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

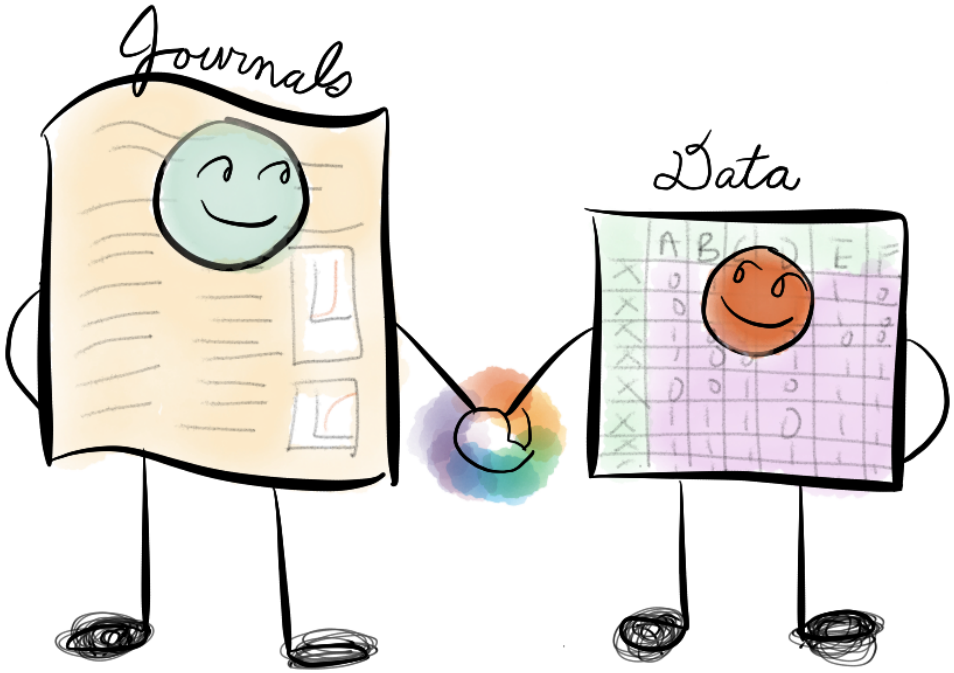
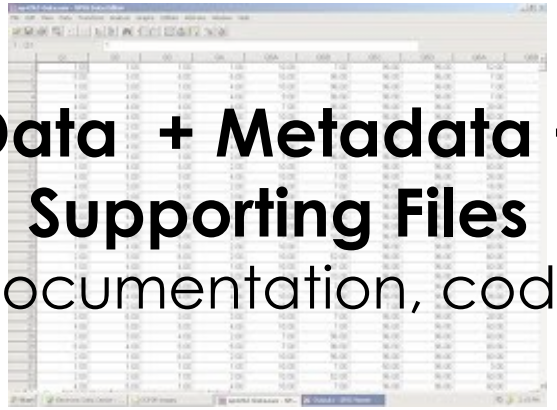


Photo: Jean Liu

Why Connect Published Work to Data

+



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

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):**

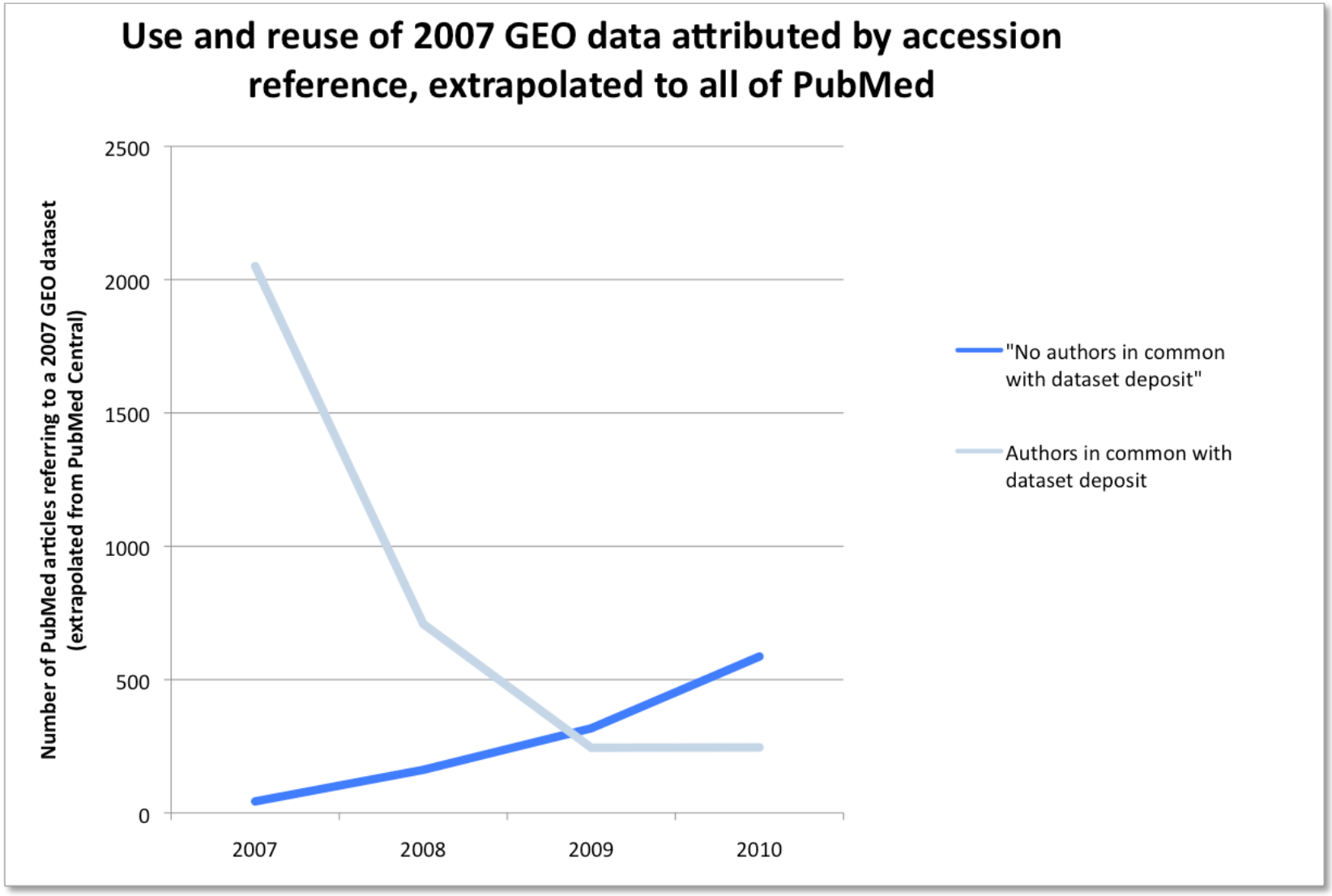
  "All research data collected with the use of SSHRC funds must be preserved and made available for use by others within a reasonable period of time. SSHRC considers "a reasonable period" to be within two years of the completion of the research project for which the data was collected.“

- **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)."



Source: Heather Piwowar http://researchremix.wordpress.com/2011/02/18/early_results/

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

The Dataverse Network™ 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**.

Photo: Flickr Commons

Integrating Open Source Systems

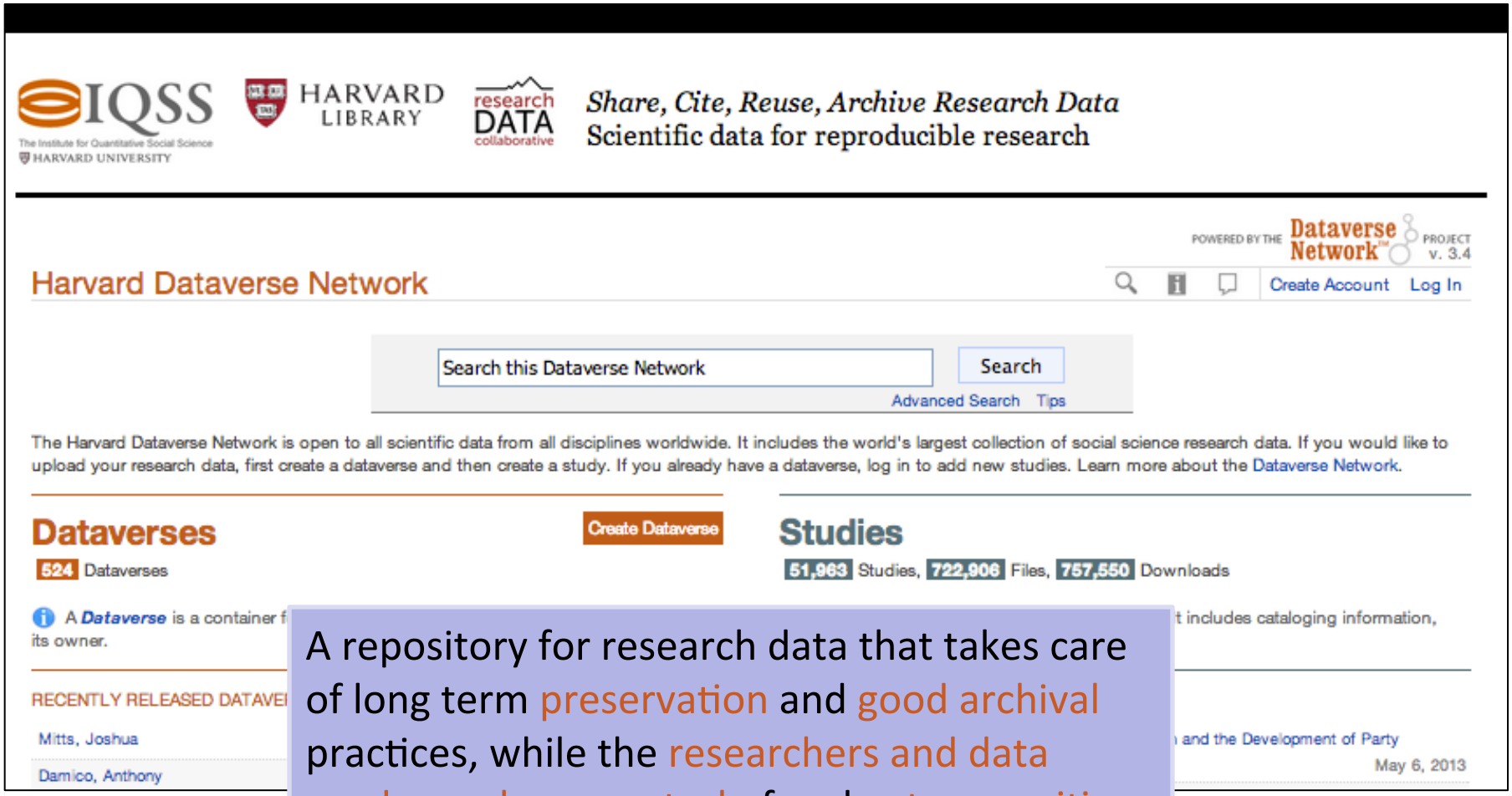
We plan to do this by integrating two well-established open-source systems:

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

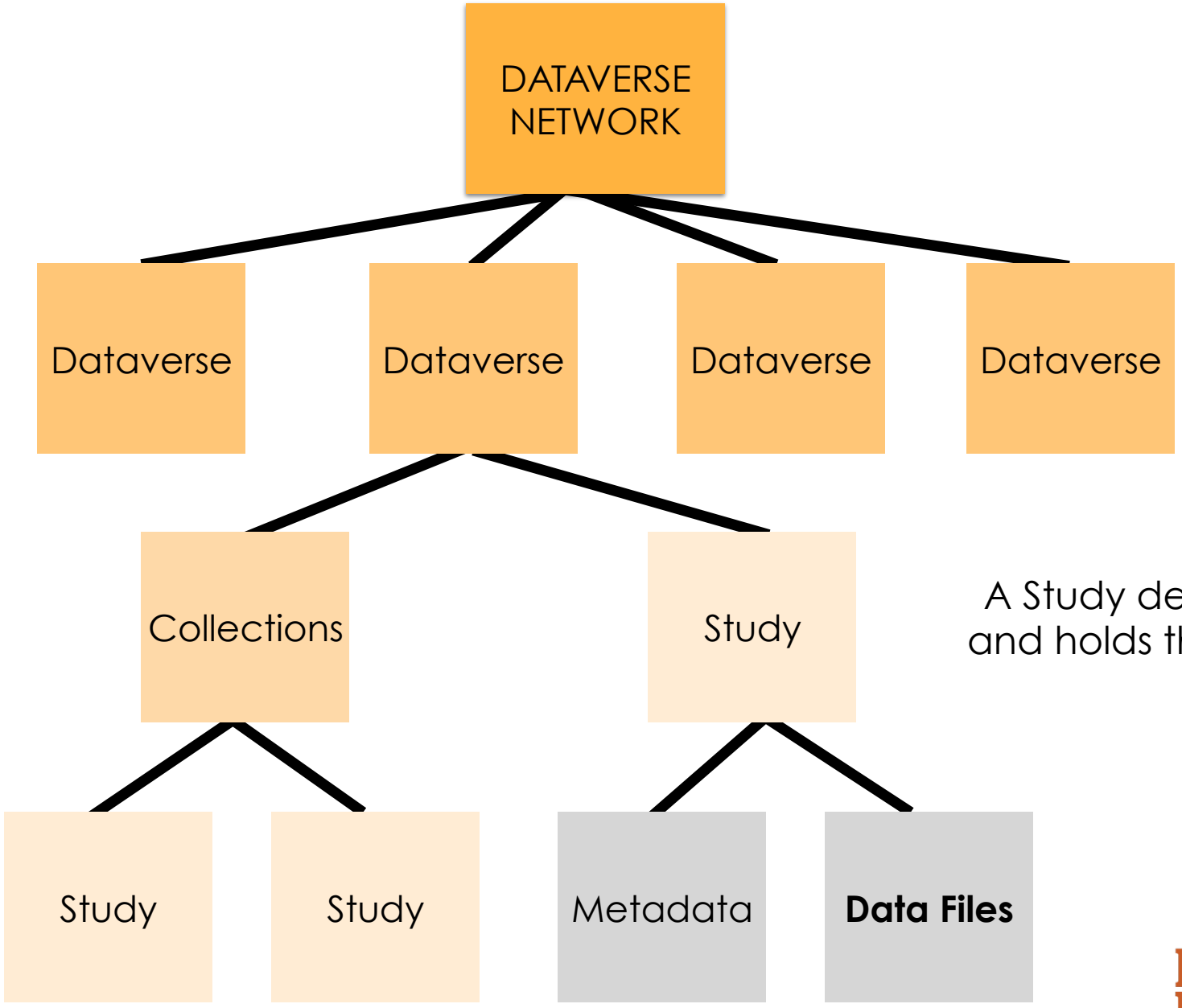
2. Dataverse Network
   [King 2007; Crosas 2011]

Dataverse Network

The Dataverse Network Project



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

The Dataverse Network™ Project

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.

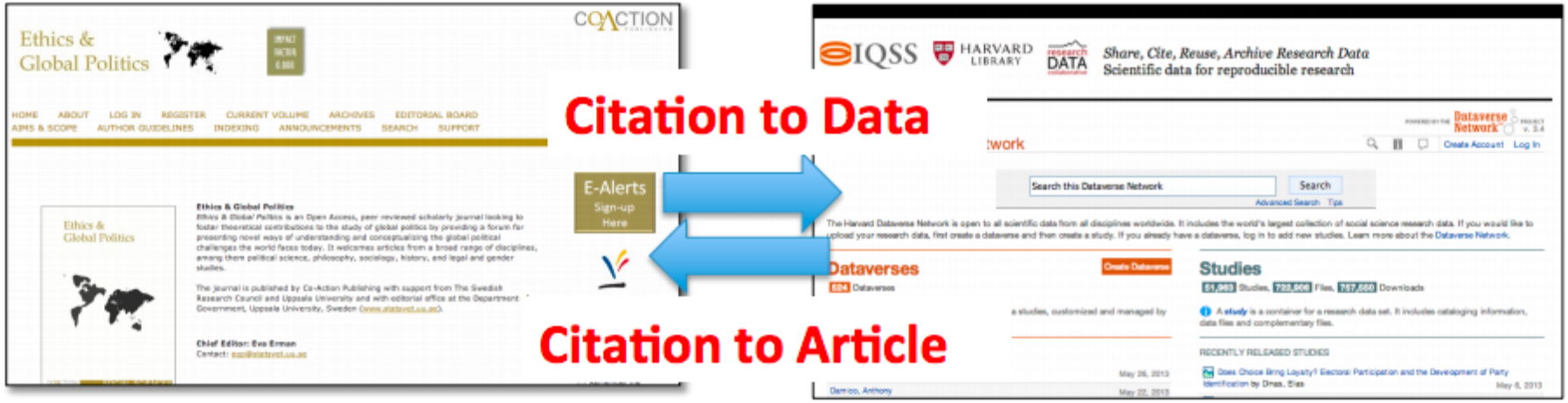
The Dataverse takes care of the archival infrastructure ("plumbing") for you!

The Dataverse Network™ Project

OK, so what is the integration going to do?

OJS Journal

Harvard Dataverse Network



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

Which Workflow?

Slide acknowledgement: Merce Crosas

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.

Scholars Portal
A Service of the Ontario Council of University Libraries

ubiquity press
open scholarship

COACTION PUBLISHING

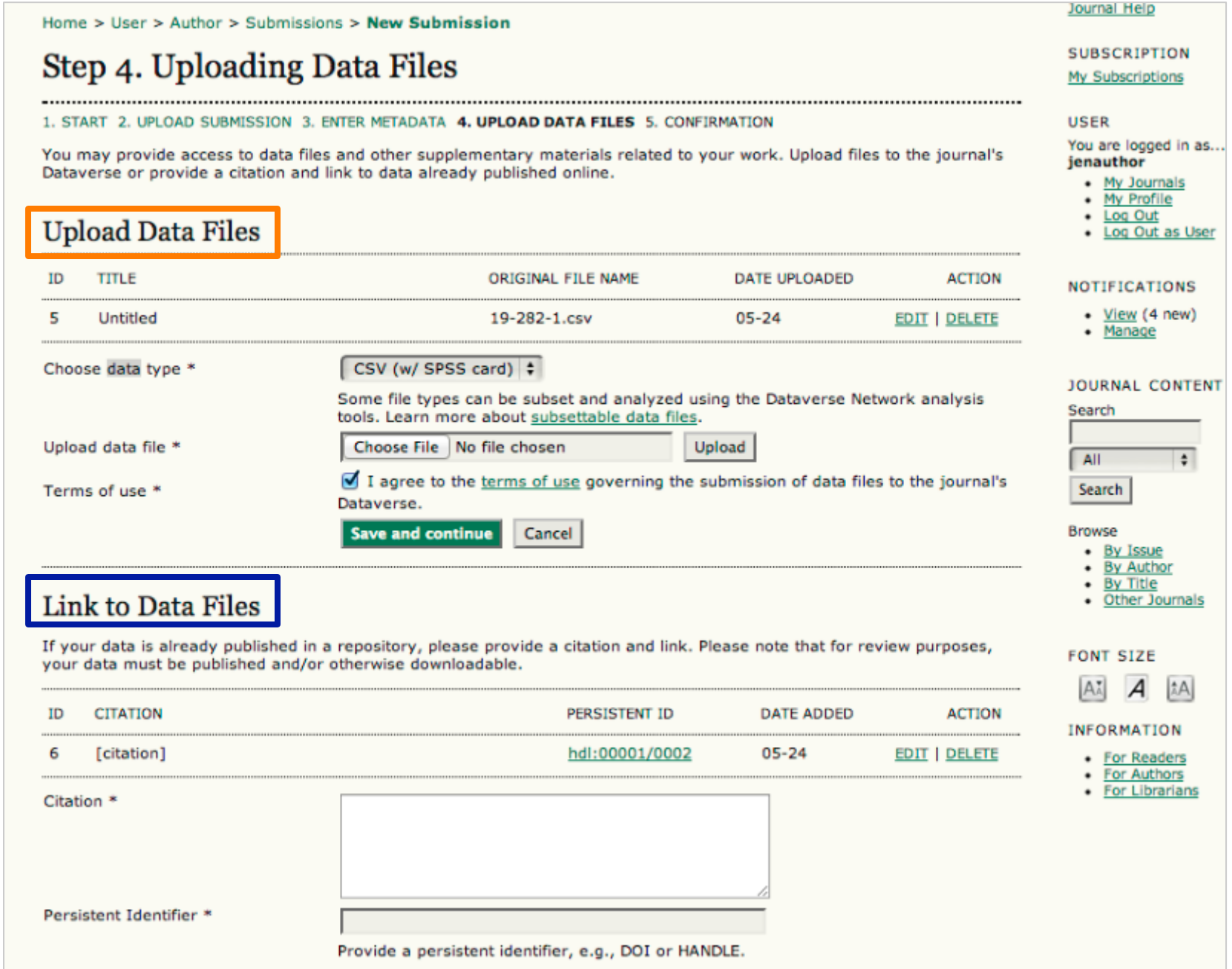
UNIVERSITY OF PITTSBURGH
ULS
UNIVERSITY LIBRARY SYSTEM

UNM UNIVERSITY LIBRARIES

International Journal of Education Policy & Leadership

Mockups of OJS Plugin: Journal Setup

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
Search

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* Untitled

Author(s) of file

Keywords

Brief description

Category Data Files

Date 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

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

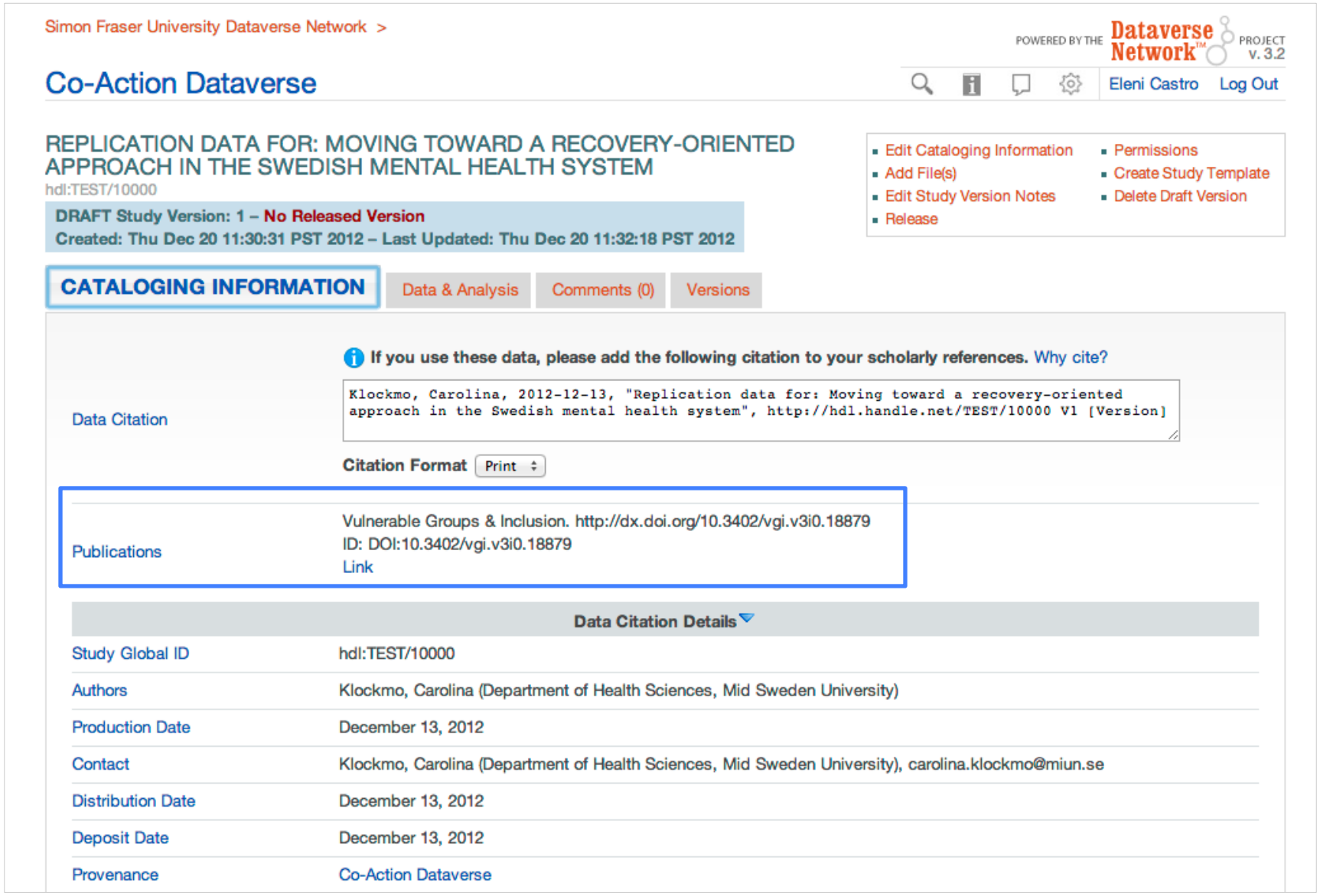
Replace file Browse... Use Save to upload file.

Save and continue Cancel

* Denotes required field

Mockup of Published Article + Link to Data

Data in the 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.

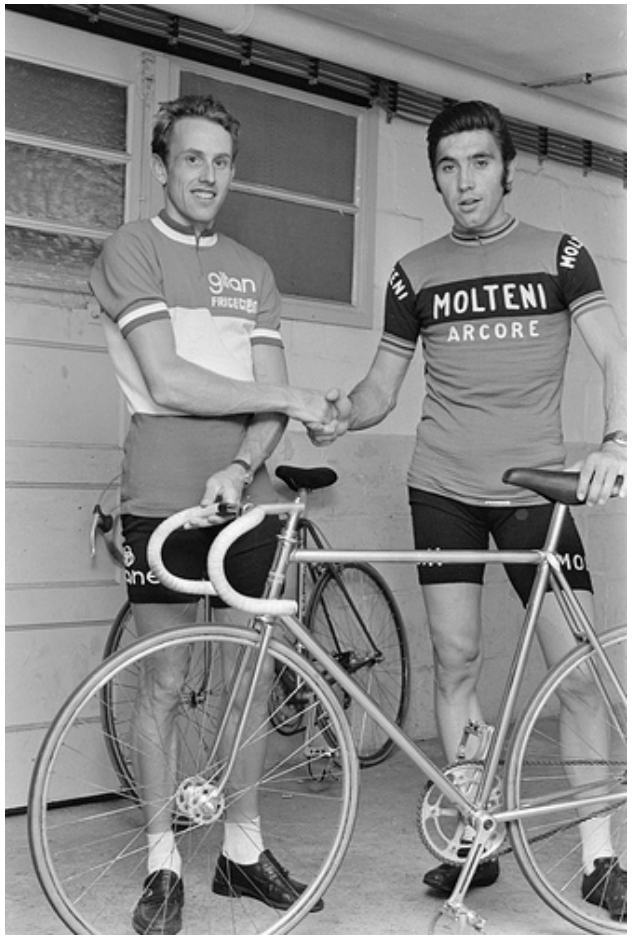


Photo: Flickr Commons

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.

Photo: Flickr Commons