Eduardo Péis and Antonio Ruiz  
Department of Library and Information Science, University of Granada, Granada, Spain

Francisco José Muñoz Fernández  
European Documentation Center, University of Granada, Granada, Spain

Francisco de Alba Quiñones  
Documentation Center, Tourism and Sport Office, Andalucia Government, Spain

Practical Method to Code Archive Finding Aids in Internet

Abstract: Actually information services cannot work ably without web functionalities. Archives are paradigmatic examples of information systems with low use of technological possibilities. A small but hopeful step forward is the electronic encoding of descriptive information. One tool made suitable for it is a DTD - SGML for instruments of archive description named EAD (Encoded Archival Description). Having a base in an international norm like SGML, this tool, makes the information interchangeable. Possible. In this work Microsoft Active Server Pages encounter a concrete method for the automatic labeling of instruments of file description utilizing the displaced to one side programming technology of the server. Although the method develops in aggregate utilizing Microsoft technology, it can be utilized in another platforms, since the result generated is a document labeled in ASCII, to plain text, and therefore easily transferable to other platforms.

1. Introduction

If archives were adjusted to the new technological possibilities, they might be exemplary in making efficient use of resources and increasing the degree of use of the information. One key step on this road to adaptation is the design of structured finding aids that would facilitate web diffusion of archival information, as well as the re-utilization of this material in the form of varied informational objects.

To start with the process of adaptation, these functionalities can be developed by creating a “digital” version of the finding aids, encoding the structured descriptive information with a “mark-up language” that conforms to international standards.

Precisely an international standard like SGML (Standard Generalized Markup Language) is the best option for conceptually codifying structural units of documents, in the broadest conception of the word, for many reasons. First of all, it is a generalized public standard, independent of manufacturers and distributors, which ensures its permanence. Secondly, it does not impose a fixed set of components: the structures can be defined by the user, under the concept of
"document type." Therefore, it is remarkably flexible and consistent, adapting to whatever alphabetical system may have been used to write the text. And finally, it breaks down the structure of presentation, as it is a descriptive codification rather than a procedural one. For these reasons, more and more applications are using SGML as a basic tool (Cover).

It is easy to understand why the first attempt to apply a standardized code to instruments of archival description --initiated by the University of California at Berkeley and directed by Daniel Pitti-- selected SGML as the ideal technique for carrying out such a codification.

This project came to design a DTD (Document Type Definition) to describe a class of documents which consists of an optional title page, the description of a unit of archival material, and appendixes, which are also optional. The title page could include varied elements such as the identification of the repository or the type of descriptive tool. The descriptive component, in agreement with the DTD, would offer a brief description of the unit (using markable elements analogous to those employed in a MARC catalographic record), a broader narrative description of the unit and any of its separate parts (including markable elements such as title, dates, scope and contents) and a formatted list of the parts that contain that unit.

Logically at the present time the EAD DTD is usually developed in XML (eXtensible Markup Language). XML is a subset of SGML, but while SGML is mostly used for technical documentation and much less for other kinds of data, with XML it is exactly the opposite, being it more usable for distributing materials on the Web (Goldfarb and Prescod, 1998).

Therefore, as SGML, XML provides the rules for defining a markup language based on tags. It has been developed to keep up the proliferation of proprietary formats in use for electronic document processing and representation. It is a "descriptive" system that gives a declarative and machine-independent description of the document structure using codes that simply offer names to categorize and identify the parts of a document. This means that XML is a protocol devised to articulate structures of contents of documents instead of the appearance of documents.

The current EAD model (version 1.0) offers the option of using XML. It is possible to activate/de-activate the variant sections SGML/XML option using the SGML feature called "marked sections."

Despite the apparent advantages of applying this model, many archival systems can run into formidable difficulties in implementing the XML descriptive system. Our proposal consists of the development of a customizable template which can computerise the process of input of the descriptive information, convert and adapt this information to the EAD model and later handle the informational XML object with complementary technologies.

2. Methodology

To develop this method we need a base in an existent archive, the archives of "Patria\" newspaper, applying to the ISAD (G) standards. In this way, we can count on an accessible and close form of a archive description instrument, strengthened and made use of a real file, and we can apply the encoding methodology according to EAD directly, without the previous process of the construction of the description instrument.
The description comes true on three levels: funds, section and document. In regards to this communication, we focus on becoming centered in the document level, given the limitations in extension impede a complete analysis. In any case the methodology is the same on the distinct levels; only ISAD’s fields vary in that they get into stake encoding.

Example: we present a document's concrete description, in a photo, following the structure ISAD (G):

1. MENTION AREA OF IDENTITY
   1.1. Referential code: ACT I/C3
   1.2. Title: La casa de Ángel Ganivet
   1.3. Dates: 11 enero 1979 (fecha de publicación)
   1.4. Description level: Unidad documental
   1.5. Characteristic day girls: 13 x 18,5 cm, positivo, papel, B/N

2. CONTEXT AREA
   2.1. producers/autor name: Ferrer

3. CONTENTS And STRUCTURE area
   3.1. Catch up with and contents
      Photo of press. Casa-Molino de Ángel Ganivet, on the Cuesta de los Molinos (Granada)
      Photo caption: "Uno de los rincones granadinos de mayor sabor romántico es este de
      la Cuesta de los Molinos y se centra de manera especial en el rincón que crea la casa
      de Ángel Ganivet. Ahora se ha levantado la voz de alarma porque al parecer la Casa
      del Molino de la Zagra, en la que está la lápida con el busto del escritor y diplomático
      granadino, está en situación ruinosa y cualquier día se nos puede venir abajo a todos.
      Porque la casita debería conservarse como reliquia que nos aproxima a la formidable
      figura del autor de "Granada la bella"

4. DOCUMENTATION ASSOCIATED AREA
   4.1.1. Publication note
   5. Fotografía publicada en el Diario PATRIA el 11 de enero de 1979, p.20 (contraportada)

6. NOTES AREA
   6.1.1. Note: Se ha tomado como título el que acompaña a la fotografía
   publicada.

KEYWORDS / INDEXING TERMS
   URBANISMO, INDUSTRIAS, CASAS-MOLINO, GANIVET, ÁNGEL,
   CUESTA DE LOS MOLINOS, GRANADA.

Anyone can notice that there is information that is not described, but it was done in this way because this information appears in finding aids on superior levels. Granted that this description instrument is already in use, we have a record data base with descriptions of distinct documents. We for applying automatic encoding to this data base. Thus, our objective will be to have a data base, and to encode it with EAD in time of to take it to Internet and as about this business in automatic form.

There can be several forms, but the method begins with a same starting point: The equivalency among the ISAD fields and the EAD labels. Obviously, it would be possible to construct independent labels from XML, but then we would not get this encoding's other objective: normalization. In this way, we would begin by constructing a document

<?xml version="1.0" encoding="ISO-8859-1"?>
structure starting from the labels EAD, encoding ISAD fields. In short it will be only an equivalency definition among ISAD fields and EAD labels. Authentic encoding would be made upon in Internet retrieval. For example, our document would get coded in recovery it as follows:

To accomplish the automatic codification process, we used the Active Server Pages technology of Microsoft, a programming language of the server's side geared for Internet Information Server and Personal Web Server. The reason for selecting this option is its implementation facility in Windows systems, as well as the accessibility of most of the archives without having to resort to more complicated systems.

Thus, departing from the base of relational data base where the records are stored with the descriptions ISAD (this data base can be Access, SQL Server or anyother accessible by means of ODBC), encoding would be effected. Then several alternatives can be presented. A first would be to connect the data base with a Web fill-out form, to consult the data base with sql entences. Thus, in our project we created a fill-out form to consult the data base, and the results will format themselves using xml, that is, the file will be in asp language but the contents would be completely xml; with the distinct EAD labels to encode contents. Asp language permits this possibility, though right now an asp file is no more than a file with format labels and code labels that must process the server. This way we can include labels as xml without problem: the processing instruction of document xml are given and the patch of IE5s shows it on the screen correctly. In this way, the automatic encoding of the description instrument for the distinct records is effected on the one hand, and in addition, the same ones are shown in an easily exportable format, in plain text, totally compatible text with anyother system that follows EAD DTD without errors. The xml and css style sheets combination and the use of name spaces allow us the appropriate data presentation in the usual format of the current
web sites. This method avoids the IE5 default presentation in form of pull-down nodes. Also possible is the use of introduction data technology to include the XML code with the search results among the HTML code. This second option improves on the previous one in that the XML code is clearly defined, facilitating the exchange between systems. The XML code is not mixed with labels of different name spaces and the identification of the EAD content is easier.

A second alternative is the use of a complete database in one XML file. The first step is coding the complete database in XML files following the label EAD list. This is an automatic process that we did using the same previous method, but now without a form to select records because we code the whole database. In this case, the selection of records is carried out using XSL, with which we can control the localization of the documents EAD nodes, as well as place conditions for data output. To have the database with EAD labels and in plain text format allows use of the DOM technology to edit and change the EAD document. With this method, we can control from one page web different types of data, applying different XSL templates to our EAD label database. Here more programming is needed to show the EAD records in the web, but we have more compatibility between different systems.

3. Conclusions

It is absolutely necessary a normalize in the handling of archive information to face to the new challenges that are presented by diffusion of information in Internet.

The EAD DTD is a XML document type definition that results very appropriate to encode archive finding aids according to the ISAD standard.

It is possible to devise a customizable input template which allows automatic EAD encoding of archivist finding aids.

The use of server programming technologies like active server pages allow the code of databases made according to the ISAD standard. These technologies allow data output in a EAD DTD labeled format, guaranteeing a higher compatibility between systems.

The trend to EAD document database models is a way of assuring the compatibility of the current information in the new systems, and can be a more effective form of freeing data of applications, avoiding software obsolescence derived problems.

We should consider that the application of the EAD is more transcendent than the application of the ISAD (G) standards, since the electronic document is today a reality in the archive.

References


Investigation project for the recovery of the "Patria" newspaper photographic archive carried out by Dña. Antonia Salvador directed by Dr. Antonio Ángel Ruiz.


Berkeley University EAD web page http://sunsite.berkeley.edu (visited April 16, 2002)