

# **Queen's University Library**

## **Research Data Management (RDM) Workflow**

Alexandra Cooper  
Jeff Moon  
Data Services, Open Scholarship Services  
Queen's University Library  
February 2018

## Table of Contents

RDM Planning.....	2
Data Ingest .....	2
Data Deposit Form .....	3
Data Processing.....	3
Completed Studies .....	4
Embargoed Data .....	5
Data Publishing .....	5
Appendix 1 – Queen’s University Library – OCUL License and Addendum .....	7
Appendix 2 – Scholars Portal Dataverse 4.x Guide .....	8
Appendix 3 – Creating Codebooks .....	13
Appendix 4 – Publishing Studies with Sensitive Data .....	15
Figure 1. QUL Data Services Disclosure Control Flowchart .....	17
Figure 2. DataFirst Disclosure Control Flowchart .....	18

## RDM Planning

Initiate contact and conversations with researchers about research data management. Arrange a meeting with the researcher(s). Meeting is best done in person, with both the Data Librarian and Data Services Coordinator in attendance. The goal is to get a sense of what data is available, the nature of the data, and any other information that might influence how we proceed with data management.

After the meeting, a link to the online Data Deposit form (<http://library.queensu.ca/data/data-deposit>) is sent to researcher – this form includes a request for metadata as well as consent to archive and disseminate data and documentation.

## Data Ingest

Receive data, documentation and Data Deposit Form. Update the data management spreadsheet in the DGIC/RDM folder (restricted folder) in QShare (DataManagement-database.xlsx). When data are published, this file is updated again with links to the archive(s) the data have been published in.

During this process, via the web-based deposit form, we receive notification from the researchers about:

- Embargo dates, if applicable
- Level of release permission (e.g. World, Canada, Ontario, Queen's, Restricted...)
- Basic metadata in XML format
- Permission to disseminate the data

Data and Documentation received is added to the DGIC/RDM folder in QShare.

- name of the folder follows the following format –  
*QULyear-sequential#-name-of-dataset*  
e.g. QUL2014-008-drone-survey
- if there is an embargo, add the date the embargos end –  
*QULyear-sequential#-name-of-dataset-embargoed-date-embargo-ends*  
e.g. QUL2014-008-drone-survey-embargoed-2016-12-31
- when the study is completed and published, the folder name is edited to include “-completed” at the end; if the embargo is still in place, the file name is edited to include “-completed-embargoed-2016-12-31”  
e.g. QUL2014-008-drone-survey-completed-embargoed-2016-12-31

In this folder, create three folders

- original – all original data and documentation goes here; **do not make any changes to these files**
  - include in this file the email, text, and xml files from the Data Deposit form
- working – all files that are being worked on go here
  - make sure that if there are multiple versions of a file that they are labeled correctly
  - if there a master file and public file are created, make sure to label the datasets properly; if needed, create master and public folders within the working folder

- master files are not published but will be made available if permission is received from the author
- final – all files that will be published
  - only add files here when the dataset and documentation files are complete and ready to be published
  - include non-proprietary file formats (csv, ascii and command code)
  - only add files that are to be published
  - include all Master files also, even if they are not being published. If they are requested, this is the folder where they should be pulled from.
- update the RDM LibGuide (<http://guides.library.queensu.ca/qul-research-data-archive>) with the new study; mark as “In progress” until the study is published

## Data Deposit Form

Researchers fill out the Data Deposit Form ([library.queensu.ca/webdoc/ssdc/data-deposit](http://library.queensu.ca/webdoc/ssdc/data-deposit)) which, in addition to the metadata, gives Queen’s the right to archive and disseminate the survey and related documentation. Dissemination of the survey is based on the level of access (Queen’s, Ontario, Canada, World) chosen by the researcher, and whether an embargo is placed on the data.

The Data Deposit Form is sent to the Data Services Coordinator. Data files and documentation are emailed separately to the Data Librarian and/or the Data Services Coordinator.

When the Data Deposit form is received:

1. copy the metadata information in the email into a text file and save
2. save this file, along with the email, in QShare in the *original* folder of the appropriate survey folder
3. input the metadata into Nesstar Publisher
4. complete rest of metadata, save as .Nesstar file and export the new .xml file in the working folder
  - when the .sav file is ready to be imported into Publisher it will be merged with this file

## Data Processing

### Assess data and documentation

- **Triage:** After a quick initial assessment, prioritize data project(s) based on factors such as: complexity of the data file(s), completeness of documentation, availability of researcher(s) to answer questions, embargo period declared, stage of the research life cycle, etc.
- **Review:** Check data carefully for problems such as:
  - disclosure risk
  - invalid variable labels
  - undeclared missing values
  - miscoding
  - alpha codes, etc.
- **Document:**
  - Gather, organize, and enrich documentation, create User Guide and Codebook

- User Guide – links to documents, etc. in the User Guide should point to the Dataverse record for the study
- Create a DDI-compliant XML file using Nesstar Publisher (or any XML editor)  
Creation of metadata such as the User Guides and Codebooks happens independently of the creation of the DDI compliant XML. It is important to ensure that metadata in the User Guides and Codebooks match that provided in the XML file. When the User Guide and Codebook are sent to the researcher for review, any changes to will have to be reflected in the XML file.
- **Copyright:** Vet related publications for Copyright – are articles open access or does copyright need to be obtained?
- **Follow-up** with researchers as needed
- **Studies with sensitive data** (i.e. data cannot be released) – see Appendix 5

## Completed Studies

Once the data files and documentation are compiled and ready to be published, send the files to the researcher to review for corrections/comments. The following files are generally sent:

- dataset file – both master and public file
- user guide, codebook, questionnaire, any other documentation created

In the email to the researcher include the following information

- list all files attached and what they are
- changes made to the original dataset – what variables have been suppressed, regrouped, derived variables created, any other new variables added
- explain that Master files will not be published but available if requested and user has researchers permission to use the files
- confirm embargo date (if there is one)

Make modifications and send documentation for final review. Once the researcher approves the documentation files, publish the study to its appropriate archive(s). If there is an embargo date, follow the steps below under “Embargoed Data”.

- If publishing to <odesi>
  - fill out the Addendum for the *Queen’s University Library – OCUL License* and sent to <odesi> before publishing; see Appendix 1 for information on the Addendum
  - tell <odesi> level of access for the study  
This is set manually by the <odesi> staff and can take 24-48 hours before it is set; if the access level is set to Canada or the world, do not announce the study until the access has been tested.
- update the RDM folder in QShare
  - update the data management spreadsheet – in the status column mark as “completed”
  - changed the dataset folder name adding “-completed” to the end of the folder name
  - the final versions of the files are copied to the “final” folder in the study folder
- update the RDM LibGuide (<http://guides.library.queensu.ca/qul-research-data-archive>) with links to where the study was published

- if the study has an embargo, note when the study will be available  
e.g. **Embargoed:** data and documentation will be available in January 2016

**File types** – files are published in the following formats

- if data is received in a proprietary format, convert it to a non-proprietary format if possible.
- Data files are saved in SPSS and ASCII+command code. If data is in Excel format, it will also be saved as a CSV. Assess and deal with other formats as they arrive.
- Documentation is saved in PDF format.

## Embargoed Data

If a study has an embargo, follow the steps above in **Completed Studies** but do not publish the files.

- add all the files to be published to the final folder in the study folder
- edit study folder name to include “-completed-embargoed-date-embargo-ends”  
e.g. QUL2013-002-epidmeology of pain-completed-embargoed-2016-01-01
- update the data management spreadsheet – in the status column mark as “completed-embargoed”

## Data Publishing

Data may be published to multiple locations depending on the level of access:

- **QSpace** – <http://qspace.library.queensu.ca/>
- **Dataverse** - QUL Research Data Archive Dataverse  
<http://dataverse.scholarsportal.info/dvn/dv/qulrda>

Main repository for data files and documentation; if QUL hosts a local website, link all files to the Dataverse record

- In theory, all data can go into Dataverse since they can be locked down at any level
- Restrictions
  - Open
  - Fully locked down – without permission, only the abstract is available
  - Individual files locked down – users can see metadata files, but data files are restricted (restriction can be applied to one or multiple files)
  - IP-based Access – only those coming from authorized IP addresses or domains have access to restricted surveys and files. For example, only users coming from Queen’s IP addresses (i.e. on campus or proxied in).
  - restricted to individuals, or groups of individuals
    - for access to files, user must have a Dataverse account and the userid and password for the restricted files
    - groups are created by the Dataverse Administrator at Scholars Portal; once created, people can be added
- DOIs – Dataverse automatically assigns DOIs to datasets once they are created.
  - If a researcher needs a DOI to include in an article submission before publication, create the dataset record but leave unpublished. The DOI is available but the link will not work

until the dataset is published. Once the embargo is over, publish the dataset and the DOI link will work.

- **<odesi>** – <http://odesi.ca>
  - All microdata files can go into <odesi> if they are open to the OCUL level and above (OCUL, Canada, World)
  - Restrictions
    - Open to OCUL member schools
    - Open to Canadian Universities and Colleges subscribing to <odesi> (includes OCUL member schools)
    - Open to the world
    - OCUL License Addendum must be updated and emailed to <odesi> (see Appendix 1)
    - <odesi> must be informed about the level of access for the survey (they set this manually and takes 24-48 hours to be set; default is OCUL member schools)
- **Website** – if appropriate
  - If a website is created, all data and documentation are linked to their record in Dataverse
- **Discipline-based Archive**
  - If there is an appropriate discipline based archive, datasets will go here first. Depending on the dataset, consideration will then be given to adding it to the above archives or adding a metadata record to Dataverse or QSpace with a link to the archive housing the data.

## Appendix 1 – Queen’s University Library – OCUL License and Addendum

This license is between Queen’s University Library and OCUL and allows Queen’s to deposit local researcher data into <odesi>. The license states that Queen’s holds the dissemination rights to deposit the data. The license was signed by the Queen’s University Vice-Provost and University Librarian, Executive Director of OCUL, and the Director of Scholars Portal in May 2013.

### Addendum

The license includes an Addendum which must be updated before a survey is added to <odesi>. Once the Addendum is updated and emailed to <odesi>, the survey can be added. <odesi> needs to also be informed of what the level of access is for the survey – OCUL subscribing institutions or world.

The Addendum lists all surveys that Queen’s has permission to add to <odesi>. Revise the date at the top of the Addendum and add the information below to the Addendum. Do not remove the surveys previously added.

Collection	Title of Survey
Date Added to Addendum	Date added to <odesi>
Metadata Access	Open OR Restricted to <odesi> subscribers only
Data Access	Open OR Restricted to <odesi> subscribers only
Terms and Conditions for Access and Local Loading	QUL holds exclusive dissemination rights. For data restricted to ODESI subscribers, requests from non-subscribers should be forwarded to QUL to ensure the requests are valid under QUL agreement with the DATA PRODUCER.
Term of Contract	Perpetual

The Addendum file is kept in the DGIC/RDM folder in QShare (queens-odesi-addendum-*monthyear*.doc; add the month and year that the new files are added - queens-odesi-addendum-march2015.doc).

The Addendum and level of access is emailed to the the Data Services Metadata Librarian at Scholars Portal.

If the survey is part of a collection of surveys, i.e. Angus Reid Global, then the Addendum does not have to be filled out for each survey. When the survey is added to <odesi> inform the Data Services Metadata Librarian at Scholars Portal that the survey is part of a collection already in the Addendum.



## Appendix 2 – Scholars Portal Dataverse 4.x Guide

Scholars Portal LibGuide - <http://guides.scholarsportal.info/dataverse/>

### Add new dataset

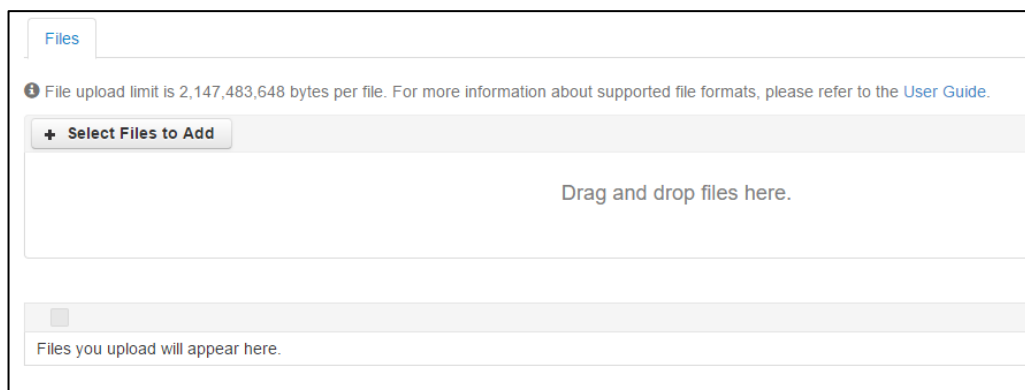
- 1) Log in to Dataverse
- 2) Go to Dataverse that dataset is to be added to
- 3) Click on **Add Data** dropdown (top right of screen)
- 4) Click on **New Dataset**
  - a) Choose template (if available)
  - b) Fill in citation metadata (basic metadata; the rest of the metadata will be added in the next step)
    - \*Asterisks indicate a required file
    - Author and Contact always default to logged in name
    - Description – this is the abstract

### Notes

- Only the fields with asterisks have to be filled in to create a dataset.
- If there is a plus (+) beside a field, additional lines can be added
- Limited HTML characters are supported in most fields
- once the dataset is created, you will receive an email from Dataverse that your dataset is created
- a dataset will remain *Unpublished* (not public) and in *Draft* mode until it is **Published**

### 5) Add data files

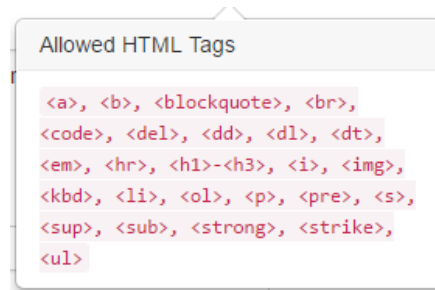
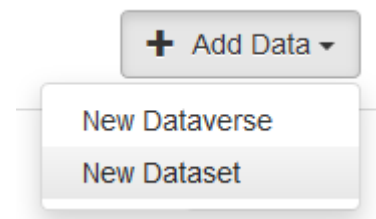
- a) Click on **+Upload Files**
- b) Drag and drop the files or click on **+Select Files to Add**
  - Restricted tabular data only (CSV, XLSX, SAV (SPSS), RData (R), or Stata)
    - Zip up data file and then zip the zipped file. This is done because Dataverse, upon ingest of the zipped file, will automatically unzip a file.
    - See “**Doubling zipping**” restricted data files section for more information



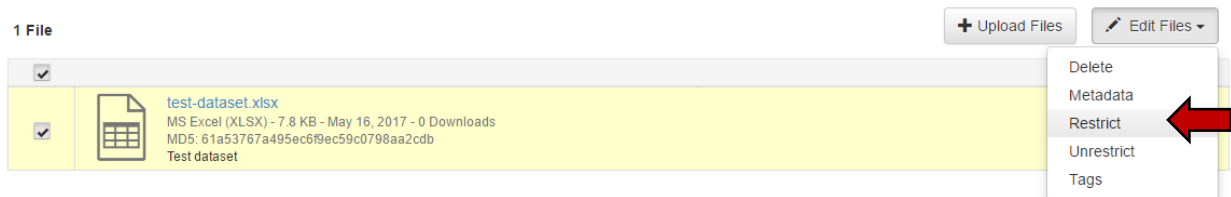
- c) Click on **Save Dataset**

### Notes

- Maximum file size for upload is 2GB



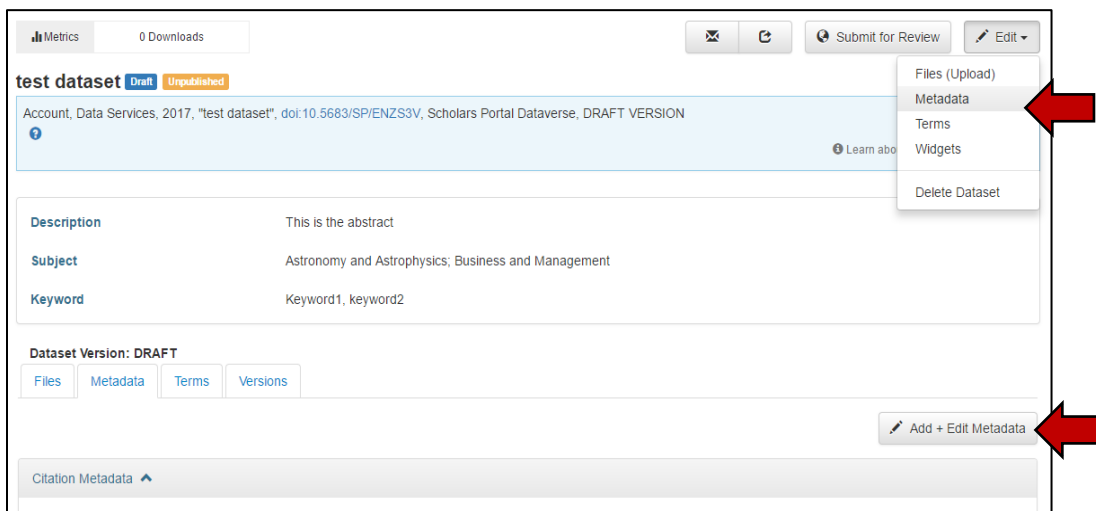
- 6) Restrict data files
  - a) Select files to be restricted
  - b) Click on **Edit Files**
  - c) Click on **Restrict**
  - d) **File Restrictions** box opens
    - Enter **Terms of Access** if applicable
    - Click on Enable access request
    - Click on **Continue**



## Editing datasets

### Metadata

Open dataset and **Edit** → **Metadata** at the top of the page or  
Click on the **Metadata** tab → **Add + edit Metadata**



### Fill in fields

- Metadata can be added for the following metadata standards
  - Geospatial
  - Social Science and Humanities
  - Astronomy and Astrophysics
  - Life Sciences
  - Journal

Click on **Save Changes**

## Adding new Data files

Click on **+Upload Files**

- 1) Drag and drop the files or click on **+Select Files to Add** and upload
  - Restricted tabular data only (CSV, XLSX, SAV (SPSS), RData (R), or Stata)
    - Zip up data file and then zip the zipped file. This is done because Dataverse, upon ingest of the zipped file, will automatically unzip a file.
    - See **“Doubling zipping” restricted data files** section for more information
- 2) add Description and tags
- 3) Click on **Save Dataset**

## Editing Data files

- 1) Choose the files to be edited
- 2) Click the **Edit files** button and choose an option



- 3) Click on **Save Dataset**

## Terms of Use

Terms of use, citation requirements, restrictions, terms of access, etc. can all be set for a dataset. The default is CC0 – “Public Domain Dedication”

## Publish/Submit for Review

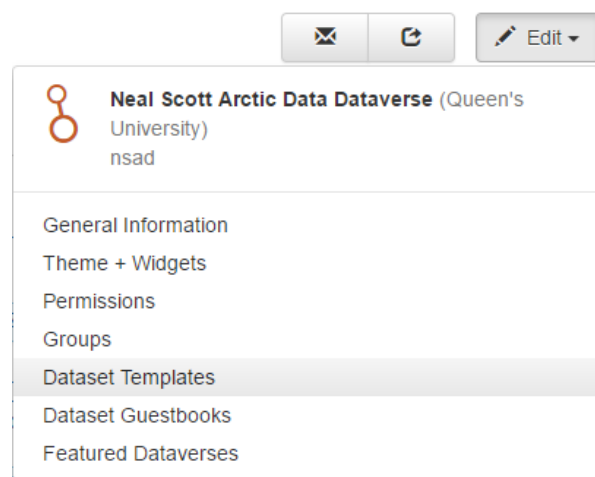
Once the dataset is ready to be made public, click on **Publish** or **Submit for Review** at the top of the screen (button available will depend on permissions)

- **Submit for Review** – sends an email to Administrator to review dataset. Administrator can either Publish the dataset or Return to Author.

## Dataset templates

Have to be an Administrator to add Templates

1. Click on **Edit** and choose **Dataset Templates**
2. Click on **Create Dataset Template**
3. Give the template a name and fill in the fields
4. **Save and Add Terms** – adds Terms of Use and Terms of Access
5. **Save Dataset Template**



## Setting permissions

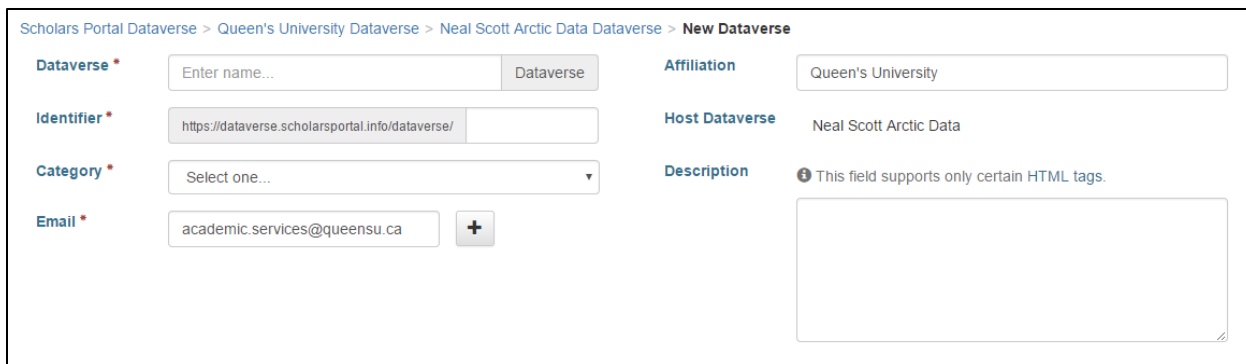
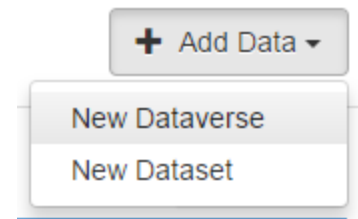
Have to have Administrator permissions for this

## Create a dataverse

Open the Dataverse that new datavers is to be added to

Fill in the fields

- Dataverse – name of your Dataverse
- Identifier – short name used for URL of this dataverse
- Email – email address for the contacts of the dataverse; click on plus (+) to add more emails



Scholars Portal Dataverse > Queen's University Dataverse > Neal Scott Arctic Data Dataverse > New Dataverse

**Dataverse \*** Enter name... Dataverse

**Identifier \*** https://dataverse.scholarsportal.info/dataverse/

**Category \*** Select one...

**Email \*** academic.services@queensu.ca +

**Affiliation** Queen's University

**Host Dataverse** Neal Scott Arctic Data

**Description** ⓘ This field supports only certain HTML tags.

Metadata Fields and Browse/Search Facets sections – leave as set by default

Click on Create Dataverse

## DOIs

A digital object identifier is created for each dataset created. Once the dataset is saved, the DOI is created. When the dataset is published, the DOI becomes a permanent link to the dataset.

## Versions

Once a dataset is published, any changes made will result in a new version.

- Small changes to the metadata will result in minor version – V1 becomes V1.1
- File changes, citation changes, or big metadata changes result in a major version change – V1 becomes V2
- All changes to the dataset are listed and can be viewed from the **Versions** tab.

## **“Doubling zipping” restricted data files**

Used for data files that are to be restricted and uploaded to Dataverse in a tabular data format such as CSV, XLSX, SAV (SPSS), RData (R), or Stata.

Upon ingest of a file in any of these formats, Dataverse automatically creates file level and variable level metadata (DDI Codebook 2.5 - Sections 3 and 4). Since you cannot restrict access at the dataset or metadata level, any user is able to view and download the metadata. If a user downloads the metadata for a dataset, they have access to all the name, label and descriptive statistics for each variable in the restricted file. Many researchers do not want this summary information available.

To get around file/variable level metadata being created for the restricted data files, upload the files within a zipped uncompressed archive folder. Since Dataverse will automatically unzip an archive folder upon upload, double zip the archive folder (create a zip archive of the original zip archive). When you ingest the double zipped folder, it is extracted and you're left with the original zip archive.

[from email received from Carrie Breton, University of Guelph, May 2, 2017]

## **Additional guides**

Havard Dataverse Guide - <https://dataverse.scholarsportal.info/guides/en/4.6.1/user/>

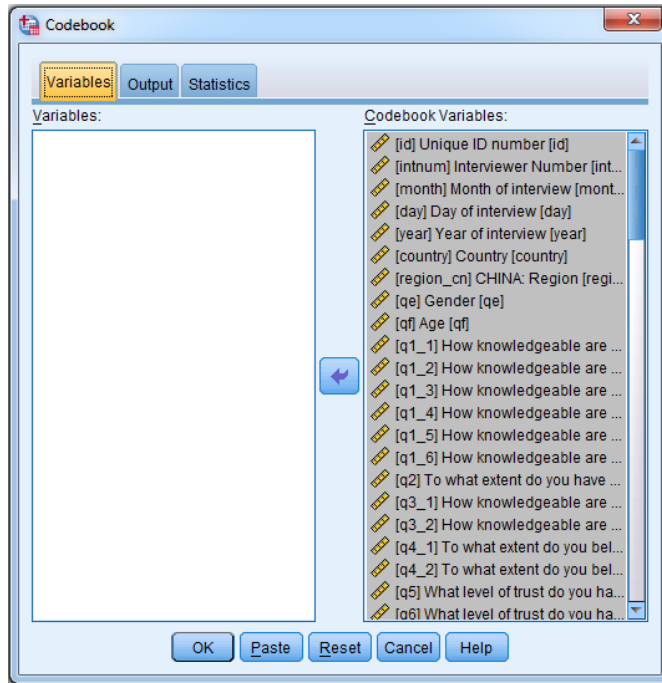
Scholars Portal LibGuide - <http://guides.scholarsportal.info/dataverse/>

## Appendix 3 – Creating Codebooks

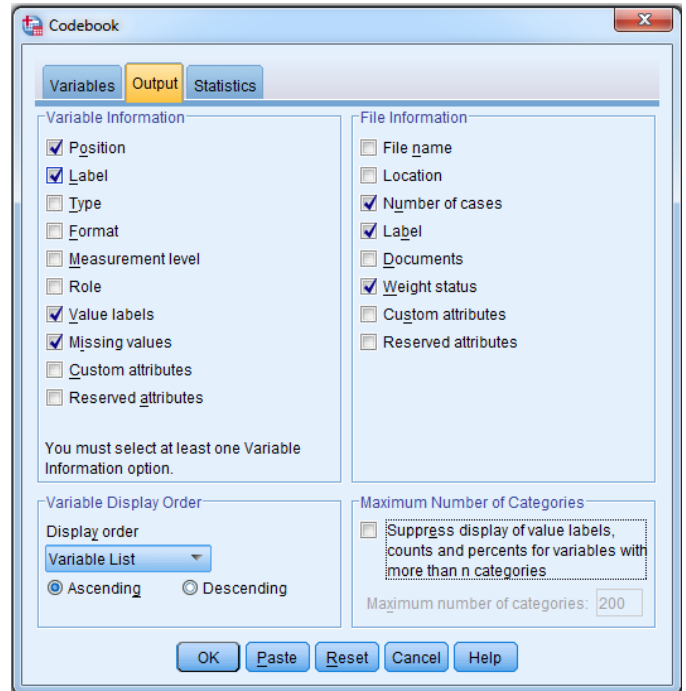
In SPSS open dataset

- Analyze → Reports → Codebook...

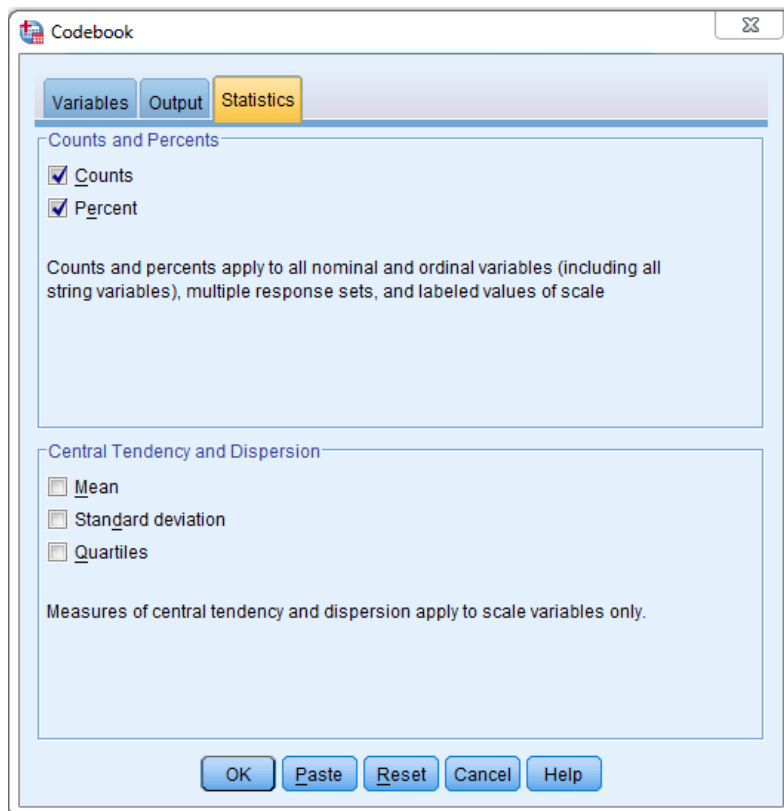
Choose the variables



Choose the output



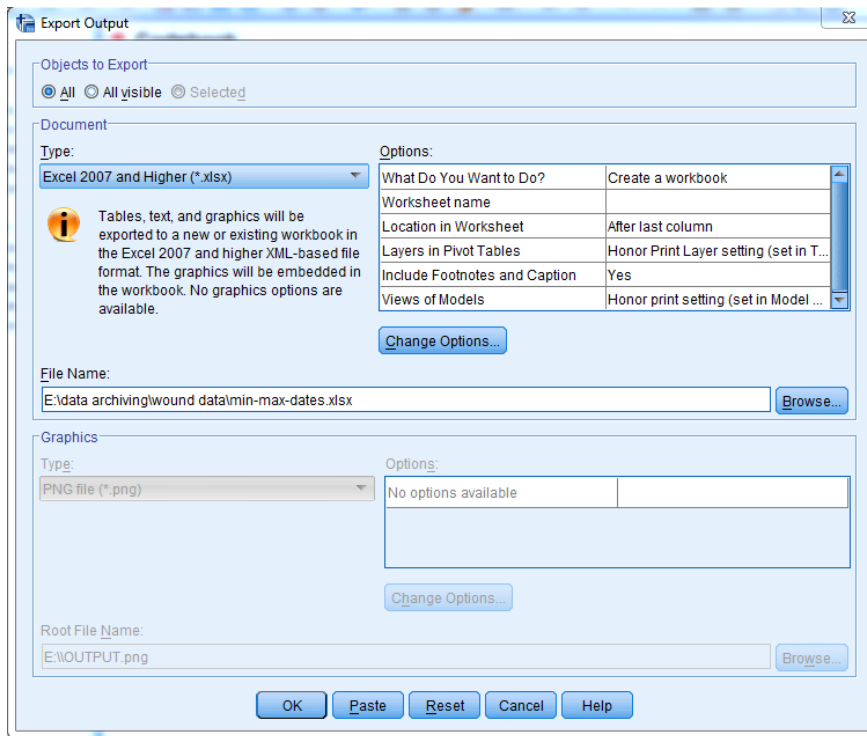
Choose the statistics



## Export to Excel

With the codebook open in the output window.

- File → Export
- File Name – browse to where the file is to be saved (this will default to last file saved)
- OK



## Export to Word

- If your columns are too wide in Excel the tables won't display properly in word. I have found these to be the best sizes:
  - Column A – 10.71
  - Column B – 18.71
  - Column C – 50.29
  - Column D, E – 8.71
- Left-align column C
- Select the whole codebook
  - Format → AutoFit Row Height – makes sure all the rows are the same size
  - Wrap text
- In word, set all margins to 0.5 inches
- Select all in excel and paste into word
- Add section headings to match the variable grouping in the documentation

## Appendix 4 – Publishing Studies with Sensitive Data

Sensitive data are data that contain identifying information (names, emails, detailed geography, direct identifiers, etc.). Such files, which risk respondent privacy, directly or indirectly, cannot be made public. These files will be kept as restricted files, with access only granted with permission. See Figure 1 Disclosure Control Flowchart for an example of the decision process in releasing data.

When the data is received and it is determined that it is sensitive, put all datasets in a folder named *sensitive-data-do-not-release* within the *original* folder.

- **Review:** Check data carefully for problems such as:

- invalid variable labels
- undeclared missing values
- miscoding
- alpha codes, etc.

While the data will not be released, they need to be as complete as possible.

- **Document:**

- Gather, organize, and enrich documentation, create User Guide
- Create a Codebook if appropriate and it can be created without disclosing any data
- User Guide – links to documents, etc. in the User Guide should point to the Dataverse record for the study
- Create a DDI-compliant XML file using Nesstar Publisher  
Creation of metadata such as the User Guides and Codebooks happens independently of the creation of the DDI compliant XML. It is important to ensure that metadata in the User Guides and Codebooks match that provided in the XML file. When the User Guide and Codebook are sent to the researcher for review, any changes to will have to be reflected in the XML file.

- **Copyright:** Vet related publications for Copyright – are articles open access or does copyright need to be obtained?
- **Follow-up** with researchers as needed

Once the documentation is compiled and ready to be published, send the documentation files to the researcher to review for corrections/comments. Make modifications and send documentation for final review. Once the researcher approves the documentation files, publish the study (documentation only) to the appropriate archive(s). If there is an embargo date, follow the steps below under “Embargoed Data – Sensitive Data”.

- if the data cannot be released, the documentation alone will be published to one or more repositories (e.g. Dataverse, QSpace and/or discipline specific archives)
- update the RDM folder in QShare
  - update the data management spreadsheet – in the status column mark as “completed”
  - change the dataset folder name adding “-completed” to the end of the folder name
  - copy the final version of the files to the “final” folder in the study folder



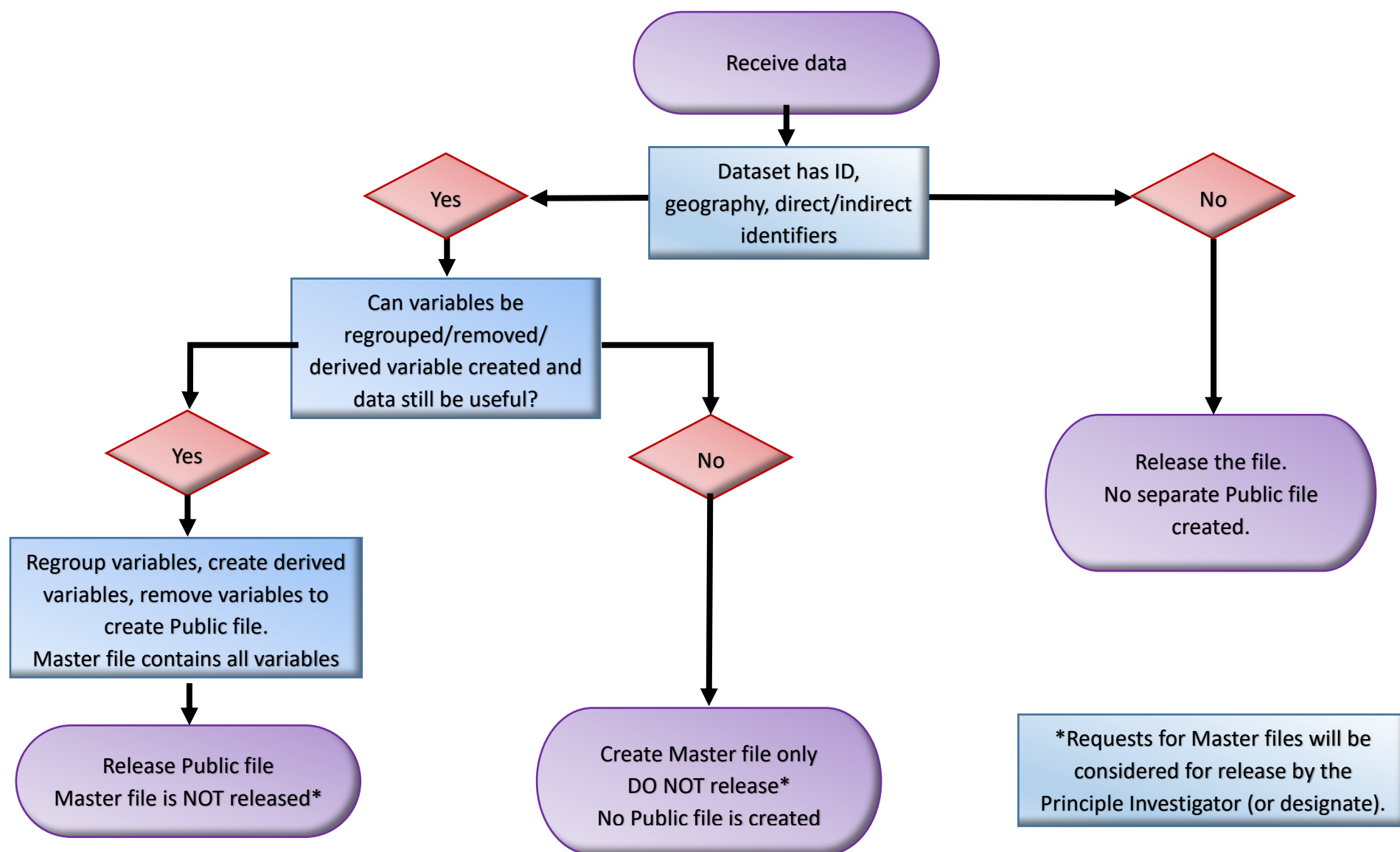
- update the RDM LibGuide (<http://guides.library.queensu.ca/qul-research-data-archive>) with links to where the study was published
  - if the study has an embargo, note when the study will be available  
e.g. **Embargoed:** data and documentation will be available in January 2016

### **Embargoed Data – Sensitive Data**

If a study has an embargo, follow the above steps above but do not publish the files.

- add all files to be eventually published to the “final” folder in the study folder, including the data files which should be put in a subfolder named *sensitive-data-do-not-release*.
- edit study folder name to include “-completed-embargoed-date-embargo-ends”  
e.g. QUL2013-002-epidmeology of pain-completed-embargoed-2016-01-25
- update the data management spreadsheet – in the status column mark as “completed-embargoed”

Figure 1. QUL Data Services Disclosure Control Flowchart<sup>1</sup>



<sup>1</sup> Modelled on O'Rourke, JoAnne. 2012. *ICPSR Disclosure Risk Decision Tree and The five essential steps for risk reduction*. See Figure 2. February 2018

Figure 2. DataFirst Disclosure Control Flowchart

