# STORAGE DISCUSSION GROUP

Facilitator: David Bloom

Location: Cypress

<https://docs.google.com/document/d/1xdRMh-GwBV7ElrtMm_d-zqfsq8TZRRG2vadiBuOvGp4/edit>

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

Funding

Temporary, primary, backups, vs. archival

Sustainability

**TCNs**

* Invertnet
* NE Vascular Plants
* Lichens and bryophytes- label imaging only
* Macrofungi- 10% of specimens being digitized
* Tritrophic- specimens imaged

Redundancy, integration, and longevity of data

What are the expectations of iDigBio/Non-TCN to help with data storage?

iDigBio Specs: <https://www.idigbio.org/content/idigbio-image-file-format-requirements-and-recommendations>

- Handle storage of public images only, high resolution jpeg images for redistribution

- iDigBio’s role in storage of data- temporary, facilitate TCN

- Provide outside backup of stored files, guaranteed support for duration of iDigBio

-Storage via Amazon- large space, expensive

From the NEVP grant:

*“We view longer term storage and preservation of these images as a community sustainability issue and trust that a community level solution to this problem will be in place before the funding period of this grant expires.”*

- Common scientific data storage system- funded system to benefit all of science

- Storage that allows access vs Just storage

-InvertNet- live data and data stored long term

- Stored data- charge groups per their access to the data

 - charge model- scaling, charge groups different costs for the access of the data.

- Once data archive is established contract out access to the archived data

Many online databases are free to access- how to warrant charging for portal access to biodiversity collections.

 - Loan fees for herbariums and collection use?

 - Fund storage by selling computational access

Biodiversity Heritage Library- all the material being scanned have copies going into the internet library- archival

- Grant writing- plan ahead to afford digitization supplies to support their research.

 - NSF grant proposal data management plan- pushing to add criteria

 - Plans of data management and sharing of the products of research

 - Dissemination and sharing of results to community, public

 - Researchers collecting data should prep objects for digitization

 - ABDC- future collections the responsibility of the researcher to digitize

<http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp>

TCN- plan to store data once complete?

- everyone the steward of their own data. - University function?

Sustainability of iDigBio results- national and global resource

Consider data storage for the renewal of iDigBio funding from NSF- hidden costs

- \*Professional academic societies- raise dues small amount to support someone to take care of TCN data\*

- Societies that use and support the data

- Provide members with something that they would have to pay for normally

-TCNs collaborate with small institutions (PUIs- Primarily undergraduate institution) in collection digitization effort proposals to NSF- increase funding rate for TCN

Small institution has to be the only submitter, can indicate collaboration with other large institutions- grant to be used collaboratively
-Worth pursuing to see if will work or not

-Look for small institutions with well established IT departments

- Build in education and other opportunities for undergraduates