Coupling Research Data to Publications in the Life Sciences

Todd Vision
University of North Carolina at Chapel Hill
ICSTI 'Delivering Data in Science' workshop
Paris, 5-Mar-2012
[ @tvision, @datadryad ]
What happens to the data behind the 1.5 million STM articles published each year?
It’s the dodgy stuff that doesn’t get shared

Reporting errors concerned with $p < .05$:

- Not shared: 1 error (18%), 2 errors (4%), 3 errors (3%), no errors (75%)
- Shared: no errors (100%)

Wicherts et al. (2011) doi:10.1371/journal.pone.0026828
Information Content

Time of publication
Specific details
General details
Retirement or career change
Accident
Death

Time

(Michener et al. 1997)
Today's Menu

Coordinated data archiving policy

A repository infrastructure to support that policy

Strengths and weaknesses of the Dryad model
Joint Data Archiving Policy
Small science ➔ long tail ➔ orphan data

Specialized repositories (e.g., EMBL)

Volume

Rank frequency of datatype

after B. Heidorn
A diagram outlines the process for publishing data alongside an article. It starts with an author preparing a manuscript and related data files. The manuscript is submitted to the journal, where it undergoes peer review. If accepted, the article is published with a data citation. If not accepted, the process returns to manuscript preparation. The data is then uploaded to Dryad, where it is packaged and curated. If accepted, the data is published with an article citation. If not accepted, the data is sent back to the author.
Dryad is an international repository of data underlying peer-reviewed articles in the basic and applied biosciences. Dryad enables scientists to share their data, and makes it possible to compare, reanalyze methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. By collaboratively promoting data archiving and ensuring the sustainability of the repository.

As of Nov 13, 2011, Dryad contains 1070 data packages and 2539 data files, associated with articles in 93 journals.

Recently Published Data


When using this data, please cite the original article:


Additionally, please cite the Dryad data package:

• http://datadryad.org
• http://blog.datadryad.org
• http://datadryad.org/wiki
• http://code.google.com/p/dryad
• dryad-users@nescent.org
• @datadryad
• Dryad