ESDW guidelines and programme III

E-CAM Deliverable 5.3
Deliverable Type: Report
Delivered in February, 2018

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	D5.3
Deliverable Nature	Report
Dissemination Level	Public
Contractual Date of Delivery	Project Month 27(31st December, 2017)
Actual Date of Delivery	8th February, 2018
Description of Deliverable	Updated guidelines for format, content and coding styles in the ESWD, and programme for year 3. Drafted jointly by CH, JUELICH, ICHEC and NUID UCD. Includes updated on-line training modules on use of a structured Wiki-like page, basic parallel programming, scripting tools, use of version control tools.

Document Control Information

Document	Title:	ESDW guidelines and programme III
	ID:	D5.3
	Version:	As of 8th February, 2018
	Status:	Accepted by WP leader
	Available at:	https://www.e-cam2020.eu/deliverables
	Document history:	Internal Project Management Link
Review	Review Status:	Reviewed
	Action Requested:	Submit
Authorship	Written by:	Ana Mendonça(EPFL)
	Contributors:	Alan Ó Cais (Juelich Supercomputing Centre), Donal Mackernan (NUID UCD)
	Reviewed by:	Simon Wong (ICHEC)
	Approved by:	Ana Mendonça (EPFL)

Document Keywords

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

8th February, 2018

***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

Executive Summary **1**

1 Introduction **2**

2 Purpose and Structure **3**
2.1 Attendees . . . 3
2.2 Logistics . . . 4
2.3 Funding . . . 4

3 ESDW Program **6**

4 ESDW Content **7**
4.1 Scientific Content . . . 7
4.2 Software Development Training . . . 7
4.3 Technical Training Content . . . 7
4.3.1 General Computing Competencies . . . 8
4.3.2 Parallel Computing Competencies . . . 8
4.3.3 Training Beyond State-of-the-Art . . . 8
4.4 Preparation . . . 9
4.4.1 Timeline for ESDWs . . . 9
4.4.2 Recommendations on the day to day running of an ESDW . . . 10
4.4.3 The role of the *Programmers* . . . 10
4.5 On-line training infrastructure . . . 11
4.5.1 Online training modules . . . 11
4.5.2 Other training material available . . . 12

5 ESDW Outputs **13**
5.1 Software Modules . . . 13
5.1.1 Acceptance Criteria . . . 13
5.1.2 Certification . . . 14
5.2 Key Performance Indicators . . . 14

References **15**

List of Figures

1 The E-CAM Software Library . . . 13
2 Example of one of the four scientific WPs repository of modules . . . 13

List of Tables

1 Extended software development workshops taking place during the third year of the project (URLs for each event are embedded in the workshop title). . . . 6

Executive Summary

The present document is an updated version of deliverable D5.2 [1] submitted in December 2016, on the guidelines for content, structure and output for our Extended Software Development Workshops (ESDWs). These guidelines are intended to be a living document which evolves to reflect experience gained in running the Extended Software Development Workshop (ESDW)s and thus they are subject to further revision based on the outcomes of each year's activities, with this present document being the third iteration. This updated version of the guidelines, valid from January 2018 to March 2019, will help to ensure that the workshops run consistently across the scientific Work-Packages (WPs) and meet the quality standards for E- CAM software.

The solicitation to incorporate High Performance Computing (HPC) more clearly into our training program has provided us with the opportunity to expand the purpose of our ESDWs, and elaborate on a number of points. Hence, in addition to refining the guidelines of D5.2, this deliverable defines:

1. the revised structure for ESDWs;
2. the timeline for the organization of an ESDW;
3. the E-CAM online training infrastructure;
4. the updates done to the E-CAM software repository.

The program of ESDWs for the third year of the project is also defined within this document, and it already implements the changes mentioned in point 1. above. Points 2. to 4. will also be integrated in this year's events.

1 Introduction

E-CAM delivers four ESDWs every year focused on software development and training in one of its four core scientific areas:

- WP 1: Classical Molecular Dynamics,
- WP 2: Electronic structure,
- WP 3: Quantum dynamics,
- WP 4: Meso- and multi-scale modelling.

In the first two years of the project, our program of ESDWs has served to bring together people from within the project working in the areas above, and also interested participants, to develop software modules and integrate them in their workflows, while at the same time teaching best practices programming in the development of scientific software.

Building on the previous experience, we will strengthen E-CAM's training capabilities with a strategy that enhances its focus on extreme-scale challenges within E-CAM and for our external target groups. We will bring training and best practices in HPC more clearly into the work practices of the participants of our ESDWs and support them on the development of software modules that incorporate these features.

Thus, the format of our ESDWs will be extended as summarized below and further discussed in section 2:

- We will introduce application-specific ESDWs. These events will outline the utility and scientific characteristics of an application, e.g. by including training of participants on efficiently utilizing it in a variety of hardware configurations, and providing example use cases. These ESDWs will foster the development of software modules for the application and must include components that relate to HPC and extreme computing. Additional HPC training content will be provided targeting the specific module developments intended;
- ESDWs will be restructured to include transversal workshops addressing topics of interest for more than one work-package such as software best practices, scalable computing frameworks, modern language standards, scalable IO, performance analysis, etc. These workshops will be developed and coordinated in collaboration with the PRACE Advanced Training Centres (PATCs) and other CoEs (where appropriate).

Furthermore, the length of an ESDW will be 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 will be captured and made available through the online training portal. (Limited) remote participation will be possible to make these events more attractive to industry. Section 4.5 speaks about this in more detail.

The exact composition of an ESDW is being refined year on year, with this present document being the third iteration. It includes elaboration on the following points:

- purpose and structure of an ESDW,
- on-line training platform and on-line training modules,
- ESDW program, content and preparation.

It is hoped that these improvements will help to significantly sharpen our ESDW events, with the particular goal of incorporating HPC more clearly into our work program.

2 Purpose and Structure

An Extended Software Development Workshop (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. To these we will add a third purpose, which is to train on extreme-scale technologies. E-CAM modules should be written in such a way that they can potentially take advantage of anticipated hardware developments in the near future. We will use ESDWs to create a top-down approach for training for next-generation architectures. Attendees will 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. Accordingly, we will expand the scope of our ESDWs, and introduce the following two types of events:

- **Application-specific** ESDWs

  E-CAM will organize events focused on co-designing and development of software applications. These efforts will 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;

- **Transversal** ESDWs

  E-CAM will also organize 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. Potential training topics for transversal ESDWs include, among others, software best practices, scalable computing frameworks, modern language standards, scalable IO, 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 works hop 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. We are currently trialling remote participation to a subset of the ESDW in order to facilitate increased industrial participation.

[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

- **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. He or she 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 the application-focused ESDW, acting as a support team for the attendees and supporting the event organization. This PDRA should also be responsible for following-up on module production after the event, ensuring that the ESDW meets its software production targets (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-boards. 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 newly developed training infrastructure (sec. 4.5). This will constitute our online training modules. Preparatory material will 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

Funding for the ESDW is provided via the CECAM contribution to E-CAM: 2/3 funded from CECAM Headquarters, 1/3 from the local beneficiary hosting the event including possible funding from external sources and co-sponsors.

Participation 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.

3 ESDW Program

The locations, organisers, dates and indicative contents of the workshops in the third 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 third year of the project (URLs for each event are embedded in the workshop title).

Number	WP	Title	Dates	Location	Organizers
ESDW12	1-4	Atomistic, Meso- and Multiscale Methods on HPC Systems	14.05.18 – 25.01.19	CECAM-DE-JUELICH, Germany	Godehard Sutmman, Burkhard Duenweg, Ignacio Pagonabarraga
ESDW11	3	Quantum Dynamics	18 – 29.06.18	CECAM-FR-MOSER, France	Federica Agostini, Basile Curchod, Ari Paavo Seitsonen
ESDW9	1-4	Intelligent high throughput computing for scientific applications	16 – 20.07.18	CECAM-IT-SIMUL, Italy	David Swenson, Alan Ó Cais
ESDW10	2	Scaling Electronic Structure Applications	4 – 17.02.19	CECAM-IRL, Ireland	Emilio Artacho, Volker Blum, Fabiano Corsetti, Micael Oliveira, Nick Papior, Yann Pouillon

The topics for ESDWs should be 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 Executive Board and the Industry Management Group. 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 years program of events is composed of three application-focused ESDWs (for a definition see section 2), namely ESDW10, ESDW11 and ESDW12; and a transversal ESDW, ESDW9. ESDW12 has also a transversal component related to load-balancing in general which will be covered in an initial brainstorming meeting, and aspects that are application specific, which will be covered in a 2nd meeting, with trying to implement these techniques in 3 to 4 codes from WP4. ESDW9 is related to High Throughput Computing (HTC) methods, and there will also be a Python course given by a PRACE instructor. This workshop will have its specific use case in an application from WP1, OpenPathSampling, where we have already had two application-specific ESDWs in the previous two years of the project. The application focused events include applications previously identified as of interest to industrial partners in previous Scoping and State-of-the-Art Workshops.

In addition, we will organize a transversal ESDW in C++ programming language, which will be an introductory course to C++, that will take place at the University of Barcelona, 5-8 March 2018. In 2018, E-CAM has also committed to collaborating with PRACE on some of the training courses in it's PATC courses. In particular, we collaborate with Juelich Supercomputing Centre in 2018 on an upcoming "High-performance scientific computing in C++" and "Python in High Performance Computing" courses.

E-CAM events are part of the annual CECAM flagship program, and are hosted at the different CECAM Nodes locations. As such, E-CAM flagship events must be submitted through the normal CECAM process for validation by the scientific advisory committee and independent referees. This insures a very high standard for our program.

Although specific dates for the ESDW workshop have been decided (see Table 1), it will be possible to choose different ones defined by the beneficiaries responsible for their delivery, in agreement with the Software Management Group and CECAM.

The program for each year is published by the end of the year, for the following year of the project.

4 ESDW Content

4.1 Scientific Content

Funding for ESDW events is provided by CECAM and are the CECAM contribution to the E-CAM project. Every ESDW event 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.

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.

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 [2]. This deliverable has been migrated to a living document (the E-CAM Scientific Software Best Practices) and is beginning to include 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 extensive on-line 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. The 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 workflow (version control, source code documentation, code review,...).

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

- version control using Git (this is a mandatory inclusion for an ESDW),
- 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 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.

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. 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 months 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. From the participant feedback from the 2016 and 2017 program, there have been some very noteworthy points.

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 learn 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).

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.

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.

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. 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.

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 2months 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) . The programmers will implement a Performance Analysis workflow for the applications which includes:

- Integrate the application into EasyBuild,
- Scaling analysis with JUBE,
- 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 On-line 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 Partnership for Advanced Computing in Europe (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 now in place at https://clowder.e-cam2020.eu/ with the initial content collected during the ESDW program of 2017 currently being added. The goals of our training infrastructure are to:

- **collect the content captured at our application-focused and transversal ESDWs**, allowing participants to revisit 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 (specialy 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 in Clowder (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 D6.5:E-CAM Software Platform III available at https://www.e-cam2020.eu/deliverables/, for a detailed description of our development efforts in 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. Access to particular content can be restricted based on user roles. For training material already covered by other community-created online training infrastructures, E-CAM will provide documents and/or URLs (see 4.5.2). The training infrastructure will aggregate and filter material and lectures into distinct categories with appropriate access points.

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 Git repository is the principle access point for users wishing to interact with E-CAM, including training. There, they are encouraged to download and upload software, through a structured scheme of quality control and what is effectively a support infrastructure. This is facilitated through an extensive set of E-CAM services: Redmine, Etherpad, 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 on-line 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 software manager and senior E-CAM scientists leading it 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 in future years 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*[4] Each individual modification of the repository automatically causes the associated documentation on ReadTheDocs.org to be rebuilt.

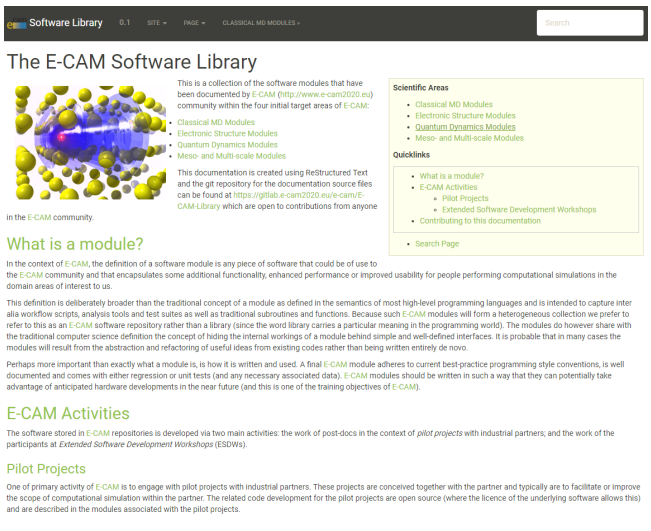

Figure 1: The E-CAM Software Library

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

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.

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

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

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 [2].

5.2 Key Performance Indicators

The output of the ESDWs will also be 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, from this year on we will 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.

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

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/17
ana.mendonca@epfl.ch ... mailto:ana.mendonca@epfl.ch
http://creativecommons.org/licenses/by/4.0 ... http://creativecommons.org/licenses/by/4.0

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/

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 6

E-CAM event calendar ... https://www.e-cam2020.eu/events/
CECAM Workshops Program ... https://www.cecam.org/workshops.html
Atomistic, Meso- and Multiscale Methods on HPC Systems ... https://www.cecam.org/workshop-1591.html
Quantum Dynamics ... https://www.cecam.org/workshop-1641.html
Intelligent high throughput computing for scientific applications ... https://www.cecam.org/workshop-1650.html
Scaling Electronic Structure Applications ... https://www.cecam.org/workshop-1640.html
ESDW10 ... https://www.cecam.org/workshop-1640.html
ESDW11 ... https://www.cecam.org/workshop-1641.html
ESDW12 ... https://www.cecam.org/workshop-1591.html
ESDW9 ... https://www.cecam.org/workshop-1650.html
ESDW12 ... https://www.cecam.org/workshop-1591.html

ESDW9 . . . https://www.cecam.org/workshop-1650.html
CECAM flagship program . . . https://www.cecam.org/workshops.html

Page 7

CECAM procedures and guidelines for submitting a workshop proposal . . . https://www.cecam.org/submitting.html

Page 8

Deliverable 6.2.: E-CAM Software Development Tools . . . https://www.e-cam2020.eu/deliverables/
Scientific Software Best Practices . . . http://scientific-software-best-practices.readthedocs.io/en/latest/
Deliverable 6.2 . . . https://www.e-cam2020.eu/deliverables/

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
Scalasca . . . http://www.scalasca.org/

Page 11

https://clowder.e-cam2020.eu/ . . . https://clowder.e-cam2020.eu/
Clowder . . . https://clowder.ncsa.illinois.edu/
NCSA . . . http://www.ncsa.illinois.edu/
https://www.e-cam2020.eu/deliverables/ . . . https://www.e-cam2020.eu/deliverables/

Page 12

E-CAM Software Library Git repository . . . https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library
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

Citations

[1] A. O. Cais, A. C. Mendonça, and D. Mackernan, "ESDW guidelines and programme II," Nov. 2016. [Online]. Available: https://doi.org/10.5281/zenodo.841731

[2] A. O. Cais, "ESDW Technical Software Guidelines I," Mar. 2016. [Online]. Available: https://doi.org/10.5281/zenodo.841735