Columbia University Libraries / Information Services

Digital Asset Management
Digital Preservation
Digital Publishing
Stephen Paul Davis
Director, Libraries Digital Program,
Columbia University Libraries (2002-present)

Previously:
- Director, Library Systems Office, CUL
Introductions

Columbia University Libraries / Information Services

– One of the top five academic research library systems in North America.
– The collections include over 10 million volumes, over 100,000 journals and serials, as well as extensive electronic resources, manuscripts, rare books, microforms, maps, graphic and audio-visual materials.
– The services and collections are organized into 22 libraries and various academic technology centers.
– The Libraries employs more than 470 professional and support staff. The website of the Libraries at www.columbia.edu/cu/lweb is the gateway to its services and resources.
Introduction

- LDPD: Libraries Digital Program Division
- PRES: Libraries Preservation and Digital Reformatting Division
- CDRS: Center for Digital Research and Scholarship
- CCNMTL: Center for New Media Teaching and Learning
- LITO: Libraries Information Technology Office
Key Issues

• Why does a research library need digital asset management?

• Which options are available?

• What is Columbia University Libraries’ approach?
WHY DOES A RESEARCH LIBRARY NEED DIGITAL ASSET MANAGEMENT?
Research Libraries must:

. . . manage, preserve and provide access to unique digitized content created from their print, manuscript and multimedia collections

E.g.,

- papyri, medieval manuscripts, image and object collections, rare books and journals, archival collections, useful reference and curricular material
Research Libraries must:

. . . collect and provide ongoing preservation of and access to *University-generated content* of all kinds (working papers, conference proceedings, theses, preprints, data sets)

E.g.,

*Academic Commons (Columbia' Institutional Repository)*
Research Libraries must:

... accept, process, preserve and provide access to *born-digital* personal and organizational archival collections (e.g., of authors, political figures, publishing houses, philanthropic organizations)

E.g.,

*PricewaterhouseCoopers Records, 1891-2000*
Research Libraries must:

. . . harvest, preserve and provide ongoing access to significant and at-risk Web sites of potential value to scholars and researchers of the future

E.g.,

Columbia Human Rights Web Archive
WHICH OPTIONS ARE AVAILABLE?
Options for Asset Management, Preservation & Access

*Commercial systems:*

Enterprise systems; focus on facilitation of content re-use within large organizations. Content is often marketing- or sales-related, e.g., product imagery, logos, marketing collateral or fonts.

*or:*

Production asset management systems focused on managing assets as they are being created for digital media production (video game, 3D feature film, animation, visual effects shots, etc.); may include workflow features.
Options for Asset Management, Preservation & Access

*Home-grown systems:*

Anything from basic “file system / file naming” techniques, to locally-developed database applications
Options for Asset Management, Preservation & Access

Open Source Systems:

- ResourceSpace
- Razuna
- EnterMedia
- Notre DAM
- Etc.

-Fedora . . .
Fedora Commons Repository Software

Fedora provides a repository system and robust application development platform for:

– Digital asset management
– Digital asset ‘curation’
– Long-term digital preservation
– Controlled access to digital assets and collections
Fedora Commons Repository Software

- Store all types of content and its metadata
- Scale to millions of objects
- Access data via Web APIs (REST/SOAP)
- Provide RDF search (SPARQL)
- Rebuilder Utility (for disaster recovery and data migration)
- Entire repository can be rebuilt from digital object and content files.
- Content Model Architecture (define "types" of objects by content)
- Many storage options (database and file systems)
- JMS messaging (your apps can "listen" to repository events)
- Web-based Administrator GUI (low-level object editing)
- OAI-PMH Provider Service
- GSearch (fulltext) Search Service
- Multiple, customer driven front-ends
Fedora Commons Repository Software

- Robust open-source development community
- Supported by Duraspace consortium & several funding agencies
- Broad adoption within higher education (see User Registry)
- *Columbia is a “gold” member of the Duraspace and one of our programmers is a Fedora “committer”*
Stone Soup
WHAT IS COLUMBIA’S APPROACH?
Columbia’s Approach

– Began Fedora implementation in 2008
– Released “Academic Commons” in 2009
– Began ingest of legacy data in 2010
– Implement “Staff Collection Viewer” in 2010
– Implement “Preservation Repository Functionality” 2011
– Really “just the beginning”
Columbia’s Approach

Digital Archiving Overview

Columbia’s Approach

Digital Library Content Management & Publishing

Columbia’s Approach

Columbia’s Long-Term Preservation Plan

Preservation Storage Infrastructure
Columbia’s Approach

Repository Tools

• Metadata creation & editing tool (Hypatia)

• Staff Collection Viewer

• Command line admin tools
Columbia’s Approach

Fedora Repository Content

• Digital Resources (all formats)
• Object Relationship Information
• Metadata types: *descriptive, technical, structural, administrative & rights*
• Metadata formats: MODS, PREMIS, MIX, PBCore, etc.
Columbia’s Approach

Fedora Repository Content - 2

• ca. 180,000 objects ingested or staged for ingest
• ca. 50 TB
• ca. 95 different projects / collections
Columbia’s Approach

Future CUL Fedora Developments

• Columbia public collections viewer
• Website preservation functionality
• Digitization workflow-management tools
• Scientific data set ingest and curation
• Many new content projects & collections
Now You Know

• Why a research library needs digital asset management

• Which options are available

• What Columbia University Libraries’ current approach is
QUESTIONS?

daviss@columbia.edu