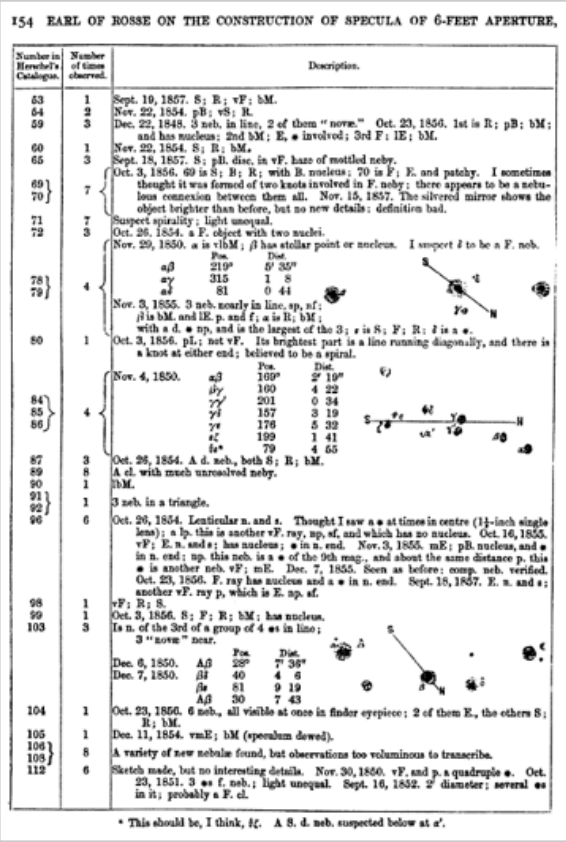
Research Data Management

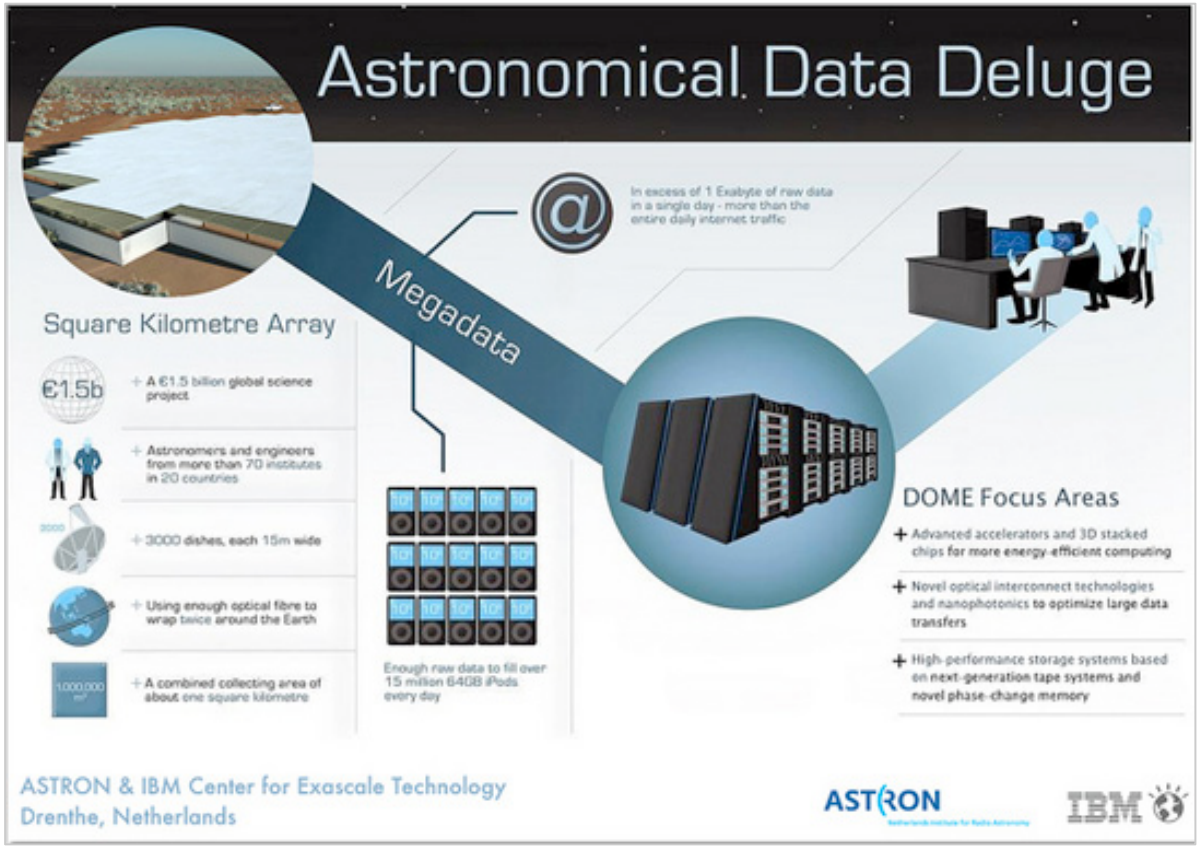
Introduction and overview



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; lE; bM.
60	1	Nov. 22, 1854. S; R; bM.
65	3	Sept. 18, 1857. S; pB. 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.
78, 79	4	Nov. 29, 1850. a is vibM; β has stellar point or nucleus. I suspect δ to be a F. neb. Pos. Dist. αβ 219° 5′ 35″ αγ 315 1 8 αδ 81 0 44
80	1	Nov. 3, 1855. 3 neb. nearly in line, sp, sf; β is bM. and lE. p. and f; a is R; bM; with a d. ● sp, and is the largest of the 3; γ is S; F; R; δ is a ●.
84, 85, 86	4	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. Nov. 4, 1850. Pos. Dist. αβ 109° 2′ 19″ βγ 100 4 22 γγ′ 201 0 34 γδ 157 3 19 γε 176 5 32 δζ 199 1 41 δη* 79 4 55
87	3	Oct. 26, 1854. A d. neb., both S; R; bM.
89	8	A cl. with mueh unresolved neby.
90	1	bM.
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 (1½-inch single lens); a lp. this is another vF. ray, sp, 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; sp. 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. sp. sf.
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 "novæ" near. Pos. Dist. Dec. 6, 1850. Aβ 28° 7′ 36″ Dec. 7, 1850. βδ 40 4 6 δε 81 9 19 Aβ 30 7 43
104	1	Oct. 23, 1856. 6 neb., all visible at once in finder eyepiece; 2 of them E., the others S; R; bM.
105	1	Dec. 11, 1854. vmE; bM (speculum dewed).
106, 108	8	A variety of new nebulæ 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, δζ. A S. d. neb. suspected below at α′.



About this course

Short presentations with exercises and discussion…

- Quick quiz on funder expectations (Martin)
- The Research Data Management Landscape (Martin)
- Introduction to Data Sharing and Data Management Planning (Sarah)
- Data management planning exercise (Sarah and Martin)
- DMPonline demo (Sarah)

PART 1

QUICK QUIZ ON FUNDER EXPECTATIONS

Questions

1. How long do RCUK funders typically expect research data of long-term value to be preserved?

a. 10+ years

b. In perpetuity

c. 5 years

2. Which RCUK funders expect researchers to submit data management and sharing plans as part of their grant application?

a. All of them

b. None of them

c. Some of them

Questions

3. When are researchers in receipt of RCUK grants expected to make their data available?

a. As soon as possible, typically on publication of results

b. Within 3 years of the end of the award

c. It's up to the researcher to decide

d. Never

Questions

4. Will RCUK funders consider withholding the final grant payment if data are not offered for deposit at designated data centres?

a. Yes

b. No

5. Will RCUK funders provide additional funds to cover the costs associated with data management and sharing?

a. Yes

b. No

Answers

Question 1

a. Correct! Most funders expect data to be preserved for 10 years, or longer if it has particular significance.

b. Wrong answer. Only STFC asks organisations to retain data which can not be re-measured 'in perpetuity'. And the EPSRC requirement can in effect mean keeping data forever, if there are regular access requests. However, most funders state 10 years as a reasonable minimum period.

c. Wrong answer. Only Cancer Research UK notes a minimum period of 5 years. Typically funders expect data to be preserved for 10 years or more

Answers

Question 2

a. Wrong answer.

b. Wrong answer.

c. Correct answer. The majority of RCUK funders now expect a data management plan at the grant application stage. BBSRC has asked researchers to submit data sharing plans since 2007. STFC introduced a requirement for data management plans in 2011. EPSRC does not ask for DMPs to be submitted in grant applications but expects that one will exist locally.

Answers

Question 3 – trick question, as all are technically correct

a. Correct - Research funders typically expect data to be available as soon as possible.

b. Correct - Embargo periods are allowed and should be based on common practice for the particular discipline. This may mean that data can't be shared for a number of years, however researchers should endeavour to share as soon as possible.

c. Correct - researchers need to determine whether their data can be shared as this may not be possible due to legal or ethical reasons. However, where data can be shared, they should be made available as soon as possible.

d. Correct - Most funders have expectations for some data to be shared unless there are very clearly defined reasons that this is not possible. These should be made explicit at the grant application stage to avoid problems toward the end of the project.

Answers

Question 4

a. Correct answer. In cases where funders sponsor a data centre, researchers are required to offer selected research outputs for deposit as part of the condition of funding. Final grant payment can be withheld if this does not happen. But note that funders do not require all data produced in the project – just selected outputs.

b. Wrong answer.

Answers

Question 5

a. Correct answer. RCUK funders have signed up to the RCUK Common Principles on Data Policy, which say it is appropriate to use public funds to support the management and sharing of publicly funded research data. However, the costs need to be very explicitly described and justified in the bid. ESRC will review any costs associated with implementing the data plan as an integral part of the funding decision, and based on this decision, provide appropriate funding for data management. At the moment, the only UK funder who explicitly will not fund RDM and data sharing are CRUK as they consider timely and appropriate data management and sharing an integral component of the research process, so will not provide additional funds for these activities.

b. Wrong answer.

PART 2

THE RESEARCH DATA MANAGEMENT LANDSCAPE

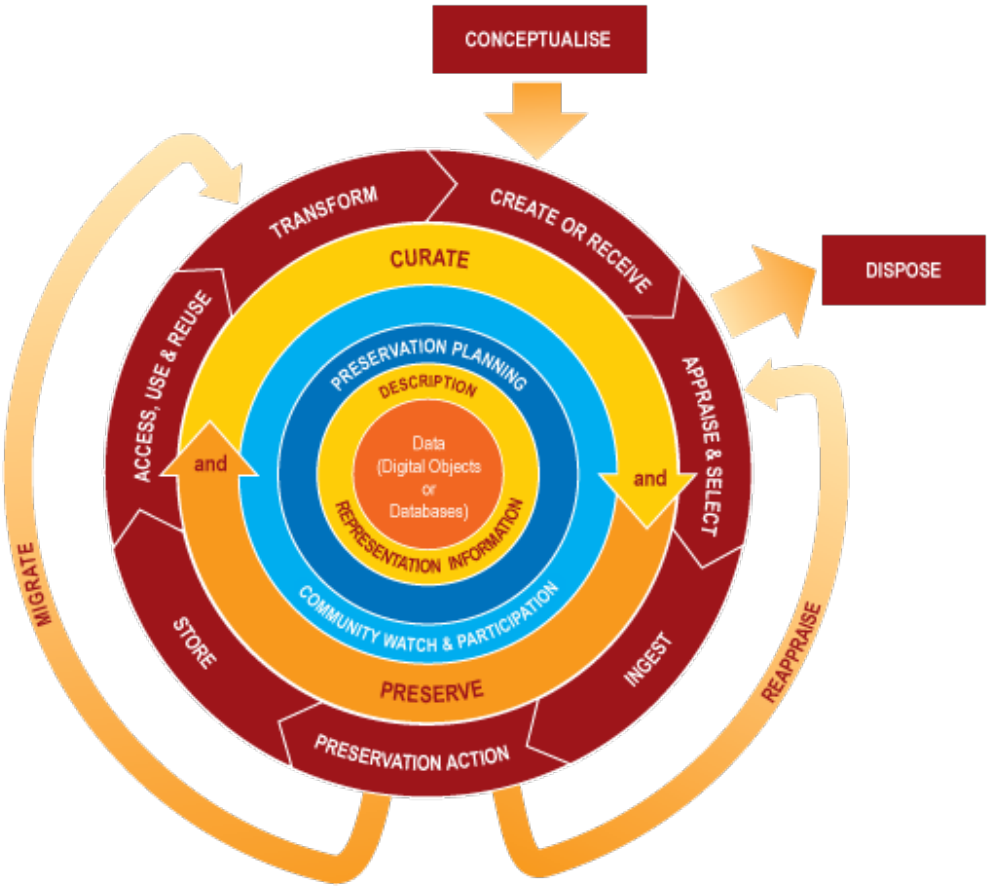
Outline for Part 2

1. About the Digital Curation Centre (DCC)
2. What is research data (management)?
3. Drivers for change
4. Recommendations, expectations and requirements
5. What help is available?

The Digital Curation Centre

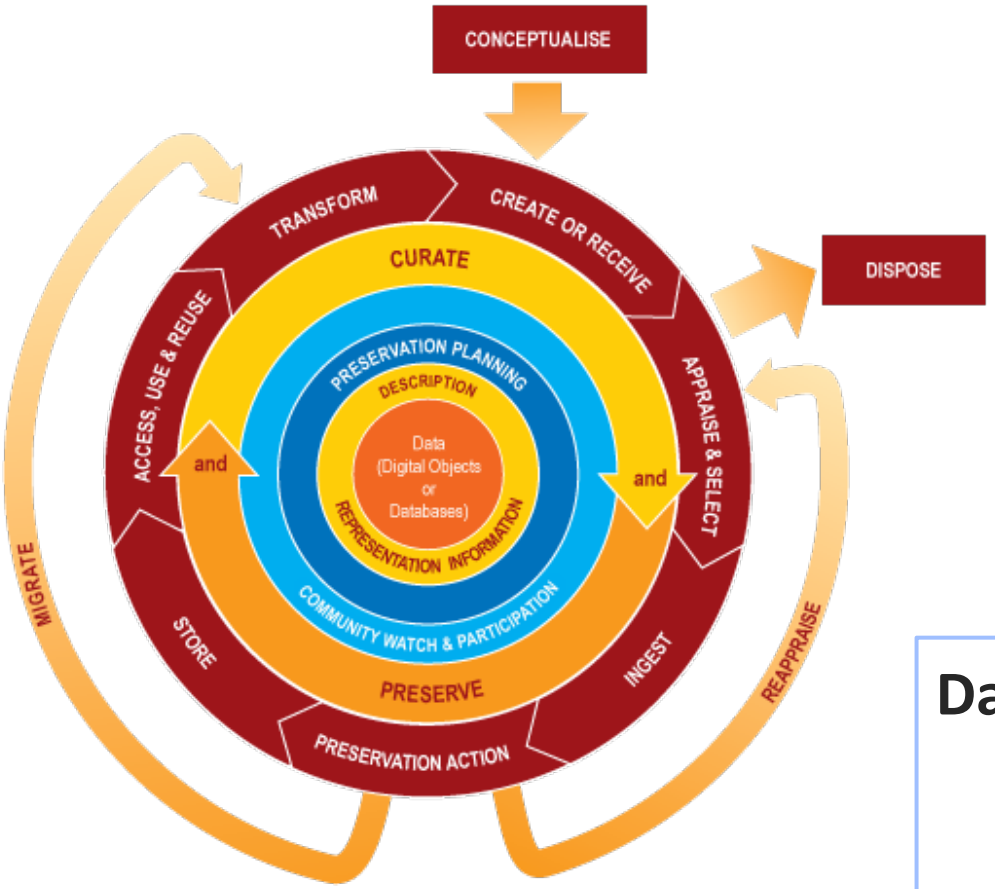
- The DCC (est. 2004) is…
  - A UK national centre of expertise in digital preservation, with a particular focus on research data management (RDM)
  - Based across three sites: Universities of Edinburgh, Glasgow and Bath
  - Working with a number of UK universities to identify gaps in RDM provision and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

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

DRIVERS FOR RDM

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

Funder expectations re. 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

Research funder data policies

● Full Coverage ◐ Partial Coverage ○ No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

www.dcc.ac.uk/resources/policy-and-legal/ overview-funders-data-policies

Ultimately funders expect…

- timely release of data
  - once patents are filed or on (acceptance for) publication
- open data sharing
  - minimal or no restrictions (where appropriate)
- preservation of data
  - typically 5-10+ years

See the RCUK Common Principles on Data Policy:
www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

EPSRC

"Research organisations will ensure that effective data curation is provided throughout the full data lifecycle, with 'data curation' and 'data lifecycle' being as defined by the Digital Curation Centre. The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated..."

i.e. **institutional** responsibility

www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx

RDM costs and other support

Research funders' policies

● Full Coverage ◐ Partial Coverage ○ No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

Key differences in policies

- Retention periods range from 3 years to forever
- ESRC and NERC may withhold the final grant payment if data are not offered for deposit in their data centres
- Cancer Research UK states explicitly that it will **NOT** provide additional funds for RDM

Which RDM costs can be included?

It's necessary to distinguish between costs that are incurred during a project and those that arise afterwards.

- In-project (direct) costs:
  - covers hardware, staff, expenses, costs of preparing data & metadata…
- Post project (largely indirect) costs:
  - existing services should be used where possible
  - where an institution is going to provide a data repository, costs should be met through FEC
  - outsourcing to a third-party is also an option

Owing to its charitable status, the Wellcome Trust in general only pays directly incurred costs.

How should costs be included?

- In-project costs should be included in the direct costs for a project
- Post-project costs could be direct (e.g. charges levied by data centres) but typically fall into indirects as universities should provide infrastructure to support RDM
- The Justifications of Resources should, where possible, separate out the following RDM cost elements:
  - cost of collecting data
  - the cost of curating data
  - the cost of analysing data
  - the cost of preservation and sharing

KEY MESSAGES

- Research data management is increasingly expected by funders: it should not be regarded as an optional extra
- The earlier you consider RDM, the better it will be (see DMP session coming up…)
- DMPs should make clear what is provided and what activities are being charged against a grant
- The costs of RDM are project-specific and will depend on the type of work involved
- It may be possible to set up small research facilities to recover the cost of RDM (e.g. similar to provision of HPC), possibly as a cross-institutional service… perhaps something for discussion later

Thank you

Martin Donnelly
Digital Curation Centre
University of Edinburgh

martin.donnelly@ed.ac.uk
@mkdDCC

For more about DCC tools, services and resources see www.dcc.ac.uk
or follow us on twitter @digitalcuration and #ukdcc

Thanks to Joy Davidson for the use of her slides, upon which this presentation builds.