DATAVERSE FOR JOURNALS

Mercè Crosas, Ph.D.
Director of Data Science
IQSS, Harvard University
@mercecrosas

Society for Scholarly Publishing 37th Meeting, 28, May, 2015

About Dataverse

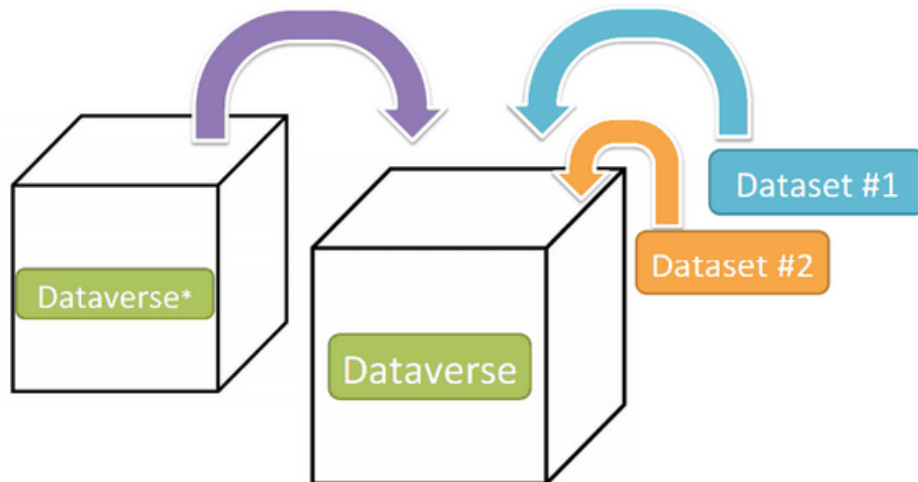| Science requires community access to data	Technology Solution →	Open source software for publishing, citing and archiving research data

The Dataverse Project

- ✓ Gives credit and control to data authors and distributors
- ✓ Follows best practices, standards for data management and archiving
- ✓ Dataverse development started in 2006 at Harvard's IQSS
- ✓ Widely used around the world, with a vibrant development and user community
- ✓ Helped instigate and is at the center of a cultural change toward open, reproducible, and transparent research

The New Dataverse

Dataverse 4.0 is a full rewrite that improves usability, defines a rigorous and standardized data publishing workflow, and leverages the latest technologies.
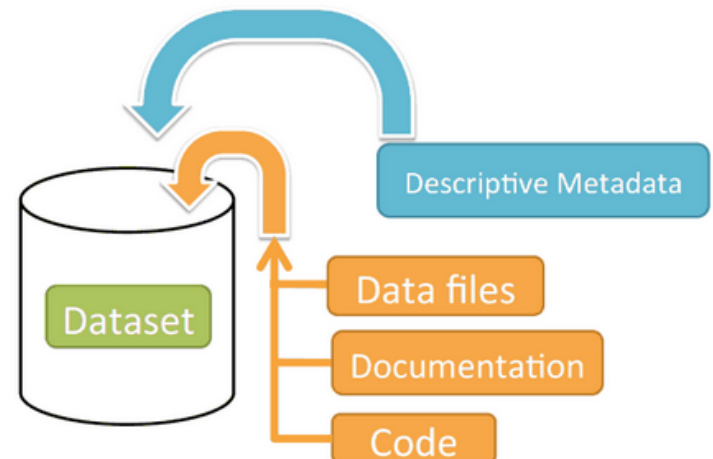
Schematic Diagram of a **Dataverse** in Dataverse 4.0

Dataverse*

Dataverse

Dataset #1

Dataset #2

Container for your **Datasets** and/or **Dataverses***

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

Schematic Diagram of a **Dataset** in Dataverse 4.0

Dataset

Descriptive Metadata

Data files

Documentation

Code

Container for your data, documentation, and code.

Software Features and Technology

Features

- ✓ Standard, persistent data citation
- ✓ Branding for each dataverse
- ✓ Standard, extensible metadata:
    - ✓ citation metadata
    - ✓ domain-specific metadata
    - ✓ file-level metadata
- ✓ Faceted search for all metadata
- ✓ Multiple levels of access control
- ✓ Multiple roles and permissions
- ✓ Re-formatting of tabular data files
- ✓ Extraction of file metadata
- ✓ APIs for search, deposit, access

Technology

- ✓ Uses open source solutions
- ✓ Leading-edge UI framework
    - ✓ PrimeFaces and Bootstrap
- ✓ Widely used, community driven enterprise software platform
    - ✓ Java EE and Glassfish
- ✓ Reliable, scalable search platform
    - ✓ Solr
- ✓ Web standard programmatic interfaces
    - ✓ RESTful APIs
- ✓ Standards for archiving and interoperability
    - ✓ OAI-PMH, LOCKSS

Dataverse Repositories Installations

Dataverse.org

1. DANS - Netherlands
2. Fudan University
3. OCUL Scholars Portal
4. Abacus
5. University of Alberta Libraries
6. Harvard Dataverse
7. Heidelberg University
8. ODUM Institute
9. UiT The Arctic University of Norway
10. UnBraL Fronteiras

@dataverseorg
1,480 Followers

IQSS
The Institute for Quantitative Social Science
HARVARD UNIVERSITY

@IQSS
www.iq.harvard.edu
datascience.iq.harvard.edu

Dataverse software installations through out the world serve as:
- public data repositories open to all research data (Harvard and ODUM Dataverse),
- or institutional research data repositories, with data from the institution only.

Harvard Dataverse

- Free and open to research data worldwide
- Containing:
  - > 1000 dataverses
  - > 58,000 datasets
  - > 270,000 files
  - > 1.3 million downloads
  - > 10,000 registered users
- Includes dataverses for:
  - individual researchers
  - research teams
  - journals
  - large institutions or organizations.
- Increase of about x30 deposits/month since last year

Harvard Dataverse (dataverse.harvard.edu)

Options for Journals and Journal Editors

- Researcher Dataverse:
  - Authors publish a dataset in their own Dataverse
  - Authors provide the data citation generated by Dataverse to journal:

Chiba, Daina; Metternich, Nils W; Ward, Michael D, 2015, "Replication data for: Every Story Has a Beginning, Middle, and an End (But Not Always in That Order): Predicting duration dynamics in a unified framework", http://dx.doi.org/10.7910/DVN/26972, Harvard Dataverse, V1

- Journal Dataverse:
  - Journal has a Dataverse where authors deposit their dataset
  - Journal reviews and publishes the data, when ready

- System Integration:
  - Journal management system integrated with Dataverse through the Dataverse deposit API.
  - Authors submit an article with the dataset in the Journal system
  - Dataset is automatically deposited in the Journal Dataverse

Workflow for Journal Dataverses

Author submits dataset	→	Journal reviews dataset	→	When Journal approves paper, it publishes dataset

Dataset is stored as *Draft* in Journal Dataverse

- Dataset gets CC0 license by default
- CC0 can be changed, if needed.
- Files can be restricted, if needed.
- Once published, dataset can't be unpublished (only deaccessioned)
- New versions can be submitted, reviewed and published.

Journal Dataverses Examples

American Journal of Political Science

ajps ▮ American Journal of Political Science

American Journal of Political Science (AJPS) Dataverse (Michigan State University)   ajps.org

Harvard Dataverse > **American Journal of Political Science (AJPS) Dataverse**

✉   ⤤

The American Journal of Political Science is committed to significant advances in knowledge and understanding of citizenship, governance, and politics, and to the public value of political science research.

| Search this dataverse... | 🔍 Find | Advanced Search | ➕ Add Data |

☑ ⚙ **Dataverses (0)**
☑ 📄 **Datasets (184)**
☐ 📊 Files (1,111)

Affiliation
Princeton University (10)
Stanford University (10)
Harvard University (9)
Columbia University (8)
New York University (8)

More...

1 to 10 of 184 results   ⇅ Sort ▾   «   < Previous   **1**   2   3   4   5   Next >   »

📄 **Replication data for: The Logic of Collective Inaction: Senatorial Delay in Executive Nominations**   📄
May 22, 2015

Ostrander, Ian, 2015, "Replication data for: The Logic of Collective Inaction: Senatorial Delay in Executive Nominations", http://dx.doi.org/10.7910/DVN/IYZWUD, Harvard Dataverse, V1

While most executive nominees are successfully confirmed, this success masks wide variation in how long it takes the Senate to decide. Delay of critical nominees influences the character and effective...

📄 **Replication data for: An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions**   📄
May 14, 2015

Economics: The Open-Access, Open-Assessment E-Journal

Journal of Information Technology & Politics

Dataverse

🔍 About   Guides ▾   Support   Sign Up   Log In

JITP Journal of Information Technology & Politics Dataverse (University of Massachusetts Amherst)   Home Page

Harvard Dataverse > **Journal of Information Technology & Politics Dataverse**

This is the Journal of Information Technology and Politics Dataverse for posting of replication datasets. Please refer to the standard for replication dataset citation guidance.

Search this dataverse...   🔍 Find   Advanced Search   ➕ Add Data

☑ 👥 **Dataverses (0)**
☑ 📄 **Datasets (11)**
☐ 🗎 **Files (50)**

Affiliation
Johns Hopkins University (2)
Drexel University (1)
National University of Singapore (1)
University of Athabasca (1)
University of Maryland (1)

More...

Publication Date
2012 (6)
2010 (3)
2008 (1)
2009 (1)

1 to 10 of 11 results   ↕ Sort ▾   «   < Previous   **1**   2   Next >   »

Replication data for: Wireless Protesters Move Around: Informational and Coordinative Use of Information and Communication Technologies for Protest Politics
Feb 6, 2012

K. Hazel Kwon, 2012, "Replication data for: Wireless Protesters Move Around: Informational and Coordinative Use of Information and Communication Technologies for Protest Politics", http://hdl.handle.net/1902.1/15455, Harvard Dataverse, V1

This study explored information and communication technology (ICT) uses for protest politics, focusing on the case of a 2008 protest in Korea. Based on a survey of citizen activists (N = 322), it exam...

Replication data for: Digital Media in the 2008 Canadian Election
Feb 6, 2012

Peter John Chen; Peter Jay Smith, 2012, "Replication data for: Digital Media in the 2008 Canadian Election", http://hdl.handle.net/1902.1/15407, Harvard Dataverse, V1

This paper examines the use and impact of digital media in the 2008 Canadian federal election. It examines the extent to which different digital media were employed by candidates and parties. The rese...

Journal of Open Psychology Data (Ubiquity Press)

Dataverse
Q   About   Guides ▾   Support   Sign Up   Log In

UPmetajournals   Journal of open psychology data   ]u[ubiquity press open access

Journal of Open Psychology Data (JOPD) Dataverse (Ubiquity Press)   Home Page

Harvard Dataverse > Ubiquity Press Dataverse > **Journal of Open Psychology Data (JOPD) Dataverse**

✉   ↪

Datasets from data papers published in the Journal of Open Psychology Data (JOPD) .

Search this dataverse...   🔍 Find   Advanced Search   ➕ Add Data

☑ 👤 **Dataverses (0)**
☑ 📄 **Datasets (4)**
☐ 📊 Files (12)

Affiliation
Bar Ilan University (1)
Bristol Speech and Language
Therapy Research Unit (1)
Department of Psychology,
University of Tartu, Estonia (1)
Estonian National Defence College,
Estonia (1)

1 to 4 of 4 results   ↕ Sort ▾

Data from paper "Demographic influences on disgust: Evidence from a heterogeneous sample."
Jan 13, 2015

Berger Uri; Anaki David, 2015, "Data from paper "Demographic influences on disgust: Evidence from a heterogeneous sample."",
http://dx.doi.org/10.7910/DVN/27285, Harvard Dataverse, V1

The dataset contains the data of large heterogeneous sample of participant who filled out the Hebrew version of the Disgust Scale Revised (DS_R).
Participants were chosen from the general population i...

Normative data for email writing
Dec 18, 2014

Data In Brief (Elsevier)

DATA IN BRIEF Data in Brief (DiB) Dataverse (Elsevier) Home

Harvard Dataverse > **Data in Brief (DiB) Dataverse**

✉ ⮺

Datasets for papers published in the Data in Brief (DiB) journal. For author guidelines on how to upload your data to DiB please visit the DiB homepage for step-by-step instructions.

| Search this dataverse... | 🔍 Find | **Advanced Search** | ＋ Add Data |

☑ ⑧ **Dataverses (0)**   1 to 4 of 4 results   ⇅ Sort ▾

☑ 📄 **Datasets (4)**

☐ 📊 Files (25)

High resolution 3D laboratory x-ray tomography data of femora from young, 1-14 day old C57BL/6 mice
Apr 6, 2015

Affiliation

Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada (2)

Bortel, Emely L; Duda, Georg N; Mundlos, Stefan; Willie, Bettina M; Fratzl, Peter; Zaslansky, Paul, 2015, "High resolution 3D laboratory x-ray tomography data of femora from young, 1-14 day old C57BL/6 mice", http://dx.doi.org/10.7910/DVN/29628, Harvard Dataverse, V1

This data article contains high resolution (1.2 μm effective pixel size) lab-based micro-computed tomography (μCT) reconstructed volume data of young C57BL/6 mouse femur bone midshafts. This data form...

Mass Spectrometry and Proteomics Group, National Research Council of Canada, Halifax, Nova Scotia, Canada (2)

Julius Wolff Institute, Charité-Universitätsmedizin Berlin, Campus

Putative N-Terminal Semi-Tryptic Peptides in mitochondrial proteome of a unicellular eukaryote (Acanthamoeba castellanii, supergroup Amoebozoa)
Aug 7, 2014

Gawryluk, Ryan M.R.; Chisholm, Kenneth A.; Pinto, Devanand M.; Gray, Michael W., 2014, "Putative N-Terminal Semi-Tryptic Peptides in mitochondrial proteome of a unicellular eukaryote (Acanthamoeba castellanii, supergroup Amoebozoa)", http://dx.doi.org/10.7910/DVN/26980, Harvard Dataverse, V1

Putative N-Terminal Semi-Tryptic Peptides in mitochondrial proteome of a Acanthamoeba castellanii

Dataverse Growing Community

- Software and project information: http://dataverse.org
- Harvard Dataverse repository: https://dataverse.harvard.edu
- Dataverse in Twitter: @dataverseorg
- Dataverse community mailing list: https://groups.google.com/forum/#!forum/dataverse-community
- Dataverse support: support@dataverse.org
- Dataverse issue tracking: https://github.com/IQSS/dataverse
- First International Dataverse Community Meeting:
  - June, 9, 10, 11 at Harvard University
  - http://projects.iq.harvard.edu/dataverse-community-meeting

Other questions:
@mercecrosas