

RISING TO THE IMPOSSIBLE CHALLENGE

THE STORAGE OF ALL RESEARCH DATA....

....FOREVER??

Tim Banks, University of Leeds
Wendy White, University of Southampton
Matthew Addis, Arkivum

Arkivum
Every bit archived

UNIVERSITY OF
Southampton UNIVERSITY OF LEEDS



Photo Credit: [PNNL - Pacific Northwest National Laboratory](#) via [Compfight cc](#)

Data Repository



Repository
Platform



Disk
Cache



Tape
Archive

EDPaaS

ARCHIVE IN THE CLOUD?



- Efficient at scale
- Flexible
- POSF* cost recovery
- Little local effort required
- No large capital investment needed
- Data escrow
- Finance escrow (for POSF)
- Purchase Tax (e.g. VAT)
- Bandwidth
- Data security
- Ingress/egress charges
- Storage location

*Pay once, store forever (25yrs)

COSTS OF LOCAL TAPE LIBRARY

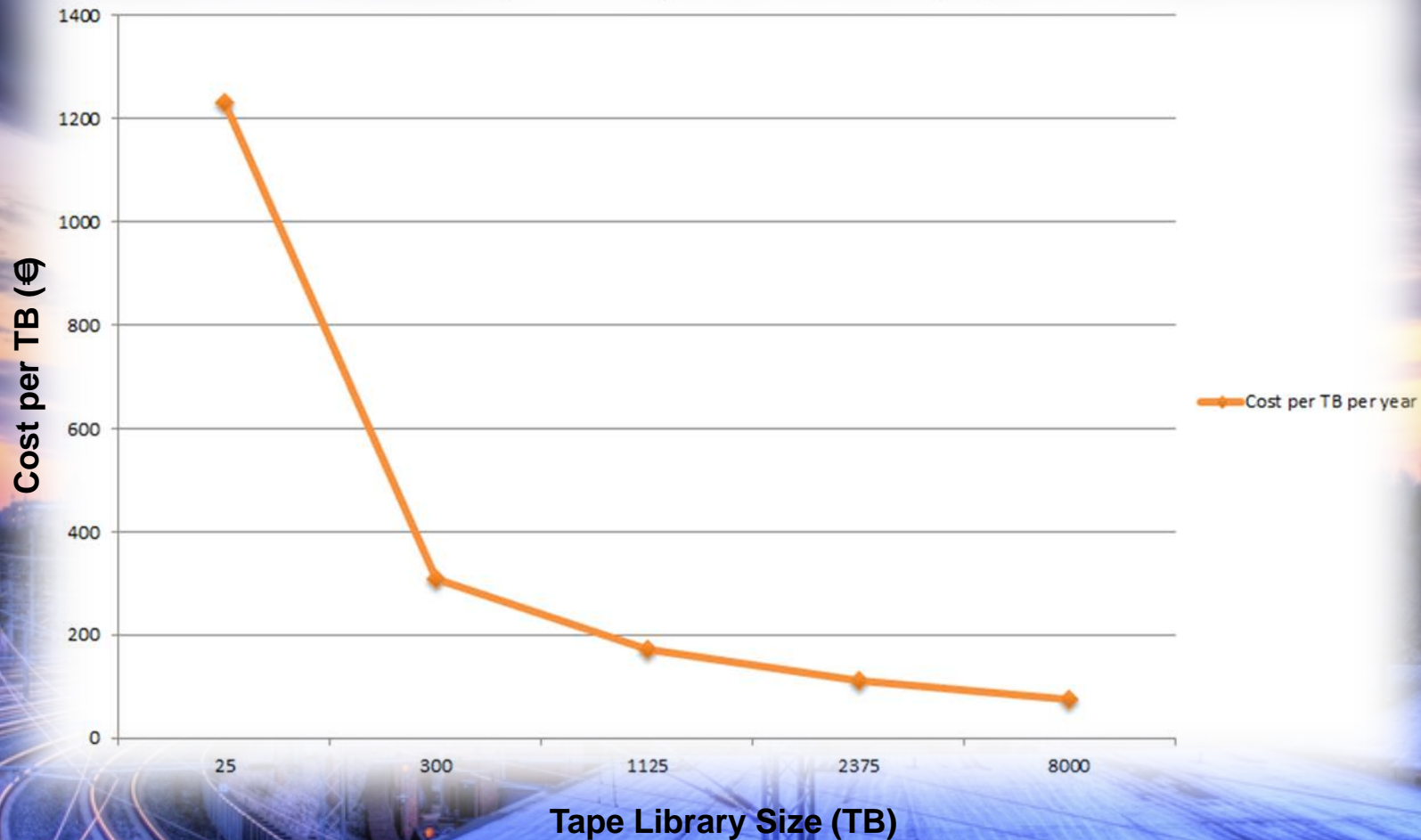
Item

Cost

Base frame inc. drives, robot etc.	€146,950
5yrs support for base frame	€ 55,250
Professional Services (installation)	€ 8,250
LTO-6 media inc. cleaning tapes	€ 32,750
120TB Spinning Disc cache	€ 50,000
5yrs support for disc cache	€ 22,250
QStar Archive manager software	€215,100
Infrastructure (fibre channel switches)	€ 31,500
1x full time archive manager post (5yrs)	€308,000
Machine Room costs; power, cooling (5yrs)	€123,200
TOTAL	€993,250
(€176.50 per TB per year for single copy)	

1.125PB raw; 450x LTO6; 4x drives; 1x robot; archive management software; 120TB disc cache

Cost per TB per year for tape libraries of varying size



The Data Archive Cloud Model

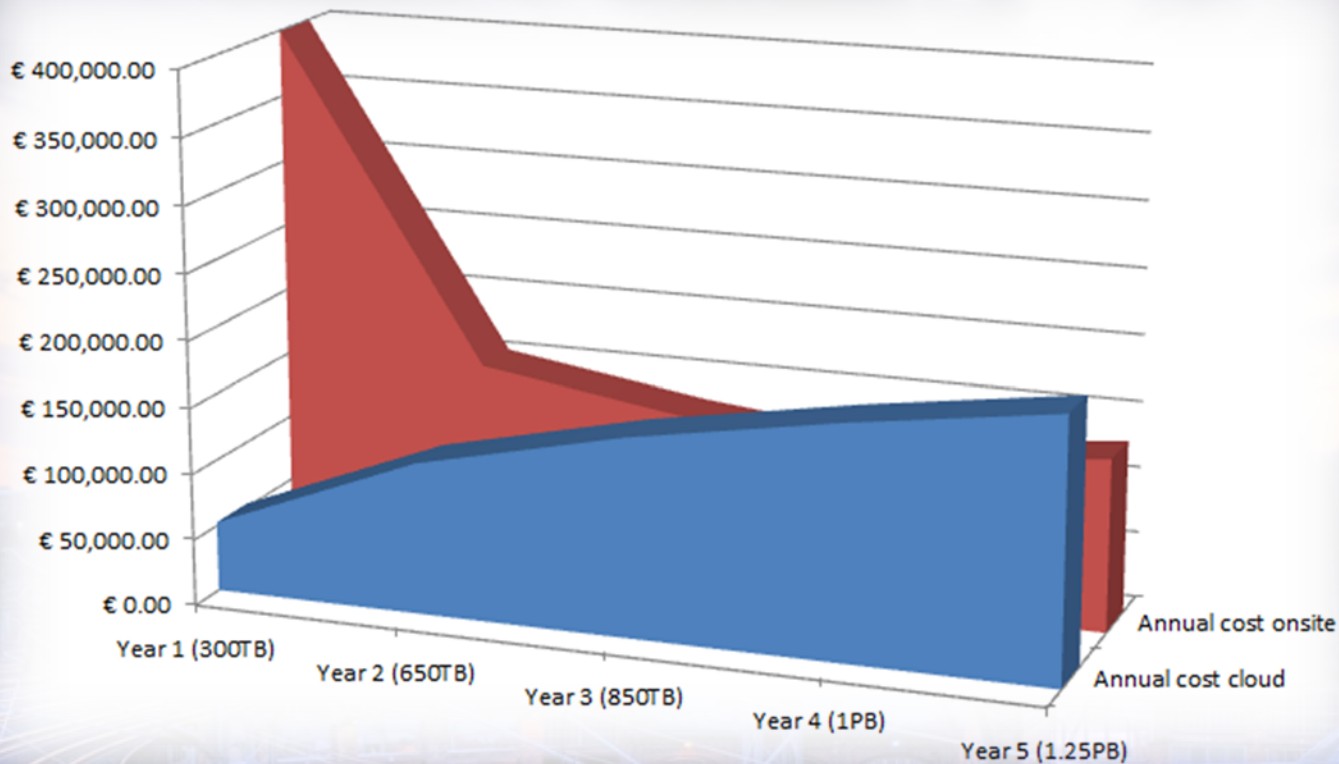
***Data
Archive
as a
Service***

Submission of
data for ingest



Request data for
retrieval

Cost profile for on-site vs. cloud data archive

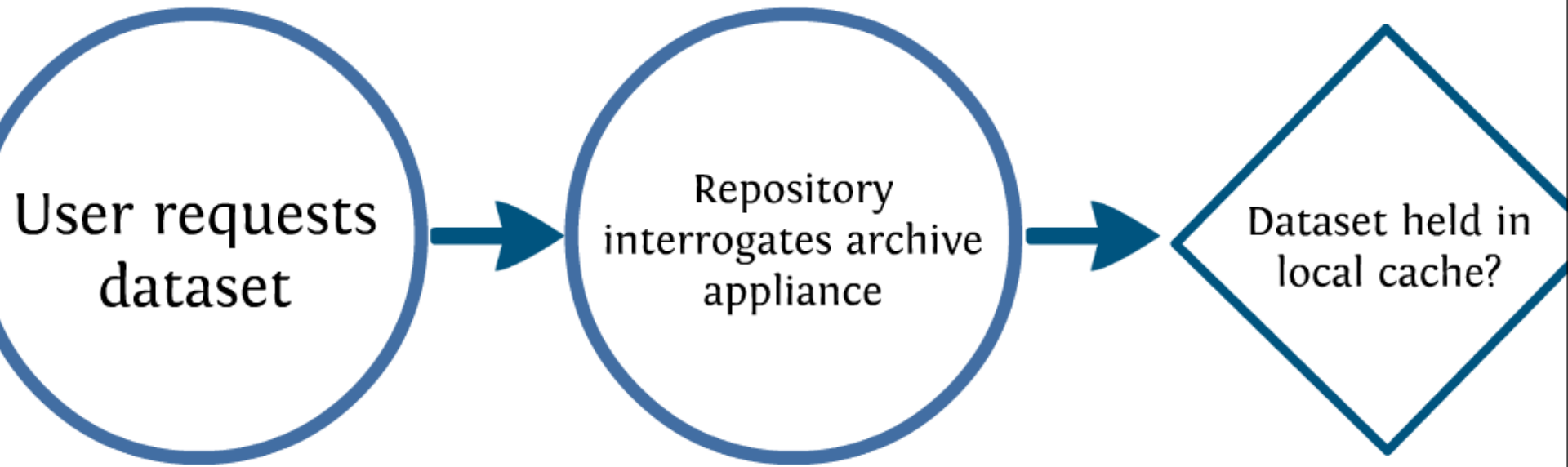


For same annualised cost per TB per year, with this storage profile, cloud service is **26% less expensive**

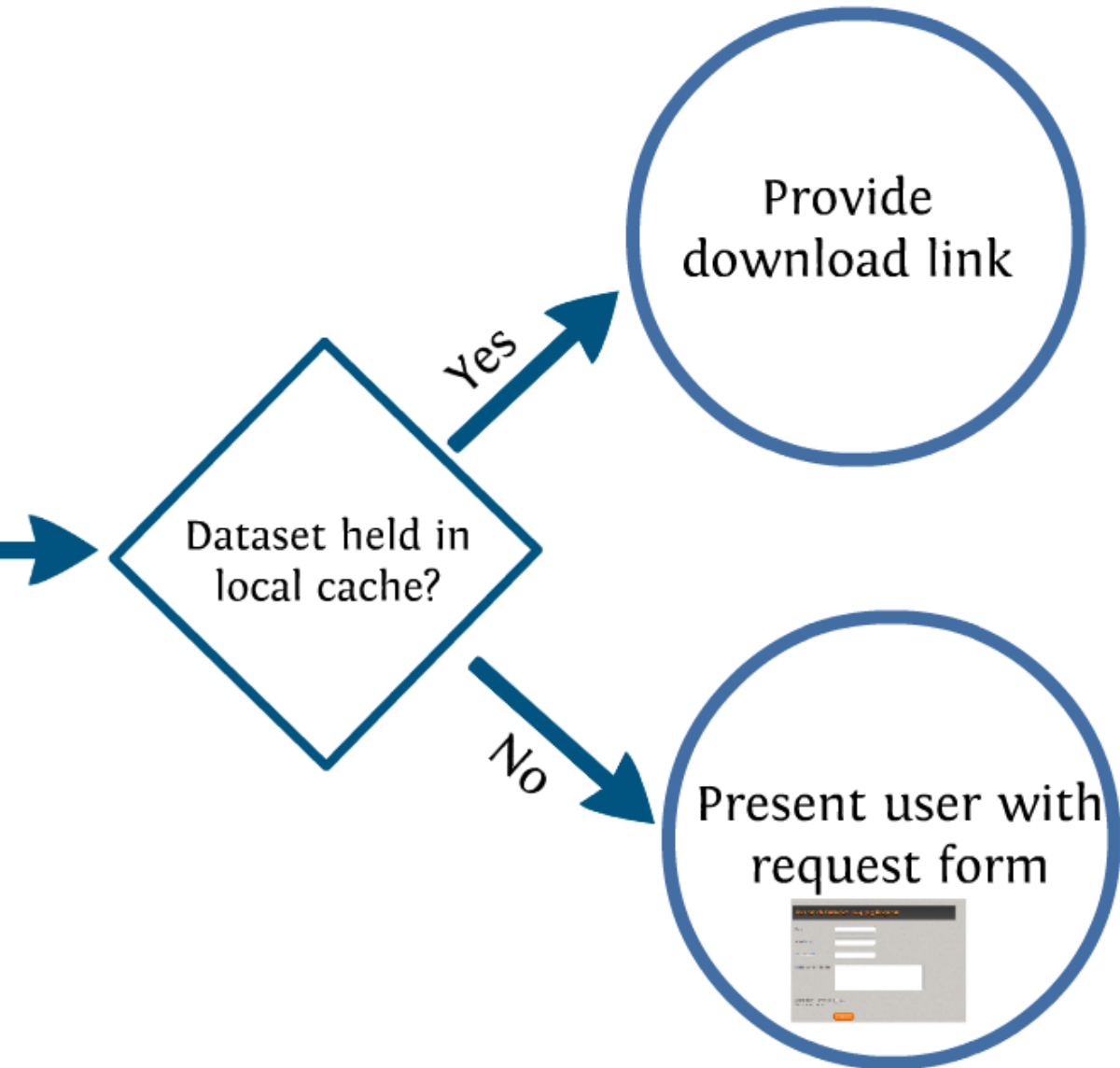
Assumptions:

- Archive manager is employed full time for 5-years
- Any fixed maintenance costs are paid up-front for 5-years

Integration with Repository (1)



Integration with Repository (2)



Integration with Repository (3)

Research Data Set 1104423 Request

Name

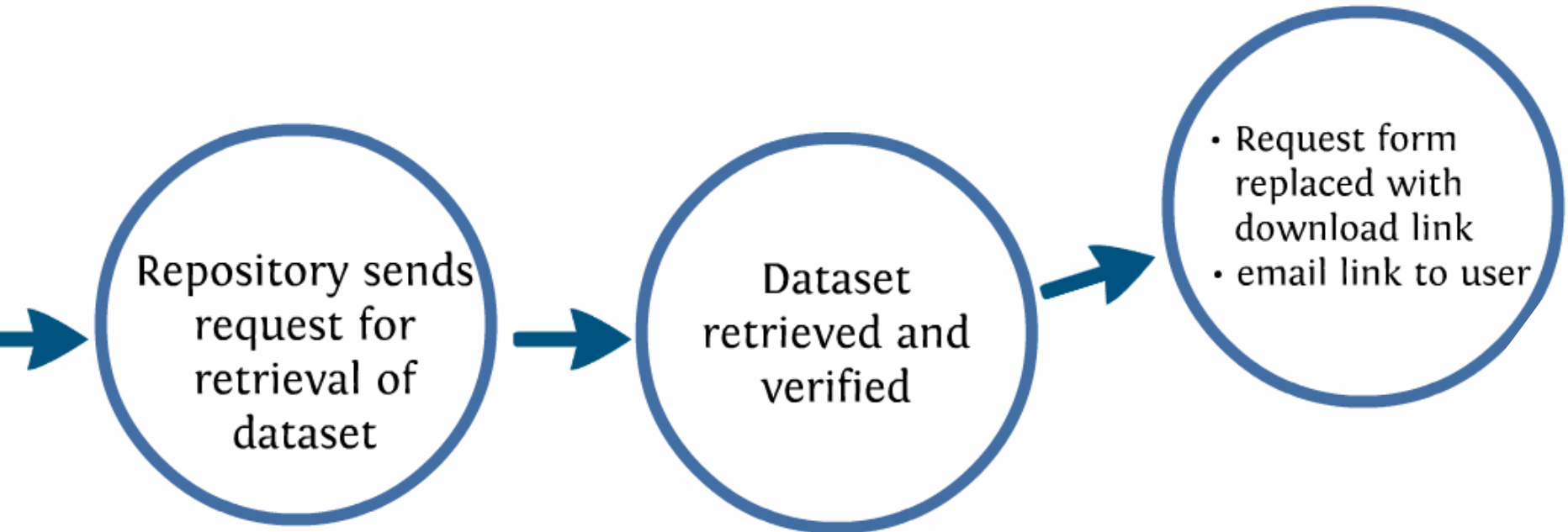
Organisation

e-mail address

Reason why data required

I agree to the Terms and Conditions of access I agree

Integration with Repository (4)



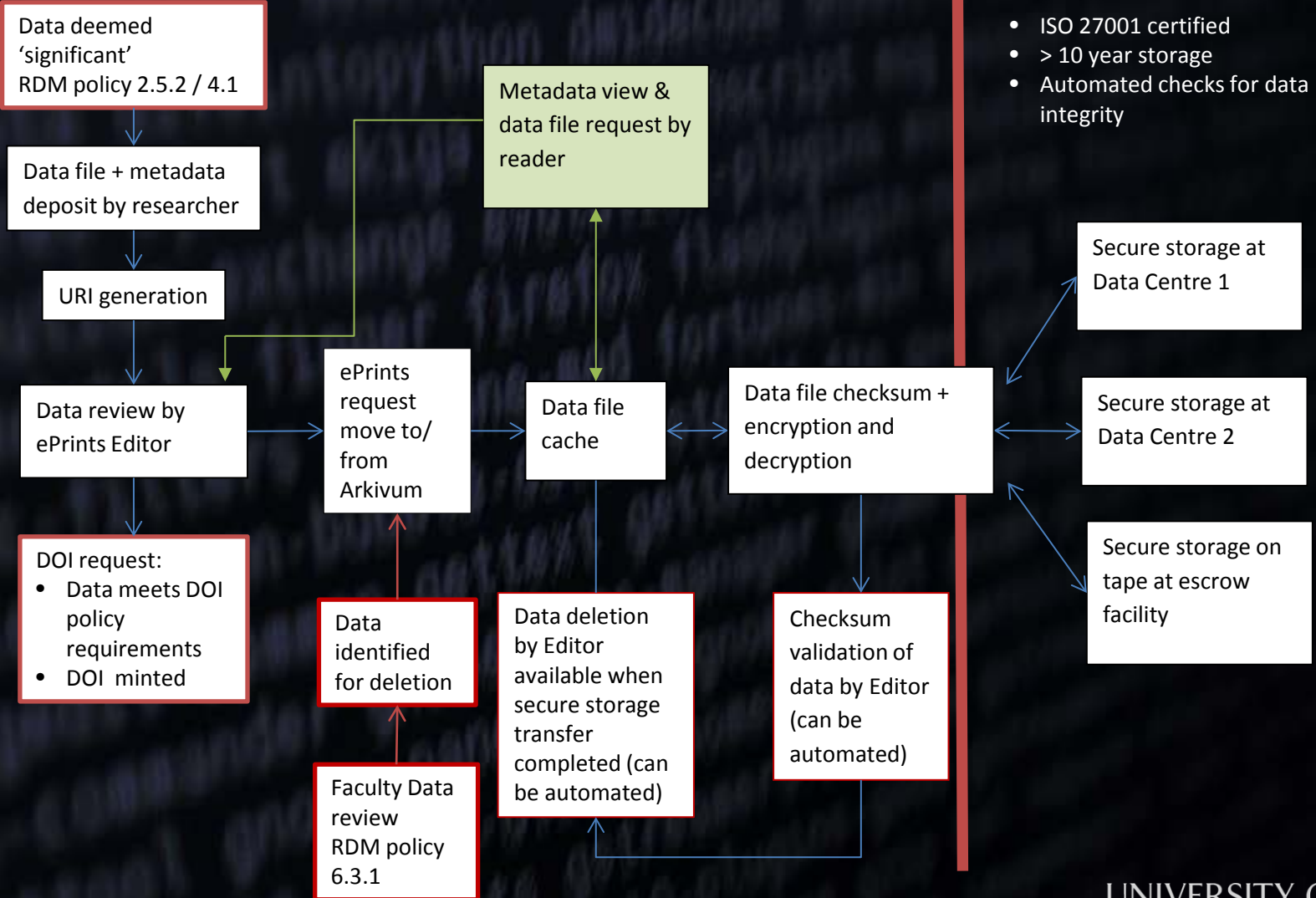
ePrints

Arkivum

Data Deposit

Data Request

Data Storage



Final Considerations

- It is important to fully understand the real costs of running a local archiving service
- Consider where your data is stored (Data Protection)
- Understand the full costs of cloud archiving service (including ingress and egress costs and limitations).
- Look at your likely 5-year archive volume profile
- Understand the protection level that you require
- Have an exit strategy