FOSTER

Facilitate Open Science Training for European Research

Focus on data management plans – policies, requirements, resources, examples

Martin Donnelly, Digital Curation Centre, University of Edinburgh

FOSTER

Facilitate Open Science Training for European Research

OVERVIEW

1. Data Management Plans and Planning
2. Policies overview (UK, USA, Australia)
3. DMP in Horizon 2020 (pilot overview and specifics)
4. Support and resources

FOSTER

1. Data management plans and planning

FOSTER

1a. DMP overview

- Research funders often ask for a data management plan (DMP) to be **submitted alongside grant applications** and/or developed over the course of the research project. (HEIs are increasingly asking their researchers to do this too…)
- A DMP is (usually) a short document detailing **specifics of the data** that will be created during a research project, together with information on **how it can be accessed** and utilised
- Data management planning is the process of **planning, describing and communicating the activities** carried out during the research lifecycle in order to…
  - Keep sensitive data safe
  - Maximise data's reuse potential
  - Support longer-term preservation
- Data management planning **underpins and pulls together** different strands of data management activities, often across multiple project partners

FOSTER

1b. Benefits of data management planning

- It is intuitive that planned activities stand a better chance of meeting their goals than unplanned ones. **The process of planning is also a process of communication**, increasingly important in interdisciplinary/multi-partner research. Collaboration will be more harmonious if project partners (in industry, other universities, other countries…) are on the same page
- In terms of data security, if there are **good reasons *not* to publish/share data**, in whole or in part, you will be on more solid ground if you flag these up early in the process
- DMP also provides an ideal **opportunity to engender good practice** with regard to (e.g.) file formats, metadata standards, storage and risk management practices, leading to greater longevity of data, and improved quality standards…

FOSTER

Limits of data management planning

What can a plan *not* do? It can't do the work for you.

The map is not the territory (Korzybski)

or

Chalk's no shears (Scottish saying)

It is important to remember that the human challenges in data management are often more difficult to meet than the technological ones.

Communication is vital.

FOSTER

1c. What does a data management plan look like?

A brief statement defining:

- how data will be captured/created
- how it will be documented
- who will be able to access it
- where it will be stored
- how it will be backed up, and
- whether (and how) it will be shared and preserved long-term
- etc

DMPs are often submitted as part of funding applications, but will be useful whenever researchers are creating (or reusing) data, especially where the research involves multiple partners, countries, etc…

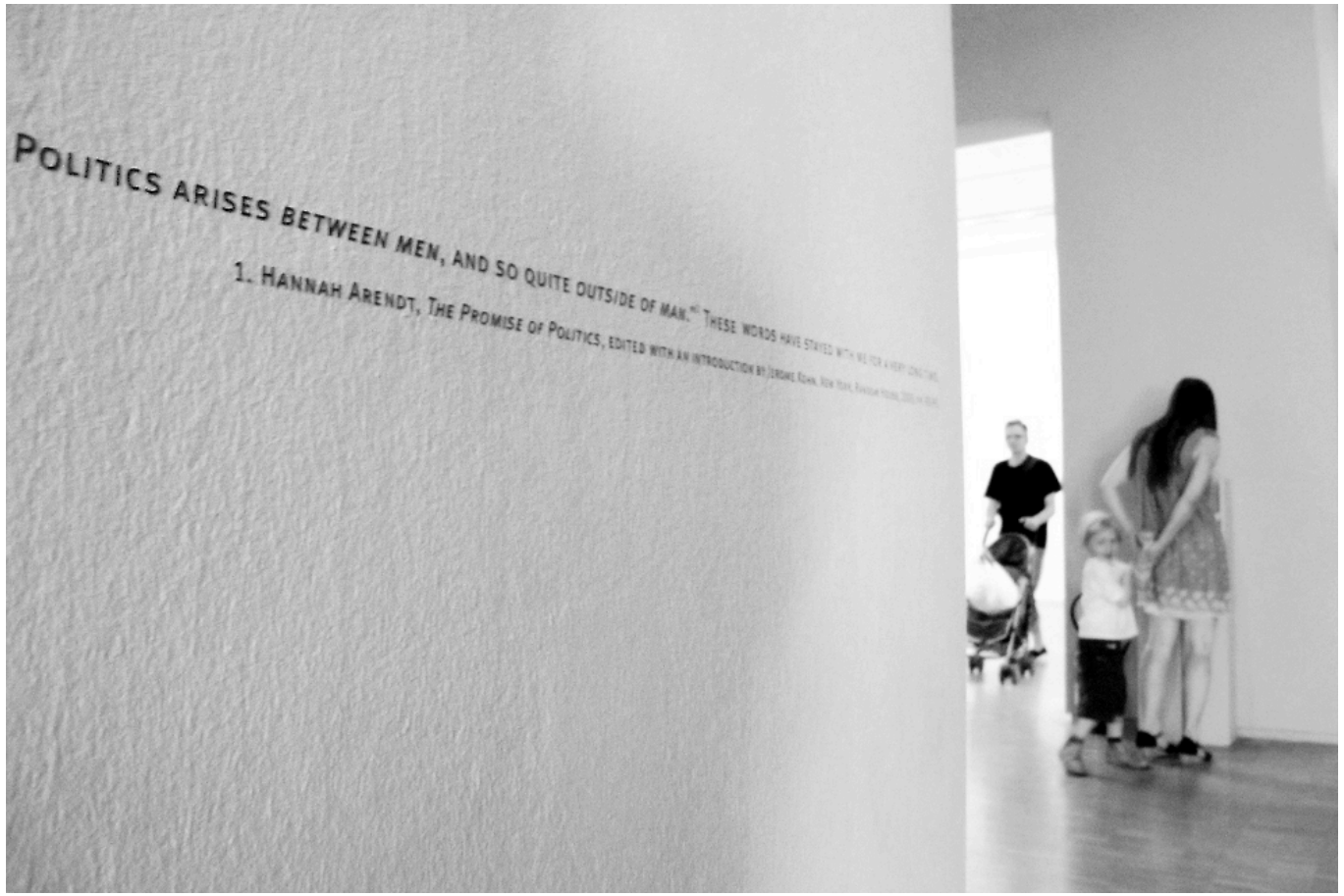
1d. Roles and responsibilities

Who's involved? It's worth bearing in mind that **RDM is a hybrid activity**, involving multiple stakeholder groups…

- The principal investigator (usually ultimately responsible for data)
- Research assistants (may be more involved in day-to-day data management)
- The institution's funding office (may have a compliance role)
- Library/IT/Legal (The library may issue PIDs, or liaise with an external service who do this, e.g. DataCite.)
- Partners based in other institutions
- Commercial partners
- etc

FOSTER

2. Policies



FOSTER

2a. Data-related policies (UK)

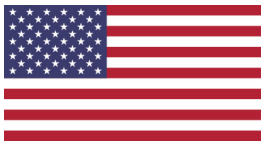
- Seven “Common Principles on Data Policy” – Data as a public good; Preservation; Discovery; Confidentiality; Right of first use; Recognition; Public funding for RDM
- Six of the seven RCUK funders require data management plans, or equivalent, at the application stage, as do Wellcome & CRUK
- The other council (EPSRC) requires nothing short of an institutional data infrastructure (by May 2015). They also expect that DMP will be a key component of this…

RESEARCH COUNCILS UK

EPSRC
Engineering and Physical Sciences Research Council

FOSTER

2b. Data-related policies (USA)



- The National Science Foundation (NSF) announced a DMP requirement in 2010, taking effect early in 2011
- White House Office of Science and Technology Policy requirement for DMPs announced March 2013 (programmes awarding >$100m annually)
- White House requirements include mechanisms covering compliance with plans and policies, and also cover costs of implementing plans

FOSTER

2c. Data-related policies (Australia)

The Australian Research Council (ARC) this year released new instructions for applications for Laureate Fellowships (http://www.arc.gov.au/ncgp/laureate/fl_instructions.htm) and Discovery Grants (http://www.arc.gov.au/ncgp/dp/dp_instructions.htm) Both include the following requirements when describing a proposal…

- COMMUNICATION OF RESULTS: Outline plans for communicating the research results to other researchers and the broader community, including scholarly and public communication and dissemination
- MANAGEMENT OF DATA: Outline plans for the management of data produced as a result of the proposed research, including but not limited to storage, access and re-use arrangements

FOSTER

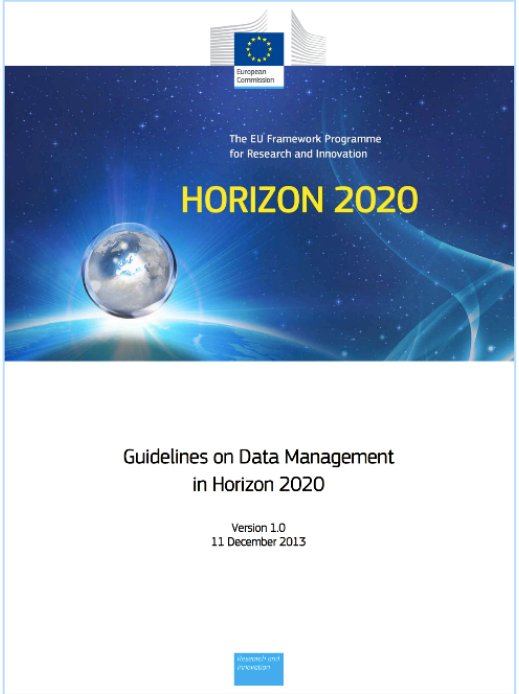
Australian Government
Australian Research Council

3. DMP in Europe

FOSTER

3a. DMP in Europe: overview

- Horizon 2020 includes a data management (planning) pilot
  - http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Proposals covered
  - “Innovation actions” and “Research and innovation actions”
- DMP contents
  - Data types; Standards used; Sharing/making available; Curation and preservation
- Multi phase approach
  - Initial DMP due within first 6 months
  - Mid-term DMP
  - Final review stage DMP
- There are opt-out conditions. A detailed description and scope of the Open Research Data Pilot requirements is provided on the Participants’ Portal



FOSTER

3b. The Horizon 2020 DMP pilot

v1: Within Six Months

For each data set specify the following:

- Data set reference and name
- Data set description
- Standards and metadata
- Data sharing
- Archiving and preservation (including storage and backup)

My plan (Horizon 2020 DMP)

Admin Details

Project Name: My plan (Horizon 2020 DMP) - DMP title

Principal Investigator / Researcher: Martin Donnelly

Funder: European Commission (Horizon 2020)

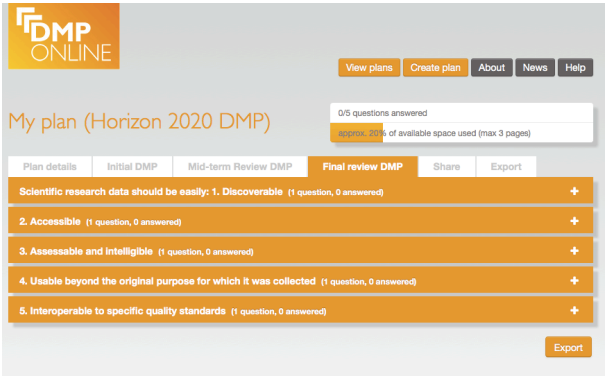
For each data set specify the following:

- **Data set reference and name**
- **Data set description**
- **Standards and metadata**
- **Data sharing**
- **Archiving and preservation (including storage and backup)**

.docx output from DMPonline

FOSTER

3b. The Horizon 2020 DMP pilot



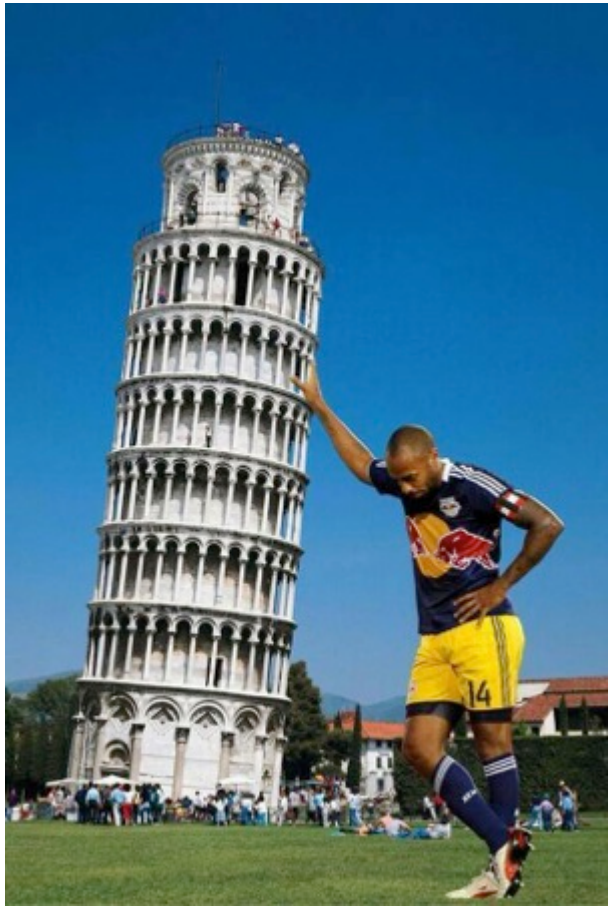
v2 and v3: Mid-Term and Final Reviews

Scientific research data should be easily:

1. Discoverable
   - Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?
2. Accessible
   - Are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses?
3. Assessable and intelligible
   - Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review?
4. Usable beyond the original purpose for which it was collected
   - Are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data?
5. Interoperable to specific quality standards
   - Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc?

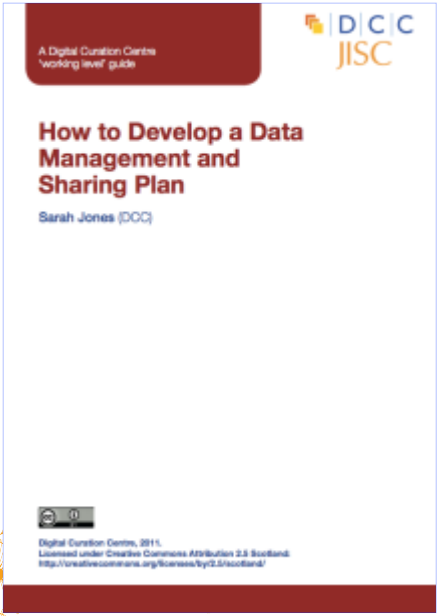
FOSTER

4. Support and resources



FOSTER

4a. DCC resources



- Book chapter
  - Donnelly, M. (2012) "Data Management Plans and Planning", in Pryor (ed.) *Managing Research Data*, London: Facet
- Guidance, e.g. "How-To Develop a Data Management and Sharing Plan"
- DCC Checklist for a Data Management Plan: http://www.dcc.ac.uk/resources/data-management-plans/checklist
- Links to all DCC DMP resources via http://www.dcc.ac.uk/resources/data-management-plans
- DMPonline: https://dmponline.dcc.ac.uk/

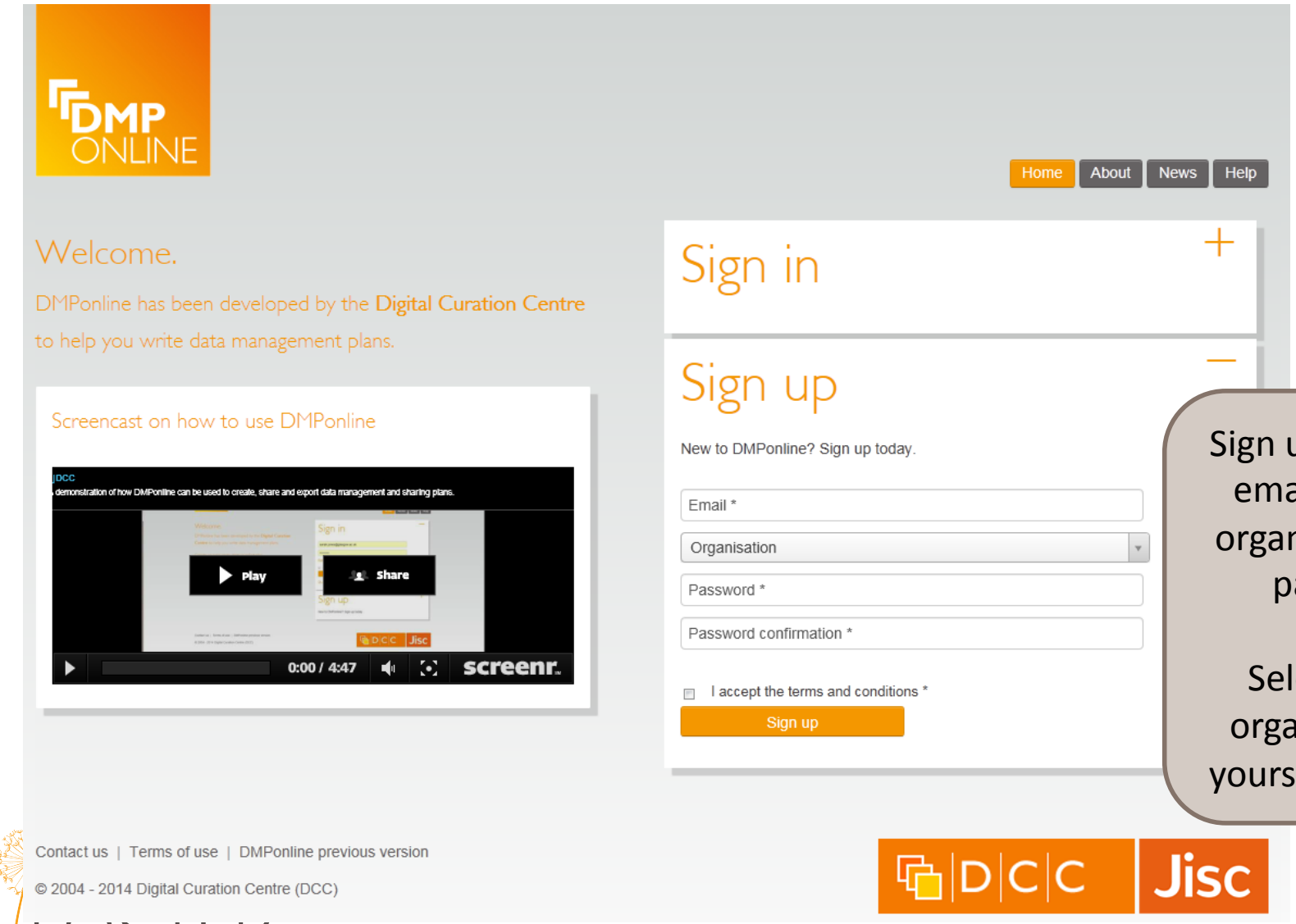
FOSTER

DMPonline: overview

- Helps researchers write DMPs
- Provides funder questions and guidance
  - Includes a template DMP for Horizon 2020
- Provides help from universities
- Examples and suggested answers
- Free to use
- Mature (v1 launched April 2010)
- Code is Open Source (on GitHub)

https://dmponline.dcc.ac.uk

Registration

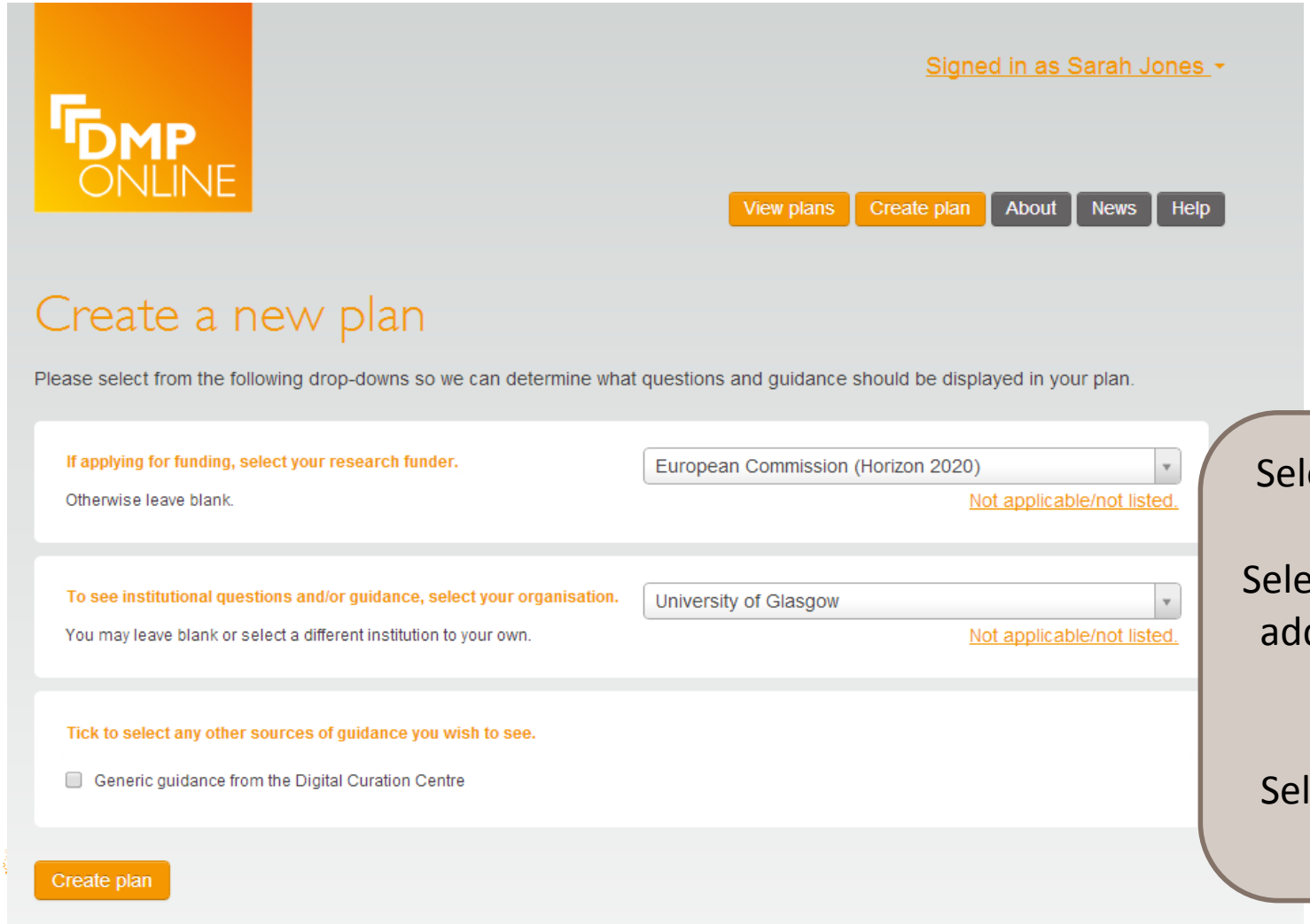


Sign up with your email address, organisation and password

Select 'other organisation' if yours is not listed

FOSTER

Creating a plan



Select funder (if any)

Select organisation for additional questions and guidance

Select other sources of guidance

FOSTER

Plan details: summary

This plan is based on:

Funder | Economic and Social Research Council

Answer questions | Export

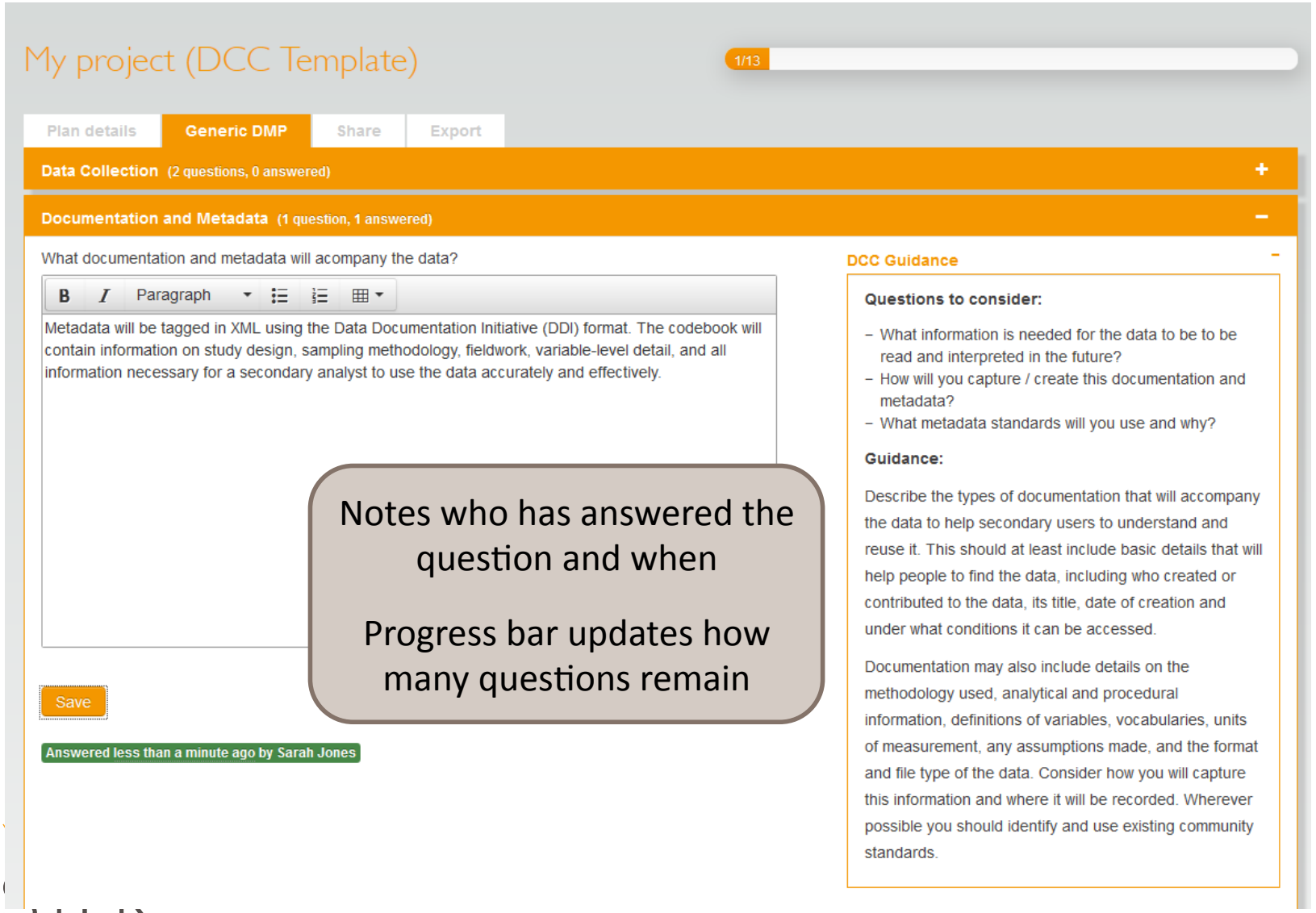
The ESRC requires that all applicants seeking ESRC funding include a statement on data sharing in the relevant section of the Je-S application form. If data sharing is not possible, the applicant must present a strong argument to justify their case.

Sections	Questions
Existing data	- An explanation of the existing data sources that will be used by the research project (with references) - An analysis of the gaps identified between the currently available and required data for the research
Information on the data that will be produced	- Methodologies for data collection - Data volume and data type, e.g. qualitative or quantitative data - Data quality, formats, standards documentation and metadata
Planned quality assurance and back-up procedures (security/storage)	- Quality Assurance - Back-Up
Management and archiving of collected data	- Plans for management and archiving of collected data
Overcoming data sharing difficulties	- Expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties.
Consent, confidentiality, anonymisation and other ethical considerations	- Explicit mention of consent, confidentiality, anonymisation and other ethical considerations
Copyright and intellectual property ownership of the data	- Copyright and intellectual property ownership of the data
Responsibilities for data management and curation	- Responsibilities for data management and curation within research teams at all participating institutions

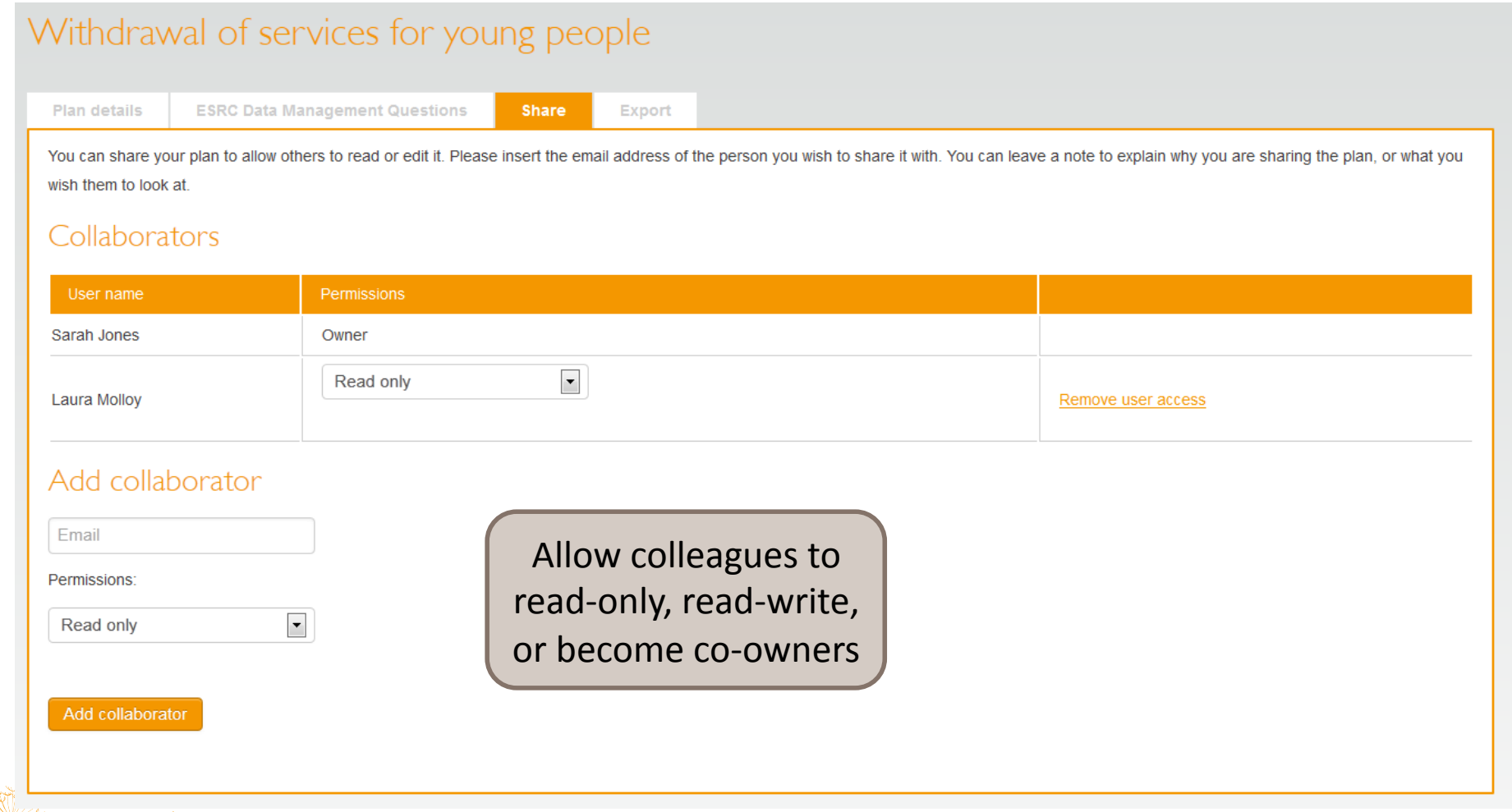
Summary of the sections and questions in your DMP

FOSTER

Answering questions

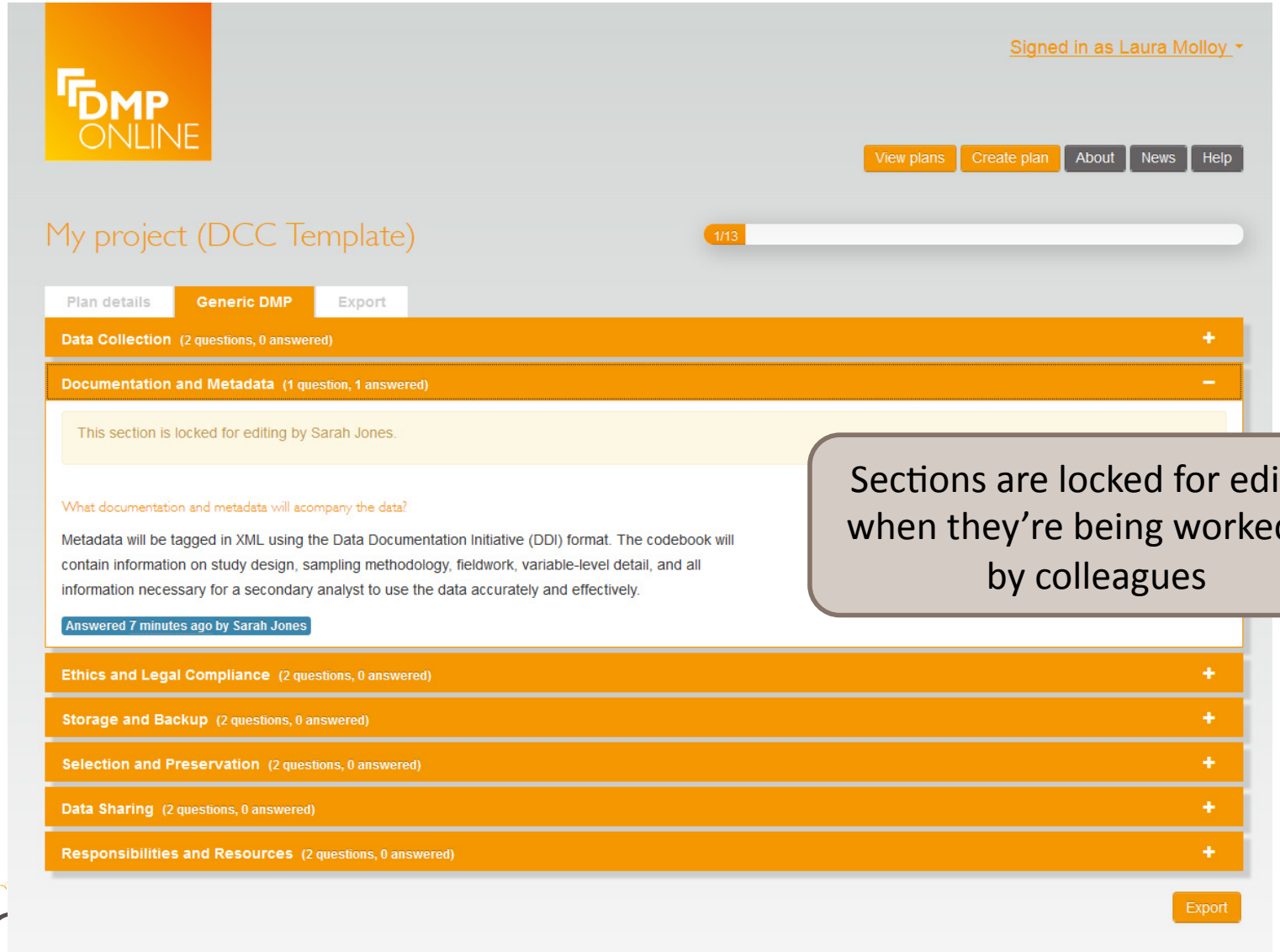


Sharing plans



FOSTER

Co-writing DMPs



Sections are locked for editing when they're being worked on by colleagues

FOSTER

Exporting DMPs

Can export as plain text, docx, PDF, html...

Withdrawal of services for young people

ESRC Data Management Questions

Existing data

Questions	Answers
- An explanation of the existing data sources that will be used by the research project (with references)	The ESDS archive has been systematically searched using a series of search terms related to and derivative of 'public service withdrawal', 'impacts', and/or 'children' and 'young people'. Our overall assessment is that there are no datasets that will adequately address the aims of this project. The following datasets are tangentially related to this project.
- An analysis of the gaps identified between the currently available and required data for the research	Given the contemporary nature of the proposed project, we know of no datasets that cover users' (especially young people's) views and experiences of austerity measures and service withdrawal. The proposed project will therefore capture new and unprecedented data, for which there is an evident demand among national and regional stakeholders, decisionmakers and service-providers (see 'Pathways to Impact' attachment). Moreover, whilst several datasets incorporate longitudinal data, none includes data gleaned from oral history and multigenerational family interview methods central to the proposed project. The proposed project therefore represents an extension to the methods and data quality of the tangentially-related projects listed in section 1. Finally, the novel use of 'impact' activities to generate data (e.g. workshops, mapping software) exceeds the scope of all extant and even tangentially-related datasets.

Information on the data that will be produced

Questions	Answers
- Data volume and data type, e.g. qualitative or quantitative data	The project shall generate new quantitative data (Bristol Online Survey outputs, SPSS data and outputs), qualitative data (digital audio files, audio transcripts, digital photographic and video data, workshop outputs, NVivo files), and mapping data (TIFF files). Metadata, in the form of pdfs and Excel spreadsheets, shall be used to facilitate the management and archiving of these data. Data shall be stored in password-protected folders on the host institution's secure servers. Data transfer between the PI and Co-I shall take place via face-to-face meetings.
	Quantitative data shall be generated from an anonymous online survey (target 10,000 responses). The survey will be administered via Bristol Online Surveys (BOS) software: a secure, quality-assured, widely-used online survey tool. Data will be exported to SPSS for analysis.

FOSTER

Institutions can customise the tool by…

- Adding templates
- Adding custom guidance
- Providing example or suggested answers
- Monitoring usage within their organisation
- Offering non-English language versions

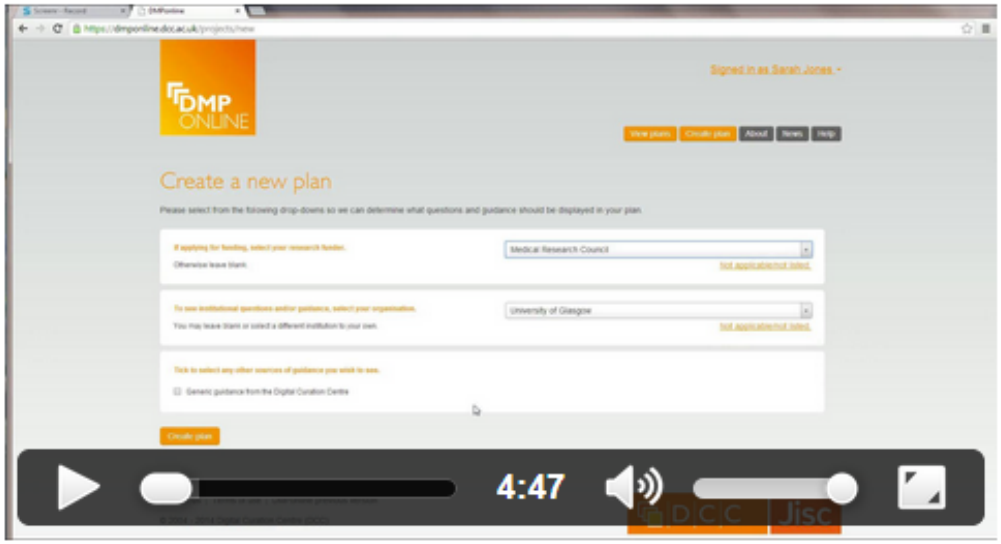
www.dcc.ac.uk/news/customising-dmponline-admin-interface-launches

FOSTER

More information

Screencast on how to use DMPonline

http://www.screenr.com/PJHN



Customising DMPonline

www.dcc.ac.uk/news/customising-dmponline-admin-interface-launches

GitHub

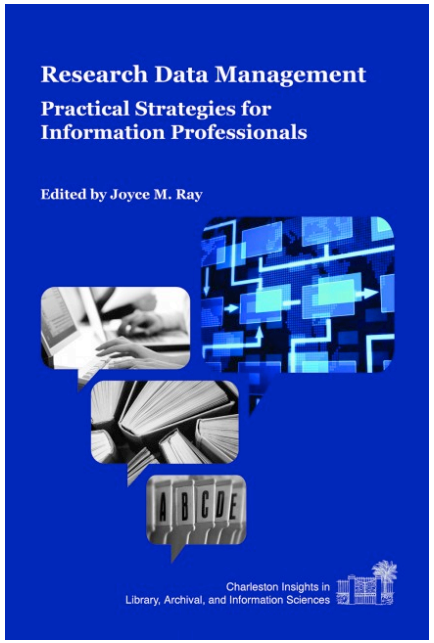
Get the code, amend it, run a local instance, flag issues, request features...

https://github.com/DigitalCurationCentre/DMPonline_v4

FOSTER

4b. Non-DCC tools and resources

- Book chapter
  - Sallans, A. and Lake, S. (2014) “Data Management Assessment and Planning Tools”, in Ray (ed.) Research Data Management, Purdue University Press
- DMPTool
- UKDA guidance
- NERC guidance
- European Union resources
- Resources from other universities, e.g. Oxford (http://researchdata.ox.ac.uk/)



DMPTool

UK • DATA ARCHIVE

FOSTER

4c. Examples

- There are lots of data management plans available on the Web. The DCC provides links to a number of DMPs via http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples
- Remember that there is no magic bullet, and no one-size-fits-all solution!
- Much of the benefit of data management planning lies in *the process of planning*, over and above the plans produced at the end of the process
- DMP is a communication activity, between the data collectors and their contemporaries (project partners and funders) and with future data re-users

FOSTER

THANK YOU

Martin Donnelly
Digital Curation Centre
University of Edinburgh

martin.donnelly@ed.ac.uk
Twitter: @mkdDCC

www.dcc.ac.uk

www.fosteropenscience.eu

FOSTER

Image credits: slides 3 and 6, https://www.flickr.com/photos/jayd/; slide 9, https://www.flickr.com/photos/junglearctic/; slide 17, https://twitter.com/AFC_Fisher/