

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

Eleni Castro

Institute for Quantitative
Social Science (IQSS)
Harvard University



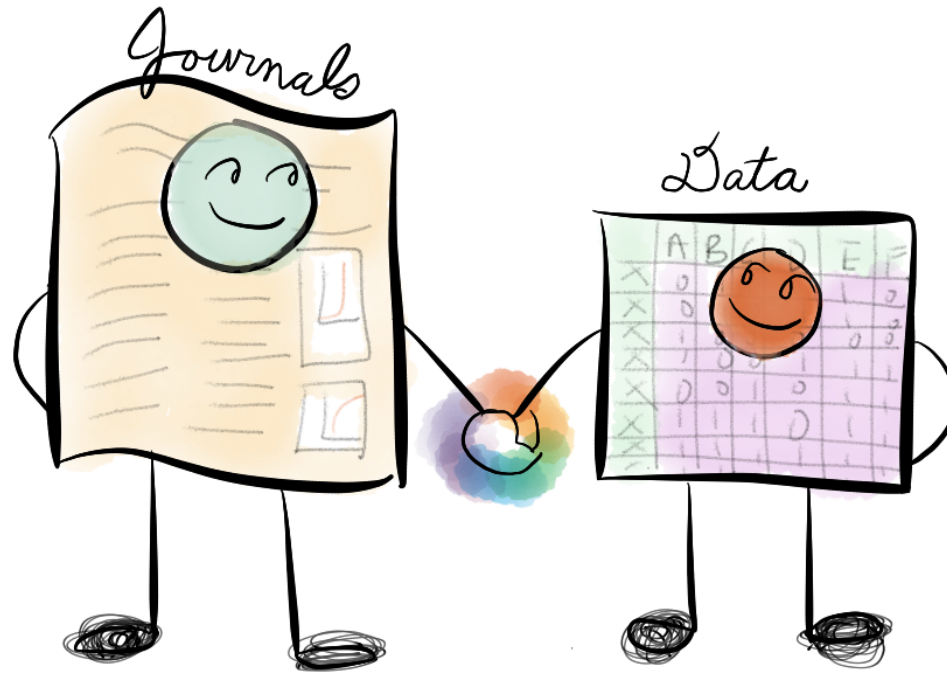
@thedataorg

Alex Garnett

Public Knowledge Project
Simon Fraser University

@axfelix

Motivation



Why Connect Published Work to Data



+

**Data + Metadata +
Supporting Files**
(documentation, code)

The image is a screenshot of a data table, likely from a spreadsheet application. It shows a grid of numerical data with multiple columns and rows. The data appears to be organized into several columns, possibly representing different variables or measurements over time or across different conditions. The table is displayed in a window with a standard operating system interface.

=

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

United States

- **National Science Foundation:**

- “The expectation is that all data will be made available after a reasonable length of time. “
- “... will be determined by the community of interest through the process of peer review and program management.”

- **National Institutes of Health (NIH)**

- 2008 mandated requirement for researchers to deposit their peer-reviewed, NIH-funded research articles in PubMed Central
-

United Kingdom

Research Council of the UK

- Publicly funded research data are a public good
- Data management plans should be developed in accordance with relevant standards
- Metadata should be deployed to ensure data discoverability
- Data should be cited appropriately

Engineering and Physical Sciences Research Council

- Effective data curation principles will be employed
 - Data will be preserved for a minimum of 10 years
-

Canada

- **Social Science and Humanities Research Council (SSHRC):**

"All research data collected with the use of SSHRC funds must be preserved and made available for use by others within a reasonable period of time. SSHRC considers "a reasonable period" to be within two years of the completion of the research project for which the data was collected."

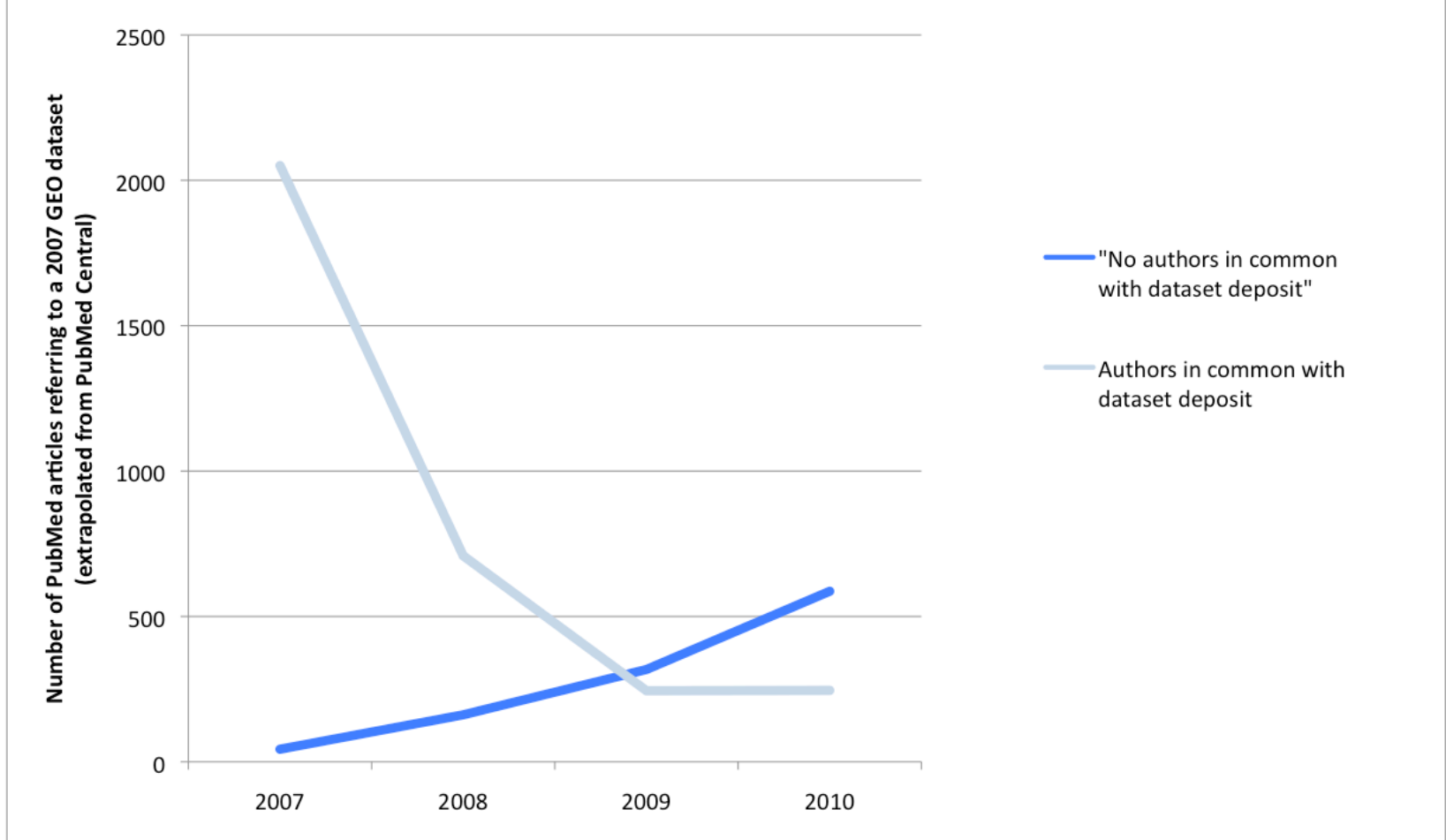
- **Canadian Institutes of Health Research (CIHR):**

"deposit bioinformatics, atomic, and molecular coordinate data into the appropriate public database (e.g. gene sequences deposited in GenBank) immediately upon publication of research results."

and

"retain original data sets for a minimum of five years (or longer if other policies apply)."

Use and reuse of 2007 GEO data attributed by accession reference, extrapolated to all of PubMed



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



PKP

PUBLIC
KNOWLEDGE
PROJECT



Project Proposal

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]



Dataverse Network



Share, Cite, Reuse, Archive Research Data
Scientific data for reproducible research

Harvard Dataverse Network

POWERED BY THE **Dataverse Network** PROJECT v. 3.4

[Search](#) [i](#) [Comment](#) [Create Account](#) [Log In](#)

[Advanced Search](#) [Tips](#)

The Harvard Dataverse Network is open to all scientific data from all disciplines worldwide. It includes the world's largest collection of social science research data. If you would like to upload your research data, first create a dataverse and then create a study. If you already have a dataverse, log in to add new studies. [Learn more about the Dataverse Network.](#)

Dataverses

524 Dataverses

i A **Dataverse** is a container for its owner.

RECENTLY RELEASED DATAVERSES

Mitts, Joshua

Damico, Anthony

[Create Dataverse](#)

Studies

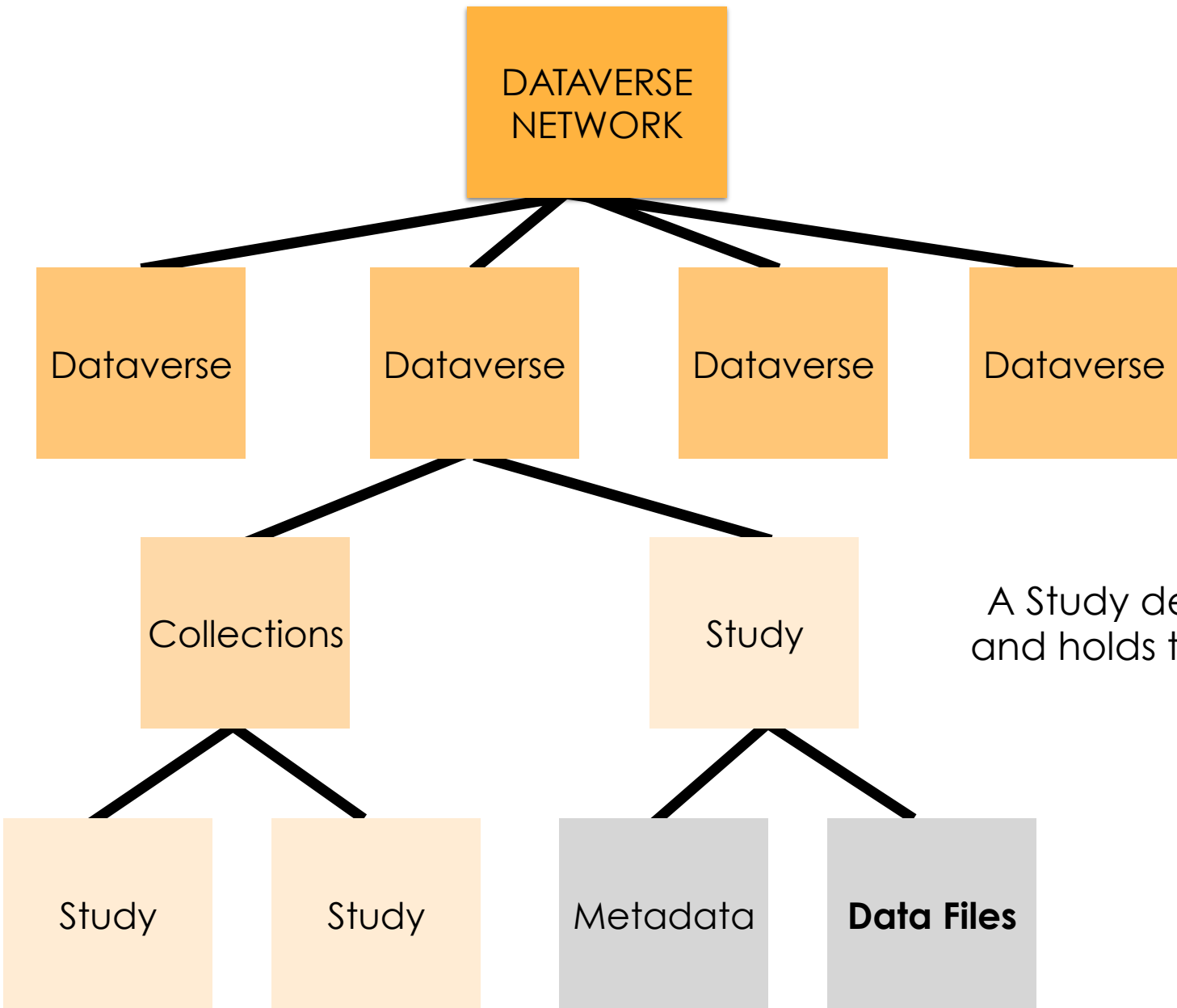
51,963 Studies, **722,906** Files, **757,550** Downloads

... includes cataloging information,

... and the Development of Party

May 6, 2013

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...

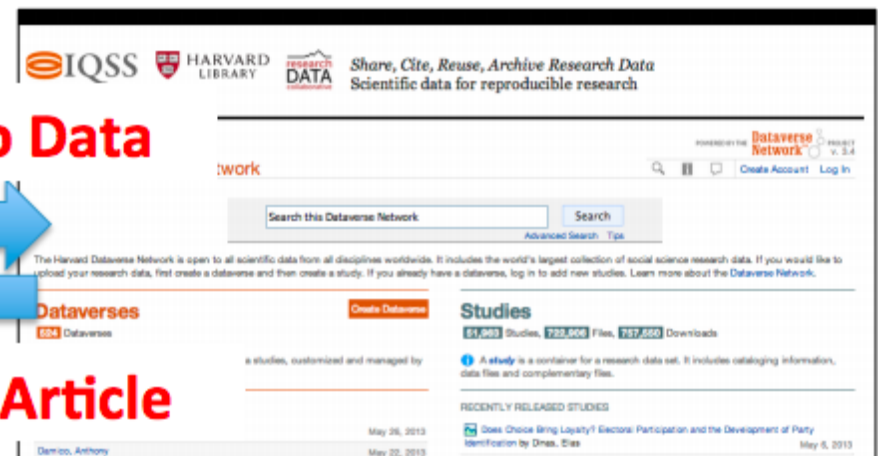
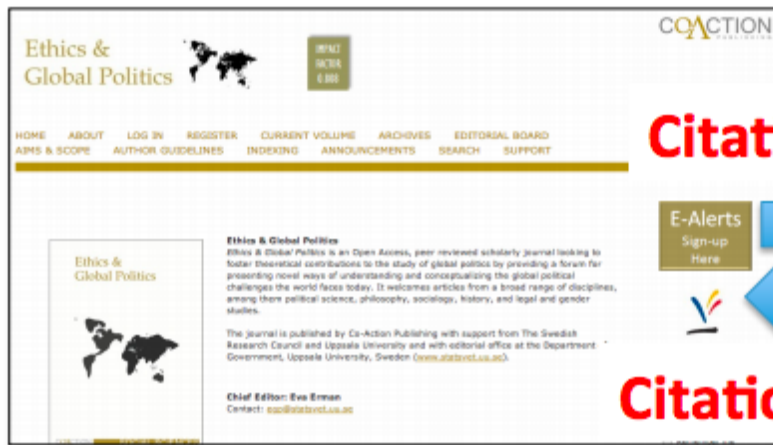
- ✓ Option for backups and replication of data in different locations (LOCKSS) so data is never lost.
- ✓ Re-format for long term accessibility so data never become obsolete.
- ✓ Extract Variable Metadata from data sets.
- ✓ Universal Metadata standards (DDI, Dublin Core).
- ✓ Inter-operability with 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 Journal

Harvard Dataverse Network



Citation to Data

Citation to Article

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,

Progress to-date

Compiled a list of potential journals (>400) that we can work with. Contacted a small sample of publishers to be our 1st round of pilot testers (50+ confirmed journals as of 06/22).

Publishers reviewed our [plugin workflow](#) and [mockups](#) to provide feedback before beginning development.



Mockups of OJS Plugin: Journal Setup

Dataverse Demo Journal

HOME ABOUT USER HOME SEARCH CURRENT ARCHIVES

Home > User > Journal Management > **Dataverse Plugin**

Dataverse Plugin

DATAVERSE SETTINGS

Configure Dataverse connection.

Dataverse URL *

Username *

Password *

* Denotes required field

[OPEN JOURNAL SYSTEMS](#)

[Journal Help](#)

SUBSCRIPTION

[My Subscriptions](#)

USER

You are logged in as...
admin

- [My Journals](#)
- [My Profile](#)
- [Log Out](#)

NOTIFICATIONS

- [View](#)
- [Manage](#)

JOURNAL CONTENT

Search

Mockup of Data Deposit (in OJS)

Home > User > Author > Submissions > **New Submission**

Step 4. Uploading Data Files

1. START 2. UPLOAD SUBMISSION 3. ENTER METADATA 4. **UPLOAD DATA FILES** 5. CONFIRMATION

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

ID	TITLE	ORIGINAL FILE NAME	DATE UPLOADED	ACTION
5	Untitled	19-282-1.csv	05-24	EDIT DELETE

Choose **data** type *

CSV (w/ SPSS card) ▾

Some file types can be subset and analyzed using the Dataverse Network analysis tools. Learn more about [subsettable 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.

ID	CITATION	PERSISTENT ID	DATE ADDED	ACTION
6	[citation]	hdl:00001/0002	05-24	EDIT DELETE

Citation *

Persistent Identifier *

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

[Journal Help](#)

SUBSCRIPTION

[My Subscriptions](#)

USER

You are logged in as...
jenauthor

- [My Journals](#)
- [My Profile](#)
- [Log Out](#)
- [Log Out as User](#)

NOTIFICATIONS

- [View \(4 new\)](#)
- [Manage](#)

JOURNAL CONTENT

Search

All ▾

Browse

- [By Issue](#)
- [By Author](#)
- [By Title](#)
- [Other Journals](#)

FONT SIZE

INFORMATION

- [For Readers](#)
- [For Authors](#)
- [For Librarians](#)

Step 4a. Upload Data Sets

1. START 2. UPLOAD SUBMISSION 3. ENTER METADATA 4. **UPLOAD DATA SETS** 5. CONFIRMATION

[<< Back to Supplementary Files](#)

Metadata

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

Title*	<input type="text" value="Untitled"/>
Author(s) of file	<input type="text"/>
Keywords	<input type="text"/>
Brief description	<div style="border: 1px solid #ccc; height: 150px; width: 100%;"></div>
Category	<input type="text" value="Data Files"/>
Date	<input type="text" value="2012-12-11"/> YYYY-MM-DD Date when data was collected or instrument created.

Metadata fields will be selected ahead of time by journal admin.

Supplementary File

File name	3-18-1-SP.pdf
Original file name	19-282-1-PB.pdf
File size	248KB
Date uploaded	2012-12-11 02:09 PM
<input type="checkbox"/> Present file to reviewers (without metadata), as it will not compromise blind review.	

Replace file Use Save to upload file.

* Denotes required field

Mockup of Published Article + Link to Data

Dataverse Demo Journal








[HOME](#) [ABOUT](#) [USER HOME](#) [SEARCH](#) [CURRENT](#) [ARCHIVES](#)

Home > Vol 1, No 1 (2012) > **Klockmo**

Moving toward a recovery-oriented approach in the Swedish mental health system

Carolina Klockmo

ARTICLE TOOLS

-  [Print this article](#)
-  [Indexing metadata](#)
-  [How to cite item](#)
-  [Review policy](#)
-  [Email this article](#)
-  [Email the author](#)
-  [Data Tools](#)

Reading Tools

journals.sfu.ca/dataverse/index.php/demo/rt/suppFiles/1/0

demo Reading Tools

Data

Moving toward a recovery-oriented approach in the Swedish mental health system

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", hdl:99999/00000
Full Record Available In:	COACTION DATAVERSE NETWORK
Data Tools:	INDEXING METADATA ANALYZE DOWNLOAD

[Close](#)

USER

You are logged in as...
pkpadmin

- [My Journals](#)
- [My Profile](#)
- [Log Out](#)

NOTIFICATIONS

- [View \(10 new\)](#)
- [Manage](#)

JOURNAL CONTENT

Search

All

Browse

- [By Issue](#)
- [By Author](#)
- [By Title](#)
- [Other Journals](#)

Data in the Dataverse

Simon Fraser University Dataverse Network >

POWERED BY THE **Dataverse Network** PROJECT v. 3.2

Co-Action Dataverse

🔍 ⓘ 💬 ⚙️ [Eleni Castro](#) [Log Out](#)

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 11:32:18 PST 2012

- [Edit Cataloging Information](#)
- [Add File\(s\)](#)
- [Edit Study Version Notes](#)
- [Release](#)
- [Permissions](#)
- [Create Study Template](#)
- [Delete Draft Version](#)

CATALOGING INFORMATION

[Data & Analysis](#)

[Comments \(0\)](#)

[Versions](#)

i 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

Next Steps

1. Complete pre-release version of plugin+API ([SWORD2-compliant](#)) (Fall 2013).
2. Additional journals (so far 50+) will test + provide feedback through a survey (Late 2013).
3. Provide best practices for data review/sharing policies and data citation (Late 2013).
4. Test and release OJS plugin + updated version of Dataverse Network (Spring 2014).
5. Make code & docs available for everyone.

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.

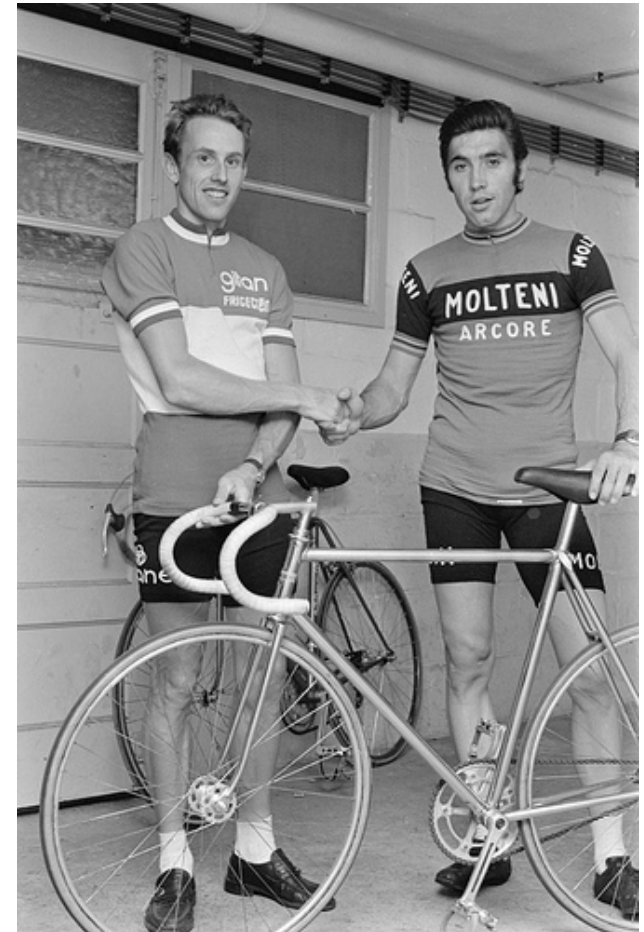


Photo: [Flickr Commons](#)

Thank you!



Project Website

<http://projects.iq.harvard.edu/ojs-dvn>

References

[Crosas, M., The Dataverse Network™: An Open-Source Application for Sharing, Discovering and Preserving Data, D-lib Magazine 17\(1/2\). 2011.](#)

[King, G. "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing." Sociological Methods and Research, 32\(2\), 173–199. 2007.](#)

[Willinsky, J.. Open Journal Systems: An example of open source software for journal management and publishing. Library Hi-Tech 23 \(4\), 504-519. 2005.](#)