



Digital Collections: Storage and Access

Jon Dunn
Assistant Director for Technology
IU Digital Library Program
jwd@indiana.edu



Storage

- Why is storage an issue?
 - Space requirements
 - Persistence
 - Accessibility
- Needs depend on purpose of storage
 - Capture/encoding
 - Access/delivery
 - Preservation



Storage: Working Space

- Space for storage of digital files during capture/encoding/quality control process
- Possibilities
 - PC hard drive
 - File server / LAN
- Issues
 - Capacity, backup, speed, accessibility



Storage: Access/Delivery

- Storage of derivative files for web delivery
 - Image, audio, video, text files, etc.
- Possibilities
 - Local web server
 - Commercially-hosted web site
 - Consortial service provider
- Issues: capacity, backup, performance, software integration, maintenance/migration



Storage: Preservation

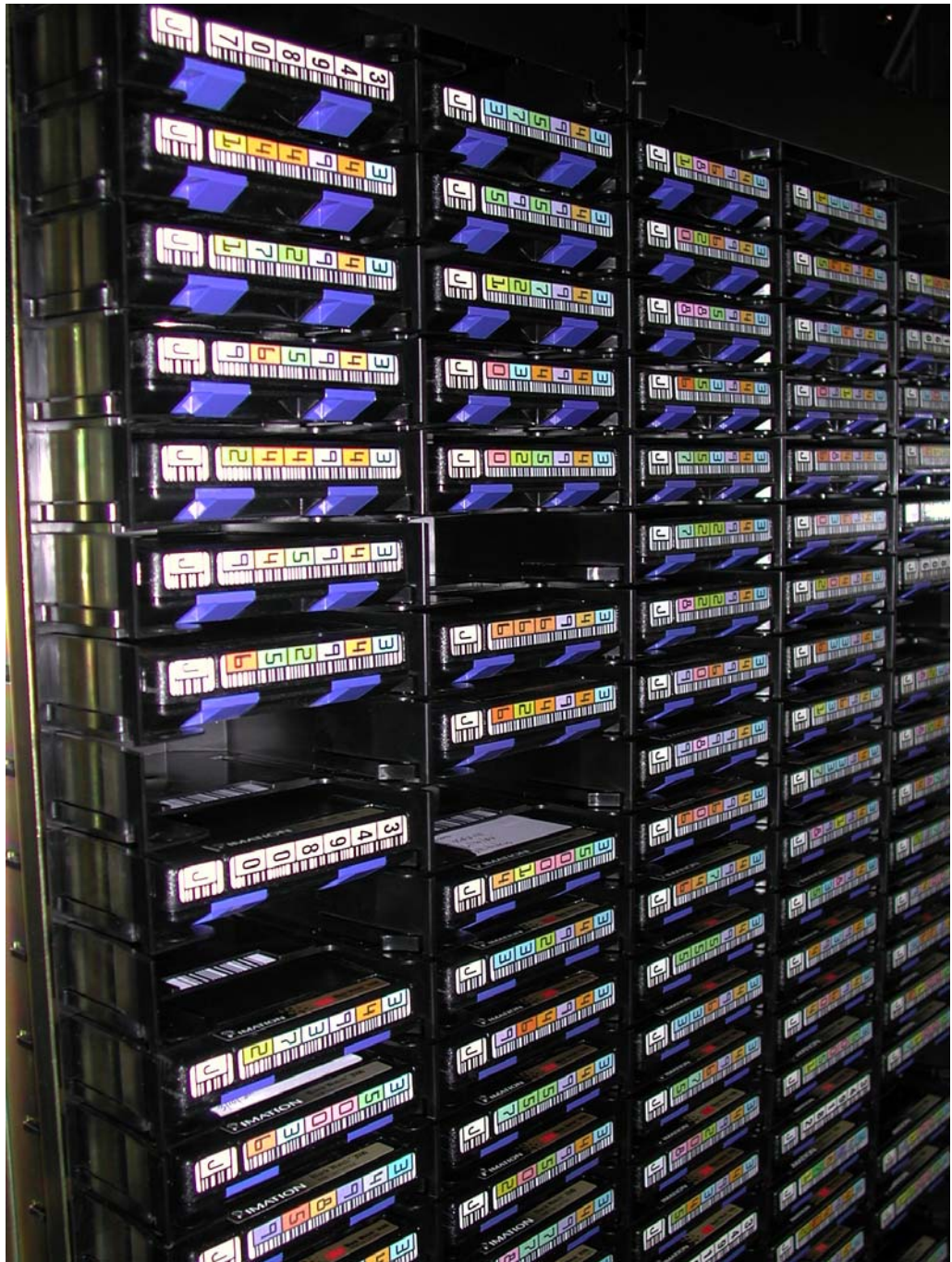
- Much harder problem
- Longer term
 - Issues of longevity of media, hardware, file format
 - “Where did we put the files?”
- Larger files
 - Hard disk storage, traditional backup methods not cost-effective
- Infrequency of access
 - Problems do not become immediately evident



Long-Term Storage Options

- Removable media stored offline
 - Optical
 - CD-R (CD-Recordable)
 - DVD-R (DVD-Recordable), DVD+R, DVD+RW, DVD-RW, ...
 - Tape
 - DLT, 8mm, DAT, ...
 - Pros: cheap, easy, produces tangible item
 - Cons: Low capacity, physical space requirements, unknown longevity, migration, potential format obsolescence
- Online/nearline storage systems
 - HSM: Hierarchical Storage Management
 - Combine disk and automated tape storage with software to keep track of where files are located
 - Locally managed or remote provider
 - Pros: high capacity, migration can be handled by software,
 - Cons: expensive, complex, network bandwidth issues, must trust service provider, potential single point of failure



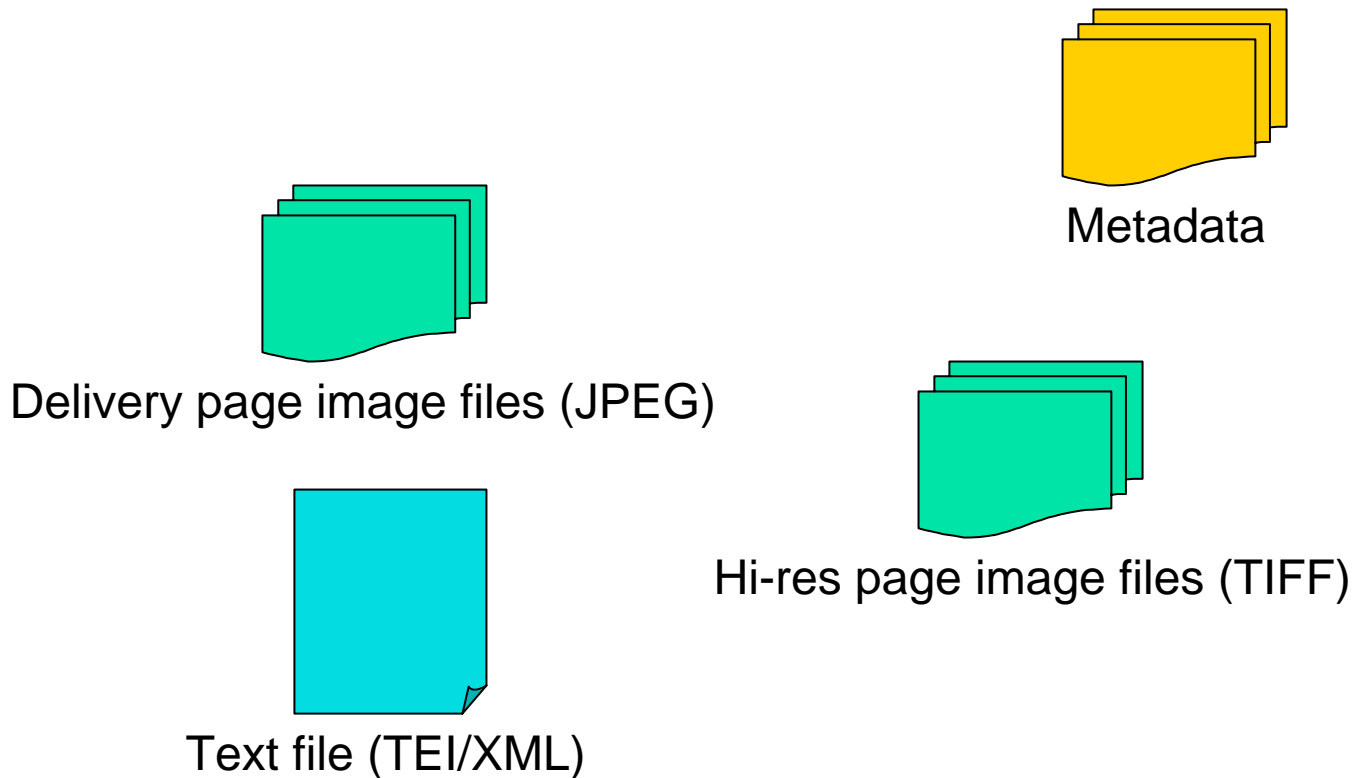




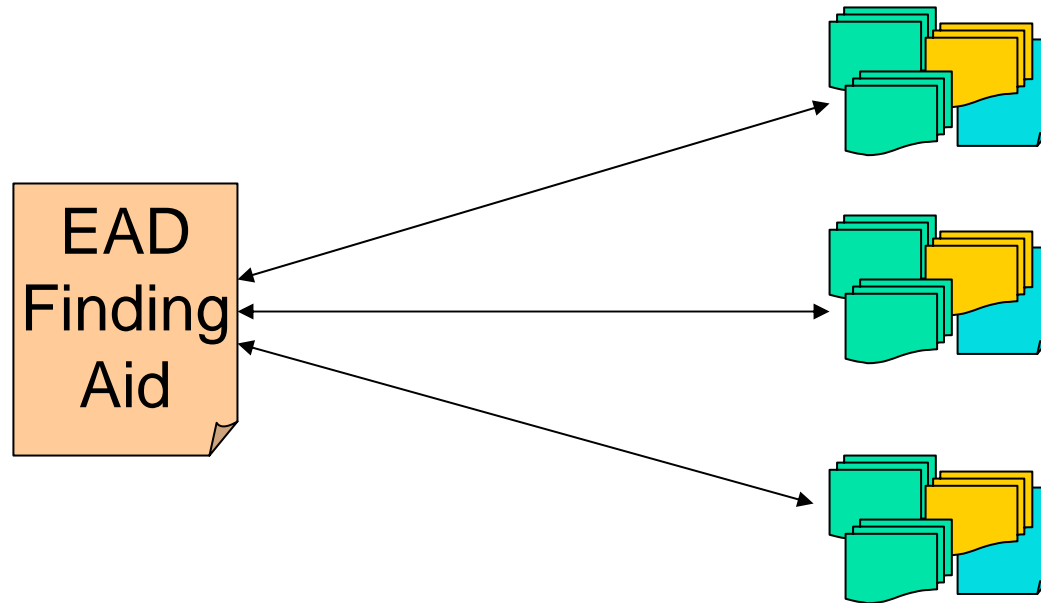
HSM Example: IU's Massive Data Storage Service (MDSS)

- HPSS (High Performance Storage System) software
 - Developed as collaboration of IBM and US national labs
- Four tape robots
 - 2 in Bloomington, 2 in Indianapolis
 - Data can be mirrored
- 540 terabytes (TB) total storage
 - ~75 TB used as of April 2001

A digital object is more than just a file!



A digital object is more than just a file!





DL Objects

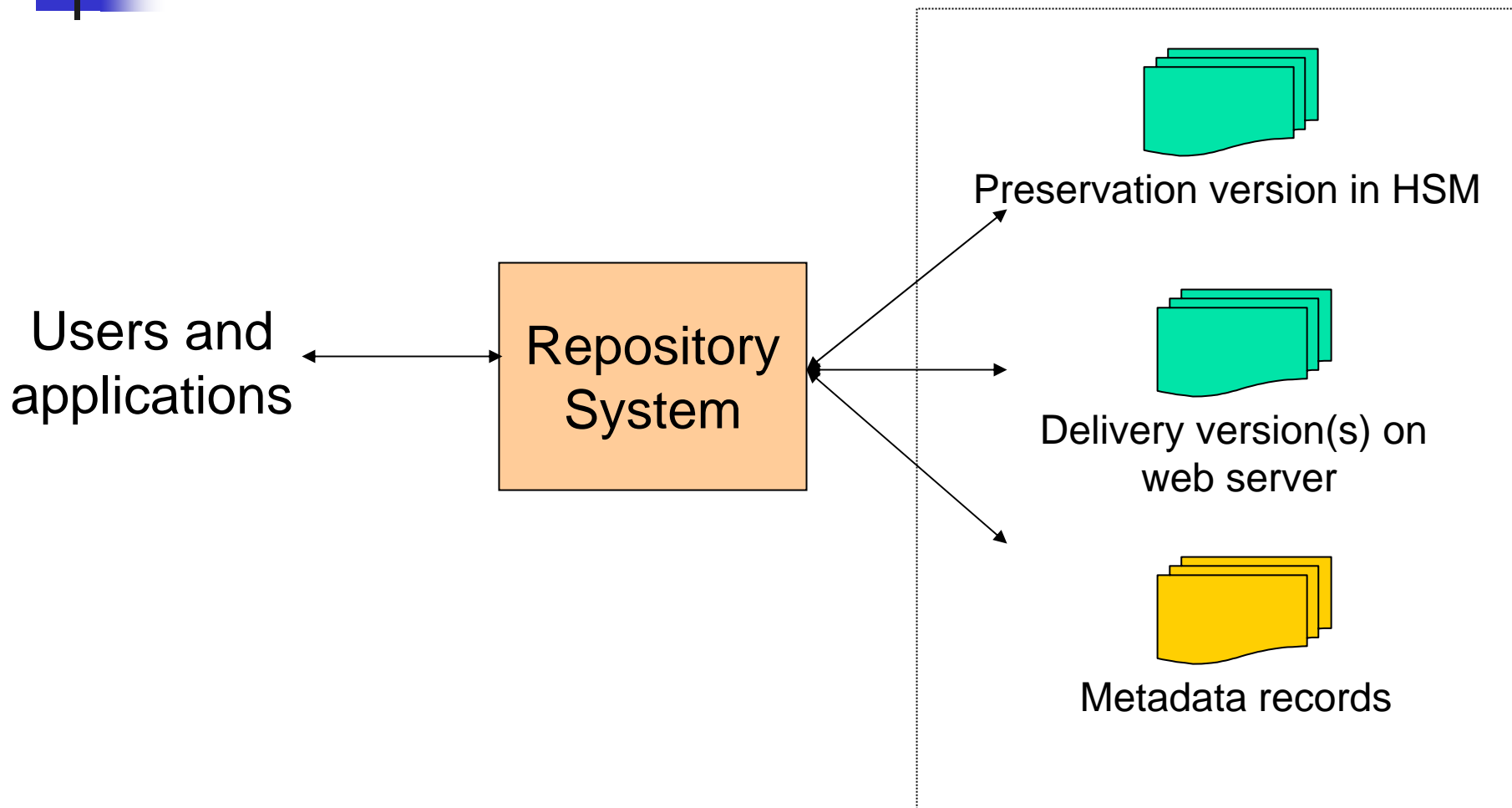
- Digital library “objects” have many parts
 - Metadata
 - Preservation/archival files
 - Delivery files
- How do we keep them connected?
 - Now: Good practice in file naming, directory organization, project documentation -not scalable!
 - Future: Digital object repository



Data Persistence

- Key is **migration**
- Keeping the bits alive
 - Physical media
 - Logical media format
- Keeping the bits understandable
 - File format
 - Metadata
- Small “pockets” of digital content pose a problem for migration

DL Object Repository





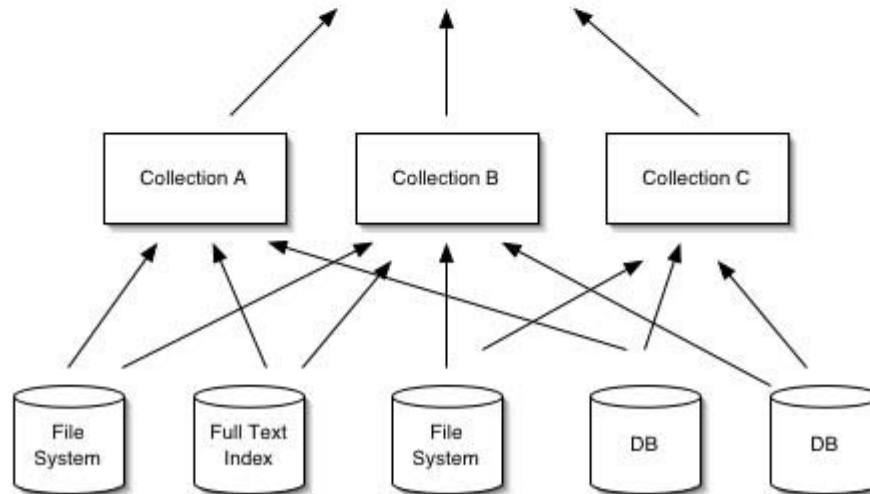
Web Delivery Functions

- Searching
 - Metadata
 - Full text
- Browsing
 - By subject, date, author, ...
- Navigation
 - Page turning, image panning/zooming, ...
- Streaming
 - For audio/video
- Reuse
 - Downloading, format conversion
 - Linking, persistent naming
- Access control
 - If necessary



Digital Collection Delivery Software

- Very complex systems
- Need to integrate data from databases, full-text search engines, file systems, and other sources
- Cross-collection searching
- Commercial
 - ContentDM, Luna Insight, various library management system addons
- Open source
 - UMich DLXS, Greenstone, Eprints, MIT DSpace, ...
- Homegrown



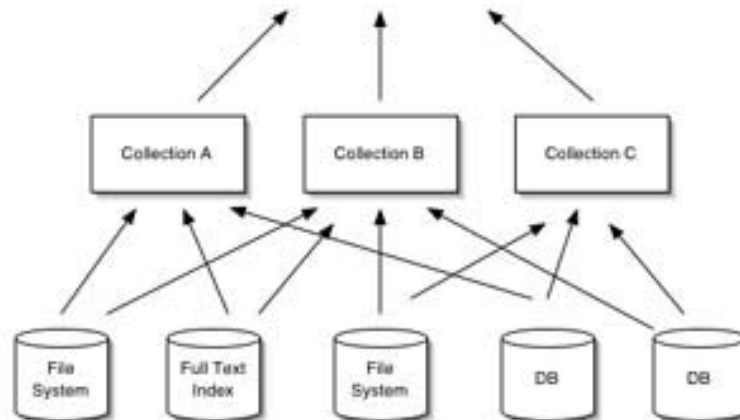


Demonstration

- Hoagy Carmichael Collection,
IU Digital Library Program
- <http://www.dlib.indiana.edu/collections/hoagy/>

Google™

IUCAT



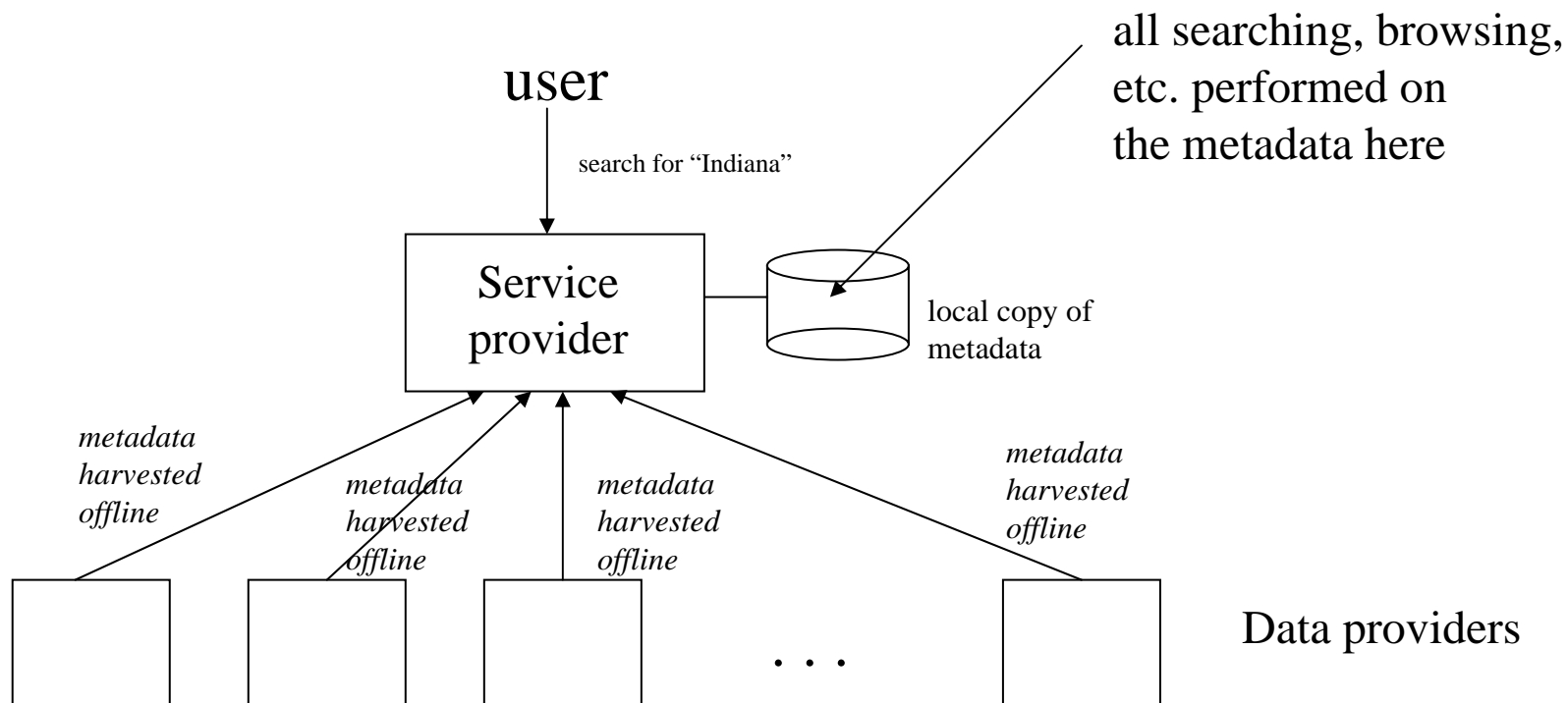


Exposing Digital Resources Broadly

- Pay services
 - RLG Cultural Materials, Archival Resources
- Free services
 - University of Michigan OAIster
 - www.oaister.org
 - UIUC Digital Gateway to Cultural Heritage Materials
 - oai.grainger.uiuc.edu
- OAI-PMH
 - Open Archives Initiative Protocol for Metadata Harvesting
 - www.openarchives.org
- Google

OAI Metadata Harvesting

- Extract metadata from various sources
- Build services on local copies of metadata





More Information

- Bibliography to be made available at:
 - <http://www.dlib.indiana.edu/workshops/alioc03/>