A digital twin component for atmospheric volcanic dispersal forecasts

Next-generation HPC workflows for natural hazards (eFlows4HPC workshop)

Barcelona, 13 September 2023

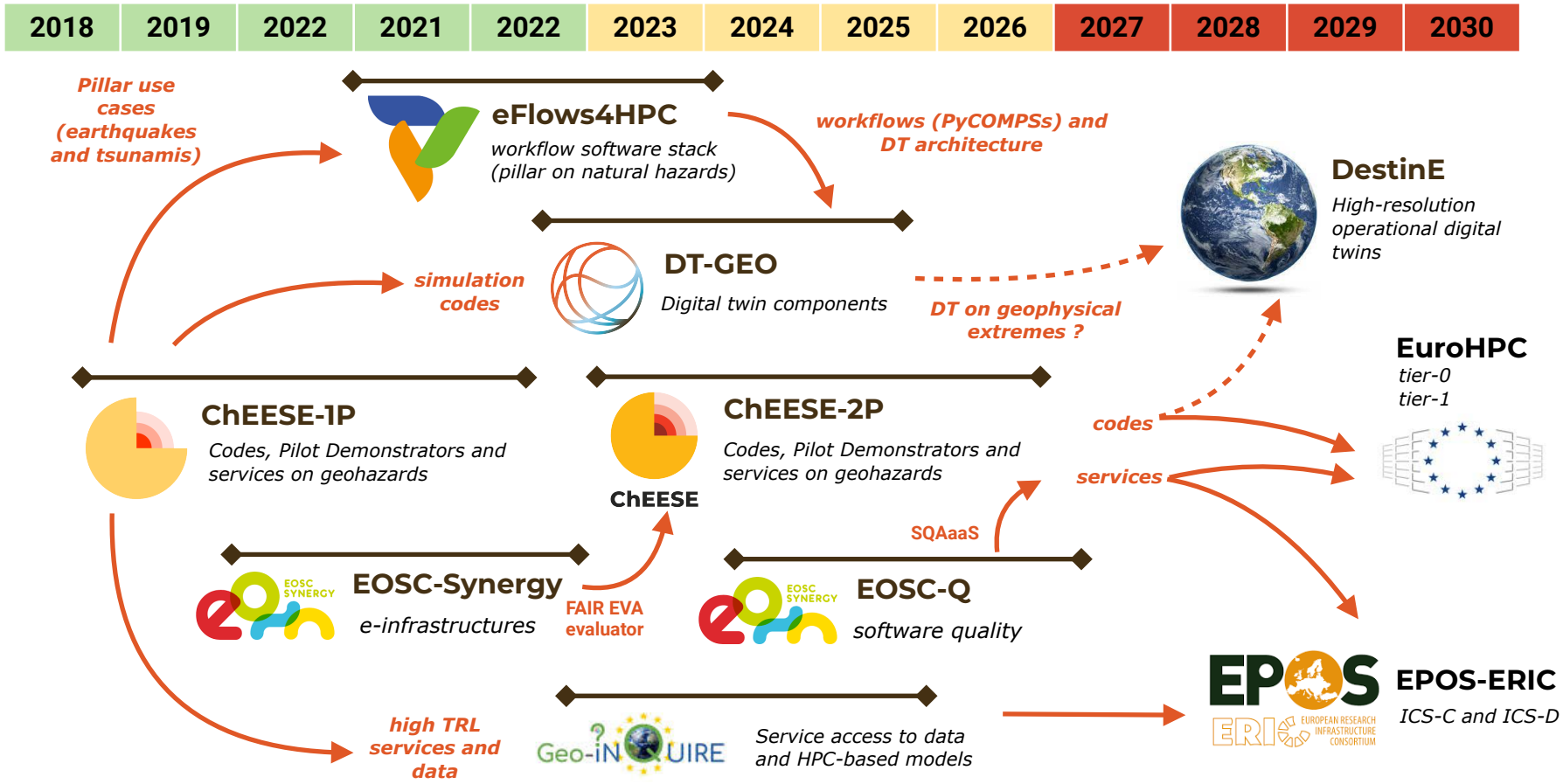
Arnau Folch
Geosciences Barcelona (GEO3BCN-CSIC)

eFlows4HPC
Enabling dynamic and Intelligent workflows in the future EuroHPC ecosystem

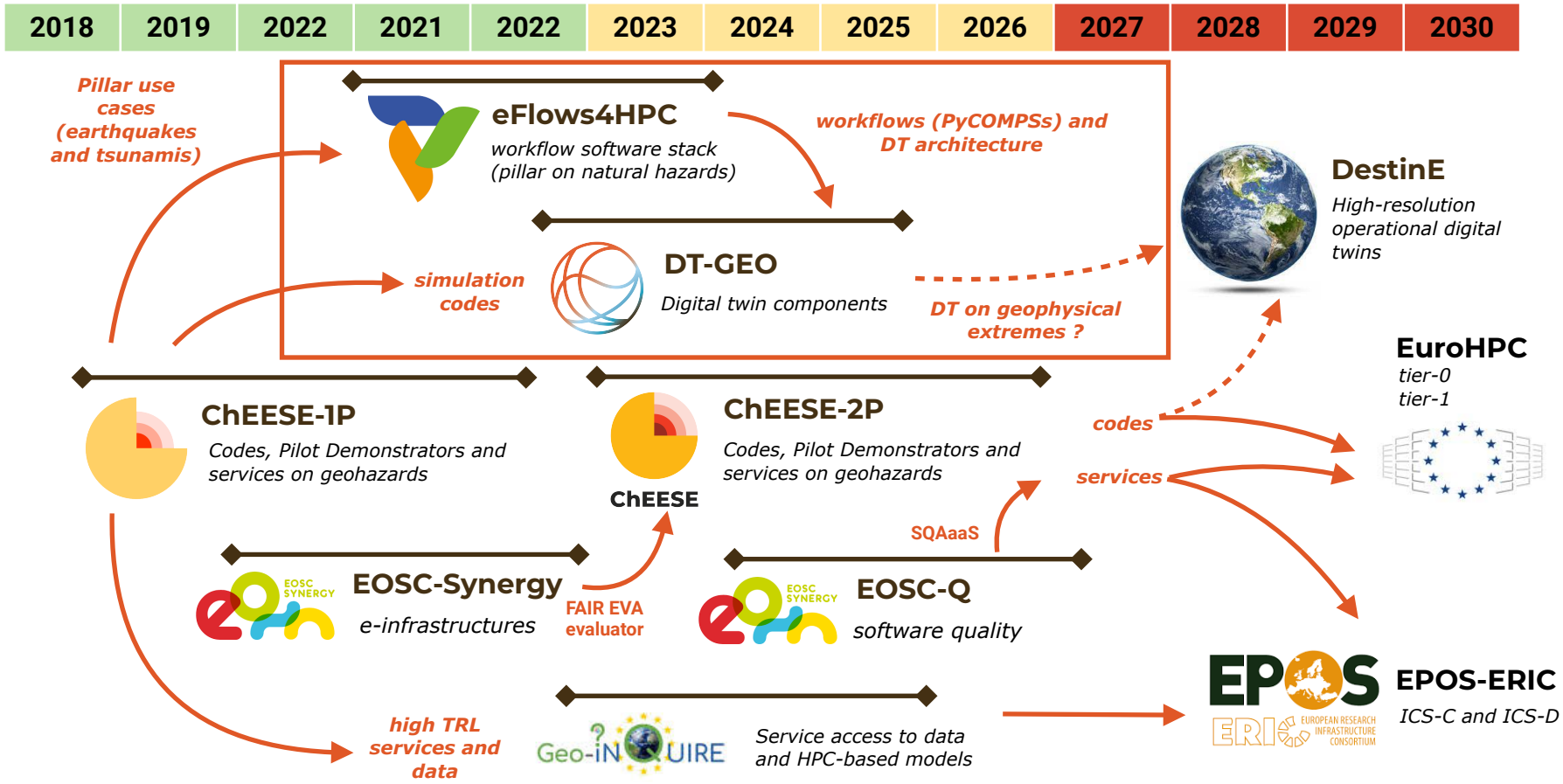
DT-GEO

This project has received funding from the European Union's Horizon research and innovation programme under the grant agreement No 101058129

HPC and geosciences: a vast ecosystem of projects



HPC and geosciences: a vast ecosystem of projects



DT-GEO project: A Digital Twin for GEOphysical extremes

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 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 EOSC-enabling and FAIR-validation of project assets and outcomes and its integration in 2 Research Infrastructures (RIs):

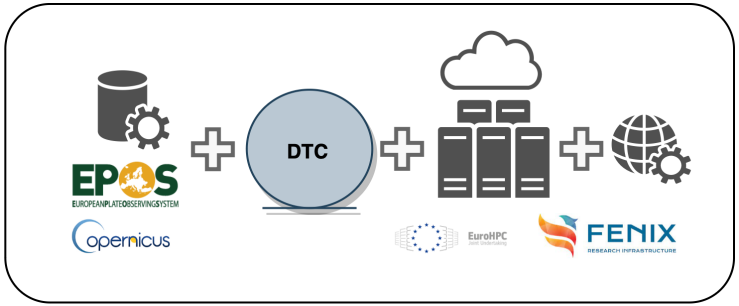
1. The European Plate Observing System (**EPOS**)
2. HPC/virtual cloud computing (**EuroHPC/FENIX**)

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

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 schema)
  - DTCs can be deployed on 3 different levels: local, cloud (virtual), and HPC (EuroHPC/FENIX)

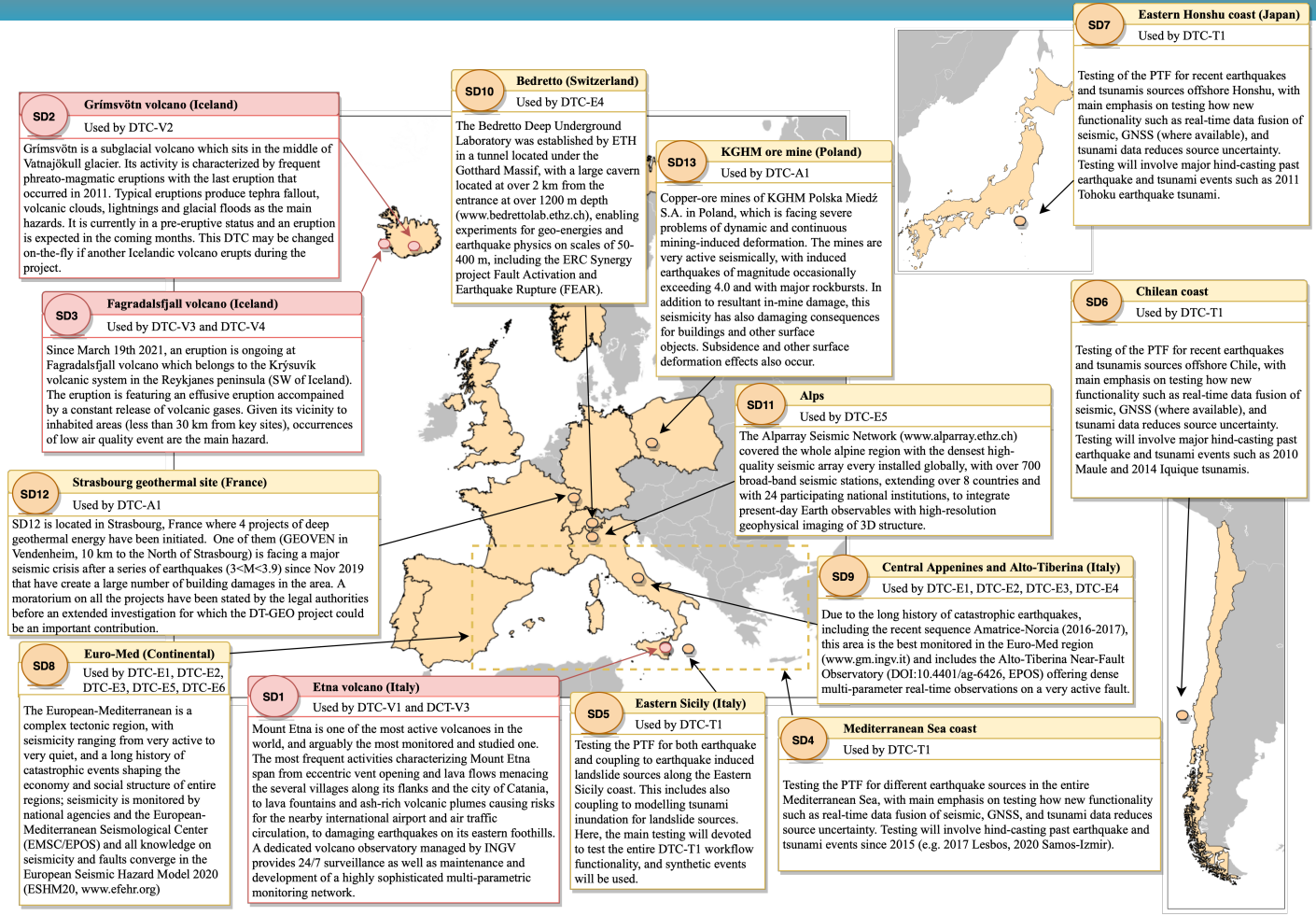
- DTCs are composed by a series of **Building Blocks (BBs)** following a modular approach:
  - A BB can be a physical/AI model execution, a data process, a data assimilation step, etc.
  - BBs are micro-service oriented
  - Containerized
  - Facilitate reusability and interoperability across DTCs
  - Facilitate coupling of the different DTCs in DT-GEO
  - FAIRness and QA



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 Links between eFlows4HPC and DT-GEO

02 The example of DT-GEO DTC-V2 (volcanic ash dispersal forecasts)

03 What’s next?

eFlows4HPC and DT-GEO

Objectives	eFlows4HPC	Deliver a workflow software stack and added value services
	DT-GEO	Deploy a pre-operational DT (model-data fusion)
Coordination	eFlows4HPC	BSC
	DT-GEO	CSIC
Common partners	6	BSC, INGV, ETH, UMA, UPV, NGI
Common goals	Both projects share a community of users (the eFlows natural hazards pillar) and a long-term ambition (the Destination Earth innitiative)	
Synergies	 One-way feedback: DT-GEO inherits part of the eFlows4HPC architecture and software stack	

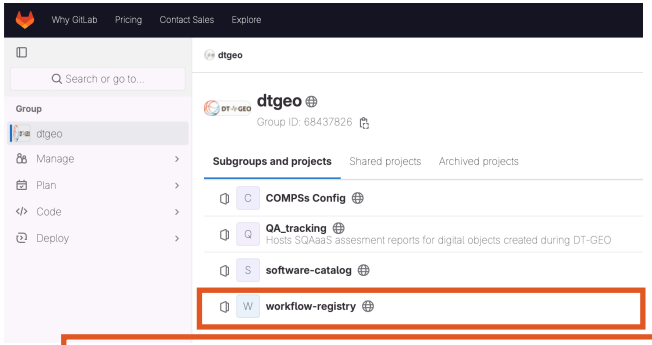
DT-GEO software components (3 layers or levels)

1. Workflow definition, reusability and QA	Workflow registry	git repository hosting workflow descriptions using spack
	Software catalog	git repository hosting software (libraries) using Spack
	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)
2. Workflow deployment	Infrastructure manager	Automated deployment of cloud virtual infrastructures
	Container image creation	Creates container images for target machines
	Container image library	Hosts container images for workflow (DTC) components
3. Workflow execution	COMPSs runtime	Parallel workflow orchestration in distributed systems
	udocker	Executes docker containers in user space
	SLURM	Workload manager (HPC)

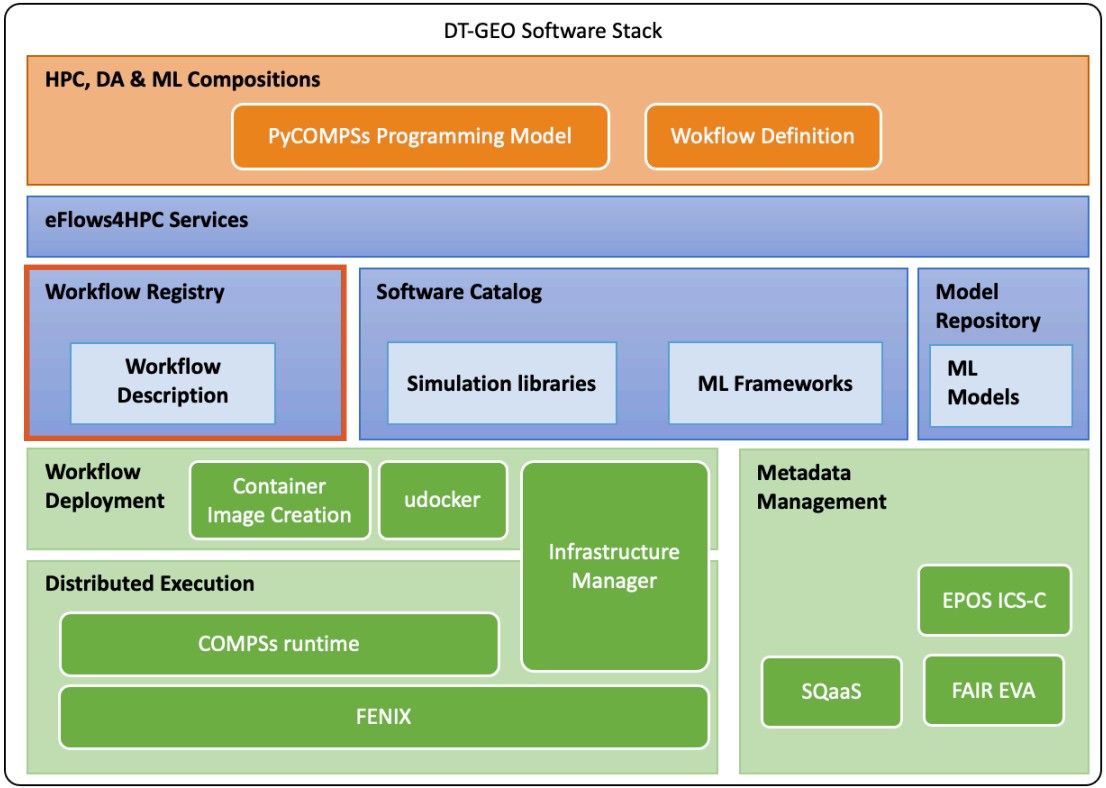
DT-GEO software stack (layer 1)

Workflow definition, reusability and QA

- **Workflow registry**
- Software catalog
- Data catalog
- Model repository
- SQAaaS



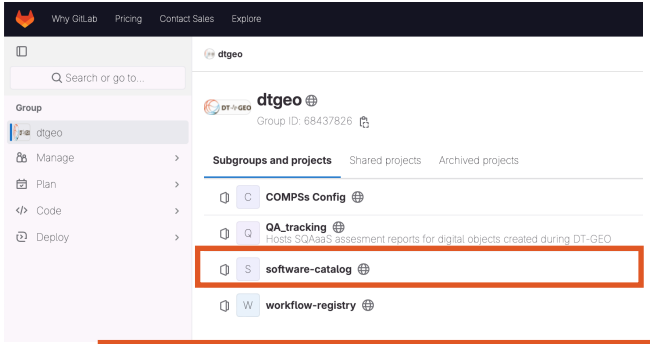

```
workflow-registry
 |- workflow_1
 |    |- tosca
 |    |    |- types.yml      TOSCA description of the different components involved in the workflow
 |    |     ...
 |    |- step_1
 |    |    |- spack.yml      Software requirements for this workflow step as a Spack environment specification
 |   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



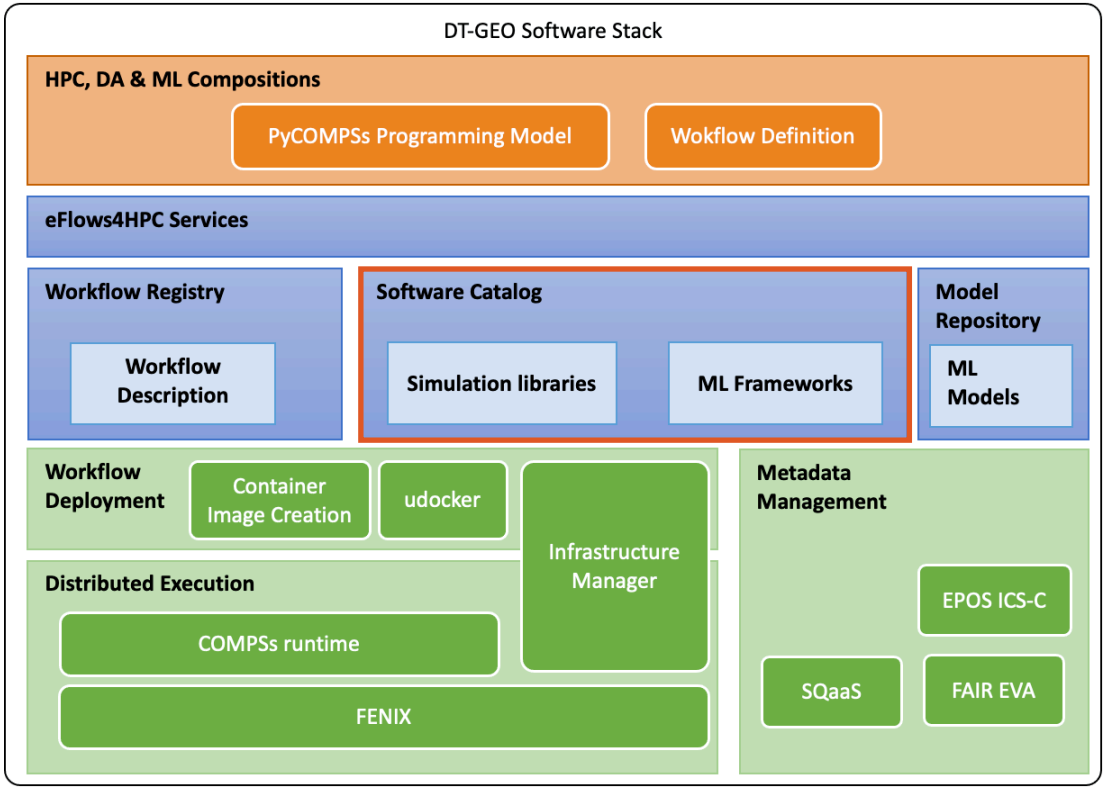
DT-GEO software stack (layer 1)

Workflow definition, reusability and QA

- Workflow register
- **Software catalog**
- Data catalog
- Model repository
- SQAaaS



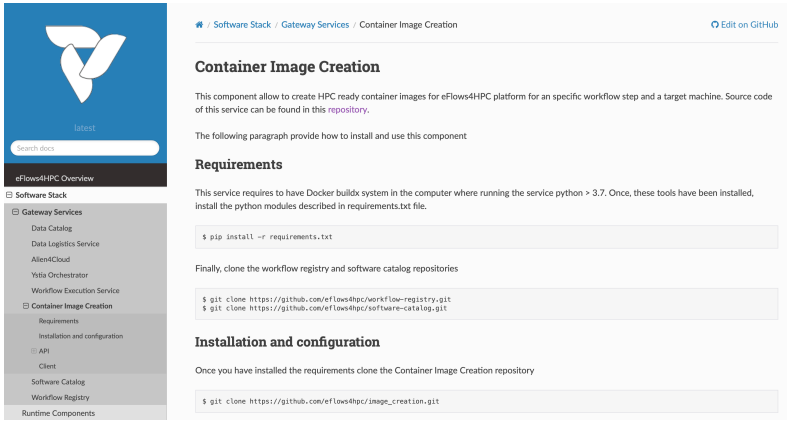

```
software-catalog
 |- packages
 |    |- software_1
 |    |    |- package.py  Installation description following the Spack package format
 |    |    ...
 |    |- software_2
 |         ....
 |- cfg       Spack configuration used by the Image Creation Service
 |
 |- repo.yaml    Spack description of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

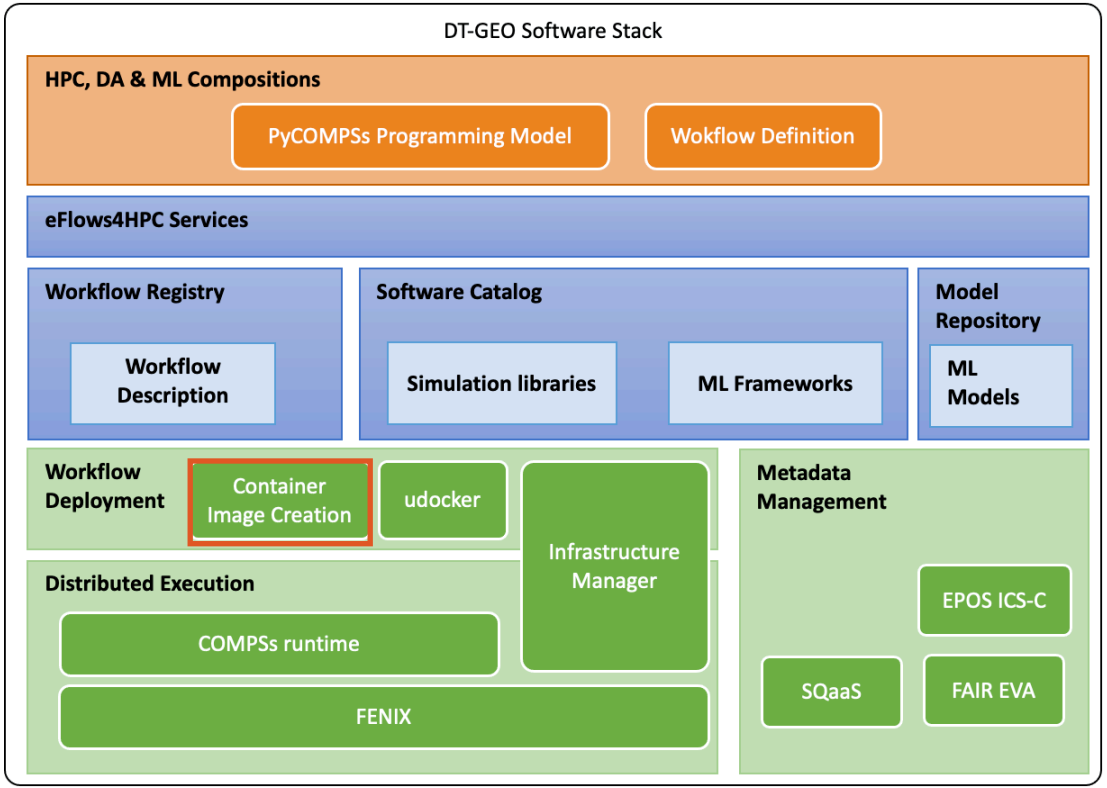


DT-GEO software stack (layer 2)

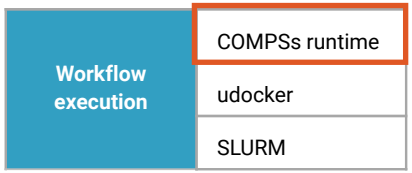
Workflow deployment	Infrastructure manager
	Container image creation
	Container image library

The Container Image Creation service offers a web dashboard and a REST API to manage the creation of container images

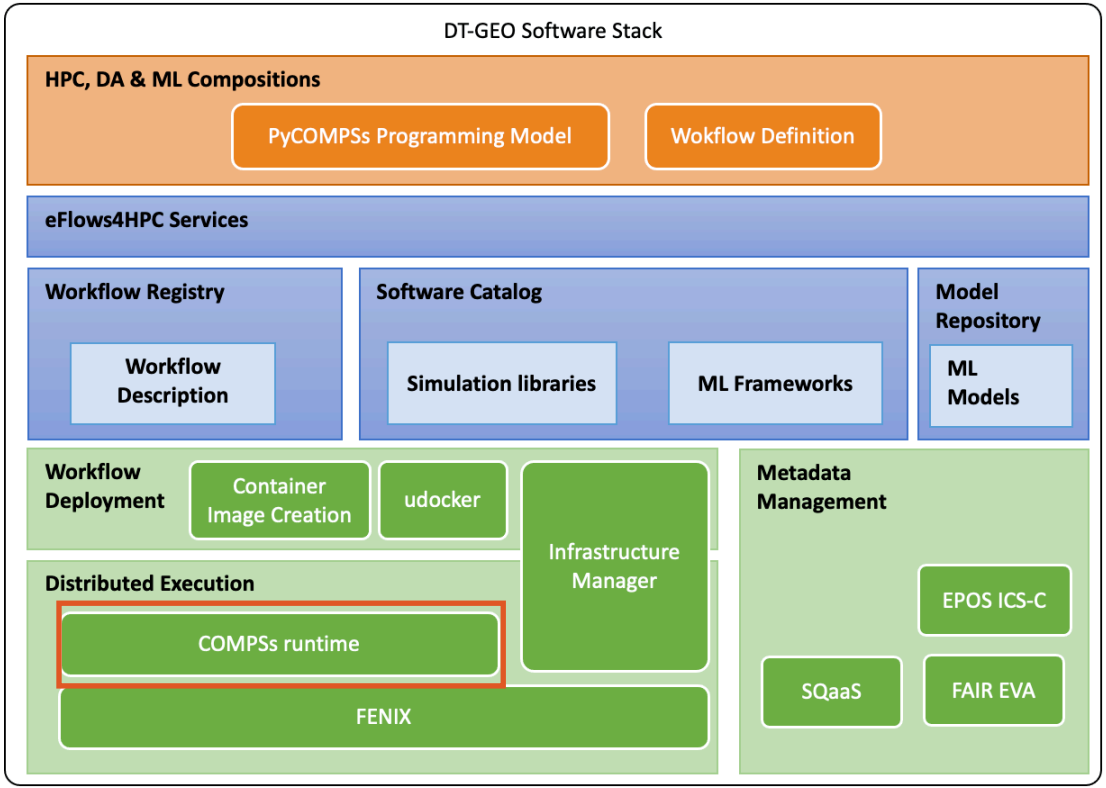




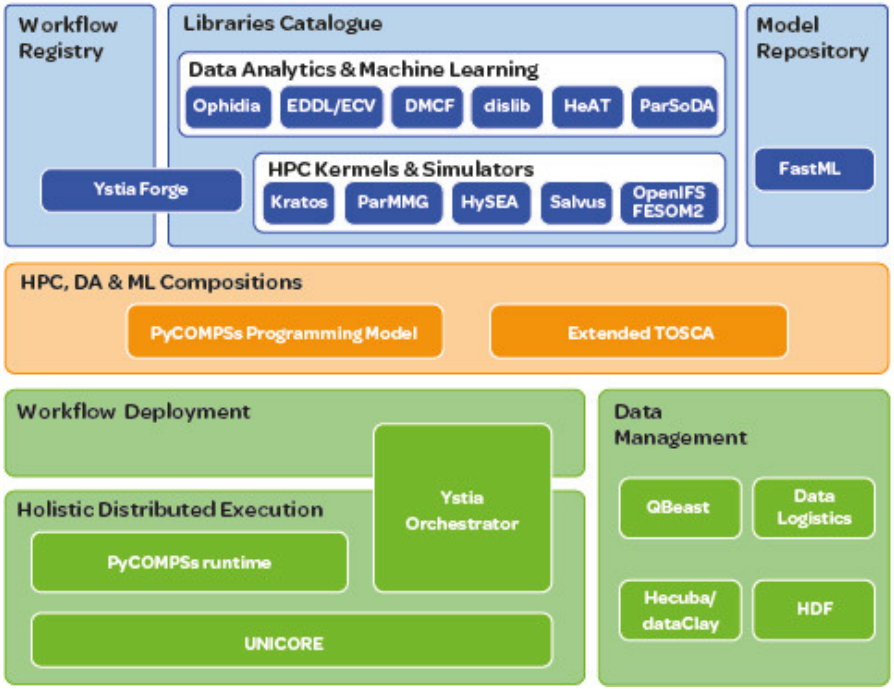
DT-GEO software stack (layer 3)

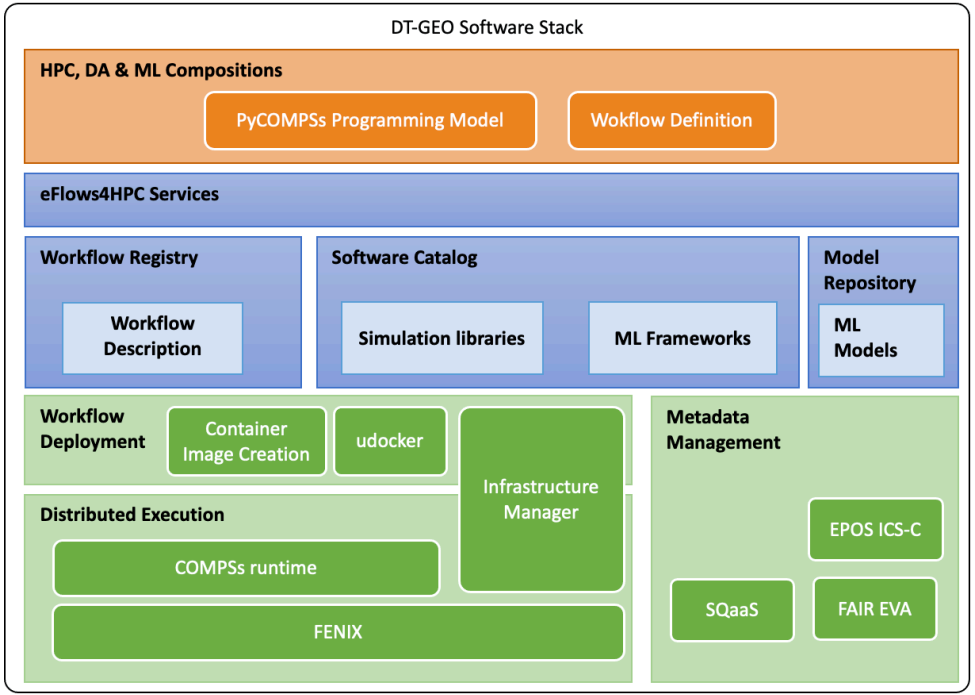


COMPSs runtime	
Task Dependency Analysis	Builds a task dependency graph
Task Scheduling	Scheduled in distributed resources
Resource Management	For cloud environments, ellastically adapt resources
Job and data Management	Perform remote execution of tasks and the data transfers



eFlows4HPC versus DT-GEO



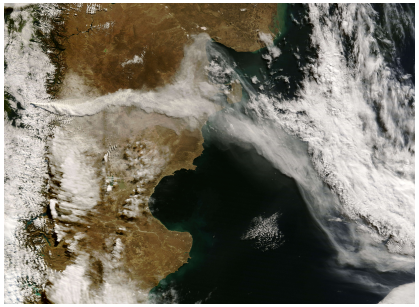


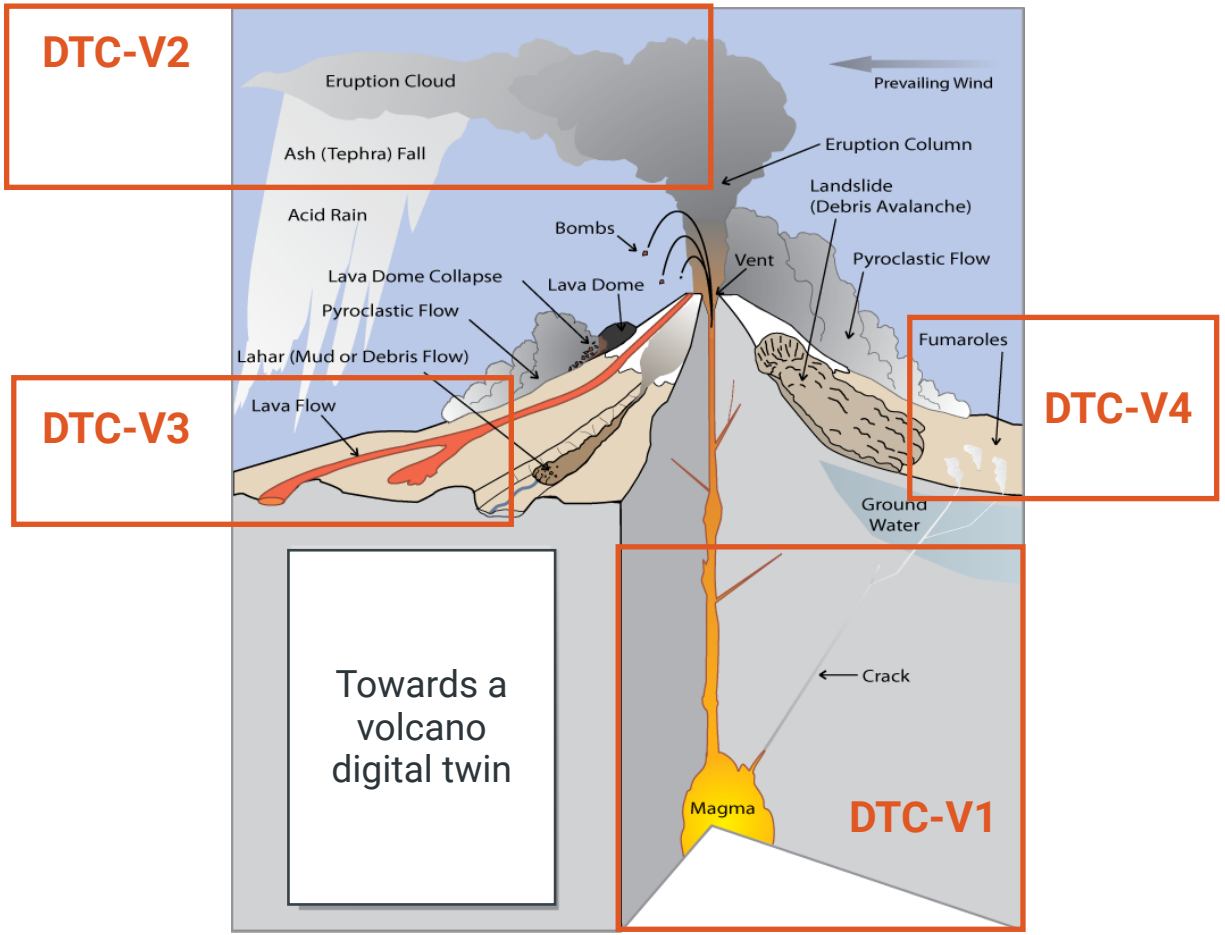
01 Links between eFlows4HPC and DT-GEO

02 The example of DT-GEO DTC-V2 (volcanic ash dispersal forecasts)

03 What’s next?

The DT-GEO volcano digital twin



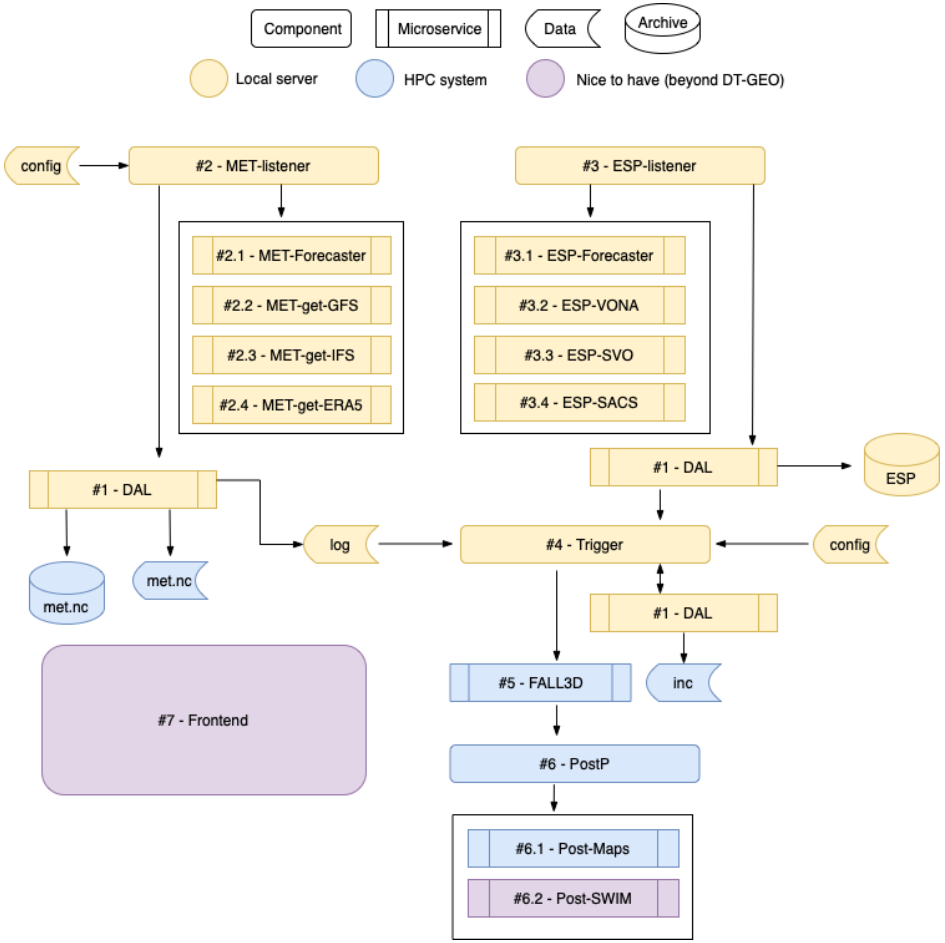


The DTC-V2 workflow and Building Blocks

DTC-V2 objective

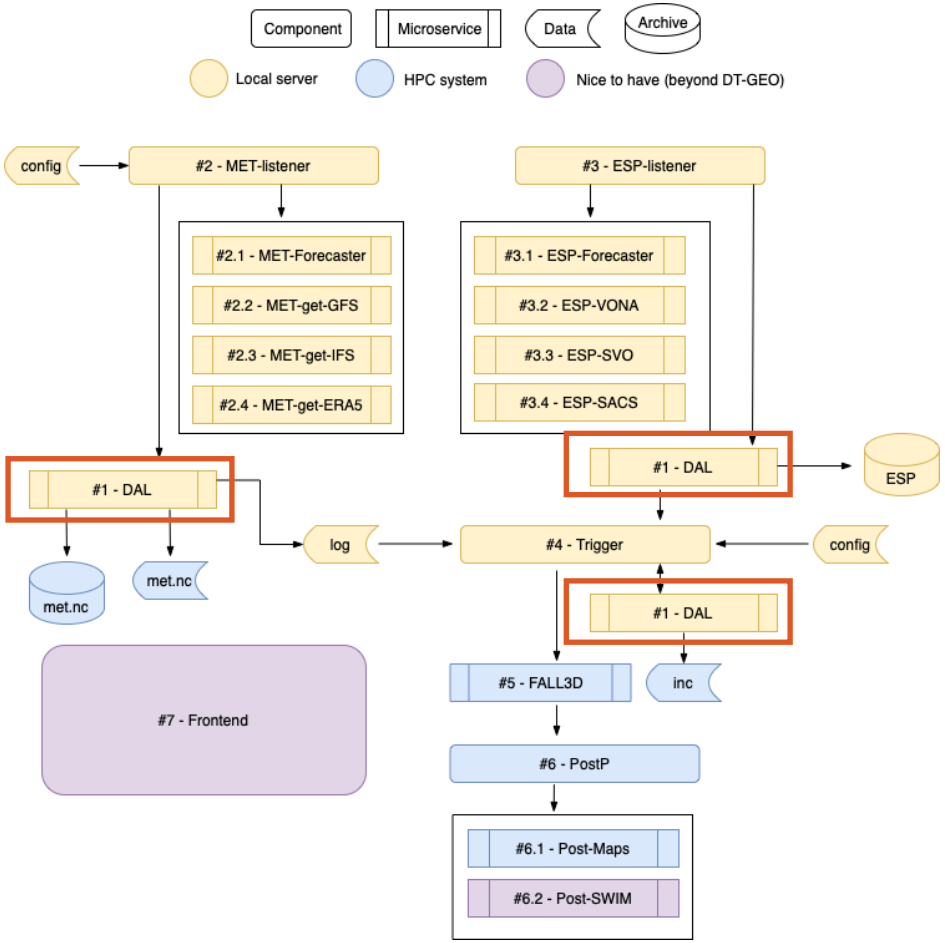
Merge real-time ground-based and satellite observations with the FALL3D model to generate ensemble-based deterministic and probabilistic volcanic ash forecast maps and products

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



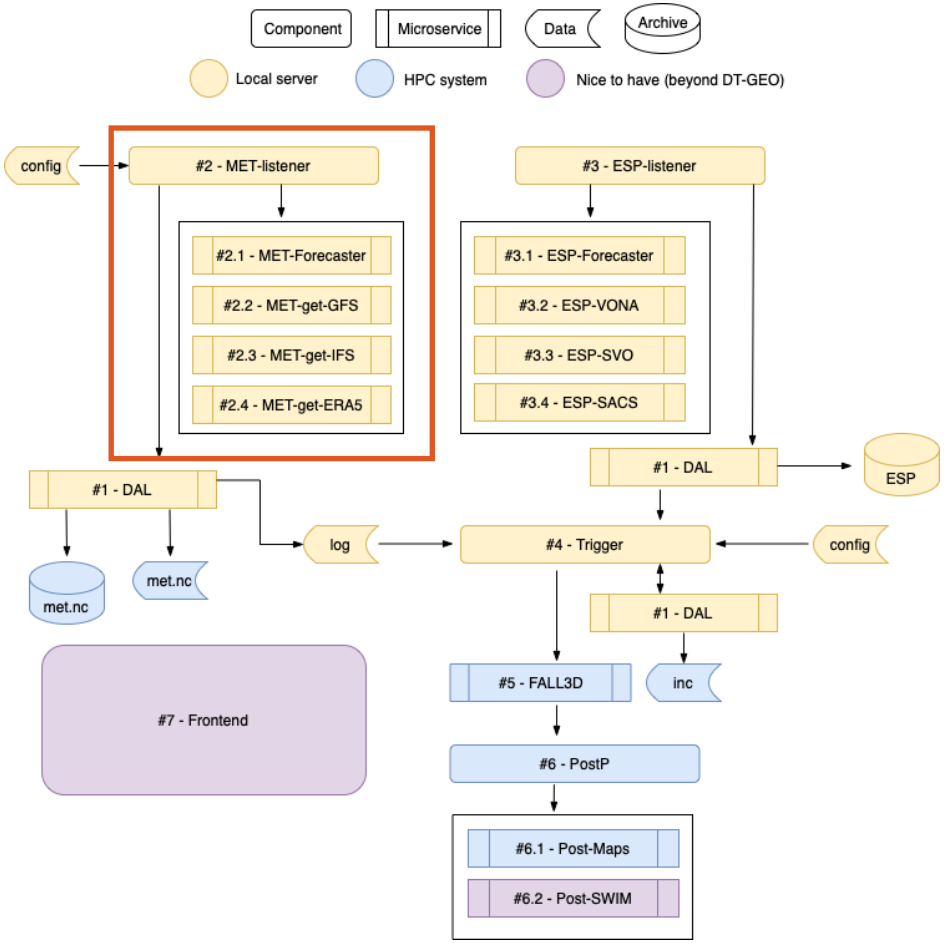
The DTC-V2 workflow and Building Blocks

BB number 1	Data Access Layer (DAL)
Type	Microservice. Python script
Deployment	Local server
Description	• Track and interact with intermediate information stored in a database or in a json log file • Manage data transfer and authentication • Manage local storage processes (archive) • Interact with the Frontend component (dashboard) for real-time update



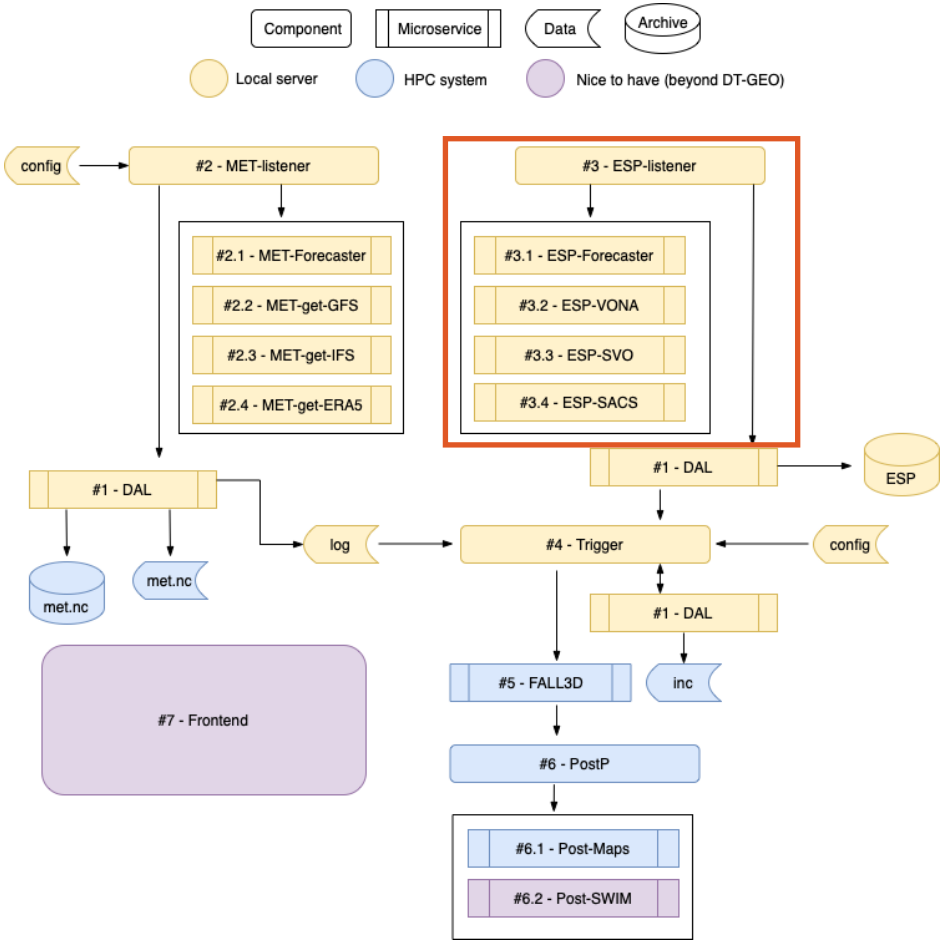
The DTC-V2 workflow and Building Blocks

BB number 2	MET-listener
Type	Set of micro services
Deployment	Local (can also be deployed at HPC system)
Description	• Activated at specific times (e.g. daily at 06:00 am). • Runs independent of the rest of the workflow • Calls a given MET-get micro service among different options • Invokes DAL to upload the met file data to the HPC system(s) and, eventually, archive met data locally



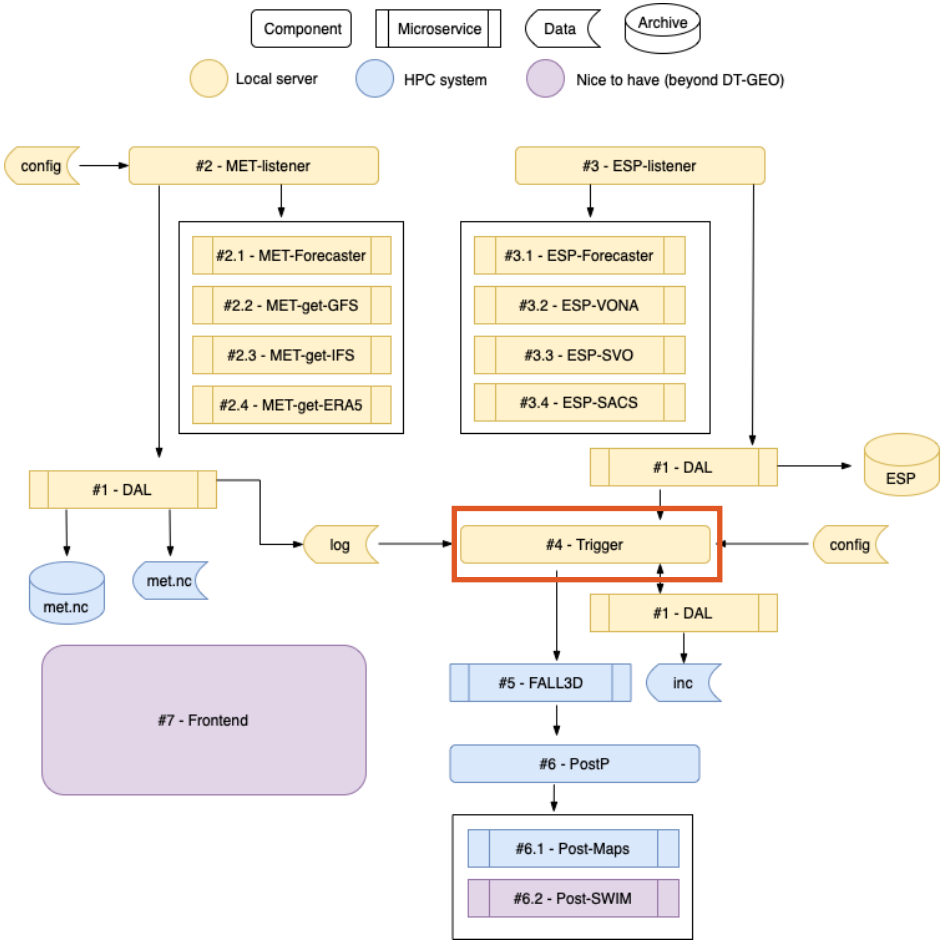
The DTC-V2 workflow and Building Blocks

BB number 3	ESP-listener
Type	python script
Deployment	Local
Description	• Permanently running in the background (daemon) • Queries different data sources to check the start (or evolution) of an eruption and to get the so-called **Eruption source Parameters** (ESP) • In case of new info, invoke DAL to archive ESP data locally and update the status • Invokes the Trigger workflow component (BB#4)



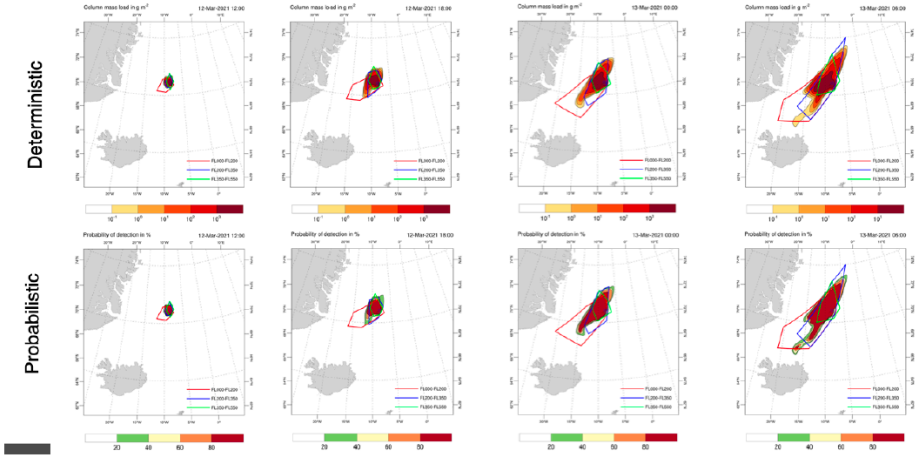
The DTC-V2 workflow and Building Blocks

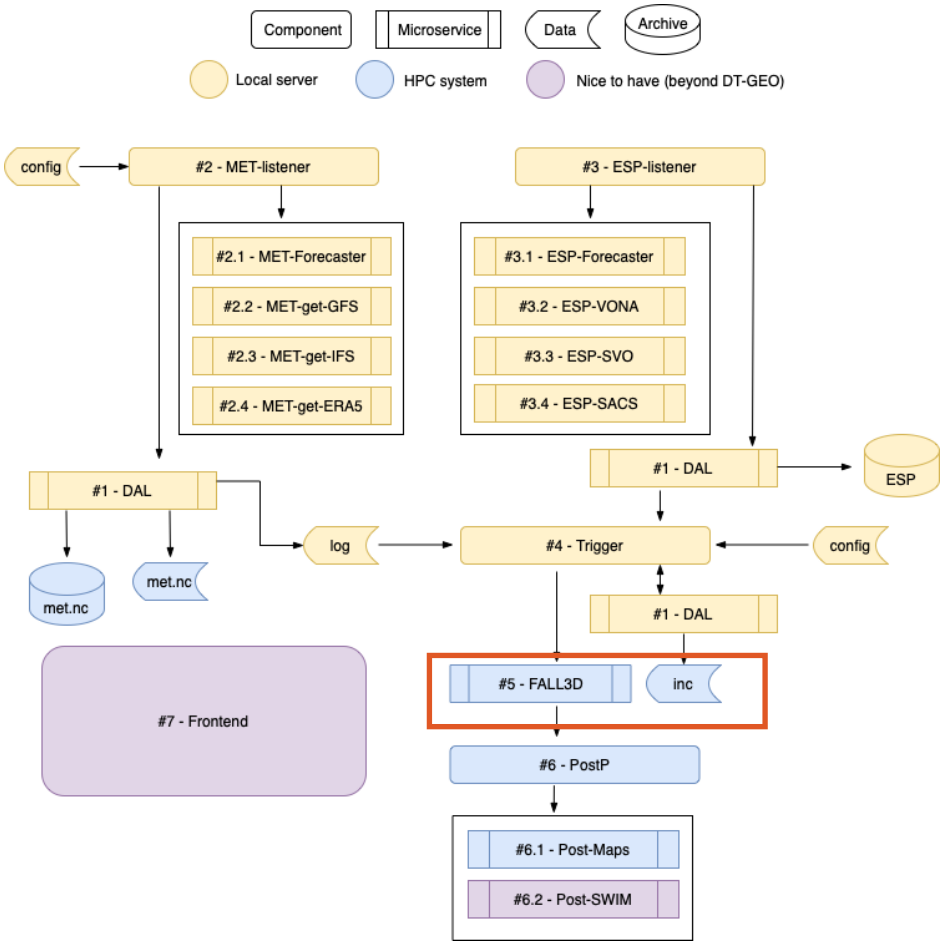
BB number 4	Trigger
Type	Python script
Deployment	Local
Description	• Activated when new information from ESP-Listener exists • Contains all the logics to decide if a new forecast has to be launched (first time), re-launched (new relevant data available), re-started (e.g. long-lasting events) • Invoke DAL to upload the FALL3D input file to the HPC system(s) and run FALL3D



The DTC-V2 workflow and Building Blocks

BB number 5	FALL3D
Type	microservice. HPC code (pure MPI or MPI+OpenACC)
Deployment	HPC
Description	• Runs a FALL3D ensemble and related auxiliary programs





01 Links between eFlows4HPC and DT-GEO

02 The example of DT-GEO DTC-V2 (volcanic ash dispersal forecasts)

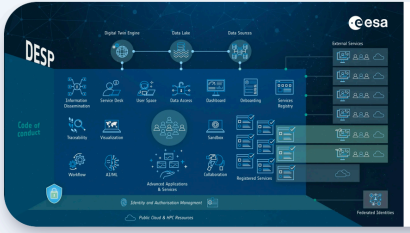
03 What's next?

The future of the DTCs beyond DT-GEO

DTC	Code	Name	DestinE Twin	DestinE use case (DESP)	EPOS
1	DTC-V1	Volcanic unrest dynamics			
2	DTC-V2	Volcanic ash clouds and deposition			
3	DTC-V3	Lava flows			
4	DTC-V4	Volcanic gas dispersal and deposition			
5	DTC-T1	Probabilistic Tsunami Forecasting (PTF)			
6	DTC-E1	Probabilistic Seismic Hazard and Risk Assessment			
7	DTC-E2	Earthquake short-term forecasting			
8	DTC-E3	Tomography and Ground Motion Models (GMM)			
9	DTC-E4	Fault rupture forecasting			
10	DTC-E5	Tomography and shaking simulation			
11	DTC-E6	Rapid event and shaking characterization			
12	DTC-A1	Anthropogenic geophysical extreme forecasting (AGEF)			

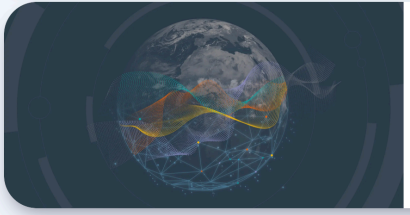
Towards Destination Earth

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



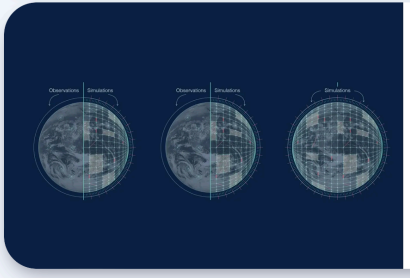
Core Service Platform

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.

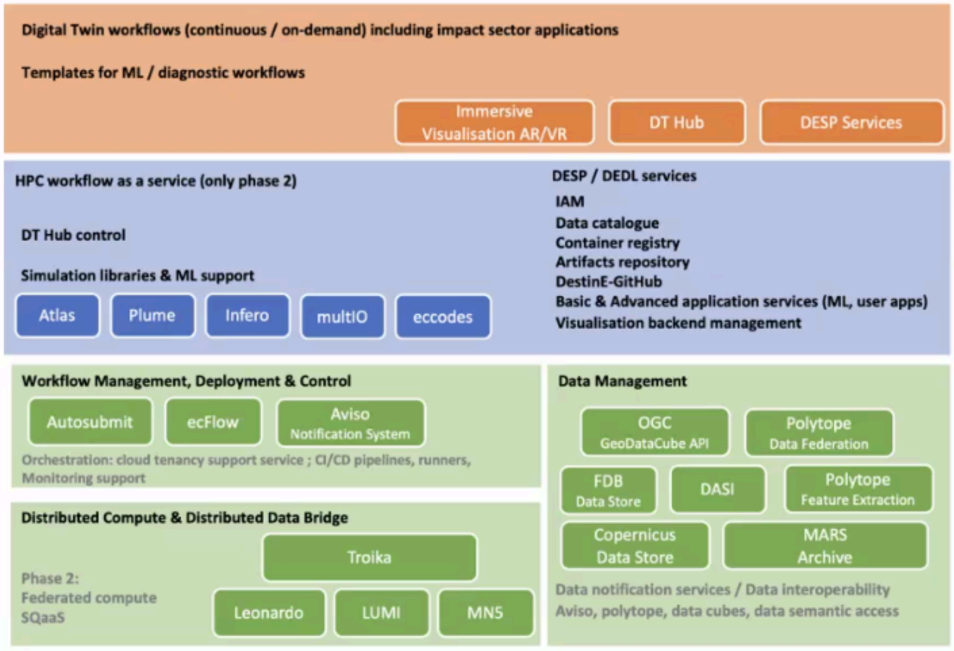
Other DTCs will follow later on (although not all the DTCs in DT-GEO target at DestinE)

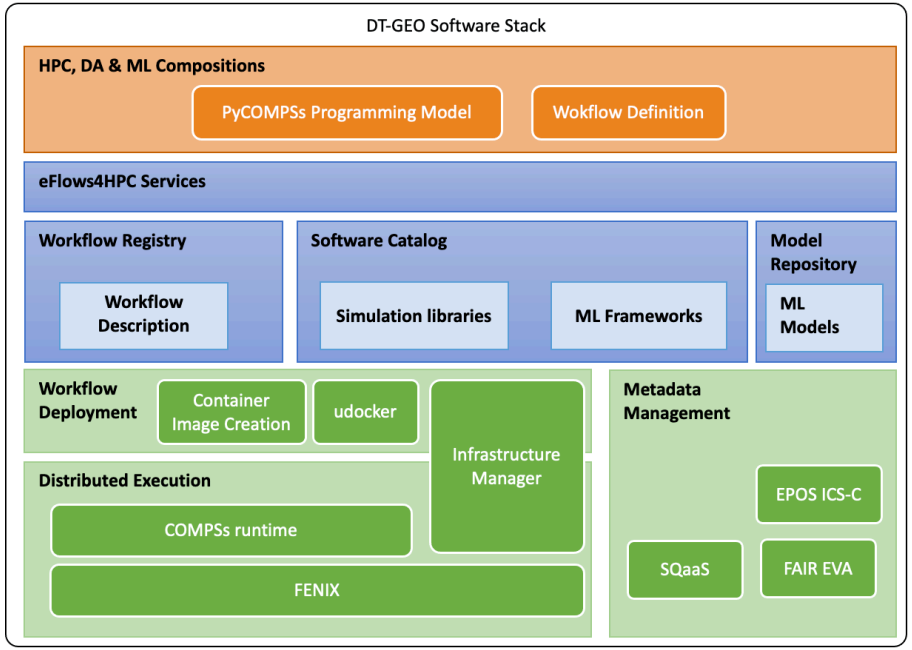
Towards Destination Earth

- DT-GEO proposed 3 high-TRL DTCs focussed on **Urgent Computing** as proofs of concept for earthquakes, tsunamis and volcanoes (others may follow) based on their high TRL

Code	Name	Concept	Coupling with other DestinE twins	
			Level	Mechanism
DTC-V2	Volcanic ash clouds and deposition	Merge real-time observations with the FALL3D dispersal model to generate **ensemble-based deterministic and probabilistic forecasts**	Strong	One-way, dispersal model driven on-line from the databridge (atmosphere)
DTC-T1	Probabilistic Tsunami Forecasting (PTF)	Faster than real-time probabilistic **forecast of tsunami inundation** following an earthquake, exploiting both real-time data and long-term seismotectonic knowledge	Weak	One-way, ocean data assimilated into the tsunami model
DTC-E5	Tomography and shaking simulation	G**enerate ground shake-maps for the most relevant seismic sources**; as new quakes are recorded	Decoupled	Could serve to pioneer the exposure/vulnerability data structure/interface (across all twins!)

Similar architectures (tool choice differ)





Foreseen couplings with other twins

DTC	Hazard	Code	DestinE (phase 1)		DestinE (future)			
			climate	weather	InterTwin	BioDT	Ocean	Anthropogenic?
1	Volcano	DTC-V1	freatomag.		?			
2		DTC-V2			?		fertilisation	
3		DTC-V3			?			
4		DTC-V4			?			
5	Tsunami	DTC-T1		meteotsun.	?			
6	Earthquake	DTC-E1	seis. activity		?			
7		DTC-E2			?			
8		DTC-E3			?			
9		DTC-E4			?			
10		DTC-E5			?			
11		DTC-E6			?			
12	Anthropogenic	DTC-A1	CO2 seq.		?		off-shore	

THANK YOU

@dtgeo_eu

linkedin.com/company/dt-geo/

This project has received funding from the European Union's Horizon research and innovation programme under the grant agreement No 101058129