



ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

ARCHIVER

Archiving and Preservation for Research Environments

RDA Virtual Plenary event

João Fernandes (CERN)
Project Leader

19th of April 2021



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 cloud services 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: 42 Months

Coordinator: CERN (Lead Procurer)



European Commission



Consortium

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



EMBL-EBI



PIC
port d'informació
científica

Buyers



Consortium

Experts

addestino
innovation delivered.








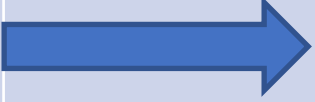
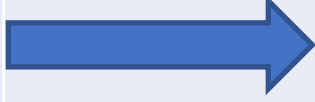
The “Buyers Group”: Public organisations committing funds to contribute to a joint-R&D-procurement, research data use cases and R&D testing effort

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

ARCHIVER & ESFRI science cluster projects

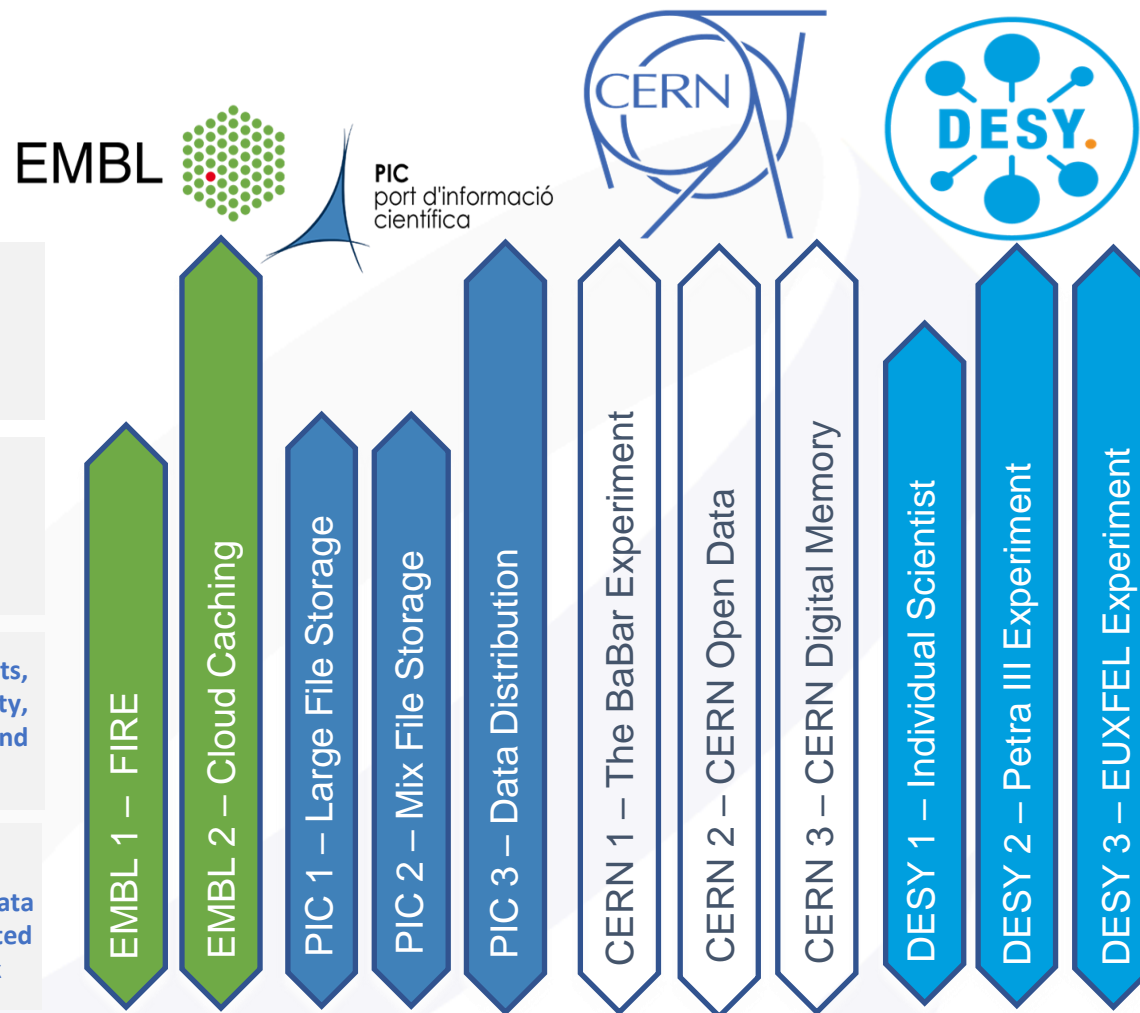
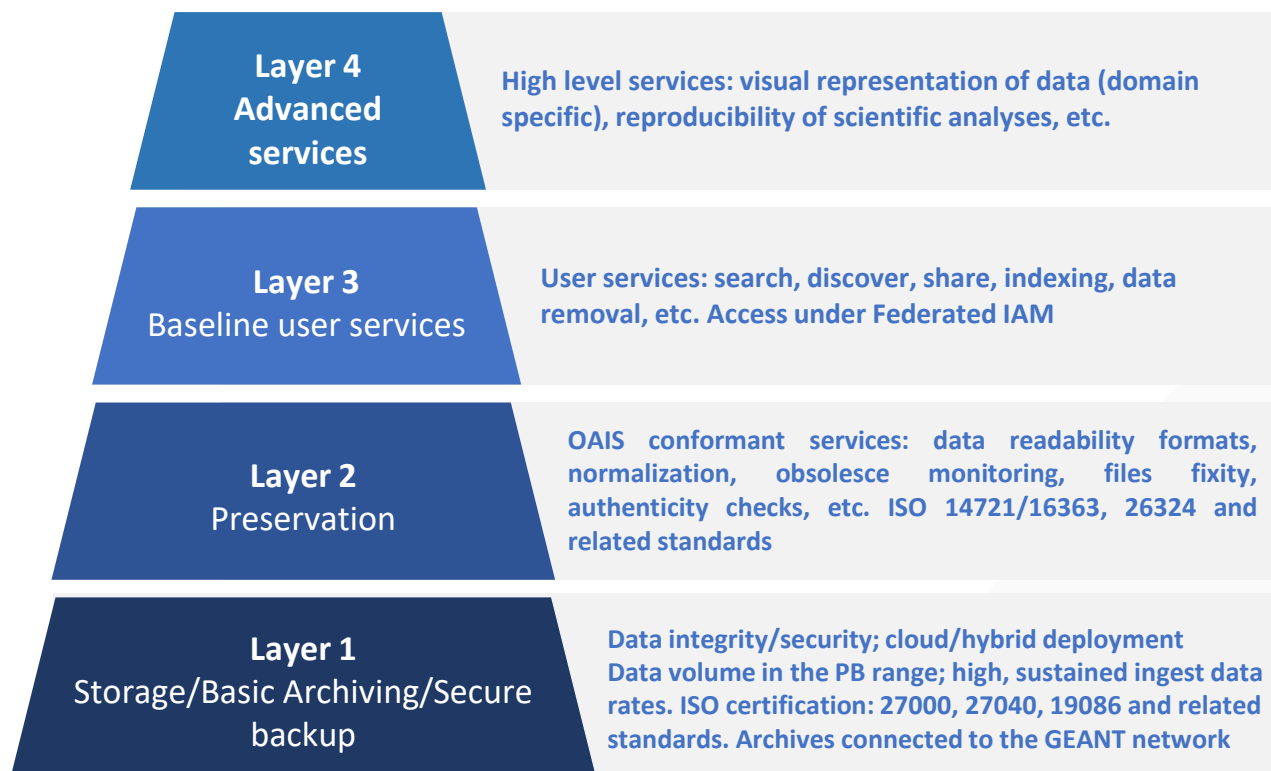


Progress Beyond the state of the art

<i>Current Scientific Data Repositories</i>		
<i>Growing data volumes</i>		<i>PB scale demonstration of scientific data repositories</i>
<i>Basic bit preservation capabilities</i>		<i>European SaaS providers in digital preservation</i>
<i>Most of research data not published</i>		<i>Best practices: FAIR, TRUST, DPC RAM</i>
<i>Technology lock-ins concerns (tape), Business Continuity plans needed (COVID-19)</i>		<i>Promote FOSS, open standards & demonstrate exit strategies</i>
<i>Fragmentation across scientific disciplines & countries</i>		<i>Pan-European: resulting services to be available in the EOSC portfolio</i>
<i>Cost underestimation at the planning phase</i>		<i>Cost model adapted to public research</i>

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

Demand Side Requirements



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

Early Adopters <https://archiver-project.eu/early-adopters-programme>

• Participants:

- Public sector organisations; Can join at each transition. Next one: Pilot Phase;

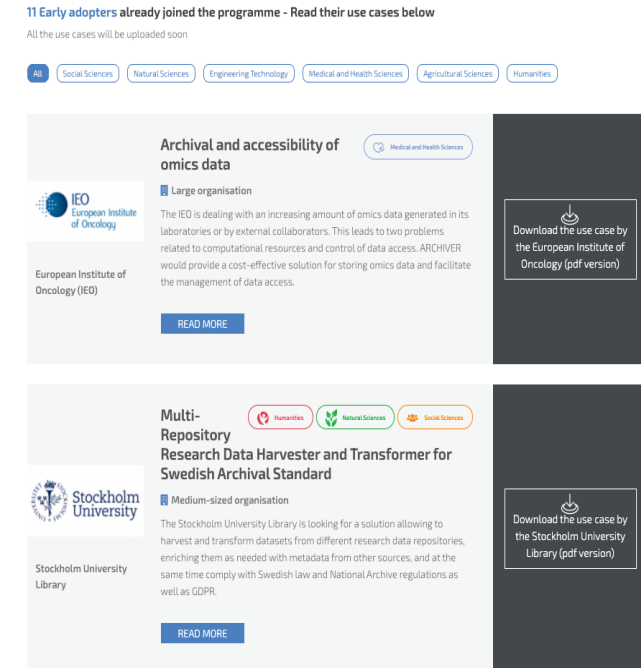
• Key advantages

- 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
- Have the option to purchase pilot-scale services by the end of the project

• Confirmed 12 organisations

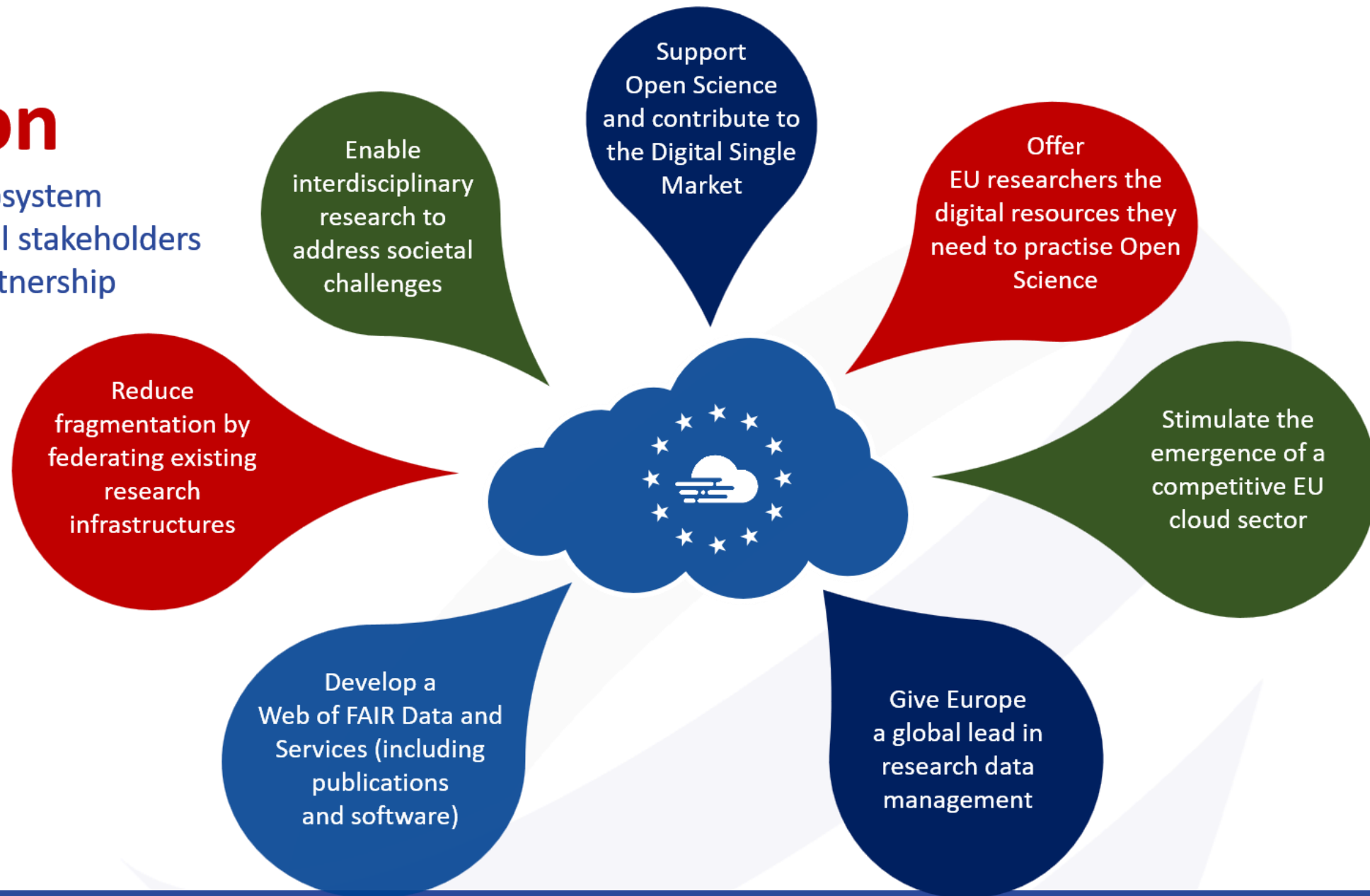


Friedrich Miescher Institute
for Biomedical Research

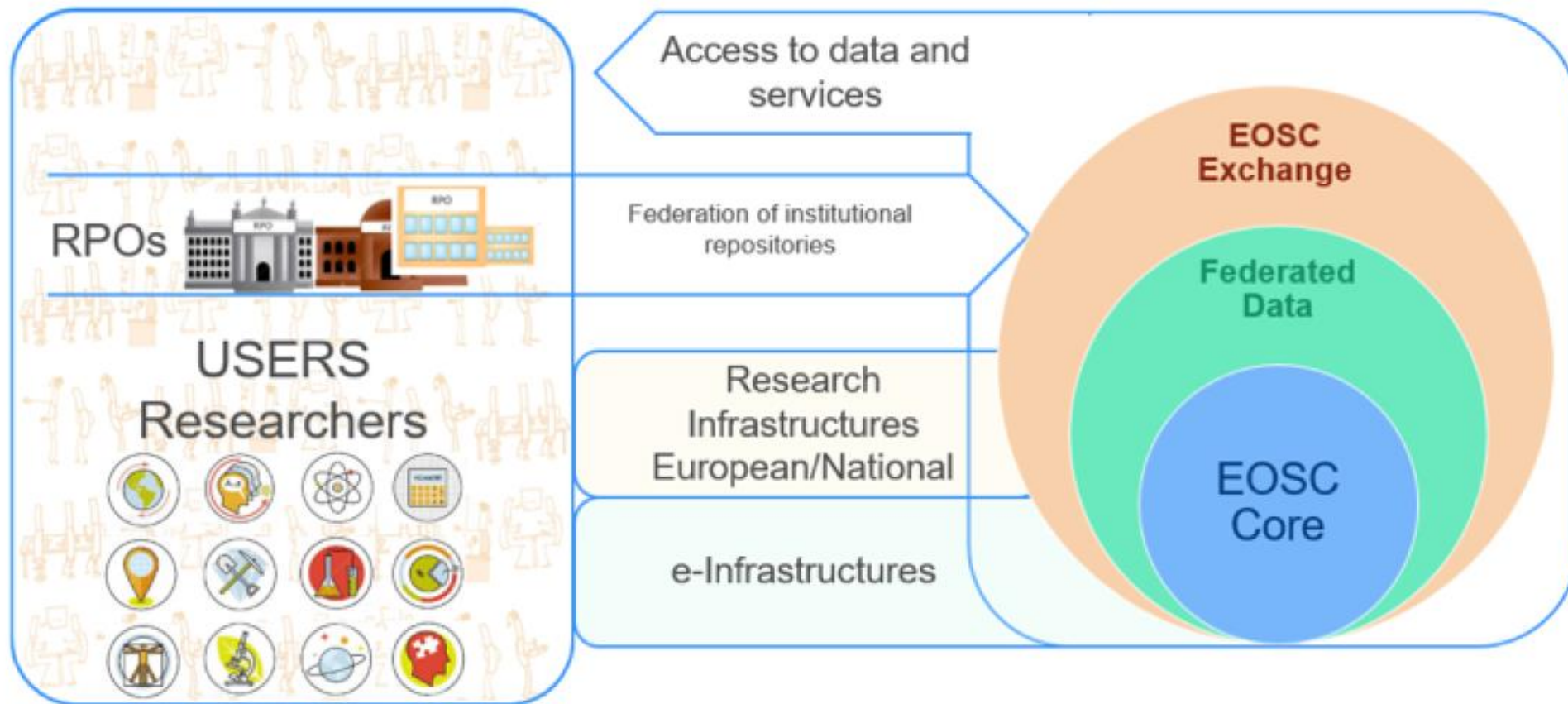


The Vision

Building the EOSC ecosystem collaboratively with all stakeholders through the EOSC Partnership

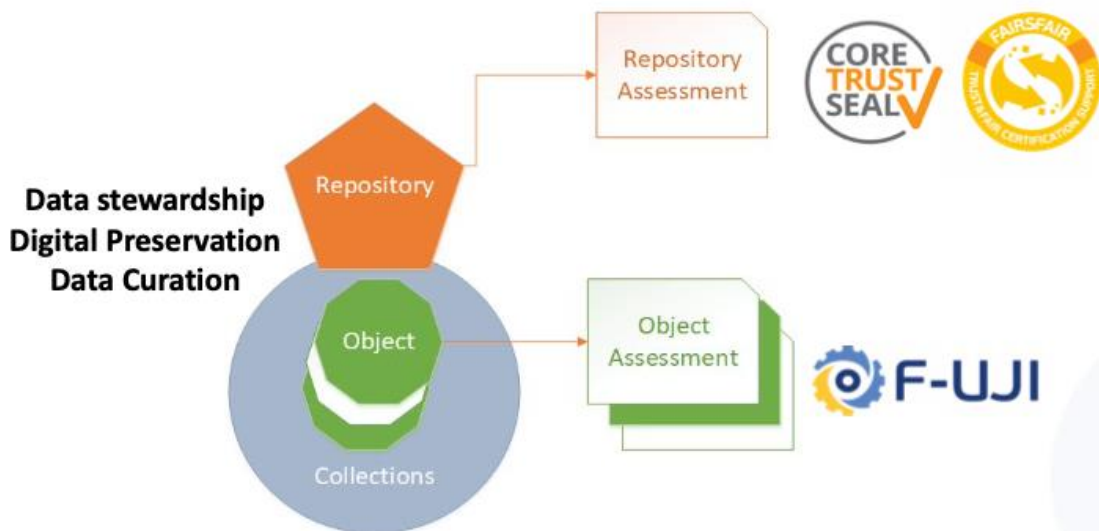


EOSC ecosystem



The challenge is not limited to linking datasets, federating infrastructures or aligning policies. It starts by linking people and organisations across the EOSC ecosystem.

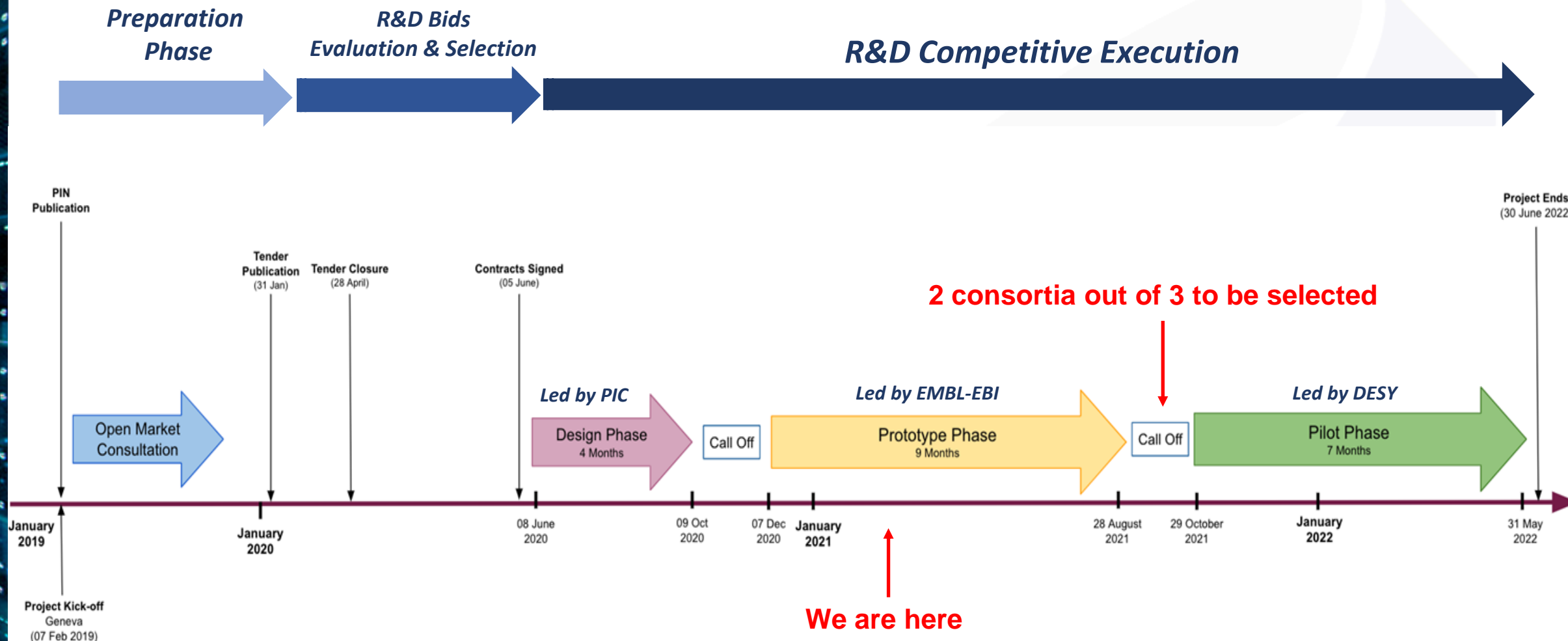
Synergies with EOSC projects



Proposed Collaboration with EOSC-FUTURE:

- Allocation of procurement funds for the deployment of the additional data preservation use cases
- Implement a procurement model for a broader set of data preservation use cases in the EOSC, including TB volume

Project Timeline



Selected Consortia for Prototypes

T-Systems



ONE DATA



libnova



UNIVERSITAT DE
BARCELONA



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



**Giarretta
Associates**



voxility



arkivum

Bringing archived data to life



Google Cloud

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

<https://twitter.com/ArchiverProject>

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

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

- Archiving much more than storing: keeps intellectual control of data and associated products, renders research outputs reusable.
- Developing innovative services addressing long-term preservation reproducibility of research results of PB data, with demonstrators of exit strategies from cloud /on-premise.
- R&D challenge co-addressed between public research performing organisations and European specialist innovative SaaS companies, with yearly cost estimations at the end of Prototypes (June 2021).
- CERN, EMBL-EBI, DESY & PIC allocating significant effort assessing and testing platforms, ingesting data, showcasing current capabilities and pushing the state-of-the-art.
- ARCHIVER acts as a template to commoditise archiving and preservation in the research domains.
- Sustainable model aligned with the EOSC Rules of Participation.
- Series of Webinars to be organised from Q2 2021 (EOSC clusters, FAIRsFAIR)