

# Categories of Digital Preservation Service from Institutional Repositories: Preliminary Thoughts

Digital Preservation in State Government:  
Best Practices Exchange  
Wilmington, North Carolina  
March 27-28, 2006

Cal Lee

School of Information and Library Science  
University of North Carolina, Chapel Hill



# Outline

- UNC-CH Digital Curation / Institutional Repository Committee
- Concept of Levels of Service for Digital Preservation
- Categories of Service
- Our Initial Approach

# UNC-CH Digital Curation / Institutional Repository Committee

- DC/IRC chaired by Helen Tibbo, UNC-CH SILS
- Sponsored by Executive Vice Chancellor & Provost, Robert Shelton
- Charge: design & develop pilot repository, explore digital curation approaches
- Steering Committee, chaired by Sarah Michalak, University Librarian & Associate Provost for University Libraries, first met Sept 6, 2005
- 22 official participants from 11 departments

# DC/IRC Structure

- Subcommittees:
  - Governance
  - Technologies & Technical Infrastructure (TTI)
  - Repository Management (I chair)
  - Community & Digital Assets (CDA)
  - Guidance, Training & Engagement (GTE)

# Sustainability & Trust as Essential Conditions

- Institutional repository is “an organizational commitment to the stewardship of these digital materials” (Lynch, 2003)
- OAIS accepts “responsibility to preserve information and make it available for a Designated Community” over the “long-term”

# Levels of Service for Digital Preservation

## Levels of support for file formats (DSpace):

- Supported: Fully support the format & preserve it
- Known: Can recognize the format, but can't guarantee full support
- Unsupported: Can't recognize the format, but can keep the bits as "application/octet-stream"

# Differences in Demand for Digital Preservation (Lavoie, 2003):

- “High-end” – perpetual access, preserving “look and feel”)
- “Low-end” – short-term, “intellectual content only”

# Levels of Service for Digital Repositories (Lefurgy, 2002):

- Optimal – “maintenance in perpetuity without significant alteration of content, structure, or any other significant characteristic,” “retain their original context,” “discoverable through multiple attributes”
- Enhanced – “some persistent qualities but that lack others,” e.g. “structural rules are variable or metadata are incomplete”
- Minimal – “few, if any, persistent characteristics”

# Categories of Service

- Categories of contributors or content types that carry similar set of promises for:
  - Pre-ingest & ingest work
  - Validation
  - Creation of metadata & documentation
  - Intellectual property protection
  - Security
  - Access Controls
  - Long-term preservation of content, context, structure, behavior

# Factors Defining Categories

- **Institutional arrangements & agreements**
- Appraisal & collection development considerations
- Complexity of digital objects & relationships
- Significant properties of objects & relationships
- File formats: availability of documentation of formats, openness & industry support
- Availability & sustainability of:
  - Technical resources (hardware, software, systems)
  - Human resources (attention & expertise)

# Evolution of Categories

- Need to start with some working categories, but attend to ones that emerge
- Analogous to faceted classification

# Examples of Service Categories

- Self-submissions of complex objects by students
- Data sets from research partners sitting on a public web server
- University records submitted from the President's office in PDF, along with documentation, according to retention schedule

# Our Initial Approach

- Building from “Requirements of a Trustworthy Electronic Records Preservation System in a College or University Setting” (Glick & Wilczek, 2005)
- Within given service categories, identify each requirement as a “must,” “should,” or “may”