Evolution or revolution?
The changing data landscape

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

4th DCC Regional Roadshow, Oxford, September 2011

UKOLN is supported by: **JISC**

D|C|C

www.ukoln.ac.uk

UNIVERSITY OF BATH
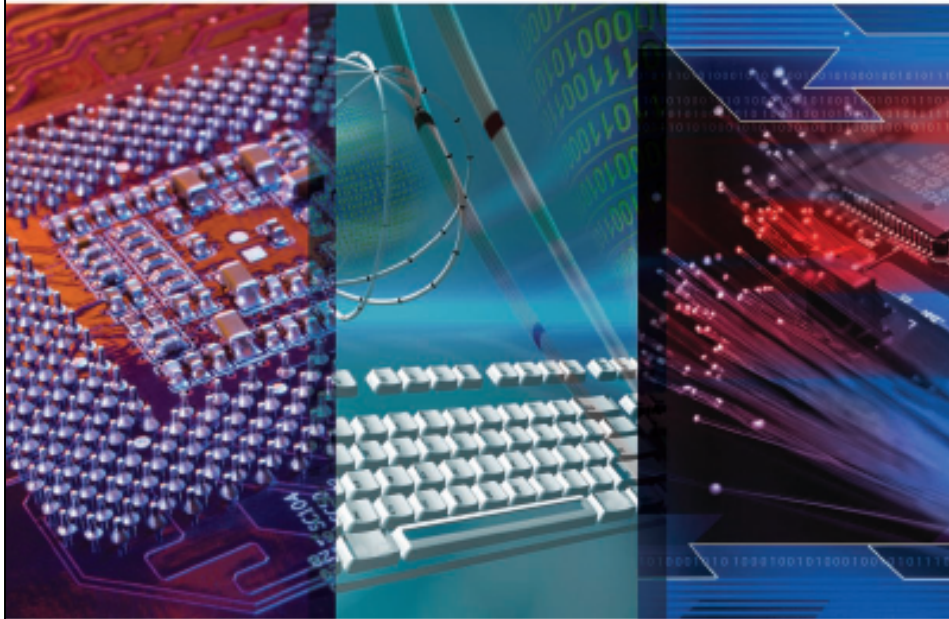
A centre of expertise in digital information management

RCUK Review of e-Science 2009

BUILDING A UK FOUNDATION FOR THE TRANSFORMATIVE
ENHANCEMENT OF RESEARCH AND INNOVATION

RESEARCH COUNCILS UK

THE ROYAL SOCIETY

"Data sets are becoming the new instruments of science"

Dan Atkins, Univ Michigan

Digital data as the new special collections?

Sayeed Choudhury, Johns Hopkins

Research data : institutional crown jewels?

guardian.co.uk

News | Sport | Comment | Culture | Business | Money | Life & style

News › Technology

Give us back our crown jewels

Our taxes fund the collection of public data - yet we pay again to access it. Make the data freely available to stimulate innovation, argue Charles Arthur and Michael Cross

Charles Arthur and **Michael Cross**
The Guardian, Thursday 9 March 2006
Article history

http://www.flickr.com/photos/lifes__too_short__to__drink__cheap__wine/4754234186/

Perspectives

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

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

"Surfing the Tsunami"
Science: 11 February 2011

"The costs of sequencing DNA has taken a nosedive...and is now dropping by 50% every 5 months".

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

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

**"I worry there won't be enough people around to do the analysis."
Chris Ponting, University of Oxford**
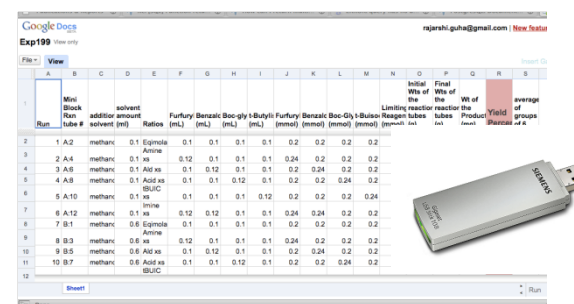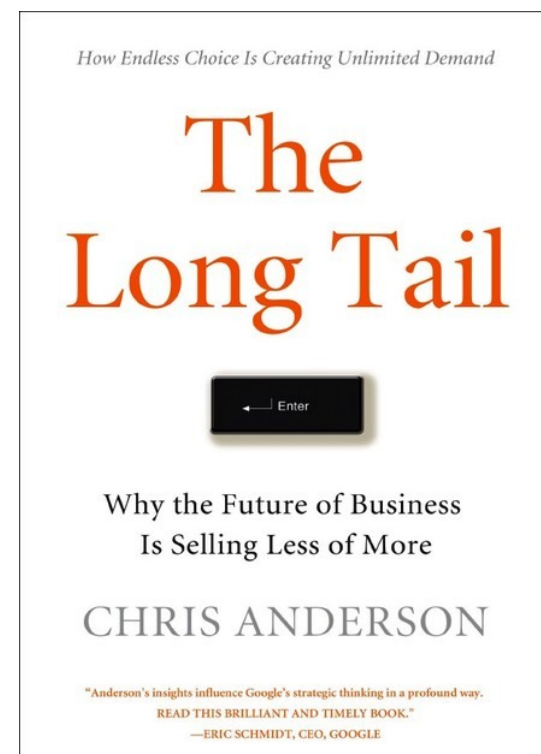
Data collections

GenBank

PDB

UniProt

Pfam

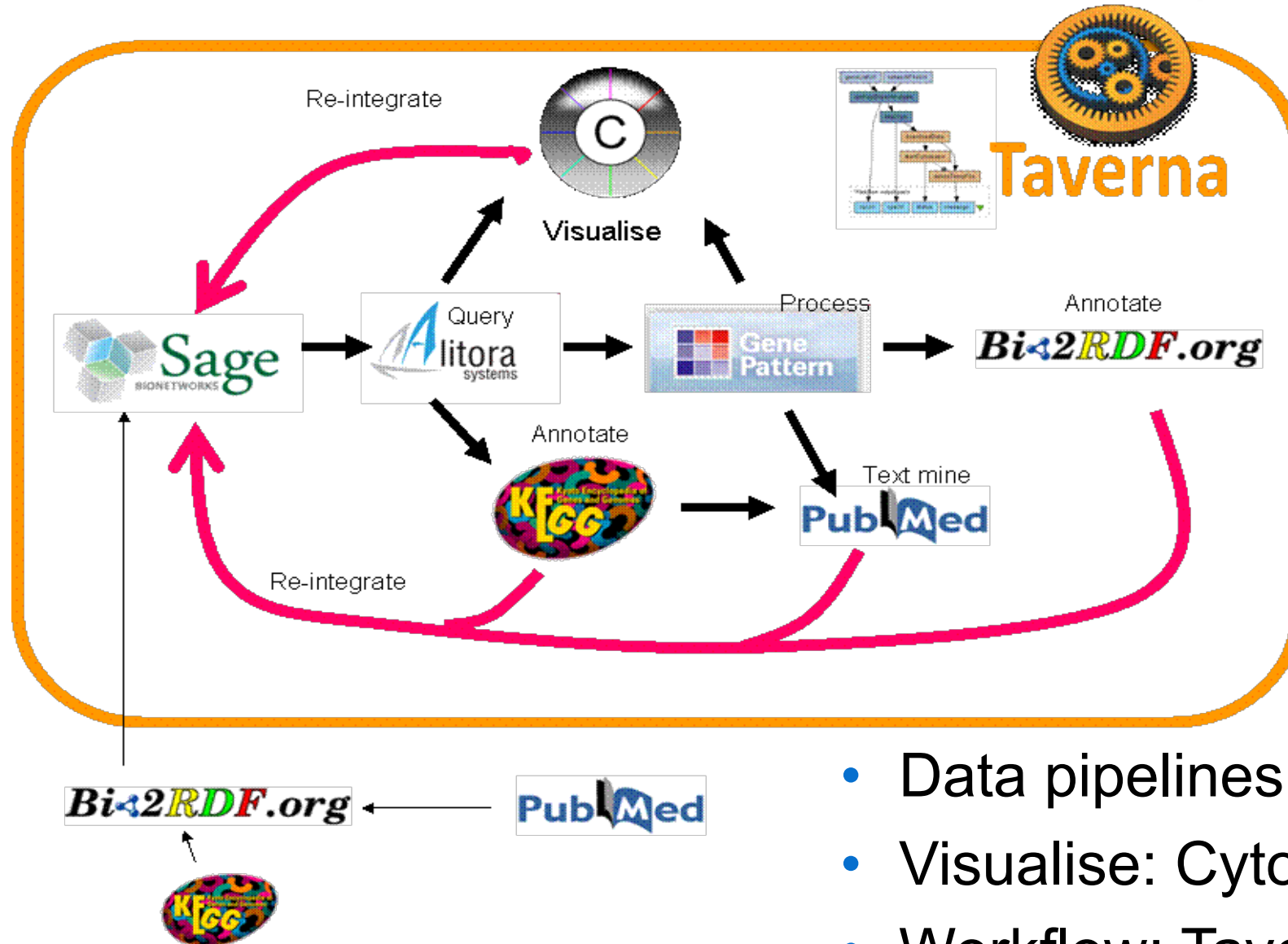ChemSpider

CATH, SCOP
(Protein
Structure
Classification)

Spreadsheets, Notebooks
Local, Lost

Slide: Carole Goble
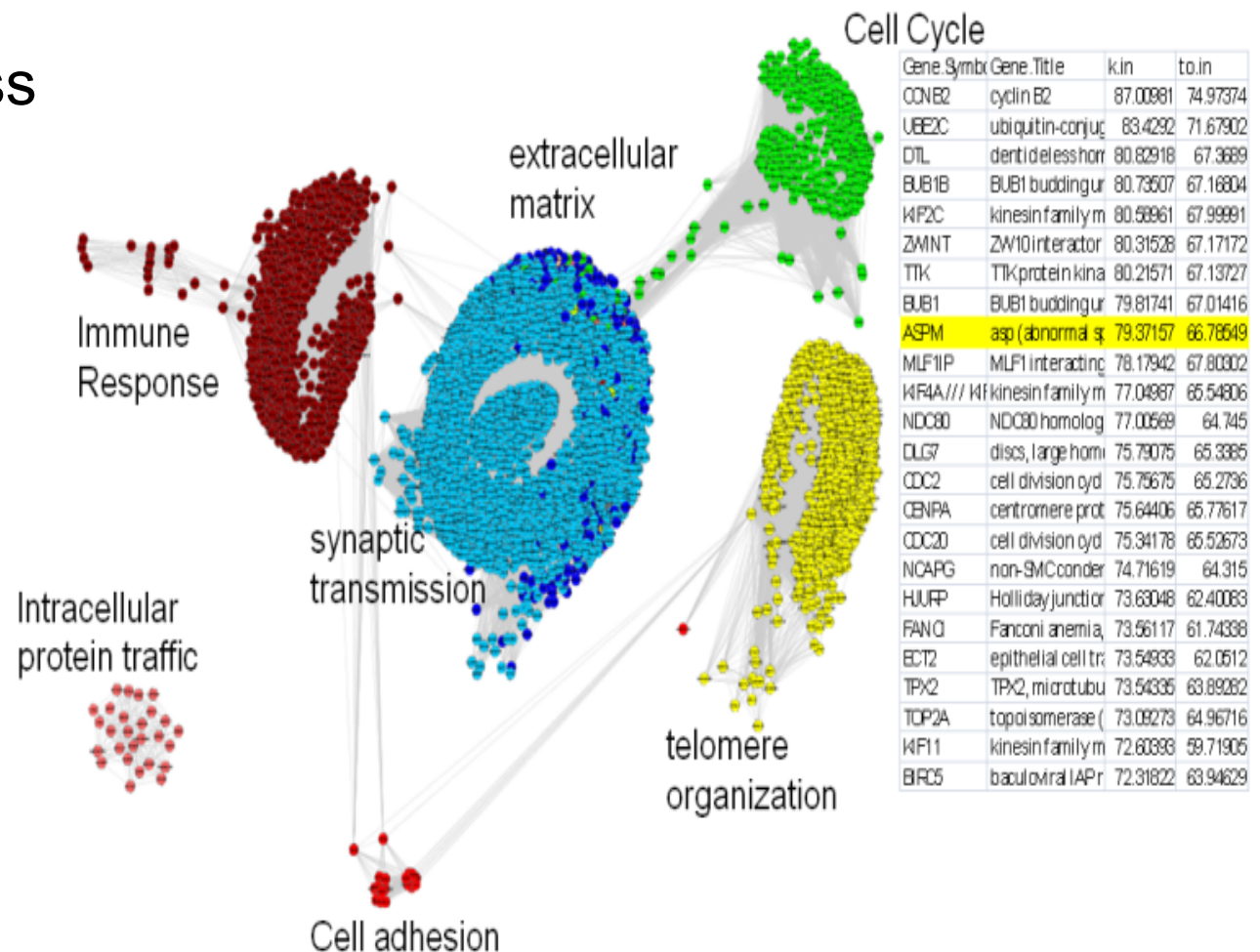
Complexity challenges

- Data pipelines
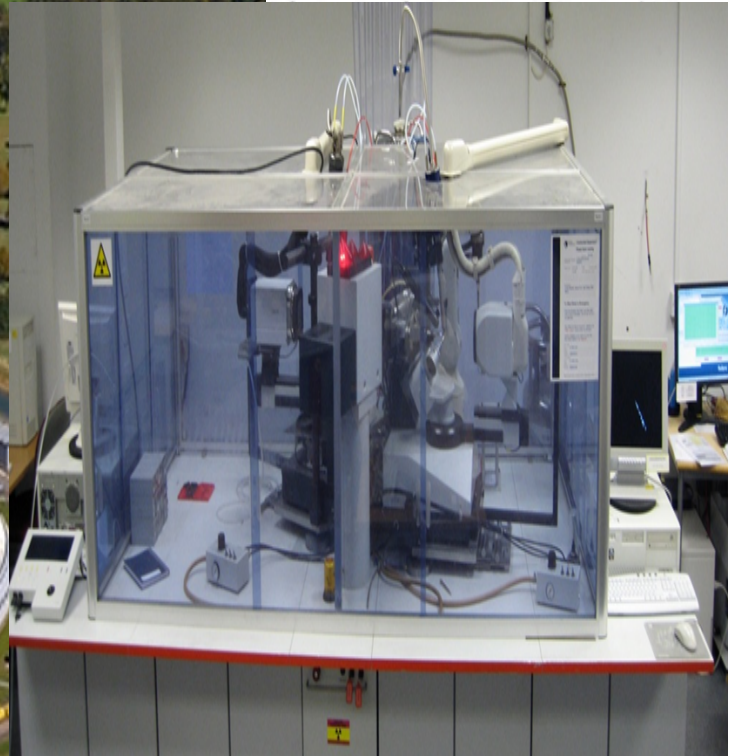- Visualise: Cytoscape
- Workflow: Taverna

- Distributed gene expression & clinical traits data
- Workflows capture the complex model construction process
- Derive large-scale bionetwork models
- Use to predict disease patterns

TCGA GBM Coexpression Network

Structural Sciences Infrastructure

Infrastructure Roadmap

Infrastructure Roadmap

Cross Disciplines
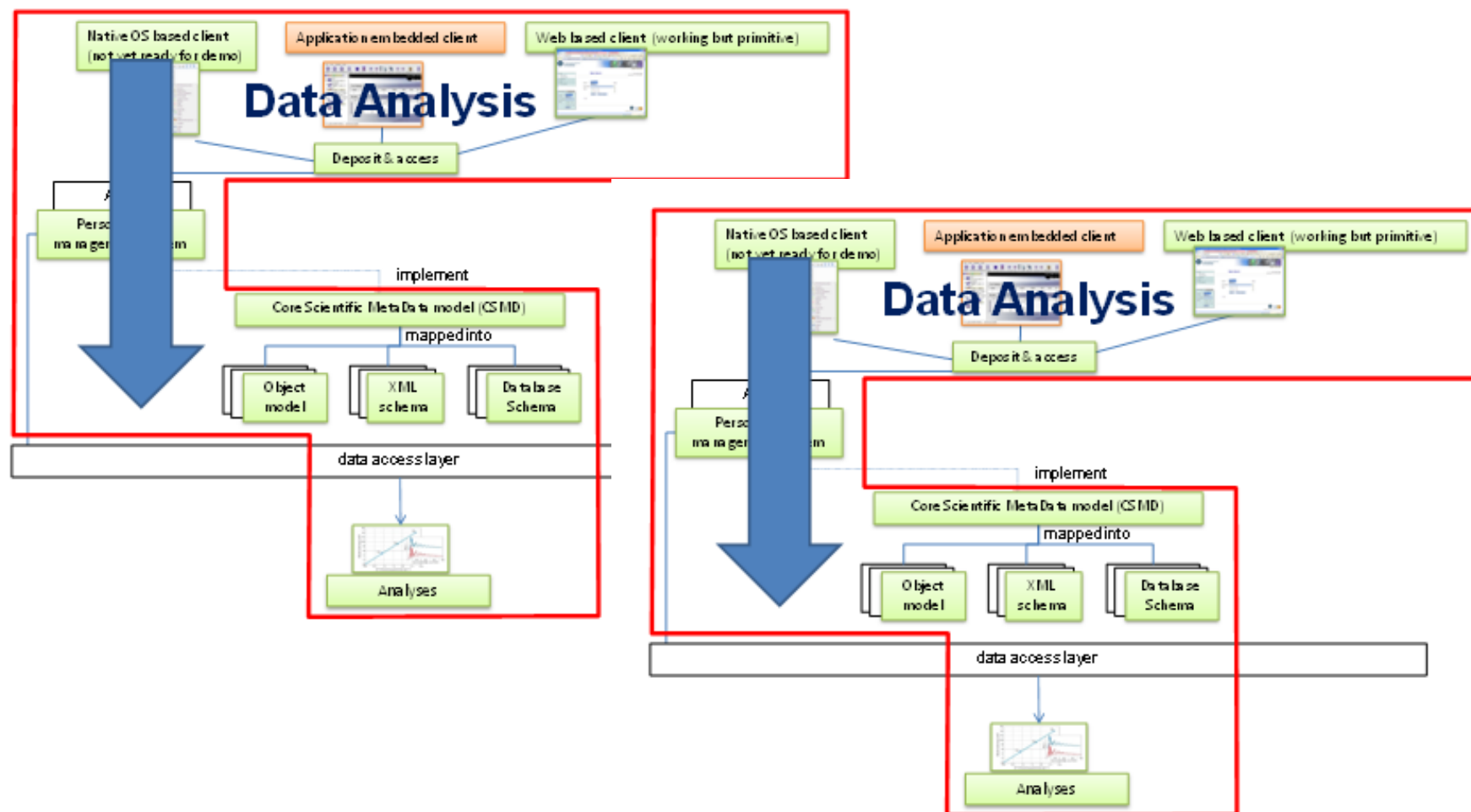
Infrastructure Roadmap
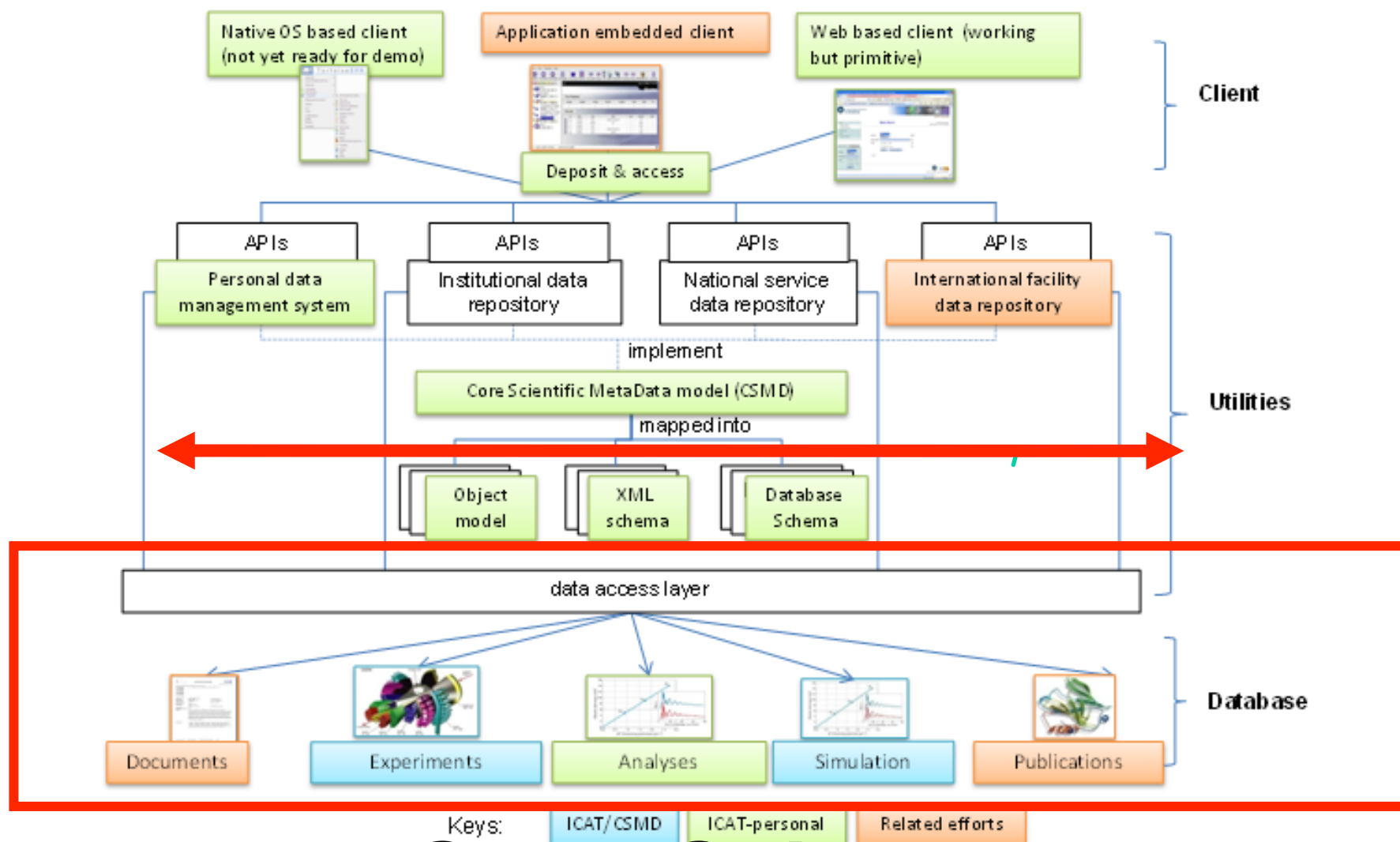
Panton Principles

Principles for Open Data in Science

OPEN **KNOWLEDGE**

OPEN **DATA**

OPEN **CONTENT**

OPEN **SERVICE**

| D | C | C

JISC**legal**
information

How to License Research Data

Alex Ball (DCC)

Open Definition

Defining the Open in Open Data, Open Content and Open Services

2011: Citizens getting involved in science

project noah

Document nature with your mobile phone.

army of **citizen** scientists

Yasser Ansari

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

My Noah

My Spottings   My Missions

My Patches   Settings

New Spotting

Available on the
App Store

Download for
Android

Citizen as scientist

scienceforcitizens.net

GALAXY ZOO

HUBBLE

Home   The Story So Far   How To Take Part   Classify Galaxies   Explore Galaxies   The Science   FAQ

Forum   Blog   Contact Us

Pictures

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.

The New York Times

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.

RECOMMEND
TWITTER
COMMENTS (18)
SIGN IN TO E-MAIL
PRINT
SINGLE-PAGE
REPRINTS
SHARE

Enlarge This Image

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

MARTHA MARCY MAY MARLENE

Working with academics

Validate results data and publish

Letter

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

Patients Participate!

Citizen-patients producing crowd-sourced lay summaries of UK PubMed Central papers

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

- Bridging the Gap
- Feasibility pilot study
- Stem cell research
- Case studies
- Guidance for academics and citizens
- Final Report forthcoming
- Talk Science event at British Library 18 October

Community Capability Model for Data-Intensive Research
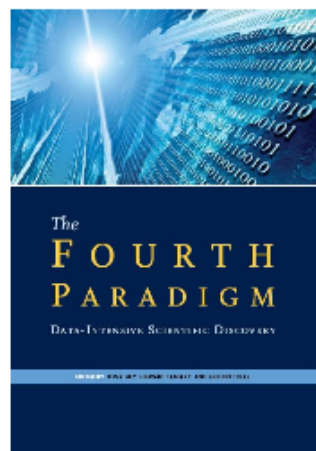
Microsoft Research Connections and UKOLN are working in partnership on an exciting new project to develop a Community Capability Model for Data-Intensive Research, building upon the principles described in *The Fourth Paradigm*. This second consultation workshop will focus on discussing and describing scholarly communications to enable data-intensive research, such as collaborative authoring platforms, common data formats and identifiers, data-sharing, data citation and socio-legal issues.

The ultimate aim is to provide a framework that is useful for researchers and funders in modelling a range of disciplinary and community behaviours with respect to the adoption, usage, development and exploitation of cyber-infrastructure for data-intensive research.

Find out more on the project details page.

Events

UK e-Science All Hands Meeting, 27-29 September 2011, York, UK [Sept. 28 workshop agenda]

2011 Microsoft eScience Workshop: Transforming Scholarly Communication, 23-25 October 2011, Cambridge, MA, USA [Oct. 23 workshop agenda]

UKOLN

UNIVERSITY OF BATH

Microsoft Research Connections

Disciplinary focus
Capability? "Maturity"?

Policy

RESEARCH COUNCILS UK

Excellence with Impact

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

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

○ This website
○ All Research Councils

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

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.

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

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

DCC

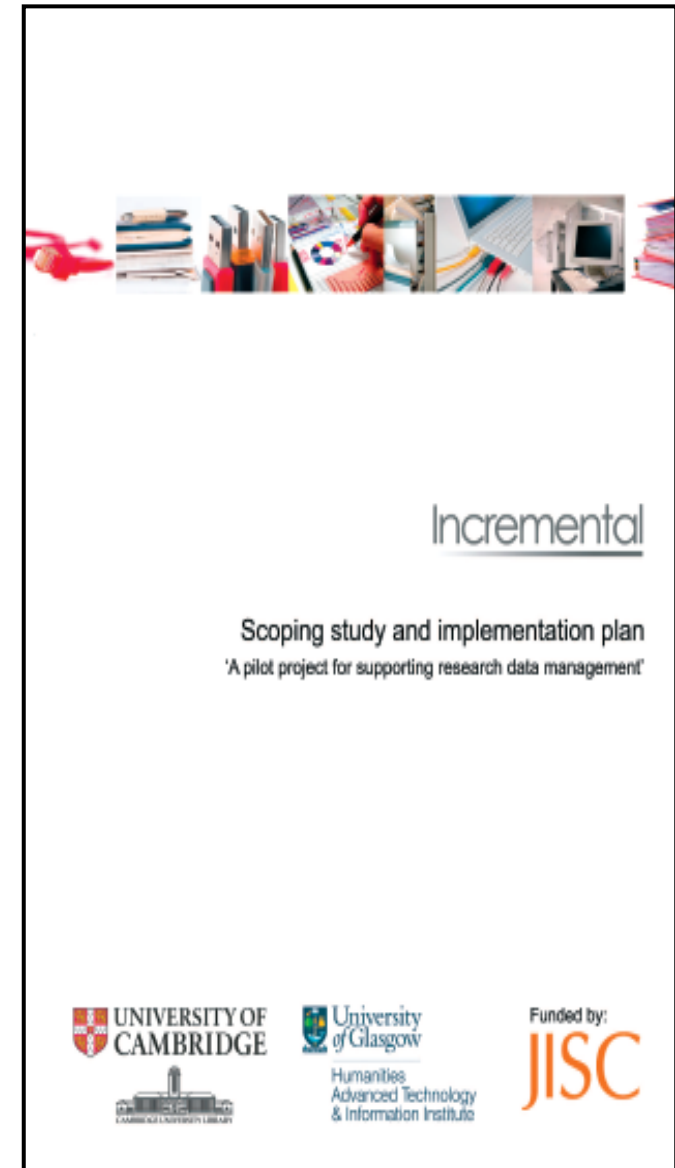- Creating & organising data
- Storage and access
- Back-up
- Preservation
- Sharing and re-use

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

The University of Edinburgh

Search term | Search the full site ▾ | Go   **Contact us**

Information Services

- Home
- Search IS
- Students
- Researchers & teachers
- Support staff
- About IS
- Services
- Library
- Computing
- Learning technology
- Research support

Related links
- Help
- Search IS
- Your feedback about this page

University Homepage › Schools & departments › Information Services › About › Policies and Regulations › Research Data Management Policy

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 next issue Int J Digital Curation

Institutional Policy

RESEARCH DATA MANAGEMENT
UAS

Enter search term — Search

- This site
- University of Oxford
- People

UNIVERSITY OF OXFORD

- 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.

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)

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

Training, Advice and Support

Institutional Policy

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.

Policy Summary from DCC

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

Policy summary from ANDS

ands
AUSTRALIAN NATIONAL DATA SERVICE

nd research data:

About ANDS
- Projects & Funding
- Our Approach
- Events

For Researchers
- Manage Data
- Publish Data
- Find Data

For Partner Institutions
- Make Connections

Managing Data
- Guides

Publishing Data
- Licensing
- Online Services
- Content Providers Guide
- Technical resources

News
- Newsletter

Community Bulletin Board

Institutional policies and procedures

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

- 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

UNIVERSITY of CALIFORNIA | *It starts here*

Data management planning tool in development

A group of major research institutions is partnering to develop a flexible online tool to help researchers generate data management plans. This effort is in response to demands from funding agencies, such as the National Science Foundation (NSF) and the National Institutes of Health (NIH), that researchers plan for managing their research data.

The partners in this project include the University of California Curation Center (UC3) at the California Digital Library, the UCLA Library, the UCSD Libraries, the Smithsonian Institution, the University of Virginia Library, the University of Illinois at Urbana-Champaign, DataONE, and the United Kingdom's Digital Curation Centre (DCC).

International collaboration around the DCC DMPOnline tool

D|C|C
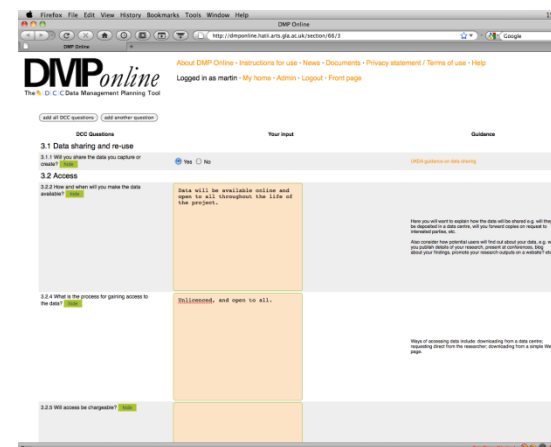
"It's hard to overcome your personal investment... it's like giving away your baby"

"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."

DCC

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

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

Data is headline news

FOI & RESEARCH DATA: RESEARCHERS' QUESTIONS AND ANSWERS
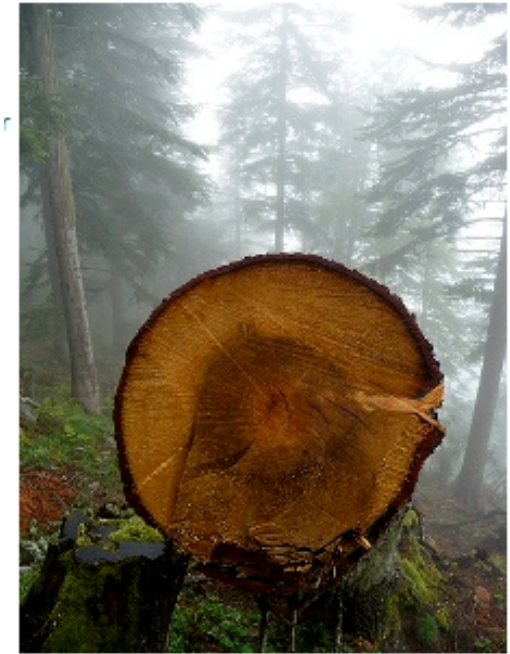
Posted by Daniel Cressey on April 15, 2010

JISC FoI FAQ

Table of Contents   Comments by Section   Comments by Users   General Comments   Login

TABLE OF CONTENTS

There are 51 comments in this document

The general rule is that the Data Protection Act trumps FoI/EIR. Both FoI Acts make personal data, of which the requester is the subject, exempt information (there is a similar exception under EIR). The requester should apply under the UK-wide Data Protection Act (for which different rules, timescales and fees apply). If the requester is not the subject of the personal data, the exemptions become more complicated, although our "general rule" above is likely to apply. Always discuss such c [...]

3 Comments

1000 Genomes Project Releases Data from Pilot Projects on Path to Providing Database for 2,500 Human Genomes

Freely available data supporting next generation of human genetic research

1000 Genomes
A Deep Catalog of Human Genetic Variation
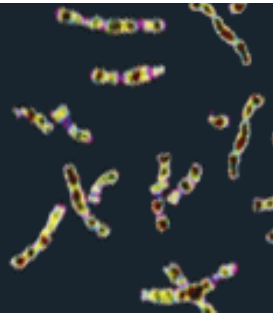
Desmond Tutu's genome sequenced as part of genetic diversity study

Archbishop Desmond Tutu has had his genome sequenced in research to reveal the true breadth of human genetic diversity

Ian Sample, science correspondent
guardian.co.uk, Wednesday 17 February 2010 18.02 GMT
Article history
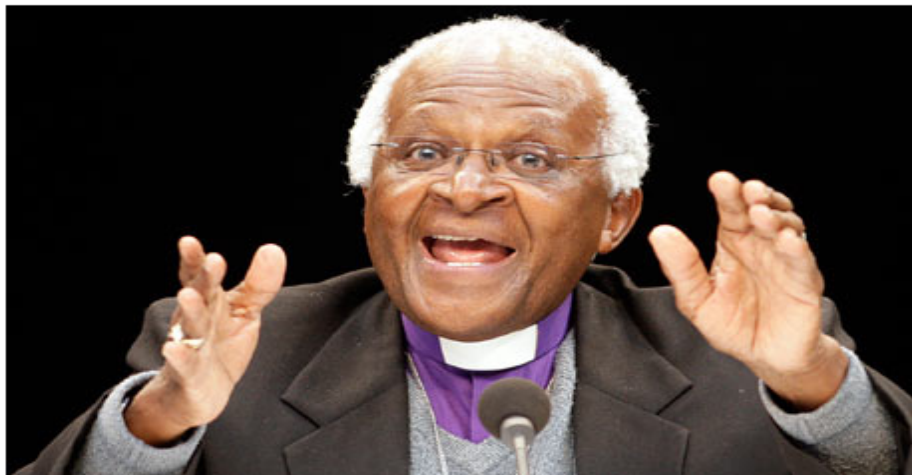
P4 medicine: Predictive, Personalised, Preventive, Participatory.

Leroy Hood –
Institute for Systems Biology

Your genome is basis for your medical record

Open data and ethics

23andMe

Buy a DIY kit?
Share <u>your</u> data?

Get the latest on your DNA with $399 and a tube of saliva.

Here's what you do:

1. Order a kit ($399 USD) from our online store.

2. Claim your kit, spit into the tube, and send it to the lab.
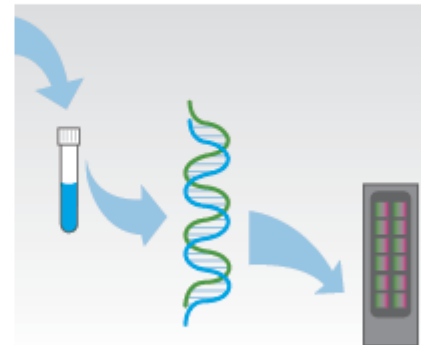
3. Our CLIA-certified lab analyzes your DNA in 2-4 weeks.

4. Log in and start exploring your genome.

Open data and ethics

- **Bring your genes to CAL**
- UC Berkeley personalised medicine initiative in 2010
- >700 new students have submitted a genetic sample and a consent form
- Aggregate analyses for three genes related to nutrition
- Constrained by State Law
- Implications for UK HE & staff?

Berkeley
UNIVERSITY OF CALIFORNIA

ON THE
Same Page
BRING YOUR **GENES** TO CAL

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".

"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
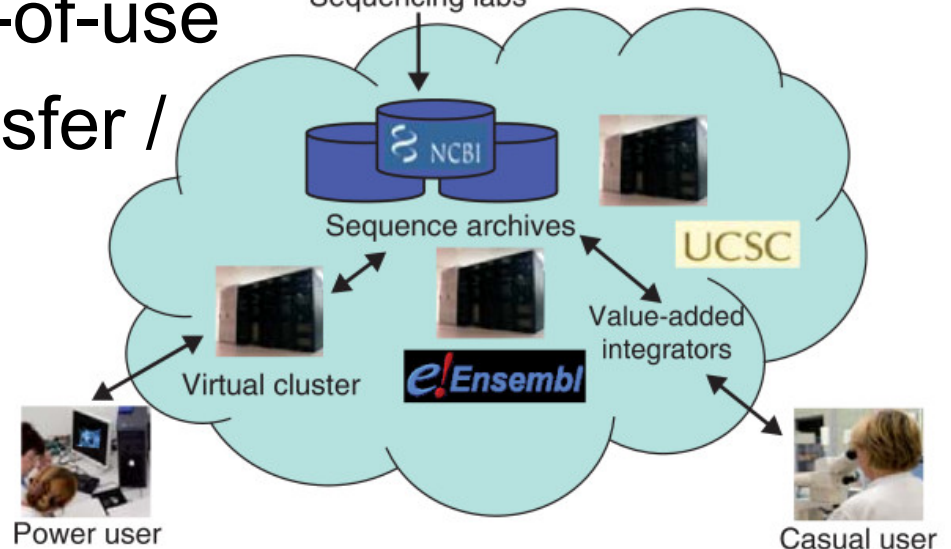
• Cloud services?

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

Genome **Biology**

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

Virtualisation and the Cloud: Realising the benefits of shared infrastructure

Your data in the cloud

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

UMF Shared Services & Cloud Programme

Community Services

DCC Services	EduBox	Disaster Recovery	VM launch pad	...

Access Control

HEFCE UMF cloud infrastructure model : new DCC role

Common Cloud Service Bus (CSB)

Public Clouds
Amazon AWS	Microsoft Azure

Janet Brokerage & Connectivity Services

JISC Community CloudConsortium
Eduserv	MIMAS	Other

Private Clouds
University A	University B	University C	University D	University E	University F	University G

DataFlow

ViDaaS — Virtual Infrastructure with Database as a Service

University of Leicester

BRISSkit: Biomedical Research Infrastructure Software Service kit

My Lab Notebook

eduserv | Eduserv UMF Cloud Pilot Helpdesk

HOME   KNOWLEDGEBASE   SUBMIT A REQUEST   CHECK YOUR EXISTING REQUESTS

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

hefce — HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND

Incentivising data management

nature cell biology

Sharing data

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

nature

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**.

PLoS COMPUTATIONAL BIOLOGY

Scholar Factor (SF)

Philip E. Bourne, J. Lynn Fink

Accreditation and attribution in data sharing

Gudmundur A Thorisson[1]

1. Department of Genetics, Univer

nature biotechnology

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

Demonstrator:
Citation of
bionetwork models
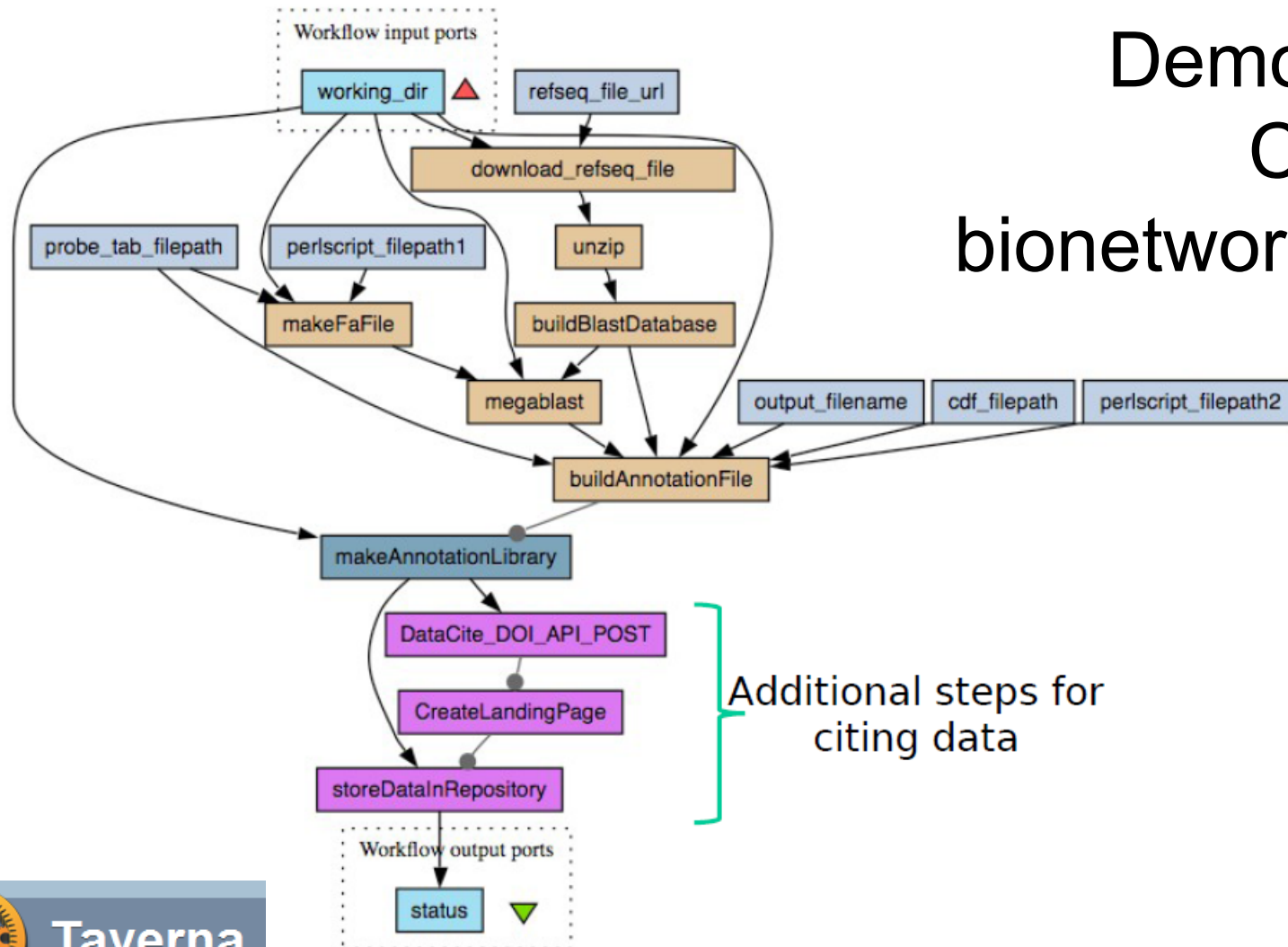
Citation Requirements

- 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

I cite others

I make my work citable

Others cite me

Costs, Benefits Value: KRDS
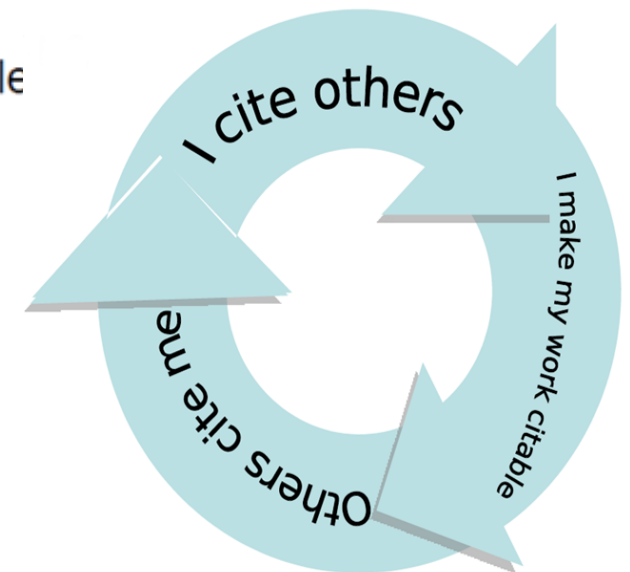
Keeping Research Data Safe Factsheet

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%

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)

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)

5-Yearly & Cumulative Cost for Refreshment in ADS — Refreshment Cost, Cumulative Cost
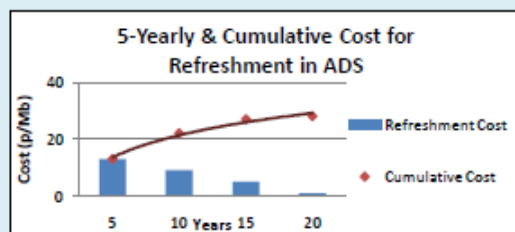
Costs for archival storage and preservation ("refreshment") decline to a minimal level over 20 years

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)

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)

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)

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

KRDS Activity Model Benefits & Metrics
Use Case 1 : National Crystallography Service
Use Case 2 : Researcher in the lab

- KRDS/I2S2 Project
- Extending the Benefits Framework
- Developing Value Chain and Impact Analysis tool
- Applied to different domains
- Toolkit now available

http://beagrie.com/krds-i2s2.php

Infrastructure for Integration in Structural Sciences

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
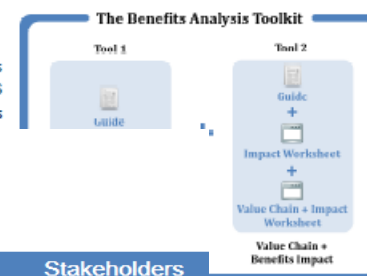
KRDS Value-Chain and Benefits Impact Worksheet

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

7th International Digital Curation Conference Dec 5-7, Bristol

DCC

because good research needs good data