

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

Report submitted by: chdietri@illinois.edu

Report Submitted on: 07/05/2013 - 15:36

Progress in Digitization Efforts

So far, images and metadata have been ingested for 1097 vial racks and 2670 slide boxes. An image annotation tool was created to allow users to tag individual specimens in an image with metadata, including basic DarwinCore specimen-level data.

The robotic system for digitizing whole drawers of pinned insects has been undergoing rigorous testing. Some minor bugs in the machine control software have been fixed and adjustments were made to the hardware to reduce wear on parts. Work is progressing on automating the image capture workflow to reduce the possibility of human error. Stitching and image compression algorithms have also been optimized to improve speed and accuracy while reducing storage space requirements.

Share and Identify Best Practices and Standards

Nothing to report on this yet, but we anticipate having a complete workflow for drawer digitization by the end of summer. We will identify some best practices

Identify Gaps in Digitization Areas and Technology

Existing "off the shelf" stitching technologies don't work well for high resolution images of pinned insect drawers because there are often large gaps (white spaces) with few features that the stitcher can use as reference points. By using information on camera position combined with our use of a low-distortion telecentric lens, we should be able to overcome these problems.

Share and Identify Opportunities to Enhance Training Efforts

Nothing to report yet.

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)

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

Report submitted by: blieber@ku.edu
Report Submitted on: 07/09/2013 - 11:56

Progress in Digitization Efforts

Regarding the University of Kansas portion of the project, led by PI Bruce S. Lieberman and co-PI Una Farrell, we now have a total of 74,476 specimens databased. Of these, there is a total of 65,745 specimens databased that have clean, proofed localities. Further, we now have a total of 23,861 specimens that are georeferenced. In addition, a total of 2,799 localities have been georeferenced. In other relevant news, Co-PI Farrell presented an invited talk at the most recent SPNCH meeting at Rapid City, South Dakota in late June that described our Paleoniches project and also the work we have been doing associated with it. In addition, she and new post-doc Michelle Casey will be attending the georeferencing workshop at Yale University in September which will enhance our ability to train others.

Since the last update, PI Hendricks (San Jose State University; SJSU) and his undergraduate student assistants have continued to generate content for the “Digital Atlas of Ancient Life” and have put it online (currently accessible at <http://www.geosun.sjsu.edu/~jhendricks/AtlasTemp/>).

The current focus at SJSU has been on adding fossil invertebrate species to the Neogene page (<http://www.geosun.sjsu.edu/~jhendricks/AtlasTemp/neogene.html>) in collaboration with staff at the Florida Museum of Natural History. Six molluscan families have been added since the last update, including Muricidae, Personidae, Ranellidae, Strombidae, Tonnidae (all gastropod taxa), and Tellinidae (bivalves). In total, 137 species-level pages have been completed to date and a summary of these (with links to species-specific webpages) may be accessed at http://www.geosun.sjsu.edu/~jhendricks/AtlasTemp/neogene-all_species.html.

On May 3, 2013, student assistants (one undergraduate and two graduate students) involved with this project at SJSU presented a poster on the Digital Atlas of Ancient Life at the SJSU College of Science Research Day.

For the Ordovician portion of the project led by PI Alycia Stigall (Ohio U.), digitization of the Jack Kallmeyer Collection at Ohio University was completed on May 29, 2013. The collection currently contains 2397 lots with a total number of 12,271 individual specimens representing 268 species (from 156 genera and 100 families). The digitization was completed by two undergraduate students and one graduate student over the course of 19 weeks, averaging 126 lots per week.

From June until August, graduate student Hannah Brame will be quality checking the database, including the revalidation of specimen identification, adding additional details to locality entries, and fixing any visible errors in the collection object records. In addition, development of web content for the digital atlases will begin mid June. This work will include planning the organizational structure for the website images and content, generating datasets for range maps, producing identification guides for each major species (with special focus on bryozoa), and selecting representative specimens for photography. So far the initial data structure for the website has been developed and content pages for 14 species (primarily bryozoans) and geologic framework were generated.

At Miami University they have continued georeferencing the Shideler localities, although at a slower rate than during the school year as the undergraduate employees are off campus for the summer. About 200 localities were georeferenced.

The Cincinnati Museum Center has focused on databasing their back-log of paper catalog records. Over 1,200 new catalogue records were entered into KeEmu. Photography of an additional cabinet of type material is nearly completed.

Share and Identify Best Practices and Standards

Nothing new to add at this time.

Identify Gaps in Digitization Areas and Technology

Nothing new to add at this time.

Share and Identify Opportunities to Enhance Training Efforts

A new post-doc, Michelle Casey, as joined us at KU. She is working to expand our outreach efforts, and also help train undergraduates and others in georeferencing, and is also involved with integrating databasing across the different projects.

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

Nothing new to add at this time.

Share and Identify Opportunities and Strategies for Sustainability

Nothing new to add at this time.

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

Nothing new to add at this time.

MOBILIZING NEW ENGLAND VASCULAR PLANT SPECIMEN DATA TO TRACK ENVIRONMENTAL CHANGE

Report submitted by: p_sweeney@att.net
Report Submitted on: 07/09/2013 - 15:46

Progress in Digitization Efforts

Capture of collection level-information (i.e., “pre-capture”) and pre-digitization are the the primary activities. At institutions where pre-digitization activities are complete, pre-capture rate is proceeding apace, with more than 475,000 specimens with at least current identification captured. The development of the high-throughput digitization apparatus was completed in June 2013. The installation of one apparatus at is scheduled for mid-July. Light-boxes will be deployed in August. After installation, continued testing will occur when primary digitization will begin at some institutions by the end of August.

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

The project has received advice from NYBG (LBCC, MCC, TTD) about the use of light-box digitization set-ups. NEVP institutions that will use light-box digitization set-ups will meet with Michael Bevans (NYBG) in August to get an overview of the system and discuss installation. The project has worked with Ed Gilbert (LBCC) to deploy and test code added to Symbiota and the TcnImageTools to support the ingest of data transfer documents into Symbiota. We worked with the team at iPlant to develop a mechanism to transfer images to the primary image repository for NEVP (i.e., iPlant/TACC).

Share and Identify Opportunities and Strategies for Sustainability

nothing to report

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

nothing to report

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: 07/10/2013 - 09:41

Progress in Digitization Efforts

Specimens newly digitized: 18,773
Specimens added to the MycoPortal: 28.938
Total specimens in the MycoPortal: 1,424,718

Share and Identify Best Practices and Standards

1. We continue to deal with issues of data standardization, particularly names of collectors and collection numbers. We have distributed a new policy about the format for collector names [(First Initial). [Second Initial, if present]. [space] [Last name]. Collection number continues to be a problematic area. Many collectors use prefixes or suffixes – in some cases these are part of the unique identifier for the specimens, but in other cases not, and in some cases we have no idea. The Symbiota software does not provide separate prefix and suffix fields; we are hoping to convince the developer of the necessity for such fields.
2. We have changed our reporting protocol, requiring reports from participants every two months instead of every month. We are hopeful that after the experience of compiling their annual reports, participants now are more aware of the need to submit reports. We reiterate our request to have reporting best practices taken up by iDigBio so that it is more clear to all participants that regular record keeping and reporting.

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 is proving to be a very good source of help with label transcription. However, there is no U.S. based website that is currently ready to accept all of the images for transcription that our projects need. Notes from Nature is beautiful in design and appearance, however there is currently no support to help with the upload of images, or other technical issues. I have funding through the Macrofungi project to further develop Notes from Nature, but probably only enough to get my own project data on the site. It seems to me that iDigBio should try to help Notes from Nature realize its potential.

Share and Identify Opportunities to Enhance Training Efforts

1. We have hired two undergraduates summer interns for the project; in addition to learning digitization techniques, they are also taking weekly field trips with P.I. Halling to collect and identify macrofungi. The interns have created a website intended for public interest in the project (<http://www.themacrofungiproject.com/>)
2. An organizational meeting with the outreach coordinator Vernia, NCSLG curator Grand, Cubeta, and Middle Georgia State College (MGSC) collaborator Mozley-Standridge was held to discuss the logistics for the summer high school biology teacher's workshop. The workshop will be held at NCSLG 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.

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

- 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. Project Coordinator gave a presentation about the project at the annual meeting of the Society for the Preservation of Natural History Collections in Rapid City, SD in June 2

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

Report submitted by: moon@begoniasociety.org

Report Submitted on: 07/10/2013 - 10:51

Progress in Digitization Efforts

* 368,158 plant bug specimens digitized

* 595,320 plant records (more information and details

<https://docs.google.com/spreadsheet/ccc?key=0AiljXeu3fRq7dEJSMXVuTWVmdGNXWEp4dEdNV21EVEE#gid=0>)

* 2 major meeting occurred regarding insect record data upload to idigbio. One for images and one for data. We will be submitting our data first, and then associated images. Our external subcontracts (those not using our centralized database, Arthropod Easy Capture) are uploading either directly to idigbio or through SCAN TCN portal.

* Insect imaging workflow has been created. Images are collected from multiple institutions on AMNH server to deliver to idigbio.

Share and Identify Best Practices and Standards

* We are spearheading a host/association modeling and data standard MISC group to start at the beginning of August in collaboration with Symbiota and SCAN.

Identify Gaps in Digitization Areas and Technology

* Stronger support needed for sharing vocabularies. This can be many things but would include: taxon catalogs and shared georeferenced data.

Share and Identify Opportunities to Enhance Training Efforts

* Focus on graduate students. We have a proven successful outline for specimen data handling short course (<http://tcn.amnh.org/home/specimen-course>) that could be the foundation of a idigbio sponsored workshop.

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

* NYBG has begun a crowdsourcing pilot project for transcribing herbarium specimen labels from images, utilizing the Atlas of Living Australia's Biodiversity Volunteer Portal, which includes a portal dedicated to transcribing specimens for our Tri-Trophic Project (see email from Barbara Thiers below). This pilot project was initiated by Barbara in cooperation with the Osborne Association (<http://www.osborneny.org>), and will involve several participants at the Association's South Bronx location transcribing labels from images while developing their computer skills. We trained the participants earlier this month, and yesterday was their first day of work. Anyone anywhere is welcome and encouraged to transcribe specimens through the portal. For our project page, see this link: <http://volunteer.ala.org.au/project/index/2277905>

* AMNH (insects) collaborating with SCAN project to facilitate TAMU data upload, using SCAN-Symbiota as a pathway for TAMU data ingestion. Several discussions have already occurred to great benefit for both projects. It is helping with the assimilation and sharing of data as well as review of insect related data structures in both projects. We are spearheading a host/association modeling and data standard MISC group to start at the beginning of August.

Share and Identify Opportunities and Strategies for Sustainability

* Suggest funding opportunities for assimilated collection data for research. Create a RCN network for training graduate students to use these data for research.

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

1. Christine Johnson attended the SPNHC meeting in South Dakota, June 18-21, where I presented AEC in the DemoCamp session.
2. AMNH curation as a result of project momentum (not directly funded by the TCN). Christine Johnson working on organizing/updating cicadellid collection according to current taxonomy with intern Devin Kreitman.
3. AMNH REU student Jeremy Frank started this month. He is being supervised by Christine Johnson, Randall Schuh and Katja Seltmann on using a specimen-level database to examine ecological patterns.
4. Melissa Tulig and Kim Watson attended the SPNHC meeting in South Dakota, June 18-21. Stephen Gottshalk (a NYBG co-worker) and I presented on the overall workflows being used for our rapid digitization projects, including how we're working to populate skeletal records. See TTD-TCN website under products:
<https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbX0dGR0Y258Z3g6MzEzNzlmN2ViYzFhOWJmMA>
5. Rob Naczi presented at the Monocots V conference yesterday at the informatics symposium.
6. Our Facebook page now has over 100 followers. Please visit: <https://www.facebook.com/pages/Insects-Plants-and-Parasites-Digitizing-Natural-History-Collections/330373400373054>

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: 07/10/2013 - 14:27

Progress in Digitization Efforts

Our digitization efforts are continuously updated at <http://symbiota1.acis.ufl.edu/scan/portal/collections/misc/collstats.php>
As of July 8, 2013 we have the following basic statistics on specimen records in SCAN Symbiota to report:

- 469,420 specimen records in database
- 271,525 (58%) records georeferenced
- 307020 (65%) records identified to species
- 1,014 families
- 7,861 genera
- 16,689 species

Share and Identify Best Practices and Standards

Previously Reported

Identify Gaps in Digitization Areas and Technology

Previously Reported

Share and Identify Opportunities to Enhance Training Efforts

Previously Reported

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

We have collaborated with Tri-Trophic and iDigBio on a regular basis. Specifically, we have worked with Pam Soltis to promote the creation of a new online journal that would provide a focused venue for publishing research utilizing TCN data. We have recently collaborated with Katja Seltmann (Tri-Trophic) working on data schema and data ingestion workflows among SCAN, iDigBio, and Tri-Trophic.

Share and Identify Opportunities and Strategies for Sustainability

Previously Reported

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

We are currently planning a production mode launch of specimen images by the end of July. We will implement a network-wide beta test for FilterPush by the end of July.

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

Report submitted by: cgries@wisc.edu
Report Submitted on: 07/11/2013 - 14:50

Progress in Digitization Efforts

from our annual report:

In this second year of funding more digitization centers have come on line while a few were finishing up their commitments. After initial hurdles the project is now in the production stage and the overall process has become an efficient routine for most imaging centers. Workflows and general requirements are well defined and are documented on our project website: <http://lbcc.limnology.wisc.edu/node/3> . Most digitized records are available in one of the two portals <http://lichenportal.org> and <http://bryophyteportal.org>. Only a few institutions are still in the process of quality controlling before uploading images and are currently storing them locally. Most of the digitizing centers that were funded from the beginning have now moved on to include specimens from small collections in their digitization process. This past year saw the addition of ~800,000 new records (lichens: 320,000, bryophytes: 480,000), 600,000 of which were imaged, the other 200,000 were integrated from previously existing local databases. Overall, the lichen portal now provides access to almost 1 million specimen (989,367 as of June 2013) records and the bryophyte portal to over 1.4 million (1,432,775 as of June 2013) records. Overall the number of herbaria represented has increased to 44 institutions across the two portals. Staff at WIS has been assisting these new collections with setting up their image upload routine, collections profiles in the respective portals and integration of existing databases. All images, the databases and portals are currently housed on servers at iDigBio in Florida.

Share and Identify Best Practices and Standards

nothing to share at this point in time

Identify Gaps in Digitization Areas and Technology

Since improving OCR seems so hard, maybe more emphasis should be put on voice recognition. Reading a hand written label out loud is still faster than typing the information.

Share and Identify Opportunities to Enhance Training Efforts

We are planning extensive Symbiota training, following iDigBio's example of train the trainer.

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)

Not really progress, but here is the discussion of digitization costs from this year's annual report: Estimating cost of digitization has been attempted by several collaborators. However, the numbers need to be discussed carefully as it is very difficult to standardize everything that should be considered 'cost of digitization'. Currently digitization efforts include the pulling of specimens, minimal preparation for imaging (e.g. sorting by species and locality, barcoding), actually taking one or more photos of the label, entering some (but varying amounts of) label information into a metadata file (species, collector, collection number, country, state) and finally returning the specimen to the collection. It does not include the process of transcription yet, i.e., the record is not complete in a database. Included in most calculations is time required to do some quality control and uploading of the images. The cost estimates range from \$ 0.3 to \$ 1 per specimen

for places where the digitization workflow has become routine and is mostly performed by undergraduates with some supervision from a senior person. Before this stage is reached and while workers are trained, the cost per specimen is higher. It is not clear if these calculations include University overhead and benefits in all cases.

These numbers do not include the support effort that is expended by the LBCC technical staff or iDigBio's expenses for setting up and maintaining the server. In addition, the cost of developing software before and adapting the software during this project is not included. However, most digitization centers seem to operate well within the proposed budget.