Introducing the Data Curation Tool

The Scholars Portal Dataverse team has been hard at work on the new Dataverse Data Curation Tool as part of our Canarie RDM grant project. Development on this project is being led by Victoria Lubitch, Programmer/Analyst at Scholars Portal.

The Data Curation Tool (DCT) allows data owners and curators to create and edit variable-level metadata for any tabular file in a dataset. Users can access this tool as a modular application once they've uploaded a tabular file (e.g., SPSS, R, Excel, CSV) to a dataset in Dataverse.

The Data Curation Tool

Similar to tools like SPSS, the DCT allows users to view summary statistics about their data, add variable information like 'Interviewer Instructions' or 'Notes', create variable groups, and indicate weighting variables.

Q1A: 1a. If a federal election were held today, which party are you most likely to vote for?
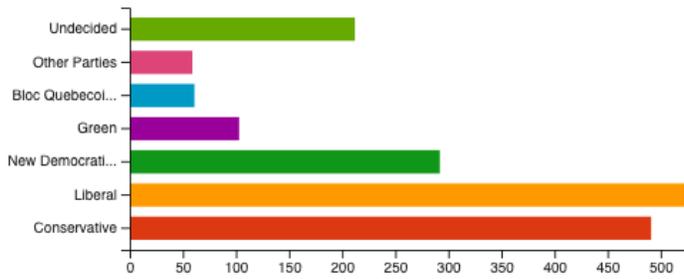
Values	Categories	Count	Count Percentage(%)	Weighted Count
1	Conservative	491	28.202	
2	Liberal	523	30.04	
3	New Democratic	292	16.772	
4	Green	103	5.916	
5	Bloc Quebecois	61	3.504	
6	Other Parties	59	3.389	
7	Undecided	212	12.177	

Summary statistics in the DCT

Variable editor in the DCT

Once edits have been completed and saved back to Dataverse, these changes can then be downloaded as an XML file or exported to a codebook.

Variable Description

Variable Groups

- [Number of hours](#)

Number of hours

Variables within Number of hours

- [Weeks absent from work](#)
- [Usual hours per week at main job](#)
- [Actual hours per week at main job](#)
- [Usual hours per week at all jobs](#)
- [Actual hours per week at all jobs](#)
- [Hours away from work, employees, part-week only](#)
- [Paid overtime hours in ref. week](#)
- [Unpaid overtime hours in ref. week](#)
- [Number of overtime or extra hours worked,](#)
- [Usual hourly wages, employees only](#)

Variables

Order of record in file

f1469 Location:

Summary Statistics: Min. 1.0; Max. 101887.0; StDev 29412.387775674837; Mean 50944.0; Valid 101887.0

Variable Format: numeric

Notes: UNF:6:cp7T3PSRxL5jaCcrZIwbeA==

Survey year

f1469 Location:

Summary Statistics: StDev 0.0; Min. 2016.0; Mean 2016.0; Valid 101887.0; Max. 2016.0;

Variable Format: numeric

Notes: UNF:6:+iqbSJOmZQUnkEFuZxYDUA==

Notes: Processing

Survey month

f1469 Location:

Summary Statistics: StDev 0.0; Valid 101887.0; Max. 1.0; Mean 1.0; Min. 1.0

Variable Format: numeric

Notes: UNF:6:teqYV6FrUBhyz4Dy6AAa5Q==

Example of a codebook in Dataverse

Usability testing sessions were recently completed with 5 participants, who worked through a series of tasks and helped us identify areas where the user experience could be improved in the tool. We're now working on translating this tool to be used in French, with translations provided by the University of Ottawa.

A demo of this tool is available online, and the code can be accessed on GitHub. The Data Curation Tool will be launched with the next Scholars Portal Dataverse upgrade, currently scheduled for the end of October, and will be available for community testing soon.

If you have any comments or suggestions, contact us at dataverse@scholarsportal.info. If you would like to see all the updates and have a SpotDocs account, click the "Watch this blog" button on the top right corner of the page to receive notifications.