

# BUILDING SUPPORT FOR A DISCIPLINE-BASED DATA REPOSITORY

Ryan Scherle<sup>1</sup>, Sarah Carrier<sup>2</sup>, Jane Greenberg<sup>2</sup>, Hilmar Lapp<sup>1</sup>, Abbey Thompson<sup>2</sup>, Todd Vision<sup>1,3</sup>, Hollie White<sup>2</sup>

<sup>1</sup> National Evolutionary Synthesis Center (NESCent), Durham, NC;

<sup>2</sup> School of Information and Library Science, University of North Carolina, Chapel Hill, NC;

<sup>3</sup> Department of Biology, University of North Carolina, Chapel Hill, NC

Dryad data repository

<http://datadryad.org>

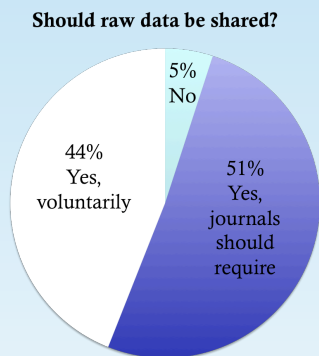
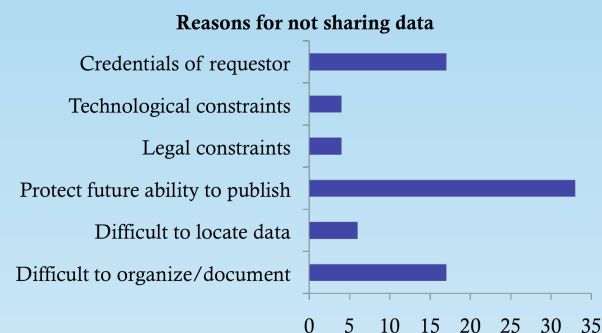
[admin@datadryad.org](mailto:admin@datadryad.org)

## SUMMARY

Dryad is a repository of scientific data from evolutionary biology and related fields. To be certain the repository meets the community's needs, we are gathering information on community practices and attitudes through a variety of methods. To further build community support, we are integrating the repository into the normal research workflow via partnerships with journals, exchange agreements with other repositories, and the solicitation of high-use datasets.

## SURVEY

A survey is currently underway to assess trends and attitudes within the community. Preliminary results are shown below.



## USE-CASE STUDIES

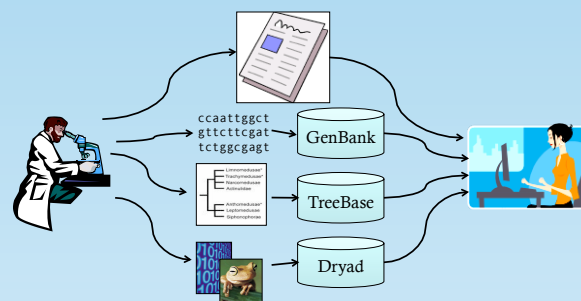
We are conducting detailed interviews with a small number of researchers in the field. These results will provide great depth, and allow us to better interpret results from the survey. Some preliminary results:

- **Data is sometimes lost** due to poor preservation habits.
- There is surprising **agreement about the proper levels of credit** for dataset creators (citation, acknowledgement, or co-authorship).
- Authors agree that data should be shared after results are published, but there is **great variety in the amount of data shared and the circumstances under which it is shared**.
- **Existing data deposition policies need to be enforced better.** GenBank submissions are typically not verified.

## WORKSHOPS

Two workshops were held to introduce the repository to the community and gather initial requirements. Results of these workshops include:

- **Immediate data collection is top priority.** Detailed metadata curation and analysis can be introduced later.
- **Metadata collection must be automated** as much as possible.
- To ensure credibility, there must be a **focus on data associated with published articles**.
- The repository **must not dictate any specific data format**.
- **Journals must be responsible for encouraging deposition** and verifying the quality of deposits.
- **All data should be openly available.**
- It is critical to **develop good practices for data citation**.



## JOINT ARCHIVING POLICY

<<Journal title>> requires, as a condition for publication, that data used in the paper should be archived in an appropriate public archive, such as GenBank, TreeBase, or Dryad. The data should be given with sufficient details that, together with the contents of the paper, allows each result in the published paper to be re-created.

## PARTNER JOURNALS

### Full partners



### Agreed to archiving policy



### Expressed interest

