Introduction to Research Data Management

Mary Kattuman, Claire Sewell, Marta Teperek

11/05/2016
Today:

Mixture of activities and talking

Introduction

1. Backup and exchange strategies
2. How to organise your data well
3. Data sharing
4. …how to avoid problems => data management plans

We will send you the slides
To start with…

- Do you have any questions about data management that you hope will be addressed during this workshop?
Part 1: Data backup and data exchange strategies
Disastrous data loss...

Department of Chemistry, University of Cambridge

Credit: Peter Murray-Rust
August 2011, CC-BY

---

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: XXXX or XXXX
(at [name]@cam.ac.uk)
Thank you!!
How much of your data would you lose if...?
How much of your data would you lose if…?

- your laptop got stolen
- your lab/office burnt
- you've lost your USB stick
- your portable hard drive got damaged
- data from your Dropbox/Googledrive account disappeared

5 mins
Backup strategies:

- Departmental backup system
- External drives
- Online backups

- At least two backups, at two different locations
At least 2 backups at 2 locations:

Every Monday morning

Everyday at 10am (automated!)

Free software to manage backups (there is plenty of free software):
http://www.2brightsparks.com/download-syncbackfree.html
At least 2 backups at 2 locations:

Store at home!

Your departmental server

Free software to manage backups (there is plenty of free software):
http://www.2brightsparks.com/download-syncbackfree.html
At least 2 backups at 2 locations:

Store at home!

Shiny new exciting data!

Your departmental server

Free software to manage backups (there is plenty of free software):
http://www.2brightsparks.com/download-syncbackfree.html
At least 2 backups at 2 locations:

Store at home!

Copy ASAP!

Your departmental server

Free software to manage backups (there is plenty of free software):
http://www.2brightsparks.com/download-syncbackfree.html
File sharing:

- Google Drive/Dropbox… - cautious!
  - Do not use cloud storage to store restricted data
- E-mail
- Website/Moodle
- Sharepoint
- FTP/SFTP
- University of Cambridge Microsoft
  OneDrive: 1TB of space for everyone

http://www.uis.cam.ac.uk/ees/onedrive
University OneDrive filespaces use the OneDrive for Business solution which gives you 1TB of cloud-based file storage space for your work and personal files. Find out how to access your online filestore, and start syncing your local files.

http://www.uis.cam.ac.uk/ees/onedrive

Questions: service-desk@uis.cam.ac.uk
Part 2: Data organisation
Data organisation:

- Examples A and B
- Which example is better, and why?
- What are good and bad features?
- Your own example – how can it be improved?

3 mins
Which example is better?
Data organisation:

- consistent
- meaningful to you and your colleagues
- allow you to find files easily

would you be able to easily get hold of your own data?
**Organisation of physical samples:**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Box #1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Name of the sample</td>
<td>date in lab book</td>
<td>quantity (if aliquots)</td>
<td>size (if aliquots)</td>
</tr>
<tr>
<td>3</td>
<td>Fibroblasts protein extract</td>
<td>23/03/2014</td>
<td>1 aliquot left</td>
<td>~40ul</td>
</tr>
<tr>
<td>4</td>
<td>Fibroblasts protein extract</td>
<td>05/04/2014</td>
<td>7</td>
<td>~40ul</td>
</tr>
<tr>
<td>5</td>
<td>Fibroblast protein extract for mass spec</td>
<td>19/08/2014</td>
<td>4</td>
<td>~50ul</td>
</tr>
<tr>
<td>6</td>
<td>Fibroblast RNA extract</td>
<td>20/09/2014</td>
<td>12</td>
<td>~10ul</td>
</tr>
<tr>
<td>7</td>
<td>Fibroblast RNA extract</td>
<td>23/09/2014</td>
<td>2 left</td>
<td>~10 ul</td>
</tr>
</tbody>
</table>
| 8 | Anti-actin antibody    | N/A                              | 18                    | ~5ul                                   | Abcam: ab8227
LOT: GR47300
Concentration: 0.6mg/ml
Rabbit polyclonal IgG |
Organisation of physical samples:

- create maps of your samples
  - can be simple Excel spreadsheets
  - and keep them up to date!

- reference your samples:
  - date in the lab books
  - supplier’s name/code

- add any relevant notes
File naming conventions – why matter?

Copyright: http://10pm.com/
File naming conventions – why matter?

Would you know in 3 years time what are all these?
File naming convention:

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

http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf
Part 3: Data sharing
What is your opinion?
It would be useful if research data underpinning publications was available
I (/my group) regularly share research data underpinning publications
Question: Why it might be a good idea to share data?
Open Access is a ‘good thing’:

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

Danny Kingsley @dannykay68 · Apr 18
NEGATIVE: the reality of all the studies into anti depressants. Negative results not published #force2016
Non-positive results need to be shared

p-value 0.05: who is going to publish their results?
Protection against misconduct

Stimulus-triggered fate conversion of somatic cells into pluripotency

Haruko Obokata, Teruhiko Wakayama, Yoshiki Sasai, Koji Kojima, Martin P. Vacanti, Shigenobu Yonemura, Hitoshi Niwa, Masayuki Yamato & Charles A. Vacanti

Bidirectional developmental potential in reprogrammed cells with acquired pluripotency

Haruko Obokata, Yoshiki Sasai, Hitoshi Niwa, Mitsutaka Kadota, Munazah Andarabi, Nozomu Takata, Mikiko Tokoro, Yukari Terashita, Shigenobu Yonemura, Charles A. Vacanti & Teruhiko Wakayama
Protection against misconduct

Japanese Researcher, an Author of a Discredited Stem Cell Study, Is Dead

By HIROKO TABUCHI  AUG. 5, 2014

THE RISE AND FALL OF STAP

Two papers published in Nature in January 2014 promised to revolutionize the way stem cells are...
Less time wasted

LKB1 and AMPK maintain epithelial cell polarity under energetic stress.
Mirouse V¹, Swick LL, Kazgan N, St Johnston D, Brenman JE.

Retraction in
LKB1 and AMPK maintain epithelial cell polarity under energetic stress. [J Cell Biol. 2013]

Abstract
LKB1 is mutated in both familial and spontaneous tumors, and acts as a master kinase that activates the PAR-1 polarity kinase and the adenosine 5-monophosphate-activated kinase (AMPK). This has led to the hypothesis that LKB1 acts as a tumor suppressor because it is required to maintain cell polarity and growth control through PAR-1 and AMPK, respectively. However, the genetic analysis of LKB1-AMPK signaling in vertebrates has been complicated by the existence of multiple redundant AMPK subunits. We describe the identification of mutations in the single Drosophila melanogaster AMPK catalytic subunit AMPKalpapha. Surprisingly, ampkalpapha mutant epithelial cells lose their polarity and overproliferate under energetic stress. LKB1 is required in vivo for AMPK activation, and lkb1 mutations cause similar energetic stress-dependent phenotypes to ampkalpapha mutations. Furthermore, lkb1 phenotypes are rescued by a phosphomimetic version of AMPKalpapha. Thus, LKB1 signals through AMPK to coordinate epithelial polarity and proliferation with cellular energy status, and this might underlie the tumor suppressor function of LKB1.

PMID: 17470638 [PubMed - indexed for MEDLINE]  PMCID: PMC2064817  Free PMC Article

Dystroglycan and perlecan provide a basal cue required for epithelial polarity during energetic stress.
Mirouse V¹, Christoforou CP, Fritsch C, St Johnston D, Ray RP.

Abstract
Dystroglycan localizes to the basal domain of epithelial cells and has been reported to play a role in apical-basal polarity. Here, we show that Dystroglycan null mutant follicle cells have normal apical-basal polarity, but lose the planar polarity of their basal actin stress fibers, a phenotype it shares with Dystrophin mutants. However, unlike Dystrophin mutants, mutants in Dystroglycan or in its extracellular matrix ligand Perlecan lose polarity under energetic stress. The maintenance of epithelial polarity under energetic stress requires the activation of Myosin II by the cellular energy sensor AMPK. Starved Dystroglycan or Perlecan null cells activate AMPK normally, but do not activate Myosin II. Thus, Perlecan signaling through Dystroglycan may determine where Myosin II can be activated by AMPK, thereby providing the basal polarity cue for the low-energy epithelial polarity pathway. Since Dystroglycan is often downregulated in tumors, loss of this pathway may play a role in cancer progression.
Less time wasted

2010
Start of the PhD

2011
1 year of PhD
gone
results not reproduced

2012
2 years of PhD gone
results not reproduced
Looking for the original data…
Less time wasted
Less time wasted

Retraction

LKB1 and AMPK maintain epithelial cell polarity under energetic stress

Vincent Mirouse, Lance L. Swick, Nevzat Kazgan, Daniel St Johnston, and Jay E. Brenman


The editors of The Journal of Cell Biology have been notified by Dr. Daniel St Johnston and Dr. Jay E. Brenman that they and the other authors of the paper referenced above retract the paper. As a result of this retraction, no data in this paper should be cited in the scientific literature.
Less time wasted

RETRACTED: Dystroglycan and Perlecan Provide a Basal Cue Required for Epithelial Polarity during Energetic Stress

Vincent Mirouse, Christina P. Christoforou, Cornelia Fritsch, Daniel St. Johnston, Robert P. Ray

3 Present address: GReD, CNRS UMR 6247, Clermont Université, Faculté de Médecine, Clermont-Ferrand 63001, France

Open Access

DOI: http://dx.doi.org/10.1016/j.devcel.2008.11.006

Open access funded by Wellcome Trust
Less time wasted

- It took 6 years from the time of the original publication (2007) to the final retraction (2013)

- Time & resources wasted because data was not available (not to mention people’s careers!)
Get access to shared data

Research Data Discovery Service (Alpha)

Search datasets...

Order by Relevance

Advanced Search

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

http://www.rcuk.ac.uk/research/datapolicy/
How to share data?

- 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


  - For sensitive data:

    - UK Data Service: [reshare.ukdataservice.ac.uk/](http://reshare.ukdataservice.ac.uk/) or EGA: [www.ebi.ac.uk/ega/home](http://www.ebi.ac.uk/ega/home)

    - Or other repositories (including Cambridge repository): [www.data.cam.ac.uk/repository](http://www.data.cam.ac.uk/repository)
Exemptions

- Personal/sensitive data
- IP protection/commercial data

Appropriate statement in the publication needs to explain the reasons for restrictions
Resources for working with personal/sensitive data

- **University Ethics website:**
  - [www.research-integrity.admin.cam.ac.uk/research-ethics/](http://www.research-integrity.admin.cam.ac.uk/research-ethics/)
  - Dr Rhys Morgan, Research Governance and Integrity Officer:
    - rhys.morgan@admin.cam.ac.uk

- **MRC guidelines:**

- **ESRC consent form, anonymisation guide, and access control:**
  - [http://ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation](http://ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation)
  - [http://ukdataservice.ac.uk/manage-data/legal-ethical/access-control](http://ukdataservice.ac.uk/manage-data/legal-ethical/access-control)

- **Our website (University resources):**
  - [http://www.data.cam.ac.uk/sensitive-data](http://www.data.cam.ac.uk/sensitive-data)
Some funders actually check it...

Random checks on all publications from 1 May 2015 that acknowledge EPSRC ± sanctions for not sharing
What do I need to do?

• For every new publication – share what is shareable & add a link to your data

• Be aware of help available to you at the University of Cambridge
Cambridge support for data management and sharing
Research Data Management

Upload your data

Have a question? E-mail info@data.cam.ac.uk

www.data.cam.ac.uk
<table>
<thead>
<tr>
<th>Funder</th>
<th>Key policy highlights</th>
<th>Date of the last update/policy check</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>
<td>April 2016</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” Data needs to be deposited into a NERC data centre within 2 years of collection</td>
<td>April 2016</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>
<td>April 2016</td>
</tr>
</tbody>
</table>
| NIHR   | "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 the report.  
- Can be obtained from the corresponding author via the (name of) repository.  
If you have deposited (or intend to deposit) data from your study into a data sharing repository, please supply the URL to the data archive so that when your report | April 2016 |
How to share research data?

What shall I do to make my data available?

To make your data available first make sure that it is properly organised and labelled, and then simply upload your data to a data repository. You can:

- upload your data to the University repository
- choose an alternative data repository
- sometimes your funder (e.g. ESRC or NERC) will require that you deposit your data into a specified data repository

www.data.cam.ac.uk/repository
Discipline-specific repositories preferred

www.re3data.org
Cambridge data repository

www.data.cam.ac.uk/upload
We will check your data, upload it into the repository and send you a link for your paper
Non-positive results can be shared as well

p-value 0.05: who is going to publish their results?
Each submission gets a separate record
Repository is well ‘googleable’
Part 4: How to avoid problems with data management
Data management plan:

...roadmap to help you not to get lost with your data
Data management plan:

- You now have 3 mins to write your own data plan

- Fill in just the top section and leave the ‘Comments’ section blank
Data management plan:

- Now work in pairs and exchange your plans
- You have 3 mins to write down comments on each other’s plans
Data management plan:

- Now you have 2 minutes to exchange feedback
Data management plan:

- You have created your first data management plan with comments from peer-review

- (This will be extremely useful when applying for grants)
Today’s summary:

We have covered the following:

1. Backup and exchange strategies
2. How to organise your data well?
3. Data sharing
4. How to avoid problems: data management plans
Take-home message:
Final conclusions:

- Data management plan can save you from a lot of trouble
- www.data.cam.ac.uk
- info@data.cam.ac.uk
THANK YOU

Feedback forms + certificates

Questions: info@data.cam.ac.uk

Follow us on Twitter: @CamOpenData