**PG Monthly Meeting 2/11/15**

**Attending:** MOBOT (Trish Rose-Sandler, Mike Lichtenberg, William Ulate., Doug Holland, Mike Blomberg), Cornell (Marty Schlabach, Holly Mistlebauer), Harvard (Joe DeVeer, Connie Rinaldo, Patrick Randall), NYBG (Andrew Schinkel),

Absent: Susan Lynch

Cornell (Marty)

Getting Macaw setup to begin processing back from scanner. Shipments 1&2 came back from vendor (15k pages). Sent 3rd shipment this week (10k) and 4th is being compiled. QA looks good of stuff returned.

Some firms don’t have MARC records yet and are being created. Holly’s work to dedupe – NALs list of holdings was not complete but they completed their cataloging of American firms to 1922 and sent list to Holly.

Fev 21st talk at Genesee Country Museum about PG project and Cornell’s seed and nursery catalog collection. Marty will share powerpoint after talk for our PG site.

NYBG (Andrew)

Idenfied there was no OCR text for many of the seed and nursery catalogs scanned at NYBG and uploaded using Macaw.  In addition, the associated PDF files weren’t searchable. With the help of Joel Richard, Mike Lichtenberg and Hank Bromley at IA the problem was determined to be the combination of 48 bit color and jpeg2000 format. The ABBYY OCR software used by IA is unable to extract text from images with these characteristics. With lots of help from Joel and Mike, all of NYBG’s submissions were corrected in both IA and BHL. Our workflow was modified to prevent future occurrences.

NYBG’s total submissions are slightly more than 25K pages. This is a combination of IA and in-house scanning.

Paul Silverman left the project and NYBG because grant funding ended.

Harvard

Transcription (Joe)

ALA (DigiVol) – Brewster journals 100% complete and diaries 100% complete; FromThePage – journals 56% complete and diaries only 15% done. Transcribers prefer journals over diaries because they are more legible.

Hiring another library assistant to help with transcription.

Has Joe identified catalogs to upload to transcription tool? Has identified about 6 catalogs (finding them via Flickr). Graphic heavy catalogs are good candidates. Hope to have all loaded by March 20th (BHL Garden stories campaign begins).

Susan Lynch suggested NYBG’s scanned Emma White catalogs.  See <http://biodiversitylibrary.org/bibliography/99330#/summary>.  The OCR text is present but bad enough to be good input to the game.

Do we have a target number to upload? Not right now. Guidelines will be tricky – harder to do than handwritten pages. How stringent do we need to be with getting order of text? OCR misses a lot of the scientific names. Genus names often only has first letter and laid out all over the page mixed with images and text.

Communications ( Patrick)

Blog post on Brewster in January. Another blog post on March 20th seed catalogs coming up from Patrick.

Put Article about PG in Harvard Gazette – got invite from professor to talk about project at Ludics seminar in Sept. Mary Flanagan was also asked to talk and will attend.

MOBOT

Scanning (Mike B.)

Remaining 13 items added to BHL collection.

OCR outputs - Mike L has begun generating the 2nd outputs in Tesseract and is testing the architecture that was setup to manage all the outputs.

TILT2 update – contracting with Desmond to finish development of his tool by end of March

Tiltfactor progress (Trish) Tested the 2nd version of the gamer game (they are calling Smorball) Sent them names of BHL users interested in OCR correction and some PG staff (Patrick, Paul Silverman, Connie, Susan Lynch) What did folks think of the game?

Tiltfactor sent us output from the test and said it looked very promising – there was a lot of consensus among the players answers.

Other

Webwise dates set? Still no word on dates for Webwise although it looks like the Webwise site is under construction so perhaps they are about to announce it.

Events at our libraries around the game release in May – clarified with Tiltfactor that the only game that will be playable on mobile devices will be the non-gamer game and that it won’t be optimized for mobile but for online

Cornell Reunion weekend in June – theme agro and horticultural biodiversity. Marty will do a talk and could introduce PG to the alums.