

奈文研版遺跡データベースの構築

森本晋

奈良文化財研究所

全国文化財データベース構想

文化庁は全国文化財データベース構築計画を1987年度から進め設計を開始した。文化庁管下機関では1988年度を予備調査の第1年とし、基本構想・基本計画の立案にあたった。当初の構想では、文化財のうち土地と結びついている「不動産文化財」と移動が可能な「動産文化財」とに分け、不動産文化財全国センターシステムを奈良国立文化財研究所（当時、以下「奈文研」）に、動産文化財全国センターシステムを東京国立博物館に設置することになっていた。

動産文化財にかかわる機関は当時の構想内であってさえ、東京国立博物館、京都国立博物館、奈良国立博物館など多数にわたっているのに対し、不動産文化財にかかわる機関は奈文研だけであった。このため、動産文化財データベースの構築には、各機関の調整が不可欠であったが、不動産文化財は情報の中核機関としての奈文研そのものの取り組みを検討することが第一という違いがあった。建造物や遺跡といった不動産文化財の中心となる対象を直接取り扱う機関は全国の都道府県市町村教育委員会など数が多いが、このころは、コンピュータの普及、インターネット接続ともに進んでいなかったため、奈文研がセンターとして機能すること自体には問題は少なかった。

不動産文化財の基本理念の検討を進め、まず遺跡データベースの構築を進めることになり、そのための検討が行われた。遺跡データベースの項目検討を行い、全国遺跡データベース検討会を開いて教育委員会・埋蔵文化財センターなど関連する組織における情報処理の動向を調査した。

第2年目である1989年度は予備調査の2年目で、1990年2月27日に全国から関係者を集めて「全国遺跡データベース検討会」を開催した。広域遺跡の扱いについての検討を中心に進めた。インターネット接続をどうするのかという点に関心が向きがちであったものの、データベースを整備する必要性などは共通認識とすることができた。

その後、奈文研内部で検討を続ける時間が流れる。第3年目である1990年度は概念設計を外部委託で進めた。動産文化財部門を担当する関連機関との打ち合わせも進めた。また、奈文研内部職員の情報処理システムに対する要求を調査した。

第4年目の1991年度はセンターシステム実施設計の初年度と位置づけ、全国不動

産文化財データベースセンターシステムの基本設計を外部委託で行った。内容は、センターシステムに必要な機能およびデータ構造の概要の設計、テストデータの入力、奈良県に関する遺跡データの収集、入力用ワークシートの作成である。このほか、関連機関との打ち合わせも続けた。また、データの入力部分についてパソコン用データベースソフトを用いて入力部分のアプリケーションを作成した。

第5年目はローカルシステム実施設計の初年という位置づけである。奈文研ローカルシステムの設計を外部委託で進めた。内部職員にデータベースシステムに対する要求を調査し、業務で使用するデータベースについて既存も構築中も含めて、項目設定、情報の発生契機、データ構造の関連性について調査した。また、ネットワークを通じての提供が必要な共用データベースとすることができるものを選択した。

1992年9月にはこれまでの検討成果を『埋蔵文化財ニュース75号』として公表した。これは、「全国文化財データベースについて」という特集号で、不動産文化財データベースを構成するデータベースの構造を記述している。データベースを実現するためのハードウェア・ソフトウェアという環境は変化を続けているが、フィールドの基本的構成やいくつかのコードは現在も当初のものを継承している。

第6年目となる1993年度はセンターシステム実施設計の2年目という扱いである。センターシステムのプロトタイプを構築する設計を外部委託で行った。テストデータを追加入力するためのアプリケーション作成とテストデータの追加入力を内部で行った。また、外部委託で、各種データベースマネージメントソフトウェアについて性能評価を行い、データベースサーバ機器の性能評価を参考にしてプロトタイプ用のハードウェア、ソフトウェアを決定した。

データベースマネージメントソフトウェア上での遺跡データベースの構築とテスト運用を外部委託で行い、検索応答速度、ユーザインターフェイスについて評価を行った。

1994年度は第7年目でありセンターシステムプロトタイプの施行を行った。これは、データ収集、データ入力、データ追加修正、通信機能テスト、プロトタイプのグレードアップ、評価からなる。通信機能テストは、奈文研所内LANの機能を強化して接続すること、通信回線経由での接続テストであり、最終的な接続先を100箇所程度と推定して必要となる無電話回線、デジタル電話回線（ISDN）を導入した。

長く続いた構想と設計の時期には機関の間の調整会議も数多く開かれて、他機関の動向もよく把握することができた。

遺跡データベースの立ち上げ

構想を現実のものとするためには設計を進めるとともに予算の裏づけが必要である。全国文化財データベース構想全体は膨大であり、なかなか実現が困難であった。1995年度末の補正予算により、3館（東京国立博物館・京都国立博物館・奈良

国立博物館) 2所(東京国立文化財研究所・奈良国立文化財研究所)と文化庁本体に予算がつき一気に構想を進めることになった。

全国センターシステムを築くためには、各機関内部の情報システムを更新することも必要である。奈文研においても、機関内ネットワークの拡充、対外接続の再整備を基幹として改良を行った。奈文研は構成部局が地理的に分散しており機関内のネットワークを整備するのに多大な費用がかかる。整備の基本は、ネットワークケーブルのような代替が困難なものほど将来を見越した設計をするというものであった。

奈文研のシステムは、不動産文化財全国センターシステム関係、その他のデータベース関係、ネットワーク関係の3分割で競争入札し導入した。

全国遺跡データベースは、サーバにUNIXマシン、データベースマネージメントソフトウェアにOracleを用いる計画であった。

データの入力

データの輸入は1996年度の予算で開始した。ハードウェアは購入から1年間は保守料がかからない機器が多く、その分を活用することによって初年度にできるだけ多くのデータを外注入力する方針であった。遺跡情報は一括入力が見込み対象であるという印象が強いが、全国レベルの広範な資料はほとんど存在しない。文化庁主導で作成された遺跡地図がその唯一のものといってよく、これは47都道府県の全体で19万件からなる。まず、この資料をデータベース化した。文化庁の遺跡地図は、項目立てや記入が簡単である。また、刊行が長期にわたったために、新旧で内容に格差が生じていて調整が必要である。

データ入力を外注することによる弊害も大きい。文字を追って入力できる形に原資料を調整しなくてはならず、そのためには内容についての知識が必要である。業者に文化財に関する極めて初歩的な内容から、詳細な説明を何度も繰り返すこととなって、日数も要し価格的にも利点が少なくなる。実際には入力業者は注文伝票などの入力には慣れていても、学術的な表には不得手で、入力のやり直しが発生し、納品されたものを内部で再度検品するための手間ひまが必要であり、結果はまったく満足できるものではなかった。入力時に発生する、数々の疑問にそのつど答えてよりよいデータを入力していく過程を考えるならば、外注入力は効率が悪く精度も低く費用もかかり、公費の濫用といえるだろう。

試験版の公開

データベースの中身は外注と内部の手直しの組み合わせによって徐々に充実してきた。問題はそれを動かす器である。導入時の計画では、サーバソフト、ミドルウェア、クライアントソフトという3層モデルを考えていた。各部分の負荷を軽減

して高速な動作をめざした設計ではあったが、OSを含めた各ソフトの対応関係が複雑となった。バージョン合わせにあっては、いつも「近日発売」や「対応予定」という業者の文言に振り回されてしまった。完全な稼働はどんどんと遅れいつと明言できる状況ではなかった。

このような情勢では、情報公開も研究支援も見通しがたたない。急速に変化するハードウェア・ソフトウェア環境のもとでは、当初の構想そのものの設計に固執しても得るものは少ない。実を取るべきと判断し、簡易なUNIXマシンでもととの役目はずれたものが所内にあったのでそれを活用し、整備してきたデータそのものをそのまま提示するシステムを手作りした。電子化が完了したデータを市町村別に手作業で分割し、UNIXのOSの機能で簡単に検索するものである。インターネットを介したこの試験版の公開は1996年11月のことである。

第2代システムとその挫折

いつまでも簡単な機能で満足しては情報公開と名乗ることができないので、代替案を提示しなくてはならないことになった。当初設計のシステムを棚上げしている以上、新しいシステム用に大きな予算を使うことはできない。限られた予算内で、ハードウェアは既設のものを流用し、別のソフトウェアによって全文検索を行うこととした。

採用したものは、Live Link Searchである。Live Link Searchは強力な全文テキスト検索エンジンであるOpen Textの簡易版ともいえるもので、アプリケーション部分は作成しなくてはならない。そうすると業者のプログラミング能力が直接に響いてくる。採用したソフトウェアが強力であったのがかえってあだとなり、奈文研のような小規模業者が利用している例がなく、業者から最適のサポートが受けられない状態が続いた。アプリケーションの細かな改良は対応時間が問題となる。わずかな手直しでよい部分について、業者の対応が数日から数週間後というのではシステムの完成はおぼつかない。

システムが軌道に乗る

システム構築は再び頓挫したのである。しかし、そのまま放置してよいはずはなく、より安価に確実な結果を出さなくてはならない。安価にするためにサーバをPCとしOSにWindowsを採用した。データベースエンジンにはM言語を用い、データを移行した。今回は文化財にも詳しい業者が担当することになり、システム的设计・改良が軌道に乗るようになった。

これらシステム入れ替えの期間にもデータ入力も営々と続けていた。ハードウェアも廉価なPCにしたため更新が比較的容易になった。いかに高度な機器でも情報関係機器の進歩はあまりに速く、常にシステムを最適な状態に保つには膨大な

予算と知識が必要で、変更の多いものほど簡単に扱えることが望ましい。整備にもっとも時間と労力がかかるのはデータそのものであり、かつデータそのものが重要で引き継いでいくべき資産である。

新しいシステムでのデータベースの公開は、1999年9月28日に開始した。

データソース

文化庁の全国遺跡地図は外注で一括入力したが、その後の入力資料は個別に対応することになった。情報源は大きく3つに分けることができる。

- 1 遺跡地図
- 2 遺跡地名表
- 3 発掘調査報告書（報告書抄録）

遺跡地図は都道府県教育委員会、市町村教育委員会が発行する遺跡に関する包括的な情報源として最も大切なものである。地域によって精度、更新頻度が異なり、近接自治体との情報格差の問題が存在する。地図が付加されていて位置情報の取得も可能な資料である。

遺跡地名表には、遺跡地図と同等のひとつの地理的範囲に対して包括的にすべての遺跡を網羅するものと、特定の主題に基づいて集成したものとがある。後者には、例えば、銅鐸出土地名表や前方後円墳集成といったものがある。ひとつの種類にかかわる遺跡については、非常に網羅性が高くまた情報も詳細である。ただ当然、適切な地名表が公刊されていない分野も多く、分野間の情報格差は大きい。

発掘調査報告書は遺跡に関する最も詳しい情報源である。ただし基本的に個別の資料であり全国の遺跡データベースを作成するという立場からは利用にあまりに労力がかかる。大部な報告書が多く、一瞥しただけではどのような情報が盛りされているのかわからないという問題もある。そこで、抄録を添付してもらってデータベース化を容易にしてもらうことになった。抄録の整備によって、抄録データベース以外に遺跡データベースにも貢献が大きい。

入力時の確認作業

各種の資料からの入力を継続していくと新たな問題が発生する。一番大きな問題は、遺跡のマッチングである。既入力情報のどれが、今手にしている情報にあたるのかを判定しなければならない。何を基準にするのか。

遺跡にユニークな番号が割り振られていたらと考えることは簡単である。しかし、遺跡は調査によってその範囲も頻繁に変更となる。遺跡が地理的な位置にいわば貼りついて存在しているのであれば、場所が異なれば別の遺跡ということになる。しかし、極端な変更でなければ同一遺跡として扱うべきであろう。

結局のところ判断の基礎となるのは、遺跡名である。遺跡名も頻繁に変更され

るがその履歴を正確に記録して提示することが大切である。その上で、ひとつのレコードとして扱うのか別レコードとして登録するのかを判断する。

いずれにしてもマッチング作業は「手作業」となる。奈文研版遺跡データベースに対して、他機関が作成した遺跡データベースとのデータの相互交換を持ちかけられたことがある。遺跡の同一性の認定を経緯度の近似のみで行うという構想では、とても確実な実現は望めない。既入力情報の仕分けや典拠の確認など、やはり遠回りのようでも熟練者による手作業が最も速いやり方でもある。

遺跡データベースにおける位置情報

遺跡がどこにあるのかは基本的で重要な情報であり、データベースに欠かすことはできない。位置情報の源もほかのデータと同じである。すなわち、

- 1 遺跡地図
- 2 遺跡地名表
- 3 発掘調査報告書（報告書抄録）

である。

遺跡地図は地図が主要な構成要素となっていて遺跡の位置が記されているのは当然のことである。では、それらは利用しやすい状態になっているだろうか。遺跡地図は紙に記録されていて、そのままではデータベースに取り込むことはできない。そこで、地図入力用のプログラムを作成して対応することとなった。

ここで問題となったのは遺跡の位置とは何かという点である。奈文研版の遺跡データベースでは、遺跡について、代表点と範囲を定義した。

代表点

代表点は、ひとつのレコードに対して、ひとつだけ決められる位置である。対象となる遺跡の面積が極めて小さい場合などでは、遺跡は代表点のみを持ち、範囲を持たない。現実には範囲は存在するのだが、表現しようとする縮尺がよほど大縮尺でない限り点として扱っても同じである。範囲が示されているのに代表点がないという遺跡は存在しない。

正確な位置が不明な遺跡については代表点を定められない場合がある。消滅した古墳について存在した市町村名しかわからない場合であっても、日本地図に位置を落とすのであれば代表点を定義できるが、通常対象としている25000分の1では地図上に表現することができない。

範囲

範囲には遺跡地図に遺跡の範囲として描かれる線を採用する。範囲の線が何を意味するかは遺跡の種類によって異なってくるし、その根拠が何なのかを吟味する

のが困難である。従って入力の基本資料の記述をそのまま採用することになる。

遺跡の範囲の意味するものについて詳細な考察はあまり成されてきていないようである。例えば、寺院の範囲と古墳群の範囲では意味が異なる。寺院の範囲はその内部の領域は一般にすべて寺院に属するものと考えられるが、古墳群の範囲として遺跡地図などに記されている線は、ほとんどの場合、点で示した古墳の位置を包含するように適当に引かれたというだけのものであって、多少ずれていてもあまり問題とならない。

一般に遺跡では遺構の密度は場所によって異なる。遺構が密集する部分と希薄な部分とを遺跡範囲として同等に扱ってよいとは考えがたい。しかし、奈文研版遺跡データベースのデータソースの制約がありそこまでの情報を収集できない。

標準化

奈文研版遺跡データベースの改良には複数の課題がある。ひとつは、内容の精度を上げる努力である。より正確でかつより詳しい情報を盛るにはどうすればよいであろうか。

正確な情報を使いやすい形で掲載するために、それぞれの遺跡のもっとも詳しい情報源である発掘調査報告書からの入力を進めている。それに合わせて、遺跡を遺跡群（集合）、遺跡（個別）、第〇次調査（調査）などの階層に分けて情報を入力するように改良した。階層化の作業は、作業量が膨大でまだその恩恵にあずかるどころまでは来ていないが重点的に進めている。

正確さを向上させるために、用語の統制も必要である。すべての用語について規制することは不可能であり有害でもあるが、ある程度の基準を提示することは大切である。用字、用語、表現法の各レベルにおいて基準の作成と改良を行っている。

標準化によって、ほかのデータベースとの協業も視野に入れることができる。以前より何度も指摘されていることではあるが、メタデータの整備、文化財情報センターへの展望を開けるような改良の道筋を作ることが重要で、そのためには他機関との共同研究が不可欠であろう。

	文化庁調査 による遺跡数 (1993.3)	奈文研版遺跡データ ベース入力済み件数 (2004.12.10)	対1993年文化 庁調査比率	奈文研版のうち経緯度情報有り 実数	比率
北海道	10,108	5676	56.2%	287	5.1%
青森	3,802	4649	112.3%	347	7.5%
岩手	8,554	6022	70.4%	405	6.7%
宮城	5,600	6252	111.6%	6053	96.8%
秋田	3,680	2036	55.3%	643	31.6%
山形	4,360	2528	58.0%	189	7.5%
福島	7,907	4829	61.1%	556	11.5%
茨城	6,609	7136	108.1%	414	5.8%
栃木	7,362	10743	145.9%	6186	57.6%
群馬	12,656	6018	47.6%	935	15.5%
埼玉	11,657	6672	57.2%	693	10.4%
千葉	25,657	6098	24.3%	1435	23.5%
東京	4,607	4901	106.4%	1039	21.2%
神奈川	7,599	6804	89.5%	739	10.9%
新潟	10,432	5419	51.9%	376	6.9%
富山	3,126	1797	57.5%	408	22.7%
石川	6,947	6874	98.9%	4056	59.0%
福井	2,929	7776	265.5%	54	0.7%
山梨	3,533	2235	63.3%	198	8.9%
長野	12,604	9162	72.7%	509	5.6%
岐阜	8,328	9135	109.7%	5065	55.4%
静岡	14,149	4508	31.9%	474	10.5%
愛知	9,319	8389	90.9%	5635	67.2%
三重	11,617	9415	81.0%	414	4.4%
滋賀	4,746	5746	121.1%	5356	93.2%
京都	12,844	19865	154.7%	6515	32.8%
大阪	7,201	7690	106.8%	3348	43.5%
兵庫	23,406	27996	119.6%	17044	60.9%
奈良	9,029	11752	130.2%	4473	38.1%
和歌山	2,862	2596	90.8%	1796	69.1%
鳥取	13,192	7779	59.0%	5649	72.6%
島根	7,313	10982	150.2%	8845	80.5%
岡山	9,331	27744	297.3%	16235	58.5%
広島	12,756	6527	51.2%	260	4.0%
山口	2,517	963	38.3%	92	9.6%
徳島	1,166	968	83.0%	93	9.6%
香川	3,551	2686	75.6%	616	22.9%
愛媛	3,549	1952	55.0%	349	17.9%
高知	2,071	2522	121.8%	2143	85.0%
福岡	16,968	19686	114.3%	1123	5.8%
佐賀	4,986	5299	106.3%	3638	68.7%
長崎	3,632	4012	110.5%	2860	71.3%
熊本	5,552	9272	167.0%	7411	79.9%
大分	11,247	5046	44.9%	3662	72.6%
宮崎	3,557	2648	74.4%	308	11.6%
鹿児島	5,319	2803	52.7%	1507	53.8%
沖縄	1,393	1821	130.7%	941	51.7%
全国	370,804	333131	89.8%	131374	39.4%

An Archaeological Sites Database Project

Morimoto Susumu

National Research Institute for Cultural Properties, Nara

An initiative to build a comprehensive database of Japan's bunkazai (cultural properties), launched in 1987 by the Agency for Cultural Affairs, has now progressed to the design development stage. In this project, bunkazai have been divided into two categories, fudosan (immovable) and dosan (movable), and a "national center system" for the handling of the data related to fudosan bunkazai has been set up at the National Research Institute for Cultural Properties, Nara (Nabunken). The system for processing the information on dosan bunkazai, on the other hand, is to be established at the Tokyo National Museum. At Nabunken, we decided to target archaeological sites in the first phase of our database-building efforts, and in February 1990 invited municipal officials in charge of bunkazai-related affairs from across the country to discuss relevant issues. Based on the results of this meeting and the findings from our own research, we selected the materials to be included in the archaeological database and announced the details in the 75th issue of *Maizo Bunkazai* (Buried Cultural Properties) News published in September 1992.

Our work commenced in earnest when funding was provided in a supplementary budget approved in late 1995. Data entry, meanwhile, was started with the implementation of the fiscal 1996 budget. One of the issues encountered at this stage was the scarcity of comprehensive archaeological resources covering the entire nation. There were, however, 190,000 sites on maps produced under the initiative of the Agency, and we decided to concentrate first on compiling their contents into the database. We have since steadily worked on data entry while also, with the limited funds available, managing to proceed with the task of migrating to a new system. Our highest priority in these efforts has been the preservation of the archaeological data, which in itself represents an important national asset.

The sources of archaeological information available for this project comprise three different format types: maps and lists of registered archaeological sites and excavation reports (summaries). At Nabunken, we have been striving to create a database that is reliable enough to serve as a primary resource by ensuring that the format differences in our data entry operation are handled.