

国立公文書館におけるオンラインサービスの展望

国立公文書館業務課

1. はじめに

現在、国立公文書館のホームページでは、一部デジタル画像データも閲覧可能な公文書目録情報検索システムや江戸時代から伝わる国絵図等の大判資料などを高画質・大容量のデジタル画像を始め、年表を利用した資料閲覧システムなど、多彩なコンテンツを提供している。さらに館に設置されたアジア歴史資料センターでは、国立公文書館が所蔵する公文書の他、外務省外交史料館所蔵の外交記録及び、防衛省防衛研究所が所蔵する旧陸海軍関係資料の中から「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」（アジア歴史資料）をデジタル画像化して提供している。これらの3機関が所蔵する資料をデジタル化し「アジア歴史資料」として、単一のデータベースからオンラインで提供することによって、これまで物理的に分断され、統一的なアクセスが不可能であった、多様な公文書資料をインターネットに接続されてさえいれば「いつでも」「どこでも」「だれもが」「自由に」「無料で」利用できる環境を整備してきている。本レポートでは国立公文書館がオンラインで提供する諸サービスの概要を報告するとともにオンラインサービスによって生まれた可能性や将来の展望と、オンラインサービスを運営維持していく上での問題や課題について報告する。

2. 国立公文書館のオンラインサービス

1971年7月に開館した国立公文書館には、1,105,496冊の資料が所蔵されている。その資料内訳は、大きく二分されている。まず、第1は、

明治から平成の現代まで国の各機関が作成した歴史資料として重要な公文書で625,996冊を数える。第2のカテゴリーは、江戸時代の徳川幕府の統治と関連する資料を中心とする「内閣文庫」の古書・古文書で、その数は479,500冊となっている。

これらの所蔵資料の目録は全て公開されている。インターネットを通じて提供されている「国立公文書館デジタルアーカイブ・システム」には、1,102,028冊の所蔵資料（全所蔵資料中、99.7%に当たる。）の目録情報がデータベースに登載されている。国立公文書館では、その所蔵資料の目録情報のみならず、定期刊行物など多様な情報を様々な形でデジタル化しホームページ上で公開している。「デジタルアーカイブ推進にあたっての基本的な考え方」として館が提唱した3つの基本的な考え方がある。(1)デジタルであること、つまりオンラインで「いつでも」「どこでも」「あれもが」アクセスできること。(2)提供する画像資料が「パブリックドメイン」（公有）であること、および提供するためのシステムに採用された技術が可能な限りオープンな技術であること。(3)自由に、無料で利用できることは、利用者から高い評価を受けているだけでなくインターネット上で情報を提供する図書館や博物館などの類縁機関のサービスのあり方にも一つのモデルを提供しているといえる。ここでは館がホームページで提供するオンラインサービスの概要を紹介する。

2.1 ホームページ

館では1998年2月から「国立公文書館ホームページ」を開設し所蔵資料の目録データベースや展示会等の情報を公開してきた。しかしながら、これ

らの情報提供は、データベースの蓄積の少なさ、画像処理技術、電送回線容量、利用者側端末の処理能力等がいずれも未発達不十分であったことから、極めて限定的であった。このような状況の中、政府が2002年6月に策定した「e-Japan 重点計画2002」において「重要な公文書等の歴史資料等に関する情報が、地理的な制約を受けずにどこにおいても入手・利用できる環境を整備する。」ことが明記され2005年度までにデジタルアーカイブ化の推進が謳われた。これを受けて館として、「国立公文書館デジタルアーカイブ」の具体的な検討に着手し、2004年4月には「独立行政法人国立公文書館デジタルアーカイブ推進要綱」を制定した。これによって、「国民が、国立公文書館デジタルアーカイブにアクセスすることにより、国立公文書館に来館することなく、インターネットを通じて所蔵資料の検索、電子化された画像の閲覧・印刷が容易に行える」システムを構築することを目標に掲げた。現在は、次に詳述する「国立公文書館デジタルアーカイブ」だけでなく館ホームページ上で利用者の必要性や興味に応ずる多様な「入り口」を設け、そのためのコンテンツの充実に務めている。ホームページに新たに設けた主なコンテンツとしては、業務の内容や資料の概要を映像で提供する「図解国立公文書館」、非公開の書庫内部など施設内部を見ることの出来る「バーチャルツアー」、1867年から1972年までの重要資料のカラー画像を年表の歴史事項に沿って閲覧できる「公文書に見る日本のあゆみ」、公文書や公文書館の業務について馴染みのない学生や一般の人たちがキャラクターを使って学べる「ぶん蔵」などである。

2.2 「国立公文書館デジタルアーカイブ」(DA)

2005年4月1日公開された「国立公文書館デジタルアーカイブ」(Digital Archives : DA) は館のホームページで提供するサービスの根幹をなすものである。DAは、2つのサービスから成り立っている。1つは、国立公文書館所蔵の公文書及び

古書・古文書の目録情報の検索を中心とするものであり、一部には、目録情報の検索結果と連動して当該資料のデジタル画像も提供する「デジタルアーカイブ・システム」(DAS)である。もう1つは、国指定重要文化財や原本の閲覧が困難である大判の資料、色彩豊かな巻物や絵図・ポスター等の画像を高精細カラーデジタル画像として提供する「デジタル・ギャラリー」(DG)である。この2つのサービスが、インターネットを通じて、「いつでも」「どこでも」「だれもが」「無料で」「自由に」利用できる。

2.2.1 「デジタルアーカイブ・システム」(DAS)

DASの導入にあたっては、事前作業としてISAD(G)(General International Standard Archival Description)の第2版の記述要素に対応させながら、移管元省庁毎に資料群概要を作成した。移管元省庁の業務又は公文書の内容に応じて「資料群」を構成し、「資料群」相互の関係を階層的に整理した。システムへの登載にあたっては「EAD2002: Encoded Archival Description 2002」(符号化アーカイブ資料記述)に準拠したXML形式のほか、他の歴史資料保存利用機関や図書館等のデータベースとの横断検索を可能にするZ39.50、SOAP等を採用した。DASには、4つの検索方法が用意されている。

第1の検索方法は、「キーワード簡易検索」であり、公文書や古書・古文書の標題や件名、作成部局、旧保存機関、著者名などに含まれると考えられる任意のキーワードを入力して検索するものである。

第2の検索方法は、資料の移管元省庁別等に分類された階層構造を辿りながら検索することができる「階層検索」である。国立公文書館所蔵資料は、まず第1階層で「公文書」と「内閣文庫」に分類される。以下、「公文書」では、第2階層では移管元省庁別に、第3階層以下は資料群ごとに分類されている。この階層構造をEAD2002の適用という観点から見ると、第1階層をfondsとし、

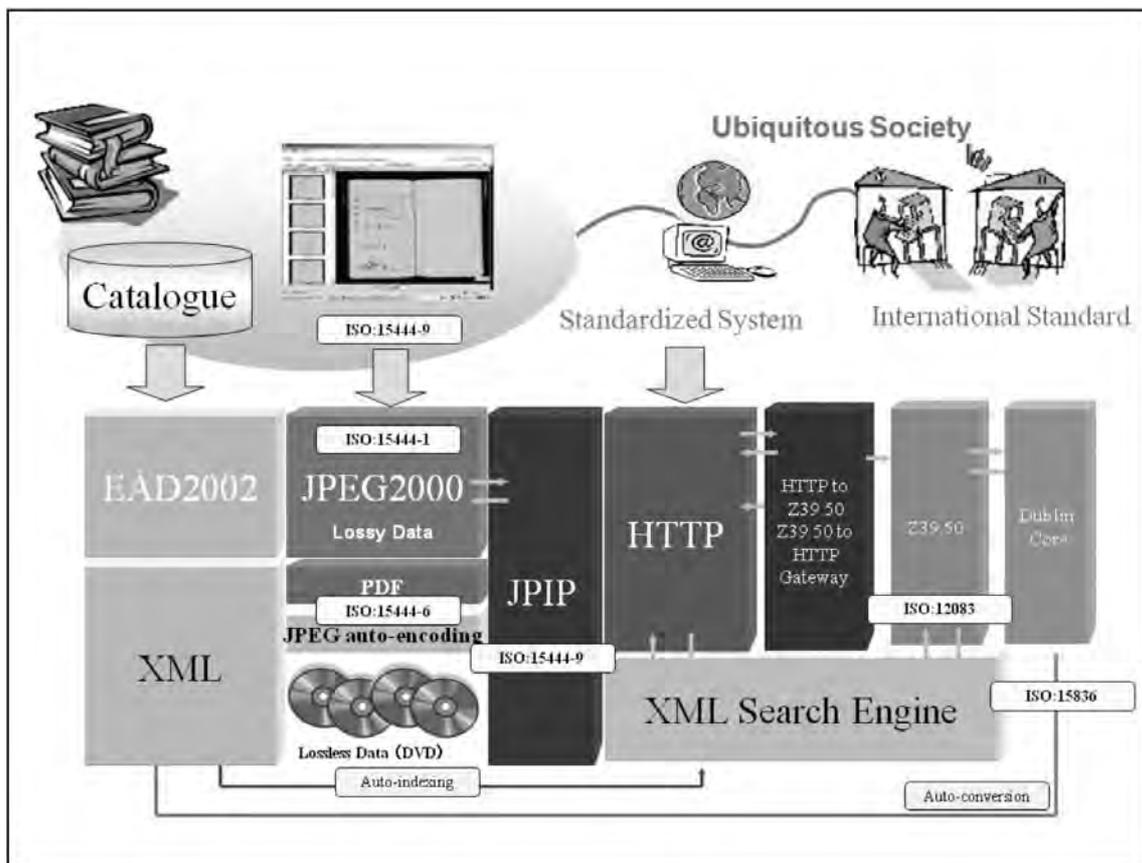


図1 : DAS の概念図

その下に最大4階層の series を置き、第6階層に file、第7階層に item を置いている。

第3の検索方法は、検索項目別に多様な条件を設定できる「キーワード詳細検索」である。「タイトル (簿冊標題・件名)」、「作成者名称」及び「関連事項」ごとにキーワード検索が可能であるほか、項目間の条件を「AND (すべてを含む)」、「OR (どちらかを含む)」及び「NOT (対象を除く)」という形で設定することができる。

第4の検索方法は、国立公文書館所蔵資料の情報以外に他機関のデータベースとの間で、共通のキーワードによる検索ができる「横断検索」である。「横断検索」は、共通検索プロトコルである Z39.50 を実装しているデータベースとの間で可能である。2008年3月現在、DAS からは、アジア歴史資料センター、岡山県立記録資料館及び国立情報学研究所総合目録データベース WWW 検索サービスとの間で横断検索が可能である。

画像フォーマットの採用に当たっては、利用者

が将来にわたって安定的に利用できるよう、画像符号化に関する規格として国際標準化機構 (ISO) 及び国際電気通信連合・電気通信標準化セクタ (ITU-TS) によって ISO/IEC 15444 シリーズとして標準化された JPEG2000 を標準画像フォーマットとして採用した。ただし、利用者のパソコン環境によっては最新の JPEG2000 に対応出来ないことがあるため全ての画像を一般的な文書フォーマットである PDF (ISO/15444-6) と JPEG でも提供している (図1)。

2.3.2 「デジタル・ギャラリー」(DG)

国立公文書館では、公文書の外に明治政府が徳川幕府から引き継いだ、漢籍を多く含む紅葉山文庫 (もみじやまぶんどこ) 等、江戸時代の資料や明治政府が収集した古典籍、古文書、欧米の出版物など貴重な資料が「内閣文庫」として保存されている。その中には、徳川幕府が諸大名に命じて作成させた国絵図など大判なために展示が不可能な

ものや絵巻物や古地図が含まれており、これらの貴重資料の多くは国の重要文化財に指定されている。これらの資料を高精細カラーデジタル画像で公開しているのが「デジタル・ギャラリー」(DG)である。現在、710画像(531点)を提供している。

DGでは、3つの検索方法を用意している。第1の方法は、「カテゴリー別」の検索である。現在、デジタル・ギャラリーでカラー画像を利用可能な資料は、「文書」・「絵図」・「絵巻物」・「写真」・「図面」・「ポスター」の6種類に分類されており、任意のカテゴリーを選択すると、該当する資料のサムネイル画像が表示されるようになっている。第2の方法は、「50音順」の検索であり、資料の名称があらかじめ分かっている場合に索引的に使用できるものである。第3の方法は、「地域別」の検索であり、地方別、都道府県別に資料が分類されている。

いずれの検索方法も極めて容易に国立公文書館所蔵資料の高画質なカラーデジタル画像を利用することができるものであり、いわば「インターネット展示室」というべきサービスとなっている。小・中学校や高等学校の学習用教材としても活用が期待される。

3. アジア歴史資料センターのオンラインサービス

アジア歴史資料センター(以下、JACAR)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」である「アジア歴史資料」をインターネットで「いつでも」「どこでも」「だれもが」「自由に」「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日に独立行政法人国立公文書館に設置された。それ以降、国立公文書館、外務省外交史料館、防衛庁防衛研究所図書館が所蔵公開している「アジア歴史資料」をデジタル化し、データベースを構築して提供している。現在の提供数は、電子画像約1500万画像、目録データ約100万件であるが近い将来には、提

供画像数を3000万に増加することを目指している。インターネットで公開されている公文書の画像データベースとしては既に現時点で世界最大級のものである。

サービス開始から6年目を迎えてJACARの資料は学术论文や出版物での引用も増えている。また、遠隔地から一次資料の利用が可能になったことで地方の大学の歴史教育で積極的に利用されるようになった。さらに、日本政府が中国や韓国とそれぞれに設置した歴史共同研究の場でも活用されるなど積極的に利用されてきている。

2006年のJACARシステム更新にあたっては、2005年に「国立公文書館デジタルアーカイブ・システム」(DAS)で実証された情報システムをセンターの基幹システムとして採用した。さらに最先端の技術や大幅なオープンソースソフトを採用した。DASシステムをJACARの基幹システムに採用した背景としては、同じ国立公文書館が運営するインターネットによるデータベースシステムを極力共通化することで汎用性を高めることで設計コスト等を削減するだけでなく利用者にとっても同じ手順で利用できるなど利便性が高まることを目的とした。ここではDASとは異なるセンター独自の部分を中心に紹介する。

3.1 センター情報提供システムの特長

新システムが旧システムと大きく異なるのは、国立公文書館デジタルアーカイブに採用された検索目録データベースにEAD2002に準拠したXML形式や他機関との横断検索を可能とするZ39.50、SRW(Search/RetrieveWebService:<http://www.loc.gov/standards/sru/srw/index.html>)等のプロトコル等、先進的かつ国際標準技術を採用した点である。さらに、画像フォーマットには旧システムから継続利用するDjVuだけでなく国立公文書館で採用した国際的な標準フォーマットであるJPEG2000(<http://www.jpeg.org/jpeg2000/index.html>)も利用出来るようにした。

3.1.1 デジタル画像変換仕様

センターに提供される画像資料は、各所蔵機関で16mmまたは35mmでマイクロフィルム化されたものであるが、これらの画像データをデジタル化する仕様は、歴史研究者や画像技術者が歴史研究に耐えうる画像の階調と解像度の水準を検討した結果、文字が研究用途として判読可能であり、かつ、データ量を出来る限りコンパクト化するため、モノクロ2値、400dpiを標準とした。また、画像ファイル形式については、機器や特定ソフトに依存性が少なく、欧米の公文書館でも標準的に利用されているTIFF (Tagged Image File Format) 形式とした。さらにインターネットでの公開形式については、オリジナルのTIFF画像が大きな情報量であっても高圧縮が可能でスムーズに画像が再現できるDjVu形式を採用した。DjVu形式は、モノクロ2値の文書情報圧縮の場合、従来用いていたJPEG形式の10～20倍の圧縮率となり、同じ通信環境であればJPEGやPDF等の一般的な画像フォーマットと比較して極めて短時間で文書画像の閲覧が出来る、現状でも最高レベルの圧縮性能を持っている。

3.1.2 同義語・関連語・英語を含む辞書機能の導入

異なるレベルの歴史知識や歴史観に立つ利用者に対して、自由かつ容易に目的とする資料を検索できるようにするための機能である。データ検索の対象となる目録データを増やしても、利用者が使用する検索用語と実際に公文書に使用されている歴史的用語とが乖離している場合が多く、検索できないことが多い。たとえば、教科書や一般的な歴史書に使用される「太平洋戦争」で検索してもわずかししかヒットしない。関係資料を検索するためには当時の公文書に使用された閣議決定による正式名称である「大東亜戦争」を用いなければならない。そこで一般的に利用されている現代用語を当時の歴史用語に結びつける特殊な「歴史用語辞書」を編纂している。現在、基本語5600語

が辞書としてシステムに搭載されている。基本語には英語検索のための英訳またはローマ字読みが付与されている。辞書システムを応用して提供されているのが五十音検索とA-Z Index検索である。

「歴史用語辞書」は、東アジア域内での資源共有のために不可欠な手段といえる。例えば、「満州事変」は中国では「九一八事変」として認識されている。「歴史用語辞書」は、単に歴史用語と現代用語とを結びつけるだけでなく、時代や国の文化と共に変化する地名や呼称の差異、日本、中国、韓国で漢字文化を共有していても起こり得る人名や地名の音表記の違いなどを克服する有効な手段といえる。

3.1.3 特別展

アジア歴史資料センターはインターネット上のみ存在するデジタルアーカイブである。提供する画像データは一般には馴染みのない公文書である。そこで歴史資料としての公文書をより多くの人に知ってもらうためにセンターでは「公文書に見る」インターネット特別展を提供している。現在、「日露戦争展」、「岩倉使節団」、「日米交渉」、「写真週報」の4つの特別展を提供している。なお、本年の4月からは年表を使った検索システムを開発し、年表に関連する条約や御署名原本を閲覧することにより日本と諸外国との歴史的関係をたどることが出来る「条約と御署名原本に見る日本の近代史」を開催している。

4. 電子政府化と国立公文書館オンラインサービスの今後の展開

2001年1月、政府の高度情報通信ネットワーク社会推進戦略本部 (IT戦略本部) が策定した「e Japan戦略」では、電子政府化について、「行政内部や行政と国民・事業者との間で書類ベース、対面ベースで行われている業務をオンライン化し、情報ネットワークを通じて省庁横断的、国・地方一体的に情報を瞬時に共有・活用する新たな

行政を実現するものである」と位置づけられた。各省庁は行政情報のインターネット上での公開を推進し、行政組織・制度等に関する基礎的情報のほか、行政活動の現状に関する情報、予算・決算に関する情報、行政評価に関する情報などを省庁のホームページを通じて積極的に提供している。一方、各省庁で作成される記録のほとんどが電子的に作成されながら、電子媒体により保存されている比率は、全ての記録の4%を占めているにすぎない(2005年度調査)。ただし、今後の電子政府化の更なる進展に伴い、電子的に作成・保存され電子媒体のまま公文書館へ移管される文書が増大することは不可避である。

このような状況を受けて、公文書館制度を所掌する内閣府と国立公文書館は協力してその対応策を検討してきた。2007年度、内閣府は、電子記録の移管と長期保存を実現するという観点から、電子記録のメタデータ及びフォーマットのあり方について、実証実験に着手した。2008年度には、プロトタイプにより、電子記録の移管・保存・利用システムのあり方を総合的に検証し、電子記録の移管・保存・利用に関する基準、ルール、フォーマット等の策定を行う予定で予算を計上したところである。さらに、2009年度から2010年度にかけては、これらの実証的研究の上に立って、国立公文書館が電子記録の移管・保存・利用システムの設計と構築を行い、2011年度に電子記録の移管・保存を開始するという計画を立てている。本システムでは、受け入れた電子記録について、ウィルスチェック、長期保存フォーマットへの変換、必要なメタデータの抽出・付与、保存システムへの格納、一般公開システム(DAS)へのデータ搭

載などの多岐にわたる作業を迅速かつ確実にできるようにするものである。現在国立公文書館では、紙媒体資料のデジタル画像化による提供を更に進めて、ポーン・デジタルの公文書の保存・公開もオンラインで行うことを可能とするよう文書のライフサイクルを電子的に運営管理する「デジタルアーカイブ」の実現を目指している。

5. 終わりに

政府の説明責任に果たす公文書館の役割が重要視される今日、自由なアクセスを提供出来るかは公文書館の存在意義を問う最重要課題といえる。情報インフラの基盤整備が進みオンラインサービスが可能になったことでこれまで距離や時間の条件により物理的にアクセスが制限されていた公文書へのアクセスは根本から改善された。その一方で、ホームページ上の情報をどう維持管理・保存するかというウェブアーカイビングの実施方法などの新たな問題が生まれている。また、電子文書の保存についても漢字文化を共有する東アジア文化圏での文字コードへの対応という重要な問題がある。各国が進める文字コード化と国際標準(ISO-10646) UNICODER の関係など、今後 EASTICA メンバー間での漢字で書かれた歴史資料のネットワーク上での共同利用を進める上で避けて通れない問題である。また今後のオンラインサービスにとって重要な、音声、動画、画像等のフォーマットの標準化など関係各国が協力して進めることの出来る課題も多い。EASTICA は、オンラインサービス関連技術の標準化などの意見交換の場となることを期待する。