Jisc Research Data Shared Service (RDSS)

Meeting needs outside the research data community
Why a Shared Service?

There is no single “solution” easily available and that meets requirements for Universities to enable better management of research data.

research data network: http://researchdata.network
web: https://www.jisc.ac.uk/rd/projects/research-data-shared-service
github https://github.com/JiscRDSS
University Pilot Institutions

- bucks new university
- Middlesex University
- University of Cambridge
- University of Surrey
- University of St Andrews
- Cardiff University
- Plymouth University
- University for the Creative Arts
- The University of Winchester
- Imperial College
- Royal College of Music
- University of Glasgow
- University of York
- University of Lincoln
- Lancaster University
- St George's University of London
Preservation Partners

» Core Systems

Preservica
Digital Preservation

Arkivum

artefactual systems inc.

» Preservation Development

THE UNIVERSITY
of EDINBURGH

DATA VAULT

Open Preservation
Foundation

» Records and Archives Pilot

University of Glasgow

UNIVERSITY OF WESTMINSTER

» Information, Advice, Guidance and Community Support

Digital Preservation Coalition

Software Sustainability Institute
Institutional or external services

Repository

National research data aggregation

Researchers find and reuse data

Preservation service

Reporting and analytics

Use of data and service is monitored

Data added to aggregation

Data is automatically preserved

Researchers deposits data

Other services are updated
First priority is research data

However research data is not good enough on its own for research publication the ideal is to store, or link to the complete research package

- Research output (Article/Thesis etc.)
- Research data
- Research software/code
- Provenance metadata (method)

Without this research is often not re-usable or repeatable.

But.....we are aware that we are providing preservation systems that are tailored to deal with all kinds of digital objects and data.

- We are exploring use cases and pilots for objects beyond research data.

https://creativecommons.org/licenses/by/2.0/

https://www.flickr.com/photos/cogdog/
Research Data is often the hardest

Can we identify our research data?
We ran Droid* over the research data deposited with us over the past year.
Out of 3752 individual files:
- only 37% (1382) of the files were identified (with varying degrees of accuracy)
- there were 34 different identified file formats in the sample

* Droid is a free tool from The National Archives that can be used to automatically identify file formats

Identified research data files
Files identified by Droid (listed by file type)
...note that native files of the software in the previous graph of research data applications are not represented

Unidentified research data files
- Files not identified by Droid (listed by file ext)
- 107 different file extensions not identified
  - huge number with no extension (help!)
  - how do we solve the .dat file problem?

Count of file extensions of unidentified files
(includes extensions represented 10 or more times)
Automated preservation workflow:
- lack resources (staff, skills, budget, time) to do a comprehensive job of preserving all their research datasets and will instead
- want a low-cost, fully automated, 'black box' approach to digital preservation of at least some of their data.
- They want a 'preservation sausage machine' whereby research data is fed in at one end and out of the other comes 'preservation packages' containing the research data in a form that is better described and structured for long-term usability.

Interactive preservation workflow:
- Institution will want to work closely with both the Research Data and the Researcher as part of an iterative process of quality control and digital
- This is a more interactive and resource intensive process than 'automated' preservation, but can yield better results and may be more appropriate for specific types of research or institution.

The most appropriate workflow to use will depend on many factors, e.g. the experience an institution has with digital preservation, the resources at its disposal, the research discipline or type of data involved, the requirements of the research funder, the institutions policy and so on.
First use case achieved: Migrate Royal College of Music’s Open Access Repository into a Jisc repository.

» RIOXX profile is already in place

» Jisc roadmap includes OA integrations to be incorporated until early 2019 including

› Jisc services
› External services
› Additional Reporting
» Jisc services
  › Jisc Publications Router
  › Sherpa Policy Services
  › Jisc Monitor
  › CORE
  › IRUS UK

» Integrations with external services
  › CrossRef API’s
  › ORCiD Premium API
  › Publisher services

» Additional Reporting
  › Compliance checking (e.g. Research Excellence Framework)
Research4Justice academic institutions using RDSS infrastructure to build a discipline specific repository of undergraduate and postgraduate theses

- Initially centred on forensic science, but is for any content that could be useful within criminal justice
- Wide range of depositors (multi-institutional)
- Different audience: Home Office, Police, Legal, practitioners

https://creativecommons.org/licenses/by-nd/2.0

www.flickr.com/photos/wolfgangfoto
RDSS suppliers Arkivum and Artefactual systems have been examining best practice for software preservation and we will produce a report in April.

Software Sustainability Institute are working with RDSS to provide community building and information, advice and guidance.

› 1st workshop held in March 2018, 2 more to come before July 2018

› Briefing papers and reports pre-July 2018
RDSS working with Universities of Glasgow and Westminster:

- Preservation use cases for records management
  - University archives
  - University Websites
  - Special collections
  - Other records
- Recommendations around workflows, systems and future integrations
- Engage with the community through workshops
Glasgow are using RDSS infrastructure:
  › Preservica and Archivematica with Jisc storage

Westminster are using the Arkivum Perpetua service
  › Archivematica and Arkivum storage

Both organisations are due to report findings back by July 2018
RDSS pilots are also using RDSS infrastructure to examine their non-RDM use cases:

- Cambridge – Polonsky fellowships
- St Andrews – University records and archives
- Lancaster – Special collections
- Middlesex – Design Museum

Other pilots have expressed desire to work with content outside research data.

We have an MOU with The National Archives and will be engaging with their Higher Education Archives Programme (HEAP) to involve more archivists in RDM
3 standard service options

- End-to-end service
- Repository service
- Preservation service

+ +

To be launched September 2018.

All 3 options include:
- Financial benefits
- Standards
- Advisory
- Network membership
» Please talk to us if you think we can help meet your use case:

› Different content types
› Different types of organisation

» If you interested in a hosted preservation or repository service, then please contact us:

› tamsin.burland@jisc.ac.uk
› john.kaye@jisc.ac.uk
Supporting UK open research

Information sources such as:
- Publications Router
- Publishers
- Crossref
- ORCID
- DataCite
- PubMed
- Sherpa policy tools
- ...

University systems (Single Sign-On, Finance, HR...)

Jisc open science repository service

Information destinations such as:
- Google (etc)
- Discovery services
- Jisc CORE (largest global OA aggregation)
- Jisc Monitor (for compliance checking)
- Jisc Collections
- UKRI systems?
- OpenAIRE + for EU
- ...

Reports and dashboards

Preservation services

Open Access publications

Research datasets

University X repository

University Y repository

University Z repository
Thank you and questions!

Rachel Bruce
Director Open Science and Research Lifecycle
Rachel.bruce@jisc.ac.uk

John Kaye
Head of Change Research
John.kaye@jisc.ac.uk