interTwin

And interdisciplinary Digital Twin for Particle Physics

Dr. Isabel Campos Plasencia

CSIC Consejo Superior de Investigaciones Científicas

IFCA Instituto de Física de Cantabria

POWERED BY egi POWERED BY

L International Meeting on Fundamental Physics & XV CPAN days

Santander, October 2023

Funded by the European Union

The interTwin project is funded by the European Union - Grant Agreement Number 101058386

IFCA Computing

Ongoing and recent Projects and initiatives
Participation in a wide range of projects and initiatives in areas such as:

- Quality assurance
- Software management
- Linux containers
- Data repositories
- Federated/distributed computing
- High Throughput (Grid) Computing
- Cloud Computing
- High Performance Computing
- Digital Twins

Digital Twin

A <u>virtual replica designed to accurately reflect real world</u> objects, systems or processes.

- a high-fidelity model of a system which can be used to emulate the actual real system

A <u>digital twin goes beyond mere simulation</u>:

- is much a richer virtual environment that mimics the physical one
- enables study and prognosis of multiple processes
- can be feed with real data

First Digital Twin: born at NASA

Apollo Simulators installed at NASA - Front - Lunar Module Simulator (green)
Mid- Mission Effects Projector-Lunar surface (green)
Back - Command Module Simulator (tan)

"Keep 'em Running," screamed Astronaut Eugene A. Cernan as he strode into the simulator room late on that fateful Monday night."

(Connecting Link magazine, Summer 1970 issue)

1960s as a "living model" of the Apollo mission. In response to Apollo 13's oxygen tank explosion

- NASA employed simulators to evaluate the failure and extended a physical model of the vehicle including digital components.

The "digital twin" was able to rescue the mission by simulating

Destination Earth

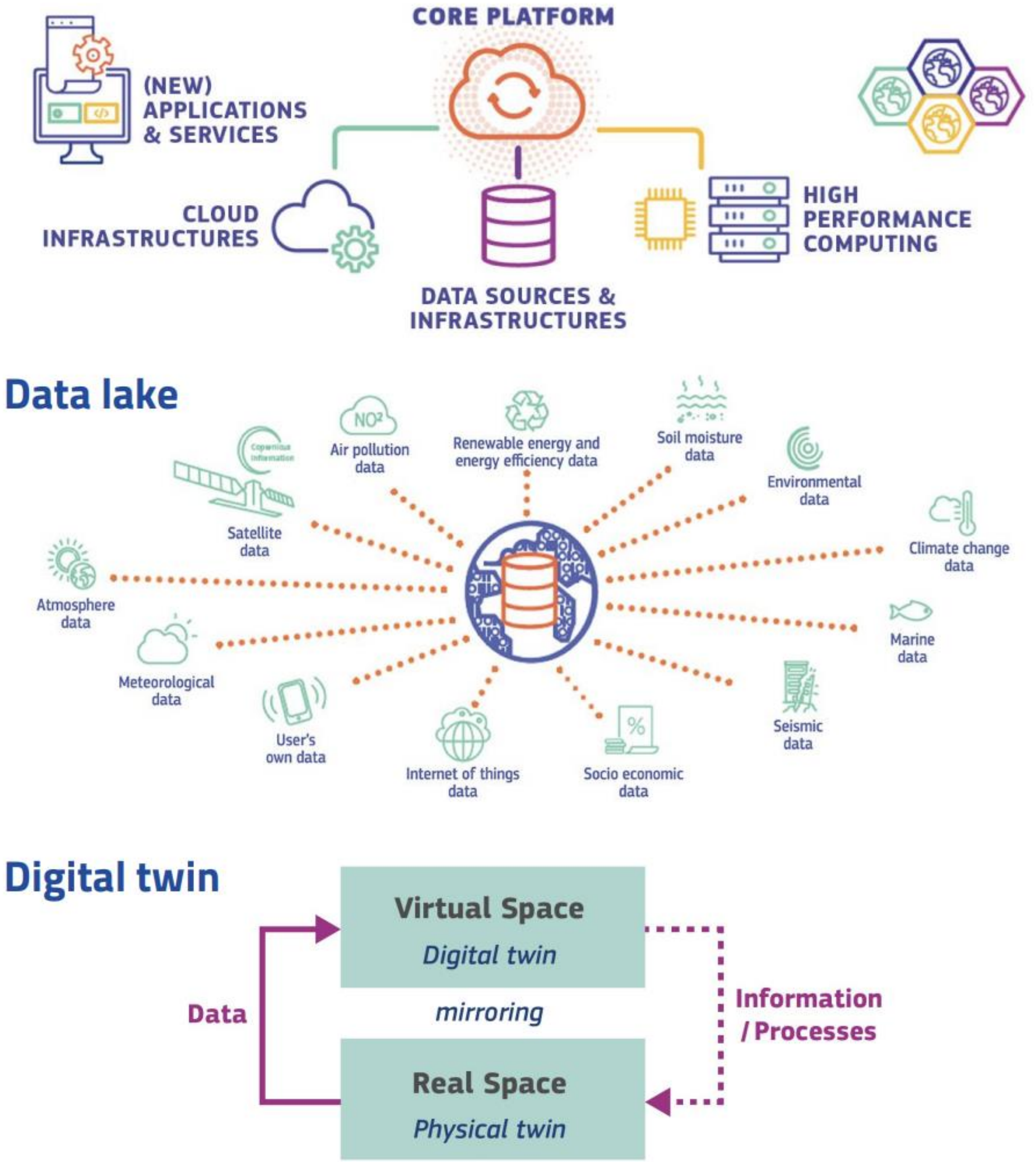
Is a flagship initiative of the European Commission to develop a highly-accurate digital model of the Earth: **a digital twin of the Earth**

Understand and simulate the evolution and behavior of the Earth system components.

Digital modelling of the Earth system

Provide evidence-based decision-making tools.

Based on an open, flexible, and secure cloud-based computing infrastructure with access to data and HPC.



www.intertwin.eu

Horizon Europe

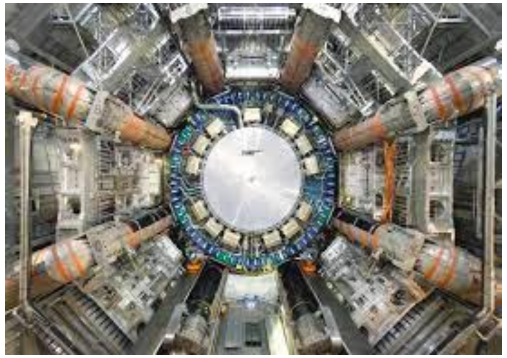
THE NEXT EU RESEARCH & INNOVATION PROGRAMME (2021 – 2027)

www.dtgeo.eu

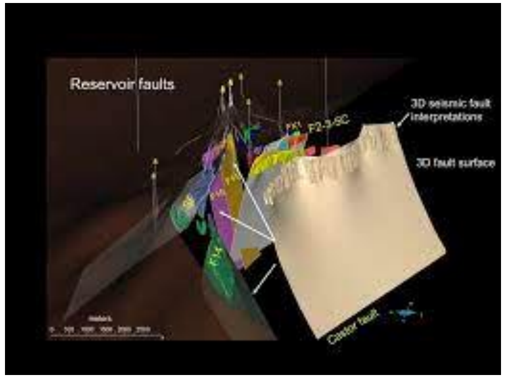
interTwin

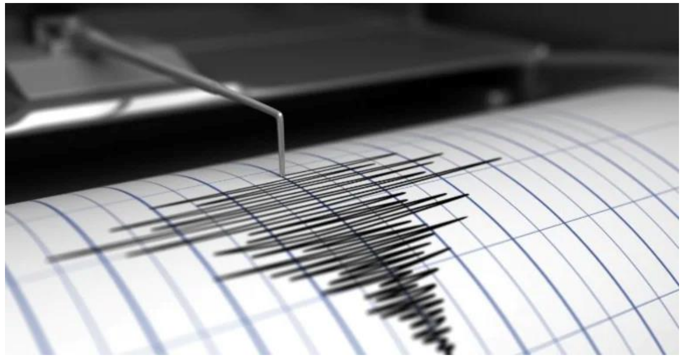
DT-GEO

InterTwin → Interdisciplinary

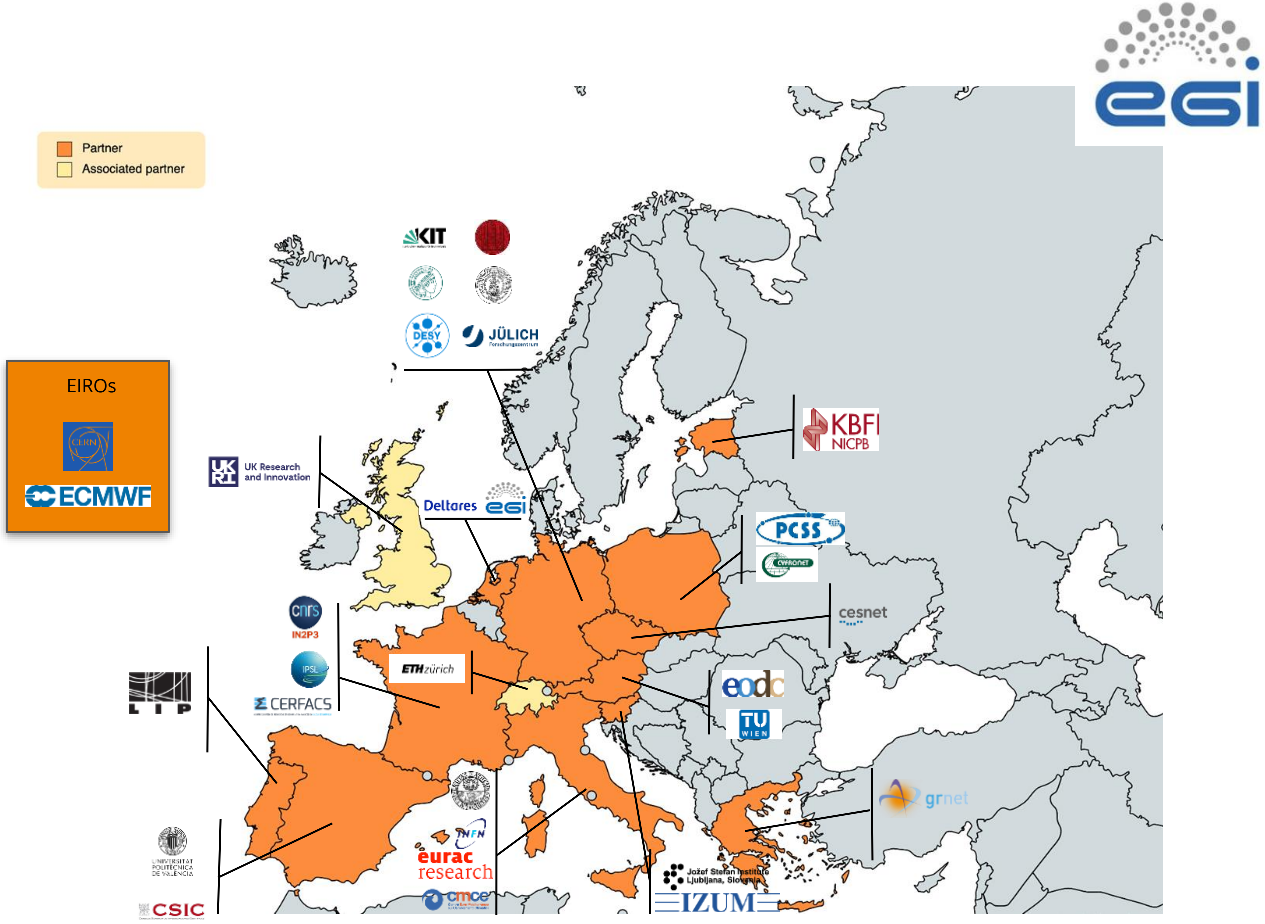


DT-GEO → Geophysical extremes





Consortium



1.09.22 - 31.08.25

Budget 11,7 M euro

EGI Foundation as coordinator

29 **Participants**, including 1 affiliated entity and 2 associated partners

Consortium at a glance

10 Providers	11 Technology providers	14 Community representants
cloud, HTC , HPC resources and access to Quantum systems	delivering the DTE infrastructure and horizontal capabilities	from 5 scientific areas; requirements and developing DT applications and thematic modules

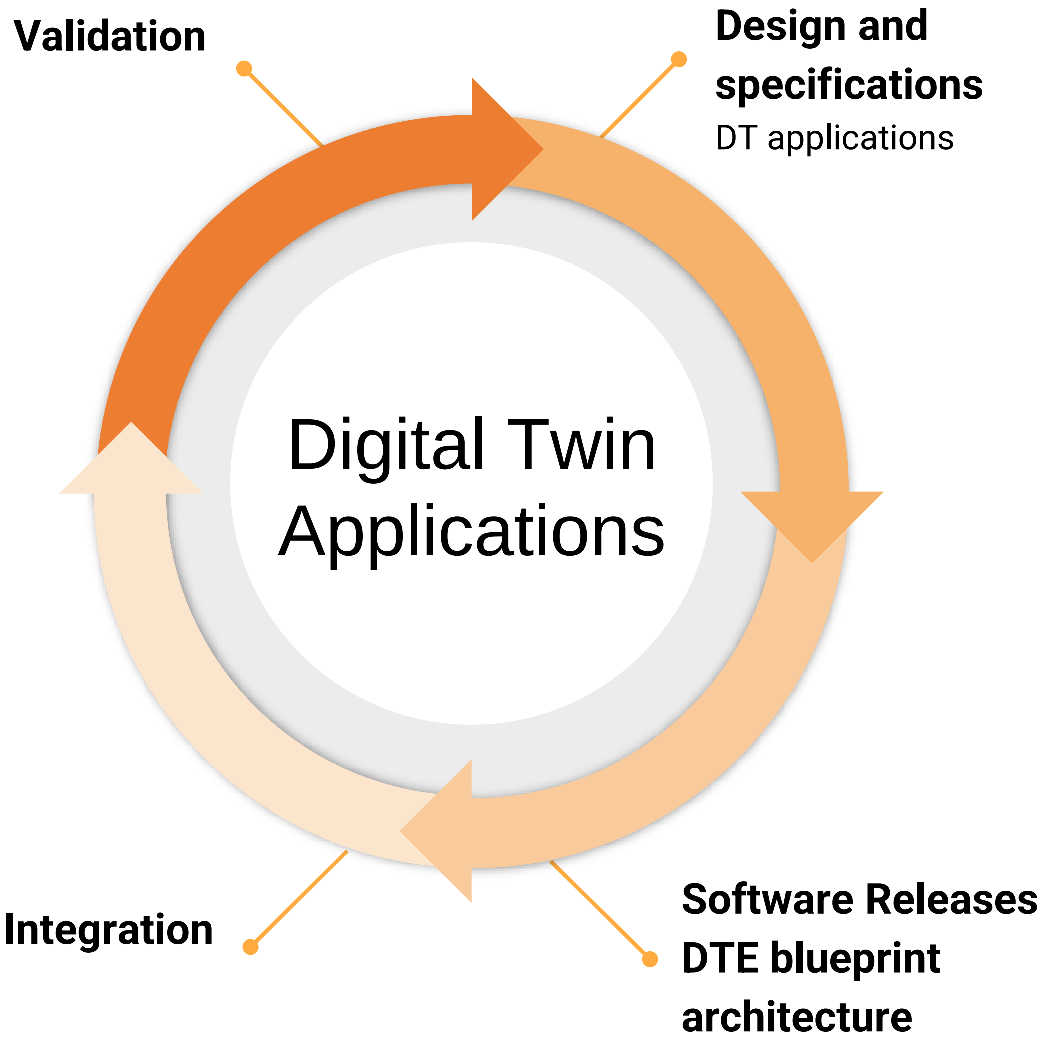
Interdisciplinary Digital Twins

InterTwin (Sept. 2022 - Sept. 2025)

- Develop a common approach to the implementation of Digital Twins (digital twin engine - DTE)
- Applicable across the whole spectrum of scientific disciplines
- Open-source and interoperable platform
- Software components for modelling and simulation to integrate application-specific DTs
- Blueprint architecture for DTs
- Liaison with Destination Earth

Contribution from IFCA - CSIC

- Software release and management
- Quality and validation for applications, models and services
- Lattice QCD



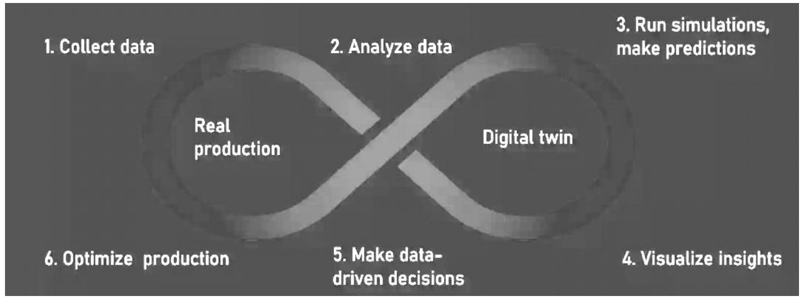
Elements of a Digital Twin

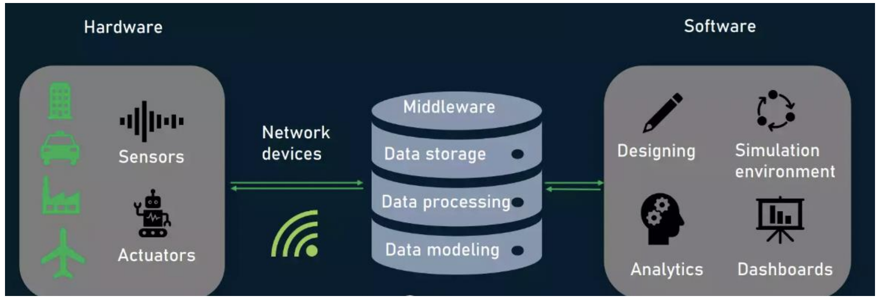
Hardware components. Measuring devices, sensors, etc… that initiate the exchange of information between reality and their software representation.

Data management middleware.A repository to accumulate data from different sources. Ideally, also takes care of data integration, processing, data quality control, etc… by interacting with software components

Software components.

- The analytics engine that turns raw observations into predictions by the simulation software
- In many cases, powered by machine learning models.
- Other: dashboards for real-time monitoring, design tools for modeling etc…





Digital Twins versus Simulations

How does one formulate a new Physics law?

1. First, we guess it
2. Second, we compute the consequences of that guess in measurable terms
3. Third, we compare with nature: i.e. compare with experimental observations

- Our "core business" is advancing computing technologies to support scientific challenges.
- Developing solutions for specific Scientific & Technological challenges notably for steps (2) and (3)
- Some are even trying to apply Artificial Intelligence to (1). The value of knowledge without understanding is however very limited.

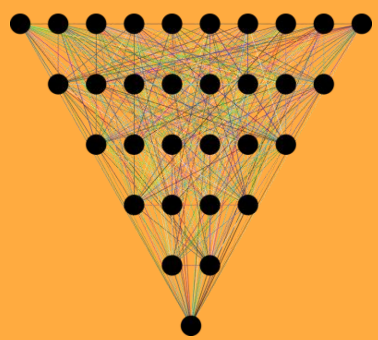
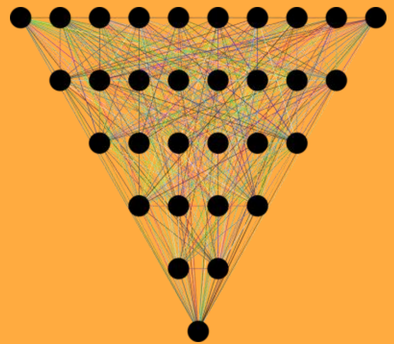
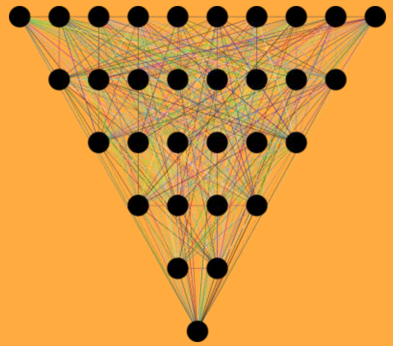
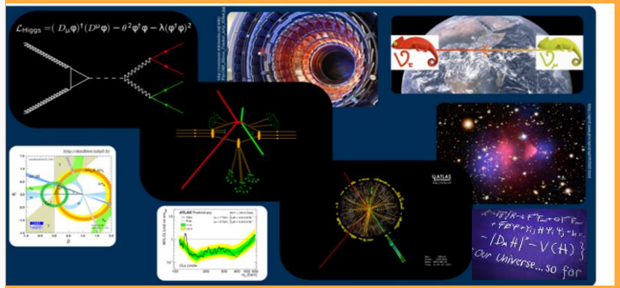
We are trying to exploit the technologies associated with Digital Twins (an industrial concept born in aeronautics) to improve research capabilities in HEP, Observational Astrophysics, Extreme phenomena, Geophysics, Climate predictions, etc…

How are "our simulations" related to the concept of Digital Twins

Digital Twins in HEP and Astro

- Fast Algorithms for Detector Design based on GANs
- Theoretical Computations and Simulations CSIC Consejo Superior de Investigaciones Científicas
- Machine Learning tools to accelerate the simulation with respect to conventional Monte Carlo
- Machine learning to help detecting noise in signals



Elementary Particles

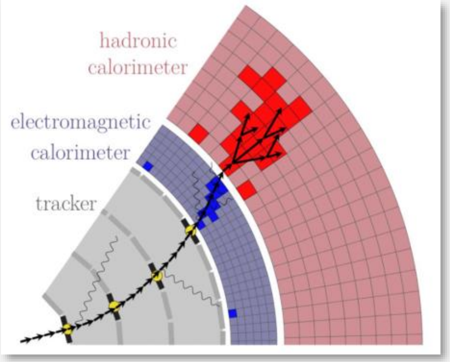
Gravitational Waves

Dark Matter

Digital Twin of a LHC detector

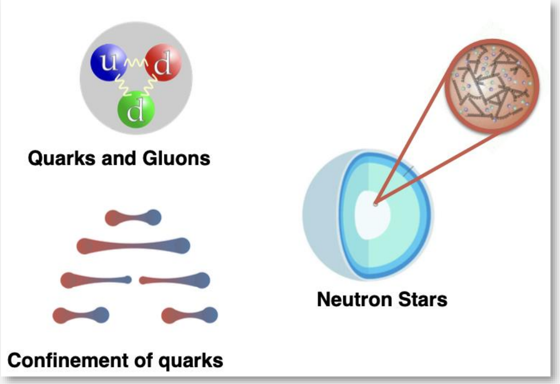
DT in High Energy Physics

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



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***.

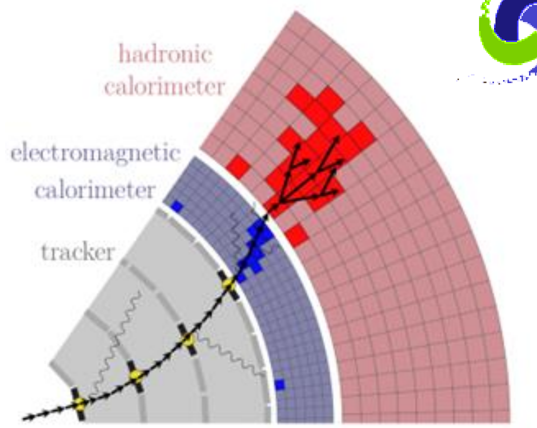
High Energy Physics: DT of a HEP detector

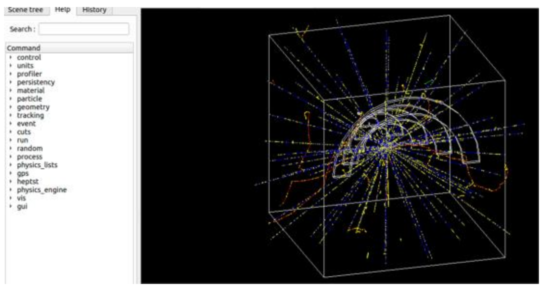
Detector Simulation Software = Software that simulates the physics processes happening at the detector level
GEANT4 is the golden standard software for detector simulations:

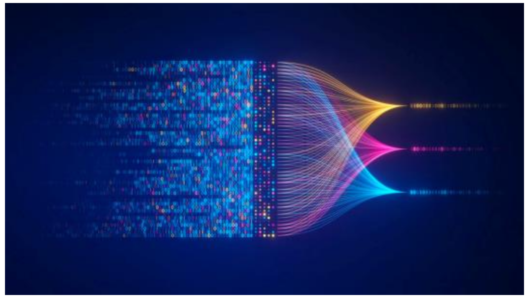
→ Passage of particles through matter, in particular through the materials a particle detector is made of. Expensive to run (40% of the time spent in all the MC simulations needed)

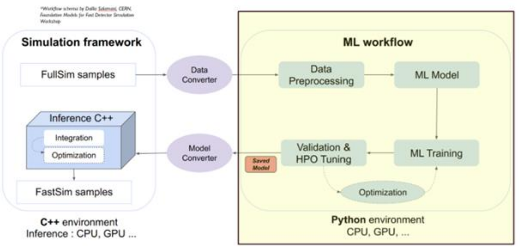
→ **Machine learning (ML) provides a complementary approach: by learning a map between incident particles and detector response using GEANT4 as reference result.**

- **Using ML to accelerate detector simulation is a popular option currently under investigation**







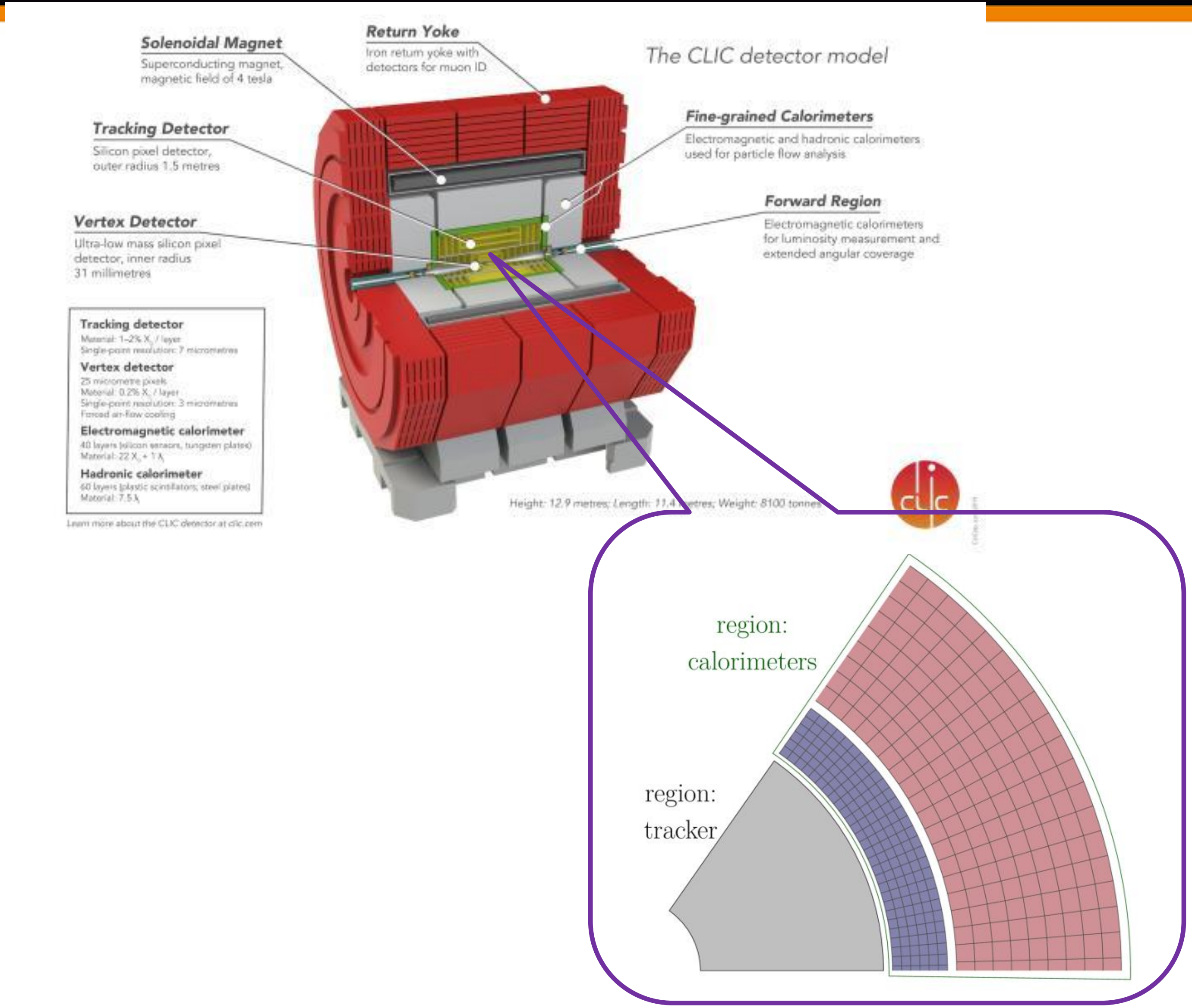


DT Particle Detector application - Scope

- **Detector Prototyping & Optimization**
  - Build **data-driven tool that simulates detector response** and integrates operation conditions from experimental setups (test-beams)

- **Online ML for Detectors**
  - adapt **real-time** detector and/or data acquisition configuration with respect to run conditions

- **Quality verification & Validation frameworks**
  - model **convergence and accuracy** of the generated data should be monitored
  - development of **sample-based validation framework** in collaboration with HEP community



Technical requirements based on capabilities

- **Generate input data for training**
  - Run Monte Carlo simulations locally using GEANT4 software → Output: ROOT files (https://root.cern.ch/)
- **Pre-process ROOT files before feeding the data into the ML model**
  - Input: ROOT files → Output: HDF5 files (decreased volume)
- **Store input and output data**
  - Object or local storage
- **Distributed training with multiple GPUs**
  - GANs, to be tested: Transformer-based models

- **Model inference**
- **Validation/Quality Check**
  - Comparing generated data with Monte Carlo data
  - Sample-based metrics
- **Continuous re-training**
  - Current state of use case (as described in proposal) is a static synthetic model of a detector.
  - Exploring of extending to an application capable of modelling in real time the detector's output in different operation conditions (beams and accelerator configurations) → continuous re-training on real data

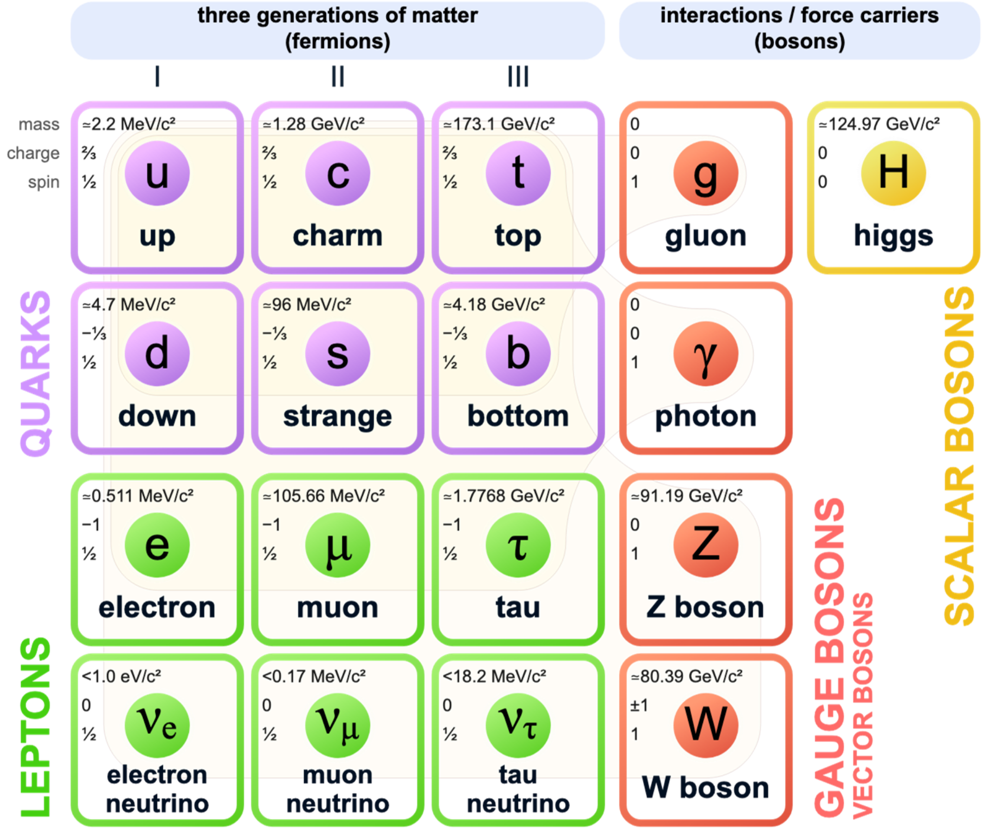
Theoretical Simulations: Lattice QCD

CSIC
Consejo Superior de Investigaciones Científicas

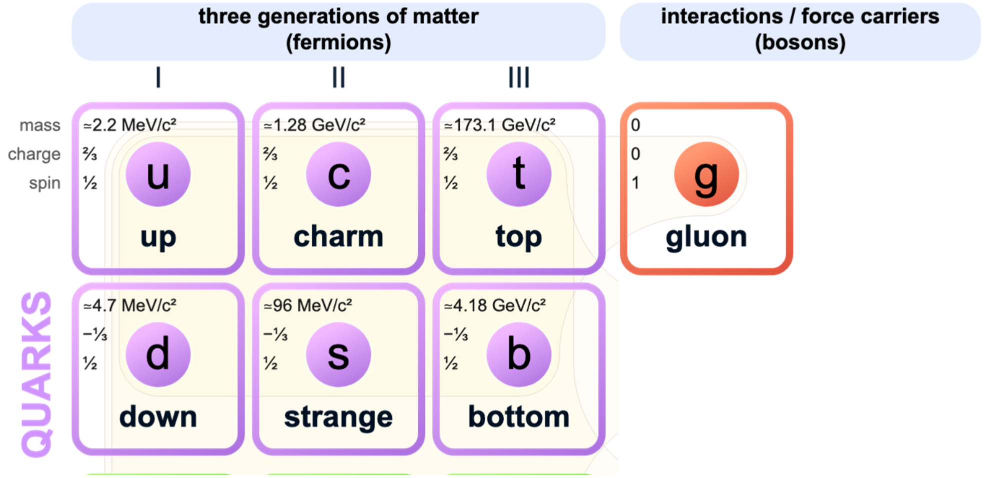
ETH zürich

High Energy Physics: the quest for reality

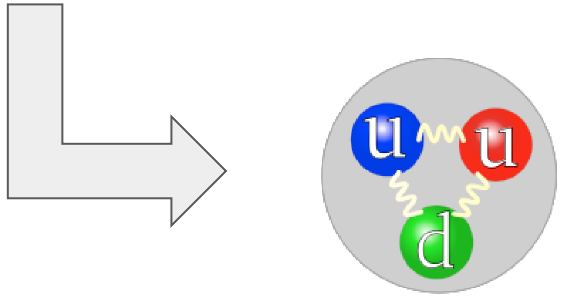
Standard Model



Quantum Chromo Dynamics part



Strong force



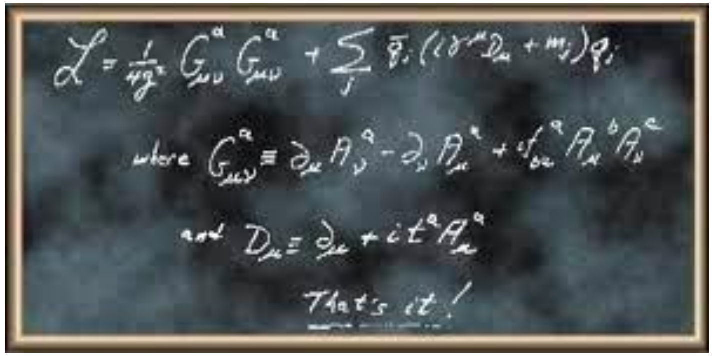
High Energy Physics: Lattice QCD
“First, we guess it”

“Reality of interest”: Quantum Chromo Dynamics

QCD is believed to explain why quarks are confined inside hadrons →

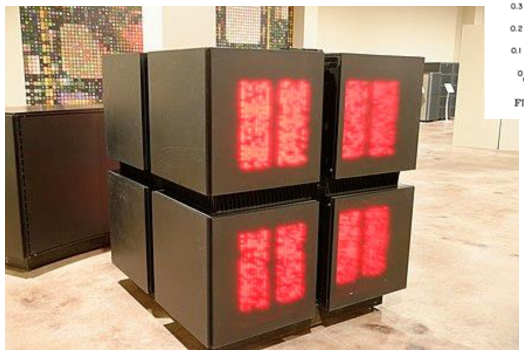
Real data: from particle accelerators experiments

Matching with simulations is the subject of an entire subarea of HEP : Phenomenology.



Is Lattice QCD actually a Digital Twin of the underlying particle physics real world ?
The question is impossible,, because we do not really know how the real world looks like

High Energy Physics: LQCD > 30 years of development
“Second, we compute the consequences of that guess”



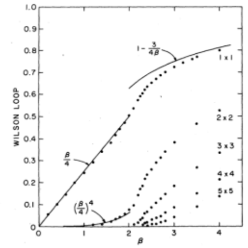


FIG. 4. Wilson loops as a function of β.

2022 - LUMI EuroHPC

← From the toy model, to

1985 - Connection Machine

probing the Standard Model →

g-2 = 0.00233184110
+/- 0.00000000043 (stat.)
+/- 0.00000000019 (syst.)

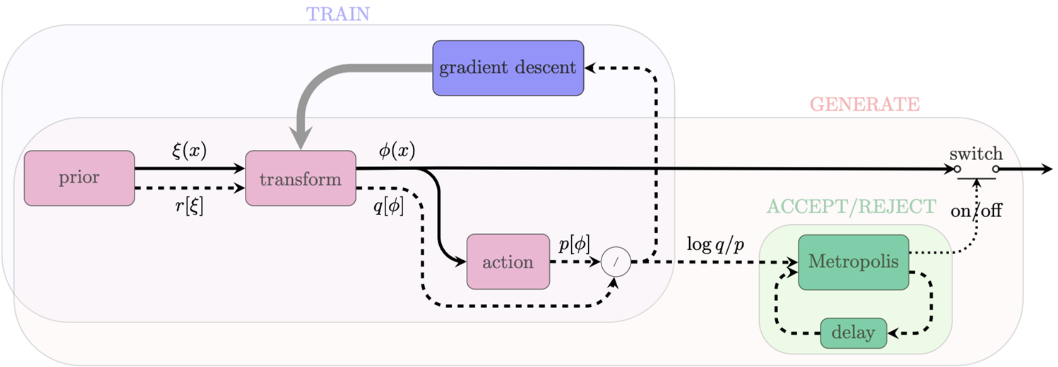
*“Feynman had to write a computer program to test the machine. Since he only knew BASIC, invented a parallel version of BASIC, and programmed Lattice QCD. He was excited by the results. **"Hey Danny, you're not going to believe this, but that machine of yours can actually do something useful !** “* *(Daniel Hillis, in “Richard Feynman and the Connection Machine”)*

Machine Learning in Lattice QCD

- Speeding up the generation of QCD configurations in complex areas of the parameter space.

- The Machine Learning process learns on the results of the full Monte Carlo simulation, comparing the results of reference observables (eg. energy) with those obtained using reference codes (eg. OpenQCD).

- Designing better architectures for Machine Learning models so that the acceptance rates become reasonable (~50% or more) as the volume of the lattice increases.



We gather data to measure (g-2):

A. In experimental facilities: the answer of nature. Several experiments ongoing
B. In the HPC systems by simulating the Standard Model (the QCD part)

if (A) is compatible with (B) the Standard Model stands correct


BBC Sign in Home News Sport Reel Worklife

NEWS

Home | War in Ukraine | Climate | Video | World | UK | Business | Tech | Science | Entertainment & Arts

Science

Scientists at Fermilab close in on fifth force of nature

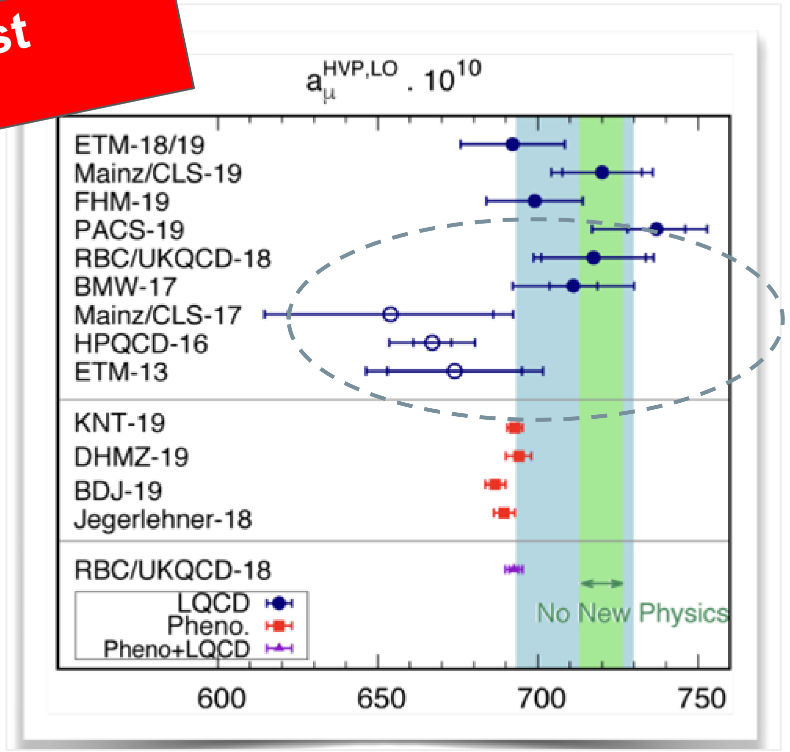
10 August

REIDAR HAHN / FERMILAB

August 2023

Interesting race !

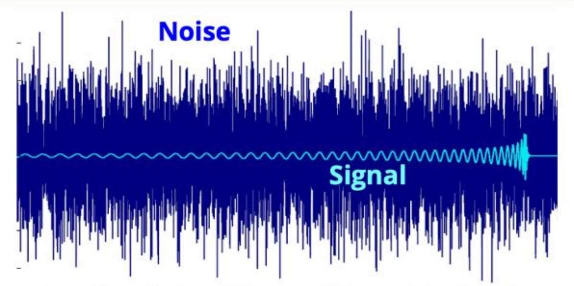
August 2023



Digital Twin in Radioastronomy and Gravitational Wave detectors

DTs of Radio astronomy and GW astrophysics

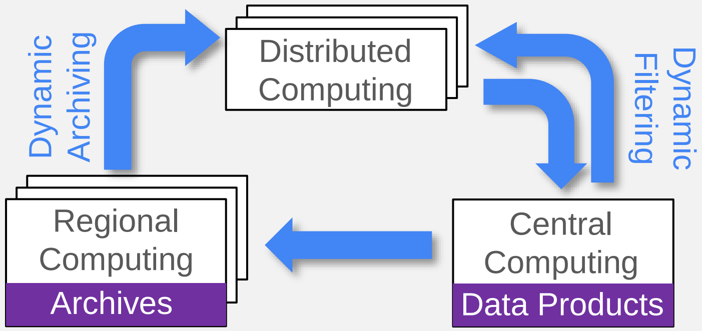
DT of the VIRGO Interferometer



Simulated signal from a binary black hole merger that has been embedded in noise.

It is 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.

DT for noise simulation of next-generation radio telescopes



Meant to provide 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 in real-time).

VIRGO experiment

Measuring the disturbance of space-time following a galactic "cataclysm" (eg. black holes colliding, neutron stars collisions, etc…)

Digital Twin of the Interferometer

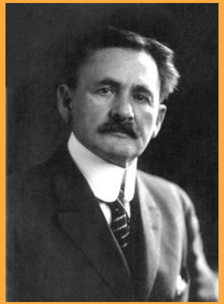
→ To realistically simulate the noise in the detector to study how it reacts to external disturbances

→ To be able to detect noise in quasi-real time to inform the online pipelines, which is currently not possible (--> to feed multimessenger astronomy)

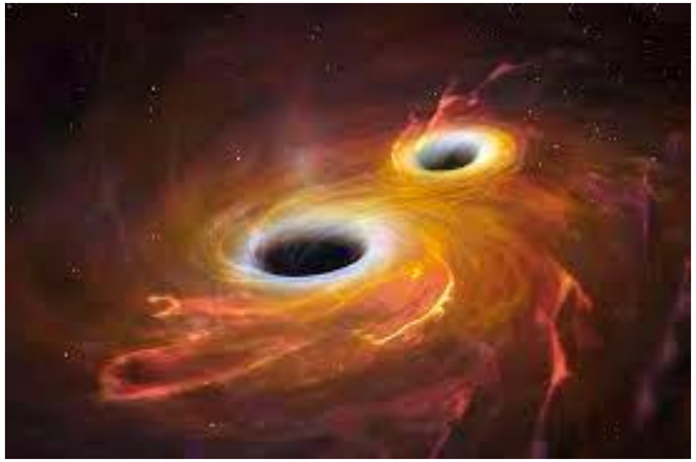
Following true cosmic cataclysm
eg. collision of two black holes

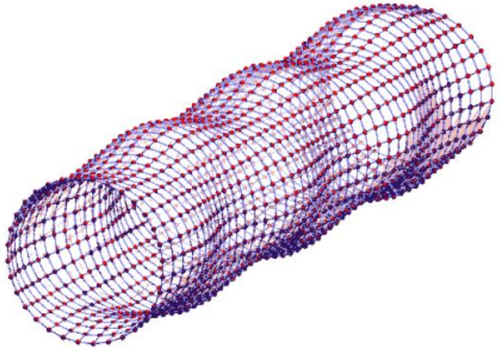
Gravitational Waves are generated:
extremely non-intuitive but true →

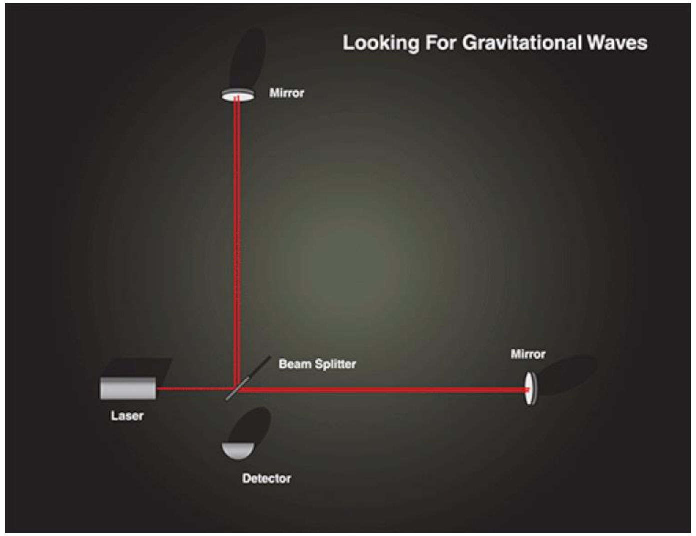
The detection of the distortion of space-time is a challenging experiment here on Earth



1907 Nobel Prize,
Albert Michelson





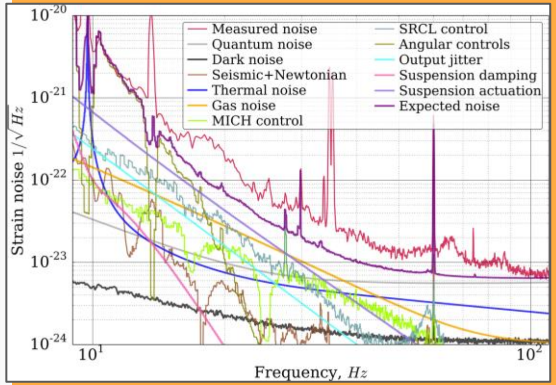


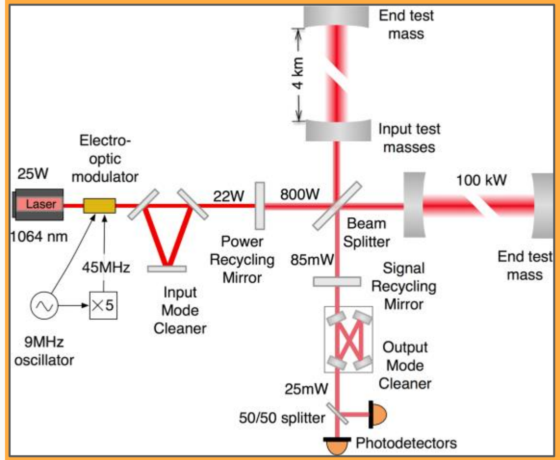
Digital Twin of a 3km long-arms Michelson interferometer

The experiment needs to measure a very faint signal

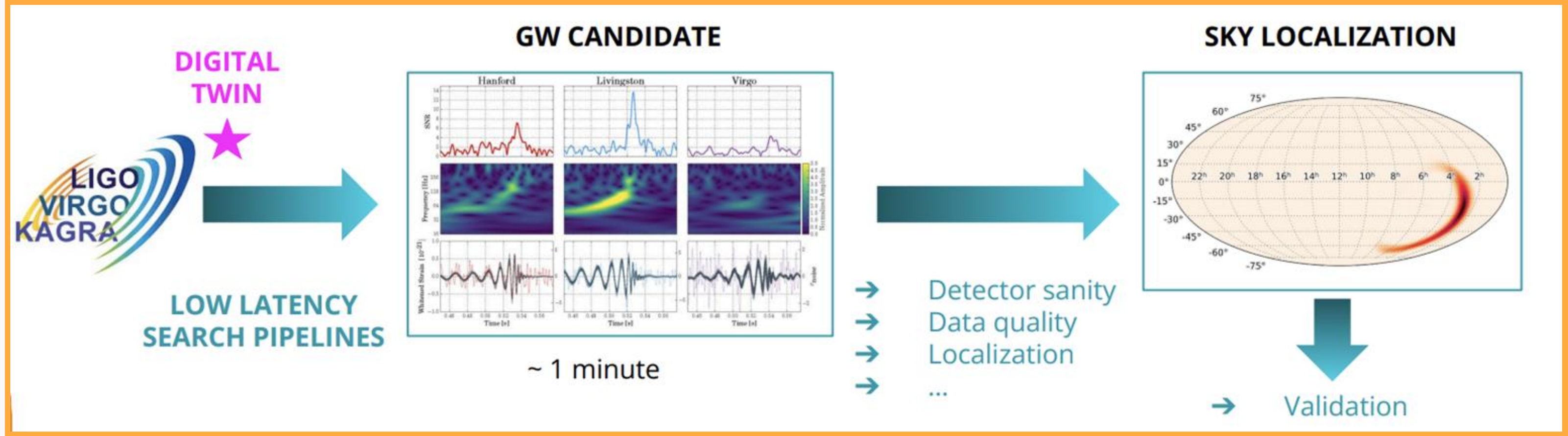
Removing the noise in the experiment is key:

A proper simulation of that noise is crucial to increase the sensitivity of the detectors



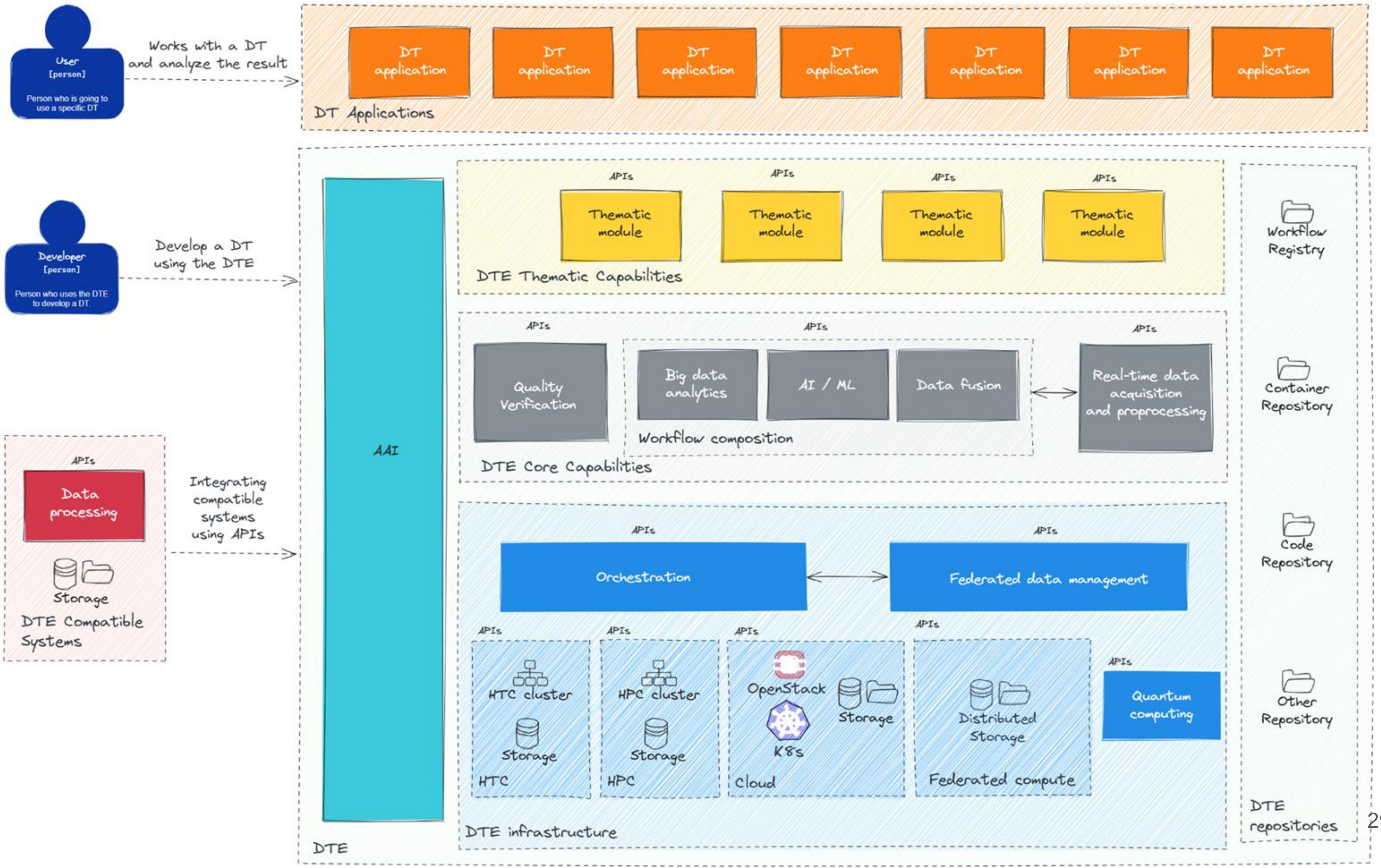


- **Seismic noise**, movements earth vibrations
- **Thermal noise**, from the microscopic fluctuations of the individual atoms in the mirrors and their suspensions.
- **Quantum noise**, due to the discrete nature of light
- **Gas noise**, from the interactions of the residual gas particles in the vacuum enclosure with the mirrors and the laser light.
- **Laser noises**, small variations in the laser intensity and frequency.
- **Beam jitter**, or slight variations in the position and angle of the laser beam in the detector,
- **Scattered light**, generated by tiny imperfections in the mirrors of the interferometers
- And finally, **electronics noise**



The elements of the Digital Twin

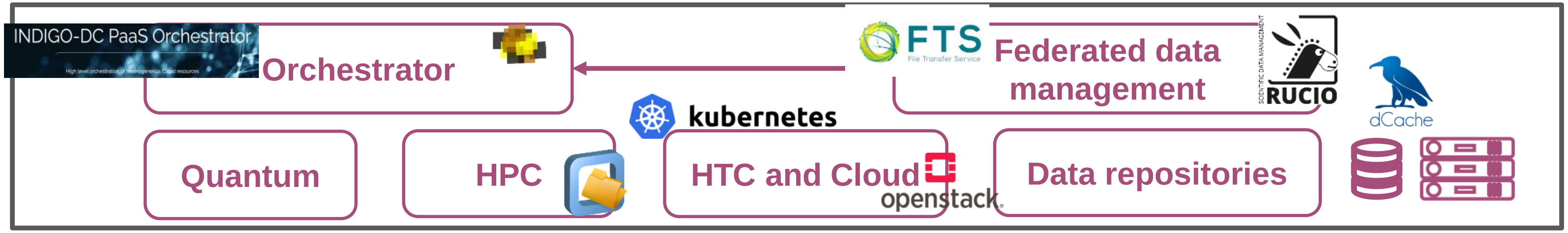
interTwin components



DTE components - Infrastructure side

Orchestrator	**PaaS Orchestrator** + **Infrastructure Manager** elaborating deployment requested expressed in **TOSCA** to be extended to deal with HPC and AI based orchestration
Federated data management	Based on **ESCAPE Data Lake** architecture and services, **Rucio**, **FTS** and HTTP accessed caches/storages. Data lake concept extended to HPC facilities
Federated Compute	Use of single-sign-on in complex simulation and modelling tasks to access data and different compute facilities, including offloading to HPC. Automated modelling and simulation fused with data repositories and computation with containers on HTC, Cloud and HPC

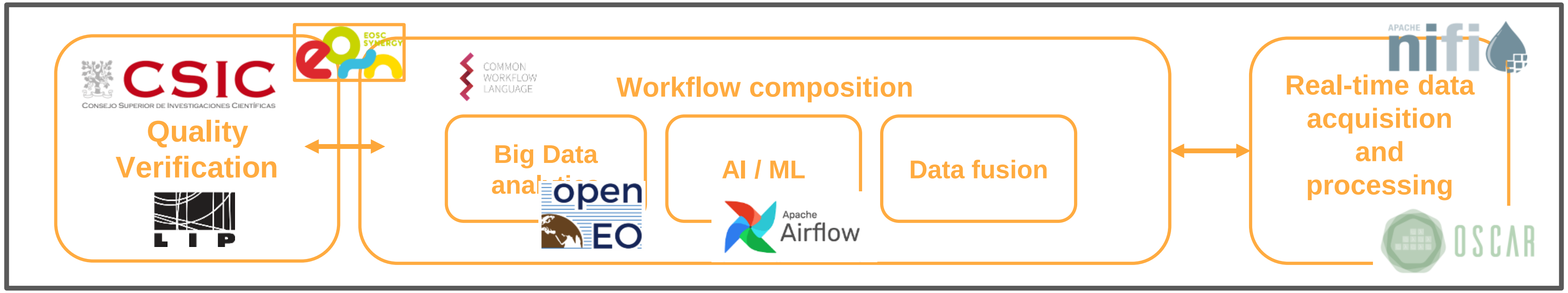
DTE Infrastructure



DTE components - Core

Workflow composition	Workflow definitions based on **CWL,** run on a workflow execution system (e.g. **StreamFlow**, **AirFlow**) and able to combine self-contained execution steps from other workflow engines (e.g . **ecFlow, Ophidia, Delft-FEWS**), different back-ends, distributed big data analysis tools (e.g. **openEO, Dask, Spark**) and ML/DL training platforms (e.g., **Horovod**, **HeAT**, **PyTorch DDP**). Data Fusion as one of the workflow steps to merge observational and modeled data and different data sources.
Real-time data acquisition and processing	Generic framework for real time acquisition and processing that builds on event-triggered execution of workflow engines and exploit serverless computing. Based on **Apache NIFI** and **OSCAR Framework**
Quality Verification	Specific module for quality assurance (QA) that aims at tackling the early validation of the DTs, before being deployed as a “living DT”. Based on the **SQaaS** developed in **EOSC Synergy** project

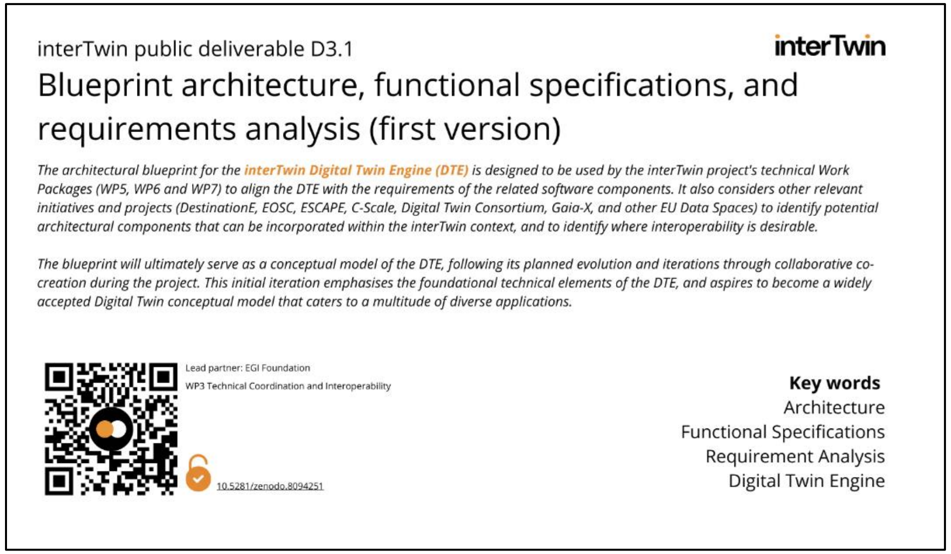
DTE Core Modules



DTE Blueprint and co-design

First version of the Blueprint architecture and design specifications are available in Github and Zenodo. Second and Third versions are planned in 2024

- https://github.com/interTwin-eu/architecture-diagrams/tree/main/Blueprint%20architecture
- https://zenodo.org/communities/intertwin/



interTwin public deliverable D3.1

interTwin

Blueprint architecture, functional specifications, and requirements analysis (first version)

The architectural blueprint for the interTwin Digital Twin Engine (DTE) is designed to be used by the interTwin project's technical Work Packages (WP5, WP6 and WP7) to align the DTE with the requirements of the related software components. It also considers other relevant initiatives and projects (DestinationE, EOSC, ESCAPE, C-Scale, Digital Twin Consortium, Gaia-X, and other EU Data Spaces) to identify potential architectural components that can be incorporated within the interTwin context, and to identify where interoperability is desirable.

The blueprint will ultimately serve as a conceptual model of the DTE, following its planned evolution and iterations through collaborative co-creation during the project. This initial iteration emphasises the foundational technical elements of the DTE, and aspires to become a widely accepted Digital Twin conceptual model that caters to a multitude of diverse applications.

Lead partner: EGI Foundation

WP3 Technical Coordination and Interoperability

10.5281/zenodo.8094251

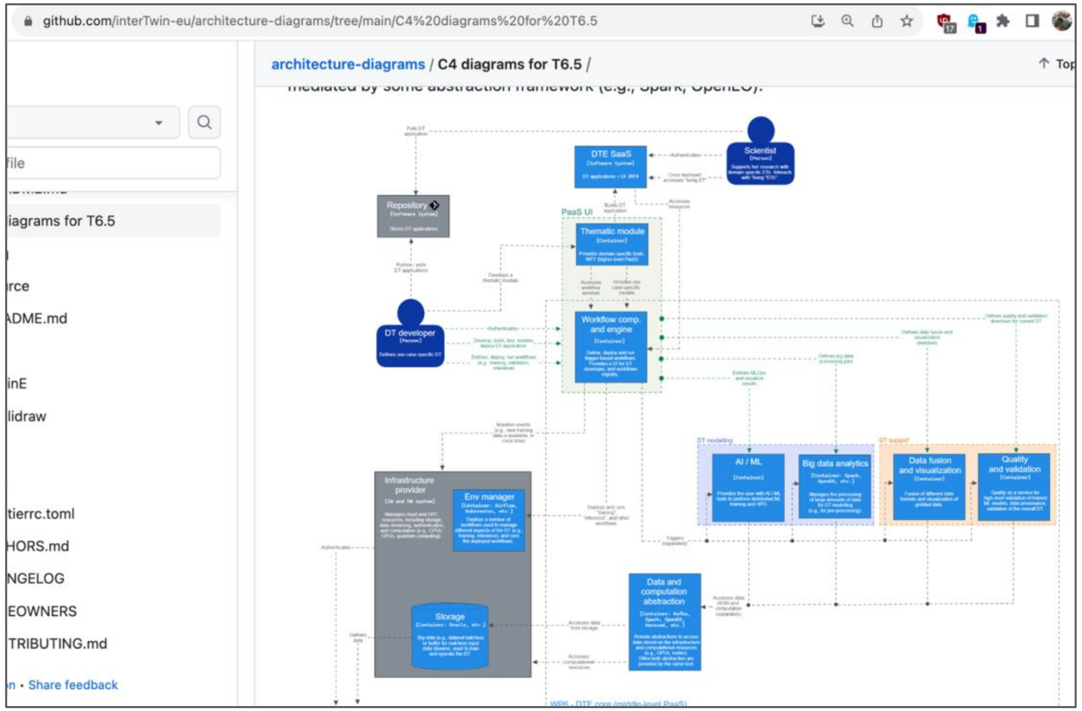
Key words

Architecture

Functional Specifications

Requirement Analysis

Digital Twin Engine



Timeline

Project Year 1

06.2023

DTE Design and specifications

Deliverables: Report on requirements for all use cases

DTE blueprint architecture

Deliverable: DTE blueprint architecture, functional specifications and requirements analysis v1

Project Year 2

11-12.2023

Software Releases

Deliverables: Software releases for all use cases and modules

01.2024

DTE blueprint architecture II

Deliverable: DTE blueprint architecture, functional specifications and requirements analysis v2

04.2024

Validation

Deliverables: DT application development and integration report

09.2024

Design and specifications II

Deliverables: Updated report on requirements for all use cases

Project Year 3

10.2024

DTE blueprint architecture III

Deliverable: DTE blueprint architecture, functional specifications and requirements analysis v3

01-04.2025

Software Releases II

Deliverables: Final Software releases for all use cases and modules

07-08.2025

Validation II

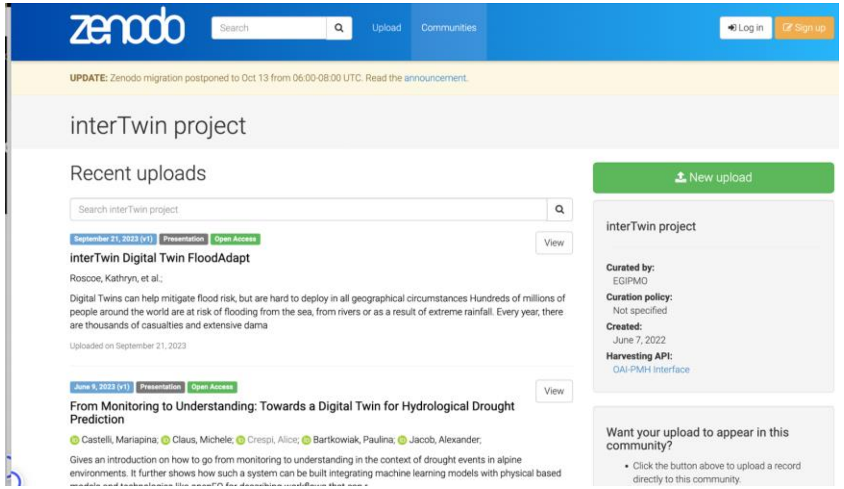
Deliverables:
DT application development and integration report
Report on software architecture concepts based on DestinE and InterTwin
Final Architecture design of the DTs capabilities

2023

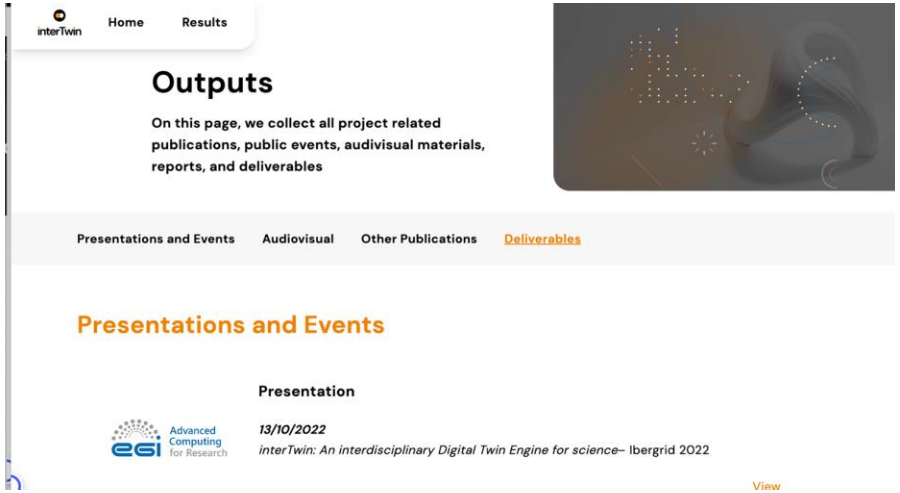
2024

2025

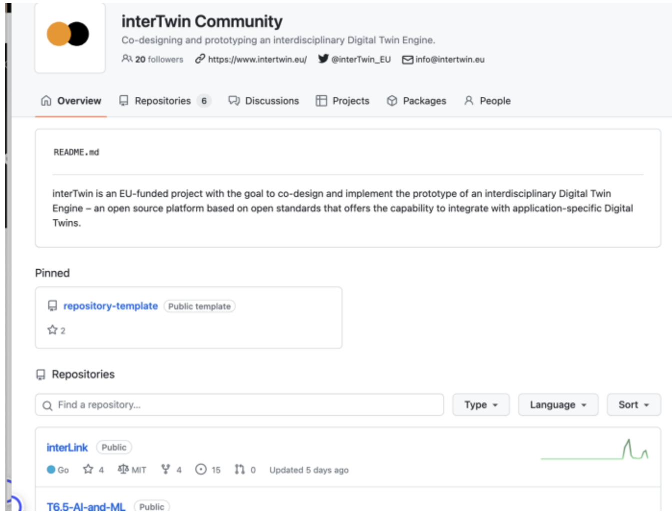
References



Zenodo Community
https://zenodo.org/communities/intertwin/



https://www.intertwin.eu/outputs/



https://github.com/interTwin-eu

First Set of Technical Documents

- interTwin D3.1 DTE blueprint architecture, functional specifications and requirements analysis first version
- interTwin D4.1 First Architecture design of the DTs capabilities for climate change and impact decision support tools
- interTwin D4.2 First Architecture design of the DTs capabilities for High Energy Physics, Radio astronomy and Gravitational-wave Astrophysics
- interTwin D5.1 First Architecture design and Implementation Plan
- interTwin D6.1 Report on requirements and core modules definition
- interTwin D7.1 Report on requirements and thematic modules definition for the environment domain
- interTwin D7.2 Report on requirements and thematic modules definition for the physics domain first version

Thank you!

www.intertwin.eu

info@intertwin.eu

intertwin_eu

intertwin