

# 5th International Digital Curation Conference

December 2009

## The Dryad Repository: Designing a Curation Workflow

Jane Greenberg, University of North Carolina at Chapel Hill, USA  
Hilmar Lapp, National Evolutionary Synthesis Center, USA  
Ryan Scherle, National Evolutionary Synthesis Center, USA  
Todd Vision, National Evolutionary Synthesis Center, USA  
Hollie White, University of North Carolina at Chapel Hill, USA  
Sarah Carrier, University of North Carolina at Chapel Hill, USA  
Peggy Schaeffer, National Evolutionary Synthesis Center, USA

August, 2009

### Abstract

Dryad is a digital repository for data underlying published works in ecology, evolution, and related fields. Dryad's curation workflow integrates automatic and human metadata generation techniques and leverages depositor/scientist and professional curator expertise. Dryad's curation workflow has been informed by results from survey involving 400 prospective Dryad depositors, intensive semi-structured interviews with 17 evolutionary biologists, a metadata content analysis of eight schemes, a vocabulary mapping study including nearly 600 terms, and stakeholder feedback. This poster will highlight how key findings from these studies and stakeholder preferences have informed the design of Dryad's curation workflow.



## The Dryad Repository

Dryad (<http://datadryad.org>) is a digital repository for data underlying published works in ecology, evolution, and related fields. Dryad is supported by a collaboration involving the National Evolutionary Synthesis Center (NESCent); the Metadata Research Center at the School of Information and Library Science, University of North Carolina at Chapel Hill (UNC/CH); North Carolina State University; University of New Mexico; and Yale University. Additional partners include major societies and journals in the field of evolutionary biology. Dryad allows investigators to validate published findings, explore new analysis methodologies, and repurpose the data for research questions unanticipated by the original authors. An effective curation procedure is significant for Dryad's success. This poster reports on several studies providing results helpful to developing the repository's curation workflow.

## Research Supporting Curation Plans

The Dryad development team has undertaken a series of analyses to develop an effective curation plan:

- A *survey* of 400 evolutionary biologists provides empirical data on their data sharing attitudes and behaviors. Results documenting biologists' interactions with existing data archives; their data sharing practices; and their dependency on digital media for research and reporting indicate various curatorial skills and limitations of prospective Dryad depositors.
- *Intensive semi-structured interviews* with 17 evolutionary biologists provide examples of how scientists describe data (White, 2008), and point to system features and functionalities that can aid depositor curation.
- A *metadata content analysis* of eight schemes and a *vocabulary mapping study* including nearly 600 terms (Greenberg, 2009) indicate where automatic metadata generation techniques can expedite Dryad's curation workflow.

## Conclusions

Findings from the studies outlined above, and results from Dryad stakeholder meetings, support a curation workflow integrating automatic and human metadata generation techniques, and leveraging depositor/scientist and professional curator expertise. This poster will highlight key findings informing Dryad's curation workflow.

## Acknowledgements

This work is supported by NSF EF-0423641 and the IMLS/NLG program.

## References

- [journal article] Greenberg, J. (2009). Theoretical considerations of lifecycle modeling: an analysis of the Dryad repository demonstrating automatic metadata propagation, inheritance, and value system adoption. *Cataloging & Classification Quarterly*, 47, 3: 380-402
- [proceedings] White, H.C. (2008). Exploring evolutionary biologists' use and perceptions of semantic metadata for data curation. *International Conference on Dublin Core and Metadata Applications. Berlin, Germany.*