

Zephyr Advisory Group
August 19, 2015

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

Apologies: Gary Charbonneau
Todd Grappone

Announcement: In the interim before a new HathiTrust management position is filled that will take on an ex officio role on the ZAG, Mike Furlough or Angelina Zaytsev will join future ZAG calls as they are able.

1. General Zephyr questions

John Mark Ockerbloom asked if there is version control in Zephyr. Kathryn confirmed that yes, we can access previous versions of records.

Naun asked how the ZAG relates to the other HathiTrust metadata policy advisory group that is forming. Todd Grappone, who will be on both groups, may be positioned to address this. Jon Rothman noted that there is such a group forming, but we don't know details of their charge. His impression is that it will be concerned with more general issues about metadata throughout the HT systems.

Kathryn proposed that we could invite Zephyr team members, perhaps technical lead, Stephanie Collett, do a technical walk-through of the system.

John Mark Ockerbloom suggested that it would be useful to have someone on hand to answer feasibility questions. Ryan Rotter suggested that it could be useful to have that walk-through further down the line and Jon Rothman proposed that we can accumulate questions about Zephyr until we have enough to merit such a presentation.

ACTION: Kathryn will compile questions and determine when to invite the Zephyr team to an upcoming call, to present on/answer questions about Zephyr functionality, especially is it relates to the procedural and policy issues the ZAG is working on.

2. Zephir development updates

Kathryn reported on Zephir progress since the last ZAG call and will continue to include this as a standing agenda item so that the ZAG is apprised of work that happens as each quarter progresses (these reports supplement the Zephir quarterly roadmaps also shared with the ZAG).

The Zephir developers have moved to 3-4 week development cycles tied to the quarterly roadmap, and are just launching their second cycle as part of this new approach.

The biggest job for the team right now is to migrate to AWS, and they have successfully migrated metadata preparation processes and the item API auxiliary service. Zephir databases (dev, stage, production) had already been migrated earlier this summer, so the team is now working on migrating the core Zephir file system, record ingest and exporting processes, which should be done in September.

The Zephir team agreed with the UM/HT operations team and Mike Furlough on revisions to our security documentation, reflecting changes predicated by moving the infrastructure to AWS. Also with UM/HT operations team, the Zephir team agreed upon service expectations for the item API auxiliary service, which currently has one user and provides item-level record metadata.

The team will be working on improving the existing metadata corrections workflow and, to this end, have initiated conversation with Angelina Zaytsev on revising our corrections work and communication-flows.

3. Google site access, document sharing

Kathryn will ensure that all permissions are set on shared documents such that anyone with the link can edit. When you share a document on the Google site, please set sharing permissions so anybody with the link can edit.

Kathryn proposed this process for sharing meeting notes: the designated recorder will share notes with her and she'll review, and add information about Zephir as needed, before posting to the Google site. She'll request that ZAG members review the meeting notes prior to sending these on to Melissa Stewart at HathiTrust to post on the HathiTrust public website (on the ZAG page). The group agreed on this approach.

SLA 4.3.1-4 policies and procedures

- review Pre-ZAG Procedures for Zephir Data and Functionality Requests (no outstanding comments or questions at this time)

- report back on identified issues and options

1. Bibliographic data requests (Todd, Jon)

Jon Rothman noted that ownership issue affects everything else. There is an assumption that contributors “own” the data, but it’s not clear what all this entails. For example, do partners retain veto rights? We need clarification.

Naun asked if it is true that the contributing institution retains ownership of their data. Jon suggested that “ownership” may be the wrong word - the main point is that HT does not have the definitive record. Getting permission from contributors would be burdensome and probably unnecessary.

We could use clarification on what metadata can be shared. Kathryn noted that a lot of HT metadata is already shared to DPLA (the preferred base records for public domain resources), which shares contributed metadata as CC0.

Jon noted that much HT metadata is also available via APIs and partly via OAI output. Can we make this clear and explicit as a policy?

John Mark Ockerbloom suggested this would be a good idea, though it isn’t within the power of this group.

Jon observed that there’s quite a lot of complexity in the way metadata travels. It’s probably not realistic to expect that people making requests would understand where the parts of the request come from. We should have a central place/person at HT to field requests.

Kathryn asked if we could consider some rubric/tools/guidelines for routing or prioritizing requests, though Jon wondered if the volume of work would justify this?

John Mark Ockerbloom wondered which requests can be met by the API, and which need to be met by special requests. Could an item-record API provide self-service access to meet some requests?

Jon suggested that there should be a policy on who we will provide metadata output to, and what we will do for whom. We need input on what kinds of requests can be addressed on a basic operational level and which ones need a more involved response, e.g. an enhancement request.

2. Ad hoc reporting requests (John, Kathryn)

John Mark Ockerbloom noted that we could clarify the distinction between ad hoc reports and requests for bibliographic data, and also between basic requests and requests needing additional development work. It's possible that requests could come in for data we haven't been tracking - will we sometimes need to reconsider what data we collect?

Kathryn gave examples of operational requests (e.g., requests for all records selected as the preferred base). It seems that these should be within scope.

John noted that handling requests could depend who makes the request, and maybe who the intended audience is. Potentially some requests will be recurring and that may need a policy. Also, which systems will need to be tapped to meet the requests, and how should this be coordinated? Do any requests involve sensitive data?

Kathryn noted that some requests may need to be handled with a broader perspective than Zephir.

3. Error reports, updates, and corrections (Naun, Gary)

Naun summarized the main points from the document he and Gary compiled. Some highlights:

- Should we focus on just corrections, or value-added metadata enhancement as well?
- The current corrections flow is cumbersome, requiring feedback from the contributor; could this be centralized at HT?
- What is the relationship between the HT record and OCLC Master?
- How could/should metadata changes be propagated?
- Are there alternate clustering approaches?
- Address issues with duplicates (e.g. differences in serials cataloging)?

Tim Cole asked if there could be more systematic detection of errors. Jon Rothman suggested that further exploration of this approach would depend on HT having its own record. And there would need to be an investment in staff and software to do it.

John Mark Ockerbloom voiced hesitancy because HT has limited resources. Version control would make it easier to share work.

Naun suggested that we may also want to enable members to make changes across all of HT.

4. Addressing and implementing enhancements (Tim, Ryan)
We didn't get to discussing the document that Tim and Ryan compiled.

ACTION: ZAG members should add their comments to this document in preparation for our next meeting.

- discuss and coordinate next steps

ACTION: Kathryn will start a collaborative document about commonalities among the issues and options that the teams of two have surfaced, and request ZAG members to review and add to it in preparation for our next call (September 16, 10am PT/1pm ET).

ACTION: In preparation for our next call, please continue to consider:

- areas where we need clarification or more information
- if/how the Pre-ZAG procedures sufficiently address issues
- policy implications for each procedural/policy area