



Stuart Lewis

Deputy Director, Library & University Collections  
The University of Edinburgh

Thomas Higgins

RDM Service Software Developer  
The University of Manchester

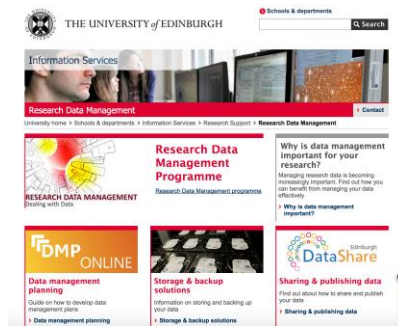
 @JiscDataVault



THE UNIVERSITY  
of EDINBURGH



# Research Data Management Services



## Data Management Support



Data  
Management  
Planning



Active Data  
Infrastructure

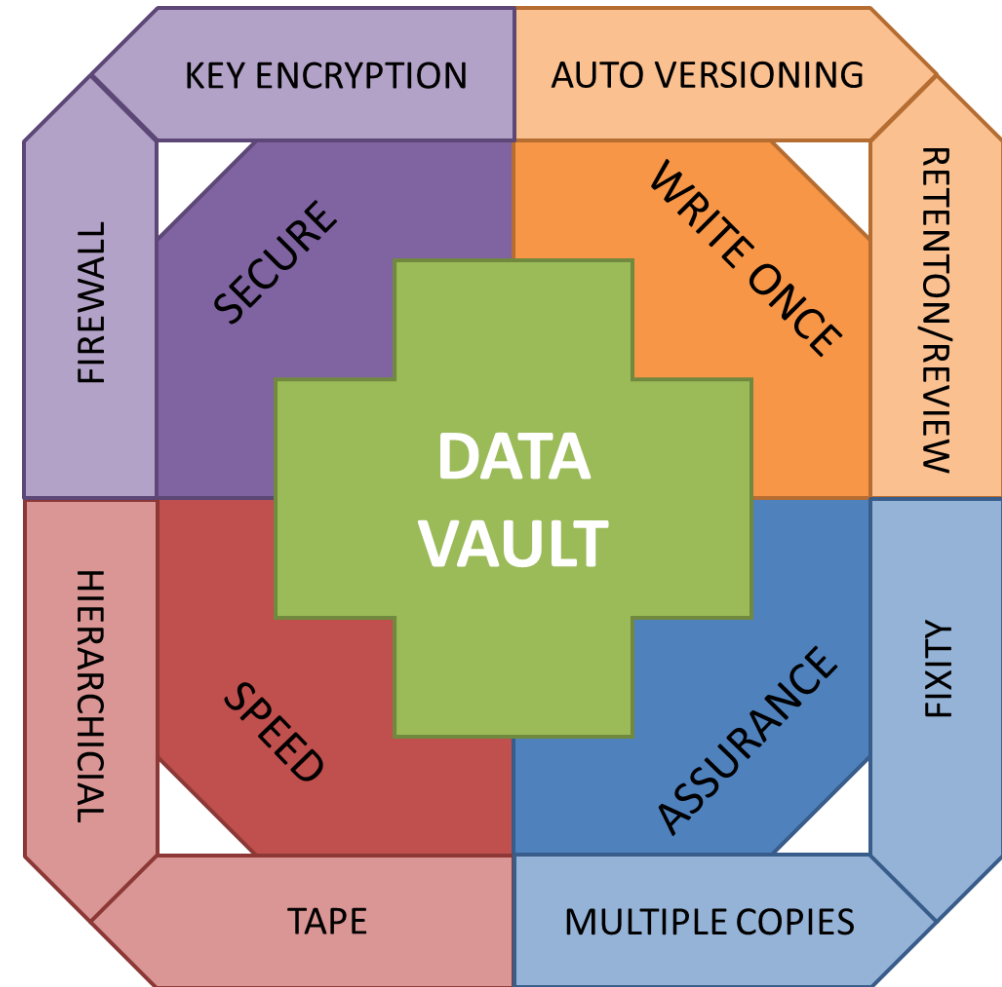


Data  
Stewardship



# Data Stewardship

- DataVault
  - Long term archival storage
  - First envisaged a few years ago...



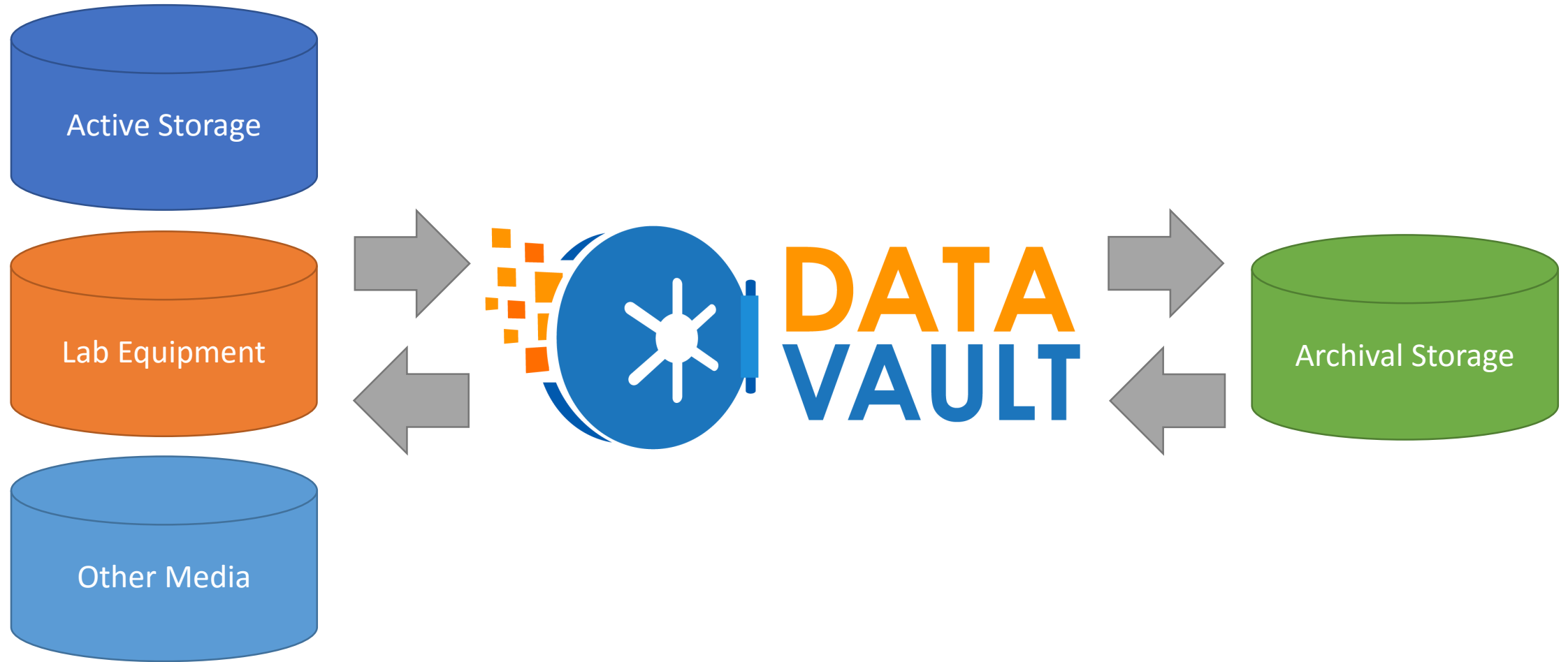
# What is the DataVault - Analogies



# What is the DataVault - Analogies



# Where does it sit?





# What is the DataVault?

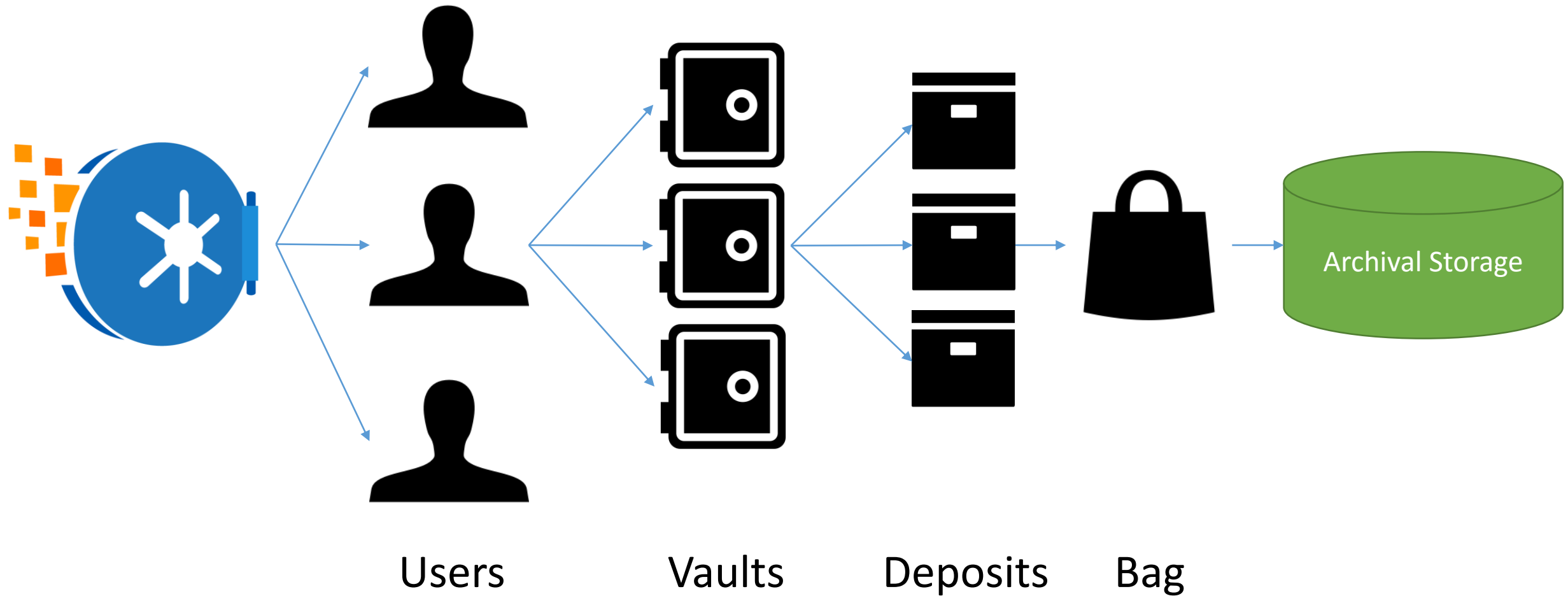
- It's a platform (a bit of web-based software), not a storage system
  - You still need an archival storage system
- It is provided via a RESTful API
  - But also comes with a default web user interface application
- It doesn't undertake any 'active' preservation such as format migration
  - It does collect metadata and package the data for long-term storage

# DataVault use cases

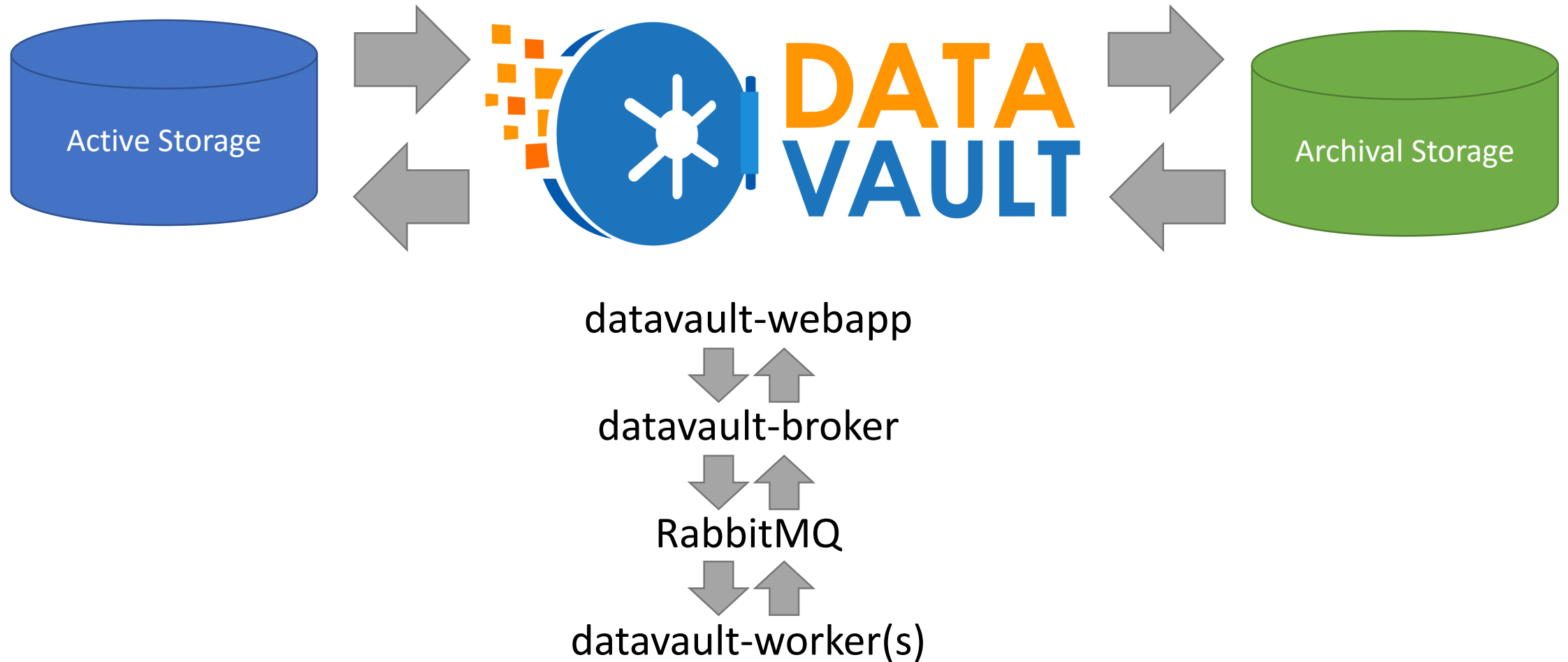
- Store golden-copy data so that it can't be changed
- Storing finalised data that can't be openly shared (or 'everything else')
- Retaining old data in case it can be re-analysed later on
- Retention to meet funder requirements



# Information Architecture

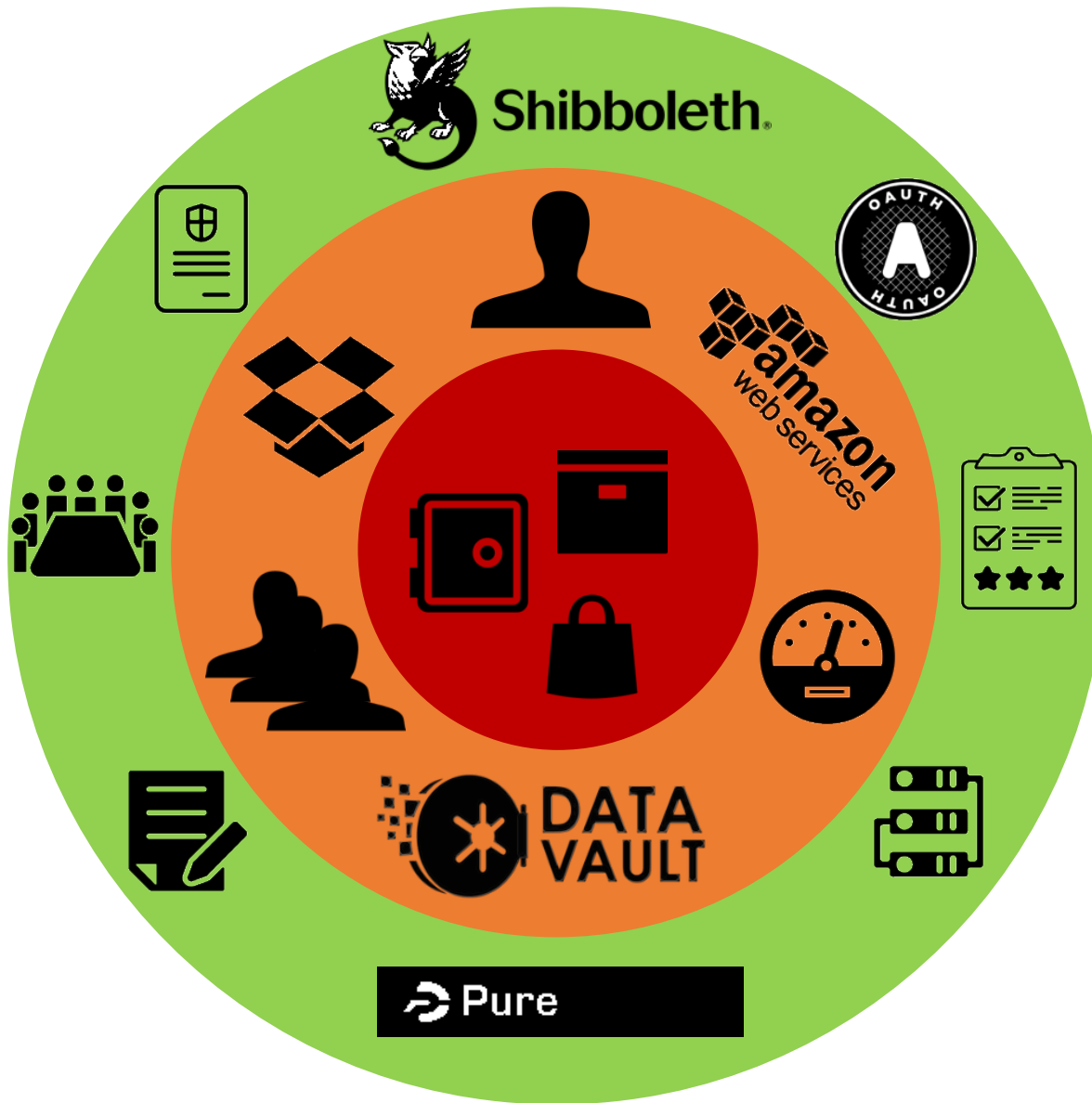


# Technical Architecture





....and for 50%  
stake in our  
project we're  
are looking for  
investment  
of....



- Phase 1 (3 months)
- Phase 2 (4 months)
- Phase 3 (6 months)

# 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

**Vault Name** ⓘ

My research project

**Description** ⓘ

Images of plants and organisms with associated metadata.

**Relates to** ⓘ


Sample dataset 1 ▼

**Retention Policy** ⓘ

EPSRC ▼

**Group** ⓘ

Science and Engineering ▼

 Create new Vault

Cancel

# My research project

Images of plants and organisms with associated metadata.

---

**Owner:** user1

**Retention Policy:** EPSRC

**Group:** Science and Engineering

**Created:** Jun 14, 2016 8:30:02 AM

**Review date:** Jun 14, 2026 8:30:02 AM (Status: OK)

**Size:** 0 bytes


 [Deposit data](#)





Deposit Note: ⓘ


First set of data


Deposit file or directory: ⓘ


▼  Default filesystem (local)


 .DS\_Store


▶  active


▼  bioimages


 images.csv


 license.xml


 names.csv

 organisms.csv


 samples.bin

▶  sparql

▶  xquery

▶  sample data

Deposit size: 1 GB

 Deposit data

Cancel

# First set of data



Transferred 700 MB of 1 GB (69 MB/sec)




Deposit

Contents 0

Events 0

Retrieves 0

Note	Status	Size	Timestamp
First set of data	 In progress	0 bytes	Jun 14, 2016 8:38:56 AM

# First set of data

Deposit

Contents 30

Events 6

Retrieves 0

Note	Status	Size	Timestamp
First set of data	✓ Complete	1 GB	Jun 14, 2016 8:38:56 AM

 Retrieve data

# ↓ First set of data

[Deposit](#)[Contents](#) **30**[Events](#) **6**[Retrieves](#) **0**





File	Type	Checksum
bioimages/license.xml	application/xml	7bfaf745438d1bd27200f67d7f40705d <a href="#">md5</a>
bioimages/xquery/index-files.xq	application/xquery	efc6c7648410eafc70d838f6428585e2 <a href="#">md5</a>
bioimages/xquery/image-rdf-all-hack.xq	application/xquery	a14223f7541240120352f422cc0b7f4b <a href="#">md5</a>
bioimages/sparql/counting-query.sparql	text/plain	fee33e1b43017b928f638d5674085027 <a href="#">md5</a>
bioimages/images.csv	text/csv	6e3accf46f1fba3a3b7415bb331294b5 <a href="#">md5</a>
bioimages/sparql/.DS_Store	application/octet-stream	fa4ade8b137c27ae09fe503ed8c32cf7 <a href="#">md5</a>
bioimages/xquery/.DS_Store	application/octet-stream	41d509a2c82675c794f77087ed54a3f5 <a href="#">md5</a>
bioimages/organisms.csv	text/csv	349384c29dc3ded0a96b6c95d6ca149a <a href="#">md5</a>
bioimages/xquery/organism-html.xq	application/xquery	bfd7f162d046660276557077c6cd2ee9 <a href="#">md5</a>
bioimages/sparql/inverse-reasoning.sparql	text/plain	027f404d125aba81df236c19116fc5c2 <a href="#">md5</a>

Describe the reason for this retrieve request (who and why) and choose a working directory to retrieve data from the archive

**Retrieve Note:** ⓘ

Restoring my data|

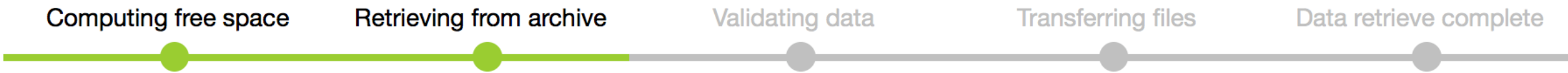
**Target directory:** ⓘ

- ▼  Default filesystem (local)
  - ▶  active
  - ▶  bioimages
  - ▶  sample data

 Retrieve data

Cancel

# First set of data



Transferred 840 MB of 1 GB (141 MB/sec)



Deposit

Contents 30

Events 6

Retrieves 1

Note	Status	Size	Timestamp
First set of data	✓ Complete	1 GB	Jun 14, 2016 8:38:56 AM

 Retrieve data

### Vault Size

 1 GB

### Deposits / Retrieves In Progress

 0  0

### Deposit / Retrieve Queues

 0  0

### Vaults

 1

### Deposits

 1

### Retrieves

 1

### Users

 31

### Groups

 3

### Review

 0

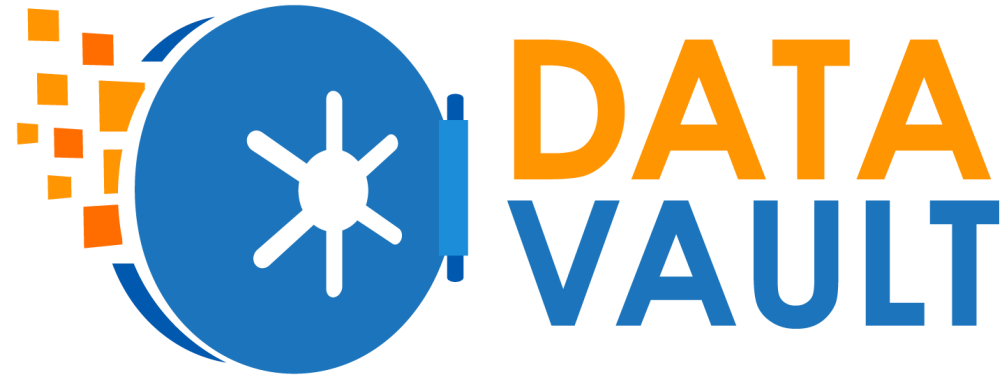
### Events

 23

### Retention Policies

 5





# Data Vault Implementation Workshop

Wednesday 29<sup>th</sup> June 2016, Central London (Islington)

15 spaces (1x RDM + 1x Tech)



THE UNIVERSITY  
*of* EDINBURGH





# DATA VAULT

## Stuart Lewis

stuart.lewis@ed.ac.uk

 @stuartlewis

## Thomas Higgins

thomas.higgins@manchester.ac.uk

 @JiscDataVault



THE UNIVERSITY  
of EDINBURGH

