

Arkivum Briefing to the DPC

Matthew Addis, Simon Bostock & Tom Lynam

Arkivum: Archiving and Digital Preservation Solutions and Services

- Founded in 2011 out of the University of Southampton
 initial focus on Higher Education Research Data
- VC backed and funded
- Headquartered in Reading, UK
- 25 full time employees and growing!
- ISO 9001 and 27001 certified
- Full SaaS offering launched in 2017
- +70 customers as of 2021 across Higher Education, Heritage, Pharma and Life Sciences
- Selected for Design and Prototype phases of the ARCHIVER project



Arkivum Archiving and Preservation Solution



arkivum

We apply and contribute to Archiving and Digital Preservation Good Practice



Our commitment to every Arkivum customer...

() 2(3)

We provide a simple and easy-to-use solution to overcome customers' long-term data management challenges.



Our customers' data safety, security and integrity is the core of our business.



The ownership and portability of our customer's data is non-negotiable – customers' data is their data.



Years of cross-industry archiving and preservation experience helps save time, effort and costs for our customers.

We will continually strive to improve our solution in line with market drivers and customer feedback.

Case study: Irish Traditional Music Association (ITMA)

Sector: Heritage

User Case: Special Collections

About: Irish Traditional Music Archive (ITMA) is the national public archive and resource center for Irish traditional music, song and dance.

Requirements: ITMA holds both physical and born digital material including large volumes of digital surrogates as a result of years of digitisation work. The agency currently has close to 31 terabytes of audio, score and image files in their digital collections, made up of over 62,800 audio files, just under 120,000 image files and 13,000 scores (music notation files). Video files currently occupy about 66 terabytes of space.

Solution: For several reasons including compliance with the OAIS standards, 100% data integrity guarantee, the customer support on offer, the cost, the use of open-source products, and the end-to-end nature of the system (that is preservation as well as access) Perpetua was chosen.

icma Irish Traditional Music Archive Taisce Cheol Dúchais Éireann

"After several years refining scalable workflows for the ingest of assets into Arkivum-Perpetua, ITMA is now well enscanced on its digital preservation journey. We are safeguarding an extensive digital collection while also providing meaningful access to a national public archive for Irish traditional music, song and dance."

Maeve Gebruers, Archivist, ITMA

arkivum

Solution Overview & Demo

Arkivum Archiving and Preservation Solution



arkivum









Digital Preservation Education

Ο

O

0

0

 \mathbf{O}

0

We're a content driven organisation

- Blog posts:
 - FAIR Data Management Principles
 - ALCOA+ three part series
 - Open Data Workshop sharing scientific research
- Research:
 - 2020 State of Digital Preservation in Higher Education
 - TMF Futures Report Update coming in May/June 2021
- eBooks:
 - Building the business case for digital preservation
 - Guide to eTMF archiving and preservation
- All available on <u>www.Arkivum.com</u> and updates via our Twitter and LinkedIn channels



Customer and Community Engagement

- Industry events
- Community events:
 - DPC
 - HSRAA
 - NHS R&D Forum
- Arkivum User Group
- Arkivum webinars



Q8A with the DPC

0

•

0

•



THANK YOU

Questions?

Matthew Addis – Matthew.Addis@Arkivum.com

Simon Bostock – Simon.Bostock@Arkivum.com

Tom Lynam – Tom.Lynam@Arkivum.com

www.arkivum.com

hello@arkivum.com

Our work with regulated data

The Problem – Electronic Trial Master File (eTMF) Archiving and Preservation

In a clinical trial involving human subjects, a set of documents known as a trial master file (TMF) must be produced.

Historically paper-based, the clinical trial data has now taken the form of an electronic trial master file (eTMF) as part of the wider digital transformation of the clinical trial process.

Industry regulations across the UK, US and Europe* require this clinical trial data to be archived and ready for inspection for up to, and in some cases beyond, 25 years.

Throughout this retention period, drug sponsors can be called on to share and make available, any part of the documents to investigators.

If they are unable to do so, the sponsor is liable for heavy fines, reputational damage and delays to drug development and manufacturing.



Data Integrity



Data Integrity is the degree to which data are complete, consistent, accurate, trustworthy, reliable and that these characteristics of the data are maintained throughout the data life cycle. The data should be collected and maintained in a secure manner, so that they are attributable, legible, contemporaneously recorded, original (or a true copy) and accurate. [MHRA GXP Data Integrity Guidance and Definitions 2018]

Data governance should address data ownership and accountability throughout the lifecycle, and consider the design, operation and monitoring of processes / systems to comply with the principles of data integrity including control over intentional and unintentional changes to data. [MHRA GXP Data Integrity Guidance and Definitions 2018]

Validated Digital Preservation and Archiving

arkivum



Bringing archived data to life

Case study: Neuraly

Sector: Pharma and Life Sciences

User Case: eTMF and Clinical Trial Data

About: US based clinical-stage company Neuraly are translating foundational research into therapies for patients suffering from Parkinson's disease and Alzheimer's disease.

Requirements: Requirement to initially archive trial data and quality documents (including the eTMFs) from two Phase II studies in line with GCP guidelines, FDA regulations and the ALCOA+ principles. This was part of a wider project to establish a QMS within Neuraly.

Solution: Arkivum helped Neuraly to migrate their eTMF data via the Exchange Mechanism Standard (EMS) into our system. Once uploaded, the data was archived and preserved in line with OAIS and ALCOA+. Neuraly are now confident that they are 'inspection ready' and compliant with GCP guidelines and FDA requirements.

NEURALY

"Arkivum's team was knowledgeable, very helpful and responsive from beginning to end. We are looking forward to working with the team for subsequent eTMF archiving needs."

Celia Baula, Head of Quality and Director, R&D Program and Portfolio Management.

Access Examples

0

•

0

0

DCÚ



α

Holdings Quick search

- Collection C1 Seán Lester Collection
- Item 1 Diary: October 1935 January 1...

2 more...

Part c1_1_0002 - Diary: October 1935 - J... Part c1_1_0003 - Diary: October 1935 - J... Part c1_1_0004 - Diary: October 1935 - J... Part c1_1_0005 - Diary: October 1935 - J... Part c1_1_0006 - Diary: October 1935 - J... Part c1_1_0007 - Diary: October 1935 - J... Part c1_1_0008 - Diary: October 1935 - J... Part c1_1_0009 - Diary: October 1935 - J... Part c1_1_0010 - Diary: October 1935 - J... Part c1_1_0010 - Diary: October 1935 - J...

Part c1_1_0006 - Diary: October 1935 - January 1936, p0006

Seán Lester Collection » Diary: October 1935 - January 1936 » Diary: October 1935 - January 1936, p0...



Identity area

Reference code	IE DCUA C1/1/c1_1_0006
Title	Diary: October 1935 - January 1936, p0006
Date(s)	October 1935-January 1936 (Creation)
Level of description	Part
Extent and medium	1 digital image
Context area	
Name of creator	Lester, Seán (1888-1959)
Repository	Dublin City University Library



	Conditions of	access and	use area
--	---------------	------------	----------



RECORDS AND **ARCHIVES**

HOME ABOUT > HOW CAN I SEE THIS ITEM? RESEARCH GUIDES ACCESSIBILITY STATEMENT



ARCHIVE

Holdings Quick search

- **v** Collection CPC China Visual Arts Project...
- **Series** CPC/1 Chinese Poster Collection
- Subseries CPC/1/A Agriculture

Item CPC/1/A/1 - Struggle to realise and s... Item CPC/1/A/2 - Criticise hard, work har... Item CPC/1/A/3 - Go all out to achieve bas...

Item CPC/1/A/4 - Man's determination wil...

Item CPC/1/A/5 - Go full stream ahead an...

61 more...

Item CPC/1/A/1 - Struggle to realise and surpass the magnificent 10 year plan by 1985

China Visual Arts Project Archive » Chinese Poster Collection » Agriculture » Struggle to realise and surpass the magnifi...

♥ Other languages available ▼



Identity area

Reference code CPC/1/A/1

Title Struggle to realise and surpass the magnificent 10 year plan by 1985

Clipboard N Add Explore Reports Explore Browse as list Browse digital objects Export Lin Chenghan / 林成翰

(Creator) Tianjin Municipal New China Bookshop / 天津市新华书店 (Subject) Tianjin People's Fine Art Publishing House Printing

Our work with very large-scale data

European Open Science Cloud - The Vision

"We are **creating a European Open Science Cloud now**. It is a trusted space for researchers to store their data and to access data from researchers from all other disciplines. We will create a pool of interlinked information, a **'web of research data'**. (...) The idea is that once we have the rules of the game ready, then **we will open this up to the broader public sector and to business as well**. So that companies can come in, store the data and use the data."

Special Address by Ursula von der Leyen, President of the European Commission, 22nd January, WEF DAVOS 2020 https://www.youtube.com/watch?v=QN476nVbFVs&feature=youtu.be&t=682

Role of the EOSC:

To ensure that 1.7 million European researchers and 70 million professionals in science and technology reap the full benefits of data-driven science

- Federated virtual environment
- Open services for storage, analysis and re-use of research data
- Promote a pan-European and multidisciplinary research approach

Phase 1 investment of EUR 300 Million on EOSC core services EOSC legal entity expected to be created by the end of 2020











FAIR Forever?

Long Term Data Preservation Roles and Responsibilities, Final Report

February 2021 (V.7)

Dr Amy Currie and Dr William Kilbride



EOSC FAIR Forever has received funding from the European Union under the EOSC Secretariat project. EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, grant Agreement number 831644. Europeana is an initiative of the European Union, financed by the European Union's Connecting Europe Facility and European Union Member States (https://pro.europeana.eu/ and https://www.europeana.eu)

ARCHIVER: archiving and preservation, from long-tail to large-scale

- 1. Scalability and Performance
 - Ingest up to 100TB per day
 - Leverage elastic infrastructure and 'serverless computing'
 - Apply massively parallel processing to preservation workflows
- 2. Cost-efficiency
 - Minimize unnecessary processing, pay-per-use
 - Scale to zero, deep archive storage
 - Multi-tenanted SaaS
- 3. Portability
 - Data lives longer than platforms, services, vendors
 - Migrate between hyperscalers
 - Run on-premise or private data centres
- 4. Integration
 - Archiving and preservation available as web services
 - Integrate with Repositories and other parts of the FAIR environment



Portable FAIR Factories



Arkivum / Google Solution:

- Scalable storage and compute
- High speed ingest and access
- Policy based cost optimization
- OAIS workflows and packages
- Digital Preservation rules and actions
- FAIR datasets and access
- Hosted scientific applications
- Open standards and specifications
- Exit and migration strategies



arkivum

Scalability, Cost-efficiency, Portability, Environmental Sustainability



Run	Size (TB)	Files	Start	End	Duration	TB per day
EBI ITU	1.2	671	2021-04-19T19:46:26.505Z	2021-04-19T21:05:28.899Z	01:19:02	21.86
CERN CMS	2.7	2916	2021-04-19T22:17:03.159Z	2021-04-20T00:36:35.608Z	02:19:32	27.86

Benefits and outcomes available to all Arkivum customers

- 1. High-speed Minimal Effort Ingest / Minimal Viable Preservation
- 2. Automation and Scalability: Preservation Factories for organisations of all sizes
- 3. Environmental sustainability: efficient use of compute, storage and networking
- 4. Trusted Digital Repositories: FAIR, Trust, CTS, DPC RAM





