

考古資料情報の XML 記述とその応用に関する事例研究

○小笠原 和慶 八重樫 純樹

静岡大学大学院情報学研究科

An Experimental Study of XML Description of Archeological Information and its Application to Web Data Base System.

Kazuyoshi Ogasahara, Jyunki Yaegashi

Abstract

This paper shows an example of converting XML from RDB data technique as archeological information. In addition, it shows the technique of developing Web-DB application that used XML as an example of application of the XML data.

1. はじめに

XML は、1998 年に W3C によって勧告された文書記述規格であり、データの汎用性、再利用性が高いことから、情報流通の記述媒体として注目を集めるようになってきている。近年においては、データの共通交換フォーマットとして、ビジネスシーンを含め様々な目的やアプリケーションにおいて利用されることが多くなった。XML は一つのデータ記述の枠組みであり、その記述文法に則った多くの記述規格がここ数年急速なペースで策定されてきている。

このような流れの中で、考古学分野においても XML をベースとした情報記述への取り組みが見られるようになりつつある。本研究では考古学情報の一つである土偶とその関連情報（時間情報・地理情報・参考文献情報等）を XML により記述し、考古情報における XML 記述の一例を示す。さらに、XML 化したデータの利用応用例として、XML をベースとした Web アプリケーションを開発し、その応用に関する一例を示す。

2. 土偶関連情報の XML 記述化

2. 1 データ

対象となるデータは、静岡大学 八重樫純樹が中心となって収集・整理を進めた、全国の土偶とその関連データである。これは、推定年代、出土状況、遺存部位、寸法、実測図等の土偶の実体情報、出土遺跡名、緯度経度等の地理情報、参考文献や著作情報などから構成され、総レコード件数は 10257 件である。

これらの情報は「土偶調査カード」(図 1) という紙媒体のカードに記入され、一意の土偶 ID によって識別・管理されている。土偶 ID は 11-12 桁であり、自治体コード(4-5 桁)、遺跡コード(4 桁)、出土 No.(4 桁)の組み合わせから構成される。静岡大学情報学部八重樫研究室では、これらの土偶関連情報を Microsoft のデータベースアプリケーションソフト、Access に入力・電子化し、リレーショナルデータベースの形に格納して保存している。本研究では、このリレーショナルデータベース内に格納された土偶関連データを XML に変換することで、土偶情報の XML 記述化を行った。

2. 2 スキーマ

XML への変換の際に大きな障害となるのがスキーマの決定であった。出力先のスキーマが未決定の状態であると、XML として出力できないためである。

図 1 土偶調査カード

スキーマは、資料は必ず時間情報、地理情報を持つという観点に基づいて、土偶の実体情報と地理情報を分類し、文献管理情報や参考文献情報等の関連情報とともに、遺跡、土偶、図表、管理、参考文献の項目に分け順次格納することとした。

用語は土偶調査カードに使用されたものと同じものを使用することとした。なお、変換元である Access 内に格納された土偶データは、電子化の過程で数箇所記述項目の用語の変更が為されていたことを付記する。これは、Access 内において検索をより効率的に行うために為された変更であるが、本研究ではこの記述用語を改め、オリジナルのデータである土偶調査カードのものに準拠させることとした。これは、土偶調査カードに記載されている用語が、実際にデータを扱う考古学分野の関係者間での深い議論の末に決定されたものであることを受けてのものであり、そちらの方がより確度が高いと判断したためである。

2.3 XML への変換

リレーショナルデータベースの XML 化には多くの手法が存在する。その多くはデータベースサーバに接続し、リレーショナルデータベースのレコードに逐次アクセスすることによって取り出したレコードを、事前に決定したスキーマに基づくオブジェクトツリー内に順次格納していくことによって可能となる。本研究においてもその手法を採用した。プログラミング言語には VBA を使用し、Access 内にマクロの形式で記述し、XML ファイルを出力した。

具体的な手法は以下の通りである。まず、リレーショナルデータベース内のテーブルをクエリ機能により一つのテーブルに結合した。この理由は、リレーショナルデータベースのスキーマが、正規化によって複雑に細分化されたテーブル構造となってしまうためである。各テーブルの内容をそれぞれに XML に出力した後、XSLT を使って一つの XML 文書に統合する手法も考えられるが、この場合、一度出力した XML を再び XSLT を使って再変換しなければならず、VBA、XSLT コード記述に関する省力化と効率の観点から、先に Access のクエリ機能を使って一つのテーブルにまとめたあと、一括でファイルに変換する方法を採用した。なお、クエリを使ったテーブル統合の際に、コード化されていた情報を事前に決定したスキーマに適合するように、文字データに修正したことを付記しておく。数字コードから文字データへの置き換えは、当該コードの意味する文字データを Access の置換機能を利用して行った。

テーブル統合後、VBA コードによりテーブル内に格納されたデータを XML へと書き出す作業を行った。具体的な手法は次の通りである。

- ① VBA を使い ADODB に接続し、ADODB を介して Access 内のテーブルにアクセスする。
- ② レコードセットから取り出したレコードを MSXML に渡し、レコードを順次 MSXML の DOM ツリー内に格納する。
- ③ メモリ内に生成された DOM ツリーを XML ファイルとして保存する。

このようにして XML ファイルを出力し図 2 を得た。

3. Web データベースアプリケーションの開発

3.1 システムの実装

システムは Microsoft Windows(R)の Web サービスである IIS 上に ASP の形で実装した。ASP のスクリプト言語には Jscript、XML パーサ

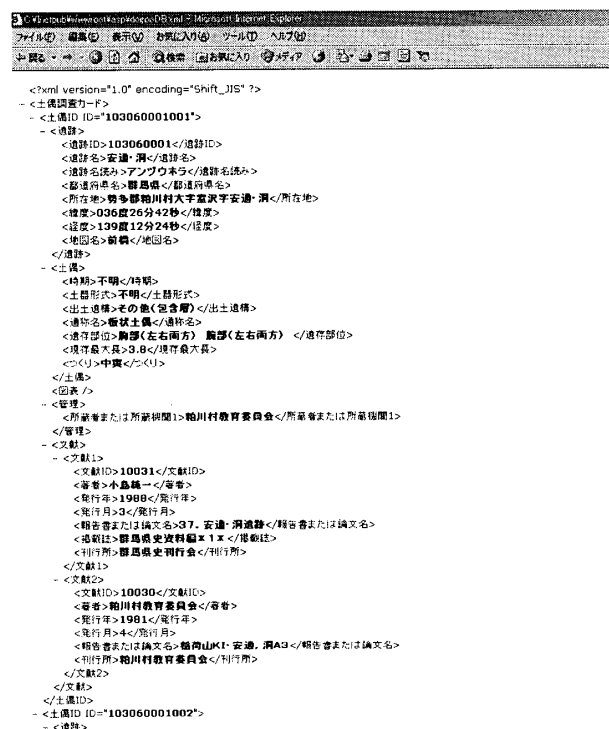


図 2 出力済み XML データ

一には Microsoft Internet Explorer 6.0 に同梱の XSLT 1.0 を使用した。開発環境は次の通りである。

- ・ハードウェア、OS
 - PC/AT 互換機 (Pentium3 933MHz RAM 512MB)
 - Microsoft Windows 2000 Server
- ・アプリケーションソフトウェア
 - Microsoft Internet Information Service 5.0
 - Internet Explorer 6.0 (XSLT 1.0)

システムは、DB 本体の XML ファイルを DOM オブジェクトとしてメモリ上に展開し、Web フォームから入力される検索キーやパラメータを XSLT が受け取り、メモリ上の DOM オブジェクトに対して XSLT 変換を行い、その変換結果を表示・保存することで機能を実現した。検索の項目としては、八重樫研究室において先行開発された Access 版土偶データベースにおける検索フォームの検索項目を参考に作成した。具体的には、遺跡名、所在地、都道府県名、時期、通称名、遺存部位、特記事項による検索を行うことができる。

システムは検索項目ごとに3つのファイルのセットから構成される。すなわち、

- ①検索フォーム用 HTML ファイル
 - ②メイン処理用 ASP ファイル
 - ③ノード抽出用 XSL ファイル
- である。

①は検索キーを ASP ファイルに渡すためのものである。②は HTML フォームよりキーワードを受け取り、③にキーワードを渡し、その後③によって抽出されたノードをユーザーへ表示させるためのものである。③は受け取ったキーワードを元に XSLT の抽出関数の機能を使って、メモリ上に展開された XML ツリーの中からキーワードに該当するノードのみを抽出するためのものである。これら3つのファイルを使い、ASP 上で XSLT を動作させ、DB 本体である XML ファイルから任意の情報を検索する。具体的な処理とシステムモデル (図3) は次の通りである。

- ① 検索フォームに入力された検索キーワードを検索キーとしてメイン処理用スクリプトを介して検索用 XSLT に渡す。
- ② データベース本体である XML ファイルを読み込み、サーバのメモリ上に DOM ツリーとして展開する。
- ③ 受け取ったキーワードを元に XSLT は抽出関数を使ってサーバ上に展開された DOM ツリーに XSLT 変換を行い、キーワードに該当するデータのみで構成される DOM ツリーを生成する。
- ④～⑤ 抽出したデータの DOM ツリー内に格納されたデータに Xpath を使ってアクセスし、HTML タグに埋め込んで表示する。検索結果数によって条件分岐を行うことで、分割・一覧表示を使い分け、ユーザーに負担をかけない表示ができるようになっている。

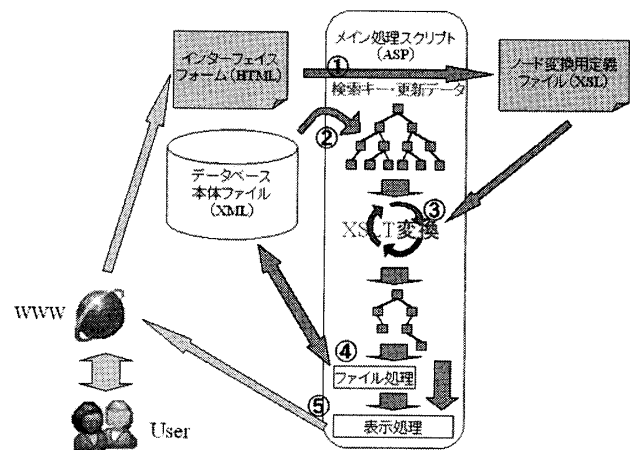


図3 システムモデル

データの削除・更新等は、検索キーの代わりに更新データや削除データのキーを XSLT に渡し、そのキーに基づいて XSLT 変換された修正データを DB 本体である XML ファイルに書き戻すことで実現した。

3. 1 システムの課題と今後の目標

上記実装によって、インターネット経由での XML-DB アプリケーションが実現できることが確認できた。ネイティブな XML ファイルを、そのまま DB ライクに扱えることが本システムの大きな特徴であり、一定の結果を得たものとする。システムの課題としては以下のことが挙げられる。これらは、今後の課題として取り組んでいく。

- ・複数の検索キーによる検索、検索結果に対する絞込み検索
- ・認証、セキュリティ処理
- ・横断検索
- ・地理的、空間的データとの連携

4. 考察

RDB 形式で保存されているデータをあえて XML に変換した上で DB として使用する、つまりネイティブ XML ベースのデータベースとする意味合いは、スキーマの変動に柔軟に対応できることにある。考古情報分野における XML 記述はまだ初期段階であり、考古学業界全体で標準となるようなスキーマは今のところ存在しない。データ内には最低でもダブリン・コア、ISAD(G)等の国際的な枠組みへの準拠や遺跡・遺物の位置情報を含ませるべきとの議論もあるが、明確な指標が決定されていないのが現状である。本研究において使用したスキーマも独自のものであり、考古学分野における他の XML 記述関連の研究も独自のものが使用されており、今後もスキーマが変動することが予想される。あらかじめ XML でデータを記述しておけば、XSLT による変換をかけることで、これに柔軟に対応できるのではないかと考える。

さらに、XSLT によるスキーマ変換は、柔軟な情報の再利用の一助になるのではないかと考える。本研究のデータは土偶データベースを念頭に設計されたため、主キーを土偶データとし、その関連情報として遺跡データや文献データを所持している。しかし、XSLT 変換を適用することによって、本研究の XML データを遺跡データベースや文献データベースのものに変換し、その一部として組み入れたり横断検索の要求に応えたりすることができるようになると考えられる。

5. おわりに

データを XML により記述することの魅力は、プラットフォームに依存しない情報の流通や再利用が行えるところにあるとされることが多いが、そればかりではなく、XSLT によるスキーマ変換の利点を享受できるところが大きいのではないだろうか。

XML による DB の現状では、速度・普及率の等の観点からみて、XML による DB はまだ性能の上で RDB と比較して分が悪い。しかし、RDB も登場当初は従来の DB には太刀打ちできなかった事実がある。今、XML による DB もこの段階にあるのではないか。今後、XML に対する研究が深まることで、これらの問題も解決されてくるのではないだろうか。目先の現象に捉われるだけでなく、将来を見据えた情報記述をしていくのが大切なのではないのではないかと考える。

参考文献

1. 杉本豪:考古学における XML の可能性ーヘリフォード州遺跡記録管理所の航空写真画像データベースを例に-, 情報考古学 Vol. 9 No. 1, pp13-26, 2003
2. 石川正敏, 波多野賢治, 天笠俊之, 植村俊亮:歴史文献画像に関する注釈の管理と再利用, 情報考古学 Vol. 9 No. 2, pp11-23, 2003
3. 後藤真, 柴山守:正倉院文書復原過程の XML/XSLT による記述, 情報知識学会誌, Vol. 11, No. 4, pp2-16, 2002
4. 國島丈生, 鈴木美沙, 宮川由香, 横田一正:構造化文書の異種性解消のための「文書群」の導入と検索機能の実現, 情報処理学会研究報告 2000-DBS-121, 2000-FI-58, pp137-144, 2000
5. 五島敏芳:「国際標準:記録資料記述の一般原則; ISAD(G)」と XML の利用ーISAD(G)第2版準拠資料記述による XML 利用実験を中心に-, 古典学の再構築:情報処理(A03)班主催「XML pro/con -XML で書く文献学的データ」報告, 九州大学, 2001/10/27