

本格的デジタルアーカイブを目指して：アジア歴史資料センターの実験

牟田昌平

国立公文書館アジア歴史資料センター

〒102-0093 東京都千代田区平河町2-1-2、住友半蔵門ビル別館4階

Tel:03-3556-8801, Fax:03-3216-1521, E-Mail: smuta@archives.go.jp

概要

アジア歴史資料センターは、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（旧陸海軍文書所蔵）の3館が所蔵する、明治初期から太平洋戦争終了期までの「アジア歴史資料」をインターネットで提供している。2003年9月末現在、画像データ数は約440万、目録データ約23万件、将来的には2800万画像の提供を目指す、本格的デジタルアーカイブとして注目を集めている。センター開設の最大の課題は、(1)画像処理・配信の仕様決定(2)目録仕様と作成手順の検討であった。本稿では、センター開設に当たって直面した諸課題に、いかに対処したかを実践的事例として紹介する。

キーワード

デジタルアーカイブ、画像、マイクロフィルム、歴史資料、画像圧縮、メタデータ、画像データベース、コンテンツ、電子辞書、Dublin Core、DjVu、ISAD(G)

Towards A Full-fledged Digital Archives: Experiment at the Japan Center for Asian Historical Records

Shohei MUTA

Japan Center for Asian Historical Archives, National Archives of Japan

Sumitomo Hanzomon Building, Annex-4F

2-1-2 Hirakawa-cho, Chiyoda-ku

Tokyo 102-0093, Japan

Phone: +81-3-3556-8801, Fax:+81-3-3216-1521, E-Mail: smuta@archives.go.jp

Abstract

The Japan Center for Asian Historical Records of the National Archives of Japan (JACAR) was established on the 30th of November 2001 as a part of an organ of the National Archives of Japan. Its purpose is providing people of all over the world access through the Internet to official records from collections kept by various ministries and agencies of the Japanese Government on modern history of relations between Japan and various foreign countries, primarily neighboring Asian countries. In order to fulfill its mission, the center introduced very unique technology and system such as the DjVu document compression technology; a catalogue system relying on the General International Standard Archival Description (ISAD (G)) and Dublin Core; and a digital dictionary for synonymous, related words, and English translations specialized for Japanese modern history. As of September 2003, the center provides about 4.4 million pieces of image data and a catalog database of 230,000 items, which will be augmented on an ongoing basis.

Keywords

digital archives, historical records, image compression technology, DjVu, Dublin Core, ISAD (G), digital dictionary, synonymous

※本論文は第24回デジタル図書館ワークショップの論文です。

1. はじめに

アジア歴史資料センター（以下「センター」：www.jacar.go.jp）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」（閣議決定：「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」「無料」で利用出来る、本格的な公共デジタルアーカイブとして2001年11月30日に開設された。2003年9月末現在、センターでは、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（旧陸海軍文書所蔵）の3館が所蔵する明治初期から太平洋戦争終了期までの「アジア歴史資料」を電子データで収集・提供している。画像データ数は約440万、目録データ約23万件、将来的には2800万画像の提供を目指している。日本語での検索手段を持たない利用者に対応するため、試行的な英文検索システムも稼働中で、海外の研究者からも日本の近現代史の貴重な知的資源として注目されている。本論では、情報提供システムの根幹をなす画像提供システムや目録検索システムの概要について紹介するとともに、急速な発展を遂げる情報技術を背景にセンターの今後の展望について言及する。

2. 公文書館制度が抱える課題とセンター開設

2.1 閣議決定によるデジタルアーカイブとしての開設

センター開設は、1994年8月31日、村山総理（当時）が、戦後50周年を記念する「友好交流計画遂行過去の歴史を直視するために、歴史図書・資料の収集、研究者に対する支援等」を行うセンターの設立検討を内外に確約したことに端を発する。翌年、15名からなる有識者は、「日本とアジア近隣諸国等との間の近現代史に関する資料及び資料情報を、幅広く、片寄りなく収集し、これを内外の研究者をはじめ広く一般に提供すること」を目的とする「アジア歴史資料センター」の設立を提言した。有識者会議提言の具体的検討のため、内閣外政審議室に担当部署が設置され、内外類似機関や歴史資料の保存公開に関する調査、センターの具体的な組織やシステムについて調査・検討が行われた。そして、1999年11月30日「アジア歴史資料整備事業」の一環としてまず、国の諸機関が所蔵する「アジア歴史資料」をインターネットで提供する「アジア歴史資料センター」を国立公文書館の組織として開設することが閣議決定された。有識者提言では、資料をマイクロフィルムで収集公開、目録のみをインターネットで提供することになっていたが、一気にデジタル画像をインターネットで提供する本格的なデジタルアーカイブとなったのは、公文書館や次に紹介する日本の近現代史を取り巻く多くの課題、情報技術やブロードバンドなどの情報基盤の急速な発展、小淵内閣が打ち出した国策としてのIT戦略など複雑な経緯があった。

2.2 公文書館制度の未整備と近現代資料の保存・公開を巡る課題

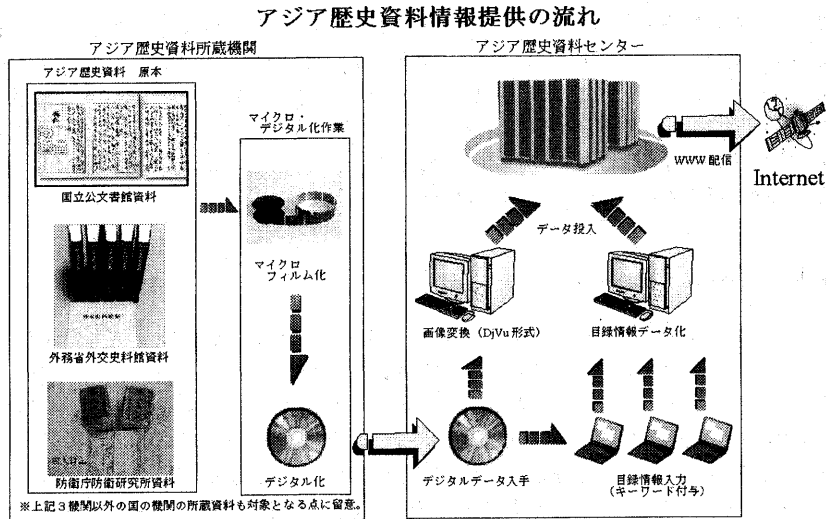
閣議決定は、センターが「国立公文書館、外務省外交史料館、防衛研究所図書館等の国の機関が保管するアジア歴史資料を電子情報の形で蓄積するデータベースを構築し、インターネット等を通じて情報提供を行う」機関であるとした。しかし、当時まだブロードバンドの普及は著についたばかり、インターネットでの画像提供については、ダイヤルアップを前提とするような状況であった。2000年4月に総理府（当時）内にセンター開設準備室が設置され、情報提供システムの具体的な検討が始まった。このような大量の画像データや公文書目録データベースをインターネットで提供した先行事例もなく、準備作業は試行錯誤であった。基礎調査の結果、センターの成功の鍵は、画像圧縮と目録検索システムにあることがわかった。しかし、本来このような先進的な試みに必要とされたパイロットシステムによる試験運用は行わず、閣議決定から2年後に本格的稼働とすることになった。システムの基本仕様が決定し、初年度予定された約画像200万、目録17万件を処理し後に紹介する辞書を作成するために当てるのが出来るのは6ヶ月程度と考

えられた。この状況は単年度予算、作業のアウトソーシングのための公開入札の原則等の制約があり日程的には厳しい状況が現在も続いている。抜本的な改善には年度予算のあり方や入札契約方法等の見直しが必要である。

さらに、本年4月に内閣府官房長が設置した「歴史資料として重要な公文書等の適切な保存・利用等のための研究会」が「中間とりまとめ」で指摘したように「公文書館制度の重要性について政府として十分な認識を持ってきたとはいえ、国民の理解も一部の人にとどまっていた」現状があった。(注1) 特に、歴史資料としての公文書を整理し利用できるようにするために不可欠な目録作成が出来る専門家(アーキビスト)の不足は、想定された大量の目録情報処理のボトルネックである。また、日本の近現代資料を整理する基本的な「史料学」や、図書における日本十進分類表のような整理体系の未発達は、アーキビストの絶対的な不足と相まって、センターの情報提供システムを設計する上での課題であった。特に対象となった3つの所蔵機関は、総理府(当時)、外務省、防衛庁とそれぞれ行政上別組織であり所蔵資料の整理分類だけでなく公開方法についても独自の方法をとっていた。これらの資料を有識者の提言にあるようにマイクロフィルムで収集し目録のみをインターネットで公開しただけでは、利用者にとって、各所蔵機関を訪問する代わりに同じ東京にあるセンターを訪問するだけであり、利便性はあまり向上しないと考えられた。そこで何らかの形で資料を統合し横断検索が出来るだけでなく、「いつでも」「どこでも」「だれもが」資料を利用できるようにと提案されたのがデジタルアーカイブである。

3. 画像提供システム

3.1 マイクロフィルムからのデジタル画像変換



図表1：アジア歴史資料情報提供の流れ

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る。当初、原本から直接デジタル撮影

を行うことも検討された。しかし、調査の結果、当時（2000年）の状況では直接デジタルカメラによる撮影よりも技術的に確立しているマイクロフィルム化を行いスキャナーによって機械的にデジタル化をしたほうがコスト的にも品質的にも優位であることがわかった。特に資料の状態や質、サイズが不均等である歴史資料には、オーバーヘッド型デジタルカメラでの撮影となるが、現状でも年間200万画像を処理するに十分な撮影機材や技術者を確保することは困難である。

さらに対象となる資料は、紙の劣化、破損、変退色によって紙面が不要な濃度を持っているために閾値の決定が困難であることが判明した。そこで一旦撮影技術が確立しているマイクロフィルムで撮影し、閾値を平均化、その後マイクロフィルムを機械的にスキャナーでデジタル化するほうがはるかに省力化できることが実証された。デジタル機材がさらに進歩した現状では直接デジタル化も検討の対象となっている。しかし、急速に進歩する画像圧縮技術で問題となってくる知的財産所有権・工業所有権等による画像ファイルフォーマットの将来性（データマイグレーションコスト等）を考慮すると、確立されたメディアであるマイクロフィルムで原資料の画像を保存しておくことはコストもかかり原資料を傷める再撮影を回避する上でも有効と考える。

3.2 画像ファイル仕様の決定

画像ファイル作成について、どのような解像度や階調の画像ファイルを標準とすべきかが検討された。画像ファイルの仕様は、出来る限り原本に近いことが望ましいことは言うまでもない。当然、高精細な画像ファイルはデータサイズが大きくなる。しかし、当初想定された3000万点近い画像データの保存、バックアップ、インターネットでの提供を考慮すると、設計当時の一般的な記憶媒体の容量やインターネットへの接続速度を考えると、出来る限り高度圧縮が可能な方法が不可欠であった。また、資料一点あたりのページ数も多いものでは数百ページにわたるものがあり、この点からもデータサイズを小さくすることが必須であった。（注2）

そこで、画像ファイル仕様決定にあたっては、歴史研究に耐えうる階調と解像度を歴史研究者や画像技術者が協力して各種の組み合わせの比較検討を行い、文字が研究用途として原本に忠実に判読可能であり、サイズを出来る限りコンパクトにする解決策としてモノクロ2値、400dpiを標準とした。画像ファイル形式については、画像ファイル作成業者からの納入形式は機種や特定ソフトウェアの依存性の少ない（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。インターネットでの公開形式は次に紹介するDjVu形式を採用した。

3.3 高度画像圧縮技術の検討

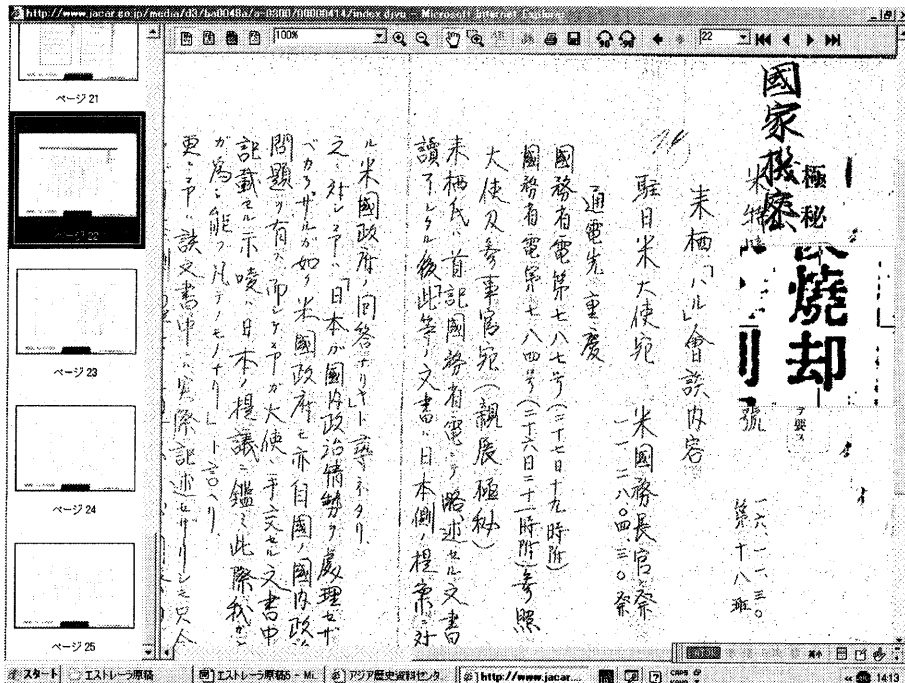
各所蔵機関から提供されるTIFFフォーマットは、インターネットで標準の閲覧用ブラウザ（Internet Explore, Netscape）が対応していない。さらにTIFFのままではデータサイズが大きく、インターネットでの公開は不可能である。そこで当時まだ実証事例がなかったが多のファイル形式と比較検討の結果、圧倒的な圧縮比を示した米国のAT&Tが開発した超高圧縮フォーマットDjVuを採用した。（注3）本形式はモノクロ2値の文書情報圧縮ではJPEG形式の10~20倍の圧縮率となり、現状でも最高レベルの圧縮性能を持っている。画像ファイルを白黒の文字画像部分（前景）とその他カラー画像部分（背景）に分解し、それぞれの部分に最も適合した圧縮方式を使用するため、非常に高い圧縮率が実現可能となっている。特にデータ量が少ない文字部分は可逆圧縮がされており文字が判読出来る状態での高度圧縮が可能でセンターの文字圧縮の機能要件に適合する技術である。画像を見る場合は、これらのファイルを見る側の専用ビューアー（無料で提供）で合成して表示する。

約 14cm × 21cm (B5 より一回り小さい程度) のサンプルを 300dpi, フルカラーでスキャンすると以下のようなファイルサイズとなる。DjVu 形式は一般的に画像を 1/100~1/1000 にまで圧縮可能であるとされている。(図表 2 参照)

画像形式	ファイルサイズ
非圧縮 TIFF イメージ	約 11.5MB
JPG イメージ	約 640KB
DjVu イメージ	約 66.8KB

図表 2：ファイルサイズ比較

DjVu 形式の画像を表示するには専用ビューアーが必要となる。専用ビューアーは単体でも動作するが、ブラウザにプラグインとしても組み込むことが可能であり、ホームページ閲覧時においても違和感なく DjVu 形式画像を閲覧することが可能である。ビューアー画面はサムネールの表示・非表示や拡大縮小印刷などに対応し虫眼鏡の用に一部分を拡大する事も可能である。その結果、公文書綴りの内容をブラウジングする感覚で見ることが出来るようになっている。(図表 3)



図表 3：DjVu 画像サンプル：サムネール機能 (画像左)、虫眼鏡機能 (「焼却」の拡大文字) 等を示す
(外交史料館所蔵：レファレンス番号 B20020307503)

4. 情報検索システムの概要

4.1 ISAD(G) に準拠した整理分類と Dublin Core を基本とした目録項目

目録項目は、記録資料を検索する上で重要な指針である。歴史研究者の中に「目録作成作業は困難ではあっても、史料学の専門家にゆだねる必要がある」とする意見がある。(注4)しかし、史料専門家の絶対的不足と方法論の未整備という現実がある。そこで提案されたのがセンターの情報提供システムは、目録内容と画像データと合わせて目的の資料にたどり着くためのツール(検索システムの一要素)と見なすという考え方であった。

まず、国際公文書館会議(ICA)が提唱する「国際標準記録史料記述：一般原則」(ISAD(G))(General International Standard Archival Description)(注5)とわが国の公文書整理の基本単位である簿冊(主題別や時系列に整理され綴じられたもの)を基本の共通単位として7階層からなる共通整理分類体系を設定した。7階層の構造は「目録データ階層構造モデル」(図表4)に示すとおりである。これによって、文書資料整理の国際的な規則となっている「原秩序尊重の原則」を壊すことなく、異なる所蔵機関の目録データの横断検索が可能となった。さらに、ISAD(G)の記述項目とインターネット対応型書誌項目 Dublin Core(注6)に我が国の文書管理の実態を考慮して15目録項目を決定した。管理項目を除き利用者が利用できるのは図表5の目録データサンプルに表示された項目である。(注7)階層構造は「機関」「出所」等と「記述レベル」((6)(7)階層)として記述されている。(図表5参照)

「アーキビスト」による要約が必要とされる「内容」のデータ作成にあたっては、各資料の先頭から300文字程度を原文のまま抽出することを原則とした。(図表5「内容」部分を参照)手書き文書や書体が異なるためにOCRによる読み取りは実用的なく採用せず、人的作業で手間や費用はかかるが、専門家の手を煩わすことなく内容検索対象となるデータを大量に増やす事が可能となった。勿論、300文字以外に重要な語が存在する可能性がある。しかし、これまでのアンケート等の回答から目録の精度を上げるための手間をかけるより、公開を優先すべしとの意見が多くこの方法は支持されている。

目録データの記述にあたっては将来的にXML形式等の標準化に対応できることとしCSVファイルでの作成とした。XML形式採用は時期尚早として見送った。目録作成作業は毎年20万件近い量を処理することを考慮して基本的に専門業者へのアウトソーシングである。このように大量の文書目録を専門業者が行った事例はなく手書き文字の判読から目録項目の抽出に当たっても歴史専門家からはかなり厳しい批判もあるが公開を優先し誤字脱字や項目内容の間違え等については、ホームページに「不具合情報」欄を設置し利用者の協力を得て修正を加えている。

4.2 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、自由かつ簡易に目的とする資料(当時の歴史用語で記述されている)を検索できるようにするための機能である。既存の歴史資料の検索システムを検討した結果、データ検索の対象となる目録のデータ量を増やしても、現在利用者が使用する用語と実際に公文書に使用されている用語とが乖離しているために当時の用語を知らない限り検索できないことが明らかになった。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45件しかヒットしない。しかし、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」で検索すると8103件がヒットする。このような歴史用語と現代用語を結びつける特殊な歴史用語辞書が必要であることが明らかとなった。問題は、歴史用語を編纂すること自体が歴史認識の問題と深くかかわっているために、いかに中立性を保って「辞書」を作成するかが課題であった。そこで「太平洋戦争」「日中戦争」のような現在利用されている用語を基本語とし、約5000語を典拠が明らかな歴史事典や教科書用語集から同義語

レベル	階層名	各所蔵機関の目録レベル		
①	Super Fond	アジア歴史資料センター		
②	Fond 所蔵機関名	国立公文書館	外交史料館	防衛研究所図書館
③	Sub-fond 出所	(例) 内閣	外務省記録 調書	陸軍 海軍
④	Series シリーズ	太政類典 公文録 公文類聚 公文雑纂 その他	旧記録 明治・大正期8門式分類 「1門 政治/1類帝国外交」 新記録 昭和戦前期16門分類法	陸軍省大日記類 陸軍資料 海軍省公文備考類 海軍資料
⑤	Sub-series サブシリーズ	各シリーズの 細分類項目 「外国交際」 等	項 「1項 一般政策」等	大日記類79シリーズ 陸軍資料75シリーズ 海軍公文備考類18シリーズ 海軍資料42シリーズ
⑥	File 簿冊	・簿冊名 「太政類典・ 第二編～」等	・ファイル(簿冊)件名 「幕末外交関係雑件」等	・簿冊名 「密大日記明治27年」等
⑦	Item 件名	・簿冊内の個別 件名 「清国二韓川国書」	・ファイル内の件名や細目 次等「遣欧使節一件」等	・簿冊内の個別の件名 「参謀本部より軍事停車場建設の 件」等

図表4：各階層と各所蔵機関の目録構造比較対照表

や関連語を抽出した。その後、事典の解説内容や入手できた外務省と公文書館の所蔵資料の件名データ約30万件のデータの形態素解析を行い日本近代史、近代アジア史などを専攻する大学院生の協力を得て同義語、関連語、英訳からなる辞書を作成した。昨年度は、基本語600を追加、現在、より検索精度を上げるために辞書内容の見直し作業を行っている。

4.2.1 同義語

同義語は、当該語句を置き換えても文書(概念)の意味が変わらないものである。アジア/亜細亜/亞細亞、柳条湖事件/柳条溝事件(誤字であるが歴史的に使用されている)/柳條湖事件のように表記の違い。南満州鉄道株式会社/満鉄のように略称、日露講話条約/ポーツマス条約のような異称などがあげられる。

4.2.2 関連語

関連語は、当該語から類似・関連・連想される語で普通名詞および固有名詞を指す。例えばある事象(真珠湾攻撃)から連想される場所(真珠湾)、組織(アメリカ太平洋艦隊、連合艦隊)、日時(昭和16年12月8日または米国時間の同7日)が関連語となる。

4.2.3 英語等

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。Pacific Warで例を取ると次のようにシステムは機能する。英語検索シス

目録	
	ヘルプ
	English
	レファレンスコード: A20030100497
表題:	大東亜戦争ノ呼称ヲ定メタルニ伴フ各法律中改正法律ヲ定ム (公文類聚・第六十六編・昭和十七年・第八十四卷・軍事一・陸軍・ 海軍・防空・雑載)
作成者:	内閣
作成年月日:	昭和17年02月10日
内容:	両院ノ議決ヲ経タル大東亜戦争ノ呼称ヲ定メタルニ伴フ各法律中改正法律案ヲ審査スルニ右ノ衆議院議長上奏ノ通裁可ヲ奏請セラレ 可然ト認ム 上諭案 朕枢密顧問ノ諮詢ヲ経テ帝國議會ノ協賛ヲ経タル大東亜戦争ノ呼称ヲ定メタルニ伴フ各法律中改正法律ヲ裁可シ並 ニ之ヲ公布センム 御名御璽 昭和十七年二月十七日 内閣総理大臣 法律第九号(上奏ノ通) 衆議院ノ議決ヲ経タル大東亜戦争ノ呼称ヲ定メタルニ伴フ各法律中改正法律案ノ裁可ヲ奉請ス 衆議院議長田子一民 衆議院書記官長大木操 勅令ヲ以テ別段ノ定メ為シタル場合ヲ除ク外各法律中
先頭300文字程度	
ISAD(G)の階層に対応	
機関・出所・シリーズ・サブシリーズ:	国立公文書館 内閣 公文類聚 軍事
記述レベル:	件名
組織歴/履歴:	内閣
複製の存在:	ru:647041132
資料の使用言語:	ja
規模:	33

図表5：目録データサンプル

テムの同義語辞書を展開すると英語の歴史用語（Pacific War）に該当する日本語の基本語「太平洋戦争」の同義語（大東亜戦争）で日本語の目録データ（先頭300文字程度の内容も含む）も検索する。（詳細は次項）

5. 情報提供システム設計の基本機能要件と仕様の考え方

センターの情報提供システムを具体化するにあたり決定された基本機能要件は、「いつでも」「どこでも」「だれもが」「無料」でデジタル化された画像データを自由に利用出来ることであった。また、「歴史認識」の問題とも絡み、資料検索の中立・客観性をいかに確保するかが検索システムを設計する上で重要な課題であった。そこで具体的なシステム設計に当たっては、それぞれの基本機能要件を次のように定義した。

5.1 「いつでも」「どこでも」

インターネット 24 時間接続による情報提供でこの要件を満たすこととした。そのために高速回線の確保と高度画像圧縮技術を採用したことは既に述べたとおりである。また、インターネットでの提供を前提としたことで使える文字コードを原則インターネット標準の Shift-JIS とした。さらに、原則 24 時間 365 日停止させないことを基本的な要件とし次のように対応した。

- 画像、目録データの追加・修正等のオンライン作業が可能なシステムの導入
- ハッカー等の攻撃から守るための 24 時間の監視体制の導入

- 事故等によるサービス停止を最小限にするためのシステム上および組織的な対応
- 回線の2重化等によるインターネットとの接続の冗長性の確保

5.2 「だれでも」

閣議決定に「国民一般及び関係諸国民の利用」を容易にするため、基本機能要件として「だれもが」容易に利用できることとした。しかし、提供するのは既に紹介したような歴史的公文書で、これまで公開されてはいたが一部専門家が利用するに止まっていた。資料利用の最大の障害は既に指摘したように目録や整理分類の不備であった。そこで今後の展開も考慮して中高生も含めた一般利用者と研究者をサービスの対象として想定、さらに「関係諸国民の利用」を念頭に、英語による検索手段の可能性・有効性を計るための試行システムを検討した結果、次のような機能が盛り込まれた。

5.2.1 5つの検索システムの導入

一般、学生、歴史研究専門家、外国人研究者など多様な利用者を想定して次の5つの検索システムを提供している。

(1) 階層検索

対象となる資料は、各所蔵館独自の方法で整理分類されており、取替えて、それらを統一的に整理し直すことは文書資料の取り扱いの基本原則となっている「原秩序維持」を壊すことになる。そこで各所蔵館の資料を国際標準記録史料記述ISAD(G)が提唱する階層構造を応用し、アジア歴史資料センターを頂点(スーパーフォンド)とする7階層(レベル)からなる体系にそって各所蔵機関の資料を検索できるようにしたものである。図4に沿って説明する。国立公文書館(レベル(2)所蔵機関名)を例にとると「内閣」(レベル(3)出所)→「太政類典」(レベル(4)シリーズ)→「外国交際」(レベル(5)サブシリーズ)→「レベル(5)のサブシリーズに含まれる簿冊のリストの表示。例えば太政類典第二編等」(レベル(6)簿冊)→「選択した簿冊に含まれる件名リストの表示」(レベル(7)件名)となる。これによってサブシリーズ単位や簿冊単位での閲覧が可能になった。

(2) キーワード検索

インターネットの検索サイトで一般的に利用される自由語検索に辞書機能と年代域での絞り込み機能を付けた一般利用者を念頭に置いた検索システムである。3館が所蔵する資料を(7)レベル(件名)で横断的に検索することが出来る。歴史用語に不慣れな一般の利用者も辞書機能を利用することで容易に当時の資料を検索することが可能となっている。例えば「太平洋戦争」(現在の歴史用語)で検索すると45件、同義語辞書(大東亜戦争、大東亞戦争等)を展開して検索すると8198件(いずれも9月末現在の検索数)となる。

(3) キーワード詳細検索

検索する所蔵機関の自由な組み合わせ(1館または複数)、年代域での絞り込み、さらに検索項目(すべて、表題、作成者、内容、組織歴・履歴)からの選択、辞書機能の展開および同義語・関連語の選択など多くの検索絞り込み機能を備えており研究者を対象とした検索システムである。

(4) レファレンスコード検索

レファレンスコードは(7)レベル(件名)の一件ごとに付与された半角英数文字12文字の識別記号(例: B20020307503)で該当資料を直接検索する方法である。論文への引用や既知の資料を利用する場合の便利

性を考えて追加された機能である。検索結果から各資料の「前資料」や「次資料」への移動が可能となり、前後に関連資料がないか探すことが可能である。

(5) 英語検索

日本語検索環境がない利用者（基本的には海外の日本研究者を想定）に対して検索を可能にするための英語検索システムである。階層検索、キーワード検索、レファレンスコード検索を提供している。英語目録には簿冊名、件名などの基本情報が英訳されている。ただし、日本語目録にある「内容」はコストなどの制約があり英訳していない。しかし、英語件名のみでは検索対象となるデータが不足し日本語検索結果とずれが生じる。例えば、「Pacific War」で検索すると10件しか検索されない。これは「Pacific War」に対応する用語である「太平洋戦争」が英語訳の対象となる件名（表題）に10件しかないためである。そこで辞書を展開すると「Pacific War」の訳である基本語「太平洋戦争」の同義語辞書に含まれる「大東亜戦争」等で日本語目録データを検索するようになっていた。その結果、7760件が検索される。日英の目録構造の違いから日本語検索で同義語辞書を展開した場合と英語検索で辞書を展開した場合に若干の差が生じるが95%近い検索結果を得ることが可能となった。今後は、検索に頻繁に利用される英文用語に対応する日本語の基本語を増加させることで辞書展開による英語検索の精度を上げるように努める。

5.3 「無料」で

基本的に提供する資料は公文書のため著作権はない。インターネットでの提供を原則とし資料が広く利用されることが目的であることから「無料」での提供が原則となった。また、「無料」にすることで利用者の利用承認も不要とし、自由に利用出来る環境を確保することで利用者の匿名性を確保した。さらに、利用制限も最小限に止め、改ざんや不当な複製の販売などに対しては、システム上のセキュリティーを強化するだけでなく、無料でオリジナル画像データをいつでもだれもが自由に利用できることで対応することとした。インターネット上での画像や情報提供について「知的財産」を保護しようとするのが潮流ではあるがインターネットの本来のあり方である「自由」に利用できることを基本とした。(注8)

6. センターの将来展望と課題

6.1 急速な技術革新との競争

センターは本格的なデジタルアーカイブとして内外の注目を集めている。しかし、試行テストを行うことが出来なかったために最初の1年間はシステムの改善と作業手順の見直しに費やされてきた。主な改善箇所だけでも次のようなものがあげられる。

- 英語検索システムの導入：日本語環境がなくても利用が可能となった。
- レファレンスコードによる検索および検索資料の前資料、次資料移動を可能とする：既知の資料への直接アクセスが可能となった。その資料の前や次の資料が何であるか確認が可能となり、資料の原秩序を確認できるようになった。
- 「初心者のための利用方法」の掲載（和英）
- 中国語とハンダ語によるセンター解説ページ：検索システムまで各国語で対応するかはコスト、手間、必要性について調査中。とりあえず各国語での簡単な解説ページを提供。
- 「不具合情報（誤字脱字を含む）フォーム」の設置：利用者の声を恒常的に反映させるためと協力を得るために設置。

- 「NOT」検索の導入:「インド」NOT「インドネシア」での絞込みが可能となり不用なノイズを排除することが出来るようになった。
- 検索使用用語の表示:検索に利用した言葉を確認することが可能になり絞込みの用語選択が効率的に出来るようになった。

しかし、予想以上に早いブロードバンドの普及は、画像数が600ページを超えるファイルの提供をダイヤルアップアクセスの限界である50画像程度に分割することなく提供することが可能となった。そのため、現在画像スペックの基本仕様となっている400dpi、2値のままでもよいのか検討する時期に来ている。

また、ウィンドウズXPが組織的に導入されたところでは、管理者権限が強化されたためにこれまで自由にインターネットからダウンロードが可能であったDjVuビューアーのダウンロードが出来なくなったとの苦情が報告されている。また、画像データを保存するストレージ容量が巨大化しておりコスト的にも安価になってきている。画像仕様全体について見直すことも視野に入れて検討する時期が予想以上に早い段階で来ている。

知的所有権保護の下に囲い込みが進んでいる最先端技術に頼りすぎるのは、将来的な画像データ等のマイグレーションコストの必要性を考慮すると公的な機関として無料でサービスを提供するセンターにとっては重大な問題である。将来的には、システム全体をオープンな規格を基本とする方向性も検討対象となろう。

6.2 デジタルであることによる課題

先頭300文字程度の資料内容をデジタル化することで、これまで件名や付与された「キーワード」検索では出来なかった資料検索が可能となった反面、想定していなかった問題が発生した。提供している資料は、現在の情報公開法で開示請求の対象となる現用行政文書ではなく、歴史資料として基本的には公開が原則の資料である。しかし、インターネット等の情報技術の発達で注目されている個人情報保護の問題を検討する必要がある。現在でも犯罪歴、病歴、戸籍などについては各所蔵機関の公開原則に沿って必要に応じて該当箇所のみをマスキング（黒く塗りつぶす）が行われている。しかし、歴史資料に含まれる個人情報をどこまで保護すべきかについて十分な国民的な合意があるとはいえない状況である。そのためこれまで所蔵機関にいても探ることが出来なかった個人に関する情報がたまたま先頭300文字に個人名が含まれていたために検索することが可能になった。これは現在、米国でも裁判記録をインターネットで公開することの是非で問題とされている点である。つまり、「時の経過」によって忘れられていた過去の犯罪記録がインターネットで検索できるようになったことで利害関係もない不特定多数の人が「時の経過」で忘れられた事実を知ることが出来る点である。祖父の戦争での自慢話を確認するつもりが犯罪記録に出くわすこともありうる。このような問題に対して検討していく必要も出てきている。電子化された社会での人権保護と情報公開という新しい問題である。

センターの情報提供システムの特長は、利用者の利便性を考えた多様な検索手段と大量の画像データ配信機能を持つ目録と画像データを合わせることでこれまで不可能であった歴史資料の探索を「いつでも」「どこでも」「だれもが」「無料」で出来るようにした点である。開設以来、可能な限り利用者がセンターのサービスを継続的かつ安定してつかえるように常に利用者の視点に立ち情報技術の発展を踏まえてシステム改善に努めている。ホームページへのアクセス件数は開設当初の一日平均220件から現在では800件を超えるまでになった。開設以来1年8ヶ月のアクセス累計も25万件を超えている。提供するコンテンツ自体は一般に馴染みがない歴史的な公文書であるがその重要性は着実に理解されてきている。

注

注1：内閣府の研究会及び「中間報告」については<<http://www8.cao.go.jp/chosei/koubun/>>を参照。

注2：牟田昌平、小林昭夫「アジア歴史資料センター」『情報管理』（科学技術振興事業団、vol.45, No.7, 2002）

注3：DjVuの技術的な詳細については<<http://www.lizardtech.com>>または<<http://www.lizardtech.co.jp>>を参照。

注4：歴史学者の立場からセンターの目録のあり方に関して問題の指摘がなされている。例としては、檜山幸夫「台湾史史料の共用化への模索」『台湾の近代と日本』（台湾史研究部会編、中京大学社会科学研究所、2003年）

注5：『記録史料記述の国際標準』（アーカイブズ・インフォメーション研究会編訳、北海道大学図書刊行会、2001年）

注6：Dublin Coreについては<<http://www.dublincore.org>>を参照。

注7：小川千代子「ISAD(G)の実装：アジア歴史資料センターの階層検索システム」『レコード・マネジメント』（記録管理学会、No.45, Nov.2002）

注8：Lawrence Lessig "The Future of Ideas: The Fate of the Common in a Connected World" (Vintage, N.Y. 2001)