

デジタルアーカイブの現在と データ持続性

後藤 真（国立歴史民俗博物館）

1. はじめに

本章では、歴史情報学と関係の深い「デジタルアーカイブ」の現在について説明をし、歴史資料のデジタル化のありようを考えます。極めて関連の深いデータ持続性についてもここで触れます。歴史情報学・人文情報学と「デジタルアーカイブ」は、極めて近い関係ですが、少し違いがあります。人文情報学は、その機械的な研究活用とそこから何らかの人文的成果を求めるのに対し、「デジタルアーカイブ」は、より広い文化的な活用の文脈などまで含めることが多くなります。また「アーカイブ」の名前の通り、長期的な保存を意識することもあります。

2. 「アーカイブズ」と「デジタルアーカイブ」

まず、言葉の意味から考えてみましょう。「アーカイブ（ズ）」とはどのような意味を持つ言葉でしょうか。「アーカイブズ」とは、（若干迂遠な説明がありますが）以下のように定義されると考えられます。「（平面的媒体に）文字で、かつ、リニアな言語で書かれた歴史的資料・資料群もしくはその資料群を長

期的に保存する機能（組織）。日本において、かつて「アーカイブズ」に関する研究は、主として「古文書学」もしくは「史料学」という分野で扱われることが多くありました。古文書学は、戦前に厳密に史料を読み解くことをはじめた時期からすでに起こっている学問であり、その意味において、非常に歴史のある学問です。また、歴史学を本格的に勉強しようとする、必ず読むことになる佐藤進一氏の『古文書学入門』（法政大学出版局、1971年初版、2003年新装版）なども意識されるのではないのでしょうか。「古文書学」はその名の通り、主たる研究対象はいわゆる「古文書」であり、現用文書はもちろんのこと、現用文書が非現用化してすぐのような時代の浅い文書が対象とされることは多くありませんでした。

しかし、1990年代に入り、特に欧米の研究文脈の中で存在した「アーカイブズ」という学問のあり方が日本で紹介されるに至り、その学問の重要性とともに、概念もよく知られるようになってきました。その学問自体は、大きくは記録資料学などといった言葉への読み替えも試みられましたが、その広範な概念を覆いきれないということで、「アーカイブズ学」と呼ばれ、現在に至っています。現在は、アーカイブズ学会や全国資料保存機関連絡協議会（全史料協）、記録管理学会などが中心となり、その研究を進めています。このアーカイブズ学は、多くは公文書の保存と活用を中心として取り扱う学問でした。

このように考えると、基本的には、少なくともデジタルではないものが持つ「アーカイブ」は比較的狭い厳密な概念であることがわかるでしょう。対象とするものは文字の書かれたものであり、考え方としては、長く未来にいかにかその情報を伝えるかを主眼にしています。しかし、「アーカイブ」に「デジタル」が付くと、その様相は少し異なります。それは歴史的な経緯が大きく関係しています。そこを見てください。

3. 「デジタルアーカイブ」とは

「デジタルアーカイブ」なる語を最初に文化資源について用いたのは、当時、東京大学に在籍していた^{つきおとしお}月尾嘉男氏です。1990年代半ば、月尾氏を中心とした「デジタルアーカイブ構想」が出されました。その中では、以下のように述べられています。「有形・無形の文化資産をデジタル情報のかたちで記録し、その情報をデータベース化して保管し、随時閲覧・鑑賞、情報ネットワークを利用して情報発信」することが目的だったそうです。この目的自体、インターネットが爆発的に普及する以前に提唱されていたのは、まさに画期的といえるでしょう。一方で、この表現からは「デジタルアーカイブ」が持っていた主たる目的が「閲覧・鑑賞・発信」であったことが読み取れます。つまり「アーカイブズ」と「デジタルアーカイブ」はそもそも大きくズレがあったのです。

その後、2003年に出されたe-Japan構想の中でも「デジタルアーカイブ」への言及がなされ、以下のように述べられています。

2005年度までに、放送・出版、映画等のコンテンツや、美術館・博物館、図書館等の所蔵品、Web情報、地域文化、アジア諸国との関係に関わる重要な公文書等について、デジタル化・アーカイブ化を推進し、インターネットを通じて国内外に情報提供が行われるよう必要な措置を講ずる（e-Japan構想2003より）

ここでも、主たる対象は、映画や博物館コンテンツであり、公文書については一部を対象としているのみであることがわかります。なお、この「アジア諸国との関係に関わる重要な公文書」の表現が対象としているのは、2001年に開設されたアジア歴史資料センターの公文書類であろうと考えられます。また、この表現からも、「デジタルアーカイブ」は、情報提供を中心に行われるものであり、いわゆる「アーカイブズ」やその理念である保存を主たる対象としていないということがわかります。つまり、「デジタルアーカイブ」は記録資料よりは文化財や文化コンテンツを発信することを意識した用語と

して用いられてきたという背景があるのです。

また、第1章で触れた情報処理学会・人文科学とコンピュータ研究会が年に一度行うシンポジウム（じんもんこん）でも、中心テーマに「デジタルアーカイブ」を据えていました。2007年までは一貫してその中心テーマに「デジタルアーカイブ」を冠し、2011年には再度「デジタル・アーカイブ再考」をテーマとしています。このシンポジウムでのテーマ設定は、必ずしも内容全体を規定するものではありませんが、当時の研究動向を大まかに読み取ることにはできるでしょう。人文情報学研究が、その黎明期において人文学の研究との学際研究を可能とするためには、基礎データの蓄積が重要であるとの認識のもと、このテーマが据えられてきたと思われます。

この20世紀から21世紀に変わる時期、特に2000年代の前半はまさに「デジタルアーカイブの第一次黄金期」ともいえる時代でした。総務省が自治体に向けて大きく旗を振ったこともあり、各地で多数のデジタルアーカイブができあがったのです。

その後、2000年代後半にはこの流れは急速にしぼみ、後述する「第二次黄金期」までの約10年間は比較的低調な時期が続きます。その間に、「アーカイブ」といいながらもなくなってしまったものも多数あります。当時の様子がどのようなものであり、現在までにどのような経緯をたどったか確認することで「デジタルアーカイブ」のあり方を考えるヒントにできればと思います。

以下のようなものが、当時には存在しました。現在、どのような状況であるかも含め、簡単に紹介します（笠羽晴夫『デジタルアーカイブの構築と運用』〈水曜社、2004年〉に掲載されているものなどを中心にピックアップ）。

▶ **石川新情報書府**（当時のドメイン <http://shofu.pref.ishikawa.jp/>）

現在は、石川県産業創出支援機構がサイトを持っています。平成27年度に事業は終了し、当時のドメインからはアクセスできません。HTML・Flash・Quicktimeによるデータがあったようですが、現在はこれらのデータ

は見られなくなっています。

▶ **上田市デジタルアーカイブポータルサイト** (<http://museum.umic.jp/>)

現在も存続しています。「メディアランド上田」という組織の中にあり、Googleによるサイト内検索とHTMLによる紹介などがあります。また、ここからリンクが張られている先には長野県のデジタルアーカイブをまとめた「信州デジくら」が存在し、こちらも継続的にアーカイブを公開しています。

▶ **青森デジタルアーカイブ推進協議会** (<http://www.acci.or.jp/adaa/>)

推進協議会自体のページは存在がなくなっています。ただし、青森県商工会議所のサーバの中にあるため、ドメインは存続しています。Webのスナップショットを保存するサービスを展開するInternet Archiveによる限りは、サイトは2007年8月までは存続していたようです。なお、このページはあくまでも推進協議会の紹介サイトであり、内部に歴史資料のコンテンツはなかったようです。

▶ **山梨県地域資料デジタル化研究会** (<http://www.digi-ken.org/>)

現在もNPO法人として精力的に活動中です。HTMLとFlickrを活用して地域資料のレスキューの様子をデジタル化するなど、さまざまな動きが進められています。

▶ **デジタルアーカイブやまぐち** (<http://www.dayi.or.jp/>)

この事業自体は、特定非営利活動法人「歴史の町山口を甦よみがえらせる会」と合併し、解散しています。ドメインは維持。Internet Archiveで確認する限り、HTMLに加え、wmv形式での方言の記録に関する動画があったようですが、現在はWeb上では存在が確認できません。

▶ **京都市デジタルアーカイブ研究センター** (<http://www.kyoto-archives.gr.jp/>)

2004年に解散しています。ドメインも維持できていない状態です。なお、メインのコンテンツであった、二条城の障壁画については、特定非営利法人・京都文化協会 (<http://archives.kyo-bunka.or.jp/>) で現在も購入することができます(長辺12000ピクセルで6万円)。しかし、これ以外のコンテンツについては現

時点では不明です。

▶ **山形県デジタルコンテンツ利用促進協議会** (<http://www.archive.gr.jp/>)

継続的に活動が行われています。HTML による写真データと、ram 形式のビデオファイルが現在でも閲覧可能になっています。

▶ **東北デジタルアーカイブ研究会** (<http://www.digital-museum.gr.jp/>)

名称を変えて、NPO 法人地域文化アーカイブスとして現在に続いています。支部の数を広げて、東京・山形・島根に支部を持っています。動画などの情報を持つとともに、この研究会が運営を行っている「地域文化資産ポータル」(<http://bunkashisan.ne.jp/>)は、検索システムを持った動画閲覧のサービスを提供している状況です。

▶ **愛知デジタルアーカイブ推進協議会** (<http://www.adaa.gr.jp/>)

2006年に解散し、ドメインは存続していません。デジタルコンテンツの中身は確認できない状況です。

▶ **デジタルアーカイブ推進協議会 (JDAA)** (<http://www.jdaa.gr.jp/>)

2005年に解散しています。ドメインも維持されず、自動車保険の見直しサイトになっており、存在を直接的に確認する方法がなくなっています。

なお、その後継団体とされている、デジタルコンテンツ協会もドメインが変更されており、過去のドメインは広告サイトで運用されています (<http://www.dcaj.org/>)。なお、変更後のドメインでは、現在も活動が確認できます (<http://www.dcaj.or.jp/>)。ただし、当時、どのようなアーカイブの活動を行っていたのか、残念ながらはっきりとした記録は残されていません。

▶ **デジタル・アーカイブ・アライアンス (DAJA)** (<http://www.daja.gr.jp/>)

ドメインなどは残されておらず、詳細は不明です。現時点ではその組織があったということだけは確認できます。

▶ **地域資料デジタル・アーカイブ化協議会** (<http://www.gijodai.ac.jp/circ/index.html>)

岐阜女子大の中で運用されている協議会でした。現在はサイトでの存在は確認されません。なお、岐阜女子大学の中にはデジタルアーカイブ研究所も

存在し、現在もデジタルアーカイブの教育を強力に推し進めているところのひとつです。後述する「現在のデジタルアーカイブ」でも大きな役割を果たしているといえるでしょう。

以上、12の「デジタルアーカイブ」について、現状を見てきました。現在も精力的に活動しているところも、残っていないところもあり、さまざまではありますが、「アーカイブ」が現在まで続くことの課題が少し見えてくるのではないのでしょうか。このような「第一次黄金期」を踏まえ、現在の状況を見てみたいと思います。

4. 「デジタルアーカイブ」の現在

「デジタルアーカイブ」に関しては、『アーカイブ立国宣言』（ポット出版、2014年）の出版がまさに大きなひとつの出発点であるといえます。これ以来、2019年現在に至るまで、「デジタルアーカイブ」は各地で大きな動きとなりつつあり、まさに「第二次黄金期」が来ているといえるのではないのでしょうか。

現在の「デジタルアーカイブ」を改めて推し進めるきっかけには東日本大震災の存在があげられます。東日本大震災で失われた多くの資料や文化財を未来に継承するために、デジタル技術をもっと用いるべきではないのか、といった問題意識から、「現在のデジタルアーカイブ」はスタートしています。その点では、「第一次黄金期」のころとスタートの問題意識は異なるといえるかもしれません。

この大きな潮流のうち、学術的な流れは、第1章で触れた通り、現在「デジタルアーカイブ学会」としてひとつのかたちになっています。これ以外にも、日本の政府や国会を含めた大きな動きになっていますので、まずはその状況を把握するようにしたいと思います。

まず、2003年のものと同様、政府の大方針を見てみましょう。

③ 文化芸術立国の実現

「文化芸術推進基本計画」や「文化経済戦略」に基づき、2020年までを

文化政策推進重点期間と位置づけ、文化による国家ブランド戦略の構築や稼ぐ文化への展開、文化芸術産業の育成などにより文化産業の経済規模（文化 GDP）の拡大を図るとともに、文化財の高精細レプリカや VR 作成など文化分野における民間資金・先端技術の活用を推進する。（中略）文化資源について、各分野のデジタルアーカイブ化を進めるとともに、内外の利用者が活用しやすい統合ポータル¹の構築を推進する。また、インターネット上の海賊版サイトに対して、あらゆる手段の対策を強化する。また、我が国の誇るマンガ、アニメ及びゲーム等のメディア芸術の情報拠点等の整備について指定法人による取組を促進する。（「経済財政運営と改革の基本方針 2018」より）

e-Japan 計画と比べてみて、皆さんはどのように考えるでしょうか。

次に関連する動きについて、説明をします。

ひとつは国会議員が中心となって進めている「デジタル文化遺産推進議員連盟」です。自民党衆議院議員の古屋圭司^{ふるやけいじ}氏を会長とした与野党を超えた議員連盟で、この教科書を書いている現在では「デジタルアーカイブ推進基本法」の提出などに向けた動きを進めています。企業に関しては、デジタルアーカイブ推進コンソーシアム（DAPCON）という組織が存在します。DAPCON は、NTT データ、大日本印刷などを中心とする企業による「デジタルアーカイブ推進」のためのグループとなっています。また、これ以外にもデジタルアーカイブ機関連絡協議会（DARA）という、デジタルデータを保有する研究機関を中心とする組織も存在します。

また、特にデジタルアーカイブを推進するために、内閣府知的財産戦略本部の中に「デジタルアーカイブジャパン推進委員会」が作られており、この下部には実務者委員会が作られています。この実務者委員会では、関係する組織などを中心として、特に後述する「ジャパンサーチ」に関する実務的なデータの集積方法の検討や、合わせてデータの流通手法の工夫などに関する議論が行われています。現在、この実務者委員会のもとで「中間まとめ」が

出されており、また、これに合わせて「デジタルアーカイブ機関評価のためのガイドライン」が公開されています。これらの動きの中で、最も大きなシステムとして構築されているのは「ジャパンサーチ」でしょう。

「ジャパンサーチ」は、国会図書館が現在開発している日本の文化資源情報をまとめて検索し、公開しようというものです。ヨーロッパの文化資源統合ポータルである Europeana (<https://www.europeana.eu/portal/en>) を当初はお手本としていたため「日本版ヨロピアーナ構想」とも呼ばれました。「ジャパンサーチ」の基本モデルは、日本全国の「アーカイブ機関」を分野・地域ごとの「つなぎ役」が仲介して「ジャパンサーチ」にデータを提供、一元的に公開することを目指しています。「ジャパンサーチ」の対象は、歴史的なものにとどまらず、漫画やアニメ・ゲーム・放送などもその範囲とし、広く日本の文化資源全体を公開しています。まさに現在進行形のプロジェクトで、2019年2月27日にベータ版が公開されました (<https://jpsearch.go.jp/>)。

「ジャパンサーチ」と「第一次黄金期」の「デジタルアーカイブ」の大きなシステム上の違いは、バラバラで出すのではなく、総体として見せようという動きが強まったことではないでしょうか。e-Japan 構想の中で見られた「放送・出版、映画等のコンテンツや、美術館・博物館、図書館等の所蔵品、Web 情報、地域文化、アジア諸国との関係に関わる重要な公文書等」については、その方向性は大きく変わらないかもしれませんが、これらを統合的に見せる動きになってきているのは重要なことでしょう。

そして、もうひとつ極めて大きな動きは「オープン化」です。つまりオープンデータの傾向が強まったことで、文化資源のデジタルデータもオープンなものとして出していこうという動きが、この「第二次黄金期」の中では極めて重要なモチベーションになっています。オープンデータ、つまり自由に複製・加工し、機械的な再利用などを文化財の情報についても自由に行えるようにしていこう、という動きがあります。特に、Creative Commons (CC) を用いたデータ公開が進められています。現在の「デジタルアーカイブジャ

パン実務者検討委員会」では、CC0（すべての権利放棄）や CC BY（著作者の氏名を表示することで、複製・加工などを許諾）などを特に推奨し、一部のものについては Right Statement.org を応用する方向で議論が進められており、近くまとまった推奨版が出されるようです。

この文化資源情報のオープン化は、歴史情報学にとっても極めて重要なことです。コンピュータで歴史資料を解析する場合に、権利などで複製や加工が認められないとなると、その研究自体が困難になることも想定されます（2019年1月施行の改正著作権法でだいぶ利用が便利になりましたが）。しかし、オープンデータであれば、それらを気にせずに進めることができます。オープン化の流れは今後も促進していくと思われます。オープン化によって、資料へのアクセス増加が見込まれ、資料を提供する側もメリットが大きいことが最近浸透してきました。一方で、オープン化の懸念であった「悪用される」といった事例は、現時点でも出てきておらず、現在はメリットのほうが大きくなっているといえるでしょう。

5. 「デジタルアーカイブ」の持続性

このように大変に活発な動きのある「第二次黄金期」ですが、課題がないわけではありません。それは「持続性」です。第一次デジタルアーカイブに対して、第二次デジタルアーカイブは、保存を課題としていました。しかし、その「保存」に関してはまだ手探りな部分が多いのが実情です。最後に、その持続性について「アーカイブ」とはどのようなものかを考えておければと思います。

特に「第一次黄金期」の結果から見えることは以下の通りです。

1. 当時の「デジタルアーカイブ」が、「アーカイブズ」や博物館の要請という具体的な課題に応えたものであるというケースは必ずしも多くなく、むしろ補助金を中心とした、伝統産業振興策の一環として位置づけられたものが多かったようです。これは、政策全体がそのようなものであったこと

は否めませんが、その中で、いわゆる GLAM や大学などが関与しているものの数は多くありません。ただし、それらが関わっている「デジタルアーカイブ」のほうが多少残りやすい傾向はあります。

2. 政府系であったり自治系であることが、必ずしも安定した「アーカイブ」になることを意味しないのも特徴です。特に、当時の「デジタルアーカイブ」の中でも、大きな予算を確保したであろう組織からなくなっているようにも見受けられます。一方、NPO などで着実に進めているところは残っており、データも存在しているものが多い傾向があります。公的機関がデジタル系の仕組みを運用することが、必ずしも長期持続を意味しないことは先行研究もありますが、「デジタルアーカイブ」も同様の傾向であるといえるでしょう。
3. 当時のものは動画が多いです。多くの第一次デジタルアーカイブでは動画を運用しており、古い物品類の画像やメタデータに興味に向いている例が多くないことがわかります。しかし、動画の形式が安定していないこともあってか、現在の環境では閲覧が困難になり、そのままなくなってしまったものも少なくないようです。また、動画に適切なメタデータがついておらず、移転してしまった結果、見つけられないものもありそうです。
4. ドメインの維持と、コンテンツの維持には密接な関係があるようです。サーバを変え、データを移行しても、その際に適切なポインタを示せるかどうか重要になっています。

このように考えると、「デジタルアーカイブ」を「アーカイブ」として保存するための考え方はいくつか見えてくるようです。

A. URL ないし適切なランディングページの維持

とりわけ、URL をいかに維持するかは、重要な課題となります。最低限、そこから移行するにせよ、ポインタを示すことが重要であり、それにより残されるということは、十分に考えられる必要があります。その点において、具体的にはデータへの DOI (Digital Object Identifier) の付与は重要な解決方法

になり得るでしょう。これ以外にも、なるべく時限的なドメインは使わない、もしくは適切な移行方法を確保するなどが求められる手法となります。最近ではドメインその物をハッシュ関数による計算値で持ってしまうおうといった技術的検討も見られます。

B. データ形式の整理

2000年代前半に比べると、デジタルデータの中で長期的に維持できるものと、そうでないものの傾向がある程度見えてきているのも事実です。少なくとも Web 上での資料公開が進む段階で、以下のファイル形式については、おおむね下位互換を維持しつつ、長期的に利用可能でしょう。

- ・テキスト（もしくはXMLなど）
- ・PDF
- ・TIFF などの非圧縮の画像ファイル（解像度などの問題はなお残る）

これらのデータ形式は、細かいバージョンや文字コードなどの変容を含みつつも、おおむねデータだけのマイグレーションを施すことで、維持が可能なものとなっています。一方で、以下のようなデータ形式は、まだ安定しているとはいいがたく、長期的保存という観点からは、なお慎重な立場をとるべき部分であるといえるのではないのでしょうか。

- ・3D のデータ形式
- ・動画
- ・いわゆる「高画質」の画像（画質そのものが、時間が進むにつれて陳腐化する）

これらのデータそのものが持つ、時間的特性に即して、データを作成する必要があります。動画は、多くのデータが「残っているような残っていないような」状況になっています。また、画像もサムネイルなどがある場合にも、現在の画面解像度でどこまでの情報があるかという、やはり厳しいのではないかといわざるを得ない部分があります。また、これらのデータ形式については「デファクトスタンダードがどれになるか」といった情報を常に得おくことが必要でしょう。

C. オープンデータはどれほど持続性の解として持ち得るか

第二次デジタルアーカイブの重要な成果として「オープンデータ」をあげました。そのメリットは極めて大きなものです。オープンデータは持続性という点も見べき点があるのです。それは、多くの人が複製を持つ可能性があるということで、ある「デジタルアーカイブ」がなくなっても、その複製を誰かが持っていればデータそのものは生き残る、ということです。では、持続性はこれで解決するのでしょうか。Aのように適切な URL を失い、ポインタのないデータだけが残ることで、どこまでそれを持続的といえるかは難しい部分があります。また、データ保存という観点では消極的な解決手段（誰かが持っていてくれるかもしれない）ともいえます。データはあればよい、はひとつの重要な解ですが、その際にはやはり一定のコンテキストも保存する必要があります。そうです。

D. データとシステムの分離の重要性

長期的なデジタルアーカイブシステムとして維持する場合には、データとシステムをいかに分離して持つかも、重要な課題となります。とりわけ、画像や目録などのデータは、システムの平均的な更新期間である5年より長いことは、おおむね了解されるでしょう。その際にシステムに依存したデータを作ることで、システムの陳腐化とともに、データそのものが陳腐化してしまい、結果的にデータがアーカイブされないという事態が起こりえます。そのため、システムとデータとは分離すべきであり、かつ長期的に保存するためには、データ形式を長く持たせるように、同時に、形式をよりデファクトスタンダードに近いものとしておくことで、マイグレーションを容易にしておくことが肝要になります。

これらの解決のためには、OAISモデルというデータの持ち方が提唱されています。データを原データと提供データに分け、原データを保存し、マイグレーション可能にすることで、長期的な維持を目指しています。

6. デジタル「アーカイブ」を考えるために

このようなデータ保存については、iPres という学会があり、そこではデジタル保存（Preservation）について、多くの研究が行われています。ここでも触れられていることは、データを保存するためには技術だけではなく、総合的な運用モデルが必要である、ということです。システムだけではなくデータなどのスタンダード、データ形式やその持ち方、場合によっては予算などに至るまでの総合的な運用が求められるとともに、そこには関わる「人」自体も重要な要素となっています。歴史情報学という中で、歴史資料のデータを長く維持し、未来の人にも開かれたデータとしていくためには、技術を知った上で、さまざまな要素を学ぶ必要があります。