ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

A Research, Management and Preservation Platform

A Research, Management and Preservation Platform combining existing technologies and new components, to overcome obstacles for research dataset management (including preservation).

The proposed solution

The solution developed by the consortium led by LIBNOVA provides a Research, Management and Preservation Platform combining existing technologies and new components, to solve obstacles for research dataset management (including preservation) identified in the ARCHIVER project. The solution proposed is based on pre-existing digital preservation platforms already in use by many leading organizations across the world. The solution is designed for the whole organization and for the whole data life-cycle, completely aligned with OAIS, ISO16363, FAIR and TRUST principles, with powerful and really innovative capabilities in all functionality layers.

The R&D potential

- **Scalability** (sustained high throughput in the 100s of PBs range)
- **Digital Preservation Best Practices** (OAIS, ISO 16363, PSC-Preservation Storage Criteria, Best Practices recommendations and implementations – OAIS Information Model including RepresentationInformationand Preservation Description Information components, problem detection such as duplicates, hidden encryption, format migration/evolution, exit strategy)
- **Metadata management** (import/creation and preservation), following OAIS
- **Data integrity management** (integrity chain, integrity at rest)
- **FAIR principles** (F: containers, customized metadata, structured hierarchy, A: multiprotocol access, public sharing, discovery solutions,
- I: Data policies, research data Representation Information, R: Integrated active integrity control, Representation Informationand Provenance Information)
- **Cost efficiency** (flexibility on deployment, several computation/storage options)

Architecture overview

The overall architecture is built on two main components:
- **Core software components (Group A)** running inside Kubernetes containers. The number of containers running in parallel of each class can be adjusted manually or automatically by the platform based on service demand to ensure full scalability.
- **Auxiliary services (Group B)** based on the Core services above. When running in "on-prem" mode, organizations will need to provide them for the platform to work. When running "as a service", the service provider (LIBNOVA) will provide them.

Five assets are deployed on the components above:
- **Containers** – keep content accessible with several protocols, organized and protected. These containers keep metadata, data and code together to ensure usability (OAIS-aligned).
- **Dynamic Insights** – help users when dealing with personal information, digital preservation and emissions reduction, with the following components: Data Policies Assistant, GDPR Assistant, Emissions Optimizer, Digital Preservation
- **Budget assistant** – helps users to plan and follow expenditures
- **Content gateway** – connects the platform with repositories for discovery solutions, such as Invenio or Dataverse
- **Digital Preservation, OAIS and FAIR conformance** – as support for the OAIS Information Model and for the Mandatory Responsibilities, and the results will fully support repositories in OAIS conformance. The focus on usability is also critical for the "Interoperability" and "Reusability" required by the FAIR principles

Comparison between current baseline and R&D being performed

Date - January 2021

Baseline before ARCHIVER	ARCHIVER R&D
Storage/basic archiving/secure backup (Layer 1) Deployments over private, public, hybrid, community and special purpose clouds in the single PB range infrastructure	Infrastructure agnostic for multiple PB; Multitenancy; sustained data ingest rates 1-10Gb/s for multiple use cases with different access patterns.
Preservation (Layer 2) Preservation services of files at basic level of redundancy, limited API support	Richer API set (essentially all capabilities available via the GUI for seamless integration); Active monitoring of data integrity in order to detect unwanted changes such as file corruption or loss on top of infrastructure services. Support for handling unstructured or missing metadata, test models to map responsibilities for local support, responsibilities for long-term data management planning;
Baseline user services (Layer 3) Volumes of hundreds of TBs with support of Indexing, elastic search, deduplication.	Software development for search, look up or filter potential datasets rapidly, to access dataset metadata and decide on its relevance.
Advanced Services (Layer 4) Basic support of retention and integrity of certain types of data	Container Orchestration engine support based on Kubernetes for the compute capabilities to allow scientific analyses to be carried out off-prem. Interfaces from infrastructure layer integrated on the overall design (allow access to data, no matter where stored);

Watch the video > Prototype Phase winners interview - Antonio Guillermo Martínez (LIBNOVA)
https://www.youtube.com/watch?v=yEeoLNncfCI