A digital twin for geophysical extremes: interim results from the DT-GEO project

Arnau Folch (1), Ramon Carbonell (1), Antonio Costa (2), Beata Orlecka-Sikora (3), Piero Lanucara (4), Rosa M.Badia (5), Finn Lovholt (6), Jorge Macías (7), Sasha Brune (8), Alice-Agnes Gabriel (9), Sara Barsotti (10), Jorn Behrens (11), Jorge Gomes (12), Jean Schmittbuhl (13), Carmela Freda (14), Joana Kocot (15), Domenico Giardini (16), Michael Afanasiev (17), Helen Glaves (18)

(1) Geociencias Barcelona (GEO3BCN-CSIC), Spain
(2) Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy
(3) Institute Geophysics - Polish Academy of Science, Poland
(4) CINECA Consorzio Interuniversitario, Italy
(5) Barcelona Supercomputing Center (BSC-CNS), Spain
(6) Stiftelsen Norges Geotekniske Institutt (NGI), Norway
(7)Universidad de Málaga (UMA), Spain
(8)German Research Centre for Geosciences (GFZ), Germany
(9) Ludwing-Maximilians Universitaet (LMU), Germany
(10) Icelandic Meteorological Office (IMO), Iceland
(11) Universität Hamburg, Germany
(12) Laboratorio Fisica Experimental de Partículas (LIP), Portugal
(13) Centre National de la Recherche Scientifique (CNRS), France
(14) European Plate Observing System (EPOS), Italy
(15) Akademia Gorniczo-Hutnicza (CYFRONET), Poland
(16) Federal Institute of Technology (ETH), Switzerland
(17) Mondaic AG, Switzerland
(18) U.K. Research and Innovation (UKRI), U.K.

DT-GEO

Project funded by Horizon Europe under the grant agreement No 101058129.

Presentation outline

01 DT-GEO in a nutshell

02 Development of DT-GEO components: interim results

03 What’s next?

HORIZON-INFRA-2021-TECH-01-01

4 Digital Twins projects funded

1

Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities (BioDT)

Digital Twin providing advanced modelling, simulation and prediction capabilities across relevant research infrastructures, the BioDT project will be able to more accurately model interaction between species and their environment.

2

A Digital Twin for GEOphysical extremes (DT-GEO)

Deploy 12 Digital Twin Components (DTCs) embedding flagship simulation codes, AI layers, large volumes of (real-time) data streams, data assimilation methodologies, and overarching workflows for deployment and execution in centralised HPC and virtual cloud computing RIs. Lead by CSIC.

3

An interdisciplinary Digital Twin Engine for science (interTwin)

Prototype of an interdisciplinary Digital Twin Engine (DTE), an open source platform that provides generic and tailored software components for modelling and simulation to integrate application-specific Digital Twins (DTs). Use cases for high-energy physics, radio astronomy, astrophysics, climate research, and environmental monitoring. Lead by EGI, the Consortium shares 4 partners with DT-GEO (CSIC, CNRS, LIP, UPV)

DestinE

4

eBRAIN-Health - Actionable Multilevel Health Data (eBRAIN-Health)

Deliver a distributed research platform for modelling and simulating complex neurobiological phenomena of human brain function and dysfunction in a data protection compliant environment.

Human brain

A Digital Twin for GEOphysical extremes (DT-GEO)

Action	Horizon-RIA
GA No	101058129
Duration	3 years
Start Date	Sep 2022
End Date	Aug 2025
Budget	15,1 M€
Partners	26
Consortium	HPC RI Data RI Monitoring Research Academia Private

01 Deploy a pre-operational prototype of **Digital Twin (DT) on geophysical extremes** for its future integration in the Destination Earth (DestinE) initiative

02 Implement 12 **Digital Twin Components (DTCs)** addressing specific hazardous phenomena from volcanoes, tsunamis, earthquakes, and anthropogenically-induced extremes in order to conduct **data-informed**:

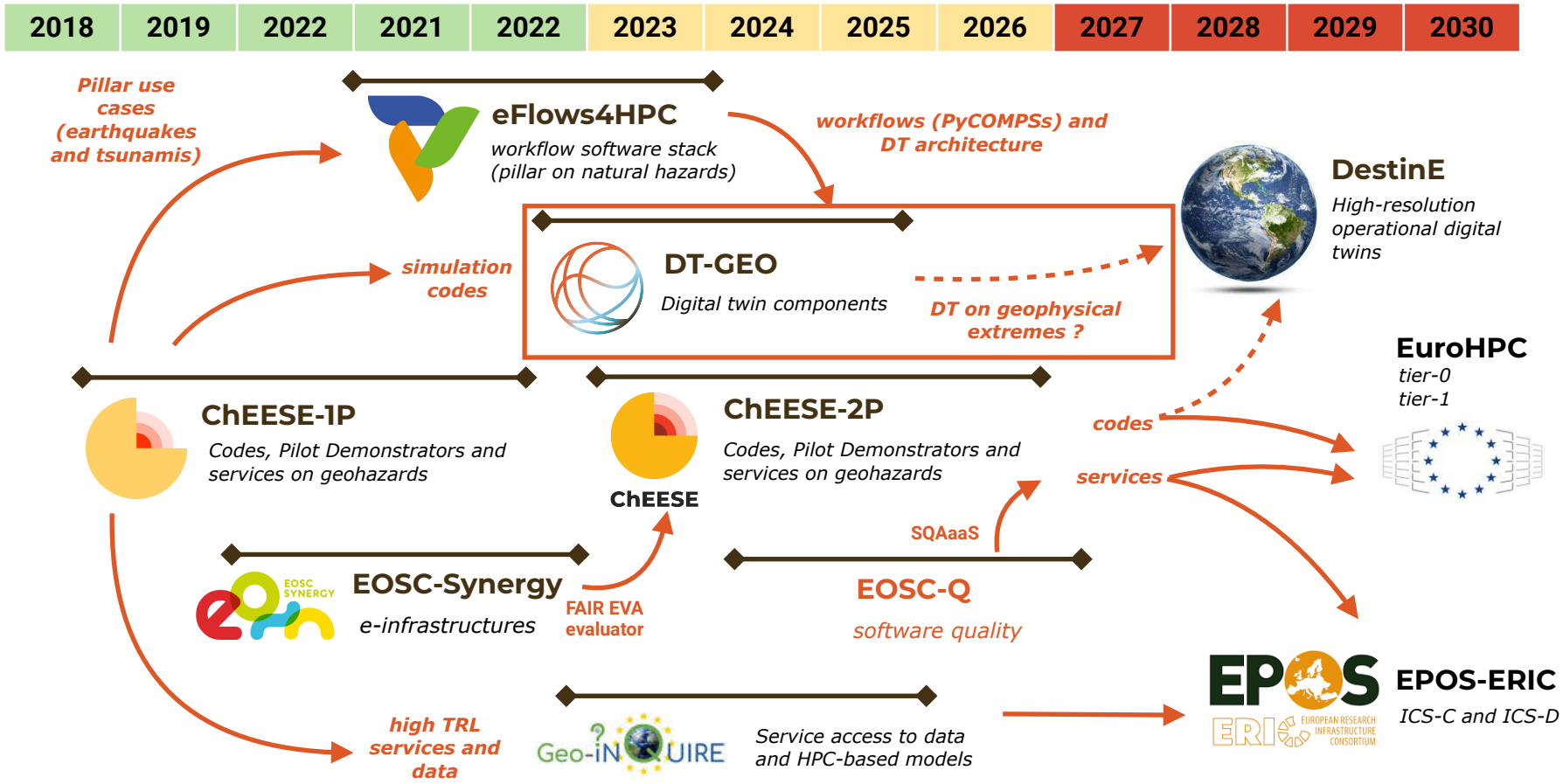
1. Early warning systems
2. Forecasts
3. Hazard assessments across multiple time scales.

03 Provide a flexible framework for FAIR-validation of project digital assets and outcomes and its integration in 2 Research Infrastructures (RIs):

1. The European Plate Observing System (**EPOS**)
2. HPC/virtual cloud computing (**EuroHPC/FENIX**)

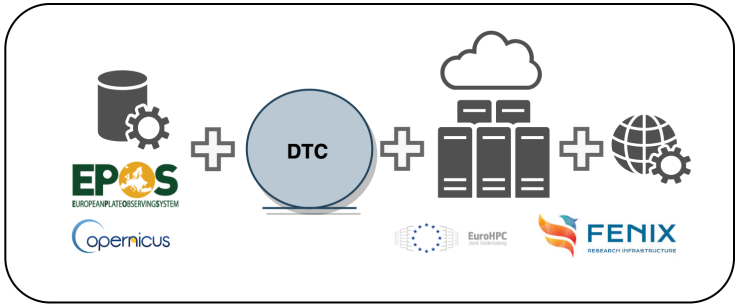
04 Verify the DTCs in operational environments at 13 **Site Demonstrators** (SDs) of particular relevance located in Europe and beyond

HPC and geosciences: a vast ecosystem of projects



The concept of Digital Twin Component (DTC)

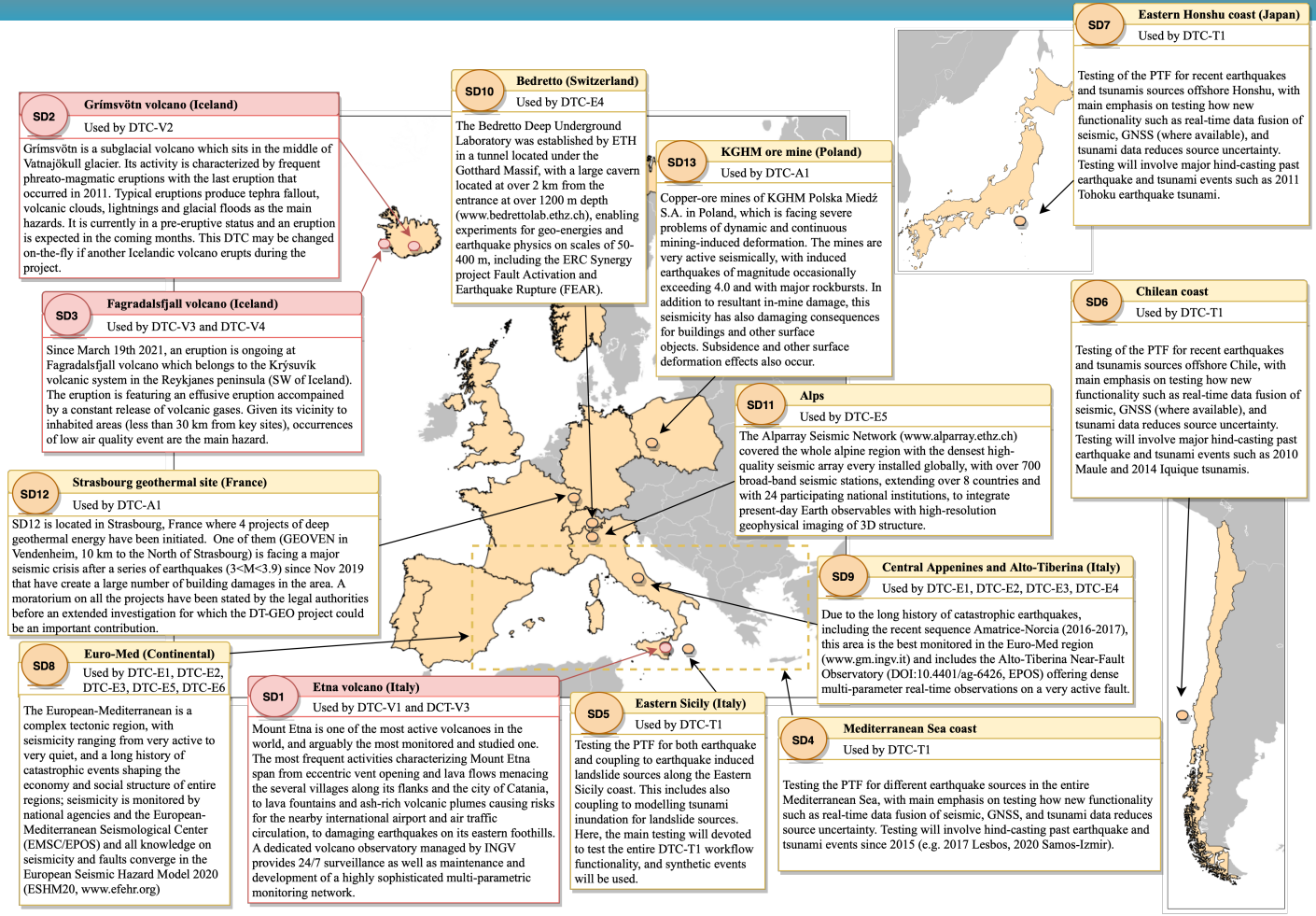
- Essentially, a **DTC is a workflow** that handles data streams and can run in distributed infrastructures:
  - A collection of coupled DTCs forms a **digital twin** (plus a set of “downstream” services or use-cases)
  - All DTCs in DT-GEO share the same architecture
  - Data in the DTCs described with rich metadata (extension of the EPOS ICS-C metadata schema)
  - DTCs can be deployed on 3 different levels: (i) local (desktop/servers), (ii) cloud (for developing stages) and, EuroHPC/FENIX HPC systems (for operational production)
- DTCs are composed by a series of **Building Blocks (BBs) or Steps** following a modular approach:
  - A step can be a physical/AI model execution, a data process, a data assimilation step, etc.
  - Some steps are micro-service oriented
  - Steps are containerized (container image creation service)
- Such modular approach facilitates:
  - Reusability and interoperability across different DTCs
  - Coupling of the different DTCs in DT-GEO
  - Automated FAIRness and QA evaluations



12 Digital Twin Components (DTCs)

DTC	Code	Hazard	Name	Target TRL	Site Demonstrator
1	DTC-V1	Volcano	Volcanic unrest dynamics	6	SD1
2	DTC-V2		Volcanic ash clouds and deposition	7	SD2
3	DTC-V3		Lava flows	6	SD1, SD3
4	DTC-V4		Volcanic gas dispersal and deposition	7	SD3
5	DTC-T1	Tsunami	Probabilistic Tsunami Forecasting (PTF)	7	SD4, SD5, SD6, SD7
6	DTC-E1	Earthquake	Probabilistic Seismic Hazard and Risk Assessment	7	SD8
7	DTC-E2		Earthquake short-term forecasting	7	SD8, SD9
8	DTC-E3		Tomography and Ground Motion Models (GMM)	7	SD8, SD9
9	DTC-E4		Fault rupture forecasting	7	SD9, SD10
10	DTC-E5		Tomography and shaking simulation	6	SD8, SD11
11	DTC-E6		Rapid event and shaking characterization	7	SD8
12	DTC-A1	Anthropogenic	Anthropogenic geophysical extreme forecasting (AGEF)	6	SD12, SD13

13 Site Demonstrators (SDs)



SD2 — Grímsvötn volcano (Iceland)

Used by DTC-V2

Grímsvötn is a subglacial volcano which sits in the middle of Vatnajökull glacier. Its activity is characterized by frequent phreato-magmatic eruptions with the last eruption that occurred in 2011. Typical eruptions produce tephra fallout, volcanic clouds, lightnings and glacial floods as the main hazards. It is currently in a pre-eruptive status and an eruption is expected in the coming months. This DTC may be changed on-the-fly if another Icelandic volcano erupts during the project.

SD3 — Fagradalsfjall volcano (Iceland)

Used by DTC-V3 and DTC-V4

Since March 19th 2021, an eruption is ongoing at Fagradalsfjall volcano which belongs to the Krýsuvík volcanic system in the Reykjanes peninsula (SW of Iceland). The eruption is featuring an effusive eruption accompained by a constant release of volcanic gases. Given its vicinity to inhabited areas (less than 30 km from key sites), occurrences of low air quality event are the main hazard.

SD12 — Strasbourg geothermal site (France)

Used by DTC-A1

SD12 is located in Strasbourg, France where 4 projects of deep geothermal energy have been initiated. One of them (GEOVEN in Vendenheim, 10 km to the North of Strasbourg) is facing a major seismic crisis after a series of earthquakes (3<M<3.9) since Nov 2019 that have create a large number of building damages in the area. A moratorium on all the projects have been stated by the legal authorities before an extended investigation for which the DT-GEO project could be an important contribution.

SD8 — Euro-Med (Continental)

Used by DTC-E1, DTC-E2, DTC-E3, DTC-E5, DTC-E6

The European-Mediterranean is a complex tectonic region, with seismicity ranging from very active to very quiet, and a long history of catastrophic events shaping the economy and social structure of entire regions; seismicity is monitored by national agencies and the European-Mediterranean Seismological Center (EMSC/EPOS) and all knowledge on seismicity and faults converge in the European Seismic Hazard Model 2020 (ESHM20, www.efehr.org)

SD1 — Etna volcano (Italy)

Used by DTC-V1 and DCT-V3

Mount Etna is one of the most active volcanoes in the world, and arguably the most monitored and studied one. The most frequent activities characterizing Mount Etna span from eccentric vent opening and lava flows menacing the several villages along its flanks and the city of Catania, to lava fountains and ash-rich volcanic plumes causing risks for the nearby international airport and air traffic circulation, to damaging earthquakes on its eastern foothills. A dedicated volcano observatory managed by INGV provides 24/7 surveillance as well as maintenance and development of a highly sophisticated multi-parametric monitoring network.

SD10 — Bedretto (Switzerland)

Used by DTC-E4

The Bedretto Deep Underground Laboratory was established by ETH in a tunnel located under the Gotthard Massif, with a large cavern located at over 2 km from the entrance at over 1200 m depth (www.bedrettolab.ethz.ch), enabling experiments for geo-energies and earthquake physics on scales of 50-400 m, including the ERC Synergy project Fault Activation and Earthquake Rupture (FEAR).

SD13 — KGHM ore mine (Poland)

Used by DTC-A1

Copper-ore mines of KGHM Polska Miedź S.A. in Poland, which is facing severe problems of dynamic and continuous mining-induced deformation. The mines are very active seismically, with induced earthquakes of magnitude occasionally exceeding 4.0 and with major rockbursts. In addition to resultant in-mine damage, this seismicity has also damaging consequences for buildings and other surface objects. Subsidence and other surface deformation effects also occur.

SD11 — Alps

Used by DTC-E5

The Alparray Seismic Network (www.alparray.ethz.ch) covered the whole alpine region with the densest high-quality seismic array every installed globally, with over 700 broad-band seismic stations, extending over 8 countries and with 24 participating national institutions, to integrate present-day Earth observables with high-resolution geophysical imaging of 3D structure.

SD5 — Eastern Sicily (Italy)

Used by DTC-T1

Testing the PTF for both earthquake and coupling to earthquake induced landslide sources along the Eastern Sicily coast. This includes also coupling to modelling tsunami inundation for landslide sources. Here, the main testing will devoted to test the entire DTC-T1 workflow functionality, and synthetic events will be used.

SD9 — Central Appenines and Alto-Tiberina (Italy)

Used by DTC-E1, DTC-E2, DTC-E3, DTC-E4

Due to the long history of catastrophic earthquakes, including the recent sequence Amatrice-Norcia (2016-2017), this area is the best monitored in the Euro-Med region (www.gm.ingv.it) and includes the Alto-Tiberina Near-Fault Observatory (DOI:10.4401/ag-6426, EPOS) offering dense multi-parameter real-time observations on a very active fault.

SD4 — Mediterranean Sea coast

Used by DTC-T1

Testing the PTF for different earthquake sources in the entire Mediterranean Sea, with main emphasis on testing how new functionality such as real-time data fusion of seismic, GNSS, and tsunami data reduces source uncertainty. Testing will involve hind-casting past earthquake and tsunami events since 2015 (e.g. 2017 Lesbos, 2020 Samos-Izmir).

SD7 — Eastern Honshu coast (Japan)

Used by DTC-T1

Testing of the PTF for recent earthquakes and tsunamis sources offshore Honshu, with main emphasis on testing how new functionality such as real-time data fusion of seismic, GNSS (where available), and tsunami data reduces source uncertainty. Testing will involve major hind-casting past earthquake and tsunami events such as 2011 Tohoku earthquake tsunami.

SD6 — Chilean coast

Used by DTC-T1

Testing of the PTF for recent earthquakes and tsunamis sources offshore Chile, with main emphasis on testing how new functionality such as real-time data fusion of seismic, GNSS (where available), and tsunami data reduces source uncertainty. Testing will involve major hind-casting past earthquake and tsunami events such as 2010 Maule and 2014 Iquique tsunamis.

Presentation outline

01 DT-GEO in a nutshell

02 Development of DT-GEO components: interim results

03 What’s next?

3 transversal pillars

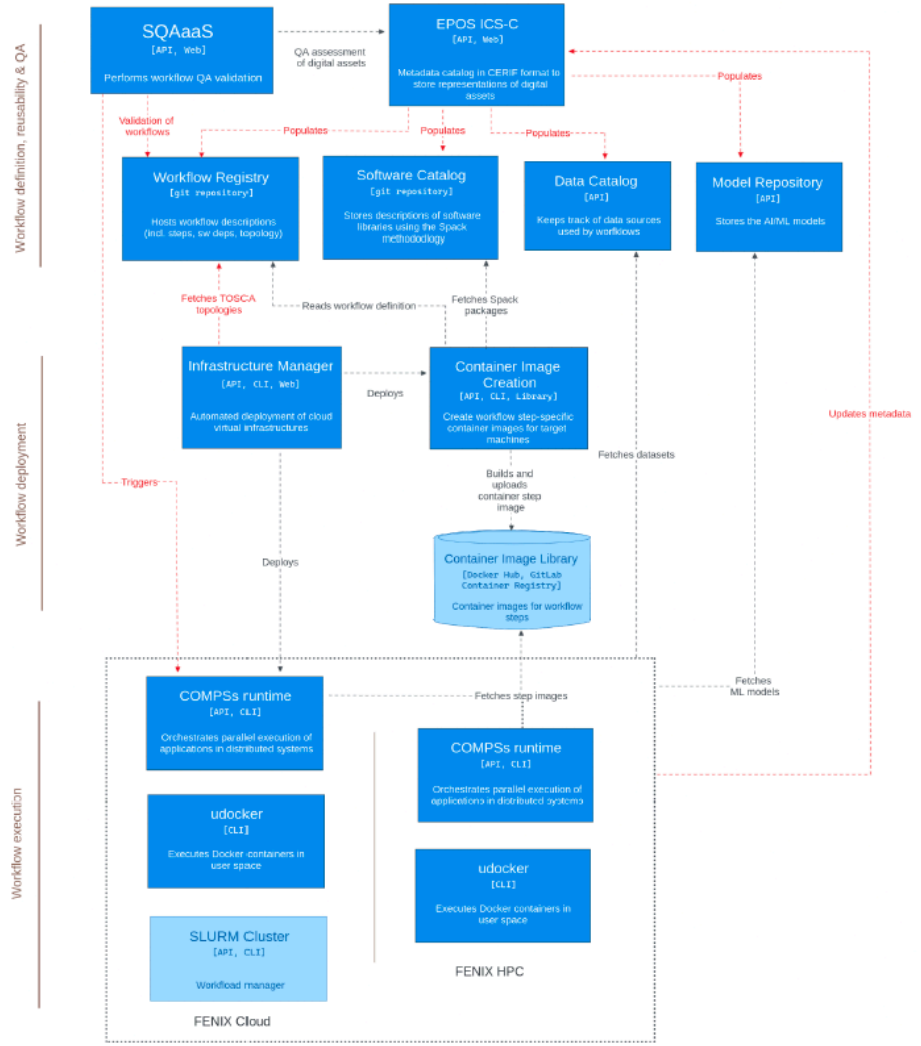
Pillar 1 Workflows and data architecture	Pillar 2 Computational infrastructure and AI	Pillar 3 EOSC-enabled data management plan and exploitation
• Enable the deployment and **execution of workflows** in HPC systems and virtual cloud environments • Workflow development (HPCWaaS) • Workflow deployment as HPC-ready containers and registered as services ready to use • Relay on the PyCOMPSs runtime to orchestrate workflow execution in the HPC systems (FENIX/EuroHPC)	• Provide access to the DT-GEO virtual cloud (development) and HPC (production) infrastructures • S**upport to containerisation and execution** on of DTCs and SDs. • Manage the **on-demand execution** of DT-GEO workflows including technical user-support and direct engagement with the 4 vertical pillars • Adapt the DT-GEO infrastructure to run on external Elastic Cloud Computing Clusters	• Provide a flexible environment to check the quality of the generated Digital Assets (DAs) repositories and integration of data/software quality management through an **automatic FAIR validation**. • Ensure quality of the DAs extending the EOSC-synergy project results on SQAaaS using CI/CD pipelines • FAIR validation (DSpace-CRIS metadata model and FAIR validator)

DT-GEO workflow components on 3 levels

1. Workflow definition, reusability and QA

2. Workflow deployment

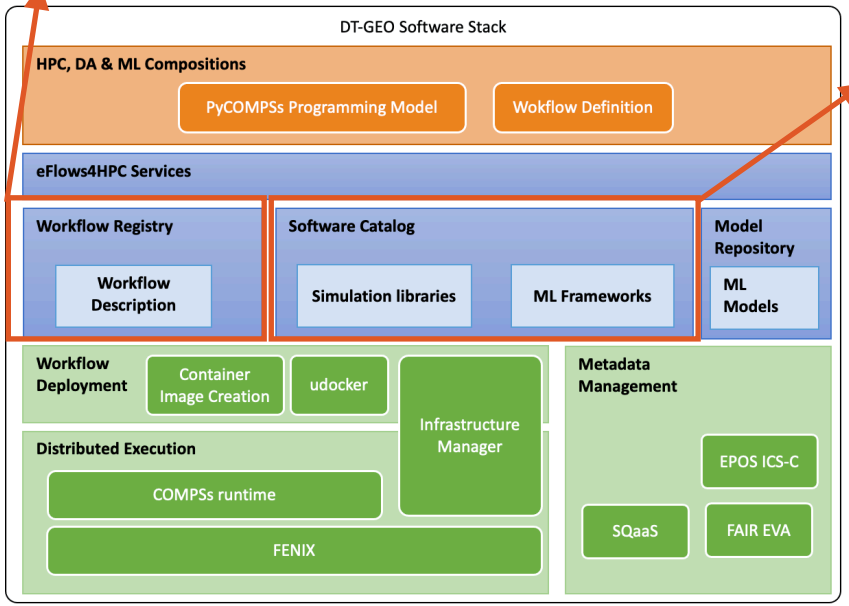
3. Workflow execution



Level 1: workflow definition, reusability and QA

Workflow Registry	git repository hosting workflow descriptions using spack
Software Catalog	git repository to store the description of the software to be used in DTC workflows using the Spack structure
Data Catalog	Keeps track of workflow data sources (meta-data)
Model repository	Repository with underlaying physical and AI/ML models
SQAaaS	Automated workflow QA validation (EOSC-synergy) and provenance (eFlows4HPC)

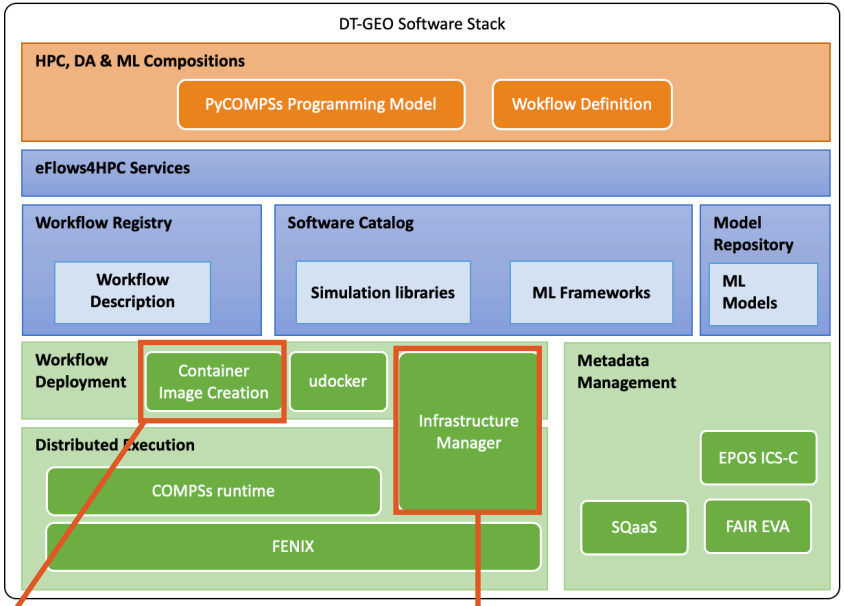
```
workflow-registry
 |- workflow_1
 |    |- tosca
 |    |    |- types.yml      TOSCA description of the different components involved in the workflow
 |    |    ...
 |    |- step_1
 |    |    |- spack.yml      Software requirements for this workflow step as a Spack environment specification
 |    |    |- src            PyCOMPSs code of the workflow step
 |    |    ...
 |    |- step_2
 |         ....
 |- workflow_2
 |                            ...
```

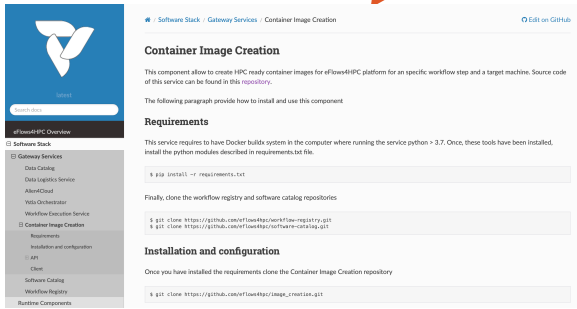


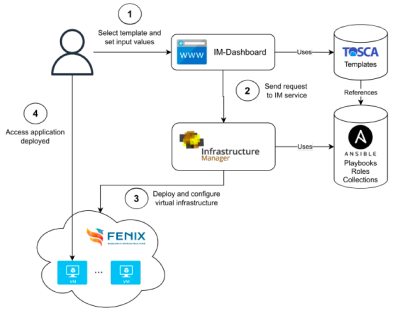

```
software-catalog
 |- packages
 |    |- software_1
 |    |    |- package.py  Installation description following the Spack package format
 |    |    ...
 |    |- software_2
 |         ....
 |- cfg       Spack configuration used by the Image Creation Service
 |
 |- repo.yaml    Spack description of for this repository
```

Level 2: Workflow deployment

Infrastructure Manager (IM) service	IM-Dashboard to select the kind of virtual infrastructure to deploy from a set of TOSCA templates available in a repository
Container image creation	Creates container images for WF steps on target machines
Container image library	Hosts container images for the different WF steps

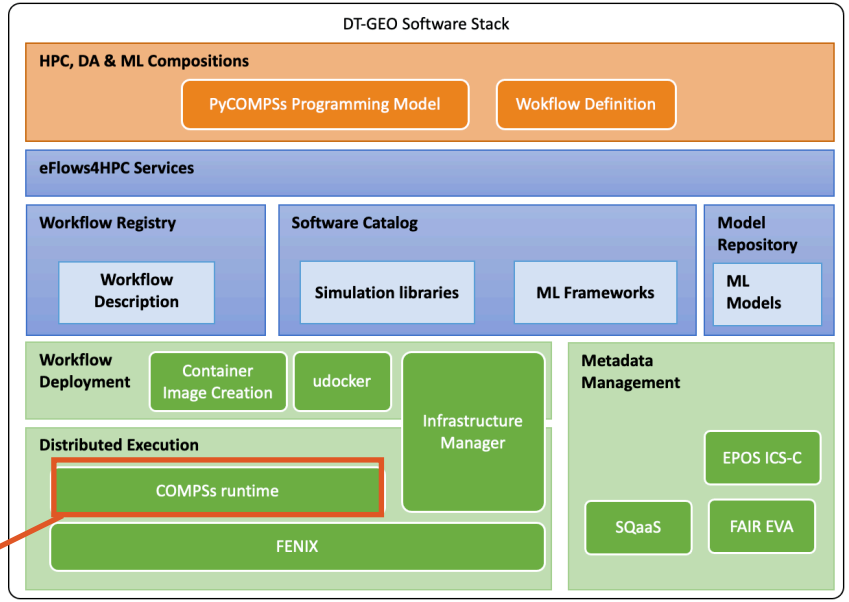






Level 3: Workflow execution

COMPSs runtime	Parallel workflow orchestration in distributed systems
udocker	Executes docker containers in user space (FENIX cloud or federated HPC-systems)
SLURM	Workload manager (HPC)



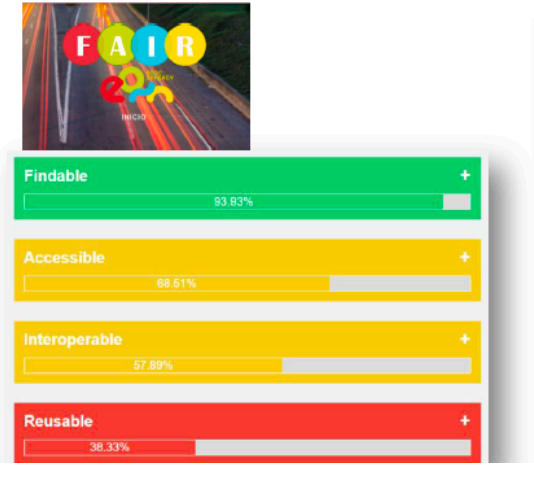
COMPSs runtime	
Task Dependency Analysis	Builds a task dependency graph
Task Scheduling	Scheduled in distributed resources
Resource Management	For cloud environments, ellastically adapt resources
Job and data Management	Perform remote execution of tasks and the data transfers

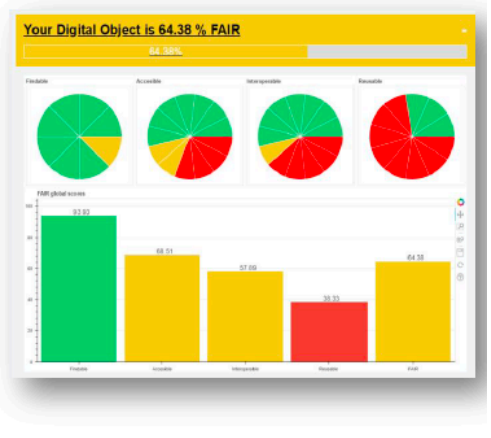
FAIR evaluation of Digital Assets

- DT-GEO is commited with the compliance of the diverse digital assets (data, source code and workflows) with FAIR principles.
- Leverage developments from other projects (EOSC-synergy and eFlows4HPC)

1. The EOSC-Synergy **FAIR Evaluator, Validator and Advisor** (FAIR-EVA; DOI: 10.20350/digitalCSIC/14559) tool:

  - Can be deployed as a stand-alone application or in a docker container (implemented also as a web service with an API)
  - FAIR EVA will be extended to the EPOS metadata scheme





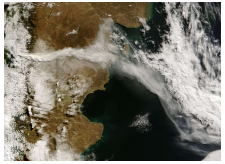
EOSC SYNERGY

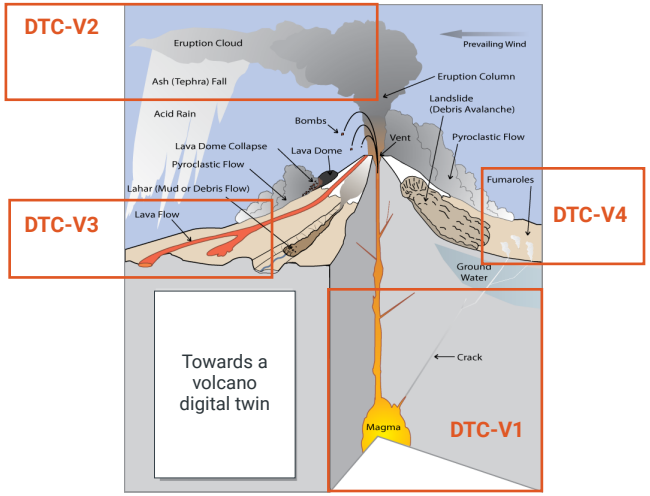
Workflow provenance

2. Achievement of FAIRness for computational workflows through the **Workflow Provenance recording mechanism implemented in PyCOMPSs** (ensures the reproducibility of workflows defined with this programming model)

- Lightweight provenance recording mechanism in PyCOMPSs
- The runtime records in a log file all accesses to individual data files and datasets in the workflow and generates the corresponding metadata associated with the workflow and its execution in the corresponding resources
- Produces a machine-readable JSON Linked Data (JSON-LD)
- Main vocabulary based on schema.org
- Based on RO-Crate, with a strong ecosystem including:
  - The ro-crate-py library
  - WorkflowHub FAIR workflow registry

The steps of the DTC-V2 workflow



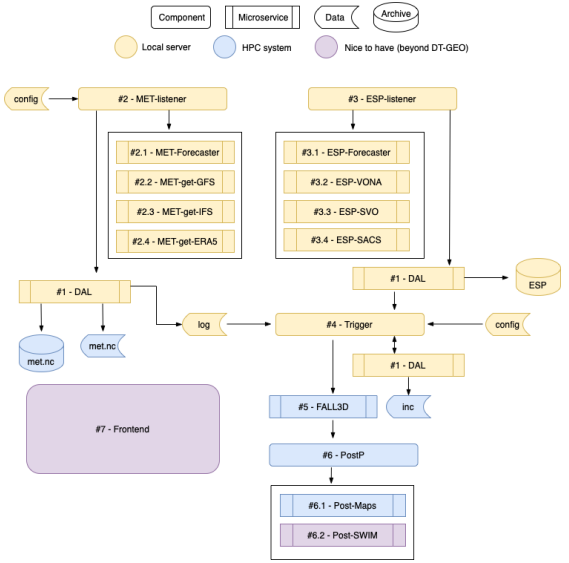


Digital twin	Triggers	Impacts	
DTC-V1		**DTC-V2**	Ash dispersal
		DTC-V3	Lava flows
		DTC-V4	Gas dispersal
State of the volcano based on geophysical monitoring and models		Urgent computing impact assesment combining physics-based models (HPC) and observations	

DTC-V2 objective

Merge real-time ground-based and satellite observations with the FALL3D model to generate ensemble-based deterministic and probabilistic volcanic ash forecast maps and products

Step	Step name
1	Data Access Layer (DAL)
2	MET-listener
3	ESP-listener
4	Trigger
5	FALL3D
6	Post Process



ChEESE

Presentation outline

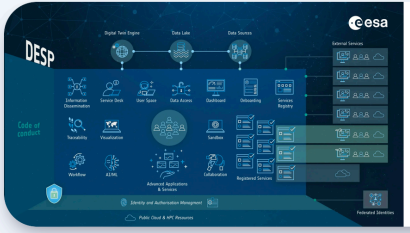
01 DT-GEO in a nutshell

02 Development of DT-GEO components: interim results

03 What's next?

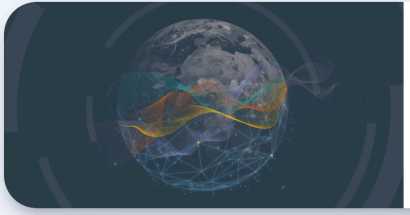
Towards Destination Earth

Destination Earth (https://destination-earth.eu/) is a flagship initiative of the European Commission to develop a highly-accurate digital model of the Earth



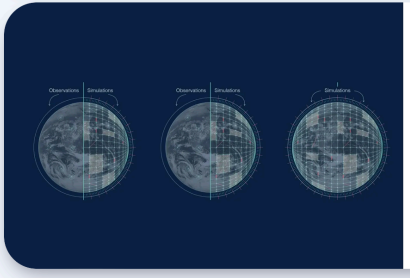
Core Service Platform

The platform will provide evidence-based decision-making tools, applications and services, based on an open, flexible, and secure cloud-based computing infrastructure.



Data Lake

The data lake will bring together data from ESA, EUMETSAT, ECMWF as well as from Copernicus, and many other diverse sources, with new data from the Digital Twins. It will allow discovery and data access as well as big data processing in the cloud.



Digital Twins and Digital Twin Engine

DestinE is creating several digital replicas covering different aspects of the Earth system and based on state-of-the-art simulations and observations. ECMWF is implementing the Digital Twin Engine, the complex software and data services needed for Earth System digital replicas, as well as the first two digital twins; Climate Change Adaptation, which will provide multidecadal simulations, and the Weather-induced Extremes twin, with both high-resolution forecasts and on-demand simulations.

Other DTCs will follow later on (although not all the DTCs in DT-GEO target at DestinE)

A Digital Twin for GEOphysical extremes

Towards Destination Earth?

Destination Earth

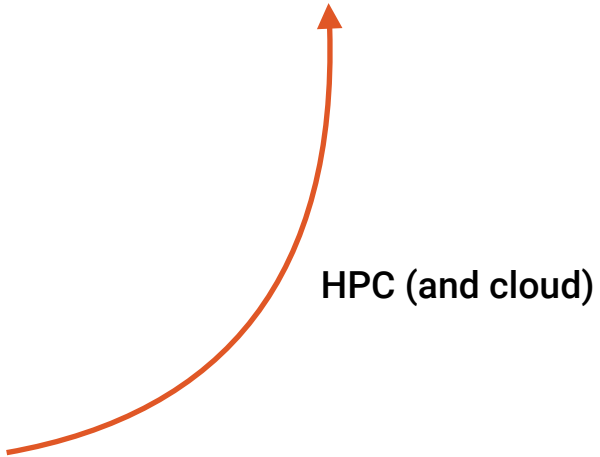
Phase 1 (Sep 2022-May 2024)			Phase 2?		Phase 3?			
2022	2023	2024	2025	2026	2027	2028	2029	2030

Climate Adaption digital TwinS (lead by CSC)	
Destination Earth On-Demand Extremes (lead by Meto France)	

Extend to new applications
Integrate components

ECMWF	Digital Twins, fusion of real-time observations and high-resolution predictive modelling
ESA	Core Service Platform, a user-friendly entry point for DestinE users
EUMETSAT	Data lake; access to the data needed for the Digital Twins and the Core Service Platform operations

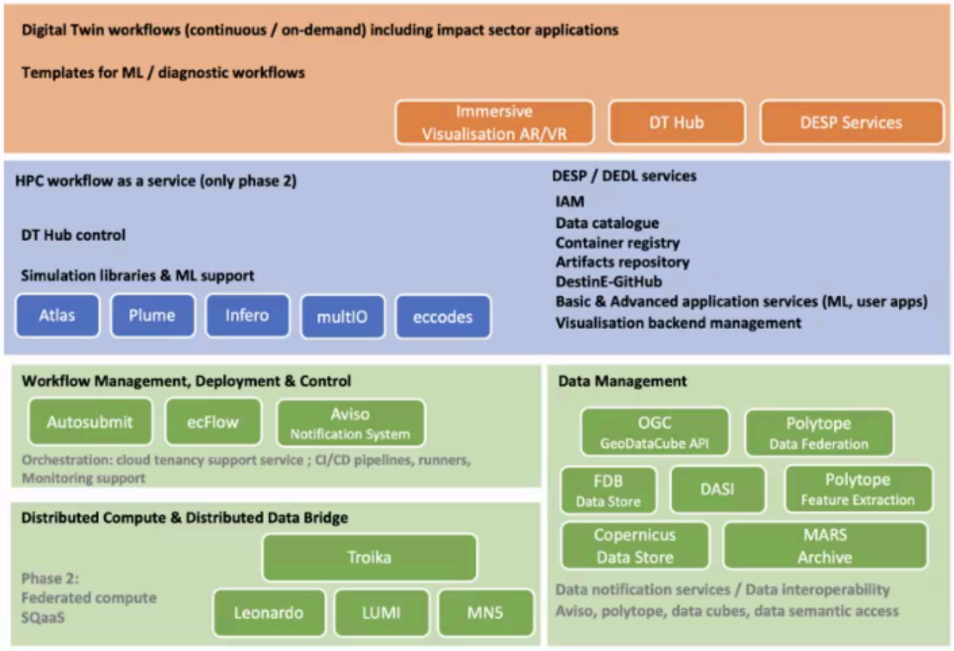
Addition of other DTs and (dowstream) services

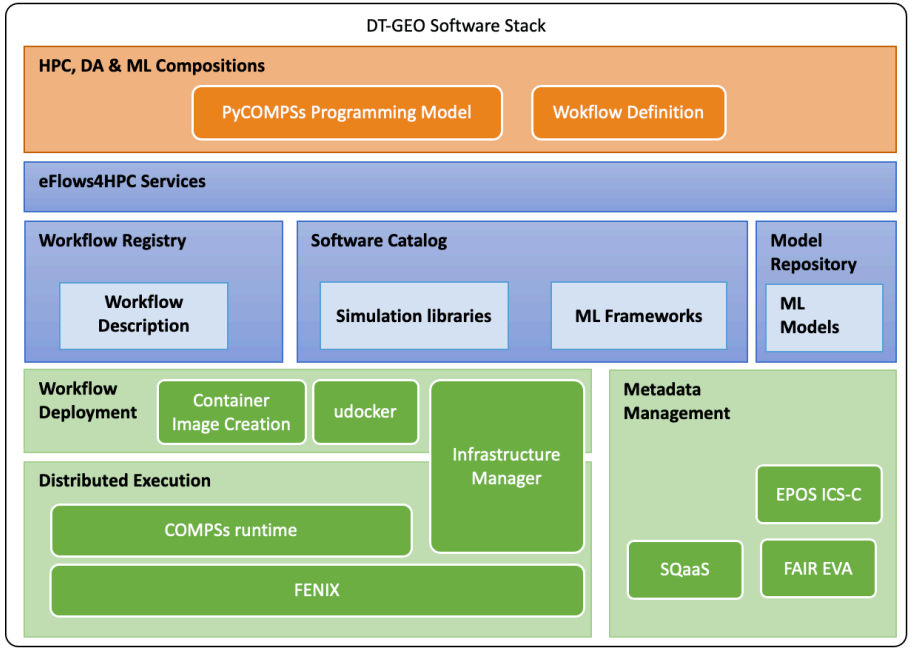


HPC (and cloud)

HORIZON-INFRA-2021-TECH-01-01	**DT-GEO**:	Geophysical extremes (lead by CSIC)
	BioDT:	Biodiversity (lead CSC)
	InterTwin:	Interdisciplinary engine (lead by EGI)
H2020-EU.3.2. - SOCIETAL CHALLENGES	**ILIAD**:	Ocean (leal by NETCOMPANY)

Similar architectures (tool choice differ)





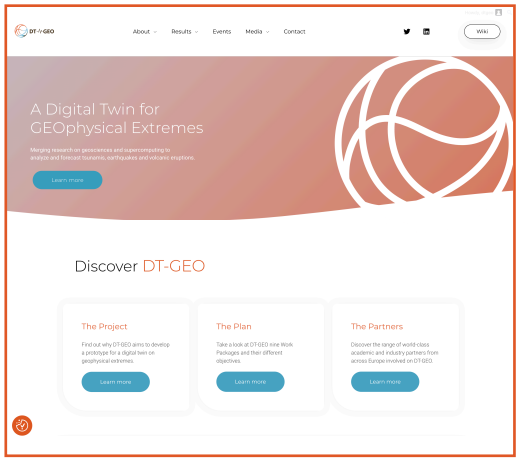
A Digital Twin for GEOphysical extremes

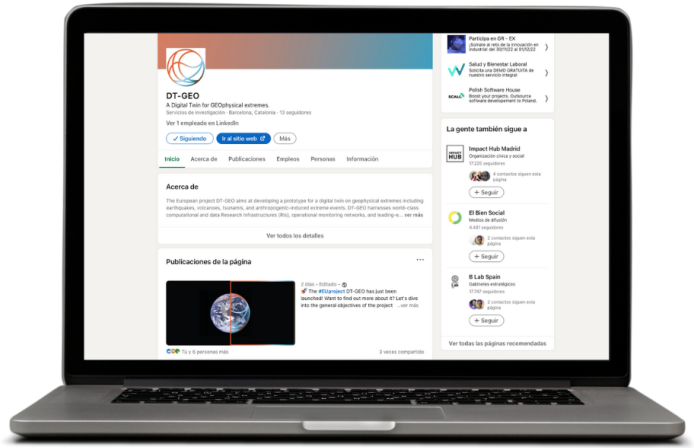
Dissemination

https://dtgeo.eu

@dtgeo_eu

linkedin.com/company/dt-geo/





THANK YOU

afolch@geo3bcn.csic.es

@dtgeo_eu

linkedin.com/company/dt-geo/

Project funded by Horizon Europe under the grant agreement No 101058129.