

# HOW TO PRESERVE RESEARCH DATASETS

## *LABDRIVE Tutorial*

**Antonio Guillermo  
Martinez**

*LIBNOVA SL*  
*Spain*  
*a.guillermo@libnova.com*

**Maria Fuertes**

*LIBNOVA SL*  
*Spain*  
*mfuertes@libnova.com*

**Abstract** - LABDRIVE is a Research Data Management and Digital Preservation platform that allows organizations to capture the research data they produce, helping them to properly manage, preserve and allow access to it, during the whole research data lifecycle, unifying and simplifying their research data management strategies.

The purpose of this tutorial is to introduce the design principles of LABDRIVE as well as explain how it works through a tutorial (a guided demonstration).

**Keywords** - Research Data Management, Digital Preservation, Software

**Conference Topics** - Immersive information, From theory to practice

### I. INTRODUCTION

The EU-funded Archiver project [1] initiated a market consultation project looking for Research Data Management platforms capable to scale to the 100's of PBs in 2019. The conclusion of the market consultation was that there were neither viable nor cost-efficient platforms in the market at the time.

With the objective of helping software/platform providers to meet the need and after requesting approx. 6M€ of EU funding, a consortium led by the CERN (European Center for Nuclear Research), EMBL (European Molecular Biology Laboratory), PIC (Port d'Informació Científica - MAGIC Radio telescopes) and DESY (Deutsches Elektronen-Synchrotron) created a set of large scale data sets and use cases and initiated a Pre-Commercial Procurement (PCP) approach to competitively procure R&D services

from firms in three stages, covering design, prototyping and pilot over the following 3 years.



LIBNOVA has been one of the winners over all three phases of the project (design, prototype and pilot) [2], producing the LABDRIVE platform as the project result. LABDRIVE is a Research Data Management platform that supports organizations in their data management endeavors [3].

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 as 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.

## II. THE LABDRIVE PLATFORM

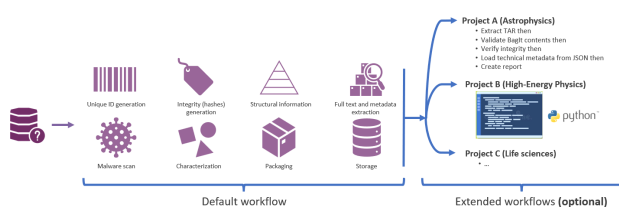
LABDRIVE is a Research Data Management and Preservation platform. It allows organizations to capture the research data they produce, helping them to properly manage, preserve and allow access to it, during the whole data lifecycle.

### 1. Design principles

1) *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.



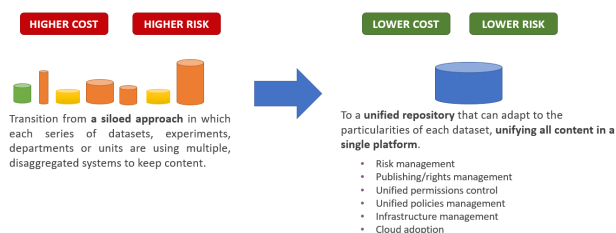
2) *LABDRIVE works with many research disciplines and content types*: It includes a default processing workflow, but it can be extended –using python– to support any other use case. Metadata schemas, data structures, permissions, storage, etc. can also be defined per project, so it can be adapted to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



3) *LABDRIVE is fully aligned with most relevant and open standards*: Fully aligned to the FAIR and TRUST principles [4]. Fully conformant with OAIIS [5] and fully aligned with the ISO 16363 [6]. Likewise, ISO 27001, ISO 27017 and ISO 27018-certified. GDPR compliant.

4) *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.

5) *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:



## TUTORIAL CONTENT

The contents would be divided into 3 blocks and would be roughly as follows:

1. LABDRIVE Introduction
2. How it works: LABDRIVE Configuration
3. How to preserve research data: LABDRIVE Operations

## 2. REFERENCES

- [1] ARCHIVER Project <https://www.archiver-project.eu/>
- [2] ARCHIVER PROJECT | PILOT PHASE AWARD - THE TWO WINNERS <https://archiver-project.eu/pilot-phase-award>
- [3] 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>
- [4] LABDRIVE support for FAIRness <https://docs.libnova.com/labdrive/concepts/oais-and-iso-16363/labdrive-support-for-fairness>
- [5] LABDRIVE support for OAIIS Conformance <https://docs.libnova.com/labdrive/concepts/oais-and-iso-16363/labdrive-support-for-oais-conformance>
- [6] LABDRIVE - ISO 16363 certification guide <https://docs.libnova.com/labdrive/concepts/oais-and-iso-16363/iso-16363-certification-guide>