

04/06/2015

## Research data management in UK universities: a collaborative venture

Rachel Bruce, Deputy Chief Innovation Officer, Jisc

- » What is Jisc ?
- » Growing pressure of data
- » The promise of open data
- » Where the UK is on this journey
- » What is the role of the library ( & others)

## Vision

To make the **UK** the most **digitally advanced education** and **research** nation **in the world**

## Mission

To enable people in **higher education, further education** and **skills** in the UK to perform at the **forefront of international practice** by exploiting fully the possibilities of **modern digital empowerment, content** and **connectivity**

# Four key pillars to Jisc activities

---



**Advice & Engagement**



Digital **Resources**

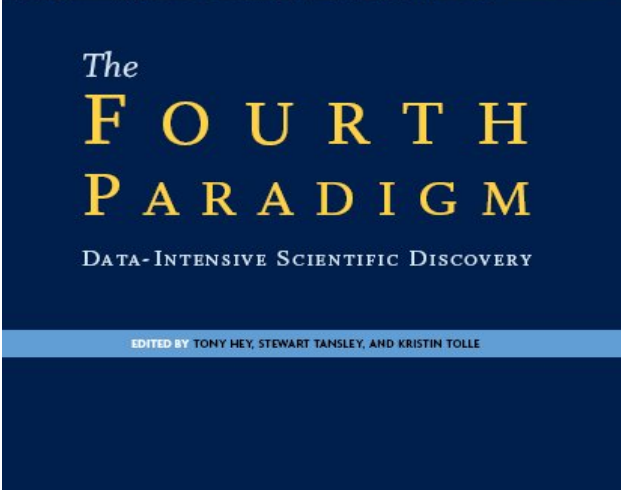


**Research & Development**



**Network & Technology**





## Science Paradigms

- Thousand years ago:  
science was **empirical**  
*describing natural phenomena*
- Last few hundred years:  
**theoretical** branch  
*using models, generalizations*
- Last few decades:  
a **computational** branch  
*simulating complex phenomena*
- Today: **data exploration** (eScience)  
*unify theory, experiment, and simulation*
  - Data captured by instruments or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files using data management and statistics

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G \rho}{3} - \frac{K c^2}{a^2}$$



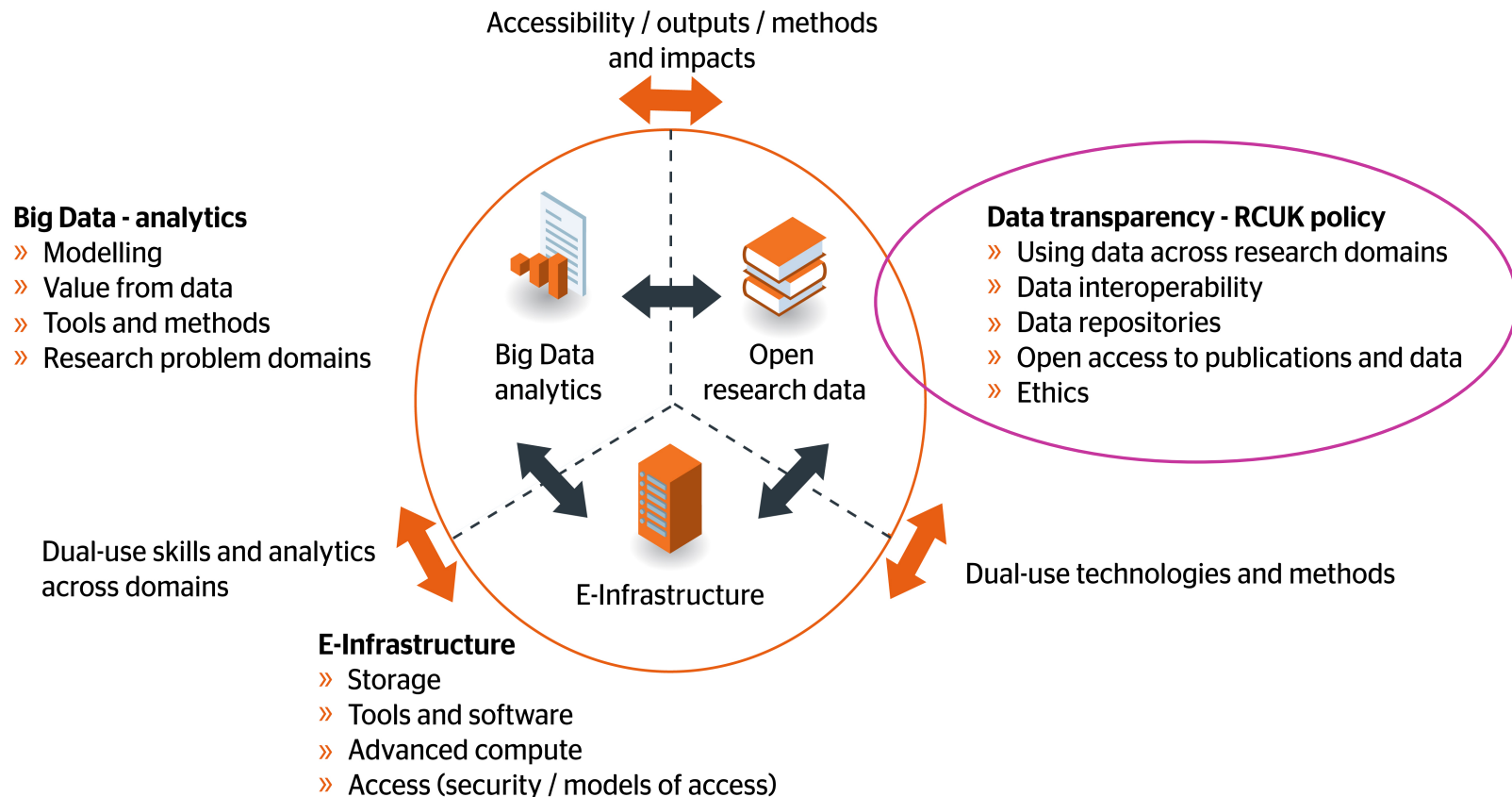
[research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_complete\\_lr.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf)

- » Develop an international framework for a Collaborative Data Infrastructure
- » Earmark additional funds for scientific e-infrastructure
- » Develop and use new ways to measure data value, and reward those who contribute it
- » Train a new generation of data scientists, and broaden public understanding
- » Create incentives for green technologies in the data infrastructure
- » Establish a high-level, inter-ministerial group on a global level to plan for data infrastructure



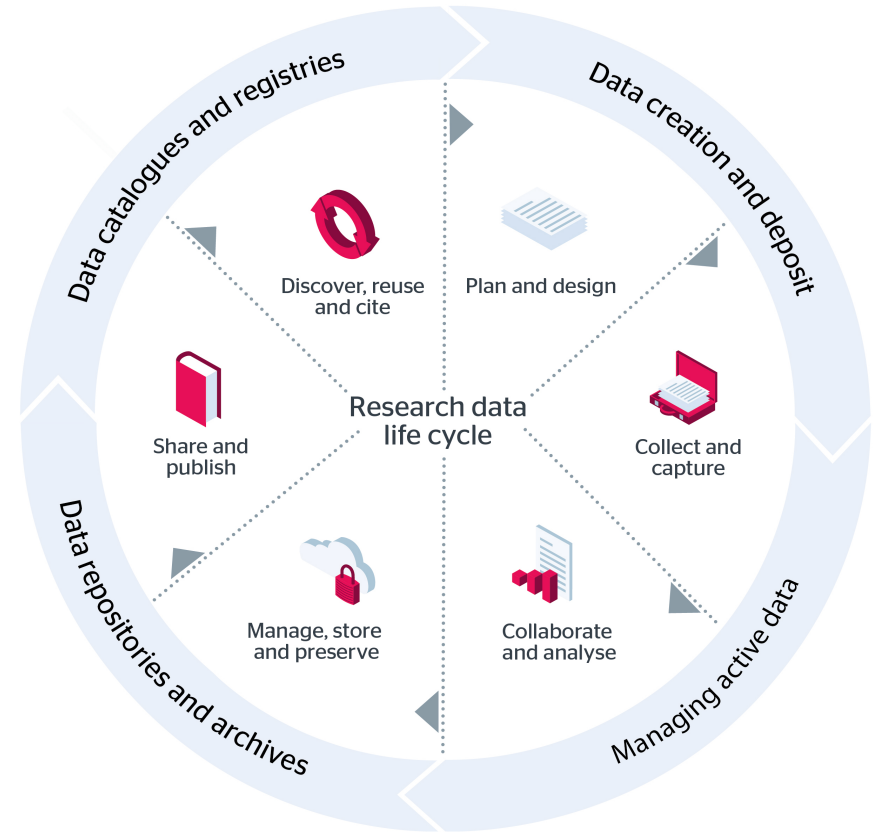
- » 'how the conduct and communication of science needs to adapt to this new era of information technology'.
- » **'As a first step towards this intelligent openness, data that underpin a journal article should be made concurrently available in an accessible database.** We are now on the brink of an achievable aim: for all science literature to be online, for all of the data to be online and for the two to be interoperable.'
- » Royal Society June 2012, Science as an Open Enterprise,  
[royalsociety.org/policy/projects/science-public-enterprise/report/](http://royalsociety.org/policy/projects/science-public-enterprise/report/)





- » “Research data is defined as **recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings**; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.” (Epsrc)
- » “Research data’ refers to **information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation**....examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.” (H2020)

- » Research Integrity and transparency
- » Re-use, new research and innovation



## RCUK Common Principles on Data Policy



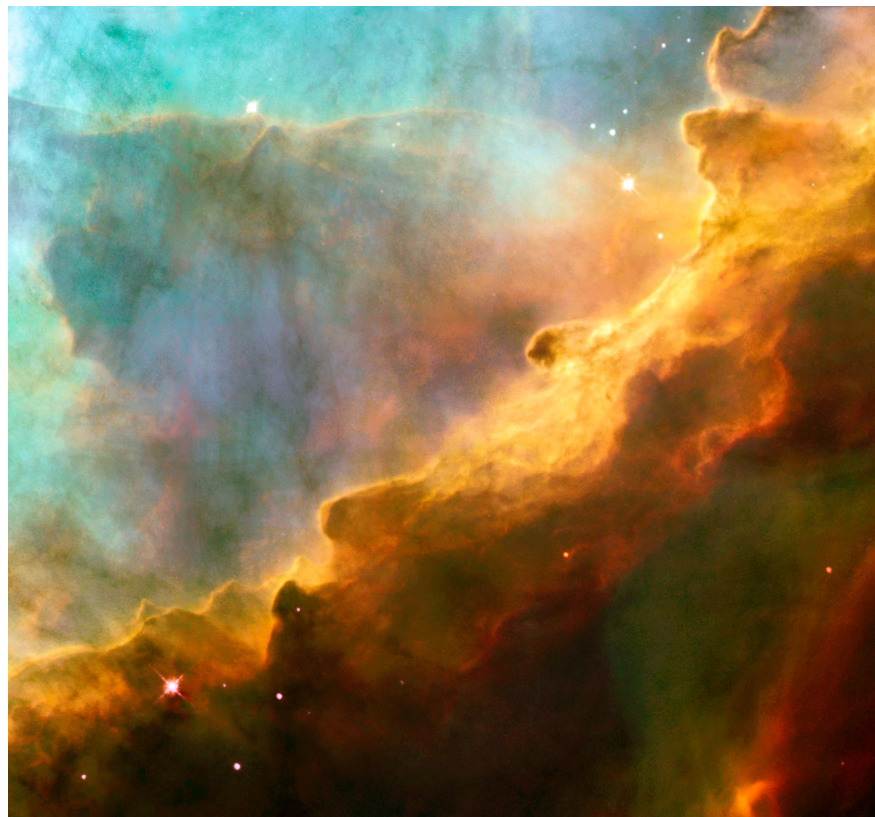
- » **Public good:** Publicly funded research data are produced in the public interest should be made openly available with as few restrictions as possible
- » **Planning for preservation:** Institutional and project specific data management policies and plans needed to ensure valued data remains usable
- » **Discovery:** Metadata should be available and discoverable; Published results should indicate how to access supporting data
- » **Confidentiality:** Research organisation policies and practices to ensure legal, ethical and commercial constraints assessed; research process should not be damaged by inappropriate release
- » **First use:** Provision for a period of exclusive use, to enable research teams to publish results
- » **Recognition:** Data users should acknowledge data sources and terms & conditions of access
- » **Public funding:** Use of public funds for RDM infrastructure is appropriate and must be efficient and cost-effective

[rcuk.ac.uk/research/datapolicy/](http://rcuk.ac.uk/research/datapolicy/)



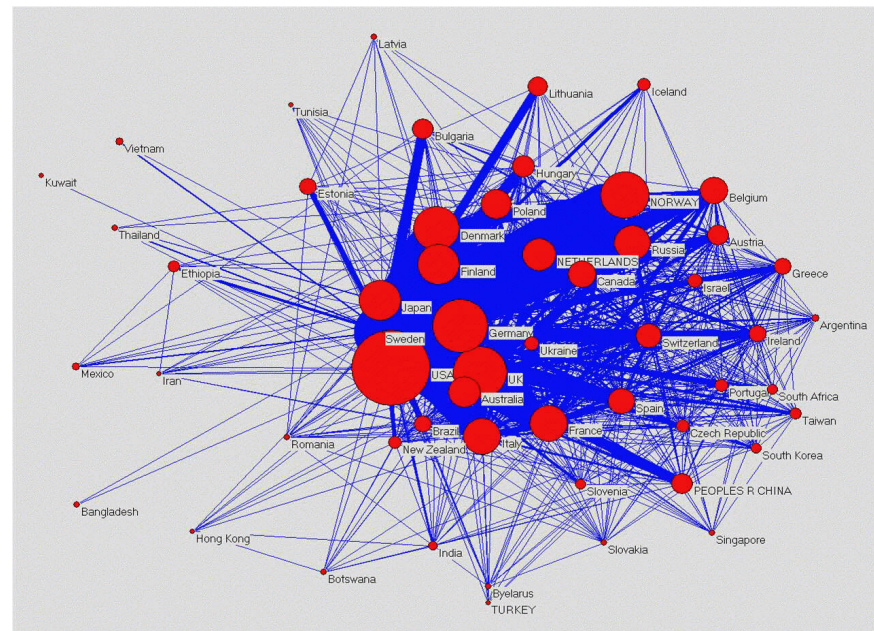
» The number of science papers written based on Hubble archival data ...has eclipsed the number of papers from new observations

[NASA](#) [ESA](#) and J. Hester (ASU)






» Microarray clinical trial publications examination of citation – 48% with publicly available data received 85% of the aggregate citations.



» The data centre under the desk (or in a back pack) is not adequate.



**CASH REWARD**  
for returning my lost backpack



- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is VERY important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: 07804430054 (ar456@cam.ac.uk)

Thank you!!

**Bio-scientist**

"why would I put my data in a repository? I share it informally with my peers, no-one else would understand it."

**Engineer**

"I have lots of data but you need a licence to this bespoke software to use it"

**Social Scientist**

"I have notes, photos & video and audio of subjects it would take way too long to anonymise it "

**Philosopher-**

"I don't have data, I annotate books"

**Informatics researcher**

"yes I will share my data but people should register; & why change to the Open Data Commons licence I have a bespoke licence I have always used"

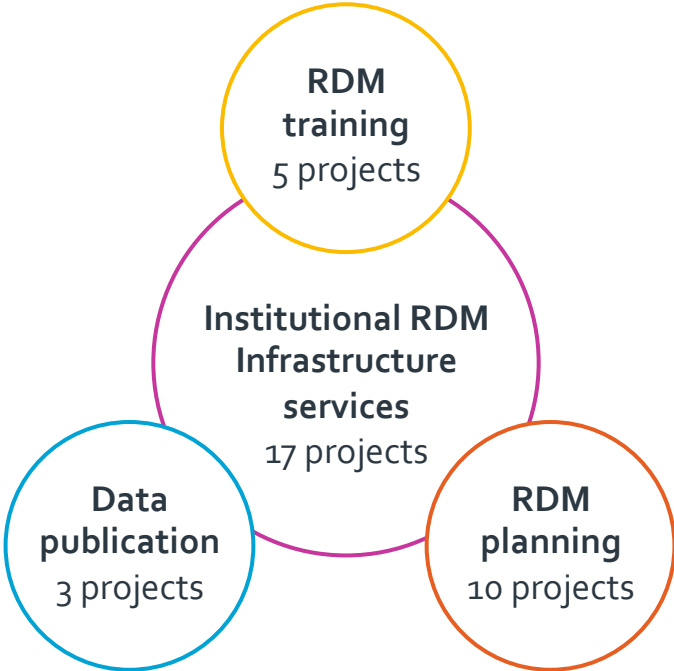
## Jisc first MRD Programme, 2009-2011 (but also 2004 DCC)

### 5 Strands

1. Research Data Management Infrastructure (RDMI) Projects
2. Research Data Management Planning (RDMP) Projects
3. Support and Tools Projects
4. Citing, Linking, Integrating and Publishing Research Data (CLIP) Projects
5. Research Data Management Training Materials Projects

## Jisc second MRD Programme, 2011-2013

- » Encouraged to reuse outputs from first programme and elsewhere.
- » Mix of pilot projects and embedding projects.
- » Holistic institutional approach to RDM.



- » **Ownership:** High level ownership of the problem.
- » **Sustainability:** Develop business cases to sustain work.



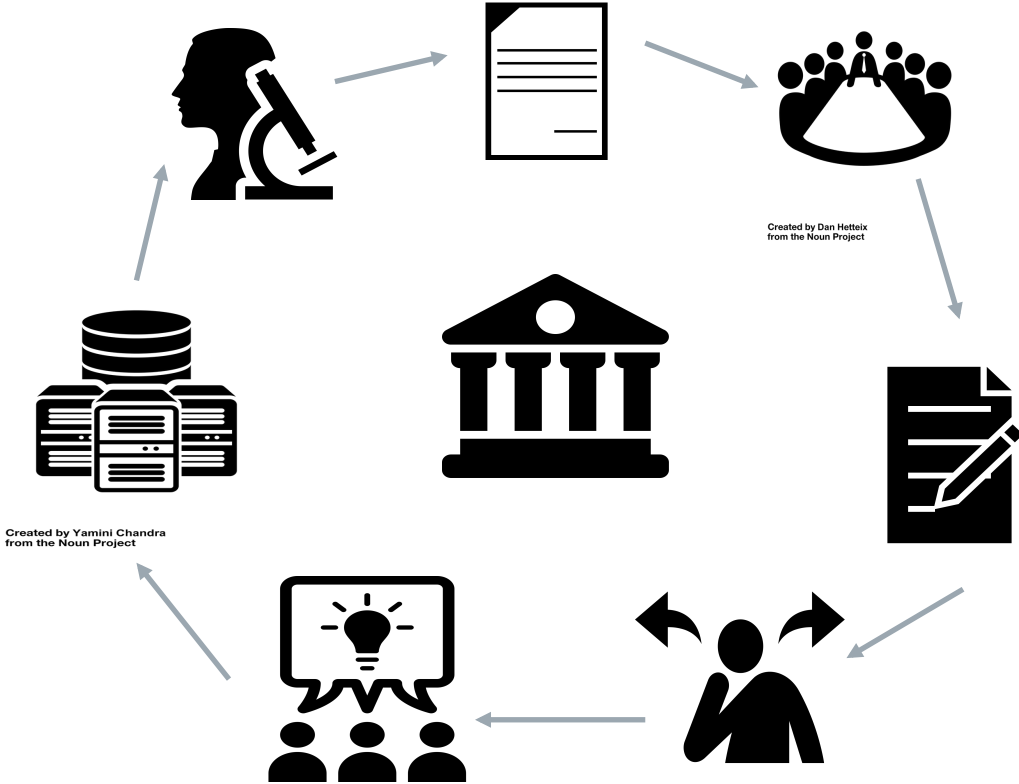
MANTRA is a free online course designed for researchers or others who manage digital data as part of a research project.



UNIVERSITY OF LEEDS



University  
of  
St Andrews



- » Module and course profiles
- » Level
- » Themes

Content

People

- » PGR Directors
- » Programme and module leaders
- » Supervisors
- » PGR & ECR researchers

- » Workshops
- » Labs
- » Online

Mode

Time

- » Point of need
- » Progression pathway
- » Modular



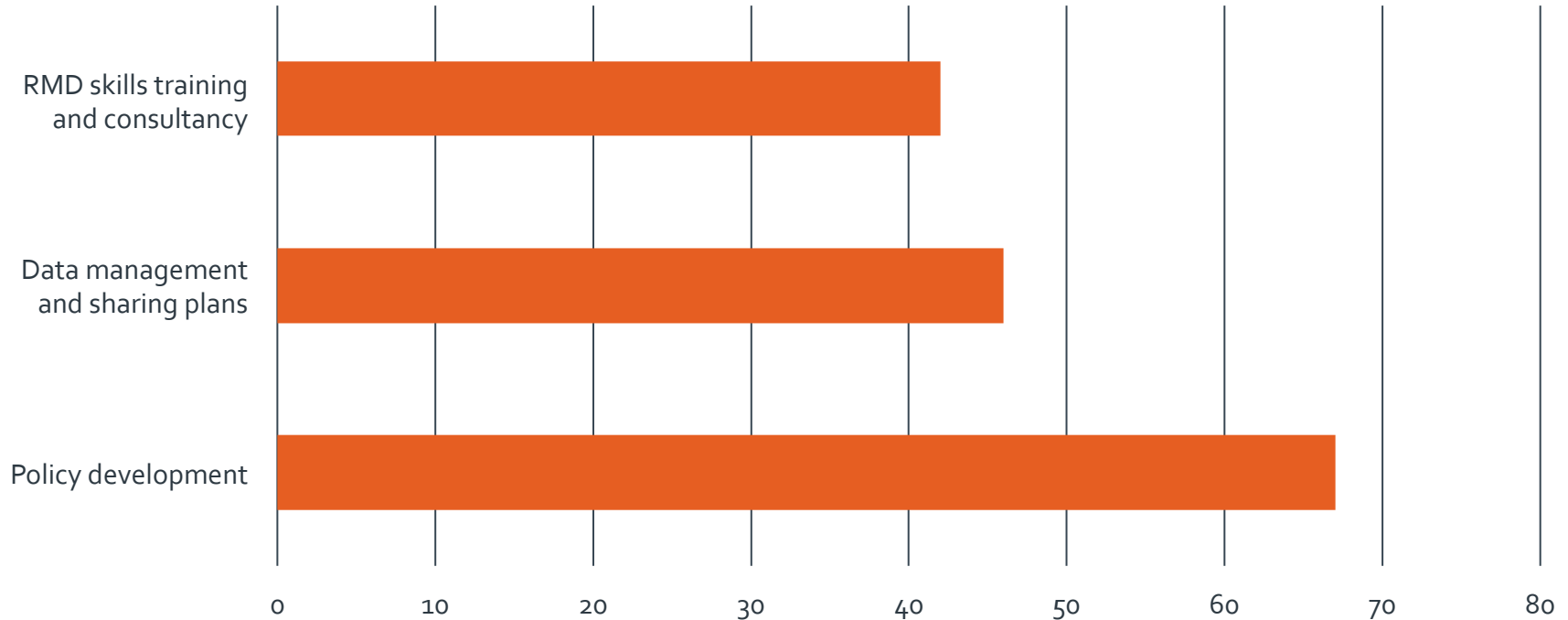
■ Russell Group (39)

■ Others 10%+ (35)

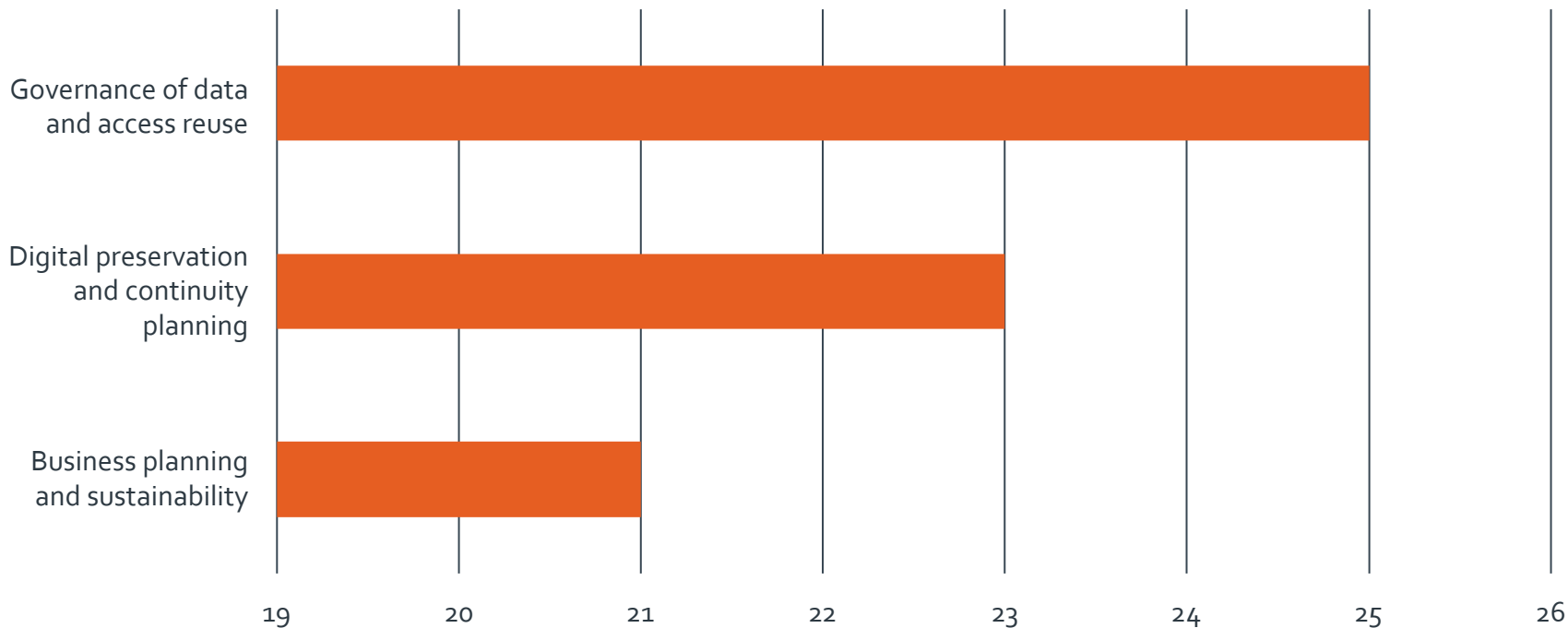
■ Others (13)



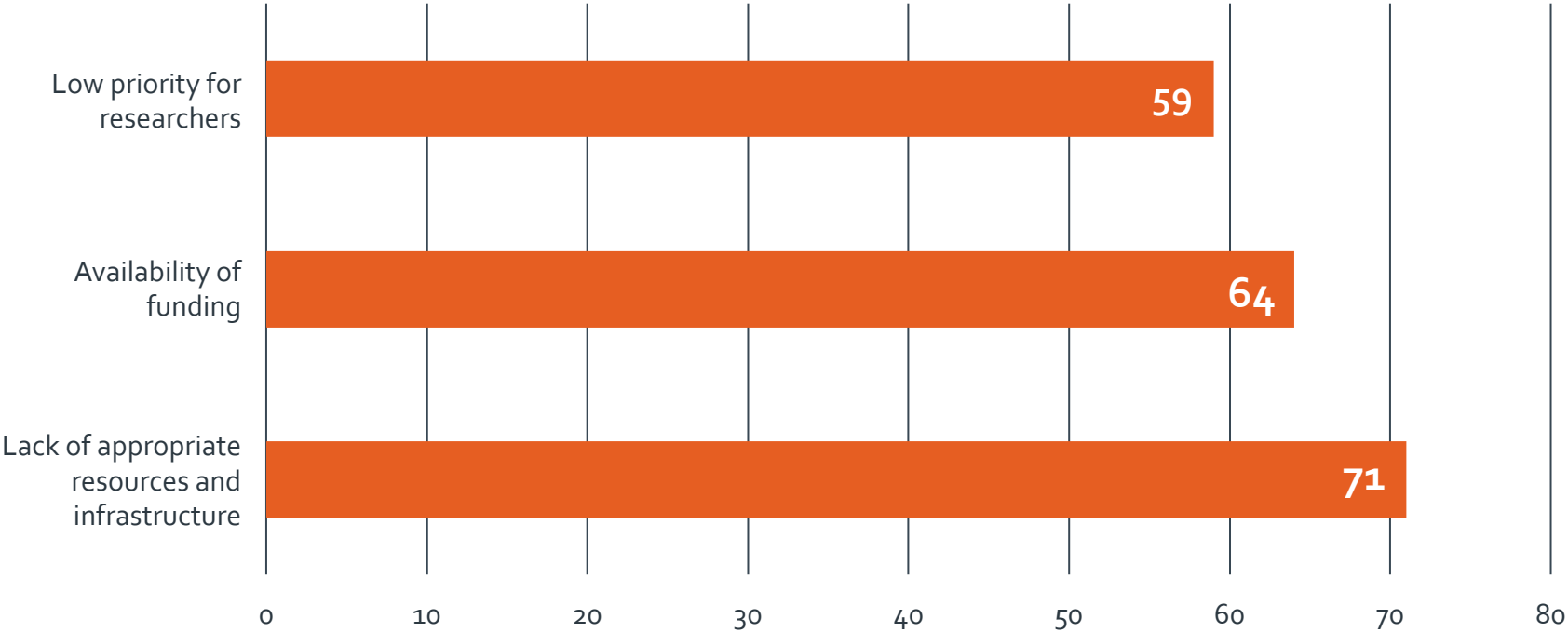
## % indicating piloting or live



## % indicating piloting or live



% citing



# Research at risk

“ This must be sorted and pulled together. Jisc can offer leadership here. ”

John Shemilt - Director of ICT, Imperial College, RUGIT



Realising a robust and sustainable research data management infrastructure and services to enrich UK research.

## Challenge

There is currently an inefficient and fragmented approach to research data management throughout the UK. This leads to underperformance and cost inefficiencies across universities. There is a real risk to universities in failing to comply with government and funder requirements.

## Who it affects

- >> Researchers need to comply with government and funder requirements
- >> IT directors need to store and manage research data
- >> Librarians need to ensure other researchers can access the data

## Why it matters

- >> Ability to attract and retain researchers
- >> Reputation of universities can be affected by mismanagement of research data
- >> Support better research collaboration (internationally)
- >> Comply with requirements of research funders

## Success criteria

- >> Cost effective national brokered infrastructure as a service
- >> Research outputs discoverable and reusable
- >> Fewer impediments to doing research
- >> Research Data Management is business as usual

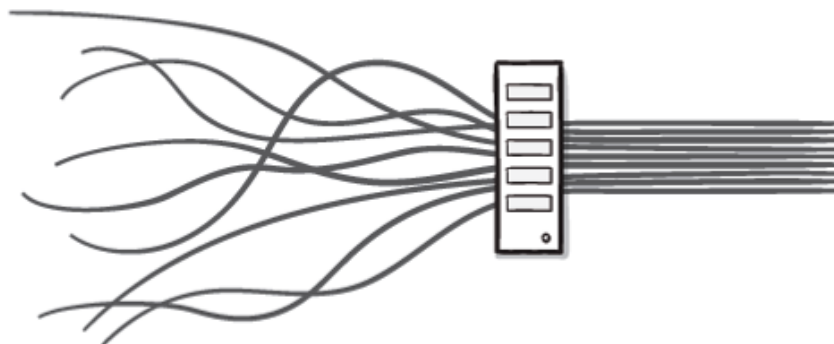


Illustration by Lisa Webb Studio

## Jisc audience

Which members of Jisc's audience will be most interested?



Vice Chancellors



College Principals



Finance Directors



Librarians



IT Directors

## Timescale

How long until we can expect to see the benefits?



1 YEAR

Quick wins, requirements and plan



2 YEARS

## Commitment

Who in the steering group pledged to commit two days of their time?



• Mike Fraser - Director of Infrastructure Services, University of

## Strategic framework

How does the challenge fit with Jisc's strategic framework?



## Safe share:

- » Encrypted VPN infrastructure between organisations; enhanced confidentiality and integrity per ISO27001
- » Requirement to move electronic health data securely and support research collaboration
- » Working with biomedical researchers at Farr Institute, MRC Medical Bioinformatics initiative, ESRC Administrative Data Centres



## File sync and share:

- » Dynamic purchasing system
- » Offering file sharing products typified by vendors such as Dropbox
- » Enhanced products available offering:
  - › EEA (European Economic Area) storage - for Data Protection compliance
  - › user managed encryption - encryption keys are managed by the customer
  - › integrated federated access - end users can access services using institutional issued credentials
- » Box; Microsoft; Q Associates; Capito currently on the system

## Data archiving:

- » A single supplier framework with Arkivum
- » secure, easy-to-use and cost-effective data archiving service for research and education
  - › 100% data integrity guarantee
  - › 2 UK stored data copies accessible online
  - › 1 UK stored copy held with 3rd party ESCROW data holding company.
  - › £5m - £100m professional indemnity insurance
  - › ISO 27001 compliance
  - › 10 year framework to December 2023

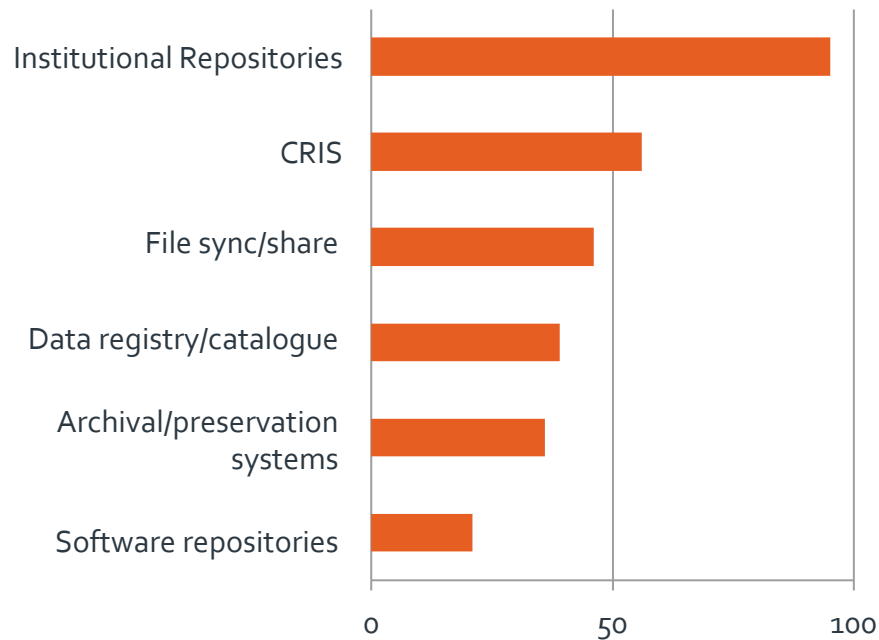
## A gap?

- » Many different research domain and institution approaches to storage provision
- » Is there a requirement for national scale storage facilities and services?
- » e.g A national large scale, on-net, 'on-line / near on-line', possibly complementing close-coupled on-site storage, preservation?
- » To complement the focus on data standards, interoperability and discoverability

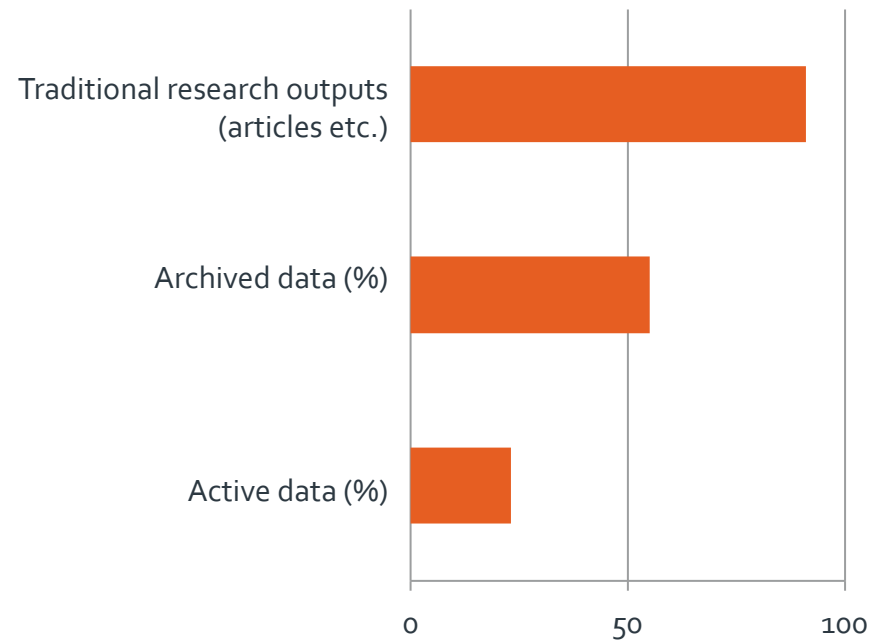




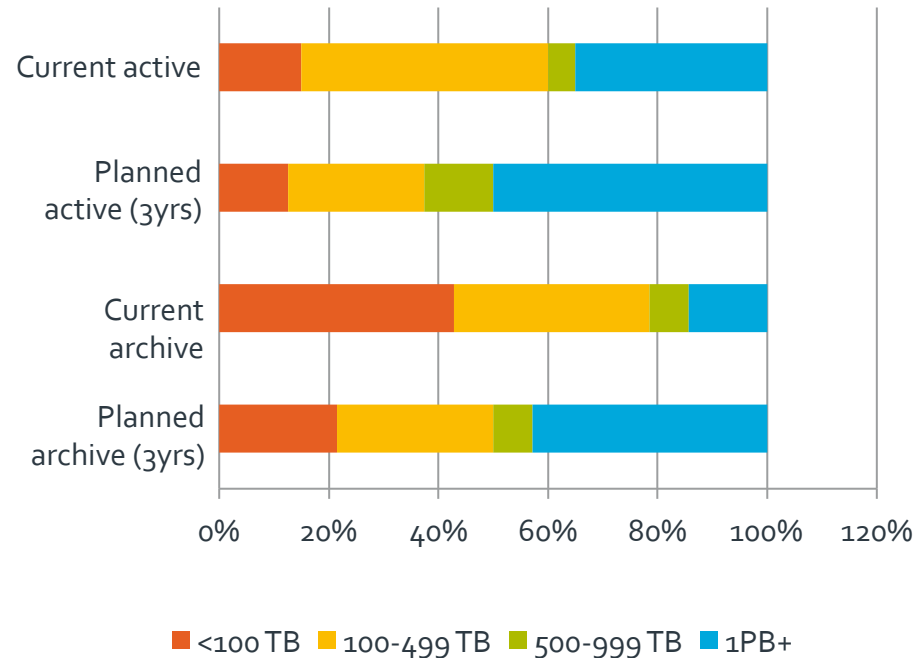
## Research Systems Provided



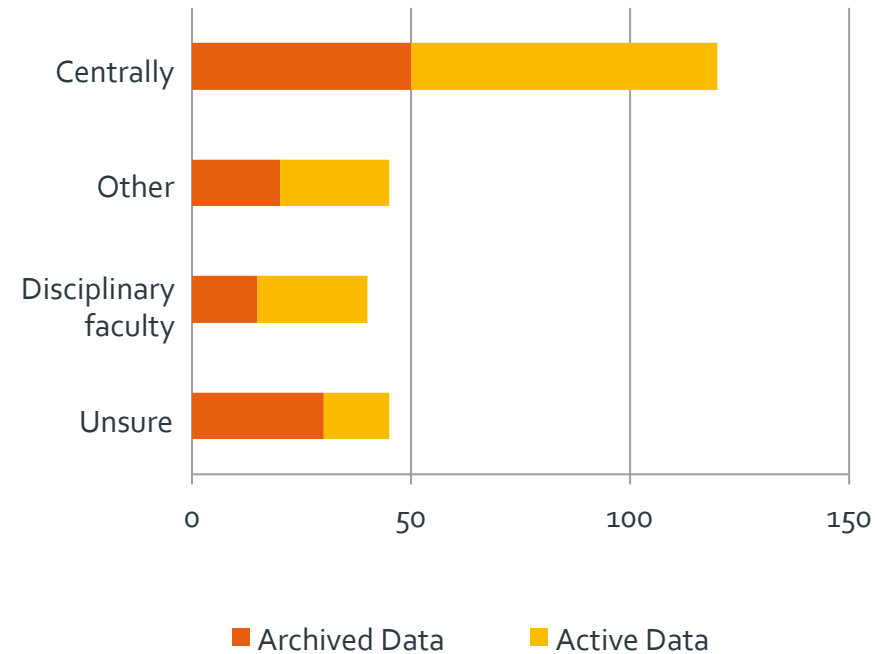
## Outputs Registered in CRIS

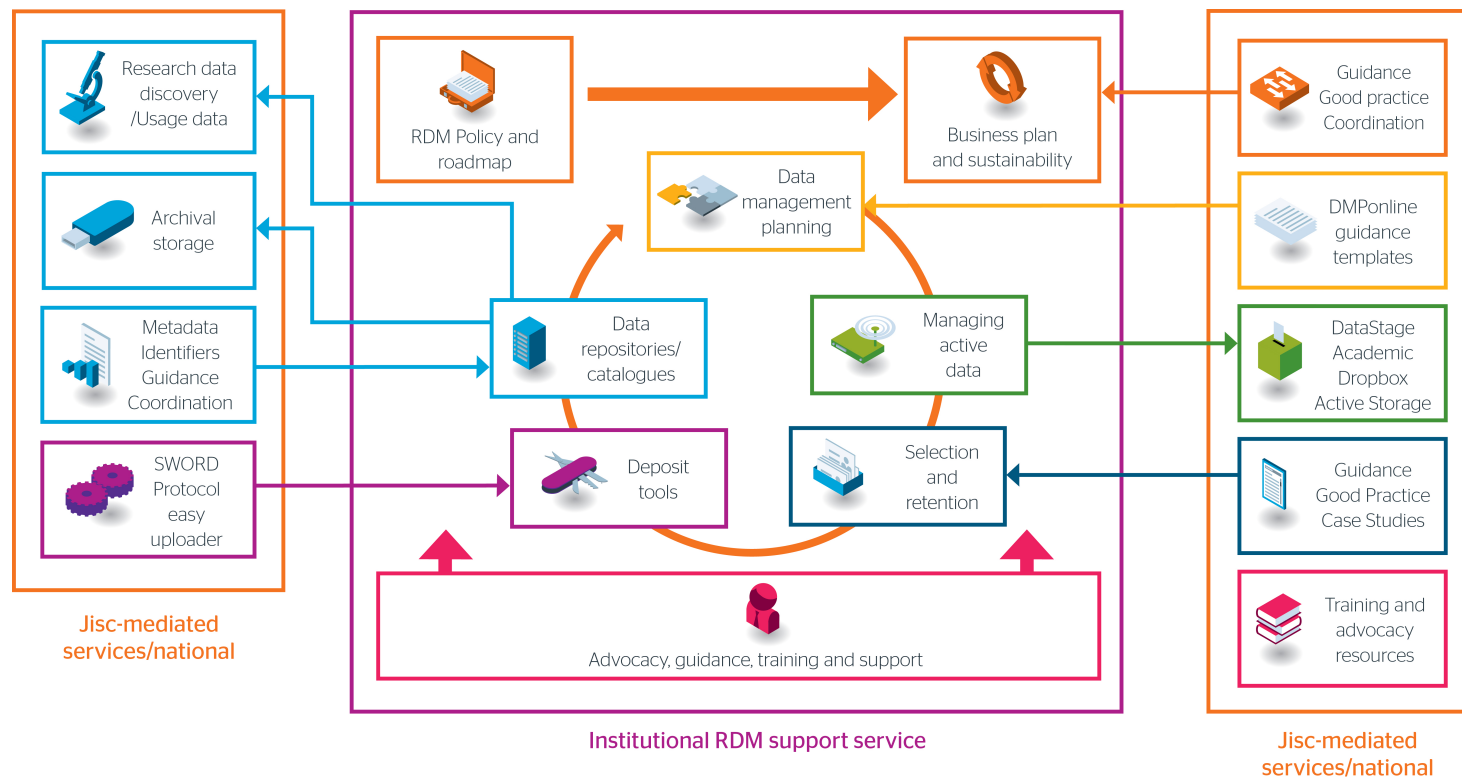


## RD storage capacity



## RD storage provision



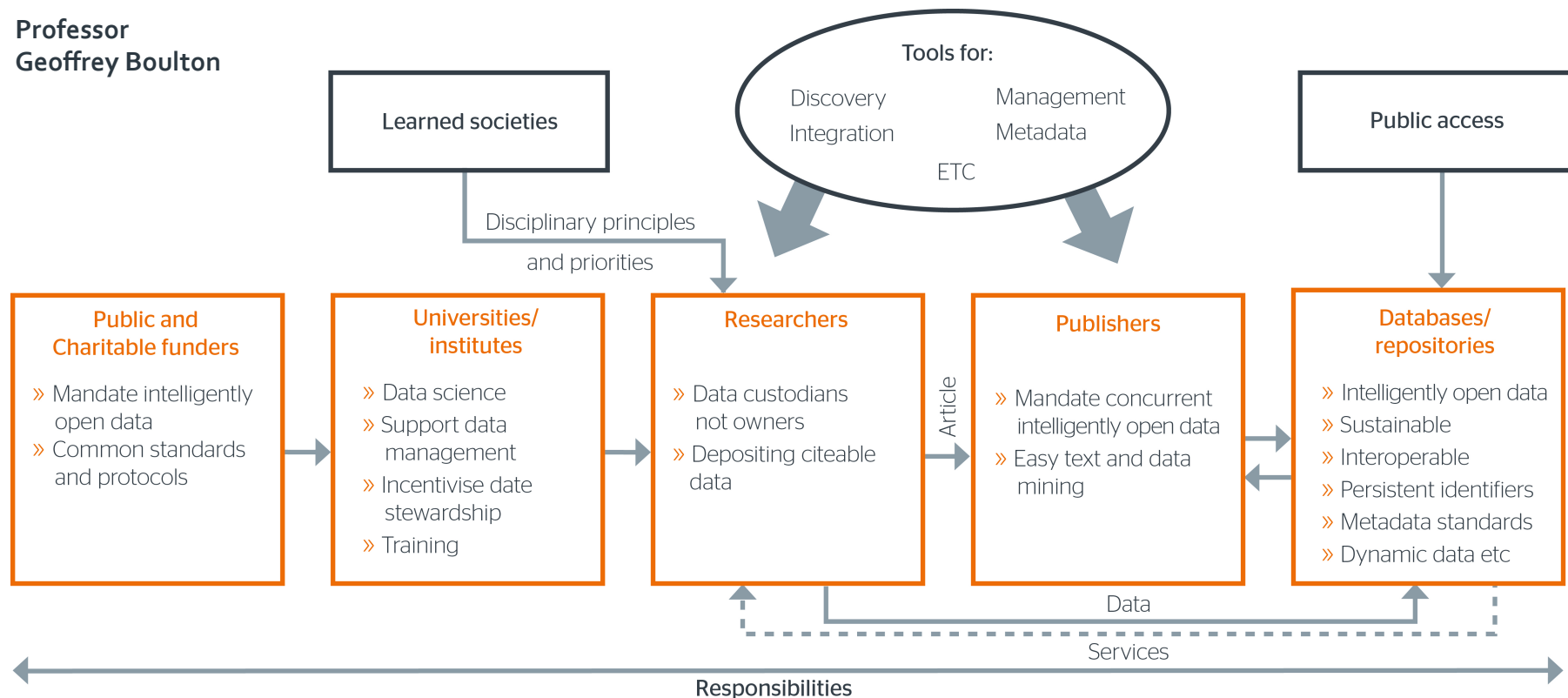


## Institutional coordination and partnerships

1. Requirements
2. Planning
3. Informatics
4. Citation
5. Training
6. Licensing
7. Appraisal
8. Storage
9. Access
10. Impact

Liz Lyon, 'The Informatics Transform: Re-Engineering Libraries for the Data Decade', International Journal of Digital Curation (2012), 7(1), 126–138; [dx.doi.org/10.2218/ijdc.v7i1.220](https://doi.org/10.2218/ijdc.v7i1.220)

Professor  
Geoffrey Boulton



## “Science is broken” Examples:

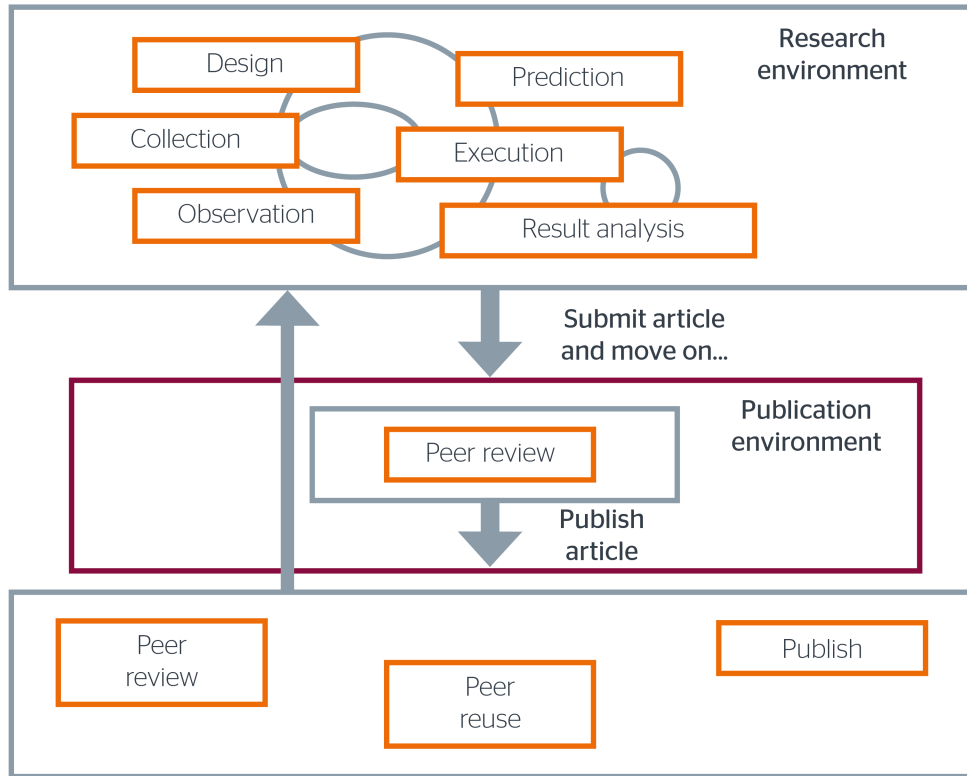


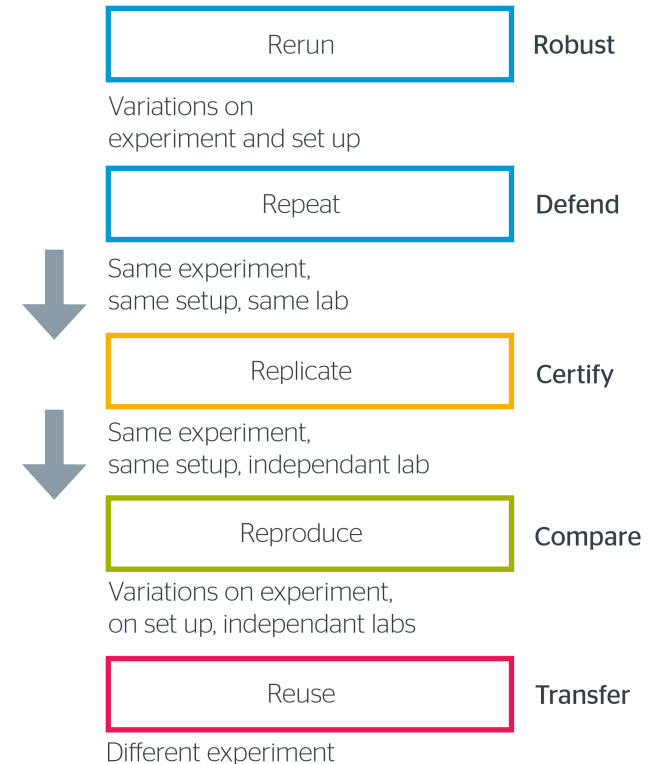
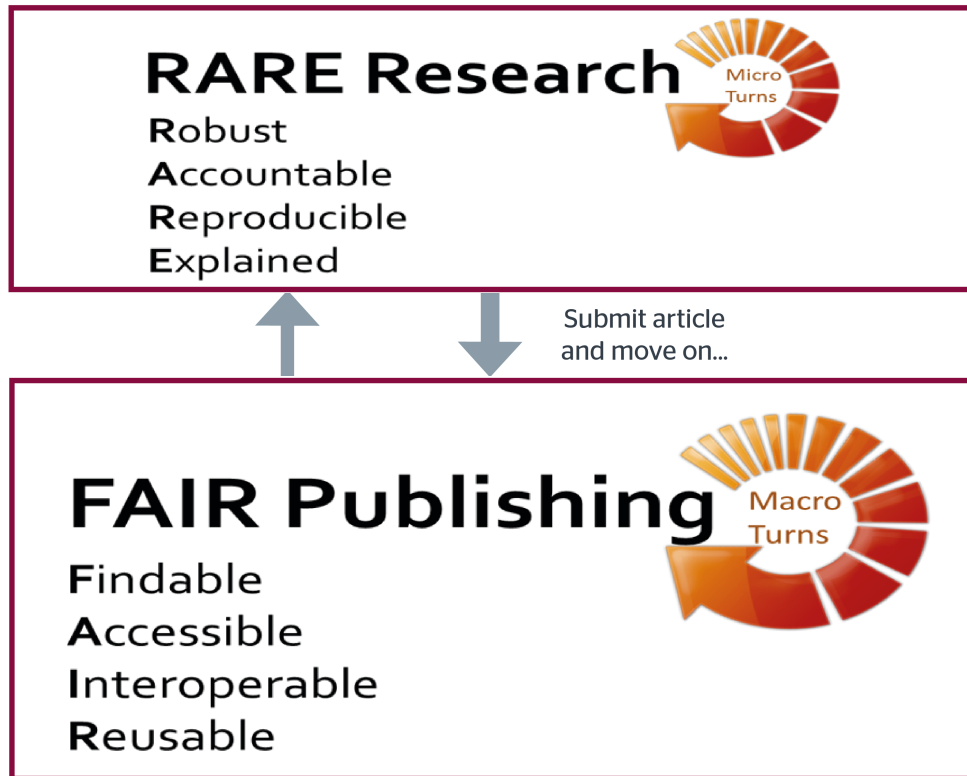
- » psychology academics making up data,
- » anaesthesiologist Yoshitaka Fujii with 172 faked articles
- » Nature - rise in biomedical retraction rates overtakes rise in published papers



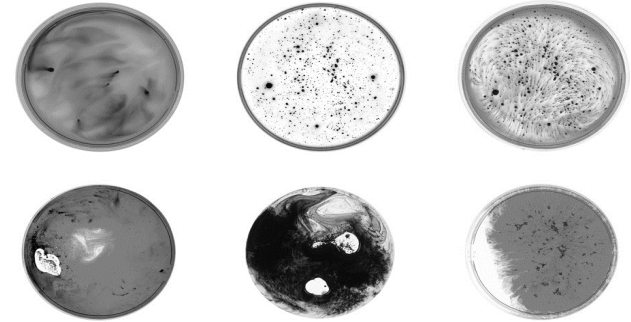
“economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors. Another surprise is that the mistakes, by two eminent Harvard professors, were spotted by a student doing his homework”

[bbc.co.uk/news/magazine-22223190](http://bbc.co.uk/news/magazine-22223190)









**Thank you!**

Rachel Bruce

Email: [email@jisc.ac.uk](mailto:email@jisc.ac.uk)

Twitter: [jisc.ac.uk](https://twitter.com/jisc.ac.uk)

Blog: [researchdata.jiscinvolve.org/wp/](http://researchdata.jiscinvolve.org/wp/)



Except where otherwise noted, this work is licensed under CC-BY-NC-ND