

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

RCUK Review of e-Science 2009

BUILDING A UK FOUNDATION FOR THE TRANSFORMATIVE
ENHANCEMENT OF RESEARCH AND INNOVATION

*“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

Benefits of data sharing (1)

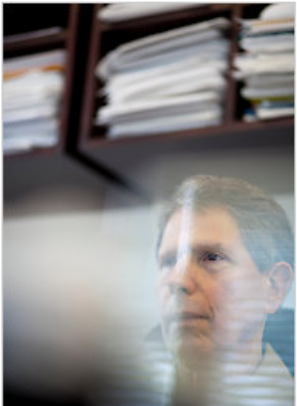
Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA

Published: August 12, 2010


In 2003, a group of scientists and executives from the [National Institutes of Health](#), the [Food and Drug Administration](#), the drug and medical-imaging industries, universities and nonprofit groups joined in a project that experts say had no precedent: a collaborative effort to find the biological markers that show the progression of [Alzheimer's disease](#) in the human brain.

 Enlarge This Image



Now, the effort is bearing fruit with a wealth of recent scientific papers on the early diagnosis of Alzheimer's using methods like PET scans and tests of spinal fluid. More than 100 studies are under way to test drugs that might slow or stop the disease.


And the collaboration is already serving as a model for similar efforts against [Parkinson's disease](#). A \$40 million project to look for biomarkers for Parkinson's, sponsored by the [Michael J. Fox Foundation](#), plans to enroll 600 study subjects in the United States and Europe.

 RECOMMEND

 TWITTER

 LINKEDIN

 COMMENTS
(155)

 SIGN IN TO
E-MAIL

 PRINT

 REPRINTS

 SHARE



"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

www.nytimes.com/2010/08/13/health/research/13alzheimer.html?pagewanted=all&_r=0

... scientific breakthroughs

Benefits of data sharing (2)

“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

The error that could subvert George Osborne's austerity programme

The theories on which the chancellor based his cuts policies have been shown to be based on an embarrassing mistake

Charles Arthur and Phillip Inman

The Guardian, Thursday 18 April 2013 21.10 BST



George Osborne says that Ken Rogoff, the man whose economic error has been uncovered, has strongly influenced his thinking. Photograph: Stefan Wermuth/PA

www.guardian.co.uk/politics/2013/apr/18/uncovered-error-george-osborne-austerity

Benefits related to sharing data

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

(Piwowar H. and Vision T.J 2013 , <https://peerj.com/preprints/1.pdf>)



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



Structured databases



Community initiatives



Data Catalogues



Home >

Enlighten: Research Data

About

Latest Additions

Search

Browse

Research Data Management

My Research Data

Welcome to Enlighten: Research Data

Atom RSS 1.0 RSS 2.0

Latest Additions

View items added to the repository in the past week.

Search Repository

Search the repository using a full range of fields. Use the search field at the top of the page for a quick search.

Browse Repository

Browse the items in the repository by subject.

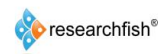
About this Repository

More information about this site.

Institutional

Enlighten: Research Data supports OAI 2.0 with a base URL of <http://researchdata.gla.ac.uk/cgi/oai2>

Funders



Username *

Password *

Request new password

Log in

Home

About Us

Our Members

Help

Government
Related

Researchfish is an easy-to-use research outcomes system for researchers and funding organisations.

Researchers and Delegates

Find Out More

Funding Organisations

Find Out More

Research Organisations/HEIs

Find Out More

Research Data Registry & Discovery

About Collections Parties Activities Services Themes



Search for Research Data



Browse by Subject Area



Browse by Map Coverage



Advanced Search

What's in the Research Data Registry and Discovery Service



Collections (49)

Research datasets or collections of research materials.



Parties (36)

Researchers or research organisations that create or maintain research datasets or collections.



Activities (0)

Projects or programs that create research datasets or collections.



Services (0)

Services that support the creation or use of research datasets or collections.

Spotlight on research data

National

Who contributes to the Research Data Registry and Discovery Service?

5 research organisations from around the UK contribute information to Research Data Australia.

See All

Share

Tips for improving data reuse

A Digital Curation Centre and JISC Legal
'working level' guide



How to License Research Data

Alex Ball (DCC)

www.dcc.ac.uk/resources/how-guides/license-research-data

A Digital Curation Centre 'working level' guide



How to Cite Datasets and Link to Publications

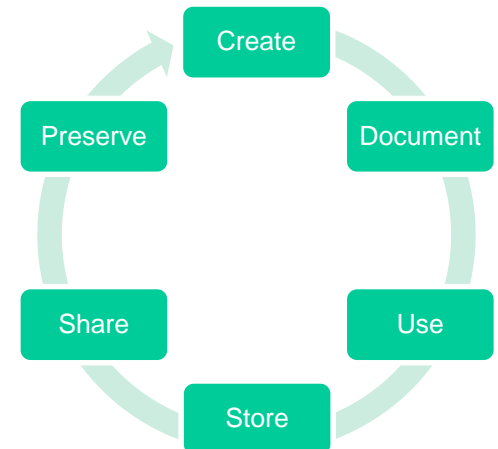
Alex Ball (DCC) and Monica Duke (DCC)

<http://www.dcc.ac.uk/resources/how-guides/cite-datasets>

Planning to Share

What is a data management plan (DMP)?

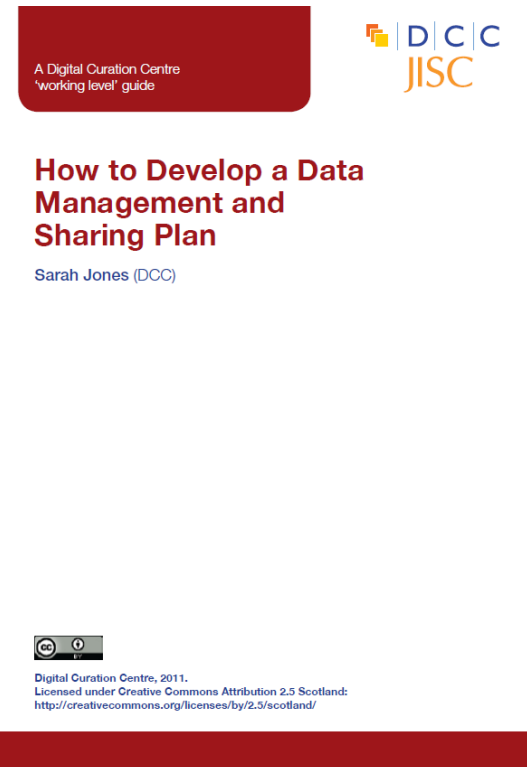
- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation



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.

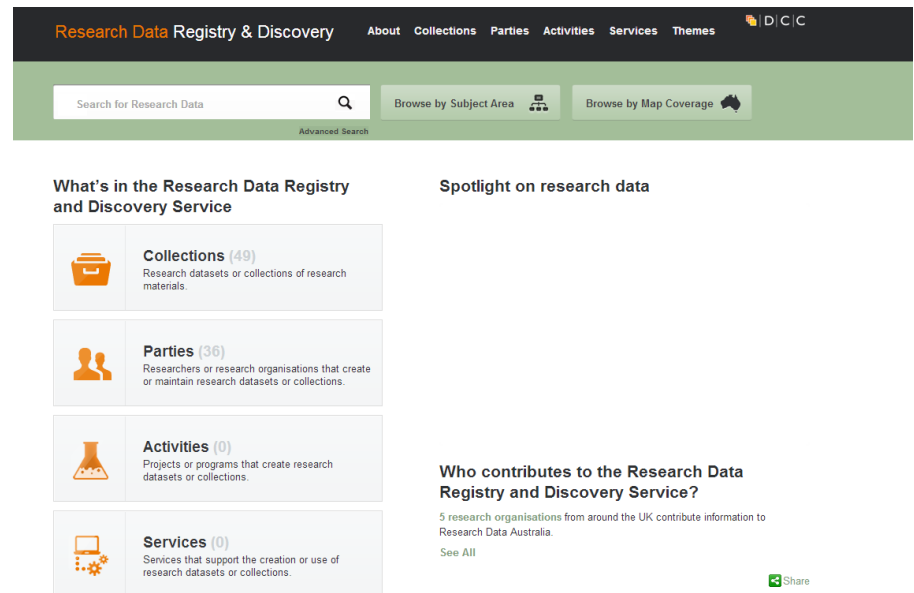


<https://dmponline.dcc.ac.uk/>

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://www.re3data.org/>

Literature and data review, data management planning



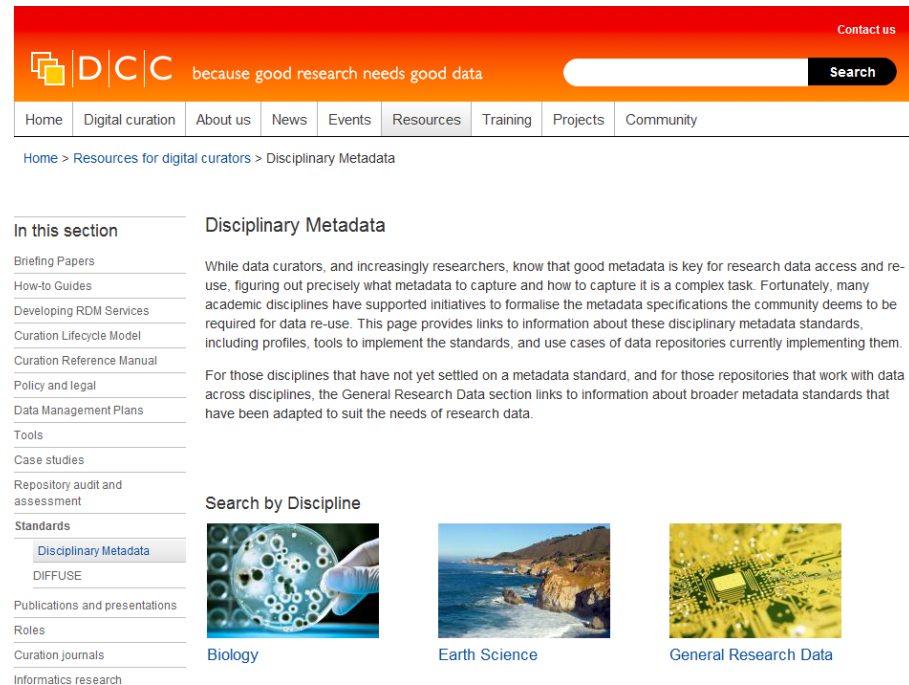
<http://rdrds.cloudapp.net/registry/>

#2 methodology

Development or design of methodology; creation of models.

Standards, formats, consent

Informed consent



DCC because good research needs good data

Contact us

Search

Home Digital curation About us News Events Resources Training Projects Community

Home > Resources for digital curators > Disciplinary Metadata

In this section




- Briefing Papers
- How-to Guides
- Developing RDM Services
- Curation Lifecycle Model
- Curation Reference Manual
- Policy and legal
- Data Management Plans
- Tools
- Case studies
- Repository audit and assessment
- Standards
 - Disciplinary Metadata**
 - DIFFUSE
- Publications and presentations
- Roles
- Curation journals
- Informatics research

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Search by Discipline

-  [Biology](#)
-  [Earth Science](#)
-  [General Research Data](#)

<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/>



DATA CARPENTRY

MAKING DATA SCIENCE MORE EFFICIENT

<http://datacarpentry.org/>

#6 investigation

Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

The image shows a four-panel interface for selecting a Creative Commons license. The panels are arranged in a 2x2 grid, connected by arrows indicating a sequential flow from top-left to top-right, then down to bottom-right, and finally down to bottom-left.

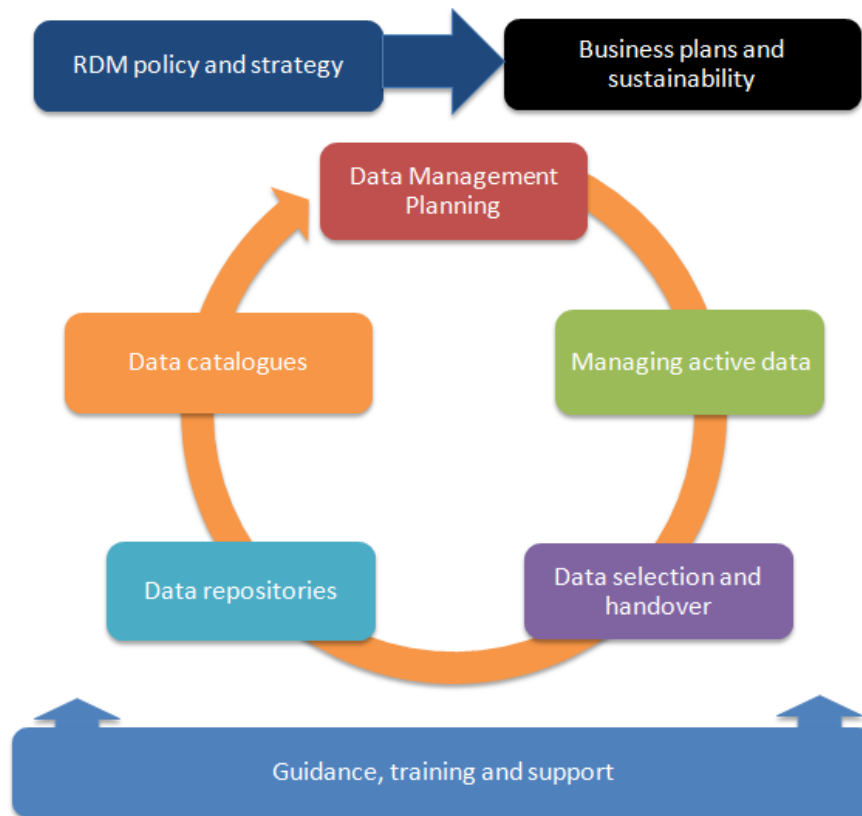
- License Features:** Contains two questions with radio button options.
 - "Allow adaptations of your work to be shared?" with options: Yes (selected), No, Yes, as long as others share alike.
 - "Allow commercial uses of your work?" with options: Yes (selected), No.
- Selected License:** Displays the chosen license, "Attribution 4.0 International", with the CC and BY icons. It includes the text "This is a Free Culture License!" and a green circular badge that says "Free Culture APPROVED FOR WORKS".
- Help others attribute you!:** A section for optional metadata with input fields for "Title of work", "Attribute work to name", "Attribute work to URL", "Source work URL", and "More permissions URL". It also has dropdown menus for "Format of work" (set to "Other / Multiple formats") and "License mark" (set to "HTML+RDFa").
- Have a web page?:** Shows the CC BY icon and states "This work is licensed under a Creative Commons Attribution 4.0 International License." It provides a code block for embedding the license on a website and radio buttons for "Normal Icon" (selected) and "Compact Icon".

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.

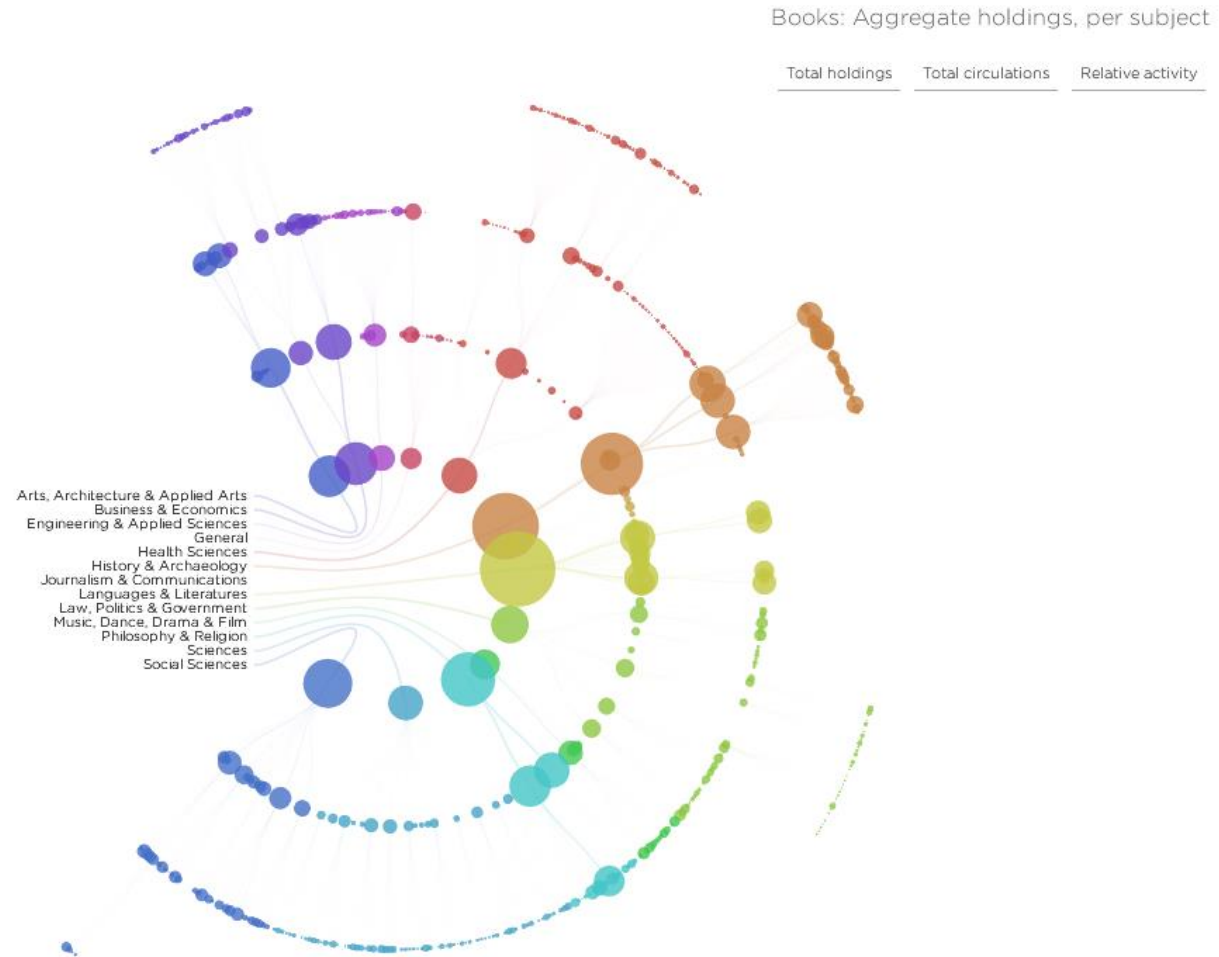


Determine curation requirements

#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.



<https://www.fosteropenscience.eu/>

Training, supervising and
advising other researchers



<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

The image shows two screenshots of research-related websites. The top screenshot is the 'Gateway to Research' website by Research Councils UK. It features a purple header with the Research Councils UK logo and the text 'Gateway to Research'. Below the header is a large banner with the text 'Welcome to the RCUK gateway to publicly funded research' and 'Search for and analyse information on the latest innovative research in the UK'. There is a search bar and a 'Log in' button. The bottom screenshot is the 'researchfish' website. It has a blue header with the 'researchfish' logo and a navigation menu with 'Home', 'About Us', 'Our Members', and 'Help'. The main content area has a blue background with the text 'Researchfish is an easy-to-use research outcomes system for researchers and funding organisations.' Below this are three white boxes with orange buttons: 'Researchers and Delegates', 'Funding Organisations', and 'Research Organisations/HEIs'. To the right of the Researchfish screenshot is the OGL (Open Government Licence) logo and text: 'All content is available under the Open Government Licence v2.0, except where otherwise stated'. Below the OGL text is the URL 'http://gtr.rcuk.ac.uk/'.

https://www.researchfish.com/

UCT POLICY for RESPONSIBLE CONDUCT OF RESEARCH

[Last revised August 2012]

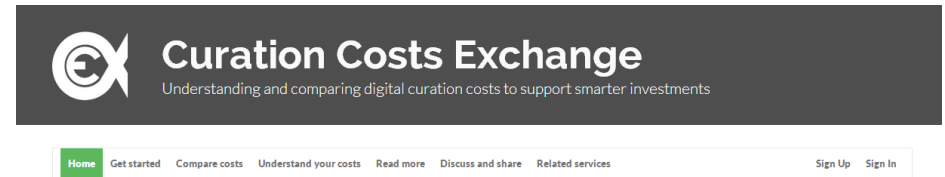
This document must be read with the UCT Authorship Practices Policy, the UCT Research Ethics Code for Research Involving Human Participants, the UCT Research Ethics Code for Use of Animals in Research and Teaching, the [DRAFT] UCT Policy for Conflicts of Interest and of Commitment in Teaching and Research, 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>

#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



All about the costs of curation

Understand what you and others are and should be spending in digital curation

Compare costs

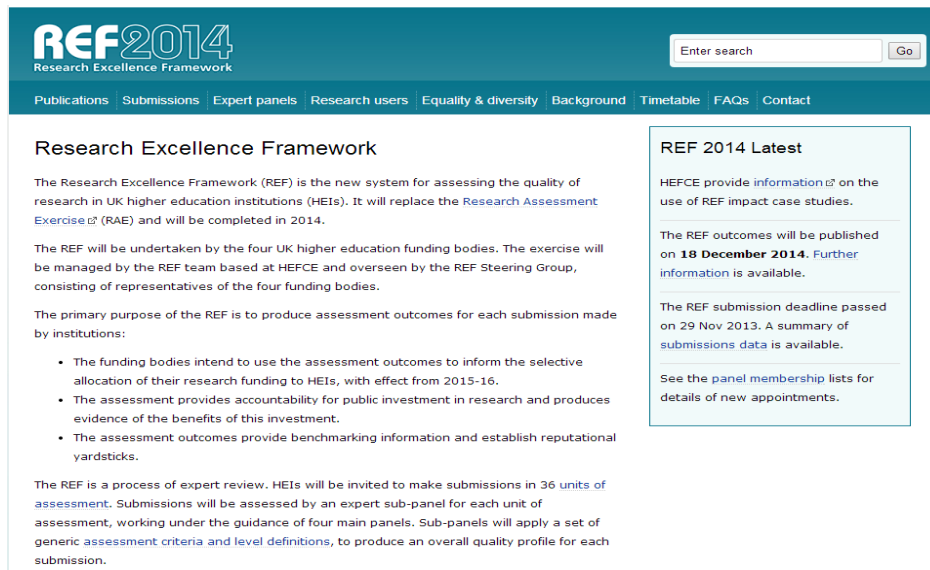
Understand curation costs and see how they compare with others



Understand costs

Understand how to assess your curation costs and how to make use of cost models to help you invest

<http://www.curationexchange.org/>

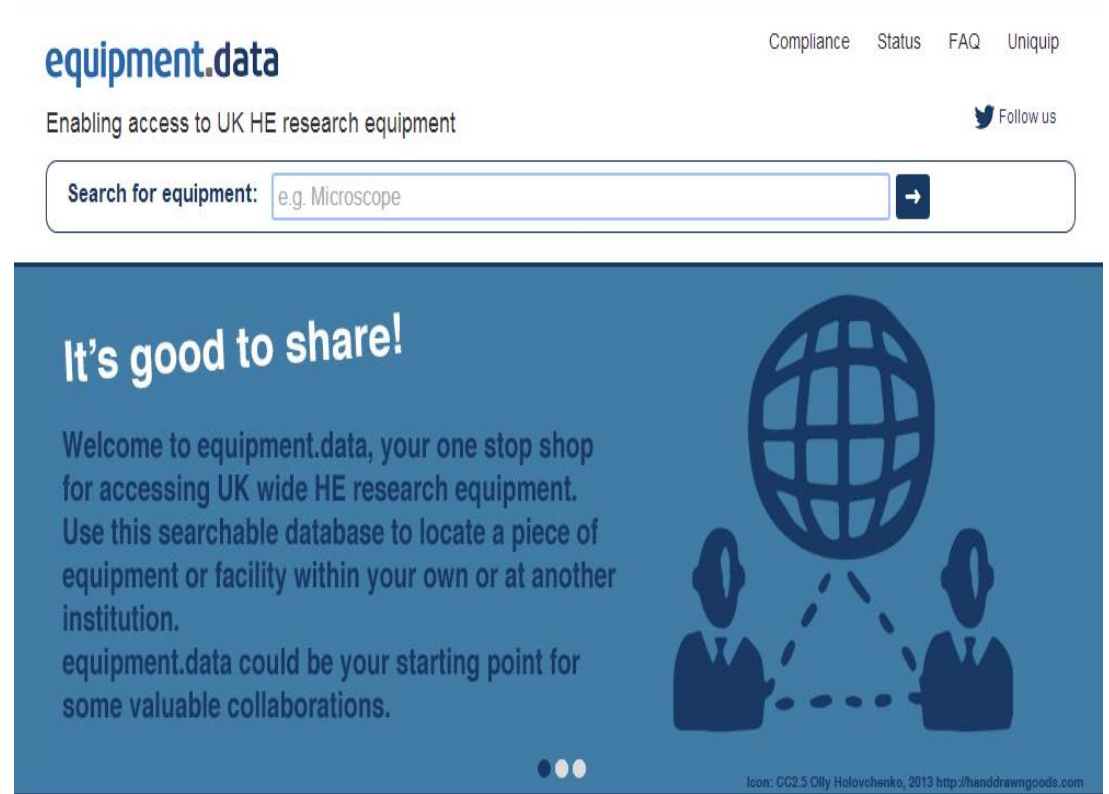


<http://www.ref.ac.uk/>

#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.



ORCID
Connecting Research and Researchers

FOR RESEARCHERS FOR ORGANIZATIONS ABOUT HELP SIGN IN

DISTINGUISH 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 linkages between you and your professional activities ensuring that your work is recognized. [Find out more.](#)

- 1 REGISTER** Get your unique ORCID identifier. [Register now!](#)
Registration takes 30 seconds.
- 2 ADD YOUR INFO** Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn).
- 3 USE YOUR ORCID ID** Include your ORCID identifier on your Website, when you submit publications, apply for grants, and in any research workflow to ensure you get credit for your work.

<http://orcid.org/>



DC¹
Data Citation Principles

<https://www.force11.org/datacitation>


Publishing



DataCite

Helping you to find,
access, and reuse data

<https://www.datacite.org/>



UP metaJournals beta


ubiquity press **open access**

Home About



DATASETS

Data papers highlight openly archived data with high reuse potential, and provide recognition for the producers of the data.



SOFTWARE

Software papers help you to locate openly archived, reusable code relevant to your research, and provide a mechanism for citing its use.



RESEARCH

Research reports provide concise summaries of key developments in a field that the community do not otherwise have access to.

Journal of Open Archaeology Data Journal of Open Psychology Data Open Health Data Journal of Open Research Software

Copyright Ubiquity Press |  All content is licensed under a Creative Commons License

<http://www.metajnl.com/>

Guidance and tools



INICIO CREAR UN PGD PGDOnline DOCUMENTOS FAQs ASESORAMIENTO

PA GO DA - Plan de Gestión de Datos

Crear su Plan de Gestión de Datos

El Plan de Gestión de Datos lo solicita un agente financiador como parte de las condiciones del contrato de subvención para un proyecto científico.

El Programa Horizonte 2020 requiere que los proyectos que formen parte del Piloto 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 [H2020 Model Grant Agreement: Multi-beneficiary General MGA: December 2013](#) se establecen las obligaciones de los participantes en el Piloto de Datos de Investigación en Abierto en lo que respecta a la gestión de los datos.

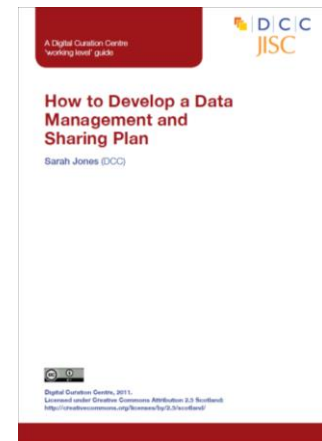
El documento [Directrices sobre la Gestión de los Datos en Horizonte 2020](#) se dirige a los solicitantes y beneficiarios de los proyectos en el Marco del Piloto de Datos de Investigación en Abierto y su objetivo es proporcionar indicaciones sobre cómo pueden cumplir con sus responsabilidades con respecto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

Para saber más sobre el Piloto de Datos de Investigación en Abierto, consulte nuestras [Faq's](#)
Para saber más sobre el PGD como parte del Piloto 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.consorciomadrono.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