

Metadata Optimization Project

Project purpose and goals

Library services rely on accurate metadata to ensure that faculty, students, and researchers can discover, access, and use library materials needed for research and education. 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. As Harvard Library embarks on a migration to a new library processing platform, the Access and Discovery Standing Committee spearheaded a project to enhance and upgrade metadata in Aleph, SFX, Verde to achieve two key objectives:

- 1) optimize the ease and accuracy of metadata migration from Aleph to Alma
- 2) increase the quality and efficacy of resource sharing for patrons at Harvard and beyond

	Alma Migration	OCLC Synchronization	Other Corrections (most will use automated scripts)
Target date for subproject completion	Dec. 2017	Aug. 2017	Dec. 2017
No. error conditions identified	68	130	292
No. error conditions defined as priorities		100	
No. error conditions resolved	4	4	0
Estimated no. records impacted	unknown	300,000	Some error conditions will impact all bibliographic and holdings records (approx. 30 million)
Target for no. of record corrections	400,000	90,000	
No. record corrections completed	11,000	53,400	0
Revised forecast for no. corrections by end of project	400,000	90,000	30 million records (all)

Contact Team

Members:

mop@hulmail.harvard.edu

Milestone Timeline

May-September 2016	Information Gathering
October 2016	Finalize Phase 1 Priorities
November - December - 2016	Define and Execute Phase 1 Remediation Options
January 2017	Finalize Phase 2 Priorities
February 2017	Define and Execute Phase 2 Remediation Options
February 2017	Finalize Phase 3 Priorities
May 2017	Define and Execute Phase 3 Remediation Options
May 2017	Finalize Phase 4 Priorities
April-May 2017	Final Report
June 2017	Define and Execute Phase 4 Remediation Options

Weekly Update & Dashboard

June 28, 2017

Estimated number of Aleph bibliographic records to be corrected: 300,000

Modifications to Date

Type	Est. Count	% by Batch
Bibliographic	1,840,942	99
Holding	9,946	98
Item	1,546	62
Orders	1034	96

In the works:

See [prior updates](#) and [project status](#) for more details.

Phase / Issue	Status / % Complete	Est. No. of Records	Aleph Records Corrected
1: Paired field errors (subfield \$\$6)	DATA CORRECTION 84%	Bibs: 29,396	25,749
1: 'Sparse' records	DATA CORRECTION 51%	Bibs: 21,538	11,144
1: Missing/Problematic titles (245)	COMPLETE 98.5%	Bibs: 855	842
1: Invalid identifiers	PLANNING & ANALYSIS	Bibs: > complete db check	Bibs: 0
1: Form of item conflict: Invalid combination LDR/06 & LDR/07	PLANNING & ANALYSIS	Bibs: 8,071 (revised est.)	8,291
1: Form of item conflict: Non-standard or obsolete codes in LDR/06 (Record Type)	PLANNING & ANALYSIS 29.28%	Bibs: 15,413	4,578
1:Form of item conflict: Non-standard or obsolete codes in 008/23 (BK)	PLANNING & ANALYSIS	Bibs: 87	0
1:Form of item conflict: Non-standard or obsolete codes in 008/23 (CF)	PLANNING & ANALYSIS	Bibs: 61	0
1:Form of item conflict: Non-standard or obsolete codes in 008/23 (MX)	PLANNING & ANALYSIS	Bibs: 3	0
1:Form of item conflict: Non-standard or obsolete codes in 008/23 (MX)	PLANNING & ANALYSIS	Bibs: 12	0
1:Form of item conflict: Non-standard or codes in 008/29 (VM)	DATA CORRECTION	Bibs: 42,229	0
1:Form of item conflict: Non-standard or obsolete codes in 008/29 (MP)	DATA CORRECTION	Bibs: 20	0
Alma: Active holdings on deleted/suppressed bibs	DATA CORRECTION Bibs: 24.86% Hol: 17.66%	Holdings: 15,575 Bibs: 16,031	2751 / 3625
1: Set lending byte	PLANNING & ANALYSIS	Holdings 13,588,541 - Unknown policy 514,505 - Invalid values	0

1: Set reproduction byte	PLANNING & ANALYSIS	Holdings 13,669,916 – Unknown policy 515,184 – Invalid values	0
1: Set retention byte	PLANNING & ANALYSIS	Holdings 140,442 – Unknown policy 9,045,492 - Invalid values	0
0: LDR - Invalid length	DATA CORRECTION 95%	Bibs: 376	348
0: LDR/05 (Record status) - Invalid codes	COMPLETE 90.4 9%	Bibs: 10,551	9,548
0: Orphan bibs	PLANNING & ANALYSIS 4.75%	Bibs: 41,000	1,949
0:008 (Fixed field) missing	COMPLETE 100%	Bibs: 36	36
0: 008 (Fixed field) - Invalid length	DATA CORRECTION 11%	Bibs: 918	102
0: 008/06 (Date Type) - Invalid codes	DATA CORRECTION 97%	Bibs: 20,398	19,786
0: 008/15-17 (Language - Invalid & obsolete codes) - All formats except music/sound	DATA CORRECTION 27%	Active Bibs: 8,554	2,270
0: 008/15-17 (Language - Invalid codes or blank) - Sound recordings	DATA CORRECTION 72%	Bibs: 30,256	21,683
0: 008/21 (Music Parts) - Invalid codes	PENDING HL	Bibs: 1,146 (1,143 have obsolete code 'a')	0
0: 008/33 LitF (BK) - Invalid codes	PLANNING & ANALYSIS	Active Bibs: 21,858 contain invalid codes 1217 contain an obsolete code	0
0: 008/33 Alph (SE) - Invalid codes	PLANNING & ANALYSIS	Active Bibs: 406	0
0: 008/33 TMat (VM) - Invalid codes	PLANNING & ANALYSIS	Active and suppressed bibs: 42,257 contain invalid or obsolete codes	0
0: 008/35-37 Place of publication - Invalid & obsolete codes	PLANNING & ANALYSIS	Active Bibs: 152,697	0
0: 260 Imprint missing	PLANNING & ANALYSIS	Judaica	Manual by unit
0: Obsolete 440 convert series statement to 490/830	COMPLETE 100%	Bibs: 1,161,189, Approx. 1.2m tags	1,724,769
Alma: Item Processing Status - invalid codes	COMPLETE 100%	Items: 1	1

Alma: Item Status - invalid codes	PENDING HL 19%	Items: 226	43
Alma: Items: Material Type - invalid codes	COMPLETE 100%	Items: 930	929
Alma: Order Status - invalid codes	COMPLETE 100%	Orders: 1010	1,010
Alma: Acquisition Method - invalid codes	COMPLETE 100%	Orders: 23	23
Alma: Open Orders	COMPLETE	Orders to be closed to be identified by ASWG as part of test load process.	
Alma: RLN conversion	COMPLETE	Holdings: 10,260. Debbie corrected 9,263 holdings. She will forward list of holdings that could not be resolved to contact at sublibraries for resolution.	9,263
Alma: LDR07 with Harvard-defined '9'	COMPLETE 100%	Bibs: 17	17

Project Approach

This project is funded by the Harvard Library and will be managed by a project team from Library Technology Services. The project team will gather information from Harvard Library staff, the OCLC Data Sync Working Group, and Metadata Standards Working Group about current and historic coding practices and workflows in Aleph, SFX and Verde to identify and recommend metadata candidates for normalization and improvement, define and recommend remediation objectives and strategies, and complete remediation of prioritized data issues.

An oversight committee representing Library Technology Services, Access & Discovery Standing Committee, and Harvard Library ITS and Access Services will prioritize work and approve data mediation strategies.

Objectives, strategies and recommendations will be informed by information gathering from experts across Harvard Library.

- Gather information on current and historic data and workflows
- Conduct metadata analysis to identify candidates for correction
- Prioritize candidate data elements to be corrected
- Design and execute data correction processes - use automated processing as much as possible

Deliverables

- A list of key areas to correct 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 correction projects defined as highest priority have been completed.

Project Team

Lynn Stram, Metadata Migration Analyst, Library Technology Services (lead)
 Corinna Baksik, Library Technology Services
 Michael Edwards, Library Technology Services
 Laura Morse, Library Technology Services

Allison Powers, Library Technology Services
Additional Term Project Resources (tbd, Library Technology Services, Information and Technical Services)

Oversight Committee

Michelle Durocher, Information and Technical Services
Laura Morse, Library Technology Services
Ken Peterson, Access Services
Tracey Robinson, Library Technology Services
Scott Wicks, Information and Technical Services
Suzanne Wones, Harvard Library

Timeline

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

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

Charter

[Metadata Optimization Charter](#)

Resource Links

[PowerPoint Slides from Public Presentations \(October 2016\)](#)

- [Metadata Optimization Project - Weekly Update](#)
- [Metadata Optimization - Project Status](#)