Moving Beyond APPM to a Global Standard:
Using DACS for Cataloging and Finding Aids

Susan Hamburger, Ph.D.
Manuscripts Cataloging Librarian
The Pennsylvania State University
MARAC Fall Conference
October 28, 2006

[SLIDE 1] I am not a religious person but I believe in two things: Bruce Springsteen and cataloging standards. Now I'm not here to convince you to join the Church of Bruce even though we're in the great state of New Jersey and what would be more appropriate; instead I want to exhort you to become followers of my other passion—standards for creating archival catalog records and finding aids.

At Penn State we are working on a way to export finding aids from an Access database into an EAD document that will be ingested into DLXS, CONTENTdm, or another platform for federated, cross-collection search and retrieval. Standards, of course, are a big component of that and we are using *DACS* (*Describing Archives: A Content Standard*) and the EAD (Encoded Archival Description) tag library to ensure that the fields are populated with the appropriate information that comes out tagged correctly.

My work with MARC cataloging and EAD predates *DACS* and it is heartening to see that DACS reinforces almost all of the practices I've been doing for the last sixteen years when I veered off the *APPM* (*Archives, Personal Papers and Manuscripts*) path slightly (particularly in how to name a collection) because of OPAC displays. **[SLIDE 2]** When *DACS* came out and I analyzed the differences between *APPM* and *DACS*, I only had to do a few tweaks to my practices to comply (such as spelling out cubic foot instead of abbreviating cu. ft., for example). I also like that *DACS* is platform-neutral and can be applied to MARC as well as EAD.

Today I want to focus on the switch my institution made from using *APPM* to *DACS*, due in part to developments in the technology (first OPACs and then EAD for finding aid presentation). I will discuss the determining factors in using *DACS*, and relate how easy—or hard—it was to implement those changes. Changes in the technology available to the profession make it necessary to at least re-evaluate and restructure our standards and will provide an impetus to get more institutions to standardize using the tools at hand.

I first began using the first edition of *APPM* when I worked in Virginia in the late 1980s. At the Virginia State Library and Archives (now the Library of Virginia), we were part of the RLG Government Records Project (1989-1991) funded by NHPRC to find commonalities among state government records and to create catalog records and agency history records (an early attempt at corporate name authority that is morphing into EAC—Encoded Archival Context) in RLIN (Research Libraries Information Network). Before we'd finished that grant project, I engaged us in an NEH grant (1991-1992) to catalog selected personal papers. It was during the first project that Steve Hensen published the second edition of *APPM* which had a profound impact on our cataloging.

We followed *APPM*'s rules but soon found that the one on how to construct a title for the collection brought us endless amounts of problems in both RLIN and our OPAC, VTLS. The convention of calling a collection, simply "Papers" or "Records," or "Papers of Thomas Jefferson" or "Records of the Virginia Auditor General," failed us when we wanted to search either database. **[SLIDE 3]** The display of titles either returned a long list of collections called "Papers" without displaying the creator or truncated the titles to "Papers of Thom" thus forcing us or a researcher to look at each record to ascertain the

full title. When I moved to the University of Virginia in 1992, we had this same display problem with the NOTIS software in the OPAC. **[SLIDE 4]** But I had already fixed the problem by defying *APPM* and naming all collections with the creator's name first in the title, for example, "Thomas Jefferson letters to James Madison" rather than simply "Letters to James Madison." **[SLIDE 5]** I knew that many OPACs displayed their title lists differently and I wanted to be sure the researcher, whether looking in RLIN, OCLC, or the local OPAC, would find the collections with the least hassle. **[SLIDE 6]** How was I to know that thirteen years later the *DACS* folks would come around to my way of thinking? **[SLIDE 7] [SLIDE 8] [SLIDE 9]**

As a hybrid librarian/archivist, I was originally trained to catalog books and in this world abbreviations run rampant. This is a legacy of typed catalog cards where the least amount of information that a cataloger had to type, the more efficient. **[SLIDE 10]** Thus we ended up with. n.p., n.d. (later changed to s.l., s.d.), p., ed., ca., and the archival catalogers joined forces by supplying cu. ft. *DACS* says to spell out all of these abbreviations so we now say undated, page(s), editor, and circa. **[SLIDE 11]** Another disconcerting change for a growing collection with accretions is to abandon the use of 1979-(ongoing) in favor of closing the dates for the materials in hand and continually updating the catalog record (everywhere, whether RLIN, OCLC, or local OPAC) and finding aid. This last rule is annoying because, at Penn State, we tapeload to RLIN and can't change the record, and the changes to our OCLC and local OPAC records not only create extra maintenance work but no longer match the holdings shown in RLIN.

Incorporating the changes from *APPM* to *DACS* was simply a matter of reading the new standards and writing a page-and-a-half field-by-field explanation of—what for

us were minor—differences. On September 2, 2004, I called a meeting of the Special

Collections Cataloging Team consisting of another faculty librarian cataloger for rare

books, one full-time staff original cataloger, one half-time staff original cataloger—all of

whom assist me in cataloging manuscript and archival collections—and one-half-time

staff copy cataloger (who was excused as not relevant to her work). I gave them a copy

of the differences, explained them, and we moved on to implementing *DACS* that

afternoon. Remembering to spell everything out took a little time to embed in our

routine, but since I proofread all the manuscript records the other team members create, I

was able to catch and correct the slip-ups. They are a marvelous group—willing and able

to switch gears between book cataloging with its abbreviations and manuscripts

cataloging with everything spelled out. One of them also catalogs maps, and I taught

both staff original catalogers how to catalog architectural drawings and graphic materials.

So from the cataloging standpoint implementing *DACS* has been easy and painless.

For EAD finding aids, on the other hand, the staff and procedures are different.

The Special Collections Library formed in 1999 from three separate units—Rare Books

Room, Historical Collections and Labor Archives, and Penn State Room—with different

administrative heads, foci, staffing, hours, locations, and ways of doing things. When I

was hired in 1994 to process literary collections and catalog for all three units, I was the

only person who crossed the boundaries. Starting with the catalog records, and later

EAD, I brought standardization to description without the authority to dictate the rules. In

preparation for becoming one library in one physical space, we began planning the

merger of everything from call slips to stack naming conventions. The staff members,

not the librarians or archivists, were interested and active in creating databases to create

box lists for their collections. Unfortunately, because we were all in separate locations

and the librarians and archivists were not consulted about designing the databases, each

unit's database was different. Each used some of the same named fields but put different

data into them. And there were no rules for data entry. Such simple decisions as putting

the creator's last name then first name in a separate field were not made. **[SLIDE 12]**

Dates, if entered at all, went in in text, numbers, separated by dashes, slashes, or not at all

(19980622, for example). **[SLIDE 13]** You have to realize that we were working first

with a Q&A database, and then migrated that to Microsoft Access. Moving into the

combined library space proved an opportune time to reevaluate the databases (by this

time there were five—the three units plus the Fred Waring's America archives database,

and a separate accessions database not linked to any of the collections' databases).

This time, determined to bring standards into play, I became a member of the

Database Redesign Committee and argued for separate fields that could be mapped to

EAD. Over the course of almost two years, we pulled apart the fields in each of the

databases and designed a new integrated one still in Microsoft Access that stepped up to a

more robust SQL server. The new version contained the limitations of the old ones—

plenty of room for box and folder lists but no room for narrative biographical/agency

history notes or scope and content notes. As a stopgap, we labeled the series, box, and

folder entries with EAD tags and generated a marked-up document that I then married to

the textual parts of the EAD document I'd created in XMetaL to generate a complete

EAD finding aid. One problem, though, was finding the biographical, scope and content,

and arrangement notes for these box lists. Some of the older collections had typed paper

finding aids, others had floppy disks in Word. Many were written so that all of these

separate elements were intermingled. Many also lacked the date span of the collection and/or total cubic feet. And often the output wasn't in the proper order and I had to cut-and-paste for hours to get back to Box 1 Folder 1, Box 1 Folder 2, Box 1 Folder 3, rather than the system generated Box 1 Folder 1, Box 2 Folder 1, Box 3 Folder 1. There had to be a better way.

Because EAD was written to accommodate all the varieties and flavors of finding aids, its very fluidity presented problems for institutions deciding what tags to use and in what order. Luckily, Penn State sent me to Rare Book School twice to take Daniel Pitti's week-long courses, "Implementing Encoded Archival Description" in 1997 and "Publishing EAD Finding Aids" in 2003. Here we learned best practices and I've used Daniel's teachings as my model template for finding aids ever since 1998. Prior to the publication of *DACS*, there were no rules for required elements or the content of them. While I happily chugged along hand-encoding or copying and pasting bits of finding aids into my EAD template and cranking out maybe twelve a year along with my other responsibilities, we soon realized that an improved database could provide the solution.

The Access database on the SQL server became overwhelmed by the amount of data and periodically and frequently ground to a halt. With the help of a programmer in Information Technologies, we migrated from SQL to Cold Fusion on an Oracle platform. Finally we had text fields available and started populating them with the biographical and scope and content notes. In 2005 we hired a Processing Coordinator to finally oversee standardization of input and she—with input from two others of us on the Processing Committee—wrote our first processing manual that spelled out not only how to label a folder but most importantly implemented *DACS* in populating the database. **[SLIDE 14]**

During the summer and fall of 2006 she and I worked with the programmer to go over the database code line by line to insert EAD and MARC tags to generate a report that will be a fully-formed, validatable EAD document. **[SLIDE 15]** We're including a screen of error messages that will tell me what data is lacking in a particular field before it creates the EAD. **[SLIDE 16]** I then will only have to transform these XML documents into HTML and PDFs for printing and post them on our Web site. **[SLIDE 17]** We're also investigating and evaluating federated searching software to make searching across finding aids available to researchers.

What does all this have to do with global standards? Plenty. Our finding aids now are being harvested by RLG's ArchiveGrid to make them findable and searchable along with thousands of other finding aids. But what good does that do a researcher if we don't use standards in creating our finding aids? All of the standards that I use in my catalog records I also use in my finding aids: the Library of Congress Name Authority File for formulating personal and corporate names—creating and contributing new ones that don't exist already; the Library of Congress Subject Headings for topical and geographical terms; and the Art and Architecture Thesaurus for form/genre terms. **[SLIDE 18]** All of these provide the researcher with one choice of form of name, subject, or genre to search rather than having to guess whether to search for Jeb Stuart, J.E.B. Stuart, or James Ewell Brown Stuart—or all of these. **[SLIDE 19] [SLIDE 20]** **[SLIDE 21] [SLIDE 22]** If we order our finding aids in a logical fashion, include the required tags, and write the notes (using keywords that a researcher may look for that are not in the subject headings) in clear, concise, and appropriately detailed prose, then

researchers don't have to guess how we're going to present our finding aids differently from the repository across town or around the world.

Speaking about around the world, one of the new required EAD tags is Language of Material to explicitly state the language(s) and scripts of the materials in the collection, as well as a tag for the language of the finding aid. Our compatriots outside the United States are embracing and implementing EAD and EAC (Encoded Archival Context) for personal and corporate names. For compatibility's sake, we would be wise to adopt standards so all of our finding aids can play together in the global sandbox.

Why *DACS* and EAD and not HTML? **[SLIDE 23]** As browsers become more sophisticated they will soon catch up with XML and be able to find the IDs within the finding aid tags for <scopecontent>, for example, and search only that tag. HTML only tells you that this word is in 24-point Arial and it is blue, not that it is a Scope and Content Note. HTML is the presentation not the content. Whether you choose to go online with your finding aids or continue to produce them only on paper, following a standard not only helps the researchers but gets you and your staff in the mindset of making sure you have not only all of the required information about the collection, but that it is in the correct fields. If you follow the standards, then once you do decide to move your finding aids into an electronic environment, you won't need to spend months marking up stray sentences within paragraphs for the encoder, nor spend precious time proofreading the final output. An excellent book to read is Elizabeth H. Dow's *Creating EAD-Compatible Finding Guides on Paper* (Scarecrow Press, 2005). Wiz assuages all the panic you might feel about getting started with EAD. **[SLIDE 24]**

Now that the archival profession has matured and developed content standards for finding aids, I urge everyone here who has not already joined the parade to step up even if you can only play the triangle. **[SLIDE 25]** With practice you'll soon move up to the drum and bugle and we can march off into the archival world of standards with everyone playing the same song in the same key at the same tempo.