Towards geophysical extremes twins: status and next steps of the DT-GEO project

Arnau Folch

Geociencias Barcelona (GEO3BCN-CSIC), Spain

3rd Destination earth User eXchange
Darmstadt, 15-16 October 2024

DT-GEO project Information

| Type of Action	HORIZON-RIA	
Call	INFRA-2021-TECH-01	Next generation of scientific instrumentation, tools and methods
Topic	INFRA-2021-TECH-01-01	Interdisciplinary **digital twins** for modelling and simulating complex phenomena at the service of **research infrastructure** communities
Dates	From Sep 2022 to Aug 2025	
Budget	15,1 M€	
Consortium	Geophysics domain	Research, Academia, Private
	Target Research Infrastructures (RIs)	

INFRA-2021-TECH-01-01

The concept of Digital Twin (DT)

A Digital Twin is a **virtual lifecycle environment** that contains a data-informed **replica** of a real system, model-based **prediction** capabilities (scenarios) and, ideally, can provide **feedback** (decisions) to modify the real system (i.e. to close the lifecycle loop)
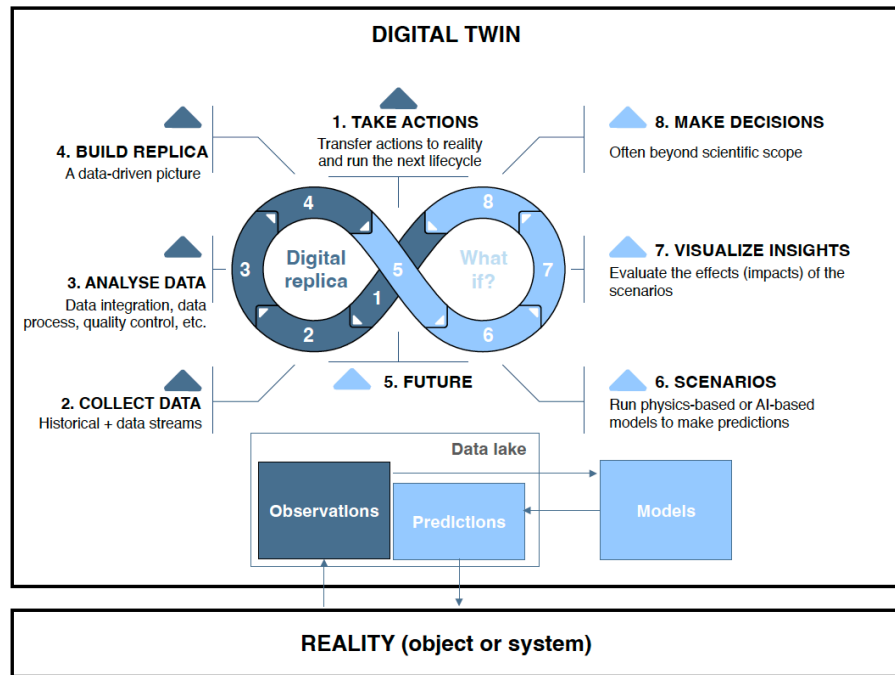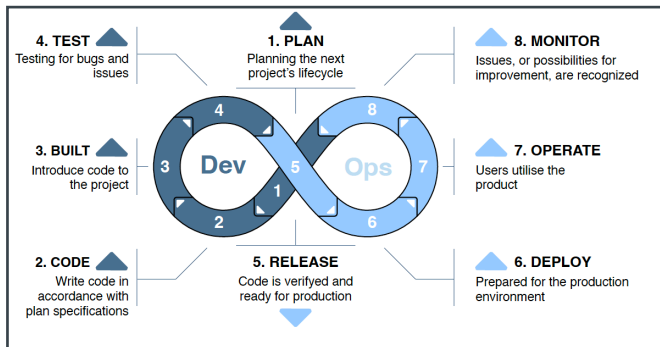
DT-GEO general objectives

01 Deploy a pre-operational prototype of **Digital Twin (DT) on geophysical extremes** (potential integration in the Destination Earth flagship 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 (EWS)
   2. Short-term forecasts
   3. Long-term hazard assessments

03 Provide a flexible framework for **automated FAIR-validation** of Digital Assets (DAs) and its integration in 2 Research Infrastructures (RIs)

04 Verify the DTCs in operational environments at 13 **Site Demonstrators** (SDs) of particular relevance located in Europe and beyond (TRL= 6,7)

DTC	Hazard	DTC name
1	Volcano	Volcanic unrest
2		Volcanic ash clouds
3		Lava flows
4		Volcanic gas dispersal
5	Tsunami	Tsunami Forecasting
6	Earthquakes	Seismic Hazard
7		Earthquake forecasting
8		Tomography
9		Fault rupture
10		Shaking simulation
11		Aftershocks
12	Anthropogenic	Anthropogenic seismicity

The DT-GEO structure

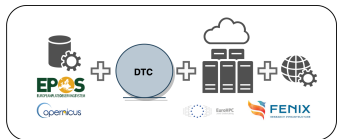| WP2	Workflows and data architecture
WP3	Computational infrastructure
WP4	Metadata and automated FAIRness evaluation of Digital Assets

Transversal elements: provide homogeneity and interoperability across the DTCs (**leveraged from other projects**)

12 Digital Twin Components (DTCs)

13 Site Demonstrators (SDs)

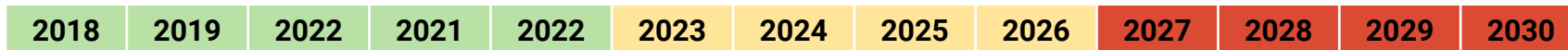| WP5	WP6	WP7	WP8
Volcanoes	**Tsunamis**	**Earthquakes**	**Anthropogenic**
• 4 DTCs for volcanoes • 3 SDs	• 1 DTCs for tsunamis • 4 SDs	• 6 DTCs for earthquakes • 4 SDs	• 2 DTCs for anthropogenic • 2 SDs

Vertical pillars

An ecosystem of European projects

DT-GEO: current status

Phase 1 — From M1 to M12
Phase 2 — From M13 to M24
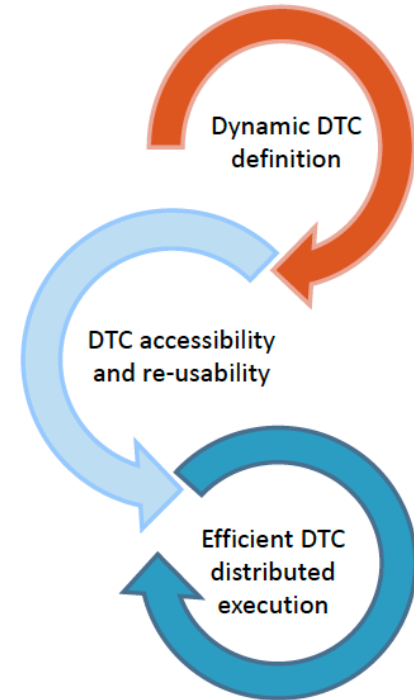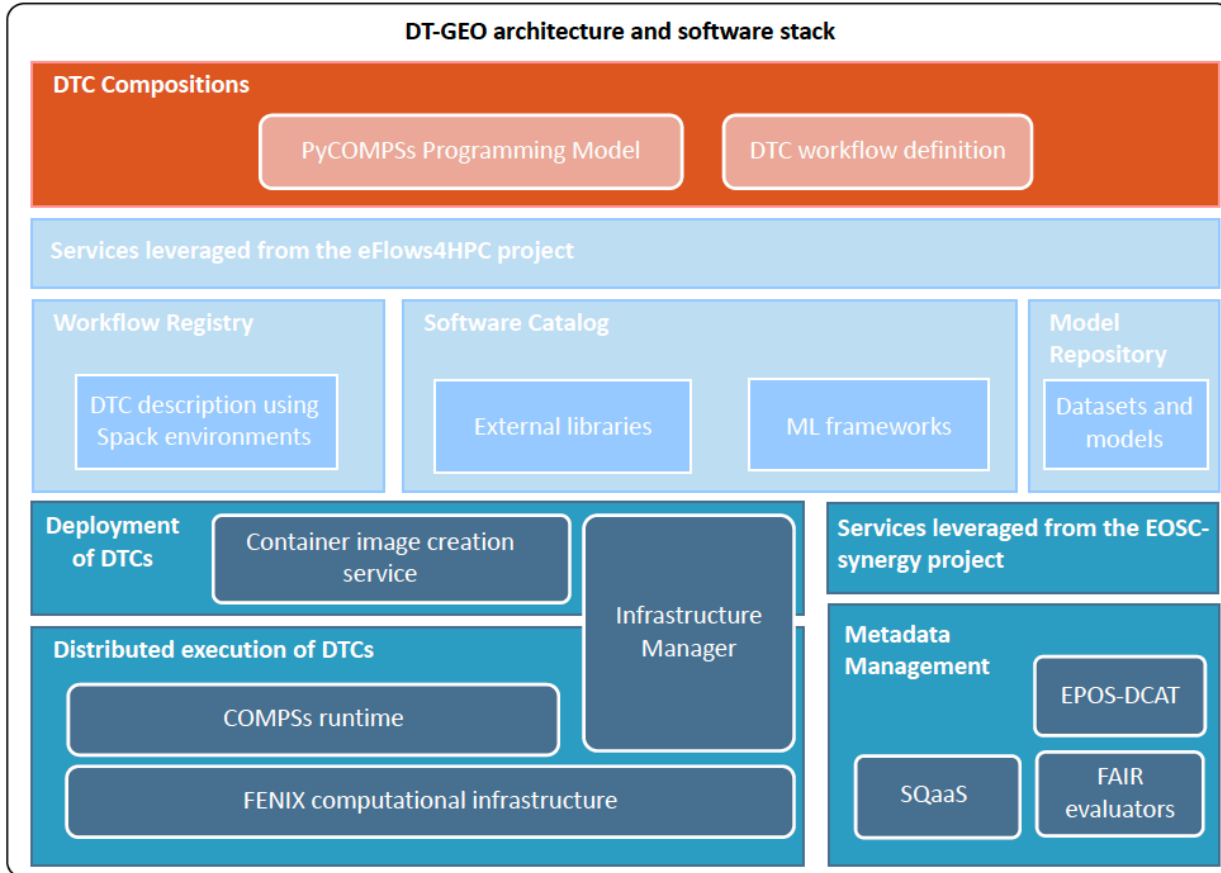**Phase 3** — From M25 to M36

Phase 1
- ✓ Project setup
- ✓ Collection of requirements
- ✓ Define the computational infrastructure
- ✓ Blueprint of DT architecture
- ✓ Metadata scheme
- ✓ FAIR quality-based ecosystem

Phase 2
- ✓ Beta implementation of the DTCs
- ✦ Early execution of SDs in the FENIX cloud infrastructure (testing/staging)
- ✓ Design and development of the DT architecture, second iteration loop
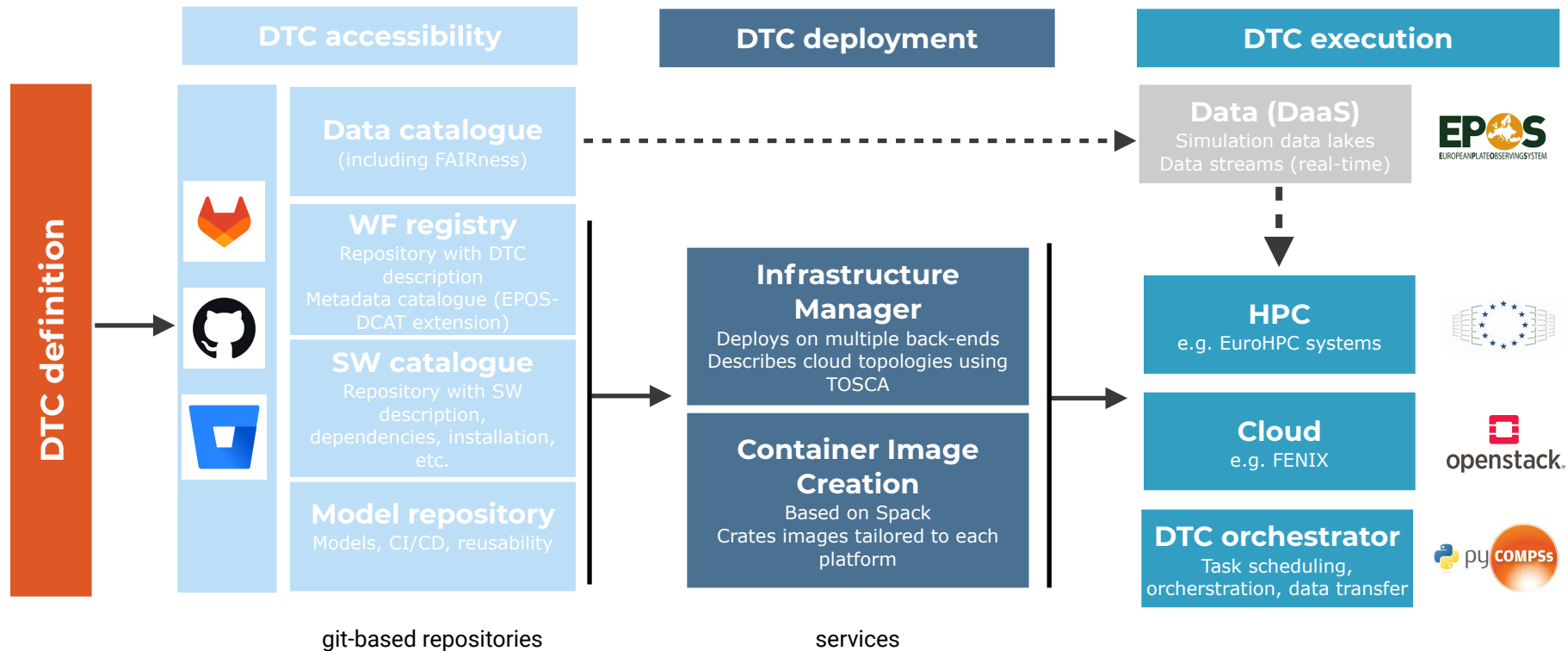
Phase 3
- • Further implementation of the DTCs (some coupled)
- • Run the SDs in operational environments using HPC

DT-GEO blueprint architecture

DT-GEO blueprint architecture

DTC definition

DTC accessibility

Data catalogue
(including FAIRness)

WF registry
Repository with DTC description
Metadata catalogue (EPOS-DCAT extension)

SW catalogue
Repository with SW description, dependencies, installation, etc.

Model repository
Models, CI/CD, reusability

git-based repositories

DTC deployment

Infrastructure Manager
Deploys on multiple back-ends
Describes cloud topologies using TOSCA

Container Image Creation
Based on Spack
Crates images tailored to each platform

services

DTC execution

Data (DaaS)
Simulation data lakes
Data streams (real-time)

HPC
e.g. EuroHPC systems

Cloud
e.g. FENIX

DTC orchestrator
Task scheduling, orcherstration, data transfer
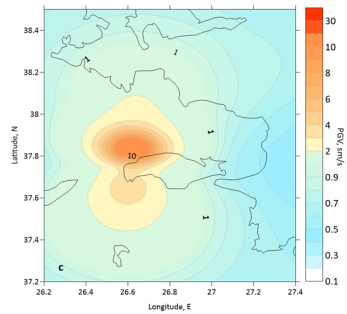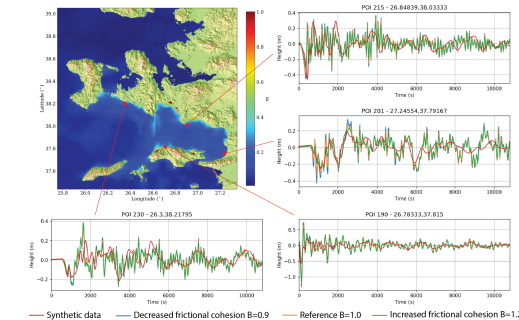
What's next? : rise TRL to 6,7

Figure 1.2.4. Location and definition of the 13 Site Demonstrators (SD) used to validate the 12 DTC in relevant and operational environments (TRL 6 and 7)

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 accompanied 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.

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 under scales of 50-400 m, including the ERC Synergy project Fault Activation and Earthquake Rupture (FEAR).

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.

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.

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.

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.

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.

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.

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.

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 be devoted to test the entire DTC-T1 workflow functionality, and synthetic events will be used.

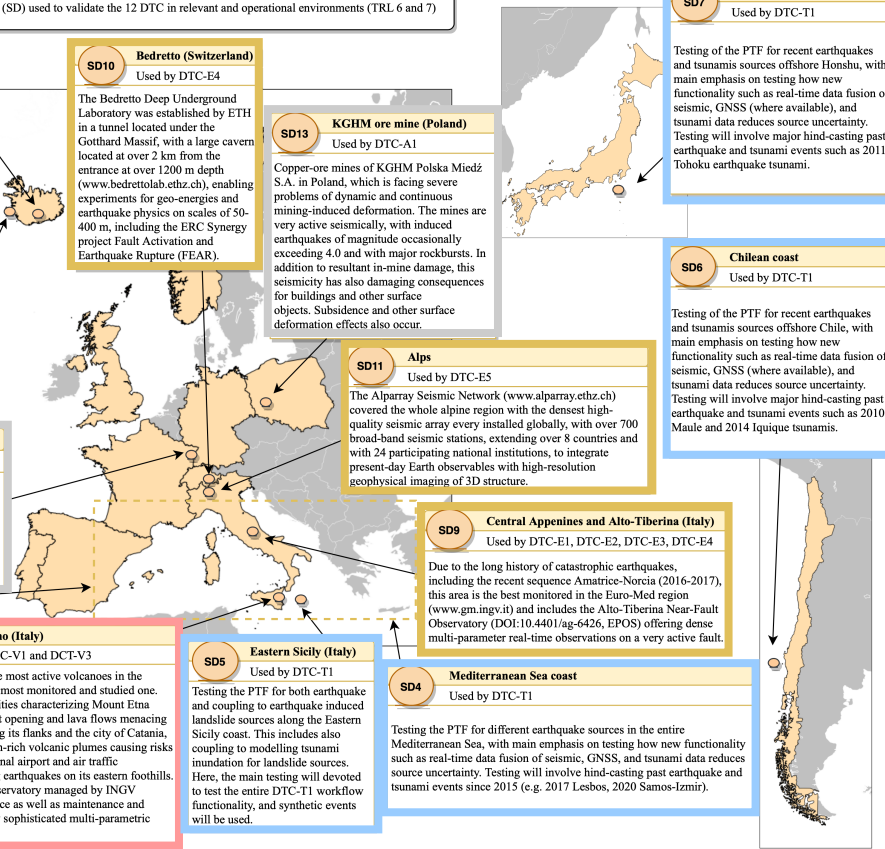**SD4 — Mediterranean Sea coast**
Used by DTC-T1

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).

WP5	3 SDs
WP6	4 SDs
WP7	4 SDs
WP8	2 SDs

POI 215 - 26.84839,38.03333

POI 201 - 27.24554,37.79167

POI 230 - 26.3.38.21795

POI 190 - 26.78333,37.815

— Synthetic data — Decreased frictional cohesion B=0.9 — Reference B=1.0 — Increased frictional cohesion B=1.2

DestinE integration continuum

DestinE Integration continuum

Full integration mode
Directly integrated in the DestinE system

Coupling mode
Own simulation and data fusion tasks

WF compatible with DTE
Same WFMS, HPC and data handing

WF independent from DTE
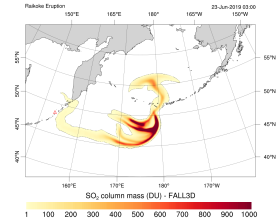Different WFMS, HPC and data handing

Postprocessing mode
Data post-process application
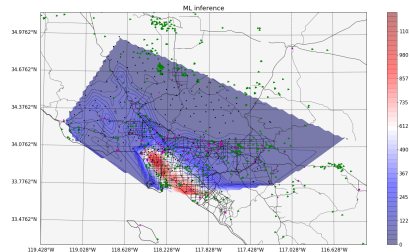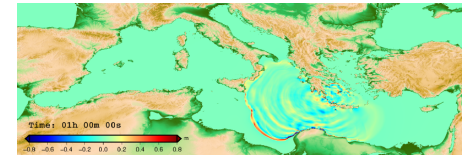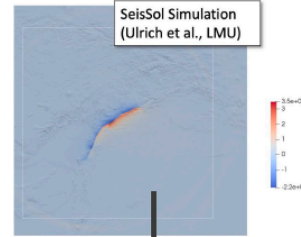
DEDL data producer

DEDL data consumer

DTC-V2 — Volcanic ash clouds and deposition

DTC-T1 — Probabilistic Tsunami Forecasting (PTF)

SeisSol Simulation (Ulrich et al., LMU)

DTC-E5 — Tomography and shaking simulation

THANK YOU

 https://dtgeo.eu

 @dtgeo_eu

 linkedin.com/company/dt-geo/
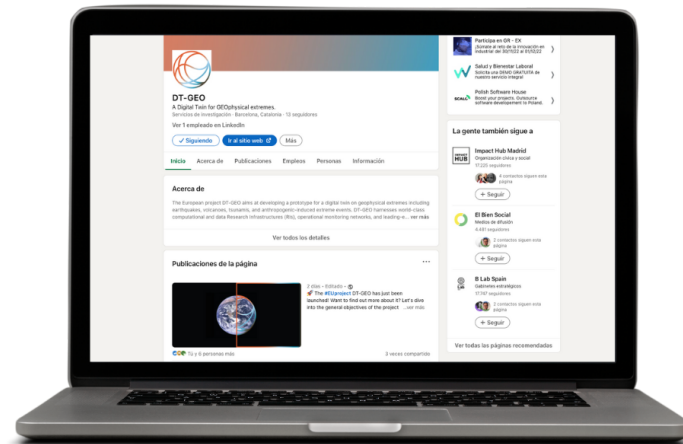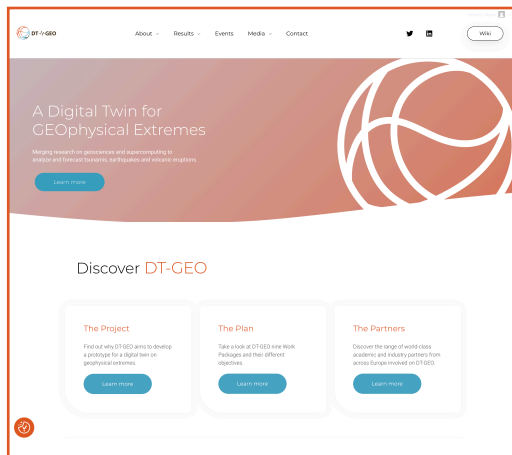
Workflow architecture

- Set up an organization (https://gitlab.com/dtgeo) in GitLab including **Workflow Registry** and the **Software Catalog**.
- Set up of Container Image Creation tool (for specific target machines).
- CI/CD approach to validate the DTCs in the staging environment and generate push-driven container images.

SQAaaS reports for the different DAs

SQAaaS pipeline templates

Configuration files to submit COMPSs jobs to Slurm in different clusters

Description of the software from DTCs using the Spack methodology

Store the WF descriptions using the Spack environment methodology (at least the TOSCA description of the WF), the code of the different steps, and the required software per step

Computational architecture

- DT-GEO uses an **Infrastructure Manager** (IM) to deploy DTCs execution in cloud environments.
    - Open stack @CINECA ADA cloud service (IaaS) for DTC testing (cloud).
- The DTC execution (SDs) will rely on HPC.
- Configuration of Spack modules on the HPC clusters and support for PyCOMPSs module configuration.

ADA OpenStack dashboard: https://adacloud.hpc.cineca.it
Documentation: link

Interactive computing service: https://jupyter.g100.cineca.it/hub
Documentation: link
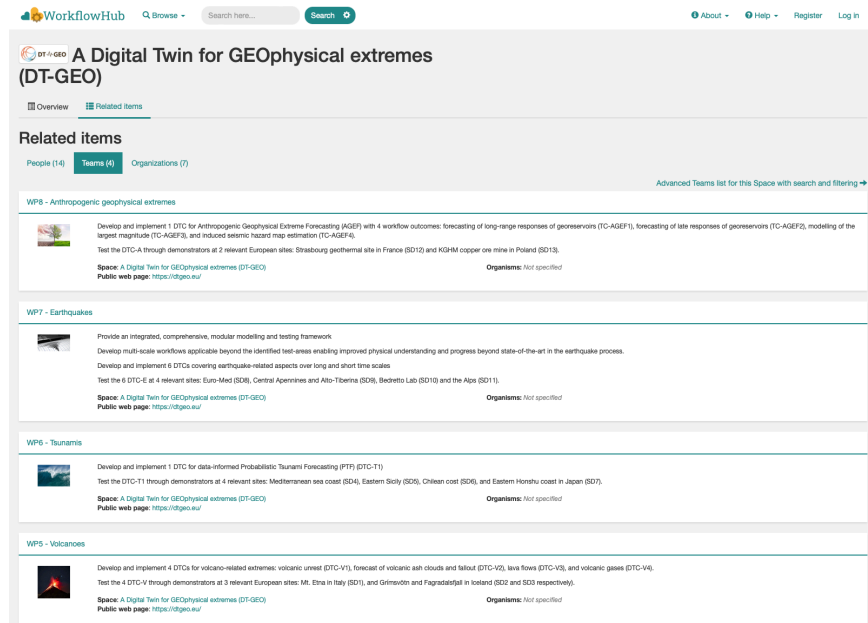
FAIRness evaluation of DTCs

- Services for DTC reusability (catalogues, registries, and repositories), including the creation of a **Workflows Hub Registry**.
- Recording of workflow provenance: **capture the details of a DTC execution as metadata**.
- The interoperability of the generated metadata is guaranteed by the use of the RO-Crate specification.
- RO-Crate (Research Object Crate) is a method for aggregating and describing research data with associated metadata.

https://workflowhub.eu/programmes/36#projects

Data architecture

- Metadata schemes to characterise the DTCs and their relationships (aligned with geo-INQUIRE):
  ‣ **DCAT**: Data Catalog vocabulary for publishing data catalogs on the web (https://www.w3.org/TR/vocab-dcat/).
  ‣ **EPOS-DCAT-AP**: an extension of the DCAT Application Profile for Research Infrastructures in the solid-Earth domain based on the Common European Research Information Format (CERIF).
  ‣ **DT-GEO**: further extension of EPOS-DCAT to accommodate metadata for new entities in the DT-GEO schema such as Workflow and Step.
- Adoption and extension of the EOSC-synergy SQAaaS platform for assuring quality on DAs (e.g. FAIR-EVA has been extended with a specific plugin to integrate with the EPOS-DCAT catalog).