

# 発掘調査報告書全文データベース「全国遺跡報告総覧」の開発 —遺跡情報のプラットフォームを目指して—

高田 祐一・森本 晋 (奈良文化財研究所)

田中 俊二・昌子 喜信・福山 栄作・矢田 貴史 (島根大学)

永島 幹大 (ENU Technologies)

遺跡の発掘調査報告書は、貴重な学術資料でありながら、流通範囲が限られ一般に利用しづらい。いわゆる「灰色文献」となっている。必要とする人が誰でも手軽に調査・研究や教育に利用できるよう奈良文化財研究所では、発掘調査報告書の全文データベース「全国遺跡報告総覧」を公開した。これまでの開発の経緯、システムの概要、開発による効果などを報告する。

## Construction of the Comprehensive database of archaeological site reports in japan

Yuichi Takata/Susumu Morimoto(Nara National Research Institute for Cultural Properties)  
Shunji Tanaka/Yoshinobu Shoji/Eisaku Fukuyama/Takafumi Yada(Shimane University)  
Yorihiro Nagasima(ENU Technologies)

The excavation report of the archaeological sites is hard to use because its distribution range is limited while it is a precious academic document. It becomes so-called "gray literature". Nara National Research Institute for Cultural Properties has released a full text database of the excavation report, "Japanese Archaeological Sites Report Conspectus". I report the process of the past development, the summary of the system and an effect by the development.

### 1. まえがき

2015年6月25日、奈良文化財研究所（以下、奈文研）では、発掘調査報告書の全文データベース「全国遺跡報告総覧」（以下、遺跡総覧）（<http://sitereports.nabunken.go.jp>）を公開した[図1]。これまでの開発の経緯、システムの概要、開発による効果、実運用に関する課題や解決方法などを報告する。

### 2. 遺跡の発掘調査報告書をめぐる状況

近年、日本全国で毎年約8000件の発掘調査が実施され、約2000冊の発掘調査報告書が発行される(1)[図2]。土木工事等の開発事業によって、やむをえず遺跡を現状保存できない場合に発掘調査を実施し、記録保存調査を実施する。また遺跡の保存目的で調査を実施することもある。その調査結果は発掘調査報告書にまとめられ、紙媒体として印刷される。考古学では発掘調査報告書を1次資料に近い扱いをしている。文化庁の『発掘調査のてびき』では「報告書は、発掘調査全般の成果を的確にまとめた埋蔵文化財の記録であることから、将来にわたって保存されるとともに、広く公開されて、国民が共有し、活用できるよう

な措置を講じる必要がある」とされる(2)。

発掘調査報告書は、既に膨大な数を発行されている。現在までの正確な報告書数は把握できておらず、推定20万冊といわれる。報告書は商業出版ではない上、発行部数は300部程度であり、流



図-01 全国遺跡報告総覧トップページ

通範囲に限られる。いわゆる「灰色文献」化している。また図書館書庫が飽和しつつある状況で、利用頻度が低い他地域の報告書を収蔵する余地がないケースが存在する。ある大学付属図書館では、図書の収容量が限界に近づいたため、利用頻度に応じ、処分候補を抽出した。その候補には発掘調査報告書が含まれていたという(3)。

また文化財政策の着実な推進によって発掘調査報告書が発行され、貴重な学術成果が蓄積されていく一方、膨大な情報量に阻まれ、報告書の全体像を把握し個々の報告書に適切にアクセスすることは困難になりつつある。



図-02 平城宮跡第224次発掘調査風景

表-01 連携大学および直接参加自治体 (2015年8月時点)

参加府県及び連携大学	参加自治体(自治体直接参加モデル)
宮城県(東北大学)	北海道 厚沢部町
山形県(山形大学)	北海道 奥尻町
秋田県(秋田大学)	青森県 青森市
茨城県(筑波大学)	岩手県 岩手県埋蔵文化財センター
富山県(富山大学)	埼玉県 熊谷市
新潟県(信州大学)	埼玉県 深谷市
長野県(信州大学)	石川県 野々市市
山梨県(信州大学)	岐阜県 岐阜県文化財保護センター
滋賀県(滋賀大学)	岐阜県 飛騨市
大阪府(大阪大学)	岐阜県 下呂市
兵庫県(神戸大学)	岐阜県 海津市
奈良県(奈良女子大学)	静岡県 静岡県埋蔵文化財センター
鳥取県(鳥取大学)	静岡県 浜松市
島根県(島根大学)	静岡県 熱海市
岡山県(岡山大学)	静岡県 磐田市
広島県(広島大学)	静岡県 伊豆の国市
山口県(山口大学)	愛知県 愛知県埋蔵文化財センター
徳島県(徳島大学)	愛知県 常滑市
香川県(香川大学)	愛知県 東海市
愛媛県(愛媛大学)	福岡県 筑後市
高知県(高知大学)	長崎県 長崎県埋蔵文化財センター
福岡県(九州大学)	熊本県 御船町
宮崎県(宮崎大学)	鹿児島県 出水市

### 3. 「全国遺跡報告総覧」開発の経緯

#### 3.1 全国遺跡資料リポジトリ・プロジェクト

国立情報学研究所の最先端学術情報基盤(CSI)整備事業の委託を受けて、2008(平成20)年度~2012(平成24)年度の5年間にわたって、全国の21の国立大学が連携して「全国遺跡資料リポジトリ・プロジェクト」(以下、遺跡リポジトリ)[事務局:島根大学]を推進した。遺跡リポジトリでは、発掘調査報告書を電子化・公開することで、報告書の流通と利活用の促進を目的としていた。連携大学図書館を中心に、自治体文化財担当部署、関係機関との連携・協力により運営し、約1万4000冊、150万ページを越える報告書を電子化して公開した。年間約50万件の報告書データのダウンロードがあり、活発に活用され大きな成果をあげた。

しかし、CSI事業の終了やサーバの老朽化など、プロジェクトの継続に課題があった。そこで、奈文研では、当初より発掘調査報告書のメタデータを提供していたことや遺跡リポジトリで共同研究してきた経緯もあって、各大学の遺跡リポジトリシステムと報告書の電子データを奈文研にて統合し、移管することとした。

#### 3.2 奈文研へのシステム・データ統合

2014年11月、大阪大学付属図書館にて、全国の連携大学の遺跡リポジトリ担当者が集い、「平成26年度実務者連絡・調整会議」(以下、実務者連絡会議)を開催した[表1]。実務者連絡会議では、質疑応答の後、奈文研へのシステム・データ統合、今後の方針などが承認された。

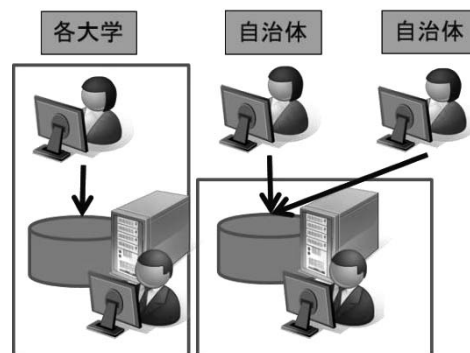


図-03 統合前の運用モデル

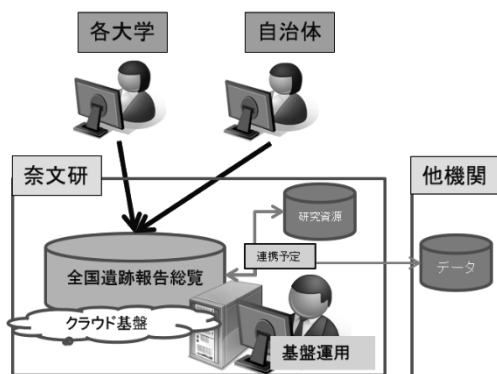


図-04 統合後の運用モデル

統合前は、各連携大学がそれぞれにリポジトリシステムを運用し、各自治体から提供された報告書データを代行登録するモデルであった。あるいは、連携大学がない都道府県は、広域版システムに自治体担当者が直接登録するモデルであった[図3]。統合後は、全て一本化しすべての連携大学や自治体が直接、奈文研システムに登録する[図4]。

## 4. システムの構成

### 4.1 収録しているデータの内容

#### (1) 発掘調査に関する情報

遺跡の発掘調査では主に、発掘区の設定、表土剥ぎ、遺構検出、遺構発掘、遺構実測・写真撮影の各工程からなる。調査中には遺物や分析試料が出土する。これらの遺構や遺物・分析試料に対し、実測図や写真撮影で情報化し、所見を文章記述する(4)[図5]。

発掘調査で得られる情報は、①遺跡情報、②遺構情報、③遺物情報、④調査情報・文献情報がある[図6]。

- ① 遺跡情報：地域の中での当該遺跡に関する情報であり、地形図・測量図が該当する。
- ② 遺構情報：個々の遺構や遺構群に関する情報であり、実測図・写真・テキストとなる。
- ③ 遺物情報：個々の遺物や一括遺物に関する情報であり、実測図・写真・テキストとなる。
- ④ 調査情報・文献情報：調査の経緯や関連文献に関する情報であり、テキストとなる。

発掘調査報告書は主に以上の情報で構成する。情報の形式として画像(地形図・測量図・実測図・写真等)とテキストになる。後付として、調査組織や得られた成果などを記載した報告書抄録をテキスト形式で添付する[図7]。

#### (2) 発掘調査報告書の書誌情報

発掘調査報告書は300部程度印刷され、関係機関や図書館に配布される。図書として扱われるため、メタデータとして書名や著者名など書誌情報

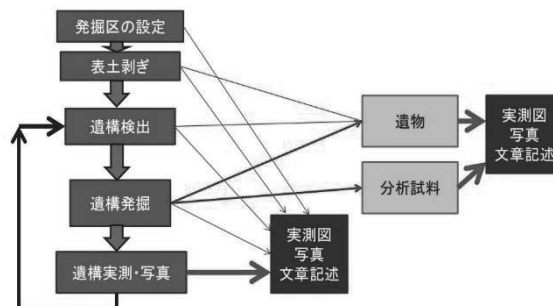


図-05 遺跡の発掘調査

(「全国遺跡資料リポジトリ・シンポジウム 文化遺産の記録をすべての人々へ! —発掘調査報告書の電子化と公開を考える—」2012から転載)

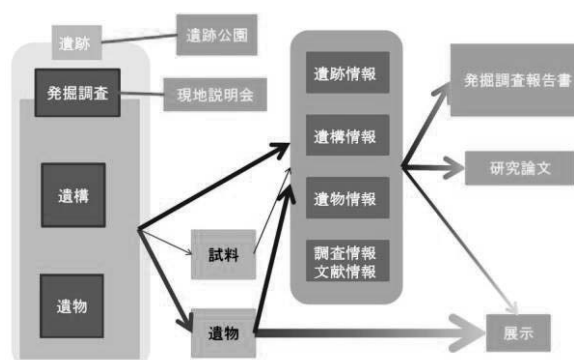


図-06 発掘調査の情報

(「全国遺跡資料リポジトリ・シンポジウム 文化遺産の記録をすべての人々へ! —発掘調査報告書の電子化と公開を考える—」2012から転載)

遺跡名	若草伽藍跡
遺跡名かな	わかくさからんあと
本内順位	
遺跡所在地	奈良県生駒郡斑鳩町
所在地ふりがな	ならけんいこまぐんいかるがちょう
市町村コード	29344
遺跡番号	
北緯(日本測地系)	343637
東経(日本測地系)	1354409
北緯(世界測地系)	
東経(世界測地系)	
調査期間	19680816-19680919 19690930-19691125
調査面積	902.75
調査原因	学術調査
種別	寺院
主な時代	飛鳥時代
主な遺構	金堂・塔の基壇 堀込地業
主な遺物	軒瓦、道具瓦、丸瓦、平瓦、土師器、須恵器、石製品、金属製品など
特記事項	金堂と塔が南北に並ぶ伽藍配置。整地土の重複関係から、まず金堂を造営し、次に塔を建設したことが判明した。

図-07 報告書抄録の一部



を持つ。利用者が発掘調査報告書にアクセスする際は、基本的に通常の図書と同じように書誌情報を使用する。

(3) 発掘調査報告書の PDF 化

発掘調査報告書は、将来にわたって保存され、広く公開し活用されるべきものである。内容は画像情報とテキスト情報で構成する。そのため発掘調査報告書を電子的に扱うためには、画像とテキストを扱え、特定の OS 環境に依存せず、広く社会で使われているデファクトスタンダードのデータ形式である必要がある。ISO32000-1 として標準化されている PDF (Portable Document Format) が要件を満たしており、PDF 形式を採用している。紙媒体の報告書を PDF 形式で電子化し、OCR 処理を施している。近年の発行分に関してはボンデジタルとして DTP ソフトから PDF を作成している。

4.2 システムの構成

遺跡総覧では、Web システムとして一般公開している。利用者はインターネットを介した HTTP 通信にてローカル PC のブラウザ上で操作する。システム構成図を[図8]に示す。

データ登録時には夜間処理にて PDF データから本文の全文テキスト抽出処理と、報告書表紙のサムネイル画像作成処理がある。テキスト抽出処理の結果は、全文検索エンジンと連携する。サムネイル画像作成処理で生成された画像は、遺跡総覧の画面で発掘調査報告書のサムネイル画像として使用する。

5. 開発のメリットおよび課題

5.1 技術的観点 (情報基盤)

(1) ネットワークおよびシステム負荷

奈文研によるシステム・データ統合の以前は、各連携大学図書館が個別にサーバ・システムを運用していた。当然、ネットワークやシステム負荷は分散していたため、性能面でそれほど懸念することはなかった。しかし、各連携大学および広域版システムの 22 のシステムを統合することによって、負荷もすべて集中することになる。

過去の実績から 1 ファイルあたり平均 30MB 程度の PDF ファイルに対して年間 50 万のダウンロードあることが判明した。奈文研の通信回線が 100Mbps であることや、多数の外部公開用 Web システムを公開していること、研究所業務に影響の恐れがあることから、所外にシステムを設置することを決定し、将来的に CDN (キャッシュサーバ) の利用も可能なクラウドプラットフォームを採用した。

(2) ストレージおよびデータ保全

システム開発時、データ統合後に発掘調査報告書のデータ登録がどれだけのペースで進捗するか予期することが困難であった。クラウドプラットフォームを採用することで、柔軟にストレージを増強することができ、過不足のない情報基盤の構成が可能になった。

ディザスタリカバリとして、遠隔地サイトへバックアップデータを複製する遠隔地バックアップを導入した。東日本・西日本のサイトにデータを保管している。

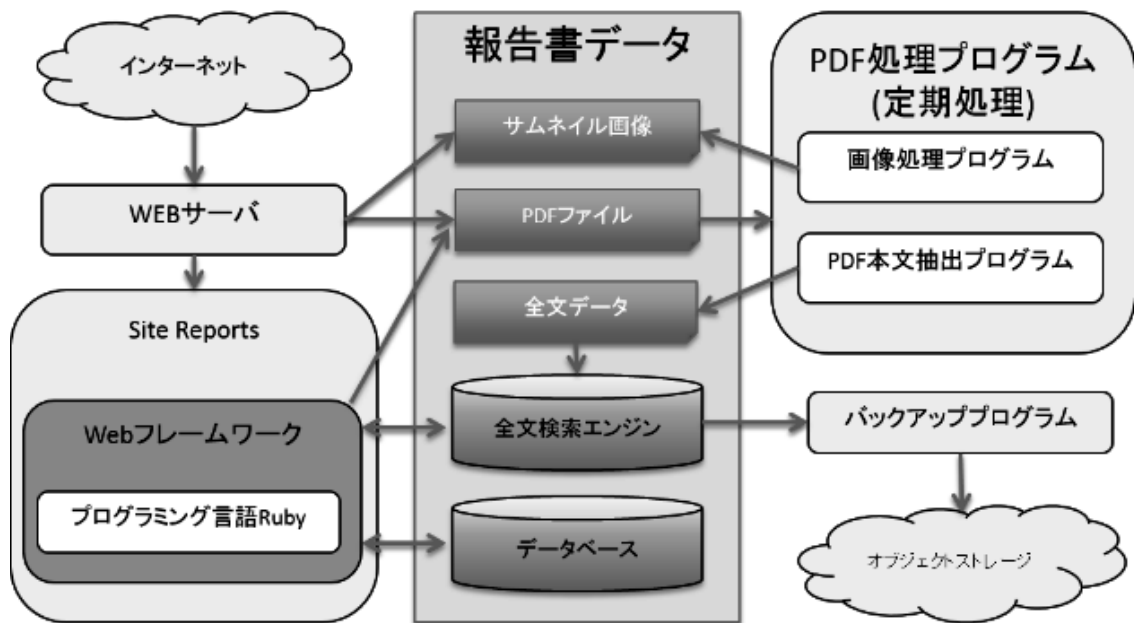


図-08 全国遺跡報告総覧 システム構成図

(3) 他サービスとの連携

国際標準 OAI-PMH によりメタデータを公開しているため、国内外の様々な検索サービス（サービスプロバイダ）を通じてデータ連携が可能である。2015年9月、ディスカバリーサービスである ProQuest 社の Summon と連携を開始した。ディスカバリーサービスとは電子ジャーナル、OPAC（オンライン蔵書目録）、データベース、機関リポジトリ等の多様な膨大な情報資源に対し、一括検索できるサービスである。既に島根大学付属図書館が Summon をカスタマイズして運用しているディスカバリーサービス「ビビッとサーチ」から、遺跡総覧のデータを検索することができる。他システム連携によって、より幅広い利用者層に発掘調査報告書のデータを提供することができる。

5.2 システム運用の観点

(1) 維持管理のコストおよび機能改善

システム統合前は、連携大学ごとにサーバを運用していたため、システム管理者がそれぞれ必要であった[図3]。サーバ保守の維持管理コストが重複していたが、システム統合によってサーバの維持管理が一本化されたこと、スケールメリット

(規模効果)が得られたことによってコストや手間を削減することができた。また機能改善のためのシステムの追加開発についても効率化された。

(2) データ入力規則の統一

システム統合前は、連携大学ごとにデータ登録の独自ルールがあり、若干の表記の揺れがみられた。データの品質を高め、検索精度を向上させるために入力規則を定めデータクレンジングを実施した。特に編著者名の異体字の誤りや、(株)と株式会社などの法人名表記の不統一を正し、名寄せを実施した。新規データ登録時に再度、揺れが発生することを防止するため、システムとして入力チェック機能を実装することや登録マニュアルを整備することで、データ品質の低下を防止している。

このようにデータ品質を高め、情報資源化することが可能になったのはシステム統合のメリットのひとつである。

5.3 利用者の観点

(1) 考古学研究の発展に寄与

2015年9月時点では、33道府県・315機関の報告書が登録されており、件数1万4594冊、PDFページ数169万6328、文字数10億5492万7789である。この約10億字に対し、1度の検索で全国の報告書全文を検索できる[図9]。該当部分は本文をマーカー表示することで、わざわざPDFファイルを開かなくとも文脈を確認できるようになっており、効率的な閲覧が可能である[図10]。

考古学は過去の事例を蓄積し、分析・研究を行う蓄積型の学問であり、研究史の把握が重要である。登録済みの発掘調査報告書に限られるものの効率的に類例・前例を調査できるようになった意義は大きい。既に一部の地域では発掘調査報告書のデータ登録が網羅されつつあり、そのような地域では網羅的な事例調査を可能にする環境が整いつつある。遺跡総覧によって調査漏れが減少することで研究史の整理が精緻化し、研究の高度化に寄与する。研究情報のWebシステムによる公開によって、他分野研究者からの情報アクセスの敷居が下がり、学際的な研究を促進する。

しかし、現在は単純なキーワード検索となっている。発掘調査報告書の本文では用語の表記に揺れが存在する。例えば石切場、石切り場、採石場、石切丁場、石丁場などである。歴史的用語として用法に使い分けはあるものの、意味として同じ場合がある。当該分野に精通していれば、用語の使い分けは可能であるが、そうでない場合はシソーラスが必要となる。シソーラスを活用したあいまい検索を導入することが今後の課題である。

(2) 容易な情報入手

既に述べたように発掘調査報告書は300部程度の発行のため、一般市民にとって入手すること



図-09 検索結果画面

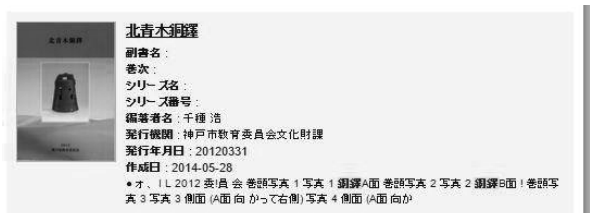


図-10 該当の検索文字をマーカー表示

が困難である。遺跡総覧によってウェブ上で閲覧可能となり、情報入手が容易となった。研究者はもちろん、一般の方も報告書を通して新たな歴史的認識を得ることができる。また近年、地域おこしに遺跡を活用するケースがある。既に地域おこしの一環として、遺跡の情報発信で活用したいという問い合わせもあり、期待が高い。公益的なシステムとして地域学習や調査成果の社会への還元に貢献するであろう。

### (3) 大規模災害への備え

菅野智則氏は、大規模災害への被災地支援の一つとして「研究環境復旧への支援」が必要と指摘している(5)。その理由に地元文化財の価値を見極めるには研究環境が必要だとし、震災復興関連の調査では発掘調査報告書の使用機会が増えるとしている。津波等の大規模災害では、図書館や研究機関が甚大な被害を受け、発掘調査報告書そのものが消失するケースがある。遺跡総覧は、情報の消失を防ぎ、研究の基礎資料を提供するプラットフォームとして震災復興に貢献できる可能性がある。

## 6. あとがき

遺跡総覧は、約10億文字のテキストデータを保持している。考古学ではこの規模の情報量を集約した例はない。将来的に自然言語処理の技術を展開することで、発掘調査報告書の新たな学術的利用が可能になり、考古学研究の高次化・効率化に貢献できる情報基盤となる可能性がある。

当面はすべての発掘調査報告書の登録を網羅するためのスキームの検討と、実際に使ってもらえるプラットフォームを目指して機能改善を進めることが課題である。

## 参考文献

- 1) 国立文化財機構奈良文化財研究所埋蔵文化財センター：埋蔵文化財ニュース, No161, (2015).
- 2) 文化庁文化財部記念物課：発掘調査のてびき-整理・報告書編-, (2010).
- 3) 北条芳隆：発掘調査報告書の閲覧環境整備に向けて-遺跡資料リポジトリの活用-, 考古学研究, Vol. 57, No. 4, pp.6-10 (2011).
- 4) 森本晋：遺跡情報記録と発掘調査報告書, 「全国遺跡資料リポジトリ・シンポジウム 文化遺産の記録をすべての人々へ！ -発掘調査報告書の電子化と公開を考える-」予稿集, 全国遺跡資料リポジトリ・プロジェクト, (2012).
- 5) 菅野智則：被災地の復興支援と遺跡資料リポジトリ, 4)と同じ。菅野友則・永井伸：遺跡資料リポジトリと震災復興支援, 宮城考古学, No. 15 (2013).