Implementing DSpace at Rice: Lessons Learned

Lisa Spiro
Director, Digital Media Center
Fondren Library, Rice University

Digitization 101: TLA Pre-conference
April 2008
Scholarly Communication: The Long View

- 20th C research dissemination: Rice Institute Pamphlets:
  - Quarterly publication containing talks given at Rice, papers by faculty
- 21st C research dissemination: Rice Digital Scholarship Archive
  - Available to anyone with an internet connection
  - Constantly updated
  - Built on DSpace software

The Meaning of the New Institution

I

The foundation: its source

It is a common saying in drawing-room and market-place that we are living in a wonderful age. Perhaps no known period of the past towers up to it, unless it be the age of Pericles, or that in which the Roman Empire was consolidated, or that of the Reformation. No features of the age are more striking than the handsome foundations which have been provided by private donation for lengthening the days of man and enriching the content of his spiritual life. Everyone of ten years knows the names of Alfred Nobel and Cecil Rhodes, of Mr. Carnegie and Mr. Rockefeller, of Girard and Peabody, of Johns Hopkins, Leland Stanford, and Cornell: the names of these gentlemen are household.
Overview of Presentation

- Why Rice adopted an Institutional Repository (IR)
- Why Rice uses DSpace
- How Rice coped with challenges in implementing DSpace
  - Content: developing collections
  - Technical: Extending DSpace’s capabilities
- The Growth of DSpace
- Lessons Learned
The Rationale for an IR

- Make the university’s research accomplishments more **visible**
- **Manage** the university’s digital assets in a single place
- **Preserve** the university’s digital assets
- Provide **open access** to research materials
- Disseminate **gray literature** (e.g. presentations, white papers, pamphlets)
Why to Use DSpace

Top Reasons to Use DSpace

- Largest community of users and developers worldwide
- Free open source software
- Completely customizable to fit your needs
- Can manage and preserve all types of digital content -- text, images, moving images, mpegs, datasets
- Used by educational, government, private and commercial institutions
- Can be installed easily "out of the box"

Also:

- Support for digital preservation
- OAI-PMH (Open Archives Initiative- Protocol for Metadata Harvesting) support means content widely accessible
- Standards-based: Dublin Core, etc.
- Scalable, robust
Encouraging Submissions to DSpace

- Studies have shown that faculty members are reluctant to deposit their works in an IR because of “redundancy with other modes of disseminating information, the learning curve, confusion with copyright, fear of plagiarism and having one's work scooped, associating one's work with inconsistent quality, and concerns about whether posting a manuscript constitutes ‘publishing’.” (Davis & Connolly of Cornell, 2007)

- Rice’s strategy has been to:
  - Partner with Rice research groups & researchers to put their existing collections into DSpace
  - Use DSpace for collections of research materials digitized by the library, often in partnership with scholars & research groups
Electronic Theses & Dissertations

- Why ETD collection?
  - Rice theses & dissertations among most requested ILL items
  - Represents university’s intellectual strengths (nanotech, bioengineering, etc)

- Building ETD collection
  - Contains 6684 theses & dissertations from 1918-2007
  - Purchased PDFs directly from ProQuest--instant collection
  - Working with Graduate Studies on requirement for grad students to deposit their work w/ ETD collection
Partnerships with Academic Departments

- Target departments that already have digital collections but want to offload management to the library

- Example
  - Electrical & Computer Engineering research publications database

Recent Submissions

- The geometry of low-dimensional signal models
  - Wakin, Michael B. (Rice University, 2007)

- Distributed multi-scale data processing for sensor networks
  - Wagner, Raymond S. (Rice University, 2007)

- Distributed redundant representations in man-made and biological sensing systems
  - Rozell, Christopher John (Rice University, 2007)

- Physical layer algorithms with limited feedback: Power control and coding strategies
  - Khoshnevis, Ahmad (Rice University, 2007)

- Code design for the relay channel
  - Chakrabarti, Arnab (Rice University, 2007)
Partnerships with Scholars

- Indigenous Australian languages
  - Linguist’s collection of materials related to her research into native Australian languages
  - Selling point: avoid complete loss of data in event of natural disaster

- Ancient Rome
  - Images of ancient Rome
  - Selling point: enable collaborative use of digital images
  - Required Digital Library Initiative to be flexible in using DSpace for protected materials
Library Projects (1)

- **Our Americas Partnership**
  - IMLS funded project to build collections & tools for scholars of the Americas
  - Partnership with Maryland Institute of Technology & the Humanities & Rice’s Humanities Research Center

- **Travelers in the Middle East Archive** *(TIMEA)*
  - IMLS & KIIT funded project focused on travel in Egypt
  - Partnership with scholars
Library Projects (2)

- Rice Institute Pamphlets
  - Record of university’s intellectual output brought into 21st C
- Collections from Rice’s Special Collections & Archives
  - Rice history: William Ward Watkin
  - Early Houston histories and documents
  - Robert Schumman collection
- Rice Thresher (student newspaper)
- Shepherd School of Music performances
Promoting Rice Digital Scholarship Archives

- Working with collections development librarians to make case for IR to departments and faculty
- Building on momentum from recent mandates by the NIH & Harvard to promote open access
Staffing for Rice’s DSpace

- 1 full-time programmer/sys admin (primary technical responsibility)
- Part-time support from:
  - Executive director of Digital Scholarship
  - Director of Digital Media Center
  - Digital project manager for Our Americas Archive
  - Digital media specialist to advise on digitization
- With additional support from
  - Special collections/archives for developing collections
  - Library IT for training and technical support
  - Technical Services for metadata creation & QC
  - Reference for encouraging faculty participation
DSpace’s Limitations for Rice

- DSpace was originally designed for born-digital, simple items (like PDFs).
- Lacks support for deeply hierarchical collections
- Import tool could be better
- Only supports as-is content presentation
  - Transformation to HTML: TEI-encoded XML texts
  - Enhanced features: zoomable images, streaming media
Adapting to DSpace’s Structure

- Can represent structured content in metadata or in DSpace Communities, Sub-Communities and Collections
- DSpace’s hierarchies are based on organizational units, not content
- DSpace hierarchies not sufficiently flexible for representing structured content from some collections
  - Collections only hold Items
  - Communities only hold Collections and other Communities
Adapting vocabulary and structures for DSpace

- **The Rice Institute Pamphlet**
  - Index to Rice Institute Pamphlet, Vols. 1-46
- **Volume 01**
  - Index to Volume One
  - Volume One, No. 1
  - Volume One, No. 2
  - Volume One, No. 3
  - Volume One, No. 4
- **Volume 02**
  - Index to Volume Two
  - Volume Two, No. 1
  - Volume Two, No. 2
  - Volume Two, No. 3
  - Volume Two, No. 4
- **Volume 03**
  - Index to Volume Three
  - Volume Three, No. 1
  - Volume Three, No. 2
  - Volume Three, No. 3
  - Volume Three, No. 4
- **Volume 04**
  - Index to Volume Four
  - Volume Four, No. 1
  - Volume Four, No. 2
  - Volume Four, No. 3
  - Volume Four, No. 4

Community: all RIP materials

Sub-community: each volume

- Collection: volume index
  - Single index PDF for all four issues
- Collection: each issue
  - Items for each article
Enhancing DSpace

Our programmer has been busy developing &/or implementing various improvements to DSpace, including:

- Batch importing
- XML publishing support
- Customized user interface
- Support for streaming media
- Support for zooming/panning images
Importing content into DSpace

- Somewhat clunky web-based interface not well-suited to importing large collections
- DSpace’s batch import tool limited
- We often get metadata in Excel spreadsheets from
  - Research groups
  - Library projects
  - Vendors

Our programmer developed a batch import script that:
- Normalizes dates
- Reformulates metadata as needed (splits fields, transforms to Dublin Core)
- Matches metadata to objects, separates thumbnails, generates reports
Integrating XML Support

- Need to transform XML (TEI Light) to HTML for display in web browser
- DSpace treats XML as “supported” format, but raw XML isn’t appropriate for presentation to users.
- Could store pre-generated HTML, but the HTML is just for presentation, not archival
- Programmer developed mechanism for associating XML document with stylesheet and delivering HTML to the browser
**XML vs HTML reading version**

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
- <TEI2 xmlns="TEI.2">
  - <teiHeader xmlns="teiHeader" status="new" type="text">
    - <fileDesc xmlns="fileDesc">
      - <titleStmt xmlns="titleStmt">
        - <title xmlns="title">Egypt through the stereoscope: a journey through the land of the Phara</title>
        - <author xmlns="author">Breasted, James Henry, 1865-1935</author>
        - <respStmt xmlns="respStmt">
          - <resp xmlns="resp">Creation of machine-readable version</resp>
          - <name xmlns="name">TechBooks</name>
          - <resp xmlns="resp">Creation of digital images</resp>
          - <name xmlns="name">Electronic Resources Center, Fondren Library, Rice University</name>
        </respStmt>
        - <resp xmlns="resp">Conversion to TEI 2-conformant markup</resp>
        - <name xmlns="name">TechBooks</name>
        - <resp xmlns="resp">Paging and proofing</resp>
        - <name xmlns="name">Electronic Resources Center, Fondren Library, Rice University</name>
      </titleStmt>
      - <pb xmlns="pb" n="2" s="27">The hum of voices reaching the koran, the shouts and the gay laughter and the juggling of the poet at the gate of the cemetery, as they receive the food distributed by the rich, all this, with the citadel and its splendid mosque outlined against the blue of skies in the distance, gives the traveler a typical scene of oriental life, as it is found only at Cairo. Behind us at the ibn en-Nasr, which is also on this side of the city, the jubilation and merry-making are even more marked than here, and the temporary booths, with piles of sweets, the merry-go-rounds, the dancers and the rejoicing multitudes give one the impression of a large country fair. In a few days these same people will be following the pieces of the sacred carpet or "kowsir" from the citadel to the Mosque of Hasanidin, in a rejoicing procession to which all Cairo will turn out. That procession is one of the most interesting public events at Cairo, and we shall later have the opportunity of observing it.</pb>
    </fileDesc>
  </teiHeader>
</TEI2>
```
Creating Customized User Interfaces: Manakin

- Rather being limited to DSpace interface, “brand” each collection with unique look and feel
  - Serve community’s need to market itself and develop own identity
- For TIMEA (timea.rice.edu):
  - Portal integrating different systems required
    - TIMEA specific navigation being built into item-level records
  - Usability testing indicated that users had difficulty navigating if the item results lacked the same look & navigational elements
Creating User Interfaces with Manakin

- XML-based customizable user interface developed by Texas A&M’s Digital Initiatives team
- Relatively easy to create custom designs and interfaces
- Helps with structural problems (e.g. render “Collection” as “Issue”)

Title: Cairo, Egypt
Creator(s): Arbuckle Bros. (Copyright); Knapp, Joseph F. (Printer)
Summary: Six drawings of stereotypically Egyptian motifs
URL: http://hdl.handle.net/1911/2324
Date: 1691
Original: Original trade card: “Cairo, Egypt” (Arbuckle Bros.). From the collection of Paula Sanders, Rice University
Subject: Streets—Egypt—Cairo
Subject: Animals Egypt
Subject: Nile River
Subject: Cairo (Egypt)—Buildings, structures, etc.
Subject: Egyptians
Growth of DSpace

- Organizational
  - DSpace Foundation created in 2007
  - Developing a network of service providers
  - More training planned
  - Building active support community

- Technical
  - DSpace keep getting better, with active community of 80+ developers
  - DSpace 1.5 just released; DSpace 2.0 coming soon
  - New features include customizable UI, configurable submissions, & improved browsing
Benefits of Implementing DSpace

- Research output more **visible**: top rankings in Google Scholar
- Provide **open access** to research to global community—users represent 144 countries
- Enables library to take leadership role in **long-term management** of digital content

Global Community Using Rice’s DSpace
[Tan=non-users]
Lessons Learned

- Important to participate in DSpace community by subscribing to listserv, attending conferences, etc.
- Technical expertise is essential in implementing DSpace (whether in-house or via service provider)
- Need to be strategic in building collections
- Customization can make the software work better for you, but it has costs:
  - Some performance trade-offs
  - When DSpace is upgraded, must make customizations work w/ new release
Acknowledgments

- IMLS and KIIT for support of TIMEA
- IMLS for support of Our Americas Archive
- Fondren Library staff
  - Geneva Henry
  - Sid Byrd
  - Monica Rivero
  - Jane Zhao
  - Diane Butler
  - Amanda Focke
  - Debra Bailey Kolah
  - And many others….
For more information

- Visit the Rice Digital Scholarship Archive at [http://scholarship.rice.edu/](http://scholarship.rice.edu/)
- Contact Lisa Spiro at [lspiro@rice.edu](mailto:lspiro@rice.edu)
- View bookmarks related to DSpace at [http://www.diigo.com/user/lspiro/dspace](http://www.diigo.com/user/lspiro/dspace)