Experiences Implementing OAI Provider Services

13 September 2001  -- ACM SIGIR, New Orleans
Open Archives: Communities, Interoperability and Services

Timothy W. Cole\((t-cole3@uiuc.edu)\)
Thomas G. Habing\((thabing@uiuc.edu)\)
University of Illinois at Urbana-Champaign
Our Observations & Generalizations

• Creation & Management of Metadata Much More Difficult than Implementation of OAI PMH

• Wide Range of Viable Implementation Options, But Some Workflow Accommodation Is Required

• Scalability & Value of Services That Can Be Built On Top of OAI PMH Remain To Be Proven

• Greatest Strength May Be Content & Format Flexibility for Metadata / Data shared via OAI
Our Architecture Approach

- Descriptive & Structural Metadata
- OAI Administrative Metadata
- Application
- Transport

In Development:
- XML Files
- HTML Meta Tags
- XSLT
- DOM

Relational DBMS (MS Access / SQL Server)
- OAI Provider Service (CGI / MS ASP)
- HTTP (Apache / MS IIS 4, 5)

Z39.50
Our Technical Choices

• Use of Sets:
  Hierarchical, Non-Overlapping, Comprehensive
  Set Name Reflected in Identifier (colon delimited)

• Repository Updates:
  Scheduled Time of Day for Routine Updates
  Repository Taken Offline During Update (Respond 503)

• Flow Control:
  Use with ListRecords & ListIdentifiers Only
  ResumptionTokens Expire with Repository Update
Additional Metadata Formats

- Other Metadata Formats of Interest:
  - Qualified DC in RDF (More Structured)
  - MARC (Have Existing Data)
  - EAD (Have Existing Data)

- Crosswalks to DC
  - Better Done by Provider or Harvester?
  - Complex Crosswalks (e.g., from EAD) a Concern

- Use of Multiple Namespaces with XSD
  - Canonical Namespaces, But No Canonical XSDs
  - XSD Import, ##any / ##other, lax Processing
Our List of Open Issues & Concerns

• IP Rights Management
  – Need Machine-Readable “<about>” Formats
  – Restricted Access Harvesting & Use Limitations

• Use of Sets
  – Consensus on Best Practices
  – Concept of Null Set or Set Information as Part of Record Header Element

• OAI Record Timestamp Granularity