Facilitate Open Science Training for European Research

Digital Resources for Open Science

Martin Donnelly
Digital Curation Centre
University of Edinburgh (Scotland)

European Medical Students Association
Berlin, 14 September 2015
Overview

1. Background and context

2. Recap
   a. What is Open Science?
   b. What are the main benefits?
   c. What are the main problems?

3. Open Science in practice
   a. What does it mean for researchers?
   b. OA and data sharing policies

4. Open Science digital resources
   a. Open Access resources
   b. Data management tools and infrastructure
   c. Training and support initiatives
Overview

1. **Background and context**
2. **Recap**
   a. What is Open Science?
   b. What are the main benefits?
   c. What are the main problems?
3. **Open Science in practice**
   a. What does it mean for researchers?
   b. OA and data sharing policies
4. **Open Science digital resources**
   a. Open Access resources
   b. Data management tools and infrastructure
   c. Training and support initiatives
Background and context

• Open Science is situated within a context of ever greater transparency, accessibility and accountability

• The impetus for Openness in research comes from two directions:
  • **Ground-up** - OA began in the High Energy Physics research community, which saw benefit in not waiting for publication before sharing research findings (and data / code)
  • **Top-down** - Government/funder support, increasing public and commercial engagement with research

• The main goals of these developments are to lower barriers to accessing the outputs of publicly funded research (or ‘science’ for short), to speed up the research process, and to strengthen the quality, integrity and longevity of the scholarly record
The old way of doing research

1. Researcher collects data (information)
2. Researcher interprets/synthesises data
3. Researcher writes paper based on data
4. Paper is published (and preserved)
5. Data is left to benign neglect, and eventually ceases to be accessible
The new way of doing research

The DataONE lifecycle model

Plan
Collect
Assure
Describe
Preserve
Discover
Integrate
Analyze

PUBLISH

...and RE-USE
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. (“80 Percent of Scientific Data Gone in 20 Years” HNGN, Dec. 20, 2013, http://www.hngn.com/articles/20083/20131220/80-percent-of-scientific-data-gone-in-20-years.htm.)

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
Overview

1. Background and context

2. Recap
   a. What is Open Science?
   b. What are the main benefits?
   c. What are the main problems?

3. Open Science in practice
   a. What does it mean for researchers?
   b. OA and data sharing policies

4. Open Science digital resources
   a. Open Access resources
   b. Data management tools and infrastructure
   c. Training and support initiatives

5. About the FOSTER project
Open Access, Open Data, Open Science

• The Internet lowered the physical barriers to accessing knowledge, but financial barriers remained - indeed, the cost of online journals tended to increase much faster than inflation, and scholars/libraries faced a cost crisis

• Open Access (OA) originated in the 1980s with free-to-access Listserv journals, but it really took off with the popularisation of the Internet in the mid-1990s, and the subsequent boom in online journals

• As Open Access to publications became normal (if not ubiquitous), the scholarly community turned its attention to the data which underpins the research outputs, and eventually to consider it a first-class output in its own right. The development of the OA and research data management (RDM) agendas are closely linked as part of a broader trend in research, sometimes termed ‘Open Science’ or ‘Open Research’
  • “The European Commission is now moving beyond open access towards the more inclusive area of open science. Elements of open science will gradually feed into the shaping of a policy for Responsible Research and Innovation and will contribute to the realisation of the European Research Area and the Innovation Union, the two main flagship initiatives for research and innovation”
    [http://ec.europa.eu/research/swafs/index.cfm?pg=policy&lib=science]

• Open Science encourages - and indeed requires - heterogeneous stakeholder groups to work together for a common, societal goal
What is RDM?

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

Data management is a part of good research practice.

- RCUK Policy and Code of Conduct on the Governance of Good Research Conduct

What sorts of activities?

- **Planning** and **describing** data-related work before it takes place
- **Documenting** your data so that others can find and understand it
- **Storing** it safely during the project
- **Depositing** it in a trusted archive at the end of the project
- **Linking** publications to the datasets that underpin them
RDM: who and how?

- **RDM is a hybrid activity**, involving multiple stakeholder groups...
  - The researchers themselves
  - Research support personnel
  - Partners based in other institutions, commercial partners, etc
- **Data Management Planning (DMP) underpins and pulls together** different strands of 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 re-use potential
  - Support longer-term preservation
- **Data Management Plans are a means of communication**, with contemporaries and future re-users alike
Benefits of Open Science

- **SPEED**: The research process becomes faster

- **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

- **IMPACT and LONGEVITY**: Open publications and data receive more citations, over longer period

- **TRANSPARENCY and QUALITY**: 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
Benefits of Open Science: Impact and Longevity

“In genomics research, a large-scale analysis of data sharing shows that studies that made data available in repositories received 9% more citations, when controlling for other variables; and that whilst self-reuse citation declines steeply after two years, reuse by third parties increases even after six years.” (Piwowar and Vision, 2013)

“Data is necessary for reproducibility of computational research, but an equal amount of concern should be directed at code sharing.”

Benefits of Open Science: Financial

“Conservatively, we estimate that the value of data in Australia’s public research to be at least $1.9 billion and possibly up to $6 billion a year at current levels of expenditure and activity. Research data curation and sharing might be worth at least $1.8 billion and possibly up to $5.5 billion a year, of which perhaps $1.4 billion to $4.9 billion annually is yet to be realized.”

More open data for more users...

40+
Number of countries with government open data platforms*

90,000+
Data sets on data.gov (US site)*

1.4 million
Page views for the UK open data site in the summer of 2013

102
Cities that participated in 2013 International Open Data Hackathon Day

1 million+
Data sets made open by governments worldwide

... can lead to more value

$3 trillion
Approximate potential annual value enabled by open data in seven “domains”

3 billion
Metric tons of carbon dioxide equivalent emission reductions from buildings that could be identified through the use of open data

35
Hours per year could be saved by commuters from schedule changes based on open data

100,000+
Medical, health, and fitness apps for smartphones

50%+
Consumer share of potential value of open data

* As of 2013

J. Manyika et al. "Open data: Unlocking innovation and performance with liquid information" McKinsey Global Institute, October 2013
Benefits of Open Science: Speed

“If we are going to wait five years for data to be released, the Arctic is going to be a very different place.”

Bryn Nelson, Nature, 10 Sept 2009

http://www.nature.com/nature/journal/v461/n7261/index.html

https://www.flickr.com/photos/gsfc/7348953774/ - CC-BY
Overview

1. Background and context
2. Recap
   a. What is Open Science?
   b. What are the main benefits?
   c. What are the main problems?
3. Open Science in practice
   a. What does it mean for researchers?
   b. OA and data sharing policies
4. Open Science digital resources
   a. Open Access resources
   b. Data management tools and infrastructure
   c. Training and support initiatives
What does it mean for researchers?

• A disruption to previous working processes
• Additional expectations / requirements from the funders (and sometimes home institutions)
• But! It provides opportunities for new types of investigation
• And leads to a more robust scholarly record
What do I need to do?

1. Understand your funder’s policies (e.g. the EC Guidelines)
2. Check your intended publisher’s OA policy (e.g. via Sherpa Romeo)
3. Create a data management plan (e.g. with DMPonline)
4. Decide which data to preserve using the DCC How-To guide and checklist, “Five Steps to Decide what Data to Keep”
5. Identify a long-term home for your data (e.g. via re3data.org)
6. Link your data to your publications with a persistent identifier (e.g. via DataCite)
   • N.B. Many repositories, including Zenodo, will do this for you
7. Investigate EU infrastructure services and resources, e.g. EUDAT, OpenAIRE Plus, FOSTER, etc…
H2020 Open Data Pilot: specifics (i)

AIM
The Open Research Data Pilot aims to improve and maximise access to and re-use of research data generated by projects. It 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 participating 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 2014-2015.

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.
H2020 Open Data Pilot: specifics (ii)

STEP 1
• The data should be deposited, preferably in a dedicated research data repository. These may be subject-based/thematic, institutional or centralised.
• EC suggests the Registry of Research Data Repositories (www.re3data.org) and Databib (http://databib.org) for researchers looking to identify an appropriate repository.
• Open Access Infrastructure for Research in Europe (OpenAIRE) will also become an entry point for linking publications to data.

STEP 2
• So 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.
• EC suggests attaching 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.
H2020 Open Data Pilot: specifics (iii)

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), e.g. EUDAT and OpenAIRE, alongside support measures such as FOSTER.

OPT-OUTS
Opt outs are possible, either totally or partially. Projects may opt out of the Pilot at any stage, for a variety of reasons, e.g.
- if participation in the Pilot on Open Research Data is incompatible with the Horizon 2020 obligation to protect results if they can reasonably be expected to be commercially or industrially exploited;
- confidentiality (e.g. security issues, protection of personal data);
- if participation in the Pilot on Open Research Data would jeopardise the achievement of the main aim of the action;
- if the project will not generate / collect any research data;
- if there are other legitimate reasons to not take part in the Pilot (to be declared at proposal stage)
Overview

1. Background and context
2. Recap
   a. What is Open Science?
   b. What are the main benefits?
   c. What are the main problems?
3. Open Science in practice
   a. What does it mean for researchers?
   b. OA and data sharing policies
4. Open Science digital resources
   a. Open Access resources
   b. Data management tools and infrastructure
   c. Training and support initiatives
Open Access resources

• SHERPA services
• Pasteur4OA
• OpenAIRE
SHERPA services

SHERPA is investigating issues in the future of scholarly communication. It is developing open-access institutional repositories in universities to facilitate the rapid and efficient worldwide dissemination of research. SHERPA services and the SHERPA Partnership are both based at the Centre for Research Communications at the University of Nottingham.

SHERPA Services

- **RoMEO** - Publisher's copyright & archiving policies
- **JULIET** - Research funders archiving mandates and guidelines
- **OpenDOAR** worldwide Directory of Open Access Repositories
- **SHERPA Search** - simple full-text search of UK repositories

SHERPA Resources

- SHERPA speaking events - for future events see [CRC](#)
- Guidance on depositing material; copyright; and open access for authors
- Advocacy Materials for administrators, including past SHERPA presentations
- Links to other initiatives and related background information

Current SHERPA & Allied Projects

The SHERPA services and partnership are based at the Centre for Research Communications at the University of Nottingham. For more information on current projects see the [CRC website](#).

News

- Upgrade to SHERPA/JULIET Released
- RoMEO API - Manual, REST Requests, and Wish List
- Hungarian Version of RoMEO Released
- Útnak indul a RoMEO magyar változata
- Country Statistics for SHERPA/RoMEO
- Access Keys Available for SHERPA/RoMEO API v.2.9
- more news...

Jobs
PASTEUR4OA

• PASTEUR4OA supports the aim of encouraging the development of compatible and coherent policies on Open Access and Open Data in the European Union, according to the European Commission’s Recommendation on “Access to and preservation of scientific information” (July 2012) and in view of maximizing alignment with the Horizon 2020 policy on access to the research funded by the Commission.

• The project supports the development and/or reinforcement of Open Access strategies and policies at the national level and facilitate their coordination among all Member States. It will build a network of centres of expertise in Member States that will develop a coordinated and collaborative programme of activities in support of policymaking at the national level under the direction of project partners.
OpenAIRE

- A large scale initiative, with 50 partners from all EU countries, collaborating to promote open scholarship and improve the discoverability and reusability of research publications and data.
- Brings together stakeholders from research libraries, open scholarship organisations, national e-Infrastructure and data experts, IT and legal researchers
- National Open Access Desks (NOADs) will collect H2020 project outputs, and support research data management. The OpenAIRE platform is the technical infrastructure that pulls together and joins these large-scale collections of research outputs across Europe.
- The project will create workflows and services on top of this valuable repository content, enabling an interoperable network of repositories via the adoption of common guidelines, and easy upload into an all-purpose repository (i.e. Zenodo).
- OpenAIRE2020 will assist in monitoring H2020 research outputs and will be a key infrastructure for reporting H2020’s scientific publications as it will be loosely coupled to the EC’s IT backend systems.
Data tools and resources

- DMPonline
- EUDAT
- Zenodo
DMPonline

- Helps researchers write DMPs
- Provides funder questions and guidance
  - Includes a template 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](https://dmponline.dcc.ac.uk)
EUDAT

- EUDAT offers **common data services** through a geographically distributed, resilient network of 35 European organisations. These **shared services and storage resources** are distributed across 15 European nations and data is stored alongside some of Europe’s most powerful supercomputers.

- The EUDAT services address the full lifecycle of research data, covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data.

- The vision is to enable European researchers and practitioners from any academic discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure (CDI) conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe’s largest scientific data centres.

- Seeks to bridge the gap between research infrastructures and e-Infrastructures through an active engagement strategy, using the communities in the consortium as EUDAT beacons, and integrating others through innovative partnership approaches.

- Jisc and DCC are partners, and we’re working to integrate DCC’s DMPonline tool with the EUDAT suite of services / infrastructure.
Zenodo

- Zenodo is a free-to-use data archive, run by the people at CERN
- It accepts any kind of data, from any academic discipline
- It is generally preferable to store data in a disciplinary data centre, but not all scholarly subjects are equally well served with data centres, so this may make for a useful fallback option
Other data management resources (DCC)

- 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](http://www.dcc.ac.uk/resources/data-management-plans)
Data management resources (Non-DCC)

- Book chapter

- DMPTool
- UKDA guidance
- NERC guidance
- European Union resources
- Resources from other universities, inc. Oxford (http://researchdata.ox.ac.uk/)
Training support and resources

• FOSTER
• DCC training
• UK Data Archive
• Open Access training courses
OBJECTIVES

- To support different stakeholders, especially younger 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.

- To integrate open access principles and practice in the current research workflow by targeting the young researcher training environment.

- To strengthen institutional training capacity to foster compliance with the open access policies of the ERA and Horizon 2020 (beyond the FOSTER project).

- To 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 developing/creating/enhancing contents as required

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

• Delivery of face-to-face training, especially training trainers/multipliers who can deliver further training and dissemination activities, within institutions, nations or disciplinary communities
  
  • The EU is also funding other specific technical and professional support services via the e-Infrastructures WP, e.g. EUDAT and OpenAIRE
DCC training

Digital curation training for all

Our training programme aims to equip researchers and data custodians with the skills they need to share and preserve data effectively. Ultimately, tools and approaches will evolve over time, but if all stakeholders understand the bigger picture they will be in a better position to make critical decisions that best reflect their individual needs.

With this in mind, DCC training courses make use of the curation lifecycle model as a means of contextualizing the range and nature of roles and activities required to maintain access to data over time. The DCC also encourages the transfer of knowledge and best practice among data custodians, producers and users. In this way, you will be able to share your skills and the responsibility for data curation with others on your research team.

We've developed a range of training courses that meet a variety of training needs - from the absolute beginner to the more experienced data curator.

Arrange a DCC training session at your institution!
# UK Data Archive training

The UK’s largest collection of digital research data in the social sciences and humanities.

## CREATE & MANAGE DATA

<table>
<thead>
<tr>
<th>CREATE &amp; MANAGE DATA</th>
<th>ADVICE &amp; TRAINING</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAINING PROGRAMME &amp; EVENTS</td>
<td></td>
</tr>
</tbody>
</table>

### Guided walk through ReShare
- 22 September 2015
- Online 12.00-13.00

### Research Data Management
- 6 - 7 October 2015
- University of Manchester

### Human rights insights from data
- 29 - 30 October 2015
- Wivenhoe House Hotel, University of Essex

## EVENT ARCHIVE

### MANAGING AND SHARING RESEARCH DATA

SAGE publication

### CREATE & MANAGE DATA FAQ
Open Access training

UKeiG Training: Open Access - How it will change your (professional) life

David Ball

Date & Time:
Tuesday, 23 September 2014 - 9:30am to 4:30pm

This one-day workshop explores the definitions, economics and growth of Open Access and explores practicalities, challenges and opportunities for the future. Delegates will be encouraged to raise questions and debate issues of immediate concern throughout the day.

Who should attend:
Information and library professionals keen to understand the impact of the growth of Open Access on their work, their institutions and the current and future services they provide for their users.
Thank you / Danke

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

• My contact details:
  • Email: [martin.donnelly@ed.ac.uk](mailto:martin.donnelly@ed.ac.uk)
  • Twitter: @mkdDCC
  • Slideshare: [http://www.slideshare.net/martindonnelly](http://www.slideshare.net/martindonnelly)

This work is licensed under the Creative Commons Attribution 2.5 UK: Scotland License.