Press release ARCHIVER Pilot phase Award

Arkivum and LIBNOVA consortia selected to deliver pilot long-term data preservation solutions to the European Open Science Cloud

Within the framework of the ARCHIVER pre-commercial procurement tender, between December 2020 and August 2021 three consortia competed to deliver innovative, prototype solutions for long-term data preservation. Two of them have been selected to continue with the pilot phase and deliver research ready solutions for long-term data preservation of research data, therefore supplementing a lack in the current European Open Science panorama. They have been announced at a Public Ceremony on the 29th of November 2021, virtually hosted by DESY.

2nd December 2021

According to the UNESCO definition[1] "*digital preservation consists of the processes aimed at ensuring the continued accessibility of digital materials. To do this involves finding ways to re-present what was originally presented to users by a combination of software and hardware tools acting on data.*" Digital preservation has emerged in recent years as a fast-moving and growing community of practice that is of ubiquitous relevance, but in which capability is unevenly distributed. Research disciplines have made a substantial and early contribution to the field, not least through investments of the European Commission. Digital preservation in the research community has a close alignment to the FAIR principles and is delivered, albeit unevenly, through a complex specialist infrastructure comprising not simply technology but also capacity of staff and 'know why' of policy[2].

The European Open Science Cloud (EOSC) initiative has extensively worked to promote and enable access to Open Science data with the stated aim of ensuring that researchers can maximize the value of their research processes and shared large-scale Research Infrastructures (RIs). **The importance of advanced long-term preservation to allow reproducibility of research results is emphasised by the EOSC Strategic Research and Innovation Agenda (SRIA) and different reports of relevant bodies such as the Digital Preservation Coalition.**

ARCHIVER, funded from the European Union's Horizon 2020 research and innovation programme, is providing a substantial contribution to this vision.

---

[1] https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation

[2] Currie, Amy, & Kilbride, William. (2021). FAIR Forever? Long Term Data Preservation Roles and Responsibilities, Final Report (Version 7). Zenodo. https://doi.org/10.5281/zenodo.4574234

Started in January 2019, ARCHIVER is a unique initiative currently running in the EOSC framework that is competitively procuring R&D services for archiving and digital preservation. The ARCHIVER tenderers were selected through an open and competitive procurement process. Between December 2020 and August 2021 three consortia worked on innovative, prototype solutions for long-term data preservation, in close collaboration with CERN, EMBL-EBI, DESY and PIC. Many NRENs (National Research and Education Networks) and research organisations - such as the European Institute of Oncology (IEO), ARDC, the Stockholm University Library, Jisc, SURF and others - have already expressed their interest in the ARCHIVER solutions by joining the group of Early Adopters, which will be part of the EOSC Marketplace.

Florian Pappenberger, Director of Forecasts at ECMWF
> *"I remember when getting forecast data was a long journey for researchers and here at ECMWF we could only collaborate with probably a handful of scientists. So we decided to have part of a new strategy in order to foster open data and collaborate with a wider scientific community and that's really where the ARCHIVER project is an essential cornerstone for the wider vision and that dream"*

Two consortia were selected to implement the Pilot solution
The selection process for proceeding to the next phase ended on the 3rd November with the final decision of the ARCHIVER PMT Board. The consortia selected to continue with the pilot phase were officially announced at the public ceremony on the 29th of November, virtually hosted by DESY, the German Electron Synchrotron based in Hamburg, one of the ARCHIVER Buyers.

The solutions which have been selected for the third phase of the project are the ones developed by the following consortia:
- Arkivum, using the Google Cloud, has proposed a solution for long-term data management and online access to address the challenges of how cloud hosted services can be used to store, manage, preserve and provide access to petabyte scale datasets.

"Digital content needs to last longer than any particular technology than any vendor than probably any data centre or anything else. For that matter, because we are talking about the open scientific community, I believe it's necessary to have a solution that isn't locked into a single vendor that can be deployed in various locations and in numerous different ways. The portability and no vendor lock-in actually provide our customers with genuine content sustainability." Chris Sigley, Arkivum CEO

- LIBNOVA, leading the CSIC, University of Barcelona, Giaretta Associates, AWS, Voxility and Bidaidea consortium, providing a Research Data Management and Preservation solution for the entire research content lifecycle and capable of managing large scale datasets in the order of 100s of Petabytes with low operational costs.

"Our solution is impacting the sustainability of digital preservation in multiple ways. First of all, it is increasing and improving the efficiency of the digital preservation process. We are using the most affordable technology with low storage costs but we are also allowing researchers to do way more with fewer resources because of the way the platform is organised. Second, we are also making the platform really easy to be used adopting the best possible practices; the trust principles and the best ISO standards are deeply embedded in the platform" [Antonio Guillermo Martinez, LIBNOVA](#)

The R&D produced by the consortia during the Prototype phase was validated by more than 200 tests executed across the contractors' platforms. The main criteria considered in the R&D review process took into account aspects such as performance to scale correctly in the PB data region, R&D validation progressing from functional to "Go-To-Market" ready, expertise in supporting Data Stewards achieving certification of scientific repositories and clear commercialisation and environmental strategies as a sustainable path for the resulting services after the end of the project.