Zephir Advisory Group

September 16, 2015

Present: Tim Cole (recorder)

              Chew Chiat Naun

              Jon Rothman

              Kathryn Stine

              Todd Grappone

              John Mark Ockerbloom

              Angelina Zaytsev

Regrets: Gary Charbonneau

              Ryan Rotter

1. Zephir Development Updates:

Kathryn Stine noted that the Zephir team continues to work on migrating core processes over to AWS, working now on setting up the staging environment before moving production processes over later this fall. A new developer on the team, Jing, has been cross training on metadata preparation processes and the developers are updating technical documentation. The team is migrating git repositories to GitHub and working on integrating git commits into our ticketing system.

The team is also refining processes regarding configuration set up and validating filenames for contributed metadata. This is part of an ongoing, longer term plan to re-imagine the submissions process over the next year.

With Angelina Zaytsev, the team is looking at pain point and opportunities to improve the metadata corrections workflow, which involves the Zephir operations team, HT staff at UM, and members of the HT User Support Working Group (HTUSWG) working with bib corrections. There is likely a role for the Zephir Item API auxiliary service to play here and the team is starting to think about both the operations and policy implications of opening this up more broadly. The team is currently documenting use cases and will develop/confirm requirements for the HTUSWG, which whom we'll share this work as a next step towards ensuring we meet their needs and successfully introduce efficiencies into the workflow both on their end and for Zephir operations staff.

2. Metadata Policy, Strategy, Use, and Sharing Advisory Group (MUSAG) Launch

Todd Grappone (MUSAG co-chair) provided an introduction to this new advisory group, considering how this group will interact with the ZAG. The background on the MUSAG is that the HT Program Steering Committee have been talking about various aspects of metadata within the HT, including policy and sharing issues. For example, what are policies regarding use of HT metadata (which is deposited by member libraries)? The PSC thinks there probably need to be multiple groups looking at this, both at a practical level and at the policy level. There is intentional overlap across MUSAG and ZAG. Once a new HT DSO is hired, should have more complete overlap (4 positions will overlap total). With MUSAG there are strategic and policy issues that need to be addressed: questions about who owns the metadata, how does it corrected, how does it get shared. MUSAG will help develop that policy and will focus on strategic directions for metadata being used in HT systems. Their first call has not yet happened. There will be a 2-year review cycle to confirm that this structure is working for HT.

Questions:

Naun: Can we share agendas of the MUSAG?

Kathryn: It might be useful to have a standing MUSAG update item on our ZAG agenda. We may find that the ZAG is surfacing broader policy issues that MUSAG needs to look at.

Todd: There may be a learning curve as we figure out who does what between MUSAG and ZAG.

Kathryn: The Zephir team has been thinking about a lot of the issues identified in the MUSAG charge. Having more help with this going forward will be good, and there may be a need to revisit some of the current policy and operational decisions currently in place.

Todd: The MUSAG will conduct an environmental scan to identify the different pools of HT metadata -- bibliographic, holdings, rights, provenance, etc., and we need to clarify where this metadata resides and how it interacts.

Kathryn: We want to clarify how things will work in this new advisory landscape. The Zephir team will likely have questions going forward, and will identify and walk through sample scenarios to clarify where and how recommendations and decisions are made and how policy could be developed with both current and anticipated Zephir functionality in mind. As part of this process, we can surface questions that we can pose to both the ZAG and MUSAG.

ACTION: The Zephir Operations Team will share sample scenarios that illustrate potential points of engagement and coordination between Zephir Operations, the ZAG, and MUSAG.

2. Policies and Procedures for Zephir data and functionality requests

Topic: Zephir Enhancements - Issues and Options (Tim and Ryan)

Tim Cole reported back on the issues and options that he and Ryan Rotter identified relating to requests for Zephir functionality enhancements. Some questions:
- Who makes requests? Who can make requests (e.g., HTRC, ZAG, external folks?)? Is there a formal way to submit requests?
- How could we prioritize/order these requests? What kind of consulting, triage is needed to field these?
- How could we get feedback from users on the metadata exposure methods?
- How might enhancement requests streamline existing services?
- How to address migrations to new metadata standards, formats, rules?
- What external factors should/could be taken into consideration when addressing requests?
- Should we include scenarios in order to illustrate the path for recommending how different types of requests might be handled?

Requests may come in from the HT community (partners, contributors, management). We need an understanding of the pipeline for requests and how to evaluate them, assessing costs/benefits so that they can be put in priority order. We should consider what level of resource requirements will escalate requests to be addressed by the ZAG and not just the Zephir Operations team.

Kathryn acknowledged that requests for data or functionality may rarely be satisfied by Zephir alone. Zephir is integrated into the work and data flows of other HT system.

Jon Rothman provided this example to illustrate: impacts on users of the APIs - scoping who might need to address a request is tricky as the APIs are outside of Zephir, so Zephir would be implicated if the content originating from Zephir isn't available.

Tim Cole observed that there is an interesting dependency for bib metadata exposed via HT: the bib API and hathifiles are provided outside of Zephir but use Zephir data, and, to get access profile, a user needs to go to the data API. Users need to go to many places to access information about data. It could be useful for the ZAG to consider user-facing issues.

Todd Grappone identified his role as ensuring that issues that arise during Zephir discussions get onto the PSC radar.

Naun asked about data enhancement?

How could this be addressed in the new advisory group structure?

Todd suggested that this issue could be addressed by a MD sharing policy.

Kathryn brought up the idea of providing a Zephir presentation and/or Q&A session to illustrate how implications arising from these policy issues relate to operations (what the current system design affords and what might need to be considered for future development). It might it be worthwhile to have a walk through by the Zephir team for ZAG and MUSAG so the advisory groups can better understand the current state of the art and anticipated enhancements.

Jon suggested that we could also at same time talk about sources for the metadata disclosed via HathiFiles, Bib API, Data API.

ACTION: Kathryn to work with ZAG and MUSAG on setting up an opportunity for presenting on Zephir.

Topic: Commonalities

Kathryn suggested the group look at commonalities already surfaced as starting point for moving forward with development of policies and procedures documents. The pre-ZAG document can serve as a starting point. To move us forward, the entire group can identify commonalities together, now and on the next call, and then allow the sub-groups to return to each respective area for specific policy or procedures needed.

Kathryn shared the following document that brings common questions, procedural and policy issues together, confirming this could be a useful step in getting us towards recommended policies and procedures documentation. The team can work on this collaboratively up to the next call:
https://docs.google.com/document/d/1pJwrASi3T0pus5x7-oAHjLij0FWnfspA7VBVqavQ3Zk/edit

Jon suggested that it seems unrealistic for partners to know where to route requests that span the system. There may need to be more centralized triage. Requests can be fielded centrally and then distributed appropriately.

Todd noted that there is a new DSO in place, it may be difficult to clarify.

Kathryn asked how HT-wide development requests are being handled now?

Angelina Zaytsev said that a lot come through her. Requests get added to lists, added to the HT ticketing system, and the lists are reviewed quarterly.

Kathryn confirmed that to-date requests to Zephir typically from internal HT sources.

Angelina shared an example request that has recently come in, relating to Zephir: to add particular linked data (VIAF data?). This request is queued and not yet handed off to Zephir. HT would do preliminary prioritization to address it. Once a new DSO is on board, assume that HT will review a lot of its procedures.

Angelina noted that if Zephir starts including linked data (e.g., URIs), this would have implications for the HT beyond Zephir - for example, for the bib-search user interface.

Kathryn suggested that this could serve as a good example to make concrete how interrelated the HT systems are. Zephir is not a silo, so we need to think about how procedures in Zephir interact with larger HT procedures.

Jon asked if any of us could conceive of an external request for enhancement in Zephir functionality that wouldn't impact more than just Zephir?

Kathryn replied that Zephir is so tightly connected to other HT systems. This is a good question that highlights the need to have a central point of contact for incoming requests.

Kathryn noted that another commonality across the issues and options that ZAG members had worked in is was who can make requests, and how who makes the request might influence prioritization? Again, this may require talking to PSC or other part of the HT governance structure?

Jon acknowledged that it may be hard to make Zephir policy about these cross-cutting issues.

4. Next Steps

Kathryn asked, where do we need more clarification? Where do we have more questions?

ACTION: In preparation for the next call, ZAG member will annotate the commonalities document, and also identify what clarification/more information we need to move forward in drafting procedures and policies.

On the next call (October 21) we can address these common issues and then after next call break back into sub-groups.

Kathryn noted that the ZAG will need to provide the HathiTrust Board of Governors, the Executive Director and the Program Steering Committee with a quarterly report on our activities.

ACTION: Kathryn will provide a draft quarterly report for the ZAG to review, improve and sign off on before sharing this with the Board of Governors, Program Steering Committee, and Executive Director.