Meeting minutes 9/10/14

Present

Trish Rose-Sandler, Mike Lichtenberg, Joe DeVeer, Paul Silverman, Doug Holland, Mike Blomberg, William Ulate, Marty Schlabach, Holly Mistlebauer

Absent: Connie Rinaldo

Agenda

*Cornell update (Marty)*

4900 pages shipped to scanning vendor, approx. 6-8 weeks until images return and can upload them to IA

Getting ready 10k pages for shipment #2, will send out after first shipment comes back

Need Macaw to work with – contact Joel Richard for support. Just need an account from Joel. Joel has some documentation. Joe can help, he is using now.

Met with seed and nursery catalog group with NAL – NYBG, Cornell and NAL. Looking at cataloging differences to minimize duplication. Holly came up with approach to compare fields. 4 fields: Firm name, holdings, local ids, etc.

NAL has digitized over 10k items which are in IA and BHL

Can probably meet the December deadline – not sure of how much up by then but something.

Review Cornell deadlines:

15k uploaded by Aug 2014
25k uploaded by Dec 2014
30k uploaded by Mar 2015

 *NYBG update (Paul Silverman )*

71 volumes 5700 pages digitized but not uploaded to IA yet. Lots of problems with Macaw – freezes up halfway through, no programmer right now to look into. John’s replacement will start in 1-2 weeks. Similar job as John’s

Review NYBG deadlines:

5k uploaded by May 2014
5k uploaded by Dec 2014
5k uploaded by Mar 2015

Should be able to meet December deadline

*Harvard update (Joe)*

ALA portal - how much complete? Of 2000 pages 1441 are transcribed

FromThePage site – how much complete? 144 transcriptions completed. Will have staff members to focus more on FromThePage for now

Any additional outreach for tools? None yet but will talk to Grace about options. Connie will get in touch with head of ornithology dept to see if we can utilize grad students.

When does Tiltfactor need transcriptions complete by? Trish will look into

Review Harvard deadlines:

* Investigate, select, implement transcription tool (Dec 13-Apr 14) done
* Transcription of field notebooks, horticultural catalogs (Apr 14-Mar 15)

Mechanical Turk implementation? Will look into soon. Probably will have them work on FromThePage tasks.

MOBOT

Scanning (Doug and Mike B.)

Continuing to scan seed lists – 70% complete. Are they cataloged? Yes. Cataloged by source? Yes. Cornell would love to see some records examples. Mike B will send on to Marty.

Starting to generate some OCR outputs from Tesseract

Gaming (Trish)

Game design moving along well with Tiltfactor. Sent them samples of our OCR both good and bad and from that they have a sense of the types of issues that their software will either need to compensate for and guidance we will need to give players.

They’re first deadline for them is end of September to develop 2 prototypes for the games/tools that need to be developed. We’ve seen a rough prototype of the first gamers for altruistic players. The second game/tool we have to make a decision by the end of this week as to what we want them to create.

Options for second game/tool

Create a game for gamers (BHL would need to manage a tool that generates the coordinate data for Tiltfactor)

Tiltfactor creates a tool to generate the coordinate data (manual drag and drop tool that has users telling us where the word goes on the page)

Phone call with developers of a tool called TILT2 developed in Australia that looks promising for automating the coordinate information. Sent some samples today of pages that had been transcribed within FormThePage William Brewster’s. Desmond will upload them to tool

Other - did everyone have a chance to fill out the Crowdsourcing Consortium for Libraries and Archives survey? https://www.surveymonkey.com/survey-thanks/?sm=ceLXLKyV%2b%2fCbRf4BV7O2Bf8ij4s%2fpNhQ8lSmTwJ7S8U%3d