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

RESEARCH COUNCILS UK

THE ROYAL SOCIETY

"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

RESEARCH COUNCILS UK

EU Policies

The EU Framework Programme
for Research and Innovation

HORIZON 2020

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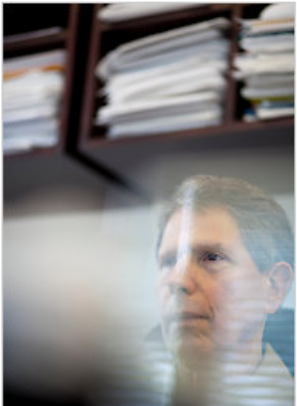
Now, the effort is bearing fruit with a wealth of recent scientific papers on the early diagnosis of Alzheimer's using methods like PET scans and tests of spinal fluid. More than 100 studies are under way to test drugs that might slow or stop the disease.

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

TRANCE
A DANNY BOYLE FILM

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

"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

... 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

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

ads ARCHAEOLOGY DATA SERVICE

esds Economic and Social Data Service

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Structured databases

ChemSpider
The free chemical database

RCS PDB
PROTEIN DATA BANK

Community initiatives

zenodo
https://zenodo.org

figshare
http://figshare.com

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
http://www.re3data.org

Data Catalogues

Institutional

National

Funders

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
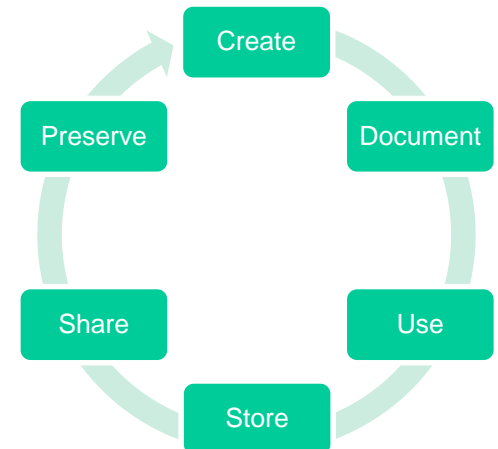
Why bother?

- 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

DMP ONLINE

DCC Checklist for a Data Management Plan | D|C|C

A Digital Curation Centre 'working level' guide

| D|C|C JISC

How to Develop a Data Management and Sharing Plan

Sarah Jones (DCC)

Digital Curation Centre, 2011.
Licensed under Creative Commons Attribution 2.5 Scotland:
http://creativecommons.org/licenses/by/2.5/scotland/
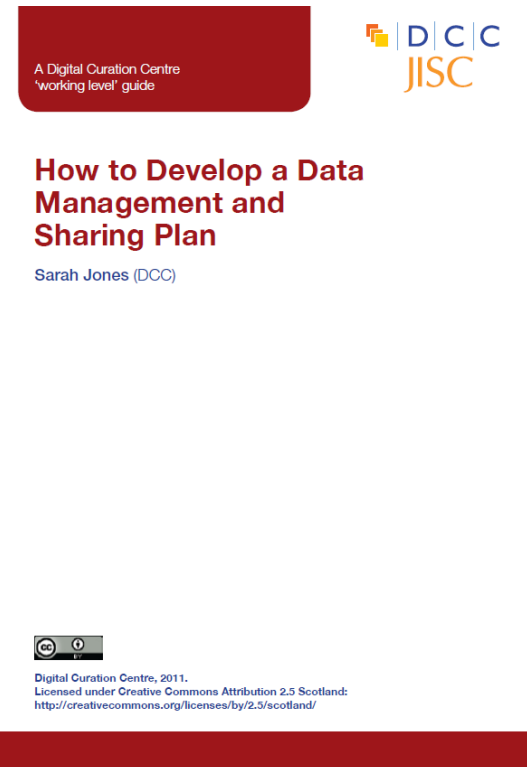
www.dcc.ac.uk/resources/data-management-plans

Specific RDM roles for support staff

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/

Literature and data review, data management planning

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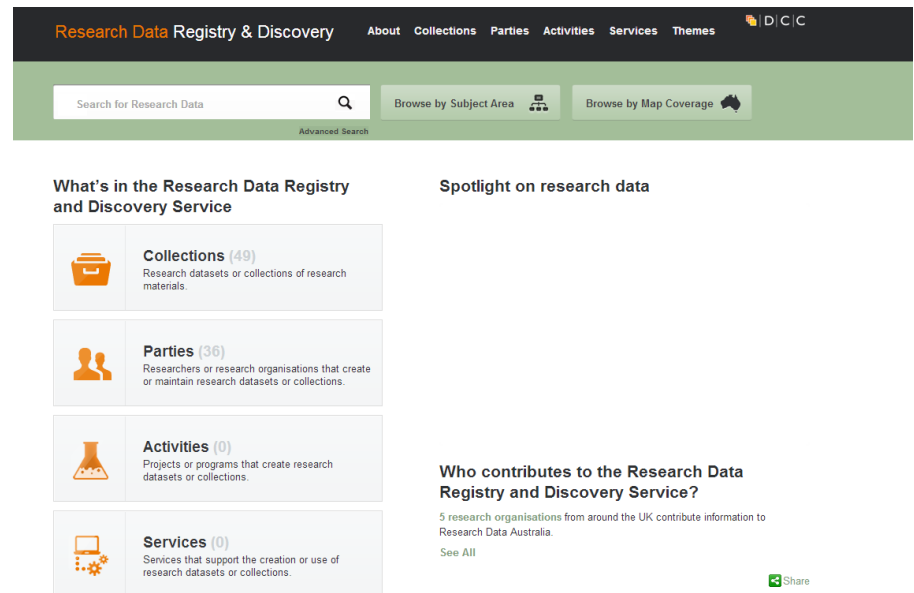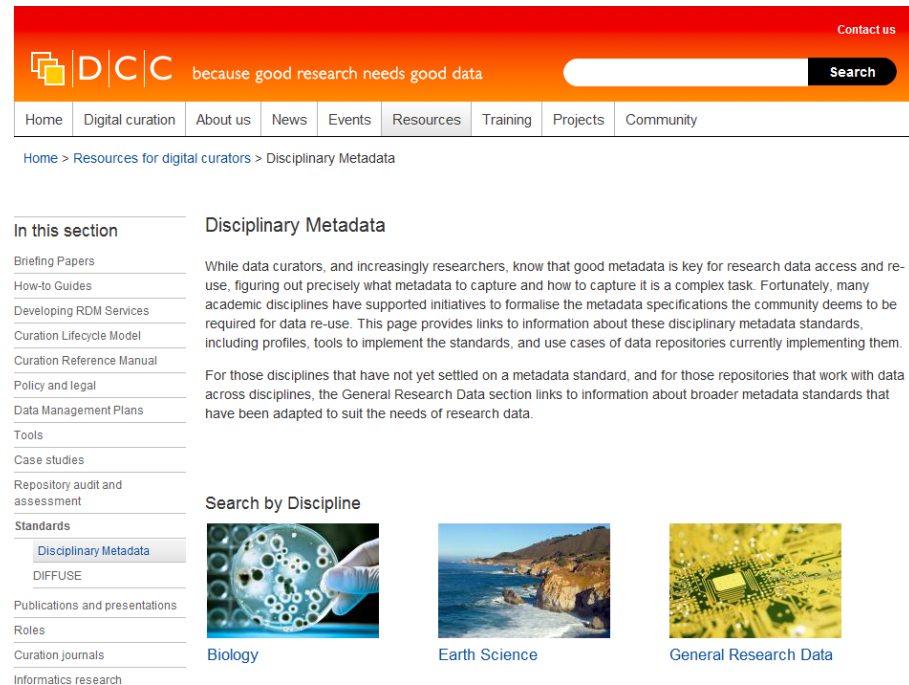
#2 methodology
Development or design of methodology; creation of models.

Standards, formats, consent

Informed consent

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.

http://creativecommons.org/choose/

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

#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.
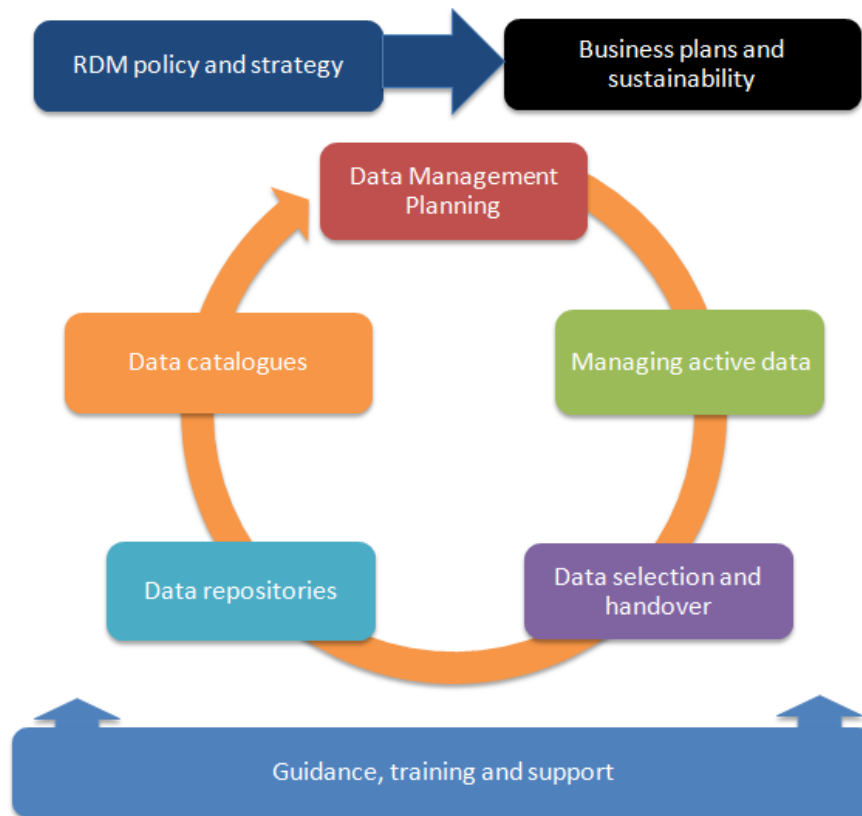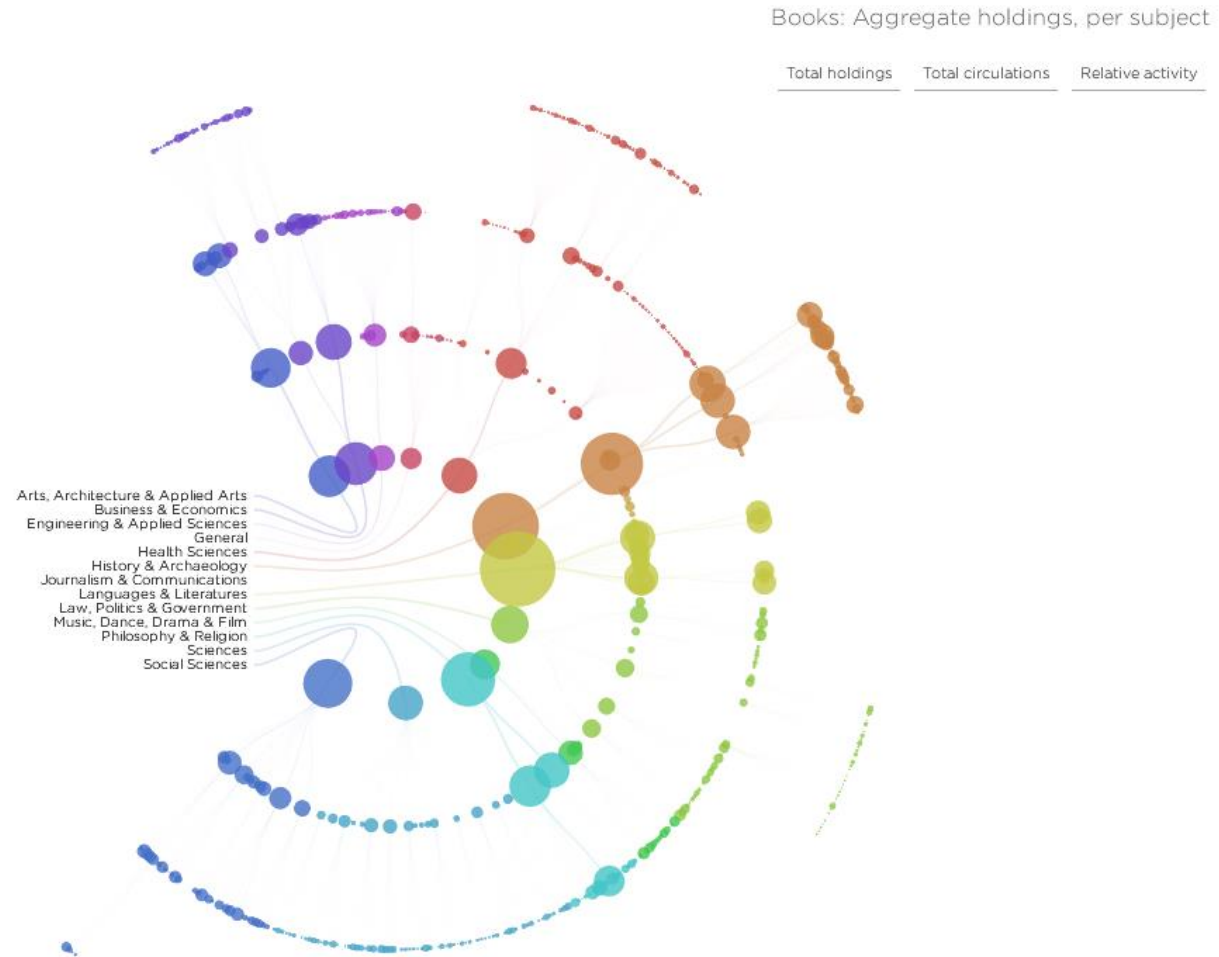

Determine curation requirements

#11 visualization
Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.

Data models and visualisations

Books: Aggregate holdings, per subject

Total holdings   Total circulations   Relative activity

Arts, Architecture & Applied Arts
Business & Economics
Engineering & Applied Sciences
General
Health Sciences
History & Archaeology
Journalism & Communications
Languages & Literatures
Law, Politics & Government
Music, Dance, Drama & Film
Philosophy & Religion
Sciences
Social Sciences

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

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

https://www.researchfish.com/

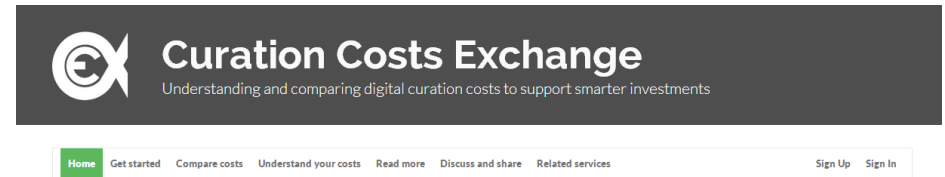**UCT POLICY for RESPONSIBLE CONDUCT OF RESEARCH**

[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

Curation Costs Exchange
Understanding and comparing digital curation costs to support smarter investments

Home   Get started   Compare costs   Understand your costs   Read more   Discuss and share   Related services   Sign Up   Sign In

All about the costs of curation
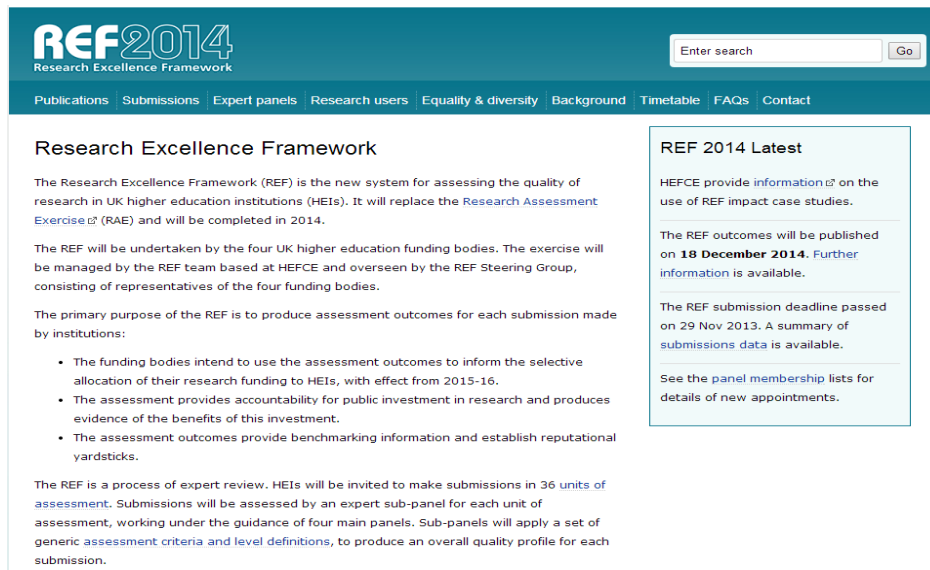Understand what you and others are and should be spending in digital curation

Compare costs
ation 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
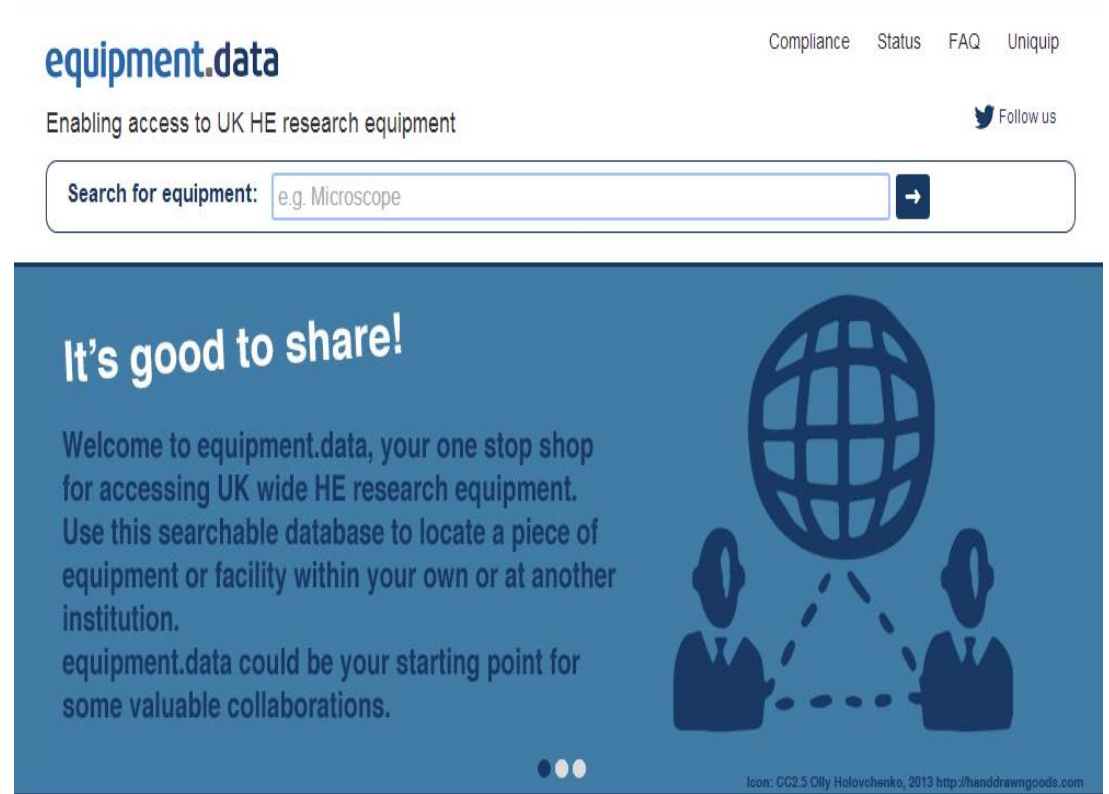help you invest

REF2014
Research Excellence Framework

Enter search   [Go]

Publications   Submissions   Expert panels   Research users   Equality & diversity   Background   Timetable   FAQs   Contact

Research Excellence Framework

The Research Excellence Framework (REF) is the new system for assessing the quality of research in UK higher education institutions (HEIs). It will replace the Research Assessment Exercise ⬀ (RAE) and will be completed in 2014.

The REF will be undertaken by the four UK higher education funding bodies. The exercise will be managed by the REF team based at HEFCE and overseen by the REF Steering Group, consisting of representatives of the four funding bodies.

The primary purpose of the REF is to produce assessment outcomes for each submission made by institutions:

- The funding bodies intend to use the assessment outcomes to inform the selective allocation of their research funding to HEIs, with effect from 2015-16.
- The assessment provides accountability for public investment in research and produces evidence of the benefits of this investment.
- The assessment outcomes provide benchmarking information and establish reputational yardsticks.

The REF is a process of expert review. HEIs will be invited to make submissions in 36 units of assessment. Submissions will be assessed by an expert sub-panel for each unit of assessment, working under the guidance of four main panels. Sub-panels will apply a set of generic assessment criteria and level definitions, to produce an overall quality profile for each submission.

REF 2014 Latest

HEFCE provide information ⬀ on the use of REF impact case studies.

The REF outcomes will be published on **18 December 2014**. Further information is available.

The REF submission deadline passed on 29 Nov 2013. A summary of submissions data is available.

See the panel membership lists for details of new appointments.

ttp://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

Publishing

http://www.metajnl.com/

Helping you to find, access, and reuse data

DataCite

https://www.datacite.org/

Guidance and tools

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