**SPNHC 2014 Symposium:**

**Update on Initiatives and Progress in Digitization of Natural History Collections**

Collected Abstracts for Thursday 26 June, 2014

Speakers names are in **bold** below (please check).

**8:50 – 9:00 am**

**Introduction** to purpose and format of the symposium and its relation to the [Friday Special Session](http://www.museumwales.ac.uk/media/31181/SIG-Digitization-International-Collaboration-SPNHC2014.pdf) on Collaboration.

**Conveners:** Deborah Paul (Florida State University – iDigBio), Elspeth Haston (Royal Botanic Garden Edinburgh), Elizabeth Ellwood (Florida State University – iDigBio Post Doc)

**P1**

**9:00-9:20 am**

**Developing and Testing Tools and Processes for Creating a Swedish Digital Natural History Collection, e-BioColl.se**

Per Ericson1, Fredrik Ronquist1, Anders Telenius1, Stefan Daume1, Kevin Holston1, and **Karin Karlsson**1

(Swedish Museum of Natural History)

**Abstract:**

Natural history collections constitute the major source of data for research in phylogenetics and systematics, and provide crucial information for analyses of long-term trends in ecology and environmental sciences. Recent mass-digitization efforts have opened up such collections abroad for transformative e-science approaches, but research dependent on Swedish objects is left behind. Out of an estimates 33 million Swedish specimens, almost 90% are unavailable digitally, and it will take hundreds of years to eliminate this backlog at the current pace of digitization.

To address this problem and provide a proof-of-concept version of an e-science platform for Swedish natural history collections allowing complete digitization within 10 years we plan to develop and test standardized, re-usable, industrial-level digitization processes where innovation and automation will bring down the cost of obtaining specimen images and geospatial and taxonomic metadata from specimen labels. Robotic techniques employed to image insect collections, and automated text recognition and citizen-science based crowd sourcing for extracting information from specimen labels during the digitization of 1.4 million specimens in the Gothenburg herbarium will provide a full-scale feasibility test of tools and concepts. The digitized information will immediately support outstanding research in Sweden and elsewhere on this unique material, and boost international initiatives (e g GBIF, LifeWatch and IPBES).

**P2**

**9:20-9:40 am**

**From massive digitisation of Paris Herbarium to a nation-wide program**

**Simon Chagnoux** 1, Marc Pignal 2 (Muséum national d'Histoire naturelle of Paris)

**Abstract:**

The Muséum national d'Histoire naturelle of Paris (MNHN) completed a massive digitisation program of the herbarium specimens (P, PC). Nearly 6 million images went online between 2008 and 2012. The resulting database records contain only a minimum set of attributes (scientific name, catalog number and continent of origin). As many more information is most often available from the photographed labels, optical character recognition and a crowdsourcing site have since been used to enrich the database without physical access to the specimen.

Building on the experience gained from these projects, a national program “e-ReColNat” was launched to extend the digitisation to the remaining herbariums and most zoological and paleontological types. The result will be available online for annotation and label retranscription to a broad audience mixing the general public and the scientific experts.

**P3**

**9:40-10:00 am**

**Barcodes, conveyor belts and laser scanners: putting the contents of the Geological Museum (South Kensington, 1935 - 1985) on the web**

**Mike Howe,** Bob McIntosh, Simon Harris, Michela Contessi & Graham Tulloch (National Geological Repository, British Geological Survey)

**Abstract:**

The Museum of Practical Geology (Jermyn Street, 1851-1935) displayed most of the British Geological Survey’s specimen collection. With the move to the Geological Museum and then to Nottingham (Keyworth, 1985 onwards) most items were consigned to drawers. The internet has provided the means to re-exhibit these collections. BGS has been at the forefront of collections digitization since the 1970s. By 2000, the collections had digital catalogues of varying completeness, and in October 2000 the borehole and rock catalogues went on line. By 2006, all the main collections had online text and GIS searches.

During the past five years, BGS has made high resolution digital images of various collections available, including building stones and 4000 early twentieth century glass plates of the Kidston palaeobotanical collection. 125,000 7216 x 5412 pixel images of UK Continental Shelf cores1 were made available online in 2012, a project requiring a conveyor belt and extensive use of barcodes. 100,000 rock thin sections2 were added in 2013. The most innovative project to date has been a collaboration between BGS, the National Museum Cardiff, the Sedgwick Museum Cambridge, the Oxford Museum of Natural History and the Geological Curators’ Group (representing other national, university and local museums) to create an online database3 of British macrofossil types. The web portal provides data about each specimen, searchable on taxonomic, stratigraphic and spatial criteria. High resolution photographs, stereo anaglyphs and many 3d digital models are available. We are keen to share our experiences, particularly of 3d digital models.

1 <https://www.bgs.ac.uk/data/offshoreWells/wells.cfc?method=searchWells>

2 <http://www.largeimages.bgs.ac.uk/iip/britrocks.html?id=250000/258489>

3 <http://www.3d-fossils.ac.uk/>

**P4**

**10:00-10:20 am**

**Automated mass-digitization line for individual insect specimens**

**Riitta Tegelberg**1, Janne Karppinen, Tero Mononen, Mira Sääskilahti, Hannu Saarenmaa

(Digitarium, SIB labs, University of Eastern Finland)

**Abstract:**

We present a novel solution for mass-digitization of individual insects. Each specimen will receive a unique ID, and the basic metadata including taxon name will be databased with the ID. A specially designed conveyer belt is being used in the imaging. On the imaging line, plastic pallets made with 3D-printers carry the specimens to the imaging unit. The insect is automatically imaged from above, and the labels attached to the pin are imaged from the side. The underside of labels will be visible through a mirror. At the moment, imaging line is applied to objects between 0.5 – 30 mm in length, and the peak performance is around 500 specimens /day.

This system is being used to process collections donated to the Finnish Museum of Natural History. Digitarium receives collections, and disassembles them, driving them through the automatic digitization line and placing them in units. After digitization, the specimens and produced data will be delivered in taxonomical order to FMNH. The original collection cabinets and drawers will not follow the specimens to the museum. From images, further metadata from the labels can be transcribed.

**P5**

**10:20-10:40 am**

**Image segmentation in high throughput digitisation workflows**

**Vladimir Blagoderov**1, Laurence Livermore1, Ben Price1, Stéfan van der Walt2, Pieter Holtzhausen2 and Vince Smith1 (Natural History Museum, London, UK)(Stellenbosch University, Stellenbosch, South Africa)

**Abstract:**

Advances in imaging technologies have made the acquisition of high throughput, multi-specimen images a trivial task. However, the subsequent movement of specimens (e.g. due to curatorial or research activity) breaks the synchronisation between the physical collection and its digital representation, degrading the value of multi-specimen images over time. Furthermore, the data record of a single specimen cannot be unambiguously linked to its digital image with this approach. In order to realise the full potential of multi-specimen imaging technology, software is required to automate or semi-automate the segmentation of images into individual specimens and support subsequent annotation. We demonstrate a potential approach for two types of specimens: microscopic slides and pinned insects. This approach will form part of the large scale digitisation programme being developed at the NHM which has the ambitious goal of digitising 20 million specimens in the next five years.

**10:40-11:00 am Break**

**P6**

**11:00-11:20 am**

**Workflows in the cooperative IMLS Silurian Digitization project between Milwaukee Public Museum and The Field Museum.**

**Patricia Coorough Burke**, A. Caywood, M. James, E. Malueg & D. Miller. (Milwaukee Public Museum)

**Abstract:**

Milwaukee Public Museum (MPM) is implementing a three-year IMLS grant in conjunction with The Field Museum to digitize our Midwest Silurian collections and create a shared online database. The first workflow was between the institutions. The joint nature of the project required information exchange between museums. To coordinate a shared online database an effort was made to use shared vocabulary and Internal Record Numbers (IRN) for lithostratigraphy, chronostratigraphy and taxonomy data. Internal Records were generated at The Field Museum and shared with MPM. After the first large data export smaller updates have been made. MPM maintains a shared Google Drive spreadsheet with corrections and new data to be updated at The Field. Sharing data creates an error check point.

The second workflow for the digitization project is within MPM. The collection is organized stratigraphically. The specimens all have a locality number but most do not have a unique specimen number. Workflow starts with data entry into KE EMu using specimen labels. Specimens without numbers are assigned the KE EMu Internal Record Number as specimen number. The photography station is divided into two parts, fossil photography and photo processing. The fossils are processed by drawers and include a photo of the drawer and photos of each fossil in the drawer. A batch process is used to convert the raw image to .dng, .tif and .jpeg formats. The .jpeg images are then associated with the KE EMu catalogue data and the .dng and .tif images are saved in MPM’s image repository.

**P7**

**11:20-11:40 am**

**Designing a Workflow to Help with Error Detection in a Paleontology (IMLS Silurian Reef) Digitization Project**

**Paul S. Mayer**, L. Connolly, N. Karpus, & A. P. Layng (The Field Museum)

**Abstract:**

The Field Museum is implementing a three-year IMLS grant using summer interns to digitize approximately 15,000 Silurian fossil lots. Workflow for this project has each intern working on six drawers of fossils at a time; the fossils are all from the same geologic period and taxonomic group (as they are arranged in the collection). Interns cycle their fossils through three workstations: a label photography station, a fossil photography station, and a KE EMu data entry station. The average time per specimen to complete this cycle is 9.5 minutes with times ranging between 4 to 18 minutes. After cycling their cart through each workstation I edit, upload, and link the images to the catalogue records. In the first year 7,785 image were generated and linked to catalogue records.

During the process of connecting images to the catalogue there are three opportunities to check for errors. First, KE EMu generates an error log of images that fail to link during the batch input. The error rate is less than 2.2%. Common errors include simple format errors in the batch upload, specimens not entered into the catalogue, and typos in the specimen’s catalogue number. Second, because specimens in each batch are from the same geologic period and taxonomic group it is easy to spot any inconsistencies or missing data when reviewing the records in a table format. Finally each record should have two images linked to it. Any missing images are easy to add.

**P8**

**11:40-12:00 pm**

**Making molehills out of mountains: crowdsourcing digital access to natural history collections**

**Laurence Livermore**, Vince Smith and John Tweddle (Natural History Museum, London, UK)

**Abstract:**

The development of cheap mass imaging techniques has made transcription the bottleneck in digitising natural history collections. Internet crowdsourcing has been successfully used across different scientific communities and has the potential to become an effective method for transcribing our natural history data. We review current crowdsourcing platforms and communities, analyse their efficacy and discuss their potential role in improving digital access to natural history collections worldwide.

**P9**

**12:00-12:20 pm**

**Exploitation of digital collection data at the Museum für Naturkunde Berlin**

**Saskia Jancke**, Dirk Striebing, Frieder Mayer saskia.jancke@mfn-berlin.de

(Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science)

**Abstract:**

Item information of many collections in the Museum für Naturkunde Berlin (MfN Berlin) - as in many other museums - is often stored locally by each curator in different formats such as Excel spreadsheets. These files are often accessible exclusively by collection staff members. Within the project “Exploitation of digital collection data” funded by the Deutsche Forschungsgemeinschaft (DFG) the data from the MfN Berlin mammal collection were transferred from Excel spreadsheets to an SQL server using scripts developed by a database specialist at the MfN Berlin. The data were subsequently standardised, for example in terms of their taxonomic and geographical information. The revised data were then transferred from the SQL server into Specify - a museum database software application. Further aims of this project are the development and implementation of common transfer tools to achieve data migration to open-access databases such as BioCASE and GBIF as well as information retrieval like the current distribution of the specimen from databases like the IUCN redlist. This will allow external information retrievals of collection data and thus will open new avenues for scientific exploration of the collections. We have successfully applied our data transfer pipeline to the mammal collection of the MfN Berlin which is the 4th largest worldwide of its kind. These methods and tools can be used for the data migration of other collections of the MfN Berlin with its approximately 30 million collection objects, and also by other museums.

**P10**

**12:20-12:40 pm**

**ZooSphere - A tool for automated spheric image capturing and interactive 3D visualization of biological collection objects**

**Alexander S. Kroupa**1, Martin Pluta1, Bernhard Schurian1 & Falko Glöckler1 (Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science)

**Abstract:**

Entomological collections, particularly the type specimens therein, are the fundament of the taxonomic research on insects. Since the beginning of collecting activities in the 18th century, some million type specimens (e.g. holotypes, lectotypes, paratypes.) are stored in museums worldwide. There are various threats to these collections: Valuable type specimens are damaged or even destroyed by museum beetles (e.g. Anthrenus verbasci) or moths (e.g. Teneola biselliella). Also changes and fluctuations in light exposure, temperature and high humidity may have negative effects on the objects. Especially the types are even more in danger by shipping them from the museums to specialists for their taxonomic work.

To give full scientific access on the type specimens without shipping the valuable material across the globe, we developed a tool (ZooSphere) for automated image capture from about 400 positions around one specimen. A high resolution image with extremely high depth of focus is produced by using stacking technology in each position. The images will be freely available in the world wide web and will be displayed by a web-based tool developed for rotating and zooming the digital object in browsers. The same technology will be used for an interactive 3D presentation of the objects.

The next steps are 1) joining the ZooSphere images with microCT images from internal specimen structures and 2) creating 3D models of the specimens by applying photogrammetric methods on the ZooSphere images.

**12:40 – 14:00 pm Lunch**

**P11**

**14:00-14:20 pm**

**Using optical character recognition (OCR) output in digitization: see your data before it's in the database and after**

**Deborah L Paul**1, Andrea Matsunaga2, Miao Chen3, Jason Best4, Sylvia Orli5 & Elspeth M Haston6

(Florida State University, University of Florida, Indiana University School of Informatics and Computing, Botanical Research Institute of Texas, Smithsonian Institution, The Royal Botanic Garden Edinburgh)

**Abstract:**

Before iDigBio, that's Integrated Digitized Biocollections, others began the work of figuring out how to use OCR output with machine learning (ML) and natural language processing (NLP) to improve the efficiency and speed with which data from museum specimen label images can be captured and validated. The Augmenting Optical Character Recognition (aOCR WG) Working Group at iDigBio is pleased to be building on that foundation. Improvements have been realized in parsing algorithms and visualization of data. Recently, researchers at The Royal Botanic Garden Edinburgh (RBGE) successfully used word clouds from OCR output to reveal useful data, otherwise dark, until a specimen label was digitized completely (Haston, et al TDWG 2013). Their work indicates this method results in greater transcriber job satisfaction. Inspired by this work, the OCR Integration Track Team at the iDigBio Citscribe Hackathon showed how indexing, scoring, and visualizing OCR output reveals otherwise hidden search terms, uncovers errors, and can improve the data transcriber and data validator experience. Using open-source software, we presented these ideas to those with transcription tools up-and-running including Notes From Nature, Biodiversity Heritage Library (BHL), ALA Biodiversity Volunteer Portal, Smithsonian Digital Volunteers, and the Lichen, Bryophytes and Climate Change Volunteer Portal using Symbiota software. The aOCR WG is working collaboratively with the Joint Research Activity (JRA) Synthesys3 Project to share expertise for automated data collection for digital images. Got text in your images? How might OCR output work for you? Come talk to our aOCR WG to find out and to share your expertise.

**P12**

**14:20-14:40 pm**

**Incorporating OCR into a digitisation and curation workflow**

**Elspeth Haston**, Robyn Drinkwater & Robert Cubey (Royal Botanic Garden Edinburgh)

**Abstract:**

The digitisation of natural history collections is a priority for many institutes for reasons including opening access to users around the world, incorporation of specimen data in research, disaster planning, etc. However, digitisation of natural history specimens is expensive and labour intensive. For this reason, digitisation of specimens has moved towards minimal data capture and imaging as part of large scale processes. The workflow can be summarised in the following steps: 1) minimal curation; 2) attach barcode as a unique identifier; 3) minimal data entry; 4) image specimen; 5) additional data entry. There has been some investigation by institutes into the use of Optical Character Recognition (OCR) within the digitsation and curation workflow. The Royal Botanic Garden Edinburgh (RBGE) now routinely processes all specimen images through OCR software. The OCR process is integrated into the overall digitisation and curation workflow and has been used to speed up the process of adding data to over 100,000 specimens. The following additional steps have been incorporated into the workflow at RBGE: 4a) assess condition of specimen; 4b) process image through OCR software; 5a) additional curation. The incorporation of OCR into digitisation workflows is being explored by the Synthesys project funded by the European Union within Framework 7, and by iDigBio funded by the United States Government within the National Science Foundation programme. The work being carried out at the Royal Botanic Garden Edinburgh to integrate OCR into the digitisation and curation workflow is discussed as part of the work of Synthesys and iDigBio.

**P13**

**14:40-15:00 pm**

**The virtually browseable collection: connecting GIS to whole drawer imaging**

**Ann Molineux**1, Robert W. Burroughs2, & Faye Geigerman1 (Non-vertebrate Paleontology and Jackson School of Geosciences at The University of Texas at Austin)

**Abstract:**

Accessibility is one of the most problematic aspects of offsite storage for collections. This project approaches that issue by creating virtual access to a large collection stored 8 miles from the main institution. To create virtual access, we used two elements that already exist for our collection. A GIS map of all cabinets in the repository and a Specify 6 database of records, which relates the specimens in those cabinets to their spatial locations. We combined database and spatial data using ArcGIS Online, because it has cross-platform compatibility. We related a third aspect, drawer and specimen images that are linked to the specimen records and can be viewed alongside the data and spatial records.

Our final result is an online resource that allows the user to virtually examine the collection, query a database file and view the drawer content, individual specimens, and label images when they exist. This product is useful for researchers, teachers and the general public. The latter group is important from an educational stand point but also because many of the collections were amassed using public funds. It fulfills an additional role in that it is a time stamp marker for each drawer. What was located in that drawer, what was its condition? Over time additional images of drawers will be taken providing an archival record of their content and condition.

**P14**

**15:00-15:20 pm**

**Georeferencing Fish Collections from the FishNet Network: An Update of Progress and Evaluation of Collaborative Georeferencing Techniques**

**Nelson E. Rios**1, Henry L. Bart2 & Michael H. Doosey3 (Tulane University)

**Abstract:**

FishNet2 (www.fishnet2.net) is a global network of fish collection databases that gives researchers access to data on roughly 4 million fish species lots, representing over 30 million specimens. Before this project began, only about two-thirds of the records in Fishnet2 were georeferenced. The goals of this collaborative project were to georeference all localities without geographic coordinates and repatriate the results to data providers. The Collaborative Georeferencing Client (CoGe)of the GEOLocate Platform was used to georeference and verify the roughly 250,000 localities in FishNet2 lacking coordinates at the start of the project. Each of the twelve collaborating institutions hired a full-time georeferencing technician to verify and correct the CoGe generated geographic coordinates for localities assigned to his or her institution. Since January 2013, when the georeferencing phase of the project began, more than 244,000 localities of have been processed (214,000 corrected), and over 1 million specimens records have been processed (935,000 corrected). Records that could not be georeferenced were skipped and reasons for not correcting the coordinates were recorded. Every corrected record includes geographic coordinates in decimal degrees, an uncertainty radius, and most of the records have a user generated error polygon to define the uncertainty in the determinations.

**15:20-16:00 pm Poster Session and Break**

**P15**

**16:00-16:20 pm**

**iDigBio's Biospex System for Engaging the Public in Biodiversity Research Specimen Digitization**

**Elizabeth R. Ellwood**1, Austin R. Mast1, Robert Bruhn2, Jeremy Spinks2, Greg Riccardi2 (Florida State University)

**Abstract:**

New web resources provide scientists opportunities to engage the public in ways and at scales not previously possible. Many ecological and environmental citizen science projects focus on generating present-day occurrence data on populations, species, and communities to address urgent societal challenges, such as the extinction crisis and biotic responses to climate change. Biodiversity research collections provide the opportunity to produce the important historical and present-day baseline data on distributions with which to compare the new observations and project future change. However, the majority of information about the specimens in these collections has yet to be digitized. Public engagement might provide an important strategy to accelerate digitization. Out of a 2012 workshop emerged the idea of an iDigBio public participation in digitization management system, which would permit the creation of record sets of incomplete specimen data and/or images from the iDigBio Cloud, management of their digitization using collaborating citizen science tools, monitoring of digitization progress, advertisement of the projects on the go-to sites for members of the public interested in citizen science, and return of the new data to the data providers and those involved in the digitization. We will introduce this emerging system, called Biospex for Biodiversity Specimen Expeditions. We will provide an overview of the management system and its interoperability with biodiversity specimen data management systems and citizen science tools via Darwin Core Archive and Ecological Markup Language files.

**P16**

**16:20-16:40 pm**

**Integrating High Throughput Digitization with Distributed Software: Supporting Data Flows in the New England Vascular Plant Network with FilteredPush Technologies**

**Paul J. Morris**1,2, James Hanken1,Maureen Kelly1, David B. Lowery1,5, Bertram Ludäscher4, James A. Macklin3, Chuck McCallum1, Robert A. Morris2,5, Tianhong Song4, Patrick Sweeney6

(Museum of Comparative Zoology, Harvard University Herbaria, Agriculture and Agri-Food Canada, UC Davis Genome Center, Computer Science Department at University of Massachusetts, Yale Peabody Museum of Natural History)

**Abstract:**

FilteredPush is supporting digitization in two Thematic Collections Networks (TCNs), the Southwest Arthropods Collections Network (SCAN) and the New England Vascular Plant project (NEVP). In NEVP, minimal data records, including current taxonomic identification, state and town of collection, and date collected are created during imaging at high-throughput digitization stations at digitization sites. Current identification is obtained at the folder level in a pre-capture step and associated with the specimens upon imaging. FilteredPush transports data from the digitization sites by wrapping DarwinCore terms in Open Annotation ontology documents, including metadata about the when, where, and who of digitization, typing as New Occurrence records, and minimal AudubonCore for the images. These annotations are ingested into the NEVP Symbiota portal. The annotations are then available for ingest into the databases of record through new occurrence ingest tools in the Specify 6 collections management system. Additional data will be transcribed, in Symbiota, from the specimen images, and, supporting the science goals of the project, flowering and fruiting state will be coded for some taxa. These assertions will be wrapped in annotations, typed to reflect domain business operations, and transported to the relevant collections for ingest. Records harvested from network participants into datastores within FilteredPush will be subject to quality control from an Akka workflow that tests the taxon name, georeference, and date collected values of each record. Quality control issues (including proposed corrections) are reported in response to queries by researchers and transported through annotations to both Symbiota and the databases of record.

**P17**

**16:40-17:00 pm**

**Using Complementarity to Improve Plant Specimen Digitization**

**Rusty Russell**, Elspeth Haston & Nicola Nicolson (Smithsonian Institution, Royal Botanic Garden Edinburgh, Royal Botanic Gardens, Kew)

**Abstract:**

Unlike most zoological specimens, botanical collections distinguish themselves by having duplicates, that is, multiple specimens generated from the same collecting event. These are either from the same plant or from a close population and are distributed among multiple herbaria. Over the last two decades, as digitization has become a significant part of doing business in systematics collections, it’s become clear that a great deal of “duplication” of effort has followed “duplication” of plant collections. Aggregators of specimen data, such as GBIF and iDigBio, recognize this phenomenon, as do individual herbaria. In 2005, a GBIF Workshop in Crete, led by the late Larry Speers, addressed this topic and Rusty Russell was tasked with investigating the issues surrounding “duplication”. More recently, a joint effort is underway at RBG-Kew, RBG-Edinburgh and the Smithsonian Institution to develop tools to harness the potential that plant duplicates offer toward making digitization a more efficient enterprise. Past and present activities to address this issue will be described.

**P18**

**17:00-17:20 pm**

**Seaweed Collections Online: Mobilising data from national and regional museums**

**Jo Wilbraham**1& Juliet Brodie(Natural History Museum, London, SW7 5BD, UK)

**Abstact:**

The seashores and shallow seas around Britain support an important component of UK biodiversity with over 650 species of red, green and brown seaweeds which represent c. 7% of the described seaweed flora of the world. However, over 55% are Data Deficient according to IUCN criteria, there is increasing evidence that large brown habitat-forming seaweeds (kelps and fucoids) are disappearing and invasive seaweed species are increasing. Consequently there is an urgent and increasing need for good quality, verifiable data on past and present species occurrence to inform research into these issues. Museum collections provide crucial evidence points for mapping changing patterns in species distribution around the

UK. For this project we aimed to capture seaweed collections data from UK national and regional museums, focussing on our target list of seaweed species which we prioritised by their relevance to environmental change and conservation research. The project also enabled us to develop a network of persons responsible for seaweed collections at the participating institutions. These collections data are disseminated online via the project website. We used a Scratchpad for this purpose (http://scratchpads.eu) which provided tools for delivering specimen data and associated information in a customisable website: http://seaweeds.myspecies.info/. Fifteen institutions participated in the project over the course of one year and over 8000 records were received, 4334 of which were newly generated as part of the project. This project was funded by the Esmée Fairbairn Collections Fund.

**P19**

**17:20-17:40 pm**

**From Museum Specimen Database to Ecological Statement**

**Christine A. Johnson1**, Richard K. Rabeler2, & Charles Bartlett3

(American Museum of Natural History, University of Michigan Herbarium, University of Delaware, Department of Entomology and Wildlife Ecology)

**Abstract:**

The goal of the Thematic Collections Networks, Tri-trophic Digitization Project (TCN-TTD) is to database and image, in part, plant-associated Hemiptera (“true bugs”), parasitoid insects associated with Hemiptera, and the plants on which Hemiptera feed, totaling approximately 2.5 million specimens from 18 insect and 14 plant institutional collections. Integrating these data with other extant datasets, particularly from diverse sources (databases, institutions) with diverse historical workflows and standards such as these, provide a unique opportunity to generate and test various biogeographical and ecological hypotheses. Here we present our progress to date on the digitization efforts and demonstrate the usefulness and some of the inherent challenges of these data with a relatively small dataset of treehoppers and their associated oak species derived from our efforts. We also demonstrate some simple tools to examine database data for quality problems.

**Wrap Up! See you at The Banquet tonight where we can continue our conversations started here.**

**We look forward to seeing you on Friday at our SIG meeting to discuss how we can collaborate across our unique digitization projects.** Thank you everyone for your participation.

**Important bits.**

Please see this iDigBio Wiki (*a work in progress)*.

* https://www.idigbio.org/wiki/index.php/Progress\_in\_Digitization
  + A downstream resource, before, during and after this symposium.
  + You may log in and edit your entry there – as needed.
* Presentations are being recorded and those recordings found on the wiki.
* All powerpoints (or PDFs) of presentations will be on the wiki too.