Data: opportunities for libraries

Kevin Ashley
Digital Curation Centre
www.dcc.ac.uk
@kevingashley
Kevin.ashley@ed.ac.uk

The DCC is supported by Jisc

DCC because good research needs good data

A summary

- A recap of why anyone cares about research data
- A focus on research data: library roles, institutional roles
- A nod to other data-related roles for libraries
- Note absence of BIG
  - A large collection of small data is big data
- But first, an apology

I am a

blogging about digital archives, scholarship & repositories since 2007

Home | Digital Archives | Repositories | Training | Web Archiving | Digitisation | Links | About the blog

B	C	D	E	F	G	H	I	J	K	L
1281	8203	2	HARTLEPOOLS WATER COMPANY	8104	1	60	1	1	4	T
1281	8303	1	HARTLEPOOLS WATER COMPANY	8204	1	60	1	1	4	T
1281	8403	1	HARTLEPOOLS WATER COMPANY	8304	1	60	1	1	4	T
1281	8503	1	HARTLEPOOLS WATER COMPANY	8404	1	60	1	1	4	T
1281	8603	1	HARTLEPOOLS WATER COMPANY	8504	1	60	1	1	4	T
1521	8112	2	WREXHAM AND EAST DENBIGHSHIRE WATER COMP	8101	1	60	1	1	3	T
1521	8212	1	WREXHAM AND EAST DENBIGHSHIRE WATER CO	8201	1	60	1	1	3	T
1521	8312	1	WREXHAM AND EAST DENRIGHSHIRE WATER CO	8301	1	60	1	1	3	T
1521	8412	1	WREXHAM AND EAST DENBIGHSHIRE WATER CO.	8401	1	60	1	1	3	T
1521	8512	1	WREXHAM AND EAST DENBIGHSHIRE WATER CO	8501	1	60	1	1	3	T
1581	8109	2	MID-SUSSEX WATER COMPANY	8010	1	60	1	1	2	T
1581	8109	4	MID-SUSSEX WATER COMPANY	8010	1	60	1	1	2	T
1581	8209	1	MID-SUSSEX WATER COMPANY	8110	1	60	1	1	2	T
1581	8403	1	MID-SUSSEX WATER COMPANY	8210	1	60	1	1	4	T
1581	8503	1	MID-SUSSEX WATER COMPANY	8404	1	60	1	1	4	T
1581	8503	4	MID-SUSSEX WATER COMPANY	8404	1	60	1	1	4	T
1581	8603	1	MID-SUSSEX WATER COMPANY	8504	1	60	1	1	4	T
1781	8203	2	LEE VALLEY WATER COMPANY	8104	1	60	1	1	4	T
1781	8303	1	LEE VALLEY WATER COMPANY	8204	1	60	1	1	4	T
1781	8403	1	LEE VALLEY WATER COMPANY	8304	1	60	1	1	4	T
1781	8503	1	LEE VALLEY WATER COMPANY	8404	1	60	1	1	4	T
1781	8603	1	LEE VALLEY WATER COMPANY	8504	1	60	1	1	4	T
1891	8112	2	MERSEY DOCKS & HARBOUR COMPANY	8101	1	70	1	1	3	T
1891	8212	1	MERSEY DOCKS & HARBOUR COMPANY	8201	1	70	1	1	3	T
1891	8312	1	MERSEY DOCKS AND HARBOUR COMPANY	8301	1	70	1	1	3	T

pp. 5-13 **Prenormative Research into Standard Messaging Formats for Engineering Materials Data**
Tim Austin, Chris Bullough, Dimitri Gagliardi, David Leal, Malcolm Loveday
ABSTRACT | FULL TEXT PDF | doi:10.2218/ijdc.v8i1.245

pp. 14-28 **Can Persistent Identifiers Be Cool?**
Barbara Bazzanella, Stefano Bortoli, Paolo Bouquet
ABSTRACT | FULL TEXT PDF | doi:10.2218/ijdc.v8i1.246

pp. 29-41 **Data Management and Preservation Planning for Big Science**

FONT SIZE

INFORMATION
- For Readers
- For Authors
- For Librarians

My home – the DCC

- Mission – to increase capability and capacity for research data services in UK institutions
- Not just a UK problem – an international one
- Training, shared services, guidance, policy, standards, futures

What is data curation ?

- “Maintaining, preserving and adding value to research data throughout its lifecycle”
- More than preservation:
  - Active management – dealing with change
- Less than preservation:
  - Lifecycle sometimes involves destruction
- Sometimes, not always, about sharing, publication or citation

Why care?

- Data is expensive – an investment
- Reuse:
  - More research
  - Teaching & Learning
  - Planning
- Impact – with or without publication
- Accountability
- Legal & regulatory requirements

Why does this matter?

- Research quality
  - How close can we get to the truth?
- Research speed
  - How quickly can we get to the truth?
- Research finance
  - How much does the truth cost?

- Improving one or more of these is of interest to all actors:
- Researchers as data creators
- Researchers as data reusers
- Research institutions
- Funders – hence government and society

G8UK - Endorses OA Open Data Charter Policy Paper 18 June 2013

G8UK - Billigt offenen Zugang Eine offene Daten Charter Strategiepapier.

Cabinet Office

Policy paper

G8 Open Data Charter and Technical Annex

Published 18 June 2013

Contents

1. Principle 1: Open Data by Default
2. Principle 2: Quality and Quantity
3. Principle 3: Usable by All
4. Principle 4: Releasing Data for Improved Governance
5. Principle 5: Releasing Data for Innovation
6. Technical annex

Funder requirements

- UK

http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx

- USA – NSF, NEH, NIH
- Europe

- Most place burden on researcher – some on the institution

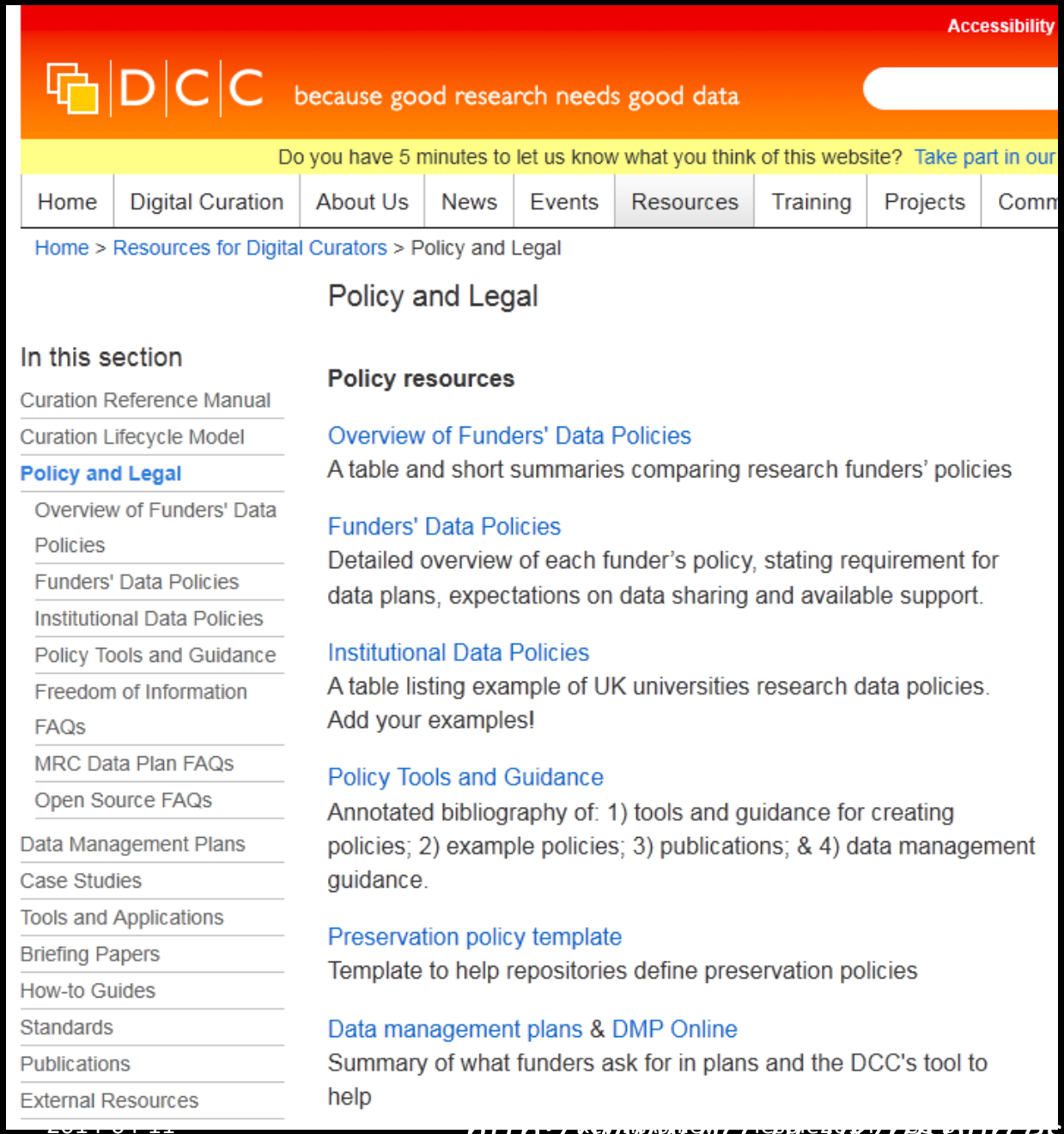
RCUK policy - The 1-minute version

- Research data are a public good – make openly available in timely & responsible way
- Have policies & plans. Data with long-term value should be preserved & usable
- Metadata for discovery & reuse. Link publications & data
- Sometimes law, ethics get in the way. We understand.
- Limited embargos OK. Recognition is important – always cite data sources
- OK to use public money to do this. Do it efficiently.

EPSRC policy points

- Awareness of regulatory environment
- Data access statement
- Policies and processes
- Data storage
- Structured metadata descriptions
- DOIs for data
- Securely preserved for a minimum of 10 years **from last use**

Compliance expected by 2016



DCC Policy Summary

http://www.dcc.ac.uk/resources/policy-and-legal

Data reuse stories

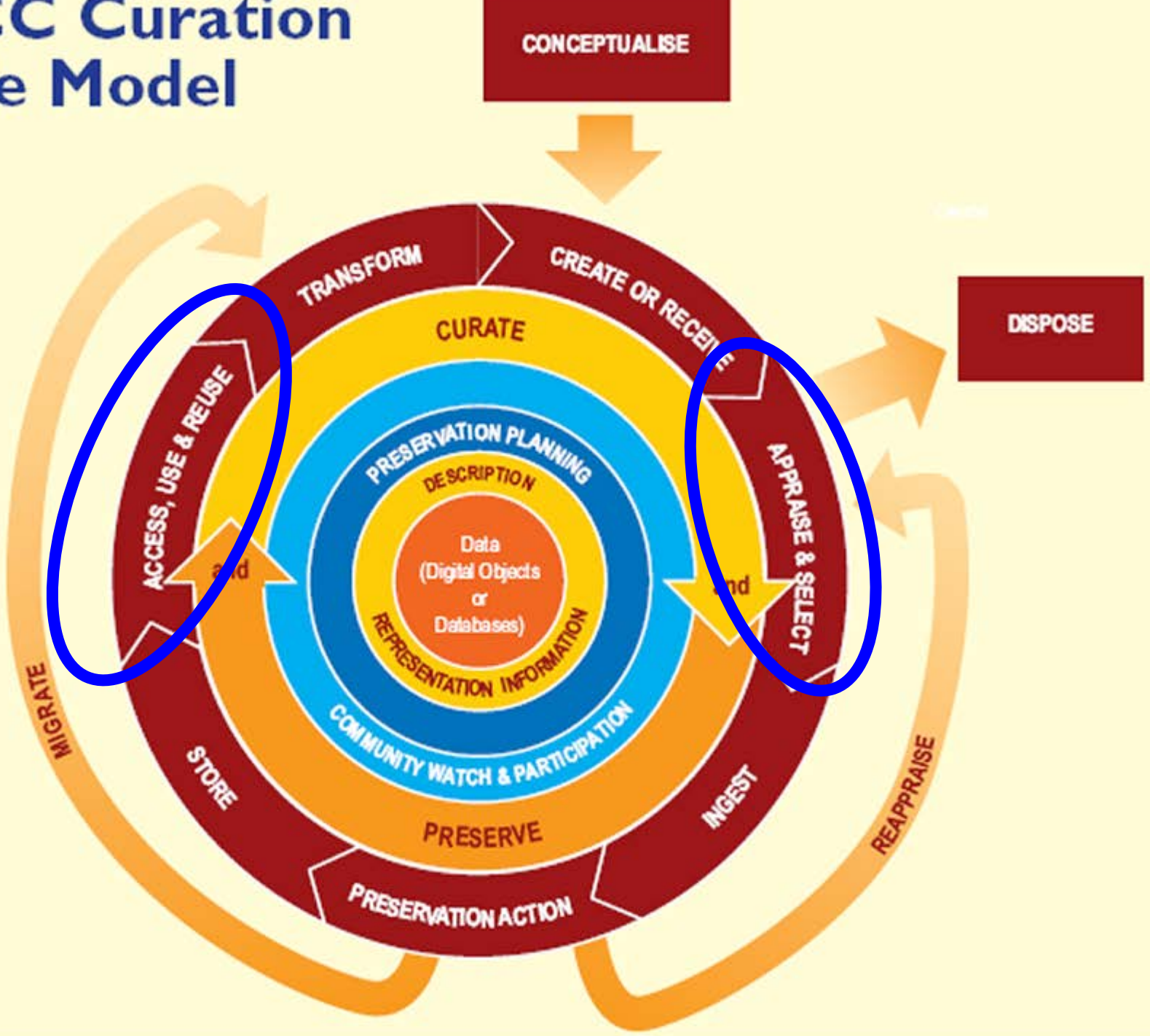
- The palaeontologist who saved years of work with archaeological data
- The ‘noise’ from research radar that mapped dust from Eyjafjallajökull
- The 19th-century ships logs that help us model climate change

Often your data tells stories that your publications do not

Compliance

Benefits

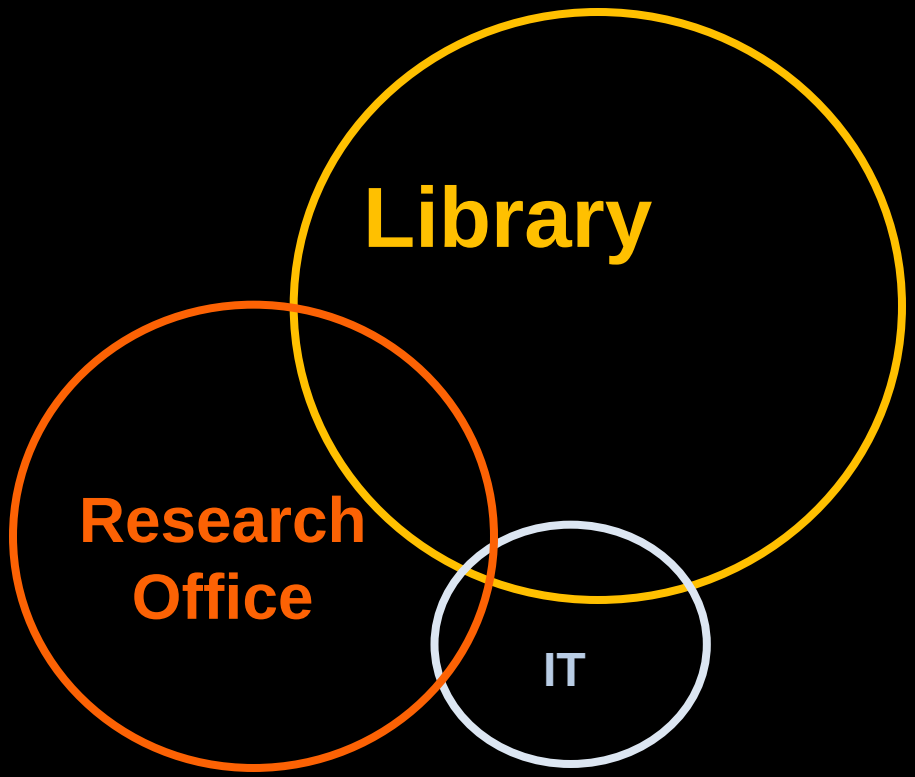
The DCC Curation Lifecycle Model



Some library roles

- Leadership – coordinate action
- Audit – who has what, where does it go?
- Advice on access – data, wherever it is
- Preservation – permanence
- Citability
- Data/publication linking
- Promoting data in teaching
- Selection
- Education – early career researchers

Who (in the UK) is addressing RDM ?



How?

- Create policy – collaborate with others
- Develop existing digital services
- Learn about audit tools (DCC & others)
- Learn about data & sources
- Reskill subject librarians
- Learn about your own data
- Bridge between publishers & researchers

Understanding Data Requirements

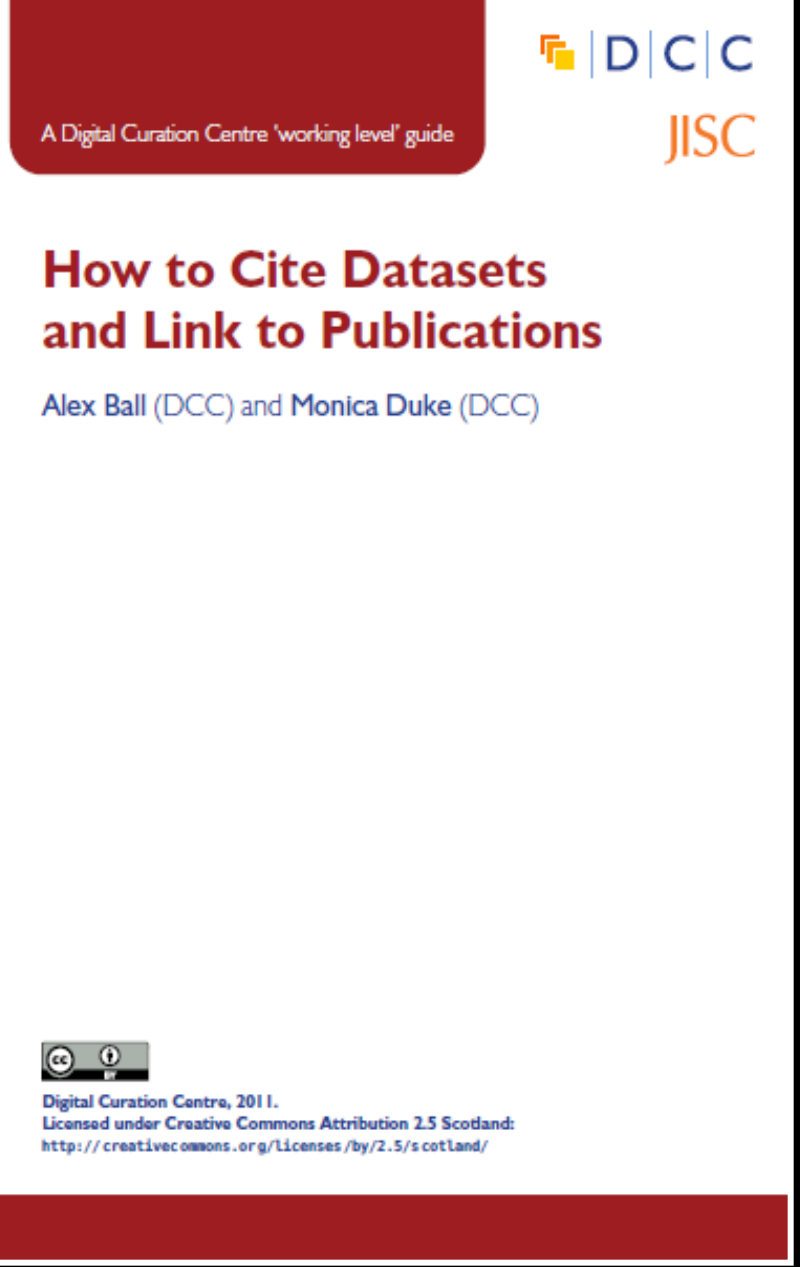
Data Asset Framework

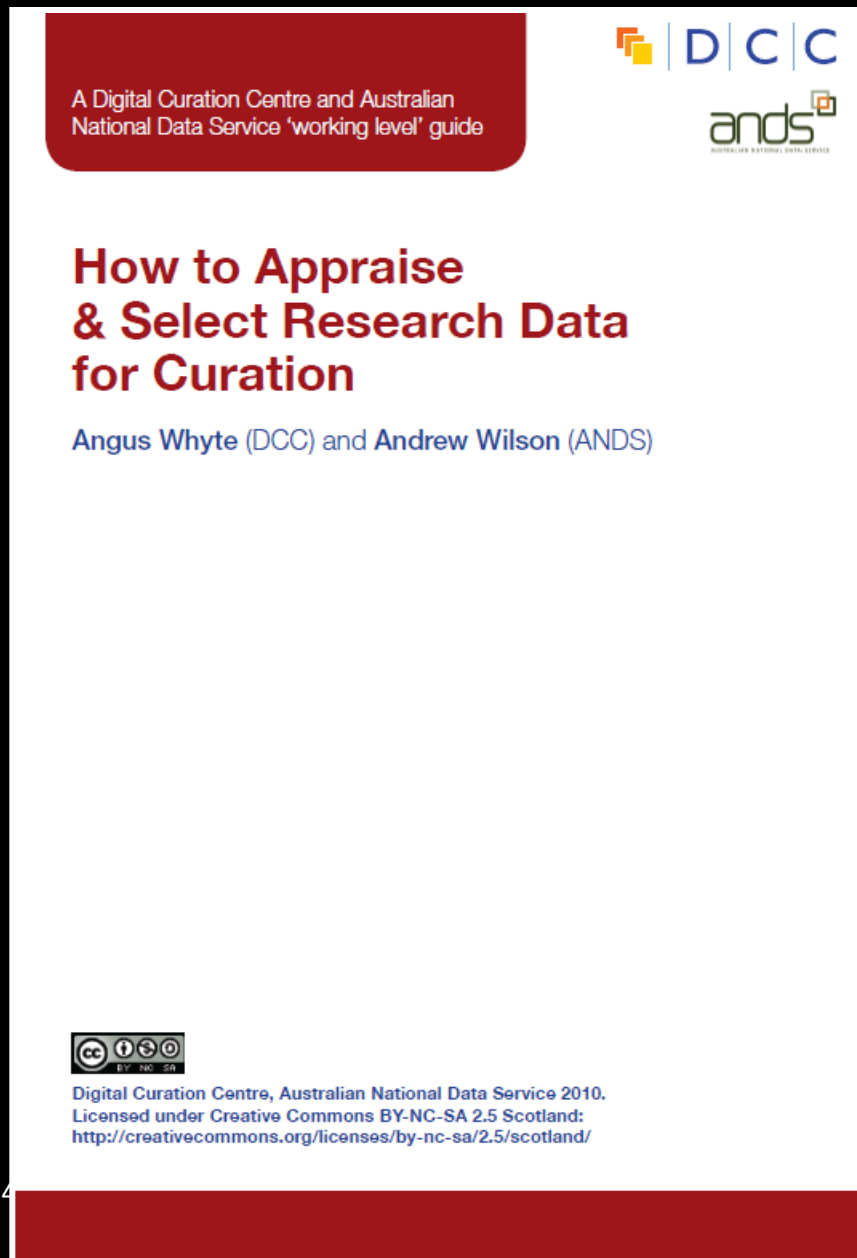
CARDIO
collaborative assessment of research data infrastructure and objectives

If research data lies at the heart of your organisation, you need to know that you have adequate infrastructure, staff skills and resources, and senior management support in place to ensure that your data is effectively managed for validation, reuse and evidential purposes.

CARDIO enables you to:

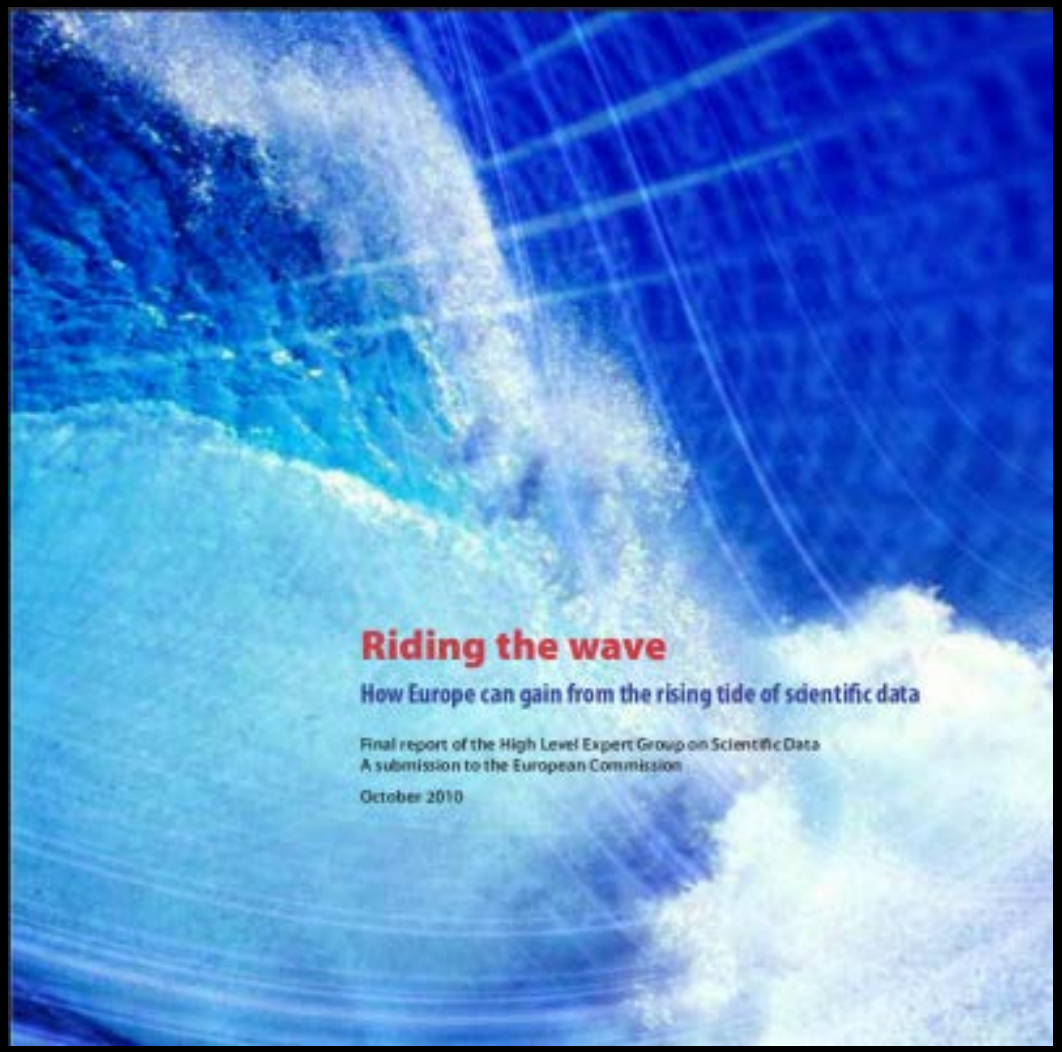
- collaboratively assess data management requirements, activity, and capacity at your institution
- build consensus between data creators, information managers and service providers
- identify practical goals for improvement in data management provision and support;
- identify operational inefficiencies and opportunities for cost saving;
- make a compelling case to senior managers for investment in data management support





How to cite data

The Data Deluge is upon us



Sensor's ability to produce data outstrips IT's ability to process it

Findable, citable data has value

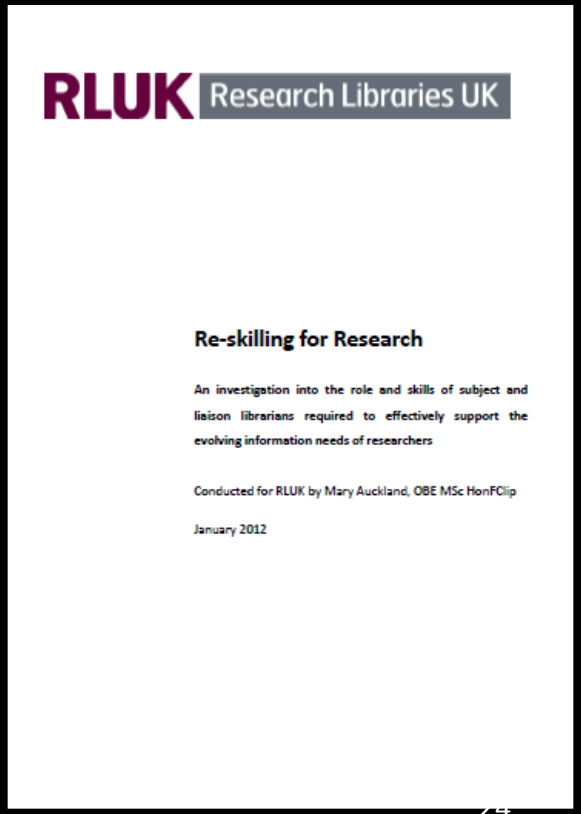
- Important to link publications to data (and vice versa)
- Increases citations – of data & publication
- Increases reuse (hence value)
- But effects exist even without publication, if data is:
  - Archived
  - Citable
  - Discoverable
- All benefit – researcher; institution; publisher

MORAL: build a data registry

How?

- Create policy – collaborate with others
- Develop existing digital services
- Learn about audit tools (DCC & others)
- Learn about data & sources
- Reskill subject librarians
- Learn about your own data
- Bridge between publishers & researchers

RLUK/Mary Auckland: Reskilling for Research 9 areas are skill gaps for subject librarians


RLUK Research Libraries UK

Re-skilling for Research

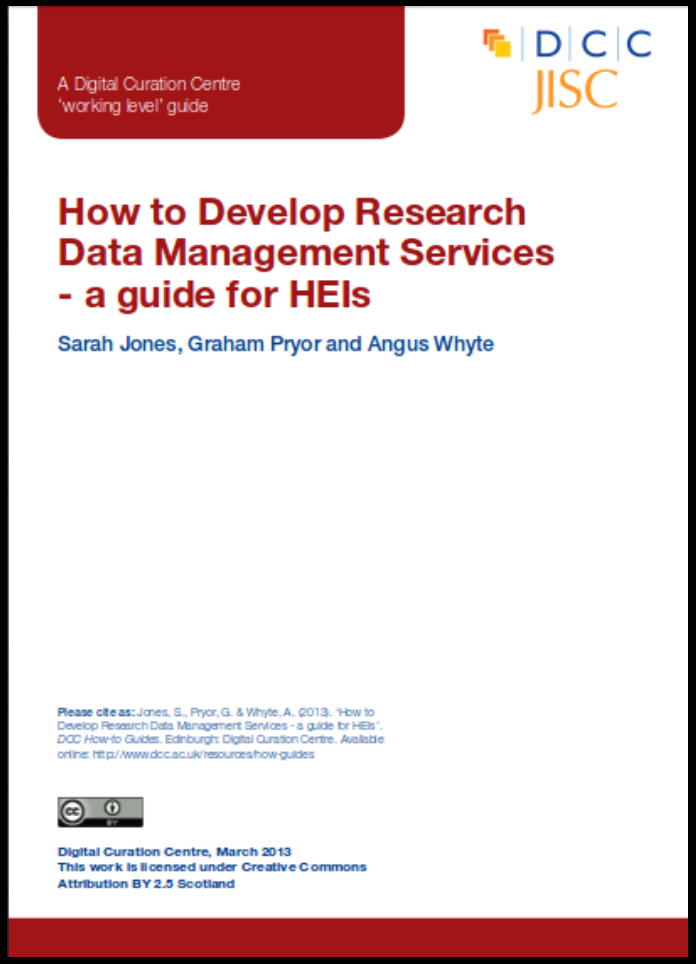
An investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers

Conducted for RLUK by Mary Auckland, OBE MSc HonFCilp

January 2012

Sheila Corrall: Libraries, Librarians and Data
Many action exemplars

Other helpful guides



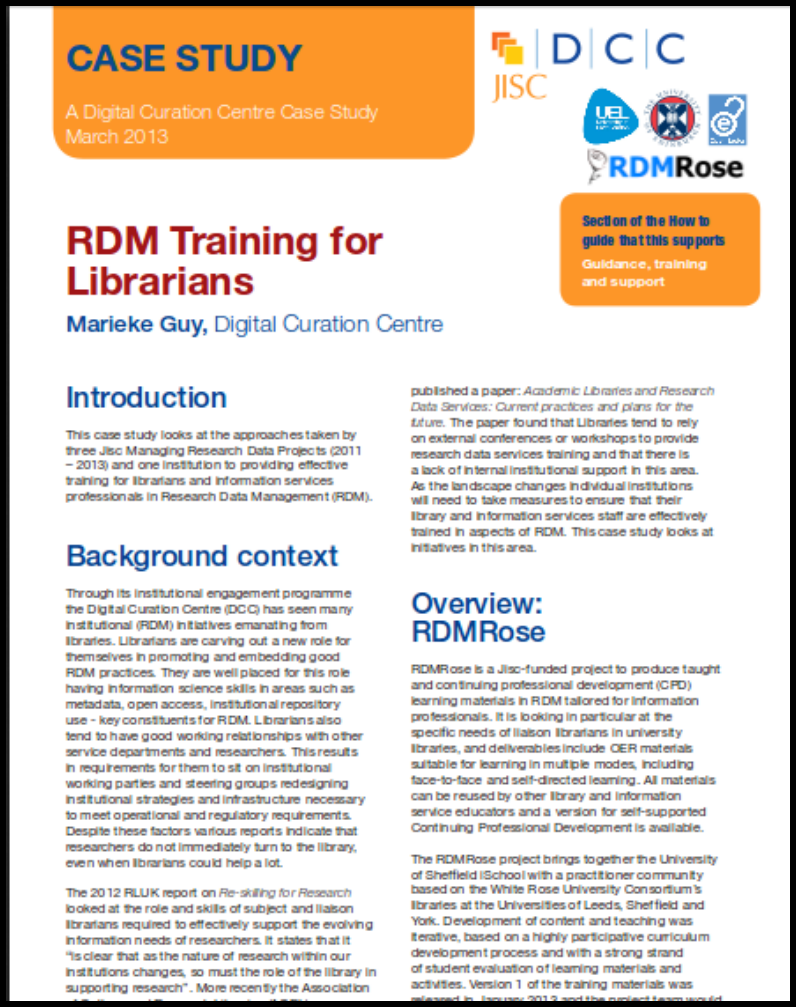
A Digital Curation Centre
'working level' guide

DCC JISC

How to Develop Research Data Management Services - a guide for HEIs

Sarah Jones, Graham Pryor and Angus Whyte

Please cite as: Jones, S., Pryor, G. & Whyte, A. (2013). 'How to Develop Research Data Management Services - a guide for HEIs'. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides



CASE STUDY

A Digital Curation Centre Case Study
March 2013

DCC JISC

RDMRose

Section of the How to guide that this supports
Guidance, training and support

RDM Training for Librarians

Marieke Guy, Digital Curation Centre

Introduction

This case study looks at the approaches taken by three Jisc Managing Research Data Projects (2011 – 2013) and one institution to providing effective training for librarians and information services professionals in Research Data Management (RDM).

Background context

Through its institutional engagement programme the Digital Curation Centre (DCC) has seen many institutional (RDM) initiatives emanating from libraries. Librarians are carving out a new role for themselves in promoting and embedding good RDM practices. They are well placed for this role having information science skills in areas such as metadata, open access, institutional repository use - key constituents for RDM. Librarians also tend to have good working relationships with other service departments and researchers. This results in requirements for them to sit on institutional working parties and steering groups redesigning institutional strategies and infrastructure necessary to meet operational and regulatory requirements. Despite these factors various reports indicate that researchers do not immediately turn to the library, even when librarians could help a lot.

The 2012 RLUK report on *Re-skilling for Research* looked at the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers. It states that it "is clear that as the nature of research within our institutions changes, so must the role of the library in supporting research". More recently the Association published a paper: *Academic Libraries and Research Data Services: Current practices and plans for the future*. The paper found that Libraries tend to rely on external conferences or workshops to provide research data services training and that there is a lack of internal institutional support in this area. As the landscape changes individual institutions will need to take measures to ensure that their library and information services staff are effectively trained in aspects of RDM. This case study looks at initiatives in this area.

Overview: RDMRose

RDMRose is a Jisc-funded project to produce taught and continuing professional development (CPD) learning materials in RDM tailored for information professionals. It is looking in particular at the specific needs of liaison librarians in university libraries, and deliverables include OER materials suitable for learning in multiple modes, including face-to-face and self-directed learning. All materials can be reused by other library and information service educators and a version for self-supported Continuing Professional Development is available.

The RDMRose project brings together the University of Sheffield iSchool with a practitioner community based on the White Rose University Consortium's libraries at the Universities of Leeds, Sheffield and York. Development of content and teaching was iterative, based on a highly participative curriculum development process and with a strong strand of student evaluation of learning materials and activities. Version 1 of the training materials was

How?

- Create policy – collaborate with others
- Develop existing digital services
- Learn about audit tools (DCC & others)
- Learn about data & sources
- Reskill subject librarians
- Learn about your own data
  - Help promote data literacy
- Bridge between publishers & researchers

DCC because good research needs good data

Search

Do you have 5 minutes to let us know what you think of this website? Take part in our survey

lome | Digital Curation | About Us | News | Events | Resources | Training | Projects | Community | Contact Us

ome > Projects > Incremental

Incremental

this section

rremental

sing the Digital Curation
p

RDIO

ARP

se Studies in the Life
ences

IM

IS

2

en Science Case Studies

The Incremental project is a collaboration between Cambridge University Library and HATII at the University of Glasgow. The project is about engaging with researchers to understand their concerns and needs with regards to data management, namely - how should you create data to make sure in can be found, accessed, understood and reused in the long-term - and

Incremental project

All UK HEIs must take steps to improve data curation activities. The Incremental project is working with researchers at the Universities of Cambridge and Glasgow to build their knowledge of data mangement, identify their requirements for support and fill any gaps. Findings will be shared to help support other HEIs.

Internet

http://www.flickr.com/photos/mattimattila/3003324844/

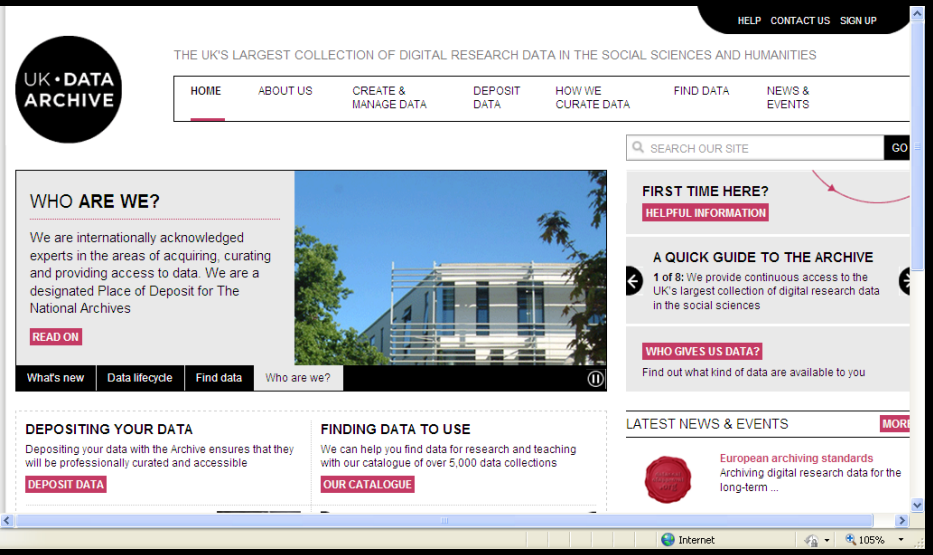
"Departments don't have guidelines or norms for personal back-up and researcher procedure, knowledge and diligence varies tremendously. Many have experienced moderate to catastrophic data loss"

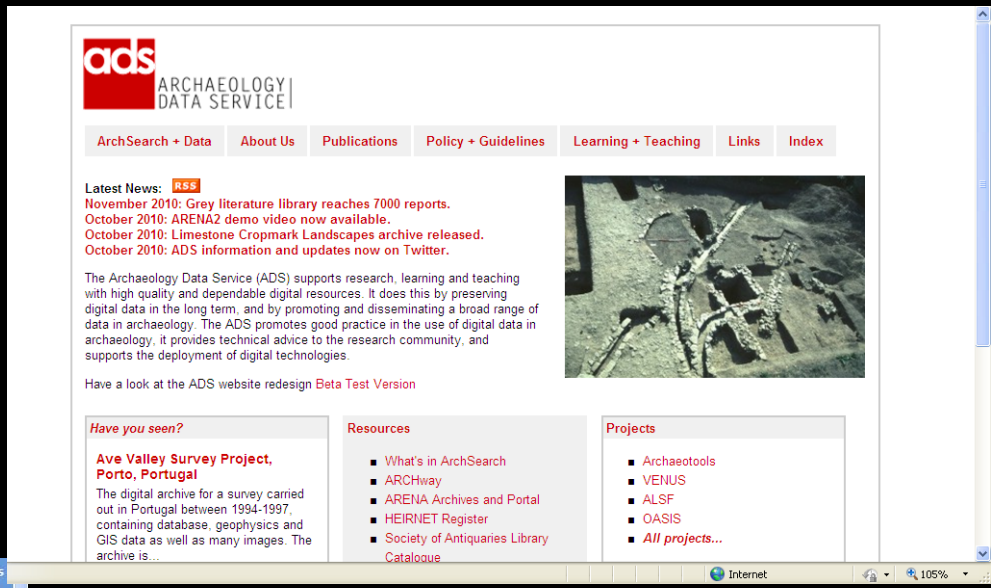
Incremental Project Report, June 2010

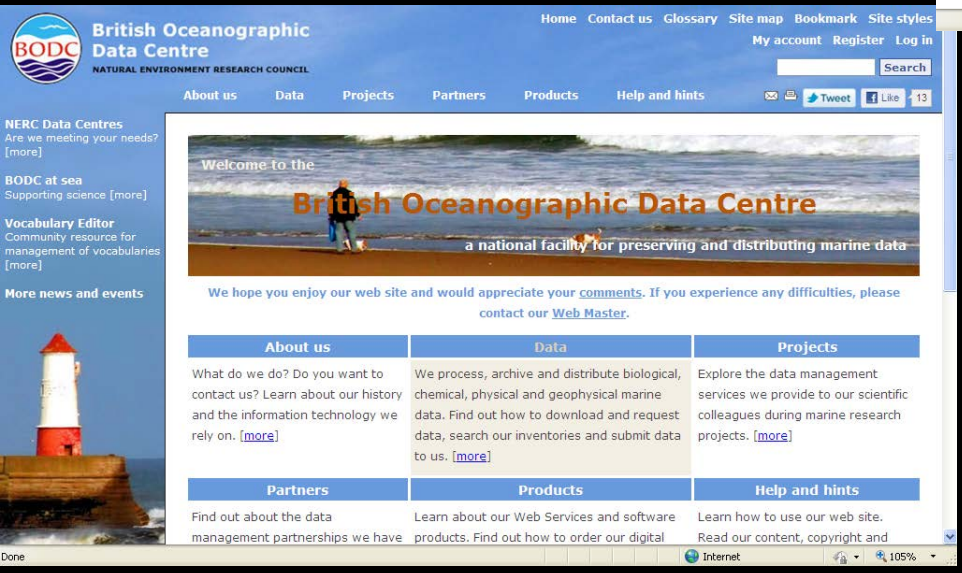
Excuses – and responses

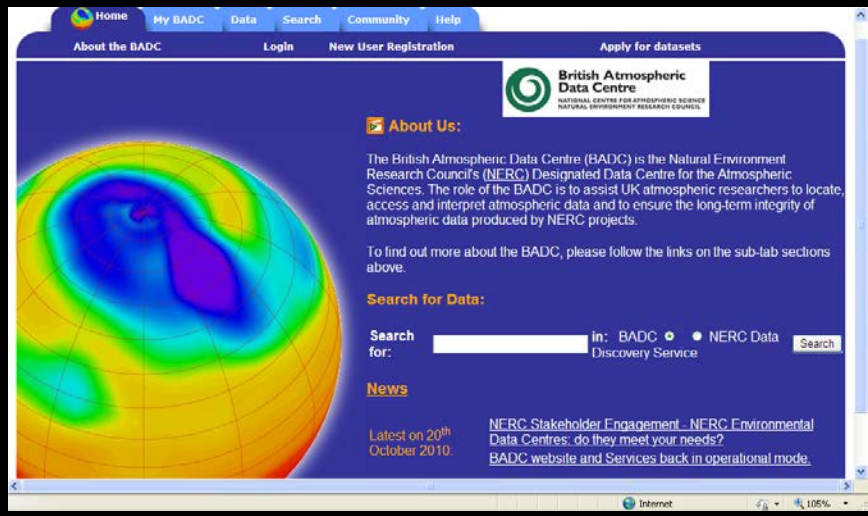
- "People will ask questions"
  - So use a data centre or repository
- "It will be misinterpreted"
  - Stuff happens. Also, openness encourages correction
- "It's not interesting"
  - Let others be the judge – your noise is my signal
- "I might get another paper out of it"
  - Up to a point. We might get more research out of it
- "I don't have permission"
  - A real problem. But solvable at senior level
- "It's too bad/complicated" –see above
- "It's not a priority"
  - Unfortunately, funders are making it so. But if you looked at the evidence, it would be your priority as well

See e.g. Carly Strasser's blog:
http://datapub.cdlib.org/2013/04/24/closed-data-excuses-excuses/


HELP CONTACT US SIGN UP
UK • DATA ARCHIVE
THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES
HOME
ABOUT US
CREATE & MANAGE DATA
DEPOSIT DATA
HOW WE CURATE DATA
FIND DATA
NEWS & EVENTS
SEARCH OUR SITE
GO
WHO ARE WE?
We are internationally acknowledged experts in the areas of acquiring, curating and providing access to data. We are a designated Place of Deposit for The National Archives
READ ON
What's new
Data lifecycle
Find data
Who are we?
FIRST TIME HERE?
HELPFUL INFORMATION
A QUICK GUIDE TO THE ARCHIVE
1 of 8: We provide continuous access to the UK's largest collection of digital research data in the social sciences
WHO GIVES US DATA?
Find out what kind of data are available to you
DEPOSITING YOUR DATA
Depositing your data with the Archive ensures that they will be professionally curated and accessible
DEPOSIT DATA
FINDING DATA TO USE
We can help you find data for research and teaching with our catalogue of over 5,000 data collections
OUR CATALOGUE
LATEST NEWS & EVENTS
European archiving standards
Archiving digital research data for the long-term ...
Internet
105%

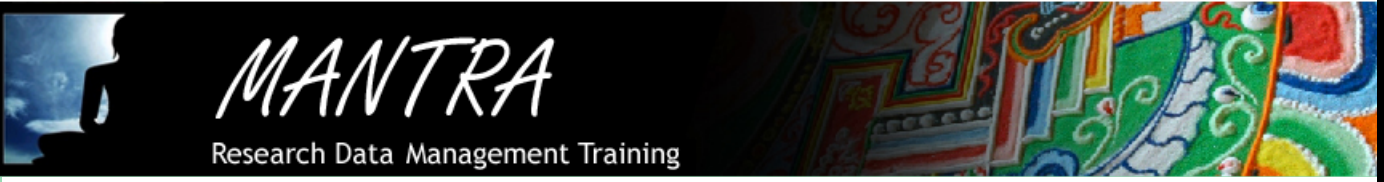

ads
ARCHAEOLOGY DATA SERVICE
ArchSearch + Data
About Us
Publications
Policy + Guidelines
Learning + Teaching
Links
Index
Latest News:
RSS
November 2010: Grey literature library reaches 7000 reports.
October 2010: ARENA2 demo video now available.
October 2010: Limestone Cropmark Landscapes archive released.
October 2010: ADS information and updates now on Twitter.
The Archaeology Data Service (ADS) supports research, learning and teaching with high quality and dependable digital resources. It does this by preserving digital data in the long term, and by promoting and disseminating a broad range of data in archaeology. The ADS promotes good practice in the use of digital data in archaeology, it provides technical advice to the research community, and supports the deployment of digital technologies.
Have a look at the ADS website redesign Beta Test Version
Have you seen?
Ave Valley Survey Project, Porto, Portugal
The digital archive for a survey carried out in Portugal between 1994-1997, containing database, geophysics and GIS data as well as many images. The archive is...
Resources
What's in ArchSearch
ARCHway
ARENA Archives and Portal
HEIRNET Register
Society of Antiquaries Library Catalogue
Projects
Archaeotools
VENUS
ALSF
OASIS
All projects...
Internet
105%


BODC
British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL
Home
Contact us
Glossary
Site map
Bookmark
Site styles
My account
Register
Log in
Search
About us
Data
Projects
Partners
Products
Help and hints
Tweet
Like
13
NERC Data Centres
Are we meeting your needs? [more]
BODC at sea
Supporting science [more]
Vocabulary Editor
Community resource for management of vocabularies [more]
More news and events
Welcome to the
British Oceanographic Data Centre
a national facility for preserving and distributing marine data
We hope you enjoy our web site and would appreciate your comments. If you experience any difficulties, please contact our Web Master.
About us
What do we do? Do you want to contact us? Learn about our history and the information technology we rely on. [more]
Data
We process, archive and distribute biological, chemical, physical and geophysical marine data. Find out how to download and request data, search our inventories and submit data to us. [more]
Projects
Explore the data management services we provide to our scientific colleagues during marine research projects. [more]
Partners
Find out about the data management partnerships we have
Products
Learn about our Web Services and software products. Find out how to order our digital
Help and hints
Learn how to use our web site. Read our content, copyright and
Done
Internet
105%


Home
My BADC
Data
Search
Community
Help
About the BADC
Login
New User Registration
Apply for datasets
British Atmospheric Data Centre
About Us:
The British Atmospheric Data Centre (BADC) is the Natural Environment Research Council's (NERC) Designated Data Centre for the Atmospheric Sciences. The role of the BADC is to assist UK atmospheric researchers to locate, access and interpret atmospheric data and to ensure the long-term integrity of atmospheric data produced by NERC projects.
To find out more about the BADC, please follow the links on the sub-tab sections above.
Search for Data:
Search for:
in: BADC NERC Data Discovery Service
Search
News
Latest on 20th October 2010:
NERC Stakeholder Engagement - NERC Environmental Data Centres: do they meet your needs?
BADC website and Services back in operational mode.
Internet
105%

What about collaboration?

- Collaborate within the university
- Collaborate with partners
- Collaborate with regional, national services
- Not everything can be done well locally
- Some examples…

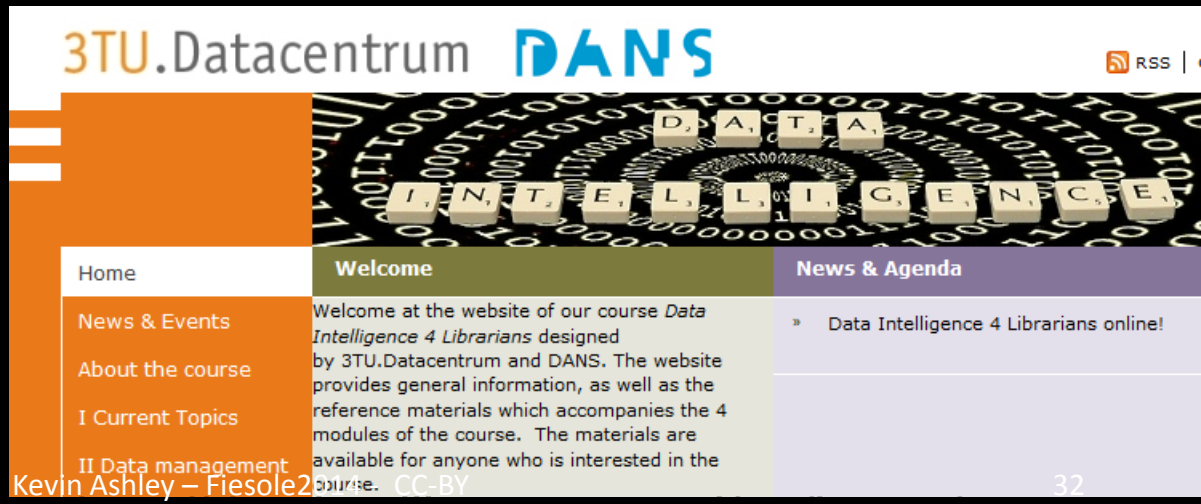


http://datalib.edina.ac.uk/mantra/libtraining.html

Do-It-Yourself Research Data Management Training Kit for Librarians

The University Of Sheffield. Information School

RDMRose



Choice of RDM training materials for librarians


Store your valuable data.
Show it to the world.
Share it with others.
http://datacentrum.3tu.nl

Data is variable

- Not always textual
- Not always tabular
- Not always fixed
- Not always clearly authored – think of archival provenance
- Not always associated with publication
- There’s more to the world than scholarly research
- Hidden data is wasted data

Closing thoughts

- Library roles: selecting content, protecting it, enabling and encouraging reuse
- All apply equally when the content is data
- Helping users find the most relevant content – much research data does not come from research
- Description can be very different where data is concerned

“Institutions will try to preserve the problem(s) to which they are the solution”

Clay Shirky

Questions

- How does data management align with institutional mission?
- When is library a coordinator, and when is it a service provider?
- What will you do alone, and what will you coordinate with others?
- What skills must you acquire?
- What do you want from national level?

Some conundrums

- Releasing genome data is OK when it's:
  - An identified human subject
  - An anonymous human subject
  - Your pet dog
  - Another mammal
  - An insect
  - A plant
  - A virus

notes

- No collections – show some data
- Not librarian etc
- Case for data – general, tamiflu
- ‘phd students don’t know what to do’
- Discovery, preservation, permanent ids, selection
- What does it cost/who pays
- Not a publication
- Research data as treasure
- Imperial example of similar time series
- Govt/academic data – the data librarian
- Libraries as providers of data
- Making better use of their own data