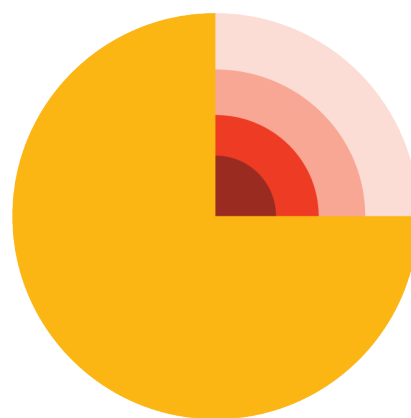




ChEESE

Deliverable D7.4 Training Plan



Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Spain, Italy, Iceland, Germany, Norway, France, Finland and Croatia under grant agreement No 101093038.



ChEESE-2P Deliverable

D7.4 Training Plan

Document Information

Contract Number	101093038
Project Website	https://cheese2.eu
Contractual Deadline	30/08/2023 (M8)
Activity	WP7
Dissemination Level	PUBLIC
Nature	R

Abstract

This document introduces the training plan of ChEESE-2P project. It outlines the key aspects of the training program, including the target audiences, topics, and expected learning outcomes. Additionally, the plan includes a tentative timetable for the delivery of the courses, ensuring a structured approach to their implementation.

I. Copyright

Copyright Members of ChEESE collaboration, 2023/2026.

II. Delivery Slip

Role	Name	Partner/Activity
Author	María Paz Baghetti	BSC-WP7 leader
Contributor 1	Varvara Vedia	CSIC, Communication Specialist
Contributor 2	Arnau Folch	CSIC, ChEESE coordinator
Reviewer 1	Maria Jose Garcia	CSIC, WP1 leader
Reviewer 2	Dr. Boris Kaus	UM, PSB member

III. Version control

Version	Date	Description of change	Author/Partner
V1	17/07/2023	Initial draft	María Paz Baghetti
V2	06/08/2023	Second draft including comments from contributors	María Paz Baghetti
V3	28/08/2023	Third draft including comments from reviewers	María Paz Baghetti
V4	31/08/2023	Final version	María Paz Baghetti

IV. List of Acronyms

Acronym	Description
AI	Artificial Intelligence
CSIC	Consejo Superior de Investigaciones Científicas
ChEESE-1P	ChEESE First Part
ChEESE -2P	ChEESE Second Part
CoE	Center of Excellence
GA	Grant Agreement
EGU	European Geosciences Union
HPC	High Performance Computing
KPI	Key Performance indicator
NCCs	National Competence Centers
SE	Solid Earth
SO	Specific Objective
PSB	Project Supervisory Board
WP	Work Package

Index

1. Introduction	5
2. Context	6
2.1 Galileo Conference	6
2.2 Social Media Campaign	8
3. Training events	8
3.1 Training List	10
3.2 Brief description of thematic courses	12
3.2.1 Computational volcanology: Modelling Volcanic plumes using European flagship codes	12
3.2.2 Computational geodynamics	12
3.2.3 Tsunami and Meteo-tsunami modelling and numerical simulation	13
3.2.4 Computational seismology I and II	13
3.2.5 Artificial Intelligence and Machine Learning for Geosciences	14
4. Special series of training courses focused on equality	14
5. Training assessment: quality assurance	15
6. Training Collaboration with CASTIEL-2, CoEs and NCCs	15
7. Conclusions	16

List of Tables

Table 1. Training list.....	12
-----------------------------	----

List of Figures

Figure 1. Attendees at the Galileo Conference endorsed by ChEESE.....	8
Figure 2. ChEESE-2P wiki Training Events screen.....	9

1. Introduction

This document presents the training plan for the second phase of the Centre of Excellence for Exascale in Solid Earth (ChEESE-2P). The preparation of this document falls under Task T7.3 Training Activities, which aims to identify the training needs of researchers and other target audiences, as well as on the organisation and logistical support of different courses.

The GA includes one specific objective related to training (SO number 11): Coordinate with the NCCs in the organization of open, free, and gender and country-balanced (EU13) training events and seminars on the specific domains of SE. The KPI to this SO is to organise at least 8 training events, which include two more editions of the successful [Advanced Training on HPC for Computational Seismology conducted during ChEESE-1P](#).

It is worth noting that at least 4 training courses have already been carried out in the first semester. While they will be mentioned, this document will primarily focus on future training sessions. We will present their objectives and provide estimated dates for their implementation.

In line with ChEESE's commitment to fostering collaboration, this training plan has been designed with the consideration of organising courses alongside other NCCs under the CASTIEL-2 project.

While the main emphasis will be on delivering the courses in a face-to-face format, there is also the possibility of conducting online or hybrid sessions: virtual/hybrid sessions are more cost-efficient and inclusive, since it allows more people to participate. This said, face to face meeting training settings are more suitable when the hands-on component of the training event is very relevant for the achievement of the training goals.

ChEESE community recognises that there is a need to contribute to the knowledge and expertise within the SE community, especially in the preparation to transition to the pre-exascale and exascale era of supercomputers. Therefore, the SE and geosciences community was consulted through surveys and social media polls in order to gain a better understanding their training needs. The answers were then matched to the skills and expertise of ChEESE partners, creating a co-developed and bespoke training plan that would be truly relevant to our stakeholders.

2. Context

In ChEESE, there is a strong belief that the input from the SE and Geosciences Community is crucial in creating an effective training plan. Therefore, to develop this training plan, three mechanisms were established to identify the topics of interest of this group.

1. One approach involved obtaining feedback from all participants of the Galileo Conference.
2. The second entailed conducting a social media campaign to collect user opinions.
3. Additionally, a post-event survey was sent to trainees after each training event, allowing the community to express further training needs that might arise throughout the project.

Based on the gathered information from all these mechanisms, ChEESE's expert partners were consulted regarding their interest in specific topics and their potential involvement as organizers and/or trainers in any of the courses. This served as the basis to elaborate a preliminary training plan and the training proposal embedded in the current deliverable.

2.1 Galileo Conference

ChEESE organised the 11th EGU Galileo Conference: Solid Earth and Geohazards in the Exascale Era, which took place in Barcelona, Spain on 23-26 May 2023.

The Galileo conferences of the European Geosciences Union (EGU) are small-scale and informal events to promote in-depth discussion among experts on cutting-edge topics at the frontier of geoscience research. The event organised by ChEESE gathered 78 participants from 15 different countries to address both the challenges and opportunities that Exascale computing presents for geosciences. Most attendees came from European-based research institutes and academia, although other institutions worldwide were also represented by individuals (e.g. from the USA, China, and Latin America).

The conference was articulated around 4 sessions linked to corresponding trending topics: preparation and optimization of High-Performance Computing (HPC) applications (Session 1), edge-to-end data workflows (Session 2), state-of-the-art in computational geosciences (Session 3), and

research infrastructures and EuroHPC access modes and policies (Session 4).

Considering the valuable presence and diverse backgrounds of the attendees, it was deemed appropriate to conduct a post-event survey. The survey aimed not only to measure the satisfaction of the participants, but also to gather insights into their specific interests and preferences regarding future training courses. This training plan has been carefully prepared, taking into account the comments received in the survey.

Some of the topics that emerged from the participants' responses include:

- Workflow projection and tools
- Simulation software
- AI, ML and HPC programming
- [Julia](#) programming language
- Modelling of extreme weather events
- Computational and numerical earthquake seismology, floods, tsunami, landslides, volcanology, and geophysics for groundwater resources
- Preparing/adapting multiphysics codes for parallel computation and running on supercomputers
- Code optimisation
- Urgent computing
- Large-scale visualisation
- Multi-physical/coupled models
- Use of ChEESE codes
- Containerization
- Performance portability



Figure 1. Attendees at the Galileo Conference endorsed by ChEESE.

2.2 Social Media Campaign

A social media campaign around the community's training needs was launched to gather feedback from ChEESE's 1100 Twitter and 460 followers on LinkedIn. Some of the answers are presented below:

- HPC for geosciences
- Use of codes developed within ChEESE
- Numerical modelling of earthquakes and tsunamis
- HPC for geosciences
- Access and launch HPC simulations
- ML & AI approaches applied to Geosciences
- AI post processing
- Use of codes developed within ChEESE
- Computational seismology and workflows
- Large-scale visualisation
- Computational geophysics for groundwater resources
- PyCOMPSs framework, Machine learning practical courses using HPC
- Finite Element Methods
- Landslide simulation

3. Training events

Considering the training needs of the ChEESE community and the skills and expertise possessed by the project partners, the WP7 team defined a list of training activities. These events will take place throughout the duration of

ChEESE and will be organised by the same team in collaboration with the instructors.

After an event is organised, we will use the ChEESE general mailing list to disseminate the event details to the project's community members. When the training course or webinar is opened to the scientific community outside of ChEESE, we will utilize social media channels to advertise and raise awareness about the event. Moreover, with the consent of the instructor, the majority of ChEESE webinars will be uploaded to the project's [YouTube Channel](#), ensuring accessibility and knowledge sharing.

The dissemination through social media will enable these webinars to reach a broader audience and foster collaboration with the wider scientific community. Furthermore, for comprehensive information sharing, we will also use the project's WIKI to promote the upcoming events and publish the recordings (if applicable) or course materials after each webinar.

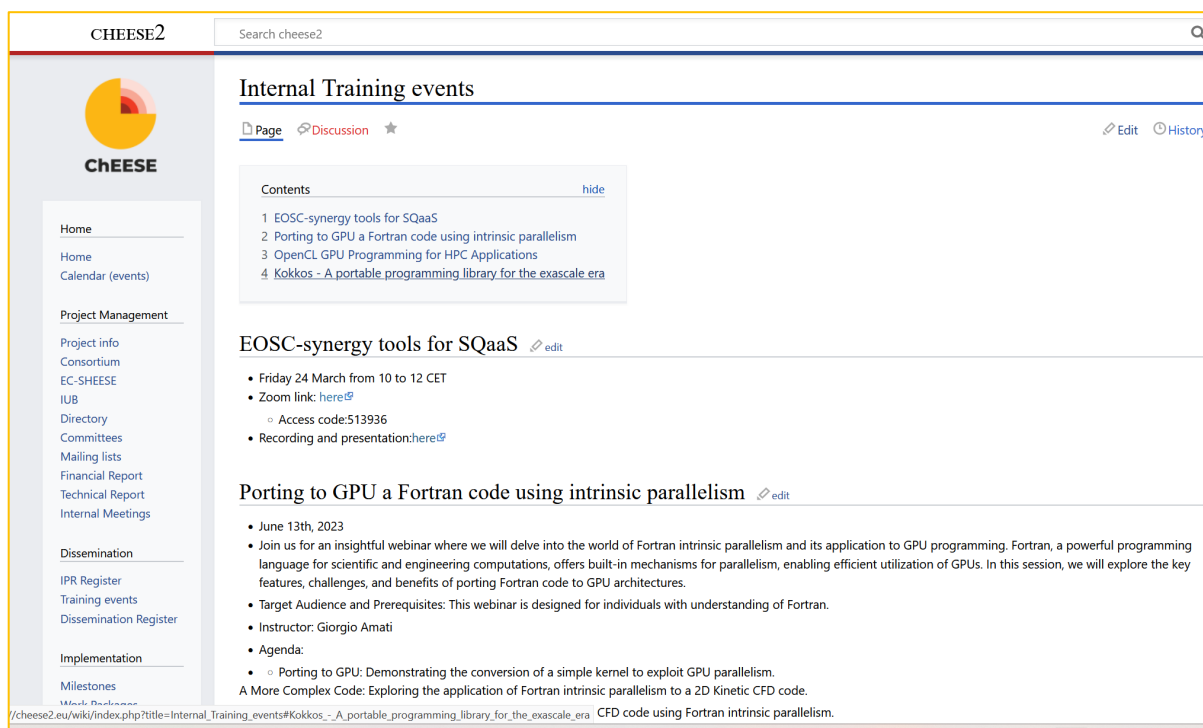


Figure 2. ChEESE-2P wiki Training Events screen

Since there is considerable interest among the community to learn more about computational geosciences and the use of ChEESE codes, we decided to create an HPC for Geosciences series with the following 4 courses:

- Computational seismology (this course will be carried out twice during the project lifetime)

- Computational volcanology
- Computational geodynamics
- Tsunami and meteo-tsunami modelling

ChEESE has also been in touch with CASTIEL-2, which has announced the intention to create a series of thematic workshops, with AI being one of the potential courses. ChEESE has already expressed a deep interest in collaborating with CASTIEL-2 on this course, with our expertise being in Artificial Intelligence and Machine Learning for Geosciences. ChEESE partners are highly motivated to teach this course since the geosciences and SE community has mentioned this topic several times in surveys. Moreover, the collaboration with CASTIEL-2 will be ongoing in order to avoid duplicating training events and efforts.

Finally, we observed a high interest in HPC applications, which are very important for understanding computational geosciences. Since it is a quite popular topic that goes beyond geoscience, the ChEESE team plans to collaborate in organising this type of event with other NCCs and members of the large HPC community.

3.1 Training List

This section presents the list of foreseen training events, which includes those that have already taken place during the first year of this second phase of ChEESE.

It is worth noting that this plan was devised by considering three types of events, each categorized according to its specific structure.

1. **Workshops:** Interactive gatherings for experts to share knowledge and collaborate on specific topics in SE sciences.
2. **Training Courses:** Programs providing practical learning and skill development in computational methods for SE sciences.
3. **Webinars:** Online sessions focusing on specific topics or techniques in SE sciences, enabling participants to acquire new knowledge and insights.

In addition, these might be either internal or open to others depending on several factors:

- If the event has a large hands-on component

- If it targets a very small group and a topic only relevant to ChEESE-2P goals
- If the information presented is sensitive and is not yet ready to be shared with the public

In all other cases, our training events will be open to other stakeholders to maximize the impact and promote collaboration of the SE ecosystem.

Date	Theme	Title/Topic	Type of event/	Location
24 March 2023		ChEESE training event on EOSC-synergy tools for SQaaS	Internal Webinar	Online
16 June 2023	Programming interfaces and languages	Webinar: Porting to GPU a Fortran code using intrinsic parallelism	Webinar	Online
28 June 2023		Webinar: OpenCL GPU Programming for HPC Applications	Webinar	Online
7 July 2023		Kokko: A portable programming library for the exascale era	Webinar	Online
9-12 May 2023	HPC for Geosciences Series	2023 High-Performance Seismology (HPS) CyberTraining	Training	Online
Q4 2023		Computational volcanology	Training	TBC (Online or at BSC facilities)
Q2 2024		Computational geodynamics	Training	In person (Mainz, Germany)
Q2 2024		Tsunami and Meteo-tsunami modelling	Training	TBC (online or at UMA facilities)
Q3/Q4 2024		Computational seismology I	Training	Online
Q3/Q4 2025		Computational seismology II	Training	Online
Q2/2024	Artificial Intelligence	Artificial Intelligence and Machine Learning for Geosciences	Training	In person (BSC)

Table 1. Training list

3.2 Brief description of thematic courses

3.2.1 Computational volcanology: Modelling Volcanic plumes using European flagship codes

This course will focus on modelling one of the most impactful natural hazards: volcanic eruptions. The aim is to provide a brief theoretical overview and then introduce students to the application of different HPC flagship codes used in ChEESE. Specifically, participants will learn about OpenPDAC, an offline Eulerian-Lagrangian model for the simulation of volcanic fluids. The model solves the conservation equations of mass, momentum and energy for a Eulerian-Eulerian compressible multiphase mixture of gasses and fine (up to mm size) particles, and the Lagrangian transport equations for coarser solid particles. The solver is based on the equations of the PDAC model, ported and optimized on the open-source OpenFOAM C++ toolbox for Computational Fluid Dynamics (CFD).

Additionally, the course will cover FALL3D, a Eulerian model for the atmospheric transport and ground deposition of volcanic tephra (ash). FALL3D is commonly used in operational volcanic ash dispersion forecasts to avoid aircraft encounters with volcanic ash clouds and to perform re-routings to avoid contaminated airspace areas.

3.2.2 Computational geodynamics

This course will be delivered face-to-face and will cover the general basics of computational geodynamics. Students will not only acquire theoretical knowledge, but will also gain practical experience through activities such as the creation and visualisation of model setups using Julia and Paraview. In addition, participants will have the opportunity to work with two 3D Geodynamics codes, LaMEM and pTatin3D, which are both ChEESE flagship codes.

Participants will receive hands-on training on how to use Julia, how to create model setups for geodynamic models, and on how to run the geodynamics software packages on both their own machines and on HPC-systems. There will also be time to help the participants with their own geodynamic modelling projects that uses these tools.

3.2.3 Tsunami and Meteo-tsunami modelling and numerical simulation

The aim of the present course is to provide the participants with the knowledge and numerical tools required to perform the mathematical modelling and numerical simulation of tsunamis and meteo-tsunamis. A brief theoretical overview on the processes and mechanisms producing this kind of natural events will be first provided. Then, the European flagship code, Tsunami-HySEA, will be used to simulate various tsunami and meteo-tsunami events. Tsunami-HySEA solves the 2D shallow water equations on hydrostatic and dispersive versions. Based on a high-order Finite Volume (FV) discretisation (hydrostatic) with Finite Differences (FD) for the dispersive version on two-way structured nested meshes in spherical coordinates. MeteoT-HySEA is a version of Tsunami-HySEA code able to include meteorological forcing (atmospheric pressure and wind stress) as input, more precisely as boundary conditions.

The training course will mainly consist of hands-on sessions, the course will also tackle pre-process and post-process strategies based on Python scripts.

3.2.4 Computational seismology I and II

During these courses, participants will receive introductory hands-on tutorials for ExaHyPE, Salvus, SeisSol and SPECFEM3D, which are four different high-performance computational seismology software packages focused on waveform modelling and inversion. The workshop will provide an overview of their similarities, differences and application areas. In the ExaHyPE module, participants will learn to implement the acoustic wave equation and explore more complex seismic modelling possibilities. The SPECFEM3D module covers setting up diverse seismology simulations, using internal mesher MESHFEM3D, forward modelling, and providing an overview of inversion capabilities, including gradient computation and full waveform inversion. The course includes an interactive overview of the full-waveform inversion (FWI) process based on Salvus. Participants will learn about data processing, adjoint-source generation, gradient computation, workflow management, and nonlinear model updates. By the end of the course, participants will be equipped to set up and solve realistic FWI problems independently. Participants will gain a comprehensive insight into the usage of SeisSol which is a software package based on the arbitrary

high-order accurate derivative discontinuous Galerkin method for simulating wave propagation and earthquake rupture.

3.2.5 Artificial Intelligence and Machine Learning for Geosciences

This course will provide attendees with an introduction to the fundamental tools underpinning Artificial Intelligence (AI) algorithms, from linear regression to deep neural networks. The machine learning cookbook, that comprises training, testing and subsequent generalization will be illustrated by a series of hands-on tutorials. These will cover the problems of classification and regression using a series of homemade notebooks deployed over hybrid CPU and GPU architectures.

4. Special series of training courses focused on equality

Since the beginning of the project, ChEESE has demonstrated a strong commitment to equality and inclusion. This commitment has been reflected in actions such as the creation of an Equality Committee and the successful [ShEESE](#) initiative, which was created in ChEESE -1P, with the aim of gender balance within the project.

In addition to these actions, we are currently preparing a series of specialized training courses focused on promoting equality and gender balance in the field of high-performance computing (HPC). ChEESE will be responsible for organising these courses and will extend invitations to other Centers of Excellence (CoEs).

While some of these workshops will be open to all, we will also dedicate specific sessions to addressing the issue of advancing gender balance in HPC, with a particular focus on encouraging female participation.

Some of the topics of these workshops are as follows:

- How to move towards inclusiveness in European Projects
- Overcoming the impostor syndrome as a female scientist
- Leadership tools for women in HPC

The specific details of these workshops are currently being discussed in the Equality Committee meetings. However, the general plan is to offer three or four workshops. The “Special series of training courses focused on equality” has been included in D7.3: Collaboration Plan with Complementary Projects.

5. Training assessment: quality assurance

The post-event survey is an ongoing mechanism that plays a vital role in assessing the quality and impact of the training. The information collected is regularly sent to the trainers, enabling them to enhance the delivery of future training activities.

Additionally, we consistently review the data and carefully consider any suggestions to improve the organization of these events.

As previously mentioned, the survey also helps us define new training needs through the projects, since we include one question that covers this aspect. This is particularly relevant, because it helps us ensure we are offering all ChEESE members the necessary skills to achieve the project goals.

Finally, the information of the assessments will be compiled and presented as part of the Training Report (D7.5), which will be submitted at M48.

6. Training Collaboration with CASTIEL-2, CoEs and NCCs

It is in everyone's best interest to promote synergies with other projects to enhance our training impact. The collaboration goals for ChEESE-2P are defined in the GA and in Task T7.6 in particular: Collaboration task with complementary projects.

In this line, we will rely on CASTIEL-2 pivotal role to coordinate NCCs and CoEs interactions. For this reason, ChEESE-2P training champions, Maria Jose Garcia (CSIC) and Maria Paz Baghetti (BSC) will collaborate with CASTIEL-2 WP3: "Training, Twinning, Mentoring".

More specifically, ChEESE-2P Training champions will attend the monthly Training Coffee breaks to interact with the NCC Training champions and deputies. We will also look for opportunities to organise joint training courses with the NCCs. In the future, as it is being developed at the moment, we are also planning to promote our training events in the training registry (HPC in Europe portal). Finally, we will participate in CASTIEL-2 task force on NCCs-CoEs training-related collaborations and exchange of materials.

7. Conclusions

With this document we have tried to outline a preliminary plan that will serve as a reference during the project.

SE hosts large Exascale Computational Challenges. As we expect new training needs to come up as the project progresses, when new questions are posed, we will react accordingly to provide the best training solutions to match those needs.

With a flexible approach that includes face-to-face, online, and hybrid training sessions, the project aims to reach a broader audience and maximize the impact of its training events. By recording virtual sessions and disseminating information through various channels, ChEESE-2P ensures accessibility and knowledge sharing, benefiting both project members and the wider scientific community.

In summary, the training plan is dynamic, responsive to evolving needs, and designed to equip participants with essential skills and knowledge to meet the challenges of computational geosciences and the pre-exascale and exascale era of supercomputers. Through continuous assessment and collaboration, the project aims to deliver high-quality training events that contribute significantly to advancing research in the field.