

発掘調査報告書の電子公開による情報発信とその新たな可能性

高田祐一（奈良文化財研究所）

New possibilities for the dissemination of information via electronic publication of
archaeological reports

TAKATA Yuichi (Nara National Research Institute for Cultural Properties)

・発掘調査報告書／Archaeological excavation reports
・電子公開／Electronic Publication・データベース／Data base

1. はじめに

2018年12月時点において、新聞で「人工知能(AI)」「ビッグデータ」「自動運転」等のキーワードを見ない日はない。これは、高度な情報技術を活用する環境が、社会に急速に普及していることを示す。18世紀の産業革命のように、社会の在り方さえも変える情報技術が社会基盤となりつつある今、文化財情報に対しても、適切にフォローアップしていく必要がある。

昨今の情報技術の根幹は、データである。自動運転技術の実現には、走行映像や地図等の膨大なデータが必要とされる。SNS系企業は、個人情報のプラットフォームとして、収集したデータを活用・分析することで、サービス改善に役立てている。

そうした技術進展や社会の変化を前提に、本稿では、データの活用という観点から、発掘調査報告書（以下、報告書とする）の情報発信と、次世代への新たな可能性について述べる。

2. 考古学における報告書の役割

考古学・歴史学は、蓄積型の学問である。考古学的遺構、遺物、歴史資料を調査研究し、知見を蓄積していくことによって、調査研究を深化させていく。そのため、情報が蓄積されるほど善であり、遺跡調査の成果物である報告書が、重要な情報資産となる。報告書は、考古学においては貴重な学術資料

であり、埋蔵文化財行政においては基礎情報となる。文化庁『発掘調査のてびき』によれば「報告書は、(中略)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(文化庁2010)。報告書は保存し、公開共有し、活用しなければならないものである。

3. 報告書の刊行冊数状況と情報爆発

2016年度における日本の報告書の刊行冊数は、1,492冊であった(文化庁2018)。近年の年度ごとの刊行冊数について、文化庁による統計調査以降は判明している。しかし、総数については、不明である。戦前を含めて、おそらく10数万冊から20万冊程度の報告書が刊行されているとみられる(高田2018)。奈良文化財研究所(以下、奈文研)では、全容を把握すべく、都道府県別に報告書の総目録を整理しており、2018年4月に兵庫県版を公開している(奈文研a2018)。

報告書の把握が困難になることは、過去から指摘されていた。1970年代後半に、既に文化財関係資料が膨大となり「資料の全貌は、もはや誰にも把握しきれない。このため現在、研究、文化財・保護の仕事にたずさわる者が、過去の資料の蓄積を適切に選択して利用するのは、大変に難しいという状況にあり、将来この傾向がさらに甚だしくなることは目にみえている」と指摘されている(岩本1977)。

1980 年台でも、依然「多量の考古学資料の蓄積、膨大な情報量が、そのまま素晴らしい研究成果を生むものになっているとはいいがたい」状態であり、「発掘調査のもたらす多量の情報に対処しうる情報処理システムの確立」が必要とされた（田中1982）。

1994 年には、報告書の内容を要約した抄録を、報告書に添付する行政的取り組みが開始した。文化庁記念物課から、1994 年 4 月 27 日付け「埋蔵文化財発掘調査報告書の抄録の作成について」（6 保記第 16 号）が、都道府県あてに通知される。その後、抄録のデジタルデータを全国的に収集し、データベース化することも開始した（奈文研 2017）。関係機関の協力によって、抄録は現在大半の報告書に添付されている。抄録には、遺跡の位置情報、時代情報、遺構遺物の情報が掲載されており、有効に活用できれば、必要とする情報にピンポイントでアクセスできる、画期的で重要な取り組みである。

しかし、「発掘が終了しても手つかずのままの考古資料が各地の収蔵庫に山積みされ、年度末に刊行される発掘調査報告書も、その活用度はけっして高くはない。いわば制御できないほどの情報を、日本考古学は抱えてしまった」（広瀬 2015）という状況である。蓄積型学問である考古学において、情報蓄積が進むのは歓迎すべきであるが、人間が制御できる量を超え、情報爆発の弊害が起きているのが現状である。

4. 遺跡資料リポジトリの展開

全国遺跡報告総覧（以下、遺跡総覧とする）（<https://sitereports.nabunken.go.jp/>）の前身となる「全国遺跡資料リポジトリ・プロジェクト」（以下、遺跡資料リポジトリとする）の経緯を説明する。

遺跡資料リポジトリは、2008 年度から 2012 年度にかけて、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、島根大学を中心とする全国 21 の国立大学が連携し取り組んだプロジェクトである。遺跡資料リポジトリでは、約 1 万 4,000 冊の報告書が電子化され、年間約 50 万件的のダウン

ロードがあるなど、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から、プロジェクトの継続に課題があった。

奈文研では、報告書のメタデータを提供し、共同研究してきた経緯もあり、各連携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移管・統合し、2015 年 6 月から遺跡総覧として運用している。そのような経緯から、大学・自治体・法人調査組織・学会等と共同推進する事業であり、代表機関を奈文研、事務局を島根大学附属図書館としている。

5. 報告書電子公開をめぐる動向

2017 年 8 月 31 日付けの日本学術会議史学委員会文化財の保護と活用に関する分科会が公開した「提言 持続的な文化財保護のために－特に埋蔵文化財における喫緊の課題－」では、遺跡調査情報のオープンアクセスを原則とする公開、および ICT を用いた新たな活用策の研究・開発の推進を提言している（日本学術会議 2017）。遺跡総覧への言及では「文化庁としても事業の永続性の確保に努めるとともに、大学等の教育研究機関を含めて我が国で遺跡調査を行うすべての組織が、報告書の積極的な登録を進め、調査成果の公開に意を払うよう、強く要望するものである」としている。

同年 9 月 25 日には、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）（以下、文化庁デジタル報告 2）が公表された。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、報告書の電子公開による活用促進はこれまで位置付けが曖昧だった。

文化庁デジタル報告 2 によって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、全国地方公共団体に遺跡総覧への積極的な登録が呼びかけられた（文化庁 2017）。

文化庁記念物課より事務連絡「『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）』の送付について」（2017 年 9 月 25 日付け）では、全国地方公共団体に「全国遺跡報告総覧登録意向調

査」が実施された。

これらの一連の動向を受け、事務局は文化庁デジタル報告2によって、報告書のデジタルデータの役割が明確化され、その積極的活用としての全文データ公開は、行政課題として位置付けられたと受け止めた。よって、今後は発行主体である、地方公共団体等による遺跡総覧への直接参加・直接登録（セルフアーカイブ）を原則とし、事務局及び連携大学がそのサポートを行うという基本方針を発表した（全国遺跡報告総覧プロジェクト2017）。

2017年度からは、関係機関による直接登録を支援するため、全国5会場で実務的な説明会を開催している（主催：奈文研、後援：文化庁・各都道府県教育委員会等）。

このような行政的な枠組みの整理と説明会の実施によって遺跡総覧への参加機関（登録IDを持つ機関）は864機関となり（2018年5月18日時点）、遺跡総覧が持続的に発展していくための行政上の枠組み

が確立しつつある。

6. 全国遺跡報告総覧の基本機能

遺跡総覧では、メタデータ及び報告書本文を対象とした簡易検索（キーワード検索）によって、登録されているすべての報告書のテキストデータを、1度に検索することが可能である。通信状況によるが、報告書類約23000件、約17億文字、約280万ページを約3秒で検索できる。検索結果画面に文章を表示し、該当キーワードをマーカー表示することで、ユーザが必要としている報告書かどうか、文脈で判断できるようにした。報告書はダウンロードして閲覧可能である。類例調査が重要である考古学研究において、大量の報告書を全文検索できる効果は大きい。またNACSIS-CATの書誌ID（NCID）をメタデータに設定しているため、CiNii Booksを参照して当該報告書の所蔵機関を調べることも可能である。2017年度の遺跡総覧のダウンロード件数は約100万

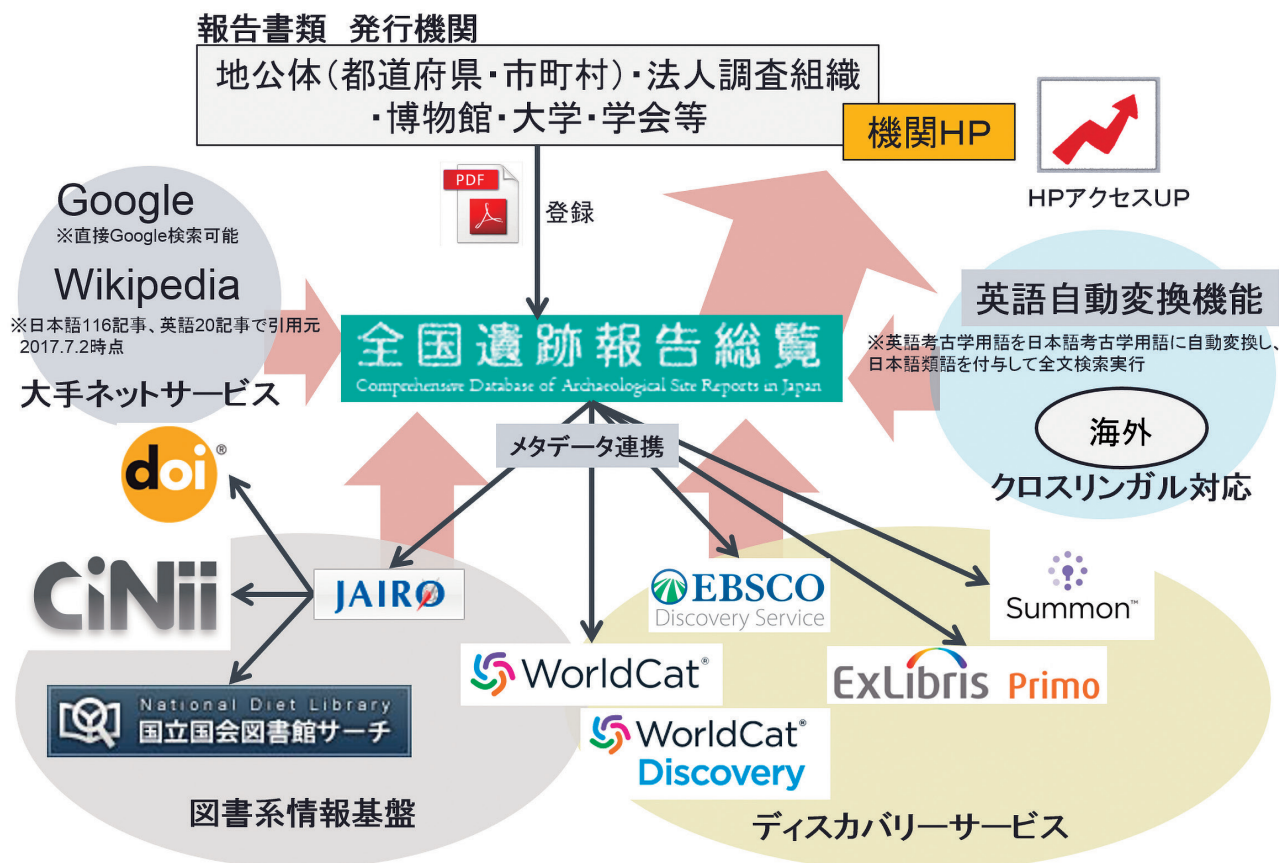


図1 全国遺跡報告総覧の外部システム連携

件、ページ閲覧数は7277万件であった。既に多数のユーザが活発に報告書を利用しているという証左である。

7. 外部情報基盤とのメタデータ連携

残念ながら、すべてのユーザが奈文研や遺跡総覧を知っているわけではない。システムの存在を知らなければ、活用されることはない。遺跡総覧を知らなくとも、気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、外部システムとの連携は重要である（図1）。

図書館系システムでは、CiNii Books、国立国会図書館サーチ、WorldCatと連携している。様々なリソースを同一インタフェースで検索できるディスカバリーサービスでは、ProQuest社のSummon（2015年9月）、WorldCat Discovery Services（2017年2月）、EBSCO Discovery Service（2017年4月）、Primo（2018年12月）と連携を開始した。

海外の考古学情報基盤との連携による日本の学術研究成果発信にむけても動き始めている。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へのアクセスを容易に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機関が参画し、欧州以外の国（米国・日本・アルゼンチン）についても事業に初参画する予定となっている。

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）では、報告書の出典情報に遺跡総覧を使用しており、報告書情報を提供するプラットフォームとしても現在機能している（野口2018）。

8. 統計的自然言語処理技術の活用

遺跡総覧には、報告書類約23000件、約17億文字、約280万ページのデータが登録されている。これら

の膨大なデータについて、適切なアクセスを提供するために下記の機能を提供している。

（1）報告書ワードマップ（頻出用語俯瞰図）

現在遺跡総覧に登録されている報告書類17億文字に対して、こういった考古学関係用語が頻出しているかを可視化したもので、遺物関係（桃色）・遺構関係（黄色）・その他（青色）の3つの種別を付与している（図2）。土器に関する用語が頻出しているのが特徴的である。

（2）各都道府県版 報告書特徴語ワードマップ

当該都道府県内にて頻出する用語、かつ他都道府県では出現頻度が低い希少用語は重要であることを勘案することで、当該都道府県の強い特徴を示す用語を可視化した。自然言語処理技術のベクトル空間モデルのTF（索引語頻度）とIDF（逆文書頻度）を組み合わせたTF-IDFにて算出している。その地域で特徴的な用語を提示することで、キーワードの入力に頼らない検索方法を提供している。専門用語を知らなくとも検索する事が可能になる。

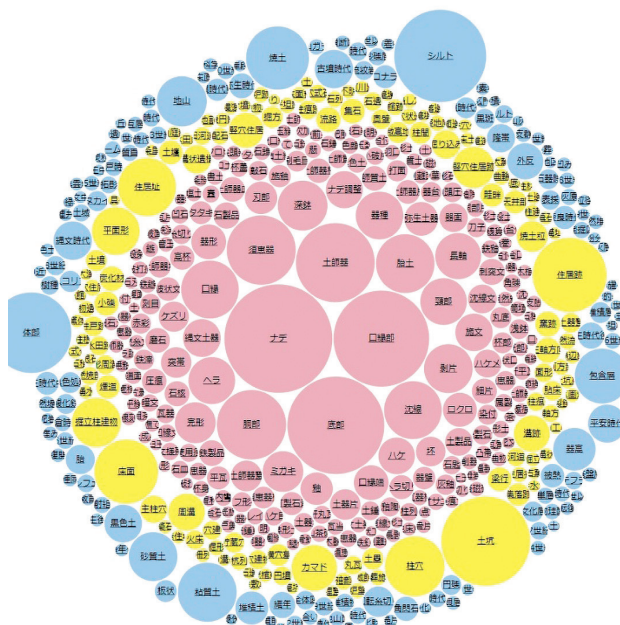


図2 報告書ワードマップ（頻出用語俯瞰図）

（3）類似報告書の自動揭示とイベント情報の連携

調査研究のために類似例を探す場合であっても、全てを通読することは困難である。遺跡総覧では、頻出用語から類似の用語で構成された報告書が、自

大坂城石垣石丁場跡小笠原石丁場跡の海中残石分布調査	
URL	https://nabunken.go.jp/21923
DOI	https://doi.org/10.24454/nabunken.21923
ファイル	21923_1_大坂城石垣石丁場跡小笠原石丁場跡の海中残石分布調査.pdf ダウンロード (42.8 MB) モバイル版 (4.6 MB)
書名	大坂城石垣石丁場跡小笠原石丁場跡の海中残石分布調査
発行(管理)機関	奈良文化財研究所
有償提供・配布ページ	https://www.nabunken.go.jp/publication/goods.html ※ 有償提供・配布していない場合があります
書名かな	おおさかじょういしがいいしちようばあとしようどしまいしちようばあとのかいちうぜんふんぶふんぶちようさ
新書名	2017年度公益財団法人福成財団神戸内海文化研究・活動支援助成(調査・研究助成)『小笠原における石垣築造技術の研究』調査成果報告書
巻次	
シリーズ名	
シリーズ番号	
著者	高田 雄一 進藤 昌 の 蔵 集

図3 報告書の頻出用語(画像右部)

動提示される。(図3)。さらに、報告書内容と類似したイベントを表示することで、文化財事業と報告書の相乗効果を見込んでいる。

(4) 海外から日本の報告書を活用するために

日本考古学の成果に関心を示す海外の研究者には、言語の壁や報告書を手にとって閲覧できないという、情報アクセスの問題がある。日本語を習熟しても、日本の考古学関係用語には多くの類語がある。遺跡総覧では、日英の考古学用語の対訳と日本語の考古学用語の類語をデータベース化し、英語自動変換機能を実装した、文化財関係用語シソーラスを構築した。(図4)。英語の用語を投入するだけで日本語用語に自動変換し、類語を自動で付与する。簡単に言語の壁を超えて網羅的な検索が可能である。

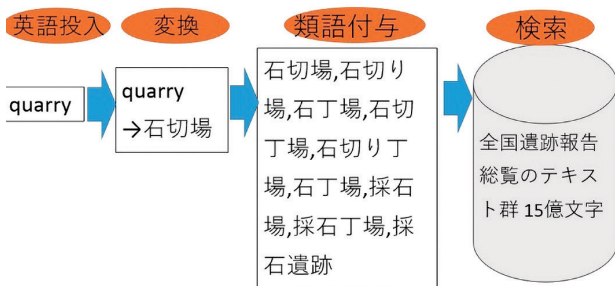


図4 用語の英語自動変換機能

9. 人工知能(機械学習)による画像認識

報告書の情報は、大きくテキストと画像(図面・写真)にて構成される。特に考古学においては、画像情報が重要となる。これまでのデータベースは、

キーワードによるテキスト検索が主であった。検索を利用するためには、事物を言語化し、用語を習得する必要がある、市民や初学者にとっての障壁となっている。しかし、近年の画像認識技術の向上によって、画像で画像を検索することを可能とする技術環境が整ってきた。画像であれば、日本語を母語としないユーザも検索が容易である。

人文系分野での実践適用例では、くずし字画像を投入すると、類似するくずし字画像を結果として表示する「木簡・くずし字解読システム MOJIZO」(<http://mojizo.nabunken.go.jp/>)がある。遺跡総覧でも同様の機能を整備中であり、実現すれば、類例調査に資することができるだろう(図5)。

軒丸瓦 1

PL 48



図5 軒丸瓦の画像自動抽出状況

10. おわりに

考古学や文化財行政において、情報蓄積は命である。そして、蓄積した情報に飲み込まれないよう、適切にアクセスできる環境を整備することも同時に

重要である。文化財情報のプラットフォームとして遺跡総覧を拡張整備することで、考古学研究および文化財行政の推進に資する事ができる。

【註】

- 文化庁 2010「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 p. 2.
- 文化庁 2018「埋蔵文化財関係統計資料.平成29年度」文化庁文化財部記念物課編 p. 33
- 高田祐一 2018「全国遺跡報告総覧と考古学ビッグデータ」『デジタル技術で魅せる文化財－奈文研とICT－』奈良文化財研究所
- 奈良文化財研究所 a 2018『発掘調査報告書総目録：兵庫県編』<http://doi.org/10.24484/sitereports.21896>
- 奈良文化財研究所 b 2017『遺跡情報交換標準の研究第4版』
- 岩本圭輔 1977「埋蔵文化財関係用語の収集と整理」『奈良文化財研究所年報』
- 田中琢 1982「考古学、みかけだけのはなやかさ」『同朋』
- 広瀬和雄 2015「解説」『考古学で現代を見る』
- 日本学術会議 2017「提言 持続的な文化財保護のために－特に埋蔵文化財における喫緊の課題－」日本学術会議 史学委員会文化財の保護と活用に関する分科会
- 文化庁 2017『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）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 59p.
- 全国遺跡報告総覧プロジェクト 2017「全国遺跡報告総覧へのデータ登録について」（平成29年9月25日付け）<https://sitereports.nabunken.go.jp/files/statics/%E5%85%A8%E5%9B%BD%E9%81%BA%E8%B7%A1%E5%A0%B1%E5%91%8A%E7%B7%8F%E8%A6%A7%E3%81%B8%E3%81%AE%E3%83%87%E3%83%BC%E3%82%BF%E7%99%BB%E9%8C%B2%E3%81%AB%E3%81%A4%E3%81%84%E3%81%A6.pdf>
- 野口淳 2018「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』 38