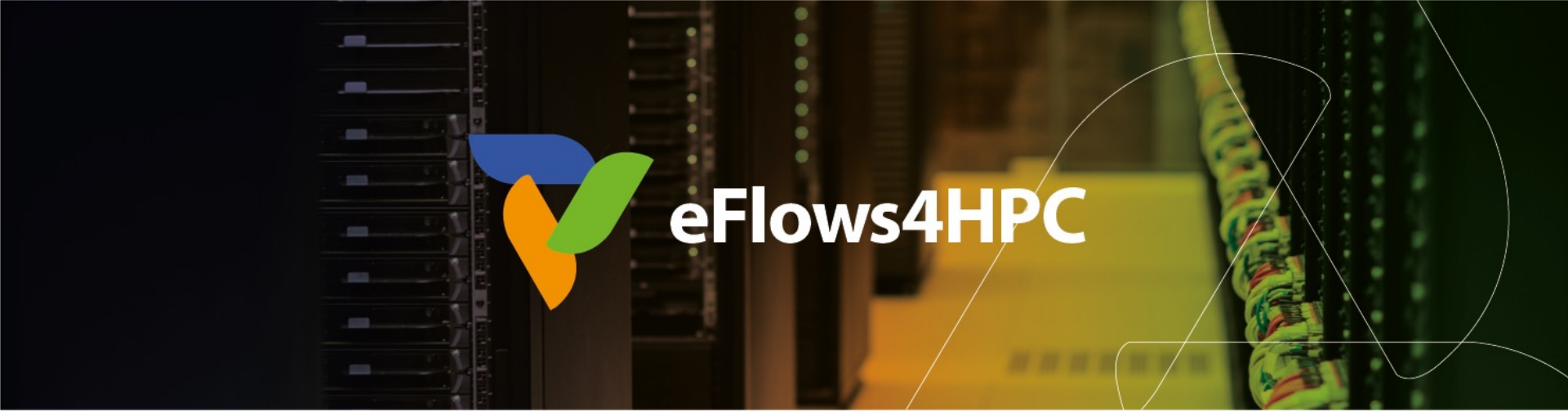



# eFlows4HPC workshop
# Innovative HPC workflows for industry

Rosa M Badia (BSC)

Innovative HPC workflows for industry
Munich Urban Colab, Munich (Germany)– 25th October 2023

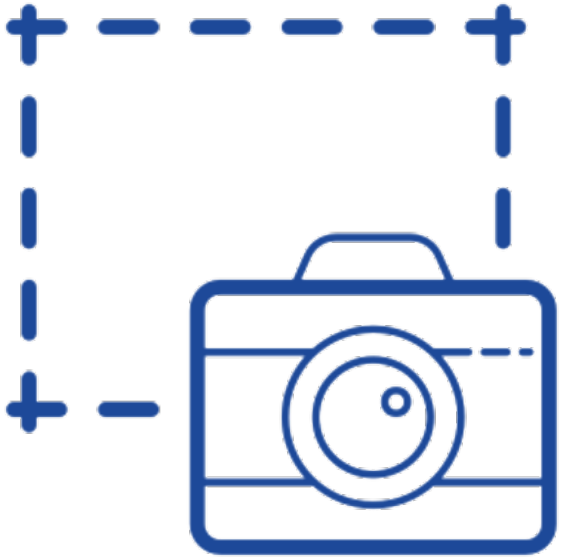
This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955558. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Germany, France, Italy, Poland, Switzerland, Norway. MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR (PCI2021-121957)

# Agenda

| | | |
|---|---|---|
| 9:00 - 9:30 | Registration and arrival | |
| 9:30 – 10:00 | eFlows4HPC project overview | Rosa M Badia (BSC) |
| 10:00 – 10:30 | Overview of Pillar I workflows | Riccardo Rossi (CIMNE)<br>Stefan Boschert (Siemens) |
| 10:30 – 11:00 | Coffee break | |
| 11:00 – 12:300 | ROM demo and hands-on session, part I | Raul Bravo, Sebastian Ares (CIMNE) |
| 12:00 – 13:00 | Lunch break | |
| 13:00 – 14:30 | ROM demo and hands-on session, part II | Raul Bravo, Sebastian Ares (CIMNE) |
| 14:30 - 15:00 | Demo about HPCWaaS | Jorge Ejarque(BSC) |
| 15:00 - 15:30 | Provenance with PyCOMPSs | Raül Sirvent (BSC) |
| 15:30 - 16:00 | Conclusions and farewell | Rosa M. Badia (BSC) |

# Photos

- For project dissemination purposes, photos will be taken during the event to appear on eFlows4HPC and BSC social media accounts and websites and project reports.



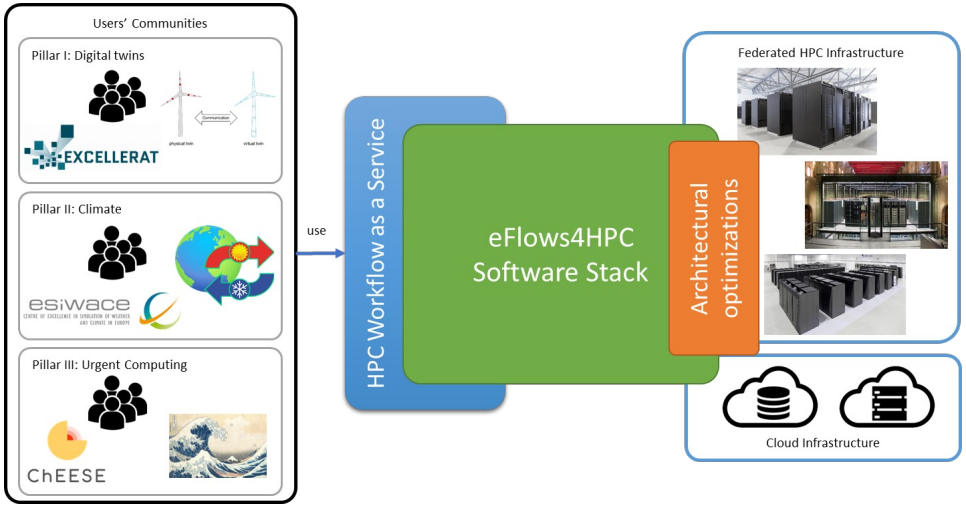
# EFLOWS4HPC OVERVIEW

# Complex workflows and complex infrastructures

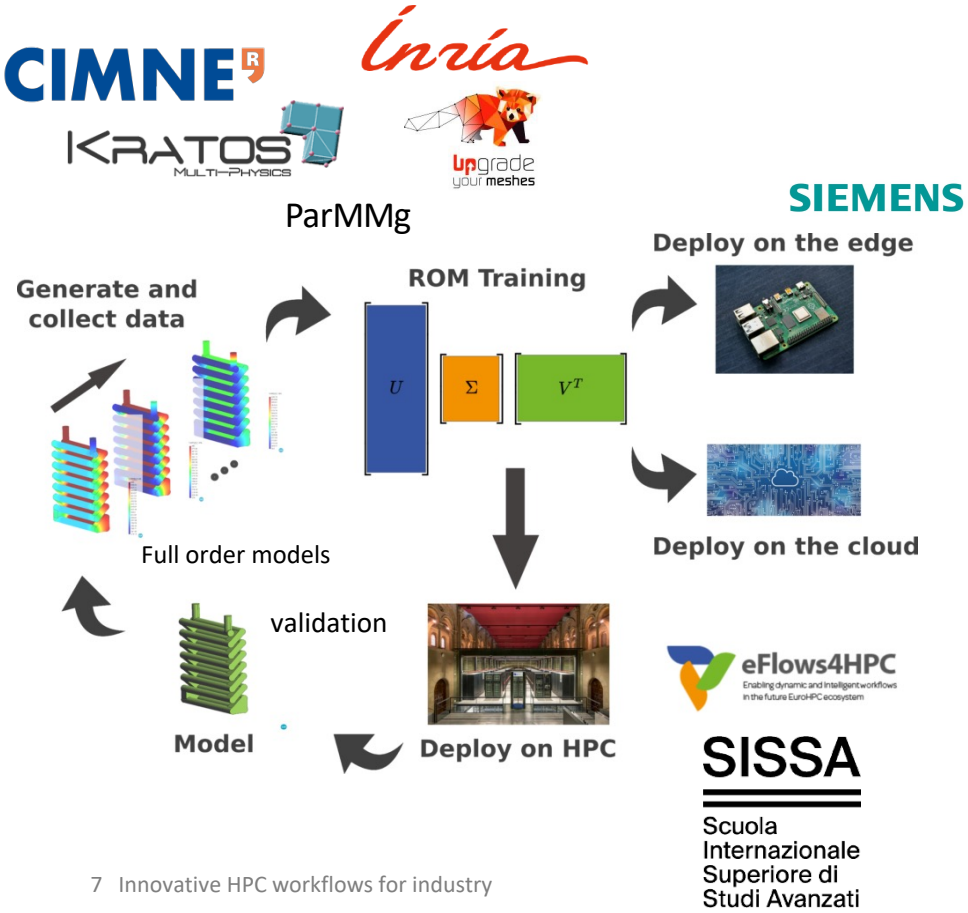
- EuroHPC aims at developing a World Class Supercomputing Ecosystem in Europe
  - Procuring and deploying pre-exascale and petascale systems in Europe
- These systems will be capable of running large and complex applications
- Applications demand the composition of HPC, artificial intelligence and data analytics
- The development, installation, execution and of workflows is manual and error prone:
  - New tools and methodologies are needed

# eFlows4HPC in a nutshell

- **Software tools stack that makes easier the development and management of complex workflows:**
  - Combine different aspects
    - HPC, AI, data analytics
  - Reactive and dynamic workflows
    - Autonomous workflow steering
  - Full lifecycle management
    - Not just execution
    - Data logistics and Deployment
- **HPC Workflows as a Service:**
  - Mechanisms to make easier the use and reuse of HPC by wider communities
- **Architectural Optimizations:**
  - Selected HPC – AI Kernels Optimized for GPUs, FPGA, EPI
- **Validation Pillar's**
  - End-user workflows linked to CoEs

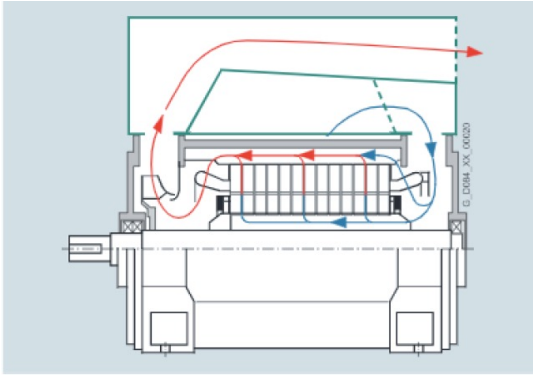


# Pillar I: Manufacturing

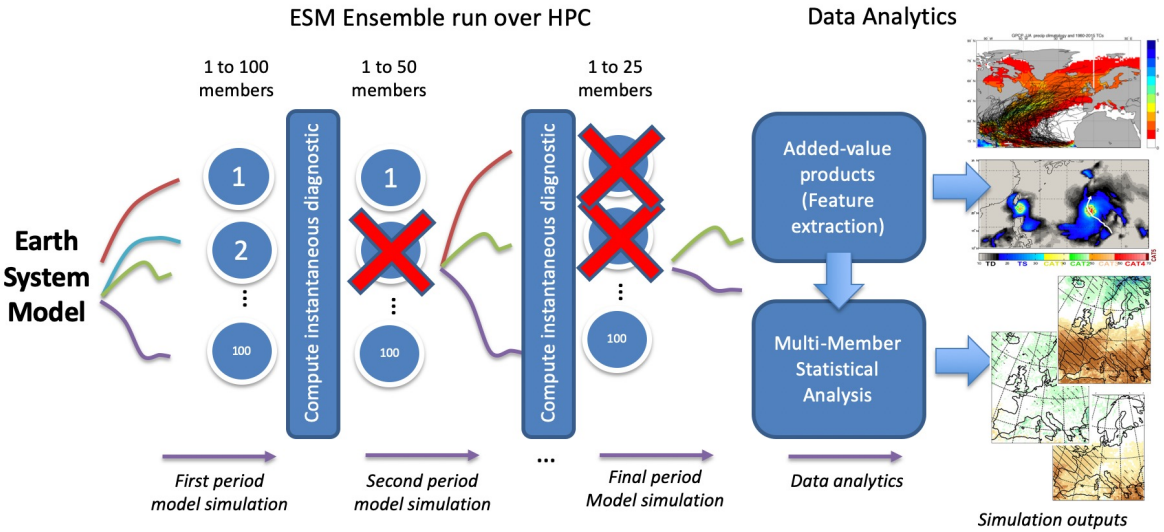


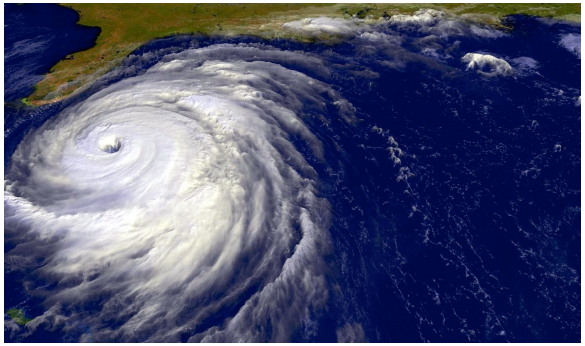

Pillar I focuses on the construction of DigitalTwins for the prototyping of complex manufactured objects:

- Integrating state-of-the-art adaptive solvers with machine learning and data-mining
- Contributing to the Industry 4.0 vision



# Pillar II: Climate



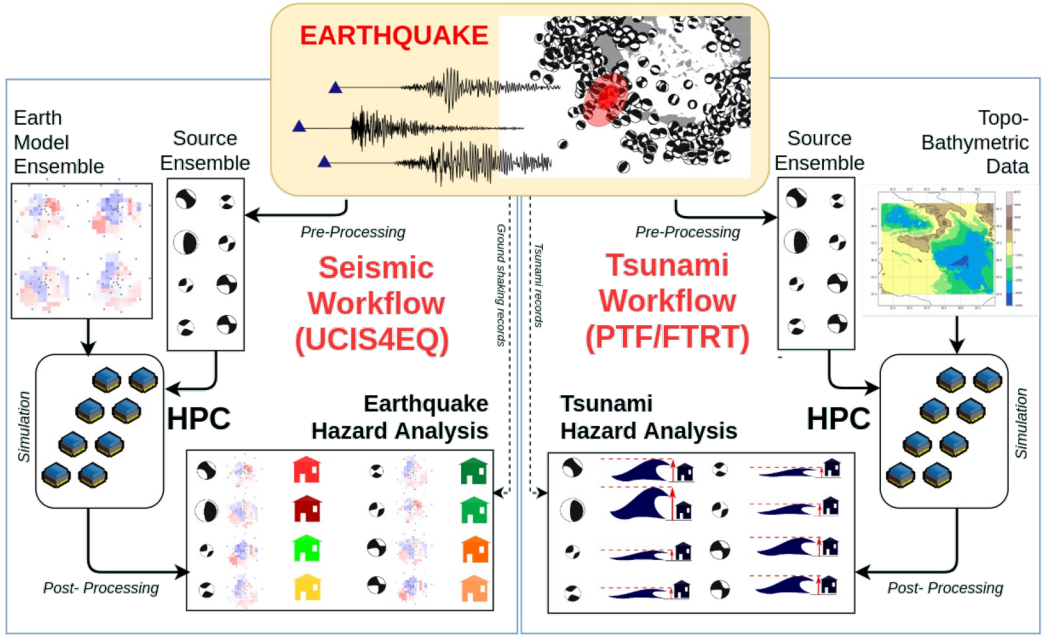



Dynamic (AI-assisted) workflow

HPDA & ML/DL

- Perform climate predictions: temperature, precipitation or wind speed
- AI-assisted pruning of the ESM workflow
- Study of Tropical Cyclones (TC) in the North Pacific, with in-situ analytics

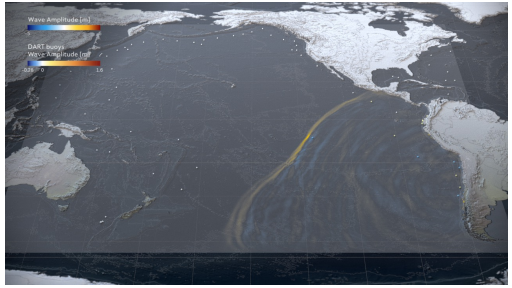
# Pillar III: Urgent computing for natural hazards



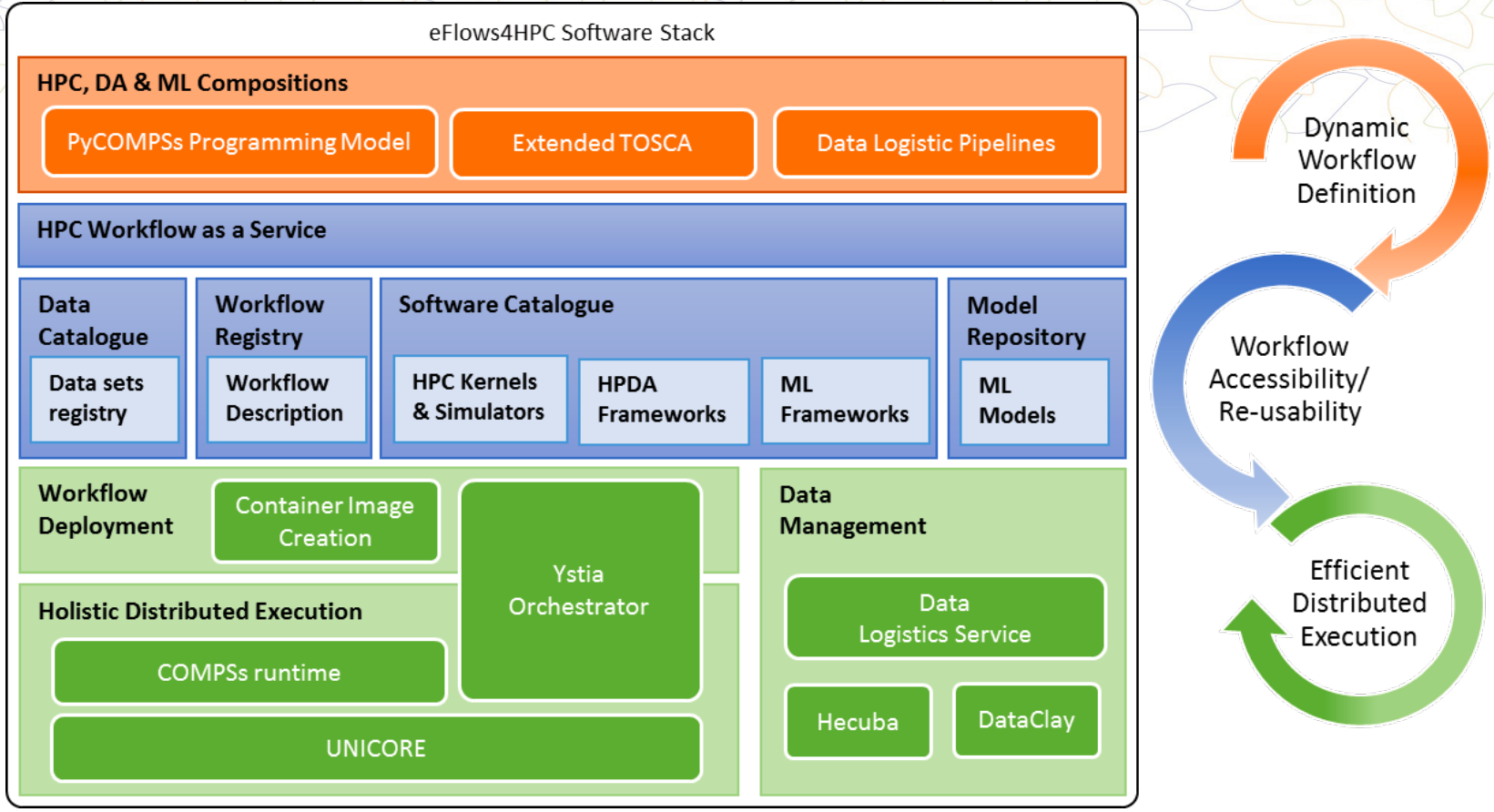

Pillar III explores the modelling of natural catastrophes:

- Earthquakes and their associated tsunamis shortly after such an event is recorded
- Use of AI to estimate intensity maps
- Use of DA and AI tools to enhance event diagnostics
- Areas: Mediterranean basin, Mexico, Iceland and Chile

*Tsunami-HySEA GPU-based code*



*25/10/2023*


eFlows4HPC Software Stack
HPC, DA & ML Compositions
PyCOMPSs Programming Model
Extended TOSCA
Data Logistic Pipelines
HPC Workflow as a Service
Data Catalogue
Data sets registry
Workflow Registry
Workflow Description
Software Catalogue
HPC Kernels & Simulators
HPDA Frameworks
ML Frameworks
Model Repository
ML Models
Workflow Deployment
Container Image Creation
Ystia Orchestrator
Holistic Distributed Execution
COMPSs runtime
UNICORE
Data Management
Data Logistics Service
Hecuba
DataClay
Dynamic Workflow Definition
Workflow Accessibility/ Re-usability
Efficient Distributed Execution

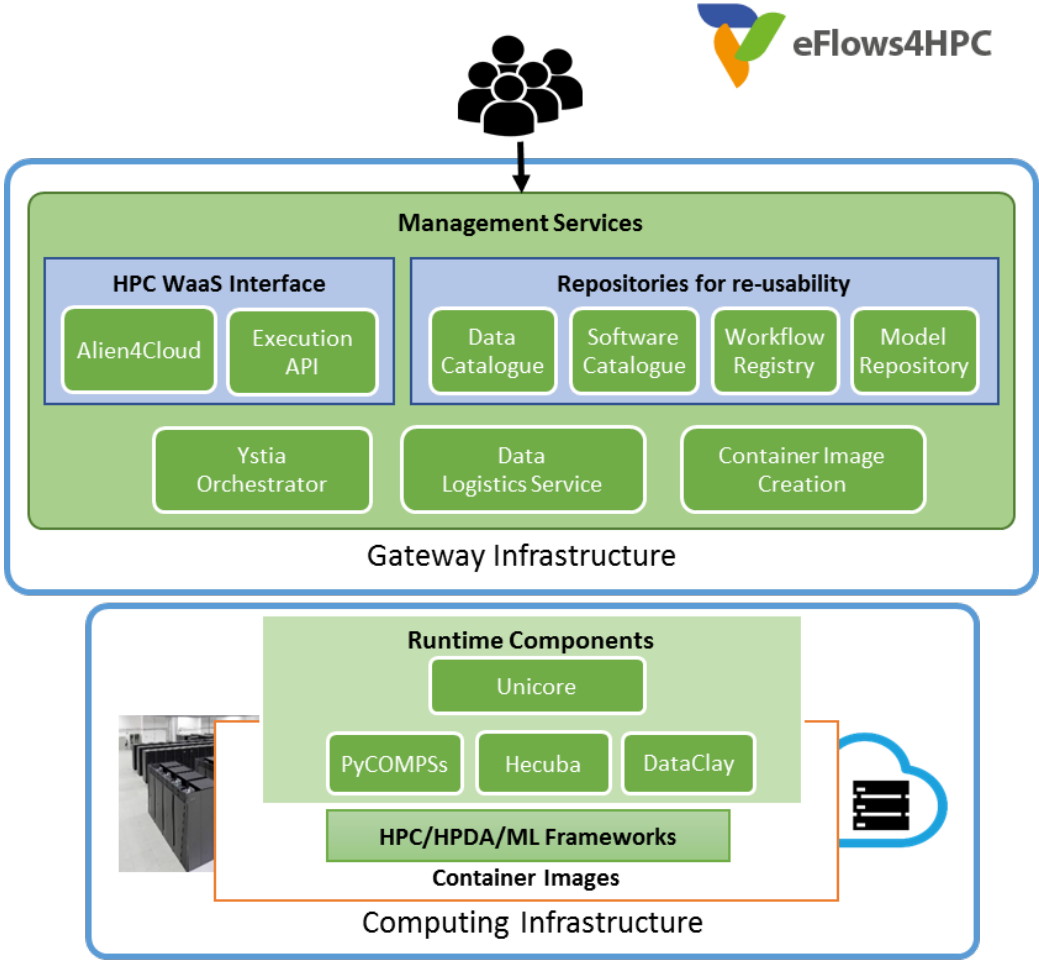
# Software stack deployment

## Gateway services

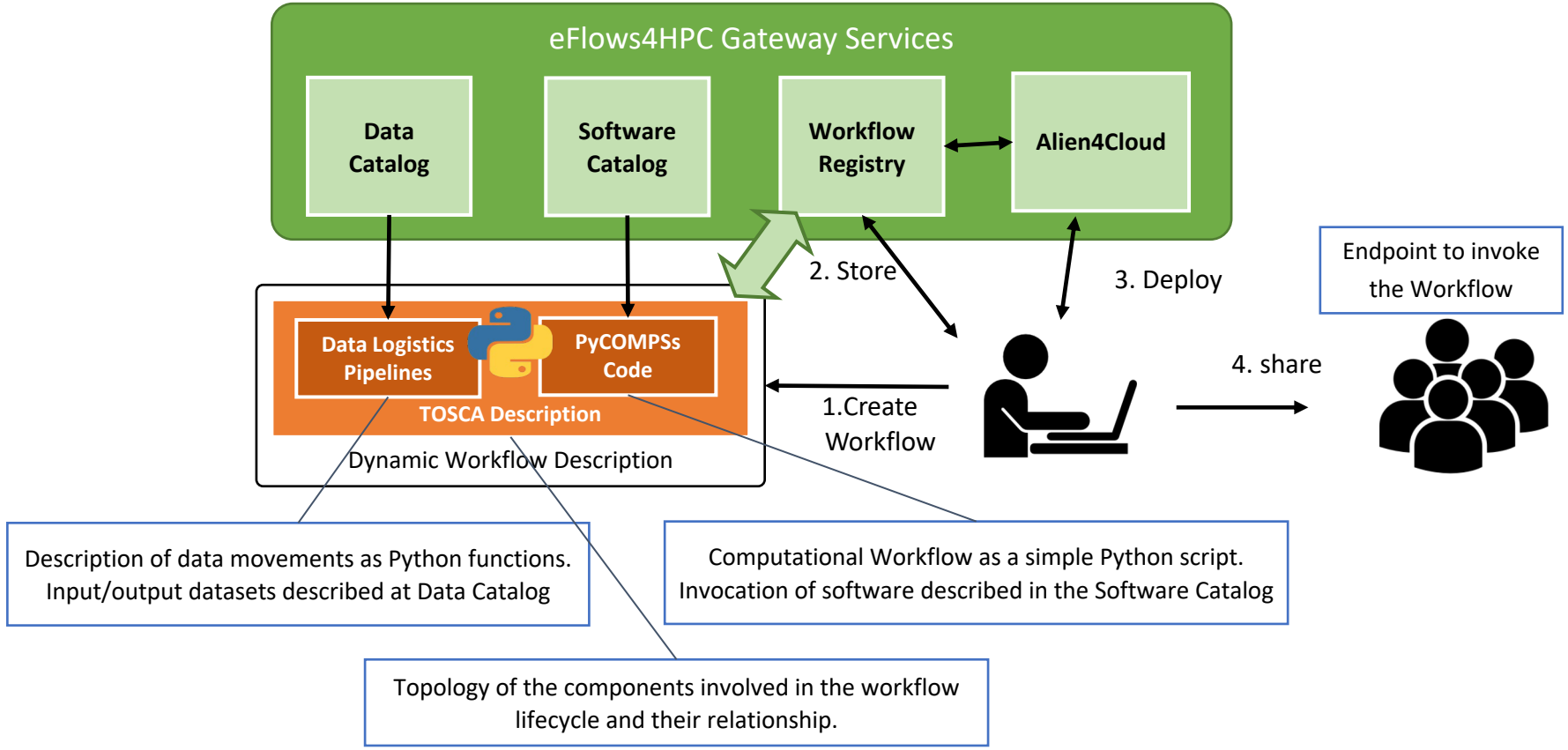
- Components deployed outside the HPC infrastructure.
- Managing external interactions and workflow lifecycle

## HPC and runtime Components

- Deployed inside the HPC infrastructure to manage the workflow execution



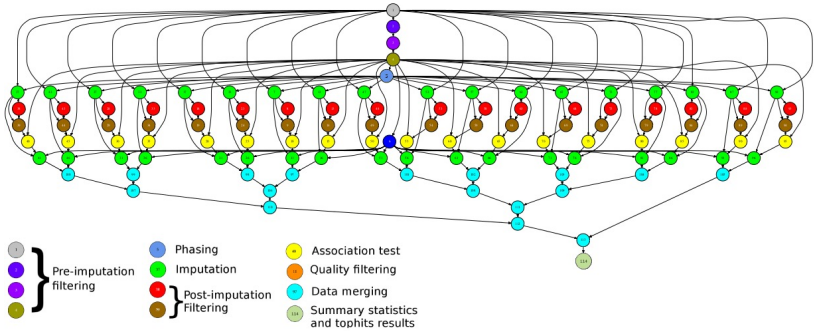
# Workflow development overview



# Main element: Workflows in PyCOMPSs

- Sequential programming, parallel execution
- General purpose programming language + annotations/hints
  - To identify tasks and directionality of data
- Builds a task graph at runtime that express potential concurrency
- Tasks can be sequential and parallel (threaded or MPI)
- Offers to applications the illusion of a shared memory in a distributed system
  - The application can address larger data than storage space: support for Big Data apps
- Agnostic of computing platform
  - Enabled by the runtime for clusters, clouds and container managed clusters

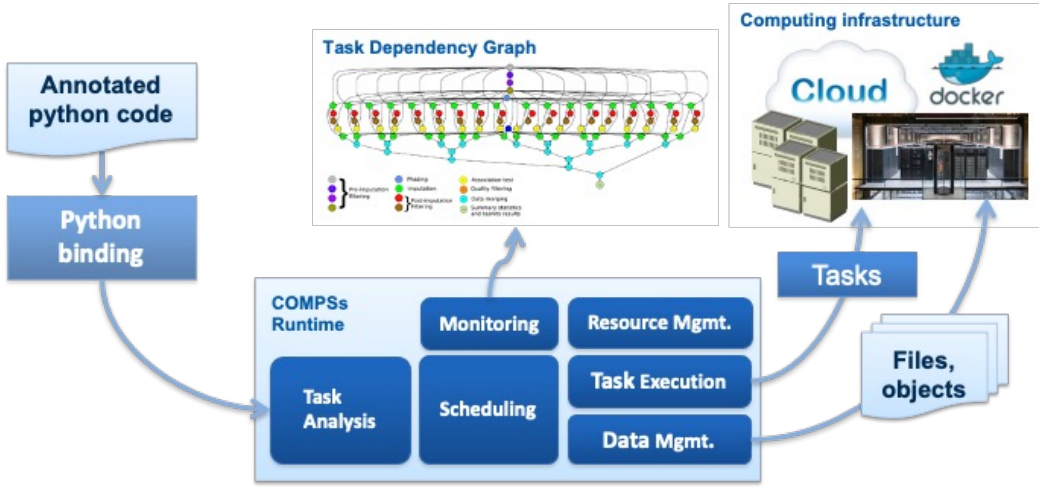
```
@task(c=INOUT)
def multiply(a, b, c):
    c += a*b
```




Barcelona Supercomputing Center
Centro Nacional de Supercomputación

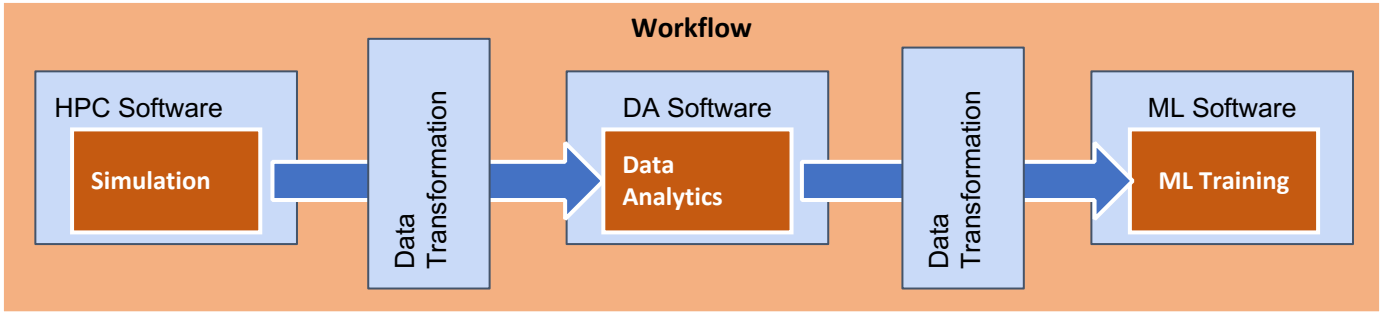
# PyCOMPSs features and runtime

- Support for tasks' constraints – support for heterogeneous infrastructure
- Support for tasks' faults and tasks' exceptions
  - Enlarges the dynamicity of the type of workflows that we support
- Streamed data
  - ... and many others
- Runtime deployed as a distributed master-worker
- All data scheduling decisions and data transfers are performed by the runtime
- Support for elasticity




Addressing challenge 2 and 3

# Interfaces to integrate HPC/DA/ML




- Goal:
  - Reduce the required glue code to invoke multiple complex software steps
  - Developer can focus in the functionality, not in the integration
  - Enables reusability
- Two paradigms:
  - Software invocation
  - Data transformations

```
#workflow steps defined as tasks
@data_transformation (input_data, transformation description)
@software (invocation description)
def data_analytics (input_data, result):
    pass

#workflow body
simulation (input_cfg, sim_out)
data_analytics (sim_out, analysis_result)
ml_training (analysis_result, ml_model)
```



Addressing challenge 1 and 3

# Data Catalogue and Data Logistics Service

## Data Catalogue:

- Lists datasets used and created by the workflow according to FAIR principles
- Provides metadata to make data movement pipelines more generic

## Data Pipelines:

- Formalization of data movements for transparency and reusability
- Stage-in/out, image transfer

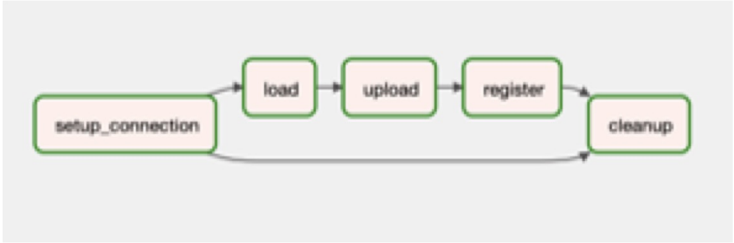
## Data Logistics Services (DLS):

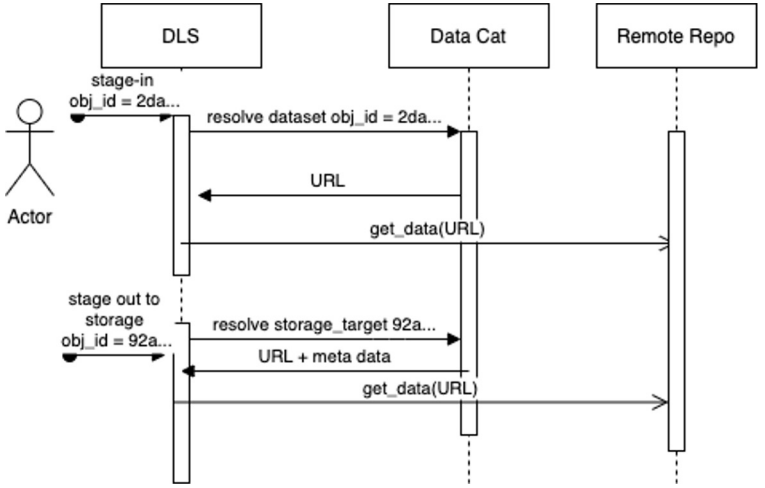
- Performs the execution of data pipelines at deployment and execution time

## Production Ready Services:

- https://datacatalogue.eflows4hpc.eu
- https://datalogistics.eflows4hpc.eu/

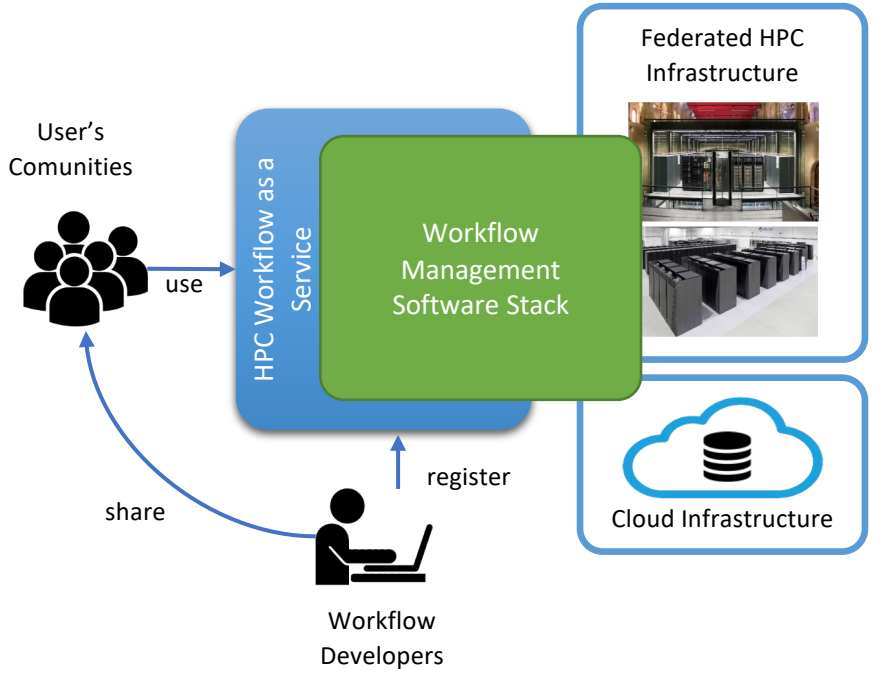
Data pipeline



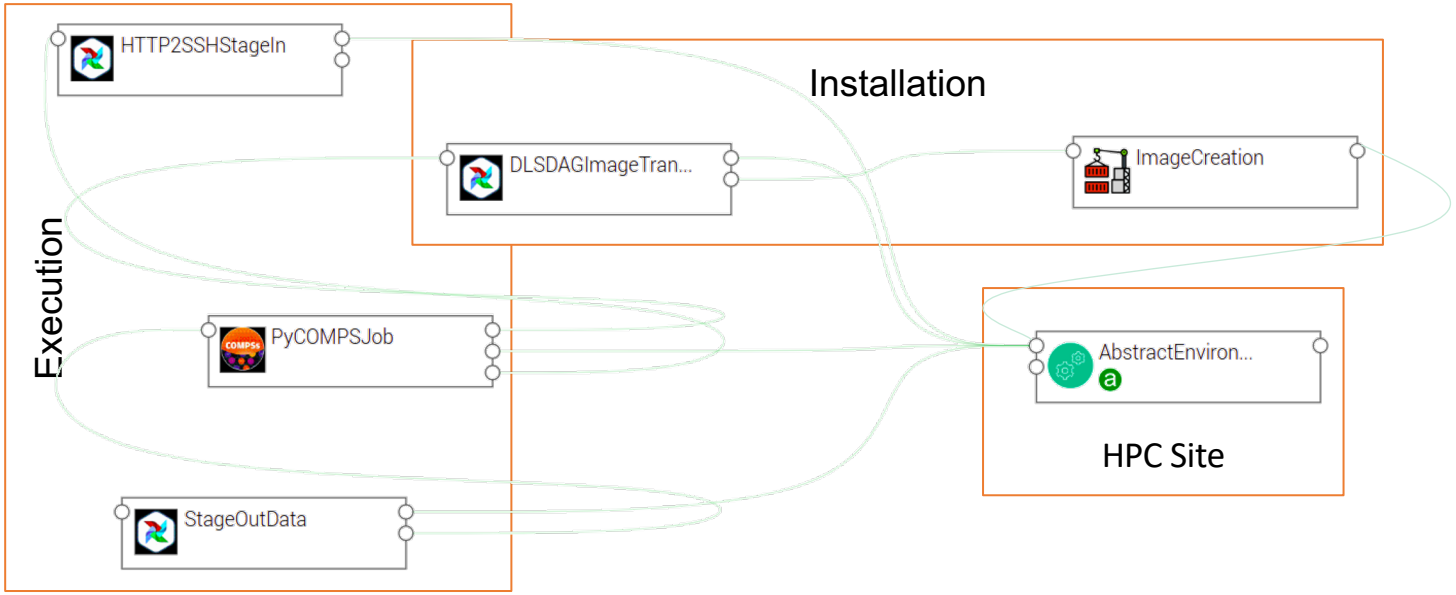
# Top-level workflows approach

- Requires a description for workflow lifecycle management
  - TOSCA:
    - Model to describe cloud application topologies and its lifecycle orchestration
- Interface for deploying and running the workflows
  - HPC Workflows as a Service (HPCWaaS):
    - Deployment based on containers
    - Execution: HPCWaaS API



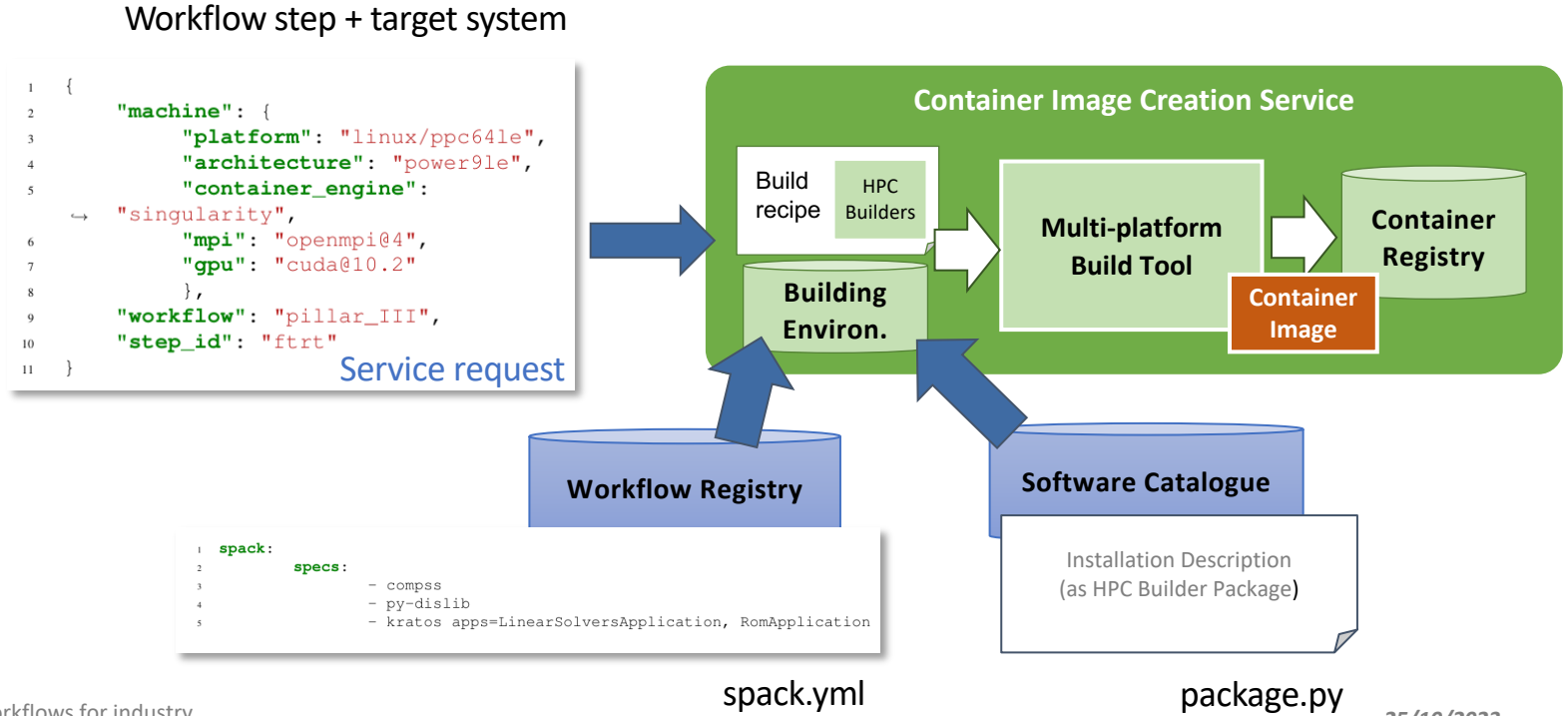
# TOSCA Modelization

Topology of the different components involved in the Workflow lifecycle management



# HPC Ready Containers

- Methodology to allow the creation containers for specific HPC system

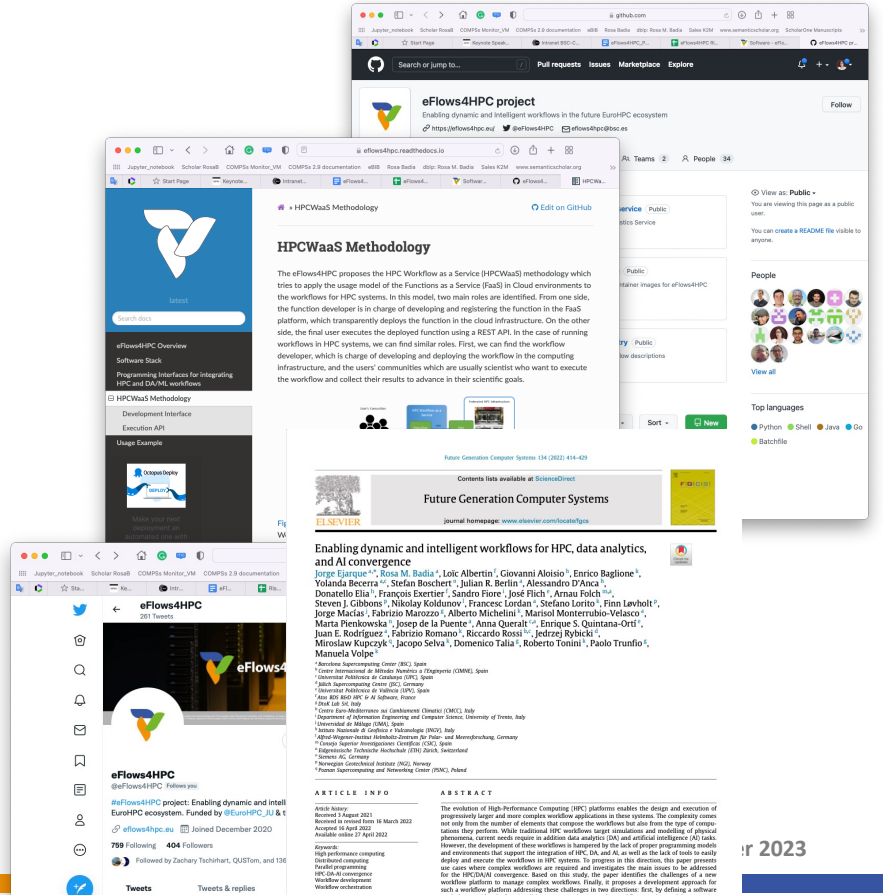


# Project main achievements

https://eflows4hpc.eu/software/

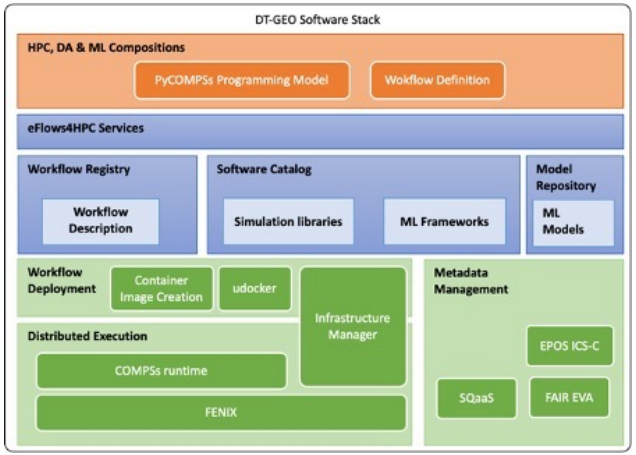
- Requirements and software architecture. Reviewed at beginning of second iteration
- Definition and implementation of abstractions to support the integration of different stack components
- Design and development of a minimal workflow. Development of a step-by-step example.
- Design and implementation of the HPCWaaS API
- Design and implementation of project services: Data Catalogue, Workflow Registry, Software catalogue
- Design and implementation of two versions of Pillars' workflows.
- Two releases of project software and documentation available
- Set of internal trainings about software stack components and HPCWaaS. ICS-HPC tutorial
- Good visibility: articles, keynote presentations, media

https://eflows4hpc.eu/software/



# Adoption of eFlows4HPC tools in other projects

DT-GEO: prototype for a digital twin on geophysical extremes




DT-GEO

CAELESTIS: simulation ecosystem of next generation aircraft structures

CAELESTIS

CEEC: Center of Excellence for Exascale CFD

CEEC

# Project partners

eFlows4HPC
















# eFlows4HPC

Enabling dynamic and Intelligent workflows
in the future EuroHPC ecosystem

www.eFlows4HPC.eu

@eFlows4HPC  eFlows4HPC Project

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955558. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Germany, France, Italy, Poland, Switzerland, Norway.