



Project Abstract: Phase 1 Arkivum

About Arkivum:

Arkivum provides a long-term data management solution that includes the safeguarding, digital preservation and online access of digital content. The solution is delivered as a fully managed service, supports a wide range of content types, and can be deployed onto public or private cloud.

Arkivum is recognised internationally for its expertise in the archiving and digital preservation of valuable data and digitised assets in large volumes and multiple formats. The long-term security, integrity and accessibility of data is crucial for all of our customers and partners, who share a commitment to good practice in its stewardship and governance.

The solution has been selected by organisations across a range of sectors, including life sciences, corporate archives, research organisations (including higher education) and memory institutions in culture and heritage. With the aid of the Arkivum solution, each customer is able to derive optimum long-term value from their data, collections and intellectual property. In addition, they benefit from 100% data integrity guarantee.

Headquartered in the UK, with US offices in Boston and New York City, Arkivum advocates and adheres to best practice in research data management, including the FAIR principles for data: Findable, Accessible, Interoperable, Reusable. In particular, Arkivum helps organisations to build and operate Trusted Digital Repositories for their content.

Proposed solution:

Partnering with Google in the project, Arkivum will develop a solution for long-term data management and online access, to be deployed on the Google Cloud Platform (GCP). While Arkivum's particular expertise lies in reliably keeping large volumes of complex archive data secure and accessible, thus maximising compliance, insight and discovery, the new collaboration with Google Cloud enables the team to manage data on the immense scale required by ARCHIVER.

The Arkivum solution and workflows are all about enabling content, which in the case of ARCHIVER includes scientific data from a range of sources, to be captured, ingested, preserved and made accessible to those who need to use that content in the future.

This includes the ability to ingest, validate, organize and manage content as it comes into an archive. The content will then go through appropriate preservation and safeguarding processes, including generating OAIS archiving packages to ensure it is properly protected and remains usable. This will also ensure that the data is searchable, discoverable and



ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

accessible for users both today and in the future so people can find and use the content in the archive that they need, when they need it.

This in turn helps research organisations achieve their research data management objectives, for example ensuring that scientific datasets are Findable Accessible Interoperable and Reusable.

Meeting the ARCHIVER requirements:

The combined Arkivum and GCP solution will meet all layers of the ARCHIVER project requirements;

- In layer 1 of ARCHIVER (storage/basic archiving/secure backup), GCP fulfils the requirements of high-volume data storage with fast ingest and access, all to the petabyte level range.
- In layer 2 (preservation) Arkivum's solution addresses the need for long-term digital preservation following the OAIS model, including obsolescence management, file fixity and authenticity checks, and packaging for preservation and access.
- In layer 3 (baseline user services), the Arkivum solution provides the ability to search, share and index large and complex research datasets
- Then finally in layer 4 (advanced services), GCP provides the basis of hosting and running scientific applications that can be executed against archived and preserved datasets.

In short, we'll be combining and extending the solutions from Arkivum and Google to deliver the full ARCHIVER requirement stack.

R&D:

The first phase of the project will focus on the design/architecture of a long-term digital preservation and archiving solution to meet the Buyer requirements. This specifically includes:

- Using GCP to meet the baseline requirements of supporting very large datasets, with high speed ingest and access, but also with the ability to host and run scientific applications against this data.
- Using digital preservation and data archiving services to meet the requirements for OAIS and to provide organisations with a hosted solution for operating their Trusted Digital Repositories.
- Ensuring that research data can be described, organized, sliced/diced, tagged and published in a flexible way that meets FAIR principles.
- Total Cost of Service models and the ability to optimize cost against data volumes, access frequencies, data safety, data processing and retention periods.
- Pervasive use of open standards, open specifications, open source and open APIs to ensure portability, interoperability, exit strategies and removal of vendor and solution lock-in.
- Detailed models and commercialization plan that includes Service Level Agreements,



ARCHIVER

HELIX
NEBULA

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

User Support, Licensing, Service Configuration and Pricing for new commercial services based on ARCHIVER to be provided to the Buyers Group, the European Open Science Cloud (EOSC) and beyond.