Project Name: Metadata Optimization Review

Project Manager: Laura Morse

I.   Problem/Value Statement

Library services rely on accurate metadata describing library collections and scholarly resources to ensure that faculty, students, and researchers can discover, access, and use the materials needed for research, teaching, and learning.   Core repositories of library metadata maintained by the Harvard Library need systematic review and remediation in order to improve resource sharing, discovery, enterprise resource management, inventory control, and collaboration. Library metadata standards, strategies, and schemas are undergoing dramatic change due to the changing nature of library collections, data creation workflows, and evolving expectations for interoperability and reuse of metadata. Syndication of high quality metadata ensures that Harvard metadata is optimized for collaborative partnerships within and beyond Harvard.

Metadata stored in Aleph, SFX and Verde will be evaluated.

1) **Aleph** is used by library staff for acquisitions, inventory control and material processing. It also provides public access to a database of Harvard library holdings - over 15 million books, journals, sound recordings, music scores, materials, maps, manuscripts, government documents, and materials on order. The Project team will analyze bibliographic and holdings metadata, recommend steps to improve it by a) reducing coding errors in preparation for OCLC reclamation, b) modifying data elements to improve resource discovery, and, c) preparing metadata for potential use in future systems. To promote data consistency, the group will also review current cataloging workflows and recommend changes to Metadata Standards Working Group to promote optimal metadata.

2) **SFX** allows researchers to link directly from an article citation or abstract in an external database to a variety of related resources determined by Harvard's electronic subscriptions. The SFX KnowledgeBase, a central data repository of information about Harvard's electronic collection, is the key to linking researchers to the material they need. The Project team will examine the SFX Knowledge Base and recommend steps to synchronize it with Harvard's access rights to electronic resources.

3) **Verde** is used by library staff to administer Harvard's electronic resources. It is a centralized repository tracking acquisitions, licenses, usage, cost, access, administrative data and Harvard libraries participating in cost sharing and stewardship of e-resources. Verde interoperates with SFX and includes information about the same full text content as SFX, as well as e-resources that are not in SFX. The Project team will recommend steps for linking Verde ERM records to Aleph and SFX.

Each of these repositories contributes to library services that depend on accurate, consistent metadata. Inaccurate, inconsistent metadata has a negative impact on discovery, access to collections and content holdings, resource sharing tools, collection analysis tools, cooperative metadata creation, and enhancement services.

The goal of this Project team is to gather information about current and historic coding practices and workflows to identify metadata candidates for normalization and improvement, define remediation objectives and strategies, and complete remediation of prioritized data issues.

II.   Vision and Approach

Objectives, strategies and recommendations for will be informed by the findings of the Harvard Library's ongoing review of metadata standards, as well as new tools and services from OCLC.

This project includes several key tasks:

a) Gather information on current and historic data- and work- flows
b) Conduct metadata analyses to identify objectives
c) Prioritize candidate data elements to be remediated
d) Design and execute data clean up processes

These tasks will result in several deliverables:

- A list of key areas to remediate or improve, prioritized by impact on users
- A communication strategy and dashboard to monitor progress
- Final summary report

The project will be complete when accepted recommendations have been implemented and prioritized remediation projects are complete.

III.   Stakeholders/People

This work will benefit Harvard Library patrons, staff, and external partners as well as to the larger research library community.

This work will be accomplished by a collaboration of HUIT LTS in consultation with Harvard Library staff. The LTS Data Migration Analyst will focus on the analysis and project management of the metadata analysis and database remediation project. Additional funding has been approved for additional project resources for metadata editing and remediation activities.

Additional Harvard Library and LTS resources, in the form of the OCLC Data Sync and WorldShare Working Group and Metadata Standard Working group, and direct consultation with library staff, will inform service and metadata review and recommendations.

Metrics, informed by the LTS Data Migration Analyst, will be defined by a working group, and progress will be reviewed by an Oversight Commitee.

IV.   Schedule and Cost

A preliminary schedule has been created, but may need adjustment based on outcome of the analysis, hiring patterns for project staff, and OCLC resources:

Information gathering, data review, development of database remediation project list.   May 2016 – September 2016

Finalize priorities for remediation, communication plan, and dashboard. October 2016

Determine options for data remediation initiatives. November – December 2016

Execute data remediation. December 2016 – April 2017

Prepare final report. April 2017 – May 2017

Project staffing costs have been budgeted and approved by the Harvard Library for FY16 & 17.

V.   Other

Assumptions

New, term positions for HUIT LTS and HL ITS will be posted and filled on schedule.

Harvard Library resources to inform analysis and planning will be made available to work on the project.

Dependencies

Project schedule may be impacted if project resources are not allocated. These include:

- Harvard Library resources to inform analysis and planning projects
- LTS operational resources
- OCLC resources

Risks

Required input from Metadata Standards Working Group and OCLC Data Sync and WorldShare Working Group are required before recommendations for metadata remediation can be finalized

- Decisions about how to optimize OCLC may not be timely
- Decisions about reconciling obsolete data fields may not be timely
- Decisions about OCLC symbol consolidation may not be timely
- Decisions about switching to a multiple record approach for items in multiple formats may not be timely