



# Introduction to Dataverse

Kaitlin Newson  
Digital Projects Librarian, Scholars Portal

Laure Perrier  
Research Data Management Librarian, UTL

UTL Senior Staff Meeting, December 10 2019

<http://tiny.cc/rpshhz>

# What is Dataverse?

- Open-source software developed by the Institute for Quantitative Social Sciences (IQSS) at Harvard
- Initially developed with social sciences data in mind, but expanding to other areas
- 52 installations worldwide





# Dataverse at Scholars Portal

- Research data repository hosted at Scholars Portal since 2012
- National service used by 48 institutions across Canada, including OCUL & BCI
- 2300+ published datasets
- <https://dataverse.scholarsportal.info>



## Benefits for Researchers

- Web-based tool with no software installation required
- Meet grant/journal requirements for sharing research data
- Share research data publicly, privately, or within a single institution
- Improved data discoverability through harvesting
- Make data explorable in the Data Explorer interface (for certain file types)
- Create a persistent identifier (DOI) for datasets
- Automatically generate citations for datasets
- Dataset versioning




## When to use Dataverse

- Small to medium-sized data; 2.5 GB per file limit
- Dependent on context and domain
  - Are there popular domain-specific repositories for the field?
  - How much control does the researcher want?
  - Does the data need to be restricted? Are there sensitive data files?
  - Funding availability
- Portage Data Repository options in Canada guide:  
[https://portagenetwork.ca/wp-content/uploads/2019/06/Repository-Options-Guide\\_EN.pdf](https://portagenetwork.ca/wp-content/uploads/2019/06/Repository-Options-Guide_EN.pdf)


# Institutional Dataverses

Scholars Portal **Dataverse** Search ▾ User Guide Support English ▾ Log In




**University of Toronto Dataverse** (University of Toronto)


Scholars Portal Dataverse > **University of Toronto Dataverse** ✉ Contact ↻ Share



Map & Data Library Dataverse





University of Toronto Mississauga Library Dataverse




University of Toronto Scarborough Library Dataverse

Search this dataverse... 🔍 Find [Advanced Search](#)

 **Dataverses (36)** 1 to 10 of 541 Results ⬆️ Sort ▾

 **Datasets (505)**

 **Files (7,256)**

Replication code and data for "Can Tolling Help Everyone? Estimating the Aggregate and Distribution Consequences of Congestion Pricing"  
Nov 18, 2019

# Data Explorer

Drone Awareness and Perceptions: A Three Country Study 2014  
Drones-Canada.tab  
Angus Reid Global, 2014, "Drone Awareness and Perceptions: A Three Country Study 2014"

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

Chart View Table View

Variable Education\_Level: Education

Values	Categories	N
1	Elementary/grade school or some high school	281
2	High school graduate	913
3	Some college / technical school	332
4	Completed college / technical school	611
5	Some university	200
6	University undergraduate degree	435
7	Some post-graduate school	79
8	Post-graduate degree	194

Records Per Page 10

# Data Curation Tool

## Labour Force Survey (Curated)

LFS2016-01\_PUMF\_EN.tab

Tester, Curation, 2019, "Labour Force Survey (Curated)", <https://doi.org/10.5072/FK2/YLJJAY>, Scholars Portal Dataverse, V1, UNF:6:EerkQFr2ySzwCu4oV5jExA== [fileUNF]

< Hide Groups

Download

Save

Add Group



Search



Items per page: 25

1 - 25 of 75



All Variables

Number of hours



<input type="checkbox"/>	ID	Name	Label	Weight	View	
<input type="checkbox"/>	v14110	REC_NUM	Order of record in file			
<input type="checkbox"/>	v14130	SURVEAR	Survey year			
<input type="checkbox"/>	v14106	SURVMNTH	Survey month			
<input type="checkbox"/>	v14127	LFSSTAT	Labour force status			
<input type="checkbox"/>	v14165	PROV	Province			
<input type="checkbox"/>	v14168	CMA	3 largest CMAs			
<input type="checkbox"/>	v14155	AGE_12	Five-year age group of respondent			
<input type="checkbox"/>	v14136	AGE_6	Age in 2 and 3 year groups			





## Other Features

- Shibboleth Login
- Bilingual application
- Canadian-hosted data



# Ongoing Developments

- Larger data support (as part of Canarie grant)
- User experience improvements in coordination with UTL UX Lab
- File preview tools
- Upgrades to align with Harvard releases



# Communities & Engagement

- Scholars Portal Dataverse community
  - Monthly community calls with contacts at institutions
- Portage Dataverse North
  - Sub-groups focusing on training, metadata & policy
- OCUL Data community



## Resources

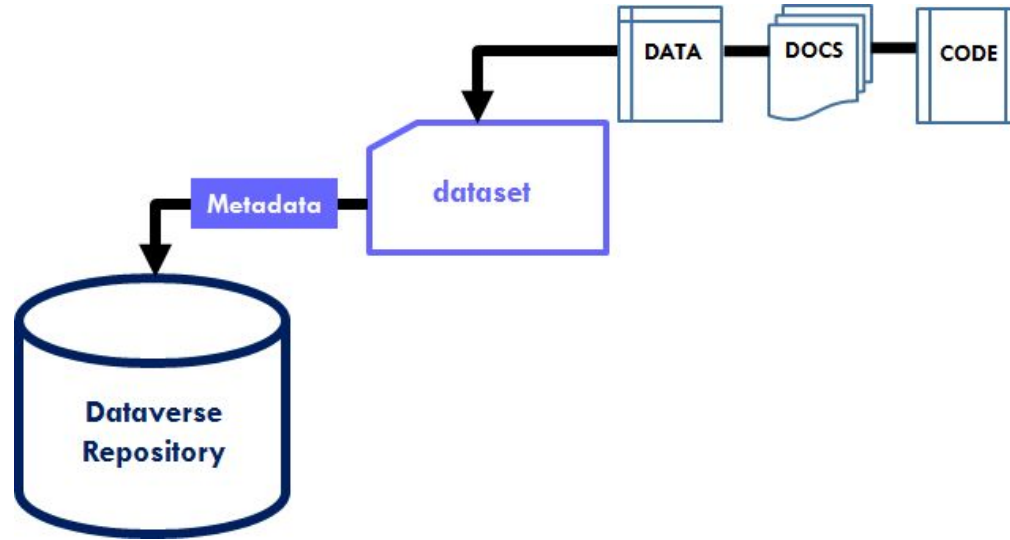
- Portage Dataverse Brief Guide:  
[https://portagenetwork.ca/wp-content/uploads/2019/08/Brief-Guide\\_Dataverse\\_EN.pdf](https://portagenetwork.ca/wp-content/uploads/2019/08/Brief-Guide_Dataverse_EN.pdf)
- Portage Dataverse Primer:  
[https://portagenetwork.ca/wp-content/uploads/2019/08/Primer\\_Dataverse\\_August2019\\_EN.pdf](https://portagenetwork.ca/wp-content/uploads/2019/08/Primer_Dataverse_August2019_EN.pdf)
- Scholars Portal Dataverse user guide:  
<https://learn.scholarsportal.info/all-guides/dataverse>



# UToronto Dataverse

A Focused Look

# How Does Data Go In To Dataverse?





# Key Features

- MD5 checksum - cannot upload a duplicate file (within a dataset)
- Versioning mechanism for published files within a dataset
- Tabular data (e.g., Excel, SPSS, etc.) converted to tab-delimited preservation copy
- Data usage license default: CCO - “Public Domain Dedication”



# Key Features

## Metadata

- Default: citation and descriptive metadata - Dublin Core and Datacite compliant
- Option to add domain-specific, project-based metadata
  - Social Science: DDI (Data Documentation Initiative)
  - Astronomy & Astrophysics: VO Discovery and Provenance Metadata
  - Life Sciences: ISA-TAB Specification



# Replication code and data for "Can Tolling Help Everyone? Estimating the Aggregate and Distribution Consequences of Congestion Pricing"

Version 3.0

Hall, Jonathan D., 2019, "Replication code and data for "Can Tolling Help Everyone? Estimating the Aggregate and Distribution Consequences of Congestion Pricing"", <https://doi.org/10.5683/SP2/RZS1FL>, Scholars Portal Dataverse, V3, UNF:6:udhpVGecTKPwoeer7mUy8g== [fileUNF]

Cite Dataset

[Learn about Data Citation Standards.](#)

Dataset Metrics

0 Downloads

## Description

This repository contains the online appendix and replication code and data for "Can Tolling Help Everyone? Estimating the Aggregate and Distribution Consequences of Congestion Pricing". Please see [readme.pdf](#) for more details. Also note that "tree view" is much more user friendly than the default "list view".

Abstract: Economists have long advocated road pricing as an efficiency-enhancing solution to traffic congestion, yet it has rarely been implemented because it is thought to create losers as well as winners. This paper uses survey and travel time data, combined with a structural model of traffic congestion, to estimate the joint distribution of agent preferences and evaluate the aggregate and distributional effects of road pricing. I find that adding tolls on half of the lanes of a highway yields a Pareto improvement. Further, the social welfare gains from doing so are substantial—up to \$1,740 per road user per year.

## Subject

Social Sciences

Files

Metadata

Terms

Versions

Change View

Table

Tree

Search this dataset...

Find

Filter by

File Type: All

Access: All

Sort

1 to 10 of 67 Files

Download








bootstrapped\_counterfactual\_results  
Replication Files/Intermediate Output/  
Plain Text - 1.8 KB - Nov 15, 2019 - 0 Downloads  
MD5: e8e0608c4f923aaa31a2a456c006a787

Download

- Restrict access
- Add tags

1 to 5 of 5 Files

<input type="checkbox"/>		<a href="#">NHES_SchoolReadinessSurvey_2007_Codebook.pdf</a> Adobe PDF - 219.6 KB - Oct 3, 2019 - 0 Downloads MD5: 7ccd0ea75493171b2575b1a7ac928125 <b>Documentation</b>
<input type="checkbox"/>		<a href="#">NHES_SchoolReadinessSurvey_2007_Dataset.tab</a> Tabular Data - 38.9 KB - Oct 3, 2019 - 3 Downloads 9 Variables, 2139 Observations - UNF:6:+FKZ0IWH98MxWWSv2aj7Fg== Information from 298 participants. <b>Data</b>
<input type="checkbox"/>		<a href="#">NHES_SchoolReadinessSurvey_2007_Questionnaire.pdf</a> Adobe PDF - 234.1 KB - Oct 3, 2019 - 0 Downloads MD5: f6f229e68a8f2d934915be154269d928 Questionnaire with 27 questions. <b>Survey</b> <b>Documentation</b>
<input type="checkbox"/>		<a href="#">NHES_SchoolReadinessSurvey_2007_Rscript.R</a> R Syntax - 6.5 KB - Oct 3, 2019 - 0 Downloads MD5: 8a88bda0f12094b2b03a05d3f8270c1e <b>Code</b>
<input type="checkbox"/>		<a href="#">README.txt</a> Plain Text - 2.4 KB - Oct 3, 2019 - 0 Downloads MD5: 30b64c5620ed1c2713d8d9e7bd9f1595 <b>Documentation</b>



# UToronto Dataverse

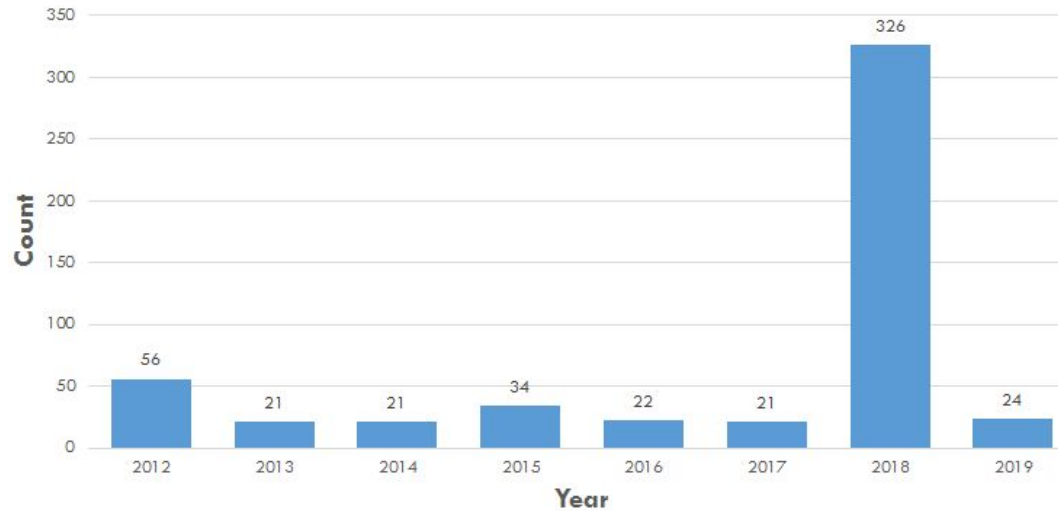
- Dataverses: 38
- Datasets: 512
- Files: 7,263

Date: December 10, 2019



# UToronto Dataverse

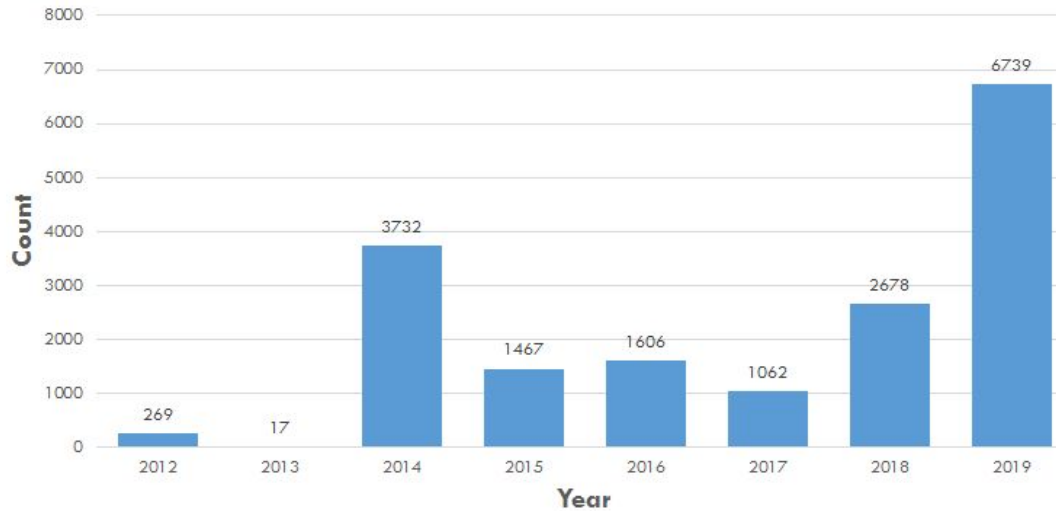
## Dataset Deposits by Year





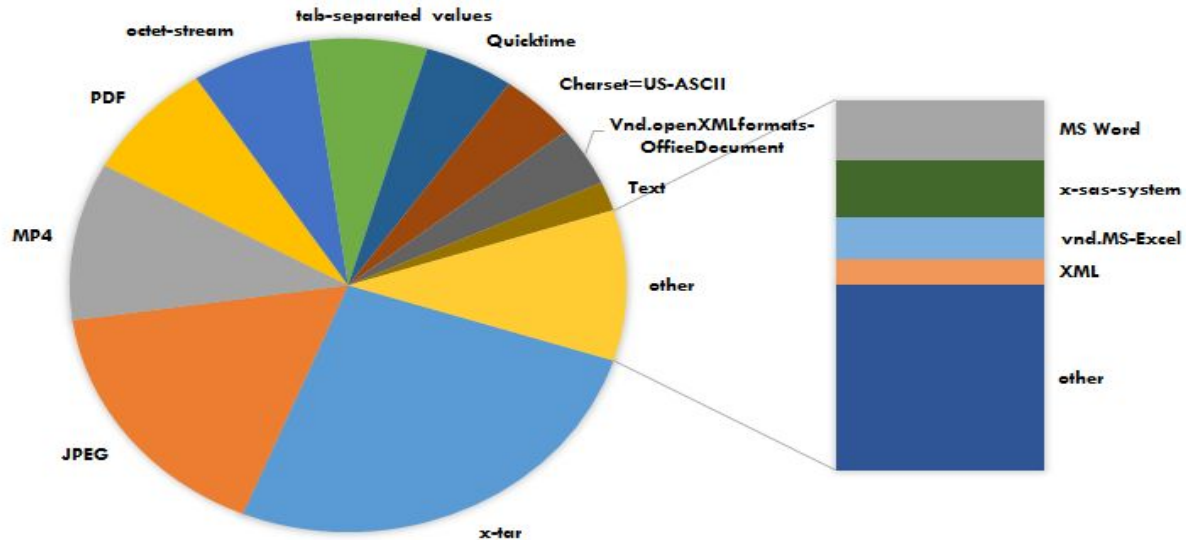
# UToronto Dataverse

## Dataset Downloads by Year



# UToronto Dataverse

## Data Types





# Questions?

Dataverse at UTL: [dataverse.library.utoronto.ca](https://dataverse.library.utoronto.ca)

Scholars Portal team: [dataverse@scholarsportal.info](mailto:dataverse@scholarsportal.info)