The Dataverse Network and OJS Project to Encourage Data Sharing & Citation in Academic Journals

Eleni Castro (Project Coordinator)
Institute for Quantitative Social Science (IQSS)
Harvard University
@thedataorg

IASSIST 2013 - Cologne, Germany
Motivation
Why Connect Published Work to Data

Published Results

Data + Metadata + Supporting Files
(documentation, code)

= A third party can replicate or reuse, thus validate and advance science
Quotes for “Why?”

"The most immediate of these obstacles is the lack of a consolidated infrastructure for the easy sharing of data" - JORD Project results via Edawax blog

"Any moves towards data sharing are dependent upon the cooperation of journals.”* – Sergiu Ghergina and Dr. Alexia Katsanidou

*from European Political Science 2013: Data Availability In Political Science Journals
A team was assembled...

2 year Sloan Foundation grant (05/12-05/14):

- Public Knowledge Project (PKP)
  - Simon Fraser University
  - Stanford University (John Willinsky)

- Dataverse Network Project
  - Harvard University’s Institute for Quantitative Social Science (IQSS) (Gary King & Merce Crosas)
  - Micah Altman – Director of Research at MIT
Who?: Address the needs of journal publishers and editors in addition to researchers and data managers.

What?: We propose to enable journals to: Seamlessly manage the submission, review, and publication of data associated with published articles.

How?: We will help build the needed technology and create awareness among journal editors and publishers regarding the importance of data sharing and preservation.
The End Result?

Help increase the replicability and reusability of published work in social science (and other disciplines) by improving the infrastructure for, practice of, and incentives related to data publication and citation.
Integrating Open Source Systems

We plan to do this by integrating two well-established open-source systems:

1. **Open Journal Systems** (OJS)  
   [Willinsky 2005]

2. **Dataverse Network**  
   [King 2007; Crosas 2011]
Open Journal Systems (OJS)

Open source journal management and publishing system created by PKP to expand & improve access to research.
S publishes peer-reviewed essays on Lacanian and related topics from the fields of art, film and literary criticism, political, philosophical and ideological critique. With permission, S also re-publishes hard-to-obtain essays and translations from seminal thinkers in Lacanian studies whose work deserves the worldwide dissemination open access publishing affords.

Vol 5 (2012): Topology
Special topic on topology edited by Benjamin Bishop

Table of Contents

Editorial
Topology and the Inscription of the Clinic
Benjamin Joseph Bishop

Articles
A Method of Reading a Knot
Jean-Michel Vapereau

Generalized Placement: Elements of Analytic Knot Theory
Robert Groome

Topological Objects and the Current State of Mathematics
Pierre Soury

A Year in the Company of Knots
Pierre Soury

Dialogues
Homology: Marx and Lacan
Samo Tomšič

From Representation to Class Struggle: A Response
Pietro Bianchi

Interventions from the Clinic
Current Controversies in the Treatment of Autism in France
Marie-Jean Sauret, Sidi Askoarfé, Pascale Macary-Garipuy
OJS Journals w/ at least 10 articles per year

2012 = 5117

Credit: Juan Alperin, PKP
OJS provides...

✓ Technical infrastructure for an entire editorial management workflow (submission, peer review, indexing).

✓ ‘Plugin' architecture (a la WordPress, Drupal, etc) to easily add new features without the need to change the entire core code base.

✓ Universal Metadata standards (expandable Dublin Core).

✓ Inter-operability with other systems through standard protocols (such as OAI-PMH, APIs).

✓ Option for backups and replication of journals in different locations (LOCKSS) so access is never lost.
Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researchers and data producers keep control of and get recognition for their data.
A Dataverse is a virtual data archive with its own branding.

A Study describes and holds the Data.
Dataverse Network provides...

✓ Option for backups and replication of data in different locations (LOCKSS) so data is never lost.

✓ Re-format for long term accessibility so data never become obsolete.

✓ Extract Variable Metadata from data sets.

✓ Universal Metadata standards (DDI, Dublin Core).

✓ Inter-operability with other systems through standard protocols (such as OAI-PMH, APIs).

✓ Generates a Handle for permanent linking to datasets.

The Dataverse takes care of the archival infrastructure (“plumbing”) for you!
OK so what is the integration going to do?

OJS plugin for:
Data + metadata + supporting files,
sent via SWORD API
to the Dataverse
Which Workflow?

1. Submit paper
   - Approve paper + ask for data
   - Publish paper in OJS + release data in Dataverse

2. Submit paper + data
   - Approve paper + review data
   - Publish paper in OJS + release data in Dataverse

3. Submit paper + data
   - Reject paper + review data
   - Release data in Dataverse

+ 4, 5, 6, ....

Slide acknowledgement: Merce Crosas
Progress to-date

Compiled a list of potential journals (~100) that we can work with. Contacted a small sample of publishers to be our 1st round of pilot testers (30+ journals).

Publishers reviewed our plugin workflow and mockups to provide feedback before beginning development.
Mockups of OJS Plugin: Journal Setup

Dataverse Demo Journal

Dataverse Plugin

DATAVERSE SETTINGS
Configure Dataverse connection.

Dataverse URL *

Username *

Password *

Save Cancel

* Denotes required field
Mockup of Data Deposit (in OJS)

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 submittable 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:

Provide a persistent identifier, e.g., DOI or HANDLE.
Metadata

To index this supplementary material, provide the following metadata for the uploaded file.

<table>
<thead>
<tr>
<th>Field</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title*</td>
<td>Untitled</td>
</tr>
<tr>
<td>Author(s)</td>
<td></td>
</tr>
<tr>
<td>Keywords</td>
<td></td>
</tr>
<tr>
<td>Brief description</td>
<td></td>
</tr>
</tbody>
</table>

Category: Data Files

Date: 2012-12-11

Date when data was collected or instrument created.

Supplementary File

<table>
<thead>
<tr>
<th>Field</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>File name</td>
<td>3-18-1-SP.pdf</td>
</tr>
<tr>
<td>Original file name</td>
<td>19-282-1-PB.pdf</td>
</tr>
<tr>
<td>File size</td>
<td>248KB</td>
</tr>
<tr>
<td>Date uploaded</td>
<td>2012-12-11 02:09 PM</td>
</tr>
</tbody>
</table>

Present file to reviewers (without metadata), as it will not compromise blind review.

Replace file

Save and continue | Cancel

* Denotes required field
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 Citation

Full Record Available In: COACTION DATaverse NETWORK

Data Tools:
- INDEXING METADATA
- ANALYZE
- DOWNLOAD
Data in the Dataverse

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

1. Begin development of plugin and API (*SWORD2-compliant*) (Summer 2013).

2. Additional publishers will test + provide feedback through a survey (Fall 2013).

3. Provide best practices for data review/sharing policies and data citation (starting Fall 2013).

4. Test and release updated versions of OJS + Dataverse Network (Spring 2014).

5. Make code & docs available for everyone.
# High Level Project Roadmap

<table>
<thead>
<tr>
<th>Phase</th>
<th>Milestones/Deliverables</th>
<th>Timeframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Contact OJS Journals for feedback on OJS Dataverse Plugin Workflow and Use cases.</td>
<td>Fall 2012-Winter 2013</td>
</tr>
<tr>
<td>2</td>
<td>OJS-Dataverse team develops plugin/API.</td>
<td>Spring-Summer 2013</td>
</tr>
<tr>
<td>3</td>
<td>Journals test + give feedback.</td>
<td>Fall 2013</td>
</tr>
<tr>
<td>4</td>
<td>Gather feedback from pilot testers.</td>
<td>Fall 2013</td>
</tr>
<tr>
<td>5</td>
<td>Test + Release new versions of OJS and Dataverse.</td>
<td>Winter-Spring 2014</td>
</tr>
<tr>
<td>6</td>
<td>Technology made available for others. Release best practices, and other documentation.</td>
<td>Spring 2014</td>
</tr>
</tbody>
</table>
Some Advantages to Integration

1. Streamlining authors’ article and data deposit process.

2. Permanent 2-way linking of the published article with its archived data.

3. Increase visibility/access, and encourage data citation, replication and re-use.
Thank you!

Contact
Eleni Castro (Project Coordinator)
ecastro@fas.harvard.edu

Project Web site / References
http://projects.iq.harvard.edu/ojs-dvn

