



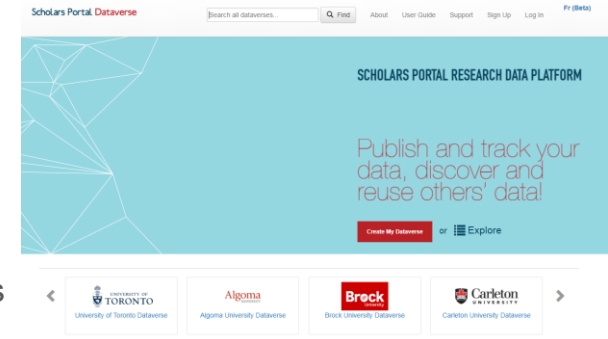
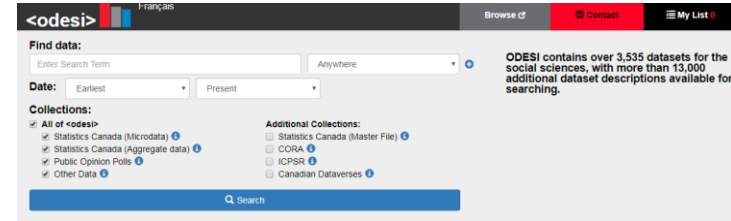
# A Metadata-Driven Data Curation Tool

By: Amber Leahey & Kevin Worthington, Scholars Portal



# Background

- Scholars Portal is a digital library project of the Ontario Council of University Libraries (OCUL), consortium of 21 university libraries
- In 2008, SP launched ODESI a social science survey data repository <http://odesi.ca>
  - Built with the Nesstar repository
  - DDI-XML metadata store (variable-level) in MarkLogic
  - Collection curated by OCUL libraries
- In 2012, SP hosts Dataverse for Ontario libraries and researchers
  - Open-source repository for research data
  - DDI is supported; not full DDI
  - Not as much curation support
- Would like to bring curation + functionality to Dataverse for research data!





# Data Curation 101

[The University of Illinois' Graduate School of Library and Information Science](#) defines data curation as

“the active and ongoing management of data through its life cycle of interest and usefulness to scholarship, science, and education. Data curation activities **enable data discovery** and **retrieval**, **maintain its quality**, **add value**, and **provide for reuse** over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation.”

In OCUL libraries -- data curation involves managing data with a view to reuse, and often can focus on documentation, metadata enhancement, and quality review, therefore it makes sense to develop tools to assist researchers, data curators, and librarians, with these tasks.



# Who are the data curators?

- Librarians
- Library support staff
- Student support staff
- Researchers
- Research managers
- Research assistants
- Data analysts

Really isn't one job title for data curators!

## PREQUESTION TEXT

And now, how do you feel about the political parties? Using a scale from 0 to 100, zero means you REALLY DISLIKE the party and one hundred means you REALLY LIKE the party.



# Data Documentation Initiative (DDI)

- Internationally recognized metadata standard for describing social science and behavioural research data
- Heavily used by data archives (ICPSR, GESIS, UKDA, CESSDA)
- Increasingly used by researchers / large research groups for improved metadata capture and reuse (MIDUS, SOEP, CLOSER)
- XML and RDF representations
- Implemented and supported by a number of repository tools (Nesstar, Dataverse, Colectica)

### Mode of Data Collection

Online survey

### Kind of Data

Survey data

### Unit of Analysis

Individuals

### Geographic Unit

Province

### Time Method

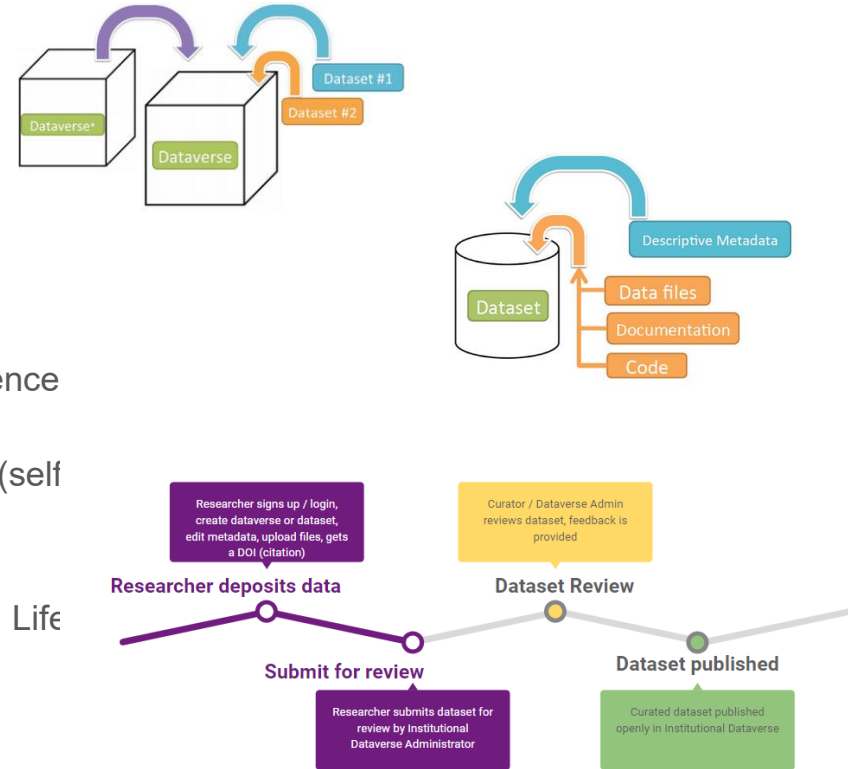
Cross section

## LITERAL QUESTION

On election day, are you certain to vote, likely, unlikely, or certain not to vote?

# Dataverse

- Open-source repository developed at Harvard University's Institute for Quantitative Social Science (IQSS)
- Offers flexible data deposit and access models (self deposit, mediated, curated, restricted)
- Supports general and disciplinary metadata standards (Dublin Core, Social Science, Health, Life Science, Astronomy)
- API-driven
- Modular application support





# Dataverse APIs

- APIs are sets of defined protocols and terms for programatically getting two systems to talk to one another
  - Useful for automating workflows (such as data upload / ingest)
  - Useful for developing applications (such as new discovery tools that rely on data / metadata retrieval from one system to another)
- SWORD API (upload)
- Data Search API
- Data Access API
- OAI-PMH (open archives standard for metadata harvesting)

# Modular Applications / Integrations

- Modular configuration of the Dataverse code + APIs for data access allow for really awesome new and custom developments that integrate with DV !

At SP we've been able to develop:

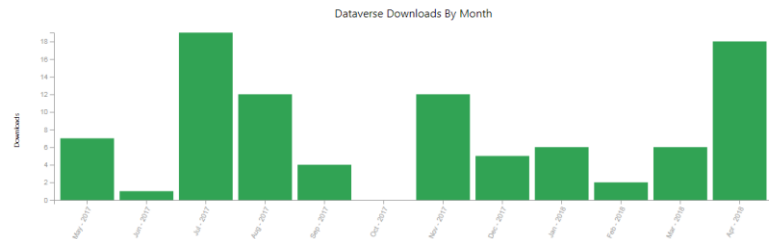
- Dataverse-specific metrics tool
- Data Explorer
- Data Curation Tool

## Dataverse Report

McMaster University Dataverse

May - 2017 to Apr - 2018

Download



<https://dataverse.scholarsportal.info/metrics/?selection=>

Other integrations: OSF, Archivematica, TwoRavens, WorldMap, OJS, and others...

 archivematica®

  
OPEN JOURNAL SYSTEMS





# Data Explorer

- Framework: AngularJS
- UI: Bootstrap
- Chart: Google Charts API
- Dataverse API: Data Access

## Forum Research Political Poll - Federal Issues (Canada) 2016

Forum Research Inc., 2016, "Forum Research Political Poll - Federal Issues (Canada) 2016", <https://doi.org/10.5683/SP/X8COC1>, Scholars Portal Dataverse, V4, UNF:6:MAyAB1dfIAYDMjNibPi0kQ==

Q

×

33 Results

Download

↓

ID	Name	Label	C
<div><div>=</div><div>140349</div><div>II</div></div>	Q1A	If a federal election were held today, which party are you most likely to vote for?	7
<div><div>=</div><div>140353</div><div>II</div></div>	Q1B	Even though you may not have made up your mind, which party are you leaning towards at this time?	7
<div><div>=</div><div>140343</div><div>II</div></div>	Q2	Do you approve or disapprove of the job Justin Trudeau is doing as Prime Minister?	3
<div><div>=</div><div>140346</div><div>II</div></div>	Q3	Do you approve or disapprove of the job Rona Ambrose (RON-na) is doing as interim leader of the opposition?	3
<div><div>=</div><div>140336</div><div>II</div></div>	Q4	Do you approve or disapprove of the job Tom Mulcair is doing as interim leader of the NDP?	3
<div><div>=</div><div>140335</div><div>II</div></div>	Q5	Regardless of which party you plan to vote for, who do you think would make the best Prime Minister?	7
<div><div>=</div><div>140330</div><div>II</div></div>	Q6	Do you approve or disapprove of Canada's anti-terrorism legislation,	3

First

«

1

2

3

4

»

Last

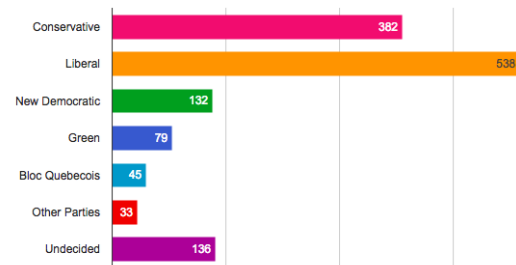
Records Per Page

10

Chart View

Table View

### Variable Q1A: If a federal election were held today, which party are you most likely to vote for?



Values	Categories	N
1	Conservative	382

# Data Curation Tool (Prototype)

- Framework:  
Angular 5
- UI: Google Angular  
Material
- Chart: d3
- Dataverse API:  
Data Access

Forum Research Political Poll - Federal Issues (Canada) 2015						
Add Group	Search					
All	ID	Name	Label	Weight	View	
Demographics	⋮ + v134374	Q1A	1a. If a federal election were held today, which party are you most likely to vote for?		○	✎
	⋮ + v134345	Q1B	1b. Even though you may not have made up your mind, which party are you leaning towards at this time?		○	✎
	⋮ + v134615	Q2	2. Do you approve or disapprove of the job Stephen Harper is doing as Prime Minister?		○	✎
	⋮ + v134478	Q3	3. Do you approve or disapprove of the job Tom Mulcair is doing as Leader of the Opposition?		○	✎
	⋮ + v134600	Q4	4. Do you approve or disapprove of the job Justin Trudeau is doing as leader of the Liberal Party?		○	✎
	⋮ + v134634	Q5	5. Which party do you expect to win the next federal election?		○	✎
	⋮ + v134338	Q6	6. Do you agree or disagree prostitution should be legal for sex trade workers and their clients?		○	✎
	⋮ + v134522	Q7	7. Parliament recently passed Bill C36 which makes the advertisement and purchase of sexual services illegal. While performing sexual services would still be legal for sex workers, they are prohibited from charging for their services, advertising or emplo		○	✎
	⋮ + v134310	Q8	8. The government has pledged \$20 million dollars to encourage sex workers to leave the sex trade and earn their living in another manner. Do you approve or disapprove of this?		○	✎
	⋮ + v134373	Q9	9. Do you agree or disagree \$20 million dollars is an adequate sum to encourage sex workers to leave the sex trade?		○	✎




# View Frequency Metadata

Add Group

All

Demographics

liberal

ID	Name	Label	Weight	View	
<input type="checkbox"/>					
<input type="checkbox"/>	v134466	Q17	17. Is Canada becoming more conservative or is it becoming more liberal?		
<input type="checkbox"/>	v134600	Q4	4. Do you approve or disapprove of the job Justin Trudeau is doing as leader of the Liberal Party?		

Items per page: 10 1 - 2 of 2

Q17: 17. Is Canada becoming more conservative or is it becoming more liberal?



Values	Categories	N	NW
1	More conservative	669	
2	More liberal	433	
3	Neither	456	
4	Don't know	183	

# Group Variables

## Forum Research Political Poll - Federal Issues (Canada) 2015

Add Group



Search



All

Demographics



	ID	Name	Label	Weight	View	
⋮	- v134374	Q1A	1a. If a federal election were held today, which party are you most likely to vote for?		👁	✎
⋮	- v134345	Q1B	1b. Even though you may not have made up your mind, which party are you leaning towards at this time?		👁	✎
⋮	- v134615	Q2	2. Do you approve or disapprove of the job Stephen Harper is doing as Prime Minister?		👁	✎
⋮	- v134478	Q3	3. Do you approve or disapprove of the job Tom Mulcair is doing as Leader of the Opposition?		👁	✎
⋮	- v134634	Q5	5. Which party do you expect to win the next federal election?		👁	✎
⋮	+ v134600	Q4	4. Do you approve or disapprove of the job Justin Trudeau is doing as leader of the Liberal Party?		👁	✎
⋮	+ v134338	Q6	6. Do you agree or disagree prostitution should be legal for sex trade workers and their clients?		👁	✎

per 1000,

# Select Weight Variable(s)

All

Demographics

☐

ID

☐

v134542

☐

v134351

☐

v134575

☐

v134576

☐

v134565

☐

v134436

☐

v134441

☐

v134566

☐

v134333

Variable Information

ID

v134351

Name

D2

Label

D2. How old are you?

Literal Question

Interviewer Instructions

Universe

Notes

Group

Weight Variable

Is Weight

☐

Weight

View

Items per page: 10

31 - 39 of 39

<

>

# Select/Edit Multiple Variables

Forum Research Political Poll - Federal Issues (Canada) 2015

 Save

Add Group



Search


























Add Selected to Group



All

Demographics



	ID	Name	Label	Weight	View	
	v134374	Q1A	1a. If a federal election were held today, which party are you most likely to vote for?			
	v134345	Q1B	1b. Even though you may not have made up your mind, which party are you leaning towards at this time?			
	v134615	Q2	2. Do you approve or disapprove of the job Stephen Harper is doing as Prime Minister?			
	v134478	Q3	3. Do you approve or disapprove of the job Tom Mulcair is doing as Leader of the Opposition?			
	v134600	Q4	4. Do you approve or disapprove of the job Justin Trudeau is doing as leader of the Liberal Party?			
	v134634	Q5	5. Which party do you expect to win the next federal election?			
	v134338	Q6	6. Do you agree or disagree prostitution should be legal for sex trade workers and their clients?			

# Search and reorder variables

Forum Research Political Poll - Federal Issues (Canada) 2015

Save

Add Group

≡+

Search

liberal

Q

All

Demographics

<input type="checkbox"/>	ID	Name	Label	Weight	View	
<input type="checkbox"/>	v134486	Q17	17. Is Canada becoming more conservative or is it becoming more liberal?			
<input type="checkbox"/>	v134600	Q4	4. Do you approve or disapprove of the job Justin Trudeau is doing as leader of the Liberal Party?			

Items per page: 10 1 - 2 of 2 < >







# Curation / Metadata Editing Interface

<https://scholarsportal.github.io/Dataverse-Data-Curation-Tool/?uri=https://dataverse.scholarsportal.info/api/access/datafile/40620/metadata/ddi>

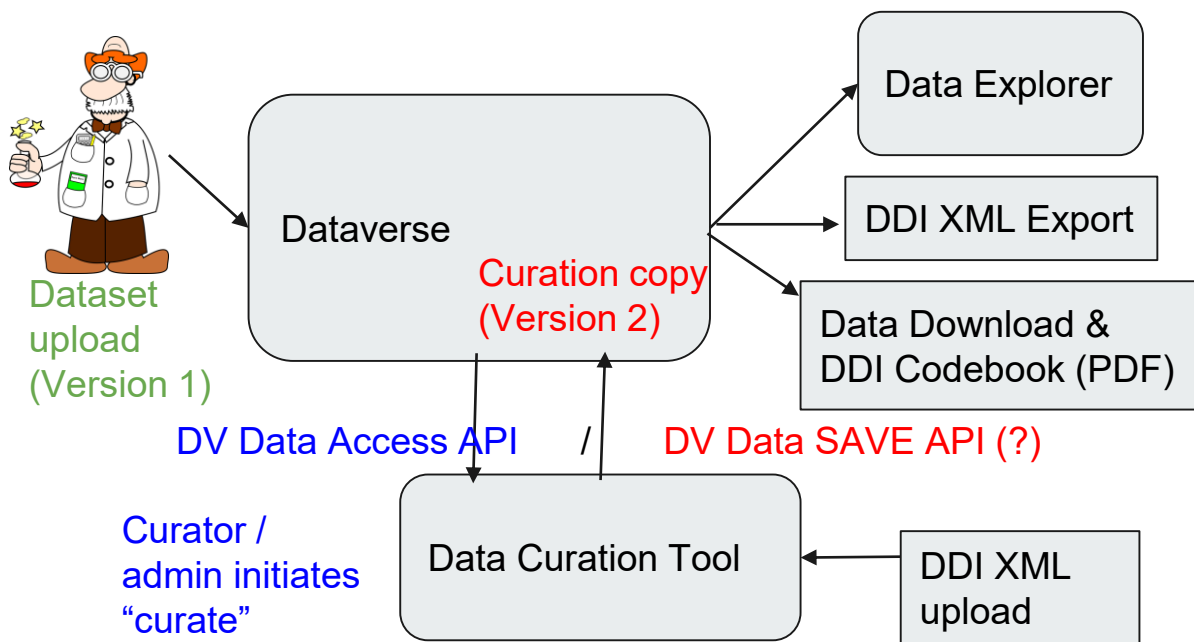
DEMO time!



# Discovery & Reuse Use Cases

- View question text, universe statements, coding instructions, questionnaire skip logic, etc. in context of variables
- Apply / tag weight(s) variables in a dataset (used for analysis)
- Group variables to make it easier to find a use large datasets with many variables
- Search on richer sets of metadata; e.g. question text provides original research question vs. label for variable
- Metadata is structured in DDI standard and can be exported for reuse in other data curation projects

# Possible curation system architecture





# Next Steps

- Connect to Dataverse DDI metadata saving API
- Develop phase two enhancements
  - Multilingual support
  - DDI xml upload / reuse
  - Additional DDI fields?
  - Enhancements to Data Explorer to support discovery and reuse

Conduct UX testing!!



Questions?

[amber.leahey@utoronto.ca](mailto:amber.leahey@utoronto.ca)