

EAD at the PRO and the Access to Archives Programme (A2A)

Bill Stocking

Introduction

The Public Record Office (PRO: the National Archives of England and Wales) was an 'early implementer' of Encoded Archival Description or EAD as it is more commonly known. Meg Sweet, a member of the EAD Working Group, first introduced it in 1997. She developed a small pilot using the beta version of the new standard, which showed that EAD was applicable to English records and archival practice. It also persuaded her colleagues that further development with EAD was a good idea.

Since then, a number of projects at the PRO have used EAD and we have learnt a great deal about its advantages and disadvantages as a tool in the development of standardised and searchable online archival catalogues. This experience is being put to good use in an important national project, Access to Archives (more commonly known as A2A), which has EAD at its very heart. This project can be used, therefore, as an example of both the advantages of EAD, which led us to use it in the project, and its disadvantages, which we have attempted to overcome in the A2A database.

Access to Archives – A2A

The purpose of A2A is to take the paper catalogues of archives that the public uses in the searchrooms of archive offices throughout England, convert them to electronic form and display them on the internet in a searchable database. The converted data is also to be delivered back to each archive to use and enhance in their own systems. By March 2002 (the period for which we have initial funding), we aim to have the equivalent of 400,000 pages of archival catalogues, or to put another way, about 8 million descriptive archival records, on the Internet.

The problem

The nature of the project leaves us with a fundamental problem. Many different archivists, in many different archive offices have prepared the catalogues we are converting over many years. Moreover, most of them were prepared without any reference to consistent standards. There has never been an accepted national standard for archival description in the United Kingdom, for example, and international standards have only gained acceptance in recent years. How then were we to achieve our twin goals of a searchable database and re-use of data by local archives? The answer to our problem was of course standards. We needed to find a method to standardise the data as part of the process of conversion to electronic form.

Standards

The central standard for the project, that for archival description, is ISAD(G). To standardise names of people, families, corporate bodies and places for authority controlled indexing we have adopted our national standard, the UK National Council on Archive's *Rules for the Construction of Personal, Place and Corporate Names* (1997). These rules are compatible with the international standard for such authority records, ISARR(CPF). For subject indexing at a high conceptual level we are using the UNESCO thesaurus (Paris, 1995), which is now being used in a number of archival contexts in the UK. Finally, we wished to develop standardised metadata, based on Dublin Core for each of the catalogues.

Technical options

Having decided that we would standardise our data we also needed to look at the technical options for its storage and display. The initial thoughts were to utilise the investment we had in developing the PRO's own online catalogue, a custom built relational database known as PROCAT.

This system had been designed to conform to the standards noted above, but as the data was under the PRO's control we could enforce a common structure. This is not the case with the A2A project, as the data is inconsistent and ultimately owned by each separate archive. Also, the funding environment for the project is such that have not had the time to develop a relational database that would suit our needs. We needed to think again, so we looked at the option that had not been possible when we developed PROCAT. That is an EAD based document management system with a search engine.

EAD: advantages and disadvantages

From our experience of using EAD, we knew that in some ways it was ideal for our purposes. It has three particular advantages:

- One of the great strengths of EAD is that it is an open Standard Generalized Markup Language (SGML) Document Type Definition (DTD). Our catalogues encoded in EAD are, therefore, future-proofed as SGML is an ISO standard and we are not tied to the whims of any particular manufacturer. This gives as much protection for our data as is possible in an ever changing information technology environment.
- The fact that EAD already conforms to Extensible Markup Language (XML) is also an advantage as this is the up and coming web language. This also means that we can use the powerful transformation and display language developed for XML, which is Extensible Stylesheet Language (XSL).
- EAD is a display standard rather than an archival description standard in its own right, but it explicitly allows conformance to our standards. EAD can represent, for example, the archival hierarchy and data elements implicit in paper catalogues. Importantly for us it also allows the representation of the authority terms and metadata we wish to apply to each catalogue.

EAD then is the *lingua franca* at the heart of A2A. Offshore contractors rekey the paper catalogues into the EAD templates that we have developed. This enables the content in the catalogues to be standardised and captured in a consistent manner so that we are in a position to display and search the catalogues as a consistent whole.

It is at this stage then that we come across the two main disadvantages of EAD:

- EAD documents are not readily delivered over the Internet. Even current versions of the most popular web browsers (Microsoft's Internet Explorer and Netscape Navigator) do not support SGML files, users needing to buy extra software to read such files. Even XML is only now supported in the current version of Internet Explorer. EAD documents are also larger than comparable Hypertext Markup Language (HTML) documents and large catalogues can take a very long time, many minutes, to download. It is our experience that users will simply switch off and we will not be achieving the major aim of our project, which is enabling the public to get access via the catalogues to the archives.
- EAD does not of itself provide a way to search finding aids. One has to attach a search engine that can index and search across all the documents. Until recently, products that could hold SGML or XML documents and provide searching of them were relatively expensive and untried.

In both these areas, recent developments in XML have come to our aid. First, as noted, XSL stylesheets can readily transform EAD XML documents and present them as many different HTML documents. Secondly, there are now cost effective XML document managers with search engines that can process the volumes of data we have in A2A.

A2A Database – <http://www.a2a.pro.gov.uk/>

We have then used one of these systems, developed by Ixiasoft, for the A2A database. This holds the EAD catalogues, as returned from our keying contractors. Using the standards that have been captured in EAD it indexes them to provide free text searches that can be qualified by metadata, such as repository or region. There is also the option to search by the authority controlled names and subjects that have been captured by EAD. The results of searches are provided in a hit list and catalogues are then presented to the user in two ways representing different XSL views of the same data.

There is a 'context' view that quickly returns all the hits from a search, usually on a single page. The user can then choose to look at an entire catalogue, which may take longer to download.

The database is in its first version and we hope to further enhance it. We believe, however, that it represents a successful use of EAD and XML technologies that achieves our aim of increasing access to archives.