The Informatics Transform:

Re-engineering Libraries for the Data Decade

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

VALA2012, Melbourne, Australia

UKOLN

DCC

UKOLN is supported by:

JISC

UNIVERSITY OF BATH

THE SUNDAY TIMES UNIVERSITY OF THE YEAR 2011-12

UKOLN

www.ukoln.ac.uk

A centre of expertise in digital information management

“Data is the new oil.”

Andreas Weigend, Stanford (ex Amazon)

“The future belongs to companies and people that turn data into products”

Mike Loukides, O’Reilly Media

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/

flickr.com/photos/orlandaswelch/3597438481/

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

Oceans: last unmapped frontier?

http://www.wired.com/wiredscience/2011/09/ocean-sensor-network/

http://bohemianadventures.blogspot.com.au/2010/06/bering-sea-day-dutch-harbor.html

BBC Mobile News | Sport | Weather | iPlayer | TV

NEWS POLITICS

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Entertainment & Arts

5 December 2011 Last updated at 21:22

1.3K Share

Everyone 'to be research patient', says David Cameron

..using personal data for research

"Let me be clear, this does not threaten privacy, it doesn't mean anyone can look at your health records, but it does mean using anonymous data to make new medical breakthroughs.

Share your genome data?

- Buy a DTC kit
- Join a project

Personal Genome Project

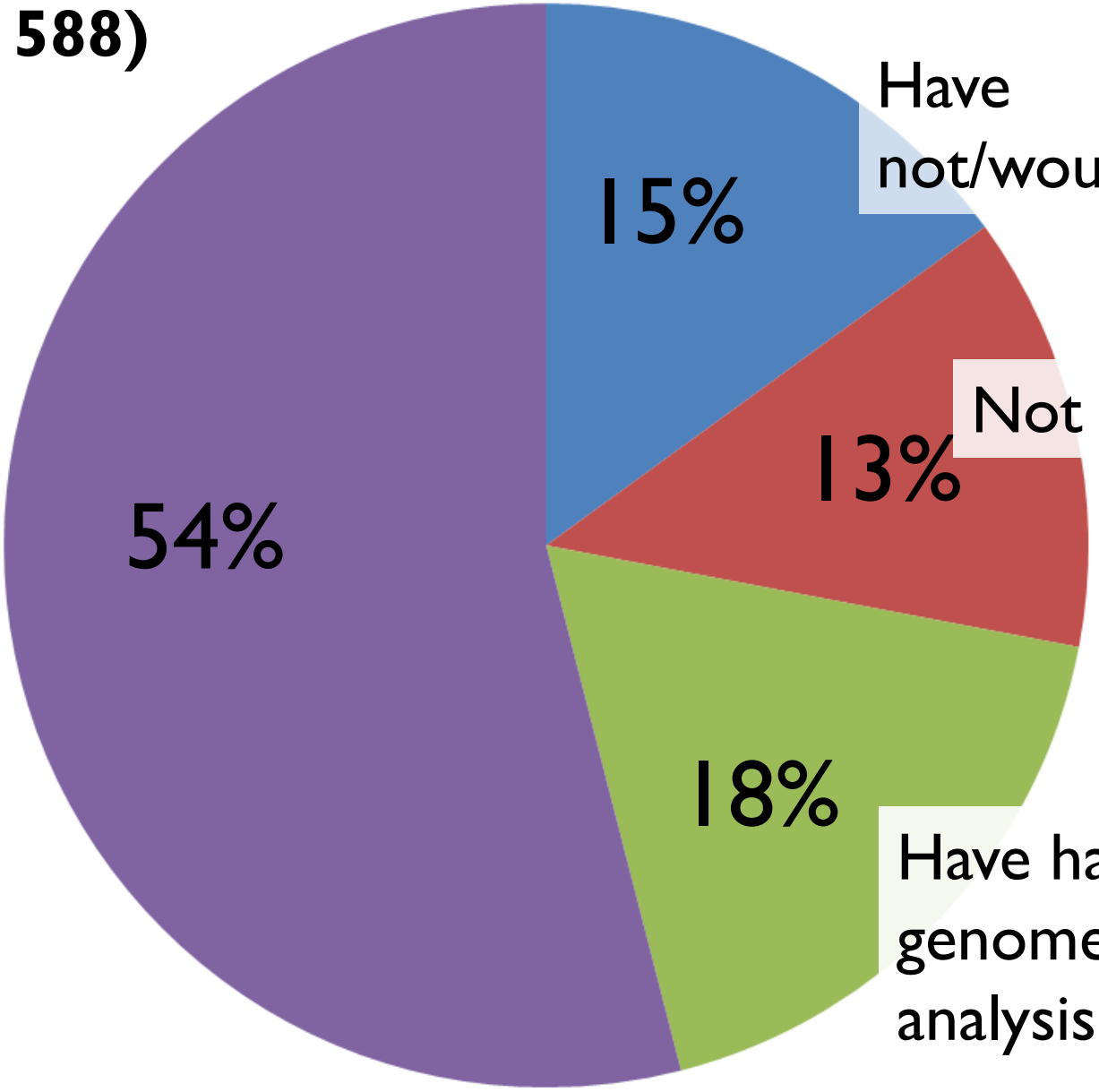
Personal Kit

Order a kit for yourself today.
Requires 1-year subscription to 23andMe's Personal Genome Service® billed at $9 per month.

$99 (1-year subscription required)

add to cart

In a recent 2011 survey, *Nature* asked its readers whether they had, or would consider, a genome analysis (n=1588)



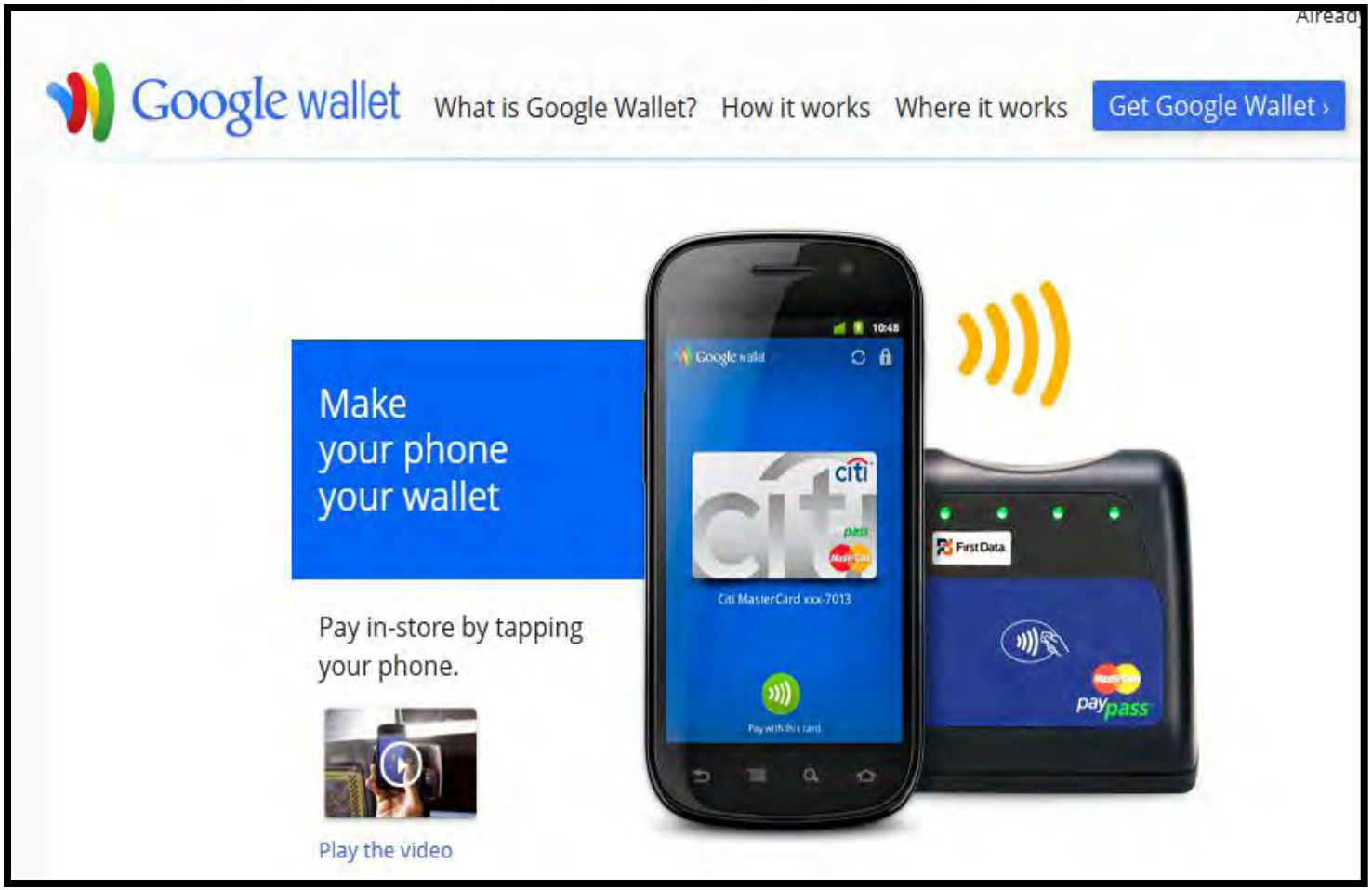
nature

Consumer data…

BUSINESS INSIDER
SAI Contributors

Companies Are Gearing Up To Track Every Retail Transaction Through Your Smartphone

Michael Griffin



One in every nine people on Earth is on Facebook

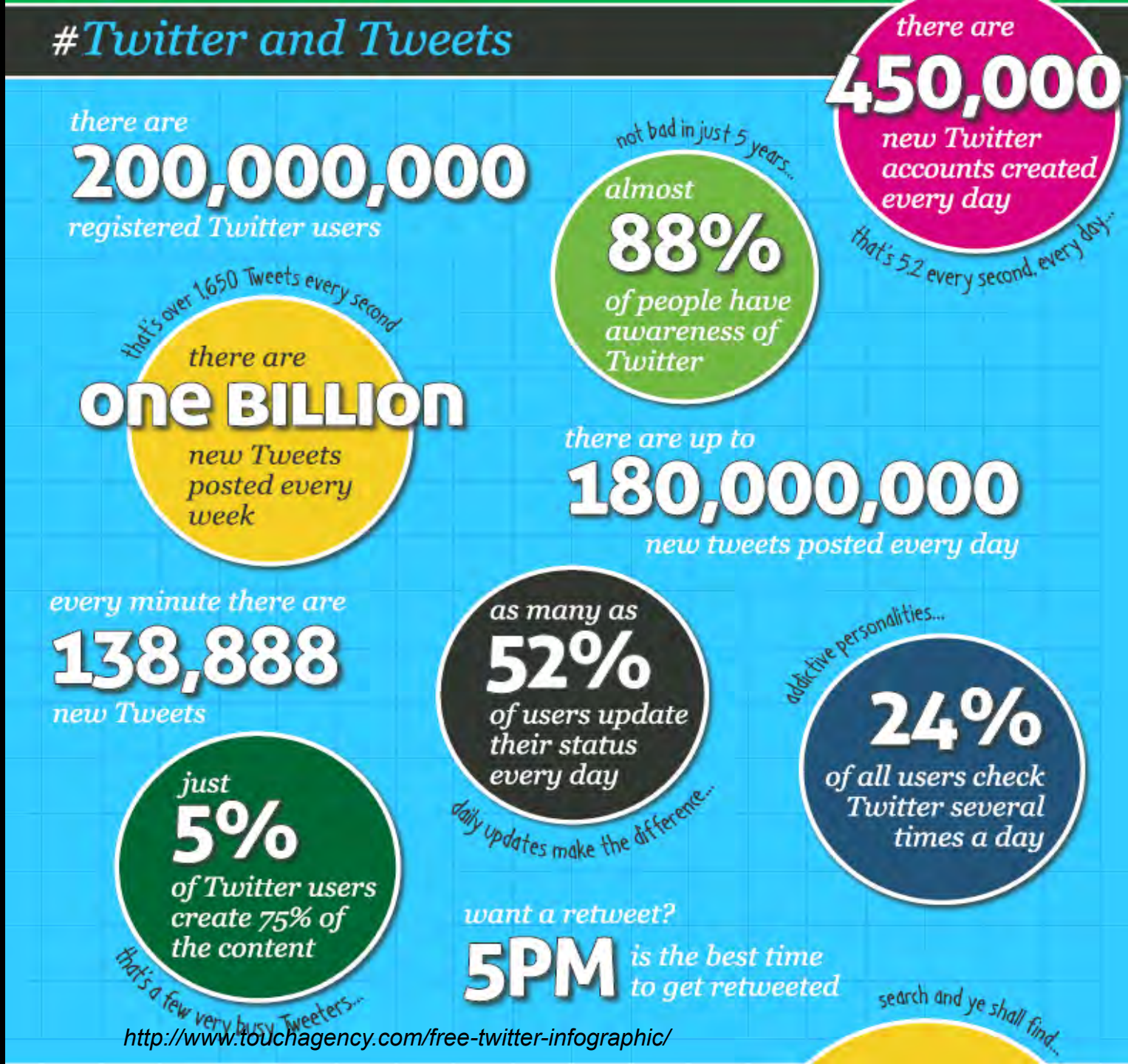
30billion pieces of content are shared on Facebook each month

People upload 3000images to Flickr every minute

Google+ has > 25million users

From 20 Social Media Statistics (Jeffbullas)

...and conversations



"Data is the new oil."

Andreas Weigend, Stanford (ex Amazon)

Data is more like soup – its messy and you don't know what's in it….

Quantified Self

self knowledge through numbers

“DIY”

Human physiology data

http://www.technologyreview.com/biomedicine/37784/

“Herculean” and “Heroic”

Particle physics data

BBC Mobile NEWS

13 December 2011 Last updated at 17:20

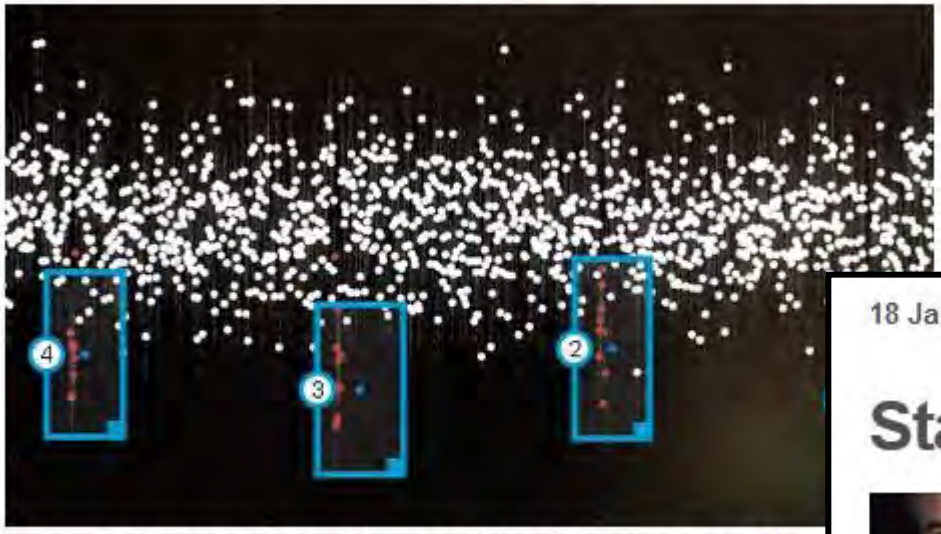
15K

LHC: Higgs boson 'may have been glimpsed'

By Paul Rincon
Science editor, BBC News website, Geneva

Two teams at the LHC have seen hints of what may well prove to be the Higgs

Volunteers wanted for planet hunt



Time-lapsed images of a single star show dips in brightness as a plane

Members of the public are being asked to join the hunt for planets that could support life.

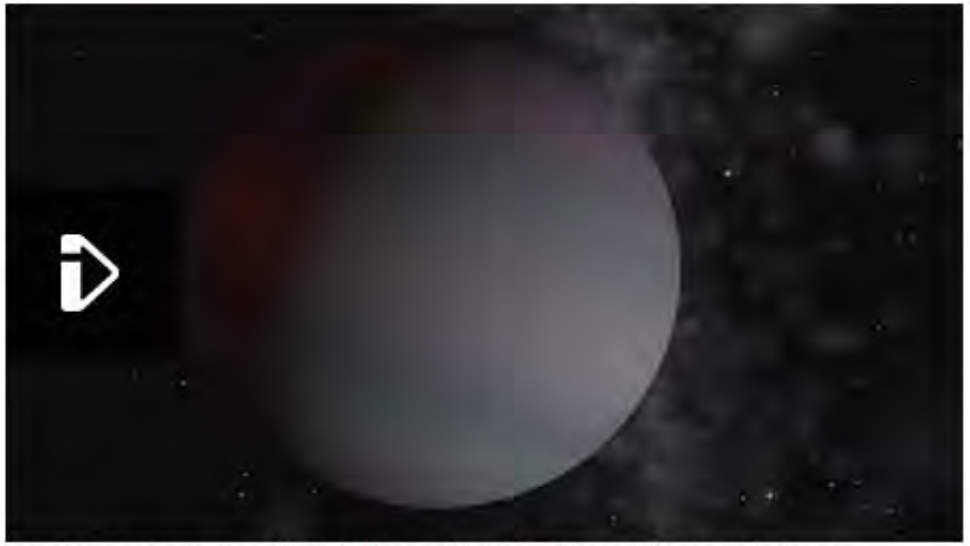
BBC Mobile NEWS

18 January 2012 Last updated at 22:16

251

Stargazing viewer in planet coup

By Jonathan Amos
Science correspondent, BBC News



Amateur astronomer Chris Holmes form Peterborough stumbled upon SPH10066540

“Crowd-sourced” astronomy

Researchers need help to manage their data.

This is a really exciting opportunity for libraries…..

with a bit of re-engineering

http://www.flickr.com/photos/49397559@N02/5899381202/

1. Leadership

(Getting attention…)

Six reasons why *you* should care about managing your research data

1. Risk: where is <u>your</u> 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

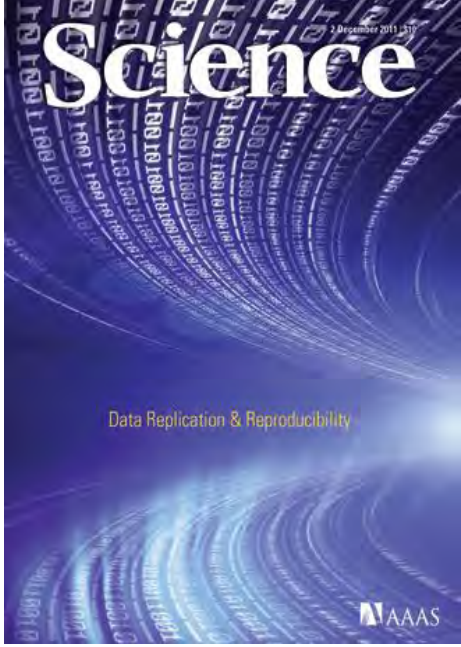
Tobacco firm demands university's research on children and smoking

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

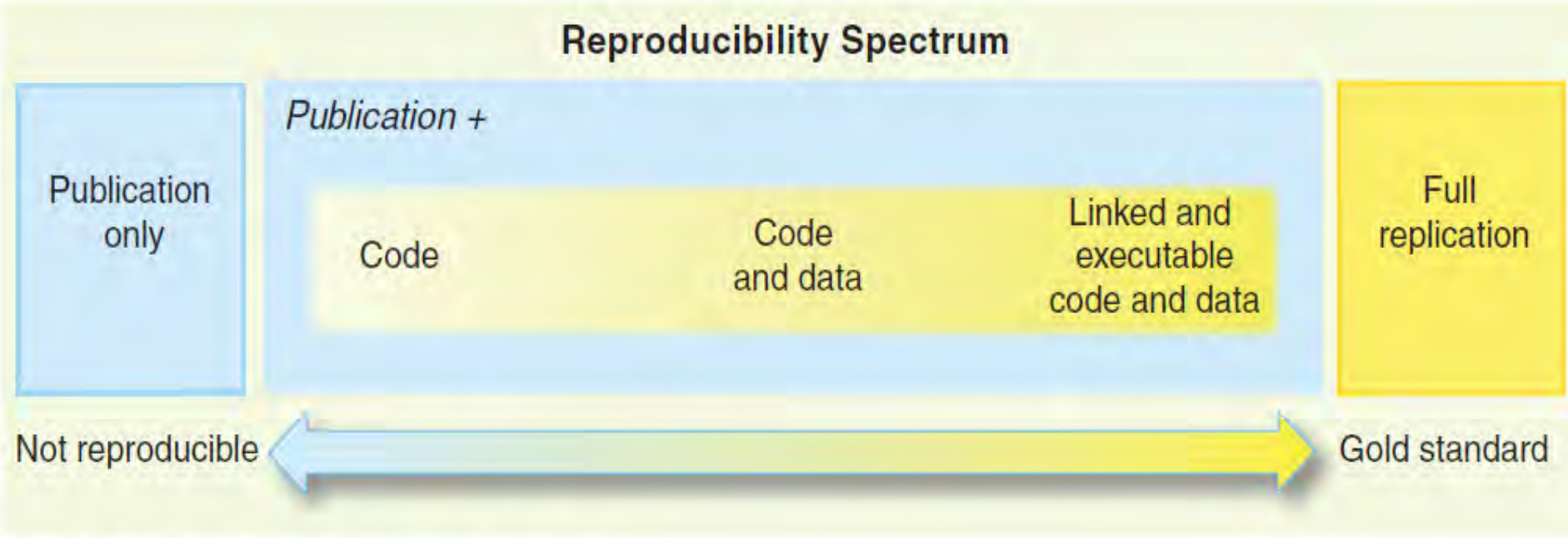


Data Replication & Reproducibility

PERSPECTIVE

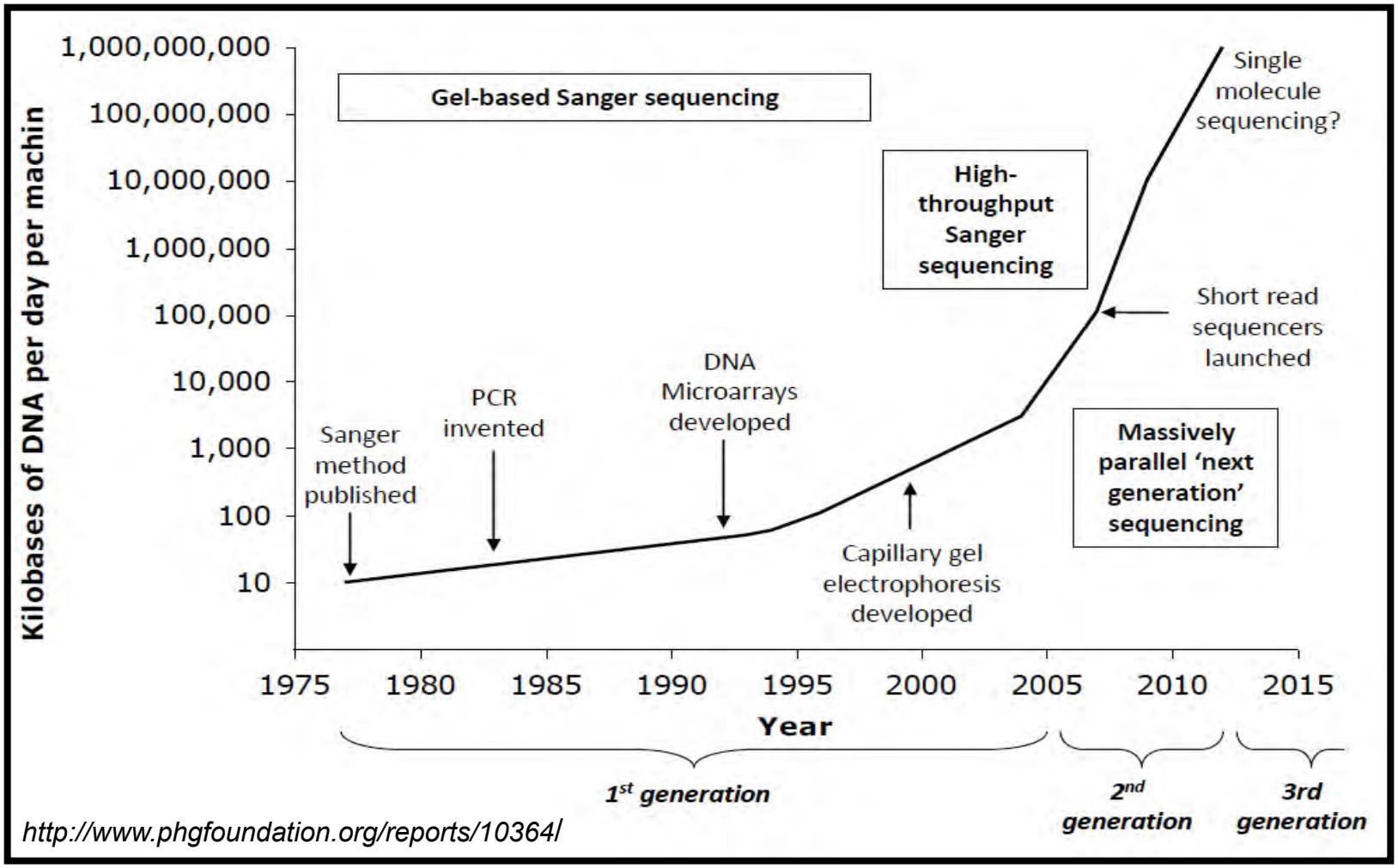
Reproducible Research in Computational Science

Roger D. Peng



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"

5.Partnerships

The New York Times

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
Pioneering research and skills

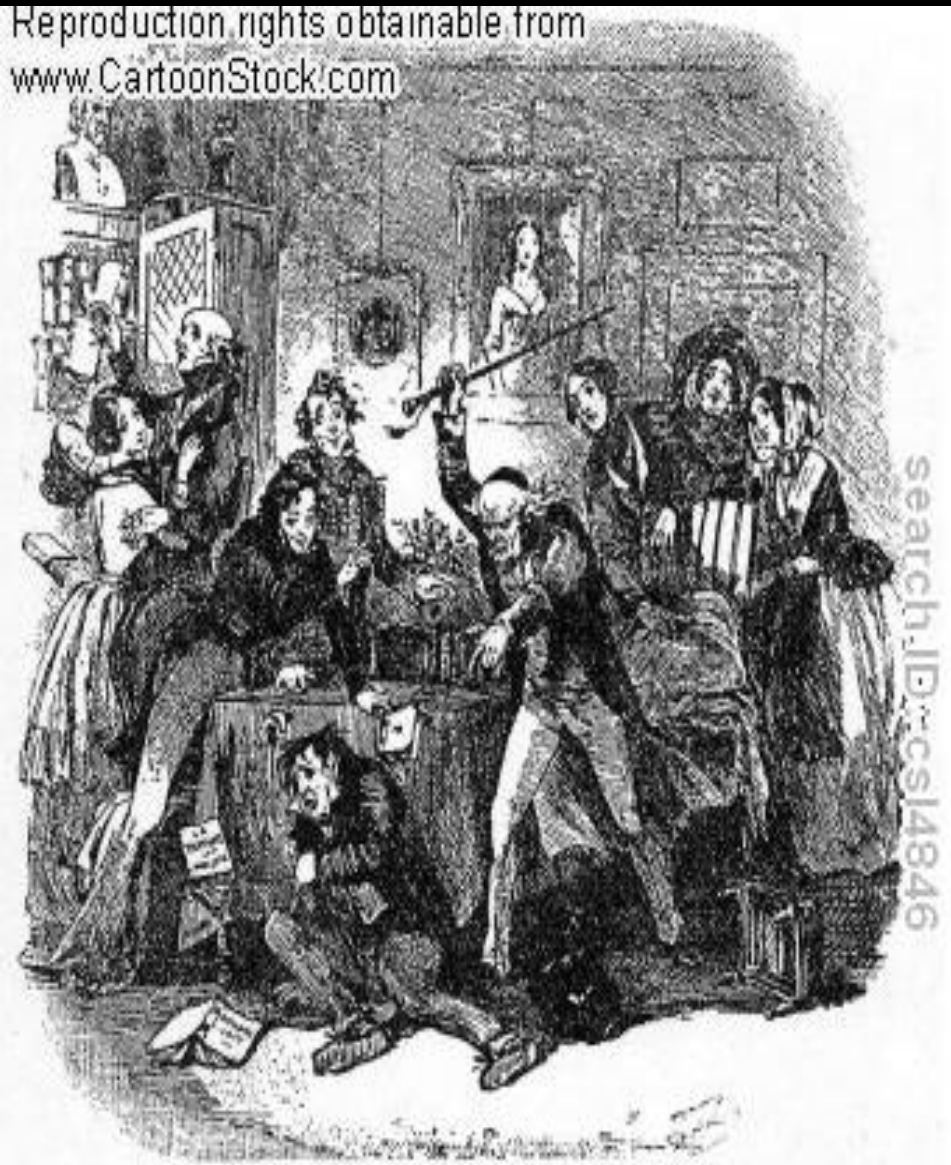
Engineering and Physical Sciences Research Council

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

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



WARM RECEPTION OF MR. PECKSNIFF BY HIS VENERABLE FRIEND

Old Martin, with his burning indignation crowded into one vehement burst, struck him down upon the ground with a well-directed blow.
Chuzzlewit, chap. lii

http://www.cartoonstock.com/lowres/csl4846l.jpg

Sticks

…and Carrots

http://www.flickr.com/photos/darshan-shah/6237564870/

2. Research Data Management 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

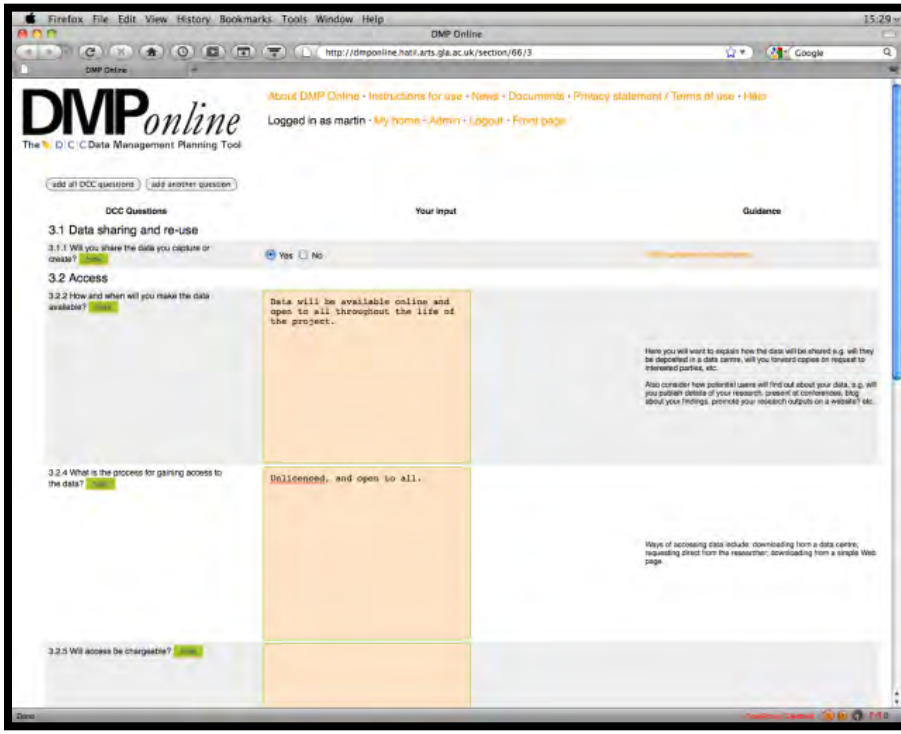
DCC

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

Data management plans

DMPonline

The DCC Data Management Planning Tool

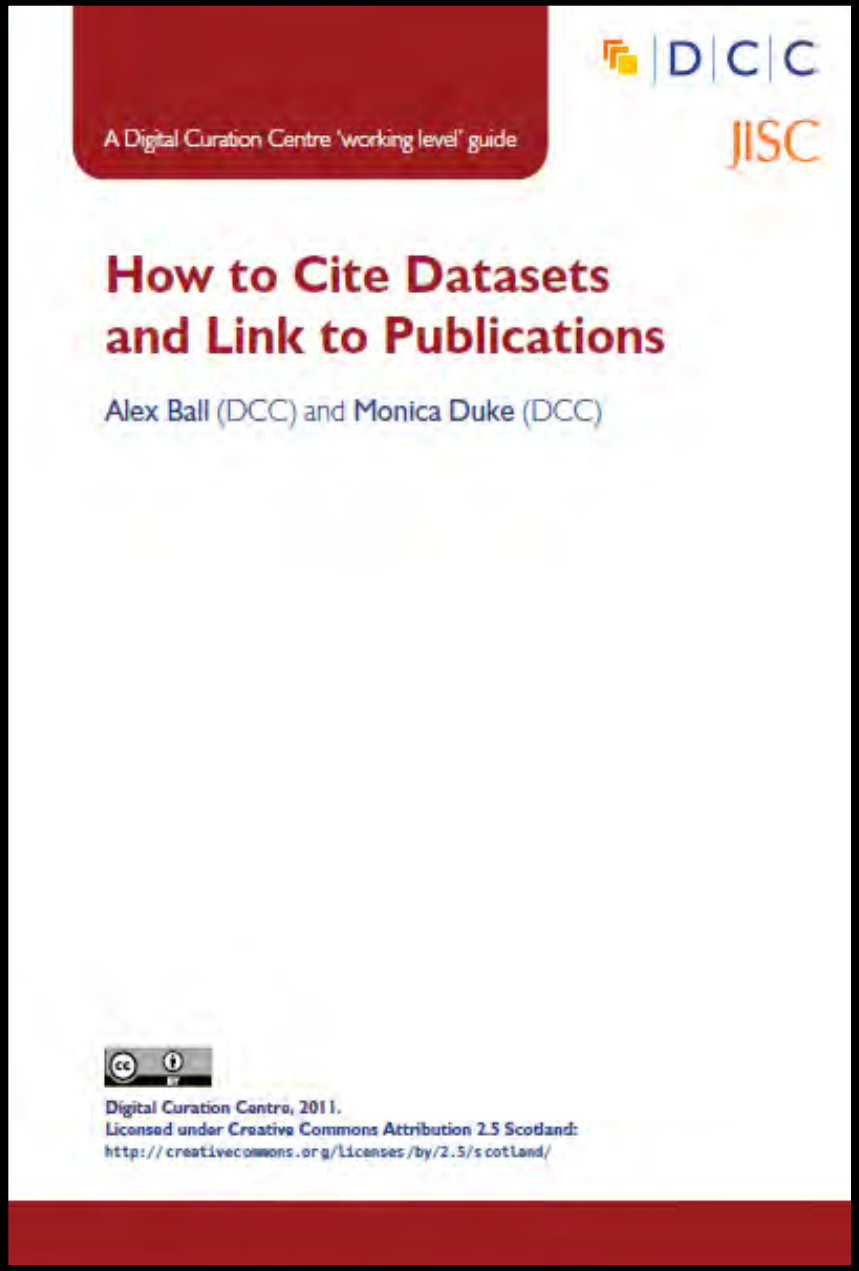


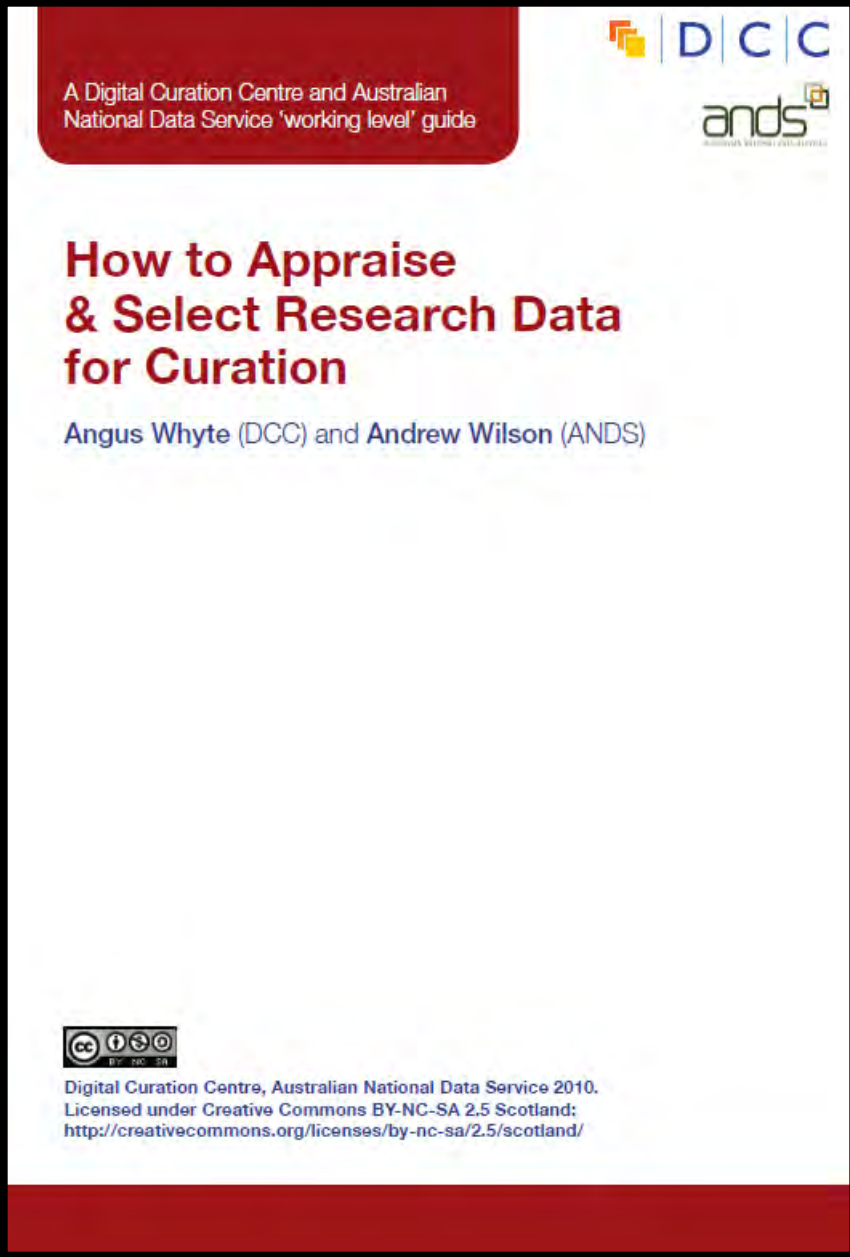
DMPTool

Guidance and Resources for your Data Management Plan

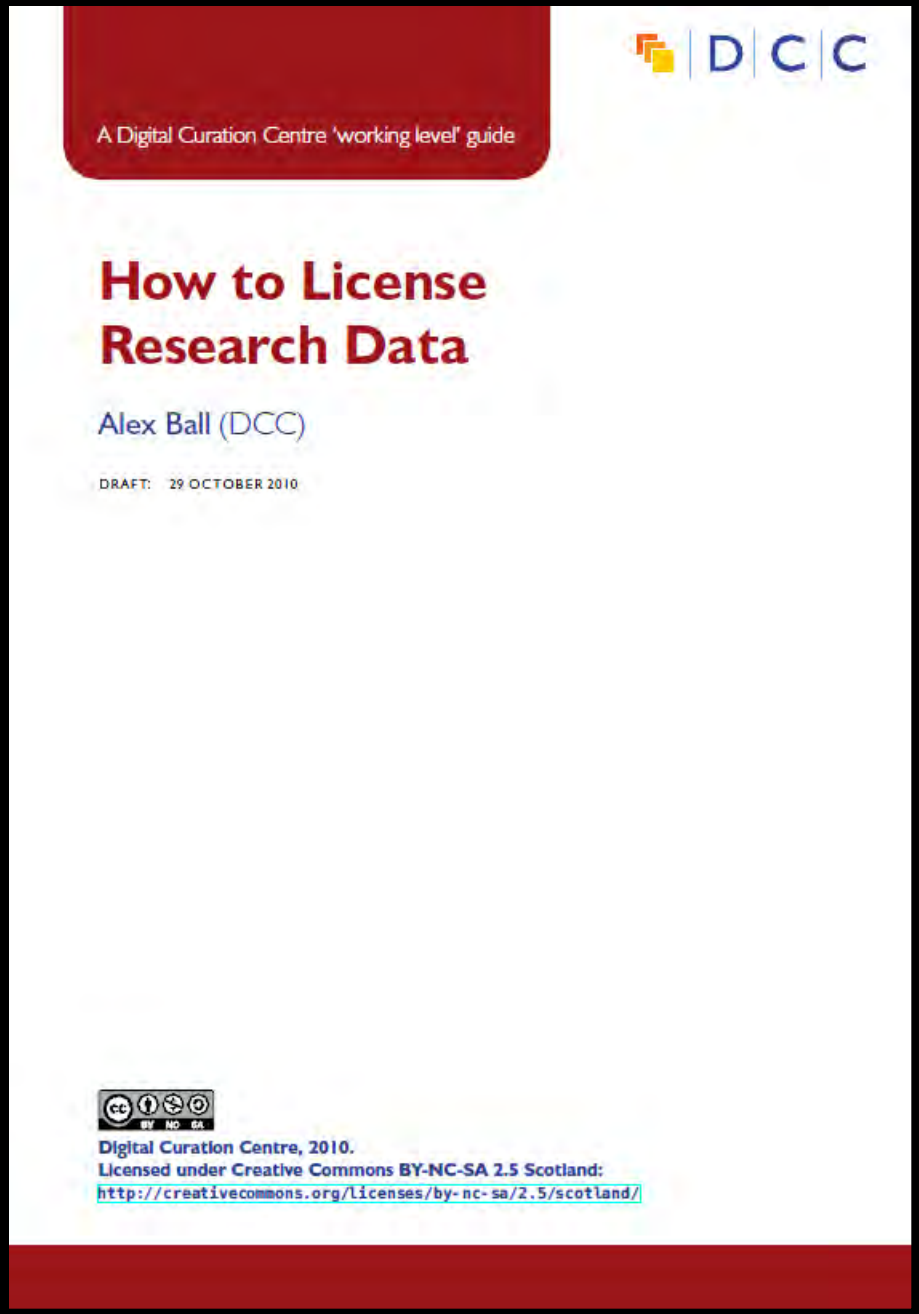
- Advocacy & Training
  - **Informatics:** disciplinary metadata schema, standards, formats, identifiers, ontologies
  - **Storage:** file-store, cloud, data centres, funder policy
  - **Access:** embargoes, FOI

What data to keep





How to cite data



Data Licensing

Bespoke licences
Standard licences
Multiple licensing
Licence mechanisms

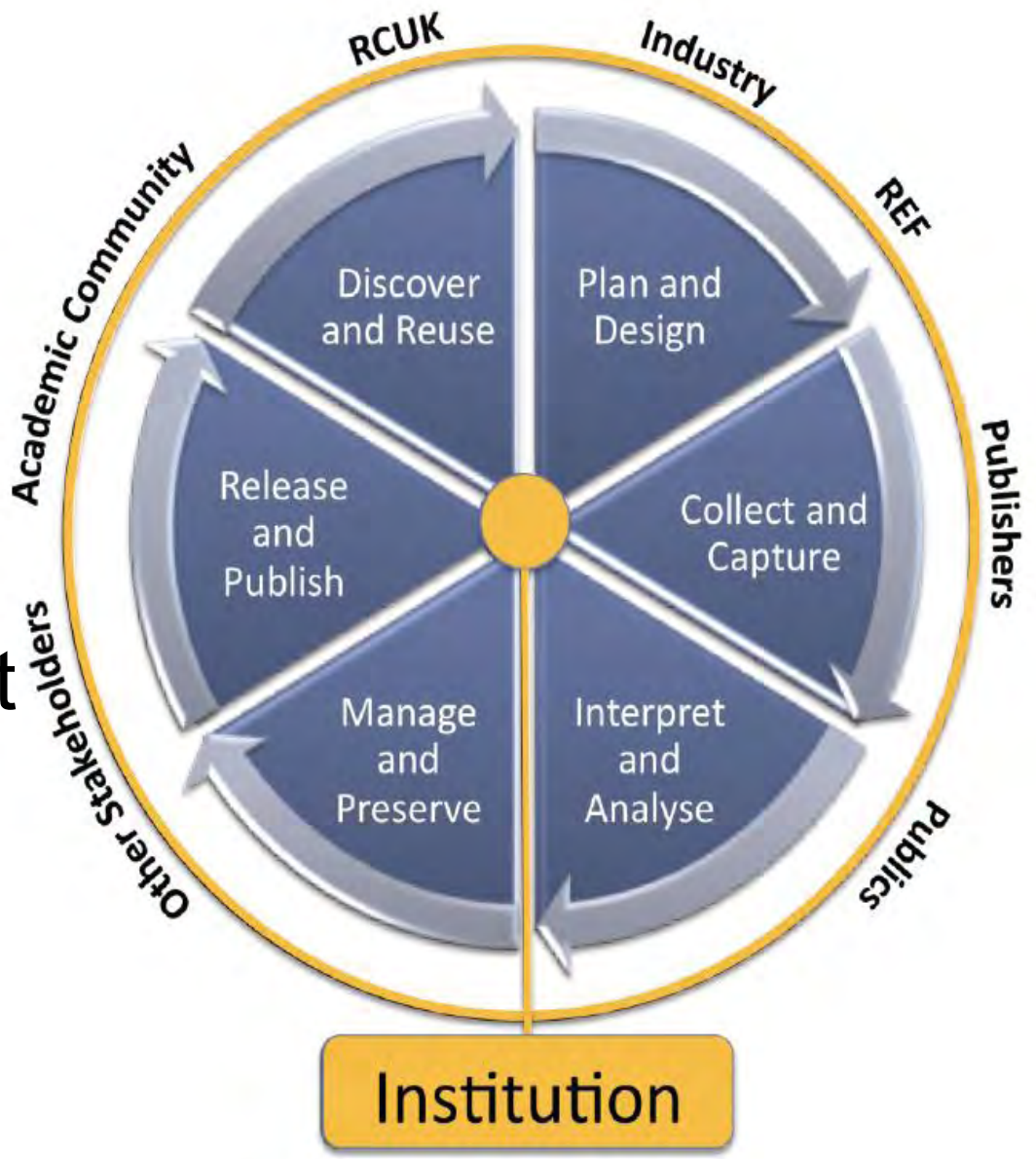
Tools to track impact

http://total-impact.org/

Research360@Bath

JISC

- **Partnership approach**
  - UKOLN-DCC
  - Library
  - IT services
  - Research Support Office
  - Doctoral Training Centres



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

Partnership approach

Library & institutional stakeholders

- 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, Ariadne Issue 68, 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

Data roles

Liz Lyon, Informatics Transform, Ariadne Issue 68, 2012

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

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

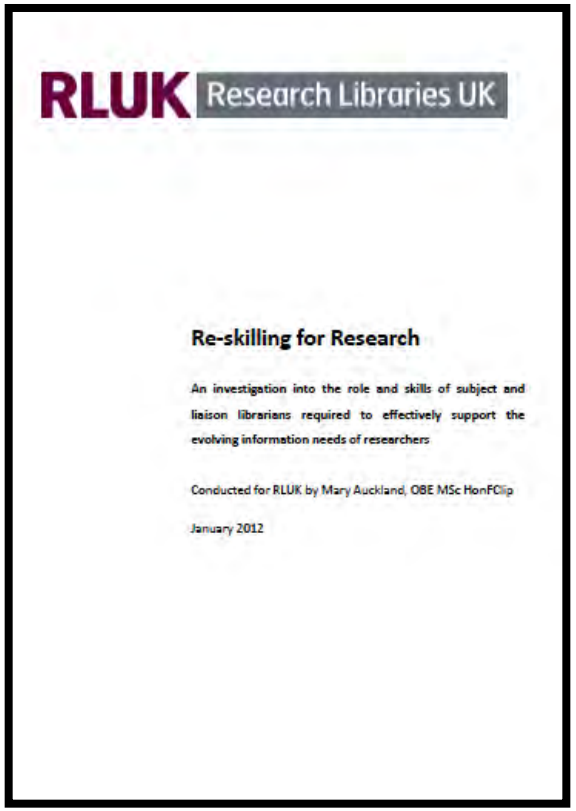
Full mapping : Informatics Transform, Ariadne Issue 68, 2012

3. Developing 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

Pause for reflection….

- Skills shortage for data informatics?
- Reposition LIS curriculum?
- LIS entry requirements?
- Get credit for informatics work?

Lyon, Informatics Transform, Ariadne 2012

Play for action….

1. Define core components of data informatics

   - Visualisation e.g. VisTrails
   - Workflow e.g. Taverna
   - Analysis e.g. R

Lyon, Informatics Transform, Ariadne 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

Play for action….

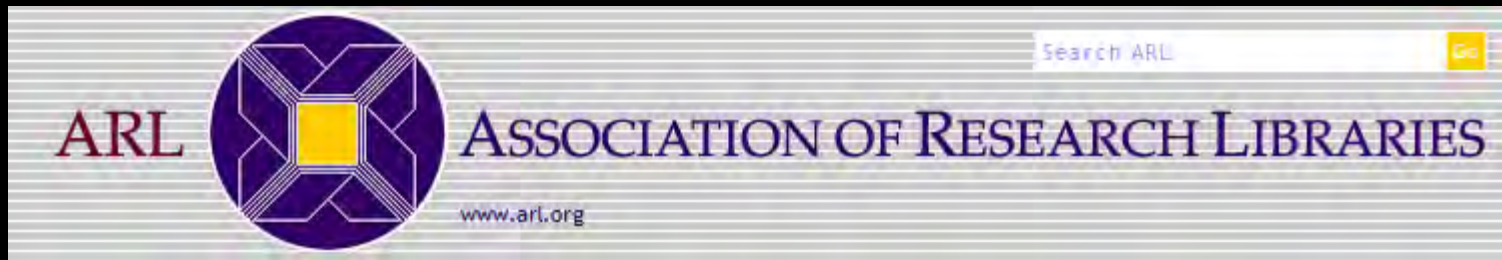
2. Analyse LIS entry qualifications & increase STEM entrants

Target

- Biologists
- Chemists
- Mathematicians

Lyon, Informatics Transform, Ariadne 2012

Let's get together



Play for action….

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

- Global awareness campaign
- Career incentives
- Benchmark good practice

Lyon, Informatics Transform, Ariadne 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

Data journalist?

New York Times Data Artist in Residence, Jer Thorp Joins Stellar Cast of Speakers at TEDxVancouver 2011

1 · No Comments »

Jer Thorp: Hope / Crisis, NYT Word Frequency

The New York Times

Data artist?

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

“Big Data” Data scientist

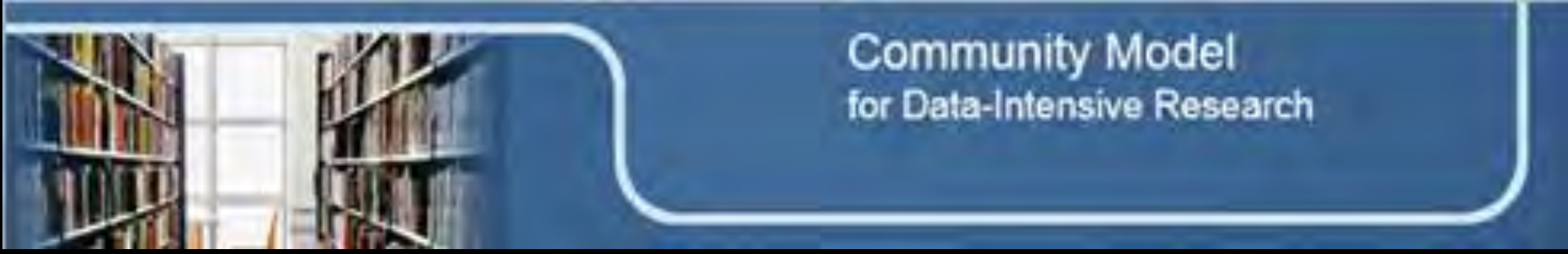
EMC²

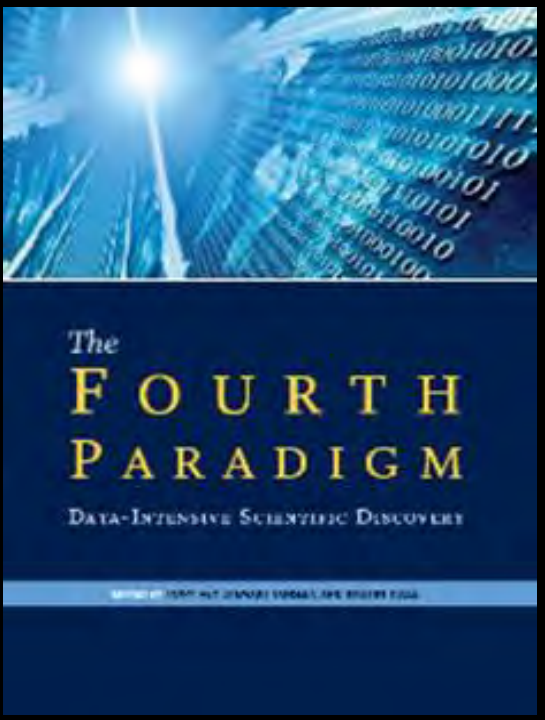
Data Science Revealed community survey

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

For a University, research data is a key element of “Big Data”.

Managing research data effectively will give business advantage.





Data-intensive research

- Intelligence
- Decision-making
- Planning
- Investment
- Capacity
- Capability

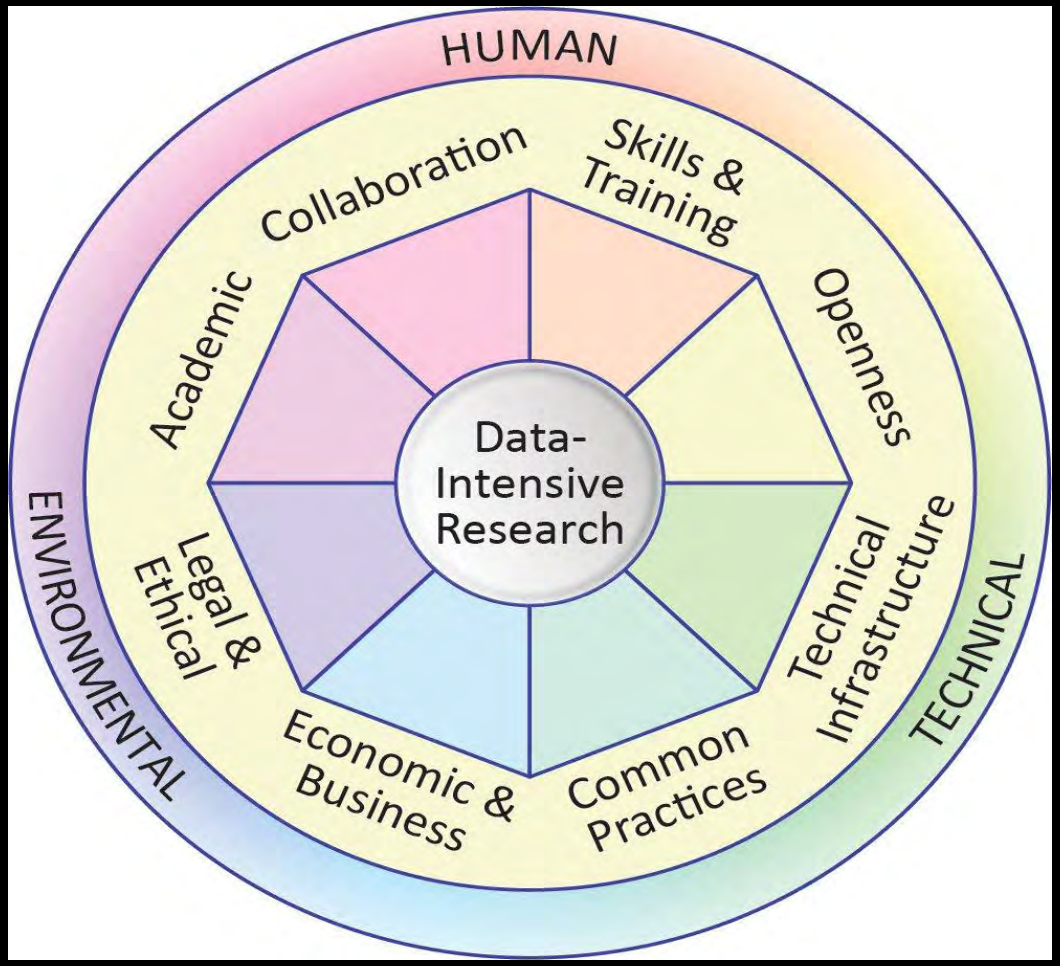
UKOLN

UNIVERSITY OF BATH

Research Connections

http://communitymodel.sharepoint.com/

Community Capability Model Framework CCMF



- Research Funders
- Institutions
- Research leaders/PIs

“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

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

Thank you!

Slides

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

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