Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

One workflow to rule them all: introducing DAGonStar, yet another workflow engine for Python developers, designed for HPC and AI.

Prof. Raffaele Montella
University of Naples "Parthenope"

DagOn CAPIO
Directed Acyclic Graph on Cross-Application Programmable I/O

Hi-WeFAI Project

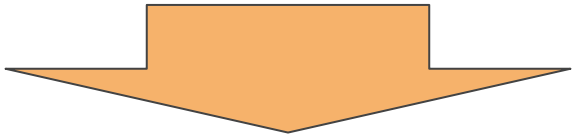
GL++

workflows community initiative
FIDES
Università degli Studi di Napoli Parthenope
ITDM your technology partner
MytilEX
Mytil
HPSC LAB HIGH-PERFORMANCE SCIENTIFIC COMPUTING LABORATORY

Introduction and motivation

Running unattended scientific workflow applications

- **Routinary:** running each time initial data are available
- **On demand:** footprint for external software components execution.
- **Orchestrating:** external software (diverse and different, producing and consuming large files -- or group of files).
- **Failsafe:** failure have not to be an issue: even partially produced outputs have to be enough to continue the production.
- **Replicability, reproducibility, reusability, FAIRness:** not just buzzword, but rocksolid cornerstones.

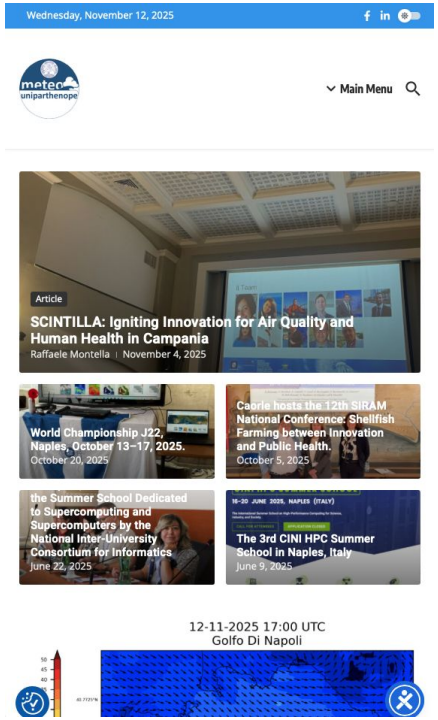


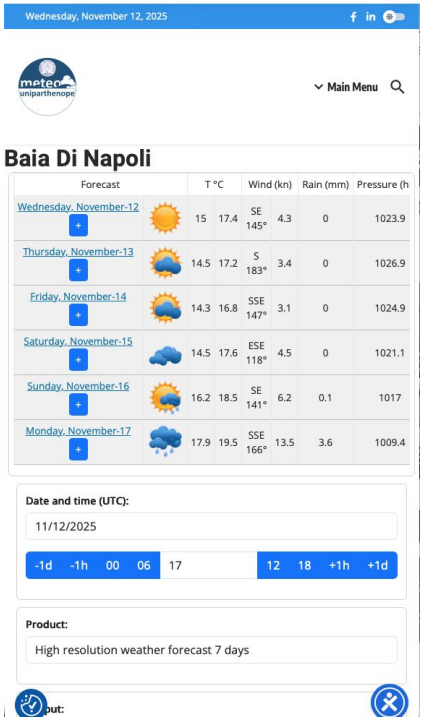
- **Co-design:** from the application requirements to the middleware features and vice versa.

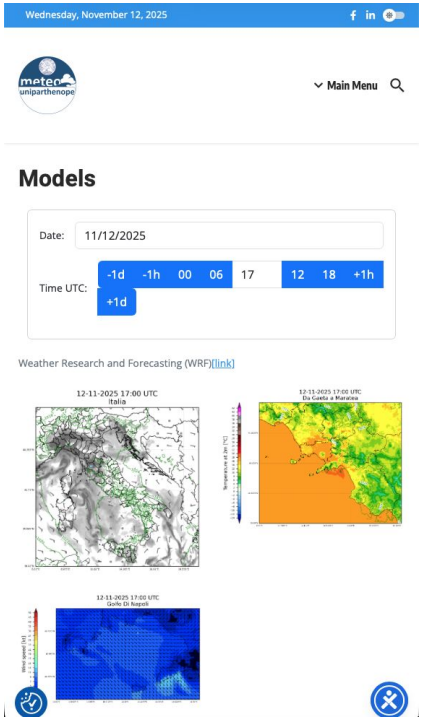
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

meteo@uniparthenope

http://meteo.uniparthenope.it







Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

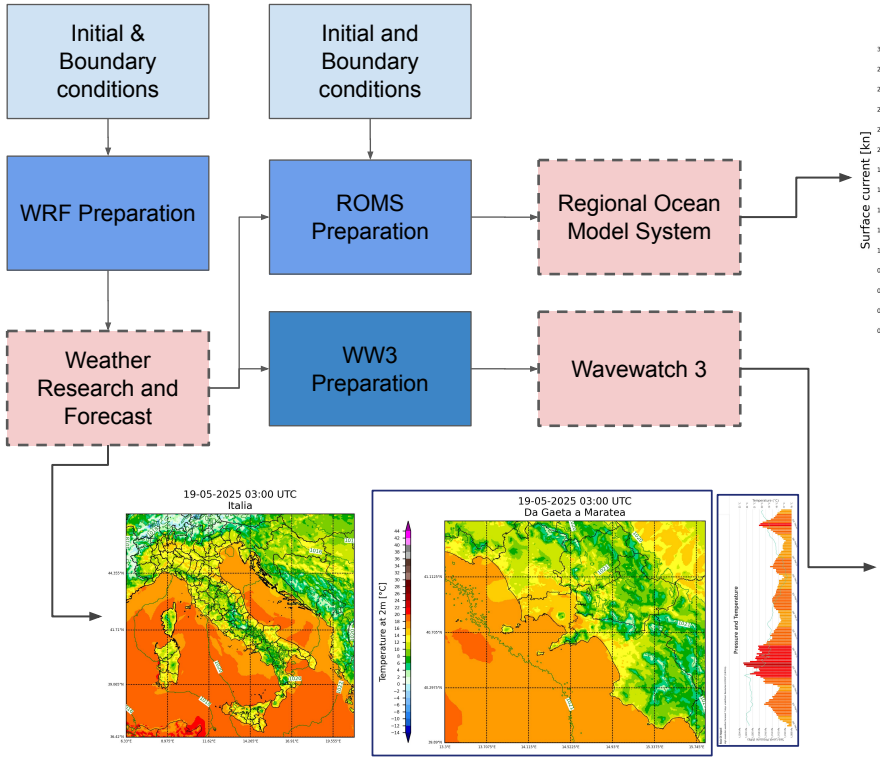
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

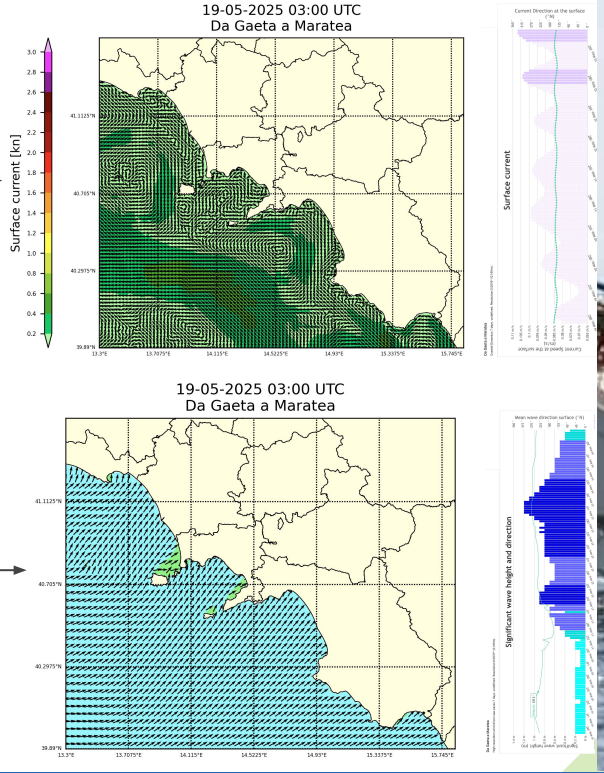
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

meteo@uniparthenope

http://meteo.uniparthenope.it

meteo uniparthenope





Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Predicting bacteria contamination in farmed mussels

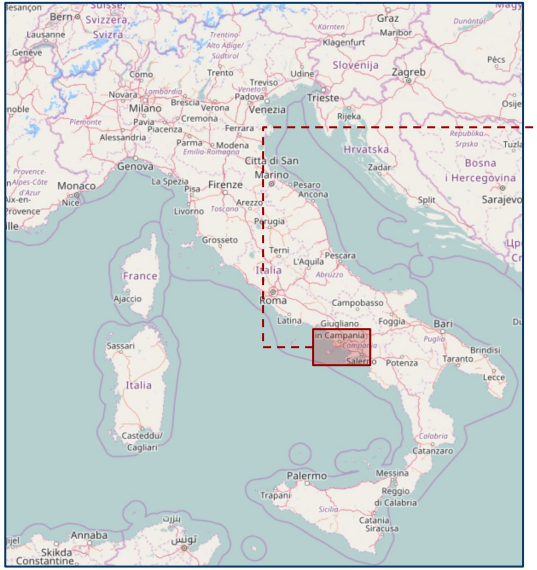
Mussels farming is an outstanding business cornerstone in the most part of Italian coastal regions.

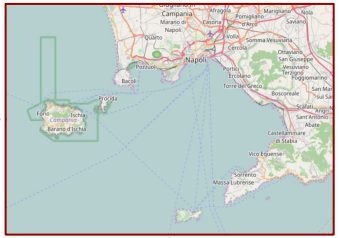
MARKET

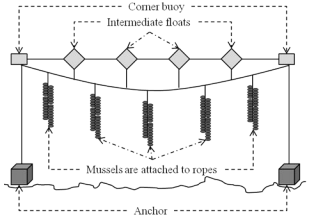
- Companies: 263(d)/886(r)
- Tons: ~64235 (⅔ EU prod - ISPRA)
- Euro/Kg: ~1.75 (average)

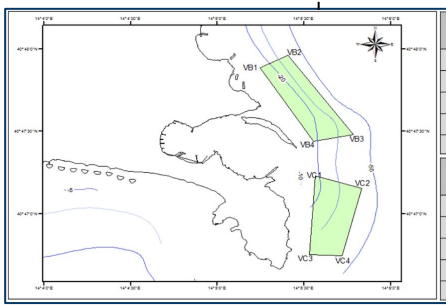
~112M€ (2013, Italy)

Making predictions about the pollutant concentration in mussel farms areas in order to limit human diseases.









Mussel farms Punta Cento Camerelle Bacoli (NA)

Vertex	Latitude	Longitude
VB1	40°47'49".38	14°05'11".64
VB2	40°47'53".94	14°05'21".30
VB3	40°46'25".14	14°05'43".98
VB4	40°46'22".74	14°05'30".66

Mussel farms Punta Terone Bacoli (NA)

Vertex	Latitude	Longitude
VC1	40°47'10".14	14°05'30".66
VC2	40°47'05".64	14°05'46".80
VC3	40°46'41".10	14°05'42".49
VC4	40°46'41".64	14°05'28".62

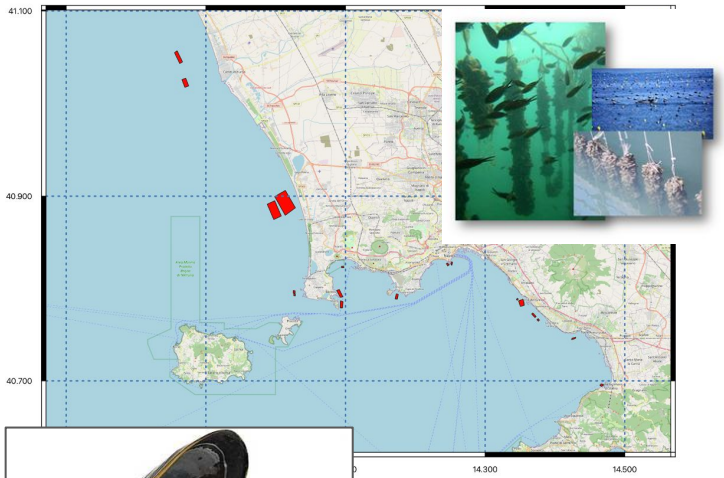
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Predicting bacteria contamination in farmed mussels

- Use case: Campania – Italy, *Mytilus galloprovincialis*
- Farming banks types: Artificial, Natural

Microbiological	Escherichia coli
	Salmonella
Chemical	Mercury
	Lead

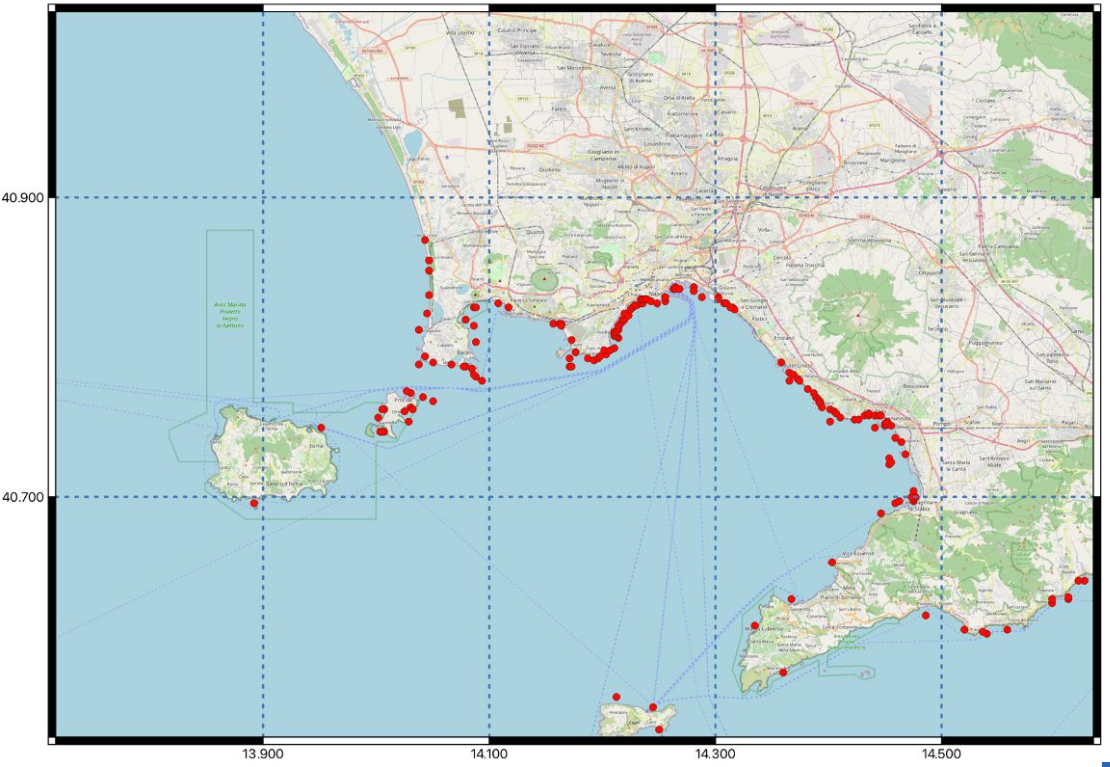
Ecotoxicological	Complesso DSP
	Yessotossine
	Azaspiracides
	ASP
	PSP
	Radioactive nuclides



Mytilus Galloprovincialis

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Study domain: the Campania Region, South of Italy



Faecal contamination
may arise from a variety of sources:

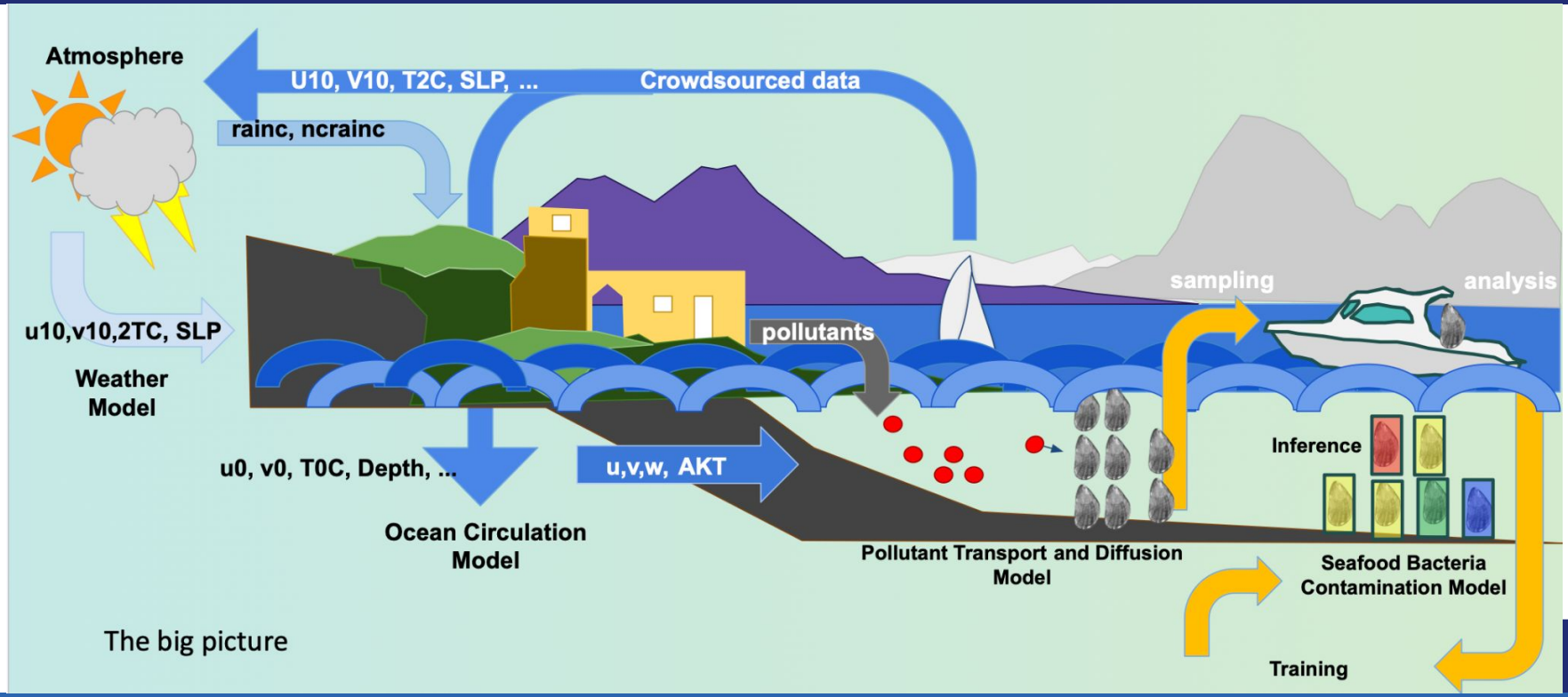
- Sewage discharges (continuous or discontinuous)
- Farmed animals
- Wildlife shipping.

Transported and diffused by sea currents.

Challenges:
High-resolution weather forecasts
Accurate bathymetric model
High-resolution numerical model
Initial conditions.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

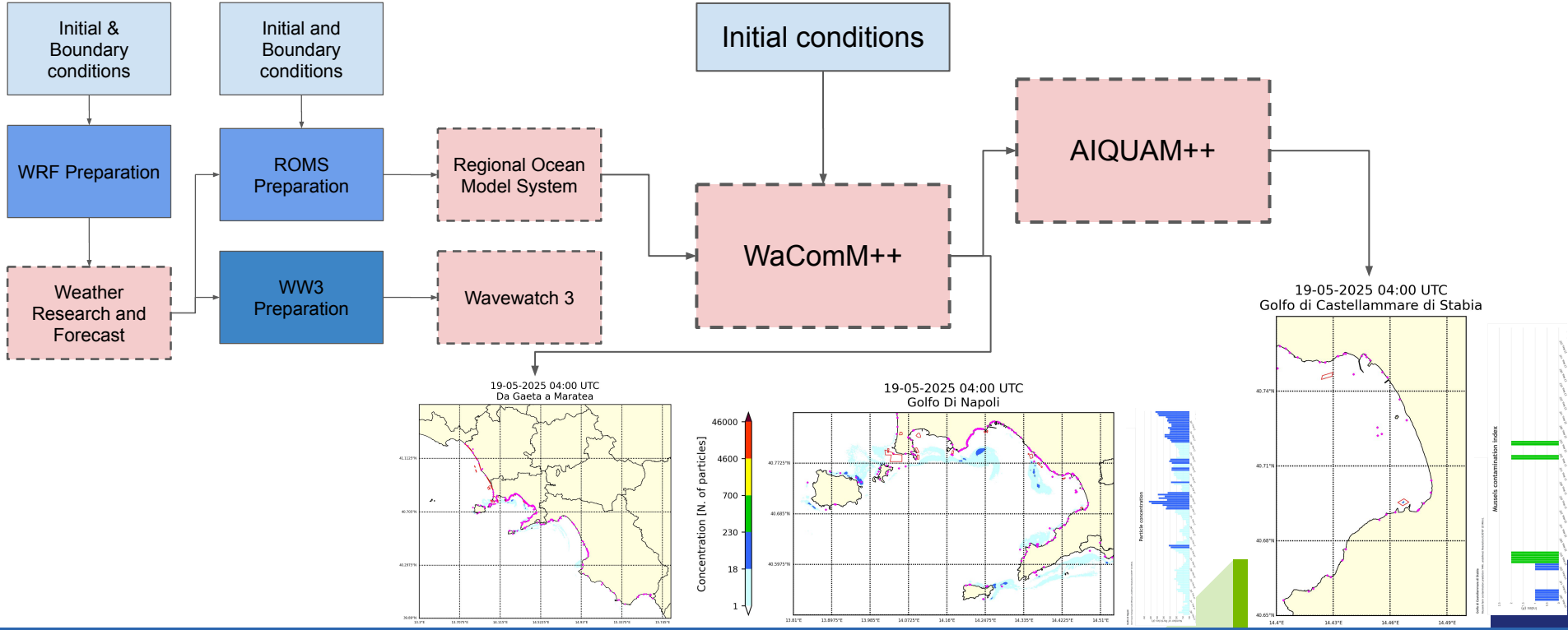
The big picture



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

MytilEx

https://meteo.uniparthenope.it/mytilex/



Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

MytilX (on demand)

Initial & Boundary conditions

Initial and Boundary conditions

Initial conditions

WRF Preparation

ROMS Preparation

Regional Ocean Model System

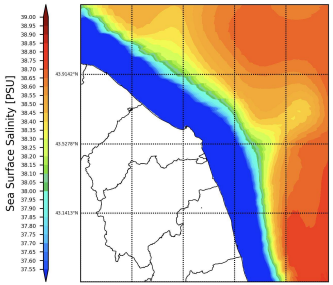
AIQUAM++

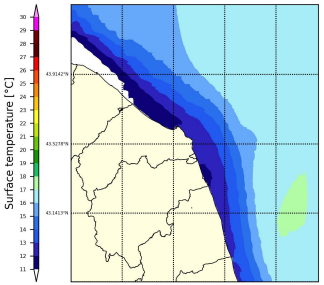
Weather Research and Forecast

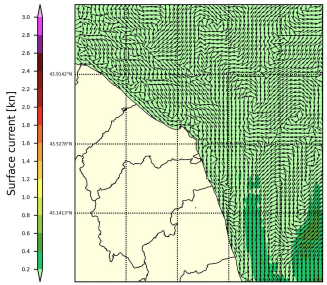
WW3 Preparation

Wavewatch 3

WaComM++







Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

SmokeTracer (on demand)

https://smoketracer.uniparthenope.it/

A numerical model-based on demand workflow HPC application to forecast the smoke plume pattern during a wildfire

REGIONE CAMPANIA

orsa
Osservatorio Regionale Sicurezza Alimentare

Università degli Studi di Napoli Parthenope

Istituto Zooprofilattico Sperimentale del Mezzogiorno

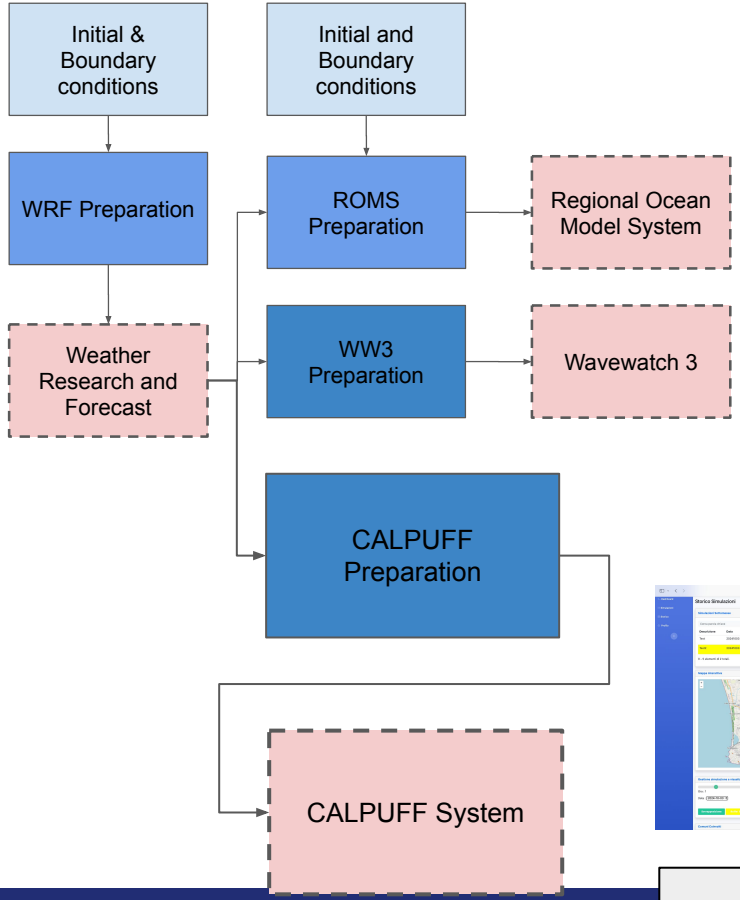
Finanziato dall'Unione europea NextGenerationEU

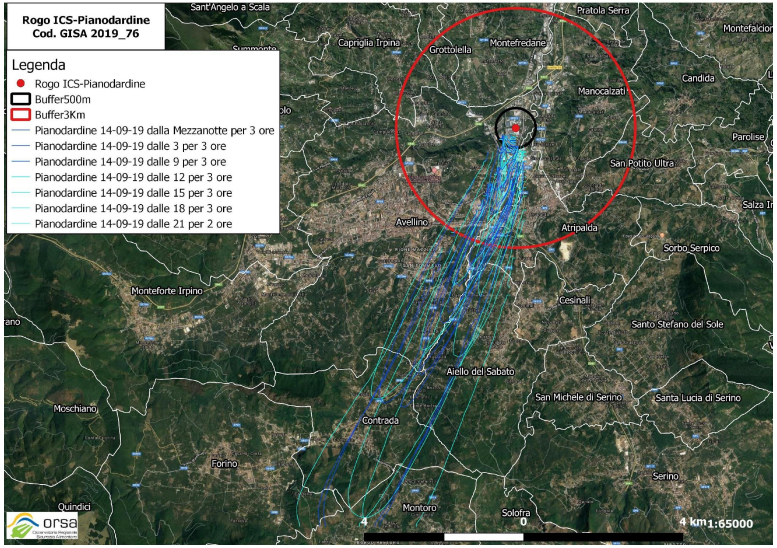
Ministero dell'Università e della Ricerca

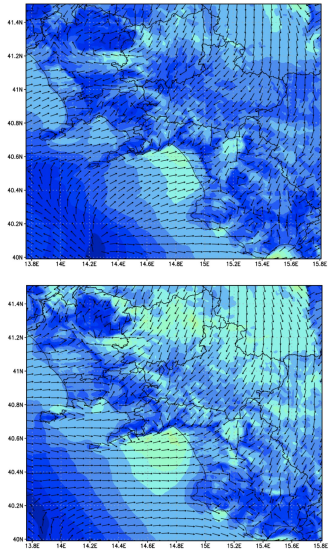
Italiadomani
Piano Nazionale di Ripresa e Resilienza

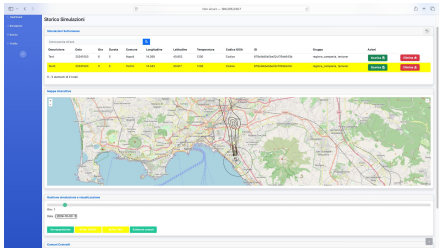
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

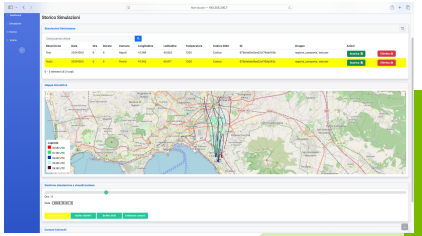
SmokeTracer (on demand)

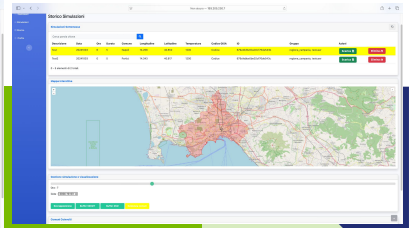












Buffer 500 mt - 3 km

Simulation evolution

Highlights municipalities

Let me to introduce the hero of the day…

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Direct Acyclic Graphs as parallel jobs on anything

DAGonStar is a production-oriented workflow engine:

- **Integrated** in the Python environment.
- **Minimal** footprint for external software components execution.
- **Avoiding** any workflow engine **centered data management**.
- **Checkpoints** for failover and execution recovery.
- **Straightforward** definition of tasks:
  - Python scripts.
  - Web interaction.
  - External software components.
- **Execution** sites **independence**:
  - Local / scheduler (SLURM).
  - Containers (Docker).
  - Clouds (AWS, OpenStack, DigitalOcean).

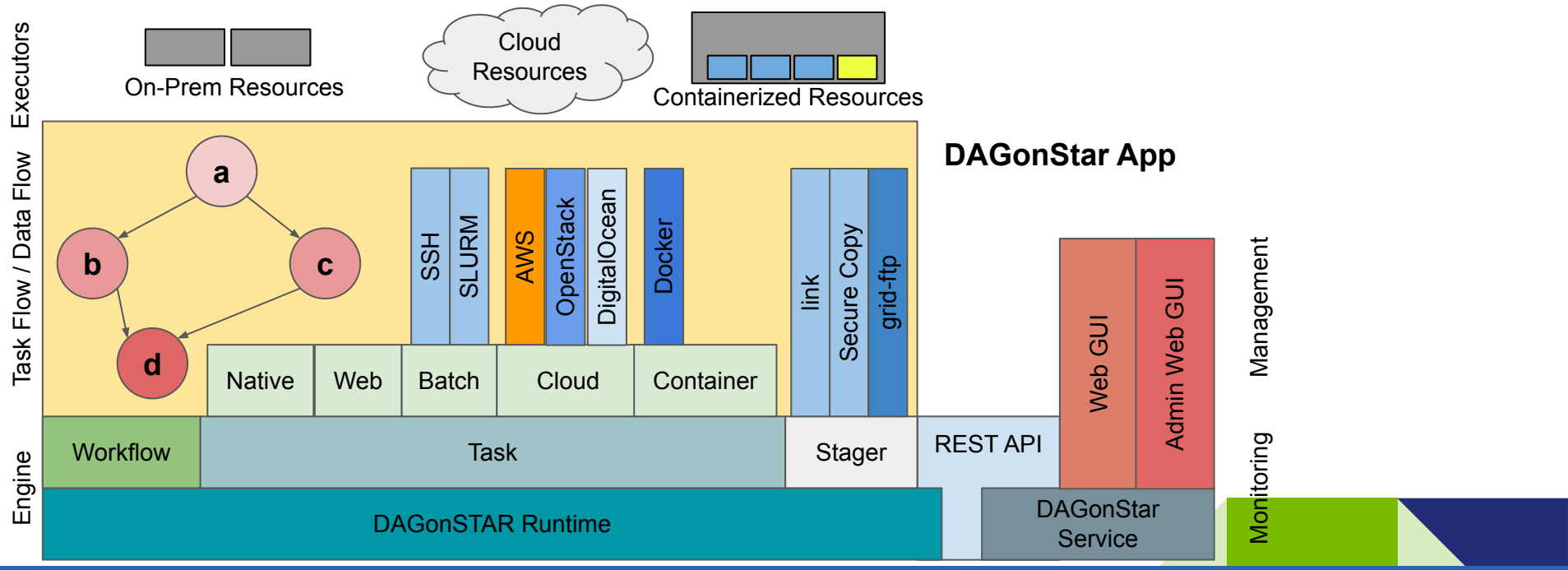
https://github.com/dagonstar/

Named after the Phoenician god-fish *Dagon* known by ancient Greeks as *Triton*.

NB: The Star (*) symbol is the wildcard for **anything**.

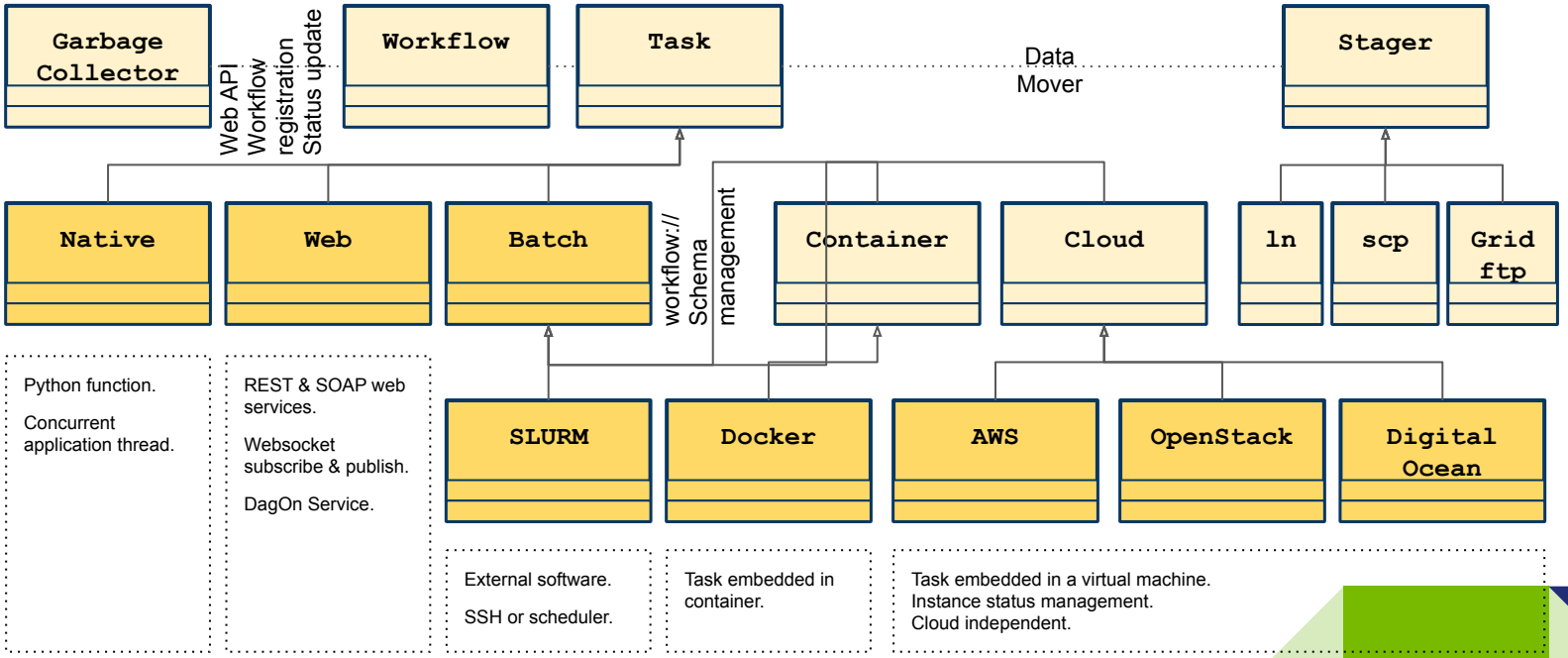
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Architecture



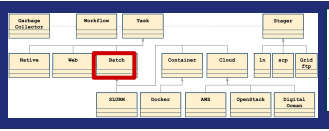
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Components

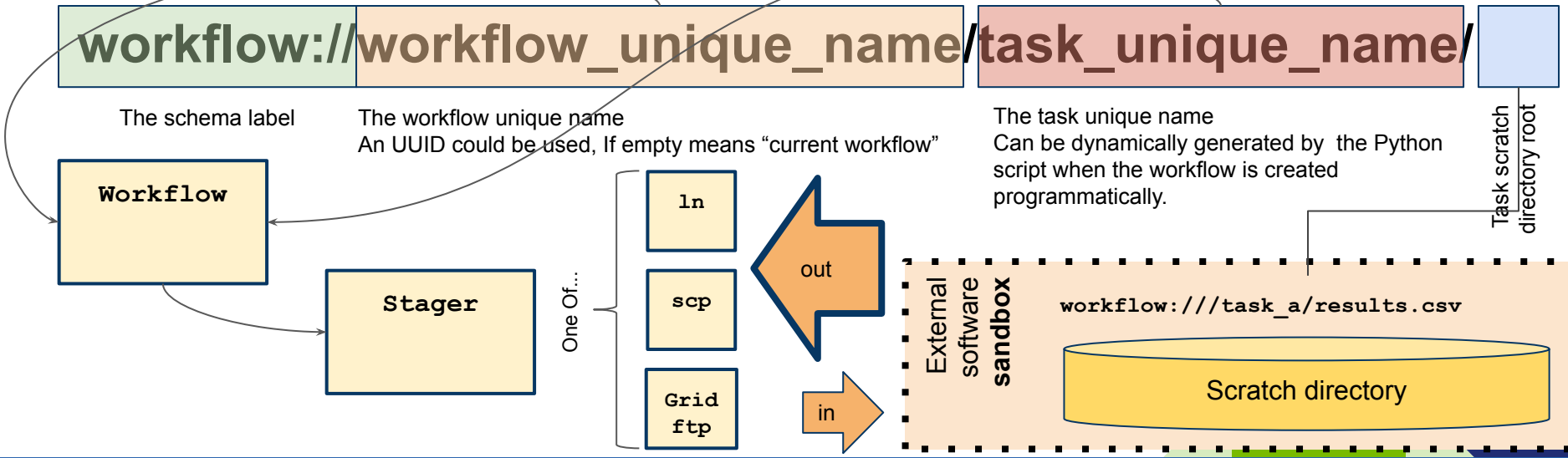


Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

workflow:// schema

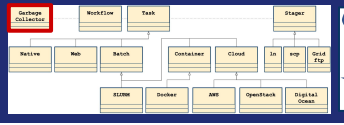


The **Batch** component takes charge of the management of data dependencies using the **workflow:// schema**.



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

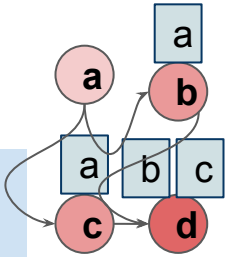
Garbage Collector



- Tracks the storage and computational resources allocated during tasks execution.
- Proceeds to dispose them when no longer needed.

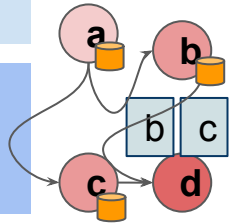
Make Dependencies

```
For each batch task in the <workflow> ...
     For each workflow://<workflow>/<task>/ reference in the task command line ...
          Increment the number of reference to <task>
```



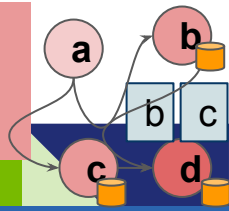
On Task Finish

```
For each workflow://<workflow>/<task>/ reference in the task command line ...
     Decrement the number of reference to <task>
     If the number of reference to <task> is 0, clean up the involved resource
```



Clean Up

Local, remote or shared file system: Remove the scratch directory.	**Virtual machine instance:** Stop the instance.	**Container:** Stop the container.

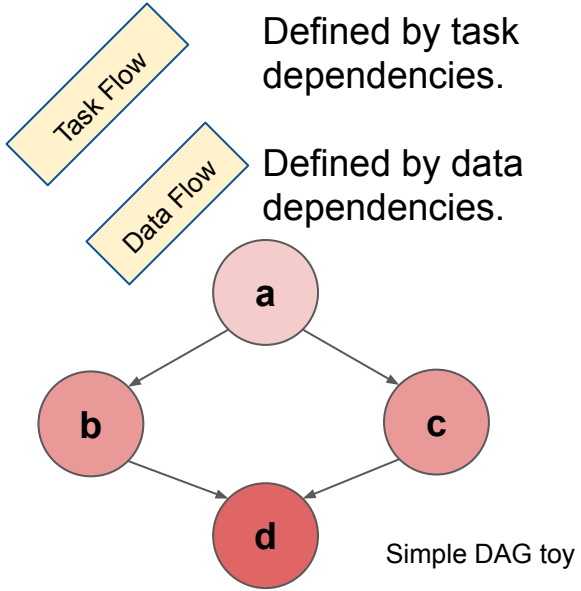


Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Programming Model

Python Script: “DAGonStar Hello World App”

```
import dagon
…
workflow=Workflow(“myapp”,settings)
workflow.add_task(new Task(“a”,”...”))
workflow.add_task(new Task(“b”,”workflow:///a”))
workflow.add_task(new Task(“c”,”workflow:///a”))
workflow.add_task(new Task(“d”,”workflow:///b workflow:///c”))
workflow.run()
sys.exit(0)
```



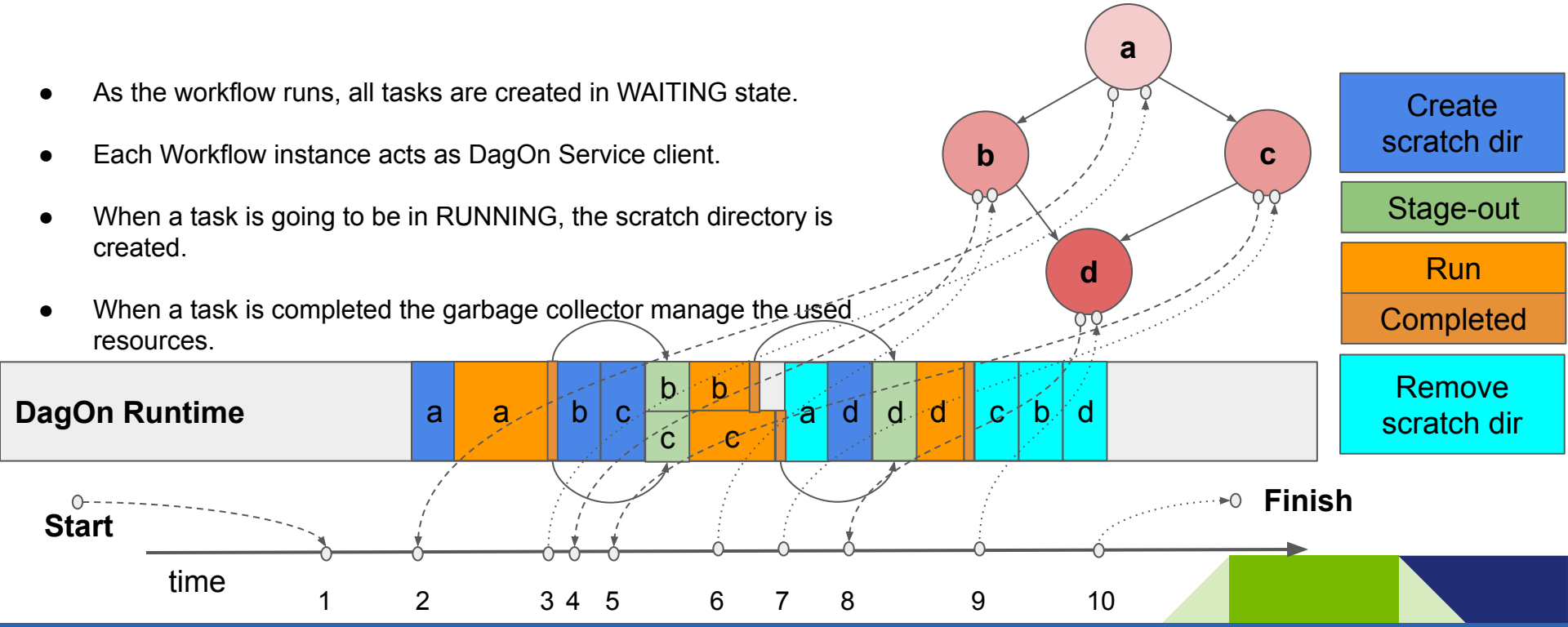
Simple DAG toy

- Dealing with actual data files instead of high-level defined datasets.
- Performing backward data references in order to create dependencies.
- Having more Workflow instances in the same Python application.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Application Lifecycle

- As the workflow runs, all tasks are created in WAITING state.
- Each Workflow instance acts as DagOn Service client.
- When a task is going to be in RUNNING, the scratch directory is created.
- When a task is completed the garbage collector manage the used resources.

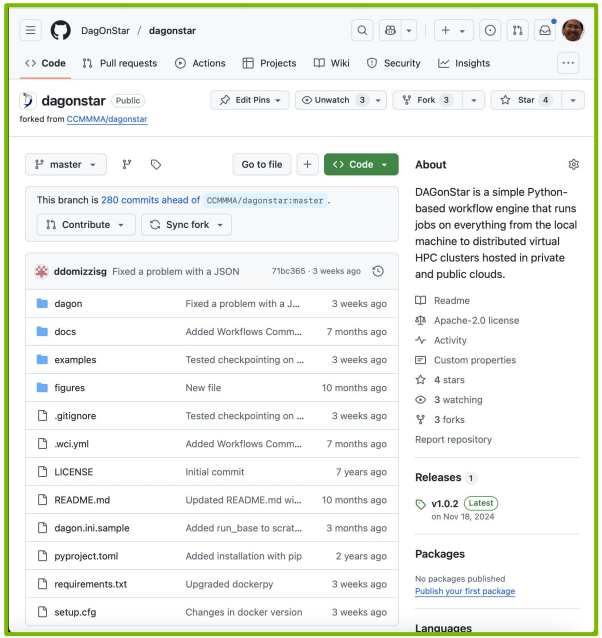


Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

It’s tutorial time!

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Docs and tutorials



How to install

```
git clone https://github.com/DagOnStar/dagonstar.git
cd dagonstar
python3 -m venv venv
. venv/bin/activate
pip install -r requirements.txt
export PYTHONPATH=$PWD:$PYTHONPATH
```

Demo

```
cp dagon.ini.sample examples/dagon.ini
cd examples/dataflow/batch
python dataflow-demo.py
```

https://github.com/dagonstar

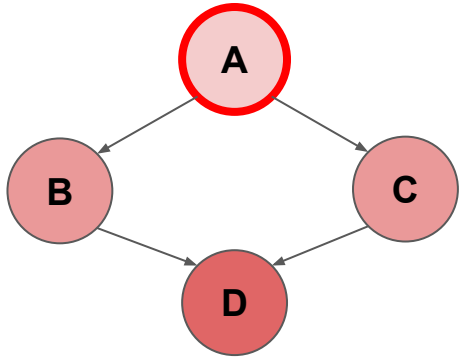
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

dataflow-demo.py (1/3)

```
from dagon import Workflow
from dagon.task import DagonTask, TaskType

# Check if this is the main
if __name__ == '__main__':
     # Create the orchestration workflow
     workflow=Workflow("DataFlow-Demo")

     # The task a
     taskA = DagonTask(TaskType.BATCH, "A", "mkdir output; hostname > output/f1.txt")
```



https://github.com/DagOnStar/dagonstar/tree/master/examples/dataflow/batch

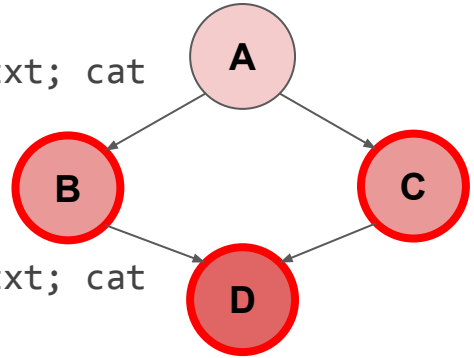
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

dataflow-demo.py (2/3)

```
      # The task b
      taskB = DagonTask(TaskType.BATCH, "B", "echo $RANDOM > f2.txt; cat
workflow:///A/output/f1.txt >> f2.txt")

      # The task c
      taskC = DagonTask(TaskType.BATCH, "C", "echo $RANDOM > f2.txt; cat
workflow:///A/output/f1.txt >> f2.txt")

      # The task d
      taskD = DagonTask(TaskType.BATCH, "D", "cat workflow:///B/f2.txt >> f3.txt; cat
workflow:///C/f2.txt >> f3.txt")
```



https://github.com/DagOnStar/dagonstar/tree/master/examples/dataflow/batch

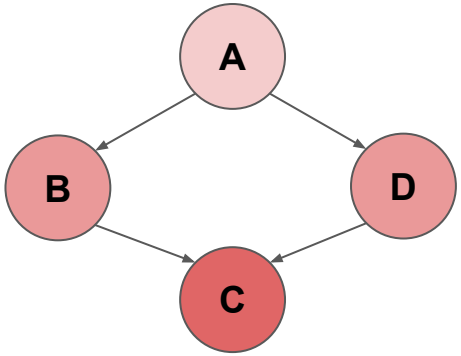
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

dataflow-demo.py (3/3)

```
# add tasks to the workflow
workflow.add_task(taskA)
workflow.add_task(taskB)
workflow.add_task(taskC)
workflow.add_task(taskD)

workflow.make_dependencies()

# run the workflow
workflow.run()
```



https://github.com/DagOnStar/dagonstar/tree/master/examples/dataflow/batch

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Tasks as Docker containers
…made easy…

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

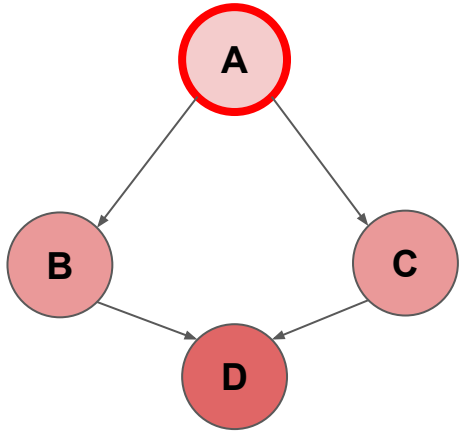
dataflow-demo-docker-remote.py

```
from dagon import Workflow
from dagon.task import DagonTask, TaskType

# Check if this is the main
if __name__ == '__main__':
     # Create the orchestration workflow
     workflow=Workflow("DataFlow-Demo")

     # The task a
     taskA = DagonTask(TaskType.DOCKER, "A", "mkdir output;hostname > output/f1.txt",
image="ubuntu:latest", ip="", ssh_username="")
```

DOCKER
APPTAINER
KUBERNETES



https://github.com/DagOnStar/dagonstar/blob/master/examples/dataflow/docker

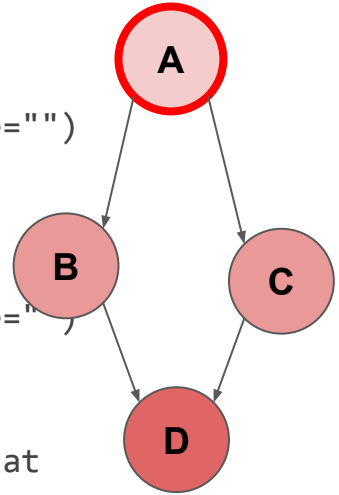
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

dataflow-demo-docker-remote.py

```
    # The task b
    taskB = DagonTask(TaskType.APPTAINER, "B", "echo $RANDOM > f2.txt; cat
workflow:///A/output/f1.txt >> f2.txt", image="ubuntu:latest", ip="", ssh_username="")

    # The task c
    taskC = DagonTask(TaskType.KUBERNETES, "C", "echo $RANDOM > f2.txt; cat
workflow:///A/output/f1.txt >> f2.txt", image="ubuntu:latest", ip="", ssh_username="")

    # The task d
    taskD = DagonTask(TaskType.DOCKER, "D", "cat workflow:///B/f2.txt >> f3.txt; cat
workflow:///C/f2.txt >> f3.txt", image="ubuntu:latest", ip="", ssh_username="")
```

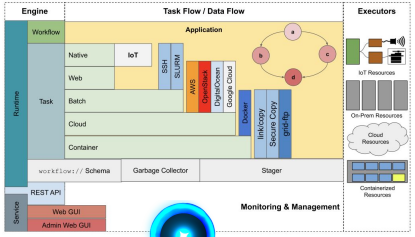


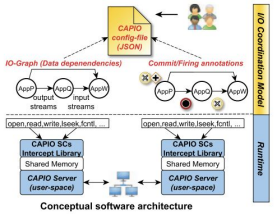
https://github.com/DagOnStar/dagonstar/blob/master/examples/dataflow/docker

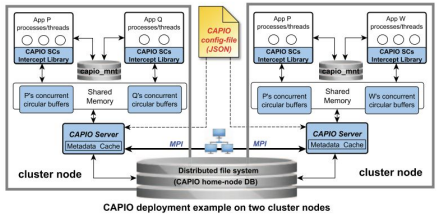
The extended universe

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

DAGonStar







HPSC LAB
HIGH-PERFORMANCE SCIENTIFIC COMPUTING LABORATORY

DagOn CAPIO
Directed Acyclic Graph on Cross-Application Programmable I/O

- DAGonStar batch tasks generate the JSON file based on the dependencies between tasks, identified thanks to the workflow:// Schema;
- The JSON file is used by the CAPIO server for configuration;
- Tasks make up a pipeline in which one produces files and the other reads them;
- Posix calls made on these output files will be intercepted by the CAPIO server, allowing it to process this data in RAM.

Partners

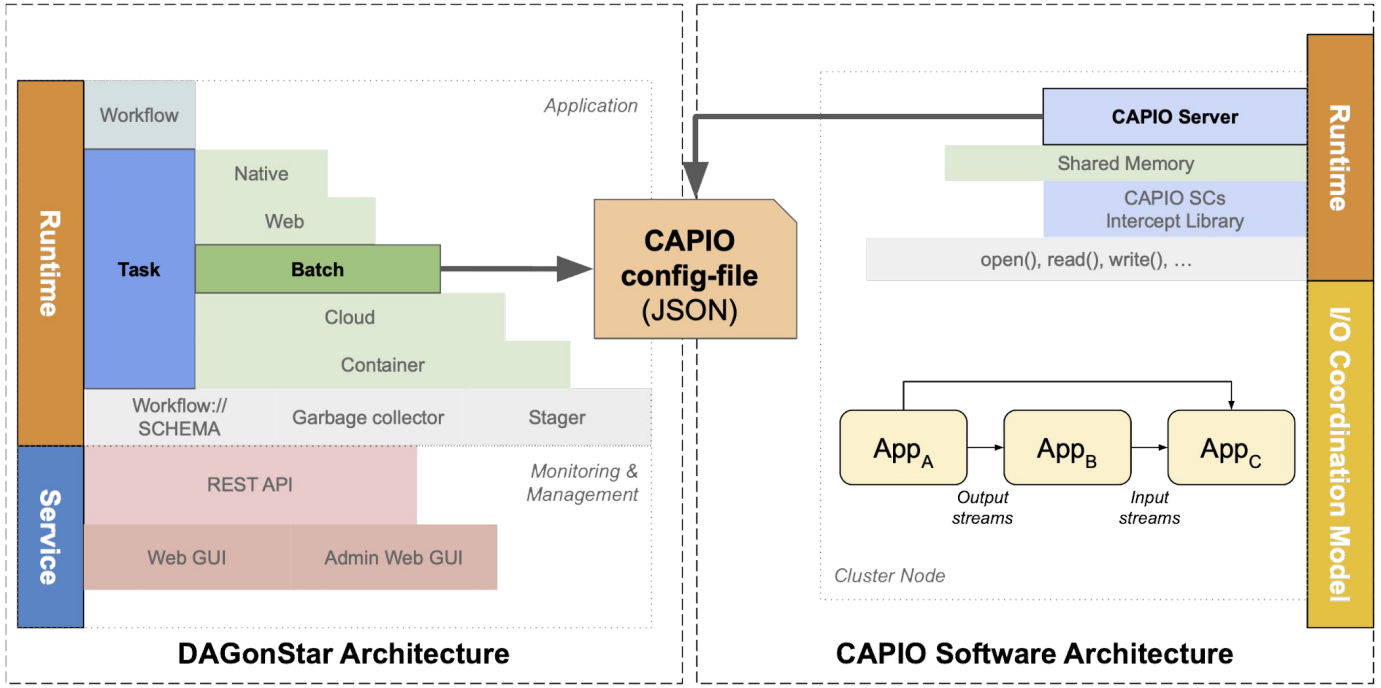
UNIVERSITÀ DEGLI STUDI DI NAPOLI PARTHENOPE

F I D E S

ITDM your technology partner

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

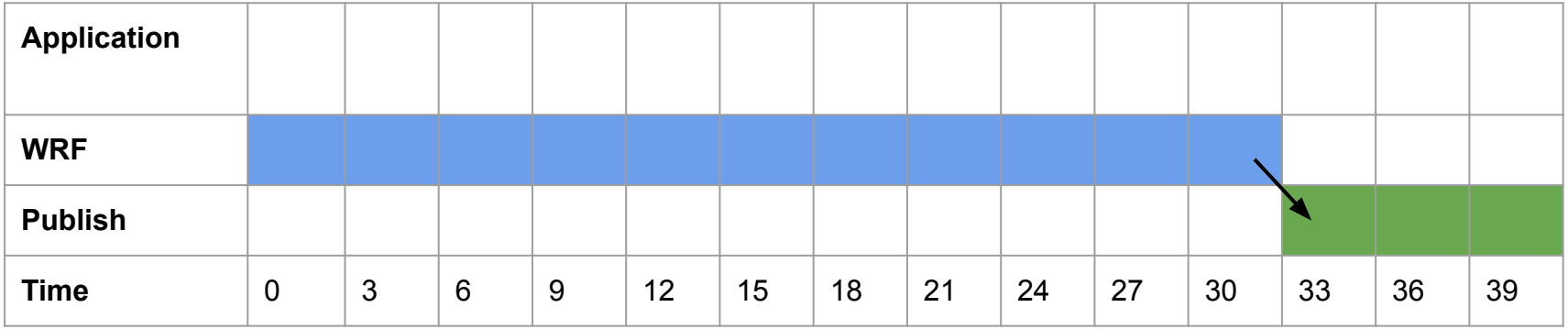
DAGonStar and CAPIO integration



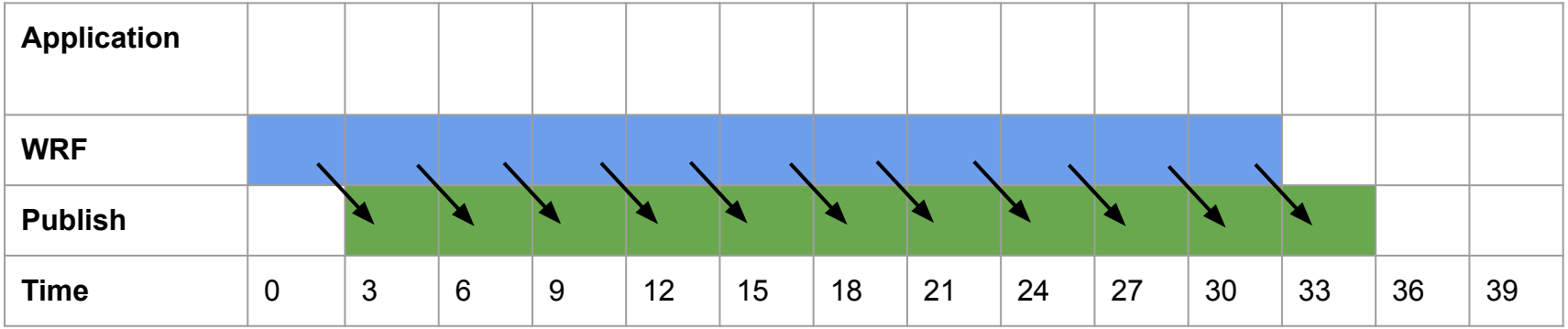
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

DAGonStar and CAPIO integration

DAGonStar



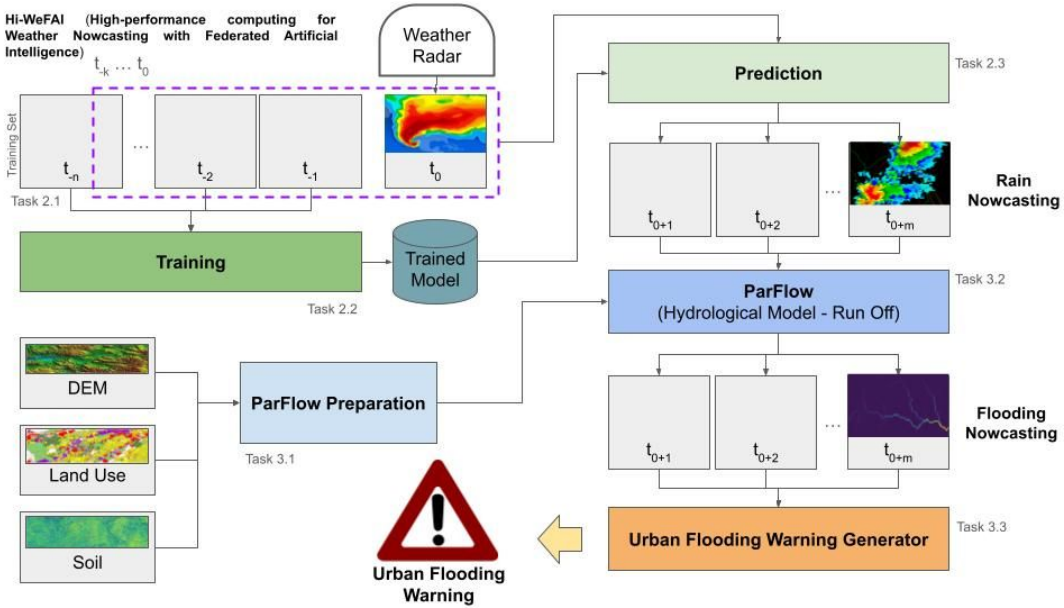
DAGonCAPIO

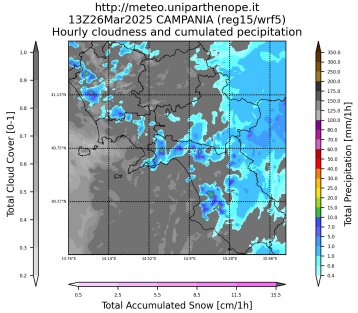


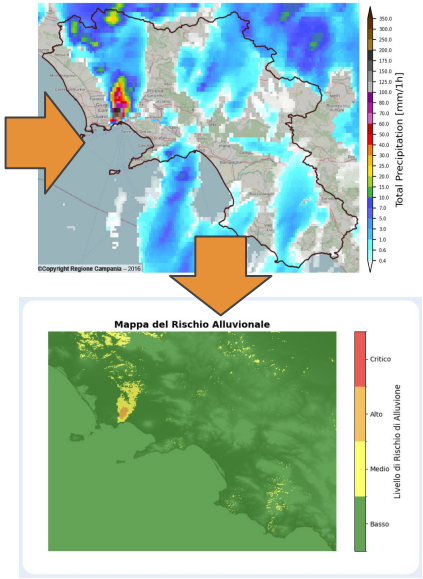
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Hi-WeFAI

https://hiwefai-project.org







HPC-based **application** for Weather nowcasting with Federated AI

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

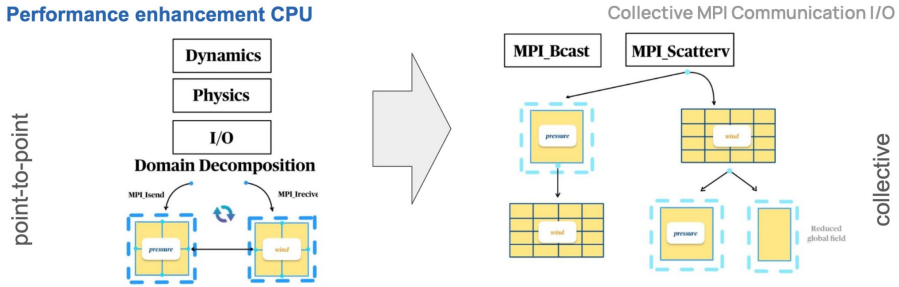
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

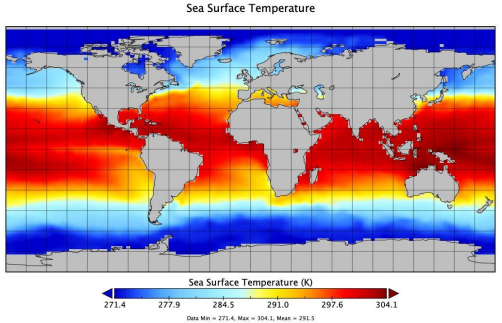
GLOWPP

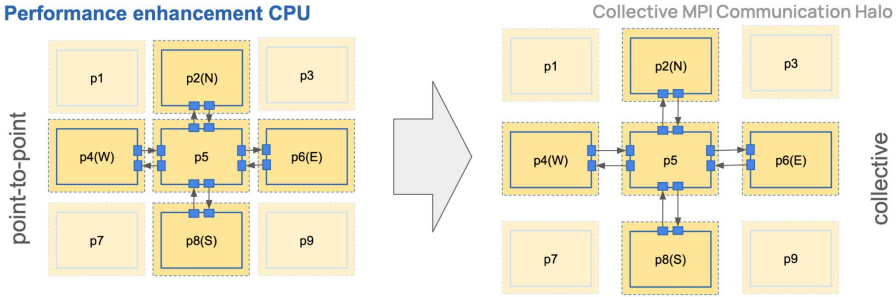
https://glowpp-project.org

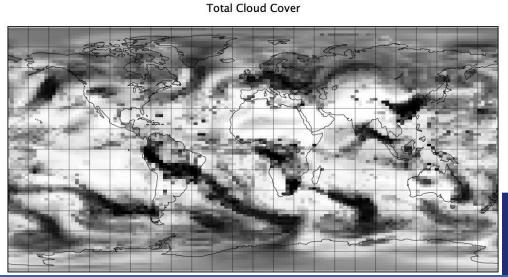
Refactoring, Optimization, and “Production-level” of GLOBO model developed by 'ISAC-CNR as an evolution of BOLAM (BOlogna Limited-Area Model).

https://glowpp-project.org









Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Time to wrap up

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Contributors

DAGonStar is open-source and open to the workflow community

Prof. José Luis Gonzalez Compean

Dr. Dante D. Sánchez Gallegos
Prof. Javier Garcia Blas

Prof. Iacopo Colonnelli
Marco Santimaria

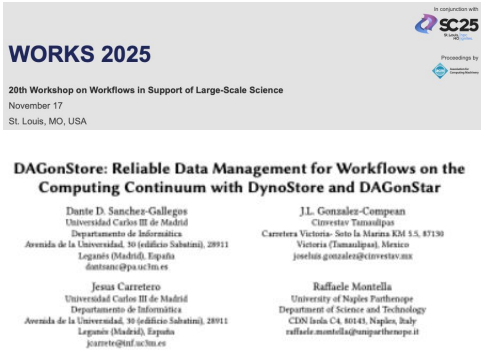
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Conclusion…

…and future directions:

- Develop DAGonStar for HPC+AI environmental applications
- Make DAGonFS as stable and reliable ad-hoc file system
- Extend the workflow:// schema to support data streaming as staging system.
- Deploy DAGonStar use cases on public cloud to democratize open-source tools and open-data products.

https://works-workshop.org/



https://www.escience-conference.org/2026/

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing