

国文学研究資料館の歴史資料関係データベース
 ——大学共同利用機関附属施設の履歴の一つとして——

Archives databases at the National Institute of Japanese Literature:
 a part of the history of an organization attached to the NIJL

五島 敏芳 人間文化研究機構国文学研究資料館
 Haruyoshi GOTOH E-mail: gotoh@nijl.ac.jp

いま国文学研究資料館では、かつて同館へ付属施設のかたちで置かれていた組織から引き継いだ、いくつかの歴史資料(アーカイブズ archives, 記録史料) 関係データベースを公開している。それらは、大きく 2 つの種類に分けられる。

1 つは館外の記録史料情報に関するデータベースであり、1 つは館内(館蔵)の記録史料情報のデータベースである。前者のデータベースは、当館では歴史的に早い時期から整備され、記述の標準類を意識しながらも独自のデータ構造を持つ。

後者のデータベースについては、2002 年より電子的検索手段の事実上の国際規格 EAD, Encoded Archival Description (符号化記録史料記述) に準拠したデータを XML 形式で構築し蓄積している。今後、当館は、両者のデータベースについて、検索手段系のデータを EAD/XML で構築し、その他のデータをその内容に即した適切なメタデータ標準類により XML で構築し、公開していくつもりである。

There are several archival databases at the National Institute of Japanese Literature now. These databases contain the part that was constructed by the "Shiryokan", Department of Historical Documents attached to the NIJL. These databases can be divided into two main types: 1st types are about archival information outside our institute, 2nd types are about archival information of our holdings. The former

one originates from the union catalog of information on the location of Japanese archives that our institute has investigated since about 1970. So, these databases have an original data format. The latter one has been constructed by XML format since 2002. This database is based on EAD, *Encoded Archival Description* that is de-facto international standard for electronic archival finding aids. From now on our institute will construct both databases based on the EAD/XML format.

キーワード: 記録史料, 歴史資料, アーカイブズ, データベース, 所在情報, 概要情報, EAD, 電子的検索手段, XML.
 Keywords: archives, database, information on the location of archives, information on the guide of archives, EAD, electronic finding aids, XML.

はじめに

国文学研究資料館*は、大学共同利用機関として2番目に設置された。その設置は、最初の大学共同利用機関の設置の翌年のことであり、大学共同利用機関の歴史のなかでは古い時期のできごとといえる。同館は、教官制の組織ではあったが、その名称があらわすとおり図書館的機能を期待されていたようだ。人文社会科学系の共同利用として一般的に想起される内容は、大規模実験設備ではなく、文献資料を主とする大量の研究用データであったからだろう。文献資料の調査収集・保存・公開提供が大学共同利用機関の業務となったことや運営形態等などについて、先行組織の影響があるのではないか。私見では、その組織は、国文学研究資料館のなかへ置かれることになった「史料館」（後述）である。

もっとも、国文学研究資料館や史料館そのものの歴史については、すでに年誌類^[1]があるのでここではふれない。本稿では、これまであまりくわしくは述べられてこなかったと思われる、国文学研究資料館（とく

*National Institute of Japanese Literature

に史料館)の担ってきた記録史料(アーカイブズ archives, 歴史資料)関係のデータベースについて紹介していく。

1. 国文学研究資料館における記録史料管理

1-1. 経緯

国文学研究資料館における記録史料(アーカイブズ archives, 歴史資料)の管理は、1951年に設置された(通称)文部省史料館に起源を持つ。第二次大戦後の混乱と社会変動によって、全国的に過去の文書または記録が、散佚湮滅の危機に瀕していた。このとき歴史学(社会経済史学)研究者が中心となって^[2]、過去の文書・記録——「歴史資料」の保存運動を展開し「史料館」の設置を求め、全国の「歴史資料」のいわば緊急避難先として文部省史料館が設立された。

その後、1972年に国文学研究資料館が設置され、文部省史料館は改組によって国文学研究資料館史料館(以下「史料館」と略し総称)となった。この前後の時期から散見される(公)文書館・資料館等の新設や、1987年の公文書館法制定により、記録史料全般の管理についての専門的知識・技法とそれらを身につけた専門職員の必要性が高まった。史料館では、自らの実務経験に基づいて1952年以来実施してきた「近世史料取扱講習会」を廃止し、1988年より「史料管理学研修会」(のち「アーカイブズ・カレッジ」と改称)を開始した。以来、記録史料管理専門職員養成の社会的要請に応え、記録史料管理研究の体系化と実践的成果の公表に努めてきた^[3]。史料館は、記録史料の研究・教育・実践を標榜した日本で最初の機関であった。

2004年4月、国文学研究資料館は、国立大学・大学共同利用機関等の法人化(民営化)により史料館を合併・改組した。記録史料管理研究を進める組織としてアーカイブズ研究系が新設され、史料館の収蔵してきた歴史資料の管理は国文学研究資料館へ継承された。国文学研究資料館における記録史料管理は、旧史料館職員を中心に整備されている。また、

受講資格として所属を問わない日本国内唯一の専門職員養成研修会である「アーカイブズ・カレッジ」も継続している。

1-2. 収蔵資料の概容と管理

国文学研究資料館収蔵の記録史料は、文部省史料館の設立経緯・状況に影響を受けており、文部省史料館設置前の1947年より日本全国から収集されてきた。三井文庫や渋沢敬三に関わる比較的大きな規模のコレクション資料も、関係各所から寄贈されている。1960年代後半より、歴史資料の現地保存の考え方が普及したため、実物資料の収集を原則的に停止し、マイクロフィルム撮影によって収集してきた。近年は、新たに発生するようになった受入先のない民間の記録史料——たとえば、倒産してしまった全国的企業の記録など——について、実物資料の収集をおこなっている。現在、所蔵の記録史料は、実物が402件（50万点超）、マイクロフィルム資料が172件（5,108リール）、寄託資料が18件（8,890点）、である。年代的には、江戸時代以降の地域の資料（「町方・村方文書」や「戸長役場・村役場文書」など）が多い。

史料館では、記録史料の受入時に、まず記録史料群としての概要を把握し（「基本カード」と呼ばれたカードへ記入）、1点ないし1件ごとの簡略な目録（「購入史料目録」や、登録簿または財産目録である『学術史料蒐集簿』、カード目録など）を作成し、資料を装備し、それら簡略な目録のコピーなどを仮目録という検索手段とし、資料を閲覧提供してきた。所蔵資料の分は、順次詳細な調査をおこない、編成・記述し、解題または集合的記述を含む詳細な目録（『史料館所蔵史料目録』・『史料目録』現在82集）を作成・公刊してきた。公刊目録は、とくに史料館が記録史料研究・記録史料管理研究の中心的担い手として自覚した以後、基本的検索手段の模範型であることが求められた。これらの目録・仮目録類は、閲覧利用だけでなく、実物保存の制御手段としても用いられた。

史料館ないし国文学研究資料館は、記録史料そのものだけでなく、記録史料に関わる文献資料——記録史料関連資料も、積極的に収集してい

る。その文献資料とは、記録史料研究・記録史料管理研究やそれらの関連諸学の研究図書・逐次刊行物、全国各地の地方誌史・自治体史類、全国各地で保存される記録史料の目録類、である。これらは、一般図書・逐次刊行物と同じように管理しているが、限定された収集方針・範囲により独自の配列・分類も持っている。

1-3. 当館の記録史料情報の概容

上述のような実物の記録史料・記録史料関連資料の管理とともに、史料館ないし国文学研究資料館は、記録史料情報・記録史料関連情報の管理にも努力してきた。とくに国文学研究資料館が記録史料管理を継承してからは、史料館時代には予算上の限界で進展しなかった記録史料情報・記録史料関連資料情報の電子化、データベース化が充実しつつある。

国文学研究資料館における記録史料情報・記録史料関連情報は、2つに大別することができる。すなわち、(1)館の収蔵資料に関する情報と、(2)館外所在資料の情報である。史料館ないし国文学研究資料館では、長い間、(2)の情報の蓄積ないし電子化が、(1)のそれよりも優先されてきた。それは、1948年から1953年にかけて実施された「近世庶民史料調査委員会」の「近世庶民史料」所在調査事業^[4]を文部省史料館が引き継いだ経緯があったためである。以後、(少なくとも江戸時代以降の)記録史料の所在情報に関するセンターとしての役割が期待されている。

(1)は、構造化された形式による電子化が1996年頃から進められてきた。(1)は、さらに(1-1)収蔵記録史料の情報と、(1-2)収蔵記録史料関連資料の情報に分けることができる。記録史料関連資料には地方誌史類・目録類を含むことから、(2)の多くは(1-2)を情報源としている。(1-1)の電子化は、収蔵記録史料の概要(ガイド)集である『史料館収蔵史料総覧』^[5]作成が契機の一つにあげられる。このほか、『史料館所蔵史料目録』印刷の電算化にともなう印刷業者からのデータ取得や、同目録作成時のパーソナル・コンピュータ(以下「PC」と略)利用の普及など、技術的環境の変化が大きく影響している。

2. 館外の記録史料情報

2-1. 「史料所在情報検索システム」

江戸時代以降の日本は、文書による統治体制が全国化し、文書を必要不可欠とする社会となった。そのため江戸時代以降の記録史料は、大量に伝存した。それらの記録史料は、しかしながら第二次大戦後に急速に失われようとしたため、その以後、現在に至るまで日本国内の記録史料の所在の把握が試みられてきた。その最初が「近世庶民史料」所在調査であり、文部省史料館が1966年度まで各都道府県に原則1名の地方調査員を置いて進めてきた所在調査へ引き継がれた^[6]。

各地の「史料保存運動」により文書館が設置され、また地方誌史類の編纂が活発になってくると、文書館や地方誌史類の編纂室で記録史料の目録が作成されていく。これらの目録は、調査済みの記録史料の所在を示す基本的情報として第一に把握する必要がある。文部省史料館は、1970年に全国の都道府県立関係機関へアンケート調査を実施し、当時所蔵の目録類の倍近い件数の未所蔵目録類の情報を得た。以来これらの所在情報を基に、未調査の記録史料の所在調査と、未所蔵目録類の収集を続けてきた。史料館は、1980年に『史料館所蔵目録一覧（近世史料・郷土資料の部）』を、1992年に『近世・近代史料目録総覧』（三省堂）を公刊した。

これらの記録史料の所在情報は、第二次大戦後の早い時期から蓄積されてきた。目録類の書誌的事項だけでなく、（おそらく1980年代より）記録史料の出所または所蔵者ごとのデータシートとしても情報の蓄積が進められてきた。そのデータシートの形式は、独自の内容を持ち^[7]、1980年代以降整備された欧米の記録史料記述の手引きや規則または国際標準とおおよそ影響関係がない。これらのデータシートが、電子化され、後に国文学研究資料館から「史料所在情報検索システム」として公開されるデータベースとなる。

データシートとデータベースの項目は、後掲【表1】のようになる。

【表1】「史料所在情報検索システム」の項目

表示(検索)項目	データ要素(タグ名)
	<レコード>
識別番号	+<カード番号>
[出所に関する情報]	+<出所>
+<出所>	+<個別出所>+<出所名>
+<現住所>	+<出所現住所>*+<都道府県名><郡市名><町村名><字以下>
+<旧地名>	+<旧地名>*+<国名><郡名><町村宿名>
+<旧支配>	+<旧支配>*
+<旧職業・階層>	+<旧階層>*+<階層>
+<関係地>	
[内容に関する情報]	+<伝来>
+<所蔵者・機関>	+<収蔵者>+<史料収蔵者>+<収蔵組織名>
+<現職業>×	+<収蔵者現職業>
+<所在地>	+<収蔵者所在地>+<都道府県名><郡市名><町村名><字以下>
+<所蔵関係>×	+<所蔵関係>*
+<寄贈・寄託者>×	+<寄贈者>+<史料寄贈者>
+<寄贈・寄託者住所>×	+<寄贈組織名> +<寄贈者住所>* +<都道府県名><郡市名><町村名><字以下>
+<年代 上限>	+<内容>+<上限年代> (和暦)
+<下限>	+<下限年代> (和暦)
+<主な年代>	+<主年代>
+<数量>×	+<件数>
+<点数>	+<点数>
+<保存状況>×	+<保存状況>
+<利用状況>×	+<利用状況>
+<内容(解説)>	+<解説>
[情報源に関する情報]	+<情報源>
+<出典>	+<出典名>
+<出典請求記号>	+<出典請求番号>
+<調査年月日>×	+<調査年月日>
+<出典刊年>	+<出典刊年>
+<調査機関>×	+<調査機関名>*
+<調査者>×	+<調査者>*
	+<付加情報>+<検索上限年代>(西暦)
	+<検索下限年代>(西暦)

注)

1. 各項目または要素の間の階層的関係は、項目等の前に付した「+」によって示した。項目等の後に付した「*」は、それが繰り返し出現しうることを示す。()内は、注記である。
2. []内に示した項目は、情報の領域で、項目としての実体はない。
3. 表示(検索)項目の後に付した「×」は、検索対象外であることを示す。

住所や地名、所在など意味的位置データの充実により、出所(所蔵者)ごとの所在情報の提示を意図している。出所を中心にした設計には、例えば、つぎのような特徴がある：仮に1つの出所の記録史料群が分割されて別々の機関に収蔵され別々に目録が作成されたとしても、それらの目録がデータに収録されていれば、出所を鍵にして分割された記録史料群をデータ上で(再)集約して提示できる。なお、旧地名・旧支配・旧階層といった繰り返し出現可能な項目は、とくに前近代の記録史料の所在情報提示の便宜を図るために設けられ、前述の所在調査の経緯からの影響と思われる。

これらの項目の情報は、目録類から抽出され、データシートの紙面へ転記される。さらに、各項目に対応する簡易タグがデータの前へ付され、PC上のテキストファイルへ電子化される。データ構造や記載内容について数度の校正を経た後、データ変換工具を用いて、SGMLタグ付きテキストファイル(XML形式とも互換)へ整形される。このSGMLタグ付きテキストファイルが、「史料所在情報検索システム」の基本データである。その中の<レコード>という単位が1つの出所の情報(所在情報データ)で、この<レコード>の集合が1つのファイルへ格納される。表には示していないが、<レコード>の上位に<史料所在情報>という単位を設け、全体を包含する。所在情報データは、目録の解説・解題を全文収録することもある「内容(解説)」項目のため、しばしば長大となる。それ故に1つのファイルの容量は、巨大となる^[8]。

附表. 所在調査書の記載項目

	1948-1953「近世庶民史料調査書」	1953-1967「近世史料所在調査書」
項目	[1]現蔵地 [2]現蔵者 [3]旧地名 [4]旧身分 [5]旧支配 [6]所蔵関係 [7]所蔵状況 [7-1] (一) 公記録 (の数量等) [7-2] (二) 私記録 (の数量等) [7-3] (三) 編著書類 (の数量等) [7-4] (四) 史料の年代 [8]史料内容 [9]保存状況 [10]整理状況 [11]利用状況 [12]備考	[1]所在地 [2]現蔵者 [5]旧地名 [4]旧身分 [6]旧支配 [3]所蔵関係 [7-1]数量 [7-2] 公 [7-3] 私 [7-4] 編著 [8-1]年代 (上限～下限) [8-2] (主年代) [9]内容 [10]保存 [11]整理 [12]利用 [13]備考
欄外	[13] (調査年月日) [14] (調査員名・押印)	[14] 史料館地方調査員 (氏名・押印)

注) []内の数字は、調査書内の記載順を示す。

この基本データを用いる検索システムは、1996年～1998年に開発された(2000年に一部機能を修正)。基本データ中の検索項目を対象とした全文検索によって結果を表示し、その検索エンジンにOpenTextを使用している。インターネットを通して情報を提供するため、利用にはWWW(World Wide Web)ブラウザが必要である。このシステムは、簡易版と詳細版、また、それぞれの一般公開用と館内用(非公開データを含む管理用)に分けて運用されている。

簡易版と詳細版の区別は、利用者登録の有無にある。各地の記録史料

保存活動を阻害しうる利用^[9]を排除するため、簡易版では検索項目と提示情報を限定している。簡易版の検索項目は、出所、旧地名、年代、出典、出典請求記号だけである。詳細版は、全検索項目を利用でき、全情報を提示する。

一般公開用と館内用の区別は、インターネット上でのデータの公開可否を各収蔵機関・所蔵者へ照会した結果、設けられた。その照会結果は、法的に要求される条件への対応の必要性和、インターネットへの漠然とした不安・危機感の存在を明らかにした。データに混在する（現在の）個人情報のため、データの一部ないし全部の非公開措置は了解できる。しかし多くの目録公刊主体は、インターネットの情報流通の速度・範囲を——いわば、その目録がベストセラーになる可能性を考慮していなかったようである。日本の記録史料保存が、しばしば一過性的対応と出版物公刊の功名心により無責任に進められてきた（進めざるをえなかった）ことを示している。

各個の所在情報データは、識別番号の末尾に「:」と数字が付され、つぎのように区別された:「:0」は、簡易版・詳細版とも公開;「:1」は、簡易版のみ公開;「:2」は、簡易版・詳細版とも非公開。

このデータベースは、2001年より公開されているが、その維持や、現在の記録史料管理の研究水準から見て、多くの問題を抱えている。それらの問題は、例えば以下の通りである。(1)テキストファイルによるデータ維持は、データの長期保存性への配慮であったが、アプリケーション・ソフトウェア（以下「応用ソフト」と略）の排除は、分かりやすいインターフェイスをも排除し、小規模・機動的データ修正を困難にした。(2)公刊目録類から独自の形式へデータを抽出・編集しなければならず、しばしば特定の研究分野のみが利用する情報の生成に作業時間を割かれる。(3)データ形式が、記録史料の国際標準類と対応関係を持たないため汎用性に欠け、国際標準類へのデータ変換が難しい^[10]。(4)緯度・経度の絶対的位置情報の項目がなく、現在のデータ維持のために市町村合併等の行政区画の変更に際しての諸データ変更が必要となる。(5)記録史料群の名称が独立の項目として用意されていない。もっとも記録史料群名の命名

の全国的統一的基準は、日本に存在しない。

2-2. 「史料情報共有化データベース」

前述の「史料所在情報検索システム」の問題点を克服し、各地の記録史料情報の共有・集約を目指すシステムとして、「史料情報共有化システム」が構想された^[11]。それは、記録史料の収蔵機関または所蔵者が、自ら収蔵する記録史料とその情報を公開・維持する意志を持って参加し、共同で共有可能な記録史料情報を構築する、という内容を持つ。システムを具現化するデータベース機能「史料情報共有化データベース」は、国文学研究資料館のサーバ単体でも稼働するし、各収蔵機関等に置くことのできる複数のサーバを結合しても稼働する。いずれのサーバにも、記録史料情報の入力・編集から公開までを実現する機能があり、国文学研究資料館のサーバにのみ存在する機能は、複数のサーバそれぞれの提供する記録史料情報を横断的に検索し、結果を集約して提示する部分(横断検索機能)である。もちろん国文学研究資料館の情報システムと同等の性能があれば、横断検索機能を使用できる。

対象とする記録史料情報は、記録史料群を単位とする記述の情報であり、主として記録史料群全体の概要記述の情報である(記録史料群内の小群、1点ないし1件の各記述も扱うことは可能)。データ形式は、記録史料記述の国際標準 ISAD(G)第2版(当時の最新版)の記述要素をそのまま採用し、収蔵機関の情報等を項目として追加した(別掲【表2】)。

データベースのシステムは、はじめはリレーショナル型データベース管理システム HiRDB と検索用テキスト・データベース(全文検索エンジン) Bibliotheca2 を併用している(いまは Infolib・Oracle・OpenText へ移行した)。サーバはデータ管理に応用ソフトを用いているが、利用者は特別な専用のソフトウェアを必要としない。データの閲覧・検索・入力・編集・提供は、インターネットを通して実行されるため、WWW ブラウザさえあればよい。記録史料情報を閲覧・検索するだけの利用者は、とくに利用者登録を必要としない。ただ、記録史料情報を入力・編集・

提供したい利用者（編集ユーザ）——多くは記録史料の収蔵機関または所蔵者——は、編集ユーザとしての登録が必要である（サーバ設置機関も同じ）。

【表2】「史料情報共有化データベース」項目とISAD(G)第2版要素

項目名	ISAD(G)第2版
識別記号	3.1.1 参照記号;レファレンス・コード
史料群記号	3.1.1 参照記号;レファレンス・コード
名称	3.1.2 標題
年代上限	3.1.3 年代
年代下限	3.1.3 年代
主年代	3.1.3 年代
年代注記	3.1.3 年代
記述レベル	3.1.4 記述レベル
書架延長	3.1.5 記述単位の大きさと媒体(量, 容積または寸法)
数量	3.1.5 記述単位の大きさと媒体(量, 容積または寸法)
数量媒体等詳細	3.1.5 記述単位の大きさと媒体(量, 容積または寸法)
出所作成	3.2.1 作成者名
履歴	3.2.2 経営の/伝記的歴史
関係地	3.2.2 経営の/伝記的歴史
領主(のち主題)	3.2.2 経営の/伝記的歴史
役職等	3.2.2 経営の/伝記的歴史
伝来	3.2.3 記録史料の歴史
入手源	3.2.4 取得あるいは譲渡の直接の源泉
構造と内容	3.3.1 範囲と内容
評価選別等スケジュール	3.3.2 評価・廃棄と日程計画の情報
追加受入情報	3.3.3 追加受入
整理方法	3.3.4 整理の体系
公開条件	3.4.1 利用可能性[/公開]を規定[/統制]する条件
複写条件	3.4.2 複製を規定[/統制]する条件
使用言語	3.4.3 史料実体の言語/書体
史料状態	3.4.4 物的特徴と技術的要件
検索手段	3.4.5 検索手段
原本の所在	3.5.1 原本の存在と所在
複製の存在	3.5.2 複製の存在と所在
関連史料	3.5.3 関係する記述単位
出版物	3.5.4 出版についての注記
摘要	3.6.1 備考
収蔵名称	3.6.1 備考
収蔵名称ヨミ	3.6.1 備考
収蔵所在(都道府県)	3.6.1 備考

収蔵所在(地名)	3.6.1 備考
収蔵所在(番地)	3.6.1 備考
収蔵所在(都道府県コード)	3.6.1 備考
収蔵所在(郵便番号)	3.6.1 備考
入力者	3.7.1 記録史料管理者の注記[/アーキビスト-ノート]
記述担当者	3.7.1 記録史料管理者の注記[/アーキビスト-ノート]
更新日	3.7.3 記述の日付

注) すべてのデータベース項目が表示されるわけではない。

データとなる記録史料情報は、およそ記録史料の目録類を情報源とするだろう。ただ、直接に記録史料の実物を情報源とすることもできる。編集ユーザは、記録史料記述という記録史料管理の業務の一部に、このデータベースを活用できる。データの作成に手書きの過程が含まれるならばデータベースの項目と同じデータシートを用意してもよいが、WWWブラウザ上に入力フォームが存在するので、インターネットを使用できる環境があるならば直接入力することができる。

また、サーバ設置機関（現在は国文学研究資料館のみ）へ、データベース項目と同じ順序で適切に区切られたデータ（データベースの指示する形式のカンマ区切りテキストファイルやXML ファイル）を、一括登録するよう依頼することもできる。カンマ区切りテキストファイル等での入力・出力は、サーバ設置機関にしか認めていないためである。国文学研究資料館では、同館のサーバを利用する編集ユーザから、カンマ区切りテキストファイル等での入力・出力の要請があれば対応することになっている（編集ユーザ登録がなくても希望があれば受付けている）。2001年の公開以後、国文学研究資料館や編集ユーザを含め216の機関等からの参加を得て、4980件弱の記録史料群の概要記述の情報が、閲覧・検索可能になっている。なお、ほとんどの機関は、編集ユーザとして登録しておらず、入力・公開代行のかたちで参加している。

「史料情報共有化データベース」ないし「史料情報共有化システム」は、前述の「史料所在情報検索システム」の問題のいくつかを解決した

が、限界や問題を有している。それらの問題等は、例えば次の通りである。

- (1) このデータベースで管理する記録史料情報は、基本的に公開用であり、内部用の管理的情報との連結には、サーバ設置と別に応用ソフトや専門知識が必要となり、その費用がかかる。
- (2) このシステムへの参加は、記録史料情報の公開と維持の意志を要求するため、その意志を持たない記録史料の収蔵機関または所蔵者の参加を得られない。とくに個人所蔵者の情報は、ほとんど収録されない^[12]。また、収蔵機関において、担当者の個人的能力に大きく依存しており、担当者の交替時にデータが維持されないことがある。
- (3) 日本の記録史料管理では伝統的に目録での記述の力点だが、概要記述（解説、解題）よりも1点ないし1件にあったため、概要記述の作成に不慣れである。
- (4) 記録史料の関係機関の主要な顧客である歴史学研究者も、しばしば記録史料群の概要記述より1点ないし1件ごとの詳細記述の提示を要求し、必ずしも専門的知識を持たない一般市民への記録史料情報提供の過程として重要な概要記述提示に理解を示さない。

記録史料情報共有のための枠組みは用意できたが、記録史料管理主体各々の認識の深化を必要とする状況は、大きな課題である。

3. 館内の記録史料情報

3-1. 館蔵記録史料情報データベース

史料館は、1996年の『史料館収蔵史料総覧』作成で、ようやく収蔵記録史料の概要を斉一的に把握した。その際のデータは、簡易タグの付されたテキストファイルで、PC上のワードプロセッサ（以下「ワープロ」と略）によって維持されていた。しかし、記載データがない場合、簡易タグごと省略され、ワープロによるデータ維持が、データ構造の不定形

化に拍車をかけていた。そこで、そのデータを加工し、応用ソフト（具体的には Microsoft Access）を用いて 1998 年頃までに『史料総覧』データベース」を作成した。

『史料総覧』データベース」は、あくまで館内用の管理的データベースであったが、情報の再利用を可能にした。ちょうど史料館のホームページ作成の時期と重なり、ホームページ提供用のデータベースが求められていた。そのため『史料総覧』データベース」からデータを抽出し、簡略なデータベース「収蔵史料概要データベース」を作成した。その項目は次の通り：通し番号、文書群名、文書群記号、出所、地名、伝来、年代、旧支配、役職等、歴史、内容と構造、数量、注意；これらに都道府県コード等の制御項目を加えた。このデータベースは、都道府県ごとの収蔵記録史料群の一覧（年代順）を動的に生成し、また直接に 1 件の記録史料群の概要を表示できる。検索（抽出）は、1 語だけ可能である。CGI スクリプトを用い、検索エンジン等は使用しない、ごく簡素な機能としたためである。

「収蔵史料概要データベース」は、史料館ホームページとともに 2000 年から公開された。〈概要から詳細へ〉という記録史料の情報提示の基本形による、当館でのナビゲーション・システムの端緒といえる。しかし、館内外において、1 つまたは複数の記録史料群の資料 1 点ないし 1 件の詳細情報を自由に検索でき即物的検索結果が示されることへの欲求が高く、不十分との批判は大きかった。

一方、『史料館収蔵史料総覧』と前後する時期から、史料館が公刊してきた『史料館所蔵史料目録』の作成過程が変化してきた。その本文原稿作成には、PC の普及もあって、市販応用ソフトのデータベース管理システムが使用され、わずかながら印刷業者の電算写植用データの回収も見られた。1997 年には、記録史料への調査で採取された基礎データを編成し記述するための支援工具が、市販応用ソフトによって開発された。その支援工具は、史料館員の過半に（部分的利用も含めて）採用され、そのなかのデータベースの項目は、館内で記録史料の 1 点ないし 1 件の記述（目録本文）の最低限の要素として了解された。これによって、新規

に作成される『史料館所蔵史料目録』の目録本文のデータの形式は、ほぼ定形化された。

しかし、当時すでに日本で紹介されていた ISAD(G)でマルチレベル記述の概念が示されているが、『史料館所蔵史料目録』の解題・解説部分と目録本文との記載項目の共通化は図られていなかった。現実の電子データも、目録本文はデータベース等の応用ソフトの固有ファイルかカンマ区切り等のテキストファイルとして構造化されている一方、解題・解説部分は記載項目の区切りすらも定形化されていないワープロまたはフラットなテキストファイルだった。

解題・解説部分と目録本文とを統合する課題は、当初 Dynamic HTML でのデータ統合・変換によって解決が試みられた。しかし、記録史料群ごとに異なる構造を個別に数値化・階層化し目次・本文見出しのリンクを構築する労力は、極めて多大なため、2 件の目録を「収蔵史料目録データベース」として作成することとどまった。なお前述 2 件のデータベースは、1 点ないし 1 件の情報への検索要求に応えるため、「収蔵史料概要データベース」と同じ簡素な検索機能を持っている。

その後の「史料情報共有化データベース」の開発経験は、記録史料情報提供における ISAD(G)利用が実現可能であるとの認識をもたらした。さらに館蔵の記録史料群から得た ISAD(G)準拠の記述例を用いた実験により、極小のソフトウェア構成の PC 環境においては XML と XSLT の原初的利用で十分マルチレベル記述を実現しえることを確認した。

しかし、ISAD(G)の記述要素またはそれを一部拡張した「史料情報共有化データベース」の項目だけでは、検索利用を想定したときに不十分であることも確認した。1 つまたは複数の記録史料群の全体・各小群・1 点ないし 1 件の全記述単位にわたる検索・抽出・並び替えを「史料情報共有化データベース」で実験すると、元の出所を追跡することが困難な結果が集約されて示された。断片化された各記述単位のデータを、各記録史料群の階層的構造に即した脈絡へ再配置するための管理的データが欠如していたのだ。ISAD(G)は、あくまで考え方の枠組みでしかなく、記録史料記述を表現した検索手段、それも電子的に表現された検索手段のた

めには、ISAD(G)とは別の枠組みが必要であった。

そこで、既に記録史料の電子的検索手段の事実上の国際標準として欧米を中心に普及している、EAD, Encoded Archival Description (符号化記録史料記述)をXMLの利用により適用する試みに挑戦することになった。その試行錯誤の末、2002年5月、おそらく日本では初めて日本の記録史料記述に対するEAD適用とその初歩的表現に成功し、館のホームページから公開した。以後いくつか類型が存在する館蔵記録史料の目録のうち、簡略な一覧表やカード形式の仮目録類を中心にEAD/XML事例を増やしていった。仮目録類は、公刊された目録と異なり来館しなければ見ることができないためである。当初は、XMLファイルそのままをXSLTスタイルシートによって表示オブジェクトへ変換し、目録の解題・解説部分から1点ないし1件までの全文を提供していた(HTML変換済みファイルも提供)。

EADは、群全体から1点ないし1件までの記述データ全てを、その記録史料群の階層的構造に即して、階層的に格納できる(別掲「(参考)EAD最小必要タグとISAD(G)第2版記述要素」参照)。ここに至って、ようやくデータ構築的には、解題・解説部分と目録本文を統合できる道筋が開けたのだ。当初「収蔵史料概要データベース」の下に「収蔵史料目録データベース」それぞれを配し相互にリンクを張って、(概要から詳細へ)と導く体系を構想していたが、それぞれのデータを別個に生成・維持することは、もはや不要である。解題・解説部分も目録本文も、ともに同じEAD/XMLデータへ格納すれば、XSLTの本来的機能を使用して、概要から詳細までの各記述レベルのデータを抽出・加工できるためである。

そのXSLTの表示オブジェクト以外の変換機能を埋め込んだスタイルシートとPHPを活用して、まず2003年に概要と全文だけを表示できるシステムを開発し、ついで2004年に概要から詳細まで絞り込んでいくように特定の記述データを表示できるシステムを開発してきた^[13]。この時点で、さしあたり「収蔵史料概要データベース」の検索機能による検索結果から当該EAD/XMLデータの部分・全文の表示へ画面が遷移するようにし、2005年から「収蔵アーカイブズ情報データベース」と改称して、公

開・提供している。さらに2006年から、1つの記録史料群内、または複数の記録史料群におよぶ、自由な検索機能が実現した(これについては、稿をあらためて詳述したい)。

このほか記録史料画像については、「収蔵史料画像データベース」として錦絵のみを抽出した電子的コレクションのデータベース(後の日本実業史博物館設立準備室旧蔵絵画資料データベース)を予定していたが、2005年に至るまで公開できなかった。公開の遅れは、技術的問題点にもよるが、錦絵等の資料を含む現実のコレクションからの位置付けをなかなか提示できなかったことも理由の一つに挙げられよう。なお前述の絵画資料データベースは、市販応用ソフトFileMakerPro6 Unlimitedを使用し、webコンパニオン機能によりインターネット上へ公開している^[14]。

電子的コレクションのデータベースよりむしろ、記録史料の概要から詳細までの記述が提示される中で(主に)1点ないし1件の記述に対応する記録史料画像へリンクするほうが、画像データの提供形態として相応しい。この目録からの画像へのリンクは、既にEAD/XML適用実験において成功している。現在EAD/XMLデータのナビゲーション的表示システムのなかで、画像データが存在する一部の記録史料群では、その画像データを公開している。これらの画像データは、いずれも調査時の現状記録を兼ねた記録史料画像であり、記録史料管理のなかで発生しうるデータを公開用に再利用したものである。今後は、こういった画像の中から一般の興味を引くような分を抽出し、電子的展示または電子的コレクションのデータベースを構築する手順を確立したい。

(参考)EAD 最小必要タグと ISAD(G)第 2 版記述要素

配置順など	EADタグと説明 (太字は必須タグ)	ISAD(G)第2版との対応
1	1 2 3 4 5...15+ -タグ階層の深度→	
1	<ead>EADタグ	
2	<eadheader>EADヘッダ	
2-1	<eadid>EAD識別子 (EADファイル識別子)	3.1.1 参照記号
2-2	<filedesc>ファイル記述 <titlestmt> (ファイル全体/検索手段の) 標題表示 <titleproper>検索手段の本標題, <author>著者 (ファイル作成者/組織) <publicationstmt>出版表示 <publisher>出版者, <date> (出版) 年代 <notestmt>注記表示	
2-3	<profiledesc>概観記述 <creation>作成 <date> (EAD化) 年代 <language>言語使用 <language> (記述の) 言語 <revisiondesc>改版記述	
3	<frontmatter>前付付 <titlepage> (ファイル全体の) 標題頁 (とびら) <titleproper>検索手段の本標題 ... (適切な下位記述要素) <div>テキスト区分 (本文章節区分) ... (凡例・解説など; 適切な下位記述要素)	
↑順序変更 ↑可能		
4	<archdesc level="~" >記録史料記述; 記述レベル属性ほか	3.1.4 記述レベル
4-1 fonds記述	<did>記述識別 <repository>収蔵 <corpname>団体名 <originatoin>出所 (または作成者) <persname>人名, または<corpname>団体名, または<famname>家名 (家族名) (いずれが適切なもの) <unittitle>記述単位標題 <unitdate>記述単位年代 (<unittitle>への内包もあり) <physdesc>物的記述 (物的状態記述) <extent>数量, <dimensions>寸法, <genreform>類形/物的特性... <unitid countrycode="~" >記述単位識別; 国コード属性ほか <abstract>要約 (<descgrp type="admininfo" >管理情報 (-) <accessrestrict>閲覧条件 (-) <userrestrict>使用条件 (-) <acqinfo>取得 (-) <altformavail>利用可能な形式の選択肢 (-) <custodhist>記録史料保管の歴史 (-) <processinfo>資料操作情報 (-) <appraisal>評価に関する情報 (-) <accruals> (記録史料の) 追加受入 ... (適切な下位記述要素)	3.2.1 作成者名 3.1.2 タイトル/標題 3.1.3 年代 3.1.5 単位の大きさと媒体 (量、容積、寸法) 3.1.1 参照記号 3.4.1 公開条件 3.4.2 複製条件 3.2.4 取得の直接の源泉 3.5.2 複製の存在と所在 3.2.3 記録史料の歴史 3.7.1 アーキビスト注、3.7.3 記述の日付 3.3.2 評価、廃棄と日程計画 3.3.3 追加受入
↑順序変更 ↑可能		
	<bioghist>履歴または歴史 <scopecontent>範囲と内容 <arrangement>編成 (<scopecontent>への内包もあり) <controlaccess>接近制御標目 (<descgrp type="add" >付加的記述 (-) <relatedmaterial>関連資料 (-) <bibliography>参考文献一覧	3.2.2 経営の伝記的な歴史 3.3.1 範囲と内容 3.3.4 整理の体系 3.5.3 関連資料 3.5.4 出版物についての注記

注)EAD アプリケーション-ガイドライン 1.0 版「付録 A: 最小の推奨検索手段要素」より作表(上掲は部分). EAD2002 版の一部を反映.

3-2. 記録史料関連図書等目録データベース

史料館は、1983年から記録史料に深く関わる都道府県史・郡史といった地方誌史類の一部を公開し、1994年から全ての地方誌史類を、翌1995年には関連の研究雑誌や紀要等の逐次刊行物を公開した。2005年に記録史料やその関連諸学の研究図書類を公開し、旧・史料館の記録史料関連資料の全てを一般に提供できた。これらの公開は、とくに地方誌史類の全面公開以降に蓄積されてきた図書等目録データベースや逐次刊行物目録データベースに大きく依存している。

逐次刊行物目録データベースは、その雑誌等の単位で把握できるため、件数2,800弱と、比較的数が限られる。しかし、文部省史料館以来の蓄積のある図書等の資料は、56,000余冊と膨大である。多数の不備が残るもののデータベースへ登録しえた理由は、国文学研究資料館の民営化時に作成された財産目録の活用による。その目録の作成には、極めて膨大な労力が費やされた。

これらのデータベースは、日常業務に使用する管理用で、市販応用ソフトを使用していた。しかし、ここ数年の間に、図書等目録のデータを館内だけでも共有する要求があり、市販応用ソフトの中でも比較的容易にインターネット/イントラネットへ公開できるFileMakerProの使用へ転換した。これにより、館内に限られるが、記録史料関連図書等目録データベースとして公開が可能になった。

より直接に記録史料のデータベースの検索結果との相互リンクは、今後の課題であるが、最後にふれる文化資源共有化システムの利用は解決策の一つと考える。

4. EAD・XMLによる展開構想：むすびにかえて

館内の記録史料情報の根幹を成す館蔵記録史料情報データベースにおいて、そのデータ形式にEAD/XMLを採用した結果、そのデータを多様に変換し活用できることが判明した。EADの情報の要素は、ISAD(G)の記述

要素の全てを含み、相互の情報要素の対応表 crosswalk が用意されている。「史料情報共有化データベース」は、そのデータ項目に ISAD(G)を採用しているため比較的容易に EAD のデータへ変換できる。また、「史料所在情報検索システム」は、そのデータ項目が国際標準類の記述要素と相当に異なるものの、SGML/XML 形式のデータとなっているため、項目の適切な対応表が用意できれば、XSLT による変換が可能であると推測する^[15]。以上から、館内外の記録史料情報のほとんど（図書等関連資料は除く）のデータが、EAD/XML 形式へ統一される可能性を指摘できる。

その形式的統一または規格化の効果は、現在収蔵記録史料の目録を EAD/XML 化する手法の汎用化や、データ表示・検索システムまたはデータ管理システムの単一化・共通化、等などに表われるだろう。単一の規格へ依存することの危険性は、データ管理システムに使用される応用ソフトを除けば、低いと考えてよい。基本的に XML で構造化されたテキストファイルであるから、情動的には長期の保存が期待できる。館外の記録史料情報の EAD/XML 化は、仮にデータが館外へ移動したとしても、当館のデータと合わせて横断的検索が可能であるかもしれない。

さらに、EAD がアーカイブズの世界の代表的メタデータとして他の世界——とくに図書館や情報学の世界に認知されているので、ネットワーク上の情報資源の 2 次情報集約・交換のために了解されたメタデータのセット DC, Dublin Core Metadata Elements Set (ダブリンコア) との項目対応表を通して、広く他の分野の情報資源との共有の可能性が生まれよう。実際に国文学研究資料館では、文化情報資源の共有化を DC と Z39.50 を用いて実現している。DC と Z39.50 の間には、各領域のメタデータを仲介させ、検索時のノイズの低減を図っている^[16]。このシステムへ、記録史料情報の諸データベースのデータが提供される予定である。

【付記】

本稿は、もと「韓・日 歴史資料情報化関連機関専門家 FORUM」（韓国・国史編纂委員会，2005年11月23日－26日）の予稿として用意した内容に、加筆・修正したものである。文中の一部について、最新の情報に訂正しきれていない可能性があることを、お断りしておく。

参考文献

- [1] 国文学研究資料館については、つぎの3誌をあげる：『十年の歩み』（国文学研究資料館〔以下2誌も同じため略す〕，1982年），『国文学研究資料館の20年』（1992年），『国文学研究資料館30年誌』（2002年）。史料館については、つぎの2誌をあげる：『史料館の歩み四十年』（国文学研究資料館史料館〔以下同じため略す〕，1991年），『史料館の歩み50年』（2001年）。
- [2] 過去の文書・記録は、日本では明治時代以降しばしば歴史学の関係者により「史料」として対象化され保存されてきた（「史料」は、本来のアーカイブズ archives の概念と完全には一致しない）。1949年に国会へ提出された「史料館設置」請願では、請願趣意文冒頭にだけ「歴史資料」という総称がある。
- [3] ちょうど「史料管理学研修会」開始前の同年5月には、国文学研究資料館史料館編『史料の整理と管理』（岩波書店，1988年）を刊行している。また、国文学研究資料館史料館解体前年の10月には、同編『アーカイブズの科学』（上下巻，柏書房，2003年）が公刊されている。なお小稿は、これら史料館編の両書に大きく依拠している。
- [4] その活動成果は、近世庶民史料調査委員会編『近世庶民史料所在目録』第一輯・第二輯・第三輯（学術振興会，1952・1954・1955年）として公刊されている。

- [5] 国文学研究資料館史料館編, 名著出版, 1996年.
- [6] その成果は、『近世史料所在調査概要』(文部省史料館, 1970年)として報告された.
- [7] このデータシートは, 注[4]・注[6]の文献の示す「近世庶民史料調査書」「近世史料所在調査書」の書式と類似し, 両調査書の書式から発展したものである. 記載要素については, 附表を参照のこと.
- [8] 現在すでに電子化されている約46,000件弱の所在情報データが, SGMLファイルで45Mbytes余となる.
- [9] 具体的には, 記録史料群の出所の一体性を破壊しがちな古書店・古道具屋・骨董屋等の利用を制限している.
- [10] 本データベースは, 出所名の典拠レコードを提供できるが, 典拠レコード記述の国際標準の要素は満たしていない.
- [11] つぎの研究助成によって開発とデータ蓄積が進められた: 科学研究費補助金基盤研究(B)(2)「歴史史料情報の共同集約と共有化に向けてのシステム構築に関する研究」(研究代表者: 鈴江英一, 1999-2001年度, 課題番号: 11410096), 科学研究費補助金基盤研究(B)(2)「アーカイブズ情報の集約と公開に関する研究」(研究代表者: 大友一雄, 2003-2005年度予定).
- [12] これは, たとえば文化財所蔵者の自己負担によって文化財維持を求めるような日本の文化財行政の貧困さの問題といえる.
- [13] スタイルシートは, 少なくとも切り替わって表示される画面ごとに必要で, 当初数枚であったが, 現在は数十枚を使用している. EAD/XMLデータ内にある管理的データを抽出し, PHPを介して引数で渡し, 記述レベルの深度や表示する情報の一式を区別する.
- [14] ブラウザで使用する画面は, 既存のインスタントwebの画面ではなく,

全て専用に再設計した画面である。

- [15] 人名・団体名等の部分には、EAC, Encoded Archival Context (符号化記録史料脈絡, 策定中) を適用することになろう。
- [16] 国文学研究資料館の文化情報資源共有化についてはたとえば、つぎを参照のこと：原正一郎, 安永尚志「メタデータによるマルチメディアデータ統合の試み」『情報処理学会研究報告』Vol. 2001, No. 67 (人文科学とコンピュータ 51-7), 2001年7月；Shoichiro HARA, Hisashi YASUNAGA, "Resource Sharing System for Humanity Researches", LREC(3rd International Conference on Language Resources and Evaluation), 3. LREC2002, 2002