The Dryad Data Repository

Summary
Dryad is a repository of data underlying scientific publications, with an initial focus on evolution, ecology, and related fields. Dryad allows investigators to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies such as formal meta-analyses.

Basic search interface
Dryad allows data to be searched using standard publication information such as title and authors. Searches can also include more detailed information, such as taxonomic names and geological timespans.

Submission system
Dryad's submission system is optimized for quick and easy submissions. Only a few pieces of information about a publication and dataset are required. Users have the option to enter more detailed descriptions, making data easier for others to find and reuse (and thus more likely to receive subsequent citations).

Harvesting and searching related content
Using the OAI-PMH and OAI-ORE protocols from the Open Archives Initiative, Dryad will harvest content from related repositories, including the Knowledge Network for BioComplexity and the Long Term Ecological Research Network. Harvested content can be searched alongside native Dryad content, providing a single place to search multiple related repositories.

Journal integration
Partner journals forward metadata about accepted publications to Dryad. Authors can import this information, greatly reducing the time required to submit data.

Joint Data Archiving Policy
Partner journals have agreed to jointly enact a data archiving policy. This policy will ensure that all data associated with papers in participating journals is saved in appropriate repositories. The current draft of the policy states:

"<<Journal>> requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive, such as <<list of approved archives>>. Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. Authors may elect to have the data publicly available at time of publication, or, if the technology of the archive allows, may opt to embargo access to the data for a period up to a year after publication. Exceptions may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species."

Machine-readable interfaces
Dryad will provide multiple interfaces for researchers and other systems to access content in Dryad. Content can be monitored via RSS feeds, searched via the SRU searching standard, and harvested via the OAI-PMH protocol.

Modifications to DSpace
The implementation of Dryad has required many changes to the core DSpace platform, including grouping of search results by publication and the ability to embargo datasets for up to one year. When these modifications meet the needs of the larger DSpace community, they are integrated into the core DSpace software.

Handshaking with specialized databases
For databases that are widely used by Dryad’s audience (e.g., TreeBASE and GenBank), Dryad will work with the database to mirror data submitted to the database and/or facilitate automatic deposit of Dryad material into the database.

Related projects
The HIVE project is developing tools for integrating controlled vocabularies and ontologies with repositories. HIVE will integrate with the Dryad submission system.

Dryad is a member of the DataONE consortium of repositories, which is developing tools for wide-scale data sharing, mirroring, and analysis.