Archiving and Preservation for Research Environments

(ARCHIVER)

Joao.Fernandes@cern.ch
CERN

Bob.Jones@cern.ch
CERN

Jamie.Shiers@cern.ch
CERN

Abstract

Data has both a value and a cost and modern research data management makes many promises in terms of capacity, scalability, ease-of-use and security. The stewardship of research data involves not only all data-related tasks during the active lifetime of a project itself but also preparing the data and associated information for later re-use. The period during which research data remains valuable can stretch into decades. Currently, many research projects cannot manage their data, as the archiving and preservation services are inadequate and fall below expectations while data stewardship costs are frequently underestimated during the planning phase.

Using the PCP instrument and building on results of recent projects, ARCHIVER's goal is to fulfill 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. ARCHIVER will combine multiple ICT technologies, including extreme data-scaling, network connectivity, service inter-operability and business models, in a hybrid cloud environment to deliver end-to-end archival and preservation services that cover the full research lifecycle. The use-cases driving the consortium's need for research and development of innovative data preservation services will extend the preservation ecosystems of the procurers to create more dynamic solutions using a hybrid model combining on premise capacity with external services operated by commercial suppliers that will be enhanced to comply with the OAIS (ISO 14721) series of standards. One of the main benefits of such a hybrid approach is that it can be implemented in a way that is transparent to data producers and (re-)users. This transparency will address issues that cross discipline and national boundaries, such as findability and interoperability of datasets, as well as reduce costs. The potential uptake for the services resulting from this proposal are many-fold, including supporting the needs of ESFRI and related research infrastructures as well as the results of short-term research projects funded at the regional, national and European-level. The European Open Science Cloud is a major European undertaking that will provide this project with a privileged engagement channel with Europe's research communities who seek reliable and scalable solutions that satisfy the obligations of data management plans required by funding agencies.