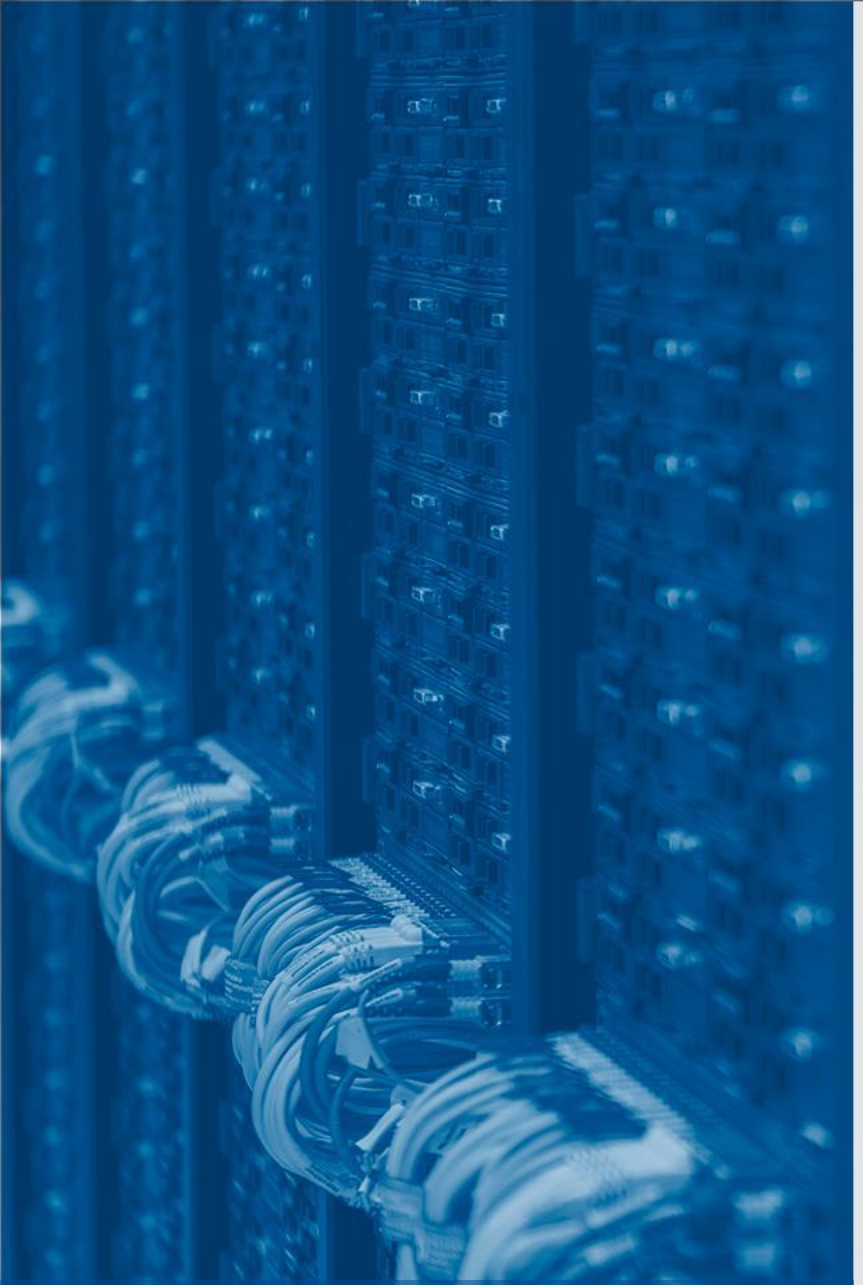


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth & European Digital twin of the Ocean

Miguel Castrillo, Bruno Kinoshita

17 October 2023

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Funded by the European Union

eFlows4HPC
www.eFlows4HPC.eu

Outline

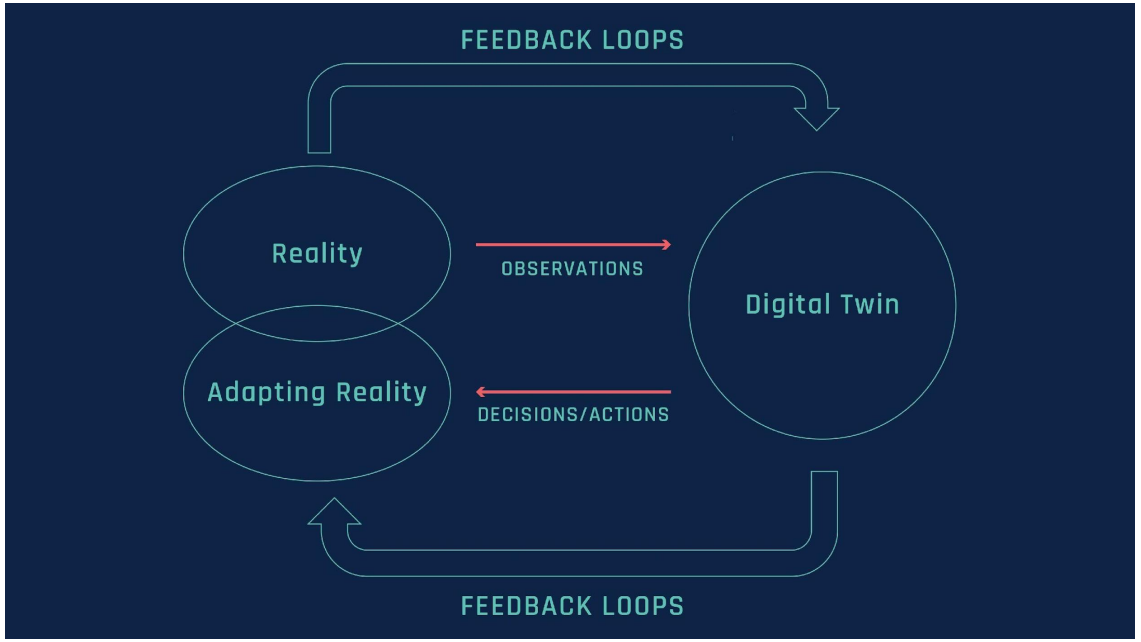
- Destination Earth
- European Digital Twin of the Ocean (DTO)
- Digital twins of the Earth
- Climate DT (DestinE)
- EDITO (European DTO)
- Infrastructure Backend / Workflow managers
- Conclusions

Destination Earth

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Destination Earth (DestinE), a European Commission flagship initiative for a sustainable future



Perform highly accurate, interactive and dynamic simulations of the Earth system, informed by rich observational datasets.

Improve prediction capabilities to maximise impact.

Support EU policy-making and implementation.

Exploit the potential of distributed and high-performance computing (HPC) and data handling at extreme scale.

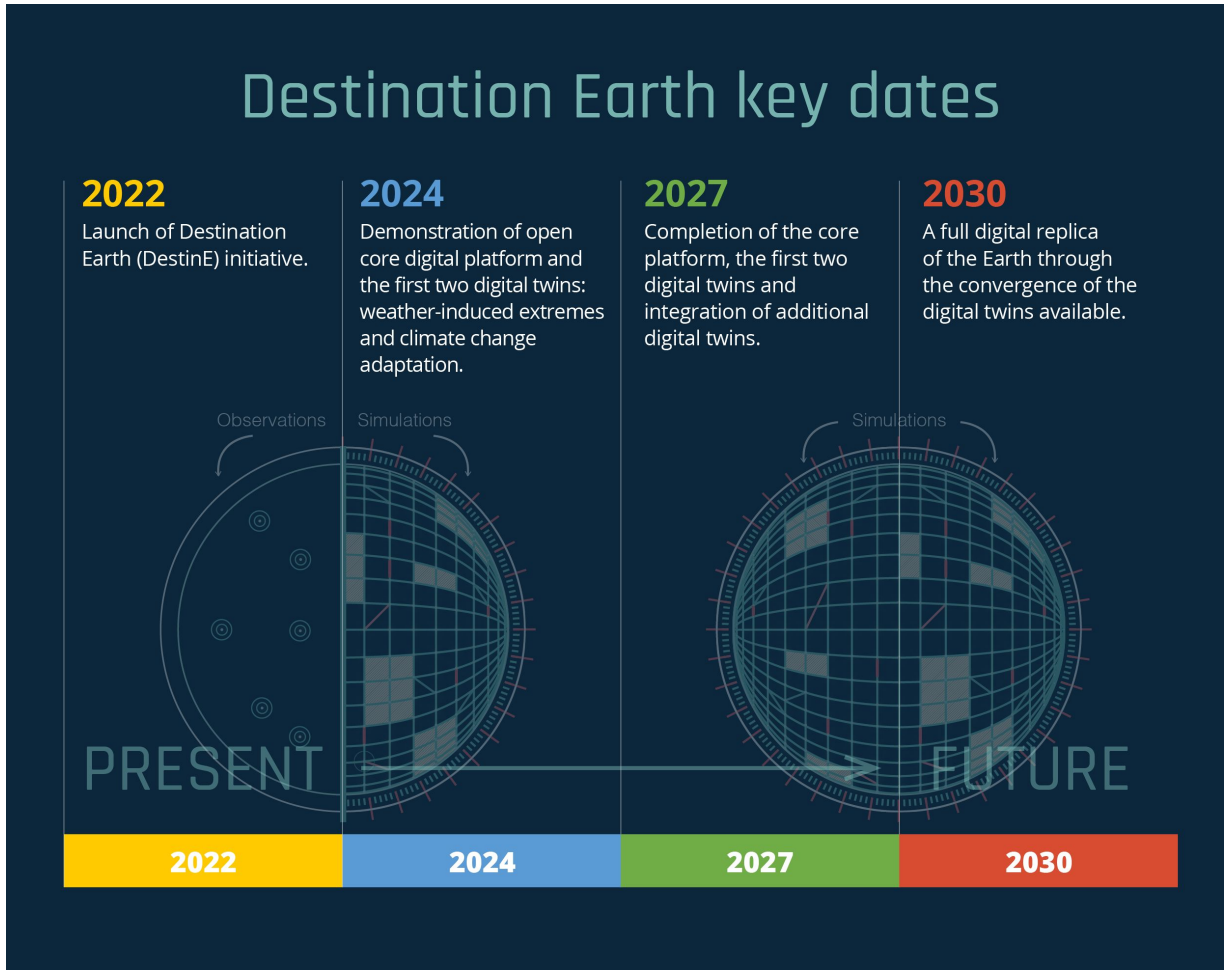
https://digital-strategy.ec.europa.eu/en/policies/destination-earth

https://stories.ecmwf.int/explainer-digitaltwins/index.html

Funded by the European Union

Destination Earth implemented by ECMWF esa EUMETSAT

Destination Earth timeline



Funded by the European Union
Destination Earth
implemented by
ECMWF
esa
EUMETSAT

European Digital Twin of the Ocean

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

European Digital Twin of the Ocean

Aims to model the ocean's multiple components, provide knowledge and understanding of the past and present and create trustable predictions of its future behaviour.

The Digital Twin Ocean is a place of digital co-creation, bringing together different disciplines and communities.

Builds on CMEMS, DIAS and EMODnet, connects them with similar systems focused on inland waters, and further integrates the whole knowledge value chain.

Core DTO as a baseline, a huge bulk of data, generic ocean models and AI processors as toolboxes, on top of which a multitude of tailor-made applications, or 'local twins' can be plugged in.

https://digitaltwinocean.mercator-ocean.eu/

https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/restore-our-ocean-and-waters/european-digital-twin-ocean-european-dto_en

Digital Twins of the Earth

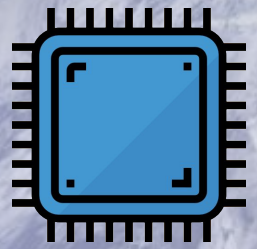
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Earth Digital Twins' components

Digital Twin Engine

Based on state-of-the-art simulations and observations.

Made up from different components and twins.



Data lake

Including data from diverse sources.

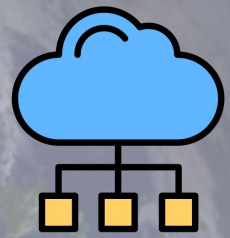
Discovery and data access.

Data processing in the cloud.

Service platform

Providing decision-making tools, applications and services, including visualization and interactivity.

Based on cloud-based computing infrastructure.



Climate DT

A Destination Earth Digital Twin

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CLIMATE DT TEAM – 13 ORGANISATIONS

CSC

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

MAX-PLANCK-INSTITUT FÜR METEOROLOGIE

UNIVERSITY OF HELSINKI

AWI ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

ISAC

Politecnico di Torino

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Deutscher Wetterdienst
Wetter und Klima aus einer Hand
DWD

HELMHOLTZ

UCLouvain

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

Hewlett Packard Enterprise

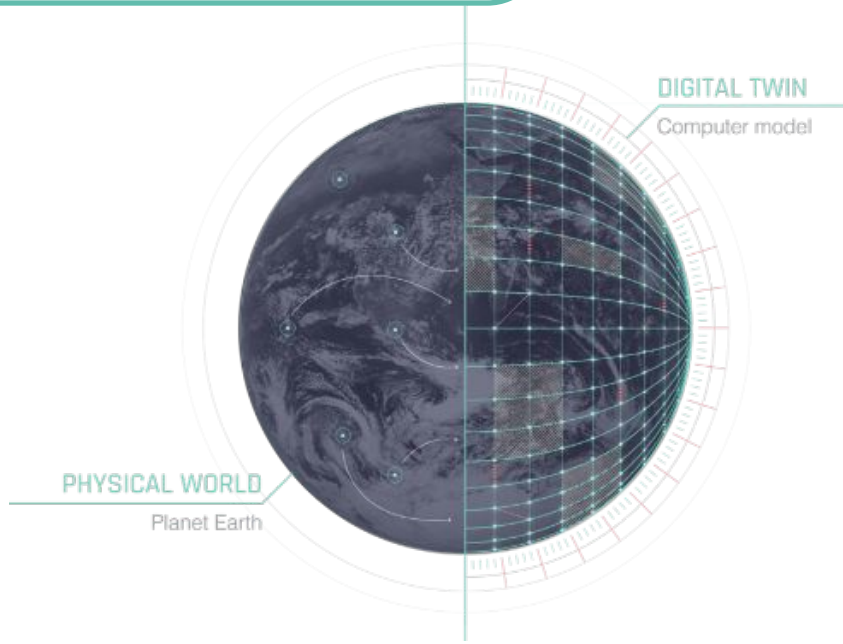
ECMWF

CLIMATE ADAPTATION DIGITAL TWIN (CLIMATE DT)

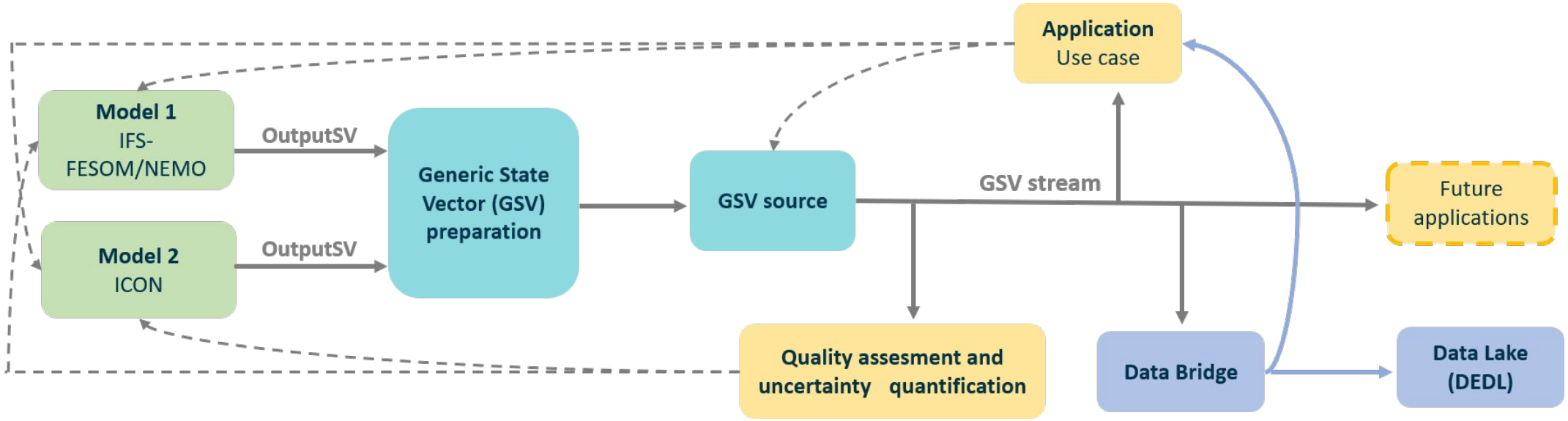
Climate DT is a new type of climate information system that focuses on **assessing the impacts of climate change and different adaptation strategies** at local and regional levels over multiple decades using a strategy where **user requests drive the whole production chain.**

Climate DT encompasses

- **Global climate simulations** at an unprecedented horizontal resolution
- Novel approach with **streaming of climate model output to impact models**
- **Quality assessment and uncertainty quantification** based on observations
- Deployment on **two European pre-exascale supercomputers** (LUMI and MareNostrum5)
- **Integration of all relevant European research** (Horizon programmes, national, private).



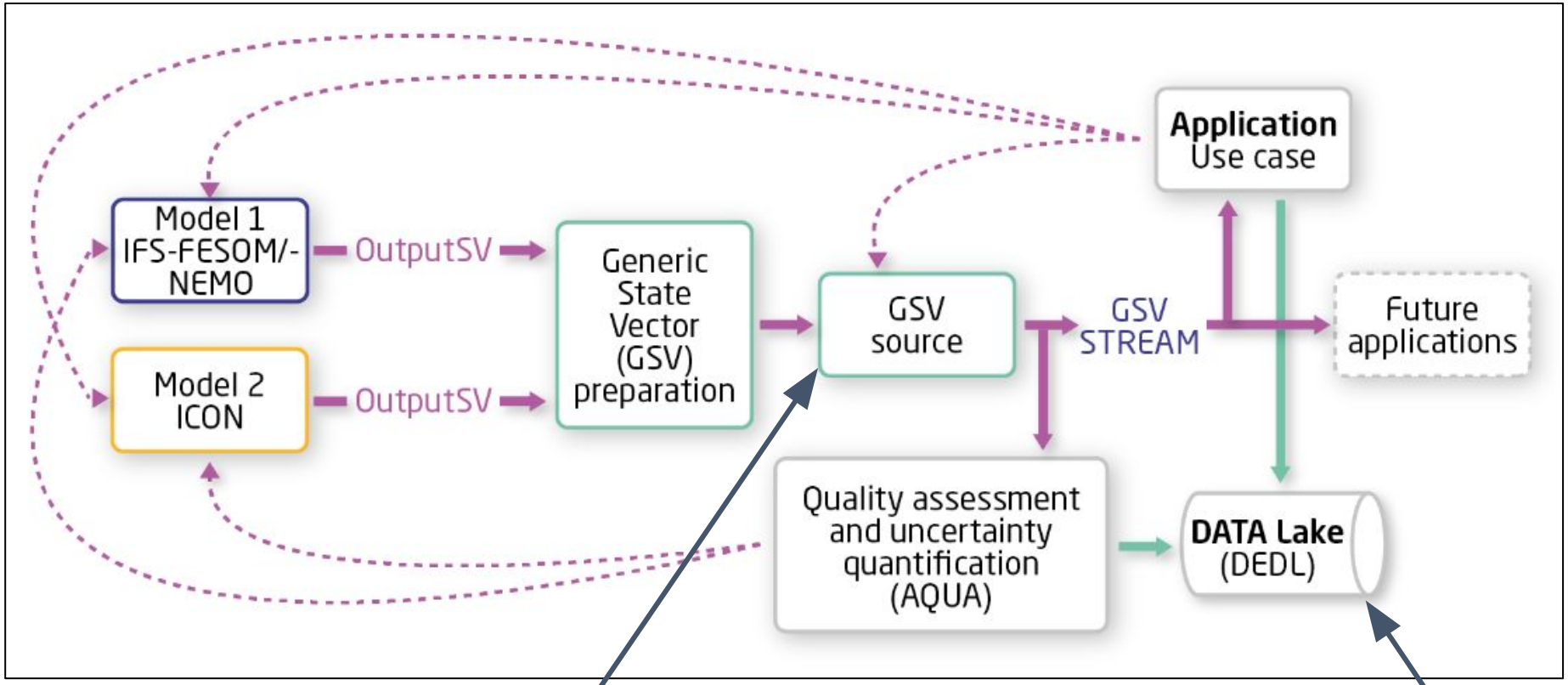
CLIMATE SIMULATION WORKFLOW RETHOUGHT



Streaming of climate model output in a standardized form (*generic state vector, GSV*) enables

- **users to access the full model state** as soon as it is produced
- **interactivity** – development to allow simulations and variables on demand in the next DestinE phase
- **scalability** – new applications and requirements can be added

4) The workflow: climate models and data consumers



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Funded by the European Union
Destination Earth
implemented by ECMWF esa EUMETSAT

EDITO

A European Digital Twin of the Ocean

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

European Digital Twin Ocean
EDITO Model Lab

A consortium based on ocean modeling expertise

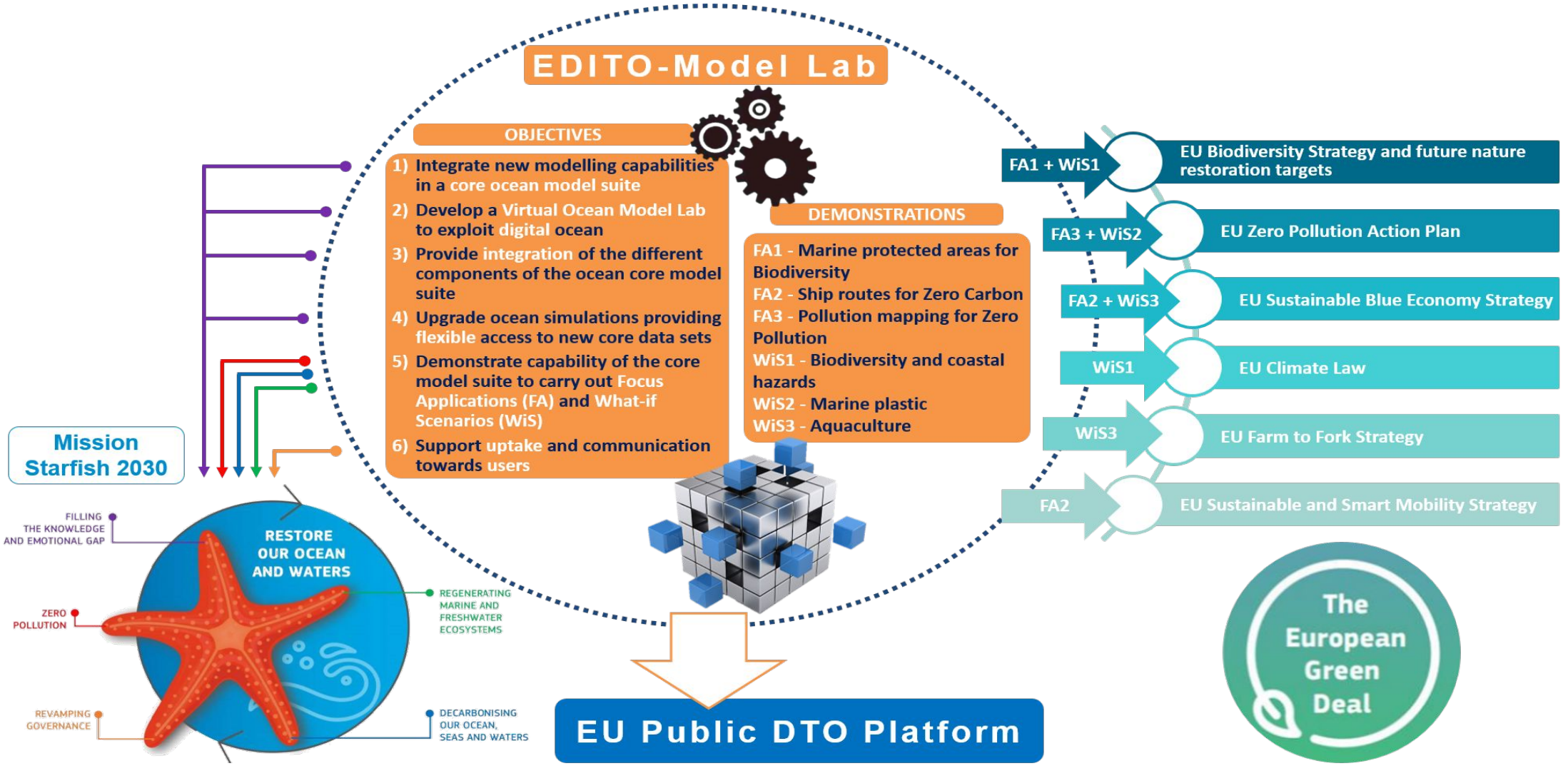
7M€
3-year project
Starting date January 2023
Kickoff meeting 21-22 Feb 2023

14 partners from 9 countries with expertise in :

- **Ocean modeling** from global scale to coastal, for ocean physics, biogeochemistry and marine environment
- **Supercomputing** including experts from computing centers and GPU conceptor
- **Artificial Intelligence** applied to ocean application
- **Software development,** model and tools co-development
- **Operational oceanography** with strong links with Copernicus Marine, Ocean Predict and UN decade
- Intermediate to final **User applications**

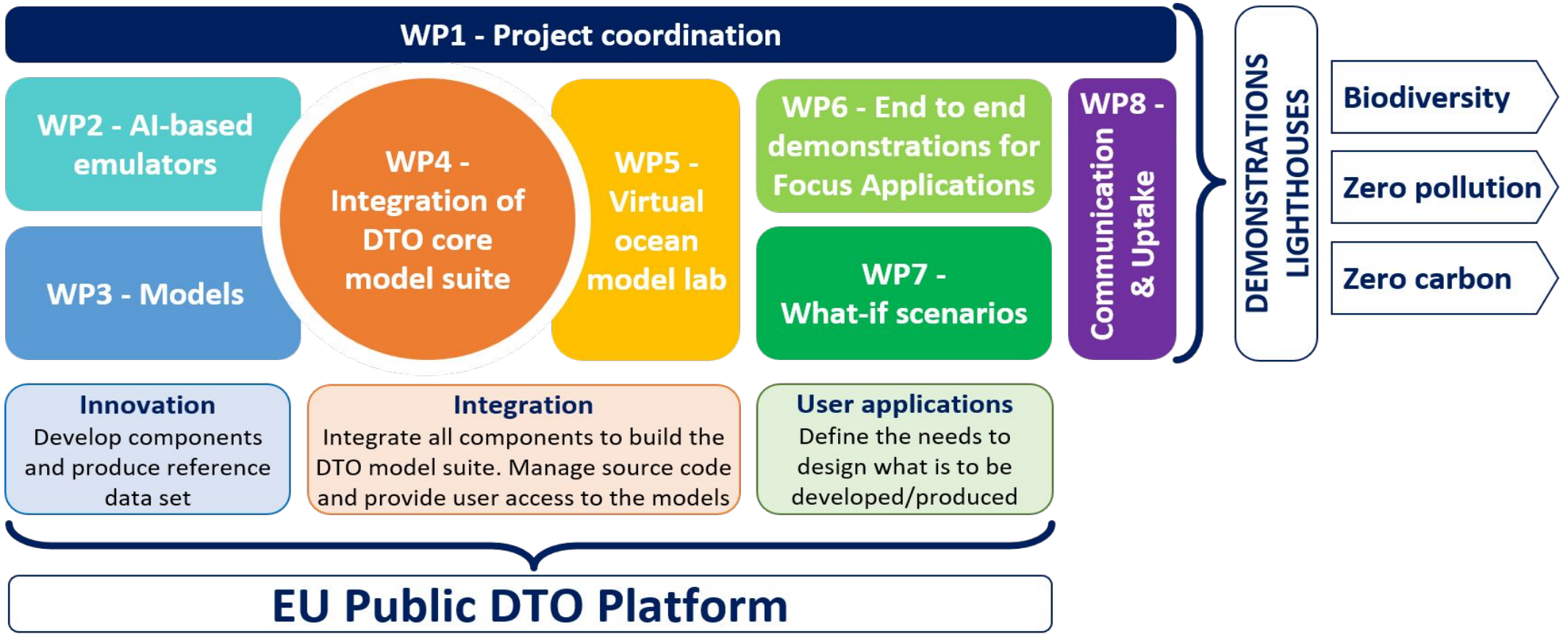
EDITOModelLab

Contributing to the European Green Deal



Yann Drillet (MOi)

EDITO Model-Lab - project organisation



Yann Drillet (MOi)

Infrastructure backend

Workflow manager

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

(Workflow) Infrastructure requirements

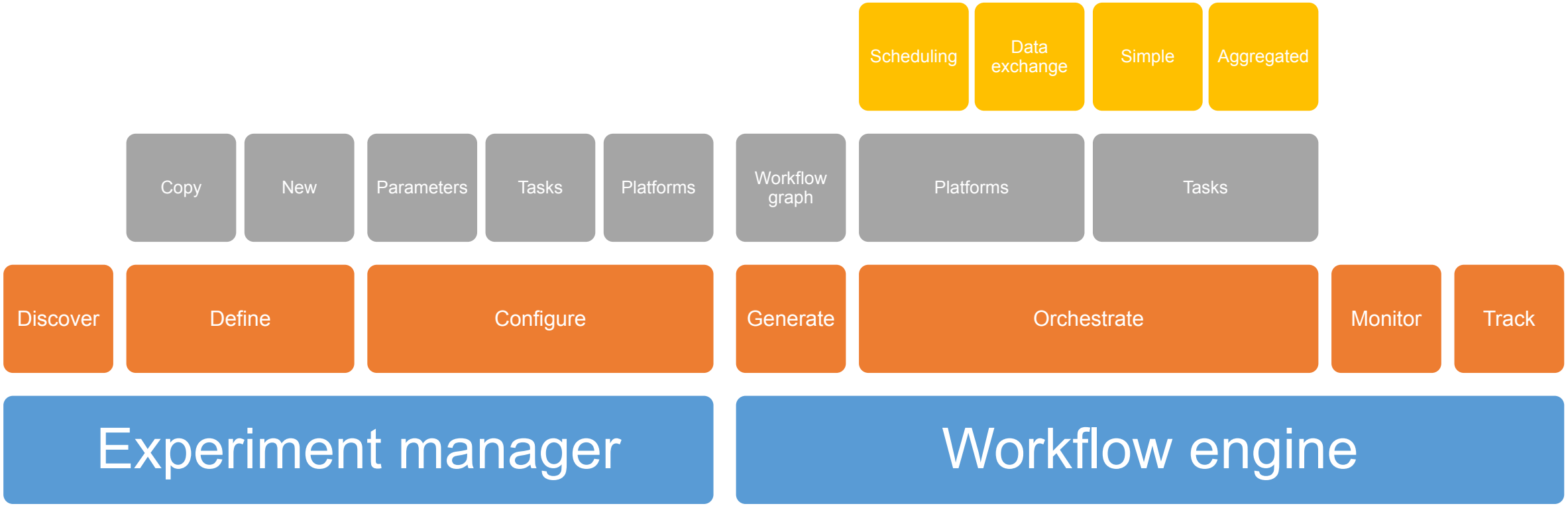
- **Co-development** → Discover and replicate experiments
- **Model-agnostic** experiment **configuration**
- **Hide complexity** → Scientific language interface

Experiment manager

- Automatic and reproducible workflow **orchestration**
- **Portability** → Python, better if containerized
- **Monitorization**
- **Robustness and efficiency** in using **shared resources** →Job aggregation
- **Interactivity**

Workflow manager

Workflow manager



The workflow manager

AUTOSUBMIT

The Autosubmit workflow manager has been designed to meet climate research necessities. It supports workflows based on different hierarchical levels (once, startdate, member, chunk) and provides multiple features developed after years of operation on climate investigation.

Automatization

High-level workflow definition

Based on task **dependencies**

Seamless communication with remote job **scheduler**

Automatic **retries** in case of error

Multi-platform

Combines different **platforms** and **partitions** in the same workflow

Centralized **user authentication**

Efficiency

Possibility to refine granularity (**chunks**)

Wrap individual tasks in bigger allocations to better suit running rules

Robustness

Scalable database

Use **multiple login** nodes

Auto-recovery (after network or filesystem issues)

Mail **notifications**

Full **traceability**

Portability

Python tool (container available)

Shared database across platforms with option to copy experiments

Web-based **GUI**

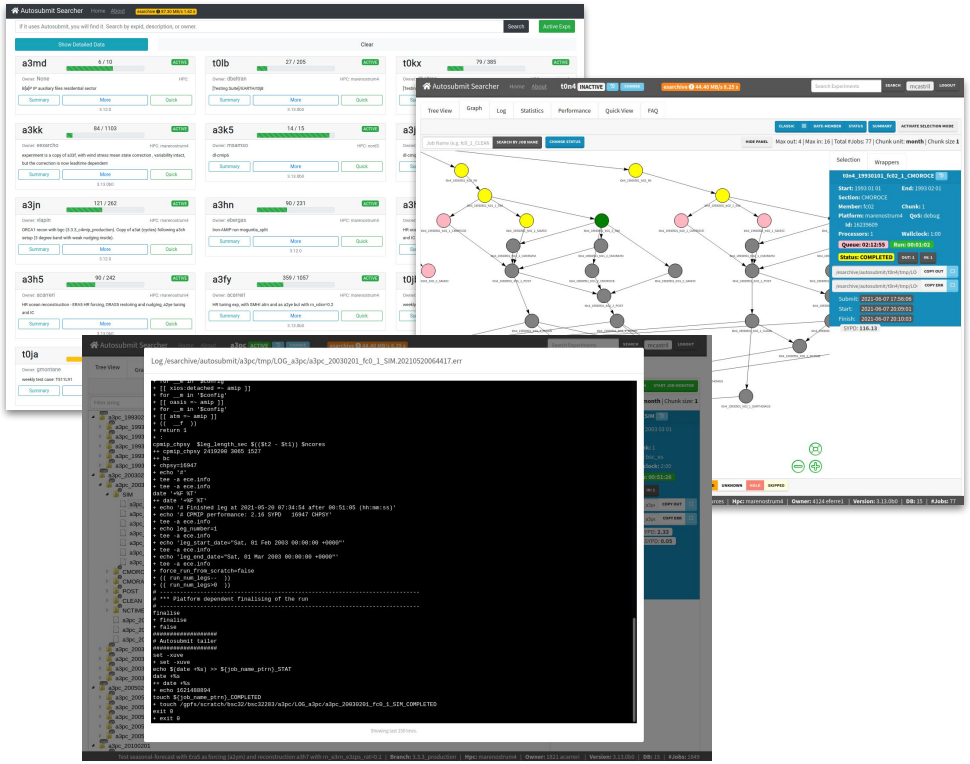
Monitoring

Real time workflow status

Unique end point to all jobs with polling method

Access **job log files** from the GUI

Stats and **performance metrics**

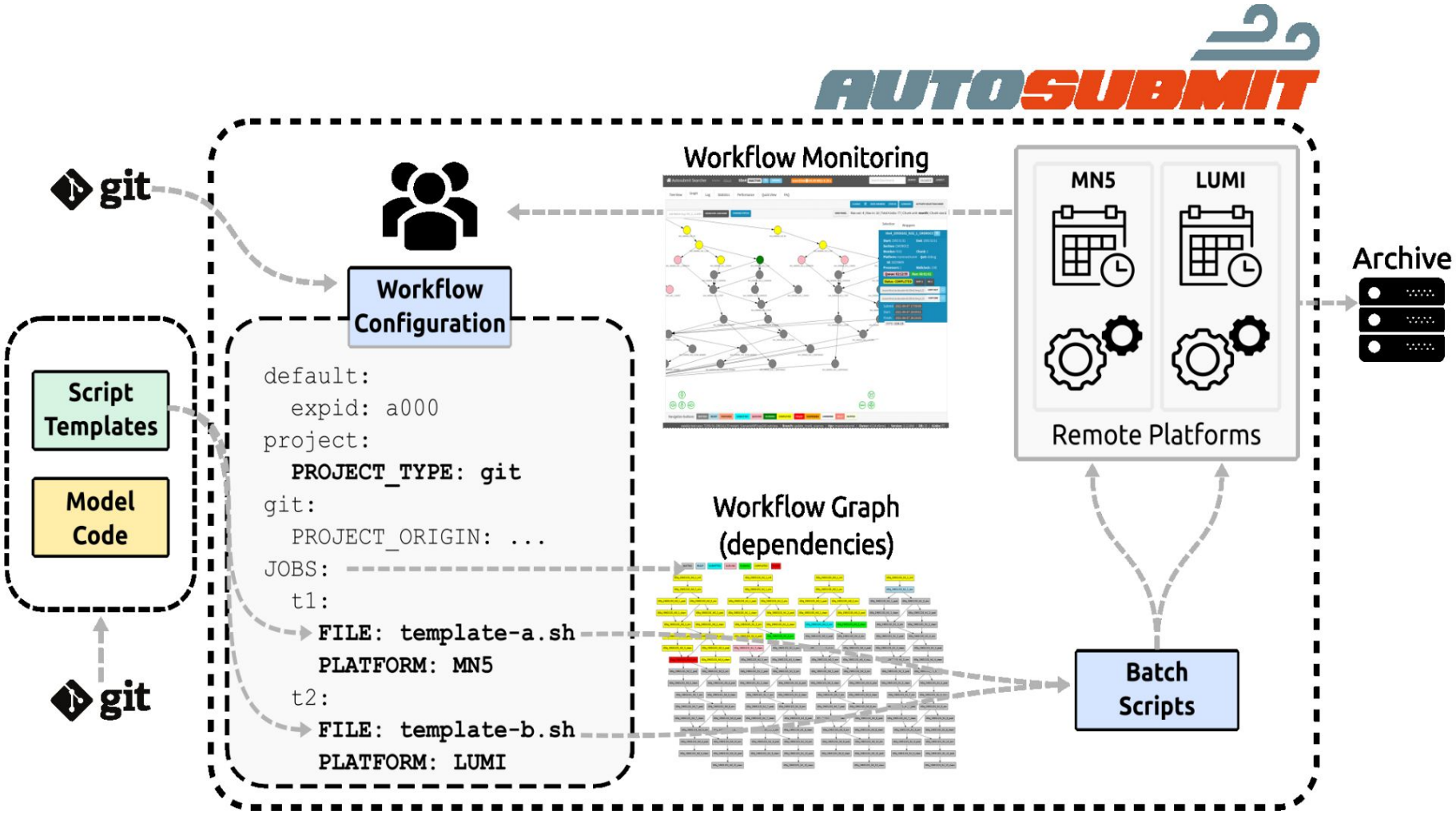


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

D. Manubens-Gil, J. Vegas-Regidor, C. Prodhomme, O. Mula-Valls and F. J. Doblas-Reyes, (2016). “Seamless management of ensemble climate prediction experiments on HPC platforms”, 2016 International Conference on High Performance Computing & Simulation (HPCS), Innsbruck, pp. 895-900. https://doi.org/10.1109/HPCSim.2016.7568429

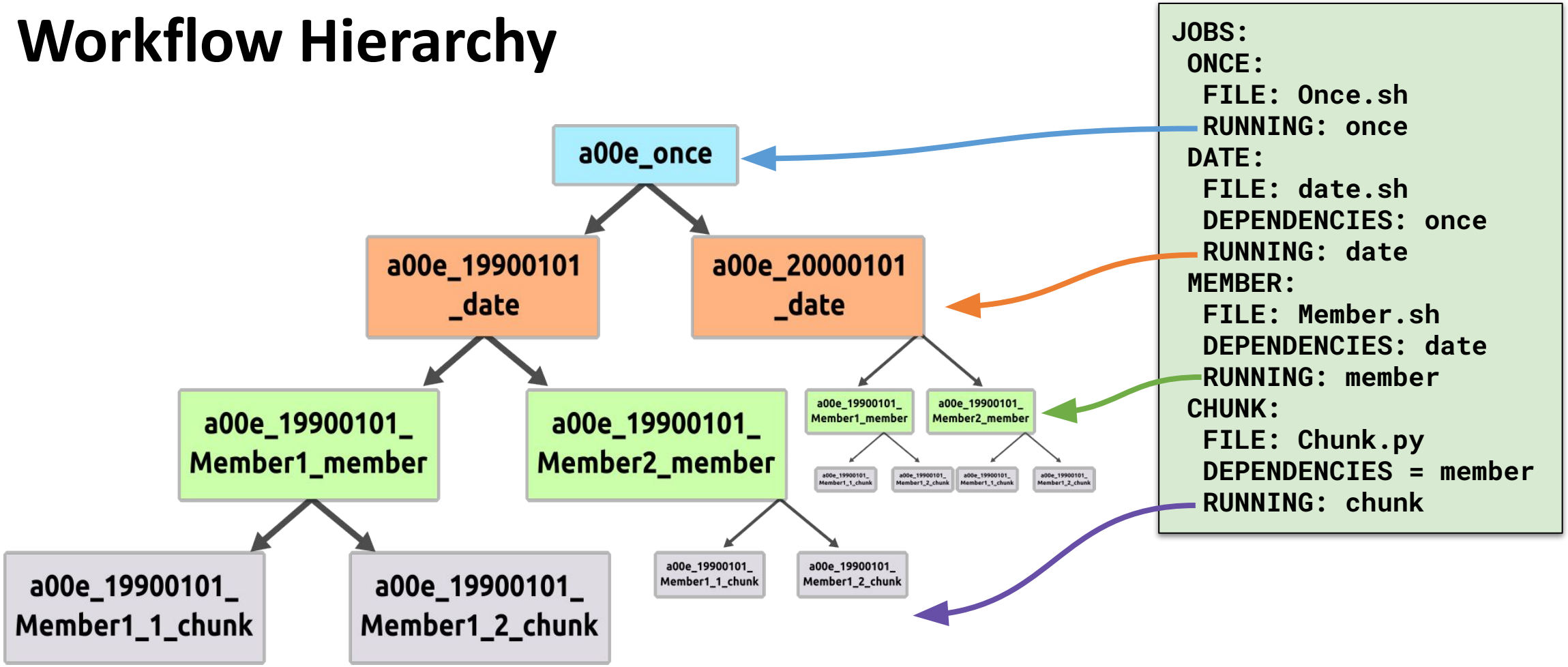
W. Uruchi, M. Castrillo and D. Beltrán, (2021). “Autosubmit GUI: A Javascript-based Graphical User Interface to Monitor Experiments Workflow Execution”, Journal of Open Source Software, 6(59), 3049. https://doi.org/10.21105/joss.03049

Autosubmit architecture



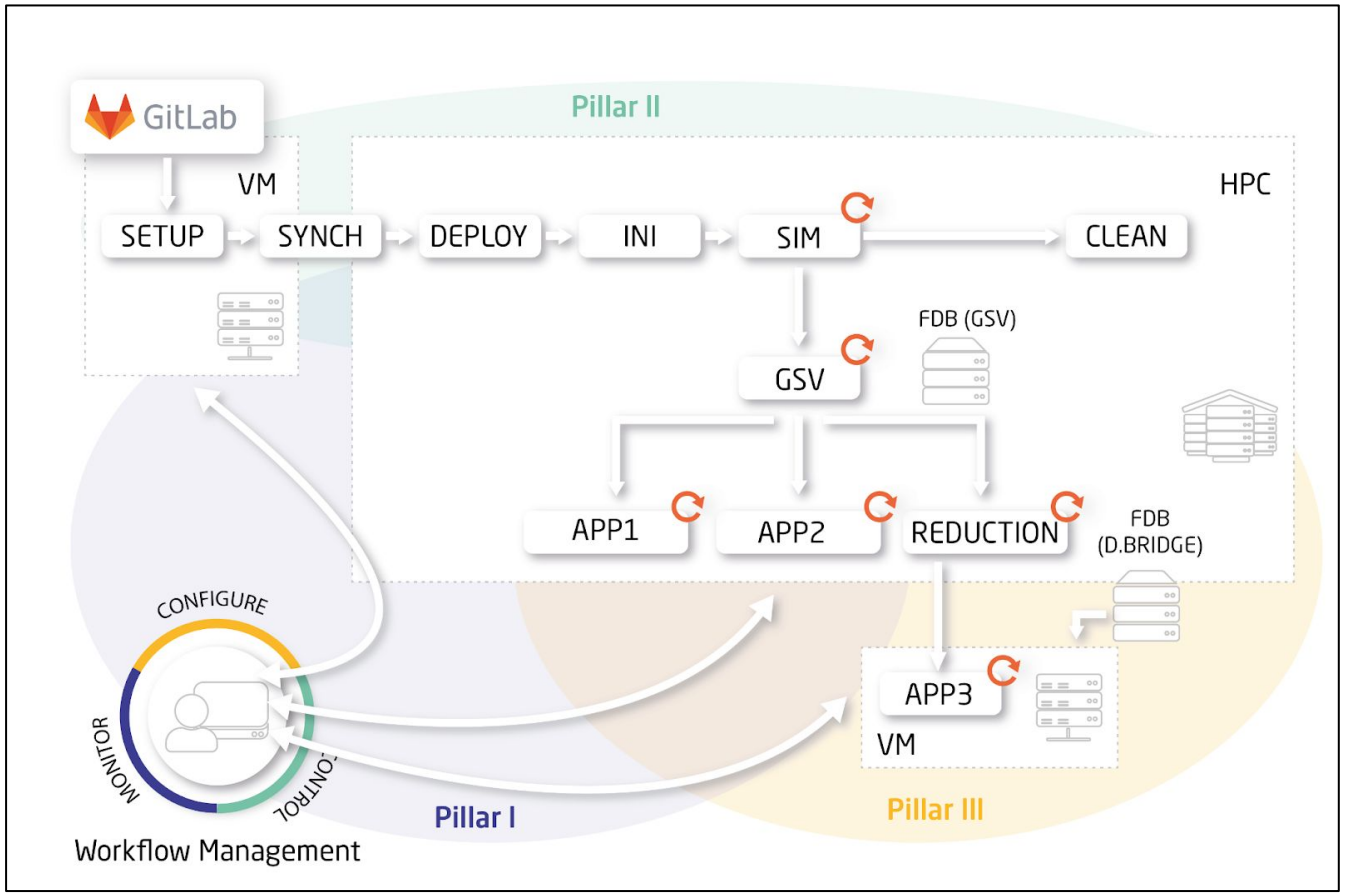
AUTOSUBMIT Configuration

Workflow Hierarchy

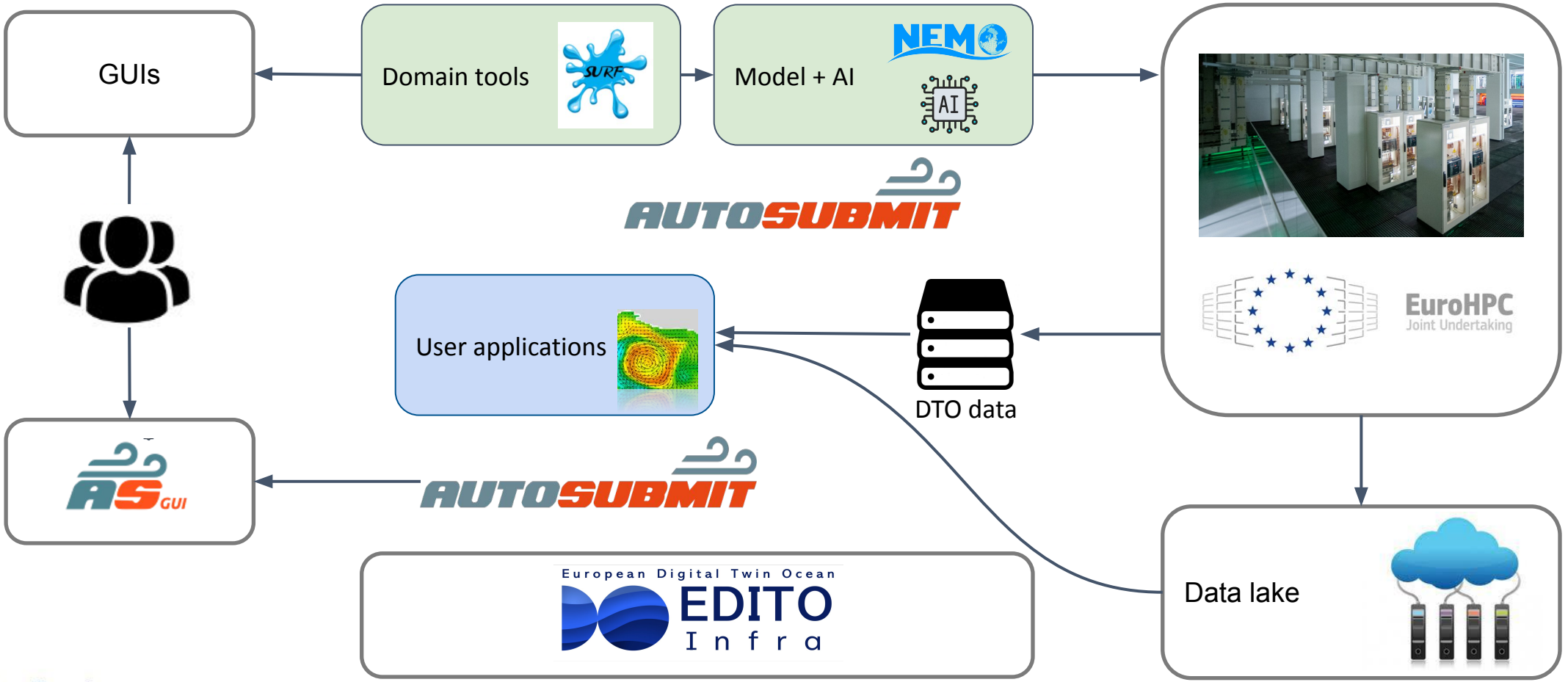


Climate DT workflow

The workflow makes possible the generation and streaming of the DT data by orchestrating the model components and the data consumers in real time.

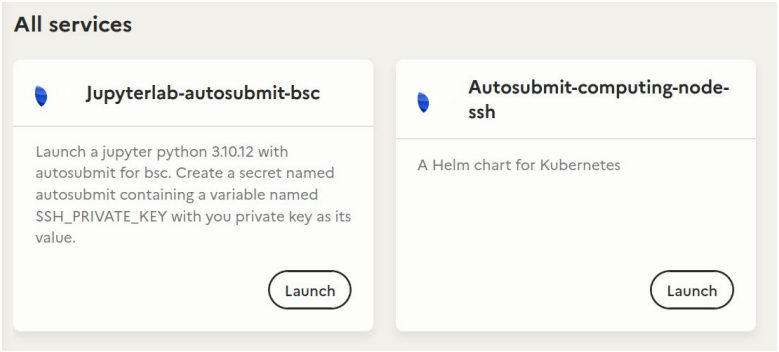


EDITO infrastructure

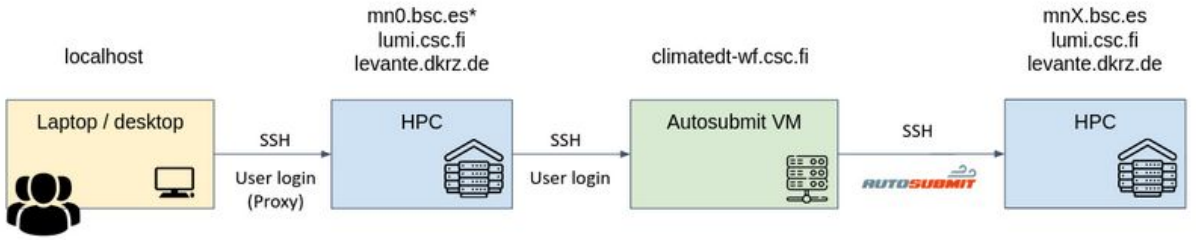


Autosubmit v4 features exploited by DTs

- New, **flexible**, YAML-based **configuration** system, allowing distributed experiment configuration.
  - Pre-defined configurations under CVS.
  - Users switch between configurations and customize them.
- Increased **interoperability** and **portability** (containers)
  - Python program in a dedicated VM (Climate DT).
  - Cloud service running in Jupyter Hub (EDITO Model Lab).
  - Export workflows to other backends.
- Increased **flexibility** in defining workflows
  - Dependencies customization.
- **FAIR principles**
  - Conforms the RO-Crate standard.

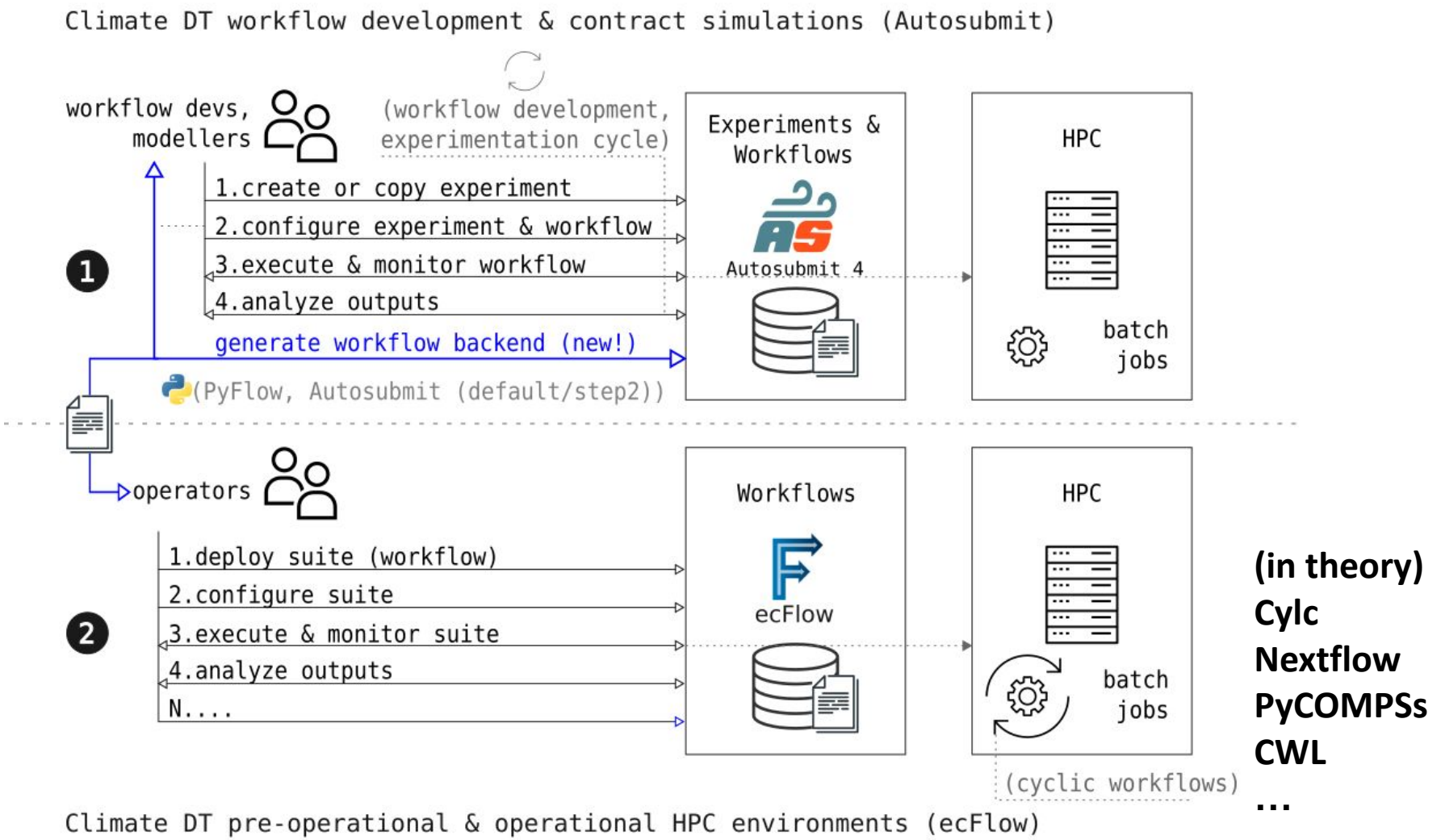


Autosubmit service in EDITO Infra



Autosubmit VM in ClimateDT

Autosubmit + PyFlow (ecFlow)



(in theory)
Cylc
Nextflow
PyCOMPSs
CWL
…

Infrastructure requirements

- **Co-development** → Discover and replicate experiments
- **Model-agnostic** experiment **configuration**
- **Complex** workflows, simple usage → Scientific language interface

Experiment manager

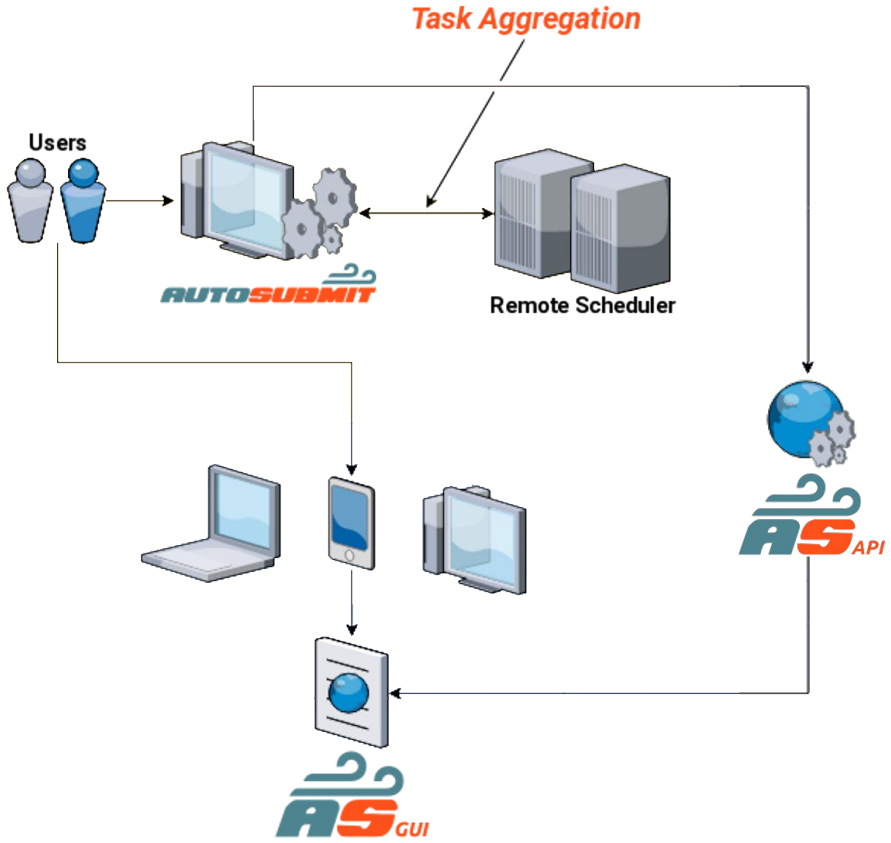
- Automatic and reproducible workflow **orchestration**
- **Portability** → Python, better if containerized
- **Workflow monitorization**
- **Robustness and efficiency** in using **shared resources** →Job aggregation

Workflow manager

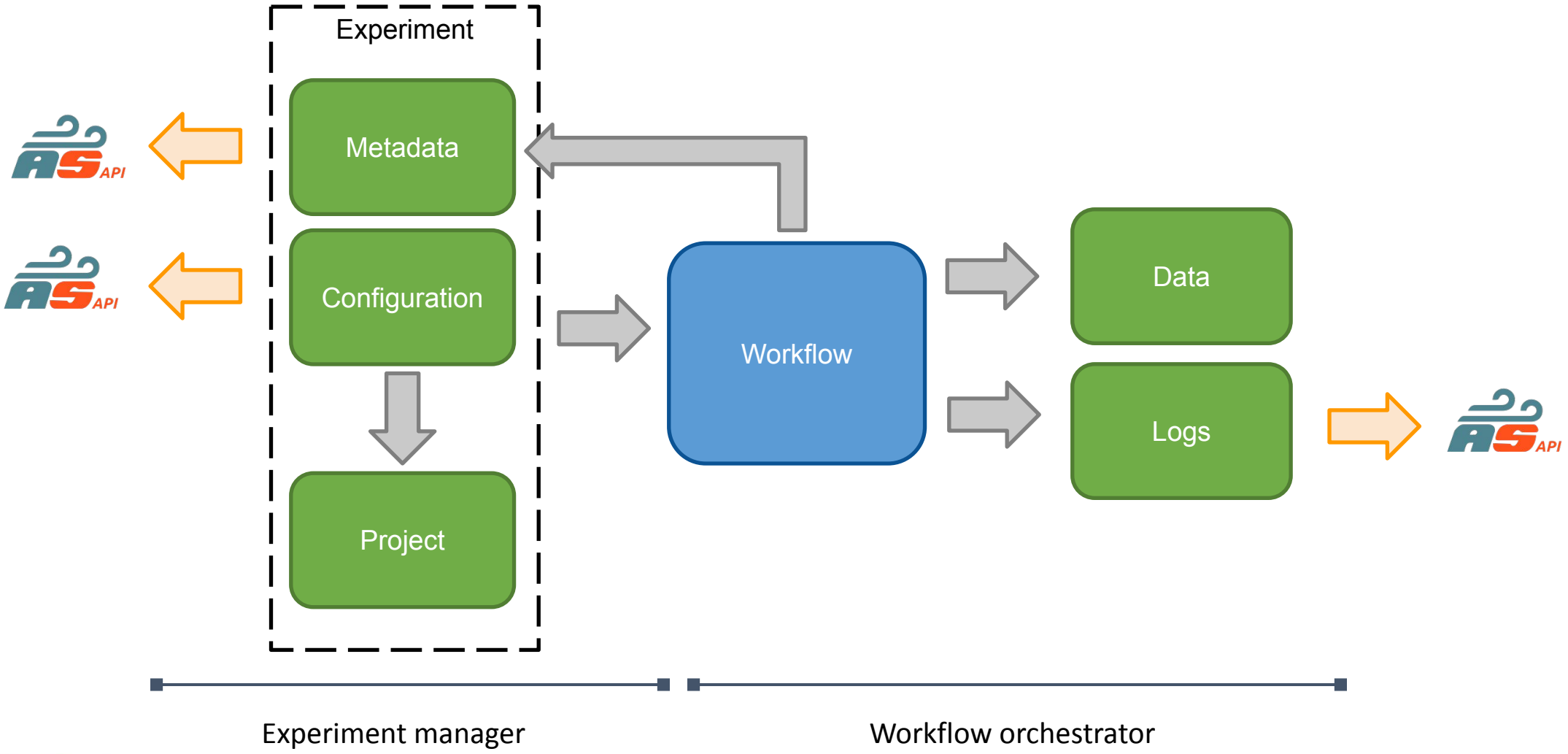
- **Interactivity**

Achieving interactivity with the system

- **Interactivity** is an essential feature of the Digital Twins.
- Users must be able to **monitor** the core engine, the available data, and their applications.
- Users must be able to **interact** with the system through requests: additional data, different scenarios, etc.
- The **workflow manager**, as backend component, must **enable** this interactivity.



Workflow operation



Conclusions

- Earth Digital Twins are **new systems** integrating **different components** and **relevant research** that can be used for **decision making**.
- They are based on **high-resolution** simulations, **impact modeling** and **high-performance computing**.
- **Experiment managers** are needed to define, configure, share and track the DT executions.
- DTs require backends allowing **automated** workflows with **user interaction** and **on-demand** computation.
- **Integration** and **interoperability** between different systems (software and hardware, including workflow managers) and paradigms (AI/Cloud, HPC) is critical (APIs).

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Questions?

Icons credit to www.flaticon.com

miguel.castrillo@bsc.es

EDITO-Model Lab has received funding from the European Union's Horizon Europe research and innovation programme under the grant No 101093293

esiwace CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE
Funded by the European Union
eFlows4HPC www.eFlows4HPC.eu
Funded by the European Union Destination Earth
implemented by ECMWF esa EUMETSAT

Autosubmit GUI