

Digital Curation for Modern Universities

Dr Liz Lyon,

Associate Director, UK Digital Curation Centre,
Director, UKOLN, University of Bath, UK

MURG Workshop, London, September 2012



This work is licensed under a Creative Commons Licence
Attribution-ShareAlike 2.0



UKOLN is supported by:



www.ukoln.ac.uk

A centre of expertise in digital information management



THE SUNDAY TIMES
UNIVERSITY OF THE YEAR 2011-12



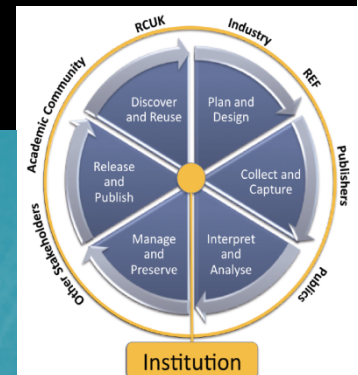
because good research needs good data

Running order.....

- Focus on institutional perspective
- RDM services the DCC can offer
- University of Bath experience

Research360

Managing data across the institutional research lifecycle





Royal Society Report

Science as an Open Enterprise

June 2012

10 Recommendations

Cites the DCC

Report sound-bytes

- “intelligently open data”
- “Scientists are increasingly turning to their university libraries and institutional repositories for support for their data.....”
- “familiarity with ... tools and principles of data management should be an integral part of the training of scientists in the future....”
- “The skills of data scientists are crucial in supporting the data management needs of researchers and of institutions.”



Research data management?

Getting attention...

1. Risk: where your data?



Photo credits: Harvey Rutt <http://www.ecs.soton.ac.uk/regenesis/pictures/>

2. Reputation : data access, FOI

University told to hand over tree ring data - April 15, 2010



theguardian

[News](#) | [Sport](#) | [Comment](#) | [Culture](#) | [Business](#) | [Money](#) | [Life & style](#)

[News](#) > [Society](#) > [Smoking](#)

Tobacco firm demands university's research on children and smoking

Stirling University fighting attempt by Philip Morris to gain access to research under freedom of information laws

[Severin Carrell](#), Scotland correspondent
[guardian.co.uk](#), Thursday 1 September 2011 15.02 BST
[Article history](#)



Philip Morris International, which makes Marlboro cigarettes, has asked for Stirling University's research on teenagers and smoking. Photograph: Paul Sakuma/AP

3. Quality: data gold standard

Data Replication & Reproducibility

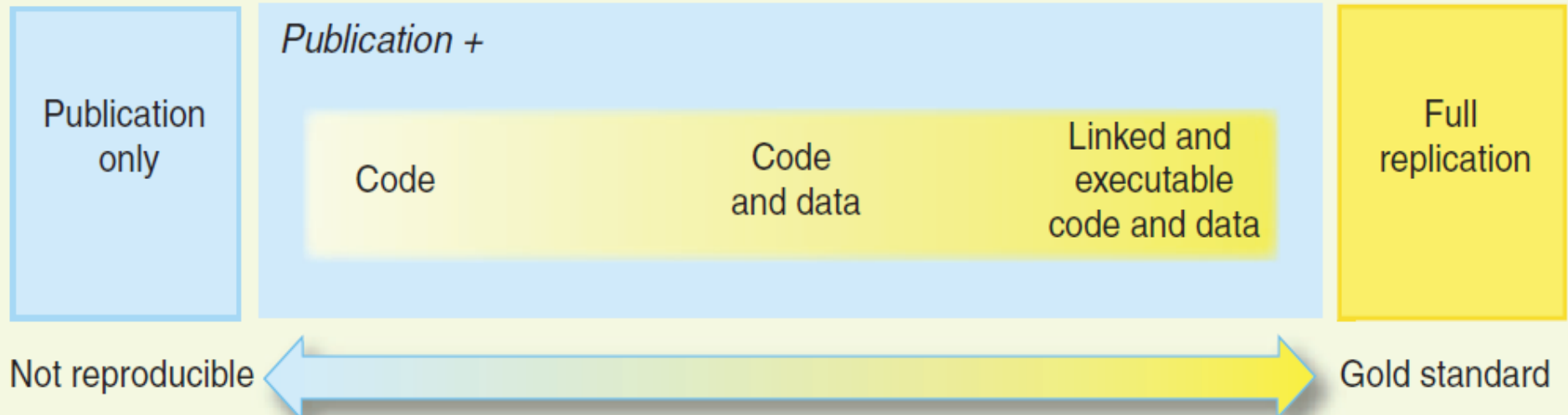
PERSPECTIVE

Reproducible Research in Computational Science

Roger D. Peng

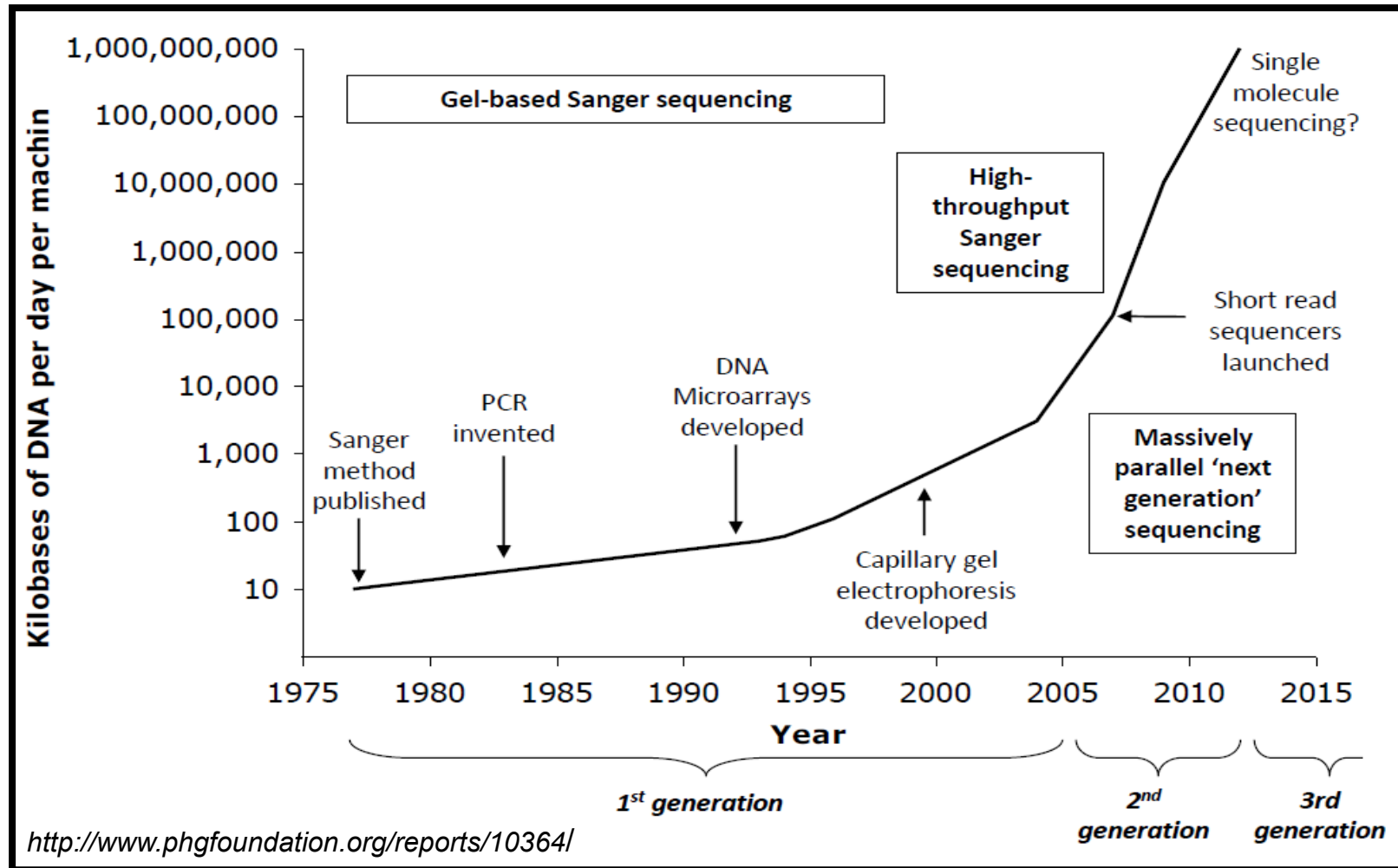


Reproducibility Spectrum



<http://www.sciencemag.org/content/334/6060/1226.full.html>

4. Scale: an explosion of data



"A single sequencer can now generate in a day what it took 10 years to collect for the Human Genome Project"

The New York Times 5.Partnerships

Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA

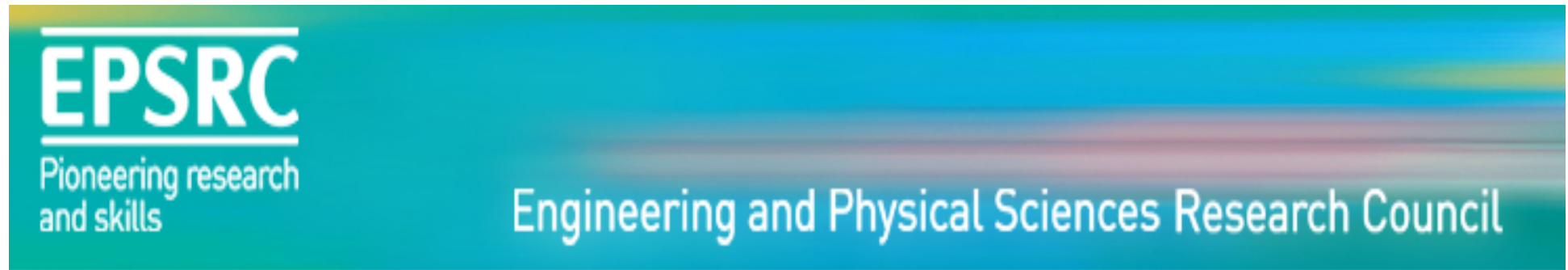
Published: August 12, 2010

Alzheimer's Disease Neuroimaging Initiative:
a unique (open) \$60M partnership between
NIH, FDA, universities and drug companies.

"It was unbelievable. Its not science the way most of us have practiced in our careers. But we all realised that we would never get biomarkers unless all of us parked our egos and intellectual property noses outside the door and agreed that all of our data would be public immediately."

Dr John Trojanowski, University of Pennsylvania

6. Funding



- EPSRC expects all those institutions it funds
- to develop a roadmap that aligns their policies and processes with EPSRC's **expectations** by **1st May 2012**;
 - to be fully compliant with these **expectations** by **1st May 2015**.

10 RDM questions to consider

1. What research data do we have?
2. Where is it stored?
3. What state is it in?
4. How long should we keep it?
5. How much will it cost ?
6. Do we have a data strategy?
7. Do we have a data policy?
8. Who has responsibility for RDM?
9. Are researchers and services staff
“data-aware” and “data-savvy”?

10. What should I do next?

Talk to the DCC:
*funded by the JISC to
provide resources and tools
to support institutional
research data management*



|D|C|C

because good research needs good data

Engaging with institutions

- Response to contacts & invitations
- > 20 live / pending
- Co-operating on:
 - Strategy
 - Policy
 - Roadmap development
 - Data assessments and audits
 - Data management plans

Some current IE activities

Assessing needs



Piloting tools e.g. DataFlow



RDM roadmaps



Policy development



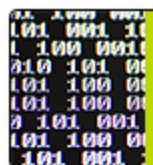
Policy implementation



10 RDM questions to consider

1. *What research data do we have?*
2. Where is it stored?
3. What state is it in?
4. How long should we keep it?
5. How much will it cost ?
6. Do we have a data strategy?
7. Do we have a data policy?
8. Who has responsibility?
9. Are research and services staff
“data-aware” and “data-savvy”?
10. What should I do next?

Data Requirements at your HEI?



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



<http://www.dcc.ac.uk/>



University of Bath Research Data Survey results



- 210 respondents (3.5% response rate) : PIs, ROs, postgrads
- ***Some preliminary findings:***
- Most have not had to produce a data management plan (81%)
- Much data is confidential, anonymised, under non-disclosure agreements, commercially sensitive, DPA, encrypted
- Data is also in non-digital form: lab notebooks, interviews
- Researchers store data on Univ Bath shared filestore 😊
- They also use Dropbox, USB sticks, home computers 😞
- Data loss: accidental deletion, hardware failure, obsolescence
- Open data is not the norm – often shared informally
- Lack of recognition for data sharing and reuse is an issue

- Research Council data requirements
- Institutional Roadmaps for EPSRC
- <http://www.bath.ac.uk/rdso/University-of-Bath-Roadmap-for-EPSRC.pdf>



University of Bath Roadmap for EPSRC

Compliance with Research Data Management Expectations

26th April 2012, Version 1.1

Authors: Dr Liz Lyon, UKOLN, & Dr Catherine Pink, UKOLN

Status: Submitted to Research Data Steering Group	5 th April 2012
Approved, with amendments, Research Data Steering Group	17 th April 2012
Submitted to Vice-Chancellor's Group (VCG)	23 rd April 2012
Submitted to VCG with revisions	30 th April 2012
Approved, with amendments, by VCG	30 th April 2012

Acknowledgement

We would like to acknowledge the leadership of Monash University in the area of research data management. The *Monash University Research Data Management Strategy and Strategic Plan 2012-2015*, released under a CC-BY licence, was highly influential in the development of this document.

Institution	Policy name	Date released
University of Oxford	Statement of commitment to Research Data Management (formal policy forthcoming through the DaMaRO project)	2010
University of Edinburgh	Research Data Management Policy	16 May 2011
University of Northampton	Research Data Policy	June 2011
University of Hertfordshire	Data Management Policy (see s.7 on research data and the appendix 'Guide to RDM')	1 Sept 2011
University of Warwick	Research Data Management Policy	7 Nov 2011
Glyndwr University	Policy on the Management of and Access to Research Data	20 December 2011
University of Southampton	Research Data Management Policy	February 2012
University of East London	Research Data Management Policy for UEL	15 March 2012
Brunel University	Research Data Management Vision	20 March 2012
Queen Mary, University of London	Research Data Management Policy	7 June 2012
University of Sheffield	Research Data Management Policy	July 2012

Draft policies

[University of Leeds](#) - via the RoaDMaP project
Timeline of developments including draft policy text

[University of Lincoln](#) - via the Orbital project
Blog post with link to the draft policy text

[University of Manchester](#) - via the MiSS project
'Towards a Research Data Management Policy' document outlining progress

[University of Exeter](#) - Open Access and Research Data Management Policy

[University of Exeter](#) - Open Access and Research Data Management Policy for PGR Students



Institutional data policy development

- Aspirational?
- Pragmatic?
- Emergent?
- High-level?
- With teeth?

Developing a research data policy: reflections from Bath

- Keep it as succinct as possible
- It does need to have teeth
- Consult widely – many stakeholders
- Essential to discuss with the legal office
 - Definitions of terms are helpful
 - Provide detail in supporting procedures
 - Research Data Steering Group approval
 - PVC Research lead




Data management plans

DMP*online*
The  D|C|C Data Management Planning Tool

Firefox File Edit View History Bookmarks Tools Window Help

DMP Online

http://dmponline.hatl.arts.gla.ac.uk/section/66/3

DMP*online*
The  D|C|C Data Management Planning Tool



About DMP Online · Instructions for use · News · Documents · Privacy statement / Terms of use · Help

Logged in as martin · My home · Admin · Logout · Front page


add all DCC questions add another question

DCC Questions Your input Guidance

3.1 Data sharing and re-use

3.1.1 Will you share the data you capture or create?   [UKDA guidance on data sharing](#)


3.2 Access

3.2.2 How and when will you make the data available? 

Data will be available online and open to all throughout the life of the project.


Here you will need to explain how the data will be shared e.g. will they be deposited in a data centre, will you forward copies on request to interested parties, etc.

Also consider how potential users will find out about your data e.g. will you publish details of your research, present at conferences, blog about your findings, promote your research outputs on a website? etc.

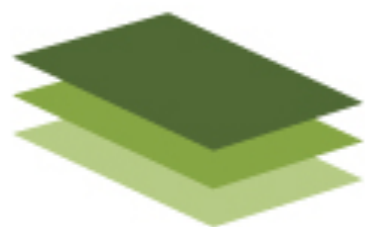
3.2.4 What is the process for gaining access to the data? 

Unlicensed, and open to all.

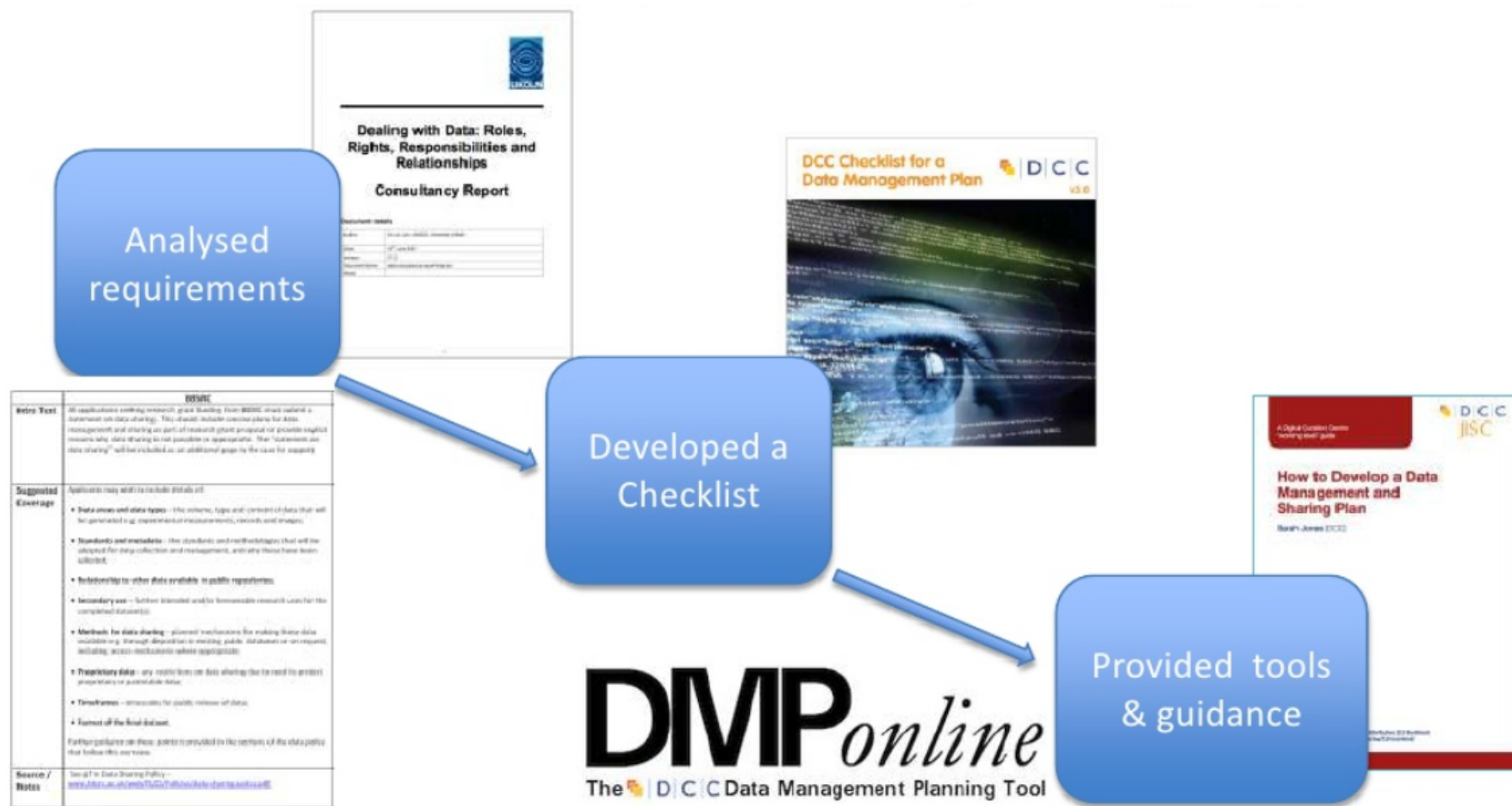
Ways of accessing data include: downloading from a data centre; requesting direct from the researcher; downloading from a simple Web page.

3.2.5 Will access be chargeable? 

Firefox Disabled



DMPTool
Guidance and Resources for your Data Management Plan



Links to all DMP resources via <http://www.dcc.ac.uk/resources/data-management-plans>

A Digital Curation Centre 'working level' guide



Some DCC guides

How to Cite Datasets and Link to Publications

Alex Ball (DCC) and Monica Duke (DCC)



Digital Curation Centre, 2011.
Licensed under Creative Commons Attribution 2.5 Scotland:
<http://creativecommons.org/licenses/by/2.5/scotland/>

<http://www.dcc.ac.uk/>

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



How to Appraise & Select Research Data for Curation

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/>

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/>

Digital Curation 101 training materials



If you are not able to attend DC 101 workshops and receive our training materials, why not buy a copy of training materials as a handy reference guide for your workplace?

Order materials

Doctoral Training
Centres
Research360 Project
at Bath

MANTRA
Research Data Management Training

Home | Software practicals | Project overview | University of Edinburgh guidance | Acknowledgements | Feedback

Online learning units

- Introduction to the course
- Research data explained
- Data management plans
- Organising data
- File formats & transformation
- Documentation & metadata
- Storage & security
- Data protection, rights &

Research Data MANTRA is a course designed for PhD students and others who are planning a research project using digital data.

The online learning units on the left cover a number of important topics. To begin, click on the first unit, [Introduction to the course](#) and page through.

After, work through one of our software-specific [data handling practicals](#).

Feel free to view our project website, guidance for research staff, find out who contributed content, or give us feedback on the course, using the menu along the top.

This course is an Open Educational Resource that may be freely used by anyone. It is available through an open license for re-using, rebranding, repurposing. Last updated: January, 2012

DCC resources

JISC projects

RDM Training



Study | Research | Faculties & Departments | Business | About

search site... go

Centre for Sustainable Chemical Technologies



CSCT

- News
- Events
- Research
- People
- Doctoral Training Centre

About the Centre

Established in 2008, the **Centre for Sustainable Chemical Technologies** brings together academic expertise from the University of Bath with international industrial, academic and stakeholder partners to carry out **research, training and outreach** in sustainable chemical technologies. In less than two years, we have attracted nearly **£20m in funding** for our activities and the centre has rapidly become an important hub for sustainable chemistry in the UK.

McKinsey Global Institute



May 2011

Big data: The next frontier
for innovation, competition,
and productivity



Implications of
“Big Data” and
data science for
organisations in
all sectors

Predicts a
shortage of
190,000
data scientists
by 2019

Institutional data scientist role

- **Co-ordination and Collaboration**
 - Liaison / subject librarians
 - Repository manager
 - IT/Computing Services
 - Research Support & Development Office
 - Doctoral Training Centres
 - Researchers
- **Advocacy**
- **Training**



Research360

Managing data across the institutional research lifecycle

*Liz Lyon, Informatics Transform,
IJDC Current Issue, 2012*

Upcoming DCC events

DCC Research Data Management Forum 9

Shaping the Infrastructure

Cambridge 14-15 November 2012

Thematic presentations and discussion;

Programme: http://www.dcc.ac.uk/sites/default/files/RDMF%20%20agenda_rev2.pdf

DCC Regional Roadshow

London LSE - Rosebery Hall

20 November 2012

Awareness raising; good practice exemplars, case studies

Programme: http://www.dcc.ac.uk/sites/default/files/LondonRoadshowAgenda2_Sept20.pdf





*Infrastructure, Intelligence, Innovation: driving
the Data Science agenda*
8th International Digital Curation Conference,
Amsterdam, 14-16 January 2013

Thank you.



DCC Resources can be downloaded from
<http://www.dcc.ac.uk>

Slides at
<http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/presentations.html>

Informatics Transform paper at
<http://www.ijdc.net/index.php/ijdc/article/view/210/279>

