Facilitate Open Science Training for European Research

SEMINAR:
Preparing research data for open access

December 10th 2014, Social Science Data Archives, Faculty of Social Sciences, University of Ljubljana
DATA MANAGEMENT PLANNING
HORIZON 2020 AND THE DMPONLINE TOOL

Martin Donnelly, Digital Curation Centre
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
The Digital Curation Centre (DCC)

The DCC (est. 2004) is...

- A UK national centre of expertise in digital preservation, with a particular focus on research data management (RDM)
- Based across three sites: Universities of Edinburgh, Glasgow and Bath
- Working with a number of UK universities to identify gaps in RDM provision and raise capabilities across the sector
- Also involved in a variety of international collaborations
# DCC networks and partnerships

<table>
<thead>
<tr>
<th>4C</th>
<th>Collaboration to Clarify the Costs of Curation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CODATA</td>
<td></td>
</tr>
<tr>
<td>DataCite</td>
<td></td>
</tr>
<tr>
<td>3TU</td>
<td>University of California</td>
</tr>
<tr>
<td>CDL</td>
<td>California Digital Library</td>
</tr>
<tr>
<td>SPARC</td>
<td></td>
</tr>
<tr>
<td>The Association of Commonwealth Universities</td>
<td></td>
</tr>
<tr>
<td>ANDS</td>
<td>Australian National Data Service</td>
</tr>
<tr>
<td>DataOne</td>
<td>Data Observation Network for Earth</td>
</tr>
<tr>
<td>OpenAIRE</td>
<td>Open Access Infrastructure for Research in Europe</td>
</tr>
<tr>
<td>SIM4RDM</td>
<td>SUPPRT INFRASTRUCTURE SCHEMES FOR RESEARCH DATA MANAGEMENT</td>
</tr>
<tr>
<td>DCC</td>
<td>Digital Preservation Coalition</td>
</tr>
<tr>
<td>RDA</td>
<td>Research Data Alliance</td>
</tr>
<tr>
<td>KE</td>
<td>Knowledge Exchange</td>
</tr>
<tr>
<td>GRDI2020</td>
<td>A Vision for Global Research Data Infrastructures</td>
</tr>
<tr>
<td>UKSG</td>
<td></td>
</tr>
<tr>
<td>COAR</td>
<td>Data Archiving and Networked Services</td>
</tr>
<tr>
<td>International Records Management Trust</td>
<td></td>
</tr>
<tr>
<td>nestor</td>
<td></td>
</tr>
<tr>
<td>DANS</td>
<td></td>
</tr>
<tr>
<td>SOD</td>
<td>Swedish National Data Service</td>
</tr>
<tr>
<td>OR2012</td>
<td>The 7th International Conference on Open Repositories</td>
</tr>
<tr>
<td>DEFF</td>
<td>Denmark's Electronic Research Library</td>
</tr>
<tr>
<td>CSC</td>
<td></td>
</tr>
<tr>
<td>RLUK</td>
<td>Research Libraries UK</td>
</tr>
<tr>
<td>CNI</td>
<td>Coalition for Networked Information</td>
</tr>
<tr>
<td>RUGIT</td>
<td>Russell Universities Group of IT Directors (RUGIT)</td>
</tr>
</tbody>
</table>
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
Data sharing and publication

Benefits of sharing / publishing data...

• TRANSPARENCY: The evidence that underpins research can be made open for anyone to scrutinise, and attempt to replicate findings. This leads to a more robust scholarly record.

• EFFICIENCY: Data collection can be funded once, and used many times for a variety of purposes.

• ACCESSIBILITY: Interested third parties can (where appropriate) access and build upon publicly-funded research resources with minimal barriers to access.

• PRESERVATION: Some data is unique. If lost, it can’t be replaced. Publishing data serves to preserve it.
Without intervention, data + time = no data

Vines et al. “examined the availability of data from 516 studies between 2 and 22 years old”

- The odds of a data set being reported as extant fell by 17% per year
- Broken e-mails and obsolete storage devices were the main obstacles to data sharing
- Policies mandating data archiving at publication are clearly needed

“The current system of leaving data with authors means that almost all of it is lost over time, unavailable for validation of the original results or to use for entirely new purposes” according to Timothy Vines, one of the researchers. This underscores the need for intentional management of data from all disciplines and opened our conversation on potential roles for librarians in this arena.


Vines et al., The Availability of Research Data Declines Rapidly with Article Age, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2013.11.014
Data management plans and planning (DMP)

• Data management planning underpins data management activities

• DMP 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
Benefits of data management planning (i)

• 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 in accord

• Plans help researchers to manage their data, to make informed decisions, to identify and avoid future problems, and to communicate with current and future users

• In terms of data security, if there are genuinely 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
Benefits of data management planning (ii)

• 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 higher quality standards
• More concretely, many funders ask for DMPs as part of the grant award process
• DMPs are increasingly required by universities too, e.g. “All new research proposals must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication”
What does a DMP look like?

Usually a brief statement defining...

- how the 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
- etc

_DMPs are often submitted as part of funding applications, but are useful whenever researchers are creating data, especially when the research involves multiple partners, disciplines, countries, etc..._
Who’s involved?

• RDM is a hybrid activity, involving multiple stakeholder types. DMP underpins RDM, so the same types of people will be involved...

  • The principal investigator (usually ultimately responsible)
  • And what about the research assistants? (who may be more involved in day-to-day data management)
  • And the institution’s funding office?
  • And Library/IT/Legal?
  • What about partners based in other institutions?
  • And commercial partners?
  • Etc
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
UK policies

• 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...
US policies

• 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
Australian policies

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
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
Open Access > Open Data

• FP7 featured an Open Access pilot, and this is now an across-the-board requirement in FP8 (Horizon 2020 / H2020)

• H2020 includes an Open Data pilot, and it seems likely that it will become an across-the-board requirement in FP9...

• This section of the presentation focuses on the data management planning aspects of this policy development, which underpins all RDM activities and provides wider benefits beyond European-funded research
DMP in Horizon 2020

- Horizon 2020 includes a data management (planning) pilot, spanning three phases
- Proposals covered: “Research and Innovation actions” 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
Horizon 2020: more detailed reqt’s (i)

- The Open Research Data Pilot aims to improve and maximise access to and re-use of research data generated by projects.
- The Pilot on Open Research Data will be monitored throughout Horizon 2020 with a view to further developing EC policy on open research.
- **SCOPE:** for the 2014-2015 Work Programme, the areas of Horizon 2020 that participate in the Open Research Data Pilot are:
  - Future and Emerging Technologies
  - Research infrastructures
  - part e-Infrastructures
  - Leadership in enabling and industrial technologies
  - Information and Communication Technologies
  - Societal Challenge: ‘Secure, Clean and Efficient Energy’
  - part Smart cities and communities
  - Societal Challenge: ‘Climate Action, Environment, Resource Efficiency and Raw materials’ - except raw materials
  - Societal Challenge: ‘Europe in a changing world - inclusive, innovative and reflective Societies’
  - Science with and for Society
- This corresponds to about €3 billion, or 20% of the overall Horizon 2020 budget in 2014 and 2015.
- Opt outs are possible, either total or partial (reasons: commercial exploitation, confidentiality, security, personal data etc)
- **COVERAGE:** The Open Research Data Pilot applies to two types of data:
  1) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
  2) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan.
Horizon 2020: more detailed reqt’s (ii)

• **Step 1:** Participating projects are required to deposit the research data described above, preferably into a research data repository. 'Research data repositories' are online archives for research data. They can be subject-based/thematic, institutional or centralised. Useful listings of research data repositories include the Registry of Research Data Repositories (www.re3data.org) and Databib (http://databib.org). In addition, it is expected that the Open Access Infrastructure for Research in Europe (OpenAIRE) will become an entry point for linking publications to underlying research data.

• **Step 2:** As far as possible, projects must then take measures to enable for third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) this research data.

• One straightforward and effective way of doing this is to attach Creative Commons Licence (CC-BY or CC0) to the data deposited (http://creativecommons.org/licenses/, http://creativecommons.org/about/cc0).

• At the same time, projects should provide information via the chosen repository about tools and instruments at the disposal of the beneficiaries and necessary for validating the results, for instance specialised software or software code, algorithms, analysis protocols, etc. Where possible, they should provide the tools and instruments themselves.

• **COSTS:** Costs relating to the implementation of the pilot will be eligible. Specific technical and professional support services will also be provided (e-Infrastructures WP).

(There’s a Horizon 2020 DMP template in DMPonline...)

FOSTER
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
DMPonline

A web-based tool to help researchers create and share Data Management Plans

• Provides funder questions and guidance
• Provides help from universities
• Can include examples and suggested (boilerplate) answers
• Free to use
• Mature (v1 launched April 2010)
• Code is Open Source (on GitHub)

https://dmponline.dcc.ac.uk
Welcome.
DMPonline has been developed by the Digital Curation Centre to help you write data management plans.

Sign up with your email address, organisation and password.
Select ‘other organisation’ if yours is not listed.
Creating a plan

Select funder (if any)

Select organisation for additional questions and guidance

Select other sources of guidance
Plan details: summary

This plan is based on:

Funder | Economic and Social Research Council

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.

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

Summary of the sections and questions in your DMP
Answering questions

My project (DCC Template)

Data Collection  (2 questions, 0 answered)

Documentation and Metadata  (1 question, 1 answered)

DCC Guidance

Questions to consider:
- What information is needed for the data to be read and interpreted in the future?
- How will you capture/create this documentation and metadata?
- What metadata standards will you use and why?

Guidance:
Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.

Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.

Notes who has answered the question and when
Progress bar updates how many questions remain
Sharing plans

Withdrawal of services for young people

You can share your plan to allow others to read or edit it. Please insert the email address of the person you wish to share it with. You can leave a note to explain why you are sharing the plan, or what you wish them to look at.

Collaborators

<table>
<thead>
<tr>
<th>User name</th>
<th>Permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarah Jones</td>
<td>Owner</td>
</tr>
<tr>
<td>Laura Molloy</td>
<td>Read only</td>
</tr>
</tbody>
</table>

Add collaborator

Email
Permissions:
Read only

Add collaborator

Allow colleagues to read-only, read-write, or become co-owners
Co-writing DMPs

Sections are locked for editing when they’re being worked on by colleagues.
## Exporting DMPs

- **Withdrawal of services for young people**

### ESRC Data Management Questions

#### Existing data

<table>
<thead>
<tr>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>- An explanation of the existing data sources that will be used by the research project (with references)</td>
<td>The ESOS 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.</td>
</tr>
<tr>
<td>- An analysis of the gaps identified between the currently available and required data for the research</td>
<td>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, decision-makers, 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.</td>
</tr>
</tbody>
</table>

#### Information on the data that will be produced

<table>
<thead>
<tr>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Data volume and data type, e.g. qualitative or quantitative data</td>
<td>The project shall generate new quantitative data (Bristol Online Survey outputs, SPSS data and outputs), qualitative data (digital audio files, audio transcripts, digital photographs, 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.</td>
</tr>
<tr>
<td></td>
<td>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.</td>
</tr>
</tbody>
</table>
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
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

For more details, see the blog post: [www.dcc.ac.uk/news/customising-dmponline-admin-interface-launches](http://www.dcc.ac.uk/news/customising-dmponline-admin-interface-launches)
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
Overview

1. Introduction / the DCC
2. Data management planning - short intro and benefits of a planned approach
3. DMP policies and expectations in different countries
4. The H2020 data management requirements
5. The DMPonline tool - overview and walkthrough
6. Customizing DMPonline for local use
7. Other DMP resources and guidance
DCC resources

- Book chapter
- Guidance, e.g. “How-To Develop a Data Management and Sharing Plan”
- DCC Checklist for a Data Management Plan
- Links to all DCC DMP resources via http://www.dcc.ac.uk/resources/data-management-plans
Non-DCC resources

• Book chapter

• DMPTool
• UKDA guidance
• NERC guidance
• European Union resources
• Resources from other universities, inc. USA
ABOUT THE FOSTER PROJECT

Facilitate Open Science Training for European Research
OBJECTIVES

• Support different stakeholders, especially young researchers, in adopting open access in the context of the European Research Area (ERA) and in complying with the open access policies and rules of participation set out for Horizon 2020;

• Integrate open access principles and practice in the current research workflow by targeting the young researcher training environment;

• Strengthen institutional training capacity to foster compliance with the open access policies of the ERA and Horizon 2020 (beyond the FOSTER project);

• Facilitate the adoption, reinforcement and implementation of open access policies from other European funders, in line with the EC’s recommendation, in partnership with PASTEUR4OA project.
METHODS

• Identifying already existing content that can be reused in the context of the training activities and repackaging, reformatting them to be used within FOSTER, and develop/create/ enhance contents if/where they are needed.

• Creation of the FOSTER Portal to support e-learning, blended learning, self-learning, dissemination of training materials/contents and Helpdesk.

• Delivery of face-to-face training, especially training trainers/multipliers that can carry on further training and dissemination activities, within their institutions, countries or disciplinary communities.
Thank you / Hvala

Martin Donnelly
Digital Curation Centre
University of Edinburgh

martin.donnelly@ed.ac.uk
@mkdDCC

For more about DCC tools, services and resources see www.dcc.ac.uk
or follow us on twitter @digitalcuration and #ukdcc