

消えてゆくデジタルコンテンツのために － 保存を目的としたテキストデータによるデータ記述法研究の提案 －

大矢 一志

千葉大学大学院自然科学研究科

丸善株式会社

ohya@cogsci.l.chiba-u.ac.jp, ohya@maruzen.co.jp

デジタルコンテンツを収める電子図書館やデジタルアーカイブが社会基盤となるには、手法や利便性の研究に加え、対社会サービスである保存の研究が欠かせない。本稿では、保存対象を digitized information と digital-born information とした場合に、前者を保存対象とするいわゆる hybrid libraries で実現するデータ保存を目的としたデータ表現法 Text-based Data Model(TBDM) 研究を提案する。

For disappearing digital contents

－ A proposal to study text-based data descriptions for data preservation －

OHYA Kazushi

Graduate School of Science and Technology, Chiba University

Maruzen co.,ltd.

In order for digital libraries and archives to be infrastructures, in addition to a study of digitization and enhancement of usability, there must be a study of data preservation as social services. In this paper, I propose to study data preservation in hybrid libraries, which comprise not digital-born but digitized information. And, as the way of data preservation, I propose to use a text-based data model(TBDM).

1 はじめに

デジタル化しただけではその対象を保存したことにはならない。デジタルオブジェクト自体が保存されその再現性が保証されて、はじめてデジタル化による対象の保存が成される。これは、デジタル化の手法、データを収める媒体の寿命や、媒体変換の手法のみで解決されるものではない。再現性対象を明確にし、手段を選択する必要がある。すなわち、デジタルオブジェクトのなにを保存するかが問題となる。

本稿では、デジタルオブジェクトの保存研究の重要性を示し、特に記述による保存を提案したい。その研究領域の概観と具体的な対象を示す。はじめに、電子図書館やデジタルアーカイブの役割と、そのひとつである保存について整理する。次に、デジタルオブジェクトの記述法研究について概観し、具体的な課題として複合データ単位を例に検討する。

2 デジタルオブジェクト

デジタルオブジェクトには、digitized information (objects, contents) と digital-born information (objects, contents) の2種類ある。digitized objectsとは、従来の情報媒体から特定のデジタル化の手法を用いて作られるもので、現在多くの研究が、人文科学研究と学際的に推められている。digital-born objectsとは、実現される環境を除いては特定した媒体を持たず、字義通り始めからデジタルオブジェクトのものである。digital-born objectsとして現在は、パッケージ系と配信系の2種類が存在している。

3 電子図書館の役割

この2種類のコンテンツを収め扱う主体をデジタルアーカイブや電子図書館(以下DL)とする。DLの役割は、従来の図書館、文書館、博物館等が担ってきた2つの役割、すなわち対人サービスである情

報提供サービスと、対社会サービスである保存サービスを引き続き担うべきである。それには、デジタルオブジェクトの特性を生かして利便性を高めた環境を提供することへの研究と、デジタルオブジェクトそのものを保存し社会資産化するための研究が必要となる。

DLはコンテンツの消費者であるから、digital-born objectsは館内利用の形態に限定された利便性の向上が図られる。恐らく、情報の統合化がサービスの主流になるとと思われる。

DLでの保存は、2つの点、すなわち、既存図書館がdigital-born objectsに対応していない点と、DLが社会基盤として機能する条件である点で重要である。

現行の図書館や文書館は、digital-born objectsに手を拱いている。特に、配信系情報は、日常会話と同様の消費情報として位置づけられ、取り扱う対象にはなっていない¹。従って、知的資産の多くがネットワーク上でデジタルコンテンツとして公開される中では、デジタルコンテンツは消えてゆき、そして現行図書館の役割は既存の媒体の保存となる。従って、digital-born objectsの保存は重要なDL研究になる。

現行大学図書館や国家の中央図書館が担う「知的資産の保存」の役割を、基幹DLも担うことになりはしないだろう。そのDLが社会基盤となるには、社会基盤としての普遍性と運用コストが求められる。それが従来よりも高い場合、例え学術上重要であっても、商業的な需要がないものは扱う対象にはなり得ない。電子化が進んだとしても、コスト上、保存対象は紙媒体でしかなくなる。これでは、保存は埋葬と同じ意味になる。従って、デジタルオブジェクトそのものの保存法の研究は重要である。

4 DLに於ける保存

従来、保存には3つの手法が採られてきた。実物保存(original preservation)、内容保存(content preservation)、実存性保存(existence preservation)である。

実物保存とは、劣化するもの対象そのものを保存する手法で、古来から行われている、直接的かつ究極の手法である。内容保存とは、形態や内容を記述説明することで保存してきた手法で、例えば写本や形態を記述した目録の作成等になる。実存性保存とは、対象そのものを疑似的な姿で保存する手法で、具体的には今世紀に書籍の保存法として開発されたマイクロフィルムによる保存法を示している。これは、内容と形態を同時に保存する点で新しい保存法である。

この3つの手法は、デジタルオブジェクトにおいても当てはまる。実物保存の場合、保存の定義を「劣化する対象の現状維持」とすると、デジタルコンテンツで劣化するはその環境であるから、保存対象は「再生環境」となる。これは従来にない保存の手法が必要となり、重要な研究対象である。デジタルコンテンツの内容保存は、「テキストデータによる記述」と考えている。メタデータによる記述の他、テキストデータとして記述できるものを全て記述することになる。詳細は後述する。デジタルコンテンツの実存性保存は、「メタプログラムによる記述」と考えている。移植、エミュレーション等が行われるものと仮定している。

デジタルコンテンツは、この3手法による保存の対象となる。すなわち、デジタル化が従来の保存の手法に加えた第4の手法に直接なるわけではない。従来の3手法で使われる媒体自らがまた保存の対象となるよう、デジタルコンテンツもこの3つの手法によって保存される。その上で、「再現性²」が保証されて、はじめて保存の第4の手法として確立されることとなる。デジタルコンテンツの保存法研究は、今後重要な課題である³

本稿では、「内容保存」について考えてみたい。

5 ハイブリッドライブラリ

本稿で想定するDLはdigitized objectsを扱うもの

に限定し、それをハイブリッドライブラリと呼

¹私としては、例えば、遅く受信料を徴収するNHKのコンテンツは国会図書館に収め、然るべき著作権処理ルールの下で自由に利用できるのが、来るべき中央図書館の役割と本来のコンテンツ利用の姿であると思っている。

²これが、実物保存なのか実存性保存になるのかは、正直良く判らない。今後実現される「再現」の程度に依るのだろう。

³私としては、本シンポジウムでうたうようなデジタル化の目的を保存と捉えるのは、DL研究としては不適切ではないかと思っている。世間に正確な情報を伝えることが専門家の役割とするならば、今後電子図書館やデジタルアーカイブの研究で「保存目的のデジタル化」を匂わせるキーワードを使うべきではないのではないだろうか。

ぶ⁴。例えば、既存図書館が収蔵する書籍をデジタル化し、管理・利用・保存するDLになる⁵。このハイブリッドライブラリでは、書籍や絵画等は静止画像データとして電子化される。制作当初に一度にテキスト化するのにはコスト上非現実的であり、異字体の規定が曖昧な文字コード体系を利用することは実用上また理論上も困難である⁶。この画像データは、メタデータで管理される。既存のメタデータは書籍単位で構成されたMARCを基にしている。従って、ハイブリッドアーカイブが完成した後、必要なメタデータは追加され、同時にメタデータが示すデータ単位は細分化してゆく。例えば、論文集は、各論文が操作単位となる方が望ましい。メタデータの精緻化に伴って書籍単位より詳細な意味区分を画像データのページ単位まで扱うことは容易である。

このように基本単位が細分化し、新規データが常に構造の変化を伴うデータを扱う場合、SGML/XMLに代表されるメタ言語を使ったデータ表現(TBDM[12])が有効である。尚、SGML/XMLを使った本格的な電子図書館研究の枠組が確認されたのは、1998年LCでの会議TEI and XML in Digital Librariesであると認識している[21]。

5.1 データ記述例

ハイブリッドライブラリでは、データ更新に伴う構造の変化に容易に対応する為と、データ寿命を高めるため⁷、SGML/XMLによるデータ記述が第一次データとして利用される⁸。例えば、以下のようなデータ記述が利用できる⁹。

```
<images>
<book name="urn:maruzen:bookA">
  <link location="urn:maruzen:001.jpg">Image 1
  </link>
  <link location="urn:maruzen:002.jpg">Image 2
  </link>
```

⁴ハイブリッドライブラリとは、既存の図書館と新しい保存技術を集合させた姿をHybrid Systemsとして示したD.Willis[20]に依っている。

⁵digital-born objectsのみを扱うDLと異なり、ハイブリッドライブラリには既存の情報があることが、digital objectsの扱いに影響を与える。本稿が想定する書籍のデジタル化オブジェクトの扱いでは、書誌データの作りや、内容のテキストデータの扱い等が、DL全体のデザインに大きく影響する。

⁶網羅的文字コードがあっても、状況は変わらない。高々100年を越える以前の字形情報は重要で、またコード化の作業手順上も画像が必要となる。

⁷本来、ここは慎重な論議が必要であるが、本稿では棚上げする。

⁸データ提供サービスでのデータが第二次データとなる。

⁹正確なデータ記述はOhya and Tutiya2000[12]にある。

¹⁰このように、言語の線状性をデータの順序表現に利用するなど、自然言語の特性を利用した記述規則や記述様式が重要になってくる。特定記述言語の表現力に頼る記述ではなく、自然言語自身が持つ性質を利用することで、データ記述は般化する。S.Abiteboul et al.[1]等、TBDMの性質をどう捉えるか、様々な試みが行われているが、自然言語自身が持つ性質からみたTBDM用メタ言語の表現力を探るアプローチもあると考えている。

```
<!-- contd. -->
  <link location="urn:maruzen:100.jpg">Image 100
  </link>
  <link location="urn:maruzen:101.jpg">Image 101
  </link>
<!-- contd. -->
</book>
</images>
```

例 1

例1では、書籍を最上位単位として、ページ単位に相当する画像データが、記述の順序をそのままデータの順序関係として記述されている¹⁰。画像データの実体は記述中に埋め込まれることなく、その関連性がリンクにより宣言されている。この例1がbookが論文集の場合、後日収録された各論文を分割する為の書誌情報が用意できた時、次の例2の用にデータ記述は変更される。

```
<images>
<book name="urn:maruzen:bookA">
  <article id="urn:maruzen:article01">
    <link location="urn:maruzen:001.jpg">Image 1
    </link>
    <link location="urn:maruzen:002.jpg">Image 2
    </link>
  <!-- contd. --> </article>
  <article id="urn:maruzen:article02">
    <link location="urn:maruzen:100.jpg">Image 3
    </link>
    <link location="urn:maruzen:101.jpg">Image 4
    </link>
  <!-- contd. --> </article>
</book>
</images>
```

例 2

ここでは、<book>と<link>との単位構造間に新たに単位<article>を挿入している。この際、元のデータ単位<book><link>には何ら変更はない。従って、この変更の事実を知らない外部から、変更前のデータ単位で求められた要求に対しても、問題なく対応することが可能となる。

またこのデータ記述は、終端単位が細分化する場合も容易に対応する。例えば、画像データを例2

の様に直接参照するのではなく、メタデータを持つデータ単位を介して参照する場合、以下のようになる。

```
<images>
  <book name="urn:maruzen:bookA">
    <article id="urn:maruzen:article01">
      <link location="urn:maruzen:001">Image 1</link>
    <!-- contd. --> </article>
  <!-- contd. -->
</book>
</images>
<!------->
<!------->
<!-- A referent of urn:local:f01 -->
<frame id="urn:local:001">
  <right>
    <desc>白</desc>
  </right>
  <left>
    <desc>表紙</desc>
    <text><location>urn:local:001:t1</location>
    </text>
    <img><location>urn:local:001:left</location>
    </img>
  </left>
</frame>
```

例 3

例 3 では、画像データの参照名をもつ新たなデータ単位<frame>の中に、画像データの書誌情報と、画像データの実体への参照が記述されている。これにより、元の記述に変更を加えることなくデータの更新が可能になる。終端データ単位の細分化の度に全体のデータ構成情報を確認する必要はない。

6 複合データ単位

TBDM は、データ記述自体でデータ内容と自らの構造を記述する。従って、先の例 3 にあるよう、リンクにより参照するだけでその実体は持たないデータ単位同士が、それぞれリンクで参照され、リンクによるデータ構造を構成する。このようなデータ単位を SGML/XML では「ハブ (SGML/XML) 文書」と呼ぶ。ハブ文書は SGML/XML が抱える課題を容易に示す要素として、様々な提案と論議が交わされてきた¹¹。ハブ文書とは、リンクにより複数の独立データ単位を、実体を持たずとも構成要素として持つデータ単位であることから、複合データ単位といえる。

データ記述の寿命を高める為には、記述内容ごとにデータを独立させることが望ましい。この個々に独立したデータ単位をひとつの意味単位として統合する際に、ハブ文書による複合データ単位は有効である。

従来、ハブ文書による複合データ単位には、以下の課題が挙げられている。

1. 構造衝突
2. 独立データ単位の認定
3. 関連性の保証
4. データ単位間の調整
5. メタデータの機能
6. リンクの制御

6.1 構造衝突

TBDM では、複合文書内で文書構造間の衝突が懸案である。構造衝突には、(1) 文書構造が異なる (論理構造の衝突)、(2) 相当するコンテンツを示す場所が無い (意味の衝突)、(3) 要素名が示す意味や構造上の位置が異なる (名前の衝突) の 3 種がある。

名前衝突の対策として、意味レベルの予約語を DTD で現実名に定義する Architecture Form の手法、現実名を予約する Namespace の手法¹²が提案されている。論理構造衝突の対策としては、中間抽象構造を緩衝して変換テーブルを用意する手法 [14]、木構造の構成規則を独立させて統合する手法 [8] 等が提案されている。前者は、異なる構成要素集合間でも有効な手段で意味衝突にも適応するが、手続きは繁雑になる。後者は、syntax に限定した手法で意味衝突には対応しないが、簡潔である。意味衝突は、言語意味世界内のもの、処理や記述様式等様々なレベルで観察される。言語意味世界とメタ言語に関する解決策には、タグを記述説明する TEIP3 の tagDoc [17]、DCD [3]、DDML [2] 等がある。

6.2 独立データ単位の認定

本稿では、SGML/XML で表現される TBDM の最小データ単位をインスタンスと呼び、ハブ文書で

¹¹ある意味、XML は、SGML が陥ったこの課題を棚上げし実用面を追求する研究を前進させる為のものであったといえる。その意義は大きい。但し、全ての課題が過去の枠組のものとして捨てられるものではない。私は、ハブ文書は依然として新しいデータ管理法として重要であると思っている。XML が広げた研究コミュニティへでも、SGML に携わった人々が検討してきた過去の課題を、生きる資産として検討する価値はある。

¹²DTD 無しを想定する XML では現実名自体を予約することになる。

構成される意味単位としての複合データ単位をユニットと呼ぶ [12]. インスタンスは、複合データ単位の構成要素となる他、自身の構成要素としてインスタンスを持つことがある。インスタンスは、現行規格で実名か構成情報等の弁別素性を利用して参照される [17][10]. 実名は名前衝突に弱く、弁別素性の記述的名前は環境の変化に弱い為、共に保存の点で問題がある。

TBDM で抽象名は、各処理系から独立した名前管理や、名前予約の課題に有効である。名前の予約は、デリミタとしてのタグの柔軟性を損なうためデータ保存の点では不適切であるが、意味を共有する共通表現としては必要である。例えば、Namespace で予約されている現実名を要素名として利用する場合、Architecture Form 上で Namespace の予約語を抽象名と見做し別名を付与することになる。従って、名前予約には抽象名を利用することで名前衝突が回避できる。例えば属性に抽象名を付与することで、要素名の変更なく意味の共通表現として別名を付与することが可能になる。抽象名サービスとして「名前付与」「名前付与サービスを知る」「名前登録」「名前登録サービスを知る」「名前解決」「名前解決サービスを知る」「連携」「認証」の 8 つが想定される [16].

本稿では独立データ単位を、抽象名が与えられたデータ単位とする¹³。現在、特定サイト内での抽象名解決サービスを検討している。

6.3 関連性保証

複合データ単位 (ユニット) は複数のデータ単位 (インスタンス) で構成される為、それらの関連性を保証する仕組みが必要になる。現行規格 SGML/XML ではこれに関する規定がない。処理系毎に関連データ単位の所在を内部規定するか、データ単位内で直接所在を宣言することになる。前者は強い処理依存、後者は関連性が固定的になる点で、保存を目的とする TBDM では共に問題である。そこで、抽象名を利用した参照関係で関連性を記述する手法を検討している。利点は、データ記述レベルから関連性保証の仕組みに関する問題を除けることである。欠点は、データ寿命の為に処理独立を目

指した TBDM が、唯一とはいえ名前解決の処理系に依存することである。

6.4 データ単位間の調整

ユニットを構成するインスタント間の調整が必要になる。調整には、1) タイプの認定と、2) 調整の方針が必要となる。

タイプには、意味内容部、メタデータ部、スタイル部、メタ記述部、等が想定される。メタ記述部とは、DTD 等あるデータ記述に対する記述規則を規定するインスタンスになる。記述規則にも、語彙表、様式、意味の規定も必要となる [11]. 調整には、置換え、連結、変換、選択、代入等、様々なものが想定される。部分データ単位の代入とスタイル情報に関する研究としては DeRose1997[5]Maden1998[7] がある。この課題の先行研究は少い。

6.5 メタデータの機能

抽象名は、必ずしも独立データ単位に一意であるとは限らない。抽象名が付与される対象には、実データ単位の他 ISBN の様な抽象 (意味) 単位が想定されている [15]. 抽象名が後者に付与された場合、その実データ単位を選択するためのメタデータが必要となる。しかしメタデータを実データに埋め込むことは、データの再利用性を高める為にタイプ毎に独立させ複合データ単位で意味単位を表現する TBDM の方針から外れる。更に、抽象名が必要とするメタデータと従来の書誌情報に相当するメタデータは異っている。従って、抽象名に必要なメタデータは、記述対象となるデータ単位から独立したデータ単位となる。実データを持たず参照情報のみで構成されるハブ文書は、ユニットの構成情報を示すデータ単位であり、これはメタデータの機能に近い。また、参照情報には、インスタンスのタイプ等のメタ情報も備える必要がある。この機能を拡張し、ハブ文書にユニット成立に必要な最小限のメタデータ CMD 持たせるとする。これにより、ハブデータ単位を基準に形成されるデータ単位は、データ単位の階層構造は関連性としてデータ単位自体に示される。単位成立に必要なメタデータ CMD はハブデータ単位にあるため、従来のよう

¹³これは、URN が独立単位に抽象名を与えることと逆になる。従って、抽象名が与えられる対象は、インスタンスの様な、従来ではファイル内のエレメントの様に親データ単位名との相対的名前しか与えられなかったデータ単位も含まれ、それらを意味単位として抽象名で管理することを目指している。

に階層間でデータ単位のメタ情報の関連性を取る必要はない。これにより、MARC や TEI IH で懸案であった、終端データ単位の細分化に階層上位のデータ単位のメタデータを連動させる問題が解消する。また、ユニットが意味単位を構成することから、Dublin Core[19] の 1:1 ルールに適う。

6.6 リンクの制御

TBDM で参照はリンクとして実現される。複合データ単位とリンクには以下の課題がある。

1. リンクの挙動制御
2. 複数リンク間の調整

6.6.1 挙動制御

ハブデータ単位は単位を認定するもので、本来の対象データはリンクの参照先にある。従って、単位への認定と同時に対象データを得るには、リンクは自動的に展開される必要がある。また、ハブデータ単位を構成要素として持つ場合、リンクへのリンクを自動展開する必要がある。この場合、下位に幾つのハブデータ単位があるか予測がつかない。従って、先読みの要らないリンク制御が必要となる。

リンクの規定

本研究では、リンクを以下のように規定している [9][12]。目的は、従来のリンクの規定に対して、1) 記述レベルとリンク経路レベルを明確にし、2) TBDM で示される処理規定と関連する点を改善することにある。

```

[Link](Reference, Referents) = {
  Link(Reference, Referents)
  Referents = Referent+
  Reference ⊃ ReferTo+, Instantize, Activate, LinkDef
  Referent ⊃ Instantize, Activate, LinkDef
  ReferTo(Locator, Referent)
  Instantize(Locator, Instance)
  Activate(Actant, Bind(Process, ReferTo))
  Locator(LocatorPoint, LocatorVale)
  LinkDef ⊃ Actant, Process
}

```

リンクを、指示参照 (Reference) から被参照 (Referent) の関係とした。Reference の属性 (リンク素性) 指示参照関係 (ReferTo) は、所在情報 (Locator) を伴い Referent と関係する。ReferTo 関係は、Locator と Referent との関連性を示す参照行為になる

参照行為自体は、名前による関係宣言であるが、実際に関連するデータ単位は関係 Instantize で示される。ReferTo 関係にあるリンク要素間の構造をリンク構造と呼び、その深さを単位ステップで表現する。ReferTo 関係は、任意の処理 (Process) と関連し (Bind)、それは活性化 (Activate) によって実現される。Activate 関係は、活性化の主体 (Actant) を待つて成立する。

独立リンクとリンク経路

従来の独立リンクは、記述レベルの参照先が、実現するリンクでは参照元となる参照関係の逆転を利点とした。しかし、これは 1) TBDM で関連性を示すリンクの機能と、2) 処理時のデータ単位の振舞を規定する機能とを混在させている点で問題がある。従って、記述レベルと実現するリンクレベルを一致させる為、ReferTo 関係で成立する選択可能な静的な経路「静的リンク構造」と、Activate 関係で成立する選択された経路「動的リンク経路」を想定している。動的リンク経路をリンクの先読みをせずに成立させることがハブデータ単位では必要となる。本研究では、動的リンク経路を再帰的な指示で確立するリンク素性 Activate として、属性 activate_timing と属性 activate_cue を導入している [9]。

リンク間の調整

リンクへのリンクは、場合によって参照先に非リンク要素が含まれる [9]。従って、リンク属性間で値の受渡しが必要となる。属性値を継承する対象の特定には、要素名か属性名を予約する必要がある。属性名として実名の予約は、要素名と同様 TBDM では不適切であるため、この手法は採らない。要素は、属性に抽象名を持つことで別名を持つことが可能であるが、属性は Architecture Form の様に DTD 内で抽象名を持つことで別名が持てる。本研究では別名の予約手順を明確にする為、TBDM での属性を参照にのみ利用する記述とし、属性名を ID か IDREFs として予約する手法を採った [12]。これにより、属性の役割は「関連性の強い同時に起こる情報を示すものではなく、他のメタ記述部と関連する為のメタ言語側の記述」となり、SGML/XML が自然言語の記述をデリミタとして導入した利点の副産物である「コンテンツとメタ言語記述の混乱」

を解消することが可能となる。処理を規定するインスタンスを参照するリンクは、活性化により処理と関連する (Bind) ため、インスタンスの独立が処理独立を保証することになり、保存を目的とした TBDM のデザインが容易になる。

7 デジタルコンテンツの内容保存

デジタルコンテンツの形態は、技術が進展する以上、安定する運命にない。それでも尚記録する手法を確立する重要であり、それは実物保存が後追いになる可能性を含む限り、内容保存は、少くともメタデータの記録を行うことで、デジタルコンテンツを知的資産化できる点で、重要である¹⁴。デジタルコンテンツの内容保存、すなわちメタデータと記述の研究のうち、後者は今までにない新しい試みとなる。

内容記述を必要とする理由には、先に挙げたハイブリッドライブラリでの必要性和、実物保存を補完する手法として内容記述は欠かせないことを挙げたい。その手法として TBDM を採用する理由には、1) 最も高い記述力を持つ自然言語を利用していること、2) それが人文科学者が従来採ってきたデータ整理のアプローチと親和性があること [11]、3) それで且つデータ処理の対象となるデータ形式となり、4) それがデータの再利用性とソフトウェアの相互利用性を保証するインターフェースとなり

得ること [1][18] を挙げたい¹⁵。

現在、TBDM では、実用性と実効性を重視して規格化を急ぎ、既成事実を作る傾向がある。技術革新の速度や、需要の高さ、新しい政治力の示し方を考えれば、新しい規格化のプロセスとして、これは極めて正しいと思う。現在欠けているのは、これを牽制する、理論的背景を持った要求や、理念的に必要とされる要求である。その一つの領域として、DL における保存の研究は、重要な役割を果たし得る。すなわち、デジタルコンテンツを記述する意味と、その要求を明確にすることが重要になる。従来、ハードやソフトを含めた処理環境に強く依存していた情報管理から、情報 (データ) を主体とした管理法並びにその環境の構築を目指すことが可能となり、その為の論議を進める視点として、デジタルコンテンツの保存は有効であると思われる。

8 さいごに

現在、本稿で挙げたハブ文書の実現形 ETO を研究している。その管理サーバを、THTTP サーバ [4][13] を拡張して作成している。その過程で、ローカルサイト内での抽象名サービスと独立データ単位管理サービスの必要性を強く認識した。またこれは、別名管理を棚上げしている現在の URN の論議に貢献すると思われる。

参考文献

- [1] S. Abiteboul, P. Buneman and D. Suciu, 2000, *Data on the Web*, Morgan Kaufmann Publishers
- [2] R. Bourret, J. Cowan, I. Macherius and S. S. Laurent: "Document Definition Markup Language (DDML) Specification, Version 1.0", W3C Note, W3C (1999)
- [3] T. Bray, C. Frankston and A. Malhotra: "Document Content Description for XML", W3C Note, W3C (1998)
- [4] R. Daniel: "A Trivial Convention for using HTTP in URN Resolution", RFC2169 (1997)
- [5] S. J. Deroose and C. R. Maden: "Problems with Dynamically Assembled Document Portions, and Some Solutions", Proceedings of SGML / XML'97 (1997)
- [6] N. Ide and J. Véronis eds: *Text Encoding Initiative: Background and Context*, Kluwer Academic Publisher (1995)
- [7] C. R. Maden: "Hypertext Link Support Requirement for XSL", manuscript (1998)
- [8] M. Murata: "DTD Transformation by Patterns and Contextual Conditions", SGML / XML'97 (1997)
- [9] 大矢, 土屋: 「複合データ単位に於けるリンク要素間リンクについて」『情報処理学会研究報告』 Vol.99 No.48 (1999)

¹⁴私としては、RDF はこの点を見失ってしまったように感じられ、残念である。関連性保証にも、メタデータの普遍的格納部としての機能にも使えない RDF は、DL には利用できないのではないか。

¹⁵勿論、SGML/XML のデータフォーマットとしての普遍性は、今後も重要な課題であり、その為に、先に挙げた記述の般化を目指す研究、例えば自然言語の特性を利用した記述規則の研究などが大切になってくる。

- [10]K.Ohya and S.Tutiya: “For Internet Minority Language – a metadata format to include multiple language cultures–”, *Proceedings of International Symposium on Digital Libraries 1999(ISDL99)* (1999)
- [11]大矢, 土屋: 「システムが決まらなければデータベースは出来ないのか-テキストベースデータモデル利用の提案」『第2回アートドキュメンテーション研究フォーラム報告集』アートドキュメンテーション研究会 (2000)
- [12]K.Ohya and S.Tutiya: “Descriptive rules of SGML/XML for data preservation”, *Proceedings of the 3rd International Conference of Asian Digital Libraries(ICADL2000)*(2000)
- [13]M.Mealling and R.Daniel: “URI Resolution Services Necessary for URN Resolution”, RFC2483 (1999)
- [14]G.F.Simons: “Using architectural forms to map TEI data into an object-oriented database”, TEI Tenth Anniversary Users’ Conference (1997)
- [15]K.Sollins and L.Masinter: “Functional Requirements for Uniform Resource Names”, RFC1737 (1994)
- [16]K.Sollins: “Architectural Principles of Uniform Resource Name Resolution”, RFC2276 (1998)
- [17]C.M.Sperberg-McQueen and L.Burnard eds: *Guidelines for Electronic Text Encoding and Interchange(TEI P3)*, Text Encoding Initiative (1994)
- [18]C.M.Sperberg-McQueen and B.T.Usdin: “Welcome to Markup Languages: Theory & Practice”, *Markup Languages: Theory & Practice* Vol.1 No.1, MIT Press (1999)
- [19]S.Weibel, J.Kunze, C.Lagoze and M.Wolf: “Dublin Core Metadata for Resource Discovery”, RFC2413 (1998)
- [20]D.Willis: “A Hybrid Systems Approach to Preservation of Printed Materials”, CPA (1992)
- [21]TEI and XL in Digital Libraries, 1998, Library of Congress, <http://www.hti.umich.edu/misc/ssp/workshops/teidlf/>