Incremental Change or Revolution? Libraries and the Informatics Transform

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

LIDA Conference, Zadar, June 2012

UKOLN is supported by:

JISC

www.ukoln.ac.uk

A centre of expertise in digital information management

UNIVERSITY OF BATH

THE SUNDAY TIMES UNIVERSITY OF THE YEAR 2011-12

Running order…..

- Headlines, Trends, Reports
- Roles and responsibilities
- Skills and competencies
- Gaps and opportunities

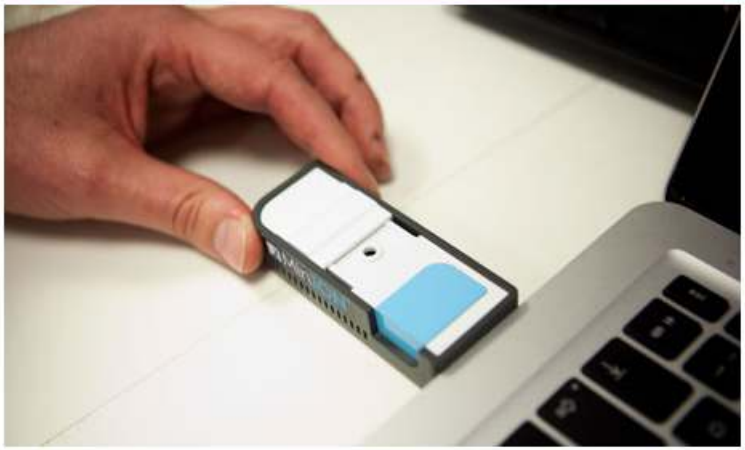
Changing research practice

DNA machine can sequence human genomes in hours

Oxford Nanopore has come up with a DNA sequencing machine the size of a USB memory stick that can decode the building blocks of life within hours rather than days

Julia Kollewe
guardian.co.uk, Friday 17 February 2012

theguardian



The MinION, a tiny DNA sequencing machine made by a firm spun out of Oxford University. Photograph: Nigel Chapman

Real-time gene sequencing used to combat superbug

REUTERS

Recommend | Be the first of your friends to recommend this.

An employee displays MRSA (Methicillin-resistant Staphylococcus aureus) bacteria strain inside a petri dish containing agar jelly for bacterial culture in a microbiological laboratory in Berlin March 1, 2008.
Credit: Reuters/Fabrizio Bensch

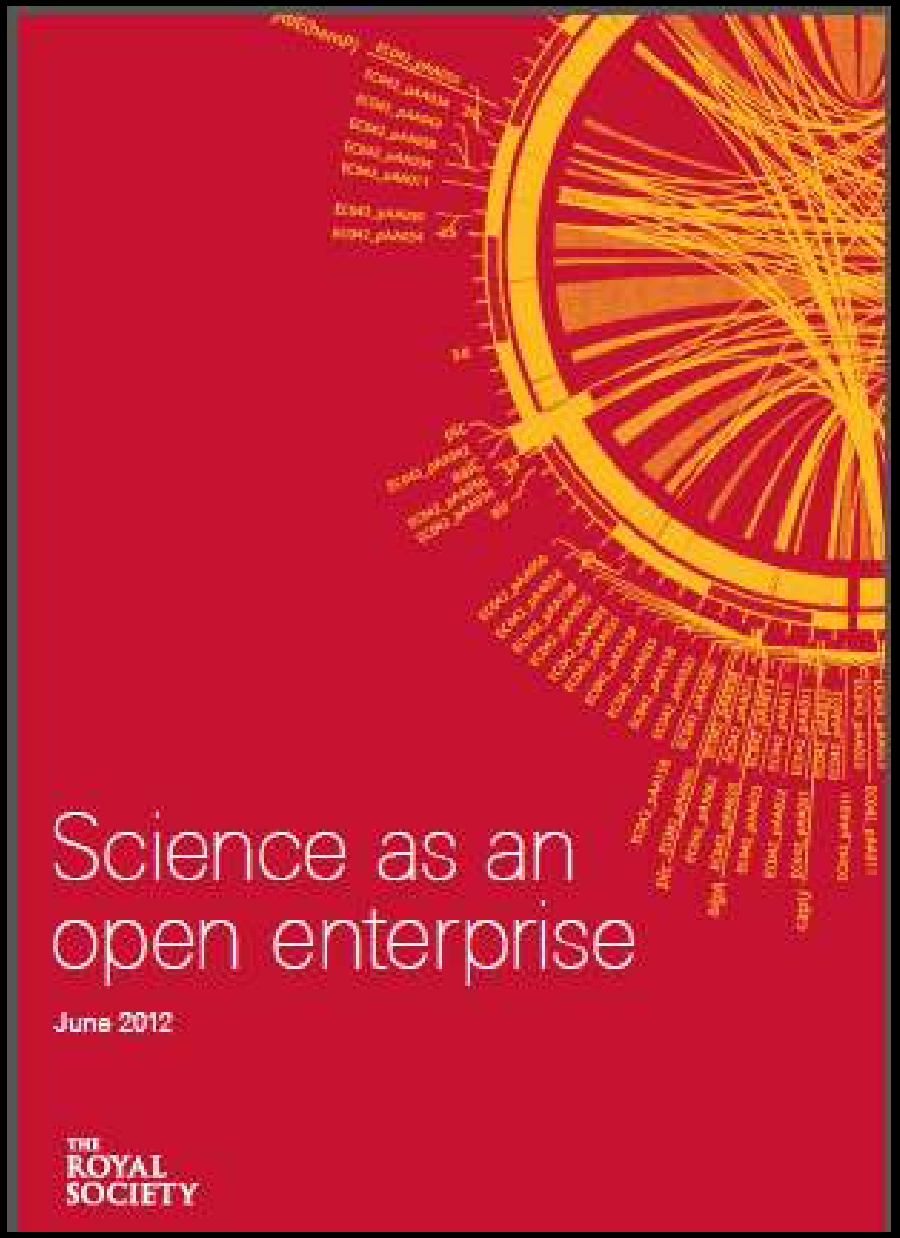
By Chris Wickham
LONDON | Mon Jun 18, 2012 12:19pm IST

(Reuters) - Scientists have used genome sequencing technology to control an outbreak of the superbug MRSA in a study that could point to faster and more efficient treatment of a range of diseases.

Tweet 2
Link this
Share this
Digg
Email
Print

Related Topics
Science »

Stocks
Illumina Inc
ILMN.O
$39.45
-0.36 -0.90%
19:01:12 IDT



Royal Society Report

Science as an Open Enterprise

June 2012

10 Recommendations

Published today!

http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

“ Recommendation 6
As a condition of publication, scientific journals should enforce a requirement that the data on which the argument of the article depends should be accessible, assessable, usable and traceable through information in the article.”

Science as an Open Enterprise Report, Royal Society, UK

“A particular dilemma for universities is to determine the role of their science libraries in a digital age. ……

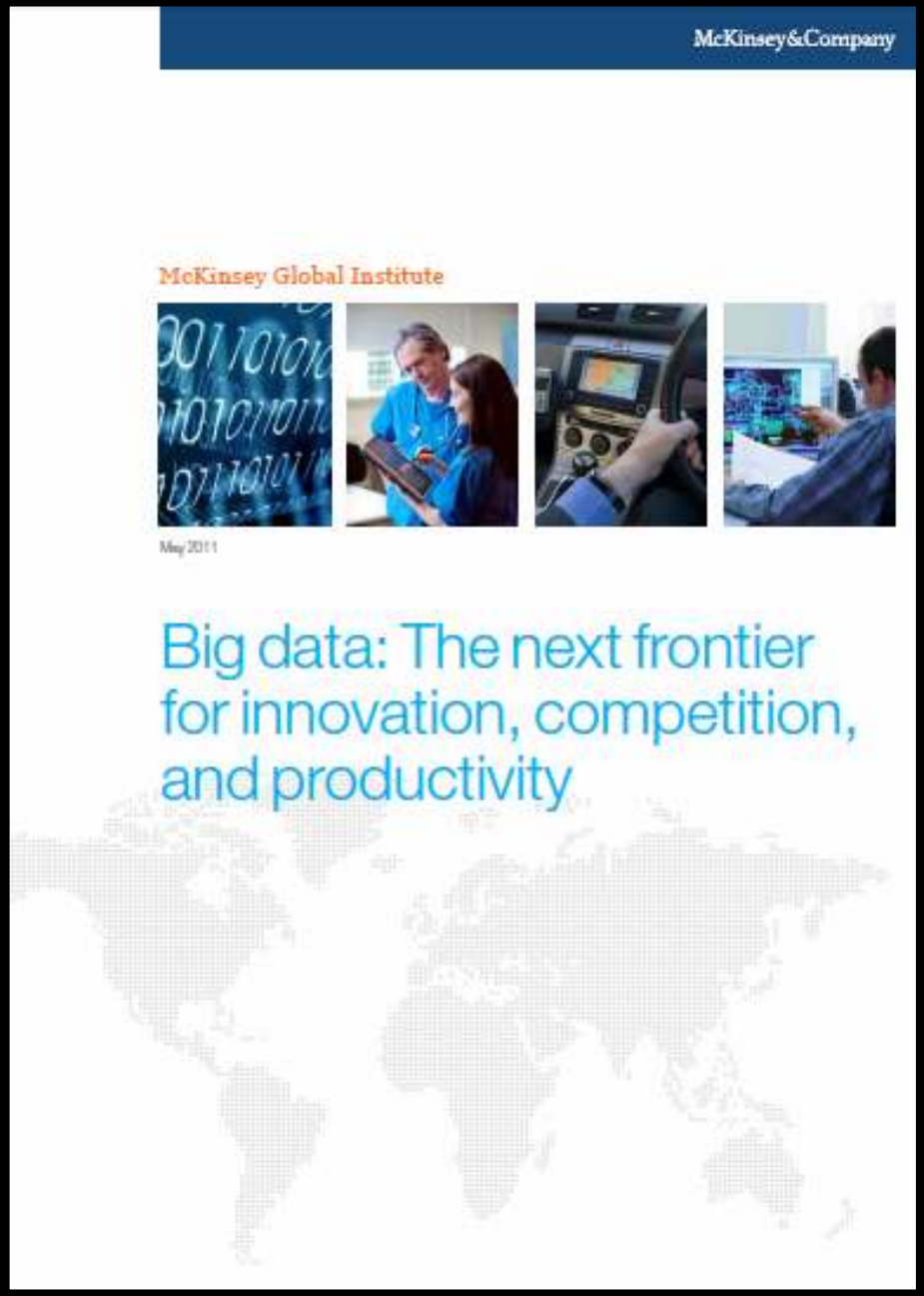
The traditional role of the library has been as a repository of data, information and knowledge and a source of expertise in helping scholars access them.

That role remains, but in a digital age, the processes and the skills that are required to fulfil the same function are fundamentally different. ”

Science as an Open Enterprise Report, Royal Society, UK

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.”

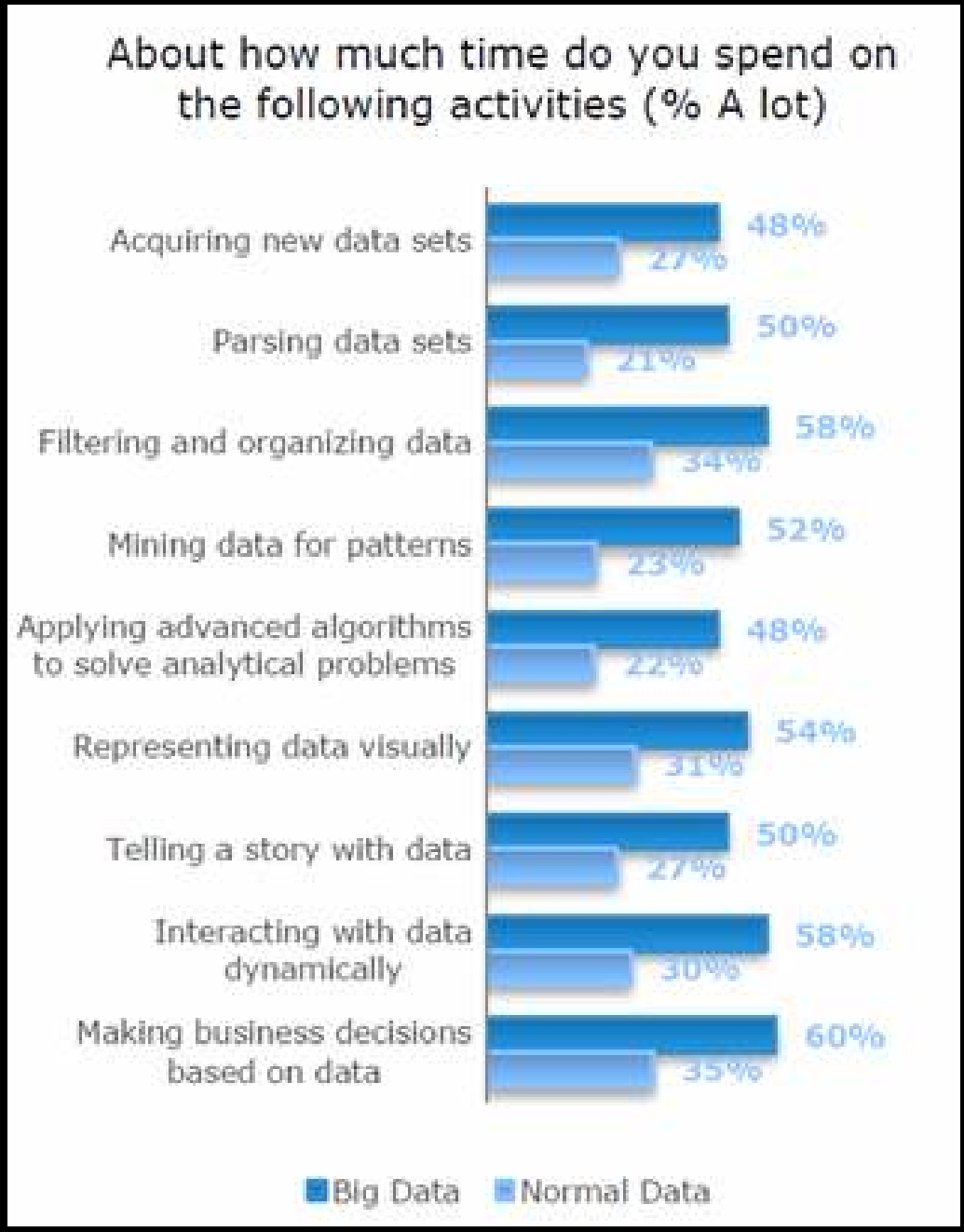


http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

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

Predicts a shortage of 190,000 data scientists by 2019

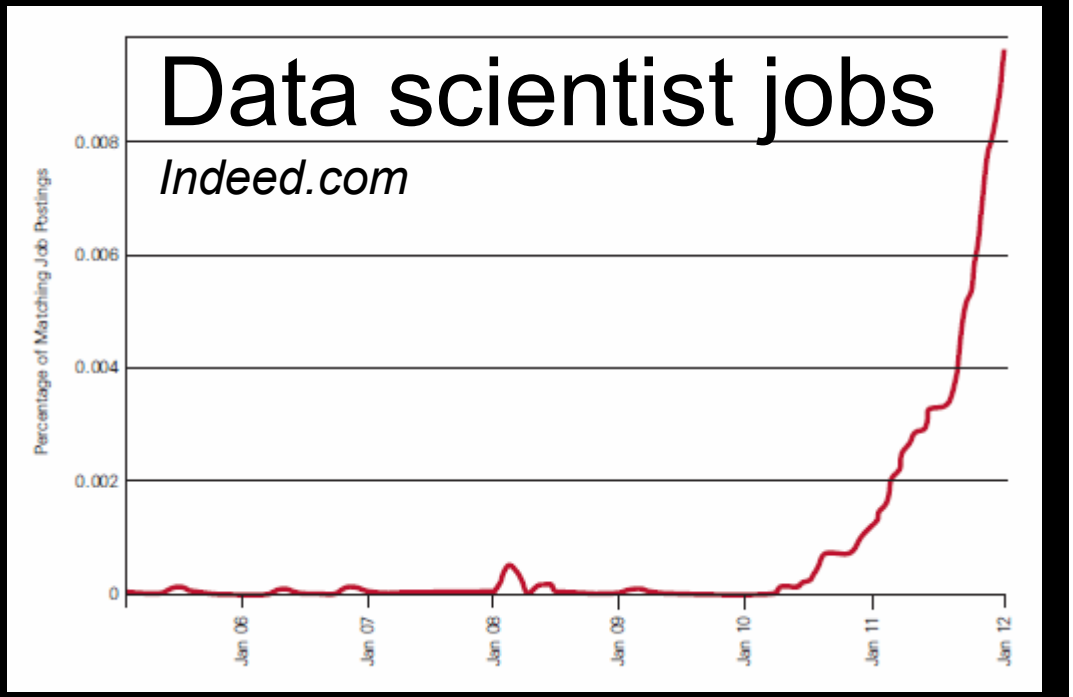
“Big Data” Data scientist



EMC²

Data Science Revealed community survey

http://www.emc.com/collateral/about/news/emc-data-science-study-wp.pdf



theguardian

News | Sport | Comment | Culture | Busine

News › Datablog

DATABLOG

Facts are sacred

What is a data scientist?

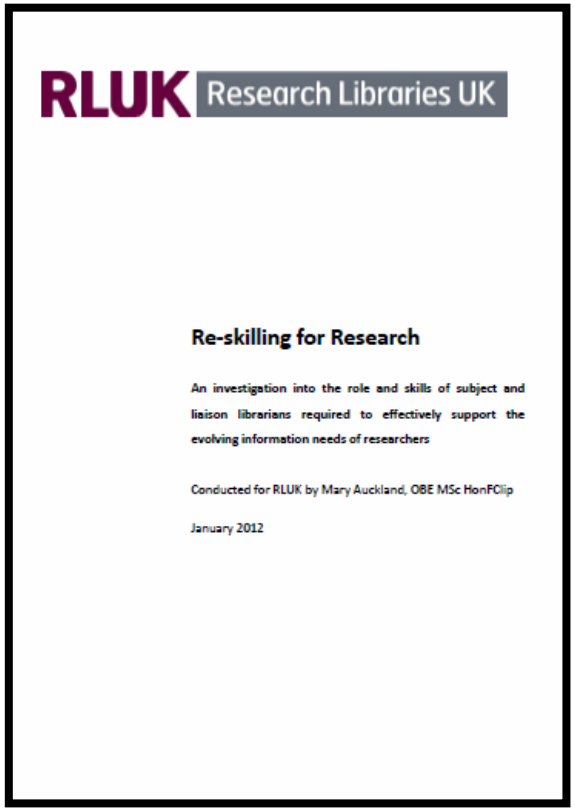
It's the job of the moment. But what exactly is a data scientist?

Data-related roles in Libraries?

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

DCC

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



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%

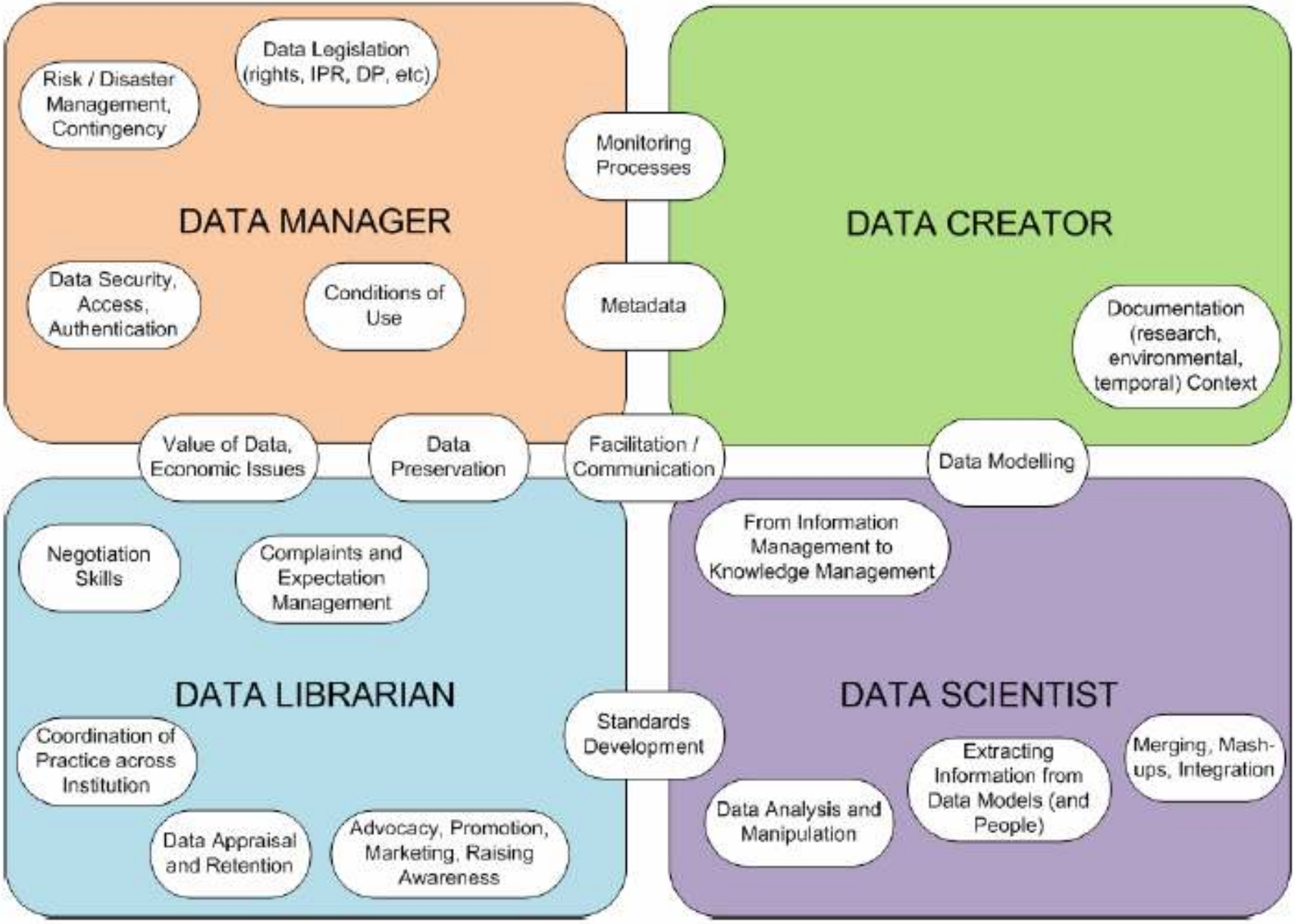
Data from RLUK/Mary Auckland: Reskilling for Research 2012

“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

CORE SKILLS FOR DATA MANAGEMENT

A follow-up from the second DCC Research Data Management Forum (November 2008)



- Leadership & co-ordination
- Strategy and planning
- Policy
- Legal and ethical (FoI, Data Protection)
- Advocacy (data informatics)
- Data repositories
- Data storage
- Data analysis
- Data visualisation
- Data mining
- Data modelling
- Data licensing
- Training….

Libraries and research data management

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

Liz Lyon, Informatics Transform, IJDC Current Issue, 2012

Data roles

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

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

Full mapping : Informatics Transform, IJDC Current issue, 2012

April 2011 - EPSRC Letter to VCs

EPSRC
Pioneering research and skills

Engineering and Physical Sciences Research Council

EPSRC expects all those institutions it funds

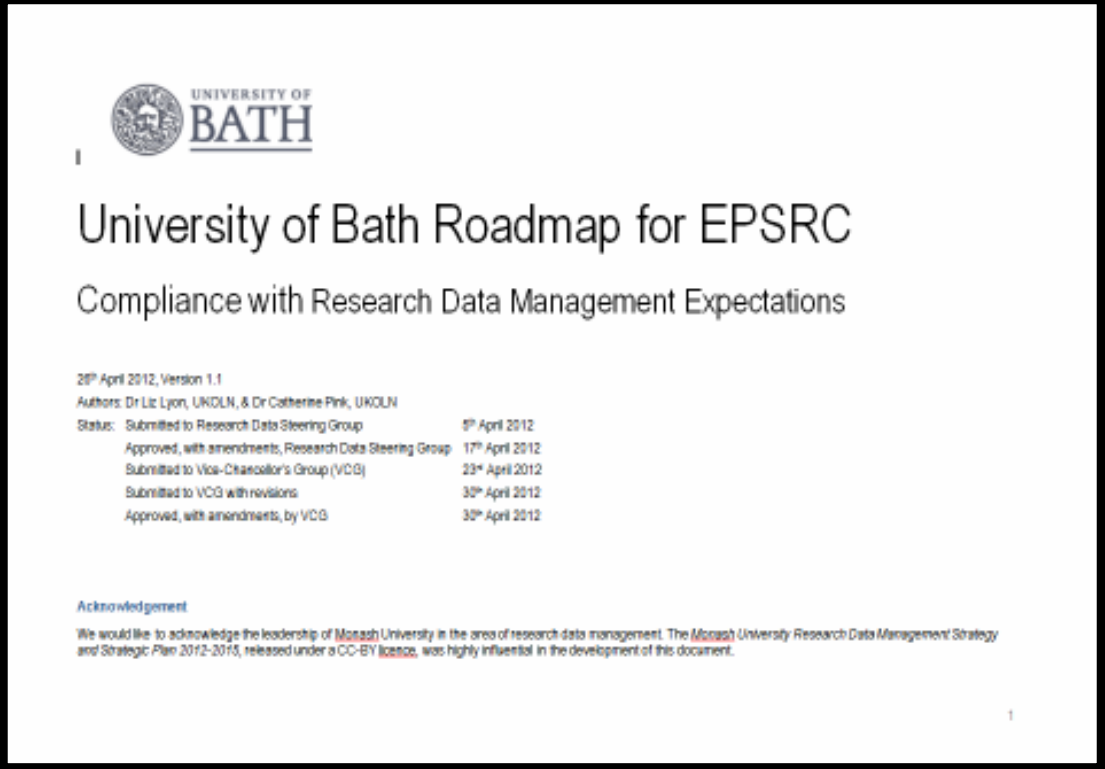
- to develop **a roadmap** that aligns their policies and processes with EPSRC's **expectations** by **1st May 2012**;
- to be fully compliant with these **expectations** by **1st May 2015**.

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

EPSRC
Pioneering research and skills
Engineering and Physical Sciences Research Council

- Awareness of regulatory environment
- Data access statement
- Data policies and processes
- Data storage
- Structured metadata descriptions
- DOIs for data
- Data securely preserved for a minimum of 10 years

- Leadership opportunity for Libraries?
- Pan-institutional perspective
- Wider strategic alignment
- Collaborate to develop Operational Plan
- *http://www.bath.ac.uk/rdso/University-of-Bath-Roadmap-for-EPSRC.pdf*



UNIVERSITY OF BATH

University of Bath Roadmap for EPSRC

Compliance with Research Data Management Expectations

28th April 2012, Version 1.1

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

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

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
	Advocacy	Knowledge of appropriate disciplinary data centres,	Data centre staff

Full mapping : Informatics Transform, IJDC Current issue, 2012

Advocacy, Library support services?

- ***Data requirements:*** legacy data
- ***Data management plans*****:** tools
- ***Informatics*****:** disciplinary metadata schema, standards, formats, identifiers, ontologies
- ***Citation:*** links to publications
- ***Reuse*****:** tracking your data

JISC

DCC because good research needs good data

Understanding Data Requirements

Data Asset Framework

BOS
Bristol Online Surveys

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
collaborative assessment of research data infrastructure and objectives

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

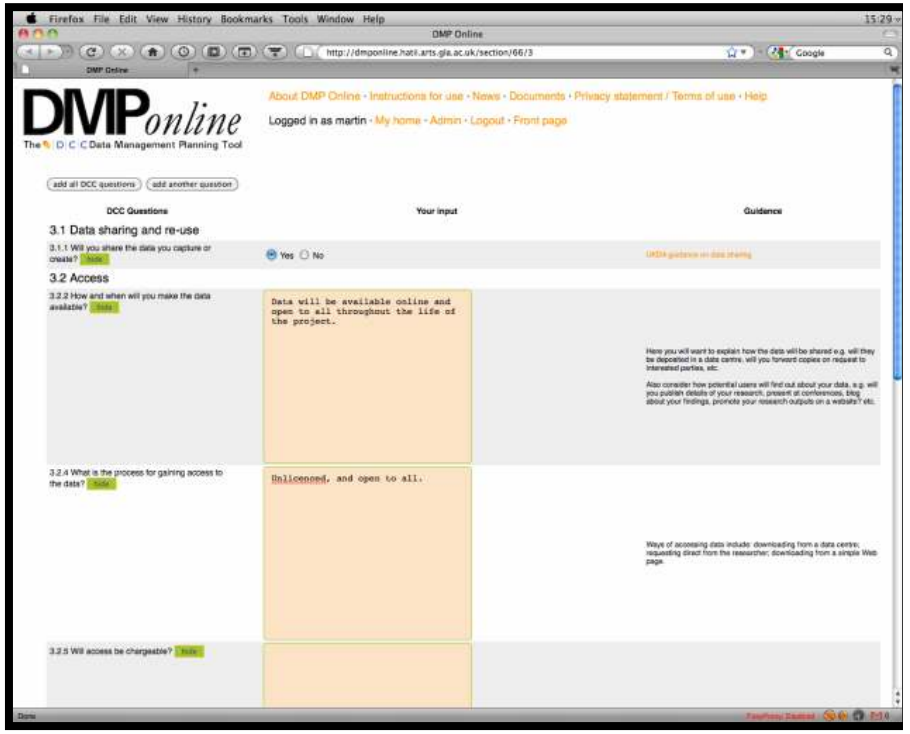
DCC

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

Data management plans

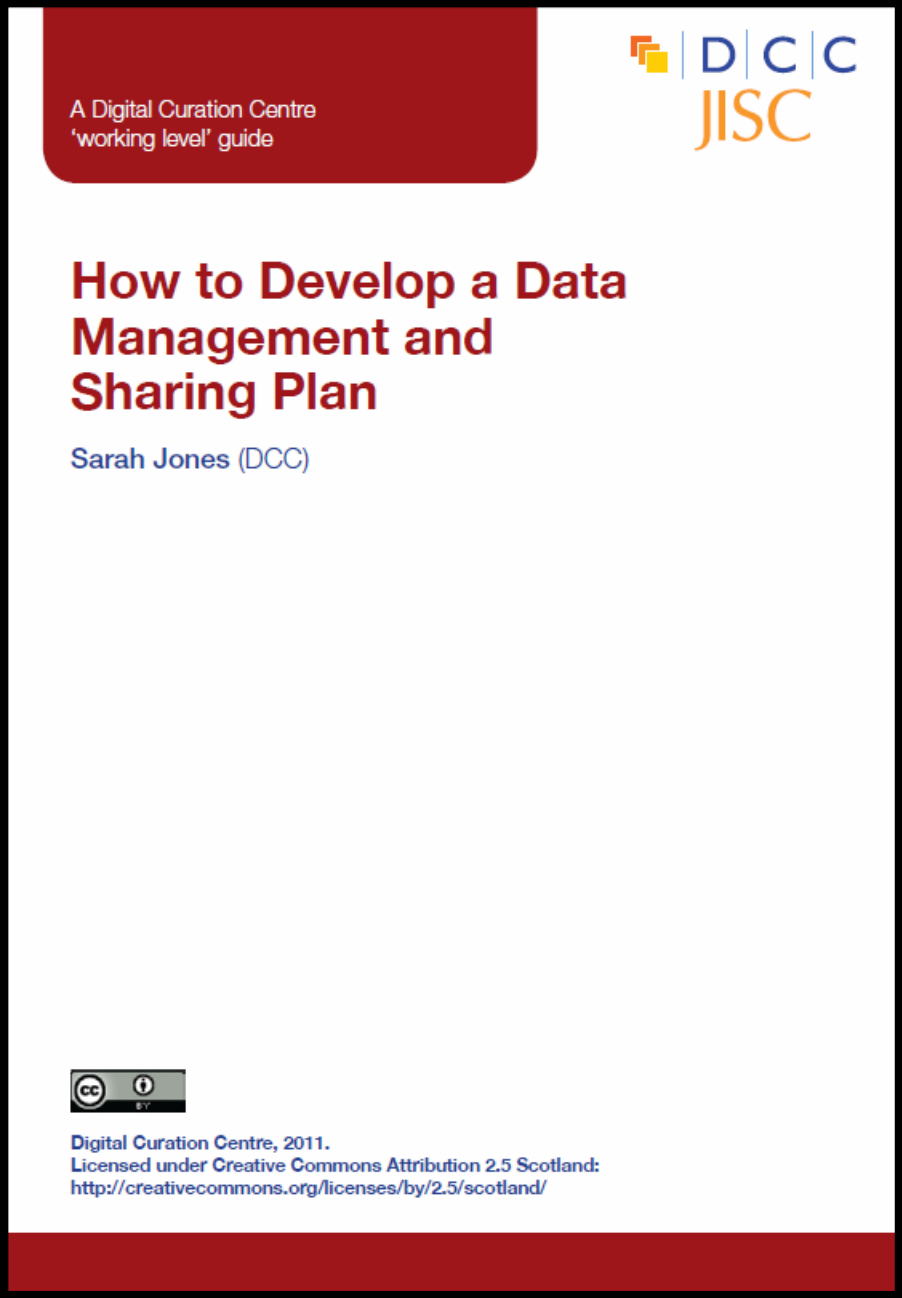
DMPonline

The DCC Data Management Planning Tool



DMPTool

Guidance and Resources for your Data Management Plan


A Digital Curation Centre
'working level' guide
DCC
JISC
How to Develop a Data Management and Sharing Plan
Sarah Jones (DCC)
Digital Curation Centre, 2011.
Licensed under Creative Commons Attribution 2.5 Scotland:
http://creativecommons.org/licenses/by/2.5/scotland/

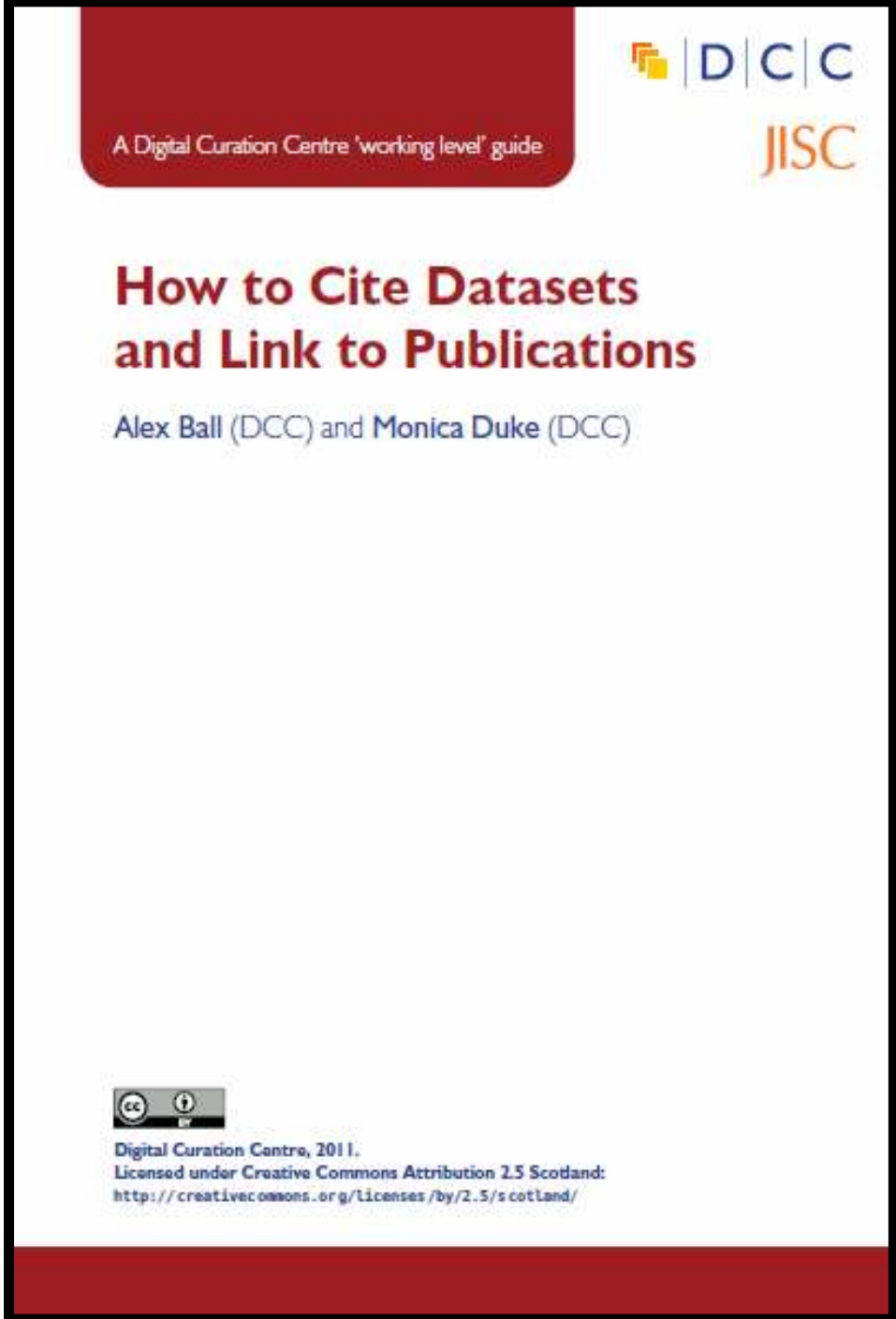

DCC Checklist for a Data Management Plan
DCC

DMP guides

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

Full mapping : Informatics Transform, IJDC Current issue, 2012



How to cite data

DataCite

Helping you to find, access, and reuse data

Using DOIs

What data to keep

Roles and Responsibilities

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

DCC

ands

How to Appraise & Select Research Data for Curation

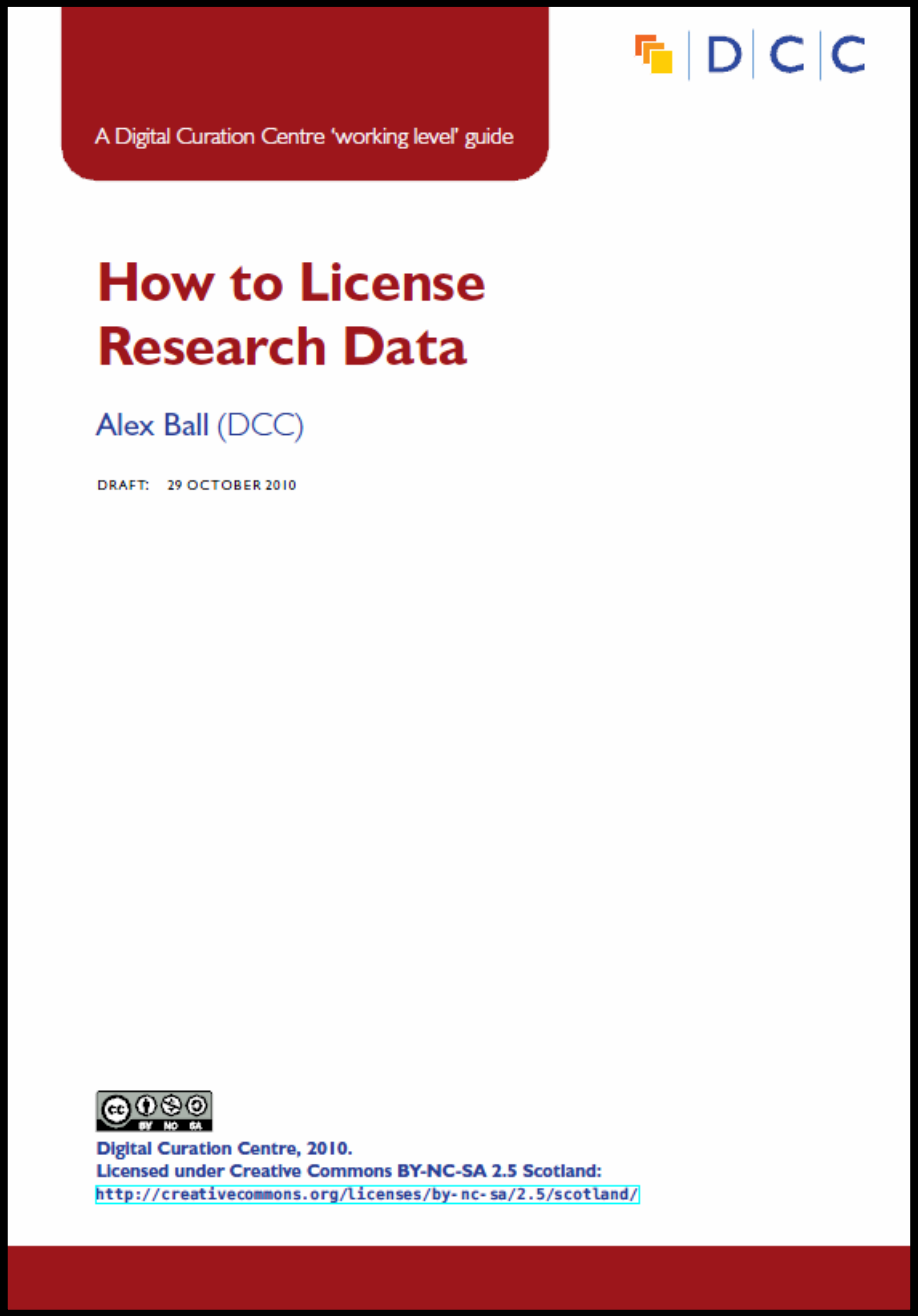
Angus Whyte (DCC) and **Andrew Wilson** (ANDS)

Researcher ('data creator')

- Provide enough information for others to assess the research data's scientific and scholarly quality and compliance with disciplinary or ethical norms.
- Provide relevant information for the repository to identify who will use the data and how i.e. the 'designated community', and any specific access requirements or constraints.
- Provide the research data in formats recommended by the data repository.
- Provide the metadata requested by the repository.

Data centre or repository

- Make explicit its mission in the area of digital archiving, and its selection policy for digital objects.
- Ensure compliance with legal regulations and contracts.
- Ensure the authenticity and integrity of the digital objects and the metadata.
- Assume responsibility from the data producer for ensuring the digital objects are accessible and available to a defined 'designated community'.
- Plan for long-term preservation of the digital assets.



Data Licensing

Bespoke licences
Standard licences
Multiple licensing
Licence mechanisms

science commons

How to track impact

http://total-impact.org/

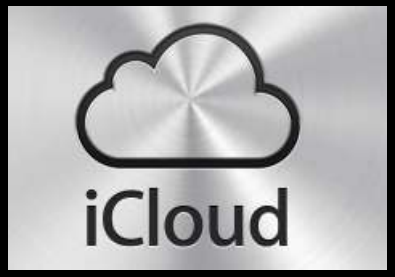
Role	Responsibilities	Requirements	Relationships
Repository managers	To ensure research papers have persistent links to underlying research data	Knowledge of persistent identification mechanisms and publisher requirements	Data librarians / Data scientists / Liaison /Subject / Faculty Librarians
IT / 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

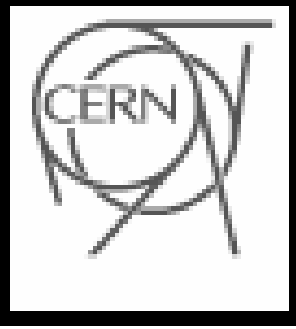
Storage

CRIS

Full mapping : Informatics Transform, IJDC Current issue, 2012

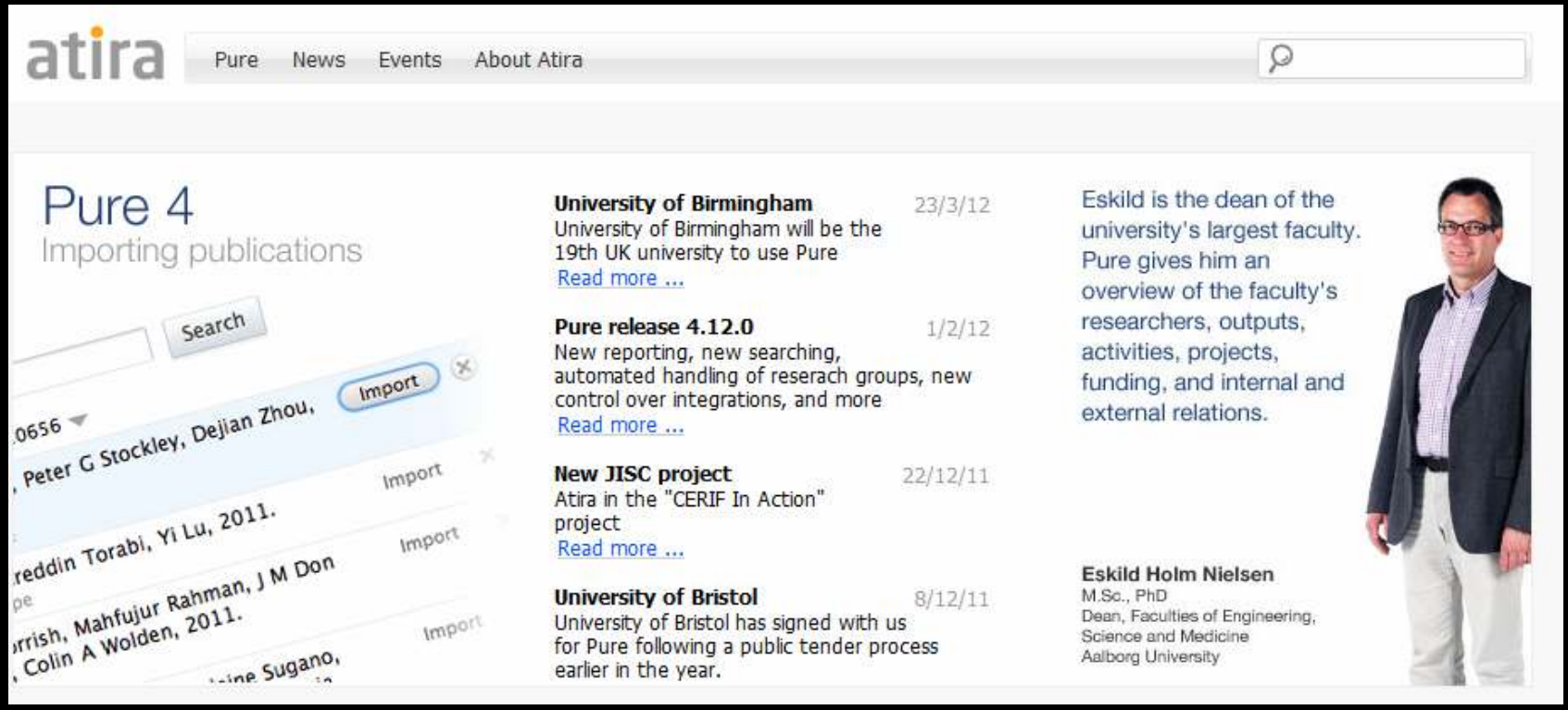
- **Storage:** file-store, quotas, cloud, data centres, funder policy
- **Access:** embargoes, FoI





CERIF for Datasets

C4D is a project funded under JISC's Managing Research Data Programme



CRIS integration, CERIF and data

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

Institution	Policy name	Date released
University of Oxford	Committment to Research Data Management	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
University of East London	Research Data Management Policy for UEL	15 March 2012

Draft policies

University of Leeds - via the RoaDMaP project
Timeline of developmentments including draft policy text

University of Lincoln - via the Orbital project
Blog post with link to the draft policy text

University of Southampton - via the DataPool project
Blog post outlining how the policy is developing at Southampton

University of Manchester - via the MiSS project
'Towards a Research Data Management Policy' document outlining progress

Queen Mary University of London
Draft Research Data Management Policy

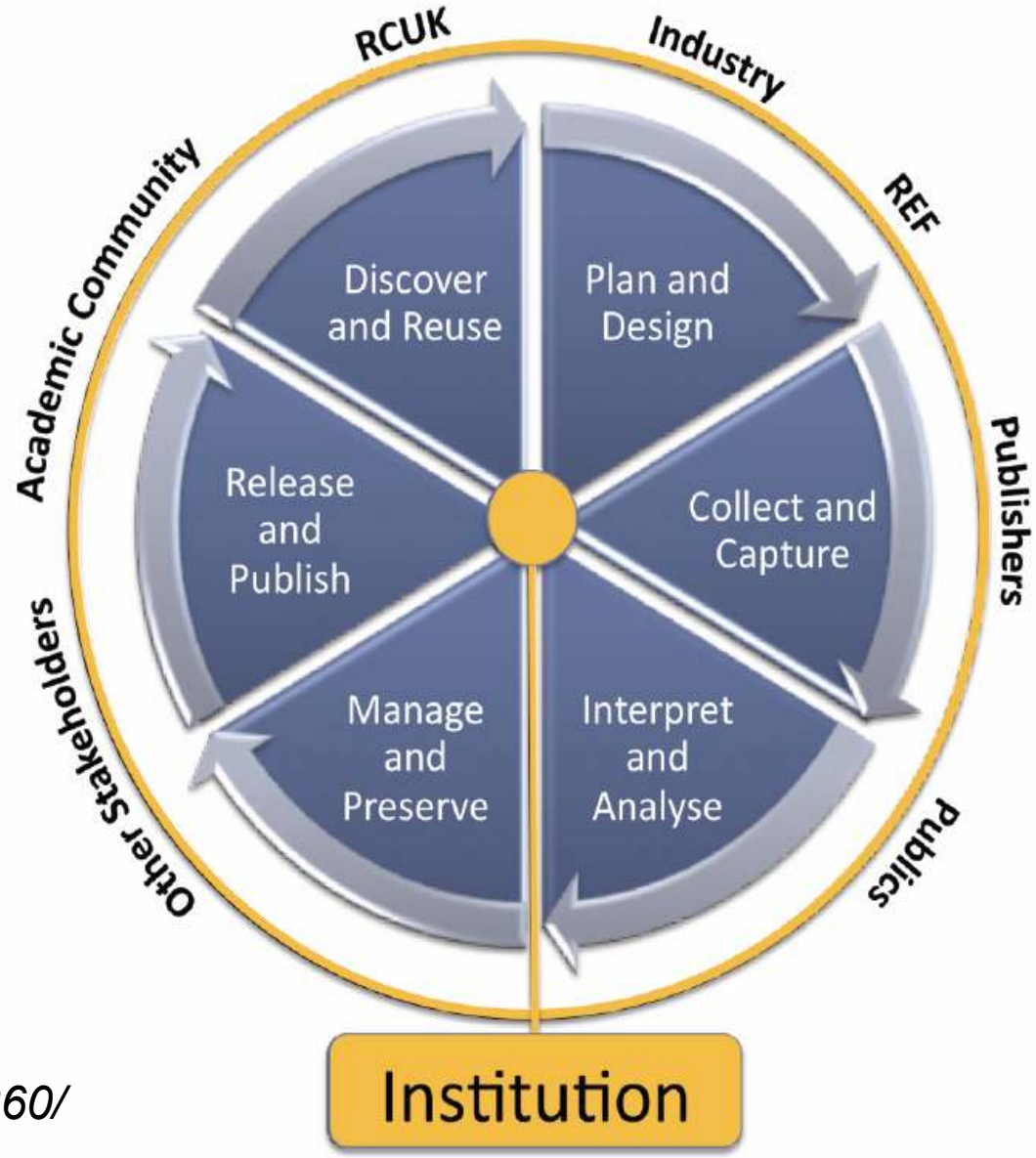
Libraries can inform institutional data policy development

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

Research360@Bath

JISC

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



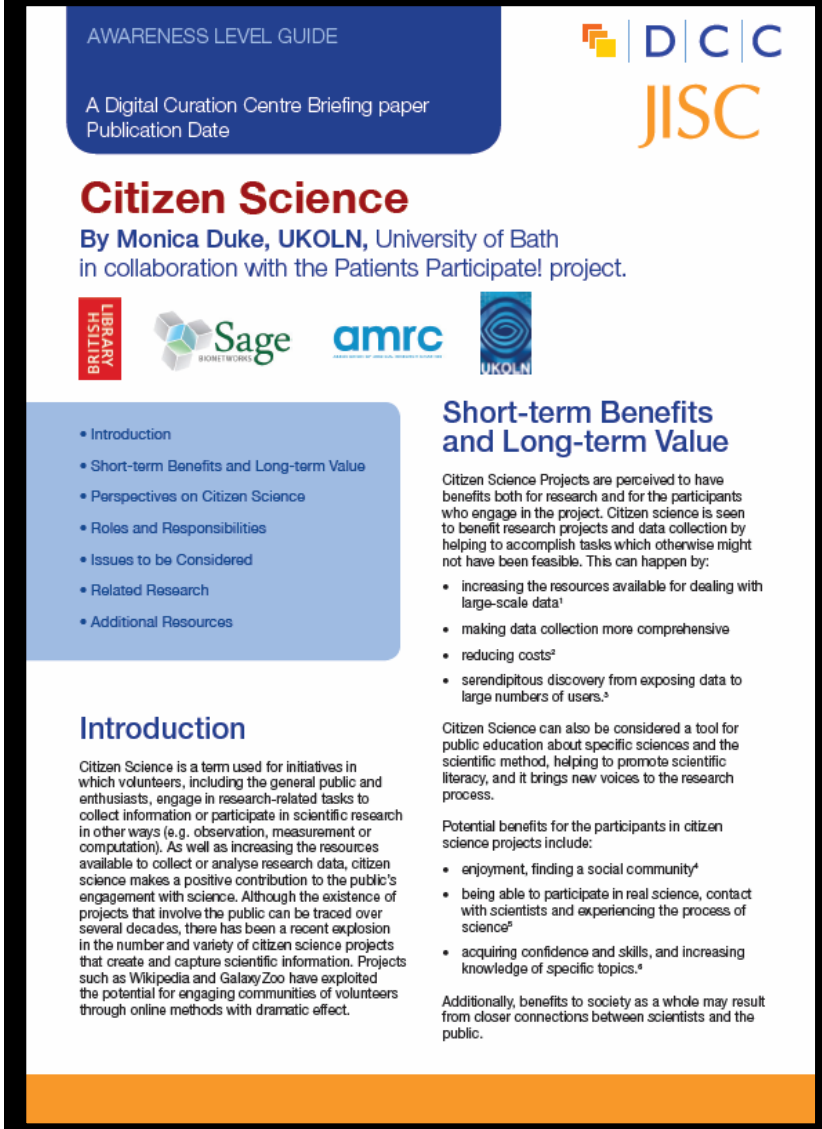
http://blogs.bath.ac.uk/research360/

Working with Doctoral Training Centres: Research360 Project @ Bath

JISC projects

DCC resources

Patients Participate!



AWARENESS LEVEL GUIDE

A Digital Curation Centre Briefing paper
Publication Date

DCC JISC

Citizen Science

By Monica Duke, UKOLN, University of Bath
in collaboration with the Patients Participate! project.

BRITISH LIBRARY · Sage BIONETWORKS · amrc · UKOLN

Introduction

Citizen Science is a term used for initiatives in which volunteers, including the general public and enthusiasts, engage in research-related tasks to collect information or participate in scientific research in other ways (e.g. observation, measurement or computation). As well as increasing the resources available to collect or analyse research data, citizen science makes a positive contribution to the public's engagement with science. Although the existence of projects that involve the public can be traced over several decades, there has been a recent explosion in the number and variety of citizen science projects that create and capture scientific information. Projects such as Wikipedia and GalaxyZoo have exploited the potential for engaging communities of volunteers through online methods with dramatic effect.

Short-term Benefits and Long-term Value

Citizen Science Projects are perceived to have benefits both for research and for the participants who engage in the project. Citizen science is seen to benefit research projects and data collection by helping to accomplish tasks which otherwise might not have been feasible. This can happen by:

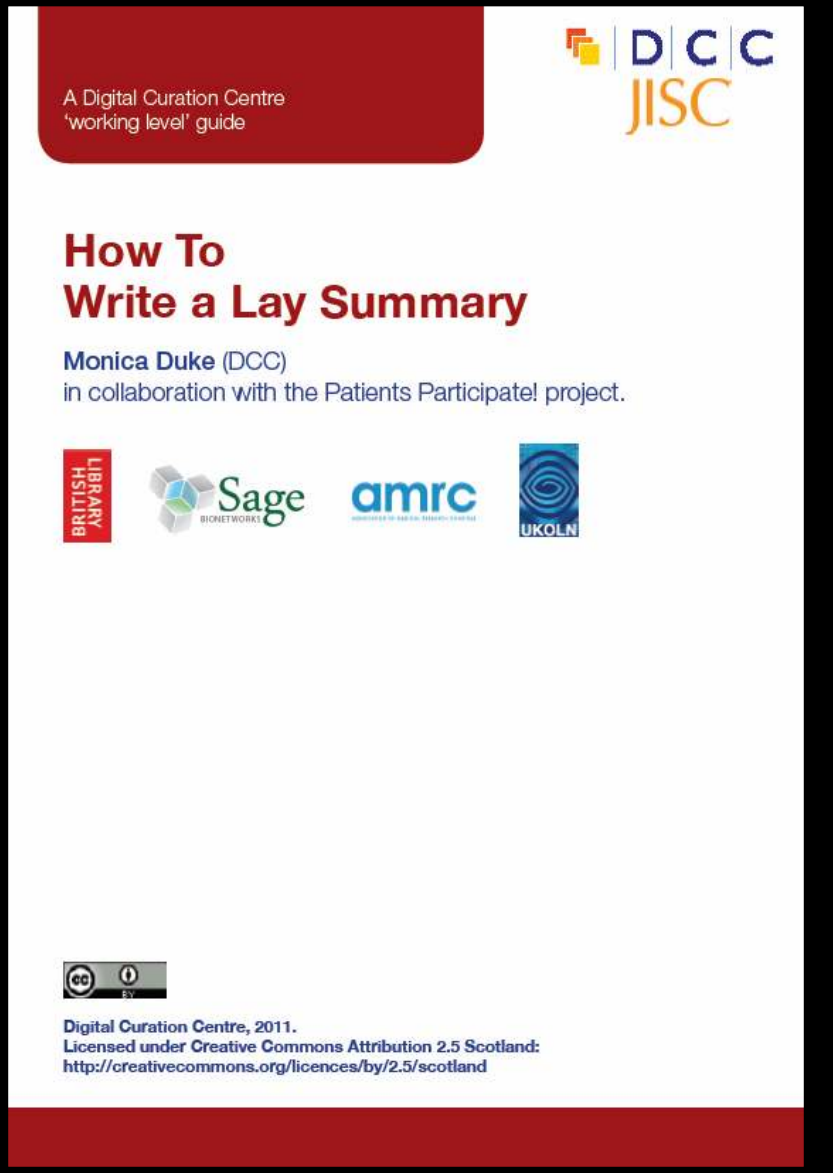
- increasing the resources available for dealing with large-scale data[1]
- making data collection more comprehensive
- reducing costs[2]
- serendipitous discovery from exposing data to large numbers of users.[3]

Citizen Science can also be considered a tool for public education about specific sciences and the scientific method, helping to promote scientific literacy, and it brings new voices to the research process.

Potential benefits for the participants in citizen science projects include:

- enjoyment, finding a social community[4]
- being able to participate in real science, contact with scientists and experiencing the process of science[5]
- acquiring confidence and skills, and increasing knowledge of specific topics.[6]

Additionally, benefits to society as a whole may result from closer connections between scientists and the public.



A Digital Curation Centre
'working level' guide

DCC JISC

How To Write a Lay Summary

Monica Duke (DCC)
in collaboration with the Patients Participate! project.

BRITISH LIBRARY · Sage BIONETWORKS · amrc · UKOLN

http://blogs.ukoln.ac.uk/patientsparticipate/

People Participate!@Bath

- Engaged360 Project at Bath
- RCUK Public Engagement Catalyst project
- Advocacy guidance to research staff & students
- Create a Bath Lay Summary Template
- Lay summaries for all new research articles in OPUS repository
- Training for new postgraduates – Library role?
- Embed in research lifecycle

- Leadership & co-ordination
- Strategy and planning
- Policy
- Legal and ethical (FoI, Data Protection)
- Advocacy (data informatics)
- Data repositories
- Data storage
- Data analysis
- Data visualisation
- Data mining
- Data modelling
- Data licensing
- Training….

Gaps? Opportunities??

Analyse LIS entry qualifications & increase STEM entrants

Target
- Biologists
- Chemists
- Mathematicians

Lyon, Informatics Transform, IJDC 2012

Gaps? Opportunities??

Define core components of data informatics and data science

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

Lyon, Informatics Transform, IJDC 2012

Gaps and Opportunities??

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

- Global awareness campaign
- Career incentives
- Benchmark good practice

Lyon, Informatics Transform, IJDC 2012

http://www.flickr.com/photos/50542505@N08/5723947474/

Data scientist flavours

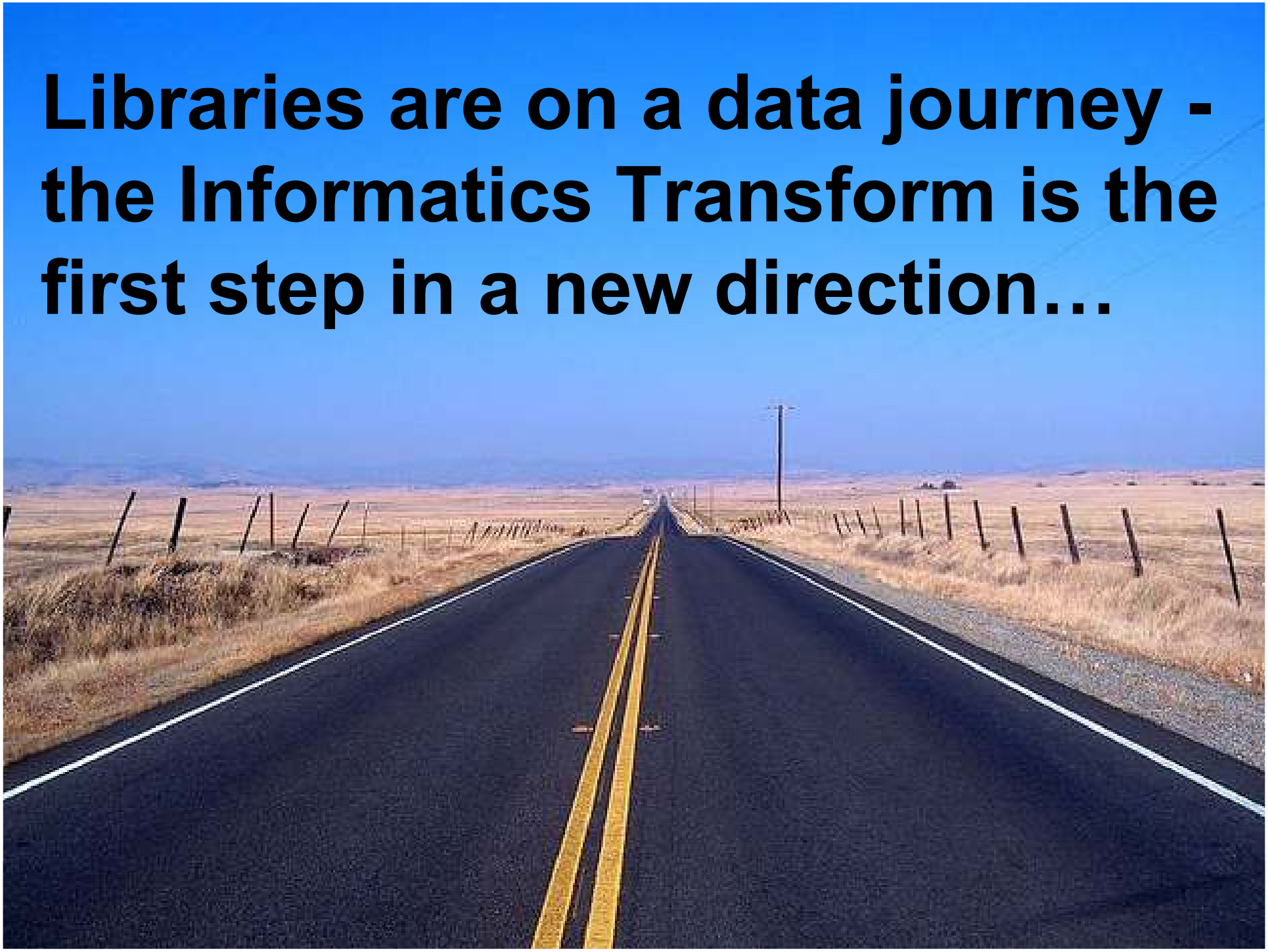
- Analysis, mining, modelling
- Visualisation, simulations
- Informatics, advocacy, training
- Repositories, preservation

Scientist
Data
Librarian

Infrastructure, Intelligence, Innovation: driving the Data Science agenda

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

Libraries are on a data journey - the Informatics Transform is the first step in a new direction…



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