ESDW guidelines and programme V

E-CAM Deliverable 5.5
Deliverable Type: Report
Delivered in April, 2020

E-CAM
The European Centre of Excellence for
Software, Training and Consultancy
in Simulation and Modelling

Funded by the European Union under grant agreement 676531

Project and Deliverable Information

Project Title	E-CAM: An e-infrastructure for software, training and discussion in simulation and modelling
Project Ref.	Grant Agreement 676531
Project Website	https://www.e-cam2020.eu
EC Project Officer	Juan Pelegrín
Deliverable ID	D5.5
Deliverable Nature	Report
Dissemination Level	Public
Contractual Date of Delivery	Project Month 51(1st January, 2020)
Actual Date of Delivery	15th April, 2020
Description of Deliverable	Updated guidelines for format, content and coding styles in the ESWD, and programme for year 5.

Document Control Information

Document	Title:	ESDW guidelines and programme V
	ID:	D5.5
	Version:	As of 15th April, 2020
	Status:	Accepted by WP leader
	Available at:	https://www.e-cam2020.eu/deliverables
	Document history:	Internal Project Management Link
Review	Review Status:	Reviewed
Authorship	Written by:	Ana Mendonça(EPFL)
	Contributors:	Alan Ó Cais (Juelich Supercomputing Centre)
	Reviewed by:	Alan Ó Cais (Juelich Supercomputing Centre), Sara Bonella (EPFL)
	Approved by:	Ana Mendonça (EPFL)

Document Keywords

Keywords:	E-CAM, HPC, CECAM, ESDWs, ...

15th April, 2020

***Disclaimer**:This deliverable has been prepared by the responsible Work Package of the Project in accordance with the Consortium Agreement and the Grant Agreement. It solely reflects the opinion of the parties to such agreements on a collective basis in the context of the Project and to the extent foreseen in such agreements.*

[1] ana.mendonca@epfl.ch

Contents

List of Figures

List of Tables

Executive Summary

This deliverables outlines E-CAM's Extended Software Development Workshop (ESDW) programme for 2020/2021. In addition, it provides the most recent guidelines for the organisation of these events, including:

1. the scope of training at ESDW events
2. the structure of ESDWs;
3. the timeline for the organization of an ESDW;
4. the capture of lectures at ESDWs;
5. the role of the E-CAM online training infrastructure;
6. the role of E-CAM programmers at ESDW events;
7. the concept of a module in E-CAM and its acceptance criteria.

The present document is an updated version of deliverable D5.4 [1] submitted in March 2019, on the guidelines for content, structure and output for our ESDWs. These guidelines are intended to be a living document which evolves to reflect experience gained in running the ESDWs and thus they are subject to further revision based on the outcomes of each year's activities, with the present document being the fifth, and last iteration. This updated version of the guidelines, valid from April 2020 to March 2021, will help to ensure that the workshops run consistently across the scientific Work Packages (WPs) and meet the quality standards for E-CAM software.

In addition to refining the guidelines for our ESDWs, this deliverable also analyses the profile of the participants to our ESDWs and the results of our satisfaction surveys. Specifically, we will report on the analysis of

- the participants profile (country of origin, gender, qualification)
- the satisfaction surveys
- training needs highlighted by the participants of our ESDWs.

When appropriate, output from these surveys has been used to refine the reported guidelines.

1 Introduction

Our ESDWs are divided into two categories:

- **Application-specific ESDWs**, in one of the core scientific areas of E-CAM:
  - WP 1: Classical Molecular Dynamics,
  - WP 2: Electronic structure,
  - WP 3: Quantum dynamics,
  - WP 4: Meso- and multi-scale modelling

  These events foster the development of software modules for an application in one of the scientific areas above and must include components that relate to HPC and extreme computing, e.g. by including training of participants on efficiently utilizing it in a variety of hardware configurations, and providing example use cases. Additional HPC training content is provided targeting the specific module developments intended;
- **Transversal ESDWs**, of potential interest to all four scientific areas of E-CAM:

  These workshops are of common interest to the different scientific areas, and address topics such as software best practices, scalable computing frameworks, modern language standards, scalable IO, performance analysis, etc. These workshops are when possible developed and coordinated in collaboration with the PRACE Training Centres (PTC) and other Centres of Excellence (CoE) (where appropriate).

E-CAM delivers 4 ESDWs every year, and the length of each event is adapted to the specific needs of each activity, with significant advance preparation by the organisers and the E-CAM team attached to the event.

The relevant lectures at transversal and application-specific ESDWs are captured and made available through the E-CAM online training Infrastructure. Depending on the licensing option chosen by the lecturer, presentations are made publicly available on the E-CAM website, or remain available to participants via the online infrastructure. Remote participation is available to facilitate industry participation to these events (Section 4.5 discusses this in more detail).

At the end of each workshop the participants and organizers fill in a participants survey. A workshop report is produced by the organizers, which is stored on our website and allows E-CAM to follow up on module development from the ESDW.

E-CAM events are part of the annual CECAM programme, and are hosted at the different CECAM Node locations. Our events also fall under the umbrella of EuroHPC.

The exact composition of an ESDW is being refined year by year, with this present document being the fifth iteration. With this iteration being the last one of this deliverables' series, we also make an analysis of the participants profile and of the satisfaction surveys. The training needs collected by the survey are highlighted in this document, as well as how we intend to address them.

2 Purpose and Structure

An ESDW is structured such that it serves two purposes. Firstly, to be a mechanism for generating software modules[2] for inclusion in the E-CAM Software Library[3]. Secondly, to be an integral part of the E-CAM training program and represent the primary "training by doing" component of the program. E-CAM modules should be written in such a way that they can potentially take advantage of anticipated hardware developments in the near future. Furthermore, we use ESDWs to create a top-down approach for training for next-generation architectures. Attendees come to the ESDW with a particular module to develop, we will advise them on a workflow and expose them to the tools that will allow them to create these modules using programming best practices and with an eye on the future of the hardware where they will run such modules.

E-CAM organizes **application-specific ESDWs**, focused on co-designing and development of software applications. These efforts include components that relate to HPC and extreme computing. Application-focused ESDWs will be typically one-week long events organized together with the application developers. Applications can be part of the codes that are used within the E-CAM community or they may be external to the project. Training at these events will come in the form of lectures on the scientific motivation and scope of the application; how to use it efficiently on varied hardware platforms; and will be followed by a walk-through of the application structure and the workflow for contributing to the project (both given by the application owners). Subsequently, the meeting will focus on implementing/improving aspects of the application, in particular components related to HPC and extreme computing. These developments will contribute to the modules production in E-CAM.

E-CAM also organizes additional complementary ESDWs of interest to all four scientific areas of E-CAM (cross-WP collaboration effort), focused on teaching best practices and application development for next generation machines. These are **Transversal ESDWs**. Potential training topics for transversal ESDWs include, among others, software best practices, scalable computing frameworks, modern language standards, scalable IO and performance analysis. Topics will be selected for their potential impact on the larger E-CAM community but the timing of particular events will be influenced by the specific work agenda of E-CAM. This type of workshop will promote also a strong link between E-CAM and the wider Centre Européen de Calcul Atomique et Moléculaire (CECAM) community, which constitutes a primary user basin in which to embed coding best practices, extreme computing exploitation, and hardware awareness. Full engagement of this very broad scientific user base will allow us to reap the benefits of collaborating with a community that has active means to transfer this knowledge to industrial contacts, multiplying E-CAM's impact.

2.1 Attendees

The typical size of the workshop will be 10-15 trainees plus staff:

- **Trainees**

  The trainees typically consist of a mixture of post-doctoral research assistants or senior PhD students, and young industrial researchers. These researchers will have already trained in simulation and modelling and, for industrial researchers, would be considered part of the companies expertise base in this field. The ESDW will extend their skill range and produce the required training boost to keep them at the leading edge of their discipline. These workshops will also be of potential value to academic staff from European institutions, in the process of changing or extending their current research fields. Trainees should come from a diverse range of backgrounds and a conscious effort will be made to avoid biases, implicit or explicit, in their selection, especially in the area of gender.

  All trainees must commit to the full period of the workshop both in residence and back at their home bases. In the case of industrial participants this will require the written permission of their line-managers and in the case of academics the written permission of their supervisors. Remote participation to a subset of the ESDW in order to facilitate increased industrial participation is possible.

- **Staff**

  WP1-4 senior academics and members of the E-CAM consortium will co-organise all application-specific ESDWs, ensuring that the guidelines are followed and that training material is captured and made available online. They will be supported by the project software manager, an E-CAM programmer and, if envisaged, an application developer (in the case of an application-specific ESDW). Other senior stakeholders and guests will be invited to lecture and mentor during the workshops when required, and in accordance to the workshop program. At least one Postdoctoral Reserach Associate (PDRA) from the relevant Work Package (WP) should attend

[2] A module is any piece of software that could be of use to the E-CAM community and that encapsulates some additional functionality, enhanced performance or improved usability for people performing computational simulations in the domain areas of interest to us. A complete definition at https://www.e-cam2020.eu/what-is-a-module/.

[3] the associated Git repository is available at https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library

the application-focused ESDW, acting as part of a support team for the attendees and supporting the event overall organization. This PDRA should also be responsible for following-up on module production after the event, ensuring that the ESDW meets its software module production targets (which is an important KPI for an ESDW).

ESDWs will normally be available to European applicants but applications from the rest of the world will be considered and accommodations will be made when the non-European applicant can add high value to a workshop. Given the significant time requirement for participation in one of these workshops, early and binding commitments will need to be obtained from perspective trainees. It is important that people appreciate that the workshops are indeed "extended" in that significant amounts of work will be done outside the face-to-face meeting (more below). Conversely, in view of this considerable commitment by participants when back at their home intuitions, organisers should ensure where practical, that modules are such that participants can tune them to their own research priorities on their return to their home bases, so as to allow and facilitate substantial further development work there.

2.2 Logistics

Depending on the specific subject area, the workshops will have a variable total duration. Generally, they will consist of

- Preliminary work by the attendees to ensure minimum common expertise and identify the modules to be developed
- One or two-week residential workshop
- Variable period of work by participants at their home bases, off-line and on-line
- Final wrap-up residential workshop.

Trainees will normally be divided into teams to develop the software, and will need office space, with projector(s) and white-board(s). Code development will be performed on desktop machines at the location of the ESDW or, preferentially, on portable computers belonging to the trainees.

When necessary, remote access to massively parallel and heterogeneous platforms will be ensured, including specialised compilers, profilers and debuggers. **Such access requires significant time to arrange and must be signalled months in advance of the workshop**.

Preparatory material will be sent to the participants in advance of the meeting through our online training infrastructure (see Section 4.5). This will constitute our online training modules. Preparatory material can include a resource library of leading edge texts, reports and software development requirements for each workshop. Specifications for the minimum requirements in terms of operating system and installed software and compilers will also be detailed before each ESDW.

Assistance will be provided, as with all CECAM schools, to the attendees and staff for accommodation during the residential parts of the ESDW.

Participants will be advised of the remote collaboration tools that are used by E-CAM and that can enable the teams to continue to work on their software projects when dispersed back to their home institutions. Host institutions must provide necessary system level support to enable such remote collaboration (e.g. opening designated ports on firewalls).

2.3 Funding

CECAM provides funding for the ESDWs that are part of the official CECAM flagship programme[4]. Two thirds of this funding derives from CECAM Headquarters and one third derives from the local beneficiary hosting the event (normally, a CECAM Node). Funding from external sources and co-sponsors is also possible. Participation to these events is free of charge for academic participants and for industrial participants from one of the current or future industrial partners of E-CAM. An attendance fee of up to 1000 Euro may be charged to industrial trainees from outside the E-CAM partnership. Special provisions for licensing and upload to the repository of software developed by paying trainees will be made if necessary.

Workshops outside of the flagship programme of CECAM are funded through the project.

[4]Events that have been submitted through the normal CECAM Call for proposals and have been evaluated by a scientific advisory committee and independent referees, and approved the the CECAM Council.

3 ESDW Program

The locations, organisers, dates and indicative contents of the E-CAM Extended Software Development Workshops (ESDWs) in the fifth year of the project are described in Table 1. They are advertised on the E-CAM website under the E-CAM event calendar, and on the CECAM website under the CECAM Workshops Program.

Table 1: Extended software development workshops taking place during the fifth year of the project (URLs for each event are embedded in the workshop title).

Number	WP	Title	Dates	Location	Organizers
ESDW18	4	HPC for mesoscale simulation	01.09.20 – 11.09.20	CECAM-UK-Daresbury, Hartree Centre, UK	Jony Castagna, Silvia Chiacchiera, Leon Petit, Michael Seaton
ESDW19	2	Improving bundle libraries	15.02.21 – 26.02.21	CECAM-HQ-EPFL, Switzerland	Emilio Artacho, Volker Blum, Fabiano Corsetti, Micael Oliveira, Nick Papior, Yann Pouillon
ESDW20	1-4	PRACE E-CAM Tutorial in Simulation and Machine Learning	10.03.20 – 13.03.20	CECAM-IRL, UCD, Ireland	Donal Mackernan, Simon Wong, Jony Castagna

The topics for ESDWs are chosen taking into account E-CAM's objectives in terms of development of methods and software scaling towards the high end of HPC systems in preparation for the next generation of exascale machines, as outlined in section 1. This is done under the supervision of the E-CAM Executive Board and the E-CAM Industry and Software Management Groups. In addition to this, the choice for ESDW topics and the software modules to be developed in each area are also selected via:

- the software development needs of the PDRAs associated to the Pilot Projects in Work Packages 1-4;
- direct requests from current industrial partners;
- the outputs of the E-CAM scoping workshop through the respective scientific reports;
- the outputs from the E-CAM state-of-the-art workshops through the respective scientific reports;
- requests collected within the participants of our ESDW events.

This year's program of events is composed of two application-focused ESDWs (for a definition see section 2), namely ESDW18 and ESDW19; and one transversal ESDW, more specifically ESDW20.

The ESDW HPC for mesoscale simulation , will be focused on introducing students to the parallel programming of hybrid CPU-GPU systems applied to mesoscale simulations. The intention is to port mesoscale solvers on GPUs and to expose the E-CAM community to this new programming paradigm. The course will be focused on the GPU version of DL_MESO, based at the Daresbury Laboratory, and a good participation from local researchers is expected. Moreover, due to DL_MESO's large use in industrial research, we expect potential students from industrial companies like Unilever, Syngenta and Infineum to attend.

The goal of the ESDW on Improving bundle libraries is to integrate the ESL bundle, which is a set of utilities commonly used in electronic structure, into electronic structure software packages that can be ported to new computer hardware and/or architecture. This effort is extremely important for the transfer of HPC best practices to industry.

The PRACE E-CAM Tutorial in Simulation and Machine Learning will focus on providing the participants with a concise introduction to key machine and deep learning (ML & DL) concepts, and their practical applications with relevant examples in the domain of molecular dynamics (MD), rare-event sampling and electronic structure calculations (ESC). By the end of the 4-day school, participants are expected to gain an understanding of the fundamental concepts of ML and DL, including how neural networks function, different types of topologies, common pitfalls, etc.; be able to implement basic deep learning workflows using Python; leverage existing framework to discover molecular mechanisms from MD simulations; and utilise the PANNA toolkit developed within E-CAM to create neural network models for atomistic systems and generate results that can be integrated with MD packages.

Three additional transversal ESDWs will take place during the fifth year of the project. More specifically:

- workshop focused on the E-CAM High Throughput Computing (HTC) library, `jobqueue_features` and on its adoption by codes from the E-CAM community. The HTC library is a joint development effort between E-CAM and PRACE.

- workshop focused on the E-CAM load balancing library (ALL) and on its implementation in codes from the E-CAM community
- workshop focused on teaching scientists how to use the well-known software package for molecular dynamics simulations, LAMMPS, on large-scale HPC resources using training material developed within E-CAM. This work is in collaboration with the FocusCoE.

These workshops are intended to work as dissemination-oriented events focused on tools developed in E-CAM and leveraging the expertise that exists in the project. The date for these event is still being defined and their announcement will be made at https://www.e-cam2020.eu/calendar/.

4 ESDW Content

4.1 Scientific Content

Every ESDW that is part of the CECAM flagship programme must follow the normal CECAM procedures and guidelines for submitting a workshop proposal. Therefore, E-CAM is not directly responsible for the scientific evaluation of ESDW proposals but can only shape the scientific content of proposals based on the guideline parameters mentioned in Section 3. ESDW proposals submitted to CECAM must, in addition to these guideline parameters, ensure that they align themselves with the stated goals of CECAM flagship program in each calendar year. For this reason, we do not discuss scientific content here but focus instead on the technical training components present in ESDW events. Coherence between the E-CAM and CECAM requirements is ensured by the E-CAM's Technical Manager, who is also the Director of CECAM.

4.2 Software Development Training

A great part of an ESDW is spent developing software modules for inclusion in the E-CAM repository, with the support of the programmers in place.

In the E-CAM context, a software module is defined as :

> Any piece of software that could be of use to the E-CAM community and that encapsulates some additional functionality, enhanced performance or improved usability for people performing computational simulations in the domain areas of interest to us.

This definition is deliberately broader than the traditional concept of a module as defined in the semantics of most high-level programming languages and is intended to capture inter alia workflow scripts, analysis tools and test suites as well as traditional subroutines and functions. Because such E-CAM modules will form a heterogeneous collection we prefer to talk about the E-CAM software repositories rather than library. The modules do however share with the traditional computer science definition the concept of hiding the internal workings of a module behind simple and well-defined interfaces. It is probable that in many cases the modules will result from the abstraction and refactoring of useful ideas from existing codes rather than being written entirely de novo.

Perhaps more important than exactly what a module is, is how it is written and used. An important function of the ESDWs, where modules are produced, is to bring modern programming standards and techniques into the work practices of the participants. A final E-CAM module will adhere to current best-practice programming style conventions, be well documented and come with either regression or unit tests (and any necessary associated data). ESDWs will contain an explicit introduction session at the start explaining what an E-CAM module is, why good software style is important, and how to use modern software development tools and methods.

The software development projects and prospective modules to be developed at the meeting are defined in advance of the meeting and shared with the ESDW participants, and finalized in the first day of workshop. Trainees will normally be divided into teams to develop the software.

Attendees can also come to the ESDW with a particular module to develop, and we will advise them on a workflow and expose them to the tools that will allow them to create these modules using programming best practices and with an eye on the future of the hardware where they will run such modules.

4.3 Technical Training Content

Each ESDW event should contain training components relating to

- general computing competencies,
- parallel computing competencies and
- training beyond state-of-the-art,

which are discussed in more detail in this section.

We further categorise training components at three different proficiency levels:

- *awareness*, where a presentation is considered the appropriate medium
- *working knowledge*, where a tutorial with hands-on components are appropriate
- *specialist knowledge*, where a dedicated workshop is required.

Given that, for typical ESDW events, technical training forms only part of the overall program, here we only consider training at the *awareness* or *working knowledge* level. If *specialist knowledge* technical training is required, then a

specific ESDW event can be considered in cases where such specialised training is not already provided elsewhere (for example, as part of the PATC program of training events) or where there is particularly high demand within the community for tailored content (the PRACE E-CAM Tutorial in Simulation and Machine Learning described in Section 3 would be an example of such a workshop).

A brief overview of some of the tools used within the expected context of technical training can be found in Deliverable 6.2: E-CAM Software Development Tools [2].

4.3.1 General Computing Competencies

The ESDWs will teach skills to ensure that the E-CAM repository embodies long term coding best practices. Software development methods will be close to the approach successfully used in open-source projects, and a description of the general software standards that we strive for the workshops is contained in Deliverable 6.1 [3]. This deliverable has been migrated to a living document (the E-CAM Scientific Software Best Practices) and includes content relevant to Subsections 4.3.2 and 4.3.3.

It is expected that software will normally be written in C or C++, modern versions of FORTRAN (e.g Fortran 2008) and/or Python. The interoperability of the software in each language will be an important component of the development. All code should contain sufficient documentation. Subroutines and functions should be stored collectively in modules, with the ultimate goal to construct a variety of programs in a common application space from the same building blocks. Each module developed should come with appropriate test cases and including specimen results where necessary.

A more detailed description of the available tools and the recommended workflow is given in Deliverable 6.2 [2]. All tools are hosted on the E-CAM server at CECAM.

Each ESDW should include a session by the Software Manager or the Programmers that will cover the workflow necessary for the participants to submit their modules to the E-CAM repository (as described in Subsection 4.2). This contribution process has been designed in such a way that it also simultaneously touches on many aspects of the recommended development workflow (version control, source code documentation, code review,...).

Examples of the appropriate type of training content for this topic includes:

- version control using Git,
- introduction to Python,
- source code documentation,
- technical skills specific to the ESDW.

If training content in this category is required, then it **must** be provided at the *working knowledge* level. Any necessary material at the *awareness* level should be provided to the participants well in advance. See Section 4.4 for more details.

4.3.2 Parallel Computing Competencies

In addition to coding sessions, the typical workshop should consist of training lectures in parallel computing such as: techniques in parallel software development; lectures on computer hardware and advances in new architecture relevant to the applications being developed; parallel programming techniques (MPI, OpenMP, OpenACC, CUDA,...); and accessing large-scale High Performance Computing (HPC) resources.

Access to HPC resources will be provided where necessary, subject to notification well in advance. See Section 4.4.3 for the E-CAM approach to performance analysis with the applications under development.

Each ESDW event must include training content that covers some aspects of the topic of Parallel Computing, though the material given can be at either the *working knowledge* or *awareness* level as appropriate for the audience and the topic of the ESDW.

4.3.3 Training Beyond State-of-the-Art

WP 7 of E-CAM connects the project to the overall development of HPC hardware/software and the people/projects in Europe that operate in this space. Given that "beyond state-of-the-art" is by definition unavailable hardware, WP 7 will create *awareness* level training content that can be presented at all ESDW events. This content will be updated annually and presented by the programmer assigned to the ESDW and will cover

- European HPC hardware available within a 5 year horizon

- Programming methods, models and tools to leverage this hardware
- Algorithm considerations for efficient scalability.

The evolution of this topic in turn leads to recommending practical content for Section 4.3.2. Examples to date of these have been our workshops in C++ and KOKKOS, where the language and libraries are being designed to give some level of future-proofing to software with respect to emerging architectures.

If additional training content in this space is required it needs to be requested months in advance to ensure that appropriate instructors for training at this level can be found.

4.4 Preparation

The impact of an ESDW to a participant critically depends on an appropriate level of preparation prior to the meeting. For this reason it is essential that the scientific and technical content of the ESDW is confirmed *a minimum of 3 months in advance*. This is to ensure that the organisers, participants and programmers will have adequate time to prepare and that there is sufficient time to request external trainers where necessary.

Trainees will be provided with training material sufficiently in advance of the workshop through our online training platform (see Section 4.5). Any necessary assistance with this material will be provided through the organisers and programming team as appropriate.

4.4.1 Timeline for ESDWs

We have built a timeline for ESDWs, with important steps that the workshop organizers and co-organisers should follow to achieve a successfully training event:

1. Before the workshop
   (a) 3 months before
      - Define workshop scientific content
      - Build up a list of software development projects and prospective modules
      - Evaluate the need for additional technical training, either developed as part of the ESDW itself or organized by another training organization (e.g. Partnership for Advanced Computing in Europe (PRACE)) and leveraged by the ESDW

   (b) 2 months before
      - Discuss software requirements and applications that will be used with the software manager
      - If required, discuss also the need for HPC resources
      - Define the technical training content and according to that create the online training modules for the workshop in the E-CAM training platform; give access to the participants (more in Section 4.5.1)

   (c) 1 month before
      - Define scientific training material to be added to the E-CAM training platform, and (where appropriate) an assessment form to evaluate software development skills before the event
      - Outline the lectures that should be recorded (from both the scientific content and the technical training content)
      - Prepare setup to allow for remote participation (if applicable)
2. At the workshop
   (a) Finalize the module development plan and distribution among the participants within the first three days of the workshop

   (b) At least one PDRA from the relevant E-CAM work package should attend the meeting
      - to provide technical support
      - if requested by the organizer, to be able to follow up on modules development after the meeting

   (c) Record the lectures (audio/video with slides) as defined in advance of the meeting

   (d) Fill in an intermediate workshop report (one page form), with the list of modules to be developed and already in progress and with a provisional date for the 2nd meeting (if envisaged)

3. After the workshop
   (a) Respond to the workshop survey (sent by CECAM)
   (b) Monitor software development (important KPI for the ESDW)
   (c) Participants that develop modules which are certified by E-CAM receive a certification for software development and enter a database that can be shared with industrialists for future corporate positions. This is managed by the leader of *WP 5 - Training*
   (d) Organize the follow-up meeting (if envisaged)
   (e) Produce a short workshop report in the CECAM website

4.4.2 Recommendations on the day to day running of an ESDW

In consideration of the output of surveys from past ESDWs, this set of guidelines has been prepared and will be provided to the organisers to ensure optimal management of the event. Specific points arising from participant's feedback are:

- **Balance of the program:** there is always a risk that organisers eagerness to cover as much material as possible during workshops may be counter-productive, and exhaust participants so much that they cannot absorb additional material. The fact that ESDW's may cover material that is very new to a participant and also require that the participant simultaneously codes new modules is intellectually very demanding. It is therefore essential to give participants time to assimilate what they have learned, by not over-charging the number of hours of work/lectures each day or presenting too may different topics within a short period, and including a long break between the morning and afternoon session (or the equivalent).
- **Logistics:** ideally, participants should arrive the day before the beginning of ESDW's, particularly if travelling to a training site takes very long. If a session is required on the day of the arrival, it should be kept short and not be essential for the rest of the meeting.
- **Preservation of material:** it is also advisable to record, where possible, audio/video with slides, allowing participants to revisit lectures or demonstrations in their own time, both during and after the meeting. Such material also can be used by people who did not have the opportunity to attend the ESDW in person, thereby increasing the potential number of future users, and the impact of the ESDW.
- **Screening of documentation:** where a substantial amount of software already exists, to which the participants are adding modules, it is critically important that such software be fully documented, with a legible glossary of terms and definitions. Organisers should liaise with programmers and the software manager to ensure that appropriate documentation is prepared and communicated to participants well in advance of the meeting.
- **Selection of projects:** modules should be identified such that participants can tune them to their own research priorities on their return to their home bases, so as to allow and facilitate substantial further work there.
- **Coordination and preparatory work for software needs:** it is also highly important that all the software requirements (applications that will be used) are discussed with the software manager at least 2 months in advance, which will help to avoid any technical problems at the event. Furthermore, if HPC resources are required, this needs to be flagged at least 2 months beforehand. Participants should be able to arrive on site and start coding without having to deal with any problems relating to installing the software, getting accounts or having the right permissions. For these reasons, guidelines for software installation and resource access need to be created in collaboration with the software manager.

4.4.3 The role of the *Programmers*

Typically applications within an ESDW will not be developed from scratch and so, prior to the workshop, the programmers role is to gain some familiarity with these applications. The applications to be used must therefore be decided at the time the program is finalised (see section 4.4.1) . Where appropriate, programmers will implement a Performance Analysis workflow for the applications which includes:

- Integrate the application into EasyBuild,
- Scaling analysis with JUBE or ReFrame,
- Performance analysis with Scalasca.

During the workshop, the programmers are there to provide instruction and support in the tools used by the ESDW and assist the participants where necessary with these tools. They can also leverage the performance analysis workflow

that they have prepared to help analyse the performance impact of the work undertaken during the ESDW on the PRACE HPC resources to which E-CAM has access.

4.5 Online training infrastructure

E-CAM wishes to develop an appropriate online training infrastructure over the course of the project. To this end it is establishing strong partnerships with PRACE and leading HPC centres in Europe and the US in order to connect to appropriate and innovative training content that can bring the E-CAM user communities to the exa-scale.

The infrastructure to support these efforts is located at https://training.e-cam2020.eu/ with content collected during the ESDW program of 2017, 2018 and 2019 already stored there. The goals of our training infrastructure are to:

- **collect the content captured at our application-focused and transversal ESDWs**, allowing participants to re-visit lectures or demonstrations in their own time, both during and after the meeting. Such material can also be used by people who did not have the opportunity to attend the ESDW in person (in particular our industrial partners);
- **generate online training modules for each ESDW**, which will be a set of preparatory material shared with the participants and that will allow everyone to acquire the same basic knowledge before the meeting (see Section 4.5.1);
- **be a repository for the data associated to our events** (captured lectures, lecture materials, reading materials, tutorial content and software requirements);
- **build tutorials on programming best practices to develop software for extreme-scale hardware**, that we can propose them to the extended CECAM community that has active means to transfer this knowledge to industrial contacts, multiplying E-CAM's impact;
- **associate with other groups and projects with similar training scope** such as PRACE, other CoEs and MolSSI, to cover for different and broader training material.

This infrastructure is built upon the Clowder application (developed at NCSA), which is a research data management system designed to support any data format and multiple research domains. Our choice for this system is because creating and maintaining, for example, a set of Massive Online Open Courses (MOOCs) is simply too labour intensive and E-CAM does not have the resources for this. The automation possibilities that Clowder presents allow us to drastically reduce the end-user overhead for adding content (or at the very least bring it to a level that E-CAM can afford to sustain).

We strongly recommend that the reader refer to deliverable D6.7: E-CAM Software Platform V[4], for a detailed description of our development efforts for the Clowder platform, and most importantly, how we will manage the information on this platform.

4.5.1 Online training modules

The online training infrastructure will deliver content-specific training modules, focused on tools that can assist the participants of ESDWs, our post-docs and other interested groups to develop software for extreme-scale hardware. Contents of the such preparatory material may include:

- software best practices
- E-CAM modules/workflow, Git and GitLab
- Scalable computing frameworks
- Basic parallel programming (MPI, OpenMP)
- modern language standards
- scalable IO
- performance analysis
- version control
- scientific background material associated to the event.

The specific content of the online modules will be defined after the technical training part of the workshops has been agreed among workshop organizers and software manager, as mentioned in Section 4.4.

Our online training modules will originate from the transversal and application-focused ESDWs, where we will record target lectures and provide supporting material for these lectures. At the event itself we circulate consent forms among the presenters, that allow to label lectures with a given licensing options as follows:

1. Public (CC BY 4.0);
2. Private, but available for the people singing up to the online training portal;
3. Private, but available for the workshop participants and organizers;
4. Private or Public, after a certain embargo period.

E-CAM's training portal allows us to easily handle these licensing requirements, and access to particular content can be restricted based on user roles. The training infrastructure aggregates and filters material and lectures into distinct categories with appropriate access points. Furthermore, we now use tags for different scientific topics, and users can aggregate results by topics. For training material already covered by other community-created online training infrastructures, E-CAM will provide documents and/or URLs (see Section 4.5.2).

When appropriate, an assessment form will be part of the preparatory material, allowing access to information with respect to the participants' skills prior to the event.

The programmers associated with each ESDW will help create and organise the material for the online modules, and upload it on the training platform for the event, under the supervision of the Software Manager.

4.5.2 Other training material available

The E-CAM Software Library GitLab instance is a principle access point for users wishing to interact with E-CAM, including training. There, they are encouraged to access (and contribute to) software described in the E-CAM Software Library, through a structured scheme of quality control and what is effectively a support infrastructure. This is further facilitated through an extensive set of E-CAM services: Redmine, CodiMD, ShareLatex, and in particular GitLab. The provision, use and further development of these services is an integral part of ESDW's, and also one of the important means by which E-CAM will deliver online material.

E-CAM is also making extensive use of community-created online training infrastructure. In the development of it's own training material, E-CAM will take the example of the Software Carpentry Foundation who are leaders in the field of developing collaboratively created, open source training content for teaching researchers computational skillsets. Software Carpentry maintains a set of core lessons that form the basic toolbox of a computational researcher, and from which they can develop more complex skills. Their webpage also contains video presentations of the core lessons, as well as links to all of the source material. In particular, given the the main E-CAM software development service is our GitLab repository, we recommend that all members of the E-CAM community are familiar with the version control system Git. Git is core content within the Software Carpentry syllabus, with the both basic and more advanced material available.

5 ESDW Outputs

The core output of these workshops is a number of software modules, skills acquisition by the trainees and, where relevant, a performance analysis workflow for the applications considered during the event.

At the end of each ESDW, the senior E-CAM scientists leading the workshop will prepare a report on the contents of the workshop and related actions. These reports will be used to monitor and improve the structure and planning of future ESDWs, as anticipated in WP 5 (Training) and in WP 6 (Software Infrastructure).

Furthermore, we also collect attendee's feedback through a survey form that allow us to improve the organization of ESDW events and help us shaping the present guidelines.

5.1 Software Modules

One of the primary outputs of E-CAM are the software modules produced by the postdoctoral researchers of the project and the participants of E-CAM ESDW events. The number of expected modules from each ESDW will be established during the event planning, leading also to the definition of its contents. This will contribute to meet the projected number of modules for each year in each scientific area.

Software modules are contributed to E-CAM through the documentation of the E-CAM Software Library. The sources for the documentation are stored on the E-CAM Software Library Git repository.

Contributions to the repositories are made through *Merge Requests*[5] Each individual modification of the repository automatically causes the associated documentation on ReadTheDocs.org to be rebuilt.

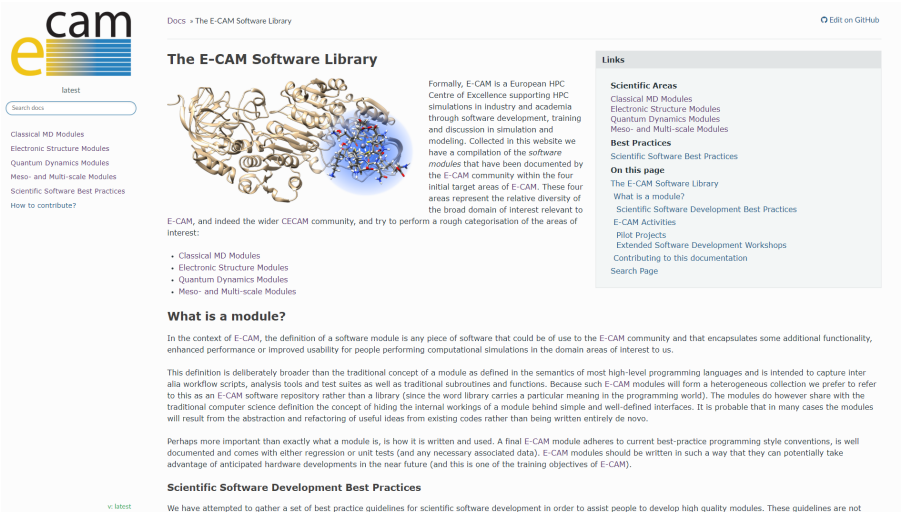

Figure 1: The E-CAM Software Library

An example of the landing page for the E-CAM library and the repository of the Classical MD WP can be seen in Fig. 1 and 2, respectively. The current template format for the submission of modules is available through the documentation pages as well as a description of how the software module is submitted.

5.1.1 Acceptance Criteria

The software manager (together with the programmers) is responsible for the evaluation of the software quality of the modules and their upload to the E-CAM repository under the appropriate software licence conditions.

An acceptance criteria checklist is applied to the submitted modules before accepting the merge request. This list covers aspects including:

- Coding style and naming conventions
- Sufficient source code documentation
- Passing unit and/or regression tests

Typical contribution workflows to the E-CAM repository can be seen in Fig. 3, where we can see that each contribution is reviewed at least once, and in some cases a number of times, by the E-CAM programmers to ensure the contribution meets our acceptance criteria mentioned above.

[5] Merge or pull requests are created in a git management application and ask an assigned person to merge two branches.

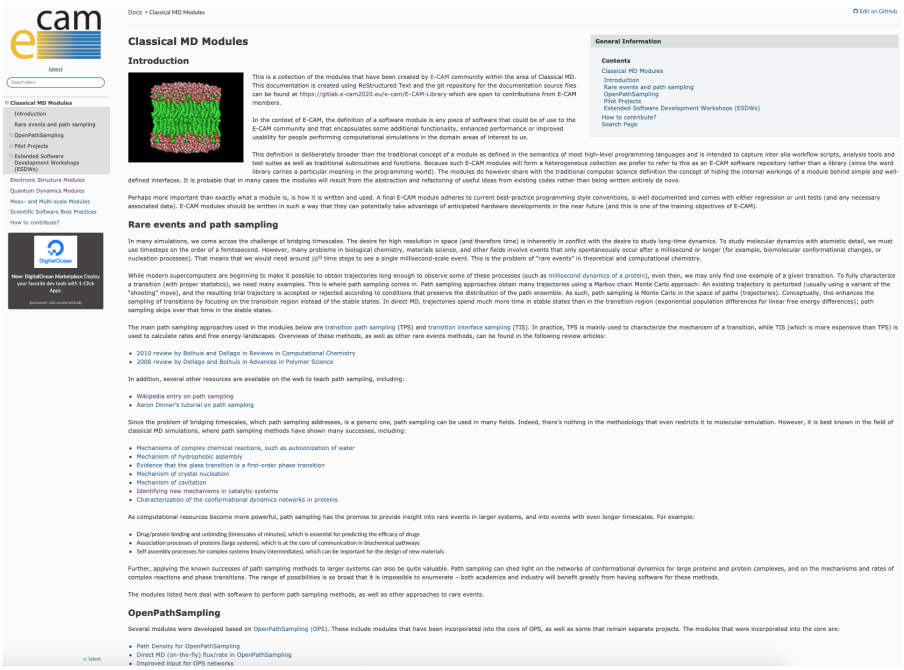

Figure 2: Example of one of the four scientific WPs repository of modules

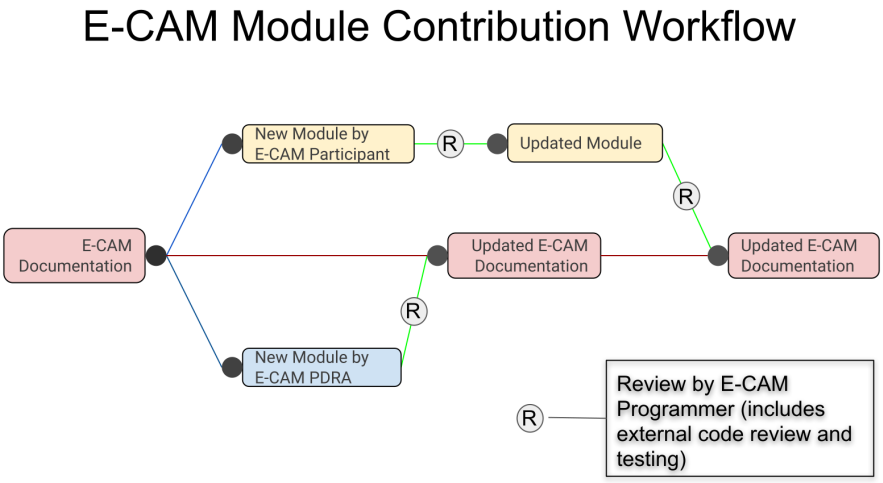

Figure 3: The contribution workflow for the E-CAM Software Library.

5.1.2 Certification

The project has decided on 2 levels of certification stemming from participation at ESDW events. The first is a general certificate of participation in the event.

The second is a certification of the acceptance of a software module generated by the participant into the repositories of E-CAM. This certification requires that the participant will have achieved a *working knowledge* of the approach to software development taken by E-CAM, which is defined above and in more detail in deliverable D6.1:ESDW Technical Software Guidelines [3].

5.2 Key Performance Indicators

The output of the ESDWs is also measured via a set of Key Performance Indicators (KPI) :

- Number of students participating;
- Number of software modules developed and uploaded to the E-CAM repository, meeting the E-CAM quality standards;
- Number of certificates of competence in scientific software development (obtained after module certification).

In addition, we also monitor the number of accounts having access to the E-CAM training platform; the number of

lectures captured and stored there and the number of topics covered. These numbers are analyzed at the end of each reporting period and allow to evaluate the overall project performance, monitor the impact of the CoE, and if necessary take corrective actions.

5.3 Analysis of the ESDW participants profile and of the ESDW satisfaction surveys

5.3.1 Profile of the ESDW participants

We have analysed responses from a total of 260 people participating to 13 of our events taking place between 2016 and 2019. Collected data included

- country of residence,
- gender,
- current position.

The country of residence for the people participating at our ESDWs is highlighted in Table 2. Germany, United Kingdom and France represent the countries where the majority of people come from, followed by different countries spread around Europe. A total of 5.4% of our participants come from countries outside the EU, and among these the US (4.2%).

In respect to gender, we have 16% of female participation to our ESDWs (see Table 3). If we take into account that the participants to our events are also part of the CECAM community, this number is close to the percentage of female participation at CECAM workshops and schools in 2018, which was around 20%.

Table 2: Country of residence for the participants of our ESDWs.

Country	**%**
Germany	20
United Kingdom	16
France	15
Ireland	8.5
Netherlands	6.5
Spain	5.8
Switzerland	5.4
Italy	5.0
Austria	4.6
United States	4.2
Norway	2.7
Denmark	1.9
Slovenia	1.5
Other countries in EU	1.9
Other countries outside EU	1.2

Table 3: Gender balance at our ESDWs.

Gender	**%**
Female	16
Male	84

The majority of people attending our ESDWs are scientists (post-docs, researchers, programmers, research software engineers) or senior scientists (professors, assistant professors, lecturers), followed by PhD students, master students and industrial researchers. See Figure 4 for more details on this distribution.

Based on these outcomes, we have identified a set of recommendations for perspective organisers and referees of ESDWs proposals. Specifically:

1. Increase efforts to attract students from EU13 countries;
2. Involve researchers and HPC centers from a larger number of EU members;
3. Increase efforts to attract female organisers and participants.

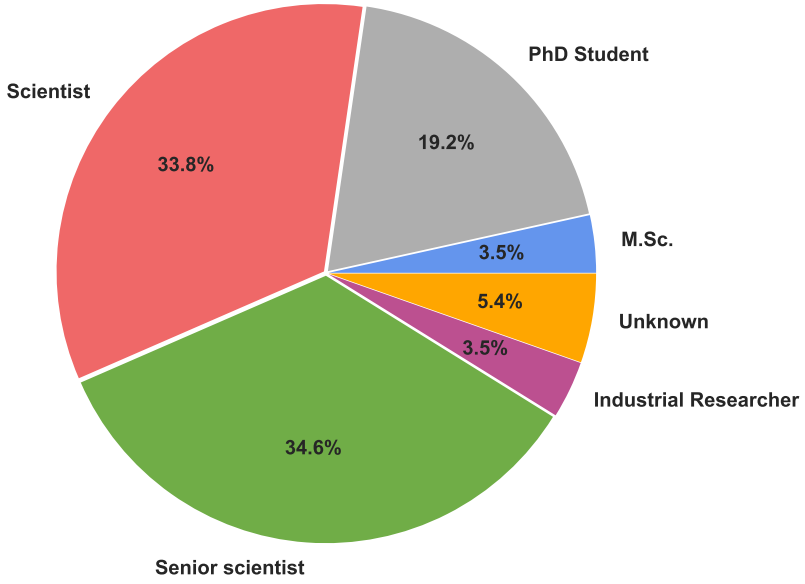

Figure 4: Position occupied by the people attending our ESDWs.

As for point 3, we also plan to include some short interviews/videos of female participants on the E-CAM website to encourage a more gender balanced set of applications.

5.3.2 Participants surveys

E-CAM sends out surveys to the participants of its ESDWs, at the end of each event. The survey is anonymous and helps E-CAM to improve the quality of E-CAM extended software development workshops by reviewing the participants' replies to 15 key questions. There are also spaces for people to make additional comments if they wish.

Figure 5 summarizes the answers from 66 participants to 10 of our ESDWs, to the following questions selected from the survey:

1. Impact of the meeting on your research
2. Was there enough discussion time at the meeting
3. Evaluate the science presented
4. Evaluate the software developed
5. How much did you learned during the meeting
6. Evaluate the quality of the training material provided
7. Did you started the development of one or more software modules to include in the E-CAM repository
8. Was there enough support for the development of your modules.

The participation rate to the survey was about 1/3 of the total number of participants to these workshops. The feedback that we got clearly showed that workshops will trigger interesting new approaches on people's research (with 15% of the people expressing that it will have a major influence on their research). 94% of the people acknowledged the time dedicated to discussions at the meeting, whether discussions were guided by a specific topic, and their relevance and topicality. The majority of participants evaluated the science as being of a really good quality and 17% considered it as being leading edge. The software developed at the meeting was considered better than what they would have expected by 41% of the participants. It was confirmed that the majority of people learned software best practices at the meeting, and that the training material provided was above average to excellent. 74% of the participants to our ESDWs started the development of one or more software modules to include in the E-CAM Software repository, and almost the totality of people judged that there was enough support for the development of software.

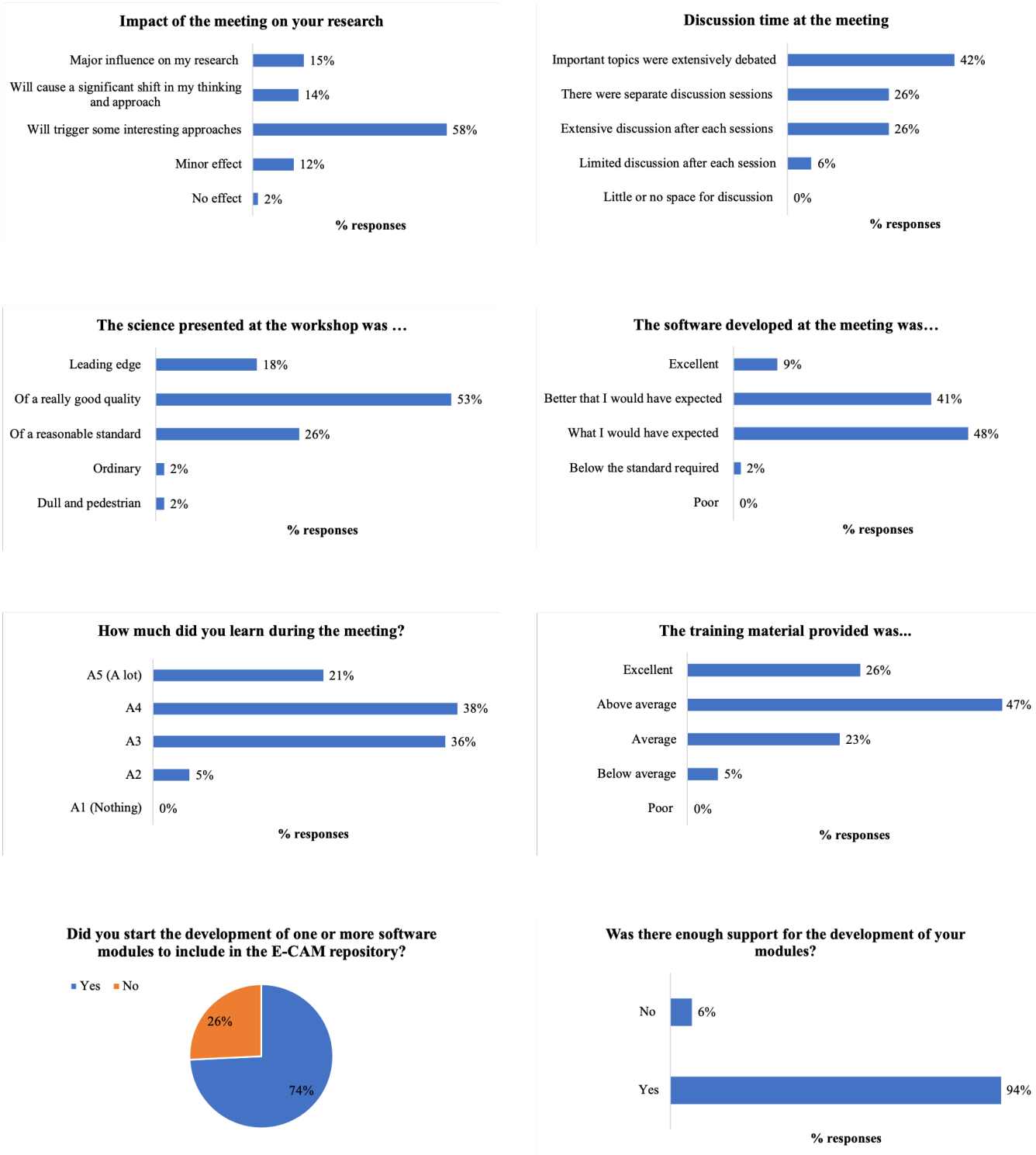

Figure 5: Results of the ESDW participants' surveys.

The collected data indicates that we are reaching the E-CAM targets for these events. Based on the responses, some improvements are under consideration. In particular, analysis of the learning curve during the workshops suggests that the level of satisfaction strongly depends on the entry level of the participant and reflects the difficulty in providing equally interesting training to people with significantly different entry skills. We will consider enhancing activities in small groups of more homogeneous level in future activities.

Since the fall of 2019, surveys also ask participants about their training needs. The following items have been collected so far:

1. Hands-on on code optimization
2. Code debugging

3. Code testing
4. Git, GitHub, GitLab training
5. Writing bindings (python as a syntactic glue)
6. Include discussions on tools and tricks/hints that the software developers can use in their everyday workflows.

The first four training needs are normally covered in all our ESDWs, showing adequacy of training coverage at these meetings. Item 5 has already appeared in one of our ESDWs in 2018 (and captured on our training portal), and we may well revisit this topic in our 2020 program. Item number 6 is a topic that we have introduced into the agenda of some of our ESDWs in 2019, and that has certainly caught interest. We will ensure future meetings include discussions of this kind and study the possibility to organize online sessions on this topic too.

We also take this opportunity to mention our effort to address extreme computing within the wider user/developer community of E-CAM. Many of the scientists connected to E-CAM were identified, via the surveys and one-on-one interactions, as *added-value developers*, that build upon software applications and libraries developed by the community. They may well not be aware of the technical internal details of this foundational software but they overlay their own scientific expertise on top to produce significant scientific added value. The Tuning LAMMPS tutorial being developed by E-CAM in collaboration with FocusCoE is a pilot example of how to expose the extreme computing capabilities of such software to this category of users and developers.

6 Conclusions

The present document is the last iteration of this series of deliverables entitled "ESDW guidelines and programme". It is an updated version of deliverable D5.4 [1], on the guidelines for content, structure and output for our ESDWs, and the programme of ESDWs for 2020/2021.

Besides refining the guidelines (Section 4) and programme (Section 3) of our training workshops, the present document also analysis the profile of our participants and the results of satisfaction surveys launched at the end of each event (Section 5).

Results show that from 260 participants to our events, 95% of them come from Europe. The majority are scientists (post-docs, researchers, programmers, research software engineers) or senior scientists (professors, assistant professors, lecturers), followed by PhD students, master students and industrial researchers. We have 16% of female participation.

The results of our satisfaction surveys are very good, with a majority of people giving positive feedback to the software developed; the science presented; and the training provided. We have collected feedback on software needs, and actions based on most of them are already integrated into our events.

References

Acronyms Used

CECAM Centre Européen de Calcul Atomique et Moléculaire
CoE Centres of Excellence
HPC High Performance Computing
HTC High Throughput Computing
PRACE Partnership for Advanced Computing in Europe
PATC PRACE Advanced Training Centres
ESDW Extended Software Development Workshop
KPI Key Performance Indicators
WP Work Package
PATC PRACE Advanced Training Centre
KPI Key Performance Indicators
PDRA Postdoctoral Reserach Associate
MolSSI Molecular Sciences Software Institute
MOOCs Massive Online Open Courses
PTC PRACE Training Centres

URLs referenced

Page ii

https://www.e-cam2020.eu ... https://www.e-cam2020.eu
https://www.e-cam2020.eu/deliverables ... https://www.e-cam2020.eu/deliverables
Internal Project Management Link ... https://redmine.e-cam2020.eu/issues/19
ana.mendonca@epfl.ch ... mailto:ana.mendonca@epfl.ch
http://creativecommons.org/licenses/by/4.0 ... http://creativecommons.org/licenses/by/4.0

Page iii

E-CAM Software Library ... http://e-cam.readthedocs.io/en/latest/

Page 2

Classical Molecular Dynamics ... https://www.e-cam2020.eu/classical-molecular-dynamics/
Electronic structure ... https://www.e-cam2020.eu/electronic-structure/
Quantum dynamics ... https://www.e-cam2020.eu/quantum-dynamics/
Meso- and multi-scale modelling ... https://www.e-cam2020.eu/meso-and-multi-scale-modelling/
E-CAM online training Infrastructure ... https://training.e-cam2020.eu/
E-CAM website ... www.e-cam2020.eu
online infrastructure ... https://training.e-cam2020.eu/
website ... https://www.e-cam2020.eu/scientific-reports/
CECAM programme ... https://www.cecam.org/program/
CECAM Node ... https://www.cecam.org/nodes
EuroHPC ... https://eurohpc-ju.europa.eu/

Page 3

https://www.e-cam2020.eu/what-is-a-module/ ... https://www.e-cam2020.eu/what-is-a-module/
E-CAM Software Library ... http://e-cam.readthedocs.io/en/latest/
https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library ... https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library
CECAM ... www.cecam.org

Page 4

CECAM flagship programme ... https://www.cecam.org/program/
CECAM Node ... https://www.cecam.org/nodes

Page 5

E-CAM event calendar ... https://www.e-cam2020.eu/calendar/
CECAM Workshops Program ... https://www.cecam.org/program/
HPC for mesoscale simulation ... https://www.cecam.org/workshop-details/8
Improving bundle libraries ... https://www.cecam.org/workshop-details/23
PRACE E-CAM Tutorial in Simulation and Machine Learning ... https://www.cecam.org/workshop-details/1
ESDW18 ... https://www.cecam.org/workshop-details/8
ESDW19 ... https://www.cecam.org/workshop-details/23

ESDW20 ... https://www.cecam.org/workshop-details/1
HPC for mesoscale simulation ... https://www.cecam.org/workshop-details/8
Improving bundle libraries ... https://www.cecam.org/workshop-details/23
PRACE E-CAM Tutorial in Simulation and Machine Learning ... https://www.cecam.org/workshop-details/1
PANNA toolkit ... https://gitlab.com/PANNAdevs/panna
`jobqueue_features` ... https://github.com/E-CAM/jobqueue_features

Page 6

https://www.e-cam2020.eu/calendar/ ... https://www.e-cam2020.eu/calendar/

Page 7

CECAM procedures and guidelines for submitting a workshop proposal ... https://www.cecam.org/submit-flagship-pr

Page 8

Scientific Software Best Practices ... https://e-cam.readthedocs.io/en/latest/best-practices/index.html#best-practices

Page 9

E-CAM training platform ... https://clowder.e-cam2020.eu/
E-CAM training platform ... https://clowder.e-cam2020.eu/

Page 10

CECAM ... www.cecam.org
EasyBuild ... http://easybuild.readthedocs.io/en/latest/
JUBE ... http://www.fz-juelich.de/ias/jsc/EN/Expertise/Support/Software/JUBE/_node.html
ReFrame ... https://github.com/eth-cscs/reframe
Scalasca ... http://www.scalasca.org/

Page 11

https://training.e-cam2020.eu/ ... https://training.e-cam2020.eu/
Clowder ... https://clowder.ncsa.illinois.edu/
NCSA ... http://www.ncsa.illinois.edu/
D6.7: E-CAM Software Platform V ... https://www.e-cam2020.eu/wp-content/uploads/2020/01/D6.7_20122019.pdf

Page 12

E-CAM Software Library ... https://gitlab.e-cam2020.eu
E-CAM Software Library ... https://e-cam.readthedocs.io
E-CAM services ... https://www.e-cam2020.eu/resources/
community-created online training infrastructure ... https://www.e-cam2020.eu/resources/training/
Software Carpentry Foundation ... http://software-carpentry.org/about/

Page 13

E-CAM Software Library ... http://e-cam.readthedocs.io/en/latest/
E-CAM Software Library Git repository ... https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library
ReadTheDocs.org ... https://readthedocs.org/
template format for the submission of modules ... http://e-cam.readthedocs.io/en/latest/example_module/readme.html#example
description of how the software module is submitted ... http://e-cam.readthedocs.io/en/latest/contributing.html

Page 14

E-CAM Software Library ... http://e-cam.readthedocs.io/en/latest/

Page 16

E-CAM Software repository ... https://e-cam.readthedocs.io/en/latest/

Page 18

captured on our training portal ... https://training.e-cam2020.eu/datasets/5b4c85e3e4b0d62a750aa24a
Tuning LAMMPS ... https://github.com/FZJ-JSC/tuning_lammps

Citations

[1] A. Mendoncca, A. O. Cais, and D. Mackernan, "D5.4: Esdw guidelines and programme iv," Mar. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.2586966

[2] A. O. Cais, “D6.2: E-cam software development tools,” Sep. 2017. [Online]. Available: https://doi.org/10.5281/zenodo.1247259

[3] ——, “Esdw technical software guidelines i,” Mar. 2016. [Online]. Available: https://doi.org/10.5281/zenodo.841735

[4] A. O’Cais, “D6.7: E-cam software platform v,” Dec. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3598331