

Everything you need to know about data management plans

Marta Teperek

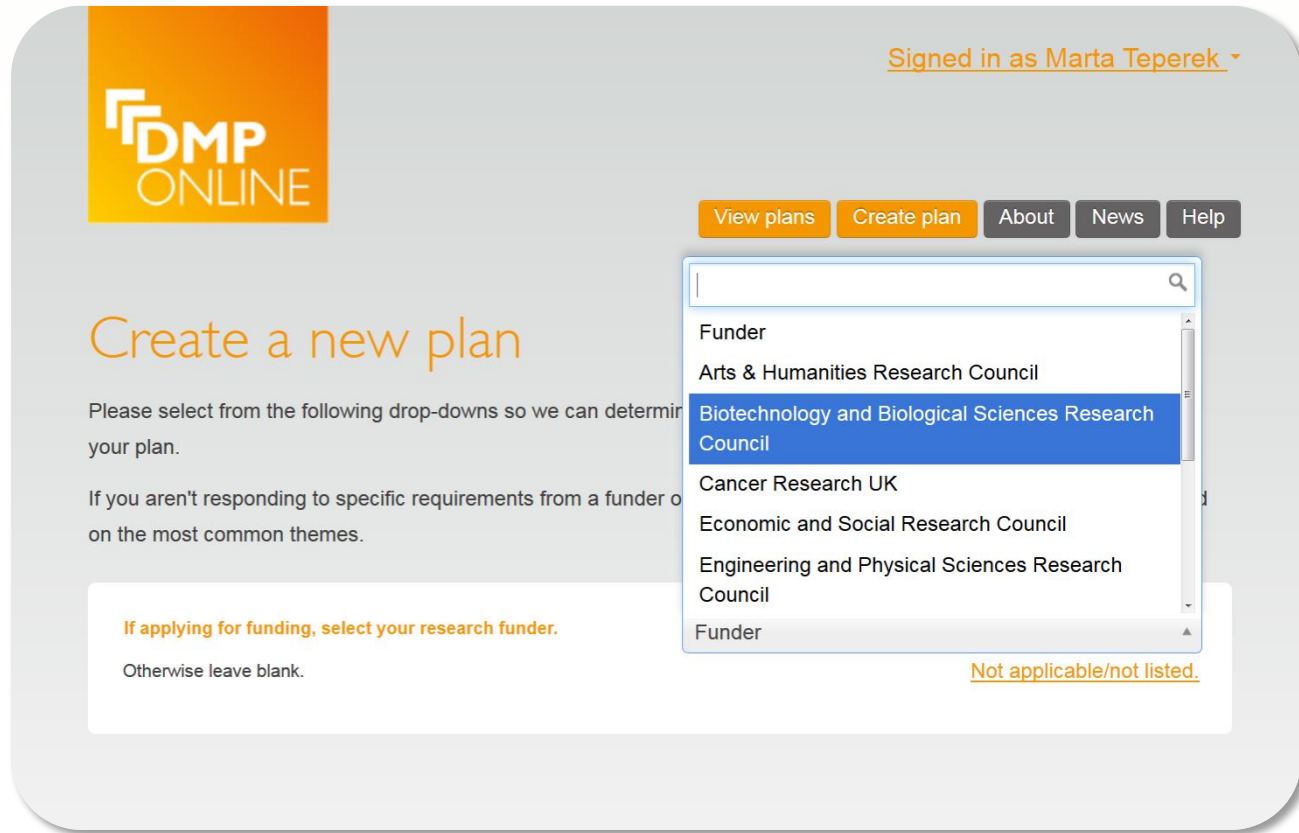
Office of Scholarly Communication, University of Cambridge

mt446@cam.ac.uk

 @martatepererek

16 November 2016

Does everyone have an account with DMPonline?



The screenshot shows the DMPonline website interface. In the top right corner, it says "Signed in as Marta Teperek". Below the navigation bar, there are buttons for "View plans", "Create plan", "About", "News", and "Help". The main heading is "Create a new plan". Below this, there is a text box with instructions: "Please select from the following drop-downs so we can determine your plan. If you aren't responding to specific requirements from a funder on the most common themes." A dropdown menu is open, showing a list of funders: "Funder", "Arts & Humanities Research Council", "Biotechnology and Biological Sciences Research Council" (highlighted), "Cancer Research UK", "Economic and Social Research Council", "Engineering and Physical Sciences Research Council", and "Funder". Below the dropdown, there is a note: "If applying for funding, select your research funder. Otherwise leave blank." and a link: "Not applicable/not listed."

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

- Fire alarms
- Refreshments
- Comfort breaks
- Interactive session – computers needed in part 2
- Ask questions at any time
- Slides will be available

Outline of this session:

- **Part 1:** Introduction to data management plans:
 - Why?
 - What?
 - How?
- **Part 2:** Introduction to DMPonline
 - Create a data management plan using an online tool
 - Peer review your colleague's data management plan

Who are you?

- Your name and role
- Your department/institution
- Why did you come to this workshop?



Have you created a data plan before?



Why do we suddenly need to have data management plans?



“Publicly funded research data are a public good (...), which should be made openly available with as few restrictions as possible...”



Concordat on Open Research Data



Why do funders require this?

1. Ensure better data management practice (reproducible research)
2. Ensure that researchers are prepared to share their data

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/>

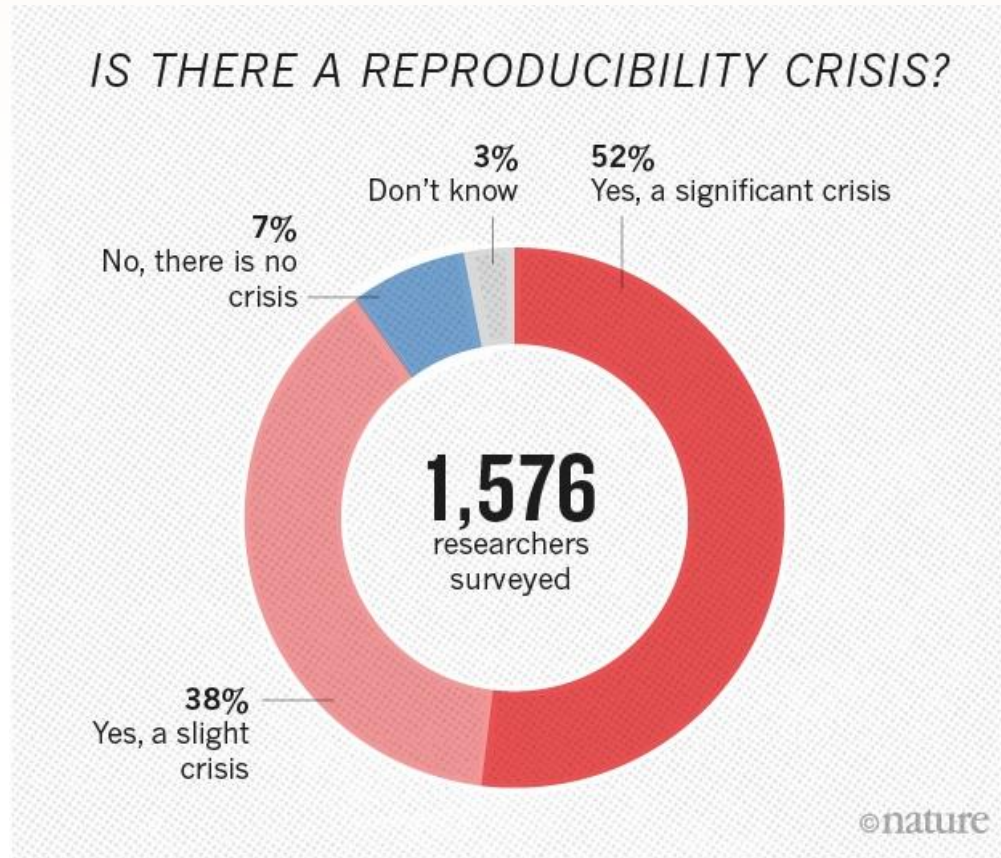
Immediate reaction:

- It would take me 5 years to find all my data!
- The postdoc who had the data had left
- Should be writing a paper

No surprise if researchers did not know how to effectively manage their data

- ...or time
- ...to understand my data





Research relies on the principle that we share
our findings

Without the supporting data, publications are
just claims

Report

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines,^{1,2,*} Arianne Y.K. Albert,³ Rose L. Andrew,¹
Florence Débarre,^{1,4} Dan G. Bock,¹ Michelle T. Franklin,^{1,5}
Kimberly J. Gilbert,¹ Jean-Sébastien Moore,^{1,6}
Sébastien Renault,¹ and Diana J. Rennison¹

sets (23%) were confirmed as extant. [Table 1](#) provides a breakdown of the data by year.

We used logistic regression to formally investigate the relationships between the age of the paper and (1) the probability

*Correspondence: vines@zoology.ubc.ca

Report

The Availability of Research Data Declines Rapidly with Article Age

- Data availability decreases by **17% per year**
- Chance of email address working decreases by **7% per year**

Report

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines,^{1,2,*} Arianne Y.K. Albert,³ Rose L. Andrew,¹
Florence Débarre,^{1,4} Dan G. Bock,¹ Michelle T. Franklin,^{1,5}
Kimberly J. Gilbert,¹ Jean-Sébastien Moore,^{1,6}
Sébastien Renaut,¹ and Diana J. Rennison¹

sets (23%) were confirmed as extant. [Table 1](#) provides a breakdown of the data by year.

We used logistic regression to formally investigate the relationships between the age of the paper and (1) the probability

*Correspondence: vines@zoology.ubc.ca

Accession Numbers

The analysis code and data are available on Dryad under DOI number [10.5061/dryad.q3g37](https://doi.org/10.5061/dryad.q3g37).

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.cub.2013.11.014>.

Data management plan = assurance to the funder:

- You are aware of their data management and sharing expectations
- You will manage you data well
- You will be prepared to share your data
- You will make appropriate resource allocation for this
- Selfish reason: you will find your data easily

We got the 'why'
Now the 'what'

What needs to be covered in a data
management plan?

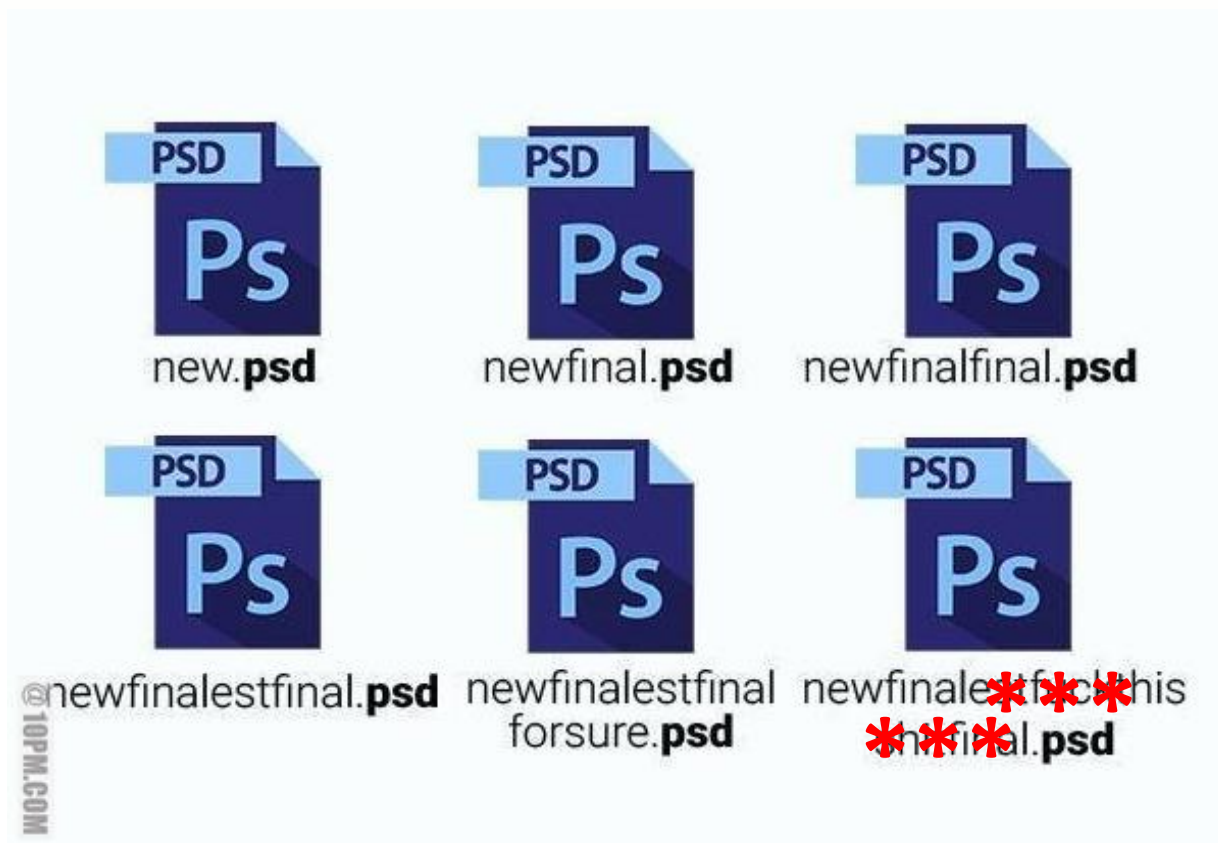
Core elements of a data management plan:

- Identify the types of data you are working with
- Decide on your data organisation strategy and data standards
- Day to day management of data
- What are your plans for data sharing?
- Will you experience any problems with data sharing?
- Will you require any additional resources?

Identify the type of data you are working with

- Genomic data?
- Proteomic data?
- Patient data?
- Will you generate your own data?
- Will you re-use somebody else's data?

Decide on your data organisation strategy and data standards



Example A

Documents library

PhD data

- 25July
- Documents
- Experiments
- Experiments2
- From desktop
- Important
- Other
- PhD
- Talks
- Experiment 1
- Lab meeting FINAL
- Meeting notes
- My talk
- Paper submission
- PhD revised

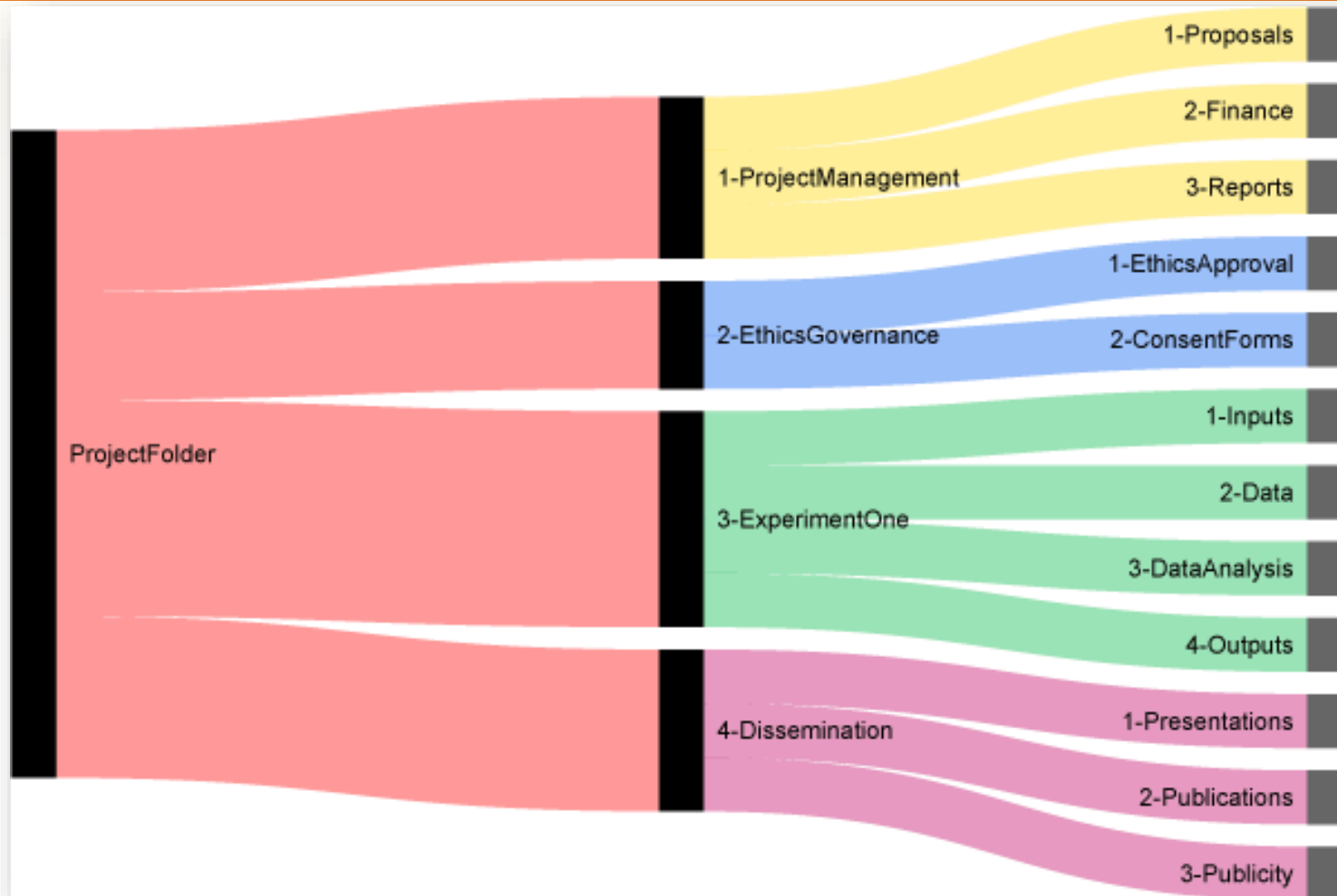
Example B

Documents library

PhD data2

- Conferences
- Downloaded publications
- Experimental data
- Financial documents
- PhD thesis
- Presentations
- Protocols
- Reagents
- Reports
- Training

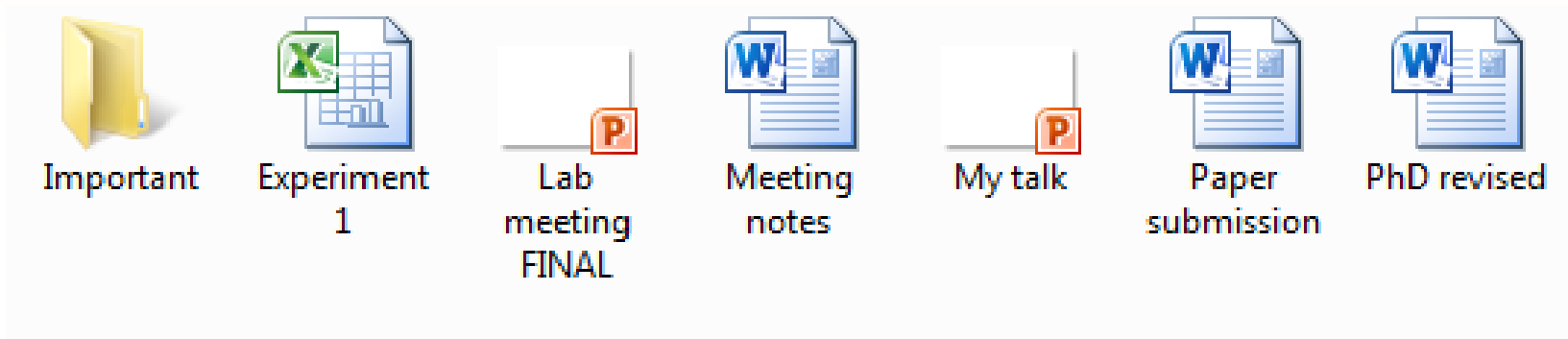
Example of data organisation structure



Data organisation strategy should be:

- Consistent
- Meaningful to you and your colleagues
- Allow you to find files easily
- Including physical samples as well!

Will you use any file naming conventions?



In 3 years time would you know what these are?

TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:

GDL - **TILSDocNaming** - **V1** - **20090612**.docx

A prefix shows the document type

The document title describes the content

The version number

The date in the format yyyymmdd

Will you adhere to any discipline-specific metadata standards?



Search Standards

Search

Search

Reset

Advanced

<https://biosharing.org/standards/>

How will you store your data during the project duration?

- Is your data going to be safe?
- Will you back up your data?

Unsure? Help available:

<http://www.information-management.uis.cam.ac.uk/research-data-compliance>

Funder *

Research Project overview & Types of Data *

Does the data contain any Personal Identifiable Data? *

Yes - Non Medical

Yes - Medical

No

Names, Addresses, Email Addresses, DNA, Medical Data, Data of a sensitive nature

Where is the data currently stored? *

Size of data (Appx) *

Submit

Cloud services comparison



Space included	1 TB	Unlimited	15GB
Price (per annum)	Free	£55-66	Free (1 TB is £95.88)
File history	Yes	30 days	Yes
Support available	Webpages from UIS	Webpages from UIS	Unsupported
Where are files stored?	Within EU	Within EU	Anywhere
Live editing	Yes – dependent on browser	Only with an Office 365 licence	Yes
What types of data can be shared?	Unclassified, Cambridge only, or confidential data	Unclassified, Cambridge only, or confidential data	Unclassified or public data

How are you going to ensure proper data management?

- Will there be any quality checks for data collection/analysis in your group?
- Who will be responsible for what?

Will you share your data?

- When will you share your data?
- How are you going to share your data?
- Are your plans in line with your funder's expectations?

Most funders expect:

- Make your data available upon publication
- Store data for (at least) 10 years
- Describe your data
- Deposit your data in suitable data repositories and add a link to your data in your publication

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

<p>EPSRC</p> <p>Read the EPSRC Expectations</p>	<p>EPSRC has the strictest policy on research data sharing. We have prepared separate, dedicated guidelines to help you achieve compliance with the EPSRC expectations.</p> <p>Additionally, please have a look at our list of FAQ, which was developed in consultation with researchers at the University of Cambridge, and with Ben Ryan from the EPSRC.</p> <p>As a result of our discussions with Ben Ryan we have also published a blog post clarifying EPSRC policy on data management and sharing.</p>	September 2016
<p>ESRC</p> <p>Read the policy</p>	<p>Research data has to be deposited within three months of the end of the award. ESRC UK Data Service (UKDS) is the approved data repository for ESRC-funded research.</p> <p>Where a delay in dissemination of deposited data is needed to allow grant holders to publish their research findings, an embargo period can be applied to the data.</p> <p>Researchers have to consider legal, ethical and commercial constraints on release of research data at the initiation of the research process and throughout both the research and data life cycles, which shall be described in the data management plan.</p> <p>The ESRC refers researchers to these pages for information on obtaining a proper consent for data collection from human participants.</p>	September 2016
<p>Isaac Newton Trust</p>	<p>At the moment, the Isaac Newton Trust does not have dedicated research data policies in place.</p>	September 2016
<p>Leverhulme Trust</p>	<p>The Trust has currently no dedicated research data policies in place.</p>	September 2016
<p>MRC</p> <p>Read the policy</p>	<p>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</p>	September 2016

Key policy highlights

Date the policy was last checked or updated.

What is a repository?

- A place where things can be stored and shared



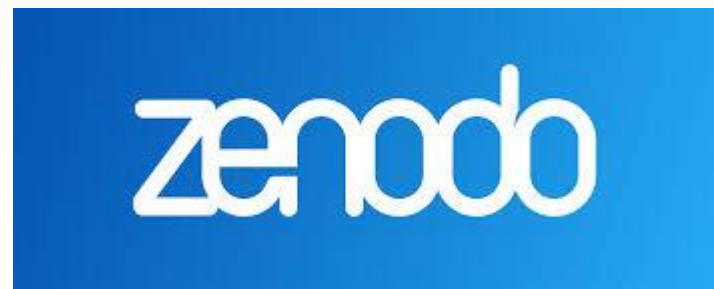
There are different kinds of repositories:

- for 'everything'
- for datasets
- for software
- for protocols
- for institutions

And many, many more!

Repositories for ‘everything’

<https://zenodo.org/>



Publications (30524)

Books (2301) Book sections (0) Conference papers (9432) Journal articles (13701) Patents (3) Preprints (204)
Project Deliverables (122) Project Milestones (12) Proposals (28) Reports (653) Theses (339) Technical notes (79)
Working papers (222) Other (867)

Posters (481)

Presentations (961)

Datasets (2416)

Images (457)

Figures (338) Plots (10) Drawings (11) Diagrams (27) Photos (23) Other (48)

Videos/Audio (150)

Software (8463)

Lessons (72)

Repositories for datasets

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://www.re3data.org/>

General purpose



Discipline-specific



Repositories for software

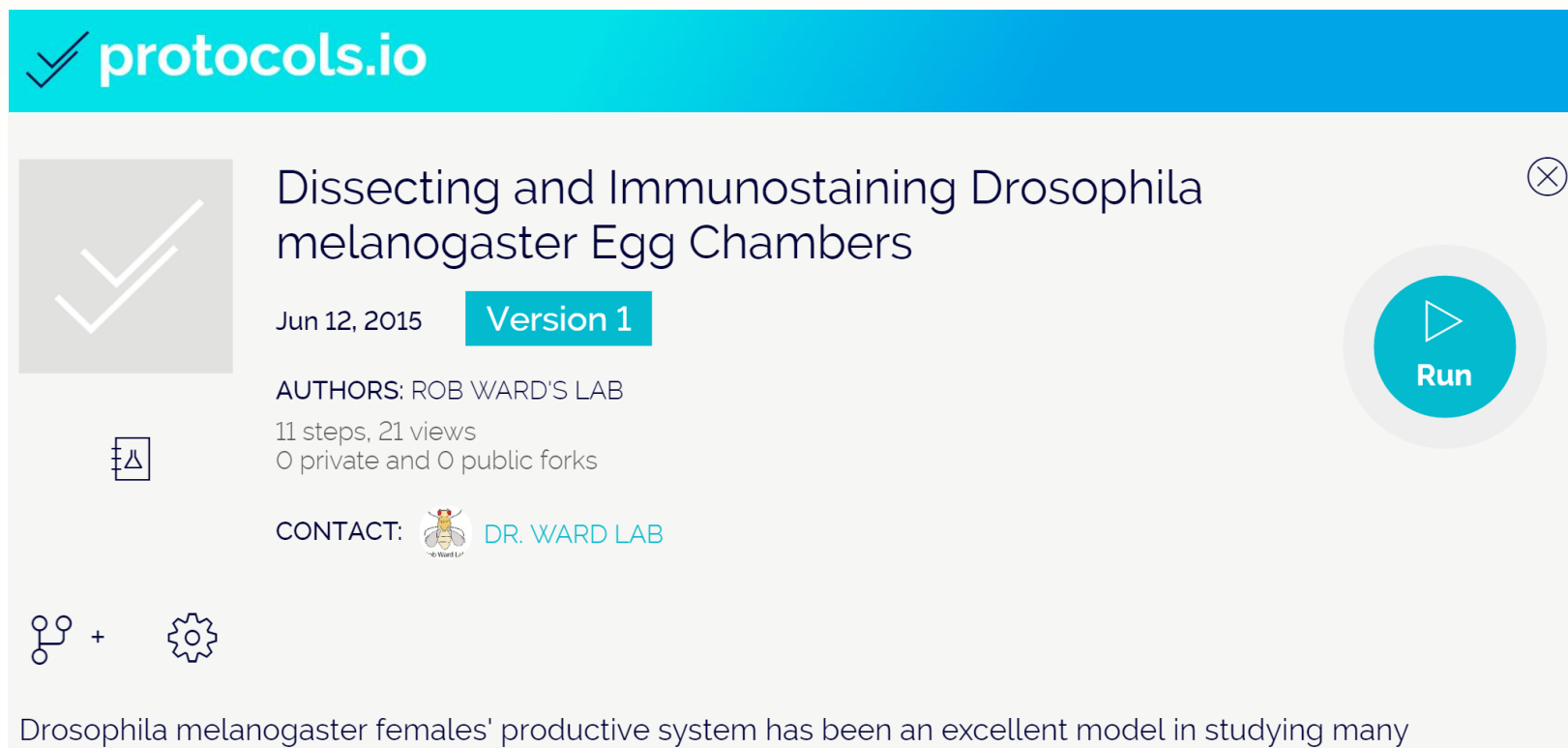
GitHub



+

zenodo

Repositories for protocols



The screenshot shows a protocol page on protocols.io. The header is a blue bar with a white checkmark icon and the text "protocols.io". Below the header, there is a grey square icon with a white checkmark. To the right of the icon, the title "Dissecting and Immunostaining Drosophila melanogaster Egg Chambers" is displayed. Below the title, the date "Jun 12, 2015" and a blue button labeled "Version 1" are shown. The authors are listed as "AUTHORS: ROB WARD'S LAB". Below the authors, it says "11 steps, 21 views" and "0 private and 0 public forks". The contact information is "CONTACT: DR. WARD LAB" with a small bee icon. On the right side, there is a circular button with a play icon and the word "Run". At the bottom of the screenshot, there is a partial text: "Drosophila melanogaster females' productive system has been an excellent model in studying many".

<https://www.protocols.io/view/Dissecting-and-Immunostaining-Drosophila-melanogas-mw9wh5>

<https://www.protocols.io/>

Repositories for images – under construction



Image Data Repository

OVERVIEW

Welcome to the Image Data Repository (IDR). This online, public resource seeks to store, integrate and serve up image datasets that have been collected and are continuing to receive existing and newly created “reference” image datasets that are important because they will be frequently accessed and cited or because they can serve as a basis for re-analysis and



Take a look at the data...



<http://idr-demo.openmicroscopy.org/about/>

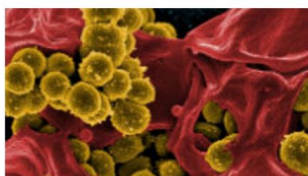
Repositories for institutions

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.



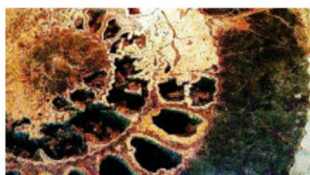
School of Arts and Humanities



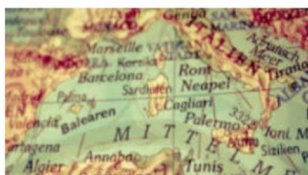
School of Clinical Medicine



School of Technology



School of the Biological Sciences



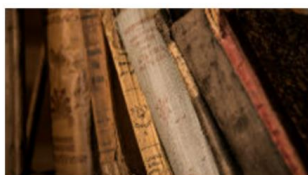
School of the Humanities and Social Sciences



School of the Physical Sciences



Colleges



Other Communities

Share your
research



Microfluidic Droplet-Facilitated Hierarchical Assembly for Dual Cargo Loading and Synergistic Delivery

Ziyi Yu[†], Yu Zheng[§], Richard M. Parker[†], Yang Lan[§], Yuchao Wu[§], Roger J. Coulston[§], Jing Zhang[†], Oren A. Scherman^{*§}, and Chris Abell^{*†}

Corresponding authors:

*E-mail: ca26@cam.ac.uk, *E-mail: oas23@cam.ac.uk.

Or





Additional data related to this publication is available at the University of Cambridge data repository (<https://www.repository.cam.ac.uk/handle/1810/248506>).

<http://dx.doi.org/10.1021/acsami.6b00661>

Data for article "Microfluidic Droplet-Facilitated Hierarchical Assembly for Dual Cargo Loading and Synergistic Delivery"



View / Open Files

-  [determination-critical-micelle-concentration-cmc-p1-micelles.zip](#) (application/zip, 227Kb)
-  [size-distribution-microdroplets.zip](#) (application/zip, 32Kb)
-  [tem-p1-micelles.zip](#) (application/zip, 11Mb)
-  [description-data.docx](#) (Microsoft Word, 18Kb)

Authors

Yu, Ziyi
Zheng, Yu

Citation

Yu, Z., Zheng, Y., Parker, R. M., Lan, Y., Wu, Y., Coulston, R. J., Zhang, J., et al. (2015). *Data for article "Microfluidic Droplet-Facilitated Hierarchical Assembly for Dual Cargo Loading and Synergistic Delivery"* [dataset]. <http://www.repository.cam.ac.uk/handle/1810/248506>

Description

Original or unprocessed data is provided in support of the article "Microfluidic Droplet-Facilitated Hierarchical Assembly for Dual Cargo Loading and Synergistic Delivery". The data is structured into three folders named: (1) Determination of Critical Micelle Concentration (CMC) of P1 micelles; (2) TEM of P1 micelles; and (3) Size distribution of microdroplets.

File formats and software

.pdf; .jpg; .xls; .png; .zip

Relationships

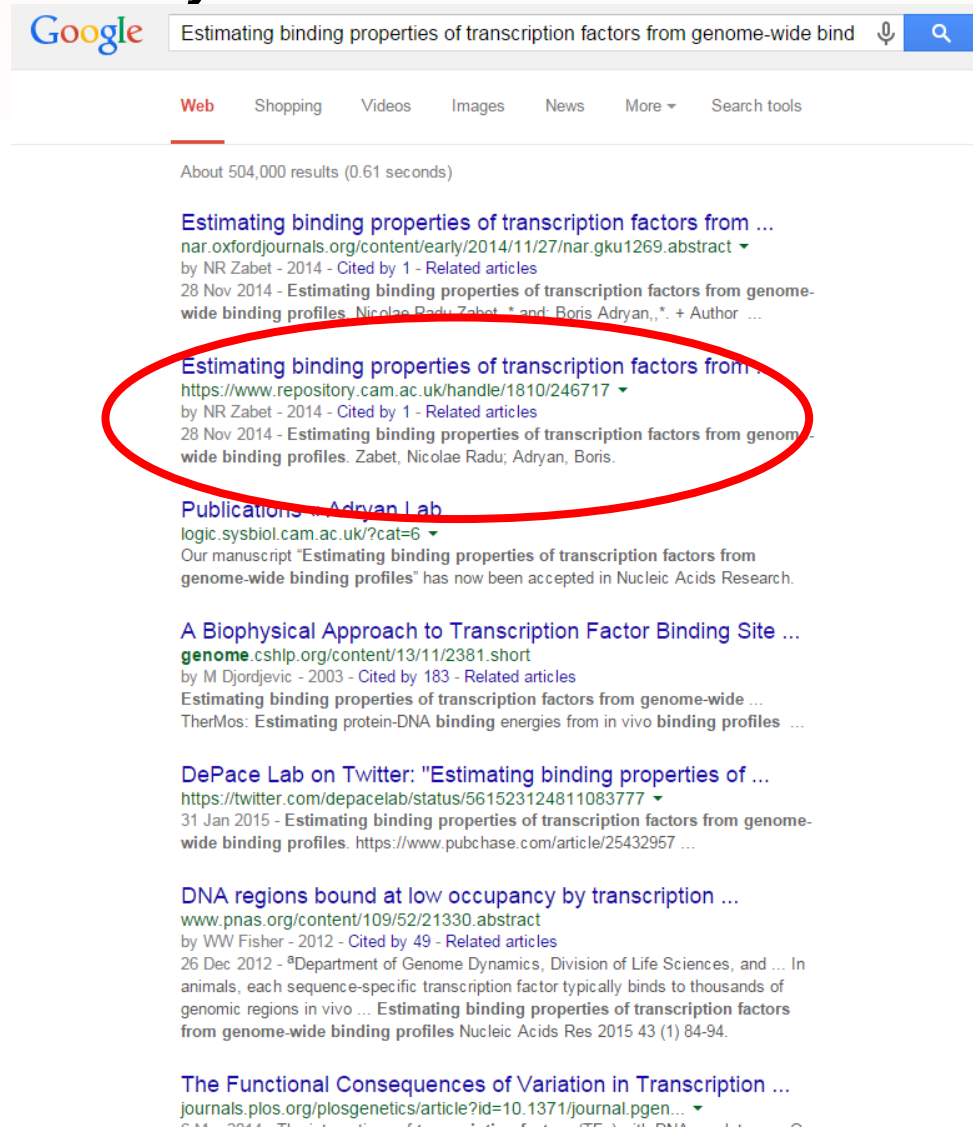
Publication Reference:

<http://dx.doi.org/10.1021/acsami.6b00661>

Sponsorship

<https://www.repository.cam.ac.uk/handle/1810/248506>

The repository is well indexed in Google



Google Estimating binding properties of transcription factors from genome-wide binding profiles

Web Shopping Videos Images News More Search tools

About 504,000 results (0.61 seconds)

Estimating binding properties of transcription factors from ...
nar.oxfordjournals.org/content/early/2014/11/27/nar.gku1269.abstract
by NR Zabet - 2014 - Cited by 1 - Related articles
28 Nov 2014 - Estimating binding properties of transcription factors from genome-wide binding profiles. Nicolae Radu Zabet, * and; Boris Adryan,*, + Author ...

Estimating binding properties of transcription factors from ...
https://www.repository.cam.ac.uk/handle/1810/246717
by NR Zabet - 2014 - Cited by 1 - Related articles
28 Nov 2014 - Estimating binding properties of transcription factors from genome-wide binding profiles. Zabet, Nicolae Radu; Adryan, Boris.

Publications - Adryan Lab
logic.sysbiol.cam.ac.uk/?cat=6
Our manuscript "Estimating binding properties of transcription factors from genome-wide binding profiles" has now been accepted in Nucleic Acids Research.

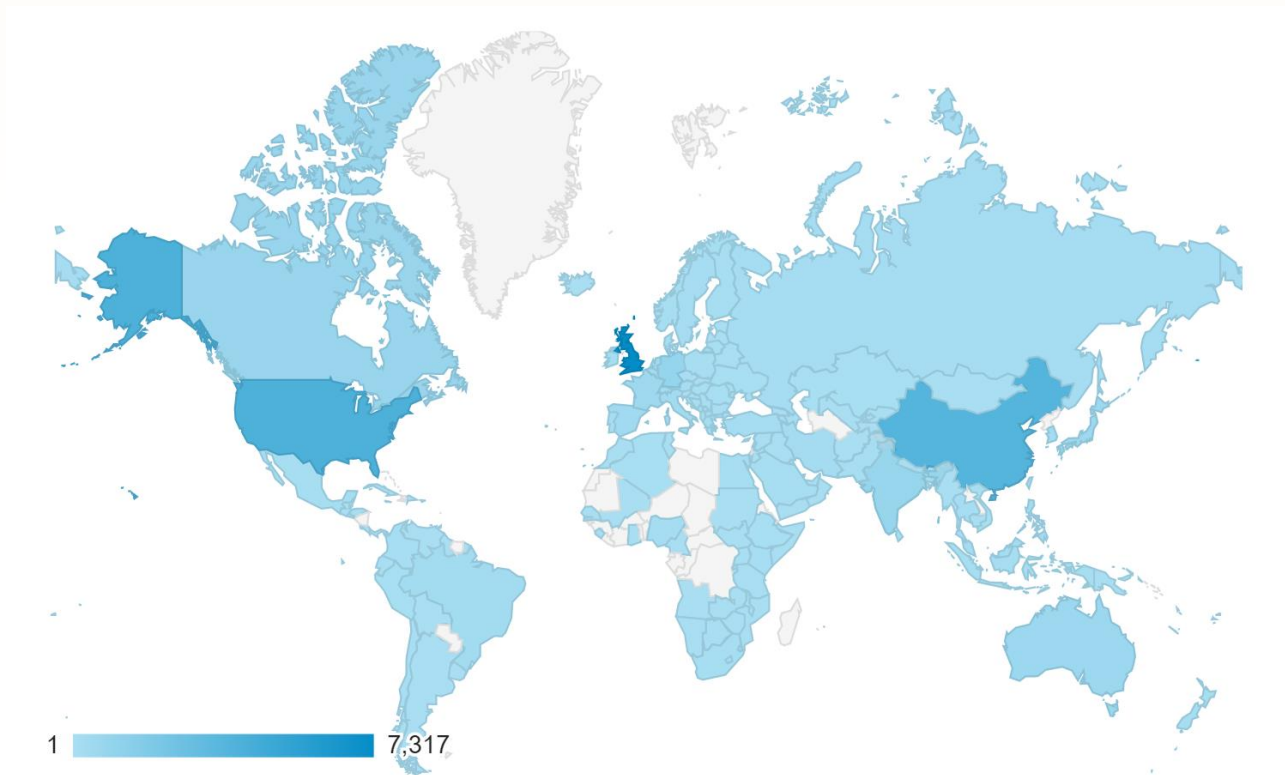
A Biophysical Approach to Transcription Factor Binding Site ...
genome.cshlp.org/content/13/11/2381.short
by M Djordjevic - 2003 - Cited by 183 - Related articles
Estimating binding properties of transcription factors from genome-wide ...
TherMos: Estimating protein-DNA binding energies from in vivo binding profiles ...

DePace Lab on Twitter: "Estimating binding properties of ...
https://twitter.com/depacelab/status/561523124811083777
31 Jan 2015 - Estimating binding properties of transcription factors from genome-wide binding profiles. https://www.pubchase.com/article/25432957 ...

DNA regions bound at low occupancy by transcription ...
www.pnas.org/content/109/52/21330.abstract
by WW Fisher - 2012 - Cited by 49 - Related articles
26 Dec 2012 - ^aDepartment of Genome Dynamics, Division of Life Sciences, and ... In animals, each sequence-specific transcription factor typically binds to thousands of genomic regions in vivo ... Estimating binding properties of transcription factors from genome-wide binding profiles Nucleic Acids Res 2015 43 (1) 84-94.

The Functional Consequences of Variation in Transcription ...
journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen...

Repository use 14 October 2016 – 13 November 2016



26,274 visits

If you share your data – use it to your benefit

- Persistent link(s) – DOIs - enable data citation
- Impress your funder – make your data discoverable and impactful – ideas:
 - indicate statement of work
 - deposit data in a public domain repository(ies)
 - provide a data statement to accompany the dataset
 - Wellcome Open Research data from Nature (<http://www.nature.com/sdata/>)
 - Wellcome Trust Open Research (<https://wellcomeopenresearch.org/>)
- providing reference on project/institutional websites
- publicising information about the data on social media
 - Cambridge does this for you

**Funders like statements like these
Make sure to include them in your data plan**



Wellcome Open Research

A new way for Wellcome-funded researchers to rapidly publish any results they think are worth sharing.

LEARN MORE

SUBMIT YOUR RESEARCH

<http://wellcomeopenresearch.org/>



SOFTWARE TOOL ARTICLE

SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics [version 1; referees: awaiting peer review]


✉ Przemyslaw Stempor, ✉ Julie Ahringer 

+ Author affiliations

+ Grant information

Abstract

Experiments involving high-throughput sequencing are widely used for analyses of chromatin function and gene expression. Common examples are the use of chromatin immunoprecipitation for the analysis of chromatin modifications or factor binding, enzymatic digestions for chromatin structure assays, and RNA sequencing to assess gene expression changes after biological perturbations. To investigate the pattern and abundance of coverage signals across regions of interest, data are often visualized as profile plots of average signal or stacked rows of signal in the form of heatmaps. We found that available plotting software was either slow and laborious or difficult to use by investigators with little computational training, which inhibited wide data exploration. To address this need, we developed SeqPlots, a user-friendly exploratory data analysis


METRICS

75

 VIEWS

9

 DOWNLOADS Get PDF Get XML Cite Export Track Email

<https://wellcomeopenresearch.org/articles/1-14/v1>

Will you experience any problems with sharing your data?

- Personal/sensitive data
- Commercially-confidential data
- Big data?

If so, you need to explain this to your funder from the very beginning.

Will you require any additional resources?

- people infrastructure: **data managers?**
- costs of active data storage: **consult your IT support**
- costs of licences for software to support data management
 - **Electronic Lab Notebooks?**
- costs of data ingestion by the repository/long-term preservation:
 - Repositories might charge to ensure sustainability
 - Cambridge is charging at £4/GB (for datasets above 20GB) – it is a one off charge

We got the 'why'
We got the 'what'
Let's look for the 'how'

Is there any help available?

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 your [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 your data management plan.

- [AHRC](#)
- [BBSRC](#)
- [EPSRC](#)
- [ESRC](#)
- [MRC](#)
- [NERC](#)
- [Royal Society](#)
- [STFC](#)
- [Wellcome Trust](#)

Funders' guidance

Template not there?

E-mail info@data.cam.ac.uk

Example data management plans

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

Funder's name *

- Select -

Deadline for the submission of the grant proposal *

Day Month Year 

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

Signed in as [Marta Teperek](#) ▾

DMP ONLINE

[View plans](#) [Create plan](#) [About](#) [News](#) [Help](#)

Create a new plan

Please select from the following drop-downs so we can determine your plan.

If you aren't responding to specific requirements from a funder or on the most common themes.

If applying for funding, select your research funder.

Otherwise leave blank.

[Not applicable/not listed.](#)

Funder

- Arts & Humanities Research Council
- Biotechnology and Biological Sciences Research Council**
- Cancer Research UK
- Economic and Social Research Council
- Engineering and Physical Sciences Research Council

Funder

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

- Introduction to DMPonline
 - Think of a project you want to apply with/you are working on now
 - Select the appropriate funder
 - Use DCC guidance
 - Answer the questions - **be prepared to share**

Review:

- Invite your colleague as a collaborator (with *edit* access)
- Review each other's plans and give suggestions
- Give feedback to each other
- Q&A

Today we have covered:

- Why funders require data management plans
- What to write in data management plans
- How to create them
- Practical experience of DMP online

14
NOV**Post-Publication Sharing: Publishing your Research Effectively for STEM PhD Students**

Monday, 14 November, 2016

J J Thomson Seminar Room, Maxwell Centre, Cavendish Laboratory, J J Thomson Avenue, Cambridge CB2 0HE

16
NOV**All you need to know about data management plans**

Wednesday, 16 November, 2016 at 12:00 - 14:00

Seminar Room G, 17 Mill Lane

17
NOV**Understanding Creative Commons, And Other Copyright Adventures**

Thursday, 17 November, 2016 at 14:30 - 16:30

Glass Room, Betty & Gordon Moore Library

18
NOV**Books: Publishing your Research Effectively for PhD Students in Humanities, Arts and Social Sciences**

Friday, 18 November, 2016 at 10:00 - 11:00

Lecture Room 4, 8 Mill Lane, Cambridge

18
NOV**Understanding Creative Commons, And Other Copyright Adventures**

Friday, 18 November, 2016 at 10:30 - 12:30

Glass Room, Betty & Gordon Moore Library

25
NOV**Post-Publication Sharing: Publishing your Research Effectively for PhD Students in Humanities, Arts and Social Sciences**

Friday, 25 November, 2016 at 10:00 - 11:00

Seminar Room G, 17 Mill Lane, Cambridge CB2 1RX

25
NOV**Managing Your Research Data**

Friday, 25 November, 2016 at 14:30 - 16:30

Glass Room, Betty & Gordon Moore Library

28
NOV**Wrangling Your Citations Like A Pro**

Monday, 28 November, 2016 at 14:30 - 16:30

Glass Room, Betty & Gordon Moore Library

30
NOV**Maximising Metrics With Your Research**

Wednesday, 30 November, 2016 at 10:30 - 12:30

Glass Room, Betty & Gordon Moore Library

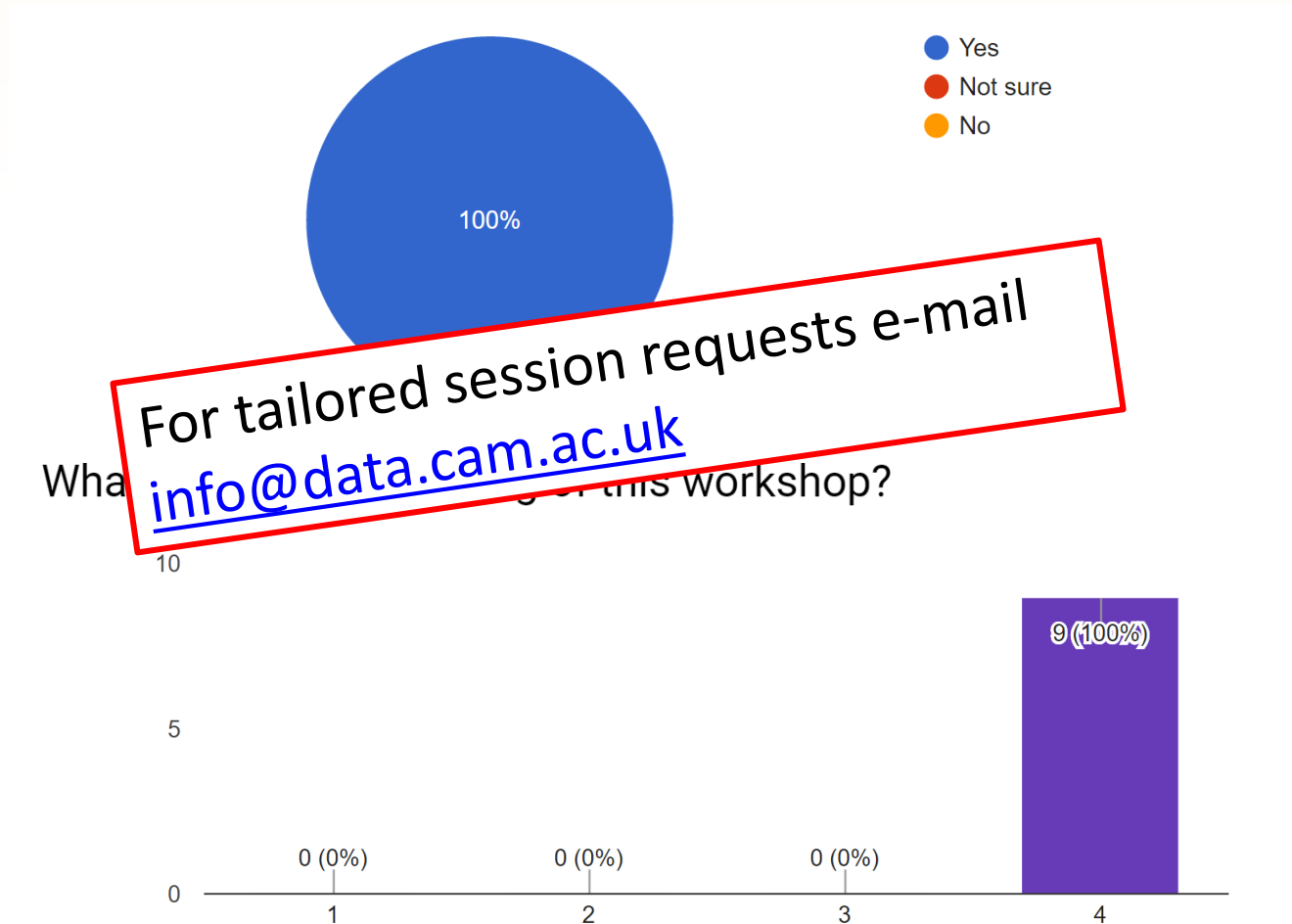
01
DEC**2nd Institute of Mathematics and its Applications Conference on the Mathematical Challenges of Big Data**

Thursday, 1 December, 2016 to Friday, 2 December, 2016

Mary Ward House, London

www.data.cam.ac.uk/events

Do you think this session would be useful for first year PhD students?

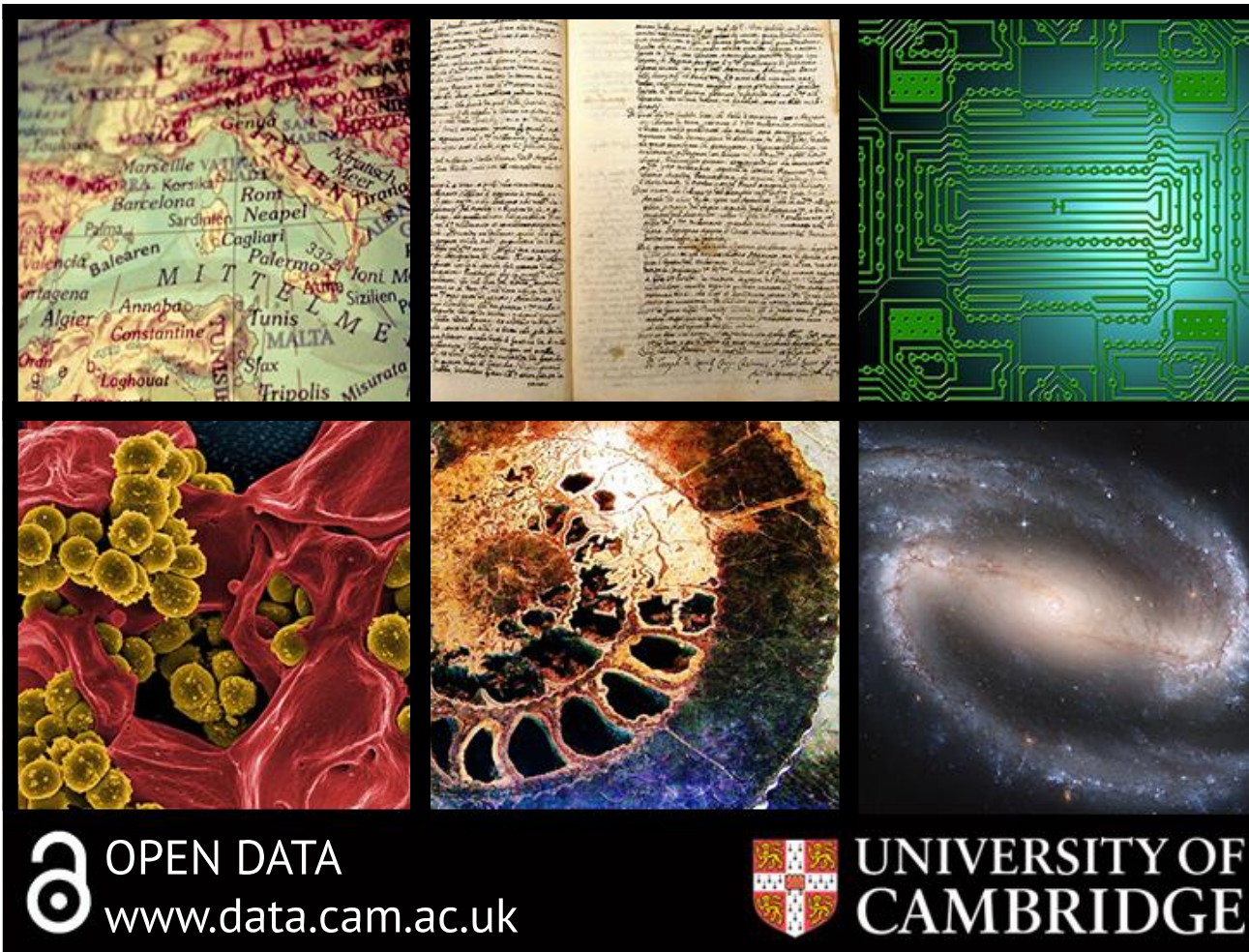


From RDM workshop for GSLS on 11 November

This is our first workshop on data management plans...

...please help us improve it

Take home message:



Thank you

Questions: info@data.cam.ac.uk

www.data.cam.ac.uk



@CamOpenData

<http://www.data.cam.ac.uk/newsletter/signup>

- Don't collect it!
- Gain informed, preferably open and written, consent
- Anonymise data
 - Remove identifiers
 - Aggregate results
 - Generalise a variable
 - Remove outliers
- Use managed access repositories

- Clinical School website:
 - Carolyn Read
- University Ethics website
 - Dr Rhys Morgan
- Our website (University resources)
- MRC guidelines
- ESRC guidelines

- Clinical School:
 - Compliance requirements: Carolyn Read
 - Storage requirements: Richard Bartlett
 - Contacts:
<https://www.medschl.cam.ac.uk/about/whos-who/>
- Other parts of the University:
 - Information Management team:
<http://www.information-management.uis.cam.ac.uk/>

Information Management team support form

To find out how you should be saving, sharing, backing up and securing your research data electronically, please fill in the following form.

A member of our team will be able to give you advice, as soon as possible.

Name *

Department/institution/organisation *

Contact Telephone *

Email Address *

Funder *

Research Project overview & Types of Data *

Or e-mail:

informationmanagement@uis.cam.ac.uk