

博物館資料に対するXMLメタデータスキーマの提案

2003MT027 岩田光将 2003MT036 加藤真一

指導教員 河野浩之

1. はじめに

Web システムを構築する各種技術,特にコンテンツ交換技術の発展に伴いデジタルアーカイブの構築が進んでいる.図書館においては,台湾で NDAP というデジタルアーカイブプロジェクトを進めている[1].しかし博物館においてはこうした電子化が進んでおらず,メタデータの共通化の定義を確立しようと試みられているのが現状である.定義が確立していないため,電子化を行っている博物館においても検索方法が各館で異なり,同じ資料でも名称が異なるという問題が出てくる.検索方法が異なることで,図書館においては各館同士で連携が行われているが,博物館においてはこの連携が難しくなる.そこで本研究ではこれらの問題に注目し,博物館資料に対してメタデータを付加するためのXMLメタデータスキーマの提案を行う.

2. メタデータフォーマットの動向

この章ではメタデータフォーマットの動向として,CIDOC,ダブリン・コア,MPEGについて述べていく.

2.1 CIDOC

CIDOCとは博物館コレクションのドキュメンテーション専門の機関である.主にCIDOC CRMモデリング,CIDOC情報カテゴリ等の開発やガイドラインの作成を行っている[2].

- Acquisition Information (取得情報)
- Condition Information (状態情報)
- Deaccession and Disposal Information (償却・処分情報)
- Description Information (記述情報)
- Image Information (画像情報)
- Institution Information (組織情報)
- Location Information (所在情報)
- Mark and Inscription Information (記号・刻印情報)
- Material and Technique Information (材料・技法情報)
- Measurement Information (測定情報)
- Object Association Information (資料関連情報)
- Object Collection Information (資料コレクション情報)
- Object Entry Information (資料受入情報)

- Object Name Information (資料名情報)
- Object Number Information (資料番号情報)
- Object Production Information (資料制作情報)
- Object Title Information (資料タイトル情報)
- Part and Component Information (部品・構成情報)
- Recorder Information (記録者情報)
- Reference Information (参照情報)
- Reproduction Rights Information (複製権情報)
- Subject Depicted Information (主題描写の情報)

この情報グループにおいて,「画像情報」や「資料関連情報」は詳しい内容がないため,情報が入れられないという問題点がある.

2.2 ダブリン・コア

Dublin Core Metadata Initiative が制定した,インターネット上にある多様な情報資源の記述に共通して使えるようデザインされたメタデータ規則である.エレメントと呼ばれる15個の基本項目が存在する.ただ機能を増やすことよりも,限定された少数の要素を標準化し,現実に見えるメタデータを目指している点が特徴である.ダブリン・コアは柔軟性と利便性を意識して,全てのエレメントをオプションとしている.つまり,自分で使いたい分のエレメントを利用できる.博物館資料に対しては適切ではないが利用はできるため,本研究ではダブリン・コア15項目のうち博物館資料の定義に必要な項目を自分たちで選び実際に使用していく.

2.3 MPEG-7

MPEG-7 (Moving Picture Experts Group), ISO/IEC 15938 は,マルチメディア用メタデータ表記方法の国際標準化規格である.MPEG-1,2,4 などとは異なり,動画データのエンコードが目的ではなく,XMLをベースとしたメタデータ記述によるマルチメディアデータの高速な内容検索を目的としている.

2.4 MPEG-21

MPEG-21, ISO/IEC 21000 は将来のマルチメディア流通を想定して,著作権保護やコンテンツ保護を中心とした規格である.MPEG-7と同様,動画の圧縮を定めた仕様ではない.メタデータを用いたアプリケーションは一般にそれ

を生成する工程およびそれを利用する工程に大別され、両者の間を標準化された表記方法が仲介するという形式でモデル化される。MPEG-21 は標準化対象とした標準表記方法である。換言すれば、MPEG-21 は特徴データの抽出方法やその利用法については規定していない。これは MPEG-21 が、これら特徴抽出や利用法は今後も更なる発展の余地があると考えられているからである。

3. XML メタデータスキーマの提案

博物館ではメタデータの共通化について確立した定義がまだ成されていないのが現状である。22 の情報グループを提案しているが、これは基本的なこととこれだけでは不足してしまう部分がある。そこで本研究では、実際に定義されているメタデータスキーマの一部を組み込むことで、不足していた部分を補うことを提案した。

3.1 MPEG を用いた要素の提案

博物館の資料を動画や静止画で納めて、それらのデータをメタデータスキーマに格納することも本研究課題の1つである。実際には第2章で挙げた MPEG-7, MPEG-21 の規格を参考にして、メタデータスキーマの構築に利用した。MPEG-7, 21 は動画や静止画などのマルチメディアコンテンツを表記するための国際基準になっている。これらの規格を参考にすれば、どのような動画や静止画等のマルチメディアコンテンツが用いられたときでも全てにおいて対応できるスキーマを作成することが出来るからである。

3.2 MODS を用いた要素の提案

本研究の言語フィールドは MODS を参考としている。MODS とは図書館での利用を重視したメタデータスキーマである。XML 言語で記述可能で、全く階層を持たないダブリン・コアのような非常に簡易な形式とも違った新たな規格である[3]。

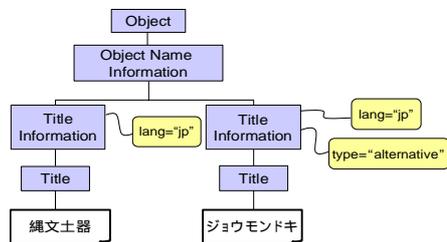


図1 多言語フィールドを用いた木構造

実際に資料名情報<Object Name Information>で使われている他言語フィールドの木構造を図1に示す。これは<Object Name Information>要素が複数の子要素<Title

Information>を持っている。この<Title Information>は属性<lang>と<type>を有しており、この属性でその要素の言語やタイプを指定する。

3.3 CMS と XML スキーマの連携

図2において、博物館資料に対しXML化を行い、においてXML化したデータをデータベースに保存する。データベースとして格納すれば、のようにデータベースとの間でデータのやり取りが可能となり、好きなときにデータの追加・削除が可能となるため、資料の整理が容易となる。また、のようにWebサーバを介することで資料や語句の検索を行なうことができ、いつでもすばやく欲しい情報を得ることが出来る。さらに、においてWebサーバ同士が連携しあうことで他の博物館との情報交換を行うこともでき、博物館の質を高めることが出来る。

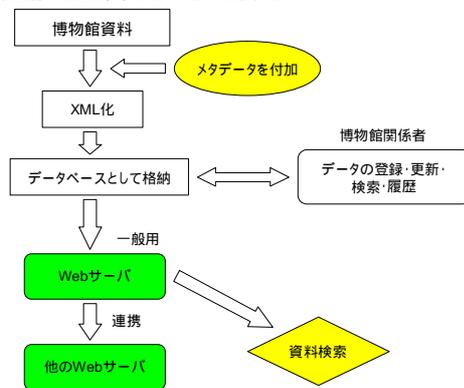


図2 XMLメタデータスキーマの利用法

また、ではCMSを利用することができる。CMS (Contents Management System) とはWebサイトを構築・編集するソフトウェアである。「博物館WebサイトのCMS利用によるユーザビリティ向上」において、CMSのデータベースを拡張しスキーマを導入することで、メタデータを入力できるようにする[4]。そうすることでWebサイトの管理を容易にし、詳細なデータ管理が出来るようになる。このように、XMLメタデータスキーマを提案することで様々な利点が生まれ、博物館を利用しやすくなる。

4. 南山大学人類学博物館との検討

本研究では、南山大学人類学博物館が所蔵する50点の博物館資料のデータを使用する。2006年8月3日と12月6日に人類学博物館においてスキーマの項目についての打ち合わせを行なった。スキーマの項目の基本となるのは、CIDOCが定義している22の情報グループで、その中から付加出来るデータ、出来ないデータを検証し、必要のない部分の削除、変更、情報の追加を行った。例えば、

「Condition Information(状態情報)」というものがあがる、この状態情報には以下の情報がある。

- Condition (状態)
- Condition summary (状態要約)
- Condition date (状態日付)

この情報に新たに「処理者」という情報を付け加え、状態情報の内容を 4 つに変更した。また、新たに「貸出し情報(履歴)」という情報を作成し、以下の内容にした。

- 貸出し日付
- 運搬方法
- 受け取り者
- 返却日 確認者

このように、南山大学人類学博物館にふさわしい情報のみを選び、50 点の資料のデータがすべて入るようなスキーマの設計を行なった。また、今回使用した 50 点の博物館資料はすべて土器である。それ以外の資料においても本研究で提案する XML メタデータスキーマが妥当であるか検証する必要があるため、いくつか別の資料を用意し検証を行なった。

5. XML スキーマの検証と実験

作成したXMLスキーマを用いてその妥当性の検証と実験考察を行った。今回の検証に用いた環境に使ったXMLパーサはMSXML4.0を用いた。これはInternet Explorer を利用してウィンドウをクリックするだけでそのXMLに関連するXSLTの結果を表示し、スキーマの妥当性を判定する。XSLTで画像を表示する実験を行うためInternet Explorerを容易に利用できるMSXMLを使用する。

5.1 XML 文書の妥当性の検証

XML Schema は単なる文書定義言語に過ぎない。実際に Schema データを XML 文書と照合し、エラーを出力したり、その後の処理をするのは DOM の役割となる。DOM (Document Object Model) とは XML 文書に含まれる各ノードを操作するためのオブジェクト群である。この DOM を介することで XML 文書へアクセスすることが出来る。その仕組みを図3に記す。

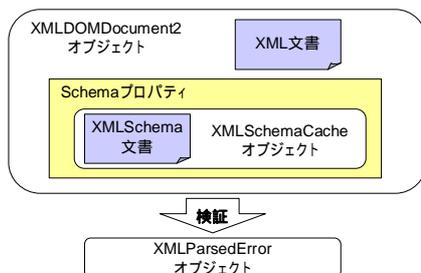


図3 妥当性検証のアルゴリズム

MSXML の DOM には XML Schema 文書を格納するためのオブジェクトとして XMLSchemaCache が用意されている。ここでは XMLSchemaCache オブジェクトを生成し、これを変数 objScm に格納している。ただしこの生成されたばかりのオブジェクトは空なので、add メソッドを実行することで XMLSchemaCache オブジェクトに XML Schema 文書が格納されることになる。

XML 文書を格納する器は XMLDOMDocument2 である。まずその器を宣言し、変数 objDoc に格納する。そして Schema プロパティに、XMLSchemaCache オブジェクトを格納することでこれから呼び出される XML 文書を検証するための元となる XML Schema 文書が XMLDOMDocument2 オブジェクトに関連付けられる。

load メソッドで呼び出された XML 文書はデフォルト状態でスキーマ文書(xsd ファイル) によって検証されエラーがあった場合には内部的にエラー情報が保持される。

保持されたエラー情報は parseError メソッドによって取得することが可能である。parseError メソッドは一連のエラー情報を保持する XMLDOMParseError オブジェクトを返す。errorCode プロパティはその XMLParseError オブジェクトに属するプロパティで、XML Schema の検証によるエラーコードが格納される。ここではエラーコードが 0 であった場合はエラーがなかったものとみなし、それ以外の場合にはそれに伴う詳細なエラー情報をダイアログ表示する。どのプロパティがどういった情報を持つかは表 1 に記す。

表1 XMLDOMParseError オブジェクトの主なプロパティ

プロパティ	説明
errorCode	エラーコード
line	行数
reason	エラー発生理由

5.2 XPath を利用したデータ抽出実験

XML 文書を書き、それを XSLT にかけて HTML で表示出来るようにした。XML で書いた情報をそのまま利用者に見せても、利用者が有益だと思える情報は少ない。そのため利用者が見て有益だと思える情報のみを XML の中から抽出して結果を XML 文書と言う形で HTML に表示させる。実際に HTML に XML のデータを抽出するために XPath を用いた。XPath とは XML 文書中の要素や属性等を特定するための記法で、1999 年 11 月に W3C から勧告された規格である。その表示結果を図 4 に示す。



図4 表示結果

```
<Object_Name_Information>
<Title_Information lang="jp">
  <Title>縄文土器</Title>
  .
</Object_Name_Information>
```

図5 XMLによる記述

```
<html>
<dd>大別名称: <xsl:value-of select="Object/Object_Name_Information/
  Title_Information[attribute::lang='jp']/Title" /></dd>
<dd>型式/器種: <xsl:.....
```

図6 XSLTによる記述

今回の実験で図5の資料の名前<Title_Information>は<Object_Title_Information>のノードに複数入っている形になる。そのため<Title_Information>を見分けるためには<Title_Information>に付いている属性を指定する必要がある。そのためのXPathは図6に記述してある。ルートノードである</>から始めて、<Title_Information>のあとに<attribute::lang='jp'>属性を持つノードを指定する。このノードの横並びとなるノードを指定するには属性指定の際にこのノードのみが有する属性、この場合は<lang='jp'>をXPathで指定する。

5.3 50点の資料に対する検証

南山大学人類学博物館所蔵の50点の資料をメタデータとしてデータに打ち込み、そのデータがXMLSchemaの妥当性に従うか検証した。この資料は全て同じ項目の情報を持っているわけではなく、一点一点そのメタデータが微妙に異なっている。それらの微妙な違いにも対応できるように空要素になるかも知れない要素の定義を利用してXMLSchemaを拡張した。しかし、全ての要素に空要素を定義し

てはいいない。その資料にとって必ず必要となる情報(例えば資料名)は記述しなければパーサはエラーを返す様になっている。また、博物館資料といっても土器ばかりではない。人類学博物館には他にも昭和の玩具や生活用品等も所蔵している。それらの資料も土器と同様の結果が得られるようにXMLSchemaを拡張した。同じ土器同士ならばその資料に対するメタデータのゆらぎは少なくなるが、全く種類の異なる資料ではメタデータのゆらぎは大きくなる。具体的にどの程度メタデータにゆらぎがあるか一部抜粋で図7に記した。

```
<!-- 7:所在情報の現所在・現所在日付・現所在タイプ・標準所在-->
<Location_Information>
  <Current_Location>第一展示室</Current_Location>
  <Current_Location_Date>2006-10-02</Current_Location_Date>
  <Current_Location_Type>展示中</Current_Location_Type>
  <Normal_Location xsi:nil="true"></Normal_Location>
</Location_Information>
```

```
<!-- 7:所在情報の現所在・現所在日付・現所在タイプ・標準所在-->
<Location_Information>
  <Current_Location>西倉庫</Current_Location>
  <Current_Location_Date xsi:nil="true"></Current_Location_Date>
  <Current_Location_Type>収蔵中</Current_Location_Type>
  <Normal_Location xsi:nil="true"></Normal_Location>
</Location_Information>
```

図7 内容の異なるデータの比較

6. おわりに

本研究では、南山大学人類学博物館所蔵の博物館資料を用いて提案したスキーマが妥当であるか検証した。スキーマの全体のグループ数は23となり、カテゴリの要素は全部で103となった。これらの情報をXMLで記述し、パーサで誤りを検証した。その結果、与えられたすべての資料を入れることができるスキーマを生成することができた。

今後の課題として博物館資料すべてに対する妥当性の検証、博物館での実用化が挙げられる。

参考文献

- [1] National Digital Archives Program, http://www.ndap.org.tw/index_en.php, 2006/1/10.
- [2] ICOM, <http://www.museum.or.jp/icom-j/>, 2006/11/23.
- [3] MODS Official Web Site, <http://loc.gov/standards/mods>, 2006/12/15.
- [4] 草苺智子, 大原千明, “博物館WebサイトのCMS利用によるユーザビリティ向上,” 南山大学数理情報学部情報通信学科卒業論文, 2006.