

Integration of archiving into RDM infrastructure

Matthew Addis

Arkivum

IDCC 2015

Overview

- Intro
- EPrints/Arkivum demo
- Archivematica/Arkivum demo
- Summary

All about what we've done, what we found in practice

Arkivum in less than 60 seconds



SLA with 100% data integrity guaranteed



World-wide professional indemnity insurance



Long term contracts for enterprise data archiving



Fully automated and managed solution



Audited and certified to ISO27001



Data escrow, exit plan, no lock-in

Example Customers



Data Archiving - Integrations



D SPACE

@archivematica®



figshare
credit for **all** your research

SYMPLECTIC
Elements

Three common themes

- Ensuring the right stuff makes it into the archive
 - Complete, correct, quality, chain of custody
- Managed access to archived data
 - User experience, monitoring and control
- Simple to use for researchers
 - Easy to use, one place to go, automation

Archiving



Long-term safety and security

- Data reuse or compliance
- Data is immutable, replicated, managed
- Auditable integrity and authenticity



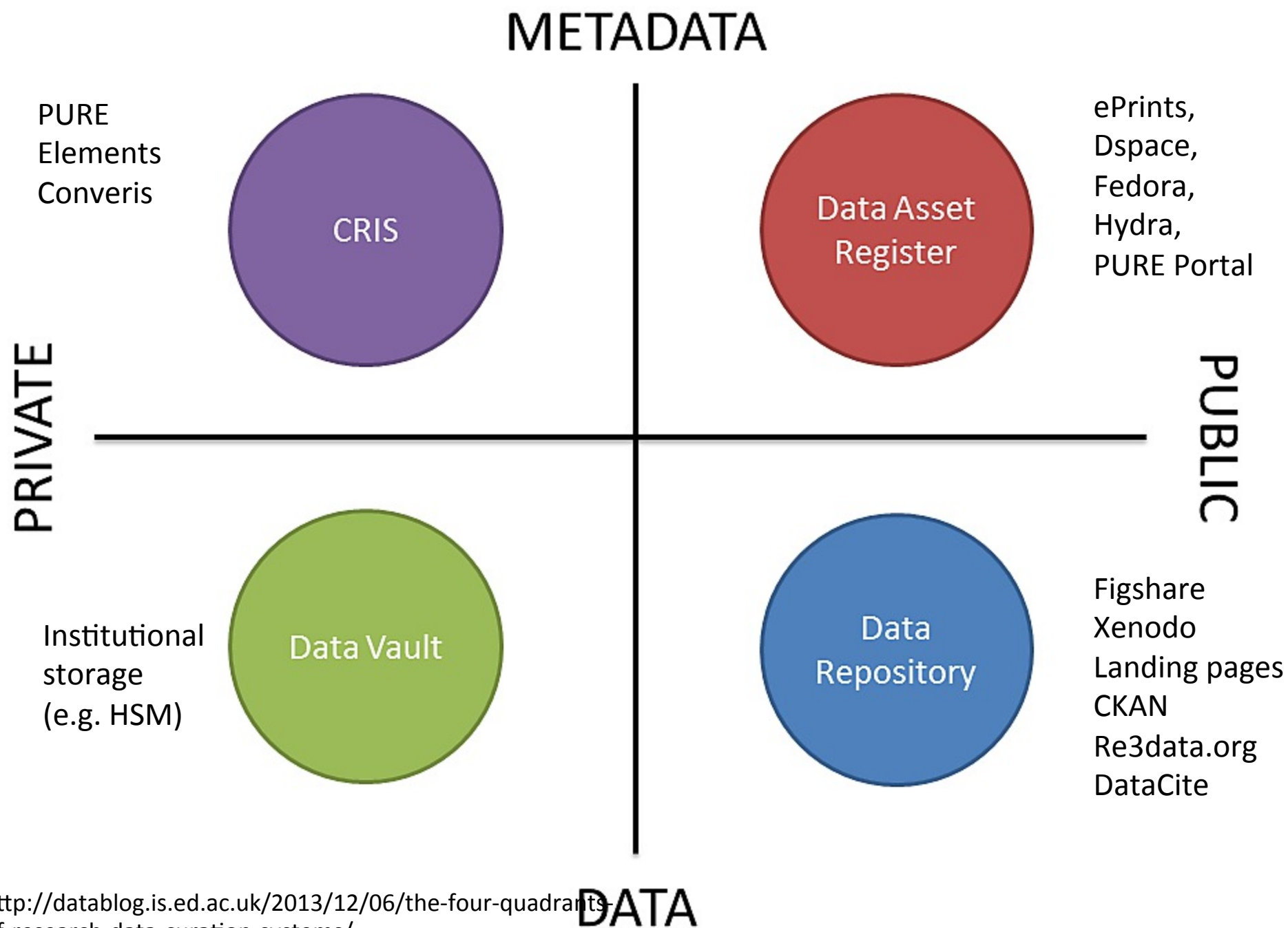
Easy access

- Quick to retrieve data when its needed
- Doesn't have to be instant

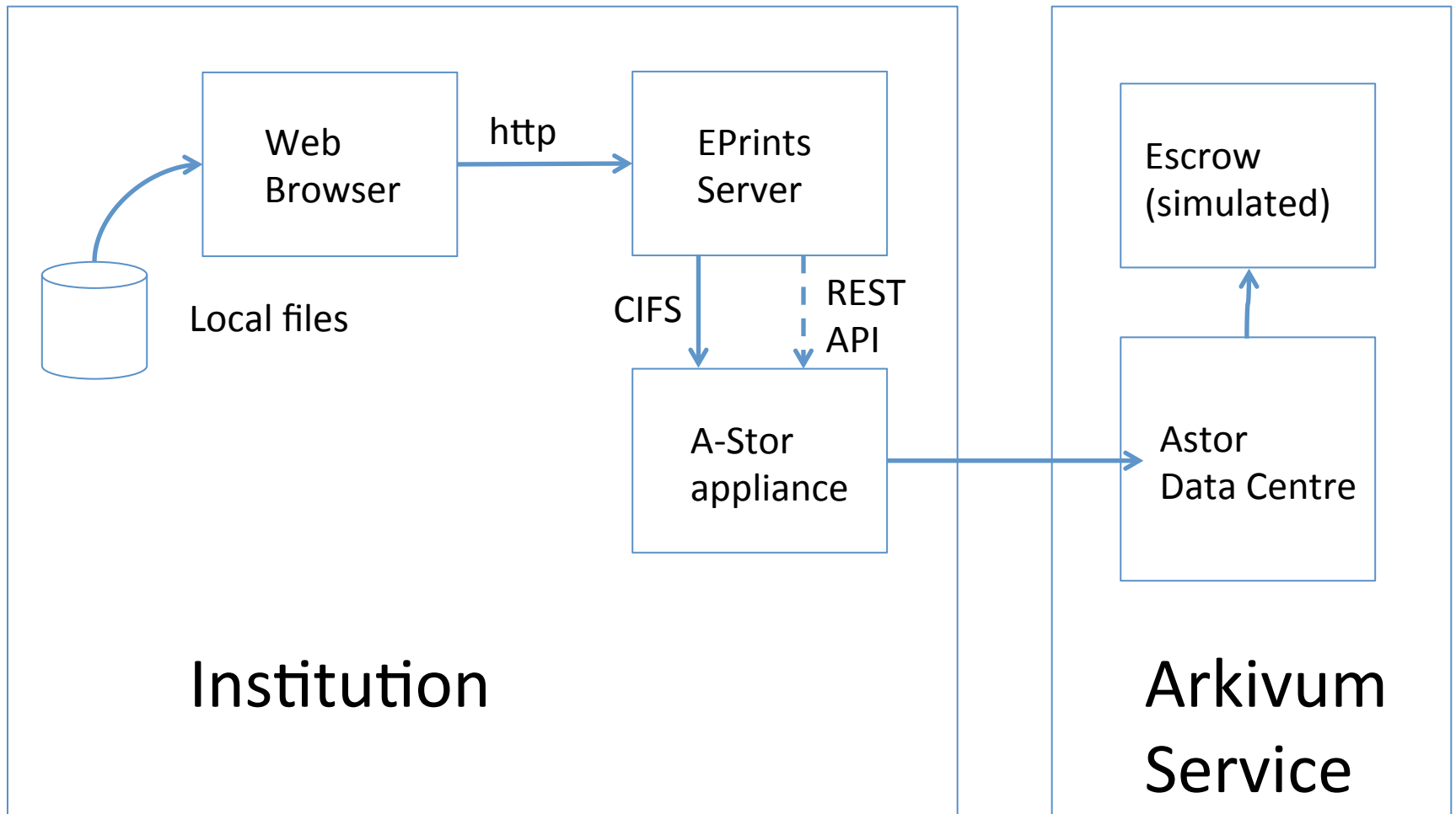


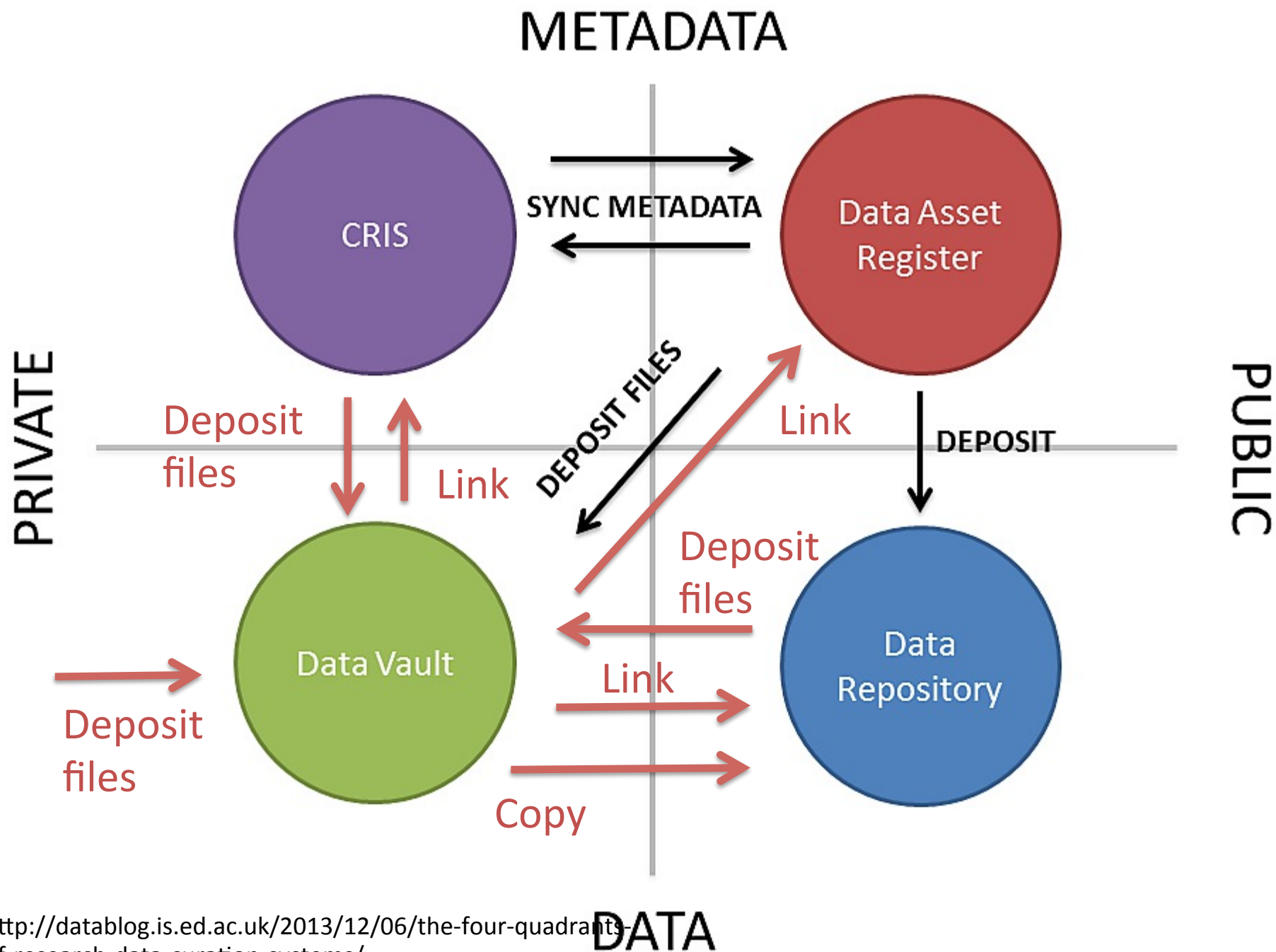
Low cost

- Way to cope with data growth
- Free up expensive resources
- Eliminate the 'cost of loss'



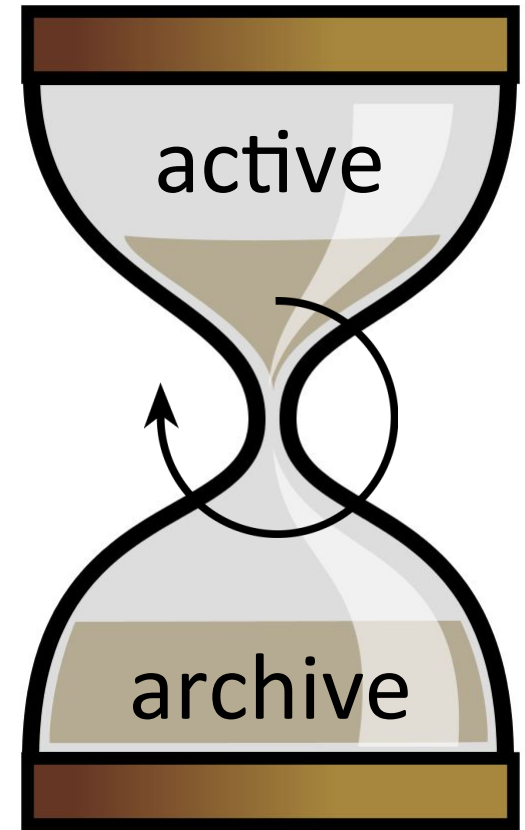
EPrints demo



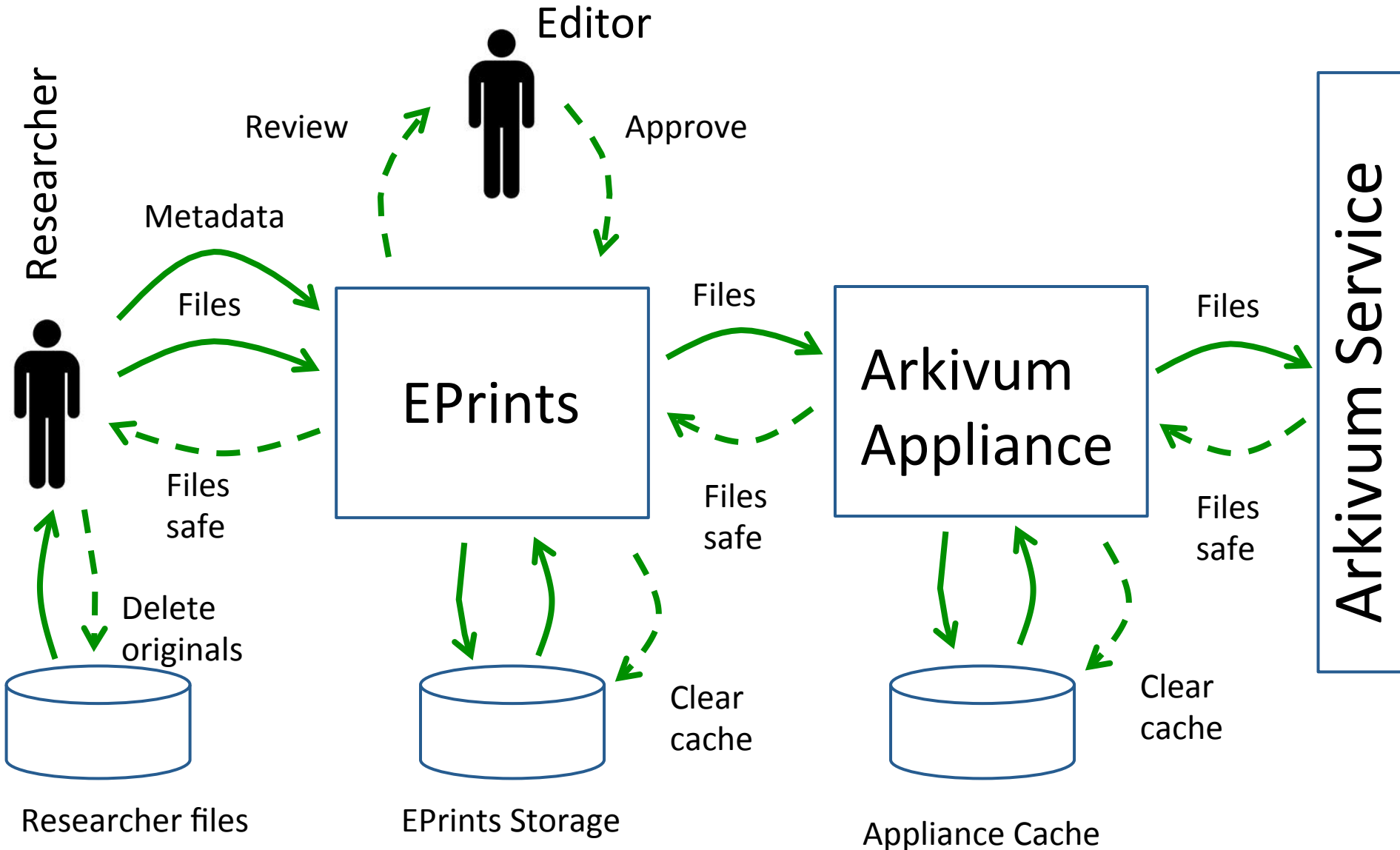


Role of the CRIS or IR

- Gateway active \leftrightarrow archive
- Single place to go
- Management of metadata
- Quality control
- Decision points and workflows
 - Effective use of resources
 - Review and approval
 - Budget and charging
 - Ethics and compliance



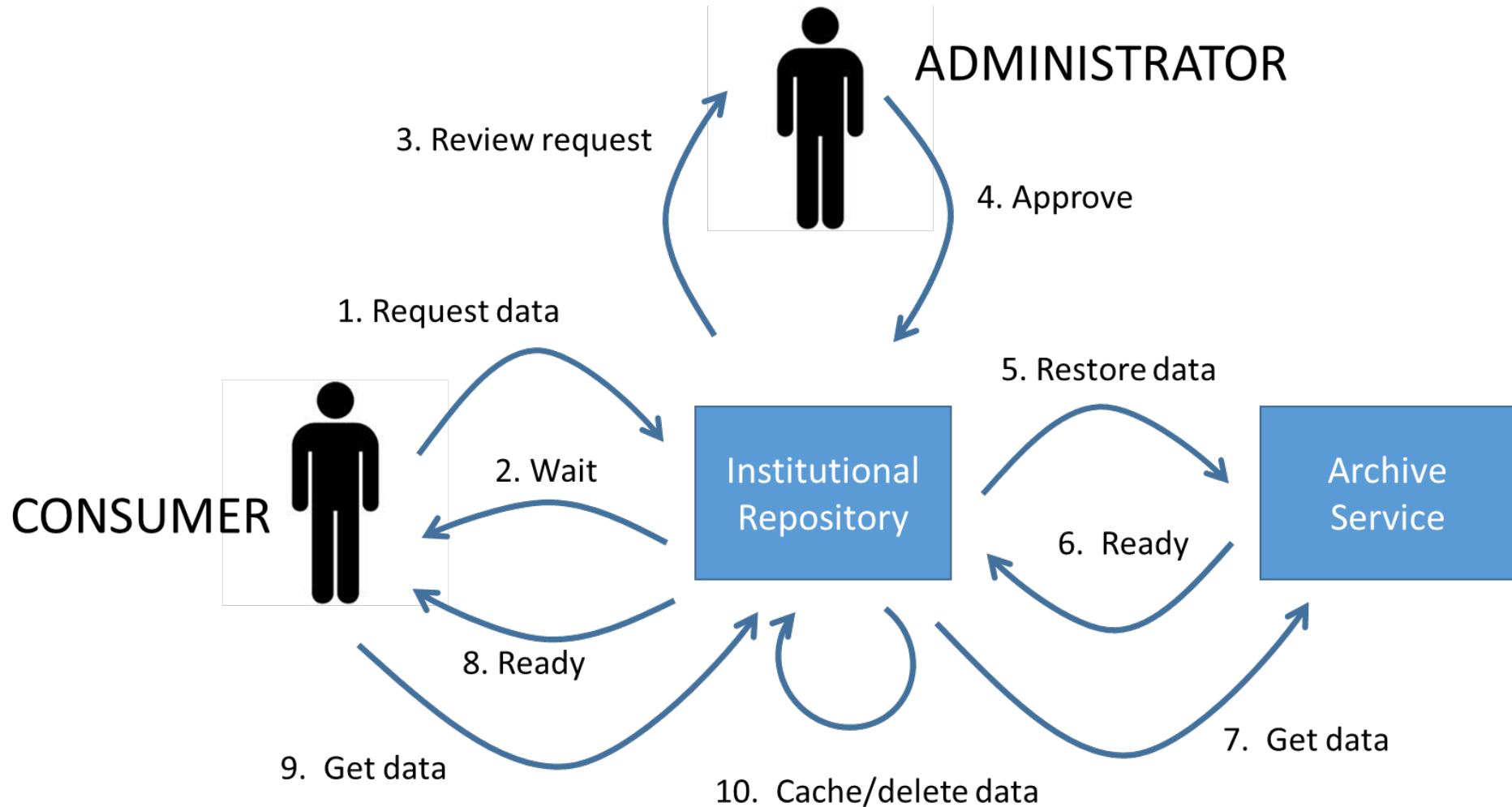
EPrints: deposit into the archive



Putting data into the archive: chain of custody

- Send data to archive
- Generate checksum for data in archive
- Compare with checksum in IR or CRIS metadata
- Update metadata that transfer is successful
- Apply archive safety mechanisms
 - Replication, WORM, access control
- Update metadata to complete the audit trail
- Remove local copy (if desired)

EPrints: getting stuff out



Getting the data out of the archive: managed access

- Open access: no license, no barriers
- Unrestricted use, but request for access
- Unrestricted use, but charge for access
- Restricted use: request/approve access
- Embargoed: no access for set period
- Locked down: no access

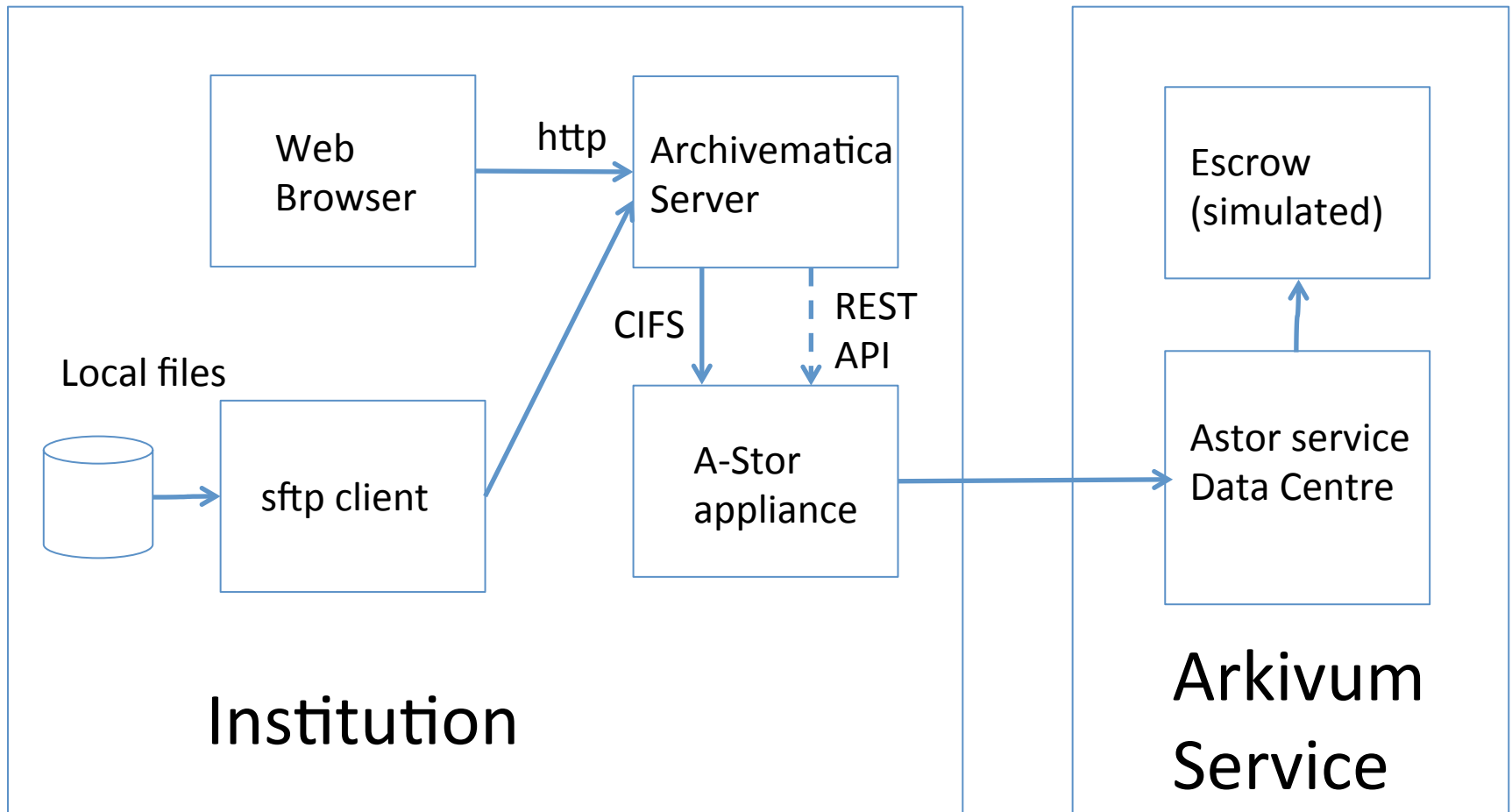
“It may be reasonable for research organisations to require persons requesting access to specific research datasets to do any of the following before granting access: register; create an account; prove identity; accept terms and conditions of use of the data.”

Archivematica

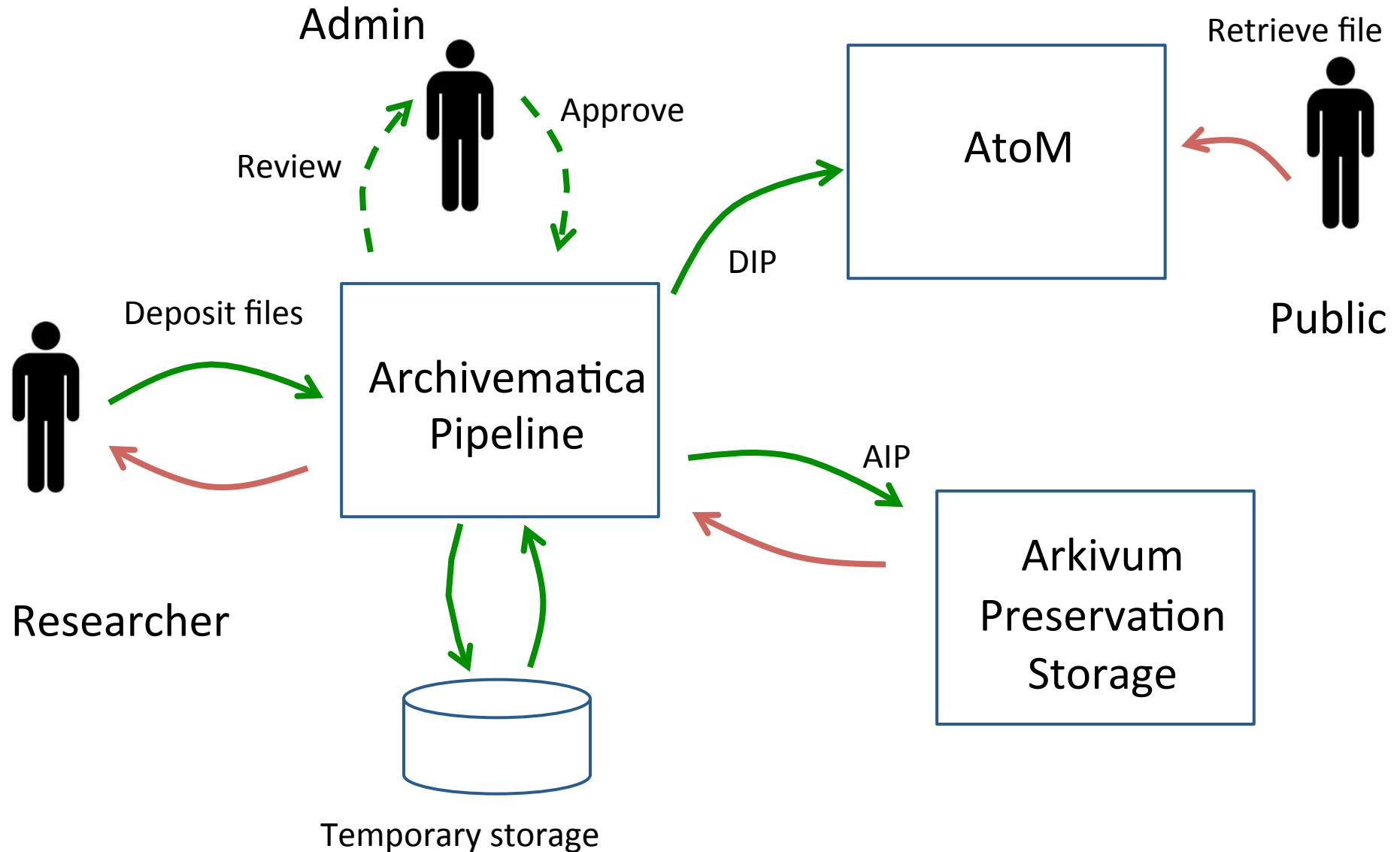
- Preparing content for preservation and access
- Automatic deposit into the archive
- End-to-end chain of custody



Archivematica demo



Archivematica



Summary

- Managed deposit
 - Check data and metadata is complete & correct
 - Chain of custody
- Managed access
 - Help user understand what's happening
 - Monitoring and control
- Simple to use
 - Easy to use, one place to go
 - Automation plus checkpoints

Questions

matthew.addis@arkivum.com