SGML→XML →HTML for Displaying EAD Finding Aids

Susan Hamburger, Ph.D.
Penn State University
Forum on Digital Library Practices
July 17, 1999
Washington, D.C.

Penn State is participating in the Making of America II Testbed Project. I signed up for Daniel Pitti's first Rare Book School course on EAD to learn how to implement EAD before the grant project was even proposed. Between the time I completed the course and when the library bought the editing and publishing software, EAD went from beta to version one. In that interim I also forgot a lot of the crucial intricacies of where to put all the files in the right pathways. Technical support has been very minimal since no one else knows EAD.

Part of the delay in choosing software is how the finding aids would be delivered on the Web. Penn State has 24 campuses across the Commonwealth plus branch libraries at the main campus in University Park, several reference points within the main library, and faculty offices, computer labs, dormitories, and at home connections—any of which may be an entry point for someone doing research in archival and manuscript collections. And then there is the issue of remote access by non-Penn State patrons.
Delivery: Interleaf’s Panorama

- PRO
  - Searchable
  - Customizable display

- CON
  - Proprietary software
  - No longer offers downloadable free viewer for remote patrons
  - Not networkable for distributed campus system

Because I had learned on Softquad’s Author/Editor and Panorama Publisher—and the price was a modest investment that I felt the library would be willing to make—I recommended we purchase the Panorama Publishing Suite.

At the time of purchase, Softquad provided a free viewer for anyone to download to see the finding aids. Unfortunately, they didn’t offer a networkable version and I was reluctant to go the Yale route and buy site licenses for hundreds of computers across our statewide system. The Softquad threw a monkeywrench in the works by selling Panorama to Interleaf which rescinded the free viewer download.
Delivery: HTML

■ PRO
  - Anyone can view
  - Easy to create
  - No special software needed

■ CON
  - Lose navigation
  - Lose sophisticated searchability
  - Displays linear not hierarchical

Our remote patrons have all levels of browsers. By now almost everyone has migrated to a windows-based environment yet the difference versions of the browsers read--or don’t read--present choice of delivery problems.
Delivery: XML

- PRO
  - Cutting-edge technology
  - Compatible with EAD dtd

- CON
  - Most browsers, except Internet Explorer 5.0, cannot read it

If we choose to ignore the people with low-end browsers and expect them to catch up with us, we can opt for XML. But currently only Internet Explorer 5.0 can read XML documents. And, from what a technical support person told me this week, IE and Netscape don’t get along when mounted on the same server.
Solution: compromise

- Create finding aid in SGML
- Map SGML to XML
- Export XML files to HTML
  - PRO
    - Mimics Panorama display with frames and navigation
  - CON
    - Lose customized font color
    - When you choose one, all subheads in navigator frame display

Although I was frustrated for months when I couldn’t get the files in the right places and the viewer couldn’t find my EAD finding aids, I can look back on it now as a fortuitous delay. Rather than creating a slew of finding aids in beta version, I only had three. When version 1.0 was released, I only needed to convert these instead of having a big conversion project or running two parallel sites.

Because of the lack of knowledge of EAD on campus, I requested that Daniel Pitti come here as a consultant to help me fix my problems. In addition to converting my three beta finding aids to version 1.0, he came up with a solution to my delivery dilemma: continue to create the SGML files but map them to XML and export to HTML for display.
Solution: alpha test conversion program

- Daniel Pitti wrote a script to convert SGML files into XML
  - SGML files must contain attributes XML can read:
    
    ```xml
    <HEAD ID="desc">Container List</HEAD>
    <C01 LEVEL="series">
    <HEAD ID="s1">Photographs Series</HEAD>
    <DID><UNITTITLE>Photographs, <UNITDATE>1891-1915</UNITDATE></UNITTITLE></DID>
    
    - XML files contain java scripts
    ```
XML to HTML

- XML to HTML script
  - Creates five files to display one SGML finding aid
    - [filename].html = frames
    - [filename1].html = navigation frame for Table of Contents
    - [filename2].html = headers in body linked from Table of Contents
    - [filename3].html = headers in body linked to text
    - [filename4].html = text of finding aid
How it works at Penn State

- Create finding aid using Author/Editor
- Use EditPad for parsing
- Save finding aids being worked on in
  - working/findaids/inprocess/ae
  - working/findaids/inprocess (for SGML version)
How it works at Penn State

- When finding aid done, export to XML
  - Parse in NSGMLS: \texttt{pwork} [filename]
  - Save to XML: \texttt{sx} [filename]
  - In XMLIN folder open [filename].xml and \texttt{system.txt} in EditPad
    - Replace \texttt{<DOCTYPE EAD SYSTEM...} with string in \texttt{system.txt} : \texttt{<DOCTYPE EAD SYSTEM ".../shared/ead/eadx.dtd" [}
    - Change \texttt{entityref=psuseal} to \texttt{href="psuseal.gif"}
    - Remove extraneous character Æ at copyright symbol
How it works at Penn State

- In Parse folder run xp1
  - `<xml version="1.0" encoding="ISO-8859-1">`
  - `<!DOCTYPE xsl:stylesheet>`
  - `<xsl:stylesheet
    xmlns:xsl=http://www.w3.org/TR/WD-xsl`
    - `xmlns:html=http://www.w3.org/TR/REC-html40/>`
    - `results-ns="html">`
    - `xsl:define-constant name='name'`
    - `value='nassh'/>`
    - `</xsl:stylesheet>`
  - Change value to current [filename]
How it works at Penn State

- In HTMLOUT folder open [filename4].html and do a find and replace <br/> with <br/>

- Save in DONE folder:
  - catalog, entityrc, psuseal.gif
  - [filename].sgm, [filename].html (all five files for each finding aid)

- Move all but SGML file to FindingAids folder on libraries' website
Making of America II Testbed

Inventory of the Pennsylvania Bridges Collection, 1884-1915

Descriptive Summary

+ Administrative Information

Scope and Content

+ Index Terms

+ Container List

7/17/99

Susan Hamburger: SGML--XML--HTML

This is the way the HTML display looks, without the frame and navigator bar.
I’ve highlighted the Photographs Series as an example of how we will embed a digital archival object within the finding aid.
Links to individual digital objects

Box: 3
[Item 5241]
Simple truss bridge, Pittsburgh, Pa., built 1903. Pittsburgh, Ft. Wayne and Chicago Railroad bridge crosses Allegheny River. Photo by Pittsburgh, Ft. Wayne and Chicago Railroad. <DAO SHOW="EMBED" ACTUATE="USER">"</DAO>

Box: 1
[Item 5242]
Three-hinged arch bridge, Schenley Park, Pittsburgh, Pa., built 1898? Bridge crosses Panther Hollow. Photo by Steiner and Co.