ARCHIVER PROJECT: A SUCCESSFUL PUBLIC-PRIVATE COLLABORATIVE PROJECT

Antonio Guillermo Martinez
LIBNOVA SL
Spain
a.guillermo@libnova.com

Maria Fuertes
LIBNOVA SL
Spain
mfuertes@libnova.com

Abstract – The ARCHIVER Project has brought together customers, vendors, and infrastructure providers in an outstanding successful public-private collaborative project, which has also been recognized with the *Award for Collaboration and Cooperation* which celebrates significant collaboration across institutional, professional, sectoral and geographical boundaries at the Digital Preservation Awards 2022.

This paper will review, from the perspective of one of the project winners, the success story of the ARCHIVER Project, highlighting the benefits achieved by leveraging the commercial digital preservation solutions for scientific research data through a pre-commercial procurement process, where end users were able to directly influence the expected functionalities in the platform and how they are expected to operate.

Keywords – Digital Preservation, Research and Development, Collaboration, public-private.

Conference Topics – We're All in this Together.

INTRODUCTION

The ARCHIVER Project (Archiving and Preservation for Research Environments) is the only European Open Science Cloud (EOSC)-related H2020 project focusing on commercial long-term archiving and preservation services for petabyte-scale datasets across multiple research domains and countries [1].

On 29 January 2020, the ARCHIVER project launched its Pre-Commercial Procurement Request for Tenders [2] with the purpose to award several Framework Agreements and work orders for the provision of R&D for hybrid end-to-end archival and preservation services that meet the innovation challenges of European Research communities, in the context of the European Open Science Cloud.

COLLABORATIVE PROJECT

The ARCHIVER project is a clear example of public-private collaboration. Four of Europe's leading research organizations: CERN, EMBL-EBI, PIC/IFAE, and DESY formed a consortium to launch this project in which R&D was performed competitively by commercial providers LIBNOVA and Arkivum [3], through different implementation phases.

In the case of LIBNOVA, the public-private collaboration of the project was twofold, as in the first phase companies/organizations were invited to combine their skills and resources to form viable consortia to achieve the required results.

In this context and based on this recommendation, LIBNOVA formed a Consortium [4] that was enriched throughout the project with the incorporation of new members with expertise in the specific needs of each phase, forming a multidisciplinary cooperative and collaborative team, combining public sector organizations such as the University of Barcelona and the Spanish National

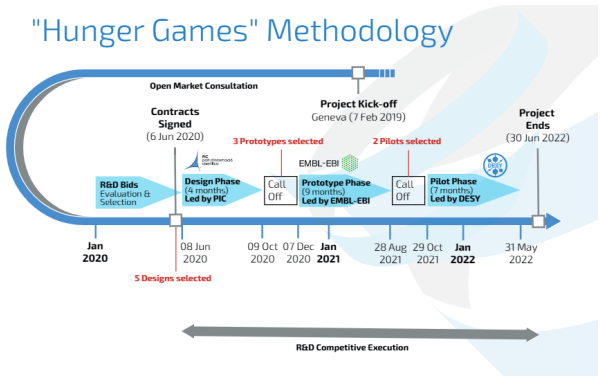
iPRES 2023: The 19th International Conference on Digital Preservation, Champaign-Urbana, IL, US.

https://hdl.handle.net/2142/121119

Research Council (CSIC), with private consulting, infrastructure and cybersecurity companies such as Giaretta Associates, Amazon Web Services, Voxility, and Bidaidea.

The R&D Methodology and Phases

In the Pre-Commercial Procurement model, R&D is divided into three phases (design phase, prototype phase, and pilot phase). Post-phase evaluations progressively identify solutions that offer the best value for the money and meet customer needs. Following a "Hunger Games" Methodology [5], where firms were selected or qualified for the next phase, or eliminated. This phased approach allows selected contractors to improve their bids for the next phase, based on lessons learned and feedback from buyers in the previous phase.



The work done at ARCHIVER, which has given rise to the LABDRIVE range at LIBNOVA, changes the approach taken to long-term research data management, both in terms of mindset and technology, i.e. what data researchers keep, how to maintain intellectual control of it, and what data stewards need to do to ensure that value can be derived from it in the long term. The companies selected by ARCHIVER promote environmentally sustainable solutions by providing the means to analyze and reduce the carbon footprint in the digital domain (big data centers).

A key component of sustainability is to ensure that the innovation developed during the project has broad exposure to potential buyers within the European research community and other business sectors. To achieve this, the project has initiated an onboarding process to make the resulting services available to early adopters. Making ARCHIVER services available through the EOSC marketplace will give researchers and contracting organizations the possibility to have sustainable access to these services, being able to test them, evaluate their functionality and purchase them with a clear cost model.

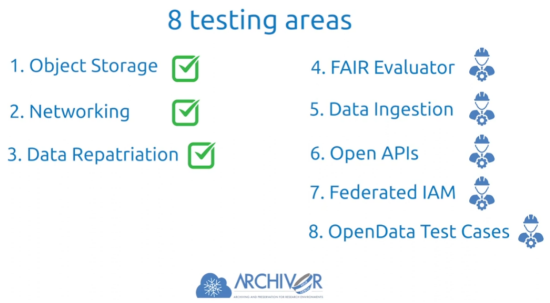
The ARCHIVER effort has resulted in services that can be used immediately by the public research sector in Europe. This will immediately expose novel service offerings, relevant to at least 18 pan-European infrastructures, to the 1.7 million European researchers and 70 million science and technology professionals, public and private sectors combined, who are expected to make use of the European Open Science Cloud (EOSC).

LIBNOVA has demonstrated the outcome of the ARCHIVER R&D activity to a wide group of potential users, both of the services developed and their potential for exploitation by the research community in EOSC [6].

LABDRIVE, the solution resulting from the ARCHIVER project

LIBNOVA has been the winner over all three phases of the project (design, protype and pilot), producing the LABDRIVE platform as the project result. LABDRIVE is a **Research Data Management** platform, that supports organizations in their data management endeavors.

During the ARCHIVER project, LABDRIVE has been tested and confirmed to work with High Energy physics, Astrophysics, Life Sciences and other types of large datasets (millions of files and tens of PBs) against 176 combinations of use cases, volume tests, researcher needs and organization requirements, **confirming suitability and scalability of the platform for multiple Research Data Management use cases and needs.**



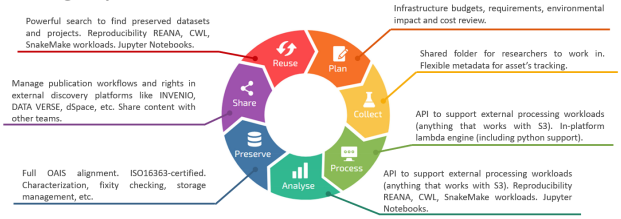
LABDRIVE is cloud-native, allowing Organizations to leverage the public/private cloud adoption if this is an objective. If not, the platform can also be deployed on premises or hybrid cloud/on premises scenarios.

While the LIBNOVA LABDRIVE platform has been re-architected for massive scalability and specific Research Data Management use cases during the Archiver project, LIBNOVA has been the community's trusted partner for digital preservation and data management for several years. Organizations like Stanford University (HILA), Princeton University, Oxford University, The British Library, Pennsylvania State University, Bayer and many other organizations in 17 countries are already LIBNOVA customers.

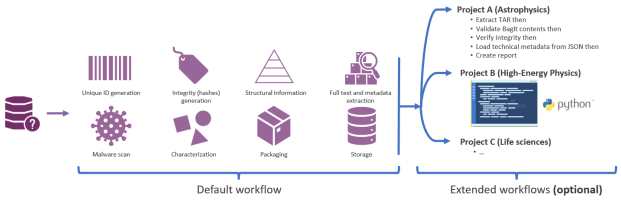
LABDRIVE is a Research Data Management and Preservation platform. It allows organizations to capture the research data they produce, helping them to properly manage, preserve and allow access to it, during the whole data lifecycle.

Design principles.

LABDRIVE provides support over the whole data lifecycle: It allows organizations to capture the research data they produce at the initial stages of the project ("shared folder"), enabling them to properly manage, preserve, reuse and allow access to it:



LABDRIVE works with many research disciplines and content types: It includes a default processing workflow, but it can be extended –using python- to support any other use case. Metadata schemas, data structures, permissions, storage, etc. can also be defined per project, so it can be adapted to multiple scenarios:



LABDRIVE is fully aligned with most relevant and open standards: Fully aligned to the FAIR and TRUST principles [7]. Fully conformant with OAIS [8] and fully aligned with the ISO 16363 [9]. Likewise, ISO 27001, ISO 27017 and ISO 27018-certified. GDPR compliant.

Built on open standards. Aligned with the FAIR and TRUST principles.

Full OAIS/ISO14721 and ISO16363 Alignment. No exit barriers: 100% of the information can be easily migrated.

High security and reliability: ISO27001, ISO27017, ISO27018 certified and externally audited yearly.

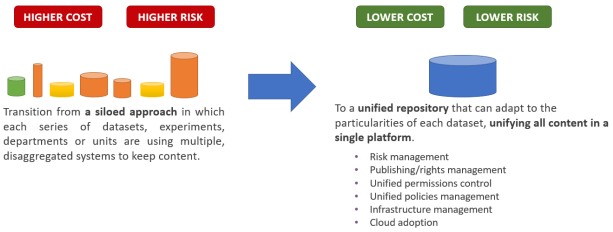
GDPR compliant (DPIA for sensitive content).

LABDRIVE equally supports power users and simplified use cases: Every action in the platform can be carried out using the easy-to-use web browser interface or the 300-ish Open API methods and 80+ CLI tools available.

Easy to use web interface for users looking for a simplified experience.

Powerful API support, compatible with 80+ CLI tools. Sample Bash scripts, python libraries, etc. available.

As a result, LABDRIVE allows organizations to organize, unify and simplify their research data management strategies, transitioning from a siloed approach to a unified and cohesive platform, obtaining lower risks and lower costs back:



LESSONS LEARNED AND BENEFITS FOR PUBLIC-PRIVATE COLLABORATION

Based on the gathered practical experience, a set of lessons learned and best practices can be taken as reference for future PCPs covering aspects such as the procurement process, R&D execution and dissemination of the R&D activities for maximization of results impact by the end of the project [5].

The highlights can be summarized as follows:

- Procurement would benefit if reduced in time and complexity, and focused more on the R&D challenge, as European innovative software SMEs "think" in months rather than years.
- Structured feedback across all parties is found essential, in order to allow full understanding of the challenge.

- The Agile software development methodology can prove to be very effective if a roadmap for the R&D strategy is produced as a wider frame for expectations, with feature prioritization to avoid possible mismatches in the understanding of the challenge.
- Effort for the tasks of requirement gathering, tender evaluation, assessment and testing of the R&D remains very significant.
- A dissemination plan articulated between the project participants boosts visibility and reach across different communities, sectors and stakeholders.
- PCPs for software services would very much benefit from structured incentives to ramp up the results (for example in the EOSC context), sustaining access to the resulting SaaS and fund trials from researchers in view of purchasing the services if trial deployments are successful.

Overall, the project has demonstrated how the PCP instrument can incite expert SMEs to develop innovative services that can satisfy the needs of Europe's research communities and paves the way to explore more effectively the integration of commercial services into the EOSC marketplace

The work accomplished in ARCHIVER is considered a game-changer for the approach taken to long-term Research Data Management both from a mindset and technological perspective, i.e. what data do researchers retain, how to keep intellectual control of it and what data stewards must do to ensure long-term value can be realized from it.

Thanks to the ARCHIVER Project, the winning companies gain experience from working within public procurement. These are the relevant benefits and tangible results of this collaboration:

- Shorter development life cycle leading to faster time to market, from 5 to 2 years, giving these participating companies an advantage in relation to other competitors.
- Increased customer base portfolio not only in Europe but with contracts signed with universities and other institutions in North America.
- Maximization of the understanding of requirements being able to work with multidisciplinary use cases.
- Pushing the boundaries of what digital preservation is, incorporating innovations that improve the products and empower other organizations to preserve data, consequently creating stronger relationships and incremental business.
- Partnership agreement with hyperscalers (e.g. AWS and Google) strengthening the business perspectives of these European SMEs.
- Increase of services sustainability with a special focus on environmental sustainability.
- Acceleration and de-risking of the ability of these companies to enter a new market with innovative services that address the problem of long-term digital preservation and access to scientific research datasets.

To summarize, the ARCHIVER project has accomplished significant work on technological solutions, its economics and business models, in a holistic manner across scientific domains, public/private sectors and geographies, consistent with the evolving Open Science policies in Europe.

By working directly with the public sector organizations, LIBNOVA and Arkivum were able to receive ongoing input and feedback into their product development to serve the mission of scientific research within Europe, enabling these SMEs to quickly and effectively develop fit-for-purpose products. This resulted in innovative commercialization approaches for the resulting services, improving their degree of FAIRness as an aspect of utmost importance for the ultimate objective of the reuse of research data.

The focus of initiatives such as ARCHIVER is Data, in particular research data, that is set to live for longer than any vendor, system or technology.

Conclusion

The ARCHIVER Project has brought together customers, vendors and infrastructure providers in an outstanding successful public-private collaborative project, which has also been recognized with the Award for Collaboration and Cooperation which celebrates significant collaboration across institutional, professional, sectoral and geographical boundaries at the Digital Preservation Awards 2022 [10]. The award was given to the European ARCHIVER project for what the

judges called "important public-private partnership work that could pave the way for the long-term digital preservation of research data."

The ARCHIVER Project has been a technological breakthrough in the solutions offered by LIBNOVA. In addition to shortening the development times involved in the creation of digital preservation software of the characteristics of LABDRIVE, it allows LIBNOVA to reach market segments that had not been addressed before and to face the design of sustainable digital solutions from a solid position.

REFERENCES

[1] Archiving and Preservation for Research Environments | ARCHIVER Project | Fact Sheet | H2020 | CORDIS | European Commission https://cordis.europa.eu/project/id/824516

[2] ARCHIVER launches its Pre-Commercial Procurement Tender https://cordis.europa.eu/article/id/413444-archiver-launches-its-pre-commercial-procurement-tender

[3] ARCHIVER PROJECT | PILOT PHASE AWARD - THE TWO WINNERS https://archiver-project.eu/pilot-phase-award

[4] ARCHIVER Project | Consortium 3 https://archiver-project.eu/consortium-3

[5] ARCHIVER White Paper. Zenodo. https://doi.org/10.5281/zenodo.7691976

[6] EOSC Marketplace - LIBNOVA LABDRIVE: The Ultimate Research Data Management and Digital Preservation Platform https://marketplace.eosc-portal.eu/services/libnova-labdrive-the-ultimate-research-data-management-and-digital-preservation-platform

[7] LABDRIVE support for FAIRness https://docs.libnova.com/labdrive/concepts/oais-and-iso-16363/labdrive-support-for-fairness

[8] LABDRIVE support for OAIS Conformance https://docs.libnova.com/labdrive/concepts/oais-and-iso-16363/labdrive-support-for-oais-conformance

[9] LABDRIVE - ISO 16363 certification guide https://docs.libnova.com/labdrive/concepts/oais-and-iso-16363/iso-16363-certification-guide

[10] Digital Preservation Awards 2022 – Winners Announced! https://www.dpconline.org/events/digital-preservation-awards/the-winners