

2021-05-27 Whitman College Migration Pilot Check-in

Attendees

- Alan Stanley
- Amy Blau
- Dana Bronson
- Dara Virks
- David Wilcox
- Kun Lin
- Paige Morfitt

Agenda

Topic
Stand-up
<ol style="list-style-type: none">1. What has everyone been working on?2. What will you be working on over the next couple weeks?3. Are there any blockers that are preventing you from getting work done?
Sample Data issues and questions
Priorities and areas of focus
<ol style="list-style-type: none">1. Timeline for remaining work2. Possible cut-over date
Wrap-up and next steps

Notes

1. Dara
 - a. Some issues with configs being overwritten on staging
 - i. BD doing an update so this won't happen anymore
 - ii. Planning to freeze staging this afternoon for the update
 - b. Don will be working on the Access Control work next week
2. Alan
 - a. Ingesting newspapers from laptop is very slow (about an hour per issue vs 3-4 minutes on staging)
 - i. Ingesting on the server fails to create OCR. Alan is investigating.
 1. Hypercube grabs the image and sends to Tesseract, seems to be timing out here
 - a. Nginx may be timing out at 64 seconds, Alan will try doubling or tripling this number
 - b. Could be due to server load
3. Amy
 - a. Created a spreadsheet of sample data issues
 - i. Title length is an issue - there is a 255 character limit
 1. Whitman team will discuss
 - b. Checklist for spreadsheets
 - c. parent_id vs member_of
 - i. parent_id is used for ingest, member_of is used for relationships in the repository
 - ii. one of the fields has a space in the spreadsheet, which workbench fails on
4. Timeline
 - a. Dara to investigate when prod site will be ready for production ingests
 - b. Should be able to start production ingest within the next 2 weeks
5. To discuss
 - a. VTT transcripts

Actions