

# RESEARCH DATA MANAGEMENT (RDM)

Michelle Edwards & Carol Perry

## Starting Your Research on the Right Foot Part 2





# Objectives

- Understand the components of research data stewardship
- Apply concepts to your research project

# WHAT IS RESEARCH DATA MANAGEMENT?

- Management of research data?
  - Management of data?
- We collect data, save it on our computers, analyze it using some software, write-up our results, and hopefully publish
- What do we need to manage????

---

# Research Life Cycle - Analyzing data



# VARIABLE NAMES INSIDE MY FILES

- Information or data that we are collecting:
  - Price of fibres
  - Feed consumption
  - Length of fibre
  - Quality of fibres
- Often used as headings in our Excel file
- Move these to our Stats packages – confusion and errors set in!

# BEST PRACTICES FOR VARIABLE NAMES

- Survey of commonly used Statistical packages to review variable naming practices.
  - SAS, SPSS, Stata, R, and Matlab
- List of best practices for variable names

# VARIABLE NAME RESTRICTIONS AND LIMITS

## Length of Variable Name

- SAS: 32 characters long
- Stata: 32 characters long
- Matlab: 32 characters long
- SPSS: 64 bytes long
  - 64 characters in English
  - 32 characters in Chinese
- R: 10,000 characters long

## 1<sup>st</sup> Character of Variable Name

- SAS: MUST be a letter or an underscore
- STAT: MUST be a letter or an underscore
- Matlab: MUST be a letter
- SPSS: MUST be a letter, an underscore or @, #, \$
- R: No restrictions found

# VARIABLE NAME RESTRICTIONS AND LIMITS

## Special Characters in Variable Names

- SAS: NONE
- Stata: NONE
- Matlab: No restrictions found
- SPSS: NONE except Period, @
- R: NONE except Period

## Case in Variable Names

- SAS: Mixed case – Presentation only
- Stata: Mixed case – Presentation only
- Matlab: Case sensitive
- SPSS: Mixed case – Presentation only
- R: Mixed case – Presentation only

NO BLANKS (SPACES) allowed in any of the Statistical Packages

Beware of Function names in all Statistical Packages – these cannot be used as Variable Names



# BEST PRACTICES FOR VARIABLE NAMES

1. Set Maximum length to 32 characters
2. ALWAYS start variable names with a letter
3. Numbers can be used anywhere in the variable name AFTER the first character
4. ONLY use underscores “\_” in a variable name
5. Do NOT use blanks or spaces
6. Use lowercase

# VARIABLE NAMES INSIDE MY FILES

- Information or data that we are collecting:
  - Diet A → `diet_a`
  - Fibre length in centimetres → `fibre_cm`
  - Location of farm → `location`
  - Price paid for fleece → `price`



# Exercise – Variables

# DMP CHECKLIST

- √ Organizing the data you've collected
  - √ Documenting your work
  - √ Managing your files – processing and analyzing your data
- Storing, backing up, and securing your files
- Preserving your data
- Accessing, sharing, and reusing your data

---

# Take a Break



# STORING and BACKING UP DATA – 3-2-1 Rule

Keep at least **three** copies of your data

Store the copies on **two** different media

(Department server, external hard drive, USB, etc.)

Keep **one** backup copy offsite

Keep a ‘master file’ – original untouched – for emergencies

U of G Research Data Classification – CCS

<https://www.uoguelph.ca/ccs/infosec/rdc>

# STORING and BACKING UP DATA

Backup all files on regular basis

Keep backup copy in separate location

Create 'master files' for raw data files & important document files

Store in physically separate, secure location

If working files are lost – make a copy from the 'master files'

Synchronize files on regular basis

Encrypt sensitive data – see CCS services

<https://www.uoguelph.ca/ccs/encryption>

# STORING and BACKING UP DATA

## ‘Master files’

- Secure location

- Separate folder with ‘master file’ in name

- Could include original documents, raw data files, final output files

## ‘Working files’

- Based on original master files

- May have multiple versions as you run tests, edit content, etc.



# STORING and BACKING UP DATA – File Versioning

AFS\_Data\_201806\_Suri.xls


- If you edit the file ...
  - Do you change the name? Do you include date of change?
    - V1.0, V1.1,
    - V1, V2, V3
- Be consistent
- Be clear
- Document changes made in a readme file or other note

# STORING and BACKING UP YOUR DATA

- Every few minutes – SAVE the file you are working on
- Synchronize your files on a regular basis
  - To avoid multiple versions of same file (changed in one place but not updated in other copies)
  - Departmental servers automatically back up files on a regular basis
    - Check frequency!
- Purpose of a backup?
  - You will always have a copy to go back to if files are messed up/lost

# SECURING YOUR DATA

- What kind of data are you collecting?
- Is there any sensitive data?
  - Direct identifiers – names, SIN, Registration #s, ...
- Who can access your data? Who should have access to your data?
  - Can we limit access? How?
  - See CCS Information Security Policies  
[https://www.uoguelph.ca/ccs/infosec/policies\\_and\\_procedures](https://www.uoguelph.ca/ccs/infosec/policies_and_procedures)



# Exercise - security

# SECURING YOUR DATA

- Login passwords are NO longer enough!!
  - Strong passwords – hackers can find a way through
- Encryption
  - Takes your information and turns it into a stream of what appears to be random symbols
  - You need a digital key to unlock – without the key the data is unusable
  - [Information and Privacy Commissioner of Ontario](#)
  - CCS offers [encryption services](#)

# SECURING YOUR DATA - WHEN THINGS GO WRONG

- Parks Canada bans wildlife photographers from using radio receivers to locate animals
  - Photographers were disturbing the natural environment of bears, elk, and wolves
  - New fines up to \$25,000
- Scientific data used to track and protect animals is vulnerable to hacking
  - Bison reintroduction project in Banff – 5 are equipped with GPS collars or VHF radio collars
- Troubling issues at the frontier of animal tracking for conservation and management
  - Tracking animals for poaching

# RESEARCH DATA MANAGEMENT

- Managing our research data resources
- Enabling us to keep our project data organized and well-documented
- Recommending best practices for file and variable naming conventions
- 3-2-1 Storage and Backup
- Preserving data for future research use
- Creating a Data Management Plan to guide us through RDM

# Contact

- Michelle Edwards
- [edwardsm@uoguelph.ca](mailto:edwardsm@uoguelph.ca)
  
- Carol Perry
- [carolp@uoguelph.ca](mailto:carolp@uoguelph.ca)
- [lib.research@uoguelph.ca](mailto:lib.research@uoguelph.ca)



By Funkipickle Retrieved from: <https://www.flickr.com/photos/funkipickle/6137260892/>  
Used under CC BY-NC-ND 2.0