Zephir Advisory Group
February 17, 2016

Present:
Jon Rothman (recorder)
Chew Chiat Naun
Kathryn Stine
John Mark Ockerbloom
Ryan Rotter
Tim Cole
Gary Charbonneau

Absent:
Todd Grappone

KS: Todd won't be joining us today.
MUSAG has created a draft charge for a subgroup/working group on Bibliographic Metadata Management Policy. Patti Martin (on MUSAG) shared the draft with Kathryn who made it available to ZAG members shortly before this meeting. ZAG members should review and comment on the charge soon, since MUSAG is moving quickly. Patti and Kathryn talked about how there is likely a role for ZAG in that group's work, though it has yet to be defined.

ACTION: Review and comment on the MUSAG Bibliographic Metadata Management Policy Working Group draft charge by 3/7 if possible.

TC: MUSAG hasn't talked about that. MUSAG is working between now and March to help figure out who should be involved: ZAG, representatives from HT partner institutions, perhaps OCLC? They are trying to wrangle the charge in early March. Another "stay tuned" item. If HT is going to take over (responsibility for data maintenance), what does that mean for relationship with OCLC.

JR: There has been ongoing discussion in CDL/UMICH Zephir huddles about the possibility of making corrections to WorldCat records in order to get corrections into Zephir without waiting for individual partners to make corrections. Hope those discussions will be incorporated into working group discussions.

TC: MUSAG update highlights:

- There were four items on the MUSAG agenda. (see MUSAG February meeting notes)
    - What should be the metadata policy regarding correction and redistribution? The Bibliographic Metadata Management Policy subgroup is being formed to address this question. They started with documents written by CDL a couple of years ago (*KS note: The report Tim mentions assessed how other like entities,*

such as DPLA, Europeana, have developed metadata use policy and making recommendations for a path forward should Hathi choose to go in this direction.). They asked if decisions made regarding the content of the hathifiles and the OAI output records were influenced by OCLC's policies. John Wilkin told them if hasn't been consistent. Current policy is what's developed over time.

JMO: Does that policy review influence data rights, e.g. OAI? There have been changes in OCLC policies over last few years, becoming more liberal.

TC: Mike is working with OCLC.
HT has shared entirety of our metadata for public domain materials with DPLA - we've done this with knowledge that DPLA treats metadata they've received as CC0.
The goal of group moving forward is to understand what's happened and then propose a consistent approach that can go through approval channels.

MUSAG's Environmental Scan Working Group met earlier this month; has a draft survey of what metadata exists, where it is, gaps and breakdowns between systems, etc. This should be shared in early March.

KS: Zephir Updates

- Brief, since had extensive knowledge share at last meeting (both the 2015 Q4 reporting and 2016 Q1 roadmap are available on the ZAG site: https://sites.google.com/a/umich.edu/hathitrust-zag/zephir-documents---backgroun).
    - This month, team has been involved in HT searches for Director of Services and Operations and two Program Officer (for the U.S. Federal Government Documents and Shared Print Monographs initiatives) searches. With all of the candidates for PO positions, had interesting and lively discussions about the role of bib metadata in those projects. Likely a role for Zephir data in supporting the initiatives.
    - The team is on track with work that had been plotted out for Q1. Currently planning work for clustering issues - matching algorithm fix was done in fall and now a plan is in place to fix incorrect clusters in the system that pre-date the fix.
    - Successfully shared item level API with HTUSWG Bib Corrections Group to support their work.
    - Migrating a (internal) data interface that has not seen regular use, but might have future use for Bib Corrections, etc. to AWS.
    - Will be meeting with CDL UX folks to map out how they gather data about the submissions process, learning from metadata contributors: how it works, issues, roadblocks, etc. Once compiled will help with planning adjustments to workflows and any associated development work.

Policies and Procedures Work

- Procedures:

KS: CDL Zephir operations team feels that using the pre-ZAG procedures is working pretty well. Requests/issues that come up generally would take less than a week so are addressed internally. That said, could the ZAG be ready to affirm the pre-ZAG document (i.e., remove “pre-”)? Are there other steps we could take to resolve potential ambiguities in how data and functionality requests are handled?
JMO: Are we to incorporate our work so far into the pre-ZAG, or just use the pre-ZAG as is?
Kathryn: That’s a question; do we need to re-write, or can we use the existing document with only minor adjustments (e.g., removing “pre-”)? Do we need more explicit, less ambiguous procedures and definitions? Do we need be more explicit in how requests come to the ZAG (beyond the two criteria in our charge: New features or service enhancements that 1) Have strategic impact on and/or implications for the broader HathiTrust community and/or 2) Require resources beyond the current allotment for running Zephir.)?
JMO: what does the Zephir operational team think?
KS: Pre-ZAG procedures as written seems to work well. Yet to encounter a case that would require bumping up to the ZAG.
RR: It’s hard to write policy speculatively. Let’s call it a “beta”, wait to see if anything turns up that tests the cases that haven’t turned up so far.
CCN: Maybe put a review date on it.
JMO: Agree with Ryan. Does Kathryn want to take a stab at updating and give us a chance to review whether the result is something we can sign off on?
KS: I’m happy to work on adjusting the language, etc. to come up with a version we can sign off on. Also, maybe it’s best to wait on any big changes until a new DSO is onboard.
JR: Yes on waiting for DSO. Also, it’s worth recognizing that there needs to be some level of ambiguity, working room for discretion.
KS: Yes, we’d not want to paint ourselves into a corner. I’ll aim to make updates to the current document and send it out for approval prior to the next meeting.

- Policies:

Some discussion between JMO and TC re: metadata, rights determination, etc. JR provided some background info on differences between bibliographic rights determination vs. other rights reasons included in rights database determinations.
TC: This is important information for MUSAG.
KS: I’ve for some time thought that it could be useful to do a Zephir presentation; in the context of this newly proposed working group, it feels like a good time to present to both the MUSAG and ZAG an overview of how corrections are currently handled, highlighting operational issues and impacts so that both groups are well-positioned to engage in a review of the current policy and processes. I’d recommend doing that in April in an off-cycle meeting that members of both advisory groups could attend.
JR: Should include parts that are outside of Zephir (bibliographic rights determination, hathifiles, catalog, OAI, etc.).

KS: Who would be best to include, Tim (Prettyman)?
JR: That would be good.
TC: From a timing standpoint, MUSAG would like to finalize the charge for this new working group sooner rather than later. Would be helpful to give information about workflows soon, and to provide comments on draft charge before March 10.

KS: I've been combing through our previous work on assessing procedures and policies for Zephir. There have been multiple discussions of whether Zephir-specific policies/procedures are needed and I'm hoping we can resolve this question for now - do we see any potential for policy that would be Zephir-specific?
JR: Only thing I see would be requests for data that would come only from Zephir (prioritization/resources). Is there enough demand to justify a policy?
TC: Who handles what requests for Zephir/UM/HTRC datasets - from outside, no-one knows what would go where.
KS: In cases of external requests, we might need to resolve this. Do we need an interim policy?
JR: I'd say wait for DSO - that's where triage, etc. needs to happen.
KS: That seems a sound suggestion. Are there any other policy issues that we could identify that wouldn't involve our working with MUSAG to address?
TC: Waiting for a DSO makes sense. [Tim said something else here, but the audio was breaking up and I didn't get it.]

New Business

KS: I've thought about discussing the structure/frequency of our meetings, but am now thinking we should leave things as they are, especially as we continue working with MUSAG and get going with the new DSO.
KS: One area of new business is establishing a watch-group that tracks emerging/changing metadata standards and formats (e.g., the availability of linked open datasets, RDA, etc.). Is there interest among the ZAG in pursuing this?
TC: Of interest, but may be premature. Not sure it's as critical as other work we're doing. Not a big rush.
JMO: Do we know of member libraries thinking of changing their underlying formats at this point?
KS: No, we haven't directly gone after that information.

ACTION: If you are interested in being involved in a bibliographic metadata standards watch-group, let Kathryn know by email.