## デジタルアーカイブへの期待と課題

## ―コミュニティの違いを越えた知的資源の保存に向けて―

筑波大学大学院図書館情報メディア研究科教授

杉本 重雄 すぎもと・しげお

#### 1. はじめに

我々の生活を支える情報基盤は急速に進化し続けている。情報基盤というとネットワークやモバイル端末のような情報機器を思い浮かべる方が多いかもしれないが、使えるコンテンツがなければ情報基盤とは呼べない。また、情報基盤には、コンテンツの流通機能(フロー)のみならず蓄積機能(ストック)を持つことが求められる。本稿で述べるデジタルアーカイブ(Digital Archive)は、情報基盤が持つストック機能の重要な要素であり、文化的、学術的資源を蓄積、提供することで現在の様々な知的営みを支えると同時に、それらを将来に残し、伝えていくための重要な役割を持つものである。

デジタルアーカイブの基本的な価値は、いつでも、どこからでも貴重な資料に簡単にアクセスできることである。90年代に開発がはじまったが、実際にその価値が理解されるまでには、それなりの時間がかかったように思う。デジタルアーカイブの価値が認められるようになったのは、これまでの継続的な努力によって大量の資料が蓄積され、そして、インターネットが我々の生活基盤となったことによって、いつでもどこからでもいろいろなコンテンツにアクセスできることから生まれる価値を実感できるようになったことが大きな要因であると思われる。

従来のデジタルアーカイブは、文化財を電子化して、蓄積提供するものというイメージが強かった。しかしながら、デジカメやワープロ、メールをはじめとして、我々が普段用いる道具は既にデジタル化しており、今後の情報資源の多くが最初

からデジタル形式で作られることは明らかである。 たとえば、先の東日本大震災では、ビデオやデジカメ等で撮られた映像、災害発生時とそれに続く 時期にインターネットに送りだされたニュースや 映像、ブログなど様々なデジタル資料がある。非 デジタル資料も含めて、震災資料を収集・蓄積し、 将来にわたって長く災害の経験を伝えるには、デ ジタルアーカイブ化が有力な方法である。

筆者は、1990年代からのデジタルライブラリ (Digital Library) に関わる研究活動に携わってきた。以下本稿では、これまでの経験を通して、筆者自身がデジタルアーカイブについて学び、考えたことをいくつかの視点から述べる。そのため、具体的なデジタルアーカイブについては述べない。デジタルアーカイブの概念や開発の歴史、実例に関しては笠羽 [1] が詳しい。また、国立国会図書館「デジタル情報資源ラウンドテーブル」[2] や総務省「知のデジタルアーカイブに関する研究会」(知デジ研) [3] の議論もいろいろな視点を与えてくれる。

#### 2.「デジタルアーカイブ」ということば

はじめに、デジタルアーカイブということばを 敢えて定義しておきたい。デジタルアーカイブと は、有形無形の資源をデジタル化して作ったデジ タル資源やもともとデジタル形式で作られた資源 を収集・蓄積・保存・提供するサービス、あるい はそのサービスのために組織化されたデータの集 合体のことである。そこでのキーは、デジタル情 報技術を用いて表現した価値ある知的資源を選択、 収集、蓄積、保存提供することである。

アーカイブ (Archive) と言うことばを辞書で引

くと、歴史文書の集まりもしくはその保管所とい う意味と、コンピュータ上の複数のファイルを保 存目的でまとめて作ったファイルという意味が出 てくる。後者はシステム用語の意味合いが強いの で、ここでは前者だけを考えることにする。上の デジタルアーカイブの定義は、前者の意味を拡張 し、文書に限定せず価値ある文化的、学術的資源 をデジタル形式で残すものを意味している。歴史 文書を扱う専門家であるアーキビストの方々には、 アーカイブという語の意味を広げすぎであると言 われるかもしれないが、現在の我が国において、 デジタルアーカイブは文書に限定したものではな い。たとえば、明治期以降の図書を電子化して提 供する国立国会図書館の近代デジタルライブラリ、 明治維新から太平洋戦争までの歴史公文書を提供 するアジア歴史資料センターの電子公文書データ ベース、アンコール遺跡や故宮他を高品位デジタ ル化して作成したコンテンツ、これらはすべてデ ジタルアーカイブである。

我が国ではデジタルアーカイブということばが 広く受け入れられている。お隣の台湾では、 National Digital Archive Program (數位典藏) や Taiwan e-Learning and Digital Archive Program(數位典藏與數位學習)が進められてい る。ところが、欧米では必ずしもそうではない。 国立国会図書館件名標目表 (NDLSH) や日本語の Wikipedia ではデジタルアーカイブの項を見つけ られるが、英語の場合、米国議会図書館件名標目 表(LCSH)や英語の Wikipedia 上では一人前の項 とはなっていないようである。何年か前、アメリ カの研究者から、デジタルキュレーション(Digital Curation)ということばを聞いた。イギリスでは、 Digital Curation Centre (DCC) という公的組織 が作られている。DCC のサイトでは、デジタルキュ レーションを、「価値ある情報資源を、収集、蓄積、 保存し、そして長期にわたって提供し続けること によって新たな価値を生み出すこと」と説明して いる。保存対象となる資源には、文化的貴重資料 だけではなく自然科学領域等のデータも含まれる。 キュレーションということばは、多種多様かつ大 量の情報あるいは資源の中から、真に価値のある ものを見つけ出し、適切に組織化、解説すること の意味でも用いられる。筆者は、それだけではな く価値ある資源を長く保存し、利用に供すること も重要であるととらえている。デジタルキュレー ションの方が、デジタルアーカイブより少し広い 概念のようにも思えるが、本質的な意味は同じで あると考えている。

#### 3. 公文書館とデジタルアーカイブ

#### 3.1 公文書館にとってのデジタルアーカイブ

公文書管理法では、公的活動の記録として残すべき文書であれば、紙か電子かといった区別はなされない。電子政府や電子自治体の推進に伴う日常業務での電子文書化が進んでいけば、おのずと電子的記録文書の割合は高まっていくことになる。公文書管理法では、レコードスケジュールに基づく公文書の管理(移管、廃棄)を行うことを定めている。長い場合には、原課において30年間文書を保管するケースがあるが、これまでの現場では、電子文書の長期にわたる保管・保存の経験があるとは言えない。公文書館に移管されるまでの期間も、デジタルアーカイブとしてきちんと管理する機能が求められており、現場と公文書館の協力が求められている。

「もの」としての価値がある文書を除くと、公文書の保存は、基本的に記録文書の中身の保存であると言える。記録文書として求められる真正性、完全性を保つことができるのであれば、ファイル形式は必ずしも原本のままである必要はない。安定した電子文書ファイル形式での保存にはファイル形式の変換を必要とする場合があるが、変換によって何らかの情報が失われることもある。そのため、どのような変換が許容されるのかについてのコンセンサスを得ること、変換のためのガイドラインが求められる。

国立公文書館では、アジア歴史資料センターや デジタルアーカイブシステムでの豊富な経験を 持っている。その一方、電子文書の受け入れはま だ始まったばかりであり、これからいろいろな知 見が蓄積されていくことになると思われる。大規模な公文書の蓄積の経験を持つのは、我が国では 国立公文書館のみであり、蓄積された知見の共有 が今後の課題となっていくと思われる。

#### 3.2 図書館や博物館・美術館との連携 (MLA 連携)

公文書館には、重要な歴史公文書を専門家のみならず、一般人や学生・生徒にも利用しやすく提供することが求められる。この視点からは、デジタルアーカイブの重要性は理解しやすい。ただし、資料を単に探して、閲覧できればよいというだけではなく、さまざまな形での資料の利用支援や利用促進が必要である。それには、公文書館の外のコミュニティとの連携が重要である。

博物館・美術館(Museum)、図書館(Library)、公文書館(Archives)は、いずれも価値ある文化資産を将来に残し、利用者に提供するという共通の役割を持つ一方、所蔵する資料の相違に起因するさまざまな違いがある。デジタルアーカイブの場合は、MLAのいずれにとっても、貴重な資料へのアクセス性を高めるという目的が共有できるので、MLA連携によるメリットが期待できる。そうしたMLA連携の大規模例としてEUで進められるEuropeanaがある。

MLA 連携を進めるには、多くの人に連携の重要 性を理解してもらわねばならない。それには、 MLA 連携による、より魅力的なコンテンツやサー ビスの創出が必要である。国立国会図書館による デジタル情報資源ラウンドテーブルはそうした課 題を議論してきた。また、今年7月22日に開催さ れたシンポジウム「文化情報の整備と活用~デジ タル文化財が果たす役割と未来像」(主催:一般財 団法人デジタル文化財創出機構)での吉見俊哉東 京大学副学長による、「MLAに加えて大学 (University) と産業 (Industry) を加えた MALUI 連携を」という視点は、デジタルアーカイ ブの魅力の向上のための重要な視点であると思う。 筆者は、これに加えて利用者、特に学校での利用 者の視点を大事にした開発が重要であると考えて いる。学習利用に関しては、検索方法等の技術的 側面だけでなく、著作権等に関する権利管理の問題に関するサポートも必要である。こうした要求を満たすには、MLAが外の世界との連携を進めることが必要であろう。

#### 4. デジタルアーカイブの基盤的課題

### 4.1 デジタルアーカイブの長期利用とコンテンツの 長期保存

デジタル資料の長期利用・長期保存は重要な課題であると同時に、情報技術の進化の速さゆえに難しい課題であることもよく知られている。「デジタルアーカイブに求められるコンテンツの保存期間は何年か?」はよくある質問である。デジタル化とは、基本的に中身(コンテンツ)の入れ物(蓄積媒体)からの分離である。したがって、入れ物のための適切な情報技術選びの側面を持つと同時に、効率的かつ適切な複製技術選びの側面を持つと同時に、効率的かつ適切な複製技術選びの側面を持つと同時に、効率的かつ適切な複製技術選びの側面を持つとは難しい。そのため、筆者は、デジタル資料の長期保存とは、保存すべき資料の内容(コンテンツ)を、適切な形で次世代につなぐことと考えている。

デジタルアーカイブの長期利用は、単に技術的な問題ではなく、総合的な問題としてとらえるべきである。筆者はおおざっぱに、次の3つの視点でとらえている。

- (1)組織の視点:組織は永遠の命を持つわけではなく、組織の消滅、統廃合は必ず起きる。そうした組織とその周りのコミュニティの変化に対応できるようにしなければならない。
- (2)デジタルアーカイブとそのコンテンツに関する情報 (メタデータ) の視点:保存されているコンテンツに関する情報が失われるとコンテンツを再生できなくなったり、コンテンツの真正性や正統性が失われたりする。したがって、コンテンツに関するメタデータもコンテンツとともに長期保存しなければならない。そして、時がたってもメタデータの内容が理解できるようにメタデータに関する情報も残さねばならない。
- (3) コンテンツの視点: デジタルコンテンツを、オリジナルのまま残すべきか、保存しやすいと考

えられるフォーマットに変換して残すのかと いったことを決めねばならない。

#### 4.2 デジタルアーカイブの裾野の拡大

デジタルアーカイブの構築を進め、その裾野を広げていくことが重要である。その一方、それなりのコストがかかるのでそう簡単に開発を進められるわけでもない。裾野を広げるには、高機能・高品位のデジタルアーカイブを作り、全体の頂を高くすることも必要であろう。重要な文化財の高品位のデジタル化にはそうした役割が期待される。その一方、デジタルアーカイブの価値は、コンテンツの量によっても決まるので、それなりのコストで作ることのできるデジタルコンテンツを地道に増やしていくということも求められる。

地道にコンテンツを増やす仕事は、地域に根差す中小のMLA機関にも求められる役割である。そうした機関にとって、デジタルアーカイブ化に関する指針を与えられることは重要であろう。総務省の知デジ研におけるデジタルアーカイブ作りやコンテンツの長期保存に関する議論や、国立国会図書館、国立公文書館等でのこれまでの取り組みに基づく報告書などは、デジタルアーカイブ化を進める上での有用な情報源になると思う。

# 4.3 情報基盤としてのデジタルアーカイブへのアプローチ

情報基盤としてのデジタルアーカイブの理想形は、いつでも、どこからでも、誰でも、任意の文化学術資源にアクセスし、閲覧できることである。しかしながら、現実には理想の実現は困難である。理想に近づくには、電子化を地道に続けるということも必要であるが、それ以上に重要なこととして、資源に関する情報に、いつでも、どこからでも、誰でもアクセスできるようにすることである。すなわち、所蔵資源に関する「は多いかもしれない。しかし、たとえ紙の台帳やパソコンの文書の形であったとしても、所蔵資源に関するメタデー

タは作られているはずである。メタデータを電子 的に作り保存することによって情報の利用性はず いぶん高まる。それをインターネット上でアクセ スできるようにすることでそれ自体の利用性が高 まるだけではなく、第三者の力を借りて価値を付 加することも可能になる。

相互運用性の確保、特に異なるコミュニティ間での相互運用性の確保は難しい課題である。基本的に、原資料のデジタル化の方法、データベースの構成方法、メタデータの構造等はデジタルアーカイブごとに異なる。これらを無理に統一するようなことはできない。その一方で、いろいろなデジタルアーカイブを横断的に使えるようにすることや、第三者によるデジタルアーカイブの付加価値サービスを可能にすることも重要である。こうしたとは、特に目新しいものではないが、なかなか実現されていないことでもある。こうした課題の解決には、標準規格の利用の促進に加えて、デジタルアーカイブに関わる種々のメタデータに関する情報の共有が必要不可欠である。

#### 4.4 デジタルアーカイブ基盤の高度化

クラウドと言うことばをよく耳にする。デジタルアーカイブが多様であるとはいっても、基盤部分になるほど要求される機能は共通化してくると考えられる。筆者は、デジタルアーカイブ間に共通の機能とそれぞれに固有の機能をうまく階層化することでクラウドの利用が進められると期待している。クラウドを利用することで総合的なコストを下げること、特にMLA機関では常に問題になるIT技術に詳しい人材に関する課題の解決となることが期待される。

従来のデジタルアーカイブでは、デジタルコンテンツの充実が中心的な課題であったが、それに加えて提供する情報資源の意味内容に応じたサービスの高度化も重要である。それには、デジタルコンテンツに関する情報を表す種々のメタデータを使い回せるようにする必要がある。この点においては、WWWのLinked Open Data の活動等から生み出されるメタデータ技術を利用して、より

高度なコンテンツアクセスを実現するためのメタ データ情報基盤を作ることが必要であろう。

#### 5. おわりに

東日本大震災では、文化財や紙文書のみならず、 サーバに格納された文化財に関するデータや記録 文書も被害を受けた。文化財を将来に残していく には、文化財そのものだけではなく文化財に関す る情報もきちんと残していかねばならない。そう した情報を失った文化財の価値を取り戻すのは容 易なことではない。

デジタルアーカイブ化は、災害に備えた文化的 資源、公的記録の保存の視点から重要であると思 う。紙であれデジタルデータであれ、物理的な「も の」として一か所にしか存在しないものは、失わ れる危険性を常に持つ。デジタルデータの場合は 複製が作りやすいという強みを持つ。とは言って も、バックアップデータを本体と同じ所に置いた のでは意味がない。バックアップのためのネット ワークが必要となるが、それを個別の機関が連携 しながら広域で実現するのは容易ではない。その ため、個別のデジタルアーカイブを支えるクラウ ド的な情報基盤について、国レベルあるいは複数 の都道府県にまたがる地域レベルで考えることが 重要であろうと思う。また、公文書管理法が、国 の規則を決めるのみならず自治体における公文書 管理の促進も求めていることを考えると、我々の 生活を支える公文書管理のためのクラウド的情報 基盤の重要性はより高まると思う。

Digital Librarian、Digital Archivist、Digital Curator と言ったことばを耳にする。これは、ネットワークを指向したサービスのための人材育成が重要な課題であることを暗示しているのであろう。クラウド化が進めば、情報基盤に関する技術的知

識をクラウド側に任すことができるので、MLA側はよりコンテンツに焦点を当てた仕事ができるようになると期待できる。とは言っても、現状ではなかなかそうはいかないであろうし、組織の大小によって一人がカバーしなければならない範囲が大きく異なることも理解できる。いずれにしても、一人だけで、それなりの規模の所蔵資料の価値を理解し、それを多くの人が利用できるように組織化し、インターネット上で提供することは難しい。そのため、筆者には、デジタルアーカイブの協働的開発に参加できる人材を育成することが必要であると思われる。

最後に、言うまでもないことであるが、デジタルアーカイブは使われるものでなければならない。それには、それなりのボリュームのコンテンツを持つこと、簡単に使えるようにすることが重要である。利用者はアーカイブを使いたいと思っているのではなく、コンテンツを欲しているのであるという視点を忘れてはならない。その意味では、提供者だけではなく多様な利用者も参加して、使えるデジタルアーカイブを作るための努力を続けることが必要である。

#### 参考資料

- [1] 笠羽晴夫. デジタルアーカイブ 基点・手法・ 課題, 水曜社, 204p., 2010.10
- [2] 国立国会図書館・デジタル情報資源ラウンド テーブル、http://www.ndl.go.jp/jp/aboutus/ honkaigi.html (2011年9月4日)
- [3] 知のデジタルアーカイブに関する研究会(デジタル・ネットワーク社会における出版物の利活用の推進に関する懇談会の下部委員会)、http://www.soumu.go.jp/main\_sosiki/kenkyu/shuppan/index.html (2011 年 9 月 4 日)