Research Data Management: What’s it all about?

Joy Davidson
Digital Curation Centre
Overview

- Overview of drivers pushing data management and sharing agenda and benefits
- What is data sharing?
- Planning to share (data management planning)
- Specific support roles
- Tips for developing RDM services
Drivers
“Data sets are becoming the new instruments of science”

Dan Atkins, University of Michigan
Funders’ expectations of public access

“Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.”

RCUK Common Principles on Data Policy

http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
EU Policies

Guidelines on Data Management in Horizon 2020

Version 1.0
11 December 2013

Ultimately funders expect:

- **timely release of data**
  - once patents are filed or on (acceptance for) publication

- **open data sharing**
  - minimal or no restrictions if possible

- **preservation of data**
  - typically 5-10+ years if of long-term value

See the RCUK Common Principles on Data Policy: [www.rcuk.ac.uk/research/Pages/DataPolicy.aspx](http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx)
“It was unbelievable. Its not science the way most of us have practiced in our careers. But we all realised that we would never get biomarkers unless all of us parked our egos and intellectual property noses outside the door and agreed that all of our data would be public immediately.”

Dr John Trojanowski, University of Pennsylvania
“It was a mistake in a spreadsheet that could have been easily overlooked: a few rows left out of an equation to average the values in a column.

The spreadsheet was used to draw the conclusion of an influential 2010 economics paper: that public debt of more than 90% of GDP slows down growth. This conclusion was later cited by the International Monetary Fund and the UK Treasury to justify programmes of austerity that have arguably led to riots, poverty and lost jobs.”

... validation of results
Benefits related to sharing data

- Avoids duplication
- Scientific integrity
- More collaboration
- Increased citation

Data Sharing
Who is involved in data sharing?

- the data sharer (researcher)
- the data repository
- support staff
- research participants
- the secondary data user
- commercial partners
Considerations

Ethics
- Informed consent – cover current and future use
- Confidentiality – is anonymisation appropriate?
- Access control – who, what, when?

IPR
- Clarify copyright before research starts
- Consider licensing options e.g. Creative Commons
Use appropriate repositories - institutional

http://datashare.is.ed.ac.uk

Research Data at Essex and DataPool at Southampton

www.dspace.cam.ac.uk

https://databank.ora.ox.ac.uk
Use appropriate repositories - external

**Research funders’ data centres**
- [ads](https://archaeologydataservice.ac.uk)
- [esds](http://www.esds.ac.uk)
- [British Atmospheric Data Centre](https://badc.nerc.ac.uk)

**Structured databases**
- [ChemSpider](https://www.chemspider.com)
- [PDB](https://www.rcsb.org)

**Community initiatives**
- [Zenodo](https://zenodo.org)
- [Figshare](http://figshare.com)
- [RE3Data](http://www.re3data.org)
Data Catalogues

Institutional

National

Funders
Tips for improving data reuse

How to License Research Data
Alex Ball (DCC)
www.dcc.ac.uk/resources/how-guides/license-research-data

How to Cite Datasets and Link to Publications
Alex Ball (DCC) and Monica Duke (DCC)
http://www.dcc.ac.uk/resources/how-guides/cite-datasets
Planning to Share
What is a data management plan (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
Why bother?

- compliance
- to help you manage your data
- to provide guidelines for everyone to work to
- to anticipate and avoid problems e.g. data loss, breaches
- to avoid duplication
The DCC offers DMP guidance and tools

www.dcc.ac.uk/resources/data-management-plans
Specific RDM roles for support staff
CRediT
An open standard for expressing roles intrinsic to research

http://credit.casrai.org/
#1 conceptualization
Ideas; formulation or evolution of overarching research goals and aims.

https://dmponline.dcc.ac.uk/

http://www.re3data.org/

http://rdrds.cloudapp.net/registry/
#2 methodology
Development or design of methodology; creation of models.

Standards, formats, consent

Informed consent

http://www.dcc.ac.uk/resources/metadata-standards
#4 validation
Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.

Selecting and appraising data to be retained
#5 formal analysis
Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.

Describing techniques in data catalogue records

http://www.software.ac.uk/

http://datacarpentry.org/
#6 investigation
Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

Data collection; reuse of third party data; searching for existing datasets for reuse

http://creativecommons.org/choose/
#8 data curation
Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.

Determine curation requirements

http://www.dcc.ac.uk/resources/developing-rdm-services
#11 visualization
Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.

Data models and visualisations

http://spatialinformationdesignlab.org/project_sites/library/catalog.html
#12 supervision
Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

Training, supervising and advising other researchers

https://www.fosteropenscience.eu/

http://www.researchinfonet.org/infolit/ridls/
#13 project administration
Management and coordination responsibility for the research activity planning and execution.

Understanding institutional and funding body requirements

[Image of Research Councils UK]

UCT POLICY for RESPONSIBLE CONDUCT OF RESEARCH

[Last revised August 2012]

This document must be read with the UCT Authorship Practices Policy, the UCT Research Ethics Code for Research Involving Human Participants, the UCT Research Ethics Code for Use of Animals in Research and Teaching, the [DRAFT] UCT Policy for Conflicts of Interest and of Commitment in Teaching and Research, the [DRAFT] UCT Policy and Procedures for Breach of Research Ethics Codes and Allegations of Research Misconduct in Research, the [DRAFT REVISED] Policy for Avoiding Plagiarism, the [DRAFT] Whistle-blowing for Academic Misconduct Policy.

https://www.uct.ac.za/downloads/uct.ac.za/about/policies/UCTresearchconductpolicy.pdf

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

https://www.researchfish.com/
#14 funding acquisition
Acquisition of the financial support for the project leading to this publication.

Aware of funding body mandates relating to RDM and data sharing and related costs

http://www.curationexchange.org/

http://www.ref.ac.uk/
#7 resources
Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.

Documenting research methods and processes

http://equipment.data.ac.uk/
#9-10 writing – original draft; review and editing, Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).

Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.

http://orcid.org/

Data Citation Principles
https://www.force11.org/datacitation

DataCite
Helping you to find, access, and reuse data
https://www.datacite.org/

http://www.metajnl.com/

Publishing

Distinguishing Yourself in Three Easy Steps
ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated changes between your professional activities and your personal activities. Find out more.

1. Register
2. Add Your Info
3. Use Your ORCID ID

http://www.datacite.org/
PA GO DA - PLAn de Gestión de DAtos

Crear su Plan de Gestión de Datos

El Programa Horizonte 2020 requiere que los proyectos que formen parte del Plano de Datos de Investigación en Abierto entreguen un Plan de Gestión de Datos completo durante los 6 primeros meses del proyecto.

Los planes de gestión de datos son una parte integral de las solicitudes de subvenciones - no pueden ser una idea de último momento; los revisores buscarán evidencia de que la gestión de datos está incluida en su propuesta, y que forma parte integral de su proceso de investigación. En el artículo 29.3 del M&G Model Grant Agreement: Multiple beneficiaries General MGA. December 2013 se establecen las obligaciones de los participantes en el Plano de Datos de Investigación en Abierto en lo que respecta a la gestión de los datos.

El documento Directrices sobre la Gestión de los Datos en Horizonte 2020 se dirige a los solicitantes y beneficiarios de los proyectos en el Marco del Plano de Datos de Investigación en Abierto y su objetivo es proporcionar indicaciones sobre cómo pueden cumplir con sus responsabilidades con respecto a la calidad de los datos de investigación, su intercambio y su seguridad.

Para saber más sobre el Plano de Datos de Investigación en Abierto, consulte nuestras FAQ's. Para saber más sobre el PGD como parte del Plano de Datos de Investigación en Abierto, consulte nuestras FAQ's. Para saber más sobre el Horizonte 2020, consulte estos documentos.

Recomendaciones para la preparación de su Plan de Gestión de Datos (PGD):

Antes de empezar a crear su Plan de Gestión de Datos es muy conveniente que previamente considere una serie de cuestiones relacionadas con la gestión de los datos en su proyecto.

Un punto de partida útil es la lista de control DCC (Digital Curation Centre) para un Plan de Gestión de Datos (en inglés) que podemos encontrar dentro de la selección de recursos del DCC (en inglés). Presenta los principales temas y preguntas que los investigadores pueden querer cubrir a la hora de escribir un Plan de...

http://pgd.consortiomadrono.es/
www.dcc.ac.uk/resources/data-management-plans
Tips for developing RDM Services

- Get senior management buy-in or champion
- Do not work in isolation – seek input from colleagues across the institution (ethics, IT, library)
- Reuse free tools and guidance wherever possible
- Be realistic - base planned services around existing infrastructure and available skills
- Plan to develop your strategy incrementally – make use of pilots and be sure to test emerging services and support systems
Thanks – any questions?

RDM guidance, tools and resources:
www.dcc.ac.uk/resources/

Follow us on twitter:
@digitalcuration and #DMPonline