Arkivum data archiving service, research data management and access

Steve Mackey
Arkivum
Arkivum introduction

- 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
Arkivum/100

- 2 Data centre copies
- 1 Escrow copy
- 100% data integrity guarantee
- Scalable upwards from 1TB / yr
- When data integrity, longevity and indemnity insurance are paramount
Arkivum’s active archive

A-Stor creates an encrypted copy on the gateway and copies to Arkivum’s Data Centre One

A-StorDC creates the escrow copy in the Tape Vault

A-StorDC creates the second copy in Arkivum’s Data Centre Two

Once all three copies are confirmed safe and secure the original copy can safely be deleted
20 years of keeping content alive

- Media
- Maintenance
- Integrity checks
- Hardware refresh
- Software migration
- Hardware migration
- Format migration
- Staff migration
- Supplier migration

21 October 2014
Ready made exit plan

- Metadata and data on a file system
- Open standards and formats
- 3-way agreement on ownership/access
- Reduce costs and risks
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!
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
The four quadrants of research data curation systems:

- **PRIVATE**
  - CRIS
  - Data Vault
  - Institutional storage (e.g. HSM)

- **PUBLIC**
  - Data Asset Register
  - Data Repository
  - DataCite
  - Re3data.org

- **METADATA**
  - EPrints, DSpace, Fedora, Hydra, PURE Portal
  - JISC Data Registry

- **DATA**
  - Figshare
  - Zenodo
  - Landing pages
  - CKAN

http://datablog.is.ed.ac.uk/2013/12/06/the-four-quadrants-of-research-data-curation-systems/
Arkivum - integrations

- Eprints
- Dspace
- Figshare
- Symplectic Elements
- Archivematica
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
Handling of large datasets
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
Questions

www.arkivum.com

steve.mackey@arkivum.com