**PG Monthly Meeting 11/12/14**

Present:

Trish Rose-Sandler, Mike Lichtenberg, William Ulate, Doug Holland, Mike Blomberg, Marty Schlabach, Holly Mistlebauer, Joe DeVeer, Connie Rinaldo, Patrick Randall, Susan Lynch

Welcome Patrick Randall!

Working PT at Harvard – helping with transcription and communication/outreach plan

Cornell (Marty)

No meeting of seed and nursery catalog this week

Creating separate collection in IA for these items that Cornell and NYBG are scanning to facilitate moving into BHL.

Working on Interim report

Marty will be doing a presentation on our project in Feb at the Heirloom Gardening Symposium at Genesee Country Village and Museum in Rochester NY

26k pages prepared to send to scanner for shipment #2. Shipment #1 is still not back yet

Holly is merging the catalog records from 3 participating libraries. Tried to make them look alike to determine what Cornell has that is unique. Each library is Cataloging differently which makes it hard to compare as is.

Seed lists – wondering whether to put in seed catalog collection in BHL or separate collection? Will be on agenda for next collections call. Are any seed lists scanned by MOBOT in BHL? Mike B. says yes and will send examples to Marty.

Kevin Nixon – co-PI on this project. Marty met with him to update. He is interested in transcription tools. Wants to make sure the Hortorium is identified in records in BHL. Is there a way to identify cultivar names like we do scientific names? No registry or one place to identify these names. Some grower societies might have them but would be fractured. Having these would greatly improve searching of this content in BHL.

NAL is now a BHL affiliate member!

NYBG (Susan Lynch)

Continue to upload to IA and then BHL with Macaw. Uploaded 3500 pages or 60 catalogs and in seed and nursery collection in BHL.

Wondering why scientific names not showing up for NYBG? Mike L says may have to do with quality of OCR. Susan will send some examples to Mike L for further investigation

Working on interim report

Harvard

Transcription (Joe)

ALA (DigiVol) – Brewster journals 52% complete, diaries 88% complete; 4 people in library working on transcriptions and validations. FromThePage – journals 29% done, diaries 5% done.

Request for reports - don’t need to send to Jim Hanken

Social media and marketing plan by Patrick Randall

Breaks down audiences for citizen scientists, librarians. Lays out strategy for social media. Partners with existing BHL social media. Partners can forward it to their constituents. Marty wondered is goal to promote transcription or awareness of whole project? Patrick says both.

Marty asked When will horticultural catalogs be pushed to transcription tool? Trish said once more catalogs are uploaded to BHL and Mike B creates second OCR then we can assess the quality

MOBOT

Scanning (Mike B.) just a few more items to scan – 18k pages done so far.

Gaming (Trish)

Tiltfactor has been working on the backend database for storing data produced from the games and nothing new to report on the games themselves.

TILT2(tool to get coordinate info from the transcriptions)– developer has sent us some output but still needs some improvements which he will work on. He originally had a deadline of Oct 31st to finish but is taking longer

Architecture and managing multiple OCR versions (William) – installing mongDB and mediaWiki to handle different representations of documents. Want to track all changes for a word and who did them (e.g manual vs automated). Mike L is working on tools to assess the quality of OCR

Other

BHL blog post last week on crowdsourcing efforts in different projects.

Trish sent out reminders about the financial and narrative interim reports from partners due to MOBOT Nov 30th

Webcast put on by the [Crowdsourcing 101 webinar](http://www.crowdconsortium.org/webinar-crowdsourcing-101-fundamentals-and-case-studies/) Oct 29th put on by the Crowdsourcing Consortium for Libraries and Archives (CCLA)

Takeaways? Marty and Trish attended

Small repeatable tasks more successful in crowdsourcing

How much Consensus on a task to know its correct answer? – depends on type of task. Classification task – do it a lot and adds to machine learning. Somewhere between 3-7 people Transcription is different – don’t want to overdue effort but no magic number.

Some may want to go beyond simple classification to engaging in dialogs, to their own research.

Provide ways to move on to roles with greater responsibility – arbitrator when public doesn’t agree.