

19/01/2018

## Research Data Shared Service – Needs a new name

RDSS: A tool for co-ordination



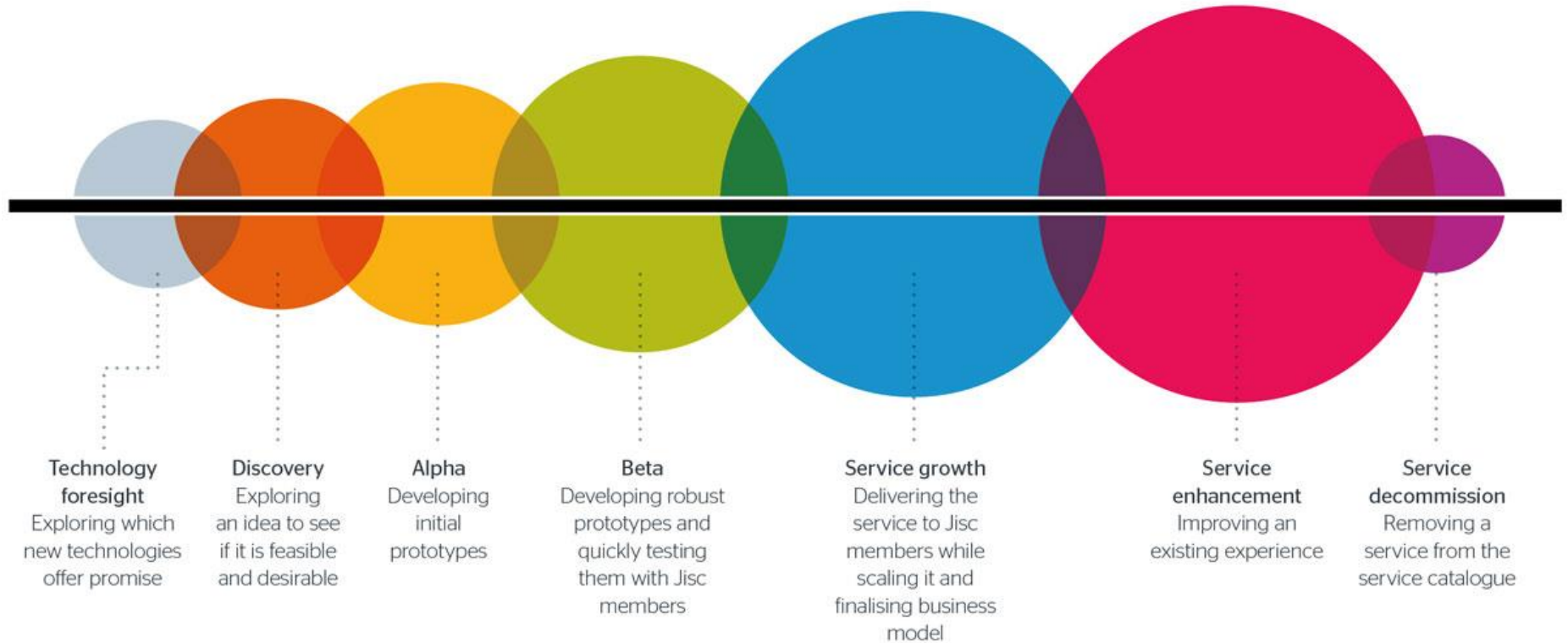
## Vision

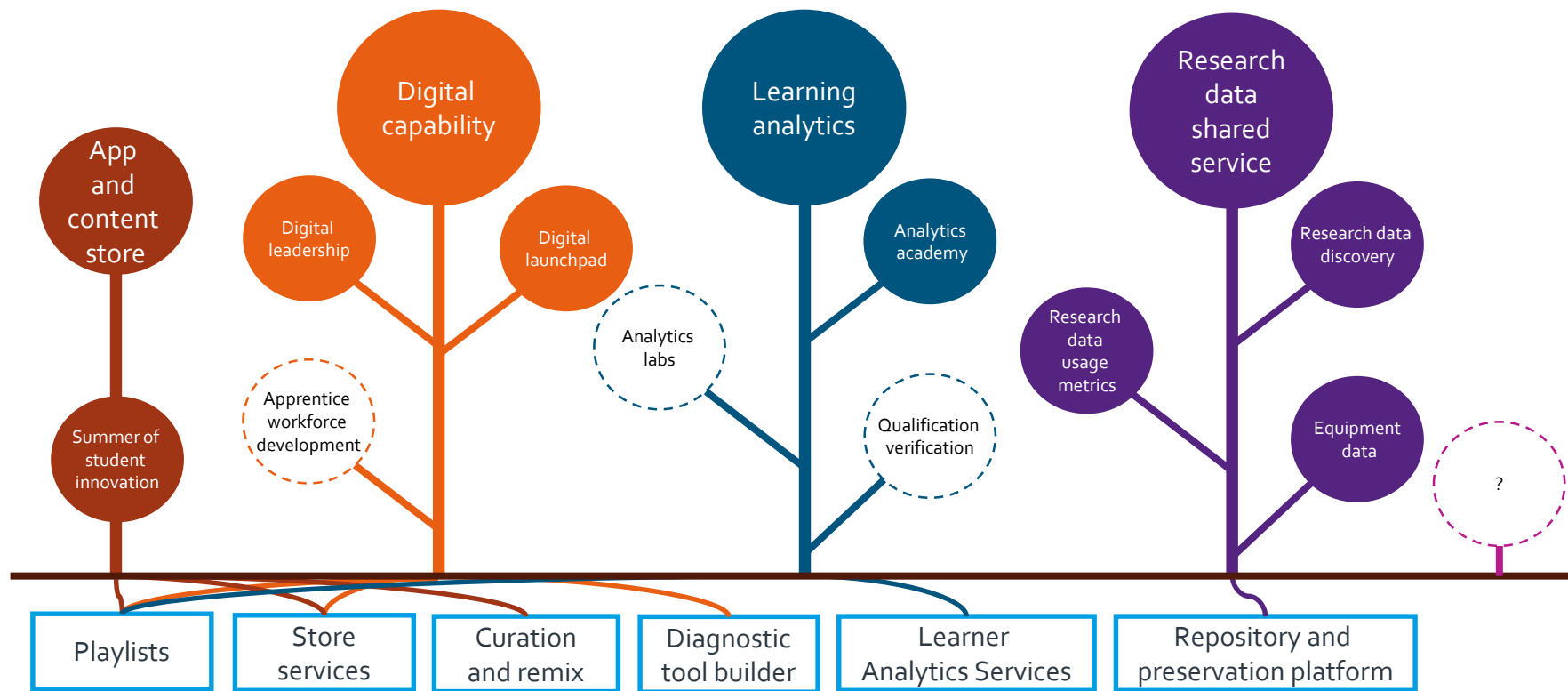
To make the **UK** the most **digitally advanced education and research nation in the world**

## 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**







## Research Data Discovery Service (Beta)

Datasets

Organisations

About

FAQ

Home &gt; Datasets

Filter by location

[Clear](#)

Map data © OpenStreetMap contributors  
Tiles by Stamen Design (CC BY 3.0)

## Organisations

Natural Environment... (5264)

University of Edinb... (1937)

Science and Technol... (1698)

Search datasets...

Search

Order by

Relevance

[Advanced Search](#)

11,796 datasets found

## eHistology Kaufman Atlas Plate 14a image I

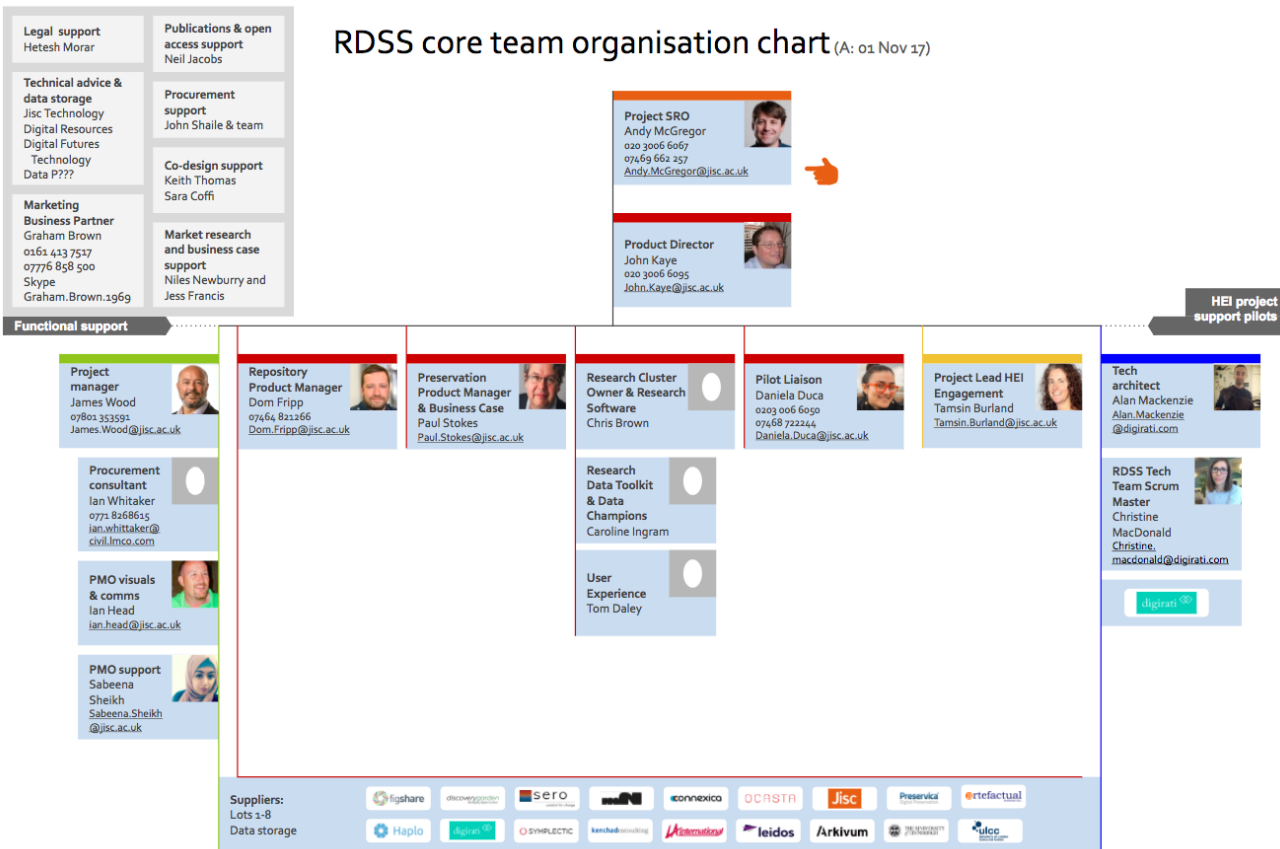
The eMouseAtlas team have re-digitised the original histological sections used for the "Atlas of Mouse Development" by Matt Kaufman. The resource is published available on the...

## 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 catechol-TiO2 surface complex. The coated films display enhanced visible region

<http://researchdiscoveryservice.jisc.ac.uk/>

## RDSS core team organisation chart (A: 01 Nov 17)









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)

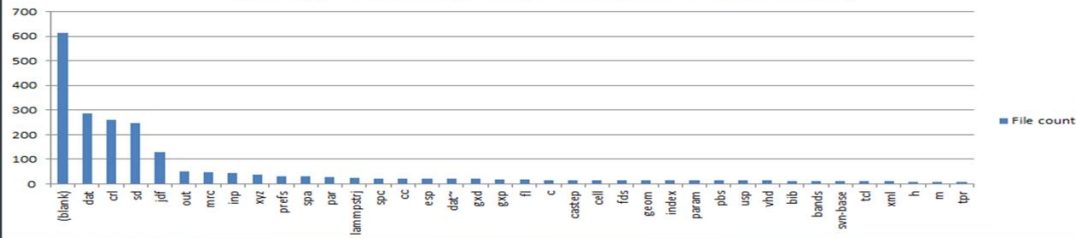
...note that native files of the software in the previous graph of research data applications are not represented



# 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)**



## Implementing Archivematica for research data preservation at York and Hull

Jenny Mitcham  
(Digital  
Archivist) -  
University of  
York

# Key researcher issues drive RDSS functionality

## 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



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:

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."*

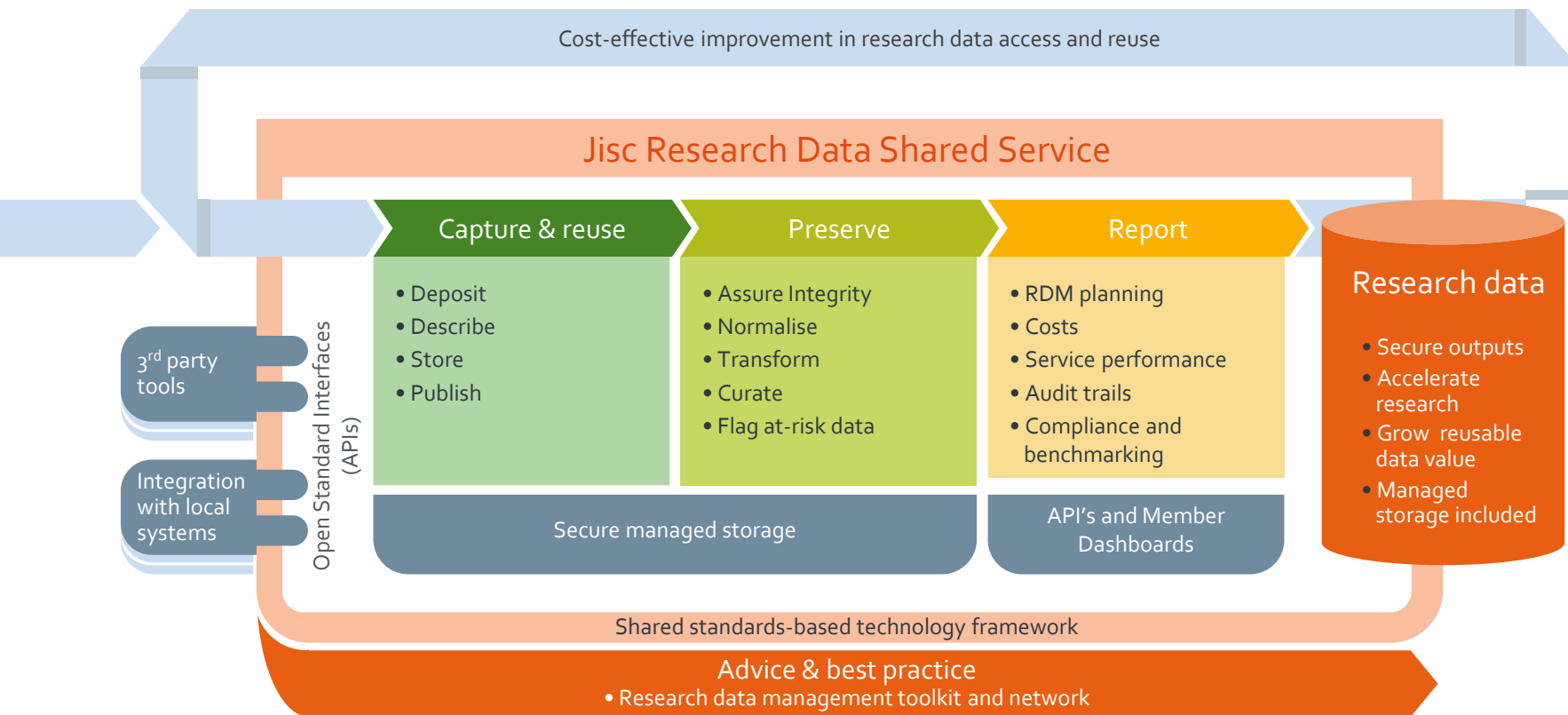
*"Please, individualise the support. Workshop are useless, emails with information are useless, brochures are useless, posters are useless."*

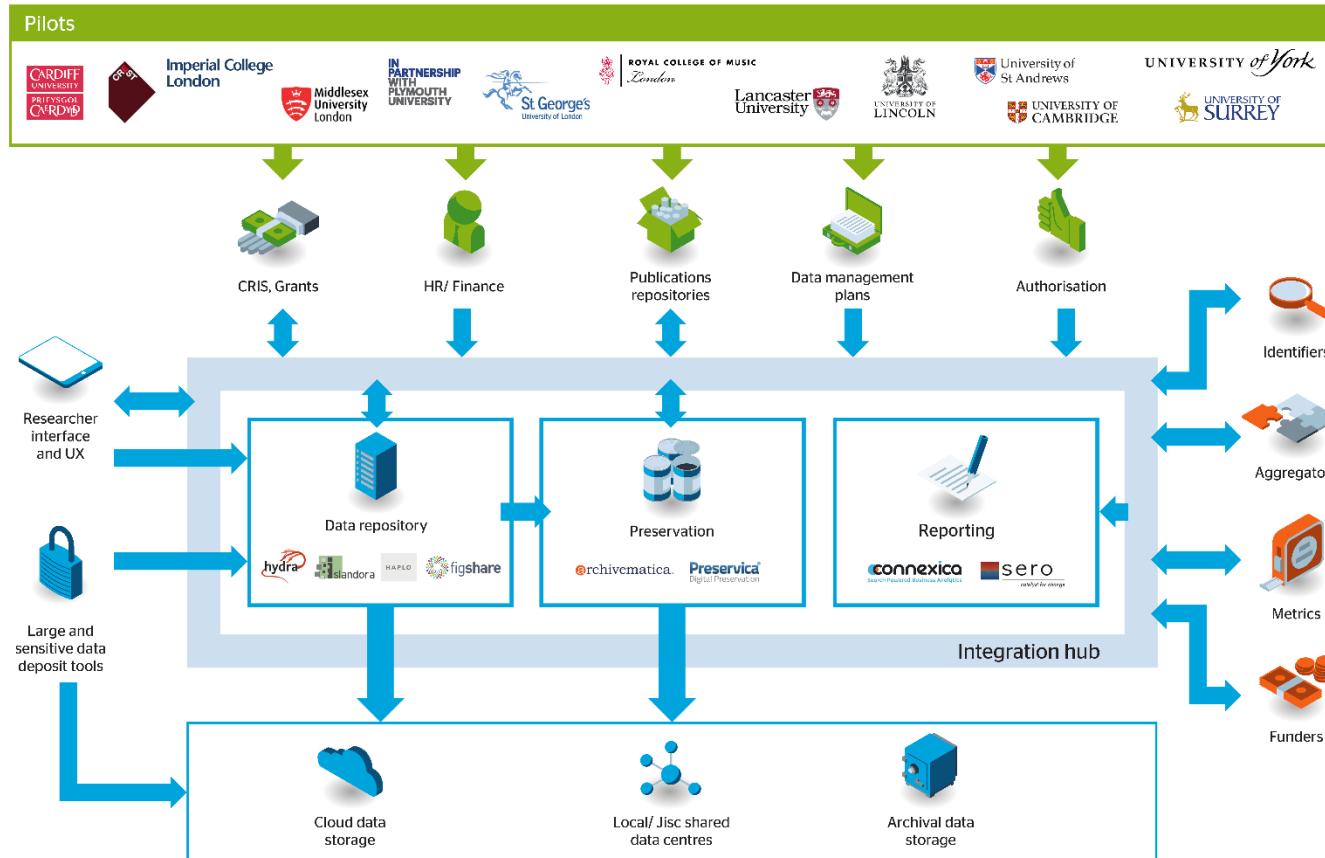
*"Please, individualise the support. Workshop are useless, emails with information are useless, brochures are useless, posters are useless."*

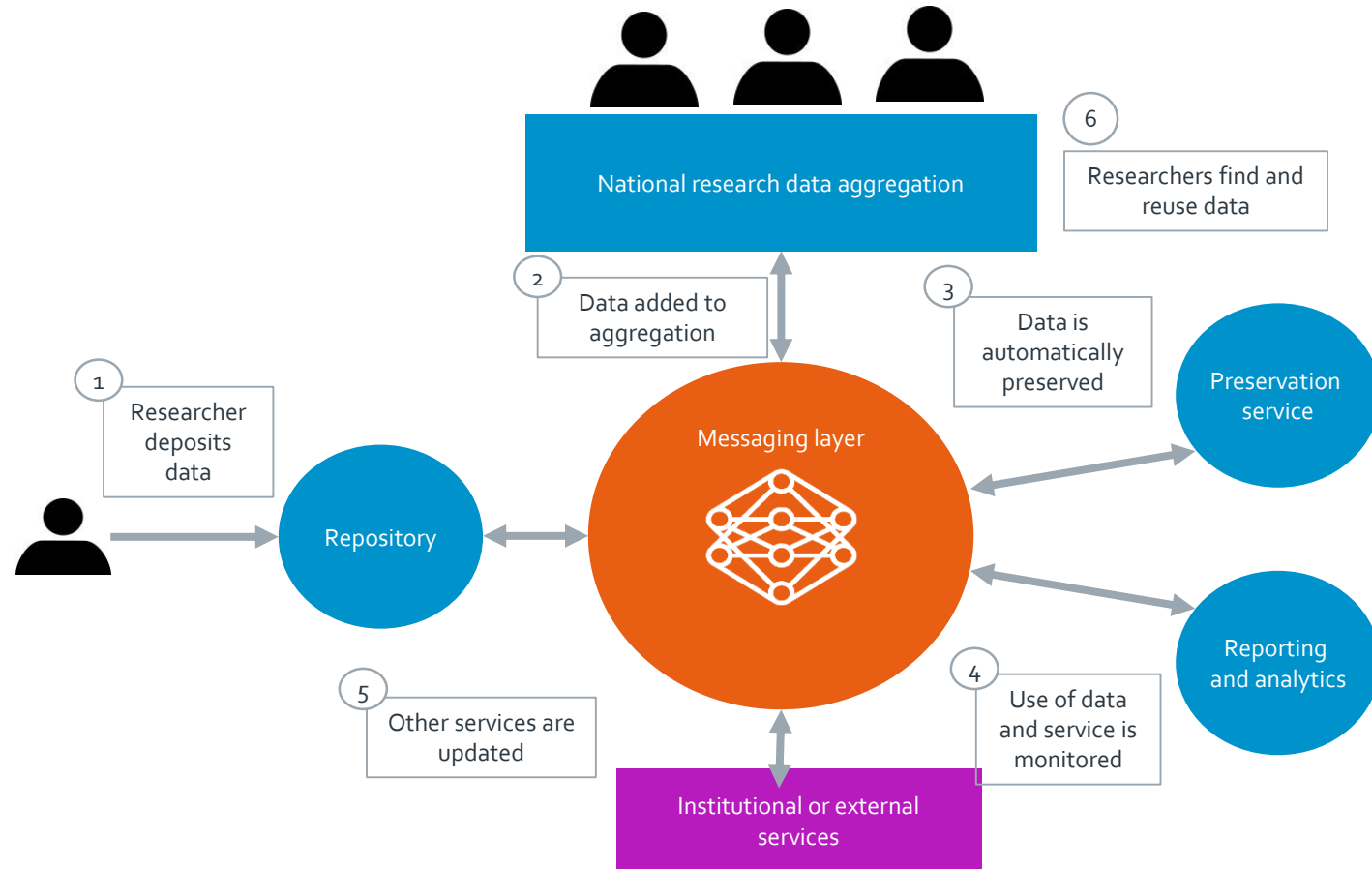
*"I currently spend about £1,200 pa on data storage from my own salary. I have the highest data needs in my School, and there is no plan in place for storing my data."*

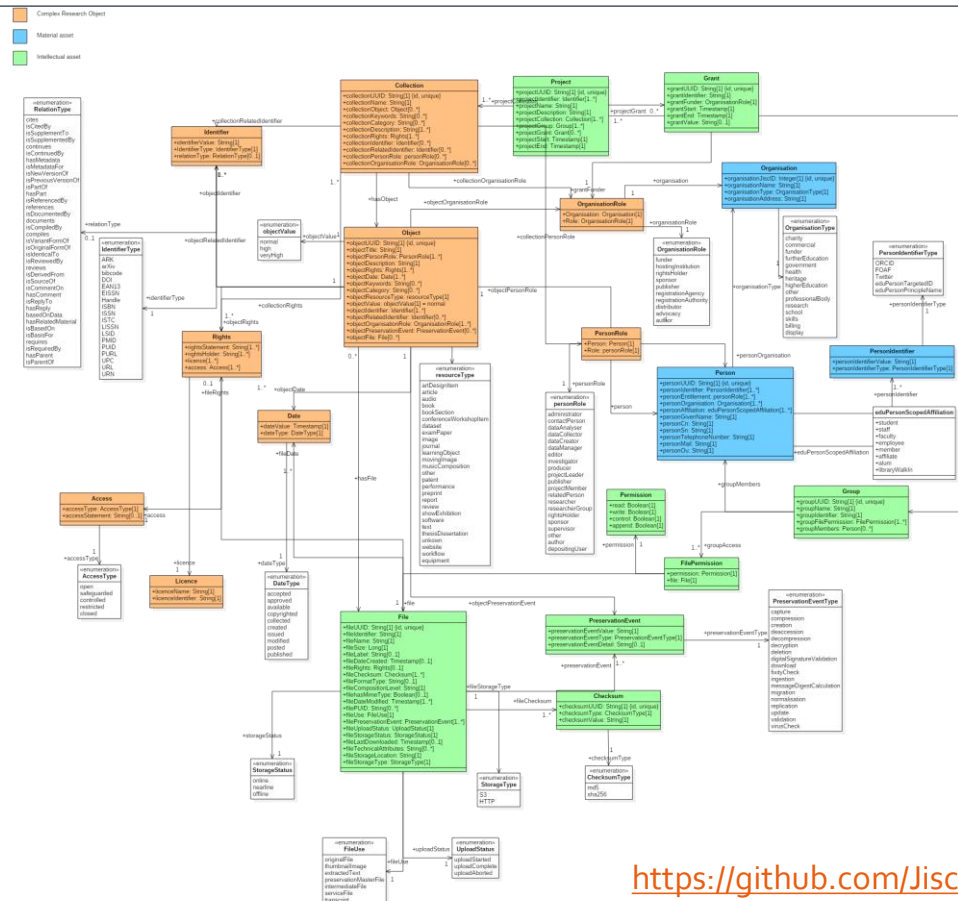


- » Diverse practices, no-one is the same
- » Pockets of good disciplinary practice, where served by existing disciplinary or institutional infrastructure
- » Generally at a central function most institutions are at the start of implementing systems and practice
- » Researcher engagement is key
- » Desired 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









<https://github.com/JiscRDSS/rdss-canonical-data-model>

## » Core Systems



## » Preservation Development

THE UNIVERSITY  
of EDINBURGH

## » Records and Archives Pilot

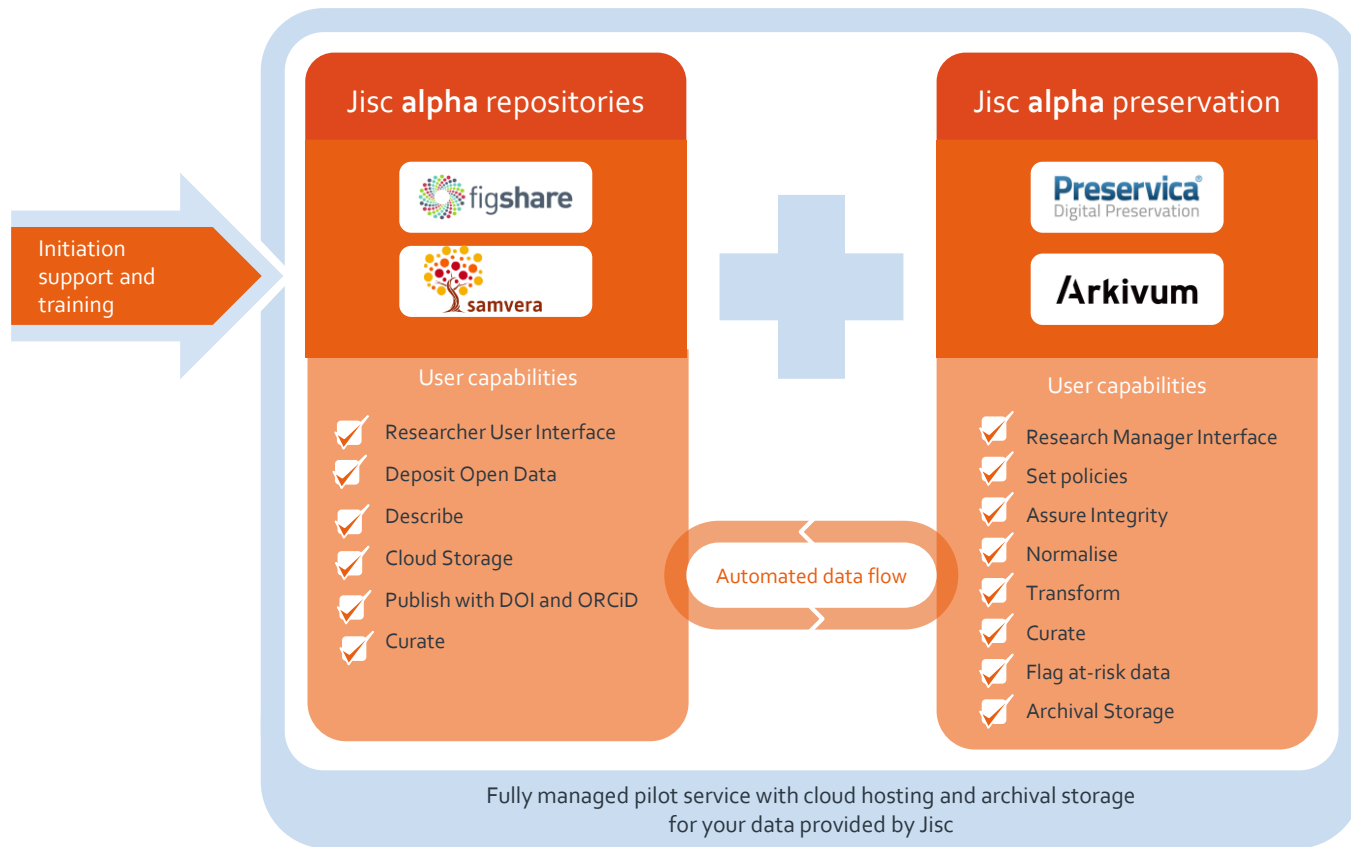
University  
of GlasgowUNIVERSITY OF  
WESTMINSTER

## » Information, Advice, Guidance and Community Support



Digital Preservation Coalition

Software  
Sustainability  
Institute



Cluster	Organisation	Status	Archivematica	Preservica	Repository	Preservation Experts Cluster	
1	Samvera and Archivematica	Cambridge	Already have access to Preservica and Archivematica. Repository to be delivered	<div></div>	<div>Samvera</div>	<div>Y</div>	
		Lancaster	Already have access to Preservica and Archivematica. Repository to be delivered	<div></div>	<div>Samvera</div>	<div>Y</div>	
		Lincoln	Already have access to Archivematica. Repository to be delivered	<div></div>	<div>Samvera</div>		
		York	Already have access to Archivematica. Repository to be delivered	<div></div>	<div>Samvera</div>	<div>Y</div>	
2	Figshare and Preservica	Middlesex	Both systems delivered - Production	<div></div>	<div>Figshare</div>		
		St George's	Both systems delivered - Production	<div></div>	<div>Figshare</div>		
		Surrey	Both systems delivered - Test	<div></div>	<div>Figshare</div>		
3	Samvera and Preservica	Cardiff	Preservica delivered, Repository to be delivered	<div></div>	<div>Samvera</div>		
		Imperial	Preservica delivered, Repository to be delivered	<div></div>	<div>Samvera</div>		
		Plymouth	Preservica delivered, Repository to be delivered	<div></div>	<div>Samvera</div>		
		Royal College of Music	Preservica delivered, Repository delivered and migrated in October	<div></div>	<div>Samvera</div>		
4	Figshare and Archivematica	CREST	St Mary's	Archivematica delivered, repository currently being implemented	<div></div>	<div>Figshare</div>	
			UCA	Archivematica delivered, repository currently being implemented	<div></div>	<div>Figshare</div>	
			Bucks New	Archivematica delivered, repository currently being implemented	<div></div>	<div>Figshare</div>	
			Winchester	Archivematica delivered, repository currently being implemented	<div></div>	<div>Figshare</div>	
5	Preservation only	Glasgow	No Repo,sitory both prservation systems deliivered	<div></div>		<div>Y</div>	
		St Andrew's	Already have access to Preservica. Archivematica and Pure interface prototype to be delivered in November	<div></div>		<div>Y</div>	

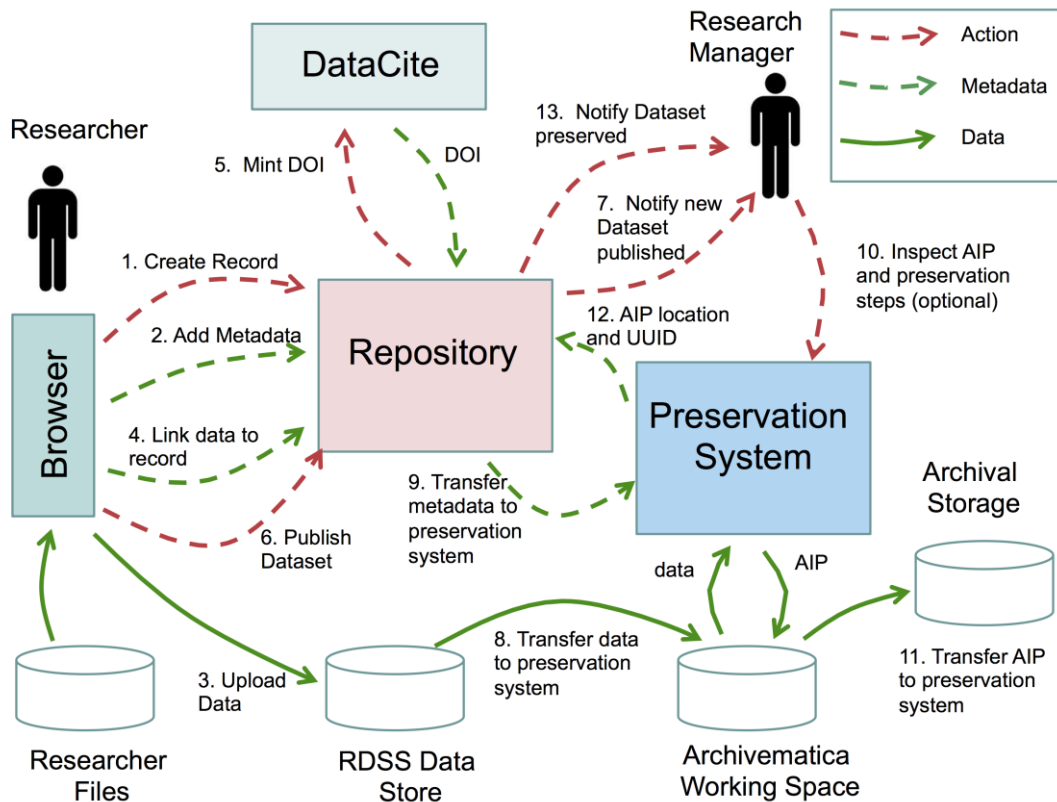


» **Automated preservation workflow:**

- › 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.





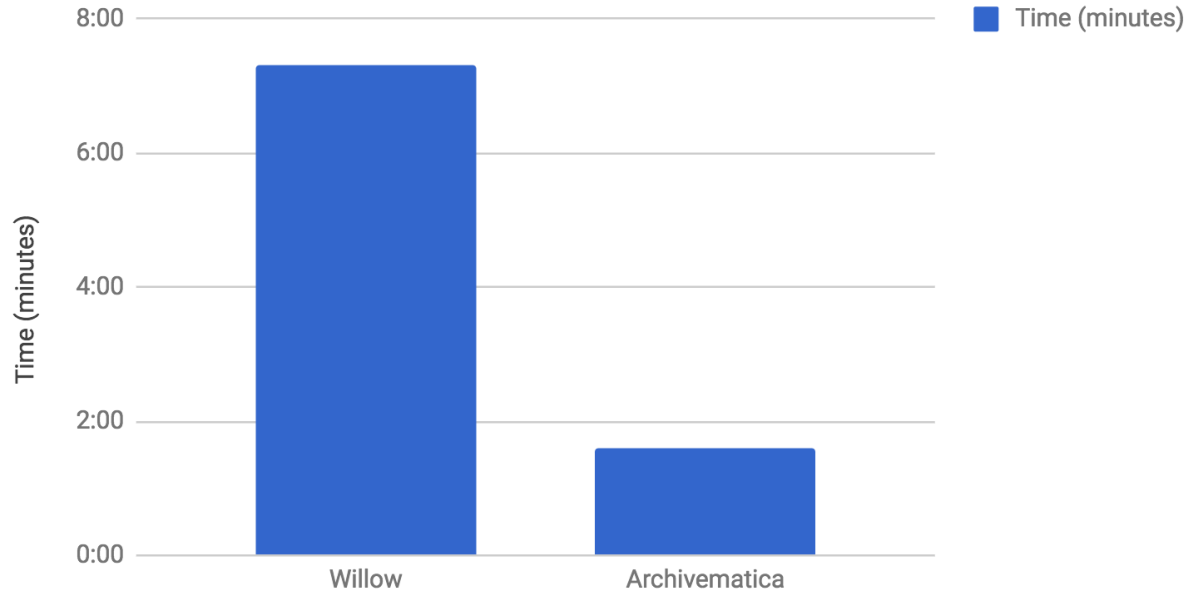
Workflow	No of Participants	Score 0 / 100	Grade
SIP Creator & Preservica Interactive	2	43.8	F "Poor"
Figshare & Preservica Automated	3	43.3	F "Poor"
Nextcloud & Archivematica Interactive	3	55	F "Poor"
<u>Samvera &amp; Archivematica</u> Automated	4	61.9	D "OK/Fair"



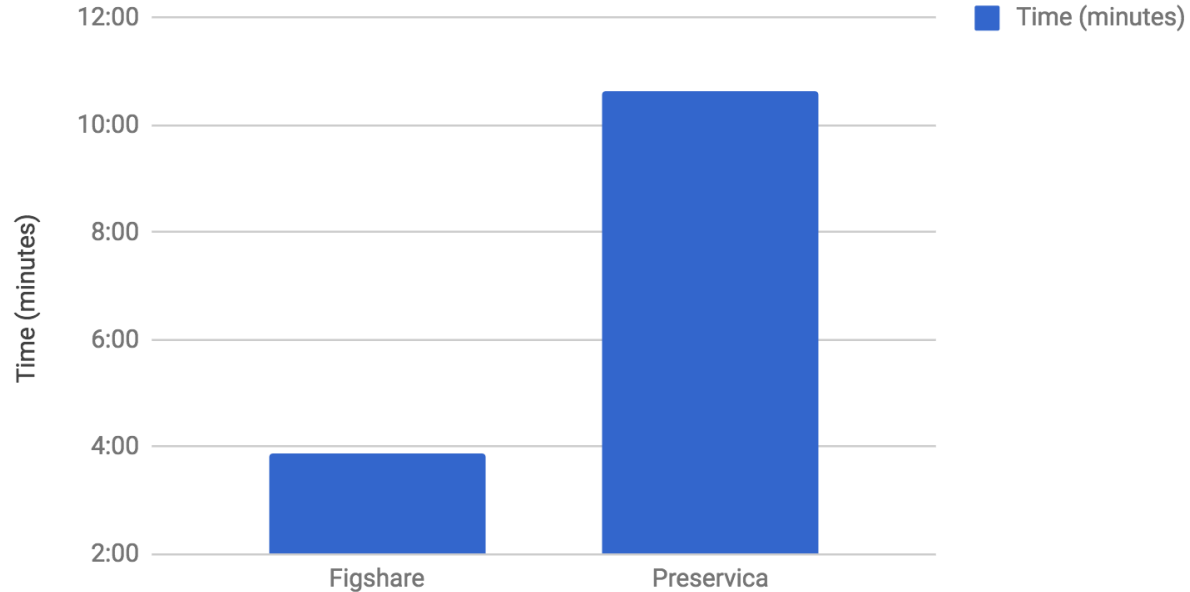
Workflow	No of Participants	Score -100 / 100	Result
SIP Creator & Preservica Interactive	2	-50	Mostly Detractors
Figshare & Preservica Automated	3	-100	Mostly Detractors
Nextcloud & Archivematica Interactive	3	-100	Mostly Detractors
<u>Samvera</u> & <u>Archivematica</u> Automated	4	-75	Mostly Detractors

- » 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

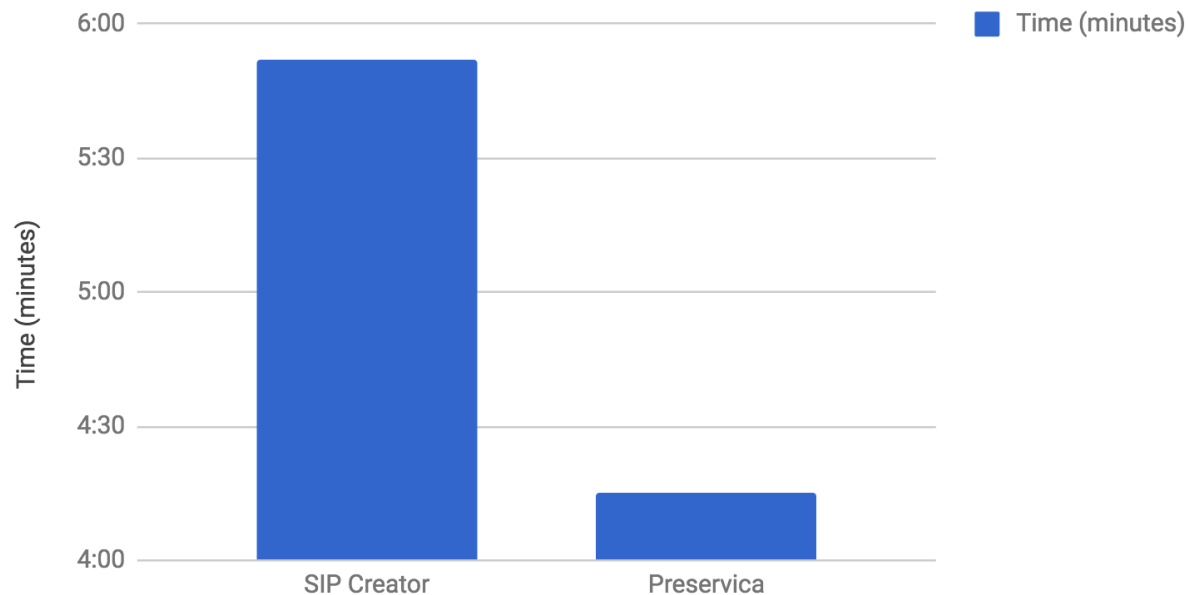
Average time spent on each technology, if successful, for the automated Willow & Archivemática test



Average time spent on each technology, if successful, for the automated Figshare & Preservica Test

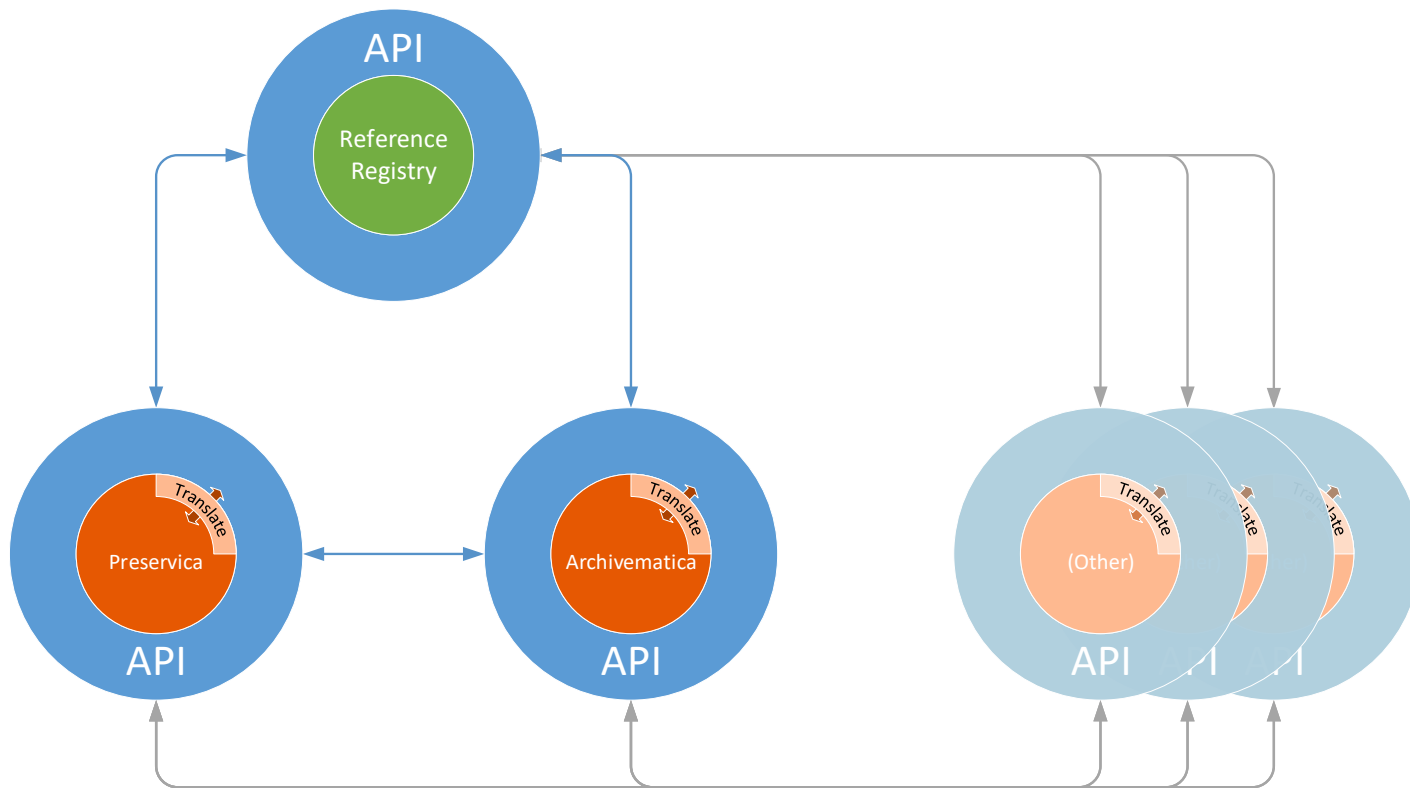


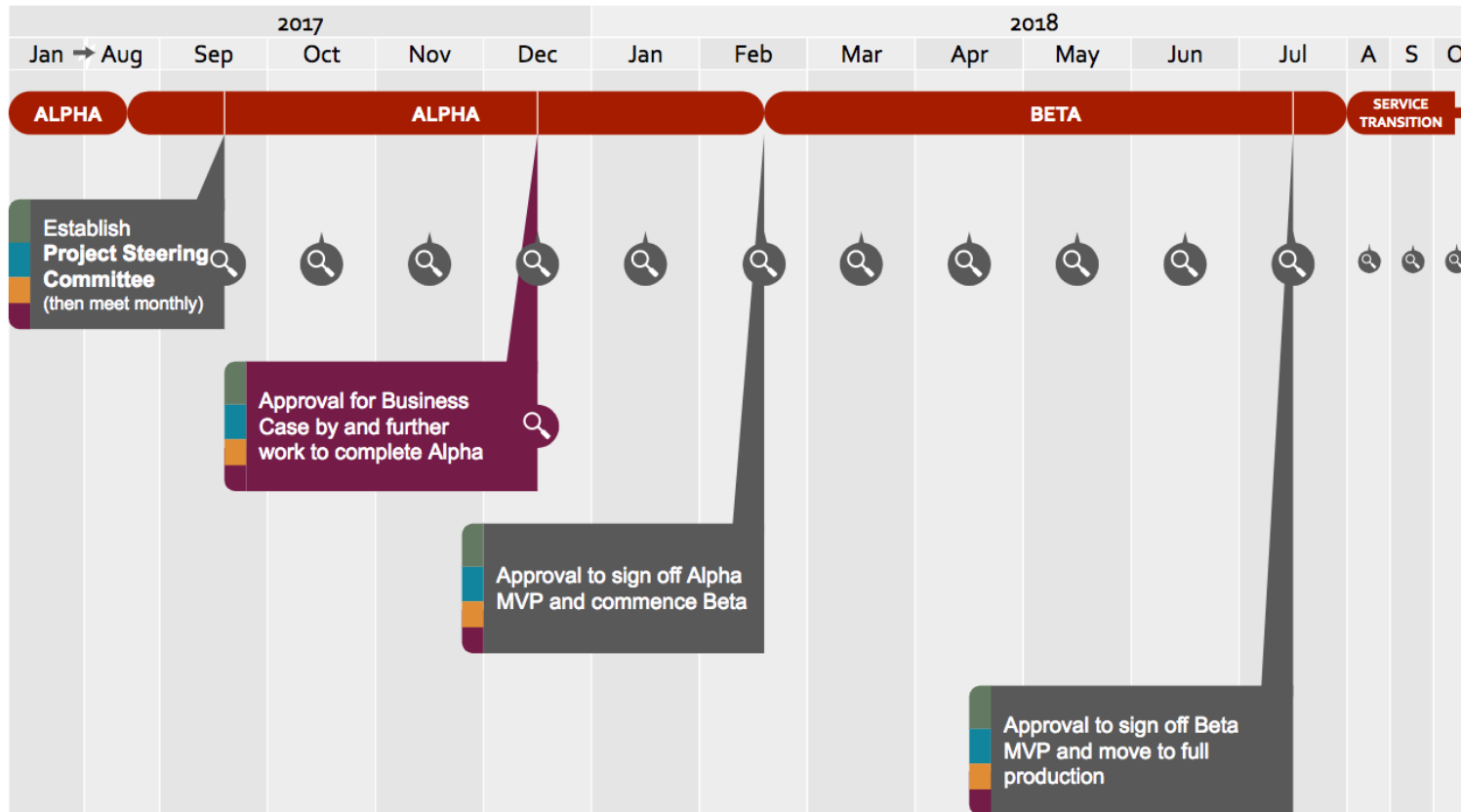
Average time spent on each technology, if successful, for the interactive SIP Creator and Preservica Test





- » 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





End-to-end  
service



Repository  
service



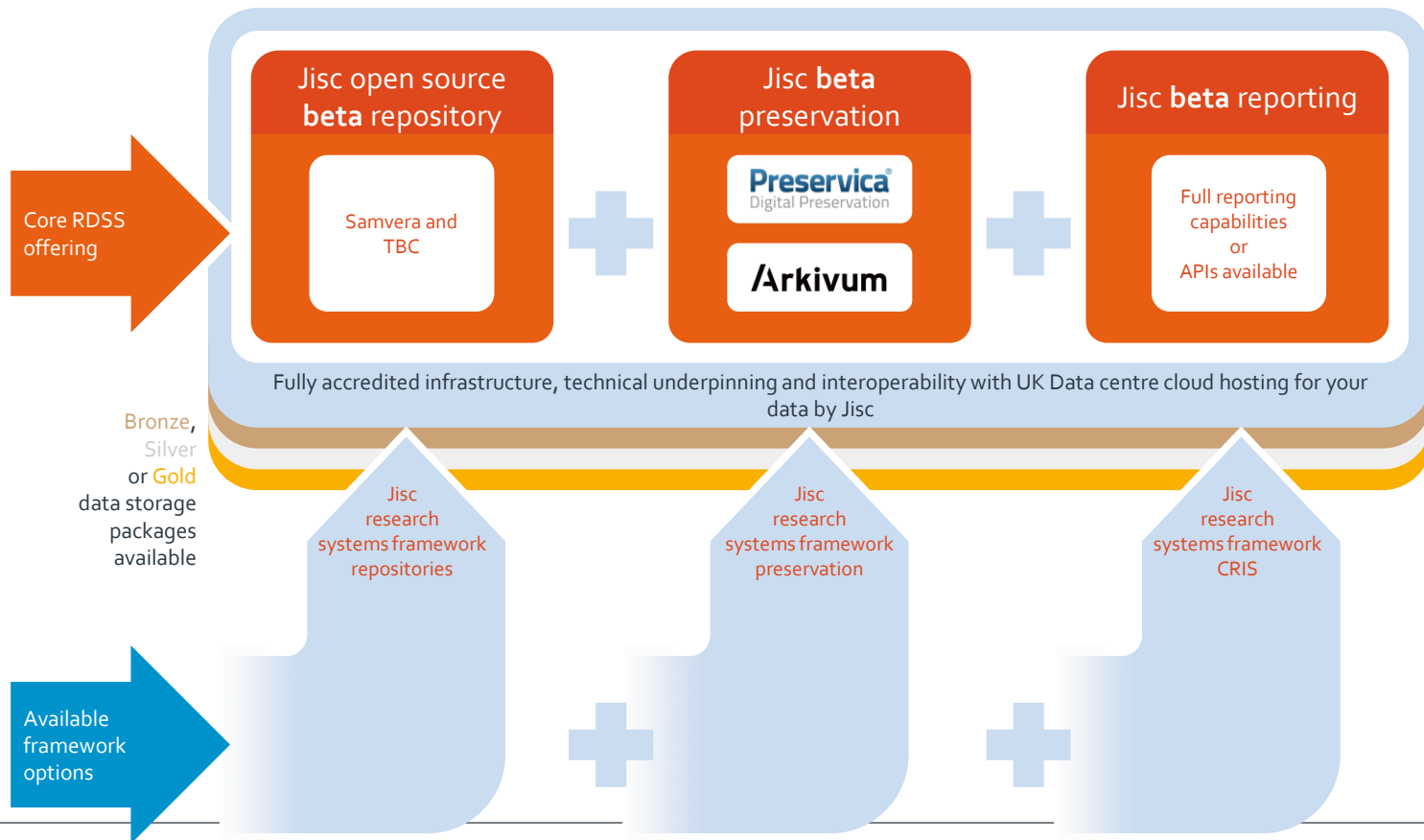
Preservation  
service



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



RDSS Potential Integrations	Number of Pilots (total =17)
Eprints (Repository)	12
Dspace (Repository)	4
Samvera (Repository)	2
Symplectic (CRIS)	4
Pure (CRIS)	3
Converis (CRIS)	1
Authentication - UKAMF	17

- » Jisc or 3<sup>rd</sup> 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).



- » 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



- » Integration with Jisc Safe Share
- » Integration with Jisc Tiered Storage
- » Integrations with additional 3<sup>rd</sup> Party integrators
- » Integration with updated DMP Online
- » Integrations with active data storage e.g. Institutional storage area network, file sync and share services.





---

**.....don't fit local budgets**



- » 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!



**John Kaye, Head of Change, Research**

[john.kaye@jisc.ac.uk](mailto:john.kaye@jisc.ac.uk)

**Dom Fripp – Product owner, Samvera**

[paul.stokes@jisc.ac.uk](mailto:paul.stokes@jisc.ac.uk)

**Paul Stokes, Senior Co-Design Manager**

[paul.stokes@jisc.ac.uk](mailto:paul.stokes@jisc.ac.uk)

**Alan Mackenzie, Chief Architect**

[alan.mackenzie@digirati.com](mailto:alan.mackenzie@digirati.com)



Except where otherwise noted, this work is licensed under CC-BY-NC-ND