TCN / iDigBio Monthly Standing Meeting (Internal Advisory Committee)
August 9, 2012 1:00 PM – 3:00 PM
Onsite: Kevin Love, David Jennings, Joanna McCaffrey, Shari Ellis, , Alex Thompson, Cathy Bester
Adobe Connect: Katja Seltmann, Corinna Gries, Rob Naczi, Chris Dietrich, Bruce Lieberman, Barbara Thiers, Patrick Sweeney, Neil Cobb, Gil Nelson, Deb Paul, Andrea Matsunaga,
Absent: Larry Page, Toby Schuh

Action Items

- **Aug 2012: Larry and Shari will send out a pre-Summit topic questionnaire to the TCN PIs**
- **Aug 2012: Katja will send a draft of the Digitizers Manual to Neil Cobb**
- **Aug 2012: Deb and Shari will prepare and send out a pre-workshop questionnaire to the Geoferencing "Train The Trainers" workshop participants**
- **Aug-Sept 2012: Joanna, Deb, and Andrea will meet to discuss EMu**

Minutes

Welcome Everyone / iDigBio New Staff Introductions
Joanna McCaffrey is the new iDigBio Bioinformatics Manager
David Jennings is the new iDigBio Project Manager

New TCN Introductions
Neil Cobb – Southwest Collections of Arthropods Network (SCAN) (see attached doc for more detail)

- The SCAN icon and two web sites have been created at: http://scanbugs.org/index.html (Main site) and http://symbiota1.acis.ufl.edu/scan/portal/index.php (Data portal).
- SCAN member museums designed a shared schema for ground dwelling arthropods.
- Symbiota Portal software installed at the iDigBio Hub at UF.
- 9 of 10 SCAN collections have been added to this portal.
- Specimen imaging protocols are being developed.
- Prototype php/Specify/Zoomify web page developed by NAU to display extremely high res images, planning to integrate with the symbiota portal.
- Have 5.5 TB of online storage for image files for all member institutions to store images
- Developing SCAN Working Groups
- SCAN Meeting planned for Aug 14-16 with 25-30 participants

Patrick Sweeney – Mobilizing New England Vascular Plant Specimen Data
Goal: The goal is to digitize 15 herbaria in New England including aggregating data and images via Symbiota. There are three components of the digitization process: collection precapture (currently working on this), primary digitization (actual image capture and precapture of label data), and the final component which includes human key stroking label data.
Progress: The TCN web site has been set up. There will be a kick-off meeting to be held at Yale in September (someone from iDigBio might attend). Progress is being made on workflow development and primary digitization.

Barbara Thiers – Macrofungi Collection Consortium Update
Goal: This TCN includes 35 collaborating institutions US-wide working to digitize mushroom collections which includes 1.6 million records, 700,000 images of living organisms, field notes, and preserved specimens. The portal has data ready to be added (via Symbiota). There is a strong outreach component

to the citizen mycology community who will be able to contribute to digitization efforts. A crowdsourcing app. will be developed to help correct and add data.
Progress: Barbara presented at Botany 2012 as well as at the Mycological Society of America Annual Meeting. They had a technology kick-off meeting that included Vizzuality and Symbiota. Subcontracts have been activated and are in the process of setting up.
Challenges: Challenges include good workflows for OCR and imaging. There is a need for an automatic means of parsing data in workflow post-OCR to populate the database fields semi-automatically. They may be working on this at Symbiota; also working with the OCR working group on this need.

Bruce Lieberman – PALEONICHES
Goal: PALEONICHES involves databasing and georeferencing collections with 6 collaborating institutions focusing on three time periods: Ordovician, Pennsylvanian, and Neogene and three bioregions Cincinnati, American mid-continent and Gulf/Atlantic Coastal Plains. There are 450,000 specimens available online; image collection and digital range maps will be used to create digital atlases online along with versions for portable devices. Goals also include training post-docs, grad and undergrad students in the field of digitization. This TCN will facilitate climate change studies and ecological niche modeling.
Progress: They have a good start on the databasing and georeferencing aspects of the digitization process. Among the tools being used are Specify and stratigraphic tree. In mid-Sept, they will host their first meeting with collaborators. In the near future, they will be advertising for a 2 year post-doc position.
Challenges: Locality concept (temporal and geographic) is an issue. iDigBio will capture the stratigraphy info from the Specify database. In addition, data needs to be broadly distributed – there doesn't exist a good web-based portal that accomplishes this. There is a need to facilitate synergies between biologists (neontologists) and paleontologists. There isn't enough paleo data available online, any opportunities to share – broadly distributed on widely accessed portals. An extension could be added to Symbiota, need to discuss with Ed Gilbert however resources ($) will be required. Bruce will submit a call for a post-doc to iDigBio to post on web site. Katja suggested that an extension for the paleo community be addressed at the Symbiota hackathon.

Action Item: Larry and Shari will send out a pre-Summit topic questionnaire to the TCN PIs
Shari has developed a pre-summit questionnaire for TCNs to be sure the summit meets everyone's expectations; she will email this questionnaire to the TCNs for feedback. Goal is to have the questionnaires back by mid-September.

Chris Dietrich – InvertNet Update
Goal: The goal of Invertnet is to digitize 22 arthropod collections located in the midwest US which includes 55 million specimens. Data consists of high quality specimen images and data labels. The data will be grabbed off the images later so it can be shared in the online portals. ZooKeys publication is now available online, providing further information on InvertNet (http://www.pensoft.net/journals/zookeys/issue/209/).
Progress: They are currently optimizing workflows for slides, vials, and pinned specimens. These workflows are available in their web site. This includes work on pinned specimens in drawers using a ($5,000) robotic arm to shoot images of drawers. The optimal number of images for the best reconstruction of drawers as a stitched image has yet to be determined.
Challenges: They need an efficient method to get specimen label data into the database. Crowdsourcing will require coordination between the TCNs and iDigBio. They are using a semantic repository tagged with minimal metadata in their database (higher taxon), not incorporating fine scale

fields into the repository (using Medici data repository).

Corinna Gries – North American Lichens & Bryophytes Update
Progress: Approximately 300,000 images have been uploaded with the assistance of lots of students working thru the summer. The work on the natural language parser is progressing. They have hired a volunteer and outreach coordinator to focus on crowdsourcing. They also have plans for a newsletter, advertisements, and volunteer events along with outreach. There will be a TCN meeting in November at Wisconsin.
Challenges: How to deal with different vocabulary sets when parsing OCR. In data capture off of slides, this is a possible opportunity for other groups based upon the authority files however this depends primarily on the quality of the OCR results.

Rob Naczi – TriTrophic Project Update (botanical)
Progress: Entomology is using direct data capture while the botanical side is imaging first and creating skeletal records. Rob attended Botany 2012, presented at the symposium and attended the workshop. He met with John Pickering from Discover Life who will be involved with this TCN to show representation of tri trophic interactions and data cleaning. Georeferences will be checked against subdivisions within their own datasets, and eventually other databases.
Challenges: Botany side challenges include the storage of raw images. Does iDigBio want to archive these raw images? Small institutions do not have storage space on servers - approximately 40-60 TB of storage is needed. Alex stated that iDigBio can accommodate this need. Rob will get a better estimate of the exact need and contact Alex. Note: jpgs are derived from the raw images and used in the actual database.

Katja Seltmann – TriTrophic Project Update (insects)
Progress: They are currently working on georeferencing. The portal database is being utilized with over 100,000 records so far. They have completed the Digitizers Manual for specimen database and transforming data as it is entered into database. These decisions are important regarding data quality. They held a two-day outreach at Colombia University, teaching middle school students how to identify insects and identify trophic interactions.
Challenges: They need iDigBio to assist with data export. John Pickering from DL is going to assist with checking data quality issues. The Authority File Working Group is merging with the MISC working group and will be using authority files with insects from online catalogs and actively taking names from scientific literature. What should be done with these lists afterwards? Support taxonomists in taxon-specific level catalogs. Katja will be attending the iDigBio Public Participation in Digitization Workshop in September. Katja will send a draft of the Digitizers Manual to Neil Cobb.

iDigBio Project Update
Gil – MISC working group need to address concerns from the paleo community. They are working on authority files and on working groups based on digitization of a collection type level with future module creation. We are organizing a workshop on herbarium digitization to be held Sept 2012 in Valdosta GA with the intent to share knowledge base with those who need to learn. Gil will attend the upcoming SCAN meeting. He is also planning to host more collection digitization workshops at other appropriate locations in the near future.
Deb – She is organizing the OCR working group planning meeting and finalizing the OCR hackathon. A German OCR expert will be invited to future OCR workshops. The TCNs should let Deb know if they would like to send anyone to future OCR workshops. Georeferencing needs issues should be covered in the upcoming Georeferencing "Train The Trainer" Workshop in October. A pre-workshop survey will be

sent out this week or early next week to determine the exact content of this workshop to meet participant needs.
Andrea – She is working with the MISC working group; they have a place for geology data although the focus is on specimen data. They have received 96,076 occurrences of PALEONICHES data. Action item for next month's IAC meeting: Joanna, Deb, and Andrea will discuss EMu offline. Export to EMu has been done with Darwin Core but the data dump isn't clean. A script has been written to clean data. Issues include different versions of Darwin Core that are being used; they are working with EMu on this issue.
Alex – The new TCN PIs should check out the preview of the specimen portal on the iDigBio web site. Kevin offered to walk the PIs through the portal, if interested, please contact Kevin via email.

General Open Discussion

- David Jennings asked that any documents/presentations prepared for future meetings be emailed to Cathy Bester (cbester@flmnh.ufl.edu) for inclusion in the meeting minutes.
- TCNs requested that iDigBio staff update their profiles to make sure there is a brief description of their role on the project.

Meeting Adjourned

Southwest Collections of Arthropods Network (SCAN

SCAN Current Progress

1. SCAN icon and two websites created http://scanbugs.org/index.html (Main site) and http://symbiota1.acis.ufl.edu/scan/portal/index.php (Data Portal)

2. SCAN member museums designed a shared schema for ground dwelling arthropods. Implemented SCAN Schema for Specify 6 at NAU. Implemented SCAN Schema to Symbiota (Darwin Core).

3. Symbiota Portal software installed at the iDigBio Hub at University of Florida (Ed Gilbert & Paul Heinrich)

4. Nine of 10 SCAN collections have been added to the portal. Data from three collections, >10,000 specimen records online.

5. Specimen imaging protocols using Visionary Digital Imaging systems being developed (31 YouTube videos and 26 Camtasia Powerpoints). http://www.mpcer.nau.edu/cpbc/digital/howto.html .

6. Prototype php / Specify / Zoomify web page developed by NAU to display extremely high resolution images on the web. Planning to integrate with the symbiota portal. Example at http://cpbc.bio.nau.edu/CPMAB/NPS/order_image.php?ord=0,2,8&page=0&item=0&side=

7. Have 5.5TB of online web storage for image files (NAU). All member institutions will be able to store images here.

8. SCAN working groups developing (Outreach, IT, Taxa Files, Georeferencing)

9. SCAN meeting August 14-16, 25-30 participants.

TCN : Macrofungi Collections Consortium : *Unlocking a Biodiversity Resource for Understanding Biotic Interactions, Nutrient Cycling and Human Affairs* EF-1206197

The New York Botanical Garden

Narrative Monthly Report for July 2012

Highlights:

- Project staff hired
- Participant meeting held in New Haven, 17 July
- Three presentations given on the projects
- Technology kick-off meeting held at NYBG, July 2012 Digitization Equipment ordered
- Training documents prepared
- Approximately 2000 photographic slides prepared for digitization
- Xx specimens digitized
- Mycoportal enhanced with observational record archive of the Northeast Mycological Foray; 40,000 records from the Field Museum
- Taxon-based links created between the Mycoportal and Mushroom Observer

Training:

Training of NYBG staff:

- NYBG staff was trained in use of the Symbiota portal by Scott Bates and Ed Gilbert
- Training documents prepared for project participants (preparing photographic slides for digitization, placement of barcodes, imaging of specimens)

Digitization Activities:

1) So far, only The New York Botanical Garden has actually conducted any digitization activities. During this first month we created 967 records at a rate of about 55.25 per hour, and we imaged 684 labels, at a rate of 78.15 records per hour. We also create about 2000 records from field books, and created metadata for the 1996 photographic slides that we are sending out for digitization.

Research: nothing to report

Education and Outreach:

1) B. Thiers Gave presentations about the project:
    - Botany 2012, Wednesday 12 July 2012, Columbus, OH as part of the iDigBio Symposium, sponsored by NSF

- FESIN workshop, 15 July 2012 New Haven CT. The theme of this two-day workshop was the development of a Mycoflora of North America. Attending were professional as well as amateur mycologists who are potential contributors to this initiative.
- Mycological Society of America annual meeting, New Haven, CT., Fungal Conservation Round Table.

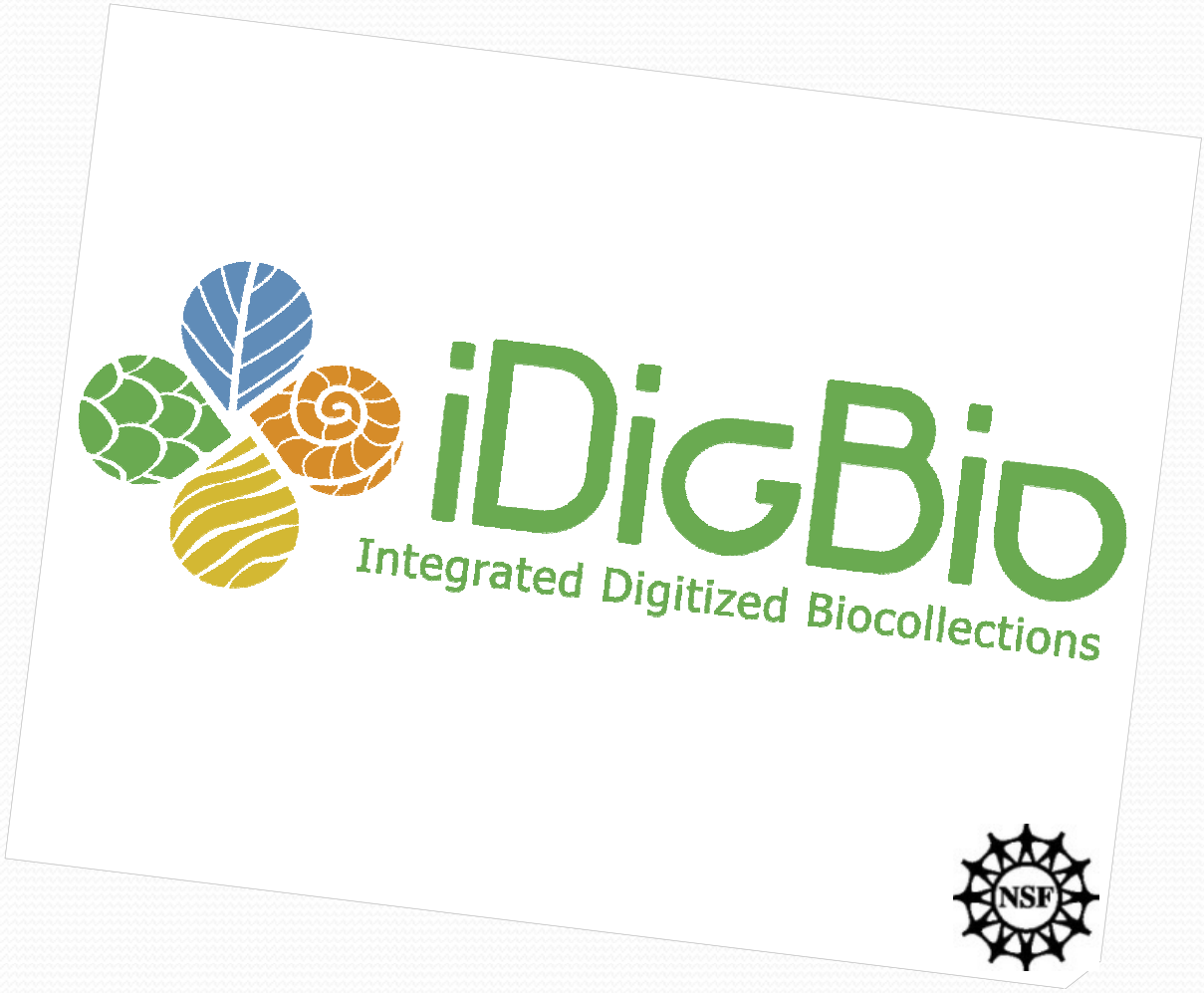
2) The following enhancements were made to the Mycoportal:
   - Linking of the Mycoportal and Mushroom Observer based on taxon name
   - Lists of observed species from the Northeast Mycological Forays (NEMF) from the past XX years were added to the Mycoportal, as an example for how the citizen mycology community can use the Portal for their own projects.
3) Site Visit by Ellen Bloch to the University of North Carolina Botanical Garden Herbarium to advice of curation needed prior to digitization of the Coker Herbarium.
4) Project Meeting Held for participants at New Haven, CT, in conjunction with the Mycological Society of America meeting at Yale University, 15—19 July 2012.

Other:

- Customized letter sent to each participating institution, reviewing budget, activities, suggested sequence of activities
- Hired Shannon Asencio, Record Creation Coordinator
- Hired Gene Yetter, Record Creation Assistant
- Hosted a technology kick-off meeting for the project, including NYBG staff as well as Portal Manager Scott Bates and Symbiota developer Ed Gilbert, and Javier de la Torre and Andrew Hill from Vizzuality to discuss crowdsourcing application funded by this project.

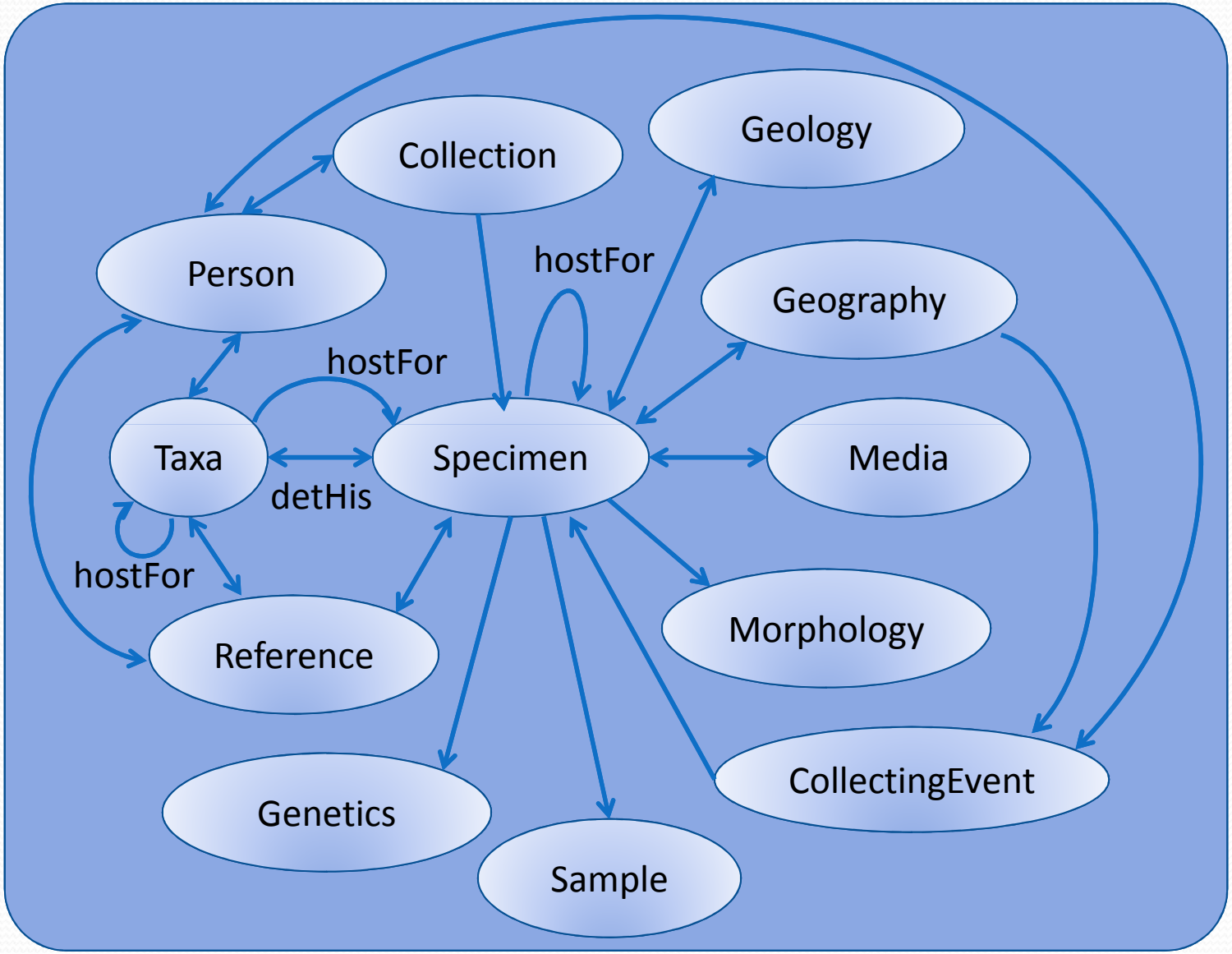
iDigBio
IAC

Andréa Matsunaga



IAC meeting
August 8th, 2012

Potential MISC



Databases/DwC-A received

Dataset	Date	Format	Occurrences	Media	Taxon
TCN-Bryophytes	Jun/01/2012	Symbiota-MySQL	961881	56217	49882
TCN-Lichens	Jun/01/2012	Symbiota-MySQL	691967	59438	10647
TCN-Mycology	Jun/01/2012	Symbiota-MySQL	279529	1179	415812
TCN-InvertNet	Mar/14/2012	DwC-A	631388	0	0
TCN-TTD-AMNH	Jun/21/2012	AMNH-MySQL	643432	4195	61655
TCN-TTD-NYBG	Apr/26/2012	CSV	1469089	905	0
TCN-PALEONICHES	Jul/12/2012	Specify-MySQL	96079	0	6128
FLMNH-Ichthyology	Dec/19/2011	DwC-A	213361	0	0
FLMNH-Ichthyology	Apr/27/2012	DwC-A	214487	0	0
Valdosta	Apr/16/2012	Specify-MySQL	14827	12291	96817
Morphbank	Nov/22/2011	DwC-A	193704	250442	0
Morphbank	Jun/29/2012	DwC-A	194015	252303	0
ITIS	May/31/2012	ITIS-MySQL	0	0	603897
ITIS	Jun/27/2012	ITIS-MySQL	0	0	606131
Total			**5,196,694**	**386,528**	**1,247,072**