iDigBio Summit IV : Future Uses of Data in Research Discussion Group

Monday, October 27, 2014, 1:30pm - 2:30pm

Organizers: Pam Soltis

Do you have ideas of novel ways to use specimen data and images in research? Are there tools that you would like to see integrated into iDigBio? Or are you curious about how to use the data to address research questions? Please come and share your ideas as we develop new research applications for digitized data.

Attendants:

Pam Soltis (iDigBio) - psoltis@flmnh.ufl.edu
Shari Ellis (iDigBio) - sellis@ufl.edu
Katja Seltmann (TTD TCN) - kseltmann@amnh.org
Claudia Segovia (iDigBio) claudia@ufl.edu
Blaine Marchand (iDigBio) - dbmarchant@ufl.edu
Petra Sierwald - psierwald@fieldmuseum.org
Keping Ma (Chinese Academy of Sciences) - kpma@ibcas.ac.cn
John LaSalle (ALA) - john.lasalle@csiro.au
Joey Shaw - joey-shaw@ufc.edu
Elizabeth Martin - elizabeth_martin@usgs.gov
Paul Johnson - paul.johnson@sdstate.edu
Chris Dick - cwdick@umich.edu
Christy Bills - cbills@nhmu.utah.edu
Matt Collins (iDigBio) - mcollins@acis.ufl.edu
Lauren Gonzalez (iDigBio) - laurenagonzalez@ufl.edu
Andrea Matsunaga (iDigBio) - ammatsun@ufl.edu
Mary Klein (NatureServe) - mary_klein@natureserve.org
Ann Molineux- annm@austin.utexas.edu
Charlotte Germain-Aubrey (notes) - cgermain@ufl.edu
Rafe Brown - rafe@ku.edu
Chris Neefus - chris.neefus@unh.edu
Molly Phillips (iDigBio) - M.phillips@flmnh.ufl.edu
Jenny Kluse (LSU herbarium) - jskluse@gmail.com
Herrick Brown (SERNEC TCN) - brownh@biol.sc.edu (via AdobeConnect)
Donald Hobern - dhobern@gbif.org

Discussion Points

Some TCNs are wrapping up the data collection and starting to use the data for research.
Collective needs ? Where should investments be made to contribute to research ?
We're not funded to do research but research will bring sustainability to the digitization effort.
NSF and other funding agencies will be happy to see iDigBio in proposals. in fact, one of the 3 categories for NSF postdoc call.

Visualization
Looking for ways in which organisms or groups of organisms are changing geographically AND temporally.
Any visualization tools will help.

Mechanism for deep time for paleontology. able to replot your localities at the geological time the specimens were alive. G-plate (Australian algorithm), Texas, and others try to work something for Specify to be able to do that. Aggregated data environment to pull lots of data in and plot it in geological time.

Data cleaning process
Raw data downloaded from several aggregators and then need to be cleaned in several steps.
Customizing the raw data is a lengthy process.
Sharing experience ? In workshops ?
Problem of getting the clean data back to the original dataset. Email.
Ways of cleaning the data might not be interesting to other users.
iDigBio ran into this problem to standardize the datasets, but not all collections want to have modified data. Example of VertNet - do a lot of cleaning before putting the data out there.
iDigBio was told by NSF - put the data out there as soon as you get it, even if there are mistakes or is not standardized. Some categories are easily standardized (country code), but for others, need to present the original data and the 'clean standardized' one (like taxonomic name).
Also, problem of storage space to integrate metadata or research results to the specimen.

Use of collections data for identification in the field. Automating the process of sorting through specimens, either new or from poorly curated collections, to aggregate similarly-looking specimens to facilitate the annotation. Putting images of drawers online, or digitizing audio material (for frogs) to download and take in the field will help for identification.
Making the databases more user friendly for using in the field.
Morphotypes is a problem - could we implement some algorithm to be able to take plasticity into consideration ?
iDigBio data carpentry workshop will help in using tools to input their data directly at the time and place of collection.

Visualisation tools for identification. Visipedia in Cornell for identification of birds. identification of trees of North America. in Australia, not one tool fits all. UK tools for moths and butterflies. Citizen science BugGuide for entomology, very efficient. Some automated way to sort through images to get the few hundreds most related to one image to help for identification.
Google-style. NSF call for biological informatics is a good place to get money for that type of research.

Integration with existing tools

iPlant collaborative environment (access to storage space and cluster for analyses, but also access to a set of data to an entire working group)- Can we format the data from iDigBio to input in the iPlant environment ? Maybe have a user-friendly to enquire the iDigBio dataset through an API search.

Integration of tools used by the community into the iDigBio portal is a great step to help the research community. Any other tools that are useful would be great - suggestions ? tools for field identification for example.

Integration with other databases / data types

Atlas of Australia started very centralized about biological collections - Atlas has shifted its attention to a different kind of data (observation data), to make it more useful to other communities such as conservation, policy, etc…. NSF has set limits to what data iDigBio can host and aggregate. Is there a way to track the origin of the data ? Useful for funding agencies and for quality check, etc…

GBIF have to have their data attached to a specimen (DOI), which is provided by the source, so metadata document comes with each download. Encourage citation of this metadata DOI, then we can track the use of the original data source. If we can ensure uniqueness and consistency, then we will have a way to track down data usage. This is the same plan that iDigBio has been discussing. Having journals requesting the DOI would work well.
Atlas of Australia feeds back the collections with the number and categories that their data was being used for.

Future Directions

Research Working Group using iDigBio data and others to address questions that could not be answered before. Virtual meeting ? Face to face meetings are great, maybe envisioning a series of small meeting that are focused on a question/methods.