感謝する。

## XML とメタデータ — メタデータの基本概念

筑波大学・図書館情報学系  
 知的コミュニティ基盤研究センター  
 杉本重雄

インターネット上での情報資源の組織化、発見、利用といったいろいろな場面でメタデータが利用されている。本稿では、ネットワーク上の情報資源の記述のためのメタデータに関する基礎概念といくつかのメタデータのモデルについて述べる。また、インターネット上の情報資源の発見のために提案されたメタデータ規則である Dublin Core に関し、いくつかの重要な基礎概念について考察する。

## 1. はじめに

セマンティックウェブやデジタルライブラリを中心としてメタデータが注目を集めている。メタデータはデータに関するデータと定義され、目録や索引、抄録、辞書・事典から、書評や地理情報まで、広義には様々なものが含まれる。ネットワーク上では、情報資源を探し、評価し、アクセス・利用するというすべての過程でメタデータを必要とする。一方、こうした作業を実現するために情報資源に関する記述(メタデータ記述)をしなければならない。こうした記述の目的も、資源の管理と提供、保存、知的財産権や利用環境の管理など様々である。そのため、目的に応じた様々なメタデータのモデルとメタデータ規則が提案されている。本稿では、情報資源に関する属性の記述を行うためのメタデータを対象としてその基本概念を考察する。こうしたメタデータ規則の構成の枠組みは下の要素からなっている。

- (1) 記述対象属性の定義と属性値として持つ値の統制語彙、値の表現形式に関する定義などメタデータを表現するために用いる語彙の定義。
- (2) 複合構造の定義や、省略可能性や繰り返し回数など、記述対象属性ごとに決められる制約など、メタデータの構造的定義。
- (3) 具体的なシステムごとに決められるメタデータの記述構文や入出力形式など実現上の定義。
- (4) 実際の対象情報資源に対して、どのように記述すべき内容を抽出し、メタデータを記述するかに関する指針。

具体的にシステムを実現するには(1)~(4)すべてが決まらねばならないが、上の要素には、(1)のように特定の応用領域やシステムに依存せずに決められるものもあれば、(3)や(4)のように強く依存するものもある。応用領域にまたがった利用を目的とする Dublin Core の場合には主として(1)のみを決めている。

インターネット上での流通のため、メタデータの表現には XML がよく利用されている。たとえば、Dublin Core では XML や RDF (Resource Description Framework)による推奨形式を決めている [1][2]。また、MARC をもとに開発された MODS (Metadata Object Description Schema)や、デジタル化資料のアーカイブを指向した METS (Metadata Encoding and Transmission Standard)では XML による記述形式を決めている [3][4]。RDF は特定のメタデータ規則に限定せずに WWW 上でのメタデータの共有・交換のために定義されたメタデータの記述形式である [5]。また、RDF Schema (あるいは RDF Vocabulary Description Language) を用いてメタデータ表現のための語彙として準備された属性や属性値の記述のための項目の定義を与えることができる。

以下、本稿では、デジタル情報資源の発見、管理、保存のためのメタデータについて紹介する。特に、Dublin Core に関して、その基本的データモデルについて述べる。<sup>1</sup>

## 2. メタデータの基礎概念

<sup>1</sup> メタデータを含むデジタルライブラリ (電子図書館) に関して、平成 15 年度大学図書館職員長期研修資料の拙著資料 ([http://www.tulips.tsukuba.ac.jp/pub/choken/youkou/3\\_3.pdf](http://www.tulips.tsukuba.ac.jp/pub/choken/youkou/3_3.pdf)) に詳しく述べた。この資料には参考文献を多数含んでいる。本稿では紙面の都合で参考文献の数は少なくしている。

はじめに述べたように、メタデータは、「データに関するデータ」とだけ定義されるため様々なものがメタデータの範疇に入れられる。以下では、書誌情報のように、情報資源を組織化・管理し、利用に提供するために、直接的に情報資源の性質を表すものについて述べる。

## 2.1 メタデータの目的と種類

メタデータを見る際に、大きく分けて、情報資源を組織化・管理するという提供者側・管理者側の視点と、情報資源を探し、利用するという利用者側の視点がある。たとえば、図書目録などの書誌データは主に前者の視点で作られてきており、検索システムが扱うメタデータは後者の視点から決められると言える。Dublin Core は情報資源の発見 (Resource Discovery) のためのメタデータと言われるため、基本的には後者の始点から決められてきているものである。

たとえば、図書館と博物館などでメタデータ記述に関する要求は異なる。ネットワーク上の情報資源と従来の冊子体資料でも要求は異なる。そのため、記述対象や応用目的にあわせて情報資源の性質を記述するためにメタデータ規則は決められる。

情報資源の記述といっても、(1)情報資源(の内容)に関する記述であり、かつ情報資源を探し出すために利用されることが主であるもの (Descriptive Metadata), (2)情報資源の保存やアクセス制御など管理方法に関するもの (Administrative Metadata), (3)情報資源の物理的・論理的内部構造に関するもの (Structural Metadata), (4)情報資源の利用に必要な技術的要件に関するもの (Technical Metadata) など多様である。たとえば、Dublin Core は情報資源の内容の記述が主である。デジタル資料のアーカイブを指向しているメタデータの場合には、これらの性質をすべて含んでいる。また、メタデータに関するデータ、すなわちメタ・メタデータ(たとえば、メタデータの作成日や作成者など)、異なるメタデータ規則を横断的に用いるためのメタデータ規則間の対応関係 (Crosswalk) などもある。

メタデータの記述に当たっては、記述対象をどのようにとらえるかも重要な視点である。たとえば、次節に示す IFLA の FRBR (Functional Requirements of Bibliographic Description) は記述対象を Work (作品), Expression (表現されたものとしての作品), Manifestation (表現された作品を具現化したもの), Item (Manifestation の個々の一点) の 4 段階でとらえるモデルを与えている。デジタルコンテンツの保存のための参照モデルを定義している Open Archival Information System (OAIS) は、保存対象となる情報資源を、それを表示するための情報とともにパッケージ化し、保存のために必要な情報を 4 つのカテゴリ (来歴, コンテキスト, 参照, 不変性) に分けて記述するモデルを提案している。

## 2.2 メタデータ規則の枠組み

はじめに書いた情報資源記述のためのメタデータ規則の枠組みを少し細かく分けると、メタデータ規則の構成要素を以下のように一般化してとらえることができる。

- (1) 記述対象属性の定義。たとえば、資源につけられた名前や作者など。
- (2) 記述対象の属性値の定義。これには、属性値として持つ値の統制語彙、値の表現形式が含まれる。
- (3) メタデータの構造の定義。たとえば、著者の記述は名前(姓・名)、所属、連絡先からなる、といった構造。
- (4) 省略可能性や繰り返し回数など、記述対象属性ごとに決められる制約。
- (5) 具体的なシステムごとに決められるメタデータの記述構文。たとえば、XML Schema や DTD を用いて決められる構造。
- (6) 実際の対象情報資源に対して、どのように記述すべき内容を抽出し、メタデータを記述するかに関する指針。

上のうち、(1)と(2)はメタデータの属性と属性値を表すことばの意味定義、言い換えるとメタデータのボキャブリティ定義である。(3)と(4)は特定の具象構文によらない構文定義、あるいは特定の記述形式によらない構造定義(すなわち、抽象構文定義)、(5)は具象構文定義、と言い換えることができる。(6)はそれ以外とは異なり記述上の指針を与えるものである。たとえば、目録規則にはこの要素が含まれている。そこで、本稿では、これ以降、(6)を除いたメタデータ規則を、特に断らない限り、メタデータスキーマと呼ぶことにする。

メタデータの属性と属性値のボキャブリティをメタデータボキャブリティと言うことにする。メタデータボキャブリティは属性のボキャブリティと属性値のボキャブリティに分けられる。属性、属性値のボキャブリティともに語の間



の意味的な関係を定義することも可能である。RDF Schema では Property と Class の概念を持ち、それぞれについて super-property と sub-property, super-class と sub-class の関係を持っている。たとえば、Dublin Core の場合には、Element Refinement Qualifier (エレメント詳細化限定子) という概念を持ち、属性とそれをより詳細化した属性を表現している。限定子によって意味が限定されたエレメントとものエレメントの間の関係定義のために Property 間の関係が利用できる。

このようにメタデータのボキャブラリと構文を分離してとらえることで、メタデータスキーマ間の関係付けが可能になる。こうすることで異なったメタデータスキーマで書かれたメタデータの横断的利用や、複数のメタデータスキーマを基礎にして応用に特化した規則を作るなどが容易になる。

### 3. メタデータのためのモデル

情報資源に関する記述をするには、どのような目的で、情報資源をどのようにとらえるかに関するモデルと、また表現の基盤となるデータモデルが必要である。以下に、記述対象のとらえ方に関するモデルとして IFLA の Functional Requirements for Bibliographic Records (FRBR), デジタルコンテンツの保存のためのモデルとして Open Archival Information System (OAIS) 参照モデル, メタデータ表現のための基本データモデルとして WWW コンソーシアムの Resource Description Framework のモデルを示す。ここで示したものの他にも、情報資源の集まり(コレクション)を対象として記述するもの[6], 「作成」、「修正」といった情報資源に起きる状態変化の視点から記述するもの[7]などがある。

#### 3.1 Functional Requirements for Bibliographic Records (FRBR)

図1に示す FRBR モデルは、IFLA の報告書[8]の中で示されているリソースに関する4階層のモデルである。Work は表現方法を問わない知的内容, Expression は何らかの表現方法によって Work を表したものの, Manifestation は Expression を何らかの媒体上に実現したもの, Item は Manifestation の個々のものを意味する。小説の場合で考えると、作者の考えたストーリーとしての小説は Work であり、それを具体的に表現したもの(小説の場合はテキスト中心に表現されたもの)が Expression である。同じ Work であっても、たとえば大人向けと子供向けといったように異なる Expression を持つ場合がある。また、ある ISBN で表されるものは Manifestation であり、一冊一冊の本がひとつの Item である。

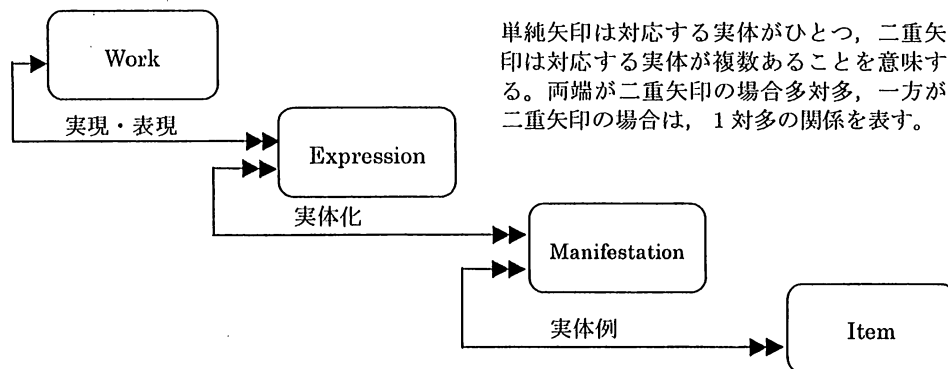


図1 IFLA FRBR における知的内容を表す実体 (Group 1 Entity) 間の関係

非デジタル資料の場合と同様、デジタル資料を組織化し、検索対象とし、また保存対象とするために、メタデータの記述対象が何であるかを正確にとらえる必要がある。ところが、デジタル情報資源の場合、その動的な性質ゆえにこのモデルをそのまま当てはめることは必ずしも容易ではない。情報資源の内容であるビット列が Expression から Item のいずれに対応するのか必ずしも明確ではない。たとえば、情報資源を利用環境毎にインストールして利用する場合、Item、すなわちコンテンツの実体(ファイル、あるいはビット列)は置かれている場所に依存する個別の実体であると考えられる。一般の Web 上の資料の場合、URL を ISBN と同様に実体化された著作物に与えられた識別子とすると、URL で指示される資源のソースファイルは Manifestation ととらえられる。この場合、ブラウザ上にダウンロードされたものが Item

にあたるにとらえられるが、その実体に永続性が保障されない。また、スタイルシートによって表現形式 (HTML 文書) を動的に作り出すことのできる Web 文書の場合、もとの XML 文書を Expression, 表示用の文書を Manifestation ととらえることもできるし、XML 文書とスタイルシートを 1 組のものとして Manifestation ととらえることもできる。

デジタル情報資源の場合、ソフトウェアで動的に作り出すことができるので、資源の提供者側で持つ資源の実体、言い換えるとネットワーク上で資源として識別される実体と、利用者が受け取る資源の実体が異なり得る。資源の発見と保存のどちらの視点においても、動的な資源に対するメタデータの記述のための基本モデルが求められる。

### 3.2 OAIS 参照モデル — デジタルコンテンツの保存のためのメタデータ[9]

デジタル情報資源の保存のモデルとして Consultative Committee for Space Data Systems (CCSDS) の開発した OAIS 参照モデル (Reference Model for an Open Archival Information System) の国際標準化が進められている [10]。OAIS 参照モデルでは、情報オブジェクトを保存するために情報パッケージの概念を導入している。情報パッケージは図 2 に示すように内容情報 (Content Information)、保存記述情報 (Preservation Description Information, PDI)、およびパッケージ化情報 (Packaging Information) からできている。ここで、内容情報は、保存対象のデータオブジェクトとその表現情報 (Representation Information) からなる情報オブジェクト (Information Object) と呼ばれる。また、情報パッケージに関する記述 (Descriptive Information about Package) も必要とされる。この 4 要素のうち内容情報以外はメタデータと言える。PDI は保存対象に関する情報を持つのにに対し、他はパッケージに関する情報をもつ。PDI の内容は次の 4 種類の情報に規定されている。

- ・ 来歴 (Provenance) : コンテンツ (保存対象の情報資源) の出处や保存に関わる処理の履歴。
- ・ コンテキスト (Context) : 情報パッケージ外のオブジェクトとの関係。
- ・ 参照 (Reference) : コンテンツを一意に識別するための識別子、もしくは識別のためのシステム。
- ・ 不変性 (Fixity) : コンテンツを保護するための情報。

デジタル情報資源の保存のためのメタデータスキーマに関する検討が進められ、Cedars や NEDLIB、OCLC と RLG の共同ワーキンググループでは OAIS 参照モデルに基づくメタデータスキーマスキーマの報告を出している。

メタデータだけではデジタルコンテンツの保存の問題をすべて解決することはできない。しかしながら、適切なメタデータなしには保存は困難になる。

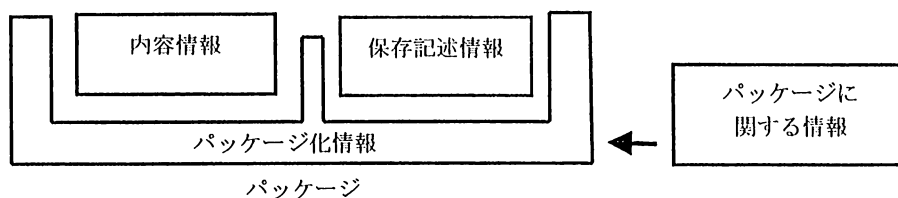


図 2 情報パッケージの概念構造

### 3.3 Resource Description Framework

XML はテキスト化可能な任意のメタデータの表現に利用することができる。しかしながら、メタデータの流通性や相互運用性のためには、異なるメタデータスキーマに対する統一的な記述構文を与える必要がある。RDF はこうした目的の下に開発されたもので、下に挙げる基本的性質を持つものである。

- (1) 情報資源が持つ性質を表すために、たとえば、「この論文の著者は杉本重雄である」というように「ある情報資源が、ある属性としてなんらかの属性値を持つ」という3つ組 (情報資源, 属性, 属性値) で表す。この例では {情報資源 = この論文, 属性 = 著者である, 属性値 = 杉本重雄} の3つ組になる。複合した性質を書く場合には3つ組を組み合わせて表現する。(図 3 および図4に例を示す。)
- (2) XML 上での表現形式を決めている。
- (3) 属性や属性値の表現のためのボキャブラリは RDF Schema として定義する。すなわち、XML 表現

にタグとしてあらわされる属性の名前と意味が別途(XML形式で)定義される。  
 (4) Namespace を用いて複数のメタデータスキーマで定義されたボキャブラリを用いることができる。  
 図4をRDFの構文に基づきXML表現すると下記ようになる。

```
<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://www.dl.ulis.ac.jp/DLjournal/
    No_21/2-sugimoto/2-sugimoto.html">
    <dc:title xml:lang="ja"> Dublin Core についてー最近の動向, 特に qualifier について
      </dc:title>
    <dc:creator xml:lang="ja">杉本重雄</dc:creator>
    <dc:date>2000-9-27</dc:date>
  </rdf:Description>
</rdf:RDF>
```

図5は Warwick Framework の概念図である。Warwick Framework は Dublin Core の開発の過程で提案されたもので複数のメタデータスキーマに基づく記述の枠組みを与える。RDF は Warwick Framework を, XML 上に実現したものということもできる。

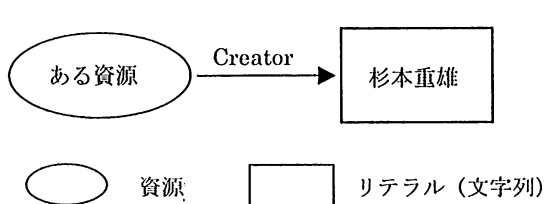


図3 RDF のデータモデルー基本構造

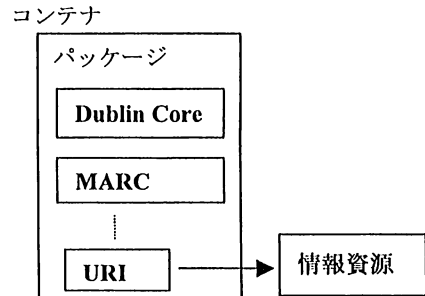


図5 Warwick Framework

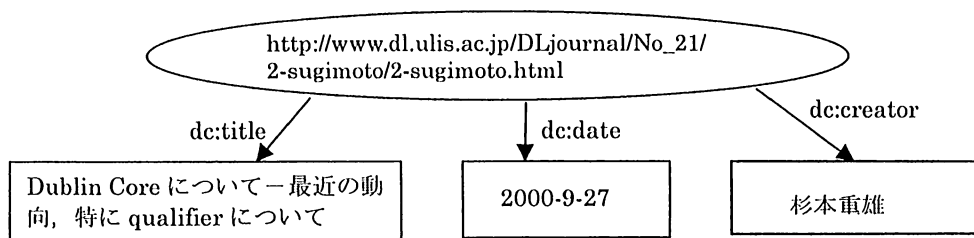


図4 RDF のデータモデルーSimple Dublin Core

#### 4. Dublin Core におけるいくつかの重要な基礎概念

##### 4.1 概要

Dublin Core Metadata Element Set (DCMES)はインターネットにおける様々な情報資源の記述と発見のためのメタデータスキーマとして広く認められている。Semantic Interoperability が Dublin Core のキーワードとして述べられることから理解できるように, Dublin Core は多様な分野で作られるメタデータを相互に利用できるようにすること, 分野の違いを越えて情報資源を検索できるようにすることを目標として開発されてきたものである。そのため, Dublin Core のエレメントセットには, 多様な分野にお

いて共通に利用できるエレメントのみが含まれている。こうした視点から最初に決められたエレメントセットは 15 の基本属性(エレメントと呼ぶ)のみからなるもので、Simple Dublin Core と呼ばれる。Simple Dublin Core は ISO 15836 として標準化されている。

Dublin Core の開発組織である Dublin Core Metadata Initiative (DCMI) は、だれでもが参加できるボランティアベースの組織である。DCMI には、いろいろな応用領域、世界の様々な国と地域からの参加者がいる。そのため、それぞれの分野からの要求を基に新しいエレメントなどの記述要素が加えられている。

分野にまたがったメタデータの Semantic Interoperability の視点から、筆者は、(1) Warwick Framework, (2) Dumb-Down 原則, (3) Application Profile の 3 点が、Dublin Core を理解する上での重要な基礎概念であると考えている。また、これに加えてネットワーク上でのメタデータボキャブラリの流通・共有を図るために、メタデータスキーマレジストリ(Metadata Schema Registry)が重要な役割を果たすことになる。以下の節では、これらについて詳しく述べる。

## 4.2 Warwick Framework

Warwick Framework は 1996 年春に開かれた Dublin Core の第 2 回のワークショップで提案された概念である。Warwick Framework は、図 5 に示したように、複数のメタデータスキーマを利用してメタデータを記述するための枠組みを示している。従来、メタデータ規則は、ある目的のために閉じたものとして設計されてきた。一方、ネットワーク上で様々な領域の情報資源を、様々な利用者に提供するための十分に詳細な記述能力を持つ閉じたメタデータスキーマを実現することは現実的ではない。そのため、いくつものメタデータスキーマから必要な記述要素を取り出して利用し、それらをまとめた記述を行うことが望まれる。こうすることで、それぞれの領域で十分に理解されている既存のメタデータスキーマを適切に利用することができる。一方、領域ごとのメタデータスキーマだけでは領域にまたがるメタデータの利用が難しくなるので、Dublin Core を領域間にまたがる共通のスキーマとして利用すればよい。

Warwick Framework は、複数のメタデータスキーマを用いたメタデータ記述のための基本概念を与えたもので、WWW コンソーシアムが開発した RDF はこの概念を具現化したものであるといえる。また、同様に、後述の Application Profile も複数のスキーマに基づくメタデータ記述のための仕組みを与えている。

## 4.3 Dumb-Down 原則

Dublin Core を実際の情報資源記述のために利用するには基本エレメントだけでは不十分であり、より詳細で正確な記述のための記述要素として限定子を導入することが必要であると言う議論が早くからあった。現在の DCMES には、情報資源の属性を表すエレメント、エレメントの意味の詳細化をするエレメント詳細化限定子(Element Refinement Qualifier)、および属性値(すなわち、エレメントの値)の記述形式あるいは記述に用いる語彙を指示するコード化スキーマ限定子(Encoding Scheme Qualifier)が含まれている。ところが、「著者」を表すために「著者の名前、所属、連絡先」を構造として表すための、構造を表すエレメントはこの中に含まれていない。従来からある目録規則や他のメタデータ規則にはこうしたエレメントがあるのに対して、この点は Dublin Core の特徴であると言える。DCMES の開発過程においては構造を持つエレメントの必要性も議論されてきたが、下に示す Dumb-Down 原則に適合しないという理由で除外された。

Dumb-Down 原則は、DCMES の Semantic Interoperability を保証するために導入されたもので、限定子を導入する際の適合性判断のためのルールである。Dumb-Down 原則は、限定子を含めたメタデータ記述から限定子を取り除いても、すなわち、基本エレメントだけでの記述にしても、メタデータとして矛盾してはならない、という原則である。この原則により、領域のニーズに合わせた限定子を導入したとしても、領域にまたがった利用の際には限定子を取り去れば領域間での矛盾が生じないことを保証できる。たとえば、Date(日付)エレメントに対して Submitted(投稿)や Revised(修正)という限定子をつけて、Date Submitted(投稿日付)や Date Revised(修正日付)という意味的に詳細化されたエレメントをつけた場合、限定子の部分を取り除いても、投稿や修正と言う概念を持たない日付エレメントとの矛盾は生じない。一方、たとえば、Creator(作成者)エレメントに対してその名前、所属、連絡先といった属性を限定子として定義し、それらを用いた記述を考えると、単純に限定子を取り除いてしまうと矛盾を生じることが

あることが容易に理解できる。作成者エレメントに対して、名前、所属、連絡先という限定子の組と、第1作成者、その他の作成者という限定子の組を考えてみよう。「作成者. 名前=杉本重雄, 作成者. 所属=筑波大学, 作成者. 連絡先=つくば市春日」という記述から、単純に限定子を取り除くと「作成者=杉本重雄, 筑波大学, つくば市春日」となる、一方、「作成者. 第1作成者=杉本重雄, 作成者. その他の作成者=永森光晴, 作成者. その他の作成者=Thomas Baker」という記述からは「作成者=杉本重雄, 永森光晴, Thomas Baker」という記述が得られる。この二つの記述が矛盾していることは明らかである。構造を持つエレメントに関して限定子を導入しない代わりに、Dublin Core ではそのエレメントの値を表現するための適切なエレメントセットを利用すればよいとしている。

DCMI では Usage Board と呼ぶ委員会を設け、そこで新しいエレメントや限定子の導入に関する認定を行っている。限定子の導入に関してはDumb-Down原則に反しないものは適合(Conforming)とされ、さらに領域にまたがって利用性が高いと判断されるものは推奨(Recommended)とされる。

#### 4.4 Application Profile

Application Profile は、応用ごとに、どのようなメタデータボキャブラリを利用するのか、どのような構造的・構文的制約(たとえば、必須・省略可能の別、繰り返し回数の制約など)を持つのかを定義する。加えて、応用ごとに決まる記述上の指針を含む場合もあるが、ここではその点は考えない。図6にApplication Profileの概念図を示す。このように、Application Profileは複数のメタデータスキーマから、応用にとって必要な記述要素のみを選び、適切な構造的制約を与えるために用いられる。Application Profileは応用毎のメタデータスキーマを与える一方、メタデータボキャブラリに関しては既存のメタデータスキーマで決められたものを用いるので、メタデータをエレメントにばらしてとらえると、異なるApplication Profileから作成されたメタデータの間での意味的互換性が保たれる。

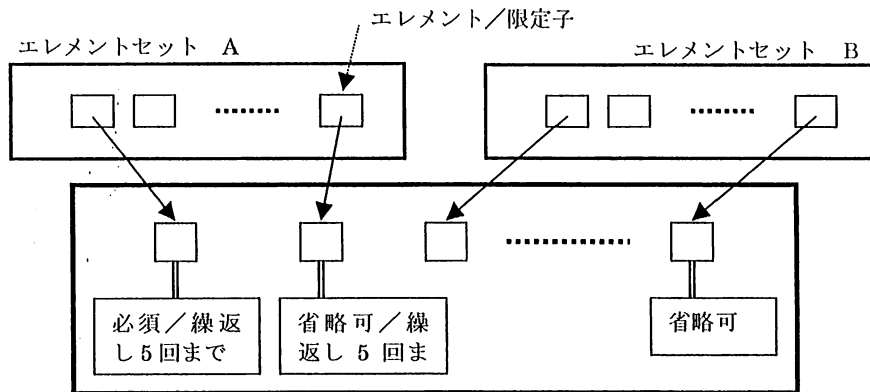


図6 エレメントセットとアプリケーションプロファイル

#### 4.5 メタデータのボキャブラリと Schema Registry

2.2 節に示したメタデータの構成要素のはじめのふたつはメタデータを記述するために用いるためのボキャブラリを決めている。メタデータの流通性を高めるにはメタデータボキャブラリの流通性を高める必要がある。この課題を解決するために、メタデータボキャブラリを登録し、ネットワーク上で提供するサービスであるメタデータスキーマレジストリの開発が進められている[11]。DCMI でもメタデータスキーマレジストリの開発を進めてきた[12]。現在のサービス(DCMI レジストリと呼ぶ)は、DCMES のボキャブラリ、すなわちエレメントと限定子の参照記述を提供している。DCMI レジストリでは、語の定義を RDF Schema を用いて定義し、人間と機械の両方に対するインタフェースを提供している。また、このレジストリでは 23ヶ国語に翻訳された参照記述を提供している。

ネットワーク上において、応用に応じた多様なメタデータ規則が必要とされることを考慮すると、メタデータのボキャブラリ定義の再利用性を高めること、ボキャブラリを共有することで、メタデータの相互利用性を

高めることが求められる。こうした要求を満たすためにメタデータスキーマレジストリは重要な役割を持っている。また、セマンティックウェブの視点からは、オントロジーの共有を支えるものと理解することもできる。

## 5. おわりに

本稿では、情報資源記述のためのメタデータ、特に Dublin Core を中心として、その基本となるモデルについて述べた。XML はインターネット上での文書の記述、データ交換のために広く用いられており、ここで述べたメタデータも XML によって表現されることはいうまでもない。たとえば、MODS は MARC を基にして XML での記述を指向したものであり、METS は、SGML を用いて作られた EAD (Encoded Archival Description) を基礎にして開発されたものである。また、学術文献のリポジトリ間の協調を進める Open Archives Initiative でも Simple Dublin Core を基礎とするメタデータの XML での表現を決めている。DCMI では XML および RDF による記述の推奨形式を決めている。

Interoperability はインターネット上におけるメタデータにとって非常に重要な要件である。従来であれば個々の応用ごとのメタデータ規則で十分であったと考えられるが、インターネット上では個々の、いわば草の根コミュニティがその応用に適したメタデータ規則を必要とし、しかもそうした規則に基づくメタデータを異なるコミュニティの間で相互に利用できるようにすることが求められる。その視点からは、メタデータボキャブラリ(すなわちセマンティクス)と構造(すなわちシンタックス)の定義の分離、それに基づくメタデータスキーマの共有機構が重要であると考えられる。これらを実現していく上で、XML は基盤技術であることは疑えないし、RDF Schema や OWL といったボキャブラリの記述形式も大きな役割を持つと思われる。

参考文献 (以下の URL については 2003 年 10 月時点で確認。)

- [1] Dublin Core Metadata initiative, <http://dublincore.org/>
- [2] 杉本重雄, "Dublin Core について(2回連載)", 情報管理, Vol.45, no.4, 2002.7, pp.241-254, no.5, 2002.8, pp.321-335
- [3] MODS: Metadata Object Description Schema (official web site), <http://www.loc.gov/standards/mods/>
- [4] METS: Metadata Encoding & Transmission Standard (official web site), <http://www.loc.gov/standards/mets/>
- [5] Resource Description Framework (RDF) / W3C Semantic Web Activity, <http://www.w3.org/RDF/>
- [6] Hirle, P.B. (ed.), Special Issue on Collection Description, D-Lib Magazine, vol.6, no.9, 2000.9, <http://www.dlib.org/dlib/september00/09contents.html>
- [7] Lagoze, C. and Hunter, J., "The ABC Ontology and Model, Journal of Digital Information", vol.2 issue 2, 2001.11, <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/>
- [8] IFLA, "Functional Requirements for Bibliographic Records", <http://www.ifla.org/VII/s13/frbr/frbr.pdf>, 1998
- [9] 杉本重雄, Maria Luisa Calanag, デジタルアーカイブとメタデータ, 人工知能学会誌, 2003.5, Vol.18, No.3, pp217-223
- [10] Consultative Committee for Space Data Systems, "Reference Model for an Open Archival Information System (OAIS)", CCSDS 650.0-B-1 Blue Book, 2002.1 <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>
- [11] Baker, T., et al. "Principles of Metadata Registries, A White Paper of the DELOS Working Group on Registries", <http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/Registries.pdf>, 2003
- [12] DCMI Registry Working Group, <http://www.dublincore.org/groups/registry/>

# オーディオビジュアルコンテンツのためのメタデータ国際標準

早稲田大学 大学院国際情報通信研究科

亀山 渉

## 1 はじめに

21世紀のキラアプリケーションはコンテンツであるという認識の下、コンテンツ流通が様々な角度から検討されている。通信の分野ではコンテンツ配信ネットワーク (CDN: Contents Distribution/Delivery Network) の検討が盛んであり、実用的な見地からの検討もかなり本格化している [1]。また、放送の分野ではデジタル放送と蓄積メディアの進展により、PDR (Personal Digital/Data Recorder) を利用した蓄積型放送方式 (日本では「サーバ型放送」と呼ばれる) も実用化が間近に迫っている状況にある [2]。

このようにコンテンツが流通するための技術やインフラが整備されれば、氾濫するコンテンツの中から自分が興味のあるコンテンツのみを検索して利用するためのユーザインターフェース構築が重要になることは当然で、この観点から、オーディオビジュアルコンテンツのメタデータは主として検討されていると言って良い。しかも、サービスの性質上、様々な機器間やサービス間でのインターオペラビリティが重要であり、この意味で、国際標準化が1990年代の終りから活発に進められてきた。

国際標準化は、主として、ISO/IEC JTC1/SC29/WG11[3] (以下 MPEG と呼ぶ) と TV-Anytime フォーラム [4] によって進められおり、前者は MPEG-7 と俗称される国際標準、後者は TV-Anytime メタデータとして有名である。本稿では、これらの標準の概要について述べる。

## 2 メタデータの役割

オーディオビジュアルコンテンツにおけるメタデータの役割は、他のマルチメディアデータに対するメタデータの役割と基本的に変わることはない。しかしながら、人に視聴されることが最終目的であるコンテンツに付随するメタデータであるため、多少特筆すべき点がある。以下にその役割をまとめる。

### 2.1 検索情報記述としてのメタデータ

一般に、メタデータに主として求められるのはこの役割である。コンテンツをコンパクトに記述し、様々な利用側面から検索が可能となるように、細分化された構造を持ったメタデータが各種の国際標準では規定されている。

### 2.2 権利情報記述としてのメタデータ

コンテンツ利用時における権利処理のための情報記述が、とりわけオーディオビジュアルコンテンツでは重要である。これには買い取りや課金のための情報だけでなく、どのような利用が許されるのかといった情報まで記述できる必要がある。

例えば、蓄積型放送においては、あるコンテンツは蓄積しても視聴が許される、あるコンテンツはリアルタイムの放送時にのみ視聴が許され蓄積は許されない、といった利用条件の区別がコンテンツ毎に必要な。この他にも、ある番組を見た場合にのみ別のある番組の視聴が許される、といったような特定のビジネスモデルに直結したような利用条件も記述できることが望ましい。

このような権利情報や利用条件に関する情報記述を TV-Anytime フォーラムでは RMPI (Rights

Management and Protection Information) と呼び、いわゆる DRM システムが様々な利用許諾処理を行うための情報を提供するものと定義している。RMPI の詳細はまだ定まっていないが、tvax[6] と呼ばれる XrML (Extensible Rights Markup Language[5]) の語彙データが TV-Anytime フォーラムで検討されている。また同様の語彙の検討は MPEG-21[7, 8] として MPEG でも検討されており、こちらの語彙拡張は mx と呼ばれている。

### 2.3 誘引情報記述としてのメタデータ

コンテンツの側面を記述する上述した 2 種類のメタデータがあれば利用に関して特に問題がなさそうに思えるが、実際にはこれだけでは不十分で、コンテンツ流通は期待できない。

一般にコンテンツを流通させるためにはメタデータこそ流通させる必要があり、コンテンツ流通のための技術やシステムの研究開発も重要であるが、メタデータを流通させるための技術やシステム、更にはビジネスモデルの検討も重要であると言える。このようなメタデータ流通環境では、コンテンツの魅力や余すところなく記述できるメタデータを定義し、それを流通させることがビジネス上重要となってくる。

このことから、TV-Anytime フォーラムでは、メタデータをアトラクタ (Attractor) とも呼んでおり、コンテンツがエンドユーザに利用されるための最重要情報であると位置付けている。この意味から、ユーザを誘引するための様々な付加的な情報も有用なメタデータとなり得る。つまり、コンテンツそのものの情報ではないが、コンテンツの運用や利用に関わる情報もメタデータとして利用価値が非常に高いと言える。

例えば、あるコンテンツを 1 回視聴する料金は 50 円であるというようなものは、従来の考えでは先に述べた RMPI の一部であり、権利処理機器の中で利用時に内部的に処理できる情報であれば良いと考えられていたが、このような情報はエ

ンドユーザを誘引するための情報として、コンテンツ利用時だけでなく、エンドユーザを視聴に誘うために常時有益な情報である。このようなエンドユーザを誘引するためのメタデータ情報を TV-Anytime フォーラムでは E-Plier (電子チラシといったような意味) と呼び、メタデータの新しい部分情報として標準化の作業を始めたばかりである [9, 10]。

### 2.4 本稿の範囲

以上のように、一口にメタデータといっても、オーディオビジュアルコンテンツでは様々な側面があり、2.2 や 2.3 で述べたような特有のメタデータ記述も標準化が進んでいる。そのどれもが興味深く、技術的にも面白い内容を含んでいるが、紙面の都合と標準化途中であるという理由から、本稿では取り上げない。以下、本稿では 2.1 の観点から、国際標準方式について述べて行く。

## 3 MPEG-7

### 3.1 概要

MPEG が標準化した初めての非圧縮符号化方式として注目されたのが、俗に MPEG-7 と呼ばれるマルチメディア用メタデータ記述のための国際標準である。正式には “Multimedia Content Description Interface” と呼ばれ、2003 年 9 月末現在、下記に示す複数のパートから成り立っている (いくつかの Amendment 文書や Corrigendum 文書が発行されているがここでは省略する)。

1. ISO/IEC 15938-1:2002, Part 1: System
2. ISO/IEC 15938-2:2002, Part 2: Description definition language
3. ISO/IEC 15938-3:2002, Part 3: Visual
4. ISO/IEC 15938-4:2002, Part 4: Audio
5. ISO/IEC 15938-5:2003, Part 5: Multimedia description schemes
6. ISO/IEC 15938-6:2003, Part 6: Reference software
7. ISO/IEC FDIS 15938-7, Part 7: Conformance testing



8. ISO/IEC TR 15938-8:2002, Part 8: Extraction and use of MPEG-7 descriptions
9. ISO/IEC WD 15938-10, Part 10: Schema definition

この内、パート1はMPEG-7で記述されたデータを圧縮、伝送、操作するための規格、パート2はメタデータを記述するためのスキーマ言語定義、パート3からパート5までがメタデータの詳細な記述方法の規格となっており、これらがMPEG-7国際標準の重要な柱を構成している。これらの関係を図1に示す。

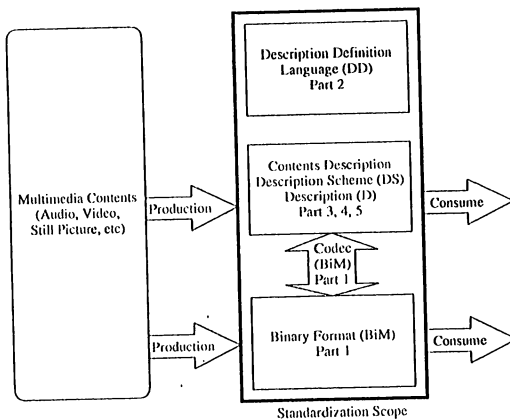


図1: MPEG-7の標準化範囲

### 3.2 システム

後で述べるように、MPEG-7のメタデータはXMLを基にして書かれるが、一般に、XMLで書かれた文書は情報量の観点からは必ずしも小さいものではないという問題点がある。特にMPEG-7では、スキーマ定義だけでも7000行を越すという膨大な規定であるため、詳細にメタデータを記述するとその量は非常に大きなものとなる可能性がある。MPEG-7そのものは汎用性のあるマルチメディア用メタデータ標準であるが、それでも、デジタル放送などでこれを用いる場合、貧弱なメモリしか持たないセットトップボックスでこれを扱わなければならないといったことも十分考えられるため、何らかの対策が要求

条件として上げられていた。

パート1で主として規定するのは、この問題を解決するためのXMLバイナリ化方式であり、BiM (Binary Format for MPEG-7) と呼ばれている。BiMは大変興味深いバイナリフォーマットで、XMLデータを圧縮するだけでなく、バイナリ状態のままXML中の部分木を選択的に復号、伝送、削除、追加することができるような仕組みを備えている。このことから、先に述べたような環境でも十分実用的なXML処理を行えるようになってきている。

### 3.3 DDL

パート2で規定されるMPEG-7 DDL (Description Definition Language) は、XMLスキーマに若干の拡張を施して定義されたスキーマ言語であり、その拡張は以下のようなものである。

#### 1. 行列を表すデータタイプの付加

次元を示す正の数をアトリビュートとして指定することで、行列表現が可能になるようにするスキーマ定義。例えば、図2のようなインスタンスの作成が可能。オーディオビジュアルの信号特徴は行列で記述される必要があることから追加された。

#### 2. 時間を示す単位の追加

ISO 8601に準拠するフォーマットとして、時刻は“YYYY-MM-DDThh:mm:ss.nnn.ff FNNN±hh:mm”、時間長は“(–)PnDTnHnMnSnNfnF±hh:mmZ”と表現。“nnn.ff NNN”は一秒当たりNNNフレーム存在しそのnnn.ff番目という意味であり、“nNnfnF”は同様に一秒当たりn(3番目のn)フレーム存在しそのn.n(1番目と2番目のn)フレーム番目という意味になる(“5N1f30F”で、一秒当たり30フレームで5.1フレーム目の意味)。映像ではフレーム数で時間を指定することが多く、その表現が可能となるようにした。

```
<IntegerMatrix mpeg7:dim="2 4">
  1 2 3 4
  5 6 7 8
</IntegerMatrix>
```

図 2: 2×4 の行列の定義

### 3.4 低レベルメタデータ

MPEG-7のメタデータは大きく分けて、低レベルメタデータと高レベルメタデータとに分類することができる。ここでいう高低は、人間の意味レベルにおいての高低を示している。つまり、低レベルメタデータはオーディオビジュアル情報から自動的にあるアルゴリズムによって抽出できる原信号の特徴を直接的に表す情報であり、高レベルメタデータはこのような自動抽出ができないオーディオビジュアル情報の意味内容を表す情報である。この様子を図3に示す。

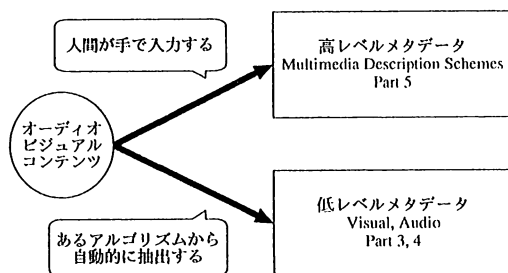


図 3: 低レベルメタデータと高レベルメタデータ

低レベルメタデータは、パート3に映像情報に関するものが、パート4に音声情報に関するものが定義されている。これらのメタデータは、それぞれ、ビジュアル記述子、オーディオ記述子とも呼ばれている。

#### 3.4.1 ビジュアル記述子

映像信号の特徴を効率良く表すものとして、以下のものが規定されている。

1. 色特徴記述子  
画面中の支配色、色の使用頻度、画面上での色配置といった特徴量を示す。
2. テクスチャ記述子

一様な模様が存在、エッジの含まれる割合といった特徴量を示す。

#### 3. 形状記述子

映像中の形を持つオブジェクトのシルエットの特徴量を示す。

#### 4. 動き記述子

被写体の動きパラメータ、被写体の重心の軌跡、動きの激しさや方向、カメラモーションといった特徴量を示す。

#### 5. 位置指定記述子

映像の時空間的な位置を指定し、指定した位置に他の情報を付与する。

#### 6. 顔特徴記述子

人間の顔を表現する記述子で、顔認証などに使用される。

#### 7. コンテナツール

様々な方法で映像情報の部分集合を指定し、それに対してビジュアル記述子を付与する。

先にも述べたように、このような特徴量を抽出するアルゴリズムが標準文書中に定められており、ある特徴量を記述すれば、誰でも矛盾のない均一な検索結果が得られるのがビジュアル記述子の特徴である。

#### 3.4.2 オーディオ記述子

オーディオの低レベルメタデータでは、少し意味内容の高いものも含まれており、ビジュアル記述子とは少々異なった内容のメタデータが定義されている。単なる記述子だけではないことから、オーディオ記述スキームと呼ばれている。

1. オーディオシグニチャ  
オーディオ信号の統計的な特徴量からオーディオ固有の値を指定する。
2. ティンブラ  
楽器の音色の印象を音の信号特徴と組み合わせで記述する。
3. サウンド認識とインデキシング  
サウンドのクラス分類やインデキシングのためのメタデータを記述する。

4. 発話内容

発話認識を行うための補助データの記述を可能とする。

5. メロディ検索

メロディの推移外形や詳細を記述する。

ビジュアル記述子と違い、ある程度 of アプリケーションを想定したようなメタデータが利用できるのがオーディオ記述スキームの特徴である。

3.5 高レベルメタデータ

MPEG-7 の高レベルメタデータは、MDS (Multimedia Description Schemes) と呼ばれ、パート 5 に詳細が規定されている。規定されている MDS の種類を図 4 に示し、以下に説明を加える。

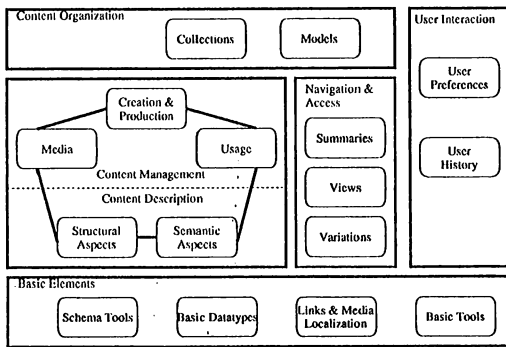


図 4: MPEG-7 MDS

1. 基本要素 (Basic Element)

MPEG-7 メタデータの基本的な構造や要素型の規定の集合を含む。

2. コンテンツ管理 (Content Management)

コンテンツの外部情報として、制作関連、メディア関連、利用形態関連の規定を含む。

3. コンテンツ記述 (Content Description)

コンテンツの構造とその意味内容について記述する規定を含む。

4. その他

ユーザの嗜好やコンテンツ使用履歴などを記述する規定などがある。

MPEG-7 MDS で記述できるメタデータは非常に範囲が広く、スキーマ定義も膨大な量である。これは、MPEG-7 MDS が汎用的に作られているためである。例えば、B2B で交換されるメタデータと B2C で交換されるメタデータでは、記述すべき内容が異なり、様々なメタデータ要素の重要性も異なっていると考えられる。しかしながら、先にも述べたように、MPEG-7 MDS は汎用的であるため、あるアプリケーションにとって必ずしも重要でないメタデータ要素を含んでしまっている。そこで、MPEG-7 MDS を実用的に使用するためには、アプリケーション毎のプロファイル作りが重要であると考えられるが、MPEG での議論はそれほど進んではない状況にある。

4 TV-Anytime メタデータ

4.1 概要

1999 年に設立された TV-Anytime フォーラムでは、近未来に PDR が放送の視聴形態を根本的に変えてしまうであろうことを念頭に、その環境における重要な技術をインターオペラビリティが可能なように規定することを目標に設立された業界標準化団体である。2003 年 9 月末現在、業界横断的に約 70 の団体がメンバとなり、活発な活動を続けている。

TV-Anytime フォーラムでは、技術仕様をフェーズ 1 とフェーズ 2 と呼ばれる 2 つのフェーズに分けて標準化活動を進めており、フェーズ 1 は先に述べた PDR を中心とした蓄積型放送アプリケーションを前提としている。フェーズ 2 の詳細は本稿の範囲を越えるので、文献 [9, 10] を参照して頂きたい。

2003 年 3 月にフェーズ 1 仕様は検討を終了し、規格文書としての安定性と受け入れを考え、ETSI (European Telecommunication Standards Institute) の技術文書として発行することをメンバー一致で選択した。このため、ETSI のサイトから無償で規格書をダウンロードすることができるようになっている [11]。

TV-Anytime フォーラムがフェーズ1で規定する技術要素は、大きく分けて以下の4つに分けることができる。

1. メタデータ
2. コンテンツを参照するための参照ID規定とIDを同定する方法
3. 権利処理と保護のための技術規定
4. メタデータ交換プロトコルとその保護手法

メタデータは、MPEG-7とは異なり、PDR環境におけるB2Cでのメタデータ利用を明確なアプリケーションとして想定しており、実用性の非常に高い規定となっている。事実、このメタデータ仕様は、ヨーロッパのデジタル放送規格を推進するDVB (Digital Video Broadcasting)、北米のデジタル放送規格を推進するATSC (Advanced Television Systems Committee)、及び日本の電波産業会で採用され、放送アプリケーションでは広く受け入れられるメタデータ仕様となっている。以下、このメタデータの構造の概要とメタデータ交換プロトコルについて述べる。

#### 4.2 メタデータの構造

MPEG-7を基にして作られているが、具体的なアプリケーションをあらかじめ設定しているため、不要なものは削り、記述が足りない場合には新たに要素を定義し直し、欠けているものは新たに追加を行っている。TV-Anytimeメタデータは、<TVAMain>をXMLのトップ要素として、いくつかのテーブルを中に包含する構造となっている。以下にそれらのテーブルが何を情報として含むことができるかを示す。

1. ProgramInformationTable  
コンテンツに対する情報を保持しており、いわゆる「番組」の情報を持っている。
2. GroupInformationTable  
複数のコンテンツに対する情報を保持しており、シリーズ番組等を表現できる。関連性のある番組を一括して検索する場合等に有効な情報を持っている。

#### 3. ProgramReviewTable

コンテンツに対する批評情報を保持し、評価の結果やレーティング情報を付加することができる。

#### 4. ProgramLocationTable

コンテンツがどこにあり、どのようなプロトコルを使用してアクセス可能かについての情報をもっている。TV-Anytimeのコンテンツモデルでは、同一タイトルのコンテンツに対して複数のサービスがあることを許しているため、このような情報が必要となる。放送として放映スケジュールが組まれたコンテンツなのか、オンデマンドでアクセスできるコンテンツなのか等が記述できる。

#### 5. SegmentInformationTable

コンテンツのハイライト視聴やブックマークの張り付けを許すため、コンテンツの部分(これをセグメントと呼ぶ)を指定する情報を保持する。MPEG-7にも同様の構造があるが、新たに定義し直している。

#### 6. CreditsInformationTable

コンテンツのクレジット情報を保持しており、誰がどこで制作したかやコピーライトの情報等を保持している。

#### 7. ServiceInformationTable

コンテンツサービスに関する情報を保持しており、どのようなサービスが、どの期間、誰によって行われるか等の情報を保持している。

#### 8. ClassificationSchemeTable

メタデータの様々な記述のクラス分けや、別名の付与等に使われる情報を保持している。

この他、MPEG-7から視聴嗜好メタデータ (mpeg7:UserPreferencesType)、視聴履歴メタデータ (mpeg7:UsageHistory) を一部流用しており、これらの情報はユーザメタデータと総称されている。

4.3 メタデータ交換プロトコル

以上のようなメタデータは、例えば MPEG-2 TS の中に含まれてコンテンツと併せて送られたり、オフラインでセットトップボックスが何らかの媒体から読み込んだりすることが想定されている。基本的に、TV-Anytime フォーラムはこのようなメタデータの送り方に対して何ら規定を設けてはいない。というのも、放送は地域的な規定によって運用されていることから一つの方式に絞ることが困難であることや、方式を決めることによって限定的なビジネスモデルしか与えられないことを避けるためである。

一方で、近未来の放送とインターネットの融合を考えると、インターネットは先に述べた地域的な限定性がなく、また、放送を補完する重要なサービスを提供するものと考えられることから、TV-Anytime フォーラムはインターネット上でのメタデータ交換方式については規格を定めた。これが “Metadata Service over a Bi-directional Network” と呼ばれる規定で、パート 6 として標準化された [11]。ここで規定されている方式は基本的に Web サービスのフレームワークを使用するものであり、先に述べた TV-Anytime メタデータを様々な方法で検索できるインターフェースを備えている。概要を図 5 に示す。

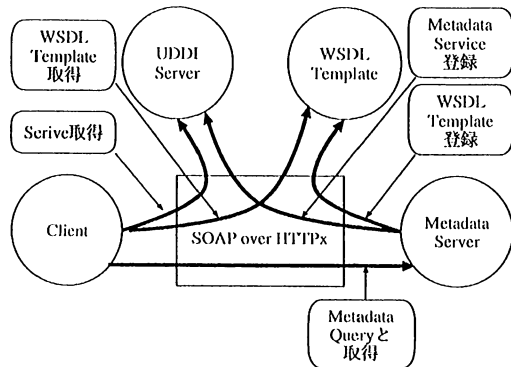


図 5: Web サービスを利用したメタデータ交換

また、このサービスでは、メタデータサーバからの一方向的な情報の流れを仮定しているだけ

でなく、エンドユーザからサーバに対する情報のアップロードも考えられている。例えば、あらかじめ自分のプロフィール情報などをサーバに送っておけば、より自分に適した検索をサーバで行える可能性がある。このため、一部のデータは安全に送る必要があり、図 6 に示すようなセキュアプロトコルスタックを使用する方式が規格化されている。

|                    |
|--------------------|
| TVA XML (Metadata) |
| SOAP               |
| HTTP               |
| TLS/SSL            |
| TCP                |
| IP                 |

図 6: メタデータ交換プロトコルスタック

いずれにせよ、このようなメタデータ交換プロトコルは TV-Anytime フォーラムが初めて規格化したものであり、将来、様々な応用や拡張によって、メタデータ交換とその利用に基づいた新しいサービスが期待できると思われる。

5 その他のメタデータ標準

その他のメタデータ標準として重要なのは、Pro-MPEG フォーラム [12] と AAF アソシエーション [13] が共同で開発した MXF (Media Exchange Format) である。これは、B2B におけるオーディオビジュアル素材やオーディオビジュアルパッケージの交換を目的に作成されたメタデータであり、いわゆるボスプロなどの処理に大いに利用されるものと期待されている。事実、BBC や CNN では、MXF 対応の放送機器を大量に導入しようとしており、放送業界では急速に普及するメタデータ標準となることが予想される。

一方で、このような B2B メタデータを B2C メタデータ、つまり TV-Anytime メタデータにどう

やって効率的に変換するかという課題もあり、この点に関して、TV-Anytime フォーラムと Pro-MPEG フォーラムは情報交換を始め、その可能性や具体的な手法について検討を開始しようとしている。

## 6 むすび

本稿では、オーディオビジュアルコンテンツのためのメタデータ国際標準として、MPEG-7 と TV-Anytime メタデータについて概説した。

MPEG-7 は汎用的に作られた規格であるため、様々なメタデータを交換するためのピボットとしての利用が考えられるが、現在までにそのような動きは顕著化していない。一方で、映像や音声の特徴量を利用して、映像データベースや音声データベースでの検索方法として利用する方法はいくつか事例が報告されている。

これに対し、TV-Anytime フォーラムは、PDR を利用したコンテンツの利用という明確なアプリケーションを目標とし、MPEG-7 を基にして B2C で使用できる実用的なメタデータ規格を作り上げた。また、メタデータを交換するプロトコルも積極的に検討しており、近未来に予想されるリッチなメタデータサービスの基盤技術を提供していくことと思われる。

また、B2B においても MXF というメタデータ標準が今後広まり、オーディオビジュアルコンテンツが円滑に流通するためのメタデータ基盤が整いつつあると言うことができるだろう。

以上のように、オーディオビジュアルコンテンツ市場では、メタデータを使用したコンテンツ流通とコンテンツ利用が急速に広まることが予想され、ここで使われるメタデータ標準が、他の分野で検討あるいは使用されているメタデータ規格に大きな影響を及ぼすことが考えられる。

インターオペラビリティの困難なメタデータは利用価値が低いため、様々なメディア間やビジネス間でのメタデータ交換が今後大きな問題となってくると考えられる。この課題をどう克服するか

が、今後の焦点の一つとなってくるであろう。

## 参考文献

- [1] 奥村, 青柳, 篠原, “CDN と通信・放送融合”, 電子情報通信学会誌, Vol.86, No.3, pp.193-197 (2003 年 3 月)
- [2] (社) 電波産業会, “サーバ型放送における符号化、伝送及び蓄積制御方式標準規格”, 電波産業会標準規格, ARIB STD-B38 (2003 年 2 月)
- [3] <http://www.chiariglione.org/mpeg/index.htm>
- [4] <http://www.tv-anytime.org/>
- [5] <http://www.xrml.org/>
- [6] TV-Anytime Forum, “RMP Specification Drafting Process, Specification Workbook”, WD550 (2002 年 3 月)
- [7] ISO/IEC FDIS 21000-5, “Multimedia framework (MPEG-21). Part 5: Rights Expression Language”. (2003 年 8 月)
- [8] ISO/IEC FDIS 21000-6, “Multimedia framework (MPEG-21) Part 6: Rights Data Dictionary” (2003 年 8 月)
- [9] TV-Anytime Forum, “Call for Contributions for TV-Anytime Phase 2 Technologies on Sharing, Synchronization, Packaging, Targeting & Interactivity”, TV179r3 (2003 年 8 月)
- [10] TV-Anytime Forum, “TV-Anytime in a Connected World, A Vision Paper for TVA Phase 2”, TV191r2 (2003 年 8 月)
- [11] ETSI, “Broadcast and On-line Services: Search, select and rightful use of content on personal storage systems (“TV-Anytime Phase 1”)", ETSI TS 102 822-1, -2, -3-1, -3-2, -4, -6, -7 (2003 年 9 月予定)
- [12] <http://www.pro-mpeg.org/>
- [13] <http://www.aafassociation.org/>

## 電子カタログにおけるネイティブ XML データベース『Sonic XIS』の活用

プログレス ソフトウェア株式会社

武末 徹也

### 1. はじめに

B2B やコンテンツ管理、Web サービスなど、XML という技術は様々な分野で応用されはじめている。それに伴い、蓄積、管理、再利用すべき XML データの量も増加しており、専用のデータベースへのニーズが高まっている。

ここでは、ネイティブ XML データベースの代表的製品である SonicXIS のアーキテクチャ、機能、特徴を、RDBMS 等他のデータベースと比較することで説明する。また、SonicXIS の代表的なアプリケーションイメージの一つである電子カタログを例として、SonicXIS の適用シーンと SonicXIS 活用により得られるメリットについて説明する。

### 2. SonicXIS

SonicXIS は、米国プログレスソフト社のビジネスユニットであるソニックソフトウェア社により開発、販売されているネイティブ XML データベースである。1998 年に最初のバージョンが発表され、現在のバージョンは 3.1.3 となっている。

#### 2.1 SonicXIS の特徴

SonicXIS は、オブジェクト指向データベースである ObjectStore をデータベースのコアに持ち、その上に XML 処理エンジンや API、XML 関連標準の統合開発環境を搭載した製品である。以下に SonicXIS の特徴を示す。

##### (1) メモリデータベース

ObjectStore はデータを RDBMS のような表形式に正規化されたデータとして格納、利用するのではなく、プログラムで定義されたオブジェクトモデル、すなわちオブジェクトのインスタンスそのものとして格納する。この格納されたオブジェクトはメモリ上にマッピングされ、クライアントプログラムから利用されるため、ある意味メモリデータベースと言える。このため、データに対してメモリスピードでのアクセスが可能となり、ディスクアクセスがメインとなる他のデータベースと比較して、格段のハイパフォーマンスを提供することができる。

SonicXIS の場合、XML データをテキスト形式で格納するのではなく、W3C (World Wide Web Consortium) で規定されている XML 操作の標準インタフェースである DOM(Document Object Model) オブジェクトの形式で格納する(XML パーシスタンス)。これにより、複雑に階層化されたデータを、マッピングなどの構造変換手段を用いることなく、そのまま表現することができる。

また、SonicXIS はメモリ上の DOM オブジェクトを直接操作することができるので、データ構造そのものの更新処理(RDBMS で言うところのフィールドの増減やテーブルの増減など)を、サービスの中断なしにオンラインで行うことができる(ダイナミックモデリング)。これにより、構造変更の頻繁なデータの取り扱い、すなわちデータの拡張性を飛躍的に向上させることができる。構造変更は XML ドキュメントのレベルではなく、必要なノードのレベルで行うことができるため、プログラム処理に対するインパクトは非常に小さい。

SonicXIS では、DOM オブジェクトのハンドリングと XML 処理を行うプロセスを XML キャッシュと呼ぶ。XML キャッシュはメモリ上に展開された DOM オブジェクトと、XML 処理エンジン(DXE: Dynamic XML Engine)とで構成される。そのため、SonicXIS アプリケーションから見ると、XML キャッシュがデータベースそのものとなる。

SonicXIS ではすべてのデータベース処理を XML キャッシュを介して行う。SonicXIS では XML データの格納時点で DOM オブジェクトとするため、テキスト形式の XML データを毎回メモリ上に DOM オブジェクトとしてロード(パース)することなく、データベース上の DOM オブジェクトを直接操作することができ、メモリスピードでのリアルタイム処理が可能となる。

## (2) ロードバランスとスケーラビリティ

SonicXIS には、大量アクセスに対しての負荷分散とスケーラビリティを確保するための手段として、XML キャッシュプロセスを複数分散できる機能がある(分散 XML キャッシュ)。この機能は、ObjectStore の米国特許技術である、キャッシュフォワードアーキテクチャを継承しているものである。

SonicXIS では、データベースそのものの役割となる XML キャッシュを、複数起動することができ、さらに複数の XML キャッシュは、同一のデータベースサーバ機だけではなく、複数のマシンに分散することができる。これにより、データベースがアプリケーションパフォーマンスのボトルネックとなることを回避することができる。



(3) 高機能ツール群

SonicXIS には、Windows エクスプローラのようなルック&フィールの管理ツール「DXEManager」を持つ。

また、XML 関連標準を統一したビューで開発するための IDE(統合開発環境)である「StylusStudio」も標準でバンドルしている。

RDBMS 等、他のデータソースに対するデータ連携を行うツールとして、XConnects がある。XConnects はオプションである。

(4) サポートプラットフォーム

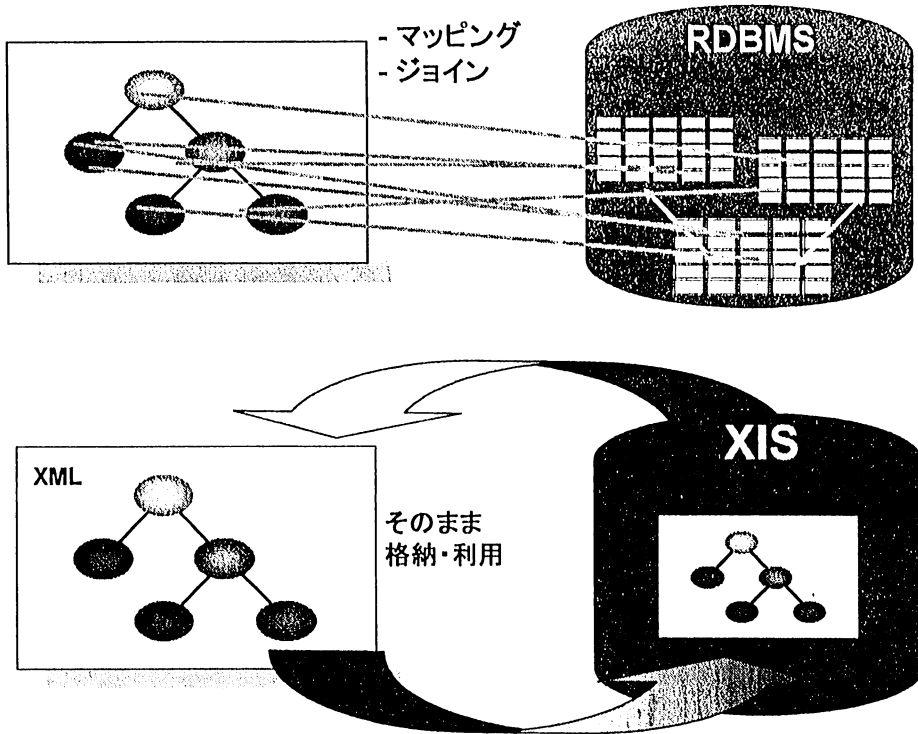
SonicXIS のサポートプラットフォームは、以下の通りである。

- Windows(NT/2000)
- Soralis
- HP-UX
- Linux

2.2 XML パーシスタンス

RDBMS の場合、XML データを利用しようとする、表形式のデータ構造と XML の階層構造とのマッピングや、複雑なテーブルジョインが必要になる。これは、テーブルジョインの多発によるパフォーマンスの低下だけでなく、表現したいデータモデルが変更された際に、毎回マッピングやプログラムのやり直し、影響範囲の調査など、開発効率の低下を招く。

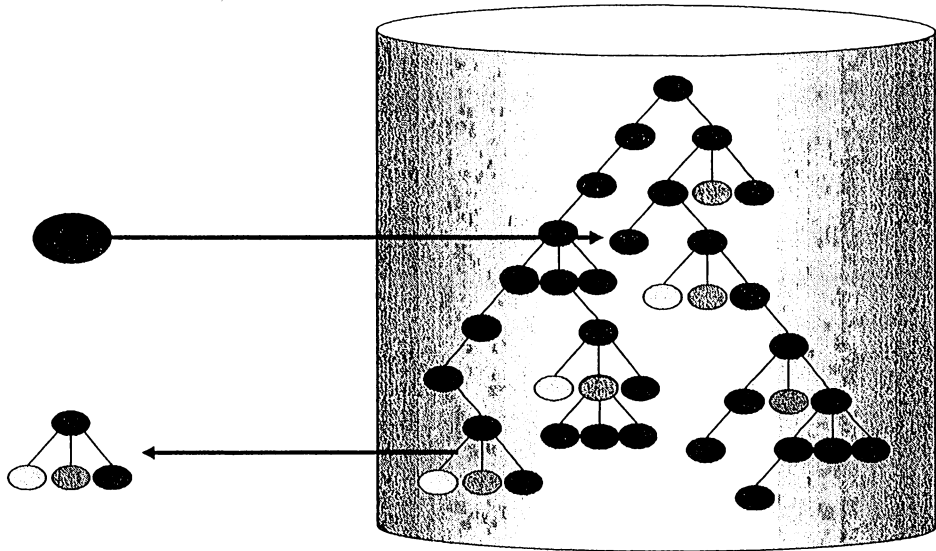
一方、SonicXIS の場合、XML の階層構造をそのまま格納することができるため、複雑なデータモデルをマッピングなしで格納することができ、さらにデータモデルへの変更にも柔軟に対応することができる。データアクセスに関しては、メモリスピードで高速アクセスができ、データへのアクセスごとに DOM パースが不要となる。



### 2.3 ダイナミックモデリング

SonicXIS の最大の特徴として、データモデルをオンライン/オンザフライで更新できるという特徴がある。ここで、オンラインとはシステムのサービスストップをすることなく、運用状態のままデータモデル更新ができることを言い、オンザフライとは、データベース上のデータをアプリケーションで更新し、再度データベースに上書きするような処理ではなく、直接更新を行うことができることを言う。

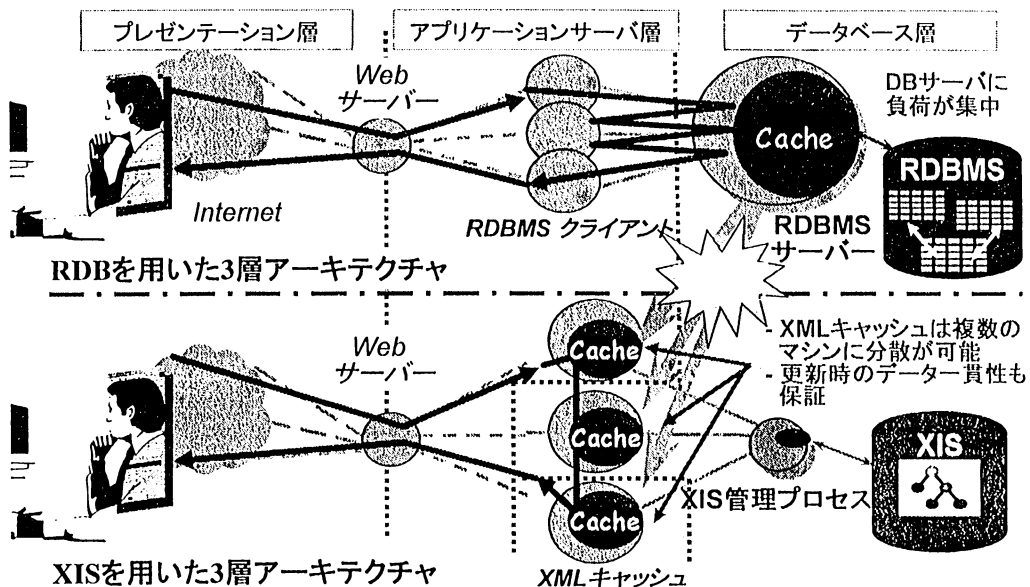
この特徴はデータそのものの拡張性を向上できるだけでなく、システム全体の可用性を向上できるという特徴を持つ。



## 2.4 分散 XML キャッシュ

SonicXIS は、高負荷システムへのスケーラビリティを確保するために、分散 XML キャッシュという負荷分散の機能を持つ。XML キャッシュをデータベースサーバ上、アプリケーションサーバ上、あるいはキャッシュ専用構築したサーバ上に複数個配置することができ、これによりデータベースサーバ機の CPU やメモリ等ハードウェアリソースを使うことなく、データベース処理を行うことができる。

分散した XML キャッシュ上のデータはトランザクションの一貫性が保証される。これは、ある XML キャッシュ上のデータに対して更新が発生した際に、同一リソースを担当する別のキャッシュに対して、メッセージを送ることにより実現する。データの更新が成功した場合、次に別の XML キャッシュへの読み出しアクセスがあると、アクセスされた XML キャッシュは、自身がキャッシュしている古いデータを捨て、再度データベースファイルから最新のデータをフェッチしなおすため、読み出されるデータは最新のものになる。



同一マシン上で分散 XML キャッシュを設定した場合は、クライアントからのアクセスに対して、担当が設定されたキャッシュに自動的にルーティングを行うことができる。

## 2.5 SonicXIS の API

SonicXIS の API は、大きく 2 種類に分かれる。1 つは DXE クライアント API と呼ばれる API で、クライアント上、例えばアプリケーションサーバ上に配置したクライアントモジュールが提供する API である。DXE クライアント API には Java API とマイクロソフト社の COM API がある。通常のデータベース操作は、この DXE クライアント API を使用して行う。

DXE クライアント API はクライアントモジュールと XML キャッシュとの間に CORBA による通信が発生する。この際にやりとりするデータのオーバーヘッドや、きめ細かいトランザクション制御を行いたいような場合には、サーバサイドの API を利用する。サーバサイド API は DXE サブレットと言い、Java API となる。DXE サブレットはいわゆる http サブレットではなく、サブレット API に準拠した処理モジュールで、XML キャッシュプロセスをコンテナとして起動、実行される。

## 2.6 SonicXIS のデータアクセス

SonicXIS のデータアクセス手段は、RDBMS のように SQL 言語を使用しない。検索手段は W3C の XML データのアドレッシング言語である XPath もしくは、XML への検索に関する次世代標準言語である XQuery(現在は Working Draft)を用いて行う。検索結果は XML として返却される。

更新は、XML ベースで書かれた独自の更新言語である XIS Updategram(XUG)を用いて行う。XUG は XPath でノードをアドレスし、更新内容(Insert/Replace/Delete)を指示することにより非常に簡潔に行うことができる。これにより、DOM インタフェースを使用せずに格納されている XML データに対する更新を 1 メソッドコールで行うことができる。もちろん、DOM インタフェースを用いた更新も可能である。この場合、サーバサイド API を使用して、XML キャッシュ上の DOM オブジェクトを直接編集することになる。

## 3. 電子カタログへの活用

上記のように、SonicXIS は複雑なデータモデルを持ち、構造変更が頻繁となるようなデータを管理するのに最適なデータベースである。そのようなケースで電子カタログは SonicXIS の特徴を最大限に利用することのできるアプリケーションの 1 つとなる。

例えば、社内購買システムのようなアプリケーションでは、文房具からパソコン、オフィス家具までを透過的に管理するようなイメージとなる。これはカテゴリが非常に複雑なツリーになることを意味する。あるカテゴリについては階層が深くなり、また、あるカテゴリでは浅い階層となる。このようなカテゴリを従来の RDBMS で管理するのは非常に困難になるが、SonicXIS では、カテゴリのツリーをそのまま格納することにより容易に管理することができる。当然、カテゴリツリーの構造変更にもオンラインで対応することができる。

また、各カテゴリで管理されるデータは管理項目がカテゴリごとに異なる。XML データで見ると、末端が微妙に異なるデータ構造となる。さらに、パソコンなど構造変更が頻繁なデータになると、サービスをストップしてデータのメンテナンスを行うことは通常データベースでは困難である。そこで、SonicXIS のダイナミックモデリング機能を使うことにより、そのようなケースに対して柔軟に対応することが容易となる。

さらに、ユーザごとに異なるビューを用意するような機能を実現したいような場合は、XML のワンソースマルチビューの機能を利用して、ユーザ独自のフィルタ(XSLT ファイル)をダイナミックに生成するような機能を実装することにより、最小限のコーディングで実現が可能となる。

電子カタログサイト等のSonicXISアプリケーションを構築するためのフレームワークとして、eCatalog Plus for XISがある。eCatalog Plus for XISは、Servlet、JSP(Java Server Pages)、JavaBeansを用いたMVCモデルで構成されるWebアプリケーション用途のPure Javaコンポーネント群で、SonicXISアプリケーション構築をクイックスタートできるようなサンプルコードである。

#### 4. まとめ

このようなSonicXISを用いたアプリケーションのアイディアは、電子カタログだけでなく、組織というカテゴリツリーと人それぞれのデータ項目を管理するヒューマンリソースマネジメントや、文書構造をモジュール化したドキュメント管理など、多数のソリューションを提供することができる。

SonicXISのアプリケーションを考える上で重要となるのは、既存の資産をどのように活かすか、ということも一つである。例えば、既存のRDBMSのシステムがあり、そこにXMLの特徴を活かした機能を追加するような場合、RDBMSをすべてリプレースするのではなく、RDBMSの長所とSonicXISの長所をそれぞれ活かしたハイブリッドな構成とすること、などである。例えばマスターデータや在庫データなどの固定的なスキーマを持つデータに関してはそのままRDBMSを用い、製品スペックなど、ダイナミックに構造が変わるようなデータに関してはSonicXISを作る、などのアプローチである。

このように、SonicXISは、XMLのメリットを最大限に活かし、従来のアプローチでは不可能だったアプリケーションを、機能だけではなく運用・管理のフェーズにまで提供できる唯一のデータベースであると言える。

## EsTerra による汎用システムとの連携 (活用事例)

株式会社メディアフュージョン

中村 真二 (nakamura@mediafusion.co.jp)

### 1 はじめに

W3C から XML1.0 が発表された当初は、電子商取引などが中心となり利用され始めた。当初、XML データベースは数える程しか発売されていなかったが、2000 年からは次第に Native XML データベース (NXD) が発表され、幾つかの利用事例が発表され始め、次第に XML に注目が集まり始めた、Yggdrasil もそんな中の 2000 年に製品化を行い、日本製の Native XML データベース (NXD) として注目を頂いた。

当初の XML の利用方法はデータ管理としての事例が多く、データ容量も余り大容量ではなかった。しかし、e-Japan 重点計画が政府から発表されると、申請文書、入札文書などに XML の利用が盛り込まれ、データ管理から文書管理へと XML の適用分野に広がりを見せてきた。当然、管理するデータも大容量となり NXD にも大容量のデータ管理機能が要求され始めてきた。

弊社においても、大容量のデータを扱えるように Yggdrasil のエンジン部分から見直し、大容量のデータ管理に向く製品 “EsTerra” を 2003 年 4 月に発表した。今回発表する内容はこの EsTerra を活用した事例を発表させて頂く。

### 2 XML の利用シーン

#### 2.1 導入が進むフロントオフィス・システム

XML を導入する際、比較的導入し易いのがフロントオフィス・システムではないだろうか。

既に、フォーム作成ツールなど多くの製品が XML 対応されており、Microsoft の Office2003 の XML 対応で、今後更なるフロントオフィス製品の XML 対応が進められていくものと考えられる。

フォーム作成ツールの XML 対応方法には 2 つのパターンに分けることが出来る、① CSV 形式の変わりに XML 形式を採用した製品。② XML の特性を生かした製品。である。

①に対応した製品には幾つかの問題がある、つまり、帳票の様式をベースにフォームを作成したため、ドキュメントの構造を余り重要視しなかったが為、既存のシステムとの連携が図れなかった事などが挙げられる。

その結果、XML を採用しても、期待以上の結果が見られず、XML の導入を諦めてしまった会社も多くあるのではないだろうか。

しかし、XML ドキュメントを解釈、バックオフィスで稼働している RDB を中心としたシステムにデータが登録できれば、柔軟な XML の特性を活かしたフロントオフィス・システムの構築が可能となる。最近、このようなシステムをニーズの増加とそれに  
応え構築した例がある。

## 2.2 実績の少ないバックオフィス・システム

バックオフィス・システムに XML を導入するのに各社は慎重になっている。

比較的導入しやすい分野は BtoB などの電子商取引の分野ではないだろうか。XML が普及をし始めた時期は、BtoB などのシステムが中心で XML は BtoB のための言語であるかのように雑誌や本などで紹介されていた。

インタフェースが異なる企業間連携には XML はマッチングしていた。しかし、それらのシステムの多くは、バックオフィス・システムとは別の位置付けで構築されているケースが多く、バックオフィス・システムと連携してシステムを稼働した例は殆ど聞く事が出来ない。

その背景には、XML への知識、技術者の不足や、登場して間もないことからの不当に低い NXD に対する信頼性や、認識されていないが故の安定性・高速性などへの不安が挙げられるが、一番は一度構築したシステムを再構築する事に対しては慎重に成らざるを得ないのは当然である。

## 3 EsTerra (NXD) と汎用業務システム (RDB) 連携

### 3.1 EsTerra を汎用業務システム (RDB) に利用する有効性

現在、新たなシステムや製品を導入して新システムを構築する際、どうしても考えなければならないのが XML 対応である。

電子政府では、電子入札や電子申請など多くの窓口業務に XML を採用している、全てを新しくシステムを構築するのであれば、最新の技術で構築する事は可能であるが、既存システムを稼働させながら、新しい技術を取り入れたシステムを構築する際には、既存システムとの連携を考慮する必要がある。

これは、一般企業でも同じ事が当てはまる、寧ろ一般企業の方が既存システムと連携（新規構築であっても、システム思想が既存システムと同様の場合も含む。）することを考慮して、新システムを構築する必要があるのである。

その際、考えられる方式としては、① NXD を導入して、既存システムとはインタフェースを統一する。② XML から直接、既存システムに登録する。が考えられる。SI 会社の特性によってもどちらを選択するかは違ってしまうが、今後のシステム拡張を



考慮すると①の方がベストであると思う。その理由としては、NXD を活用する理由である。DBMS を導入すると、そこにデータを蓄えるイメージが先行しがちだが、NXD はデータを蓄えるだけでなく、XML の構造を解釈するのにも便利なのである。つまり、NXD を DBMS として使う傍ら、RDB とのインタフェースとして利用するのである。

その際、XML の構造と RDB の構造をマッピングする定義ファイルを作成する必要があるが、定義ファイルさえ作成する事によりどのような XML ドキュメントが登録されても、正確に RDB へ登録する事が可能となるのである。

この様に既に稼動しているバックオフィス・システムを変更することなく、フロントオフィス・システムのみを XML 対応が容易に変更する事が可能となるのである。バックオフィス・システムを変更しないという事は、他のシステムに影響を与える事はないのである。

XML に対応している RDB は Well-Formed にも対応しているものもあるが、それはあくまで RDB に XML を格納することであり、既存の RDB スキーマに XML データを格納するわけではない。RDB スキーマに XML データを格納するためには、XML インスタンスを解析し、解析結果とマッピング定義によって、RDB スキーマに格納する必要がある。

弊社では、この様な異なる DBMS 間を連携するためのサブシステムとして“異種 DB 接続ツール”を開発し、既に導入している。

### 3.2 EAI ツールとの比較

“フロントオフィス・システム”を導入した時のメリットを“EAI ツールだけで処理する場合”と比較する。

- ・ データの 2 重化  
フロントにも DBMS の管理を持ったシステムを入れることで、バックアップが出来る (EAI は流すだけ)。
- ・ 安全性の向上  
EAI で直接接続するよりも、フロントシステムを入れることで、外部と完全に切り離せる。
- ・ 安定性の確保  
外部からのトランザクションの変動にも、フロントオフィス・システムが緩衝となり、直接基幹システムへの負荷が無い。
- ・ 外部変化への柔軟な対応

外部は、速い速度で変改が起こっている、取引についても、時代に合わせるには、変化が必要だが、フロントの変更だけで基幹への影響を最小限に対応できる、しかも早い。

また、DB 連携は、EAI ツールでも可能だが、弊社が提供している“異種 DB 接続ツール”の機能拡張で対応が可能である。本事例では、“異種 DB 接続ツール”の機能拡張によるデータマッピングを行っている。

### 3.3 フロントオフィス・システムと汎用システムを連携する際の問題点

フロントオフィス・システムで生成された XML データを汎用システム (RDB) と連携する際の問題点は、① XML データのデータタイプと RDB スキーマのデータ型の違い。② 1つの XML データを複数の RDB のテーブルに分散する。などが挙げられる。

XML データはテキストデータであるため、全て文字列 (マルチメディアデータなどは除く) である。例えばタグで数値と指定しても、その中身は文字列である。そのデータを RDB に格納する際には数値で処理しなければ、RDB へ登録時にエラーになってしまう。

このように、XML データの構造からノードタイプを洗い出し、ノードタイプに合う RDB 側のテーブル名やカラム名を定義する事により、スムーズに XML データを RDB にデータ移行が可能となるとともに、開発効率も向上する。

図 1 にマッピング処理の流れをしめす。

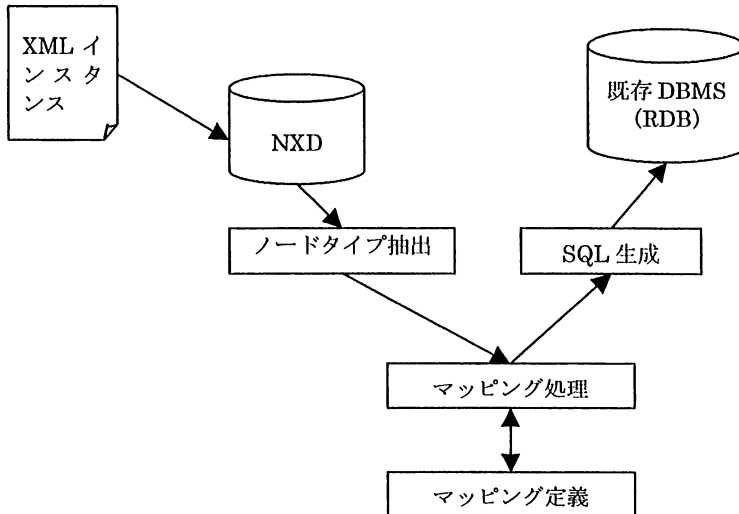


図 1 マッピング処理の流れ

### 3.4 懸念点・注意すべき点

マッピングには、XML ドキュメントでの構造解析が重要となる。XML は柔軟性があるから、どのような XML ドキュメントでも対応が可能であると思われるが、本システムでは矛盾が発生してしまう。

確かに、XML ドキュメントは柔軟に対応する事が出来るが、この中で書かれた内容は“汎用 DBMS に登録する”のである。つまり、データ型が限定しているのである。

例えば、納品日の場合、RDB では日付型を使う。この場合は必ず“2003.9.15”や“2003.9.23”と記述が決められるのである。これを XML で入力すると、様々な書き方をする。“2003年9月15日”や“9月23日”これならば、後処理で変換する事は可能だが、“極力早く”など日付ではなく、文書列で入力されているケースが実際にあった。これはフォームツールでのチェック処理が甘いと言えればそれまでだが、システム全体を把握しないで XML の柔軟性だけに着目してシステムを構築するとう言うケースは起きてしまう。

### 3.5 異種 DB 接続システムのメリット

異種 DB 接続システムを導入する際は、システムで処理するデータをよく検討する必要がある。

異種 DB 接続システムを導入する事により、すべて RDB でシステム構築するのは無く、XML との柔軟な組み合わせで、システムの最適化、安全性、安定性の向上、開発効率の向上、拡張性の向上などが図る事ができる。また、汎用的な EAI ツールは汎用的であるがゆえに、高コストである。本事例では異種 DB 接続ツールによる専用処理を開発する事で、コストダウンにもつながった。

## 4 XML データベースの今後の展望

NXD の活用場面は今度増加傾向にあると考えている。

今回の事例のように汎用システムとのミドルウェア（インタフェース）として利用価値や、NXD を DBMS としての機能として純粋に利用する、Web サービスが普及するとそこには必ずといっていいほど NXD は必要となってくる。

更に、XML はデータ交換のみならず、文書（特にマニュアル）での活用も多く使われ始めてきている。

マニュアル作成現場で求められる機能としては、“履歴管理”“差分管理”“ワークフロー”“高度な検索”“自動マニュアル生成”“画像管理”などが求められるのが一般的

である。更にデータ容量も増加傾向にあり、弊社も上記 XML 処理に最適化したアプリケーションを提供しています。

従来からの NXD はオブジェクト指向 DBMS の考えや高速性を考えメモリ上のみでデータ処理を行ってきたが、ディスク・インタフェースでのデータ処理が必要となり、さらに大容量データでも高速に処理が行われる必要がある。

EsTerra では、いち早くこの要求にもこたえるべく、テラバイト級（論理値 16TB 以上）の大容量 XML データにも対応。従来の高速性能とあいまって、より高度なシステム開発を可能にしています。

#### 4.1 フロントオフィス・システムのモバイルへの展開

モバイル機器（PDA・携帯電話）が充実してきて、フロントオフィス・システムをモバイル機器にも提供するシステムが増えてきている。その中に地図データを取り入れたいとの要求が高い。

地図データを使う業務には、防災関連、不動産、天気予報、エンターテイメントなど利用シーンは様々であるが、地図データそのまま表示するのではなく、地図データと付加情報を合成して表示する事が要求されてくる。

このような案件に対して弊社では、地図データに SVG 形式を採用した。

SVG データと POIX データを EsTerra (NXD) にて管理し、サーバ側にて地図データと POIX データを統合する方式で検証を行い、モバイル端末から表示したい地域と観光案内などのポイント情報を入力しリアルタイムで SVG データを生成し、配信を行う事が可能である。

弊社のデモサイトでは、地図データと POIX データを EsTerra で統合し SVG データを生成しただけという単純なものであるが、今後はモバイル端末から送られてくる写真データや位置データなどを並行して受信し、EsTerra にて管理し、本来のリアルタイムでの情報配信が可能となるであろう。この考え方は GPS 付きモバイル端末に拡張する事が可能となる。

また、様々なマルチメディア・コンテンツを時間軸や空間軸で統合配信を規格化している“SMIL”では、マルチメディアと属性情報を EsTerra で管理を行い、自分が知りたい情報や場面などを EsTerra に検索を行い、SMIL 対応プレイヤーで再生することが可能である、特に教育分野では既に多く使われており、今後はテレビの編集・制作や携帯電話での利用など応用分野に広がりを見せている。

これからの NXD は、XML データや XML ドキュメントを管理だけではなく、マルチメディア・コンテンツ管理など新たな管理要素が増えてくると考えている。弊社も

EsTerra を核にこれらのコンテンツが最適な環境で管理ができ、開発効率の向上、システムの拡張性の向上などを考慮して製品群の開発を進めている。

## 5 おわりに

汎用システムとして、NXD が利用可能なのかとの不安を抱いている人、企業が多くあると思います。しかし、全てを NXD で構築するのではなく、XML が適している処理やそのデータの特性などから判断する必要がある。

今回の事例ではまさに、適材適所のシステムであったと思う。フロントは構築がしやすい XML 対応のフォーム作成ツールを導入し、今後の項目の変更などに対して柔軟なシステムへと作り上げている。一方のサーバ側は、様々な既存システムと連携されており、容易に変更することが出来ない、したがって XML ドキュメントから既存の汎用システム (RDB) にデータを登録する際、XML ドキュメントのタグ単位で処理を行うことができる EsTerra をミドルウェアとして利用し、汎用システムにデータだけを登録する事が可能となったのである。

これにより、XML の導入がより一層進められる事を期待したい。

## Oracle XML DB の活用事例

作田淳子・有賀洋平／日本オラクル(株)

### 1. Oracle XMLDB の特徴

Oracle XML DB は、ネイティブ XML データベースの要件を満たしたものであるといえます。

また、Oracle XML DB では、通常の insert 文や select 文、または FTP、WebDAV 経由での操作で XML 文書の格納・取り出しが可能です。ユーザは変換・マッピングなどをする必要は一切ありません。

さらに Oracle XML DB では、XML Schema をサポートしています。XML Schema をデータベースに登録することで、XML DB に格納された XML 文書は自動的に分解され、データベース内部でオブジェクト型を使用した表に格納されます。XPath を使用した、XML 文書の構造に沿った検索が可能になります。XML 文書が XML Schema に基づいた内容になっているかどうかの検証も、データベース内部で自動的に行われます。

XML 文書のデータベースへの格納は、前述のようにオブジェクト型を使用した表に分解して格納することも可能ですが、単なる文字列として格納することも可能です。また、オラクルの最大の特徴は、今までのリレーショナルデータベースの項目属性で XML 機能を提供していることです。つまり、1つのデータベース内で、リレーショナルデータベースとネイティブ XML データベースの機能を共存、連携できます。システムを構築する上で、既存のデータと XML のデータの連携ができるので、開発コストの削減、システム導入費の削減になります。

また、すでに Oracle Database がもつ可用性、堅固なセキュリティ、スケーラビリティ、高い運用性もそのまま生かされます。

### 2. 活用事例

昨今、東証様、国税庁様等で採用され話題を集めております XBRL についても Oracle XMLDB をリポジトリとして利用できます。タクソノミーをオブジェクト型に登録しインスタンス文書を格納すると、データベース内部でオブジェクト型を使用した表に格納されますので、ピンポイントの検索、高速な処理が可能となります。すでに事例としましても、XML Cities 社におけるパッケージ製品の格納先、投稿された論文をオンラインで XML フォーマットに変換、格納し、公開するために利用した事例、Dell Computer において発注システムのデータ記述に XML を利用し、そのデータの格納先とした事例等、さまざまな事例をご紹介します。

### Oracle XML DBの機能について

Oracle XML DBは、ネイティブXMLデータベースの要件を満たしたものであるといえます。Oracle XML DBでは、通常のInsert文やselect文、またはFTP、WebDAV経由での操作でXML文書の格納・取り出しが可能です。ユーザは変換・マッピングなどをする必要は一切ありません。

Oracle XML DBでは、XML Schemaをサポートしています。XML Schemaをデータベースに登録することで、XML DBに格納されたXML文書は自動的に分解され、データベース内部でオブジェクト型を使用した表に格納されます。XPathを使用した、XML文書の構造に沿った検索が可能になります。XML文書がXML Schemaに基づいた内容になっているかどうかの検証も、データベース内部で自動的に行われます。

XML文書は前述のようにオブジェクト型を使用した表に分解して格納することも可能ですが、単なる文字列として格納することも可能です。

また、オラクルの最大の特徴は、今までのリレーショナルデータベースの項目属性でXML機能を提供していることです。つまり、1つのデータベース内で、リレーショナルデータベースとネイティブXMLデータベースの機能を共存、連携できます。

今回は、機能詳細としてXML文書の格納、取り出し、RDBデータとの融合について詳細を述べさせていただきます。またその活用事例をご紹介します。

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

### XMLデータベースの活用事例

日本オラクル株式会社  
アドバンスソリューション本部  
作田 淳子

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

### Agenda

#### ⇒ ネイティブXMLデータベースとは

- Oracle XML DB 機能詳細
  - 各種標準への対応
  - XML文書の格納
  - XML文書の取り出し・操作
  - XML文書とOracleデータの融合
- 活用事例
  - XBRL
  - その他事例

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

3

### ネイティブ XML データベースとは NXD: Native XML Database

- XML文書を明示的な変換・マッピング・操作をすることなく、単純にXML文書として格納・取り出しできる
  - アプリケーションによる解析や変換を行う必要がない
- XML文書を論理的なモデルとして定義し、格納し、検索することができる。要素、属性、PCDATA、ドキュメント順序を最低限サポートしなければならない。
  - XML SchemaによるXML文書構造の定義
  - XPathによる文書内のデータへのアクセス
- 物理的な格納モデルに制限はない
  - 階層型ツリー
  - オブジェクト型
  - バイナリデータ

参考: DB Initiative - Native XML Databaseの定義  
<http://www.xml-db.org/faq.html>

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

4

### Oracle XML DB と NXD の関係

- XML文書を明示的な変換・マッピング・操作をすることなく、単純にXML文書として格納・取り出しできる
  - 従来の(Oracle9i Release 1で実装したXMLType)に対する操作のSQLによる操作に加え、FTPやWebDAVによる文書の格納・取り出しが可能。
- XML文書を論理的なモデルとして定義し、格納し、検索することができる。要素、属性、PCDATA、ドキュメント順序を最低限サポートしなければならない。
  - XML Schemaに対応した構造化マッピングのXMLTypeデータベース
  - 内蔵されたSchema ProcessorによるXML文書の自動検証
  - XPathなどを含む柔軟なSQL操作
- 物理的な格納モデルに制限はない
  - 以下の2つの格納モデルを実装
    - 構造化マッピング
    - テキスト形式(CLOB型)



ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

6

### Oracle XML DB のアドバンテージ

- リポジトリの実装
  - XML文書をファイルシステムとして格納・管理できる。
- 既存の表データとの結合
  - SQL文によるXML文書の操作を実装しているため、既存のOracleデータベースに格納された表データの結合検索を行うことができる。
- XML文書に対する整合性の実現
  - XML文書の要素、属性に対して一意制約や外部参照制約(参照される表は従来のOracleデータベースの表をサポート)を付与することができる。
- 高速な全文検索
  - XML文書にテキスト索引を付与することにより、Oracle Textによる高速な全文検索を行うことができる。
- 容易なメンテナンス
  - 従来のOracleデータと同様のバックアップやリカバリにも対応

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

6

### Agenda

- ネイティブXMLデータベースとは
- Oracle XML DB 機能詳細
  - 各種標準への対応
  - XML文書の格納
  - XML文書の取り出し・操作
  - XML文書とOracleデータの融合
- 活用事例
  - XBRL
  - その他事例

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 7

### Oracle XML DB でサポートされている標準

- Oracle XML DBは以下のようなXMLに関連するあらゆる標準に対応しています。
  - XML Schema: W3CのXML Schema1.0勧告
  - XPath: W3CのXPath1.0勧告
  - XSLT: W3CのXSL1.0勧告
  - XML DOM: W3CのDOMレベル1.0および2.0コア勧告
  - SQLX関数: XML関連仕様 (SQL/XML) のISO ANSI草案 (ISO/IEC 9075 Part14 and ANSI)

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 8

### Agenda

- ネイティブXMLデータベースとは
- Oracle XML DB 機能詳細
  - 各種標準への対応
  - XML文書の格納
  - XML文書の取り出し・操作
  - XML文書とOracleデータの融合
- 活用事例
  - XBRL
  - その他事例

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 9

### Oracle XML DB Repository

- データベース内に構築された仮想的なファイルシステム
- XML文書を効率よく管理することを目的
- 以下の機能をサポート
  - フォルダリング
    - パスによる管理
    - 様々な手段でコンテンツにアクセス可能
  - バージョニング
    - 変更履歴の保存
  - ACL
    - アクセス制御の管理

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 10

### Oracle XML DB Repository が実現する多様なアクセス方法

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 11

### Oracle XML DB の XML 文書の格納方法

- Oracle XML DBでの格納方法
  - XMLType (CLOB型)
    - XML文書として文法的に正しく記述されている文書(Well-formed)を格納
    - XML Schemaの定義は必須ではないので、要素や属性の変更などにも柔軟に対応できる
    - タグを含めすべてのテキストが格納される(容量大)
  - XMLType(構造化マッピング)
    - 特定のXML Schemaに基づいた文書(Valid)を格納
    - XML Schemaに合わせたオブジェクト型として格納されているので、Schemaの変更が発生するようなXML文書には向いていない
    - 一部の要素・属性のみの操作が可能のため、DML処理のパフォーマンスがよい
    - Schemaの定義はXSDファイルに基づき内部的に格納されているので、個々のXML文書はタグやスペースを挟いだ値の部分のみが格納される(容量小)

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 12



### 構造化マッピングとは

- XML文書を細分化し、データのみをオブジェクト型の属性にマッピングして格納する機能

XMLスキーマ → DBMS\_XMLSCHEMA.registerSchema() → オブジェクト型  
 オブジェクト型が使用される  
 XML文書 → CREATE TABLE → XMLType列/XMLType表  
 INSERT

ORACLE  
Copyright © Oracle Corporation Japan, 2003. All rights reserved. 13

### 格納方法の違い CLOB型 vs 構造化マッピング

| CLOB型                                                                                                                                    | 構造化マッピング |            |          |
|------------------------------------------------------------------------------------------------------------------------------------------|----------|------------|----------|
|                                                                                                                                          | (NUMBER) | (VARCHAR2) | (NUMBER) |
| <pre>&lt;emp&gt;   &lt;empno&gt;1100&lt;/empno&gt;   &lt;ename&gt;SCOTT&lt;/ename&gt;   &lt;sal&gt;300000&lt;/sal&gt; &lt;/emp&gt;</pre> | 1100     | SCOTT      | 300000   |
| <pre>&lt;emp&gt;   &lt;empno&gt;1200&lt;/empno&gt;   &lt;ename&gt;SMITH&lt;/ename&gt;   &lt;sal&gt;350000&lt;/sal&gt; &lt;/emp&gt;</pre> | 1200     | SMITH      | 350000   |

ORACLE  
Copyright © Oracle Corporation Japan, 2003. All rights reserved. 14

### XMLTypeのクエリー・リライト

- XPathを使用した問合せは、構造化マッピングの該当するオブジェクトに直接アクセスするドット表記の問合せにリライトされる

リライトされたSQL文

```
SELECT e.xmlDoc.XMLData.ename
FROM emp_clob_e;
```

ORACLE  
Copyright © Oracle Corporation Japan, 2003. All rights reserved. 15

### 検索のパフォーマンスの違い CLOB型 vs 構造化マッピング

全部を読み込み、解析 (Parse) する (CLOB型)  
 オブジェクト型  
 一部を読み込む  
 Oracle要素を検索 (/emp/ename)  
 メモリ

ORACLE  
Copyright © Oracle Corporation Japan, 2003. All rights reserved. 16

### Agenda

- ネイティブXMLデータベースとは
- Oracle XML DB 機能詳細
  - 各種標準への対応
  - XML文書の格納
  - XML文書の取り出し・操作
  - XML文書とOracleデータの融合
- 活用事例
  - XIBRL
  - その他事例

ORACLE  
Copyright © Oracle Corporation Japan, 2003. All rights reserved. 17

### URL 指定による XML 文書へのアクセス

<http://localhost:8080/home/SCOTT/purchaseOrders/1999/APR/ADAMS-20011127121040988PST.xml>

ORACLE  
Copyright © Oracle Corporation Japan, 2003. All rights reserved. 18

### リレーショナルデータの XML 文書化

http://localhost:8080/oradb/SCOTT/EMP

| EMPNO | ENAME  | JOB      | MGR  | ... |
|-------|--------|----------|------|-----|
| 7369  | SMITH  | CLERK    | 7902 | ... |
| 7499  | ALLEN  | SALESMAN | 7698 | ... |
| 7521  | WARD   | SALESMAN | 7698 | ... |
| 7566  | JONES  | MANAGER  | 7839 | ... |
| 7654  | MARTIN | SALESMAN | 7698 | ... |
| 7698  | BLAKE  | MANAGER  | 7839 | ... |
| 7782  | CLARK  | MANAGER  | 7839 | ... |
| 7788  | SCOTT  | ANALYST  | 7566 | ... |

```

<?xml version="1.0"?
<EMP>
  <ROW>
    <EMPNO>7369</EMPNO>
    <ENAME>SMITH</ENAME>
    <JOB>CLERK</JOB>
    <MGR>7902</MGR>
    <HIREDATE>17-DEC-80</HIREDATE>
    <SAL>800</SAL>
    <DEPTNO>20</DEPTNO>
  </ROW>
  <ROW>
    <EMPNO>7499</EMPNO>
    <ENAME>ALLEN</ENAME>
    <JOB>SALESMAN</JOB>
    <MGR>7698</MGR>
  </ROW>
  <ROW>
    <EMPNO>7521</EMPNO>
    <ENAME>WARD</ENAME>
    <JOB>SALESMAN</JOB>
    <MGR>7698</MGR>
  </ROW>
  <ROW>
    <EMPNO>7566</EMPNO>
    <ENAME>JONES</ENAME>
    <JOB>MANAGER</JOB>
    <MGR>7839</MGR>
  </ROW>
  <ROW>
    <EMPNO>7654</EMPNO>
    <ENAME>MARTIN</ENAME>
    <JOB>SALESMAN</JOB>
    <MGR>7698</MGR>
  </ROW>
  <ROW>
    <EMPNO>7698</EMPNO>
    <ENAME>BLAKE</ENAME>
    <JOB>MANAGER</JOB>
    <MGR>7839</MGR>
  </ROW>
  <ROW>
    <EMPNO>7782</EMPNO>
    <ENAME>CLARK</ENAME>
    <JOB>MANAGER</JOB>
    <MGR>7839</MGR>
  </ROW>
  <ROW>
    <EMPNO>7788</EMPNO>
    <ENAME>SCOTT</ENAME>
    <JOB>ANALYST</JOB>
    <MGR>7566</MGR>
  </ROW>
</EMP>
    
```

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 19

### XPath 式による XML 文書へのアクセス

http://localhost:8080/oradb/SCOTT/PURCHASEORDER /ROW/PurchaseOrder[Reference="ADAMS-20011127121040988PST"]?contentType=text/xml

```

<PurchaseOrder>
  <Header>
    <PurchaseOrderID>ADAMS-20011127121040988PST</PurchaseOrderID>
    <PurchaseOrderDate>2001-11-27</PurchaseOrderDate>
    <PurchaseOrderType>STANDARD</PurchaseOrderType>
  </Header>
  <Items>
    <Item>
      <ItemID>1</ItemID>
      <ItemDescription>IBM PC</ItemDescription>
      <ItemQuantity>1</ItemQuantity>
      <ItemUnitPrice>1000</ItemUnitPrice>
      <ItemTotalPrice>1000</ItemTotalPrice>
    </Item>
  </Items>
  <Shipping>
    <ShipToAddress>
      <AddressLine1>123 Main St</AddressLine1>
      <AddressLine2>Springfield, MA 01103</AddressLine2>
      <City>Springfield</City>
      <State>MA</State>
      <Country>USA</Country>
    </ShipToAddress>
  </Shipping>
</PurchaseOrder>
    
```

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 20

### スタイルシートの適応 (XSLT Processor)

http://localhost:8080/oradb/SCOTT/EMP?transform=/public/emp.xsl&contentType=text/html

スタイルシート

| EMPNO | ENAME  | JOB      | MGR  | ... |
|-------|--------|----------|------|-----|
| 7369  | SMITH  | CLERK    | 7902 | ... |
| 7499  | ALLEN  | SALESMAN | 7698 | ... |
| 7521  | WARD   | SALESMAN | 7698 | ... |
| 7566  | JONES  | MANAGER  | 7839 | ... |
| 7654  | MARTIN | SALESMAN | 7698 | ... |
| 7698  | BLAKE  | MANAGER  | 7839 | ... |
| 7782  | CLARK  | MANAGER  | 7839 | ... |
| 7788  | SCOTT  | ANALYST  | 7566 | ... |

```

empno ename sal
-----
7300 SMTH 800
7400 ALLEN 1600
7521 WARD 1250
7560 JONES 2075
7654 MARTIN 1250
7698 BLAKE 2850
7782 CLARK 2450
7700 SCOTT 3000
    
```

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 21

### SQL で操作するための XML 関数群

- XML文書の更新: UpdateXML()
  - XMLTypeインスタンス(Oracle XML DB に格納されたXML文書)の特定の属性値や要素値をXPathシンタクスに基づき変更することができる
  - 構造化マッピングの場合はデータベース内部のメモリーにDOM展開する必要が無いので、高速な処理を行うことが可能
- SQLX 関数
  - 複雑な構造のXML文書をSQL文で作成

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 22

### XMLElement 関数

- XML文書の要素を生成
- 入れ子構造をもったXML要素も生成可能
- XMLAttributes句で属性名を付与

XMLElement(要素名, [XMLAttributes(列名[, as 別名]), (子ノード)...])

```

SELECT XMLElement('Emp',
  XMLAttributes(empno as 'EmpNo'),
  XMLElement('Name', ename),
  XMLElement('SALARY', sal)
)
FROM EMP WHERE empno = 7369;
    
```

```

EmpNo="7369"
<Name>SMITH</Name>
<SALARY>800</SALARY>
    
```

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 23

### Agenda

- ネイティブXMLデータベースとは
- Oracle XML DB 機能詳細
  - 各種標準への対応
  - XML文書の格納
  - XML文書の取り出し・操作
  - XML文書とOracleデータの融合
- 活用事例
  - XBRL
  - その他事例

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 24

## XMLType ビュー

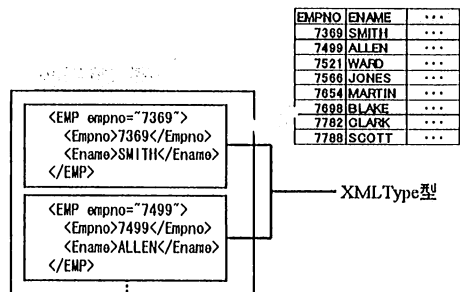
- XMLTypeをオブジェクトに持つビュー
  - 既存の表に格納されたデータをXML文書として表示
- XMLTypeビューに対する検索およびDML操作が可能
  - XMLTypeビューに対するDML操作で、元表を操作可能
- XMLスキーマに対応
  - XMLスキーマによりデータの妥当性を自動的に検証

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

26

## XMLType ビュー



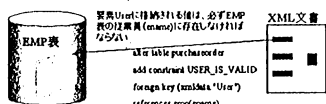
ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

26

## XML 文書に対する整合性

- Oracle XML DBはXML文書の特定の要素に対する一意制約や既存のデータとの外部参照整合性を制約として追加することが可能
- 既存のシステムで利用していたアプリケーションデザインをそのままXMLデータベースの世界に持ち込むことができる。



ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

27

## 格納されるXML文書に対する制約

- デフォルト表(XMLType データ型による特別なオブジェクト表)に付与された制約は、特別なアプリケーションを作成することなく、SQL/FTP/WebDAVといった格納時のプロトコルを問わずXMLスキーマ検証と合わせて自動的に実行される
  - 堅牢なXMLデータベースを実現するための必要なアプリケーションがデータベース内に格納され自動実行されているのと同じ

少ないコストで企業システムを支える  
XMLデータベースシステムを構築できる

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

28

## XML Schema 定義による制限と外部参照

- XML Schema定義ファイルによる値の制限
  - データ型(NumberやCHAR)
  - データ形式(混合型や記号コードなど)
    - ただし、営業所コードなど更新や追加しやすい値をXML Schema定義で制限していた場合は、XML Schemaの再設計およびそれに基づき文書の再検証が必要になる。
- Oracle XML DBの外部参照による値の制限
  - 従業員コードや営業所コードなどの値
  - マスターディテール表といった親子表
    - 変更・追加が発生した場合でもXML Schemaの再定義は必要ない

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

29

## Oracle Text による全文検索

- Oracle Textを使用して全文検索を行うために必要となるテキスト索引をXML文書に対して作成することができる
- キーワードを含むような要素にText索引を付与することでさらに高速で柔軟な検索を行うことができる

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

30

## Agenda

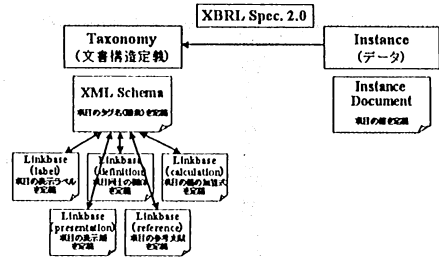
- ネイティブXMLデータベースとは
- Oracle XML DB 機能詳細
  - 各種標準への対応
  - XML文書の格納
  - XML文書の取り出し・操作
  - XML文書とOracleデータの融合
- 活用事例
  - ➡ - XBRL
  - その他事例

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

31

## XBRLの構成



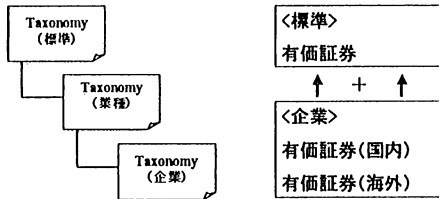
ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

32

## Taxonomyの拡張の例

XML SchemaのImport要素により、Taxonomyを拡張することができる。  
 →標準で策定されているタグの他にタグを追加することができる。



ORACLE

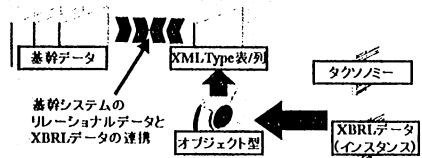
Copyright © Oracle Corporation Japan, 2003. All rights reserved.

33

## XBRLによるXMLDBの活用

オラクルのアドバンテージ:

既存する基幹システムのデータとXBRLデータの連携を実現することができる。



ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

34

## Agenda

- ネイティブXMLデータベースとは
- Oracle XML DB 機能詳細
  - 各種標準への対応
  - XML文書の格納
  - XML文書の取り出し・操作
  - XML文書とOracleデータの融合
- 活用事例
  - ➡ - XBRL
  - その他事例

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

35

## XMLCities, Inc.



### • 会社概要

- XML を使用したコンテンツを自動的に変換やインターネット上での公開を行うための製品やソリューションを提供する会社
- <http://www.xmlcities.com>

### • Oracle XML DB 利用方法

- パッケージ製品「XML Capture Suite」での XML データの格納先、変換および変換したデータの保存先として利用

### • Oracle XML DB のメリット

- 大量データの格納
- 大量データの変換

### • 参照先

- <http://www.xmlcities.com/xc2.asp?pres/presy031902>

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved.

36

### BioMed Central Ltd.

( )  
BioMed Central

- 会社概要
  - 生物医学研究論文に関する雑誌・書籍などの出版および電子出版  
- <http://www.biomedcentral.com/>
- Oracle XML DB 利用方法
  - 投稿された論文をオンラインで XML フォーマットに変換、格納し、公開するために利用

- Oracle XML DB のメリット
  - Oracle XML DB の多くの拡張された機能
  - RAC と組み合わせた高可用性

- 参照先
  - <http://www.oracle.com/customers/profiles/PROFILE9085.HTML>
  - [http://www.biomedcentral.com/info/authors/startajournal\\_jp](http://www.biomedcentral.com/info/authors/startajournal_jp)

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 37

### エルゼビア・サイエンス (Elsevier Science)

- 会社概要
  - 科学・技術・医学・社会科学分野に関する研究情報の雑誌・書籍などでの出版および電子出版  
- <http://japan.elsevier.com/index.html>
- Oracle XML DB 利用方法
  - 出版用データのコンテンツやメタデータとして XML を使用し、そのデータの格納先として利用

- Oracle XML DB のメリット
  - SQL および XPath による検索
  - ファイルシステムへの格納と比較した場合の堅牢さ

- 参照先
  - [http://www.oracle.com/features/ow/index.html?0626\\_ow\\_dbcust.html](http://www.oracle.com/features/ow/index.html?0626_ow_dbcust.html)

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 39

### Dell Computer Corporation

**DELL**

- 会社概要
  - 世界一の PC 製造会社  
- <http://www.dell.com/>
- Oracle XML DB 利用方法
  - 免注システムのデータ記述に XML を利用し、そのデータの格納先として

- Oracle XML DB のメリット
  - SQL と XML の両立

- 参照先
  - [http://www.oracle.com/pls/ooow/ooow\\_user.show\\_public?p\\_event=13&p\\_type=session&p\\_session\\_id=38275](http://www.oracle.com/pls/ooow/ooow_user.show_public?p_event=13&p_type=session&p_session_id=38275)

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 39

### Oracle XML DB は、、、

- ネイティブXMLデータベースとして以下のような機能を実現します

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 40

### Oracle XML DB は、、、

- SQLとXMLの両方に透過的な環境を提供します

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 41

### Oracle XML DB は、、、

## Databaseはひとつで良いんです！

ORACLE

Copyright © Oracle Corporation Japan, 2003. All rights reserved. 42

## 土木施工における XML の活用 —情報を活用した施工管理とライフサイクルマネージメントへの展開—

(株) 大林組 古屋 弘

### 1. はじめに

我が国の土木施工における IT の活用は、先端的な大規模工事の一部に留まっている。これまでの土木工事における情報化のアプローチはデータ計測や品質管理、また GPS をはじめとする測量機器やセンサーの高度化に付随する施工管理等が主体で、取得された情報の施工プロセスを超えた横断的な活用の試みはあまり事例がなかった。また、近年はコンピュータの現場への普及と高性能化、および CALS/EC の推進に伴い、各種データの電子化も大幅に進みつつある。このような環境で、以前から多くの土木技術者が、測量・設計・施工管理データを活用することで、様々な業務の効率化が可能になることを指摘しているが、いまだに多くの現場では、依然として膨大な工事関係書類の受渡しが必要で、帳票や図面の作成を人手作業に依存している非効率な実態がある。加えて、これらの電子データは、定義や構造等について多種多様な形態で存在するため、多目的利用するためには、利用者間で共通化をはかり、無駄な運用コストを減らしていく必要がある。特に土木工事においては、国や自治体が測量や監督業務でこれらのデータを扱うことで業務改善できる可能性があることから、施工業者と発注者は共にデータの有効利用を検討して行く必要があることは各方面で検討され始めている。この時に有用な情報交換ツールとなるものが XML 技術であり、新しい情報技術はこれらを進めるために欠かせない技術となる。

本報文では、土木施工における XML の活用例と3次元データ利用に関する新しい試みについての研究アプローチ、および現場実装例を紹介する。

### 2. 土木建設プロジェクトにおける問題点

建設工事における施工管理は、工事管理の中で安全管理や原価管理等とともに重要な管理であり、その管理項目としては、①工程管理、②品質管理、③出来型管理、④現場計測による盛土の沈下・安定管理がある<sup>1)</sup>。この中で②の品質管理は、構造物の本来保有すべき性能を完成後において保証するために重要な管理項目であり、建設工事においても ISO2394 (構造物の信頼性に関する一般原則) に基づく信頼性設計法が適用されつつある<sup>2)</sup>。なかでも建築基準法<sup>3)</sup>や土木学会のコンクリート標準示方書では性能規定の方向をいち早く取り入れた施工管理を行いつつある。

このような社会の流れの中で、土木工事においても性能規定による構造物発注は必然的となりつつあるが、土木工事は単品受注で現地生産であるという特性から、完成後の製品(構造物)の性能を支配する品質管理に以下のような難しさがある。

- ① 構造物全数量の管理は困難である
- ② 施工途中のプロセス管理によらざるを得ない
- ③ 施工のやり直しが困難である

さらに、土工事に関しては、一定の物性を持った材料を経済的に入手することは困難であり、土の強度や変形特性を経済的にコントロールすることが困難であることに起因し、

- ④ 室内試験や事前調査等との対応が不明確

という不確定要素も加わり、従来最低限の性能を満足する仕様規定を施工管理手法としていた。これは、土

構造物を対象とした工事における品質管理を、事前の調査結果や試験施工をもとに管理値を設け、それに従って施工を行い、施工終了後にサンプル試験を実施するという方法である。このような管理手法では試験およびその結果を出すまでに時間がかかることが多く、その間に施工が進んでしまい結果をフィードバックすることができない場合もしばしばあり、必ずしも構造物の品質を保証するものではないといった問題点もあった。

さらに、工事における施工管理は、上記の品質管理のみではなく、前述のような安全管理をはじめとする種々の管理項目があり、この管理を行うための多くの情報が施工プロジェクトの中で発生し、利用されている。すなわち、発注者側から受け渡された設計図書(設計図ならびに仕様書等)をもとに施工計画を作成し、施工を行うわけであるが、施工中はこれに伴う報告書や管理図書を作成し、監督検査を受けるのが一般的である。これを発注者側と受注者側の施工プロセスにおける情報量という観点から模式的に示したものが図-1である。施工に伴い発生する情報量は、建設プロジェクトの中で施工時に受注者側でピークとなり、この部分の情報化を行うことにより業務の効率化を行うことができ、建設工事というプロジェクトの中で、ワークフローの改善できる余地があるものとする。しかし、実際には情報化はスムーズに行われているとは言い難く、特に施工中の管理図書や発注者への報告書等を作成するためだけの情報や作業も多数存在し、無駄な労力(コスト)をかけていることになる。これは、測量機器や計測器、コンピュータが高性能化したにもかかわらず、これらの特性を生かさず、旧態依然の施工管理を行っていることに起因すると思われる。例えば、各種図面や報告書は電子化されているが、これらは施工中に得られた情報がある出力形式に則ってコンピュータを用いて転写しているにすぎず、無駄な労力(コスト)がここにつぎ込まれていることになる。

### 3. 土工事を例とした情報の連携

このように、土木工事の施工における様々なデータは、設計図書から発生し、施工のプロセスで参照され、様々に活用される。この流れを土木工事の中で筆者が主として担当している盛土工事に特化して模式図で示したものが図-2である。ここで、図の中心は施工プロセスを示しており、入力(発注者側から提供される情報)は出力(施工者が発注者へ提出する情報)である。このような施工の流れの中で、近年のコンピュータの進歩とCALS/ECの推進から、かなりの部分の電子化が行われつつある。

しかし、この電子化に関しては、「施工の合理化をはかる」目的とともに、情報の流通を円滑にし、「コンピュータや通信技術などを活用した合理的な生産システムの確立」することによる省力化を目的として現場にも積極的に導入されているが、他産業の状況と比較したとき、土木工事において次のような問題があることがわかる。

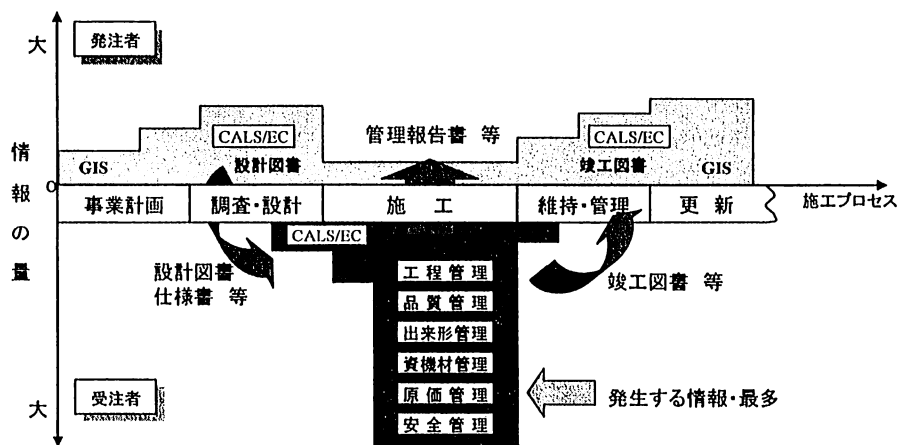


図-1 施工プロセスの中で発生する情報量の変化(模式図)

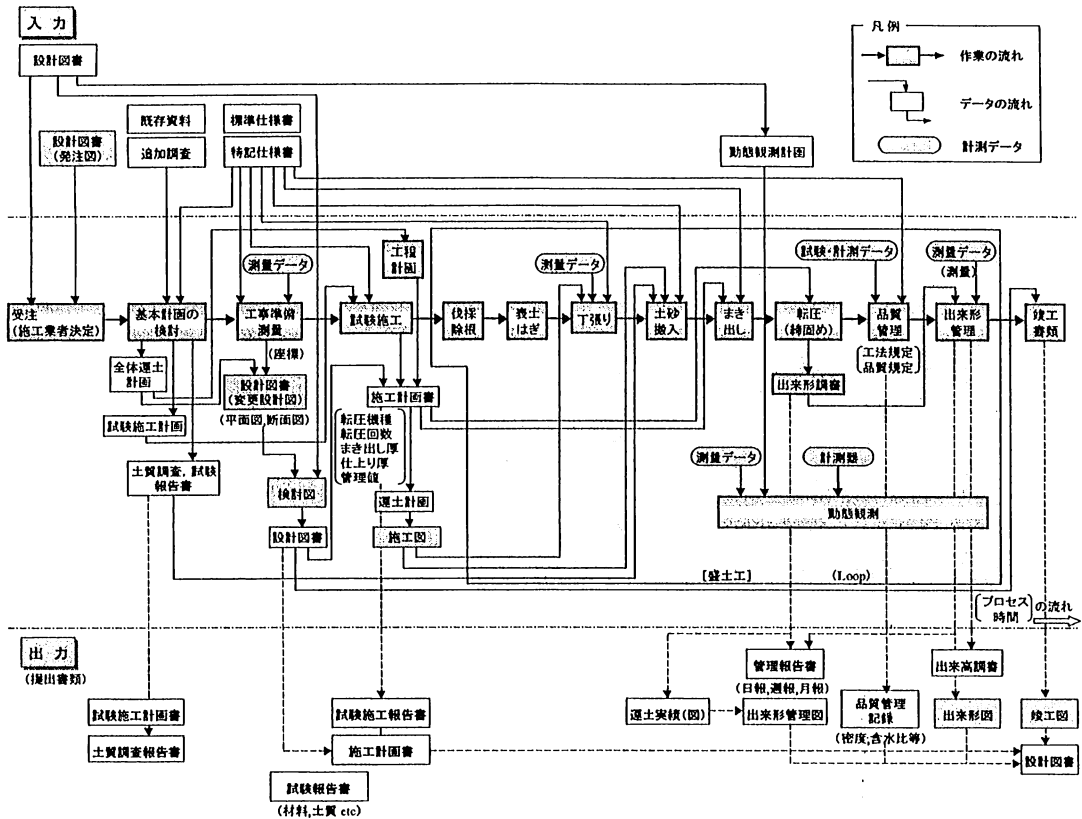


図-2 盛土工事の施工プロセスと情報連携(模式図)

製造業等では CAD 等の異種ソフトウェア間でのコンカレント作業を可能とするデータの運用ルールが整備されているのに対し、土木分野では、設計と施工・施工管理の間で、共通のデータを再利用することを想定した電子納品の標準化が現在のところ考えられていない。土木工事を1つのプロジェクトととらえた場合構造物完成までに多くのプロセスがあるが、各プロセスで構造物の機能を満足させるための管理が行なわれており、この管理において取り扱うデータは一過性のものが多いため、共通データを使う利点は発生しないためであると考えられる。こうした背景もあり、データの共用・再利用は進んでいない。しかし、図-1, 2で示すように、情報(データ)が共有できるとするならば、情報共有のメリットを各プロセスで享受でき、施工の合理化や新しい電子発注・納品の方法への展開が期待できるものとする。

#### 4. 土木施工におけるXMLの活用事例

このような状況で、各方面でDMや情報交換ルールが検討されつつあり、筆者らも数々のアプローチを試みている。以下にその事例を示す。

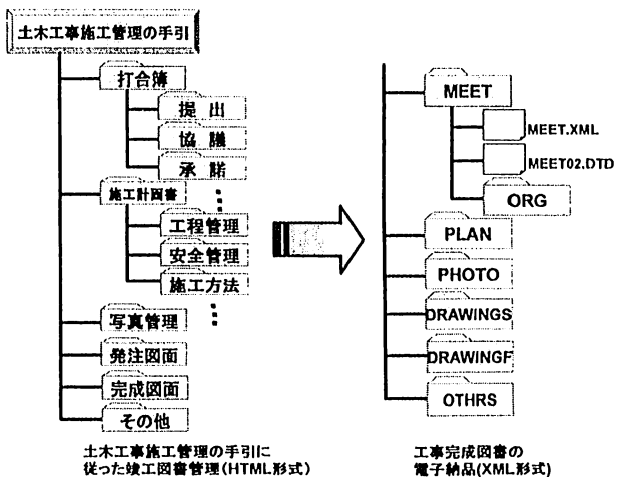


図-3 電子納品におけるXMLの活用例





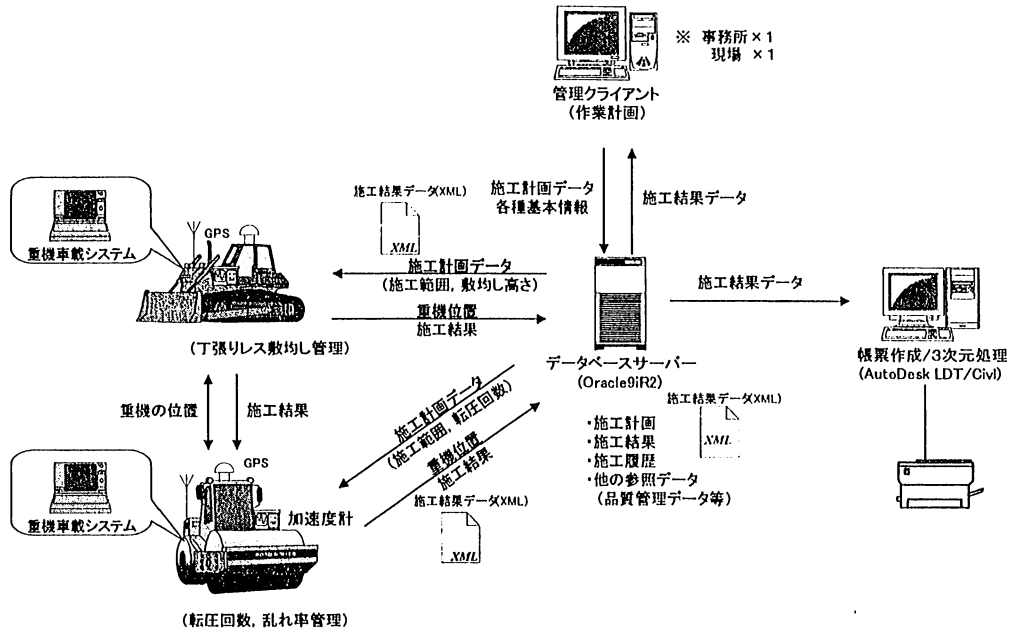


図-6 重機土工におけるXMLの利用例

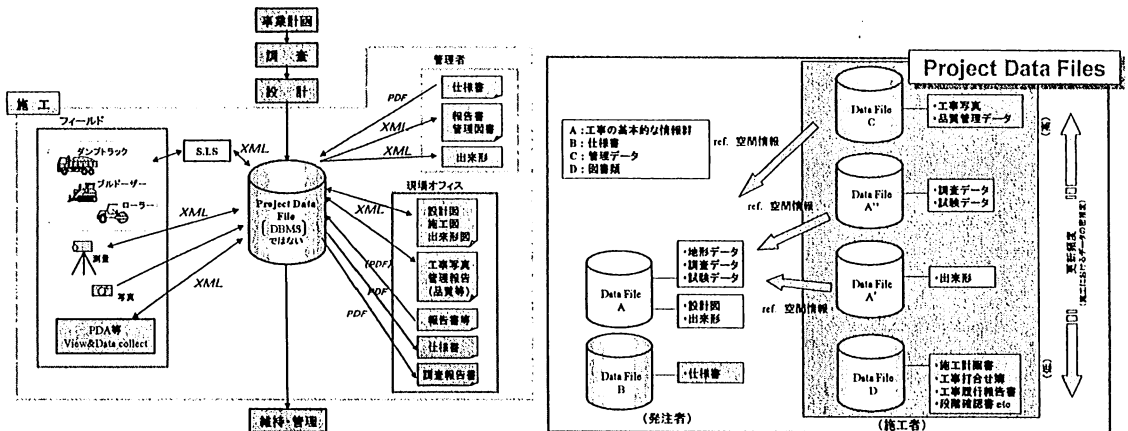


図-7 プロジェクトデータ(左:Physical Model , 右:Logical Model)

4.3 重機土工におけるデータを統合と施工管理におけるXMLの利用例

土工の施工管理は、これまで色々な施工管理システムが提案・実装され成果を上げてきている<sup>5) 6) 7)</sup>。特に大規模土工においては重機の管理とそれに付随する施工計画の立案が、整然とした施工を行い所定の品質を満足するためにも重要である。当然そのような施工を行うことは施工の合理化にもつながる。

筆者らは空港島の造成工事において図-6に示すような施工管理システムを構築し、現在現場適用を行っている。システム構築の目的は以下の通りである。

- (1) 施工計画支援：CAD上で重機土工計画の立案
- (2) 重機土工支援：施工指示・支援(GPSと無線LANによる施工支援) → 施工の合理化

- (3) データベースを用いた施工データ管理
- (4) 3次元 CAD による柔軟な帳票作成
- (5) XML 技術によるデータ標準化：維持管理データとしての活用を考慮

ここで XML は重機からの施工結果データと、上記(5)に示す工事管理データの中で活用している(図-7参照)。なお、図-6における重機からの施工データに XML を利用することは、一般にはクローズドシステム内の XML 利用で、データ量を増やすことになり一般には行われないことであるが、今後の建設機械における情報の標準化(ISOの動向)も踏まえ、XMLスキーマをデータ辞書に容易に置き換えられるよう考慮したものである。また、工事管理の中で品質管理は、これまで材料品質管理や施工品質管理について材料使用場所や施工位置の位置情報の管理が難しかったため、これらの品質管理を別々に帳票で管理していたが、メタデータ(XML)で空間情報を参照するデータ管理により、図-7に示すようなプロジェクトデータ管理を行い、3次元データ上で各施工地点における使用材料品質、施工品質のデータ管理が可能となり、これまで帳票類等の作成に費やしていた作業時間を大幅に効率化できる。更に、空間情報と一体となった品質管理データが容易に取り出せることから、性能設計に対応した品質管理データを維持管理に利用することも可能であると考えられる。

## 5. まとめ

土木工事で使用される情報は、発注者から提示される契約図書を基本としており、その中に含まれる設計図面や測量成果、施工数量表等を利用して施工や施工管理、工事関係書類の作成等を行っている。現状では施工においては現場の測量や出来形管理や品質管理、数量確認といった施工管理資料の多くは、人手による転記やデータの手入力による計算処理を必要とし、数量計算根拠提示のための図面への色塗りなど、膨大な手間を必要としている場合が多い。

今回、紹介したようなXMLの活用は、単にデータ交換を容易にすることのみならず、3次元空間データ利用も考慮したシステム構築等を行い運用した場合、空間情報の正確喝迅速な表示(View)だけではなく、出来高管理や工程管理等、施工中の様々な管理での利用も考えられる。さらに、設計・施工のみならずGISを利用した維持管理においてもデータを活用することが可能となると考えられる。

## 参考文献

- 1) 佐用泰司：工事管理，鹿島出版会，1978.
- 2) ISOにおける性能照査型設計とCENにおけるCPD（建設製品指令）：「ISOへの対応」に関する第2回シンポジウム，1998.
- 3) 日本建築学会：建築工事標準仕様書・同解説 JASS5 鉄筋コンクリート工事，1997.
- 4) 新田恭士，古屋 弘：土木施工の情報化と業務改善（その1）－情報を活用した建設事業のライフサイクルマネジメント－，第58回土木学会年次学術講演会，2003.
- 5) 古屋 弘：Web-GISを用いた土工事の施工管理システム，2002年度土木情報システム論文集，Vol.11，pp.67～76，2002.10
- 6) 古屋 弘ほか：加速度計を利用した締め固め管理システムにおける解析手法の比較，第54回土木学会年次学術講演会，1999.
- 7) 例えば，古屋 弘ほか：振動ローラーを用いた締め固め管理手法に関する考察（その2：加速度データと現場密度の関係），第36回地盤工学研究発表会，2001.

## 財務情報の XML 化事例

ピー・シー・エー株式会社  
水谷 学(mmiz@pca.co.jp)

日本における財務情報の電子文書化は 1998 年 4 月に稼働開始した東京証券取引所の TDnet\*1 と 2001 年 6 月に稼働開始した金融庁の EDINET\*2 により大きく前進した。しかしながら、TDnet では CSV 形式および PDF 形式、EDINET では HTML 形式であるため情報の再利用性では紙媒体での財務情報提供と大差のない状況である。

国税庁は 2002 年 11 月 8 日に 2003 年度から開始予定の国税電子申告の添付書類となる財務諸表部分に XBRL 形式を採用する仕様公開をした。XBRL は財務情報のための XML であり、国税電子申告での XBRL 採用が日本における XML 化を大きく促進することとなった。

### 1 財務情報の電子文書化の現状

現在のところ年次の財務情報は有価証券報告書、決算短信、商法計算書類（事業報告書・決算公告）、法人税申告書添付書類、融資申込書添付書類の形式で企業から発信されている。決算短信は 2 ページ目以降に有価証券報告書の財務諸表相当部分が速報として記載される。融資申込書添付書類の財務諸表は法人税申告書添付書類の控えを利用するのが一般的である。作成を要請する法令との関係では、証券取引法(財務諸表規則)、商法(商法施行規則)、法人税法の 3 グループとなり、様式として少なくとも 3 種類存在する。

これらの財務情報の電子文書は、有価証券報告書については 2004 年 6 月より全上場会社を対象に HTML 形式で EDINET のホームページ上に提供される。提供のタイミングは財務局閲覧室では即時であるが、ホームページでは提出の翌日である。しかも株主総会の終了後に提出されるので決算短信を元にした情報配信よりも 1 ヶ月近く

遅れることになる。TDnet による決算短信の情報配信では、決算短信の 1 ページ目の情報が CSV 形式で日経新聞、時事通信社等の報道機関に専用線で提供されている。ピーク時には 1 分間に 100 社を超える情報配信がある。この情報の株価に与える影響は非常に大きく、我が国で最も利用されている財務情報といっても過言ではない。

他方、未公開企業の財務情報の電子文書化は進んでいない。2002 年 4 月より電子公示による商法決算公告(インターネットでの決算公告)が認められているが、普及しているとは言い難い。電子公示の方法による場合は、要旨による決算公告は認められないという制度上の問題が普及の障害と言われるが、そもそも従前より官報での決算公告をしていない企業が多いという悪い慣行があるのである。商法決算公告の電子公示では HTML、PDF の形式に加え、XML 形式も容認されるが、実態は HTML、PDF での公示のみである。

公益法人関係では、2001 年 8 月にインタ

\*1 TDnet(Timely Disclosure network) <http://www.tse.or.jp/DISCLOSE/html/daily/today.html>

\*2 EDINET (Electronic Disclosure for Investors' NETwork)  
<http://info.edinet.go.jp/InfoDisclosure/main.htm>

ーネット上での決算書類等の公開が義務付けられている。形式は HTML 形式である。

このような例をあげると、ここ数年で電子文書化が大きく進展しているようにも受け取れるが、TDnet が配信している CSV 形式の決算短信情報を除いて財務情報サプライチェーンを構成できていない。HTML や PDF といった電子文書形式では、正確な情報交換を保証することは困難だからである。

TDnet が採用している CSV 形式は効率的な情報交換が可能であるが、PDF 形式の関連資料との不整合を検証できない問題や、業種別の勘定科目のように多様性のある記載項目の追加が困難であるといった問題があるので機能的な限界に達していると認識されている。元になる情報を XML 形式にすれば、スタイルシートを使って PDF や CSV 形式に変換が可能である。XML 形式を採用することは、従来の情報形式に対応し、なおかつ不整合を起こさない解決策となるのである。

米国では日本のように証券取引所によるディスロージャのレポジトリサービスは行なわれておらず、監督機関である SEC が行なっている。EDGAR と呼ばれるシステムがあり、誰でもインターネットでアクセス可能である。SGML、テキスト、エクセル等の形式の財務情報が提供されている。日本と状況は異なるが、XML 化(XBRL 採用)の必要性は日本と同じで、エンロン、ワールドコム事件に関連して議会証言にも登場したほどである。

## 2 XBRL の登場

XBRL は AICPA(米国公認会計士協会)が

中心となって設立した XBRL ORG によって開発された財務情報交換用 XML であり、現在は XBRL International という非営利法人により開発が継続されている。1998 年に米国公認会計士チャールズ・ホフマンにより発案されたプロジェクトが前身であるので、普及までに長期を要している。日本には 2000 年 8 月に第一回 XBRL シンポジウムの開催により紹介された。

## XBRL の特徴

XBRL は XML の一種であるが、以下の特徴を持っている。

- ・ 記載項目辞書であるタクソノミーの採用
- ・ 表示項目(要素)と名称(ラベル)の分離
- ・ 勘定科目の種類毎に要素を定義
- ・ 記載項目の背景情報(コンテキスト)の採用
- ・ 定義、計算、表示、名称、参照のリンクベース(W3C から勧告されている XLink によるリンク情報の集合)の採用
- ・ 英文財務諸表への変換を含む、多言語対応

ある会社のある年度の売上高をインターネットで検索する場合、Yahoo や Google といった著名な検索エンジンでも、この答えはでてこない。PDF か HTML 形式の財務諸表を検索して人間が読み取るしかないのが現状である。財務情報が XBRL 形式になれば、会社(Entity)、年度(Period)、売上高(Sales)を指定して正確に検索できるようになる。売上高が「製品売上」のような別の名称であっても概念(この場合、Sales)の同一性で判定されるのである。しかも、この検索処理はパソコンの表計算ソフトで

十分可能である。上場会社の財務諸表は EDINET や TDnet により、誰でもインターネット経由で無料で入手できるようになったが、XBRL 形式になればデータとして利用できることとなり、今までとは次元の異なる用途が生まれるであろう。ノートパソコン1台で、全上場会社を対象にした分析が可能となるのである。

XBRL 以外の財務情報用 XML 仕様が存在しない訳ではない。例えば、米国 Intuit 社の qbXML や ArapXML といった XML 仕様が存在するが、個々の会計・ERP ソフトウェア・ベンダーの XML 仕様は他のベンダーが採用しないし、会計基準の設定主体(団体)が関与していない団体が作成した XML 仕様では財務諸表部分の要素定義の保守ができないといった問題があるので、少なくとも財務諸表に関連する XML 仕様としては XBRL が最有力であることは間違いない。XBRL には国際会計基準の設定主体である IASB や、AICPA を始めとする各国の会計士協会などが参加している。

XBRL 仕様は用途により XBRL FR (Financial Reporting) と XBRL GL (General Ledger) に分けられる。前者は財務諸表を対象とする仕様であり、後者は入金伝票・仕訳帳・総勘定元帳・試算表・勘定明細等を対象とする仕様である。XBRL GL には勘定科目要素について XBRL FR へのリンクを設定できる。財務諸表をクリックすると、その明細を表示するような機能を備えているのである。XBRL FR に関しては競合する XML 仕様がないと言って良いが、XBRL GL は EDI を含むので他の XML 仕様と競合する。財務諸表情報との連動が必要な用途では XBRL GL にアドバン

テージがある。XBRL International は HR-XML・ebXML・EDIFICAS・UBL・OMG 等との間でインターオペラビリティの協議を継続している。XBRL はコンテンツ・モデルであり、トランスポートに関する仕様は持っていない。従って SOAP を使用したり、国税電子申告のように、別の XML 文書の中に取り込んでから送信を行なうことになる。セキュリティ技術に関しても、XML の一般的なセキュリティ技術を併用可能である。

### 3 XBRL の採用状況と今後の用途

海外ではオーストラリア金融監督局 (Australian Prudential Regulation Authority : APRA)、アメリカの連邦預金保険公社 (Federal Deposit Insurance Corporation : FDIC)、ヨーロッパの国際信用保険協会 (International Credit Insurance Association : ICIA)、イギリスの内国税歳入庁 (Inland Revenue) など多方面で採用されている。XBRL Japan のホームページで公開されている FACT BOOK に最新の事例紹介がある。

日本においては 2002 年 11 月に東京で開催された XBRL 国際会議で東京証券取引所、三井住友銀行、ワコール等から採用の発表があったほか、国際会議の直前に国税電子申告での採用が国税庁から公開された。

2003 年 4 月 18 日に産業構造審議会 (経済産業相の諮問機関) の産業金融部会から発表された報告書に、電子ファイナンス市場の創設にむけた XBRL の必要性が記されており、今後の用途として大きく期待されている。国税電子申告に合わせて多くの会計・ERP パッケージ・ベンダーが XBRL 対

応ソフトウェアを提供するものと予想され、XBRL 形式の申告データ(インスタンス文書)は容易に作成できるようになるであろう。国税庁は受理した申告データを他の機関の用途に流用することはしないが、納税者側が電子申告用に作成したインスタンス文書を他の機関に提出する形態での再利用が期待されている。電子ファイナンスでは、融資債権の売買の道が開かれる。XBRL 形式の財務情報と他の申請情報に基づく信用評価情報が添付されて売買が行なわれるだけでなく、年々更新される財務情報と連動する仕組みが必要となるが、XBRL を利用すればこのような仕組みも実現可能である。

近年、連結納税や商法連結決算公告といった改正があり、連結経営の必要性が増している。単独決算であれば、財務情報を月次の段階で把握し舵取りが可能である。連結経営の場合、子会社の月次決算が固まってから財務情報を入手したのでは、舵取りが一步後れてしまう。大企業の中には親会社が採用している ERP システムを子会社の全てに使わせる対処をしている場合があるが、子会社の中には高速な通信回線が確保できない地区にある会社や小規模な会社もあるのでコスト面で不利だけでなく、子会社の規模に合わない仕様のソフトウェアを使用する業務上の不効率が生じる場合がある。また、革新的な業務システムが登場しても採用できなくなる硬直化の問題もある。このような場合、子会社が報告する財務情報形式に XBRL を採用すれば、異種システム間での情報交換が可能となるし、業務システムのリプレースによる影響を最小限抑えることも可能である。

連結対象会社間の財務情報交換だけでな

く、企業内の情報システムの統合のために XBRL を使用することも可能である。ワコールは、2002 年 11 月の XBRL 国際会議でそのような情報システムの統合のために XBRL GL を採用した発表をしている。業務システムは定期的に機能改善が行なわれる。企業情報の一元化は必要であるが、業務に合ったサブシステムが他の基幹システムと相互に情報交換できれば大きな業務効率の改善を期待できる場合もある。そのような情報交換形式として XBRL GL は有力な選択肢の 1 つとなりえる。

企業内の情報システムでの利用例として、モニタリングという用途がある。電子認証が普及すると、やがては領収証や請求書等の証憑書類が電子的な証明になると予想される。商取引の实在証明は、時刻認証で発行されるハッシュ値によりソフトウェアで検証可能となる。財務情報との関連を検証するには、勘定科目のような財務概念の ID 定義が必要となるが、XBRL のタクソノミーに定義される要素名は、その ID として機能することになる。財務情報の発生と同時に、関連情報との検証を行なうことが可能になれば、人件費を中心とする情報処理コストの低減だけでなく、情報公開のスピードアップ、不正防止にも貢献することになる。

物流関係では、RFID が普及しつつあるが、RFID の記憶容量の増大により、XBRL 形式の財務情報が利用可能となりつつある。通関や納品に必要な情報を XBRL 形式で RFID に記録できれば大きな業務改善の可能性がある。

#### 4 財務情報の比較可能性

XBRL Japan は財務情報の記載項目辞書であるタクソノミーとして、有価証券報告書用・商法計算書類用・税務用等の種類に分けて開発している。これらは、基本財務諸表部分を共有する構造となっているので、そのレベルで比較可能性が保たれている。つまり、有価証券報告書の中の財務諸表や、商法の決算公告、融資資料として入手した財務諸表といった具合に入手目的や経路が異なっても、一定の比較可能性は確保されているのである。上場会社の財務諸表は別として、中小企業の財務諸表を入手することは極めて困難であるので当然の配慮であるが、世の中で使用されている全ての勘定科目を比較できるようにすることは困難である。ある財務情報提供者は、50万勘定科目のデータベースを持っているそうである。財務情報の利用者が必要としている情報は、もっと要約整理されたものであることは言うまでもない。XBRL形式の財務情報の利用を検討しているのであれば、XBRL Japan が作成した基本財務諸表が、比較可能であるべき項目を満たしているか調べておく必要がある。XBRL Japan では taxonomy@xbrl-jp.org という e メールアドレスで、タクソノミーに対する項目追加の要望や質問を受けている。

よくある質問として、「XBRL を採用すれば日本の財務諸表と海外の財務諸表は比較可能になるのか？」がある。答えは厳密には NO であるが、YES と言えるレベルの場合もある。XBRL は比較可能性を重視した仕様であることは間違いないが、会計基準の違いを吸収することは不可能である。従って、国際会計基準を採用する企業の情報は比較可能であるが、日本のように独自

の会計基準を採用している場合のインスタンス文書は比較可能性が低い。近年、日本では税効果会計や金融商品会計を国際会計基準に合わせる対応をしている。企業結合会計と減損会計を導入すれば、ほぼ国際会計基準並みになるところまで来ている。しかしながら、それら残る 2 つを導入しても国際会計基準ではないので注意してもらいたい。日本は欧米から会計の仕組みを導入したのだが、長い歴史の中で独自の会計慣行を形成している。そのため、国際会計基準や米国 SEC 基準で作成された財務諸表と日本の財務諸表は内容的に異なる。タクソノミーに登録されている要素の数で示せば 90% 以上の要素名が異なるのである。

この問題に対し、XML 技術でできることはマッピング（組み替え）である。マッピングには A の勘定科目を B に組み替えるという単純な操作と、A と B を合計したものを C に組み替えるという操作などがある。マッピングに関しては、ドイツ銀行から XBRL International に仕様の提案がされている。他方、ヨーロッパの国際信用保険協会である ICIA では、独自技術による XBRL インスタンス文書のマッピングシステムを用意している。日本の財務データもその対象に含まれている。信用保険の用途では厳密に財務諸表を比較することが目的ではなく評価することが目的であるので、会計基準の違いは許容範囲のようである。YES と呼べる例ではないであろうか。

##### 5 提出企業側で対応すべき事項

XBRL では、企業が使用する個々の勘定科目に要素名を定義している。XBRL Japan が提供している基本財務諸表タクソ



Japan が提供している基本財務諸表タクソノミーに登録されている勘定科目体系と現在使用している勘定科目体系との互換性を把握し、組み換えができるレベルに調整しておくことが望ましい。過去の会計慣行では、企業固有の勘定科目を尊重するあまり、企業間での財務情報の比較可能性が損なわれている。紙媒体での財務諸表は、罫線や均等割付といった美観を重視していたが、財務情報の高度利用の時代には、データの比較可能性の方が重視されるのである。因みに XBRL 関係者の間では、前者をプレゼンテーション・セントリック（表示中心主義）、後者をデータ・セントリック（データ中心主義）と呼んでいる。XBRL は後者である。

## 6 むすび

以上のように、財務情報の XML として XBRL がデファクト・スタンダードの地位を固めつつある状況である。国税庁のような公的機関での採用は、対応ソフトウェアの開発を一気に加速させることが予想される。東京証券取引所での XBRL 採用と異なり、その影響は 170 万社を超える日本の全企業に及ぶ可能性がある。しかしながら、電子申告自体の普及は、コストの問題や、慣れの問題があり、場合によっては電子帳簿保存制度のように 1% 程度の数の企業で採用さえるに止まる可能性もある。電子ファイナンス市場で、電子申告用に作成した財務諸表データが再利用されれば、保証料の割引のようなインセンティブが提供されることによって、電子申告の普及に貢献することになる。融資業務では従来より税務申告の控えの財務諸表を利用しており、電

子ファイナンスにおいても電子申告の控えの財務諸表データの利用が期待されているのである。

XBRL を利用できる法人税電子申告は 2004 年 3 月より名古屋国税局管内において開始される予定であり、他の地域では 2004 年 6 月からとなる。2004 年が日本における XBRL の本格利用開始の年となる。XBRL Japan は 2003 年 11 月末に電子申告対応簡易ソフトウェアを添付した入門者用書籍を発行するほか、マイクロソフトからは新しい Office 製品用として、XBRL 用アクセラレータが発表されるなど、一般の人が安価に入手できる対応ソフトウェアが出揃うことになるので、関連情報に注意しておくべきである。

## 参考文献

Charles Hoffman, CPA, Carolyn Strand, CPA, Ph.D., "XBRL ESSENTIALS", AICPA, ISBN: 0-87051-353-2

「電子メディアによる情報開示に関する研究」2000(平成 12)年 9 月 20 日 日本会計研究学会

「XML ビジネス標準開発のための設計情報アーキテクチャ」湯浦克彦 小柳義夫 大野 邦夫、デジタルドキュメント 31・3(2001.11.30)

## 特許事務処理システムのXML化

特許庁情報システム課  
安久 司郎

## 1. はじめに

特許庁では1990年から特許電子出願を基礎としたペーパーレスシステムを稼働し、出願手続き、庁内事務処理および審査、特許公報の発行、電子図書館の構築等の一連の業務の流れを電子データによる処理で行える環境を構築し、実際にも業務の効率化を進める中核となってきた。このペーパーレスシステムは、1985年当時のシステム設計段階で考えられる技術の中で最新かつ世界標準となる可能性の最も高いものを採用し、将来に渡ってシステムが維持できることを目標にしたものであった。例えば、データ通信プロトコルはOSIを採用し、電子文書の仕様は当時のCCITT規定のミクストモードデータ文書交換標準仕様をベースに構築していた。しかしながら、今やこれらの技術のデファクトスタンダードはWeb技術、HTTPをベースとしたインターネット環境であり、また、電子文書の標準もXMLが大きく注目され、Webとの連携の上では必須のものとなりつつある。そこで、これらの新たな技術に対応するために、まずは2003年からペーパーレスシステムで扱う電子文書仕様をXML対応とし、出願から特許公報までの一連の処理を、XMLをベースとした形態で行えるようにすることとした。

## 2. 特許XMLデータの世界標準の策定

XMLを採用するにあたり、最初に検討しなくてはならない点は、その標準化ということである。特許の世界の場合、扱う文書データは、出願人、発明者等の氏名、住所に代表される書誌的データと特許の技術内容を説明する明細書データに大別されることになる。書誌的データについては、特許法に規定された非常に複雑な手続き処理を行う必要があるため、特許法および規則で規定された事項に関する非常に多くのボキャブラリが必要となる。また、明細書の場合、通常のテキストデータの他に、図面、表、数式、化学式、遺伝子配列等の多岐にわたる種類のデータを扱う必要がある。

これらの内容について、標準化を行い使用するXMLを規定することになったが、今や特許データの利用は日本国内で閉じたものでは全くなく、全世界的に通用するものでなくてはならず、標準化にあたっては、世界標準として策定し、それを採用する必要がある。そのため、世界知的所有権機関(WIPO)を中心として、この標準が取りまとめられることとなり、日本特許庁もその主要メンバーとして標準策定に参加した。

この国際標準策定にあたって、基本的に対象とするモデルとしては、PCT(特許協力条約に基づく国際出願)の様式が基礎とされた。その理由としては、現在、各国はそれぞれ独自の特許法を持っているため、使用される書類の様式も各国特許庁ごとに大きく異なり、それら全てを網羅したボキャブラリ、スキーマ等を策定するのは容易なことではないので、現状では、唯一、全世界で統一様式として利用されているPCT国際出願の様式により、基本のボキャブラリ、スキーマを策定した。そのため、本標準の中には、この標準を各国で使用する場合のこの標準の国内仕様への拡張方法に関するルールも規定することとした。

この標準は、WIPOにより特許協力条約実施規則付属書F<sup>(1)</sup>として2001年12月に発効し、現在は、PCT国際出願を電子的に行う場合については、各国特許庁が守らなくてはならない基準書となっている。この標準の中でXMLに関して規定されたものの概要は以下のとおりである。

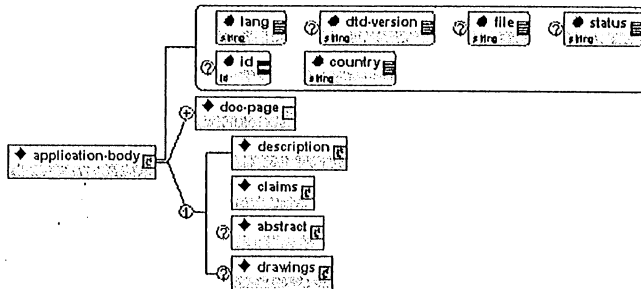
(1) 文書DTD

中核をなす DTD は以下のものがある。(最新のバージョンでは以下のものを含め、約 20 の DTD が規定されている)

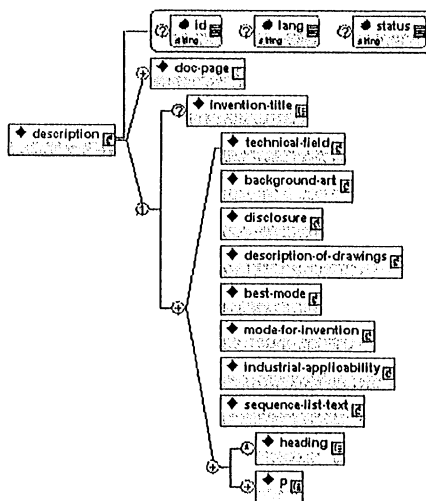
- package-data.dtd : 文書パッケージエンベロープ用 DTD
- request.dtd : 願書データ DTD
- fee-sheet.dtd : 料金表 DTD
- application-body.dtd : 出願明細書等 DTD

application-body は明細書データであるため、文字飾りを含むテキスト、表、数式、図面等に対応しており、作成にあたっては以下の点が考慮されている。

1. 使える部分については、XHTML で定義されているコンポーネントを利用
2. PCT 条約で規定されている項目は独自タグとして規定
3. 数式は W3C 規定の mathml2 をモジュールとして組み込む (オプション)
4. 表は OASIS の Open XML Table Model を組み込む (オプション)
5. 外部参照実体として、イメージファイル (形式は TIFF、JFIF、PDF) を参照可能



application-body.dtd の全体概要



application-body 内の明細書部分の概要 (例えば、p はパラグラフ用のタグ)

(2) 文字コード

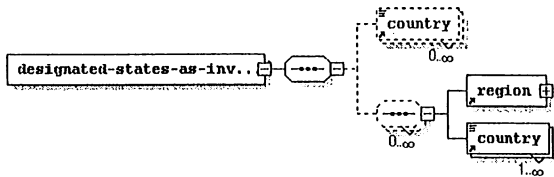
全世界での利用を考慮し、文字コードセットは Unicode3.0、エンコードは UTF-8 を標準とした。ただし、各国ごとに他のエンコードの利用も許されている。(例：日本の場合は Shift-JIS)

(3) スタイルシート

PCT 国際出願用に、標準のスタイルシートが規定されている。PCT 規則の関係で、明確なページレイアウトが要求されているため、XSL-FO によりそれぞれの文書スタイルシートが規定されている。

この PCT 国際出願の標準仕様内で定義された DTD に基づいて、標準スキーマコンポーネントセットの定義も別途行っている。このコンポーネントは、XML Schema を利用して、個々のタグを分解し、コンポーネントとして定義し直したものである。新たな文書 Schema を作成する場合を、このコンポーネントを組み合わせることで作成することができる。

以下は、"designated-states-as-inventor" というコンポーネントの例で、子供のコンポーネントとして "country" と "region" を別の XML Schema コンポーネントとして取り込んでいる。



```
<xsd:schema xmlns:xsd = "http://www.w3.org/2001/XMLSchema">
  <xsd:include schemaLocation = "country.xsd"/>
  <xsd:include schemaLocation = "region.xsd"/>
  <xsd:element name = "designated-states-as-inventor">
    <xsd:annotation>
      <xsd:documentation>
        PCT Rule 4.6(c) :
        "The request may, for different designated states, indicate different persons as inventors where, in
        this respect, the requirements of the national laws of the designated States are not the same. In such
        a case, the request shall contain a separate statement for each designated State or group of States in
        which a particular person, or the same person, is to be considered the inventor, or in which particular
        persons, or the same persons, are to be considered the inventors."
      </xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref = "country" minOccurs = "0" maxOccurs = "unbounded"/>
        <xsd:sequence minOccurs = "0" maxOccurs = "unbounded">
          <xsd:element ref = "region"/>
          <xsd:element ref = "country" maxOccurs = "unbounded"/>
        </xsd:sequence>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

```

</xsd:sequence>
</xsd:sequence>
</xsd:complexType>
</xsd:element>
</xsd:schema>

```

現在のところ、XML Schema については、この標準コンポーネントセットの定義にのみ使用しており、文書スキーマはあくまでもDTDを基本としている。新たな文書DTDを作成する場合は、まず、このXML Schema コンポーネントを組み合わせることで1つの文書用のXML Schemaを作成し、最後にDTDに変換する手法を用いている。

### 3. 日本国内特許データへの適用

日本国特許庁においては、この標準コンポーネントから国内特許出願用のDTDの作成を行っている。この場合、国内出願で規定されている規則に対応した独自のタグが必要となるため、標準化の作業の中で作成されたルールに基づき、ネームスペースを利用して独自のタグを定義している。ただし、DTDはネームスペースを基本的にはサポートしていないので、現状では、DTD内に `jp:` というプレフィックスを最初から埋め込むことで実現している。以下は、サンプルDTDの一部である。

```

<!ELEMENT jp:pat-amnd (jp:amendment-a51 | jp:amendment-a523 | jp:amendment-a524 |
    jp:amendment-a525 | jp:amendment-a526 | jp:amendment-a527 | >
<!ATTLIST jp:pat-amnd
    lang CDATA #REQUIRED
    dtd-version CDATA #IMPLIED
    file CDATA #IMPLIED
    status CDATA #IMPLIED
    xmlns:jp CDATA #FIXED "http://www.jpo.go.jp" >
<!--information regarding Agents or common representatives JPO expanded-->
<!ELEMENT jp:agents (customer-number | jp:agent+ ) >
<!ATTLIST jp:agents
    jp:error-code CDATA #IMPLIED >

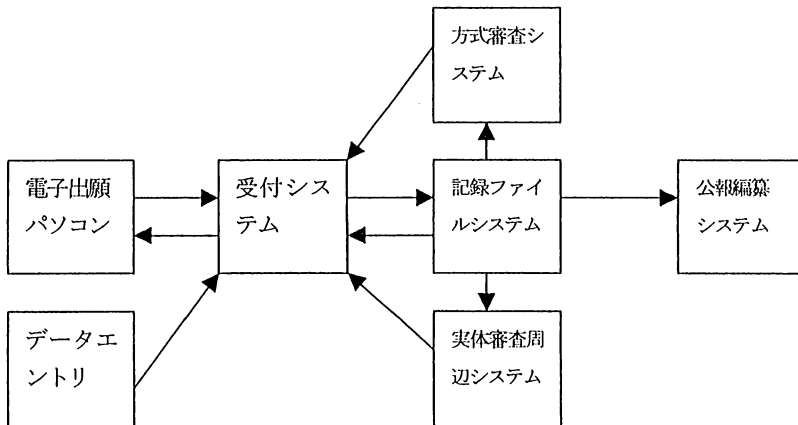
```

なお、日本特許庁では、`mathML2` については現状では一般的なツールがまだ十分でないとの判断から、採用を見合わせており、複雑な数式は当面はイメージデータを利用することとなっている。また、スタイルシートについては、国内出願の場合はXSLTによるHTML変換を利用している。(PCT国際出願についてはXSL-FOを利用しているが、出願件数の多い国内出願では現在のところ利用していない。)

### 4. 特許庁内事務処理システムへの実装

特許庁の事務処理システムの大まかな流れは下図の通りであり、すでに全ての処理は電子データで行わ

れている。電子出願の受付、もしくは、書面による出願を XML データ化したものを受け付けシステムが受領し、各種処理を行った後に、電子公報として同様に XML で発行されるまでの流れを対象としている。これらのシステムを全て、これまで述べてきた標準に準拠した XML 対応のシステムに変更、もしくは新規システムとして開発を行った。



- 電子出願パソコン (パソコン出願ソフト)

従来は、CCITT 規定のミクストモードデータ文書交換標準仕様をベースとした独自のフォーマットを用いて電子出願を行っていたが、2003年7月から、その基本フォーマットを XML に変更し、業務運用を開始した。出願人は、出願データを作成するにあたり、はじめに HTML 形式で願書や明細書を作成し、パソコン出願ソフトが持っている HTML→XML 変換ツールで日本特許庁仕様の XML にそれぞれの文書データを変換する。必要な XML を生成した後、それらの XML と添付されるイメージファイル等の外部ファイルをパッケージ情報管理 XML と共に ZIP により 1 ファイルとし、ISDN 回線を通して、パソコンから特許庁の受付サーバに対して送信、電子出願を行う。

また、特許庁で作成された各種通知類も XML データとして作成され、同様の形式により電子出願パソコン側で受領することが可能である。

- データエントリ (書面出願対応)

特許出願は、電子出願によるもののみではなく書面によるものも受領する必要があるため、受領した書面出願の内容を電子出願と同様な形式の XML データに変換して、その後の処理を行っている。書面による出願は、特許庁での受領後 OCR によりコードデータに、図面はイメージデータに変換し、規定されているタグを付与して標準の XML を生成する。このデータを受付システムに投入して、電子出願されたものと同様の処理を行う。

- 受付システム

ISDN を通じて受領した XML ファイルを含む ZIP ファイルは受付システムで解凍され、DTD によるパースチェック、その他のデータ形式チェックを行い、後続システムに引き渡される。また、受付システムは、特許庁内の方式審査システム、実体審査周辺システムで作成される各種の通知、起案類の XML データを電子出願パソコン側に渡す機能も実装しており、オンラインによる発送、書類閲覧も XML により行うことができる。

- 記録ファイルシステム

XML データとして受領した出願データは、全て記録ファイルシステムで管理される。このシステムで記録、保管される XML データが、その特許出願の原本データとなるため、それぞれの書類ごとに XML データを管理サーバ内のデータベースにそのまま蓄積し、管理を行っている。また、XML データ内に存在する各種の主要データ項目（出願人名、住所、代理人情報等）は XML データから抽出され、別途データベースが構築される。これらのデータベース内の情報を、後続の処理システムが利用することとなる。なお、これらの XML データによる文書を照会する場合、照会用の PC において、対象となるスタイルシートを適用して、規定されたレイアウトで照会することができる。

- 方式審査システム、審査周辺システム

方式システムは、XML による出願データが法律、規則に適合しているかのチェックを行うための支援システムである。XML データ受領直後に、受付システムでパースチェック等の基本的なチェックは行われているが、この方式審査システムでは、タグ内に記載された内容同士の対比チェックや、他の XML 文書データ内の項目との対比による妥当性チェックもシステム上可能な範囲で自動的に行っている。方式上の違反がある場合は、その指摘内容を記載した通知類を担当官が XML で作成し、出願人が受領できるようになっている。また、実態審査周辺システムでは、審査官が特許審査に関する各種の通知文書を同様に XML で作成し、出願人に受付システムを通して送付している。

- 公報編纂システム

特許出願されたものは、基本的には全て公開公報という形で公開され、さらに、特許となった場合には特許公報という形態で、その特許された内容が発行される。これまで、簡易な SGML 形式で、これらの公報は電子的に CD-ROM で発行されてきたが、今回の XML 化に伴って、これらの公報も XML 化し、DVD-ROM による発行を行う予定である。（2004 年 1 月を予定）XML 公報を発行する場合、提出された XML データをそのまま発行するわけではなく、各種の編集が必要となる。そのため、公報用の DTD は出願用に定義されたものとは異なる別の DTD を使用しており、それらは出願用 DTD よりも多くの付加情報を含んでいる。また、特許庁における審査の段階で、出願人はその出願内容の補正を行うことが認められているため、公報編纂時にはそれらの補正された内容を差し替えて編集する必要がある。例えば、ある段落をそのまま補正された新たな内容に入れ替えたり、指定された図面を入れ替えたり、削除するといったものである。これらの編集について、今回の XML 公報編纂システムでは XML の特質を極力生かして、可能な限り自動的に行うことを目指している。

## 5. 今後の展開

ここまで述べたように、国内出願に関しては、基本的には従来からのシステムの基本文書仕様を XML に置き換え、新たなシステムの運用を実施することができた。今後はこれらの機能をさらにレベルアップしてより XML の特徴を生かせるようにシステム化を進めることとなる。また、来年には、そもそもの XML 化の標準化のきっかけとなった国際出願の電子出願開始を予定しており、これが実現した後は、国際出願としての XML 出願データをこの標準に準拠して受領し、世界各国に展開されていくことになり、さらなる発展が期待される。

参考文献・情報

1. WIPO, "Standard for the electronic filing and processing of international applications: Annex F"  
[http://pcteasy.wipo.int/efiling\\_standards/EFPage.htm](http://pcteasy.wipo.int/efiling_standards/EFPage.htm), [http://www.wipo.int/pct/en/texts/pdf/ai\\_01\\_part7.pdf](http://www.wipo.int/pct/en/texts/pdf/ai_01_part7.pdf)
2. Paul Brewin , Shiro Ankyu, "Filing and Processing Patent Data Using XML - A World Standard", XML Europe 2002, Barcelona, Spain.



## これからの電子自治体システムと人材

福岡県高度情報政策課

溝江 言彦

### 1. はじめに

福岡県の IT 化への取り組みは、2,000 年度から本格的に始まった。その IT 戦略は次のようなポリシーより成り立ち、それぞれのポリシー内の政策のスパイラルな実施により体系化される。

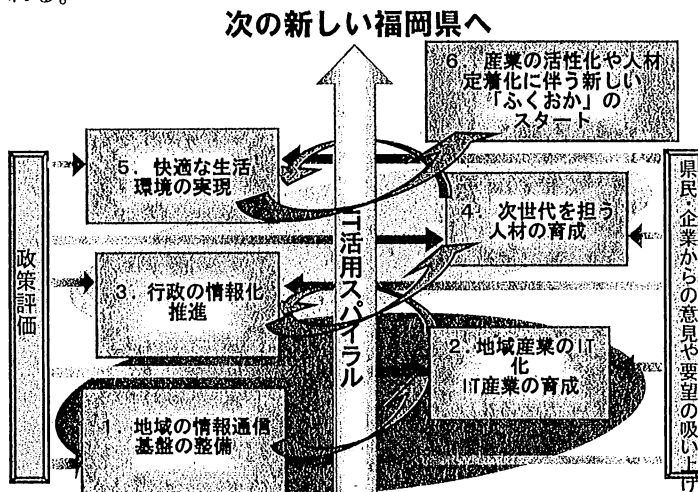


図1 IT活用のスパイラル

### 2. ふくおか電子自治体共同利用センター構想について

現在地方自治体は税政のひっ迫、住民ニーズの多様化などにより今までのサービスでは満足を得られないというといったことが大きな課題となっている。また、市町村合併に伴う問題も加わり、こういった問題に対応するために電子自治体の推進は業務の効率化・コスト削減・住民サービスの向上などが期待できる大変有効な対策である。

しかし、電子自治体の構築は国・県だけではなく市町村まで連携して推進していくことが重要である。とはいえ理想的な情報基盤を単独自治体で整備するのは、財務的・人的制約により非常に難しいため、複数の自治体が共同して構築することが解決策と考える。福岡県は県が事務局となって市町村電子自治体共同運営協議会を立ち上げて、平成 17 年度には共同アウトソーシングサービスの開始を行えるよう進めている。

## 2. 1 電子自治体化推進の背景

### ①外部要因

- ✓ 国内外における電子自治体構築による住民サービスの向上への先進的取り組みが拡大
- ✓ インターネットの普及により、物理的制約（時間、空間）を受けない住民サービスニーズの高揚
- ✓ 住民サービス満足度等の指標が流通し、自治体の比較評価尺度の認知（東洋経済、ダイヤモンド、日経等）
- ✓ 情報公開法施行にともなう文書管理運用の重要性増大

### ②内部要因

- ✓ e-Japan 重点計画による達成目標の明確化
- ✓ 総合行政ネットワーク稼動によるコミュニケーション形態の変化、公的認証基盤による真正保証の導入
- ✓ 財政状況の悪化にともなう、業務効率化の必要性増大（行財政改革の必要性）

### ③ 県内市町村の現状と課題

- ✓ システムの維持運用に専門スタッフを配置する必要性が増大
- ✓ 技術高度化の進展が早い一方、専門ノウハウを有する人材確保が困難
- ✓ 情報の機密性増大にたいするセキュリティ対策へのコスト拡大
- ✓ 多様化した住民サービスへの対応が困難になってきている

自治体のあるべき姿(ToBe)実現のためどのような手段を用いればよいか。

#### ◆自治体のあるべき姿(ToBe)のための達成目標

- 業務改善・効率化によるコスト削減
- 行政サービス品質の向上

## 2. 2 ふくおか電子自治体共同運営協議会の取組

福岡県が事務局となり、平成14年に立ち上げた「ふくおか電子自治体共同運営協議会」（任意協議会）には県内68市町村が加入（平成15年9月10日現在）しており、平成17年度中のASPサービス開始を目標に、検討を行っているところである。

### ① 事業目的

各市町村が共同運営し、費用負担の軽減と行政サービス向上の両立を図る

### ② 目標設定（共同運営事業の定性的、定量的な最終目標を設定する：具体的には平成15年度決定予定）

- ✓ 各アプリケーションシステムの構築、運用管理費用を、単独で実施する場合と比較して平均70%に節減する

- ✓ 24 時間 365 日ノンストップでのサービス提供を行う
- ✓ 最高レベルのセキュリティを確保する
- ✓ サービスの追加、変更を迅速かつ柔軟に行う
- ✓ 利用者（自治体）が利用したいサービスを任意に選択することができる

## 2. 3 ふくおか電子自治体共同利用センター

### 第1ステップ

- ✓ コミュニケーションインフラ、サービスインフラの共同利用
- ✓ LGWAN 接続サービス
- ✓ 行政機関などにおける安全なメール環境文書交換、公的認証基盤の活用
- ✓ インターネット接続サービス
- ✓ 外部チャネルの拡大、情報公開の推進

### 第2ステップ

- ✓ アプリケーションシステムの共同開発、共同運営によるサービスの拡充
- ✓ 電子自治体アプリケーションの共同開発複数市町村によるアプリケーションの共同開発によるコストの削減
- ✓ システムの共同運営
- ✓ 共同利用センターにおけるシステム運用の共同アウトソーシングによる人的負担の軽減

## 2. 4 ふくおか電子自治体共同利用センターの全体像

ふくおか電子自治体共同運営協議会での電子自治体共同アウトソーシングは、公共 iDC（ふくおか電子自治体共同利用センター：以後共同利用センター）を定め、そこへの電子自治体システムの共同アウトソーシングとなるが、共同利用センターは、単なるシステムの共同ハウジングではなく、電子自治体システムの共同運営、共同利用を実現し、各市町村の電子自治体化における負担の軽減と効果的な課題解決を図ることとしている。

各市町村は、それぞれが必要とするアプリケーションを選択し、利用することが可能であり、各自治体が所有するレガシーシステムとの連携も可能とする。また、共同利用センターで利用するシステムの開発仕様、運用・保守、契約、SLA（サービス水準保証）等の標準化、共通化を行い、システム機能における重複を排除し、効率性の向上を図る（システムにおける全体最適の実現）ことも目標とし、そのために福岡県が構築した電子自治体共通基盤の活用を想定している。

### 3. 業務システムの最適化手法への取組について

#### 3. 1 電子自治体の実現イメージ

電子自治体構築においては、今までは総合的に電子自治体システムを議論することが欠けていた。全体が最適化された電子自治体システムは、その土台の部分はどう作るかが重要なポイントである。電子自治体システムの各サービスをどのように連携するかだけではなく、自治体の IT 調達・開発運用に関する問題もある。それをどのように解決していくかは業務システムの最適化手法（EA：Enterprise Architecture）を用いることが解決策と考える。その手法を用いて構築したものが電子自治体共通基盤である。

#### 3. 2 業務システムの最適化手法の考え方

福岡県の電子自治体共通基盤とは、業務システム最適化手法（EA：Enterprise Architecture）のモデルにおける適用処理体系、技術体系を技術参照モデル（TRM：Technical Reference Model）としてまとめたもので、今年度（平成15年度）に整備を行っている

#### 3. 3 業務システムの最適化手法（EA）とは何か？

それは組織の目標とミッション、ビジネス、それを支える情報、アプリケーションシステムとテクノロジーを定義する業務・システム全体の設計図であり、日本では経済産業省が e-Japan 実現のための IT ガバナンス手法として紹介している。中央府省では採用を表明しているほか、東京三菱銀行など一般企業でも採用の動きがある。それは大規模組織で、バラバラ個別に企画・構築・管理されてきた情報システムを、その調達プロセスから見直し、組織全体の調達フレームワークに移行させることを意図するからである。

#### 3. 4 なぜ EA なのか？

それは EA のメリットはそのまま「電子自治体」のメリットであるからである。

システム基盤を統一化し、その統一した技術に対する知識を得るとともに継続して最新動向を確認し、必要に応じて基盤を発展させる。それはオープンシステム環境での IT 調達～運用～管理には不可欠なものである。

技術基盤を自治体が管理することで、システム調達のたびに必ず基盤と業務ロジックを新規導入して委託規模を膨張させる必要はなくなる。調達の都度、必要な規模で調達できるため、地元 IT 企業の参入障壁が下がり、地元のノウハウ蓄積に繋がる。技術基盤の統一と、運用まで含めた各種標準ガイドラインが定まることによって、サービスの品質が安定し、これによって共同利用センターなどの集中管理は結果的に大いに安定する。

### 3. 5 共通基盤システム の概念図と TRM の位置付け

共通基盤の考え方をシステム化した共通基盤システムの全体概念と、TRM（技術参照モデル）の位置づけを図に示す。

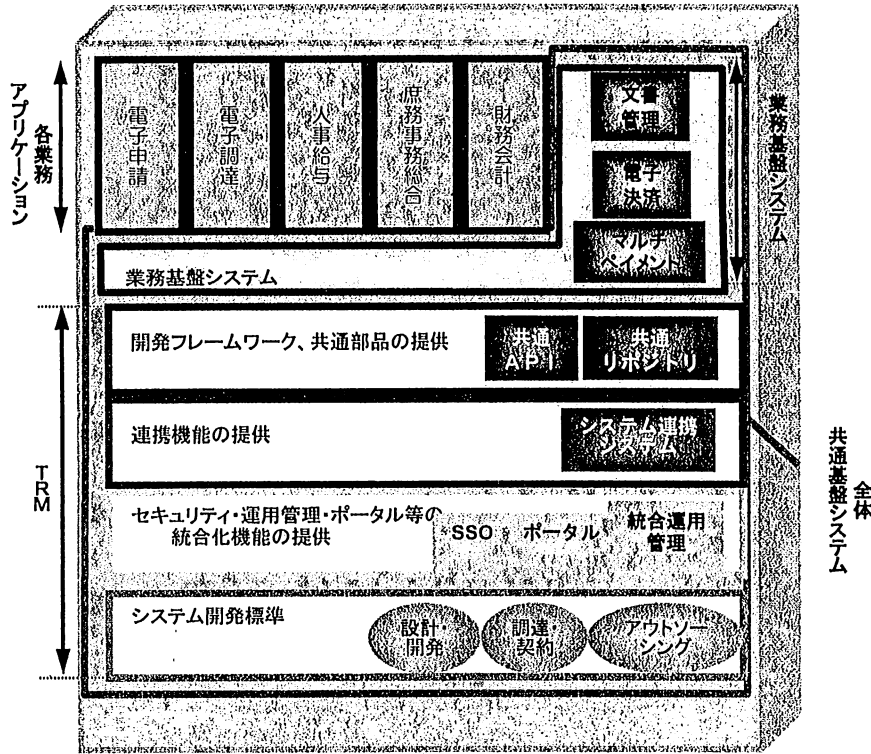


図2 共通基盤システム

TRMは、システムの作り方の約束事であり、これに業務的な共通機能（業務基盤）を組み合わせて、共通基盤システムの全体像となる。

- ①業務基盤システムは、文書管理や電子決済などの、電子自治体システム開発に必要な業務的な共通機能の提供を行う。
- ②TRM（技術参照モデル）は、主にシステム開発の決めごと（標準＋フレームワーク）を提供。

これにより、運用統合化、保守作業の地場への移行の容易化等が実現される。

### 3. 6 共通基盤によって実現させるべきもの

共通基盤によって実現させるべきものとして次の4つが挙げられる。

- ①J2EEを利用した、コンポーネント指向のシステム

共通基盤アーキテクチャを導入したアプリケーションは、J2EEを利用した、コンポーネント指向のシステムとなるため、システムの変更が必要となった場合に、コンポ

一ネットごとに変更を行うことが可能となる。

たとえばコンポーネント指向でないシステムの場合、一部の変更の場合でも、システムの広範囲で影響を受けたり、影響調査に多大なコストが生じたりするが、共通基盤を利用したシステムであれば一部のコンポーネントの変更のみで対応できる。つまりシステムの一部変更が生じた場合、共通基盤を導入していれば、ビジネスロジックの変更のみで対応が可能となり、かつコンポーネント指向のシステムとなっていると、ビジネスロジックの一部のコンポーネントの変更のみで対応することができる。

### ②システム連携等、共通機能、運用の一元化

各システムの、システム連携や、帳票出力などの共通機能の構成を、共通基盤にて一元化することによって、運用の一元化が可能となり、システムの運用コストの大幅な削減が図ることができる。そして運用管理の一元化、インターネット接続による脅威を防ぐなどのセキュリティ上の配慮、職員作業の高度化を図るため、全システム共通で職員ポータル、セキュリティ管理、運用管理機能の導入を図ることにより、統合的で高度な IT ガバナンスをきかせることができる。この統一は iDC 活用のためには必須となり、さらにセキュリティ機器、ストレージなどの iDC 提供機器の積極利用により、機器購入費の圧縮を図ることが可能となる。

### ③開発技法・技術の標準化

共通基盤によって、各システムにおける開発技法・技術を標準化することができ、高品質で効率的なシステム構築が可能になる。地場企業においても、その標準化された技術を習得することによって、システム開発への参入が容易となる。

福岡県では開発技法・技術を標準化し、統一するためのガイドラインとして、次のようなドキュメントを用意する。

- システム開発規定文書
- システム間連携規定文書
- システム品質管理規定文書
- 運用保守規定文書 等

各システムの開発者は、これらのガイドラインに従い、共通アーキテクチャを利用してビジネスロジック部分を構築する。ガイドラインに沿って構築を行うため、各システムの開発技法・技術が標準化でき、全体として統一されたシステムとすることが可能となる。

### ④アプリケーションシェア

福岡県構築の共通基盤システムは今後、電子自治体（県内市町村共同アウトソーシ

ング)への展開や県外自治体への展開を積極的に働きかけており、これにより、県内市町村および県外自治体の構築したシステムを相互に利用することが可能となる(アプリケーションシェアモデルの実現)など、多大なメリットの享受が可能になるものと考えている。

#### 4. 人材育成への取組について

##### 4. 1 高度 IT 人材アカデミー (AIP) : 誕生の背景

- ▶ 自治体のみならず、IT をより有効に活用する為には・・・・
- ▶ 最先端の IT を駆使出来る技術エキスパートが不足している。さらに・・・・
- ▶ 自治体および企業の IT 調達・活用に関わるガバナンスが不足している
- ▶ その原因の一つとして、現状のアセスメントや IT 調達の為の企画、対投資効果の高いシステム基盤やアプリケーションの設計・開発・運用など、システムライフサイクル全般を見通す人材・スキルの不足が指摘されている

##### 4. 2 AIP のミッション

IT を使ってビジネスプロセスを改善し、企業価値を上げてゆくために必要となるスキルを実践的・経験的に身に付けた最高位 IT 人材を戦略的に育成し、しいては育成される人材や知の集約をもとに、IT 活用の高度化と、地域の活性化を促進する。

つまり自治体や企業の IT 調達・保有のあるべき姿を、IT ライフサイクル全般に渡って見直し、次にあげるような IT ガバナンスを実現する人材を育成する。

- ✓ IT 戦略と IT 投資の決定
- ✓ 目的に合致する IT の適正な導入と構築
- ✓ 導入・稼働後の評価と適切なフィードバック

##### ①今後の企業ビジネスに必須であるエキスパートを幅広く育成

E-Business を実践する上で欠かせない e 人材。本アカデミーでは、IT 基盤整備から経営までの幅広い領域をカバーするエキスパートとして、戦略 IT 活用のリーダーとしてプロジェクトのマネジメントやプロセス改革を推進する「ビジネスマネージャ」、戦略 IT システムのリーダーとして業務分析に基づくシステム設計、プロジェクト管理を行う「ソリューションエキスパート」並びにビジネスに強い技術スペシャリストとして最先端の戦略 IT システムを構築・運用する「基幹エキスパート」を育成する。

##### ②実践力を備え即戦力となる技術者を育成

大規模なラボ機材を活用した実機演習や実例ベースのワークショップなどを中心としたカリキュラムを提供することにより、「最高レベルの技術力」と「多様なスキル」「ニーズ変化への柔軟な対応力」「リーダーシップ」「実践に即したビジネス適応能力」

などを兼ね備えた IT 技術者を育成する。

③オープンな IT 調達、運営スキームの整備

全運用期間全体にわたって最適なシステムを柔軟にかつコストを押さえて調達、運営するための枠組みを開発、研修を介して広く、行政、公益事業、民間事業に適応させて行く。

④各分野のリーディングカンパニーの協力による最先端のカリキュラムを提供

IT 業界をリードするシスコシステムズ (株)、日本オラクル (株)、サンマイクロシステムズ (株) やビジネスコンサルティング分野で豊富な実績を持つキャップジェミニ・アーンスト&ヤング、地元企業である九州電力 (株) や麻生グループが協力してカリキュラムを提供することにより、最先端かつ最高レベルの技術力とより広範囲で実践的なスキルを身につけさせる。

#### 4. 3 人材育成と EA

(自治体および地元企業の) EA の確立によって、(地元 IT 企業が) 身に付けるべきスキル領域が明らかになるため、人材育成の方向性が確立し、育成が効果的に行うことができる。

それぞれのレイヤで、参照モデル作成や標準化を進め、実際の IT 調達 (及び評価) を行うためのスキルと、そのレベルが ITSS で規定されている状況にすることにより、効果が高く高度な人材育成が可能となる。

#### 4. 4 共通基盤導入による効果波及

人材については、今まで自治体のアプリケーションを提供できるのは大手企業などであったが、共通基盤というリファレンスモデルを出して、それをもとに企業に構築を依頼する形にし、地元企業に構築方法を身に付けてもらって積極的に参加してもらいたい、と考え AIP を立ち上げ、そこで実践的な研修を行って最高級の人材を育成する事業を実施している。電子県庁・電子自治体、それに人材育成を含めたところで地域の IT 企業の発展モデルとしてとらえるのが福岡県の IT 戦略である。

共通基盤の導入と、AIP とを有機的に融合させ、次のような良循環が生まれることを期待している。

- ① 応札企業が習得すべき個別業務システム開発関連の技術を標準化、特定化
- ② 特定化された技術やシステム開発に必要な管理能力などを地域 IT 企業に習得させるために体系化し、育成
- ③ 地域企業に対して必要な技術と管理能力を習得させ、大手競合エリアの位置に押し上げる
- ④ 技術力を兼ね備えた地域 IT 企業を自治体アプリケーション開発に参入させる



## 電子書籍制作における XML の活用

大日本印刷(株) 前川 真二\*

### 1. はじめに

インターネット書店、プリントオンデマンドサービス、電子書籍ダウンロードサービスの登場に見られるように、インターネットの普及に伴い、出版物の流通形態が著しく変化している。特に、いわゆる「活字離れ」による出版物の売り上げ低迷、返品率増加によるコスト上昇は、出版社にとって深刻な問題となっており、在庫を抱えないプリントオンデマンドサービス、電子書籍ダウンロードサービス等新しい出版流通に対する取り組みが盛んに行なわれるようになってきている。

こうした状況の中、DNP でもこの新しい出版流通に対して、企画・製造から販売に至るまで一貫したアプローチを行なっている。(図1参照)

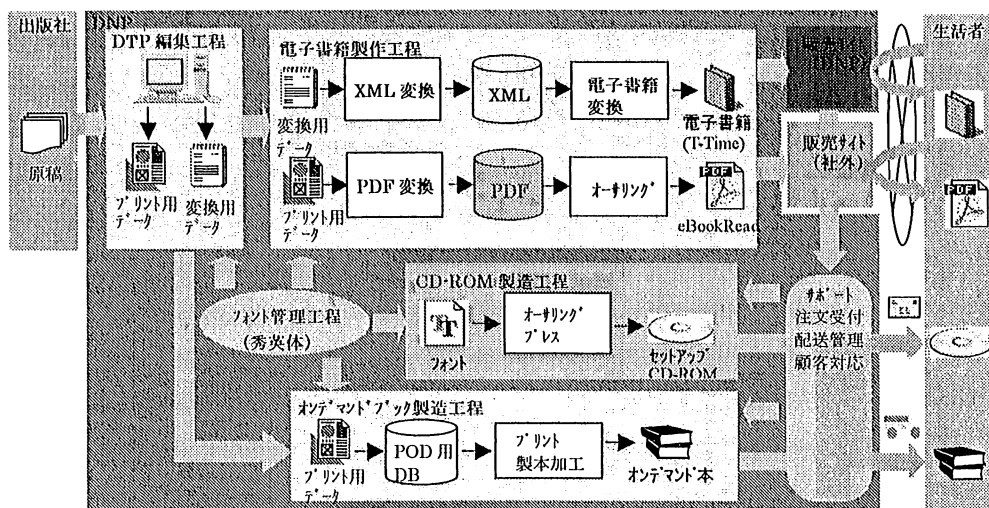
販売面では、書籍そのものを取り扱っている「専門書の社」(<http://www.semmonsho.ne.jp>) や、電子書籍及びオンデマンドブックを取り扱った「ウェブの書齋」(<http://www.shosai.ne.jp>) といった自社の販売サイトを構築・運営している。また、最近では、書籍だけでなく、写

真や映像まで幅広い電子コンテンツを扱うモールサービス「AnyStyle」(<http://www.anystyle.jp>) を立ち上げている。コンテンツの制作面では、既存の資産を有効に活用しながら、オンデマンドブックや電子書籍に対応できることが理想的である。例えば文庫本のコンテンツ制作において、一般書籍と同様に、既存の DTP 工程にて編集を行い、そこからのデータ流用が可能となれば、印刷会社だけでなく校正を行う出版社にとっても大きなメリットとなる。また、電子書籍フォーマットが多様化した場合、単に DTP のデータからダイレクトに電子書籍に変換するような工程では対応ができない。こうした課題を解決する手段として、XML の活用に着目している。

しかしながら、データ管理の効率化、生産性の向上、データの再利用などの実現が期待されているものの、具体的にどう取り組んでいくかについては、まだまだ模索の段階といえる。

現状、想定される課題として、XML を核とした様々なデータ交換を可能とする工程構築や、得意先の保有する多様なデータをニーズに応じて加工・配信する等、よりフレキシ

図1.DNP の出版コンテンツ流通ビジネス全体概要  
fig1. The outline of the publication contents circulation business in DNP



ブルなコンテンツ流通を実現する必要がある。今回の取り組みは、DTP から電子書籍へ変換する工程において、XML を中間フォーマットとして採用することにより、最初の実用

\* C&I 事業部 IT 開発本部  
〒162-0066 東京都新宿区市谷台町 6-3 市谷大東ビル  
C&I Operations, IT Development Headquarters.,  
6-3 Daimachi Ichigaya Shinjuku, Tokyo 162-0066

的な取り組みとして重要な意味を持つものとする。

当社では、実際にこの仕組みを構築・運用することにより、文庫系コンテンツにおいて、オンデマンドブックと電子書籍の同時進行を実現している。

本資料は、電子書籍を制作する過程でのXML採用にあたり、DTPデータを活用する為の編集ルールやDTPデータからXMLを経由し電子書籍データへ変換するプロセス及び開発したツール類の機能等を説明し、当社のXMLを活用した制作工程構築への取り組みについて、具体的に説明を行ったものである。

## 2. XMLと印刷工程

### (1) XML化の状況

実際のコンテンツ制作部分の説明に入る前に、印刷工程にXMLを活用することに対する、一般的な状況について触れておくことにする。

XMLの登場以来、その話題性から急速に普及したように思われている面もあるが、実際の印刷現場においてXMLを活用した事例が目立つような状況には至っていない。しかしながら、ADSL等の最新技術による低価格な高速常時接続環境の急速な普及等により、各企業のインターネットを活用した様々な情報提供ビジネスがより盛んになり、ユーザの必要とする情報を望まれる形式で提供できるようになると考えられる。そうなれば、豊富な情報を一元管理し、そこから必要な部分を取り出して、様々なスタイルに自動変換することが必須条件となってくる。このような機能を実現するにあたっては、XMLの採用が有力視される。また、電子政府構想においても、文書の構造化が推進されており、世の中の様々な分野の情報がXMLで管理されるようになりつつある。

このような情報の構造化の動きは、出版業界にとっても例外ではない。出版物の流通の活性化を目的に、コンテンツを紙だけでなく、様々な形態に展開することに取り組んでいる出版社にとって、XMLの採用は重要な要素であり、業界の中で様々な検討が開始されている。例えば、電子出版市場の普及に備え、様々なフォーマットへ容易に交換を実現できるよう、XMLをベースとしたJepaXを策定し提供している日本電子出版協会の取り組みがあげられる。

こうした出版業界の動きに対して、印刷会社としてもXMLを中心とした制作、製造工程への取り組みの必要性が高まっている。

### (2) XML採用における課題

電子書籍制作工程にXMLを採用するにあたっては、いくつかの課題が存在する。

レイアウト情報主体の既存制作工程に、文書構造情報主体のXMLを持ち込むことは、オペレータの作業手

順が現状と大きく変わってしまう為、それに伴った作業負荷の軽減、人材の育成、新たな工程の構築など、様々な問題をクリアする必要がある。

また、XMLに対応した編集ソフトや、エディタなど、XMLを扱う環境もまだまだ不十分な状況である。

こうした状況において、全ての書籍を網羅するような取り組みも必要であるが、そのことを視野に入れつつ、まずは最適な部分に対してXMLを活用した事例に取り組み、しっかりとした効果を検証するべきである。その上で拡張展開するというアプローチが重要であると考えられる。

## 3. 電子書籍制作フローの設計

前章で示した課題を解決した、新たな電子書籍制作フローを設計するにあたり、いきなり全ての書籍と全ての工程をXML対応にすることは、コストや実現時間の面から考えても現実的でないと判断し、最適な部分にXMLを採用し、今後拡張展開していくこととした。また、既存のDTP、印刷工程をなるべく変更することなく、低コストかつ現状の現場オペレータのスキルレベルにて電子書籍制作ができることを目的とした。

従って、書籍制作のための既存のDTP工程をそのまま活用し、DTPからのデータを変換して電子書籍データを生成するフロー(図1参照)を設計した。その過程において、XMLを中間フォーマットとして採用し、自動処理を行うところ、今回の取り組みのポイントである。

中間データにXMLを採用したことにより、今後電子書籍市場が成長し、様々なビューワ(フォーマット)が登場したとしても、そのフォーマットに対応したスタイルシート(XSL)に切り替えることで多フォーマットへの対応が可能となり、その場合のコストや対応期間を低く抑えることにつながっている。

この電子書籍制作フローにより以下のようなメリットが挙げられる。

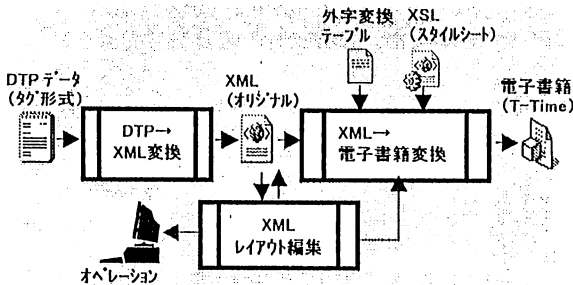
- 1) 印刷物と同じ環境(ゲラ)で得意先校正が行える
- 2) XMLを自動的に生成される中間データという位置付けにすることで、オペレータはXMLを意識することなく作業が行える
- 3) DTPの完全データは、そのままオンデマンドブック製造工程で活用できる
- 4) データ変換を行うプログラム以外は、特別なシステム等の導入は必要ない

## 4. 電子書籍変換処理

電子書籍変換処理では、DTPデータからXML、電子書籍データへと変換させていくが、それぞれの持つ特性が異なるため、工程設計においては、その利点を生かすような工夫と、外字情報の取り扱いには、十分な検討を行った。

紙への印刷を目的として優れた日本語組版機能を持つDTPと、パーソナルコンピュータ上での表示と操作性を重視した電子書籍とでは、表現力と機能性に大きな格差があり、DTPで編集した全てのレイアウト情報を電子書籍に反映させること、或いは、電子書籍が要求する機能をDTP編集にて指示することは不可能である。また、XMLは文書構造を定義するものであり、レイアウトに関する情報を内包することはないが、逆にDTPや電子書籍はレイアウト主体であって構造情報を持つことはない。こうした格差を工程のどの部分でどのように埋め、効率良く制作するかが大きな課題であった。また、書籍を扱う上で外字の活用は必要不可欠であり、異なるフォーマットを経由させる過程において、いかにして外字情報を正確に伝えていくかも重要な要素である。これらを十分に検討したうえで、今回の工程設計を行っている。こうした検討をもとに、図2に示す通り、DTPデータからXMLを介して電子書籍へ変換するにあたり、2種類の変換プログラムと1種類の編集ツールを開発している。変換プログラムは入力と出力のファイル名を指定することで、自動的にフォーマット変換を行うものである。編集ツールは、自動変換では対応できない細かな修正箇所に対応する目的で開発している。

図2.DTP→電子書籍変換フロー  
fig2. The conversion flow from DNP



読み込み対象としているDTPデータは、DTPが持つ自動組版機能にて使用されるトリガーコマンドと呼ばれるタグ付テキストデータである。また、中間生成するXMLは、日本電子出版協会が策定したJepaXをベースに、当社が必要な情報を付加している。JepaXは出版業界において電子出版コンテンツの蓄積や交換を行う為の交換フォーマットとして利用されることが想定されており、将来的に業界標準的なフォーマットになる可能性が考えられることや変換ツール等の充実が予想されることから、採用している。

変換対象としている電子書籍フォーマットは、現段階ではボイジャー社のT-Timeとしているが、XML→電子書籍変換プログラムにおけるXSLを切り替えることにより、様々なフォーマットへの展開を可能としている。

5. 電子書籍制作工程詳細

(1) DTPでの編集

通常の印刷物制作におけるDTPの編集は、レイアウト指示に基づき作業を進行すれば良いが、構造情報主体のXMLを生成するにあたっては、入稿原稿から構造化を検討したうえで、その結果を何らかの形でDTPデータに反映させる必要がある。

原稿から構造化を検討するにあたっては、見出し部分のスタイルと階層の深さで判断し振り分けるというルールを設定し、オペレータにリリースしている。

また、DTPデータに構造情報を反映させるには、DTPの持つ「浮動枠」という機能を活用している。これは、任意の範囲をボックス単位で区切り、そのボックスに対してレイアウトには反映されないような属性情報を指定できる機能である。前述したルールに基づき構造化検討した結果から、それぞれの構造を判別できるような属性を付加する。このような編集を行うことにより、DTPデータに構造情報を反映させることを実現している。

表1に示したのは、スタイルと浮動枠属性の関連ルールである。

また、実際のレイアウトに対する属性値の設定のイメージを図3に示す。

表1 浮動枠属性と原稿スタイルの関連

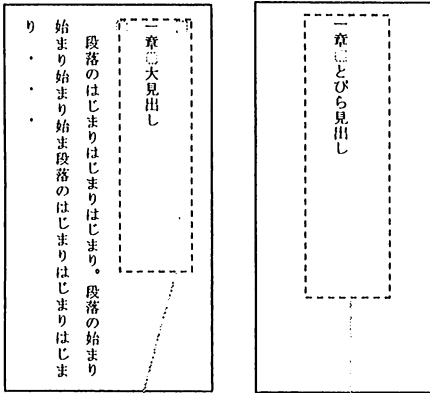
Table 1. Relation between floating frame and a layout

|    | DTP浮動枠属性 |     |     | 原稿スタイル | XMLタグ名 |
|----|----------|-----|-----|--------|--------|
|    | 属性値      | 表題  | レベル |        |        |
| 1  | 大        | 見出し | 1   | 改頁見出し  | 改頁見出し1 |
| 2  | 中        | 見出し | 1   | 改行見出し  | 改行見出し1 |
| 3  | 小        | 見出し | 1   | 改行見出し  | 改行見出し2 |
| 4  | 独立       | 見出し | 2   | 改行見出し  | 改行見出し3 |
| 5  | 独立       | 見出し | 3   | 改行見出し  | 改行見出し4 |
| 6  | 独立       | 見出し | 4   | 改行見出し  | 改行見出し5 |
| 7  | 独立       | 見出し | 5   | 改頁見出し  | 改頁見出し2 |
| 8  | 独立       | 見出し | 6   | 改頁見出し  | 改頁見出し3 |
| 9  | 独立       | 見出し | 7   | 改頁見出し  | 改頁見出し4 |
| 10 | 独立       | 見出し | 8   | 見出し予備  | null   |
| 11 | 独立       | 見出し | 9   | 見出し予備  | null   |
| 12 | 独立       | その他 | 1   | 扉見出し   | 扉見出し1  |
| 13 | 独立       | その他 | 5   | 扉見出し   | 扉見出し2  |
| 14 | 独立       | その他 | 6   | 扉見出し   | 扉見出し3  |
| 15 | 独立       | その他 | 7   | 前書き等   | 独立文章枠  |
| 16 | 独立       | その他 | 8   | 通常文章   | 文章枠    |
| 17 | 独立       | その他 | 9   | 余白     | 余白枠    |

※レベルが文書階層の深さを示す

図3.浮動枠属性設定例

fig3. The example of an attribute setting of a floating frame



本文中ははじめに出現する「大見出し」なので、変換テーブルNo1.No7～No.9の「改頁見出し」を使用する  
 とびら見出しなので、変換テーブルNo.12～No.14の「とびら見出し」を使用する

DTP 編集における入力ルールは、構造化（浮動枠）だけでなく、欧文、ルビ、連数字などの各種組版に対しても設定している。例えばルビは、親文字との文字列長との関係、親文字の位置（行頭、行末など）など様々な条件によって組版規則（JIS X 4051）が定められている。更に出版社によっても、より細かな規則が定められている場合もある。こういった細かな規則は DTP では表現できるが電子書籍では表現できない。また DTP では、見た目に基づき、スペース等を使用しながら微妙な調整を行っているが、出力されたデータをプログラム処理する際には、それらスペース等の意味を解釈することは不可能であり、結果的には崩れたデータを生成してしまうことにつながってしまう。こうしたことを回避するために、DTP での編集に対して、ルビであれグループルビは使用せずモノルビとし、スペース等による調整は行わないなど、ある程度制限を設定し運用を行っている。こうしたルールを DTP 編集現場にリリースすることで、その後のプログラムによる自動処理を可能としている。

(2) プログラムによる自動変換

DTP→XML→電子書籍への変換は、図4に示すようなメニューに対して、入力するDTPと出力する電子書籍のファイル名を指定するだけで自動に行えるようなプログラムを開発している。基本的には、中間に生成されるXMLを意識することなく、DTPから直接電子書籍に変換されるイメージであるが、内部の処理においては、必ずXMLを生成したうえで電子書籍に変換を行っている。

図4. 自動変換メニュー画面

fig4. The menu of an a conversion program

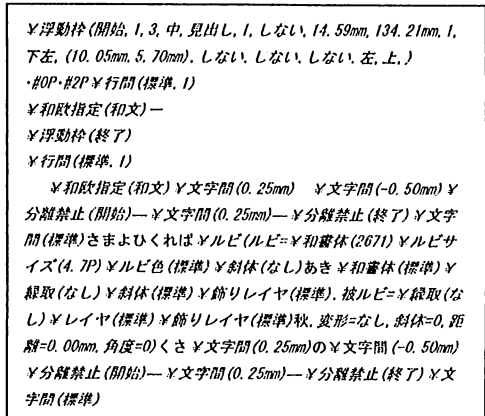


(2) - 1. DTP から XML への変換

DTP より出力されるトリガーとは、図5に示すように、様々な組版や制御に関するコマンドタグを含んだテキストファイルである。1つのコマンドタグの中に、そのコマンドの条件を示す要素が設定されており、コマンドの機能によっては、複数の要素が設定されているものもある。

図5. トリガー例

fig5. An example of the trigger command outputted from DTP



プログラムでは、このファイルを読み込み、組版や制御に関する情報部分のテキスト処理にて XML のタグ付けを行っている。おおまかな処理のロジックは、以下のとおりである。

- ①単一のコマンドタグ毎へ分解する
- ②単一のコマンドタグの要素を取り出す
- ③①~②の結果を解析シタグ形式変換を実行する
- ④①~③までの処理を繰り返し、全体のタグ変換を実行して一時的なファイルを作成する（構造情報は反映されておらずフラットな状態）
- ⑤構造情報を示すタグを解析し、全体の構造マップを作成する
- ⑥構造マップに従い一時ファイルのタグ変換を実行し、構造情報を反映させて最終的な XML を出力する

このような処理の結果、図6に示すような XML ファイルを作成している。

本来 XML は構造情報を定義するものであるが、この処理の過程においては、DTP で編集されたレイアウト情報を電子書籍に反映させる必要がある。そこで、構造情報以外にレイアウトに関する情報を定義するタグを設定し、DTP でのレイアウト情報を格納するようにしている。この XML の内容については、後の項で説明する。

図 6. XML 例  
fig6. The example of XML

```
<?xml version="1.0" encoding="Shift_JIS" ><jepax>
<bookinfo>
<book-title>海燕2</book-title><book-author>小島政二郎
</book-author></bookinfo>
<front></front>
<body>
<div class="改頁見出し1" id="body_00001">
<head>
<title>
<style br_before="2" h_level="H2" br_after="2" index="false"
pb_before="true"></style>
その<ruby><rt>ひ</rt></rt></ruby>
</title>
</head>
<div class="改行見出し1" id="body_00002">
<head>
<title>
<style br_before="1" h_level="H3" index="false"
pb_before="false"></style>
—
</title>
</head>
<p> —さまよひくれば<ruby><rt>あき</rt></rb>秋
</p></div></body></xml>
```

(2) - 2. 外字への対応

処理の過程において、外字についても考慮が必要なポイントである。トリガーファイル内の外字は、DTP 独自の書体及び文字コードで定義されている。電子書籍側の外字は、T-Time であれば PC 上の書体名とシフト JIS コードにて定義されているが、今後登場が予想されるビューワについては必ずしもシフト JIS とは限らない。また外字は、品目毎に新たな文字が必要となり、それに伴い追加作成されており、追加外字についても変換処理に確実に反映させるよう、管理・運用を行わなければならない。当社では、文字管理部門にて

外字の作成と一元管理（独自のユニークなコード体系）が行われ、DTP 編集工程、オンデマンドブック製造工程、電子書籍閲覧ユーザに提供されており、外字を扱っていくうえでは文字管理部門との連携が必要不可欠である。こうした点から、XML 内部での外字は CTS コード体系にて表現し、各ビューワの文字体系に併せた変換テーブルを用意することで、随時対応していく体制としている。

(2) - 3. XML から電子書籍への変換

図7に例を示すように、電子書籍 T-Time のデータは HTML に近いタグ付きテキスト形式の仕様となっている。

XML から電子書籍への変換は、XSLT に基づいたプログラムを開発して行っている。おおまかな処理のロジックは以下の通りである。

- ①XSLT 変換にて XML から T-Time タグへ変換する
- ②特定文字列の置換処理を行う
- ③外字コード変換処理を行う（CTS→シフト JIS）
- ④T-Time ソース整形処理（改行挿入）
- ⑤特定置換処理（欧文アクセント対応）

XSLT に関しては、XML 技術を利用した Apache 関連ソフトウェアの開発プロジェクトである Apache XML Project から提供されている Xalan を活用し、開発負荷の軽減をはかっている。

図 7. 電子書籍データ例 (T-Time)  
fig7. The example of digital book data(T-Time)

```
<HTML>
<HEAD>
<TITLE>海燕2</TITLE>
<T-TIME charspace=0 hommonsiz=16 linespace=12 indent=0
tatoyoko=2 midashisize=18 stroke=vertical start=0L
hashiravisible=true></T-TIME></HEAD>
<!------->
<BODY valign=top>
<P></P>
<!------->
<T-PB valign=middle>
<H1>海燕2</H1>
<T-PB valign=top>
<!------->
<H2>H次</H2>
<T-PB valign=top>
<T-PB valign=top hashiravisible=true>
<BR>
<H2>その<T-R>H (ひ) </H2>
<BR>
<H3>— </H3>
<BR>
—さまよひくれば<T-R>秋 (あき) くさの—<BR>
フラ<T-CODE src="glfgaiji/22227222.gif"><T-CODE
src="glfgaiji/22227223.gif">と、<T-R>三 (み) <T-R>千 (ち) <T-R>
代 (よ) は<T-R>海 (うん) <T-R>野 (の) の<T-R>家 (うち) の<T-R>
前 (まへ) に<T-R>来 (き) てゐた。<BR>
```

(3) XMLでのスタイル編集

自動変換で生成された電子書籍のスタイルに関して編集や修正を加える場合、専用のエディタ等がない為に、テキストエディタ等により、電子書籍のフォーマットを理解したうえで行う必要があり、オペレータにそのスキルが要求される。また、章や節単位に同じレイアウトがあっても、電子書籍上では文字列単位でしか扱えない為、作業負荷や修正ミスの可能性が高くなる。こうしたことに対応する目的で、変換の途中に生成されるXMLに対して、XMLタグを意識することなくスタイル編集が行えるツールを開発した。

このツールの特徴は以下の通りである。

- ①XMLの構造情報をツリー表示する (図8参照)
- ②ツリー表示された構造タグを指示することで、その文章内容が確認できる (図8参照)
- ③各階層の意味にあった編集メニューが自動表示される (図9参照)
- ④同階層に対して、共通・個別のどちらの単位でもスタイル指示が行える (図9参照)
- ⑤スタイルシートを追加することにより、様々なフォーマットへの変換が可能となる

このようなツールを開発しリリースすることで、XML及び電子書籍に関しての知識が少ないDTPオペレータでもXMLを介した電子書籍編集作業を可能としている。

図8. 文書構造ツリー表示

fig8. The tree display of document structure

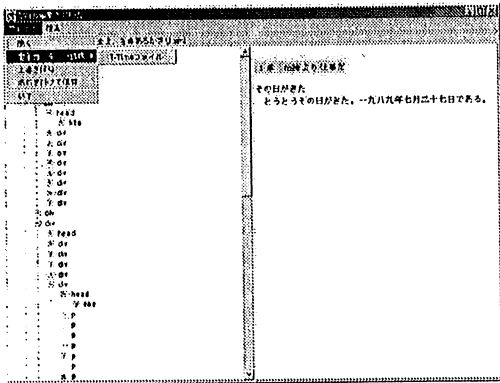


図9. 各編集メニュー例

fig9. The difference in the menu by structure

■JepaX 階層

- 目次
  - 目次表示に関する内容等
- ページスタイル
  - 組み方向など全体スタイルに関する内容等

■title 階層

- 改行
  - 見出しの直前直後の改行や改頁の設定等
- 目次・柱
  - タイトルを目次に表示した場合の設定

- 同じパスに同じスタイルを設定
  - ・チェックボックスをチェックすることで同一階層に同じスタイルが設定される
  - ・チェックしなければ個別の設定が行われる

(4) XMLフォーマットについて

これまでの中でも説明してきたが、DTPから電子書籍に変換する過程において生成しているXMLは、JepaXをベースに、スタイルに関する独自タグを加えた仕様としており、スタイル設定が必要な構造タグの直後に、

```
<style 要素1 要素2 要素3・・・></style>
```

といったように、<style>というタグを挿入することで、その構造に対して、「要素」の部分に指定した値のレイアウトが設定できるようになっている。

内部にスタイルに関する情報を設定することは、本来構造情報(XML)とスタイル情報(XSL)を分けて管理するといったXMLの定義からは外れてしまうものであるが、電子書籍への変換の効率化を最優先に考え、最低限必要なスタイル情報を把握可能な状態で残す方法としている。仮に、出版社等に純粋なJepaXの提供を望まれた場合、XSLT変換によりスタイル情報を取り外すことにより提供することが可能である。

このように、XMLの交換フォーマットとしての利点を十分に活かし、従来工程のフローをできるだけ変えることなく効率的に生産が行えることを優先し、スタイルを残したXMLを採用するに至っている。

なお、表2は<style>に関する要素の一覧、図10は<style>の指定例、である。

表2. style 要素一覧  
Table2. A list of the element in a styletag

| 設定項目         | 属性名            |                |             |
|--------------|----------------|----------------|-------------|
|              | jepax          | Title          | p           |
| フォントサイズ(本文)  | honmonsize     |                |             |
| フォントサイズ(見出し) |                | Xsize          | xsize       |
| 行間           | linespace      |                |             |
| 字間           | charspace      |                |             |
| 本文フォント       |                |                | fontname    |
| 見出しフォント      | fontname       | fontname       |             |
| インデント        | indent         |                |             |
| 先頭文字         |                | top_string     | top_string  |
| 見出しインデント     |                | br_flag1       | br_flag1    |
| 行揃え          |                | align          | align       |
| 前の改行数        |                | br_before      |             |
| 後の改行数        |                | br_after       |             |
| 前改頁          |                | pb_before      |             |
| 前改頁の位置       |                | pb_position    | pb_position |
| 後改頁          |                | pb_after       |             |
| 柱レベル         |                | h_level        |             |
| 段落前制御        |                |                | p_control   |
| 目次行揃え        | valign         |                |             |
| 目次出力(全体)     | index          |                |             |
| 目次出力(見出し)    |                | index          |             |
| 目次のインデント     |                | index_indent   |             |
| 目次のフォントサイズ   |                | index_xsize    |             |
| 目次の後の改行数     |                | br_after_index |             |
| 柱表示          | hashiravisible |                |             |
| 組み方向         | stroke         |                |             |
| 縦中横有効文字数     | tateyoko       |                |             |

図10. XMLにおける style タグ使用例  
fig10. The example of style tag use in XML

```

<jepax>
<body>
<div id="PBtitle1_1">
  <head>
  <title>
    <style title_indent="2" index_inde="3" hyouji="true">
    </style>
    XML技術
  </title>
  </head>
</div id="PBtitle1_2">
  <head>
  <title>
    <style title_indent="3" index_inde="1" hyouji="false">
    </style>
    XMLとは?
  </title>
  </head>
</div id="PBtitle1_3">
  <head>
  <title>
    <style br_before="2" h_level="H2" br_after="2"
    index="false" pb_before="true">
    </style>
    XMLって? </title>
  </head>

```

## 6. 今後の課題

これまで述べてきた内容は、電子書籍制作の一部において、従来の印刷物製造工程からXMLを介してデータ変換を実施するといった取り組み事例である。

今後は、初めにXMLを作成・管理してから、印刷物も含め様々なデータ展開を行うフローの実現が重要な課題であると考えられる。こうしたことが実現すれば、大量の情報の効率活用(串刺し検索等)や2次の展開が実現し、付加価値の高いサービス提供を可能とすると共に、製造工程コストの大幅な低減につながるものと予想される。

しかしながら、全ての書籍や印刷物を対象としてXMLありきの工程を実現するにあたっては、全体を確実に網羅できる標準的な構造化フォーマットの設定やそれを制作・管理・編集するためのツール(XMLエディタ、XMLデータベース、DTP等)の開発が必要となり、実現へのハードルは高いものがある。また、書籍や印刷物の全てのデータが構造化されることは理想的ではあるが、書籍の種類によっては、構造化の効果が望めるものばかりとは言いきれない。

よって、現段階においては、目的やコストに応じて、例えばJepaXのように、特定の目的に対して開発されているフォーマットを活用し、それに対して最小限の手を加えることで実現しておき、必要に応じてプレーンな状態或いは標準的なフォーマットに移行することも可能である、という道筋をつけておくことが次善の策であると考えられる。

但し、こうしたアプローチはあくまでも「標準」というグローバルな拠り所があつての取り組み方であり、当社としても、標準的なXMLフォーマットの開発を前提とした印刷工程の構築に取り組んでいく必要があると思われる。

## 文 献

- (1) 日本印刷出版協会出版データフォーマット標準化研究委員会:「JepaX - JEPa 電子出版交換フォーマット」, Ver0.9, 全44頁
- (2) 網モリサワ:「印刷とXML2002」, Ver1.20, 全14頁
- (3) 日本工業標準調査会 審議、日本企画協会 発行:「日本語文書の組版方法 JIS X 4051」, 全69頁

## XMLパブリッシングのための 出版標準フォーマットへの取り組み

田原 恭二†  
KYOJI TAHARA

インターネットの発達により、ネットワーク全体が大きな一つのデータベースになりつつある。ブロードバンド時代に対応するためには、データベース上の情報を、いつでも、誰でも、どこからでも取り出せ、利用できることが重要である。そのためには、データベース上の情報表現ルールが統一されていることが望ましい。XMLの利用が各分野において非常な勢いで拡大している最大の理由である。

出版分野においてもXMLの利用は広がっている。XMLを利用することにより、「コンテンツ」「文書構造」「文書体裁」を分離して持つことが可能となる。データベースに構築し、一元管理による資産管理が容易になるとともに、検索性の向上、多くのメディアへの展開、処理の効率化、時間短縮、コスト削減、正確性の向上などが可能となる。

### 1. はじめに

#### 1.1 インターネットの進展

図1.に示すように、総務省の速報<sup>[1]</sup>によると2003年7月末時点で国内のインターネット接続サービス加入数は9,624万件となり、インターネット利用人口は6,942万人(2002年末)となった。中でもブロードバンドの利用は、1年間で2.4

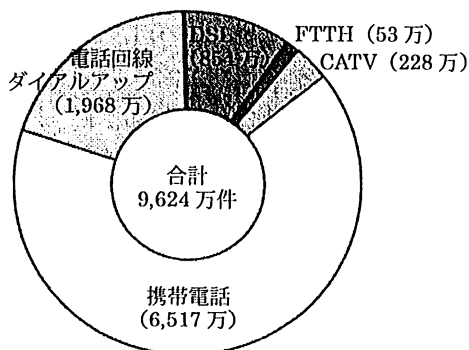


図1. 国内のインターネット接続サービス加入数  
(2003年7月末現在)

倍強の飛躍的な進展をみせており、現在はインターネット利用者の4人に1人以上がブロードバンドの利用が可能となっている。また、「カメラ付き携帯」等のヒットにより、携帯インターネットも発展しており、現在、携帯電話加入数7,780万件のうち84%(6,517万件)が、インターネットへのアクセス可能となっている。

このような値や、最近特に街角で携帯電話の画面を見ている人々を多く見かけるようになった、ということも含めて、インターネットがより身近なものとなりつつあることを実感させられる。

#### 1.2 インターネットコンテンツ量の推移

インターネットコンテンツの情報量も増大している。4年前の1998年(平成10年)では、国内のインターネットコンテンツの総データ量は664ギガバイトで、総ファイル数は3,648万ファイルであったが、2002年末(平成14年)には、総データ量が10,150ギガバイト、総ファイル数が27,421万ファイルとなった。これは、この4年間で、総データ量が約15倍、総ファイル数が約8倍増加したことになり、今後もますます増大していくことが予測<sup>[2]</sup>されている。このよう

† 凸版印刷株式会社Eビジネス推進本部研究開発部  
Research and Development Department,  
E-Business Operations, TOPPAN Printing Co.,Ltd.



なインターネットコンテンツ量の推移からも、インターネットが着実に、大きな一つのデータベースになりつつある、ということが窺える。

### 1.3 XMLが注目されてきた理由

図2に、インターネットの発達に伴う、新しい情報流通市場の創出、これまでの情報システムの基本的な仕組みの変化、インターネットというグローバルネットワークが大きな一つのデータベースへと進化する、という潮流を示した。

このような中では、データベース上の情報を、いつでも、誰でも、どこからでも取り出せ、利用できることが大変重要になってくる。そのためには、データベース上の情報をどのアプリケーションからでも簡単に取り扱える、共通な情報表現ルールが必要となる。このことが、現在、各分野においてXMLの利用が注目されている最大の理由である。

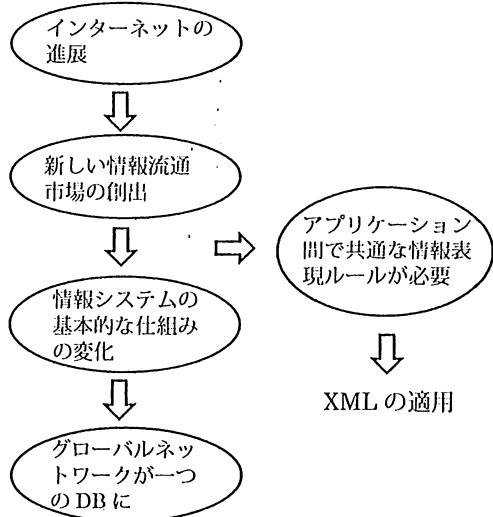


図2. インターネットの発達による潮流

## 2. メディア・ソフトの流通市場規模

図3.は、国内におけるメディア・ソフトの流通市場規模<sup>[3]</sup>を示したものである。

このうち、出版分野に大きく関わるテキスト系ソフトについては、市場の約5割を占めている

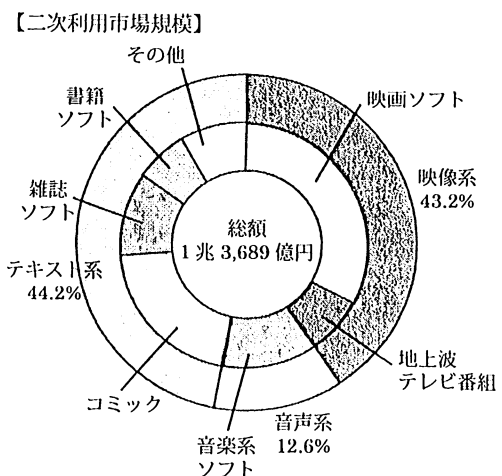
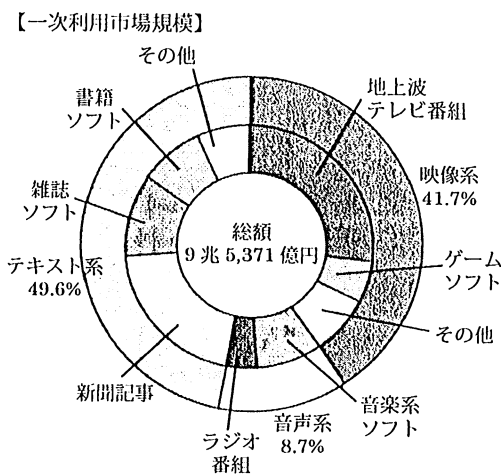
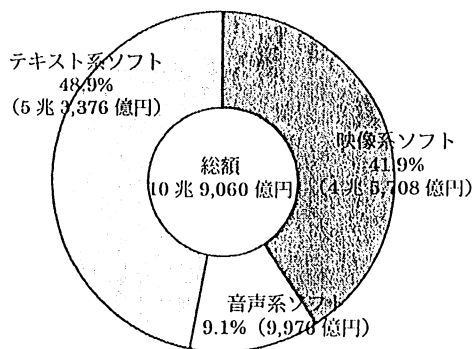


図3. メディア・ソフト流通市場規模 (2002年末)

ことがわかる。更に、コンテンツの一次利用と二次利用という視点で詳細を見てみると、テキスト系ソフトについては、新聞記事など、より即時性が求められるコンテンツは一次利用が主体であり、週刊漫画雑誌からコミック化の流れが定着している漫画コンテンツは二次利用性が高い。

また、これら以外の書籍ソフトや雑誌ソフト等の、いわゆる“文字が主体のコンテンツ”は、全体シェアは少ないものの、一次利用と二次利用の比率変化が少ないことから、再利用性が高い、及び多メディア展開が行い易いといった性質があると思われる。

### 3. 市販組版システムの現状

現在、書籍や雑誌の制作で多く使われている市販の組版システムでも、XML対応のものが数多く出てきた。ネイティブなXMLデータベースと連動し、コンテンツ管理を容易にするとともに、自動組版による処理の効率化、時間短縮、多メディアへの展開などが実現されている。

しかしながら各システムは、従来の組版システムを活かした内部フォーマットを利用しており、コンテンツの受け渡しには変換プログラムが必要である。これは、データ流通という面から見るとまだまだ問題である。あわせてXMLデータを組版システムに取り込み、しかるべき組版編集を行った後で、組版データを元のXMLデータへ自動的に完全復元できないといったことも、資産の一元管理やワンソース・マルチユースといった面で、大きな問題となっている。また、XMLそのものを入力、或いは編集するツール(XMLエディタ)の機能が弱く、複数ページにまたがるような修正が可能なシステムが少ないのも現状である。

### 4. XMLパブリッシングモデル

図4は、現在凸版印刷が取り組んでいる多メディア対応のXMLパブリッシングのモデルである。当社が考える理想のモデルは、入力から出力まで、いかなる出力媒体に対しても、XMLを中心とした運用が行えることであり、その中心とな

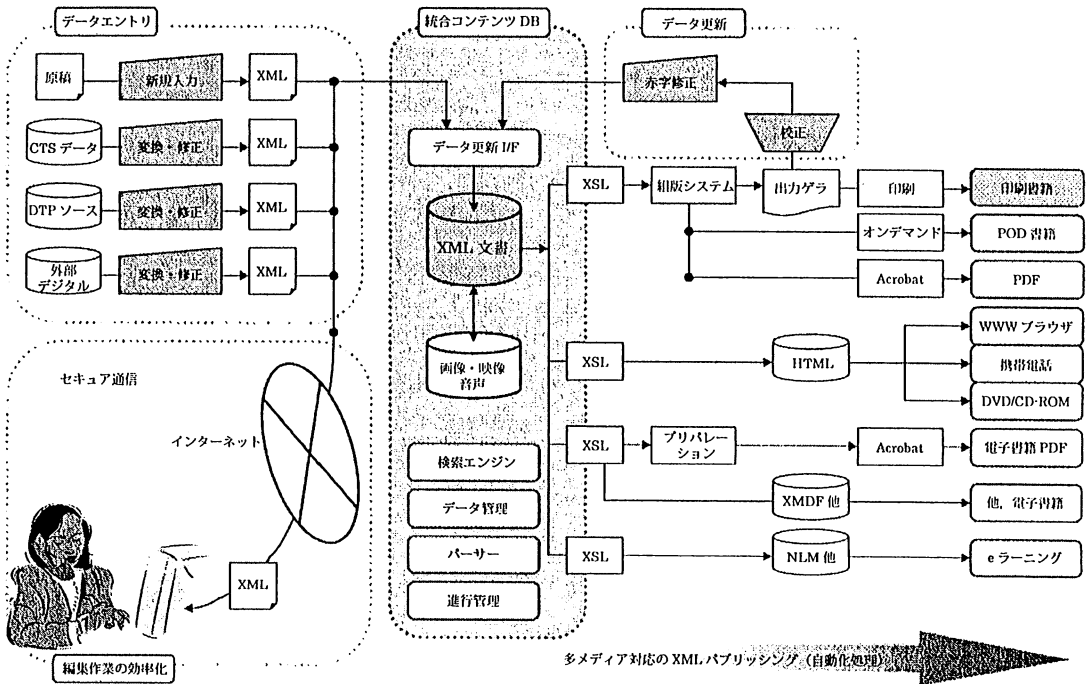


図4. 多メディア対応のXMLパブリッシングモデル

るのは、データの展開が容易に行える標準フォーマットによる XML データベースである。

入力側の対応は、紙原稿などからの新規入力による XML 化、既存の CTS データや DTP データからの変換、スキーマの異なる外部デジタルデータの取り込みなどを考えている。また、出力側は組版システム経由で、書籍印刷、オンデマンド印刷、PDF 出力のほか、Web 展開、CD-ROM・DVD-ROM などのパッケージ展開、電子書籍などへはスタイルシートを介して展開しており、最近では e ラーニングへの展開も出てきている。

## 5. 出版標準フォーマットの現状

### 5.1 出版標準フォーマットの現状

現在、出版業界における XML による標準フォーマットは、残念ながら存在しておらず、先に述べた通り XML 対応の組版システムが利用しているフォーマットは、いずれも独自のものである。電子出版分野においては、日本電子出版協会により 1999 年に策定された JepaX (バージョン 0.9)<sup>[4]</sup>があるものの、現状活発に利用されているとは言いがたい。また、JepaX が中立的な位置づけのデータ交換フォーマットであるため、一般書籍印刷用としては不足している要素があり、印刷会社でそれら不足分を追加する必要がある。

他のシステムで作成されたデータを利用する場合にも、各々のシステム毎に変換アプリケーションソフトを作成し、対応しなければならない。よって、できるだけ早く業界の出版標準フォーマットが決定され、負荷なくデータ流通がなされることが望まれている。

### 5.2 出版標準フォーマットの作成

当社では、このような状況を踏まえ、内部生産性の向上・コンテンツ二次展開へのスピードアップ・処理の自動化によるコスト削減・コンテンツマスターの一元化による品質保証・データ流通性の向上等の必要性から、出版標準フォーマットを作成した。この出版標準フォーマットは、出版コンテンツに対して、アクセス性はもちろんのこと、

正確性、拡張性などに加え、社内の展開性、及び業界のガイドラインとして利用することも考慮するとともに、JepaX や JIS X 4052<sup>[5]</sup>等、既存の標準フォーマットと互換性が取れるように考慮しながら作成した。

## 6. 出版標準フォーマット解説

### 6.1 概要

図 5. は、出版標準フォーマットの外観を示したものである。このフォーマットは、JepaX をベースにし、JIS X 4052, HTML 4.0<sup>[6]</sup>, OEB1.0<sup>[7]</sup>などを参照しながら拡張を行った。

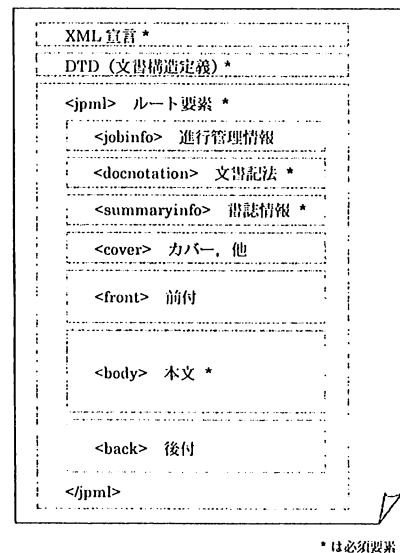


図 5. 出版標準フォーマットの外観

対象としているコンテンツは、現状は、一般書籍、学術論文誌、法律書など、編・章・節・項・段落・見出し、といった文書構造や、表現体裁がパターン化できるものに限定している。

また、図 4. に示した XML パブリッシングモデルを実現するために、コンテンツの論理構造と、スタイル指示の分離に努めており、スタイル指示は利用目的に応じ、スタイルシートで指示することを前提としている。要素は、論理構造要素・ブロック要素・インライン要素とがあり、要素数は、合計 102 種類となっている。

## 6.2 拡張した仕様

このフォーマットが、他の出版関連仕様と比較して強いところは、日本語文書の論理構造単位を JapaX の持つ 26 種類から 47 種類へ細分化している点や、索引バリエーションへの対応・組方向への順応・言語混在・欧文ハイフネーションなど数多くあり、これらの拡張は、主に出力側で組版処理を容易に行えるようにするという前提によるものが大きい。次に幾つかの拡張仕様を解説する。

### 索引バリエーションへの対応

日本語書籍における索引は、人名索引・分野別索引・地名索引など、様々な索引が存在する。また、索引の配列も、読み五十音順・姓名五十音順・姓名五十音+字画順・電話帳配列など、多くのバリエーションが存在する。これらに対応するためには、索引種類の別や、姓名であれば読みの姓・名の別、漢字表記の姓・名の別、漢字表記一文字に対応した読みなどを、何らかの形で把握できなければ機械的に行うことは不可能である。

今回、例 1. に示すように、索引項目になる本文中の文字列に対して、配列用の「読み」「表記」属性を持たせられるようにした。

#### 例 1. 索引配列を想定した人名のマークアップ

```
<name type="人名" reading="もり、た/いち、
ろう" file-as="森、田/一、郎">
森田一郎 </name>
```

```
<name type="人名" reading="もり/た、いち"
file-as="森/太、市">
森太市 </name>
```

```
<name type="人名" reading="もり/た、ろう"
file-as="森/太、郎">
森太郎 </name>
```

#### [配列イメージ]

姓名五十音順 (左) と、読み五十音順 (右)

|      |      |
|------|------|
| 森 太市 | 森 太市 |
| 森 太郎 | 森田一郎 |
| 森田一郎 | 森 太郎 |

### 組方向への順応

書籍の大幅な紙面変更や、コンテンツを二次利用する場合など、縦組から横組へ (またはその逆) など、組方向が変わる場合があり、本文中の「漢数字」と「算用数字」や、「上」と「右」など、表記文字そのものを変更しなければならない場合がある。このようなケースの対応負荷を軽減させるために、例 2. に示すような、組方向に応じたテキストの切替えができるようにした。

#### 例 2. 組方向に応じたテキスト切替え

```
<swtext>
<st type="ht">上の </st>
<st type="vt">右の </st>
</swtext> 表において…
```

```
<swtext>
<st type="ht">24 日 </st>
<st type="vt">二十四日 </st>
</swtext> の午後、中野区の路上で…
```

[組版イメージ] 横書き (左), 縦書き (右)

上の表において

24 日の午後、  
中野区の路上  
で

右  
の  
表  
に  
お  
い  
て

中  
野  
区  
の  
路  
上  
で  
二  
十  
四  
日  
の  
午  
後、

### 言語混在・欧文ハイフネーションの対応

日本語文章中には、欧文など、日本語以外の言語が混在する場合がある。日本語文字組版では、このような言語が混在する場合について、多くの組版ルールがあり、そのルールに基づいた細かな

対応が求められる。また、欧文組版では、単語が行末にかかる場合、分綴（ハイフネーション）と呼ばれている特別な処理を行う場合もある。

これらの対応を各出力側で行うと、その都度負荷が発生するため、今回、例 3. に示すように、言語の別（lang 要素）と欧文ハイフネーション位置（shy 要素）について文書構造の一部として捉え、これらの対応が自動的に行えるようにした。

例 3. 和欧文混在と欧文ハイフネーション

```
<div type="項 1">
  <head><title>つけぐすり </title></head>
  <p>
    付け薬
    <lang class="en">
      a medicine for ex<shy/>ternal
      application.
    </lang>
  </p>
</div>
```

[組版イメージ]

付け薬（つけぐすり） a medicine for external application.  
 告げ口（つけぐち）. tell (on); tell tales (about).

同一内容への対応

例 4. に示すように、名簿や財務諸表等の紙面によく見かける「同左」「同上」「〃」等の表記箇所について、表記テキストと、対応する元のテキストとを合わせて保持できるようにした。

これにより、名簿などにおいて都道府県名が同じ間は、先頭（およびページの変り目の先頭）のみ表示させるような体裁や、財務諸表を単年で利用したい場合など、データの取り扱いが容易になっている。

例 4. 名簿における同一内容への対応

```
<same-as>
  <sab> 東京都 </sab><sab> 〃 </sab>
</same-as> 千代田区
<same-as>
  <sab> 東京都 </sab><sab> 〃 </sab>
</same-as> 江東区
```

```
<same-as>
  <sab> 東京都 </sab><sab> 〃 </sab>
</same-as> 葛飾区
<same-as>
  <sab> 東京都 </sab><sab> 〃 </sab>
</same-as> 北区
```

7. 用途サイクル

図 6. は、出版標準フォーマット文書を使った用途サイクルとして、書籍制作の例を示したものである。

大本の出版標準フォーマット文書を、スタイルシートを使って変換し、組版システムの入力データを作成する。組版システムから組版ゲラが出力され、赤字指示があればデータ修正を行い、最終的にフィルムや CTP 出力を行う。

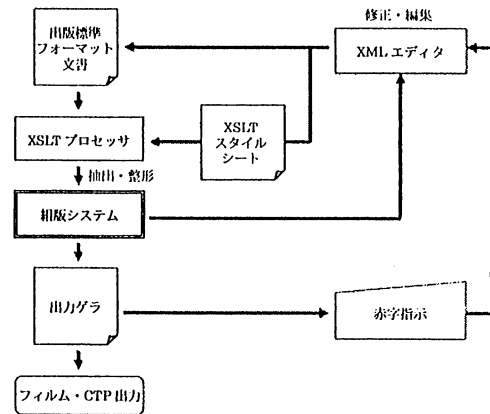


図 6. 用途サイクル概念図

このサイクルにおけるポイントは、データ修正が、常に元の XML 文書に対して行われることであり、これによってデータの一元化と、二次利用性を高めている。

また、出力ゲラ上の赤字指示に対する XML 上の修正箇所が特定しにくいいため、XML データの各要素と、出力ゲラの見・段・行を対応づけ、容易に特定できるようにし、作業効率の向上を図っている。

これらを実現するために、今回、組版システムの拡張と、XML エディタの開発を行った。

## 8. XML エディタの開発

今回、求めている XML エディタの要件は、

- 書籍データが入力できること
- 初期入力（エントリ）と、データ修正を同一のツールで行えること
- 様々なスキーマを把握できる汎用性を持っていること
- 制作体制の構築において、大多数へツールの展開が容易に行えること
- 進行管理システムなど、既存のシステムと親和性を高める I/F が取れるなど、拡張性があること

であり、これに対して市販製品も十分検討を行ったが、要件を満たす XML エディタがなく、社内開発し、検証を行った。

### 8.1 XML エディタの概観と機能

図 7. は、開発した XML エディタのメインウィンドウを示したものである。書籍制作に関わる対象入力者の平均スキルを踏まえ、XML 文書を直に編集していくシンプルなテキストエディタとし、Windows® の「Notepad」並みの軽快さでテキスト編集を行うことができる。更にキーバインド設定（入力キーと入力データの関連づけ）が行えるようにしてあり、タグパターン等を簡単な

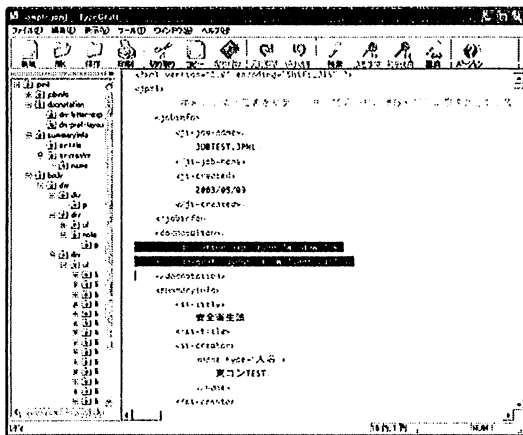


図 7. XML エディタ

キー操作で入力できるようにした。また、入力対象となるスキーマを読み込み、入力可能な文字の制限や、タグ属性値の選択肢表示、入力データの検証等、パーサー関連の機能を搭載し、入力精度が向上するようにした。

出力ゲラの赤字指示に対する XML データの修正箇所の特定は、組版システムが生成する関連づけの情報を XML エディタが読み込み、エディタ上で赤字箇所の頁・段・行を指定することで、編集画面が自動的スクロールし、該当箇所にジャンプするようになっている。

## 9. 検証・評価実験

これまで述べた出版標準フォーマット文書を使った用途サイクルと XML エディタについて、運用面での課題や問題点を把握検証すべく、書籍制作入力者 1 名に参加してもらい、約 2 ヶ月間にわたり評価実験を行った。

### 9.1 評価実験環境

実験に使用した環境は、XML データ入力と、赤字修正作業は、Windows® のパーソナルコンピュータを使って行った。

組版システムは、社内で保有するシステムを用い、実験用のサンプルデータはパターン化しやすい構造を持ち、且つ、組版体裁が比較的高度なものを選定し、データ量は、出力ゲラで約 100 ページ分のデータを使用して行った。

### 9.2 新規データ入力作業

新規データの入力作業について、次に示す三つの方法で実施し、作業負荷・作業効率・入力精度等の観点から比較を行った。

- 原稿を見ながら、入力者がタグを判断し、タグとテキストを同時に入力する方法（方法 1）
- 事前に原稿にタグを記入しておき、入力者は原稿通りにタグとテキストを同時に入力する方法（方法 2）
- 先にテキストを入力してから、後でタグを入力する方法。（方法 3）

表 1. に、比較結果を示した。方式 3 のテキストとタグを別々に入力する方式が最も効率の良い結果となった。

表 1. 新規データ入力における入力方式の比較

| 方式   | 負荷 | 効率 | 精度 | 総合評価 |
|------|----|----|----|------|
| 方式 1 | ×  | ×  | △  | ×    |
| 方式 2 | △  | △  | ×  | △    |
| 方式 3 | ○  | ○  | ○  | ○    |

○ 良い △ふつう ×悪い

### 9.3 赤字修正作業

赤字ゲラを修正原稿とし、段落内のテキスト挿入・削除・移動をはじめ、文書構造の階層の移動、別紙原稿の挿入、ルビをはじめとしたインライン要素の変更、表の挿入など、修正作業を行った。新規入力と比較すると、赤字修正作業のほうは、作業負荷が少なく、XML データを直接修正する作業も十分対応可能であった。これは、既にデータ中に参照可能なタグが入力されていることにより、新規入力よりもタグの仕様を意識する必要が少ないためであると思われる。なお、ルビや数式などの、細かなタグ付けが必要な要素の修正では、入力者の思考する傾向が強くなり、多くの時間を必要とした。

### 9.4 XML エディタの評価

現在、書籍制作では、データ入力・修正ツールとして、ワードプロセッサやテキストエディタを使用しており、今回の XML エディタの操作に対する使いづらさはなかったが、画面上でタグとテキストを見ながらの編集になるため、入力者が、現在編集中の階層位置などの、編集状況が分かりづらさがあった。よって、もっと入力者の思考を促さない、作業に集中できるユーザーインターフェースへ、改良すべきであると思われた。また、入力エラーに関して、パース機能を搭載し、エラーの防止を図ったわけだが、入力者に対し、具体的なエラー箇所の位置・原因・回避方法等が明確に通知されないメッセージが多く、入力者が

その場で対応を考えなければならず、一部の箇所では、対処方法が判断できず、作業効率を低下させる要因となった。

### 9.5 出版標準フォーマットの評価

現在の仕様では、文書構造の定義に論理構造単位の div 要素を多様することになってしまった。div 要素が type 属性を使って階層意味を表すため、今回のようなタグとテキストを同時に表示する XML エディタでは、視覚的に階層が判りづらくなってしまった。また、プログラムによるデータ検証も負荷が高くなるということが判明したため、この点に関しては、フォーマットそのものの再考が必要であると思われる。

### 10. おわりに

今回の検証・評価実験で、出版標準フォーマットを中心とした用途サイクルをシミュレーションし、有効性や合理性を幾つか確認できた。実運用のためには、フォーマットの最適化や、XML エディタの機能拡張など、課題も把握できた。これらは引き続き検討していきたい。あわせて、進行管理システムへの組み込みや、コンテンツデータベースとの連携等、検証を行っていく予定である。なお、この出版標準フォーマットでは、今後検証がまとまった段階で、多方面から評価していただく意味でも、オープン化したいと考えている。

### 参考文献

- [1] 総務省. インターネット接続サービスの利用者数等の推移【平成 15 年 7 月末現在】(速報). 8, 2003.
- [2] 総務省. 情報通信統計データベース. 情報通信白書. 平成 15 年度版. 5, 2003.
- [3] 総務省. 情報通信統計データベース. 情報通信白書. 平成 14 年度版. 5, 2002.
- [4] 日本電子出版協会. JepaX. 3, 1999.  
<http://www.jepax.org/>
- [5] 日本工業規格. 日本語文書の組版指定交換形式. JIS X 4052, 10, 2000
- [6] WORLD WIDE WEB consortium, HTML 4.0 Specification, 4, 1998
- [7] Open eBook Forum, Open eBook Publication Structure Specification Version 1.2, 8, 2002.

## 「情報知識学会誌」投稿規定

2002年8月27日 制定

2003年3月19日 一部改訂

0. 情報知識学会誌編集規程による本会機関誌「情報知識学会誌（以下、会誌という）」への投稿に関する事項は、この規定の定めるところによる。

### 1. 投稿資格

投稿者の少なくとも1人は本会員でなければならない。ただし、編集委員会による依頼原稿の場合にはこの限りではない。

### 2. 投稿原稿

2.1 広い意味での情報知識学に関連し、またその発展に貢献するもの（情報／知識の収集、整理、蓄積、検索および各種解析、利用などに関するもの）とする。刊行時において未発表の原著でなければならない。本会誌の記事の種類を以下に示す。

2.2 投稿者は会誌記事の種類を明記して投稿しなければならない。ただし、編集委員会を変更することがある。

- (1) 研究論文 (Research Paper) : オリジナルな研究論文で、内容の主要な部分が学術論文として他に公表されていないもの。
- (2) 事例／調査報告 (Report) : 情報知識学に関連したシステムなどの開発、利用、調査に関するもの。資料も含む。
- (3) 解説／展望 (Review) : 情報知識に関連した特定分野の論文や学説などを総括、解説、紹介、あるいは技術動向などを展望したもの。技術、研究上の処理、解析方法などに関する解説。
- (4) 論談 (Proposal Paper) : 情報知識学に関連した新たな意見の表明、提案など。
- (5) 討論 (Discussion) : 本会誌に掲載された論文についての学術的な討論。
- (6) 研究速報 (Notes) : 技術、手法、新事実などの簡単な報告。
- (7) 講座 (Lecture) : 情報知識学の各分野に関する基礎理論、技術の適用などについて、テーマを定めて系統的に説明するもの。
- (8) 学会記事 (News) : 本会の事業、運営などの報告、記事、資料など。
- (9) ニュース、お知らせ (News) : ニュース、お知らせ。最近刊行された単行本やモノグラフの紹介。
- (10) 講演 (Lecture) : 特別号などにおける講演資料。
- (11) その他 : 編集委員会が適当と判断したもの。

2.3 会誌記事の種類のうち、(1) から (6) までは査読を行う。その他については編集委員会編集を行う。

### 3. 投稿原稿

#### 3.1 原稿の形式

##### (1) 投稿時の原稿

以下の A、B のいずれかの体裁でプリントされたワープロ原稿（横書き）4部の提出とする。その他、執筆に関する詳細は「執筆要領」を参照のこと。



A 刷り上り原稿を想定したレイアウト (A4判, 2段組, 20字 × 46行 × 2段).

図, 表は希望の位置に配置すること.

B ベタ打ち原稿 (A4判, 40文字 × 40行).

図, 表は, 1枚ずつ別の用紙に印刷すること. ベタ打ち原稿右余白に図表の挿入位置を朱書きすること.

(2) 採択決定後の原稿

以下の C, D の両方の形式で記録された電子媒体1部の提出とする.

原稿の送付にあたってはフロッピーディスクなどの適当な電子媒体とする. 詳細は提出時に事務局に相談のこと.

C Microsoft Word, 一太郎, DVI, PDF などの代表的なフォーマット.

D 図, 表は充分な品質で印刷できる形式 (JPEG, GIF など).

### 3.2 原稿の制限

(1) 原稿の長さを原則として次のように制限する.

研究論文, 事例/調査報告, 解説/展望, 論談 : 刷り上がり 20ページ以内

討論, 研究速報, 講座 : 刷り上がり 6ページ以内

ニュース他 : 刷り上がり 2ページ以内

(2) 図原稿 (原図) の大きさは A3判を越えないものとする.

(3) 原則として, 図版も含めてモノクロ印刷とする. ただし, カラーでなければならぬ図版を使用する場合は, 別途編集委員会と相談する. なお, カラーページやページを超過する分については, 印刷費を著者の全額負担とする.

(4) 使用言語は日本語または英語とする.

### 4. 原稿の採否

投稿原稿の採否は, 専門家による査読の後, 編集委員会において決定する. 不採択となった原稿は, 編集委員長より理由を付して通知する.

### 5. 査読のプロセス

学会員の中から編集委員会が指名した査読者2名によって査読を行う. 内容によっては, 編集委員会は著者に照会し, 原稿の修正を求めたうえで, 再査読を行うことがある.

### 6. 校正のプロセス

採択が決定した投稿原稿は, 掲載原稿として著者に校正を依頼する. 著者による校正は原則として1回とする. その際, 字句の修正以外は原則として認めない.

### 7. 別刷

別刷 (抜刷) は著者の実費負担とする. 希望部数を事務局に申し出ること.

### 8. 投稿の手続き

原稿投稿時には下記の書類を添え, 原稿送付先に郵送する.

#### 8.1 必要書類

(1) 最初の投稿時 (a, b, cの3つが必要です.)

a. 投稿原稿整理カード: ホームページからコピーして, 必要事項を記入し, 印刷したものを1部. 掲載原稿整理カードと兼ねるので, コピーを保存しておくこと.

b. 紙媒体の原稿 (図, 表を含む): 4部.

なお, 投稿者は著者校正用に原稿のコピーを保存しておくこと.

c. E-mail による連絡票

- ・論文種別, 標題 (和文, 英文), 著者名 (和文, ローマ字), 所属機関/住所 (和文, 英文), 要旨 (和文, 英文), キーワード (和文, 英文), 刷り上り予定ページ数
- ・連絡先: 著者 1 名の連絡先 (氏名, 所属機関/部局, 所属機関住所, 電話番号, Fax 番号, E-mail アドレス)

なお, 投稿後の連絡は主として E-mail で行う。

(2) 採択決定後の投稿

- a. 掲載原稿整理カード: 投稿時のカードに追加事項を記述し, 印刷したものを 1 部。
- b. 3.1(2) に指定した電子媒体: 1 部。
- c. 3.1(2) の C の印刷出力 (プリントアウト): 1 部。

8.2 原稿の送付先

〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内  
情報知識学会事務局  
電話: 03-3835-5692 Fax: 03-3837-0368  
E-mail: jsik@nifty.com

8.3 原稿の受付

事務局が原稿を受け取った日を受付日とする。受付の確認を 1 週間以内に投稿者の連絡先に E-mail で通知する。不備のある投稿原稿は返送し, 再提出するものとする。

9. 原稿提出期日

投稿は随時とする。ただし, 特集号などは除く。

10. 著作権

- 10.1 機関誌『情報知識学会誌』に掲載された論文 (電子版を含む) の著作権 (著作財産権, copyright) は情報知識学会に帰属する。
- 10.2 掲載論文は冊子による出版の他, 電子的に蓄積し, 本会が行う情報提供サービスなどを通じて公開する。
- 10.3 本学会誌に掲載された執筆内容が第三者の著作権を侵害するなどの指摘がなされた場合には, 執筆者がその責任を負う。

11. 規定の改訂

- 11.1 本規定の改訂は, 編集委員会の議を経て, 理事会の承認を得なければならない。

12. 施行

- 12.1 本規定は 2002 年 8 月 27 日より施行する。
- 12.2 本規定の施行により, 現行規定 (第 4 版 (暫定板) 2002 年 3 月) は廃止する。

13. 改訂履歴

2003 年 3 月 19 日 一部改訂。「10. 著作権」に, 10.3 項を追加。

## 「情報知識学会誌」執筆要領

2002年8月27日 制定

2003年5月2日 一部改訂

### 1. 一般的な事項

本会誌への投稿は、「投稿規定」に従い、投稿原稿は本執筆要領に従って作成されなければならない。

本会誌の投稿原稿の種類には、研究論文、事例／調査報告、解説／展望、論談、討論、研究速報、講座、本会記事、講演、ニュース、その他がある。

### 2. 日本語原稿の構成

#### 2.1 全体構成

##### (1) 第1ページ（査読者には見せない）

- ・ 標題（和文および英文）
- ・ 著者名（和文およびローマ字、ローマ字による著者名は、名、姓の順で、姓は全て大文字を使用する。）
- ・ 所属（和文および英文による所属機関名）
- ・ 住所（和文による所属機関の住所、E-mail、脚注とする。）
- ・ 見出し用原稿（研究論文、事例／調査報告、解説／展望、論談の原稿には、刷り上がりページ上部欄外につける著者名および標題を30字以内で書く。）

##### (2) 第2ページ目以降（査読者に見せる）

- ・ 要旨（研究論文、事例／調査報告、解説／展望、論談の原稿には、和文および英文で要旨をつける。和文要旨の長さは400字以内とする。英文要旨の長さは200語以内とする。要旨中には、図、表、数式などを用いない。本文中の図、表、数式、文献などを番号で引用しない。）
- ・ キーワード（研究論文、事例／調査報告、解説／展望、論談、討論、研究速報、講座にはキーワードをつける。和文および英文でそれぞれ5個程度、和文と英文のキーワードは、対応することが望ましい。キーワードはカンマ（,）で区切る。）
- ・ 本文（和文または英文）
- ・ 文献、付録など（和文または英文）
- ・ その他（とくに長い論文の場合、読者の便宜を考えて内容目次を付してもよい。ただし、章、節の見出し程度とする。）

#### 2.2 本文（Body）

##### (1) 構成

章、節などの構成は、第1レベルは1, 2, …, 第2レベルは1.1, 1.2, …, 第3レベルは1.1.1, 1.1.2, …のようにする。

##### (2) 脚注

脚注はできるだけ避ける。止む無く使用する場合は簡潔な文とする。

##### (3) 図および表

- a. 図、表にはそれぞれ通し番号をつける。図1 (Fig.1), 図2 (Fig. 2), …表1 (Table 1), 表2 (Table 1), …など。
- b. 通し番号とともに説明文（キャプション）をつける。キャプションの位置は図は下部に、表は上部とする。

(4) 数式, 化学式

- a. 数式 (独立式), 化学式は, 段落外で記述されているものも本文中で一回は参照する.
- b. 数式には, 通し番号を振る.

(5) リスト (または箇条書き)

- a. 記号なしリスト.
- b. 記号つきリスト. リストの記号は, 数字, アルファベット, 記号を用いることができる. ただし, これらの混在した使用は避ける. アルファベットは1 論文中では大文字, 小文字の使い分けをしない.
- c. 複雑化を避け, せいぜい2 段 (親子関係) のリストとし, ネストを跨ぐ順序づけを用いない.

(6) 注記および参考文献

本文中で少なくとも一回は参照すること. 通し番号で参照し, タイトルなどでの参照は避ける.

2.3 後付け (End)

(1) 謝辞

本文の最後に続けて記述する. 章番号は用いない. 章題は「謝辞」とする. 最終原稿時に記述することが望ましい.

(2) 注記および参考文献

- a. 注記または参考文献には, 参照順に通し番号を付し, 本文の最後に番号順にまとめて記述する. 章番号は持ちない. 章題は「参考文献」とする.
- b. 1 つの番号には1 つの注記または参考文献を対応させる.
- c. 注記中には参考文献を含めない. 注記はできる限り簡潔に表現すること.
- d. 参考文献の記述形式は, 以下の形式を満たさなければならない.
- e. URL を参照してもよいが, 移動または削除される可能性があるため, 極力避ける. 原著が URL でのみしか参照できない場合など, やむをえない場合は用いてもよい. その場合, 参照時点でのハードコピーを保管しておくなど, 参考文献へのアクセス手段を確保するよう努力しなければならない.

【参考文献の形式】

1. 雑誌中の1 論文

[引用通し番号] 著者名: 論文名, 雑誌名, 巻号, 掲載ページ, 出版年, その他.

2. 図書1 冊

[引用通し番号] 著者名: 書名, 版表示, 出版地, 出版社, 総ページ数, 出版年, その他.

3. 図書の1 部

[引用通し番号] 著者名: 論文名, 書名, 版表示, 出版地, 出版社, 掲載ページ, 出版年, その他.

4. 会議報告

[引用通し番号] 著者名: 論文名, 書名 (会議名), 版表示, 編集者名, 会議開催地, 会議開催年, 会議開催機関, 出版地, 出版社, 掲載ページ, 出版年, その他.

5. インターネット上の論文

[引用通し番号] 著者名や標題など可能な限り詳細な書誌事項, URL, 参照年月日. (単なるホームページなどは参考文献にしないこと).

【参考文献の記述】

1. 著者名，編集者名の記述

- (1) 個人著者名は，姓，名の順に記述する．欧文著者名は，カンマ（，）で姓，名を区切る．
- (2) 複数著者の場合は，各著者をセミコロン（；）で区切る．
- (3) 翻訳図書などの翻訳者名の場合は，著者名の後に括弧（）に入れて記述する．

2. 論文名，書名の記述

- (1) 論文名，書名は，和文の場合はかぎ括弧（「」），欧文の場合はダブルクォーテーション（“ ”）に入れて記述する．
- (2) 図書中の一部を引用した場合の書名は，和書の場合は二重かぎ括弧（『』）に入れ，欧文の場合はイタリック体で記述する．

3. 掲載ページの記述

- (1) 論文の場合は，開始ページと終了ページを記述する．「pp. 開始ページ-終了ページ」とする．
- (2) 図書の場合は，総ページ数とする．「総ページ数 p.」とする．

【参考文献の記述例】

- [1] 藤原譲：「情報知識学試論」，情報知識学会，Vol.1, No.1, pp.3-10, 1990.
- [2] 原正一郎；安永尚志：「国文学研究支援のための SGML/XML データシステム」，情報知識学会，Vol.11, No.4, pp.17-35, 2002.
- [3] Fujiwara, Shizuo: “East-West Communication and Information Transfer — Coordination of Specificity”, Journal of Japan Society of Information and Knowledge, Vol.4, No.2, pp.11-18, 1994.
- [4] Ellis, David (細野公男監訳，斎藤泰則，鈴木志元，村上泰子訳)：「情報検索論」，丸善，180p., 1994.
- [5] 根岸正光：「学術情報の流通と利用」，『情報学とは何か』情報学シリーズ3，丸善，pp.43-69, 2002.
- [6] 名和 小太郎：「デジタル図書館と著作権」，デジタル図書館，No.4,  
[http://www.dl.ulis.ac.jp/DLjournal/No\\_4/nawa/nawa.html](http://www.dl.ulis.ac.jp/DLjournal/No_4/nawa/nawa.html) (2002年8月27日参照)

3. 文章と文体

- 3.1 文体はひらがなと漢字による口語常態（である調）とし，現代かなづかいを用いる．
- 3.2 漢字は当用漢字とする．ただし，固有名詞や学界で広く用いられている慣用の術語はこの限りではない．
- 3.3 句読点その他には「，」「。」を用いる．
- 3.4 本文中の人名には敬称をつけない．ただし，謝辞の人名はこの限りではない．
- 3.5 数量を表す数字はアラビア数字とする．
- 3.6 数式は印刷に便利なように十分注意して記号を記すこと．原則として数量（変化量）を表す記号はイタリックとする．
- 3.7 ローマ字の人名の姓は大文字体とする．
- 3.8 固有名詞で読み誤るおそれのあるものにはふりがなをつける．
- 3.9 英数字は原則として半角英数字で記述する．

4. 英文原稿

英文による投稿原稿の場合も、原則として和文による投稿原稿の諸規定に従う。英語圏以外の著者の場合、著者名表記にその国語による表記を認めるが、可能な限り英文表記とする。

4.1 研究論文、事例／調査報告、解説／展望、論談、討論、研究速報などの原稿は英文でもよい。

4.2 英文原稿は語学的に難点の少ないものであることを必要とし、著者の責任において完全を期する。

4.3 英文原稿には、英文による要旨 200 語程度、ならびに日本語による 400 字以内の要旨をつける。ただし、著者が日本語を理解できない場合は日本語要旨を省略できる。

5. その他

原稿は和文または英文によるものとする。文章は語学的に難点の少ないものであることとし、著者の責任において完全を期する。編集委員会は語学的校正を行わない。

6. 要領の改訂

6.1 本要領の改訂は、編集委員会の承認を得なければならない。

7. 施行

7.1 本規定は 2002 年 8 月 27 日より施行する。

8. 改訂履歴

2003 年 5 月 2 日一部改訂。英語要旨の長さを 500 語から 200 語に変更。図、表のキャプション位置を訂正。

様式1 / 様式2

## 「情報知識学会誌」投稿原稿整理カード / 掲載原稿整理カード

1. 論文種別 (投稿規定の 2.2 からお選びください。)
2. 標題 (和文)  
   標題 (英文)
3. 著者名 (和文, ローマ字)
4. 所属機関名 (和文, 英文)
5. キーワード (和文, 英文)
6. 連絡責任者 (1名)  
   氏名, 所属機関 / 部局, 同住所, 電話番号, Fax 番号, E-mail アドレス
7. 送付投稿原稿  
   テキスト部分の枚数  
   図の枚数  
   表の枚数  
   付録の枚数
8. 図の返却希望 (YES, NO)
9. カラー図の有無 (原則として認めていないが, カラーでなければならない場合など, 全額実費著者負担)
10. 投稿日

---

11. 登録番号
12. 受付日 (再受付日)
13. 受理日
14. 送付掲載原稿  
   フロッピーディスクなど 枚数  
   プレインテキストプリント 枚数 (刷り上がり見本, 図表なども貼り込んだもの)
15. 別刷り (抜刷り) の希望部数 (全額実費負担)

### 【投稿原稿整理カード】

1. 標題を「様式1 投稿原稿整理カード」とし, 1 から 10 項目 (11 項目以降は採択後) をもれなく A4 判横書き 2 枚程度に, ワードプロでお作りください。
2. 投稿原稿と一緒に送下さい。
3. なお, 投稿時には「投稿規定」にある E-mail による連絡票もお忘れなく, お送り下さい。

### 【掲載原稿整理カード】

1. 標題を「様式2 掲載原稿整理カード」とし, 全項目をもれなく A4 判横書き 2 枚程度に, ワードプロでお作り下さい。掲載原稿と一緒に送下さい。

お知らせ [論文募集]

情報知識学会第12回(2004年度)研究報告会 発表論文募集について

実行委員長 後藤 智範(神奈川大学 理学部)

情報知識学会では下記の期日・場所で、総会と共に研究報告会を開催する予定です。

期日: 平成16年5月22日土曜(予定)

場所: 都内(具体的な場所は未定)

研究報告会の発表論文を下記要領で募集いたしますので、学会員の皆様どうぞ奮ってご応募ください。

1. 募集分野

- (1) 情報知識の構造解析、モデル化、可視化、知識発見
- (2) 情報・知識の表現、生産、組織化、検索、提供
- (3) 電子出版、電子図書館
- (4) マルチメディア、電子ミュージアム
- (5) 用語、シソーラス
- (6) 知識情報の流通と知的所有権
- (7) 専門分野における品質管理、基準化
- (8) その他情報知識学に関連する諸研究・開発

2. 応募方法

発表論文題目、著者名(連名の場合、登壇発表者に○印)、所属、論文概要(200字以内)、連絡代表者の氏名、住所、電話・FAX番号、電子メールアドレスを明記の上、下記宛になるべく電子メールにてお申し込みください。

〒259-1293 平塚市土屋2946 神奈川大学 理学部 情報科学科 後藤 智範  
TEL: 0463-59-4111(ex. 2702) FAX:0463-58-9684  
E-mail: gotoh@info.kanagawa-u.ac.jp

応募期間 : 2004年2月1日 ~ 2004年3月20日(予定)  
採択可否通知 : 2004年2月20日 ~ 2004年4月1日(予定)  
発表論文受付期間 : 2004年3月20日 ~ 2004年4月20日(予定)

3. 論文執筆・発表について

- (1) 採択されたものについては論文執筆要領の詳細を別途お知らせいたします。発表論文はA4版で2~6頁です。4ページまでは無料ですが、それを超えると有料(2ページ2000円)になります。
- (2) 発表時間は、質疑応答を含めて30分を想定しています。なお、論文提出がないと発表はできません。
- (3) 登壇発表者は当学会員に限ります。当日入会も可能です。



情報知識関連新刊図書一覧 2003年7月～9月

- 『よくわかる情報システム&IT業界』 新井進編著 日本実業出版社 2003.7 1300円  
『Namazuシステムの構築と活用 改訂』 馬場肇著 ソフトバンクパブリッシング 2003.7  
2800円  
『情報技術革新の経済効果』 篠崎彰彦著 日本評論社 2003.7 3300円  
『大人のための「情報」教科書』 坂村健/清水謙多郎著 数研出版 2003.7 1500円  
『情報セキュリティ事典』 土居範久監修 佐々木良一ほか編 共立出版 2003.7 25000円  
『情報社会を理解するためのキーワード1』 国領二郎ほか共編 培風館 2003.7 4200円  
『情報社会を理解するためのキーワード2』 細野公男ほか共編 培風館 2003.7 3500円  
『情報社会を理解するためのキーワード3』 神沼靖子ほか共編 培風館 2003.7 3000円  
『アルゴリズム論』 浅野哲夫/和田幸一共著 オーム社 2003.7 2800円  
『知のオアシス』 日本医学図書館協会編 紀伊国屋書店(発売) 2003.7 1429円  
『心の科学と哲学』 戸田山和久ほか編 昭和堂 2003.7 2500円  
『ユビキタス・コンピューティング』 美崎薫著 ソフトマジック 2003.8 1500円  
『生命複雑系からの計算パラダイム』 大内東ほか著 森北出版 2003.8 3400円  
『図書館新世紀』 国立国会図書館関西館編 日本図書館協会 2003.8 1500円  
『コンピュータ科学者がめったに語らないこと』 ドナルド・E. クヌース著 滝沢徹訳  
エスアイビー・アクセス 星雲社 2003.9 2800円  
『チャンス発見の情報技術』 大沢幸生監修著 東京電機大学出版会 2003.9 4600円  
『コンピュータはどれほど賢いのか』 舩本昇竜著 すばる舎 2003.9 2400円  
『マルチメディア表現と技術』 小館香椎子/柳沼良知編著 丸善 2003.9 2800円  
『アメリカ公立図書館で禁じられた図書』 イーヴリン・ゲラー著 川崎良孝訳  
京都大学図書館情報学研究会 日本図書館協会 2003.9 5000円  
『eラーニングが創る近未来教育』 先進学習基盤協議会編著 オーム社 2003.9 1600円

(注) 図書館流通センター刊『週刊新刊全点案内』より抜粋。コンピュータの技術・操作関連のもの、初心者入門的なもの、またビジネス色の強いものは除外した。

制作者：平田 周

情報知識学会誌 第13巻4号 (第8回 SGML/XML 研修フォーラム特別号)

2003年10月30日 発行 頒布価格 3,000円

発行者 細野 公男

編集者 根岸 正光 川上 能徳

発行所 情報知識学会 〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内

©2003, Japan Society of Information and Knowledge

TEL : 03(3835)5692 FAX : 03(3837)0368

E-mail:jsik@nifty.com URL : <http://www.jsik.jp>

この特別号は、情報知識学会が平成 15 年 10 月 30、31 日に東京国際フォーラムにて開催した第 8 回 SGML/XML 研修フォーラムのテキストとして使用されました。