Extending the ESCAPE Datalake prototype to test RIs data accessibility from commercial cloud services

Bob Jones, Simone Campana, Xavier Espinal, Joao Fernandes, CERN  
29 April 2020

# Introduction

This document summarizes the possible synergies between the ESCAPE project and MECHANICS and proposes a collaboration to extend the ESCAPE datalake prototype in order to test RIs data accessibility from commercial cloud services.

# Relevance to the ESCAPE programme of work

As part of the ESCAPE project’s programme of work, WP1: MIND (Management, Innovation, Networking and Dissemination) states anetwork of industrial stakeholders already established by the different ESCAPE partners will be consolidated and considered for potential further cooperation on issues such as services and software developments, **RIs data accessibility from commercial cloud services**.

To achieve this, the objective is to propose and plan potential innovation actions with industries, that can include co-development and procurement of some limited commercial cloud service capacity for test purpose, other validation tests, etc. The budget includes 70,000€ to procure a limited commercial cloud service capacity.

ESCAPE WP2 DIOS (Data Infrastructure for Open Science) aims to create a cloud of data services, often referred to as a “Data Lake” by building on and integrating existing work from a variety of areas. Compute services would be provided either in the data centres that form the data lake, or by well-connected compute resources (e.g. existing research data centres, HPC resources, **commercial cloud resources** – for example building on the results of HNSciCloud, and volunteer computing).

Such data lake structure allows the underlying storage service to scale-out the capacity by adding additional large storage providers (data centres). This model also introduces directly the possibility to include **commercial service providers** as physical storage sites.

The data-lake model is constructed to allow data centres to join/leave the infrastructure, whilst assuring the overall security and availability of the data, not relying on any single data centre. This is an ideal mechanism by which public **commercial clouds** can participate as data storage sites, permitting a tendering process to select different providers over time.

The model is also designed to be flexible and highly adaptable to how compute resources can be made available. It is assumed that some of those resources will be cloud centres, and WP2 will validate the use of **commercial cloud resources** in this model by verifying that RIs data is accessible from commercial cloud resources.

A policy mechanism will be used to define which data sets can be stored in which data centres (including the possibility to use **public cloud data centres**), and to define replication and access policies. The work of WP2 will validate that RI data is accessible from a variety of platforms including **commercial**

**cloud resources** (Milestone 2.4 at month 24).

# MECHANICS: **M**aking **E**uropean **C**loud and **H**PC resources **A**vailable – a **N**etwork, **I**nfrastructure and **C**loud **S**ecurity project

MECHANICS is a proposal to be submitted to the ICT-40-2020 funding call with a deadline of 17 June 2020[[1]](#footnote-1).

High Performance Computing has become the underlying infrastructure of today’s digital society with its ability to manage abundant data and complexity. HPC enables personalized medicine, autonomous systems, and climate simulations, among other applications. Likewise, cloud infrastructures are now the vehicle of choice for delivering IT resources. Furthermore, energy efficiency has gained mayor importance. Integrating Europe’s forthcoming EuroHPC infrastructure into public cloud environments would make innovative, rare, resources available to large enterprises, public research organisations and SMEs. As Europe lacks hyperscalers, a federated approach providing efficient access and orchestration for end users can provide the required capacity and functionality.

The MECHANICS proposal aims to prove the feasibility of a public cloud-based approach, covering the planned European Cloud Federation (ECF) and industrial partner access to EuroHPC computing resources.

The following proposed measures lay the groundwork for providing HPC, QC and IoT services securely within European dataspaces within the European legal framework in a highly energy efficient manner. It aims to provide:

* A security model for the ECF including Secure Terabit Connectivity both for the ECF and EuroHPC will identify security requirements of the cloud-federation and EuroHPC (including Quantum Computing (QC)) to enable industrial stakeholder access.
* A Proof of Concept (PoC) prototyping aspects of the security concept will demonstrate energy efficient, high bandwidth, high QoS, secure distribution of computing resources:
  + **High Throughput Computing (HTC)** resources, physically located either on the public cloud or at CERN’s computing center, will be integrated into the cloud service offering portfolio. This will reduce the need to transport LHC collision data to each relevant institute. **(Data Lake Model)**
  + **Quantum Computing (QC)** resources will be integrated into the cloud service offering a portfolio to optimize and accelerate processes for complex industrial systems, exploring models that will make them ubiquitous and generally available for all .

The MECHANICS consortium is led by T-Systems[[2]](#footnote-2) and includes CERN & OROBIX who are members of the ESCAPE consortium.

CERN has a long-standing relationship with T-Systems who have been an active member of the Helix Nebula initiative[[3]](#footnote-3) since 2011 and provides one of the network links[[4]](#footnote-4) between the Wigner data centre in Hungary and CERN’s data centre in Geneva.

T-Systems operates the OpenStack based Open Telekom Cloud (OTC[[5]](#footnote-5)) from its 40,000m2 data centre in Germany[[6]](#footnote-6). OTC provided the basis for the T-Systems hybrid cloud infrastructure that was selected by a group of 10 research organisations (CERN, CNRS, DESY, EMBL, ESRF, IFAE, INFN, KIT, STFC, SurfSARA) as part of the HNSciCloud pre-commercial procurement[[7]](#footnote-7). Several applications from a range of RIs (including CTA[[8]](#footnote-8), LHC[[9]](#footnote-9) and LOFAR[[10]](#footnote-10)) and disciplines were successfully deployed on OTC supported by these research organisations in 2018[[11]](#footnote-11).

T-Systems also operates one of the DIAS (Data and Information Access Services) platforms providing access to Copernicus Sentinel data[[12]](#footnote-12).

# Use Case Proposal - Collaboration between ESCAPE and MECHANICS

The objective of the proposed collaboration between ESCAPE and MECHANICS is to extend the ESCAPE Data Lake prototype in order to test RIs data accessibility from commercial cloud services as part of the the MECHANICS HTC PoC.

The intention is to include T-Systems data centre in Germany as a site in the ESCAPE data lake during 2021 in order to demonstrate RI data is accessible from commercial cloud resources and thereby successfully achieve Milestone 2.4 by the end of 2021.

A range of research applications will be deployed on T-Systems OTC cloud services and access datasets from the data lake via a high-speed network link deployed by T-Systems to CERN. The network connection and application deployment will be tested by CERN using LHC-one and LHC experiment workloads then made available to other ESCAPE partners.

In this context, a caching layer deployed at CERN based on XCache[[13]](#footnote-13) technology will be the interface to the ESCAPE Data Lake scientific data. Such caching and content delivery service will be in charge of data caching, latency hiding and file streaming to the compute resources located at OTC provided by T-Systems.

Access to the OTC cloud services will be made available to ESCAPE partner end-users via a voucher-based model supported by T-Systems and successfully tested during the HNSciCloud project[[14]](#footnote-14).

The provisioning of the network connection and use of the OTC cloud services will be funded by the MECHANICS project.

The ESCAPE procurement budget will not be necessary for this activity.

1. Cloud Computing: towards a smart cloud computing continuum, <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ict-40-2020> [↑](#footnote-ref-1)
2. <https://www.t-systems.com/de/en/> [↑](#footnote-ref-2)
3. <https://www.helix-nebula.eu/> [↑](#footnote-ref-3)
4. <https://www.datacenterdynamics.com/en/news/deutsche-telekom-gives-cern-a-100gbe-link-to-budapest-data-center/> [↑](#footnote-ref-4)
5. <https://www.t-systems.com/de/en/cloud-and-infrastructure/manage-it-efficiently/open-telekom-cloud> [↑](#footnote-ref-5)
6. <https://www.t-systems.com/de/en/about-t-systems/company/innovation-management/data-center-biere> [↑](#footnote-ref-6)
7. <https://www.hnscicloud.eu/> [↑](#footnote-ref-7)
8. <https://www.cta-observatory.org/> [↑](#footnote-ref-8)
9. <https://home.cern/science/accelerators/large-hadron-collider> [↑](#footnote-ref-9)
10. <http://www.lofar.org/> [↑](#footnote-ref-10)
11. <https://www.hnscicloud.eu/sites/default/files/files/HNSC_BookletA5_November2018_21081123_web.pdf> [↑](#footnote-ref-11)
12. <https://mundiwebservices.com/> [↑](#footnote-ref-12)
13. <https://www.epj-conferences.org/articles/epjconf/pdf/2019/19/epjconf_chep2018_04008.pdf> [↑](#footnote-ref-13)
14. Voucher Schemes for Accessing Commercial Cloud Services in the Research Environment, <https://doi.org/10.5281/zenodo.2615456> [↑](#footnote-ref-14)