An interdisciplinary Digital Twin Engine for science

Małgorzata Krakowian, interTwin Project Manager

13th October 2022

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

General Information

Duration	36 months
Start date	1 September 2022
End date	31 August 2025
Budget	11,731,665 EUR
PMs	1481.5

HORIZON-INFRA-2021-TECH-01- 01: Interdisciplinary digital twins
Outcome

- prototype of an interdisciplinary Digital Twin, using a combination of the latest digital technologies, to address complex challenges;
- support interoperability of data and software, integration and collaboration across different scientific domains;
- A framework enabling Researchers to ensure the quality, reliability, verifiability of the data available through the Common European Data Spaces and the European Open Science Cloud

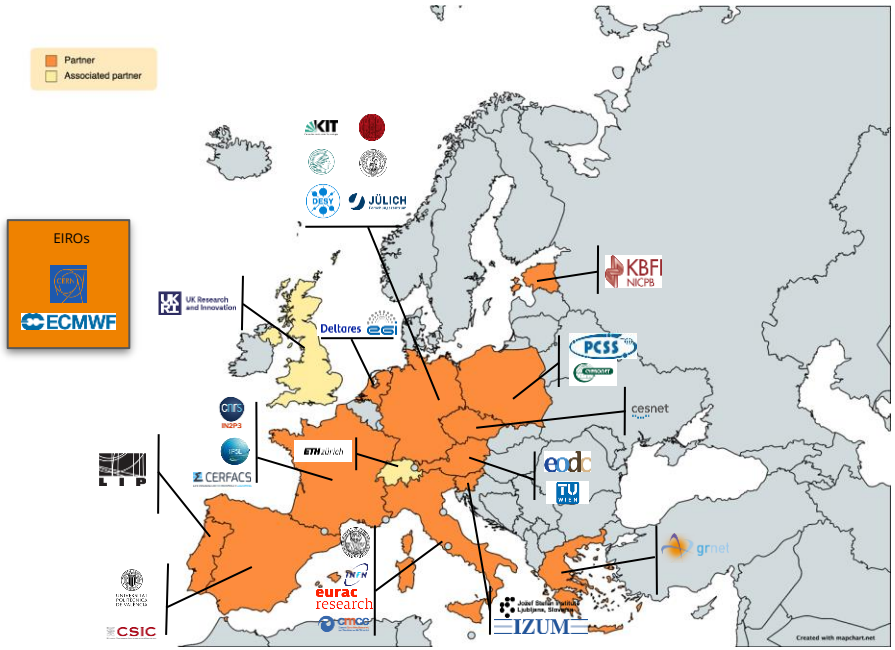
interTwin overall objective

Co-design and **implement** the prototype of an **interdisciplinary Digital Twin Engine** - an open source platform based on open standards that offers the capability to integrate with **application-specific Digital Twins**. Its functional specifications and implementation are based on a **co-designed interoperability framework** and conceptual model of a DT for research **- the DTE blueprint architecture**.

1. Co-design, develop and provide a Digital Twin Engine that simplifies & accelerates the development of complex application-specific DTs that benefits researchers, business and civil society
2. Co-design a Digital Twin Engine blueprint architecture that provides a conceptual framework for the development of DTs supporting interoperability, performance, portability & accuracy.
3. Extend the technical capabilities of the European Open Science Cloud with modelling & simulation tools integrated with its compute platform
4. Ensure trust and reproducibility in science through quality, reliability and verifiability of the outputs of Digital Twins
5. Demonstrate data fusion with complex modelling & prediction technologies
6. Simplify DT application development with tools to manage AI workflows and the model lifecycle while reinforcing open science practices

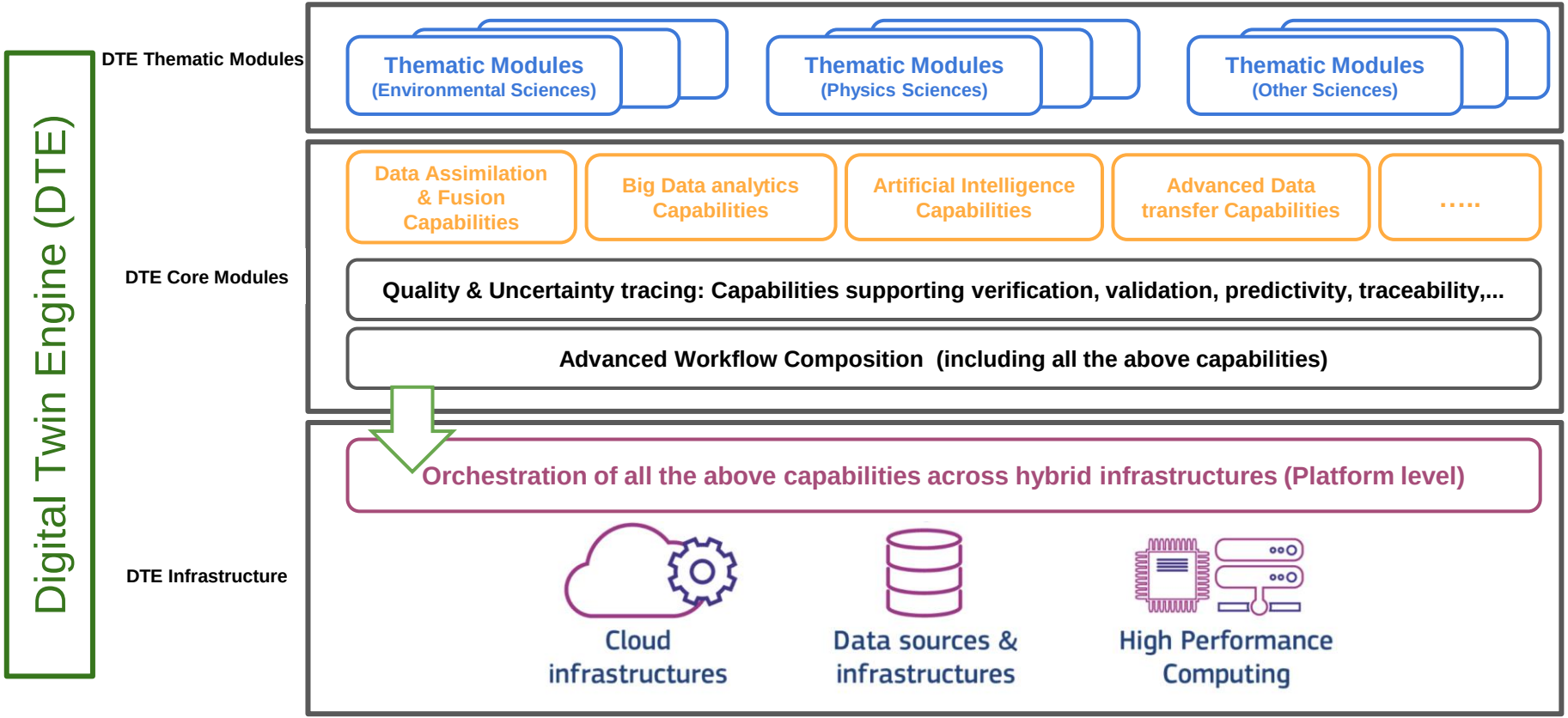
interTwin Specific Objectives

Consortium Overview



- **EGI Foundation as coordinator**
  - **30 participating partners, including**
    - 1 Affiliated entity
    - 2 Associated partners
- **Consortium at a glance**
  - **10 partners to deliver cloud, HTC , HPC resources and access to Quantum systems**
  - **11 open source technology providers delivering the DTE infrastructure and horizontal capabilities**
  - **14 partners representing research communities from 5 scientific areas bringing requirements and developing DT applications and thematic modules.**

Digital Twin Engine - Strawman Concept



Cooperation with other projects

Projects in HORIZON-INFRA-TECH-01

https://biodt.eu/
A Digital Twin prototype to help protect and restore biodiversity

https://dtgeo.eu/
A Digital Twin for GEOphysical Extremes

https://www.ebrain-health.eu/
Actionable Multilevel Health Data

Cooperation with external initiatives

https://eosc.eu/

https://digital-strategy.ec.europa.eu/en/policies/destination-earth

https://www.ai4europe.eu/

https://gaia-x.eu/

https://www.plattform-i40.de/

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 **ECMWF**

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

ECMWF

National and international research and innovation initiatives contributions [1/2]

- interTwin uses and expands the **ESCAPE Data Lake blueprint**, which will form the basis of the federated data management solution, and the **C-SCALE blueprint** for distributed cloud-based access of Copernicus data.

- The interTwin compute platform federates compute capacity from research HTC cloud providers and HPC hosting nodes of the **EGI Federation, PRACE** and **EuroHPC**

- The **openEO** community contributes with open-source software for the execution of **Earth observation workflows** on top of different cloud back-ends, **to be extended to other domains**.

National and international research and innovation initiatives contributions [2/ 2]

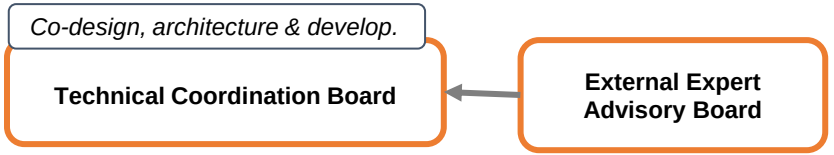
- Software components of the **INDIGO-DataCloud** community are enriched with new functionalities.

- The project leverages on solutions of the European Center of Excellence in Exascale Computing 'Research on AI- and Simulation-Based Engineering at Exascale', (**CoE-RAISE**) developing novel AI methodologies and solutions towards exascale computing along a variety of use cases.

- The DT of the Earth relies on the knowledge and data from weather and climate research networks such as **ENES**, members of the earth observation community active in **GEO** (Group on Earth Observation) and **EARSeL** (European Association of Remote Sensing Laboratories).

Cooperation approach

Proposal:



Between 5 and 7 Members.

Members:

- Key Experts across interTwin value chain, end-user experts across interTwin scientific domains and experts in other scientific/application domains especially those which have already relevant industrial adoption of digital twins
- Idea to start with key representatives of Digital Twin Projects in the same EU-call

Responsibilities:

- **Advises** the Technical Coordination Board on the project technical architecture and implementation plan. It draws membership from Digital Twin innovators and adopters from different sectors.
- **Reviews** blueprint architecture relevant for the specific scientific domains of the project (such as Destination Earth)– from an implementation point of view
- **Validates** and makes sure that what the project is delivering is useful in broader application domains.

Duties:

- Main expected duties include **sharing of project relevant documentation**, the provision of **feedback** and when possible collaborate with project related **dissemination & exploitation**

Timeline

Project Year 1

04.2023 — **Design and specifications**

Deliverables: Report on requirements for all use cases

05.2023 — **DTE blueprint architecture**

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

Project Year 2

10-12.2023 — **Software Releases**

Deliverables: Software releases for all use cases and modules

01.2024 — **DTE blueprint architecture**

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

04.2024 — **Validation**

Deliverables: First version of the thematic module

09.2024 — **Design and specifications**

Deliverables: Updated report on requirements for all use cases

Project Year 3

10.2024 — **DTE blueprint architecture**

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

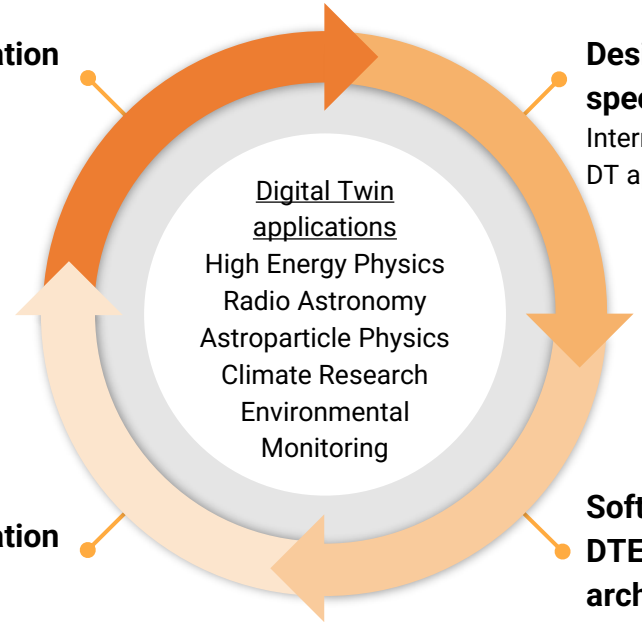
01-04.2025 — **Software Releases**

Deliverables: Final Software releases for all use cases and modules

07-08.2025 — **Validation**

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

DTE Development Cycle

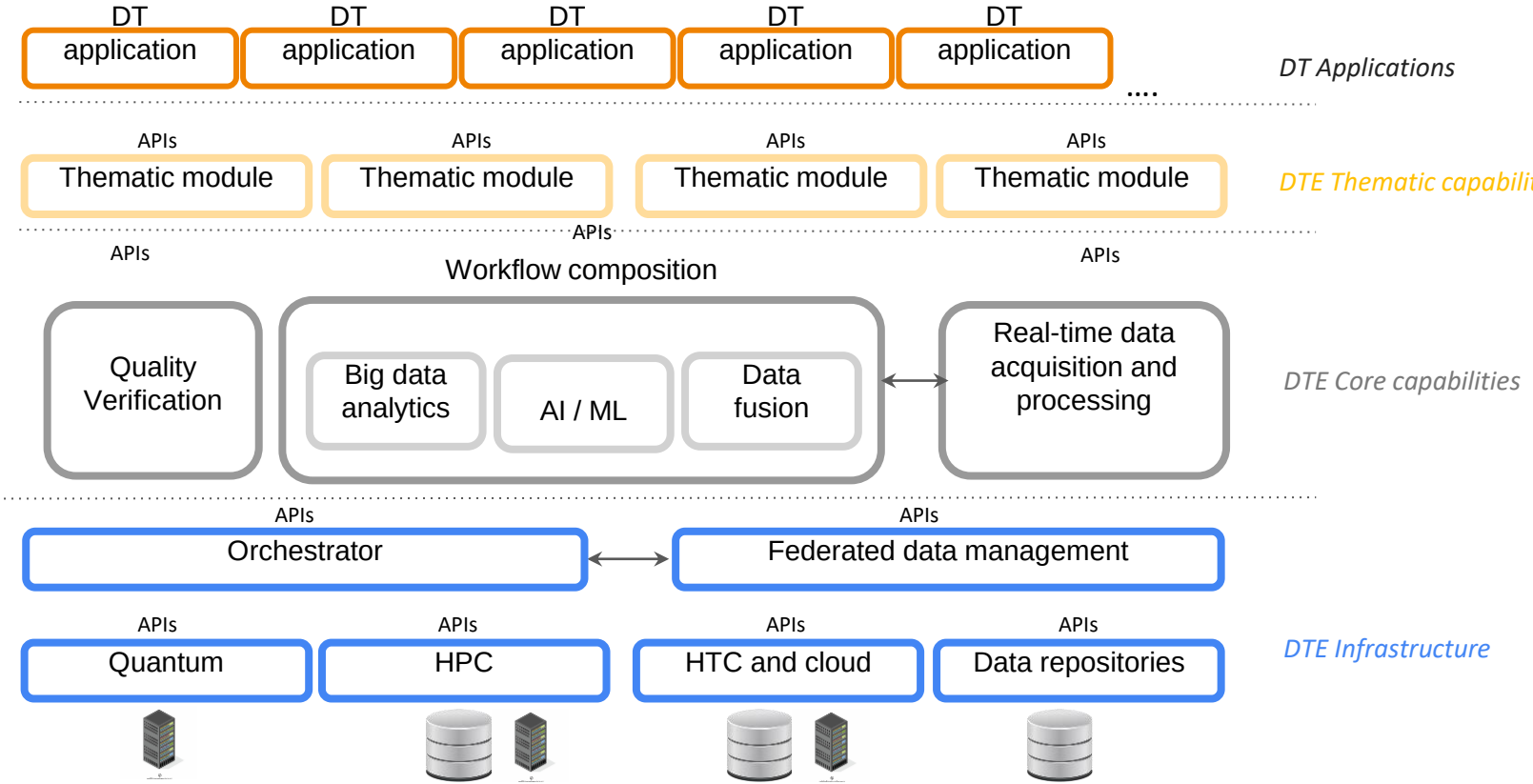


Aim:
Pre-operational software of a DTE at TRL 6 or 7 depending on the components

iBERGRID **contribution:**

CSIC, UPV Software Releases
LIP, CSIC Integration

interTwin components



DT Application: High Energy Physics

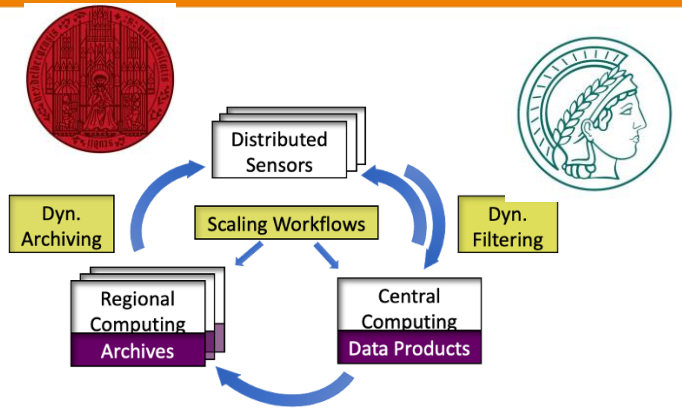
DT of Large Hadron Collider (LHC) detector components: seeking for strategies to face the increase in the need for simulated data expected during the future High Luminosity LHC runs. The primary goal is to provide a fast simulation solution to complement the Monte Carlo approach. ***Faster and deeper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

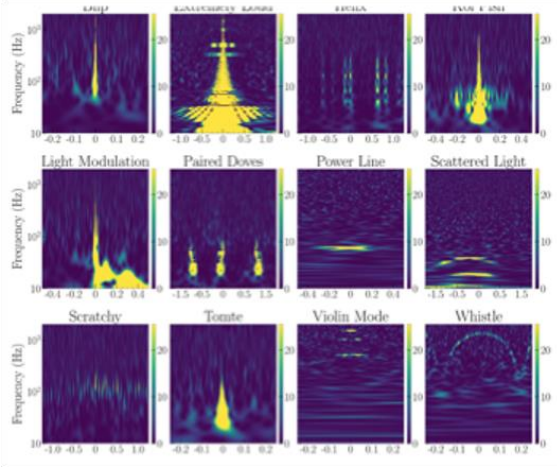
DT of the Standard Model in particle physics: competitive results in Lattice QCD require the ***efficient handling of Petabytes of data***, therefore the implementation of advanced data management tools is mandatory. On the side of algorithmic advancement, ML algorithms have recently started to be applied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

IBERGRID contribution: CSIC

DT Application: 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 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:

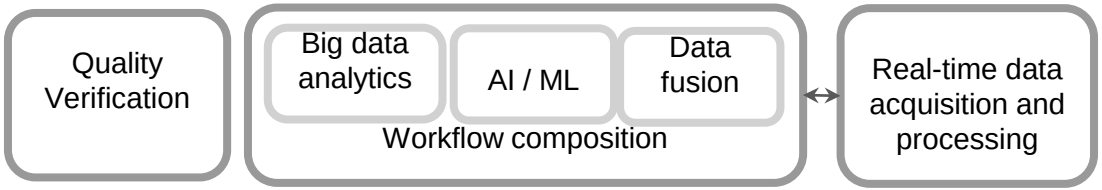
- Climate change, long-term predictions of extreme natural events (storms & fires)
- Early warning for extreme events (floods & droughts)
- Climate change impacts of extreme events (storms, fires, floods & droughts)

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** IBERGRID contribution: CSIC
- **Noise simulation for radio astronomy**
- **GAN-based 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**

DTE Core Capabilities [1 / 2]



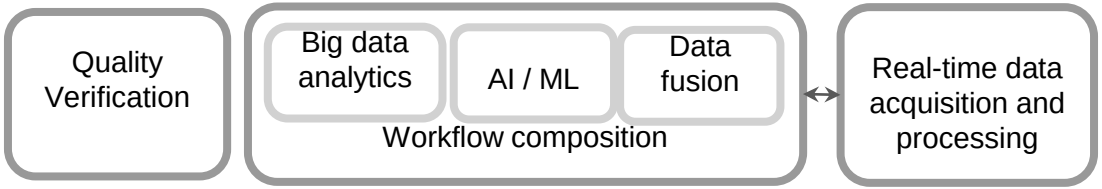
The DTE Core Modules offer horizontal capabilities to facilitate the creation and the operations of data-intensive and compute-intensive DT applications:

- **Advanced workflow composition**: executes DT workflows that can invoke other Core module capabilities. An **interdisciplinary processing graph** will serve as a link and API for the supported workflow engines guaranteeing a **common user experience** and facilitating the integration of discipline specific tools.

- **Data Fusion:** implements and integrates processes for merging datasets from different sources. This includes **linking 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.

IBERGRID contribution: CSIC

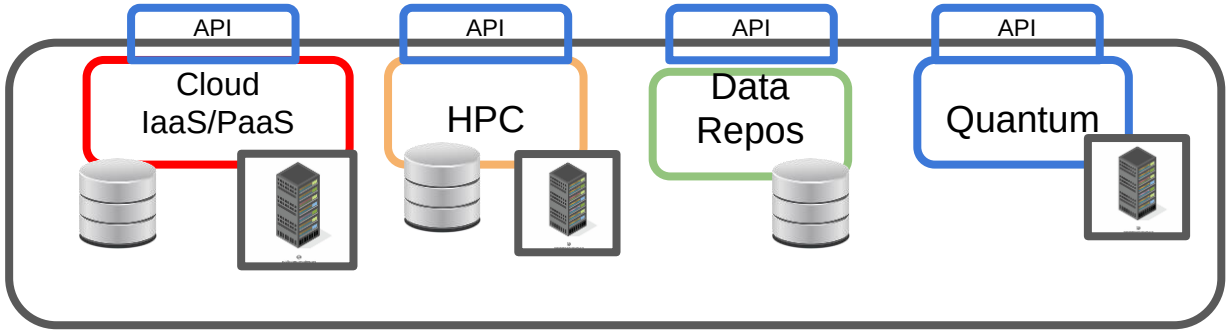
DTE Core Capabilities [2 / 2]



- **AI workflow and method lifecycle management**: toolbox for realizing complex AI setup
- **Real time acquisition and data analytics**: delivers high performance data ingestion by applying the paradigm of "**serverless**" computing to DTs. This module expands the advanced workflow composition module to trigger on the fly processing upon data acquisition.
- **Validation, verification, and uncertainty tracing for model quality**: toolkit that provides developers with the possibility to **design and validate DT Models guaranteeing quality and reliability of the DT applications outputs**. Offered as a "Model Validation as a Service" enabling customisations of best practices and standard quality measures for scientific disciplines and applications.

DTE Infrastructure: Computing, Storage, Federation services and policies

- Solutions to provision access to 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
- Development of homogeneous security and access policies and resource accounting



DTE Infrastructure: AI Orchestrator and Federated Data Management

- 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 (e.g. choose data source and compute sites, how to make data available to compute, etc)
- 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
- The data management actions are implemented by a **Federated Data Management** layer interoperable with various repository types and including data transfer, caching tools and cataloguing systems. Through this layer the DTE can manage for instance HPC data ingress/egress

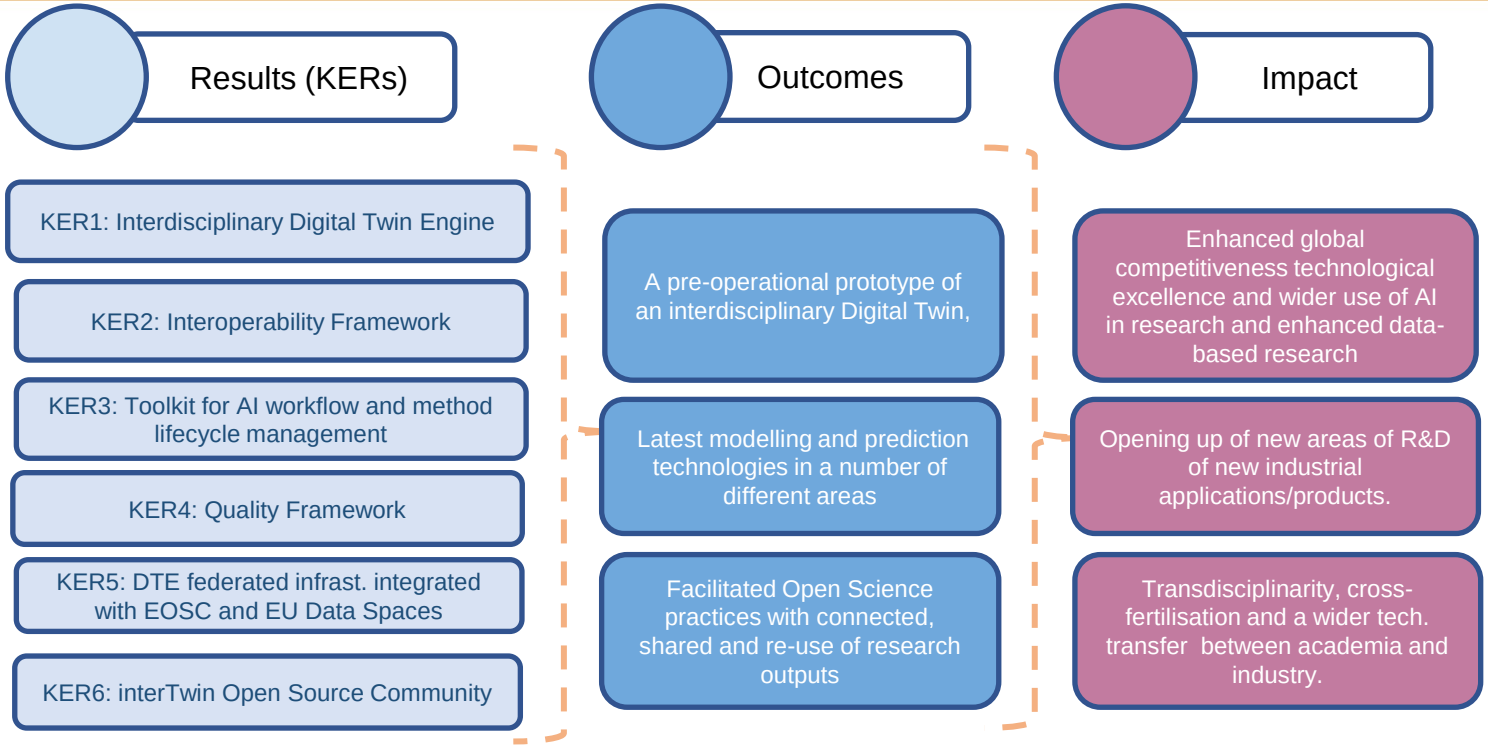
Infrastructure

HPC (6)
TU Wien
GRNET
PSNC
UKRI
JSI (EURO HPC)
JUELICH

Cloud (6+)
TU Wien
EODC
GRNET
PSNC
UKRI
JUELICH
EGI Federation

HTC (2+)
UKRI
KBFI
EGI Federation

KERs and Impact



Key Exploitable Results [1/2]

- **KER1: Interdisciplinary Digital Twin Engine**
  - **Description:** A software platform that provides generic and tailored functional modules for modelling and simulation to facilitate the development and deployment of Digital Twins that address scientific problems in different domains.
- **KER2: Interoperability Framework**
  - **Description**: The interTwin interoperability framework aligns technical approaches and foster collaboration in modelling and simulation application development across scientific domains.
- **KER3: Toolkit for AI workflow and method lifecycle management**
  - **Description**: AI-based methodologies to extract application sector specific information from research data at the exabyte-scale level in a real-time manner and increase the efficiency and accuracy of simulation and modelling outputs.

Key Exploitable Results [2/2]

- **KER4: Quality Framework**
  - **Description:** Tools for automated quality measures and trust, development of standard quality mapping and indicators for appropriately communicating differences in qualities of inputs and outputs from digital twins, addressing issues such as data and model pedigree, accuracy, and lack of knowledge.
- **KER5: DTE federated infrast. integrated with EOSC and EU Data Spaces**
  - **Description**: Federated distributed compute platform providing access to distributed data and integrating HTC, HPC, Cloud and Quantum Computing capabilities for processing.
- **KER6: interTwin Open Source Community**
  - **Description**: The community of DT application developers, users and operators that is responsible for the design, development and maintenance of the DTE code base.

Thank you!

https://www.intertwin.eu
info@intertwin.eu