

minisymposia and BOF-sessions to established conferences, in particular PASC, ISC, SC and EGU.

Dedicated PRACE allocation: A potential, dedicated computing allocation from PRACE to the CoEs will be managed, ensuring that this resource is used by project partners for developments in the framework of the ESiWACE2. Further allocations will be negotiated or applied for if necessary for the work in WP1.

Strategic interaction with HPC ecosystem and HPC industry: Liaising with the European HPC and Big Data ecosystems is planned, linking a) to relevant FET projects, Extreme Scale Demonstrators (EsD, H2020 call ICT-14-2019), and other CoEs, e.g. through active participation to European summit week; b) to the ETP4HPC, in particular by contributing to updates of the Strategic Research Agenda; c) to HPC-related hard- and software industry and foster their involvement in ESiWACE2 activities, in particular their participation to the HPC workshops; d) on the technical level with PRACE and other HPC centres, in particular with respect to the requirements arising from the developments in WP1 and WP4; e) with HPC Europa3 and similar initiatives granting use HPC resources; f) to the European BDVA, in particular by providing use-cases.

Task 6.2: Training and schools on HPC software engineering, methods and tools [Lead: CERFACS. Partners: ECMWF, BSC, METO, CMCC, UREAD, STFC, SEAGATE, ETH Zurich, MeteoSwiss, DDN, CNRS-IPSL]

This task organises the trainings offered in different HPC areas of ESiWACE2 pre-exascale expertise i.e. IO, computation (DSLs, C++ and coupling software), data analytics and containerisation. It also organises two summer schools to train scientists in these matters.

For all trainings, we embrace the creation of online digital media as *Open Educational Resources (OER)* material (<https://www.oercommons.org>) that will be available under a permissive CC-by licence and can be reused by teachers and researchers outside the consortium. Resulting material will be presented during the summer schools. All trainings also propose online sessions or face-to-face sessions at the partner's institution (Deliverable D6.2). Additionally, we will explore the collaboration with the international effort of the Universität Hamburg to establish an HPC certificate programme for scientists (<https://www.hhcc.uni-hamburg.de/en/hpc-certification-program.html>).

These trainings will result in a larger number of scientists and engineers with higher qualifications in the use and optimisation of climate and weather applications on tier-0 machines, increasing HPC awareness in that community. Ultimately, training will allow climate and weather research to be more productive, favouring European scientific excellence in that field.

Training on IO and HPC awareness: This task covers training on the IO hardware and software stack. Delivered material will target scientists accessing data on storage and developers that aim to utilise storage APIs efficiently.

We will create descriptive course material covering the co-designed software stack of WP4 as a standardised and efficient platform for IO (Cylc, XIOS, ESDM, NetCDF, HDF5) and how the analysis stack fits into this (in-depth analysis workflows are part of training on high-performance data analytics, see below), training on the underlying storage APIs and file systems (e.g., Mero/Clovis, Lustre), efficient creation of IO dominated workflows, performance and efficiency considerations when dealing with file systems, NVM, tape or object storage, and cost-considerations in storage architectures.

Material will cover roughly 8 hours course work in the form of presentations and a summarising paper covering the individual aspects. Webinars that present the created OER material and blog entries describing issues will be organised. Finally, we will harness existing community approaches like the Virtual Institute of IO (<https://www.vi4io.org/>) for promoting this training.

Training on DSL: The rapid changes in the multiple supercomputing architectures used to run weather and climate codes and the different programming models used seriously affect the development productivity and the ability to retain a single source code running efficiently everywhere. Domain-specific languages provide a solution to portability of these codes. In this training, we provide insights on DSLs considered in

ESiWACE2 (PSyclone, CLAW and GridTools ecosystem) and demonstrate how to apply them to weather and climate models.

The training will be organised in the form of a 5-day face-to-face training workshop in spring of year 2 for about 25 people. Participants will theoretically and practically learn how to use the DSL languages to implement PDE operators. During a hands-on session, participants will be encouraged to implement some of the benchmark models defined by WP2 using DSLs and to build their own toy models, followed by an in-depth evaluation of generated optimised implementation and performance benefits.

Training on C++ for HPC: ETH Zurich organises an advanced course for C++ in HPC as part of its regular trainings. Up to 12 domain scientists can further improve their software-engineering skills by participating in the autumn 2021 three-day course on "Advanced C++ programming for HPC". This course will present advanced tools for effective C++ programming in the context of HPC, such as generic programming techniques, API development, specific C++-11/14 constructs, rather than treating parallel programming primitives, such as OpenMP or MPI, a subject of widely available courses.

Online training course for code coupling with OASIS3-MCT: The objectives of this task is to create an online course that teaches the participants basic concepts in code coupling, focussing on the ocean-atmosphere context, and help them learn on how to use OASIS3-MCT. This training course is for engineers, physicists, and computer scientists wishing to use this code coupling software in their own coupled model.

The training course will be delivered once per year during the 2nd, 3rd and 4th year of the project, over 4 consecutive weeks with learning activities delivered each week and requiring about 2 hours of work per week from the participants. The material will cover theoretical concepts about code coupling, instructions on how to download, install and compile OASIS3-MCT and finally implementation of the coupling between two toy models in a hands-on tutorial session.

Training on High-Performance Data Analytics: The objective of this task is to increase scientists' expertise on scientific data analysis at scale applied to climate and weather domains, using high-performance data analytics tools available from the open source market (e.g., Ophidia). It will address several vertical training levels (e.g., intermediate, expert) from different horizontal perspectives (e.g., end-user, developer, administrator), covering from simple analytics tasks to workflows and applications (e.g., Python-based) and providing best practices and guidelines on dealing with massive scientific datasets on HPC architectures.

Delivered material, using aforementioned material on IO and HPC awareness as background, will include 8 hours of theoretical and practical aspects on big data (both introductory and advanced topics), high-performance data management (including performance and optimisation aspects), scientific data analytics at scale, and analytics workflows. As for the training on code coupling with OASIS3-MCT, the online training course (e.g. webinar) will be delivered once per year, during 2nd, 3rd and 4th year of the project, over 4 consecutive weeks, with 2 hours of work per week from the participants.

Training on Docker containerisation: Docker is an open software technology for developers to build, ship, and run distributed applications in containers. This is achieved by describing the software environment and build instructions allowing the creation of a container that can be executed on different environments and does not depend on system-wide software packages. Consisting of a rich ecosystem, Docker enables applications to be quickly assembled from components and eliminates the friction between development, QA, and production environments. As a result, software can ship faster and run the same unchanged stack on laptops, data centres, and virtual machines. With Docker, developers can build any application in any language using any toolchain; "dockerised" applications are completely portable and can run anywhere.

A 4-hour introductory tutorial followed by a 2-hour hands-on session on Docker containers will be produced, condensing the experiences from WP2, task 2.4. A recording of the material presented at the summer schools will be made available online in order to reach a wider audience.

Summer school in HPC for weather and climate: In this task, we will organise two summer schools for Earth system scientists, including PhD students and postdocs, covering the HPC aspects to run scientific workflows on large scale HPC environments efficiently. A five day summer school will take place in year 2

and 3 of the project and support about 30 students each year and fund their attendance. The selection procedure for subsidising students will be established 6 months before the summer school taking into account the experience gained with summer schools held within the IS-ENES projects. It will prioritise diversity and internationality and fund students depending on the financial support possible by their home institution.

Theoretical courses and hands-on sessions will be organised on IO, HPC computation (DSLs, C++ and coupling software), data analytics, containers, efficient programming and usage of data Centre resources, using the teaching material created in this WP, but we will also invite external experts to give tutorials about selected topics. Selected presentations will be recorded and made available online after the summer school for a wider audience. A survey will be conducted to evaluate the feedback of participants about the content of the tutorial itself (D6.1).

Deliverables

D6.1 Report on summer school and its material

Report covering the teaching material and discussing the result of the conducted survey regarding the quality and effectiveness of the summer school and the individual OER materials. Due at month 36; PU; R; Lead: UREAD

D6.2 Report on the different trainings

Report describing the different online or face-to-face training on IO, DSLs, C++, coupling, HPDA, containerisation held during the project. Due at month 42; PU; R; Lead: CERFACS

D6.3 6th ENES HPC-workshop in Hamburg

Three day Workshop continuing the successful series of HPC workshops (Hamburg, Toulouse (2x) and Lecce (2x)) on weather and climate-related subjects. Due at month 18; PU; OTHER; Lead: DKRZ

D6.4 Training OER material made available under a permissive CC-by licence

Report describing the OER material. Due at month 36; PU; R; Lead: CERFACS