The Dryad Digital Data Repository: Coupling the Publication of Scientific Articles and Data

Todd Vision[1,2] and Ryan Scherle[1]

[1]U.S. National Evolutionary Synthesis Center, Durham NC; [2]University of North Carolina, Chapel Hill, NC

Many scientific datasets are lost once the findings based on them appear in print, jeopardizing the reproducibility of results, and foregoing their potential for reuse. Dryad is a digital data repository, governed by a consortium of biology journals that seeks to promote data archiving through a tight coupling of the processes of article publication and data publication.

Dryad is designed to allow for a very low-burden submission process. This is achieved in part by obtaining bibliographic metadata automatically from publishers, rather than from individual authors. Low-burden submission is also based on a metadata curation model: minimal information is required of depositors, light manual curation ensures basic discoverability, and semi-automated metadata curation is used to enrich discoverability further.

Dryad promotes reuse of data in a number of ways. Authors agree to an Open Data license upon submission, with an option for a time-limited embargo, so that restrictions on data reuse expire when they no longer serve as an incentive to deposit. Data records are assigned permanent identifiers in the form of DataCite DOIs, enabling persistent links between each article in the partner journal and its data record in the Dryad repository. A citation policy is conspicuously promoted so that depositors may receive professional credit for instances of data reuse.

Dryad seeks to provide value-added services above and beyond that typically provided by the Supplementary Materials hosted by publishers. There is no pay wall to access. Data may be searched across journals and publishers, and searches may be based on the content of either the articles or the data. Data may be provided in author-specified formats that emphasize suitability for reuse rather than display. Standard machine formats can be migrated, and authors may update data records since Dryad supports versions through its interface and DOI assignment scheme. File size limits are generally less restrictive. Web services for query and retrieval can be supported, and content can be harvested in a number of ways. Dryad also provides support for anonymous review of pre-published data to partner journals.

This demonstration will show how these features have been built into a customization of the DSpace framework, effectively making software designed for an institutional repository into a very powerful medium for data publication.