

遺跡抄録情報の現状と可能性

はじめに 本稿では、報告書に遺跡抄録が付加されるまでの経緯と、これまで分散していたデータベースを統合し、登録作業のフローや入力・検索方法の変更点と今後の課題について報告する。また、変更にともなう遺跡抄録の登録作業における注意点と、今後の可能性についても述べる。

遺跡抄録の経緯 発掘調査報告書（以下、報告書）を採す際、まず手掛かりになるのは書名である。しかし報告書の書名はたいいてい遺跡名のみを表し、遺跡の年代や、出土物の内容まで把握できない。利用者にとって研究資料としての必要可否を判断するためには、当然ながら資料に隅々まで目を通す必要がある。よって網羅的な類別調査を行うためには、膨大な時間が必要であった。報告書の発行部数は限られており、遠隔地の報告書を取り寄せ、確認する作業も容易ではない。そこで、調査の要約を記載した遺跡抄録（以下、抄録）を報告書に付加し、大意を抄録で把握することが試みられる。そして、文化庁「埋蔵文化財発掘調査報告書の抄録の作成について」（文化庁記念物課6保記第16号、1994年4月27日付け）によって、報告書は抄録に報告書誌、発掘概要、遺構遺物等の概要が記載されることとなった。

データ統合と入力方式変更 2003年からは抄録のデータベース化が開始された（「報告書データベースの作成について」（文化庁15第財記念第67号、2003年12月16日付け））。まず抄録データベースで抄録記載内容を検索し、さらに詳細を知る必要があれば、当該報告書を取り寄せる。当時の抄録データの集約方法は、年1回奈文研に都道府県が各基礎自治体からとりまとめた複数のExcelを1つに統合し、CDで送付していた。一方、全国埋蔵文化財法人連絡協議会（以下、全埋協）加盟法人が作成した報告書の抄録データについては、上記とは別経路で、同様に年1回程度、データを奈文研に送付する方式がとられていた。（図59）この時、①抄録データベースへExcelのデータを手入力する作業を挟むことでヒューマンエラーの確率が高まる、②自由記述という性質上時代や種別表記が揃わない、③奈文研でのデータ登録に時間がかかり即時反映できない、などの課題があったため、全埋協では

早期に独自のデータベースを構築し、調査機関によるWEB入力を実現していた。2015年から奈文研でも全国遺跡報告総覧（以下、遺跡総覧）を運営していたため、一部の自治体や法人調査組織では、抄録のExcel送付と遺跡総覧へのデータ抄録入力作業が重複し、非効率であった。そのため、2019年6月には抄録登録フローを変更し、奈文研抄録データベース、全埋協抄録データベースについては遺跡総覧へすべて統合完了し、新規登録分はWEB入力化された（「報告書データベースの作成について」（文化庁文化財第二課長／奈良文化財研究所所長、事務連絡、平成31年3月15日付け））（図60）。これにより、データの即時反映が可能となったほか、システム上でも自動入力チェック機能を実装し、データ登録時に不正なデータ登録を防ぎ、ヒューマンエラーの発生を抑制することでそれぞれの問題についてはおおむね解決出来たと言える。また、これまで各データベースで検索する必要があったデータを一括で調べられるようになった。2020年4月2日時点で、抄録は遺跡総覧に130,024件登録されている。（図61）

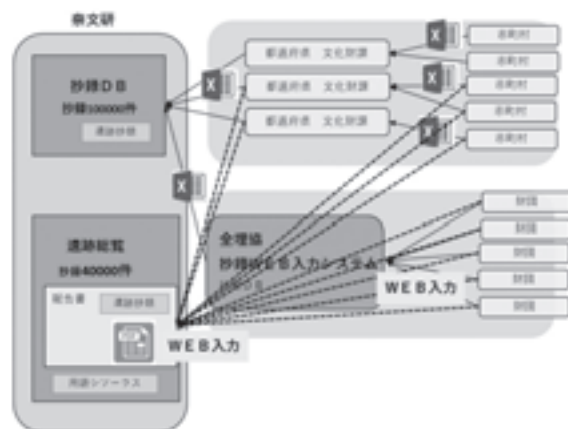


図59 2019年6月までの抄録登録フロー

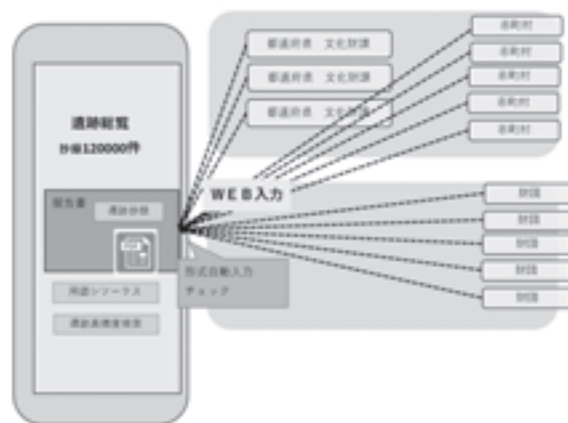


図60 2019年6月以降の抄録登録フロー

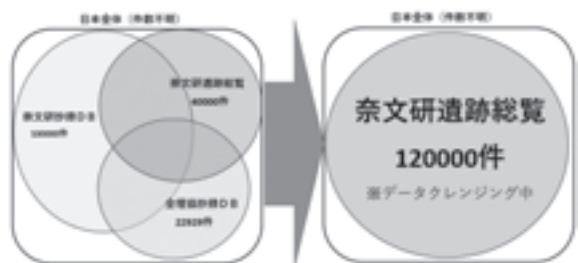


図61 各抄録データの保持状況ベン図

検索方法 データ統合の際、遺跡の種類・時代・遺跡所在地の市町村コードをすべてデータクレンジングした。時代の表記は、たとえば「旧石器～弥生」という表記の場合、これまでの奈文研抄録DBでは、「縄文」という表記がないため、縄文時代を検索することはできなかった。しかし、遺跡総覧の抄録検索では、「旧石器～弥生」を時間軸の意味として通じるように整理し、旧石器・縄文・弥生のそれぞれで検索できるようにした。しかし、「縄文後期後葉」のような詳細表記は非対応である。日本全体でデータ標準化の観点から整合性を個別に取ることは困難であると考えられる単語についても、詳細表記などは、現状では特記事項にテキストで記載し、テキスト検索で検索対象とするほかないものの、類義語を含めた検索については、専門用語の使い方は専門家の認識や研究史に基づくものの、研究成果を社会に普及する観点からは検索性を確保する必要がある。2020年2月から遺跡総覧では用語の類義関係を整理し、内部にソーラスを構築することで、類義語も含めて検索できる機能を追加した。

遺跡抄録の重複登録 遺跡総覧へのWEB入力による登録における注意点としては、データの重複登録である。報告書のデータとして、書誌・抄録・PDFの3つがあるが、そのうち、すでにいずれかのデータが登録されている可能性がある。もし既存の登録データがある場合、不足データのみを登録すればよい。登録の詳細は、遺跡総覧の参加・登録手続－各種登録マニュアル・参考資料を確認願いたい。

(<https://sitereports.nabunken.go.jp/ja/abouts/participation>)。

遺跡位置情報の注意点 遺跡総覧に登録された既存の抄録データには、いくつかの原因によって位置情報が誤っているケースがある。間違いのパターンには、一つの項目に複数の位置が入っている・桁不足・緯度経度が

逆転している・文字列が混入・60進法で記入等がある。誤りの大半は、遺跡位置を地図上に展開しなければ、気づく可能性は低い。遺跡総覧の内部機能として、遺跡総覧直接参加機関は、登録されている遺跡位置を日本地図に展開し閲覧することができる。各機関での抄録データの整備作業が進めば、将来的にユーザーへの公開・研究データとしての活用を視野に入れることもできるだろう。PDFのテキスト・画像データだけでは位置・時代の情報を取得できないため、抄録データの整備は今後ますます重要となる。

遺跡位置情報の取得と測地系の変換方法 抄録の遺跡位置は、遺跡のほぼ中心の位置を度分秒の単位で記入することとなっている（文化庁2004）。遺跡位置情報の取得は、国土地理院の地理院地図が便利である。測量法で規定されている「測量の基準」は、現在日本測地系から世界測地系に改正されている。そのため、現在日本測地系では測量してはならない。過去の日本測地系の成果を活用するために、国土地理院では経緯度変換をおこなうプログラムTKY2JGDを無償で提供している。TKY2JGDでは、日本測地系から世界測地系への変換、十進法度から度分秒への変換などを簡単な操作でおこなうことができる。(<https://vldb.gsi.go.jp/sokuchi/surveycalc/ky2jgd/main.html>)。

おわりに 現在、抄録130,024件を都道府県・遺跡種別・時代・出土遺構・遺物で簡単に検索できる。書誌や本文PDFでは実現できない検索方法であるため、抄録は重要である。今後遺跡位置情報を整備できれば、地理空間的に把握が可能となり、学術的な研究基盤としての可能性はさらに高まろう。

(国武貞克・高田祐一・村上 蛭／技術補佐員)

参考文献

文化庁『行政目的で行う埋蔵文化財の調査についての標準（報告）』文化庁埋蔵文化財発掘調査体制等の整備充実に関する調査研究委員会、2004。

なぶんけんブログ「全国遺跡報告総覧：発掘調査報告書総目録 大阪府・兵庫県・鳥根県・高知県編の書誌情報を全国遺跡報告総覧に登録」<https://www.nabunken.go.jp/nabunkenblog/2020/01/somokuroku202001.html>（2020年3月20日確認）。

なぶんけんブログ「全国遺跡報告総覧：類義語およびOCR誤認識用語検索機能の公開」<https://www.nabunken.go.jp/nabunkenblog/2020/02/ruigigo.html>（2020年3月15日確認）。