



Modern Research Data Management workshop

Sarah Jones

DCC, University of Glasgow

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC

Martin Donnelly

DCC, University of Edinburgh

martin.donnelly@ed.ac.uk

Twitter: @mkdDCC



Agenda

Time	Topic	Who
09:00-09:30	Intro to RDM	Sarah
09:30-09:45	Benefits and challenges of RDM	Martin
09:45-10:00	Dealing with sensitive data	Martin
10:00-10:30	Data sharing exercise	All
10:30-11:00	Coffee	All
11:00-11:30	Data Management Planning (including a demo of DMPonline)	Sarah
11:30-12:00	Exercise: writing a DMP	All
12:00-12:15	Other useful tools and resources	Martin
12:15-12:30	Questions and discussion	All





SOME DEFINITIONS

What is Research Data Management?

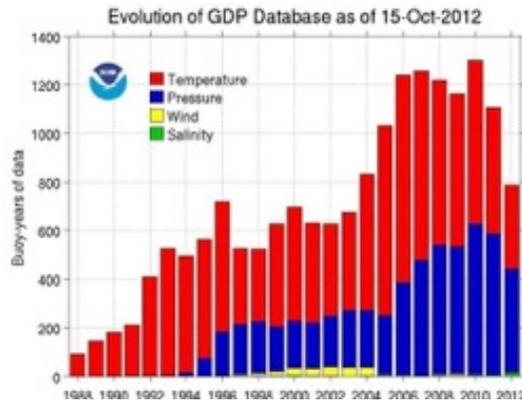
What is research data?

- Research data is defined as recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.
- Research data refers to **information**, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation.

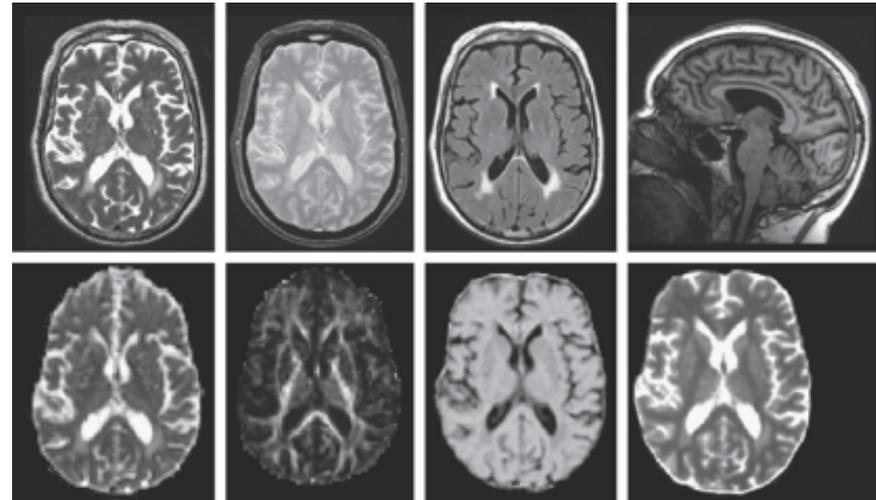
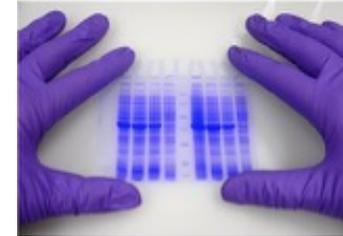
In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in **digital form**.



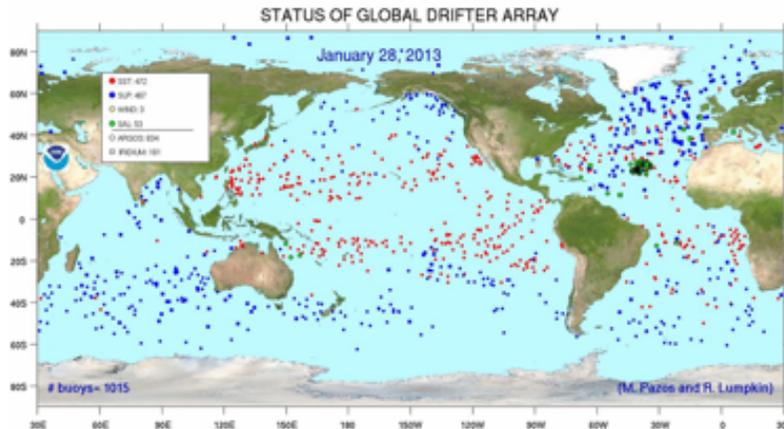
So, what might this include?



www.aoml.noaa.gov/phod/dac/array_grwth.html



www.sbirc.ed.ac.uk/documents/lbc_protocol.pdf



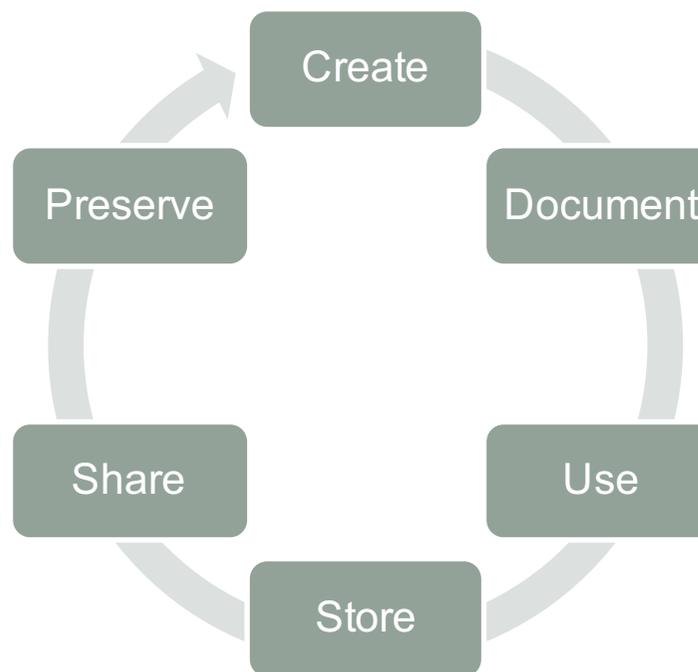
www.aoml.noaa.gov/phod/graphics/dacdata/globpop.gif

Anything & everything produced in the course of research

What is Research Data Management?

The active management of data throughout the lifecycle

- Data Management Planning
- Creating data
- Documenting data
- Accessing / using data
- Storage and backup
- Selecting what to keep
- Sharing data
- Data licensing and citation
- Preserving data



Why is RDM an issue?

- Digital technology now used very widely in research, and is enabling new research and scientific paradigms
- Research funders and publishers know that digital research data can be expensive to produce but inexpensive to share, making reuse more feasible and desirable
- The challenge is to ensure digital research findings can be reproduced and cited



Reasons to manage and share data

Direct benefits for you

- To make your research easier!
- Stop yourself drowning in irrelevant stuff
- Make sure you can understand and reuse your data again later
- Advance your career - data is growing in significance

Research integrity

- To avoid accusations of fraud or bad science
- Evidence findings and enable validation of research methods
- Meet codes of practice on research conduct
- Many research funders worldwide now require Data Management and Sharing Plans

Potential to share data

- So others can reuse and build on your data
- To gain credit - several studies have shown higher citation rates when data are shared
- For greater visibility, impact and new research collaborations
- Promote innovation and allow research in your field to advance faster

What if this was your laptop?

Why YOU need a Data Management Plan

<http://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan>



CASH REWARD
for returning my lost backpack

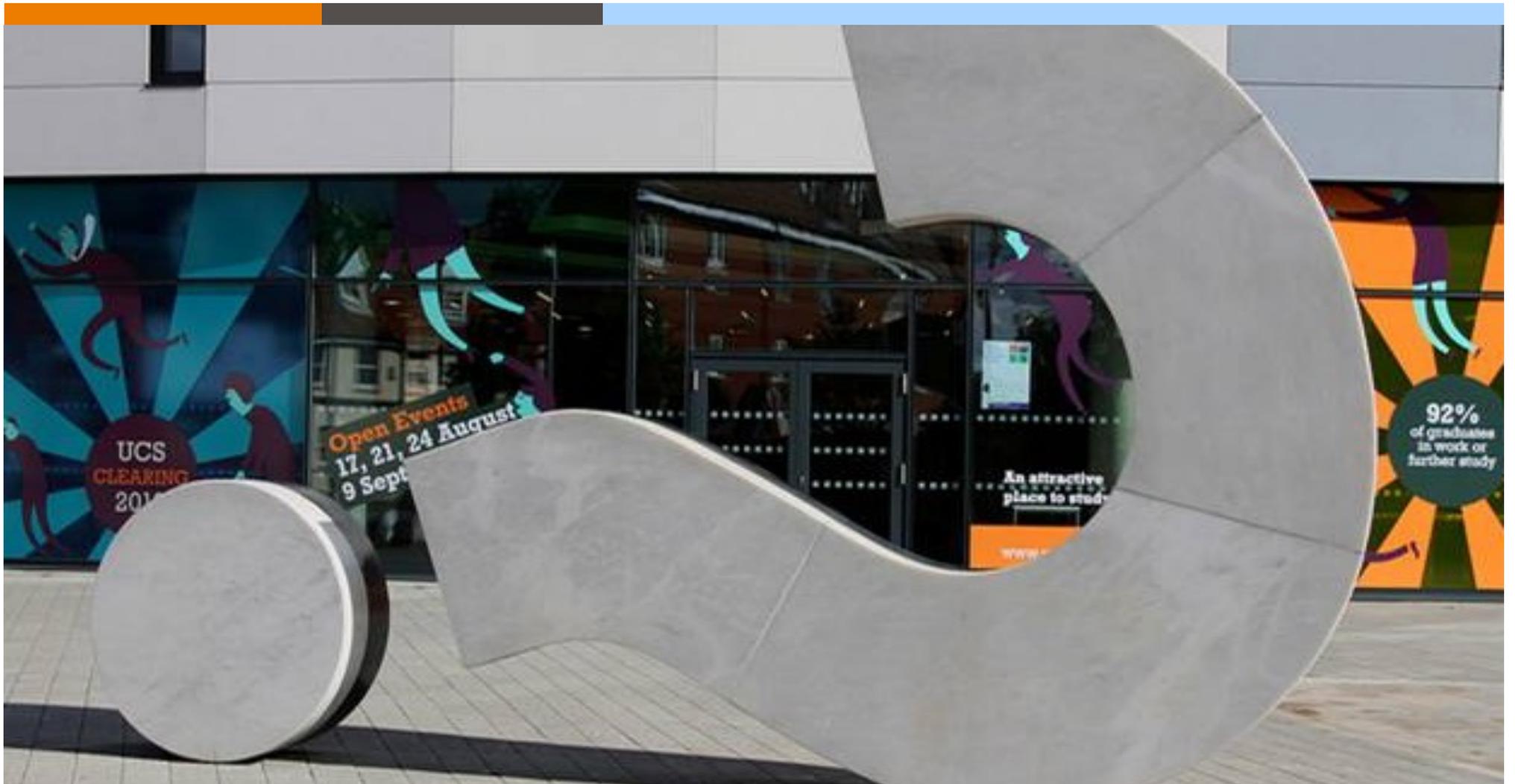


- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

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: 07804430054 (ar456@cam.ac.uk)

Thank you!!



HOW TO MANAGE DATA?

Questions to consider

What file formats will you use?

If you want your data to be re-used and sustainable in the long-term, you typically want to opt for open, non-proprietary formats.

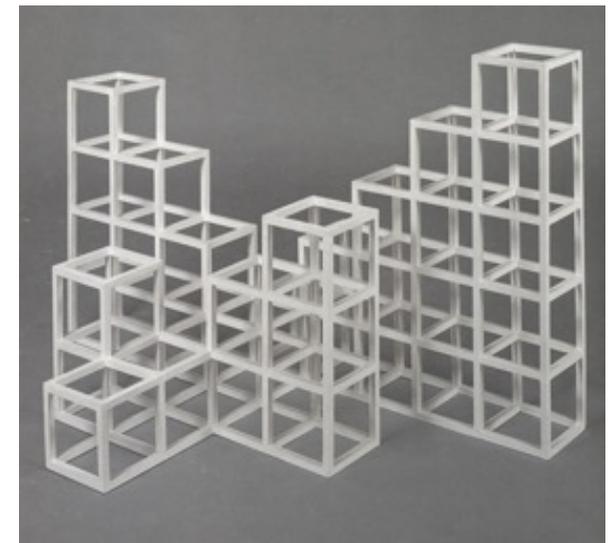
- Do you have a choice or do the instruments you use only export in certain formats?
- What is common in your field? Try to use something that is accepted and widespread
- Does your data centre recommend formats? If so it's best to use these.

Type	Recommended	Avoid for data sharing
Tabular data	CSV, TSV, SPSS portable	Excel
Text	Plain text, HTML, RTF PDF/A only if layout matters	Word
Media	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	Quicktime H264
Images	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF	RDBMS



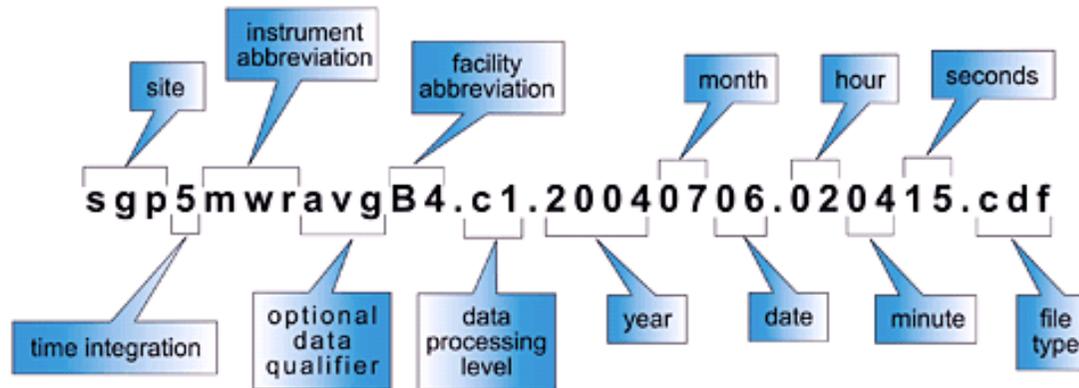
How will you organise all your stuff?

- Adopt file naming conventions:
 - <http://www.jiscdigitalmedia.ac.uk/guide/choosing-a-file-name>
- Design a good project folder structure
 - <http://research-data-toolkit.herts.ac.uk/document/research-project-file-plan>
- Develop a method for describing new versions of your files.



Good practice in file naming

An example netCDF data file name is depicted below:



Example from ARM Climate Research Facility
www.arm.gov/data/docs/plan

- Keep file and folder names short, but meaningful
- Agree a method for versioning
- Include dates in a set format e.g. YYYYMMDD
- Avoid using non-alphanumeric characters in file names
- Use hyphens or underscores not spaces e.g. day-sheet, day_sheet
- Order the elements in the most appropriate way to retrieve the record

Can others understand your data?

Think about what is needed in order to find, evaluate, understand, and reuse the data.

- Have you documented what you did and how?
- Did you develop code to run analyses? If so, this should be kept and shared too.
- Is it clear what each bit of your dataset means? Make sure the units are labelled and abbreviations explained.
- Record metadata so others can find your work e.g. title, date, creator(s), subject, format, rights...,

Where will you store the data?

- Your own device (laptop, flash drive, server etc.)
 - And if you lose it? Or it breaks?
- Departmental drives or university servers
- “Cloud” storage
 - Do they care as much about your data as you do?

The decision will be based on how sensitive your data are, how robust you need the storage to be, and who needs access to the data and when



Who will do the backup?

- Use managed services where possible (e.g. University filestores rather than local or external hard drives), so backup is done automatically
- 3... 2... 1... backup!
at least **3** copies of a file
on at least **2** different media
with at least **1** offsite
- Ask central IT team for advice



How to keep your data secure?

- Develop a practical solution that fits your circumstances
 - Ideally store your data on secure, managed servers
 - Restrict access to those who need to use / view the data
 - Keep anti-virus software up-to-date
 - Encrypt mobile devices carrying sensitive information



Which data need to be kept?

Five steps to follow

- ① Could this data be re-used
- ② Must it be kept as evidence or for legal reasons
- ③ Should it be kept for its potential value
- ④ Consider costs - do benefits outweigh cost?
- ⑤ Evaluate criteria to decide what to keep

5 steps to decide what data to keep

www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep



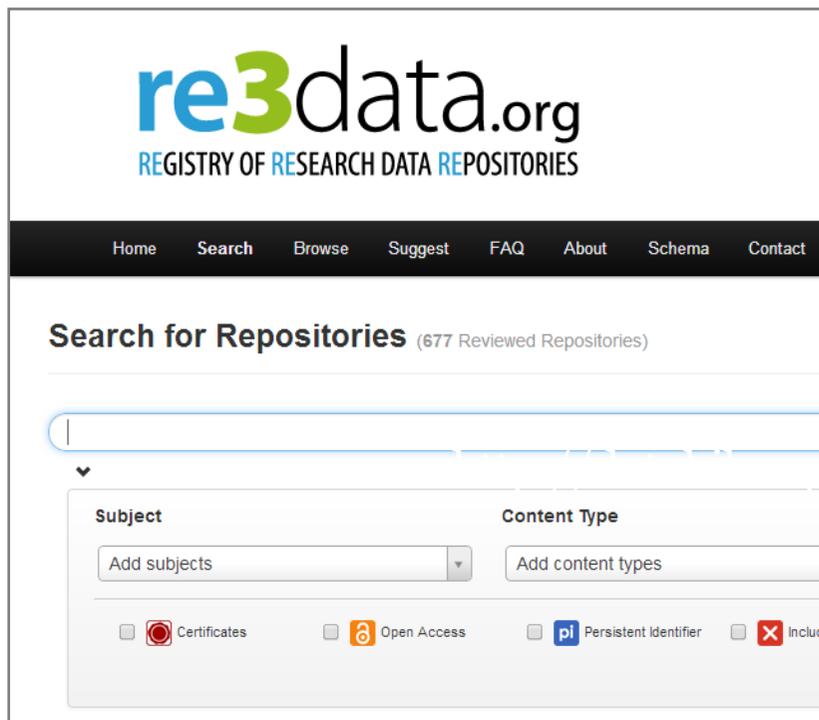
Can you publish / share your data?

- Who owns the data?
- Have you got consent for sharing?
- Do any licences you've signed permit sharing?
- Is the data in suitable formats?
- Is there enough documentation?



Where can you deposit?

- Does your publisher or funder suggest a repository?
- Are there data centres or community databases for your discipline?
- Does your university offer support for long-term preservation?



The screenshot shows the re3data.org website, which is a registry of research data repositories. The header includes the logo and the text "REGISTRY OF RESEARCH DATA REPOSITORIES". Below the header is a navigation menu with links for Home, Search, Browse, Suggest, FAQ, About, Schema, and Contact. The main content area is titled "Search for Repositories (677 Reviewed Repositories)" and features a search input field. Below the search field are two dropdown menus for "Subject" and "Content Type", each with an "Add" button. At the bottom of the search area, there are several checkboxes for filtering results: "Certificates" (checked), "Open Access" (unchecked), "Persistent Identifier" (checked), and "Include" (unchecked).

<http://service.re3data.org/search>



Zenodo

- OpenAIRE-CERN joint effort
- Multidisciplinary repository
- Multiple data types
 - Publications
 - Long tail of research data
- Citable data (DOI)
- Links funding, publications, data & software

www.zenodo.org

Managing and sharing data: a best practice guide



-  Planning for sharing
-  Consent and ethics
-  Copyright
-  Documenting your data
-  Formatting your data
-  Storing your data
-  Strategies for centres

Benefits and challenges of RDM in the digital age (MD)

1. Helicopter view
2. A few drivers
3. A few challenges



Helicopter view: What are the benefits of RDM?

- **SPEED:** Sharing data leads to a faster research process
- **TRANSPARENCY:** The data that underpins research can be made open for anyone to scrutinise, and attempt to replicate findings
- **EFFICIENCY:** Data collection can be funded once, and used many times for a variety of purposes
- **RISK MANAGEMENT:** A pro-active approach to data management reduces the risk of inappropriate disclosure of sensitive data, whether commercial or personal
- **PRESERVATION:** Lots of data is unique, and can only be captured once. If lost, it's irreplaceable.

Driver 1: Technology

- Developments in sensor technology, networking and digital storage enable new research and scientific paradigms
- As costs also fall, possibilities for data sharing, citation and re-use become much more widespread
- Journals dedicated solely to publishing data have even started to appear. That's not to say it's an entirely new thing: journals have always published data, just never before at such scale...



An earth science open access data journal

 **Geoscience Data Journal** Open Access

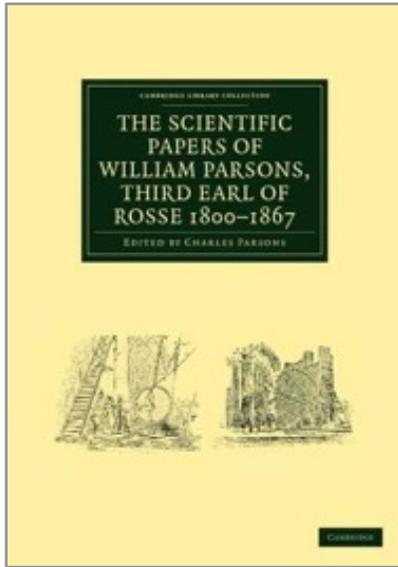
Open access to earth science data papers cross-linked to, and citing, datasets deposited in approved data centres.

Open for submissions. Geoscience Data Journal will publish its first papers in Summer 2012



www.geosciencedata.org  WILEY Open Access

Rosse



from
*Philosophical
 Transactions of
 the Royal
 Society,
 (MDCCCLXI)
 (or 1861 if
 you'd prefer)*



154 EARL OF ROSSE ON THE CONSTRUCTION OF SPECULA OF 6-FEET APERTURE,

Number in Herschel's Catalogue.	Number of times observed.	Description.
53	1	Sept. 10, 1857. S; R; vF; bM.
54	2	Nov. 22, 1854. pB; vS; R.
59	3	Dec. 22, 1848. 3 neb. in line, 2 of them "nova." Oct. 23, 1856. 1st is R; pB; bM; and has nucleus; 2nd bM; E, * involved; 3rd F; IE; bM.
60	1	Nov. 22, 1854. S; R; bM.
65	3	Sept. 18, 1857. S; pB. disc. in vF. haze of mottled neby.
69	7	Oct. 3, 1856. 69 is S; B; R; with B. nucleus; 70 is F; E. and patchy. I sometimes thought it was formed of two knots involved in F. neby; there appears to be a nebulous connexion between them all. Nov. 15, 1857. The silvered mirror shows the object brighter than before, but no new details: definition bad.
70		
71	7	Suspect spirality: light unequal.
72	3	Oct. 26, 1854. a F. object with two nuclei. Nov. 29, 1850. a is vB M; β has stellar point or nucleus. I suspect β to be a F. neb.
78	4	Pos. Dist. αβ 219° 35' αγ 315 1 8 αδ 81 0 44
79		
80	1	Nov. 3, 1855. 3 neb. nearly in line, sp. sf; β is bM, and IE. p. and f; α is R; bM; with a d. * sp, and is the largest of the 3; ε is S; F; R; δ is a *. Oct. 3, 1856. pL; not vF. Its brightest part is a line running diagonally, and there is a knot at either end; believed to be a spiral.
84	4	Nov. 4, 1850. Pos. Dist. αβ 169° 2' 19" βγ 160 4 22 γγ' 201 0 34 γγ'' 157 3 19 γγ''' 176 5 32 αζ 199 1 41 αζ* 79 4 65
85		
86		
87		
89	3	Oct. 26, 1854. A d. neb., both S; R; bM.
90	8	A cl. with much unresolved neby.
91	1	bM.
92	1	3 neb. in a triangle.
96	6	Oct. 26, 1854. Lenticular n. and s. Thought I saw a * at times in centre (1½-inch single lens); a lp. this is another vF. ray, sp, sf, and which has no nucleus. Oct. 16, 1855. vF; E. n. and s; has nucleus; * in n. end. Nov. 3, 1855. mE; pB. nucleus, and * in n. end; sp. this neb. is * of the 9th mag., and about the same distance p. this * is another neb. vF; mE. Dec. 7, 1855. Seen as before: comp. neb. verified. Oct. 23, 1856. F. ray has nucleus and a * in n. end. Sept. 18, 1857. E. n. and s; another vF. ray p, which is E. n. sf.
98	1	vF; R; S.
99	1	Oct. 3, 1856. S; F; R; bM; has nucleus.
103	3	Is n. of the 3rd of a group of 4 * in line; 3 "nova" near.
104	1	Dec. 5, 1850. Pos. Dist. αβ 28° 7' 36" Dec. 7, 1850. βγ 40 4 6 βδ 81 9 19 αβ 30 7 43
105	1	Oct. 23, 1856. 6 neb., all visible at once in finder eyepiece; 2 of them E., the others S; R; bM.
106	1	Dec. 11, 1854. vmE; bM (speculum dowed).
109	8	A variety of new nebulae found, but observations too voluminous to transcribe.
112	6	Sketch made, but no interesting details. Nov. 30, 1850. vF. and p. a quadruple *. Oct. 23, 1851. 3 * s f. neb.; light unequal. Sept. 16, 1852. 2" diameter; several * in it; probably a F. cl.

* This should be, I think, ζ. A S. d. neb. suspected below at α.

AND A SELECTION FROM THE OBSERVATIONS MADE WITH THEM.

Number in Herschel's Catalogue.	Number of times observed.	Description.	
113	2	Both have nuclei; "nova" near. Nov. 16, 1857. 113 is E. p. and f; * closely sp; 114 is R, with ragged edge and bM; "nova;" S; R; bM.	
114			
115			
121			
116	1	Oct. 3, 1850. The p. one is a pB. disc in F. outlying neby. The f. one is R; bM.	
118	1	Dec. 18, 1851. s. end of neb. is like a brush or broom with a split.	
118	2	4 neb. found, 2 have nuclei. 118 is S; R; 120 has 2 * on np. edge; E. p. and f.	
120			
119	1	Dec. 9, 1854. pL; pB; bM to a nucleus.	
123	2	Sept. 18, 1857. Rough sketch made; mE. np, sf; * F. triple * f.	
128	3	Nov. 28, 1856. L; B; mE; B. nucleus. "Nova" f.	
131	27	Pos. Dist. Nov. 29, 1850. αβ 215° 0' 51" αγ 163 0 56 αδ 160 2 56 αε 178 3 07 αζ 192 3 44 αη 206 4 14 αθ 224 4 58 Dec. 27, 1850. απ 147 5 34 αλ 179 5 56 ακ 201 6 42 αμ 143 6 28 αι 287 4 30 ακ 341 6 45 Jan. 2, 1851. α2 5 5 18 αβ 357 4 42 α3 51 11 0 α4 38 9 50 αr 58 11 16 Dec. 23, 1851. αw 161 5 20 αx 149 6 53 αy 172 6 32 αz 174 7 18 αp 205 2 22	
			For previous observations see Transactions, Part II. 1850.
			Sept. 13, 1850. Large spiral full of knots; to sf. is a S. neb. B. which on a very good night might appear attached to spiral, than which it is brighter. Oct. 11, 1850. Spiral arrangement not clearly seen. Nov. 27, 1850. Arms of spiral scarcely seen; fog. Nov. 30, 1850. Spiral form very indistinct; wind very high from s. Oct. 22, 1851. Viewed for drawing, I should not have seen the spiral arrangement had I not observed it before. Oct. 25, 1851. Neby. extends for several minutes all round, perhaps for half a degree in radius. Oct. 29, 1851. Observed for drawing. Dec. 14, 1851. Sketched. Dec. 26, 1851. Drawn. Dec. 7, 1856. This neb. reaches in length through at least a field and a half of finder eyepiece. Mr. Stoney's drawing leaves out a great deal of the neby. about the centre, and * supposed to left of centre of the tropezium of **, perhaps others also. Nov. 15, 1857. There are 3 * s about the principal nucleus. Dec. 7, 1857. Carefully observed, with a view to a new sketch. Dec. 18, 1857. Carefully observed, and my sketch proceeded with. See fig. 10, Plate XXVI.
132		1	Nov. 28, 1856. B; S; R. nucleus, a * p. and another n.
134		2	Oct. 20, 1854. Both S; R; R.
135			
139			Sought for four times; not found.
142		8	Dec. 13, 1848. Rough sketch made. Spiral? Dec. 14, 1848. Confirmed last night's observation; feel confident it is a spiral. Oct. 24, 1851. Centre formed of * s; easily seen to be such; several * s through the neb.
143		1	Oct. 3, 1856. vS; F; R; bM; had a * close to n. edge.
147		2	Nov. 30, 1856. S; R; bM. to a nucleus.

Driver 2: VfM via data re-use

Ships' log books build picture of climate change 14 October 2010

You can now help scientists understand the climate of the past and unearth new historical information by revisiting the voyages of First World War Royal Navy warships. Visitors to OldWeather.org will be able to retrace the routes taken by any of 280 Royal Navy ships. These include historic vessels such as HMS Caroline, the last survivor of the 1916 Battle of Jutland still afloat. By transcribing information about the weather and interesting events from images of each ship's logbook, web volunteers will help scientists build a more accurate picture of how our climate has changed over the last century.

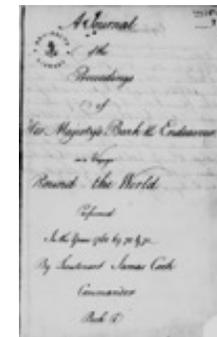
<http://www.nationalarchives.gov.uk/news/503.htm>



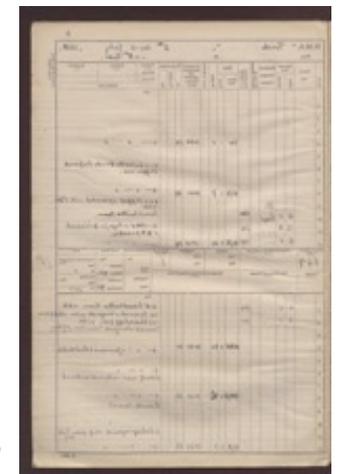
Detail from Royal Navy Recruitment poster, RNVR Signals branch, 1917 (Catalogue reference: ADM 1/8331)



HMS Beagle, 1830-34



Endeavour, 1768-71 (Captain Cook)



HMS Torch, 1918

Driver 3: Government pressure/support

6.9 The Research Councils expect the researchers they fund to deposit published articles or conference proceedings in an open access repository at or around the time of publication. But this practice is unevenly enforced. Therefore, as an immediate step, we have asked the Research Councils to ensure the researchers they fund fulfil the current requirements.

Additionally, the Research Councils have now agreed to invest £2 million in the development, by 2013, of a UK 'Gateway to Research'. In the first instance this will allow ready access to Research Council funded research information and related data but it will be designed so that it can also include research funded by others in due course. The Research Councils will work with their partners and users to ensure information is presented in a readily reusable form, using common formats and open standards.

<http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf>



Driver 4: Increasing 'openness' in public life

- Open Data is a philosophy, underpinned by pragmatism... transparency + utility.
- “Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.” - Wikipedia
- Governments, cities etc are all getting onboard
- Open Knowledge Foundation is basically the political / activist wing: <http://okfn.org/>
- From the government / industry side, we have the Open Data Institute: <http://theodi.org/>



The background of the slide is a close-up, slightly blurred image of the American flag, showing the stars and stripes in a curved, draped manner.

Meanwhile, in the USA...

May 9, 2013

United States Chief Technology Officer, Todd Park, and United States Chief Information Officer, Steven VanRoekel, discuss the importance of President Obama's executive order that takes groundbreaking new steps to make information generated and stored by the Federal Government more open and accessible to innovators and the public, to fuel entrepreneurship and economic growth while increasing government transparency and efficiency. The move will make troves of previously inaccessible or unmanageable data easily available to entrepreneurs, researchers, and others who can use those files to generate new products and services, build businesses, and create jobs.

<http://www.youtube.com/watch?v=n603rEnEGXA>

Why don't we live in a data sharing utopia?

- Five main reasons...

- i. Lack of widespread understanding of the fundamental issues
- ii. Lack of joined-up thinking within institutions, countries, internationally...
- iii. Issues around ownership / privacy
- iv. Technical/financial limitations, and the need for selection and appraisal of data
- v. Issues around reward and recognition for researchers

Overview

1. Benefits and challenges of research data management in the digital age (15 mins)
2. **Dealing with sensitive data: data protection, privacy, informed consent, commercial issues (15 mins)**
3. Exercise: Data Sharing (30 mins)
4. Other useful tools and resources (15 mins)



Dealing with sensitive data

- Data can be sensitive for two main reasons
 - **Commercially** sensitive data is information that may be used to derive economic capital. It may be closely guarded by its generators as a source of future income. Commercial R&D can be the difference between a product failing or succeeding.
 - Purely commercial data generation may not be subject to data sharing obligations, although interestingly in the case of drug trials it has started to be covered by legislation
 - Data generated as a result of public funding is increasingly expected to be shared at some stage, even if it has commercial potential. Note that it doesn't have to be shared immediately, and that expectations and norms vary from country to country.
- Furthermore, data may be **ethically** sensitive if it relates to living human subjects, or could be used to do harm (e.g. some areas of weapons and power research, disease studies, etc)
- This section of the workshop looks at different types of ethically sensitive data, and suggests ways in which this can be shared or reused appropriately...



Sharing ethically sensitive data

- Data relating to living humans is subject in many countries to data protection laws, which guard the privacy of (e.g.) the subjects of research
- In the UK, such data must not be shared without the subject's express (informed) consent, or without performing actions on the data in order to make it impossible to identify individual subjects
- Such actions include anonymising (or pseudonymising) data, aggregating it (removing a degree of detail), or restricting access to appropriate audiences
- In some cases, two or more anonymised datasets can be compared in order to 'deanonymise' and thus identify an individual. While it can never be ruled out entirely, this is highly unethical and a gross breach of trust

Overview

1. Benefits and challenges of research data management in the digital age (15 mins)
2. Dealing with sensitive data: data protection, privacy, informed consent, commercial issues (15 mins)
3. **Exercise: Data Sharing (30 mins)**
4. Other useful tools and resources (15 mins)



Interactive exercise

- Now that you know the reasons why data sharing is A Good Thing, we're going to do some role-playing (don't panic)
- We're going to put you in our shoes, not a researcher's shoes
- Sarah and I will articulate some of the most commonly heard objections to data sharing, and you're going to explain why we are wrong 😊
- After each section, we'll look at some suggested ripostes from the experts in sensitive data at the UK Data Archive...



	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
1	My data is not of interest or use to anyone else.	It is! Researchers want to access data from all kinds of studies, methodologies and disciplines. It is very difficult to predict which data may be important for future research. Who would have thought that amateur gardener's diaries would one day provide essential data for climate change research? Your data may also be essential for teaching purposes. Sharing is not just about archiving your data but about sharing them amongst colleagues.
2	I want to publish my work before anyone else sees my data.	Data sharing will not stand in the way of you first using your data for your publications. Most research funders allow you some period of sole use, but also want timely sharing. Also remember that you have already been working with your data for some time so you undoubtedly know the data better than anyone coming to use them afresh. If you are still concerned you can embargo your data for a specific period of time.

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
3	I have not got the time or money to prepare data for sharing	It is important to plan data management early in the research data lifecycle. Data management ideally becomes an integral part of your research practice, reduces time and financial costs and greatly enhancing the quality of the data for your use too.
4	If I ask my respondents for consent to share their data then they will not agree to participate in the study.	Don't assume that participants will not participate because data sharing is discussed. Talk to them – they may be less reluctant than you might think, or less concerned over data sharing! Make it clear that it is entirely their decision, whereby they can decide whether their data can be shared, independent of them participating in the research. Explain clearly what data sharing means, and why it may be important. But they are still free to consent or not. You can always explain what data archiving means in practice for their data. If you have not asked permission to share data during the research, then you can always return to gain retrospective permission from participants.

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
5	I am doing highly sensitive research. I cannot possibly make my data available for others to see.	The first thing is to ask respondents and see if you can get consent for sharing in the first instance. Anonymisation procedures can help to protect identifying information. If these first two strategies are not appropriate then consider controlling access to the data or embargoing for a period of time. Also data that is held in the UK Data Archive is not publically available. Only registered researchers can gain access to the data.
6	I am doing quantitative research and the combination of my variables discloses my participant's identity.	Quantitative data can be anonymised through processes of aggregation, top coding, removal of variables, or controlled access to certain variables (i.e. postcodes).

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
7	I have collected audiovisual data and I cannot anonymise them, therefore I cannot share these data.	Visual data can be anonymised through blurring faces or distorting voices, but this can be time consuming and costly to carry out. It can mean losing much of the value of the data. It is better to ask for consent to share data from participants in an unanonymised form, and/or control access to the data.
8	I have made promises to destroy my data once the project finishes.	Why were such promises made? Always avoid making unnecessary promises to destroy data. There is usually no legal or ethical need to do so, except in the case of personal data. But that certainly would not apply to research data in general. Also consider where you have received this advice from? You may need to negotiate with research ethics committee or ethics boards about this agreement.

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
9	My data have been gathered under complete assurances of confidentiality.	Again why was such an assurance made? It is best to avoid unnecessary promises. Anonymisation procedures can be implemented to protect identities, but confidentiality can never be completely guaranteed. You can also consider controlling access to the data.
10	My data collection and resulting transcripts are in a foreign language.	This should not be a problem. The UK Data Archive can accept foreign language transcripts although translations into English are preferred.

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
11	It is impossible to anonymise my transcripts as too much useful information is lost.	Get in touch with us at the UK Data Archive. We may be able to help and it might not be as difficult as it looks. Also, access controls on the data may be a better solution than anonymisation if too much useful information would be lost.
12	My data collection contains data which I have purchased and it cannot be made public.	It is important to know who holds the copyright to the data you are using and to obtain the relevant permissions. You need to be aware of the licence conditions of the data you are using and what you can and cannot do with the data.

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
13	Other researchers would not understand my data at all - or may use them for the wrong purpose.	Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data.
14	There is IPR in the data.	This should not be a problem if you seek copyright permission from the owner of the intellectual property rights. This is best done early on in the research project, but could be sought retrospectively.

Role playing exercise derived from the UKDA's
 “Potential barriers to data sharing - with suggested solutions” (CC-BY-NC-SA)

The original is available from <http://data-archive.ac.uk/create-manage/training-resources>





DATA MANAGEMENT PLANNING (SJ)

Building a structure to work to

Data Management Plans

It's useful to consider how you will manage and share your data in practice. Many research funders and universities now ask for these details in a DMP.

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be shared/made available?
- If not, why? e.g. ethics & IP issues, embargoes, confidentiality
- How will this data be curated and preserved?



www.dcc.ac.uk/resources/data-management-plans/checklist

Lots of health funders require a DMP



Their focus is often on data sharing

- Which data will be shared?
- When will it be shared?
- With whom?
- How will the data be shared?
- Will any restrictions or conditions govern use?
-

Cancer Research UK

The following should be considered when developing a data sharing plan:

- The volume, type, content and format of the final dataset
- The standards that will be utilised for data collection and management
- The metadata, documentation or other supporting material that should accompany the data for it to be interpreted correctly
- The method used to share data
- The timescale for public release of data
- The long-term preservation plan for the dataset
- Whether a data sharing agreement will be required
- Any reasons why there may be restrictions on data sharing



Wellcome Trust

Applicants should consider the following seven questions:

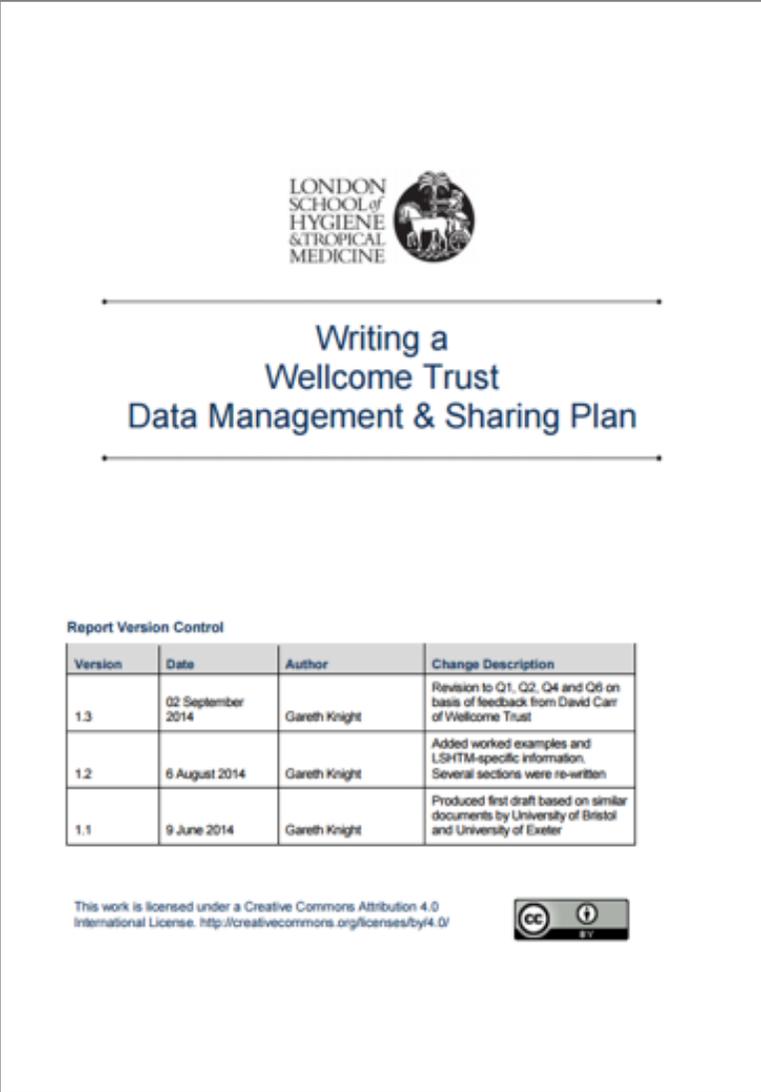
- i. What data outputs will your research generate and what data will have value to other researchers?
- ii. When will you share the data?
- iii. Where will you make the data available?
- iv. How will other researchers be able to access the data?
- v. Are any limits to data sharing required - for example, to either safeguard research participants or to gain appropriate intellectual property protection?
- vi. How will you ensure that key datasets are preserved to ensure their long-term value?
- vii. What resources will you require to deliver your plan?



Guidance on writing a DMP

- Explains what is asked for
- Gives example answers
- Suggests best practices
- Provides links to standards, tools and support

www.lshtm.ac.uk/research/researchdataman/plan/wellcometrust_dmp.pdf



LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Writing a Wellcome Trust Data Management & Sharing Plan

Report Version Control

Version	Date	Author	Change Description
1.3	02 September 2014	Gareth Knight	Revision to Q1, Q2, Q4 and Q6 on basis of feedback from David Carr of Wellcome Trust
1.2	6 August 2014	Gareth Knight	Added worked examples and LSHTM-specific information. Several sections were re-written
1.1	9 June 2014	Gareth Knight	Produced first draft based on similar documents by University of Bristol and University of Exeter

This work is licensed under a Creative Commons Attribution 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>



What data will be generated?

Why is this important?

A good description of the data to be collected will help reviewers understand the characteristics of the data, their relationship to existing data, and any disclosure risks that may apply.

Example:

Data capture will be performed during months 8-20, at which point four types of data will be collected from the (approximately) 500 study subjects:

[1] Questionnaire: An interviewer-administered CAPI questionnaire will be performed covering medical history, health and socio-demographic circumstances

[2] Clinical: A clinical examination will be conducted to measure blood pressure, anthropometry, 12-lead digital ECG assessed for abnormalities; assessment of physical functions (grip strength, chair rise, walk speed) and cognition.

[3] Biological samples: Blood samples will be taken from each subject and centrifuged within 2 hours. Together with serum and plasma, it will aliquoted into 20 bar-coded cryovials and stored at -70C

[4] Interviews: It has previously been found that participant's partners tend to report higher levels of alcohol consumption and more frequent episodes of hazardous drinking than men themselves. We will interview partners of 1 in 4 of the study participants (selected randomly) in order to estimate the extent of under or over reporting of behaviours.

All 4 data types will be usable for future research in some form, subject to appropriate measures being implemented to protect participant confidentiality. Digital and physical outputs will be made available to bona fide researchers for health-related research, irrespective of their institution (university, charity, government, commercial) or location (UK or elsewhere).



When will you share the data?

Why is this important?

Research funders are looking for timely data sharing with minimal or no restrictions if possible. Embargo periods / delays to sharing should be justified and in line with standard practice for the field.

Example

Research papers written and published during the funding period will be made available with a subset of the (anonymised) data necessary to verify the research findings, in compliance with the Wellcome Trust's OA Policy.

The study team will make digital data (outputs 1, 2 and 4) available within 6 months of project completion. This embargo period is requested to allow time for additional analysis and further publication of research findings to be performed.



How can others access the data?

Why is this important?

If the data aren't discoverable, accessible and intelligible, they won't be reused. Data should be shared in a meaningful way.

Example

To enable potential users to learn of the dataset's existence, structured metadata describing its content will be created and made available in human readable and machine processable form. The LSHTM Data Repository will publish metadata in several metadata formats, including Dublin Core, via OAI-PMH, RSS and ATOM, for indexing by search engines and harvesting by research data catalogues.

To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data.

Biological samples (output 3) will be deposited with the UK BioBank for future use, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points. Similar to the above, applicants will be required to comply with a Data Transfer Agreement prior to gaining access to data. DNA extracted from biological samples will be normalized and plated at Lab Y for use in future studies, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points.



Are any limit to sharing required?

Why is this important?

As funders expect data to be shared, any restrictions need to be valid. Protection of human subjects is a fundamental tenet of research and an important ethical obligation for everyone.

Example

To protect participant confidentiality, data outputs will be anonymised prior to deposit. To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule.

Biological samples are limited and depletable, so access will need to be carefully controlled and coordinated. The quantity of sample that is provided will be judged against the potential benefits of the research project, with advice from appropriate experts as required.

State the long-term preservation plan

Why is this important?

Digital data need to be actively managed over time to ensure that they will always be available and usable. Depositing data resources with a trusted digital archive can ensure that they are curated and handled according to good practices in digital preservation.

Example

Data will be provided in file formats considered appropriate for long-term access, as recommended by the UK Data Service (<http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx>). For example, SPSS Portal format and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview transcripts. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule.

Appropriate documentation necessary to understand the data will also be provided. This will include high-level information on the study itself and a comprehensive data dictionary, which describes the purpose of each variable and the permitted values.



More example plans

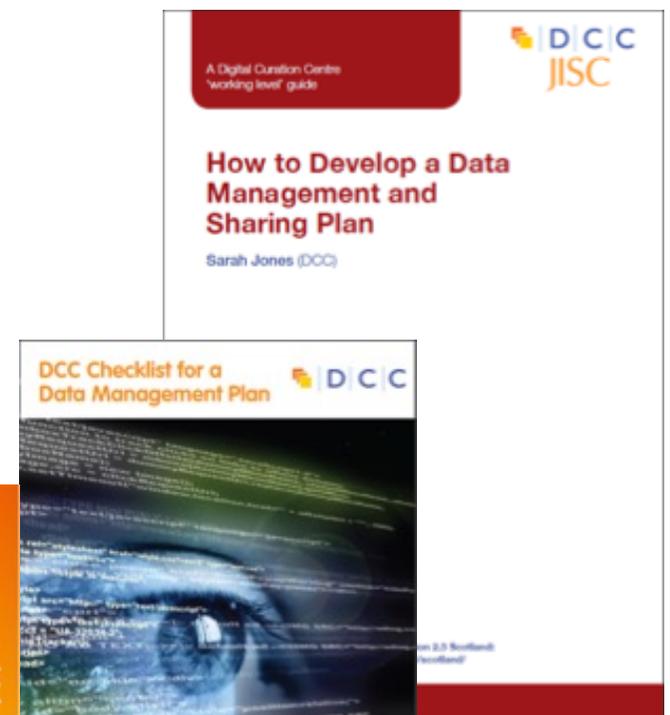
- Technical appendix submitted to AHRC by Bristol Uni
<http://data.bris.ac.uk/files/2013/02/data.bris-AHRC-Technical-Plan-v21.pdf>
- Rural Economy & Land Use (RELU) programme examples
<http://relu.data-archive.ac.uk/data-sharing/planning/examples>
- UCSD example DMPs (20+ scientific plans for NSF)
<http://rci.ucsd.edu/dmp/examples.html>
- My DMP - a satire (what not to write!)
<http://ivory.idyll.org/blog/data-management.html>
- Further examples:
www.dcc.ac.uk/resources/data-management-plans/guidance-examples



DCC support on Data Management Plans

- Checklist on what to include
- How to guide on developing a plan
- Guidance on assessing plans (forthcoming)
- Webinars and training materials
- DMPonline tool
- Example DMPs

www.dcc.ac.uk/resources/data-management-plans



What is DMPonline?

- A web-based tool to help researchers write Data Management and Sharing Plans
- Matches requirements with guidance tailored to user
- Free to use for anyone
- Developed by the DCC

<https://dmponline.dcc.ac.uk>



Main features in DMPonline

- Templates for different requirements (funder or institution)
- Tailored guidance (funder, institutional, discipline-specific etc)
- Ability to provide examples and suggested answers
- Supports multiple phases (e.g. pre- / during / post-project)
- Granular read / write / share permissions
- Comment feature for collaboration
- Customised exports to a variety of formats
- Single-sign-on facility (for UK unis)



How the tool works

Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the most common themes.

If applying for funding, select your research funder.

Otherwise leave blank.

Economic and Social Research Council ▼

[Not applicable/not listed.](#)

To see institutional questions and/or guidance, select your organisation.

You may leave blank or select a different organisation to your own.

University of Glasgow ▼

[Not applicable/not listed.](#)

Tick to select any other sources of guidance you wish to see.

DCC guidance

Create plan

Click to write a generic DMP

Or choose your funder to get their specific template

Pick your uni to add local guidance and to get the uni template if there isn't a funder one

Choose any additional optional guidance

DMPonline demo

The screenshot displays the DMPonline website interface. In the top left corner is the DMP ONLINE logo. The top right navigation bar contains links for Home, About, Roadmap, and Help. The main content area is divided into two sections. The left section, titled 'Welcome.', contains a paragraph: 'DMPonline has been developed by the Digital Curation Centre to help you write data management plans.' Below this is a video player titled 'Screencast on how to use DMPonline' which shows a 'Create a new plan' form with fields for 'Primary research project or research centre', 'Secondary research project or research centre', and 'Your contact details'. The video player shows a progress bar at 4:47. The right section contains a 'Sign in' form with fields for 'Email address *' and 'Password *', a 'Forgot your password?' link, a 'Remember me' checkbox, a 'Sign in' button, and a link for 'Or, sign in with your institutional credentials (UK users only)'. Below the sign-in form is a 'Sign up' section with the text 'New to DMPonline? Sign up today.' and a plus sign icon.

Exercise on writing a DMP

Pick one of the themes below and think about the approach you will take:

- Data creation (including standards and metadata)
- Data storage, backup and security
- Ethics and intellectual property
- Data sharing
- Data preservation

Draft some text and discuss your ideas with partners

Thanks - any questions

- DCC resources on Research Data Management
www.dcc.ac.uk/resources
- DMP guidance, tools and example plans:
www.dcc.ac.uk/resources/data-management-plans

Follow us on Twitter:
@digitalcuration
#ukdcc
#DMPonline



Overview

1. Benefits and challenges of research data management in the digital age (15 mins)
2. Dealing with sensitive data: data protection, privacy, informed consent, commercial issues (15 mins)
3. Exercise: Data Sharing (30 mins)
4. **Other useful tools and resources (15 mins)**



SHERPA services



... opening access to research

[Home](#) | [Guidance](#) | [Repositories](#) | [Projects](#) | [Links](#) | [About](#) | [Contacts](#)

[Award winning](#) SHERPA is investigating issues in the future of scholarly communication. It is developing [open-access](#) institutional repositories in universities to facilitate the rapid and efficient worldwide dissemination of research. SHERPA services and the [SHERPA Partnership](#) are both based at the [Centre for Research Communications](#) at the University of Nottingham.

SHERPA Services

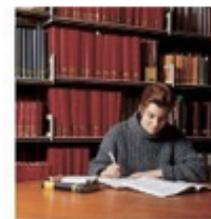
- [RoMEO](#) - Publisher's copyright & archiving policies
- [JULIET](#) - Research funders archiving mandates and guidelines
- [OpenDOAR](#) worldwide Directory of Open Access Repositories
- [SHERPA Search](#) - simple full-text search of UK repositories

SHERPA Resources

- [SHERPA speaking events](#) - for future events see [CRC](#)
- [Guidance](#) on depositing material; copyright; and [open access for authors](#)
- [Advocacy Materials](#) for administrators, including past SHERPA presentations
- [Links](#) to other initiatives and related background information

Current SHERPA & Allied Projects

The SHERPA services and partnership are based at the Centre for Research Communications at the University of Nottingham. For more information on current projects see the [CRC website](#).

Stuffy old journals -
Away with your subscriptions!
Open Access Rules!
Lisa Martin, BioMedCentral

News

- [Upgrade to SHERPA/JULIET Released](#)
- [RoMEO API - Manual, REST Requests, and Wish List](#)
- [Hungarian Version of RoMEO Released](#)
- [Útnak indul a RoMEO magyar változata](#)
- [Country Statistics for SHERPA/RoMEO](#)
- [Access Keys Available for SHERPA/RoMEO API V.2.9](#)
- [more news...](#)

Jobs



EUDAT



- EUDAT offers **common data services** through a geographically distributed, resilient network of 35 European organisations. These **shared services and storage resources** are distributed across 15 European nations, and data is stored alongside some of Europe's most powerful supercomputers
- The EUDAT services address the full lifecycle of research data, covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data
- The vision is to enable European researchers and practitioners from any academic discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure (CDI) conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres
- DCC is a partner in EUDAT, and we're working to integrate our DMPonline tool with the EUDAT suite of services / infrastructure



Zenodo



- Zenodo is a free-to-use data archive, run by the lovely people at CERN
- It accepts any kind of data, from any academic discipline
- It is generally preferable to store data in a disciplinary data centre, but not all scholarly subjects are equally well served with data centres, so this may make for a useful fallback option
- See <http://zenodo.org/> for more details



Figshare

 search figshare (titles, tags, authors, etc.)  [Browse](#) [Upload](#)

store, share, discover research

manage your research in the cloud and control who you share it with
or make it publicly available and citable

[About figshare](#) [Browse research](#)

[See how we support data management for Institutions >](#)

[See how we partner with Publishers >](#)

Arkivum

Arkivum

[ABOUT US](#) [SOLUTIONS](#) [PRODUCTS](#) [RESOURCE LIBRARY](#) [BLOG](#) [Q](#)



Long-term, large-scale data archiving.

Keeping your data available, protected and curated.

Why Arkivum?

Arkivum provides data archiving services to a range of industries including higher education, life sciences and heritage. These services assure the long-term value, trustworthiness and authenticity of data irrespective of whether it's terabytes or petabytes being archived, and irrespective of whether the retention period is a few years, a decade, or a quarter of a century.

<http://arkivum.com/life-sciences/>

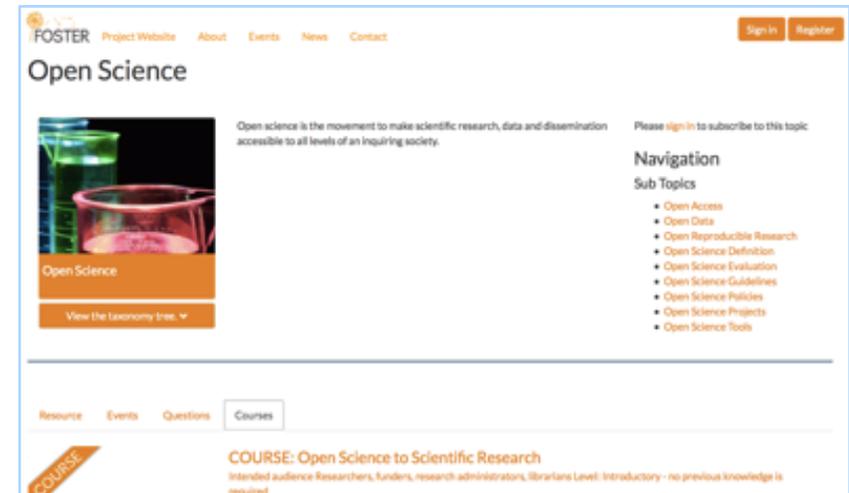




The FOSTER project

Facilitate Open Science Training for European Research

- FOSTER is developing MOOC-type content and dedicated training modules in different aspects of Open Science, including research data management
- The FOSTER Portal also provides access to existing content that can be reused and/or repackaged to support e-learning, blended learning, self-learning, etc
- There is also a Helpdesk which can direct your enquiries to appropriate experts



The screenshot shows the FOSTER Open Science portal. At the top, there is a navigation bar with links for 'Project Website', 'About', 'Events', 'News', and 'Contact', along with 'Sign In' and 'Register' buttons. The main heading is 'Open Science'. Below this, there is a featured image of laboratory glassware with the text 'Open Science' and a button 'View the Laxonomy list'. To the right, there is a definition of Open Science and a 'Navigation' section with 'Sub Topics' including Open Access, Open Data, Open Reproducible Research, Open Science Definition, Open Science Evaluation, Open Science Guidelines, Open Science Policies, Open Science Projects, and Open Science Tools. At the bottom, there is a 'COURSE' section for 'Open Science to Scientific Research' with a description of the intended audience and level.

DCC training and guidance

The screenshot displays the DCC website interface. At the top, there is a navigation bar with the DCC logo and the tagline "because good research needs good data". A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a secondary navigation bar with links to Home, Digital curation, About us, News, Events, Resources, Training, Projects, Community, and Tailored support. The main content area is divided into several sections. On the left, there is a sidebar with a list of resources under the heading "In this section". The main content area features a breadcrumb trail: "Home > Drupal > Resources > Briefing Papers". Below the breadcrumb trail, there is a section titled "Briefing Papers" with a sub-section "Briefing Papers" and a list of resources: Introduction to Curation, Making the Case for RDM, 5 Steps to Research Data Readiness, Citizen Science, Legal Watch Papers, Standards Watch Papers, and Technology Watch Papers. To the right of the list, there is a text block describing the Briefing Papers: "Among the many online learning resources that the DCC offers digital curators are high-level briefing papers and legal watch, standards watch and technology watch papers. Our digital library of resources is free to use and contains everything you need to engage effectively in digital curation and data preservation activities. The DCC provides a range of high-level papers that provide an introduction to all areas of digital curation." Below this text, there is a bullet point: "The Introduction to Curation is made up of short papers that are designed to help grow your basic understanding of key curation issues by providing a high-level introduction to a given topic." On the far right, there is a section titled "Curation Reference Manual" with a sub-section "Curation Reference Manual" and a text block: "Advice, in-depth information and criticism on current techniques and best practice. Contributions are made by our extended network of specialist partners and associates." Below this text, there is a "Read more" button.

Contact us

DCC because good research needs good data Search

Home Digital curation About us News Events Resources Training Projects Community Tailored support

Home > Drupal > Training

DCC because good research needs good data Search

Home Digital curation About us News Events Resources Training Projects Community Tailored support

Home > Drupal > Resources > Briefing Papers

In this section

Curation webinars

Digital Curation 101

DMP workshop at UCT

LIASA RDM workshop

UCT Research Data Management Policy and Strategy Workshop

Materials for Trainers

Career profiles and related data management skills

DC 101 training materials

Disciplinary RDM training

Skills frameworks

Data management courses training

Tools of the Trade training

In this section

Briefing Papers

Introduction to Curation

Making the Case for RDM

5 Steps to Research Data Readiness

Citizen Science

Legal Watch Papers

Standards Watch Papers

Technology Watch Papers

How-to Guides & Checklists

Developing RDM Services

Curation Lifecycle Model

Curation Reference Manual

Policy and legal

Data Management Plans

Briefing Papers

View Edit Outline Revisions Clone

Among the many online learning resources that the DCC offers digital curators are high-level briefing papers and legal watch, standards watch and technology watch papers.

Our digital library of resources is free to use and contains everything you need to engage effectively in digital curation and data preservation activities.

The DCC provides a range of high-level papers that provide an introduction to all areas of digital curation.

- The [Introduction to Curation](#) is made up of short papers that are designed to help grow your basic understanding of key curation issues by providing a high-level introduction to a given topic.

Curation Reference Manual

[\[edit\]](#)

Advice, in-depth information and criticism on current techniques and best practice.

Contributions are made by our extended network of specialist partners and associates.

[Read more](#)



UK Data Archive guidance

HELP CONTACT US SIGN UP

THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES

HOME ABOUT US **CREATE & MANAGE DATA** DEPOSIT DATA HOW WE CURATE DATA FIND DATA NEWS & EVENTS

CREATE & MANAGE DATA ADVICE & TRAINING

SEARCH OUR SITE GO

A QUICK GUIDE TO THE ARCHIVE
9 of 10: You won't be able to find your family records on our site. You should visit a genealogical site instead

MANAGING AND SHARING RESEARCH DATA

SAGE publication

CREATE & MANAGE DATA FAQ

UK • DATA ARCHIVE

CREATE & MANAGE DATA
RESEARCH DATA LIFECYCLE
PLANNING FOR SHARING
CONSENT & ETHICS
COPYRIGHT
DOCUMENTING YOUR DATA
FORMATTING YOUR DATA
STORING YOUR DATA
STRATEGIES FOR CENTRES
ADVICE & TRAINING
TRAINING PROGRAMME & EVENTS
EVENTS ARCHIVE

TRAINING PROGRAMME & EVENTS

 **Guided walk through ReShare**
22 September 2015
Online 12.00-13.00

 **Research Data Management**
6 - 7 October 2015
University of Manchester

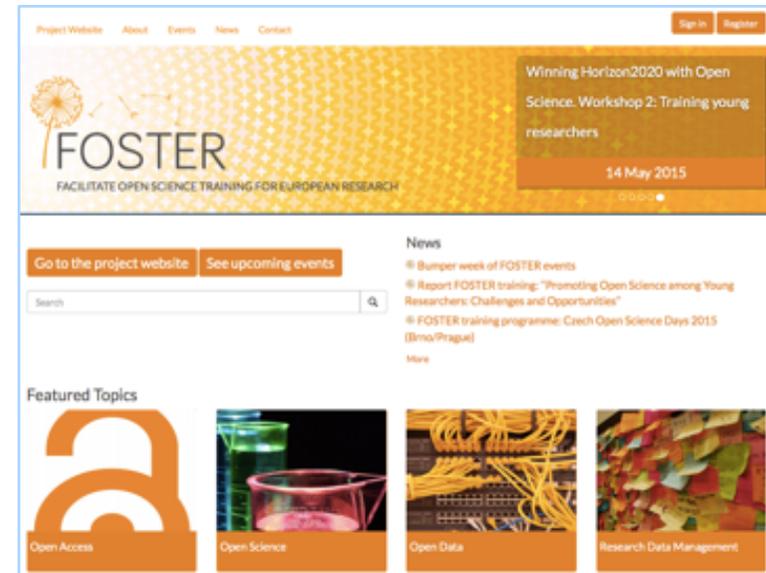
 **Human rights insights from data**
29 - 30 October 2015
Wivenhoe House Hotel,
University of Essex

EVENT ARCHIVE



Thank you / Danke

- For more information about the FOSTER project:
 - Website: www.fosteropenscience.eu
 - Principal investigator: Eloy Rodrigues (eloy@sdum.uminho.pt)
 - General enquiries: Gwen Franck (gwen.franck@eifl.net)
 - Twitter: @fosterscience
- My contact details:
 - Email: martin.donnelly@ed.ac.uk
 - Twitter: @mkdDCC
 - Slideshare: <http://www.slideshare.net/martindonnelly>



This work is licensed under the Creative Commons Attribution 2.5 UK: Scotland License.



Our contact details...

Sarah Jones

DCC, University of Glasgow

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC

Martin Donnelly

DCC, University of Edinburgh

martin.donnelly@ed.ac.uk

Twitter: @mkdDCC

