Workflow management tools exploitation at CMCC

Alessandro D'Anca
Sandro Fiore
CMCC Advanced Scientific Computing Division
Prof. Giovanni Aloisio
CMCC Strategic Council & Director of the CMCC SCC

3rd ENES Workshop on Workflows funded by ESiWACE
Brussels, 13-14 September 2018

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

enes
EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

Outline

- **The Ophidia Project**
- **SWE based indicators** (CLIP-C Project)
- **Multi-model analysis** by means of a centralized architecture (ESiWACE Project)
  - **ECAS** in the context of EOSC-hub
- **GLOB16** analysis at CMCC
- Workflow in **operational environment**: the OFIDIA2 Project
- **Videos**

The Ophidia project

Ophidia (http://ophidia.cmcc.it) is a CMCC Foundation research project addressing big data challenges for eScience

It provides support for declarative, parallel, server-side data analysis exploiting parallel computing techniques and database approaches

Exploits a multidimensional data model providing the data cube abstraction for access and analysis of scientific n-dimensional data

Ophidia

cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

Ophidia in a nutshell

- ✔ ***Big data stack for scientific data analysis***
- ✔ ***Features:*** *time series analysis (array-based analysis), data subsetting (by value/ index), data aggregation, model intercomparison, OLAP, etc.*
- ✔ *Use of parallel operators and parallel I/O*
- ✔ ***Support for complex workflows***
- ✔ *Extensible:* ***simple API*** *to support framework extensions like new operators and array-based primitives*
  - ✔ *currently 50+ operators and 100+ primitives provided*
- ✔ ***Multiple interfaces*** *available (WS-I, GSI/VOMS, OGC-WPS).*
- ✔ *Programmatic access via C and* ***Python APIs***
- ✔ *Support for both* ***batch & interactive*** *data analysis*
- ✔ *Command line interpreter for submitting operators.*

SWE – Snow Water Equivalent

✔ Snow on

First day of snow season with SWE>0

✔ 33 files* 6 MB = 198 MB

✔ Snow off

Day with SWE=0 after last day with SWE>0

✔ 33 files * 6 MB = 198 MB

✔ Lenght of snow season

(Snow off) – (Snow on)

✔ 33 files* 6 MB = 198 MB

Input dataset

✔ SWE Observation obtained through SSM/I Sensor

✔ 6341 nc files

✔ 50 GB of input data

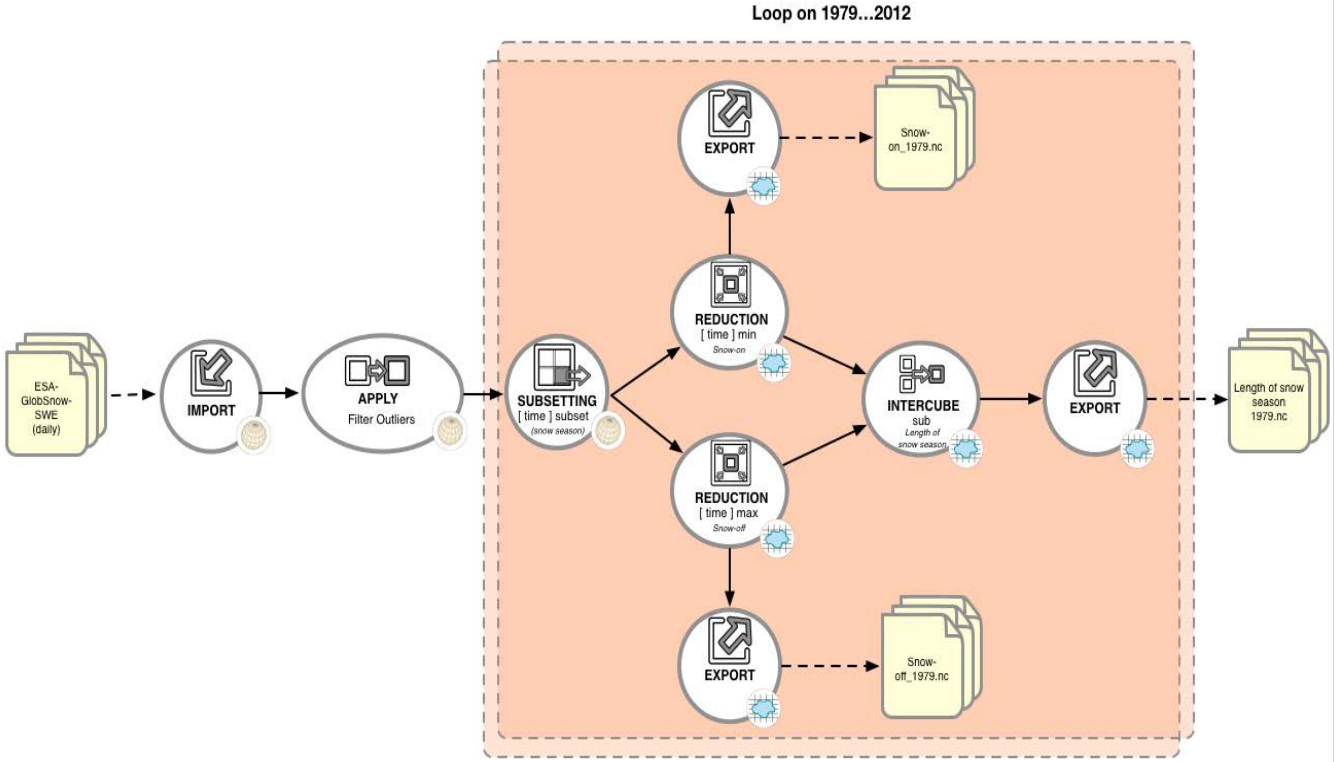
✔ 25Km x 25Km spatial resolution

✔ Latitude 35N - 85N; Longitude 180W - 180E

✔ time resolution: daily on 33 years (09/1979 – 12/2012)

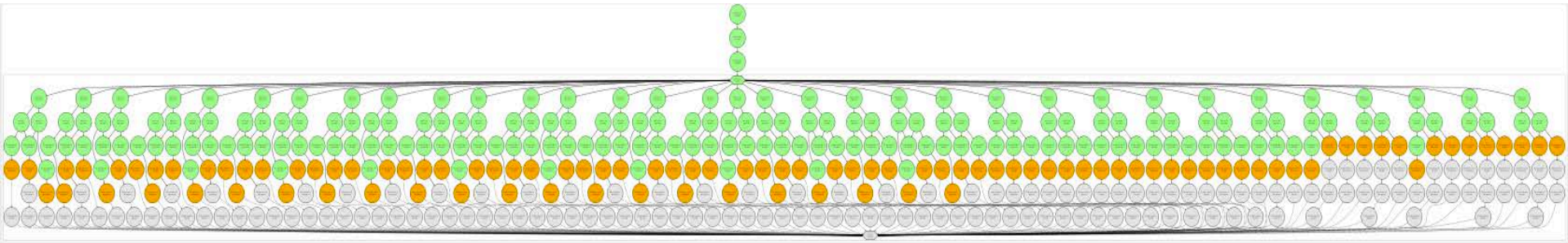
Climate Information Portal

Snow on, snow off and Length of snow season

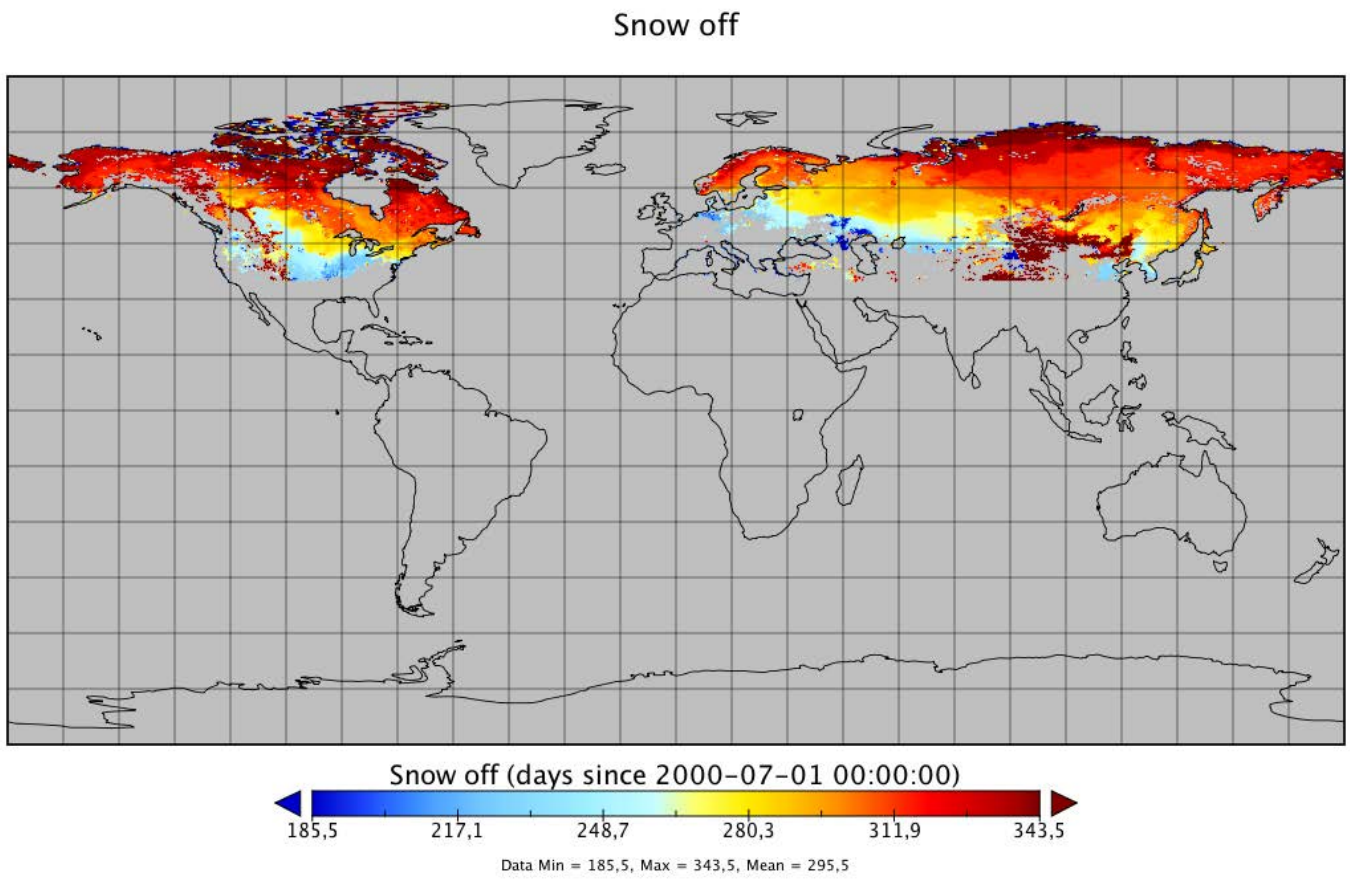


Ophidia workflow

- ✔ 599 tasks
- ✔ Defined a loop on all the cosidered years
- ✔ 99 output files (3 for each year)
- ✔ 594 MB outputs
- ✔ 3 indicators produced

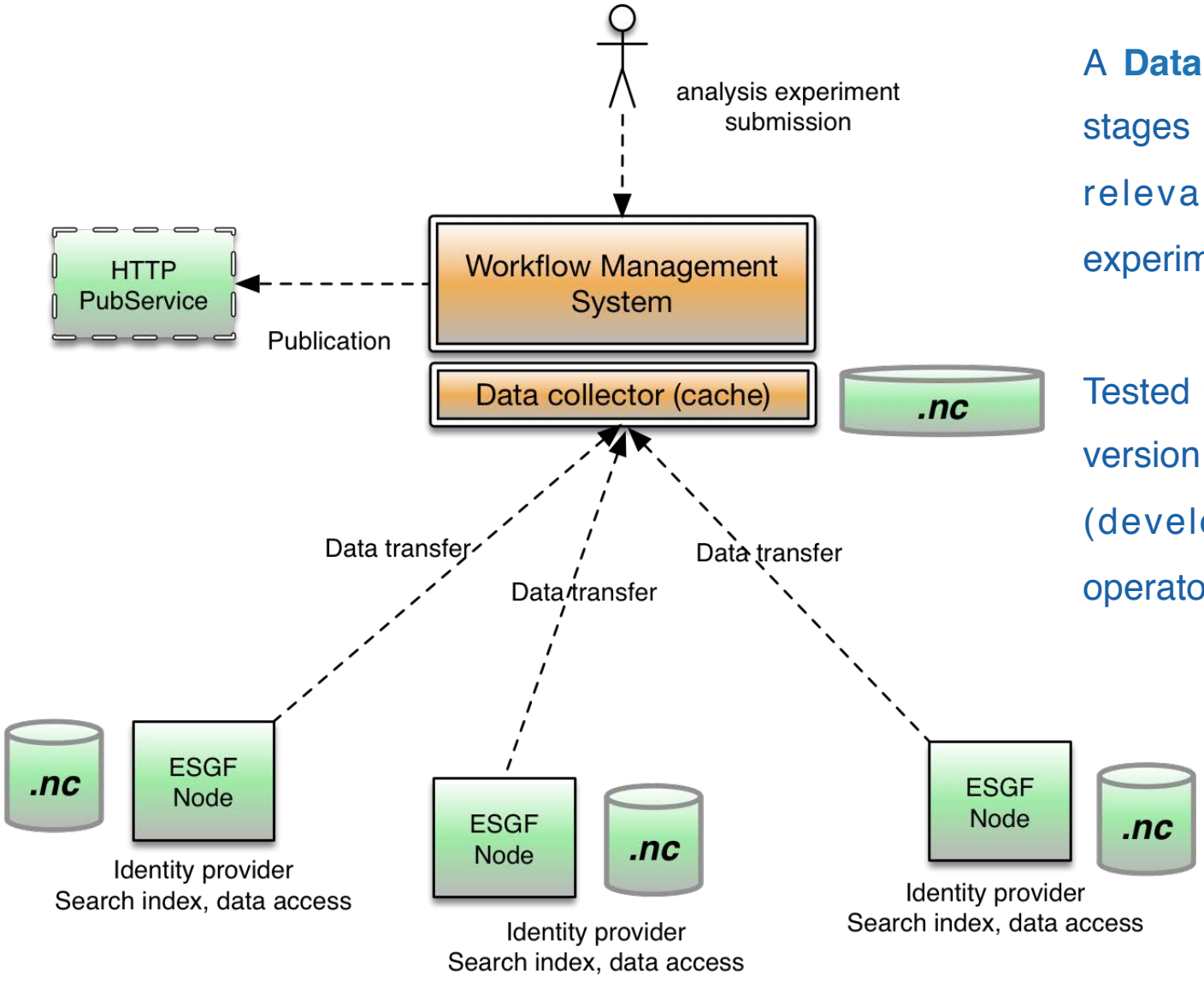


Snow on/off – Length of snow season



Outputs are stored on a File Server to allow downloading and visualization

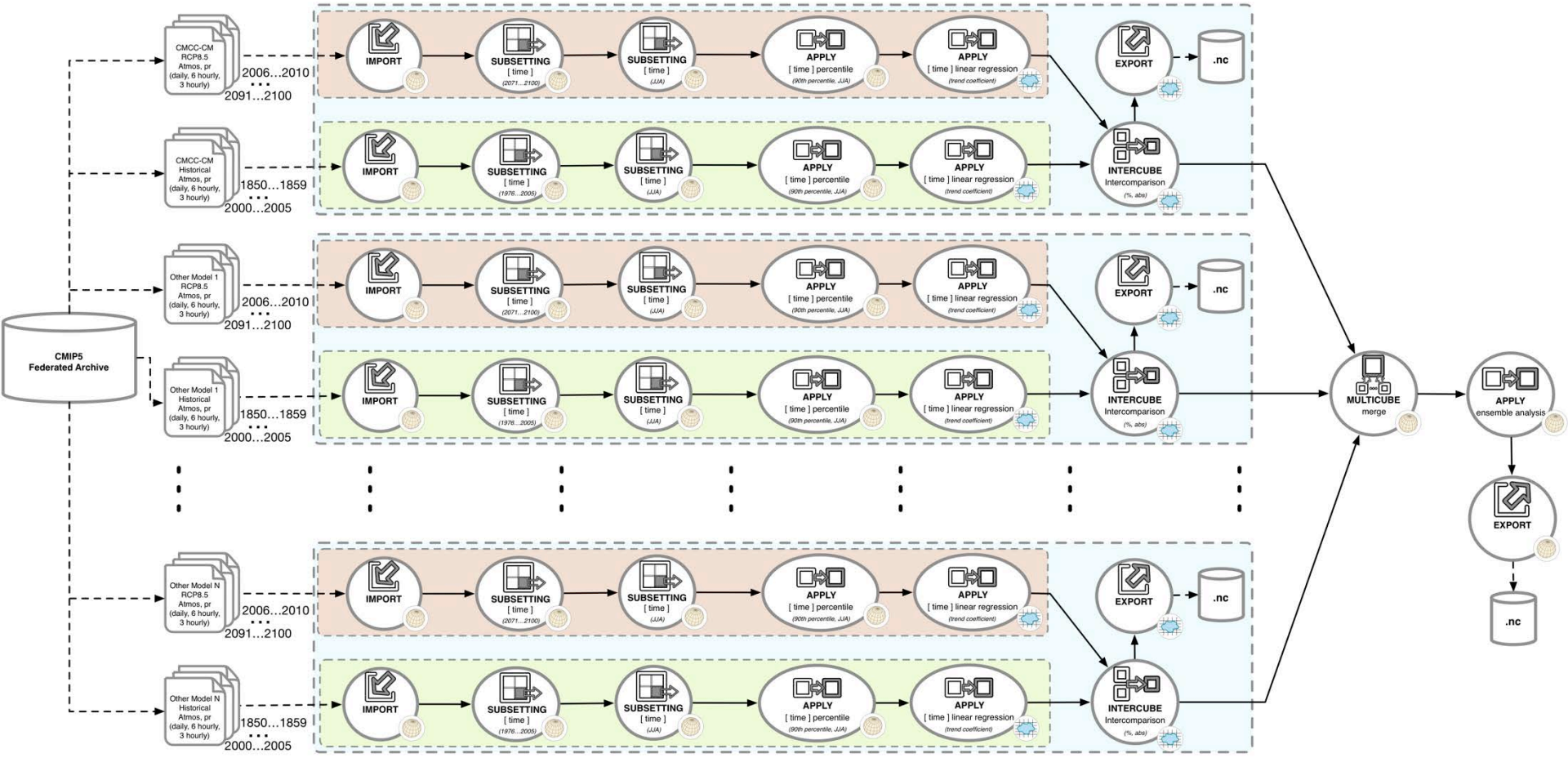
Multi-model analysis: centralized architecture



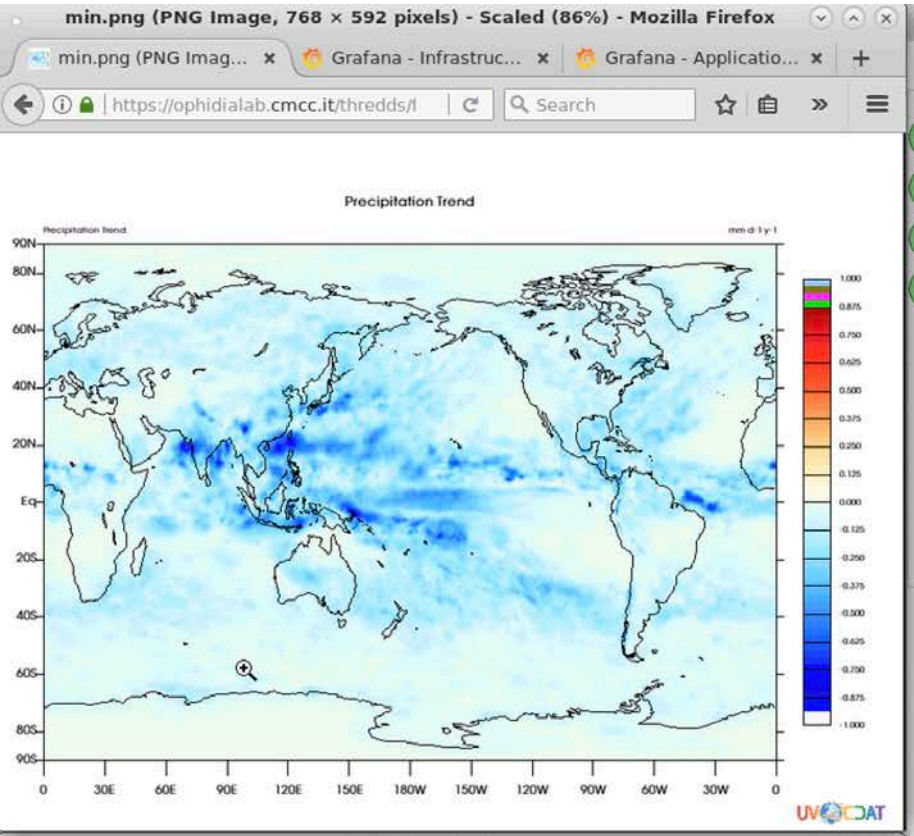
A **Data Collection Layer** pre-stages and caches the data relevant for the analysis experiment.

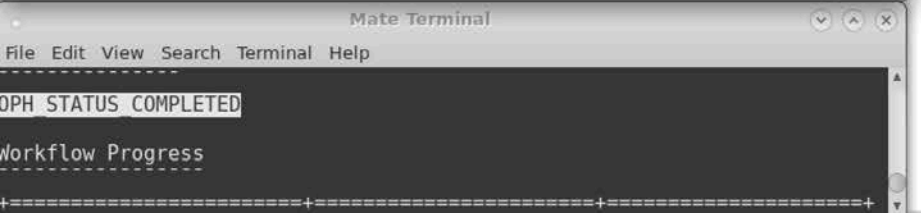
Tested through the multi-model version of the **PTA** test case (developed using Ophidia operators).

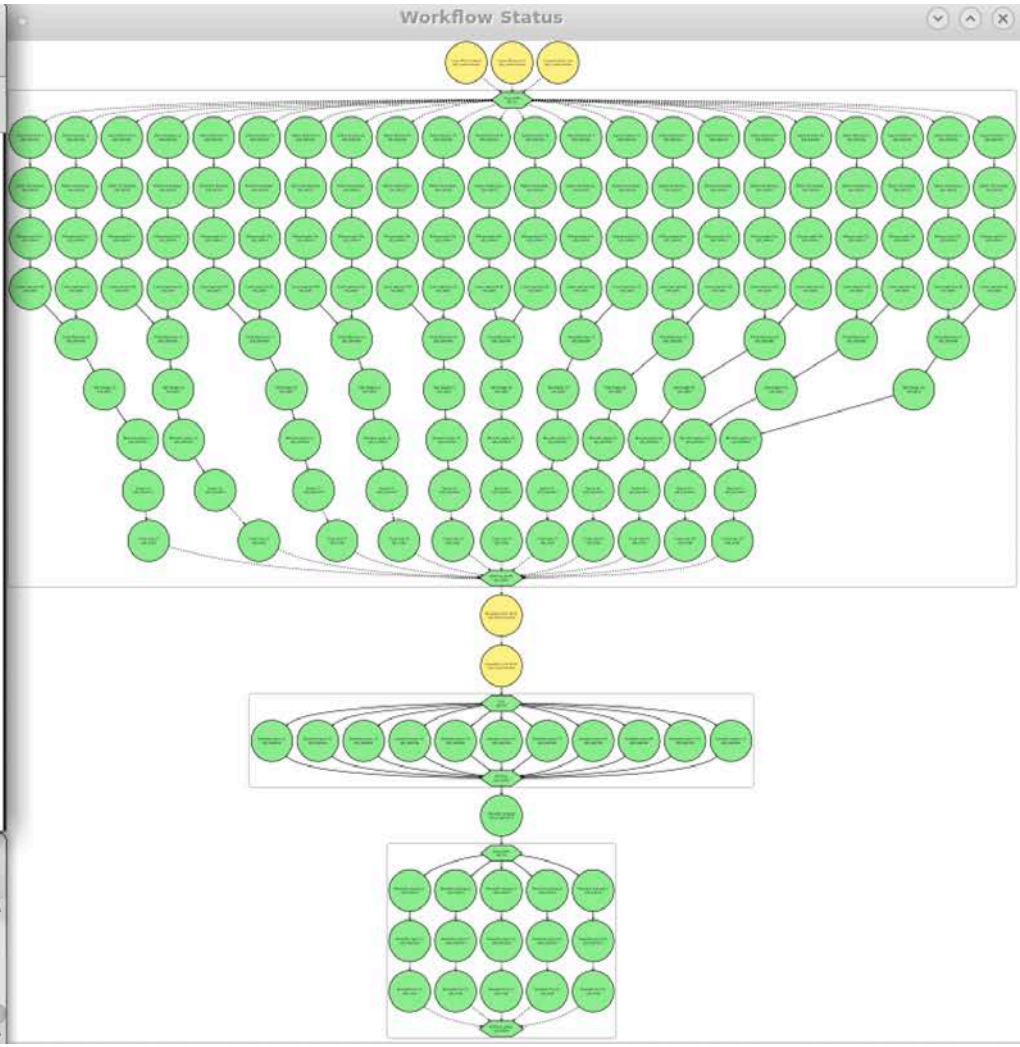
Multi-model Precipitation Trend Analysis diagram



Runtime and output

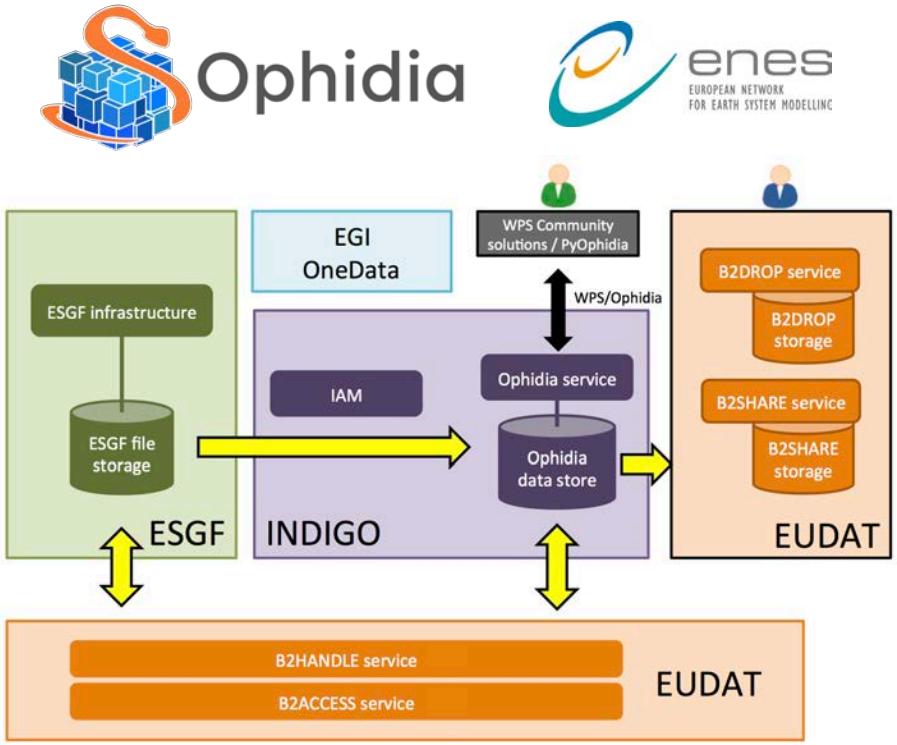






ENES Climate Analytics Service (ECAS)

- ✓ The **ENES Climate Analytics Service (ECAS)**, proposed by CMCC & DKRZ in EOSC-hub supports climate data analysis
- ✓ It is one of the **EOSC-Hub Thematic Services** and has been ranked as the **1st out of 64** Thematic Service proposals
- ✓ ECAS builds on top of the **Ophidia big data analytics framework** with components from INDIGO-DataCloud, EUDAT and EGI



DKRZ DEUTSCHES KLIMARECHENZENTRUM

cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici

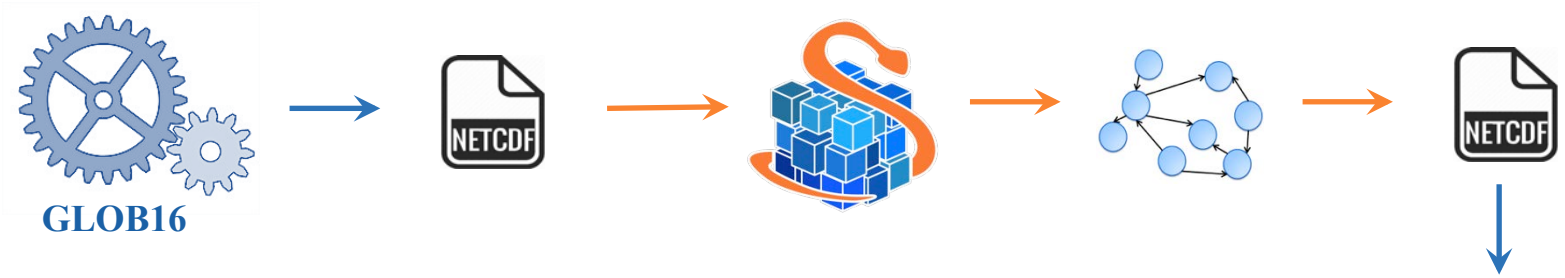
EOSC-hub

Earth System Grid Federation ESGF

The European Commission launched the European Open ScienceCloud Initiative to capitalise on the data revolution. EOSC will provide European science, industry and public authorities with world-class digital infrastructure that bring state of the art computing and data storage capacity to the fingertips of any scientists and engineer in the EU.

EOSC-hub receives funding from the EU's Horizon 2020 research and innovation programme under grant agreement No. 777536.

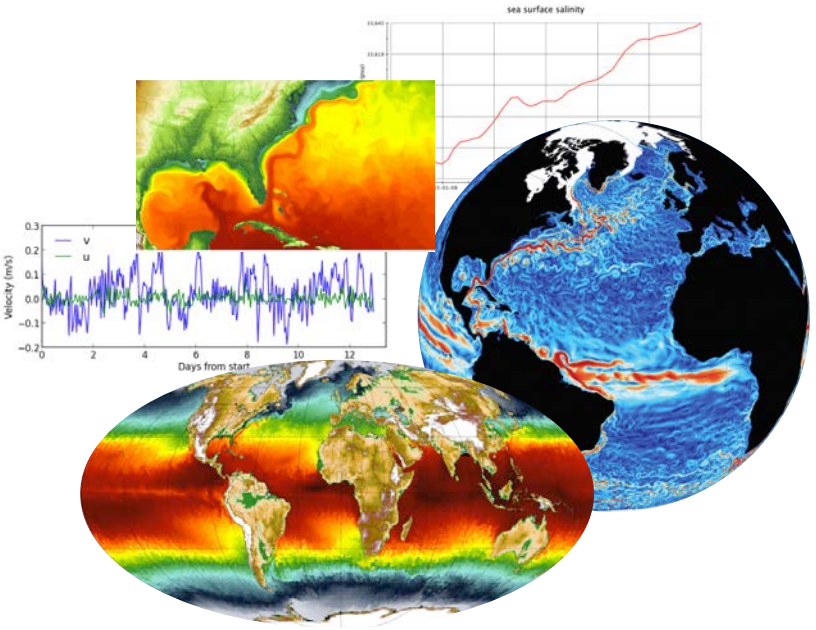
Real time indicators computation on GLOB16



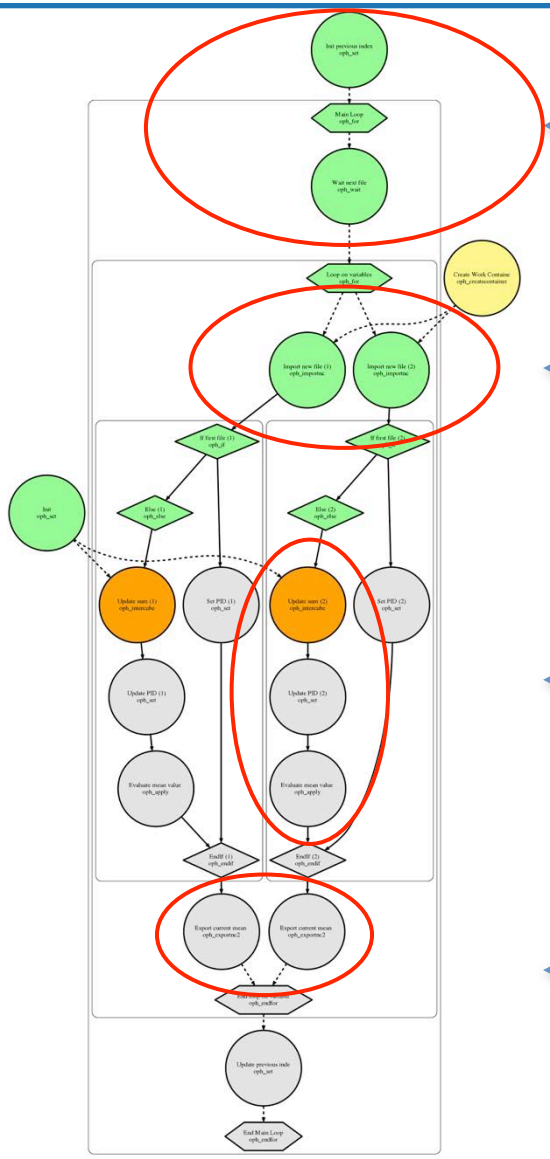
Variables: sea surface temperature and salinity
Outputs: Mean values + Timeseries on spatial mean
Domain: Global + spatial subdomain

Offline mode
Ophidia workflow run as “batch” application

On line mode
Ophidia workflow triggered by GLOB16 output – Ophidia and GLOB16 run concurrently



Ophidia Workflow Analysis



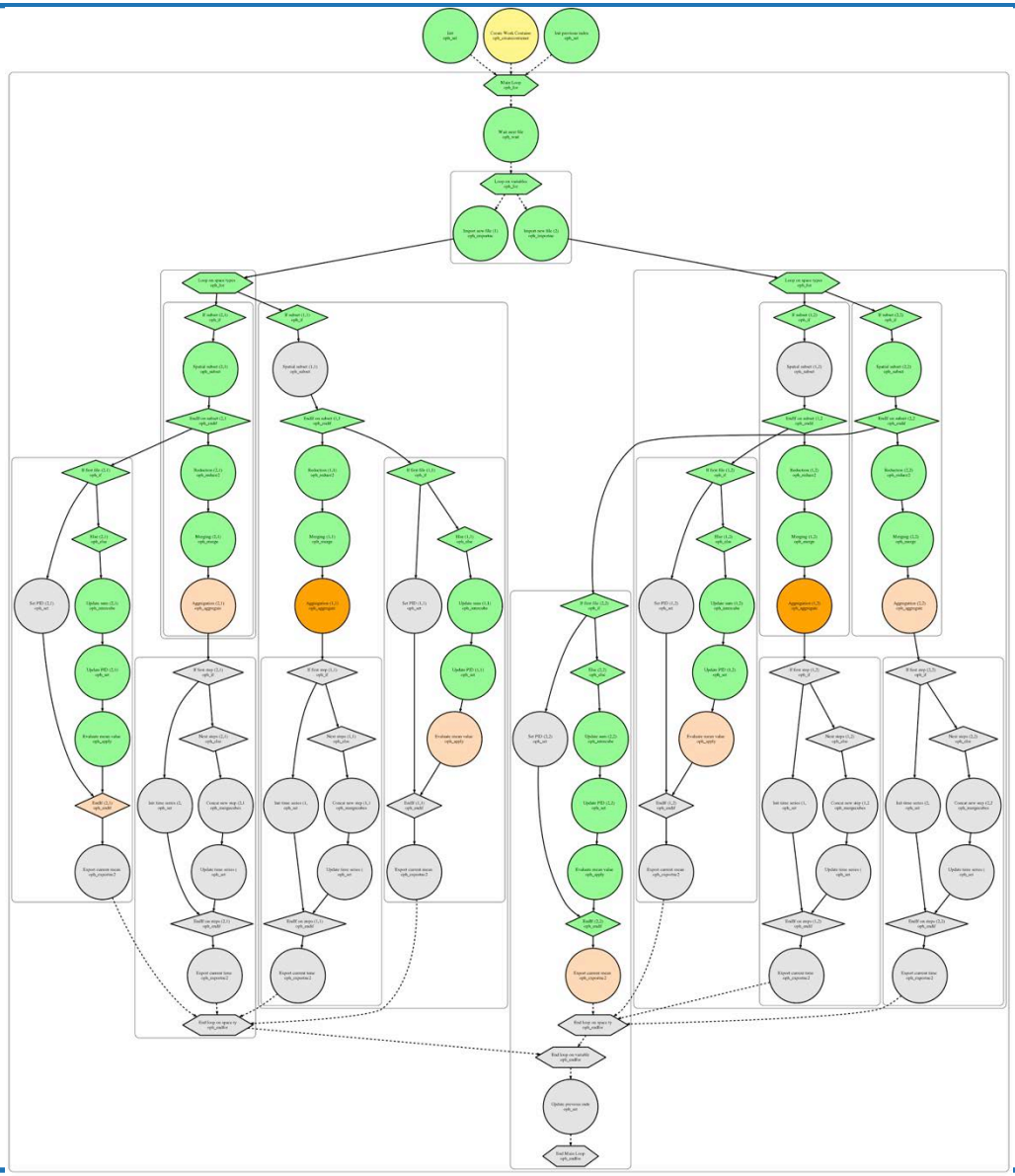
Synchronization between GLOB16 execution and Ophidia Workflow – **OPH-WAIT**

Import sosst and sosaline variables

Mean computation (update)

Create netcdf and publication on THREDDS

Runtime execution of the analytics workflow



Complete workflow execution on a single GLOB16 output:

- *sea surface temperature and salinity*
- *global and local* mean
- *global and local timeseries (spatial mean)*

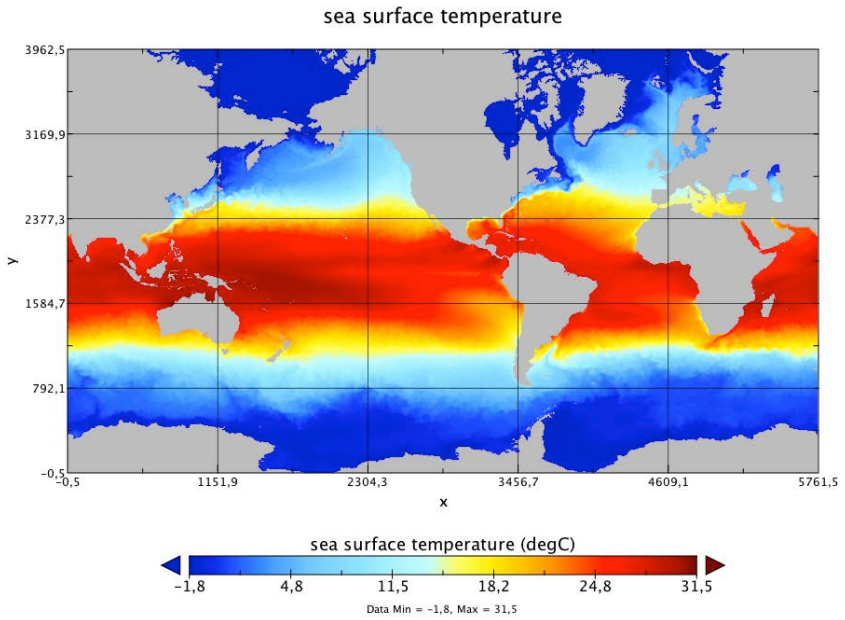
8 metrics computation for each time step

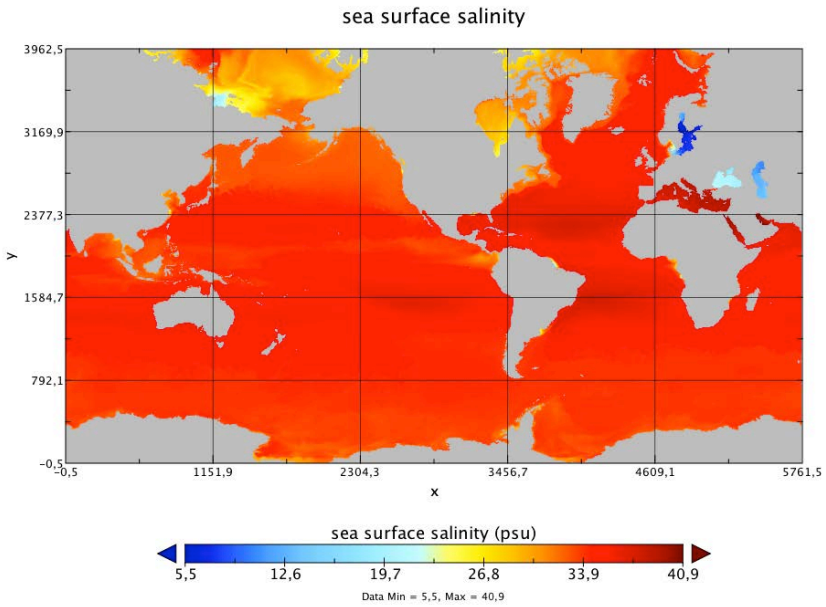
Ophidia **loops** management simplifies replication of similar “patterns” in the experiment

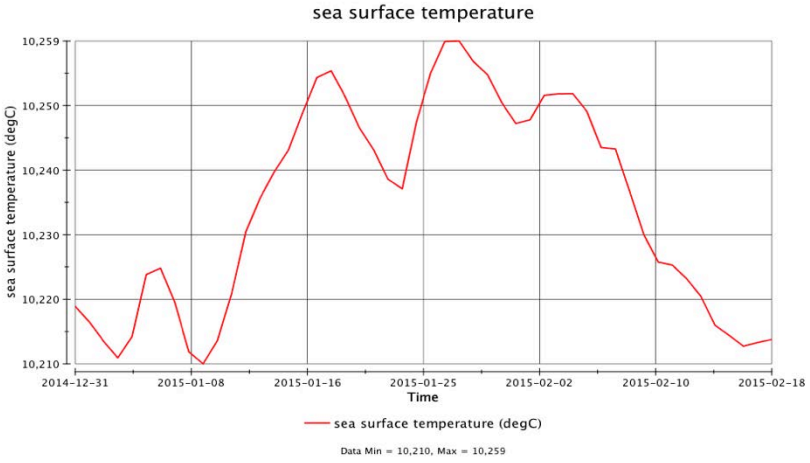
GLOB16 runs on **Athena –** Ophidia Workflow on **OphidiaLab**
Bash scripts used to run the operational chain

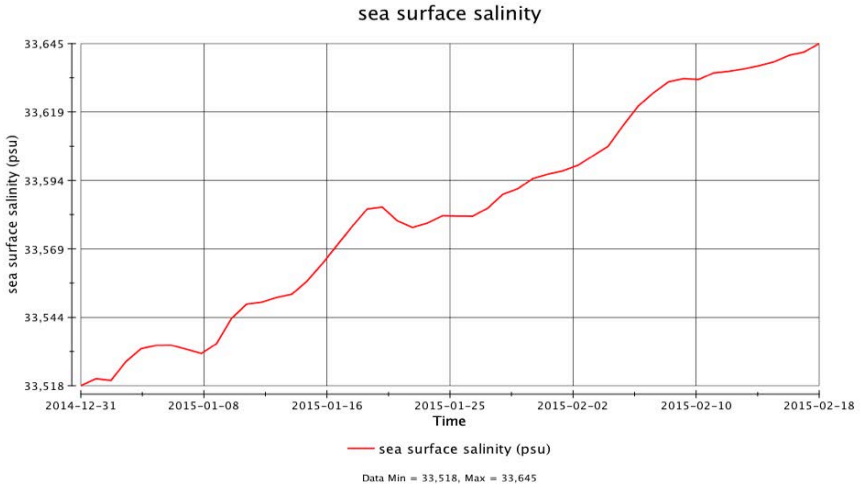
Planned to move to **Cylc**

Results at 50th step (global)



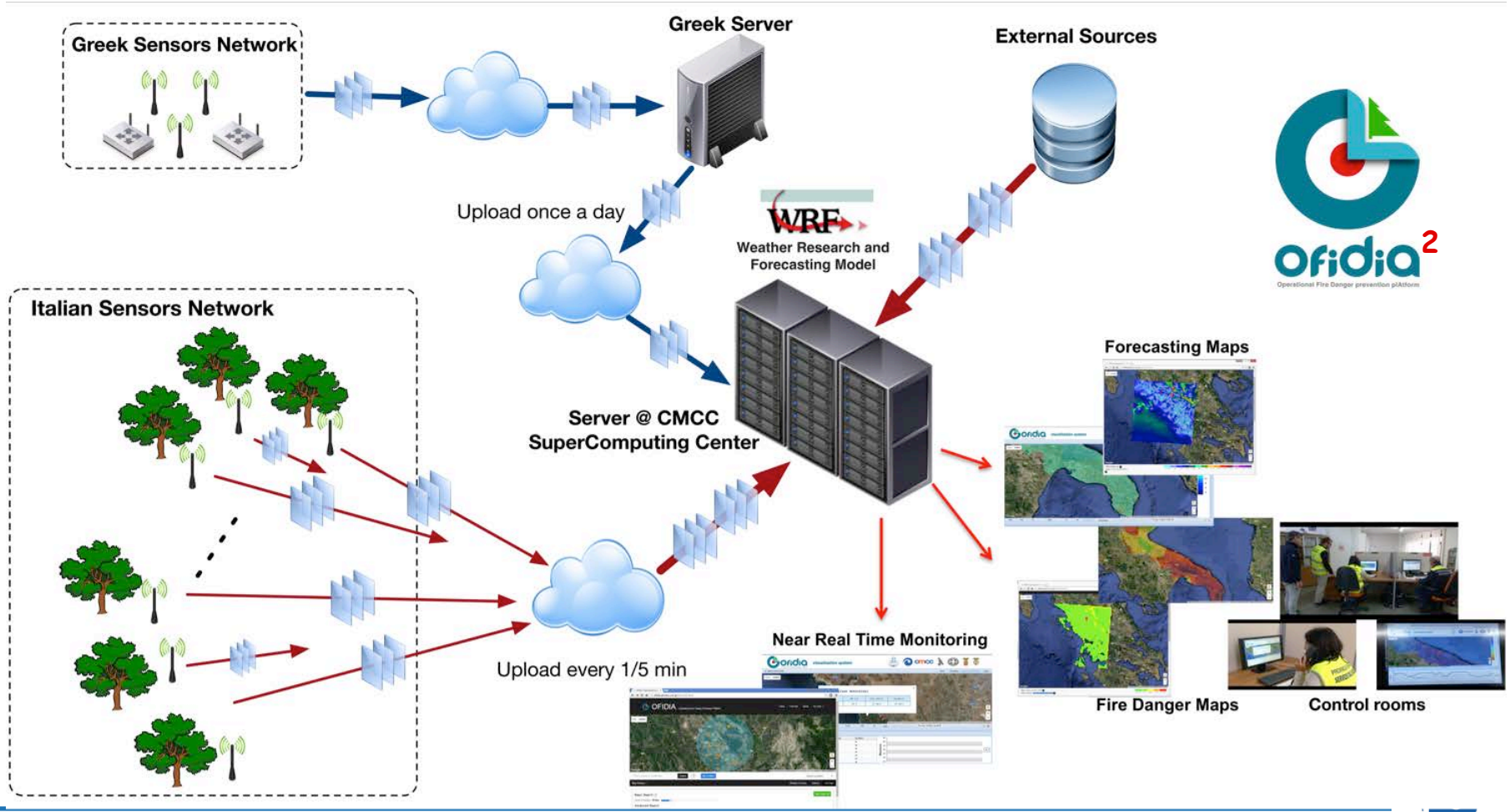




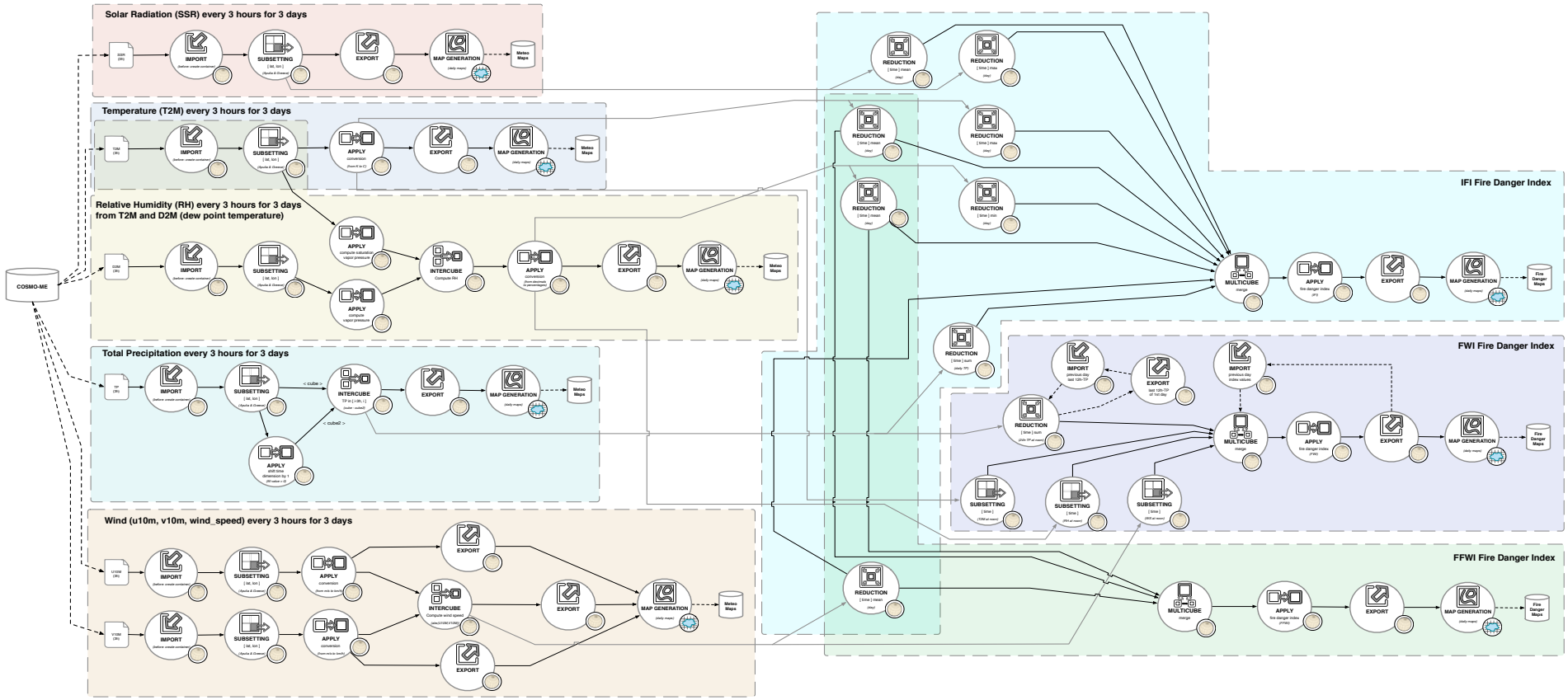


OFIDIA2 Project

The goal of the **OFIDIA2 Project** is to develop a platform for fire danger prevention that starting from the measurement of some environmental variables, is able to carry out the forecast of the risk of fire and the danger of the fire if it occurs across Apulia and Ioannina Regions.



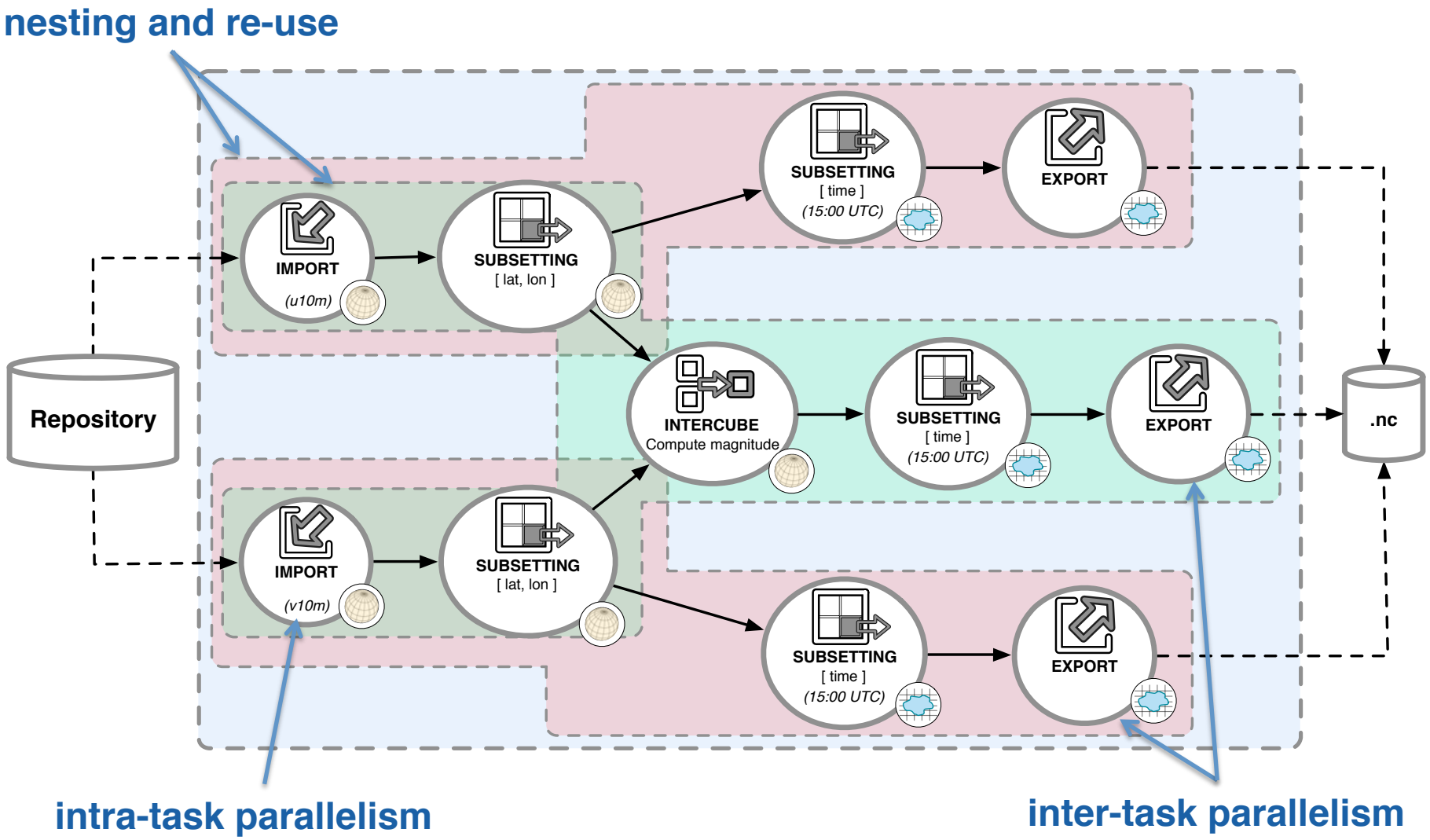
OFIDIA workflow - design



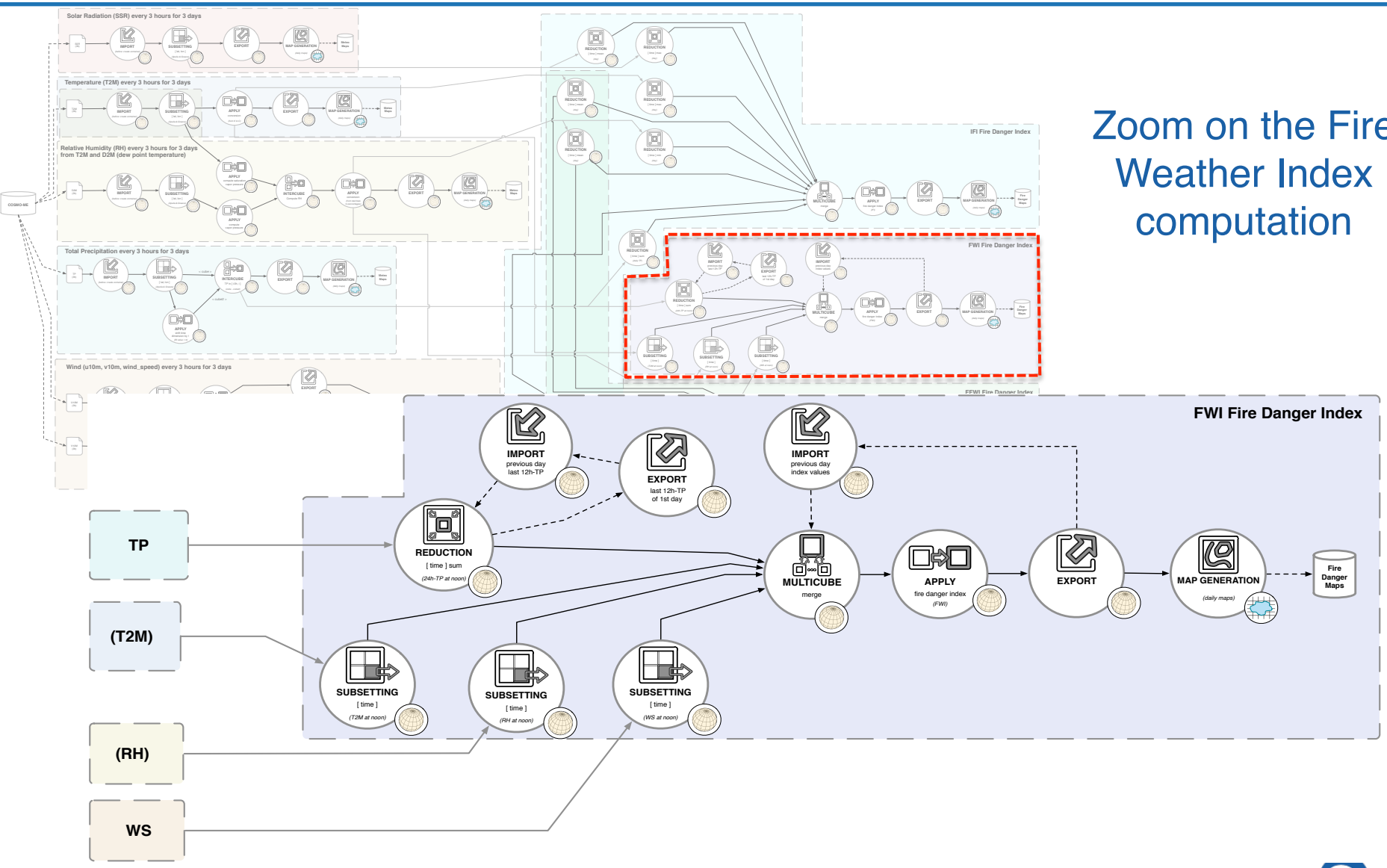
Input Variables: Solar Radiation (SSR), Temperature (T2M), Dew point temperature (D2M), Total precipitation (TP), Wind (u10m, v10m, wind speed)

Fire Danger Indexes: Canadian Fire Weather Index (CFWI), Fosberg Fire Weather Index (FFWI), Ichnusa Fire Index (IFI)

OFIDIA: 10m wind speed at 15:00 UTC

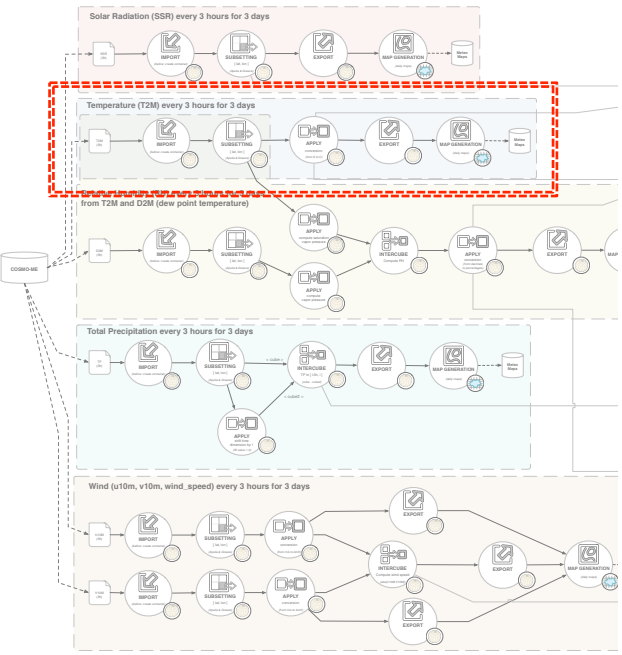


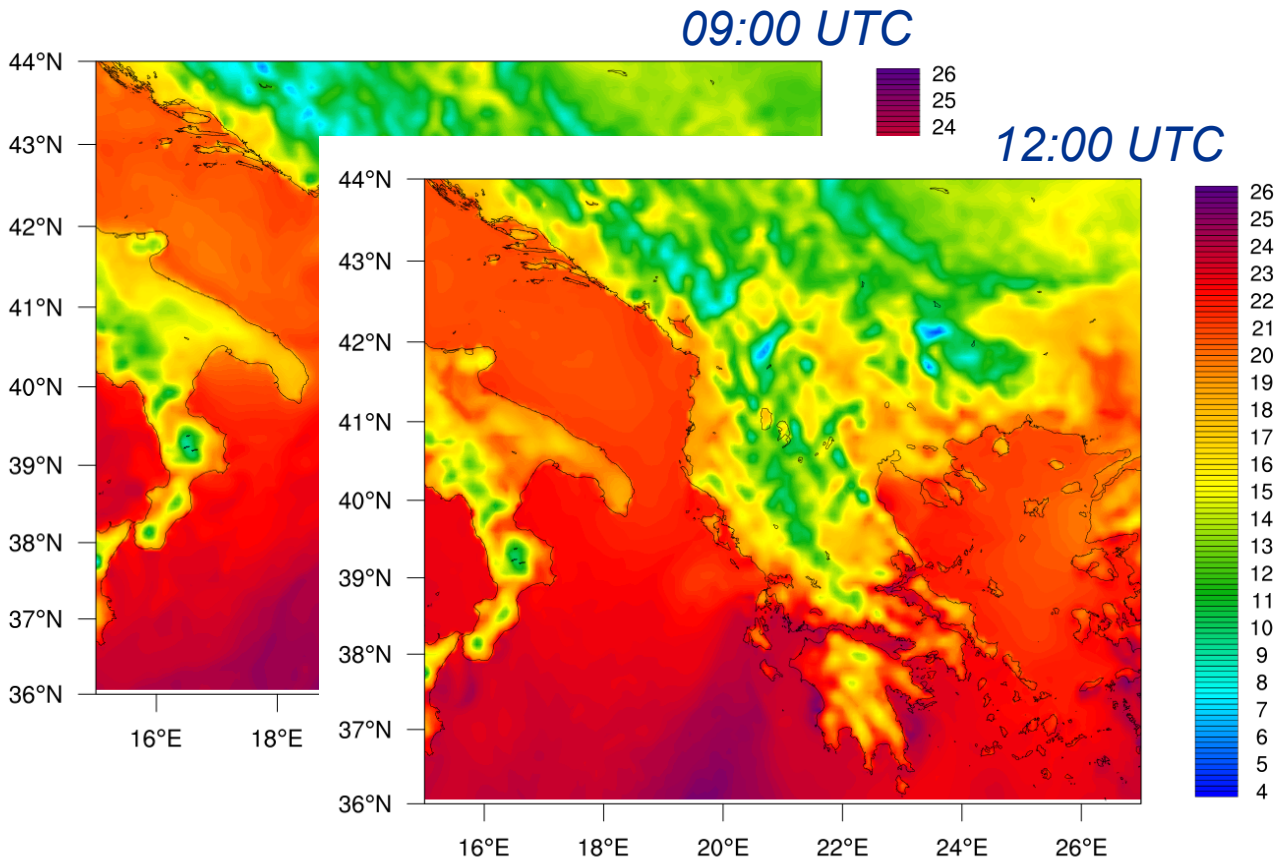
OFIDIA: Fire Danger Indices (I)

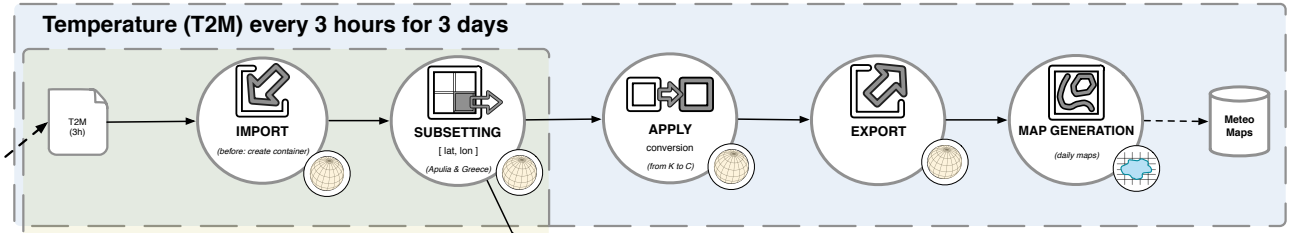


Zoom on the Fire Weather Index computation

OFIDIA: Fire Danger Indices (II)

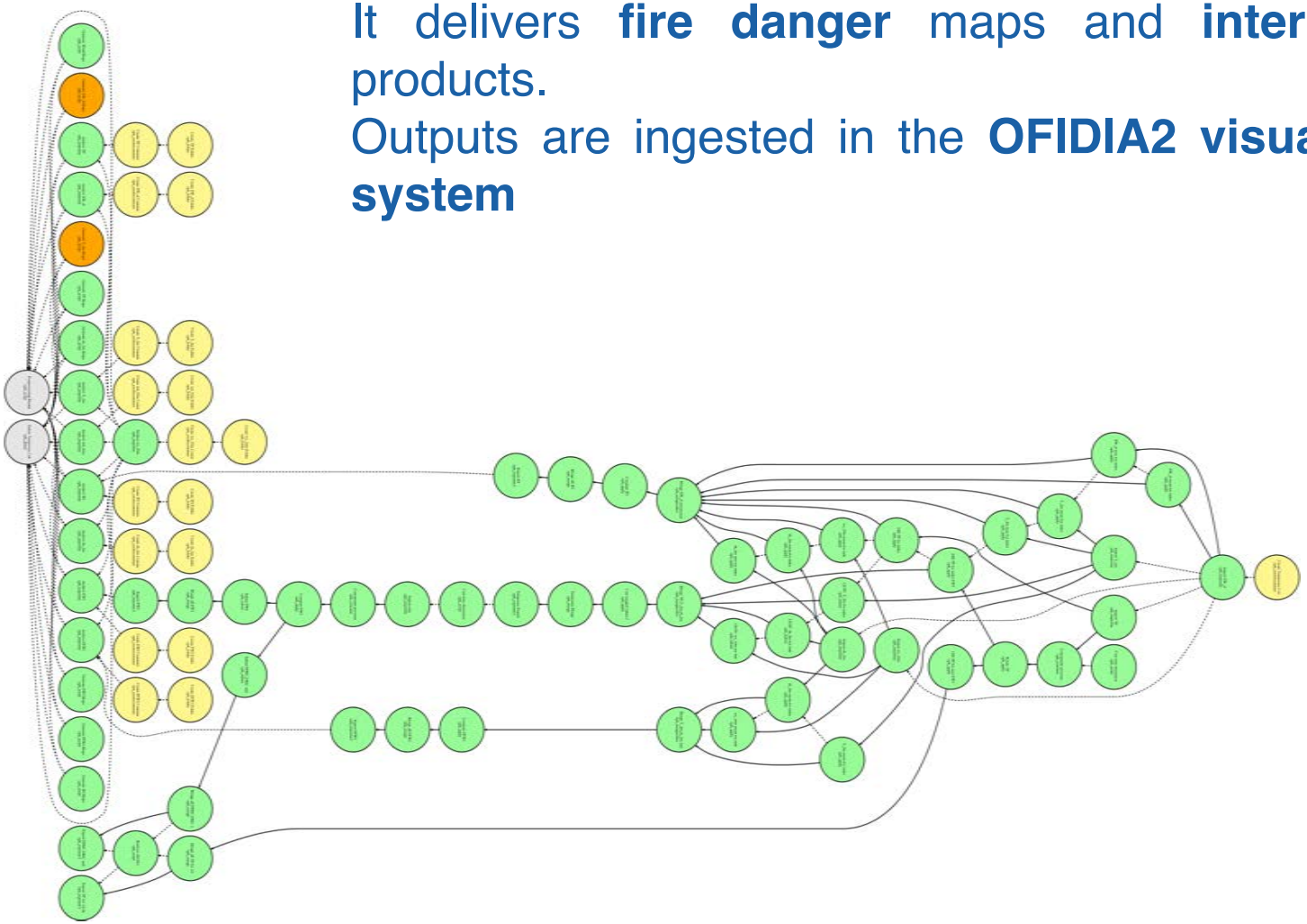






Zoom on the temperature forecast maps computation

OFIDIA workflow – runtime execution



It delivers **fire danger** maps and **intermediate** products.
Outputs are ingested in the **OFIDIA2 visualization system**

https://www.youtube.com/watch?v=vxbYF1Zhpuc&feature=youtu.be

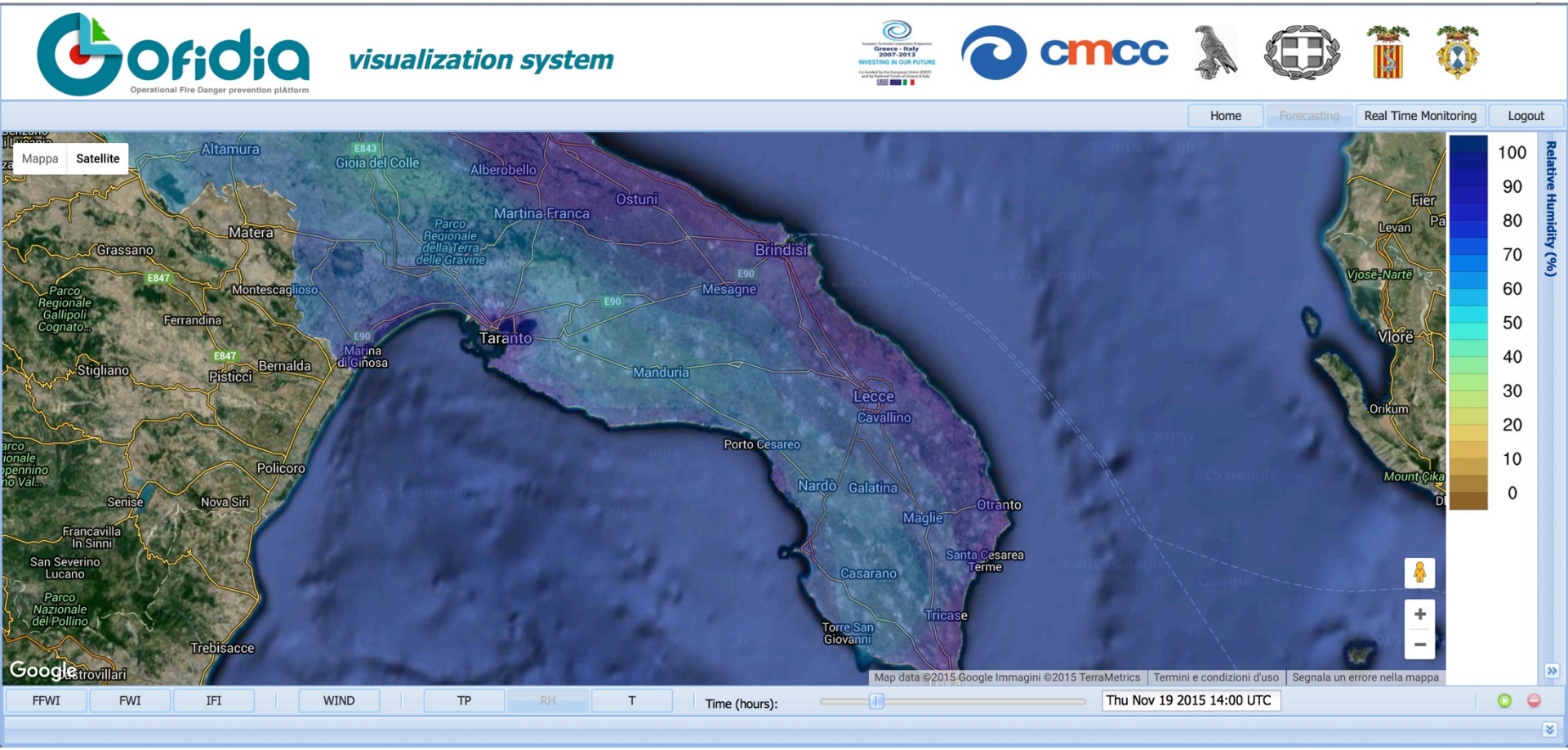
Forecasting for the next 72 hours – Fire Danger Maps

- Fire danger maps:
  - Canadian FWI
  - Fosberg FWI
  - IFI

Forecasting for the next 72 hours – Fire Danger Maps

- Fire danger maps:
  - Canadian FWI
  - Fosberg FWI
  - IFI

Forecasting for the next 72 hours – Weather maps



- Fire danger maps
- Weather forecast maps
  - ✓ T - Air temperature
  - ✓ **RH - Relative humidity**
  - ✓ PP - Precipitation (daily, cumulative)
  - ✓ WS - Wind speed & Wdir – Wind direction
  - ✓ Rg - Solar radiation flux density
  - ✓ Fuel moisture and temperature

Conclusions

- Ophidia (data analytics **operators** and **workflows management** capabilities) exploited in many project and use cases:

  CLIP-C, ESIWACE, GLOB16 indicators, OFIDIA

- Design and develop **other use cases**
- Exploit **Cylc** capabilities
- Deploy on **HPC environment** (Athena@CMCC)

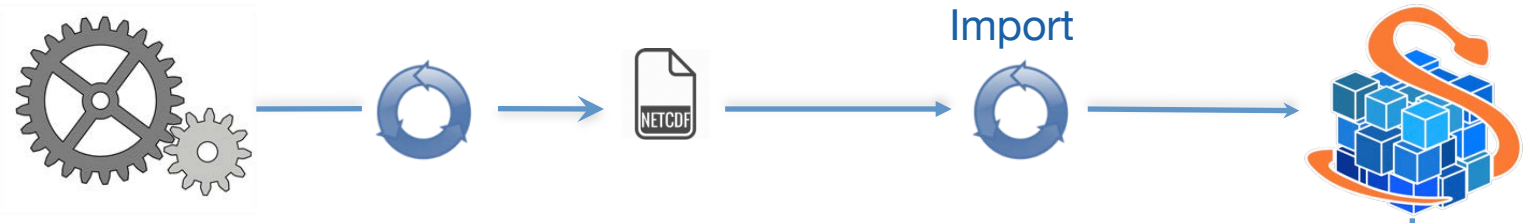
Thanks

http://ophidia.cmcc.it

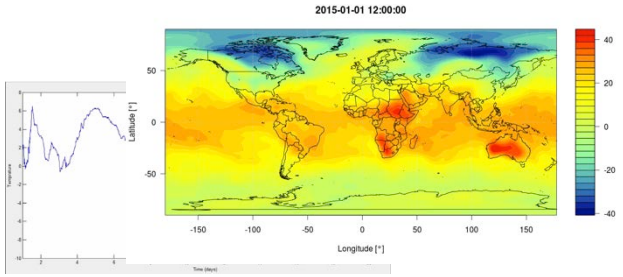
@OphidiaBigData

www.youtube.com/user/OphidiaBigData

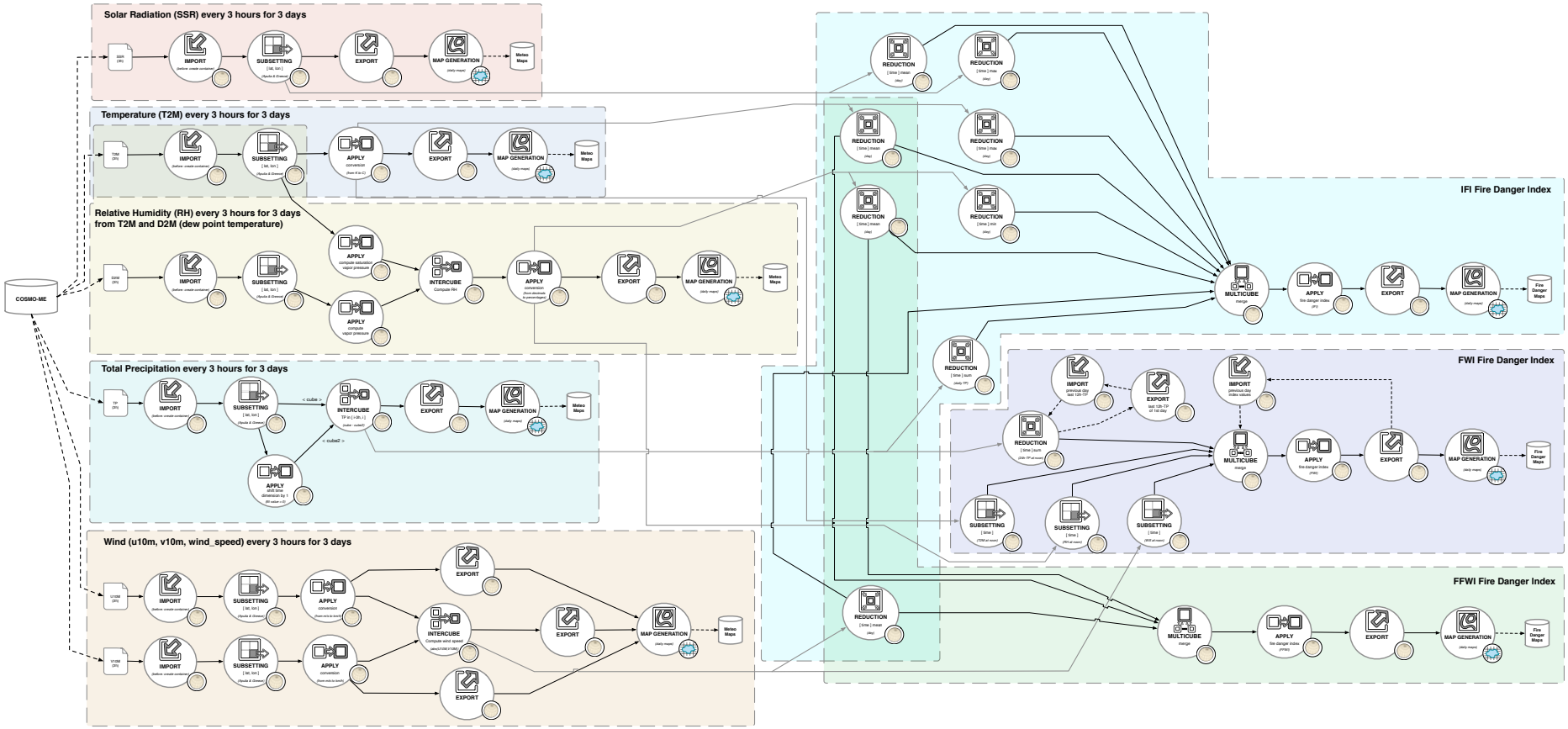
Real time parallel diagnostics workflow



- Asynchronous execution: Ophidia WF and GLOB16 model run in parallel
- Ophidia Workflow execution triggered by GLOB16
- Metrics are updated for each timestep



OFIDIA: Fire Danger Indices (I)



Delivery of fire danger maps and intermediate products (NetCDF formats)

INTERREG Project Italy-Greece on fire danger prevention (2013-2015)

Runtime and output