**Purposeful Gaming monthly partner meeting 3/5/14**

Present: John Mignault, Joe DeVeer, William Ulate, Doug Holland, Marty Schlabach, Connie Rinaldo, , Trish Rose-Sandler, Mike Lichtenberg, Mike Bloomberg

**Agenda**

* Summary of face to face meeting at Webwise (Trish)

Webwise was a worthwhile conf for Trish - Good mix of unconference and more formal presentations

Sessions on transcription and crowdsourcing that were useful and had a meeting with Mary Flanagan who was one of the plenary speakers and heads of Tiltfactor Lab at Dartmouth. She said they would be willing to consider the RFP.

Face to face meeting – just 5 of us: Trish, Joe, Mike Blomberg, John, Marty. Talked about our roles on project, did round robins, reviewed the RFP and spent most of our time reviewing and discussing the transcription tools.

Both Marty and Mike B. felt it was worthwhile. Size was good and easy to connect. John said it was his 5th year and it was much improved from last year but attendance was down. Marty and Trish had conversation about Users tagging cultivar names in seed nursery catalogs, Tom from Boston public library overheard discussion mentioned he has funding for doing digitization projects across state. They could contribute their seed catalogs to our project.

* Reminder of project page on internal BHL wiki

<https://bhl.wikispaces.com/Purposeful+gaming+IMLS+grant+project+page>

Is Cornell and NYBG putting their meeting minutes there? Marty will post those but needs to contact Trish first for how to do it.

We have a writeup in the InBrief section of DLIB magazine this month! I added this to our media coverage.

* Updates from Cornell, NYBG, Harvard, and MOBOT

MOBOT

Doug/Mike B. – talked with cataloger who has seed lists all cataloged and ready to go and started on scanning part of those. Copyright issues – mostly cleared but need some due diligence procedures for some. Doug wondered if there are codified guidelines somewhere for doing due diligence? Connie says no. Marty says up to institution doing digitization to decide how they want to interpret law. In UK is based more on death date of author rather than publication date. Marty wondered In case of seed nursery catalogs who is the author -corporation? He will seek advice on non-US copyright law. Doug says most of theirs are published outside US. andpublished by places that are happy they are available (botanical gardens and universities). Challenge is finding person at institution who you should contact to clear permissions.

Trish, William, Mike – RFP draft is finished and having different folks review it before going out (hopefully week of March 17th). Trish will be contacting the gaming companies we have identified to gauge their interest before sending to them. Just got word from IMLS that we are not limited to US based companies only!

Continue negotiations with Harvard over what type of contract they will write with us. This may delay Harvard’s ability to hire their transcribers in April.

William looking into what others have done before with OCR, particularly some projects in Europe and even previous experiences with BHL. Also, downloaded the Python code for Darwin Score from Jason Best and started looking at it. <https://github.com/idigbio-citsci-hackathon/darwin-score/blob/master/README.md>

Darwin Score is a method to evaluate text representing a natural history bio-collection object. Scores calculated in this process can help provide a rough evaluation of the quality of the text, whether it is generated by human transcription or OCR.  It was initially developed for labels and annotations applied to herbarium collections, but there’s no reason why it can’t be used for our content.

[This process will check the text for words including taxonomic names, collector names, annotator names, location names, common abbreviations and other expected words (all stored in purpose-built dictionaries) and patterns such as dates, numbers, and geo-coordinates. The text is given a score based on the number of the matches found.]

Cornell – Marty said they have bi-weekly meetings with NYBG , preparing list of firms plan to digitize, coordinate with list from NYBG to minimize duplication. Still talking with NAL, they have 3900 volumes in seed nursery collection in BHL. Has over 4500 scanned and in IA. Missing tag in IA that prevents us from harvesting all of it. Student intern working with Marty. Meeting with catalogers to talk about their needs going forward, will have to add MARC records to catalog to accompany images. Building citations for publications dealing with seed industry. Intern is building on the bibliography and putting into Refworks. Can share bibliography eventually with BHL site. Hoping to start scanning soon. Learning how to use Webex for conference calls (screen sharing). We could use Cornell webex account for these meetings if needed.

Harvard – exploring transcription tools, narrowed down to 4 options. TPEN and Transcribe Bentham were eliminated due to messy code and UI learning curve.

SI transcription tool – code isn’t available and Carolyn looking into whether they can host. Left with FromThePage, ALA, Transcribr. Formed a small task force along with TAG folks to make final decision.

NYBG- John not much to report. Working with Cornell to coordinate scanning material. Paul Silverman hired and scanning. Marty asked how is Macaw testing going? Doing pilot of Macaw. Working to streamline process. Scanned material masters are uncompressed and locally stored. Copying large files is slowing things down. May convert TIFF s to JP2 or use external drive for storage. Joe says they use the external drive approach and works well but their files are JP2s. Network may be part of the problem.