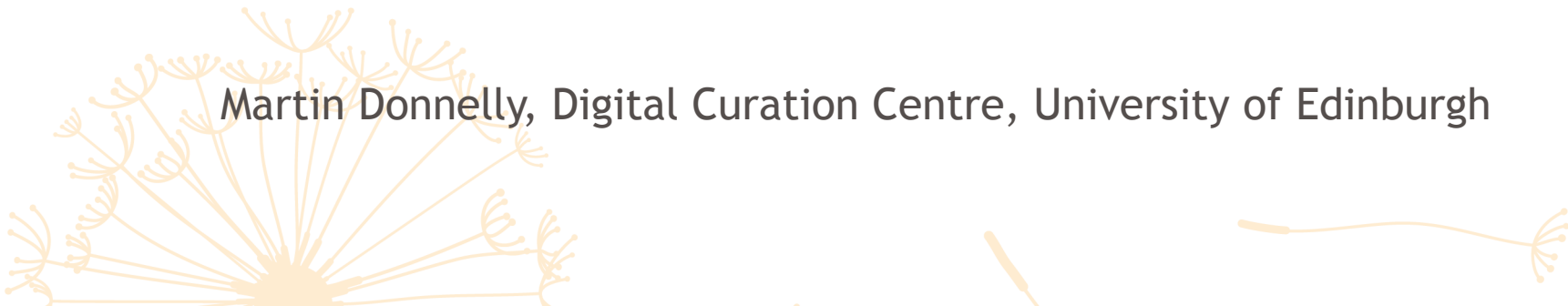




Facilitate Open Science Training for European Research

Research data management (RDM): what do support services need to know... and do?

Martin Donnelly, Digital Curation Centre, University of Edinburgh





Facilitate Open Science Training for European Research

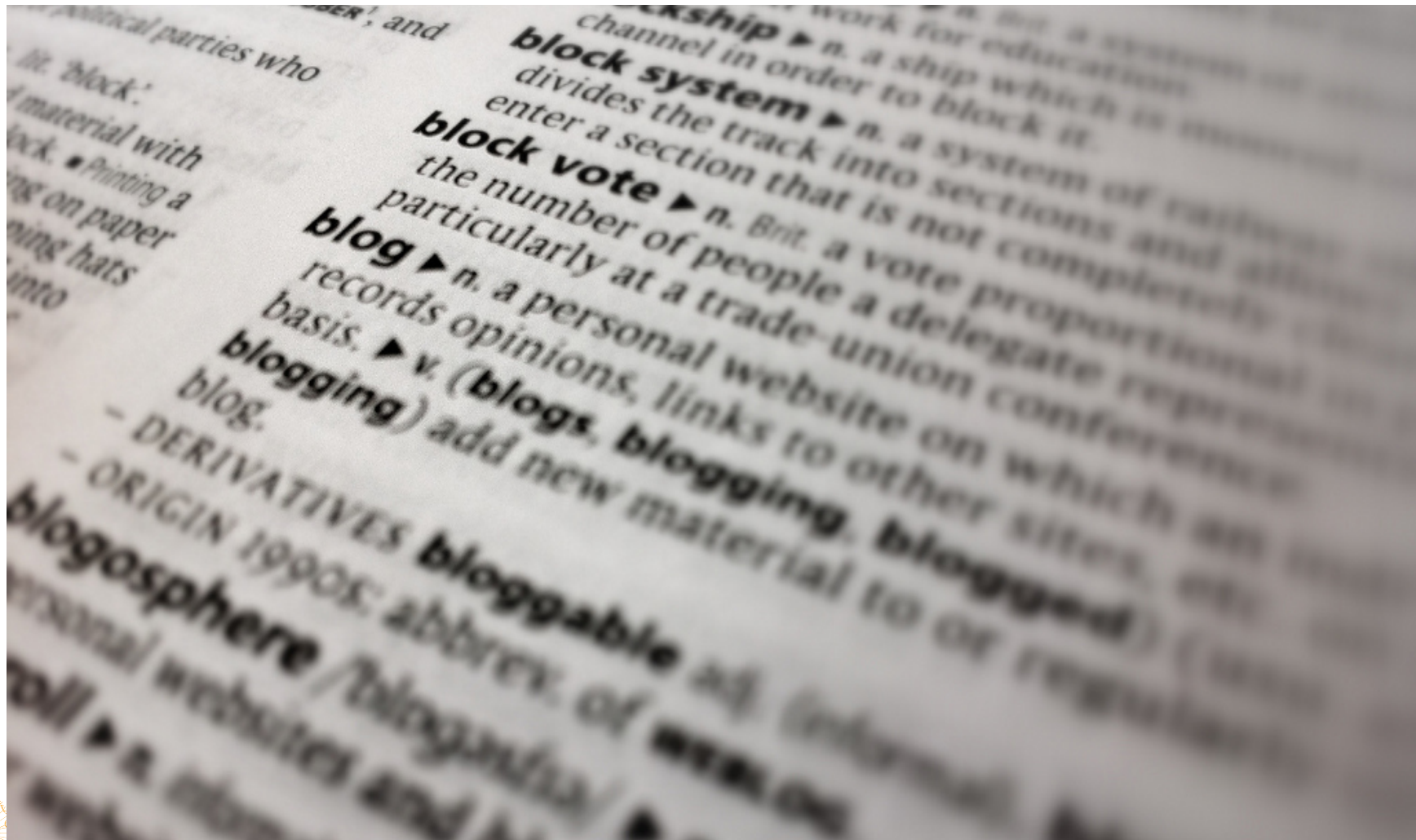


OVERVIEW

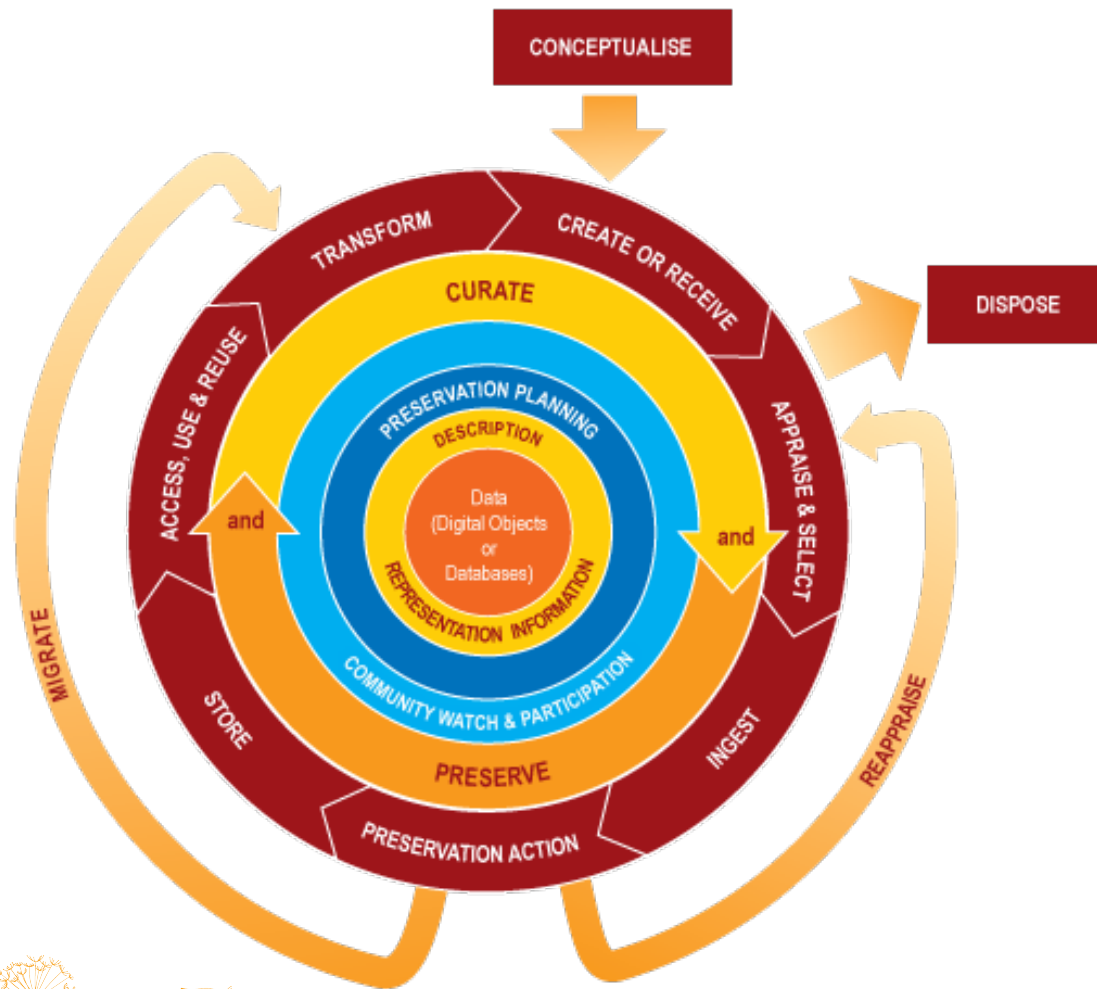
1. Introductions and definitions
2. Drivers for RDM
3. What does it mean for researchers?
4. What does it mean for support staff?



1. INTRODUCTIONS AND DEFINITIONS

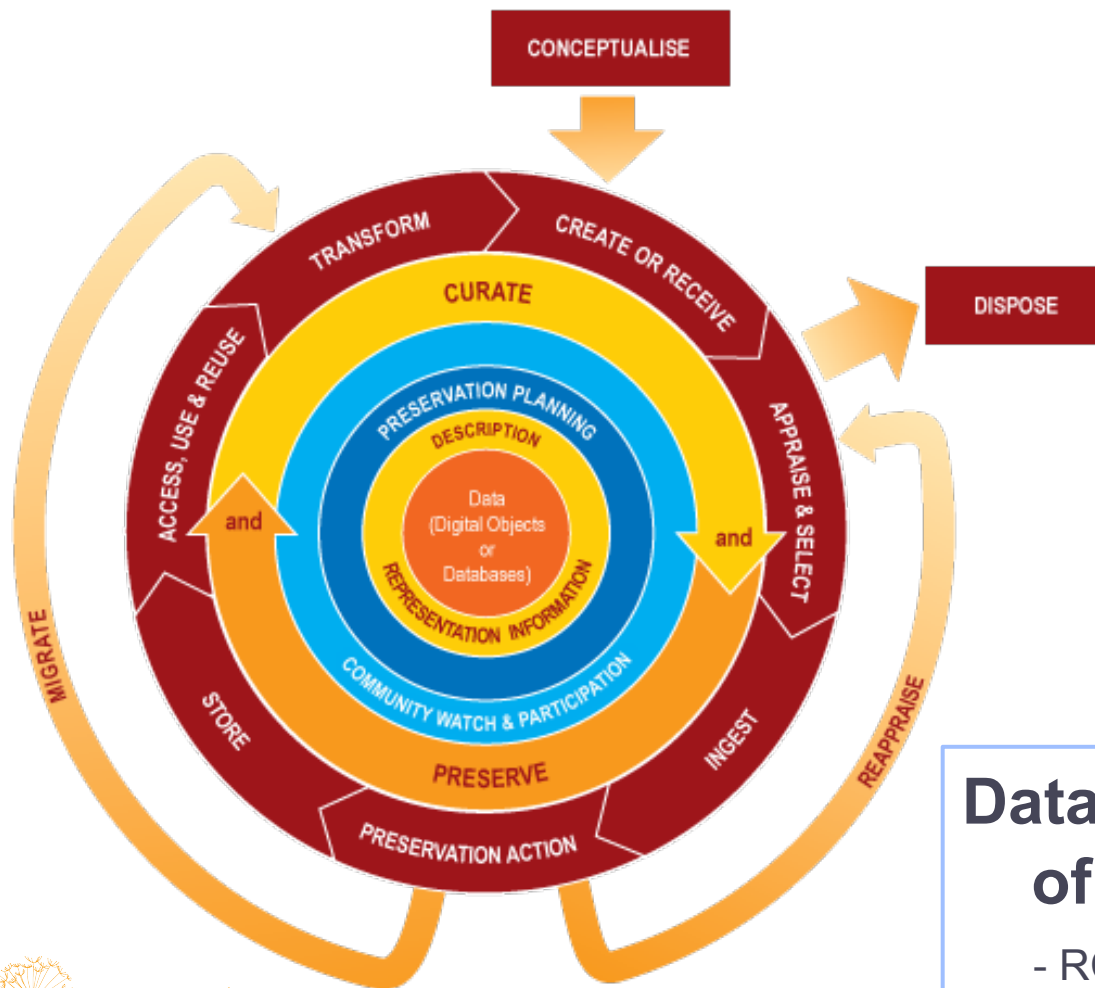


What is RDM? A definition...



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

What sort of activities?



- **Planning** and **describing** data-related work before it takes place
- **Documenting** your data so that others can find and understand it
- **Storing it safely** during the project
- **Depositing** it in a trusted archive at the end of the project
- **Linking** publications to the datasets that underpin them

Data management is a part of good research practice.

- RCUK *Policy and Code of Conduct on the Governance of Good Research Conduct*

Okay, but what is ‘data’ exactly?

- Definitions vary from discipline to discipline, and from funder to funder...
- Here’s a science-centric definition:
 - “The recorded **factual** material commonly accepted in the scientific community as necessary to **validate** research findings.” (US Office of Management and Budget, Circular 110)
 - [Addendum: This policy applies to scientific collections, known in some disciplines as institutional collections, permanent collections, archival collections, museum collections, or voucher collections, which are assets with long-term scientific value. (US Office of Science and Technology Policy, Memorandum, 20 March 2014)]
- And another from the visual arts:
 - “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/>)



From data to research objects?

- “Research object” is a term that is gaining in popularity, not least in the humanities where the relevance of the term ‘data’ is not always recognised...
- Research objects can comprise any supporting material which underpins or otherwise enriches the (written) outputs of research
 - Data (numeric, written, audiovisual....)
 - Software code
 - Workflows and methodologies
 - Slides, logs, lab books, sketchbooks, notebooks, you name it!
- See <http://www.researchobject.org/> for more info



Helicopter view: What are the benefits of active RDM?

- **TRANSPARENCY:** The evidence that underpins research can be made open for anyone to scrutinise, and attempt to replicate findings.
- **EFFICIENCY:** Data collection can be funded once, and used many times for a variety of purposes.
- **RISK MANAGEMENT:** A pro-active approach to data management reduces the risk of inappropriate disclosure of sensitive data, whether commercial or personal.
- **PRESERVATION:** Lots of data is unique, and can only be captured once. If lost, it can't be replaced.



2. DRIVERS FOR RDM

1. Technological developments
2. Value for money / Return on investment
3. Risk management
4. Transparency, integrity and good scholarly practice




Technology

- Developments in sensor technology, networking and digital storage enable new research and scientific paradigms
- As costs also fall, possibilities for data sharing, citation and re-use become much more widespread
- 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...





An earth science open access data journal

 **Geoscience Data Journal** Open Access

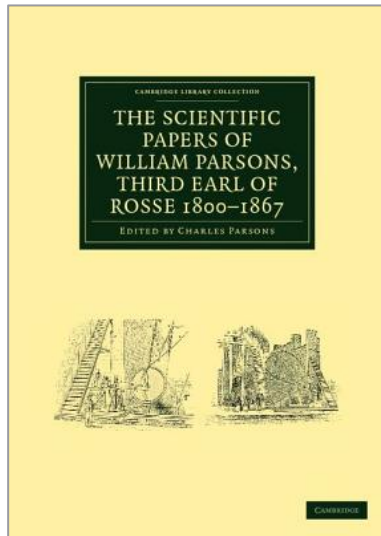
Open access to earth science data papers cross-linked to, and citing, datasets deposited in approved data centres.

Open for submissions. Geoscience Data Journal will publish its first papers in Summer 2012



www.geosciencedata.org  WILEY Open Access

Rosse



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 "novæ." 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; pB. disc. in vF haze of mottled neby.
69	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.
70		
71		
72	3	Oct. 26, 1854. A F. object with two nuclei. Nov. 29, 1850. α is vB; β has stellar point or nucleus. I suspect δ to be a F. neb.
78	4	Pos. Dist. $\alpha\beta$ 219° 5' 35" $\alpha\gamma$ 315 1 8 $\alpha\delta$ 81 0 44
79		
		Nov. 3, 1855. 3 neb. nearly in line, sp, n, f; β is bM. and IE. p. and f; α is R; bM; with a d. * np, and is the largest of the 3; ϵ is S; F; R; δ is a *.
80		1
84	4	Nov. 4, 1850. Pos. Dist. $\alpha\beta$ 169° 2' 19" $\beta\gamma$ 160 4 22 $\gamma\delta$ 201 0 34 $\gamma\epsilon$ 157 3 19 $\gamma\epsilon$ 176 5 32 $\epsilon\zeta$ 189 1 41 $\epsilon\eta$ 79 4 55
85		
86		
87	3	Oct. 26, 1854. A d. neb., both S; R; bM.
89	8	A cl. with much unresolved neby.
90	1	bM.
91	1	3 neb. in a triangle.
92		
96	6	Oct. 26, 1854. Lenticular n. and s. Thought I saw a * at times in centre (1½-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. sf.
98	1	vF; 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 "novæ" near.
104	1	Dec. 6, 1850. Pos. Dist. $\alpha\beta$ 28° 7' 36" $\beta\gamma$ 40 4 6 $\beta\delta$ 81 9 19 $\alpha\beta$ 30 7 43
105	1	Oct. 23, 1856. 6 neb., all visible at once in finder eyepiece; 2 of them E., the others S; R; bM.
106	1	Dec. 11, 1854. vmE; bM (speculum dewed).
108	8	A variety of new nebulae found, but observations too voluminous to transcribe.
112	6	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.

* This should be, I think, ζ . A S. d. neb. suspected below at α' .

AND A SELECTION FROM THE OBSERVATIONS MADE WITH THEM. 155

Number in Herschel's Catalogue.	Number of times observed.	Description.
113	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.
114		
115		
121		
116	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.
118	2	4 neb. found, 2 have nuclei. 118 is S; R; 120 has 2 *s on np. edge; E. p. and f.
120		
119	1	Dec. 9, 1854. pL; pB; bM to a nucleus.
123	2	Sept. 18, 1857. Rough sketch made; mE. np, sf; a F. triple * f.
128	3	Nov. 28, 1856. L; B; mE; B. nucleus. "Nova" f.
131	27	Nov. 29, 1850. 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 58 $\alpha\mu$ 147 5 34 $\alpha\lambda$ 179 5 56 $\alpha\kappa$ 201 5 42 $\mu\nu$ 143 6 28 $\alpha\iota$ 287 4 30 $\alpha\omega$ 341 6 45 $\alpha 2$ 5 5 18 $\alpha\psi$ 357 4 42 $\alpha 3$ 51 11 0 $\alpha\phi$ 38 9 50 $\alpha\tau$ 58 11 16 $\alpha\chi$ 161 5 20 $\alpha\chi'$ 140 6 53 $\alpha\beta'$ 172 6 32 $\alpha\gamma'$ 174 7 18 $\alpha\delta'$ 205 2 22
		Dec. 27, 1850.
		Jan. 2, 1851.
		Dec. 23, 1851.
		For previous observations see Transactions, Part II. 1850. Sept. 13, 1850. Large spiral full of knots; to nE. is a S. neb. B. which on a very good night might appear attached to spiral, than which it is brighter. Oct. 11, 1850. Spiral arrangement not clearly seen. Nov. 27, 1850. Arms of spiral scarcely seen; fog. Nov. 30, 1850. Spiral form very indistinct; wind very high from a. Oct. 22, 1851. Viewed for drawing, I should not have seen the spiral arrangement had I not observed it before. Oct. 25, 1851. Neby. extends for several minutes all round, perhaps for half a degree in radius. Oct. 29, 1851. Observed for drawing. Dec. 14, 1851. Sketched. Dec. 26, 1851. Drawn. Dec. 7, 1855. This neb. reaches in length through at least a field and a half of finder eyepiece. Mr. Stoney's drawing leaves out a great deal of the neby. about the centre, and * suspected to left of centre of the trapezium of *s, perhaps others also. Nov. 15, 1857. There are 3 *s about the principal nucleus. Dec. 7, 1857. Carefully observed, with a view to a new sketch. Dec. 18, 1857. Carefully observed, and my sketch proceeded with. See fig. 10, Plate XXVI.
132	1	Nov. 28, 1856. B; S; R. nucleus, a * p. and another n.
Nova.		Nov. 29, 1850. A S. neb. or cl. with 3 *s in it. R 1° 26'. N.P.D. 60° 35'.
134	2	Oct. 26, 1854. Both S; R; E.
135		
136	1	Sought for four times; not found.
142	8	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.

Repurposing / VfM via data re-use

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

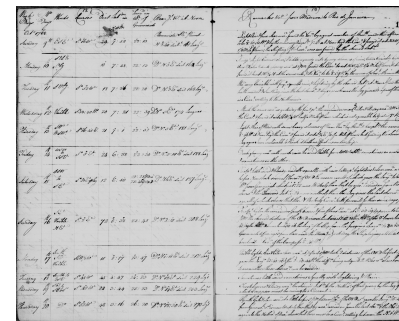
You can now help scientists understand the climate of the past and unearth new historical information by revisiting the voyages of First World War Royal Navy warships.

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.

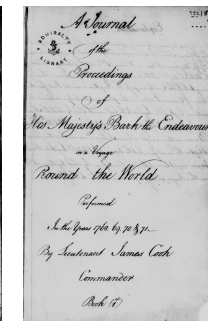
<http://www.nationalarchives.gov.uk/news/503.htm>



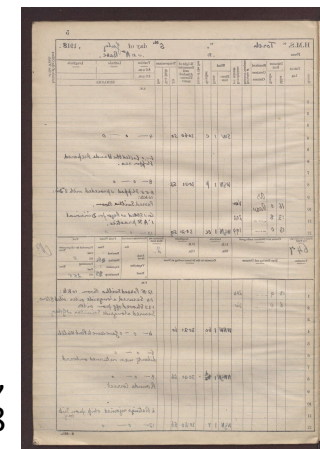
Detail from Royal Navy Recruitment poster, RNVR Signals branch, 1917 (Catalogue reference: ADM 1/8331)



HMS Beagle, 1830-34



Endeavour, 1768-71 (Captain Cook)



HMS Torch, 1918

Funder principles/expectations

- Major funders in many countries now have data management policies, and mandate data management plans (DMPs)
- In the UK, the RCUK councils have seven shared principles which underpin their individual policies...
 1. Data as a 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 European Commission is running an Open Data pilot in Horizon 2020, about which more later...



Risk management

Controversial FOI requests to...

- University of East Anglia
- Queen's University Belfast
- University of Stirling

BBC NEWS | Sign in | News | Sport | Weather | iPlayer | TV | Radio | More... | Search BBC News

NEWS | **BBC NEWS AT SIX**

Page last updated at 06:45 GMT, Monday, 19 April 2010 07:45 UK

University told to hand over tree ring data

Queen's University in Belfast has been told by the Information Commissioner to hand over 40 years of research data on tree rings, used for climate research.

Douglas Keenan, from London, had asked for the information in 2007 under the Freedom of Information Act.

Mr Keenan is well-known for his questioning of scientists who propose a human cause for climate change.

Queen's University refused his request saying it was too expensive, but it is now considering its position.

The university claimed that as the information was unfinished, had intellectual property rights and was commercially confidential.

SEE ALSO

- Q&A: Professor Phil Jones 13 Feb 10 | Science & Environment
- Climate scientist defends results 02 Feb 10 | Norfolk

RELATED INTERNET LINKS

- Douglas Keenan's website
- Queen's University
- Information Commissioner

TOP NORTHERN IRELAND STORIES

- Two in court over Twelfth riots
- Major arrests promised over riots
- Riot tactics defended by police

WIKIPEDIA The Free Encyclopedia

Climatic Research Unit email controversy

From Wikipedia, the free encyclopedia

The **Climatic Research Unit email controversy** (also known as "**Climategate**")^{[2][3]} began in November 2009 with the **hacking** of a server at the **Climatic Research Unit (CRU)** at the **University of East Anglia (UEA)** by an external attacker.^{[4][5]} Several weeks before the **Copenhagen Summit** on climate change, an unknown individual or group breached CRU's server and copied thousands of emails and computer files to various locations on the Internet.

The story was first broken by **climate change critics** on their blogs,^[6] with columnist **James Delingpole** popularising the term "Climategate" to describe the controversy.^[7] Climate change critics and others *denying the significance of human caused climate change* argued that the emails showed that global warming was a *scientific conspiracy*, in which they alleged that scientists manipulated climate data and attempted to suppress critics.^{[8][9]} The accusations were rejected by the CRU, who said that the emails had been taken out of context and merely reflected an honest exchange of ideas.^{[10][11]}

The mainstream media picked up the story as negotiations over **climate change mitigation** began in Copenhagen on 7 December.^[12] Because of the timing, scientists, policy makers, and public relations experts said that the release of emails was a **smear**.

Climatic Research Unit email controversy

Date	17 November 2009
Location	Climatic Research Unit, University of East Anglia
Also known as	"Climategate"
Inquiries	House of Commons Science and Technology Committee (UK) ^[1] Independent Climate Change Email Review (UK) International Science Assessment Panel (UK) Pennsylvania State University (US) The United States Environmental Protection Agency (US) Department of Commerce (US)

1 September 2011 Last updated at 12:31

239 Share

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 (FoI) requests to Stirling University.

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

The university research examines why teenagers start smoking

Research quality and integrity



the guardian

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



The INDEPENDENT

NEWS | VOICES | SPORT | TECH | LIFE | PROPERTY | ARTS & ENTS | TRAVEL | MONEY | IT

UK | World | Business | People | Science | Environment | Media | Technology | Education | Obituaries | Disasters

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

- 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) identified by a postgrad researcher attempting to replicate the paper’s findings
- Widespread embarrassment, but at least the errors were discovered!

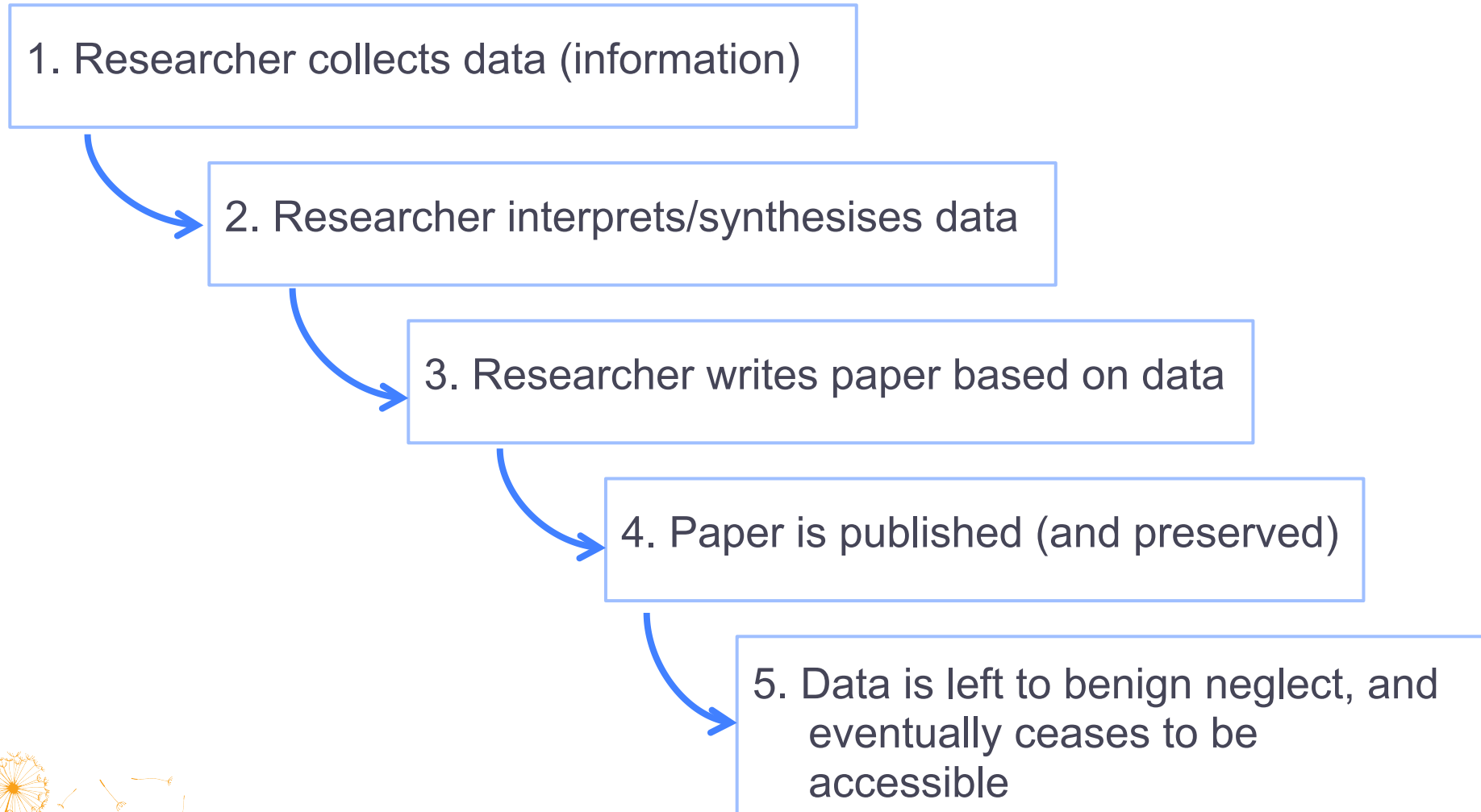
 FOSTER

3. WHAT DOES IT MEAN FOR RESEARCHERS?

- A disruption to their working processes
- Additional expectations / requirements from the funders
- But! It provides opportunities for new types of investigation
- And leads to a more robust scholarly record



The old way of doing things



Without intervention, data + time = no data

Vines et al. “examined the availability of data from 516 studies between 2 and 22 years old”

- The odds of a data set being reported as extant fell by 17% per year
- Broken e-mails and obsolete storage devices were the main obstacles to data sharing
- Policies mandating data archiving at publication are clearly needed

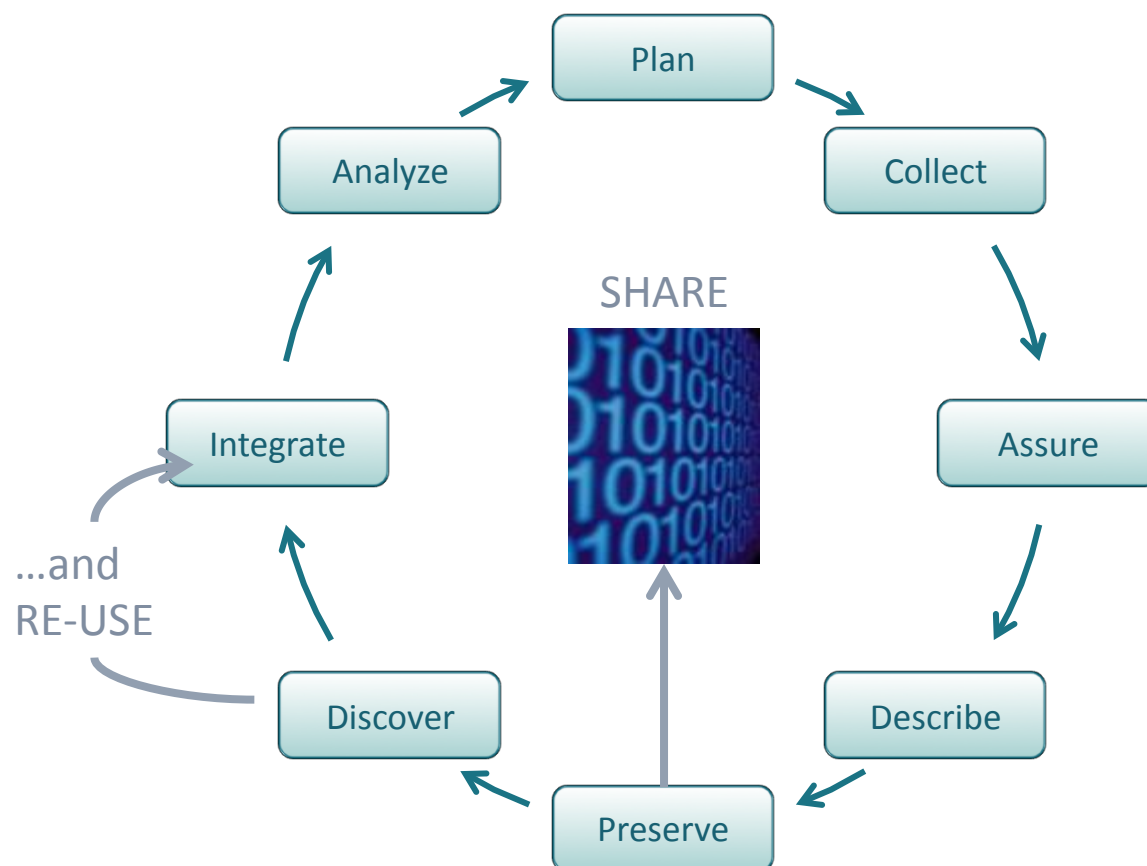
“The current system of leaving data with authors means that almost all of it is lost over time, unavailable for validation of the original results or to use for entirely new purposes” according to Timothy Vines, one of the researchers. This underscores the need for intentional management of data from all disciplines and opened our conversation on potential roles for librarians in this arena. (“80 Percent of Scientific Data Gone in 20 Years,” *HNGN*, Dec. 20, 2013,

<http://www.hngn.com/articles/20083/20131220/80-percent-of-scientific-data-gone-in-20-years.htm>.)

Vines et al., The Availability of Research Data Declines Rapidly with Article Age, *Current Biology* (2014), <http://dx.doi.org/10.1016/j.cub.2013.11.014>

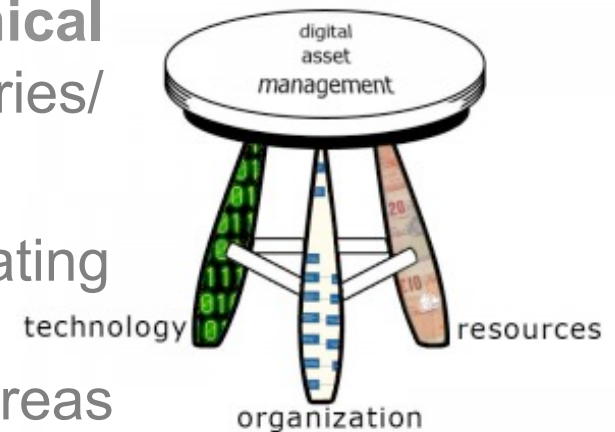


The new way of doing things



What does it mean for universities?

- Three principal areas of focus
 - Developing and integrating their **technical** infrastructure (storage space, repositories/ CRIS systems, data catalogues, etc)
 - Developing **human** infrastructure (creating policies, assessing current data management capabilities, identifying areas of good practice, data management plan templates, tailoring training and guidance materials...)
 - Developing **business** plans for sustainable services / roles
- Forming cross-function (hybrid) working groups, advisory groups, task forces, etc...



<http://blog.soton.ac.uk/keepit/2010/01/28/aida-and-institutional-wobbliness/>



FOSTER

4. WHAT DOES IT MEAN FOR SUPPORT STAFF?

- Need to understand the key elements in the process, as well as roles and responsibilities
- Understand the key points of the funders' requirements
- Expect questions from researchers...



Understand the different roles

- Three main roles for research support staff...
 - **Compliance:** checking adherence with funder policies, at both ends of the funding process (pre-award and end-of-project)
 - **Guidance:** helping researchers meet expectations and requirements
 - **Selection etc:** some staff may also have an appraisal and retention role, making decisions re. what the institution will want to keep / share, under what conditions, and for how long. There are various reasons for universities to want to keep some datasets, and to get rid of others.
- Different institutions organise their provision in different ways; there's no one-size-fits-all approach



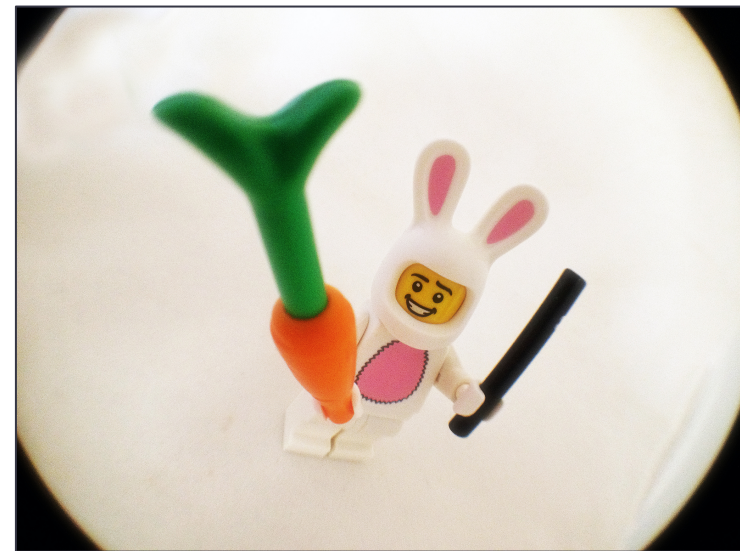
Understand funder requirements

- The DCC maintains an overview of the major UK and European funders' data-related expectations / requirements
 - <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>
- The European Commission has introduced an Open Data pilot in Horizon 2020
 - Details:
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf



Expect questions from researchers

- ...and know where to direct enquiries
- Also, expect resistance!
 - It can pay to focus on the benefits rather than hammering home the requirements, but be clear that DMPs (and good RDM practice, more broadly) are no longer optional for many funders (and publishers)



Last slide: take-home messages

- Research data management (RDM) is...
 - An integral part of doing quality research in the 21st century
 - Increasingly expected / mandated by funders, publishers and others
 - An opportunity for new discoveries and different approaches to research
 - A safeguard against inappropriate data disclosure
 - An activity that requires careful planning and consideration, and – ideally – coordination and support across many stakeholder types



THANK YOU

Martin Donnelly
Digital Curation Centre
University of Edinburgh

martin.donnelly@ed.ac.uk

Twitter: @mkdDCC

www.dcc.ac.uk

www.fosteropenscience.eu



Image credits: slide 3, <http://www.flickr.com/photos/dougbelshaw/>; slide 9, <http://www.flickr.com/photos/rpmarks/>; slide 24, <https://www.flickr.com/photos/kaptainkobold>

Thanks to Sarah Callaghan, PREPARDE, for the Rosse example.

All images are Creative Commons licensed.