

Building a Bridge Between Journal Articles & Research Data

IDCC 2014 - Feb 26, 2014

Eleni Castro

Institute for Quantitative
Social Science (IQSS)
Harvard University



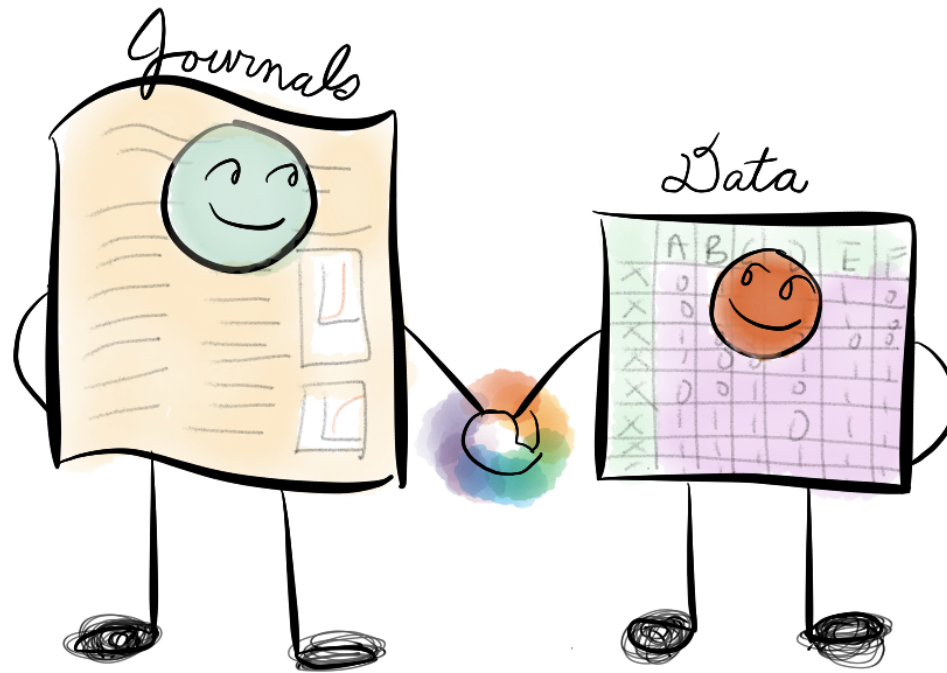
@thedataorg

Alex Garnett

Public Knowledge Project
Simon Fraser University

@axfelix

Motivation



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

A team was assembled...



2 year Sloan Foundation grant (2012-2014):

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 **journals** in addition to researchers and data managers.

What?: Enable journals to seamlessly manage the submission, review, and publication of data associated with published articles.

How?: Build the needed **technology** and create **awareness / incentives** regarding the importance of data sharing and preservation.

Why?: Help **increase** the **replicability** and **reusability** of research data.

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. 3.6.2

Search, Home, Info, Chat, Settings, Eleni Castro, Log Out

Search this Dataverse Network

Search

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. Learn more about the Dataverse Network.

Dataverses

625 Dataverses

A **Dataverse** is a container for research data studies, its owner.

RECENTLY RELEASED DATAVERSES

Jesse C. Johnson	Feb 19, 2014
Elections Project	Feb 4, 2014
Blanca de Lizaur	Feb 3, 2014

Create Dataverse

Studies

Create Study + Upload Data

Studies, 945,810 Downloads

Note: Major New Version 4.0 Coming Soon (Spring 2014)

a research data set. It includes cataloging information, data

RECENTLY RELEASED STUDIES

Replication data for: Dynamic Hierarchical Factor Models by Emanuel Moench; Ng, Serena; Potter, Simon	Feb 20, 2014
Impact Lite Dataset by Silvestri, Silvia; Douxchamps, Sabine; Teufel, Nils; Quiros, Carlos F.; Singh, Dhiraj; Ndungu, Anthony; Mutie, lanetta; Ndiwa, Nicholas; Van Wijk, Mark; Rufino, Mariana; Herrero, Mario; Kiplimo, Jasper	Feb 20, 2014

A repository for research data that takes care of **long term preservation** and **good archival practices**, while researchers can **share, keep control of** and get **recognition** for their data.

Dataverse 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, Header (FITS) Metadata from datasets.
- ✓ Universal Metadata standards (DDI, DC, DataCite).
- ✓ Interoperability with other systems through standard protocols (such as OAI-PMH, APIs).
- ✓ Generates a DOI for permanent linking to datasets.

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

Open Journal Systems (OJS)

The screenshot shows the homepage of the Open Health Data journal. The header features the 'UPmetaJournals' logo, the journal title 'open health data', and the 'ubiquity press open access' logo. A navigation menu includes links for Home, About, Editorial Board, Contact, Repositories, FAQ, Articles, and Archive. On the left, there are options to log in, register, and follow on Twitter or via RSS, along with font size controls. The main content area contains a description of the journal, a cartoon illustration of a scientist and animals on a book, and sections for 'Most recent articles' and 'Most popular articles'. A sidebar on the right features an advertisement for 'eifl Open Access in developing countries'.

UPmetaJournals

open health data

ubiquity press
open access

Home About Editorial Board Contact Repositories FAQ Articles Archive

Log in | Register

Follow on Twitter

Follow via RSS

Tweets

UCL Open Access @UCLOAfunding

UCL researchers: check out our wonderful Open Access FAQs, they explain how to use central funds to pay for your oa:
<http://t.co/PIIOFIKRL1>

Open Health Data is a peer-reviewed journal publishing papers that describe openly-accessible datasets with high reuse potential. We work with trusted data repositories to ensure that associated datasets are professionally archived, preserved, and openly available.

Open Health Data was previously published as the *Journal of Open Public Health Data*. The title and scope of the journal were amended on 1 January 2014 to support a broader range of health research.

Most recent articles

Nagulapalli
Data from "Social determinants of unmet hospitalisation need...
04 Nov 2013

Most popular articles

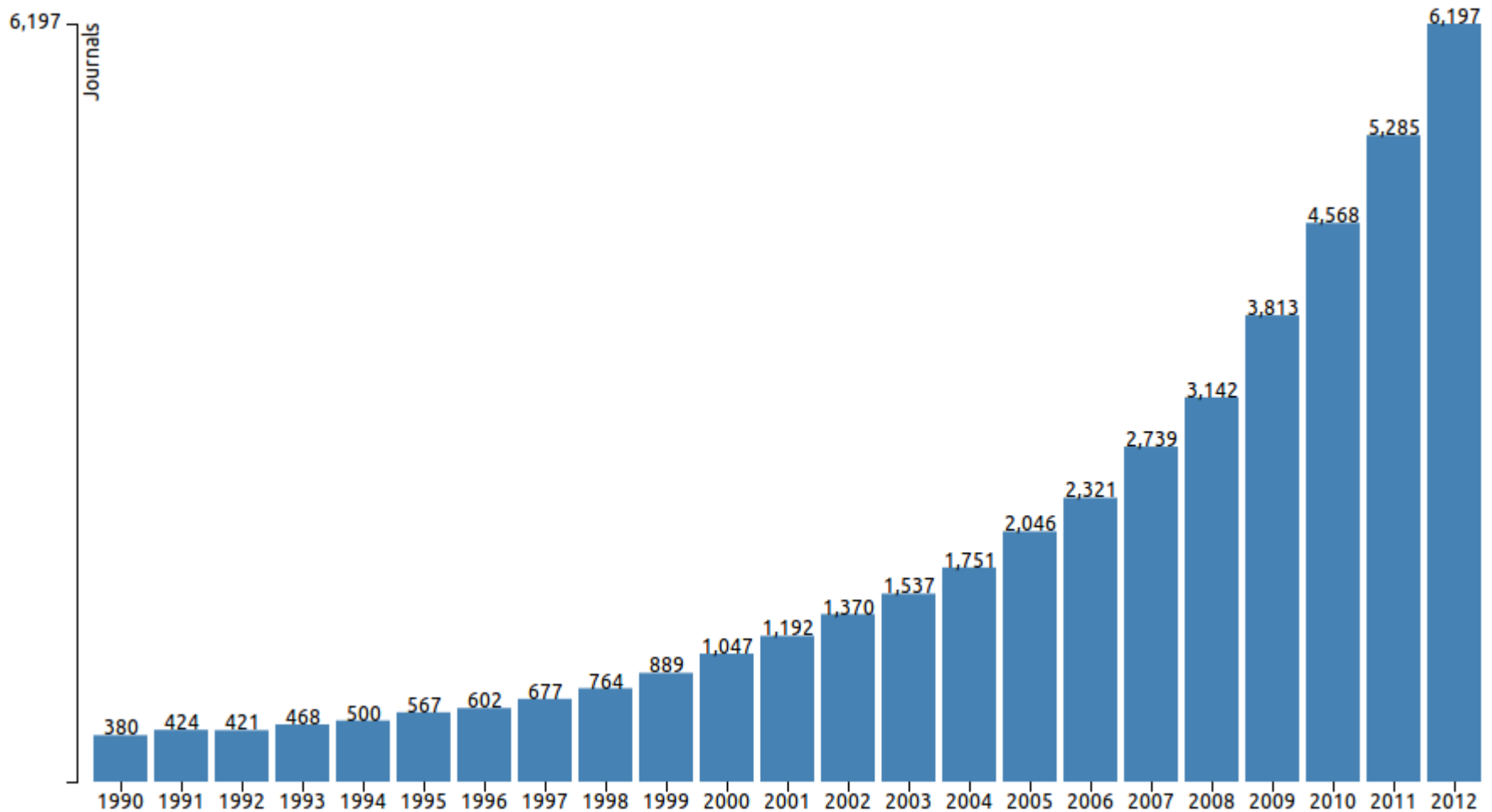
Alexander, Wint
Projected Population Proximity Indices (30km) for 2005, 2030 & 2050
11 Jul 2013

Ads

eifl
Open Access in developing countries

Open source journal management and publishing system created by PKP to expand & improve access to research.

Open Journal Systems (OJS)



Journals using OJS to publish ≥ 10 articles per year. Credit: Juan Alperin, PKP

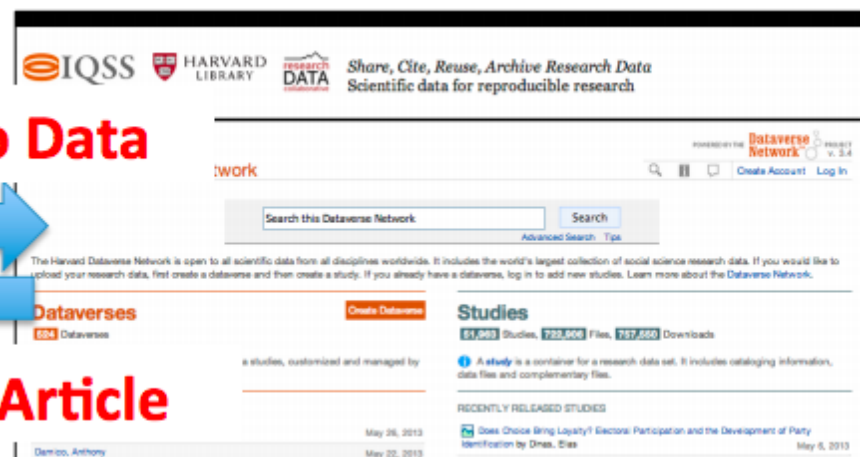
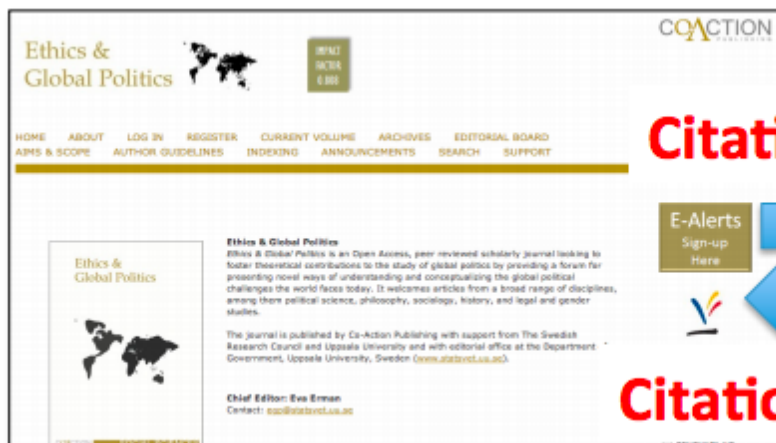
OJS provides...

- ✓ 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).
- ✓ Interoperability 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.

OK, so what is the integration going to do?

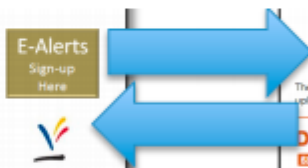
OJS Journal

Harvard Dataverse



Citation to Data

Citation to Article



OJS plugin for:
Data + supporting files + metadata
sent via API (SWORD-based)
to 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

1. Compiled a list of potential journals (~100) that we can work with.
2. Publishers reviewed our [plugin workflow](#) and [mockups](#) pre-development.
3. Completed and released first version of plugin+API.
4. Small sample of journals tested + provided feedback of integration via survey.
5. Some of our partners have begun production testing.
6. Working on improvements to UI and metadata customization for v2 release



Demo

<http://lib-hf1.lib.sfu.ca/ojs-2.4.3>

Survey Findings

- Overall **positive**: Most testers plan to implement plugin!
- OJS doesn't have the most eye-catching UI. Dataverse link & data citation must be prominent.
- **But users excited for an OJS 3 + Dataverse 4 overhaul!**
- Default dcTerms metadata defined in the SWORD spec is too limited for all of our users; the ability to customize fields is important.
- Plugin should not treat all supplemental files as Dataverse-worthy (e.g., authors' notes).



Next Steps (Spring 2014)

1. Publish best practices for data sharing policies.
2. Develop & release updated versions of plugin + Data Deposit API.
3. Continue to release all code and documentation :)
4. Expand on Collaboration.



Photo: [Flickr Commons](#)

Advantages to Integration

- Reference implementation so others can reuse and extend (extension of SWORD).
- Streamlining **authors'** article and data deposit process.
- Permanent **2-way linking** of the published article with its archived data.
- Increase **visibility/access**, and encourage data citation, replication and re-use.

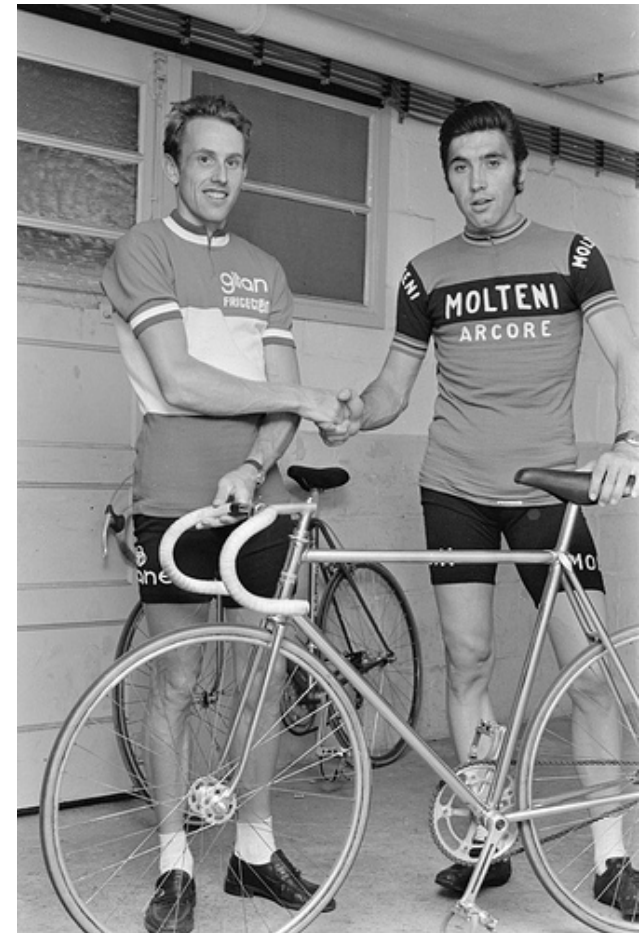


Photo: [Flickr Commons](#)

Thank you!

Contact

Alex Garnett - garnett@sfu.ca

Eleni Castro - ecastro@fas.harvard.edu

Project Website

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

References

- [Crosas M. A Data Sharing Story. *Journal of eScience Librarianship* 1\(3\), 173-179. 2013.](#)
- [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.](#)

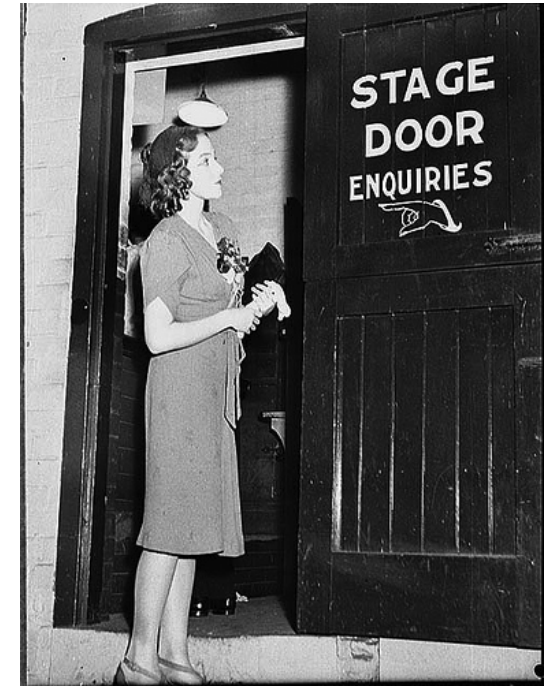


Photo: [Flickr Commons](#)