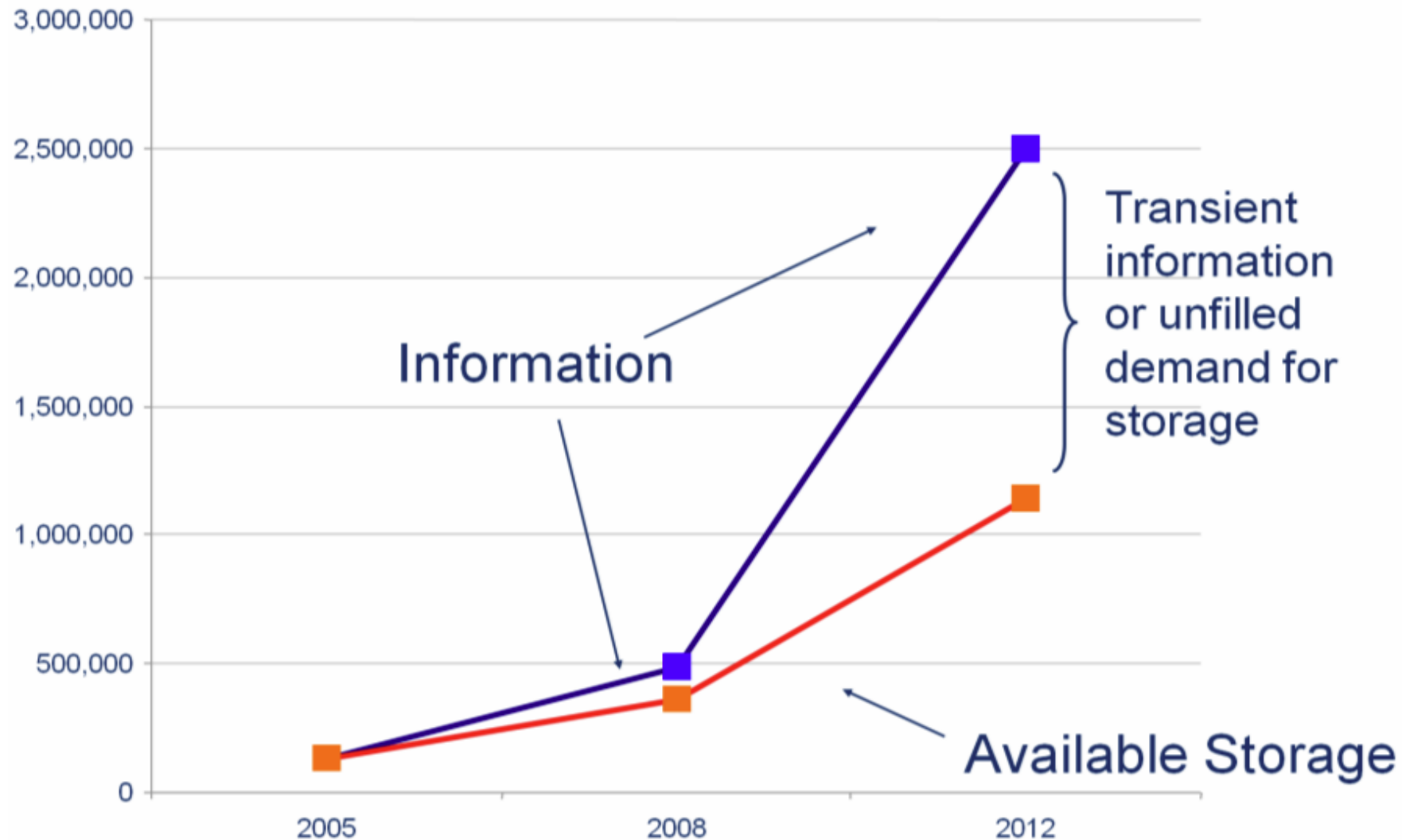


Digital Preservation

Strategies and Technologies

Stephen Paul Davis
Director, Libraries Digital Program
Columbia University
April 2011

Petabytes Worldwide



Digital Preservation Challenges

- **Uncertainty** about selection criteria for assessing long-term value, especially with large-scale data sets, small “hand-crafted” digital collections, and the emerging genres of collective authorship on the Web;
- **Misalignment** of incentives between those who are in a position to preserve and those who benefit from preservation and access;
- **Lack of clear responsibility** for digital preservation, coupled with a prevailing assumption that it is someone else's problem;
- **Little coordination** of preservation activities across diffused stakeholder communities;
- Difficulty in separating **preservation costs** from other costs, that is, in distinguishing between the processes of making things available now and making things available in the future; and
- Difficulty in valuing or monetizing the **costs and benefits** of digital preservation, which are necessary to secure funding and investment.

Ensuring Access

- Digital information is vital
- Digital information is fragile
- Access in the future means action today

Columbia University Libraries

Preservation Strategies and Technologies

Columbia University Libraries / Information Systems

- 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.

Technology at Columbia Libraries

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

Preservation Focus at Columbia

1. Local Digitization Projects
2. Institutional Repository / Data Sets
3. Born Digital Archival Content
4. Archived Web Sites

Local Digitization Projects

Preservation of unique digitized content created from 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, oral histories

Institutional Repository

Preservation of University-generated content of all kinds
(working papers, conference proceedings, theses,
preprints, **research data sets**)

[Academic Commons \(Columbia' Institutional Repository\)](#)

[NSF Data Management Plan Support](#)

Born Digital Archival Content

Preservation of born-digital personal and organizational archival collections (e.g., of authors, political figures, publishing houses, philanthropic organizations)

E.g.,

- [Human Rights Watch Records](#)
- [Bomb Magazine Records](#)
- [Carnegie Corporation of New York Records](#)

Archived Web Sites

Preservation of significant and at-risk Web sites of potential value to scholars and researchers of the future

E.g.,

[Columbia Human Rights Web Archive](#)

... archived via “Archive-It,” a service of the [Internet Archive](#).

Other Areas

- Digitized Books: [Hathi Trust Digital Library](#)
- Commercial Journal Literature: Portico, LOCKSS, JSTOR

Columbia's Digital Preservation Infrastructure

Columbia is building a repository system and robust application development platform for:

- Digital asset management
- Digital asset '[curation](#)'
- Controlled access to digital assets and collections
- Long-term digital preservation

Essential infrastructure for digital preservation.

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*

Fedora Stone Soup



Stone Soup

Once upon a time, somewhere in [Eastern Europe | China | India | etc.], there was a great famine in which people jealously hoarded whatever food they could find, hiding it even from their friends and neighbors. One day a wandering [soldier | beggar | wanderer | etc.] came into a village and began asking questions as if he planned to stay for the night.

"There's not a bite to eat in the whole province," he was told. "Better keep moving on."

"Oh, I have everything I need," he said. "In fact, I was thinking of making some stone soup to share with all of you." He pulled an iron cauldron from his wagon, filled it with water, and built a fire under it. Then, with great ceremony, he drew an ordinary-looking stone from a velvet bag and dropped it into the water.

By now, hearing the rumor of food, most of the villagers had come to the square or watched from their windows. As the soldier sniffed the "broth" and licked his lips in anticipation, hunger began to overcome their skepticism.

"Ahh," the soldier said to himself rather loudly, "I do like a tasty stone soup. Of course, stone soup with *cabbage* -- that's hard to beat."

Soon a villager approached hesitantly, holding a cabbage he'd retrieved from its hiding place, and added it to the pot. "Capital!" cried the soldier. "You know, I once had stone soup with cabbage and a bit of salt beef as well, and it was fit for a king."

The village butcher managed to find some salt beef . . . and so it went, through potatoes, onions, carrots, mushrooms, and so on, until there was indeed a delicious meal for all. The villagers offered the soldier a great deal of money for the magic stone, but he refused to sell and traveled on the next day. The moral is that by working together, with everyone contributing what they can, a greater good is achieved.

CUL/IS Fedora Architecture

- Fedora Software Platform
- Digital Preservation Storage System
- Application and authentication middleware
- Applications to support Long Term Preservation Archive

**Columbia University Libraries / Information Services
Digital Archiving Overview**

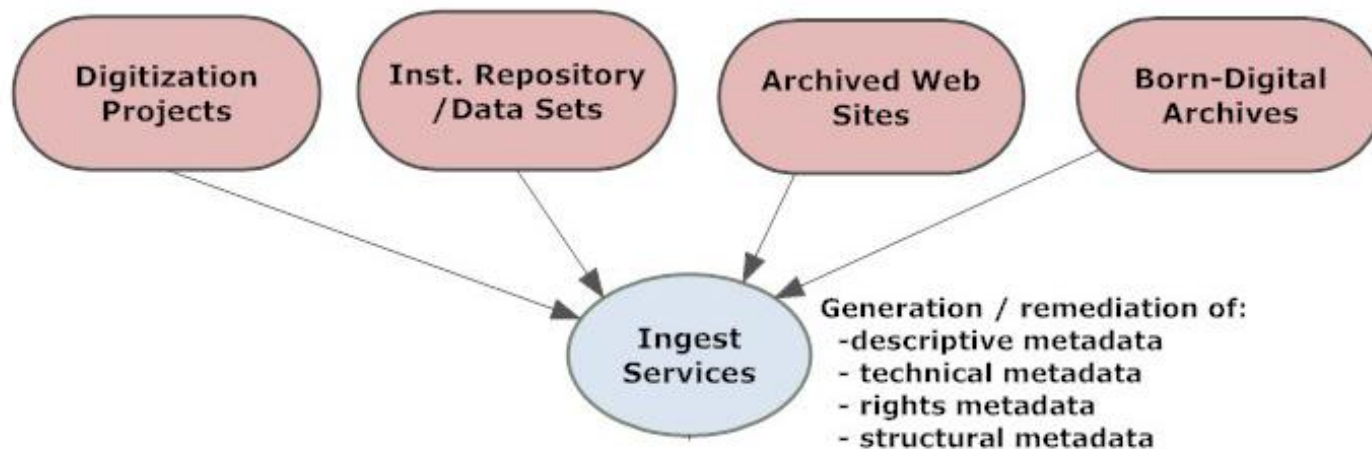
**Digitization
Projects**

**Inst. Repository
/Data Sets**

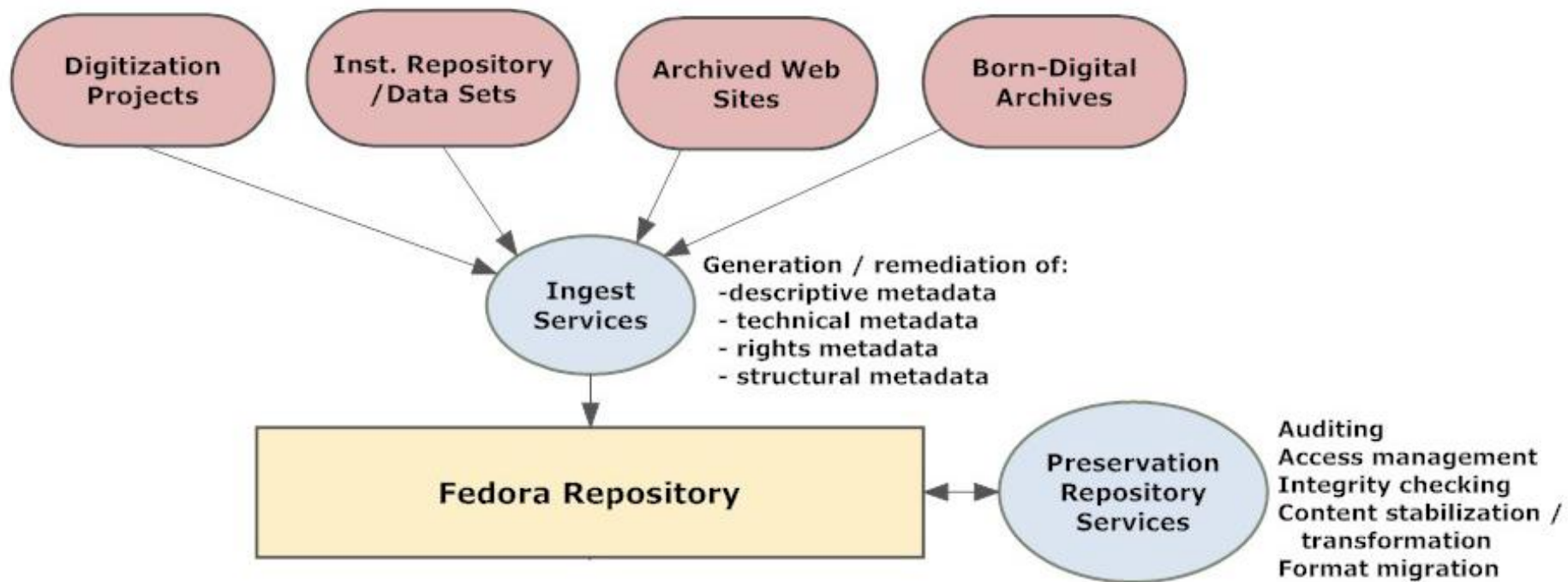
**Archived Web
Sites**

**Born-Digital
Archives**

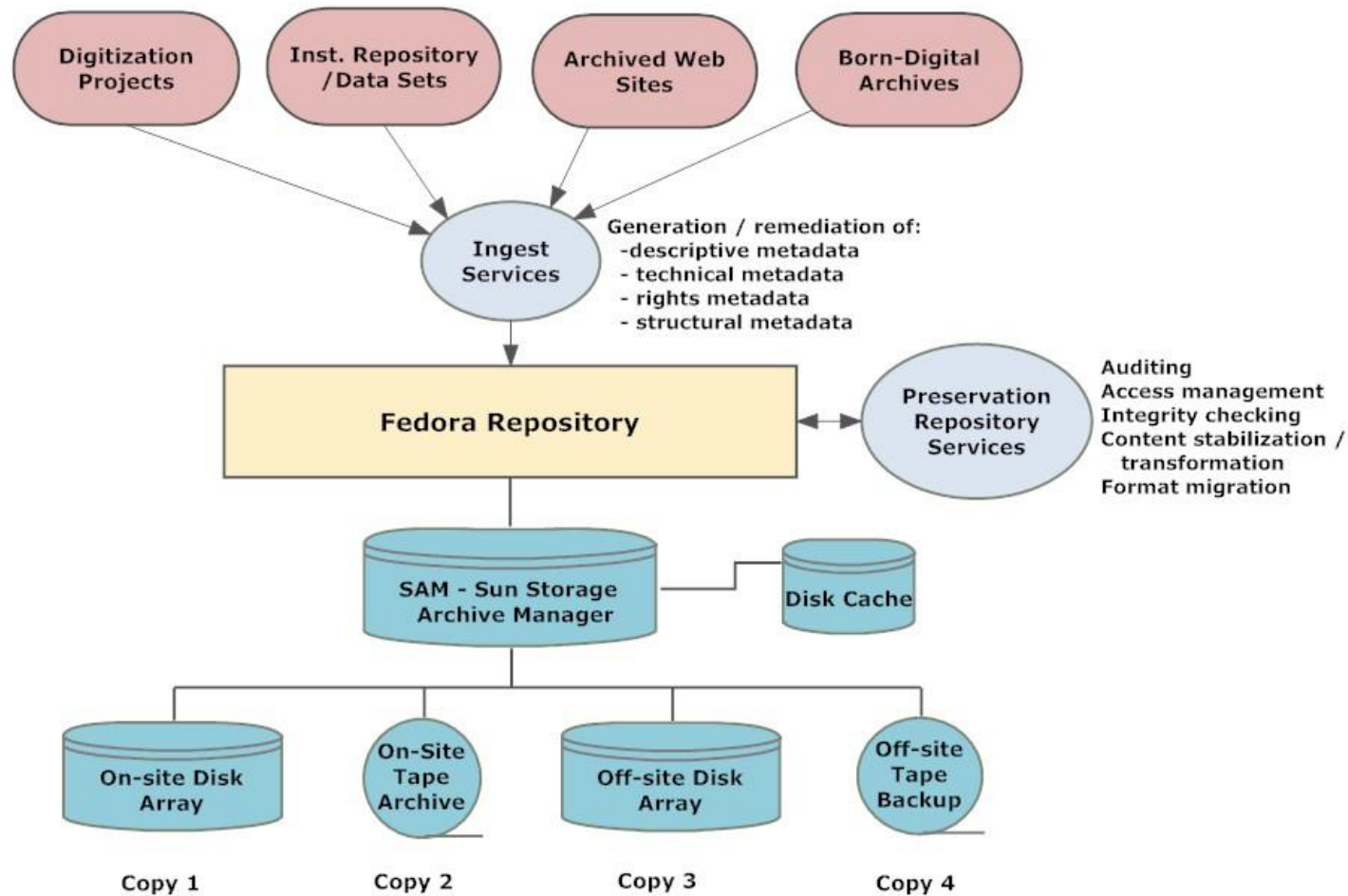
**Columbia University Libraries / Information Services
Digital Archiving Overview**



Columbia University Libraries / Information Services
Digital Archiving Overview



Columbia University Libraries / Information Services
Digital Archiving Overview



Fedora Repository Tools

- Metadata creation & editing tool (Hypatia, 2010)
- [Staff Collection Viewer](#) (2011)
- Command line admin tools

TRAC

- [Trustworthy Repositories Audit & Certification: Criteria and Checklist \(2007\)](#) – OCLC, CRL & NARA. Previous: *RLG-NARA Task Force on Digital Repository Certification (2003)*
- [Drambora: Digital Repository Audit Method Based on Risk Assessment \(2008\)](#) – *Digital Curation Centre (UK) and DigitalPreservationEurope*

Issues

- **Access** – Needed for Preservation?
- **Sustainability** – Budgetary, organizational, technological
- **Preservation** – Dark archives? Dim archives? Open archives?
- **Curation** -- Semantic and ontological continuity and comparability of the collection content

The Future

- Mixed, distributed environment
- Efforts to coordinate within subcommunities
- A few large trusted digital archives in the U.S. and elsewhere
- Many smaller digital archives that provide initial stabilization and packaging for later deposit in larger archives

What Would You Preserve ?

- A. Harlem Hospital Murals
- B. Encyclopedia Iranica
- C. Mapping Gothic France
- D. Top 100 CUL Oral Histories

Only the beginning ...

Questions:

daviss@columbia.edu

[CU Libraries Digital Program](http://www.columbia.edu/cu/libraries/inside/units/ldpd/)

<http://www.columbia.edu/cu/libraries/inside/units/ldpd/>