

# Organize, Access, Share, and Archive Geospatial Data:



## A Hands-on Case for Dataverse

Larry Laliberte  
GIS Librarian  
University of Alberta

# ACCOLEDS

Vancouver, British Columbia || November 25-27, 2014



<http://bit.ly/1xGgriK>

Based on Slides Crafted with Chuck Humphrey as Part Of Research Data Management Week March 2014

# Outline

- Researcher data management needs
- Features that address these needs
- How to use Dataverse
- Geospatial Data & Dataverse

<https://thedata.harvard.edu/dvn/>



# University of Alberta RDMS



[Find a Person](#) | [ONEcard](#) | [Bear Tracks](#) | [Maps](#) | [Email & Apps](#) | [eClass](#) | [UofA](#)

## LIBRARIES

**Ask Us**

[chat](#) - [text](#) - [email](#) - [phone](#)

[Version française](#)

[DATABASES](#)

[JOURNALS](#)

[SUBJECTS](#)

[LIBRARIES](#)

[MY ACCOUNT](#)

[SERVICES](#)

[HOURS](#)

## RESEARCH DATA MANAGEMENT

### Organize my data

How to prepare a data management plan & protect your data

### Share my data

Why you would share, data sharing tools, metadata & how to get credit

### Access existing data

How to locate data and access data tools, terms & citations

### Archive my data

How to find a repository, prepare your data & submit your data

[DMP BUILDER](#)



[DATAVERSE](#)



#### Tweets



**figshare**  
@figshare

22h

Are We Heading for a Librarian vs Publisher Data Sharing Space Race?  
[digital-science.com/blog/guest/are...](https://digital-science.com/blog/guest/are...)  
Expand



**Dryad**  
@datadryad

20 Nov

Over 1K research data repositories are now indexed by @re3data  
[re3data.org/2014/11/over-1...](https://re3data.org/2014/11/over-1...)  
Expand



**CERN**  
@CERN

20 Nov

CERN launches Open Data Portal to make public the data of LHC

Good data management is informed by funder and institutional policy and encompasses the full data lifecycle including collection, analysis, dissemination and potential reuse.

# The Dataverse Network



A Web Application for Publishing, Citing,  
Analyzing and Preserving Research Data

[ABOUT](#)[GETTING STARTED](#)[BEST PRACTICES](#)[SOFTWARE](#)[APPS](#)[GUIDES](#)

[HOME](#) / [SOFTWARE](#) /

## DATVERSE NETWORKS AROUND THE WORLD

Listed below are featured Dataverse Networks from around the world. If you would like to be listed, please [contact us](#).

- [Dutch Universities \(DANS: Utrecht, Tilburg, etc\)](#)
- [European Union Democracy Observatory \(Beta version\)](#)
- [Fudan University](#)
- [Heidelberg University \(HeiDATA\)](#)
- [Odum Institute](#)
- [Ontario Council of University Libraries \(OCUL\)](#)
- [Thailand](#)
- [UnBraL Fronteiras \(Brazil\)](#)
- [University of Alberta Libraries](#)
- [University of Copenhagen - DataBox](#)
- [UiT The Arctic University of Norway](#)

[Printer-friendly version](#)

### Software

[Latest Releases](#)[Upcoming Releases](#)[Features](#)[Architecture](#)[Software License & Social Contract](#)[Dataverse Networks Around the World](#)

<http://thedata.org/book/dataverse-networks-around-world>

# Built for Statistical Data

- SPSS
- Stata
- RData
- csv
- tab
- zip files
- Other



**Dataverse 4.0 - .shp**

# Not Long Term Preservation

“ARCHIVE YOUR DATA OR THEY DIE WITH



**Life is a Transitory Illusion - Your data need not be**

# Staging for preservation

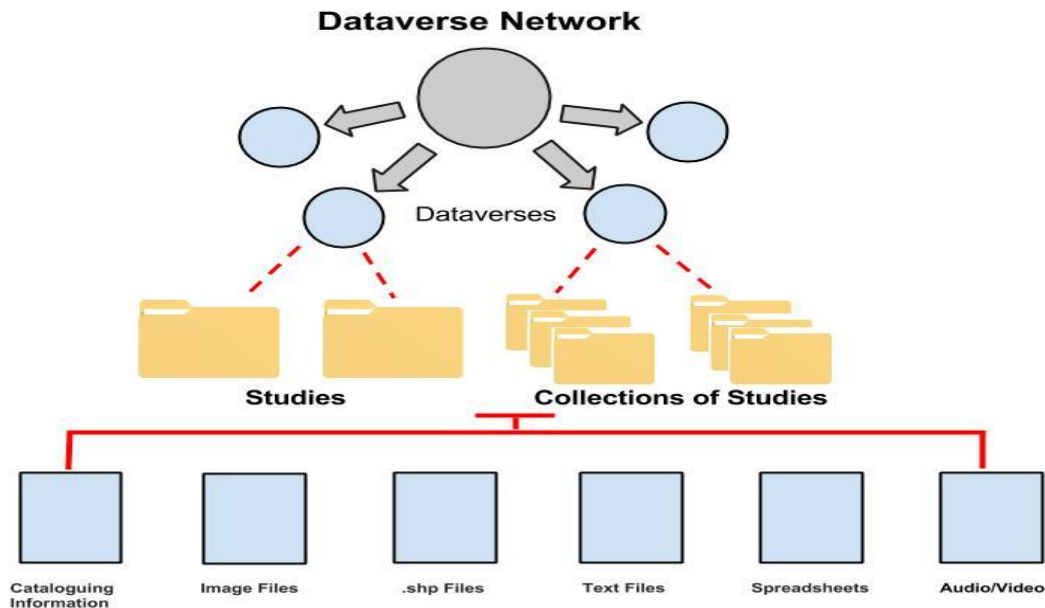
- Dataverse does not provide preservation services but can be used to organize files for ingest into a preservation service.
- A development project is underway in Canada to create an ingest link between Dataverse and [Archivematica](#).

# Meeting researchers' needs

- A way to **share data** with one, a few, or many other researchers that does not involve the use of email or Dropbox
- A way to enter **project-level metadata** that can be used repeatedly with other applications
- A way to mint **DOIs** for data files that can be used in publications
- A way to manage multiple **versions** of data files, including keeping track of changes made to the data
- A service that helps organize data to submit for **preservation** processing



# Place to store your data for sharing



- A software framework to build data repositories
- The data are organized in studies, which contain metadata, data files and complementary files. The studies can in turn be grouped into collections.
- Researchers can **share**, keep **control** of and get **recognition** for their data through an easy to use web interface

# Controlled access to data

- Can release data to the world (Open Data)
- Can release data to a set of University ID's
- Can release data to a predefined group
- Can restrict access to the data

## User Permission Settings

There are several roles defined for users of a Dataverse Network installation:

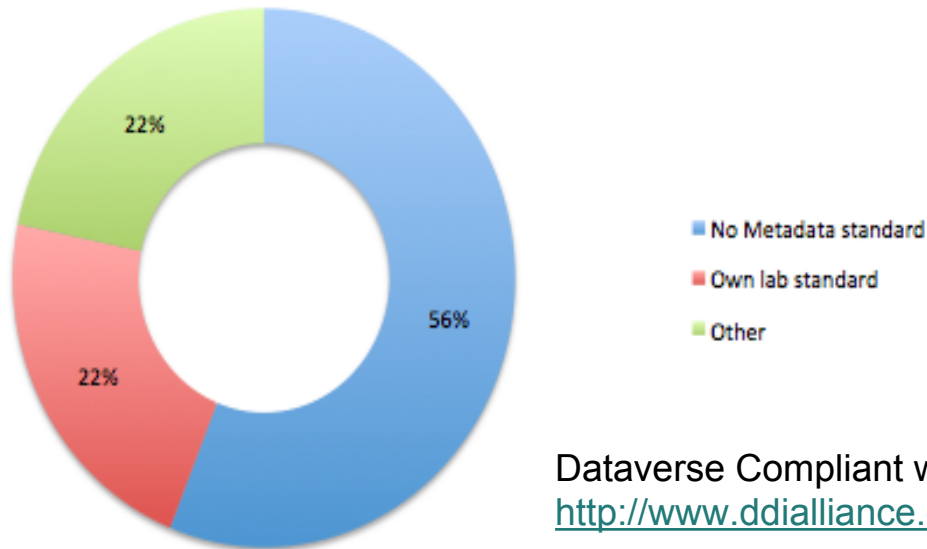
- Data Users - Download and analyze all types of data
- Contributors - Distribute data and receive recognition and citations to it
- Curators - Summarize related data, organize data, or manage multiple sets of data
- Administrators - Set up and manage contributions to your dataverse, manage the appearance of your dataverse, organize your dataverse collections

# Metadata

- Dataverse allows researchers to describe their studies using an international metadata standard.
- This metadata allows the files in a study to be discovered by search engines.

# Use of metadata

Use of a Metadata Standard

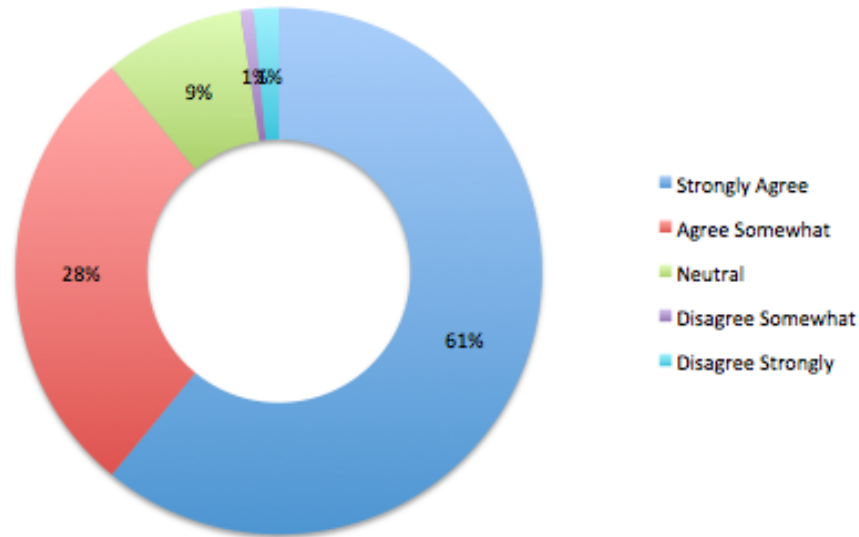


Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) *Data Sharing by Scientists: Practices and Perceptions*. **PLoS ONE** 6(6): e21101

Dataverse Compliant with the **DDI standard** (schema version 2)  
<http://www.ddialliance.org/>

# Data citation

**Important My Data Are Cited When Used by Others**



Results from the 2012  
U of A Campus Data  
Summit Survey

# Data citation

Dataverse generates a Data Citation for each study and includes:

author(s), distribution date, title, global persistent identifier and url, and optional fields such as the distributor, data subsets, and versions.

Example:

- Authors, Year, Dataset Title, DOI, Data Repository, version
- Logan Vidal, 2013, "ANES data coding ", <http://dx.doi.org/10.7910/DVN/23274>, Harvard Dataverse, V8

# DOI Generation in Dataverse

Dataverse supports the identification and sharing of research materials in the scholarly publishing process.

- *DOI assigned* when upload data
- DOI not registered until the data are *released*
- Automatically registered with DataCite

<https://www.datacite.org/>

- DataCite DOI Citation Formatter (Beta)

<http://crosscite.org/citeproc/>

# DOI Citation

Publications are the engine room of the research enterprise.

As research data is forming to be one of the first class outcomes of the research, the link between publications and data is essential to make research data discoverable, reusable and credible.

Aryani, Amir (2014): ANDS is building the Australian Research Data Commons. figshare. <http://dx.doi.org/10.6084/m9.figshare.1045389>



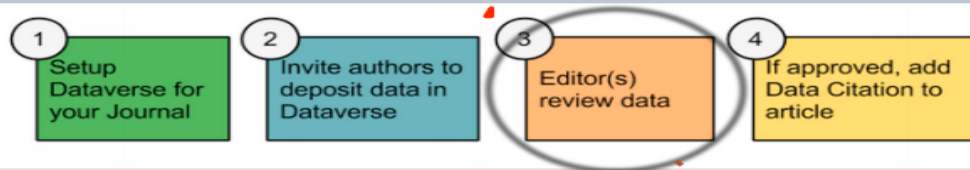
# PKP-Dataverse Integration Project

## Workflows that Connect Data to Journals

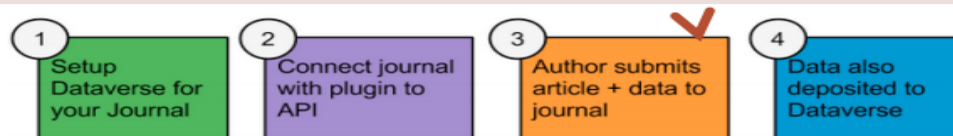
### A. Journals include Dataverse as a Recommended Repository



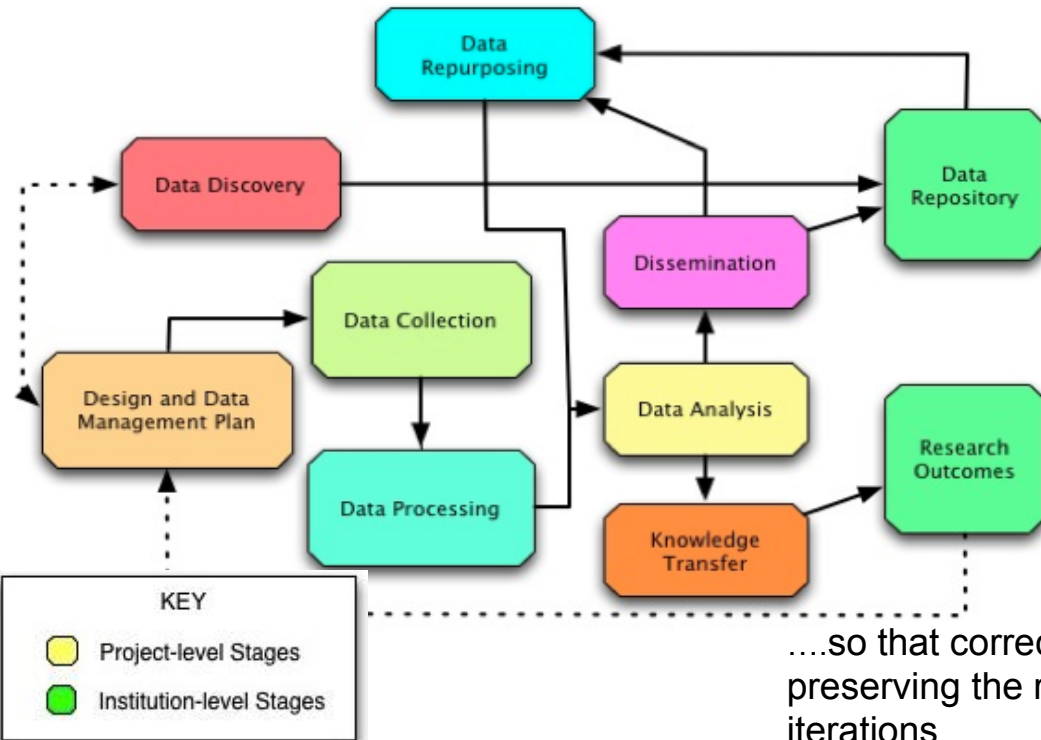
### B. Authors Contribute Directly to a Journal's Dataverse



### C. Seamless Integration btw Journal + Dataverse (e.g., OJS)



# Versioning



## Different Versions of Data Files Across the Lifecycle

Data files are transformed throughout the lifecycle of research as they go through Collection, Processing, and Analysis stages. Keeping track of versions can be a major headache for researchers.

....so that corrections can be made to published data while still preserving the replicability of results based on earlier iterations...

# Dataverse

- <https://dataverse.library.ualberta.ca/dvn/>
- <https://thedata.harvard.edu/dvn/>
- <http://dataverse.scholarsportal.info/dvn/>

# Examples:

- [World Historical Dataverse](#)
- [Harvard Design Course Collection](#)
- [Ethnic Power Relations Dataverse](#)
- [Queen's University Biological Station Data Archive Dataverse](#)
- [Agri-environmental research data repository Dataverse](#)
- [Institut de l'Environnement | Institute of the Environment Dataverse](#)

# Online Data Promotion

- Online CV: <http://gking.harvard.edu/biocv>
- Social Media



Wyman Laliberté  
@wyman10


Edmonton Annexation History 1892 - 1982  
(polygon .shp file) #HGIS  
[dx.doi.org/10.7939/DVN/10...](https://doi.org/10.7939/DVN/10...)

08:33 PM - 28 Apr 14


- <http://wyman10.tumblr.com/>

# Dataverse 4.0

## Integration with WorldMap (.shp files only)

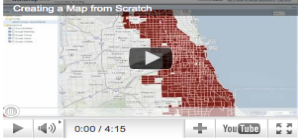
 **WorldMap** [Sign in](#) | [Create Map](#) | [View Map](#) | [Help](#)

[Create a Map](#) [View a Map](#) [About](#)



**Boston Research Map**

Build your own mapping portal and publish it to the world or to just a few collaborators. WorldMap is open source software.



[Watch the WorldMap Quick Start video](#)

<http://worldmap.harvard.edu/>

# Access Analysis + Subsetting

## SPSS, Stata, RData

An interface for quantitative analysis, that allows users at all levels of statistical expertise to explore their data, describe their substantive understanding of the data, and appropriately construct statistical models.



<http://datascience.iq.harvard.edu/tworavens>

# Other Options



<http://figshare.com/features>

(Offers 1GB of private space for research - hosted using Amazon Web Services)



<http://www.icpsr.umich.edu/icpsrweb/landing.jsp>



<http://datadryad.org/>



GitHub <https://github.com/> Citable code for academic software



# Related Material

Joint Declaration of Data Citation Principles: Implementation and Compliance in the Dataverse Repository

[http://thedata.org/files/thedata\\_new2/files/nisovirtualconference-datacitation-mercecrosas.pdf](http://thedata.org/files/thedata_new2/files/nisovirtualconference-datacitation-mercecrosas.pdf)

Dataverse User Guide: <http://thedata.harvard.edu/guides/dataverse-user-main.html>

The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data. D-Lib Magazine. January/February 2011, Volume 17, Number 1/2.

<http://www.dlib.org/dlib/january11/crosas/01crosas.html>