

## **Teetering on the Cutting Edge: Exploiting the Potential of Encoded Archival Description (EAD)**

**Elaine Bradtke**

*An attempt to explain our unconventional use of EAD to gain intellectual control over the contents of the James Madison Carpenter Collection. Originally presented at the American Folklore Society's Annual Meeting, Rochester, New York, October 16-20, 2002. Updated October 13 2008.*

### **Not what the designers of Encoded Archival Description had in mind.**

The International Council on Archives defined Archival Description as:

The creation of an accurate representation of a unit of description and its component parts, if any, by capturing, analyzing, organizing and recording information that serves to identify, manage, locate and explain archival materials and the context and records systems which produced it.<sup>1</sup>

Encoded archival description then, is doing all this, in a machine readable format. The emphasis is on representation, identification, and context. However, EAD, like most archival standards, was created with collections of historical manuscripts and public records in mind. What we have in the Carpenter collection is a complex, multimedia, ethnographic field collection. These two types of collections have different origins, functions, and uses, resulting in different indexing requirements. To accurately represent, identify, and illustrate the context of a collection that deviates from the model upon which EAD is built, we have had to deviate slightly in the application of EAD.

Historical manuscripts and public record collections originate as the accumulated papers of a person or corporate body . The material functions as a record of the corporate body or person and their work and content is narrowly

focused on this work. Because of the nature of this material, it often comes to an archive already arranged and labeled in accordance with its original use and function. This makes it fairly straightforward to describe the collection with a finding aid in a way that represents its context and structure. The inherent chronological or subject arrangement imposed upon the material by its creator often eliminates the need for detailed item identification. Users of this type of collection are less likely to be interested in extracting small segments of the collection to compare with each other or similar items outside the collection.

Multimedia ethnographic field collections originate from a collector actively seeking out and gathering material. One afternoon of fieldwork may result in sound recordings, field notes, photographs, drawings, printed documents, etc. all related to each other, yet filed separately. Because of the different media involved, the material may not be arranged by subject, or even chronologically, but rather by physical characteristics. This results in several different series of data with convoluted interrelationships.

Carpenter's collection includes, in addition to his fieldwork, correspondence and drafts of essays, material copied from published sources, items collected or created by others and material he himself created later in life. The subjects he covered were wide ranging: songs, mummer's plays and photographs of ancient buildings for instance. Items in all formats also exist in several generations; drafts of essays, rough and fair copies of song texts, duplicated recordings, multiple versions of the same image and so on. A large portion of the material is undated, and whatever original chronological order there may have been, has been disrupted over the years. Simple and straightforward are not words that come to mind when contemplating the arrangement of the Carpenter collection.

An ethnographic field collection functions not merely as a record of the collector's work, it forms a basis for research and analysis, and often provides

source material for publications. The use of this type of collection tends to be concerned with specific facets of the material, their sources and contexts. Rarely is the focus on the person who assembled the collection. This is a different type of use than might be put to say the papers of Richard Nixon, or the Royal Society.

For our purposes we can say that ethnographic field collections differ from historical manuscript and public record collections on two important points. **1)** Due to their unique origins and specialized function as raw materials, assembled for research and publication, they may be less coherent and cover a broader range of subjects and formats, and **2)** the typical end user requires a deeper and more precise level of access to the information contained within them. Thus, applying EAD to a multimedia ethnographic field collection will involve teetering on the cutting edge, as we perform a balancing act between existing standards and the requirements of the users.

#### **Overview of Finding aids vs databases:**

A database can be defined as a body of data organized especially for rapid search and retrieval. Information is entered into discreet records pooled together to be fished out as required using various search terms for hooks. A finding aid is more like a roadmap. It provides the means of navigating through the collection. Finding aids often have a title page, and an introductory section followed by a description of the contents that represents its physical arrangement. In large collections this may mean boxes that contain folders that contain papers. Like nesting Russian dolls, the information being sought may be concealed inside several layers of containers. EAD allows a researcher to electronically dig through the strata of the collection to extract buried nuggets of information.

The James Madison Carpenter Collection project team has always had the functionality of a database as our ideal. Because of our experience as researchers, we knew that it was important to elaborate on the fact that these folders contain

Scottish songs and those have mummers plays. Identifying exactly what was in a given folder, who performed it, where and when, and making it easy to zero in on a specific item was what was needed. A great deal of thought was given to how the collection would be searched and how data might be drawn together. We also felt the need for additional access points such as Child numbers for the ballads, Ordinance Survey references for geographic locations, and genre terms. It was a blessing in disguise that none of us were experts on EAD, or it might have stifled our creativity.

We have worked to combine the retrievability of a database within the structure of a finding aid. This in itself is not completely unheard of. What made our project different from other uses of EAD was the level to which we wanted to go. Because of the physical organization problems inherent in the collection and because information about a particular collecting event might be contained in several formats (recordings, photographs, music notation, field notes, even letters) we felt the need to catalogue every intellectual item. We defined intellectual item as a discrete piece of information - it need not be complete, nor independent and may be contained within a larger body of information. For example: a mummer's play that includes the play text, a song (with music notation) and the description of a dance. The parent item would be the play as a whole, and nested within it are the song, and dance description. In database terms we were creating records for each of these items. In finding aid terms, we were going into much more detail than a folder or even page level description.

If it's not really a database, then how is the data retrieved? Through the magic of tagging. Tags are used in EAD to mark pieces of information for further manipulation by a style sheet (for display) or a search engine (for retrieval). Instead of set fields in a database EAD uses tags that may be applied where needed. Personal names and place names may be found near the title, in biographical and historical information, or in scope and content notes and are tagged as appropriate.

Attributes are a useful aspect of tagging. A name can be designated as that of a photographer, or the subject of a photograph, a letter's author or its recipient and so on. One attribute we developed was 'secondary source'. This was used for phrases such as "learned from Mary Jones." The hope being that we could eventually link these people either within the collection or in other collections.

### **Why use EAD for such detailed indexing?**

- 1) EAD is the standard and a good basic foundation from which to work.
- 2) XML, on which it is based, is infinitely flexible, works on all computer operating systems and is non-proprietary. XML files are easily converted to web-friendly data.
- 3) EAD has the potential for linking to external items. Our goal is for the end user to connect directly from the search results to a representation of the item - similar to the American Memory Project's online collections at the Library of Congress (<http://memory.loc.gov/ammem/index.html>).

EAD allows for the inclusion of explanatory front matter on the history of the collection, and the rationale behind the way it is represented. It also permits identification of intellectual items while reflecting the context of the actual material. Merely creating a database record for the song (within the mummer's play) does not illustrate that this is a significant part of a whole performance. Using a hierarchical configuration shows clearly the relationship between the parent intellectual item, the child and its siblings. Finally, in EAD, the original arrangement of the collection dictates the arrangement of the document. Carpenter's assortment of boxes, packets and mail sacks is represented in the digital finding aid.

### **Exploiting the potential of EAD**

The EAD has certain elements that either suited our needs exactly, or were adapted to fit (all quotes are from the EAD tag library).<sup>2</sup>

#### ***Unit title:***

The unittitle is the heart of the matter, where information that identifies the intellectual item is placed. We envisioned a display of search results based on information contained in the unittitle that would allow the user to decide if an item was what they wanted. Our complex unittitle is flexible and comprehensive enough to accommodate the wide variety of material in the collection and to give sufficient, pertinent detail.

*Title:*

The title of the item (when present) is transcribed as given by Carpenter, idiosyncratic spellings and all. When there was no title, we used the first line of text (or incipit). Photographs don't often have titles, so a description was used. These title analogues are given attributes, so that the user understands what they are looking at, and the style sheet will display it appropriately

*Genreform:*

"A term that identifies the types of material being described, by naming the style or technique of their intellectual content (genre); order of information or object function (form); and physical characteristics." Genre terms appear following the title to show what type of item it is. Both physical format and content information appear in this list; song text and sound recording for instance. This is unavoidable if we are to show whether this instance of Barbara Allen was a recording, a song text, or music notation. We had the option of using more than one term here, but it was rarely needed.

*Persname:*

A personal name is defined as "including any or all of that individual's forenames, surnames, honorific titles, and added names". It is standard practice to create an authority list of names to support the use of a standard form of name throughout the collection. Although we do have an authority file for names, we chose to enter the names as given, and use the so called normal attribute and the correct

form of the name as a searchable, but hidden term. Each person is also given a code that will link all the items related to them, regardless of the form of name used. Attributes were used to designate the person's function in relation to the material.

*Geogname:*

Place names are very important in ethnographic field collections, especially in a collection geographically diverse as this one. Place names were entered and tagged as Carpenter gave them, even though they may be incorrect. Complications with Carpenter's place names are handled with a geographic authority file outside the EAD. This is constructed hierarchically from the country level down to the address. Whenever possible, Ordinance Survey references are given to pinpoint exact locations. In a similar fashion to the personal names, place names are given a code to link them together across the collection.

*Unitdate:*

"The creation year, month, or day of the described materials." Defined in our case as the date of collection. This is a rarely used element as Carpenter was not careful about noting the date. Because of the potential confusion between UK and US forms of the date we used Day, Month, and Year (in that order) and inserted the ISO form as a 'normal' attribute to allow for searching and sorting.

*Bibref:*

Items copied from published sources are included in bibref: "A reference element that provides a citation and/or electronic link for a published work". In this case it is a short citation including name, title and date. These references are drawn together into a formal bibliography with full publication information.

***Control Access:***

"The <controlaccess> element is a wrapper element that groups key access points for the described materials and enables authority-controlled searching across finding aids on a computer network, serving a function similar to the added entries

and subject headings in a catalogue record”. We have not used control access for subject headings or genre terms, but for something more specific to the material at hand - song classification. It was evident that Carpenter knew about and relied upon Francis James Child’s system of numbers and uniform titles for ballads. Therefore we included them as an aid to bringing together related songs. We’re probably the first users of EAD to use Child numbers in <controlaccess> but the hope is that others may follow suit. We have added Steve Roud’s more comprehensive system of numbers in the same way for many of the sound recordings, but as yet this work is incomplete.

***Biographical and Historical information:***

The tag library defines bioghist as: “A concise essay or chronology that places the archival materials in context by providing information about their creator(s).” Carpenter’s information on where the contributor acquired the item was deemed to be an example of provenance and formed the basis of the biographical historical section.

***Scope & Content:***

“A prose statement summarizing the range and topical coverage of the described materials. . .The purpose of the <scopecontent> element is to assist readers in evaluating the potential relevance of the materials to their research.” This includes information such as the extent or duration of an item, alternate titles, and information about related items.

***Genreforms:***

We felt a need for additional access points to enable users to perform complex searches. The genre terms in this area were intended to be searchable, descriptive, and include both general and specific terms to facilitate boolean searches. The images in the collection have further search terms so they may be retrieved by the type of performance, custom, building etc. they portray. The terms



for images are taken from the Thesaurus for Graphic Materials and the Proto-Thesaurus for Ethnographic Materials with a few terms added to fill the gaps. Strictly speaking, all these terms might be better located in control access. Because they related directly to the scope and content of the material, and they were not from a single, published, standard list, we put them here.

*Title:*

“The formal name of a work, such as a monograph, serial, or painting”

We’ve included not only the titles that Carpenter crossed out (as alternative titles) but the first line of a song, and the incipit of an essay or other piece of text. Although stretching the definition above, these title analogues give content information and help the user verify that they have the correct page (important when there are many items on the same page, or the page sequence has been disrupted).

*Extent:*

Because we were originally working from digital surrogates, we could not give an accurate physical description of the original, so it was omitted. The format of the item can be gleaned from the genreform in Unit title. Given that some intellectual items were short excerpts, we included information on extent, usually in the form of the number of lines or pages. Some of the photographs have been examined in person and the dimensions of these are indicated.

*Notes:*

“A generic element that provides a short statement explaining the text, indicating the basis for an assertion, or citing the source of a quotation”. This is where relationships between items in the collection are explained, including the sequence of pages in longer items. Some of the pages were shuffled and have subsequently been microfilmed and digitized in this order, so page numbers may not represent the logical order of an intellectual item.

A particularly unusual interpretation of EAD on our part was the inclusion of character and actor lists from plays in scope and content notes. The characters and actors are tagged as personal names, and given the attribute character or actor (which gives us the option to display and or search them separately from other types of persnames).

### **Conclusion**

These are just some of the tactics we used to represent, identify and put into context the many facets of the Carpenter collection with the help of EAD. While the project is not yet completed, early indicators point to a successful result. The final proof will occur when it is connected to the digital surrogates and thus fully functioning as intended. We have shown that with some modifications, it is possible to use Encoded Archival Description for complex collections such as James Madison Carpenter's. Our work is not meant to be the last word on the subject of multimedia ethnographic collections and EAD. Rather, these are only the first tentative steps, and it is hoped that discussion, evolution, and widespread implementation of this powerful tool will follow.

- 1 ISAD(G): General International Standard Archival Description, Second edition, p. 10 [<http://www.ica.org/en/node/30000> accessed 13 October, 2008].
- 2 EAD Tag Library 2002 [<http://www.loc.gov/ead/tglib/index.html> accessed 13 October 2008].