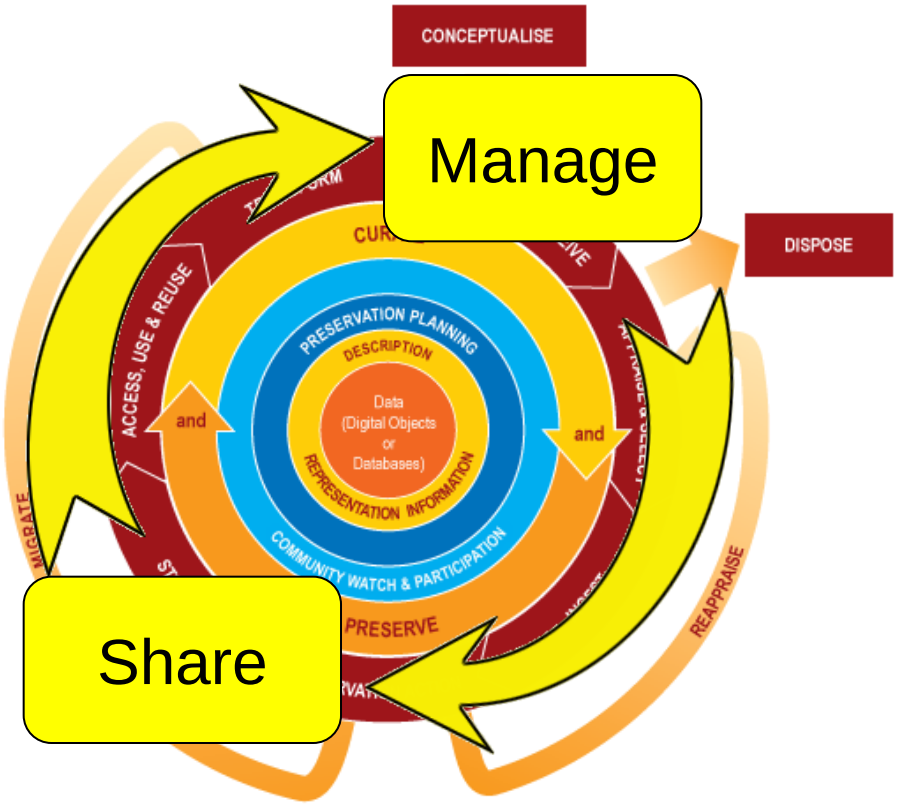
What is Research Data Management and why does it matter?

Marieke Guy
DCC, University of Bath
m.guy@ukoln.ac.uk

Making the most of your data, Northampton, 15th June 2012

Funded by: JISC

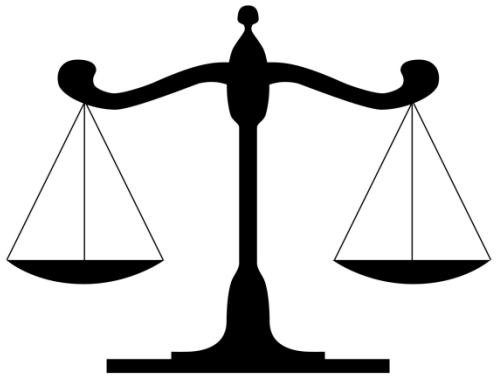
What is research data management?

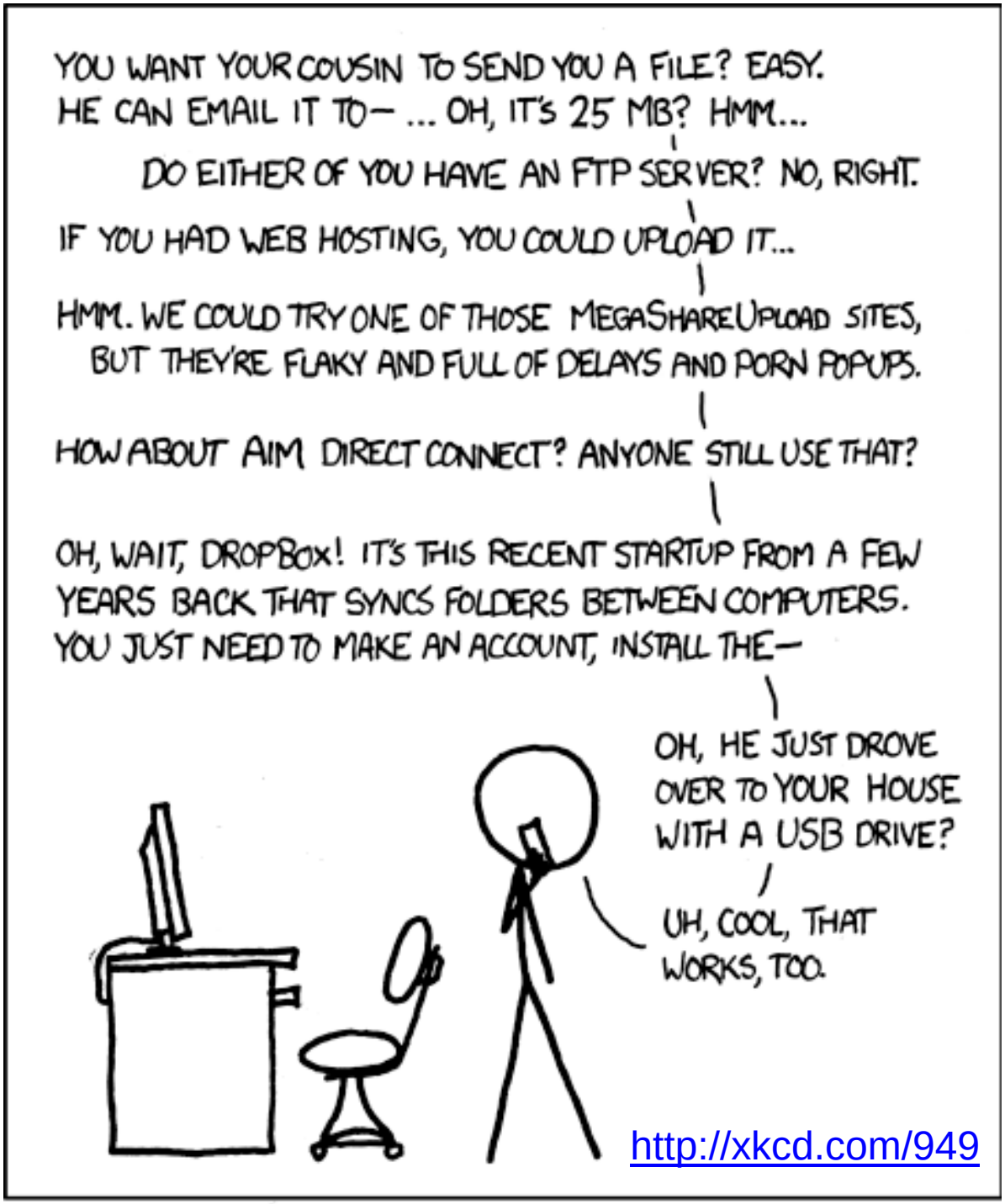


"the active management and appraisal of data over the lifecycle of scholarly and scientific interest"

Data management is part of good research practice

Good data management is about making informed decisions





http://xkcd.com/949

Why manage your data well?

- so you can find and understand it when needed
- to avoid unnecessary duplication
- so you can validate results if required
- so your research is more visible and has greater impact
- to get credit when other researchers cite your data

What is involved in RDM?

- Data management planning
- Creating data
- Documenting data
- Storing data
- Sharing data
- Preserving data

Data management planning

What do you (and others) want to do with the data?
→ make decisions that allow for this

Remember:

Data management is about making informed decisions

Talk to colleagues and support staff to see which option works best

Data Management and Sharing Plans

Funders typically want a short statement covering:

- What data will be created? (format, types, volumes etc)
- What standards and methodologies will you use? (incl. metadata)
- How will you manage ethics and Intellectual Property?
- What are the plans for data sharing and access?
- What is the strategy for long-term preservation?

DMP tool: https://dmponline.dcc.ac.uk/

How to write a DMP:
www.dcc.ac.uk/resources/how-guides/develop-data-plan

Creating data: questions

What formats will you use?

- determined by the instruments / software you have to use
- common, widespread formats to enable reuse

How will you create your data?

- What methodologies and standards will you use?
- How will you address ethical concerns and protect participants?
- Will you control variations to provide quality assurance?

Creating data: advice

Different formats are good for different things

- open, lossless formats are preferable for preservation e.g. rtf, xml, tif, wav
- proprietary, compressed formats are often in widespread use e.g. doc, jpg, mp3

You might use one for analysis & convert for preservation

Excellent guidance on creating data & managing ethics in:

www.data-archive.ac.uk/media/2894/managingsharing.pdf

Documenting data: questions

What information do users need to understand the data?

- descriptions of all variables / fields and their values
- code labels, classification schema, abbreviations list
- information about the project and data creators
- tips on usage e.g. exceptions, quirks, questionable results

How will you capture this?

Are there standards you can use?

Documenting data: advice

Create metadata at the time – it's hard to do later

Develop processes so everyone does the same

Use standards for interoperability

Storing data: questions

What is available to you?

What facilities do you need?

- remote access
- file sharing with colleagues
- high-levels of security

How will the data be backed up?

Storing data: advice

Speak to your local IT Team for advice

Remember that all storage is fallible – need to back-up

- keep 2+ copies on different types of media in different locations
- manage back-ups (migrate media, test integrity)

Choose appropriate methods to transfer / share data

- email, dropbox, ftp, encrypted media, filestore, VREs...

Sharing data: questions

Does your funder expect you to share data?

Which data can be shared?

How will you share your data?

Sharing data: advice

Know what you're expected to share (or not!)

Check if there is a repository, data centres or community initiative for sharing your data

Share you data where possible

- there are benefits!

More citations: 69% ↑

(Piwowar, 2007 in PLoS)

Preserving data: questions

Are you required to preserve (or destroy) your data?

How will you select what to keep?

Is there somewhere you can archive your data?

How can you support the reuse of your data?

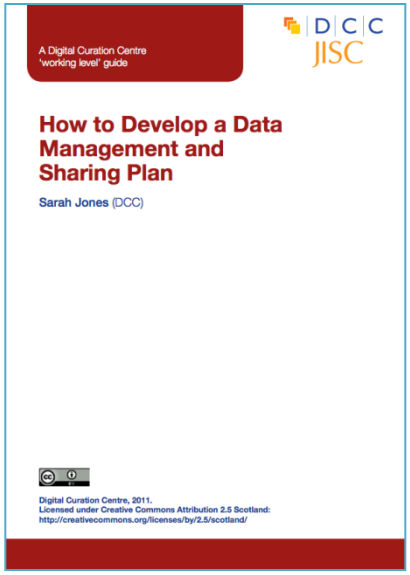
Preserving data: advice

Use available data centres - http://datacite.org/repolist

Check out the DCC's How to guides

- select and appraise research data
- licence research data
- cite datasets and link to publications

www.dcc.ac.uk/resources/how-guides



Thanks - any questions?

For DCC guidance, tools and case studies see:

www.dcc.ac.uk/resources

Follow us on twitter @digitalcuration and #ukdcc