

Minutes of Internal Advisory Committee Meeting on May 2, 2018

Date/Time: Wednesday, May 2, 2018 - 2:00 PM Eastern

TCN Attendees:

SoRo	Dina Clark	InvertEBase	Petra Sierwald
PALEONICHES	Bruce Lieberman	GLI	Ken Cameron
EPICC	Erica Clites, Charles Marshall	SERNEC	Michael Denslow
LepNet, SCAN	Neil Cobb	MiCC	Andy Miller
Cretaceous World	Bruce Lieberman	MAM	Anne Barber
oVert	David Blackburn, Oona Takano	FIC	Talia Karim
MHC	Chris Neefus		

iDigBio Attendees:

Administration	David Jennings Cathy Bester	Digitization & Training	Deb Paul
Education & Outreach	Molly Phillips	Cyberinfrastructure	Kevin Love
Research Use		Assessment & Evaluation	Shari Ellis

NSF Attendee: Reed Beaman

Facilitator: Bruce Lieberman

Meeting Location: Building 105 Room 310 and <http://idigbio.adobeconnect.com/iac/>

Recording: <http://idigbio.adobeconnect.com/pk472fytgilo/>

Agenda Items

Georeferencing

Petra would like some practical advice on georeferencing. Specifically, she is seeking data such as how long does it take for 10 or 100 US locations to be georeferenced. Has anybody run timed experiments? Has anybody checked accuracy of georeferencing done by say volunteers or students?

In the past, iDigBio has lead georeferencing training workshops, now with new TCNs continuing to come online there may be a need for refresher and/or training opportunities. Andy's TCN is hitting georeferencing pretty hard – efficiency depends on collections and localities, batch georeferencing – used counties, with an average error rate of 4%. Bruce stated that his TCN's georeferencing activities are handled on case by case with the time consuming aspect more worthwhile, with the method of georeferencing added to records.

Petra has found spent time looking at georeferencing mistakes and associated patterns in these mistakes including data entry issues, etc. She is looking for hard data from particular collections for example: pinned insects, 3700 records from Florida, ~500 have coordinates while the rest do not – how long does it take? Issues are carefully logged, noted hours for data cleaning along with georeferencing method and source, etc. For

example they have georeferenced 9,000 localities at approx. 3.44 min per record. They are observing trends and patterns of mistakes to improve efficiency and proofing. They used records with coordinates from three states and re-georeferenced them in Geolocate, noting mistakes and efficiency. Petra noted that each collection type will be very different based on how collections grow and collections management. She is keeping a list of pitfalls for now to help others in the future to avoid these with this information to be eventually shared on the iDigBio website or as a publication.

Deb provided a link to a paper called Mapping Life (Ellwood, E R et al 2016 Mapping Life – Quality Assessment of Novice vs. Expert Georeferencers. Citizen Science: Theory and Practice, 1(1): 4, pp. 1–12) although not directly what Petra is searching for, this is a piece of the puzzle.

Deb noted that while there are no georeferencing workshops being planned by iDigBio at the moment, there are related activities in the near future including symposia at the upcoming SPNHC 2018 meeting and the Digital Data Conference in Berkeley next month. David J also mentioned that perhaps a mini georeferencing workshop/symposium be included at the iDigBio Summit to which there was a good amount of agreement. TCNs should visit the Summit wiki page to suggest topics to be included during the Summit sessions.

https://www.idigbio.org/wiki/index.php/Talk:ADBC_Summit_2018

Bruce asked the participants if there had been much opportunity to share georeferencing localities. Deb stated that the TriTropic TCN knew georeferenced localities existed in the iDigBio portal so they designed their workflow to include pulling that info which saved a lot of time (extracted from iDigBio and created a resource to pull down specimen records). Anne also said that the MAM TCN uses this method while Chris noted that there wasn't much overlap for his TCN. Ken mentioned that the fish, molluscs and plants shared georeferenced localities (same body of freshwater). Petra is interested in running a trial for comparison between institutions. According to Neil, Symbiota georeferenced locations can be shared however you cannot share across portals. This was discussed with Nelson, although Geolocate can be provided a lot of localities, these would need to be vetted. Cat Chapman wrote a script used to ping iDigBio for georeferenced locations. A better API connection with the iDigBio database that we could reference and decide if that is a good geolocation, is needed. An Adobe Connect meeting would be good for people who want to move this forward, in addition to anything that may be scheduled during the Summit. Deb suggested that the georeferencing working group get involved with this effort.

Google Analytics

Neil would like to provide an update on the Google Analytics project.

Neil reported that everyone other than VertNet, InvertNet and TriTrophic has provided a link to their usage statistical data. Otherwise the TCNs are getting him the data that he needs using Data Studio and directly through Google Analytics. These provide useful metrics on engaging the community, thanks to Neil for compiling this data.

Neil will send out the Google Analytics report to the IAC listserv (IDIGBIOIAC-I@lists.ufl.edu) which goes to all TCN PIs and the PIs and staff of iDigBio. Neil will run the analytics on a quarterly basis; should this data be included in TCN quarterly reports? Reed noted that the TCN reports to NSF provide a summary of each TCN so this data would be appropriate for each TCN separately. NSF may also be able to get this information from the

iDigBio annual report. The TCN leads can take what they want from the data generated in a timely manner by Neil to include in their NSF reports.

Regarding which Google Analytics data would be most useful for TCNs, David J noted that the TCNs have different styles of reports therefore Neil would be best to give advice. Alex's Data Studio template, using four variables from the front page, provides a good snapshot on activity of web sites. Neil can generate an annual full review using these four variables. The overall goal is to find out who is interacting with web sites and data portals, who is engaged with ADBC data, and to develop a strategy to increase this engagement. Kevin noted that data users via API are not reported in Google Analytics due to lack of a mechanism to report on these users using this service.

Use of Collection Data

Charles would like some discussion of how one of the biggest challenges is making use of their collective data to advance our understanding of ecological, evolutionary and environmental change as well the underlying systematics of their taxa.

Bruce agreed that this is important and reminded the group about the upcoming Digital Data Conference next month in Berkeley being organized by Gil Nelson. Kevin gave details about yesterday's ecological niche modeling webinar which was a one hour and included the offering of four tutorial videos/accompanying data files produced by an iDigBio grad student (Blaine Marchant) along with office hours to walk people through challenges and issues working with these materials. This may be a good model to deliver such training materials. <https://www.idigbio.org/content/using-digitized-collections-based-data-research-free-hands-crash-course-ecological-niche>

Neil brought up that problems with connecting with ecologists are perhaps are due to low quality data and their need for sophisticated data sets. There is potential to engage ecologists who work at the plot level/landscape level, needing to scale up in utilizing presence data. Species distribution modeling ecologists are already on board with using digitized data. GBIF promotes monitoring data and accepts observation data while ADBC does not. It would be more useful if iDigBio took in more types of data. iDigBio always defers to NSF regarding data scope, however it would really help if iDigBio would open up to more types of data which would be more useful for ecologists. Kevin mentioned iDigBio's involvement in a new model of conference exhibit to provide help desk capabilities to conference participants, in space shared with other data providers (ESIP, EDI, Arctic Data Center, DataONE, and iDigBio coordinating a Data Help Desk at ESA). Shari informed the group that the iDigBio Data Portal user survey asking for use of the data has resulted in systematics first, followed by ecologists. Bruce has been able to pull out abundance information (rank abundance) from the iDigBio Portal, although sometimes the scale of resolution is different than the ecologists scale. Data seemed to conform to types of abundance data that ecologists and paleoecologists would expect to find.

Deb stated that the ICCER working group has been talking with ecologists on how our data can be more useful both now and in the future. In a trial project, the Denver Botanic Garden is asking collectors and ecologists to collect data for their own use, in addition for the other group.

Summit 2018

ADBC Summit 2018 agenda planning: https://www.idigbio.org/wiki/index.php/ADBC_Summit_2018.

There is a wiki page on the iDigBio web site set up for the 2018 Summit consisting of two pages: regular view and a discussion view (see tabs in the upper left of the page) where ideas will be recorded. iDigBio is starting to plan the Summit agenda, please note any discussion group topics, workshops, demos, etc. into the wiki to be included in the planning process. https://www.idigbio.org/wiki/index.php/Talk:ADBC_Summit_2018 Deb briefly discussed Cindy Skema from MAM and her experience working with the challenges of innovating in a limited area with industry regarding light boxes. A Summit session on Innovations and Challenges may be useful and of interest.

Discussion Topics

With the retirement of Joanna McCaffrey, Kevin will be assuming the responsibilities of Biodiversity Informatics Manager. These responsibilities include coordinating with data publishers, ensuring timely and accurate reporting on data mobilization and digitization progress, and managing and curating iDigBio's Collection Catalog. (Kevin's contact klove@flmnh.ufl.edu)

Bruce asked for a TCN volunteer to lead the next IAC Meeting on Aug 1, 2018 at 2pm eastern. No one stepped up; Kevin will meet with the TCN PIs at the Digital Data Conference on June 3-6, 2018 at which time someone will be nominated to lead the next IAC Meeting.

Meeting Adjourned: 3:05 pm