EGI: Advanced Computing for Research

www.egi.eu
@EGI_eInfra

European computing infrastructures for digital twins: the EGI project 'interTwin'

T. Ferrari, A. Manzi/EGI Foundation

contact@egi.eu

The work of the EGI Foundation
is partly funded by the European Commission
under H2020 Framework Programme


EGI Federation
European flagship digital infrastructure for data-intensive scientific computing

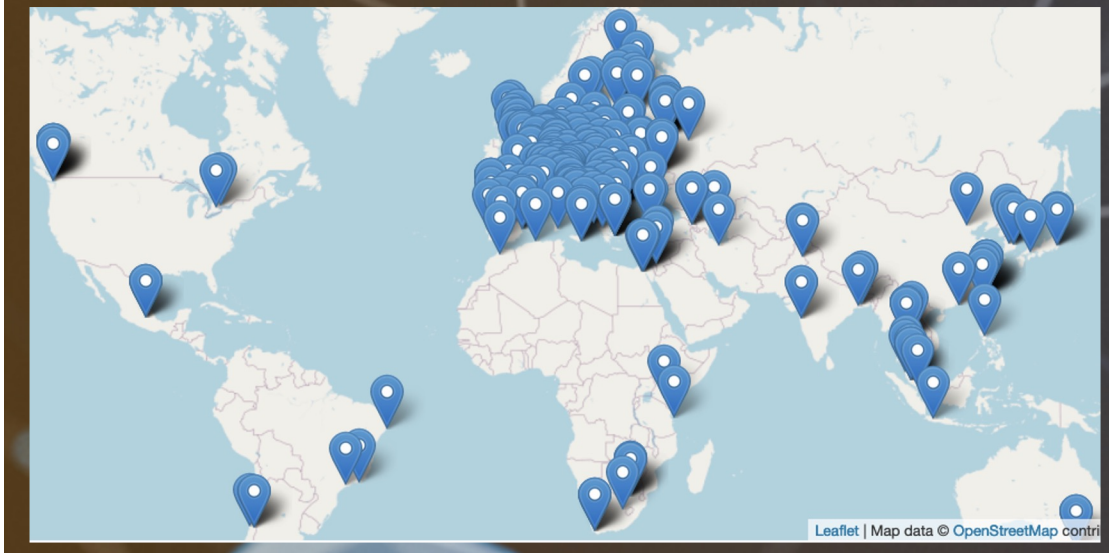
Why a federation?

Support science at international scale

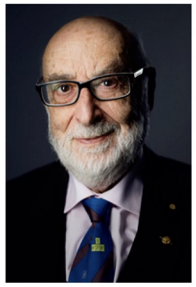
Build an hyperscale facility for research

Invest nationally, access globally

Bring computing to the data



The Nobel Prize in Physics 2013

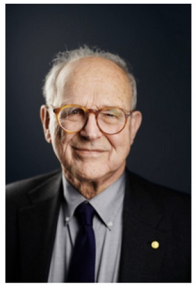


© Nobel Media AB. Photo: A. Mahmoud
François Englert
Prize share: 1/2

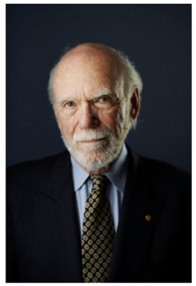
© Nobel Media AB. Photo: A. Mahmoud
Peter W. Higgs
Prize share: 1/2

The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider."

The Nobel Prize in Physics 2017



© Nobel Media AB. Photo: A. Mahmoud
Rainer Weiss
Prize share: 1/2

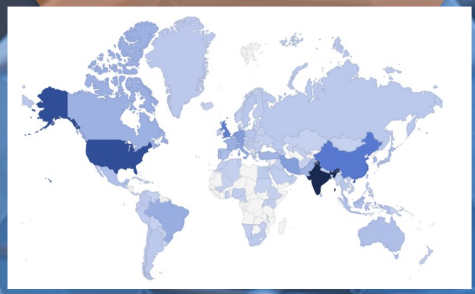


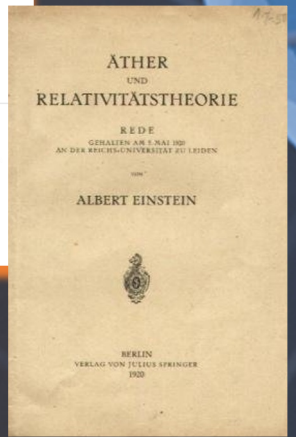
© Nobel Media AB. Photo: A.Mahmoud
Barry C. Barish
Prize share: 1/4

© Nobel Media AB. Photo: A.Mahmoud
Kip S. Thorne
Prize share: 1/4

The Nobel Prize in Physics 2017 was divided, one half awarded to Rainer Weiss, the other half jointly to Barry C. Barish and Kip S. Thorne "for decisive contributions to the LIGO detector and the observation of gravitational waves."

78,100 researchers worldwide





EGI Federation and European Research Infrastructures

Environment

Landmarks
EISCAT_3D
EMSO ERIC
LifeWatch ERIC

Health & Food

Landmarks
ELIXIR
INSTRUCT ERIC
BBMRI NEW

Projects
EMPHASIS
METROFOOD-RI NEW

Social & Cultural innovation

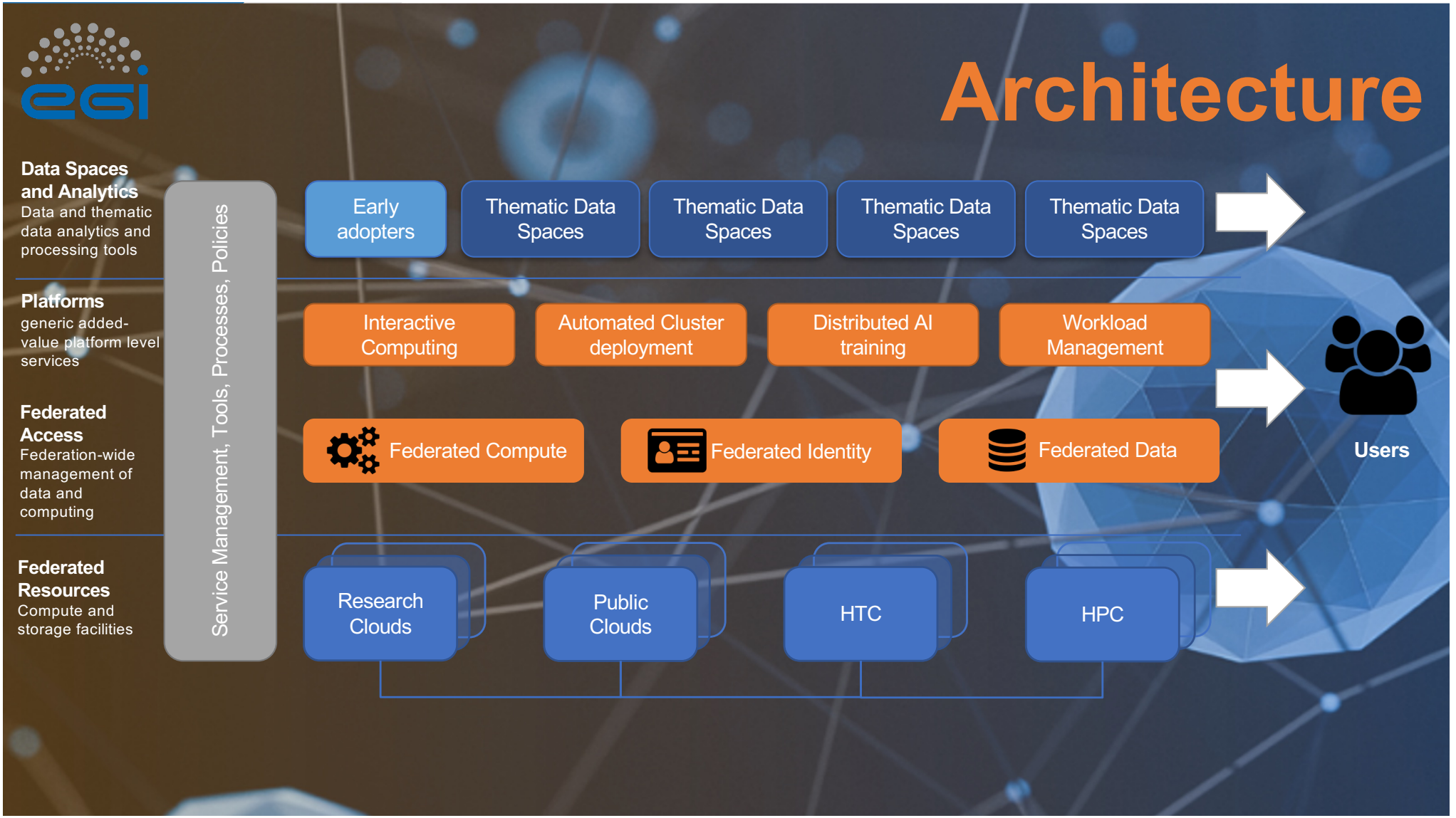
Landmarks
CLARIN ERIC
DARIAH

Projects
E-RIHS NEW
OPERAS

Physical sciences & Engineering

Landmarks
CTA
ELI ERIC
HL-LHC
SKAO

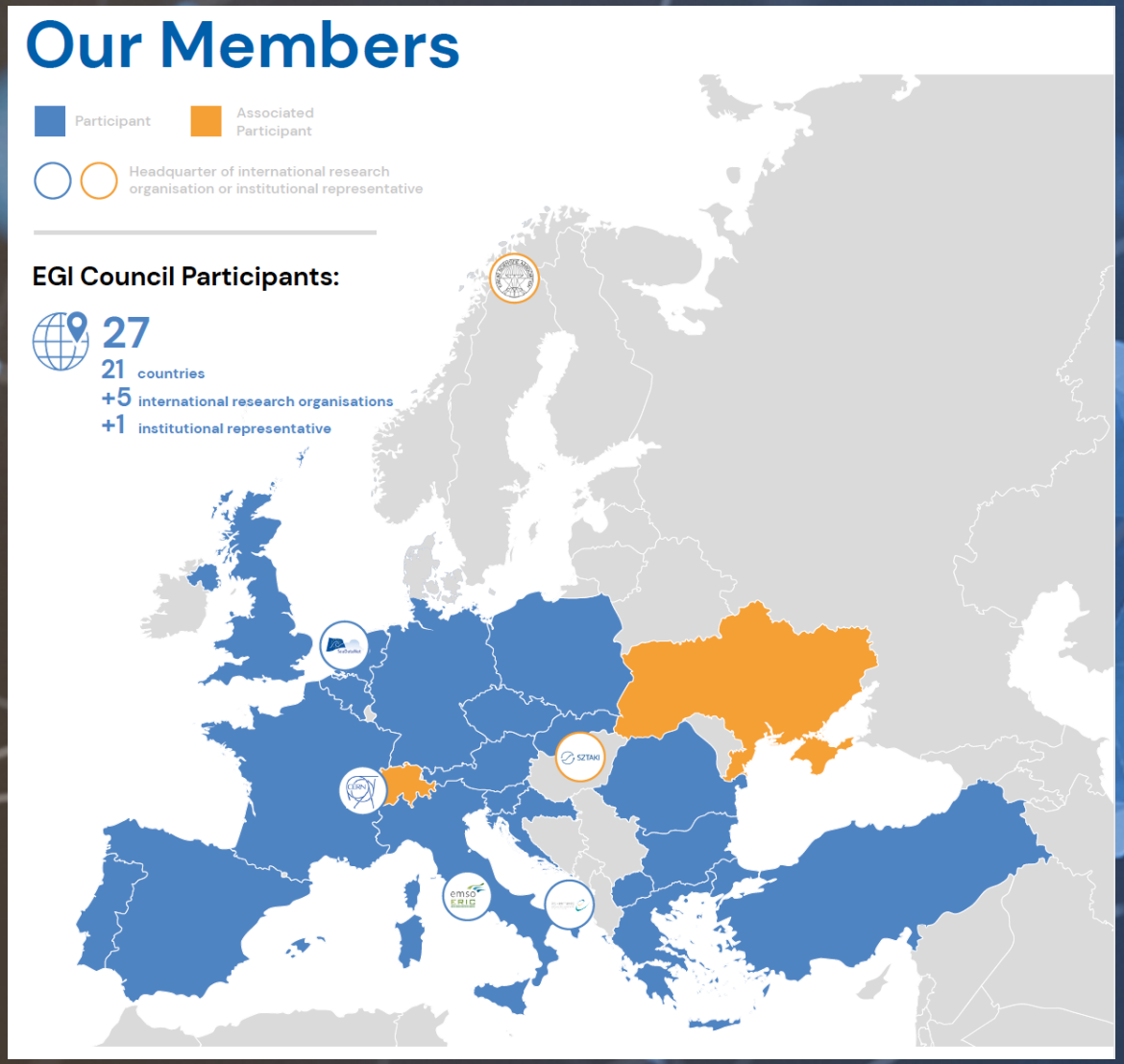
Projects
KM3NeT 2.0


Architecture
Data Spaces and Analytics
Data and thematic data analytics and processing tools
Platforms
generic added-value platform level services
Federated Access
Federation-wide management of data and computing
Federated Resources
Compute and storage facilities
Service Management, Tools, Processes, Policies
Early adopters
Thematic Data Spaces
Thematic Data Spaces
Thematic Data Spaces
Thematic Data Spaces
Interactive Computing
Automated Cluster deployment
Distributed AI training
Workload Management
Federated Compute
Federated Identity
Federated Data
Users
Research Clouds
Public Clouds
HTC
HPC

6.5
Billion
CPU
hours/year

1.3
Exabyte of
data

Our Members



European Open Science Cloud

"EOSC: A multi-disciplinary environment where researchers can publish, find and re-use data, tools and services, enabling them to better conduct their work"

builds on existing infrastructures and services supported by the EC, Member States and research communities. It brings these together in a federated 'system of systems'

EGI delivers the EOSC Compute Platform with the EGI-ACE project

Digital Twins for science

Compute-intensive science involves modelling and simulation at very high resolution for prediction and inference workflows

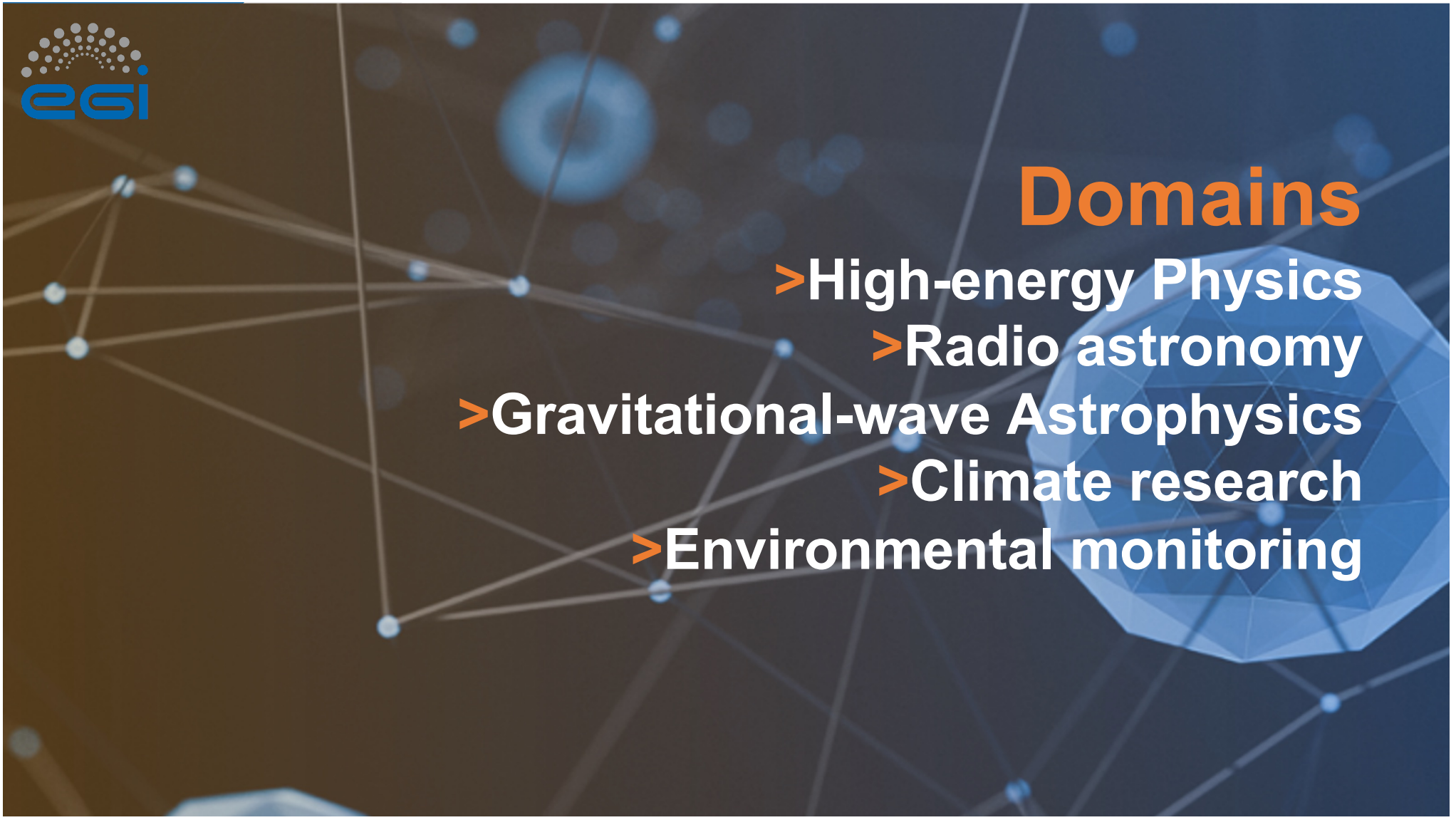
- **From Gigabytes to Petabytes of data per day**
- **Heterogeneous distributed data and computing**

Ambition of the interTwin project

Develop a Digital Twin blueprint architecture and an interdisciplinary Digital Twin Engine

Capabilities to be demonstrated

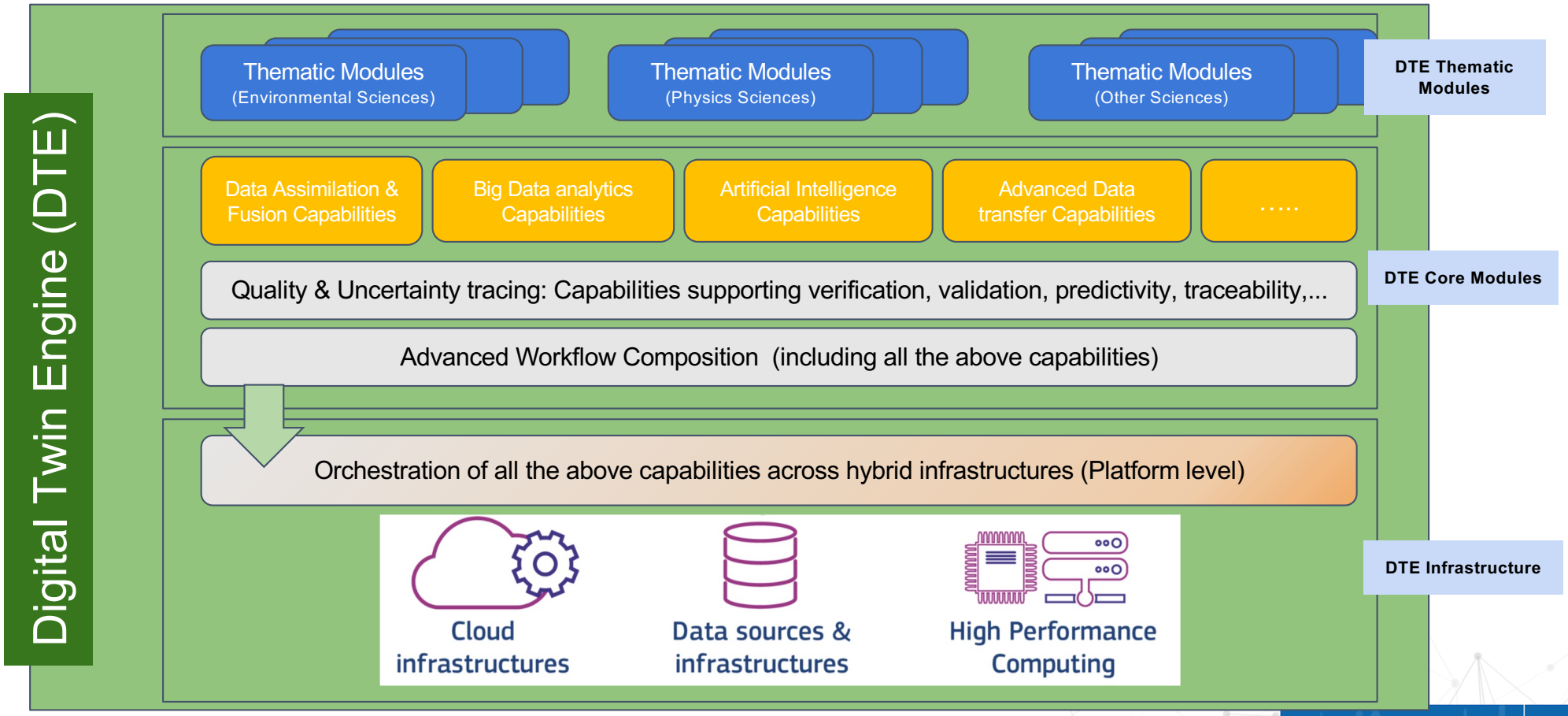
- High-volume and high-speed data acquisition & pre-processing
- Big data assimilation into models
- Collection of simulation models
- Real time processing
- Validation of accuracy


Domains
>High-energy Physics
>Radio astronomy
>Gravitational-wave Astrophysics
>Climate research
>Environmental monitoring

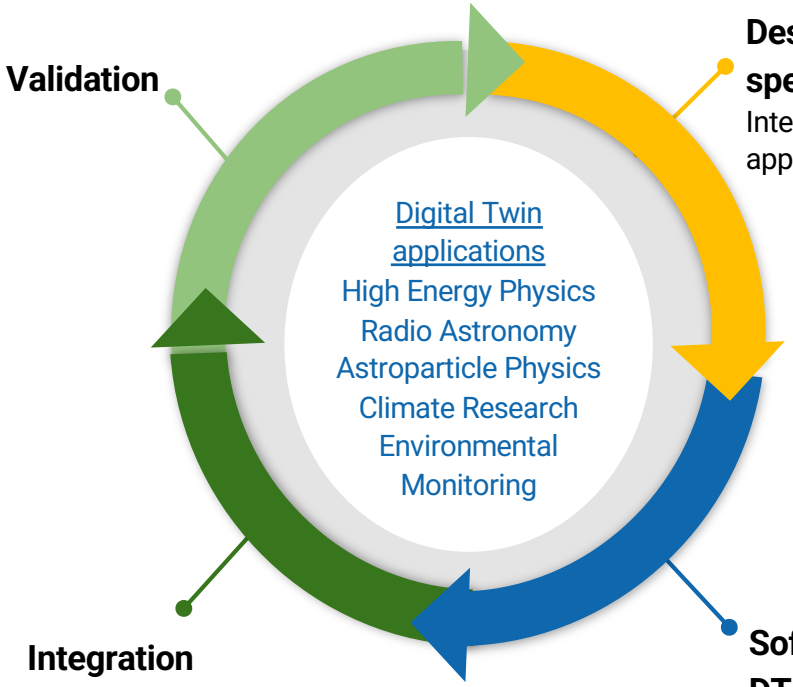
Digital Twin Engine

> Conceptual model of a DT for research - the DTE blueprint architecture
> Open source and open standards
> Integrate with application-specific Digital Twins
> Co-design ed interoperability framework >

Digital Twin Engine - Strawman concept



DTE development cycle



Pre-operational software of a DTE at TRL 6 or 7 depending on the components

2 planned releases (M18 , M31)

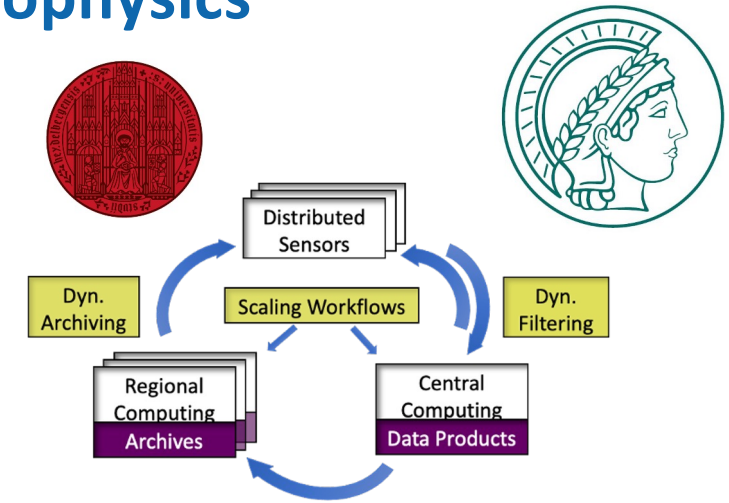
DT Applications: High Energy Physics

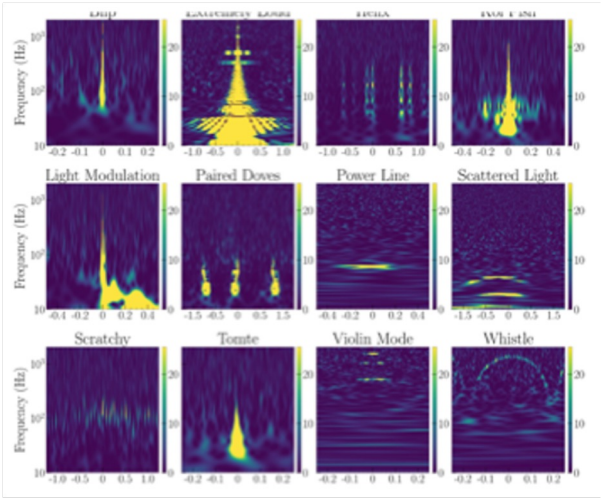
DT of Large Hadron Collider (LHC) detector components: The High Energy Physics community is investigating different strategies to face the increase in the need for simulated data expected during the future High Luminosity LHC runs. The primary goal is to provide a fast simulation solution to complement 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: 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 for generation of gauge configurations. In this project we intend to ***systematize the inclusion of ML for large scale parallel simulations***.

DT Applications: Radio astronomy and Gravitational-wave Astrophysics

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 the low-latency search pipelines to veto or de-noise the signal, sending out ***more reliable triggers*** to observatories for multi-messenger astronomy.

DT Application: Climate change and impact decision support tools

DT 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 of extreme events (storm, fire, floods & drought)

Interoperability & Link with DestinE

interTwin will conduct joint pilot activities with **DestinE** to design a compatible architecture that addresses the requirements of the largest set of user communities.
Interoperability is the aim of this activity.

Demonstrators of data handling across InterTwin and DestinE DTs for the Extremes and Climate in production-type configurations will be implemented in collaboration with **ECMVF**

Part of the collaboration with DestinE includes the development of common software architecture concepts that are also applicable to other major DTs initiatives.

ECMWF

Conclusions

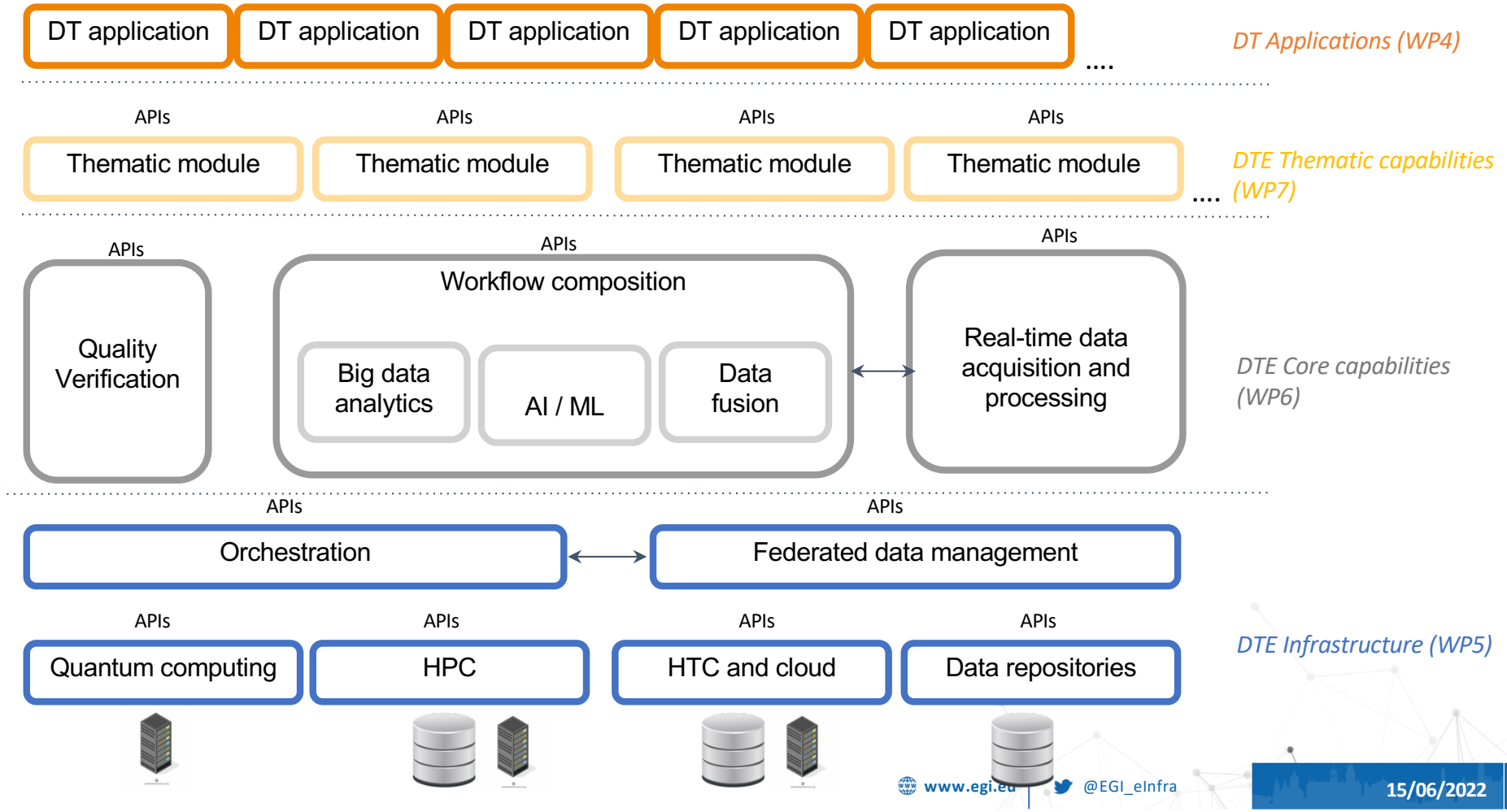
> Co-design a Digital Twin Engine blueprint architecture

> The project will advance the current EGI infrastructure offering (and EOSC capabilities)

> Aim to foster an Open Source community

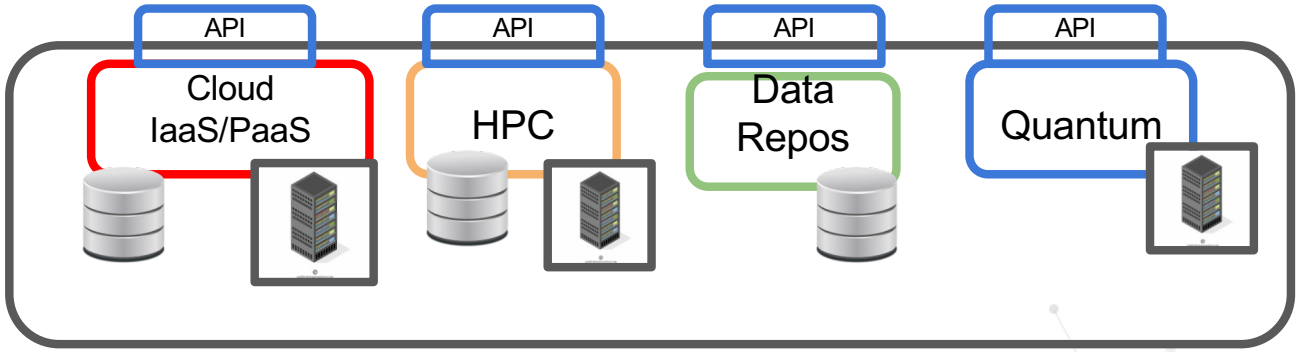
For more information

interTwin components



DTE Infrastructure: Computing and Storage

- Solutions to provision resources on a wide range of compute providers, including HTC, HPC, Cloud, Quantum systems and a stable interface to the storage systems of interest for the use cases in the project.
- Integral part of the Infrastructure are resource providers who allow access to their resources within the project: *EODC (Austria), FZJ (Germany), GRNET (Greece), IZUM (Slovenia), JSI (Slovenia), KBFI (Estonia), KIT (Germany), PSNC (Poland), TU Wien (Austria) and UKRI -STFC (UK)*
- Access to Quantum systems via the collaboration with *ETH Zurich (Switzerland)*

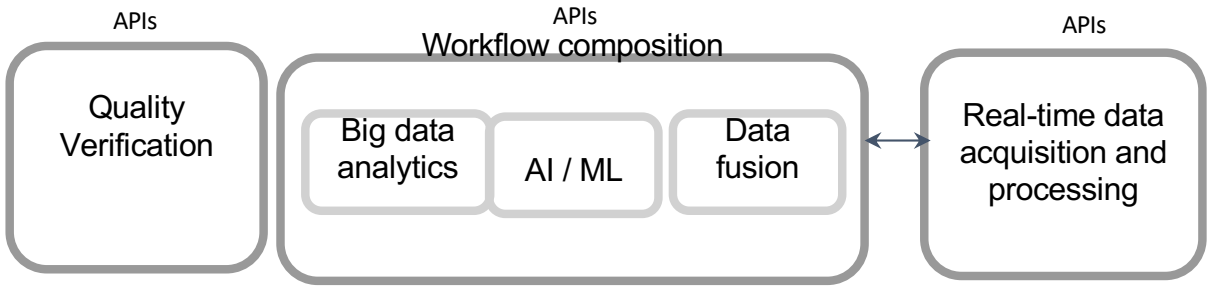


DTE Infrastructure: AI Orchestrator

- In order to match storage and compute in complex workflows like those expected in the project, a series of non trivial decision must be taken
  - Choose data source, if more than one available
  - Choose compute site, if more than one available
  - How to make data available to compute, which can mean
    - Use computing close to data
    - Transfer data closer to compute
    - Instantiate caches at/close to the compute site
    - Instantiate proxies to allow a streaming access to data

- An orchestrator with predictive AI capabilities will be developed, basing its decisions on static configurations (site description, network connections, cost, ...) as well as on the data collected in previous runs

DTE Core Capabilities



The DTE Core Modules offer horizontal capabilities to facilitate the creation and the operations of data-intensive and compute-intensive DT applications, e.g.:

- **Quality verification**: "Model Validation as a Service" enabling customisations of best practices and standard quality measures for scientific disciplines and applications. It applies DevOps practices to automate the process.
- **AI workflow and method lifecycle management:** a customizable toolkit to realize complex AI setups. It offers capabilities such as: model training implementing a pre-selection of (pre-trained) models, model evaluation to offer standard evaluation metrics and plots, a distributed training framework, distributed hyperspace optimization
- **Data Fusion** implements and integrates processes for merging datasets from different sources. This includes linking and visualization of observational and modelled data, and the harmonization of different types of observational data like gridded datasets with vector based datasets like point streams of data from ground stations.

DTE Thematic Capabilities

The **interTwin DTE Thematic modules** are addons providing capabilities tailored to the needs of specific groups of applications (i.e. of general applicability to multiple 'adjacent' communities) developed with the aim to be "promoted" as Core modules following the successful adoption by multiple resource communities from different domains:

- **Lattice QCD simulations and data management**
- **Noise simulation for radio astronomy**
- **GAN-based thematic modules to manage noise simulation, low-latency de-noising and veto generation**
- **Climate analytics and data processing**
- **Earth Observation Modelling and Processing**
- **Hydrological model data processing**
- **Fast simulation with GAN**