Facilitate Open Science Training for European Research

The role of Open Access and Open Data, and their practical implementation in research institutions

Iryna Kuchma
EIFL
Kiev, Ukraine

Martin Donnelly
Digital Curation Centre
University of Edinburgh, Scotland

FOSTER event, Riga, 20 October 2015

Agenda

Time	Topic	Who
13:00-13:10	Introductions / workshop overview	All
13:10-13:40	Open Access to publications: a detailed look at policies	Iryna
13:40-14:10	Exercise: drafting an Open Access policy	Iryna
14:10-14:30	Group discussion	All
14:30-15:00	Research data management: a bit more detail	Martin
15:00-15:20	Coffee break	-
15:20-15:50	Exercise: developing a data management plan	Martin
15:50-16:10	Group discussion	All
16:10	Ends	

FOSTER

Facilitate Open Science Training for European Research

Open Access to publications: a detailed look at policies

Iryna Kuchma
EIFL
Kiev, Ukraine

FOSTER event, Riga, 20 October 2015

FOSTER

Facilitate Open Science Training for European Research

Exercise: drafting an Open Access policy

Iryna Kuchma
EIFL
Kiev, Ukraine

FOSTER event, Riga, 20 October 2015

Facilitate Open Science Training for European Research

Research Data Management: a bit more detail…

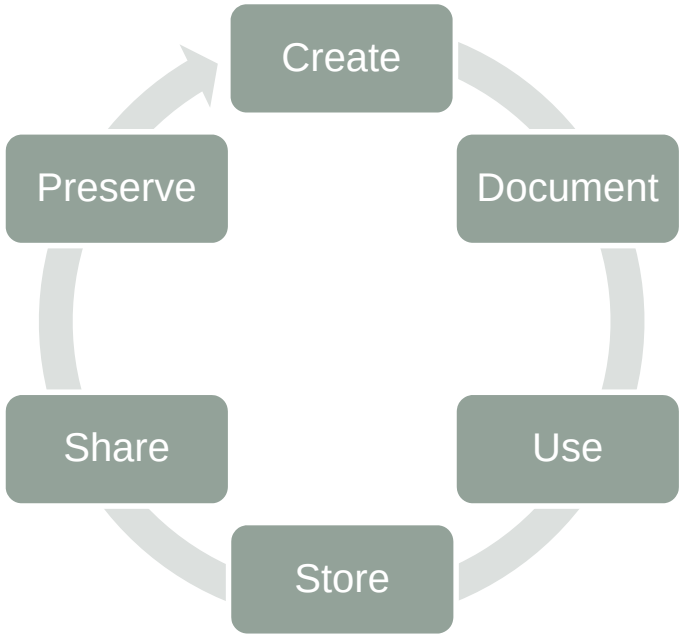
Martin Donnelly
Digital Curation Centre
University of Edinburgh, Scotland

FOSTER event, Riga, 20 October 2015

What is Research Data Management?

The active management of data throughout the lifecycle

- 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
- ...



FOSTER

Why is RDM an issue?

- Digital technology now used very widely in research, and is enabling new research and scientific paradigms
- Research funders and publishers know that digital research data can be expensive to produce but inexpensive to share, making reuse more feasible and desirable
- The challenge is to ensure digital research findings can be reproduced,cited, validated…

FOSTER

Reasons to manage and share data

Direct benefits for you

- To make your research easier!
- Stop yourself drowning in irrelevant stuff
- Make sure you can understand and reuse your data again later
- Advance your career – data is growing in significance

Research integrity

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

Potential to share data

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

FOSTER

Which data need to be kept?

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

FOSTER

Data Management Plans

It's useful to consider how you will manage and share your data in practice. Many research funders and institutions now ask for these details in a DMP…

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be shared/made available?
- If not, why? e.g. ethics & IP issues, embargoes, confidentiality
- How will this data be curated and preserved?

www.dcc.ac.uk/resources/data-management-plans/checklist

FOSTER

Lots of funders require a DMP

Their focus is often on data sharing

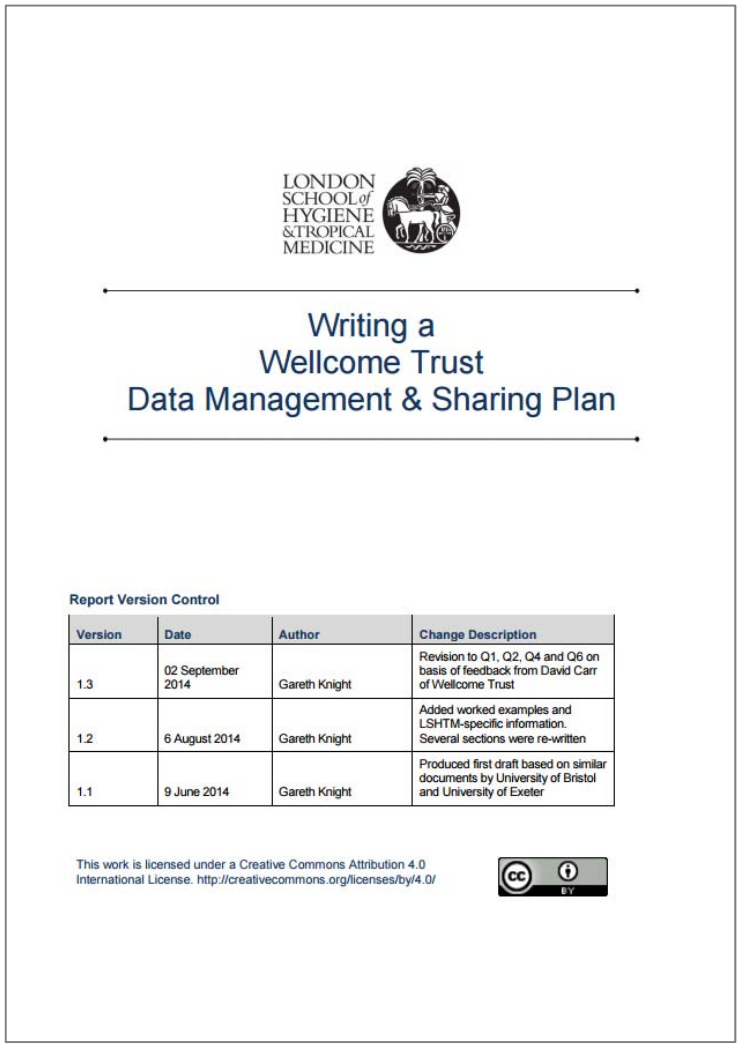
- Which data will be shared?
- When will it be shared?
- With whom?
- How will the data be shared?
- Will any restrictions or conditions govern use?
- ….

FOSTER

Guidance on writing a DMP

- Explains what is asked for
- Gives example answers
- Suggests best practices
- Provides links to standards, tools and support

www.lshtm.ac.uk/research/researchdataman/plan/wellcometrust_dmp.pdf



LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Writing a
Wellcome Trust
Data Management & Sharing Plan

Report Version Control

Version	Date	Author	Change Description
1.3	02 September 2014	Gareth Knight	Revision to Q1, Q2, Q4 and Q6 on basis of feedback from David Carr of Wellcome Trust
1.2	6 August 2014	Gareth Knight	Added worked examples and LSHTM-specific information. Several sections were re-written
1.1	9 June 2014	Gareth Knight	Produced first draft based on similar documents by University of Bristol and University of Exeter

FOSTER

What data will be generated?

Why is this important?

A good description of the data to be collected will help reviewers understand the characteristics of the data, their relationship to existing data, and any disclosure risks that may apply.

Example:

Data capture will be performed during months 8-20, at which point four types of data will be collected from the (approximately) 500 study subjects:

[1] Questionnaire: An interviewer-administered CAPI questionnaire will be performed covering medical history, health and socio-demographic circumstances

[2] Clinical: A clinical examination will be conducted to measure blood pressure, anthropometry, 12-lead digital ECG assessed for abnormalities; assessment of physical functions (grip strength, chair rise, walk speed) and cognition.

[3] Biological samples: Blood samples will be taken from each subject and centrifuged within 2 hours. Together with serum and plasma, it will aliquoted into 20 bar-coded cryovials and stored at -70C

[4] Interviews: It has previously been found that participant's partners tend to report higher levels of alcohol consumption and more frequent episodes of hazardous drinking than men themselves. We will interview partners of 1 in 4 of the study participants (selected randomly) in order to estimate the extent of under or over reporting of behaviours.

All 4 data types will be usable for future research in some form, subject to appropriate measures being implemented to protect participant confidentiality. Digital and physical outputs will be made available to bona fide researchers for health-related research, irrespective of their institution (university, charity, government, commercial) or location (UK or elsewhere).

FOSTER

When will you share the data?

Why is this important?

Research funders look for timely data sharing with minimal or no restrictions where possible. Embargo periods / delays to sharing should be justified and in line with standard practice for the field.

Example

Research papers written and published during the funding period will be made available with a subset of the (anonymised) data necessary to verify the research findings, in compliance with the Wellcome Trust's OA Policy.

The study team will make digital data (outputs 1, 2 and 4) available within 6 months of project completion. This embargo period is requested to allow time for additional analysis and further publication of research findings to be performed.

FOSTER

How can others access the data?

Why is this important?

If the data aren't discoverable, accessible and intelligible, they won't be reused. Data should be shared in a meaningful way.

Example

To enable potential users to learn of the dataset's existence, structured metadata describing its content will be created and made available in human readable and machine processable form. The LSHTM Data Repository will publish metadata in several metadata formats, including Dublin Core, via OAI-PMH, RSS and ATOM, for indexing by search engines and harvesting by research data catalogues.

To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data.

Biological samples (output 3) will be deposited with the UK BioBank for future use, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points. Similar to the above, applicants will be required to comply with a Data Transfer Agreement prior to gaining access to data. DNA extracted from biological samples will be normalized and plated at Lab Y for use in future studies, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points.

FOSTER

Are any limits to sharing required?

Why is this important?

As funders expect data to be shared, any restrictions need to be valid. Protection of human subjects is a fundamental tenet of research and an important ethical obligation for everyone.

Example

To protect participant confidentiality, data outputs will be anonymised prior to deposit. To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule.

Biological samples are limited and depletable, so access will need to be carefully controlled and coordinated. The quantity of sample that is provided will be judged against the potential benefits of the research project, with advice from appropriate experts as required.

FOSTER

State the long-term preservation plan

Why is this important?

Digital data need to be actively managed over time to ensure that they will always be available and usable. Depositing data resources with a trusted digital archive can ensure that they are curated and handled according to good practices in digital preservation.

Example

Data will be provided in file formats considered appropriate for long-term access, as recommended by the UK Data Service (http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx). For example, SPSS Portal format and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview transcripts. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule.

Appropriate documentation necessary to understand the data will also be provided. This will include high-level information on the study itself and a comprehensive data dictionary, which describes the purpose of each variable and the permitted values.

FOSTER

Some sample plans

- Technical appendix submitted to AHRC by Bristol Uni

  http://data.bris.ac.uk/files/2013/02/data.bris-AHRC-Technical-Plan-v21.pdf
- Rural Economy & Land Use (RELU) programme examples

  http://relu.data-archive.ac.uk/data-sharing/planning/examples
- UCSD example DMPs (20+ scientific plans for NSF)

  http://rci.ucsd.edu/dmp/examples.html
- My DMP – a satire (what not to write!)

  http://ivory.idyll.org/blog/data-management.html
- Further examples:

  www.dcc.ac.uk/resources/data-management-plans/guidance-examples

FOSTER

Facilitate Open Science Training for European Research

Exercise: Developing a Data Management Plan

Martin Donnelly
Digital Curation Centre
University of Edinburgh, Scotland

FOSTER event, Riga, 20 October 2015

DMP roles and responsibilities

- I mentioned earlier that RDM is a **hybrid** activity, involving multiple stakeholder types.
- So who's involved?
  - The principal investigator (usually ultimately responsible)
  - What about the research assistants? (they may be more involved in day-to-day data management)
  - And the institution's funding office?
  - And the Library/IT/Research Funding office?
  - What about partners based in other institutions?
  - And commercial partners?
  - Etc…

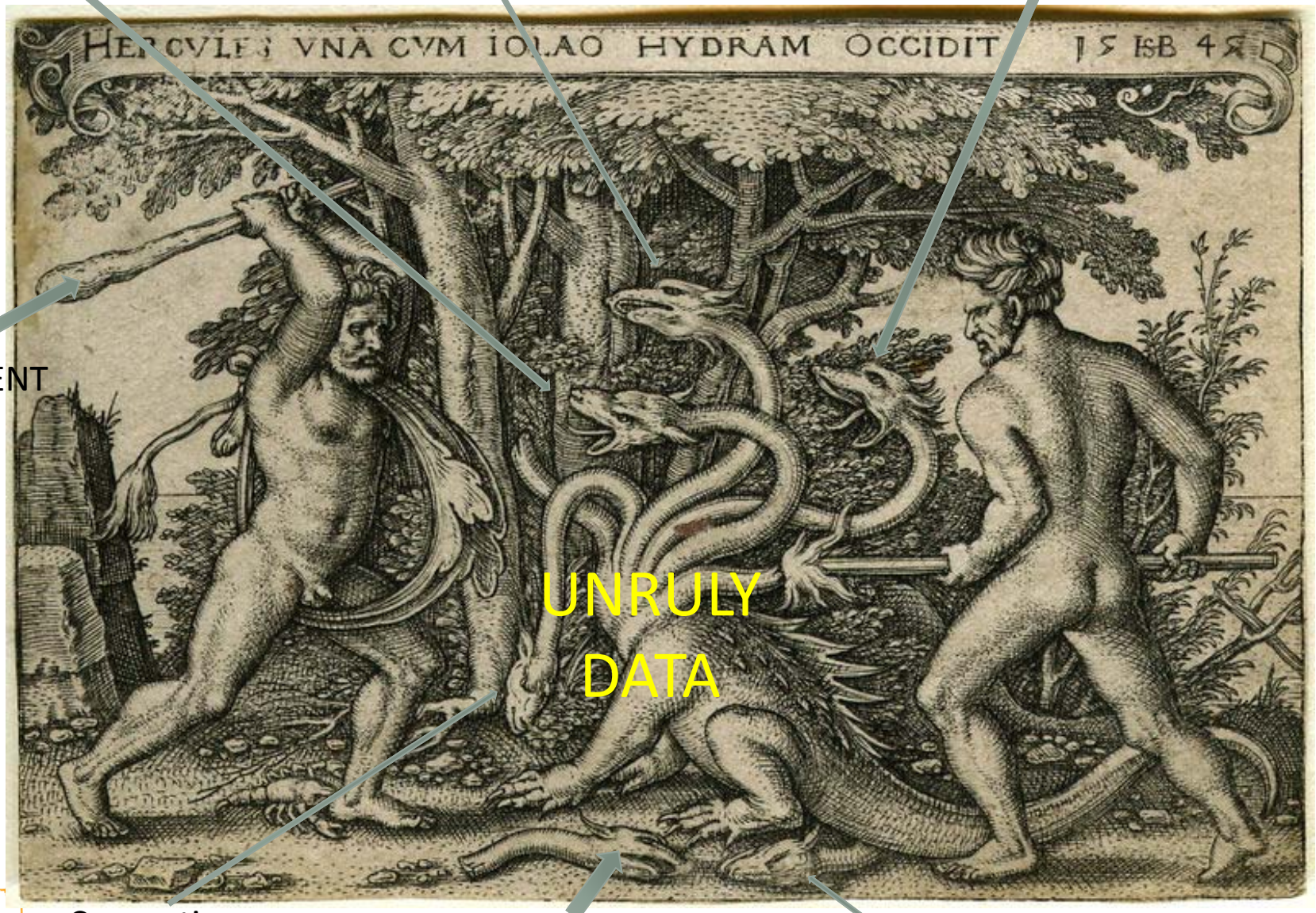
FOSTER

Researcher

Research Support Office

Data Library / Repository

DATA MANAGEMENT …PLAN?



FOSTER

Computing Support

Faculty Ethics Committee

Etc...

Interactive exercise: data management planning

§1. Administrative Data [basic details about the project]

§2. Data Collection

- ✓ What data will you collect or create?
- ✓ How will the data be collected or created?

§3. Documentation and Metadata

- ✓ What documentation and metadata will accompany the data?

§4. Ethics and Legal Compliance

- ✓ How will you manage any ethical issues?
- ✓ How will you manage copyright and Intellectual Property Rights (IPR) issues?

§5. Storage and Backup

- ✓ How will the data be stored and backed up during the research?
- ✓ How will you manage access and security?

§6. Selection and Preservation

- ✓ Which data should be retained, shared, and/or preserved?
- ✓ What is the long-term preservation plan for the dataset?

§7. Data Sharing

- ✓ How will you share the data?
- ✓ Are any restrictions on data sharing required?

§8. Responsibilities and Resources

- ✓ Who will be responsible for data management?
- ✓ What resources will you require to deliver your plan?

- Select one of the DMP Checklist headings (left), and brainstorm all the **stakeholders** you think might be involved (and how/why) – be specific!
- Remember to think of different **stages** of research: pre-award, in-project, post-project
- We'll have a short **reporting/discussion** session at the end
- http://www.dcc.ac.uk/resources/data-management-plans/checklist

FOSTER

§2. Data Collection

- ✓ What data will you collect or create?
- ✓ How will the data be collected or created?

§3. Documentation and Metadata

- ✓ What documentation and metadata will accompany the data?

§4. Ethics and Legal Compliance

- ✓ How will you manage any ethical issues?
- ✓ How will you manage copyright and Intellectual Property Rights (IPR) issues?

§5. Storage and Backup

- ✓ How will the data be stored and backed up during the research?
- ✓ How will you manage access and security?

§6. Selection and Preservation

- ✓ Which data should be retained, shared, and/or preserved?
- ✓ What is the long-term preservation plan for the dataset?

§7. Data Sharing

- ✓ How will you share the data?
- ✓ Are any restrictions on data sharing required?

§8. Responsibilities and Resources

- ✓ Who will be responsible for data management?
- ✓ What resources will you require to deliver your plan?

FOSTER

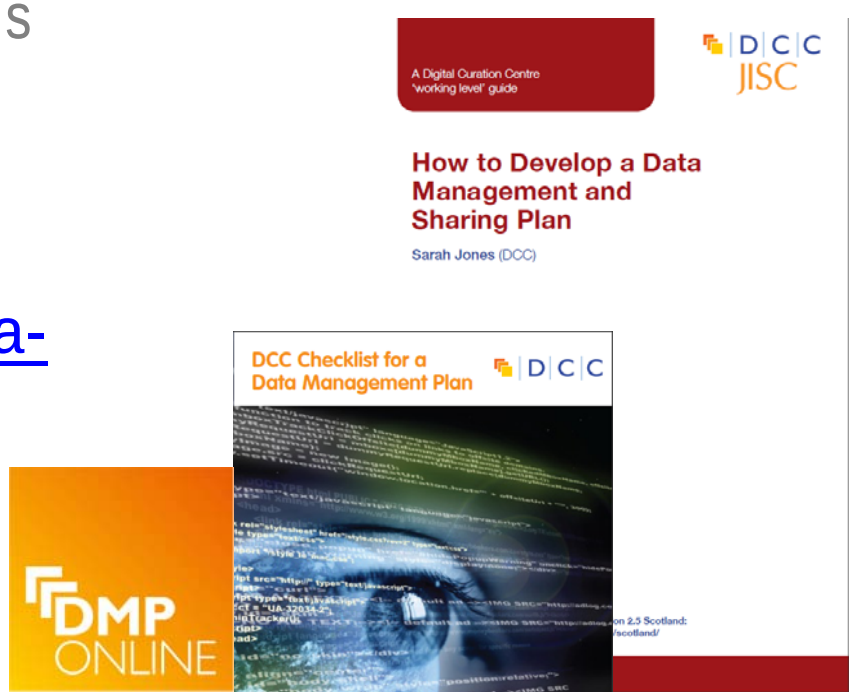
Data management planning exercise: outcomes

- It's not necessary – or even desirable – for every researcher (or librarian, or research administrator…) to become an expert in every aspect of data management
- Universities have an increasing obligation to provide infrastructure and support
- Specific expertise may be available from the research office, library, IT, departmental support staff, legal services, etc…

FOSTER

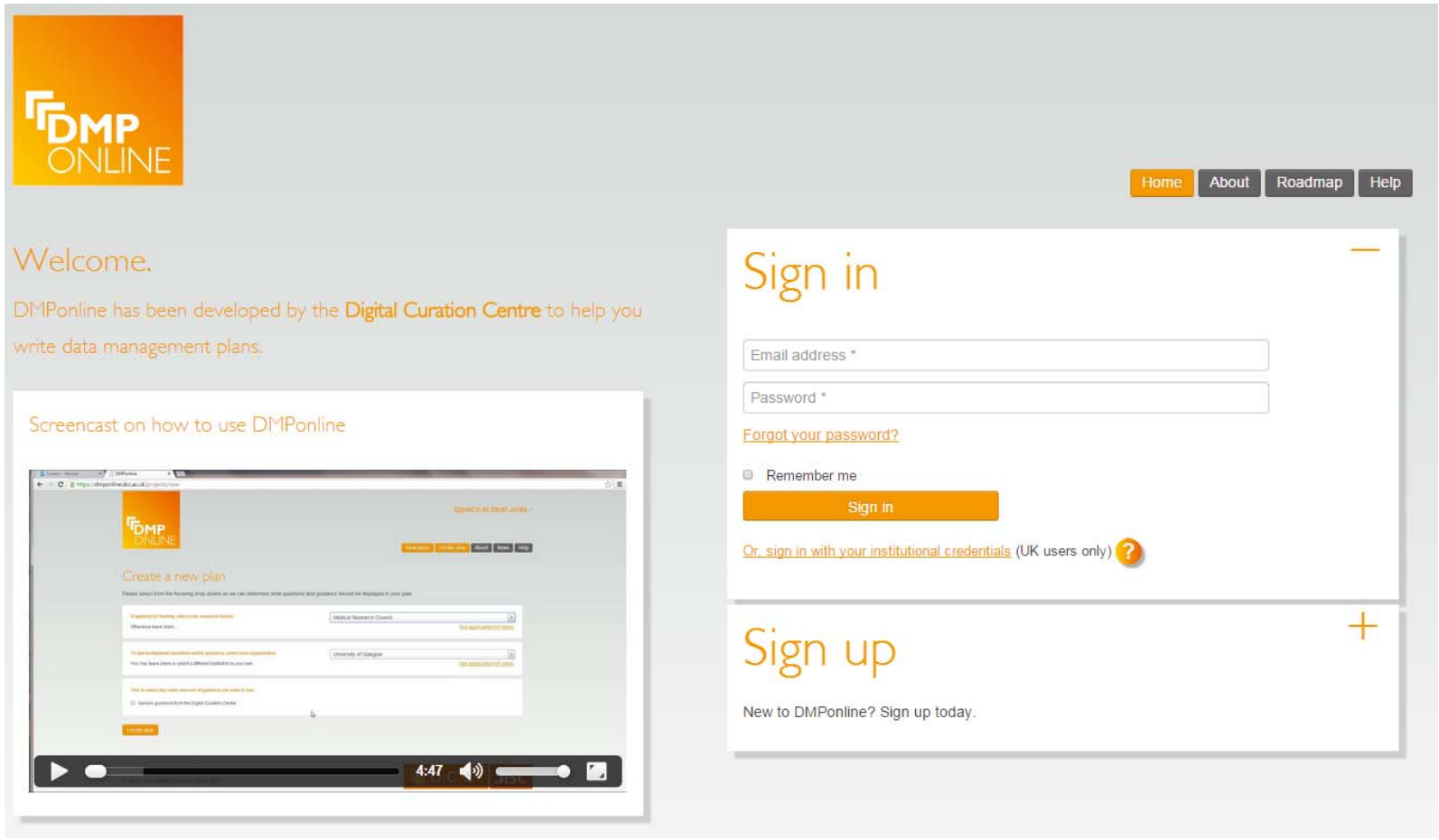
DCC support for Data Management Planning

- Checklist on what to include
- How-to guide on developing a plan
- Guidance on assessing plans (forthcoming)
- Webinars and training materials
- DMPonline tool
- Example DMPs: www.dcc.ac.uk/resources/data-management-plans



FOSTER

DMPonline



FOSTER

DMPonline: overview

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

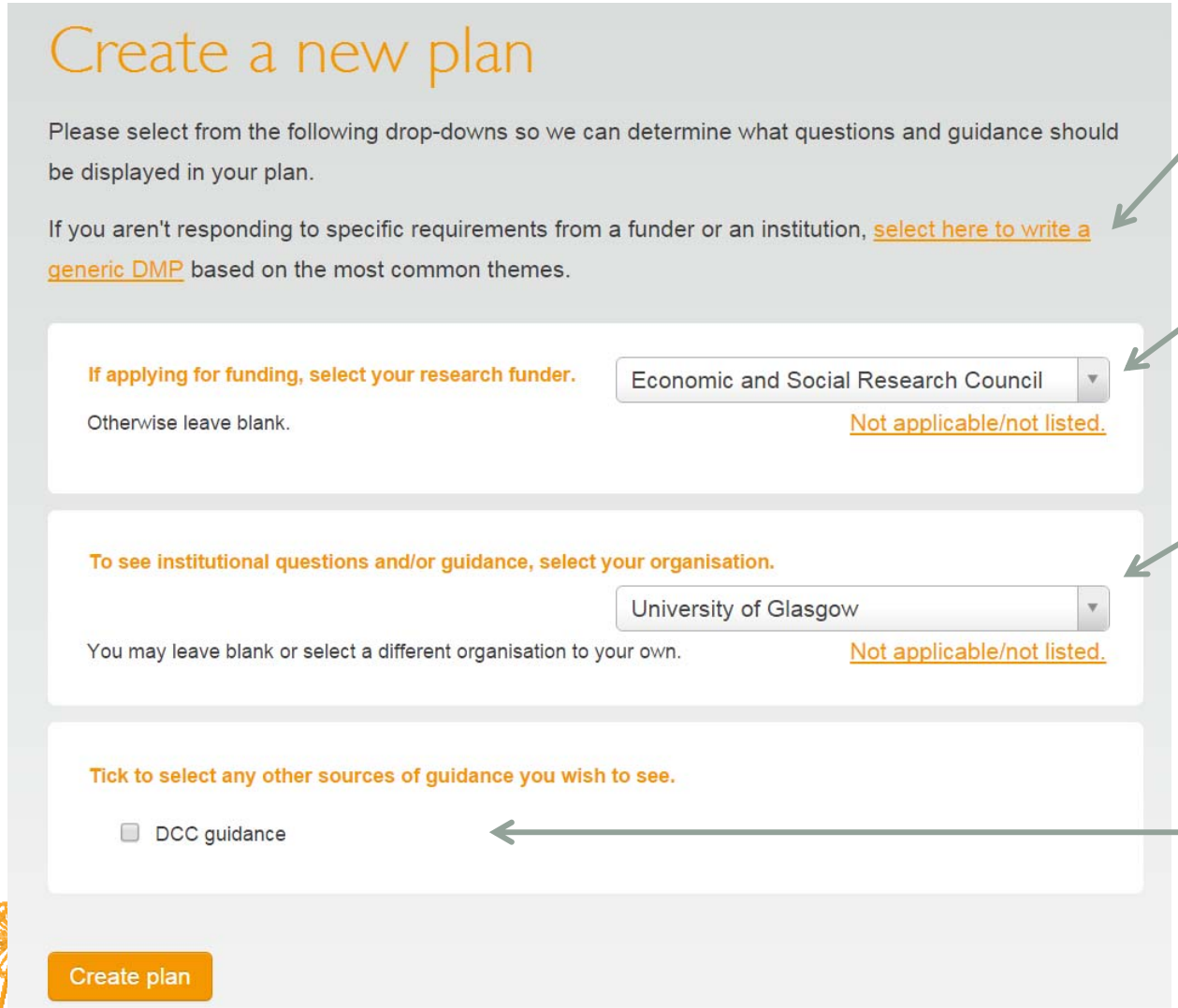
https://dmponline.dcc.ac.uk

Main features

- Templates for different requirements (funder or institution)
- Tailored guidance (funder, institutional, discipline-specific etc)
- Ability to provide examples and suggested answers
- Supports multiple phases (e.g. pre- / during / post-project)
- Granular read / write / share permissions
- Comment feature for collaboration
- Customised exports to a variety of formats
- Single-sign-on facility (for UK unis)

FOSTER

How the tool works



Click to write a generic DMP

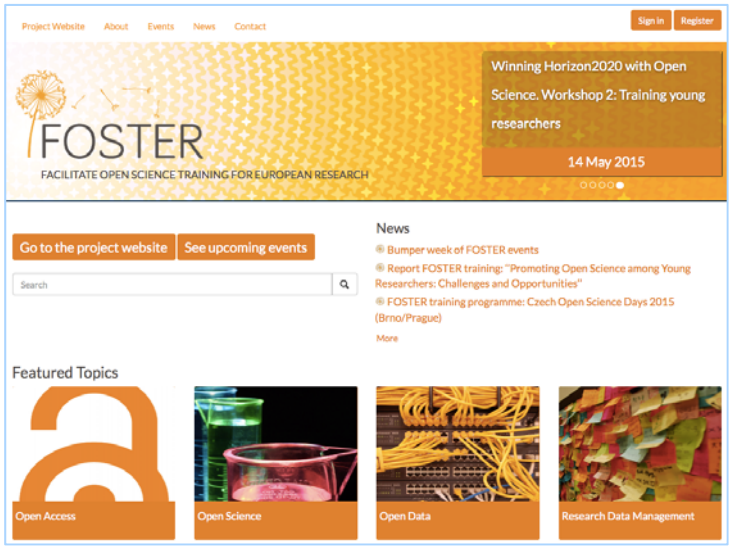
Or choose your funder to get their specific template

Pick your uni to add local guidance and to get the uni template if there isn't a funder one

Choose any additional optional guidance

Thank you: any questions?

- For more information about the FOSTER project:
  - Website: www.fosteropenscience.eu
  - Principal investigator: Eloy Rodrigues (eloy@sdum.uminho.pt)
  - General enquiries: Gwen Franck (gwen.franck@eifl.net)
  - Twitter: @fosterscience



- My contact details:
  - Email: martin.donnelly@ed.ac.uk
  - Twitter: @mkdDCC
  - Slideshare: http://www.slideshare.net/martindonnelly

FOSTER