

13/09/2017

# Research Data Shared Service

The drivers for change

Jisc is the UK **higher, further education** and **skills** sectors' **not-for-profit** organisation for **digital services** and **solutions**

Jisc

We...

Provide trusted **advice** and **practical assistance** for universities, colleges and learning providers

Negotiate **sector-wide deals** with IT vendors and commercial publishers

Operate **shared digital infrastructure** and **services**



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>



EPSRC and RCUK have both set out clear expectations that Institutions must take responsibility for Research Data Management, Preservation and Sharing

<https://www.epsrc.ac.uk/about/standards/researchdata/expectations/>

Costs & risks	Risk to research funding	Reduced income	Consequential loss
	Loss of research data	Lost value of research work (17% lost key data – DAF survey)	
	Leakage of sensitive data	Legal threat and cost (Unlimited fines with GDPR)	
	Researcher reputation	Key staff leave (75% EXPECT HEI to do this)	
	Institution reputation	Defensible integrity of research, responding to FOI etc. (e.g. Climategate £10m?)	
	Cost and risk of delivery	Inefficient research and over-expensive IT	

## Today' cottage industry

c.200 UK public sector institutions perform around £6b of funded research per year in public purse every year

No sector-wide visibility of RDM effectiveness

Only 20% of research data from the 1990s is still usable\*

Majority of research institutions want a solution which cuts the pain

Stand-alone solutions need every institution to manage and maintain on limited budgets

**Research Funders require data to be managed and EPSRC have put onus on institution to preserve and make research data available as a condition of funding**

## Jisc shared service

**Share and optimise cost and benefit across the sector**

**Grow value of research data and increase research productivity**

**A single, cost-effective, pre-integrated solution available to all**

\*<http://dx.doi.org/10.1016/j.cub.2013.11.014>

## 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

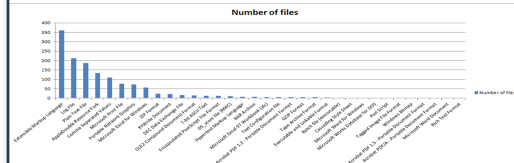


\* Droid is a free tool from The National Archives that can be used to automatically identify file formats

## Identified research data files

Files identified by Droid (listed by file type)

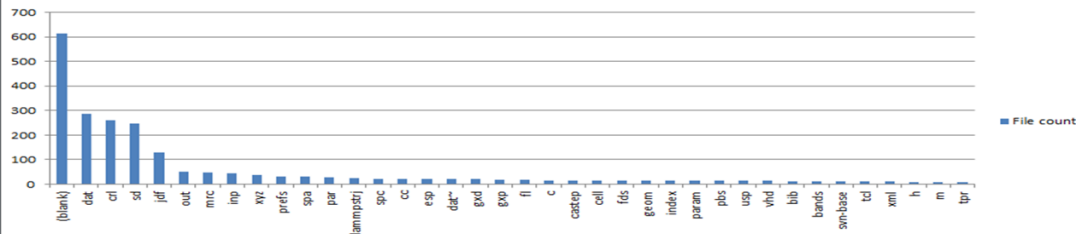
...note that native files of the software in the previous graph of research data applications are not represented



## 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



## 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:

Advice &  
best practise

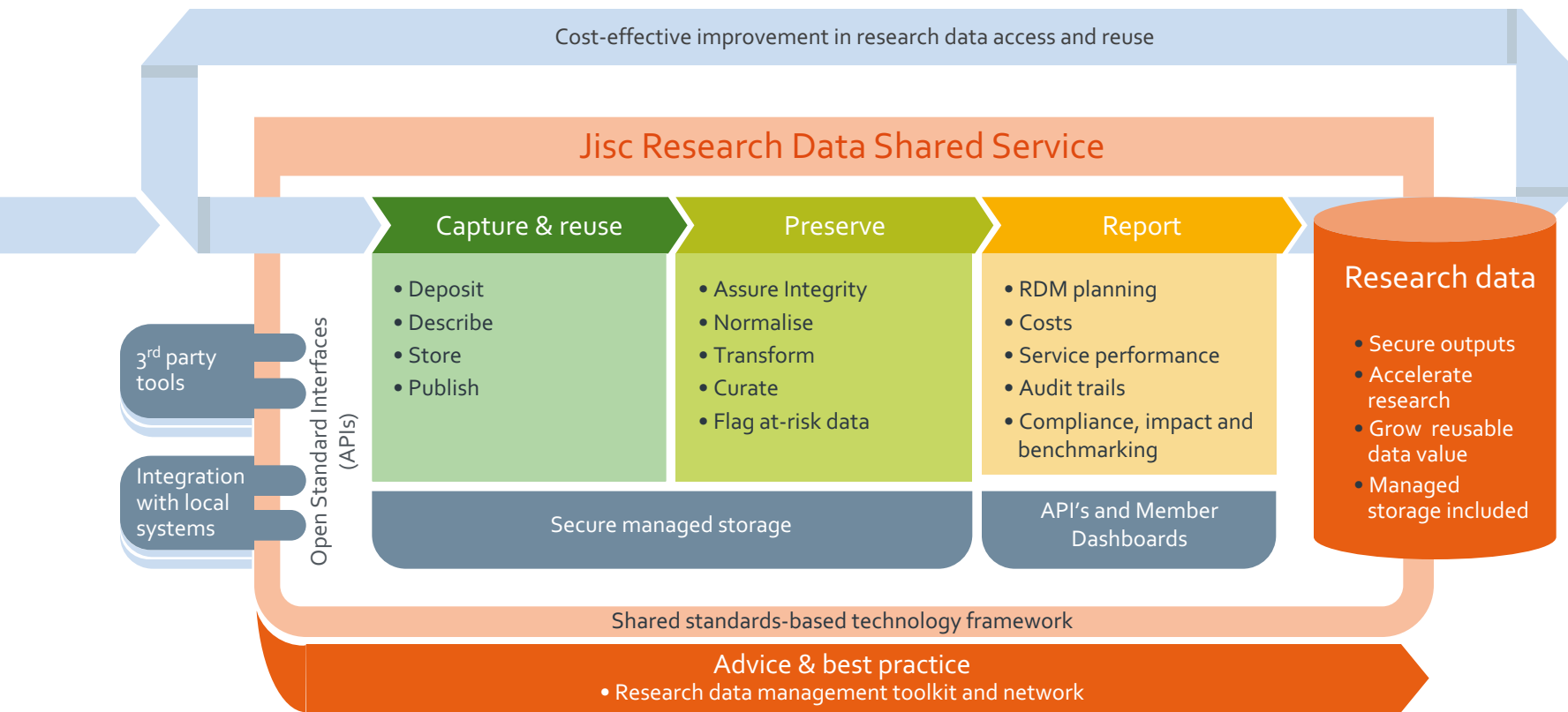
Capture & reuse

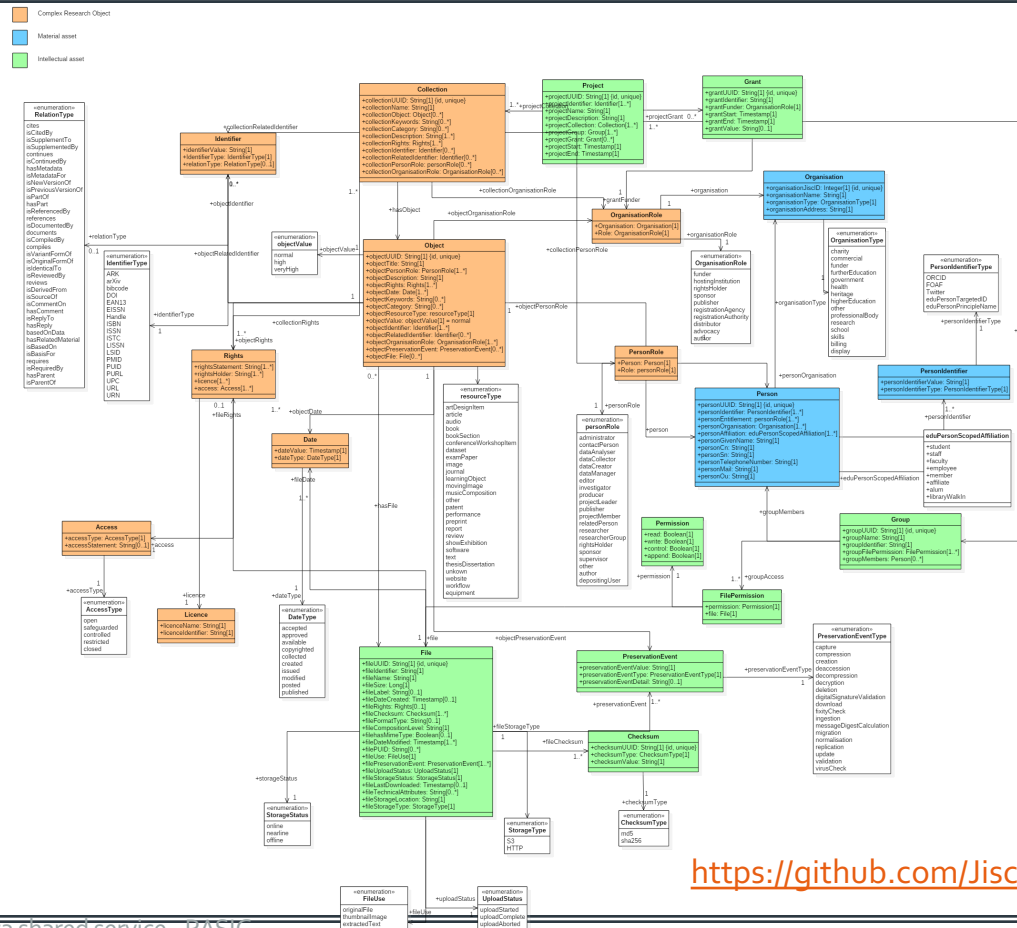
Preserve

Report

Source: Jisc DAF Survey results 2016







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

## » Core Systems



## » Preservation Development



THE UNIVERSITY  
of EDINBURGH



## » Records and Archives Pilot



University  
of Glasgow

UNIVERSITY OF  
WESTMINSTER

## » Information, Advice, Guidance and Community Support

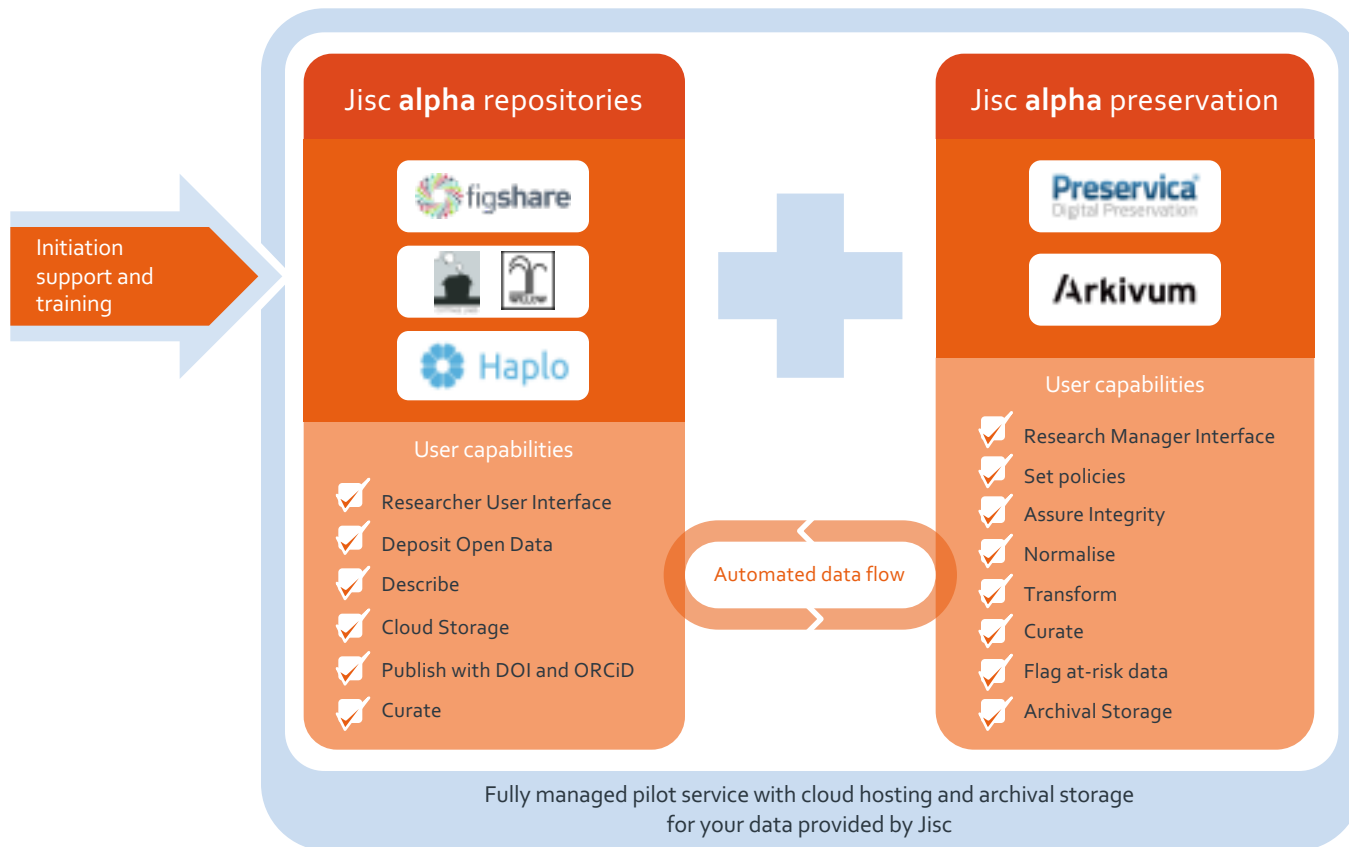


Digital **Preservation** Coalition

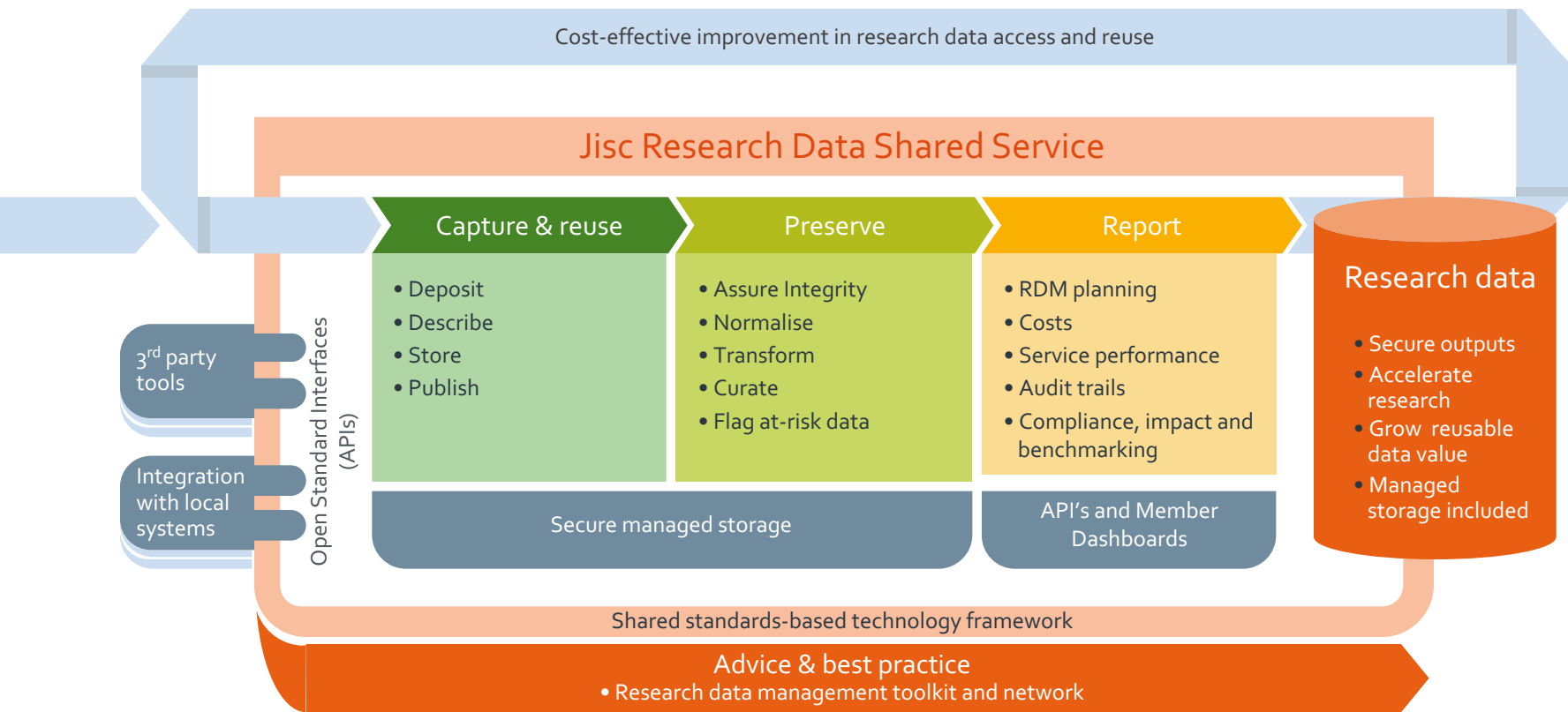


Software  
Sustainability  
Institute





Cluster	Organisation	Status	Archivematica	Preservica	Repository	Preservation Cluster
1	Willow (Samvera) and Archivematica + 2 Preservica	Cambridge	Already have access to Preservica. Archivematica and Repository to be delivered in September		Willow	
		Lancaster	Already have access to Preservica. Archivematica and Repository to be delivered in September		Willow	
		Lincoln	Archivematica and Repository to be delivered in September		Willow	
		York	Archivematica and Repository to be delivered in September		Willow	
2	Figshare and Preservica	Middlesex	Both systems delivered		Figshare	
		St George's	Both systems delivered		Figshare	
		Surrey	Both systems delivered		Figshare	
3	Haplo and Preservica	Cardiff	Preservica delivered Haplo to be delivered in September		Haplo	
		Imperial	Preservica delivered Haplo to be delivered in September		Haplo	
4	Willow and Preservica	Royal College of Music	Already have access to Preservica. Repository to be delivered in September		Willow	
5	Figshare and Archivematica	CREST	St Mary's	Figshare and Archivematica to be delivered in September	Figshare	
			UCA	Figshare and Archivematica to be delivered in September	Figshare	
			Bucks New	Figshare and Archivematica to be delivered in September	Figshare	
			Winchester	Figshare and Archivematica to be delivered in September	Figshare	
6	Preservation only	Glasgow	No Repo, Just Preservation, contract to be signed			
		St Andrew's	Already have access to Preservica. Archivematica to be delivered in September			
		Plymouth	Preservica Delivered. Repository to be Procured			



End-to-end  
service



Repository  
service



Preservation  
service

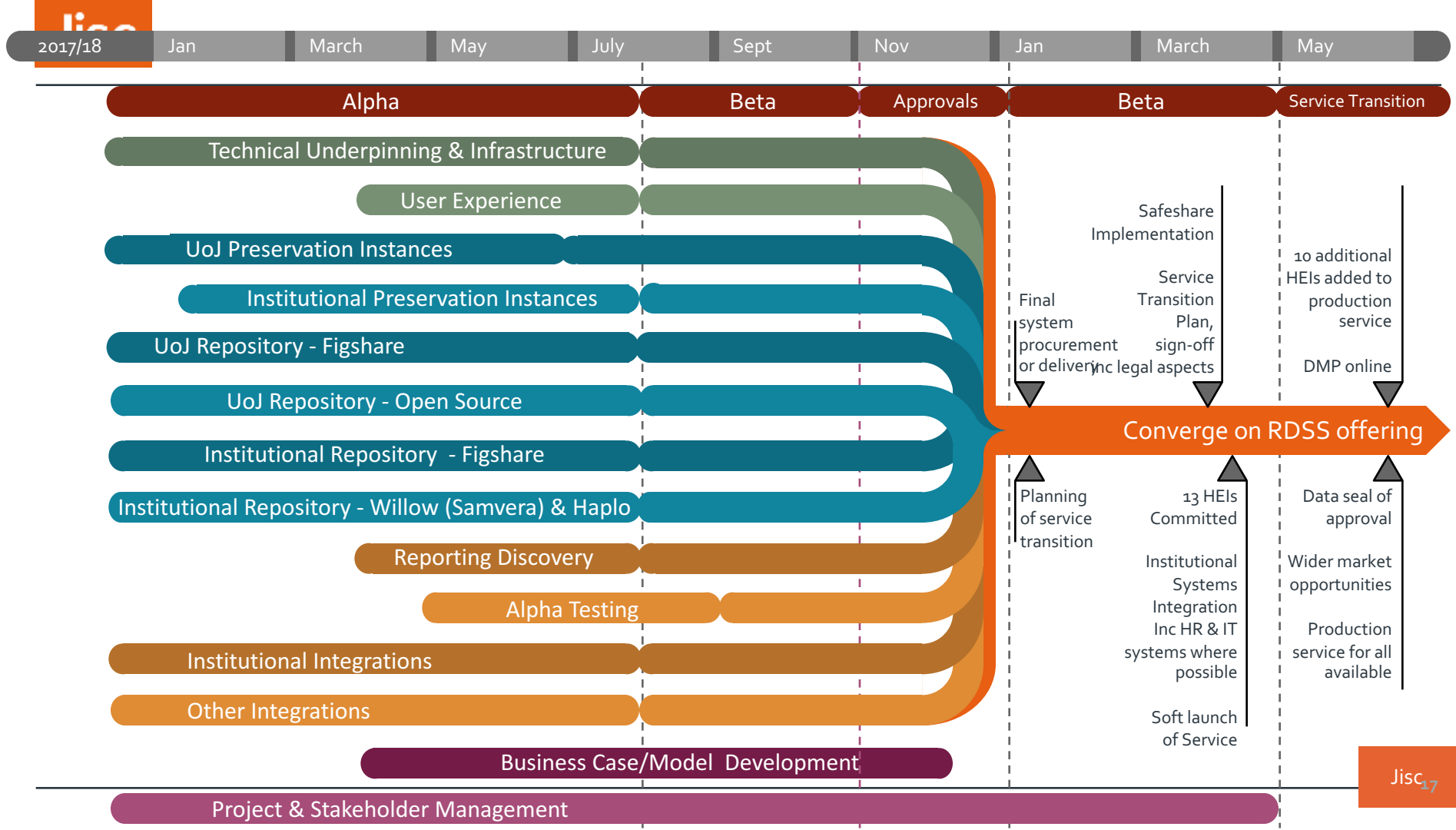


Approach supports pilot investments in a range of components, direction of travel towards standard managed solution for all over time.

All 3 options include:

- ✓ Financial benefits
- ✓ Standards
- ✓ Advisory
- ✓ Network membership





**John Kaye**

john.kaye@jisc.ac.uk @johnpkaye



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