Managing and sharing research data: an RDM training course

23 October 2019

Instructor: Nicholas Syrotiuk



http://doi.org/dcz7

Session plan

What is data management planning? 40:00 [1] Managing active data 10:00 [2] Sharing & preserving data 15:00 [3]

Re-using research data 5:00 [4]

What is data?

One definition: Anything which can be used to validate or replicate a research conclusion, or enrich understanding of the research process



2019/01/28: Reward for stolen Royal Oak wreck data laptop

BBC

Why bother with RDM?



2005/10/30: Fire destroys Southampton research centre





Estimated damage to Mountbatten Building: £50 million

CASH REWARD

for returning my lost backpack



- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is VERY important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on:

Thank you!!

years of research data lost

Photograph by P. Murray-Rust is licenced under CC-BY.

"Availability of research data declines rapidly with article age"--Vines et al.



Image created by N Syrotiuk from the orignal research data and R script.

padlet



Managing and sharing research data

Interactive session

What might be considered good RDM practice? What might be considered poor RDM practice?



https://padlet.com/sefnyn/rdm



Part one: WHAT IS DATA MANAGEMENT PLANNING?

It is a process or activity





DMP document: a communication tool



Objectives

Maximise RE-USE potential of data

Keep sensitive data SAFE



TYPE IN CONTRACTOR

THE DEPTH OF A DEPTH O

A DECK

den miller fan it staal het it de kerken aan de kerken de staal de staar de staar de staar de staar de staar de

A CONTRACTOR OF A CONTRACTOR OF

THERE IS NOT A CONTRACTOR AND A MUNICIPAL WITH A REPORT OF A DESCRIPTION OF A DESCRIPTION OF A DESCRIPTION OF A

finden ale meit fertig immelit mit i Mingenfit meiter ihr meite bilt ficher mit will seine von

THE A REAL THE COMPANY AND A REAL ADDR. ADDR

na the claim for a function of theme is a start addition of the same a contract for \$7.500 for a start as a start as

A STATUS AND A S

||福口||12月1||

. Name a final color of a final color of the state of the

In the second second

THE REPORT OF STREET, S AND DE LA MARINE DE LA MANAGERE

THE TOTAL STOCK





2. **IDENTIFY** the data to be collected

Types of data

October

Methods of data collection

Volume of data in bytes

File formats: open preferred

3. State how data will be **ORGANISED**

4. Explain how data will be **DOCUMENTED**

2001(1)

data QUALITY will be assured

6. Provide a solid STRATEGY for data storage and preservation

7. Define the project's data POLICIES

Data (and software) licencing

Plans for sharing. Embargo?

Management of sensitive data

8. Describe how data will be DISSEMINATED

WESTDEUT

STEE DUSSELDOR

Basler Zeitu A MELERAL RUHR TETTUS

Vogel Gryff begeis

9. Assign ROLES and RESPONSIBILITIES

10. Prepare a realistic B U D G E T

STO

Brookentreased Walter at Venas Brind

2.24 The section is equilated. Investme Investment Value a Year en 424 963 446 211 467 4 468 522 1 005 0 491 948 1 620 9 516 545 2 324 1 42 372 3 124 7 69 491 4 033 8 Start at monthly 5 063 6 can be do this R 35 4



funder guidelines.





Choose funder template Write detailed plan

3. Share plan

Project Title 🗧 🗧	Template 🗧	Organisation	Owner	Download
Measuring understanding of biological data models.	University of Manchester Generic Template	University of Manchester	Yo Yehudi	
Charters for Better Work Better Lives: An Indian Partner Network	University of Manchester Generic	University of Manchester	Wendy Olsen	A

https://dmponline.dcc.ac.uk

Identifying Mechanisms of Failure of Cartilage Repair Surgery	University of Manchester Generic Template	University of Manchester	Gwenllian Tawy	ß
Simulating tourism water consumption with stakeholders (SIMTWIST)	Data Management Plan NWO (English)	Wageningen University and Research Centre (Netherlands)	Bas Amelung	ß



RDM training



D M P R E V I E W











The Future:

machine-actionable DMP = maDMP



Dynamic DMPs



Interoperable DMPs: to exchange information between research systems

Persistent identifiers in DMPs



Funder IDs

Grant IDs

Org IDs

Repo IDs

DOIs

Public DMPs

Travel is to make a journey or to have an adventure to somewhere by bicycle. train. airplane, car, motorcycle. or boat. It could be an exploration to somewhere new planned or unplanned to meet new people. new things and new places. There are different types of adventures waiting for you to explore.

There are lots of places to explore. Places could be urban or suburban. Some people loves to be with nature to free their minds and refresh their souls, but some like to be in the city. You will get lots of benefits such as exploring new culture.

B TRANS CONTRACTOR

Part two: MANAGING ACTIVE DATA



Storage Options Tool

The tool on this page is designed to provide advice on the most appropriate storage for different use cases. This is not an exhaustive list and does not supersede systems such as DUO. There may also be a specific reason why the proposed solution may not be ideal in your particular situation. Chosing a non-recommended option is viable, but please think carefully when accepting an option marked as less suitable, especially if sensitive information is involved. Do you currently use solutions flagged as **avoid**?

If you require more detailed advice please contact the IT Service Desk.

오 🔰 Guidance

- Use this for advice on where to store different types of files
- Use the tool separately for each type of data you want to store
- Read each question and tick the boxes that apply to your files
- 4. Submit the form and you will be

Specify Storage Criteria

Type of Data

- My files are standard "offic
- My files are non-standard of
- My files are exceptionally la

Access

- I need high-performance a latency)
- I need to access my files wl

Storage

options

Sharing

My files need to be accessed by multiple people within the University

tool

I need to share my files with people outside the University

Purpose



 \bigotimes

 \otimes

 \otimes

0

0

 \bigotimes

0

0

0

Personal Research Storage

X

 \otimes

 \odot





B

Shared Research Storage

Storage Manager



My Access	Manage Shares	Audit Log
User Guide	Log out pzvx49	

Directories shared with you

Personal Read Write pzvx49/prs

CIFS (Windows): \\prsblue02.mds.ad.dur.ac.uk\pzvx49\prs

CIFS (Linux): //prsblue02.mds.ad.dur.ac.uk/pzvx49/prs

Linux Desktop Service: /prs/prsblue02/silver3_07/pzvx49/prs

Personal Read Write pzvx49/test01

CIFS (Windows): \\prsblue02.mds.ad.dur.ac.uk\pzvx49\test01

CIFS (Linux): //prsblue02.mds.ad.dur.ac.uk/pzvx49/test01

Linux Desktop Service: /prs/prsblue02/silver3_07/pzvx49/test01

Shares you manage

OneDrive for Business and Office 365 make it easy for you to share and work together on all your files.

OneDrive for Business



Durham University (DU) Box: Cloud storage for research

≡	box	Search files and folders				
-	All files 🕶					
_	Name		Updated 🗸	Size		>
	Research.Data	a	Mar 4, 2019 by NICH	2 Files		
9	LARGEFILES		Jan 15, 2019 by NICH	2 Files		
•	My Box Notes	i	Jan 15, 2019 by NICH	0 Files		
=/	research.data		Nov 28, 2018 by NIC	1 File		
-/	SefnynYahoo		Nov 20, 2018 by Sefn	0 Files		
7	top-level		Jul 17, 2018 by NICH	2 Files		
	DU Box Users	3	Jan 11, 2018 by Andre	1 File		









Part three: SHARING & PRESERVING DATA

Barriers to sharing

"69 % increase in citations" --Piwowar et al.

Benefits of sharing

Preparation and organisation

THIL

EDVOLT

Provide all necessary files

Ensure accessibility of files

Descriptive file names

Logical file structure



Data deposit



Illustration by Pampel, Vierkant, Scholze, Bertelmann, Kindling, Klump et al is licenced under CC BY SA. DOI: 10.1371/journal.pone.0078080 (pre-print).

Choosing a research data repository

Collections

https://collections.durham.ac.uk/



Enter search terms

.........

Q Go All -

About H

Help Contact

+) Login

Durham research data repository



Recently Uploaded

Depositor	File Details
N.G. Chancellor	<u>Finding spin-glass ground</u> <u>qwspinglass_data.tar.gz</u> <u>quantum computing, quantum walk, quantum</u> optimisation, quantum algorithms, spin glass

Tweets by @DurhamRdm

Durham RDM Retweeted

DurhamResearchOnline @DROdurham

Accepted manuscript now available on DRO: Kilby, Karen E. (2018) 'Julian of Norwich, Hans Urs von Balthasar, and the status of suffering in Christian theology.', New blackfriars., 99 (1081). pp. 298-311. dro.dur.ac.uk/22207/

ഒ

Data citation

D: VAL VA

n' an "In O Main har

9 ... 2X - 20

Format:

Creator (PublicationYear): Title. Version. Publisher. (resourceTypeGeneral). Persistent identifer Par Her 2 . B & million 10a 1 Sin 103

Example:

~

Breckon, T; Tiancheng, G (2018): Pretrained neural network models for Guo 2018 study, TensorFlow format. Durham University. (dataset). DOI: http://doi.org/10.15128/r23j333226h Mai the min . . C G. M

Source: Metadata Schema Documentation for the Publication and Citation of Research Data, version 4.1. DOI: <u>http://doi.org/10.5438/0014</u>

no me

Exceptions to sharing

Ethical reasons

Public safety reasons

Commercial reasons



Part four: RE-USING RESEARCH DATA

Data re-use #1





Data re-use #2

Combine data with other existing data: A + B = C



y

I love doing science in 2019. Saw this bioRxiv preprint via twitter. Downloaded the data, integrated it with a dataset we have. Discussed next steps with the lab on Slack. Experiments planned. All in a day, and the paper is unpublished. biorxiv.org/content/10.110...

Source: https://twitter.com/clathrin/status/1181596997272375297

Data re-use #3

Reproduce or replicate the results of a prior study



Life-saving guidance

DIA BASIC

DO'S AND DON'T

RDM rules of thumb



Online training

Research Data Management and Sharing MOOC (Massive Open Online Course): <u>https://www.coursera.org/learn/data-management</u>

110

< 11,

MANTRA: https://mantra.edina.ac.uk/

FOSTER e-learning platform https://www.fosteropenscience.eu/

Thank you

Nicholas Syrotiuk





References:

Vines, T. H., et al. (2014): "The availability of research data declines rapidly with article age," Current Biology 24(1): 94-97. DOI: <u>https://doi.org/10.1016/j.cub.2013.11.014</u>

Vines, T. H., et al. (2013): "The availability of research data declines rapidly with article age." Dryad Digital Repository. (dataset). DOI: <u>https://doi.org/10.5061/dryad.q3g37</u>

Michener, W. K. (2015): "Ten simple rules for creating a good data management plan," PLoS Computational Biology 11(10): e1004525. DOI: <u>https://doi.org/10.1371/journal.pcbi.1004525</u>

UK Research and Innovation (2018): "Guidance on best practice in the management of research data." UKRI web site.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007): "Sharing detailed research data is associated with increased citation rate," PLoS ONE 2(3): e308. DOI: <u>https://doi.org/10.1371/journal.pone.0000308</u>

Fear, K. M. (2013): Measuring and Anticipating the Impact of Data Reuse. Doctoral dissertation. University of Michigan. PID: <u>http://hdl.handle.net/2027.42/102481</u>