Partnering for research data

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

2nd LIBER International Workshop on Digital Preservation, Florence, May 2012

UKOLN is supported by:

JISC

www.ukoln.ac.uk

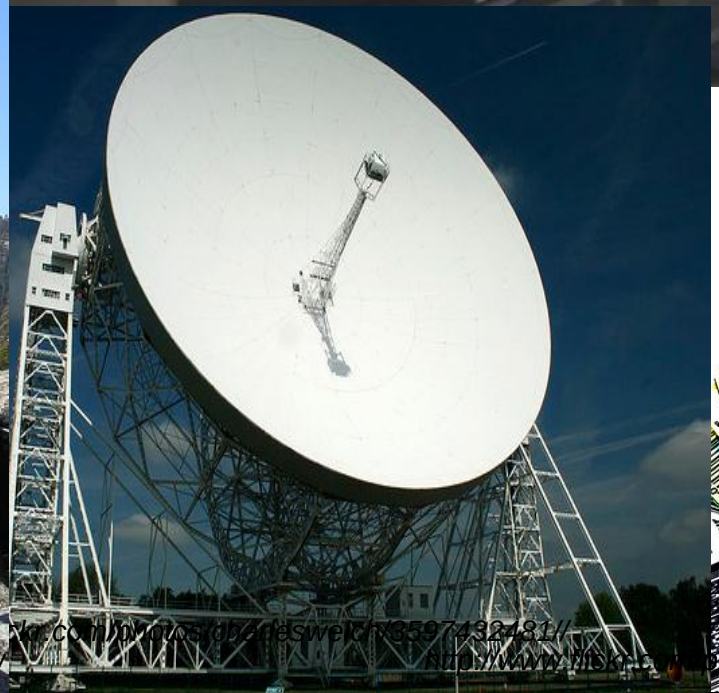
A centre of expertise in digital information management

Data…

http://www.flickr.c

http://www.google.co.uk/imgres?q=illumina+bgi&hl=en&client=firefox-a&hs=Jl2&rls=org.mozilla:en-GB:official&biw=1366&bih

http://www.flickr.com/photos/usfsregion5/4546851916/



http://www.flickr.com/photos/drondaswelch/3597432481/

http://www.flickr.com/photos/wasp_barcode/4793484478/

Research data stakeholders

- Who are they?
- Roles and responsibilities?
- Infrastructure requirements, tools, and skills?
- Gaps and opportunities?

Accademia Nazionale dei Lincei

Università degli Studi di Firenze

European Commission

elixir Data for Life

Partners: European....

DARIAH-EU

LIBER

OpenAIRE
Open Access Infrastructure for Research in Europe

ELSEVIER

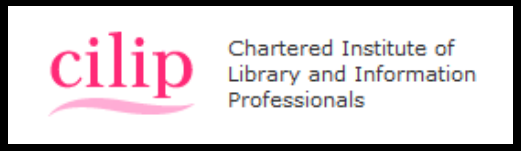
DANS

PANGAEA®
Data Publisher for Earth & Environmental Science

CERN

EMBL

flickr.com/photos/angeltree2008/3008028336/



Partners: global….

Institutional partners

Stakeholders within the University

- Roles (7 listed)
- Responsibilities
- Requirements
- Relationships

Role	Responsibilities	Requirements	Relationships
Director Information Services / CIO University Librarian	To lead and co-ordinate data informatics support	Appropriate LIS structure in place	PVC Research, Deans, Associate Deans, Faculty/School Directors of Research, IT Director, Director Research Support
		Library staff with data informatics & research data management skills	Other key institutional stakeholders
		Institutional repository with content links to underlying research data	Open Access Publishers
Data librarian / Data scientist / Liaison /Subject / Faculty Librarian	To deliver expert data informatics advice and guidance to research staff	Knowledge of data management planning and data audit and assessment tools	DTCs, post-grads, PIs
	To facilitate access to datasets for PIs, research staff, postgraduate and undergraduate students	Knowledge of selection and appraisal, metadata standards and schema, data formats, domain ontologies, identifiers, data citation, data licensing	DCC DataCite
		Knowledge of appropriate disciplinary data centres,	Data centre staff

Liz Lyon, Informatics Transform, IJDC Current Issue, 2012

1. Director IS/CIO/University Librarian
2. Data librarians /data scientist /liaison/subject/faculty librarians
3. Repository managers
4. IT/Computing Services
5. Research Support/Innovation Office
6. Doctoral Training Centres
7. PVC Research
8. *+ Public Engagement Office*

Liz Lyon, Informatics Transform, IJDC Current Issue, 2012

Data roles

Role	Responsibilities	Requirements	Relationships
Director Information Services / CIO University Librarian	To lead and co-ordinate data informatics support	Appropriate LIS structure in place	PVC Research, Deans, Associate Deans, Faculty/School Directors of Research, IT Director, Director Research Support
		Library staff with data informatics & research data management skills	Other key institutional stakeholders
		Institutional repository with content links to underlying research data	Open Access Publishers
Data librarian / Data scientist / Liaison /Subject / Faculty Librarian	To deliver expert data informatics advice and guidance to research staff	Knowledge of data management planning and data audit and assessment tools	DTCs, post-grads, PIs
	To facilitate access to datasets for PIs, research staff, postgraduate and undergraduate students	Knowledge of selection and appraisal, metadata standards and schema, data formats, domain ontologies, identifiers, data citation, data licensing	DCC DataCite
		Knowledge of appropriate disciplinary data centres,	Data centre staff

Leadership

Advocacy

Full mapping : Informatics Transform, IJDC Current issue, 2012

Role	Responsibilities	Requirements	Relationships
Repository managers	To ensure research papers have persistent links to underlying research data	Knowledge of persistent identification mechanisms and publisher requirements	Data librarians / Data scientists / Liaison /Subject / Faculty Librarians
IT / Computing Services	To provide data storage infrastructure and guidance	Knowledge of data storage options including cloud-based services	EduServ data centre. Cloud service providers National data centres
Research & Development Support Office / Research & Innovation Services	To provide RIM/CRIS capability for research outputs	Provision for non-textual outputs such as datasets, software and program code, gene sequences, models	Research funding bodies Data scientists / Liaison /Subject / Faculty Librarians

Discovery

Storage

CRIS

Full mapping : Informatics Transform, IJDC Current issue, 2012

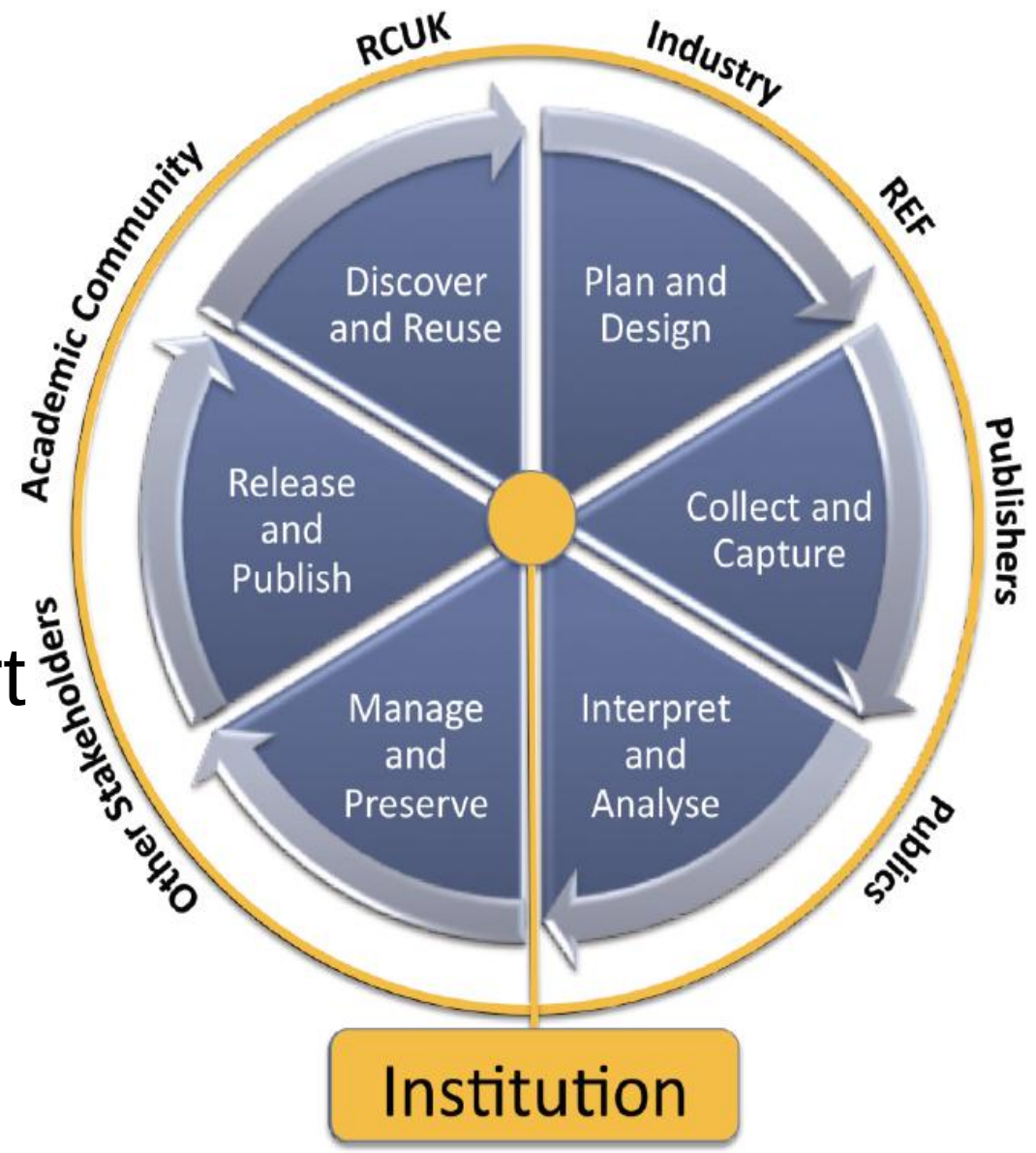
Role	Responsibilities	Requirements	Relationships
Faculty Doctoral Training Centres	To supply training to new-entrant researchers and PIs Training	Knowledge of data management planning and data audit and assessment tools Training programmes and modules	Deans & Associate Deans, PIs Data librarian / Data scientist / Liaison / Subject / Faculty Librarians
PVC Research	To develop institutional research policy and code of practice Policy	Understanding of data management compliance implications, risks including legal and ethical issues, and sustainability challenges	Deans & Associate Deans Key service directors Research & Development Support Office / Research & Innovation Services
Public Engagement Unit	To facilitate citizen participation in the research process Participation	Understanding of open science methodologies and infrastructure	PVC Research Director, Communications Deans & Associate Deans, PIs The Media

Full mapping : Informatics Transform, IJDC Current issue, 2012

Research360@Bath

JISC

- **Partnership approach**
  - UKOLN-DCC
  - Library
  - IT Services
  - Research Support Office
  - Doctoral Training Centres
  - Data Scientist



http://blogs.bath.ac.uk/research360/

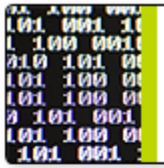
Infrastructure requirements: Delivering RDM Services

(Providing tools & support)

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

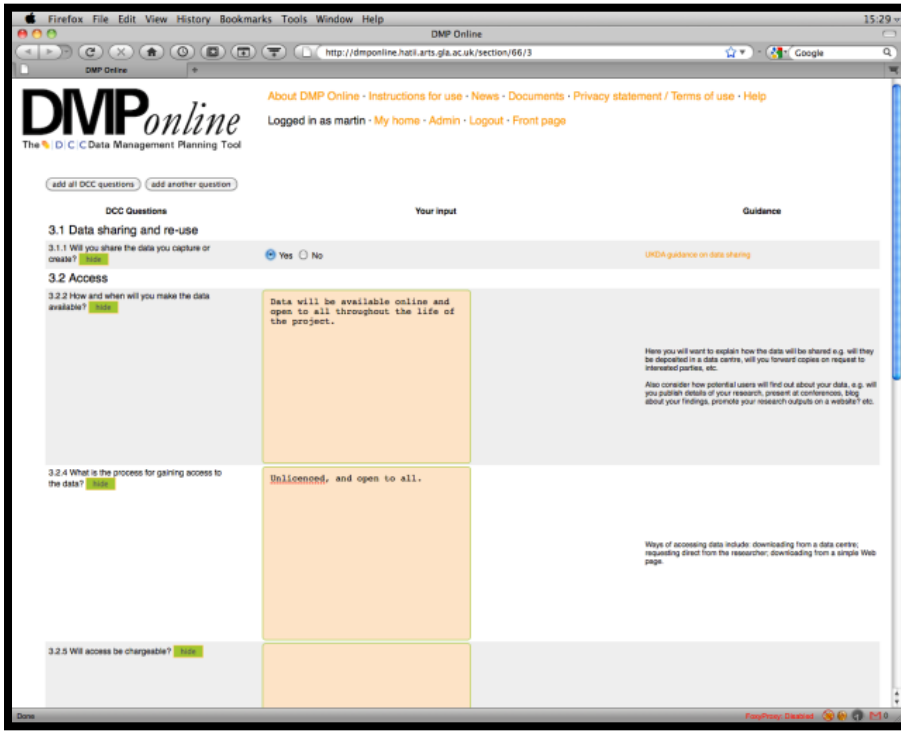
DCC

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

Data management plans

DMPonline

The DCC Data Management Planning Tool

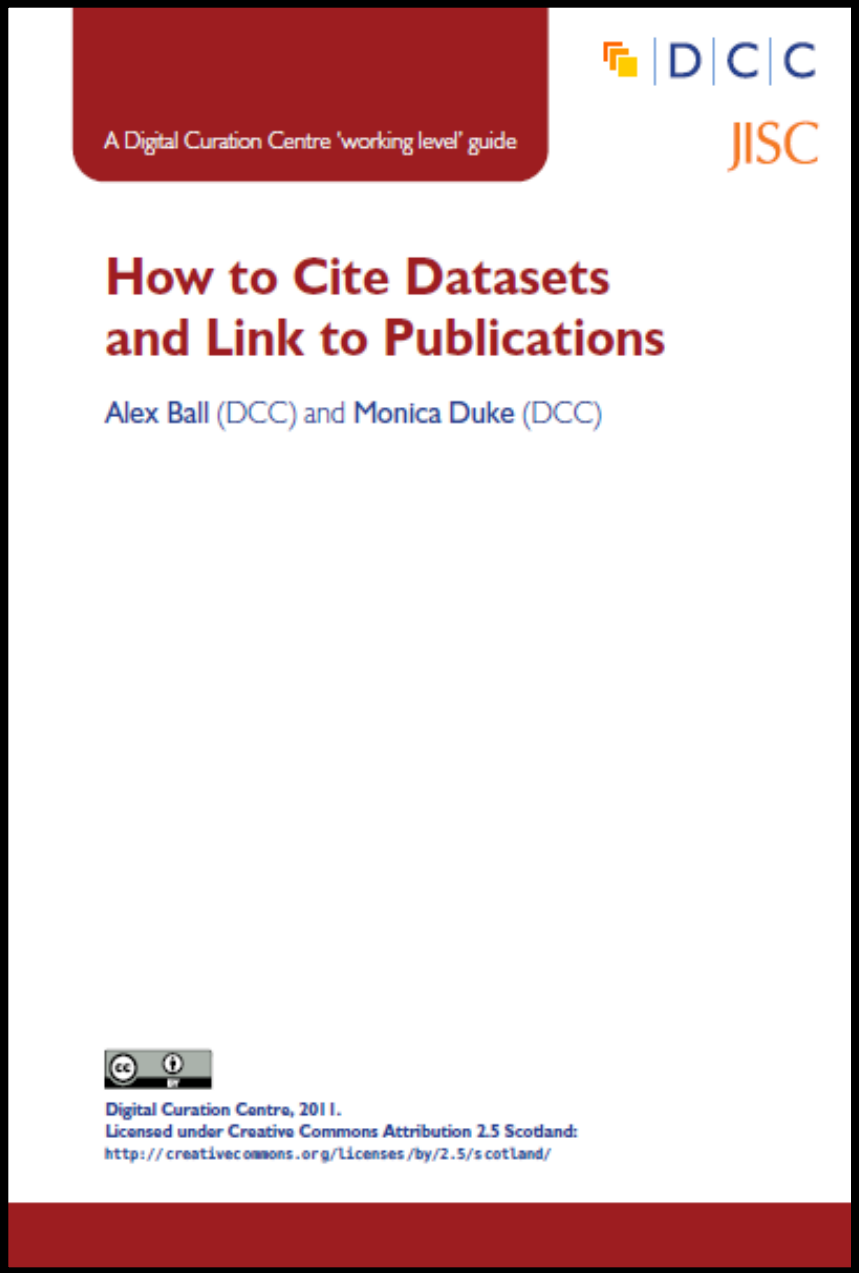


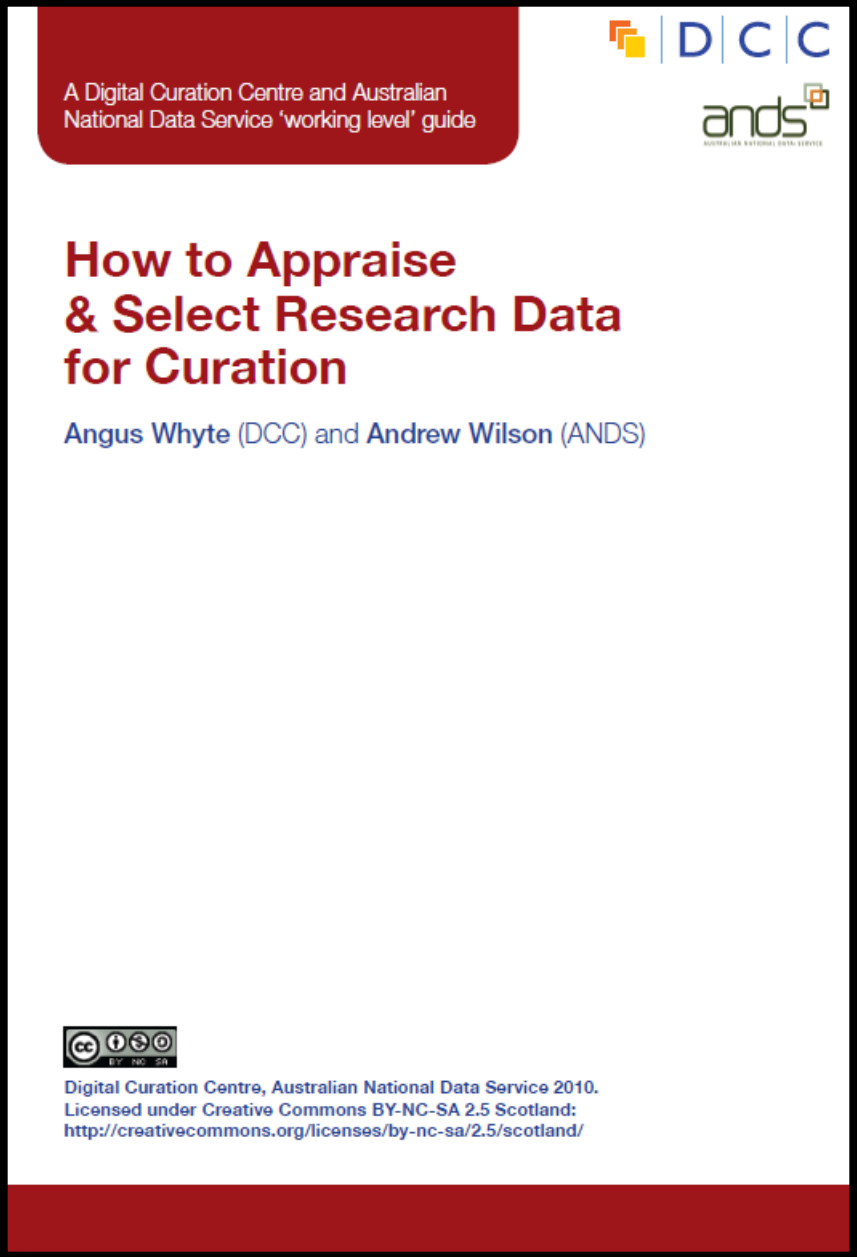
DMPTool

Guidance and Resources for your Data Management Plan

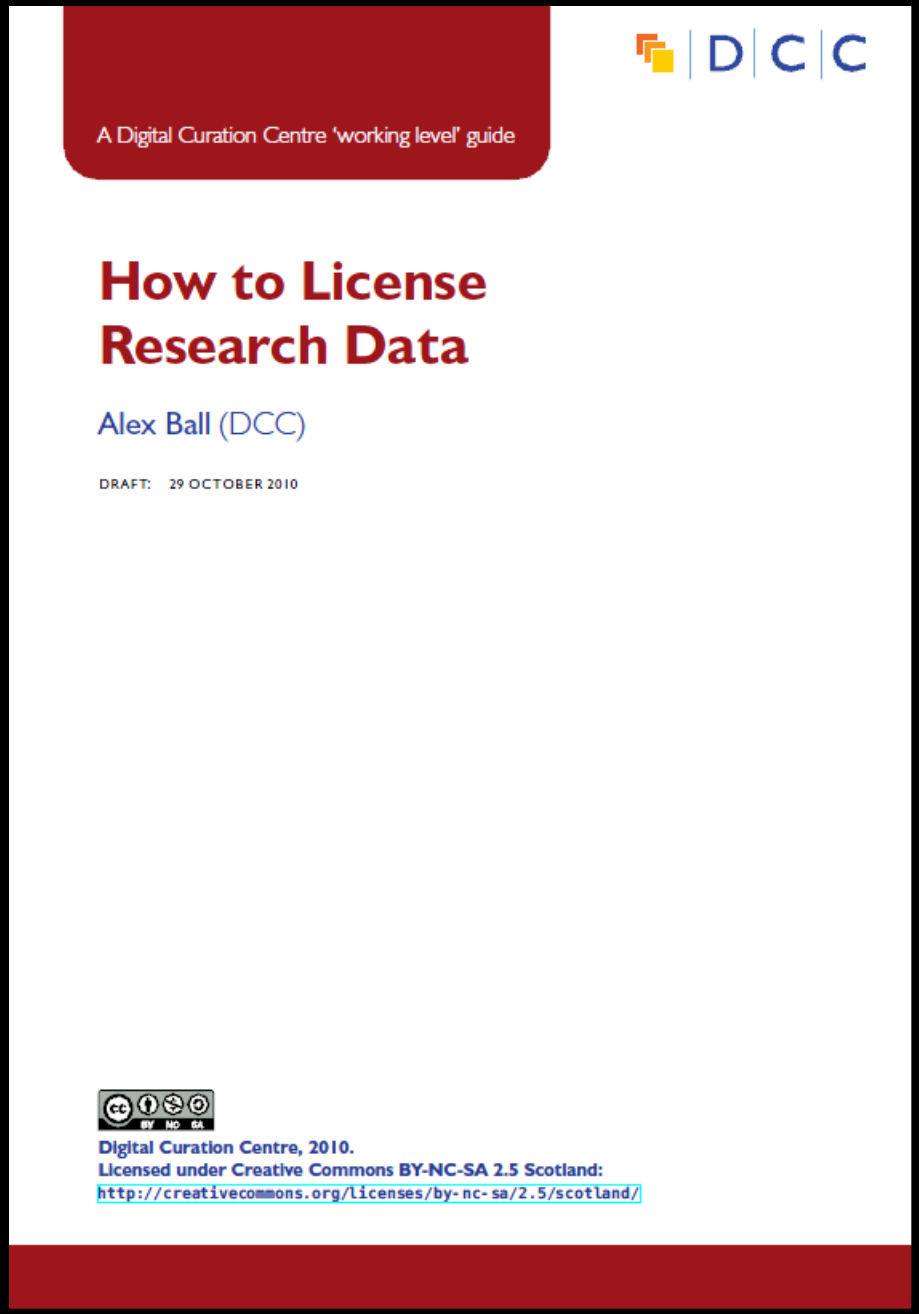
- Advocacy & Training
  - **Informatics:** disciplinary metadata schema, standards, formats, identifiers, ontologies
  - **Storage:** file-store, cloud, data centres, funder policy
  - **Access:** embargoes, FOI

What data to keep





How to cite data



Data Licensing

Bespoke licences
Standard licences
Multiple licensing
Licence mechanisms

science commons

Tools to track impact

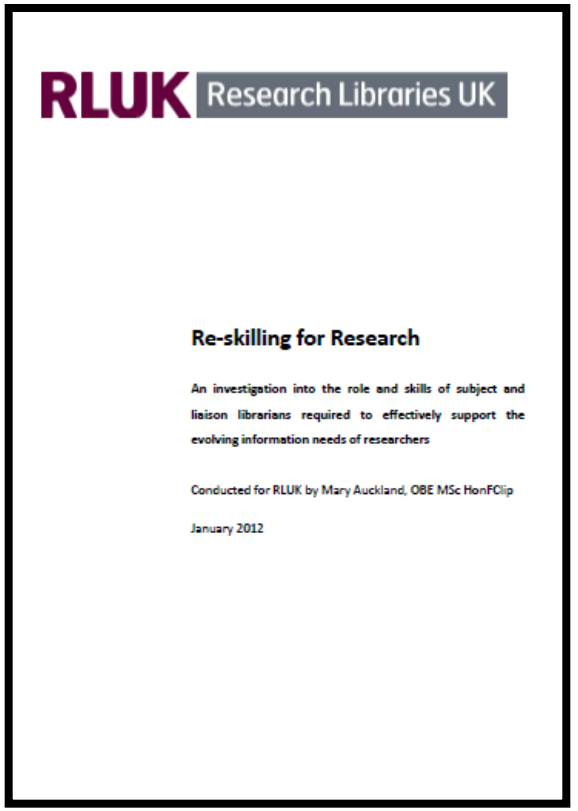
http://total-impact.org/

Data informatics capacity & capability

(Acquiring the skills….)

DCC

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

2012: Libraries in review

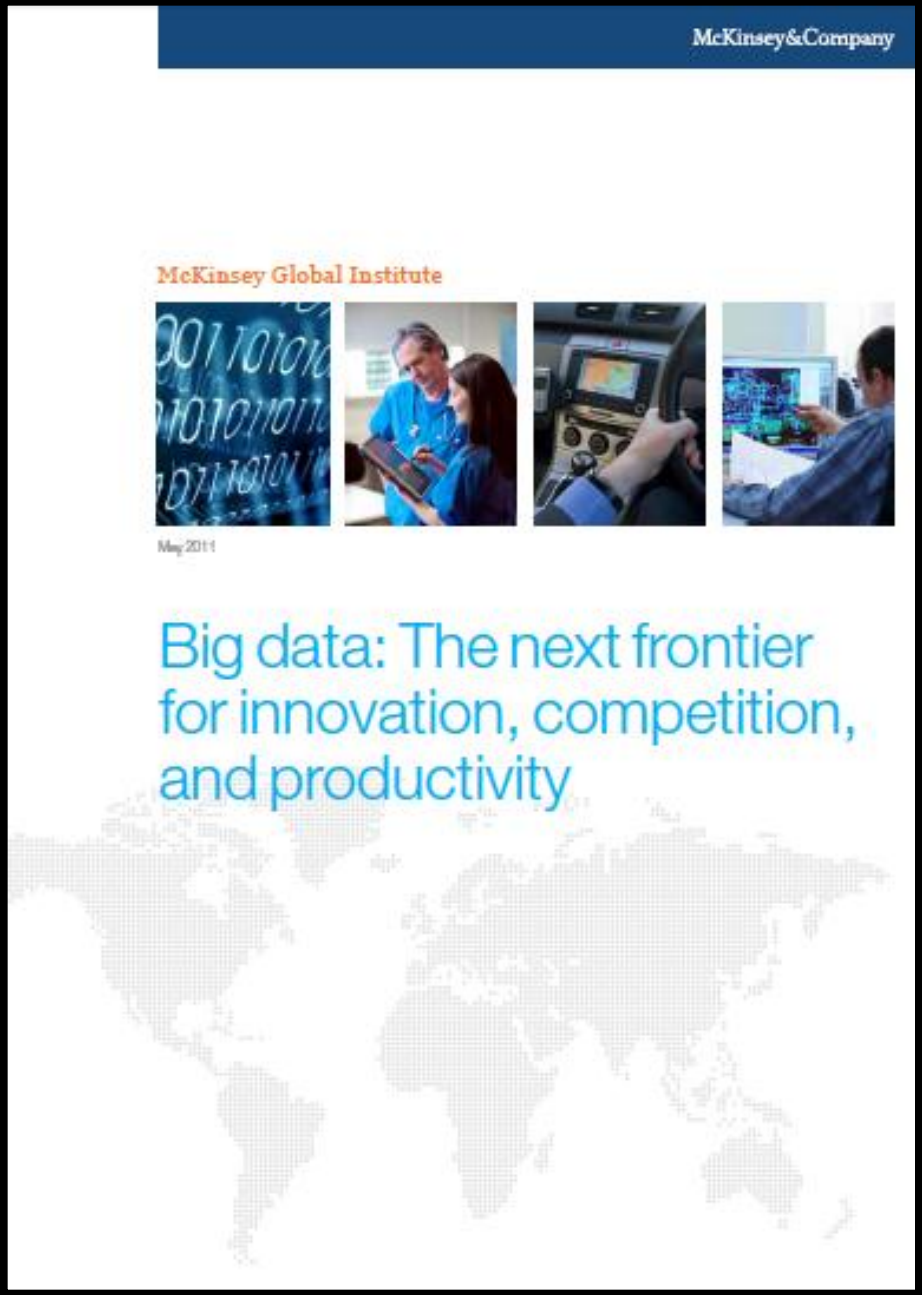
Skill gap	2-5 years	Now
Preserving research outputs	49%	10%
Data management & curation	48%	16%
Comply with funder mandates	40%	16%
Data manipulation tools	34%	7%
Data mining	33%	3%
Metadata	29%	10%
Preservation of project records	24%	3%
Sources of research funding	21%	8%
Metadata schema, discipline standards, practices	16%	2%

Data from RLUK/Mary Auckland: Reskilling for Research 2012

“Very few librarians are likely to have specialist scientific or medical knowledge - if you train as a research scientist or a medic, you probably won’t become a librarian.”

RLUK/Mary Auckland: Reskilling for Research 2012

Position	Location
Science Data Librarian	Stanford
Data Management Librarian	Oregon State
Social Sciences Data Librarian	Brown
Data Curation Librarian	Northeastern
Data Librarian	New South Wales
Research Data Management Co-ordinator	Sydney
Research Data & Digital Curation Officer	Cambridge
Data Services Librarian	Iowa
Data Analyst	ANDS
Institutional Data Scientist	Bath



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

Predicts a shortage of 190,000 data scientists by 2019

http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

Gaps? Opportunities??

1. Define core components of data informatics

- Metadata (discovery, preservation)
- Domain ontologies
- Visualisation e.g. VisTrails
- Workflow e.g. Taverna
- Analysis e.g. R

Lyon, Informatics Transform, IJDC 2012

Gaps? Opportunities??

2. Analyse LIS entry qualifications & increase STEM entrants

Target

- Biologists
- Chemists
- Mathematicians

Lyon, Informatics Transform, IJDC 2012

Gaps? Opportunities??

3. International Data Informatics Working Group to explore promotion, recognition & reward

- Global awareness campaign
- Career incentives
- Benchmark good practice

Lyon, Informatics Transform, IJDC 2012

“The ability to take data - to be able to understand it, to process it, to extract value from it, to visualise it, to communicate it - that’s going to be a hugely important skill in the next decades.”

Hal Varian, Chief Economist, Google

8th International Digital Curation Conference, Amsterdam, 14-16 January 2013

Thank you!

Informatics Transform article

http://www.ijdc.net/index.php/ijdc/article/view/210

Slides

http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/presentations.html

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