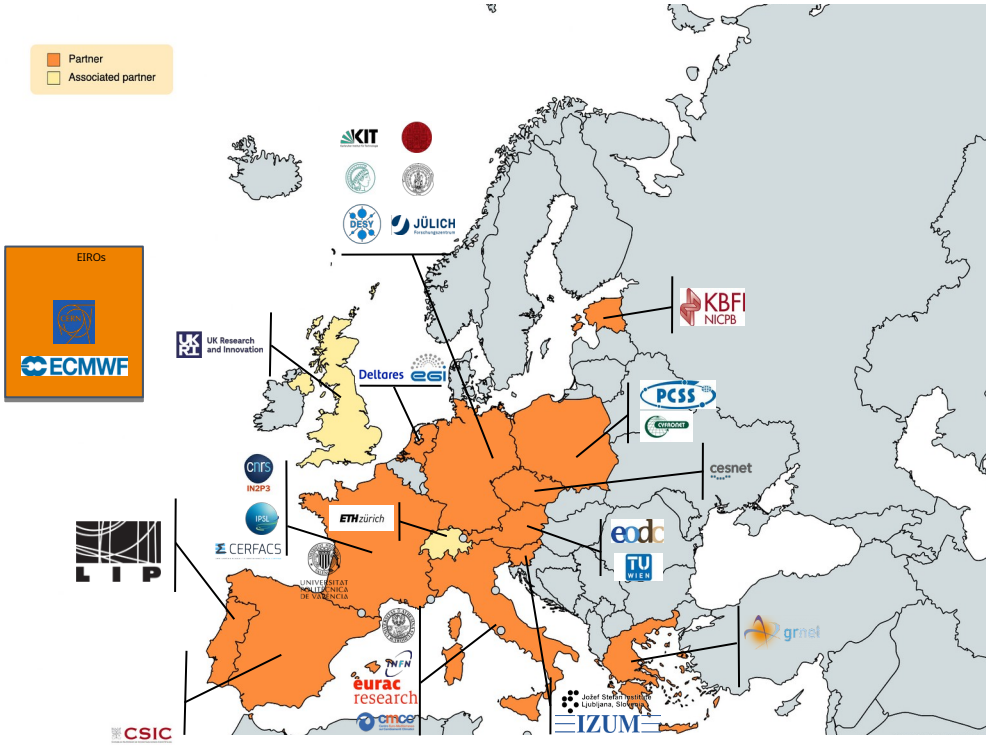
interTwin

interTwin Digital Twin Engine: Enabling Federated Scientific Workflows

Dijana Vrbanec, Paul Millar

interTwin is funded by Horizon Europe under grant agreement n° 101058386

Overview of the interTwin project



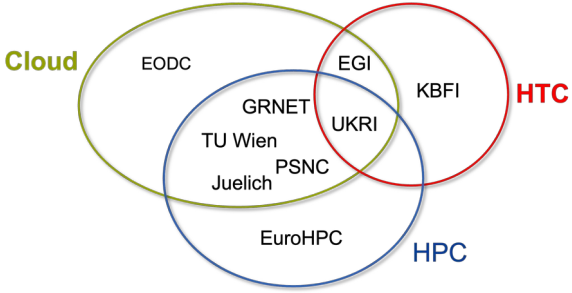
1.09.22 - 31.08.25

Budget 11,7 M €

Coordinator: eGI

30

Participants,
1 affiliated entity
2 associated partners



8
Providers

Cloud, HTC and HPC resources

11
Technology providers

delivering the infrastructure and horizontal capabilities

14 Community representatives

from 5 domains, developing DT applications and thematic modules

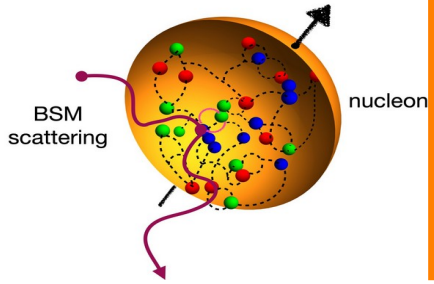
Project objective

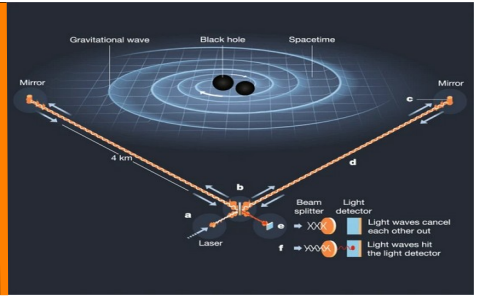
Co-design and implement the prototype of an interdisciplinary Digital Twin Engine

Digital Twin Engine

- Is an **open-source platform** based on open standards
- It offers the capability to develop **application-specific Digital Twins**
- Piloted by a large spectrum of **diverse use cases** from physics and environmental sciences
- Its functional specifications and implementation are based on:
  - a **co-designed interoperability framework**
  - conceptual model of a DT for research - **the DTE blueprint architecture**

Scientific use cases: High-energy & Astrophysics





DT of LHC detector components

- Faster and deeper cycles of optimisation of experiment parameters

DT of the Standard Model in Lattice QCD

- Systematize the inclusion of ML for large scale parallel simulations

DT for noise simulation of next-generation radio telescopes

- Contributing to realization of “dynamic filtering”

DT of the Virgo Interferometer

- Detecting “glitches” in quasi-real time sending out more reliable triggers

Scientific use cases: Climate change & Extreme Weather Events

DTs of the Earth, addressing complementary topics such as:

- Climate change, **long-term predictions** of extreme natural events (storms & fires)
- **Early warning** for extreme events (floods & droughts)
- Climate change **impacts** on extreme events (storms, fires, floods & droughts)

interTwin DTE Architecture

DT Applications | DT Applications | DT Applications | DT Applications

DTE Thematic

Thematic Modules
Environmental Sciences

Thematic Modules
HEP / GW Astro / Radio Astro

Thematic Modules
Other Sciences

DTE Core

Quality Verification

Workflow Composition

Big Data Analytics

AI / ML

Data fusion

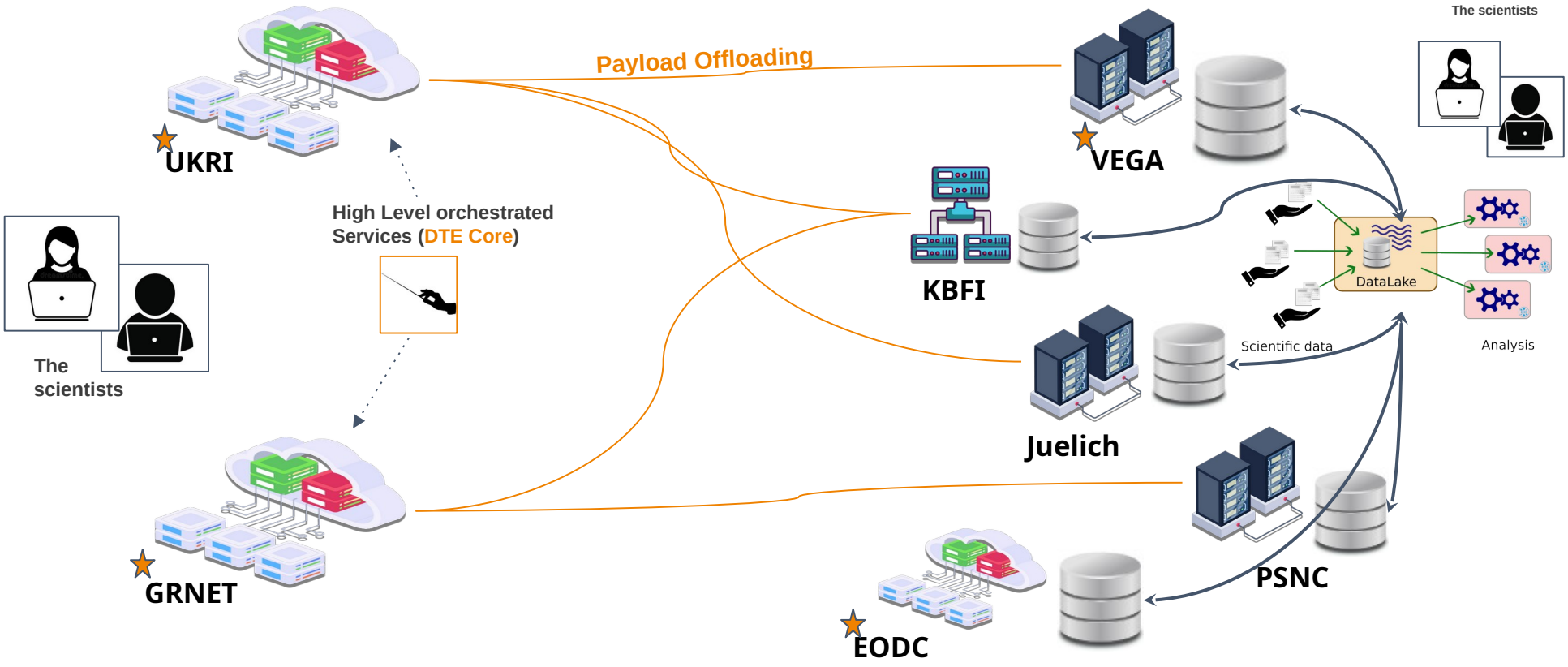
Real-time data acquisition and processing

DTE Infra

AAI | Orchestrator | Federated compute | Federated data management

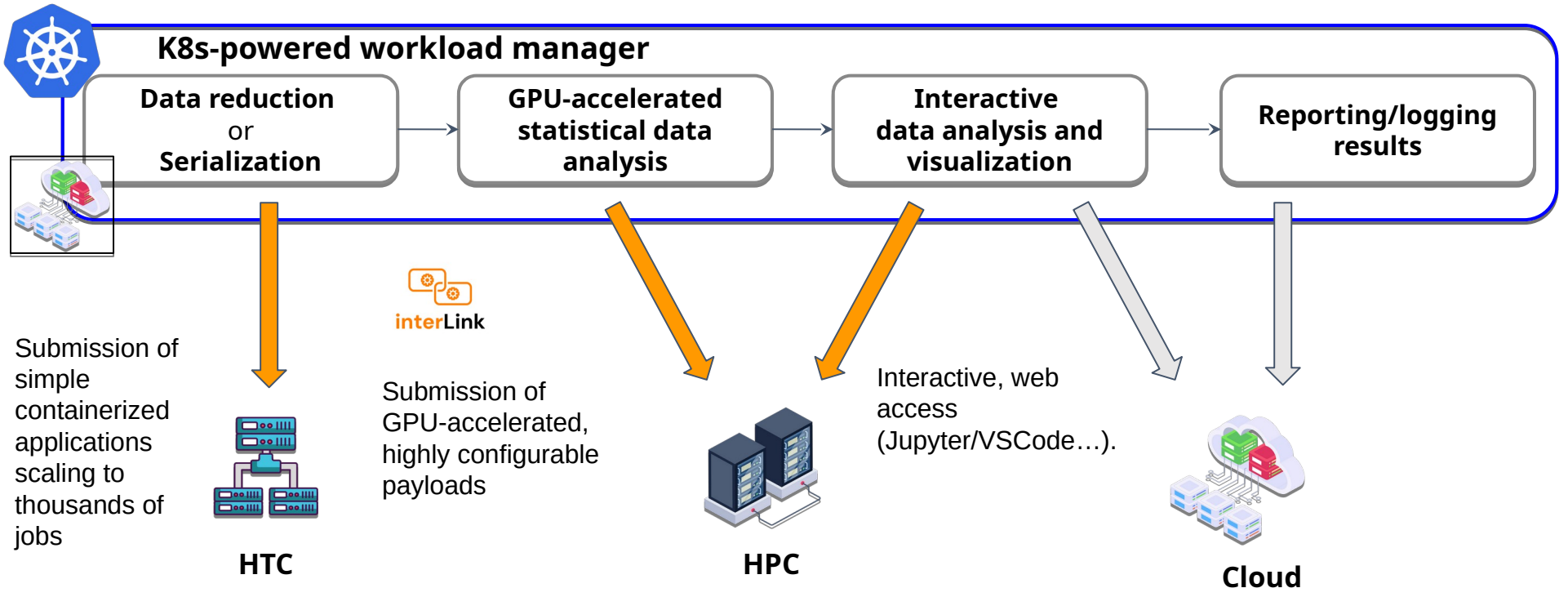
HTC | HPC | Cloud | Data Repositories

DTE Infrastructure integrated topology

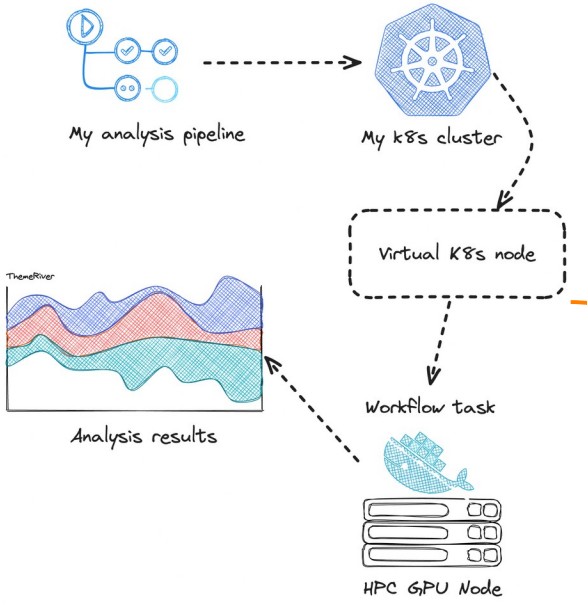


Federated compute infrastructure

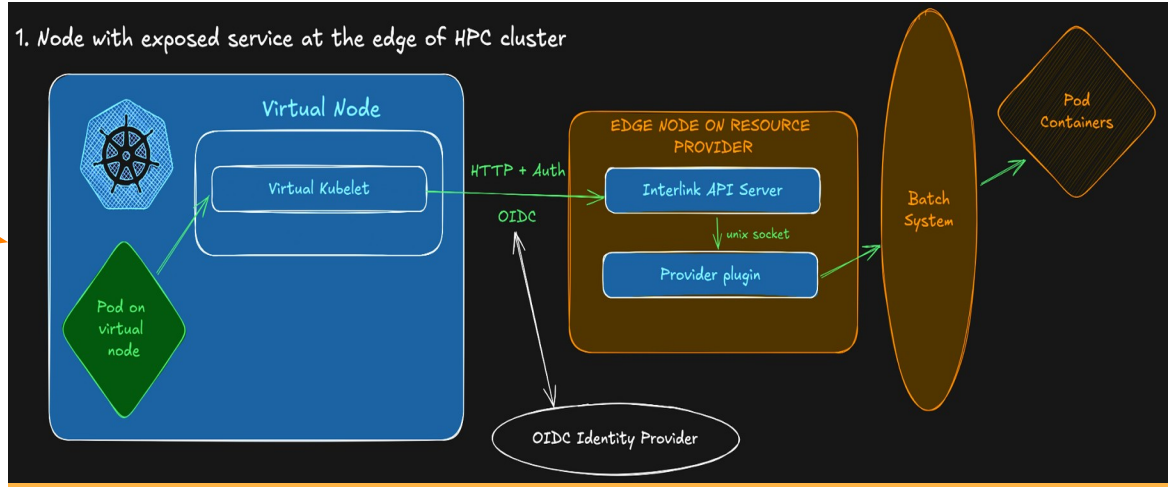
Workload Offloading: extending the container orchestration layer (K8s) to support the execution of a user payload to a remote resource instead of a physical node of the cluster



Virtual Kubelet and interLink



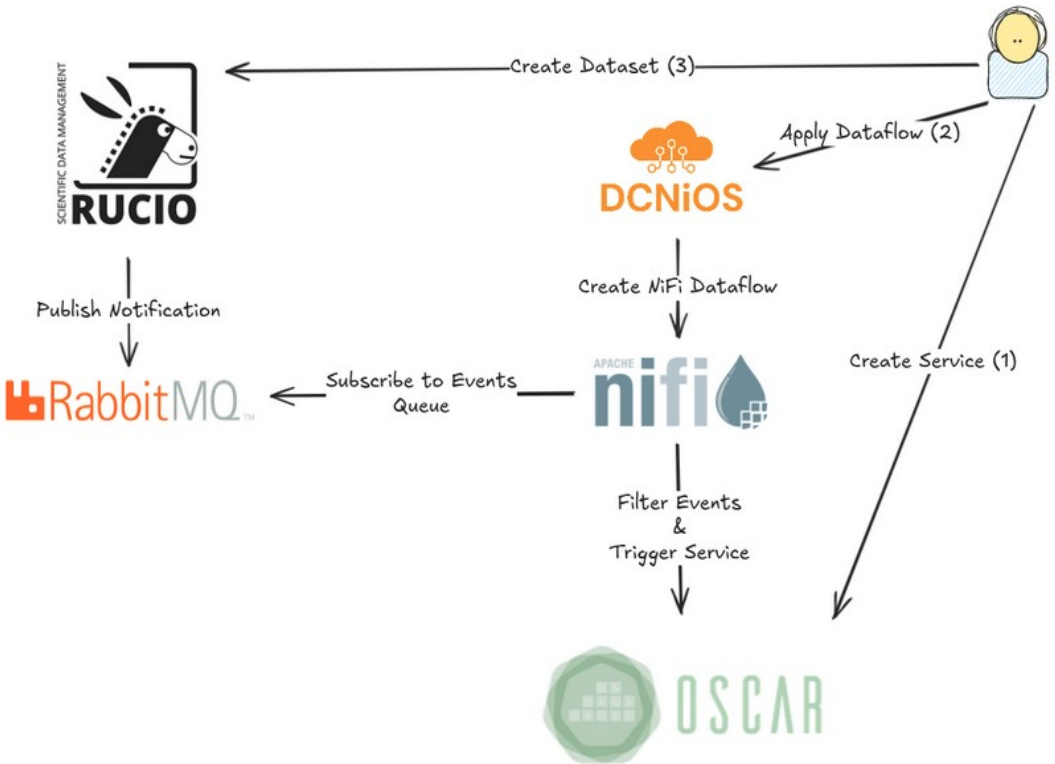
- **Virtual Kubelet** (a virtual node) translates Kubernetes pod execution requests into remote calls to the **interLink API server**
- **interLink API Server** – a modular, pluggable REST server with provider-specific plugins (sidecars) for different execution environments



Event triggered workflows

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

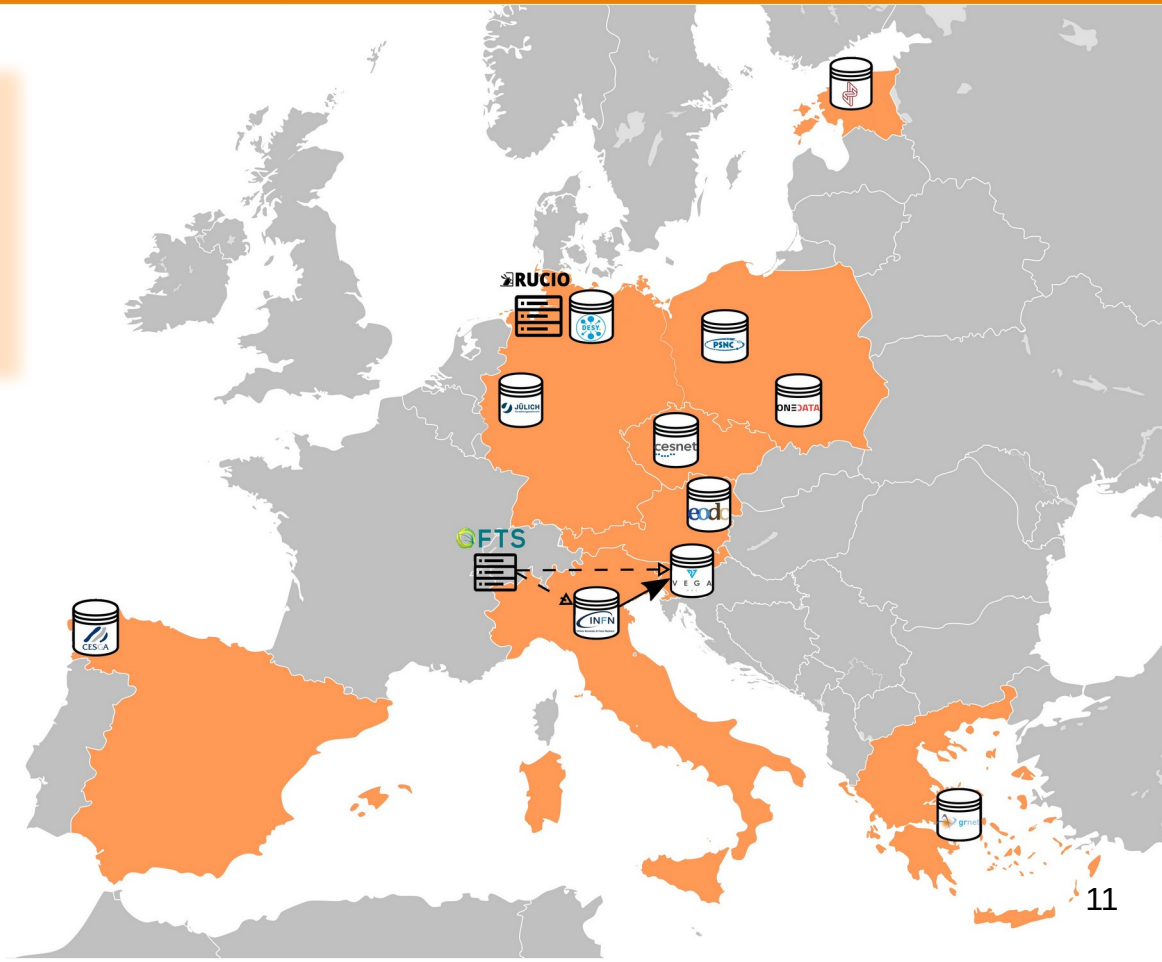
- **OSCAR** is an **open-source platform** that **supports the event-driven** serverless computing **model for data-processing applications**
- User uploads data to the data lake, creates and closes a dataset
- **Rucio sends an event when the dataset is closed using RabbitMQ** – an open-source message-broker
- **OSCAR** subscribes to Rucio's event queue using **Apache NiFi**, which filters events and upon the desired event triggers the service to execute the desired workflow



interTwin Data Lake

- The interTwin data lake consists of:
  - Federation of **storage resources** across Europe
  - Data management software **Rucio**
  - **FTS** – a File Transfer Service

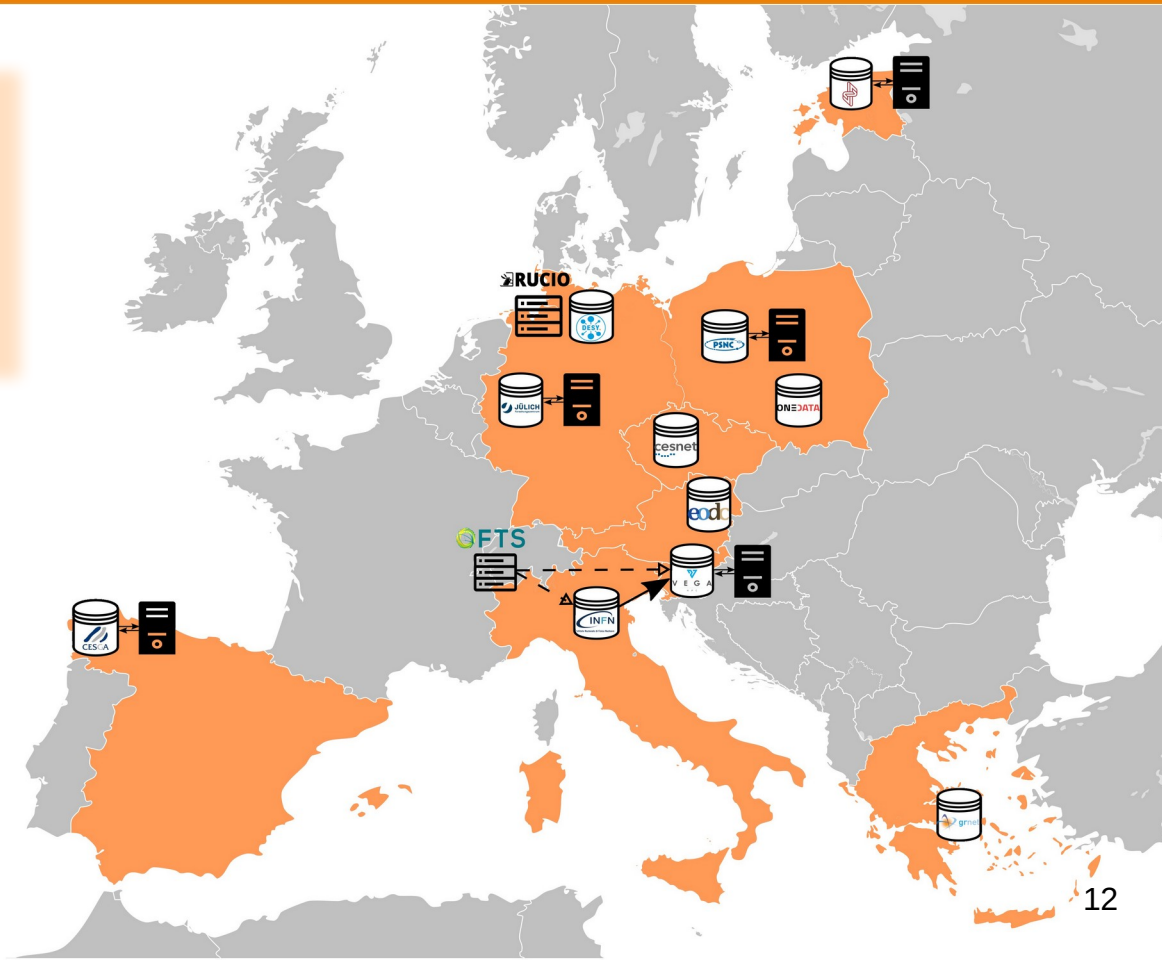
- Token-based authentication, with **EGI Check-in** as the Identity Provider
- VO-membership-based authorization
- User interacts with the data lake via Rucio's client



interTwin Data Lake – Compute integration

- The interTwin data lake consists of:
  - Federation of **storage resources** across Europe
  - Data management software **Rucio**
  - **FTS** – a File Transfer Service

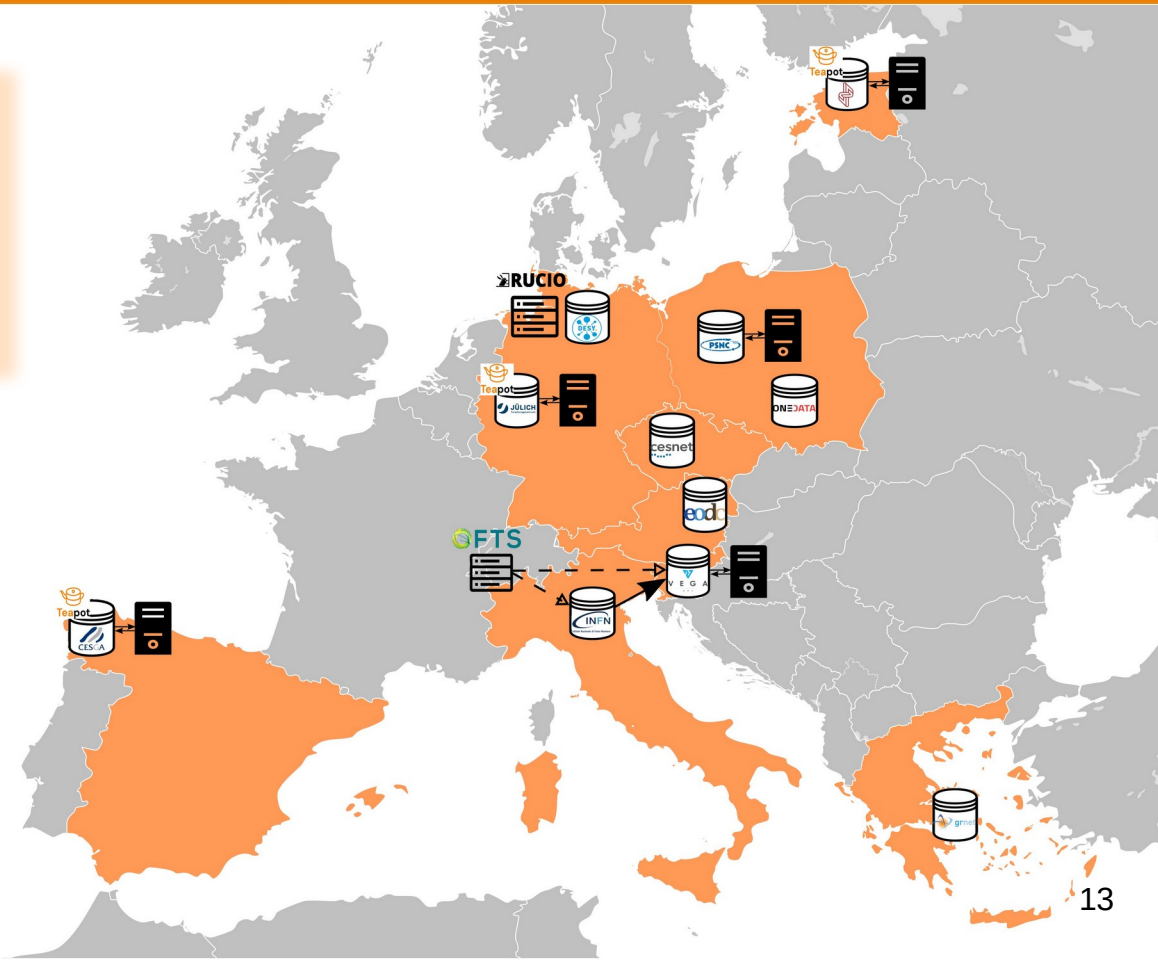
- At sites that also host compute resources, **the data at the local storage system needs to be accessible from the local compute resource** for users to have access to the data in the data lake
- We **enabled the HPC centers to mount the local dCache storage system to the compute nodes** (VEGA and PSNC)



interTwin Data Lake – Teapot

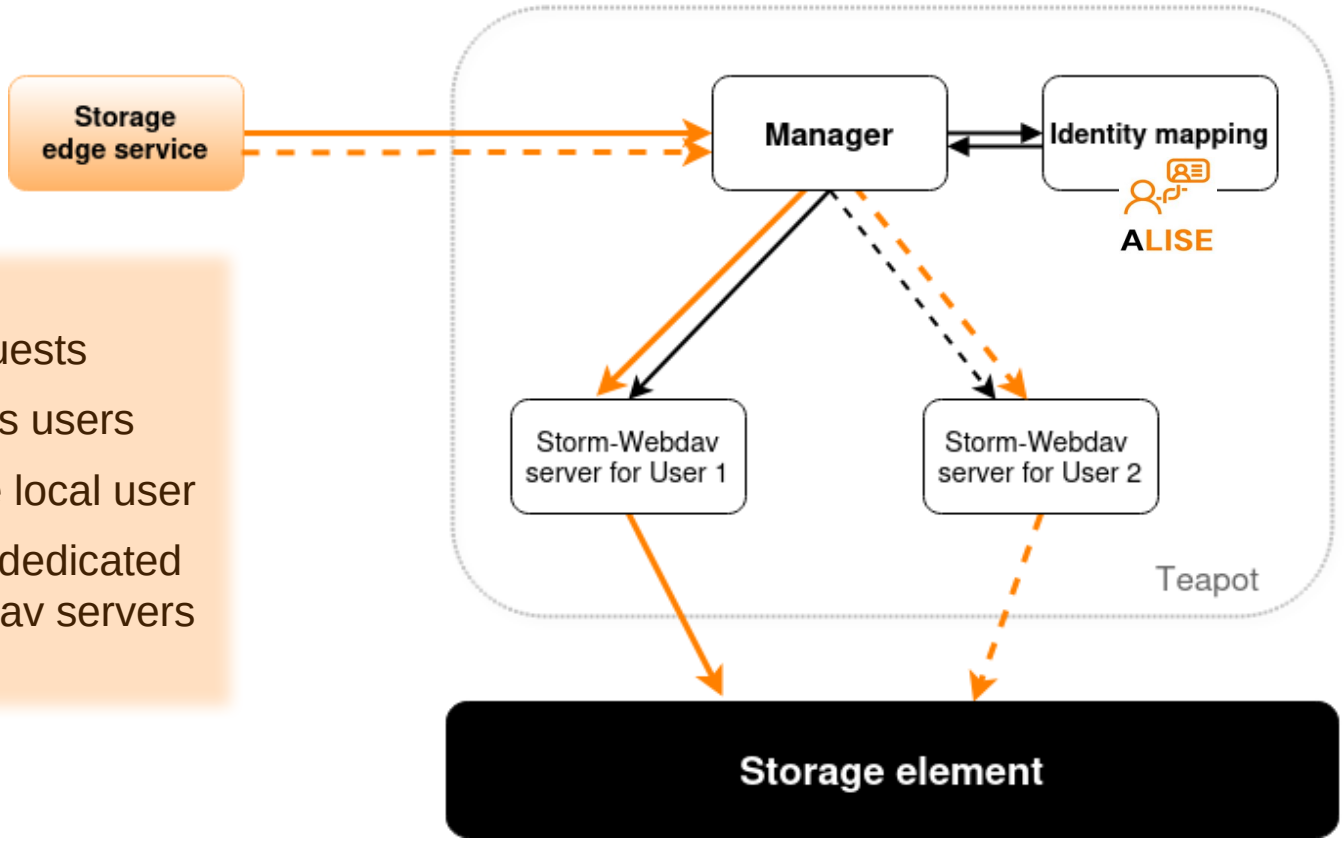
- The interTwin data lake consists of:
  - Federation of **storage resources** across Europe
  - Data management software **Rucio**
  - **FTS** – a File Transfer Service

- Teapot is a **lightweight, easy-to-deploy** WebDAV application that **supports multi-tenancy**
- Based on Storm-Webdav
- **Exposes allocated storage to the internet via WebDAV**
- **Preserves file ownership**
- **A solution suitable for HPC storage**



Teapot

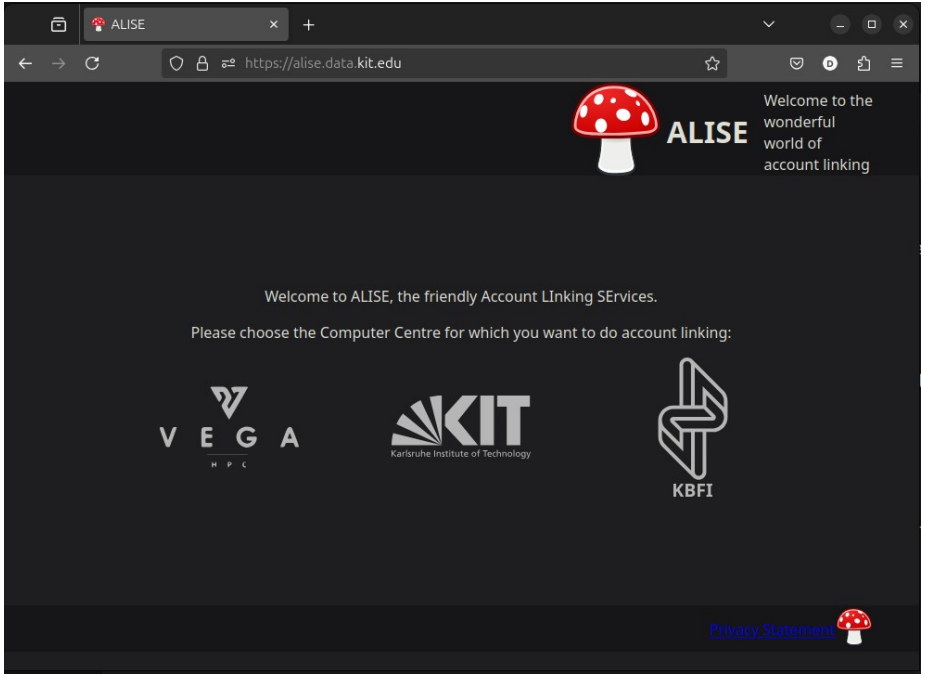
- Accepts requests
- Authenticates users
- Identifies the local user
- Starts/stops dedicated Storm-Webdav servers



ALISE – Account LInking SErvice

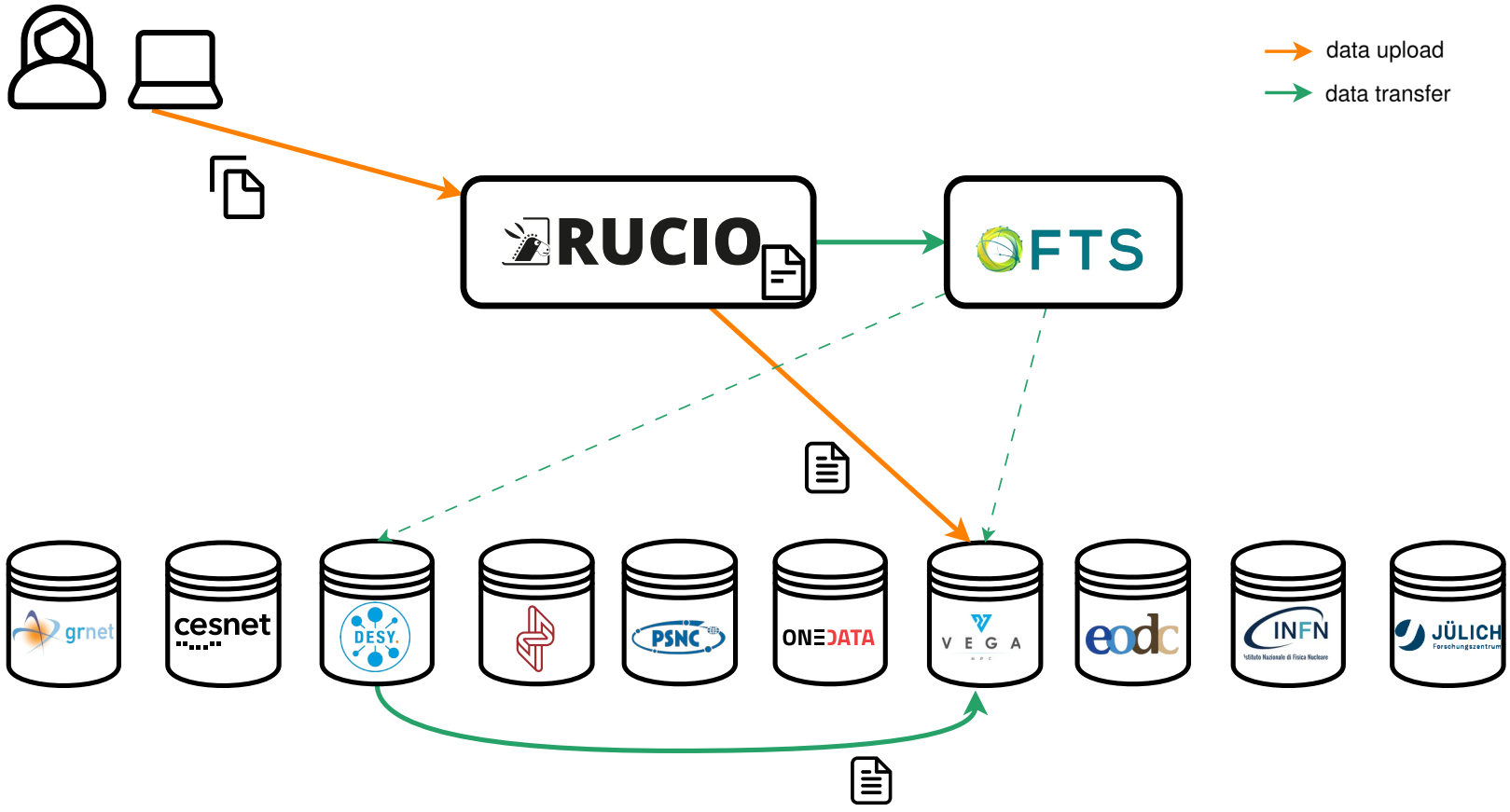
- A tool for **linking a user's federated identity to their facility account**

- An **easy-to-deploy standalone service**
- ALISE provides **an automated procedure** for users of a facility **to register their federated identities**
- ALISE instance **allows sites to deploy other services that require OIDC/token-based authentication**



Karlsruher Institut für Technologie

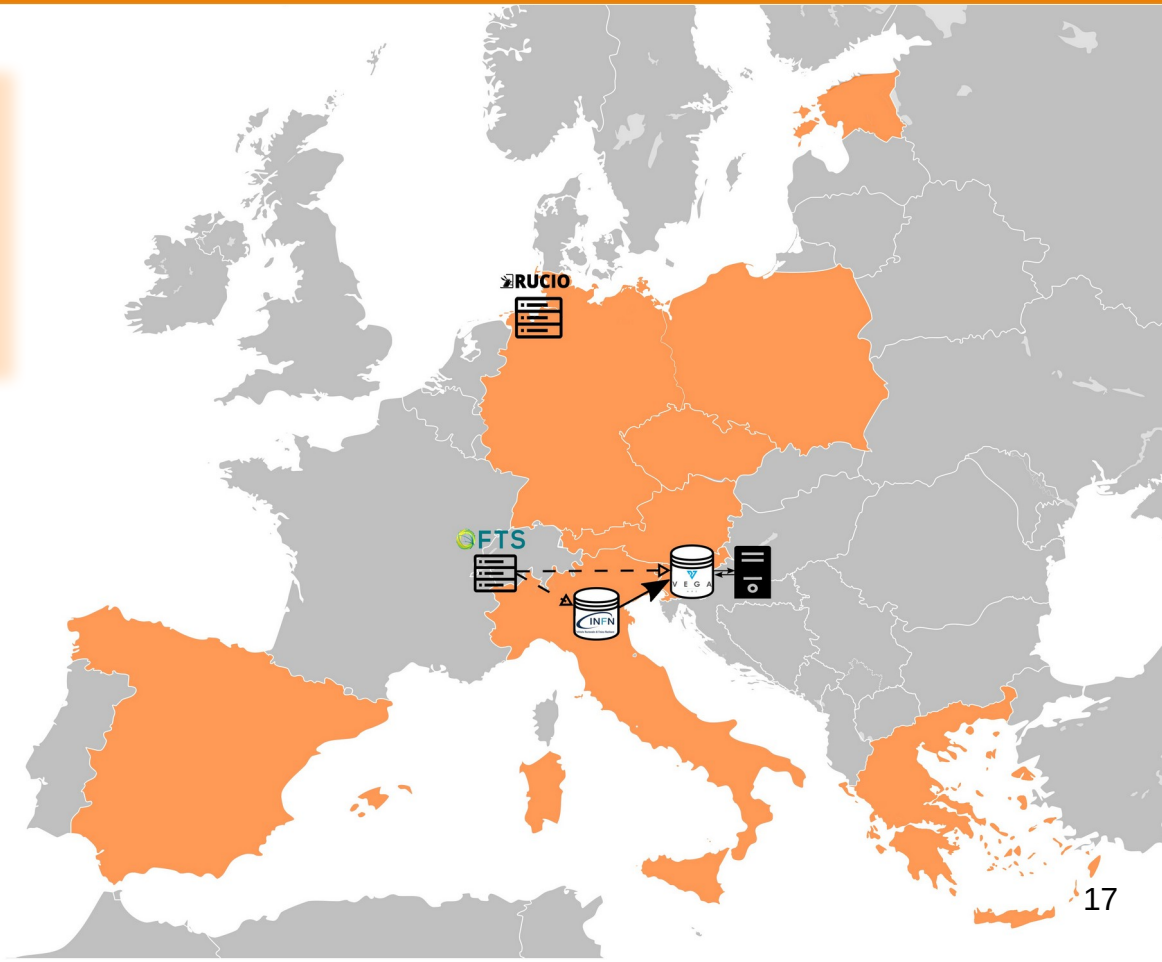
Data movement



Integration with the use cases – Virgo

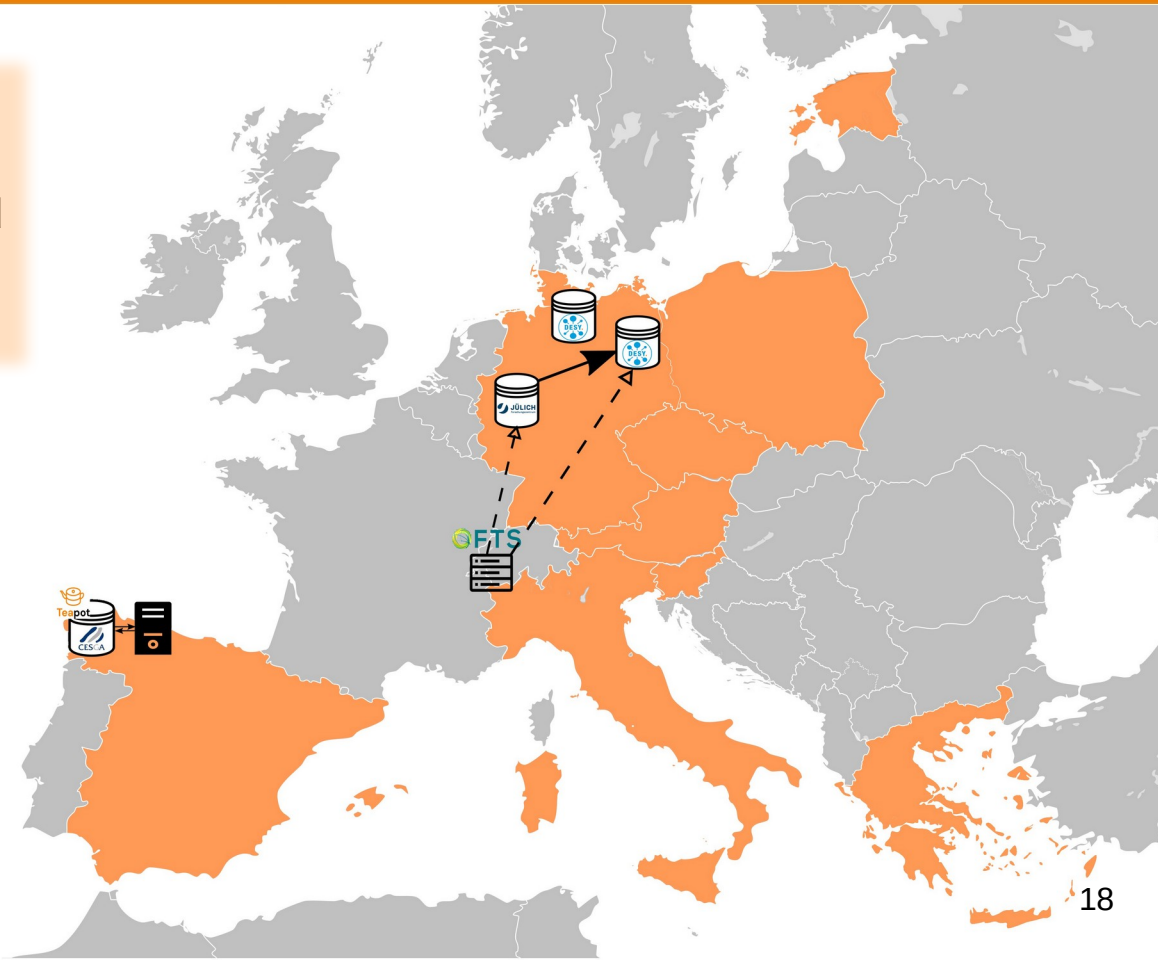
- The virgo data lake consists of:
  - Federation of **storage resources** at **INFN** and **EuroHPC Vega**
  - Data management software **Rucio**
  - **FTS** – a File Transfer Service

- **Virgo Consortium** – behind the Virgo Noise Detector use case – **is storing their private data in the data lake**
- To protect the virgo data **we created a separate data lake**
- **Authorization** to virgo data lake **based on membership in virgo.intertwin.eu VO**



Integration with the use cases – LatticeQCD

- The LatticeQCD data-lake-based solution consists of:
  - Federation of **storage resources** around Europe
  - **FTS** – a File Transfer Service
- **LatticeQCD data managed with an ILDG catalog**
- **AuthN/Z** done via ILDG catalog's **Indigo IAM**
- Using a **fine-grained authorization model**
- **Scope-based authorization policy**
- Building a **data-lake-based solution** for the LatticeQCD community **that is not managed by Rucio**



Thank you for your attention!

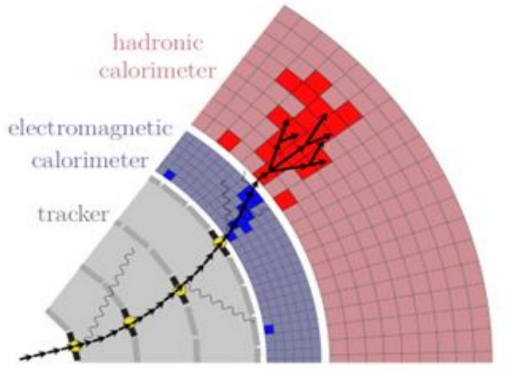
Digital Twin

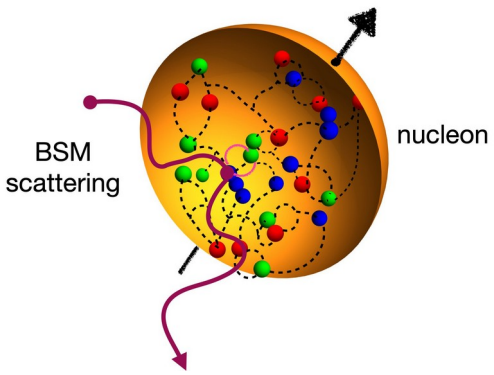
- A **digital twin** is a digital model of an intended or actual real-world physical product, system, or process (a physical twin) that serves as the effectively indistinguishable digital counterpart of it for practical purposes, such as simulation, integration, testing, monitoring, and maintenance
- It is a **virtual environment** – It can run any number of useful simulations in order to study multiple processes
- It is based on **real-time data** – By having better and constantly updated data related to a wide range of areas digital twins are able to study more issues from far more vantage points than standard simulations can

Scientific use cases: High Energy Physics

DT of Large Hadron Collider (LHC) detector components

Seeking for strategies to face the increase in the need for simulated data expected during the future High Luminosity LHC runs. The primary goal is to provide a fast simulation solution to complement the Monte Carlo approach. **Faster and deeper cycles of optimisation of the experiment parameters** in turn will enable breakthroughs in experimental design.





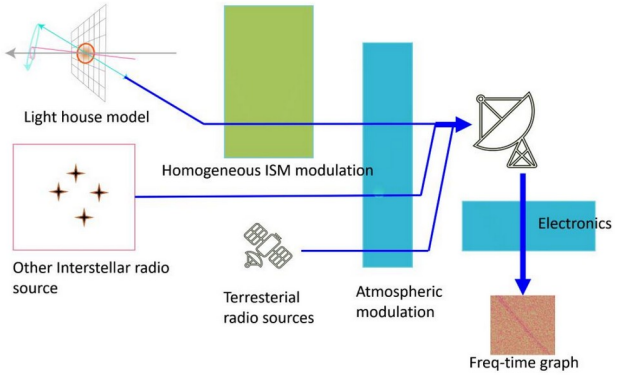
DT of the Standard Model in particle physics (Lattice QCD)

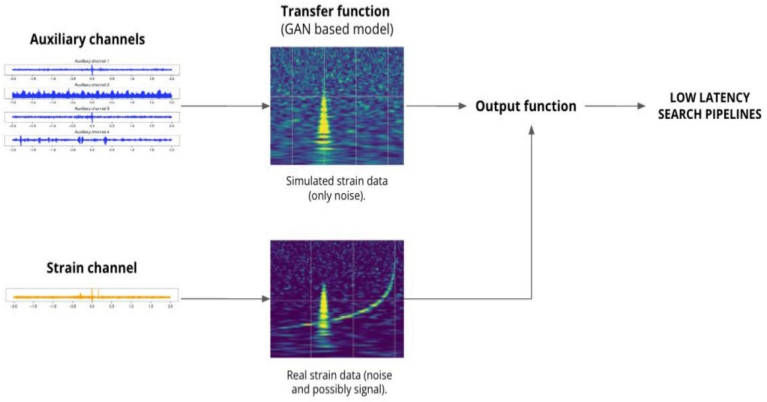
Competitive results in Lattice QCD require the **efficient handling of Petabytes of data**, therefore the implementation of advanced data management tools is mandatory. On the side of algorithmic advancement, ML algorithms have recently started to be applied in Lattice QCD. The goal is to **systematize the inclusion of ML for large scale parallel simulations.**

Scientific use cases: Astronomy

DT for noise simulation of next-generation radio telescopes

Providing DTs to simulate the noise background of radio telescopes (**MeerKat**) will support the identification of rare astrophysical signals in (near-)real time. The result will contribute to a realisation of "**dynamic filtering**" (i.e. steering the control system of telescopes/sensors in real-time).





DT of the Virgo Interferometer

Meant to **realistically simulate** the noise in the detector, in order to study how it reacts to external disturbances and, in the perspective of the Einstein Telescope, to be able to detect noise “glitches” in **quasi-real time**, which is currently not possible. This will allow sending out **more reliable triggers** to observatories for multi-messenger astronomy.

DTE Infrastructure: How is made

Middleware toolkit

interLink
Teapot Rucio FTS ALISE
INDIGO PaaS Orchestrator
APEL

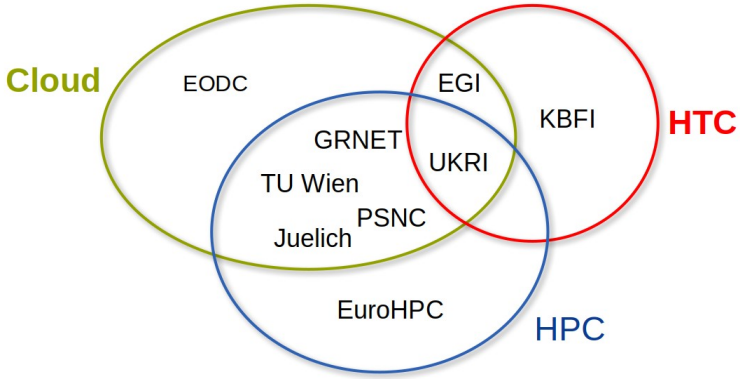
To enable the compute continuum
seamlessly integrate Cloud and word-class HPC and HTC

Abstracting storage infrastructure complexity
enabling a DataLake model offering data management capabilities even on word-class HPC

Federating and Orchestrating Clouds to deploy high level services
Instantiate end user services of the DTE on the most suitable cloud infrastructure

Effective resource usage accounting
Including specialized HW and resources provisioned via container orchestrators

Computing & Storage Capacity

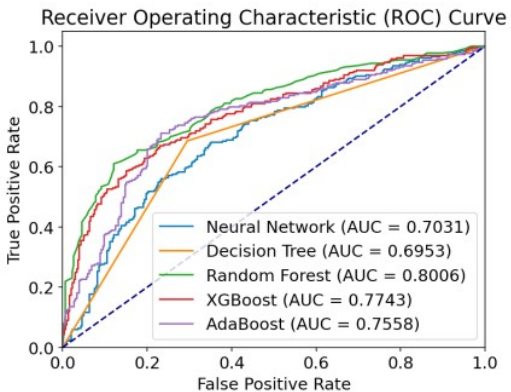


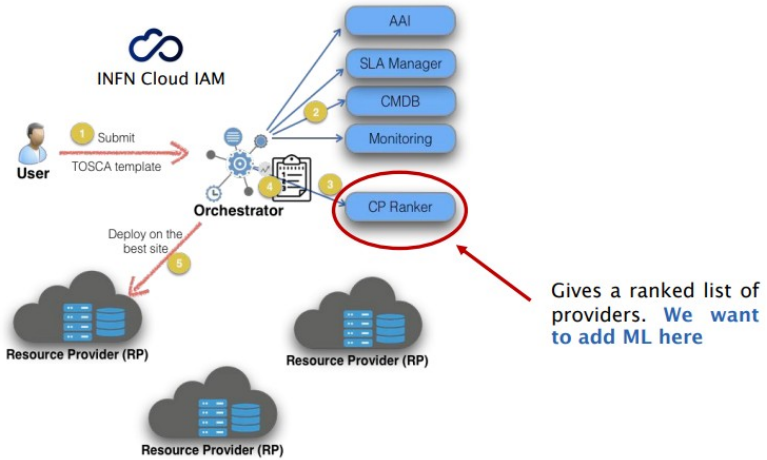
Intelligent Orchestration

The INDIGO PaaS Orchestrator receives high-level deployment requests in the form of TOSCA templates and coordinates the process of creating deployments (DTE Core high level services)

Prototyping Intelligent Orchestration

- Making the provider selection process more dynamic with AI-based models
- **A predictive model to forecast deployment success and a regression model to predict deployment times**





Replacing legacy PaaS Components

- Implemented a brand new service, the Federation-Registry, to replace the Configuration Management Database
- **Crucial for enabling matchmaking between users' deployment requests and providers' capabilities**