Advanced Technologies and Libraries, Archives, and Museums

An Introduction

Daniel Pitti
Institute
for
Advanced Technology in the Humanities
University of Virginia

§

Riga
Library School
November 2007

Overview

- Traditional Media
- Cultural Heritage Professionals
- Computer and Network Technologies
- Database and Markup Technologies
- Conclusion: Opportunities and Challenges

Traditional Media

- Books, manuscripts (medieval and modern), photographs, video and sound recordings, maps, drawings, painting, sculpture, architecture …
- The accumulated evidence of human activity (both good and bad)
- A world of physical artifacts
- Participants: creators, curators, and users

Cultural Heritage Professionals

- Librarians, Archivists, Museum curators
- Cultural heritage professionals
- Fundamental responsibility:
  - remember
  - on behalf of everyone
  - everywhere
- Collect, describe, preserve, make accessible
- A profound responsibility

Traditional Tools

- Card and print catalogs; Printed finding aids; Inventories and registers
- To organize, describe, manage, preserve in order that the artifacts be found or discovered, and used
- Computers and network technology transform the tools
- Enhance and expand professional objectives

Computer and Network Technologies

- Computers and digital media
  - New types of cultural objects constantly emerging (born digital)
  - Re-represent traditional media in digital media (born again digital)
  - Change the tools used my cultural heritage professionals for both new and old media
- Computers separate the medium of storage from the medium of access and rendering
- Network technologies enable communicating objects, anywhere, anytime

Database and Markup Technologies

- Character data (as opposed to picture, sound, 2-D, and 3-D data)
- From 1s and 0s, to sequences of 1s and 0s, to repertoire of characters: a,b,c, … A,B,C,…; 1,2,3, …
- Unicode: the entire repertoire of human characters (?)
- Two predominant technologies for representing character data
  - Database technologies
  - Markup technologies

Database and Markup Technologies

- Both technologies standardized in late 1980s, and revised in second generation in 1998
- SQL (Structure Query Language)
- XML (Extensible Markup Language)
- Complementary rather than competing
- Each optimized to perform certain task efficiently and well
- Each has complementary strengths and weaknesses

Databases and Data-centric

- Regular number of components (fields)
- Order not generally significant
- Each component restricted to data (in internal delimiters)
- Regularized structure; little or no hierarchy
- Relations of fixed number of types
- Processing of data components dependent on strict datatyping, formality, accuracy, and consistency

Data-centric Examples

- Passport application
- Student records
- Bibliographic records
- Authority records (library)
- Census data
- And many, many more

Markup Technologies and Document-centric Data

- Irregular number of components
- Serial order is significant
- Semi-regular structure
- Unbounded hierarchy
- Arbitrary mixing of data and markup
- Arbitrary number of interrelations within and among documents

Document-centric Examples

- Books
- Journals and journal articles
- Poems
- Newspapers
- And many, many more

Technology and Reality

- Not all character-based documents conform to one or the other models
- Archival finding aids, for example, have components that map well to markup model and still others to databases mode
- Which technology is best?
- Decision to be made, based on priorities and objectives
- Frequently a very difficult decision

Success of XML

- Markup technologies developed by the document-centric community
- Alternative to word processing and text processors: separating what the data is from the render and other processes applied to it
- SGML to XML in 1998
- Since 1998

Success of XML

- Use for document-centric as expected
- Unexpected:
  - Data communication
  - Computer to computer
  - Computer to people
- Database technologies make extensive use of markup technologies
- What is next?

Integration of Markup and Database Technologies

- Major database developers working on next generation databases
  - Integration
  - Relational or object-relational architectures
  - Native XML architectures
- XQuery with SQL extensions to integrate the data
- Opportunity: the strengths of both technologies in one architecture

Some Uses of XML by Cultural Heritage Communities

- Metadata (or Control data)
  - MODS, MADS, METS, EAD, EAC, Dublin Core, MIX, …
- Text representation and analysis
  - TEI and others
  - Mss transcribed and encoded
  - Books and articles transcribed and encoded

Types of Metadata

- Descriptive data (cataloging)
- Administrative data
  - Technical data
  - Rights data
  - Source description
- File or Address Data
- Structural Data
- Rendering (or behavior) Data

Conclusion

- Digital objects have become part of the cultural heritage canon
- Computers and network technologies offer cultural heritage professionals with
  - New tools
  - New opportunities to more effectively fulfill professional objectives
  - New opportunities to push the boundaries of professional activity

Discussion