発掘調査報告書から多言語対応収蔵品データベースへ

Yanase Peter • 奈良文化財研究所

From Fieldwork Report to Multilingual Database

Yanase Peter • Nara National Research Institute for Cultural Properties

キーワード／英訳　考古学／Archaeology　データベース／Databases

奈良文化財研究所（以下、奈文研）は2020年度に多言語化整備事業の一端として平城宮跡資料館（以下、資料館）の多言語化を掲げた。資料館の多言語化の目的は、展示室内における各種説明を英・中・韓の三言語で提供し、日本語ができない、または得意としない来館者が資料館を楽しめる環境を整えることと、多言語化によるWeb上の情報発信の強化であった。そのため、展示室内上映ビデオの吹き替え、言語別の木簡に関する解説シート、展示パネルの訳文、Web閲覧用データベース、スマホアプリ用のデータなどを作成した。その中から、今回は試行錯誤を重ねながら半年かけて一から作った多言語対応収蔵品データベースに焦点を当てる。

元データの問題点

今回作成したデータベースの作成過程は極めて特殊であった。なぜなら、今回は研究者と学芸員が事前に準備したテキストの翻訳を行ったのではなく、多言語データのために過去の日本語データを掘り起こして英・中・韓の三言語のデータとして再構築した。そうした理由をまず説明する。

資料館にはもともと公開データベースがなく、あるのは収蔵品管理のために作成された簡易なデータベースのみだった。内部向けデータベースなので、保管場所、貸出履歴、関連文献（報告書、図録）や関連写真の番号などの実務的な情報を中心に入力されていた。

資料館は奈文研の施設であり、奈文研の平城宮跡に関わる研究成果を中心に展示・解説している。奈文研は巨大な組織であり、遺物の種類によって扱う研究室が異なるだけでなく、各研究室で独自のデータ方式があり、登録されている情報にもばらつきがある。筆者は過去に各部屋のデータの項目を比較する機会があっ

た。同じ組織でありながら、統一的なフォーマットが未整備であることに驚いた記憶がある。今回問題となった資料館の展示品・収蔵品データベースは管理のための独自フォーマットであった。この情報をそのまま翻訳して外部に公開しても、それは来館者・Web閲覧者が求めている情報とマッチしないと判断した。そのため、多言語対応収蔵品のデータベースのために新たにデータを整理し、一からデータを作成することにした。

必要な項目の整理

今回新たに作成する多言語データの用途として二つが想定された。一つは、Web 上での閲覧、もう一つは展示室内観覧中のスマホアプリでの閲覧であった。採用しているシステム上ではそれぞれに別々のデータを表示することも可能であったが、今回はデータ作成の効率化を考えて二つとも同じ内容にした。そのため、データベースで表示する情報をこの二つのユースケースに合わせる必要があった。

資料館の多言語化においてスマホアプリを導入した最大の理由は、展示室内のスペースが限られていて、すべての情報をアナログ型式で四言語で表示するのは困難だと判断されたことである。とはいえ、資料館を楽しむことをスマホ持ち込み前提にするのは避けたかった。そのため、「スマホがあってもなくても、展示は理解できるが、スマホがあった方が楽しい」という発想の下で情報を整理することにした。これにより、展示室内のパネルを全面的に四言語対応するようにした一方、展示品のキャプション・題箋はタイトルだけを四言語で表示し、詳細な説明をスマホで表示することにした。他方、データをパソコン上で閲覧するユーザーには展示室内の説明が見えていないため、収蔵品・展示品一個一個の情報が独立するように情報を整理する必要があった。また、Web 上で提供するデータベースはデータセットとしても使用できるような形を取ることを視野に入れていた。

結果として、「名称」、「材質」、「寸法」、「時代」、「解説文」という基本情報に加えて、考古学の特有の情報として「出土場所」、またデータ連携のため、「関連発掘調査報告書」の項目を設定した。さらに、パソコン・スマホの特徴を生かし、展示室では見られない画像を多数掲載することにした。

情報収集

データベースの必要な項目が決まったら、次にそれらの項目を埋めるのに必要な情報を探し出す作業にかかった。真っ先に整理したのは関連発掘調査報告書の項目だった。ほかの項目は一番信頼にあたる情報源、つまり報告書から抽出することを決めていたので、まずそれぞれの遺物の詳細が記述されている報告書を特定する必要があった。幸いに、この情報は元データの中に含まれていた。

次にそれぞれの報告書から基本情報を探し出した。ほとんどの遺物に関しては報告書から寸法、材質などの記述をそのまま採用できた。ただ、報告書の解説文は専門家向けなので、細かくて難しい文章が多い。そのため、報告書の解説文だけをそのまま採用することは困難であった。

今回のプロジェクトではスマホのコンテンツを充実させるために収蔵品の中から100件に対して解説文を付けることにしたので、短時間で新たに100件の解説文を、しかも三言語で作る必要があった。この作業の効率化を図るために、すべての解説文を同じ構造に統一することにした。つまり、すべての解説文を「これは何？」、「どのように使われていた？」「他に何か面白いことはあるか？」の三つの質問に対して、1文ずつで答える構造に統一した。これを決めたのち、報告書と過去の図録からそれぞれの質問に対する答えをピンポイントで探し出した。そしてその情報をベースに英・中・韓の三言語でそれぞれのオーディエンスに合わせて文章を作成した。

文字データを作成した後、それを補う画像データを集めた。報告書やその他の過去の刊行物ではそれぞれの遺物に関連する画像が大量に含まれているので、ここでも解説文の文章の作成と同様に、前もって掲載したい画像の種類を整理した。今回は、展示品（多くの場合レプリカ）の表と裏、実物の記録写真（表と裏）、実測図、出土場所、出土場所の実測図、出土地区（遠景）、出土地区の実測図・地図を対象とした。資料館に展示されている実物とレプリカの撮影は展示企画室と写真室が行った。その他の画像は報告書から切り取った。

全国遺跡報告総覧との連携

発掘調査で発見された遺物の最大の特徴は、遺物単独でほとんど意味を持たないことと言える。そのため、遺物・展示品一つずつの説明をいくら充実させても、不完全になってしまう。この問題を解決するため、今回作成したデータベースに

はそれぞれの遺物が掲載された発掘調査報告書 PDF へのリンクを張った。また、PDF が公開されていなかった関連報告書を新たにデジタル化して登録した。これによってユーザーが一つの遺物から簡単に発掘調査全体の詳細な説明にジャンプできるようになり、遺物が置かれたコンテキストについて学ぶことができるようになった。

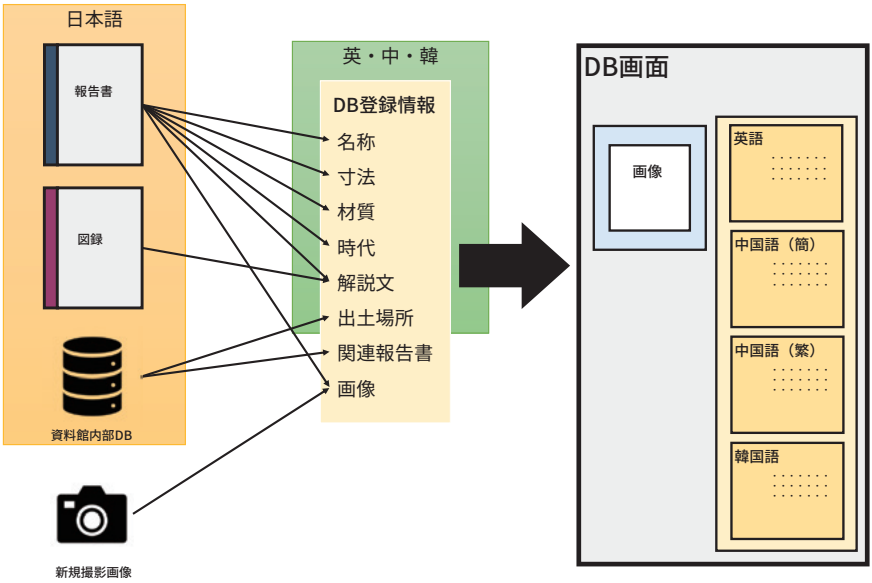

図1　多言語データ作成のワークフロー

結びに

奈文研の 2020 年度の多言語化事業の中で多言語対応収蔵データベースの作成はとりわけ大きなプロジェクトであった。その過程で最大の問題となったのは、データベース用の日本語データが整備されていなかったことであった。このような状況の中で、データベースの作成を実現できたのは、日本語テキストの作成ステップを経由せずに、直接遺跡の記録保存としての機能を持つ発掘調査報告書に基づいて多言語データを作成したからである。このような試みは奈文研としても初めてであったが、結果として発掘調査報告書の新たな活用方法が切りひらかれた。この手法は発掘調査報告書を刊行した実績のある自治体であれば今すぐ着手

することが可能であり、多言語化のコストダウンと作業時間短縮につながることが期待できる。

Ⅳ

資料