

歴史系資料情報を中心としたデータ記述とその活用に関する研究

メタデータ	言語: jpn 出版者: 公開日: 2008-10-24 キーワード (Ja): キーワード (En): 作成者: 八重樫, 純樹, 小笠原, 和慶 メールアドレス: 所属:
URL	https://doi.org/10.14945/00002674

歴史系資料情報を中心としたデータ記述と その活用に関する研究

A Research of Data Description and Application Focus on Historical Related Information

小笠原 和慶 八重樫 純樹
静岡大学大学院情報学研究科

Kazuyoshi Ogasahara Jyunki Yaegashi
Graduate school of Informatics, Shizuoka University

概要：本論文では、東北地方の縄文遺跡データと土偶関連情報に関するデータという2種類のデータをXML記述し、両者のスキーマ差異をXSLTによるスキーマ変換機能を用いて吸収しながらデータ統合を行う。さらに統合済みデータを地図表示ソフト「カシミール3D」上にマッピング表示する。これらの試行を通じ、CIDOC-CRMならびに緯度経度座標値を用いた問題解決案や歴史系資料記述のための3+1層モデルを示す。

キーワード：XML, XSLT, スキーマ変換, 考古学資料情報, データ統合, 緯度経度座標値

Abstract: This thesis will show descriptions about two types of date which are historical sites date of the Jomon era in Tohoku region and date of clay figurine, and also show date unifications with absorbing differences of each schemas using XSLT technology. Then, point each site position on map of "Kashimir 3D" by these date. Through these experiments, this thesis will show solutions by CIDOC-CRM or Coordinates of Latitude and Longitude, and also show the "3+1 layer model" for historical related information.

Keywords: XML, XSLT, Transformation of Schema, Archaeological Date Information, Date Unification, Coordinates of Latitude and Longitude

1. はじめに

一概に「歴史系資料情報」と述べてみたところで、その種類や数、フォーマットは様々である。テキスト文書、写真、図面、はては音声・映像資料等、こういった非常に多岐にわたる資料やデータをどのように扱えば、その有効性を損なわず、また、そのポテンシャルを十分に引き出すことができるのだろうか。

近年の急速なITの進展は、考古学研究を取り巻く状況をも急速に変化させるものとなった。

デジタルという世界に事象を記述し、ネットワークという搬送路を得、これまでには思いもよらなかった手法で、新しい知見が次々と生まれている。新規情報だけでなく、これまでに蓄積されてきた既存のデータに対しても、ITという新たなツールを用いることで、今までは気づかなかった事実を浮かび上がらせることが可能となった。

その反面、増え続ける多種多様なデータをどう体系的に扱えば良いのかという問題点にも直面するようになった。また、古い既存のデータをどう有効活用するのかについても、その手法は発展

段階にある。アナログ情報として依然として残っているデータはどうすれば良いのか。現在のテクノロジーを使ったデータ活用のためには、古いデータであっても現在の土俵に持って来なければ、なかなか有効な成果を導き出すことはできない。さらに、国際的なデータ流通への対応をどうつけていくのかという問題もある。これらの問題は、ITという便利な道具を得た我々に、新たな代償として降りかかっている。

本研究においては、これらの現状を鑑み、データを体系的に扱い活用するための方策を明らかにすることとした。具体的には東北地方における遺跡のデータと縄文土偶のデータのXML (eXtensible Markup Language) 記述を行い、これらデータの統合や活用に関する試行を通じて、歴史系資料情報のデータ記述に関する問題点や考察点を明らかにし、それをもとに歴史系資料情報記述の際の方策や指針を導く。

本論文においてデータの記述方式にXMLを採用する理由は次のとおりである。

- ・XSLT (eXtensible Stylesheet Language Transformations)によるスキーマ変換が利用できる
- ・主要な情報流通インフラであるインターネットとの親和性が良い
- ・スキーマの併記が可能
- ・資料記述に関する国際的な標準、RDFにも対応可能

XMLによるデータ記述を考古学の世界に持ち込み応用する研究は、後藤らによる正倉院文書データ [1] や石川らによる木簡データ [2] 等、まだまだ主流とは言えないまでも、最近では多くの事例が報告されるようになってきている。しかし、異なる主題のデータに対して、そのスキーマ差異を吸収し結合した事例はほとんど見受けられない。歴史系資料情報の持つポテンシャルを存分に発揮するためには、同一分野・同一主題のデータだけではなく、これまでは連携の図られなかった異分野のデータに対しても、

その興味を向けるべきである。こういった類のデータを連結・統合し、お互いを比較・分析することで、今までに見えてこなかった未知の知見にたどり着く可能性は大きいものと言える。

2. 歴史系資料情報のXML記述とデータ統合

2.1. データのXML記述とスキーマ

本章では、東北地方の縄文遺跡データと全国の土偶に関するデータをXML記述し、XSLTの持つ文書変換機能を用いて両データ間のスキーマ差異を吸収し、データの統合を図る。そして、統合データ活用の一例として高詳細地図表示ソフト「カシミール3D」[15]上にマッピングを行う。これらの作業を行うに伴って明らかになった事実を元に、歴史系資料のデータ統合についての現状や問題点、モデル分析を行う。

統合用サンプルとして用いたデータは、岩手大学工学部情報システム工学科横山研究室が保有する青森・秋田・岩手県の縄文時代の遺跡に関するデータである（以下、東北遺跡データ）。主なデータ項目としては、遺跡番号、遺跡名、市町村名、所在地、時代、種別、遺構・遺物などがあり、データの総件数はそれぞれ、青森県2,845件、秋田県1,746件、岩手県5,821件、計10,412件である。緯度経度座標値（以下、座標値）については旧測地系（注1）であり、その表記方式はdd.mmsssssの形式である。当該データの元データの記述方式はCSV形式（コンマ区切り）のデータであった。

次に土偶に関するデータである。これは、八重樫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（以下、土偶データ）[3]。土偶データは、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等、全40項目から構成され、総件数は10,641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、

意の土偶IDによって識別・管理されている。土偶IDは自治体ID(5桁)+遺跡ID(4桁)+土偶コード(3桁)の計12桁の数字から構成される。静岡大学情報学部八重樫研究室においては、このデータをパソコンに入力し、Microsoft(R) Access形式にて保管している。

なお、この際に土偶調査カード作成時にはほとんど記入されていなかった出土遺跡の座標値を取得していることを付記する。取得については、データ提供者に対して5万分の1の地形図を送付し、そこに該当する遺跡箇所を記入してもらうことで、当該箇所の座標を算出するという方法で行った。取得した座標値は、新測地系(注2)であり、表記はddd度mm分ss秒の形式である。本事例においては、これらのAccessデータをVBA(Visual Basic for Applications)コードを用いてXML化し、統合を行うこととした。

統合に当たってはまず、東北遺跡データと土偶データの両者のスキーマを比較することから始めた。各データのスキーマの詳細と比較結果は次の通りである。

■東北遺跡データ

遺跡番号 5桁(市町村コード(2桁)+ハイフン+遺跡番号(3桁))

遺跡名

市町村名

所在地

時代(文字)

時代(コード)4桁。@+3桁の数字。 ※注3

種別

遺物・遺構

緯度 ※注4

経度 ※注4

■土偶データ

土偶ID 12桁(市区町村コード(5桁)+遺跡ID(4桁)+出土番号(3桁)) ※注5

遺跡ID 7桁(遺跡ID(4桁)+出土番号(3桁))

遺跡

図1 土偶調査カード

遺跡名

遺跡名読み

都道府県名

所在地

位置

緯度 ※注6

経度 ※注6

地図名 ※注7

土偶

時期

土器形式

出土遺構

通称名

遺存部位

現存最大長

つくり

特記事項

備考

図表

実測図・写真
地図、実測図、写真
所蔵者、機関
実測者、撮影者
文献
報告者、著者
発行年月
報告書、論文名
掲載誌
ページ
刊行所
記入日
記入者

この結果、データの統合に対して明らかとなった問題点は次の通りであった。

- ・同一データと異なるデータの混在
- ・IDの取り扱い
- ・座標値のスキーマおよび記述方式

これらの問題点に関しては以下の方策により解決を図った。

(1) 同一データ(重複データ)の取り扱い

東北遺跡データと土偶データ内に記載された遺跡データにおいて、両者の間で同一のものと思われる遺跡のデータがそれぞれ含まれるというのは想像に難くない。そのまま機械的に単純に統合したのでは、同一遺跡に対するデータが、東北遺跡データに由来するデータと土偶データに由来するデータの2つ存在する事態となってしまう。そのため、同一と思われる遺跡に関しては、重複が発生しないように調整することが必要となる。この処理は、データ統合を行う際に、データ比較に関する条件分岐処理を行うことで解決した。具体的な判断基準については次のID(Identifire:識別番号:以下IDとする)の取り扱いで述べる。

(2) IDの取り扱い

次に、IDの取り扱いについてである。統合に際して大きな問題となったのが両者のデータにおけるIDの差異をどうするのかということである。両者のデータに振られたID形式が異なるため、どちらかの形式のIDもしくは新しい形式のIDに統一する必要があった。

本事例においては、統合後の遺跡のIDに関しては「自治体コード(5桁)+遺跡番号(4桁)」の形に統一することとした。自治体コードは、国土地理院が発行している全国市区町村コード表(JIS X 0204)に基づくものとする。これは、現状で定まっている統一的な標準があるものに関してはできるだけそれを採用するという考え方に即したものである。なお、この自治体コードを用いたID付与方式は土偶データに採用されていたものである。

ただ、自治体というものは統廃合によって常に変化するため、恒常的なデータ管理を行う上では不向きなものであることは否定できない。座標値に基づくIDの付与(後述)ができれば理想的であったが、座標値が揃っていないデータが多数あったため、本事例においては採用を見送らざるを得なかった。

ID算出に対する具体的な手法に関しては以下の通りである。4桁の遺跡番号については、遺跡データと土偶データを、遺跡名・座標値の2項目について比較し、(1)遺跡名称が同等のものは同一遺跡とみなす(2)遺跡名称が異なっても座標値が同一のものもしくは非常に近接している場合は同一遺跡とみなすという判断基準を設けた。さらに東北データの遺跡データは3桁なのに対し、土偶データの遺跡データは4桁であったため、同一とみなした遺跡のIDは桁数の多い土偶データのものを採用することとした。それ以外の、両者の比較の結果異なると判断した遺跡については、土偶データの遺跡IDの空きNo.に対して、順次機械的に4桁の数値を割り振った。

(3) 座標値の取り扱い

歴史系資料において座標値を扱う場合、考慮しておきたいことが3点ある。

まず1点目は測地系についてである。今回使用したデータにおいては、東北遺跡データは旧測地系、土偶データは新測地系であったため、どちらかのデータの測地系はもう一方の測地系に変換する必要があった。将来的な対応性の観点から、今回は座標値は新測地系のものに統一することとした。

旧測地系のデータを新測地系のデータに変換するには、旧測地点と新測地点における基準点のズレに該当する分量の緯度経度をそれぞれ加減することで可能であるが、このズレは全国で見れば一様ではない(注8) [16]。そのため、変換に際しては3次メッシュ毎の変換パラメータに基づいた変換を行わなくてはならず、作業コストが高くなることが予想される。そこで、今回の事例においては作業の省力化のため、国土地理院が作成・公開しているTKY2JGDという測地系変換ツール [18] を使用し、旧測地系で作成された東北遺跡データを新測地系のものに一括変換した。

2点目は表記方法に関するスキーマである。現状、座標値に関しては大別して10進法による表記と60進法による表記の2種類の表記法が存在する。そして、その表記法も座標値を扱うアプリケーションソフトウェアによって、様々である。例えば、「度」「分」「秒」などの文字を交えて表現したり、各度分秒の区切りに「.」(コンマ)や「/」(スラッシュ)を挿入したりという具合である。本事例においては統合済みデータをカシミール3D上にマッピングするため、60進法に統一、度と分の区切りに「.」を挿入することにした。なお、東北遺跡に関するデータの座標値は10進法により作成されていたため、座標系変換前に60進法に変換した。進数の変換は表計算ソフト上にマクロプログラムを記述し行った。

3点目は座標値の精度についてである。緯度

経度の座標値について、どれだけの精度を持っていれば実用に耐えるのかを考えてみる。緯度経度において座標値が1秒ずれば、約30mの誤差となって現れる(注9)。それゆえ、1/100秒単位までの緯度経度を算出しておけば、その誤差は30cm程度になり、高詳細な分析を行わない限りは十分使用に耐え得るだけの精度が確保できるものと思われる。

以上、3つの観点から、統合済みデータの座標値に関しては、60進法、精度は小数点以下2桁(1/100秒)、ddd.mmssssの表記形式とすることとした。

以上を踏まえ、データ統合後のスキーマを決定した。当該スキーマは独自のものであるが、統合元の両データのスキーマを参考に、遺跡データを上位階層に、そこから出土する遺物、遺構を下位階層のものとして策定した。準拠させるべき適当なスキーマが存在しなかったために、本スキーマはあくまで本事例研究用に策定されたサンプル用としてのスキーマであることを明記しておく。

次に、各データを統合するためのスキーマ変換用のXSLTファイルを作成した。XSLTファイルの実態は、xsl:ifによる条件分岐と文字列操作関数 substring () を組み合わせたものであり、元データの記述スキーマや階層構造を統合先のものになるように変換するものである。

具体的なデータ統合の手法は、次の通りである。まず、データ元のXMLファイルに対して各々のスキーマ修正用XSLファイルを適用しXSLT変換を行い、得られた変換結果を新たなXMLファイルに書き出す。さらに、別のデータ統合用XSLファイルを作成し、XSLTのdocument () 関数を用いて変換済み両ファイルを読み込み、メモリ上で統合を行う。この際に重複データの削除および遺跡IDの再付与を行う。この統合済みのデータをファイルとして書き出すことで、両データの統合済みXMLを得た。変換のシステムモデル(図2)と出力結果は(図3)を以

下に示す。なお、これら一連の作業は Jscript を用いた ASP プログラム上で行った。

2.2. カシミール3Dへのマッピング

本事例では、XMLをベースとした情報活用の事例として、統合データをさらに高詳細地図表示ソフト「カシミール3D」上にマッピング表示し、新たな知見へ繋げるための一例とする。

カシミール3DはDAN 杉本が開発したフリーの地図閲覧ソフトである [15]。使用する地図によっては一部シェアウェアとなるが、基本は無料にて使用することができる。緯度経度を参照することによって、詳細な地図上に当該データをプロットして表示させることが可能である。このソフトは元来、登山愛好家用に作成されたGPS (Global Positioning System, 以下GPSとする) 情報の地図マッピングソフトであるが、コスト面と汎用性・視認性の高さから多くのユーザーを持っている。本事例でもこれらの点に着目して連携を行うこととした。

カシミール3Dは緯度経度のマッピングにNDBファイルという独自形式のファイルをプロットファイルとして使用している。NDBファイルの実体はタブ区切りのCSV形式ファイルであり、項目としては、プロット地点名、緯度経度、測地系、標高、レイヤ情報などから構成される [4] [17]。逆に言えば、このNDBファイルさえ得られるならば、当該アプリケーションとの連携が可能ということになる。そこで、本事例ではXSLTの機能を使い、XMLからカシミール3D用のNDBファイルを出力することとした。

本事例においてはNDB出力用のXSLファイルを作成し、遺跡の名称や緯度経度に関するデータをXMLファイルの中から抽出し、NDBファイルの形式にあわせて整形したのちに出力し、NDBファイルを得た。カシミール3Dでは、マッピングをかける座標値指示地点をレイヤごとに色分けして表示させることが可能である。本事

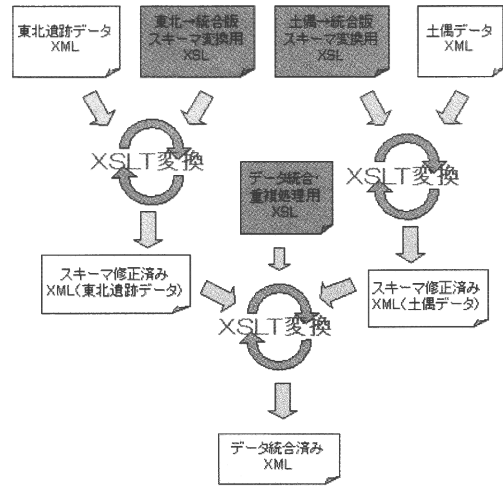


図2 データ統合のシステムモデル

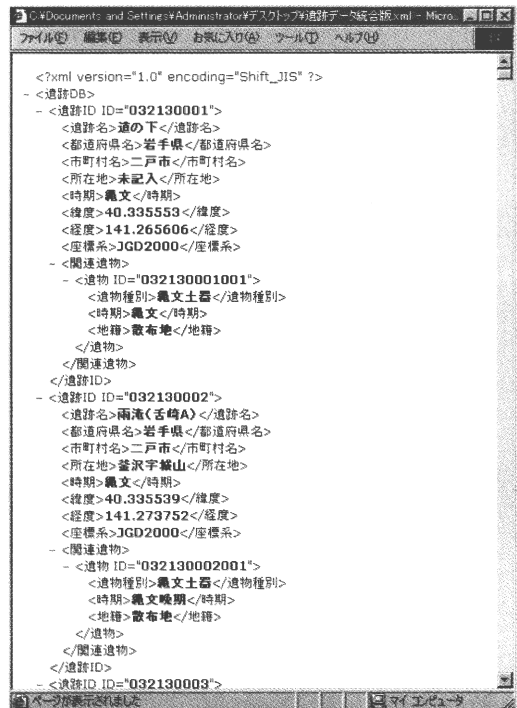


図3 出力結果

例においては、東北遺跡データを赤丸、土偶データ由来の遺跡を青三角形としてプロットした。このことにより、東北地方の遺跡の中で土偶の出土した遺跡とそうでない遺跡について視覚的に捉えることが可能となった (図4)。

3. 考察

前章の事例をもとに、本章では歴史系資料のデータ記述とデータ統合についての問題点をいくつか指摘し、所感を述べる。

3.1. スキーマ

実世界をどうデジタルの世界に表現するのか？その指標となるのがスキーマである。何をデータとして取得し、それをどのような形式で表現するのか。この出来栄が将来のデータの価値を決定づけることになる。そのため、情報記述を行う際には、スキーマについて深く議論・検討がなされなくてはならない。むしろ、IT (Information Technology) の活用が進んできている考古学の世界においても、この例は例外なく適用でき、スキーマについても多くの議論がなされるようになった。特にインターネットという世界規模での情報流通路が確保されたことに

より、データの国際流通を視野に入れた情報記述の規格統一の必要性が高まっており、それに関する議論がなされている。

これらの議論は、国際博物館会議 (ICOM: International Council of Museums) の国際ドキュメンテーション委員会 (CIDOC: International Committee for Documentation) や国際文書館評議会 (ICA: International Council on Archives) 等で行われており、その結果としていくつかの記述標準やガイドラインが策定され、これに合わせた資料記述・データ作成が世界的な主流となつつある。

その一方、日本国内の歴史系資料に関するデータ記述は、統一化という視点では非常に遅れているのが現状である。面積に対する発掘件数という発掘密度に注目すれば、日本は世界トップクラスの件数を誇る [5] が、現状では、それらの膨大な発掘データは日本国内はおろか、その各データ保有機関内のみでしか利用できな

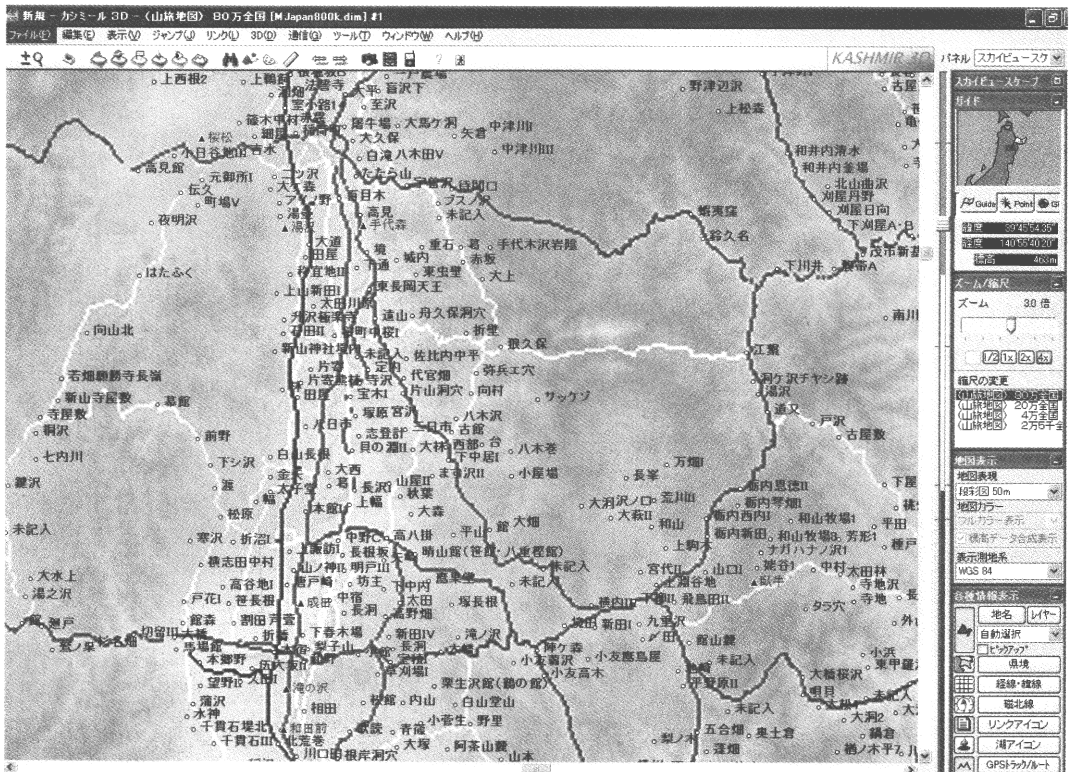


図4 統合結果のカシミアール3D上へのマッピング結果

いという状況がほとんどである(注10)。結局、そのほとんどが十分に活用できない状態で積み状態となってしまうている。膨大な量の貴重な国内データが十分に活かされないまま、世界のデータとも接続する手段を失ったまま漂流しているのは非常に遺憾なことである。

この状況を打破するためには、何かしらの国際規格に即したデータ記述・データ再整備を行う必要がある。現状、対応が図られるべき候補としては表1に挙げた国際規格が挙げられる。しかし、一概に歴史系資料情報に関する記述標準と言ったところで、準拠させる候補となる規格は表に挙げた通り、数多く存在する。そして、どの規格も単体のみでは全ての歴史系データの記述要求を満たすことはできない。例えば、DublinCoreはデータをネットワーク上で参照するためのメタデータの役割しか持たず、これを考古学研究のための資料記述や博物館におけるデータ整理のための規格として用いるには不十分である。また、考古学遺跡および歴史的建造物のための国際コアデータ標準(以下、ICOM-CIDOC案)も、対象記述分野は遺跡であって、そこから出土する遺物や遺構の記述には不十分である。

こういった議論が発生してしまう原因は、歴史資料という実体が非常に多岐に渡り、また、複数分野にまたがるものであるからに他ならない。例えば、及川は、考古学データベースにおいては、属性が遺物の種類ごとに激しく異なり細かく分類するときにきりがなく、扱う対象を厳格に共通的に定義できないため共有化が可能なデータベースの作成はなかなか実現しない、と指摘している[13]。歴史系資料という大きな枠組みで見た場合、その対象範囲は発掘した物理的資料から、古文書、ひいては衣服や生活道具等の民俗資料まで、実に多岐にわたる。さらに、例えば、古墳の壁画や資料に装飾として施された絵画は、歴史資料であると同時に美術資料としての側面を持つ。結果、全ての分野を網羅的に過不足なく対応するというようなスキーマを策定

するのは不可能に近い。

では、どれに準拠させるべきなのか。和久田らは、ダブリン・コアはあくまでも情報資源の効率的な発見のために定められた標準であって、それぞれの専門分野の情報を記述するには不十分であり、最低限MICMOへの対応が必要としている[6]。筆者は基本的にはこれに賛成であ

表1 歴史系資料記述に関連する国際規格

※資料記述に関する標準

「国際標準記録史料記述一般原則」

ISAD (G) / International Standard Archival Description (General)

→ 文書館分野の資料記述規格。記録資料用。26要素。

※データ構造上の標準

「ダブリン・コア」

DCMES / Dublin Core Metadata Element Set

→ 博物館に限らず、全ての情報資源におけるメタデータの基本。ネットワーク情報資源用。15要素。

「博物館資料の最小限情報分類」

MICMO / Minimum Information Categories for Museum Objects

→ 1994年8月ワシントンの国際会議にて提唱。博物館資料用。17要素。

「博物館資料情報の国際ガイドライン」(CIDOC情報カテゴリ)

IGMOI / International Guidelines for Museum Object Information

→ 博物館資料用。74要素。

「CIDOC概念参照モデル」

CIDOC-CRM / CIDOC - Conceptual Reference Model

→ ISO/TC46/SC4/WGとして2000年9月より策定中(ドラフト段階:ISO/CD 21127)。84概念+139プロパティ。

※考古学の核となるデータ標準

「考古学遺跡および歴史建造物のための国際コアデータ標準」(考古遺跡共通データ規格国際版(案))

CIDOC draft international Core Data Standard for archaeological sites and monuments

※水嶋英治「資料認識とナレッジマネジメント」(参考文献[9])P83「博物館情報に関する「標準」の種類」を参考に作成。規格名の後ろのカッコ内は通称名。

るが、歴史系資料を体系的に扱うという広い視点で見た場合、これだけでは不十分であり、もう少し付け加えをした提案をしたい。

筆者の考えは次の通りである。データ流通・活用のための必須メタデータ項目としてのDublinCoreを必須とし、これをベースにして、ISAD (G)、MICMO、CIDOC案等、各データの分野特性に適した記述規格を採用する。具体的な記述に関してはRDF (Resource Description Framework) と互換性を持つXML記述を用い、データ流通の際の各標準間の概念差異についてはCIDOC-CRMを用いて解決を図るべきだと考える。これら国際標準は、データ活用・流通のための最低限の基準としての規格群という位置づけであり、各データへのアクセスポイントを相互に確保しておくための一種の制約的な性格を持つものであると考えて良い。また、こういった各標準では直接的には触れられていないが、より円滑で安全なデータ流通のためには、データの取り扱いや管理ルールに対しても明確にされておく必要がある。

3.2. ID (Identifire)

次は遺物や資料に関するID付与に関する問題である。現状では歴史系資料のID付与についての標準というものが存在しない。本事例における縄文遺跡データベースのデータにせよ、土偶データベースのデータにせよ、あくまで便宜上に割り振られたIDに他ならない。今回のデータ統合を通じて、遺跡、遺物、ひいては歴史系資料のID付与について何らかのガイドライン策定の必要性を感じた。IDの取り扱いについては各国際規格においても、必須項目ではあるが、その具体的な付与方法については述べられていない。国内で統一的な付与指針を設け、さらにこれを国際的な流れとなるように情報発信していく必要もあると考える。これに関連して筆者の考えた座標値を盛り込んだIDについては後述する。

3.3. 不完全データ・歯抜け状態となったデータの取り扱い

今回の事例においても、例えば、座標値が記入されていないもの、項目に未記入・不明と記されたもの、空欄となっているもの等が散見された。また、たとえ記入がなされていても、あるデータに至っては、遺跡名:長野市 所在地:長野市とだけしか記入されていないものがあり、そのデータの実効性に疑問を抱かざるを得ないものも存在した。

データ項目が未記入となっている場合、再調査によりデータを取得しなおすか、推定によりデータを算出し補うかの手法が考えられる。しかしながら、多くの遺跡や遺構は発掘調査によって破壊されたり、事後の開発によってサイト自体が失われていたりするケースが多く、再調査が困難であることがほとんどである。報告書に記録されたデータが唯一それらを知る手がかりとなることも少なくない。肝心のデータが報告書に記載されないまま発掘が終了してしまった箇所については、データの再取得は事実上不可能であろう。また、報告書からのデータ作成時に漏れたデータは、報告書からの再入力をすれば済むことであるが、この作業には多くの労力がかかることが予想される。データ取得・作成のガイドラインをしっかりと策定し、データ作成時にきちんとしたデータ作成を行うべきことはもちろんだが、現状で残っている不完全なデータに対しても、何らかの方策を講じて再利用可能なところにまで持っていくことが必要である。

結果、現実的な手法としては、周辺データより推定が可能なものに関しては、プログラムの算出し、推定する手法が挙げられる。もちろん推定データには「推定による参考値」という具合に、正規データと区別するための何らかの標準的仕組みが必要である。

3.4. 緯度経度座標値

座標値の記載に関しては、ICOM-CIDOC案で

も必須記載項目となっている。現状では、日本では座標値の取得されていないデータは多数存在 [9] し、こういったデータに対する座標値の付与を何らかの手法で早急に行い、座標値データの整備を行うことが不可欠である。国際的な歴史系資料記述の流れからしても、最低限どのデータにおいても、座標値の付与は確保されておかなければならない事項である。また、住所による所在地指定は文字で表現された間接的な位置参照情報 [8] でしかなく、あくまで絶対的な位置情報としての緯度経度情報をデータ内に加えることの意義は大きい。

現状で未記入・未取得の座標値データに対して座標値を取得する方法については、以下の3通りの手法が考えられる。

まず、GPS 機器を用いた再取得である。現在も遺跡や遺構が発掘後も失われずに残っており、出土地点が明確にわかるのであれば、再調査が可能であり、再び当該箇所へ赴き、GPS 機器を使って緯度経度を取得することができる。GPS 測量機器の精度は年々上がっており、かなり高い精度で座標値を取得することが可能である。しかし、直接現地へ赴くことへの人的・作業的コストは莫大なものとなるのと、実際には再調査の可能な箇所は事後の開発や発掘に伴う遺跡層の物理的破壊によって、ごく一部に限られてしまうという問題点がある。

次に、デジタイザという機器を用いて、遺跡地図等より再入力する手法が挙げられる。デジタイザによる座標点のプロットは、既存の遺跡地図等が残っておれば即座に座標値の再取得が可能である。現地でのGPS測量によるデータ再取得が難しい場合には、大きな効果が上がるものと期待される。しかし、手作業による座標点のプロットは作業者の負担が大きく、誤差も発生する。例えば、10万分の1の地形図に記載された遺跡地図からマッピングを行った場合、1mmのズレが実際には100mのズレとなって現れる。これら人海戦術により座標値を再取得する手法は、結果として莫大なコストを要する作業

となってしまうため、作業にかかる人的・経済的な支援を行う法的・制度的な整備が必要となる。そのため、データ再整備には長い時間がかかることが予想される。

現状で一番現実的な手法としては、アドレスマッチングにより、所在地住所から緯度経度を算出する方法である。

国土交通省国土計画局は「JNS 住所認識システム」という所在地住所から緯度経度を算出するソフトウェアを公開し、無償提供を行っている [19]。これは、同局が整備・公開している、全国街区レベル位置参照情報を取り込み、ローカル環境内に住所と座標値の対応データベースを構築することで、入力された住所をマッチングにより座標値に変換する機能を実現している。街区情報は全国のもので作成・公開されており、同局のWebサイトより自由にダウンロードすることが可能である。この手法を用いれば比較的低コストで座標値を取得可能であるが、変換元の座標値データが街区の代表地点のものであるため、その変換精度に最大で数100mのズレが生じる可能性がある。現状では、変換率・変換精度の面で詳細な研究に耐えられるだけのレベルには至っていない。

民間では街区の代表地点だけでなく、住所の枝葉番号にまで対応したデータベースを作成し精度の高い変換を行っている企業もある [20] が、これらのサービスはすべて有償となるため、万単位の数で変換が必要な歴史系データを変換すれば、結局かなりの金銭的コストがかかることが予想される。また、一つの住所内に複数の遺跡が存在したり、複数の住所にまたがって遺跡が存在する場合があったりと、考古学データ特有の問題により、たとえこれらの高詳細なデータベースを利用しマッチングを行ったところで、どこまでの変換精度を確保できるのかは未知数である。現状では、変換率・変換精度の面で詳細な研究に耐えられるだけのレベルには至っておらず、この分野での更なる研究の進展があることを期待する。

結果、現状では、それぞれの現状に応じた座標値の再取得および活用を行うしかない。点として座標値を取得した場合は、それはあくまで対象の抽象化情報であり、小縮尺の地図における利用に限られるということを念頭に置かなければならない。さらに複数点を取得し、遺跡を範囲として認識し比較検討を行いたい場合は、その記述方法や誤差、中心点について深く議論されなければならない。

以上のような国際規格への対応を意識した座標値の整備の他に、国際規格においては触れられていないが、より活発でシームレスなデータ流通のためには、座標値自体の座標系や表記形式、精度に関しても、ある程度の体系的統一を目指すべきである。1章において行ったデータ統合事例では、緯度経度の記述フォーマットの違いが大きな作業コストとなって現れた。将来的な活用の可能性を考慮すれば、現状で保有している旧測地系データはプログラムの演算をし、新測地系のものに変換しておくべきである。また、60進10進表記を含め、各GIS (Geographical Information System: 地理情報システム) ツールやアプリケーションにおいても進数や表記に関する何らかの統一が図られることを期待する。

3.5. データアクセスポイントとしての座標値

緯度経度による地点の指定は、各分野共通の重要なデータアクセスポイントであり、異種情報を結びつけるための貴重な糊代的役割を果たすものとなる。世の中には歴史系資料情報だけでなく、座標値の記載されたデータは多種多様なものが存在する。例えば、商業施設の立地データ、防災地図、地域データベース、カーナビゲーションシステムのデータベース等が挙げられる。こういったものと、歴史系情報を座標値をもとに関連付けることで、これまで思いもよらなかった新たな知見が発見される可能性がある。

また、各種発掘調査や研究には、公共財から

の人的・金銭的投資がなされており、基本的にその成果は広く一般にも還元されるべきものである。研究結果は研究機関・研究者個々のものだけでなく、広く大衆に帰属すべき性格のものであるという意識を持ち、考古学研究におけるアカウントビリティを満たすために、歴史系資料情報の公開も積極的になされなくてはならない。これらデータの社会情報資源としての価値を決定付けるのは、そのデータの利用性に依存するが、多種情報に対するデータ連結点としての座標値の有無がその大半を決するということは想像に難くない。

さらに、座標値+出土番号という表記形式を用いることで、遺物を一意に特定できる可能性があり、資料のID的役割を果たすようになる可能性もある。将来的にはID内に座標値を取り込

<p>座標値を元にした資料ID ある遺跡のある地点から出土した資料に対するID付与例 出土地点：東経 135度 24分 11秒 21 北緯 39度 41分 23秒 55 出土番号：026 番作成 ID：E135241121.21-N394123.55-026</p>

図5 座標値を元にしたID付与例

むことも検討しても良いものと思われる(図5)。このID法の導入に関して想定される問題点としては、データをIDの昇順・降順に並べて一元管理することが難しくなる、IDの桁数が大きくなってしまふ等が挙げられる。しかし、この手法は非常にシンプルな方法で歴史系データを一意に扱うことが可能であり、今後の議論の価値がある。さらに、

保管場所・所有機関の座標値を順次別項目に入力しておけば、資料の移動に関する追跡も可能となる。

遺跡だけではなく、遺物に対してもできる限り正確な座標値を記載するべきである。この理由は以下のとおりである。

考古学に関連するデータは遺跡∈遺構∈遺物という階層関係を持ち [11] (注11)、その階層関係の論理構造を考えた場合、遺物は遺跡というサイト上に内包される。そして、基本的に遺物は遺跡より複数発掘される。

そもそも遺跡は2次元的な広がりを持った「サイト」として存在する。しかし、データ表現の問題(各標準における要素数の制限)から、遺跡の場所を表現する座標値は1箇所しか記載されないことも多い(注12)。そのため、その遺跡の代表地点もしくは中心地点、重心地点の座標地が、遺跡の「位置」として取得される場合が多い。しかし、遺跡の実態は「面」であるため、遺跡という実態を正確に電子化して表現しているとはいえない。

ここで、各遺物における出土地点の緯度経度が精密に取得・記載されていれば、複数の遺物の出土場所を地図上にプロットし、その集合を「面」として表現することが可能となる。精度やどこまでの出土範囲を遺跡とするのかの議論はもちろんなされなければならないが、少なくとも、遺跡をただの地図上の一地点として捉えるよりは遺物の出土範囲の集合を面として捉え、それを遺跡の広がりとして表現する方が、より実世界を的確に描写した表現方法と言える。これは、「遺跡は出土地点の集合である」という考えに基づくものである。

さらに、この考え方は、遺跡の地形分析にも応用できる。出土資料の標高データを取得しておけば、座標値と組み合わせることで、遺物の位置を3次元的に特定することが可能である。そのため、複数の資料の分布を分析することで、当該地域が平面地であったのか、斜面地であったのかという大まかな地形を推定することもある程度可能となる。さらに精密な標高データを取得しておけば、各遺物の層的な管理も可能となることが予想される。CIDOC案でも標高データに関しては任意記述項目であるが、その記載が規格として定められていることを付記しておく。

3.6. 歴史系資料情報のためのデータ記述に関するモデル

本章の最後に、歴史系資料情報のためのデータ記述に関するモデル考察を行い、問題解決を図る上での手がかりとしたい。ここで筆者は、歴史系資料情報のためのデータ記述に関して、「歴史系データ記述のための3+1層レイヤモデル」というものを提案する(図6)。

歴史系資料のデータ記述を円滑に進めるためには、基本的にどのような項目のデータを取得・記載すべきかというレベルの話と、データ活用のためにどのような表現形式を用いてデータ記述を行うかという話は、分けて考えた方がよい。その結果、データ記述のための基本的な階層は、下からスキーマ層、フォーマット層、プレゼンテーション層の3層構造となると考えるものである。

スキーマ層は、実世界の情報について何をデータ化するか、その取舍選択を決定する層である。次に、フォーマット層は、選択されたデータに対して、そのデータを媒介するためのメディアを決定する。その上で、プレゼンテーション層において、データをどのような形式で表現するのかを決定する。

さらに、それらを情報資源体系的に扱い、有効な利用に供するために記述系レイヤの上位レイヤとしてマネジメント層が必要となる。マネジメント層は、各データをオブジェクトとして体系的に扱うための情報を設定する層である。下位3層が主にデータの記述方法を扱う層であるのに対し、この層ではそれらを有効に活用するための管理情報を扱う層である。これらの3+1層構造により、歴史系資料情報のデータ記述について一般化できると考える。

そして下位レイヤが定まらない限り上位レイヤの活動余地はなくなる。現状ではその根本となるスキーマ層の不備が目立つ。インターネットでのOSI基本参照モデル同様、障害解決のためには、下層からの確実な解決を図って行かなければならない。歴史系資料情報の円滑な流通

のためには、まずは、その大元であるスキーマ層の問題解決を図り、それが終了した時点で、初めて上位レイヤの問題解消に取り掛かれるのであると肝に銘じなければならない。まずは早速なるスキーマ問題の解決を行うべきである。

4. 結論

歴史系資料情報をデータとして蓄積し公開することは世界的流れとして進められており、各データ標準規格はその中心をなすものである。多種多様な歴史系データを全て網羅し過不足なく画一的に記述できるデータ標準は現実的ではなく、データの特性を見ながら複数の規格をうまく取り入れ、それらを連携させながらデータ記述をしていく必要がある。データ流通のための規格準拠のレベルを保ちつつ、それぞれの研究要求に応じたデータ粒度を確保することが、歴史系資料情報記述の今後の方向性であると信ずる。

多分野にわたるデータをうまく住み分け、ま

た共存させるには、共通の思想を持った枠組みの中でそれらを体系的に扱う必要がある。本論文においては、それらをスキーマの相互参照のための CIDOC-CRM と、それを実現するための枠組みを提供する XML 及び RDF という形で提案した。そして、データを有機的に結びつけ、新たな知見を発見するための各分野共通のデータアクセスポイントとして緯度経度座標値を位置づけ、これの整備の必要性を強く訴えた。

データはそれ自体の持つ価値以外に、他のデータとの連携を取ることで初めて浮かび上がってくる価値というものが存在する。自己の持つデータのみで知見を生み出すという研究は、長い歴史を持つ考古学研究の中でそろそろ限界の位置に差し掛かりつつある。他分野、異種データとの関係比較・関連という新たな視点を持ち、マクロ的な視点から事実を俯瞰すれば、今まで想像もつかなかった未知なる知見に到達する可能性が高い。

本論文においては、そのためのデータとデータをつなぐための糊代的役割としての緯度経度

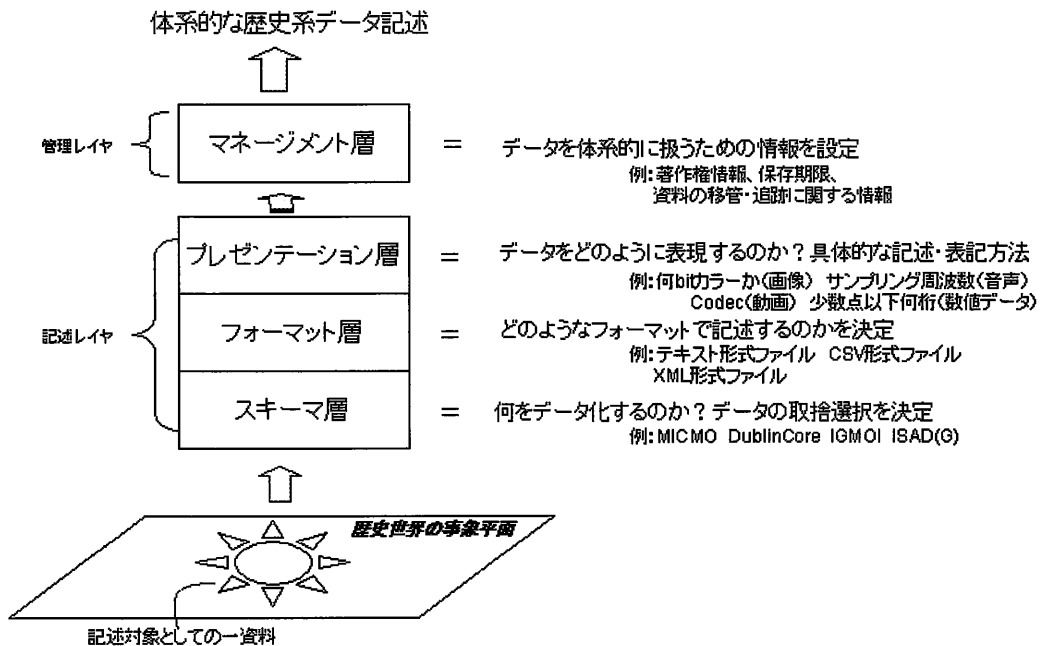


図6 歴史系資料記述のための3+1層モデル

座標値に注目した。緯度経度座標値は今後の研究スタイルやデータ利用形態を考えれば、もはやデータの価値を決定付けるものになりつつある。また、緯度経度を中心とした、より活発なデータ連携のためには現状では標準化されていない、座標値に関する表記フォーマットに関しても議論していくべきである。

[注]

- 注 1 日本測地系。いわゆる旧日本測地系 (TOKYO)
- 注 2 新日本測地系 (JGD2000・日本測地系 2000)。日本の測地系を世界測地系 (WGS84) に対応させなおしたもの
- 注 3 数字で時代を示す。複数の時代がある場合、時代コード欄には区切り文字なしで続けて記述。
(凡例)
- @020: 縄文時代とだけ記載されているもの
- @023: 縄文時代草創期
- @024: 縄文時代早期
- @025: 縄文時代前期
- @026: 縄文時代中期
- @027: 縄文時代後期
- @028: 縄文時代晩期
- @029: 続縄文
- 注 4 10進法記述。旧測地系。小数点以下7桁。必ずしもすべての遺跡の緯度経度が入っている訳ではない。
- 注 5 市区町村コードは総務省制定市区町村コード (JIS X 0402) に準拠。
- 注 6 新測地系。住所を元に国土地理院2万5千分の1数値地図より手作業にてプロット。60進法。小数点以下4桁。
- 注 7 国土地理院発行 5万分の1地図名に準拠
- 注 8 東京付近では東西方向約450m、博多付近では同約420m、北海道宗谷岬付近においては同約400m。
- 注 9 ズレは均一に起こるものではなく、ある

程度の幅を持ったズレとなる。緯度においては約31m、経度においては東京付近で約25m、札幌:約22.6m、那覇:27.8mのズレとなる。

- 注10 アジア歴史資料センターのアジア歴史資料データベース等、一部機関では国際的な対応が見られる (同センターは DublinCore, ISAD (G) に準拠) が、ほとんどが未対応。
- 注11 「 \in 」は集合に属することを意味する数学記号。「 $x \in S$ 」は、 x が集合 S の元であることを意味する。
- 注12 CIDOC案では、遺跡範囲を指定する方法として、点の順序として捉えることが必須項目と定められている。しかし、現状では遺跡の所在地に関して複数の座標が取得されている例は、日本においては稀である。

参考文献

- [1] 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.
- [2] 石川正敏; 波多野賢治; 天笠俊之; 植村俊亮; 勝村哲也: XMLを用いた木簡画像共有システム, 日本情報考古学会 第16回大会, pp91-96, 2003.
- [3] 八重樫純樹: 国立歴史民俗博物館研究報告 第37集 土偶とその情報, 国立歴史民俗博物館, 1992.
- [4] 杉本智彦: カシミール3D GPS応用編, 4.8 測地系, pp80-85, 実業之日本社, 2002.
- [5] 堅田直: 総論—21世紀を迎えた情報考古学, 電子情報通信学会誌 Vol.85 No.3 pp146-147, 2002.
- [6] 和久田聖衣, 八重樫純樹: デジタルアーカイブの調査研究—博物館情報の標準化動向を中心に—, 静岡大学情報学研究 第10巻, pp127-146, 2004.

- [7] 安藤正人 編：「用語解説 現代考古学の方法と理論Ⅲ」, 同成社, pp65-75 pp211-217, 2000.
- [8] 碓井照子：21世紀における高度情報化社会の情報インフラ整備, 機関誌「KIIS」Vol.109, (財)関西情報センター, 1998.
- [9] 水嶋英治：資料認識とナレッジマネジメント 博物館情報学の立場から, アーカイブズ学研究 No.1, pp78-86, 2004.
- [10] 牧野哲；八重樫純樹：遺跡データベースに関する海外と日本の動向, 日本情報考古学会第16回大会, pp63-pp70, 2003.
- [11] 八重樫純樹：歴史系支援情報処理研究の課題, 国立歴史民俗博物館研究報告第30集, pp3-24, 1991
- [12] 小笠原和慶, 八重樫純樹：考古学資料情報のXML記述とXSLT—RSSを用いたデータ配信への応用, 日本情報考古学会第19回大会
- 会
- [13] 及川昭文：考古学データベース —過去を復元するマルチメディア技術—, 情報処理, Vol.38, No5, pp338-391, 情報処理学会, 1997.
- [14] 小笠原和慶：考古学情報のXML記述とその応用に関する研究, 静岡大学大学院情報学研究科修士論文, 2005
- [15] カシミール3D, DAN 杉本,
<http://www.kashmir3d.com/>
- [16] 日本測地系と世界測地系, 総務省統計局,
<http://www.stat.go.jp/data/mesh/08-5.htm>
- [17] カシミール / 測地系 (DATUM) について,
http://www.kashmir3d.com/kash/manual/map_datu.htm
- [18] TKY2JGD 国土地理院 <http://vldb.gsi.go.jp/sokuchi/tky2jgd/download/agreement.html>
- [19] JNS 住所認識システム, 国土交通省国土計