

Zephyr Advisory Group
April 20, 2016

Present:

Tim Cole (recorder)
Gary Charbonneau
Todd Grappone
John Ockerbloom
Jon Rothman
Kathryn Stine

Absent:

Chew Chiat Naun
Ryan Rotter

Agenda:

1. Zephyr Updates

- [Zephyr 2016 Q2 Roadmap](#)

KS: Much of what's in this quarter's roadmap is continuing ongoing work. The Zephyr team has implemented processing changes at load time to better cluster records for volumes representing the same title, and are now preparing to remediate legacy clustering issues. Zephyr clustering uses OCLC numbers and local bib system numbers (which cluster multi-vol sets and serials) as match points. There were retrospective clustering issues in the HT db already, and the Zephyr team is beginning with low-hanging fruit. In some cases contributors may not have sent all records for a given serial (or MVM) title where their bib number has since changed.

The team is interested in better tracking process provenance data. Our primary tracking now happens at the file level, e.g., at the point of load and we'd like to move towards a more event-based approach that tracks more information at the record level to help with troubleshooting, etc.

We're looking at sharing more through Web-based delivery options (e.g., now that we've moved to AWS we are looking at S3 as a means of providing Web-based access to the reports we generate). This reports are generated post-load. Looking at whether Web-based is better for contributors and other consumers of our reports.

We continue to make improvements to metadata processing, including leader processing, reloading of records. We're also continuing to cross-train Zephyr staff (added a new staff member). Following this quarter probably will no longer highlight cross-training in our road map as this will be folded into business as usual.

We're planning on surveying our contributors to understand: what are their processes, what are their pain points, how useful is the reporting we do? We had an effort yesterday to brainstorm what should be in such a survey and we'll bring this planning to our HT/UM operations colleagues on next call before finalizing and conducting the survey.

Plans are in place to better coordinate with UM regarding how UM and Zephir are tracking administrative metadata values that support both metadata and content submissions.

As new staff join HathiTrust, Zephir will be working with them to get them oriented and up to speed.

Feel free to ask questions later (by email or otherwise), and/or ask about last quarters accomplishments.

2. MUSAG Updates

NB: No April MUSAG meeting yet (trying to schedule).

TG: There is a working group drafting a report on the environmental scan that was done earlier this year. But nothing new to really update at this time.

KS: Looking forward to seeing the environmental scan. How will this report be shared?

TG: Once ready, we'll try to give ZAG an early peek, but main distribution will be through the HathiTrust PSC.

KS: We talked about the potential benefit of a joint ZAG - MUSAG meeting. Anything to plan for yet?

TG: Probably question for the larger group.

TC: Metadata policy work is in progress and it'd be good to talk about those. Good to get early feedback from ZAG on that work if there is anything to propose.

JR: It'd also be good to get feedback from HT/UM.

KS: ZAG could benefit from seeing the environmental scan and see progress on metadata policy proposal(s). We could use these as a guide for what and when might be good for larger group interactions on one of our calls. Lots of meaty overlap between ZAG and MUSAG and Ops.

ACTION: KS and TG will stay in touch about when would be best time for joint call.

3. Policies and Procedures Work

- Review, discussion, and confirmation: [Zephir Request Procedures DRAFT document](#)
KS: Is everyone able to access the draft? (no problems identified.)

KS: It's been circulating for a while, but let's walk through it. Not a lot of major changes from what we've discussed before (e.g., procedures for evaluating requests, prioritizing, etc.). But have clarified where ZAG would be brought in to comment about requests for data and/or functionality. The biggest change from the previous iteration of this document is that I've added a preamble to provide more context and relationships with other HT entities. For example, prioritization schedule established on a quarterly basis with MF, ZAG, UM Ops, etc. - this is pulled from other documents (the Zephir SLA, ZAG charge, etc.). We note where requests come from ZAG, Zephir, MF, etc. Do people see this as the right scope? Any questions?

JO: I like the last part [of the preamble]. Clarifies that memberships can make requests, etc.

KS: Yes, but the routing of how requests come in is still not entirely settled.

JR: Should include UM Ops staff explicitly here as well.

KS: Yes, that's an oversight - corrected this in the draft.

KS: Let's talk about the next (3rd) paragraph about what the Zephir Ops Team addresses. I was trying here to combine an approach to both data and functionality. That both can require development work.

JR: Would it be helpful to indicate that this would cover data requests "outside normal operational workflow"?

KS: Thanks, yes - made suggested changes.

KS: Moving on to the 4th paragraph of the preamble. Since we are obligated (by the SLA) to address how we handle data error reports, we say here that we handle them like any other data request. Any feedback on this approach?

JO: Does this deal with systematic data corrections as well as one-off data corrections?

KS: Current metadata correction policy doesn't really address this - a little in the final paragraphs. So we don't have workflows in place to implement systematic corrections, though the last paragraph of the current metadata correction policy is intended to leave the door open.

JO: Could imagine wanting to do this for bulk updates, e.g., for copyright updates, etc.

KS: In terms of setting the stage for how requests are handled, does this draft touch on all the areas we've discussed?

One specific question is, do we need to talk about how this document will be reviewed and updated? So for example as we better define our workflows and request routing, we may want to add some of these details into this document. For time being this document should capture where we are, but we need a suggested timeline for review with an out to do sooner as warranted.

DECISION: A review timeframe for the procedures document should be specified, 2 years sounds good to all with an option to review sooner.

KS: Added language to say this.

In terms of having an option to request an early review before two years, we need to make sure we identify who can make such requests, and maybe we need to say what would be a qualifying event to trigger that review. I'd imagine this would be the ZAG, Zephir Manager, and HT Repository Manager. Added these to the document.

JR: How much this opens up may depend on the future pathway and gatekeeper for addressing requests.

TC: If we anticipate how requests pathways may expand, it is good that we acknowledge that we are in many cases wanting to respond to contributors and others in the HT ecosystem.

KS: Will work on this item (end of #2) to refine wording. Our goal today is to reach general consensus on about 90% [so that we're prepared to confirm the document the next time we meet].

JO: Is this more a comment on the document than an edit?

JR: Could go in as a parenthetical?

KS: On to #3. 3.b is new and describes when a request needs to be bumped up to ZAG, so that ZAG and Ops Team can work together. Language is pulled from ZAG charge.

Do we need to provide examples (see previous notes)? It might be better to leave it more open rather than hem it in.

JR: Yes, and I don't think it would be worth the time and effort to generate examples now.

KS: Okay. As we implement we may find examples.

KS: Last 2 paragraphs of #3. Acknowledges need to consider time required and impact beyond Zephir before doing work.

JR: Is it clear that judgment will be exercised whether a request should really be done at all?

KS: Made some edits to make this more clear.

4. Welcoming Sandra McIntyre, the new HathiTrust Director of Services and Operations

KS: Sandra begins first week of May and will be ex officio member of ZAG. Hopefully will be able to join for our next meeting.