

# DATAVERSE AT UBC LIBRARY

» Eugene Barsky and Paul Lesack, UBC Library

» November 2021



# OVERVIEW

- » Research data - Scholars Portal Dataverse
- » Licensed data - Abacus Dataverse
- » dryad2dataverse platform

# RESEARCH DATA

- » UBC Library started to work on RDM in 2014
- » First deposits were fully curated

# RESEARCH DATA

Contact Share

Search this dataverse...

Find

Advanced Search

 **Dataverses (0)**

 **Datasets (10)**

 **Files (0)**

## Publication Year

2019 (10)

## Author Name

**Ward, W. Peter (10) X**

Gagné, Monique (9)

Jahanfar, Shayesteh (1)

Lesack, Paul (1)

## Subject

Medicine, Health and Life Sciences (10)

## Keyword Term

Health (10)

Economic and Financial Statistics (9)

Country::World (4)

Country::United States (3)

Country::Canada (2)

## Deposit Date

2019 (10)

Author Name: **Ward, W. Peter X**

1 to 10 of 10 Results

Sort

### Bologna Sant'Orsola Maternity Clinic Database, 1880-1940

Mar 4, 2019 - UBC Research Data Collection

Ward, W. Peter; Jahanfar, Shayesteh; Lesack, Paul, 2019, "Bologna Sant'Orsola Maternity Clinic Database, 1880-1940", <https://doi.org/10.5683/SP2/N9LLCF>, Scholars Portal Dataverse, V1

The variables contained in the data sets are primarily concerned with perinatal outcomes and maternal health. The types of fields are wide-ranging but are primarily concerned with infant birth, maternal health, and socioeconomic status.

### Birth weight and economic growth data sets, Allgemeines Krankenhaus, Vienna, 1865-1930, [2012]

Mar 4, 2019 - UBC Research Data Collection

Gagné, Monique; Ward, W. Peter, 2019, "Birth weight and economic growth data sets, Allgemeines Krankenhaus, Vienna, 1865-1930, [2012]", <https://doi.org/10.5683/SP2/Z7NRRK>, Scholars Portal Dataverse, V1, UNF:6:ILNttT4n+7rw7LXSumLoSg== [fileUNF]

The variables contained in the data sets are primarily concerned with perinatal outcomes and maternal health. A number of variables with respect to the social and economic status of the mothers and their families were also included (ie. Occupation, Marital status, Region). While...

### Birth weight and economic growth data sets, Utrecht Hospital, 1880-1940, [2012]

Mar 4, 2019 - UBC Research Data Collection

Gagné, Monique; Ward, W. Peter, 2019, "Birth weight and economic growth data sets, Utrecht Hospital, 1880-1940, [2012]", <https://doi.org/10.5683/SP2/WNY6FG>, Scholars Portal Dataverse, V1




The variables contained in the data sets are primarily concerned with perinatal outcomes and maternal health. A number of variables with respect to the social and economic status of the mothers and their families were also included (ie. Occupation, Marital status, Region). While...


# RESEARCH DATA

- » Hired a PhD student to do data curation
- » Took a lot of time and resources, data dictionaries, recoding
- » Data management vs. data curation

# RESEARCH DATA

1 to 6 of 6 Files

<input type="checkbox"/>	 <p><a href="#">S_Orsola.tab</a> Data/ Tab-Delimited - 5.4 MB - Mar 4, 2019 - 42 Downloads MD5: a1e6c2b7e22c679bc0f4b16fca77b661 Microdata - Sant'Orsola Maternity Clinic Database</p> <p><b>Data</b></p>	<input type="button" value="Download"/> <input type="button" value="Grid"/>
<input type="checkbox"/>	 <p><a href="#">S_Orsola_Codesheet_1_Abbrev_ITA.pdf</a> Documentation/ Adobe PDF - 60.1 KB - Mar 4, 2019 - 9 Downloads MD5: 088b34ccfb70cd4d5b358d4912423538 List of abbreviations</p> <p><b>Documentation</b></p>	<input type="button" value="Download"/> <input type="button" value="Grid"/>
<input type="checkbox"/>	 <p><a href="#">S_Orsola_combined_codesheets.xls</a> Documentation/ MS Excel Spreadsheet - 113.0 KB - Mar 4, 2019 - 4 Downloads MD5: 498cf15607f0753b42bbee58bba0a57f Health and place name codes</p> <p><b>Documentation</b></p>	<input type="button" value="Download"/>



# RESEARCH DATA

- » Fast forward six years
- » Mediated deposit
  - » Users can login and deposit data
  - » RDM Librarian / Analyst review deposits and consult with users to improve
  - » Data is published

# RESEARCH DATA

## » Curation

» Every dataset needs attention

» Common issues:

» Lack of data dictionaries

» Corrupt data files

» Insufficient metadata

» No geospatial metadata



# RESEARCH DATA

» Data Curation, Level 2:

“Activities that enhance the discoverability of datasets and help ensure their usability over time. Eg. recommended metadata fields are populated and dataset includes sufficient documentation to allow a user with a similar background to understand the dataset and open and use the files.”

# RESEARCH DATA

Hi

I apologize for the delay in the publication of your material,

I had a look at the study which you submitted and I have some suggestions which I hope you will consider.

1. A data dictionary for your CSV file outlining what each variable represents would be very helpful for researchers using your data. Normally this information is included in a file traditionally called README. More information on files of this type is available here:

[https://researchdata-06oct2014.sites.olt.ubc.ca/files/2020/04/QuickGuide\\_UBC\\_readme\\_v1.0\\_20200427.pdf](https://researchdata-06oct2014.sites.olt.ubc.ca/files/2020/04/QuickGuide_UBC_readme_v1.0_20200427.pdf)

2 As well, your data set is listed as a CSV, but the separator for values is a semi-colon, which makes it not really a csv file (ie, comma-separated value). As a result, the automated conversion to a .tab file (really just a TSV) has failed because the processor is expecting a comma as a separator. You can see the result of this by viewing your data set and selecting “Explore”. You will see that it doesn’t load correctly as tabular data. You should consider recreating the file as a csv or tsv (tab-separated) and re-uploading it.

An example of a metadata enhancement request

# RESEARCH DATA

- » We do allow (and encourage) some research groups to create and manage their own Dataverses
- » Examples:
  - » Housing Research Collaborative (20 datasets)
  - » Centre for Sustainable Food Systems (22 datasets)
  - » Dr. Paul Pavlidis Dataverse (36 datasets)
  - » Many more...

**CDTP**

**dryad2dataverse**

# DRYAD2DATAVERSE

- » A platform agnostic tool which harvests material from Dryad
- » Can be automated if required
- » Requires zero researcher effort
- » Associated python package can be used for other metadata conversions

# DRYAD2DATAVERSE

- » Metadata from Dryad is preserved as much as possible
- » DOIs from Dryad are written into Dataverse records
- » JSON metadata from Dryad is automatically inserted into a Dataverse record by default.

# DRYAD2DATAVERSE

- » No coding
- » No dedicated server
- » Minimal platform requirements
- » Easy installation
  - » `pip install git+https://github.com/ubc-library-rc/dryad2dataverse.git@master`

# DRYAD2DATAVERSE

## » Code

» <https://github.com/ubc-library-rc/dryad2dataverse>

## » Documentation

» <https://ubc-library-rc.github.io/dryad2dataverse/>





# LICENSED DATA

- » UBC hosts a Dataverse installation which holds licensed (ie, commercial) data sets
- » Multi-university consortium
- » Common feel between two separate Dataverse installation platforms

# LICENSED DATA

- » Like Scholars Portal, users authenticate via Shibboleth
- » Data sets that are not the product of the university research
- » Data sets are organized by **license**
- » Participating institutions have their own dataverses

# LICENSED DATA

- » End user experience is nearly the same as for Scholars Portal
- » Most of the tool installed are the same
- » API access is available just as for research data

# QUESTIONS?

- » Eugene Barsky , Head, University of British Columbia Library Research Commons, [eugene.barsky@ubc.ca](mailto:eugene.barsky@ubc.ca)
- » Paul Lesack, University of British Columbia Library Research Commons, [paul.lesack@ubc.ca](mailto:paul.lesack@ubc.ca)