# “BIRDS OF A FEATHER” DISCUSSION GROUPS

## Topic: Research

* Research has too often been relegated to be something that will happen once our digitization workflows have run to completion. A "build it and they will come" mentality; not ideal.
* Instead, potential research uses should be considered right up front, to be sure we capture the right data and make them available in the most useful ways.
  + "Inreach" = outreach to the scientific community. Don't necessarily provide data in the way our user community currently thinks might be useful, but in ways they should fine useful.
  + Educate our scientific users to use specimen data optimally.
  + Enable and encourage new and more rigorous research approaches
  + NSF expects exemplary research results to appear right away
* Two examples where database design elements could be very helpful in enabling and guiding good research were discussed:
  + **Taxonomy/systematics** -- There are two parts to this problem, one tough but being solved by several groups (taxonomic *name* reconciliation), the other *very* tough and currently remaining unsolved by anyone (taxonomic *concept* reconcilitation).
    - There is a major need for a complete name registry so that all digitizers are using the same authority file. Some clades (e.g., plants) have good existing nomenclatorial resources (e.g., Tropicos), other clades (e.g., invertebrates) have no or poor resources. iDigBio could help the latter.
    - TNRS iPlant name service -- solves name reconciliation issue at least. Merges multiple lists. Allow people to use different taxonomy tree. iPlant or iDigBio are not arbiter of correct name; they should just present different taxonomies and let people choose. Actually providing an accepted classification for a group is a better role for a regional (e.g., Consortium of California Herbaria) or taxon-based (e.g., Solanum PBI) group.
    - It is very challenging to put "the best" taxon name on a particular specimen (unless it is a type specimen; then it is easy). Not clear what "the best" means anyway; there never is or will be a final, perfect taxonomy -- it is always subject to revision.
    - Therefore we need to include "according to the concept of \_\_\_\_\_" in the database along with a name. Two ways to specify taxon concept: characters, or a new approach using annotation history (especially specimens annotated by monographers).
    - Therefore it is very important to keep annotation history in DB.
    - Problems with complex synonymy are more or less problematic in different groups depending on biology and history of study.
    - If we can incorporate phylogenies with names it will really help. E.g., generic lumps and splits are less problematic if we can specify how names are related to each other phylogenetically. An advantage to phylogenetic nomenclature, e.g., RegNum.
    - We need to provide a rating system for specimen IDs -- e.g., gold standard is ID by monographer rated most highly, silver standard is ID by other recognized expert in the geographic area and/or taxon -- it goes down several levels from there.... This would let the user decide on their preferred level of confidence needed for the type of study they are doing.
    - The sociological downside is that labeling identification quality by publicly ranking people's expertise could lead to bad feelings depending on ones rating!
  + **Climate change modeling** -- how to ameliorate biases in specimen data? How can we allow people to "cherry-pick" the specimens they need for this particular purpose of detecting real changes in distribution.
    - One important criterion is accurate ID (see section above).
    - Another important criterion is that many researchers want to know whether they can count on completeness of an inventory of an area at a particular time, in particular whether they can infer absence of a taxon from a place. This is very tricky from museum specimens -- only possible when a good collector has done a "complete," vouchered survey of a plot or other geographic area.
    - We could have a variable in the database that would mark certain specimens as part of a "complete" survey in the past (e.g., an inventory of a plot or area) -- it could link to literature references or notes that would give the details. We could also let people search from a pick list of complete surveys, or from a literature citation.
    - This would provide a different kind of rating system for specimens; a user could select only that fraction of the specimens that result from a "complete" inventory within a geographic area.

## Topic: GUIDs

Participants: Deb Paul, Barbara Thiers, Rod Spears, Corinna Gries, David Bloom, Greg Riccardi, Andrea Matsunaga, Jose Fortes, Patrick Sweeeny, Jim Beach and scribe Joanna McCaffrey

These are opinions expressed by the attendees (and not necessarily those of iDigBio):

1. The Hub should be the source of unique names (ref: authority files, especially collector names, taxa, sites)
2. The Hub should have a resolution service for GUIDs
3. 1 GUID per representation of the physical specimen (i.e., the digital record)
4. opaque ids are good
5. use cases for GUIDS:

* write-back, annotations
* being able to follow how a specimen has been used (also a great motivator to make collectors to do some standardization of data and general collection data improvement)

1. we cannot look to or wait for the community as a whole or any of the associated standards groups (TDWG, SPNHC, etc) to make a decision in regards to GUID strategies. They have tried and failed. The last I heard DataONE was going to use EZIDs and it appears that BiSciCol is leaning that way. If we choose the same approach I think that would benefit us in the long run and maybe finally push this issue past the point of inertia. I certainly do not think we need to create something new or take a localized tactical approach.
2. the HUB should be the gatekeeper, the resolver of GUIDs, meaning everyone should register their physical specimen, Taxon, Geography and Collector GUIDs with the HUB.
3. The problem with the TCN grants is that they are designed to motivate institutions to digitize as much material as possible but they do not stress or require any data quality standards. If we do not address the quality issue now we will be paying for this for a very long time. The first step in the long road to quality data is getting unique identifiers on the important pieces of information.
4. We also need to get the 'software stragglers' (those using inadequate home grown systems or Excel, etc.) to start using contemporary software that will make the process nearly transparent for them, plus, all the other benefits Collection Management Systems will bring.

## Topic: Symbiota

Participants: Katja Seltmann, Toby Schuh, Anna Monflis, and Lucile McCook

We had a small, very informal BOF breakout group that went over the some of the data entry and management functionality built into the Symbiota software. Katja asked for this in order to determine if Symbiota was appropriate for their project. Anna was interested in using Symbiota for her mid-west project.

Across the room, Paul Morris and Patrick Sweeney met with a small group to talk about Filtered-Push, Symbiota, and other applications that will be interacting with the FP project.

## Topic: Filtered Push

Participants: Paul Morris, Matt Collins, Austin Mast, Alex Thompson

In-depth discussions of some of the broader features of the filtered push architecture and drilled down into points of implementation.

## Topic: Adobe Connect

Participants: Kevin Love, Bruce MacFadden

Kevin gave a walkthrough of the basic features of Adobe Connect. Focus was on how to get connected and participate in meetings.

* Audio setup wizard
* Microphone / Phone icon
* Speaker icon
* Webcam icon