

## 個別 DB の深化と連携確保をめぐる

馬場基 渡辺晃宏 井上幸† 中川政樹†† 未代 誠仁†††

奈良文化財研究所では『木簡字典』『墨書土器字典』を公開している。その開発では、個別の資料の特性をデータベースに反映させることの重要性を再認識した。一方、東京大学史料編纂所『電子くずし字字典』との連携検索システムも公開している。この開発では、「一般化」の必要性を強く認識した。本報告では、上記の事例を具体的に紹介し、「個別深化」「一般化」という二つの課題解決への提案を行う。

### About enabling enhancement and cooperation of individual DB

HAJIME BABA AKIHIRO WATANABE MIYUKI INOUE††  
MASAKI NAKAGAWA†† AKIHITO KITADAI†††

In Nara National Research Institute for Cultural Properties, "Mokkan Jiten" and "Bokushodoki Jiten" are exhibited. In the development of them, we have confirmed the importance of reflecting the characteristic of each data on a database. The cooperation search system with "Denshi Kuzusiji Jiten (Historiographical Institute The University of Tokyo)" is also exhibited. In this development, the necessity for "generalization" has been recognized strongly. By this report, we introduce the above-mentioned example concretely and propose the solution of two subjects of "individual enhancement" and "generalization."

#### 1. はじめに

本報告では、独立行政法人国立文化財機構奈良文化財研究所（以下奈文研と略称）で公開している『木簡画像データベース・木簡字典（以下『木簡字典』と略称）』、『墨書土器画像データベース・墨書土器字典（以下『墨書土器字典』と略称）』、および東京大学史料編纂所との連携検索システムである『木簡画像データベース・木簡字典』『電子くずし字字典データベース』連携検索システム（以下「連携検索システム」と略称）について、その特長と課題について述べる。そして、より高次の体系構築と連携確保に向けた議論の素材を提供したい。

#### 2. 機関の性格と DB の方向性

現在、様々な人文系データベースが公開されており、それぞれの性格・特徴も多様であると思われる。『木簡字典』も独自の性格を有するが、その背景には奈文研の機関としての性格や業務の特性が存在すると思われる。

##### (1) 研究拠点機能とデータベースの公共性

奈文研は、文化財の総合的研究を通じ、文化財保護行政に資することを目的として、多くの文化財が所在する奈良に設置された研究所である。現在は、独立行政法人国立文化財機構の一員として、文化財研究・保護のナショナルセンターの一翼を担い、平城宮・京、藤原宮、飛鳥地域などの発掘調査や研究、南都古社寺の古文書・建築などの総合

的調査と研究、文化財の自然科学的分析や保存方法の開発と実践、地方公共団体への支援・助言、国際共同研究や国際支援・協力など、多岐に渉る活動を行っている。

こうした性格から、奈文研のデータベースは、データベースそれ自体が「研究成果」であるという側面も皆無ではないが、「研究インフラ」整備の一環という性格がより強い。

研究拠点という性格と関連して、奈文研でのデータベース公開の大きな目的の一つとして、研究拠点（ナショナルセンター）としての機能を果たす手段の一つという側面がある。全国の情報を集約してデータベース化し、それを提供することで研究者や地方公共団体の業務を支援するという方向性である。

##### (2) 調査機関としてのデータベース構築

奈文研は、実際にフィールドで活動する調査機関でもある。調査成果は、毎年発行される『奈良文化財研究所紀要』や『奈良文化財研究所学報』『奈良文化財研究所史料』等の刊行物の他、記者発表・現地説明会等を通じて公表する。また、平城宮跡資料館等の公開施設でも文化財の公開を行っている。

奈文研のデータベースには、こうした文化財の公表・公開の一部という側面も存在する。研究者向けの研究資源としてはもちろん、広く国民への公開という役割も担っている。文化財保護法に明記されるように、「文化財は国民共有の財産」である。国民共有の財産たる文化財を、国民全体により広く公開することも重要な課題である。

また、奈文研が保管する文化財の管理や、奈文研が行う調査の為にツールという役割も存在する。公開しているデータベース以外にも、発掘調査で出土した遺物の管理台帳や画像のデータベース化にとどまらず、文化財調査の調書

† 独立行政法人国立文化財機構 奈良文化財研究所  
Nara National Research Institute for Cultural Properties

†† 東京農工大学  
Tokyo University of Agriculture and Technology

††† 桜美林大学  
J. F. Oberlin University

を直接デジタルデータとして入力する事例も存在する。文化財写真のデジタル化・データベース化も進められている。また、公開データベースも含め、研究所が抱えている喫緊の研究課題に必要なデータの入力を優先したりすることもある。

以上の様に、奈文研のデータベースは奈文研の設置目的や業務に関わって、いくつかの側面を併せ持ちながら構築されている。

### 3. 『木簡字典』の考え方

データベースの性格・目的は、特に意識する必要がない場合も多いと思われる。だが、各データベースの根幹となる役割を明確にしないで開発を行うと、中途半端なシステムに仕上がる可能性も高いと考えられる。

#### (1) 『木簡字典』開発の経緯

まず、「木簡」について簡単に説明しておく。木簡とは、墨書された木製品である。文字を書く目的で製作された場合も、木製品に結果的に墨書がなされた場合も含む。伝世品も存在するが、多くは発掘調査によって遺跡から出土する。出土点数は全国で35万点ほど、うち25万点弱を奈文研で保管している。

遺跡から出土する文字資料であり、遺跡の理解に大きな役割を果たす。特に古代史では、正倉院文書とならんで貴重なナマの資料・一次資料である。重要な資料である反面、脆弱である。また劣化等によって墨痕が不明瞭だったり、破損等によって、墨跡が不完全な事例は非常に多い。したがって、釈読（解読）や保存・保管には困難がともなう。また、時期や地域・内容による字形の変化、書写媒体の制約による省画なども多々みられる。破損により、文脈が途切れている場合もある。

木簡は、こうした特性があるため釈読作業を助ける「字書」の必要性が従来より認識されてきた。焼き付け写真を用いた字書作成の準備作業も行われたことがある。だが、検索が簡便では無く、また写真の鮮明さを確保することも困難であるなど、多くの問題点があり、作業途中で中止されてしまった。

2003年度に科学研究費補助金基盤研究S「推論機能を有する木簡など出土文字資料の文字自動認識システムの開発」（研究代表者：渡辺晃宏）の交付を受けて、その研究の一環としてデータベースによる字書開発に着手し、その後2008年度からはおなじく科研費基盤S「木簡など出土文字資料釈読支援システムの高次化と総合的研究拠点データベースの構築」（研究代表者：渡辺晃宏）によってさらに改良を加えている。

『木簡字典』の最大の目的は、「木簡の文字による「字書」データベース」の構築である。字書作成の目的は、木簡釈読作業を補助することである。

木簡全体の情報や、木簡の管理のためのデータベースと

しては、すでに「木簡データベース」が整備されており、『木簡字典』が担う役割ではない。また、木簡の「研究成果」の反映も「字書」の範疇を超えるものであり、『木簡字典』の役割ではない。

#### (2) 『木簡字典』の方向性

『木簡字典』は字書であるが、木簡の特性に合わせた「見せ方」や「引き方」が必要になる。上述の様に、木簡は、文字が不鮮明な場合が多く、破損していることも多い。文脈とは無関係に文字がとぎれる場合もある。文字が書かれる地域や場面、時代による変化もある。

そこで、見えにくい墨痕に対応するため、文字画像の種類を複数にした。出土文字資料の画像としては、白黒写真・カラー写真の他に、赤外線データ・記帳データが存在する。

赤外線データは、赤外領域で墨部の反応が特に際立つことを利用したものである。赤外線領域を利用した観察方法は出土文字資料観察で広く用いられているが、近年こうした画像を高解像度で記録することができるようになり、データの蓄積がはじまっている。

記帳データは、研究員による記帳ノートのデータである。奈良文化財研究所では、出土文字資料を読む場合、「記帳」と呼ばれる記録作業を行う。資料のアウトライン・形状・破損状況等と、墨の状況を記録する。この際、筆をどのように動かしたのか観察し、を記録する点が、通常の実測とは異なる。したがって、記帳作業は文字を読み取る最も重要な基礎的作業であり、その記録である「記帳ノート」は文字釈読のノウハウそのものといえることができる。

これらのデータを総合的に提供することで、不鮮明な文字を様々に観察することを可能にしようと考えたのである。

次に、複数文字による検索についても、現実には木簡上の文字が単語や文脈単位で残存しているとは限らない。たとえば、若狭国遠敷郡のうち、「国遠」だけが読み取れている場合もある。そこで、熟語で登録するのではなく、ユーザが指定した文字列を任意に表示できる方法を開発した。

地域差や時代性などに対応するために、メタ情報をしっかりと付加しておくことによって、その文字の地域的・時代的特性などによる絞り込みを可能にした。

一方、webによる公開を通じて、『木簡字典』を奈文研所員だけが利用する字書ツールという位置づけでは無く、木簡など出土文字資料に直面した各地の地方公共団体等が利用できる支援ツール、さらには木簡の文字を検討・研究しようとする研究者のための研究資源、さらには文化財に関心を持つ国民全般が木簡の文字に親しむためのツールとしての機能を付与する、という方向性も付与した。このため、検索画面の構成などにも工夫をおこなった。

#### (3) 『木簡字典』の波及効果

『木簡字典』の開発は、いくつかの波及効果をもたらした。一つには、『木簡字典』システムを流用した他の字書データベースの開発である。現在開発が終了し、公開に至っ

ているのは『墨書土器字典』である。メタデータを木簡から墨書土器に相応しいものに変更し、一部画像データの持ち方などを調整して開発した。『木簡字典』の枠組みは「字書」であり、字書データベースとしての汎用性の高さを示しているといえよう。

また、複数文字検索実現のためのデータの持ち方の「副産物」として、『木簡字典』は文字それぞれの位置情報が格納されている。この位置情報を利用することで木簡上の文字や語句の記載パターンを抽出する研究を行った。さらに、この研究過程でタグ付きデータを『木簡字典』に流用することによって、「文字」からだけではなく「意味」から字書を引ける可能性も検討中である。

さらに、上記の研究や『木簡字典』等の開発の経験や開発を通じて得たノウハウのもとに、木簡調査情報の取得・蓄積方法自体から大きく見直す方向性も見えてきた。これまで、各種データベース間で、相互に情報の融通や共有は進めてきた。基幹となる「木簡データベース」、字典である『木簡字典』、人名に特化した「木簡人名データベース」、補助データベースで非公開の「木簡出土遺構年代観データベース」などである。これらの構築方法を抜本的に切り替え、木簡情報を集約したデータから、それぞれのデータベースに必要なデータを「抜き出す」というようなあり方を模索するに至っている。

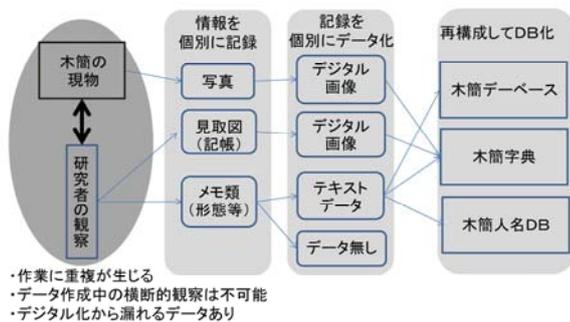


図 1 旧来のデータベース作成手順

Figure 1 The conventional database creation procedure

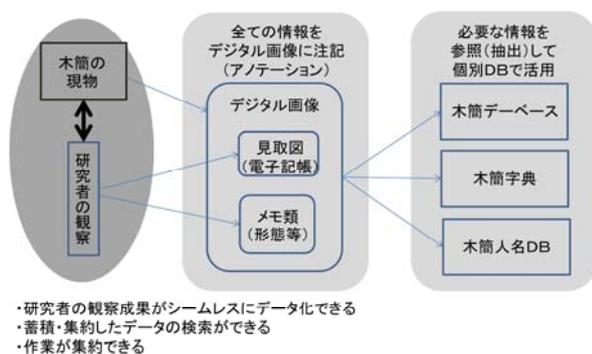


図 2 新しいデータベース作成手順

Figure 1 The new database creation procedure

#### 4. 連携に向けた課題

『木簡字典』は『電子くずし字字典』との連携検索システムの開発・公開に成功している。機関間のデータベース連携の成功事例といえよう。

##### (1) 『電子くずし字字典』との連携

『木簡字典』開発着手時から、東京大学史料編纂所でも『電子くずし字字典』を開発しているという情報を得、相互に開発状況等も含めて情報交換を行い、2009年には連携検索システムの開発・公開に至った。

この過程では、制度や運用、費用負担から現実の開発まで多様な交渉とともに、様々なデータベース開発を巡るノウハウの交換、議論が行われた。「機関同士の継続的で対等な協力関係」という大前提を確立し、この前提に基づいて費用負担の等分化、相互のデータベースへの不干渉などが決定されていった。連携検索は、非常に歓迎され、現在でも多くのアクセスを得ている。

この連携では、双方のデータベースの持つ様々な特性をあえて捨象して、「字書」として文字を引くという一点に絞り込んで開発を進めた。「電子くずし字字典」は熟語を登録している他、漢字辞書機能を持ち、部首から文字を選択できる機能がある。これらは木簡字典にはない機能である。一方、文字の位置情報は『木簡字典』側にも存在する。これらの方向性からの検索は今後の課題として留保し、一字を検索するという機能に絞り込んで開発を進めた。

結論的にいえば、この成功の背景には担当者相互から所長レベルに至るまでの各段階における密接な打ち合わせ・相互理解と信頼関係が構築されて可能になったものであり、単に「似たデータベースがある」というだけでは連携の成功は難しい。

研究者個人レベルでは、データベースの継続的・安定的な運用に不安が残り、機関同士となると相互の機関の利害なども存在する。データベース連携の確立には、技術的問題以前に、相互の信頼関係に及ぶ密接な関係をどのように構築するか、という課題が存在すると言わざるを得ない。

##### (1) 連携拡大に向けた課題

もちろん、技術的課題も多く存在する。特に、「どういう情報を共通して、どういう型式でもつか」という点について、今後議論を深めていく必要がある。

ざっと思いつくものを挙げる。文字そのものについては、文字コードにない文字をどのように扱うか、異体字をどのように扱うか、というような問題を整理しておく必要がある。文字に付与するメタデータを巡る問題は多く、どの範囲までを共有するのかというようなガイドラインの作成が必要では無いかと思う。たとえば時代・地域についての情報を、ある精度でいれるか、というようなルールは作成の必要があるように思われる。

文字画像の形式などについては、実際連携が決まった段

階から再作成しても十分間に合うように思われるが、データの中身や構成に関わる部分は、データベース構築の早い段階から共通のルールにのせる方が都合がよいように感じるがいかがであろうか。

## 5. おわりに

データベース開発を始めると、つい自分の興味や関心、研究の方向に合わせて凝ってしまう。だが、連携強化にせよ、ユーザを向いた研究インフラとしての整備にせよ、個別データベースが「特殊化」に突き進んでしまっては果たせない課題であろう。

データベースを現物からの距離感や役割に応じて作り分けること、それぞれのデータベース間でうまくデータを共有することが、重要であると考え。こうした中に、他機関のデータベースとデータを共有したり、連携していくためのしなやかさを潜ませていくと、飛躍的な研究の発展につながるであろう。