Data sharing lessons learnt at Cambridge: the whys and hows

Dr Marta Teperek  
Office of Scholarly Communication, University of Cambridge  
@martateperek  
mt446@cam.ac.uk

14th June 2016, Opendata.ch/2016
This session will cover:

• How **not** to start with RDM services
• The (forgotten?) benefits of data sharing
• RDM services at Cambridge
• Democratic approach to RDM
• Researchers’ reactions
• Reflections
How not to start with RDM
In the UK funders consider data management and sharing seriously.
You **must** do it

EPSRC do random checks to make sure data are shared in publications that acknowledge their funding

https://www.epsrc.ac.uk/files/aboutus/standards/clarificationsofexpectationsresearchdatamanagement/
Immediate reaction:

• This is not my priority
• Why would I do that?
• People will steal my results!
• Data management is a waste of time
• Nobody will understand my data
• It would take me 5 years to find all my data!
Let’s re-think the approach...
Approach number 2:

• Benefits of sharing
• Support available
• Democratic approach
Talking about the benefits:
Why would you want to share your research?
Sharing your data is ‘A Good Thing’

- More exposure for your work
- Practitioners can apply your findings
- Higher citation rates
- Your research can influence policy
- The public can access your findings
- Compliant with grant rules
- Taxpayers get value for money
- Researchers in developing countries can see your work
Some other very important reasons to share
Research relies on the principle that we share our findings

*Ideas and results need to be shared to move human knowledge forward*
Science relies on the principle that we share our findings

**FDA record of clinical trials with 12 antidepressants:**
Only positive results published

<table>
<thead>
<tr>
<th>Positive</th>
<th>Not positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 3 6 10 12 17 20 29 30 34 36</td>
<td>2 4 7 9 11 13 18 21 22 27 32 35</td>
</tr>
<tr>
<td>2 5 8 14 19 22 29 31 32 36 38</td>
<td>15 16 19 23 24 25 26 27 28</td>
</tr>
</tbody>
</table>

From Dr Eric Turner:
[https://figshare.com/articles/Peer_review_After_Results_are_Known_Are_we_PARKing_the_Cart_Before_the_Horse_/3381379](https://figshare.com/articles/Peer_review_After_Results_are_Known_Are_we_PARKing_the_Cart_Before_the_Horse_/3381379)
p-value 0.05: who is going to publish their results?

Non-positive results need to be shared not to waste time and resources.
Sperm is epigenetically programmed to regulate gene transcription in embryos

*Corresponding authors. E-mail: j.jullien@cam.ac.uk; mt446@cam.ac.uk

or

Data access

All ChIP-seq, RNA-seq, MBD-seq, and MNase-seq data sets have been submitted to the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE75164.

http://dx.doi.org/10.1101/gr.201541.115
Selfish reason: share once and don’t be bothered

Research Data Discovery Service (Alpha)

11,121 datasets found

Cathedral Close St Mary Major, Exeter 1971-1976 (Exeter archive site 40)

https://researchdata.jiscinvolve.org/wp/2016/02/04/932/
At the same time:
Development of RDM services

*Adequate provisions need to be in place for people to do the right thing.*
Please fill in the form below and we will get back to you as soon as possible to arrange a meeting.

www.data.cam.ac.uk/consultancy
Data management plan support service

- Select -

Deadline for the submission of the grant proposal *

- Select -

Your name *

Your e-mail address *

Please enter your e-mail address carefully – we will use it to get in touch with you.

Upload your draft plan *

Browse... No file selected. Upload

Submit

www.data.cam.ac.uk/DMPsupport
Events

15 JUN
Research Integrity
Wednesday 15 June, 10:00 AM - 11:00 AM, Mill Lane Lecture Room 5 - Mill Lane, Cambridge

16 JUN
Scholarship in the digital age – a survival guide for researchers
Thursday 16 June, 11:30 - 15:30, S1, Alison Richard Building, West Road, University of Cambridge

23 JUN
Working towards an Open Research future
Thursday 23 June, 14:00 to 17:30 (BST), The Cambridge Union - 9A Bridge Street, Cambridge, CB2 1UB

26 JUL
Research Data Management for Librarians
Tuesday 26 July, 14:30-16:30, Mill Lane Lecture Room 5 - 1 Mill Lane, Cambridge

28 JUL
The Data Dialogue. Time to Share: Navigating Boundaries & Benefits
Thursday 28 July, 9:30 - 17:00, Murray Edwards College, The University of Cambridge.

2 AUG
Altmetrics: A Primer for Librarians
Tuesday 2 August, 14:30-15:30, Mill Lane Lecture Room 5 - 1 Mill Lane, Cambridge

www.osc.cam.ac.uk/events
Apollo - University of Cambridge Repository

This repository holds the research output of members of the University of Cambridge. It is delivered and managed by the University Library’s Office of Scholarly Communication team.

School of Arts and Humanities
School of Clinical Medicine
School of Technology
School of the Biological Sciences
School of the Humanities and Social Sciences
School of the Physical Sciences
Colleges
Other Communities

www.repository.cam.ac.uk
Easy to use webform to upload data
Non-positive data can be shared as well
Policy development and discussions with funders

<table>
<thead>
<tr>
<th>Funder</th>
<th>Policy Highlights</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRC</td>
<td>The MRC expects valuable data arising from MRC-funded research to be made available to the scientific community with as few restrictions as possible so as to maximize the value of the data for research and for eventual patient and public benefit. Such data must be shared in a timely and responsible manner. Grant holders shall review and update their data management plans annually. MRC also provides a detailed guideline on dealing with personal data in medical research.</td>
</tr>
<tr>
<td>NERC</td>
<td>Possible sanctions: &quot;those who do not meet the data management requirements [...] risk having award payments withheld or becoming ineligible for future funding&quot; Data needs to be deposited into a NERC data centre within 2 years of collection</td>
</tr>
<tr>
<td>NIH</td>
<td>&quot;Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health.&quot;</td>
</tr>
<tr>
<td>NIHR</td>
<td>&quot;Data generated through participation of patients and the public should be put to maximum use by the research community. When you submit your final report to NIHR they will ask you to make a statement about your data archiving position. Possible responses might state that all available data: Can be obtained from the corresponding author. Is included as an appendix to</td>
</tr>
</tbody>
</table>

www.data.cam.ac.uk/funders

Funder names arranged alphabetically. Click on the hyperlink to see the full-length policy.

Key policy highlights

Date the policy was last checked or updated.
The Unlocking Research blog has clarification of funder policies and information on our negotiations with them.

https://unlockingresearch.blog.lib.cam.ac.uk/?p=151

https://unlockingresearch.blog.lib.cam.ac.uk/?p=337

https://unlockingresearch.blog.lib.cam.ac.uk/?p=285

https://unlockingresearch.blog.lib.cam.ac.uk/?p=525

https://unlockingresearch.blog.lib.cam.ac.uk/?p=528
**Huge** engagement programme:

- We have spoken with over 1,800 academics
- 75 individual information sessions about data sharing
- 37 different training events and workshops

*In 18 months since January 2015*
Frequently Asked Questions

Thanks to all the people who have attended our open data sessions during the first half of 2015. We were able to collect all the questions we received from researchers and research students and have prepared a list of Frequently Asked Questions.

We would also like to thank Ben Ryan from the EPSRC for carefully checking our advice and for his comments that helped us improve the answers.

We have arranged the list of FAQ into eight categories, thematically grouped:

What needs to be shared and when?
Questions about exemptions from research data sharing
Questions about resources to support research data sharing
Questions about data repositories and linking to data
Questions about personal/sensitive/commercial data
Questions about source code sharing

http://www.data.cam.ac.uk/data-faq

Huge engagement programme:

These are actual questions from our discussions with over 1,800 researchers
Various communication channels

January 2016 Research Data Management Newsletter

1. Wellcome Trust and Cancer Research UK coming to Cambridge for a session on sharing
2. Workshop: Strategies for managing social media research data
3. Call for abstracts - are you using social media data in your research?
4. Call for access to the RCUK Research Data Facility
5. Summer School of Research Data Science - applications open
6. Advances in Data Modelling and Knowledge Representation - call for papers
7. Ethics of Big Data in Practice - Lent Term workshop series
8. Call for researchers working with data

Twitter: @CamOpenData
Over 900 followers!

Newsletter: over 1,700 subscribers!

http://www.data.cam.ac.uk/datanews
Various communication channels

Share your data

- info@data.cam.ac.uk
- www.data.cam.ac.uk
- @CamOpenData

Research Data Services:
- Data repository
- Funding policy information
- Data management plan assistance

www.lib.cam.ac.uk/librarians/oa/oa_promo.html
DEMOCRATIC approach to RDM

Empower researchers to shape RDM services
Let researchers shape the services

1. What should research data management support look like at the University of Cambridge?

Quick questionnaire

Please help us shape research data management support available at the University of Cambridge. We have prepared a short questionnaire, and we would greatly appreciate your feedback. The questionnaire takes 1 min to fill in.
Example 2: is data management & sharing a waste of time?

Never dismiss questions and worries + be honest

“Open Data – moving science forward or a waste of money & time?”

Video:
https://upload.sms.cam.ac.uk/media/2113537

https://storify.com/CamOpenData/odpd15

https://unlockingresearch.blog.lib.cam.ac.uk/?p=392
Example 3: should Cambridge do more to encourage Open Research?

Let researchers shape the policy
They are the ones to adhere to it

The Office of Scholarly Communication would like to consult researchers on a number of ideas geared towards helping researchers to make all of their work openly available and rewarding the quality of the whole research process. We count on YOU to tell us what current problems in academic research are and what ideas would be useful to include in a planning document.

Everyone is welcome to come to this event and contribute their ideas.

https://www.eventbrite.com/e/improving-the-research-process-discussing-an-open-research-position-tickets-25391407383
Example 4: invitation to be part of the RDM Project Group at Cambridge

Let researchers work with you (empower the champions)

And take advice from people who know better

Become a member of the Research Data Management Project Group

Who can apply to be a member of the Group?

- Any member of the University of Cambridge with interest in research data management and sharing who is willing to work together towards improving RDM practices at the University
Did it work?
Researchers’ feedback on our services
In a bit more than a year 10X more data submissions than during a decade 2005 - 2015
Behaviour change – data shared earlier and earlier

• Researchers used to panic and share their data at the last minute (compliance requirements)

• More and more frequently:
  • Data shared before publication – access for peer-reviewers
  • Placeholder requests for data in preparation

• Data sharing seems to be more and more embedded in the publication process
Anonymous feedback:

“*I thought this was the best workshop I've attended so far. Very professional delivery, good balance of participant interaction.*”

“*Course should be mandatory* and done at the beginning of the year. It was excellent - thank you. *I wish I did it at the beginning.*”
“I really think it’s very important what you are doing, it is the future and you are proactively embracing it + helping shape it in this community which badly needs it, we are very grateful as a community for what you are doing.”

Group Leader, Department of Genetics

“I am really happy I got in touch with you. Your help was extremely important to me, as I previously had no idea how to do this. Thank you so much.”

Postdoc, Department of Psychology

“I meant to send this to you last week – the first paper that cites 'Kymata' data directly in the references has just been published!

This is a direct result of my discussions with you about the best way to cite this data in a paper – and I think it looks really good, so thanks again for your advice :)

Group Leader, Department of Psychology
And over 40 applications to join the RDM Project Group!
Collaboration and open communication: key to succeed with RDM provisions
Thank you

Questions: mt446@cam.ac.uk

www.data.cam.ac.uk

@martateperek
@CamOpenData

https://unlockingresearch.blog.lib.cam.ac.uk/