TCN / iDigBio Monthly Standing Meeting (Internal Advisory Committee)
September 12th, 2012 1:00 PM – 3:00 PM
Onsite: Gil Nelson, Shari Ellis, David Jennings, Joanna McCaffrey, Jessica Grosso, Kevin Love
Adobe Connect: Andréa Matsunaga, Deb Paul, Bruce Lieberman, Nahil Sobh, Patrick Sweeney, Rob Naczi, Toby

Schuh, Neil Cobb, Barbara Thiers
Absent: Katja Seltmann, Corinna Gries, Chris Dietrich, Alex Thompson, Larry Page

Action Items

- **Sept 2012:** David Jennings will send out a request to change meeting date/time for next month because of conflict with a scheduled workshop.
- **Sept 2012:** TCNs need to respond to the pre-summit survey ASAP as this information will be used to determine the agenda of the Summit II.
- **Sept 2012:** PALENICHES is conducting a search for post-doc, deadline is November 1st. The advertisement has been submitted several listservs; please send possible candidates their way.
- **Sept 2012:** Contact Gil with potential working group candidates for Entomology detailing digitization tasks list for approximately 6-8 modules.

Minutes

Welcome/Housekeeping
If anyone has TCN update documents that they please email them to Cathy or David. David will change the meeting date/time for next month because of conflict with a scheduled workshop at iDigbio on the same day.

Reviewed Previous Action Items
- Aug 2012: Larry and Shari will send out a pre-Summit Survey to the TCN PIs. Shari resent the survey on 9/12/12. TCNs need to respond to the pre-summit survey ASAP as this information will be used to determine the agenda of the Summit II.
- Aug 2012: Katja will send out a draft of the Digitizers Manual to Neil Cobb. Katja was not present to discuss.
- Aug 2012: Deb and Shari will prepare and send out a pre-workshop survey to the Georeferencing "Train the Trainers" workshop participants. It was sent out and feedback has been received.
- Aug-Sept 2012: Joanna, Deb and Andrea will meet to discuss Emu. Joanna, Deb and Andrea have been engaging in many discussions about Emu.

Nahil- InvertNet Update:
Progress: They have completed stages for vials and slides. They have also answered request for collaborators to include batch upload for the profiles. In addition pages have been sent to Joanna at iDigbio. A paper will be published about InvertNet in ZooKeys. They have the first prototype of digitizing drawers which uses a new set-up, hoping to have a multi-user website.

David Jennings- North American Lichens & Bryophytes Update:

Progress: Digitization is proceeding very well; working with natural language processing applications.

Tri-Trophic Project Update:

Rob- They have just under 56,000 specimens imaged. In preparation of imaging, they barcode the specimens with 74,000 specimens barcoded currently. Kim Watson has trained 4 of collaborating institutions on imaging: University of Wisconsin, University of Minnesota, University of Illinois/ Illinois Natural History Survey and Miami University. They have been receiving data from collaborators; Kim is working with is data including the fully digitized data from University of Minnesota, University of Iowa, University of Colorado, and University of Michigan. Also received images to fill out the data: University of Colorado and University of Michigan. Over 60,000 specimen records total with the recent transfer of data.

Toby- Major update to host data capture to web-based software which will improve the quality of botanical data being captured in association with insect specimens. Most of subcontracts are now capturing data, about 16 institutions are busy working or about to start. Quality of data capture is quite good. Data for 100,000 specimens captured. Produced and updated data entry manual that included new lessons learned from their new data entry staff. Few other software updates to put together.

Barbara- Macrofungi Collections Consortium Update:

Progress: They have just completed 2nd month of the project. This was their first experience with the outreach component; several different events for amateur mycology this past month in Pennsylvania, Alaska and Rocky Mountain Bio Blitz; trained amateurs on how to use portal for their own projects. They have imported 5 different projects since then, mostly annotated checklists from Quebec and Ohio with many more promised. The staff has mostly occupied their time this past month with developing a comprehensive, 100 page project manual (steps people through the different activities). There will be three training workshops in September, October and November. Andrea: Data field input can occur through the MISC working group.

Challenges: Many of Macrofungi Collections Consortium Institutions are not up and running yet. They are still buying equipment, hiring staff, etc., however in the past month they have been accumulating a lot of information for the portal. Almost 400,000 specimen records for the portal from pre-existing records have been imported.

Bruce- PALEONICHES Update:

Progress: Two new undergraduates hired to do databasing; one graduate student is working on databasing and georeferencing. They are doing a search for post-doc; deadline is November 1st, they have submitted ad to several listservs and would like anyone to send them possible candidates their way. They are making good progress on data aspect of project: a lot of entry focused on brachiopods and georeferencing. Almost all trilobytes have been georeferenced: localities, error estimates, notes on how georeferencing is being done (in Specify). They are beginning to work on digital atlases led by John Hendricks at San Jose State University, starting out with Neogene examples, which includes some of the best material available. The first TCN meeting will be held in Lawrence next week with people coming from different institutions. Other aspects of the database include making updates to Specify tree for startographic unit and updating taxonomic trees. They have been contacted by two institutions who are

interested in joining the Paleoniches unit in a partnership of existing network (PEN): Texas A&M and Yale University (they will proceed). They have also been contacted by other paleo people who are interested in submitting their own TCN proposals.

Neil- Southwest Collections of Arthropods Network (SCAN) Update:
Progress: They had a successful "all hands" meeting in mid-August which resolved a lot of workflow procedures. They went over the incorporation of Filtered Push in Symbiota in project but also talked about Specify. Everybody is set up to put data in directly into Symbiota database located at UF or Symbiota is set up to periodically harvest data from other databases. There is no mechanism for having Filtered Push talk to anyone using KEMu. Filtered Push is set up to be work with Symbiota or Specify. They have received data from 5 of 10 Universities. Tri-trophic manual has been sent from Katja -SCAN will incorporate this information into their own manuals and will be able to share that with others by next month.

Patrick Sweeney- Mobilizing New England Vascular Plant Specimen Data Update
Progress: They had their kick off meeting last week: Communication – using Open Atrium which is a project management communication tool that has blog features, notebook features and calendar functionality that will help keep track of progress of project and issues that pop up. The first stage of their project is "precapture stage", which is going to get data that reflects the source structure of the collection and will be associated with specimen records at a later point. Using the application developed by Paul Morris at Harvard - it is ready to go**.** They have developed a workflow that people can use and spent a portion of the meeting going through the workflow, building questions and showing people where to download it and demonstrating the application**.** This month will be the start the "pre-capture" stage of the project where people will begin going through collection and capturing collection level data with the application**.** Neil Cobb asked Patrick to go over "pre-capture" workflow (discussed last time). Herbaria are arranged by taxonomy and often geography, resulting in a bunch of folders within the collection that each have the taxon name along with many specimens identified by that name. Folders may also be organized by geography. There may be several taxon folders, each one with a different state for example and so then several things in the folder will be the same taxon collected in the same state. The pre-capture stage involves going through collection and for a folder you type in a taxon name into the application then type in the state (if collection is arranged that way) and you would do this for several folders in cabinet and print out a label that you can adhere/sheet of paper (insert into file) that has a QR code on it and at a later stage they can be scanned and that data can then be applied to all that data that is within the folder. http://datashot.sourceforge.net/precapture/ApplicationHelp.htm

iDigbio Project Update:

Andréa Matsunaga: She has been meeting with the MISC working group preparing vocabulary needed for data model that they are going to propose, further information is available on the Wiki. This is the first draft of the data model; feedback on draft is welcomed. Joanna is participating and adding descriptions. https://www.idigbio.org/wiki/index.php/MISC-Authority-File-Working-Group

Ver. 0 portal is available; currently gathering feedback from users via the iDigBio web site.

Gil Nelson: The MISC Working Group has merged with Authority Files Working group because of significant overlap. Katja Seltmann, Joanna McCaffrey and Gil Nelson will be co-chairing this working group which will have their first joint meeting in a few weeks. They hope to have a data model that Andrea has been building the database around and first set of concepts and data model will be out by the end of this year. Working group for workflows, which grew out of DROID workshop, Herbarium group called Flatsheets and Packets Workgroup is done. Almost all task list modules are on iDigBio.org documentation page. What Patrick outlined earlier about the New England Plant group is in pre-curation module list of tasks on that page. Database entry module will be out by Friday (9/14/12). Next week there will be a workshop at Valdosta State University involving about 25 participants from small Herbaria that are new and have not yet started imaging or databasing yet. Transitioning workflow digitization task group into Entomology next. The first meeting is next week. Would like to include people outside of DROID if possible, so If there are people you want involved in working groups for Entomological group detailing digitization tasks lists for about 6-8 modules, email Gil.

Deb Paul: The upcoming georeferencing workshop is almost "good to go" and there are almost completed development materials for workshop. Ten participants are on a list for next georeferencing training workshop. Funding is available to develop materials after this workshop is concluded. There will be a post-workshop survey and several members of the working group will set up an outline of what material needs to be developed to provide the community with and to come up with plan for on-line training of georeferencers. The OCR Working Group is creating content for the website, along with planning a workshop in October:

1. Create content of website; what works with Abbey, with Teserac and doesn't work, best practices for setting up software, what image types do they work best with

2. Plan a Hackathon: at BRIT in Fort Worth in conjunction with i-School conference

She plans on submitting a workshop submission and poster to engage outside the natural history collections community to find programmers interested in natural language processing and scene processing to get this output of OCR into more effective format to get into the database faster. Guest Speaker Carl Heinstinka (Germany) from Digital Herbarium Project works with Walter Berinson. They've been working for the past 5 years on OCR on how they can speed things up, success rates with handwriting, etc. They are currently working on presentations for ESA meeting (Entomological Society of America). We want to organize a biodiversity informatics workshop with collection managers, honor undergrads, and grad students interested in learning more about creating databases, moving data into and out of it, how to speak to IT people, etc. The Tri-trophic TCN plans to do two of these workshops. Vertnet is doing something similar based on a need in the community. Paul Morris is developing an informatics course at Harvard University and will discuss it at NIBA. The initial course will be offered at American Museum of Natural History with hopes to recruit outside and inside the institution. If iDigBio can help bridge the gap for students until it is incorporated into curriculums that would be great and more people asking for supports the notion.

Joanna: Usability input, to the portal team to work on the interface from going from V0 to V1. Following change of ownership in data, interested in knowing two things: What you are amassing as a TCN and

what you may be giving us as an institution, and how are duplicates being handled? (Any input on how iDigbio can handle duplication of records, possibility of overlap). Could be a concern in regards to Filter push notion… do you want to give data back from the TCN or from the institution? David Jennings said it could become an issue with the new version of portal at the beginning of the year. Trying to foresee some of the potential problems that we can run into. Data ownership issue is that many databases have data for multiple institutions. Good idea for summit is use of GUIs, ability to identify duplication in database records; there are better and worse models for data aggregation. GBIF may not be the best model. Questions about workflows, about expectations, about data coming to TCNs and them being providers. Morphbank vs. EOL – duplication between the two Identifier questions: force policy with object having only one identifier, and databases need to be able to store more than 1 identifier. One form of ID is not practical workload function. Ed Gilbert is working on Symbiota records – planning – asking provider, do you want to share with others, ex: GBIF

General Open Discussion:
Reminder about the Community page on iDigBio as a good vehicle to make announcements about positions, workshops, meetings, etc.   https://www.idigbio.org/feed/community-announcements

Kevin Love: Has been working with groups to prepare workshops in the next few months and using resources that is available at the HUB – contact if assistance is needed utilizing iDigbio technology https://seminars.adobeconnect.com/_a227210/vqs-participants/

Shari: shoot email or fill out surveys to suggest info. for summit.

David Jennings will send out email about scheduled meeting on October 10th.

Meeting Adjourned