Putting preservation into practice using shared preservation infrastructure
» Jisc Shared Service Overview – Dom Fripp

» Preservation System Supplier – Arkivum: Matthew Addis - and Artefactual (Archivematica): Peter Van Garderen and Kelly Stewart

» Preservation System Supplier – Preservica: Jon Tilbury

» Coffee

» Panel and questions
Shared Service Goals

» Policy compliance
» Efficiencies
» Better research
» Better Discovery
» Better Preservation
A KEY requirement
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)
Service Vision
Service MVP (Alpha – test systems to be delivered July/Aug 2017)

User Experience: Principles, Governance, Testing and Development

Repository Application
- Data Ingest
- Data Catalogue
- Data Discovery
- Data Publication/Download
- User Area
- Admin Area

Preservation Application
- Carry out preservation policies
- File format identification, transformation, migration
- Admin Area

Data Storage
- Online Access Data Object Store
- Archival Data Store and backup
- Integrity Checking

Reporting Application
- Discovery
- Define metrics
- Define Dashboards/API access

Institutional Systems Integrations
- Institutional Repositories
- CRIS Systems

Other Integrations
- Service API
- DataCite, ORCiD
- IRUS Data UK
- Research Data Discovery Service

Technical underpinning: data model & architecture: event notification, APIs, infrastructure as code, micro-services, authentication & security.

Provided by RDSS Framework - Platforms
Provided by RDSS Framework - Development
Provided by Jisc / Other Jisc procurement route
Service MVP (Iterations to Spring 2018)

User Experience: Principles, Governance, Testing and Development

**RD Preservation Development**
- Enhancements to systems and tools to manage research data

**Large Data Management**
- Methods for ingesting large datasets

**Sensitive Data Management**
- Methods for ingesting, managing and accessing sensitive data

**Repository Application**
- UX improvements
- Additional features e.g. extra previews
- Large data ingest
- Managing sensitive data

**Preservation Application**
- UX improvements
- Incorporate research data preservation development

**Reporting Application**
- Delivery and testing
- Additional metrics
- Additional dashboards

**Institutional Integrations**
- Active Data Storage
- HR Systems
- Finance Systems
- Other e.g. Impact, Ethics systems

**Data Storage**
- Hierarchical storage management (HSM)
- Large data capabilities
- Sensitive data capabilities

**External Integrations**
- Identify and implement additional integrations

**Technical underpinning:** Testing and improvements to data model & architecture for scalability, robustness and security

Provided by RDSS Framework - Platforms
Provided by RDSS Framework - Development
Provided by Jisc / Other Jisc procurement route

27/06/2017  Jisc Research Data Shared Service
There is no “One size fits all”
Results of the survey

**HEIs' role**
75% of respondents look first to their institution to preserve their data

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

**Public data**
>70% recognise that research is a public good and should be open to all

**Use of DMPs**
Only 40% of respondents have a Research Data Management plan

**Metadata**
Only 16% of respondents say they follow established metadata guidelines

**Sensitive data**
41% of respondents have some form of sensitive data

17% of respondents had lost data, resulting in...

- Wasted research effort
- Delay to publication
- Reduction in quality of research outputs
- Reputation damage
- Failure to meet funder requirements
- Failure to meet regulatory requirements

Top three causes for loss of data:
1. Hardware failure
2. Human error
3. Equipment stolen

Where do respondents store their research data?

- Hard disk drive of computer owned by the University
- External hard drive or memory stick / USB / Flash drive
- Hard disk drive of a privately owned computer
- University managed network storage
- Cloud storage - Dropbox

Top 5 locations

- All
- A substantial amount
- Some
- None

How long do respondents expect data to be kept?

- 39%
- 24%
- 20%
- 16%
- 5 - 10 yrs
- >10 years
- I don’t know
- 1 - 5 yrs
“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.”

“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.”

“Please, individualise the support. Workshop are useless, emails with information are useless, brochures are useless, posters are useless.”

“It would be helpful to clarify the rules for storing anonymised data on cloud services. My departmental rules say this is never OK, however this seems to contradict University rules.”
Welcome to the Research Data Network. This is a site to exchange information on research data management and related topics within the UK research information management sector. We hope that the site can build a body of useful resources, help people share up to minute developments, practices and develop new ideas.

Let's jump right in!

GET STARTED  ANNOUNCEMENTS

View All

New site features added
Google Doc for UKSG 2017 Research Data Shared

How to use the network site
Research Data Network Workshops

What features would you like to see on the RDN site?
Research Data Network – June 2017

» When: 27th to 28th June 2017
» Where: York, Ron Cooke Hub
» Cost: Free!

Registration now open
http://jisc.ly/UD2if7
More information on the RDN website
http://researchdata.network
Useful Resources

» Blog: https://researchdata.jiscinvolve.org/wp/

» Research Data Network: http://researchdata.network

» Shared Service web page: https://www.jisc.ac.uk/rd/projects/research-data-shared-service

» Github https://github.com/JiscRDSS

» Zenodo Community: https://zenodo.org/communities/rdn