

TCN / iDigBio Monthly Standing Meeting (Internal Advisory Committee)

June 13th, 2012 1:00 PM – 3:00 PM

Onsite: Shari Ellis, Kevin Love, Alex Thompson, Jason Grabon, Cathy Bester

Skype: Chris Dietrich, Toby Schuh, Rob Naczi

Absent: Gil Nelson, Deb Paul, Andréa Matsunaga (attending the SPNHC Meeting); Corinna Gries

Action Items

- **June 2012: iDigBio will hire a new Project Manager**
- **June-July 2012: TCN PIs will submit names of recommended programmer participants to Deb Paul, to attend the Augmenting OCR Hackathon (tentatively scheduled for February 2013)**
- **June-July 2012: Andréa is working with Tony Kirchgessner to obtain a full export of NYBG data, with relationships intact (not compressed into a flat CSV file) from the KE Emu database.**
- **June-July 2012: Toby Schuh has offered data digitized from previous projects. Andréa will initiate contact regarding this offer.**
- **June-July 2012: Chris Dietrich has offered data digitized from previous projects. Andréa will initiate contact regarding this offer.**

Minutes

iDigBio Strategic (High-Level) Project Summary

Jason: Final day is June 29th. The new Project Manager has yet to be hired so next month's meeting may or may not be canceled. The four new TCNs will start participating in Internal Advisory Committee meetings when they continue, either in July or August.

Deb and Gil (Jason reporting for them): Currently working out details for the Augmenting OCR Hackathon which will include participation by Xerox and other industry experts. Agenda includes OCR optimal configurations for many OCR applications, comparison of products using a common data set, and code sharing for optimization. This hackathon will be held in February 2013 – there is a request for names of programmer participants to be sent to Deb Paul. Deb and Gill continue to focus on workflow and digitization technology documentation.

Alex: Progressing toward the Specimen Portal Version 0 release at the end of June. A demonstration will be available utilizing Morphbank and FLMNH datasets in July or early August. A webinar will be provided for TCN PIs and TCN IT staff, along with NSF, to show functionality including a description of the back-end technology. This will be followed by requirements-gathering sessions for future portal development and enhancement.

When will data be ready to share and what type of data will we receive from these 2 TCNs?

- Rob – has submitted data to Andréa from the existing NYBG herbaria dataset, however the data is available only in either a flat csv file, or GBIF/IPT output as Darwin Core 1.2 (older version) which means we need to map 1.2 to the newer 1.4 which is an issue. Andréa is working directly with Tony Kirchgessner to obtain a full, relationship-intact export from KE Emu. In terms of TCN activity, NYBG is currently capturing images and later adding metadata (1-2 years from now).
- Toby – iDigBio needs to work with Katja on obtaining datasets, including the special case with host plants along with herbivorous insect data in a relational format (not currently possible with Dwc). FSU is working on extensions for Darwin Core and relational datasets. 100,000+ records are ready to be ingested once we get past the Darwin Core/data ingestion issues.

- Chris – InvertNet is emphasizing images initially; data will be added later with a focus on crowdsourcing. Other institutional databases generated by previous projects can be easily exported as Darwin Core datasets to iDigBio. InvertNet will wait for metadata addition to their database before data ingestion can commence by iDigBio for TCN-related data.

Andrea (Jason reporting for her): The Cyberinfrastructure Working Group has been formed. ACIS is currently working on an application for image ingestion into the iDigBio cloud storage.

Kevin: The Wiki, discussion Forums, and publishing mechanisms are all available at the iDigBio website. The Georeferencing Working Group is gathering and presenting information on the iDigBio Wiki; we expect this documentation base to grow for the benefit of the community.

InvertNet Strategic Project Summary

Chris: InvertNet hosted a workshop in April; the summary was included on the iDigBio web site and in the iDigBio eNewsletter this month. They are refining and testing workflows, have published protocols for slide digitization, and training collaborators. InvertNet is ingesting data which should be made public in the next few weeks. The drawer digitization protocols include the need to assemble a robot to test the most efficient method; a prototype robot has a smaller footprint and is more precise, and should be built later this week. The next step is to replicate this technology with collaborators by building and demonstrating equipment at the Illinois Natural History Survey, conducting training, and sending them home with the tools and skills needed. InvertNet is currently upgrading their cyberinfrastructure, using the Openstack cloud computing infrastructure. UF/FSU are using Openstack for their cyberinfrastructure as well. InvertNet has no storage resources, neither does iDigBio – this has been identified as a gap in NSF funding throughout ADBC. The NIBA Implementation Plan must address this issue, along with tool development gaps.

Tri-Trophic Strategic Project Summary

Toby: They have hired staff to do specimen digitization; most of the subcontracts are engaged. They have restructured and optimized the way they capture host data.

Rob: Melissa, Rich and Kim are attending the SPNHC meeting. Melissa and Kim traveled to the University of Colorado to train that subcontractor. They have hired a summer intern to assist with imaging. In two weeks, Kim will be staffed fulltime. They have been approached by three institutions interested in partnering on future proposal submissions. Digitization and imaging is ongoing.

iDigBio Summit 2

Scheduled for October 22-25, 2012 in Gainesville. The event will include representatives from all TCNs (2011 and 2012). 2011 TCNs are being asked to pay it forward, share information and challenges with new TCNs and connect road maps together, including best practices. The summit will include PIs, technical staff, and digitization staff, with approximately 5-6 participants from each TCN.

NIBA Community Implementation Plan

This was discussed earlier; the workshop will address funding gaps in tool development and storage, among other NIBA needs that are not being addressed by ADBC. The workshop is planned for Fall 2012.

Education and Outreach Coordination

TCNs have \$0 funding dedicated to education and outreach. The Tri-Trophic TCN will be holding two workshops for grad students, but that is the limit of education and outreach. iDigBio has been funded for education and outreach, has a PI dedicated to this purpose, and would like to work with the TCNs on

these items. 2012-2013, iDigBio plans to produce videos on digitization and museum collections which will be shared through TeacherTube, YouTube and the iDigBio website. 2013-2014 the focus will be on individual TCNs to produce and distribute content that advertises the TCNs.

ADBC Messaging/Marketing/Branding

iDigBio will be utilizing resources at FSU to work on marketing and branding to enable consistent communication of information regarding ADBC and iDigBio to the community. Feedback from the TCNs is encouraged to help us improve communication. iDigBio is laying the foundation for sustainability of the project by building community and public knowledge of ADBC. iDigBio will work with the TCNs to form content and messaging strategies, to be discussed at the Summit later this year.

Workflow Modeling/Optimization

There is no consensus on optimal workflows for digitization and how to measure the effectiveness of digitization. Collections are categorized into: flat sheets/2-D objects (e.g., herbarium sheets), 3-D objects (e.g., fossils, vertebrates), and pinned objects (e.g., insects). Funding, staffing, and grant requirements differ greatly, which resulted in a decision to develop workflows that are broken into modules to be used in a dynamic manner by an institution conducting digitization. The working group should have the modules finalized for herbarium sheets in July 2012, with others soon to follow. TCNs may find this information valuable when evaluating their own workflows.

General Open Discussion

Next month's meeting may or may not be canceled due to the hiring of a new iDigBio Project Manager. Confirmation to be provided.

Review Open Items and Discuss Next Steps

A webinar on the iDigBio Specimen Portal and underlying technologies will be available in the near future.