

奈良文化財研究所におけるデータベース Database in the Nara National Research Institute for Cultural Properties

森本 晋

Susumu Morimoto

独立行政法人国立文化財機構 奈良文化財研究所, 奈良市二条町 2-9-1
Independent Administrative Agencies National Institutes for Cultural Heritage
Nara National Research Institute for Cultural Properties
2-9-1 Nijocho, Nara

あらまし:奈良文化財研究所は長年に渡り、調査研究に基づいて文化財関係のデータベースを作成してきており、その多くを公開している。データベースそれぞれの特徴について、入力内容の工夫とともにデータベース構造の特色についても紹介する。また合わせて考古資料の総合的な情報化と活用についても述べる。

Summary:Nara National Research Institute for Cultural Properties has been continuing to develop the databases about the cultural heritage for a long time and disclosed them. It is mentioned about the characteristics of databases and the device of data entry. I also express the total digitization, informatization and application of archaeological data.

キーワード:文化財, データベース, 電子化, 情報化, 公開

Keywords:cultural properties, database, digitization, informatization, disclosure

1. はじめに

奈良文化財研究所(以下、奈文研)は長年に渡り、調査研究に基づいて文化財関係のデータベースを作成してきており、その多くを公開している。データベースのいくつかについて、内容を入力する時の工夫とデータベース構造の特色を紹介する。また合わせて、奈文研が取り組んでいる考古資料の総合的な情報化と活用についても述べる。

2. 公開データベース

奈文研がインターネット経由で公開しているデータベースには以下のものがある。

木簡関係のデータベース

- 木簡人名データベース
- 木簡画像データベース・木簡字典/電子くずし字典データベース連携検索
- 木簡データベース
- 木簡画像データベース[木簡字典]
- 全国木簡出土遺跡・報告書データベース

遺物関係のデータベース

- 軒瓦データベース
- 遺跡のデータベース
- 遺跡データベース
 - 地方官衙関係遺跡データベース
 - 古代寺院遺跡データベース

遺跡整備関係のデータベース

- 官衙関係遺跡整備データベース
- 遺跡の斜面保護データベース

庭園のデータベース

- 発掘庭園データベース
- Archaeologically Excavated Japanese Gardens
- Japanese Garden Dictionary

図書のデータベース

- 所蔵図書データベース
- 発掘調査報告書のデータベース
- 報告書抄録データベース
- 文書関係のデータベース

■薬師寺典籍文書データベース

■大宮家文書データベース

これらの中で、筆者が設計・入力・更新に直接関わっているデータベースのいくつかについて詳しく紹介する。他のデータベースについては、公開しているものなのでぜひ、実際にご利用いただき、ご意見をいただきたいと考えている。

<http://www.nabunken.go.jp>

遺跡データベース

遺跡データベースは、日本の遺跡全体を対象として構築を進めているデータベースである。

遺跡とひと口に言っても様々なものがあり、遺跡によっては、重層的な構造が見られる。例えば平城宮跡はひとつの遺跡としてとらえる場合もあれば、大極殿地区、朱雀門といった部分に分けて考えるべき場合もあろう。

遺跡データベースにおいては、情報の典拠となる文献の記述に従ってデータを入力している。よって、ある古墳群について「円墳5基」という情報しかなければ、古墳群としての情報しか入力できない。これに対し、群を構成している1基ずつの古墳それぞれについて情報が得られれば、個々の古墳ごとにレコードを作成して情報入力が可能である。

遺跡群とその構成要素である個々の遺跡といった重層的な構造をそのままデータベースの構成に持ち込むと、群と個別とが明確に分かれておらず、かつ十分な情報が得られないことが多い現状では、構成が複雑になりすぎる危険性がある。そこで、奈文研の遺跡データベースでは「種別」というフィールドを導入し、そこに集合・支群・個別・地区・調査といった重層構造中の想定位置を記述している。群としてとらえられる遺跡も単体としてとらえられる遺跡も同じようにフラットな構造の中で取り扱っている。

また、フィールドの設定において、その内容に関する情報が得られている遺跡数があまりに少ないものについては、当該フィールドを廃止して、「概要」の中に記述を移動した。例えば、「現状」に関する情報は「遺構概要」の中に「現状」という接頭語を付けて記述している。こうすることによって、検索結果の画面に空欄が並ぶという事態を回避している。

遺跡データベースには現在474,756レコードがあり、日々入力を続けている。典拠となる文献としては、都道府県単位の遺跡地名表・遺跡地図を優先して入力してきたが、最近では出版させているものをほぼ網羅し

たので、市町村単位の遺跡地名表・遺跡地図へも参照範囲を広げて入力を継続している。

また、遺物別の出土地名表の類も適宜参照してデータベースの充実を図っている。

発掘調査報告書抄録データベース

このデータベースは、遺跡の発掘調査報告書の内容に関するデータベースである。冊子体の報告書に添付されている抄録を電子化したものだが、抄録がない報告書についてもデータを入力するとともに、遡及入力を継続して行っている。

1レコードは、報告書内に記載されているひとつの遺跡に関する記述を単位としている。このため、1冊の本の中に複数の遺跡に関する記載があれば、複数レコードとして入力されている。

情報源は、都道府県教育委員会が、管下の市町村教育委員会を含めて取りまとめたデータと、全国埋蔵文化財法人連絡協議会が加盟機関のデータを取りまとめたものの提供を受け、奈文研が独自入力データを加えている。

外部から寄せられたデータについても、奈文研が用字の統一、用語の調整を行っている。これは検索によって確実に必要な情報が得られるようにするためである。

データ取りまとめの性質上、まとまった公開情報の更新は1年に1回であるが、1年分のデータの調整には10ヶ月ほどを要している。2011年度末には、1984年度と2010年度のデータが寄せられる予定となっており、カバーしている範囲が27年分となる。毎年、遡及入力により、日本で出版された遺跡調査報告書に関する情報の全てをカバーできる日もそれほど遠くないと期待されている。

3. 業務用データベース

奈文研内部で業務用として使用しているデータベースは3種に分けることができる。

1. 公開しているデータベースと同じデータベースで利用しているレコードも一緒のもの。
2. 公開しているデータベースであるが、公開していないフィールドや公開していないレコードを含んでいたもの。
3. 全く外部には公開していないデータベースである。

1に属するものとしては、遺跡データベースなどがある。2に属するものとしては、木簡データベースなどがある。木簡データベースにおいては、木簡に記載された本文がたいへん重要な情報であるが、古代の遺物

に記された文字なので、木簡が破損していたり、保存状態が良くなかったりして文字の読み取りが困難な場合が多い。そこで、複数の研究者による検討によって積文が確定したもののみが、公開されており、確定していない木簡については業務用データベース内でのみ利用可能となっている。

第3のカテゴリーに属するものとしては、写真データベース、航空写真データベース、図面画像データベースなどがある。またその他に、奈文研の所内には、研究者個人が作成しているもの、研究室内でのみ共有しているデータベースがある。

写真データベース

奈文研が撮影したフィルムのひとコマを1レコードとするデータベース。電子化されたものを中心としており、362,605レコード。データ中に、奈文研所蔵ではない遺物の写真などを含むため非公開となっている。

奈文研では撮影されたフィルムは登録され、恒温恒湿の保管庫で保管されており、劣化は少ないが、それでも自家処理していない一部のカラーフィルムに退色が認められ、早急な電子化が望まれる。また、8×10インチといった大判のフィルムやガラス乾板のように取り扱いが困難な原板も、通常の出版物での使用といった用途には電子化したものを利用の方が楽である。

使用頻度の高いカラーキャビネ版から、奈文研内部での電子化を開始し、次いでカラーブローニュ版、カラー35mmスライドと範囲を拡大して継続している。また、大判やガラス乾板については、設備の関係で外注して電子化を進めている。

航空写真データベース

奈文研が航空測量の会社から預かっている航空写真の撮影位置を中心とする情報のデータベース。地図上に撮影位置を表示できる。

奈文研が預かっている航空写真には次の特徴がある。昭和29年から昭和63年に撮影されたフィルムで、発注者に納品されておらず、撮影機関自身が保管しているフィルムを除いたもの。測量や地図作製のために専用のカメラで撮影された垂直写真に限る。

データベース化の手順は次の通りである。原板は1コマが23cm角のモノクロネガフィルムであり、視認性が低い。長いものでは50m近い長尺であるのに対し、短く切られたものもあり、そのままでは扱いが困難である。そこで、4500コマ程度をひとまとまりとして、マイクロフィルム化を行っている。データベースへの入力は専用入力ソフトを使用し、地形図の画像をパソコンの画面に出し、空中写真の撮影位置をクリックすることに

よって位置情報をデータベースに取り込む。入力者の左側でマイクロフィルムをリーダーに掛けて表示しながら、右側でパソコンを操作するという作業である。

空中写真には、どこを撮影したというフィルムかというメタデータが、評定図という形で添付されているのが通常であるが、添付されていないこともあり、必要な情報が欠けている場合も多い。評定図が不備な写真については、撮影の件名と写真に写っている地表面の情報のみから、どこを撮影したフィルムであるかを特定しなければならない。20万分の1地勢図や5万分の1地形図を参照しながら作業を行うが、空中写真撮影後の地形変化が著しい場合もあり、同定はたいへん困難な作業である。

このデータベースは、内容となっている航空写真が奈文研の所有品ではないために公開することができない。ただ、ある場所を撮影した写真が奈文研にあるかどうかという所在情報については、個別の照会に対応可能である。現在までに、1,277,315レコードが登録されており、航空写真に関するデータベースとしては日本最大である。

図面画像データベース

60年に及ぶ奈文研の調査で得られた実測図などの図面をスキャンした画像を検索するためのシステムで、本年度に利用開始したばかりである。

大判用紙に対応したスキャナの導入により、発掘調査時の遺構実測図から順に電子化を行っている。スキャンはカラー24bit、400dpiで行っている。スキャン後の作業として、余白の切り取り、傾きの補正を行うとともに、複数図面を一度にスキャンした場合には、個々の図面への切り分けを行って、基礎となる画像を作成する。次いで、この基礎画像を元にインデックス用の画像を作成している。

メタデータとしての文字データを作成し、画像データとのリンク情報をつけたものをデータベース化している。発掘調査の現場で作成された原図は、大判であることもあって、遺構を検討する際にも参照するのが困難であったが、このデータベースにより細部の確認などが研究者各自のパソコンから可能となった。

文化財情報研究室では、研究所内の個々の研究室内でのみ共有されているデータを、研究所全体で共有できるようにする取り組みを続けている。専門性の高いデータであっても、異分野の研究者が参照することによって新たな視点からの研究が進むことも期待される。

全所的な共有や対外的な公開に関しては、各研究室のみの努力では実現できないので、文化財情報研究室が助言やコンピュータ資源の提供を行っている。

4. 総合的な電子化と情報化

奈文研内部においては、文章の電子化、写真の電子化、図面画像の作成という方向で電子化を進めている。長い調査・研究の蓄積がある機関だけに、紙媒体・フィルム媒体で保管している資料は膨大な量にのぼっている。

これらの資料の高度な利用を図るために電子化を進めてきているが、新規発生分を除く既存部分の電子化が終了するにも、まだかなりの年月を必要とする。ただ、写真資料についても電子的なデータによる貸し出しが行われるなど、日常業務として利用できる水準にまで電子化が進んできている。膨大な資料数に臆することなく作業を日々進めることが大切であると考えている。

文字列による検索を行うデータベースを統合的に活用する道筋も導入している。それは、データベースの横断検索である。

例えば、「朱雀門」という語句を検索したい時に、どのデータベースを検索すれば、望んでいる情報が得られるかわからない場合がある。こういう時に、横断検索を活用すると、木簡データベースには〇〇件、〇〇データベースには〇〇件というヒット数が表示されるので、それ以降はそれぞれのデータベースに入って詳細な情報を得ることができる。

写真情報に関しては、近年最初からデジタルカメラで撮影されたデータ、いわゆるボーン・デジタルなデータが増加している。これらのデータは、現像・焼き付けという過程を経ない手軽さもあって、大量に作成される傾向がある。このため、予備的な整理を行ってメタデータを付与しないと、写っているものの内容が不明となってしまうやすい。撮影日時などのメタデータが自動で取得できるという利点を活かしつつ、必要な情報を加えてから、上記の写真データベースへと情報の移行を図るための、デジタル写真用のシステムを別建てで運用している。

図面の情報に関しては、現在発掘調査時の遺構実測図に限られている対象を、順次建築調査時の図面、遺物の実測図などに広げていく予定である。

次の課題として、画像情報としてのみ電子化している図に意味を付加する作業がある。これは図の構造や

構成要素をどのようにデータベース化するかということと深くかかわっている。

すなわち、掘立柱建物として認識している遺構が、複数の柱穴と溝によって構成されている場合、遺構を構成する遺構素として穴・溝を認識している訳で、そういった遺構構造を記述しなくてはならない。ひとつの柱穴を取ってみても、図として表す場合には、柱穴の上端線、下端線、遺構ケバといった図の構成要素があり、それらは単独で存在しているのではなく、相互につながりがある要素となっている。よって、図として表現する時に、柱穴位置の変更ということが、上端線と下端線の位置を個別に変更することで解決されるのではおかしいことになる。柱穴という実体としての把握がデータの構造にも必要である。

こういった遺構構造、遺構図構造を適切に情報化することによって、蓄積された膨大な情報をよりよく活用することが可能となると考えられる。先行している文字情報、画像情報の電子化と統合も必要で、この分野に関する研究をさらに進めることが大きな課題となっている。

デジタルデータの脆弱性が指摘されて久しい。電子化を進めることで、貴重な文化財を記録した、多くは唯一で代替物のない記録が失われてしまわないような配慮が必要である。遺物や美術工芸品であれば、複数回の写真撮影も可能ではあるが、遺跡の発掘調査は再現不可能な一回きりの作業の記録である。

記録内容に対する適切な到達可能性を確保した上でのデータ・マイグレーションを継続して行っていくことは、発掘調査の主体である地方公共団体単独の努力では困難な面が多いと考えられる。奈良文化財研究所が行っている、情報の電子化・情報化のノウハウを参考にしてもらうことによって、人類の宝とも言うべき文化財情報を安全かつ完全に次の世代に引き継ぎ、活用してもらえることを強く望む。