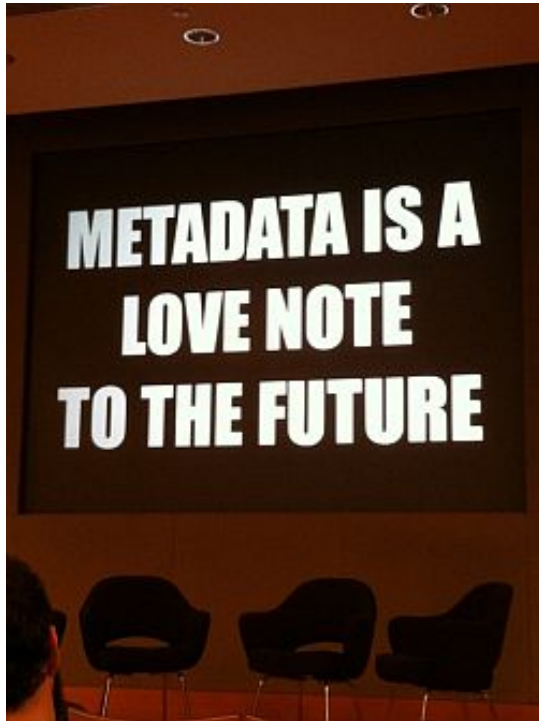


# Research data management: definitions, drivers and resources

1



Martin Donnelly  
Digital Curation Centre  
University of Edinburgh

# Overview

- I. The Digital Curation Centre
- II. Definitions
  - i. What is research data management?
  - ii. Data's role in the research process
- III. Drivers: why is RDM necessary / desirable?
- IV. Barriers, and possible roles for libraries
- V. RDM resources

# I. The Digital Curation Centre

- ì The  (est. 2004) is...
  - ì A UK centre of expertise in digital preservation, with a particular focus on research data management (RDM)
  - ì Based across three sites: Universities of Edinburgh, Glasgow and Bath
  - ì Working with a number of UK universities to identify gaps in RDM provision and raise capabilities across the sector
  - ì Also involved in a variety of international collaborations...

# DCC networks and partnerships

 <p>4C Collaboration to Clarify the Costs of Curation</p>	 <p>CODATA S U</p>	 <p>DataCite</p>	 <p>3TU University of California CDL California Digital Library</p>
 <p>SPARC<sup>®</sup></p>	 <p>The Association of Commonwealth Universities</p>	 <p>ands AUSTRALIAN NATIONAL DATA SERVICE</p>	 <p>CDL California Digital Library</p>
 <p>DataONE Data Observation Network for Earth</p>	 <p>OpenAIRE Open Access Infrastructure for Research in Europe</p>	 <p>SIM4RDM RESEARCH INFORMATION MANAGEMENT</p>	
 <p>Digital Preservation Coalition</p>	 <p>RDA RESEARCH DATA ALLIANCE</p>	 <p>KE Knowledge Exchange</p>	
 <p>GRDI 2020 A Vision for Global Research Data Infrastructures</p>	 <p>UKSG</p>	 <p>UNIVERSITY OF READING</p>	 <p>COAR Confederation of Open Access Repositories</p>
 <p>International Records Management Trust</p>	 <p>nestor</p>	 <p>DANS Data Archiving and Networked Services</p>	
 <p>SND Swedish National Data Service</p>	 <p>APA ALLIANCE FOR PERMANENT ACCESS</p>	 <p>LIGUE DES BIBLIOTHEQUES EUROPEENNES DE RECHERCHE ASSOCIATION OF EUROPEAN RESEARCH LIBRARIES</p>	
 <p>OR2012 The 7th International Conference on Open Repositories</p>	 <p>DEff Denmark's Electronic Research Library</p>	 <p>CSC</p>	
 <p>RLUK Research Libraries UK</p>	 <p>cni Coalition for Networked Information</p>	 <p>rugit Russell Universities Group of IT Directors (RUGIT)</p>	

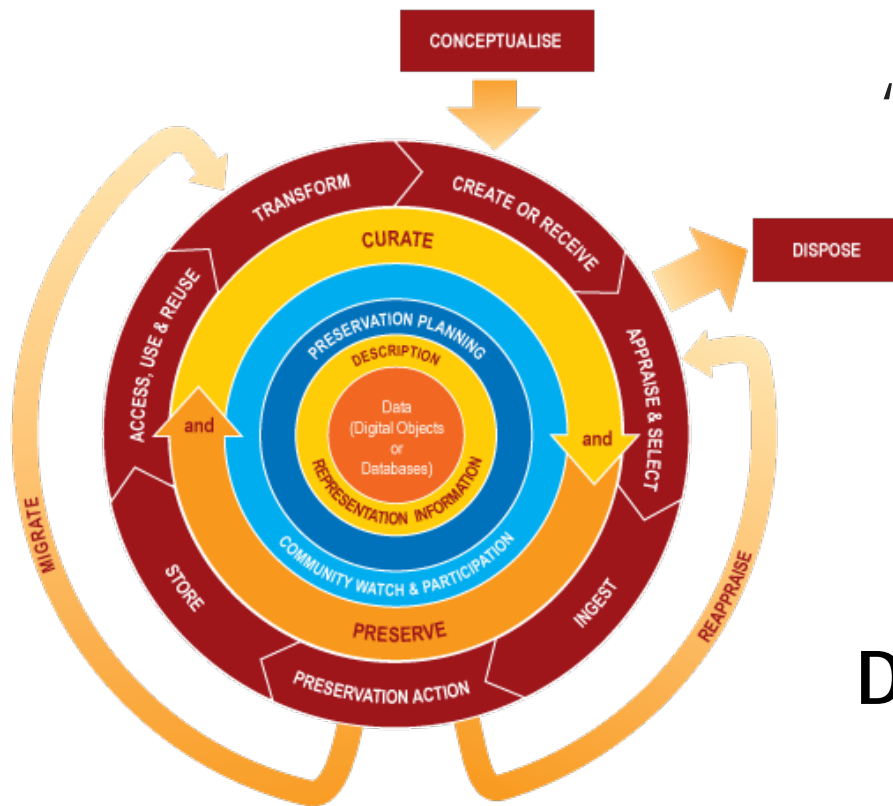


## II. DEFINITIONS





# What is RD(M)?



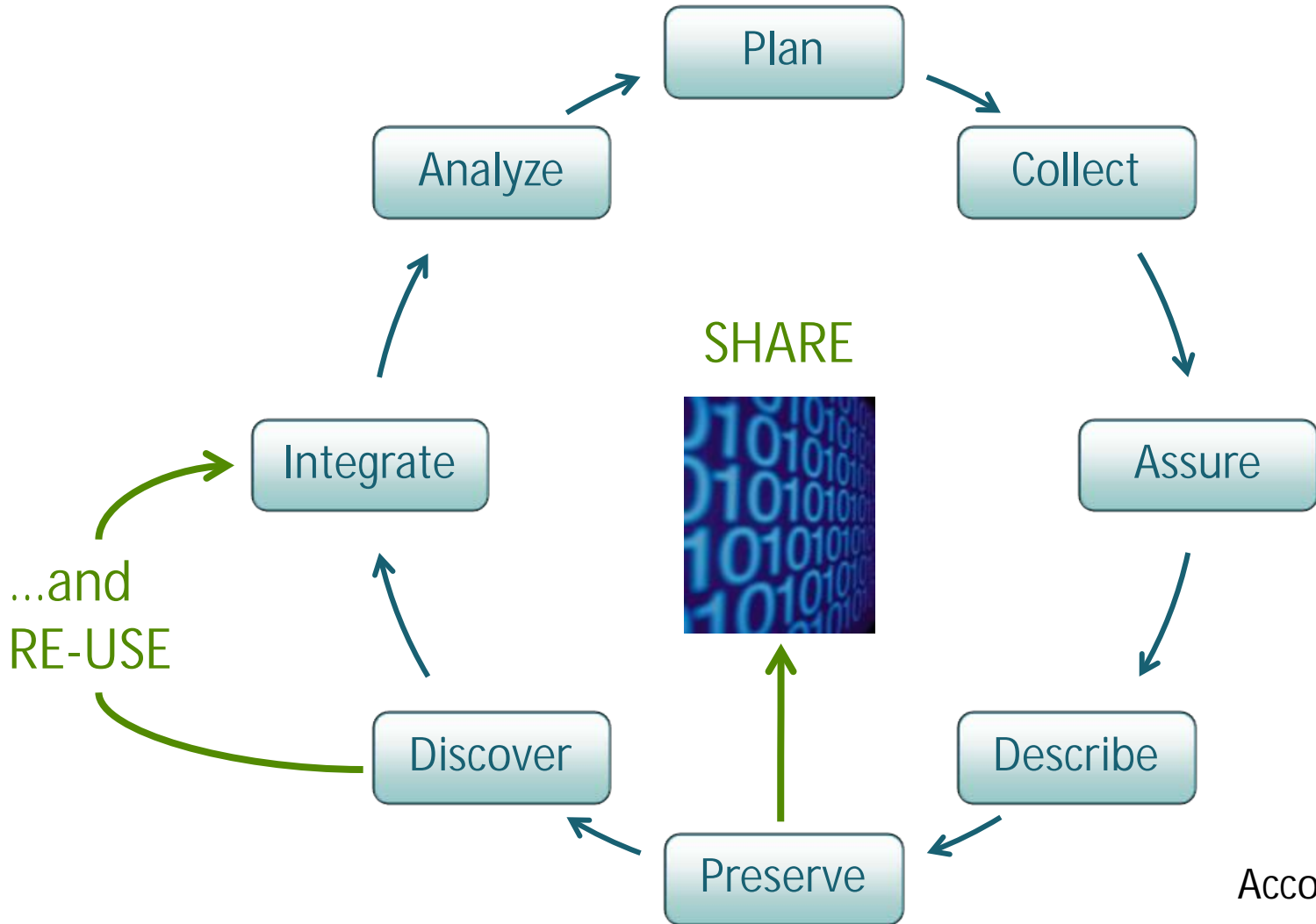
“the active management and appraisal of data over the lifecycle of scholarly and scientific interest”

**Data management is a part of good research practice.**

## ii. What do we mean by research *data*?

- î It varies from discipline to discipline, and from funder to funder...
- î A science-centric definition:
  - î “The recorded **factual** material commonly accepted in the scientific community as necessary to **validate** research findings.”  
(US Office of Management and Budget, Circular 110)
- î And another from the visual arts:
  - î “**Evidence** which is used or created to generate new knowledge and interpretations. ‘Evidence’ may be intersubjective or subjective; physical or emotional; persistent or ephemeral; personal or public; explicit or tacit; and is consciously or unconsciously referenced by the researcher at some point during the course of their research.”  
(Leigh Garrett, KAPTUR project: see <http://kaptur.wordpress.com/2013/01/23/what-is-visual-arts-research-data-revisited/>)

# Research data lifecycle



According to DataONE



# III. DRIVERS

- i. Technology
- ii. Efficiency
- iii. Expectation
- iv. Verification
- v. Protection



# i. It's now possible

- i Developments in sensor **technology**, networking and digital storage enable new research and scientific paradigms
- i As costs also fall, possibilities for data sharing, citation and re-use become much more widespread
- i Journals dedicated solely to publishing data have even started to appear. That's not to say it's an entirely new thing: journals have always published data, just never before at such scale...



# Rosse

CAMBRIDGE LIBRARY COLLECTION  
 THE SCIENTIFIC PAPERS OF  
 WILLIAM PARSONS,  
 THIRD EARL OF  
 ROSSE 1800-1867  
 EDITED BY CHARLES PARSONS



from  
*Philosophical  
 Transactions of  
 the Royal Society,*  
 (MDCCCLXI) (or  
 1861 if you'd  
 prefer)

## 154 EARL OF ROSSE ON THE CONSTRUCTION OF SPECULA OF 6-FEET APERTURE,

Number in Herschel's Catalogue.	Number of times observed.	Description.																								
53	1	Sept. 19, 1857. S; R; vF; bM.																								
54	2	Nov. 22, 1854. pB; vS; R.																								
59	3	Dec. 22, 1848. 3 neb. in line, 2 of them "nova." Oct. 23, 1856. 1st is R; pB; bM; and has nucleus; 2nd bM; E, * involved; 3rd F; IE; bM.																								
60	1	Nov. 22, 1854. S; R; bM.																								
65	3	Sept. 18, 1857. S; pl. disc. in vF haze of mottled neby.																								
69 } 70 }	7	Oct. 3, 1856. 69 is S; B; R; with B. nucleus; 70 is F; E. and patchy. I sometimes thought it was formed of two knots involved in F. neby; there appears to be a nebulous connexion between them all. Nov. 15, 1857. The silvered mirror shows the object brighter than before, but no new details: definition bad.																								
71	7	Suspect spirality; light unequal.																								
72	3	Oct. 26, 1854. a F. object with two nuclei. Nov. 29, 1850. $\alpha$ is vB; $\beta$ has stellar point or nucleus. I suspect $\delta$ to be a F. neb.																								
78 } 79 }	4	<table border="1"> <thead> <tr> <th></th> <th>Pos.</th> <th>Dist.</th> </tr> </thead> <tbody> <tr> <td><math>\alpha\beta</math></td> <td>219°</td> <td>5' 35"</td> </tr> <tr> <td><math>\alpha\gamma</math></td> <td>315</td> <td>1 8</td> </tr> <tr> <td><math>\alpha\delta</math></td> <td>81</td> <td>0 44</td> </tr> </tbody> </table> Nov. 3, 1855. 3 neb. nearly in line, sp, n, f; $\beta$ is bM. and IE. p. and f; $\alpha$ is R; bM; with a d. * np. and is the largest of the 3; $\epsilon$ is S; F; R; $\delta$ is a *. Oct. 3, 1856. pL; not vF. Its brightest part is a line running diagonally, and there is a knot at either end; believed to be a spiral.		Pos.	Dist.	$\alpha\beta$	219°	5' 35"	$\alpha\gamma$	315	1 8	$\alpha\delta$	81	0 44												
	Pos.	Dist.																								
$\alpha\beta$	219°	5' 35"																								
$\alpha\gamma$	315	1 8																								
$\alpha\delta$	81	0 44																								
80	1	Oct. 3, 1856. pL; not vF. Its brightest part is a line running diagonally, and there is a knot at either end; believed to be a spiral.																								
84 } 85 } 86 }	4	<table border="1"> <thead> <tr> <th></th> <th>Pos.</th> <th>Dist.</th> </tr> </thead> <tbody> <tr> <td><math>\alpha\beta</math></td> <td>169°</td> <td>2' 19"</td> </tr> <tr> <td><math>\beta\gamma</math></td> <td>160</td> <td>4 22</td> </tr> <tr> <td><math>\gamma\delta</math></td> <td>201</td> <td>0 34</td> </tr> <tr> <td><math>\gamma\epsilon</math></td> <td>157</td> <td>3 19</td> </tr> <tr> <td><math>\gamma\zeta</math></td> <td>178</td> <td>5 32</td> </tr> <tr> <td><math>\zeta\eta</math></td> <td>199</td> <td>1 41</td> </tr> <tr> <td><math>\eta\theta</math></td> <td>79</td> <td>4 55</td> </tr> </tbody> </table> Oct. 26, 1854. A d. neb., both S; R; bM. A cl. with much unresolved neby. 1bM.		Pos.	Dist.	$\alpha\beta$	169°	2' 19"	$\beta\gamma$	160	4 22	$\gamma\delta$	201	0 34	$\gamma\epsilon$	157	3 19	$\gamma\zeta$	178	5 32	$\zeta\eta$	199	1 41	$\eta\theta$	79	4 55
	Pos.	Dist.																								
$\alpha\beta$	169°	2' 19"																								
$\beta\gamma$	160	4 22																								
$\gamma\delta$	201	0 34																								
$\gamma\epsilon$	157	3 19																								
$\gamma\zeta$	178	5 32																								
$\zeta\eta$	199	1 41																								
$\eta\theta$	79	4 55																								
87	3	Oct. 26, 1854. A d. neb., both S; R; bM.																								
89	8	A cl. with much unresolved neby.																								
90	1	1bM.																								
91 } 92 }	1	3 neb. in a triangle.																								
96	6	Oct. 26, 1854. Lenticular n. and s. Thought I saw a * at times in centre (1 1/2-inch single lens); a lp. this is another vF. ray, np, sf, and which has no nucleus. Oct. 16, 1855. vF; E. n. and s; has nucleus; * in n. end. Nov. 3, 1855. mE; pB. nucleus, and * in n. end; np. this neb. is a * of the 9th mag., and about the same distance p. this * is another neb. vF; mE. Dec. 7, 1855. Seen as before; comp. neb. verified. Oct. 23, 1856. F. ray has nucleus and a * in n. end. Sept. 18, 1857. E. n. and s; another vF. ray p, which is E. np. af.																								
98	1	vF; R; S.																								
99	1	Oct. 3, 1856. S; F; R; bM; has nucleus.																								
103	3	Is n. of the 3rd of a group of 4 *s in line; 3 "nova" near.																								
104	1	<table border="1"> <thead> <tr> <th></th> <th>Pos.</th> <th>Dist.</th> </tr> </thead> <tbody> <tr> <td><math>\alpha\beta</math></td> <td>58°</td> <td>7' 36"</td> </tr> <tr> <td><math>\beta\gamma</math></td> <td>40</td> <td>4 6</td> </tr> <tr> <td><math>\beta\delta</math></td> <td>81</td> <td>9 19</td> </tr> <tr> <td><math>\alpha\beta</math></td> <td>30</td> <td>7 43</td> </tr> </tbody> </table> Oct. 23, 1856. 6 neb., all visible at once in finder eyepiece; 2 of them E., the others S; R; bM. Dec. 11, 1854. vmE; bM (speculum dewed). A variety of new nebulae found, but observations too voluminous to transcribe. Sketch made, but no interesting details. Nov. 30, 1850. vF. and p. a quadruple *. Oct. 23, 1851. 3 *s f. neb.; light unequal. Sept. 16, 1852. 2' diameter; several *s in it; probably a F. cl.		Pos.	Dist.	$\alpha\beta$	58°	7' 36"	$\beta\gamma$	40	4 6	$\beta\delta$	81	9 19	$\alpha\beta$	30	7 43									
	Pos.	Dist.																								
$\alpha\beta$	58°	7' 36"																								
$\beta\gamma$	40	4 6																								
$\beta\delta$	81	9 19																								
$\alpha\beta$	30	7 43																								

\* This should be, I think,  $\zeta\zeta$ . A S. d. neb. suspected below at  $\alpha$ .

## AND A SELECTION FROM THE OBSERVATIONS MADE WITH THEM. 155

Number in Herschel's Catalogue.	Number of times observed.	Description.																																																																								
113 } 114 } 115 } 116 }	2	Both have nuclei; "nova" near. Nov. 16, 1857. 113 is E. p. and f; * closely sp: 114 is R, with ragged edge and bM; "nova;" S; R; bM.																																																																								
117	1	Oct. 3, 1826. The p. one is a pB. S. disc in F. outlying neby. The f. one is R; bM.																																																																								
118	1	Dec. 18, 1851. s. end of neb. is like a brush or broom with a split.																																																																								
119 } 120 }	2	4 neb. found, 2 have nuclei. 118 is S; R; 120 has 2 *s on np. edge; E. p. and f.																																																																								
121	1	Dec. 9, 1854. pL; pB; bM to a nucleus.																																																																								
122	2	Sept. 18, 1857. Rough sketch made; mE. np, sf; a F. triple * f.																																																																								
123	3	Nov. 28, 1856. L; B; mE; B. nucleus. "Nova" f.																																																																								
131	27	<table border="1"> <thead> <tr> <th></th> <th>Pos.</th> <th>Dist.</th> </tr> </thead> <tbody> <tr> <td><math>\alpha\beta</math></td> <td>215°</td> <td>0' 51"</td> </tr> <tr> <td><math>\alpha\gamma</math></td> <td>163</td> <td>0 56</td> </tr> <tr> <td><math>\alpha\delta</math></td> <td>160</td> <td>2 56</td> </tr> <tr> <td><math>\alpha\epsilon</math></td> <td>178</td> <td>3 07</td> </tr> <tr> <td><math>\alpha\zeta</math></td> <td>192</td> <td>3 44</td> </tr> <tr> <td><math>\alpha\eta</math></td> <td>206</td> <td>4 14</td> </tr> <tr> <td><math>\alpha\theta</math></td> <td>224</td> <td>4 55</td> </tr> <tr> <td><math>\alpha\iota</math></td> <td>147</td> <td>5 34</td> </tr> <tr> <td><math>\alpha\kappa</math></td> <td>179</td> <td>5 56</td> </tr> <tr> <td><math>\alpha\lambda</math></td> <td>201</td> <td>6 42</td> </tr> <tr> <td><math>\mu\nu</math></td> <td>143</td> <td>6 28</td> </tr> <tr> <td><math>\alpha\mu</math></td> <td>287</td> <td>4 30</td> </tr> <tr> <td><math>\alpha\nu</math></td> <td>341</td> <td>6 45</td> </tr> <tr> <td><math>\alpha\omega</math></td> <td>5</td> <td>5 18</td> </tr> <tr> <td><math>\alpha\phi</math></td> <td>357</td> <td>4 42</td> </tr> <tr> <td><math>\alpha\psi</math></td> <td>51</td> <td>11 0</td> </tr> <tr> <td><math>\alpha\rho</math></td> <td>38</td> <td>9 50</td> </tr> <tr> <td><math>\alpha\sigma</math></td> <td>58</td> <td>11 16</td> </tr> <tr> <td><math>\alpha\tau</math></td> <td>161</td> <td>5 29</td> </tr> <tr> <td><math>\alpha\chi</math></td> <td>140</td> <td>6 53</td> </tr> <tr> <td><math>\alpha\theta</math></td> <td>172</td> <td>6 32</td> </tr> <tr> <td><math>\alpha\eta</math></td> <td>174</td> <td>7 18</td> </tr> <tr> <td><math>\alpha\delta</math></td> <td>205</td> <td>2 22</td> </tr> </tbody> </table> Dec. 27, 1850. Jan. 2, 1851. Dec. 23, 1851.		Pos.	Dist.	$\alpha\beta$	215°	0' 51"	$\alpha\gamma$	163	0 56	$\alpha\delta$	160	2 56	$\alpha\epsilon$	178	3 07	$\alpha\zeta$	192	3 44	$\alpha\eta$	206	4 14	$\alpha\theta$	224	4 55	$\alpha\iota$	147	5 34	$\alpha\kappa$	179	5 56	$\alpha\lambda$	201	6 42	$\mu\nu$	143	6 28	$\alpha\mu$	287	4 30	$\alpha\nu$	341	6 45	$\alpha\omega$	5	5 18	$\alpha\phi$	357	4 42	$\alpha\psi$	51	11 0	$\alpha\rho$	38	9 50	$\alpha\sigma$	58	11 16	$\alpha\tau$	161	5 29	$\alpha\chi$	140	6 53	$\alpha\theta$	172	6 32	$\alpha\eta$	174	7 18	$\alpha\delta$	205	2 22
	Pos.	Dist.																																																																								
$\alpha\beta$	215°	0' 51"																																																																								
$\alpha\gamma$	163	0 56																																																																								
$\alpha\delta$	160	2 56																																																																								
$\alpha\epsilon$	178	3 07																																																																								
$\alpha\zeta$	192	3 44																																																																								
$\alpha\eta$	206	4 14																																																																								
$\alpha\theta$	224	4 55																																																																								
$\alpha\iota$	147	5 34																																																																								
$\alpha\kappa$	179	5 56																																																																								
$\alpha\lambda$	201	6 42																																																																								
$\mu\nu$	143	6 28																																																																								
$\alpha\mu$	287	4 30																																																																								
$\alpha\nu$	341	6 45																																																																								
$\alpha\omega$	5	5 18																																																																								
$\alpha\phi$	357	4 42																																																																								
$\alpha\psi$	51	11 0																																																																								
$\alpha\rho$	38	9 50																																																																								
$\alpha\sigma$	58	11 16																																																																								
$\alpha\tau$	161	5 29																																																																								
$\alpha\chi$	140	6 53																																																																								
$\alpha\theta$	172	6 32																																																																								
$\alpha\eta$	174	7 18																																																																								
$\alpha\delta$	205	2 22																																																																								
132	1	Nov. 28, 1856. B; S; R. nucleus, a * p. and another n.																																																																								
134 } 135 }	2	Nov. 29, 1850. A S. neb. or cl. with 3 *s in it. $R. 1^p 20^m. N.P.D. 60^p 35'$ .																																																																								
136	2	Oct. 26, 1854. Both S; R; B.																																																																								
142	8	Sought for four times; not found. Dec. 13, 1848. Rough sketch made. Spiral? Dec. 14, 1848. Confirmed last night's observation; feel confident it is a spiral. Oct. 24, 1851. Centre formed of *s: easily seen to be such: several *s through the neb.																																																																								
143	1	Oct. 3, 1856. vS; F; R; bM; had a * close to n. edge.																																																																								
147	2	Nov. 30, 1856. S; R; bM. to a nucleus.																																																																								



# Five years of front pages...



Nature, 09/08



ACM, 12/08



Nature, 09/09



Economist, 02/10



InformationWeek, 08/10



Science, 02/11



Popular Science,  
11/11



Computerworld,  
11/10

# ii. VfM via data re-use / re-purposing

**Ships' log books build picture of climate change** 14 October 2010

You can now help scientists understand the climate of the past and unearth new historical information by revisiting the voyages of First World War Royal Navy warships.

Visitors to OldWeather.org will be able to retrace the routes taken by any of 280 Royal Navy ships. These include historic vessels such as HMS Caroline, the last survivor of the 1916 Battle of Jutland still afloat. By transcribing information about the weather and interesting events from images of each ship's logbook, web volunteers will help scientists build a more accurate picture of how our climate has changed over the last century.

<http://www.nationalarchives.gov.uk/news/503.htm>



Detail from Royal Navy Recruitment poster, RNVR Signals branch, 1917 (Catalogue reference: ADM 1/8331)



HMS Beagle, 1830-34



Endeavour, 1768-71 (Captain Cook)



HMS Torch, 1918



# Government pressure/support

*6.9 The Research Councils expect the researchers they fund to deposit published articles or conference proceedings in an open access repository at or around the time of publication. But this practice is unevenly enforced. Therefore, as an immediate step, **we have asked the Research Councils to ensure the researchers they fund fulfil the current requirements.** Additionally, the Research Councils have now agreed to invest £2 million in the development, by 2013, of a UK 'Gateway to Research'. In the first instance **this will allow ready access to Research Council funded research information and related data** but it will be designed so that it can also include research funded by others in due course. The Research Councils will work with their partners and users to ensure information is presented in a readily reusable form, using common formats and open standards.*

<http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf>



# iii. Funder principles/expectations

The screenshot shows the RCUK website with the following content:

- Header:** RESEARCH COUNCILS UK, Excellence with Impact, Accessibility
- Breadcrumbs:** Home > Research and Funding > RCUK Common Principles on Data Policy
- Page Title:** RCUK Common Principles on Data Policy
- Text:** Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.
- Section: Principles**
  - 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.
  - Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
  - To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
  - RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
  - To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
  - In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
  - It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

1. Public good
2. Preservation
3. Discovery
4. Confidentiality
5. First use
6. Recognition
7. Public funding

*Six of the seven RCUK councils require data management plans (or equivalent), as do Wellcome Trust, Cancer Research UK, and more...*



# The worldwide policy environment (i)

- ì USA: Lots happening. Office of Science and Technology Policy Memorandum on Open Data (February 2013)
  - ì This requires that **products** from federally-funded research be made publicly available, and that funders have a **plan** in place to make sure this happens (“Federal agencies investing in research and development (more than \$100 million in annual expenditures) must have clear and coordinated policies for increasing public access to research products.”)
    1. Maximize free public access
    2. Ensure researchers create data management plans
    3. Allow costs for data preservation and access in proposal budgets
    4. Ensure evaluation of data management plan merits
    5. Ensure researchers comply with their data management plans
    6. Promote data deposition into public repositories
    7. Develop approaches for identification and attribution of datasets
    8. Educate folks about data stewardship
  - ì The National Science Foundation and the National Institutes of Health both require data management plans to accompany funding proposals, and we are seeing these plans develop teeth as the community buys into the culture of openness and sharing

# The worldwide policy environment (ii)

- ì Canada: “Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada” – consultation document (16 October 2013) URL: [http://www.sshrc-crsh.gc.ca/about-au\\_sujet/publications/digital\\_scholarship\\_consultation\\_e.pdf](http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf)
- ì The G8 Open Data Charter (June 2013) - covers Canada, France, Germany, Italy, Japan, Russia, United Kingdom, United States of America (and the European Union) URL: <https://www.gov.uk/government/publications/open-data-charter>
- ì Internationally, the [Research Data Alliance](#) is bringing together important stakeholders from the community to ensure standards and practices are universal to ensure data is open and usable

# iv. Research quality and integrity



The INDEPENDENT

NEWS VOICES SPORT TECH LIFE PROPERTY ARTS & ENTS TRAVEL MONEY

UK World Business People Science Environment Media Technology Education Obituaries Dis

News > World > Americas

## Meet Carmen Reinhart and Kenneth Rogoff, the Harvard professors who thought they had austerity licked - and Thomas Herndon, the student who proved them wrong

Tim Walker looks at the postgraduate whose work caught out two of the biggest names in economics

TIM WALKER MONDAY 22 APRIL 2013



theguardian

News Sport Comment Culture Business Money Life &

News Politics George Osborne

## 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

- Reinhart & Rogoff (2010) "Growth in a Time of Debt" - paper not peer-reviewed, data not initially made available...
- Very influential and repeatedly cited by politicians to lend weight to economic strategy
- Multiple issues (selective exclusions, unconventional weightings, coding error) found by postgraduate researcher attempting to replicate R&R's findings
- Embarrassment all round

# v. Protection of sensitive data

There's a delicate balance between the rights of researchers, of human research subjects, of funders, and other interested stakeholders to enable or prevent access to research data...

1 September 2011 Last updated at 12:31 239 [Share](#) [f](#) [t](#) [m](#) [p](#)

## University fights Philip Morris tobacco research bid

**A Scottish university is battling a tobacco giant's attempt to gain access to its research into the smoking habits of thousands of teenagers.**

Philip Morris International (PMI), which makes Marlboro cigarettes, has submitted Freedom of Information (Fol) requests to Stirling University.

The research examines why the teenagers start smoking and what they think of tobacco marketing.



PA

The university research examines why teenagers start smoking

Philip Morris International vs University of Stirling (2011) - another example of unanticipated data re-use!



# What does this all mean?

- ì It means RDM is...
  - ì An integral part of doing quality research in the 21<sup>st</sup> century
  - ì Increasingly expected by funders, publishers and others
  - ì An opportunity for new discoveries and different approaches to research
  - ì A safeguard against inappropriate data disclosure
  - ì Something that requires active participation

# IV. BARRIERS AND ROLES FOR LIBRARIES



# Why don't we live in a data sharing utopia?

- ì Four main reasons...
  - ì Issues around ownership / privacy
  - ì Lack of understanding of the fundamental issues
  - ì Lack of joined-up thinking within institutions, countries, internationally...
  - ì Technical/financial limitations and the need for appraisal
- ì I'll now talk briefly about each of these, with some thoughts on how Libraries might get involved...

# i. Ownership and privacy

- ï **Conflicting priorities**
- ï Government/funders increasingly want data to be shared openly ASAP as the default position, **BUT** this has to be balanced against rights, both ethical and commercial. Industry may want to keep data locked down for longer periods in order to maximise profit, which can be a problem in collaborative research. (Embargo periods are generally okay, but tend to be longer in Arts disciplines than Science.)
- ï Publishers may want copyright / IPR to be handed over to them, **BUT** universities may prohibit this in their policies unless it is a condition of funding
- ï Finally, sensitive data (e.g. relating to living humans or dangerous topics) is subject to legal and ethical restrictions
- ï **ROLE FOR LIBRARIES?** Monitoring (and seeking to influence) publisher policies, and giving clear guidance to researchers on how they are affected by these

## ii. Ignorance

- ⌋ Researchers tend to find data management quite boring, and see it as an unwelcome extra bit of bureaucracy / red-tape
- ⌋ Therefore they can be reluctant to engage with the message, preferring to remain blissfully unaware
- ⌋ **ROLE FOR LIBRARIES?** Raising awareness – not only of the risks, but of the benefits of good research data management. There is a major opportunity for libraries to own the RDM space on behalf of their institutions



## iii. Lack of seamless integration

- ì Research data management is a hybrid area, and the landscape can be somewhat fragmented within institutions
- ì For example, a policy office might be in charge of compliance, the IT service in charge of storage, the Library in charge of cataloguing, HR in charge of training, etc...
- ì Mixed teams or steering groups are needed to overcome this disconnect between 'silos'
- ì **ROLE FOR LIBRARIES?** Coordinating internal activities and driving the agenda. National libraries may have a special role in monitoring international best practice

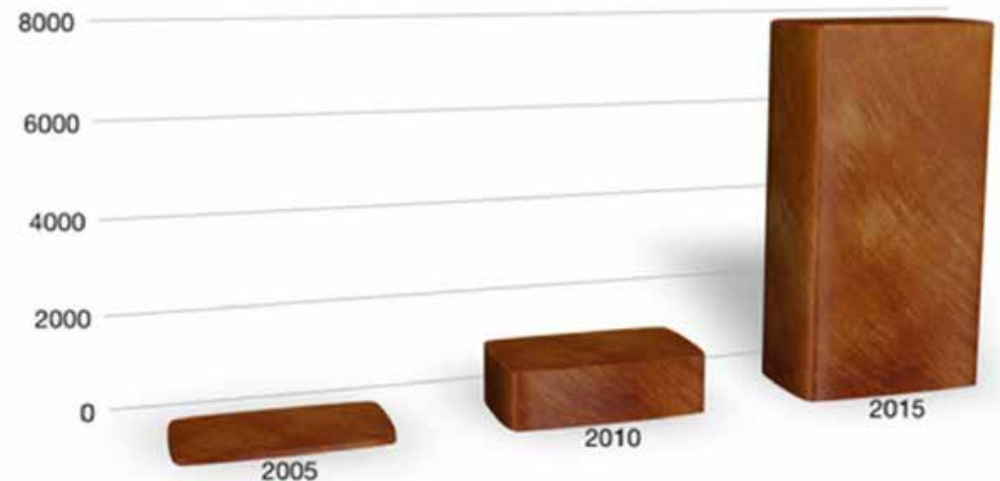


# iv. Technical and financial limitations

ì Because growth in data volume is outstripping the availability of storage, we need to *appraise* datasets, keeping only what we need...

ì David Rosenthal:  
“Keeping 2018’s data in S3 [Amazon’s Simple Storage Service] would cost the entire global GDP” <http://blog.dshr.org/2012/05/lets-just-keep-everything-forever-in.html>

A Decade of Digital Universe Growth: Storage in Exabytes



Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

According to: John Gantz and David Reinsel (2011) *Extracting Value from Chaos*, [http://www.emc.com/digital\\_universe](http://www.emc.com/digital_universe)

# V. RDM RESOURCES



# i. DCC resources

## ì Publications

- ì Briefing Papers and How-To Guides

## ì Training

- ì e.g. DC101 courses and *Curation Reference Manual*

## ì Advice

- ì e.g. Disciplinary metadata,  
[www.dcc.ac.uk/resources/metadata-standards](http://www.dcc.ac.uk/resources/metadata-standards)

## ì Events

- ì International Digital Curation Conference (next one in San Francisco, February 2014)
- ì Research Data Management Forum (themed events – next one TBC, but always held in UK... so far)

## ì Tools

- ì DMPonline, CARDIO, Data Asset Framework, DRAMBORA



## ii. Other UK resources

- ì Jisc services and resources
  - ì RDM resources, [www.jisc.ac.uk/guides/research-data-management](http://www.jisc.ac.uk/guides/research-data-management)
  - ì EDINA and Mimas (national data centres)
  
- ì JISCMRD projects – Phase 1 (2009-2011) and Phase 2 (2011-2013)
  - ì 1) Research Data Management Infrastructure (RDMI)
  - ì 2) Research Data Management Planning (RDMP)
  - ì 3) Support and Tools
  - ì 4) Citing, Linking, Integrating and Publishing Research Data (CLIP)
  - ì 5) Research Data Management Training Materials
  - ì 6) Enhancing DMPonline
  - ì 7) Events
  
- ì Universities
  - ì Good materials are available from Edinburgh, Cambridge, Oxford, Glasgow, Bristol, and many others

# iii. Further reading

- ì Two recent surveys about libraries and data...
  - ì USA & Canada – “Academic Libraries and Research Data Services: Current practices and plans for the future” - Tenopir, Birch & Allard, University of Tennessee (*Association of College & Research Libraries*, June 2012)
  - ì UK – “Research data management and libraries: Current activities and future priorities” - Cox & Pinfield, Information School, University of Sheffield (*Journal of Librarianship and Information Science*, June 2013)
- ì For more on potential future roles for librarians, see slides from Open Repositories 2013 workshop: <http://tinyurl.com/whyte-OR13>
- ì The DCC also publishes a series of themed *Briefing Papers*, *How-To Guides* and *Case Studies*, pitched at different audiences / levels of detail
  - ì <http://www.dcc.ac.uk/resources/briefing-papers>
  - ì <http://www.dcc.ac.uk/resources/how-guides>
  - ì <http://www.dcc.ac.uk/resources/developing-rdm-services>



# Thank you / Teşekkürler

Martin Donnelly  
Digital Curation Centre  
University of Edinburgh  
SCOTLAND

martin.donnelly@ed.ac.uk  
www.dcc.ac.uk  
@mkdDCC

## Image credits

Slide 1 (love note) - [http://www.edawax.de/wp-content/uploads/2013/01/Metadata\\_love250.jpg](http://www.edawax.de/wp-content/uploads/2013/01/Metadata_love250.jpg)

Slide 2 (forest) - [http://assets.worldwildlife.org/photos/934/images/hero\\_small/forest-overview-HI\\_115486.jpg?1345533675](http://assets.worldwildlife.org/photos/934/images/hero_small/forest-overview-HI_115486.jpg?1345533675)

Slide 5 (dictionary) - <http://www.flickr.com/photos/dougbelshaw/>

Slide 9 (driver) - <http://www.flickr.com/photos/rpmarks/>

Slide 10 (ocean monitoring) - <http://www.flickr.com/photos/usoceangov/>

Slide 20 (tobacco) - <http://www.bbc.co.uk/news/uk-scotland-tayside-central-14744240>

Slide 22 (barriers) - <http://www.flickr.com/photos/thetrapezium/>

Slide 23 (utopia) - <http://www.flickr.com/photos/burningmax/>

Slide 26 (silos) - <http://www.flickr.com/photos/birdwatcher63/>

Slide 28 (greenhouse) - <http://www.flickr.com/photos/mykl/>

I am indebted to Sarah Callaghan, PREPARDE, for the Rosse data publishing example



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