Research Data Shared Service – Needs a new name

RDSS: A tool for co-ordination
Jisc is the UK higher, further education and skills sectors’ not-for-profit organisation for digital services and solutions.

We...

- Operate shared digital infrastructure and services
- Negotiate sector-wide deals and conditions with IT vendors and commercial publishers
- Provide trusted advice and practical assistance for universities, colleges and learning providers
- Negotiate sector-wide deals and conditions with IT vendors and commercial publishers
- Operate shared digital infrastructure and services
Mission
To enable people in higher education, further education and skills to perform at the forefront of international practice by exploiting fully the possibilities of modern digital empowerment, content and connectivity

Vision
To make the UK the most digitally advanced education and research nation in the world
The goal is developing new national shared technology services.

Delivered by collaborating across the sector.

Facilitated by Jisc as a national body with a technology focus.

Using Jisc’s co-design innovation process.
Research Data Discovery Service (Beta)

Datasets Organisations About FAQ

Home > Datasets

Search datasets… Search

Order by Relevance Advanced Search

11,796 datasets found

eHistology Kaufman Atlas Plate 14a image I
The eMouseAtlas team have re-digitised the original histological sections used for the "Atlas of Mouse Development" by Matt Kaufman. The resource is published available on the...

Improving carbon coated TiO2 films with a TiCl4 treatment for photocatalytic …
Using a simple thermal decomposition route, carbon-TiO2 hybrid films have been synthesized from a carbon-TiO2 surface complex. The resultant films display enhanced visible exten...

http://researchdiscoveryservice.jisc.ac.uk/
University Pilot Institutions

bucksm new university
Middlesex University London
UNIVERSITY OF CAMBRIDGE
UNIVERSITY OF SURREY
University of St Andrews
CARDIFF UNIVERSITY
PLYMOUTH UNIVERSITY
University for the Creative Arts
St Mary's University Twickenham London
THE UNIVERSITY OF WINCHESTER
Imperial College London
ROYAL COLLEGE OF MUSIC London
University of Glasgow
UNIVERSITY OF YORK
UNIVERSITY OF LINCOLN
Lancaster University
St George's University of London
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
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)
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:

**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

- **Advise & best practise**
- **Capture & reuse**
- **Preserve**
- **Report**

Source: Jisc DAF Survey results 2016
“Support is woeful in the university currently, in particular long-term data archiving is critically required. Most of my non-current data is rotting on CD's and hard-drives.”
“Please, individualise the support. Workshop are useless, emails with information are useless, brochures are useless, posters are useless.”
“Please, individualise the support. Workshop are useless, emails with information are useless, brochures are useless, posters are useless.”
“I currently spend about £1,200 pa on data storage from my own salary. I have the highest data needs in my School, and there is no plan in place for storing my data.”
» Diverse practices, no-one is the same

» Pockets of good disciplinary practice, where served by existing disciplinary or institutional infrastructure

» Generally at a central function most institutions are at the start of implementing systems and practice

» Researcher engagement is key

» Desired configuration of research systems varies between institutions

» Large institutions often have lots of locally managed active data, where there could be a storage need.

» Often case for RDM has been made, but institutions will have to make the case for RDSS and we want to help
**RDSS functional solution at a glance**

Cost-effective improvement in research data access and reuse

**Jisc Research Data Shared Service**

**Capture & reuse**
- Deposit
- Describe
- Store
- Publish

**Preserve**
- Assure Integrity
- Normalise
- Transform
- Curate
- Flag at-risk data

**Report**
- RDM planning
- Costs
- Service performance
- Audit trails
- Compliance and benchmarking

**Research data**
- Secure outputs
- Accelerate research
- Grow reusable data value
- Managed storage included

**Shared standards-based technology framework**
- 3rd party tools
- Integration with local systems
- Open Standard Interfaces (APIs)

**Advice & best practice**
- Research data management toolkit and network

19/01/2018 Jisc research data shared service - RDSS: A tool for co-ordination
RDSS Components

Pilots

- Imperial College London
- University of York
- University of Cambridge
- University of St Andrews
- University of Lincoln
- Lancaster University
- St George's
- Plymouth University
- Middlesex University London

CRIS, Grants
HRU, Finance
Publications repositories
Data management plans
Authorisation

Integration hub

- Data repository
- Preservation
- Reporting

Researchers interface and UK

Large and sensitive data deposit tools

Cloud data storage
Local Jisc shared data centres
Archival data storage

Identifiers
Aggregators
Metrics
Funders
Institutional or external services

Repository

National research data aggregation

Researchers find and reuse data

Preservation service

Reporting and analytics

Data added to aggregation

Data is automatically preserved

Use of data and service is monitored

Other services are updated

1. Researcher deposits data

2. Researchers find and reuse data

3. Data is automatically preserved

4. Use of data and service is monitored

5. Other services are updated

6. Researchers find and reuse data

Simplified Workflow

19/01/2018

Jisc research data shared service - RDSS: A tool for co-ordination
RDSS Data Model
Alpha Phase

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Organisation</th>
<th>Status</th>
<th>Archivematica</th>
<th>Preservica</th>
<th>Repository</th>
<th>Preservation Experts Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cambridge</td>
<td>Already have access to Preservica and Archivematica. Repository to be delivered</td>
<td><img src="logo.png" alt="Archivematica" /> <img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Lancaster</td>
<td>Already have access to Preservica and Archivematica. Repository to be delivered</td>
<td><img src="logo.png" alt="Archivematica" /> <img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Lincoln</td>
<td>Already have access to Archivematica. Repository to be delivered</td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>York</td>
<td>Already have access to Archivematica. Repository to be delivered</td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td>2</td>
<td>Middlesex</td>
<td>Both systems delivered - Production</td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>St George’s</td>
<td>Both systems delivered - Production</td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Surrey</td>
<td>Both systems delivered - Test</td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td>3</td>
<td>Cardiff</td>
<td>Preservica delivered, Repository to be delivered</td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Imperial</td>
<td>Preservica delivered, Repository to be delivered</td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Plymouth</td>
<td>Preservica delivered, Repository to be delivered</td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Royal College of Music</td>
<td>Preservica delivered, Repository delivered and migrated in October</td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td>4</td>
<td>St Mary’s</td>
<td>Archivematica delivered, repository currently being implemented</td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>UCA</td>
<td>Archivematica delivered, repository currently being implemented</td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Bucks New</td>
<td>Archivematica delivered, repository currently being implemented</td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>Winchester</td>
<td>Archivematica delivered, repository currently being implemented</td>
<td><img src="logo.png" alt="Archivematica" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Figshare" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td>5</td>
<td>Glasgow</td>
<td>No Repository, both preservation systems delivered</td>
<td><img src="logo.png" alt="Archivematica" /> <img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
<tr>
<td></td>
<td>St Andrew’s</td>
<td>Already have access to Preservica, Archivematica and Pure interface prototype to be delivered in November</td>
<td><img src="logo.png" alt="Archivematica" /> <img src="logo.png" alt="Preservica" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
<td><img src="logo.png" alt="Preservation Experts Cluster" /></td>
</tr>
</tbody>
</table>
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.
Automated Workflows

Jisc research data shared service - RDSS: A tool for co-ordination
<table>
<thead>
<tr>
<th>Workflow</th>
<th>No of Participants</th>
<th>Score 0 / 100</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIP Creator &amp; Preservica Interactive</td>
<td>2</td>
<td>43.8</td>
<td>F &quot;Poor&quot;</td>
</tr>
<tr>
<td>Figshare &amp; Preservica Automated</td>
<td>3</td>
<td>43.3</td>
<td>F &quot;Poor&quot;</td>
</tr>
<tr>
<td>Nextcloud &amp; Archivematica Interactive</td>
<td>3</td>
<td>55</td>
<td>F &quot;Poor&quot;</td>
</tr>
<tr>
<td>Samvera &amp; Archivematica Automated</td>
<td>4</td>
<td>61.9</td>
<td>D &quot;OK/Fair&quot;</td>
</tr>
</tbody>
</table>
## User Experience

<table>
<thead>
<tr>
<th>Workflow</th>
<th>No of Participants</th>
<th>Score -100 / 100</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIP Creator &amp; Preservica Interactive</td>
<td>2</td>
<td>-50</td>
<td>Mostly Detractors</td>
</tr>
<tr>
<td>Figshare &amp; Preservica Automated</td>
<td>3</td>
<td>-100</td>
<td>Mostly Detractors</td>
</tr>
<tr>
<td>Nextcloud &amp; Archivematica Interactive</td>
<td>3</td>
<td>-100</td>
<td>Mostly Detractors</td>
</tr>
<tr>
<td>Samvera &amp; Archivematica Automated</td>
<td>4</td>
<td>-75</td>
<td>Mostly Detractors</td>
</tr>
</tbody>
</table>
» Demos of Automated transfer between Samvera (formally Hydra) and Preservica by Dom Fripp

» Other Demos coming soon:
  › Samvera > Archivematica
  › Figshare > Preservica
  › Figshare > Archivematica
  › Pure > Archivematica
  › Pure > Preservica
Average time spent on each technology, if successful, for the automated Willow & Archivematica test

- Willow: 8:00 minutes
- Archivematica: 2:00 minutes
Average time spent on each technology, if successful, for the automated Figshare & Preservica Test
Average time spent on each technology, if successful, for the interactive SIP Creator and Preservica Test

- SIP Creator: 5.5 minutes
- Preservica: 3.5 minutes
Sharing good practice “recipes” ([who does] what and why)
Collaboration between Arkivum/Artefactual/Preservica, potential to broaden it out globally
Point to point exchange of information
Working on:
- Data model and API (PAR Core [thanks Matthew Addis for that])
- Reference registry
- Proof of concept
Project timeline

2017
Jan  ↔ Aug  Sep  Oct  Nov  Dec  Jan  Feb  Mar  Apr  May  Jun  Jul
ALPHA

Establish Project Steering Committee (then meet monthly)

2018
Jan  ↔ Aug  Sep  Oct  Nov  Dec  Jan  Feb  Mar  Apr  May  Jun  Jul
ALPHA

Approval for Business Case by and further work to complete Alpha

BETA

Approval to sign off Alpha MVP and commence Beta

SERVICE TRANSITION

Approval to sign off Beta MVP and move to full production
3 standard service options

End-to-end service + 

Repository service + 

Preservation service + 

Approach supports pilot investments in a range of components, direction of travel towards standard managed solution for all over time.

All 3 options include:
✓ Financial benefits
✓ Standards
✓ Advisory
✓ Network membership
Customer side Beta MVP (post March business case approval)

- **Jisc open source beta repository**
  - Samvera and TBC

- **Jisc beta preservation**
  - Preservica
  - Digital Preservation
  - Arkivum

- **Jisc beta reporting**
  - Full reporting capabilities or APIs available

Core RDSS offering

Bronze, Silver or Gold data storage packages available

Available framework options

- Jisc research systems framework repositories
- Jisc research systems framework preservation
- Jisc research systems framework CRIS

Fully accredited infrastructure, technical underpinning and interoperability with UK Data centre cloud hosting for your data by Jisc.
## RDSS Potential Integrations

<table>
<thead>
<tr>
<th>RDSS Potential Integrations</th>
<th>Number of Pilots (total = 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eprints (Repository)</td>
<td>12</td>
</tr>
<tr>
<td>Dspace (Repository)</td>
<td>4</td>
</tr>
<tr>
<td>Samvera (Repository)</td>
<td>2</td>
</tr>
<tr>
<td>Symplectic (CRIS)</td>
<td>4</td>
</tr>
<tr>
<td>Pure (CRIS)</td>
<td>3</td>
</tr>
<tr>
<td>Converis (CRIS)</td>
<td>1</td>
</tr>
<tr>
<td>Authentication - UKAMF</td>
<td>17</td>
</tr>
</tbody>
</table>
» Jisc or 3rd Party hosted and managed Repository and Preservation system (for use for Research data and other outputs)

» Training, Support and gathering best practice

» Event based integration between the Repository and Preservation system (Alpha).

» Flexible metadata schemas for different content types
» Cloud storage on AWS and Arkivum, with additional JANET connected UK owned and located private cloud from UK Cloud coming on stream

» Discovery work on integration with institutional systems e.g. existing publication repositories, Records Management Systems

» R&D and prototype development in ‘Problem areas’ e.g. Research Data File Formats, Sensitive and large data (Beta).
Pilot Alpha Phase Definition of Done - Headlines

» Repositories (7 Figshare, 8 Samvera (Willow))
  › Released for Researchers to describe, deposit and publish data to agreed standards, meeting funder compliance
  › Administrators to QA, approve, query data deposits

» Preservation (10 Archivematica, 11 Preservica)
  › Automatically ingests repository deposits (Currently working for both systems)
  › Carry out preservation events to institutional or default plans

» User Experience Assessments and Acceptance Testing Complete

» Integration with external services: ORCiD, DataCite, IRUS Data UK, Elsevier Pure and more

» Integration with Jisc storage

» End to end Integration with Jisc event based architecture and data model enabling automated workflow between systems (Currently working for most systems in Jisc test)
Repositories updated from user feedback with UX, API and scalability improvements and ability to ingest large (5>GB Data)

Preservation updated from user feedback with UX, API and scalability improvements

Reporting service providing research data and system metrics by Dashboard and API

Single sign on for systems (UKAMF/SAML)

Service API released

Integration with more common external services e.g. Eprints, Dspace

Integration with Institutional, Private cloud and Archival Storage and DataVault for raw, un-curated data

Security assurance: web application firewall, Penetration testing, threat modelling.

Trusted Repository certification: Data Seal of Approval

Information, Advice and Guidance research data management toolkit released
» Working on new GDPR compliant DPA agreements/templates

» Planning a GDPR for Research Data workshop

» Longer term working on appropriate security certifications
Post-Beta Developments

» Integration with Jisc Safe Share
» Integration with Jisc Tiered Storage
» Integrations with additional 3rd Party integrators
» Integration with updated DMP Online
» Integrations with active data storage e.g. Institutional storage area network, file sync and share services.
Data is the universal language of research
Disadvantaged research communities can't share their data.
.....don’t fit local budgets
Collaborations

» Current/future Collaborations and opportunities
  › Figshare/Digital Science (Africa)
  › Jisc and Samvera Community
  › Jisc/Digirati/ORCiD
  › Jisc/Elsevier Open Science Agreement
  › Jisc and HPC Community via Imperial College and Pawsey Centre (Australia/SKA)
  › OpenAire Bid to enhance Samvera
  › Want to work to integrate DMPOnline (DCC/CDL)
  › Open to any other suggestions!
Thank You and Questions!

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

Dom Fripp – Product owner, Samvera
paul.stokes@jisc.ac.uk

Paul Stokes, Senior Co-Design Manager
paul.stokes@jisc.ac.uk

Alan Mackenzie, Chief Architect
alan.mackenzie@digirati.com