

# 京都府行政資料のデジタル・アーカイブ化とその課題

## —— EAD/XMLの適用の可能性と歴史学 ——

後 藤 真

### 要 旨

本稿は、近代の都市を含むさまざまな行政資料をデジタル・アーカイブ化する事例とその課題について述べる。特に、EADを用いたデジタル・アーカイブ化をその材料とした。EAD (Encoded Archival Description) とは、世界共通の文書館・公文書館といったアーカイブズのためのメタデータのセットである。本稿では、特に京都府立総合資料館に所蔵されている京都府行政資料をデジタル・アーカイブ化する際に、このEADを用いた。この事例をもとに、近代行政資料をはじめとした歴史的資料を、目録化し、データベース化、それを時間的・空間的に流通させるためのデジタル・アーカイブへと転化させる可能性について検討した。その際、特にメタデータをいかにして用いるかといった議論を中心として、実例を確認した。その結果、デジタル・アーカイブ化に際し、EADによって構造を記述することが可能な部分と、きわめて困難な部分があることが判明した。特に媒体と内容が一对一で対応しない事例などでは困難であり、デジタルデータとして構造を記述することは困難であった。このことはEADが有用でないということを直接的に示すものではない。しかし、EADが資料の完全な構造を記述しうるものではないことを示している。そして、既存の日本の歴史学の目録では当然のように入っていた項目が、EADの要素の中に存在しない点についても検討した。日本の歴史学の目録に存在しながら、EADの要素にないものを、提示し、その分析を行うことを通じ、デジタル・アーカイブという行為が、人文科学、とりわけ歴史学の資料分析の理解の深化へと進んでいく可能性があることを提示する。

キーワード：デジタル・アーカイブ、京都府行政文書、アーカイブズ、史料学、EAD

(2006年5月10日論文受理、2006年7月7日採録決定 『都市文化研究』編集委員会)

### 1. はじめに

本稿は、近代の都市に関係する情報を含む資料群をデジタル・アーカイブ化する場合の方法論の模索と、その課題について述べる。その事例として、特に京都府行政資料<sup>1)</sup>の調査に際し、作成するデータベースのデジタル・アーカイブ化について検討したい。近代行政文書をはじめ

とした歴史的資料を目録化し、データベース化する。そして、それを時間的・空間的に流通させるためのデジタル・アーカイブとするまでにいたる課題について、特にメタデータ（データを説明するためのデータ）についての議論を中心に据え、検討を加えたい<sup>2)</sup>。

## 2. 適用する事例の概要—議論の前提—

現在、京都府立大学の小林啓治氏を中心として、京都府立総合資料館の近代行政資料を総合的に検討するプロジェクト「京都府行政文書を中心とした近代行政文書についての史料学的研究」が行われている。このプロジェクトは、保存の方法がいまだに確立しておらず、劣化・消失の恐れがある近代行政史料群を媒体・資料構造の側面から、史料学的に分析し、近代行政文書を体系的に理解しようというものである<sup>3)</sup>。私は、この研究メンバーの一員として、この行政史料のデジタル・アーカイブ化を担当している。

ここで、検討の対象となる、京都府行政文書（資料）について、確認しておきたい。京都府行政文書とは、2002年に重要文化財に指定された、京都府の行政文書群である。対象は慶応4年～昭和21年までであり、総点数は、15407点である。地方公共団体作成・所有の行政文書としては、初の重要文化財指定がなされた。大きく以下のような史料群に分類可能である。

### a. 京都府庁文書（12641点）

京都府行政文書の中核である。知事部局の各部課作成の文書を簿冊としたもので、永年保存文書。社寺・土木・人事関係が多い。また、ほかに大正・昭和の大札関係や文化財保存関係がある点も特徴である。

### b. 京都府庁史料（2110点）

明治から第2次大戦終了前後までの永年保存指定以外の文書である。業務参考資料として作成された。「京都府史」「町村沿革調」などの稿本であり、「寺院明細帳」「神社明細帳」などの台帳類。

### c. 郡役所文書（249点）

明治前半から大正期まで各郡に設置されていた郡役所で作成又は受理され、管理されてきた文書である。郡役所廃止数年後の昭和4年に京都府に引き上げられるが、第二次大戦中に廃棄された。

### d. 豊岡県第十四・十五大区区務所文書（18点）

明治6～9年の大区小区制期に作成されたものである。加佐郡役所、京都府庁に保管されている。役用記録などが中心である。

### e. 宮津藩政記録（389点）

旧宮津県が県政の参考資料として、藩政時代の資料を収集したものである。与謝郡役所、京都府庁の順にて保管されてきた。藩政記録、書簡などが存在する。

今回とりあげる対象となったのは、主にaの行政の簿冊の史料群である。これらの一部をサンプルとして抽出し、簿冊の劣化状況を確認する。さらに目録化し、デジタル・アーカイブとすることによって、時間的・空間的流通性を高めると同時に、より効率的にその劣化状況を確認するための一助としている。さらに、将来、行政資料の総合的なデジタル・アーカイブ化を模索するためのテスト・ケースでもある。

具体的には、1. 所在情報の流通、2. 調査情報の永続的な保存、3. 資料情報の多様な検索方法の確保、4. 資料情報を機械的に抽出し、統計的に処理するための基礎データの整備、などを目的としている。

## 3. デジタル・アーカイブ化への技術的な方法

デジタル・アーカイブのためのメタデータを記述するための方法に関しては、論点が大きく二つ存在するとされている。基本的に、デジタルデータを入力し、それを処理し、活用するためには、二つの課題が存在する。一つは、データの形式、もう一つはデータの内容である<sup>4)</sup>。

まず、デジタル・アーカイブ化にあたっては、データの形式をいかなるものにするか、という課題が存在する。デジタル・アーカイブとはその名の通り、デジタルデータである。そのため、機械可読性と、人間の可読性の両者を備えていなければならない。そのため、基本的にはテキストのデータとして作成されることが、一般的であったといえる。しかし、テキストデータは、そのままでは、どこにいかなる情報があ

るか、といった場所を特定しにくいという欠点がある。また、データのもっている「意味」をコンピュータがそのままでは理解できない（機械可読性がない）という課題が残っていた。そのため、機械可読性をもたせるために、データのためのデータ、つまりメタデータが必要であると考えられるようになってきた。また、機械可読性を向上させることにより、検索などが容易となるため、人間の可読性も向上する。

そこで、近年、SGMLとXMLという構造化可能なメタデータが技術的に策定されることにより、あらたなデータ形式による新たな方法のデジタル・アーカイブ化が可能となった。現在、目録などの資料所在情報や、文字資料を記述するためのデジタル・アーカイブは、ほぼ、このXMLの形式をとるようになっている。

そのため、本事例においても、データの形式は、XML形式とすることとした。

次の課題は、データの内容に関する検討である。データの内容に関わる、そのタグセットには世界共通の仕様とされているものを用いることとなった。その理由は以下の通りである。

デジタル・アーカイブ化を行うにあたって、資料のデータの個別構造を把握し、その後、その構造に則したデータベースを構築することが、本来もっとも理想の形である。それは、XMLという、データの形式そのものが本質的に「意味と構造を記述する言語」として定義されているためである。そのため、データの構造を総合的に把握した後、その構造をもとにデータを作成することが望ましい。

しかし、資料の構造を検討し、その後メタデータを作成するという手法には、現実には大きく二つの問題点がある。

1. 資料構造を把握、理解した後に、データベースを作成、入力という手続きを踏まねばならず、非効率であるという、プロジェクト自体が持つ問題。
2. 資料の構造に則したメタデータを作成した場合、そのデータ構造は、その資料群に独自のものとなってしまい、後への拡張性と検索性を著しくそいでしまうという、技術的、かつデジタル・アーカイブの機能的な問題。

これら二つの制約があるため、実際の作業としては、資料構造に則したメタデータは作成しがたいという問題がある。

また、独自のデータ構造は、デジタル・アーカイブの特徴である、時間的永続性、そして、空間的な流通性を失わせることになりやすい。本来資料の構造を理解するために、メタデータがあるべきにもかかわらず、検索のため、また、メタデータを理解するために、資料構造を別の文脈で把握しなければならないという事態も考えられる。データベースがデジタル・アーカイブとして機能するためには、時空間をこえ、さまざまな人々に使ってもらうことが前提となる。そのため、極度に資料構造に依存したデータベースは避けざるを得ない。

また、このプロジェクトにて用いられたデータベースは、将来的には京都府立総合資料館の全文書群を網羅的に検索するためのデータベースへと発展されるべきであろう。そのため、国立公文書館など、他の近代行政文書・公文書と共通で検索が可能となるようなくみを目指し、この資料群のみならず、他の資料群へも適用可能なメタデータを採用することが要求される。

これらの点を総合的に考慮し、このプロジェクトにおけるデジタル・アーカイブ化においては、独自のデータの定義(タグセット)ではなく、すでに存在する共通のタグセットのメタデータを用いることとした。

基本的にデータベースがある種のデジタル・アーカイブとして機能するためには、形式のみならず、内容についても流通性が確保されるべきである、との考えのもと、いくつかのメタデータの世界的な統一が図られている。その世界共通のタグセットを用いることとしたのである。

主要なものとして、ダブリン・コアと、EAD (Encoded Archival Description) について、その可能性を検討した。両者の概略を述べる。

ダブリン・コアとは、当初、web上に流れる情報群を目録化する目的で、1995年より策定が開始された。W3Cのメンバーの呼びかけによって、図書館員・アーキビスト・人文科学者・Z39.50の関係者、SGMLのコミュニティなど

の関係者が、アメリカ・オハイオ州ダブリンに集まり、webで公表される文書を記述するための、データ要素について議論した。この地にちなんで、ダブリン・コアと名付けられることとなった<sup>5)</sup>。

その後、1997年に、15の基本的なメタデータセットからなる要素表が作成された。これらの要素は、1. コアセットとして、情報資源の本来の属性を簡素に記述する。2. 要素は、どのコミュニティで用いられようとも、意味は同一であって、いわゆる意味的な相互運用性がある。3. これまでの各メタデータとの接合を前提として、拡張性や運用上のオプションを考慮した柔軟性をもたせる。という原則の下に作成された。また、「構文独立の原則」をとまえ、構文のいかんをとわず、要素セットにもとづき、相互運用性を確保するものとして作成することとした。このダブリン・コアは、さまざまな拡張が行われ、現在では、図書館情報学などの分野では積極的に用いられているものである。

一方、EADとは、アメリカアーキビスト協会において作成された、アメリカのアーカイブズを検索し、かつ記述するためのDTDのセットであった。もとは、独自の規定であったのだが、その整備の過程において、ISAD (G) との互換性を意識し、作成されるようになった。そのため、アーカイブズ・公文書館が、目録類のデジタル・アーカイブを行うのにもっとも適したメタデータとして策定されたのである。現在、EADのDTDはISAD (G) の記述要素を含みこむかたちで作成されている。結果的に、EADによって、紙媒体での資料群の階層構造を提示した記述というイメージのまま、同時にコンピュータならではの便利な検索機能を持ちえた、との評価を得ている<sup>6)</sup>。

この両者が、現在のメタデータの記述方法の二大潮流とあってよいであろう。また、メタデータの統一の動きとしては、TEIがあげられるが、TEIは資料それ自体のメタデータの記述方法という面が強く、目録類などのデジタル・アーカイブには必ずしも向いていない現状がある。

この両者の現在の状況をあわせて考えると、デジタル・アーカイブのためのメタデータは、形式はXMLによって、内容はダブリン・コア

かEADによって作成されることが、近年の大きな潮流となってきている。

#### 4. デジタル・アーカイブ化への適用

この二つのタグセットのうち、デジタル・アーカイブ化にあたり、本事例においては、EADを基礎的なタグのセットとすることにした。EADには、以下のメリットがあるためである。

1. EADは、国立公文書館・国文学研究資料館アーカイブズ系など各地で用いられており、タグが同一となるため、共通検索が容易に可能となる。
2. アーカイブズをデータベース化するためのメタデータであり、適用が比較的容易。
3. EADをもちいることがデファクト・スタンダードになりつつあり、このメタデータを用いることがもっとも穏当。

そこで、仮にEADに準拠し、XMLデータベースとして作成することとした。

日本においても、現在、国立公文書館、国文学研究資料館をはじめとして、日本各地の公文書館・文書館が、EADを採用している。国立公文書館のEADの定義集・およびサンプル集は、国立公文書館のwebサイトより入手可能である<sup>7)</sup>。また、国文学研究資料館のEAD適用事例も同様に、国文学研究資料館アーカイブズ系のwebサイトより入手可能である<sup>8)</sup>。国文学研究資料館での適用事例は、五島敏芳が、詳細な研究をまとめている<sup>9)</sup>。

また、最近では国立国語学研究所もEADによるデジタル・アーカイブの構築を行い、その有用性を証明している<sup>10)</sup>。本事例のデジタル・アーカイブ化に際してのEAD適用にあたり、これらの事例・研究を参照した。特に、本プロジェクトで扱う行政資料という資料のもつ性質、および、京都府立総合資料館が、行政文書の受入を行う機関であるという所蔵館の性質を考慮し、国立公文書館の定義に、可能なかぎり従うこととした。

現在、プロジェクトが把握している資料構造と、その結果、作成される調査票について説明する。京都府立総合資料館の資料構造は図1の

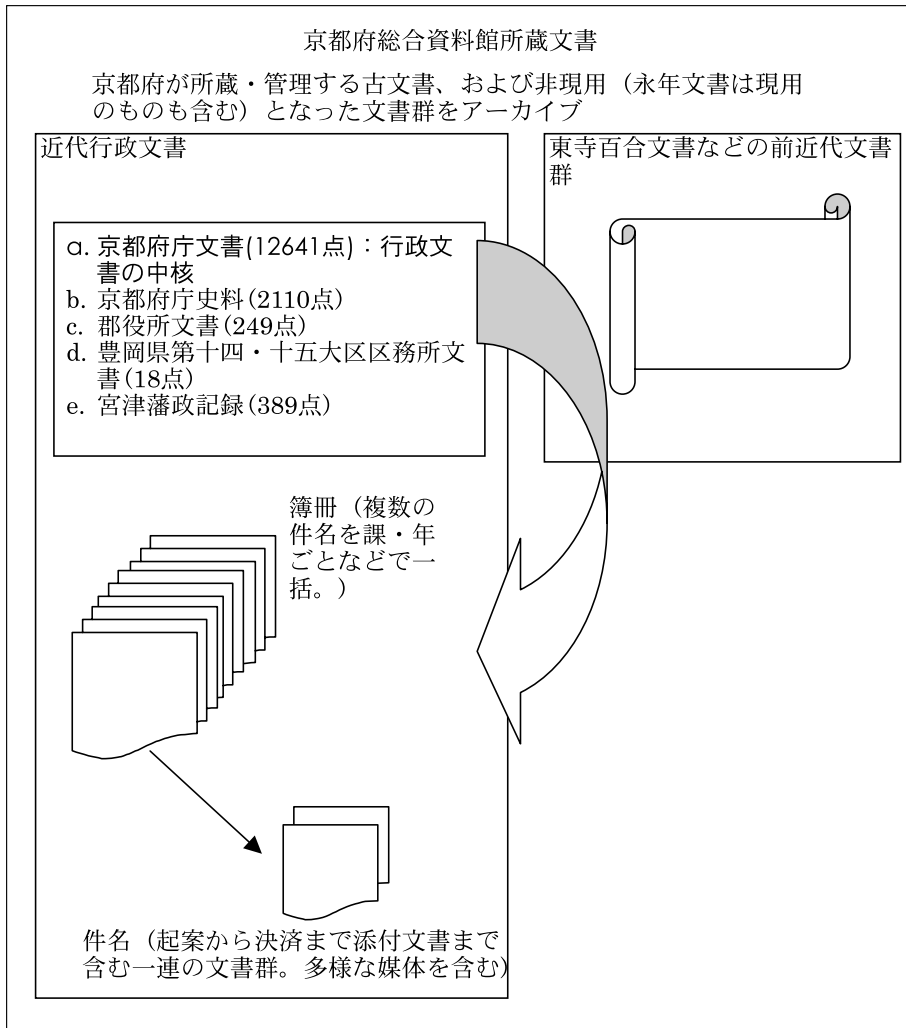


図1 京都府総合資料館の資料構造 (近代行政文書の構造のみ)

通りである。

基本的には、京都府立総合資料館所蔵の文書群、京都府行政文書、第2節にて記述した史料群・簿冊・件名という順に細分化される資料構造をもっているものとして把握した。

この資料構造が存在することを前提とし、作業を行った。史料群の構造は、資料の全体構造が判明してからとし、まずは、簿冊と、件名という二つのレベルでの目録とりから開始した。簿冊の目録は、資料自体の劣化状況などを図ることが主たる目的であるため、法量・全体の簡単な内容・簿冊全体の劣化程度についての特徴を記すものとなった。そして、簿冊の下位レベルの情報を「件名」と定義し、件名を調査票の最小レベルとして、作成することとした。件名とは、「起案から施行報告の文書と添付資料をまとめて一件とする」という定義である。

さて、本プロジェクトにおける目的は、主として、行政史料の構造と、劣化状態の把握にある。そこで、件名に関しては、歴史学における基本的な目録情報である、件名・作成者、宛、作成日付・形態、数量といった情報に加え、書いてある媒体・その罫線・インキの質・版心(文書の作成主体の調査のため)といった、詳細な媒体情報と、それぞれその劣化状況について記述することとした(表1の左部を参照)。これは、資料の劣化状況の把握のためである。媒体や、インキの質などにより、劣化の状況が異なる可能性を考慮したものである。

これをもとにXMLとEADによるデジタル・アーカイブ化をほどこした。その結果、表1のような対応関係となった。

各項目について説明し、デジタル・アーカイブ化への課題を抽出する。

作成：資料の件名を記載する。対応するタグは、unittitleである。

作成者：作成課を基準とした作成者を記載する。対応するタグはorigination。

宛：史料自体に宛が記載されている場合、宛を記載する。対応するタグに関しては後述する。

作成年月日：起案日に関わらず、含まれている文書の中で、最も古いものの年月日を記載する。ただし、近世文書の写しなどはとらない。対応するタグは、unitdate。西暦も併記する形式となっているが、両者同じ要素(タグ名)とし、label属性により、処理する。

内容：一件の文書の概要を簡略に記す。対応

表1 調査票の項目とEADの対応表

| 調査票の項目    | EADの対応タグ     | 備考  |
|-----------|--------------|---|
| 史料（簿冊）番号  | unitid?      |   |
| 件名番号      | unitid?      |   |
| レファレンスコード | unitid       |   |
| 作成        | origination  |   |
| 宛         | ?            |   |
| 史料名       | unittitle    |   |
| 作成年月日（和暦） | unitdate     | label属性に"作成年月日"と記述することにより西暦と区別                |
| 西暦        | unitdate     | label属性に"西暦"と記述することにより和暦と区別                   |
| 内容        | scopecontent |   |
| 媒体        | genreform    |   |
| 罫線の色      | physfacet    | label属性に"罫線の色"と記述することにより、他のphysfacet属性と区別     |
| 版心        | physfacet    | label属性に"版心"と記述することにより、他のphysfacet属性と区別       |
| 記法（劣化程度）  | physfacet    | label属性に"記法（劣化程度）"と記述することにより、他のphysfacet属性と区別 |
| 媒体の劣化     | physfacet    | label属性に"媒体の劣化"と記述することにより、他のphysfacet属性と区別    |
| 媒体の枚数     | extent       | 全体枚数とは、階層の差を設けて区別                             |
| 備考（特徴）    | phystech     |   |
| 全体枚数      | extent       | 媒体の枚数とは、階層の差を設けて区別                            |
| 調査日       | processinfo  |   |
| 調査者       | ?            |   |

するタグはscopecontentである。

以上は、いわゆる歴史学の目録とそれほど変わるところがない。ただ、作成年月日は、起案～決済まで、複数の書類を一件として作成するので、複数の年紀があらわれることとなるが、今回は、調査の目的などを鑑み、最も古い日付を記載することとしている。

次に、本資料調査の第二の目的である、媒体情報の項目について説明する。

記録媒体：構成する素材によってA～Jの10種類に分類し、その記号を書き込む（A～Jの詳細は、本稿では省略する）。該当するタグは、genreformを用いた。資料の物理的状态を記録するという目的、また、国立公文書館のEADのサンプルにおいて内閣文庫の物的特徴の中に料紙情報などが含まれることを考えると、physfacetに分類される可能性も検討したが、「媒体」であるということを優先し、genreform要素を用いることとした。

罫線の色・版心・記録方法・媒体の劣化程度：それぞれ、記号を決定し、記号形式で簡便に入力することとした。対応する要素は、すべて、physfacetに入力し、label属性により、区別することとした。これは、EADにおいては物的情報に関する要素として、physfacetと、phystechしか存在

しないためである。phystechは、本来の要素の意味が、“Physical Characteristic and Technical Requirements”であり、属性としてlabelを含むことができない。そのため、phystechは、次に記す特記すべき媒体情報（備考）と対応させることとし、physfacetをこれら記号で記す物的特徴に記すこととした。

備考（特徴）：その他、調査票の入力者が気づいた情報を書く。対応するタグは、前述の通り、phystechである。

枚数：一つの媒体単位での枚数を記す。対応するタグは、extentである。

全体枚数：枚数の合計である。この全体枚数は、同時に「件名」一件の枚数に該当する。対応するタグは枚数と同様、extentである。ただし、階層が一階層上となる。

調査日：調査票に入力した日付を書く。対応するタグを、アーキビストノートの中のprocessinfoとする。ただし、現在のところ、他にアーキビストノートは記述しない。

調査者：調査票に入力した、担当者の名前を記す。対応するタグは、現在のところ、存在しない。

これらのタグの対応によって調査票をEADに対応させ、共通の情報による検索が可能となるしくみを構築した。EADのデータは図2に記す。

|   |  |        |
|---|--|--------|
| <pre> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;ead audience="external"&gt;   &lt;eadheader audience="internal"&gt;     &lt;eadid 1 /&gt;     &lt;filedesc&gt;       &lt;titlestmt&gt;         &lt;titleproper&gt; 京都府総合資料館所蔵文書 &lt;/titleproper&gt;         &lt;author&gt; 京都府総合資料館 &lt;/author&gt;       &lt;/titlestmt&gt;     &lt;/filedesc&gt;   &lt;/eadheader&gt;   (中略) </pre>   |  | 史料群の情報 |
| <pre>     &lt;did&gt;       &lt;unitid identifier="000000001"&gt; 昭-17-224 &lt;/unitid&gt;       &lt;unittitle&gt; (昭和17年) 大政翼賛運動 &lt;/unittitle&gt;       &lt;unitdate label="作成年月日"/&gt;       &lt;pysdesc&gt;         &lt;extent label="枚数"&gt; 1枚 &lt;/extent&gt;       &lt;/pysdesc&gt;       &lt;origination label="作成部局"&gt; 振興課 &lt;/origination&gt;       &lt;scopecontent&gt;         &lt;p&gt; 裏あり 付箋あり &lt;/p&gt;       &lt;/scopecontent&gt;       &lt;physdesc&gt;         &lt;genreform type="記録媒体"&gt; 洋紙白紙1枚 &lt;/genreform&gt;       &lt;/physdesc&gt;       &lt;materialspec /&gt;     &lt;/did&gt;     &lt;c level="item"&gt;       &lt;did&gt;         &lt;unitid 9 /&gt;         &lt;unittitle label="件名"&gt; 講演会に関する件 &lt;/unittitle&gt;         &lt;unitdate label="作成年月日"&gt; s17/4/5 &lt;/unitdate&gt;         &lt;physdesc&gt;           &lt;extent label="全体枚数"&gt; 8枚 &lt;/extent&gt;         &lt;/physdesc&gt;         &lt;origination&gt; 作成者名 &lt;/origination&gt;         &lt;scopecontent&gt;           &lt;note label="史料内容"&gt;             &lt;p&gt; [会場・日時・主催者] &lt;/p&gt;           &lt;/note&gt;           &lt;note label="史料内容"&gt; 満州国建国十周年講演会開催の件 &lt;/note&gt;           &lt;note label="史料内容"&gt; 講演会に関する件回報 &lt;/note&gt;           &lt;note label="史料内容"&gt; 職員出張の件 &lt;/note&gt;           &lt;note label="史料内容"&gt; 封筒 &lt;/note&gt;         &lt;/scopecontent&gt;         &lt;processinfo type="記述規則"&gt;           &lt;p&gt;             &lt;date normal="20050905" type="記述作成年月日"&gt; 2005年9月5日 &lt;/date&gt;           &lt;/p&gt;         &lt;/processinfo&gt;         &lt;processinfo type="入力者のノート"&gt;           &lt;p&gt; 記述者のノート &lt;/p&gt;         &lt;/processinfo&gt;       &lt;/did&gt;     &lt;/c level="itemdata"&gt;       &lt;did&gt;         &lt;unitid 001 /&gt;         &lt;physdesc&gt;           &lt;genreform type="媒体"&gt; D &lt;/genreform&gt;           &lt;physdesc&gt;             &lt;physfacet label="罫線の色"&gt; 青 &lt;/physfacet&gt;             &lt;physdesc&gt;               &lt;physfacet label="記法 (劣化程度)"&gt; c &lt;/physfacet&gt;               &lt;physfacet label="記法 (劣化程度)"&gt; e &lt;/physfacet&gt;               &lt;physfacet label="記法 (劣化程度)"&gt; p &lt;/physfacet&gt;             &lt;/physdesc&gt;             &lt;physfacet label="版心"&gt; 青 &lt;/physfacet&gt;             &lt;physfacet label="媒体の劣化"&gt; 5-1 &lt;/physfacet&gt;             &lt;extent label="媒体の枚数"&gt; 1 &lt;/extent&gt;           &lt;/physdesc&gt;           &lt;phystech&gt;             &lt;p&gt; 起案 &lt;/p&gt;           &lt;/phystech&gt;         &lt;/did&gt;       &lt;/c&gt; </pre> |  |        |
| <pre>     (中略)   &lt;/dsc&gt; &lt;/archdesc&gt; &lt;/ead&gt; </pre>   |  | 件名の情報  |
|   |  |        |

図2 EADにより作成・構造化したデータ

(媒体情報を除き、資料の構造をそのままXMLの構造に転換している)

## 5. 課題

一方、下記のような課題があげられる。

これらの課題自体は、デジタル・アーカイブ化に際して、EADの非有益性を示すものではない。しかし、日本において、デジタル・アーカイブにEADを適用していく際の重要な論点となるであろうと考えられる。

### ・EADには、「宛」が存在しない。

表1を見れば一目瞭然なのだが、EADのタグのセットの中には「宛」が存在しない。これは、ISAD(G)にもとづいて共通化された目録を日本の目録システムにあてはめる根本的な問題である<sup>11)</sup>。ISAD(G)は、当初の目的が、公文書のアーカイブに関する原則を定義したものである。現代の公文書類は、原則として「宛」をもつことがない。そのため、宛が必要ないと理解されたものとする。

また、以下は、推測にわたるが、ISAD(G)が文書の出所単位の原則を踏襲する以上、宛は、ほぼ同一になるか、想定されないと考えるべきではないだろうか。そのように考えると、宛が想定されない状況もISAD(G)の策定時に、宛という定義がなされなかったことも無理のないことであろう。宛が出所と同一でない場合は、例外として、〈note〉などに記述されることとなるからである。

### ・媒体と内容の問題

現在の調査票においては、資料の最小レベルを件名としている。件名は、「起案から施行の文書と添付資料をまとめて一件とする」という定義である。これは、一件の中に複数の媒体を含みうるということの意味している。このこと自体は、さほど問題ではない。EADは、複数の要素を含むことを認めているからである。問題となるのは、一件あたりの内容と媒体の数が異なる事例である。図3のように、内容と媒体が一对一として対応しない事例は普通に考えうる。複数の質の異なる料紙に同一の内容を記述する。または、同一の媒体に、内容を追込み式で書くという形式である。この場合、一つの解決方法としては、より下位のレベルのデータ

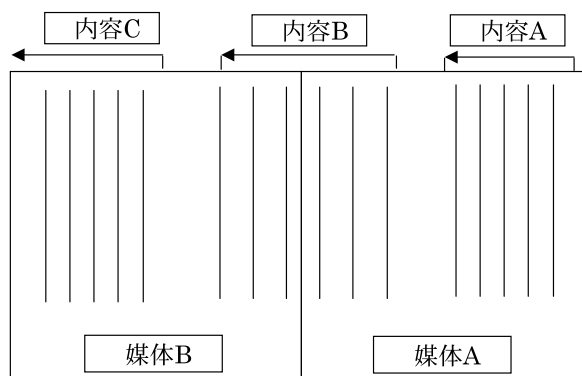


図3 媒体と内容が一对一で対応しないモデル図

をおくことが考えられる。通常のISAD(G)の資料構造の原則を鑑みれば、媒体を最小単位とした構造を作成することになるだろう。しかし、媒体を最小単位とした場合、同一内容が複数媒体にわたってしまう。逆も同様の事態が起こるのである。結果的に、どのレベルを最小単位として、XMLを記述すべきかということに混乱が起こるのである。

### ・〈unitid〉による資料記述表現の困難さ

本来的に〈unitid〉は「資料の構造を反映する」ことを前提としている。しかし、現在の国立公文書館のサンプルにおいては、〈unitid〉に、レファレンスコードが割り振られている。そして、このレファレンスコードは、請求番号などの、検索に用いるための番号として定義されているのである。そして、資料構造を反映するコード類が、資料固有情報へと割り振られている。つまり、〈unitid〉が、資料構造を反映しない構造になっているといえるのである。しかし、〈unitid〉が、EAD記述の構造の基礎となることを考慮するならば、本来的には望ましい形とは言えない。しかし、データの入力と、目録の作成の段階が異なる場合には、往々にしてありうることである。この両者に常に整合性をもたせることが重要である。しかし、作業の過程としてはきわめて困難であるといえる。

### ・メタデータのメタデータが必要

EADの定義群においては、データの作成日付・公開条件・検索手段といった、「公文書館のための情報」、「もしくは文書管理の実務のた



めの情報」は記述が求められているが、データベースの作成者・作成日付・目録の作成者・メタデータを付与する根拠を記す要素群が存在しない。これは、ISAD(G)とEADが本来的には、「公文書館など、文書管理実務で用いられる」ことを目指しているためであると考えてよいだろう。

ただ、歴史的な文書・資料群の目録や、メタデータを作成する際には、このアーカイブズ保存業務に加えて、メタデータのメタデータは必要ではなかろうか。既存の目録は、歴史学における古文書学の理解を自明の前提として作成されてきた。そのため、これまでメタデータがどのように編成されてきたのか、などと言った点は、重視されてこなかった点ではある。しかし、目録の編成が、(そのようなものがあるとすればだが)資料それ自体に対する要請に応じて、もしくは、ある種の目的をもって作られるものであり、その目的に則した知識の把握と集約の結果であることを考えると、メタデータの編成にはその根拠が必要ではないかと考える。既存の情報をEADのどの要素にあてはめたのか、その根拠は、などといった情報が、現在は、データの外にしか存在しない。これらの情報をデジタル・アーカイブの内部において記述することは、何らかの形で必要ではないだろうか。

たとえば、国文学研究資料館史料館(現在同資料館アーカイブズ系)が作成した『史料館収蔵史料総覧』はISAD(G)の記載に従いつつ、各文書群を概括し、説明した、きわめて先駆的な仕事である。この『史料館収蔵史料総覧』においては、各文書群の伝来を詳細に記しており、文書群の構造がよくわかる構造となっている(ただし、この伝来はISAD(G)の項目に入る部分である)。この場合、文書群の構造をどのように理解したか、という情報が、より仔細に載っている目録の先駆的事例としてとらえられるだろう。しかし、デジタル・アーカイブ化する際に、EADとのクロスワークを実現しようとするれば、いかなるデータをいかなるポリシーで入力したのか、を明示する必要がある。そのメタデータが明示されないかぎり、検索において、使いにくいという難点を抱えると同時に、目録自体がもつバックボ

ーンを理解することができないという弱点をもつ。現在、EADのなかでは、アーキビストノート(〈processinfo〉)を用いざるを得ない。なお、メタデータのメタデータとしてはISAAR(CPF)があるが、これはISAD(G)の各記述レベルにおける資料来歴を示すものであり、目録編成の前提となる情報を示すものではない。

## 6. 考 察

上記のような課題はある。宛がない、構造を正確に記述できないといった点は、EADが近代行政文書のデジタル・アーカイブ化のためのメタデータ記述として不適であるということではない。また、EADは、拡張を認めていないわけでもない。しかし、デジタル・アーカイブ化にあたって、いくつか考えなければならない問題があるのもたしかである。まずは、そもそも、EADを適用した上で、資料構造は記述可能かという点について考察したい。

今までの研究において、簿冊レベルまででは、EADの適用が問題なく行われることが判明している<sup>12)</sup>。本事例においても問題なく記述できる。また、今回の検討において、資料構造としてはより下位である件名レベルまで適用可能であることが判明した。〈scopecontent〉を複数記述し、〈physfacet〉、〈physdesc〉などといった、タグを複数記述すれば問題ない。この点は、調査票において、紙の目録に記載する際にも、それほど問題とならなかった。しかし、デジタル・アーカイブとする場合には、若干の問題が生じている。現在、筆者が作成しているモデルにおいては、内容を件名のレベルでの情報ととらえ、個別の媒体を、件名より、さらに一つ下位のレベルにおくことで仮に処理している。この処理方法では、どの媒体にどの内容が記載されているかが、データ上で明示されないという問題点が残る。複雑な媒体と内容を並行に記述することは、そもそもXMLにおいて困難である<sup>13)</sup>。内容、媒体どちらかの情報を分割する、もしくは削除せざるを得ない。両者の情報とも、余すことなく表現することが困難であることがわかる。

今までの紙媒体での目録記載であれば、自由度の高い記述が可能であったため、媒体と内容が複雑に関連する場合の対処も容易であったであろう<sup>14)</sup>。しかし、デジタルデータとして、それも、XML形式という線形テキストの構造を維持した形式で、この構造を表記しようとすると、無理が生じるのである。EADはSGML、もしくはXMLで記述する以上、可能なかぎり、well-formedであることを要求される。しかし、上記の構造を可能な限り忠実に再現しようとするならば、well-formed形式のXML文書としては記述できないことがわかる。EAD/XMLによるデジタル・アーカイブ化は、紙による目録編成や、既存の歴史学での目録編成では、今までさほど問題とならなかった、媒体と内容との連関性、もしくは、非連関性を如実に示すものとなったといえるだろう。

また、unitidの要素の中に記述されるべき、資料構造を反映するIDと、資料調査時に資料に付される番号が一致しないという点も、資料構造を記述するという意味において望ましいものとは言いがたい。

本プロジェクトにおける調査票においても、簿冊番号—資料固有の件名番号という独自の項目を作成し、その項目とは異なる、別個のIDとして、レファレンスコードを作成する構造としている。簿冊番号—件名番号が資料構造の基礎を示す指標であることは言を俟たない。しかし、この簿冊番号—件名番号は、プロジェクト上特有の番号として処理せざるをえないため、結果的にレファレンスコードを別に設定することとなった。レファレンスコードは、簿冊番号と件名番号の構造を配慮したうえで作成されるべきものとなるであろう。

EADの一部であり、かつ、現在ではEADの主要な要素を構成するようになっているISAD(G)によって提示されている原則は、以下の二つである。1. 原秩序尊重の原則と、2. 共通化された目録記述の2点である。そして、EADも、ISAD(G)に可能な限り対応するように作成されている。しかし、本来EADは、「多角的な検索手段」の構築を目指すものであり、EADそれ自体は、厳密な意味において資料構造を記述することが第一義とされていない。し

たがって、デジタル・アーカイブ化に際しての資料構造の記述、という点に関して、EADに過度な期待を求めることは危険なのではなかろうか。

たとえば、日本において、EADを文書に適用した過去の研究の中で「史料構造の記述を同時に行ってもなんらの不自由がない」<sup>15)</sup>という指摘が行われている。ここまでの評価であれば問題ない。しかし、EADによるデジタル・アーカイブ化が「紙媒体での史料群の階層構造を提示した記述というイメージのまま、同時にコンピュータならではの便利な検索機能を持ち得た」<sup>16)</sup>という評価へとつながるようであれば、やはり、日本におけるデジタル・アーカイブ化におけるEADの理解には、若干の誤解があるのではないだろうか。

そもそも、ある目的を果たすためのデータベースであれば、要素と構造を分離し、目的に応じて、XSLT<sup>17)</sup>により構造をつくるようなしくみの可能性も考えられる。

京都府立総合資料館の事例では、上記の調査票の項目を要素化し、構造化せずにタグとして並列に作成する。その後、

1. 京都府行政文書の媒体に関するデータベース用の構造
2. 資料館の文書構造に則し、資料の状況をデジタルデータで表示できる構造
3. そして、共通のEADのDTDにしたがった構造

の三つの構造を作成できるようにXSLTを作成することが、より効果的である可能性がある。

このように、目的と状況に応じて構造を作成するしくみであれば、資料の構造を理解し、一般化した後に、XSLTによるコンバートを行うことにより、資料の構造記述を行うことも可能である。むしろ、EADに準拠するかたちへと変更することも可能であるし、そのEADに準拠したかたちが、資料構造を反映するものであれば、なおよい。検索の用としても十分に機能する。

この意味では、EADではなく、「構文独立の原則」をもつ、ダブリン・コアのほうがより効果的である可能性もある。この可能性の検討は今後の課題であるといえる。

五島敏芳は、「電子的検索手段は、紙媒体のばあいその物質的特性にゆだねていた機能をすべて情報的に再現しなければならない」とし、今までの紙カードの構造が「アーカイブズ管理上の理由」であるかぎり、その構造・情報を、可能なかぎり反映させるべきであるとしている<sup>18)</sup>。しかし、ある種の目的に従い、自由にデータを再構成できるのはデジタルデータの本質的な強みである。また、デジタルデータに符号化をほどこす以上、紙媒体として可能であった二次元的な情報は（少なくともXMLなどの検索可能なテキストデータへと変換する以上）、線形の情報へと変換せざるをえないのである。そのため、捨象される情報は、必ず生じてくるのである。

そのように考えるならば、固定された、汎用のメタデータにより構造化されているデータより、目的に応じ可変であるデータのほうが、より永続性・流通性を高める可能性を持っているのではないだろうか。それこそが、新たな形のアーカイブズであるデジタル・アーカイブのありようの一つであろう。

もっとも、それすら、中立的なデジタル・アーカイブではありえない。要素が、ある種の観念によって決定される以上、その要素はある種の「政治性」をおびてしまうからである。しかし、そのためにも、資料記述の構造よりは、どの要素に、どの資料の知識をいれるべきか、その点に重点をおくべきではなかろうか。そのことは、今までの目録記述と、デジタル・アーカイブへ向けてのEADとの要素の共通化という新たな課題へと進むと思われる。

そこで、特に、文書の「宛」に注目したい。「宛」は、EADに記述されない要素である以上、基本的には、独自に配列することになる。今までの歴史学が作成してきた目録の中で、EADに存在しないのは「宛」のみであろう。この「宛」は少なくとも日本国内において、どの要素に記述すべきか、共通化することを考えなければならない。

「宛」に関しては、他のEAD適用事例も、それをどのタグに入れるべきか苦心しているように見受けられる。国立公文書館のEAD定義集においては、「宛」の項目がない。また、国

文学研究資料館のEADサンプルにおいては、「宛」と作成を〈controlaccess〉にしている。controlaccessは、“A wrapper element that designates key access points for the described materials and enables authority-controlled searching across finding aids on a computer network.”と定義されている。つまり、コンピュータ上で見つけるための、基本的なアクセス情報であるという扱いである<sup>19)</sup>。この事例に見られるように国文学研究資料館において提示されているように、〈controlaccess〉に記述することが妥当なのか否かも含めて議論する必要がある。

## 7. おわりに

本稿においては、文書群としての資料ではなく、より細かい、件名レベルにおけるEAD適用の方法論について検討し、デジタル・アーカイブにおける、課題の一つを分析した。

EADが、その本質において、アーカイブズのためのメタデータであることを考えるならば、EADによって記述されたデジタル・アーカイブは、「アーカイブズの目的」にもとづいたものとなる。この「アーカイブズの目的」にそったメタデータと、歴史学のため、もしくはある種の目的のためのデータベースやデジタル・アーカイブのメタデータのための情報群とが齟齬をきたすことは、当然おこりうる。アーカイブズには、資料の全き保存を目指すという文脈があり、EADはその文脈に従ったメタデータである。そのため、資料を保存するという文脈においては、EADは優れたメタデータである。しかし、そのメタデータは必ずしも、万能のメタデータではない<sup>20)</sup>。

たとえば、今まで日本の歴史学が重視してきた、「宛」のようなデータを資料個別の情報として、ある種、「日本固有の特殊な記述」として処理することを「世界共通」としなければならないのはなぜだろうか。それは、資料の全き保存のためによいことだろうか。これらを資料個別の情報として、小さく処理してしまうのではなく、なぜ、日本の歴史学では、「宛」が必

要な目録であったのか。そして、その理由を考察し、その結果を踏まえて、世界共通のタグのいかなる箇所にいれるべきか、議論する必要がある。日本の史料群に対して、デジタル・アーカイブを行うときにはその議論が不可欠であり、EADを事例として適用することにより、可能な議論となったともいえるのである。

EAD自体は、「純粋に」資料を記述するものではない。公文書のアーカイブズとして必要な、いくつかの要件を満たすためのタグセットにし過ぎない。したがって、どのようにEADを拡張し、どの部分を共通化し、どのように用いるかを、丁寧に考える必要があるであろう。これらのさまざまなメタデータの適用を通じ、資料をデジタル・アーカイブ化することは、私たちが、今まで認識してきた歴史資料をはじめとするさまざまな資料群を再考することにもつながるのである。

## 注

1. 京都府立総合資料館では、本稿で扱う文書群のことを、京都府行政文書と、定義しているが、本稿では、アーカイブズ学などの議論の共通性をはかるため、事業概要などの引用を除き、原則として、「京都府行政資料」の語を用いている。
2. なお、本稿ではデジタル・アーカイブを、さまざまな記録、およびその目録情報を電子化し、さまざまな記録情報の管理と保存の用に資するものとして作成されるものであり、時間的・空間的な流通性をもつデジタルデータとして定義している。いわゆる、一般語としてのデジタルアーカイブとは若干異なる定義として使用している。
3. 京都府立総合資料館のこのプロジェクトは、
  1. 近代行政文書に特徴的な形状・素材の多様性・多様な筆記方法などを分類整理して損傷状態の評価方法を検討する。
  2. 「京都府行政文書」の多種多様な内容を史料論的観点から整理分類する基準を検討し、本資料の中心を占める簿冊類を形態論的に検討する。年次ごとに整理された現在の目録に加えて、確定された分類基準に則

したデータベース化を推進する。また、劣化状況などの保存状態もデータとして集積する。

3. 府内各地および他府県の行政文書の保存実態や保存方法についても調査を行い、適切な保存および調査のあり方を検討する。各市町村の行政文書と「京都府行政文書」との体系的連関について解明し、近代行政文書の体系的把握を目指す。
 

という三つを主要な課題としており、京都府の行政資料を素材として、近代行政資料を総合的に、史料学的にとらえることを目的としたものである。
4. 五島芳敏「アーカイブズ情報の電子化とネットワーク」（『アーカイブズの科学』下巻、柏書房、2003）
5. 永田治樹「アーカイブズと図書館情報学」（『アーカイブズの科学』上巻、柏書房、2003）。ダブリン・コアの情報は、<http://dublincore.org/>にて入手可能（2006年5月10日現在）
6. 森本祥子「アーカイブズ情報の電子化とネットワーク」（『アーカイブズの科学』下巻、柏書房、2003）。EADの詳細な情報・タグの詳細な説明は、<http://www.loc.gov/ead/>にて入手可能。（2006年5月10日現在）
7. 国立公文書館のEAD定義集・サンプルは<http://www.digital.archives.go.jp/howto/index.html>にて入手可能。ここの定義集を参照した。（2006年5月10日現在）
8. 国文学研究資料館のEADのサンプルは、<http://archives.nijl.ac.jp/DAS/projects/EADFAprj.htm>にて入手可能（2006年5月10日現在）。ここのサンプルを参照した。
9. 五島敏芳「日本の記録史料記述EAD/XML化と記録史料管理：記録史料管理過程におけるEAD利用の位置をめぐって」（『情報知識学会誌』12巻4号、2003）
10. 森本祥子「EADを用いた資料記述システムの開発」（『アーカイブズ学研究』4、2006）
11. ISAD(G)とは「記録資料記述の国際標準」と訳されるものであり、記録資料を管理、活用する際にもちいられることが望ましいとき

れている、世界的な取り決めである。日本語訳は『記録資料記述の国際標準』（アーカイブズ・インフォメーション研究会訳、北海道大学出版会、2001）にて参照することが可能である。

12. たとえば鈴江英一『近現代史料の管理と資料認識』（北海道大学出版会、2002）などで、近現代の目録編成と、標準化について詳細に論述しており、そのなかで、ISAD(G)に準拠した形で、簿冊レベルまでの記述が十分可能であるとの認識を示している。
13. 日本古代のテキストではあるが、後藤真・柴山守「正倉院文書復原過程のXML/XSLTによる記述」（『情報知識学会誌』11巻4号、2002）にて、紙媒体に複雑に記述されたテキスト情報が、well-formedになりえない場合の対処法を検討した。
14. 今までの紙ベースでの目録編成論は、目録にいかん情報配列するかという議論が中心で、いかなる要素が必要かという検討は行われなかったように思われる。そして、EADも同様にどのような階層構造で記述するかに主眼があったようである。しかし、コンピュータ上でのデータ入力を前提とし、多角的検索

手段を提供することを目的とすれば、目録内部の配列や、構造を議論することは、あまり意味がない。

15. 注4五島前掲論文
16. 注6森本前掲論文
17. XSLTとは、XMLを変換するための言語。この機能により、XMLは自由にデータの構造を変換することができる。
18. 注4五島前掲論文
19. ちなみに、この要素は、ISAD(G)とのクロスワークでは定義されていない。国文学研究資料館は、このcontrolaccess要素を多用している。
20. EADによる記述を前提として、資料を見ること自体が、資料構造の認識に影響を与える可能性も考えなければならない。

#### 付記

本稿を書くにあたって、京都府立総合資料館の福島幸宏氏をはじめ、資料館のみなさま、その他関係のみなさまより、多大なるご助言をいただいた。ここに記して、感謝の意とさせていただきます。

## The Possibility and Problems when Making “the Kyoto Prefectural Library and Archives” into Digital Archives

Makoto GOTO

This paper describes the usefulness and problems at the time of applying EAD in the case of digital-archiving modern administration data. EAD (Encoded Archival Description) is the set of the metadata for archives, such as a document hall and the Public Record Office common to the world. Especially in this paper, the Kyoto-fu archives possessed by the Kyoto synthesis archives were used as an application example of EAD. Based on this example, historical data, including modern administration data, were made into metadata, and the possibility of converting it to the digital archives for circulating database creation and in time and space was described. Examination was added focusing on the argument of how to use especially metadata at that time. Consequently, it became clear that there was a portion with EAD able to describe structure and a very difficult portion. It was difficult to describe structure as digitized data in the example by which especially the medium and the contents are not equivalent by 1 to 1. This does not show directly that EAD is not useful. However, it is shown that EAD can not describe the structure of data neutrally. And by the table of the history of existing Japan, the point that the item which was contained that naturally did not exist in the element of EAD etc. became clear. Existing in the table of the history of Japan, what is not in the element of EAD is shown, and performing the analysis shows that it is a human science and that it is especially possible for an understanding of the data analysis of history.

Keywords : digital archive, Kyoto-fu administration document, archives,  
research for historical material, EAD