

Minutes of Internal Advisory Committee Meeting held on July 7, 2015

Date/Time: Tuesday, July 7, 2015 10:00 AM - 12:00 PM Eastern

TCN Attendees: Ken Cameron (GLI), Neil Cobb (SCAN), Michael Denslow (SERNEC), Chris Dietrich (InvertNet), Corinna Gries (LBCC), Bruce Lieberman (PALEONICHES), Chris Neefus (MHC), Petra Sierwald (InvertEBase), Dena Smith (FIC), Mari Roberts (MaCC)

iDigBio Attendees: David Jennings (Project Manager), Larry Page (Project Director), Shari Ellis (Project Evaluator), Joanna McCaffrey (Biodiversity Informatics Manager), Andréa Matsunaga (Research Scientist), Molly Phillips (E&O Coordinator), Deb Paul (Digitization Expert), Adania Flemming (Office Assistant)

Meeting Location: Building 105, Room 310

Remote Access: <http://idigbio.adobeconnect.com/iac/>

RECURRING ITEMS:

Upcoming workshops/meetings/events

- David reminded folks to look at the calendar as there is always lot going on <https://www.idigbio.org/calendar>

TCN bi-monthly progress reports to iDigBio are due

- David still needs reports from LBCC, SCAN, VACS, and SERNEC. <https://www.idigbio.org/content/tcn-bi-monthly-progress-report-idigbio>

Please maintain your TCN wiki pages

- Joanna asked everyone to send her any changes to their TCN participants. She can then update maps, counts, etc. <https://www.idigbio.org/wiki/index.php/TCNs>

DISCUSSION TOPICS:

Did you calculate cost per specimen in your annual report? If so, how did you do it?

- Chris Neefus (plants), utilized direct labor cost only, and came out less expensive than they estimated. They utilized students mainly, and found that some students became very efficient, whereas others were not and those were given different tasks.
- Bruce Lieberman (fossils), similarly utilized direct labor cost, without any equipment charges, and also came up with slightly less cost than anticipated (allotted to efficient staff cranking through the work).
- Petra Sierwald (pinned insects), calculated time per record, but found it difficult to come up with a realistic estimate. She noted there was need to compartmentalize types of specimens better (i.e., by size), and record the many benefits to pre-curation. She also noted they have summer interns who are using voice recognition to digitize source material (field notes and locality information).

- Bruce Lieberman sympathized with Petra’s struggle for a realistic count; he noted the number of specimens varied within lots, which gave very broad variance (multiple specimens makes it easier).
- Many agreed to Petra’s idea of recording the side benefits of digitization for future digitizers. Deb added that it might also help others change their collection practices. Joanna responded that she would dig up what benefits of digitization were already available, or start something, a page on iDigBio website etc. where such recordings can live. Shari has these data, and it was decided after the meeting that she would be publishing them.
- Neil Cobb’s calculation was also lower in cost than expected.
- Chris Dietrich noted they are not at a point where they can figure a cost per specimen, but they plan to divide the specimens digitized by the amount of money given.
- Corinna Gries has not yet calculated cost per specimen, but predicts it will be more expensive based on dollars from NSF divided by number in portal; they were too optimistic about crowdsourcing (online and in person) - tradeoff: pay students to do transcription OR pay someone to do volunteer coordination. She noted volunteers need a lot of “cheerleading” (i.e., handholding, personal interactions, and time).
- Mari Roberts found it difficult to keep students around for more than a summer, which is not good for traction with crowdsourcing. She agreed that volunteers need a lot of handholding and online volunteers are not working.
- Petra Sierwald suggested if anyone has had success using volunteers, those folks can give some helpful insight.
- Shari mentioned that iDigBio is working on a worldwide transcription event, WeDigBio (www.wedigbio.org), and has information on how to host a Transcription Blitz.
 - Corinna Gries responded, that the decision has to be made to either have people cheerleading volunteers or have them transcribe.
- Mike Denslow noted that there is a grey area between on-site and online volunteers, as those folks have different desires (social interaction vs. task oriented). He added with regard to Notes from Nature, the longer a volunteer takes to do a task the more bored they will become. Therefore, instead of transcribing every single detail on the label, he found it helpful to transcribe what may be most useful, which is less time consuming.
 - Corinna Gries agreed with Mike, but reasserted that in order to reduce cost per specimen more crowdsourcing is needed.
- Chris Dietrich suggested the key is prioritization (some specimens are more important than others) and felt the research community could help decide.
 - Corinna Gries agreed with Chris, but added researchers will only transcribe 100 – 200 out of the millions that need to be transcribed.
- Mike Denslow noted that Notes From Nature's output may be limited but is better than nothing. There are three tiers of information on a label: You need to decide at what level you should ask.
 - Andrea added that there is also a need to get a consensus output from Notes From Nature transcriptions. She worked with 150k transcriptions for 40k labels in finding ways to improve

the use of workforce. If someone is interested: <https://www.idigbio.org/content/reaching-consensus-crowdsourced-transcription-biocollections-information>

- Joanna asked if folks thought further discussion about transcription might be useful (possibly have WeDigBio talk to us).
 - Most folks agreed; Deb also thought we could invite Paul Flemmons and other platforms potentially.
 - Mike Denslow agreed, but asked what the possibilities were in term of changing their TCN workflow.
 - Petra thought it would be helpful and noted that they are far from digitizing everything in their collection and would be happy to use a more helpful workflow with other potential funding (having established that we could show we did this much, learnt this much, and now moving forward we will follow these steps, for more efficiency etc.).
 - Joanna suggested that WeDigBio would be a good topic for the Summit.
 - David Jennings responded that WeDigBio has already suggested a discussion topic at the summit that can be summarized as: “developing an integrated pipeline for making citizen science/crowd sourcing efforts readily adaptable for the various TCNs.”
- Corinna commented that we need to push for automation and put more effort into OCR.
 - Deb Paul noted we need to remember what is the most important info: where, what, and who. She added if we don’t have a complete record, how do we ensure it gets to completeness. She believes it is a compelling topic; a lot has happened with the NLP format and processing of OCR output since this project has started. The Europeans have taken a lot of what has been done in the U.S. and are running with it on their end.
 - Corinna responded that we need to consider what will make the data so interesting that researchers will want to add more information, which will compel them to go back to the label and add the information.

Brief summaries of the community survey and impact evaluation results

- Shari thanked everyone for participating in the survey, and thanked Neil Cobb for his heavy promotion of the survey. iDigBio ended up with 250 completed surveys. She divided the results into two categories Impacts and Survey Outcomes:
 - Survey
 - The portal comments have been translated into Redmine tickets for the ACIS team.
 - iDigBio website is good resource, but is not well organized and is hard to navigate.
 - Many do not use social media.
 - **Priorities for upcoming year:** training, software development, and bringing more data online
 - Impacts
 - **Unanticipated consequences of digitization:** improved collections management, improved quality, and improved training.
 - **Anticipated consequences of digitization:** increased sharing, increased collaboration, and increased digitization.

- Did not get a lot of detail on the above broad categories
- **New foci for iDigBio:** research use of data and more international collaboration (which is already in the works)
- The reports are available on the website if anyone would like to read them:
<https://www.idigbio.org/content/results-2015-idigbio-community-survey-are>

Report on the recent iDigBio API Hackathon

- The hackathon went very well from iDigBio's perspective.
 - Many repositories of code were created - media ingestion; research tools; augmenting data; correcting georeferences; mapping; LifeMapper plugin for QGIS, all information is on the GitHub wiki
- iDigBio launched new version of portal and API last week.
- Andrea is working on creating videos from the demo sessions and will give a demonstration of LifeMapper at the next IAC meeting.
- Reports were created from each of the four teams from the hackathon on the Wiki page:
<https://github.com/idigbio-api-hackathon/HackathonCentral/wiki>

Beta portal evaluations

- <https://beta-portal.idigbio.org>
 - The new portal and API were launched last week.
- Joanna had suggested that everyone re-do the research question using the new portal and report changes/improvements in user experience.
 - No one has done anything yet.

iDigBio Summit V

- David noted that we are trying to draft an outline each day of the Summit (see attached outline).
 - The summit will be on Thursday (November 5) and Friday (November 6).
 - Wednesday (November 4) is reserved for meetings and gatherings for iDigBio and TCNs, or is a travelling day for everyone.
 - TCNs can have meeting space on Wednesday. Please contact David to reserve a room.
 - Send discussion topic ideas to David (to help design breakout sessions).
- David has a list of proposed discussion groups, which are included on the attached Summit outline.
- Chris Neefus noted he would like time scheduled for networking, preferably during the day, to meet with people face to face not just at breaks, meals, and receptions. He will also would like a discussion group about research tool development.
 - Deb noted there should be a GBIF data quality task group report out by the Summit that can help frame conversations.
- Dena commented that the sustainability conversation needs to continue at this Summit, especially as projects are ending. She noted sustainability is a major topic and asked folks to share and communicate that information.
- Deb noted there might be a possible GBIF Task Force (<http://www.gbif.org/page/82104>) meeting in conjunction with the Summit.
- Larry asked if Corinna had a no cost extension.

- Corinna responded that she was in her first year of no cost extension, and planned to apply for another.
- Chris Dietrich was in the same category.
- Mari Roberts said that TTD has a no cost extension until January 2016.
- Larry commented that iDigBio has submitted its proposal, and the panel is meeting this week.

Workshop suggestions

- Summaries of past workshops are here: https://www.idigbio.org/wiki/index.php/IDigBio_Workshops
- Deb's main concerns are: What workshops do we want to repeat? What new ones? Any variations? What workshops do people want in Year 5 and Years 6-10?
 - Chris Neefus asked if it will be possible to participate remotely in the Niche Modeling workshop at Botany 2015.
 - Molly confirmed the meeting will be online, and will confirm if remote participation is possible.
 - Deb noted there is a Georeferencing Train-the-Trainer #3 (TTT3) being planned, and they are currently looking for a West Coast location.
 - Mike Denslow noted they put in a workshop proposal into iDigBio about concealing sensitive data (proposed by Gil to SC, but tabled for now). Dorothy Allard presented a paper about this at the recent SPNHC meeting. David will bring up again at next SC meeting.
 - Chris Neefus is working on a proposed solution for herbarium specimens.
 - Mike Denslow added that the SERNEC will be doing some Specify and GEOlocate collaborative georeferencing workshops within the next 12 months, where other TCNs and/or iDigBio members can participate.
 - Deb proposed having a five-day workshop that focuses on data skills/data carpentry and data use of collection data. The first two days would be introduction to the tools, and the last three will be for working with your own data.
 - Shari noted that Gil is working on series of webinars and one workshop on vertebrate digitization.
 - GEOlocate and Symbiota were suggested.
 - Deb noted she worked with NYGB (Mike Yost): they are blogging on our site and describing what they are doing and learning, which has downstream effects of changing data functionality for people using Symbiota. See <https://www.idigbio.org/content/steps-georeferencing-specimen-locality-data-community-examples>

Solicitation for next round of TCNs is available

- David reminded folks that solicitation for next round of TCNs is available: <http://www.nsf.gov/pubs/2015/nsf15576/nsf15576.htm>

Other announcements

- Dena Smith mentioned that their proposal was funded to link major data sources between paleo and bio (iDigPaleo/EarthCUBE/iDigBio) which would improve interoperability between the data sources and create the iDigPaleo Portal (ePANDAA).
- Joanna noted, referring to a discussion about seeking international data providers, that iDigBio is now ingesting most of the Paris Natural History Museum and we are working with NHM London and the

Atlas of Living Australia (ALA). She asked everyone to let iDigBio know of other collections that we should reach out to.

- Joanna added that we are working with the Smithsonian NMNH as well.
- Larry added we are talking to collections from China and Ecuador as well.
- Deb noted, something interesting learned from the SPNHC TCN Coffee Klatch: if a TCN still exists in "some form", then collections/institutions can apply to join the TCN as a PEN even if the funding for the TCN is complete.
- Molly was introduced as the new Education & Outreach Coordinator for iDigBio.