SEMINAR:
Preparing research data for open access

December 10th 2014, Social Science Data Archives, Faculty of Social Sciences, University of Ljubljana
INTRODUCTION TO RDM FROM THE INTERNATIONAL PERSPECTIVE

Angus Whyte, Digital Curation Centre
Overview

1. What is Digital Curation Centre?
   • Quick introduction
2. Why open science?
3. Where is the infrastructure?
4. What can you do?
Overview

1. What is Digital Curation Centre?
2. What is Research Data Management?
   • What is the problem, how do we define its scope?
   • Where is policy coming from?
   • Must all the data be open?
3. Where is the infrastructure?
4. What can you do?
Overview

1. What is Digital Curation Centre?
2. Why open science?
3. Where is the infrastructure?
   - International level
   - National level e.g. UK
   - Institutional level in the UK
4. What can you do?
Overview

1. What is Digital Curation Centre?
2. Why open science?
3. Where is the infrastructure?
4. What can you do?
   - Plan data management throughout the research lifecycle
   - Deal with personal data properly
   - Select what to keep & where to deposit
Established 2004
UK wide exchange good practice
Share good practice
Original focus on digital preservation
Since 2009 increasing focus on Research Data Management

“Helping to build capacity, capability and skills in data management and curation across the UK’s higher education research community”
How to Develop Research Data Management Services - a guide for HEIs

Sarah Jones, Graham Pryor and Angus Whyte

Guidance and support

Guidance webpages

Researchers are often unaware of the support available within their institution, so raising awareness is useful. Many institutions have RDM guidance webpages. We can help you to create something.

Guidance documents

The DCC provides a number of guidance documents. See in short, practical How to guides. We can provide you with hard copy and licence our content as CC-BY to encourage reuse.

We can also develop guidance documents to meet your purposes. University asked us to produce a FAQ on Data Management. Queen Mary, University of London asked for a Quick Guide on DMPs.

RDM support

We can help you to develop support more generally, such as contact with other institutions undertaking similar work to help others.


Digital Curation Centre, March 2013
This work is licensed under Creative Commons Attribution BY 2.5 Scotland
Supported by Jisc

Shared service provider to UK higher education

Catalogue of services

• Digital content
• Network and IT services
• Advice
  - legal aspects of ICT
  - disability and accessibility
  - research data curation and digital preservation
  - innovative use of digital media
• Research & Development
So what is the problem?

1. Researchers do what is required to manage data to pursue the immediate need

2. Ad-hoc solutions, unsupported, un-rewarded for managing research data

3. Digital research data disappears from the research record unless actively managed

4. Research cannot be scrutinised or reproduced

5. Funders’ investment is lost along with the data
So what is the problem?

1. Researchers do what is required to manage data to pursue the immediate need

2. Ad-hoc solutions, unsupported, un-rewarded for managing research data

3. Digital research data disappears from the research record unless actively managed

4. Difficulty scrutinising or reproducing research

5. Funders’ investment is lost along with the data!
Disappearing research record

MISSING DATA
As research articles age, the odds of their raw data being extant drop dramatically.

Concern about reproducibility from within the research community ...

Statistics’ Crisis of Reproducibility

By Organizing Committee of the Future of the Statistical Sciences Workshop | Published: July 21, 2014

- Studies cannot be reproduced
- Data analysis poorly carried out

Concern about scrutiny

...and from the public, to manage and share data better to make fraud easier to detect

Retraction Watch

New Dutch psychology scandal? Inquiry cites data manipulation, calls for retraction

The University of Amsterdam has called for the retraction of a 2011 paper by two psychology researchers after a school investigation concluded that the article contained bogus data, the Dutch press are reporting.

The paper, “Sense Creative! The Impact of Global and Local Vision, Hearing, Touching, Tasting and Smelling on Creative and Analytic Thought,” was written by Jens Förster and Markus Denzler and published in Social Psychological & Personality Science. It purported to find that:

“Holistic (global) versus elemental (local) perception reflects a prominent distinction in psychology; however, so far it has almost entirely been examined in the domain of vision. Current work suggests that global/local processing rather occurs across sensory modalities. As for vision, it is assumed the global

Define ‘Research Data Management’?

“An explicit process, covering the creation and stewardship of research materials to enable their use for as long as they retain value” (DCC)
What about ‘Research Data’?

“Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship”

So what research data is ‘published’?

Any combination of...

1. **Source data** collected, created, or held elsewhere that the research has used

2. **Assembled datasets** extracted or derived from (1)

3. **Referenced data** ‘supplementary material’ from which conclusions drawn (=most common*)


**Must all ‘managed’ data be ‘open’?**

No. RDM also for data that needs kept but not shared.

One definition of open science

“science carried out and communicated in a manner which allows others to contribute, collaborate and add to the research effort, with all kinds of data, results and protocols made freely available at different stages of the research process”

**RDM/ Open Science overlap = data sharing**

Case study examples

**What is open, why? Who open to?**

- Interviewed 18 researchers, 6 domains
- All claimed to be working openly to some degree
- All saw benefits in working that way
- What was made accessible and usable by others, and when?

“Degrees of openness - extremely important - different people work in different ways and have different constraints imposed upon them” (Chemistry, Senior Researcher)
You can view some of the administrative data sources that might be right for your research project.

We've grouped them into themed areas, and given some information about the sources.

This is not a comprehensive list of all possible data sources. We welcome ideas and requests for other data sources which may be valuable for research. Please get in touch if you have any suggestions.

---

### Catalogue

- **Business & Third Sector**
- **Crime & Justice**
- **Economy & Employment**
- **Education & Learning**
- **Health & Wellbeing**
- **Population**
- **Housing & Environment**

---

**888049 - Inter-Departmental Business Register, 1994/95**

**888050 - Register of Charities, 1993**

**888051 - Citizens Advice, 2008**

[http://adrn.ac.uk](http://adrn.ac.uk)
Open data policy from the ‘top down’

“Member states are invited to:

• Harmonise access and usage policies for research and education-related public e-infrastructures …

Research stakeholder organisations are invited to:

• Adopt and implement open access measures for publications and data resulting from publicly funded research”

Reinforced European Research Area Partnership for Excellence and Growth COM(2012) 392 final

http://ec.europa.eu/euraxess/pdf/research_policies/era-communication_en.pdf
Data sharing policy from the ‘top down’

“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 2011
http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

“**To the greatest extent and with the fewest constraints possible** publicly funded scientific research data should be **open**…respecting concerns in relation to privacy, safety, security and commercial interests [and] legitimate concerns of private partners.”

G8 Science Ministers Statement- June 2013
But what makes data a ‘public good’?

<table>
<thead>
<tr>
<th></th>
<th>Subtractability / Rivalry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exclusion</strong></td>
<td>Low</td>
</tr>
<tr>
<td><strong>Difficult</strong></td>
<td>Public Goods</td>
</tr>
<tr>
<td></td>
<td>General knowledge</td>
</tr>
<tr>
<td></td>
<td>Public domain data</td>
</tr>
<tr>
<td><strong>Easy</strong></td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Common-pool resources</td>
</tr>
<tr>
<td></td>
<td>Libraries</td>
</tr>
<tr>
<td></td>
<td>Data archives</td>
</tr>
<tr>
<td></td>
<td>Toll or Club Goods</td>
</tr>
<tr>
<td></td>
<td>Subscription journals</td>
</tr>
<tr>
<td></td>
<td>Subscription data</td>
</tr>
<tr>
<td></td>
<td>Private Goods</td>
</tr>
<tr>
<td></td>
<td>Printed books</td>
</tr>
<tr>
<td></td>
<td>Raw or competitive data</td>
</tr>
</tbody>
</table>
But what makes data a ‘public good’?

<table>
<thead>
<tr>
<th>Exclusion</th>
<th>Easy</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtractability / Rivalry</td>
<td>NO FIREWALL, PAYWALL OR LICENSE RESTRICTION</td>
<td>High</td>
</tr>
<tr>
<td>Public Goods</td>
<td>Toll or Club Goods</td>
<td>Private Goods</td>
</tr>
<tr>
<td>General knowledge</td>
<td>Subscription journals</td>
<td>Printed books</td>
</tr>
<tr>
<td>Public domain data</td>
<td>Subscription data</td>
<td>Raw or competitive data</td>
</tr>
<tr>
<td>Common-pool resources</td>
<td>Data archives</td>
<td></td>
</tr>
<tr>
<td>Libraries</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Christine Borgman, RDA Plenary 4 Keynote "Data, Data, Everywhere, Nor Any Drop to Drink." 23.9.2014
https://rd-alliance.org/plenary-2-session-chaired-kees-de-laat-university-amsterdam.html
But what makes data a ‘public good’?

<table>
<thead>
<tr>
<th>Subtractability / Rivalry</th>
<th>Infrasture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Common-pool resources</td>
</tr>
<tr>
<td>Difficult</td>
<td>Libraries</td>
</tr>
<tr>
<td>Public Goods</td>
<td>Data archives</td>
</tr>
<tr>
<td>General knowledge</td>
<td></td>
</tr>
<tr>
<td>Public domain data</td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>Private Goods</td>
</tr>
<tr>
<td>Toll or Club Goods</td>
<td>Printed books</td>
</tr>
<tr>
<td>Subscription journals</td>
<td>Raw or</td>
</tr>
<tr>
<td>Subscription data</td>
<td>competitive data</td>
</tr>
</tbody>
</table>

Christine Borgman, RDA Plenary 4 Keynote "Data, Data, Everywhere, Nor Any Drop to Drink." 23.9.2014
https://rd-alliance.org/plenary-2-session-chaired-cest-de-laat-university-amsterdam.html
Collaborative Data Infrastructure

How does that support RDM lifecycle?

Plan
Create
Use
Appraise
Deposit and Publish
Discover and Reuse
International infrastructure

Policy and research

Funding & rewards

Data Mgmt Planning

Registries & catalogue

Research infrastructures

Data repositories

Community support services

Policy alliances & common data services
International infrastructure

Policy and research → Funding & rewards

Data Mgmt Planning

Registries & catalogue → Research infrastructures

Data repositories → Community support services

Policy alliances & common data services
International infrastructure
### Slovenia

<table>
<thead>
<tr>
<th>RI name &amp; associated project</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>JSI TRIGA REACTOR</strong></td>
</tr>
<tr>
<td>AIDA Advanced European Infrastructures for Detectors at Accelerators</td>
</tr>
<tr>
<td><strong>SLONMR</strong></td>
</tr>
<tr>
<td>Bio-NMR NMR for Structural Biology</td>
</tr>
<tr>
<td><strong>SLONMR</strong></td>
</tr>
<tr>
<td>EAST-NMR Enhancing Access and Services To East European users towards an efficient and</td>
</tr>
<tr>
<td>coordinated panEuropean pool of NMR capacities to enable global collaborative research &amp;</td>
</tr>
<tr>
<td>boost technological advancements</td>
</tr>
<tr>
<td><strong>P3-IMI-MF</strong></td>
</tr>
<tr>
<td>EVA European Virus Archive</td>
</tr>
<tr>
<td><strong>NODC-SI</strong></td>
</tr>
<tr>
<td>SeaDataNet II SeaDataNet II: Pan-European infrastructure for ocean and marine data</td>
</tr>
<tr>
<td>management</td>
</tr>
<tr>
<td><strong>TANJS</strong></td>
</tr>
<tr>
<td>SPIRIT Support of Public and Industrial Research using Ion Beam Technology</td>
</tr>
</tbody>
</table>

International infrastructure

Policy and research

Funding & rewards

Data Mgmt Planning

Registries & catalogue

Research infrastructures

Data repositories

Community support services

Policy alliances & common data services
# Expert Seminars

CESSDA Expert Seminars are held annually. The seminars concentrate on practical information and on various aspects of data archiving.

<table>
<thead>
<tr>
<th>Year</th>
<th>Host</th>
<th>Location</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>UK Data Archive</td>
<td>Colchester, UK</td>
<td>Research integrity - a digital curation problem? [website]</td>
</tr>
<tr>
<td>2013</td>
<td>WISDOM</td>
<td>Vienna, Austria</td>
<td>Research Data Management: Cost and Incentives</td>
</tr>
<tr>
<td>2012</td>
<td>FSD</td>
<td>Tampere, Finland</td>
<td>Data Acquisition and Licence Agreements [website]</td>
</tr>
<tr>
<td>2011</td>
<td>FORS</td>
<td>Lausanne, Switzerland</td>
<td>Question data banks - towards greater integration? [website]</td>
</tr>
<tr>
<td>2010</td>
<td>SND</td>
<td>Gothenburg, Sweden</td>
<td>Cooperation between CESSDA archives and NSIs [website]</td>
</tr>
<tr>
<td>2009</td>
<td>ADP</td>
<td>Ljubljana, Slovenia</td>
<td>Towards the CESSDA-ERIC metadata model and DDI3 [website]</td>
</tr>
<tr>
<td>2008</td>
<td>DDA</td>
<td>Odense, Denmark</td>
<td>Cessda, DDI, Tools and Methods [website]</td>
</tr>
<tr>
<td>2007</td>
<td>UKDA</td>
<td>Colchester, UK</td>
<td>Developing the CESSDA Portal Further [programme]</td>
</tr>
<tr>
<td>2006</td>
<td>GSDB</td>
<td>Athens, Greece</td>
<td>Open Access to Data: Anonymisation, Data Protection &amp; Confidentiality [website]</td>
</tr>
</tbody>
</table>

---

**Data without Boundaries Training course on EU Census Microdata - IECM**

Data without Boundaries, in cooperation with Eurostat, organises a course on the Integrated European Union Census Microdata (IECM) in Barcelona 21 – 23 January 2015, hosted by Centre d’Estudis Demogràfics, Universitat Autònoma de Barcelona. [More..]

**CESSDA - Introducing the new team**

In June 2013 CESSDA was established as a permanent legal entity owned and financed by the individual member states’ ministry of research or a delegated institution. Norway is hosting CESSDA, and the main office is located in Bergen. But who are the faces behind this new infrastructure? A small team of three staff started work and moved into their office, CESSDA House in Bergen, in September 2014. [More..]

**CESSDA Expert Seminar, UK Data Archive, 13-14 November 2014**
DDI Data Description protection software

DDI Data Description Statistics Protection Software je orodje, ki omogoča enostavno anonimizacijo univariatnih statistik, ki so vključene po DDI (Data Documentation Initiative) standardu. Orodje je bilo razvito, da bi rešilo problem zaščite zaupnosti pri distribuciji agregiranih podatkov posameznih spremenljivk. Ti so javno dostopni, da bi zagotovili koristne informacije znanstvenim raziskovalcem in s tem promovirali rabo mikropodatkov uradne statistike v znanstvene namene.

Funkcije orodja za avtomatsko zaščito statistik, vključenih v opis podatkov, temeljijo na tako metodah in tehnikah anonimizacije mikropodatkov, kot tudi na tistih uporabljenih pri zaščiti tabelarnih podatkov. Sledijo pa tudi posebnostim zaupnosti podatkov uradne statistike. Tehnike zaščite so tematsko razdeljene glede na tip anonimizacijske tehnike, ki se uporablja za zaščito agregiranih statistik.

Orodje omogoča hitro in enostavno avtomatsko zaščito podatkov z uporabo metod za zaščito podatkov kot so rekodiranje, zaščita najnižjih in najvišjih vrednosti, izbris (numeričnih informacij) spremenljivk, zaščita deskriptivnih statistik številskih spremenljivk in zaščita frekvenc vrednosti spremenljivk, pri čemer se upošteva pravilo najnižje dovoljene prikazane frekvence. V nasprotju z obstoječimi orodji za zaščito mikropodatkov in agregiranih podatkov to razvito orodje zaščiti agregirane podatke direktno v XML kodi.

>> Prenesi orodje

ADP is founded by Ministry of Higher Education, Science and Technology in scope of infrastructure program "The Research Infrastructural Centres Network of University of Ljubljana." Social Science Data Archives is a member of CESSDA, the organisation for social science data archives across Europe.
CESSDA uses and supports the DDI international metadata standard.

Enabling better access and reuse, e.g. find which longitudinal studies have asked which questions.
International infrastructure

Policy and research → Funding & rewards

Data Mgmt Planning

Registries & catalogue

Research infrastructures

Data repositories

Community support services

Policy alliances & common data services

(DataCite logo)
International infrastructure

Policy and research → Funding & rewards

Data Mgmt Planning

Registries & catalogue → Research infrastructures

Data repositories → Community support services

Policy alliances & common data services
CESSDA
New pan-European Research Infrastructure

Access across national social science data repositories

http://www.cessda.net/
Access across national social science data repositories

Data repositories

Registries & catalogue

http://www.cessda.net/

ADP are the central Social Science Data Archives in Slovenia.
International infrastructure

Policy and research → Funding & rewards

Data Mgmt Planning

Registries & catalogue → Research infrastructures

Data repositories → Community support services

Policy alliances & common data services
National infrastructure

Policy and guidance → Funding & rewards

Data Mgmt Planning

Portal / catalogue → Community tools & resources

Data repositories → Outreach services

Policy monitoring & evaluation
Jisc Managing Research Data Programme

- DCC Institutional Engagement in parallel - support to further 21 Institutions to set up services 2011-13
- Ongoing tailored support to 38 since 2013

Institutional RDM Infrastructure Services
17 Projects

RDM Training
5 projects

Data Publication
3 projects

RDM Planning
10 projects

Institutional infrastructure

Policy and guidance → Business planning

Data Mgmt Planning

CRIS/ Data catalogue → Managing active data

Data repositories → Data selection & handover

Practical guidance, training & support
Institutional infrastructure

Policy and guidance → Business planning

Data Mgmt Planning

CRIS/ Data catalogue → Managing active data

Data repositories

Data selection & handover

Practical guidance, training & support
1979 national school leavers survey

Title: 1979 national school leavers survey

Data Creator: Ruus, Laine; Scottish Education Department. Scottish Office; Social Science Research Council; Manpower Services Commission

Date Available: 2014-07-22


Dataset Description (abstract):
The 1979 survey covered young people who had left school in the session 1977/78. It was based on a smaller sample than the earlier 1977 survey: 10 per cent of O and H grade leavers, and 20 per cent of non-qualified * leavers, compared with a sample of 40 per cent of all leavers in 1977. The questionnaires were based on those used in 1977, and repeated most of the earlier items. Several new questions were added and others were revised, and the non-qualified leavers' questionnaire was substantially re-designed to include extra items on job-seeking, early employment experience and the Youth Opportunities Programme. In this way the 1979 survey has extended and updated the time series in the Scottish Education Data Archive. It also had four more specific objectives: (i) to obtain information on additional topics (covered by the new questions); (ii) to obtain information on non-Certificate leavers in Regions where they were not surveyed in 1977; (iii) to obtain information on changes between 1977 and 1979 with respect to the items common to both surveys; (iv) to provide contextual information for analyses which require large sample numbers and which will therefore need to be conducted on the 1977 data.

Files

<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
<th>Format</th>
<th>View</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative_Research_Questionnaire_1979.pdf</td>
<td>1.698Mb</td>
<td>PDF</td>
<td>Download</td>
<td>Questionnaires</td>
</tr>
</tbody>
</table>

http://datashare.is.ed.ac.uk/handle/10283/571
Institutional infrastructure

Policy and guidance → Business planning

Data Mgmt Planning

CRIS/ Data catalogue → Managing active data

Data repositories → Data selection & handover

Practical guidance, training & support
Practical checklists
key points in research cycle

Repository selection
1. Policy & legal
2. Discoverable
3. Preservation
4. Reports
5. Trust

Catalogue Metadata
1. Name
2. Description
3. Identifier
4. Subject
5. URL
6. Date
7. Creator
8. Rights
9. Spatial
10. Publisher

Data Mgmt Plan
1. Collection
2. Documentation
3. Ethics & legal
4. Storage & backup
5. Selection & preserve
6. Data sharing
7. Responsibilities

Data Selection
5 Steps to decide what to keep
1. Could - benefit
2. Must - risks
3. Should - value
4. Cost factors
5. Weigh-up 1-4

Start
Writing-up
Archive
Active storage
Data selection checklist

Five steps to deciding what data to keep

DCC Checklist for Appraising Research Data

Version 1

Please cite as: DCC (2014) ’Five steps to deciding what data to keep: a checklist for appraising research data v.1 Edinburgh: Digital Curation Centre. Available online: www.dcc.ac.uk/resources/how-guides

Digital Curation Centre, October 2014

This work is licensed under Creative Commons Attribution BY 2.5 Scotland, except Section 4, which is adapted under licence CC-BY-NC-SA. from UK Data Archive (2013) Data management costing tool and checklist. Available at www.data-archive.ac.uk/create-manage/planning-for-sharing/costing

http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep
Straightforward steps

1. **Could** this data be re-used
2. **Must** it be kept to manage compliance risk
3. **Should** it be kept for its potential value and...
4. Consider costs
5. Will ✔ or won’t ✗ it be kept, shared on what terms

---

**Data Selection**
5 Steps to decide what to keep
1. Could - benefit
2. Must - risks
3. Should - value
4. Cost factors
5. Weigh-up 1-4

**Repository selection**
1. Policy & legal
2. Discoverable
3. Preservation
4. Reports
5. Trust

---

Institution or external repository
Step 1 (?) What ‘must’ be kept?

Research record includes data as evidence for e.g. ...

- Audit purposes
- Health & Safety (Lab book)
- Contractual requirement

Compliance also about data that won’t be kept, or only shared with approved researchers…

Research Ethics, Duty of Confidentiality, Data Protection Act, Human Rights Act, Statistics & Registration Services Act. UK Data Archive:

http://www.data-archive.ac.uk/create-manage/consent-ethics/legal
Step 1 (?) What ‘must’ be kept?

Data may be part of research records for compliance

- Audit
- Health & Safety (e.g. Lab book)
- Contractual obligations

Compliance also about data that won’t be kept, or only shared with approved researchers...

Even where there are legal requirements to keep, or dispose of data, investigator makes initial selection of data that fulfils the research purpose
Step 1 (?) What ‘must’ be kept?

Funder & journal data policies expect some value judgement

“Data with acknowledged long-term value” Research Councils UK Common Principles on Data Policy

“Data, information and other electronic resources of long-term interest” Economic Social Research Council UK Data Archive Collections Development Policy

“An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Nature
Step 1 (?) What ‘must’ be kept?

Funder & journal data policies expect value judgement

“Data with acknowledged long-term value” Research Councils UK Common Principles on Data Policy

“Data, information and other electronic resources of long-term interest” Economic Social Research Council UK Data Archive Collections Development Policy

“An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Nature

So the first step should be establishing
1. what are my community’s expectations on verification/ replication (journals, repositories, societies)
2. what other purposes could the data be reused for?
Step 2-1 What could it be reused for?

Any purposes (potential benefits) not already considered?

1. Verification (community expectations)
2. Further analysis (data linking, collaboration)
3. Visibility (impact, citation, credit)
4. Resource development (funding)
5. Further publications /data articles (citation)
6. Learning and teaching materials (credit)
7. Private reference (exploitation)
Step 2.1 What *could* it be reused for?

Any purposes (potential benefits) not already considered?

1. Verification (*community expectations*)
2. Further analysis (*data linking, collaboration*)
3. Visibility (*reputation, citation, credit*)
4. Resource development (*funding*)
5. Further publications /data articles (*citation*)
6. Learning and teaching materials (*credit*)

From what is **desirable**, what is likely to be **feasible**? What data, documentation, software, other resource would you need to keep to make it happen later?
Step 3 What data *should* have value

Does it meet any two of these criteria?
1. **Good quality** - data and description complete, accurate, reliable, valid, representative etc
2. **High demand** known users, integration potential, reputation, recommendation, appeal
3. **High effort to (re)produce** difficult, costly, or impossible to reproduce
4. **Low barriers to reuse** legal/ ethical, copyright non-restrictive terms and conditions
5. **Rarity value** unique copy or other copies at risk
Step 4 Consider cost factors

Costs already met may **add** to data value

Question is, can you afford to do the minimum to ensure that value is not lost when research ends?

1. Creation, collection & cleaning
2. Short-term storage & backup
3. Short-term access & security
4. Team communication & development
5. Preservation & long-term access
Step 5 Bring it all together

Balance risks, costs and value

Document the choices made

1. Dataset name, contributors, description, sensitivity - metadata
2. Reuse purposes and value - the ‘reuse case’
3. Risk of non-compliance and costs shortfall
4. Justification to keep or dispose
5. Actions to prepare for preservation or disposal
Thank you, any questions?

Any further thoughts on preparation for sharing openly?