



# ARCHIVOR

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

## Innovative Long-Term Data Preservation Services for the EOSC

**OS Fair 2021**

João Fernandes (CERN)

**Ignacio Peluaga (CERN)**

Sara Pittonet (Trust-IT)

*September 21st 2021*



ARCHIVOR - Archiving and Preservation for Research Environments project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.

# ARCHIVER Project

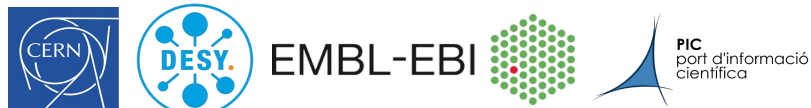
***Focus:*** Archiving and Data Preservation Services using cloud services available via the European Open Science Cloud (EOSC)

***Procurement R&D budget:*** 3.4M euro; ***Total Budget:*** 4.8M

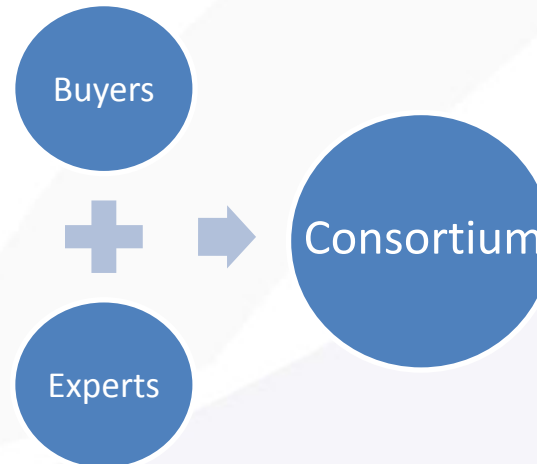
***Starting Date:*** 1<sup>st</sup> of January 2019

***Duration:*** 42 Months

***Coordinator:*** CERN (Lead Procurer)



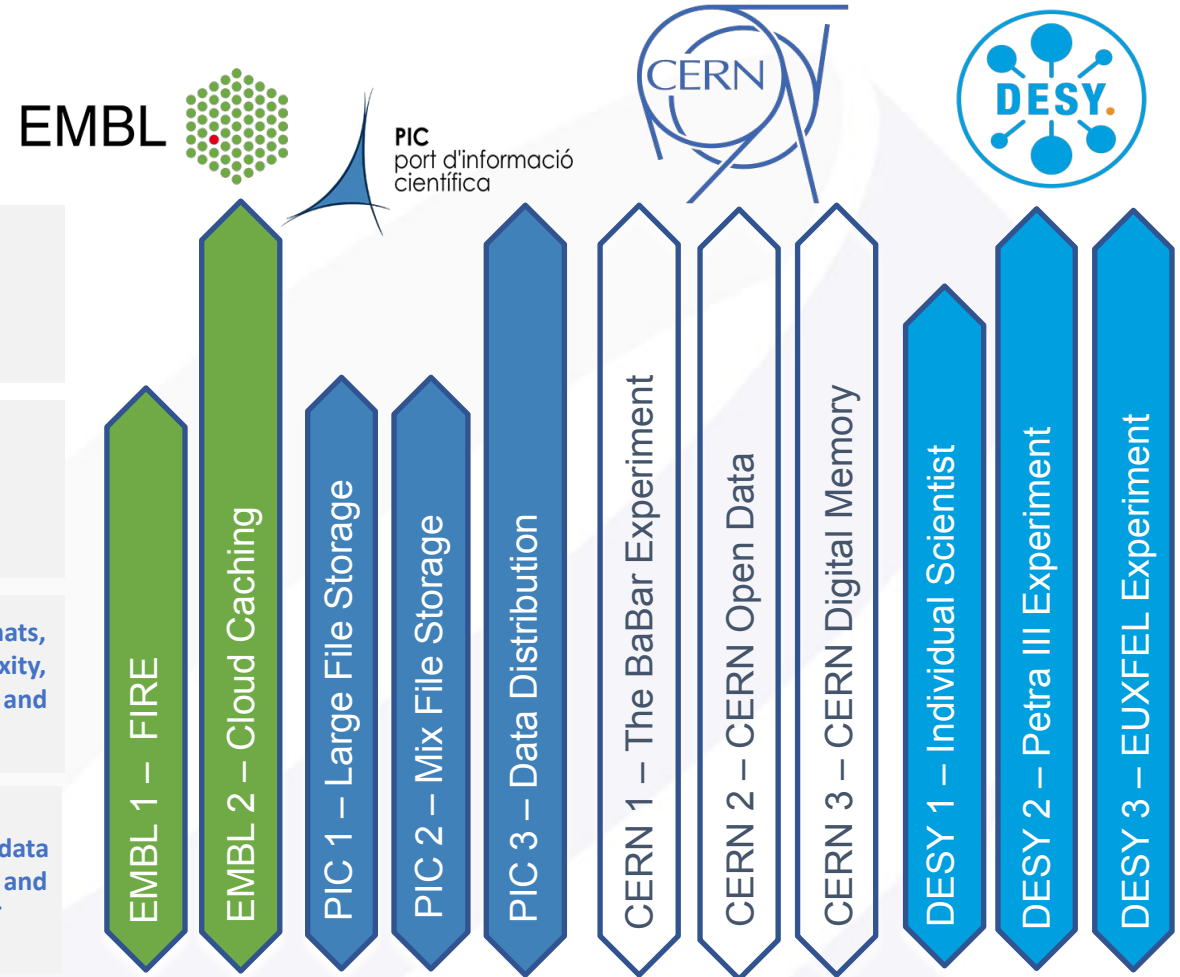
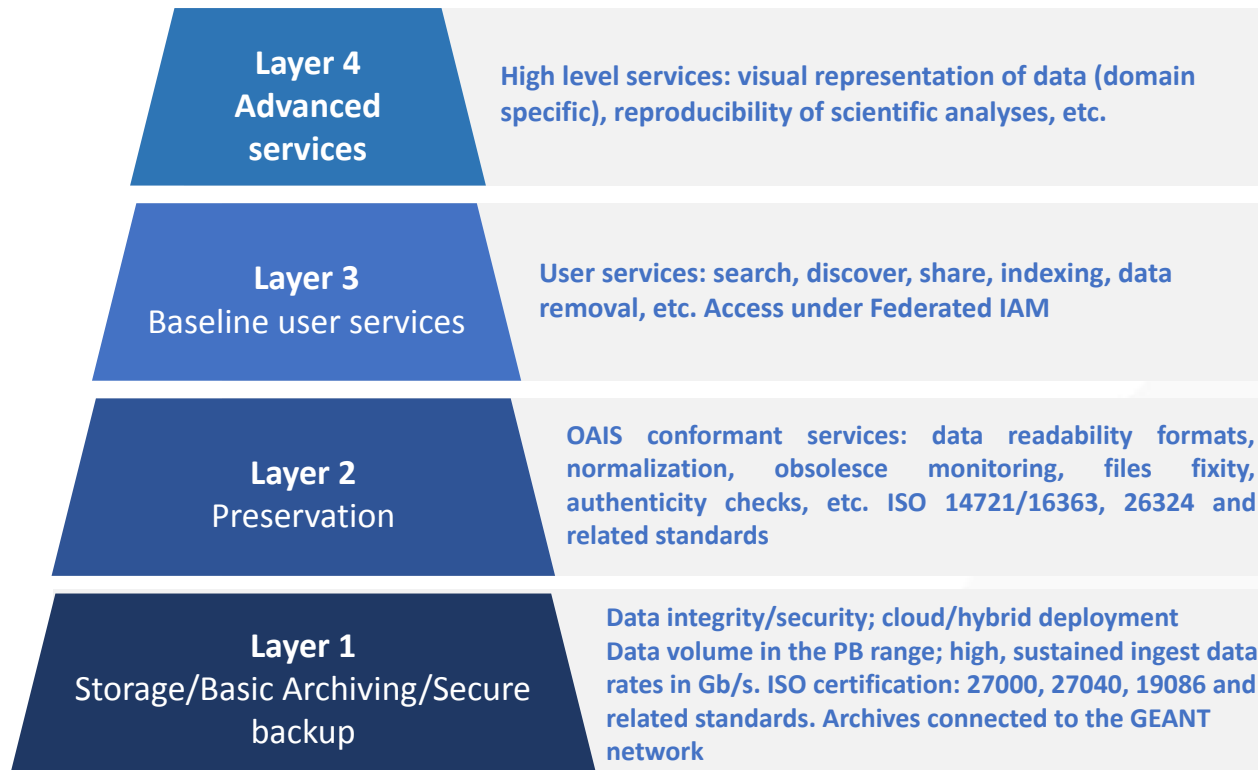
**Buyers Group (BG)** - Public organisations committing funds to contribute to a joint-R&D-procurement, research data use cases and R&D testing effort



**Experts** - Partner organisations bringing expertise in requirement assessment and promotion activities



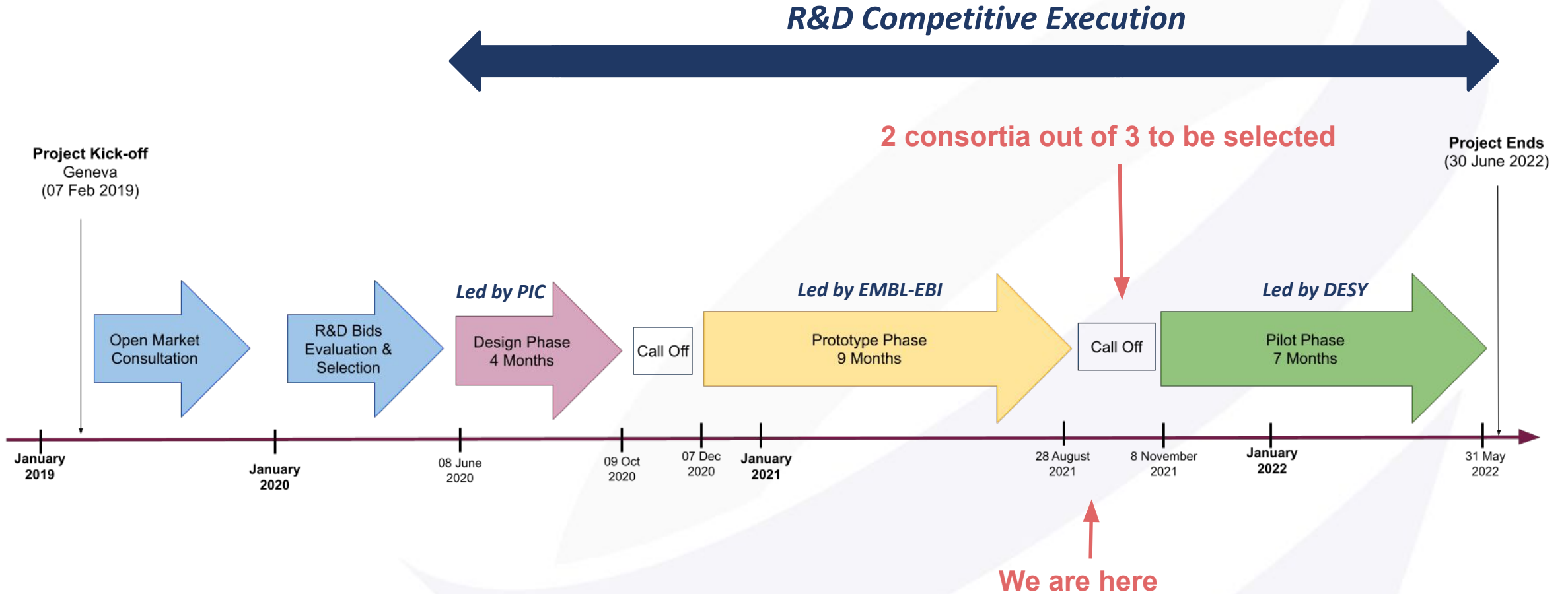
# R&D - Use Cases



Scientific use cases deployments documented at: <https://www.archiver-project.eu/deployment-scenarios>

ARCHIVER “current state of the art” report in the context of the EOSC: <https://doi.org/10.5281/zenodo.3618215>

# Project Timeline



# Prototype Phase Selected Consortia



arkivum

Bringing archived data to life



Google Cloud

  
libnova



UNIVERSITAT DE  
BARCELONA



CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS

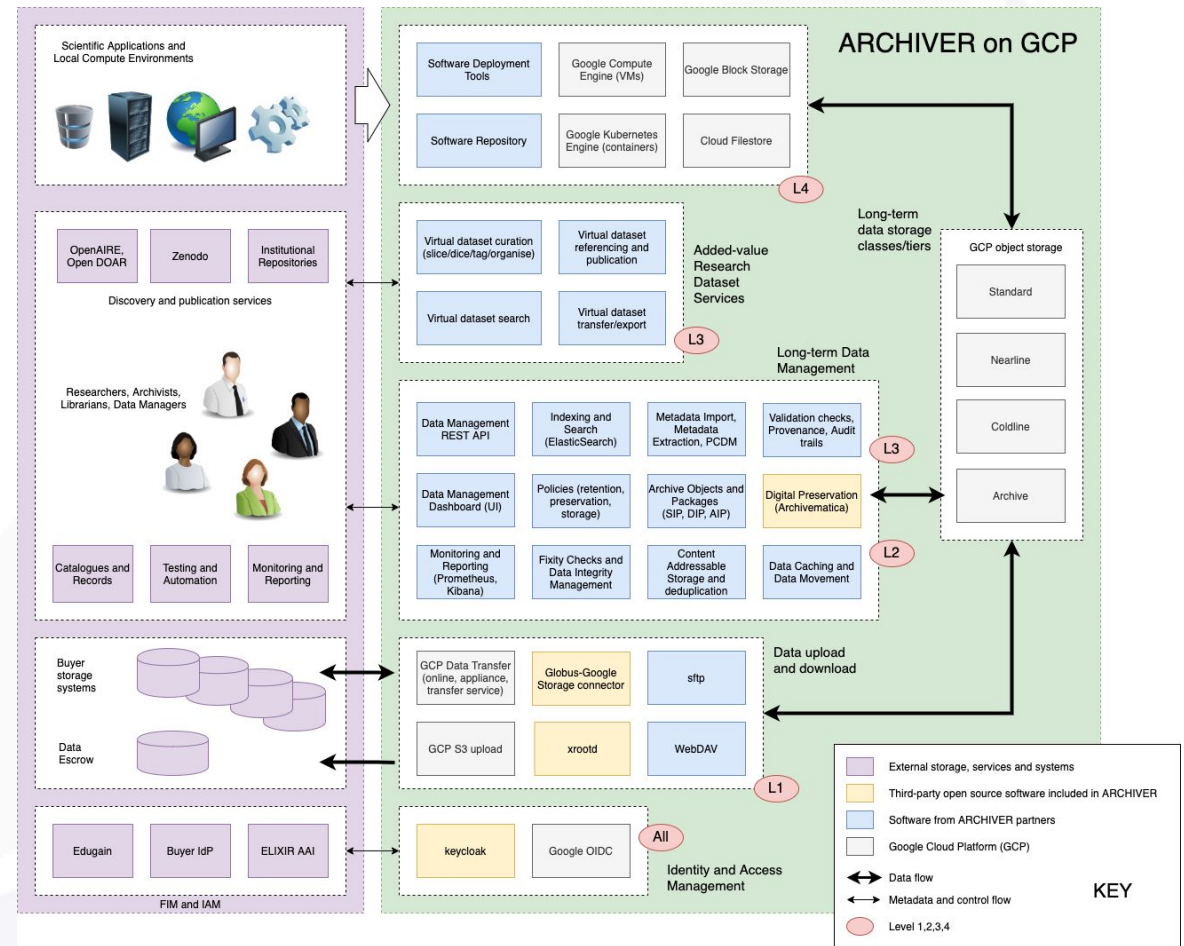


T-Systems



# Selected Consortia: Arkivum

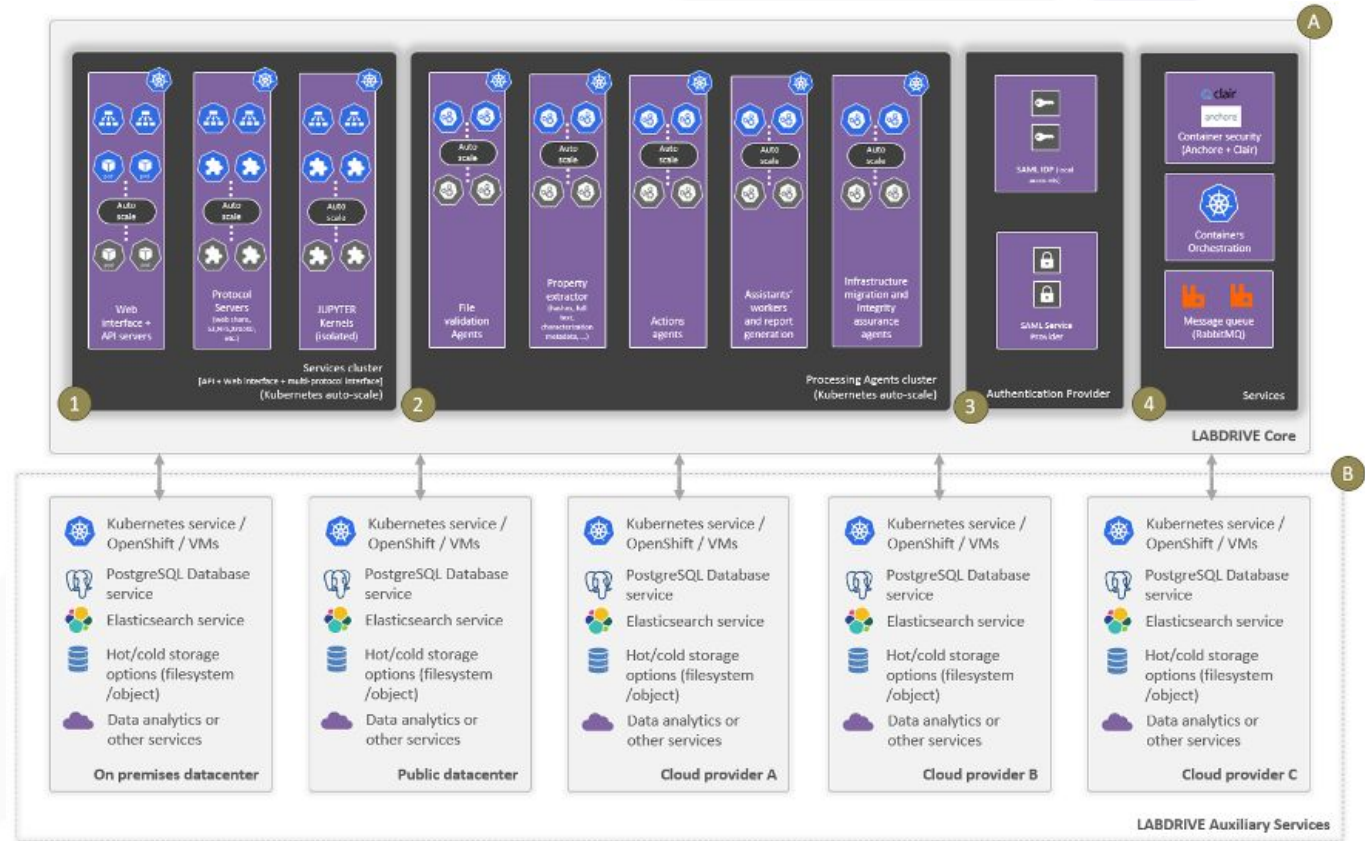
- Overall architecture composed of micro-services to scale to multi-petabyte volumes of billions of objects. Based on Kubernetes on GCP
- Service-oriented SaaS stack
- The solution can be deployed on-premises or in a hybrid cloud configuration
- Archivematica as an option



Prototype architecture of the Arkivum consortium (image courtesy of the Arkivum consortium)

# Selected Consortia: Libnova

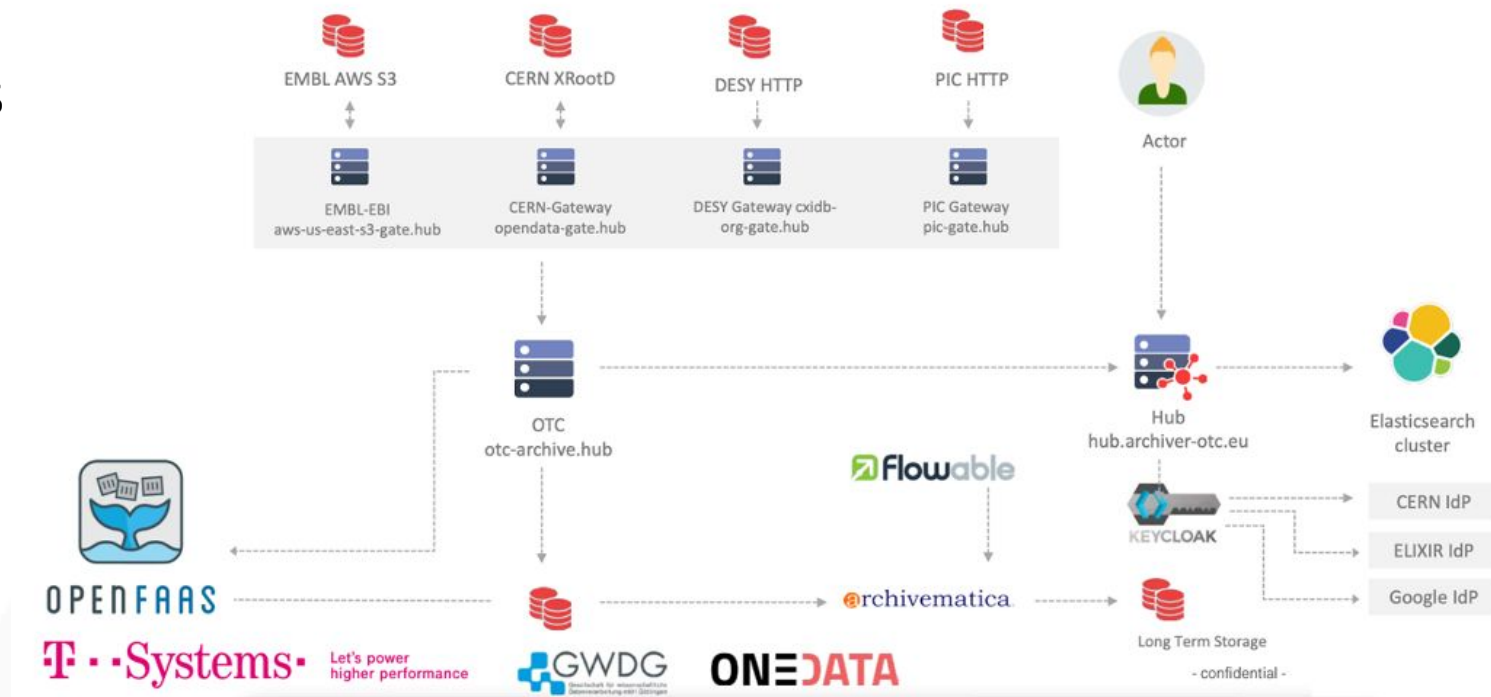
- Prototype based on proprietary LibSAFE SaaS
- Using infrastructure provided by AWS and Voxility
- All capabilities in the UI available via API
- Active monitoring to control data loss
- A: Software components running inside Kubernetes containers. Adjustable number of containers based on service demand to ensure full scalability
- B: Auxiliary services



Prototype architecture of the Libnova consortium (image courtesy of the Libnova consortium)

# Selected Consortia: T-Systems

- Full gap analysis performed during the Design phase, resulting in a modern OS architecture: Onedata, OpenFaaS, Flowable and OTC infrastructure
- Kubernetes-based platform
- BPMN based workflow Management
- The components and R&D performed will be made available with OSS licensing

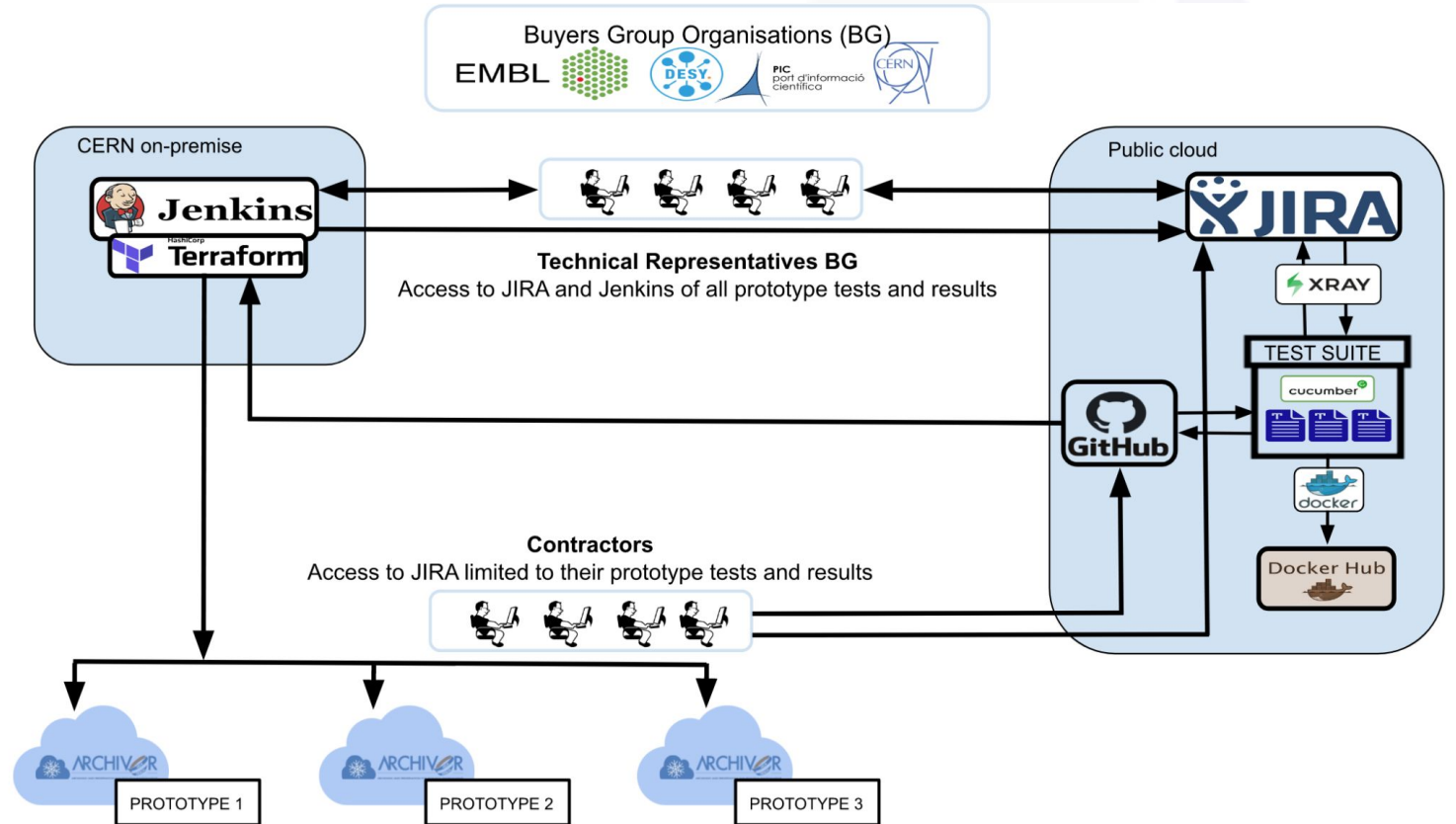


Prototype architecture of the T-Systems consortium (image courtesy of the T-Systems consortium)



# Testing during prototype phase

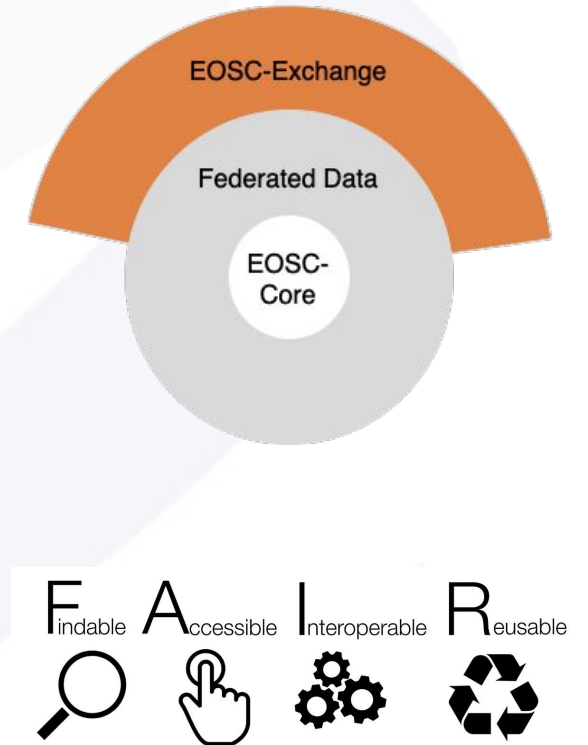
- The phase's main activity was to test the 3 proposed prototype platforms
- Following an Agile methodology, the BG carried out testing activities of all the platforms:
  - Functionalities and capabilities
  - Networking - perfSONAR
  - Federated IAM
  - On-prem deployment
  - Initial exploration of FAIRness - F-UJI
  - High Volumes (hundreds of TB)



Continuous Testing Environment for ARCHIVOR's testing activities

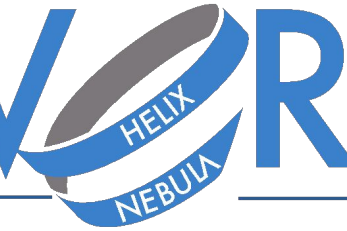
# Conclusions

- Pilot phase to start in November: **2 consortia** to be selected from the current 3
- The R&D challenge of digital archiving goes **beyond data storage**: keep intellectual control of data and associated products for decades, make research outputs reusable
- Extending **FAIR** to research associated products: software, workflows, services and even infrastructures
- ARCHIVER is acting as a template to **commoditise** archiving and preservation in research domains
- ARCHIVER is promoting a **sustainable model** with services that will exist beyond the project lifetime in the context of the **EOSC**





ARCHIVER



ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

# Thank you! Questions?



[info@archiver-project.eu](mailto:info@archiver-project.eu)



<https://www.archiver-project.eu/>



<https://twitter.com/ArchiverProject>

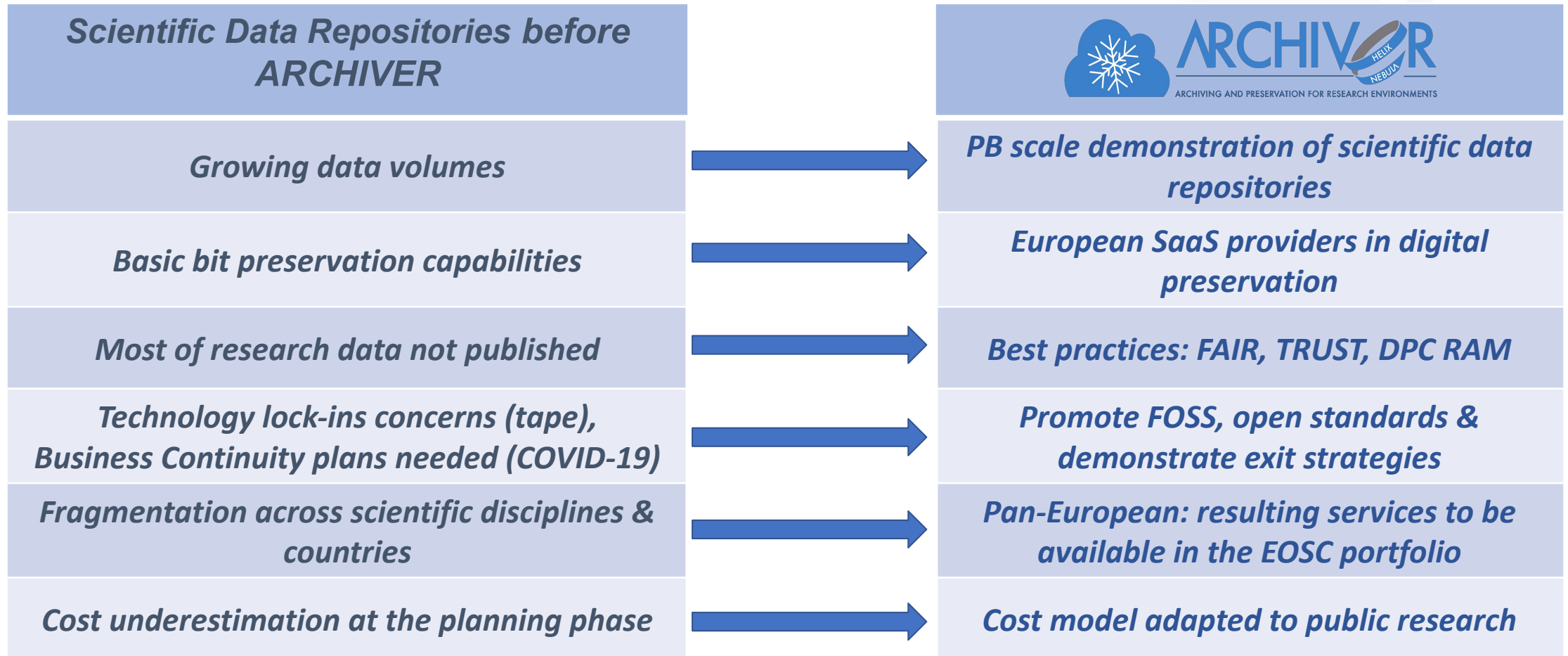


<https://www.linkedin.com/company/archiver-project/>



<https://www.youtube.com/channel/UCCBIyLpUt-hWmQatqdlhlzw>

# Backup Slide - Progress Beyond the state of the art



ARCHIVER “current state of the art” report: <https://doi.org/10.5281/zenodo.3618215>

# Backup Slide - R&D Execution

ARCHIVER is following an implementation on **three** phases with multiple competing **consortia**:

- **Phase 1 - Solution Design**
  - The 5 selected consortia develop designs including architecture and technical components
  - The activity during this phase has produced the results to be taken into account in the selection process that allows a consortium to proceed to the subsequent project phase
- **Phase 2 - Prototype Development**
  - The 3 selected consortia from the Design Phase are building prototypes based on the designed solutions
  - Make them available to the buyers group for testing purposes
- **Phase 3 - Pilot Deployment**
  - The 2 selected consortia will deploy expanded prototype services
  - These services will potentially be exposed to end-users and early adopters, to determine if they are suitable for their needs