# SUSTAINABILITY DISCUSSION GROUP

Facilitator: Neil Cobb

Location: Century C

<https://docs.google.com/document/d/1bcC8Z8yylDHij9Me_eOVR6JwA4TU1SCSeTiD4h7JWEc/edit>

[**http://tinyurl.com/summit2-sustainability**](http://tinyurl.com/summit2-sustainability)

**Sustainability**

How do we continue efforts of iDigBio and the TCNs (National Digitization Effort) beyond the specific funding provided by the ADBC program?

**Considerations**

Sustainability can be achieved best by demonstrating the success of the program in promoting research and education. For example, the we need to be able to address grand challenges (e.g., ecosystem services, biodiversity loss crisis, global warming, etc...).

Need to identify what needs to be sustained with direct continuing support? It depends on the role of the National Digitization Effort, which may change over 10 years (e.g. working more with policy makers).

Need to have a number of options for sustainability planning that range from bare bones sustainability to a full-featured program. Need to create Sustainability Working Group that can create a long-term plan (5-15 years) that has community buy-in.

Timeline: TCNs have three years, iDigBio has ten years

**Needs**

TCN Sustainability Needs:

1. Digitizing new specimens (needs funding for staff, equipment)

2. Training (need some funding, iDigBio funding)

3. Maintaining cyberinfrastructure (covered by lead institutions)

4. Adding new collections to the portal, especially legacy data (needs funding)

iDigBio Needs :

1. Maintaining portal and data archiving

2. Working Groups (includes new training)

3. Coordinating with other programs

4. Continued support for digitization

5. Making data available to research and education

**Key Steps to take that lead toward sustainability**

1. Finish the basic job in 10 years

2. Demonstrate general importance of program for research and education/outreach.

3. Make sure the data is available for as many purposes as possible.

4. Integrate collection data into research and education programs (e.g., integrating with molecular biology) (a proactive extension of 3)

5. Integrating with federal collections

6. Collaborate with national to international efforts

7. Promote public participation (e.g, crowdsourcing)

8. Promote Train the Trainers programs (e.g., from general georeferencing to more specific Darwin Core Webinars)

9. Train biodiversity informatics workforce, target Big Data program, work with DataOne

10. Need more informatics funding in future TCN budgets

11. Create foundation that is specific to national digitization effort

12. Seek other sources of funding for more project-specific activities (e.g., Foundations, NGO’s)

13. Develop fee services where it is feasible

TCN-Specific

1. Obtain buy-in by other groups for sustaining certain efforts (e.g., transfer Lichen TCN portal to the bryology and lichen society).

2. Draw upon state or regional foundation funding?

3. Share institution-specific ideas for ways to support TCN efforts

Inclusiveness with other initiatives

Can NIBA take over HUB responsibilities? NIBA is a separate process that includes a loose coalition of interests groups, so sustainability planning at this level would be at a different level.

Can the grow beyond TCNs in developing sustainability plans (e.g., Vertnet, BISON). Yes

Data Management

Storage group addressed this issue.

Funding

Public involvement

**Notes by Charlotte**

TCN needs for digitizing new specimens

Tom: There needs to be sustainability within TCNs. Series of workshops to train people on databasing rather than sending people a ready-made database.

Need to ask Judy about biological informatics training with NSF money.

Larry, Greg, Gil, Lucille and Neil: Sustainablity of the training as well to keep up with new technology and knowledge (ie georeferencing has changed a lot in the last 2 years). Sustainability after ADBC and iDigBio is done.

It is now part of the curating process, to digitize new incoming specimens and collections.

Larry: is NIBA a permanent group ? We don’t want effort to stop when the NSF money runs out.

Greg reminds us that we are responsible for implementing a sustainable plan after ADBC. The storage from iDigBio is not part of the sustainable plan, which is a problem.

TCNs have a short lifespan, so we cannot rely on them for sustainability plan.

iDigBio needs: maintaining the portal and databases

Tom: we expect our portal to be transferred to the society (bryology and lichen) for his TCN. Does any bigger society have the money to sustain the iDigBio portal ?

Niel: state funds or foundations for regional type projects. can pool smaller databases together to be supported at once.

programs like VertNet: David Bloom said unfunded after 2014.

Gil: there is an assumption that people will take over and see it as valuable enough to invest funding into keeping the database and portals going. How to measure the value of these databases ? ie. FSU kept moving even after the funding ran out.

Could enquire on how a database was sustained on the long term, with or without funding.

Niel: **create a sustainable working group.** VertNet will create a conference to talk about sustainability.

Key steps to take: evaluating collections and communicating value

Larry: FLAS funded by IFAS and had to be dealing with the library to be considered a digital collection. Collection managers need to get outside of their group and show how valuable these data are for dealing with biodiversity loss crisis, global warming, etc...

Anna: do we not need to propose a sustainability plan for the continuous serving of the data. NSF needs to make sure people actually have a plan to deal with the sustainability of the data produced.

Neil: Harvard put a 35% overhead reduction to maintain their museum collections (did I get this right?)

iDigBio needs: maintaining datavase.

Tom: working outside of academia to find sources of funding.

Neil: not necessarily outside of academia. Can work with development people to take over some other collections. That is what VertNet is doing.

Kansas and Berkeley is decidated to making this continue. It could be non academic society, from donors, etc.. through a university or not.

Pam: how can professional society contribute to the endowment ?

Lucille: since most journals are not printed anymore, can use these funds for the maintenance of databases.

Greg: PLoS should be interested.

Lucille: AIBS could be a funding partner.

Neil: reminds that AIBS was critical for NEON. NSF also was involved.

Can get AIBS to play a role.

NACA, SPNCH, …

Tom: is a journal involved in the publication of monographs ? Can refer to the specimens directly online.

Charlotte, Larry, Pam, Greg: could envision a system when journal demand authors to submit a voucher ID for readers to go see it online.

NCBI national library. GenBank maintained by NIH.

Greg: could go through the National Library of Medicine.

Gil: not a good idea to involved the libraries.

Anna, pam, Larry, how are other libraries funded ? JSTOR ?

Lucile: Library of congress ?

Neil created the goals for TNC and iDigBio sustainability needs. See at beginning of doc.

iDigBio needs:

staff needed to create things is different to the staff for long term sustainability.

Now: cyberinfrastructure development is the strongest right now.

later: informatics skills to integrate the data to applied questions and research projects.

Neil: my collections are not dependent on your server because it hosts it. It will only grow in the future, the interdependence of sustainability efforts. Need 1 truly supportive centralized group.

Gil: iDigBio should be working itself out of a job in 9 years. Training will be sustainable between institutions, can go to societies and bigger institutions to learn how to curate and manage collections. Then iDigBio will focus on storage server needs.

Pam, Neil: need to have focused goals to be able to ask societies for fundings.

Neil: VertNet is storing its data long term on the Apache service, free and permanent.

Larry and Greg: Where is the data backed up ? Not our goal. At some point, we will have to talk to Amazon or some group to host the data. It might just be a matter of paying for storage in 10 years.

greg: the training groups are not only for training but also for innovating.

iDigBio needs: coordinating with other programs. Steps to take 2; integrating with research questions

Bruce: could change the focus to reaching out to policy makers, land managers, etc... how these data can serve society. Our focus will change eventually.

Greg: integration with molecular biology is crucial.

Pam and Larry: integration with other disciplines will be the crucial goal of iDigBio later.

Neil: in terms of sustainability, if we can show the world that our product are valuable, it will be essential.

Larry: ecosystems services important for biodiversity. Society for ecosystem services annual meeting ?

BREAK

GO over the points by Neil.

Ed Gilbert; problem with having the personal to migrate data from Excel spreadsheets to higher software. He does it for different TCNs, but cannot anymore. IT people cannot be hired to do that, they are too expensive.

Gil: biodiversity informatics group. Need people with informatics skills. Ed: not a lot of courses that can train these people for biodiversity.

David Bloom: set up a workshop and got people from outside of the country interested in Biodiversity informatics.

Andy Miller: bioinformatics tends to attract genome annotation people. Biodiversity informatics better term.

Pam: How to budget for training of biodiversity informatics for institutions ?

Ed; Symbiota set up a couple of pages for what data they accept, how to clean their data. Data cleaning can take anything from 10mins to 10 hours for Ed when getting data from institutions.

Lucille: train the trainer workshops.

Ed: offer to put online training for everyone. 1-2 hours workshop online with Adobe Connect, then go away for a couple of days for homework. topics: Symbiota, data migration, darwin Core, etc...

Imaging docs and training by Mike Bevens at NYBG very efficient.

Neil: important to coordinate courses at universities.

Can have different workshops for imaging, by prep type for example (Gil).

curatorial staff in general. How sustainable for someone who was trained at biodiversity informatics ? Get a job in academia ?

Get somebody to take over part of what Ed is doing ?

Gil: informatics people in biology dpt need to be trained on new tools for database management, etc...

Ed and Lucile: could create certificates for biodiversity informatics training to add value to the resume of people going through it.

sustainability goal 4. it is not only the digitization support effort only, but the data itself. Working with institutions to keep the data flowing.

Greg: support for research ? Will increase as iDigBio advances.

Also need some staff who can help researchers to retrieve the data and also help them understand the limitations of the data, what you can do with it, and what you should not be inferring (example of absence data).

Niel (going over key steps): support effort for new effort. TCNs can evolve into more sustainable working groups supported by more regional agencies or societies (like the lichen database).

Leveraging what we are doing to obtain additional funding from additional programs, like the Big Data program from NSF. DataOne also: NSF effort to create metadata. Alex: DataOne archive for datasets, like physics and chemistry. Not really a biodiversity informatics thing, not that useful to the community. Help to write a data management plan for NSF.