Stuart Lewis
Deputy Director, Library & University Collections
The University of Edinburgh

Robin Taylor
Library Software Developer
The University of Edinburgh

@JiscDataVault
Introducing the DataVault

• Practice paper:

  1. What is the DataVault
  2. Use cases
  3. How does it work
  4. Demo
  5. What do you think?
Research Data Management Services

Data Management Support

- Data Management Planning
- Active Data Infrastructure
- Data Stewardship
Research Data Management Services

Data Management Support

Data Management Planning

Active Data Infrastructure

Data Stewardship
Data Stewardship

• DataVault
  • Long term archival storage
  • First envisaged a few years ago...
What is the DataVault - Analogies

https://www.flickr.com/photos/brookward/8457736952
What is the DataVault - Analogies

https://www.flickr.com/photos/timshelyn/4125480767
Where does it sit?
What is the DataVault?

• It’s a platform (a bit of web-based software), not a storage system
  • You still need an archival storage system

• It is provided via a RESTful API
  • But also comes with a default web user interface application

• It’s doesn’t undertake any active preservation activities
  • It does collect metadata and package the data for long-term storage
DataVault use cases

• Store golden-copy data so that it can’t be changed

• Storing finalised data that can’t be openly shared (or ‘everything else’)

• Retaining old data in case it can be re-analysed later on

• Retention to meet funder requirements
Information Architecture

Users          Vaults        Deposits     Bag

Archival Storage

Technical Architecture

Active Storage

datavault-webapp

datavault-broker

RabbitMQ

datavault-worker(s)

Archival Storage
Project funding

• Funded by Jisc

• DataSpring ‘Research at Risk’ programme

• Three phases

• Ideascale → pitches
Phase 1 (3 months)
Phase 2 (4 months)
Phase 3 (6 months)
Deposit/Vault size (bytes) incorrect - database update.

Application logging infrastructure

Use of the phrase 'retrieve' rather than restore

Investigate UoE SSH key setup

First version of help text / user guide

Arrange engagement event - 29th June (or 30th)

Onboarding: First time login UI - setup storage, connect to dropbox etc.

Remove or fix "default" file store configuration (added for testing).

System Administrator / Data Manager (sub-admin)

Tidy log file names

Shibboletih (Spring security?)

Clean-up /users endpoint

Email admin when job fails

Check boundaries and limits (e.g. max length of vault name in db, user agent length)

Technical scenario testing

Email user when deposit is complete.

Starting the worker process seems to require an internet connection to download/verify the spring XSD. I think it's possible for this to be checked against the local Jar instead.

Metadata service for integration with the CRIS system

Store raw CRIS metadata in bag metadata folder

Metadata entry page as part of vault creation flow. Automatic fill of fields from PURE.

Out of the box retention policies

User interface tooltips and inline text help.

Review file picker - confirm file
Demonstration time...

Data Vault
Keep your research data safe.

Welcome to the Data Vault!
Please use the 'Create new Vault' button below to get started...

Create new Vault
Demonstration time...

• Try it yourself...

http://demo.datavaultplatform.org/
What do you think?

- Do you agree with the use cases?
- Do you see a need for such a tool?
- Do you think the tool approaches preservation appropriately?
- Do you think it is missing any obvious features?

- Are you going to investigate it further!
  - Please talk to us!
Do you think the DataVault does enough 'preservation'?

Poll locked. Responses not accepted.

“I don't feel like I have enough information to answer the question.”
3 months ago

“No, not as yet”
3 months ago

“For its limited purpose, yes.”
3 months ago

“Its doing bit preservation which is necessary, but not sufficient.”
3 months ago

“Unsure - need more information”
3 months ago

“depends on the use cases”
3 months ago

“No...”
3 months ago

“Surely, more complex formats will become unusable over time ??”
3 months ago

“No”
3 months ago

“No, but is better than nothing. And couldn't it be expanded?”
3 months ago
Do you think it is missing any obvious features?

Poll locked. Responses not accepted.

“Warning to depositor that certain formats might not work in the long term?”
3 months ago

“Yes”
3 months ago

“So far, so good!”
3 months ago

“File format metadata?”
3 months ago

“Large datasets can be problematic to upload via https”
3 months ago

“researchers might think everything is fine now”
3 months ago

“How do you insure sustainability?”
3 months ago

“How many petabytes it can manage?”
3 months ago

“Warnings upfront to researchers”
3 months ago

“How does it record any metadata associated with events (e.g. PREMIS)?”
3 months ago

“Incremental updates.”
3 months ago

“costs?”
3 months ago

“It does what is says on the tin”
3 months ago

Poll Everywhere
Upgrade to enable moderation
Hot off the press...

Data Vault Implementation Workshop

Wednesday 29th June 2016, Central London (Islington)
15 spaces (1x RDM + 1x Tech)
Keep an eye out for the booking form / speak to me
What is a Data Vault?

Robin Taylor
robin.taylor@ed.ac.uk
@JiscDataVault