

Introduction to Dataverse

Meghan Goodchild
Scholars Portal & Queen's University
April 19, 2018



Plan

- Dataverse et la gestion des données de recherche (GDR)
- Scholars Portal Dataverse
- Dépôt des données
- Découverte des données et réutilisation
- Développements à venir
- Comment faut-il se commencer?

What is Dataverse?

- Open source research data repository software
- Store, share, publish and discover research data

DATAVERSE REPOSITORIES - A WORLD VIEW

32 Installations

2,666 Dataverses

50,999 Datasets

3,199,446 Downloads



Stats generated: 16th April 2018 10:23 EDT

What is Dataverse?

- Open source research data repository software
- Store, share, publish and discover research data



Open



Explore



Secure



Collaborative

Dataverse et GDR



Scholars Portal Dataverse

Scholars Portal **Dataverse**

🔍 Trouver

À propos

Guide d'utilisation

Soutien

S'inscrire

Se connecter

En

PLATEFORME DONNÉES DE RECHERCHE DE SCHOLARS PORTAL

Un outil pour publier et suivre vos données, et découvrir et réutiliser les données des autres.

Créer mon Dataverse

ou  Explorer



Queen's University Dataverse

Algoma
UNIVERSITY

Algoma University Dataverse



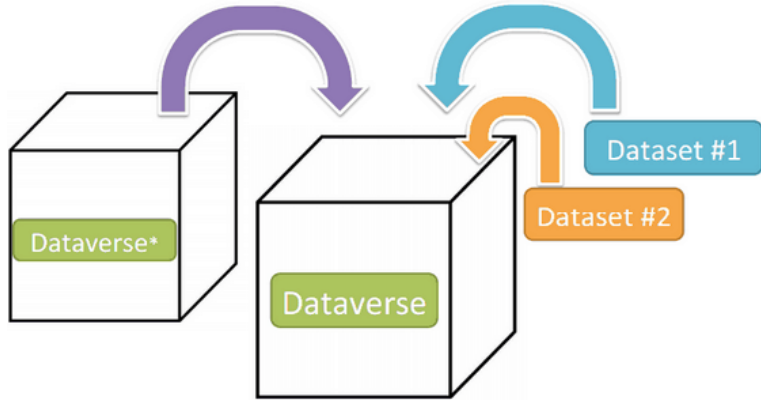
Brock University Dataverse



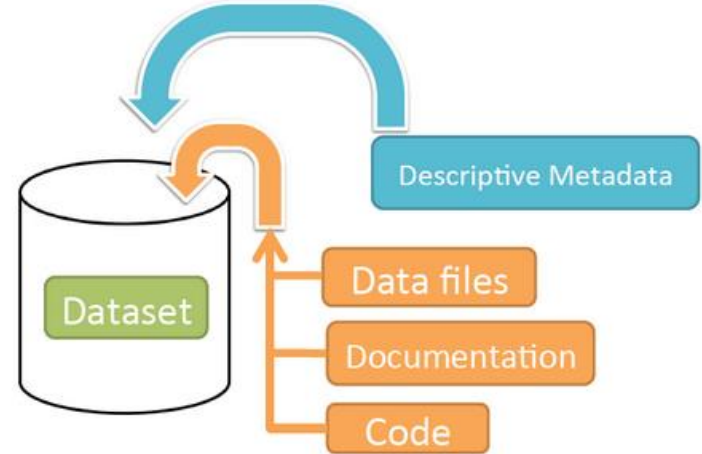
Carleton University Dataverse



Dataverses, Datasets, Files



Dataverse = Container for datasets
and/or dataverses



Dataset = Container for your data,
documentation, and code



Forum Research/Political Science Data Archive (University of Toronto)

Scholars Portal Dataverse > University of Toronto Dataverse > Forum Research/Political Science Data Archive > **Forum Research Political Poll - Provincial Issues (Ontario) 2017**

Statistiques

18 téléchargements

Personne-ressource Partager



Forum Research Political Poll - Provincial Issues (Ontario) 2017 Version 3.0

Forum Research Inc., 2017, "Forum Research Political Poll - Provincial Issues (Ontario) 2017", doi:10.5683/SP/IL6OHI, Scholars Portal Dataverse, V3, UNF:6:ysploNpettntnnWGNPy9Jw==

Citer l'ensemble de données ▾

Pour en apprendre sur le sujet, consultez le document [Data Citation Standards](#).

Description

This survey is an provincial opinion poll with general political questions as well as focus questions. Collection includes polls for Ontario (2017)

Sujet

Sciences sociales

Mot-clé

Politics, Provincial, Opinion poll

Fichiers

Métadonnées

Conditions


Versions

Chercher dans cet ensemble de données...















Trouver


[Fichiers](#)[Métadonnées](#)[Conditions](#)[Versions](#)

Chercher dans cet ensemble de données...

 Trouver

16 Fichiers

<input type="checkbox"/>		 Télécharger
<input type="checkbox"/>	 <p>Ontario - August 2017 - Codebook.docx MS Word (docx) - 292.10 KB - Nov 20, 2017 - 4 téléchargements MD5: 4fda8917d58e5559cd381b5dc020f6c8 Codebook for the August 2017 ON opinion poll. Contains metadata and questions.</p> <p>Documentation</p>	 Télécharger
<input type="checkbox"/>	 <p>Ontario - August 2017 - Dataset.tab Données tabulaires - 79.9 KB - Nov 20, 2017 - 5 téléchargements 30 Variables., 981 Observations - UNF:6:zKx4T2ggZM5aP74gFqvdbw== Survey data for the August 2017 ON opinion poll. Contains general political questions as well as focus questions.</p> <p>Survey Data</p>	 Explorer  Télécharger ▾
<input type="checkbox"/>	 <p>Ontario - February 2017 - Codebook.docx MS Word (docx) - 290.1 KB - Nov 17, 2017 - 0 téléchargements MD5: bc42eb383febe73b63a7791f889292c4 Codebook for the February 2017 ON opinion poll. Contains metadata and questions.</p> <p>Documentation</p>	 Télécharger
<input type="checkbox"/>	 <p>Ontario - February 2017 - Dataset.tab Données tabulaires - 74.4 KB - Nov 17, 2017 - 0 téléchargements 22 Variables., 1120 Observations - UNF:6:VubIZgdB7QuT0uwz15ltqQ== Survey data for the February 2017 ON opinion poll. Contains general political questions as well as focus questions.</p> <p>Survey Data</p>	 Explorer  Télécharger ▾
<input type="checkbox"/>	 <p>Ontario - June 2017 - Codebook.docx MS Word (docx) - 291.6 KB - Nov 20, 2017 - 0 téléchargements MD5: d2d1cf64fbb75b9d11761c327a97b7c9 Codebook for the June 2017 ON opinion poll. Contains metadata and questions.</p> <p>Documentation</p>	 Télécharger
<input type="checkbox"/>	 <p>Ontario - June 2017 - Dataset.tab Données tabulaires - 84.5 KB - Nov 20, 2017 - 0 téléchargements 34 Variables., 1095 Observations - UNF:6:BBFC3u7Cv5iQD2fuu1Q1A==</p>	

Identifiant pérenne de l'ensemble de données	doi:10.5683/SP/IL6OHI
Date de publication	2017-11-17
Titre	Forum Research Political Poll - Provincial Issues (Ontario) 2017
Auteur	Forum Research Inc.
Personne-ressource	 Utiliser le bouton de courriel ci-dessus pour communiquer avec cette personne. Map and Data Library (University of Toronto)
Description	This survey is an provincial opinion poll with general political questions as well as focus questions. Collection includes polls for Ontario (2017) (2017-06-12)
Sujet	Sciences sociales
Mot-clé	Politics Provincial Opinion poll
Langue	Anglais
Producteur	Forum Research Inc. http://forumresearch.com
Date de production	2017
Diffuseur	Department of Political Science (University of Toronto) http://politics.utoronto.ca
Date de disponibilité	2017
Déposant	Forum Research, Inc.
Date de dépôt	2016-06-12

Conditions d'utilisation 

Licence accordée

Les [normes de la communauté Dataverse](#) de même que les bonnes pratiques scientifiques exigent que toute source utilisée soit citée correctement. Veuillez utiliser la référence bibliographique ci-dessus générée par Dataverse.

licence CC0 - "Transfert dans le domaine public"



Usage

Scholars Portal **Dataverse** *Usage at a glance...*

71,769

Total number
of downloads

957

Number of
published
datasets

401 GB

Total size of
published
datasets

889

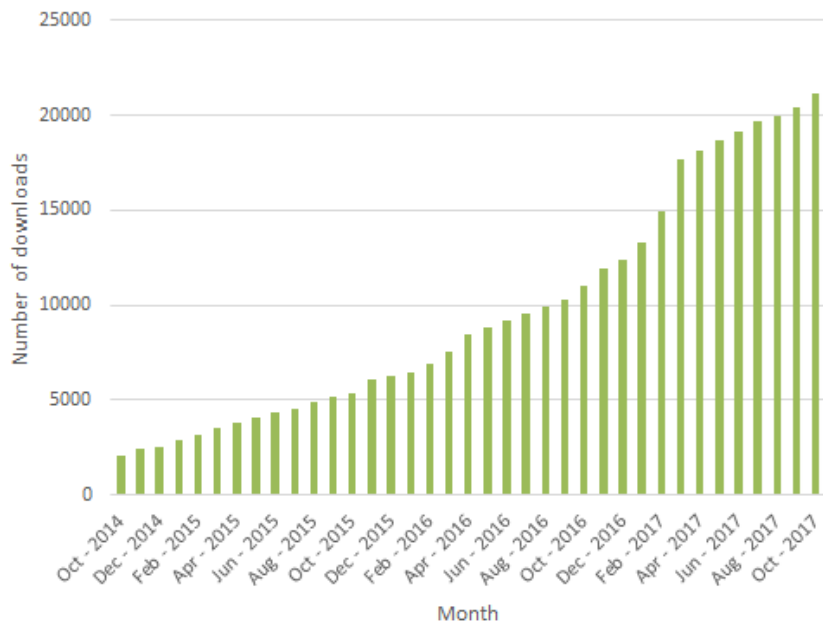
Number of
registered
users

Usage - Downloads over time

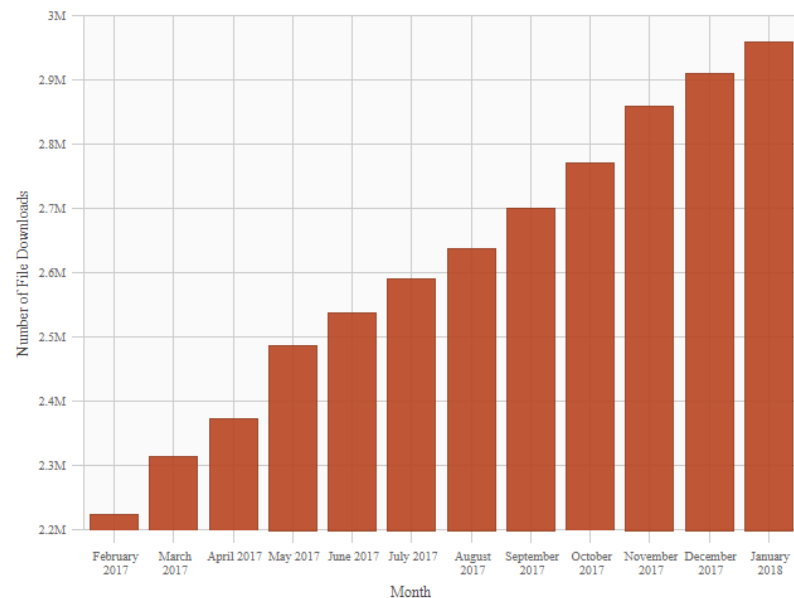
Scholars Portal **Dataverse**



Downloads Over Time




File Downloads Over Time



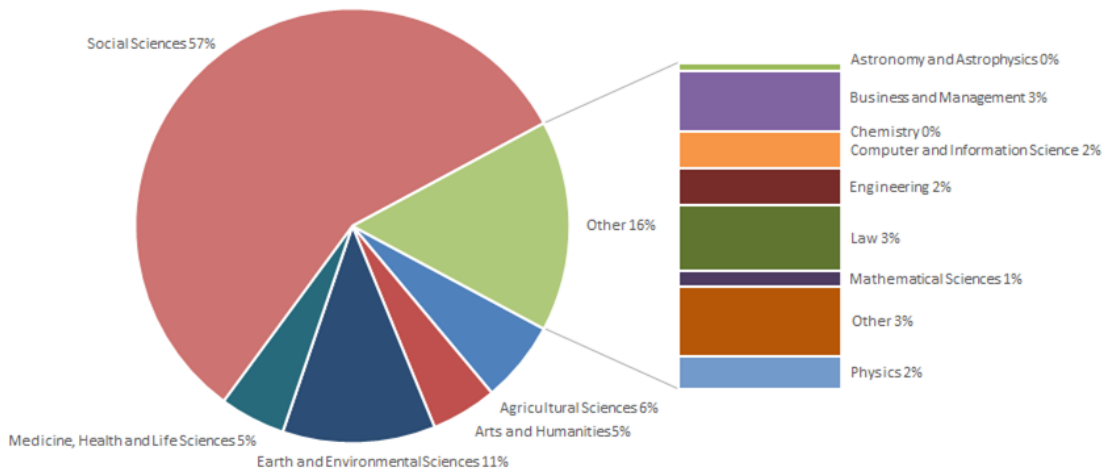
Usage statistics are for Harvard Dataverse: <https://dataverse.org/metrics>

Dataverse usage - Downloads by dataset

	Dataverse name	Dataset name
3809	University of Toronto	Forum Research Political Poll - Federal Issues (Canada) 2013
2998	University of Guelph Dataverse	Theano-based large-scale visual recognition with multiple GPUs
1369	University of Guelph Dataverse	Assessment of the fecal microbiota in beef calves
765	University of Toronto	2011 National Household Survey, Forward Sortation Area (FSA) Level [custom tabulation]
673	University of Ottawa Dataverse	Open access article processing charges May 2014

Usage - Disciplinary coverage

Scholars Portal **Dataverse**



Datasets by Subject

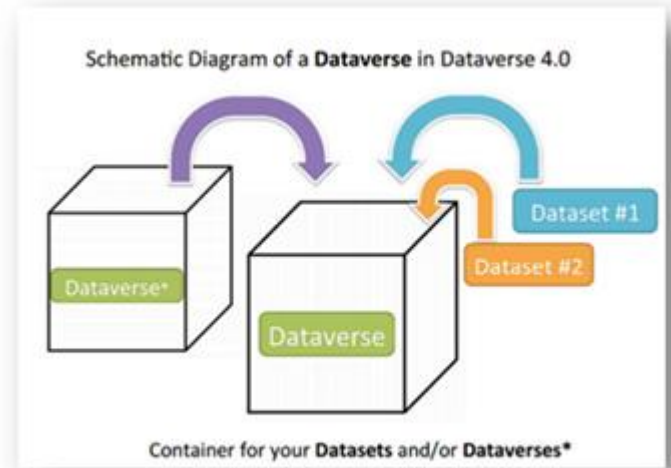


Data deposit

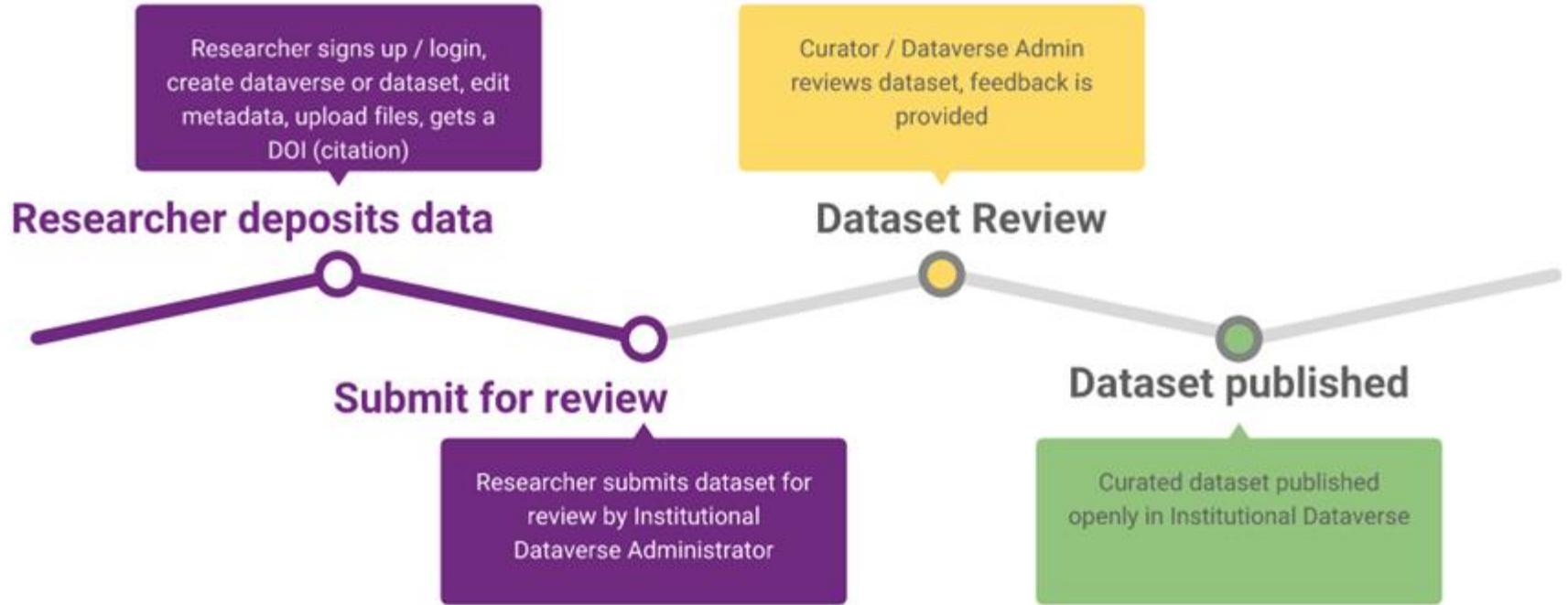
- Institutional dataverses
- Deposit types
- Metadata
- Versioning
- File formats

Institutional dataverses

- All data organized by institution (general root available for non-affiliates, multi-institutional projects)
- Researchers deposit in Institutional Dataverses (defined by user affiliation)
- Library administers institutional space
- Customizable features (branding, featured Dataverses, facets, etc.)



Mediated deposit



Self-deposit

- Open to anyone to deposit and publish data
- Usage statistics at institution-level to track published data

Note: SP Terms of Service covers removal of data if necessary (<http://guides.scholarsportal.info/dataverse>)

Metadata Standards

- Citation (DataCite, Dublin Core, DDI)
- Social Science (DDI)
- Astronomy (Virtual Observatory VOResource)
- Life Science (ISA-Tab)

Citation Metadata ▾

Geospatial Metadata ▾

Social Science and Humanities Metadata ▾

Astronomy and Astrophysics Metadata ▲

Type

Image
 Mosaic
 EventList
 Spectrum
 Cube
 Table


Facility

Instrument




Object


Spatial Resolution


Dataset Versioning


 **Canadian Media Concentration Research Project Dataverse** (Carleton University) CMCRP Project Site

Scholars Portal Dataverse > Carleton University Dataverse > Canadian Media Concentration Research Project Dataverse >
The Growth of the Network Media Economy in Canada, 1984-2016

 Metrics  3 Downloads  Contact  Share

 **The Growth of the Network Media Economy in Canada, 1984-2016** Version 2.0

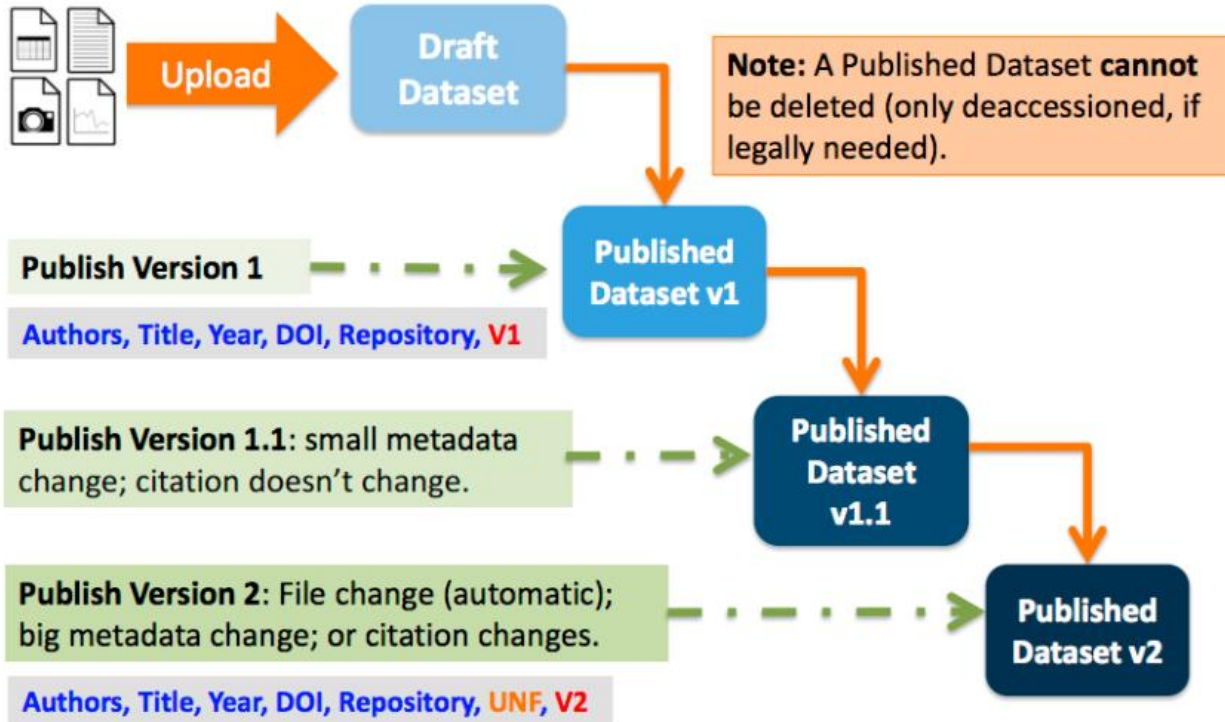
Winseck, Dwayne, 2017, "The Growth of the Network Media Economy in Canada, 1984-2016",
doi:10.5683/SP/I5ZHPR, Scholars Portal Dataverse, V2  Cite Dataset ▾

 [Learn about Data Citation Standards.](#)

Files Metadata Terms Versions

Dataset	Summary	Contributors	Published
2.0	Files (Added: 2); View Details	MacOdrum Library Cataloguing and Collections	December 13, 2017
1.0	This is the first published version.	MacOdrum Library Cataloguing and Collections, MacOdrum Library	December 7, 2017

Dataset Versioning (continued)



File formats

All file formats are supported

- Tabular data files (SPSS, R, excel, CSV)
- Geospatial files
- Images

SP Dataverse storage

- Maximum file size 2GB per file
- Unzipping, tabular ingest processing intensive

Data discovery and re-use

- Linking research outputs
- Permissions and licensing
- Discovery and access
- Collaboration
- Data explorer

ORCID and DataCite integration

- Mechanisms for linking research output
 - ORCID ID field
 - ORCID sign-in *not configured
 - DataCite Canada DOI minting
 - DataCite indexing



Cross-referencing research outputs

Obtain a DOI /
data citation **before**
publishing data

Reference and link to
publication in Dataverse
record



Provide private access
to data during review
process



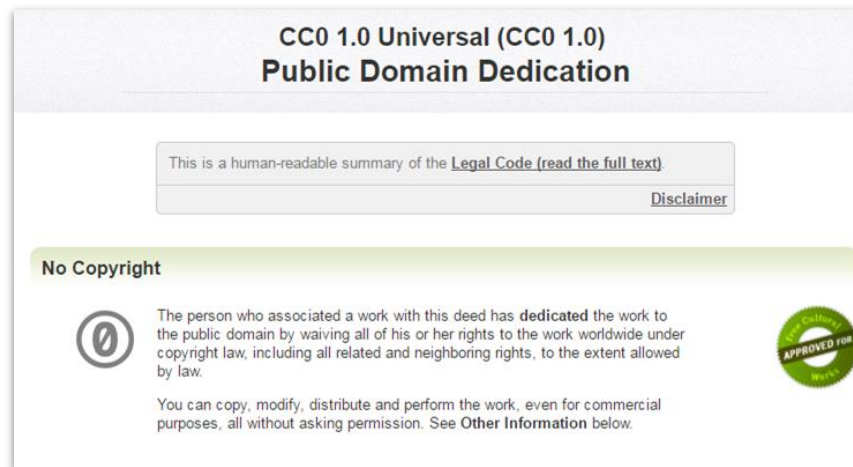
Cite and link to data in
publication

Licensing

- Default to CC0 (open data)

OR

- Custom terms of use 'Data Usage Agreement'
- Restricted files and custom terms of access



Sample Data Usage Agreement

(Note: This is a sample DUA for datasets that have de-identified human subject data)

This is an agreement ("Agreement") between you the downloader ("Downloader") and the owner of the materials ("User") governing the use of the materials ("Materials") to be downloaded.

I. Acceptance of this Agreement

By downloading or otherwise accessing the Materials, Downloader represents his/her acceptance of the terms of this Agreement.

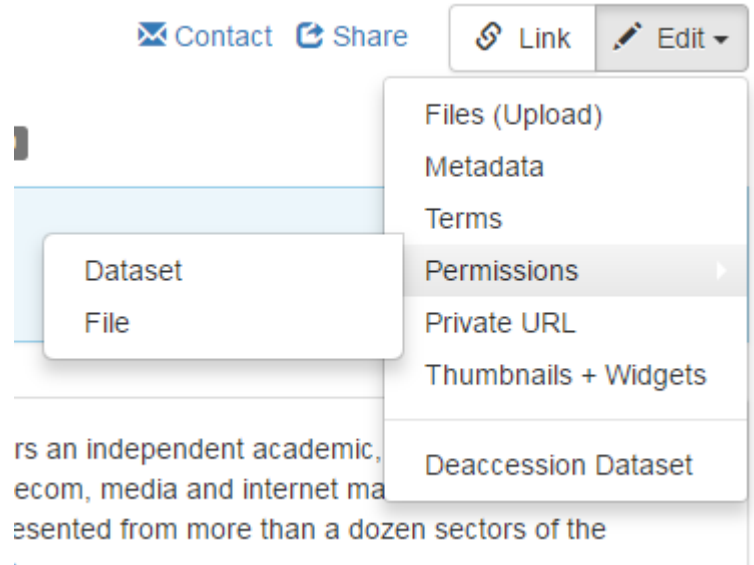
II. Modification of this Agreement

Users may modify the terms of this Agreement at any time. However, any modifications to this Agreement will only be effective for downloads subsequent to such modification. No modifications will supersede any previous terms that were in effect at the time of the Downloader's download.


Dataset and file permissions

User/groups and roles

- Assign permissions for collaborators, curators, file downloaders (access)
- Granular file-level permissions
- IP Group based permissions



Dataset and file permissions (continued)

 Canadian Media Concentration Research Project Dataverse (Carleton University) CMCRP Project Site

Scholars Portal Dataverse > Carleton University Dataverse > Canadian Media Concentration Research Project Dataverse > The Growth of the Network Media Economy in Canada, 1984-2016 > **Dataset Permissions**

Users/Groups ⤴ All the users and groups that have access to your dataverse.

[Assign Roles to Users/Groups](#)

5 Users/Groups

User/Group Name (Affiliation) ⤴	ID ⬇	Role ⬇	Action
CMCR Project (Carleton University)	@cmcrp	Admin	<i>Role assigned at Canadian Media Concentration Research Project Dataverse</i>
George Duimovich (Carleton University, MacOdrum Library)	@parser	Admin	<i>Role assigned at Canadian Media Concentration Research Project Dataverse</i>
MacOdrum Library (Carleton University)	@cudata	Contributor	✖ Remove Assigned Role
MacOdrum Library (Carleton University)	@cudata	Curator	<i>Role assigned at Canadian Media Concentration Research Project Dataverse</i>
MacOdrum Library Cataloguing and Collections (Carleton University)	@cucat	Curator	<i>Role assigned at Canadian Media Concentration Research Project Dataverse</i>

Roles ⤴ All the roles set up in your dataverse, that you can assign to users and groups.

Admin - A person who has all permissions for dataverses, datasets, and files.
[AddDataverse](#) [AddDataset](#) [ViewUnpublishedDataverse](#) [ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataverse](#) [EditDataset](#) [ManageDataversePermissions](#) [ManageDatasetPermissions](#)
[PublishDataverse](#) [PublishDataset](#) [DeleteDataverse](#) [DeleteDatasetDraft](#)

Contributor - For datasets, a person who can edit License + Terms, and then submit them for review.
[ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataset](#) [DeleteDatasetDraft](#)

Curator - For datasets, a person who can edit License + Terms, edit Permissions, and publish datasets.
[AddDataverse](#) [AddDataset](#) [ViewUnpublishedDataverse](#) [ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataset](#) [ManageDataversePermissions](#) [PublishDataset](#) [DeleteDatasetDraft](#)

File Downloader - A person who can download a published file.
[DownloadFile](#)

Member - A person who can view both unpublished dataverses and datasets.
[ViewUnpublishedDataverse](#) [ViewUnpublishedDataset](#) [DownloadFile](#)

Discovery and Access

- Open Archives Initiative (OAI) protocol for metadata harvesting (OAI-PMH)
 - Metadata from published, unrestricted datasets can be harvested
- Dataverse APIs
 - Search API
 - Data Access API
 - SWORD API (upload)
 - Native API

Guestbooks

Who is downloading my data?

- User fills out guestbook form
- Owner downloads guestbook report
- Can be used to mediate access / approval of access to data



Ryerson University Dataverse

Root Dataverse > Ryerson University Dataverse > Create Dataset Guestbook

Guestbook Name *

Data Collected Dataverse account information that will be collected when a user downloads a file. Check the ones that will be required.

- Name
- Email
- Institution
- Position

Custom Questions Create your own questions to have users provide more than their account information when they download a file. Questions can be required or optional and answers can be text or multiple choice.

Question Type	Question Text	
Multiple Cho ▾	Are you intending to use this data for your personal research?	+
	Yes	+ -
	No	+ -

Required field

Collaborative data sharing

- Group permissions
- Assign account roles
- Track changes through versioning



Data Explorer

Drone Awareness and Perceptions: A Three Country Study 2014

Français

Angus Reid Global, 2014, "Drone Awareness and Perceptions: A Three Country Study 2014", hdl:10864/10798, Scholars Portal Dataverse, V5

Q [x] 85 Results Download

ID	Name	Label
104579	PanelistIdQuestion	PanelistIdQuestion
104625	gender	Gender
104461	Q5_Federal_Vote_2011	Did you vote in the last FEDERAL election in May 2011?
104501	Q6_Federal_PartyChoice_2011	Who did you vote for in the last FEDERAL Election?
104666	Age_Rollup_broad	Age_Rollup_broad
104709	Income	Income
104442	HH_Income	Household Income
104512	Education_Level	Education
104538	Region_Rollup	Region_Rollup
104553	MACRO_REGION	MACRO_REGION

First « 1 2 3 4 5 » Last Records Per Page 10

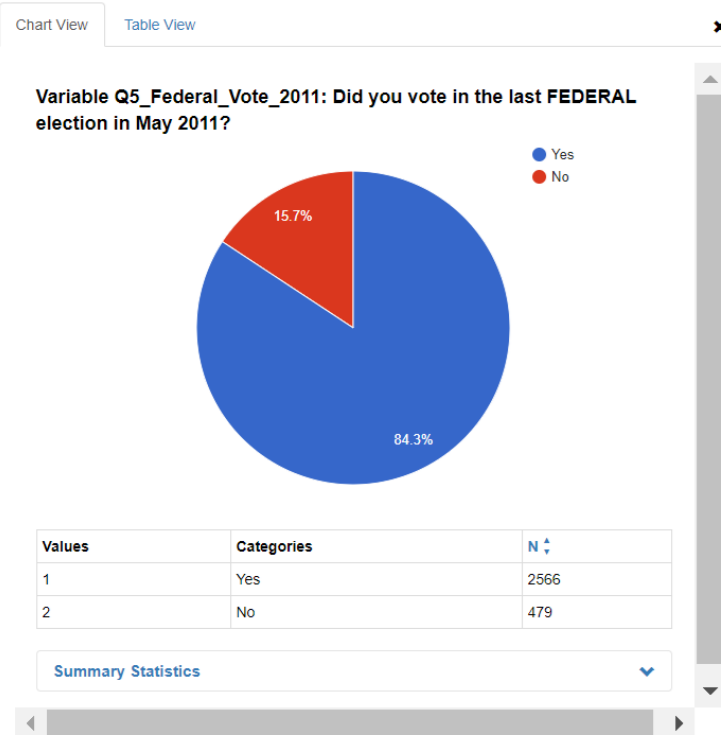


Chart
view

Data Explorer

Drone Awareness and Perceptions: A Three Country Study 2014

Angus Reid Global, 2014, "Drone Awareness and Perceptions: A Three Country Study 2014", hdl:10864/10798, Scholars Portal Dataverse, V5

Q x 85 Results Download ▾

Chart View Table View

ID	Name	Label
104579	PanelistIdQuestion	PanelistIdQuestion
104625	gender	Gender
104461	Q5_Federal_Vote_2011	Did you vote in the last FEDERAL election in May 2011?
104501	Q6_Federal_PartyChoice_2011	Who did you vote for in the last FEDERAL Election?
104666	Age_Rollup_broad	Age_Rollup_broad
104709	Income	Income
104442	HH_Income	Household Income
104512	Education_Level	Education
104538	Region_Rollup	Region_Rollup
104553	MACRO_REGION	MACRO_REGION

First < 1 2 3 4 5 > Last

Records Per Page 10 ▾

Who did you vote for in the last FEDERAL Election?	Region_Rollup	%
Conservative Party of Canada	BC	4.8
	Alberta	3.5
	SK	1.7
	MB	1.3
	ON	14.0
	QC	8.8
	Atlantic	3.6
Liberal Party of Canada	Territories	0.0
	BC	2.3
	Alberta	1.9
	SK	0.7
	MB	0.9
	ON	7.0
	QC	4.8
	Atlantic	1.6
	Territories	0.0
	BC	4.4
	Alberta	3.2
	SK	1.1

Cross
tabulation






Preservation

Current support:

- Native support for file verification
 - checksums, UNF
 - file format identification
- DOIs for persistent identification of data
- Derivative generation (tabular data - .tsv/tab)

Needed:

- Policies required
 - repositories have some responsibilities (short term)
 - institution focused long-term archiving

<input type="checkbox"/>		LaurelCreek_Invertebrate_Water_Spring2017.csv Comma Separated Values - 4.9 KB - Jan 9, 2018 - 0 Downloads MD5: 08a302dd8603f5faf71c90485f96fde3
<input type="checkbox"/>		LaurelCreek_Invertebrate_Water_Winter2017.csv Comma Separated Values - 3.2 KB - Jan 9, 2018 - 0 Downloads MD5: 899ff2e8f35326b824b06ab7e43e2595
<input type="checkbox"/>		LaurelCreek_Map.tif TIFF Image - 2.4 MB - Jan 9, 2018 - 1 Download MD5: 2b8a158126afb8d346ebb0fc9853768d
<input type="checkbox"/>		LaurelCreek_SiteFeatures_2017.csv Comma Separated Values - 11.8 KB - Jan 9, 2018 - 0 Downloads MD5: c8c0440a9e2864b15b017a850f654998
<input type="checkbox"/>		README_LaurelCreek_Water_Invertebrate_2017.txt Plain Text - 2.4 KB - Jan 9, 2018 - 0 Downloads MD5: 2b18152570ef8dca95398cac19438e34

Coming Next

- Upgrading to 4.8.6 with new French translations (UdeM)
- Dataverse-Archivematica integration
- Investigating Shibboleth login support
- Persistent identifiers at the file level
- Support for file hierarchy
- Large file support and HTTP upload support

Why use Dataverse?

- Supports FAIR data principles
 - Findable, Accessible, Interoperable, Reuseable
- Secure data management
- Effective sharing
- Long-term access and preservation
- Increase research visibility

Scholars Portal **Dataverse**

The
Dataverse
Project 

Get started with using Dataverse

Scholars Portal Demo Dataverse

<http://demodv.scholarsportal.info>

Scholars Portal Production Dataverse

<http://dataverse.scholarsportal.info>

Free, open for all researchers in Canada

SP Dataverse support team



Amber Leahey - Data & GIS Librarian

Meghan Goodchild - RDM Systems Librarian (Queen's / SP)

Kaitlin Newson - Digital Projects Librarian

Kevin Worthington - Data/GIS Programmer

Jayanthi Chengan - Senior Developer

Bikram Singh - Systems Analyst

Contact us at dataverse@scholarsportal.info

Équipe de GDR à uOttawa



Gestion des données de recherche



vous par « gestion des données de recherche »?

La gestion des données de recherche représente une activité continue tout au long du cycle de vie des données. Ce processus comprend l'organisation et la gestion des données (à mesure que les recherches avancent), et l'archivage approprié des données à la fin du projet.

[En savoir plus](#)

Qu'est-ce que les données de recherche?

Les données de recherche sont des données utilisées comme sources principales pour appuyer la recherche technique ou scientifique, les travaux de recherche, l'érudition ou les activités artistiques.

[En savoir plus](#)

Pourquoi investir temps et énergie à gérer des données?

De nombreux éditeurs et organismes subventionnaires demandent aux auteurs et titulaires de subventions de rendre accessible au public les données sous-jacentes des recherches publiées et des produits intellectuels.

[En savoir plus](#)

Planification de gestion des données

La planification est une étape de nos formations. Elle permet d'organiser les données de recherche et de garantir leur accès et leur préservation à long terme.

[En savoir plus](#)

Conférence : La gestion des données pour les projets de recherche

Équipe de gestion des données de recherche

Courriel : gdr@uOttawa.ca

Resources

Dataverse Usage Statistics (2014-2017)

- [Report](#) | [Spreadsheet](#)

User Guides

- [SP Guide](#) | [Harvard Guide](#)

Questions?