A Public Catalog of Theses and Dissertations

Thomas Hickey
Jeffrey Young

Online Computer Library Center
Outline

• What we want to accomplish
• What we have done so far
• What we would like to do
Intent

• Develop a comprehensive catalog of theses and dissertations
• Make the catalog publicly available
• Experiment with associated services
What We Start With

- **WorldCat**
  - 47,000,000 bibliographic records
  - 4,300,000 theses and dissertations
- **Excellent relationships with libraries**
  - Including national libraries around the world
- **Experience with large sets of metadata**
Related Projects

• OAI Research
  – Harvest/Server software (open source)
• Web Services
  – Name Authority files
  – SRW
• FRBR
  – IFLA’s Functional Requirements for Bibliographic Records
The Current Catalog

- 4,300,000 records
  - Dublin Core
  - NDLTD ETDMS
- OAI 2.0 harvestable (almost)
  - ETD set of 4,760 records
- Name authority linking
Experiences

• Large collection for OAI harvesting
  – Needed some optimization
  – Was one week, now 12 hours

• ETDMS problems
  – Unavailable required fields
  – Uses DC labels, but these are not in the ‘DC’ namespace
Authority Linking Features

• Desirable that links support multiple metadata formats
• Links should be persistent and current
• Links should be meaningful and useful outside NDLTD’s domain
ETDMS Resource Attribute

• The ETDMS format defines an XML resource attribute intended to allow links to authority records

• An OAI GetRecord URL seems a convenient mechanism for linking to individual records in an authority repository
Problems With GetRecord Links to Authority Records

• Pros
  – OAI GetRecord allows URL access to individual records in a repository
  – The OAI GetRecord verb allows various metadata formats to be specified

• Cons
  – Existing GetRecord links to obsolete records won’t resolve to the most current record.
  – GetRecord links require a single explicit metadataFormat to be specified. Accessing alternate formats would require reformulating the URL

PURL: http://authlink.oclc.org/ndltd/oai:laf:n2001-101817
URL: http://authlink.oclc.org/ndltd/oai:laf:n2001-101817/oai_dc

Partial Redirect: http://authlink.oclc.org/ndltd/oai:laf:n2001-101817/
URL: http://alcme.oclc.org/laf/OAIHandler?verb=GetRecord&identifier=oai:laf:n2001-101817&metadataPrefix=

PURL Link Architecture
Possibilities

- Other sets of T&D records
  - India
  - UK
  - OCLC PICA
  - Others?
Plans

• Complete OAI 2.0
• Start regular updates
• Integrate other collections
  – FRBRize them?
  – Multiple metadata schemes?
• Public HTML interface
  – SiteSearch
More Plans

• More sophisticated name matching
• Enhanced authority files?
Questions?