H2020-INFRAEDI-2018-2020



Concerted action for the European HPC CoEs

Project Number: 823964

D4.2 – Final Training Report WP4 – Training

© 2022 FocusCoE Project Consortium



Focus CoE has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement N° 823964

Workpackage:	4	Training		
Due Date	31/03/2022			
Author(s):	Simon Wong			ICHEC (NUIG)
	Alan O'Cais			FZJ
	Lore	enzo Zanon		USTUTT
	Emily Lumley			UCL
Authorized by	Simon Wong			ICHEC (NUIG)
Reviewer	Guntram Berti			SCAPOS
Reviewer	Renata Gimenez		BSC	
Reviewer	Andrea Quintiliani		ENEA	
Final Version	31/0	3/2022		
Date				
Dissemination	Public			
Level				

Date	Author	Comments	Version	Status
2022-03-26	Simon Wong	Consolidated draft	0.6	Draft
2022-03-28	Guntram Berti	Internal project review comments	0.7	Draft
2022-03-30	Simon Wong	Release candidate with internal	0.8	Draft
		review comments incorporated		
2022-03-31	Hans-Christian	Final release version	1.0	Final
	Hoppe			
2022-08-30	Simon Wong	Amended final version to address	1.1	Final
		reviewers' comments		
2022-09-28	Simon Wong	Minor revision	1.2	Final

List of abbreviations

BioExcel	HPC CoE: European Centre of Excellence for Biomolecular Research
BSC	Barcelona Supercomputing Center
CPU	Central Processing Unit
CoE	Centre of Excellence
CSA	Coordination and Support Actions
E-CAM	HPC CoE: An e-Infrastructure for Software, Training and
	Consultancy in Simulation and Modelling
EC	European Commission
ENEA	Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile
EoCoE	HPC CoE: Energy Oriented Center of Excellence
ESiWACE	HPC CoE: Center of Excellence in Simulation of Weather and
	Climate in Europe
EU	European Union
EU13	Refers to the 13 countries who joined the EU in 2004 (Cyprus,
	Czechia, Estonia, Hungary, Latvia, Lithuania, Malta, Poland,
	Slovakia, Slovenia), 2007 (Bulgaria, Romania), and 2013 (Croatia)
EuroHPC JU	European High Performance Computing Joint Undertaking
FZJ	Forschungszentrum Jülich GmbH
GPU	Graphics Processing Unit
HiDALGO	HPC CoE: HPC and Big Data Technologies for Global Systems
HPC	High Performance Computing
HTML	Hypertext Markup Language
LAMMPS	Large-scale Atomic-Molecular Massively Parallel Simulator
MaX	HPC CoE: Materials design at Exascale
MPI	Message Passing Interface
MSc	Master of Science
POP/POP2	HPC CoE: Performance Optimisation and Productivity Centre of
	Excellence in Computing Applications
PRACE	Partnership for Advanced Computing in Europe
RSS	Really Simple Syndication

Executive Summary

FocusCoE is a Coordination & Support Action (CSA) established to assist the European HPC Centres of Excellence (HPC CoEs) in fulfilling their role in the European HPC ecosystem: ensuring that extreme scale applications result in tangible benefits for addressing scientific, industrial or societal challenges. HPC training and education plays an important role in sustaining a steady stream of users and developers of these extreme scale HPC applications.

WP4 supported the training activities of the CoEs and realised transversal benefits to the CoEs. Its activities included establishing a central registry of CoE training events, examining an overview of the CoE training programmes as a collective, developing sample training materials, providing transversal training courses for CoE personnel and organising important events supporting the planning of enhanced HPC education and training in Europe.

The CoE registry developed by WP4 aggregates all CoE training events. It uses a recognised pan-European platform (the HPC in Europe portal) and is supported by the PRACE and CASTIEL/EuroCC projects. An interface to this registry is included in the hpccoe.eu website, and for the first time offers all CoE training events and materials in one place. Ultimately, it serves HPC users by making CoE courses more visible and easy to locate and browse, the CoEs themselves by increasing the circle of users of their material, and the European HPC training effort by highlighting the extent of HPC training activities across Europe and enabling further analysis to best shape future training activities.

An analysis of 152 CoE training events revealed good coverage of the scientific and technical domains represented by the CoEs. In the process, the CoEs gained a much deeper understanding of each other's training portfolios.

Based on the Carpentries approach [9], WP4 developed new interactive training materials for the molecular dynamics code LAMMPS and the blood flow simulation application HemeLB. These courses focus on problems and solutions for running HPC applications at large scales, and include presentations, actual work with code examples and knowledge tests in an engaging, interactive setting.

WP4 organised transversal training courses for the benefit of CoE researchers and staff, including an online pedagogy course covering teaching skills as well as two business sustainability and commercialisation courses. Topical workshops discussed important issues in co-design and sustainability, facilitating the CoEs to advance their joint positions.

Finally, WP4 organised two pan-European stakeholder workshops. The first one on HPC training and skills development needs attracted 32 experts and stakeholder representatives and started useful and ongoing discussions on the future of HPC education and training in Europe. The second one took place at the virtual EuroHPC Summit Week 2021 and included highlights of CoE training offers and short presentations from EU13 training representatives. The event was very well attended (over 100 unique viewers) and is seen as an important step towards an EU-wide coordinated HPC training program.

In conclusion, WP4 has coordinated and delivered a significant programme of activities and has fully achieved its objectives. It has brought into view the value of the collective CoE training offering, and provided support for increased awareness and effectiveness. Furthermore WP4 has organised courses and workshops that are of transversal benefits to the CoEs. The CoEs have a vital role in education and training within the European HPC ecosystem, and WP4 has played its part in this ongoing venture.

Table of Contents

1	Inti	roduction	7
2	De	velopment of a pan-European HPC Training Registry	8
	2.1	Development of a pan-European training registry	9
	2.2	Transition to an Alternative Solution and the HPC in Europe Portal	9
	2.3	Extending the Functionality of the HPC in Europe Portal	10
	2.4	Agreement Between Stakeholders on Metadata	11
	2.5	Instructions and Assistance to CoE Course Organisers	12
	2.6	Outlook for the Pan-European and CoE Training Registry	12
3	Ov	erview and Analysis of CoE training activities	13
4	De	velopment of Sample Training Materials	15
	4.1	Methodology and Assembly of Materials	15
	4.2	Lesson Structure	16
	4.3	Conclusion	16
5	Tra	insversal Training Courses/Workshops for CoEs	17
	5.1	Pedagogical Skills	17
	5.2	Co-design Workshop	17
	5.3	Business Development and Sustainability	18
	5.3	.1 Sustainability Workshop Day 1	18
	5.3	.2 Sustainability Workshop Day 2	19
	5.3	.3 Sustainability and Commercialisation Training - ConceptionX	20
6	Par	n-European Training Stakeholder Workshop	21
7	Co	nclusions	22
8	Bib	bliography	23
9	An	nex	24
	9.1	Mini-Guide for CoEs and HPC NCCs – Adding Training Metadata to HPC in	24
		Ешоре гона	

Table of Figures

Figure 1. The new CoE Training Registry on the hpccoe.eu website	8
Figure 2. Screenshot of the HPC in Europe portal.	10
Figure 3. Breakdown of instruction level and target audience of CoE training events	14
Figure 4. Breakdown of CoE courses according to scientific domains.	14
Figure 5. Virtual EHPCSW'21 workshop "HPC education and training"	21

1 Introduction

Skills development of researchers and application developers is key for the European HPC ecosystem as it evolves with new technologies and applications towards Exascale computing. There is a constant need to maintain pipelines of talent who will be responsible for designing, building, tuning and exploiting HPC systems and software for tackling large and complex problems in research and industry. The importance of education and training is reflected in all the European HPC CoEs who have dedicated considerable resources in the field of education and training.

WP4 of the FocusCoE project supported the training activities of the CoEs and realise transversal benefits to the CoEs. The intended activities were split over a number of objectives and tasks that are described in subsequent sections of this deliverable.

First and foremost WP4 contributed significantly towards the development of a pan-European HPC training registry as part of the HPC in Europe portal, in the interest of CoEs but also in collaboration with others, that offers a one stop shop for one to browse and search across HPC and computational sciences training in Europe, including the training offering from CoEs, PRACE, EuroCC (mediated via CASTIEL), other projects and national HPC centres (Section 2). For the CoEs, this significantly extends the reach of their training activities compared to individual dissemination actions. For parties interest in HPC training, this simplifies the task of locating the courses matching their demand and level. Other tasks provided an overview and analysis of CoE training activities collectively (Section 3), developed sample training courses and materials on running domain-specific applications at scale (Section 4), and organised transversal training and services for CoEs (Section 5). WP4 was also engaged in support activities for the strategic planning of HPC training and education for Europe. It organised a comprehensive European stakeholder workshop on HPC training and skills development needs (Section 6). Furthermore, and at the request of the European Commission representatives within the organising committee, WP4 in cooperation with WP2 and WP5 held a training workshop at the EuroHPC Summit Week 2021 which highlighted CoE training and gave at platform for EU13 countries to express their HPC training opportunities and challenges (also described in Section 6).

Two prior deliverables D4.1: "Interim Training Report" [1] and D4.3: "Report on the outcomes of the comprehensive European stakeholder workshop on HPC training and skills development needs" [2] (submitted at M18 and M12 respectively) have described the activities and outcome from the first half of the FocusCoE project across the objectives and tasks as set out above. This deliverable focuses mainly on activities carried out in the second half of the project.

2 Development of a pan-European HPC Training Registry

As described in D4.1 [1], much of the initial discussions with the CoEs concerned the development of a pan-European HPC training registry. This meant a single gateway (e.g. presented as a web page or portal) that would allow collective dissemination of training material on a pan-European basis, e.g. where HPC users can browse and search for relevant training courses in HPC and related fields. It is important to recognise that while this gateway should cater to the interest of the CoEs, i.e. CoE training events can readily take advantage and benefit from such a platform, such a gateway should be pan-European in nature that caters for HPC training offerings from other projects such as those from PRACE, EuroCC and others in order to avoid fragmentation and duplication of efforts. Therefore, WP4 did not seek to develop a CoE-specific training registry platform in isolation. However, as can be seen below, our approach to develop such a platform has enabled FocusCoE to host the *only* gateway where all CoE training events can be found in one place, including upcoming and past events at the hpccoe.eu website (see Figure 1 and [3]). Additionally, it points the users to a pan-European training registry [4] where besides all the CoE offerings accessible through hpccoe.eu, other courses can be found.



Figure 1. The new CoE Training Registry on the hpccoe.eu website developed by FocusCoE WP4 [3]. At the time of writing, this was the only web resource where all CoE training events are collected and shown together. In addition, a link to the HPC in Europe All Short Courses site [4] is provided where users could find information from a repository of European HPC courses.

2.1 Development of a pan-European training registry

During the first half of the project, T4.1 liaised with PRACE and the CoEs to advise on suitable tools and mechanisms to develop a registry of training events. Requirements were taken into account such as the need of an openly accessible platform/service, the ability to aggregate event metadata from different sources, handles for users to navigate and find relevant courses (e.g. filters) and lastly the ability for different websites, external to the gateway, to embed all or part of this registry (e.g. only events from one or all CoEs). It is important to note that these requirements cannot be met by adopting a typical "joint calendar" solution – tools or services that fulfil such requirements are few and far between. In addition, training events from different sources come tagged with different metadata, and a key requirement for a common platform is that it can support and integrate this information, and that it will enable users to locate specific events using metadata queries. This goes beyond browsing by titles and chronological order.

A third party commercial web-based event management tool, Timely [5], was identified as a platform that addressed most of the identified requirements. Hence initial efforts have been dedicated to adapt Timely into a solution for FocusCoE in collaboration with PRACE, another important stakeholder in the provision of a large programme of HPC training complementing those offered by the CoEs. As described in D4.1, this had provided a satisfactory solution that can serve a wide variety of European HPC training providers. The platform was included in the hpccoe.eu and PRACE training websites.

2.2 Transition to an Alternative Solution and the HPC in Europe Portal

At the M18 review, the reliance on a subscription-based commercial service (Timely) was flagged as an unacceptable risk, hence WP4 explored alternative solutions. It became apparent that there are no turnkey solutions similar to Timely from the open source community. There were also insufficient resources to write one from scratch. So in order to realise an alternative, a certain level of technical development work was to be required to expand/adapt the functionality of an existing training registry platform. It is important to emphasise that the extent of such technical development work, required to replicate entirely all the features and the user interface of a fullyfledged third party commercial service, is beyond that of limited WP4 resources. Nevertheless, it can be seen below that FocusCoE had been instrumental in developing a joint collaboration to build the foundations of a common training registry. While much effort has been invested in developing the current registry with basic functions (listing of events, basic search), much more is required in future to replicate some of the functions provided previously by Timely, e.g. different ways to view event listings, more flexibility for external embedding, polished user interface.

Incidentally the PRACE-6IP project had been developing a pan-European HPC services registry called the "HPC in Europe portal" [6] since 2019. While built by PRACE this portal was designed to be a "neutral" (with no overt PRACE branding) catalogue of European HPC services, to serve other European projects and initiatives. The portal was designed to cater for a variety of HPC services including training, access to supercomputing resources, support, overview of available software, hardware and applications at the European level. There was a basic and generic data schema to catalogue any HPC service to be disseminated within this portal, e.g. description, availability, an URL pointing to an external website that provides more information about the service.



Figure 2. Screenshot of the HPC in Europe portal that aggregates pan-European HPC services.

While it is somewhat basic in terms of functionality (and designed to be so), the HPC in Europe portal provides a reasonable platform, or at least one with the best potential, that would also cater for the needs to develop a CoE training registry by WP4 and promise sustainable operation. It was established as a transversal portal not tied to any particular project as prescribed by the EC. Embedding of information held in the HPC in Europe portal by external sites was one of the requirements. The design and structure of the portal had also been disseminated at the ISC and SC conferences for HPC ecosystem stakeholders, i.e. there was already a growing level of awareness of this pan-European platform. However, it soon became clear that the HPC in Europe portal needed extensions in order to satisfy our requirements for a CoE training registry. This has been the focus of T4.1 in the second half of the FocusCoE project.

The emergence of the National Competence Centres (NCCs) from the EuroCC project, coordinated by CASTIEL with a similar need to develop a registry of NCC training events, resulted in a third major stakeholder who shares an interest in extending the HPC in Europe portal for training.

Ultimately, this led to a successful collaboration between three major European HPC projects that are associated with training delivery – FocusCoE (on behalf of the CoEs), PRACE and CASTIEL (on behalf of the NCCs from the EuroCC project). The relationship was formalised in a MoU (details can be found in [7]) signed by the three parties recognising the contributions from each in the development and future collaboration of the portal. Some of the steps taken to achieve this end are described below.

2.3 Extending the Functionality of the HPC in Europe Portal

As mentioned, the HPC in Europe portal had a lightweight schema for cataloguing generic HPC services. While this may suffice for relatively simple lists of services (e.g. access to supercomputers), there are numerous events from the different stakeholders in training covering many different, diverse fields. Hence for users of this registry in search of training, further categorisation and filters using course metadata is needed, i.e. the HPC in Europe portal needs to include more information about individual courses.

WP4 created a new training service definition, with its own custom schema for metadata (see below) developed acting on CoE input, on the HPC in Europe portal. This allowed CoE course information to be aggregated on the portal in a similar manner to Timely. While this custom

schema is largely similar to that used by Timely, technical work was carried out to establish all the required data fields. For example, internal data stores were set up in the HPC in Europe portal software to deal with dynamic and extendable lists, e.g. of training providers, courses associated with different European projects. Some extra functions were also implemented, such as the ability to show on a geographic map the locations of training venues and organisers, and placeholders for potential future incorporation of schema.org metadata templates (e.g. for training events) that will make the platform more accessible for other stakeholders.

More importantly, WP4 also developed customised views of the registry that facilitate querying and displaying a subset of the registry which can then be embedded, e.g. on the hpccoe.eu website for the CoE training registry (see Figure 1 and [3]). The same feature has been adapted by CASTIEL on its EuroCC Access website to show NCC training events [8].

2.4 Agreement Between Stakeholders on Metadata

For each training event, a coherent set of metadata is collected in order to facilitate users to filter down and find relevant courses. The metadata associated with each event is a heterogeneous mix of fields that require input as free text, multiple selection from a list of options, addresses and geolocation information, or an image file. After extensive discussions between FocusCoE, PRACE and CASTIEL – the different stakeholders organise a very diverse range of courses that should be accommodated on the portal – a set of metadata was agreed up on by FocusCoE, PRACE and CASTIEL. FocusCoE consulted with the CoEs and was heavily influenced by their input during these discussions. This can be roughly split into three groups as shown below.

Events details include the following fields:

- Title
- Primary Event/Course URL
- Event/Course start
- Event/Course end
- Event/Course Format
- Online learning mode
- Venue Details Link
- Location
- Languages
- Event/Course Description + Summary
- Image

As an example of a small improvement over the first iteration of the portal, the course format ("Presence: In-person, Online") has been extended to include the option "Mixed". An additional field allows one to specify the online learning mode: Live (synchronous), Self-paced (asynchronous), Blended (mixture of live and self-paced). The venue details link allows the organiser to include a link to a physical location where the course takes place, or provide connection details in the case of an online course. The Summary is an optional field that allows entering a short abstract. Together with the Image, the user can provide an alternative text for accessibility purposes, and a title to include source and copyright information.

Metadata in addition to the core event details includes:

- Sector of the Target Audience
- HPC profile of target audience
- Country
- Organisers
- Supporting Project(s)

The target audience fields allows one to specify whether courses are targeted at a specific sector (e.g. academia, industry) and a particular cohort of HPC users (e.g. application users, developers, system administrators). Organisers are not entered as free text but a small database is kept internally to facilitate auto-completion and consistency. All the CoEs were granted access to the portal by default who would then be added to this database. Multiple organisers can be specified per event. The list of Supporting Projects include the CoEs and FocusCoE, EuroCC/CASTIEL and PRACE. This field is especially important to filter events when embedding custom views of the registry, e.g. on a particular CoE's webpage.

Finally the following fields help to define the scope of the training events that may be filtered by users:

- Scientific domain
- Technical Domain
- Level of Instruction

These have been slightly modified from the first iteration of the portal in agreement with EuroCC/CASTIEL (e.g., within Scientific Domain, "mathematics and computer sciences" has been split, "programming languages" and "quantum computing" have been added to the list of Technical Domains); it was important to ensure consistency for these taxonomic terms. Finally, the Level of Instruction has also been separated into "Basic", "Intermediate" and "Advanced".

2.5 Instructions and Assistance to CoE Course Organisers

The CoEs were provided with a walkthrough of using and entering information into the HPC in Europe Portal. WP4 presented to the CoEs the transition to this platform (away from the Timely-dependent registry) that will provide a central directory of HPC related courses in Europe for the foreseeable future. To facilitate entering events, a short guide has been produced by WP4 with the cooperation of EuroCC/CASTIEL and can be found in the Annex 9.1.

The procedure to enter events (and providers) in the Portal has been made as self-explanatory as possible, with help texts guiding through the list of fields. Still, the short guide gives some hints on data entry steps where the user might be in doubt. The procedure to embed the training calendar in each project's own webpage is also covered.

2.6 Outlook for the Pan-European and CoE Training Registry

As described above, WP4, as part of its mission to develop a CoE training registry, has always conducted this activity in collaboration with the CoEs on one side and other European training providers (e.g. PRACE, CASTIEL/EuroCC) on the other in order to avoid fragmentation and duplication of efforts. It has successfully managed to develop a single (and the only) web presence on the hpccoe.eu website [3] that shows all CoE training offerings in one place. It has also dedicated efforts to transition the back-end system from reliance on a North American commercial service to the HPC in Europe portal that was first developed by PRACE and is now recognised as a joint venture with commitments from both FocusCoE and EuroCC/CASTIEL to contribute to its development. This is reinforced by the signing of the aforementioned MoU between the three parties with specific regards to the training aspect of the portal.

With regards to the future of the HPC in Europe portal, there have been indications to the EC that this portal is an important infrastructure of the European HPC ecosystem, and FocusCoE did migrate the training registry to this portal. Again, it is important to note that this portal has only basic functions at present and will require continuous funding and effort to further develop and maintain this infrastructure in a stable manner.

3 Overview and Analysis of CoE training activities

All of the CoEs have training activities that target specific scientific domains and applications, with perhaps the exception of the POP2 project which is transversal across scientific disciplines they provide practical HPC performance optimisation and profiling courses as a service that caters for different communities, including the other CoEs. In the first half of the FocusCoE project, WP4 completed a high level overview and analysis of CoE training activities which focused on 39 training events organised by the CoEs from January 2019 to February 2020. These events consisted of training courses, workshops, or hackathons organised by the CoEs, but excluding webinars which are offered by the CoEs but are diverse in topic (e.g. there was a fair amount of scientific lectures). The analysis showed that, taking into account the nature of the CoEs that are not evenly spread among all scientific disciplines, the specific domains represented by the CoEs are well covered. With regards to technical content (i.e. HPC topics such as parallel programming, performance engineering or best practices in application usage), many of the CoE courses are focused towards tuning or using domain-specific applications. There have been good examples of collaboration between CoE on co-organising training events. The only gap was perhaps the geographical location of these predominantly in-person courses - there was a paucity of courses at EU13 countries.

In the second half of the project, the analysis of the CoE training offering was updated with the analysis of 152 events from the CoEs (again excluding short webinars). These events consisted of training courses, hackathons, workshops and a study module organised by the CoEs. The data includes 48 in-person events, 103 online events (which have largely been missing from the first analysis which is mostly pre-COVID) and 1 hybrid event. The events were annotated with respect to the fields used in the training registry to facilitate joint analysis.

Figure 3 shows the breakdown of CoE courses into instructional levels and the HPC use profile of the target audience. It can be seen that there is a good balance between basic courses (which often cover a larger percentage of the user base) and the more intermediate to advanced courses among the CoEs. There is evidence that the target audience of these courses cater more towards application users (47%) and data scientists (18%), this is expected as the CoEs provides training that is geared towards instructing users on getting the best out of HPC applications. For the courses aimed at application developers – these fall into two groups: POP2 performance engineering courses/workshops and courses offered by the CoEs in developing extensions, modules/libraries and code coupling for domain-specific applications. The latter comprises over 90% of CoE courses aimed at application developers; some of these involve collaboration with PRACE and POP2 who tend to provide more generic HPC courses (e.g. parallel programming, performance optimisation). The results appear to show an appreciable complementarity between training offerings from the CoEs (domain-specific) and PRACE/POP2 (generic HPC programming and performance tuning), as well as evidence of good collaboration between all those projects.

© 2022 FocusCoE Project Consortium



Figure 3. Breakdown of instruction level and target audience of CoE training events.

Figure 4 shows a breakdown of the CoE training courses according to their scientific domains (as defined in the training registry, which is a categorisation of scientific domains used in applications for supercomputing resources). It can be seen that CoE training is slightly more focused towards biochemistry, bioinformatics and life sciences, chemical sciences and materials and Earth system sciences. This result is more or less expected due to the nature of the CoEs that have been established. The scientific domains that are not represented at all include fundamental physics, fundamental constituents of matter and universe sciences. This is hardly surprising, as there are no CoEs in these fields. In any case, the distribution of courses among scientific domains is largely reflective of the CoE domains.



Figure 4. Breakdown of CoE courses according to scientific domains.

Finally, it was highlighted in the first reporting period that many of the CoE courses have transversal benefits, i.e. they are of particular relevance to other CoEs. Hence, and in particular in the second reporting period, WP4 regularly promoted CoE training events to

training/communication contacts of the other CoEs (via Slack and email) to keep the CoEs informed of upcoming events.

4 Development of Sample Training Materials

Each CoE typically focuses on a set of applications that can be scaled up and deployed on Exascale class systems. In order to maximise the impact of the CoEs, training should be available for the respective user communities in the caveats of running their specific applications efficiently at scale. For this type of community-oriented training activity, we advised the CoEs that there should be some uniformity of this training offering across the CoEs (with respect to length, scope, pedagogical approach, etc).

FocusCoE T4.2 leveraged the Carpentries [9] approach to create collaboratively developed, well structured, open source training content that can be delivered by "non-experts", hence enabling a much more scalable training concept. The idea was to develop pilot materials to serve as an example, the same concepts would then be transferable to developing similar training materials for other HPC applications.

LAMMPS [10] was the first HPC application that was selected for creating a first set of sample training material about. It is a classical molecular dynamics code with a focus on materials modelling, which is also of particular relevance to the E-CAM CoE who were collaborating on the task. Importantly, the application is used on a wide spectrum of hardware, from laptops to the top supercomputers in the world. While the application has the potential to scale, the user often has to determine for oneself whether optimal performance is being achieved. The set of materials on LAMMPS produced by WP4 (available at [11] online), forms the basis of a half-day course. The materials were used in a successful pilot on 9th November 2020, where FocusCoE organised a workshop for a group of users and introduced them both to the concept of this type of course and went through the materials itself.

Based on the success in building the first lesson "Running LAMMPS on HPC systems", and in order to demonstrate the general applicability of this approach, T4.2 turned to develop a similar implementation of a Carpentries style lesson for another highly scalable application. HemeLB [12] is a high performance lattice-Boltzmann solver optimised for simulating blood flow developed by the team of Prof. Peter Coveney of University College London (UCL). It was selected as the material on which the second example lesson would be based since the solver is scalable to hundreds of thousands of cores, and the scalability has been tested and validated on some of the largest supercomputers in the world (e.g. SuperMUC-NG at LRZ, and Archer2 at EPCC, SUMMIT at Oak Ridge National Laboratory).

4.1 Methodology and Assembly of Materials

As with all Carpentries workshops, the material was created and maintained in a GitHub repository, in which the GitHub Pages option is enabled to generate an interactive webpage of the repository. This was set to a particular style which is used in all Carpentries lessons which matches the teaching style and layout of Carpentries [9]. The HemeLB materials have since been made publicly available at [13].

WP4 worked collaboratively with HemeLB developers, who were introduced to the Carpentries approach, to develop the training materials. Most of the materials have been inspired by the layout provided in the FocusCoE workshop on "Running LAMMPS on HPC systems" mentioned above, so there are similarities between the two. Both codes are commonly used on large supercomputers and use both MPI and OpenMP. However, LAMMPS and HemeLB they follow different approaches to improve scalability and performance.

4.2 Lesson Structure

For a typical Carpentries lesson, there are a collection of episodes, which contain a number of sections and exercises that relate to a certain topic. These episodes and sections therein provide a transcript for an instructor to work from or indeed for someone to work through independently. The sections in a given episode cover a topic from start to finish, and where possible, an exercise is added to ensure the concepts which have been taught are understood and can be implemented by the learner.

In the case of "Running HemeLB on HPC systems", the basic framework of material was constructed based on the equivalent LAMMPS lesson [11]. The exact material varies depending on what the developers consider as important and necessary for the end-user. The topics covered should largely cover the common performance issues with the application, when and how to configure the application depending on the use case (e.g. problem size and complexity) and hardware (e.g. availability of accelerator and processor-specific features).

The structure and layout of the episodes were developed as follows:

- 1. *Why bother with performance* provides an overview of why performance is important when running high performance libraries, as well as introducing the important HPC concepts (e.g. FLOPs, CPU hours)
- 2. *Connecting performance to hardware* gives an overview of understanding the hardware that an end-user would be utilising, and how to effectively modify code depending on the system type that they are using. This serves as a brief overview of OpenMP and MPI.
- 3. *Benchmarking and scaling* this episode gives a thorough overview of the importance of benchmarking and includes exercises on the fundamental task of running a benchmark with HemeLB on a sample dataset and utilising a batch job script. The concepts of scaling are also introduced.
- 4. *Bottlenecks in HemeLB* depending on the application, different bottlenecks can occur and specific procedures need to be implemented to reduce runtime. In the case of HemeLB, many of the settings are pre-adjusted so the best settings are used as default. This also covers how to understand the HemeLB outputs and how to minimise I/O bottlenecks.
- 5. *Accelerating HemeLB* covers vectorisation and monitoring of HemeLB jobs. Most of the main acceleration mechanisms in HemeLB can be enabled or disabled by commenting out specific sections of code in .xml files, or by certain flags in the compilation script.
- 6. *GPUs with HemeLB* enables users to utilise HemeLB on GPUs by providing a background in GPU theory as well as sample sections of CUDA programming. Secondly, it shows how to modify the compilation script to enable running on GPUs.

The episodes listed above cover all necessary information about how to configure and run HemeLB on HPC systems. These materials can be easily updated to reflect improvements and new code versions For HemeLB, future modifications to the material written to date include enhanced sections on utilising the code with GPUs, or indeed on ARM hardware.

4.3 Conclusion

We have shown that an existing Carpentries style lesson (for LAMMPS [11]) can be easily modified and extended to cover a different application (HemeLB [13]). Such creation of a derivative involves the modification of a few elements, which depend on the relevant performance bottlenecks and possible improvements. Each of these courses is targeted at a specific user community and will certainly promote and increase their awareness and use on HPC systems.

5 Transversal Training Courses/Workshops for CoEs

Task 4.3 identified and organised transversal training courses/opportunities targeted at researchers and staff of the CoEs. The task has identified three topics where transversal courses/workshops are of value to the CoEs, as described in Subsections 5.1 to 5.3 below.

5.1 Pedagogical Skills

Pedagogical skills are important for trainers to teach effectively; yet they are sometimes overlooked, e.g. situations where a very competent developer may be asked to teach a course on a particular technical topic, although that person may not necessarily have sufficient pedagogical skills, which could ultimately reduce the impact of a course. Hence WP4 has organised a pedagogy focused course and made it available to CoE personnel.

While the original plan was to organise an in-person event in Q2 of 2020, this has been delayed due to its coincidence with the COVID pandemic. However, WP4 has successfully organised an online version of the event in November 2020.

The course was organised in collaboration with CodeRefinery [14] who work with students, researchers, and research software engineers from all disciplines to advance software development and management practices, so that research groups can collaboratively develop, test, review and discuss their source code, with the aim of making it shareable and reusable. It has similarities with the Carpentries organisation but adopts a slightly different approach. WP4 partnered with CodeRefinery as they readily provide instructor courses (i.e. teaching trainers how to deliver training effectively) and specialise in trainers who already know the technical details but may benefit from taking different approaches to teaching. Furthermore, they have also developed updated content to help instructor pivoting to online training.

The course was delivered over 3 half-day sessions online on 2nd, 3rd and 9th November 2020 which was attended by 14 participants from various CoEs. The course content included:

- Teaching philosophies
- Interactive teaching style
- How to teach online
- Lesson design (learner profiles, objectives, prerequisites, portability)
- Collaborative lesson development
- Online workshops

Apart from these topics, the course included useful practical sessions (teaching demos and where participants were asked to teach a simple subject and obtain feedback). The final day of the course on 9th November 2020 incorporated the training materials developed by FocusCoE discussed in section 4 above.

The course was well received by the participants who learnt valuable approaches and advice. The course was also extremely interactive with active contributions from all participants. All the materials covered by the course are referenced in the course website [14].

5.2 Co-design Workshop

In considering topics that are of highest importance and transversal value to the CoEs, co-design was identified as a particularly relevant subject for all the CoEs. But co-design can infer different meanings for different CoEs and stakeholders and hence WP4, in collaboration with HPC3 and WP2, organised a co-design workshop so the CoEs have the opportunity to share their views and discuss the important issues.

The online workshop was organised on 12th March 2021 split into two sessions, each with 5 expert panellists from the CoEs to present their views and lead discussions. The first session focused on the "conventional" levels of co-design, e.g. at the chip, network, low level software, programming model and application levels, while the second session explored the concept of co-design for new usage, i.e. Exascale workloads will be more complex that will require co-design of hardware and software tools, code couple, large ensemble runs, *in situ* data analytics. The workshop was of wide interest and attracted over 80 participants. A recording was made available to all CoEs.

5.3 Business Development and Sustainability

Business development and sustainability are key transversal challenges faced by all the CoEs. Engagement with industry is often closely related to how a particular CoE may formulate a model for sustainability. The importance of these types of issues to the CoEs is highlighted by the fact that a special working group on business sustainability has been established by HPC3 (the HPC CoE Council) to examine this issue.

WP4 originally planned to organise a workshop in Q1/Q2 2020 for this special working group to discuss the key issues, with potential expert advice presented by third parties, to help the realisation of the position paper. However, this has been curtailed due to the COVID-19 pandemic. Nevertheless, WP4 has managed to organise two workshops in November 2020 and January 2021 to facilitate CoE discussions and exchange on this topic, as well as rolled out two research commercialisation training courses for CoE staff in spring 2021 and 2022. Details of these events are described below.

WP4 members were very active in the HPC3 Business working group; findings from the WP4 business development and sustainability activities can be found in a public report authored by this working group [15].

5.3.1 Sustainability Workshop Day 1

The first Sustainability workshop was held online on 5th November 2020 and introduced the topic of sustainability. It also took an inventory of status of each of the CoEs with respect to their sustainability plans, progress and knowledge. In total ten CoEs presented their sustainability plans which represented all the established CoEs at the time and the workshop was attended by 36 people across these CoEs and FocusCoE.

The workshop started with an introduction and welcome followed by a presentation on the Innovation Management activities that were underway within FocusCoE WP5. The activities within T5.4 and reported in FocusCoE deliverable D5.5 [16], directly fed into the sustainability aspects of CoEs with innovation management giving an opportunity for sustainable income for the CoEs and their partners. There followed presentations from the following CoEs:

- HiDALGO Creating value solving Global Challenges
- EoCoE Mid-term and long-term sustainability strategy for the EoCoE CoE
- CompBioMed Long-term sustainability scenarios and challenges
- BioExcel Sustainability of Centres of Excellence: Where's my cane?
- E-CAM E-CAM's thoughts on sustainability
- EXCELLERAT Development of Sustainability for CoE EXCELLERAT
- MaX Sustainability plans @ MaX
- POP POP's Pathway to Sustainability
- ESiWACE Steps towards a sustainable infrastructure for Weather and Climate simulation
- ChEESE Long term sustainability of ChEESE CoE

At the end of the meeting a discussion of the future actions from FocusCoE, HPC3 and the CoEs was held. It was found that there are many similarities in the approach of each CoE to sustainability

with a focus on developing funding through services such as training, consultancy and support for users of the applications. However, each CoE had a unique set of challenges in their sustainability and were at different stages in their planning.

Some questions and concerns that were raised were whether the focus should be towards a profit or non-profit organisation, and whether sustainability meant support of the CoE as a whole or of individual concerns within each consortium. Information on how to produce business plans was required by the CoEs so that each could determine the best method of working for their own situation and set of codes. It was clear that sustainability was made more difficult due to the focus on scaling of applications and a drive towards Exascale computing. For these specific activities, external funding will be essential with a reduced focus on innovation and sustainability such as similar initiatives in the US and Switzerland if the EU is going to compete in a serious fashion.

One other essential component within the CoE landscape was identified as community building. The hpccoe.eu website set up by FocusCoE within WP5 helps in this as it introduces people to the diverse set of activities taking place across all the CoEs.

5.3.2 Sustainability Workshop Day 2

The second Sustainability workshop was held online on 21st January 2021 and included presentations from experts in commercialisation and sustainability to inform the CoEs of best practices and examples of success stories on the topic. It was attended by 21 individuals from various CoEs. As with the first workshop, an introduction and welcome was followed by an update on the innovation management report from WP5 of FocusCoE. This presentation included a list of the recommendations for innovation that were produced through this report and therefore, indicated some best practice initiatives that could help the CoEs.

There followed a longer presentation from Andrea Di Anselmo from the META group. The Meta group have previously assisted numerous EU projects with their sustainability and improving the impact of their outcomes. They had worked previously with EoCoE, so were familiar with the CoE ecosystem and the unique concerns of these with respect to the focus on Exascale applications. This presentation gave extensive information to the CoEs on topics such as:

- Determining what adds value
- Key Exploitable Results
- Sustainability, Exploitation and Dissemination and the differences between them
- Results, Impacts and Outcomes
- How to determine the best route for exploiting results (user-driven approach)
- Planning an exploitation roadmap
- How to pitch the idea to the users and other stakeholders

This resulted in a useful discussion between the participants related to standards and how to exploit these in the project. During the presentation and the final discussions, the concept of a "results booster" was introduced and explained, which could be a way of improving outcomes of the research results.

Finally a success story in terms of commercialisation of research results from European Projects was presented. Christopher Morton from ELEM Biotech explained how the Alya code (used in multiple CoEs, but specifically with the computational biomedicine use cases in CompBioMed) had been used to set up a company that can run simulations for in silico clinical trials related to heart models. This took the form of a question and answer session in which Christopher Morton explained the process of setting up the company and participants could ask questions on this process and barriers that were encountered.

5.3.3 Sustainability and Commercialisation Training - ConceptionX

Two training events were held with the Deep Tech Venture Programme ConceptionX (https://conceptionx.org) on 9th April 2021 and 8th March 2022. This was given by Riam Kanso, CEO of ConceptionX who introduced the principles of the company as a programme that helps researchers commercialise their work and start companies. The second presenter was Richard Anson, Academic Lead of ConceptionX who has extensive knowledge and experience both in industry settings and academia.

At the first training event there were 34 participants from six CoEs and the interaction with the training was good. The response to the training was good, so it was decided to repeat it again in 2022, however, although 20 people registered to the event, it was less well attended - just six participants, but these were from some of the newer CoEs. However, despite the reduced numbers, it was still well received and interactions with the trainers was good.

nderstanding The commercial readiness of your
Search How the market sees your technology The role of tech transfer and knowledge insfer An understanding of the relevance of IP How to create and capture value The difference between technological sh and customer pull An appreciation of business models and e stages of search vs execution An appreciation of the importance of stomer centricity How to work with companies to celerate commercialisation.

Both sessions covered the following topics with the associated learning outcomes.

The sessions have been recorded and can therefore be used in the future by other CoEs The downside of the recording is of course that you do not get the opportunity to interact directly with the trainers and get their expert advice on any matters.

6 Pan-European Training Stakeholder Workshop

One of the major tasks of WP4 in the first year of the project was to organise a comprehensive European stakeholder workshop on HPC training and skills development needs. This was to be carried out in collaboration with the European Commission (EC), bringing together experts and stakeholders to define what the training and education requirements are from different domains/target audiences, what the state of the art is as regards HPC training and education, and to examine how the variety of requirements may be met by existing or new European and national programmes on HPC training.

In close collaboration with WP2 and WP5, the workshop was held in Brussels on 7th October 2019, where a total of 32 participants attended the workshop, representing the interests of a range of stakeholders and communities [17]. A report that describes the implementation and outcome from the discussions arising from the workshop was delivered as a separate deliverable D4.3 in November 2019 [2]. Major findings from the report were also presented by the FocusCoE project leader at the EuroHPC JU Governing Board meeting in Luxembourg, February 2020.

While this task has largely been completed by M18, WP4 was also involved in organising another training related workshop at the virtual EuroHPC Summit Week 2021 (EHPCSW'21). At the request of the project officer, the theme of the event should highlight HPC education and training at the EU13 countries. Hence WP4, in collaboration with WP2 and WP5, had organised a workshop on 26th March 2021 entitled "HPC education and training – perspectives from EU13 Member States". The workshop opened with an introduction of the HPC education and landscape, with particular emphasis on the CoE training offers, followed by 12 presentations from HPC training representatives at EU13 Member States (with all presenters invited and coordinated by WP4; only Malta was not represented among the presenters). The presentations demonstrated the value of CoE training and conveyed a diverse landscape of HPC training opportunities and challenges across the EU13 Member States. The online workshop attracted over 100 unique viewers at the Summit Week.



Figure 5. Introduction to CoE training at the virtual EHPCSW'21 workshop "HPC education and training – perspectives from EU13 Member States".

7 Conclusions

Over the course of the FocusCoE project, WP4 has engaged well with the CoEs on important issues related to HPC training and has been successful in completing its various actions in support of the CoEs. These actions provided important support for CoE training activities and delivered training on key transversal topics for their benefit. Overall, the work package has achieved its objectives in full. Key results include:

- Perhaps the most significant outcome is the establishment of the CoE training registry, that makes all CoE training events available aggregated in one place within hpccoe.eu [3]. In achieving this end, WP4 also contributed towards the establishment of a wider pan-European HPC training registry in collaboration with PRACE and CASTIEL/EuroCC. This not only serves HPC users and course attendees by making CoE courses more visible and easy to locate and browse, it highlights the full extent of HPC training activities across Europe, increases distribution of materials and enables analysis to shape future training support activities.
- The collective CoE training offerings were analysed and no major scientific or technical gaps were found, except in domains where CoEs have not been established. The CoEs have gained a better understanding of each other's training portfolio, and WP4 helped to disseminate CoE courses that may be of transversal benefit to other CoEs.
- The production of a full sample set of training materials on using LAMMPS on HPC and extreme-scale systems, and the derivation of a version for the HemeLB application. While the materials are of tremendous value by themselves, they will also serve as a template for other HPC applications at other CoEs.
- WP4 organised three courses and two workshop series of transversal value for the CoEs. This includes courses on pedagogy, sustainability and commercialisation that were delivered by external experts (CodeRefinery and ConceptionX) for CoE personnel. WP4 also took an active part in organising topical workshops that are of relevance and utility for the CoEs in the areas of co-design and sustainability facilitating a forum for exchange of ideas among the CoEs in addressing these issues.
- The successful organisation and reporting of the pan-European HPC training stakeholder workshop instigated extremely useful discussions on the future of HPC training in Europe.
- At the request of the EC, a second pan-European training workshop was organised by WP4. The event, titled "HPC education and training perspectives from EU13 Member States" took place at the virtual EuroHPC Summit Week 2021. It included highlights of CoE training and short presentations from EU13 training representatives on existing activities, opportunities and challenges. The event was very well attended by over 100 unique viewers.

Facing the challenges of an unabated COVID-19 pandemic, WP4 has managed to move key CoE events and workshops to happen online, with significant participation from the targeted CoEs and impressive content. This includes the transversal courses on pedagogy and business development and sustainability, and the widely attended training workshop at the European HPC Summit Week 2021.

Similarly, the examples and instructional materials prepared for the CoEs to assist them in their training activities have shifted focus from in-person events to online meetings. Given all the well-known downsides of meeting in a virtual space, it is clear that online courses will be a significant part of training and education going forward, and the actions of WP4 have equipped the CoEs with important tools & competencies to expand their online outreach and education activities.

8 Bibliography

- [1] S. Wong and A. O'Cais, "FocusCoE D4.1: Interim Training Report".
- [2] S. Wong and A. O'Cais, "FocusCoE Deliverable D4.3: Report on the outcomes of the comprehensive European stakeholder workshop on HPC training and skills development needs".
- [3] "CoE Training Registry on hpccoe.eu," [Online]. Available: https://www.hpccoe.eu/coe-training-calendar/.
- [4] "All Short Courses HPC in Europe Portal," [Online]. Available: https://hpcportal.eu/all-events-courses.
- [5] "Timely event management online services," [Online]. Available: https://time.ly/.
- [6] "HPC in Europe portal," [Online]. Available: https://hpc-portal.eu/.
- [7] "D1.2 The Month 40 Management Report".
- [8] "Training EuroCC ACCESS," [Online]. Available: https://www.eurocc-access.eu/services/training/.
- [9] "Carpentries website," [Online]. Available: https://carpentries.org/.
- [10] "LAMMPS Molecular Dynamics Simulator," [Online]. Available: https://lammps.sandia.gov/.
- [11] "Running LAMMPS on HPC systems," [Online]. Available: https://fzj-jsc.github.io/tuning_lammps/.
- [12] "HemeLB," [Online]. Available: http://hemelb.org.s3-website.eu-west-2.amazonaws.com/.
- [13] "Running HemeLB on HPC systems," [Online]. Available: https://hemelbdev.github.io/HemeLB-Carpentries/.
- [14] "CodeRefinery instructor training for EU HPC CoEs (Nov 2020)," [Online]. Available: https://coderefinery.org/2020-11-02-instructor-training/.
- [15] ""Report on CoE business activities: opportunities, challenges and lessons learned", HPC3 Business Working Group Report on Sustainability," [Online]. Available: https://www.hpccoe.eu/hpc-coe-council/.
- [16] "FocusCoE Deliveravble D5.5 Final Innovation Management Report".
- [17] "European stakeholder workshop on HPC training and skills development needs. Brussels, Oct 2019.," [Online]. Available: https://www.focuscoe.eu/index.php/2019/11/19/european-hpc-training-stakeholder-workshop-in-brussels/.

9 Annex

9.1 Mini-Guide for CoEs and HPC NCCs – Adding Training Metadata to HPC in Europe Portal

The Portal for European HPC Services - Relaunch 2022

Login at https://hpc-portal.eu/user

Please contact Stelios Erotokritou for an account: s.erotokritou@cyi.ac.cy You will get an **Editor** account to add and edit content. **Important**: At the moment, you will be able to manage content entered by yourself only.

Landing page to add training content:

https://hpc-portal.eu/add_new_content

The default landing page to add content is <u>https://hpc-portal.eu/node/add</u>
 You can reach all pages from this link as well, but we recommend to use the first one.

Visualise existing content:

Browse "Training and events" in the main menu, or go directly to: <u>https://hpc-portal.eu/visualise_new_content</u>

In there, you will see hyperlinks to:

- <u>Calendar</u> of upcoming short courses: <u>https://hpc-portal.eu/upcoming-events-courses</u>
- <u>Calendar</u> of ongoing short courses: <u>https://hpc-portal.eu/ongoing-events-courses</u>
- <u>Calendar</u> of all short courses: <u>https://hpc-portal.eu/all-events-courses</u>
- <u>Map</u> of upcoming short courses: <u>https://hpc-portal.eu/upcoming-events-courses-map</u>
- <u>Overview</u> of providers and organisers: <u>https://hpc-portal.eu/providers</u> and map: <u>https://hpc-portal.eu/providers-map</u>
- In the calendar of short courses (upcoming, all), it is possible to **filter by project** only using "Search for keywords" (e.g., enter "EXCELLERAT"). New events are immediately visible but might be correctly filtered only after some time has passed.

Hints to enter training events:

- Venue Details Link: For technical reasons, this field is compulsory. Please read carefully the help-text. If you would like to use the link to the page on the HPC Portal itself (e.g., https://hpc-portal.eu/node/802), enter a placeholder link, save, get the HPC Portal link and edit the page entering this new link.
- All providers in the database (<u>https://hpc-portal.eu/providers</u>) must contain a geolocation to appear on the providers' map (<u>https://hpc-portal.eu/providers-map</u>).
 Warning: Provider pages entered up to end of 2021 are "owned" by the administrators. Please check in the database whether your provider already exists and let us know in case we should transfer the ownership to you to keep it up-to-date.
- At the end, you can directly **save** your entry or show the **Preview**. In the preview, use the blue button top-left to go back and save. You might have to shift the grey command bar from horizontal to vertical orientation to see the blue button "Back to content editing".

© 2022 FocusCoE Project Consortium



Embedding the calendar in your website

For **upcoming short courses**, use the following line of code:

<script id="hpc_events" type="text/javascript" src="https://calendar.learnhpc.eu/script.js" data-website="https://calendar.learnhpc.eu" data-project="CompBioMed" data-country="Ireland" data-filters="level,language,country,sector,format,project"></script>

Options:

- **data-project** : Any value in the list of projects below (or empty to include all of them). **Hint**: Select FocusCoE to include all events of all CoEs (i.e., all projects in the list except EuroCC/CASTIEL and PRACE).
- data-country : Enter a country name to display this country's events only (country name as entered in any of the forms in https://hpc-portal.eu/add_new_content).
- **data-filters** : Include any subset of the values in the script line to have those filters available in your page. **Warning**: This option will be activated soon.

Additional remarks:

- The export of data from the HPC Portal is done automatically **once a day (overnight)**: Newly entered events might be available only on the following day.
- An event feed can be downloaded as .ics file. Warning: This feature is available only when data-country is **not** specified.

The list of projects:

- EuroCC_CASTIEL
- PRACE
- BioExcel-2
- ChEESE
- CoEC
- CompBioMed
- E-CAM
- EoCoE-II
- ESiWACE2
- EXCELLERAT
- HiDALGO
- MaX
- NOMAD
- PerMedCoE
- POP
- RAISE
- TREX
- FocusCoE