

Zephyr Advisory Group
August 17, 2016

Present:

Kathryn Stine
Gary Charbonneau (Recorder)
Chew Chiat Naun
Sandra McIntyre
Ryan Rotter
Todd Grappone
Tim Cole
Jon Rothman

Absent:

John Mark Ockerbloom

Agenda:

1. Zephyr update
 - o [Zephyr Q3 Roadmap](#)

KS:Project to continue remediating retrospective record clustering issues is a big one, with several record subsets to work through. Currently, the team is addressing a subset of Zephyr clusters where some records which had included >1 OCLC number have been updated to include just one master number and resubmitted. Zephyr team is creating test records for a number of use cases and will run tests of a new program designed to re-assign clustering identifiers where OCLC numbers or bib system numbers no longer agree.

Designing new, improved approaches to tracking/reporting on data handling events in the “life of a record” as it moves into and through Zephyr to output.

Addressing HT PSC Collections Committee request for Zephyr reporting in support of (suspected) single-volume monograph duplicate analysis. Several sample reports on subsets of (suspected) single-volume monographs already shared with the Collections Committee:

- U.S. Federal government documents
- 10 or more volumes that share a clustering identifier

Working on providing a report that samples from the entire subset of (suspected) single volume monographs that will include representation of publication dates and source institutions.

ACTION: Kathryn to post duplication reports shared with the HT PSC Collections Committee to the ZAG Google site (Zephir Documents - Background page).

Launched a new daily report on digitization agent code changes - shared with HT operations staff to investigate and address unexpected changes.

Metadata contributor data gathering plan to kick off later in August/early September. Will take a three phase approach starting with structured 1:1 conversations with 6-8 metadata contributors selected from institutions that represent the range of institutions that submit content/metadata to HT. The second phase will be to solicit survey responses from metadata contributors from all 40+ contributing institutions and the final phase will be to dig in on any leads from the survey results with online focus groups. Data gathered/analyzed will be used to inform development and workflow decisions that will improve the submissions process.

Kathryn heading to Ann Arbor again in September to meet with HT staff. One goal for connecting with HT operations staff face to face will be to confirm shared reference points for metadata/content ingest coordination.

- DLF Forum / Metadata Assessment Working Group

KS: There is a new DLF Metadata Assessment Working Group that was established earlier this year: <https://wiki.diglib.org/Assessment:Metadata>. They aim “to build guidelines, best practices, tools and workflows around the evaluation and assessment of metadata used by and for digital libraries and repositories”. Last week this new working group sent out a call for metadata evaluation and assessment use cases which will be gathered and analyzed in advance of a discussion to be held at DLF Form this November in Milwaukee. The Zephir team is compiling use cases to submit and looks forward to participating in discussion of the analysis.

ACTION: Kathryn will share the use cases that the Zephir team submits to the DLF Metadata Assessment Working Group’s call with the ZAG.

2. MUSAG update.

TG: Will share update by email.

TC: Made progress on review of environmental scan. Considering audience for the environmental scan report - clear appeal to librarians, but there may be some broader appeal as well. An opportunity area that’s really emerged from the scan is the recognition that there are potential benefits to thinking through how to enhance HT metadata.

3. Metadata watch group – confirm plan

KS: Let's revisit the watchlist template that's been circulated to review the list of issues that we'd like to track, identify links/leads that aren't already in the sheet, and see who on ZAG is well-positioned to monitor issues in the list.

Goal with having a watchlist and having folks monitor issues is to stay out in front of developments that could have impact on Zephyr/HT bib metadata. We could perhaps take a more light-weight approach than what's set out in the template we have going. Ultimately, the idea is to have folks who are already tracking issues also be on the lookout for when/how/where these might be of import to Zephyr/HT bib metadata.

ACTION: We'll include watchlist updates as a standing agenda item on future ZAG calls and will further refine our approach at the next meeting (describing more specifically why ZAG/HT would be interested in a particular issue, what a trigger to action might be, etc.). ZAG members who are primary monitors on issues can share out either at meetings (or in between meetings by email) developments that may be of import to Zephyr/HT bib metadata. See watchlist for primary monitors.

NNC: There is a close relationship between **BIBFRAME** and **LD4P (and the related LD4L Gateway)**. I will take this on as the group's primary monitor.

KS: Great! We can consolidate LD4P under BIBFRAME development. We also want to try to track **BIBFRAME uptake** as well - any ideas on how to do this?

TC: There is a good listserv on BIBFRAME.

KS: I'll be monitoring that listserv to stay aware of BIBFRAME developments and uptake.

Jon Rothman had mentioned developments he'd heard about at this past ALA Midwinter on **OCLC Research work on Knowledge Vault for Libraries**. I can follow up with him to see if there is anything useful for us to track on this; might be too specific.

John Mark Ockerbloom had suggested on our last call that he could monitor developments in **consuming linked open data authority datasets**. I think the main thing here is to understand what datasets are available/become available that could be of benefit to HT services, what conditions are associated with data use and updating, and to keep track of projects that are employing these data streams, what challenges and opportunities they've encountered.

Metadata provenance: Tim's name pencilled in here. Not sure what activity has happened recently with the Dublic Core group (link in spreadsheet); perhaps there is W3C work on this to investigate?

TC: This area is pretty wide open - that work on metadata needs to be done is the consensus, but not done deal. A big issue is to determine when data is authoritative and how to deal with uncertainty/ambiguity in the data. It's hard to track. HTRC is to be doing work in this area, though I'm not directly doing that work.

There are a number of **PCC Linked Data Task Groups**, Naun is on the PCC URIs in MARC group. Additionally, Stephen Folsom (Cornell Library, MUSAG member), is active with a [PCC group looking at the work entity](#) along with Karen Coyle and Jean Godby, so we could talk with him about tracking this work.

ACTION: Todd will check in with Stephen on tracking this work.

TC: There is some ongoing work on **FRBR**, in particular, how to model relationships between collections and items.

KS: I need to get catch up with FRBR developments, so can monitor this one.

We will leave **RDA** tracking this hanging for now. If anyone has expertise and/or interest in tracking RDA developments please sign up on the spreadsheet.

On our last call, Tim, suggested that he could monitor developments on schema.org (the bibliographic extension). And we could consider consolidating **linked data in WorldCat** here.

4. URIs in MARC – discussion

KS: Zephir team has received test records received from University of Queensland which have authority identifiers in \$0 subfields (for main entries, subjects, etc.). Is there a best practice is emerging for formatting this data. How should we suggest to contributors that they share this data?

NNC: Per MARC Advisory Committee decision at ALA 2016, include URI's without parenthetical "(uri)".

KS: Great - thanks for this update.

ACTION: The Zephir team will consider adjusting the HT bib metadata specifications to guide formatting of this data so that folks adhere to best practice and so that we're able

to parse, store, and make this data actionable. Kathryn will share proposed language for the spec with the ZAG.