The changing data landscape

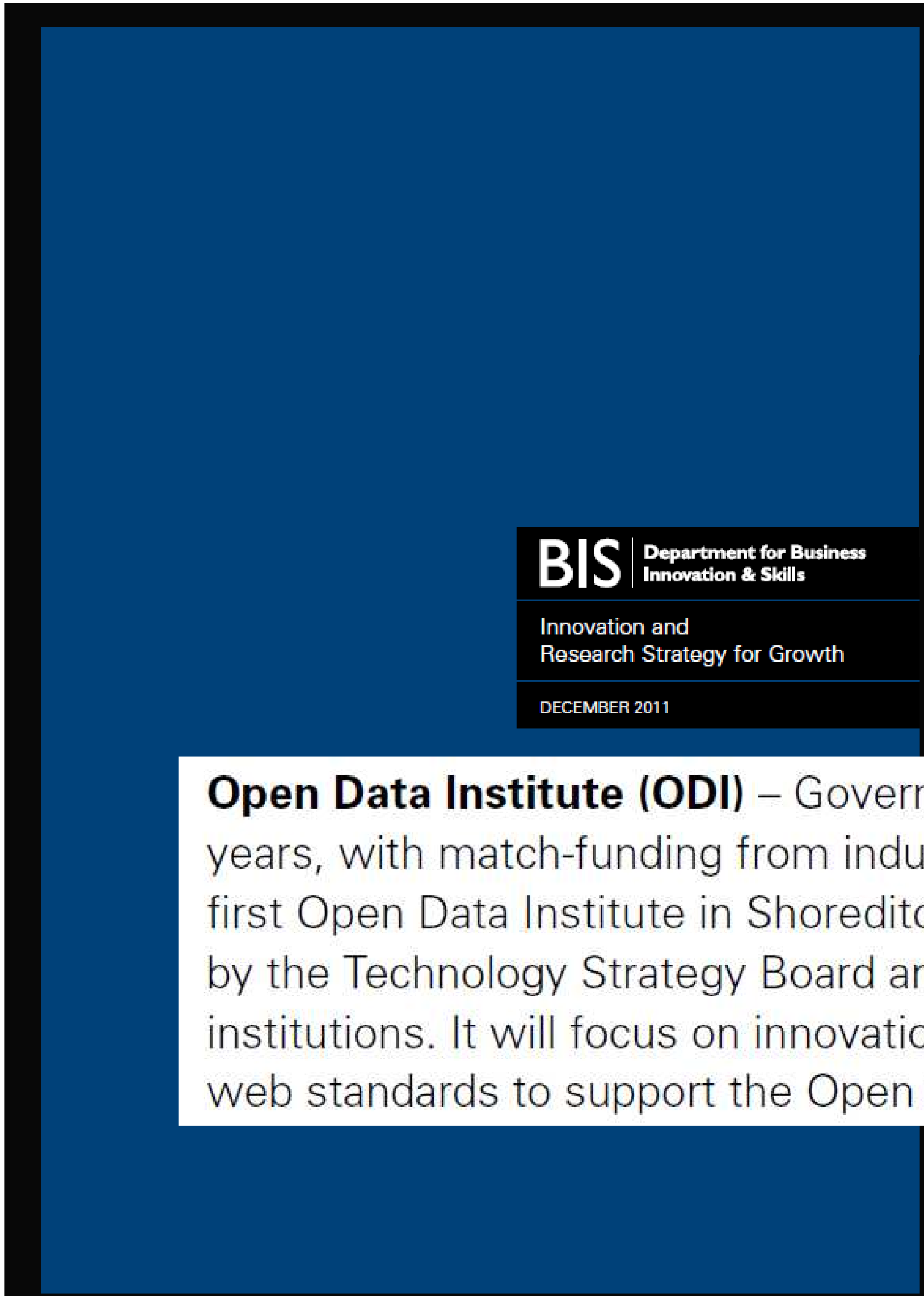
Kevin Ashley, Director DCC – director@dcc.ac.uk
Dr Liz Lyon, Associate Director, UK Digital Curation Centre
Director, UKOLN, University of Bath, UK

DCC Regional Roadshow, East Midlands, February 2012

.

BIS Department for Business Innovation & Skills

...open data



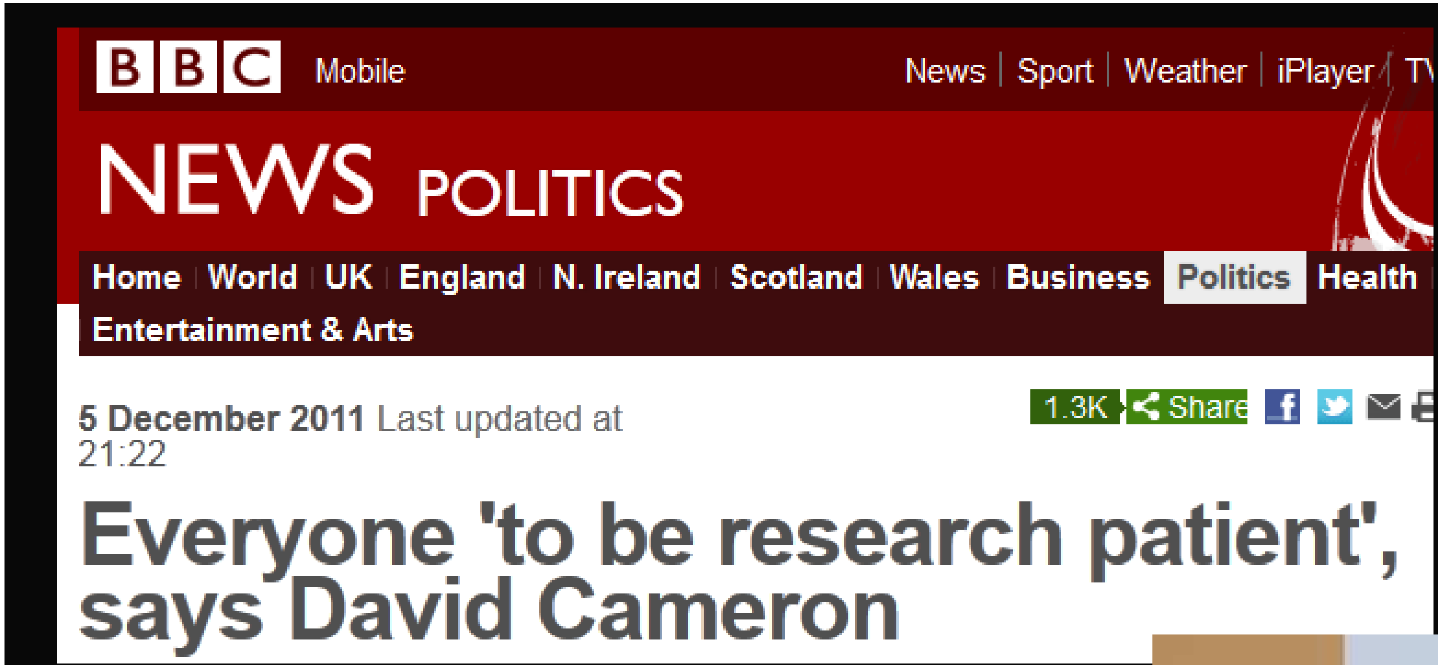
BIS Department for Business Innovation & Skills

Innovation and Research Strategy for Growth

DECEMBER 2011

Open Data Institute (ODI) – Government will provide up to £10 million over five years, with match-funding from industry and academia, to establish the world's first Open Data Institute in Shoreditch, East London. The ODI will be developed by the Technology Strategy Board and will involve businesses and academic institutions. It will focus on innovation, commercialisation and the development of web standards to support the Open Data Agenda.

http://www.bis.gov.uk/innovatingforgrowth



BBC Mobile

News | Sport | Weather | iPlayer | T

NEWS POLITICS

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

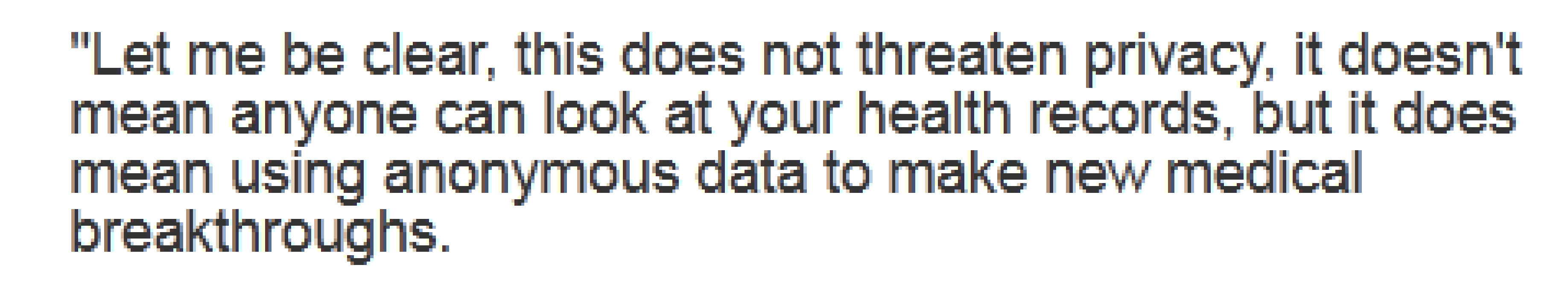
Entertainment & Arts

1.3K Share

5 December 2011 Last updated at 21:22

Everyone 'to be research patient', says David Cameron

..personal data



"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.

…data replication & reproducibility

2 Dec 2011 issue

http://www.sciencemag.org/content/334/6060.toc

SPECIALSECTION

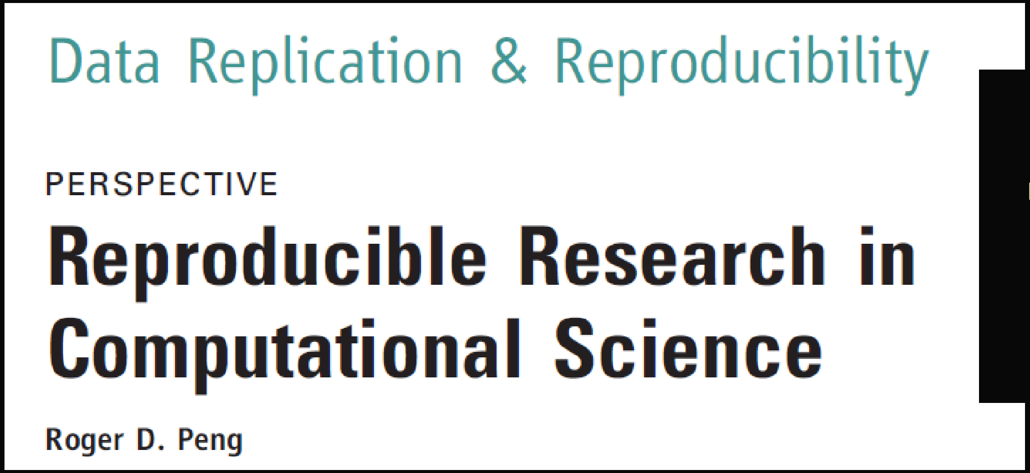
INTRODUCTION

Again, and Again, and Again …

REPLICATION—THE CONFIRMATION OF RESULTS AND CONCLUSIONS FROM ONE STUDY obtained independently in another—is considered the scientific gold standard. New tools and technologies, massive amounts of data, long-term studies, interdisciplinary approaches, and the complexity of the questions being asked are complicating replication efforts, as are increased pressures on scientists to advance their research. The five Perspectives in this section (and associated News and Careers stories, Readers' Poll, and Editorial) explore some of the issues associated with replicating results across various fields.

Data Replication & Reproducibility

...data gold standard



Data Replication & Reproducibility

PERSPECTIVE

Reproducible Research in Computational Science

Roger D. Peng

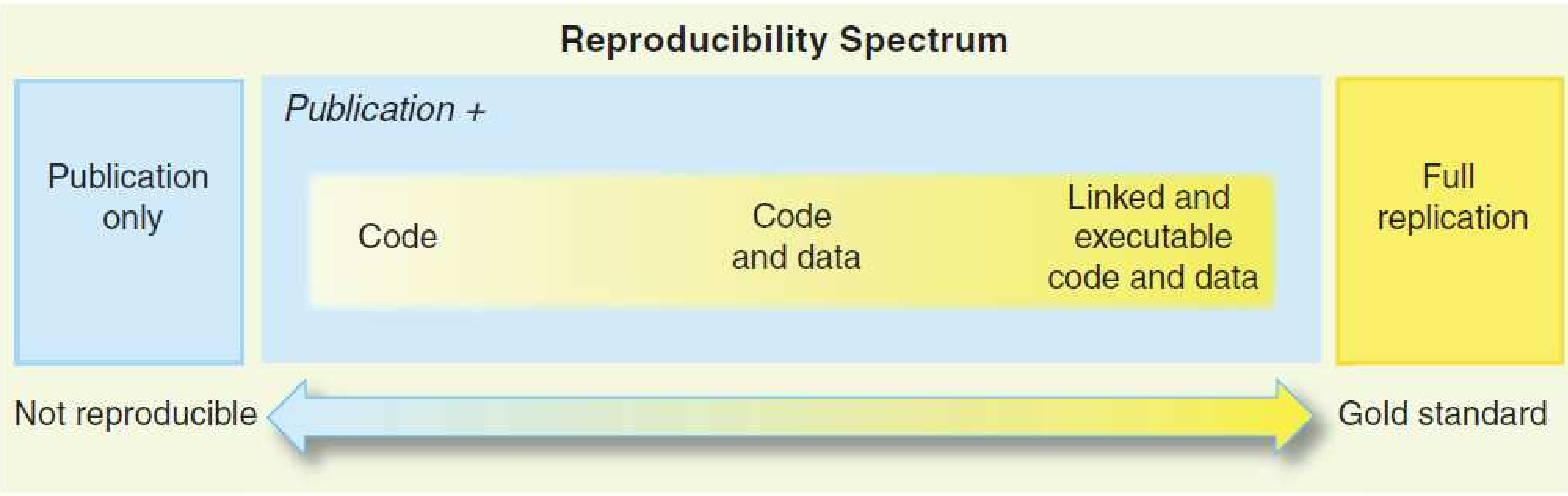


Fig. 1. The spectrum of reproducibility.

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

Perspectives

- Environmental scan
  - Scale and complexity
  - Open science
- Policy
  - Funders
  - Institutions
  - Ethics & IP
- Practice Challenges
  - Storage
  - Incentives
  - Costs & Sustainability

http://www.flickr.com/photos/thegreenalbum/3997609142/

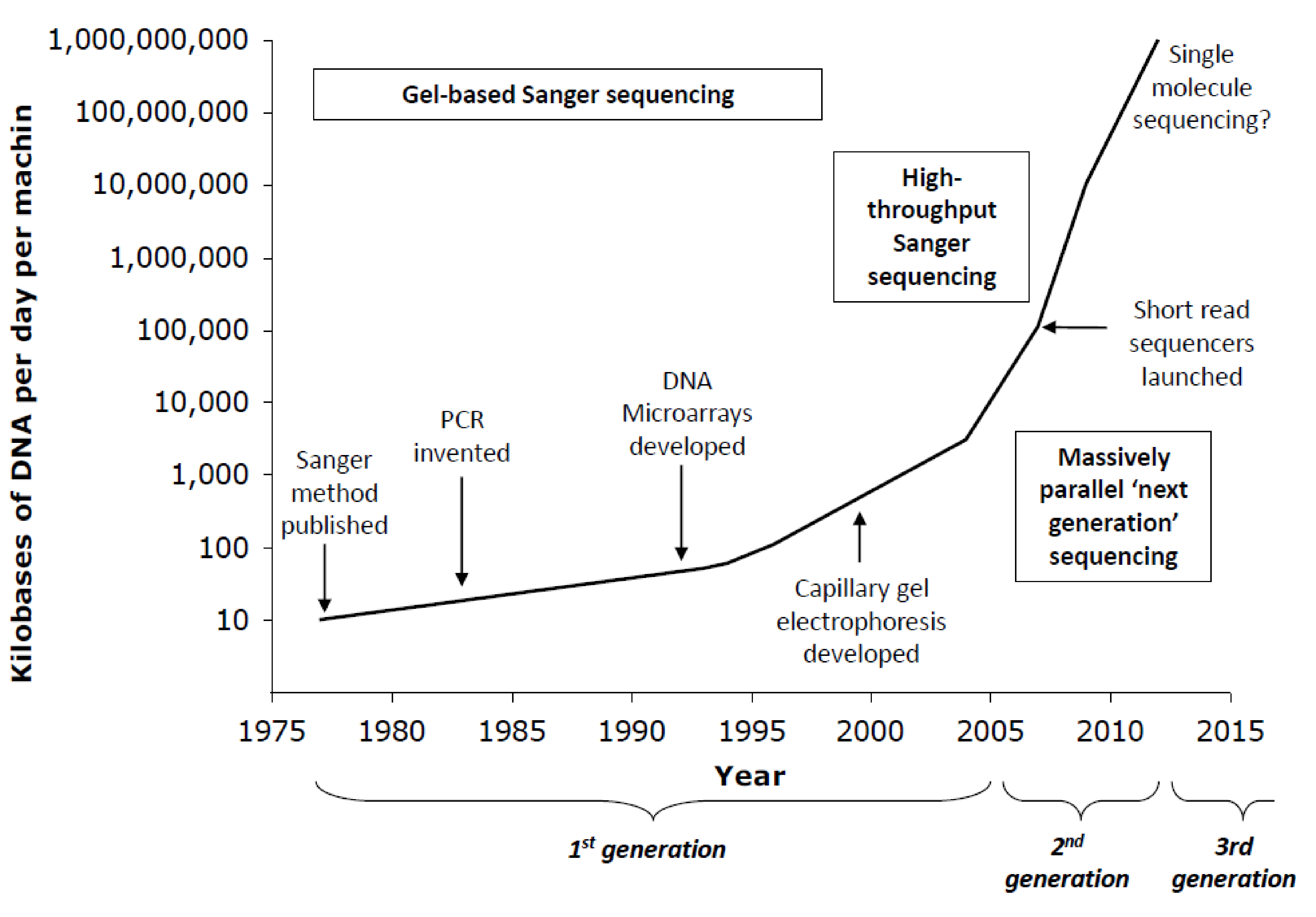
“The cost of sequencing DNA has taken a nosedive...and is now dropping by 50% every 5 months”

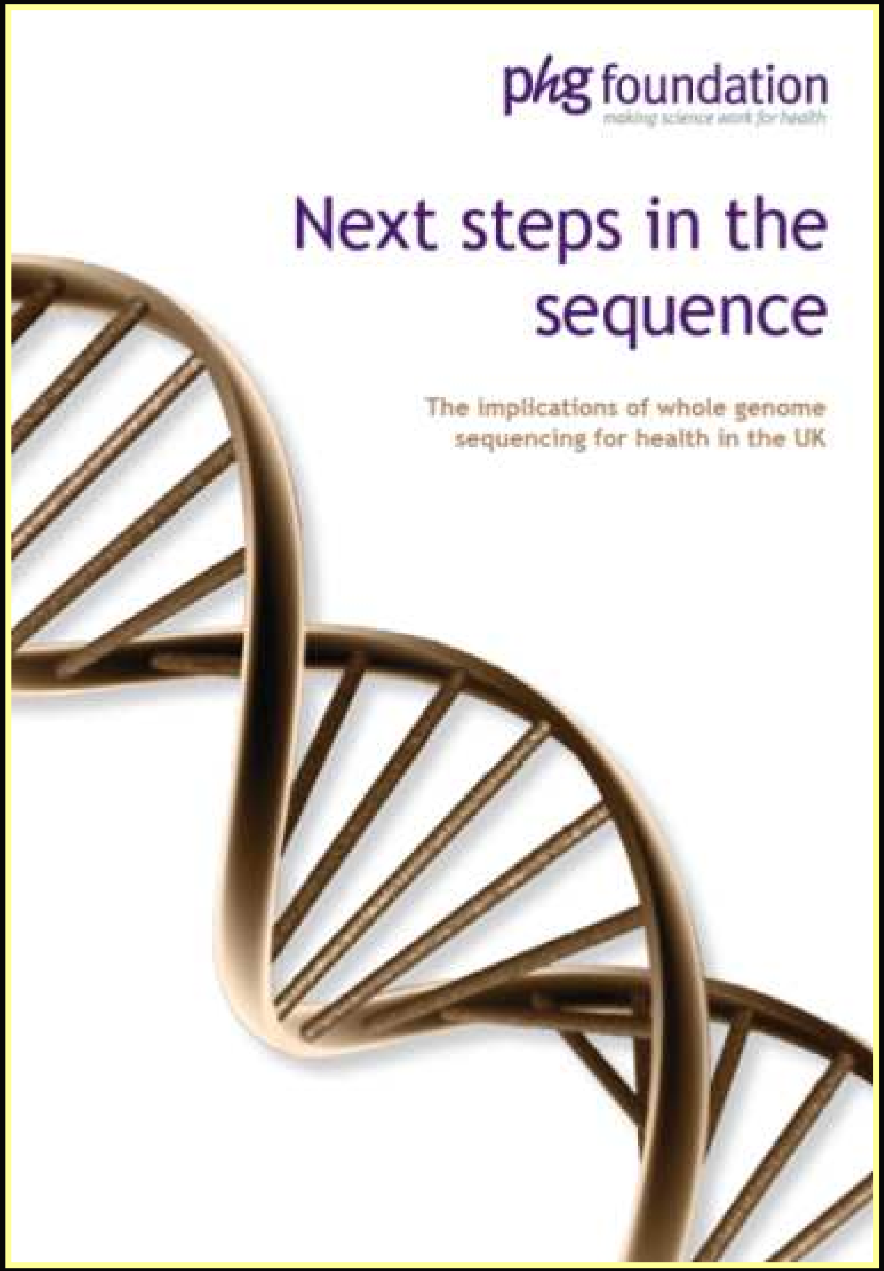
“Raw image files for a single human genome have been estimated at 28.8 terabytes, which is approaching 30,000 gigabytes”

"The 1000 Genomes Project generated more DNA sequence data in its first 6 months than GenBank had accumulated in its entire 21 year existence”

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

An explosion of data…





phg foundation

making science work for health

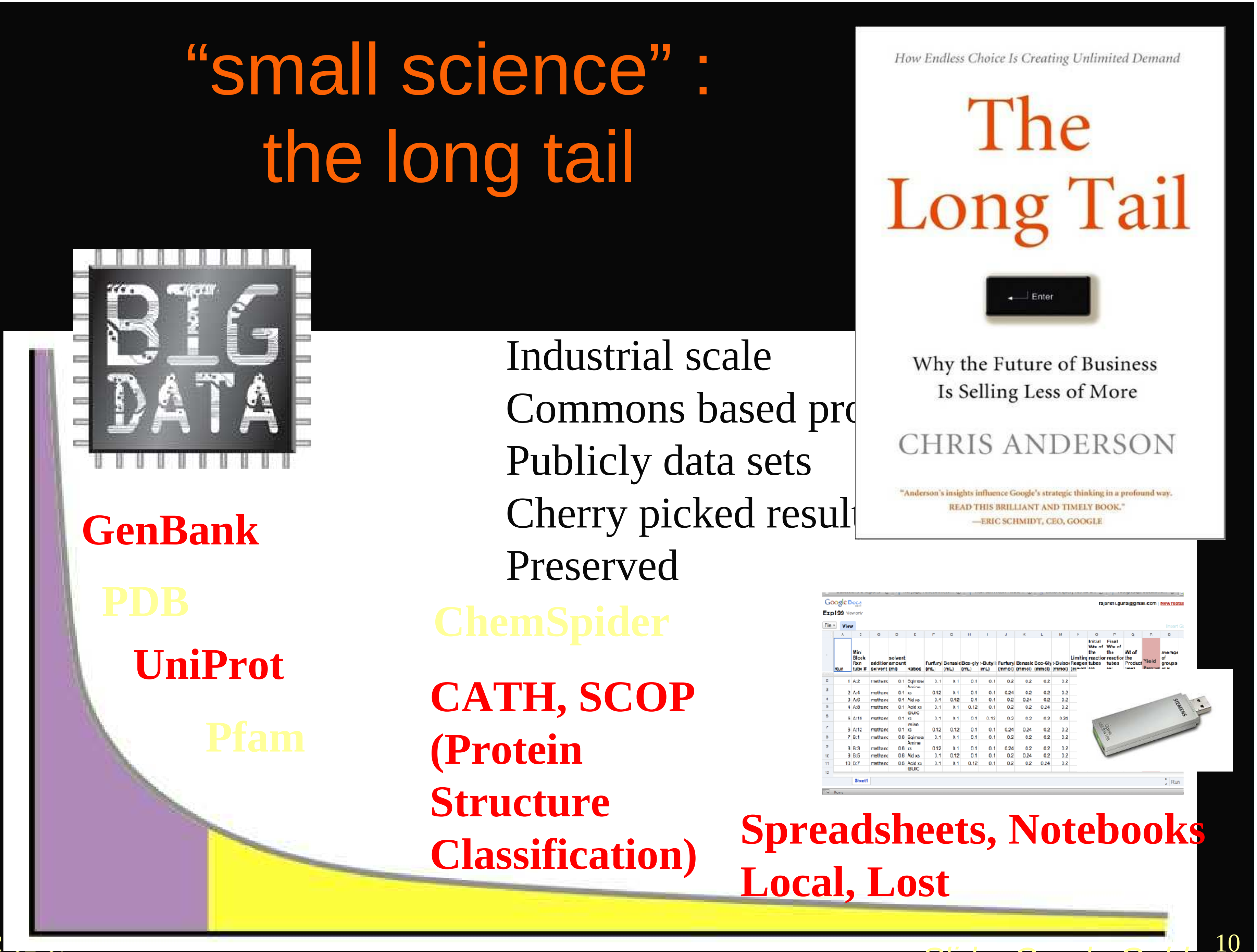
Next steps in the sequence

The implications of whole genome sequencing for health in the UK

…resulting in significant implications for the NHS

http://www.phgfoundation.org/reports/10364/

“small science” : the long tail



Industrial scale
Commons based pro
Publicly data sets
Cherry picked result
Preserved

GenBank

PDB

UniProt

Pfam

ChemSpider

CATH, SCOP (Protein Structure Classification)

Spreadsheets, Notebooks
Local, Lost

How Endless Choice Is Creating Unlimited Demand

The Long Tail

Why the Future of Business Is Selling Less of More

CHRIS ANDERSON

"Anderson's insights influence Google's strategic thinking in a profound way. READ THIS BRILLIANT AND TIMELY BOOK."
—ERIC SCHMIDT, CEO, GOOGLE

Google Docs

rajarshi.guha@gmail.com | New featu

Exp199 View only

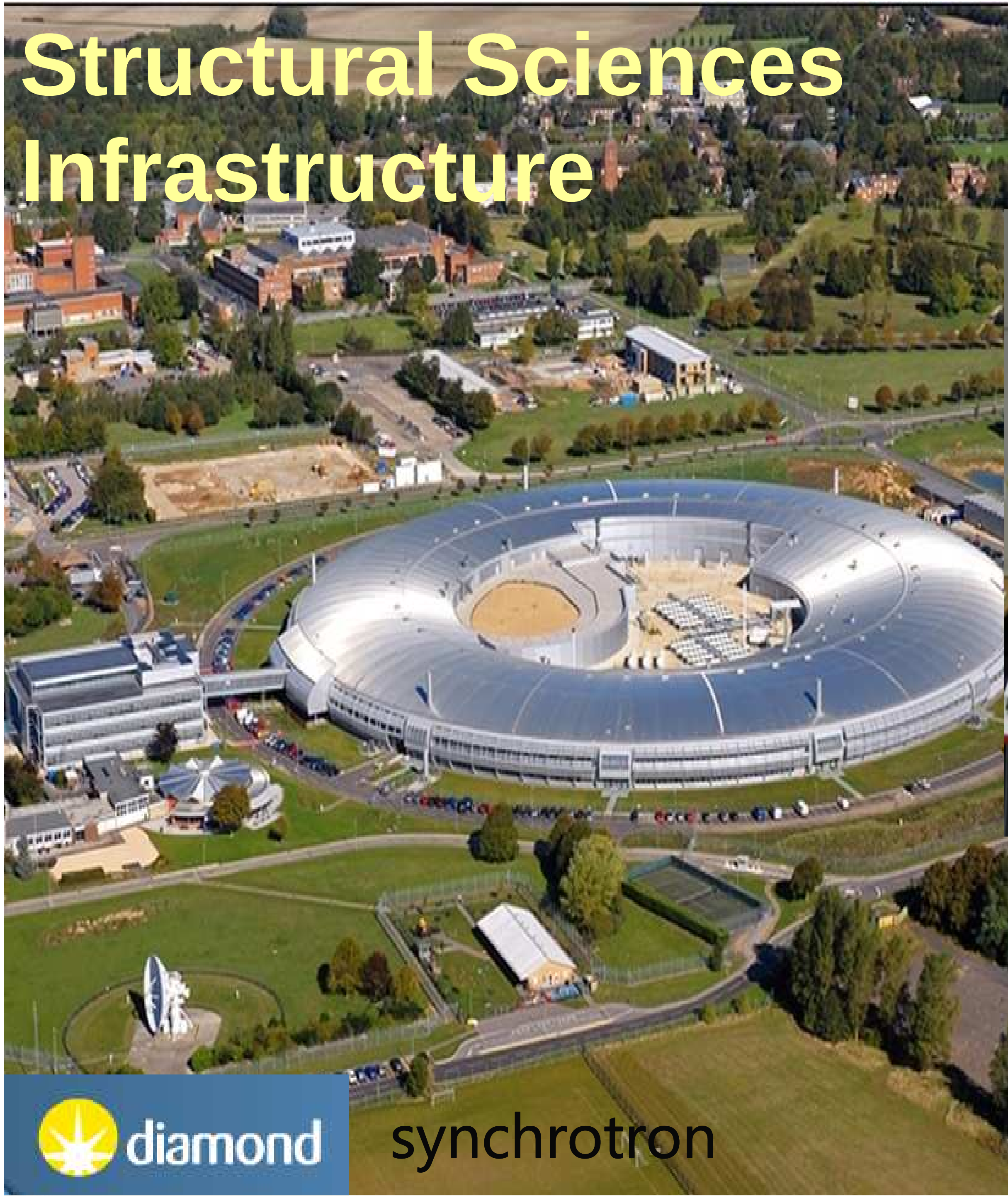
File View

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Run	Mini Block Rxn tube #	additior solvent	solvent amount (ml)	Ratios	Furfury (mL)	Benzalc (mL)	Boc-gly (mL)	t-Buty (mL)	Furfury (mmol)	Benzalc (mmol)	Boc-Gly (mmol)	t-Buiso (mmol)	Limiting Reagen (mmol)	Initial Wts of the reaction tubes	Final Wts of the reaction tubes	Wt of the Product	Yield	average of groups
2	1	A:2	methanc	0.1	Eqimola	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2						
3	2	A:4	methanc	0.1	Amine xs	0.12	0.1	0.1	0.1	0.24	0.2	0.2	0.2						
4	3	A:6	methanc	0.1	Ald xs	0.1	0.12	0.1	0.1	0.2	0.24	0.2	0.2						
5	4	A:8	methanc	0.1	Acid xs	0.1	0.1	0.12	0.1	0.2	0.2	0.24	0.2						
6	5	A:10	methanc	0.1	tBUIC xs	0.1	0.1	0.1	0.12	0.2	0.2	0.2	0.24						
7	6	A:12	methanc	0.1	Imine xs	0.12	0.12	0.1	0.1	0.24	0.24	0.2	0.2						
8	7	B:1	methanc	0.6	Eqimola	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2						
9	8	B:3	methanc	0.6	Amine xs	0.12	0.1	0.1	0.1	0.24	0.2	0.2	0.2						
10	9	B:5	methanc	0.6	Ald xs	0.1	0.12	0.1	0.1	0.2	0.24	0.2	0.2						
11	10	B:7	methanc	0.6	Acid xs	0.1	0.1	0.12	0.1	0.2	0.2	0.24	0.2						
12					tBUIC														

Sheet1 Run

Slide: Carole Goble

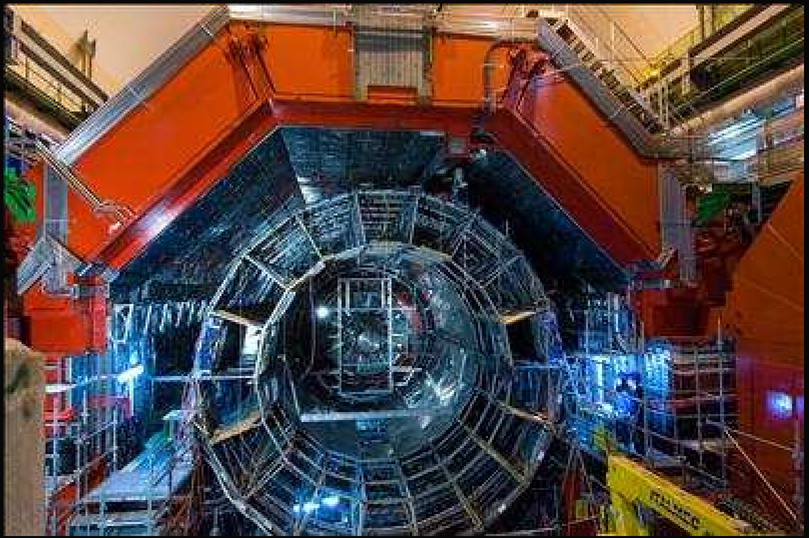
Structural Sciences Infrastructure

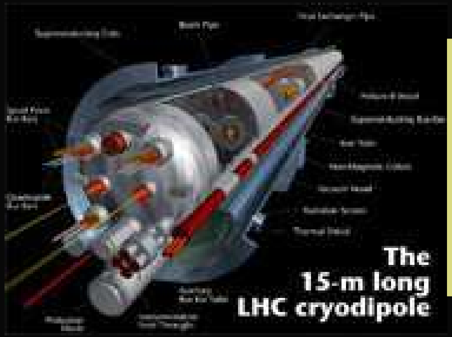


Diamond Podcast

diamond synchrotron

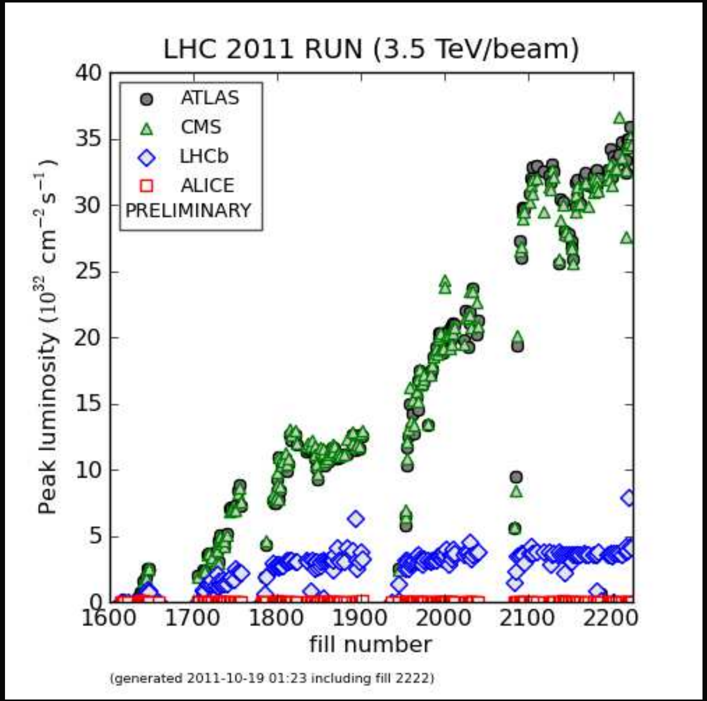
Challenges of scale and complexity – The Large Hadron Collider





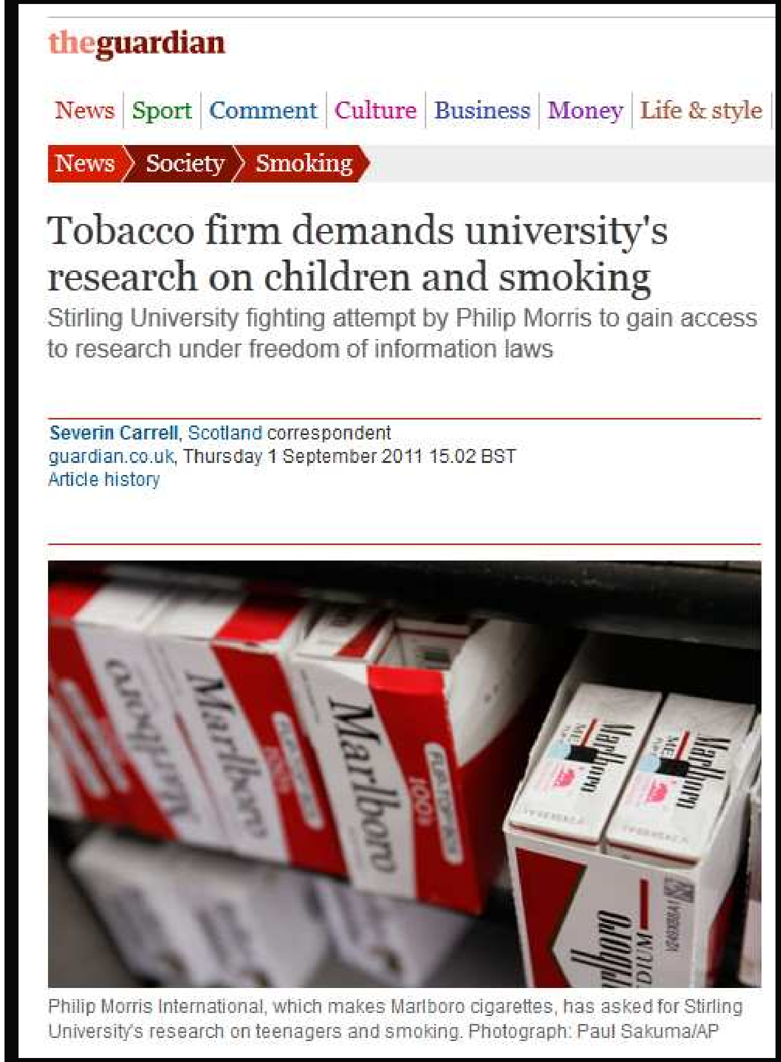
1 cryodipole X 1200 = 27km LHC

- Predicted generation of around 15 petabytes of data (i.e. 15 million gigabytes) annually



Data access is headline news

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

2011, proposed amendment to FOI legislation would enable institutions to claim exemption where research is ongoing and where data disclosure before the date of publication would substantially prejudice the research

"It's hard to overcome your personal investment… it's like giving away your baby"

"While many researchers are positive about sharing data in principle, they are almost universally reluctant in practice. ..... using these data to publish results before anyone else is the primary way of gaining prestige in nearly all disciplines."

"Data sharing was more readily discussed by early career researchers."

INCREMENTAL Project

The New York Times

Sharing of Data Leads to Progress on Alzheimer's

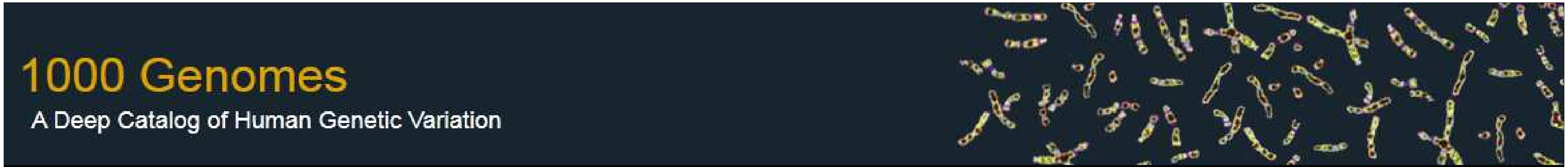
By GINA KOLATA

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

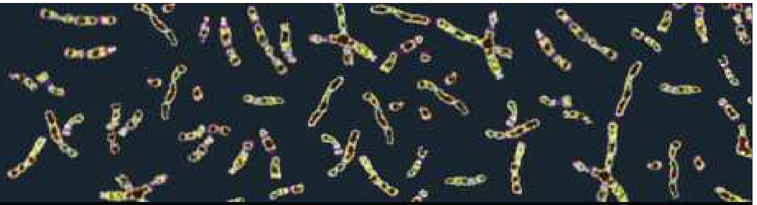


1000 Genomes

A Deep Catalog of Human Genetic Variation

A critical new component of the Project is the selection of 2,500 DNA samples from 27 populations around the world. Each participant has provided explicit consent for full and public release of DNA samples and full sequence data….

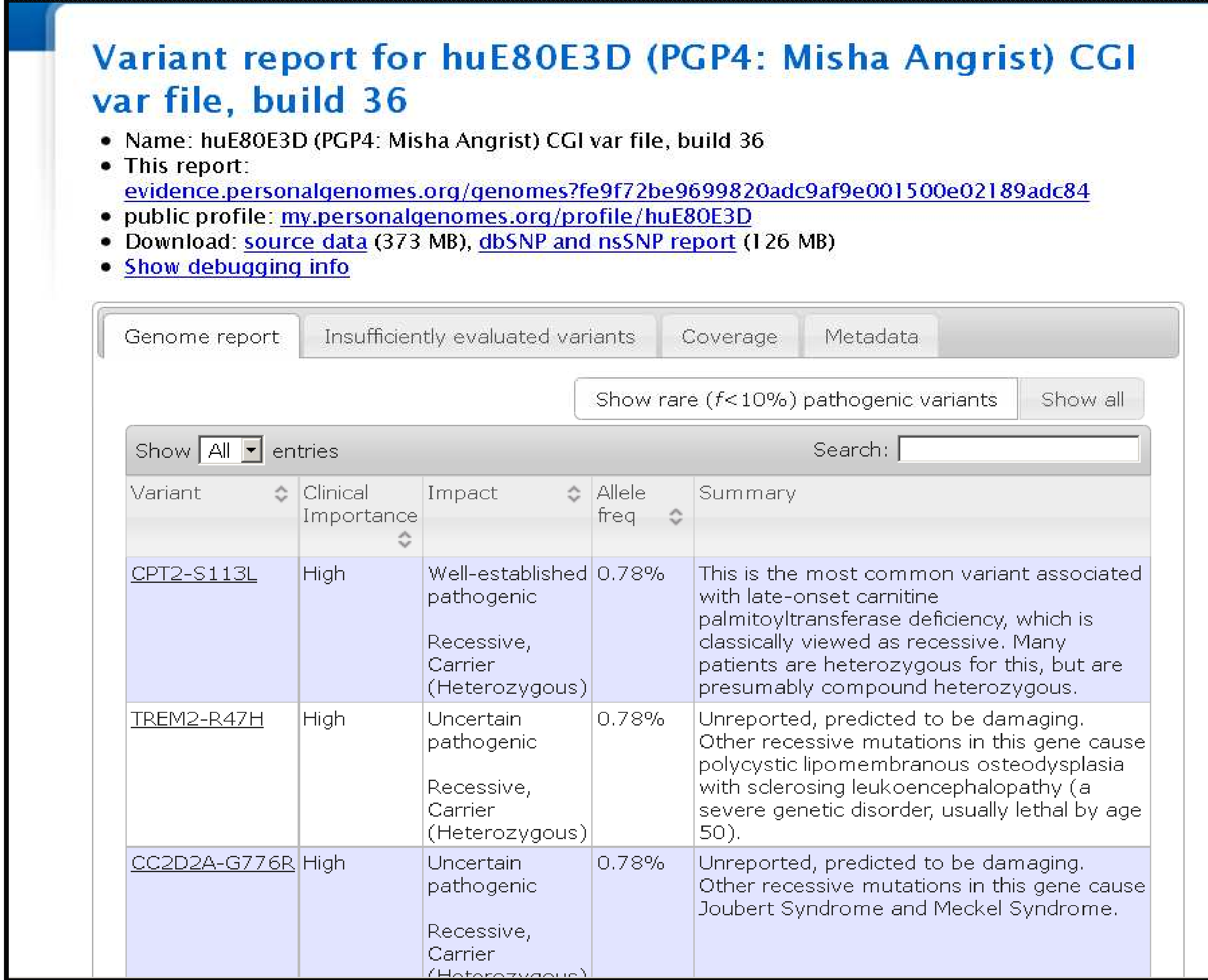
- 1000 Genomes from 27 populations around the world
- Each participant provided explicit consent for full release



"Free and open access to genome data has had a profoundly positive effect on progress."

Francis Collins, *Nature, April 2010*

Misha Angrist - a personal record...

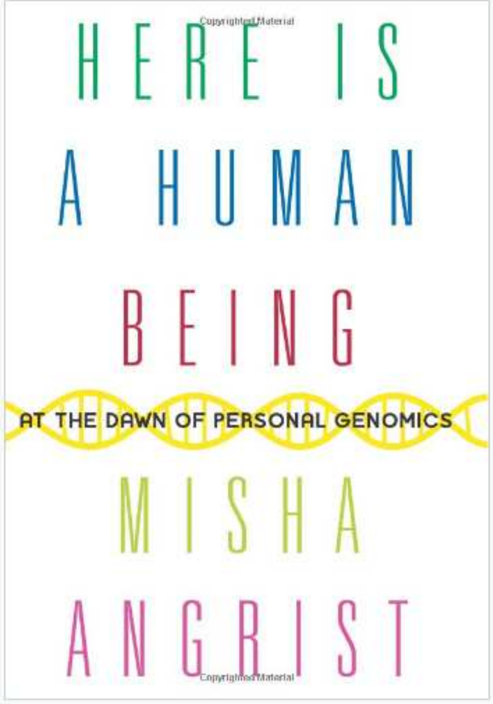


Variant report for huE80E3D (PGP4: Misha Angrist) CGI var file, build 36

- Name: huE80E3D (PGP4: Misha Angrist) CGI var file, build 36
- This report: evidence.personalgenomes.org/genomes?fe9f72be9699820adc9af9e001500e02189adc84
- public profile: my.personalgenomes.org/profile/huE80E3D
- Download: source data (373 MB), dbSNP and nsSNP report (126 MB)
- Show debugging info

Genome report
Insufficiently evaluated variants
Coverage
Metadata
Show rare (f<10%) pathogenic variants
Show all
Show All entries
Search:

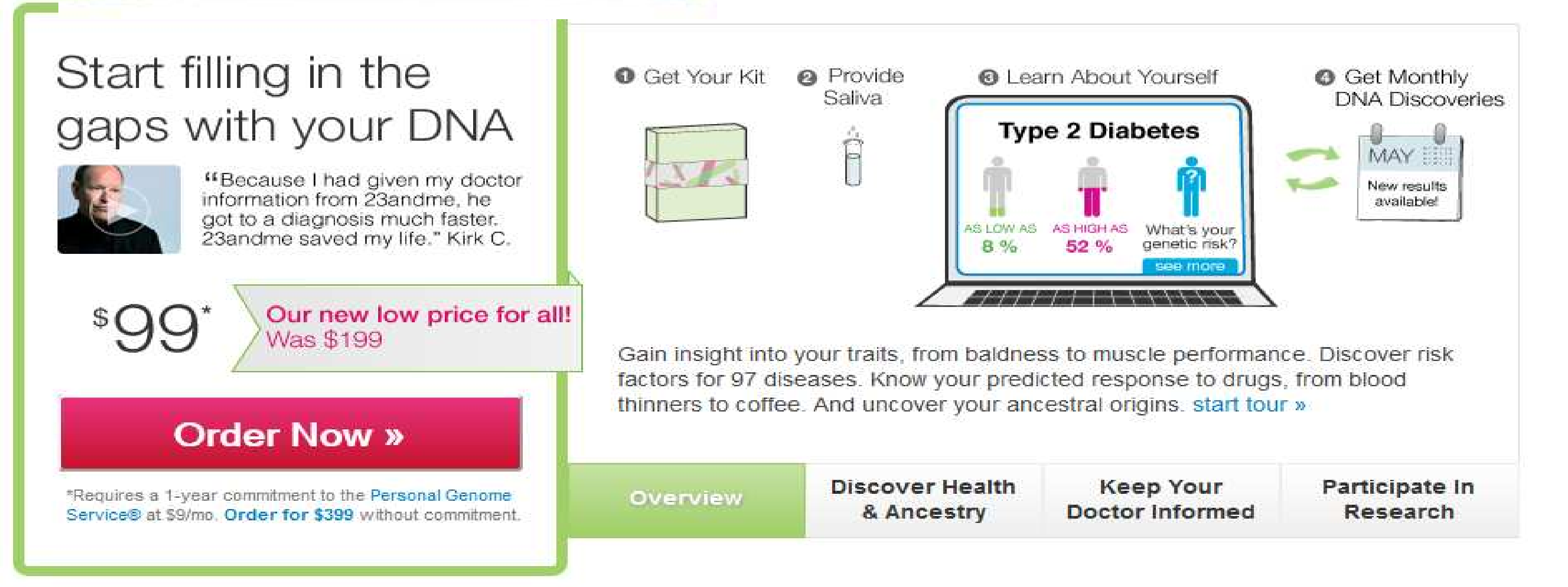
Variant	Clinical Importance	Impact	Allele freq	Summary
CPT2-S113L	High	Well-established pathogenic Recessive, Carrier (Heterozygous)	0.78%	This is the most common variant associated with late-onset carnitine palmitoyltransferase deficiency, which is classically viewed as recessive. Many patients are heterozygous for this, but are presumably compound heterozygous.
TREM2-R47H	High	Uncertain pathogenic Recessive, Carrier (Heterozygous)	0.78%	Unreported, predicted to be damaging. Other recessive mutations in this gene cause polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (a severe genetic disorder, usually lethal by age 50).
CC2D2A-G776R	High	Uncertain pathogenic Recessive, Carrier (Heterozygous)	0.78%	Unreported, predicted to be damaging. Other recessive mutations in this gene cause Joubert Syndrome and Meckel Syndrome.



Direct-to-consumer personal genomics

23andMe

Buy a DTC kit…
Share your data?



Start filling in the gaps with your DNA

“Because I had given my doctor information from 23andme, he got to a diagnosis much faster. 23andme saved my life.” Kirk C.

$99* Our new low price for all! Was $199

Order Now »

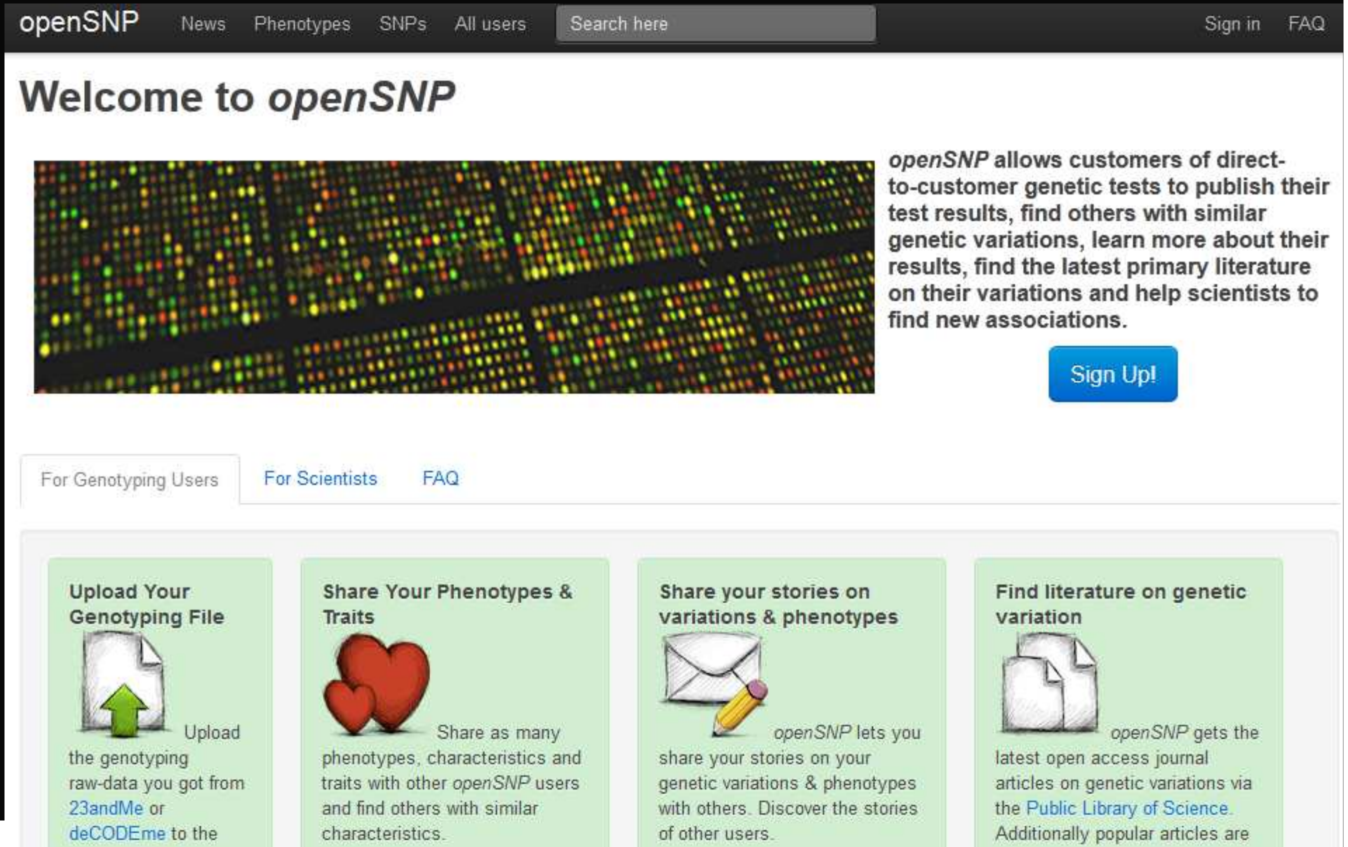
*Requires a 1-year commitment to the Personal Genome Service® at $9/mo. Order for $399 without commitment.

1 Get Your Kit
2 Provide Saliva
3 Learn About Yourself
4 Get Monthly DNA Discoveries

Gain insight into your traits, from baldness to muscle performance. Discover risk factors for 97 diseases. Know your predicted response to drugs, from blood thinners to coffee. And uncover your ancestral origins. start tour »

openSNP: share your phenotype too?

- Launched October 2011
- By 3 Masters students in Frankfurt



Welcome to *openSNP*

openSNP allows customers of direct-to-customer genetic tests to publish their test results, find others with similar genetic variations, learn more about their results, find the latest primary literature on their variations and help scientists to find new associations.

Upload Your Genotyping File

Upload the genotyping raw-data you got from 23andMe or deCODEme to the

Share Your Phenotypes & Traits

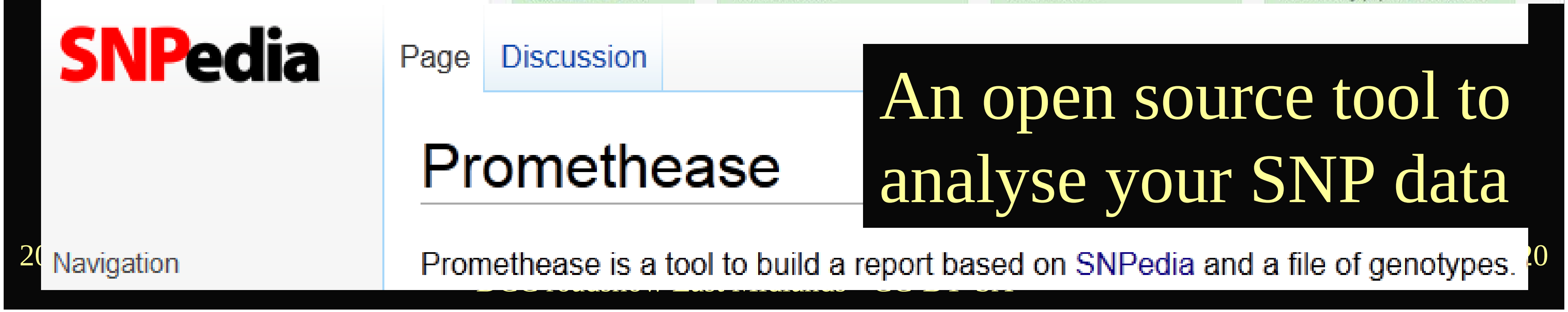
Share as many phenotypes, characteristics and traits with other *openSNP* users and find others with similar characteristics.

Share your stories on variations & phenotypes

openSNP lets you share your stories on your genetic variations & phenotypes with others. Discover the stories of other users.

Find literature on genetic variation

openSNP gets the latest open access journal articles on genetic variations via the Public Library of Science. Additionally popular articles are

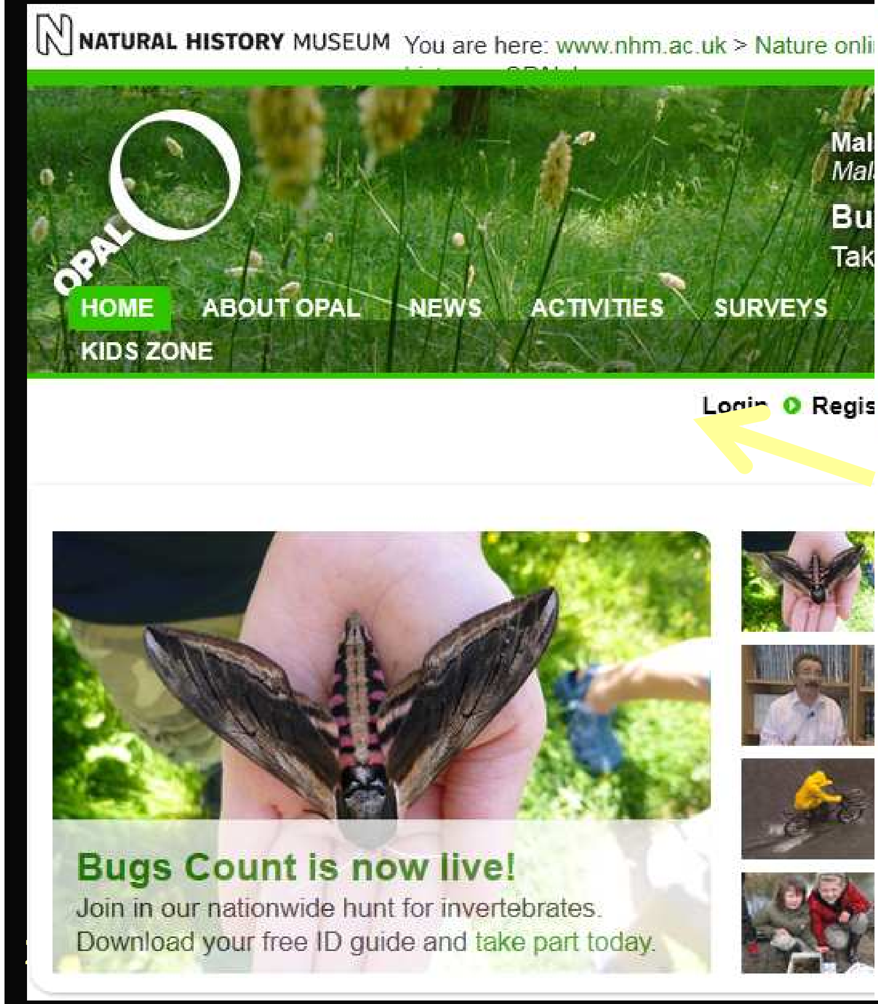


Promethease

Promethease is a tool to build a report based on SNPedia and a file of genotypes.

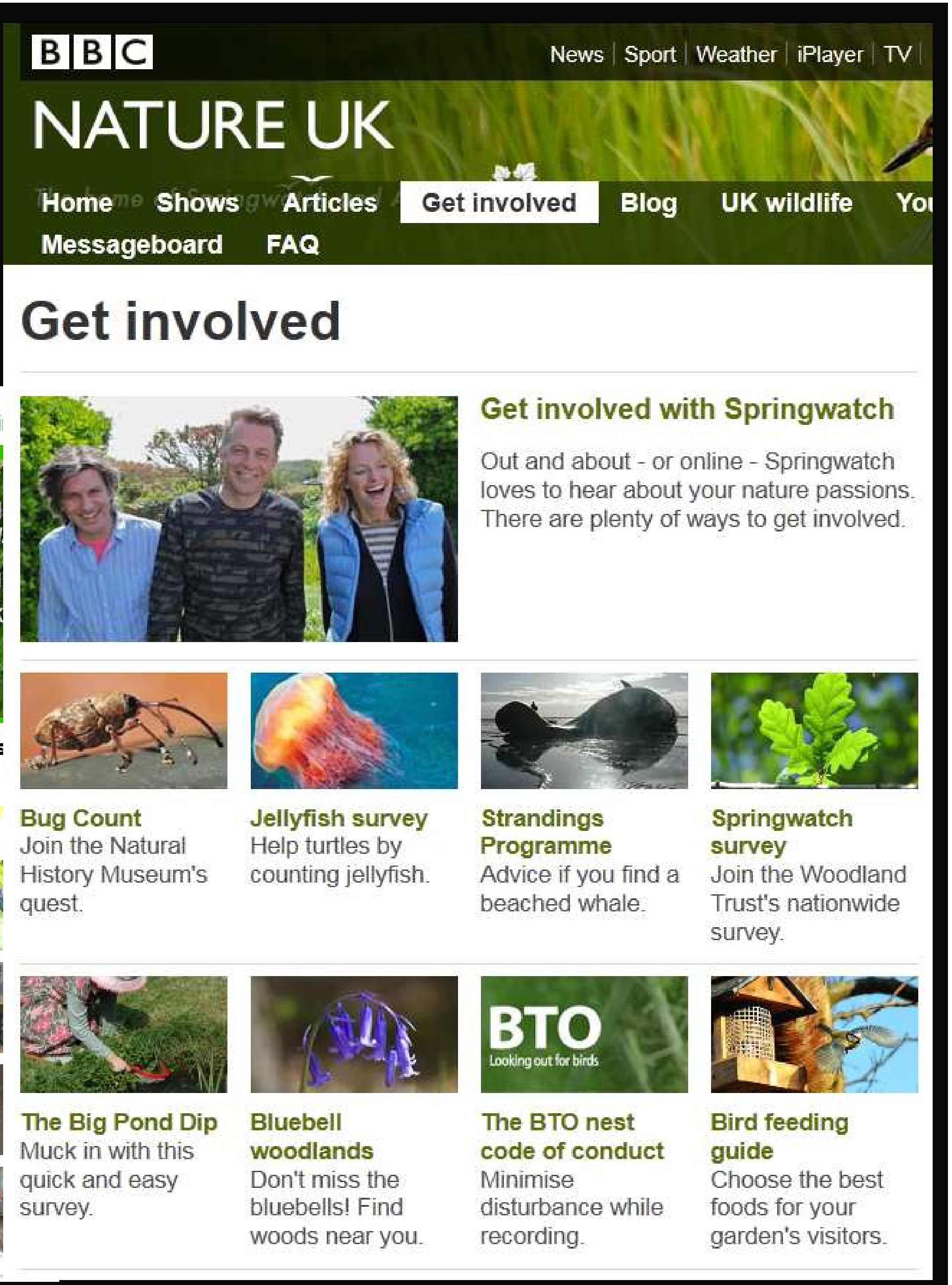
An open source tool to analyse your SNP data

2011: Citizens getting involved in science



Bugs Count is now live!

Join in our nationwide hunt for invertebrates. Download your free ID guide and take part today.



Get involved

Get involved with Springwatch

Out and about - or online - Springwatch loves to hear about your nature passions. There are plenty of ways to get involved.

Bug Count
Join the Natural History Museum's quest.

Jellyfish survey
Help turtles by counting jellyfish.

Strandings Programme
Advice if you find a beached whale.

Springwatch survey
Join the Woodland Trust's nationwide survey.

The Big Pond Dip
Muck in with this quick and easy survey.

Bluebell woodlands
Don't miss the bluebells! Find woods near you.

The BTO nest code of conduct
Minimise disturbance while recording.

Bird feeding guide
Choose the best foods for your garden's visitors.

project noah

Document nature with your mobile phone.

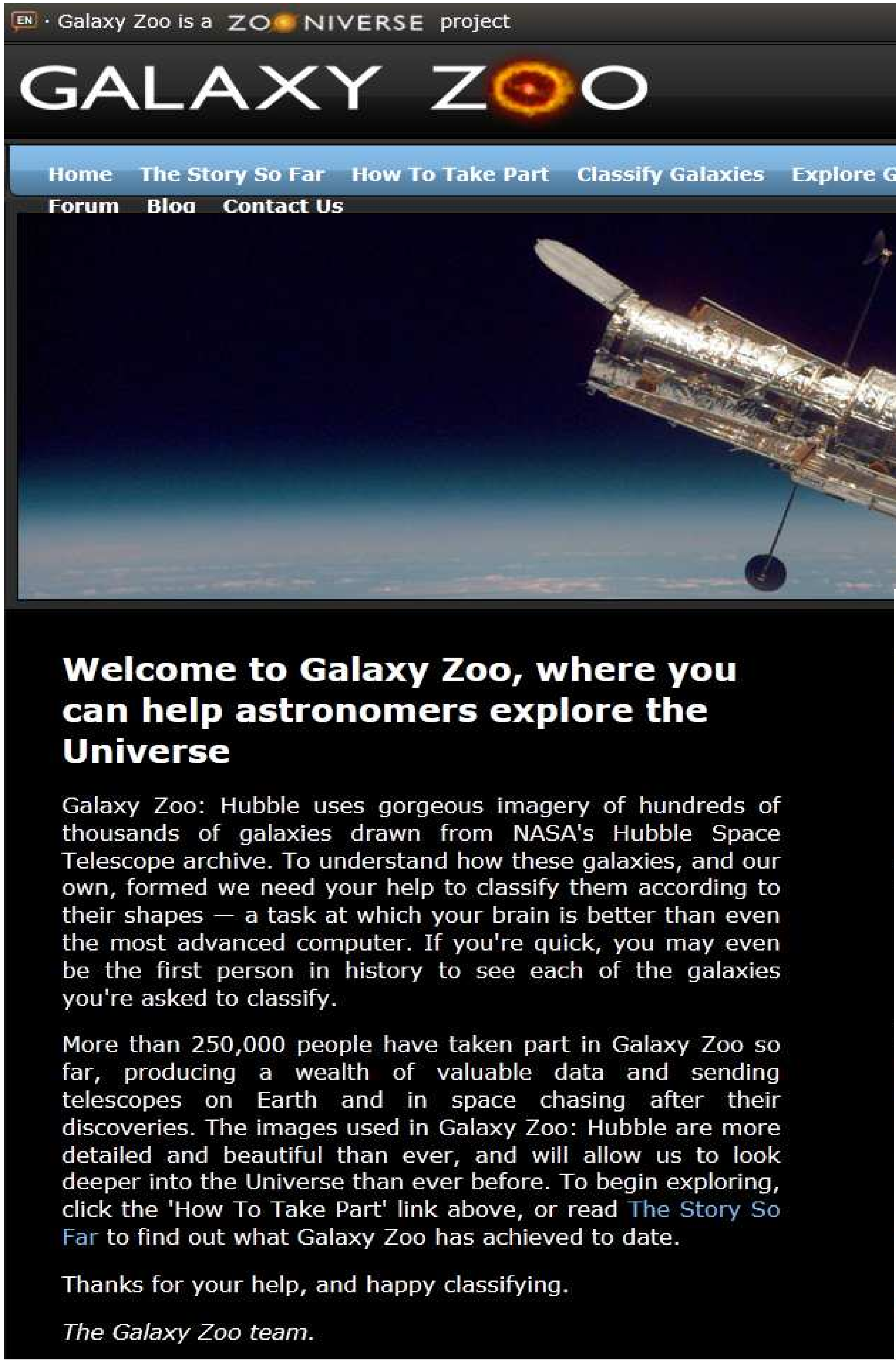
Yasser Ansari

Become a top spotter!
Grab a photograph of an interesting organism and share it with the community.

Citizen as scientist

scienceforcitizens.net

Classify galaxies…



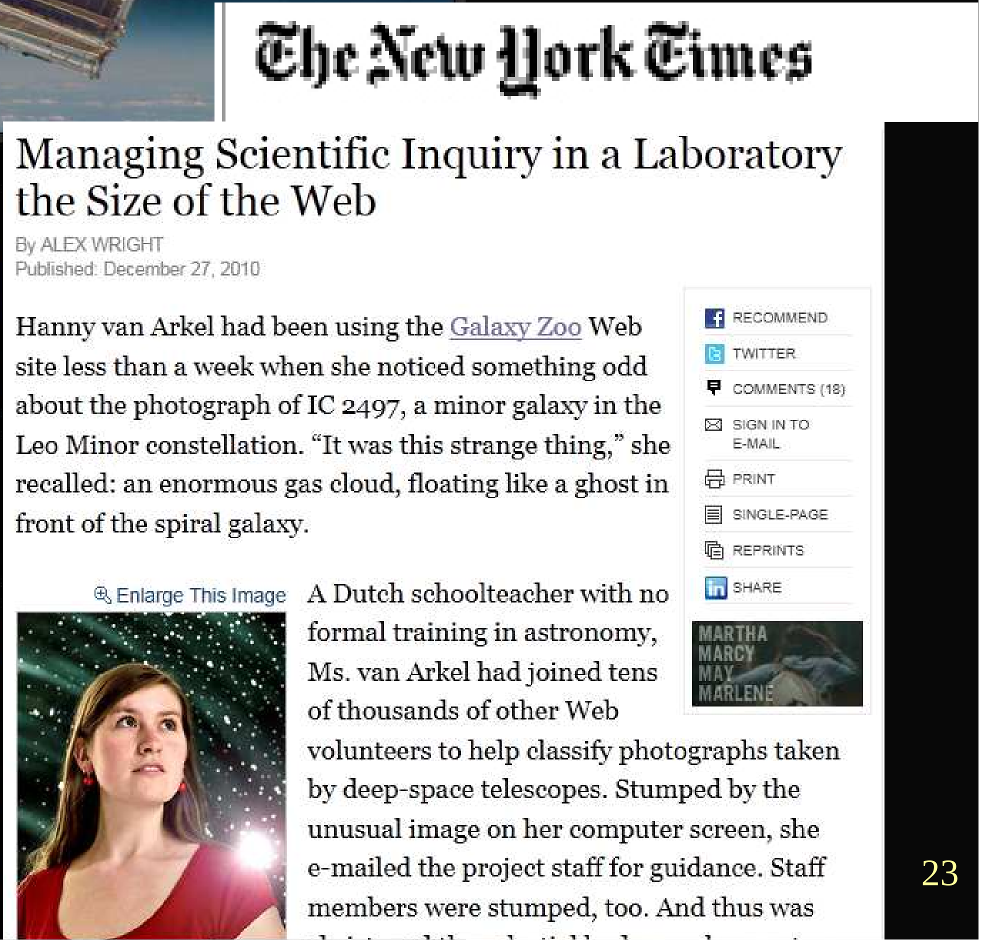
Welcome to Galaxy Zoo, where you can help astronomers explore the Universe

Galaxy Zoo: Hubble uses gorgeous imagery of hundreds of thousands of galaxies drawn from NASA's Hubble Space Telescope archive. To understand how these galaxies, and our own, formed we need your help to classify them according to their shapes — a task at which your brain is better than even the most advanced computer. If you're quick, you may even be the first person in history to see each of the galaxies you're asked to classify.

More than 250,000 people have taken part in Galaxy Zoo so far, producing a wealth of valuable data and sending telescopes on Earth and in space chasing after their discoveries. The images used in Galaxy Zoo: Hubble are more detailed and beautiful than ever, and will allow us to look deeper into the Universe than ever before. To begin exploring, click the 'How To Take Part' link above, or read The Story So Far to find out what Galaxy Zoo has achieved to date.

Thanks for your help, and happy classifying.

The Galaxy Zoo team.

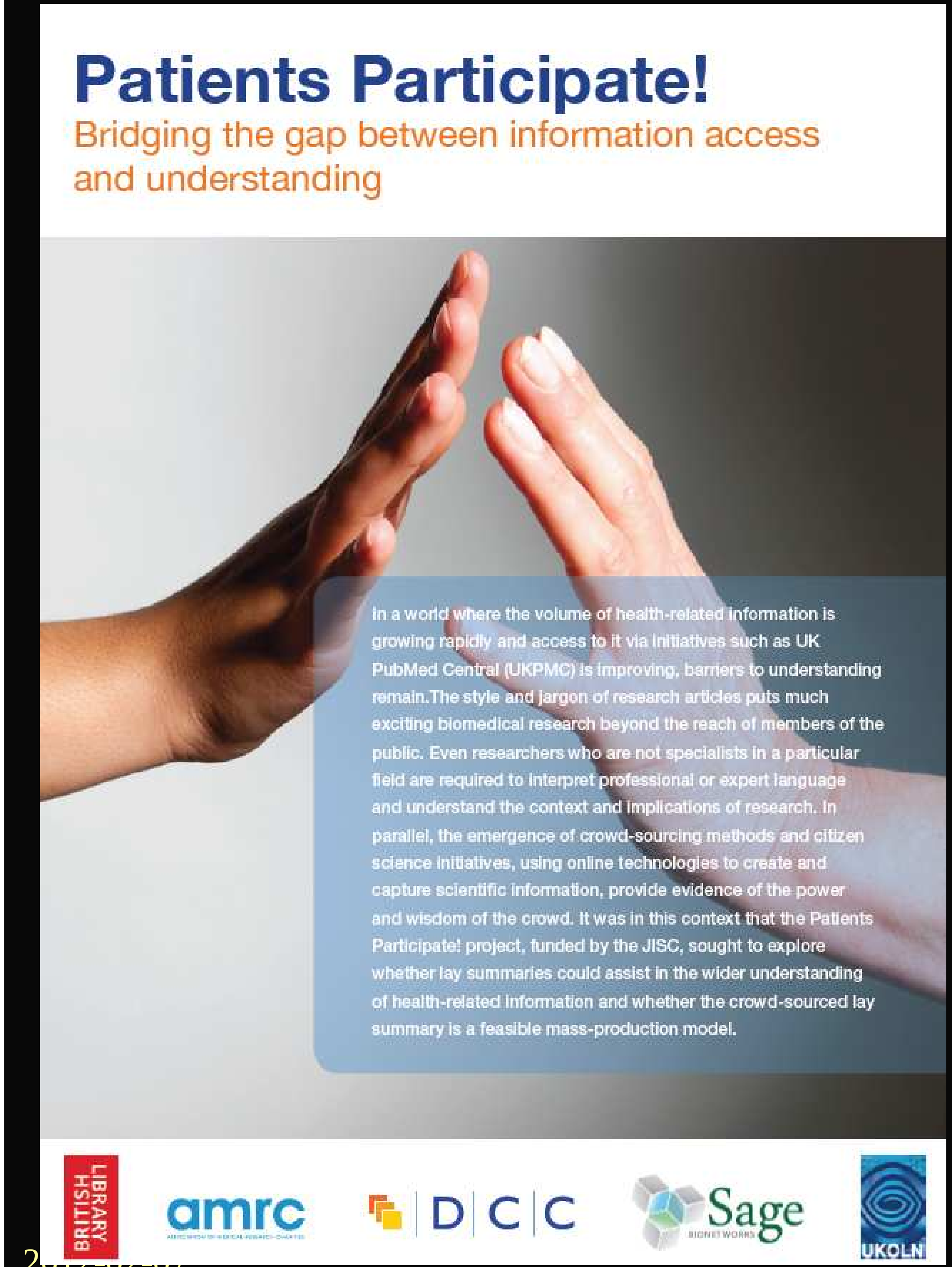


Managing Scientific Inquiry in a Laboratory the Size of the Web

By ALEX WRIGHT
Published: December 27, 2010

Hanny van Arkel had been using the Galaxy Zoo Web site less than a week when she noticed something odd about the photograph of IC 2497, a minor galaxy in the Leo Minor constellation. "It was this strange thing," she recalled: an enormous gas cloud, floating like a ghost in front of the spiral galaxy.

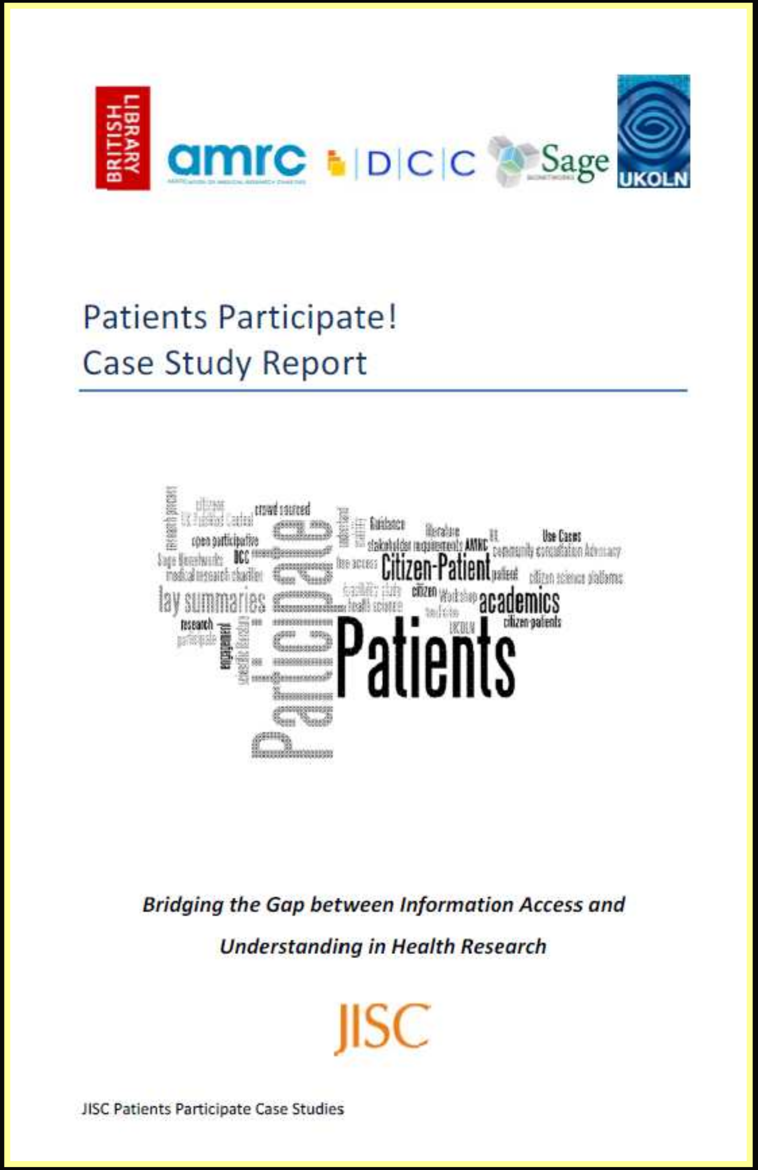
A Dutch schoolteacher with no formal training in astronomy, Ms. van Arkel had joined tens of thousands of other Web volunteers to help classify photographs taken by deep-space telescopes. Stumped by the unusual image on her computer screen, she e-mailed the project staff for guidance. Staff members were stumped, too. And thus was



Patients Participate!

Bridging the gap between information access and understanding

In a world where the volume of health-related information is growing rapidly and access to it via initiatives such as UK PubMed Central (UKPMC) is improving, barriers to understanding remain.The style and jargon of research articles puts much exciting biomedical research beyond the reach of members of the public. Even researchers who are not specialists in a particular field are required to interpret professional or expert language and understand the context and implications of research. In parallel, the emergence of crowd-sourcing methods and citizen science initiatives, using online technologies to create and capture scientific information, provide evidence of the power and wisdom of the crowd. It was in this context that the Patients Participate! project, funded by the JISC, sought to explore whether lay summaries could assist in the wider understanding of health-related information and whether the crowd-sourced lay summary is a feasible mass-production model.



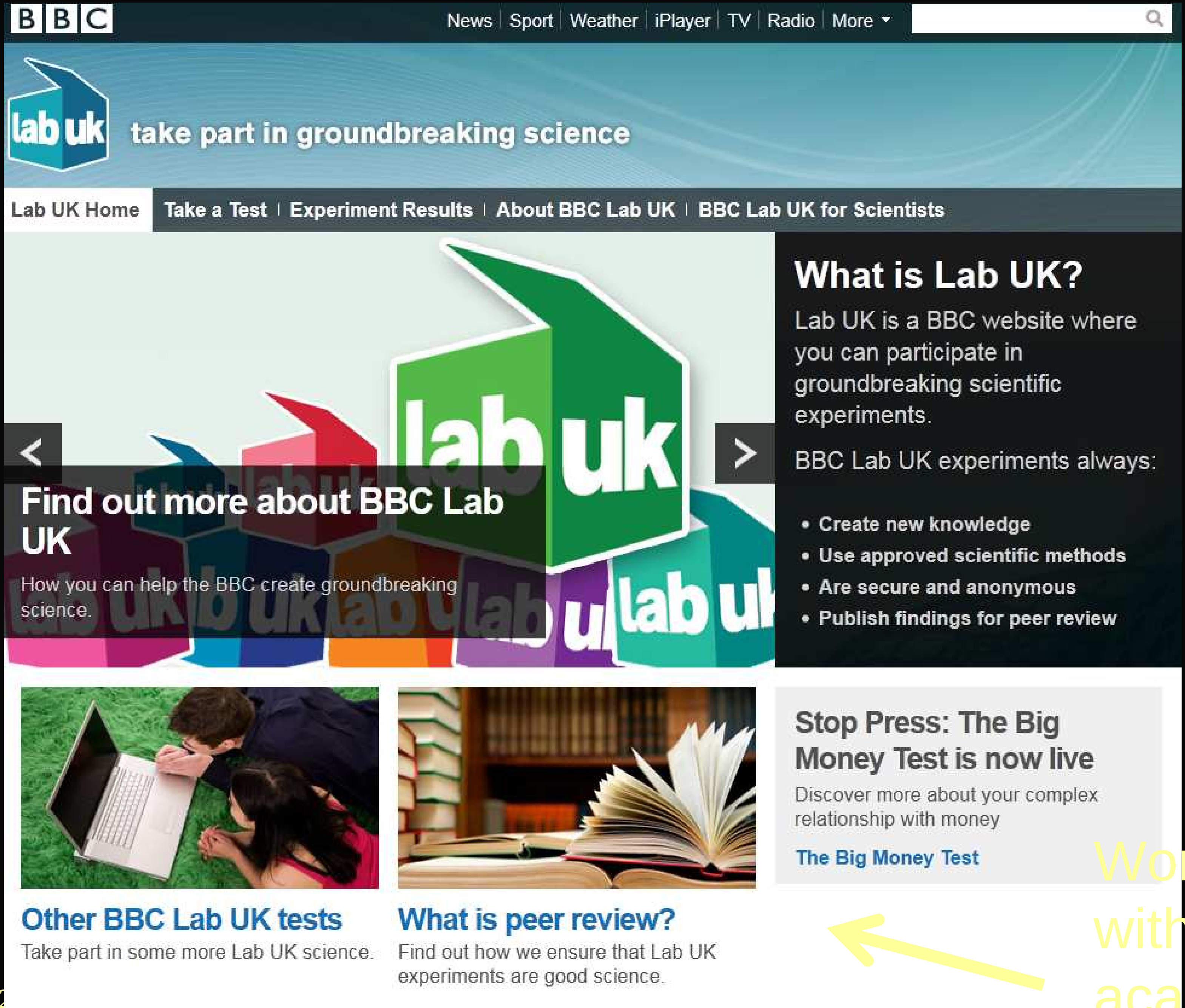
Patients Participate!
Case Study Report

Bridging the Gap between Information Access and Understanding in Health Research

JISC

JISC Patients Participate Case Studies

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


BBC
News | Sport | Weather | iPlayer | TV | Radio | More
lab uk
take part in groundbreaking science
Lab UK Home | Take a Test | Experiment Results | About BBC Lab UK | BBC Lab UK for Scientists
lab uk

Find out more about BBC Lab UK

How you can help the BBC create groundbreaking science.

What is Lab UK?

Lab UK is a BBC website where you can participate in groundbreaking scientific experiments.

BBC Lab UK experiments always:

- Create new knowledge
- Use approved scientific methods
- Are secure and anonymous
- Publish findings for peer review

Stop Press: The Big Money Test is now live

Discover more about your complex relationship with money

The Big Money Test

Other BBC Lab UK tests

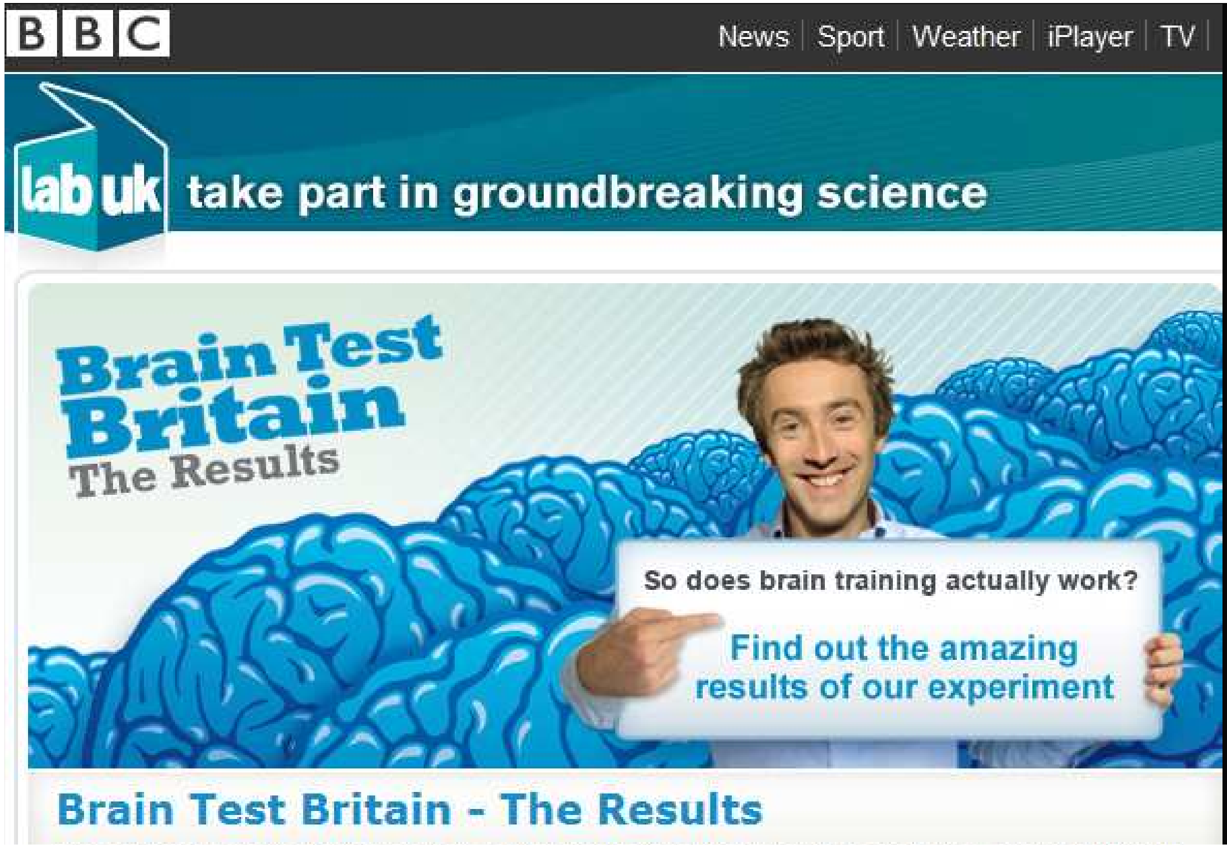
Take part in some more Lab UK science.

What is peer review?

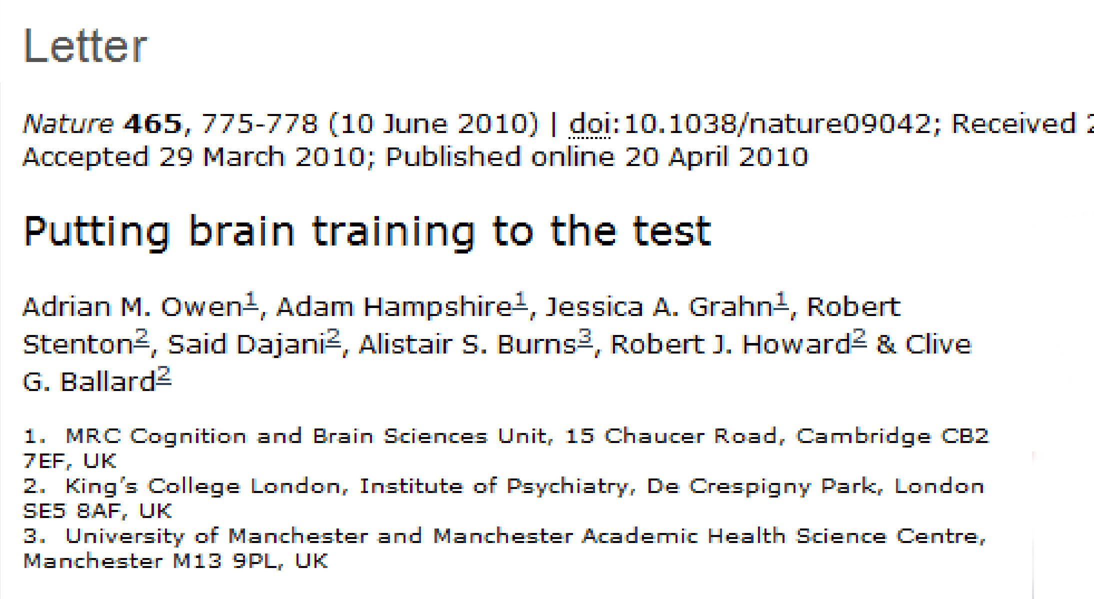
Find out how we ensure that Lab UK experiments are good science.

Working with academics

Validate results data and publish


BBC
News | Sport | Weather | iPlayer | TV
lab uk take part in groundbreaking science
Brain Test Britain The Results
So does brain training actually work?
Find out the amazing results of our experiment

Brain Test Britain - The Results



Letter

Nature **465**, 775-778 (10 June 2010) | doi:10.1038/nature09042; Received 25 January 2010; Accepted 29 March 2010; Published online 20 April 2010

Putting brain training to the test

Adrian M. Owen[1], Adam Hampshire[1], Jessica A. Grahn[1], Robert Stenton[2], Said Dajani[2], Alistair S. Burns[3], Robert J. Howard[2] & Clive G. Ballard[2]

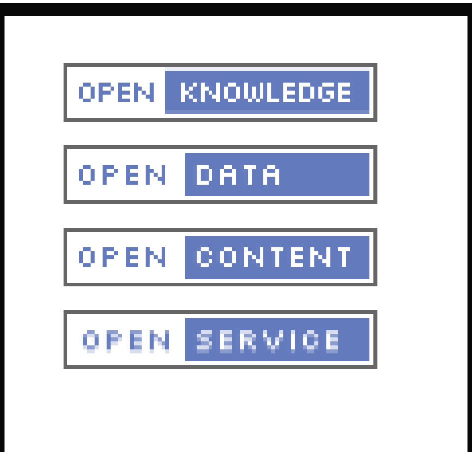
1. MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK
2. King's College London, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK
3. University of Manchester and Manchester Academic Health Science Centre, Manchester M13 9PL, UK

nature

Institution – enabler or barrier?

Panton Principles

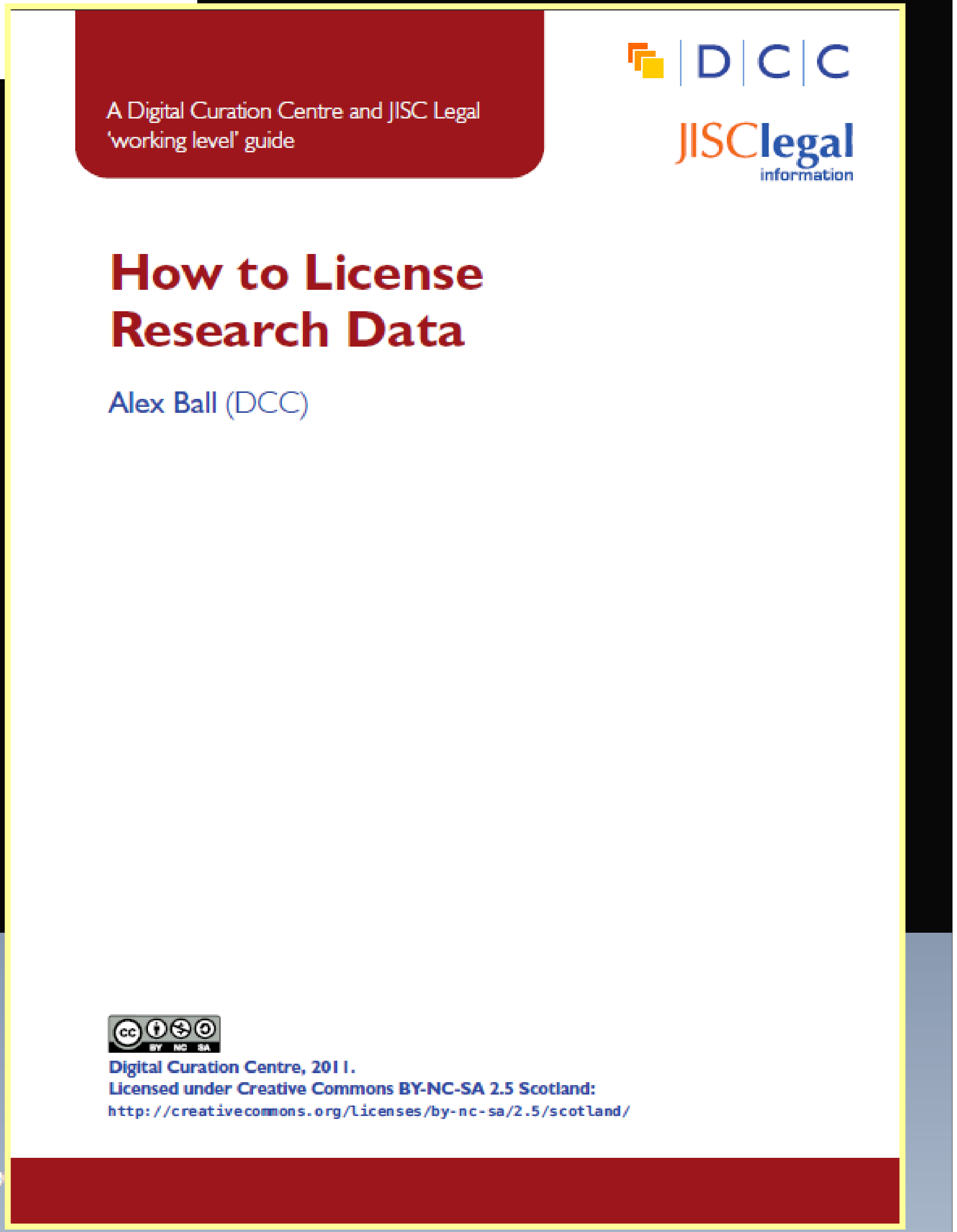
Principles for Open Data in Science



"For science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made **open**"

Open Definition

Defining the Open in Open Data,


A Digital Curation Centre and JISC Legal 'working level' guide

DCC

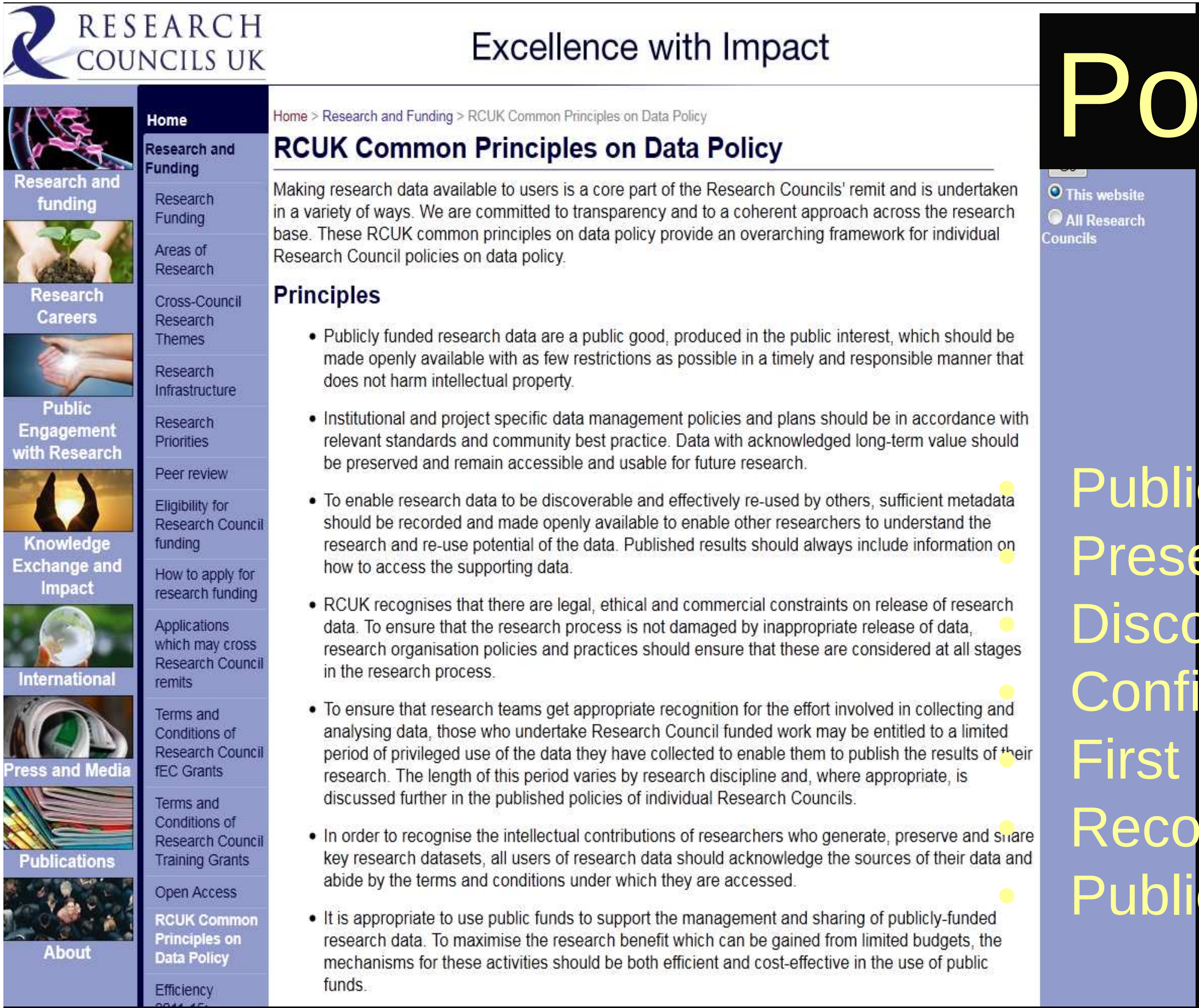
JISClegal information

How to License Research Data

Alex Ball (DCC)

Digital Curation Centre, 2011.
Licensed under Creative Commons BY-NC-SA 2.5 Scotland:
http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/

Policy



RESEARCH COUNCILS UK

Excellence with Impact

Research and funding
Research Careers
Public Engagement with Research
Knowledge Exchange and Impact
International
Press and Media
Publications
About

Home
Research and Funding
Research Funding
Areas of Research
Cross-Council Research Themes
Research Infrastructure
Research Priorities
Peer review
Eligibility for Research Council funding
How to apply for research funding
Applications which may cross Research Council remits
Terms and Conditions of Research Council fEC Grants
Terms and Conditions of Research Council Training Grants
Open Access
RCUK Common Principles on Data Policy
Efficiency

Home > Research and Funding > RCUK Common Principles on Data Policy

RCUK Common Principles on Data Policy

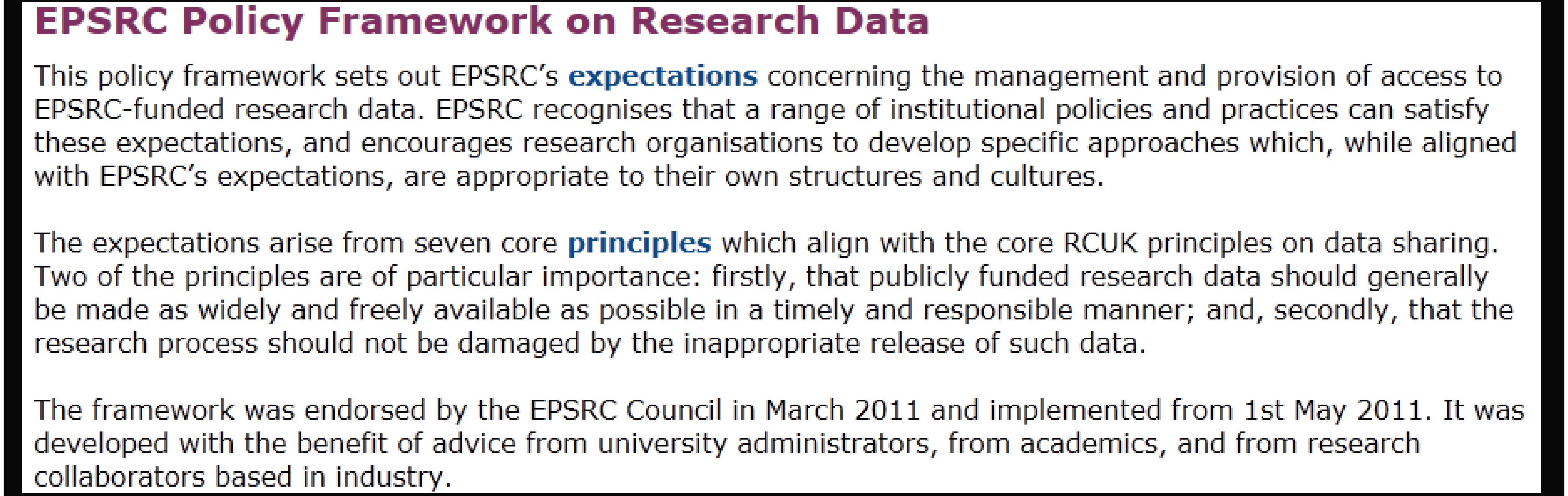
Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

- Public good
- Preservation
- Discovery
- Confidentiality
- First use
- Recognition
- Public funding

EPSRC
Pioneering research and skills
Engineering and Physical Sciences Research Council



EPSRC Policy Framework on Research Data

This policy framework sets out EPSRC's **expectations** concerning the management and provision of access to EPSRC-funded research data. EPSRC recognises that a range of institutional policies and practices can satisfy these expectations, and encourages research organisations to develop specific approaches which, while aligned with EPSRC's expectations, are appropriate to their own structures and cultures.

The expectations arise from seven core **principles** which align with the core RCUK principles on data sharing. Two of the principles are of particular importance: firstly, that publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner; and, secondly, that the research process should not be damaged by the inappropriate release of such data.

The framework was endorsed by the EPSRC Council in March 2011 and implemented from 1st May 2011. It was developed with the benefit of advice from university administrators, from academics, and from research collaborators based in industry.

EPSRC Expectations : implications for HEIs

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

EPSRC

Pioneering research and skills

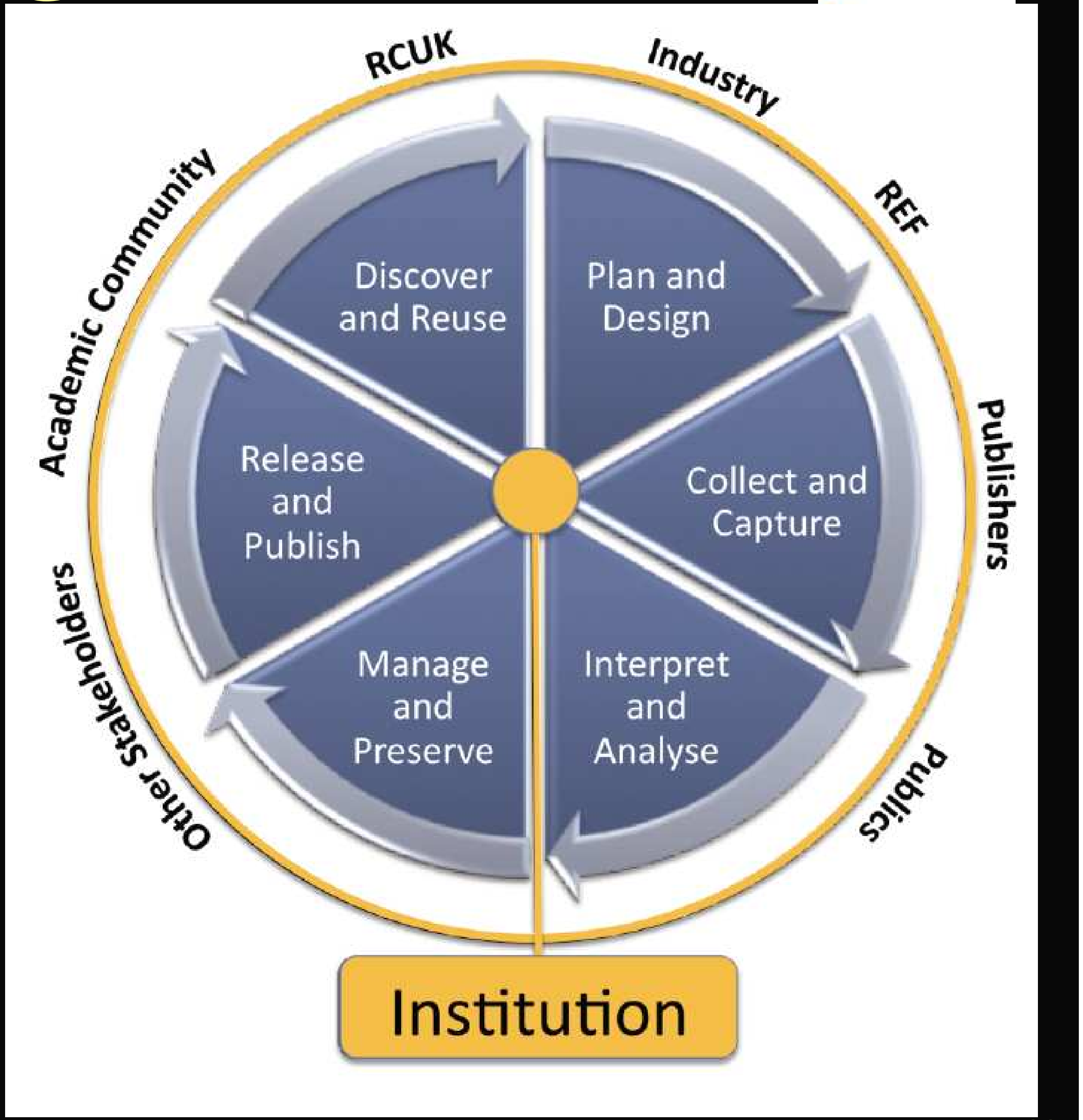
Engineering and Physical Sciences Research Council

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**.
- Compliance will be monitored and non-compliance investigated.
- Failure to share research data could result in the imposition of sanctions.

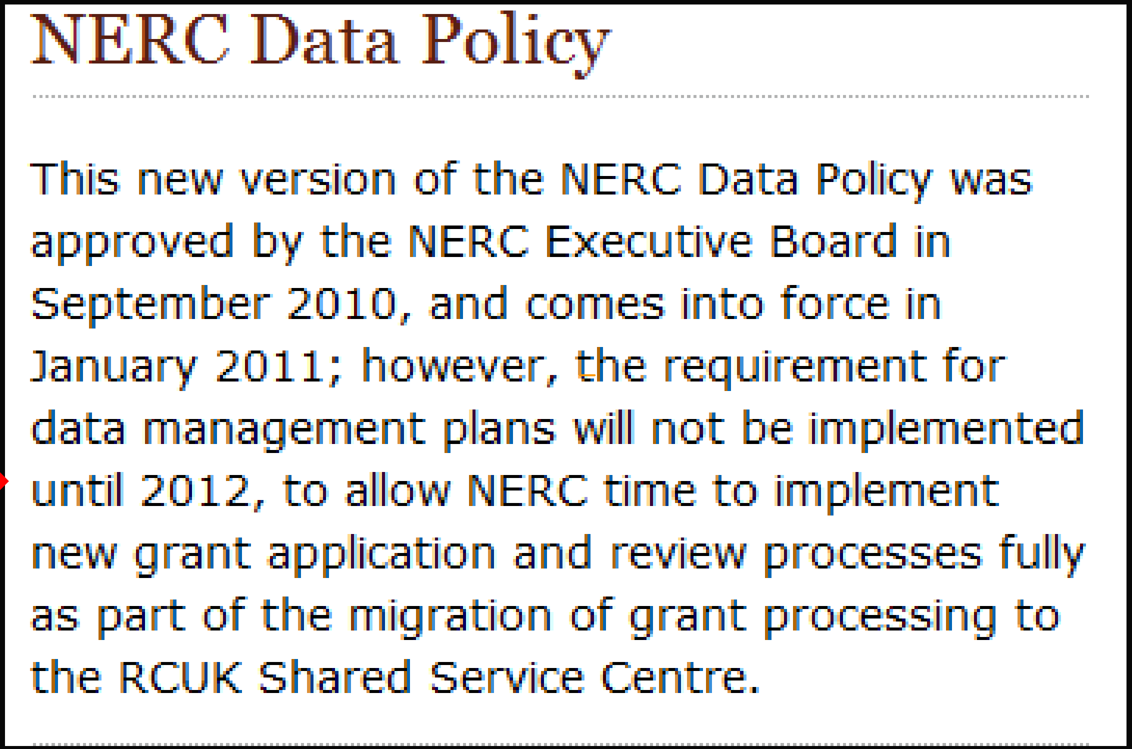
Research360@Bath

- Faculty-Industry focus
- New institutional data scientist role
- Addresses EPSRC expectations
- Doctoral Training Centre hubs
- Faculty cascade model
- Multi-team approach



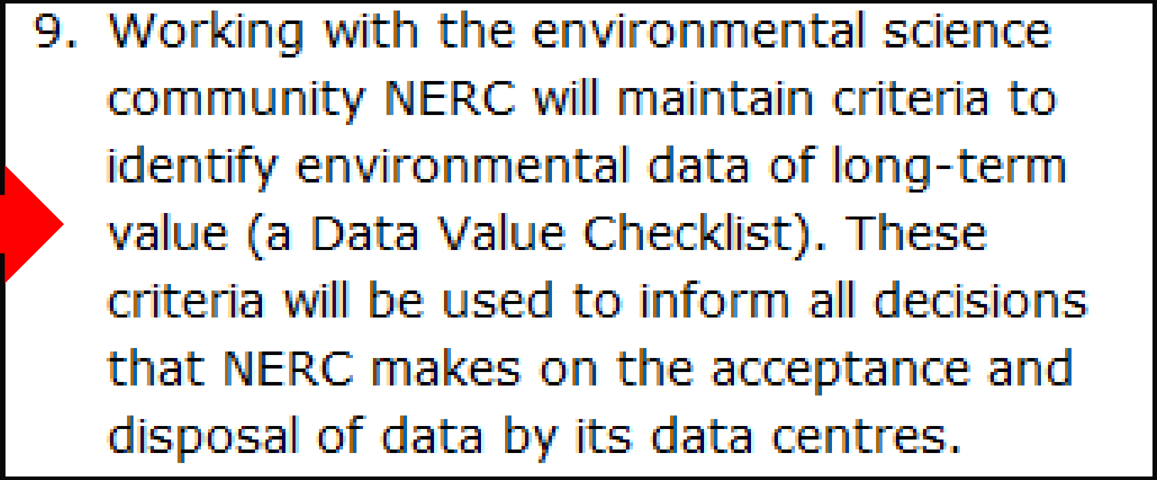
http://blogs.bath.ac.uk/research360/

Funder Policy



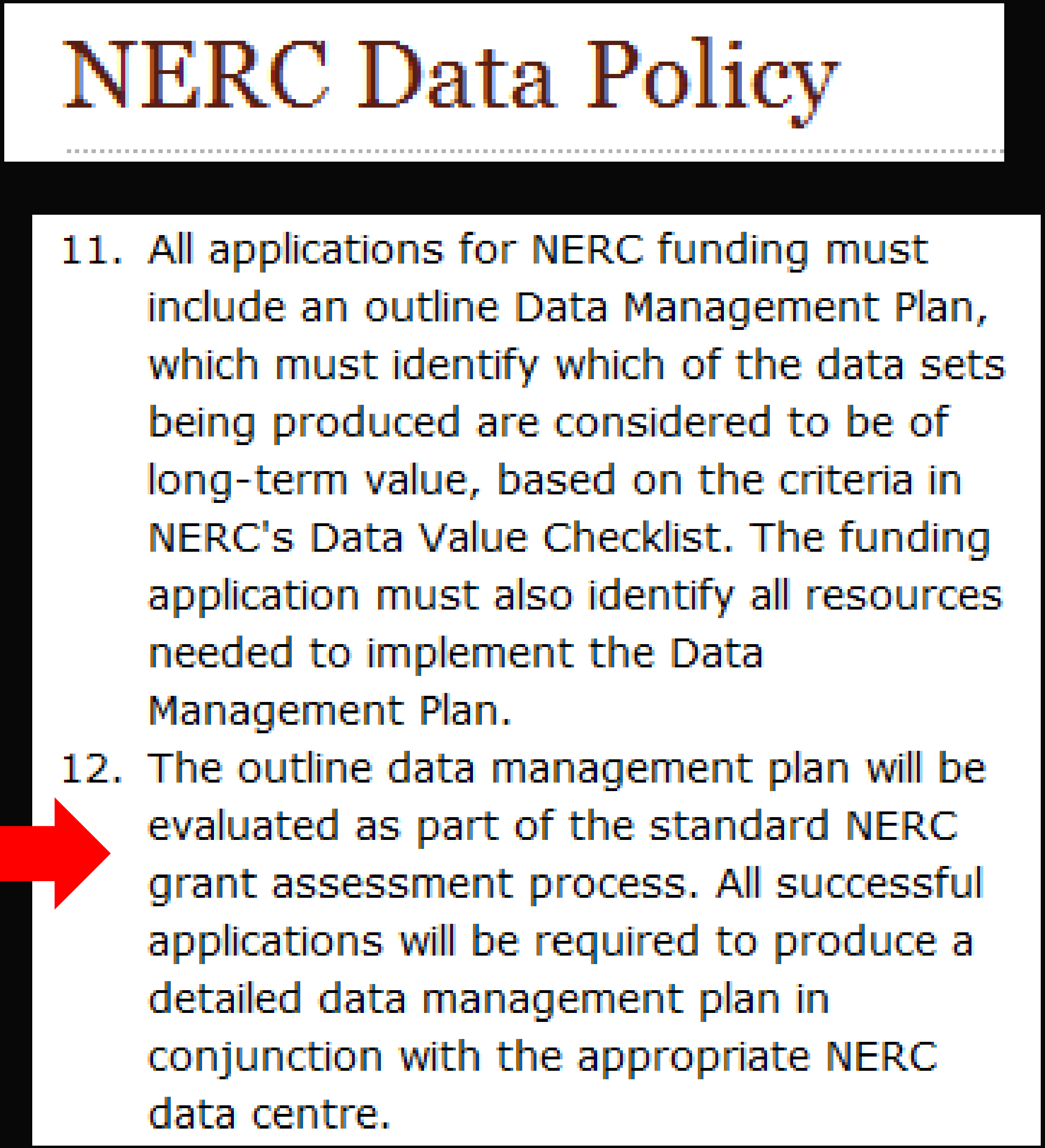
NERC Data Policy

This new version of the NERC Data Policy was approved by the NERC Executive Board in September 2010, and comes into force in January 2011; however, the requirement for data management plans will not be implemented until 2012, to allow NERC time to implement new grant application and review processes fully as part of the migration of grant processing to the RCUK Shared Service Centre.



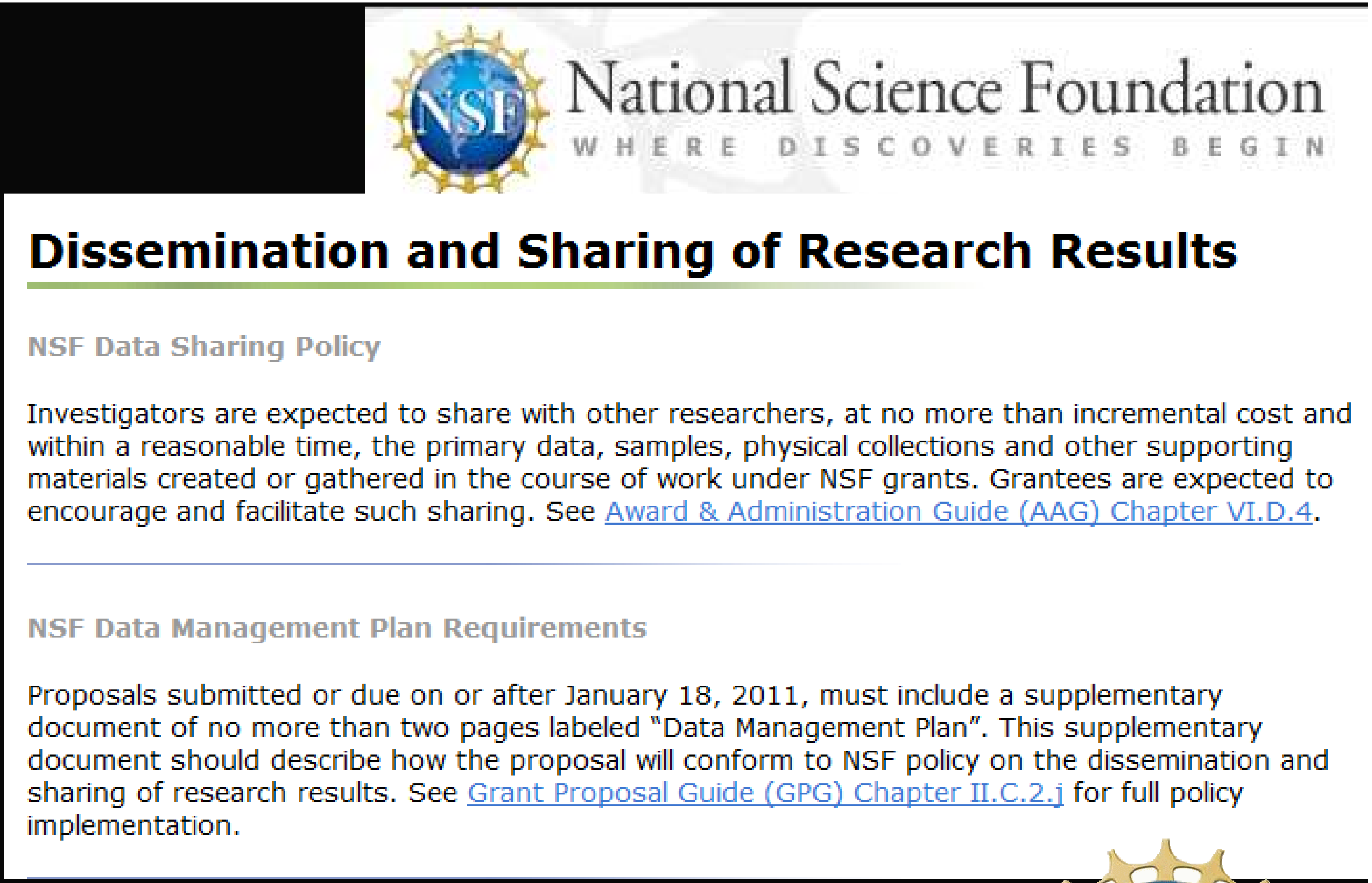
9. Working with the environmental science community NERC will maintain criteria to identify environmental data of long-term value (a Data Value Checklist). These criteria will be used to inform all decisions that NERC makes on the acceptance and disposal of data by its data centres.

Funder Policy



NERC Data Policy

11. 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.
12. The outline data management plan will be evaluated as part of the standard NERC grant assessment process. All successful applications will be required to produce a detailed data management plan in conjunction with the appropriate NERC data centre.



National Science Foundation
WHERE DISCOVERIES BEGIN

Dissemination and Sharing of Research Results

NSF Data Sharing Policy

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.

NSF Data Management Plan Requirements

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.

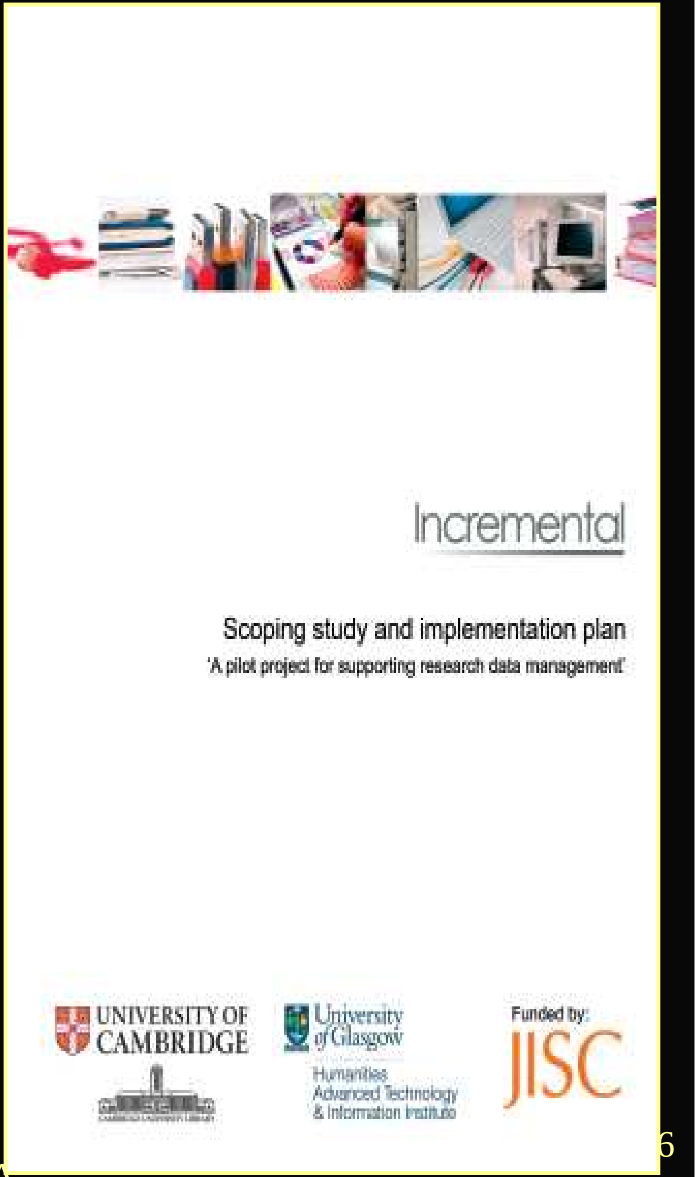
NSF-OCI TASK FORCE on
Data and Visualization : Report

http://www.nsf.gov/od/oci/taskforces/

Institutional perspective

- Creating & organising data
- Storage and access
- Back-up
- Preservation
- Sharing and re-use

The majority of people felt that some form of policy or guidance was needed....



Incremental

Scoping study and implementation plan

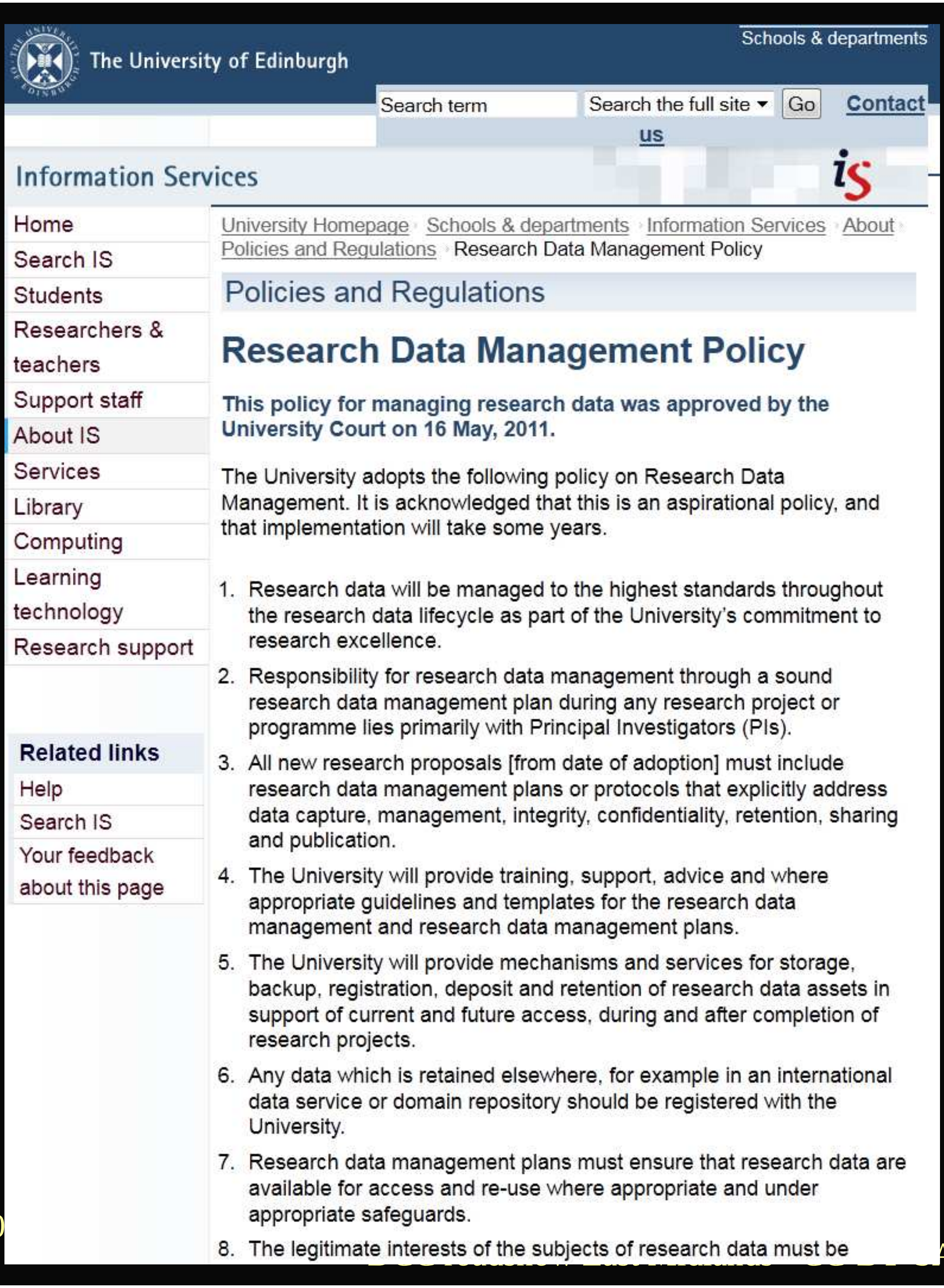
'A pilot project for supporting research data management'

UNIVERSITY OF CAMBRIDGE

University of Glasgow

Humanities Advanced Technology & Information Institute

Funded by: JISC



Policies and Regulations

Research Data Management Policy

This policy for managing research data was approved by the University Court on 16 May, 2011.

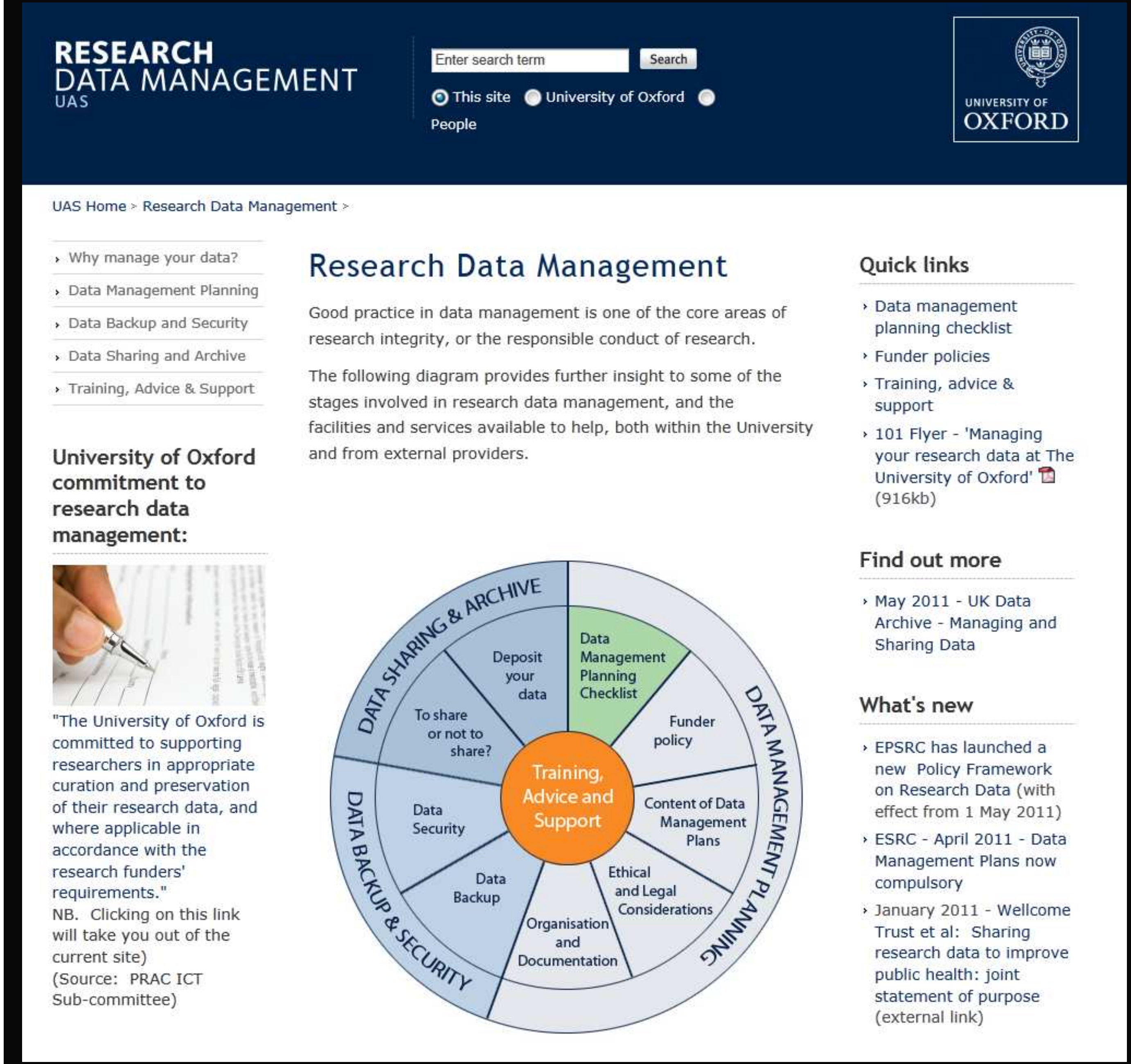
The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.
5. The University will provide 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.
6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
8. The legitimate interests of the subjects of research data must be

Institutional Policy

…article in *International Journal of Digital Curation*

Institutional Policy



UAS Home > Research Data Management >

- Why manage your data?
- Data Management Planning
- Data Backup and Security
- Data Sharing and Archive
- Training, Advice & Support

University of Oxford commitment to research data management:

"The University of Oxford is committed to supporting researchers in appropriate curation and preservation of their research data, and where applicable in accordance with the research funders' requirements."

NB. Clicking on this link will take you out of the current site)

(Source: PRAC ICT Sub-committee)

Research Data Management

Good practice in data management is one of the core areas of research integrity, or the responsible conduct of research.

The following diagram provides further insight to some of the stages involved in research data management, and the facilities and services available to help, both within the University and from external providers.

DATA SHARING & ARCHIVE
DATA MANAGEMENT PLANNING
DATA BACKUP & SECURITY
Deposit your data
Data Management Planning Checklist
Funder policy
To share or not to share?
Training, Advice and Support
Data Security
Content of Data Management Plans
Data Backup
Ethical and Legal Considerations
Organisation and Documentation

Quick links

- Data management planning checklist
- Funder policies
- Training, advice & support
- 101 Flyer - 'Managing your research data at The University of Oxford' (916kb)

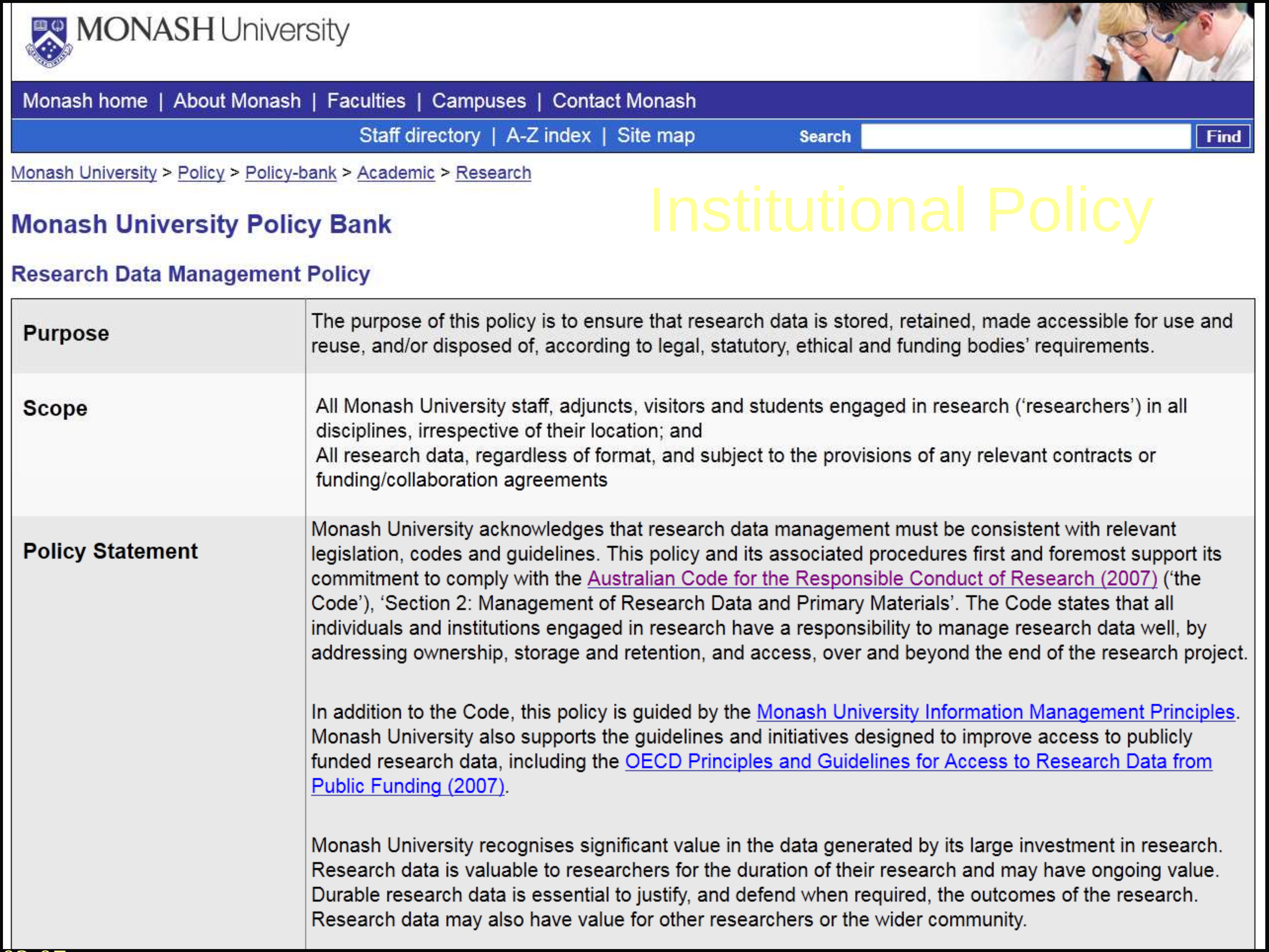
Find out more

- May 2011 - UK Data Archive - Managing and Sharing Data

What's new

- EPSRC has launched a new Policy Framework on Research Data (with effect from 1 May 2011)
- ESRC - April 2011 - Data Management Plans now compulsory
- January 2011 - Wellcome Trust et al: Sharing research data to improve public health: joint statement of purpose (external link)

Institutional Policy



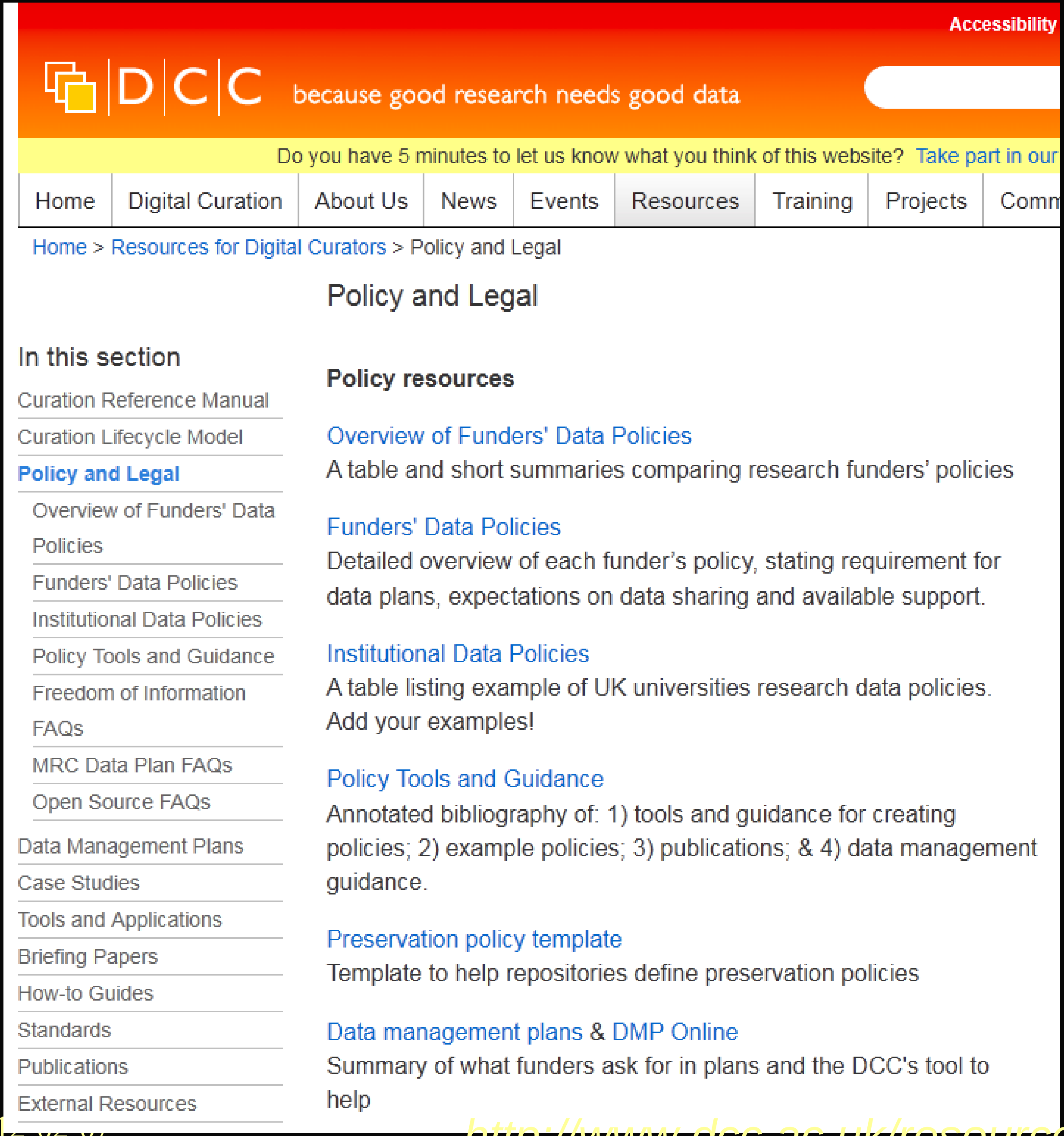
MONASH University
Monash home | About Monash | Faculties | Campuses | Contact Monash
Staff directory | A-Z index | Site map
Search
Find

Monash University > Policy > Policy-bank > Academic > Research

Monash University Policy Bank

Research Data Management Policy

Purpose	The purpose of this policy is to ensure that research data is stored, retained, made accessible for use and reuse, and/or disposed of, according to legal, statutory, ethical and funding bodies' requirements.
Scope	All Monash University staff, adjuncts, visitors and students engaged in research ('researchers') in all disciplines, irrespective of their location; and All research data, regardless of format, and subject to the provisions of any relevant contracts or funding/collaboration agreements
Policy Statement	Monash University acknowledges that research data management must be consistent with relevant legislation, codes and guidelines. This policy and its associated procedures first and foremost support its commitment to comply with the Australian Code for the Responsible Conduct of Research (2007) ('the Code'), 'Section 2: Management of Research Data and Primary Materials'. The Code states that all individuals and institutions engaged in research have a responsibility to manage research data well, by addressing ownership, storage and retention, and access, over and beyond the end of the research project. In addition to the Code, this policy is guided by the Monash University Information Management Principles. Monash University also supports the guidelines and initiatives designed to improve access to publicly funded research data, including the OECD Principles and Guidelines for Access to Research Data from Public Funding (2007). Monash University recognises significant value in the data generated by its large investment in research. Research data is valuable to researchers for the duration of their research and may have ongoing value. Durable research data is essential to justify, and defend when required, the outcomes of the research. Research data may also have value for other researchers or the wider community.



Accessibility
DCC
because good research needs good data

Do you have 5 minutes to let us know what you think of this website? Take part in our

Home | Digital Curation | About Us | News | Events | Resources | Training | Projects | Comm

Home > Resources for Digital Curators > Policy and Legal

Policy and Legal

In this section

- Curation Reference Manual
- Curation Lifecycle Model
- **Policy and Legal**
  - Overview of Funders' Data Policies
  - Funders' Data Policies
  - Institutional Data Policies
  - Policy Tools and Guidance
  - Freedom of Information FAQs
  - MRC Data Plan FAQs
  - Open Source FAQs
- Data Management Plans
- Case Studies
- Tools and Applications
- Briefing Papers
- How-to Guides
- Standards
- Publications
- External Resources

Policy resources

Overview of Funders' Data Policies
A table and short summaries comparing research funders' policies

Funders' Data Policies
Detailed overview of each funder's policy, stating requirement for data plans, expectations on data sharing and available support.

Institutional Data Policies
A table listing example of UK universities research data policies. Add your examples!

Policy Tools and Guidance
Annotated bibliography of: 1) tools and guidance for creating policies; 2) example policies; 3) publications; & 4) data management guidance.

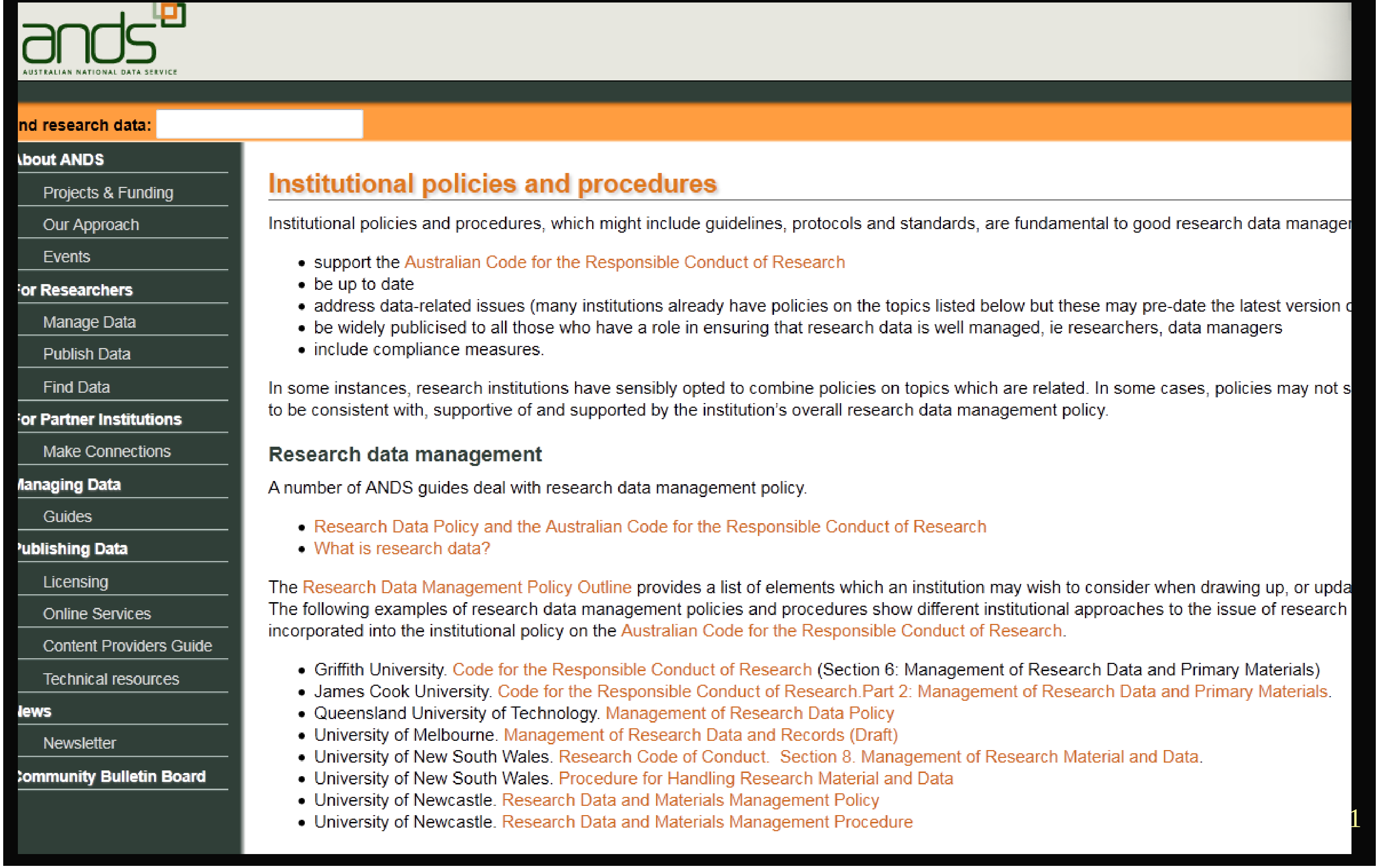
Preservation policy template
Template to help repositories define preservation policies

Data management plans & DMP Online
Summary of what funders ask for in plans and the DCC's tool to help

DCC Policy Summary

http://www.dcc.ac.uk/resources/policy-and-legal

Policy summary from ANDS



Institutional policies and procedures

Institutional policies and procedures, which might include guidelines, protocols and standards, are fundamental to good research data manage

- support the Australian Code for the Responsible Conduct of Research
- be up to date
- address data-related issues (many institutions already have policies on the topics listed below but these may pre-date the latest version c
- be widely publicised to all those who have a role in ensuring that research data is well managed, ie researchers, data managers
- include compliance measures.

In some instances, research institutions have sensibly opted to combine policies on topics which are related. In some cases, policies may not s to be consistent with, supportive of and supported by the institution's overall research data management policy.

Research data management

A number of ANDS guides deal with research data management policy.

- Research Data Policy and the Australian Code for the Responsible Conduct of Research
- What is research data?

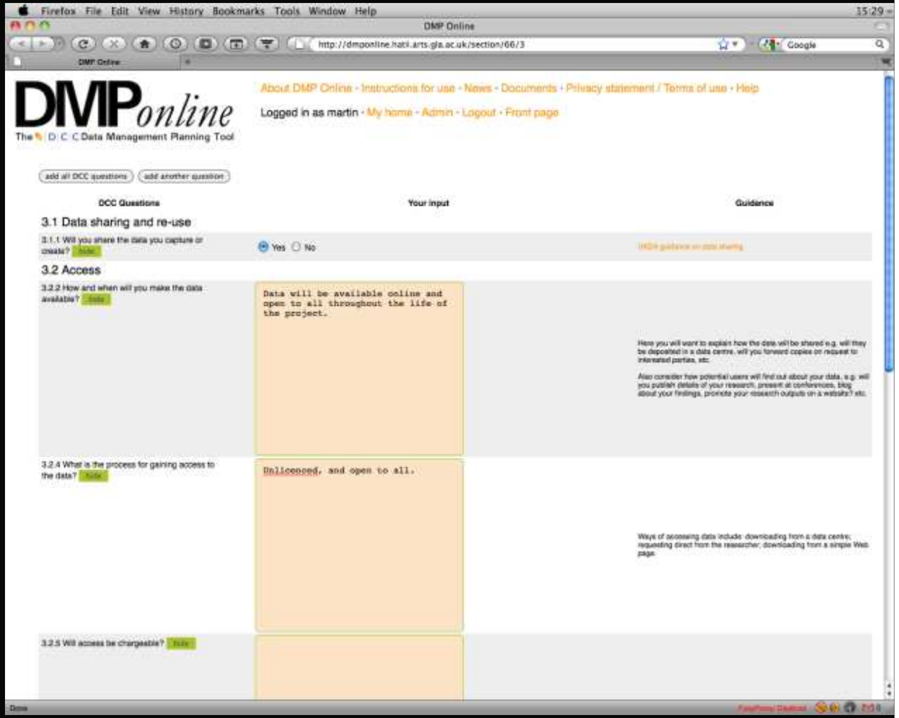
The Research Data Management Policy Outline provides a list of elements which an institution may wish to consider when drawing up, or upda The following examples of research data management policies and procedures show different institutional approaches to the issue of research incorporated into the institutional policy on the Australian Code for the Responsible Conduct of Research.

- Griffith University. Code for the Responsible Conduct of Research (Section 6: Management of Research Data and Primary Materials)
- James Cook University. Code for the Responsible Conduct of Research.Part 2: Management of Research Data and Primary Materials.
- Queensland University of Technology. Management of Research Data Policy
- University of Melbourne. Management of Research Data and Records (Draft)
- University of New South Wales. Research Code of Conduct. Section 8. Management of Research Material and Data.
- University of New South Wales. Procedure for Handling Research Material and Data
- University of Newcastle. Research Data and Materials Management Policy
- University of Newcastle. Research Data and Materials Management Procedure

DMPonline

The DCC Data Management Planning Tool

Tools to help you plan…



Firefox File Edit View History Bookmarks Tools Window Help
15:29
DMP Online
http://dmponline.hatii.arts.gla.ac.uk/section/66/3
Google
DMP Online

DMPonline
The DCC 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

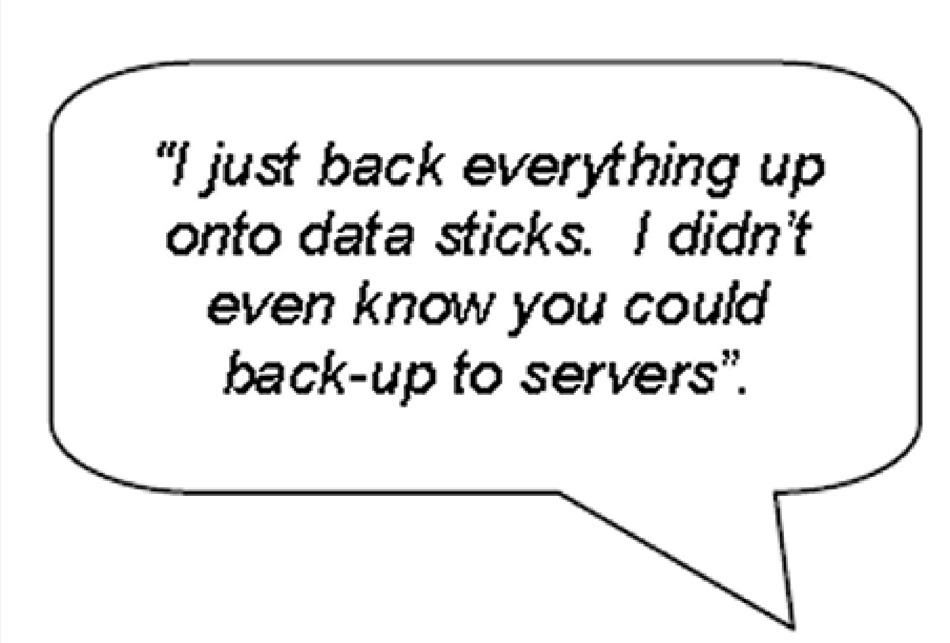
DCC Questions	Your Input	Guidance
3.1 Data sharing and re-use		
3.1.1 Will you share the data you capture or create? hide	● Yes ○ No	UKDA guidance on data sharing
3.2 Access		
3.2.2 How and when will you make the data available? hide	Data will be available online and open to all throughout the life of the project.	Here you will want 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? hide	Unlicenced, 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? hide		

DMPTool

Guidance and Resources for your Data Management Plan

Policy Gaps...

- Is Policy disconnected from Practice?
  - Data Sharing
  - Data Licensing
  - Ethics and Privacy
  - Citizen Science & Public Engagement
  - Data Storage, Selection & Appraisal
  - Data Citation and Attribution



"I just back everything up onto data sticks. I didn't even know you could back-up to servers".

http://www.flickr.com/photos/mattimattila/3003324844/

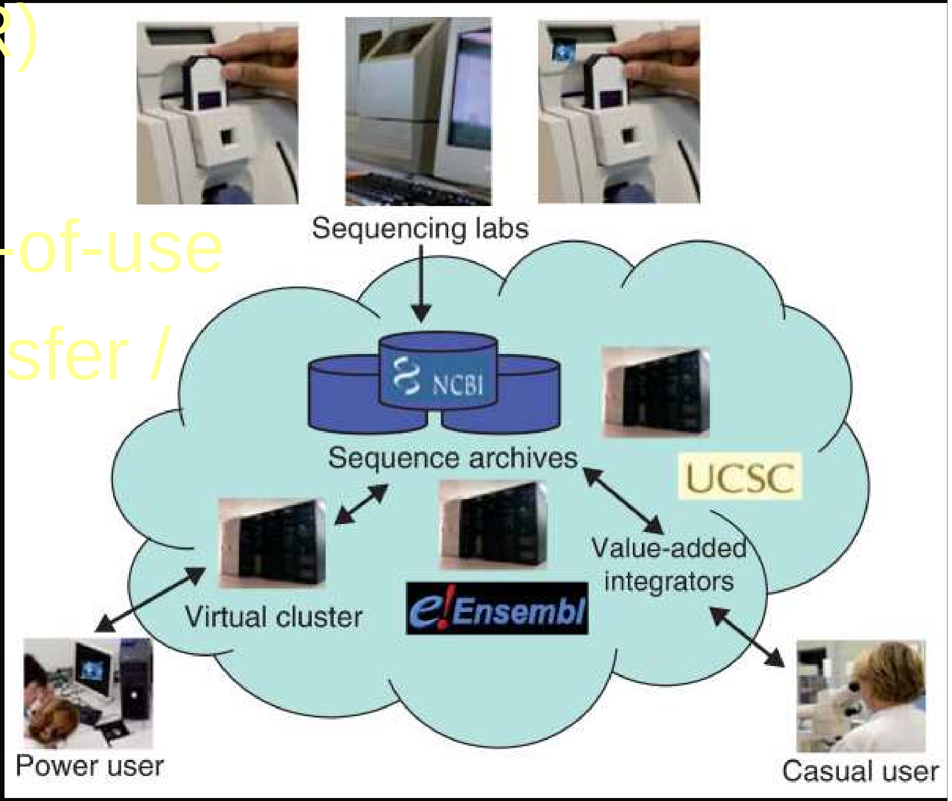
"Departments don't have guidelines or norms for personal back-up and researcher procedure, knowledge and diligence varies tremendously. Many have experienced moderate to catastrophic data loss"

Incremental Project Report, June 2010

Data storage...

- Scaleable
- Cost-effective (rent on-demand)
- Secure (privacy and IPR)
- Robust and resilient
- Low entry barrier / ease-of-use
- Has data-handling / transfer / analysis capability

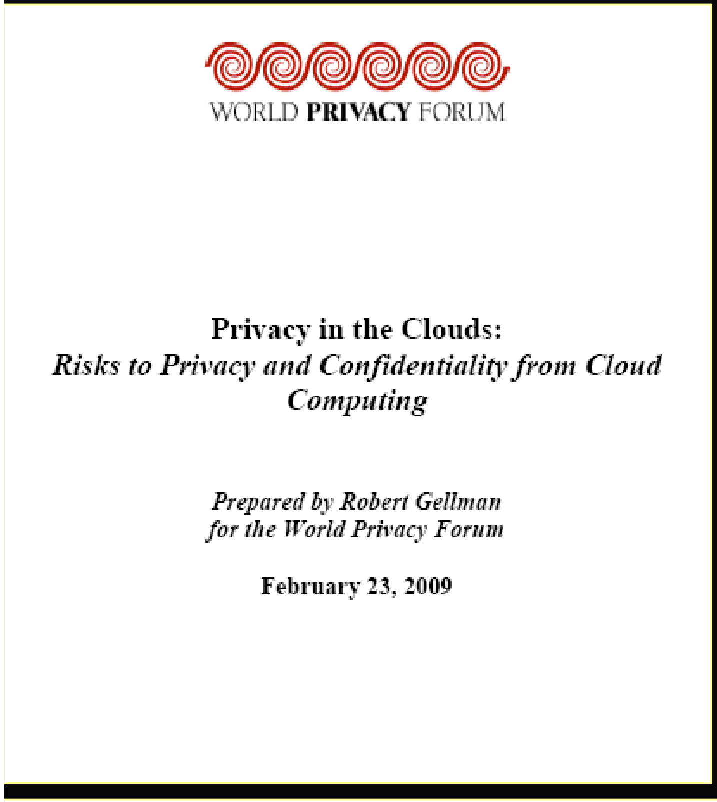
- Cloud services?



The case for cloud computing in genome informatics. Lincoln D Stein, May 2010

Genome Biology

Your data in the cloud


WORLD **PRIVACY** FORUM

Privacy in the Clouds:
Risks to Privacy and Confidentiality from Cloud Computing

Prepared by Robert Gellman
for the World Privacy Forum

February 23, 2009

Eduserv Symposium 2011

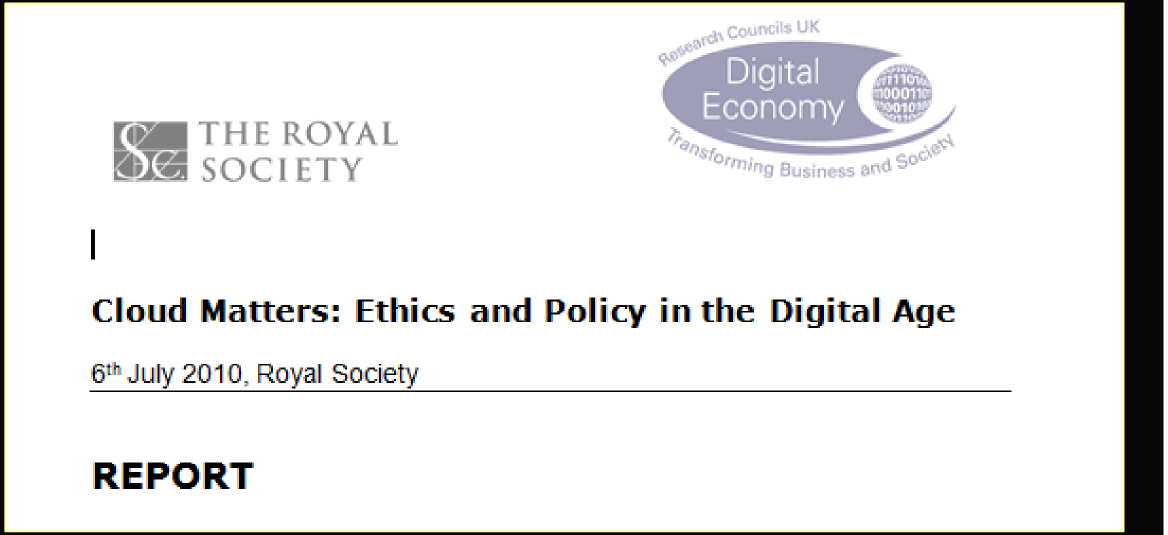
University
Collaboration
Thinking
Challenge
Modernisation
Cloud
Shared
Savings
Data Centre
Public
Education

Virtualisation and the Cloud: Realising the benefits of shared infrastructure

Cloud Computing for Research

The *Window* Conference Centre, London, Tuesday 20 July 2010

RESEARCH
COUNCILS UK

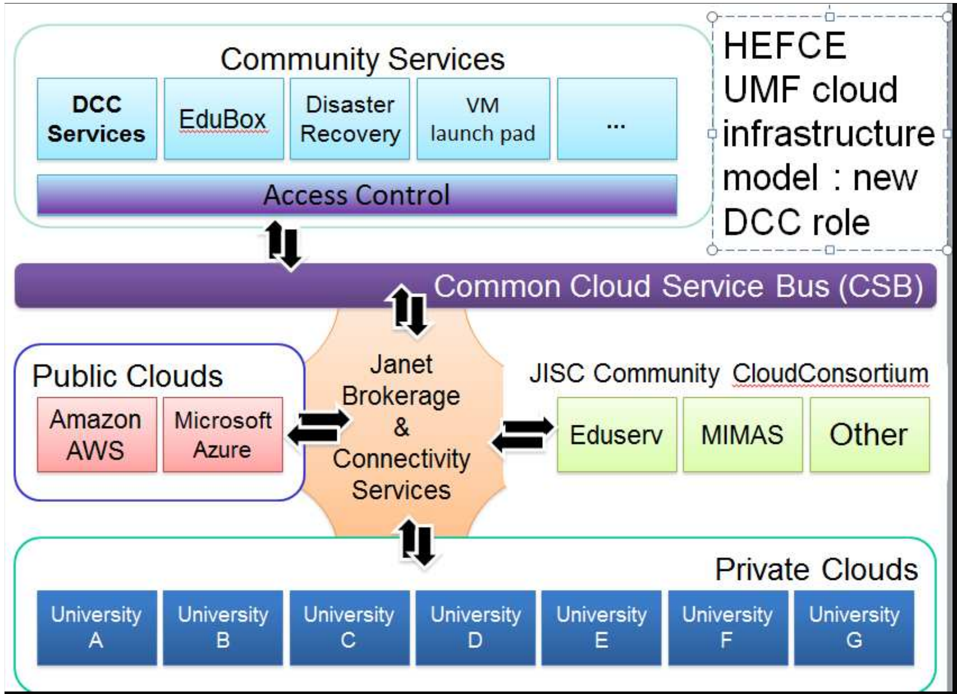

THE ROYAL SOCIETY

Research Councils UK
Digital Economy
Transforming Business and Society

Cloud Matters: Ethics and Policy in the Digital Age

6th July 2010, Royal Society

REPORT



HEFCE UMF cloud infrastructure model : new DCC role

DataFlow

ViDaaS

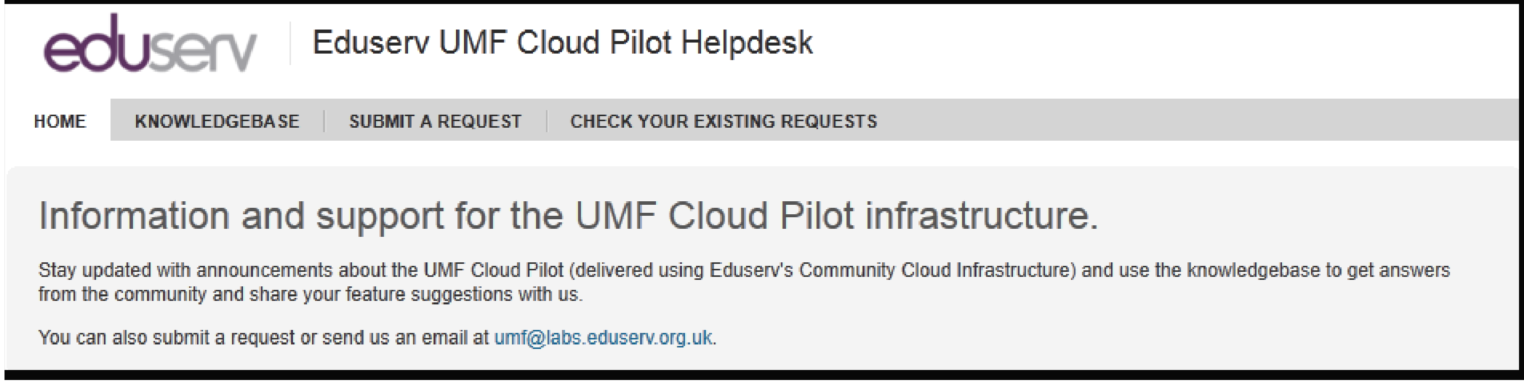
Virtual Infrastructure with Database as a Service

University of Leicester

BRISSkit: Biomedical Research Infrastructure Software Service kit

My Lab Notebook

UMF Shared Services & Cloud Programme



eduserv

Eduserv UMF Cloud Pilot Helpdesk

Information and support for the UMF Cloud Pilot infrastructure.

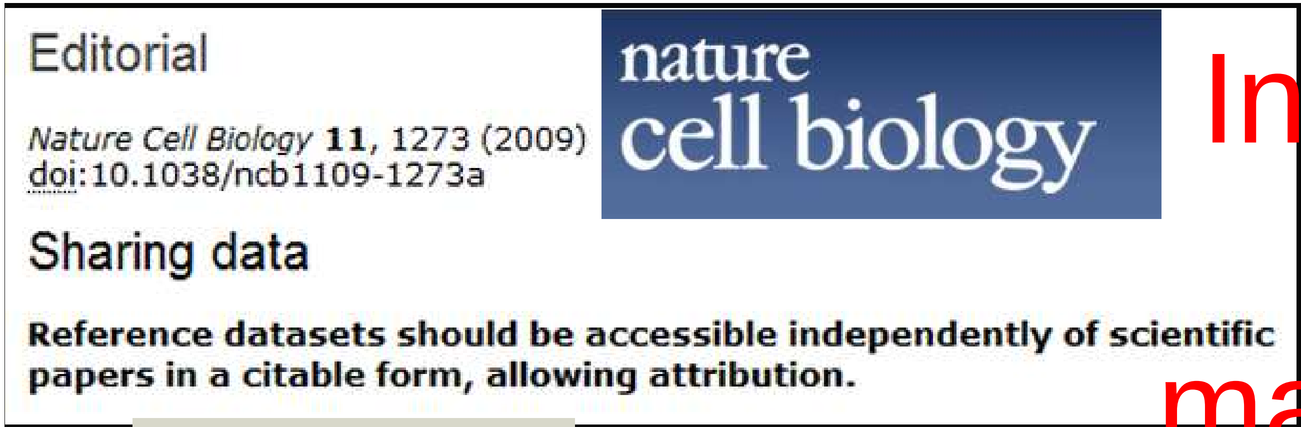
Stay updated with announcements about the UMF Cloud Pilot (delivered using Eduserv's Community Cloud Infrastructure) and use the knowledgebase to get answers from the community and share your feature suggestions with us.

You can also submit a request or send us an email at umf@labs.eduserv.org.uk.

JISC

HIGHER EDUCATION FUNDING COUNCIL hefce FOR ENGLAND

Incentivising data management

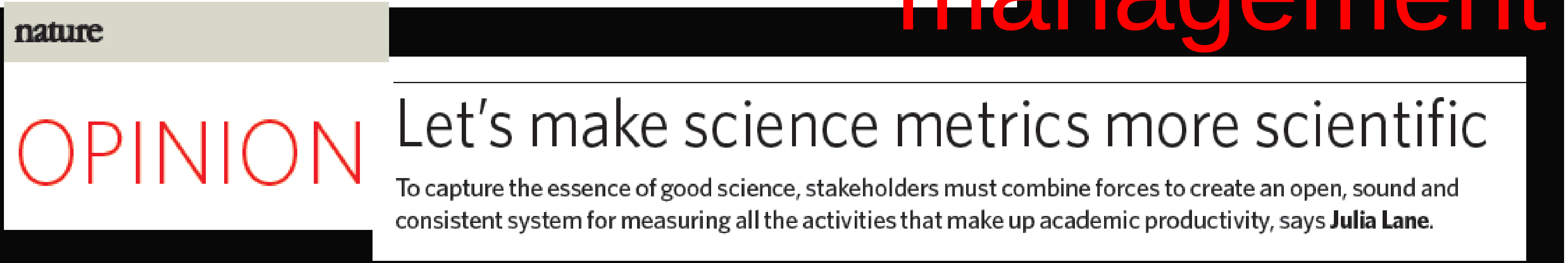


Editorial

Nature Cell Biology **11**, 1273 (2009)
doi:10.1038/ncb1109-1273a

Sharing data

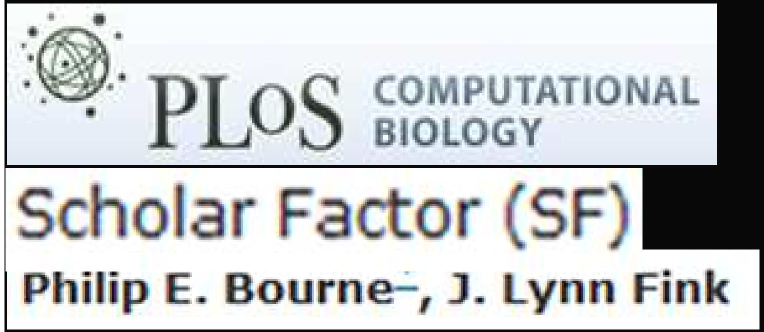
Reference datasets should be accessible independently of scientific papers in a citable form, allowing attribution.



OPINION

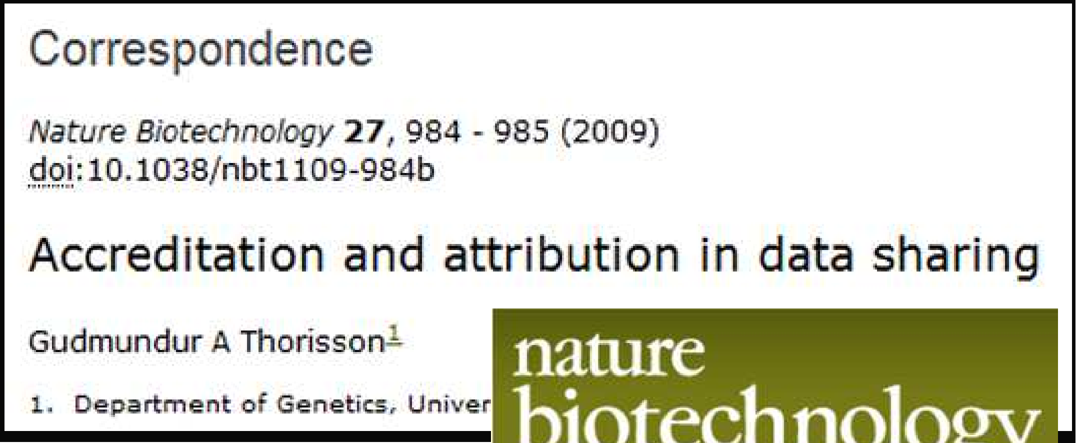
Let's make science metrics more scientific

To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.



Scholar Factor (SF)

Philip E. Bourne, J. Lynn Fink



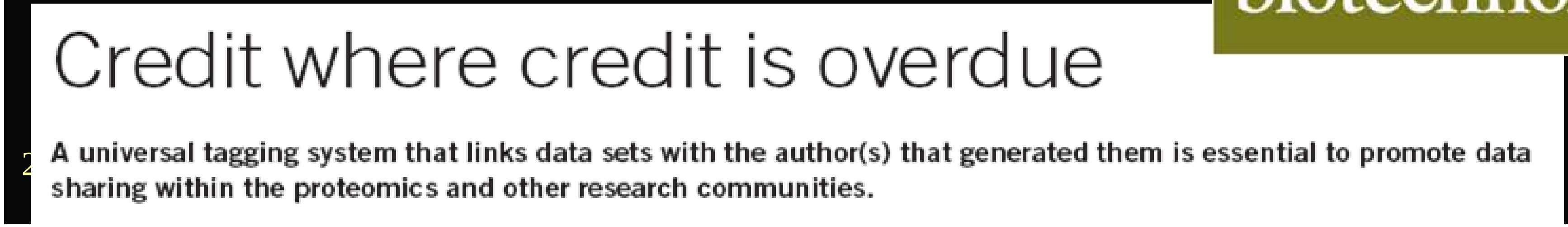
Correspondence

Nature Biotechnology **27**, 984 - 985 (2009)
doi:10.1038/nbt1109-984b

Accreditation and attribution in data sharing

Gudmundur A Thorisson[1]

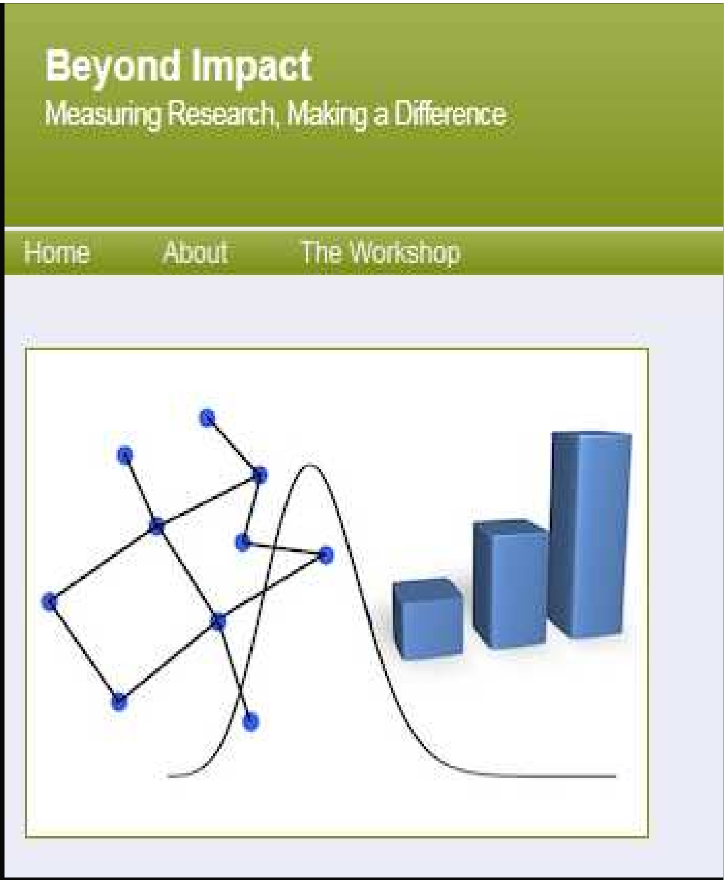
1. Department of Genetics, Univer



Credit where credit is overdue

A universal tagging system that links data sets with the author(s) that generated them is essential to promote data sharing within the proteomics and other research communities.

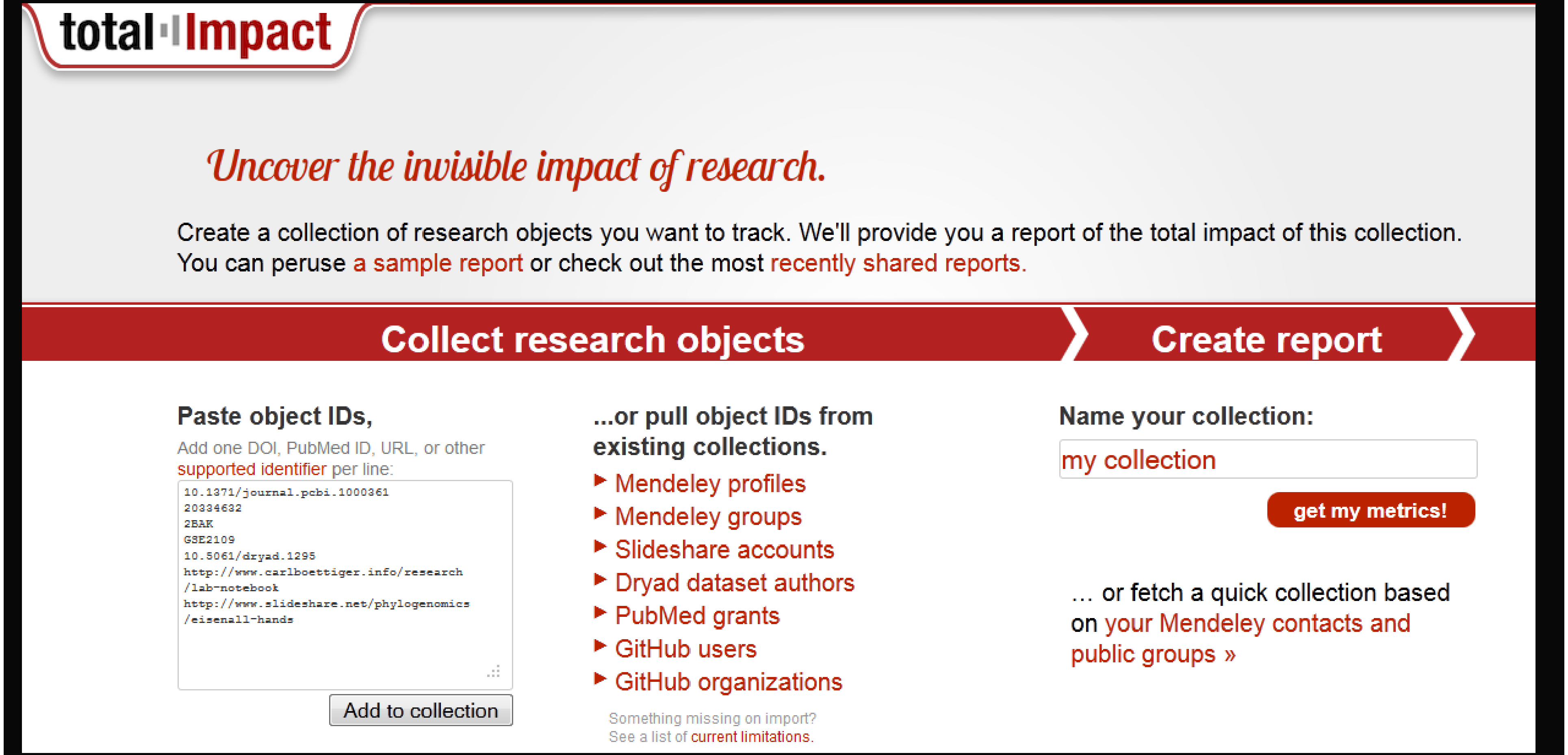
Beyond the PDF Workshop, January 2011



- Concept of “reproducibility”
- Executable papers
- Data papers
- Links to data, workflows, analyses (GenePattern) within a document
- Post-publication peer review
- Alternative impact metrics : downloads, slide reuse, data citation, YouTube views
- La Jolla Manifesto : guiding principles for digital scholarship

Jodi Schneider, Ariadne, Issue 66, January 2011

Tracking the impact of your data



total Impact

Uncover the invisible impact of research.

Create a collection of research objects you want to track. We'll provide you a report of the total impact of this collection. You can peruse a sample report or check out the most recently shared reports.

Paste object IDs,

Add one DOI, PubMed ID, URL, or other supported identifier per line:

```
10.1371/journal.pcbi.1000361
20334632
2BAK
GSE2109
10.5061/dryad.1295
http://www.carlboettiger.info/research
/lab-notebook
http://www.slideshare.net/phylogenomics
/eisenall-hands
```

...or pull object IDs from existing collections.

- Mendeley profiles
- Mendeley groups
- Slideshare accounts
- Dryad dataset authors
- PubMed grants
- GitHub users
- GitHub organizations

Something missing on import?
See a list of current limitations.

Name your collection:

my collection

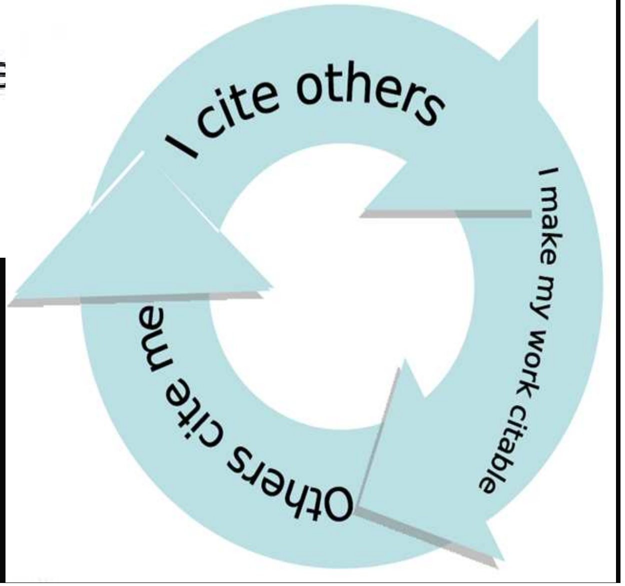
… or fetch a quick collection based on your Mendeley contacts and public groups »

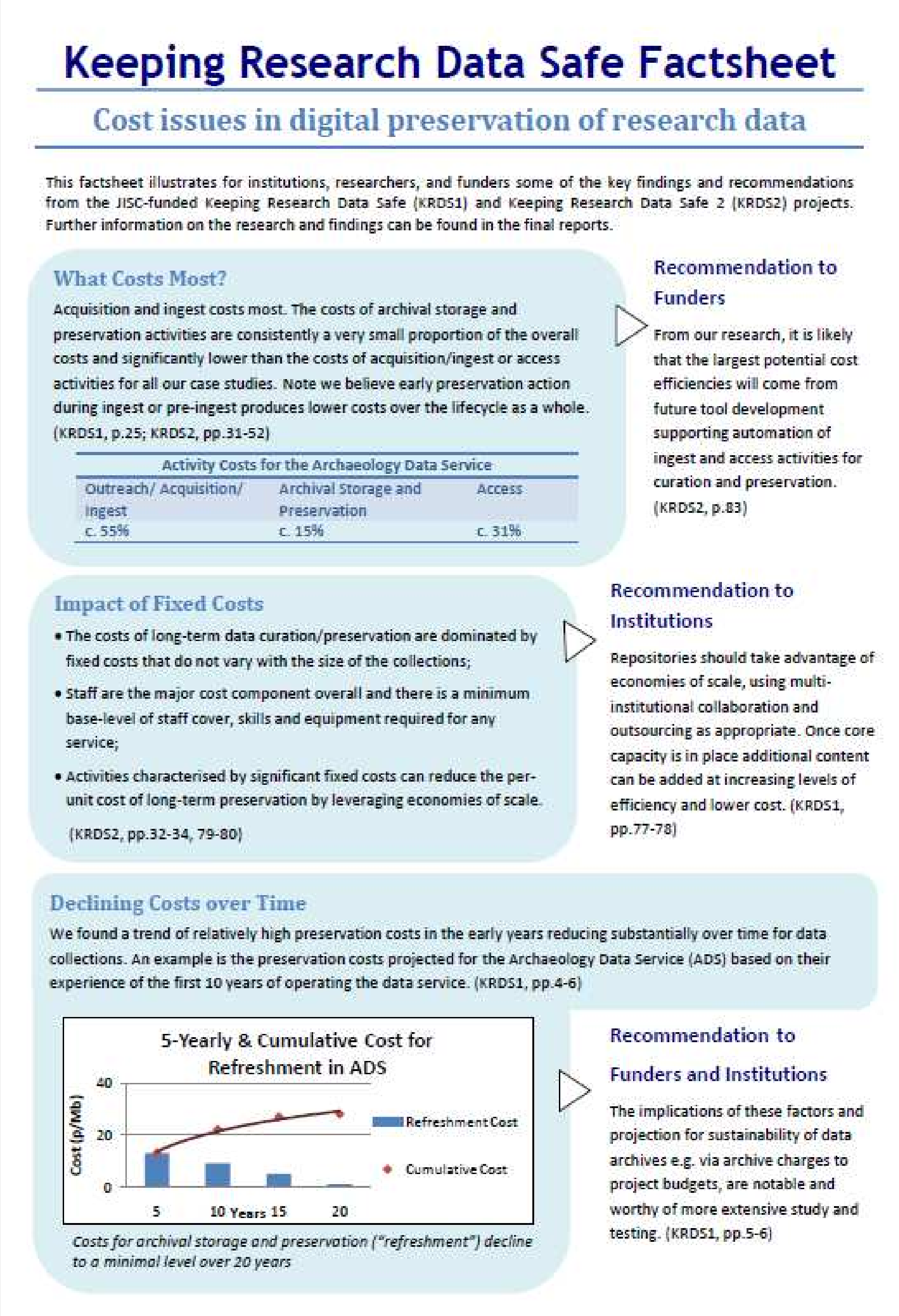
http://total-impact.org/

Citation Requirements

- Requirement 1 The Citation needs to be able to uniquely identify the object cited.
- Requirement 2 The Citation needs to support the retrieval of the cited object.
- Requirement 3 The citation mechanism must be compatible with Web infrastructure.
- Requirement 4 The citation 'system' must be able to generate a citation with all the desired fields
- Requirement 5 The citation mechanism must be identifier-agnostic and accomodation different resolution mechanisms
- Requirement 6 The citation mechanism must support gathering of metrics
- Requirement 7 The citation must be human readable
- Requirement 8 The citation must be machine processable
- Requirement 9 Support for bi-directional linking

SageCite





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 (KRDS1) and Keeping Research Data Safe 2 (KRDS2) 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/ Ingest	Archival Storage and Preservation	Access
c. 55%	c. 15%	c. 31%

Recommendation to Funders

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. (KRDS2, p.83)

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 per-unit cost of long-term preservation by leveraging economies of scale.

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

Recommendation to Institutions

Repositories should take advantage of economies of scale, using multi-institutional collaboration and outsourcing as appropriate. Once core capacity is in place additional content can be added at increasing levels of efficiency and lower cost. (KRDS1, pp.77-78)

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 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)

Costs, Benefits Value: KRDS

Charles Beagrie

KEEPING RESEARCH DATA SAFE 2

Neil Beagrie, Brian Lavoie and Matthew Woollard

with contributions by the Universities of Cambridge, Oxford, and Southampton, the Archaeology Data Service, OCLC Research, UK Data Archive, and University of London Computer Centre.

Final Report - April 2010

Prepared by:

Charles Beagrie Limited

www.beagrie.com

A study funded by

JISC

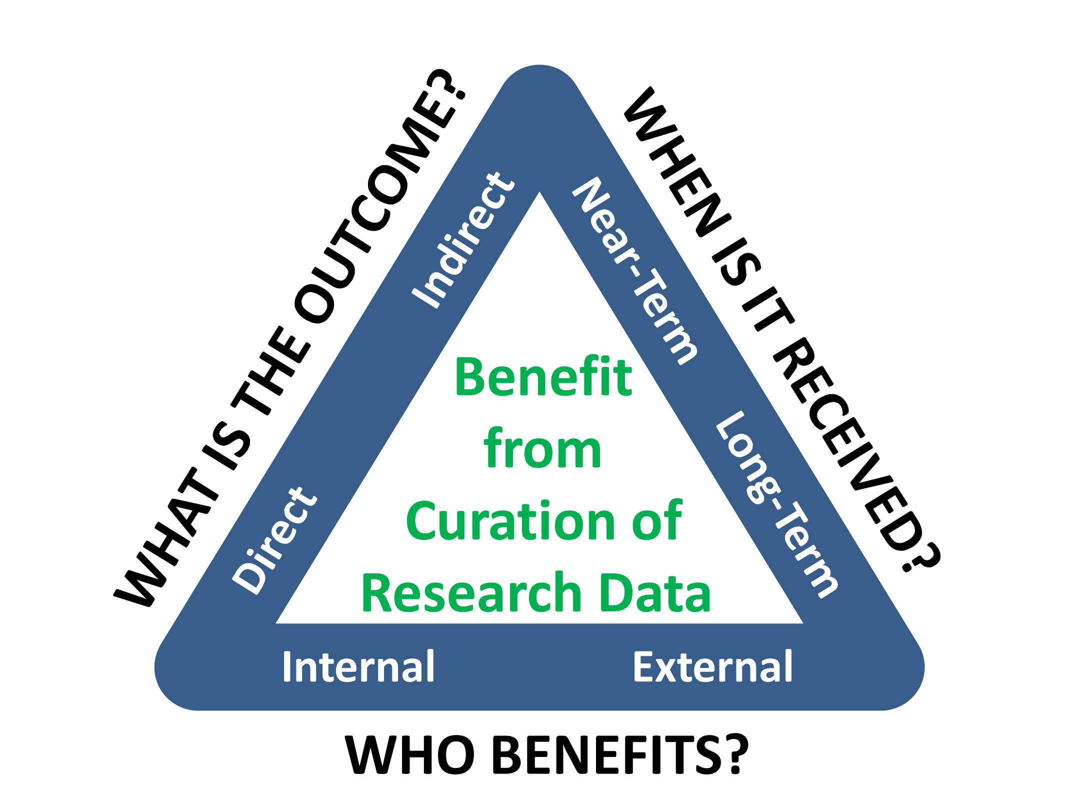
With support from OCLC Research and the UK Data Archive

Copyright HEFCE 2010

The authors have asserted their moral rights in this work

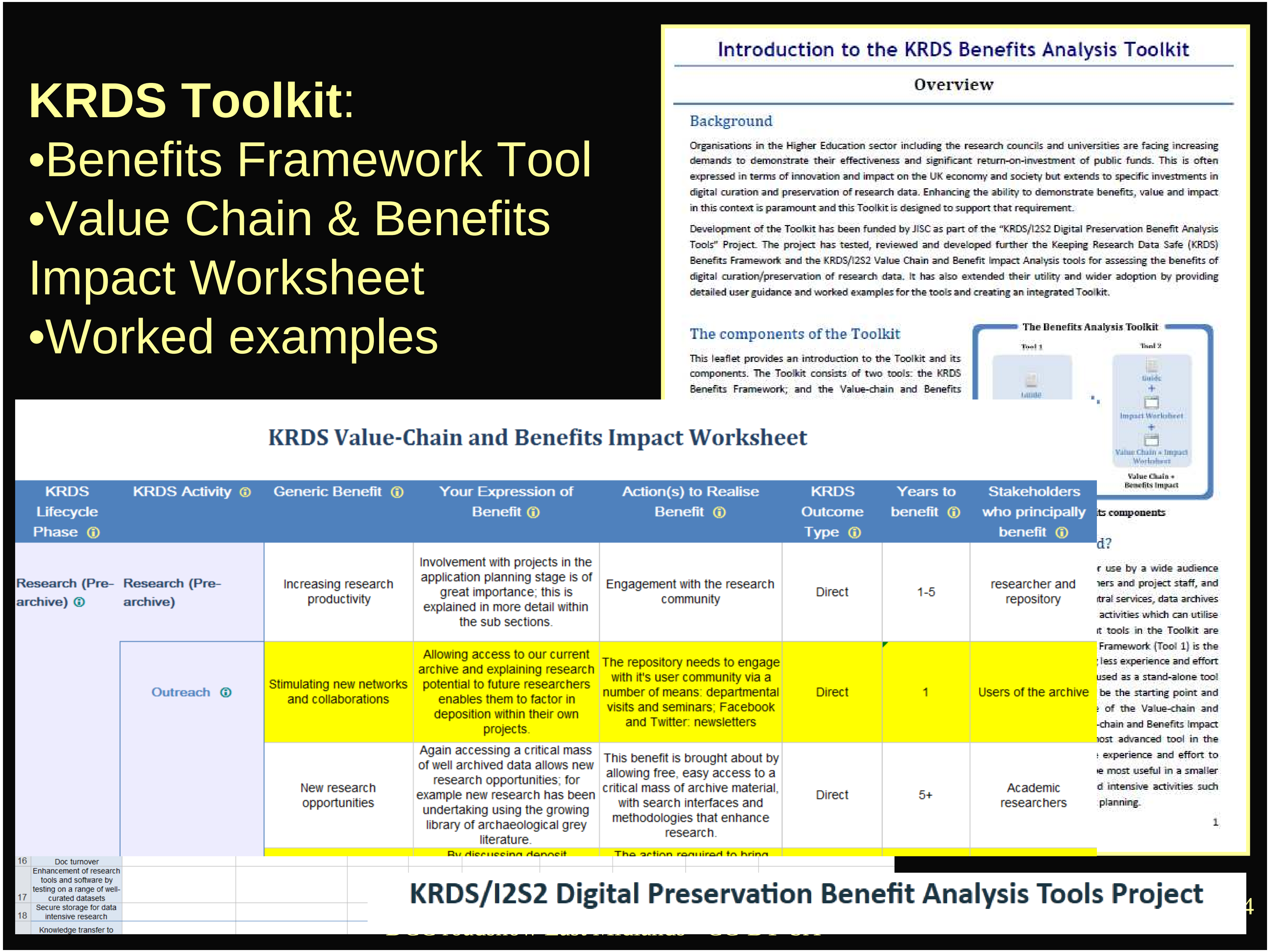
Benefits Framework

- Framework arranged on 3 dimensions with two sub-divisions each; Pick list of common generic benefits
- Individual benefits identified and assigned within this



KRDS Toolkit:

- Benefits Framework Tool
- Value Chain & Benefits Impact Worksheet
- Worked examples



Introduction to the KRDS Benefits Analysis Toolkit

Overview

Background

Organisations in the Higher Education sector including the research councils and universities are facing increasing demands to demonstrate their effectiveness and significant return-on-investment of public funds. This is often expressed in terms of innovation and impact on the UK economy and society but extends to specific investments in digital curation and preservation of research data. Enhancing the ability to demonstrate benefits, value and impact in this context is paramount and this Toolkit is designed to support that requirement.

Development of the Toolkit has been funded by JISC as part of the "KRDS/I2S2 Digital Preservation Benefit Analysis Tools" Project. The project has tested, reviewed and developed further the Keeping Research Data Safe (KRDS) Benefits Framework and the KRDS/I2S2 Value Chain and Benefit Impact Analysis tools for assessing the benefits of digital curation/preservation of research data. It has also extended their utility and wider adoption by providing detailed user guidance and worked examples for the tools and creating an integrated Toolkit.

The components of the Toolkit

This leaflet provides an introduction to the Toolkit and its components. The Toolkit consists of two tools: the KRDS Benefits Framework; and the Value-chain and Benefits

The Benefits Analysis Toolkit
Tool 1
Guide
Tool 2
Guide
+
Impact Worksheet
+
Value Chain + Impact Worksheet
Value Chain + Benefits Impact

ts components

d?

r use by a wide audience
hers and project staff, and
tral services, data archives
activities which can utilise
t tools in the Toolkit are
Framework (Tool 1) is the
less experience and effort
used as a stand-alone tool
be the starting point and
of the Value-chain and
-chain and Benefits Impact
ost advanced tool in the
experience and effort to
e most useful in a smaller
d intensive activities such
planning.

1

KRDS Value-Chain and Benefits Impact Worksheet

KRDS Lifecycle Phase ⓘ	KRDS Activity ⓘ	Generic Benefit ⓘ	Your Expression of Benefit ⓘ	Action(s) to Realise Benefit ⓘ	KRDS Outcome Type ⓘ	Years to benefit ⓘ	Stakeholders who principally benefit ⓘ
Research (Pre-archive) ⓘ	Research (Pre-archive)	Increasing research productivity	Involvement with projects in the application planning stage is of great importance; this is explained in more detail within the sub sections.	Engagement with the research community	Direct	1-5	researcher and repository
	Outreach ⓘ	Stimulating new networks and collaborations	Allowing access to our current archive and explaining research potential to future researchers enables them to factor in deposition within their own projects.	The repository needs to engage with it's user community via a number of means: departmental visits and seminars; Facebook and Twitter: newsletters	Direct	1	Users of the archive
		New research opportunities	Again accessing a critical mass of well archived data allows new research opportunities; for example new research has been undertaking using the growing library of archaeological grey literature.	This benefit is brought about by allowing free, easy access to a critical mass of archive material, with search interfaces and methodologies that enhance research.	Direct	5+	Academic researchers
			By discussing deposit	The action required to bring			

16	Doc turnover
17	Enhancement of research tools and software by testing on a range of well-curated datasets
18	Secure storage for data intensive research
	Knowledge transfer to

KRDS/I2S2 Digital Preservation Benefit Analysis Tools Project

The DCC Mission

- Helping to build capacity, capability and skills in data management and curation across the UK’s higher education research community

  – *DCC Phase 3 Business Plan*

because good research needs good data

DCC services (more later…)

- Roadshows
- Institutional engagements
- Research data management forum (RDMF)
- Training programme
- ‘How-to’ guides
- Briefing papers
- Curation tools and services
- International conference (IDCC)
- Support to JISC MRD programme

because good research needs good data