

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

Report submitted by: blieber@ku.edu
Report Submitted on: 05/04/2013 - 12:57

Progress in Digitization Efforts

Regarding the KU component of the project, we have completed databasing and georeferencing all of our trilobite collections associated with the project. We have nearly completed databasing all of our brachiopod collections associated with the project. Thus far, since the project began we have databased 64,126 collection objects. Regarding localities, over the years in our collections locality information has been recorded first on locality cards. We recognized that the most efficient way to georeference would be to proof read all of our locality cards first to make sure that they contained accurate information and to eliminate any redundancies in potentially geographically (but not stratigraphically) overlapping localities; this would allow us to avoid georeferencing the same locality multiple times and will also allow batch georeferencing, greatly speeding up the pace at which we can georeference. We have finished proofing 80% of the pertinent locality cards (several thousand) and are thus nearly completed with this task such that we have databased 55,395 specimens with clean localities. The locality proofing will allow us to greatly ramp up our georeferencing pace. Still though, since the project began 19,737 of the databased collection objects are georeferenced with lat-long and error estimates and that equals 1,590 georeferenced localities. We have also provided photographs for a limited number of species (principally trilobites and also some brachiopods) to be associated with our collections data in Specify so that they can be accessed on line and also so that they can be used in the creation of the Digital Atlases. A new post-doc has been hired and will start June 1. A new graduate student has been hired for the summer and will start working the first week of June.

Regarding the San Jose State component of the project,

Activities at SJSU have thus far been focused on the development of the "Digital Atlas of Ancient Life" webpage, which may be currently accessed at: <http://www.geosun.sjsu.edu/~jhendricks/AtlasTemp/index.html>. The major components of the Neogene portion of the website have been designed and species level pages currently exist for 101 taxa from the Neogene time period (relevant data have been provided to SJSU project participants by the staff of the FLMNH). Each individual species webpage provides an overview of the higher-level classification of the species, its geological range, its stratigraphic occurrences, comments on its paleogeographic distribution, and detailed digital maps (constructed from georeferenced data) for multiple time slices, which show how the species' range may have changed over time.

Regarding the Ordovician component of the project,

Ohio University:

In the Ohio University Kallmeyer Collection, a new Specify Database was generated

including development and augmentation of taxon, locality, and stratigraphy trees to house the newly digitized specimens. To date, 1,295 unique digital records have been created for specimens including 83 families, 121 genera, and 201 species.

Georeferencing has been accomplished on 91% of our localities. The total number of specimens digitized is 6,711, representing 67% of the total collection. Content generation has begun for the Digital Atlas pages of some species, notably bryozoans. The project was presented in two poster presentations at Ohio University, one for the Student Expo, where the poster by the students won top place in the Geology-Geography-Math Division, and a second time at the Ohio University Geological Sciences Alumni Symposium. These two opportunities showcased the project to the larger university community, including the President of Ohio University. Miami University:

The Specify program was installed on the grant-supplied computer and the Ordovician databases provided by Ohio University were installed as well. Co-PI Hauer decided to start with the Ohio localities cataloged by William H. Shideler, who from 1920 until about 1950 collected, and oversaw the collection of, the majority of the Limper Museum's Ordovician fossil specimens. To date, nearly all of Shideler's Ohio localities have been georeferenced. This amounts to approximately 250 mapped localities that encompass about 1050 distinct entries in Shideler's coded

locality list for Ordovician fossil collection locations in Ohio. A dozen or so problematic Ohio localities remain to be completed. Hauer and students have just started on Shideler's Indiana localities.

Cincinnati Museum Center:

Work on the project began in January, 2013 –The Curator of Invertebrate Paleontology (Brenda Hunda, Co-Investigator) hired two undergraduate students from the University of Cincinnati in December of 2012 and officially they began in the new year. Since January, approximately 2,400 records have

been entered into KeEmu. In addition, over 1,300 specimens have been photographed and prepared for entry into KeEmu.

Share and Identify Best Practices and Standards

Nothing new to report here.

Identify Gaps in Digitization Areas and Technology

Nothing new to report here.

Share and Identify Opportunities to Enhance Training Efforts

Participants have been involved in a variety of workshops and training endeavors. Una Farrell at KU in particular has been trained to teach folks georeferencing. Specify staff trained individuals from Ohio University and Miami University in use of Specify software. Subsequently, Ohio University graduate student Hannah Brame completed iDigBio's Train the Trainers Georeferencing Workshop in October, 2012. She hosted two formal georeferencing workshops: (1) Paleoniches TCN Georeferencing Workshop,

focusing on paleontological considerations for data digitization and georeferencing collecting localities, and training NSF Digitization TCN grant collaborators to georeference museum collections. Cincinnati Museum Center (Hosted by Brenda Hunda, collections manager in association with iDigBio) 1 February 2013; and (2) Georeferencing: Turning words into mappable data points; Determining latitude, longitude and uncertainty from written locality descriptions with considerations

for data collection and field notes. Ohio University Department of Geological Sciences 11

March 2013. Dr. Brenda Hunda, received georeferencing training through Alycia Stigall at Ohio

University. A day long workshop was held at the Cincinnati Museum Center for CMC staff as well as Co-Investigator Kendall Hauer. The Curator at CMC then trained two undergraduate students in the techniques of georeferencing, which will be ongoing. Similarly, Co-PI Hauer trained two undergraduate students in georeferencing techniques.

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

There are new PENs to our network from Yale and University of Texas

Share and Identify Opportunities and Strategies for Sustainability

Nothing new to report here.

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

"Digital Atlas of Ancient Life" webpage, which may be currently accessed at:
<http://www.geosun.sjsu.edu/~jhendricks/AtlasTemp/index.html>.

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: 05/08/2013 - 12:28

Progress in Digitization Efforts

Specimens newly digitized (includes labels digitized): 14,062
Records completed (or completed records entered): 7798
Field book records created: 522
Specimens georeferenced: 1076
Specimens added to the MycoPortal: 286,159
Total specimens in the MycoPortal: 1,386,541
Mycoportal records edited: 384,502

Share and Identify Best Practices and Standards

1. We developed some low tech but practical guides to help participants interpret collectors names. Basically this procedure consists of creating an authority file of collector names that can be inserted as a lookup table in the Mycoportal, eventually, but for now is in the form of a crib sheet for data enterers as well as a spreadsheet that can be customized for each institution.
2. We continue to try to instill best practices about by participants with regard to record-keeping for this project. It continues to be an uphill battle to get institutions to record their work on a regular basis in order to prepare monthly reports. I would really appreciate some support from iDigBio and/or NSF in the importance of record keeping, especially in light of the fact that NSF reports for this project require this information, and the calculation of a per-specimen cost for digitization.

Identify Gaps in Digitization Areas and Technology

We have identified some issues in searching in the Symbiota software that powers the MycoPortal. Now that we have a significant number of records in the Portal, we have found that searches fail if collection prefixes and suffixes are included in the search, and that searches on collector last name fail when the record includes collector initials or full first names. We are in discussions with Ed Gilbert about the problem.

Share and Identify Opportunities to Enhance Training Efforts

1. Revised the project manual, in anticipation of the upcoming training sessions

2. Organized two training sessions for seven institutions to be held in at NYBG in May
3. Submitted article for *Miclavinea*, journal for amateur mycology, entitled, "Making Scientific Vouchers of North American Macrofungi" with co-authors Nathan Wilson, of EOL, and co- PI Roy Halling.
4. Hired two new interns for the summer. Previous intern Ana Maria Ruiz was re-hired as a part time worker on the project.

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

NONE

Share and Identify Opportunities and Strategies for Sustainability

NONE.

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

Thiers gave a presentation about the project at the Mid Atlantic States Mycology Conference, held Saturday, April 27, 2013 at the USDA Agricultural Research Center in Beltsville, MD. The title was "The Macrofungi Collection Consortium: A Progress Report."

MOBILIZING NEW ENGLAND VASCULAR PLANT SPECIMEN DATA TO TRACK ENVIRONMENTAL CHANGE

Report submitted by: patrick.sweeney@yale.edu

Report Submitted on: 05/10/2013 - 10:23

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, and two institutions (UMASS, NHA) are on schedule to be completed by the next reporting period. At institutions where considerable pre-digitization curation (e.g., updating taxonomy, sorting, barcoding) are happening in conjunction with pre-capture, the pre-capture rate is slow. The imaging of specimens and capture of specimen-level data, will begin in July/August, after the conveyor belt system and light-boxes are deployed.

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 June to get an overview of the system and discuss installation. The project has worked with Ed Gilbert (LBCC) to deploy the CNH Symbiota instance on a iDigBio/UFL server. Currently working with iPlant on image and data transfer mechanism between apparatus and iPlant.

Share and Identify Opportunities and Strategies for Sustainability

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

Hardware & Software Development

The engineering group at the University of Oklahoma have been focused on testing the conveyor system and refining the controller station Server-Client application.

Software development:

We have further refined the UI interface components and have integrated Dragon Voice Recognition software to interface with entry UI elements. We have found out that with very limited software training, the software does a good job of entering collection dates and towns. Significant time savings achieved. Initial tests indicate an average of 25 - 30 seconds taken per specimen with complete specimen meta-data entry - best 15 seconds and worst 50 seconds. This worst case was when the voice recognition software made collector name entries wrong.

Hardware development:

The conveyor system, camera, master computer controllers are all working fine and well integrated. Movements of the conveyor belt and camera are coordinated with the data entry system. Contrast sensor implemented to ensure that specimens do not fall out of the conveyor belt. The Hardware/Software development team at OU will deploy the apparatus at Harvard in June/July.

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

Report submitted by: cgries@wisc.edu
Report Submitted on: 05/13/2013 - 12:57

Progress in Digitization Efforts

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

Lichens: <http://lichenportal.org>

Herbaria actively submitting images or key stroked records to the portal: 39 (up by 5 since March 2013)

Specimen records in portal: 961824 (up by 39441 since March 2013)

Specimen records with images: 935598

Bryophytes <http://bryophyteportal.org>

Herbaria actively submitting images or key stroked records to the portal: 28 (up by 3 since March 2013)

Specimen records in portal: 1380757 (up by 53694 since March 2013)

Specimen records with images: 392214

A number of images have not been uploaded by collaborating institutions and are expected in the near future, as we are assured they have been captured and are managed locally. Specimen records are increasing through imaging of labels and through integrating existing databases. In contrast to last month, from now we are reporting records that have images attached to ease the statistics as there may be several images per record.

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

Digitization of labels from lichen and bryophyte specimens is fairly straight forward, and although work intensive there are no obvious gaps in digitization technology. Optical Character Recognition of course is a different matter and progress there is slow. Quality of OCR can be improved by proper lighting and orientation of labels during digitization, but beyond that it is the OCR technology that leaves room for improvements.

Share and Identify Opportunities to Enhance Training Efforts

None in this reporting cycle

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

The LBCC is a collaborative project between 18 institutions, each of which collaborates with a number of smaller collections. Several institutions were added through successful PEN proposals. Overall, 76 collections are actively participating in the project

Share and Identify Opportunities and Strategies for Sustainability

Each collection is currently maintaining the data for their records. However, the infrastructure to make these data available could be picked up by professional societies interested in the subject. As we move from paper publications to electronic publications this is a natural shift and the cost is comparable.

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

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

Report submitted by: chdietri@illinois.edu

Report Submitted on: 05/14/2013 - 09:49

Progress in Digitization Efforts

Collaborators at 8 institutions have, so far, uploaded 6720 images representing 2308 slide boxes (up to 100 slides/box) and 642 vial racks (up to 48 vials/rack) and tagged these with basic metadata. Specimen-level metadata capture has not yet begun, but we have developed a tool that allows users to annotate and tag individual specimens within each image. This tool is currently on our test bed and we anticipate making it public within the next few weeks. Our infrastructure team is exploring different options for crowd-sourcing image annotation and capture of specimen label data. We also continue to test and refine our digitization workflow for whole drawers of pinned insects. Senior collaborator David Raila (UI Computer Science) demonstrated our prototype robotic drawer digitizing system at the Pinned Insect Digitization Workshop in Chicago. David has nearly completed work on a real-time stitcher that will allow multiple high resolution images to be combined (stitched) as they are being captured by the robot, reducing the time required to digitize each drawer to 5 minutes or less.

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

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: rnaczi@nybg.org
Report Submitted on: 05/14/2013 - 13:51

Progress in Digitization Efforts

In the last bimonthly report, we emphasized activities of the entomological half of our project. This time, we report on the botanical activities. The entomological institutions have continued to progress at a good pace, and we have no problems to report from them.

All 14 botanical institutions are actively digitizing, except for Eastern Michigan University, which plans (and has always planned) to begin its work in July 2013. Most of the effort at this time has been with imaging herbarium specimens. We report the following digitization accomplishments for the past two months.

Lead institutions

New York Botanical Garden: A total of 138118 specimens have been imaged, with 15021 during the past two months.

University of Michigan: A total of 60886 specimens have been imaged, 9880 during the past two months.

Missouri Botanical Garden: Since many of the specimens had already been imaged, the emphasis is on georeferencing, barcoding, and databasing. To date, these activities have been completed for 50,500 specimens, approximately 3100 of which were during the past two months.

Subcontracting partners

Miami U, U of Maine, and U of Texas (coordinated by New York Botanical Garden): 35595 specimens imaged to date.

U of Minnesota, U of Wisconsin (coordinated by U of Michigan): 76750 specimens imaged to date.

Illinois Natural History Survey, Iowa State University, U of Colorado, U of Illinois, U of Kansas (coordinated by Missouri Botanical Garden): 99,000 sheets imaged to date

Another area of active progress is integration of data from the three trophic levels (plant hosts, hemipteran herbivores, and parasitoid hymenopterans). Toward this end, NYBG has provided DiscoverLife access to 127000 images of plant specimens. For these images, DiscoverLife has generated derivative images at various resolutions appropriate for online posting, and labeled features on the specimen images.

On the curation front, Rob Naczi and Rich Rabeler curated grass specimens at NYBG and U of Michigan, respectively, in preparation for their digitization. In addition, they compiled a list of accepted names for North American Dichanthelium grasses (Poaceae). The genus Dichanthelium is one of the largest genera of grasses in North America, and usually judged

the taxonomically most complicated. Also undergoing curation in advance of digitization are specimens of Asteraceae at NYBG by Kim Watson and Charlie Zimmerman.

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

Kimberly Watson (NYBG) delivered several presentations to train and share best practices in digitization, especially georeferencing, as follows.

6 Mar: Assisted with georeferencing workshop at NYBG for digitizing staff, graduate students, and colleagues at AMNH.

13 Mar: Conducted georeferencing workshop at NYBG for digitizing staff, graduate students, and colleagues at AMNH.

12 Apr: Presented “Plants, herbivores, and parasitoids: Tri-trophic digitization strategies” at Association of Southeastern Biologists (ASB) meeting symposium: Workflows and Challenges in the Digitization of Biological Specimens. Charleston, WV (https://www.idigbio.org/wiki/images/2/2a/WATSON_ASB_TriTrophic_Digitization_Strategies.pdf).

13 Apr: Attended iDigBio Digitization Workshop following Association of Southeastern Biologists (ASB) meeting. Charleston, WV.

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

Rich Rabeler and Melissa Tulig are members of the iDigBio Biodiversity Informatics Manager Working Group, which held it's first meeting in March. Rich interacts regularly with U of Michigan personnel working on the Macrofungi and Lichens and Bryophytes TCNs.

Share and Identify Opportunities and Strategies for Sustainability

nothing to report

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

nothing to report

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

Report submitted by: neil.cobb@nau.edu

Report Submitted on: 05/16/2013 - 10:44

Progress in Digitization Efforts

We are on target to meet our first year quota to digitize labels from pinned specimens. We have 270,294 specimen records in the database, representing 510 families, 5,001 genera, and 9,943 species. Sixty percent of the records are georeferenced. We have an additional 141,000 specimen records in our database from museums that are not funded by

NSF-ADBC funding. These represent 728 families, 3,576 genera, and 7,386 species. We have also uploaded over 800 high-resolution specimen images.

Share and Identify Best Practices and Standards

We have our protocols available at <http://scan1.acis.ufl.edu/> .

Identify Gaps in Digitization Areas and Technology

The biggest gaps currently include a standard system for imaging specimens, creating a system for crowdsourcing data to be georeferenced, and data sharing workflows beyond the SCAN TCN.

Share and Identify Opportunities to Enhance Training Efforts

'nothing to report'

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

'nothing to report'

Share and Identify Opportunities and Strategies for Sustainability

We have identified five options for sustaining digitization efforts beyond SCAN-ADBC funding that apply to each institution. These options do not require collaborations with other museums, they would complement targeted external funding efforts to maintain a digitization effort. Individually, these options will not be adequate to meet future demands, they may be sufficient if they are collectively implemented.

1. Build into project cost (new material)
2. Museum operation costs
3. Internal institution funding opportunities
4. Biodiversity informatics related classes
5. Independent student research

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

'nothing to report'