Facing the Data Challenge : Institutions, Disciplines, Services & Risks

Dr Liz Lyon, Director, UKOLN, University of Bath, UK Associate Director, UK Digital Curation Centre

1st DCC Regional Roadshow, Bath November 2010





This work is licensed under a Creative Commons Licence Attribution-ShareAlike 2.0 UKOLN is supported by: $\prod C$



www.ukoln.ac.uk

A centre of expertise in digital information management



Overview

- 1. Facing the data challenge : Requirements, Risks, Costs
- 2. Reviewing Data Support Services : Analysis, Assessment, Priorities 🥜
- 3. Building Capacity & Capability : Skills Audit
- 4. Developing a Strategic Plan : Actions and Timeframe



Facing the Data Challenge



http://www.flickr.com/photos/mintchocicecream/7491707/

Institutional Diversity



EMBEDDING INSTITUTIONAL DATA CURATION SERVICES IN RESEARCH (EIDCSR)





Based on DCC Curation Lifecycle Model



Patterns of information use and exchange: case studies of researchers in the life sciences



Disciplinary Diversity



HSITIA



http://www.flickr.com/photos/30435752@N08/2892112112/



DCSCARP Case studies

- Atmospheric data
 - Neuro-imaging
 - Tele-health
 - Architecture
 - Mouse Atlas



μ

DCC

Data Dimensions: Disciplinary Differences in Research Data Sharing, Reuse and Long term Viability

A comparative review based on sixteen case studies

Key Perspectives Ltd With a foreword and recommendations by Chris Rusbridge and Liz Lyon

DOCSOMER SYNCHESIS REPORT

Commissioned by the Digital Commission Canara. De Noe Gable 184.11 Version No. 1.0 Status FINAL Date 18 Jahwany 2010

Data Granabas: distiglinary Gfferences in research data sharing, reuse and long term viability

Recommendations:

• JISC

- HE & Research funders
- Publishers &
 Learned societies
- HEIs and research institutions
- Researchers & scholars

http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP%20SYNTHESIS.pdf



http://www.data-archive.ac.uk/media/203597/datamanagement_socialsciences.pdf



- Crystallography community engaged
- 'Embargo' facility
- Structured foundations
- Discoverable & harvestable

Data Curation Profiles

Data Curation Profiles

View Attachments (12) In	fo				
Added by Jake Carlson, last edited by	2		Typical File		Other/
Labels:	Data Stage	Output	Size	Format	Notes
Home Project Description Data	<u>.</u>				A file is spat out
A Data Curation Profile is a resource f	"Raw"	Floating point 3-D data	181 files X 100- 200 Mbieach	NetCDF	at each time step
by a single scientist or scholar, based	1	Cartesian	Smallerthan raw.<100 Mb		
data, disciplinary materials, and stand	"Interpolated"	interpolated data	each	Raw binary.dat	181 files
Data Curation Profiles are also intende	"Intern edicte"	Compressed binary ASCII form	10.100 Mb	VicED used form of	1 file per
Data Curation Profiles:	Internieulate	Ofraw data	10-100 Wib	VISSD .vou iornial	dataset
• Atmospheric Modeling - NEW	"Analyzed"/"Statistics"	summaryofdata	trivial	.txt	ile/simulation
<u>Carbonate Sedimentology</u> <u>Human Cell Defense Systems</u>	Overview	Web pages with graphics and stats	Around 140 Mb	.html	
Human Genomics Movement of Proteins - NEW Diagt Gasessian		Statistics across multiple		Excel	
Plant Genomics Plant Nutrition and Growth	Final	simulations	small	spreadsheet	
Soil Ecology - NEW Traffic Flow	Ancillary Data	1			
Water Flow and Quality	Text	n/a	5Kb	Fortran	"namelist" file
DUDDUE	Text	n/a	8Kb	(unspecified)	"sounding" file
UNIVERSITY	Image	1.lmages 2. An imations	1. Less than 1 Mb 2. 5-6 Mb	1.NCAR graphics metacode, Computer Graphics Metafile, Adobe Illustrator files.gif 2. an im ated gif/ Quick Time.mov	

Exercise 1a: Gathering requirements

- What are the researchers' data requirements?
- What datasets exist already?
- What are their data priorities?
- Research methodologies?
- Equipment and instrumentation? Formats?
- Where are the "pain points"?
- How will you find out? Approaches to use?
- How will you use the information?

- Standards?
- Skills?
- Plans?

Exercise 1b: Motivation, benefits, risks

- What are the RDM drivers and enablers for research staff and post-grad students?
- RDM drivers and enablers for Libraries / IT / Computing Services / Information Services?
- RDM drivers and enablers for the institution?
- What are the barriers? What are the risks?
- How will you articulate the benefits?
- How will you find out? Approaches to use?
- How will you use the information?

Exercise 1c: Costs & sustainability

- What are the costs associated with RDM?
- For the researcher?
- For the institution?
- Direct / indirect costs? Fixed / variable costs?
- What cost data already exists?
- What time horizon are you considering?
- How will you find out? Approaches to use?
- How will you use the information?

Requirements gathering: Approaches and tools

- Survey e.g. Oxford, Parse.Insight
- Focus groups : semi-structured interviews
- Case studies departmental / disciplinary
- Joint R&D projects
- Data champions in departments
- Data Preservation readiness : AIDA tool
- Data audit / assessment : DAF tool

Dealing with Data Report : Rec 4 Benefits:

Prioritisation of resources

Capacity development and planning

Efficiency savings – move data to more cost-effective storage

Manage risks associated with data loss

Realise value through improved access & re-use

Scale:

Departments, institutions









Implementation Guide

October 2009

Acknowledgement

This guide is based on the lessons learned to olign DAF offer orders and early exemplars. We revery grateful to those groups for sharing their experiences with rus to help refine the methodo bgy and assist future users. They were,

niversity Humanities Advanced Technology

of Glasgow & Information Institute

- Curve Etimetologia and Roipin Rice, university of Edinburgh
- Nell Jerrome and Jonatina n Breeze, Imperial College London
- Stephen Glace and Galetin Knight, King's College London
- Panay bia Polydo etou and Martin Moyle, u niversity College condon.
- Harry Globs and Teresal WcGowan, University of Southam monit
 Lub Martinez-University of Oxford
- Los viartines university of Conord
 Alex Ball, university of Barn [carn of DAF development ream].
- Sam Searle, Monash university [carl of ANDS project]

We related indepted to the JBC, which has supported to bresearch

• DAF Implementation Guide October 2009

 Collating lessons of pilot studies

 Practical examples of questionnaires and interview frameworks

• DAF online tool autumn 2010

http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf



Methodology

http://www.data-audit.eu/DAF_Methodology.pdf









The Innovative Design & Manufacturing Research Centre at the University of Bath





Science and Engineering at The University of Edinburgh

UNIVERSITY OF



School of GeoSciences



ecology environment geography geology geophysics meteorology oceanography Centre for Computing in the Humanities



JISC		http://sudamił	.oucs.ox.ac.u	ık/docs/Us	e%20of%20the%20DAF.pd
JISC Final Repo	ort				
Project					
	Pro jec t in form	nation	Г		
Project Acronym	UCL DAF				
Project Title	UCL Data Audit Framewo				
Start Cetx	18 August 2008				
Lead in this toton	UCL Library Seruices				1 Alian
Project 🗆 rector	Pavi Ayrk		S.		(ini)
Project Manager & contact detail ii	Martin Moyle m.m.oyle @ucl.ac.uk / D2D				
Partner in Lituiion L	By Terms of Triplemeintug collaborathig with DAF: HATH Uniters by officient (http://www.kathlanb DitalCuration Cen				OXFORD
Project Web URL	http://www.votac.vk/k/da				
Programme Name (and number)	Digital Repositores progr				
Programme Manager	Dr Ne I Jacobs				
					a data da
				Use of the Da	ata Audit Framework within the
Cloc u mo et					• •
Document		Sudamih Researcher Requ	irements Report	Sudamih Pro	ject
	Document		-		-
LOCUMENTINE	Hila I Report				
re porting Period Authority 28 protestation					
Autoron (nijas projektirole Fasta	Ранауюва Рокусствой, Р	Supporting Data Management Infrastruct	ure in the Humanities (Sudamih	Supporting Data Ma	anagement Infrastructure in the Humanities (Sudamih)
		sudamih.oucs.ox.ac.uk			- , , , ,
URL .	тар:мертитв:логао.лк/15			sudamih.oucs.ox.ac.uk	
ACCOLL		Authors Dr. James A. J. Wilson			
	Cocument (Dr. Meriel Patrick			
Version D	Ca te	Affiliation Oxford University Con	puting Services	Author	Dr. Meriel Patrick
Da Marci 2	2009 Deart to Proor			Affiliation	Ovford University Computing Services
				AnniauOn	Onlord Oniversity Computing Services
Parma I of 25					
Cocument Mar. J SC Final Repol Last updated. April 2009					
		1150			
	_				
		_		1150	
tn·//onri	into unl a	n 11k/15052/1/16	5053 ndf		
μ .// U	າແຈ.นษา.ส	1. un/ 10000/ 1/ 10	ιυυυ.ρυι	-	



Some lessons learned....

"CeRch had four false starts before finding a willing audit partner"
"Pick your moment" "Timing is key" (avoid exams, field trips, Boards...)

Plan well in advance!

"Be prepared to badger senior management"

Little documentation/knowledge of what exists:"a nightmare"

Defining the scope and granularity is crucial

Collect as much information as possible in interviews/surveys



Identifying risks

- Data loss (institution, research group, individual)
- Increased costs (lack of planning, service inefficency, data loss)
- Legal compliance (research funder, H&S, ethics, Fol)
- Reputation (institution, unit, individual)



Session 2009-10 Publications on the internet Science and Technology Committee Publications

House of Commons

Science and Technology Committee - Eighth Report

The disclosure of climate data from the Climatic Research Unit at the University of East Anglia

Freedom of Information FAQ (Draft)

FOI & RESEARCH DATA: RESEARCHERS' QUESTIONS AND ANSWERS

TABLE OF CONTENTS

SENERAL COMMENTS

Table of Contents

- 1. Introduction (0)
- 2. Q1 How do I recognise a FoI or EIR request? (1)
- 3. Q2 What's the short answer on what I should do if asked for data? (4)
- Q3 Why should I make my data available? (5)
- 5. Q4 How long have I got to respond to a request? (3)
- 6. Q5 I don't want to provide my data. What must I do first? (0)
- Q6 What are the most likely FoI exemptions for research data? (0)
- 8. Q7 If there's an exemption/exception can I ignore the request? (0)
- Q8 Some of the information may identify living individuals. Must I disclose it? (2)
- Q8a What does "redacted" or "redaction" mean? (0)
- Q8b Disclosing some information could put people at risk of harm. Must I disclose it? (0)
- 12. Q9 The data are covered by confidentiality agreements, must 1

Research data can be the subject of Freedom of Information requests, as recent high profile cases have shown. As a researcher, how should you respond if faced with such a request? This document sets out to answer this question and some others you may have. Details of particular circumstances can make a major difference, so conclusions reached in an individual case may well differ from those suggested here. This document does not constitute, and should not be construed as, legal advice.

Chris Rusbridge

Andrew Charlesworth



http://foiresearchdata.jiscpress.org/

Blue Ribbon Task Force on Sustainable Digital Preservation and Access



Sustainability: Who owns? Who benefits? Who selects? Who preserves? Who pays?

4 Domain areas



Keeping Research Data Safe2 Report: April 2010

	Benefits Tax	xonomy: Summary
Charles Beagrie	Dimension 1	
KEEPING RESEARCH DATA	Direct	Indirect (costs avoided)
SAFE 2	Dimension 2	
Neil Beagrie, Brian Lavoie and Matthew Woollard with contributions b∮ the Universities of Cambridge, Cxford, and Southampton, the Archaeolog∮ Data Service, OCLC Research, UK Data Archive, and Universit∮ of London Computer Centre.	Near-term	Long-term
Final Report - April 2010 Prepared by: Charles Reporte Limited	Dimension 3	
A study funded by	Private	Public
With support from OCLC Research and the UK Data Archive Cop/right HEFCE 2010 The authors have asserted their moral rights in this work		

Keeping Research Data Safe Factsheet

Cost issues in digital preservation of research data

This factsheet illustrates for institutions, researchers, and funders some of the key findings and recommendations from the JISC-funded Keeping Research Data Safe (KRD51) and Keeping Research Data Safe 2 (KRD52) projects. Further information on the research and findings can be found in the final reports.

What Costs Most?

Acquisition and ingest costs most. The costs of archival storage and preservation activities are consistently a very small proportion of the overall costs and significantly lower than the costs of acquisition/ingest or access activities for all our case studies. Note we believe early preservation action during ingest or pre-ingest produces lower costs over the lifecycle as a whole. (KRDS1, p.25; KRDS2, pp.31-52)

Activity Costs for the Archaeology Data Service		
Outreach/ Acquisition/	Archival Storage and	Access
Ingest	Preservation	
c. 55%	c. 15%	c. 31%

Impact of Fixed Costs

- The costs of long-term data curation/preservation are dominated by fixed costs that do not vary with the size of the collections;
- Staff are the major cost component overall and there is a minimum base-level of staff cover, skills and equipment required for any service;
- Activities characterised by significant fixed costs can reduce the perunit cost of long-term preservation by leveraging economies of scale.

(KRDS2, pp.32-34, 79-80)

Declining Costs over Time

We found a trend of relatively high preservation costs in the early years reducing substantially over time for data collections. An example is the preservation costs projected for the Archaeology Data Service (ADS) based on their experience of the first 10 years of operating the data service. (KRDS1, pp.4-6)



Costs for archival storage and preservation ("refreshment") decline to a minimal level over 20 years

Recommendation to

Recommendation to

From our research, it is likely

that the largest potential cost

efficiencies will come from

future tool development

supporting automation of ingest and access activities for curation and preservation.

Repositories should take advantage of

outsourcing as appropriate. Once core

capacity is in place additional content

can be added at increasing levels of

efficiency and lower cost. (KRDS1,

economies of scale, using multi-

institutional collaboration and

Funders

(KRDS2, p.83)

Recommendation to

Institutions

pp.77-78)

Funders and Institutions

The implications of these factors and projection for sustainability of data archives e.g. via archive charges to project budgets, are notable and worthy of more extensive study and testing. (KRDS1, pp.5-6)

- Which costs?
- Effect over time?
- Benefits taxonomy
- Repository models
- Case studies
- Key cost variables
- Recommendations

• User Guide, business templates forthcoming 2010

Reviewing Data Support Services Analysis, Assessment, Priorities







http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/publications.html#november-2009

Open Science at Web-Scale



JISC

Open Science at Web-Scale: Optimising Participation and Predictive Potential

Consultative Report

Document details

Author:	Dr Liz Lyon, UKOLN/ Digital Curation Centre, University of Bath
Date:	6th November 2009
Version:	V1.0
Document Name:	open-science-report-6nov09-final-sentojisc.doc
Notes:	Revised after comments from JISC and reviewers

- 1. Scale, Complexity, Predictive Potential
- 2. Continuum of Openness
- 3. Citizen Science
- 4. Credentials, Incentives, Rewards
- 5. Institutional Readiness & Response
- 6. Data Informatics Capacity & Capability



Exercise 2:

Analysis, Assessment, Priorities

- Institutional stakeholders?
- Data support services?
- Range, scope, coverage?
- Gaps?
- Fitness for purpose?
- Timeliness?
- Resources?
- Skills?
- SWOT



	Charles Beagrie
DIGIT	AL PRESERVATION POLICIES
	STUDY
	Part 1: Final Report October 2008
Neil Beagri	ie, Najla Semple, Peter Williams, and Richard Wright
	Prepared by
	Charles Beagne Umfed
	A study funded by
	JISC
	Copyright HEFCE 2008
n	he authors have associed their moral rights in this work

Report October 2008

Digital Preservation Policies Study

High-level pointers and guidance

Outline policy model/framework

Mappings to institutional strategies

Exemplars

State-of-the-Art Report : Models & Tools (Alex Ball, June 2010)

- Data Lifecycles
- Data Policies (UK) incl DMP
- Standards & tools
- Data Asset Framework (DAF)
- DANS Seal of Approval
- Preservation metadata
- Archive management tools
- Cost / benefit tools



REVIEW OF THE STATE OF THE ART OF THE DIGITAL CURATION OF RESEARCH DATA

ALEX BALL

erim 1 rep091 103 ab1 2.pdf

ISSUE DATE: 2nd June 2010



University's draft RDM policy



It shall be the University's policy that:

- Research data should be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
- **The University** should provide training, support and advice, as well as mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
- Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with PIs.
- All new research proposals must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
- Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
- The legitimate interests of the subjects of research data must be protected.
- Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, should be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository. Such research data deposited elsewhere should be registered with the University.

Jeff Haywood, RDMF V October 2010 http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF5/Haywood.pdf



High level statements from University's Research Computing Strategy - contains 'embryonic RDS Policy'

- Researchers will have access to world-class data services which will include storage, backup, sharing and access facilities to enable re-use, curation, and archive of data that they obtain through experimentation, observation and simulation or that is purchased or procured for use in research. It will be possible share data with groups both within and beyond the University.
- Researchers will have the skills and knowledge to make best use of the computational facilities available to them. Training will be available in order to ensure this is the case.
- Flexible and timely support will be available for all researchers to help them to make the most of these services.
- Research facilities will be available to collaborators from different institutions and to independent visiting scholars and will support mobile researchers.

Jeff Haywood, RDMF V October 2010

http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF5/Haywood.pdf

Research Data: Concerns, Challenges & Necessities in a Research-led University



Embryonic RDS Strategy

- 6. Recommendations from RDS WG (Draft)
- 1: Archiving of research data [= repository+?]
- 2: Accessibility of research data to all virtual collaborators, facilitating extra-institutional collaboration
- 3: Globally accessible cross-platform file store
- 4: Backup/synchronisation of data on mobile devices
- 5: Establishing networks of knowledge
- 6: Federated structure for local data storage

Jeff Haywood, RDMF V October 2010

http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF5/Haywood.pdf



Assessing cloud options

Flexible Services for the Support of Research



Repositories and the Cloud

Tuesday, February 23, 2010 from 10:00 AM - 4:30 PM (GMT)

London, United Kingdom

3 JISC Reports in 2010 :

- Technical Review
- Cloud computing for research
- Environmental & Organisational issues



Cloud Computing for Research

The Window Conference Centre, London, Tuesday 20 July 2010



```
October 27, 2009
        Triangle Universities, RENCI Launch Data
         Initiative

    North Carolina

                   nc
                                        universities
TUCASI
                                             Cyber-
The Research Triangle Park
                             infrastructure project

    Data cloud across

                                   three campuses

    "regional"
```

Locations

Policy & practice



Policy

2.3. Data Management and Sharing Planning

2.3.1. Responsibilities of ESRC grant applicants

18. Those ESRC grant applicants who plan to generate data are responsible for preparing and submitting data management and sharing plans for their research projects as an integral part of the application.⁸ It is expected that an outline data management and sharing plan will include the following points:



Press Release 10-077 Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans

Government-wide emphasis on community access to data supports substantive push toward more open sharing of research data

DCC cumition policy report

DCC

A report on the range of policies required for and related to digital curation

> Sarah Jones Digital Curation Centre, University of Glasgow



Data Sharing in the Biosciences

• The benefits of sharing data

 How data can be made available

May 10, 2010



- Data types, formats, standards, capture
- Ethics and Intellectual Property
- Access, sharing and re-use
- Short-term storage & data management
- Deposit & long-term preservation
- Adherence and review



DMP Online Currently updating Version 1.0

http://www.dcc.ac.uk/dmponline

Checklist for a Data Management Plan

Checklist questions mapped to funder's data requirements



Slide : Martin Donnelly, DCC

DCC

DMP Online v2.0 (coming soon)

O	OMP Online	D
	The http://depositive.hatil.ans.gla.ac.uk/section/68/3	ý *) - (₫¶ Coogie
OWP Ordere 0		Saves Dates as
	About DMP Online - Instructions for use - News - Document	tis + Privacy statement / Terms of use - Help
JVP online	Logged in as martin - My home - Admin - Logout - Front pa	ça
D.C.C.Data Management Planning Tool		
(add all DCC questions) (add another question)		
OCC Qualitiona	Your input	Guidance
3.1 Data sharing and re-use		
2.1.1 Will you amone the cada you capture or create?	😁 Yes 🔿 No	1929 galaxie or contracting
3.2 Access		
3.2 1 km est allen all po nake the one analogo 7 <mark>600</mark>	Pass will be particulable online and space to all Unoragonian the life of the property.	teag provid and to page in our in our information and a set of the memory of the set of
3.2 / White is the process for galling access to the data ³	Bullement, and upon to all.	Napa of anomany data holder discussions you is data savets angles of the first the measured scientification for a sample final approx
3.2.5 Will assess he chargestric? "Nice		

- Cleaner interface
- Funder-specific guidance
- Versioning feature
- CSV output

<u></u>	DMP Online	
) (〒) (http://dmptest.hatil.arts.gla.ac.uk/section/66/3	😭 🔪 - 🚷 Google 🔍 Q
DMP Online 😵 🗋	DMP Online 🖸 +	1
DCC C Data Management Planning Tool	Logged in as martin [Admin Privacy/Statement/Terms of I	Legeut Lee Heb ly home
ICC CLAUSES	YOUR INPUT	
Access, data sharing and re-use .1 Data sharing and re-use		
1.1 Will you share the data you capture or create?	⊛ Yes ⊙ No	
hid	Here you will want explain how the da	to
2.2 How and when will you make the data available?	Data will be available online and open to all throughout the life of the project. We be deposited throughout the life of the project. Also consider how potential users will find out about your data, e.g. will you publish details of y messarch, present a soort your finding promote your	will in a u od od st s
2.4 What is the process for gaining access to the	promote your meseric outputs on the second s	na

http://www.dcc.ac.uk/dmponline



DMPs next steps?

- Embed DMPs in funder policies & research lifecycles as the norm
- Code of Conduct for Research
- Assess & review DMPs (not just the science content of proposals)
- Educate reviewers (DCC guidance for social science in prep)
- Manage compliance of researchers
- Infrastructure to share DMPs
- Integrate in institution research management information system



Building a University Data registry...



Building Capacity & Capability



Data challenges?

- 1. Data management plans
- 2. Appraisal: selection criteria
- 3. Data retention and handover
- 4. Data documentation: metadata, schema, semantics
- 5. Data formats: applying standards
- 6. Instrumentation: proprietary formats
- 7. Data provenance: authenticity
- 8. Data citation & versions: persistent IDs
- 9. Data validation and reproducibility
- 10. Data access: embargo policy
- 11. Data licensing
- 12. Data linking: text, images, software









Exercise 3: Skills Audit

- What skills do you have in house?
- What are your strengths? Core data skills?
- Gaps? Do these matter?
- Can / should they be developed?
- How? Resource implications?
- Other sources of expertise?
- Key partnerships?
- Team science roles?

Skills Audit

Skill	Source / Gap	Comment

- Be specific
- Prioritise core skills



Data Access & Re-use

"Community Criteria for Interoperability"

(Scaling Up Report 2008)

- Domain data format standard: CIF
- Domain data validation standard: CheckCIF
- Metadata schema: eCrystals Application Profile http://www.ukoln.ac.uk/projects/ebank-uk/schemas/
- Crystallography Data Commons: TIDCC Data Model in development
- Domain identifier: International Chemical Identifier
- Citation & linking: DOI
 http://dx.doi.org/10.1594/ecrystals.chem.soton.ac.uk/145
- Embargo & Rights http://ecrystals.chem.soton.ac.uk/rights.html

D C C

A Digital Curation Centre 'working level' guide

How to License Research Data

Alex Ball (DCC)

DRAFT: 29 OCTOBER 2010



Digital Curation Centre, 2010. Licensed under Creative Commons BY-NC-SA 2.5 Scotland: http://creativecommons.org/Licenses/by-nc-sa/2.5/scotland/

Data Licensing

Bespoke licences

- Standard licences
- Multiple licensing
- Licence mechanisms
- Forthcoming 2010



A Digital Curation Centre and Australian National Data Service 'working level' guide

How to Appraise & Select Research Data for Curation

DCC

ands[®]

r.

Angus Whyte (DCC) and Andrew Wilson (ANDS)



Digital Curation Centre, Australian National Data Service 2010. Licensed under Creative Commons BY-NC-SA 2.5 Scotland: http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/

What to keep?

It is NERC's policy that:

 All applications for NERC funding must include an outline Data Management Plan, which must identify which of the data sets being produced are considered to be of

long term value, based on the criteria in NERC's Data Value Checklist. The funding application must also identify all resources needed to implement the Data Management Plan.



Policy-making for Research Data in Repositories: A Guide

http://www.disc-uk.org/docs/guide.pdf The Policy-making for Research Data in Repositories: Guide is intended to be used as a decision-making and The Policy-making for Research Data in Repositories: A and in Repositories in a decision-making and in the used as a decision-making in the set of the used as a decision resituries in a decision resituries in the set of the used as a decision resituries in the set of the used as a decision resituries in the set of the used as a decision resituries in the set of the used as a decision resituries in the set of the used as a decision resiture of the used as a decision resiture of the set of the used as a decision resiture of the set of the used as a decision resiture of the used as a decision resiture of the set of the used as a decision resiture of the set of the used as a decision resiture of the set of the Guide is intended to be used as a decision-making and institutions with digital repositories in any tool for institutions with at are considering planning tool for institution development that are considering Planning tool for institutions with digital repositorie existence or in development that are collections adding research data to their digital collections existence or in development that are considering adding research data to their digital collections. Supported by llS May 2009 Version 1.2 by **DISC-UK** Ann Green Jatas Stuart Macdonald Robin Rice project

Repositories

1. Content Coverage

- a. Scope: subjects and languages
- b. Kinds of research data
- c. Status of the research data
- d. Versions
- e. Data file formats
- f. Volume and size limitations

2. Metadata

- a. Access to metadata
- b. Reuse of metadata
- c. Metadata types and sources
- d. Metadata schemas

3. Submission of Data (Ingest)

- a. Eligible depositors
- b. Moderation by repository
- c. Data quality requirements
- d. Confidentiality and disclosure
- e. Embargo status
- f. Rights and ownership

4. Access and Reuse of Data

- a. Access to data objects
- b. Use and reuse of data objects
- c. Tracking users and use statistics

5. Preservation of Data

- a. Retention period
 - b. Functional preservation
 - c. File preservation
 - d. Fixity and authenticity

6. Withdrawal of Data and Succession Plans

Data Information Specialists Committee -UK

DRAMBORA interactive



Quality Assurance

Standards

Audit and certification tools

- TRAC
- DRAMBORA
- PLATTER
- NESTOR
- DANS Data Seal of Approval



Catalogue of Criteria for Trusted Digital Repositories

> Version 1 (draft for public comment)

Published by nestor Working Group Trusted Repositiories – Certification

nestor-materials 8



urn:nbn:de:0008-2006060703



Registry/Repository RRORI

http://www.loc.gov/standards/premis/

10



Training

- Consortial
- Institutional
- Departmental
- Laboratory
- Project
- Library

ISC DCCC Digital Curation 101 5 - 10 October 2008 National #-Sciences Editivity

Why do we need a Digital Combine 10

In recent years there have been a nondex of useful transag events desclud to the preservation of digital Information. However, there is a valit a need to provide precisit, hands on training in a life cycle approach to conceptualizing cruciting preserving socials to. and re-using data over time. The DCC Digital Conceptual DCC 101 Interactive cases will contribute towards bridging the gat that concerding values between a general analyse woll preservation and curstom teams arranget the information services and substitute concenses to and their addity to proceeding woldwas them in a working exervicement.

When should account

OCI01 is streed at supporting the core establish reads of adversation services staff and scientific researchers in Higher Education. Is the first instance, the workshap will focus on singuing.

- bench sciencesta
 Binary and sciencesta
- Brary and Information profil
 computing sciencists

A key goal will be the integration of these dispersite commentation of practices. Benefits of perceptions OC (10) will help indentiate to understand the relet() they play as the digital control (16-cycle over them. The workshop will explay a rate of leasance and exercises to priorite statisticit with hard a theorismic descention of digital cursation resum as well approach, hands on experiments in splaying the latence leasance discontions workshop will offer inscheres an opportunity to extended with other assessing measurement, insertational experiment, and practitioners across disciplenery and nutsional banachinas.

When and whend The first DCId1 workshop will take piece from Plankay 8th – Finday 10th Octobe at the Nettonal e Science Clearce in Ethologist.

For Arther details see http://www.doc.ac.ai/www.digital-curation.101.300 www.dcc.ac.uk "excellent : probably the best course I have been on since starting my role as an Informatics Liaison Officer"

DCC Digital Curation 101 Lite

Research Data Management Forum

http://www.dcc.ac.uk/data-forum/

- 1) Roles & Responsibilities
 - 2) Value & Benefits
- 3) Sensitive Data: Ethics, Security, Trust
- 4) Economics of Applying & Sustaining digital curation

- Computing Services
- Research staff / postdocs
- Postgraduate students

Research Data MANTRA project

About Research Data MANTRA

September 2010 - August 2011 funded by <u>JISC</u> Managing Research Data programme.

Research Data MANTRA aims to develop online learning materials which reflect best practice in research data management grounded in three disciplinary contexts: social science, clinical psychology, and geoscience.

The resulting materials will be

embedded in three participating postgraduate programmes and made available through the Transkills programme for use by all postgraduate and early career researchers as well as made available generally through an open license.

In addition to web-based 'chapters' that students can work through at their own pace, the course will include video interviews with leading academics about data management challenges, and practical exercises in handling data in four software analysis environments: PASW, also known as SPSS, NVivo, R and ArcGIS.





- Online resources
- Includes training for
- Data handling
- Software
- SPSS, NVIVO

D|C|C
 JISC



Links

- · CAIRO Blog
- o CAiRO Wiki
- o CAiRO Team
- JISCMRD Programme
- o Email Us

Project Summary

Project Name: Curating Artistic Research Output (CAiRO)

Project Hashtag: #croprj

Duration: 12 months (Aug 2010 -Aug 2011)

Project Partners: JISC Digital Media, the Digital Curation Centre and the University of Bristol Department of Drama: Theatre, Film and Television

Curating Artistic Research Output

Research data created by the UK's performance and v nature. This work may include interconnected multim dependant upon the choices made by a viewer.

In addition to enriching our cultural heritage sector, reinterpreted by subsequent researcher-practitioner: the use of such non-standardised research data can b



Television.

Our key output will be a post-graduate teaching and l effectively self-archive or to communicate their need:

- Live arts
- Department of Drama
- Researcherpractitioner focus





DCC
JISC

MChem or BSc in Chemistry with Informatics (Computational Drug Design)



Embedding data informatics education ...faculty & LIS...

DCC Digital Curation 101 Lite

EPSRC



GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE The iSchool at Illinois

University of Bath

- Doctoral Training Centre in Sustainable Chemical Technologies
- Industrial Doctorate Centre in Digital Media, Special Effects and Animation

Doctoral Training Centres

Master of Science: Specialization in Data Curation

Developing a Strategic Plan

Optimising organisational support

- Organisational structures
- Library / IT / IS / research support structure
- Where does data management fit?
- Leadership?
- Co-ordination?
- Roles : data librarian, data manager, research support officer, data scientist, data curator...
- New roles?

New data support structures



Exercise 4: Actions and Timeframe

- Vision and Objectives: Are they clear?
- Organisational structures: Fit for purpose?
- Library / IT / IS structure : Is it optimal?
- Roles : who is best placed to take action?
- Responsibility : for each service / activity?
- Priorities : what will you stop doing?
- Resources : Do you need to bid for funding?
- Partnerships : Who do you need to talk to?
- Plan: What? Who? How? When?

Actions and Timeframe

Short-term 0-12 months	Medium-term 12-36 months	Long-term >3 years

- Identify quick wins
- What can you do tomorrow?

Take homes

- Understand the research data requirements of your campus / institutional consumers
- 2. Agree research data service delivery priorities
- 3. Define data roles and responsibilities
- 4. Collaborate and strengthen the data support provided
- 5. Be pro-active! Engage! Be part of team science!





6th International Digital Curation Conference "Participation & Practice: growing the Curation Community through the data decade"

Chicago Mart Plaza, 6-8 December 2010