Research data sharing at the University of Cambridge

Presented by Dr Lauren Cadwallader
16th June 2016

NFAIS Open Science Workshop
This session will cover:

- A little bit about the University of Cambridge
- The need for central RDM services
- Support services for research data
- Workflows for data sharing
- Researchers reactions
- Support from publishers
- The future
• 18,000 students; 9,000 staff
• 6 Schools:
  – Arts and Humanities
  – Biological Sciences
  – Clinical Medicine
  – Humanities and Social Sciences
  – Physical Sciences
  – Technology
• 150 Departments and Institutes
• 31 Colleges
• 141 Libraries
• Extremely fragmented
• Local support solutions
  – variable quality
• Whole range of research data types and requirements for RDM support
Benefits of having central RDM support services
Benefits for researchers:

• Baseline support available to everyone, regardless of their local provisions

• Researchers can get extra exposure and benefits for sharing of their data

• Researchers can easily comply with funders’ policies
Benefits for funders:

• Assurance that research data resulting from their project is properly handled

• Assurance that research data from funded projects is shared to maximise their impact and return on investment

• Assurance that funders are properly acknowledged and get exposure for their funded projects

• Research data from funded projects will be available long-term
Benefits for the University:

• Assurance that researchers adhere to good RDM practice (research integrity agenda)
• Ensure University’s compliance with funders’ requirements for RDM
• Ensure University’s leadership in RDM good practice
• Cost-efficiency and sustainability of centrally-provided solutions
RDM services in Cambridge
RDM services at Cambridge:

- Information
- Consultancy on data management
- Data management plan support service
- Training
- Policy development, discussions with funders
- Data repository (free up to 20GB)
- Advocacy and outreach
Data Management Consultancy Request Form

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
Advice on preparing a data management plan

Would you like to get advice on your data management plan? We are here to help. Please send us your data management plan by using the form below and we will get back to you within five working days.

Important notes:

By using this service you certify that you have read you funder's policies on research data management and prepared your data management plan in accordance with these policies.

Funders' guidance on data management plans:

Specific guidance on funders' requirements for data management plans is available below – you need to follow funders' guidance when preparing a data management plan.

- AHRC
- BBSRC
- EPSRC
- ESRC
- MRC
- NERC
- Royal Society
- STFC
- Wellcome Trust

Example data management plans:

We also have some example data management plans which might be useful when preparing your first data management plan:

- AHRC – from the University of Bristol
- ESRC (Social Sciences) – from the University of Leeds
- ESRC (Psychology) – from the University of York
- MRC – from the University of Bristol
- Wellcome Trust – from London School of Hygiene and Tropical Medicine

www.data.cam.ac.uk/DMPsupport
<table>
<thead>
<tr>
<th>Date</th>
<th>Event Title</th>
<th>Date &amp; Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 Jun</td>
<td>Research Integrity</td>
<td>Wednesday 15 June, 10:00 AM - 11:00 AM</td>
<td>Mill Lane Lecture Room 5 - Mill Lane, Cambridge</td>
</tr>
<tr>
<td>16 Jun</td>
<td>Scholarship in the digital age – a survival guide for researchers</td>
<td>Thursday 16 June, 11:30 - 15:30</td>
<td>S1, Alison Richard Building, West Road, University of Cambridge</td>
</tr>
<tr>
<td>23 Jun</td>
<td>Working towards an Open Research future</td>
<td>Thursday 23 June, 14:00 to 17:30</td>
<td>The Cambridge Union - 9A Bridge Street, Cambridge, CB2 1UB</td>
</tr>
<tr>
<td>26 Jul</td>
<td>Research Data Management for Librarians</td>
<td>Tuesday 26 July, 14:30-16:30</td>
<td>Mill Lane Lecture Room 5 - 1 Mill Lane, Cambridge</td>
</tr>
<tr>
<td>2 Aug</td>
<td>Altmetrics: A Primer for Librarians</td>
<td>Tuesday 2 August, 14:30-15:30</td>
<td>Mill Lane Lecture Room 5 - 1 Mill Lane, Cambridge</td>
</tr>
</tbody>
</table>

[www.osc.cam.ac.uk/events](http://www.osc.cam.ac.uk/events)
University of Cambridge Research Data Management Policy Framework

Last updated: 23 April 2015

Background

1. The University of Cambridge is committed to disseminating its research and scholarship as widely as possible. In keeping with that commitment, it supports the principle that the results of its research that have been publicly funded should be freely accessible, and therefore supports its staff in making their research available.

http://www.data.cam.ac.uk/university-policy
### Key Policy Highlights

#### MRC
- **Read the policy**
- 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.

#### NERC
- **Read the policy**
- Possible sanctions: "those who do not meet the data management requirements [...] risk having award payments withheld or becoming ineligible for future funding"
- Data needs to be deposited into a NERC data centre within 2 years of collection.

#### NIH
- **Read the policy**
- "Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health."

#### NIHR
- **Read the statement**
- "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
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.
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

These are actual questions taken from our discussions with over 1,800 researchers

http://www.data.cam.ac.uk/data-faq
Advocacy and outreach

Twitter: @CamOpenData
Over 900 followers!

Newsletter: over 1,700 subscribers!

January 2016 Research Data Management Newsletter

1. Wellcome Trust and Cancer Research UK coming to Cambridge to discuss data 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
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
How it all works in practice:
Workflows for data sharing
Easy to use webform to upload data
Non-positive data can be shared as well

www.data.cam.ac.uk/upload
Behaviour change:

• Researchers used to panic and share their data to comply with funders’ requirements at the last stage
• Now they share data before publication
• More and more frequently they request placeholder for data in preparation

[www.data.cam.ac.uk/upload]
Data curation by data service team:

• Check for restricted/confidential information
• Check the content and relevance of files (if software available)
• Check if metadata description is sufficient and correct
• Check if funders acknowledged properly
• Record administrative metadata
### Key elements of a data record in Apollo

#### Guidance on citation

Every dataset gets a DOI

#### Reference to publication

Every dataset gets a DOI

#### Sponsors acknowledged

Every dataset gets a DOI

#### Every dataset has a licence

Every dataset gets a DOI

---

**Research Data Supporting "DNA brick self-assembly with an off-lattice potential"**

- **Citation**
  - Reinhardt, A., & Frenkel, D. Research Data Supporting "DNA brick self-assembly with an off-lattice potential" (dataset). [http://dx.doi.org/10.17863/CAM.135](http://dx.doi.org/10.17863/CAM.135)

- **Description**
  - Umbrella sampling weights for structures assembled with an off-lattice potential. The DNA sequences and patch pairings associated with each structure are also included. DOI:10.1039/C6SM01031H

- **File formats and software**
  - Any text editor. The Unix end-of-line convention is used.

- **Subjects**
  - DNA bricks, self-assembly, umbrella sampling

- **Relationships**
  - Publication Reference: 10.1039/C6SM01031H

- **Sponsorship**
  - EPSRC (EP/1001352/1)
  - Center for Functional Nanomaterials: US Department of Energy, Office of Basic Energy Sciences (DE-SC0012704)

- **Identifiers**
  - This record's DOI: [http://dx.doi.org/10.17863/CAM.135](http://dx.doi.org/10.17863/CAM.135)
  - Rights
    - Attribution-ShareAlike 4.0 International
    - License URL: [http://creativecommons.org/licenses/by-sa/4.0/](http://creativecommons.org/licenses/by-sa/4.0/)

- **Creative Commons**

---

[http://dx.doi.org/10.17863/CAM.135](http://dx.doi.org/10.17863/CAM.135)
Data record for article: Adaptive study of IL-2 dose frequency on regulatory T cells in type 1 diabetes (DILfrequency): a mechanistic, non-randomised, repeat dose, open label, response-adaptive study

Citation
https://www.repository.cam.ac.uk/handle/1810/251259

Description
Supporting data for a protocol paper "Adaptive study of IL-2 dose frequency on regulatory T cells in type 1 diabetes (DILfrequency): a mechanistic, non-randomised, repeat dose, open label, response-adaptive study", published in BMJ Open, repository URI - https://www.repository.cam.ac.uk/handle/1810/250599. The data available is the pilot data shown in Figure 2, the clinical flow cytometry (FACS) assay used to ascertain the co-primary endpoints of the study. These endpoints are: the frequency of T regulatory cells (Tregs), CD25 expression on Tregs, and the frequency of autoreactive effector T lymphocytes (Teffs). As the data is individual-level human subject data, it cannot be hosted on an Open Data repository, but is instead available on completion of a licence at the authors' website - https://www.gene.cimr.cam.ac.uk/data/. While the website should be checked for a more recent version, a representative licence (current at 30th September 2015) is available here.
This record is licensed under a JDRF/Wellcome Trust DIL Data Access Agreement licence.

File formats and software

https://www.repository.cam.ac.uk/handle/1810/251259
Researchers’ perspective
• 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!
• Postdoc left with all the data on his laptop
It takes time...

...and learning needed on both sides
Data services:

- Provide adequate support – requirements without support useless
- Researchers need to see the benefits
- HUGE engagement programme
- Bottom-up approach – involve researchers in creating the service (and requirements) – BE DEMOCRATIC
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.
Become a member of the Research Data Management Project Group

What is the purpose of this Group?

- Review, evaluate and improve current provisions for RDM and data sharing support at the University of Cambridge
- Propose and create new solutions for RDM and data sharing at the University of Cambridge
- Review and propose policies for RDM and sharing
- Monitor and evaluate RDM solutions available externally
- Ensure sustainability of research data management support models at the University of Cambridge

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
Example 3: is data sharing a waste of time?

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

Unlocking Research
University of Cambridge Office of Scholarly Communication

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

On 4 November 2015, the Research Data Facility at Cambridge University invited some inspirational leaders in the area of research data management and asked them to address the question: “Is open data moving science forward or a waste of money & time?”. Below are Dr. Marta Teperik’s impressions from the event.

Great discussion

Want to initiate a thought-provoking discussion on a controversial subject? The recipe is simple: invite inspirational leaders, bright people with curious minds and have an excellent chair. The outcome is guaranteed.

https://unlockingresearch.blog.lib.cam.ac.uk/?p=392
Example 4: should Cambridge have a position on Open Research?

Improving the research process - discussing an 'open research' position

By: Office of Scholarly Communication

Could this solve some of the current problems experienced in the academic sector?

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

https://unlockingresearch.blog.lib.cam.ac.uk/?p=392
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

April 2015: Launch of the easy-to-use data upload form
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 must admit I feel greatly re-assured that there is a service and a specialist team for this [data management plan support]. “

Senior Lecturer
Department of Theoretical and Applied Linguistics

“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.

I'll let you know the result of this application, around the end of the year. :)

Postdoc, Department of Psychology
Collaboration and communication: key to succeed with RDM provisions
What could publishers do?
Collaboration with:

- ‘Scientific Data’ (events)
- Cambridge University Press (members of RDM Project Group)

Getting published in Nature's 'Scientific Data' journal

Did you know that Nature Publishing Group has a dedicated journal 'Scientific Data' for publishing scientific data? An excellent way to get citations for the research data that you share. Articles in 'Scientific Data' can describe data from new or published studies, and can be published alongside traditional research works.

Come to our event on 11 January to talk directly to 'Scientific Data' editors and to learn about getting your data published and cited.

Poor practice – Science

“I discussed your question with my editorial colleagues, and our provisional response is that we would prefer you to take the cautious option of keeping the metadata in the ‘dark archive’ until the date of publication. We appreciate, though, that this is non-ideal from your point of view, and will now be discussing the question with our Office of Public Programs in Washington, to see whether we can accommodate your preferred procedure.”

https://unlockingresearch.blog.lib.cam.ac.uk/
• Clear expectations that data should be shared via repositories
  – good practice from ‘Scientific Data’

• Data scientists as reviewers (e.g. bioinformaticians)
  – Plus in-house data specialists
The ideal world in 5-10 years time
• We make a proper use of internet

• Impact factor of journals as a measure of success forgotten

• Peer-review directly in repositories

• Researchers rewarded for quality research directly by their peers
Thank you

Questions: info@data.cam.ac.uk
lc340@cam.ac.uk
www.data.cam.ac.uk

@CamOpenData