

CA1936

全国遺跡報告総覧における 学術情報流通と活用の取り組み

たかた ゆういち
高田祐一*

1. はじめに

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)は、発掘調査報告書を中心に文化財報告書(以下、報告書)を全文電子化しインターネット上で検索・閲覧できるようにした「全国遺跡報告総覧」⁽¹⁾(以下、遺跡総覧)を2015年6月から運営している(図1)。2017年度は年間約100万件のダウンロードがあり、活発に利用されている。島根大学附属図書館を中心とした国立大学、地方公共団体(以下、地公体)・博物館・学会等と共同推進している事業である。これまでの経緯やメリットは拙稿(E1700参照)を参照願いたい。



図1 全国遺跡報告総覧

遺跡総覧は、主に遺跡の発掘調査報告書を登録している考古学・歴史学分野の主題リポジトリである。特定分野の情報基盤であるため、利用が研究者や文化財担当者などに偏る恐れがある。しかし埋蔵文化財(遺跡)は、国民共有の財産であることから、発掘成果を取りまとめた報告書においても、国民に広く公開され共有し活用されなければならない⁽²⁾。そのため遺跡総覧は、幅広く情報を流通させ、活発に利用してもらうことを強く志向したシステムである。本稿は、特定分野の専門情報を広く周知し、より活発な活用に結び付けるための実践例の報告である。

*奈良文化財研究所

2. 最近の動向

2016年度における日本の発掘調査報告書の発行点数は1,492点であった⁽³⁾。そのうち地公体による発行点数は1,334点で89%を占める。報告書の活用を進めることは文化財行政の一環であるが、報告書の電子データ公開による活用促進はこれまで位置付けが曖昧だった。2017年には、文化庁から報告書『埋蔵文化財保護行政におけるデジタル技術導入について2』(報告)が公表され、デジタルデータの報告書の行政的位置付けが明確化されるとともに、全国地公体に遺跡総覧への積極的な登録が呼びかけられた⁽⁴⁾。2017年度には、実務的な説明会を全国5会場で、文化庁の後援により奈文研が主催して開催した。このように行政的な枠組みが整理されたことによって遺跡総覧への参加機関(登録IDを持つ機関)は864機関となった(2018年5月18日現在。2017年3月時に比較して約750機関増)。遺跡総覧が持続的に発展していくための行政上の基盤が確立しつつある。

3. 情報を広く流通させるシステム連携

3.1. システム連携の目的

前述の文化庁報告では、従来の印刷物の報告書とデジタル(PDF)の報告書は、役割分担し相互補完するものとしている。デジタル技術は検索性に優れていることから、遺跡総覧を「報告書のインデックス」として、幅広い活用に道を拓くものと位置付けている。

さらに、各種の図書館系システムとの連携によって印刷物の報告書へ橋渡しする役割を果たすことができる。

遺跡総覧を様々な情報基盤と連携させることによって、遺跡総覧を知らなかった人も気が付けば遺跡総覧の情報にアクセスしているという状況を創出するためにも、外部システム連携は重要と考える。文化財に関心を持つ層の裾野を広げることにもつながるだろう。次節では、特に遺跡総覧で既に実現している図書館系のシステムとの連携を紹介する(図2)。

3.2. 図書館系システムとのデータ連携

遺跡総覧に登録されているメタデータには、国立国会図書館(NDL)のJP番号(全国書誌番号)とNACSIS-CATのNCIDを登録している。遺跡総覧の検索結果画面でこれらの書誌番号をクリックすることで、国立国会図書館サーチやCiNii Booksなどの図書館系システムに遷移して、当該報告書の印刷物の所蔵機関を調べることが可能である。報告書はシリーズや分冊の扱いなど、発行機関によって揺らぎがあり、NDLとNACSIS-CATでは、書名などの基本的な書誌情報が異なる場合がある。遺跡総覧では、前述のようにそれぞれの書誌番号を記載することで、書誌データの異同を

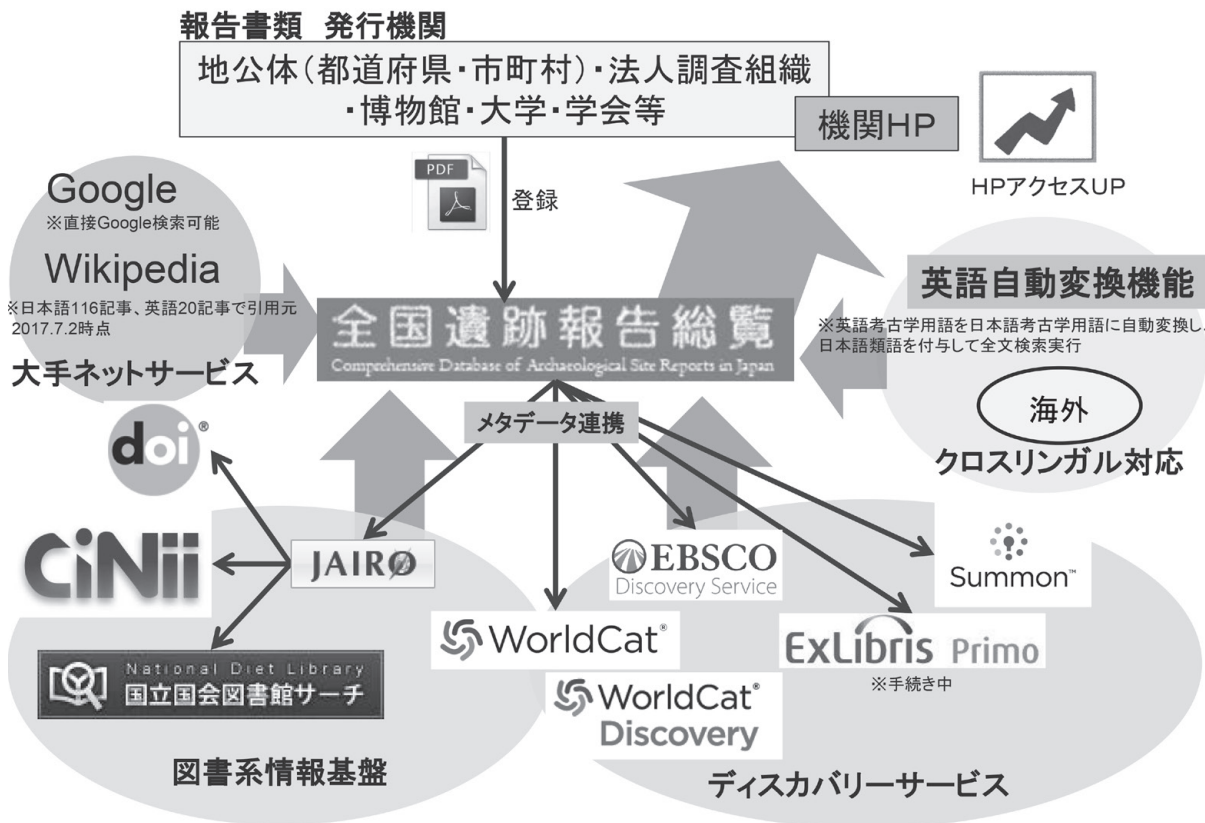


図2 全国遺跡報告総覧のデータ連携状況

確認できるようにしている。

2016年3月には、CiNii Booksとデータ連携を開始した⁽⁵⁾。CiNii Books で印刷物の報告書を検索すると、検索結果から遺跡総覧のデジタル版報告書に直接リンクしている。CiNii Books から、出版物1点1点について他の電子リソースへ直接リンクする方式は遺跡総覧が第一号となった。2017年2月には、WorldCatともデータ連携を開始した。WorldCatでは海外の人が使うことから、書名よみをローマ字化して連携している⁽⁶⁾。

3.3. ディスカバリーサービスとのデータ連携

ProQuest社のSummon(2015年9月)⁽⁷⁾、WorldCat Discovery Services(2017年2月)⁽⁸⁾、EBSCO Discovery Service(2017年4月)⁽⁹⁾、Primo(2018年10月予定)など、各種のディスカバリーサービスとの連携も開始した。ディスカバリーサービスと連携することで、遺跡総覧のコンテンツの発見可能性がさらに高まったと言える。各ディスカバリーサービスへはjunii2形式によりOAI-PMHでメタデータを提供している。

3.4. 国内外の考古学情報基盤との連携

日本の学術研究成果を海外にも発信するには、海外の専門家が使用する情報基盤とデータを連携させることが効果的であろう。ARIADNE(欧州考古学統合情報基盤)は、欧州中の考古学情報を統合し、相互連携によって情報にアクセスしやすくするシステムの構築、コミュニティの組成に取り組んでいる事業である。2019年から開始される次期計画のARIADNE Plusではコミュニティの拡大が重要課題と位置づけられており、25か国40機関が参画する見込みである。欧州以外の国(米国・日本・アルゼンチン)が初めて事業に参画する予定となっている。海外の情報基盤との連携を安定化させるため、2017年7月、遺跡総覧の報告書にDOIを付与した⁽¹⁰⁾。遺跡総覧のデータ(遺跡情報・報告書PDF)をARIADNEと連携させることで、さらなる情報流通の拡大が見込まれる。

国内においては、日本旧石器学会のデータベース『日本列島の旧石器時代遺跡』(JPRA-DB)⁽¹¹⁾において、報告書の出典情報に遺跡総覧を使用している。このように遺跡総覧は、専門的なデータベースにおいて、報告書情報を提供するプラットフォームとなっている⁽¹²⁾。

3.5. 外部システムとの連携効果検証

遺跡総覧へのアクセス流入元に関するアクセス統計から外部システムとの連携の効果を検証する。2017年1月から12月にかけてのアクセス統計から流入元別に集計すると、ユーザ数が多い順にGoogle:36%、Yahoo:29%、直接(不明含む):19%、Bing:4%、CiNii:1%、Wikipedia:1%であった。Google、Yahoo、Bingはウェブ検索エンジンであり、大半の利用者がウェブ検索エンジンの検索結果からアクセスしていることがわかる。CiNii、Wikipediaからは、ページに設定された遺跡総覧へのリンクから流入している。Wikipediaの記事に関しては出典に遺跡総覧掲載の報告書が活用される場合があり、それらの情報参照で流入していると思われる。上記の流入元の残りの10%に流入割合1%未満の各ディスカバリーサービス、JAIRO、WorldCat、多数の地公体ウェブサイト等の約500のサイトがロングテール状に分布する。国立国会図書館サーチの検索結果からはJAIROを経由して遺跡総覧のコンテンツにリンクするため、JAIROは実質的には国立国会図書館サーチからの流入だろう。図書館系サービス(CiNii Books、WorldCat)からの流入は多いとはいえない状況である。報告書は書名のみでは内容がわからないため、ユーザにとっては要否を判断できない可能性がある。また約500のサイトの大半を占めるのは、地公体ウェブサイトからである。地公体ウェブサイトでは自機発行の報告書をウェブサイトにて紹介し、遺跡総覧にリンクを貼っていることから流入が多くなっていると考えられる。

流入元の約7割がウェブ検索エンジンである。アクセス数を増やすには、ウェブ検索エンジンへの対応が必須といえる。一方、様々なサイトからリンクをたどってアクセスするユーザもいる。遺跡総覧は、動的な検索が可能なデータベースであるが、コンテンツごとに固定URLを付与している。固定化したURLであることが、他サイトからのリンク設定を容易にしておき、加えてDOIを付与することでアクセスを保証している。

4. 文化財関係用語の整理と展開

遺跡総覧の主たるコンテンツは報告書のPDFデータである。しかしPDFの単純な掲載に止まらず、印刷物ではできないデジタルならではの機能開発を進めている。目指すところは、日々増大する膨大なデータから適切に情報アクセスできる環境の提供である。遺跡総覧は全文情報を保持していることから、全文テキストを最大限活用した機能の開発を行った。本章では文化財関係用語の整理と全文テキストを活用した機能を紹介する。

4.1. 英語自動検索

日本には膨大な文化財の調査成果が蓄積されているものの海外のユーザが手軽にアクセスできる環境ではない。原因のひとつに言語の問題がある。日本の学術用語には類語が多数あり、それらを高度に習熟しなければ、網羅的な検索は困難である。そこで、文化財関係用語の日英対訳と類語を整理したデータベースを作成し、遺跡総覧に搭載している(図3)。この機能により、英語の考古学用語を検索ワードとして入力すると、日本語の考古学用語に自動変換したうえで、類語を含めて検索することが可能である⁽¹³⁾。

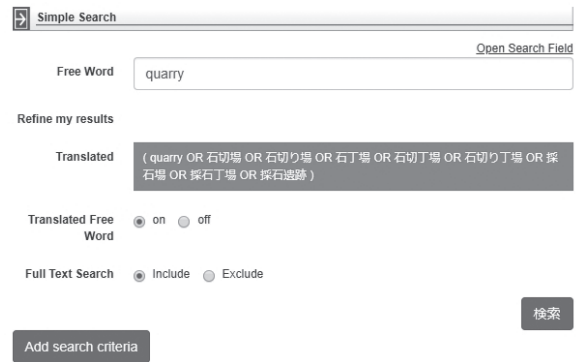


図3 英語自動検索

4.2. 頻出用語と特徴語の可視化

キーワード検索では、ユーザは予め検索対象となる事物を言い表す用語を知っていなければ、検索することができない。そこで用語を知らなくても検索できる機能を開発し、実装した。登録されている全ての報告書の全文を対象にして、日本全体あるいは各都道府県での頻出用語と特徴語(その地域では頻出語であるが、他地域では希少語)を可視化し、クリックするだけでその語を使った検索ができるようにした⁽¹⁴⁾(図4)。遺跡総覧は7万411語の文化財関係用語の辞書を内蔵しており、辞書をもとに報告書全文テキスト内の語の出現数をカウントしたものである。

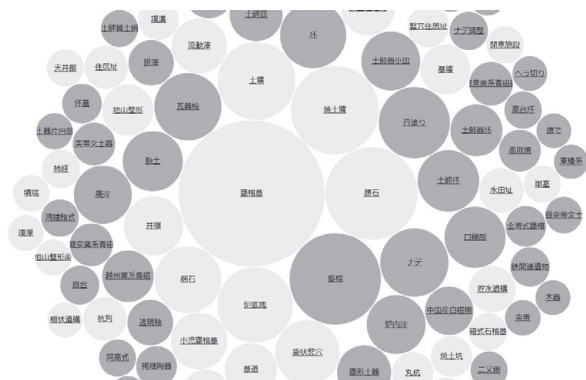


図4 特徴語の可視化(福岡県)

4.3. 類似報告書の提示

全文検索では、すべてのテキストデータを対象とするため、周辺遺跡への言及など直接的に関係のない報告書がノイズとして検索結果に含まれてしまう。そこで、報告書ごとに本文類出用語を上位40点抽出し、報告書詳細ページに表示した。報告書における類出語は、本文の内容をある程度忠実に表す場合が多いと考えられる。表示された類出語をクリックした場合、当該用語を類出語とする報告書のみが絞り込まれて表示される。若干の漏れは発生しうると思われるが、ユーザが必要とする高精度な結果を期待できる。



図5 類似報告書の表示機能

また、上位40の語の構成と類似する他の報告書を表示する機能がある(図5)。蓄積型の学問である考古学・歴史学は、網羅的な類例調査が不可欠であり、内容が類似する報告書の自動表示は、研究を手助けするツールとなりうる。この機能は、遺跡総覧のページ閲覧数を劇的に向上させた。この機能を公開する前の閲覧ページ数は2016年度は1,155万ページであったが、公開後の2017年度には7,277万ページとなった。

4.4. イベント情報の登録と公開

遺跡総覧には文化財関係イベントを公開できる機能があり地公体などのイベントが登録されている。この文化財イベントの本文情報について、特徴語を抽出し、その語の構成と類似の報告書類を自動表示している⁽¹⁵⁾。文化財イベントに参加する前後に参考となりそうな報告書類を閲覧することにより、理解を深めてもらうことが目的である(図6)。報告書成果の新たな活用方法の一つとなっている。

文化財資料と活用事業の相乗効果

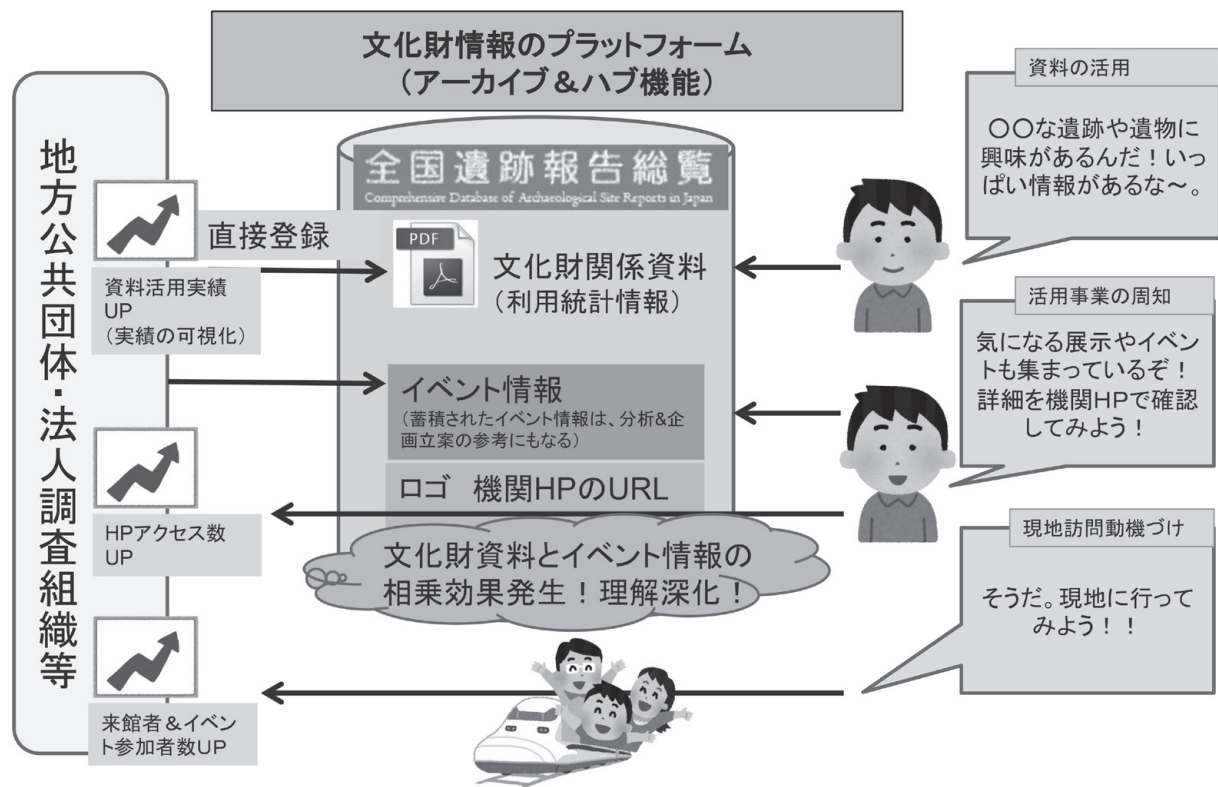


図6 文化財資料とイベント情報の相乗効果

5. おわりに

遺跡総覧事業は、島根大学附属図書館を中心とした大学と地方公共団体等が共同で進めている事業である。システム連携など情報流通のノウハウは図書館の得意とするところである。遺跡総覧が発展していくためには、今後も情報の取扱いに長けた図書館の役割は大である。事業の代表機関である奈文研は、発掘を実施し報告書を作成するデータ作成機関でありながら、他機関のデータを取りまとめて提供する機関でもある。また、筆者含め所員は、遺跡総覧を活用して研究するユーザの立場でもある。各機関の得意分野を活かしつつ関係機関の課題解決や目的達成に寄与できるシステムであることが存在意義として重要と考える。そのうえで、文化財分野における幅広い情報流通と活用の具体的施策を今後も考えていくことが必要であろう。

と文化財イベント情報連携機能の強化”。奈良文化財研究所。2017-11-06。
<https://www.nabunken.go.jp/nabunkenblog/2017/11/renkei.html>, (参照 2018-06-28)。

[受理：2018-08-06]

Takata Yuichi
 Approach to Distribution and Utilization of Academic Information by Comprehensive Database of Archaeological Site Reports in Japan

- (1) 全国遺跡報告総覧. 奈良文化財研究所. <https://sitereports.nabunken.go.jp/>, (参照 2018-07-13).
- (2) “発掘調査報告書”. 発掘調査のてびき - 整理・報告書編 -. 文化庁文化財部記念物課編. 文化庁, 2010, p. 2.
- (3) 文化庁文化財部記念物課編. 埋蔵文化財関係統計資料. 平成29年度. 文化庁, 2018, 33p.
- (4) 埋蔵文化財発掘調査体制等の整備充実に関する調査委員会. 『埋蔵文化財保護行政におけるデジタル技術の導入について2』(報告). 文化庁, 2017, 59p.
- (5) NII学術コンテンツサポート. “CiNii Booksと全国遺跡報告総覧とのデータ連携開始のお知らせ”. 国立情報学研究所. 2016-03-23. <https://support.nii.ac.jp/ja/news/cinii/20160323>, (参照 2018-07-13).
 なぶんけんブログ. “全国遺跡総覧とCiNii Booksのデータ連携開始”. 奈良文化財研究所. 2018-03-23. <https://www.nabunken.go.jp/nabunkenblog/2016/03/cinii-books.html>, (参照 2018-07-13).
- (6) なぶんけんブログ. “全国遺跡報告総覧とWorldCatのデータ連携開始”. 奈良文化財研究所. 2017-02-07. <https://www.nabunken.go.jp/nabunkenblog/2017/02/worldcat.html>, (参照 2018-07-13).
- (7) なぶんけんブログ. “全国遺跡報告総覧:ディスカバリーサービスSummonが全国遺跡報告総覧に対応”. 奈良文化財研究所. 2015-09-02. <https://www.nabunken.go.jp/nabunkenblog/2015/09/sumon.html>, (参照 2018-07-13).
- (8) “全国遺跡報告総覧 WorldCatと連携!”. OCLC News. 2017, 38(4), p. 1. https://www.kinokuniya.co.jp/03f/kinoline/1704_06.pdf, (参照 2018-06-28).
- (9) なぶんけんブログ. “全国遺跡報告総覧:ディスカバリーサービスのEDSが全国遺跡報告総覧に対応”. 2017-04-11. <https://www.nabunken.go.jp/nabunkenblog/2017/04/eds.html>, (参照 2018-07-13).
- (10) 国武貞克, 小沼美結, 高田祐一. 文化財情報の多国間連携による研究基盤の高次化. 奈良文化財研究所紀要. 2018, 2018, p. 18-19.
- (11) データベース『日本列島の旧石器時代遺跡』. 日本旧石器学会. <http://palaeolithic.jp/data/index.htm>, (参照 2018-07-13).
- (12) 野口淳. 新しい『日本列島の旧石器時代遺跡』データベース - オープンデータ・オープンサイエンス時代の考古学研究を目指して -. 日本旧石器学会ニュースレター. 2018, 38, p. 1-3. <http://palaeolithic.jp/nl/newsletter38.pdf>, (参照 2018-07-13).
- (13) なぶんけんブログ. “全国遺跡報告総覧:英語自動検索機能公開のお知らせ”. 奈良文化財研究所. 2016-08-24. <https://www.nabunken.go.jp/nabunkenblog/2016/08/post-87.html>, (参照 2018-06-28).
- (14) なぶんけんブログ. “全国遺跡報告総覧:考古学ビッグデータの定量的可視化:日本の発掘調査成果を1枚の画像で表現すると?”. 奈良文化財研究所. 2017-04-27. <https://www.nabunken.go.jp/nabunkenblog/2017/04/hinshutu.html>, (参照 2018-06-28).
- (15) なぶんけんブログ. “全国遺跡報告総覧:文化財調査報告書