ARCHIVER

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

ARCHIVER – Design Phase

Archiving and Preservation for Research Environments

Jakub Urban (CERN)
Technical Coordinator

2nd of November 2020

ARCHIVER - 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.

Project

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

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

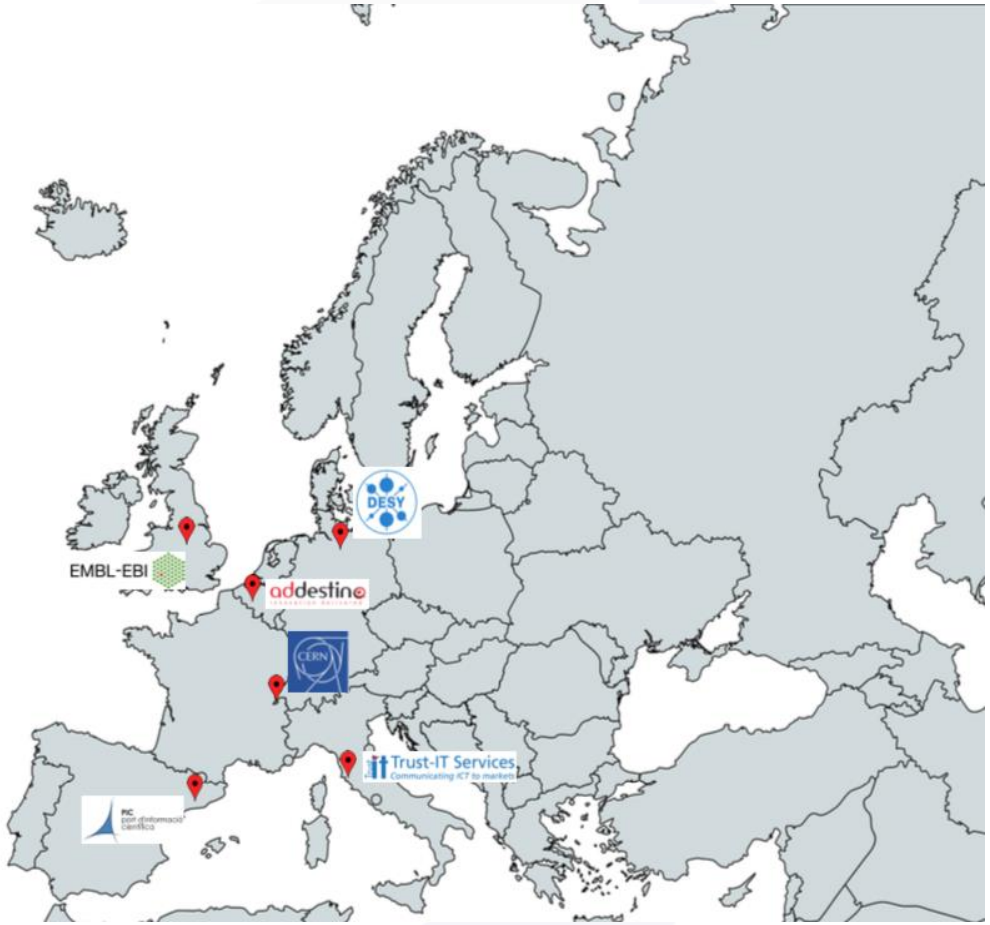
Starting Date: 1st of January 2019

Duration: 41 Months

Coordinator: CERN (Lead Procurer)

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

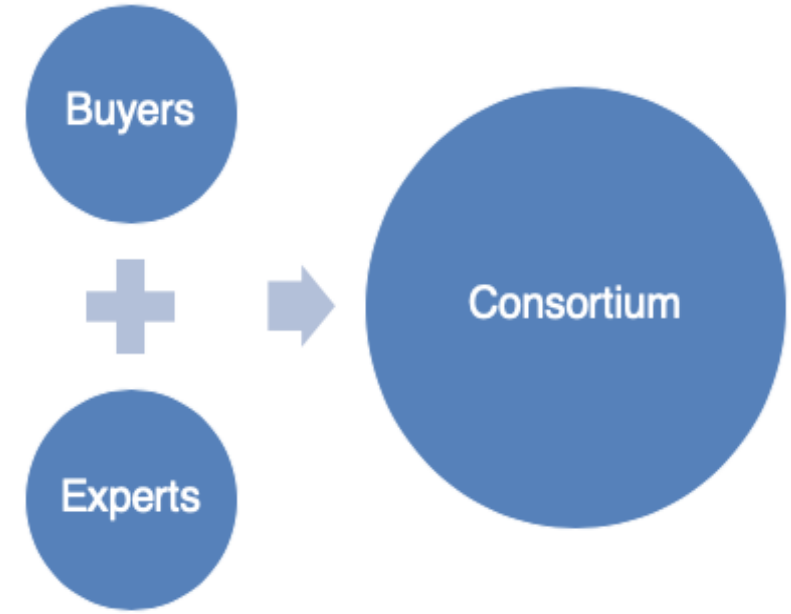
European Commission



Consortium

Includes Buyers and Experts in the preparation, execution and promotion of the procurement of R&D services

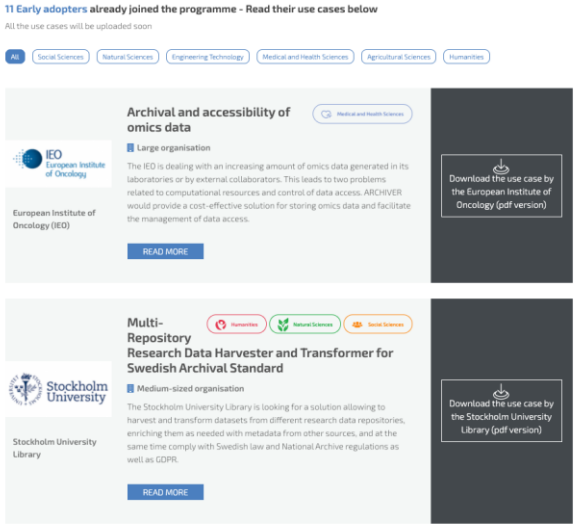
The “Buyers Group”: Public organisations committing funds to contribute to a joint-R&D-procurement, research data use cases and R&D testing effort



Experts – Partner organisations bringing expertise in requirement assessment and promotion activities, not part of the Buyers Group

Early Adopters

https://archiver-project.eu/early-adopters-use-cases



- **Participants:**
  - Demand side public sector organisations
- **Key advantages**
  - Access and assess if resulting services address archiving and preservation meet their needs
  - Contribute and shape the R&D carried out in the project, contribute with use cases and
  - Have the option to purchase pilot-scale services by the end of the project
- **Confirmed 11 organisations, more are in the process:** → ***High level of interest from the community***

CRG Centre for Genomic Regulation

IEO Istituto Europeo di Oncologia

SURF SARA

aarnet Australia's Academic and Research Network

SWITCH Serving Swiss Universities

KAUST

Universitet Stockholms

FMI Friedrich Miescher Institute for Biomedical Research

UKRI Science and Technology Facilities Council

ARDC Australian Research Data Commons

Jisc

Move from current state of the art

Current Scientific Data Repositories

- ***Growing data volumes***
- ***Basic bit preservation capabilities***
- ***Concerns: technology lock-in (tape), Disaster Recovery/Business Continuity plans needed (COVID-19)***
- ***Most of research data not published***
- ***Fragmentation across scientific disciplines & countries***
- ***Cost underestimation at the planning phase***

- ***PB scale demonstration of scientific data repositories***
- ***Profit from considerable experience of European SMEs preservation experts***
- ***Promote FOSS, open standards & concretely test exit strategies***
- ***Best practices: FAIR, TRUST, DPC(RAM)***
- ***Pan-European: resulting services available in the EOSC***
- ***Cost model adapted to public research***

ARCHIVER "current state of the art" report: https://doi.org/10.5281/zenodo.3618215

R&D Scope

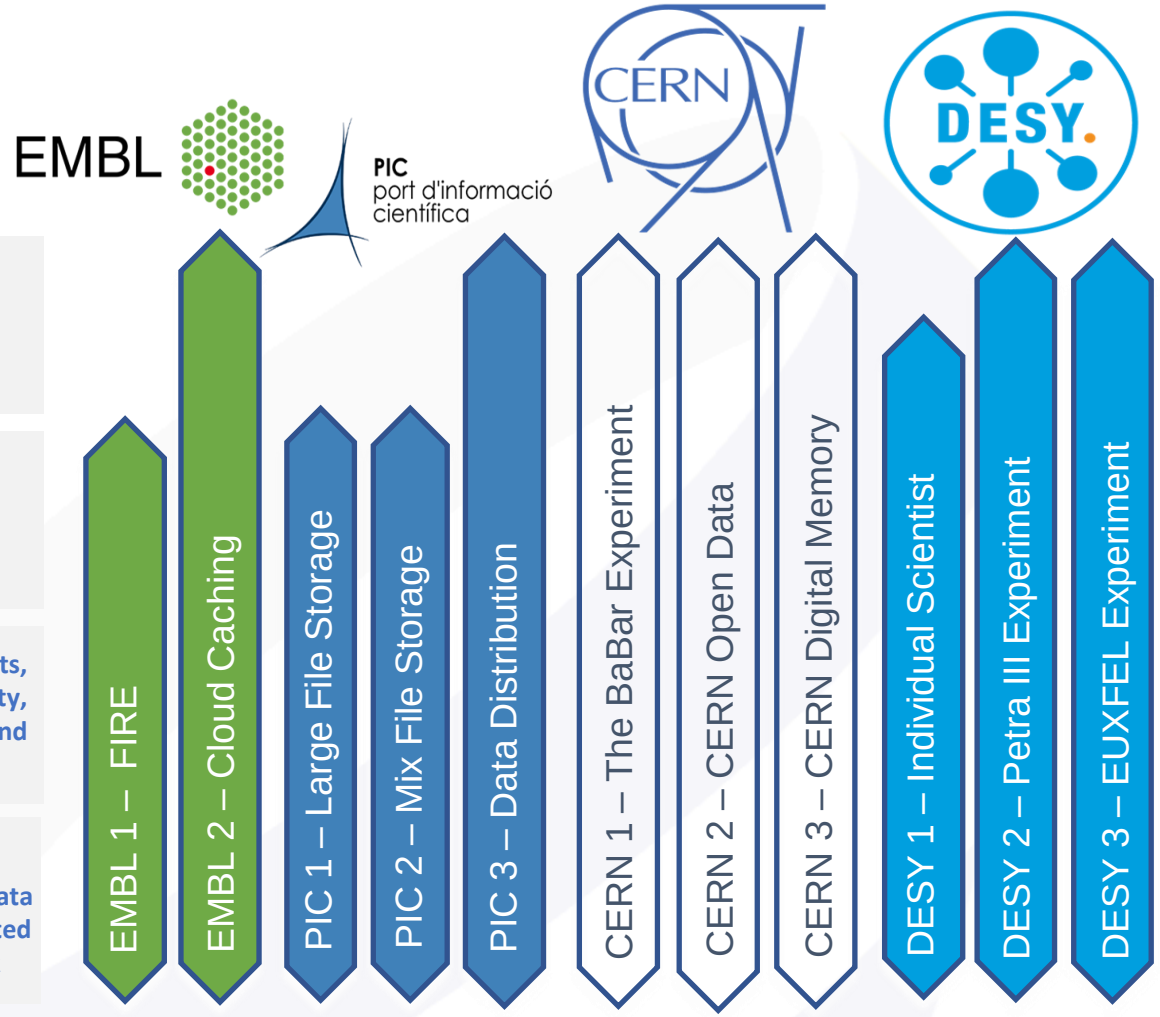
Demand Side Requirements

Layer 4 Advanced services
High level services: visual representation of data (domain specific), reproducibility of scientific analyses, etc.

Layer 3 Baseline user services
User services: search, discover, share, indexing, data removal, etc. Access under Federated IAM

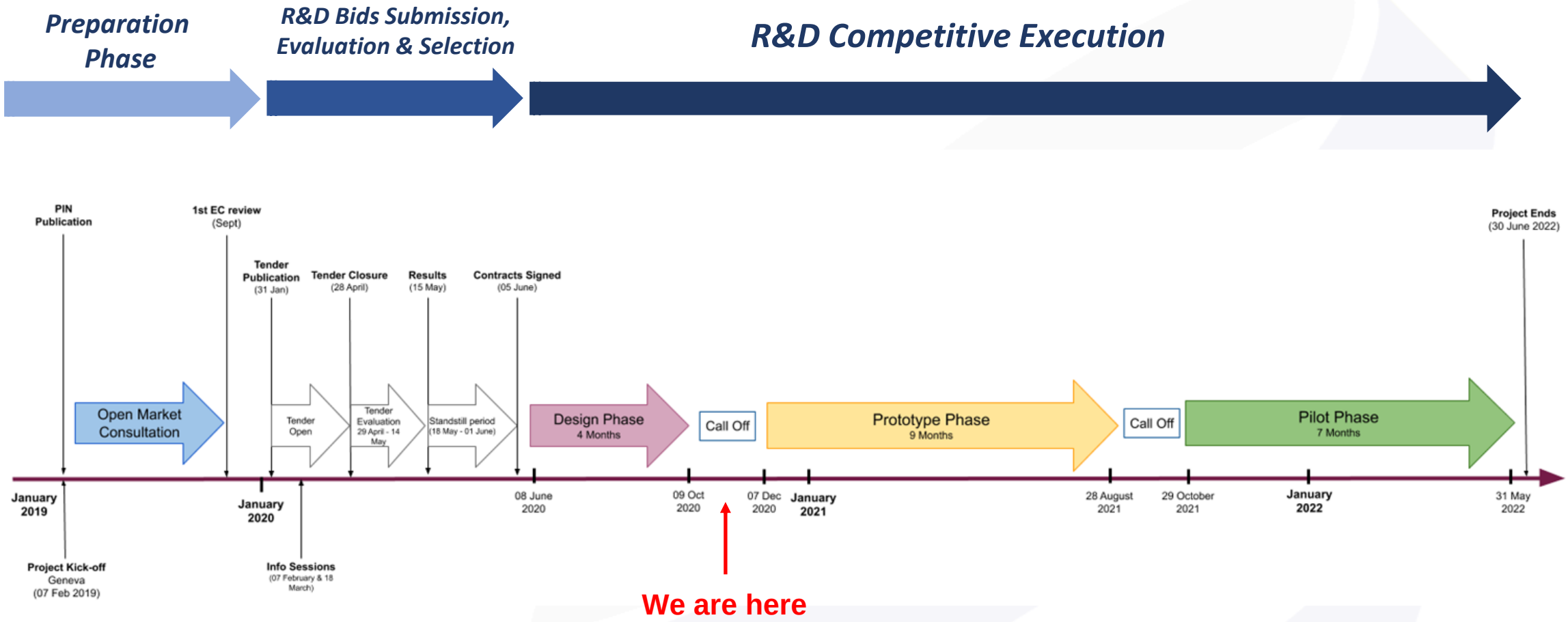
Layer 2 Preservation
OAIS conformant services: data readability formats, normalization, obsolesce monitoring, files fixity, authenticity checks, etc. ISO 14721/16363, 26324 and related standards

Layer 1 Storage/Basic Archiving/Secure backup
Data integrity/security; cloud/hybrid deployment
Data volume in the PB range; high, sustained ingest data rates. ISO certification: 27000, 27040, 19086 and related standards. Archives connected to the GEANT network

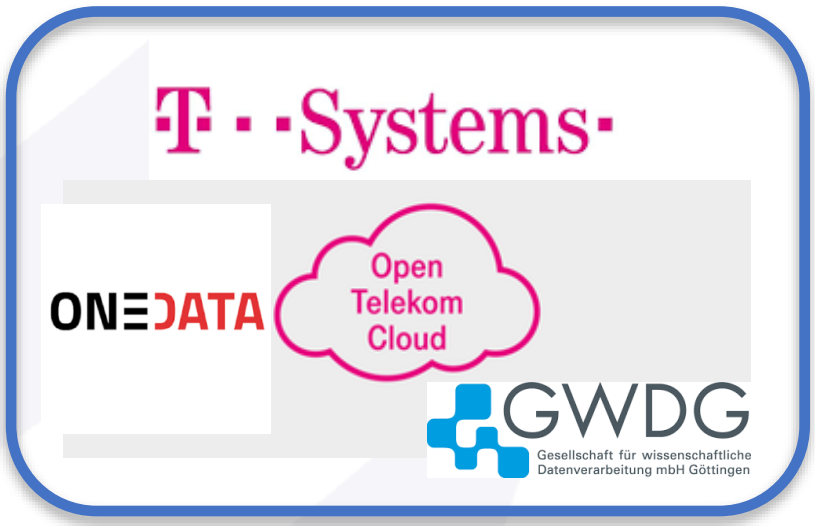


Scientific use cases deployments documented at: https://www.archiver-project.eu/deployment-scenarios

Project Timeline



Selected Consortia for the Design Phase



R&D proposal of ARKIVUM

- SaaS stack deployed on Google Cloud
- Scientific Data oriented
- OAIS Archive workflow
- FAIR Data support
- Open standards & specifications
- Trusted Digital Repository techniques
- Cost-effective archiving and preservation

https://www.archiver-project.eu/design-phase-award/consortium-1

R&D proposal of GMV

- Piql Connect and Piql UI
- Piql software on top of Archivematica
- Fed. Identity & Access Management
- AI functionalities
- Container-based workflows
- Operations over huge data scale

https://www.archiver-project.eu/design-phase-award/consortium-2

R&D proposal of LIBNOVA

- Established preservation platform
- Life-cycle OAIS compliant
- Scalability
- Containers
- Cost Efficiency
- Dynamic insights
- Budget Assistant

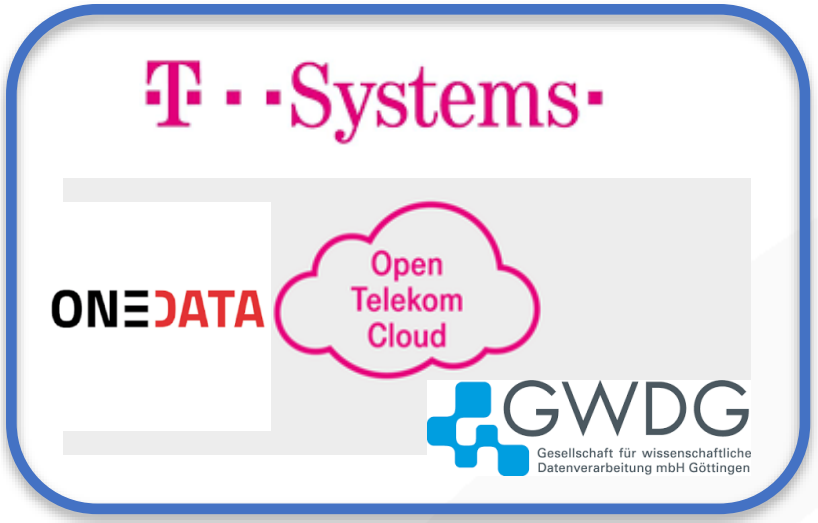
https://www.archiver-project.eu/design-phase-award/consortium-3

R&D proposal of RHEA

- Secure Service Portal
- Federated access service (SAML 2.0)
- Open Source applications
- Archivematica and AtoM tools
- Stewardship lifecycle
- Readiness XaaS services

https://www.archiver-project.eu/design-phase-award/consortium-4

R&D proposal of T-SYSTEMS



- **Petabyte-scale storage options, compliance with OAIS, PREMIS, METS and BagIT standards and new innovate functions for distributed data and workflow management, search and discovery, data representation and scientific analysis.**

https://www.archiver-project.eu/design-phase-award/consortium-5

Design Phase conclusions

- The objectives of the design phase were successfully met.
- Most of the deliverables provided by the companies were of sufficient quality.
- Good indication that the main R&D purpose and the use cases requirements were globally understood.
- CERN, EMBL-EBI, DESY and PIC allocated significant effort assessing and testing the demo platforms showcasing current capabilities and state-of-the-art.
- Feedback was systematic. Companies congratulated the project team for the excellent interaction, generating good progress when compared to other project formats, including project dissemination actions.

The next steps of ARCHIVER project

- **Selection process for the Prototype Phase**
- **Kick Off event - 9th December 2020 (TBC)**

https://www.archiver-project.eu/
https://twitter.com/ArchiverProject
https://www.linkedin.com/company/archiver-project/
https://www.youtube.com/channel/UCCBIyLpUt-hWmQatqdlhIzw