

EADを実装したアジア歴史資料センター 新情報システムによせて —EADの概要とアジア歴史への期待

人間文化研究機構国文学研究資料館
(アーカイブズ研究系)

五島 敏芳



1. EADの概要

EAD、Encoded Archival Description (符号化記録史料記述) は、アーカイブズ (記録史料、永久保存記録) の検索手段を電子的に符号化するための事実上の (デファクト) 国際標準です。検索手段は、日本では資料目録が典型的で、アーカイブズの実物へたどりつくための道具です。それを、具体的にはコンピュータで扱える電子データへ加工するときの枠組みが、EADです。

図書館の例をあげたほうが、わかりやすいかもしれませんが。図書館では、収蔵資料を検索するのに、いまやオンライン閲覧目録 OPAC が一般的で、その目録レコードは MARC、MACHINE-Readable Cataloging (機械可読目録) の形式です。この MARC のアーカイブズ版が、EAD だ、と言えます。そして EAD の開発は、図書館の世界からはじまりました。EAD は、米国カリフォルニア大学バークレイ校図書館の研究プロジェクトに端を発し、いくつかの団体と専門職協会により後援された協同の成果としてその内容が定められ、現在アメリカ アーキビスト協会 SAA と米国議会図書館ネットワーク開発・MARC 標準事務局によって維持されています。

EAD は、アプリケーション ソフトウェアに依存しないコンピュータのファイル形式であるテキストファイルを利用します。たんなるテキストファイルではなく、電子的に文書の体裁を再現し意味内容を区別できる SGML、Standard Generalized Markup Language (標準汎用マークアップ言語) および XML、eXtensible Markup Language (拡張マークアップ言語) という国際規格類にそくしたものです。はじめ EAD の第 1.0 版では SGML だけに対応していましたが、2002 版から XML へも対応しました。

EAD が SGML/XML を用いていることから、これまで不可能だったことが可能になりました。いずれもアーカイブズの記述の特徴にかかわります。図書館の OPAC の目録レコード 1 つは、ふつう 1 冊の図書や雑誌の書誌的情報をあらわすので、極端に長大な情報とはなりません。しかし、アーカイブズの資料群 1 件全体について解説すれば、その標題や数量、年代、出所 (作成者) くらいまでは箇条書きにできるかもしれませんが、その出所の履歴、内容、資料群の伝来、利用条件、等などの情報は長い文章でつづられた解説 (解題) となるのが一般的でしょう。しかも、その資料群に

はしばしば、いくつかのまとまり（小群）があり、さらに各小群にはより小さなまとまりが含まれることもあり、そのなかに1冊ずつ書類があり、各書類には複数の事柄が収められている（これらはマルチレベル記述で表現できる）。このような長大な情報で、入れ子のように階層的に構成され、その階層も伸縮自在とあっては、図書館のOPACの目録レコードの枠へ押し込めるのは無理があり、OPACでは扱えなかったのです。SGML/XMLを使えば、タグによって情報と情報の区切りを示し、そのなかにタグを用いることもできるので、1つの要素の情報が長くても、入れ子となっても、問題がない。そのSGML/XMLを用いたEADによって、アーカイブズの電子的検索手段においても、大量の記述データを扱え、マルチレベル記述が可能になりました。

EADを使って資料目録を符号化すれば、図書館のOPACのようなことをアーカイブズについても実現でき、そのデータは比較的長期の保存が可能になるのです。つまり、目録データの共有と長期保存の効果が、期待できます。

2. アジア歴史資料センターの新情報システムにおけるEAD

アジア歴史資料センター（アジ歴）の新しい情報システムでは、アジ歴のサービスを支えるデータベース（情報の格納・提供を担う部分）に、EADやその考え方が使われています。その概要は別に紹介されていることでしょうかから、ここでは新情報システム運用開始時に残っていた課題にふれておきます。

アジ歴のデータベースは、国立公文書館デジタルアーカイブを参考にしています。そのことは、ブラウザへ表示される画面デザインが似ていることから、うかがえます。もちろんデータも、ともにEADの形式に則っています。ただ、これらのデータ構造は、それまでに使用されていた情報システムからデータを移行した影響なのか、EADを採用している多くの文書館・図書館等で使われているデータ構造と異なっています。具体的には、別掲の図をご参照ください。

端的に言えば、資料群全体の記述レベルの情報（資料群情報）を、<c>というタグへ収めてしまっていることに問題があります。すべての記述レベルのデータを一括で処理するには、1種類のタグを階層的に使うほうが便利だったからと想像します。しかし、EADを採用している多くの関係機関では、資料群情報を<archdesc>というタグへ格納しています。当時RLG（現在OCLCへ合併）の助言を受け公文書のアーカイブズの検索手段へEADの適用を成功させた、英国国立文書館でも、資料群情報だけは<archdesc>へ収めています。このままだと、目録データの長期保存の効果は得られても、関係機関との共有が不十分になってしまいます。

しかし、それほど深刻な問題でもない、ともいえます。やはり別掲の図に示したとおり、所蔵機関や資料群の情報を、<archdesc>下の適切な要素へ集約すればよいだ

けだからです。データを一括で処理する機能の修正も、極端に難しいものとはならないでしょう。あるいは既に、この問題は修正されているかもしれません。

このことを、あえて取り上げた理由は、日本のアーカイブズの目録データのあり方に、誤解があるかもしれない、と考えたからです。

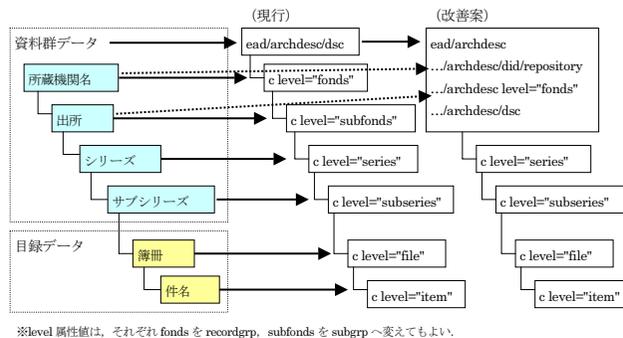
3. さらなるアーカイブズ情報の共有にむけて

目録データは、およそ(1)資料群情報、(2)資料群のなかの小群の記述の情報、(3)詳細な1点ないし1件ごとの内容目録の情報、の3つの部分で構成されます。EADを使ったアーカイブズの目録データで、どうしても必要な部分は、(1)の資料群情報です。しかし日本の資料目録では、(3)が重視されがちで、これまで(1)・(2)が軽視されてきたように思えます。(3)は、どうしても必要な部分ではありませんし、簡略でもかまわない。(2)も、分析がいたらず記述が難しければ、省略できます。ただし(1)は、簡略でもよいのですが、前項のとおり適切な位置に配される必要があります。

EAD データは、(1)だけ、(1)と(2)、(1)と(2)と(3)、あるいは(1)と(3)、の いずれかであれば、EAD を採用している多くの関係機関とアーカイブズ情報を共有できます。具体的には、横断的検索によって複数機関の収蔵資料から必要な資料をさがし出せるはずですが、しかも、ダブリンコア等による簡略な横断的検索ではなく、アーカイブズ情報だけの、アーカイブズの検索システムとして適切な、横断的検索です。

アジ歴は、すでに3機関の収蔵資料の目録データと画像データの横断的検索を実現してきました。アジア関係の歴史資料を収蔵する他の関係機関の目録データも EAD 形式であれば、それらも含めた横断的検索が、これからは可能になるかもしれません。EAD は、かならずアジア歴史資料センターの可能性をひろげてくれます。そして、日本の関係機関でのデファクト国際標準 EAD の採用が増えることを期待します。

図：目録データと EAD データの階層構造の対応



五島敏芳（ごとう はるよし）：学習院大学大学院人文科学研究科卒業。平成 16 年から人間文化研究機構国文学研究資料館助手。アジ歴新規システム導入に当たっては次期システム検討委員として EAD 採用に関する技術的アドバイスを行う。