Supporting scientific communities by publishing data

Dryad Digital Repository
Peggy Schaeffer
OpenAIRE/LIBER Workshop
May 28, 2013
Ghent, Belgium
Outline

• Introduction to Dryad
  Joint Data Archiving Policy (JDAP)
  How Dryad works with journals and publishers
• How librarians can use Dryad
• Ideas for librarians to support research data management & publication
An international repository of data underlying scientific and medical publications

datadryad.org
Dryad welcomes data in any format
Joint Data Archiving Policy

< 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.
Researchers and journals are using Dryad for archiving

DataDryad.org is a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad has integrated data submission for a growing list of journals; submission of data from other publications is also welcome.

Browse for data

Recently Published Data


Stats

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data packages</td>
<td>3052</td>
</tr>
<tr>
<td>Data files</td>
<td>8723</td>
</tr>
<tr>
<td>Journals</td>
<td>211</td>
</tr>
<tr>
<td>Authors</td>
<td>11780</td>
</tr>
<tr>
<td>Downloads</td>
<td>2174140</td>
</tr>
</tbody>
</table>
...and using the data for research

<table>
<thead>
<tr>
<th>Title</th>
<th>Sequences after initial processing in QIIME (seqs.fna, library 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downloaded</td>
<td>61 times</td>
</tr>
<tr>
<td>Description</td>
<td>Raw sequence data from the Roche 454 GS FLX sequencer, region 1 (split_library_output_1). These data are the output of the command: split_libraries.py -m 454_Map.txt -f 1.TCA.454Reads.fna -q 1.TCA.454Reads.qual -o split_library_output_1/ -l 100 -L 700 -H 9 -M 2 -b 10</td>
</tr>
<tr>
<td>Download</td>
<td>README.txt (3.755Kb) View File Details</td>
</tr>
<tr>
<td>Download</td>
<td>seqs.fna (20.96Mb) View File Details</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title</th>
<th>Sequences after initial processing in QIIME (seqs.fna, library 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downloaded</td>
<td>33 times</td>
</tr>
</tbody>
</table>
Data from: Towards a worldwide wood economics spectrum

Files in this package

To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.

Title  
Global Wood Density Database

Downloaded  
4724 times

Description  
Please direct all correspondence to G. Lopez-Gonzalez <G.Lopez-Gonzalez@leeds.ac.uk>

Download  
GlobalWoodDensityDatabase.xls (2.047Mb)  View File Details

When using this data, please cite the original article:


Additionally, please cite the Dryad data package:

# Journals benefit when data is reused

<table>
<thead>
<tr>
<th>Journal</th>
<th>Integration Date</th>
<th>Data Packages</th>
<th>Data Downloads</th>
<th>Ave. downloads per package</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Journals</td>
<td></td>
<td>2765</td>
<td>127,396</td>
<td>46</td>
</tr>
<tr>
<td>Molecular Ecology</td>
<td>2009-11-29</td>
<td>615</td>
<td>23,604</td>
<td>38</td>
</tr>
<tr>
<td>Evolution</td>
<td>2010-5-4</td>
<td>380</td>
<td>12,524</td>
<td>33</td>
</tr>
<tr>
<td>American Naturalist</td>
<td>2009-8-29</td>
<td>208</td>
<td>11,195</td>
<td>54</td>
</tr>
<tr>
<td>Journal of Evolutionary Biology</td>
<td>2010-7-12</td>
<td>205</td>
<td>5,729</td>
<td>28</td>
</tr>
</tbody>
</table>

A “Data Package” is all of the data files for a journal article. All Dryad data packages link to the associated journal article.
What makes Dryad unique

1. Aligns data deposition process with the process and business of scholarly publishing;
2. Article publication as a model for how researchers can benefit from data sharing;
3. Motivates researchers to disclose data of the greatest value for scientific reuse;
4. Article metadata = foundation of metadata for associated data.
What makes Dryad unique

5. Data files in Dryad are curated;
6. All contents are freely available via CC0 waiver;
7. Dryad is an open source enterprise, built on DSpace, with open development & open documentation; and
8. Dryad is a nonprofit organization responsive to and managed by its stakeholders.

DataDryad.org
How can librarians use Dryad

- **Ask** scientists about their data!
- **Refer** grant writers to Dryad, they can build it into a data management plan
- **Show** Dryad to researchers
- **Add** a link to Dryad when you offer resources for researchers
- **Demo** Dryad to show what making data files public looks like.
Data sharing: advantages to authors

- Visibility
- Citability
- Workload reduction
- Preservation
- Impact and opportunity
Other ways librarians can support scientific data management

Consult and share *best practice guidelines*:


Use & promote the use of good data citations with DOIs

- Data citation conventions are evolving
- Authors, journals and publishers need to see good models of data citation
  - Articles
  - CVs
  - Grant proposals
- Help make data citation and data DOIs familiar
- Use DOIs in social media
Dryad’s citation philosophy:

• **Cite both** the article and the data – they are both useful research products

• But limit data citations to one “data package” per article – this eliminates most concerns about the size/granularity of data files
Global Wood Density Database

To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.

Title: Global Wood Density Database
Downloaded: 4724 times
Description: Please direct all correspondence to G. Lopez-Gonzalez <G.Lopez-Gonzalez@leeds.ac.uk>
Download: GlobalWoodDensityDatabase.xls (2.047Mb) View File Details

When using this data, please cite the original article:

Additionally, please cite the Dryad data package:

DataDryad.org
34. We thank R. FitzJohn, A. Mooers, R. Olmstead, T. Paape, L. Popović, D. L. Shade, and R. Ree for help and discussions. Data are deposited in the Dryad database with accession no. 1888. Funded by NSF grant DEB-0919089 to B.I.
Checking citations to the data in the data-sharing article

For 338 articles associated with Dryad data:

-- 253 did include a DOI for the data (75%)
-- 85 did not (25%)

• where the DOIs were located:
  -- dedicated section (Data accessibility) n= 148
  -- in or near article header  n= 43
  -- in-text (Methods, Acknowledgments): n= 71
  -- in References: n= 28 (but: 17 are not actual full citations in the style of the other citations).
Encourage opportunities for libraries and librarians

• Scientific data positions at academic libraries, universities, & research institutes
• Scholarly communications officers
• Collaborative roles with researchers can extend to consulting on data sharing
• Other emerging roles
• ??????
To learn more:

- Repository: [http://datadryad.org](http://datadryad.org)
- News: [http://blog.datadryad.org](http://blog.datadryad.org)
- Documentation: [http://wiki.datadryad.org](http://wiki.datadryad.org)
- Twitter: @datadryad

Peggy Schaeffer
pschaeffer@datadryad.org