

Supporting Research Data Management: Providing practical advice to researchers

Anna Collins

DSpace@Cambridge Research Data and Digital Curation Officer
Cambridge University Library



"People will work at home, so things will be stored locally and shuffled between machines on memory sticks."

Why are we talking about research data management?

"How do you decide what can be deleted as it's all a record – I'm not confident to make that decision."

"PhD students lose material all the time...and they are exactly the people who want to be backing up. These are people who are creating data which are life and death important to them"



Why are we talking about research data management?

“When you take a book out of the library and there are pages missing, you bring it back to the library and expect them to fix it”.



Incremental

Aim:

- To **improve research data management** within the institutions
 - Focusing on better advice, training and support
 - Ensuring resources created can be re-used by others

Collaboration with:



University
of Glasgow

Humanities
Advanced Technology
& Information Institute

Incremental: Key findings



- Main concerns are not institution- or discipline- specific
 - Recognise need for disciplinary examples
- Simple issues often the most irksome
- Training and guidance resources must be simple, engaging and easy to access
- Points of intervention
- Language matters



<http://www.lib.cam.ac.uk/preservation/incremental/>

Incremental: Outputs



- Provision for people who cannot access face-to-face training
 - Flexibility
- Links to alternative resources and more in-depth information
 - Academic and non-academic sources (eg blogs, government websites, videos of seminars)
- Downloadable information

A screenshot of the DSpace@Cambridge website. The header includes the University of Cambridge logo, the DSpace@Cambridge logo, a search bar with 'Google Custom Search', and a 'Contact' link. Below the header is a breadcrumb trail: 'University of Cambridge > Cambridge University Library > Data Management Home'. A left-hand navigation menu lists: Home, Create, Organise, Access, Look After, About, Browse, Glossary, Training, Advice & Support. The main content area is titled 'Support for Managing Research Data' and contains an introductory paragraph: 'Research data takes many forms, ranging from measurements, numbers and images to documents and publications. These web pages will help you plan, create, organise, share, and look after your electronic research materials, in whatever form they take.' Below this are four colored boxes: 'Creating Your Data' (orange), 'Organising Your Data' (green), 'Accessing Your Data' (yellow), and 'Looking After Your Data' (purple). Each box has a corresponding paragraph of text. At the bottom, there is a footer with the text: 'This site was created by the Incremental project and is maintained by DSpace@Cambridge. Materials created by Incremental are re-useable via a CC license: BY-NC-SA - 2.0, U.K.' and links for 'Accessibility' and 'Privacy'.



www.lib.cam.ac.uk/dataman/



DataTrain

- Produced **discipline-specific resources** for face-to-face training
 - Focused on Archaeology and Social Anthropology
- Resources created by recent PhD students in Cambridge
 - Examples from the disciplines show relevance
- Resources released under CC licence 2.0 BY-NC-SA: By Attribution, Non-Commercial, Share-Alike

<http://archaeologydataservice.ac.uk/learning/DataTrain>



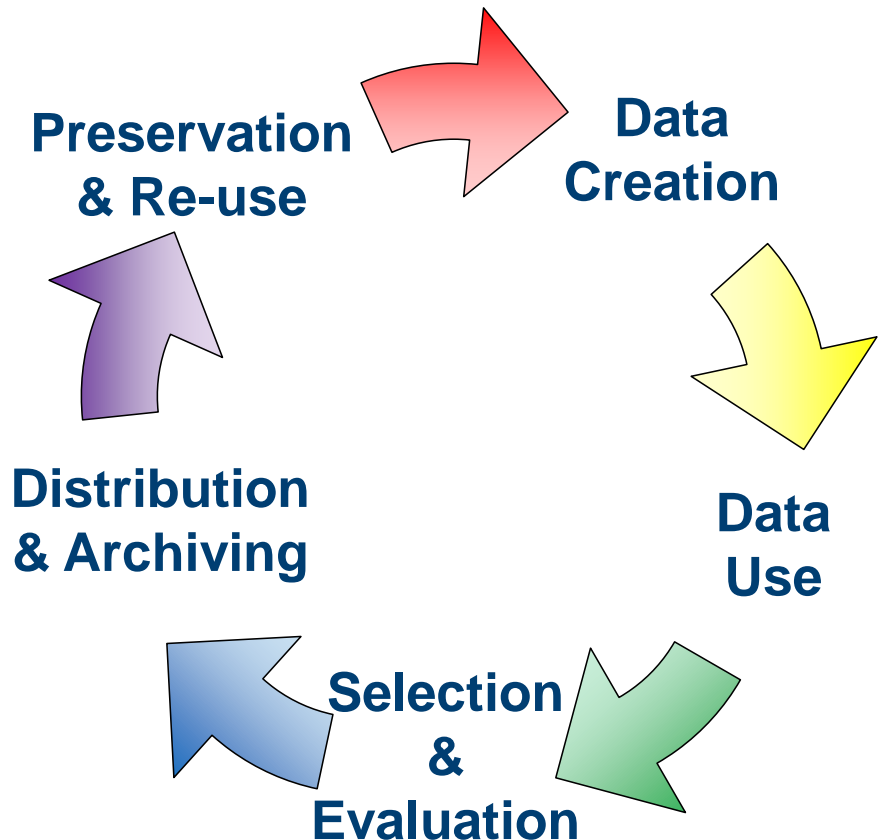
<http://www.lib.cam.ac.uk/dataman/datatrain/socanthintro.html>

Principle topics



Think in the context of their own research about:

- File structure and file naming schemes
- Hardware and software solutions
- Version control
- Strategies for backing-up
- Making decisions about what to keep and what to delete
- E-theses
- Intellectual Property Rights (IPR)
- Open access
- Freedom of Information (Fol)





Archaeology

- PowerPoint presentations provide key information
- Examples from recent research, showing good practice (and not so good practice!)
- Lots of information on file types, particularly graphics

Social Anthropology

- Discursive style highlighting points for discussion
- Examples came from old styles of managing data – (paper archives) establishing parallels
- Lots of information on tools, software, hardware

Features of the modules



- Use discipline-specific illustrations and examples where possible
- Draw comparisons with physical data/paper archives, non-work digital data
- Include exercises that relate to their own research that they can go away and use
 - File naming schemes, draft data management plans
- How much do we need to adapt for the discipline?
 - Differences between research groups, even within divisions

How does that help with Science, Engineering & Technology?



- Focus on common features of the two courses
- Transferrable skills
 - What fits with the RDF?
- What did I wish I knew before I started my doctorate (chemical crystallography)?
- What tips have I passed on to friends?
- Talk to people!



<http://www.canterbury.ac.uk/support/employability-and-careers-services/students/gain-skills/index.asp>



Creating and Managing Digital Research Data in Archaeology: An overview

Looking After Your Digital Research Data: Now, later, and long-term



Temple in Ubud's Monkey Forest
(Photographs: Kayan3 on Flickr)

Balinese Temples

are conceived of as never finished and on-going ritual and architectural projects.

However beautiful they are,
they are not a good model for
academic research data management!



And doing nothing about digital data is not an option either...



UNIVERSITY OF
CAMBRIDGE



Cambridge University Library

Be the boss of your hard drive: Managing your digital research data

Anna Collins
DSpace@Cambridge Research Data and Digital Curation Officer
Cambridge University Library

Why you need a data management plan



Electronics and Laser Research Building, University of Southampton, 2005

**What would happen to your data
if there was a fire in your
department?**



UNIVERSITY OF
CAMBRIDGE



How to get people interested...

“The whole thing is
incredibly dull.”

... and deal with conflicting aims

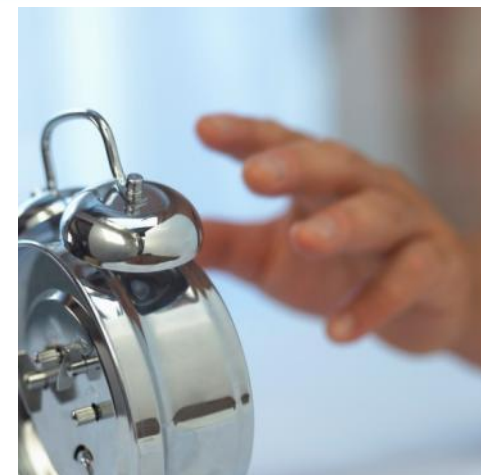
“[A] PhD student’s goal is to
get the papers written, the
thesis written in order to get
the PhD. Their goal is not
long-term storage of the
information.”

Timing is everything



That tricky balance between being too early...

“There’s no point being told all this stuff when you’re not using it because – I only learn how to do things when I need to know.”



... and being too late



“People bring in sticks with 4GB of data that simply no longer work – and nothing can be done to retrieve it.”

It's obvious when you know how



- Lots of really good tips seem obvious ...
- ...but are not necessarily intuitive
- Easy to assume a high level of Information Literacy among students and researchers
- Spectrum of knowledge
 - Covering the basics is important
 - Students find it hard to get started



The things you know you ought to do



- Things people know about but a reminder is usually a good idea
 - Eg Backing up!
- Knowing it needs to be done vs ***actually doing it***
 - Remembering, setting aside the time, appreciating its importance, etc
 - “I’m not as good at backing up as I ought to be...”
- Doing it in the best way
 - Patchy awareness of different types of storage and their limitations

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
- 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: 07700 123 456
(p.murrayrust@cam.ac.uk)

Thank you!!

Thanks to Peter Murray-Rust, who took the original picture

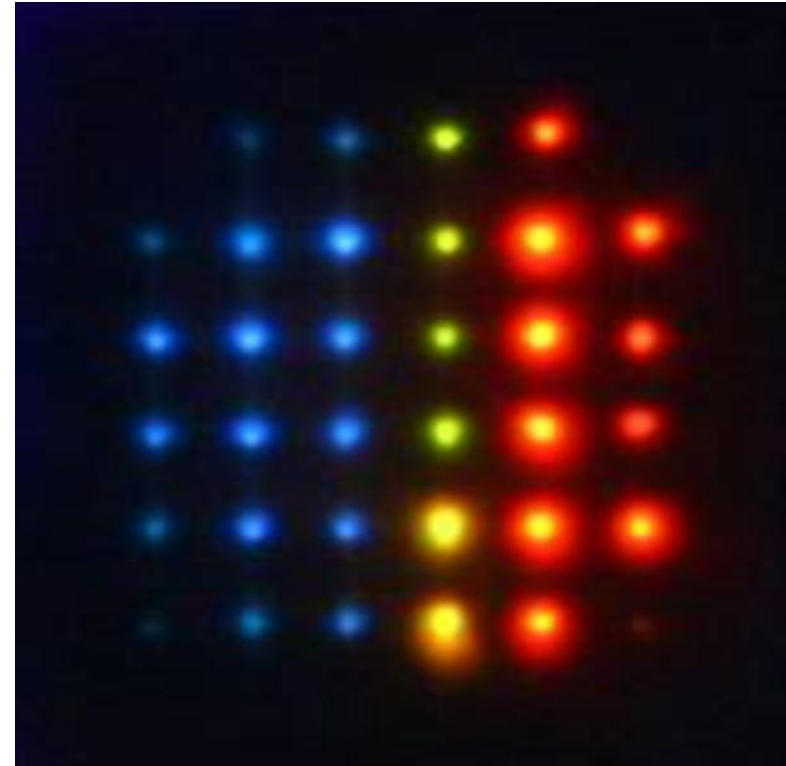


- Translate data management from specialist to non specialist
 - ‘Looking after your data’ rather than ‘Preserving your data’
 - ‘Be the boss of your hard drive’ rather than ‘Data Management Planning’
 - Suspicious of ‘policies’ which imply a mandate. More receptive to ‘guidance’ or advice’
- How much new jargon should a researcher have to learn?
 - Data management is part of being a researcher
BUT
 - We’re trying to promote a service, so need to use their language to make relevance clear

Department of Engineering



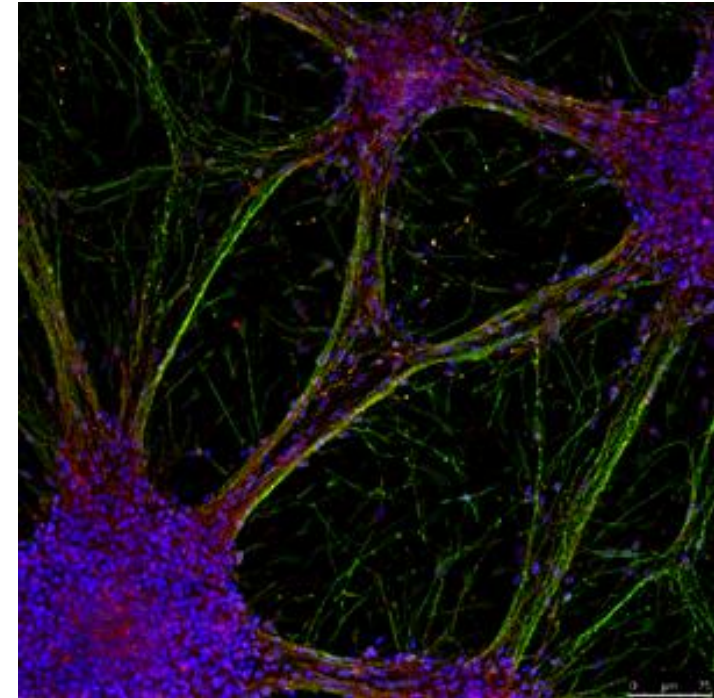
- Largest department in the University (about 10%)
- 6 academic divisions:
 - Turbomachinery, energy and fluid mechanics
 - Electrical engineering
 - Mechanics, materials and design
 - Civil, structural and environmental engineering
 - Manufacturing and management
 - Information engineering
- Transferable skills training compulsory for 1st year PhD students – try to include RDM



Stephen Morris - Liquid crystal laser array
A two-dimensional array of red-green-blue liquid crystal lasers constructed from a single liquid crystal device. This device emits all three colours simultaneously.



- Looks after educational and career needs of graduate students and early career researchers in Faculties of:
 - Biology
 - Clinical Medicine
 - Veterinary Medicine
- Wide range of training in transferable and study skills
 - Graduate Development Programme (Geraint Story)



Jignesh Tailor

Human neurons generated from neuro-epithelial stem cells.

Promotion



<http://www.lib.cam.ac.uk/datacenter/dspace/>

DSpace
@Cambridge

Prevent data

© University of Cambridge

DSpace
@Cambridge

Publish your files and other items

© Cambridge University Library

DSpace
@Cambridge

Prevent data

© Cambridge University Library

DSpace
@Cambridge

Need help with finding your files?
www.lib.cam.ac.uk/datacenter/support@repository.cam.ac.uk

Lessons learned



- Information needs to be seen to be useful
 - Need to make sure that the approach is consistent with what the researchers want and expect
 - Dialogue with researchers to check context/provide examples
 - Relevance of generic information needs to be clear
- Discussion helps to draw out the relevant points
 - Get people to think about their own research
 - Share ideas, experiences and best practice



“A good plan implemented today is better than a perfect plan implemented tomorrow.”

George Patton

Acknowledgements



DSpace@Cambridge:

Elin Stangeland, Barbara Bültmann, David Piper

Incremental:

Catherine Ward, Lesley Freiman, Sarah Jones, Laura Molloy, Kellie Snow

DataTrain:

Lindsay Lloyd-Smith, Irene Peano, Cameron Petrie, Sîan Lazar, ADS