

画像ベース電子図書館用 Metadata モデル (ETO) の提案

5 P-6

大矢 一志¹

丸善株式会社

千葉大学大学院自然科学研究科

土屋 俊²

千葉大学文学部

1 はじめに

電子図書館が現実のインフラとなるには、始めから電子化された情報の他、既存の図書館が所蔵する書籍や絵画等を電子化し、それらを統一的に扱う必要がある。その場合、既存の情報は画像データとして扱われ、メタデータベースで管理される。但し、MARCは書籍ベースで作成されている為、操作対象となるデータ単位はシステム完成後、日常的に細分化される。例えば、論文集は各論文毎で操作対象となるのが望ましく、それに必要なメタデータを作成しデータ単位を細分化することになる。

一方、データ単位に変更があっても、外からの参照に対しては当該データ単位へのアクセスを保証し続ける必要がある。これには、所在に依存しない抽象名が有効であるが、それだけでは情報単位の細分化に対処できない。

また図書館は、このような情報提供サービスという個人に対するサービスの他、保存すなわち文化の資産化という、社会に対するサービスを行う義務がある。電子図書館でもこれに変わりはない。従来の情報資産に匹敵するデータ寿命が求められる。従って、データは可能な限りアプリケーション独立が求められ、それにはテキストベースデータモデル (TBDM) の利用が望ましい。膨大なデータを TBDM で表現する場合、現実にはデータそのものの管理用とデータ提供用のシステムを分ける必要がある。

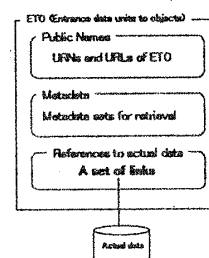
本稿では、前者で利用されるメタデータモデルを提案する。

2 メタデータモデル:ETO

ETO (Entrance data units to objects) は、関連するデータ単位を宣言して、操作対象データ単位を構成するデータ単位である。文書の電子化並びにデータ寿命からして、データ表現には TBDM を利用する。現行の TBDM 用メタ言語 SGML/XML では、中心となるデータ単位内で関連するデータ単位が宣

言される。この手法では、記述対象内容、メタ言語、これら各々のメタ記述の都合 4 レベルの内容が混合するため、データ寿命の点からすると望ましくない。ETO はこれらを中心となるデータから独立させ、その関連性を宣言する単位である。ETO は、以下の構成になる。

Independent Metadata unit models



ETO は、自らが示すデータ単位の公開名と単位構成に関連するデータを宣言し、全てのデータ単位は、ETO を介して扱われる。Public Names 部は、公開名が宣言され、URN と URL と検索用メタデータ (Core Metadata) を備える。Metadata 部は、データタイプ毎の詳細なメタデータがリンクで示される。References 部は、関連するデータ単位がリンクで示され、中心となる実データは埋め込み処理される。

3 要件

抽象名サービス

TBDM において、自らの同定手段は何らかの処理に依存せざるを得ない。そこで、少なくとも環境から独立するには、URN の様な抽象名が必要となる。抽象名は、URN の他、必ずしも global で解決が保証されないローカル名も導入する。URN サービスとして、ETO 用に拡張した THTTP サーバ (cf.RFC2169, RFC2483) を想定している。

検索用メタデータ

全てのデータ単位が ETO を介してアクセスする為、データ検索に必要なメタデータを ETO 内に置く。これを Core Metadata (CMD) と呼ぶ。CMD はメタデータ項目の集合として、URN と同様、同定情報

の役割を果たす。具体的には Dublin Core を利用するが、多言語対応の為、属性 lang、script、notation と、下位要素 original、pronouncing、meaning を導入する。

画像データ

画像データは、実データであるテキストデータ中にリンク要素として記述される。これにより、記述の線状性を使用し、その連続する構造が容易に示され、必要とあれば TBDM の構造を利用して構造を示すことが可能である (cf. TEI P3)。また、実データを分離し、ETO によってその独立単位を保証することが出来る。

```
<ImageList>
<表1Xing name="urn:local:mm001:1"/>X<表2Xing name="urn:local:mm001:2"/>X<表3>
<見廻Xing name="urn:local:mm001:3"/>X<見廻Xing name="urn:local:mm001:4"/>X<白>
<屋Xing name="urn:local:mm001:5"/>X<屋Xing name="urn:local:mm001:6"/>X<白>
<凡例Xing name="urn:local:mm001:7"/>X<凡例Xing name="urn:local:mm001:8"/>
<Xing name="urn:local:mm001:9"/>X<凡例>
<目次Xing name="urn:local:mm001:10"/>X<目次Xing name="urn:local:mm001:11"/>X<目次>
<第1章Xing name="urn:local:mm001:12"/> ~ <第1章> ~ <表4Xing name="urn:local:mm001:13"/>
```

部分文書構造定義

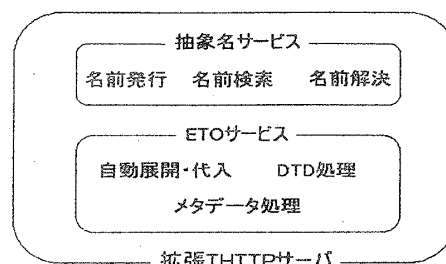
ETO は実データへの参照情報のみを持ち、利用時にそれを展開し埋め込む。ETO は複数の ETO を並列的階層的に参照できることから、部分単位毎に異なる文書構造を宣言する必要がある。それに伴い、現行の規格 SGML/XML の規格を拡張する。任意の所で要素・属性宣言を可能にし、そのスコープとして local、global、ptl(priority to the last) の 3 つを導入した。現状のブラウザで用いる際には、ETO サーバ側でデータ中の全ての宣言を外し well-formed とするか、ひとつの DTD をまとめる。

リンクと処理

情報単位が細分化すると、ETO と最終実データとの間に複数の ETO が介在する。これら ETO は全て自動展開することが望ましい。但し、細分化の階層を先読みせずに指定できることが望ましい。そこで、リンクと「処理」の関係を整理し、処理との関連性を生む「リンク活性」を、再帰的に指定する属性 activate を導入した (cf. 大矢・土屋 1999)。

4 ETO を用いたシステムの概要

ETO を用いたデータ保存・管理支援サービスは、「抽象名サービス」と「ETO サービス」から成る。前者は、抽象名の発行・確認・解決を行い、後者は、リンクの処理指定の代行、部分的構造宣言の処理を行う。検証用プロトタイプを、THTTP サーバを拡張して作成した。



ETO を利用したシステムには、1) データ作成者の責任内でデータ間の構成情報も管理できる、2) 異なる書式のメタデータを利用できる、3) 実データを得ると同時に下位要素のメタデータ全てを獲得できる、4) データ単位の細分化を当該データ単位のみの変更で可能になる、という利点がある。

尚、今後の課題として以下がある。

部分単位の独立性

TBDM の場合、従来基準であった「処理」をデータ独立の前提とし得ない為、新たな基準が必要となる。ETO が操作対象データの宣言としても、例えば、リンク要素の所在情報の書換えの際、部分要素を独立単位とするのは必要である。

リンクの定義

処理独立のデータモデルでは、全ての処理を記述する必要がある。処理の名前をどう付けるかの他、宣言方法が問題である。現行の PI は極めて直接的な処理指定を目的とする。対してリンクは「名前解決」と参照先との「関連性を何らかの処理で実現」という抽象的手順の総体である。関連性の表記法を統一させる意味では、リンクの利用が望ましい。但し、処理指定の関連事項、特に部分単位と、独立リンクと処理指定の場合、困難である (cf. 大矢・土屋 1999)。

Local URN の扱い

local 抽象名と global 抽象名 (URN) との関連が未定である。RDS の実体が不明である為、サーバ指定で local 抽象名が使えるという保証が無い。

5 さいごに

本研究は、既存の図書館収蔵情報を含む電子図書館研究と共に、TBDM を利用したデータ作成の研究の取り掛かりとなった。

参考文献

- [1] C.M.Sperberg-McQueen and L.Burnard ed., 1994, *The Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, TEI
- [2] 大矢・土屋 1999 「複合データ単位に於けるリンク要素間リンクについて」 *IPIS SIG Notes*, Vol.99 No.48, IPISJ