

付録 「アーカイブ資料の特質と EAD 採用の経緯」

五島 敏芳 京都大学総合博物館

EAD を採用するにあたっての経緯と、その前提となるアーカイブ資料の特質について、後の参考として付録にまとめた。以下にその議論を示す。なお、この付録における参考文献は、第 1 章（五島執筆分）を参照のこと。

1. 前提：アーカイブ資料の特質

いま「大学共同利用機関の歴史」に関する資料がアーカイブ資料（アーカイブズ *archive(s)*、史料、記録史料とも）¹¹を指していることは、さほど説明を要さない。ただ、アーカイブ資料の特質については、説明を加えておく必要がある。アーカイブ資料の情報共有や情報検索システムが、他の資料の場合と異なり、いかに困難か、いかにその実現が挑戦であるか、を知るためである。

1.1. アーカイブ資料のもと

アーカイブ資料には、もとになる記録・書類（または記録・文書・情報）がある。

もとになる記録・書類は、つぎのような特徴がある。

- 作成から授受、使用・保管まで、人にも似た一生（記録のライフサイクル）がある。
- その一生の末期、およそ使用されなくなったとき（非現用）、アーカイブ資料となりえる。
- それらを作成し授受し使用・保管しつづけてきた（法人）団体や個人・家（出所 *provenance*）がある。
- 出所による一体性がある。（フオーン *fonds*; 記録群 *record group*, コレクション *collection*）

- 出所のさまざまな活動を通して蓄積される。
- 出所内部の分業や機能に対応して、まとまり（小群；シリーズ series～サブシリーズ subseries, ファイル file）に分けられる。
- 分業や機能のなかの一連の業務処理（過程，手続き）によって，発生・蓄積する。（アイテム item ないしファイル）
- 分業や機能，業務処理に対応した小群が階層的に構成される。
- 分業や機能，業務処理に分けられるために，断片的である。
- 分業や機能，業務処理において頒布されたり繰り返されたりするとき，複製や内容の重複がある。
- 出所やそのなかの直接関係者（当事者）にしか理解できない。
- およそ大量に作成される。（出所の活動期間に比例し，数百，数千，ときに数万という点数。）
- それぞれは固有の存在である。（かりに複製や内容の重複があったとしても，作成・使用の過程が異なる。）
- 媒体や形式を限定しない。（紙だけでなく電磁的記録も含む；図面，表，地図，写真，録音，録画，電子記録...，とさまざま。）

もとなる記録・書類は，非現用になると，すべてが自動にアーカイブ資料となる**わけではない**。アーカイブ資料としての価値を持つ分が，アーカイブ資料となる。アーカイブ資料としての価値とは，永久保存価値 *archival value* のことである。

永久保存価値は，アーカイブズ用語事典（ICA1988）の説明によれば，つぎの各価値から構成される^[2]。

- 経営的 administrative,
- 財務的 fiscal,
- 法務的 legal,
- 証拠的 evidential, および／または,
- 情動的 informational.

これらの価値のいずれか1つないしすべてが、記録・書類（またはアーカイブ資料）の無期限 indefinite または永久的保持 permanent retention を正当化する。

これらの永久保存価値を、対象とする非現用の記録・書類すべてが有していれば、もちろんすべてがアーカイブ資料となる。しかし、ふつう大量に作成される記録・書類すべてが、永久保存価値を有することはない。永久保存価値を有するかどうか**評価**され**選別**されて、なお残る分が、アーカイブ資料となる。

永久保存価値を有するかどうかの評価・選別は非現用の記録・書類の当事者がになうべきかといえ、必ずしもそうではない。非現用の記録・書類がアーカイブ資料となるとき、その利用者は、当事者に限定されない。むしろ当事者以外のほうが多いと想定できる。非現用となりアーカイブ資料となるまでの時間は、しばしば当事者の記憶や存在を失わせる長さを持つ。また、当事者に評価・選別をまかせると、自ら関与した記録・書類を惜しんで不要分の廃棄ができなかったり、自ら関与した痕跡の残ることを怖れすべて廃棄してしまったりと恣意が働くかもしれない。

要するに、非現用の記録・書類は、当事者以外の利用を想定し、そのため当事者以外にも理解しやすく処理されなければならないし、その処理に価値の可能性を損なう恣意が入ってはならない。〈非現用の記録・書類をアーカイブ資料へ変える〉過程が必要になる。その過程が、**ドキュメンテーション**や**記述**という情報を補う処理であったり、評価・選別という不要な情報を除去する処理である。それらの処理には、アーカイブ資料管理の専門的知識・技能が必要であり、そういった専門家こそが *archivist*（アーキビスト、アーキビスト、文書司書とも）である。

1.2. アーカイブ資料管理の基本原則

アーカイブ資料管理には、その専門家にかぎらず、経験的につぎの基本原則にしたがう必要がある。

- A 出所原則／フォーン尊重原則 principle of *respect des fonds/provenance*,

- B 原秩序尊重原則 principle of respect for original order, の2つである^[4]。それぞれ、つぎのように説明される。
- A アーカイブ資料を、作成ないし収受し保管・保存してきた団体・個人・家（＝出所）ごとの一体性を持った群，ととらえ，同一の出所をもつ群は，その操作にあたり他の出所をもつ群と混同してはならない。
- B 出所を同じくする1つの資料群のなかで，個々のアーカイブ資料へ，もともと与えられている秩序（＝原秩序）が，それを生んだ団体・個人・家の活動の体系を反映しているばあい，その原秩序を尊重して残さなくてはならない。

Aにみるフォーン *fonds* は，「同一作成母体起源一体性記録群」と訳すこともできる^[4]。アーカイブズを理解し管理する際のキーワードである。アーカイブ資料の理解や管理は，このフォーンを単位とすべきである。現実には文書館の多くは，フォーンやコレクションの単位で資料を管理している^[4]。

アーカイブ資料管理は，もととなる記録・書類の管理の延長線上に位置する。記録・書類の当事者を母体組織とする文書館は，その当事者の継承者の一つであり，そこでの記録ないしアーカイブ資料の管理は，自然と出所単位が実現する。継承者という存在は，アーカイブ資料の起源にかんする確かさの根拠となり，文書館というアーカイブズ管理の専門的知識・技能の発現の場は，確かさの維持の根拠となろう。

Bの原秩序は，資料の出所によって形成された，資料の配列や，資料のまとめり・小群のまとめ方やその小群自体の配列，また資料ないし小群の作成・授受・保管等の諸過程をあらわす物質的形態や類型・形式などのことを指す。たとえば，一つひとつの文書のまとめられ方，まとめられたファイルのなかの各文書の配列，そうしたファイルのファイルキャビネットにおける配列，ファイルキャビネット自体の配置，等などである。

経験的に，原秩序は，資料ないし資料小群の，細部に残されていることが多い。その保存（物質的形状の保存）は，資料の利用を可能にするための物質的形状の改変（**装備**）と衝突することがある。装備は，たとえば，

封筒に入れる（新しい封筒に入れなおす）とか、番号ラベルを貼るとか、資料を劣化させる金属クリップを取り除いて紙縫で綴じなおすとかの保存手当での措置のことである。このとき、保存措置の原則の一つ、〈可逆性の原則〉を担保するために、資料の現状を記録するのがよい⁴⁴。「現状記録」は、原秩序を残せない可能性のある処理に際しての、必須の作業過程である。「現状」は、それが「原状」すなわち原秩序かどうか検討し、「尊重して残」さなくてはならないと判断したときに残すのではなく、必ず記録して残すものである。現状を尊重すべきかどうかの検討に時間を費やすよりも、何らかのかたちで記録するほうが、時間を浪費することなく合理的だからである。

1.3. アーカイブ資料の情報提示のあり方

こうして生まれる（再生する）アーカイブ資料は、その情報を提示しようとする、たんに外形の情報（書誌的情報）や、記された（描かれた、写された、録音された...）**内容**を示すだけでは不十分である。資料群（コレクション）内外の関係する情報（**脈絡**）が補われなくてはならない。具体的には、つぎをあげる。

- コレクションの出所と関係する他の団体・個人・家それぞれの情報、それら関係団体等を出所とするコレクションの情報。
- コレクション内の、資料のまとまり（小群）や、各資料の関係。

これらの情報へ、資料の利用者が、もとめる資料の情報からわかりやすく到達（アクセス）できなくてはならない。しかしまた、利用者が、（恣意的）連想によって資料を誤解しないよう、過去の資料としての証拠性を正しく理解する援助も必要である。

そのためアーカイブ資料の情報提示は、出所の**構造**、分業・機能・業務の体系に基づくことになる。アーカイブ資料のもととなる記録・書類は出所の構造や活動に対応して生成・維持されてきたはずだからである。

そもそもアーカイブ資料の情報は、検索簿を基本とする。検索簿とは、資料のまとまり（シリーズ等の小群）や1点ないし1件（アイテムないしファイル）がなにかを把握（識別）し、どこにどんな資料があるか探し出

せるようにする工具（ツール）である。資料の単位や所在がはっきりして、資料そのもの（実物、実体）を保存できる。資料の保存や利用を含むアーカイブ資料管理のツールとして重要な位置にある。利用者が閲覧室で目にする具体物で言えば、索引や資料目録（検索手段 *finding aid*）である。

検索手段は、つぎのいずれかの情報が必要となる。

- コレクションの概要(1)．
- コレクションの概要(1)と、シリーズ～サブシリーズの概要(2)
- コレクションの概要(1)と、収納容器一覧（リスト）またはアイテムないしファイルリスト(3)
- コレクションの概要(1)と、シリーズ～サブシリーズの概要(2)と、収納容器リストまたはアイテムないしファイルリスト(3)

このように、かならず(1)のコレクション全体を説明する概要は提示されなくてはならない。この必須条件が満たされれば、コレクション内部の情報は、小群の概要にとどまっても（(1)＋(2)）、詳細な一覧だけでも（(1)＋(3)）、小群の概要ごとに詳細な一覧があっても（(1)＋(2)＋(3)）、どんな水準でもかまわない。ただし、ふつうアイテムないしファイルリストを、検索手段の基本的情報として完備することはなく^[4]、むしろそういった詳細リストは索引（後掲）に含まれる。

なお、これらの情報を構成する単位（**記述データ**；(1)～(3)）は、コレクションまたは出所の構造や活動に対応し、〈コレクション～シリーズ／サブシリーズ～ファイル／アイテム〉といった各階層のことを**記述レベル**という。この記述データは、アーカイブズの世界の国際標準類^[5]により記述要素が定められ、どの記述レベルでも同じ記述要素を使用できる。言い換えれば、コレクションを説明する要素と、資料1点を説明する要素とが、共通である、ということである。これを**マルチレベル記述**という。

前述の検索手段の情報に、関連する参考文献（脈絡情報）や、資料の影印（写真、模写；ファクシミリ *facsimile*）・翻刻（筆写、口述筆記、読本等；トランスクリプション *transcription*）等など（内容情報）を結びつけ、その総体を資料の利用者へ提供すれば、いっそうアーカイブ資料は利用し

やすくなる。出所またはコレクションの構造にもとづいて、各記述レベルの記述データへこれらの脈絡・内容情報が結びつけられれば、より資料ないしコレクションを適切に理解できるようになる。

さらに、特定の個々の資料へアクセスできるようにするため、アクセスのための用語（団体・個人・家の名称、地名、主題、等など）を、アルファベット順・五十音順、年代順、ジャンル別、等などに配列した索引 *index*（多角的検索手段）を整備し、求める資料の（概要・リスト内の）記述データへたどりつけるようにすれば、およそアーカイブズ資料の利用環境の万全を期することができる。

さて、これらの検索手段の情報を、オンラインで提供しようと考えたとき、一見すると、検索手段の各情報を電子化すればよいだけのようにみえる。脈絡や内容の情報も、例えばハイパーリンクによって相互参照できるようにすれば、完備されるかに思える。

しかし実は、そう簡単に済まない。

オンラインの検索手段が充実している図書館の世界とくらべてみよう。およそ図書館のオンラインの検索手段、情報検索システムは、図書、雑誌、または雑誌の論文が、主たる情報となる。求める資料の検索キーワードを入れたら、その検索条件に合致する図書や雑誌か論文が、検索結果としてあらわれるだろう。この検索結果の単位は、図書や雑誌か論文であり、1冊、1篇と物的に独立した存在である。検索結果の図書、論文が、シリーズのなかの一部（分冊）であったり、論文集や雑誌に収録された1篇であったりしても、シリーズの図書全冊や、所収雑誌等の記事すべてを読まないで、その1冊、1篇を理解できない、ということはない。検索結果の図書、論文と、関連する図書、雑誌、論文があったとしても、同じくそれらの読解が、検索結果のそれらを理解するのに必須というわけではない。

つまり、図書館の情報検索システムは、構造や脈絡の提示がなくても、基本的には情報提示の目的を達成できる。

一方アーカイブズ資料の情報検索システムは、その検索結果から出所またはコレクションの構造や脈絡に即して上位・下位や関連の記述データへすぐ移動できることが求められる。出所の構造・活動に対応して階層的に構

成される記述データは、分量も一定ではなく、階層の深度も伸縮する⁴⁾。複数のコレクションをまたぐ検索において、同じ記述レベルの記述データが検索結果にあらわれるとはかぎらないし、そうして提示される検索結果の記述レベルが異なる場合、異なっていること自体の情報をも同時に提示されなければならない。

つまり、アーカイブズの情報検索システムは、どの階層の記述データもその階層の位置の情報を持ち、関連情報へ接続するための情報を持ち、検索結果として断片化された情報以外に、出所またはコレクションの（階層的）構造や脈絡の全体像を同時に提示できなければならない。矛盾する内容の実現が要件となっていて、その実現は困難に見えるのだ。

2. EAD による解決

これまで小稿筆者は、アーカイブ記述の ICA 国際標準類を情報検索システムへ実装するための試みにおいて、XML, *eXtensible Markup Language* の使用が前述の困難を解決するのではないかと予想していた。そして、XML の使用については、誤りはなかった。しかし、ICA 国際標準類の使用については、誤りだった（または、ICA 国際標準類だけでは、情報検索システムに求められる要件として十分ではなかった）。まず、この試行錯誤の過程にふれ、次にその困難を解決した EAD について説明していく。

2.1. 記述と情報検索システムのちがい

はじめアーカイブ資料制御 *archival control* の要素のなかで記述こそが基本であり、それを電子化して編集すれば、自然と（素朴な）情報検索システムができあがると考えた。そして ICA 国際標準類の記述要素をそのままデータベースのフィールドとしてみた。そのデータベース- テーブルへ、1 つのコレクションのデータを入力することから始めた。資料番号の記載を工夫すれば、表としてあらわれる一連の記述データの配列も、期待した順序になった。その順序が維持されていれば、複数の異なる記述レベルの記述データを収めても、各資料間の関係を示す脈絡の情報が損なわれることはなかった。

問題は、作業を進めるうちに顕在化した。

1 つは、検索・抽出や並び替えである。探し求める記述データの位置を

順に示していく検索なら、表の前後の記述データも視界に入る。探し求める記述データを抽出して一つの表にまとめても、抽出を解除すれば、もとの順序で一連の記述データの表に戻る。並び替えによりもとの順序に戻らなくなれば、もとの順序のための識別記号（コード、符号）を別途に用意すればよい。情報検索システムとして、データベースとして、当り前の機能の、一時的操作は、とくに問題とは思えなかった。並び替えの問題も、すこしの工夫で解決できたかと思えた。

もう 1 つは、複数のコレクションのデータの入力である。いくつかのコレクションの記述データをあわせて前述のような検索・抽出や並び替えを実行した結果は、混乱していた。検索・抽出結果となった記述データは、それぞれどのコレクションに属する資料を指しているか、一見してわからない。どのシリーズやサブシリーズに属しているか、わからないし、それらもどのコレクションに属しているか、わからなかった。結果表示された記述データの記述レベルからコレクションまでの階層的位置が表示される必要があった。さらに並び替えを考慮すると、管理用コードを別に用意するとしてもコレクション以下すべての階層にわたって記述レベルを固有に特定・識別できるように設計しなければならない。当り前の一時的操作の結果や、すこしの工夫どころではない、と認識できた。

これが、コレクションよりも上にある機関・施設をこえた記述データを扱うとなれば、さらに複雑な設計を要することは、容易に想像できた。当該の記述レベルの記述データまでの長い階層的位置の表示とそのための管理コードなくして、アーカイブ資料としての構造・脈絡の提示はできない。

何を誤ったのか、たどってみることにする。

情報検索システム以前に、その直接の基礎となる検索手段について、もうすこし考察しておく。その検索手段と、記述とは異なる。あらためて両者の関係を考えて、つぎのような指摘ができる。

- 記述をもとにして作成した利用者向け^[10]の表現が、検索手段である。
- 検索手段は、記述にない資料への多様なアクセス方法を含む。

- 記述はアーカイブ資料の基本原則にもとづいて編成（整理）された秩序に即してその構成を固定的とするのに対し、検索手段はその一式の表現（基本形式）だけでなく利用者の必要から順序・範囲が加工され記述の外部の参考情報が追加されることもある情報の一式（多様な形式）をも含む。

さきに具体物として例示した「索引や資料目録」のうち、〈完全な資料目録〉が前掲「基本形式」、〈索引〉や〈資料目録の一部〉が同「多様な形式」にあたる。〈完全な資料目録〉は、あるコレクションの全体像を描くように全体の解説から1点ごとのリストまでの情報を完備し、コレクションの体系的構造、階層的構成・機能等の秩序にしたがって配置され提示されたものを想定している。〈索引〉は、もとめる資料へのアクセスのための用語と、その用語に対応する完全な資料目録のなかの位置（該当ページとか資料番号とか）が示され、その位置の指示により、コレクションを構成する資料の単位（シリーズやファイルのまとまりのこともあれば、1点のアイテムのこともある）の記述データへたどりつく。〈資料目録の一部〉としては、ある主題や事項に関係する資料の記述データの一覧（それが1つのコレクションだけにとどまることもあれば、複数のコレクションをまたぐこともある）や、ある資料単位以下の基本形式ではカバーしない詳細な内容の一覧（ある帳簿のなかの件名目録など）を例にあげることができよう。

アーカイブズの各種の検索手段とそれら相互の関係は、つぎのようにまとめられる^[4]。

- 機関または個人の作成により準備される検索手段（記録の現用時点）
 - 必須
 - 登録簿 *Registers*
 - 索引類 *Indexes*
 - ファイリング体系の説明記述 *Filing system descriptions*
 - 分類語彙辞典類 *Thesauri*

アーカイブズ資料の特質と EAD 採用の経緯（五島）

- アーキビストにより準備される検索手段（記録の受入以後）
 - 必須
 - 受入登録簿 *Accession Registers*
 - 記述の基本目録 *Descriptive Inventories*（コレクション案内 *Collection Guides*）
 - 架橋の補助手段 *Bridging Aids*（機能や年代で区切った範囲に，作成者・出所の名称，事項等の語を，アルファベット順，50 音順に並べた一覧表）
 - 望ましい
 - 収蔵資料案内 *Guides to Holdings*（要覧 *Concise Guides*）
 - 収蔵資料報告書 *Reports of Holdings*
 - 記述の基本目録索引および収蔵資料案内索引
Indexes to Descriptive Inventories and Guides to Holdings
 - 資源が許せば
 - 内容索引 *Content Indexes*
- 編集された情報源
 - 資源が許せば
 - 特殊リスト *Special Lists*（出所，作成者の組織図や機能一覧，年表など）
 - 主題ファイル *Subject Files*
 - 分析的情報源 *Source Analyses*（特定の話題に関する資料のリスト；シリーズやときにアイテム，また出版物をも含む）

■ 参考情報案内および媒体案内 *Reference Guides and Media Guide*

このように検索手段は多様である。しかし、中心となる情報の要素は、記述の要素と大きく隔たっているわけではない。形式のちがいが大きい。この形式のちがいは、かつて図書館にあったカード目録を想起すれば、わかりやすいかもしれない。書名の50音順で並べられた基本のカードと、著者名や主題の50音順で並べられた副出のカードは、取り出した情報の並び方が異なるだけで、カードに記された情報の要素は、ほぼ同じであろう。

情報検索システムでは、こうした検索手段の形式の多様さが1つの画面で切り替わって現前する。カード目録であれば目録の種類異なるカードボックスや引き出しの前に人間が移動して利用するが、それを画面越しの要求に応じてカードボックスや引き出しを目の前に用意するようなものである。そうして用意する動作や、目録の種類が異なることを、1つの画面に区別して提示しなければならない。

また、コンピュータの世界では、とうぜん検索手段のデータは電子化され、電子化された文字列のどの部分を取り出してもアクセスのための用語となりえる。コンピュータ技術として自由に検索できる方途が準備されるとき、アクセスのための用語を特別に準備することが疑問視されるだけでなく、そのアクセスのための用語を列挙して提示することをも疑問視されるかもしれない。しかし、自由に検索できる方途には、アクセスのための用語となりうる検索用語や検索禁止語¹²⁾を収録した用語辞書も含まれていなければならない。用語辞書を伴わなければ、記述データから無意味な文字列を拾い出し、検索ノイズのある検索結果を返してしまうかもしれない。

なお、検索行動・道程を適切に誘導するようにアクセスのための用語を設定する必要もある。記述データ作成において、アクセスのための用語を、その語の単位のデータとして準備する作業は、省略できるものではない。そうして各記述データに埋め込まれたアクセスのための用語は、集積すれば用語辞書を構成するし、用語辞書と照合されるインデクスともなる。

用語辞書の構築，記述データのインデクスの構築は，従来の検索手段の作業に付け加えられる作業である．情報検索システムの質を上げようとするればするほど，その作業は重い内容となっていく．

これら検索手段および情報検索システムの特徴的要素が記述には存在しない，という〈記述と情報検索システムのちがい〉についての認識を欠いていたことが，誤りだった．端的に言えば，記述の情報要素だけでなく，検索手段としての多様な表現形式を制御する要素をも考慮に入れる必要があり，その多様な表現形式を実現するインタフェースを用意しなければならなかった，ということである．

2. 2. EAD の再発見

アーカイブ記述の ICA 国際標準類には，検索手段としての多様な表現形式を制御する要素もインタフェースのための要素も定められてはいない．それらを自力で設計し情報検索システムとして実装することは，小稿筆者の当時おかれていた環境にあっては不可能であった．しかし，アーカイブズの情報検索システムの実現は，日本のアーカイブズの世界で長らく求められてきたことであり，日本のアーカイブズ研究・実践のナショナルセンターであった小稿筆者の当時の所属機関，国文学研究資料館史料館（当時）¹⁴¹こそが取り組まなければならないことでもあった．その早期の実現は，史料館にとっても直接担当であった小稿筆者にとっても喫緊の課題であった．

アーカイブ記述の ICA 国際標準のうち ISAD(G), *General International Standard Archival Description* 第 2 版¹⁴²を用いた記述適用実験と，そのデータを XML 化し XSLT を通した表現（単純全文表示，ファイル内リンクあり）は，すでに小稿筆者の経験するところだった．それらのデータを用いて EAD へも適用し同様の XSLT による表現も確認していた（～2001 年末）．その結果，はじめ EAD を難解で複雑な存在と考えていた．

しかし，ほぼ同時期に前述の誤りも経験したことで，EAD の難解さ・複雑さの理由が記述と情報検索システムとの違いにあることを理解できた

（この経緯は，前出の五島 2003 参照）．情報検索システム特有の要素を自力で設計して付け加えるより，EAD を理解することに注力したほうがよい

ことは、作業効率の点だけでなく、事実上の標準としての普及しつつあったことから後のアーカイブ資料の情報交換・共有のためにも利益が大きい、と判断するに至ったのだ。

かくして小稿筆者にとって EAD は再発見されることになった。しかし、この時点では、EAD がどのような内容か、理解できていなかった。その理解の過程を、ここですべて振り返るわけにはいかないが、次項に簡略に触れていくことにする。

2.3. EAD とはなにか

EAD, *Encoded Archival Description* (符号化永久保存記録記述) は、アーカイブ資料の検索手段を電子的に符号化するためのデファクト-スタンダード *de facto standard*, 事実上の国際標準である¹⁵¹。EAD は、アーカイブ資料のための情報検索システムそのものではないが、その基礎となる検索手段の電子データのありようを定める枠組みである。

ここでも図書館の世界とくらべるとわかりやすい。

EAD に相当する存在は、MARC, *Machine-Readable Cataloging* (機械可読目録)¹⁵²であり、一方 ISAD に相当する存在は、国際標準：書誌記述 ISBD, *International Standard Bibliographic Description*¹⁵³となろう。EAD は、いわばアーカイブ資料のための〈機械可読目録〉といえる。

MARC によって実現している図書館世界の情報検索システムは、すでに触れたように、アーカイブ資料の書誌的情報を扱うのには限界がある。その限界を認識するのに、図書館世界の貢献があったことは、まず銘記されなければならない。図書館関係者による MARC の限界を越えようとする努力¹⁵⁴が、EAD を生んだのである。

はじめアーカイブ資料の機械可読目録の動きは、図書館世界の情報検索システムへ組み込もうとすることで進行していたといえる。それは MARC を拡張する MARC AMC, *USMARC Format for Archival and Manuscripts Control*¹⁵⁵の存在が示している。英米目録規則第2版 AACR2 をより良くアーカイブ資料へ適用するための手引書, *APPM2, Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries, 2nd ed.*¹⁵⁶もあり、そのなかで MARC AMC の適用

実例も紹介されている。

MARC AMC を適用できるアーカイブ資料の書誌的情報は、概要目録 *catalog* である。概要目録は、コレクションのレベルの情報である。図書 1 冊に対応する適切なアーカイブ資料の構成単位は、コレクションのレベルだけだからである。この対応関係は、じっさい多くの図書館の情報検索システムが採用していよう。仮にコレクション内部の階層の深度の伸縮するレベルのいずれかの記述データを対応させたとすれば、その上位のレベルとの階層的関係を提示できない。ただ、コレクションのレベルの書誌的情報であれば、かならず MARC のデータにできてオンライン目録へ組み込める、かといえそうではない。しばしばコレクションの書誌的情報の完全な一式は、各資料の構成・構造、出所・作成者の分業体系・機能・履歴、等などの長大な解説を含むため、ふつう情報量的に図書館の情報検索システムで扱えない。

外形的に図書 1 冊と類似するアーカイブ資料の構成単位は、ファイルやアイテムのレベルであろう。それらの一覧や容器（コンテナ）リスト、シリーズかファイルのレベルのリスト（欧米ではこのレベルが一般的）もあり、収蔵目録 *inventory* や登録簿 *register* という。記述データ 1 つ（1 レコード）あたりの情報量は限られるが、資料の点数または件数だけレコード数は存在し、しばしば大量となる。その記述データを市販データベース管理システムへ投入できても、すべてのレコードはフラットにあつかわれる。図書館が収蔵するアーカイブ資料で、収蔵目録や登録簿を電子化したとしても、コレクションの概要や階層的構成の情報は別に用意して単独でサービスされたことであろう。

長大な解説により 1 つひとつの記述要素における文字列数が大量となる概要目録 *catalog* と、件数として大量となり伸縮自在な階層を内包する収蔵目録 *inventory* や登録簿 *register* とを、コレクションの単位でどのように合わせて取り扱うのか。

この〈*catalog* と *inventory* の結合〉は、アーカイブ資料の機械可読目録が模索され始めてから長らく課題となっていた。

EAD は、テキストファイルの形式である。ここでの「機械」として想定されるコンピュータやネットワークの技術の急速な進展により、ハードウェアやソフトウェアの仕様が変わったとしても、最低限の技術的理解として維持される可能性が高い。市販の応用ソフトウェアに依存した形式より、よほど比較的安定した保存形式であり、長期の持続可能性を期待できる。

EAD のテキストファイルは、市販の応用ソフトウェアで使用するカンマやタブで区切られたテキストファイルではない。カンマやタブで区切られたテキストファイルは、行と列で構成される表を想定している。たとえば 1 行をレコードの単位とするならば、何番目の列に何の情報が入るか、1 行目に見出し・凡例が明示されたりされなかったりして、1 レコード内の情報の要素の区切りは不安定なことがある。

EAD では、どんな情報の要素の区切りも、かならず明示する方法を結果的に採用した。タグ（要素および属性）という目印（マーク）を付け、そのマークによってどこからどこまでが何の情報かを明示するのである。その国際規格には、テキストファイル形式によって文書一般を電子的に符号化するための種々の規約を構築できる、SGML, *Standard Generalized Markup Language* を選んだ^[21]。この方法であれば、情報の要素の区切りに間違いが起こらないだけでなく、表の枠目に入る分量の文字数までといった制限もない。また、文書一般の意味内容の区切りにも対応できるように設計されているので、たとえば小説や論文の章・節・パラグラフといった入れ子の区切りを表現できる。概要目録 *catalog* の文字数の問題を解決でき、収蔵目録 *inventory* や登録簿 *register* で階層的構成を実現でき、*catalog* の下に *inventory* や *register* を収めることができる。マルチレベル記述の階層的配列が可能になる。

つまり、〈*catalog* と *inventory* の結合〉をもたらすのだ。

EAD は、SGML や XML のサブセット（部分集合）として策定された。EAD の文法・記法は、DTD, *Document Type Definitions*（文書型定義）によって規定され、現在 2002 年版である。EAD の DTD ファイルは、1 つの DTD ファイルといくつかの関連ファイルで構成される^[22]。EAD の目的を、EAD

技術文書をかねる DTD に聞けば、

「収蔵目録 *inventories* や登録簿 *registers*——それらは、文書館、図書館、大学、団体、政府機関、およびその他の組織によって作成されるレファレンス・ツール（参照工具）であり、アーカイブ資料および手稿資料について、記述し、制御し、接近を提供する——として知られるアーカイブ資料の検索手段の類を符号化するため」

という。

アーカイブ資料の検索手段の多くの形式を取り扱う枠組みであることを表明している。

EAD では、単に対象となる検索手段を電子化・符号化データを提供するだけでなく、その検索手段を電子化・符号化した（電子的に作成した）行為を記録できる。たとえば、ある個人書類の資料目録と、その個人書類そのものは、区別されなければならない。対象となる検索手段の内容と関係ない、電子的検索手段としての書誌情報を持つことができる。SGML や XML のテキストファイルとして EAD データが作成されるとき、その環境（OS のファイルシステム）によりタイムスタンプや作成者のデータが付随するかもしれないが、それらとは別に EAD データ内に書誌情報を記録可能である。

もと研究用テキスト、文学作品の電子版作成から発達してきた、TEI, *Text Encoding Initiative* は、このような区別に配慮がある^[2]。TEI も、EAD と同じように SGML や XML のサブセットであり、EAD は一部を TEI に倣って開発されたという。

2.4. EAD の基本的構成

EAD のデータは、大きく <eadheader>EAD ヘッダ・<frontmatter>前付け・<archdesc>アーカイブ記述の 3 つの部分に分かれ、それぞれ階層的に構成されている。前述の電子的検索手段そのものの書誌記述は EAD ヘッダ部分に示し、検索手段の本文はアーカイブ記述の部分へ収める。前付け部分には、扉や（可視の）書誌事項、序文・献辞・凡例など、実物の図書と同じような記載を表現できる。

EAD のタグと ISAD(G)第 2 版の記述要素の関係は、現行タグ- ライブラリの付録 A にある EAD クロスワークという対照表から知ることができる。同表によれば、ISAD(G)の要素のすべてが EAD のタグのいずれかと対応するが、EAD で最低限必要とされるタグの一式は必ずしも ISAD(G)の要素すべてが含まれるわけではない。記述とそれを表現した検索手段との違いといえる。たとえば、あるアーカイブ資料の特徴からは不要と思われる ISAD(G)の記述要素を、ほとんど情報もないままに EAD の対応するタグで符号化するより、ISAD(G)の記述要素としては存在しないが検索時のキーワードや見出し語となるかもしれない<controlaccess>アクセス制御タグに情報を収めるほうが、検索手段としては有益であろう。

EAD におけるマルチレベル記述は、<archdesc>タグ内の各タグヘフーンまたはコレクションレベルの記述の情報を収め、同タグ下の<dsc>下位構成要素タグ内の<c0x>/<c>構成要素タグ内の各タグヘサブフーン以下の各記述データを入れ子状に収める、というかたちで実現している。入れ子状のデータ格納は、各記述レベルのデータの配置を決定する。整列秩序を与えなければ無秩序に存在するデータベースのレコードのように、異なった記述レベルのデータが混在することはなく、ファイル内にアーカイブ資料の内的階層構造に即した連続的配置が可能となっている。

なお、<archdesc>タグ下に設定された EAD データの階層レベルと、記述レベルとは、一致していなくてもよい。たとえば、<c02>のレベルはつねにシリーズ記述でなければならない、ということはない。研究の進展により今後、アーカイブ記述が改変される可能性があるばあい、記述レベルとタグの階層レベルとの対応を固定しない方がよい。またたとえば、<c0x>タグの階層レベルを示す「x」の数値を自動的に付与するような処理をしている場合、特定の記述レベルが存在しないときは、1 以上の階層レベルを省略する必要が生じ、不具合が発生してしまう。むしろ、両者は一致させない（固定しない）方がよい。

注記

[1] *archive(s)*や*archival*の語は、アーカイブズの基本概念を考慮すれば、「永久保存記録」や「永久保存」と訳すのがふさわしい。しかし、日本で定着した訳語ではなく、アーカイブズの資料（機関や、施設・設備ではなく）を区別して示す意図から、便宜的に「アーカイブ資料」の語を小稿では用いた。なお、アーカイブズ用語事典（ICA1988）の*archives*の説明（定義）のうち、アーカイブズの機関については「文書館等」の訳語をあて、施設・設備については文書館等の建物各部（閲覧室、収蔵庫等など）を示すのに一般に使用される語を用いた。

[2] *archival value* は、アーカイブズの基本概念を端的かつ適切に説明するため、小稿筆者の私訳を用いた。

[3] これらの基本原則は、アーカイブズの入門書でよくふれられる（M. J. Fox, et al.1998 など）。

[4] 説明的訳語で煩雑な記載ではあるが、おそらく最も適切な内容を示している（ICA 編・安澤訳 2006）。「一体性記録群」とも略記できる。資料の種類や慣行により、コレクション *collection* や記録群 *record group* と言いかえることもできる。

[5] 記録作成母体が巨大なばあいや、同じ機能を維持していても組織改変により担当部分が分割したり別の機能をもつ部分と統合したりすることがあるばあい、実際には、もっと下の単位（シリーズ *series* やサブシリーズ *subseries*）で管理することもある。

[6] その記録は、写真でも略図でも、文字や文章による説明だけでもよく、後日にわかるように残せばよい。

[7] 日本は例外的で、1点ごとのリストだけ存在し、しばしばコレクションの解説・解題を欠くばあいが多い。

[8] 後にもふれるが、まず国際文書館評議会 ICA, *International Council on Archives* の定める国際標準により記述の考え方や意味内容の一式が明らかにされている。この「記述」は、アーカイブズ資料管理のなかで、コレクションの内容構成を把握・理解した後の、内部的基礎データを含む。そういった「記述」をもとにして作成する、利用者向けの表現形が「検索手段」であり、検索手段作成のための意味内容の一式を示す標準類も存在する。検

索手段作成のための標準類は、しばしば国内標準ないし地域標準であり、「目録規則」は主要な具体的存在といえる。これまで目録規則は、とくに図書館の世界で鍛えあげられてきていて、アーカイブズの世界でも共通して使用する部分がある（例えば、英米目録規則 AACR と、APPM や DACS との関係のように）。なお、日本ではアーカイブズ専用の目録規則は、存在しない。

[9] 前述のとおり、図書館の世界で対象とする情報（書誌記述）は、均質なうえ、たとえば 1 冊の図書には 1～3 枚の図書カードを費やす程度で済むかもしれない。これに対して、アーカイブ資料の記述では、もちろん 1 つのアイテムだけか概要（要約）だけなら簡単でありえるものの、1,000 頁以上におよぶこともあり、平均しても 15～30 頁にはなる、という（D. V. Pitti 1999）。

[10] ここで言う「利用者」は、文書館等の外部からの一般閲覧者だけでなく、内部の資料管理のための利用者をも含む。内部の管理的利用者と記述作成担当者とは一致しない可能性を考えれば、外部閲覧者向けではない管理的（housekeeping の）情報であっても、わかりやすく表現する必要が出てくる点に注意されたい。

[11] ここでは、やや古いが、オーストラリアのアーカイブ資料管理・文書館運営の概説書（教科書）に依拠した（Ann Pederson ed. 1987）。

[12] たとえば、「a」「あ」のような、冠詞やおよそ無意味な文字列が、該当する。

[13] もと 1951 年に国会請願の成果として成立した、文部省史料館（通称）の後身。戦後の混乱のなかで急速に失われつつあった「史料」の緊急避難的保存先として活動し、1972 年に日本で 2 番目の国立大学共同利用機関である国文学研究資料館の設置にともない、国文学研究資料館の下におかれ、附置機関または付属施設として解釈された。2004 年の国立大学法人化にともなう大学共同利用機関の法人化に際し、消滅した。

[14] この第 2 版が現行版である（英語版は、つぎを参照：

<http://www.ica.org/sites/default/files/isad_g_2e.pdf>）。初版は 1994 年に公表され、日本では「国際標準：記録史料記述の一般原則」として翻訳があ

る（ア研編訳 2001 収録）。このほか典拠レコードの国際標準である ISAAR(CPF), *International Standard Archival Authority Record for Corporate Bodies, Persons, and Families*（国際標準：団体，個人，家に関するアーカイブ典拠レコード）があり（第 1 版は 1996 年公表，日本語版はア研編訳 2001 収録，英語版：<http://www.ica.org/sites/default/files/isaar_eng.pdf>），現在第 2 版（2004 年公表，英語版：

<<http://www.ica.org/sites/default/files/ISAAR2EN.pdf>>）である。

[15] EAD は，SAA, *Society of American Archivists*（アメリカ-アーキビスト協会）の EAD 作業部会と LC, Library of Congress（アメリカ議会図書館）の関係事務局が維持にあたっている。つぎの URL を参照のこと。

<<http://www.loc.gov/ead/>>.

[16] つぎの URL を参照のこと。<<http://www.loc.gov/marc/>>.

[17] つぎの URL を参照のこと。

<http://www.ifla.org/files/cataloguing/isbd/isbd-cons_2007-en.pdf>.

[18] その動機は，おそらく図書館のあつかう資料の範囲の拡大にあったと想像する。多くの検索手段のオンライン目録化の進んでいなかったであろうアーカイブズ資料は，図書館の中心的対象資料とはいえないが，オンライン目録化が進めば，資料の利用が進むと同時に，図書館にとって新しい顧客を開拓できる，格好の対象であったにちがいない。

[19] MARC AMC は，1982 年 10 月に SAA で承認されたが，1987 年以降更新されていないという。通常の MARC との主要な相違は，多くの注記の項目の設定にあるという。この注記は，つぎの文献に依拠した。（オンライン版）Victoria Irons Walch ed., *Standards for Archival Description: A Handbook*, SAA, 1994. より直接には，つぎの URL を参照のこと。

<<http://www.archivists.org/catalog/stds99/chapter3.html#umamc>>.

[20] S. L. Hensen. ed. 1989. APPM の第 1 版は，1983 年刊行。目録「規則」ではなく，「マニュアル」となっている点には留意してよい。なお，目録規則である AACR2 は，1978 年発行。

[21] なお，当初から SGML が選択されていたわけではなく，初期のプロジェクトでは，データの形式として，単純 ASCII テキスト形式，HTML タグ付

き ASCII テキスト形式, MARC タグ付きレコード, SGML, の 4 つが検討されたという.

[22] 関連ファイルには, 特定の地域的ないし国家的連合体での特殊な約束事のために要素・属性の用法や値を拡張する実体参照や, 普通のアルファベット以外の文字 (キリル, ギリシア, ラテン, その他の記号) の使用のための実体参照, SGML アプリケーションのための宣言, 等などがある. 第 1 版では, EAD のすべての要素・属性の指定も実体参照のファイルとして DTD ファイル本体から呼び出される構成になっていた (現在は, EAD のすべての要素・属性は DTD ファイル本体で指示されている) .

[23] TEI については, つぎの URL を参照のこと. <<http://www.tei-c.org>>. 本文に記した配慮は, SGML, HTML, XML といったマークアップ言語で設計すれば実現できる. 実現できてこそ, 電子記録・電子書類の真正性 authenticity・信頼性 reliability・統合性 integrity が保証される.