

ARCHIVER

Archiving and Preservation for Research Environments

João Fernandes

CERN
Switzerland
joao.fernandes@cern.ch
<https://orcid.org/0000-0002-0445-7038>

Bob Jones

CERN
Switzerland
Bob.Jones@cern.ch
<https://orcid.org/0000-0001-9092-4589>

Jamie Shiers

CERN
Switzerland
Jamie.Shiers@cern.ch
<https://orcid.org/0000-0001-8404-3750>

Sara Pitonnet Gaiarin

TRUST-IT
Italy
s.pittonet@trust-it-services.com

Abstract – Do you need to acquire standards-based, cost-effective archiving and preservation services? Are ingest rates, data volume and long-term support important to you? The ARCHIVER project aims to introduce significant improvements in these areas of archiving and digital preservation services, supporting the IT requirements of European scientists developing end-to-end archival and preservation services for data generated in the context of scientific research. With a total procurement budget of 3.4 million euros, the project will use a Pre-Commercial Procurement (PCP) approach to competitively procure R&D services from firms in three stages covering design, prototyping and pilot, over a 3-year period (Jan 2019 - Dec 2021). The resulting services will become part of the catalogue of the European Open Science Cloud (EOSC) initiative funded by the European Commission (EC). This contribution will showcase the results obtained during the project phases up to iPRES2019.

Keywords – Archiving, Data Preservation, Pre-Commercial Procurement, Tender, EOSC

Conference Topics - Designing and Delivering Sustainable Digital Preservation, Exploring New Horizons

I. INTRODUCTION

Currently, many research projects struggle to manage their data, as the archiving and preservation services are inadequate and fall below expectations while data stewardship costs are frequently underestimated.

Using the EC Pre-Commercial Procurement (PCP) [1] instrument, the ARCHIVER project's goal is to fulfil these data management promises in a multi-disciplinary environment, allowing each research group to retain ownership of their data whilst leveraging best practices, standards and economies of scale. The objective is to

address the critical gaps between what is increasingly required by funding agencies, requested by data creators and eventual (re-)users and what is currently commercially available.

ARCHIVER will procure R&D services that address the archiving and long-term preservation needs of multiple scientific domains, implemented under the OAIS reference model and related standards [2] supporting FAIR [3] data principles.

II. ARCHIVER APPROACH

Acting as a collective of procurers, the ARCHIVER consortium aims to create an eco-system for specialist ICT companies active in archiving and digital preservation, willing to introduce innovative services capable of supporting the expanding needs of research communities, under a common innovative procurement activity for the advanced stewardship of publicly funded data in Europe.

These innovative services will be ready to be commercialized, by the end of the project (December 2021) potentially benefiting the research community at large in the context of the EOSC (European Open Science Cloud) [4]. The project is coordinated by CERN [5], with a consortium of procurer research organisations (CERN, DESY [6], EMBL-EBI [7] and PIC [8]) and experts (ADDESTINO [9] and Trust-IT [10]) and receives funding from the European Union's Horizon 2020 research and innovation programme [11].

A. ARCHIVER PCP phases

The invitation to tender of ARCHIVER will follow the implementation phases foreseen in the PCP instrument:

1) **Phase 1 - Solution Design:** Provision of a design report including architecture and technical design of components. The activity during this phase, will produce the results to be taken into account in the selection process that allows a contractor to proceed to the subsequent project phase. The expectation is to select multiple designs in order to promote competition and prevent vendor lock-in.

2) **Phase 2 - Prototype Development:** selected contractors from the Design Phase will build prototypes of the designed solutions and make them available to the procurer organisations forming the buyers group. During the Prototype Phase, basic functionality testing will be performed by specialists from the Procurer organisations. The results of these tests will drive the selection process for the last project phase (Pilot Phase).

3) **Phase 3 – Pilot Deployment:** Selected contractors will deploy expanded prototype services. These pilots will be essentially pre-production services to be tested in aspects such as performance, scalability and robustness. These services will potentially be exposed to end-users and early adopters, in order to determine if the resulting services are suitable for their needs. ARCHIVER will promote the development of the business models of the resulting services across all phases of the project, requesting contractors to provide total cost of ownership (TCO) and commercialisation plans for their solutions. This aspect is fundamental in order to provide a clear cost perspective to organisations that will purchase the resulting services at the end of the project.

B. *Requirements and R&D scoping*

Before project execution, a preparation phase takes place both for requirement assessment of the demand side and public consultation of the supply side. As announced in the corresponding Prior Information Notice (PIN) [12] published in the Official Journal of the European Union (OJEU) [13], an open market consultation process (OMC) will be carried out prior to the invitation to tender, in order to engage firms in an open dialog and exchange of information with the procurer organisations forming the buyers group. The OMC is organised as an evolutionary consultation process, formed by several sessions either targeted to the demand side or to suppliers, in order to estimate value, complexity and required R&D effort to support the foreseen use cases. The OMC, making use of the planning poker technique [14], will be led by Addestino, a specialist management and technology consulting firm partner in the ARCHIVER consortium. The main goal of the OMC is the identification of the innovation potential from a technological point of view. As an additional benefit, the suppliers will also acquire important information about the project use cases, the tender process and the different phases across the project. The results obtained during the OMC will be used as a basis for the Request for Tender specification to be published in October 2019.

In addition to the OMC itself during this period, training sessions will be organized on OAIS conceptual framework for

members of the ARCHIVER consortium. The main objective is to promote the reference model and to provide consortium members with the skills and knowledge necessary to respond to emerging issues in digital preservation from an organisational perspective. The Digital Preservation Coalition (DPC) [15], an external expert collaborating with ARCHIVER of which CERN is a member, will provide a range of opportunities for digital preservation training during several phases of the project.

III. ARCHIVER INNOVATION POTENTIAL

ARCHIVER will contract ICT companies to perform the research and development of innovative services to archive and preserve scientific data of multiple scientific domains. Although OAIS conformant solutions already exist in the market, it remains to be demonstrated functionality at the petabyte scale of scientific data from multiple research domains, with very high data ingestion rates (10Gbps/day), preferentially via commons solutions, interconnected via the GEANT network. These aspects combined with federated identity management services to support authentication and authorization to the archives and, the exploitation of a hybrid deployment model combining resources of the scientific organisations on-premises and commercial archiving and preservation services, can create the opportunity to innovate in a number of promising but untested aspects in order to develop the next generation of archiving and preservation services. In parallel, these services will be required to be compliant with the current regulations and legislations (e.g. GDPR [16]) and that the associated business models are transparent and well defined taking into account several factors: archives lifetime, maximum ingestion rates, data volumes, number of copies, data portability and exit strategies.

ACKNOWLEDGEMENTS

The authors would like to thank the ARCHIVER consortium members for the production of this contribution.

REFERENCES

- [1] Pre-Commercial Procurement process: <https://ec.europa.eu/digital-single-market/en/pre-commercial-procurement>
- [2] OAIS Reference Model: <http://www.oais.info/>
- [3] FAIR Principles: <https://www.go-fair.org/fair-principles/>
- [4] EC Directorate-General of Research and Innovation, Prompting an EOSC in practice, DOI: 10.2777/112658 (2018)
- [5] CERN: <https://home.cern/>
- [6] DESY: <http://www.desy.de/>
- [7] EMBL-EBI: <https://www.ebi.ac.uk/>
- [8] PIC: <https://www.pic.es/>
- [9] ADDESTINO: <https://www.addestino.be/>
- [10] TRUST-IT: <https://www.trust-it-services.com/>
- [11] Horizon 2020: <https://ec.europa.eu/programmes/horizon2020/en>
- [12] ARCHIVER Prior Information Notice: <https://ted.europa.eu/udl?uri=TED:NOTICE:87071-2019:TEXT:EN:HTML&src=0>
- [13] OJEU: <https://www.ojeu.eu/>
- [14] Planning poker: https://en.wikipedia.org/wiki/Planning_poker
- [15] Digital Preservation Coalition: <https://dpconline.org/>
- [16] GDPR Data Retention Guide: <https://www.dpnetwork.org.uk/gdpr-data-retention-guide/>