

# TCN Progress Update

## February 2013

### Contents

Contents.....	1
InvertNet Update (by Chris Dietrich) .....	1
North American Lichens and Bryophytes Update (by Corinna Gries).....	2
Tri-Trophic Project Update (by Katja Seltmann) .....	2
PALEONICHES Update (by Bruce Lieberman) .....	3
Macrofungi Update (by Barbara Thiers) .....	4
New England Vascular Plants Update (by Patrick Sweeney) .....	6
SCAN Update (by Neil Cobb).....	7

### InvertNet Update (by Chris Dietrich)

#### 1. Progress on digitization efforts:

To date, 3,069 images have been uploaded comprising 1563 slide trays (20 slides/tray) and 247 vial racks (10-16 vials/rack).

A second prototype of the pinned insect drawer digitizing system has been assembled and is being tested. The new system differs from the previous prototype in having a square (rather than triangular) footprint and uses 4 pairs of robotic arms rather than three. This is likely to be the final prototype tested before we move to the production phase later this spring.

The HUBzero cyberinfrastructure platform has now been upgraded to the most recent release of HUBzero, which will provide added security. The infrastructure team is assembling a storage server to be delivered to iDigBio for set-up (InvertNet team members are working with iDigBio tech team members on this). This will provide a third (off-site) node for backing up our network.

## **2. Share and identify best practices and standards:**

Nothing new to report.

## **3. Identify gaps in digitization areas and technology:**

Nothing new to report.

## **4. Report on collaborations with other TCNs, institutions and/or organizations:**

Initial discussions began with personnel at The University of Hawaii, which was recently awarded an NSF CSBR grant (#1203509) to join InvertNet as part of their efforts to upgrade and improve access to their collection.

## **North American Lichens and Bryophytes Update (by Corinna Gries)**

All participants on the LBCC project are digitizing and are making steady progress. We had some discussion about when to expect uploads of images as some institutions currently are not uploading on a regular basis. So, we are expecting larger numbers of images from ALA, F, and DUKE in the near future which have been accumulating locally.

We have started the web development for our volunteer effort. Ben Brandt has been hired to support John Brinda with the setup of a Drupal website which will be hosted in Florida. John Brinda is planning to participate in the Sobefree moss foray in California and do a Symbiota training for participants there hopefully recruiting some volunteers.

## **Tri-Trophic Project Update (by Katja Seltmann)**

1. As of 11 February, NYBG has barcoded 128,354 specimens in preparation for imaging, imaged 104,907 specimens, and received 74,396 additional images from its partners at University of Colorado, University of Michigan, University of Minnesota, and Miami University.
2. We are preparing for the first export of botanical images to iDigBio. Pending resolution of some details related to GUIDs, we plan to export as many of the available images as possible.
3. We are working on first export of insect specimen records to iDigBio from Arthropod Easy Capture. We do not expect the first export to include images, or subcontracts utilizing databases other than Arthropod Easy. The first export is expected to be around 300,000 records.
4. Kimberly Watson is attending the OCR hackathon in Fort Worth, TX on 12-13 February.

5. Christine Johnson is attending the iDigBio Wet Collections digitization workshop (March 5-6, 2013).
6. Katja Seltmann will be attending the RCN Natural Language Processing meeting hosted at NESCent (February 25-27, 2013).
7. We are adding two new digitizers at AMNH. The two new people brings our total up to 8 working through AMNH, and 55 digitizers in total actively working for the insect side of the project. This number is expected to increase over the summer.
8. Interactions with the SCAN TCN occurred. We discussed data issues regarding sharing arthropod data and subcontracts we have in common. Also, database user manuals were shared between the two projects.
9. Imaging exemplar specimens for pinned insects is well underway at AMNH. We have created a user manual for selecting and imaging exemplar specimens, and made it available through the TTD-TCN Website ([tcn.amnh.org](http://tcn.amnh.org)).
10. We evaluated subcontract progress at the beginning of February, and identified those that were behind the curve regarding data-capture progress. We have worked with them to facilitate their data-capture activities, such that now all entomological subcontracts are effectively capturing data.

## **PALEONICHES Update (by Bruce Lieberman)**

This February update is divided into three parts reflecting the work conducted at the University of Kansas (KU) by PI Bruce S. Lieberman and co-PI Una Farrell, and also the work led by the PI at San Jose State University (SJSU), Jon Hendricks, and the PI at Ohio University (OU), Alycia Stigall.

Thus far at KU we have databased a total of ~ 38,000 specimens since the project began. As reported last time all of our locality data has been proofed and entered into our Specify database. Since the project began we have georeferenced a total of 1,336 localities and that translates to 12,786 Pennsylvanian specimens with lat-long. Since our last report, we have also taken some images of Pennsylvanian species though thus far the number of Pennsylvanian species photographed is relatively small and comprises trilobites and brachiopods. We will attach these images to the respective Specify records in our database and we have also sent these images to PI Hendricks for generation of the Pennsylvanian digital atlases (digital atlases are discussed more fully below). I can't remember if I reported this last time, but co-PI Farrell presented a talk at the well attended Palaeontological Association meeting in Dublin, Ireland in late December describing the work on our project (if already reported please accept my apologies). We have made an offer to a post-doc to work on the project that hopefully will accept the position (should know in about 1.5 weeks) and if we are able to convince her to

come this will further accelerate the pace of our work. Two undergraduates continue to work on databasing and we have just about completed work on databasing our brachiopods (as mentioned earlier all of our trilobites are databased) and next for the purposes of databasing we will be moving on to our important echinoderm collections.

Here is what Jon Hendricks had to say: Generation of our proposed digital atlases is well underway. This effort is being spearheaded by PI Hendricks at San Jose State University (SJSU). The overall structure of the websites has been designed and initial efforts have been focused on the development of the Neogene digital atlas. PI Hendricks has worked closely with a SJSU undergraduate student assistant and staff at the Florida Museum of Natural History on this part of the project. Two SJSU graduate students have also been hired to assist with the construction of paleo-distributional maps using GIS and also digital image processing. Digital atlas website construction is nearly finished for two Neogene bivalve groups: the Veneridae and the Lucinidae. Next, attention will shift to two gastropod groups: the Conidae and the Muricidae. We expect to be able to share with iDigBio live links to these four digital atlases by the end of March 2013.

The work on the Ordovician led by PI Alycia Stigall at Ohio University (OU) is also going very well. Here is what she had to say: We met as a group on Feb 1 for a georeferencing workshop, led by Hannah Brame of OU. So all PIs on the Ordovician grant are now trained in georeferencing. Cincinnati Museum Center (CMC) has been working on digitizing the occurrence data for their type collection and have hired a student who will be dedicated to georeferencing their collections. At Miami University, work is underway to transition the old DBS database conversion to be ready for the move to Specify. At OU, we have our Specify database completely setup and running. All of our 135 localities have been georeferenced, and approximately 300 specimen lots (involving over 3,000 individual specimens) have been digitized.

## **Macrofungi Update (by Barbara Thiers)**

### **Digitization Activities:**

#### **Assembling existing records to share through the MycoPortal:**

During January 2013, approximately 30,000 existing records were uploaded to the MycoPortal. As of 1 Feb 2013, approximately 500,000 specimen records from 22 institutions are being shared through the MycoPortal, representing about 83% of the existing records that we estimated in the original proposal – we now believe that we may have underestimated the number of existing records available by as much as 100,000. The bulk of the remaining records to be ingested will come from the U.S. National Fungus Collection in Beltsville.

### **New Digitization (all Participants)**

- 1) Skeletal records (specimen label image plus data record consisting of taxonomic name, collector and collecting number) created: 9226
- 2) Label images captured: 7018
- 3) Full records captured: 2090
- 4) Records completed in the NYBG Project Center: 2318
- 5) Specimens imaged: 262
- 6) Fieldbook records created (database record consisting of locality, collector, number, and date): 16,200
- 7) Ancillary items digitized: 3322

### **Research**

Nothing to report

### **Project Training:**

Dates for the training of Year 2 participants have been selected, and all year 2 participants have been notified. An additional 14 institutions will join the project in year 2.

The sections of the project manual pertaining to photograph digitization and metadata recording, as well as the section on transcription of field books have been revised.

### **Education and Outreach:**

#### **Directly related to the MaCC project:**

- 1) New content (10 documents) has been added to the MycoPortal related to how to collect, prepare and store scientific specimens of fungi:

<http://mycoportal.org/portal/misc/links.php>

- 2) The University of Michigan is working on plans to incorporate citizen mycologists and the public into the project. Project Coordinator Matthew Folz led a winter mushroom foray that included a visit to the herbarium for interested parties. This is the first of an ongoing series of outreach efforts to enlist volunteers for the project to local citizen mycologists.

## New England Vascular Plants Update (by Patrick Sweeney)

### Hardware & Software Development

The engineering group at the University of Oklahoma have been focused on assembling the conveyor system and refining the controller station Server-Client application. Below are some details:

#### Hardware development:

- 1) Assembled and installed 18ft conveyor system in OU laboratory.
- 2) Installed conveyor components – Image contrast sensor and Optical feedback motion sensor to the conveyor station.
- 3) Performed check-up on conveyor belt motion and electrical connections. Discovered issue with conveyor motor; issue resolved through discussion with motor vendor.

#### Software development:

- 1) Made refinements to entry system UI-Interface to key-in specimen meta-data to the database.
- 2) Conducted preliminary time studies to estimate amount of time taken by the entry user to key-in all relevant specimen meta-data using the UI interface.
- 3) Modified database to incorporate image, specimen and process level meta-data.
- 4) Held discussions with iPlant personnel regarding upload of DNG images from digitization station. Performed trial upload of DNG images from OU Wi-fi and Ethernet2 connection using the Windows iDrop (Java) interface.

#### Other activities:

Informatics team at Yale and Harvard decided on method for transferring specimen and image data and images from digitizing stations to Symbiota and iPlant. This will involve generation of rdf/xml data exchange files that will be transferred to and ingested by Symbiota and iPlant. Images (and associated data) will be transferred to iPlant via iDrop, a Java client that allows parallel uploads to the iPlant servers. Currently working on finalizing rdf/xml format for data exchange files and on method to generate data exchange files from digitizing station databases.

#### Digitization

Preparing collections for digitization and capturing collection level-information (i.e., “pre-capture”) are still the the primary activities. Collections-level data capture is occurring at most of the primary digitizing institutions with some institutions approaching 50% completion of collections-level data capture. At institutions where sorting is completed, pre-capture rate is proceeding as fast as

expected or faster. At institutions where considerable sorting and barcoding are happening in conjunction with pre-capture, the pre-capture rate is considerably slower.

## Southwest Collections of Arthropods Network Update (by Neil Cobb)

The Southwest Collections of Arthropods Network (SCAN) has focused on five areas, 1) mapping schema for each museum and protocol development, 2) digitizing specimen label data and preparing for imaging, 3) integrating Filtered Push and Symbiota. 4) enhancing Symbiota for entomological collections, 5) soliciting participation by non-SCAN museums.

We have mapped the schema for each museum in SCAN, allowing us to integrate legacy data sets and/or establish a means to regularly harvest museum data. Nine SCAN museums are now entering data directly into Symbiota. Texas Tech University is beginning to incorporate legacy data into Symbiota followed by all new entries going into Symbiota by March, 2013. For the two museums that will not enter data directly into Symbiota SCAN, UC-Boulder is set to have legacy records harvested by Symbiota on a regular basis, and we have a set protocol for harvesting Texas A&M University data. We have completed the first protocol guide for data entry and will soon submit a SCAN protocol for images and quality control. These and other documents will be served on our new Drupal site provided by iDigBio <http://scan1.acis.ufl.edu/>.

To date we have over 333,070 specimen records served through SCAN Symbiota. This includes 987 families, 7354 genera, 15,187 species.

One of our broader impact activities is to host specimen records from collections that are not funded by the NSF-ADBC program. We have just developed a search function to easily parse out records from NSF-funded SCAN collections and other museum data that have alternate sources of funding to upload their data to SCAN Symbiota. To date the three non-SCAN collections have submitted 134,399 records (part of the 333,070 total). In addition to adding BYU and Harvard ant collection to the SCANetwork, the Museum of Comparative Zoology (Harvard), the University of Utah, and Kutztown University have been entering data into SCAN Symbiota, Wichita State University will begin entering data for Scarabaeidae by March, 2013 and we plan to do the same for a portion of the Utah State University insect collection.

Symbiota now has all of the functionality that is essential for entomological collections and we will continue to add more options in 2013 as time permits.

FilteredPush has reconfigured the SCAN network instance to use the W3C Open Annotation Community Group's Open Annotation Ontology Specification. Filtered Push has demonstrated harvest of a taxonomic authority file from Symbiota into an ontology, and the use of reasoning on that ontology

to notify parties who express interests at higher taxonomic ranks of annotations making assertions about included taxa.