Introduction to data management planning

Joy Davidson

Digital Curation Centre

Funded by: Jisc

Digital Curation Centre (DCC)

The DCC Mission

"Helping to build capacity, capability and skills in data management and curation across the UK's higher education research community"

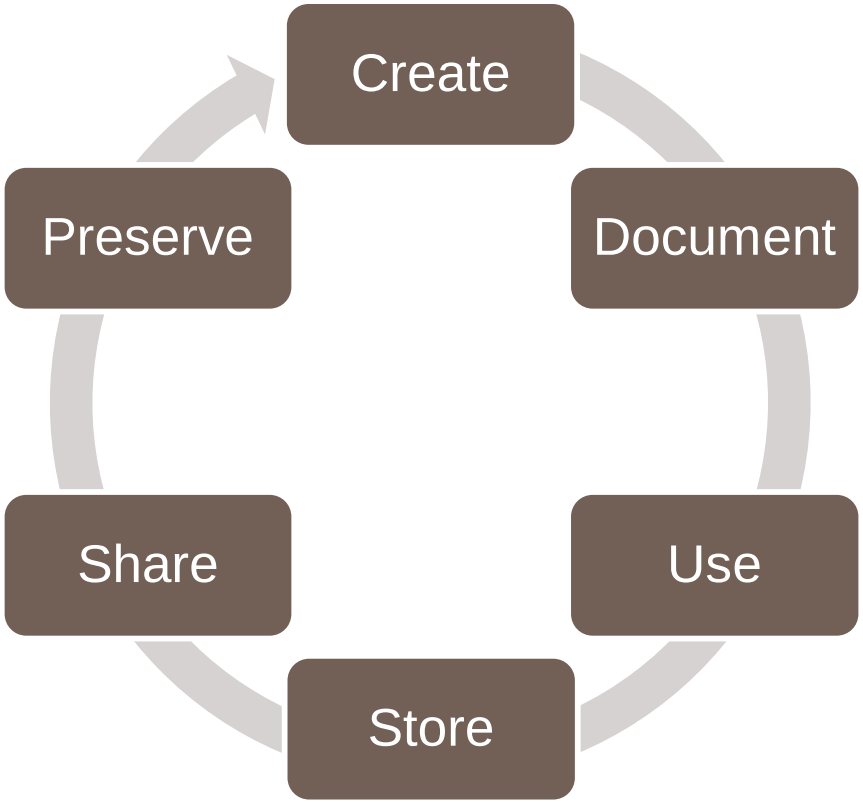
Phase 3 Business Plan

www.dcc.ac.uk

JISC

What is Research Data Management?

- Data Management Planning
- Creating data
- Documenting data
- Accessing / using data
- Storage and backup
- Selecting what to keep
- Sharing data
- Data licensing and citation
- Preserving data



Why manage research data?

Direct benefits for you

- To make your research easier!
- Stop yourself drowning in irrelevant stuff
- Have data organised so you know which versions are most up-to-date
- Make sure you can under stand and reuse your data again later

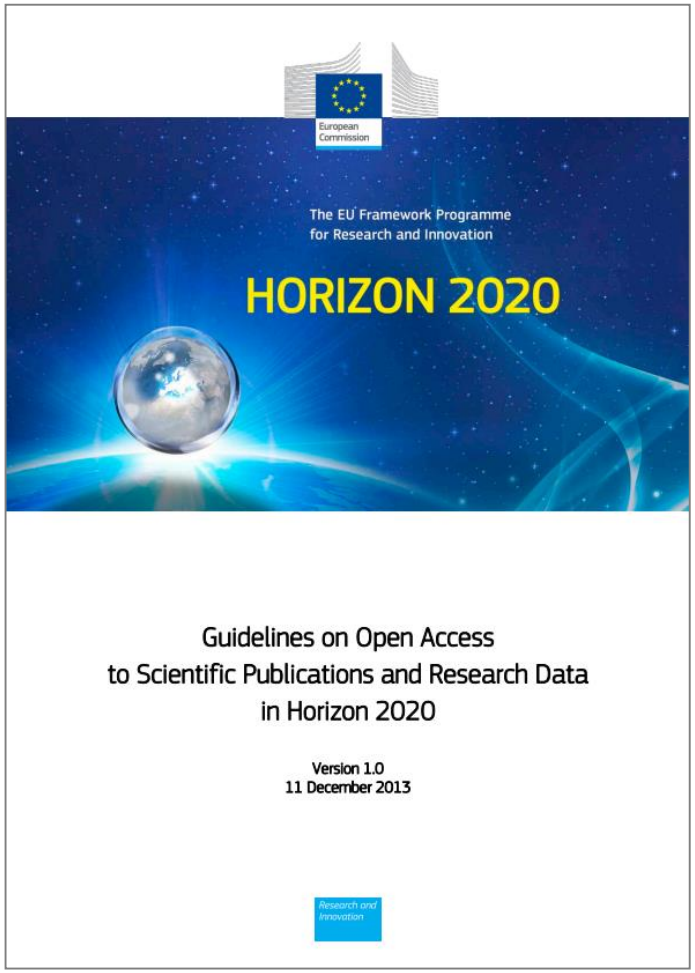
Research integrity

- To avoid accusations of fraud or bad science
- Evidence findings and enable validation
- Codes of practice on good research conduct
- Many research funders worldwide now require Data Management and Sharing Plans

Potential to share

- So others can reuse and build on your data
- To gain credit –higher citation rates when data are shared
- For greater impact and new collaborations
- Promote innovation and allow research in your field to advance faster

Funders have expectations about data sharing…

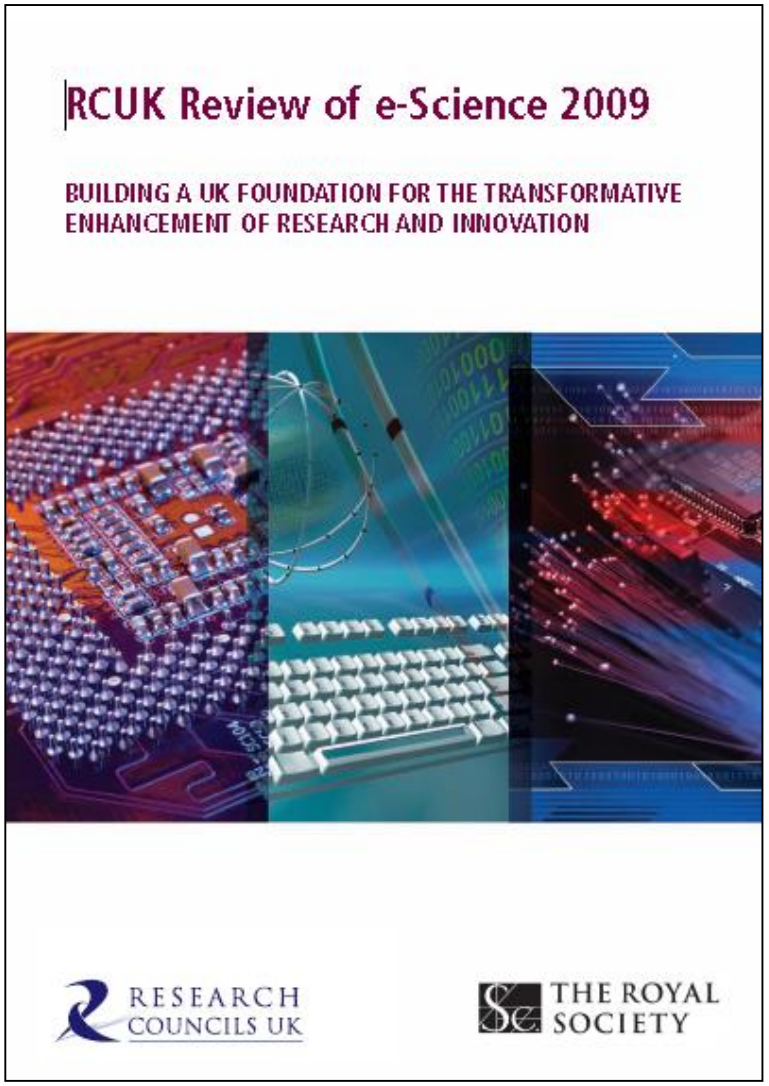


"The European Commission's vision is that information already paid for by the public purse should not be paid for again each time it is accessed or used, and that it should benefit European companies and citizens to the full."

http://ec.europa.eu/research/participants/data/ ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Data management plans requested for those participating in Open Data pilot.

…but RDM is part of good research practice!



"Data sets are becoming the new instruments of science"

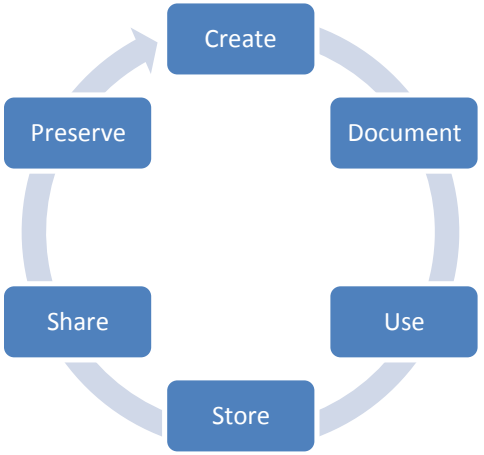
Dan Atkins, University of Michigan

Data management plans (DMPs)

What should be addressed in a DMP?

- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation

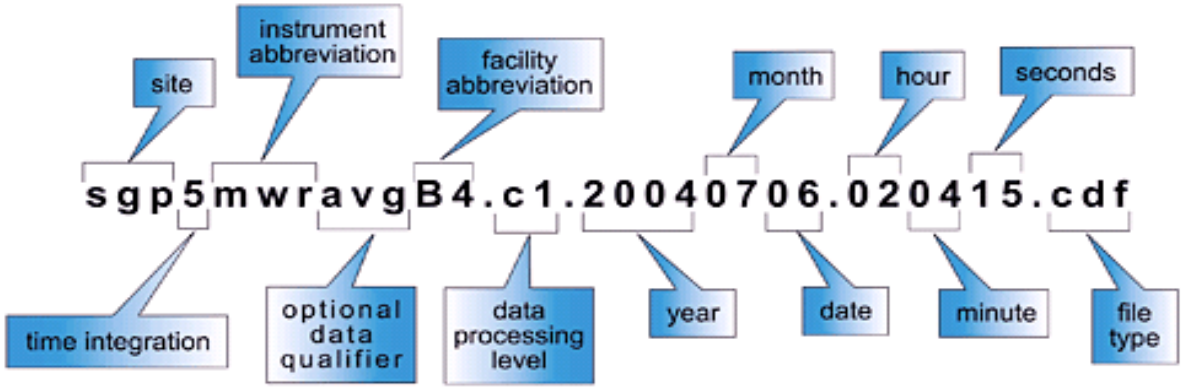
A DMP is a plan to share!



How will you name your files?

- Keep it simple!
- Agree methods with partners
- Include dates
- Avoid non-alphanumeric characters
- Use hyphens or underscores not spaces e.g. day-sheet, day_sheet
- Order the elements logically

An example netCDF data file name is depicted below:



Example from ARM Climate Research Facility www.arm.gov/data/docs/plan

www.jiscdigitalmedia.ac.uk/guide/choosing-a-file-name

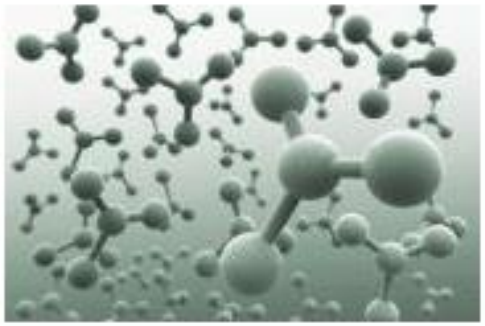
How should you describe your data?

Search by Discipline

Biology

Earth Science

General Research Data



Physical Science

Social Science & Humanities

RDA
RESEARCH DATA ALLIANCE

http://www.dcc.ac.uk/resources/metadata-standards

Where will you store the data during your research?

- Your own laptop
- University systems
- Cloud storage
- Combination

Your decision will be based on how sensitive your data are, how robust you need the storage to be, who needs access to the data, and when they need access to the data!

Which data need to be kept after the project ends?

Five steps to follow

① **Could** this data be re-used

② **Must** it be kept as evidence or for legal reasons

③ **Should** it be kept for its potential value

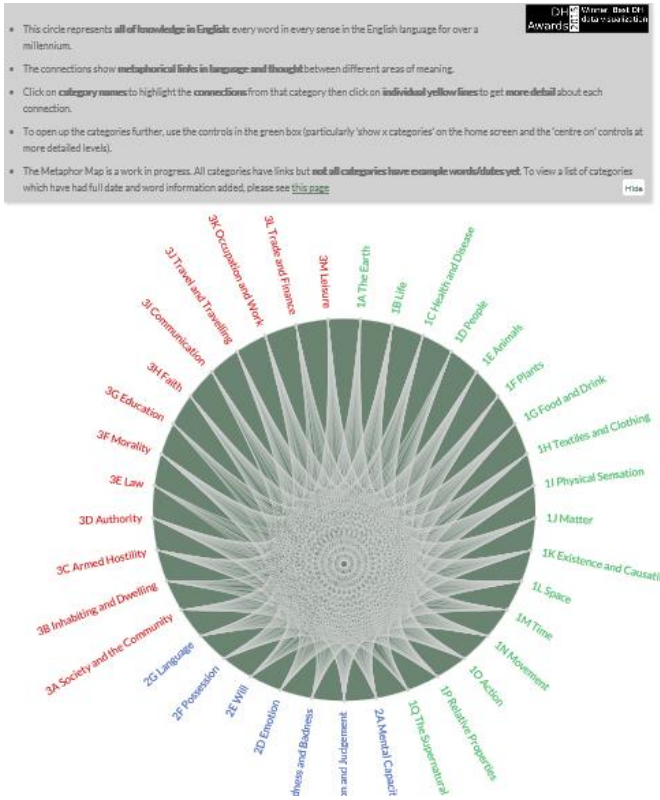
④ **Consider costs** – do benefits outweigh cost?

⑤ **Evaluate criteria** to decide what to keep

5 steps to decide what data to keep

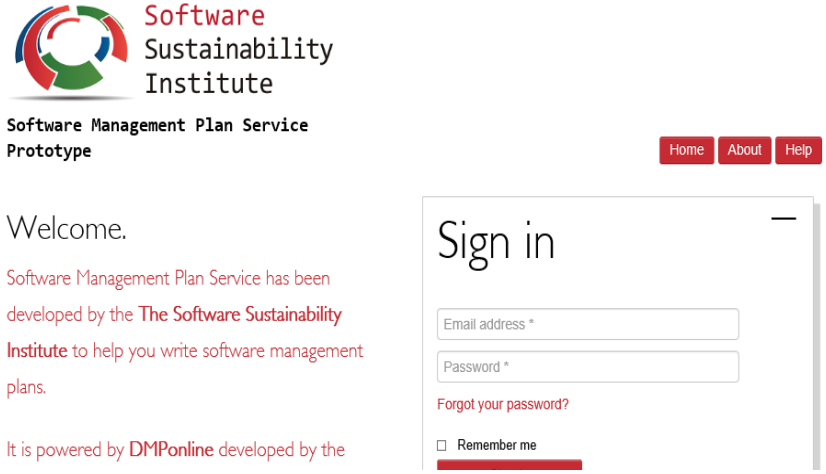
www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep

Remember to consider physical data, software and models



http://mappingmetaphor.arts.gla.ac.uk/

http://www.ukcrcexpmed.org.uk/Coventry_Warwick_CRF/PublishingImages/Tissue%20Bank%201.jpg



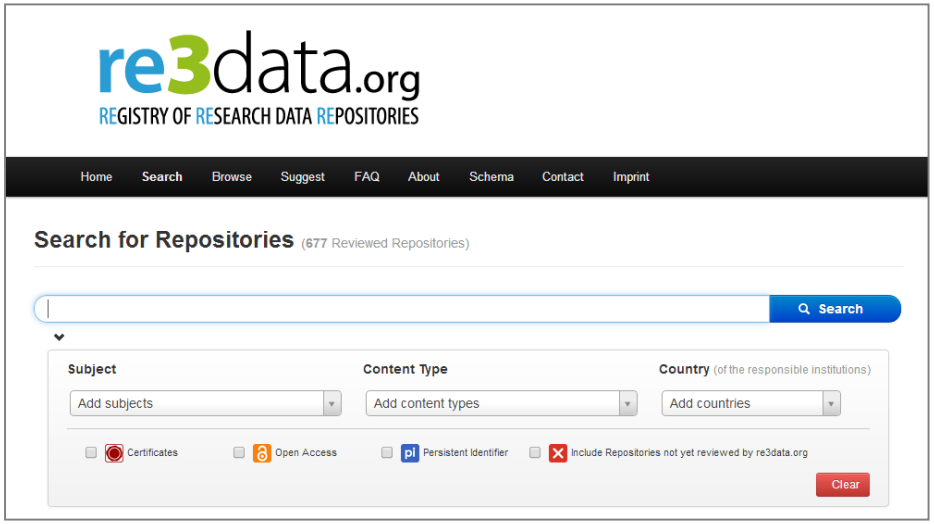
https://ssi-dev.epcc.ed.ac.uk/

Can your data be shared with others?

- PI/researcher
- Data repository and support staff
- Research participants
- Commercial partners
- Secondary data user

How will it be shared?

- Does your publisher or funder suggest a repository?
- Are there data centres or community databases for your discipline?
- Does your university offer support for long-term preservation?



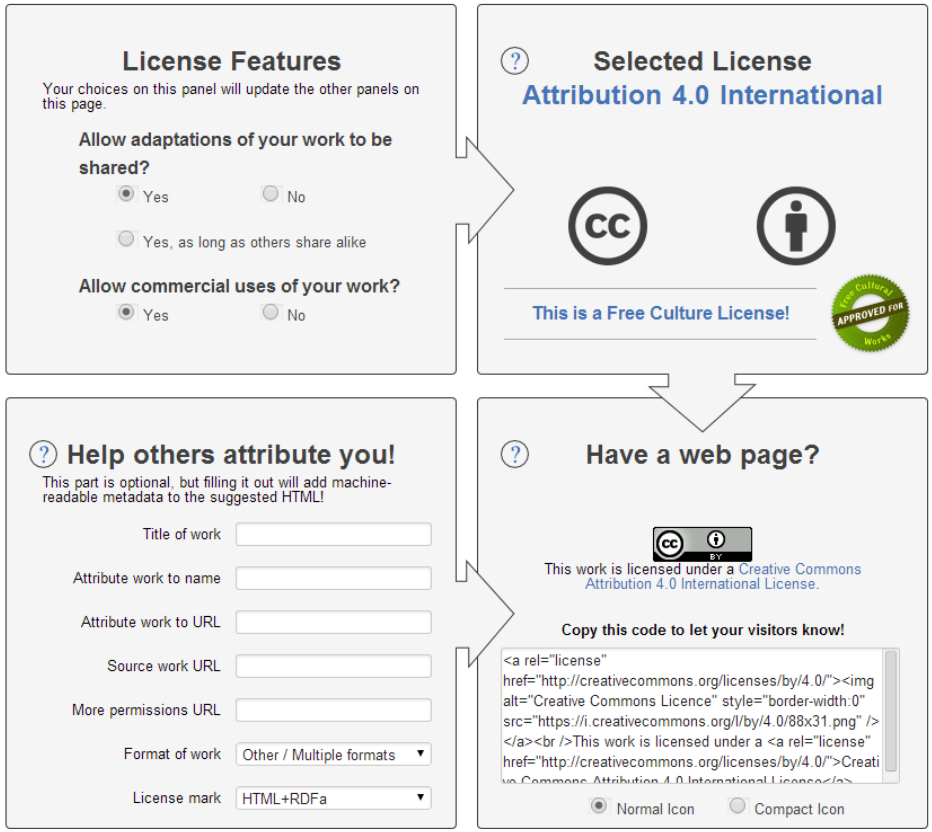
http://service.re3data.org/search

Zenodo

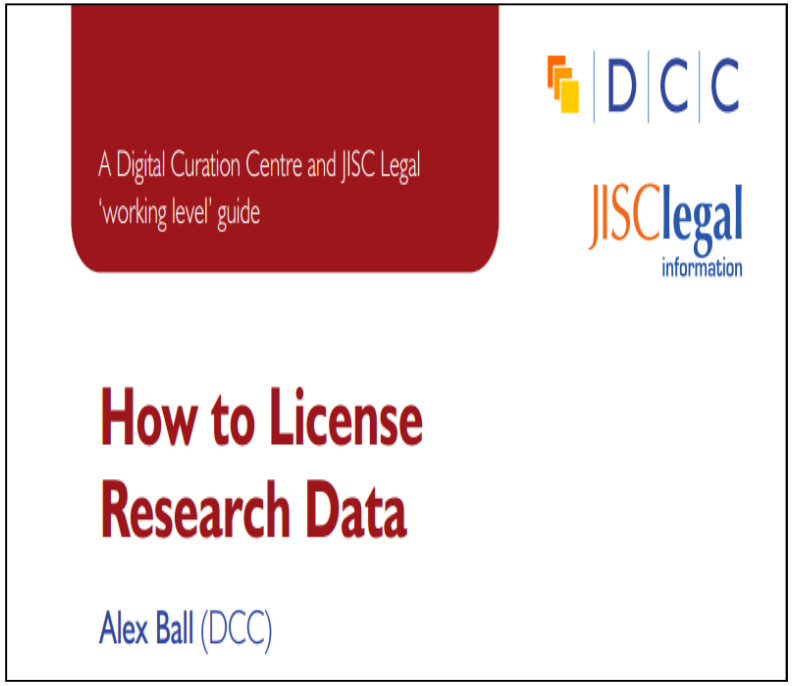
- Joint effort by OpenAIRE-CERN
- Multidisciplinary repository
- Multiple data types
- Citable data (DOI)
- Links funding, publications, data & software

www.zenodo.org

What do you want others to be able to do - or not do - with your data?

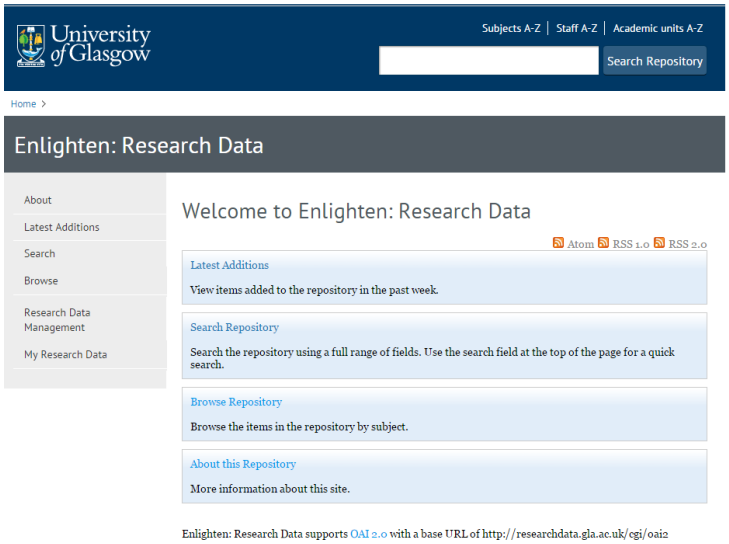


http://creativecommons.org/choose/

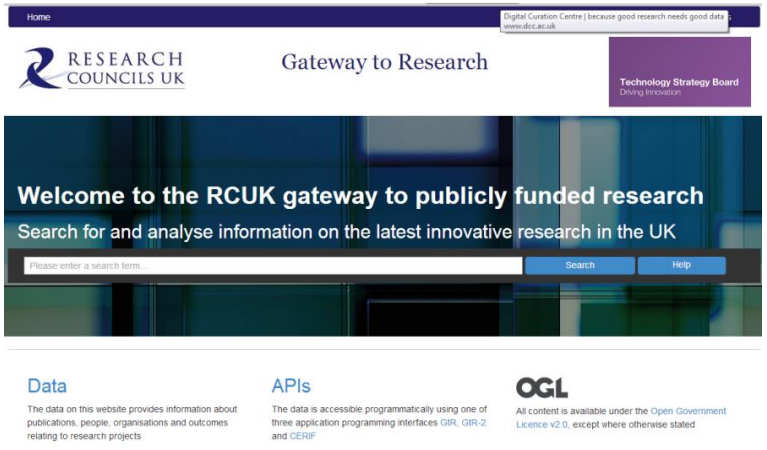


www.dcc.ac.uk/resources/how-guides/license-research-data

How will you make your data discoverable?

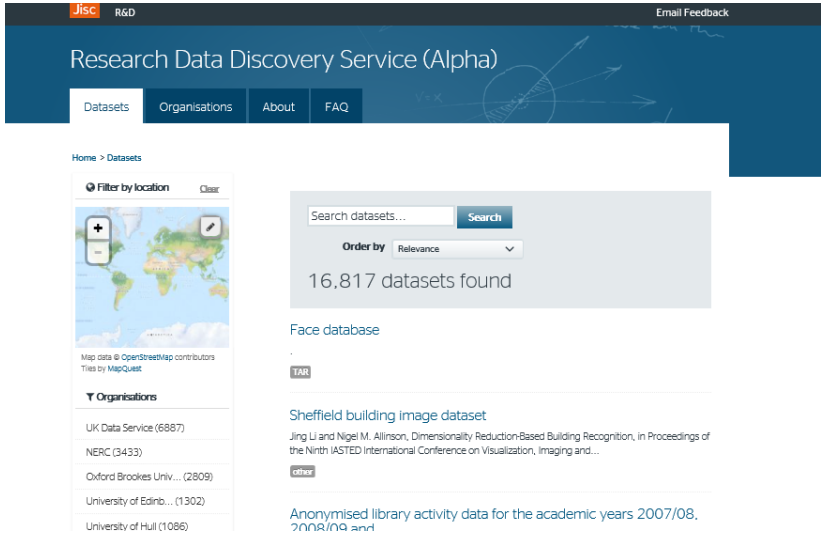


http://researchdata.gla.ac.uk/



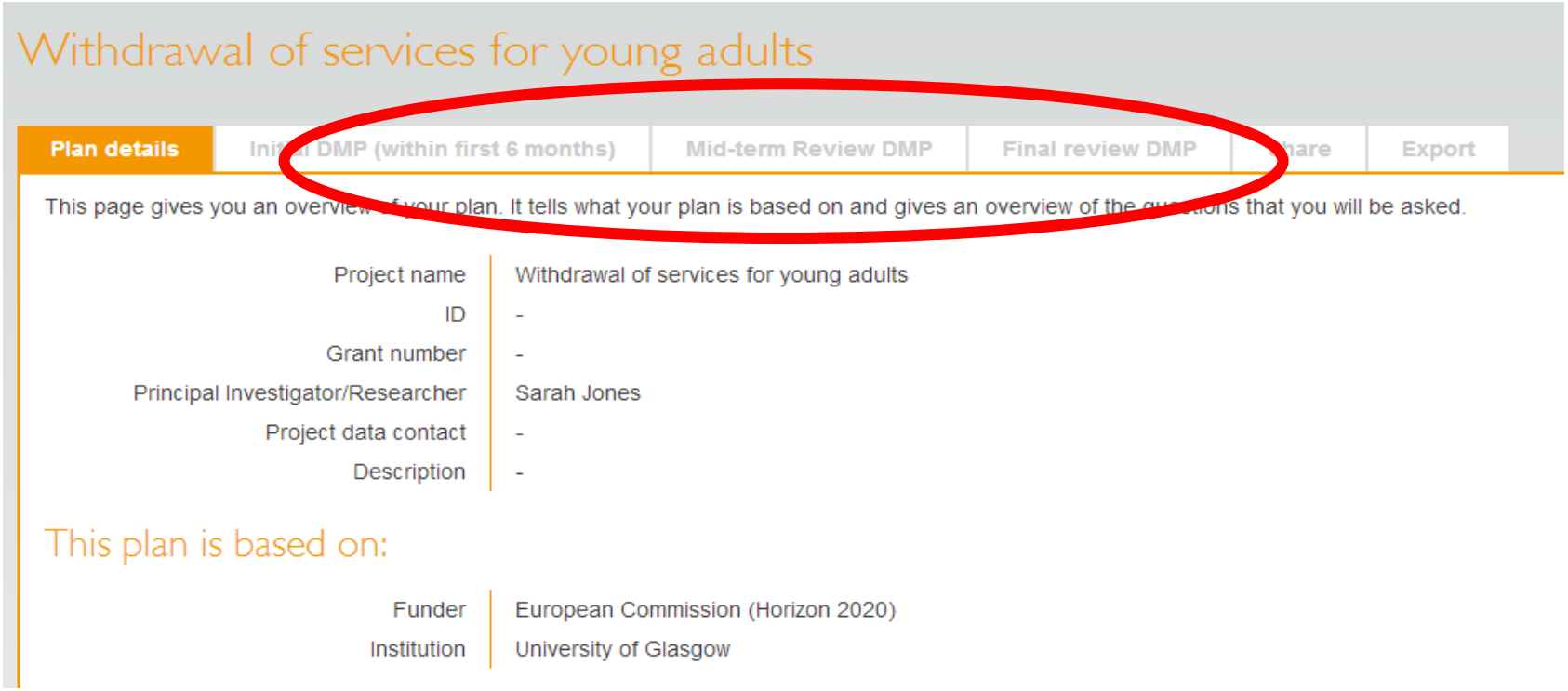
http://gtr.rcuk.ac.uk/

https://www.researchfish.com/



http://ckan.data.alpha.jisc.ac.uk/dataset

Update your plan!



… most do not update the DMP throughout the life of the project.

Good practice when creating DMPs

☑ **Start early**

☑ **Cost in sufficient effort to application**

☑ **Write the plan collaboratively**

☑ **Be realistic**

☑ **Update DMP**

Good RDM helps you comply with mandates but also leads to…

- More visible research outputs and increased impact - even for negative results
- Easier outputs reporting
- Better and more reproducible research!

Refer to free guides and briefing papers

www.dcc.ac.uk/resources/

Make use of free tools

https://dmponline.dcc.ac.uk

Thanks for listening!

joy.davidson@glasgow.ac.uk
www.dcc.ac.uk

Follow us on twitter:
@jd162a
@digitalcuration and #ukdcc