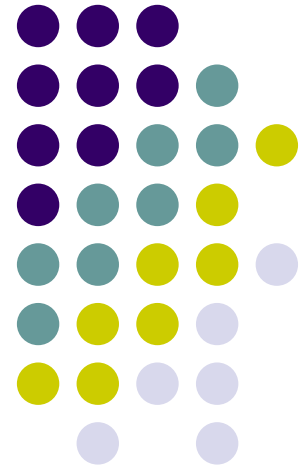
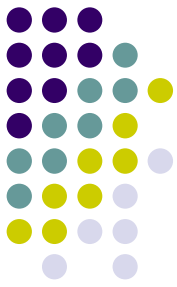


Best Practices for OAI: A Status Report

Kat Hagedorn
Sarah Shreeves

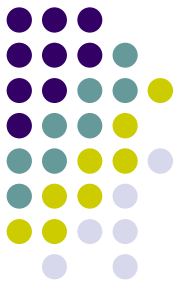
DLF Spring Forum
San Diego, CA
April 14 2005





Outline

- Purpose
- History and background
- Process
- Snippets of Best Practices
- Open issues
- Next steps



Why?

- OAI has lots of options – need guidelines
- Critical time in development of OAI
 - Implemented by many communities
 - Included in content management systems
 - Service providers have battle scars
- Wild, wild west of metadata
 - Not a shared understanding of shareable metadata
- OAI can't be the last stage in a digital project



Purpose

To establish best practices for OAI data and service provider implementations and for shareable metadata.

To facilitate communication between OAI data and service provider

To identify tools needed for the OAI community.

DLF / NSDL current focus
but meant for the wider OAI community



History of Effort

- Part of a DLF grant proposal to IMLS
- Interest in / need for best practices so high, given go-ahead by DLF
- First met in July 2004 in Oakland to hash out all potential issues and develop a plan and timeline



Participants from...

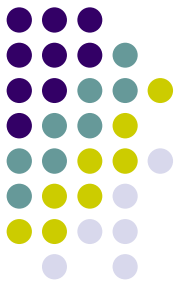
- California Digital Library
- Cornell University
- DSpace
- Emory University
- Indiana University
- Library of Congress
- National Science Digital Library
- OCLC
- Princeton University
- University of Illinois
- University of Michigan
- University of Tennessee
- University of Southern California
- University of Washington

Representatives from both data and service provider communities

Service / Data Providers



- Cognizant of the balance between service and data providers
- Part of effort is establishing and encouraging a culture of communication between data and service providers
- True for both data and service providers



How we work

- A listserv
- A wiki - <http://oai-best.comm.nsd.l.org/>
- Conference calls
- Meetings at the DLF Forums



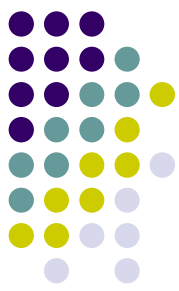
Developing the BPs

- Result of meeting: Word document
- Transcribed to a wiki
- Categorized into major issue areas
- Focus on the two areas most essential
 - Metadata content
 - DP implementation
- Chairs appointed to encourage work

Wiki



- Workspace, not a document
- Work in concert with conference calls
- Issues:
 - Familiarity with working in a wiki
 - Updating notification
 - Noting differing opinions



Example wiki page

OAI Best Practices

[TableOfContents](#)

Best Practices for Data Provider Implementations

Reference documents:

- [\[OAI protocol specifications\]](#)
- [\[OAI-PMH implementation guidelines\]](#).
- [\[OA Forum Tutorial\]](#) (which also includes a very good XML schema section)
- [\[OA Forum Glossary\]](#)
- Technical Specifications in the NSDL [\[Metadata Primer\]](#)
- [\[OAI static repository\]](#)

Issues:

- [\[Documenting Provider Practices\]](#) - Diane (this is also applicable in the Metadata section)
- [\[OAI Identifiers\]](#) - [Complete?](#)

The OAI-PMH 2.0 specification requires that each item in the repository have a unique identifier. This section addresses best practices for OAI identifiers including compliance to the oai identifier scheme.

- [\[Datestamps\]](#) - Tim
- [\[Deleted Records\]](#) - [Complete](#)

If a record is no longer available it is said to be deleted. Repositories must declare one of three levels of support (no, persistent, transient) for deleted records in the deletedRecord element of the Identify response. This section addresses best practices for deleted records including a discussion of the downstream implications for service providers of each level of support.



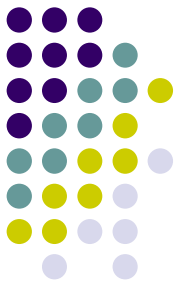
Difficulties...

- Getting things done without formal deadlines
- IMLS and BP groups haven't worked together...yet
- Differing methodologies / interpretations among participants
- Outstanding questions no one can / has time to tackle



Implementing a DP

- Besides administrative issues, i.e., support, time, money...
- Which software (off-the-shelf / build)?
- Does it do everything it's supposed to?
 - (does it follow the OAI protocol)
 - (does it follow the spirit of the OAI protocol)
- Vendor support



Set practices

- MYTH: SPs / harvesters don't use sets
- TRUTH: SPs do selective harvesting
- Set structure

Open Questions:

- Set description
- Communicating structure to SPs



Deleted records

- Three levels (none, persistent, transient)
 - Best Practice = persistent
- No deleted record status
 - SP still serves up deleted records
 - Periodic full harvests necessary = SP time drain

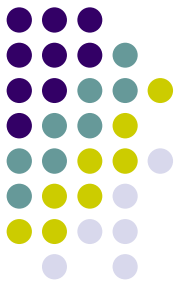


resumptionTokens

- Best practice is to implement resumption tokens
- Method for splitting up OAI responses
- Eliminate huge HTTP responses
- Allow timeout recovery

Open Question

- Goldilocks problem (what size?)



About containers

- For documenting provenance, rights, branding, descriptive info
- Rights info is about metadata record, not resource described
- Rights specification in beta



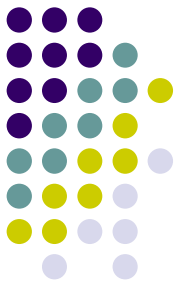
Repository lifecycle

- Registration (2 places)
 - <http://www.openarchives.org/data/registerasprovider.html>
 - <http://gita.grainger.uiuc.edu/registry/>
- Maintenance
- Communication with SP re: changes, status (defunct? temporarily down?)

A Taste of Best Practices for Shareable Metadata



- What is shareable metadata?
 - Proper context / Self-explanatory
 - Consistency
 - Use of standard vocabularies
- More than DC please!



Appropriate Metadata

- What is appropriate for your content?
- What is useful for the service provider?
- A view of your metadata

Recommendations for classes of data elements

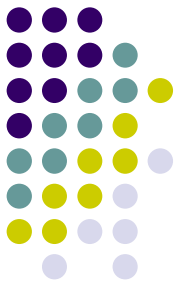


- Titles
- Names
- Dates
- Subjects / Topics
- Language
- Geographic Places
- Identifiers
- Rights for resources
- Bibliographic Citation



Other Metadata Issues

- Granularity
- Describing versions and reproductions
- Appropriate links
- Technical issues
 - Schemas and Namespaces
 - Character Encoding



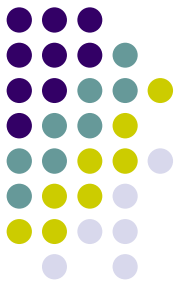
Open issues

- Documentation
 - What?
 - Where?
- Repository descriptive information
 - What?
 - Where?
- Static repositories and gateways
- Talking to the vendor community



Future Work

- Continue development
- Review and feedback from wider community
- New areas of focus: service provider implementations, communication between SP and DPs; tool sets; extending the OAI protocol
- Incorporation of best practices into the IMLS funded training



Questions and Comments?

- Kat Hagedorn khage@umich.edu
- Sarah Shreeves sshreeve@uiuc.edu

Wiki: <http://oai-best.comm.nsd.org/>

Birds of a Feather Session

4:30-6:00

Harbor B – 3rd Level