

The background of the slide is a landscape photograph. It features a calm body of water in the foreground, reflecting the sky and clouds. In the middle ground, there is a dense line of dark evergreen trees. The sky is a deep blue, filled with large, bright white cumulus clouds. In the bottom foreground, there are some wooden logs or debris floating in the water, and the tip of a boat is visible on the right side.

# Research data on the horizon

Future-proofing Digital Preservation for Research Data : Chris Keene

- » Where are we now
  - » Research data shared service
  - » What might be coming in the future
  - » Discussion
-


# OPEN ACCESS

*may also  
contain  
data*

Comment | [OPEN](#) | Published: 15 March 2016

## The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons  - [Show fewer authors](#)

*Scientific Data* **3**, Article number: 160018 (2016) | [Download Citation](#) 

### To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

### To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

### To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

### To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards



EPSRC and RCUK have both set out clear expectations that Institutions must take responsibility for Research Data Management, Preservation and Sharing

<https://www.epsrc.ac.uk/about/standards/researchdata/expectations/>



A close-up photograph of a man and a woman looking at a laptop screen. The man is on the left, wearing a dark grey button-down shirt, and the woman is on the right, wearing a brown knit sweater. Both are smiling and looking down at the screen. The background is blurred, suggesting an office or library setting.

Jisc

# Research Data Shared Service

A brief introduction

## Filling a gap

75% of respondents look first to their institution to preserve their data



## Uptake of RDM

Only 40% of respondents have a Research Data Management plan



## Advocacy

Only 16% of respondents are currently accessing university RDM support services



## Metadata

Only 18% of respondents say they follow established metadata guidelines



## Public datasets

>70% recognise that research is a public good and should be publicly released



## Sensitive data

41% of respondents have some form of sensitive data



Following input from our Expert Advisory Group, the Research Data Network, funders, and dialogue with global users and vendors, Jisc RDSS will provide the following core researcher functional needs:

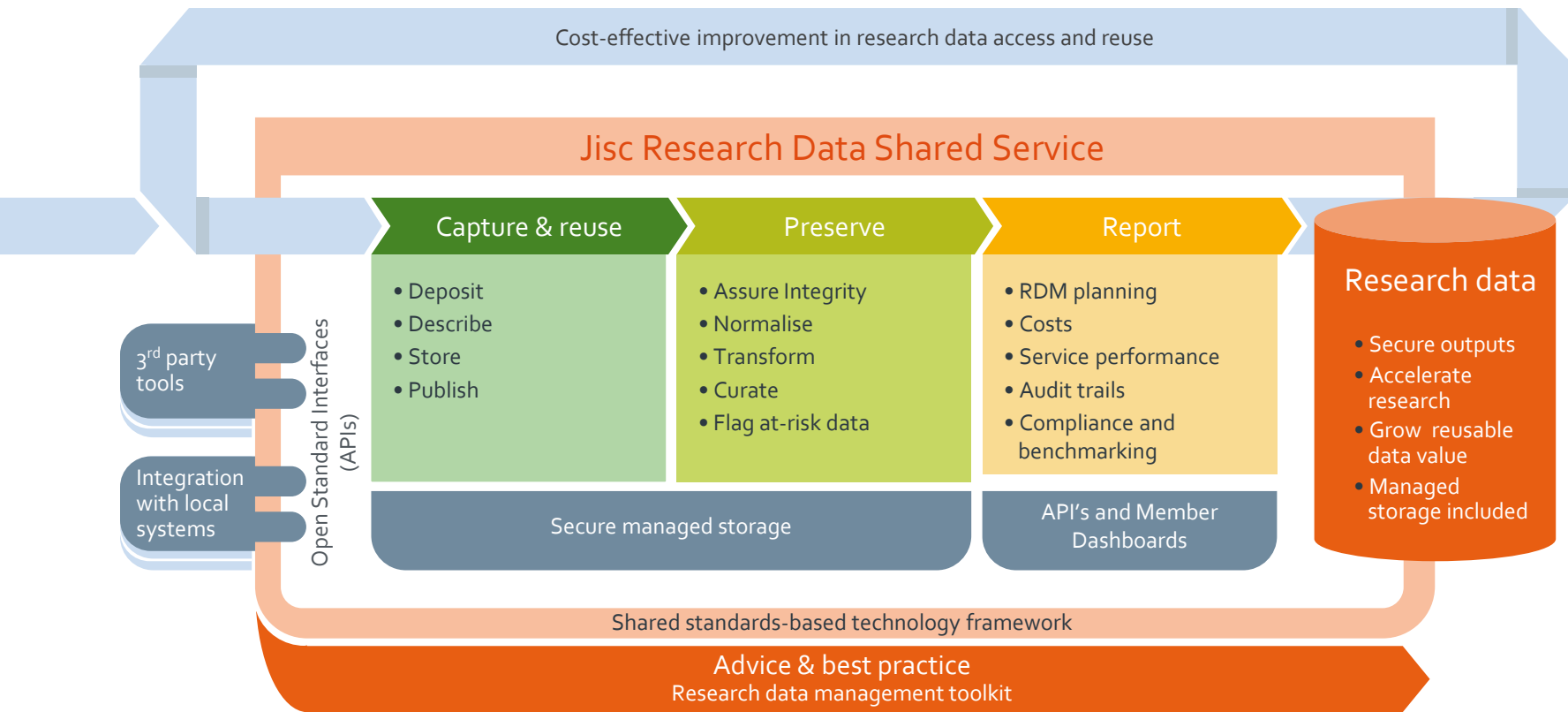
Advise &  
best practise

Capture & reuse

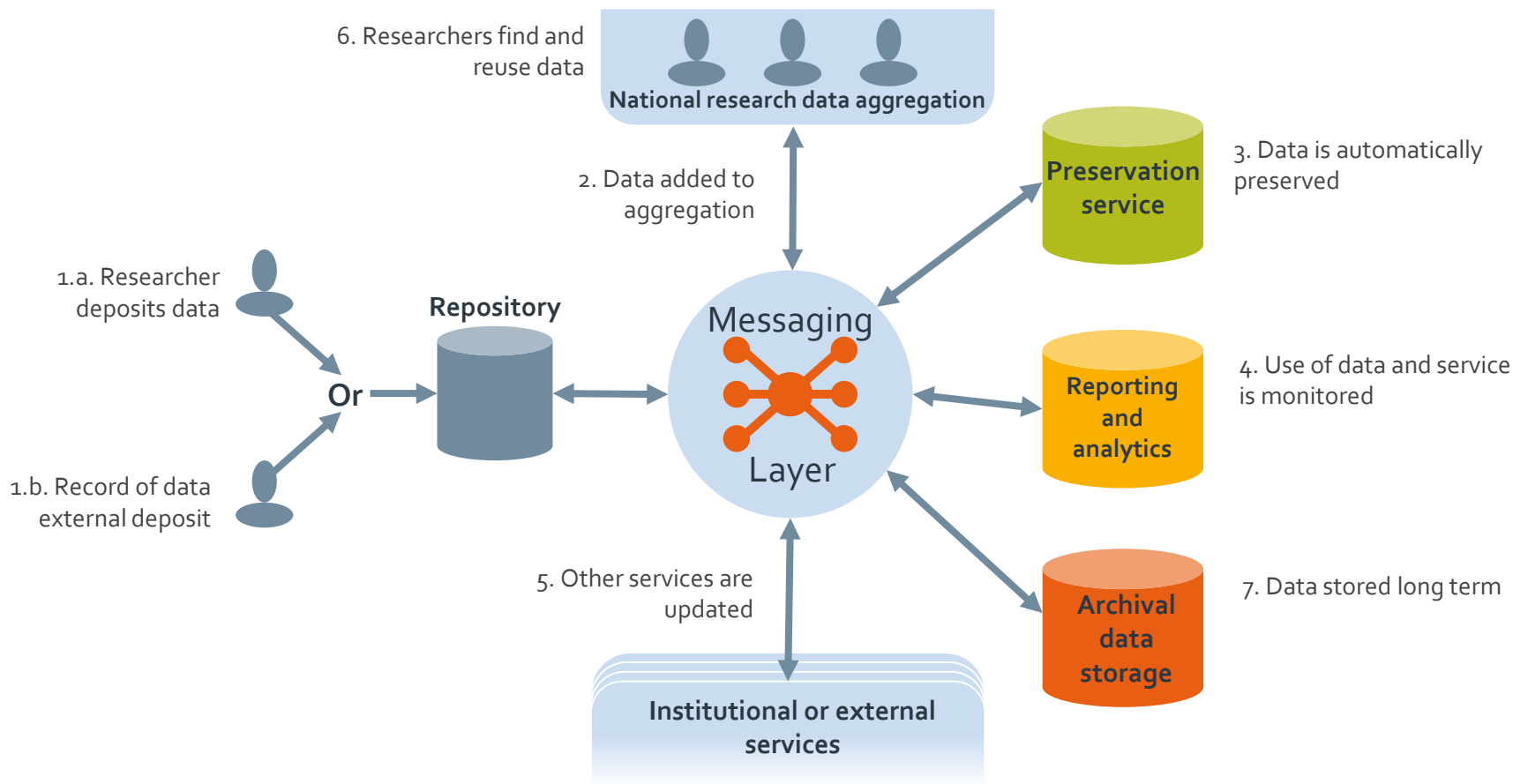
Preserve

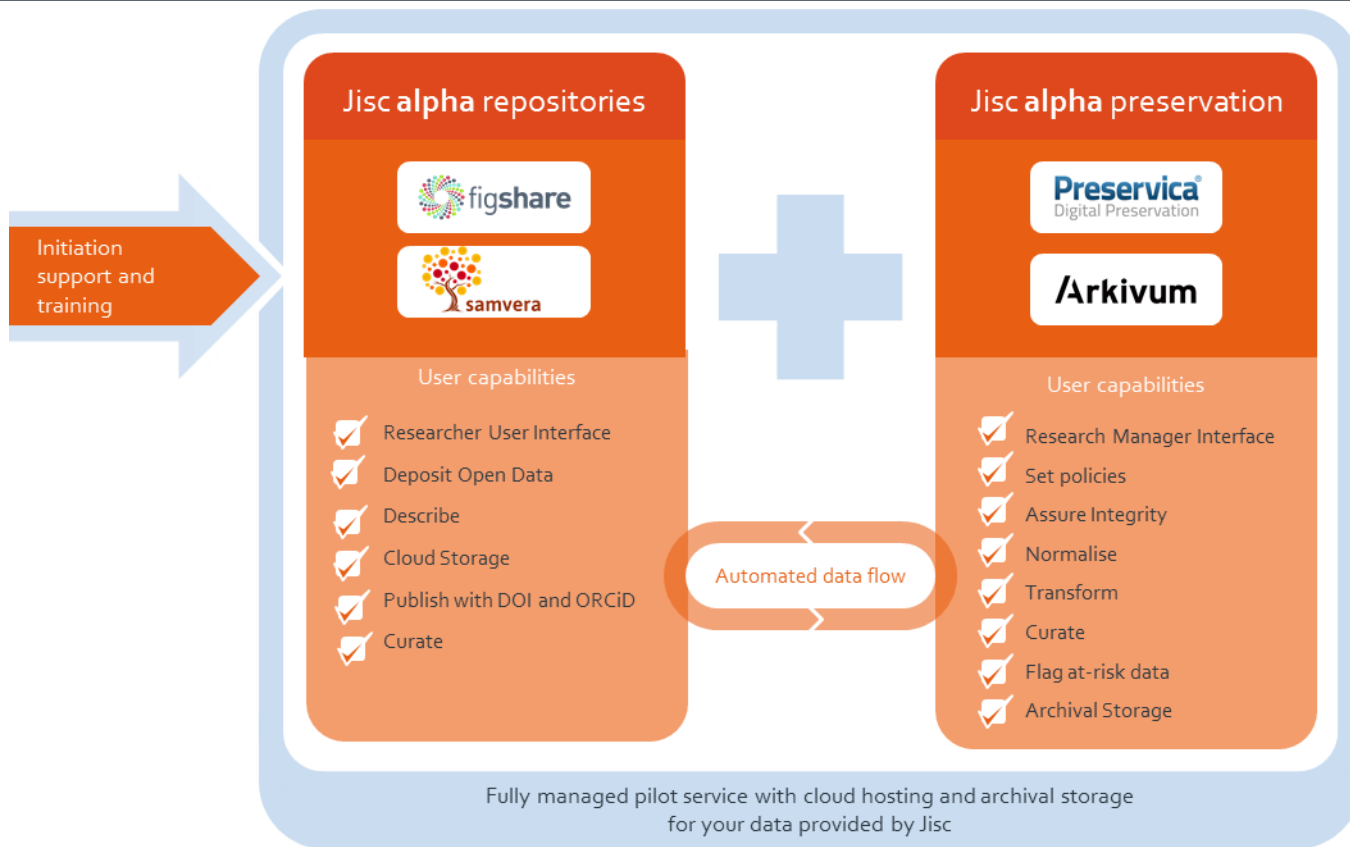
Report

Source: Jisc DAF Survey results 2016









End-to-end  
service



Repository  
service



Preservation  
service

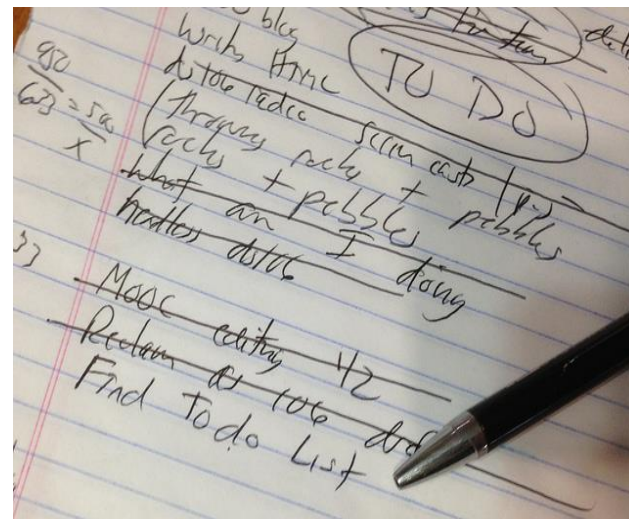


Service to be launched in  
Autumn 2018

All 3 options include:

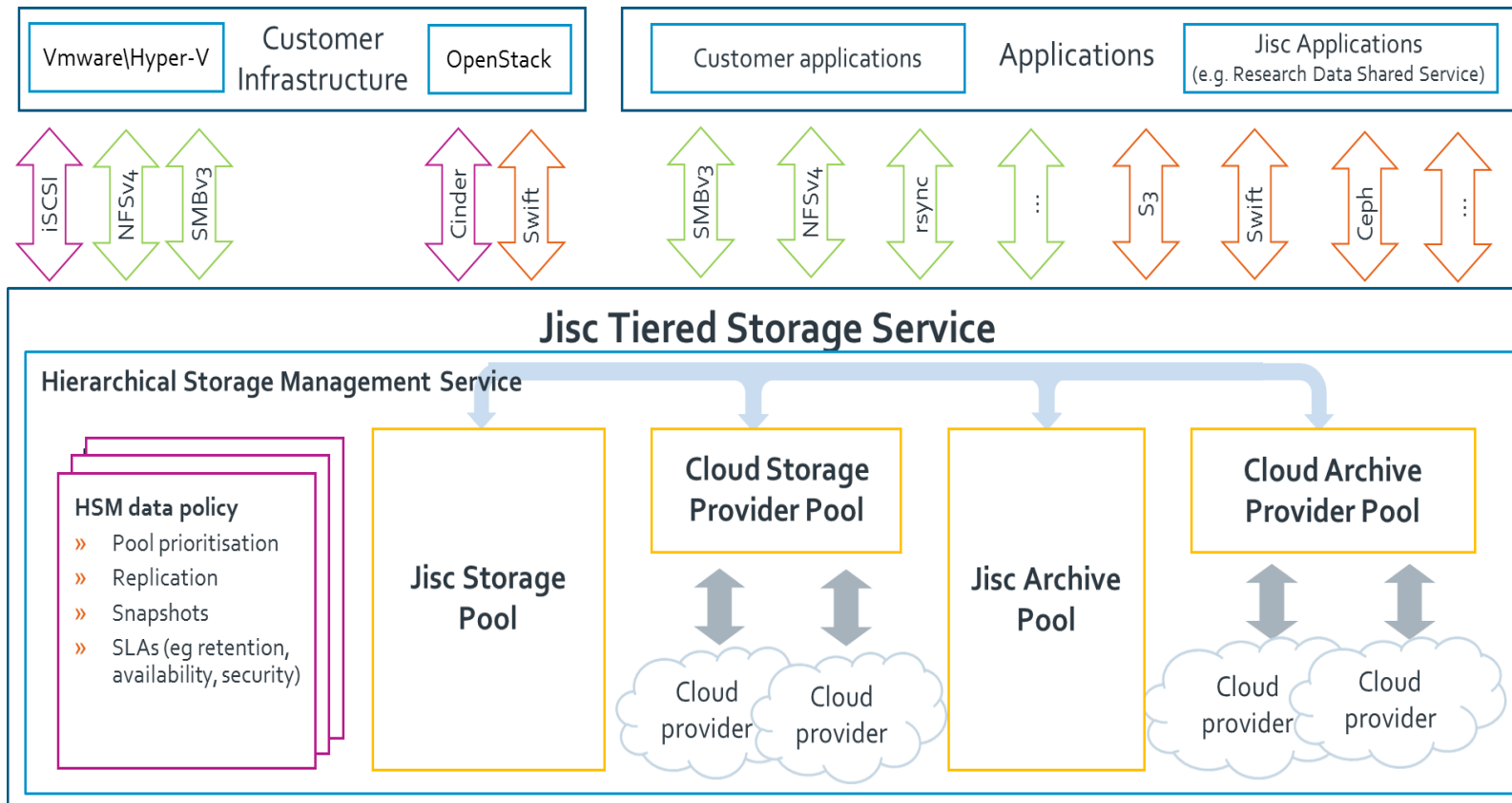
- ✓ Financial benefits
- ✓ Standards
- ✓ Advisory
- ✓ Network membership

- » First priority is research data
  - › Research output (Article/Thesis etc.)
  - › Research data
  - › Research software/code
  - › Provenance metadata (method)
- » But also.....
  - › Preservation systems tailored for multiple digital objects and data types
  - › Use cases and pilots for objects beyond research data



<https://creativecommons.org/licenses/by/2.0/>  
<https://www.flickr.com/photos/cogdog/>

# Under Development: Tiered Storage Service

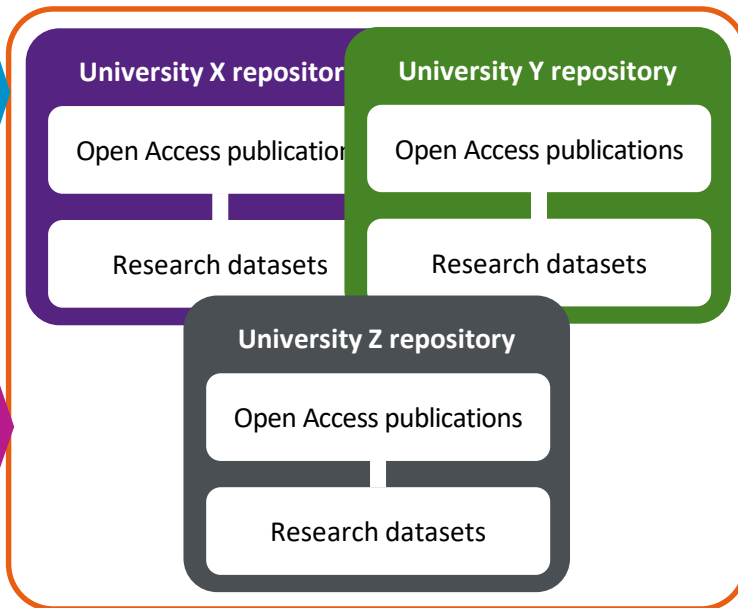


## University systems

- » (Single Sign-On, Finance, HR..)

## Information sources

- » Publications Router
- » Publishers
- » Crossref
- » ORCID
- » DataCite
- » PubMed
- » Sherpa policy tools



## Information destinations

- » Google etc.
- » Discovery services
- » Jisc CORE (global OA aggregation)
- » Jisc Monitor (compliance checking)
- » Jisc Collections
- » Funders systems
- » OpenAIRE + for EU

## The horizon – where are we going

---

**Data will redefine scholarly research**

---



- » Datasets
  - » Excel files
  - » Code/software
  - » Documents
  - » Music
  - » Audio
  - » Visual and images
  - » Evidence
  - » everything
-

# The researcher of the future...



## » Data miner

- » I want data mine relevant data. I want to find relevant datasets and mine appropriate information.
  - » Need to understand, in a machine readable way, the nature, subject, and format of the data. The data needs structure and needs to be discoverable.
  - » The data accessed in real time – rather than downloaded.
-

**» Data miner #2**

» I want to use sound recordings of speech from across the world, translated to English, and summarise discussions around my research area.



» I want to set up a bot to trawl around research datasets in my field (and beyond) to find new facts and discoveries.



## » **Big Very large data**

» The data I'm working with is too large to handle, I need a way to query and process at it's source, without needing to download the raw dataset.



## » Reproducing research

- » As I funder/reviewer/researcher I want to use the data, data (code), data (method), data (virtual machine) *'research artefacts'* and reproduce a number of experiments from a specific journal/researcher/subject area to verify their results.
- » As a researcher I want to use these research objects but modify the inputs
- » Links between data; research packages.





## » Distributed working

- » Our experiment involves people, equipment, data, labs and servers based across the world.
  - » Need to be able to access and utilise data in this distributed environment.
  - » Meeting various mandates and national laws
  - » European Open Science Cloud and other initiatives
-



How well do data repositories cater for these researchers and stakeholders?



## The burning bridge of the future

We will be holding back science if we do not take a leap forward.

We need to facilitate new ways of working to help improve research and lead to new discoveries

---

**the scholarly record of the future will be  
built on data**

