Scholars Portal Dataverse Update

By: Kevin Worthington

Date: October 27th, 2016

What is The **Dataverse** Project

- Open source research data repository
- For organizing, sharing/promoting, research replicability and archiving data
- Developed by Harvard's Institute for Quantitative Social Science (IQSS)
- Predecessor projects
  – Virtual Data Center (VDC) (1999-2006) for collaboration with MIT
 – Automatic transfer of cataloging information using FTP (1987-1999)

Dataverse Features

- Supports uploading of any file format
  - .zip, .xls, spss, sas, stata, pdf, doc, etc.
  - Version Control
- Allows Administration
  - Data management
    - Organize, Create, Update, Delete, Feature
  - Roles and Permissions at the file-level
  - Terms of Use
  - Theming
  - Metadata element inclusion
- Persistent ID support
- API's
- Sharable widgets
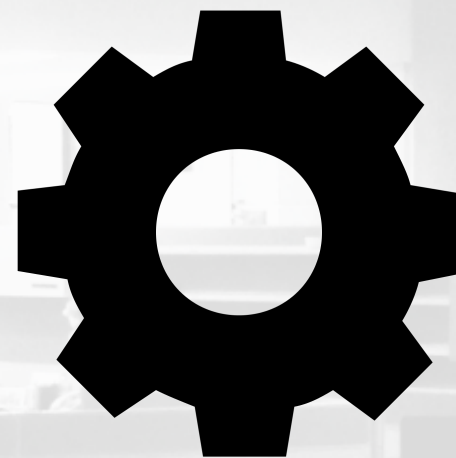
Background

- Deployed Ver. 3.6 at SP in 2011 by Steve Marks
- Self-deposit model (anyone can signup and deposit data)
- Currently have
  - 127 Dataverses
  - 688 Datasets
  - 8133 Files
  - approx. 390 GB size

New Datasets by year

250				
200				
150				
100				
50				
0				
2012	2013	2014	2015	2016

- 13 OCUL libraries using it
- Curating multidisciplinary data
- All to satisfy the growing need to deposit and share research data easily with others
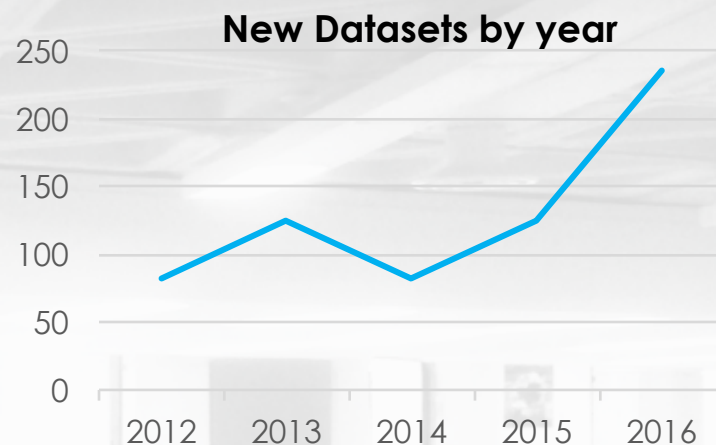
New Features from Upgrade

- DataCite Canada Digital Object Identifiers (DOIs)
  - DOI minting for new data deposits
- Extended metadata support
  - Health & Life Sciences, Geosciences, Social Sciences, Astronomy, Journals
  - Harvesting and Export support with OAI-PMH
- Nested Dataverses
- Shibboleth support (SSO)
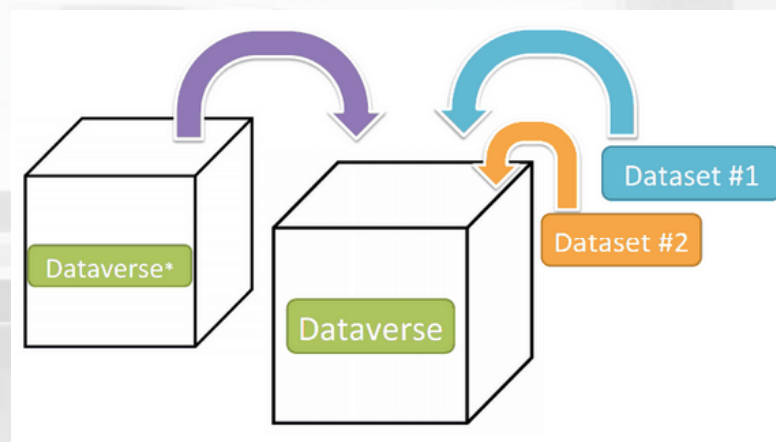
Custom Development

- Internationalization
  - Translations provided by Universite de Montreal – Beta Launched October 2016
- DDI Explorer
- Custom splash page
- Institutional
  - Dataverse mapping
  - Affiliation using IP Groups
  - Redirecting
- Released Ver. 4.5.1 On October 25th 2016

Dataverse Before

Silver Data Dataverse

🔍 ⓘ 💬   Create Account   University of Toronto IP Group   Log In

MYSPACE GENRES
hdl:10864/11402UNF:5:2W7x1+2m2IZPD68+wjb2XQ==
Version: 1 – Released: Tue Jan 19 14:43:12 EST 2016

< View Previous Study Listing

| Cataloging Information | **DATA & ANALYSIS** | Comments (0) | Versions |

ⓘ Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.

☐ Select all files   [Download Selected Files]   Total Number of Files: **2**   Total Downloads: **11**

☐ Data File

☐ MySpaceGenres.tab
Tab Delimited - 138 MB - 9 downloads + analyses
MD5 Checksum: 01c6aaf5155ad760a40b962fd5566ac9

| TABULAR DATA | 2883141 Cases | 5 Variables |

⬇   Download as... ▼   MySpace Genres

📈 Access Analysis + Subsetting   🅒 View Data Citation [+]

☐ Documentation

☐ MySpace Codebook.docx
application/octet-stream - 68 KB - 2 downloads
MD5 Checksum: 2cba90b75aad4e8bfa9fdcceaf7a7eba

⬇ Download   Codebook for MySpaceGenres.tab

POWERED BY THE **Dataverse Network**™ PROJECT v. 3.6.2

Silver Data Dataverse

🔍 ℹ 💬 Create Account   University of Toronto IP Group   Log In

MYSPACE GENRES

< Back to Study

hdl:10864/11402UNF:5:2W7x1+2m2IZPD68+wjb2XQ==
Version: 1 – Released: Tue Jan 19 14:43:12 EST 2016
Data File: MySpaceGenres.tab

| DOWNLOAD SUBSET | Recode & Case-Subset | Descriptive Statistics | Advanced Statistical Analysis |

Selected Variables

```
ID
GENRE2
```

ℹ Choose File Format to download selected variables:

- ◉ Text
- ○ R Data
- ○ S plus
- ○ Stata

[Create Zip File]

ℹ Select variables from table below (selected variables will be displayed above)

Show [10 Variables ▼]

Variable Information Table

☐	Type	Name	Label	Summary
☑	Character	ID	[label missing]	⊞
☐	Character	Name	[label missing]	⊞
☐	Character	GENRE1	[label missing]	⊞
☑	Character	GENRE2	[label missing]	⊞

Value (Label)	Frequency
A'cappella	I14448.0
Acousmatic	335.0
Acoustic	▮48655.0
Afro-beat	I3417.0
Alternative	▮61805.0

Choose a Statistical Model ▼

Choose a Statistical Model
Categorical Data Analysis
Cross-Tabulation
Ecological Inference Models
Hierarchical Multinomial-Dirichlet Ecological Inference Model for R x C Tables
Event Count Models
Negative Binomial Reg for Event Count Dep Vars
Poisson Reg for Event Count Dep Vars
Models for Continous Bounded Dependent Variables
Exponential Reg for Duration Dep Vars
Gamma Reg for Cont, Positive Dep Vars
Log-Normal Reg for Duration Dep Vars
Weibull Reg for Duration Dep Vars
Models for Continuous Dependent Variables
Least Squares Reg for Cont Dep Vars
Linear regression for Left-Censored Dep Variable
Models for Dichotomous Dependent Variables
Logistic Reg for Binary Dep Vars
Probit Reg for Binary Dep Vars
Rare Events Logistic Reg for Binary Dep Vars

Run Model

s from table below (selected variables will be displayed above)

Show   10 Va

Variable In

Type	Name	Label	Summar
Character	ID	[label missing]	▦
Character	Name	[label missing]	▦
Character	GENRE1	[label missing]	▦
Character	GENRE2	[label missing]	▦

Dataverse After

Search all dataverses... **Q Search**   About   Guides ▾   Support   Sign Up   Log In   **Fr (Beta)**

SCHOLARS PORTAL RESEARCH DATA PLATFORM

Publish and track your data, discover and reuse others' data!

Create My Dataverse   or   ☰ Explore

UNIVERSITY of GUELPH
University of Guelph Dataverse

uOttawa
University of Ottawa Dataverse

UNIVERSITY OF TORONTO
University of Toronto Dataverse

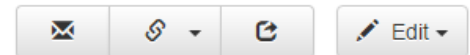UNIVERSITY OF WATERLOO
University of Waterloo Dataverse

Canada's federal research agencies are strong advocates for making publically-funded research data as accessible as possible. In 2016, the Tri-Agency released a **Statement of Principles on Digital Data Management** that outlines expectations and responsibilities for research data management and open data sharing.

Search all dataverses...   **🔍 Search**   About   Guides ▾   Support   Dashboard   👤 Kevin Worthington ▾

uOttawa

University of Ottawa Dataverse

✉   🔗 ▾   ↗   ✏ Edit ▾

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

✅ 👥 **Dataverses (9)**	**1 to 10 of 22 Results** ⇅ Sort ▾
✅ 📄 **Datasets (13)**	
☐ 📊 **Files (61)**	

Publication Status
Published (14)
Unpublished (5)
Draft (3)
Deaccessioned (1)

Publication Date
2014 (8)
2013 (4)
2015 (4)
2016 (1)

Author-Name

Dramatic Growth of Open Access 📄
Jun 29, 2016 - Dramatic Growth of Open Access Dataverse
Morrison, Heather, 2014, "Dramatic Growth of Open Access", hdl:10864/10660, Scholars Portal Dataverse, V13
Macro level data depicting the growth of open access. Updated quarterly (March 31, June 30, September 30, December 31 and usually an early December early year-end edition). This is a spreadsheet with multiple tab. The intention is to regularly update this dataverse with the lates...

Dramatic Growth of Open Access **Draft** 📄
Jun 29, 2016 - Dramatic Growth of Open Access Dataverse
Morrison, Heather, 2014, "Dramatic Growth of Open Access", hdl:10864/10660, Scholars Portal Dataverse, DRAFT VERSION
Macro level data depicting the growth of open access. Updated quarterly (March 31, June 30, September 30, December 31 and usually an early December early year-end edition). This is a spreadsheet with multiple tab. The intention is to regularly update this dataverse with the lates...

Jeanette Hatherill Dataverse

DDI Explorer

## The Globalization of Personal Data (GPD) Project International Survey on Privacy and Surveillance

Français

Surveillance Studies Centre, 2013, "The Globalization of Personal Data (GPD) Project International Survey on Privacy and Surveillance"

🔍 [_____] ✖   **293** Results   [Download My Variables] [Chart My Variables] [Download All ▼]   [Advanced Analysis ⏏]

My Variables Select All Clear	ID	Name	Label	Categories	Valid Cases
☐	49529	idx	[idx] Id number (Canada/ US)		
☐	49604	id	[id] Id number (except Canada/ US)		
☐	49956	country	[country] Country	7	
☐	49846	cma	[cma] CMA		
☐	49760	region_b	[region_br] BRAZIL: Region		
☐	49757	region_c	[region_ca] CANADA: Region	6	
☐	49688	region_f	[region_fr] FRANCE: Region	9	
☐	49982	region_h	[region_h] HUNGARY: Region	7	
☐	49598	region_m	[region_m] MEXICO: Region		
☐	49651	region_s	[region_s] SPAIN: Region	8	

First « 1 2 3 4 5 » Last

Records Per Page [10 ▼]

The Globaliz... ... ... of Personal Data (GPD) Project International S... ... Privacy and Surveillance

Surveillance Studie...

urveillance"

Advanced Analysi

My Variables
Select All   Clear

	Categories	Vali Cas
☐	9	
☐	9	
☐	9	
☐	9	
☐	9	
☐	9	
☐	9	hat
☐	9	rd
☐	9	from
☐	9	

First « 3 4 **5**

cords Per Page   10

Variable: [q11] How worried are you about providing personal information on websites? ✕
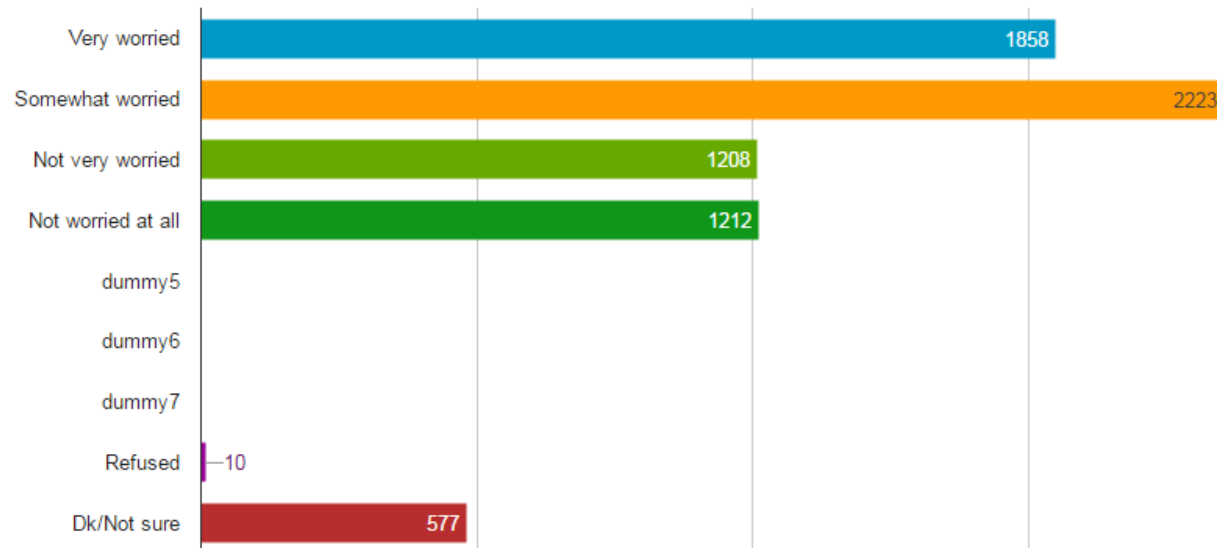
Values	Categories	N ⬍
1	Very worried	1858
2	Somewhat worried	2223
3	Not very worried	1208
4	Not worried at all	1212
5	dummy5	
6	dummy6	
7	dummy7	
8	Refused	10
9	Dk/Not sure	577

TwoRavens Advanced Analysis
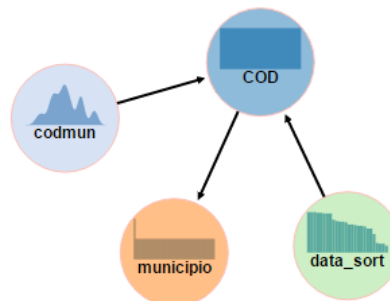
Environments/Tools

- Production
  - Load balancing →
- Demo
- Staging, Test, Dev
- GIT
- Jenkins
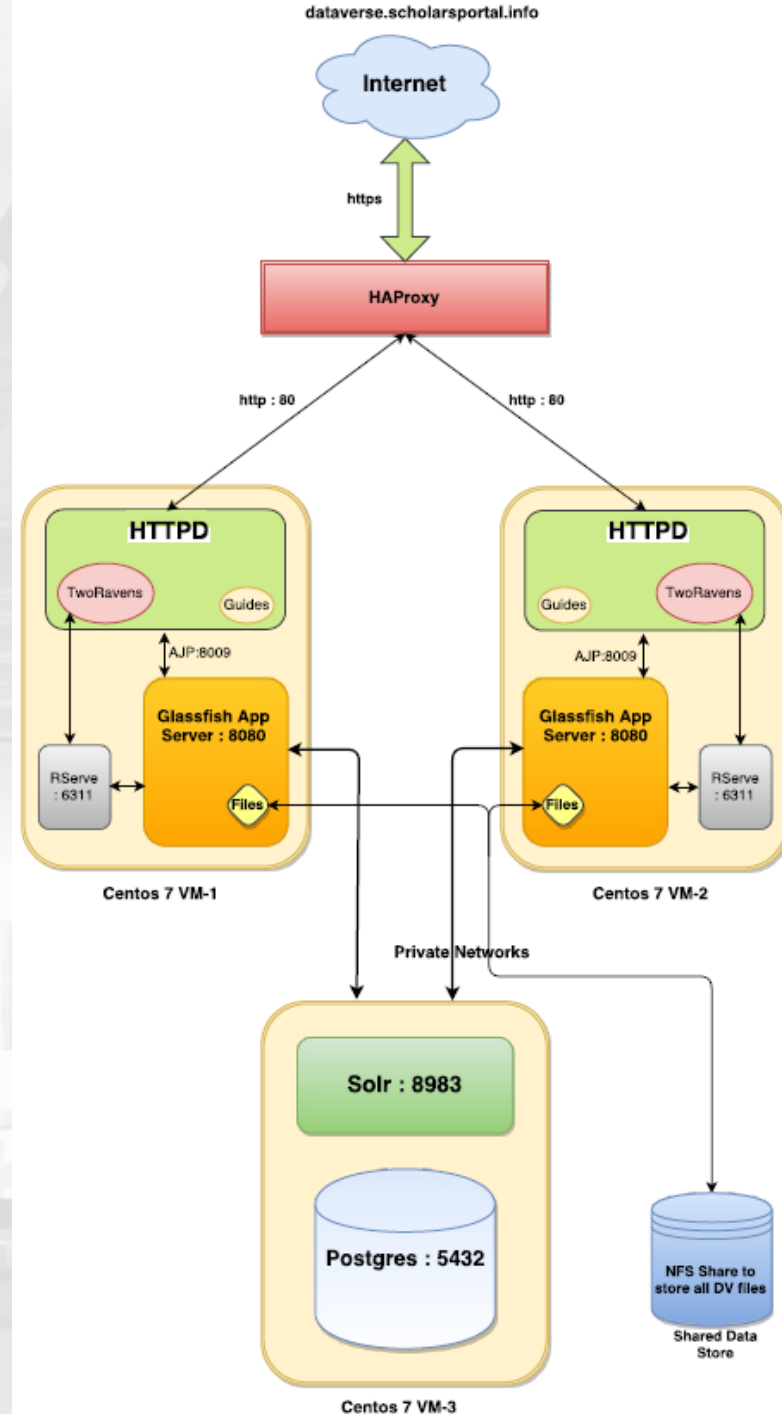
Research Data Preservation

- Addresses a preservation gap in Dataverse
- Contracted Artefactual Systems to integrate Dataverse and Archivematica
- Functional requirements were developed in cooperation with Canadian Dataverse users (e.g. UofA, SFU, UBC)
- Goal is to automate transfer of studies from Dataverse to Archivematica for long-term preservation
- The results of this test will be reported on for the broader research data community

Dataverse-Archivematica Integration Workflow

Artefactual Systems. "Dataverse-Archivematica Workflow Diagram". https://wiki.archivematica.org/Dataverse (Accessed 8/12/2016)
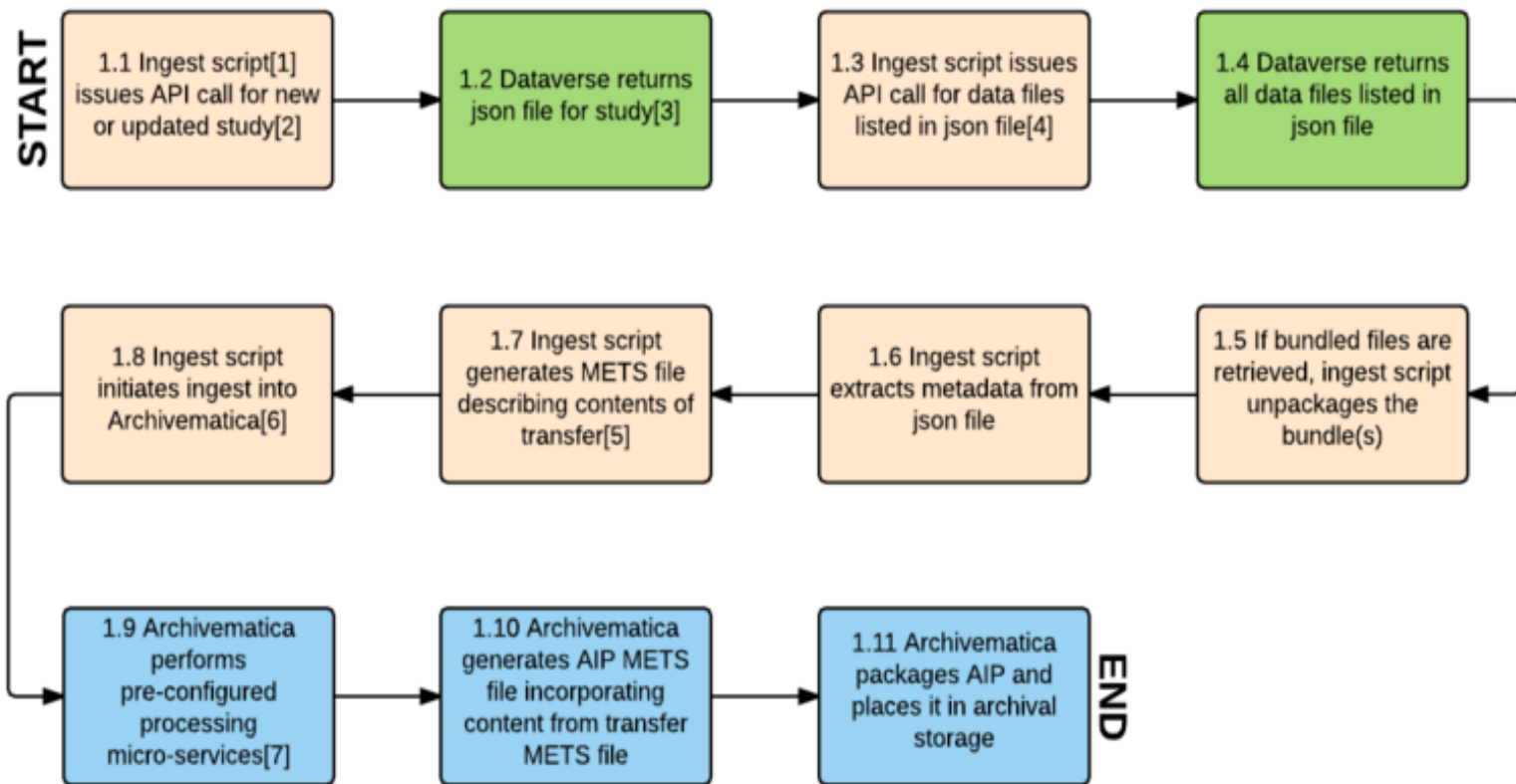
PERMAFROST Hosted Preservation Pilot

- Currently working with identified pilot partners to determine needs, develop terms, and scope services

- Vision is hosted instance of Archivematica connected to the Ontario Library Research Cloud (OLRC)

- Pilot will (hopefully) begin in Spring 2017

- The project is being led by Grant Hurley – contact grant@scholarsportal.info for more information

Next Steps

- Continue testing for scalability and preservation of research data
- Leverage IP Group Access (new in 4.5.1)
- Complete French translations
- Refine Advanced Analysis component
- Release Dataverse Bridge in ODESI
- Geospatial data support

Dataverse Upgrade Working Group

George Duimovich (Carleton)
Jane Fry (Carleton)
Carrie Breton (Guelph)
Carol Perry (Guelph)
Jason Brodeur (McMaster)
Vivek Jadon (McMaster)
Alexandra Copper (Queens)
Leanne Trimble (UofT)
Andrew Nicholson (UofT)
Steve Marks (UofT)
Jess Whyte (UofT)
Kathy Szigeti (Waterloo)
Genny Jon (York)

Kaitlin Newson (SP)
Jayanthy Chengan (SP)
Bart Kawula (SP)
Bikramjit Singh (SP)
Amber Leahey (SP)
Kevin Worthington (SP)

THANK YOU!