Building and Testing an "Off-the-Shelf" Digital Library: The Penn-Oxford University Press Digital Books Project

Presenter: *Roy Heinz - Director, Information Systems; University of Pennsylvania Library*

The project will build an online collection of newly published Oxford University Press monographs in all fields of history. Over a period of 5 years, it will provide Web access to 1500-2000 titles with full-text and graphics using the PDF format.

The project seeks to combine and modify existing tools to achieve an economic, timely and scalable digital library. It will include a formal study, with the assistance of an external consultant, of the impact of electronic versions on learning, teaching and research, on the sale of paper books and on the economics of publishing.

The full collection will be restricted to the Penn community but there will be a demonstration site available to the Internet audience.

- Building Online Documents
  - Oxford University Press monographs are delivered to Penn in PostScript format
  - Adobe® Acrobat software is used to "distill" PostScript to Portable Document Format (PDF)
  - Adobe Exchange is used to reassemble the files and create an online image of the book
  - Compose® (an Exchange "Plug-in" from Infodata Systems) automatically adds "bookmarks" and links from each monograph's Contents and Index pages
- Creating metadata and access tools
  - Bibliographic records are added to the library's OPAC with PDF file links using 856 tags
  - Extensible Markup Language (XML) and the "Dublin Core" metadata element set build a database of finding tools
  - Verity® Information Server creates access to XML and full text indexes of the PDFs
  - Verity Agent Server alerts users to new documents which match their profile of interests
- Presenting the collection

        - XML and "stylesheets" allow dynamic creation of web pages without constant page editing as well as easy stylistic changes to the pages as the project evolves
        - Predefined general categories (e.g. American History or History of Music) display quick collection summaries
        - User-defined groupings using Verity's keyword searches of Dublin Core and full-text indexes provide custom sets across the full collection
        - Adobe Reader "Plug-in" for Web browser enables free, seamless, access to documents
- Analyzing use
    - Access control and web logs provide statistics on what type of user (e.g. undergraduate, department) views what type of monographs
    - XML and "stylesheets" allow dynamic creation of web Formal surveys examine complex relationships between online access to the collection and changes to teaching and research trends as well as effects on hard copy sales