

Minutes of Internal Advisory Committee Meeting on September 15, 2016

Date/Time: Thursday, September 15, 2016 2:00-3:30 PM Eastern

TCN Attendees:

InvertNet		FIC	Talia Karim
LBCC	Corinna Gries	MHC	
PALEONICHES	Bruce Lieberman	GLI	Aaron Goldberg Ken Cameron
MaCC		InvertEBase	Liz Shea Petra Sierwald
NEVP	Patrick Sweeney	SERNEC	Michael Denslow Zack Murrell
SCAN / LepNet	Neil Cobb Steven Heydon	MiCC	Andrew Miller Rhianna Baldree Alex Kuhn Teresa Iturriaga
VACS		EPICC	Erica Clites
The Cretaceous World	Bruce Lieberman	MAM	Anne Barber Cindy Skema
		HUH	Jonathan Kennedy
			Jason Best Laura Wilson

iDigBio Attendees:

Administration		Digitization & Training	Deb Paul
Education & Outreach	Molly Phillips	Cyberinfrastructure	Joanna McCaffrey Renato Figueiredo- Kevin Love
Research Use	Shelley James	Assessment & Evaluation	Shari Ellis

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

Recording: <http://idigbio.adobeconnect.com/p4sbtehy24v/>

RECURRING ITEMS:

Meeting started: 11:03 am

Upcoming workshops/meetings/events

<https://www.idigbio.org/calendar>

TCN bi-monthly progress reports to iDigBio are due

<https://www.idigbio.org/content/tcn-bi-monthly-progress-report-idigbio>

- Thanks to the TCNs that submitted their reports in advance of the meeting. David Jennings will follow-up with the others.
- TCN PIs : your bi-monthly reports are due before the meeting. Please go to <https://www.idigbio.org/content/tcn-bi-monthly-progress-report-idigbio>
- Reports from the following TCNs are included with the minutes and are archived on the IAC wiki page (https://www.idigbio.org/wiki/index.php/Internal_Advisory_Committee):

<input checked="" type="checkbox"/> InvertNet	<input checked="" type="checkbox"/> SCAN	<input checked="" type="checkbox"/> InvertEBase
<input type="checkbox"/> LBCC	<input checked="" type="checkbox"/> VACS	<input checked="" type="checkbox"/> SERNEC
<input checked="" type="checkbox"/> PALEONICHES	<input checked="" type="checkbox"/> FIC	<input checked="" type="checkbox"/> MiCC
<input type="checkbox"/> MaCC	<input type="checkbox"/> MHC	<input checked="" type="checkbox"/> EPICC
<input checked="" type="checkbox"/> NEVP	<input checked="" type="checkbox"/> GLI	<input checked="" type="checkbox"/> The Cretaceous World

Symbiota Data

- Joanna reminded all to publish and republish your data with Symbiota.
 - iDigBio will not have access to your data if you do not republish your Darwin CORE archives
 - All records must have an occurrence ID
 - Link provided for how to publish: <http://symbiota.org/docs/darwin-core-archive-data-publishing/>

DISCUSSION TOPICS:

File schemes – Rich Rabeler

What file schemes (naming conventions, etc.) are projects using for storing their RAW/DNG images? Are they trying to file taxonomically or merely by barcode?

- Taxonomy not a good idea...can change
- Listing by barcode number seems to be best
 - Cindy Skema (MAM) – lists by barcode
 - Patrick Sweeney (NEVP) – images named by barcode and sorted by date uploaded and by collections
 - Jonathan Kennedy (HUH) - Just barcode. Barcode is searchable in database, stored in AWS with minimal metadata for backup discovery capability
 - Bruce Lieberman - Paleoniches and Cretaceous World at KU save each file with its name being specimen # then a dash, then the taxon name.
 - Alex Kuhn - (MiCC) place images in folders by barcode, but split into folders based on boxes
- Deb Paul provided reference material: <http://www.exadox.com/en/articles/file-naming-convention-ten-rules-best-practice>

Anyone not storing RAW? – Ken Cameron

- Anne Barber (MAM) - Storing RAWs (CR2, etc.) isn't necessary. You want an open source file format like DNG or TIFF for archival images

- DNG stores metadata better than any other file type (or RAW)

Where are folks storing their data? - Ken Cameron

- Jonathan Kennedy (HUH) recommends storing images in 3 places to ensure preservation
 - starting to store on Amazon service and archive Glacier Service
 - S3 will allow people to search easier
 - Glacier and Harvard <https://aws.amazon.com/glacier>
- Jason Best: We are currently using local storage and Dropbox as cloud backup (we have unlimited storage thru our biz account). Looking into Glacier for the long term.
- Talia Karim (Fossil Insect TCN): We are using research computing PETA Library storage at CU-Boulder
- Patrick Sweeney (NEVP): Yale's archival images are stored in Yale's DAMS
<http://ydc2.yale.edu/category/digital-ecosystem/technology/digital-asset-management>

One large portal or smaller portals?

- Cindy Skema – MAM's sustainability included integration in a larger portal
- Neil Cobb – SCAN is stand alone arthropod portal that can also access data sent directly to iDigBio
- Patrick Sweeney (NEVP): Symbiota can import from a DwC-A (from an IPT, for example) and custom procedures can be written to import only portal relevant data.
- Andrew Miller – Suggested referencing:
https://www.idigbio.org/wiki/images/1/17/IDigBio_SWG_MeetingSummary-04AUG2014.pdf

Asking for input – Joanna

When iDigBio receives your data we save it and additionally create index layer to better search the data. There are a lot of inconsistencies within data. The data quality flags are great to fix problems especially for locality errors but taxonomy issues can be much more challenging. iDigBio uses the GBIF Backbone (<http://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c>) for their taxonomy referencing. iDigBio is working to resolve these issues in the portal and in GBIF's Backbone. As things improve, is there anything you would be opposed to iDigBio changing?

- Scientific name - The scientific names will only be changed if it does not match the GBIF backbone
- Jason Best: If changes are made, is previous data preserved in record versions or annotation records?
 - Shelley James: the raw data is still present
- Petra Sierwald: The taxonomic names on GBIF for one of my groups, Diplopoda, are just not good enough.
- Talia Karim: her data is definitely being organized incorrectly
- Deb Paul: About the GBIF Backbone: <http://gbif.blogspot.com/2016/04/updating-gbif-backbone.html>
- Shelley James: One way to have taxonomic data sets incorporated into the GBIF backbone is to publish it with GBIF
 - Shelley James: <http://www.gbif.org/publishing-data/summary>
 - Deb Paul: Researchers need to work with the data providers at GBIF to ensure that lists and updates happen

- Given all the taxonomic name / ranks are correct, the only thing that should be changed on the index file is the name to provide a synonym

Abstracts due

- Deb Paul: SPNHC 2017! <https://www.idigbio.org/content/spnhc-2017-call-abstracts>
- Shelley James: Big data symposium at International Botanical conference.

Research Spotlight

Submit research spotlight highlights and data through the research Spotlight page

<https://www.idigbio.org/tags/research-spotlight>

2018 TDWG and SPNCH happening together in Dunedin, New Zealand

4. Summit? When, where?

Meeting Adjourned: 12:15 pm