Primo Normalization Subteam

Project Charter

EMORY

3. *Project Scope / Service Description*

Objectives: State the purpose of the project and relation to strategic themes, goals and objectives.

Contribute to the accomplishment of Goal 3 of the University Libraries Strategic Plan (…which is, "With a customer-centered approach, the Library will build distinctive research environments that integrate both print and digital resources and enable connections among users, communities, resources, and tools.") by implementing SI 3.2 Digital Library Services and Next Generation Systems Objective #1 (1.1, 1.2, 1.4) Implement Portal Services (User-centric, integrated interfaces and services for discovery and delivery of resources) by FY2009)

The team will implement Primo to provide to our users (students, faculty, staff):
- Intuitive and effective searching across multiple scholarly resources to enable improved access for researchers of all types
- Access to the rich resources of the Emory libraries, currently searched separately, in an inviting and vastly improved discovery interface

Primo harvests data from a wide range of source databases and formats. For Emory, locally ingested data sources include (but over time will not be limited to):

- Bibliographic records from Unicorn
- XML metadata records from Kirtas digitization
- XML metadata records from ETD
- XML metadata records from http://oai.library.emory.edu/ collections
- XML metadata records from Pitts Theology Library Digital Image Archive
- Insight image library metadata records
- Other local sources to be identified

Data from these sources must be normalized and optimized to facilitate efficient discovery and delivery. The various source formats must therefore be mapped to the Primo Normalized XML (PNX) record. Normalization work is expected to focus primarily on the following parts of the PNX record: display, links, search, facets, sort, duplication detection, grouping (FRBR), delivery and scoping, and ranking.

Project Justification: State the problem, issue, or opportunity this project addresses. Describe the project itself, its impact on library products and services, and its benefit to the Emory Enterprise. Include a description of assumptions, tangible and intangible benefits.

Because of their volume (~2.3 million, growing >100,000 per year), complexity, and historical depth (DOBIS and Unicorn records from OCLC, batch loaded records from multiple sources, and locally input), PNX normalization for Unicorn records is expected to require a significant amount of analysis, experimentation, and change of Primo Normalization rules for discovery and delivery to work effectively. Baseline normalization work needs to be complete by the Primo internal release date (current revised target date is March, 2008). Release of Primo to a wider audience of end users (both staff and, later, the public) can be expected to result in many problems being reported by users. Most of the of these problems can be resolved through Primo PNX normalization changes; others will require corrections to the source data. Considering that we are still encountering and correcting DOBIS data problems 13 years after bringing up Unicorn, PNX normalization work should be considered an open-ended commitment; at some level it will never end.

XML records ingested from other Emory sources, such as repositories, Insight, iTunes University, etc. will present different challenges. Initially, there will be relatively few local sources, but ingestion from these sources will grow over time. Although structurally simpler than MARC records, there are fewer guidelines (e.g. AACR2) for non-MARC record creation, so additional challenges will be presented by these records, which will have far less standardization. As with MARC XML records, normalizing these records will require a combination of Primo PNX normalization changes and corrections to the source data.

Failure to do this work will result in Emory resources being incompletely discoverable through Primo, which will lead either to frustrated users, additional work for service desk staff (to find things that are known to exist), or both.