

地域発表：EASTICA2005

重要記録の保存と利用：国立公文書館デジタルアーカイブの試み

国立公文書館

1 国立公文書館が所蔵する貴重記録

日本の歴史は、地震、台風、水害、火山の噴火などの自然災害の歴史とも言える。さらに太平洋戦争下の空襲や重要文書の疎開、そして、終戦時の廃棄によって多くの貴重な記録が失われたり散逸したりした。このような状況に危機感を頂いた日本学術会議は、1959年（昭和34年）内閣総理大臣に対して、公文書の散逸防止と公開の施設として「国立公文書館」の設立を勧告した。実は、太平洋戦争以前の政府では、各省庁がそれぞれで文書を保存管理しており、国の公文書を集中して保存管理するいわゆる「国立公文書館」はなかった。政府は、1971年、「国の各行政機関から受け入れた歴史資料として重要な公文書等を永く保存し、閲覧・展示などに供するとともに、そのための調査研究を行う機関」として現在の国立公文書館を設立した。

国立公文書館には、明治時代（1868年から）以降、内閣が授受した公文書だけでなく、最古のものでは、延喜8年（908年）の年号がある「東大寺文書」のように、明治以前の貴重な記録が移管された。さらに、各省庁からは「歴史資料として重要な公文書」が移管されている。所蔵資料には、「御署名原本」と呼ばれる天皇の御名・御璽のある憲法、詔書、法律、条約、勅令、政令など、政府の最重要文書の他、1988年に国の重要文化財に指定された、1868年から1885年までの明治政府の公文書を編集した「公文録」、1882年以来、法律や規則の原義を収録し編集された「公文類聚」など、明治以降の政府文書が保存されている。その他、江戸幕府（1603年から1867年）が収集した古書、古文書や、さらに国の重要文化財に指定され26件の古書や古文書や古地図がある。特に「御署名原本」などの最重要文書は東京の本館ではなく1998年、東京から50キロ離れた、つくば市に建設された分館の貴重書庫に保管されている。さらに重要文化財に指定された文書や地図、その他の貴重資料は、東京の本館に新たに設けられた貴重書庫に保管されている。日本国憲法のように一部の資料は精巧なレプリカによって展示室で見ることができる。しかし、これら貴重資料の多くはマイクロフィルムによる閲覧に限られてきた。また、大判の地図や巻物を一般の利用者が見る機会は特別な展示以外にはなかった。

2 国立公文書館デジタルアーカイブ構想

2001年11月、国立公文書館の組織としてアジア歴史資料センターが開設された。同センターは、日本政府が保存公開している戦前の公文書の画像をデジタル化しインターネットで公開している。戦前の公文書を「いつでも」「どこでも」「だれもが」「自由に」利用出来る様にする事で近隣諸国の国民との相互理解を深めるためのものである。2005年9月現在、1000万画像を超える戦前の公文書のデジタル画像を提供する本格的なデジタルアーカイブである。

国立公文書館では、2002年夏にアジア歴史資料センターの経験を踏まえ、公文書館所蔵資料の

デジタル化の検討に入った。アジア歴史資料センター設立の検討段階では国内全般のブロードバンドの未整備など、情報基盤や画像圧縮技術の限界からモノクロ2値の画像を提供するのが限界であった。しかし、この数年、政府のe-Japan計画など、国家戦略としての情報基盤の充実や技術の進展により、限りなく原本に近いカラーイメージを提供出来る可能性が生まれた。2003年より具体的な調査研究を行い画像フォーマットにJPEG2000を採用することで、これまで保存の観点から実物の公開が出来なかった大判の古地図や絵巻物、モノクロマイクロフィルムでの閲覧しか出来なかった日本国憲法のような最重要文書などをカラー画像で見ることが可能となった。さらに200万件を超える文書の検索を容易にするためにEAD（符号化記録史料記述）を採用するなど最先端の技術を導入した国立公文書館デジタルアーカイブを2005年4月1日に開設した。これによって原本を最良の環境で保存しながらインターネットを通じて「いつでも」「どこでも」「だれもが」「自由に」「無料で」貴重な資料を利用出来る環境を提供することが可能となった。

3 国立公文書館デジタルアーカイブを支える3つの考え方

「Being Digital」（デジタルであること）、「Commons」（コモンズ：共有財）、「WWW」（ワールドワイドウェブ）の3つを合わせて考え、コンセプトを整理したものが、国立公文書館が提唱する、「いつでも」、「どこでも」、「誰もが」、「自由」に使える「ユビキタス」デジタルアーカイブである。

「Being Digital」（デジタルであること）とは、マサチューセッツ工科大学教授のニコラス・ネグロポンティの定義によれば、アトム（実体のあるもの）がビット（デジタル情報）になるということである。デジタルであることは、光の速さで瞬時に廉価（ネグロポンティは、かならずしも「無料」である必要はなく「廉価」であればよいとする）で電子データが送信されることである。光のスピードでいろいろな情報が結びつくことが「デジタルであること」である。デジタル情報をオンラインではなく完結的なパッケージソフトとしてDVD等の媒体で売るとは、本来の意味での「デジタルであること」ではない。デジタル情報はネットワークに入って初めてデジタルであることの意味がある。そしてインターネットの世界を誰でもが簡単に利用出来るようにしたのが「ワールドワイドウェブ」WWWの技術である。

もう一つ重要なのが米国でインターネット関連の法律学者であるローレンス・レシグが提唱するコモンズ（共有財）という考え方である。企業や個人が払った税金で作られた政府の記録や情報は、基本的にパブリックドメイン（公共財）であるとされる。パブリックドメインは、原則的に著作権はないが、国によって意見が違ふ。例えば国の出版物に関して、日本の場合は国が著作権を主張している。米国の場合は、国の出版物であっても、税金で作られた場合は原則として著作権はない。パブリックドメインの出版物だけでなく、その中にある情報も含めて自由に加工、出版販売することができる。日本の場合、政府の出版物に関しては国に著作権が認められている。出版物でない公文書原本には著作権がないとされる。国立公文書館でも、基本的な考え方として同館所蔵の資料は原則的にパブリックドメインであり、著作権フリーであるという考え方をとっている。

4 国立公文書館デジタルアーカイブの紹介

国立公文書館デジタルアーカイブの特徴は、国際標準画像フォーマットであるJPEG2000を採用したことである。また、検索に関する情報の記述方式は、EAD/XMLである。EAD（Encoded Archival Description）とは、公文書館や図書館の目録記述方法をもとにして作られた機械で可読、検索できるメタデータ方式である。単に公文書館だけの資料を検索するのではなく、他の公文書館や図書館の目録を横断的に検索可能とする通信プロトコルであるZ39.50も実装している。

デジタルアーカイブ国立公文書館では、「デジタルアーカイブ・システム」と「デジタル・ギャラリー」の2つのサービスを提供している。

「デジタルアーカイブ・システム」

デジタルアーカイブ・システムは検索システムである。デジタルアーカイブ・システムでは、現在、公文書館が所蔵する主要公文書58万冊と江戸幕府から引き継いだ古書43万冊の目録データ、約240万件のキーワード検索を可能とした他、各省庁毎の所蔵資料の概要から個別資料までたどりつける、階層検索等利用者の利便性に合わせて様々な検索補助システムを組み込んでいる。目録情報の記述方式は、EAD（Encoded Archival Description）/XMLである。米国議会図書館とカリフォルニア大学が共同で開発してきたEADは、公文書館や図書館の目録記述方法をもとにして作られた機械で可読、検索できるメタデータ方式である。また、他の公文書館や図書館の目録を横断的に検索可能とする通信プロトコルであるZ39.50やメタデータの共有を可能とするSOAP等も実装している。

デジタルアーカイブ・システムは、各省庁から移管されてくる現代資料にも対応したシステムである。今後、利用度の高い文書群から資料のデジタル化を行い検索だけでなく原本の閲覧も可能としていく。全ての資料が閲覧出来るわけではないが閲覧出来ないものも、タイトルは公開し「要審査」や「非公開」と明示してある。画像閲覧に当たっては、利用者のパソコンの環境や目的に応じて選択できるように、「JPEG2000」のほか「PDF」や「JPEG」方式でも利用可能な環境を提供している。本年4月の運用開始時には、日本国憲法制定等の戦後改革の資料を中心にデジタル化した12万画像に加え、アジア歴史資料センターで公開している画像データも含め、180万画像が閲覧可能である。

「デジタル・ギャラリー」

デジタル・ギャラリーは、物理的に閲覧が困難な一辺が7メートルに及ぶ「天保国絵図」（重要文化財）や「日本国憲法」のような貴重資料、色彩豊かな巻物やポスター、主要都市の終戦直後における焼失状況を記した「戦災概況図」などの多様な資料の高精細カラー画像223点、346画像をインターネットで提供している。地域やカテゴリー、五十音順から検索し、画像の拡大・縮小、スクロール等が出来る。デジタル・ギャラリーでも利用者のパソコン環境や目的に応じて選択できるように、「JPEG2000」のほか「JPEG」形式で閲覧出来るようになっている。提供画像のうち特に「天保国絵図」82点は、南は琉球から北は蝦夷までほぼ全国をカバーしている。当時の村や町の名前も確認出来、歴史だけでなく地理の授業にも利用出来る資料である。「安芸の国」を例にとると、地図自体は3.5m×3.5mの大判で広げてもなかなか全体を見ることができない。そこで、

一度カラー写真に分割して撮影し、個々の写真をデジタル化してつなぎ合わせ、1本のデジタル画像データにしている。データ量は5ギガ程度である。これではデータが重くインターネットでの提供には無理がある。そこでデータ量を画像の劣化が肉眼で確認できない程度に圧縮して提供している。それでもコンピューターの画面の全体像では点でしかない厳島神社が画像を拡大することで詳細に見ることができる。絵巻物には、「桜町殿行幸図」等がある。長さ20メートルを超え、これまでのJPEGフォーマットでは画像を一度に見ることはできなかった。JPEG2000では全体をスクロールして見たり部分を拡大したりすることが簡単である。

JPEG2000の技術的な特長は、大量のデータの内、必要なデータのみを相互通信しながら受け取ることである。配信受信双方のコンピューターの負荷を押さえて、大量のデータを提供するwww.の利点を最大限生かしたウェブレット変換技術に基づく画像フォーマットである。しかし、国や一部の組織でプラグインソフトのダウンロードを管理者以外に認めていないところが増えているため、この特長を活かすことができない場合がある。またファイアウォールの設定によっては相互通信がうまく行かない場合が確認されている。そこでそのような制限に影響を受けないJPEG版も提供している。JPEG版は、JPEG2000と比較すると操作性が落ちるがセキュリティーが優先されている現状では避けられない方法である。

「電子媒体の保存管理」

国立公文書館では、公文書は一旦モノクロのマイクロフィルム撮影を行い機械的スキャンによるJPEG2000画像データへの変換を行っている。また、古地図、絵巻物、憲法などの重要文書はカラー撮影後JPEG2000へ変換される。マイクロフィルム、カラーフィルム、DVDに格納された非可逆（ロスレス）JPEG2000画像データはバーコードを付与され、保存用としてつくば分館にて厳重に保存管理されている。一般に利用されるのは、非可逆のJPEG2000画像データを必要に応じて圧縮し、東京にある本館のデータベースに投入されたデータである。

5 貴重記録の保存利用手段としてのデジタルアーカイブの可能性

国立公文書館が提供するデジタルアーカイブは、最先端の技術を用い、既に公文書館に所蔵されている「歴史的な公文書」だけでなく将来公文書館への移管が想定される電子的に作成された公文書（電子公文書）も提供できるように考えられた、本来的な意味での電子文書館「デジタルアーカイブ」である。国立公文書館のデジタルアーカイブの設計に当たっては、既に紹介したように電子化された国際標準の記録資料記述方式であるEAD2002や画像フォーマットにも最先端かつ国際標準のJPEG2000形式を採用した。公開され、かつ世界標準の技術を採用することで公文書館のデジタル化のモデルとするものである。紙原本のデジタル化は、原本保存の観点からも、今後各国の公文書館で採用されていくと考えられる。電子文書の保存管理も念頭にシステムは設計されている。国立公文書館では喜んでこれまで培ってきた経験やノウハウを共有していきたいと考えている。