

## **MOBILIZING NEW ENGLAND VASCULAR PLANT SPECIMEN DATA TO TRACK ENVIRONMENTAL CHANGE**

Report submitted by: p\_sweeney@att.net  
Report Submitted on: 09/09/2013 - 09:02

### **Progress in Digitization Efforts**

Capture of collection level-information (i.e., “pre-capture”) is the primary activity. At this stage approximately 650,000 specimens have been pre-captured -- with at least current identification captured. One high-throughput digitization apparatus (conveyor system) was installed at the end of July; completion of testing & configuration of the conveyor system is imminent and primary digitization will begin before the end of September. Light-boxes will begin to be deployed in September/October.

### **Share and Identify Best Practices and Standards**

nothing to report

### **Identify Gaps in Digitization Areas and Technology**

nothing to report

### **Share and Identify Opportunities to Enhance Training Efforts**

nothing to report

### **Share and Identify Collaborations with other TCNs, Institutions, and Organizations**

We are continuing to work with the team at iPlant to develop a mechanism to transfer images to the primary image repository for NEVP (i.e., iPlant/TACC). We continue to collaborate with the FilteredPush project, the Symbiota team, and iDigBio.

### **Share and Identify Opportunities and Strategies for Sustainability**

nothing to report

### **Other Progress (that doesn't fit into the above categories)**

nothing to report

## **DIGITIZING FOSSILS TO ENABLE NEW SYNTHESSES IN BIOGEOGRAPHY- CREATING A PALEONICHES**

Report submitted by: blieber@ku.edu  
Report Submitted on: 09/09/2013 - 12:40

### **Progress in Digitization Efforts**

Please see attached file

### **Share and Identify Best Practices and Standards**

Please identify attached file

### **Identify Gaps in Digitization Areas and Technology**

Nothing new to report at this time

### **Share and Identify Opportunities to Enhance Training Efforts**

Nothing new to report at this time

### **Share and Identify Collaborations with other TCNs, Institutions, and Organizations**

Nothing new to report at this time

### **Share and Identify Opportunities and Strategies for Sustainability**

### **Other Progress (that doesn't fit into the above categories)**

## **NORTH AMERICAN LICHENS AND BRYOPHYTES: SENSITIVE INDICATORS OF ENVIRONMENTAL QUALITY AND CHANGE**

Report submitted by: cgries@wisc.edu  
Report Submitted on: 09/10/2013 - 18:29

### **Progress in Digitization Efforts**

As of September 2013 the number for the LBCC are as follows:

Lichens: <http://lichenportal.org>

Herbaria actively submitting images or key stroked records to the portal: 45

Specimen records in portal: 1,070,751 (up by 81384 since July 2013 - and we surpassed the 1 million!)

Specimen records with label images: 362,393 (over 65,000 labels have been imaged since July 2013)

Bryophytes <http://bryophyteportal.org>

Herbaria actively submitting images or key stroked records to the portal: 36

Specimen records in portal: 1,567,730 (up by 134,955 since July 2013)

Specimen records with label images: 351,698 (about 78,000 labels have been imaged since July 2013)

The digitization effort has taken a major leap over the summer with 143,000 labels imaged within 2 months. We assume that this may slow down a bit again, now that the semester has started and not as much student help is available. However, we have put major efforts into developing Natural Language Processing routines and the first batch of imaged labels (Lichen from Montana) have parsed beautifully. We will continue to improve NLP for batches of similar labels. The data entry screen in Symbiota has been adapted to highlight information that has been placed programmatically rather than typed by a person, so that editing is easily done.

### **Share and Identify Best Practices and Standards**

All our material on Best Practices and Standards for the LBCC are on our project website under workflows:

<http://lbcc.limnology.wisc.edu/node/3>

### **Identify Gaps in Digitization Areas and Technology**

nothing to report

### **Share and Identify Opportunities to Enhance Training Efforts**

nothing to report

### **Share and Identify Collaborations with other TCNs, Institutions, and Organizations**

### **Share and Identify Opportunities and Strategies for Sustainability**

### **Other Progress (that doesn't fit into the above categories)**

We continue to improve our volunteer interface. The first round of testing has been conducted and suggestions have been integrated.

## **INVERTNET: AN INTEGRATIVE PLATFORM FOR RESEARCH ON ENVIRONMENTAL CHANGE, SPECIES DISCOVERY AND IDENTIFICATION**

Report submitted by: chdietri@illinois.edu

Report Submitted on: 09/11/2013 - 09:01

### **Progress in Digitization Efforts**

A total of 10,110 images, mostly of sets of slides and vials, have been uploaded. Testing of the image annotation tool that will allow users to highlight individual objects on each image and tag these objects with metadata is nearly complete and we plan to push this tool from the testbed to the live site soon. For images of slide trays, the tool automatically detects individual slides, draws a box around them, and creates a blank record into which the user may enter DarwinCore metadata. We are continuing discussions with Notes from Nature in the effort to incorporate crowd-sourcing of image annotation and label data capture into our web portal. Testing of our drawer digitization system is essentially complete and we are planning a workshop for InvertNet collaborators to be held Nov. 1-3 to train collaborators in the set-up and use of the system.

### **Share and Identify Best Practices and Standards**

Nothing to report.

### **Identify Gaps in Digitization Areas and Technology**

Our technical team is working on ways to improve software for stitching images of whole drawers to decrease errors and increase speed. Our goal is to create software that can stitch in real time, as the camera is capturing the raw images. We also have a PhD student in computer science working on the creation of improved 3D reconstructions for insects using the images captured using our drawer digitization system.

### **Share and Identify Opportunities to Enhance Training Efforts**

Nothing to report.

### **Share and Identify Collaborations with other TCNs, Institutions, and Organizations**

### **Share and Identify Opportunities and Strategies for Sustainability**

### **Other Progress (that doesn't fit into the above categories)**

## **PLANTS, HERBIVORES AND PARASITOIDS: A MODEL SYSTEM FOR THE STUDY OF TRI-TROPHIC ASSOCIATIONS**

Report submitted by: moon@begoniasociety.org

Report Submitted on: 09/11/2013 - 09:32

### **Progress in Digitization Efforts**

Insect Complete Records: approximately 457,553

Insect Images: approximately 870

Plant Images: approximately 834,836

Plant Complete Records: approximately 13,151

### **Share and Identify Best Practices and Standards**

Working with iDigBio to formulate fields for sharing association data.

Discover Life is expanding a crowd sourcing portal for transcription of plant sheet image records. This portal is intended to be open to everyone in the community.

NYBG is working with Atlas of Living Australia to transcribe plant sheet images. A summary and information of this initiative can be seen on the NYBG blog (<http://www.nybg.org/plant-talk/tag/crowdsourc>).

### **Identify Gaps in Digitization Areas and Technology**

Nothing new this month.

### **Share and Identify Opportunities to Enhance Training Efforts**

Nothing new this month.

### **Share and Identify Collaborations with other TCNs, Institutions, and Organizations**

Recently Katja Seltmann met with AntCat, GNA, ScratchPads, ZooBank and SpeciesFile regarding taxon name catalogs and data sharing. Simple workflows and action items were met to share data from already existing taxon catalogs (i.e. AntCat and Plant Bugs Catalog) to GNA and ZooBank. New web services were discussed to facilitate transfer of catalog information easier.

### **Share and Identify Opportunities and Strategies for Sustainability**

### **Other Progress (that doesn't fit into the above categories)**

## THE MACROFUNGI COLLECTION CONSORTIUM: UNLOCKING A BIODIVERSITY RESOURCE FOR UNDERSTANDING BIOTEC INTERACTIONS, NUTRIENT CYCLING AND HUMAN AFFAIRS

Report submitted by: barbara.thiers@gmail.com  
Report Submitted on: 09/11/2013 - 10:06

### Progress in Digitization Efforts

# specimens newly digitized (includes labels digitized): at least 17,481  
# records completed (or completed records entered): 5956  
#field book records created: unknown  
Field images digitized:1480  
Specimens added to the MycoPortal: 17,481  
Total specimens in the MycoPortal: 1,453,123  
Mycportal records edited: 37,747

### Share and Identify Best Practices and Standards

Since our go-to camera, the Canon EOS Mark II is no longer being made, we have been exploring other camera options. We have found that the best alternative for this project in the same price range is one the Nikon 36 MP DSLRs

### Identify Gaps in Digitization Areas and Technology

Gaps in digitization areas and technology remain with regard to interpretation of OCR results and data parsing. Apparently the SALIX approach to data parsing will soon be folded in to the Symbiota websites; this will be very welcome.

Crowdsourcing: Notes from Nature has made a lot of progress figuring out how to interact with potential new additions to the project, and has developed a list of budget and procedural items for new proposals. Because NfN is managed by already over-committed volunteers, the biggest challenge we have is how to edit and manage the data post entry, and a secondary challenge is how to manage inquiries and day to day interactions with users and interactions with it. I think discussions are underway to address how iDigBio could help Notes from Nature with some of these challenges, at least until NfN can secure its own funding.

### Share and Identify Opportunities to Enhance Training Efforts

1. Two undergraduates summer interns worked on the project at The New York Botanical Garden in July and August. In addition to learning digitization techniques, they also took weekly field trips with P.I. Halling to collect and identify macrofungi, created a website intended for public interest in the project (<http://www.themacrofungiproject.com/>), and participated in a fungal bioremediation/art project at a Superfund site in New York City (Newtown creek, between Brooklyn and Queens, <http://www.livescience.com/20573-fungal-cleanup-newtown-creek.html>).
2. A high school teacher from Aspen, Colorado, James Dula, spent a week at NYBG in July volunteering in the MaCC project and enhancing his knowledge of fungi. He is developing a course in biotechnology in which he wants to use macrofungi, probably in a bioremediation project.
- 3.. A workshop for high school biology teachers funded through this grant was held at North Carolina State in Raleigh from August 5-7, 2013. Twenty-eight applications were received for the workshop and five teachers were selected on a competitive basis. In addition to these teachers, five undergraduate students from MGSC that are preparing for a career in

teaching will also participate in the workshop. The participants made field collections, and then made microscopic examinations in the laboratory. They used the MycoPortal to record their collections and observations.

### Share and Identify Collaborations with other TCNs, Institutions, and Organizations

We have begun a collaboration with Genbank mycologist Conrad Schock to add GenBank numbers to MycoPortal records, and to add links to MycoPortal records with sequences in GenBank. We are very excited about this collaboration because it not only addresses the problem of poor citation of specimens in GenBank, but also further imbeds the MycoPortal in the standard work practices of mycologists, which is key to the sustainability of this resource

### Share and Identify Opportunities and Strategies for Sustainability

Our strategy for sustainability is as follows:

- 1) Make the MycoPortal an indispensable tool for mycological research by linking it to GenBank (see above), and making upload of specimen data into the MycoPortal a requirement for publication in *Mycologia*, the journal of the Mycological Society of America. Discussions have started on GenBank portion of this strategy, and will be started with the editor of *Mycologia* at the annual MSA meeting in August.
- 2) Continue the reach of the MycoPortal beyond macrofungi. Dr. Andrew Miller of the Illinois Natural History Survey is preparing a proposal to digitize microfungi to be added to the MycoPortal. Broadening the user base will help sustain the project
- 3). Internationalization of the MycoPortal. Soon we will add data from non North American herbaria to the MycoPortal; we hope this will stimulate continued discussion of the development of complementary projects in Asia, Europe and South America that further broaden the scope of the MycoPortal
- 4) Management of the Portal beyond the current grant: My dream is to have the MycoPortal Management become a standing committee of the Mycological Society of America, and that they will allow donations above their current membership rates to support the MycoPortal. I hope to start discussions with members of the Executive Committee of the Society about this at the MSA meeting in August.

### Other Progress (that doesn't fit into the above categories)

1. Presentations about the MycoPortal and the MaCC Project were given at two summer mushroom forays for amateur mycologists: the Gulf Coast Foray, held in Mississippi July, and the New Mexico foray, held this year in Southwestern Colorado, in late August. Dr. Clark Ovrebo of Central Oklahoma University gave these presentations
2. Three presentations were given about the MaCC project at the annual meeting of the Mycological Society of America in Austin, TX in late July, 2013. Thiers, Bates and Halling presented a poster summarizing accomplishments of the first year, Hughes and Petersen presented a poster on the digitization of the L. R. Hesler notebooks at University of Tennessee, and Matthew Foltz gave an oral presentation on the digitization of macrofungi at the University of Michigan.
3. An initial dataset of macrofungi, mostly from University of Michigan and New York Botanical Garden, has been prepared for inclusion in *Notes from Nature*, and hopefully will be posted there by the end of this month.

## **SOUTHWEST COLLECTIONS OF ARTHROPODS NETWORK (SCAN): A MODEL FOR COLLECTIONS DIGITIZATION TO PROMOTE TAXONOMIC AND ECOLOGICAL RESEARCH**

Report submitted by: neilscobb@gmail.com  
Report Submitted on: 09/11/2013 - 17:48

### **Progress in Digitization Efforts**

We are on target to meet our second-year quota to digitize labels from pinned specimens. Below are three sets of statistics as of September 11, 2013.

Specimen records in SCAN database from ADBC funded museums

- 365,915 specimens in database
- 247,075 (68%) georeferenced
- 265,467 (73%) identified to species
- 527 families
- 5380 genera
- 11,167 species

Specimen records in SCAN database from non-ADBC funded museums

National Park Collections, Museum of Comparative Zoology, Harvard University, Gregory P. Setliff Collection - Kutztown University, Entomology Collection at the Natural History Museum of Utah, and Utah Department of Agriculture and Food Entomology Collection -

- 148,756 specimens in database
- 58,787 (40%) georeferenced
- 72,489 (49%) identified to species
- 738 families
- 3,759 genera
- 7,821 species

All specimen records in SCAN database

- 506,682 specimens in database
- 298,105 (59%) georeferenced
- 336,432 (66%) identified to species
- 1,016 families
- 8,045 genera
- 17,501 species

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.

(A) Set up an OAI/PMH harvesting of the taxonomic authority file from SCAN Symbiota into the network, along with conversion to an ontology.

(B) Set up demonstration for using the ontological representation of the taxonomic hierarchy to match an annotation that asserts an identification using a lower taxon with an interest expressed on a higher taxon.

(C) We will configure the scan instance of Symbiota on symbiota1 to start submitting annotations to Filteredpush by September 13, 2013.

Symbiota continues to develop functionality that is essential for entomological collections and we will continue to add more options in 2013 as time permits. The most recent upgrade allows users to create species checklists for areas at one kilometer increments. This permits users to create specific lists for even small parks or schoolyards. We will have completed the protocol for uploading batches of images to the SCAN portal by September 13, 2013.



### **Share and Identify Best Practices and Standards**

Nothing new to report, we are working on activities already described in previous reports

### **Identify Gaps in Digitization Areas and Technology**

Nothing new to report, we are working on activities already described in previous reports

### **Share and Identify Opportunities to Enhance Training Efforts**

Nothing new to report, we are working on activities already described in previous reports

### **Share and Identify Collaborations with other TCNs, Institutions, and Organizations**

Nothing new to report, we are working on activities already described in previous reports

### **Share and Identify Opportunities and Strategies for Sustainability**

Nothing new to report, we are working on activities already described in previous reports

### **Other Progress (that doesn't fit into the above categories)**

Nothing new to report, we are working on activities already described in previous reports

## **THE MACROALGAL HERBARIUM CONSORTIUM: ACCESSING 150 YEARS OF SPECIMEN DATA TO UNDERSTAND CHANGES IN THE MARINE/AQUATIC ENVIRONMENT**

Report submitted by: Chris.neefus@unh.edu  
Report Submitted on: 09/12/2013 - 09:27

### **Progress in Digitization Efforts**

The project started on August 1 of this year. We are using the light-box imaging stations. The members of the consortium that will be digitizing in Year 1 have ordered the imaging station components. The light boxes have been taking 4-6 weeks to arrive, so some institutions have not received theirs yet. While waiting for equipment to arrive, we have been barcoding specimen folder by folder and creating a skeletal database record for each specimen that contains the barcode, filed-as name (and in some cases a couple additional fields)

### **Share and Identify Best Practices and Standards**

Digitizing institutions that do not already have a previously adopted digitization workflow in place are utilizing the tutorials prepared by NYBG for the imaging component. UNH is developing a series of short training videos for equipment setup and imaging. Once the Symbiota Portal is set up (should be within the next 2 weeks), we will develop training videos for label transcription.

### **Identify Gaps in Digitization Areas and Technology**

There is a database for algal taxonomy at Algaebase.org that is used as the "standard" reference by the phycological community. It is based on the Index Nominum Algarum at Berkeley, but with enhancements such as mapping names to currently accepted names. The Algaebase data is available as a webservice through WoRMS (World Registry of Marine Species) at VLIZ in Belgium. It would be very helpful if access to WoRMS could be built into the Symbiota for the Macroalgae Portal. The function would be to look up the filed-as name on WoRMS and automatically populate the currently accepted name, authorities, higher taxonomy, etc for each record.

Another enhancement that would be helpful would be tuning the GeoLocate for macroalgae. In initial tests, localities for marine macroalgal specimens will return a terrestrial point as the best guess. The frustrating part is that one of the alternate points it returns is correct. If it could bias its suggestions based on coastline awareness, it would improve georeferencing throughput.

### **Share and Identify Opportunities to Enhance Training Efforts**

As mentioned above, we are developing a series of training videos for each component of the workflow.

UNH sent a graduate student to the Georeferencing Train the Trainers workshop. In addition to georeferencing specimens at UNH, she will train georeferencers at other institutions in the Macroalgal Consortium.

### **Share and Identify Collaborations with other TCNs, Institutions, and Organizations**

### **Share and Identify Opportunities and Strategies for Sustainability**

### **Other Progress (that doesn't fit into the above categories)**