

## デジタル・アーカイブの現状と問題点

原田隆史

### The overview of the digital archives

Takashi HARADA

#### Abstract

The definition of the "digital archives" is ambiguous; various types of organizations, such as web archives, digital museum, data archives and so on, which collect original material in digital form has been called digital archives. This study gives some concrete examples of digital archives and discusses their good effect and problems.

Although the good effect given by digitalizing materials is acknowledged widely, the evaluation of digital archives is not easy; for instance, the cost effectiveness is difficult to estimate. One of the roles of the digital archives is to preserve cultural assets; most of them have been founded and operated by government or public organizations. The future of the digital archives might depend on whether the value of them would be given social consensus of their cost effectiveness.

#### 1. デジタル・アーカイブの定義

デジタル・アーカイブという言葉は、様々な場面で使用される。しかし、その定義はあいまいで広い範囲におよんでいる。英語でアーカイブ(Archives)は、「古記録,公文書」や「文書館,記録保管所,官文庫」を意味し<sup>1)</sup>、図書館や美術館・博物館とは別の機関と位置づけられている。ここから、デジタル・アーカイブもデジタル化された記録や文書を保存する機関であって、電子図書館やデジタル・ミュージアムとは異なる存在であるという考え方もある<sup>2)</sup>。

しかし、Googleで「デジタル アーカイブ」という検索語で検索を行った時に表示される最初の100サイト中の68%がデジタル・ミュージアムを指し示すものとして使用される(2001年4月20日現在)など、デジタル・アーカイブは単に古記録や公文書をデジタル形式に変換した機関や組織だけではなく、その対象範囲は広くとらえられていることが多い。

ただし、書誌データベースや検索エンジンをさしてデジタル・アーカイブの一種であるとする表現はみられない。すなわち、デジタル・アーカイブは二次資料のデジタル化ではなく、一次資料またはその代替物をデジタル化し、その保存および流通を目的とするものと捉えることができよう。

#### 2. デジタル・アーカイブの分類

デジタル・アーカイブの分類は、さまざまな観点から行うことができるが、ここでは対象としている資料からの分類を行う。対象となる資料は、最初からデジタル形式で作成されデジタル形式で提供することを目的とした資料(Born-Digital Material)と、元来はデジタル形式で提供することは目的としていない資料を対象とする場合がある。後者は現物の代替物の保存であるとも言えよう。また、後者は作成段階でデジタル形式のデータが作成されているか、またはそのようなものが存在しないかで分けることができる。ただし、これらは必ずしも排他的なものではなく、複数の資料をまとめて保存するアーカイブも数多く存在する。それぞれの例を以下にあげる。

(1) デジタル形式で提供することを目的とした資料

例：Web ページ，ソフトウェア，コンピュータグラフィックス(CG)，MP3 配信の音楽

(2) 資料作成の途中過程でデジタル形式のデータが作成される資料

例：印刷データ(図書，雑誌，新聞など)，CAD などのデータ，研究上の調査データ

(3) デジタル形式で作成されない資料

例：美術品，考古学資料，電子組版によらない印刷物，音楽データ，(テレビニュース)

### 3. デジタル・アーカイブ プロジェクト

今日のネットワーク社会の進展とともにデジタルデータの流通も容易になったことからデジタル化を行うための環境は整ってきた。このような変化にともなう、近年、デジタル・アーカイブを作成するプロジェクトは世界中で行われている。第2章で示した分類にもとづき、いくつかの特徴的なプロジェクトを以下に示す。

#### 3.1 デジタル形式で提供することを目的とした資料の収集

もともとデジタル形式で提供することを目的としていた資料の代表的なものとしては、ソフトウェアと Web ページをあげることができる。このうち、ソフトウェアについては古くから Anonymous FTP サイトなどの形で収集・提供が行われてきた。また、近年では(株)ベクター (<http://www.vector.co.jp>)や Impress 社の「窓の社」(<http://www.forest.impress.co.jp/>)などのソフトウェアアーカイブを中心にした企業も増加してきている。

一方、インターネット上の Web ページはその発信の容易さから急速な勢いで増加している。その中には電子ジャーナルのように雑誌論文の代替として使用されるものから、日記や掲示板のような極めて私的な文章まで多岐にわたる。しかし、その絶対量の増加にともなう、確実に研究者にとって有用な情報源となりうる Web サイトの数は増大している。このような Web 上のデータを網羅的に収集するものとして、Google(<http://www.google.com>)や Altavista(<http://www.go.com>)のような検索エンジンがあるが、検索エンジンの目的は保存を目的とするものではない。

Web 上のデータを情報源として使用する場合、Web ページが更新や削除で参照できなくなってしまうという問題があり、これは検索エンジンでは解決されない。近年、このような問題に対応することを目的として、Web ページを網羅的に収集するプロジェクトが注目を集めている。その代表的な存在としては The Internet Archive(<http://www.archive.org>)がある。

The Internet Archive は、WAIS の開発者としても有名な Brewster Kahle が設立した NPO 組織である。ただし、実際には Amazon.com の 100%子会社である Alexa International 社が実質的な作業の一切を行っており、The Internet Archive は Web データを無償提供されて保持しているにすぎない。ロボットを使って現在までに 60T バイト(2001 年 2 月現在)のデータが収集されており、さらに 1ヶ月ごとに 10T バイト程度の更新データを収集している。収集されたデータは URL から検索することができるようになっている。ただし、検索することができるのは URL のみからであり、更新日付からの検索はできない。したがって、各ページの最新(または削除時)の内容は表示することができるが、更新前のページ内容は参照することができない(更新前の古いページも保存はしている)。なお現在では、The Internet Archive は世界中の全ての Web データだけでなく、映画や音楽をそれぞれ MPEG2, MP3 形式に変換して収集・配信する活動も行っている。

また、同様の活動をしているプロジェクトとしては、スウェーデン王立図書館の The Royal Swedish Library Kulturarw3 Project(<http://kulturarw3.kb.se/html/webbarcheng.html>)がある。スウェーデンで作成された Web データの全てを収集するプロジェクトで、スウェーデン国内のサイトだけではなく、\*.com, \*.net のドメイン名であっても、その所有者がスウェーデンである場合には収集対象としている。

さらに、オーストラリアの Web ページを収集しようとする試みであるオーストラリア国立図書館の PANDORA プロジェクト(<http://www.nla.gov.au/pandora/>)もある。ただし、PANDORA プロジェクトではオーストラリアの Web ページすべての収集を行おうとするのではなく、選考委員会が決定した「有用な情報源となりえる Web ページ」だけを収集している。PANDORA プロジェクトと同様の手法をとるものとしては、日本の農林水産研究情報センターが提供する「サーチエンジン農林 1 号(仮称)」(<http://seekar.affrc.go.jp/seek.html>)がある。

### 3.2 資料作成の途中過程でデジタル形式のデータが作成される資料

現在、世界中の多くの出版社において印刷に電子組版システムが使用されている。電子組版システムでは、デジタル形式で作成された文字データを元に印刷が行われており、これを利用して容易にデジタル・アーカイブを構築することが可能である。近年、多くの学会ではその発行する学術雑誌や年次大会の報告を Web 上でも閲覧できる仕組みを提供しているところが多くなっている。たとえば、アメリカ化学会(<http://www.acs.org>)、アメリカコンピュータ学会(<http://www.acm.org>)などは大規模な Web サイトを持っている。また、新聞社や通信社では新聞の作成過程で得られる組版データを元に新聞記事の全文を Web サイト上で公開している。朝日新聞(<http://www.asahi.com>)や日経新聞(<http://www.nikkei.co.jp>)、CNN(<http://www.cnn.com>)などはその例である。

また、近年注目されているものとしては、研究者の研究過程で得られる調査データなどを公開し、多くの研究者間で共有しようとする試みである。その初期の例としては、収集した調査データを公開したジョージア工科大学のインターネットユーザ調査(GVU's WWW User Surveys)がある。ここでは、調査の集計結果だけでなく調査データファイルそのものが Web 上でダウンロードできるようになっている。また、日本においても郵政研究所が平成 11 年 1 月に実施した「インターネットの利用に関する調査」の調査データを公開している。

さらに、データを選別・分類し、利用しやすいようにアーカイブしている機関としてミシガン大学の ICPSR(<http://www.icpsr.umich.edu>)があげられる。研究者は、自分の分析に使用したデータを ICPSR に提供し、その旨を論文に記載する。他の研究者は ICPSR からデータをダウンロードして、追試やより発展的な分析を行うというものである。大学単位で ICPSR に加入することによって、他大学に所属する研究者でも、ここに収録されたデータを使用することができる。ただし、日本の大学で ICPSR に参加しているのは極めて少数である。

### 3.3 デジタル形式で作成されない資料

電子組版システム導入以前に作成された図書や雑誌、さらに美術館や博物館で所蔵している文化財や美術品のデジタル化は、米国議会図書館(library of Congress : LC)をはじめとする図書館や各地の博物館で進められている。たとえば、日本ではデジタルアーカイブ推進協議会が文化財や美術品をデジタル化するための産官学の共同プロジェクトとして 1996 年 4 月に設立され、石川県が所

有する美術品・焼き物や染め物などをデジタル化してネットワークでの発信している。また、Bill Gates の出資のもとに Corbis 社(<http://www.orbis.com/>)は世界中の 6500 万枚におよぶ写真を収集し、210 万点以上をオンラインで提供している。また、慶應義塾大学は、グーテンベルグ聖書のデジタル化を皮切りに、各種の資料のデジタル化を行う HUMI プロジェクト(<http://www.humi.keio.ac.jp/>)を発足させ、その成果をあげている。

さらにテレビニュースについても、CNN が 2003 年からのニュース画像のデジタル画像配信を予定しているほか<sup>3)</sup>、現在テープ形式でテレビニュース画像を収集し索引・抄録を Web 上で発信している Vanderbilt University TV News archive(<http://www.vanderbilt.edu/tvnews/>)も将来計画としてデジタル化を検討している。

#### 4. デジタル・アーカイブの問題点

デジタル・アーカイブの必要性は、各種マスメディアや Web ページでもとりあげられている。しかし、The Internet Archives についても多くのデジタル美術館にしても、その利用は必ずしも高くないのが現状である。将来の利用に向けて準備をすることの重要性や広報の不足ということはあるだろうが、それ以外にデジタル・アーカイブ自身にも以下のような多くの問題点がある。

- (1) 特に大量の資料をデジタル化する場合には、その作成に多額の経費がかかる。
- (2) 私企業が提供するデジタル・アーカイブにおけるビジネスモデルが確立していない。したがって、たとえば出版社などが書籍や雑誌をデジタル化して提供した場合にコストが売り上げの増加を越えてしまい経営を圧迫する可能性や、美術館などの資料のデジタル化にともなって入場料収入が減少する可能性などが指摘されている。
- (3) データ・アーカイブなどにおいてデータの提供者に対する報奨制度が確立していない。
- (4) 著作権の侵害の問題がある。特に、デジタル化された資料を収集しようとした場合、その対象となる相手が非常に多く許諾をとるのが困難であることもありえる。

これらの問題はデータの提供者に関わる事項が多く含まれている。デジタル・アーカイブの発展のためにビジネスモデルの作成も含めて今後検討していく必要があるだろう。

#### 5. 参考文献

- 1) プログレッシブ英和辞典 第3版. 小学館 (1997).
- 2) 鶴川義弘. 学術情報のインターネット・アーカイブの必要性.  
<http://www.affrc.go.jp/Cinfo/news/ric/55/55-3.html> [2001/04/10].
- 3) 東京新聞 第7面. 2001年4月25日

---

原田 隆史 慶應義塾大学文学部(〒108-8345 東京都港区三田 2-15-45)

Takashi HARADA (ushi@slis.keio.ac.jp) School of Library and Information Science, Keio University