



Project Abstract: Phase 1

GMV & Piql Consortium

The GMV & Piql consortium

The consortium is composed by **GVM** and **Piql** as main members and **AWS** and **Safespring** as subcontractors providing cloud resources.

Piql's main goal, as a company involved in the archiving business, is to define the preservation requirements and the architecture of the service and **GMV's**, as an expert company in engineering and ICT technologies, is to create a layer of services on top and below the preservation core.

About GMV

In nearly 30 years GMV has evolved from a tiny three-person engineering firm working almost exclusively in the space sector into a strong business group established in Europe, the USA and Asia, trading in several hitech sectors with a swelling order book in all five continents. The GMV-SES industrial directorate general was created to contribute to GMV's overall mission by providing secure ITC solutions for the corporate information systems as well as specialized business solutions for telecommunications and media, bank and insurance, public administration and healthcare in a global market and in cooperation with the remaining general directorates.

Our **expertise on Information and Communication Technologies (ICT)** allows us to come up with tailored solutions that perfectly fit the needs of our clients in large corporations. GMV provides a variety of IT solutions for areas of maximum specialization and innovation while maintaining total commitment to product quality. The experience of GMV's team of professionals and their passion for technology and quality ensure the best telecommunications and e-business/m-business services on the market.

As global suppliers and integrators, GMV provides customized solutions including hardware and software platforms, professional services and 7x24 support.

About Piql

Piql AS is a Norwegian company established in 2002 under the name Cinevation. In 2014 the company rebranded to Piql and today delivers services for long term preservation of digital data through a network of international partners. The company has a strong technology focus and the main office in Norway is the company's research and development department. The company has the last 15 years developed technology and services for storage of digital data together with leading Norwegian and international research institutions like Norwegian Computing Center, Norwegian Defense Research Establishment and Fraunhofer Institute in Germany. Piql main product is the Piql System, an open solution for digital preservation where the storage medium used has a documented lifetime of several hundred years. This solution uses Archivemata as its core digital preservation platform and is prone supporter of open source solutions for digital archiving. Piql has secured funding from EU commercialization projects of total 2M€, giving the company a solid financial base.



ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

About AMAZON WEB SERVICES

Amazon Web Services (AWS) is the world's most comprehensive and broadly adopted cloud platform, offering over 175 fully featured services from data centers globally. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—are using AWS to lower costs, become more agile, and innovate faster.

The leading cloud Platform with the most functionality, the largest community of customers and partners, the most secure and with the fastest pace of innovation and the most proven operational expertise.

About SAFESPRING

Safespring produce secure, reliable and flexible cloud services. Our infrastructure as a service is based on open source and open standards. We produce all our services within Europe and its therefore easier for you to follow European local laws and regulations.

Proposed solution

Our solution has been divided in 4 layers to address the different levels of design. The key features of the proposed solution are described below:

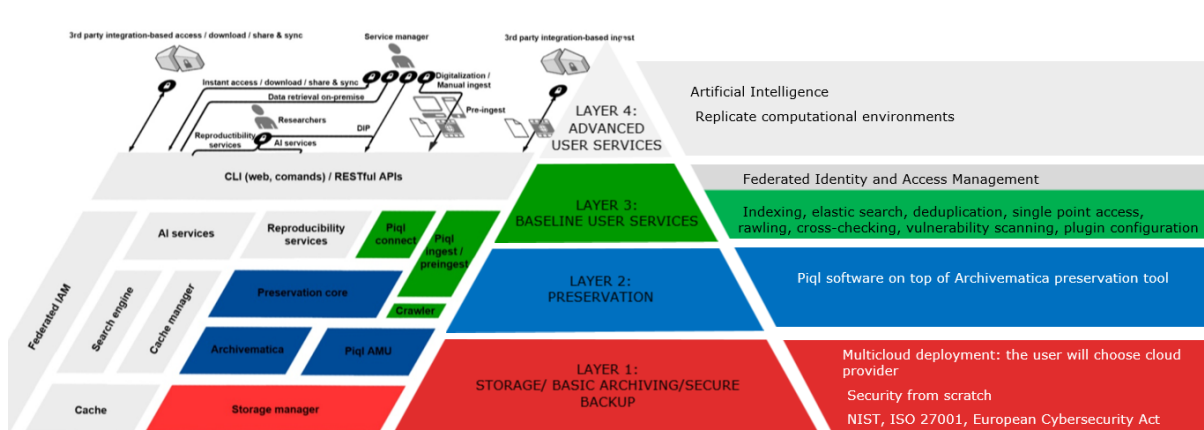
- **Layer 1** (storage, basic archive and secure backup capabilities): the solution will provide features for **multi-cloud deployment** (over private, public, hybrid, community and special purpose Clouds) where the buyer will be able to choose which cloud provider wants to use for his storage.
Security will be implemented from the scratch and grouped in clearly defined areas that help to achieve requirements and compliance requisites, such as Secure Coding, Secure CD/CI and Secure Infrastructure.
Our consortia knowledge and frameworks covers NIST, ISO27000 and European Cybersecurity Act, to protect stored, transmitted or otherwise processed data against accidental or unauthorised storage, processing, access or disclosure during the entire life cycle of the ICT product.
- **Layer 2** (preservation capabilities): the solution will provide the **preservation** of files, folders, content management packages (archives) and data types, as well as emails backup in pdf format. Migration and backup mechanisms, APIs, SFTPs, web interfaces, drag&drop ingestions, and metadata and context processing, will be also provided.
The preservation services will rely on **Archivematica**, designed around OAIS requirements, which should be extended due to the high rate, high volume requirements.
PiqI AMU is used in case of very valuable data need to be kept for 100+ years on PiqIFilm but it should be improved because of the required data sizes. The user could decide which media to be used for the storage of the data among the three available, tape, hard-disk and PiqIFilm, depending on its needs.
- **Layer 3** (user services): The solution will allow the operation with volumes of hundreds of TBs with support of **indexing, elastic search, deduplication, single point access, crawling, cross-checking, vulnerability scanning, and plugins configuration** filtering potential datasets rapidly, to access dataset metadata and decide on its relevance (e.g. citation purposes or reusing a dataset). Automated metadata indexing for several tens of PB content will be supported, including also



support for dataset filtering and information tagging, aiming at maximum interoperability and easy and broader searches for the research community. Piql Connect and Piql UI will simplify the user interaction with the whole system. Access and permission management against repositories and various collections will be also supported by the Federated Identity and Access Management schemes, as a strategy to promote open data access in the research community.

- **Layer 4** (advanced services capabilities): the solution will provide features for the retention and integrity of data over a decade ensuring its tamper-proof behaviour whilst allowing easy access and basic re-usability with a collection of **artificial intelligence** functionalities.

The solution will provide the ability for a researcher to replicate a computational experiment that was done by someone else, using the same software and data. Reproducibility of experiments will be based on container-based workflows exposed to the user taking advantage of modern cloud computing paradigms that provide scale on-demand, necessary to run large scientific experiments.



Solution design from components to layers

The R&D for Phase 1

During the Phase 1 of the Archiver project the complete design of the system will be implemented based on the components listed below. These components have been carefully thought to fulfill all the needs and to provide a complete system to become one of the leader solutions in the market of preservation.

How the objectives will be achieved

Within the consortium several dedicated teams made up of highly qualified engineers have been created to work in the different lines involved, starting from infrastructure to software architecture and engineering to Artificial Intelligence and Big Data to support services.

As part of this Phase 1, a complete set of self-assessment compliance activities will take place to ensure that the proposed design fulfills all the applicable regulations such as GDPR, FAIR, OAIS, to mention some of them.

During this Phase 1 of design, a complete environment (to scale) will be deployed involving



ARCHIVOR

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

the two different cloud providers (AWS and Safespring) to carry out proofs of concept as well as experiments and measurements to validate the proposed design.

On top of all this, the archiving and preservation business is core to Piql AS and also a very interesting new field for GMV in which we are keen to get into.