Incorporating Data Access into Journal Workflow

Mercè Crosas, Ph.D.
Twitter: @mercecrosas
Director of Data Science, IQSS
Harvard University
The Harvard Dataverse Network is a data sharing repository open to all research data from all domains.

The Dataverse Network software is open-source, installed in institutions across the world (http://thedata.org)
**Dataverse**: Container for research studies

**Study**: Container for data, documentation, and code
Data sharing and archiving with control and recognition for data authors, distributors

Persistent Data Citations
permanently linking your data to your publication (Altman, King, 2007)

Customized Branding
or embed on your site

Support for all file types
any format, max 2 GB per file

Data Restrictions
& terms of use options, although encouraging Open Data
Rich data support for some data formats

*SPSS, Stata, R Data*
metadata extraction, subsetting & analysis (R, Zelig)

*FITS Data*
metadata extraction from file header

*Social Network Data (GraphML)*
smart queries & subsetting

*Data visualizations*
for time series
Data management, standards and archival good practices

Data Cataloging
self-curated, with custom metadata templates (DDI, Dublin Core)

Log traffic & downloads
to your dataset with Guestbook

Data Versioning
preserve & cite previous versions

Permanent storage
preservation format with w/copies in multiple locations (OAI-PMH, LOCKSS)
Dataverse for an Individual Researcher
Dataverse for an Organization

International Food Policy Research Institute (IFPRI) Dataverse

In collaboration with institutions throughout the world, IFPRI is often involved in the collection of primary data and the compilation and processing of secondary data. The resulting datasets provide a wealth of information that IFPRI freely distributes as many of these datasets as possible and requires proper citation. For more information, please contact Data@cgiar.org or IFPRI-Library@cgiar.org for questions.

IFPRI Data by Type and Region

- Country Level
- Geospatial
- Household Surveys
- Institution-level Surveys
- Regional
- Social Accounting Matrix (SAM)

A 2007-2008 Social Accounting Matrix for Pakistan
by Debowicz, Dario; Dorosh, Paul A.; Robinson, Sherman; Haider, Syed Hamza
Description: This Social Accounting Matrix (SAM) was built for the year 2007–2008 for Pakistan. The proposed approach to estimating SAMs is motivated by an information theoretic approach to estimation (Judge & Mittelhammer, 2012) that takes a Bayesian approach. Continue [+]

Search Studies: 95 | Downloads: 17888

Advanced Search  Tips
### Review of Economics and Statistics Dataverse

This page contains replication data and code for the articles published in the Review of Economics and Statistics. Please notice that most articles in volumes previous to volume 92 were not subject to the journal's data availability policy.

#### Search Studies

- **Sort By:** Global ID
- **Studies:** 218 | **Downloads:** 21264

#### Volume 92

- Volume 92, Issue 1
- Volume 92, Issue 2
- Volume 92, Issue 3
- Volume 92, Issue 4

#### Volume 93

- Volume 93, Issue 4

#### Volume 94

- Volume 94, Issue 1
- Volume 94, Issue 2
- Volume 94, Issue 3

---

**Dataverse for a Journal**
Seamless Integration between Dataverse and Journals

PKP’s Open Journal System (OJS)

Harvard Dataverse Network

Citation to Data

Citation to Article

OJS plugin for:

Data + metadata + supporting files, sent via SWORD API to the Dataverse
OJS Journals and Dataverses Growth

From 1990 to 2013:
5000 active OJS Journals

From 2007 to 2013:
500 Dataverses

Credit: Juan Alperin, PKP; Gustavo Durand, Dataverse
First, set up a Datavase for each Journal
Step 4. Uploading Data Files

You may provide access to data files and other supplementary materials related to your work. Upload files to the journal's Dataverse or provide a citation and link to data already published online.

Upload Data Files

<table>
<thead>
<tr>
<th>ID</th>
<th>TITLE</th>
<th>ORIGINAL FILE NAME</th>
<th>DATE/uploaded</th>
<th>ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Untitled</td>
<td>19-282-1.csv</td>
<td>05-24</td>
<td>EDIT</td>
</tr>
</tbody>
</table>

Choose data type *

CSV (w/ SPSS card)

Some file types can be subset and analyzed using the Dataverse Network analysis tools. Learn more about subsetting data files.

Upload data file *

Choose File No file chosen Upload

Terms of use *

I agree to the terms of use governing the submission of data files to the journal's Dataverse.

Save and continue Cancel

Link to Data Files

If your data is already published in a repository, please provide a citation and link. Please note that for review purposes, your data must be published and/or otherwise downloadable.

<table>
<thead>
<tr>
<th>ID</th>
<th>CITATION</th>
<th>PERSISTENT ID</th>
<th>DATE ADDED</th>
<th>ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>[citation]</td>
<td>hdl:00001/0002</td>
<td>05-24</td>
<td>EDIT</td>
</tr>
</tbody>
</table>

Citation *

Persistent Identifier *

Provide a persistent identifier, e.g., DOI or HANDLE.
Metadata fields will be selected ahead of time by journal admin.
Moving toward a recovery-oriented approach in the Swedish mental health system

Carolina Klockmo

1. Replication data for: Moving toward a recovery-oriented approach in the Swedish mental health system

Type: Data Set
Data Citation: Carolina Klockmo, 2012, "Moving toward a recovery-oriented approach in the Swedish mental health system", doi:10.9999/00000
Full Record Available In: COAction Dataverse Network
Data Tools: INDEXING METADATA
ANALYZE
DOWNLOAD
Data in Dataverse - Linked to Published Article

Co-Action Dataverse

REPLICATION DATA FOR: MOVING TOWARD A RECOVERY-ORIENTED APPROACH IN THE SWEDISH MENTAL HEALTH SYSTEM
hdl:TEST/10000

DRAFT Study Version: 1 – No Released Version
Created: Thu Dec 20 11:30:31 PST 2012 – Last Updated: Thu Dec 20 11:32:19 PST 2012

CATALOGING INFORMATION

If you use these data, please add the following citation to your scholarly references. Why cite?

Data Citation

Klockmo, Carolina, 2012-12-13, "Replication data for: Moving toward a recovery-oriented approach in the Swedish mental health system", http://hdl.handle.net/TEST/10000 V1 [Version]

Citation Format

Publications

Vulnerable Groups & Inclusion. http://dx.doi.org/10.3402/vgi.v3i0.18879
ID: DOI:10.3402/vgi.v3i0.18879
Link

Data Citation Details

Study Global ID
hdl:TEST/10000

Authors
Klockmo, Carolina (Department of Health Sciences, Mid Sweden University)

Production Date
December 13, 2012

Contact
Klockmo, Carolina (Department of Health Sciences, Mid Sweden University), carolina.klockmo@miun.se

Distribution Date
December 13, 2012

Deposit Date
December 13, 2012

Provenance
Co-Action Dataverse
Data Publishing Workflow

Submission Article + Data

Review, Approved
- Article published in journal (OJS)
  - # of article downloads
- Data published in Dataverse
  - # of data set downloads

Review, Not Accepted
- Data published in Dataverse?
Alternative Workflow

Submission Article → Add Data Citation to Article

Data already in Repository ← Submit Article Citation to Data

Track # of article downloads and # of data sets downloads
Participating Journals and Outreach

**Beta Testers Criterion:** Current, quantitative OJS journals interested in data sharing

- **Beta Testers:** 43 OJS journals, from 6 publishers (social sciences, health, life sciences)
- **Extended Testers:** > 400 OJS journals (economics, social sciences, health, life sciences)
- **Outreach:** Beyond OJS, journals using other publishing systems

- **Summer/Fall 2013**
- **Fall 2013/Winter 2014**
Metajournals as incentives

Credit: Brian Hole, Ubiquity Press (beta tester Publisher for Dataverse integration)
Amsterdam Manifesto:

The Amsterdam Manifesto on Data Citation Principles

Preface:
We wish to promote best practices in data citation to facilitate access to data sets and to enable attribution and reward for those who publish data. Through formal data citation, the contributions to science by those that share their data will be recognized and potentially rewarded. To that end, we propose that:

1. Data should be considered citable products of research.
2. Such data should be held in persistent public repositories.
3. If a publication is based on data not included with the article, those data should be cited in the publication.
4. A data citation in a publication should resemble a bibliographic citation and be located in the publication’s reference list.
5. Such a data citation should include a unique persistent identifier (a DataCite DOI recommended, or other persistent identifiers already in use within the community).
6. The identifier should resolve to a page that either provides direct access to the data or information concerning its accessibility. Ideally, that landing page should be machine-actionable to promote interoperability of the data.
7. If the data are available in different versions, the identifier should provide a method to access the previous or related versions.
8. Data citation should facilitate attribution of credit to all contributors.