Examples of combining research data archiving and access

Matthew Addis
Arkivum
Contents

• Intro
• Institutional and researcher perspectives
• Illustrations
  – Loughborough
  – ULCC
  – Aston
• Summary
Arkivum in less than 60 seconds

- 100% GUARANTEED
- 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
Typical state of play

• Policy exists for RDM
• Some idea of what data exists and where
• RDM infrastructure under construction
• Lots of awareness raising, training and support
• Getting researchers to act is a challenge!
Data Archiving - Integrations

• Eprints
• Dspace
• Figshare
• Symplectic Elements
• Archivematica
EPSRC Expectations

• Institution policy and internal DMP
• Online description of research data (discovery)
• DOIs and details of how to access the data
• Data accessible unless a reason not too
• Data preserved for 10 years+ (usage monitoring)

Metadata v.s. data
Public v.s. private
Discovery v.s. access
http://datablog.is.ed.ac.uk/2013/12/06/the-four-quadrants-of-research-data-curation-systems/
http://datablog.is.ed.ac.uk/2013/12/06/the-four-quadrants-of-research-data-curation-systems/
Role of the CRIS or IR

- Gateway active ↔ archive
- Single place to go
- Management of metadata
- Quality control
- Monitoring and metrics
- Decision points and workflows
  - Effective use of resources
  - Review and approval
  - Budget and charging
  - Ethics and compliance
RDM begins and ends with the researcher

- Researchers create the data
- Researchers understand the data
- Researchers publish the data
- Researchers use the data

- RDM needs positive cost/benefit for researchers
Clear benefits, minimal cost

• More citations
• More downloads
• More collaborations
• More funding (and lower risk of rejection)
• Be seen to be following Good Research Practice

• One place to go, minimal training
• Part of day-to-day business, embedded RDM
• Easy to use, impact on researcher time is a big cost
CASE STUDY: LOUGHBOROUGH UNIVERSITY

FIGSHARE, ELEMENTS, ARKIVUM, DSPACE
Research data deposit via Figshare

Data and metadata → figshare web interface → figshare storage on Amazon

Data and metadata → Metadata and DOI

Data → Arkivum appliance

Data → Metadata and DOI

Data → DSpace

Data → Arkivum data centres

DOI → DataCite

DOI → Symplectic Elements
Handling of large datasets

Data and metadata

figshare desktop uploader

figshare storage on Amazon

Symplectic Elements

DataCite

Data and metadata

Arkivum appliance

Metadata and data link

Arkivum data centres

Metadata and DOI

Data and DOI

Metadata and DOI

DSpace

DOI
Simple story for EPSRC

• Data easily discoverable
• Automatic minting of DOIs
• Data easily accessible
• Data stored safely for 10 years+
• Data access stats can be monitored
• Adopted by researchers
CASE STUDY: ULCC

EPRINTS

ARKIVUM
EPrints: deposit into the archive

- **Researcher**
  - Files
  - Metadata
  - Review
  - Approve
  - Delete originals

- **EPrints**
  - Files
  - Files safe
  - EPrints Storage
  - Clear cache
  - Files safe

- **Editor**
  - Approve

- **Arkivum Appliance**
  - Files
  - Files safe
  - Clear cache
  - Mint DOI

- **DataCite**
  - Mint DOI

- **Arkivum Service**
  - Files
  - Files safe
  - Clear cache

- **Researcher files**
  - EPrints Storage
  - Appliance Cache
EPrints: data access

- Request data
  - Wait
- Review
- Approve
- Retrieve files
  - Files ready
  - Files ready
- Retrieve files
- Files

Researchers files

Arkivum Service

Files

Arkivum Appliance

Files

Appliance Cache
Getting the data out of the archive: managed access

• Open access: no license, no barriers
• Unrestricted use, but request 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.”
Direct archiving, large datasets
Easy to adopt and use

• Hosted solution: ‘RDM nursery’
• Easy for institutions to get started
• Single interface for the researcher
• Review/approve, quality control
• Adding support for ‘growing RDM’
  – Larger datasets
  – Onsite/offsite mix’n’match
Safe storage for future access

Institutional Storage service

Admin

Copy

Delete

Files

Files safe

Review

Files

Files

Files safe

Local Storage

Appliance Cache

Researcher files

Arkivum Service

Arkivum Appliance

Researcher files
Start simple

• Quick to implement
• Supports institutional storage for RDM
  – Lowers costs
  – Addresses funder expectations
  – Recoverable costs
• Build access on the top
  – Persistent names/paths/ids
  – Add links to CRIS, IR etc.
Summary

• Range of approaches
  – Start with storage and work up
  – Start with hosted storage/access as a quick win
  – Start with public data and provide a complete solution

• Researcher focussed
  – Ease of use
  – Positive benefits
Questions

www.arkivum.com

matthew.addis@arkivum.com

http://orcid.org/0000-0002-3837-2526