

TCN December Updates

December 2012

Contents

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

InvertNet Update (by Chris Dietrich)

Digitization of slide-mounted and ethanol-preserved specimens is ongoing at several collaborating institutions. To date > 2,500 images have been uploaded, representing >1500 slide boxes and >100 vial racks, together comprising >16,000 individual arthropod specimens.

Most recent activities at the lead institution involved upgrading the project web platform and development and testing of the hardware and workflow for digitizing drawers of pinned specimens.

Over the past month we upgraded our HUBZero Infrastructure to the latest released version. This will enable us to address some security issues and include some new features. Student programmers worked on a new image segmentation/annotation tool and customization of the image ingest form. We also began working with the iDigBio technical team to establish an InvertNet backup storage server at UF.

Testing of the first delta robot prototype (drawer digitization hardware), based on the original 3-arm design indicated that this system will work for the purpose of capturing images necessary to create high-resolution 2D and 3D reconstructions, but the overall size and footprint of the machine is cumbersome. The machine hardware design was redesigned based on recently published research, to a square format machine with 45 degree descending motion struts that optimizes the positioning workspace of the

system while minimizing the size and footprint of the system. This new prototype is under construction and testing will occur over the next month. The system software was updated to support modern 64-bit multicore commodity computers, the OpenCV computer vision system was integrated with the machine control software to support automated and optimized machine positioning, focus, and tray capture by detecting the geometry of the drawer and unit trays within, independent of size and without operator intervention. OpenCV integration leverages the industry standard machine vision system and includes support for feature detection, optimized stitching, label processing, and 3D reconstruction.

Computer science graduate student, JonMark Lau is continuing work on improved approaches to 3-D reconstruction. Initial results achieved using the Disney stereo reconstruction method are very good given the characteristics of the data (fine features, unsmooth and shiny surfaces, unfamiliar shapes). However, although expectations for 3-D reconstruction from multiple photographs are quite high, methods that have been successful work on retexturing very crude reconstructions (e.g. cities constructed from coarse building blocks) or very familiar shapes (e.g. faces with smooth surfaces, where detail was re-synthesized afterward). Because these previously available algorithms do not yield “realistic” reconstructions of insect specimens, Lau is working on a new model that better takes into account the variable textures, colors, and reflectance properties of arthropod exoskeletons.

North American Lichens and Bryophytes Update (by Corinna Gries)

LBCC annual meeting in Madison: November 1 -2, 2012 “Transcribing label information in Symbiota”
Participants: Barbara Thiers, Blanka Shaw, David Giblin, Donald Pfister, Ed Gilbert, James Allen, John Brinda, Julie Smith, Lance Jones, Matt von Konrat, Michaela Schnull, Monica Proulx, Robert Anglin, Tatyana Livshultz, Tiffany Bone, Tim Hogan, Tom Nash.

Advances in Symbiota over the last month:

- Added to portals the ability for collection manager to select their own usage rights. The default options follows the iDigBio documentation. See attached screenshot. This will be implemented on all Symbiota portals.
- Added to portals the ability to create Darwin Core Archive (DwCA) file extracts containing occurrence, determination, and image data. These will be setup so that collection managers will be able to refresh the DwCA files at will. All DwCA files are listed on the following RSS feed. iDigBio will harvest data from each of the data portals in this same manager.
 - <http://lichenportal.org/portal/webservices/dwc/rss.xml>
- GUID table is being added that assigns UUID to occurrence, determination, image, collection records. At the next major software update (next month), the GUIDs will be included into the DwCA files.

- Exsiccati search, display, and management tools activated. The exsiccate index has been added to the left menu of the lichen portal.
 - <http://lichenportal.org/portal/collections/exsiccati/index.php>

Ed Gilbert is involved in the OCR hack-a-thon planning and preparations

Imaging: as of today 307346 label images are on the server at FSU:

| Lichens | | | Bryophytes | | |
|----------------|------------------|---------------|-------------------|------------------|---------------|
| Acronym | Specimens | Images | Acronym | Specimens | Images |
| ASU | 114263 | 32 | BING | 747 | 3 |
| BRY | 14847 | 14419 | CMN | 11227 | |
| CMN | 31102 | | CNABH | 2 | |
| CNALH | 8587 | | DUKE | 180098 | 11413 |
| DUKE | 3801 | 1124 | F | 66150 | 8134 |
| F | 35020 | | FH | 7340 | |
| FH | 3484 | | ILL | 5660 | 9122 |
| ILLS | 10453 | | LSU | 468 | 354 |
| ISC | 4294 | | MICH | 23501 | 25475 |
| LSU | 17 | | MO | 228835 | |
| MICH | 31620 | 27508 | MSC | 2653 | 2653 |
| MIN | 144182 | | NY | 344041 | |
| MSC | 105476 | | PH | 12463 | |
| NCU | 3612 | 227 | TENN | 15767 | 11754 |
| NY | 119539 | 67653 | UBC | 167840 | |
| OMA | 18574 | | UC | 95192 | |
| OS | 2545 | 3640 | UTC | 1164 | |
| OSC | 5198 | | VSC | 4069 | 4069 |
| PH | 583 | | VT | 14343 | 13584 |
| SBBG | 31294 | | WTU | 74141 | 23082 |
| SRP | 16336 | 6622 | | | |
| TENN | 8061 | 8023 | | | |
| UCR | 13693 | | | | |
| UPS | 26226 | | | | |
| UTC | 265 | | | | |
| VT | 2291 | 2291 | | | |
| WIS | 70242 | 63111 | | | |
| WTU | 30375 | 3053 | | | |

Tri-Trophic Project Update (by Toby Schuh)

- Kim Watson of the NYBG has completed training in imaging for staff at all of the herbaria in the network, and all are capturing images of specimens in their collections.
- As of 29 Nov 2012, staff at the New York Botanical Garden completed barcoding of 100,000 specimens and imaging of 85,000 specimens in the herbarium.
- The Arthropod Easy Capture (aka PBI) database has captured data for 160,892 specimens as of 29 Nov 2012.
- We are moving forward to preparing Arthropod Easy Capture for release as an appliance.
- We have started imaging exemplar insect specimens at the AMNH.
- Two of our participants had articles appear in the recent ZooKeys volume entitled “No Specimen Left Behind: Mass Digitization of Natural History Collections”

PALEONICHES Update (by Bruce Lieberman)

Progress is going very well on the grant. For instance, just since the middle of September at KU alone we have databased more than 11,000 specimens. We have also at last finished scanning and proofing all of our several thousand Pennsylvanian locality cards so this will now facilitate much easier batch georeferencing. Another activity we have been doing is taking digital images of some of our species. We have been sending these to Jonathan Hendricks at SJSU where he has been hard at work creating digital atlases both for the Neogene and also for the Pennsylvanian material. Moreover, one of my graduate students has been working with Roger Portell at Florida, facilitated by Jon, to georeference a bit of their material. Finally, things are also going very well with the Ohio folks led by Alycia Stigall. Both her at OU and those working under her Miami U have updated their databases to Specify 6 and are working on data entry. The Cincy Museum is right now focusing on databasing their types.

Macrofungi Update (by Barbara Thiers)

Highlights

- Trained representatives from 11 participating institutions for year one of the grant (13 people in total) in 3 group and 2 individual training sessions.
- Completed a comprehensive training manual for the project
- Incorporated about 200,000 records into the MycoPortal; the portal now contains 488,815 records, approximately 75% of the existing records to be incorporated
- Created 20,641 skeletal and 36,244 fieldbook records
- Digitized 4062 photographic slides and 4143 other ancillary documentation (e.g., notes)

- Held a workshop for amateur mycologists about making scientifically documented collections and sharing the data through the MycoPortal at the annual forary of the North American Mycological Association in Santa Cruz, CA.

Digitization Activities

Assembling existing records to share through the MycoPortal: As of 18 December 2012, 488,815 specimen records from 19 institutions are being shared through the MycoPortal, representing about 75% of the existing records to be incorporated.

Creation of New Records:

- 1) Skeletal records (specimen label image plus data record consisting of taxonomic name, collector and collecting number) and Full records created: 20,641
- 3) Fieldbook records created (database record consisting of locality, collector, number, and date): 36,244

Digitization of Ancillary items:

- 1) 4062 photographic slides
- 2) 4143 specimen notes and drawings

Project Training

Training sessions held:

[26—28 September: Trained participants from Cornell University, University of Florida and Louisiana State University at The New York Botanical Garden]

2-4 October: Trained participants from U.C. Berkeley, University of North Carolina and Harvard University and Louisiana State University at The New York Botanical Garden.

6—8 November: Trained participants from N.C. State University, Denver Botanic Gardens, Illinois Natural History Survey and Harvard University at The New York Botanical Garden.

6—8 December: Trained participant from University of Michigan at The New York Botanical Garden.

10-12 December: Trained participants at San Francisco State University.

Also, prepared 200 page procedure manual for the project

Education and Outreach

Directly related to the MaCC project:

1) MycoPortal content created by or for the citizen mycology community:

- All voucher images and observational records for the Northeast Foray and the North American Mycological Association have been posted to the Portal
- 17 checklists based on geography and 9 checklists based on taxon with keys and illustrations and in some cases descriptions for species.
- Approximately 1715 images of living fungi

2) P.I.s Barbara M. Thiers and Roy E. Halling and Project Coordinator Shannon Asencio held a workshop for amateur mycologists about making scientifically documented collections and sharing the data through the MycoPortal at the annual foray of the North American Mycological Association in Santa Cruz, CA (13—16 December, 2012).

3) Else Vellinga, U. C. Berkeley, wrote an article for the December 2012 issue of Mycena News, newsletter of the San Francisco Mycological Society about the MaCC project called, “Dried Mushrooms, What’s the Deal?”

General outreach about fungi by MaCC project members:

Kathie Hodge presented a talk on Fungi for the Cornell Young Naturalists’ Club, Oct 23 Included 25 preteens and some parents.

Kathie Hodge presented a talk on Nov 14 entitled “Pretty Rotten: The Fungi that Eat Our Food” for Trumansburg High School.

New England Vascular Plants Update (by Patrick Sweeney)

Hardware & Software Development

The engineering group at University of Oklahoma have been focused on developing the controller station Server-Client application. Below are brief status updates on the application development. The controller station interfaces with all of the software and hardware components of the entire digitization system.

1. MySQL Database developed. Entity Relationship Diagrams (ERD) between the different tables and data attributes within the database are drawn up for system design and development efforts.

2. Server side interface with the Canon MarkIID DSLR camera system completed. The CANON software kit is integrated into our JAVA based application to enable 1) signal communication between software controller and camera system; b) image capture, image retrieval and image delete functions automatically done through software interface.
3. TCP/IP communication link up with entry-user computer for specimen meta-data retrieval and storage within the system database completed. Coding for match-up of specimen meta-data with specimen image done through a queue based algorithm.
4. Client-side interface for the entry user to preview all of the images captured by the system completed.

OU just completed the UI interface prototype for the entry user who will be performing the majority of data entry tasks during the digitization of specimens. The UI interface coupled with a system database TCP/IP communication link to ensure constant communication between the controller station and entry user computers. The entry user computer is a large touchscreen HP computer station.

Other Activities

- Primary data-capture fields finalized and field mappings into Symbiota established.
- Draft image administrative metadata fields identified.
- Began discussion with iPlant about image and image metadata transfer to iPlant data storage environment. Next steps are to develop a proof of concept document and validate concepts/features.

Digitization

At this time, preparing collections for digitization and capturing collection level-information (i.e., “pre-capture”) are the primary activities. Using the Java application completed in September, collections-level data capture is occurring at most of the primary digitizing institutions. Progress at the various institutions ranges from just beginning to over 28% (at HUH) complete.

Outreach

The team at Boston University has been pushing very actively to establish the groundwork for the citizen science network. They have been contacting potential collaborators and volunteers and have established a website on the National Phenology Network site: www.usanpn.org/NELOP. A major goal is to have a network of trained volunteers in place by the beginning of spring.

SCAN Update (by Neil Cobb)

SCAN has concentrated on five focus areas, 1) mapping schema for each museum, 2) digitizing specimens, 3) soliciting participation by non-SCAN museums, 4) enhancing Symbiota for entomological collections, 5) integrating Filtered Push and Symbiota.

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. Seven museums are now entering data directly into Symbiota. For the three museums that will not enter data directly into Symbiota SCAN, Texas Tech University has a functional Specify database running, 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.

To date we have over 100,000 specimens digitized in SCAN Symbiota. This includes over 4,106 species, ~407 families and from at least 48 countries. We have georeferenced 56,393 specimen records.

In addition to adding BYU and Harvard ant collection to the SCANetwork, we are set to have the University of Utah digitize ~250,000 records into SCAN Symbiota 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.

Symbiota and Filtered Push have produced beta annotations that link to Symbiota and we will begin testing annotations among museums in early February, 2013.