

Zephyr Advisory Group
July 15, 2015

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

Absent: Todd Grappone

1. ZAG transitions:

Jeremy York has left HathiTrust for another position; his last day was June 23rd.

Kathryn Stine has joined CDL and this group. She was formerly at Zephyr and has returned as Zephyr product manager after being at Berkeley.

Kathryn listed the other people who work on the CDL HathiTrust Metadata Management System team:

- Kathryn Stine

Product Manager (provides management and leadership for the team, serves as primary point of contact for Zephyr communications, and convenes the ZAG)

- Stephanie Collett

Technical Lead (provides technical leadership for the team, works primarily on developing, maintaining, and updating code that ingests bibliographic metadata to the Zephyr database and filesystem, and provides exports to HathiTrust systems)

- Mike Thwaites

Senior Developer (works primarily on bibliographic metadata submissions, develops, maintains, and updates metadata prep code and reports on record loading activities)

- Jing Jiang

Developer (will serve as technical back-up support for the metadata prep code, code that ingests records to the Zephyr database and filesystem, and code that provides exports to HathiTrust systems)

- Dana Jemison

Metadata Analyst (handles new and ongoing bibliographic metadata submissions and corrections, serves as front-line point of contact for HathiTrust bibliographic metadata contributors and members of the HTUSWG bib corrections team)

- Colleen Whitney

Discovery and Delivery Technical Manager (manages all Discovery and Delivery technical staff, including all Zephir technical staff)

- Patricia Martin

Discover and Delivery Director (manages the Discovery and Delivery team, comprised of technical, product management, and metadata analyst staff)

Some of the above people do not work full-time on the project.

Additionally, staff working on behalf of HathiTrust at the University of Michigan meet regularly with the Zephir operations team, including Jon Rothman, Tim Prettyman, and Angelina Zaytsev.

2. General Zephir questions:

When records are submitted by multiple contributors describing the same manifestation, one of the submitted records serves as the base record in HathiTrust catalog, determined by a scoring algorithm

(<https://docs.google.com/document/d/1A6lq4bQBvHmqdz3Ab-CiqPpNqcXILe1bW5JCeNuW4x0/edit>). The contributing institution is responsible for making any necessary corrections to this record, in response to requests or as part of their own local metadata improvement processes.

Metadata updates outside of those requested via the HTUSWG bib corrections workflow are contributor-driven. Zephir loads any and all updates from contributors,;there has not to-date been an effort to actively request comprehensive updates on a regular schedule.

Batch corrections of records is something that Zephir would like to explore, though current assumptions and policy regarding metadata stewardship may be “collision” with this approach.

From the HathiTrust Bibliographic Metadata Correction Policy: “HathiTrust bibliographic metadata management assumes that the definitive bibliographic record resides in contributors' respective catalogs and/or OCLC.” (https://www.hathitrust.org/bib_metadata_correction)

Per the Bibliographic Metadata Correction Policy, Zephir is empowered to make changes that would put items into limited view for copyright reasons.

OCLC # is the primary mechanism for clustering records (along with local system numbers for clustering the records for particular volumes of a work).

Rights data is not kept in Zephir, but rather in a HathiTrust rights database. Rights data (attributes, reasons, etc.) is described here: https://www.hathitrust.org/rights_database, and can be accessed via the HathiTrust Data API: <https://www.hathitrust.org/data>. Automated bibliographic Rights Determination is described here: https://www.hathitrust.org/bib_rights_determination

Currently, local notes and other local metadata from specific item records is not shared in the HathiTrust catalog, other than that included in the base record. Zephir does maintain local data for each item in HathiTrust. Local data could be of use in identifying copies of rare books, for example.

Zephir staff is sensitive to the end user impact that changes to clustering ID's (CIDs), which are used as the record ID in the HathiTrust catalog, can have (e.g., broken links) and works to minimize their changing whenever possible.

3. ZAG deliverables (section 4.3 of HathiTrust Metadata Management System Service Level Agreement):

ACTION: We will split into groups of two to work on policies and procedures in four areas. Draft policies are due by November. Kathryn will serve as a consultant for the teams as necessary. Each group should come up with issues and options for discussions at the August meeting. The four areas are as follows:

1. "The Zephir Manager will, in consultation with the Zephir Advisory Group, draft policies and procedures for receiving and handling requests for bibliographic data, including ad hoc reports from Zephir." Volunteers: Todd, Jon.
2. "The Zephir Manager will, in consultation with the Zephir Advisory Group, draft policies and procedures for receiving and handling ad hoc requests for reporting against the Zephir metadata management service." Volunteers: John, Kathryn.
3. "The Zephir Manager will, in consultation with the Zephir Advisory Group, draft procedures for handling error reports related to the data. Data updates and corrections will be made according to the HathiTrust Bibliographic Metadata Correction Policy." Volunteers: Naun, Gary.
4. "The Zephir Manager will, in consultation with the Zephir Advisory Group, draft policies and procedures for addressing and implementing enhancements to Zephir. These policies and procedures will include a prioritization schedule and guidelines for the level of changes or fixes

that can be or should be implemented by the Zephir Manager without consultation with the Zephir Advisory Group.” Volunteers: Tim, Ryan.

4. Discussion of the Zephir Q3 roadmap:

The group agreed that the format and level of detail of the report was about right.