Meeting minutes 6/11/14

Present: Trish Rose-Sandler, Mike Lichtenberg, Joe DeVeer, Marty Schlabach, Holly, William Ulate, John Mignault

Absent: Doug Holland, Mike Blomberg, Connie Rinaldo

Updates

Cornell

Scanning (Marty) Holly MIstelbauer, Mann Library IT area, Assistant Head of IT, here to Faciliate any IT assistance for Marty for the project.

Not as much as we would like – job interviews, Marty time off. Working on setting up workflow for scanning for seed & nursery catalog. Coordinating with cataloging group, Identified 1350 catalogs in 1st batch (40k pages). Focused on grape Catalogs and only those not being done by NAL and NYBG.

NYBG

Scanning (John)

Got barcodes for Paul to start with Macaw this week. Send out shipment to IA yesterday. Using a credit NYBG got from IA when they did a shoddy job previously.

Harvard

Transcription (Joe)

BHL blog post announcing our transcription tools ready for William Brewster notes. – ALA and FromThePage. Both have duplicate material. ALA transcription has started with some folks from Harvard and seeing some volunteers from ALA also

Joe has to validate the volunteer work – done according to our instructions. Will add Status of “validated”. If there are problems Joe can fill out a form with suggestions

FromThePage – no volunteers yet.

Other ways to promote? ALA will do a blog post about it, Paul Flemmons suggested looking for crowdsourcing listservs, could notify volunteer coordinators at our organizations. Ornithological clubs, local birding listservs. Could use a short writeup about the value of this collection and transcription

There are guidelines on both sites to help volunteers.

FromThePage seems to have some save issues – we should all do some transcription ourselves and identify any bugs. Let Joe know if tutorial needs improvement.

When will they be added to the BHL portal? TBD. Each page will have different areas needing correction. Tool to handle different versions of OCR will be at page level.

ALA site – 10 journals there but only 3-4 activated. IF there’s one not activated and you want to do it let Joe know.

MOBOT

Scanning (Doug and Mike B.)

Mike says “scanned approximately 3500 pages so far alone since the beginning of May, and one of our other scanning technicians has been helping out with some seed list scanning as well. I’m not 100% sure on a number on his end (he’s keeping a separately tally that I will combine with mine later on), but I would guess he’s scanned about another 1,000 pages…. So we should be right at or past the halfway point of our promised goal of 8,000 pages. The great news is that a couple of weeks ago we realized that these seed list collections can have a deceptively “large” amount of pages (in comparison to what we originally thought at least), and we’re only about 20% of the way through of what’s been pulled for scanning in this project. I could very easily see us ending up with twice as many pages as promised (possibly even upwards of 20,000 scanned pages).”

Gaming Bids (Trish)

We had a deadline of May 22nd for the gaming bids. Received only 3 bids (was hoping for 6-7 from companies who had expressed interest). Extended the deadline by 4 days but still no more bids. Expect the low response was partly due to our limited budget of 110K for the game.

Good news is of the 3 bidders at least two are very strong bids! Exocet (German based company with 140 employees and have worked with commericial and non-profit groups). Tiltfactor Labs (based at Dartmouth, research oriented, done a lot of research on gaming motivations, created a lot of crowdsourcing games)

Review taskforce (Trish, Willliam, Mike, John, Joe, Connie, Jiri Frank) will meet weekly until the June 23rd deadline for notifying a winner. Have had 1 meeting where we established our criteria we would use to rate them on.

* Did they answer all our questions from the proposal?
* What are the pros/cons of this company?
* Can they work within the budget?
* Do we know the IP rights for their products?
* Do we know if they follow scrum/agile methods in their development process?
* What platforms do they propose to build on and what are pros/cons for BHL beyond the life of the project?
* Do they have successful games they can point to? How many downloads? How long did players stay engaged in the game?
* Would the game they develop for us have any player costs?

BHL Architecture for managing OCR and Transcription versions (William) Not much updates. First step to automate the review of QA of texts – non-alpha characters does it contain? Tool found from hackathaon we could use. Looking at current OCR and where to do the comparisons? Accuracy score – run differences through the game of words. William will work on running an item through whole process.