ARCHIVER Cloud Benchmarking Test Suite

Jakub Urban (CERN) - Ignacio Peluaga (CERN / Universidad de Sevilla) - João Fernandes (CERN)

Motivation

ARCHIVER combines multiple ICT technologies, including extreme data-scaling, network connectivity, service interoperability and business models, in a hybrid cloud environment to deliver end-to-end archival and preservation services that cover the full research lifecycle. By acting as a collective of procurers, ARCHIVER creates an eco-system for specialist ICT companies active in archiving, who would like to introduce new services capable of supporting the expanding needs of research communities. This activity is performed in the context of the EOSC (European Open Science Cloud) catalogue integration in mind.

The project requires automated and transparent process of quality assurance, short feedback cycle to evaluate the technical readiness level of the ICT companies participating in the R&D competition. The IT department at CERN has been putting effort for the last years on designing and developing a testing framework in order to simplify and automate as much as possible the testing tasks on projects like Helix Nebula Science Cloud or OCRE. The lessons learned during that time are now applied to the ARCHIVER Cloud Benchmarking Test Suite. The concept is expected to be expanded and considered a best practice for the on-boarding of commercial services for research environments in the context of the European Open Science Cloud.

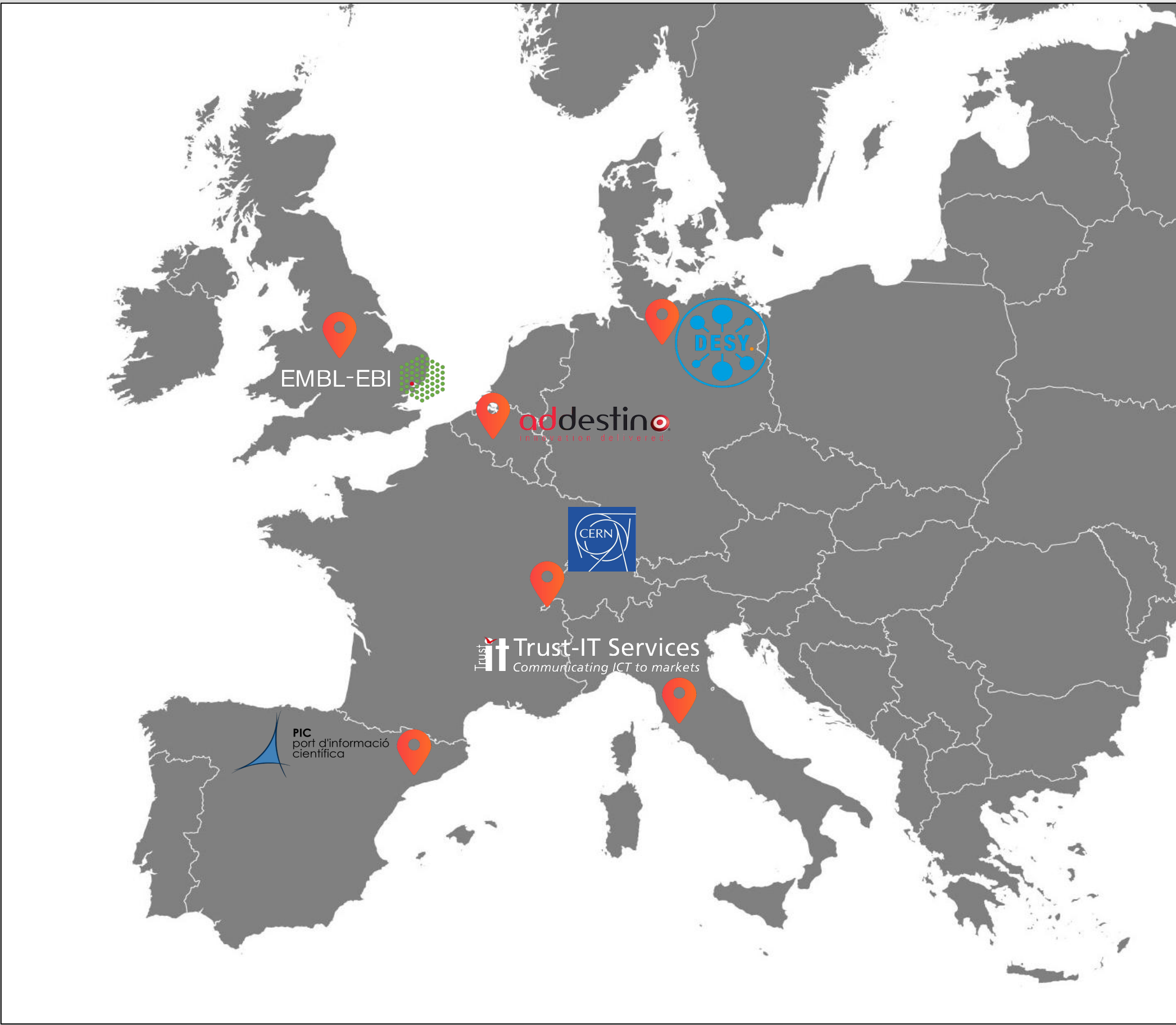## Organizations involved

Buyers Group

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

- ○ CERN (leader)
- ○ DESY
- ○ PIC
- ○ EMBL-EBI

Expert Partners

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

- ○ Addestino
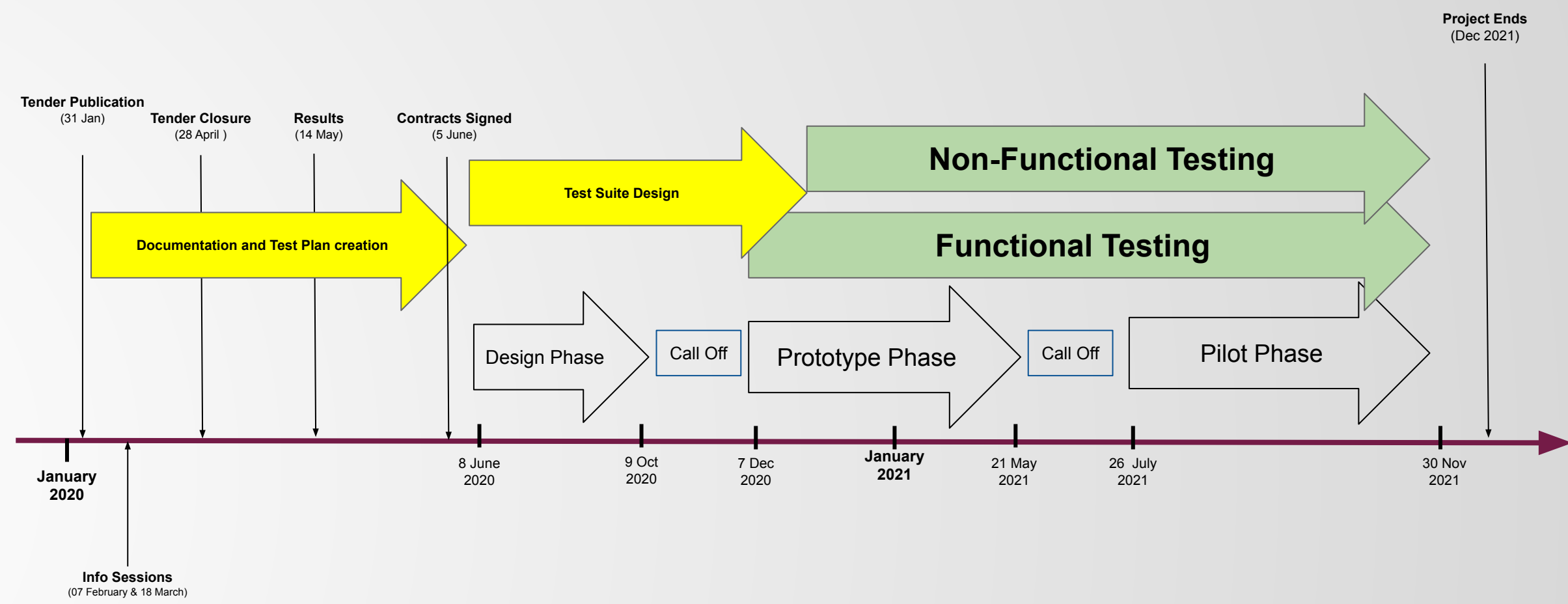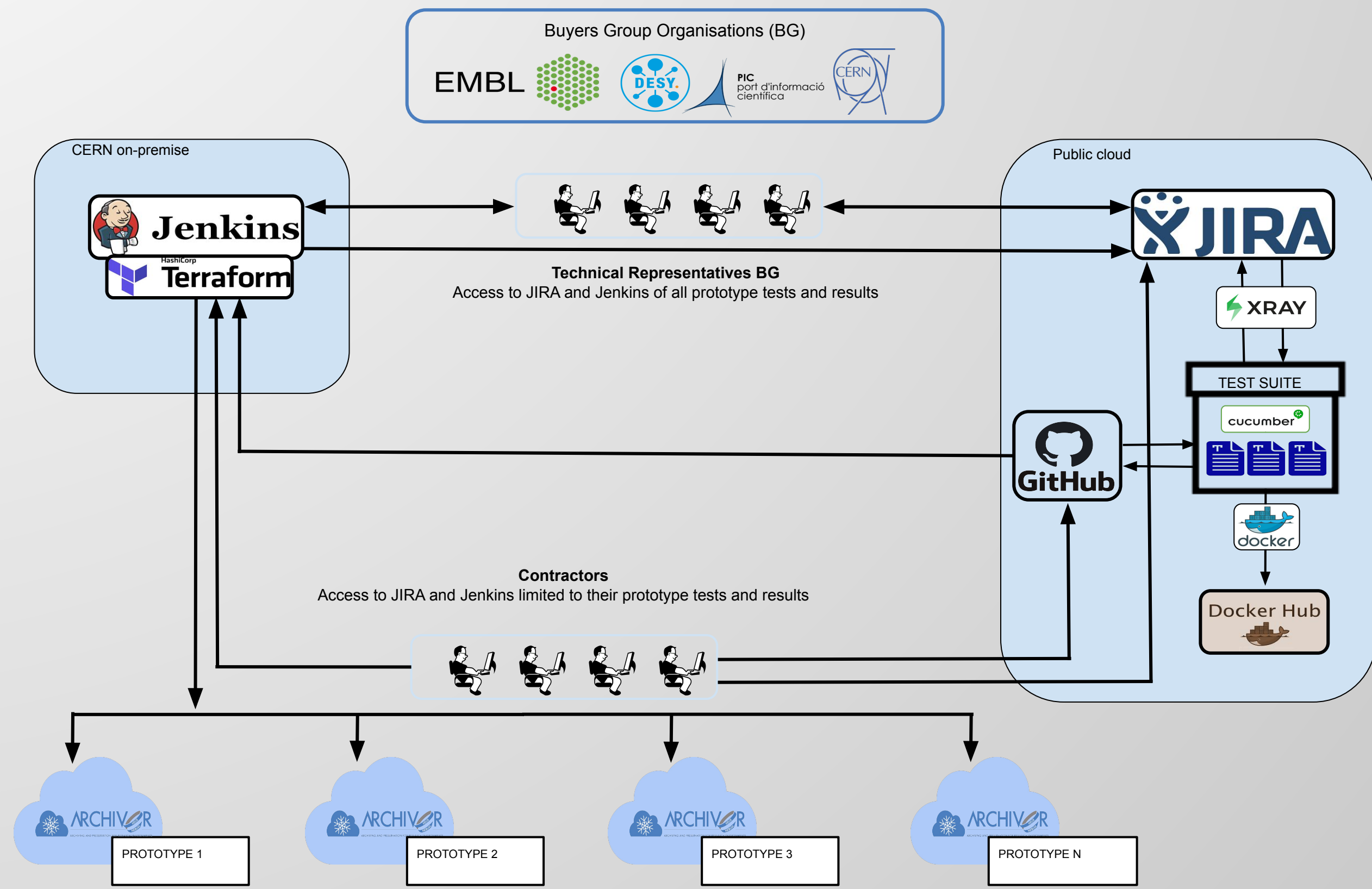- ○ Trust-IT

Project plan

- ● **May 2020**
  - ○ Hard Fork from the OCRE Test Suite: filter the out of the scope tests
  - ○ Design and implement ARCHIVER test suite specific tests, Provision CI/CD stack: ready to be used against ARCHIVER staging environment
- ● **November 2020**
  - ○ The modules scope is established for the first level validation
- ● **January 2021**
  - ○ Addition of second level modular tests and inclusion of the suite on the EOSC portal to be used for future projects.

Current status

- ➢ Code base in **Python** and **Shell** operating pipeline defined in **Jenkinsfile**
- ➢ Usage of simple **YAML** files for configuration. Raws results consumed in **JSON** to be processed for humans.
- ➢ Provisionning of raw infrastructure with **Terraform**
- ➢ Abstraction layer provided by **Kubernetes** and **Docker**
- ➢ Cluster bootstrapping by **Ansible**'s ansible-playbooks
- ➢ Custom **Jenkins** Docker image deployed on **CERN Openshift**: on builds, clones the ARCHIVER test-suite repo
- ➢ **Jira Cloud** integration for a broad management in collaboration with the buyers group and external staff
- ➢ **Cucumber/Xray** for Jira-Jenkins integration

Area	Test Description	Status
Object Storage	S3 protocol endpoint and CRUD methods: Testing of different Simple Storage Service functionalities. This test first tries to create an S3 bucket on the provider side. On success, it tries the following methods: PUT, simply GET, GET with prefix matching, GET chunk and GET chunks.	✔
Networking	Networking tests using the open source perfSONAR toolkit: includes traceroute, tracepath, iperf, nuttcp, ping and owamp. A perfSONAR endpoint is created on the provider side and from there, the networking test are launched against another perfSONAR endpoint the user choses. perfSONAR is a mature project developed by the University of Michigan among other institutions and offers endpoints deployed all around the globe.	✔
Data Repatriation	Data exportation: take data out of the cloud onto Zendo. In the chase of making the scientific community embrace a vouchers usage of cloud computing, there is always a drawback: once my cloud credits are over, what can I do with the data I have on the cloud? Zenodo is a tool developed by CERN useful in this case as it keeps data for the long term and it is free. What this test does is to verify that data can be taken from the private cloud being tested onto Zenodo.	✔
FAIR Evaluator	Verify the FAIR principle of the solution looking to the ingested dataset and their quality. The curation process and manual verification is crucial here same as the EOSC guidance. Automated testing as defined at https://github.com/FAIRMetrics/Metrics or any adapted one.	⚠
Data Ingestion	Test of the ingest process, ingests with incremental changes for high volumes of data, lifecycle from archive packages data creation, combination and/or aggregation before final archival for long-term preservation.	⚠
Open APIs	It is crucial to require open standards and well defined application interfaces for the exit strategy and preventing vendor lock-ins. We need to verify the created innovative workflow openness and reliability in accordance to the specification and integration possibilities of the hybrid premise.	⚠
Federated IAM	The federated Authentication and Authorization Infrastructure is a requirement to provide a single point access to the researchers and professionals with the divergent levels of privileges. The purpose is to check if the AAI meets all the mandatory requirements of OAuth/OpenID mechanism and to integrate within the EOSC-Hub as well as to be incorporated as a SSO feature.	⚠

For each test deployed, a Docker image has been created containing the respective software: normally python or shell scripts or third-party tools. These images are used on Kubernetes resource definition YAML files to deploy pods on the cluster. Once the test inside the container running on the Kubernetes cluster is completed, results are harvested, resources released and the next test is deployed.

Continuous Testing Environment

Get involved

The project developers seek for contributions and feedback from everyone. The code and all information concerning the project have been made available to public. Please refer to the following links to obtain useful information. Feel free to give it a try, provide your feedback or let us know if you want to integrate a test.

- ❖ Archiver Test-Suite Repository: https://github.com/cern-it-efp/archiver-testsuite
- ❖ Documentation: https://archiver-testsuite.readthedocs.io/en/latest/
- ❖ OCRE Test-Suite Repository: https://github.com/cern-it-efp/eosc-testsuite
- ❖ ARCHIVER website: https://archiver-project.eu
- ❖ OCRE website: https://ocre-project.eu
- ❖ HNSciCloud website: https://hnscicloud.eu

For more information please contact *jakub.urban@cern.ch*

CERN - 2020