

DRIADE: A Data Repository for Evolutionary Biology

Jed Dube

School of Information and Library
Science, University of North
Carolina at Chapel Hill
+01-919-962-8066

jdube@email.unc.edu

Sarah Carrier

School of Information and Library
Science, University of North
Carolina at Chapel Hill
+01-919-962-8066

scarrier@email.unc.edu

Jane Greenberg

School of Information and Library
Science, University of North
Carolina at Chapel Hill
+01-919-962-8066

janeg@ils.unc.edu

ABSTRACT

NESCent (The National Evolutionary Synthesis Center) is developing DRIADE (Digital Repository of Information and Data for Evolution) to address synthetic research challenges fundamental to advancing the field of evolutionary biology. This poster highlights results from a survey of selected repositories' functionalities, DRIADE's functional requirements, and DRIADE's functional model. We also summarize ongoing research activities, studying evolutionary biologists' data preservation practices and use requirements.

Categories and Subject Descriptors

H.3.7 [Digital Libraries] *Systems issues, user issues*; H.3.5 [Online Information Services] *Data sharing, Web-based services*.

General Terms

Design, Standardization.

Keywords

Scientific data, digital data repository, evolutionary biology, cyberinfrastructure.

1. EVOLUTIONARY BIOLOGY AND THE DRIADE PROJECT

Evolutionary biology is an interdisciplinary field drawing from ecology, genomics, paleontology, population genetics, physiology, and systematics, and integrating field, laboratory and theoretical approaches. Evolutionary biology requires the integration of diverse kinds of data, and general conclusions in the field often require the synthesis of the results from many individual small scale studies. This poses a formidable challenge to the field, exacerbated by the generation of voluminous data from new biological subdisciplines such as genomics.

Digital library and repository developments offer an obvious means for addressing this challenge. Evolutionary biologists are increasingly being required to submit data supporting publications to both journal repositories and authoritative data systems (e.g., GenBank, TreeBASE) as part of the publication process. These requirements support open science objectives.

Limited interoperability among repositories, however, makes the current process tedious.

Copyright is held by the author/owner(s).

JCDL '07, June 18–23, 2007, Vancouver, British Columbia, Canada.

ACM 978-1-59593-644-8/07/0006.

NESCent (National Evolutionary Synthesis Center), in collaboration with the Metadata Research Center at the University of North Carolina at Chapel Hill, is addressing this problem through DRIADE (Digital Repository of Information and Data for Evolution) - a repository being developed as the primary home for published data in the field of evolutionary biology.

2. FUNCTIONAL REQUIREMENTS

The first phase of DRIADE has included identifying functional requirements. We conducted a survey of selected leading digital data and resource repository initiatives. Among systems examined were Global Biodiversity Information Facility (GBIF), Knowledge Network for Biocomplexity (KNB), Science Environment for Ecological Knowledge (SEEK), National Science Digital Library (NSDL), Interuniversity Consortium for Political and Social Research (ICPSR), and Marine Metadata Initiative (MMI). We examined the scope, goals, and function of each project, with specific focus on their support for the following set of key DRIADE parameters: heterogeneous digital datasets; long-term data stewardship; tools and incentives to researchers; minimized technical expertise and time requirements; intellectual property rights; published Datasets.

Our survey allowed us to develop a framework and define DRIADE's initial functional requirements, which include: computer-aided metadata generation and augmentation; specialized modules linking data submission and manuscript review; data and metadata quality control by integrating human and automatic techniques; support for identity, authority and data security; support for basic metadata repository functions, such as resource discovery, sharing, and interoperability.

We developed a simple functional model based on OAIS. Our work also allowed us to consider goals and requirements for later phases.

3. NEXT STEPS

Among next steps, we will be conducting a two-part study, consisting of a survey and semi-structured interviews, to better understand evolutionary biologists' data use and preservation practices, and to further inform our functional requirements and data/metadata model.

4. ACKNOWLEDGMENTS

This work is supported by National Science Foundation Grant # EF-0423641.