# Innovative Computational Science by Integration of Simulation/Data/Learning on Heterogeneous Supercomputers

jh240029

Leading-PI: Kengo Nakajima (U.Tokyo, Japan), Co-PI's: Takashi Furumura (U.Tokyo, Japan), France Boillod-Cerneux (CEA, France), Edoardo Di Napoli (JSC, Germany)

Overview

- In this project, we advance computational science by integrating "Simulation/Data/Learning (S+D+L)" using heterogeneous supercomputers like "Wisteria/BDEC-01 (U.Tokyo)" and "Flow (Nagoya U.)", along with "mdx."
- In FY.2021/2022, our focus was earthquake simulation with real-time data assimilation.
- **From FY.2023, we expand the "S+D+L" concept to other fields, initiating joint research and software development with international partners (Germany: JSC, FAU, France: CEA, Croatia: RBI). In FY.2023, we conducted research in atmospheric simulations, ported international partner applications to Wisteria/BDEC-01, and investigated "S+D+L" integration using h3-Open-BDEC.**
- For FY.2024, we switch to the international project and focus on the fields of earth science, life science, and library/software/tool for integration of "S+D+L". JSC (Jülich Supercomputing Centre, Germany), one of the international partners of this project, also conducts research on heterogeneous computing based on the idea of Modular Supercomputing Architecture
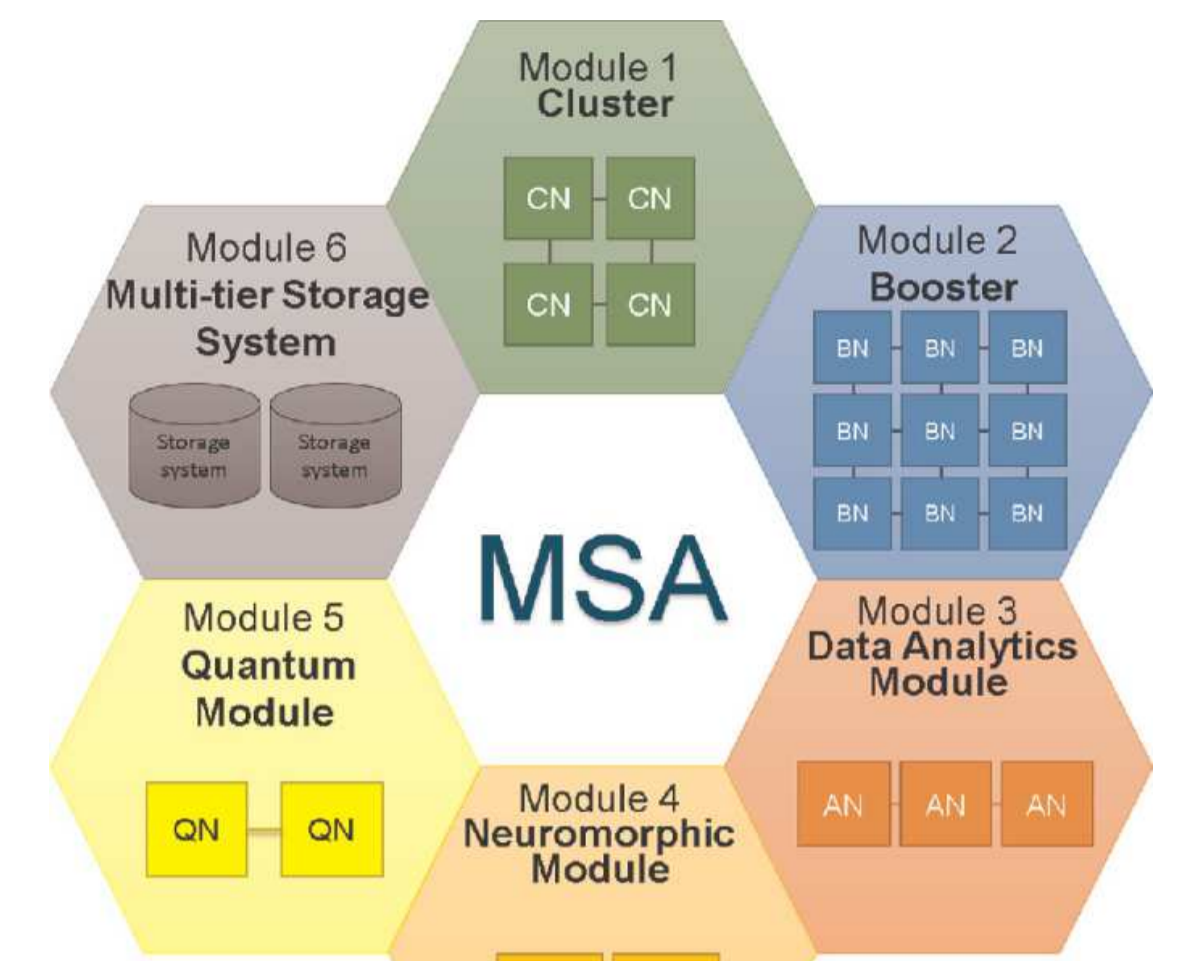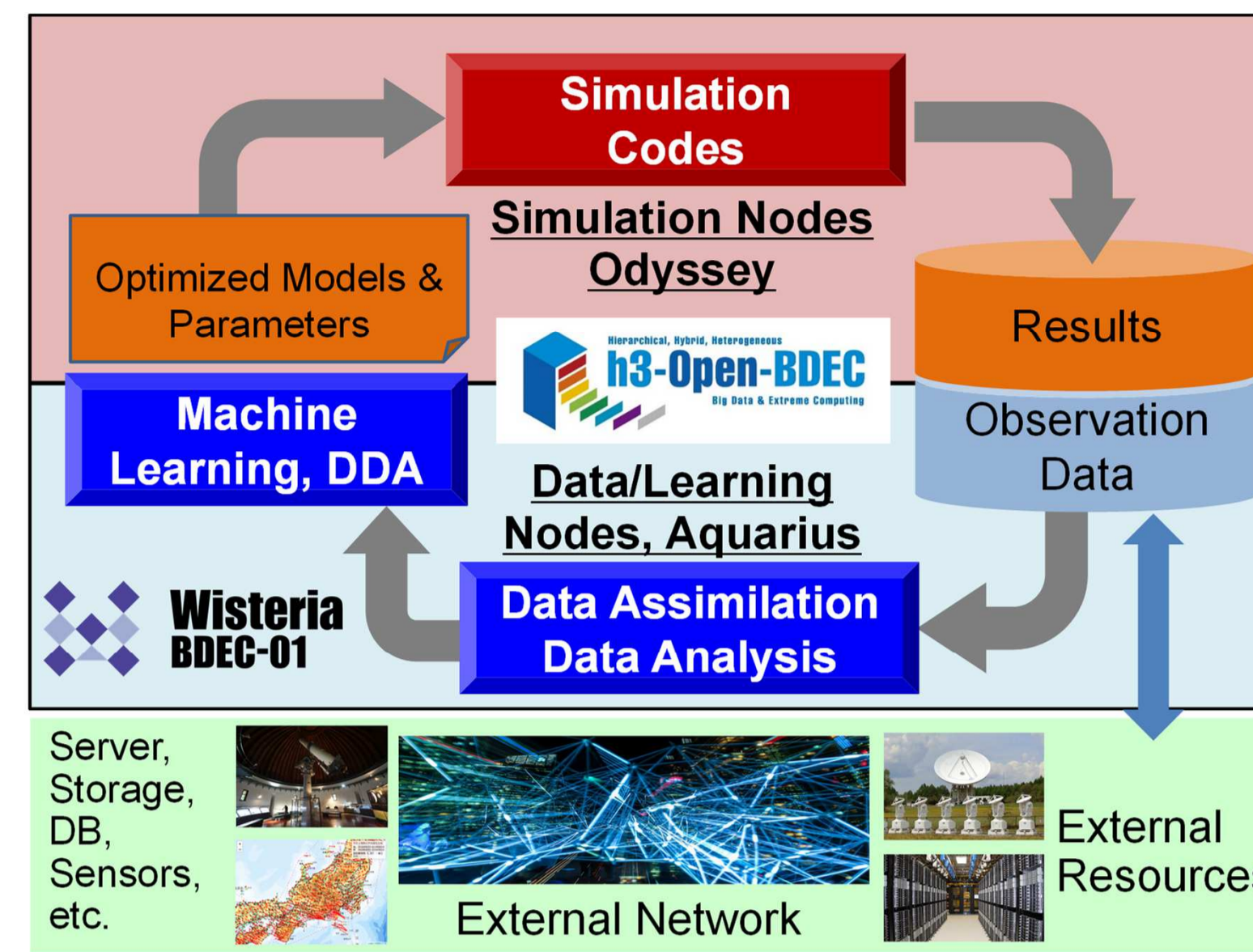

Integration of (S+D+I) by h3-Open-BDEC: WaitIO-Socket and Coupler (h3o-U/MP)

- h3-Open-SYS/WaitIO-Socket (WaitIO-Socket) is a system-wide communication library to couple multiple MPI programs for heterogeneous environments, such as Wisteria/BDEC-01, which provides an MPI like API and can easily connect multiple MPI programs.
- h3-Open-UTIL/MP (h3o-U/MP) originally provides capabilities for multi-physics coupling between different grid systems. Furthermore, it is equipped with a function for combined ensemble, and with an interface for coupling codes for large-scale simulations on Odyssey and ML/AI applications written in Python on Aquarius
  - While such coupled computing has been only possible by MPI on a single/homogeneous system, h3o-U/MP with WaitIO-Socket provides more flexible interface for integration of (S+D+L) on heterogeneous system, such as Wisteria/BDEC-01.
- **In this project, we apply h3-Open-BDEC to various types of workloads by partners' codes towards integrations of (S+D+L)**
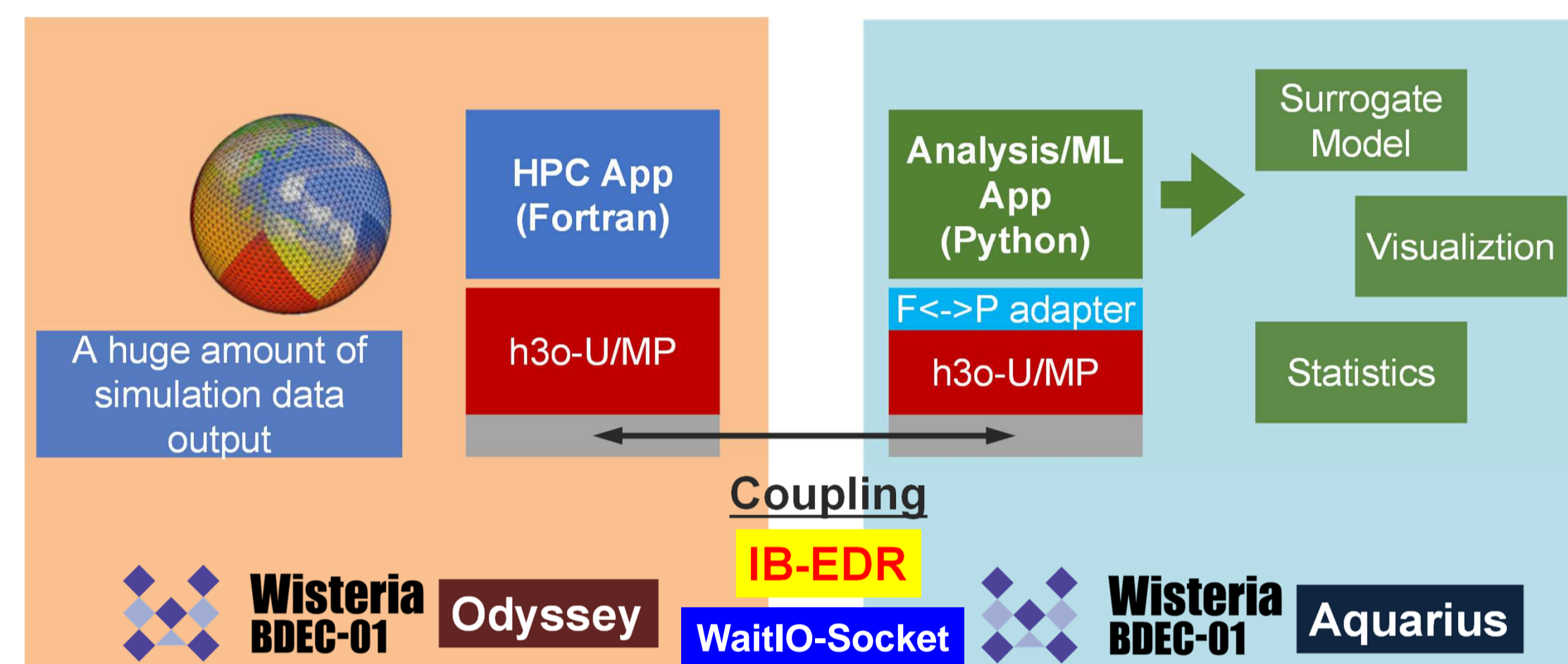

Ensemble Coupling for Global Atmospheric Simulations

- We improved the h3o-U/MP to enable *ensemble coupling* of low-resolution (224 km mesh) ensemble simulations and high-resolution model (14 km) simulations for global atmospheric simulations.
- **A simulation including a low-resolution ensemble of 64 cases was performed in FP32 using 320 nodes of Wisteria/BDEC-01 (Odyssey), and the results were compared with a high-resolution ensemble of 64 cases in FP64. In a trial simulation for 9 hours, we achieved approximately 100 times the computational efficiency while maintaining accuracy (2,240 node hours ⇒ 19.3 node hours).**
- In weather forecasting, the number of ensembles for data assimilation is approximately 50-100 due to limitations in computational resources and time in current medium- to long-term forecasts, but it is known that highly accurate forecasts can be made by performing data assimilation of approximately 1,000 ensembles. [Miyoshi et al. 2014]. By increasing the number of ensembles at a lower resolution, it is possible to issue more accurate forecasts with the same calculation time.
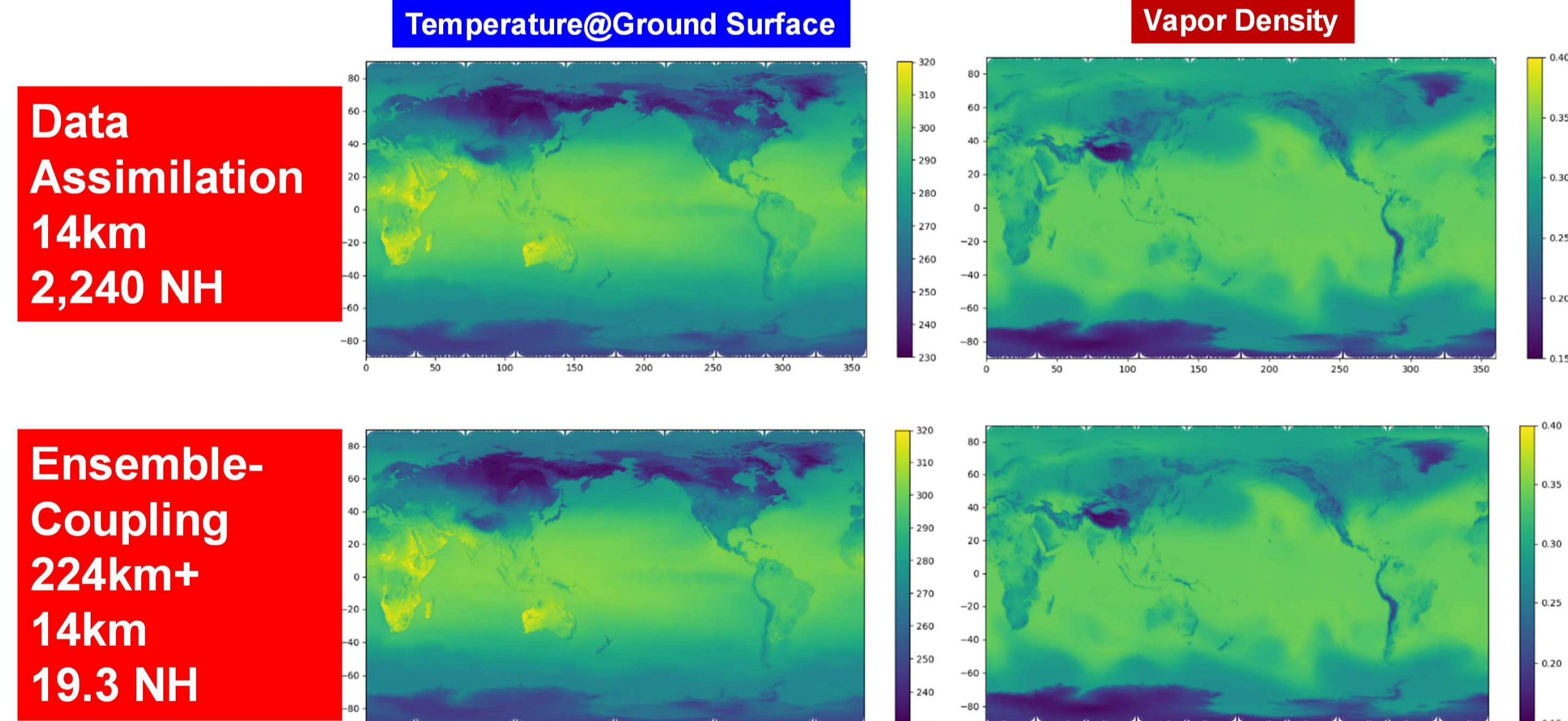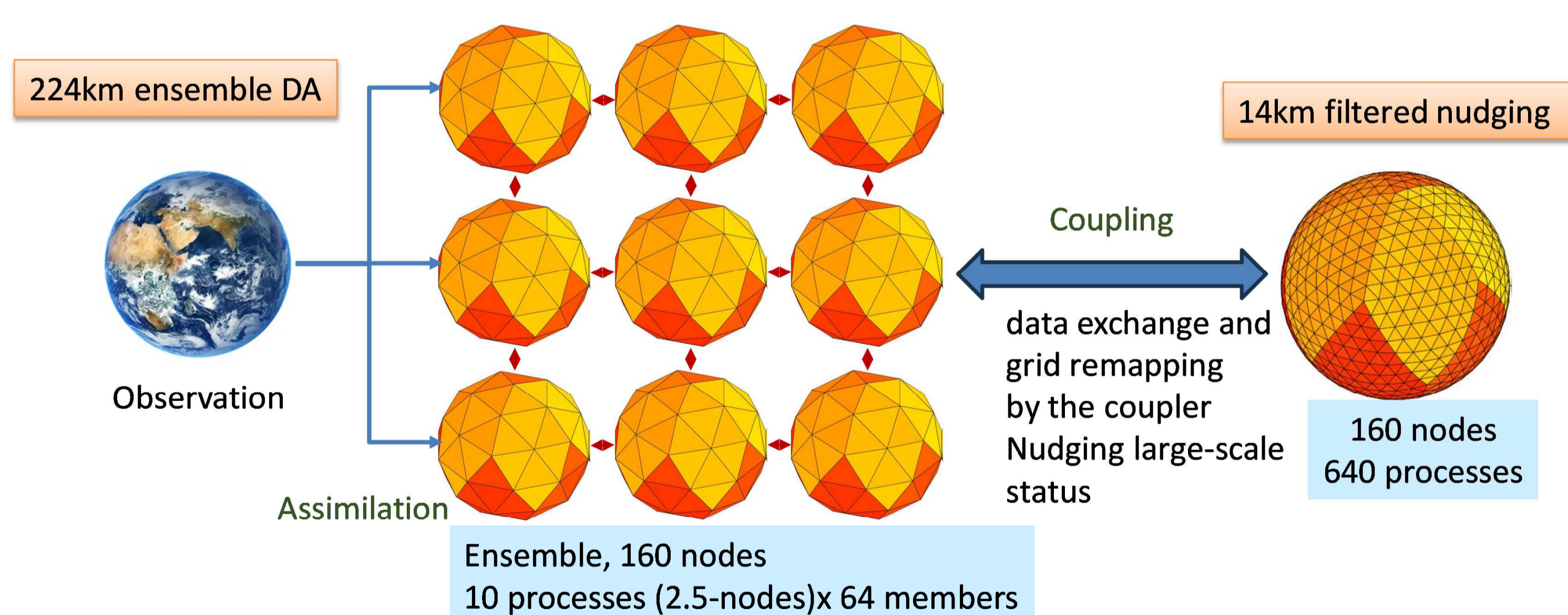
(Part of) Target Applications in FY.2024

Terrestrial Systems Modeling Platform TSMP)

- TSMP is a scale-consistent, highly modular, massively parallel, fully integrated soil-vegetation-atmosphere modeling system by JSC.
- Our target is coupling COSMO/ICON (Atmosphere)-ParFlow (Surface/Subsurface Flow)-CLM(Land Surface Model).
  - The coupling of 3 models has been already done using OASIS3 library on CPU-GPU heterogeneous environment of DEEP System in JSC based on MSA
- **In this project, we replace OASIS3 with h3-Open-BDEC, and coupled simulations will be possible on really heterogeneous systems, such as Wisteria/BDEC-01.**
  - In FY.2023, we mainly ported codes to Odyssey and made preliminary evaluations.
  - In FY.2024, we fucus on replacing OASIS3 with h3-Open-BDEC, develop preliminary version of the coupled codes, and conduct preliminary evaluations on Wisteria/BDEC-01.
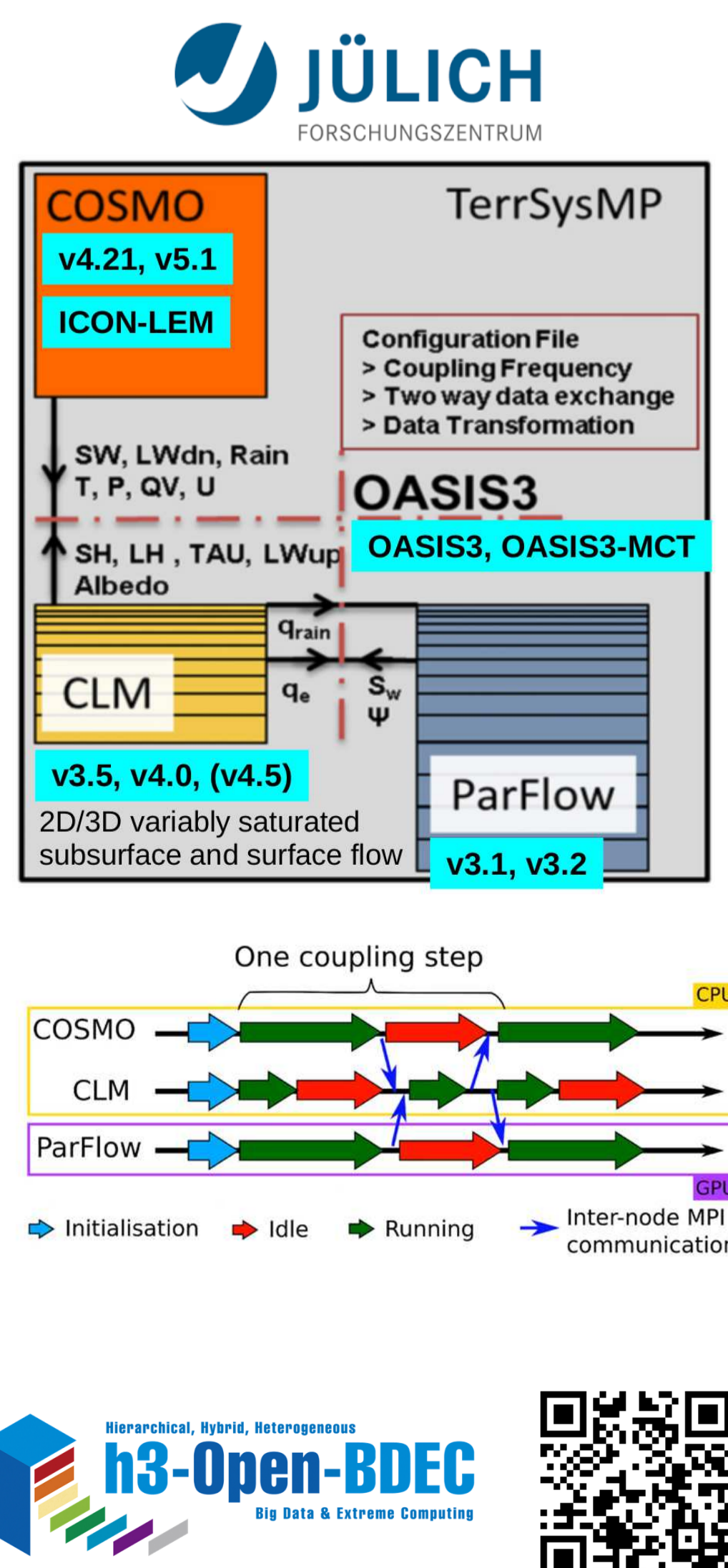
Big-DFT with GENESIS for SARS-CoV-2 Main Protease

- Developing medicines for viruses like SARS-CoV-2 faces challenges, including drug resistance (SARS-CoV-2: Virus, COVID-19: Infection)
  - Understanding and predicting drug resistance involves modeling structural changes from point mutations, utilizing long trajectories from classical molecular dynamics (MD/MM).
  - Mechanistic insight into mutation effects can benefit from quantum mechanical (QM) modeling.
- **In this project, we will exploit the heterogeneous architecture of Wisteria/BDEC-01 to build a coupled QM-MM workflow.**
  - The MM workflow will run the "GENESIS" (RIKEN) on Aquarius to exploit its GPU nodes and provide samples from a trajectory that are sent to the QM-MM workflow running "BigDFT" on Odyssey.
- **In FY.2024, we will construct preliminary version of QM-MM workflow using h3-Open-BDEC on Wisteria/BDEC-01, and make evaluations.**