

## 国立公文書館のデジタルアーカイブ：過去の記録から未来の記憶へ

牟田昌平

国立公文書館

我が国では「アーカイブ」という言葉の本来の意味である「記録資料」や、それを保存公開する施設としての「文書館」について社会的な認知が低い。本来の意味での「アーカイブ」が社会的に認知されている欧米諸国とは異なって、「デジタルアーカイブ」は「データベース化されたデジタルコンテンツ」の意味で理解されている。最近では Google のようなサーチエンジンとさえ混同されるようになってきている。本稿では、一般に馴染みのない「アーカイブ」のデジタル化を進める国立公文書館の試みを通じて、本来の意味での「アーカイブ」をデジタル化することで生まれた可能性や課題について紹介する。

### National Archives Digital Archives: From Past Records to Future Memories

Shohei Muta

National Archives of Japan

In Japan, a word 'archives' is not recognized as its original meanings of 'historical records' or a depository of such records. A word 'digital archives,' therefore, has developed with different meanings in Japan compared to the western countries, where the word 'archives' traditionally means 'historical records' or their depository. In this paper, through National Archives digital archives project, original meanings of 'archives' will be introduced, and problems and possibilities surrounding digitization of archival materials will be discussed focusing on access to the records.

#### はじめに

「デジタルアーカイブ」という言葉が日本で生まれたことはあまり知られていない。IT 産業振興に不可欠な電子化されたコンテンツ、つまり「デジタルコンテンツ」と「集積保管」というコンピューター用語である「アーカイブ」が複合して作られた言葉である。一方、「アーカイブ」とはギリシャ語の「アルケイア」（公的記録）を語源とするように、政府や組織の記録やその保管所を意味する言葉である。福田総理は、2007 年の参議院予算委員会で、公的記録を保存する意味での「アーカイブ」（公文書館）制度が世界的に見て「お粗末というか、わびしい状態」とであると答弁している。公文書館としての「アーカイブ」は日本では社会だけでなく行政内部でもその存在や役割が認知されていないのが現実である。つまり、電子化した図書館や博物館を指す「デジタルライブラリー」や「デジタルミュージアム」とは異なり、我が国では「デジタルアーカイブ」という言葉は「公文書館」とは全く関係なく広まってきた。本稿ではまず、我が国で生まれた「デジタルアーカイブ」という言葉の意味を整理し、日本では馴染みのない本来の意味での「アーカイブ」とは何かについて紹介する。そして、歴史記録のデジタル化がもたらした可能性や課題、さらに電子政府化推進による「ボーンデジタル記録」の爆発的な増大によって大きく変わりつつある「アーカイブ」やその専門家としての「アーキビスト」の機能や役割について、国立公文書館のデジタルアーカイブプロジェクトの経験を踏まえて紹介する。

## 1. 日本発の「デジタルアーカイブ」と海外での理解

1990年代の半ば、マルチメディア工学の専門家である月尾嘉男東京大学教授がある会合で「かつての図書館などの電子版」という意味で「デジタルアーカイブ」を使ったことに始まると言われている。[1] 1994年12月には通産省（当時）の外郭団体（財）マルチメディアソフト振興協会主催で国際会議「世界の文化を未来に継承するデジタルアーカイブ国際会議」が開催され、1996年4月には「人類の文化遺産を未来に継承する」ことを目的としたデジタルアーカイブ推進協議会が発足する。1990年代半ば、電子図書館（デジタルライブラリー）という言葉さえ一般的でない時代に、図書館や博物館の枠を超えて「有形・無形の文化遺産をデジタル情報の形で記録し、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発信」という「デジタルアーカイブ」の概念を官・民・学が一体となって提唱したのである。

一方、国際図書館協会連盟（IFLA）に相当する国際公文書館会議（ICA）が2002年11月にマルセイユで開催した円卓会議（CITRA）でフランス人アーキビスト Marie-Anne CHABIN は「フランスメディアにおける「アーカイブ」という言葉の利用の比較研究」という発表の中で、伝統的な意味での「アーカイブ」という言葉が急速にコンピューター用語としての「集積保管」という意味での理解に置き換わっていると報告した。発表の中で彼女は audiovisual and digital archives と表現しているように、紙の記録に対して「音声や電子記録」という一般的な意味合いで使用している。欧米では次の項で紹介する伝統的な意味での「アーカイブ」（記録資料）が電子化されたものとして特に抵抗無く使用されている。さらに最近では、ポーンデジタル（デジタル記録として作成されたファイルや電子メールなど）の保存管理が重要な課題となっており「デジタルアーカイブ」の意味は変化してきている。[2]

## 2. 本来的な意味での「アーカイブ」とは

我が国では馴染みのない「アーカイブ」という言葉について、2003年に出版された『アーカイブ辞典』ではその表記に当たって「内容の記述においては執筆者によりアーカイブ、アーカイブズ、アーカイブス、文書館、公文書館、史料館、文書、史料、記録・記録史料など呼び方はまちまちであるが、理解がまだ一応でない現状からあえて統一しなかった」との断り書きを付けている。[3] 世界的にも「アーカイブ」という言葉の表記方法や意味についても統一的なものはない。ここでは「アーカイブ」として表記する。

### 「支配や統治の記録」としての「アーカイブ」

「アーカイブ」という言葉は、本来、ギリシャ語の「アルケイア」、つまり「公的記録」に語源を発する。人類は、文明の発祥の時から「支配や統治の記録」、それを保存する場所としての「記録保管所」を持っていたことは古代メソポタミアやエジプトの遺跡から外交文書や徴税の記録等が整理された形で発掘されていることから明らかである。わが国でも既に奈良時代、『大宝律令』（701年）に行政文書の作成、保管についての様式を制定した公式令（くしきりょう）があり正倉院のような施設が設けられていた。古代から洋の東西を問わず高度の発達した統治機能を備えた文明社会には何らかの形で統治の記録を保存する施設（アーカイブ）が存在したのである。[4]

ただし、現在「アーカイブ」と呼ばれる近代公文書館制度の誕生は、人民主権に基づく共和制を打ち立てた18世紀末のフランス革命といわれる。ルイ王政を倒した革命の指導者達はルイ王朝の記録を「国民の共有物」として保存し公開することとし、1789年、革命の年に設立したのが近代公文書館の最初といわれるフランス国立公文書館（Archives Nationales）である。1794年には国立公文書館法が制定され、政府記録の保存義務と市民への記録閲覧権の保障が法律によって保障された。近代公文書館の理念が確立したのである。[5]

### 近代的文書館制度と記録の評価選別

19世紀以降欧米で発達した近代官僚制度は、大量の文書記録を生み出すようになった。出生記録のように永久に保存されなければその個人の存在そのものが永遠に失われるものから鉛筆一本の調達に関する書類のようにある時期が来れば記録としての価値を失うものもある。もし、これらの記録を全て残したとしたら行政は完全に麻痺してしまう。行政の効率化という観点から一般的に95パーセント以上の記録が廃棄される必要があるといわれている。しかし、一旦廃棄した記録は永遠に失われる。一方、民主主義の根幹とも言える行政が主権者に対する説明責任（アカウントビリティ）を果たすためには出来る限り記録を残す必要がある。この矛盾する命題について各国で試行錯誤が行われてきた。そこで生まれたのが「記録の評価選別論」である。ここでは現代評価選別論の元祖といわれるテオドール・シュレンバーグの理論を簡単に紹介する。

シュレンバーグによれば、作成された行政記録は業務のための1次的価値（「業務的価値」であり「財務的価値」、「法務的価値」を含む）と業務とは関係の無い2次的価値（「研究価値」であり「証拠的価値」、「情報的価値」を含む）を持つ。それぞれの記録は、記録保存年限規定（リテンション・スケジュール）にそって作成から5年、20年、30年等で時限保存され最終的に廃棄または公文書館に移管され永久保存されていく。その過程で行われるのが理論に基づく評価選別作業である。例えば、1次的価値の記録といえる鉛筆等の備品調達のような組織に関係なく共通の一般業務に関する記録は比較的短い保存期間で廃棄される。一方、2次的価値でも証拠的価値を持つといえる法律制定に関する記録等は、原案から政策形成に関わる関連記録を一つの固まり（ファイル）として保存し最終的にはファイル単位で公文書館へ移管・公開される。[6]

評価選別に当たって不可欠なのが記録の生成過程を全体として分類体系化することである。そして、残された記録は国際標準記録史料記述：一般原則（General International Standard Archival Description (ISAD(G))）という目録記述様式にそって整理され利用に供される。現在では、ISAD(G)をベースとしてXMLに対応した国際規格の符号化記録史料記述、EAD (Encoded Archival Description) として電子データ化も進んでいる。[7]

### 3. 「アーカイブ」後進国日本と国立公文書館の「デジタルアーカイブ」構想

国立公文書館が「デジタルアーカイブ」を構想する上でそのモデルとしたのが「アジア歴史資料センター」（以下センター：[www.jacar.go.jp](http://www.jacar.go.jp)）である。センターの基本構想を検討した「アジア歴史資料センター（仮称）の設立検討のための有識者会議」は、1995年6月、村山総理（当時）への報告の中で周辺環境の改善、改革という一項をもうけて、海外調査の結果、「痛感させられた」

点として、公文書の移管や開示等の取り扱いが法的未整備、文書記録の専門家としてのアーキビストの人材育成の必要性、公文書を作成する各省庁等の国の機関の歴史記録保存の重要性への理解不足を指摘した。[8]

### 「デジタルアーカイブ」としての「アジア歴史資料センター」開設

アジア歴史資料センターは、「近現代における我が国とアジア近隣諸国等との関係に関わる歴史資料として重要な我が国の公文書その他の記録」をインターネットで「いつでも」「どこでも」「だれもが」「無料」で利用出来る「本格的デジタルアーカイブ」として2001年11月30日に開設された。2007年12月現在、画像約1500万、目録データ約90万を提供する。

なぜ、「本格的」を謳ったかという歴史資料を電子化するという本来的な意味を広める意味で強調した点がある。当初検討されたのは、関係図書や公文書のマイクロフィルムを所蔵しインターネットで目録データベースを公開するセンターであった。それが、第1回有志者会議開催から5年後の1999年11月30日「わが国とアジア近隣諸国との歴史に関し、国が保管する資料について国民一般及び関係諸国民の利用を容易にし、併せて、これらの諸国との相互理解の促進に資することを目的」として「国立公文書館、外務省外交史料館、防衛庁（当時）防衛研究所図書館等の国の機関が保管するアジア歴史資料（近現代における我が国とアジア近隣諸国等との関係に関わる歴史資料として重要な我が国の公文書その他の記録）を電子情報の形で蓄積するデータベースを構築し、インターネット等を通じて情報提供を行う」センターを国立公文書館に開設することが閣議決定されたのである。[9] なぜ、画像のデジタル化まで踏み切れたかという1999年12月に小渕政権が打ち出した国策としての「ミレニアム・プロジェクト」（新しい千年紀プロジェクト）のIT21（情報通信技術21世紀の推進）において2005年度までに「全ての国民が、場所を問わず、超高速のインターネットを自由自在に活用して、自分の望む情報の入手・処理・発信を安全・迅速・簡単に行えるインターネット&コンピューティング環境を創造する」ことが謳われ、それまで不可能であった大容量のデジタル画像をインターネットで提供する基盤整備が行われることが明らかとなったことが理由としてあげられる。[10]

### アジアの基本コンセプト「デジタルであること」(Being Digital)

アジア情報システムの基本コンセプトは「デジタルであること」(Being Digital)である。「アトム」(実体のあるもの)から「ビット」(デジタル情報)へ社会が変化することで生活のあり方そのものが根底から覆されることを予測したニコラス・ネグロポンテは、その著書『ビーイング・デジタル』(1995年)で「デジタルであること」とは「光の速さで瞬時に廉価で電子データが送信される」と定義した。「アトム」の状態であれば、図書とは違いオリジナルは基本的に一つしか存在しない公文書を複数の人が同時に閲覧することは出来ない。しかし、デジタル化しインターネットに接続することで「いつでも」「どこでも」「だれもが」「自由に」閲覧することが可能になる。公文書のデジタル化は公文書利用の最大のネックであったアクセスの問題を解決する糸口を与えた。その一方で、オリジナルの特定が困難になるなど「原本性」の問題が発生する。原本性確保については技術的な対応だけでなく運用上の対応などが考えられる。現在のところ公文書館のデジタルアーカイブではデジタル原本をオフラインで保存し改竄等が発見された時点でデジタル原本に戻す処置を執っている。

さらにネグロポンテは、デジタル化によって技術革新の早さが予想を超え、技術の陳腐化が不可避であることを指摘し、これに対応するためには技術進歩に柔軟に対応できるシステム構成が不可欠であること、特定の技術に頼るのではなく、例えばリナックスのようなオープンで誰もが自由に利用できる「公有」（パブリックドメイン）の技術を採用することを説いた。さらに、インターネットでは既存の著作権が意味をなさなくなっておりコンテンツを自由（free）で利用できることが公共財としてのインターネットの意義であることを指摘した。これらの考えを整理したのがアジ歴システムの「基本機能要件」である『「いつでも」「どこでも」「だれもが」「無料で」自由に利用できる』情報提供システムである。[11]

この「基本機能要件」は、2005年4月に開設された国立公文書館デジタルアーカイブにも採用され「デジタルアーカイブ推進要綱」として公文書館デジタル化の指針となった。[12] 公文書館デジタルアーカイブは、センターの構想段階では時期尚早として採用を見送った検索目録データベースに EAD (Encoded Archival Description 2002: 符号化記録史料記述) に準拠した XML (eXtensible Markup Language) や実験的に採用していた他機関との横断検索を可能とするプロトコル Z39.50 の本格的運用、さらにそのインターネット対応と言われる SRW (Search/Retrieve Web Service) を採用し、国立情報学研究所の総合書誌データベース NACSIS WebCat や国立国会図書館の「デジタルアーカイブポータル」等と接続し横断検索すら可能としている。1次資料を扱う公文書館と2次資料を扱う図書館データベースとの横断検索が実現されたことで歴史研究に新しい可能性が生まれた。一昨年、システムを更新したアジア歴史資料センターのシステムではオープンソース化を一層進めた。

#### 4. デジタル化の最大の目的はアクセス

デジタル化することによって公文書は時間や空間の制約を受けずに利用できるはずである。しかし、特定の公文書にアクセスするためにはいくつかの解決されなければならない課題がある。ここでは「森鷗外」という検索語を事例にその問題点や課題を紹介する。まさにこの点こそ Google 的サーチエンジンと異なる点である。

##### 「森鷗外」では探せない

「森鷗外」で検索するとわずか1件しかヒットしない。それには理由がある。公文書には「森鷗外」ではなく彼の本名である「森林太郎」や「森一等軍医」や「森軍医監」などのように役職を冠した名字で記載されていることが多いためである。センターではこのような問題を克服する手段として「辞書」を作成している。図「森鷗外による検索結果」は「辞書」で検索用語を拡大して検索した結果である。145件ヒットしている。しかし、件名を見る「美術学校授業囑託の件」とあるだけで森の名前はない。ではなぜ検索でヒットしたかという本文の先頭から300文字程度をテキスト化して検索データに付与しているからである。もちろん、全文検索が理想的かも知れないが手書き文書の多い戦前の公文書をテキスト化して検索することは現実問題として不可能である。そのために開発したのが文書の先頭300文字程度をテキスト化する方法である。これで検索対象となるデータは一般的に件名だけに頼るデータベースよりはるかに増えた。しかし、先

頭 300 文字に含まれない場所に検索語がある場合その資料にはアクセスできない限界はある。公文書を理解するには単に検索にヒットした情報だけでなくその資料が含まれている資料群やどのような構造を持っているかを知ることによって資料が有機的に結びつき一つの資料では判らなかつた事実が判明することがある。例えば、件名の下に【階層】とある。これは EAD を利用して初めて可能となった文書資料が持つ階層構造を示している。つまり、この資料は「防衛研究所所蔵資料」という資料群の中の「陸軍」、さらに陸軍省が作成編集した「大日記類」の中の「老大日記」の明治 24 年 2 月の簿冊（ファイル）に含まれていることが判る。実はこのことから明治 24 年の「老大日記」には森鷗外に関する記録がある可能性が考えられる。そこで関係ありそうな簿冊（ファイル）を調べていくのが公文書館での一般的な資料の探索方法である。デジタル化したことによって階層を自由に選び検索することが出来るようになった。その結果、二番目の「21 年度過渡俸給返納の件」もどうように同じ明治 24 年の「老大日記」に含まれている。前者が 2 月であり後者が 11 月である。このような時系列の情報も公文書にとって重要な意味を持っている。では、この 2 つの資料が意味するのは何かというと前者は軍医として美術学校で美術解剖の講義のために囑託として派遣してほしいとの文部大臣から陸軍大臣宛の依頼状である。そして、11 月には会計検査院から陸軍大臣宛に公務員として給与の 2 重払いになっているために森に払いすぎ給与の返納を求める書類である。EAD の優れている点は公文書の持つ構造を壊すことなく資料を検索出来るだけでなく検索語や作成年月日等で検索を絞り込んだり拡大したりできることである。

「森鷗外の辞書展開による検索結果」

検索条件: 森鷗外|森林太郎|森一等軍医|森一等軍医|森軍医監

この検索結果を対象 絞り込み検索

表示件数 20 辞書 使用しない

所蔵機関:  国立公文書館  外務省外交史料館  防衛省防衛研究所

検索する資料:  資料群  簿冊  件名

---

該当件数: 145件 (21-40件目を表示)

<<先頭 <前 1 2 3 4 5 6 7 8 次> 最後>>

表示順 指定無し

検索対象	No.	資料閲覧	概要
	21		件名美術学校授業囑託の件 【階層】 <a href="#">防衛省防衛研究所</a> > <a href="#">陸軍</a> > <a href="#">陸軍省大日記類</a> > <a href="#">老大日記</a> > <a href="#">明治24年2月「壹大日記」</a> 【レファレンスコード】C03030637200【画像数】3
	22		件名21年度過渡俸給返納の件 【階層】 <a href="#">防衛省防衛研究所</a> > <a href="#">陸軍</a> > <a href="#">陸軍省大日記類</a> > <a href="#">老大日記</a> > <a href="#">明治24年11月「壹大日記」</a> 【レファレンスコード】C03030689100【画像数】5
	23		件名顧問医囑託の件

たとえフルテキスト検索が可能だとしてもこのような検索結果にはならない。公文書を理解するためには構造の理解が不可欠なのである。

昨年 11 月、図書館と文書館の協力を主要テーマとしてカナダで開催された文書館円卓会議 (CITRA) では、欧米で進められている統合的な電子図書館計画へ「アーカイブ」として如何に協力するかが議論となった。その中で「アーカイブ」(記録資料) が電子図書館に加えられることは歓迎できるが「アーカイブ」の独自性をどのように保つのかとの危惧が呈せられた。つまり、デジタル世界では絵画であれ音楽であれ動画であれ文書であれ、デジタル化された情報としては同じ「ビット」の値でしかない。その中で図書館、博物館、文書館の所蔵資料をデジタル化しインターネットで提供することそのものの問題というよりも、図書、博物、文書それぞれの資料が持つ資料としての構造体系や資料の属性を現すメタデータのあり方、つまり検索やアクセスの方法について、それぞれの独自性を如何に確保するかが議論された。少なくとも統制された伝統的な「アーカイブ」資料を Google のような無統制のサーチエンジンのみで理解することの問題点が指摘された。[13]

## おわりに

Google の 2 人の設立者、Sergey Brin と Lawrence Page の共著による *The Anatomy of a Large-Scale Hypertextual Web Search Engine* という彼らがスタンフォード大学のコンピューターサイエンス学部在籍していた 1998 年頃に発表された Google に関する論文がある。その中で彼らは当時のサーチエンジンが広告主等の影響を受けていること、利用者は検索結果のトップから数十件 *a few tens of results* しか見ないこと、一般的な統制されたデータベースとは異なりハイパーテキストで記述された無統制なウェブの大量な情報の検索結果には大量の雑音が含まれること、当時の検索エンジンの多くが検索結果を意図的にコントロールしていることを指摘している。彼らは論文の中で人為的な判断を排除し、高額な機器に頼らず市販のパソコンを大量に接続し、テキストのリンクされている量からその重要度を測るという PageRank というアルゴリズムや新しい技術を導入して全く新しい検索エンジンを開発した。現在の Google の広まりを見れば彼らの夢は叶ったかに見える。[14]

しかし、顧客からの多くの訴訟が示すように Google の検索結果の中立性については既に疑問の声が上がっている。さらに、中国市場への参入に関して政治的問題のある検索語を排除するなど中国政府の要望を受け入れたことが明らかになった。本来、自由で民主的な空間として発達してきたインターネットやウェブの存在の根幹に関わる問題である。さらに世界中の大規模図書館と協力して進める図書の電子化計画でもその秘密主義は関係者の間からも疑問の声が上がっている。Google が商業的に成功すればするほど秘密主義は一層強まっている印象を受ける。Google が提起しているのはインターネットやウェブに係わる者にとって重要な課題を定義している。[15] 昨年 11 月 CITRA での Natalie Ceeney 英国公文書館長の発言は示唆に富む。彼女は、伝統ある Keeper (記録保護官) というタイトルを捨て自らを CEO (最高意志決定者) と位置づけている。ビジネス界で成功した彼女の経験から Google の提案は公文書館の利益とは合致しないものであるというものであった。

ここでは今回のパネルディスカッションのテーマを提供するという意味でいくつかの課題を提供して本稿を終える。

無統制サーチエンジン (Google) vs. 統制されたデータベース (EAD)、商業主義 vs. 公共サービス (コモンズ)、利便性 vs. 創造性、ブラックボックス vs. オープンソース

## 参考文献

- [1] 景山幸一、「デジタルアーカイブという言葉を生んだ「月尾嘉男」」  
[http://www.dnp.co.jp/artscape/artreport/it/k\\_0401.html](http://www.dnp.co.jp/artscape/artreport/it/k_0401.html)。また、わが国での「デジタルアーカイブ」という言葉の起源と博物館や美術館との関係については日本色彩学会誌 Vol.31, No4(2007・12)に収録の国立歴史民俗博物館の鈴木卓治氏の「デジタルアーカイブとは何か」が詳細に論じているので参照されたい。
- [2] <http://old.ica.org/new/citra.php?pcitraprogramid=12&plangue=eng>
- [3] 『アーカイブ辞典』小川千代子、高橋実、大西愛編著、大阪大学出版会、2003年
- [4] 『記録史料の管理と文書館』第10章、「欧米における記録管理」(石原一則)安藤正人、青山英幸編著、北海道大学図書刊行会、1996年
- [5] 同、1章「古代・中世における文書の管理と保存」(松井輝昭)
- [6] 安藤正人、『記録史料学と現代』吉川弘文堂1998年
- [7] 五島敏芳、「日本の記録史料記述 EAD/XML 化と記録史料管理」『情報知識学会誌』Vol.12 No.4 (Jan.2003)
- [8] 『アジア歴史資料センターの設立について』平成7年6月30日、国立公文書館発行『アーカイブズ』27号、平成19年3月20日に本文を掲載
- [9] <http://www.kantei.go.jp/jp/kakugikettei/991130ajia.html>
- [10] <http://www.kantei.go.jp/jp/mille/index.html>
- [11] Negroponte, Nicholas: *Being Digital*, Alfred A. Knopf, Inc. N.Y. 1995、和訳は『ビーング・デジタル』として2001年にアスキー社から出版
- [12] [http://www.archives.go.jp/owning/d\\_archive/pdf/youkou.pdf](http://www.archives.go.jp/owning/d_archive/pdf/youkou.pdf)
- [13] 会議の詳細については今後ネットで公表される。会議の概要については <http://www.citraquebec2007.org/> を参照。
- [14] <http://infolab.stanford.edu/~backrub/google.html>
- [15] Vise, David A. and Malseed, Mark: *The Google Story*, Bantam Dell, N.Y. 2005