H2020-INFRAEDI-2018-2020

Concerted action for the European HPC CoEs

Project Number: 823964

D4.1 – Interim Training Report

WP4 – Training

Workpackage:	4	Interim Training Report	
Due Date	31/05/2020		
Author(s):	Simon Wong		ICHEC (NUIG)
	Alan O'Cais		FZJ
Authorized by	Simon Wong		ICHEC (NUIG)
Reviewer	Guntram Berti		Scapos AG
Reviewer	Renata Gimenez		BSC
Reviewer	Andrea Quintiliani		ENEA
Final Version Date	27/05/2020		
Dissemination Level	Public		

List of abbreviations

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

Executive Summary

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

WP4 of the FocusCoE project was designed to support the training activities of the CoEs and realise transversal benefits to the CoEs. Its activities are based upon a number of objectives and tasks that include establishing a central calendar or registry of CoE training events, examining an overview of the CoE training programmes as a collective, developing sample training materials, providing transversal training courses for CoE personnel and organising a comprehensive European stakeholder workshop on HPC training and skills development needs.

In maintaining an ongoing dialogue between the CoEs and PRACE via teleconferences and face-to-face meetings, WP4 has developed a forum where exchange of ideas and best practices is enabled among some of European's leading providers of HPC training. One of the key early objectives was to establish a central calendar or register that would collate European training events from all the CoEs and other projects like PRACE. This could then be used as a single gateway that allows collective dissemination and querying of European HPC training events regardless of project. In collaboration with PRACE and the CoEs, WP4 has identified and set up the Timely web service as the backend to collate the master, centralised calendar, so that all CoE training events (as well as others) are visible in one place/platform. This not only serves HPC users and course participants by making CoE courses more visible and searchable, it should in time highlight the extent of HPC training activities across Europe. WP4 has worked with the CoEs to establish an ongoing protocol for CoEs to keep the calendar up to date. The calendar will also soon be available for embedding into FocusCoE and other CoE web sites.

Another objective of WP4 was to collect and analyse data on the CoEs' training activities, which has been carried out based on 39 CoE courses from January 2019 to February 2020, before the Covid-19 pandemic restrictions were introduced in Europe. While there is no clear indication of any major scientific or technical "gap" in CoE training activities, this collective analysis has highlighted areas for improvement (e.g. in attracting participants from EU13 countries) and harmonisation (e.g. monitoring and dissemination of CoE courses that may have transversal benefits). While many CoEs had originally planned a mix of in-person training courses supplemented by online webinars, the Covid-19 pandemic has placed obvious pressure on the CoEs to pivot to online teaching. In this regard WP4 have put together some documentation for the CoEs as well as providing an ongoing platform to exchange best practices. Additional support for online teaching is planned (see below).

Another objective in WP4 aims to maximise the impact of the CoEs, by making training available and accessible for the respective user communities in the caveats of running specific applications efficiently at scale. This activity first involved the development of new training materials based around the molecular dynamics code LAMMPS. The material has been produced using the Carpentries approach that is recognised globally for its effectiveness in training many people on practical aspects. Pilot training courses are being planned using the completed material, and the development of a second set of materials on a different HPC application is ongoing.

WP4 also set out to identify and organise transversal training courses that are targeted towards researchers and staff of the CoEs. Plans have been put in place to organise some of these courses in Q1-2 2020 but these have been impacted by the Covid-19 pandemic; a pivot to online delivery

is foreseen for most courses to be organised by FocusCoE. The plan to run a pedagogical course in Q2 2020, for example, is expected to be postponed to Q3 2020 as an online event. The benchmarking workshop is being prepared as an in-person event in Q4 2020, with contingency plans being put in place in case it needs to be converted online. Finally the business sustainability course or workshop, originally planned as an in-person workshop to assist in drafting a white paper for the HPC3 special interest group on the same subject, is being adapted to be an online course or workshop for Q4 2020.

Another significant outcome from WP4 in the first half of the project is the successful organisation of a comprehensive European stakeholder workshop on HPC training and skills development needs in collaboration with the European Commission. The workshop was held in Brussels on 7th October 2019, where a total of 32 participants attended the meeting. Discussions at the workshop were based around several pre-defined themes including the needs of the different user/developer communities, and educational programmes (e.g. MSc) that are useful in generating talent. The event instigated extremely useful discussions on the future of HPC training in Europe. And not only has it produced useful output for the EuroHPC Joint Undertaking, it seeded the formation of a group of HPC training experts and stakeholders who have indicated continued interest in further discussions that may be useful for future strategies and programmes.

In summary, WP4 has made significant progress in realising some of the transversal benefits for CoEs via its activities described above. Looking forward to the second half of the project, an ongoing dialogue will be maintained with the CoEs and PRACE to enable exchanges of best practices. There will be continuation of activities such as ongoing support for CoEs in adopting the central registry, development of additional training materials, and the organisation of transversal courses on pedagogy, application benchmarking and business sustainability.

Finally, the impact of the Covid-19 pandemic has had profound effects on the CoEs' training programmes that will add pressure to deliver online courses. WP4 will maintain a forum where the CoEs may share best practices in online teaching and help identify collaborative opportunities where possible. The latter may be of particular value as online courses can be more labour-intensive (e.g. the need for more moderators compared to an in-person event) and more flexible to incorporate trainers from multiple institutions regardless of geographical location or project affiliation. Indeed, WP4 will seek to allocate resources where practical to provide transversal support to CoEs in delivering training online, e.g. by developing or sourcing new tools and materials.

Table of Contents

1 Introduction

HPC training and education plays an important role in sustaining a steady stream of researchers and developers who can exploit the full potential of High Performance Computing (HPC) and modern computational laboratories. Its role is set to grow in the coming years as Europe, via the EuroHPC Joint Undertaking (JU), looks to develop the next generation of technologies, applications and systems towards exascale, as well as enhancing uptake of HPC from industry and achieving excellence in HPC application delivery and use. The HPC CoEs have a particular important role to play in the European HPC ecosystem by ensuring that extreme scale applications result in tangible benefits for addressing scientific, industrial or societal challenges..

WP4 of the FocusCoE project was designed to support the training activities of the CoEs and realise transversal benefits to the CoEs. The intended activities are split over a number of objectives and tasks that are described in subsequent sections of this deliverable. First and foremost WP4 had set out to support the development of a pan-European HPC training registry that will offer a single point of presence for one to browse and search across HPC and computational science training in Europe, including the training offering from CoEs, PRACE, other projects and national HPC centres (Section 2). Other tasks were conceived to provide an overview and analysis of CoE training activities collectively (Section 3), to develop sample training courses and materials on running domain-specific applications at scale (Section 4), to organise transversal training and services for CoEs (Section 5). Finally, WP4 was tasked to organise a comprehensive European stakeholder workshop on HPC training and skills development needs (Section 6).

2 Development of a pan-European HPC training registry

An increasing number of projects/initiatives (e.g. EC-funded projects such as the PRACE implementation phase projects and the CoEs) and institutions (e.g. HPC centres, universities and research institutes) are involved in the provision of HPC training. While the CoEs play a critical role in HPC training, in particular on scientific domain or application specific subjects, the PRACE implementation phase projects also support a large HPC training programme that are more focused on the technical aspect (i.e. parallel programming concepts that are transversal to all domains). For the first half of the project, WP4 has organised two conference calls and two face-to-face meetings with training representatives from CoEs and PRACE. These meetings initially focused on potential synergies between FocusCoE WP4 activities with that of the CoEs (e.g. transversal courses, see Section 5) and the gathering of information about the respective CoE training programmes, which are examined in Section 3. However, one of the more significant discussions with the CoEs has been the development of a pan-European HPC training registry.

There have been increasing calls for a single gateway (e.g. that is presented as a web page or portal) that allows collective dissemination of such training on a pan-European basis, e.g. where HPC user can browse and search for relevant training courses around his/her location, regardless of which project or institution is responsible for organising the courses being queried.

While there are plans in the PRACE-6IP project to develop the underlying infrastructure to support such a pan-European HPC portal which encompasses other topics beyond training (e.g. access to supercomputing resources, support), a key task (T4.1) of FocusCoE was to liaise with PRACE and the CoEs to advise on suitable tools and mechanisms to develop a registry of training events. Preliminary investigations highlighted the following considerations and requirements for developing a registry:

- The solution must be an open platform/service where any European project or institution would have the ability to add its HPC training events to the registry.
- In order to aggregate HPC training events all around Europe which could then be browsed or queried, each event in the registry must include some key details (e.g. title of the event, dates and duration, location) as well as some metadata (e.g. scientific/technical categories, instructional level, target audience).

In collaboration with PRACE, technical solutions that may support such a registry were examined. This included the possibility of developing a custom tool from scratch, which would involve very significant software development effort, or adopting in-house training event registry tools from other European projects. There were also concerns about the sustainability of effort and expertise to maintain a highly customised in-house solution. Ultimately, a third party web-based tool, Timely [1], was identified as a platform that addressed most of our needs.

Timely is a web-based event calendar service that allows its clients to manage and promote events. Importantly, it enables the following features:

- The platform is naturally an "event aggregator" – once a calendar is set up, it is easy for others to add entries, either through a simple web form or imported via RSS feeds or other calendars (e.g. Google, iCal).
- Calendars can be configured with "filter groups" that allows categorisation of the events based on different attributes, making it easier for users of the calendar to find relevant events via the categories.
- The aggregated calendar or registry can be embedded easily on a web site via a short embed HTML, e.g. a CoE would be able to embed the aggregated calendar on its own web site that will include its training events, as well as others.

Timely was hence trialled as a solution as a collaboration between FocusCoE (that looked at CoE events and needs) and PRACE (who looked into how its many training courses can be incorporated). It was found that events can be successfully imported in bulk automatically that would suit projects/institutions with larger number of HPC events (~100 listings typically per year). With the CoEs, who typically have relatively few events per year, most have chosen to begin with manual additions to the registry initially with assistance from FocusCoE; those who wish to adopt an automatic process in future will also be supported. A publicly-viewable version of this "HPC Training in Europe" master calendar/registry can be found online at https://calendar.time.ly/5r0mgbac – it serves as a useful preview of what can be embedded into web sites that could offer a custom "view" of the calendar, e.g. a CoEs-centric view of their upcoming training activities on the FocusCoE web site [2] and upon planned roll-out of the joint HPC CoEs web site [3]. Integration into the FocusCoE web site is pending upon an investigation of whether the calendar can be used to integrate non-training events also.

After a demonstration of the calendar to all the CoEs, the registry has been populated with training events from the CoEs as well as PRACE training events; see Figure 1 for online training courses from several CoEs (MaX, BioExcel, ESiWACE) and PRACE. The pan-European nature of this aggregate calendar regardless of project or institution, as well as metadata annotation (e.g. scientific and technical domain), provides a powerful platform for HPC users to look for relevant courses. Figure 2 shows an example of a training webinar entry that has been added by the MaX CoE. Finally, it can be seen from Figure 3 that the calendar has been embedded on the HiDALGO web site; the same calendar is similarly embedded on the PRACE Training Portal [4]. Plans have been made to incorporate this training calendar into the FocusCoE web site.

Having successfully launched the registry in partnership with PRACE, FocusCoE will continue to help monitor CoE training events for additions to the registry, and provide assistance if needed.
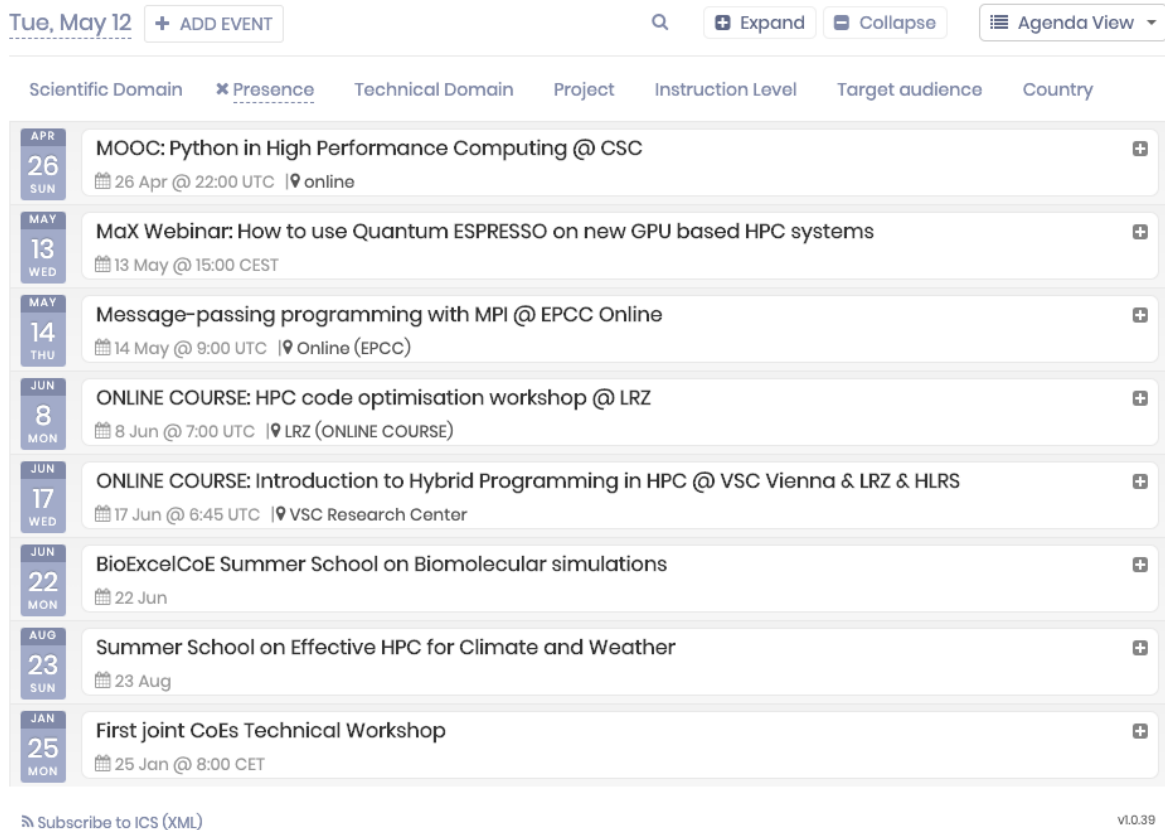
Figure 1. Training registry on time.ly service that incorporates training events from multiple CoEs and PRACE
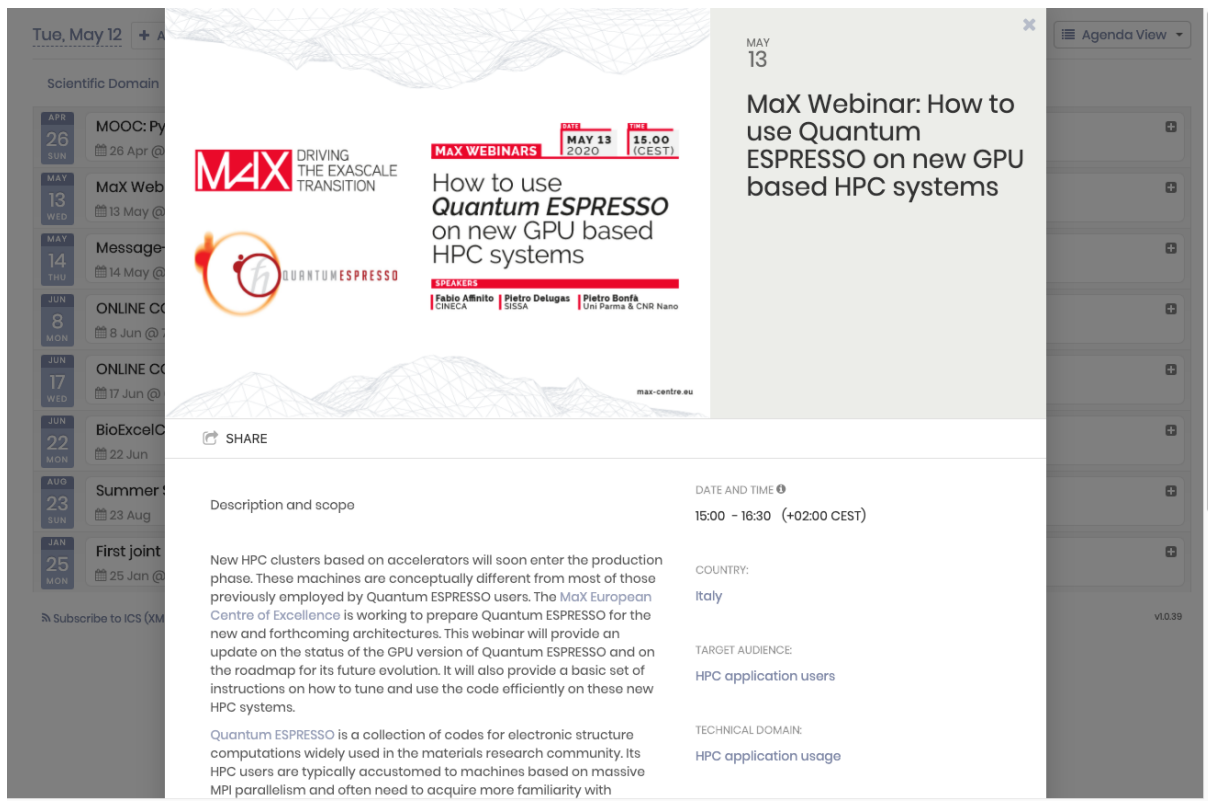

Figure 2. Display of a CoE event - an upcoming MaX CoE webinar

Figure 3. Demonstration of how a CoE (e.g. HiDALGO) may embed the pan-European calendar in its own site via embedding

3   Overview and analysis of CoE training activities

All of the CoEs have training activities that target specific scientific domains and applications. Perhaps one major exception is the POP2 project which is scientifically cross-disciplinary – they provide practical HPC performance optimisation and profiling courses as a service that caters for specific communities, including other CoEs. One of the WP4 tasks was to carry out a high level overview and analysis of these CoE training activities collectively as a whole.

Even with high level data collected by FocusCoE – for example the CoE Service Booklet compiled by T5.1 includes a description of training services provided by CoEs at varying levels of detail – it was clear that all CoEs are involved in providing training services. More detailed CoE training activity data (including similar metadata as those in the training registry discussed above) were collected by T4.2 to carry out this overview and analysis.

Annex 9.1 (Table 1) shows a list of CoE training events from January 2019 to February 2020. This list contains the training courses, workshops, hackathons organised by the CoEs that typically include a mix of talks, practical hands-on instructions and problem-solving exercises; delivered as in-person events or as online courses with remote participation. However, the list does not include an exhaustive list of all the online webinars (i.e. typically 1-2 hour online video lectures followed by questions and answers) that are offered by all the CoEs on numerous topics from scientific theory, software development and HPC applications. In the 14-month period, just prior to the Covid-19 pandemic and the introduction of mobility restrictions in most of Europe, the CoEs have delivered approximately 39 training events (excluding webinars).

Among these, the following observations and analyses can be made:

- Not taking into account online webinars, 38/39 CoE training courses are in-person courses. There was one online training course offered by the BioExcel CoE. However, 8 online courses were made available since March 2020 (see below).
- The locations of the in-person events are primarily in established (western) countries in Europe. With the exception of one event in Slovenia, Eastern Europe and EU13 countries are poorly represented. However, the increasing availability of online courses from the CoEs does provide new opportunities.
- The scientific domains of the training events largely reflect the distribution of the CoEs, i.e. life sciences, chemical sciences and materials, earth system sciences, engineering. A number of events are not domain specific; these are predominantly those provided by POP2 and discussed above.
- The technical domains of these courses are predominantly in the areas of scientific programming, performance engineering and HPC application usage. Apart from these there are small numbers of courses that deal with data science/analysis and numerical libraries and methods.
- There are no obvious major gaps in the coverage of scientific and technical domains given the nature of the CoEs that have been established.
- CoEs have been actively collaborating in delivering some events, e.g. joint events organised by POP and EoCoE, MaX and E-CAM.

The Covid-19 pandemic has caused significant disruptions to training events from March 2020 onwards. Some CoEs have since adapted by offering some of their courses online; there are at least 8 online CoE courses being planned so far, until 2022, compared to just one such event in the last 14 months. It's appreciable that there may be difficulties in transforming what used to be in-person courses to an online offering – the remote setting just doesn't suit the nature of some courses where you have extensive interaction between the instructor and participants, e.g. in scenarios where instructors "shoulder surf" to identify participants who may need help. While this has to be recognised, WP4 has collated some documentation for the CoEs on some of the best practices for online training (see Section 5.4).

With the ability to look at the ensemble of training activities across the CoEs, it has been highlighted that some CoE training events have transversal benefits for other CoEs. For example, the best practice webinar on Git and Jenkins provided by HiDALGO has real relevance to other CoEs. There is scope for WP4 to provide some assistance to identify and disseminate these transversal training opportunities within the network of CoEs.

4 Development of sample training materials

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

Many CoEs are at least aware of the Carpentries project [5], expanded from the original Software Carpentry initiative, which has developed best practices and lesson templates. The Carpentries approach is recognised globally for its effectiveness in training many people on practical aspects of software development and conducting data science. Furthermore, its approach to developing teaching material is designed to be scalable and intended for adoption by many teachers/trainers who can "give back" and subsequently improve the material itself (as it is open source).

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

LAMMPS [6] was the first HPC application that was selected on which the first set of sample training material is based upon. It is a classical molecular dynamics code with a focus on materials modelling, which is also of particular relevance to the E-CAM CoE who are collaborating on the task. Importantly, the application is used on a wide spectrum of hardware, from laptops to the top supercomputers in the world. While the application has the potential to scale, the user often has to determine for oneself whether optimal performance is being achieved. Hence the training material focuses on the following key concepts:

- What is performance and how it is measured.
- How to determine the expected performance.
- How to compare expected vs actual performance (i.e. on a particular HPC system).
- Ways to improve sub-optimal performance, e.g. leveraging accelerators such as GPUs and performance libraries such as Kokkos.

The set of materials on LAMMPS that have been produced (available at [7] online, see Figure 4), which forms the basis of a half-day course on LAMMPS. The intention is that this will be preceded by a half-day course on the HPC system, i.e. a full day induction course for existing LAMMPS users on how to effectively use the application at scale on their target HPC system. It is expected that T4.2 will conduct pilot training courses based on the material that has been produced. Demonstration of the material and training to the CoEs should garner feedback and hopefully stimulate interest in development of materials on other applications. Meanwhile, other candidate HPC applications are being assessed for their suitability in developing an additional set of sample training material in T4.2.

Figure 4. Running LAMMPS on HPC Systems training material

5 Transversal training courses/workshops for CoEs

Task 4.3 of the FocusCoE project was aimed at identifying and organising transversal training courses/opportunities that are targeted towards researchers and staff of the CoEs. The task has identified three topics where transversal courses/workshops would be of value to the CoEs, as described in Subsections 5.1 to 5.3 below. In addition, T4.3 set out to produce documentation for the CoEs on collated best practices, e.g. for online training delivery and capturing live sessions for subsequent online distribution, which is described in Subsection 5.4.

5.1 Pedagogical skills

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

While the original plan was to organise an in-person event in Q2 2020, this has since been tentatively postponed to Q3/Q4 2020 with an online offering.

5.2 Application benchmarking

Benchmarking is a transversal activity that is regularly performed by all the CoEs, hence a 2-day workshop on benchmarking and performance analyses (leveraging European tools where possible) had been scheduled to take place in Q4 2020. Our current plans is that the event will

take place in France, using the new Jean Zay HPC system that will facilitate benchmarking of both CPU and GPU codes. The workshop will encompass key performance tools such as JUBE, CUBE, Score-P, Scalasca, ARM-Forge, PGPROF / NVPROF and Cuda-GDB. Experts from ARM and NVIDIA are expected to provide instructions at some of the sessions.

5.3 Business development and sustainability

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

WP4 originally planned to organise a workshop in Q1/Q2 2020 for this special working group to discuss the key issues, with potential expert advice presented by third parties, to help the realisation of the position paper. However, this has been curtailed due to the Covid-19 pandemic.

The alternative plan is to organise an online workshop that examines key business development issues that will be offered to the CoEs. This is being explored for delivery in Q4 2020.

5.4 Best practices documentation

From the outset, WP4 recognised the role of distance learning is important for many CoEs to reach a wide audience and to cater for those who may find it difficult to participate in in-person training events. One objective of WP4 had been to develop documentation and materials for the CoEs on ways to capture training sessions (e.g. making recordings from a live training event) that may then be used for live streaming and/or subsequent online distribution (e.g. to be made available on YouTube or other channels). The importance of this has arguably been magnified by the Covid-19 pandemic where there is additional pressure to offer courses online with remote participants. Documentation on best practices in capturing training sessions for live or subsequent streaming (Annex 9.2) was produced by WP4, which had since been disseminated to all the CoEs. The expectation is that this will become a living document where it can be updated to include improvements in technology and methods.

6 Pan-European training stakeholder workshop

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

The workshop was held in Brussels on 7th October 2019, where a total of 32 participants attended the workshop, representing the interests of a range of stakeholders and communities [8]. Discussions at the workshop encompassed a variety of themes that highlighted the different needs and challenges of different communities such as HPC users, developers, system administrators. These also include discussions surrounding support for career development such as the scope for developing new MSc programmes, or even shorter diploma/certificate programmes, that may incorporate traineeship opportunities.

In conclusion, the workshop proved to be extremely useful not only to reflect on the current landscape and directions for HPC training, but for the participants themselves who had gained more awareness of national and European initiatives. There is a consensus that the workshop could be the first iteration of future meetings to facilitate communication between the many projects and institutions that conduct training and pave the way to maintain a level of pan-European coordination.

A report that describes the implementation and outcome from the discussions arising from the workshop was delivered as a separate deliverable D4.3 in November 2019 [9]. Major findings from the report were also presented by the FocusCoE project leader at the EuroHPC Governing Board meeting in Luxembourg, February 2020.

Due to interest from the training stakeholder workshop participants, preliminary preparations were being made to organise a follow-up workshop at the EuroHPC Summit Week 2020 in Porto, March 2020. While the this had to be cancelled due to the Covid-19 pandemic, there remains considerable interest from past workshop participants to convene at a future meeting to examine some of the topics that may have been overlooked previously due to lack of time, e.g. the training needs of R&D staff at technology providers and innovators.

7 Conclusions and outlook

Since the start of the FocusCoE project, WP4 has made significant inroads in engaging with all the CoEs on issues related to HPC training. The over-arching principle of WP4 has always been on finding ways to support CoE training via its activities. A summary of its achievements so far is as follows:

- Perhaps the most significant outcome is the establishment of the European HPC Training calendar (i.e. the training registry), along with an established process that makes visible all CoE training events (as well as others) in one place/platform. This not only serves HPC users and course attendees by making CoE courses more visible and searchable, it should in time highlight the extent of HPC training activities across Europe which should enable further analyses.
- A collective overview of CoE training activities was implemented and analysed. While there is no evidence of any major scientific or technical "gap" in CoE training activities, this collective analyses has highlighted areas for improvement (e.g. addressing EU13 countries) and harmonisation (e.g. monitoring and dissemination of CoE courses that may have transversal benefits).
- The production of a full sample set of training materials on using LAMMPS on HPC and extreme-scale systems. While the materials are of tremendous value by themselves, they will also serve as a template for other HPC applications at other CoEs.
- The successful organisation and reporting of the pan-European HPC training stakeholder workshop instigated extremely useful discussions on the future of HPC training in Europe. Not only has it produced useful output for the EuroHPC Joint Undertaking, it seeded the formation of a group of HPC training experts and stakeholders who have indicated continued interest in further discussions that may be useful for future strategies and programmes.

The second half of the project will continue some of the ongoing activities that have already begun. This includes maintenance and support for CoEs to integrate events into the central calendar/registry, and how the calendar can be deployed on local, CoE websites. Having an overview of CoE training activities and an ongoing relationship with the key personnel also raised the possibility of monitoring and dissemination of transversal events that may be of

interested between different CoEs, e.g. these may be published in regular CoE dissemination materials such as an electronic newsletter.

The completion of the LAMMPS training materials has set the template to develop an additional set of training material based on another HPC application. While some candidate applications are being examined, e.g. the HemeLB application that simulates blood flow [10], further discussions will be held with the CoEs to determine the application to be eventually used. This will also depend somewhat on the availability of expertise in WP4.

While there has been some setbacks in terms of organising transversal courses and workshops for the CoEs (partly due to impact from the Covid-19 pandemic), there will be a renewed focus on delivering those events in an online setting where feasible. As highlighted above, concrete plans have been made to organise the pedagogical course as an online offering in Q3 2020. The application benchmarking course in Q4 2020, while conceived as an in-person event, will be examined for potential conversion into an online course where possible. The business sustainability workshop will require some adjustments to change its focus and mode of delivery, it is scheduled for Q4 2020 but it may need to be slightly flexible to adapt to further comments from the CoEs and to avoid a clash with other events.

Undoubtedly, the Covid-19 pandemic has had a significant impact on the training programmes of many CoEs, and indeed of FocusCoE WP4, who have since strived to provide the same or similar courses and workshops online. As discussed in Section 3, this "conversion" to online teaching does not always work or makes sense. While having collected and disseminated some materials to assist the CoEs in this "conversion" process, WP4 will maintain a forum where the CoEs may share best practices in online teaching and help identify collaborative opportunities where possible. The latter may be of particular value as online courses can be more labour-intensive (e.g. the need for more moderators compared to an in-person event) and more flexible to incorporate trainers from multiple institutions regardless of geographical location or project affiliation. Indeed, WP4 will seek to allocate resources where practical to provide transversal support to CoEs in delivering training online, e.g. by developing or sourcing new tools and materials.

8 Bibliography

[1] [Online]. Available: https://time.ly/.

[2] [Online]. Available: https://www.focus-coe.eu/.

[3] "HPC CoE Website to be launched," [Online]. Available: https://hpccoe.eu/.

[4] [Online]. Available: https://training.prace-ri.eu/.

[5] [Online]. Available: https://carpentries.org/.

[6] "LAMMPS Molecular Dynamics Simulator," [Online]. Available: https://lammps.sandia.gov/.

[7] [Online]. Available: https://fzj-jsc.github.io/tuning_lammps/.

[8] [Online]. Available: https://www.focus-coe.eu/index.php/2019/11/19/european-hpc-training-stakeholder-workshop-in-brussels/.

[9] "FocusCoE Deliverable D4.3: Report on the outcomes of the comprehensive European stakeholder workshop on HPC training and skills development needs".

[10] [Online]. Available: http://www.hemelb.org/.

9 Annex

9.1 CoE training events

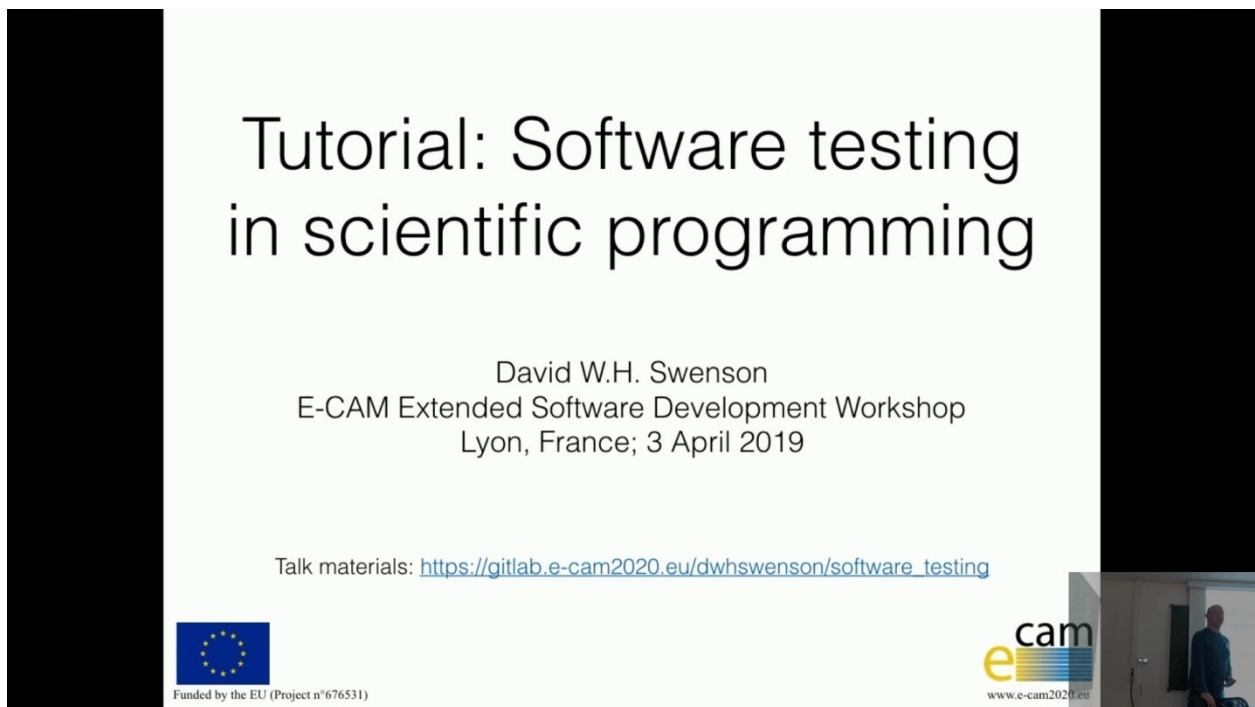| CoE(s)	Event	Date	Duration	Event Type	Venue	Country	Target Audience	Scientific Domain
E-CAM	Extended Software Development Workshop:	07/01/2019	Week(s)	Workshop	In-person	Ireland	HPC application developers	Chemical sciences and materials
POP	30th VI-HPS Tuning Workshop	21/01/2019	Day(s)	Workshop	In-person	Spain	HPC application developers	Not scientific domain specific
CompBioMed	CompBioMed Winter School 2019	13/02/2019	Day(s)	Training	In-person	Spain	HPC application users	Biochemistry, bioinformatics and life sciences
MaX	AiiDA plugin migration workshop	25/03/2019	Day(s)	Hackathon	In-person	Switzerland	HPC application developers	Chemical sciences and materials
E-CAM	ESDW: Topics in Classical MD	03/04/2019	Week(s)	Workshop	In-person	France	HPC application developers	Chemical sciences and materials
POP	32nd VI-HPS Tuning Workshop	24/04/2019	Day(s)	Workshop	In-person	UK	HPC application developers	Not scientific domain specific
MaX	Tutorial on writing reproducible workflows for	21/05/2019	Day(s)	Training	In-person	Switzerland	Academia	Chemical sciences and materials
E-CAM	Extended Software Development Workshop:	03/06/2019	Day(s)	Workshop	In-person	Germany	HPC application developers	Chemical sciences and materials
BioExcel	BioExcel/PRACE Seasonal School 2019, Sw	10/06/2019	Day(s)	Training	In-person	Sweden	HPC application users	Biochemistry, bioinformatics and life sciences
E-CAM	Electrochemical energy storage: Theory meet	12/06/2019	Day(s)	Workshop	In-person	France	Academia	Chemical sciences and materials
POP	ISC-HPC'19 turorial on Performance Enginee	16/06/2019	Hour(s)	Training	In-person	Germany	HPC application developers	Not scientific domain specific
E-CAM	High Throughput Computing Hackathon	18/06/2019	Day(s)	Hackathon	In-person	Italy	HPC application developers	Chemical sciences and materials
POP	33rd VI-HPS Tuning Workshop	24/06/2019	Day(s)	Workshop	In-person	Germany	HPC application developers	Not scientific domain specific
MaX	One day Blender course	27/06/2019	Day(s)	Training	In-person	Italy	Academia	Chemical sciences and materials
BioExcel	BioExcel Summer School on Biomolecular Sir	30/06/2019	Day(s)	Training	In-person	Italy	HPC application users	Biochemistry, bioinformatics and life sciences
E-CAM	Recent developments in Quantum Dynamics,	01/07/2019	Day(s)	Workshop	In-person	France	Academia	Chemical sciences and materials
E-CAM	ESDW in Quantum dynamics	08/07/2019	Week(s)	Workshop	In-person	UK	HPC application developers	Chemical sciences and materials
HiDALGO	First Flee Workshop	15/07/2019	Week(s)	Training	In-person	Ethiopia	Academia	Engineering
MaX	Picking flowers: Hands-on FLEUR	09/09/2019	Day(s)	Training	In-person	Germany	Academia	Chemical sciences and materials
MaX	Summer School on Advanced Materials and I	16/09/2019	Day(s)	Training	In-person	Slovenia	Academia	Chemical sciences and materials
POP & EoCo	ANF Evaluation de Performance des Applicat	16/09/2019	Day(s)	Workshop	In-person	France	HPC application developers	Not scientific domain specific
EoCoE	Terrestrial modelling and high-performance sc	23/09/2019	Day(s)	Training	In-person	Germany	HPC application users	Earth system sciences
POP & EoCo	Performance Evaluation Workshop	07/10/2019	Day(s)	Workshop	In-person	Germany	HPC application developers	Not scientific domain specific
BioExcel	Advanced GROMACS, HADDOCK + PMX Wc	09/10/2019	Day(s)	Training	In-person	Finland	HPC application users	Biochemistry, bioinformatics and life sciences
E-CAM	ESDW: Mesoscopic simulation models and Hi	15/10/2019	Day(s)	Workshop	In-person	Finland	HPC application developers	Chemical sciences and materials
BioExcel	Hands-on Introduction to HPC for Life Scienti	30/10/2019	Day(s)	Training	In-person	UK	HPC application users	Biochemistry, bioinformatics and life sciences
E-CAM	Inverse Molecular Design & Inference: buildin	01/11/2019	Week(s)	Workshop	In-person	Ireland	HPC application developers	Chemical sciences and materials
POP	2nd MareNostrum Hackathon	04/11/2019	Day(s)	Hackathon	In-person	Spain	HPC application developers	Not scientific domain specific
MaX	Hackathon: Development of common domain	25/11/2019	Day(s)	Hackathon	In-person	Italy	HPC application developers	Chemical sciences and materials
ChEESE	HPC and natural hazards modelling tsunamis	02/12/2019	Day(s)	Training	In-person	Spain	Academia	Earth system sciences
ESiWACE	Container Hackathon for Modellers	03/12/2019	Day(s)	Hackathon	In-person	Switzerland	HPC application developers	Earth system sciences
EXCELLERA	Spectral Methods with NEK5000	06/12/2019	Day(s)	Training	In-person	Italy	Data scientists	Mathematical and computer sciences
BioExcel	Computational biomolecular simulation workflo	09/12/2019	Day(s)	Training	Online		HPC application users	Biochemistry, bioinformatics and life sciences
E-CAM	State-of-the art workshop: Challenges in Mult	09/12/2019	Day(s)	Workshop	In-person	Italy	Academia	Chemical sciences and materials
HiDALGO	HiPEAC: HPC and Big Data Technologies for	21/01/2020	Hour(s)	Workshop	In-person	Italy	Data scientists	Engineering
MaX	Computational School on Electronic Excitatio	27/01/2020	Day(s)	Training	In-person	Italy	Academia	Chemical sciences and materials
CompBioMed	Short course on HPC-based Computational E	11/02/2020	Day(s)	Training	In-person	Spain	Academia	Biochemistry, bioinformatics and life sciences
MaX	Hackathon: AiiDA plugins and workflows	17/02/2020	Day(s)	Hackathon	In-person	Italy	HPC application developers	Chemical sciences and materials
MaX & E-CAM	ESL Workshop 2020	17/02/2020	Week(s)	Training	In-person	Switzerland	Academia	Chemical sciences and materials

Table 1. CoE training events from January 2019 to February 2020. These excludes the numerous online webinars provided by the CoEs.

9.2 Best practice documentation on training session capture

The following sections and pages represent a snapshot of the best practice documentation produced by WP4 for the CoEs as part of T4.3. It is written as a self-help guide for the reader to gain an understanding of the practical set-ups that are required from the hardware, software, configurations and processes. In some of the topics, e.g. "What equipment do I need?", the questions are rhetorical, i.e. the "I" refers to the reader or someone.

Capturing Training Events

If you are using a tool like Zoom to hold a training event then using it's **Record** button is maybe all you need. However, if you're doing something else; want more control over the quality, look and feel of the end result; or are eager to stream your training event on YouTube then you might be interested in this article.

Audio

Audio quality is absolutely of the highest importance which is why it appears here first. You can use specialist equipment but you can also achieve very good results with minimal equipment if you are mindful of a few principles:

- Your main priority is to the speaker/lecturer, the quality of their audio will dictate whether people will watch your content or not. If the audience audio is poor when people

ask questions, people will tolerate that because it is not frequent and is usually very short in duration.

- Some rooms have lots of mics so the entire room audio can be enhanced. This is great for the people in the room (where this just adds a boost to the sound coming directly from the speaker) but leads to a lot of unnecessary background noise in recorded audio. If you are recording audio with a plan to broadcast/release the result, you should be careful to only enable the microphones that are required at any point in time. This will save you a lot of effort in removing noise from the end result.
- Different presenters and different mic placements will give you different volume levels. You can correct for these afterwards but this is extra effort that can be avoided if you pay attention to the levels when a speaker starts and try to make them uniform.
- Microphone placement is important. Don't allow a mic to be rubbing against clothing. Don't place a mic too close to a speakers nose or mouth as you will hear them breathing too much. Mics in general, even cheap ones, are pretty good at picking up sound as long as they are pointed in roughly the right direction (and good mics are not even strict on this!), the only thing that changes as you move them further away or change the angle is the volume level (which you can fix yourself anyway). The general principle is not too close, not rubbing, roughly pointing towards the speakers mouth area.

What equipment do I need?

Maybe none! If the event you are capturing is a webinar, you don't really have any control. What you need to watch out for in that particular scenario is that if the speaker uses a headset which includes a mic, by default their laptop will most likely use this mic. If they are in a loud room then this is great since the quality will be better. However, if they are in a quiet room or the mic is on a set of earbuds, you will probably get better quality audio using the mics on the speakers laptops (which usually do some automated noise cancellation).

If you are recording an in-person event (or a hybrid in-person/online event) then you will need a mic that the speaker can move around with. A wireless *lavelier microphone* is what you should be googling for. You can spend quite a bit of money on this but you don't really need to, if you pay attention to the reviews of cheaper equipment you can probably find something adequate for less than €50 (and cheaper kit can be more portable, which is a bonus). Another option is a simple bluetooth microphone. While cheaper again, the sound quality of these is not as good in general though.

I use the AKG WMS 40 which is about €75 for the single mic version, but you will need to purchase a microphone (with a mini-XLR connector) separately (you can get something decent for about €35).

For the audience, you can use dedicated hand-held mics or simply, for example, the audio from a webcam. With webcam audio the sound won't be great, but it can be acceptable if you clean up the signal a bit and it saves you from having a complicated setup.

If you do capture the speaker and the audience separately, you are now going to need to do some audio mixing. Sometimes it is useful to have a special piece of equipment to help with this since it will allow you to tweak everything with actual knobs rather than through software. I have been very happy with the Behringer Xenyx 302USB which is about €60 (but there are probably

cheaper and more modern options available). It also acts as an external sound card when connected via USB to your laptop, which makes it trivial to integrate sound into your software package.

What software do I need?

I'm not going to discuss paid options here, there is some really great open source software out there. For cleaning and editing audio there is Audacity but you, like me, are probably short on time and would prefer to avoid having to do this at all.

In that case I would just suggest you use the features of OBS Studio to add filters to your audio inputs there to clean your signal as it is being captured. The ones you are most likely to need are:

- Gain - helps to set the volume level for quiet equipment (if you use a mixer you shouldn't need this)
- Noise gate - Kills a microphone when the noise falls below a certain level. Useful if you only want sound from a mic when it is clear someone is talking.
- Noise suppression - Makes an attempt to erase background noise from the captured audio. Don't be too aggressive with this as it can result in robotic sound.

A useful walkthrough of how to use these is available here, and a bit of experimentation will go a long way.

OBS Studio allows you to add multiple audio sources so it is a very useful tool to control and integrate audio sources on your final video.

Screen

In the types of events that I capture, there is usually a programming element where the presenter will be typing in a shell, writing code, showing a visualisation or using a Jupyter notebook. In this case, the screen of the presenter is of extreme value in the final video…and can't be explained adequately by distributing a PDF of the slides afterwards.

What equipment do I need?

If you are just doing a webinar, you will already have the screen when you capture the webinar. However, quality is a major issue in cases where someone is doing, for example, live coding. The person who looks at the video afterwards needs to be able to easily read text that could be in quite a small font. For this reason you want to ensure that you are capturing the screen in an adequate resolution and without compression software blurring the text.

Thankfully because gamers like to stream their gameplay online, there is an abundance of cheap solutions to do this with extremely high quality. I use the AverMedia Live Gamer Portable but, while capable, that is quite old now and there are smaller, more advanced solutions like the Live Gamer Mini available in a similar price bracket. Most options are probably good, but choose something that is well reviewed and that will work with your software solution (which I hope will be OBS Studio).

Note that the equipment (and software solution) you choose needs to be supported by the operating system you intend to use. Most gamers use Windows so must capture cards focus on this OS. AverMedia cards work on Macs so they are a good choice there **but I have not tested this**. While Linux is supported by the software solution I propose, it will be probably be hard to find a hardware capture card that has Linux support.

With this type of equipment we are dealing pretty much exclusively with HDMI. Converters to HDMI for other ports (VGA, DisplayPort,…) are easy to come by. A converter from HDMI to VGA can also be useful in cases where the projector in the room only has this option available (an increasingly rare scenario), you would use this on the output signal from your capture device.

Finally, you will need a long HDMI cable (2m to 3m) to give yourself some positioning freedom when you insert yourself in the connection between the projector and the speaker.

What software do I need?

Each device comes with it's own software package but I would advise you to skip that and just focus on becoming familiar with OBS Studio. As seen above, it can already juggle our audio inputs and, because it is targeted at gamers, it supports a lot of the video capture equipment that you will come across.

Integrating the hardware into OBS Studio is trivial, if it is supported by OBS it will automatically appear in a list of possibilities when adding a video capture device.

Video of the speaker

To be honest, you could easily leave out camera recording of the lecture and you would lose very little. Where video becomes useful is in giving a viewer an idea of where someone is pointing or to have an idea of what is going on in the room (e.g., if you can see that the speaker is not on stage, then having no sound is probably not a glitch in the video).

Also, in this article we've given precedence to the capture of the presenters screen. You're probably not going to put anything higher than HD quality online so, if you do do a camera recording, the resolution of a video recording is really unimportant since it is likely to just take up a corner of the main view in the final video.

One important plus of using a webcam is that decent quality ones also usually have good built-in microphones. A webcam can probably help capture a question from (or interaction with) the audience without requiring an additional microphone for audience members.

What equipment do I need?

Of course you could do something very complicated here but you want this to easily integrate into your software solution. For that reason, I would keep it cheap and simple: an HD webcam that can be mounted on a tripod, such as the Logitech C920. You don't really need a zoom lens or anything since you can do usually do basic cropping in your software solution.

What software do I need?

If you use a webcam then none. I hope at this point you have decided on OBS, if so then adding video is simply selecting an additional video capture device.

Even if you use other software, adding a webcam is a very common request so should not be difficult to do.

OBS Studio

I've promoted it pretty hard at this point, so I hope you've chosen OBS Studio as your software solution. On top of being open source, it has a large online community (meaning it is easy to find answers to your questions), is easy to use and has more features than you are ever likely to need.

Tuning the end result

There are some settings in OBS that you should be mindful of in order to tune your final video.

Firstly you want to make sure that the canvas you are integrating all your inputs into is the right size. I chose to record everything in full HD, so that means `1920x1080` pixels. You find this under `File` -> `Settings` and selecting `Video` (see the image below).
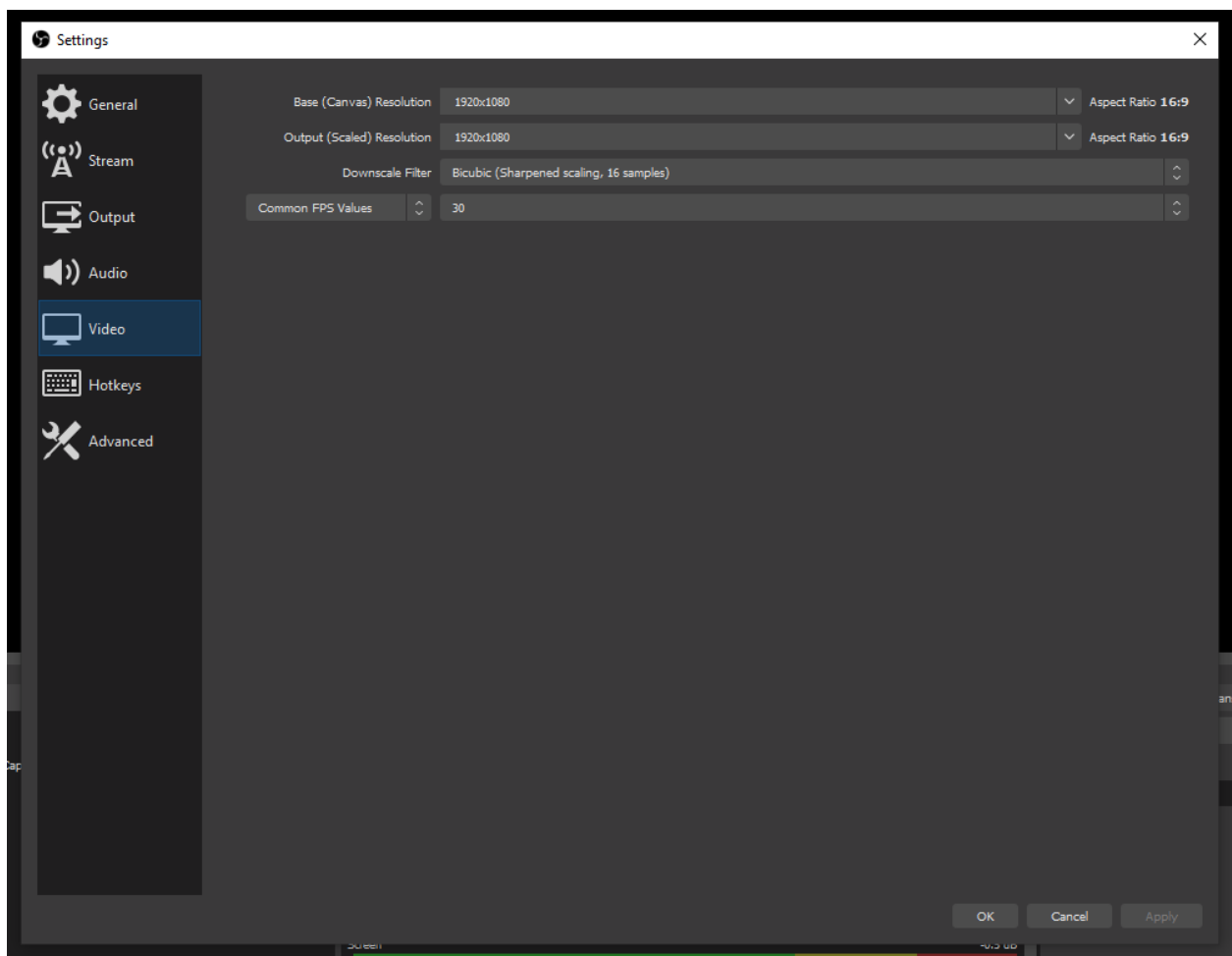
You need to select this for both your "Base Resolution" and "Output Resolution".

Audio

We've discussed the various filters you might want to use, and as you've added additional video sources you'll have noticed that each video source also comes with an audio source. You can use filters on each audio source independently (select the gear icon beside each source).
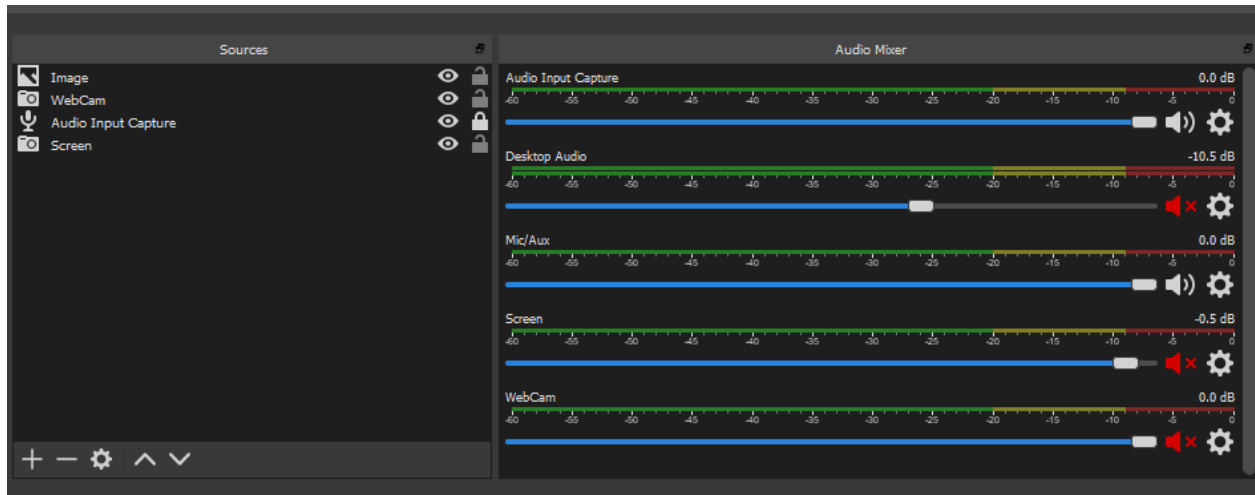

You might notice in my case that I keep most of the sources muted completely. Remember, only unmute the ones you want to record!

Layout

The layout is a matter of personal taste. In my case I like to add a watermark (typically the icon of the project who is supporting the recording) in the top right and the video in the bottom right, with the screen taking up the entire shot. That gives me a total of 4 `Sources`: an image for the watermark, a webcam, the usb audio device I use for sound, and the video capture card I use for the screen. See these in the left of the image below.
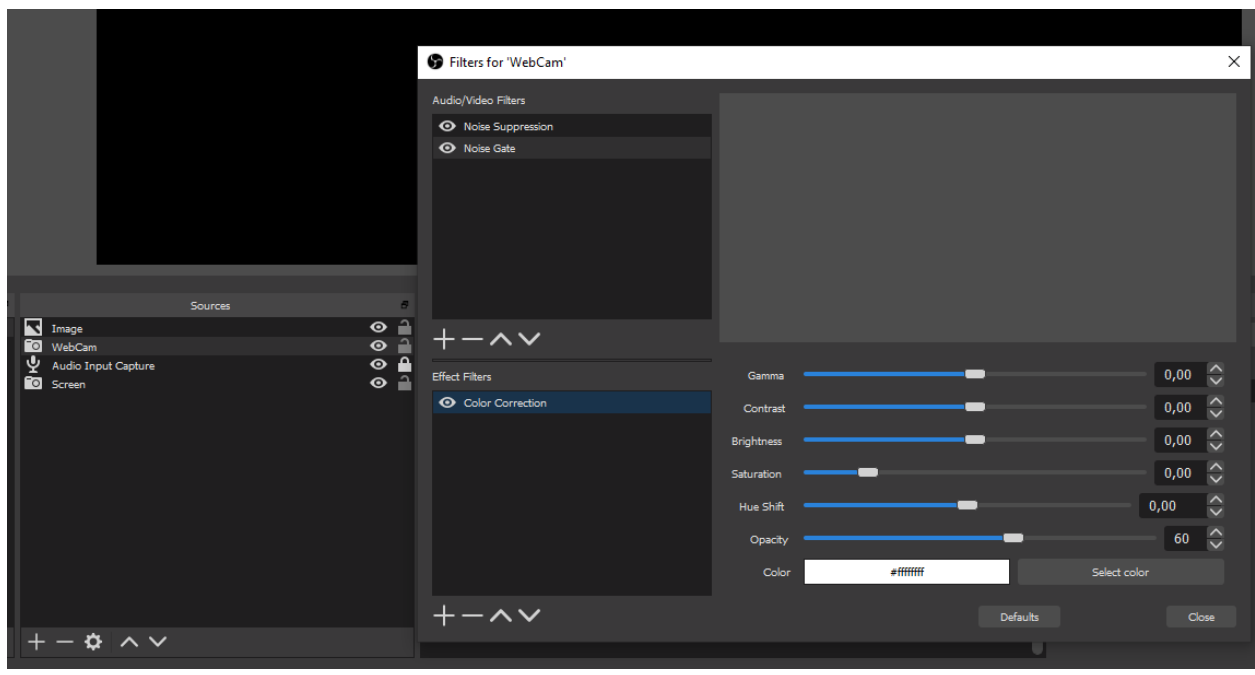
In order that the watermark and video don't obscure the screen, I make then transparent. In both cases I do this by adding a filter to the source. To add a filter you right click on the source (such as `WebCam` in my case) and select `Filters`. For the image, you will immediately see an `Opacity` setting that you can adjust for transparency. For the `WebCam` you need to add `Colour Correction` from the `Effect Filters` to configure an `Opacity` option. The end result looks like the image below (but with no watermark in this case).
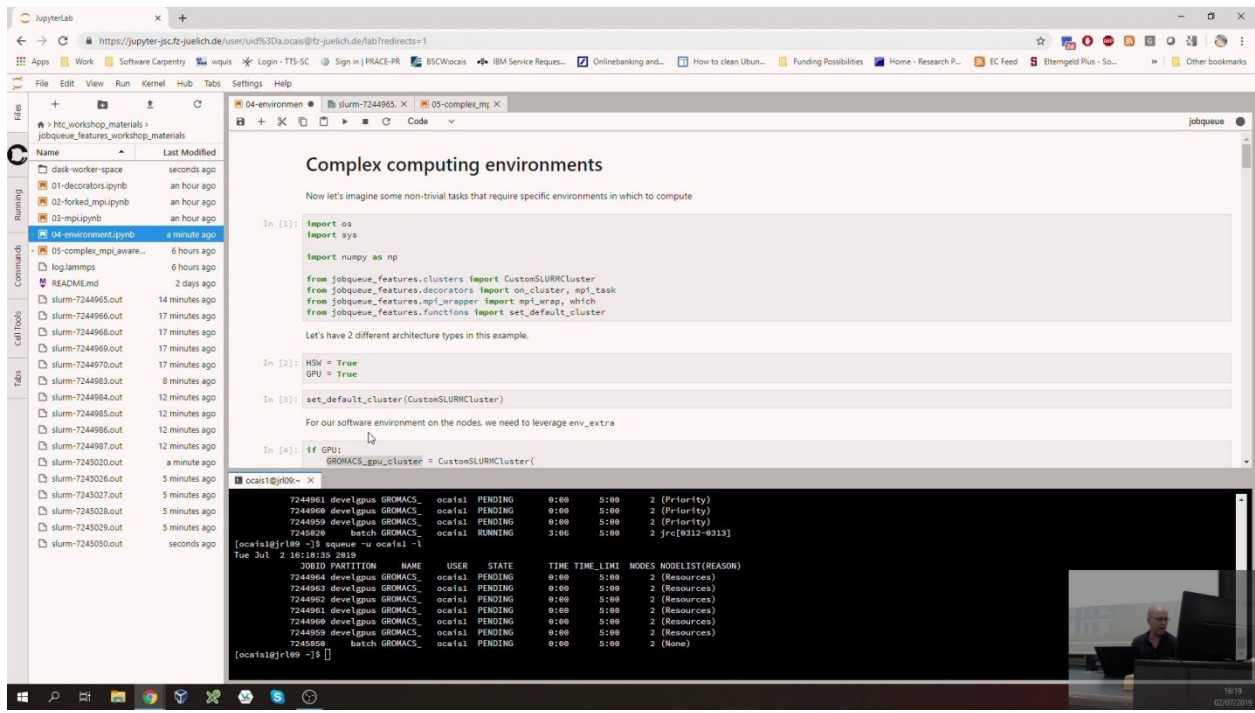
Output video

By default OBS Studio will do some compression on your video to save space. In our case we want a faithful, uncompressed copy so we need to ensure our output video is also recorded in

1080p. The selections I make here are expensive in terms of space but you can always compress the result later with something like HandBrake or FFmpeg.
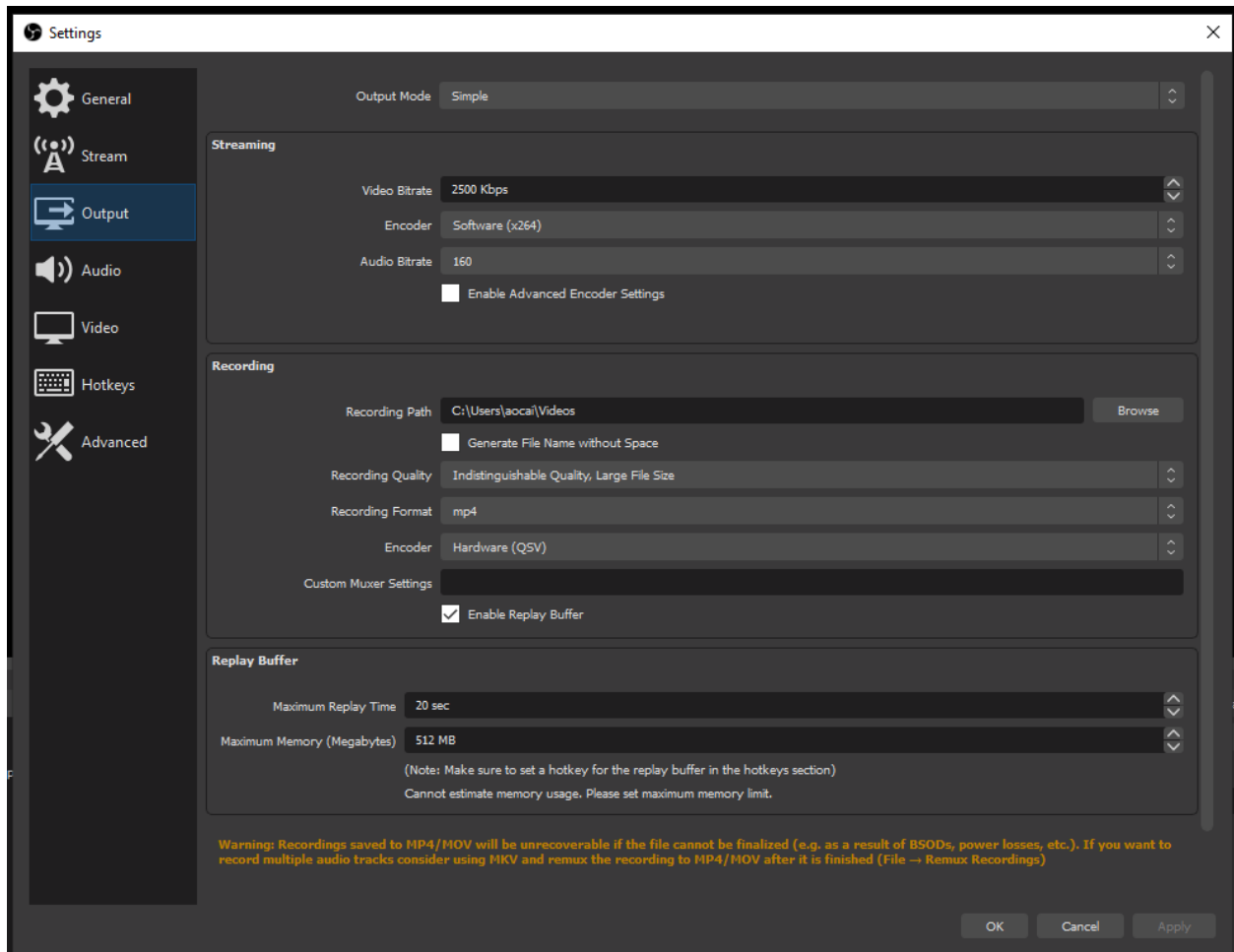
If we are streaming you can also see that this is where we set the quality for our stream. The quality you give for your stream will determine the upload bandwidth you need (about 2400Kbps at the resolution given here).

Streaming your event

OBS Studio is designed for streaming and there are plenty of tutorials on how to do this. Rather than repeat those I give the specific example of YouTube. Somebody else has already created a video tutorial for this. First you need to get your streaming key for your YouTube stream (either watch the tutorial or Google how to do this, it is a bit of a moving target as the YouTube interface may change), once you have that you simply insert the code in OBS: `File =>` `Settings` and select `Stream`, see the image below for typical settings.
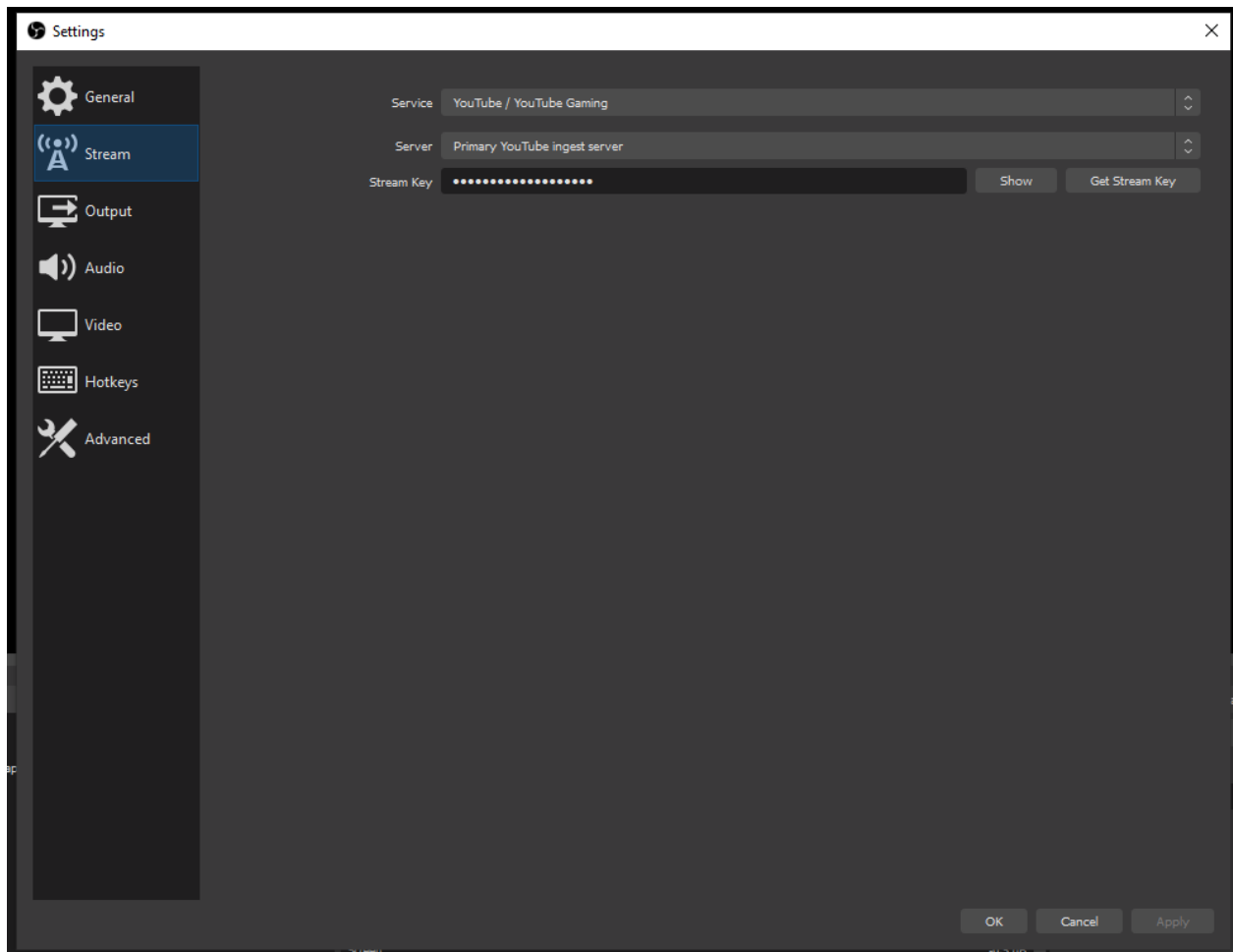
Streaming a closed webinar

A useful corner case you may want to consider is streaming a closed online event (i.e., only a select group are directly connected to the event and can ask questions etc.). This too is possible in a roundabout way. The way I do this is use a preview window for my recorded video and then share that window in the video call. See the image below for how to select this.
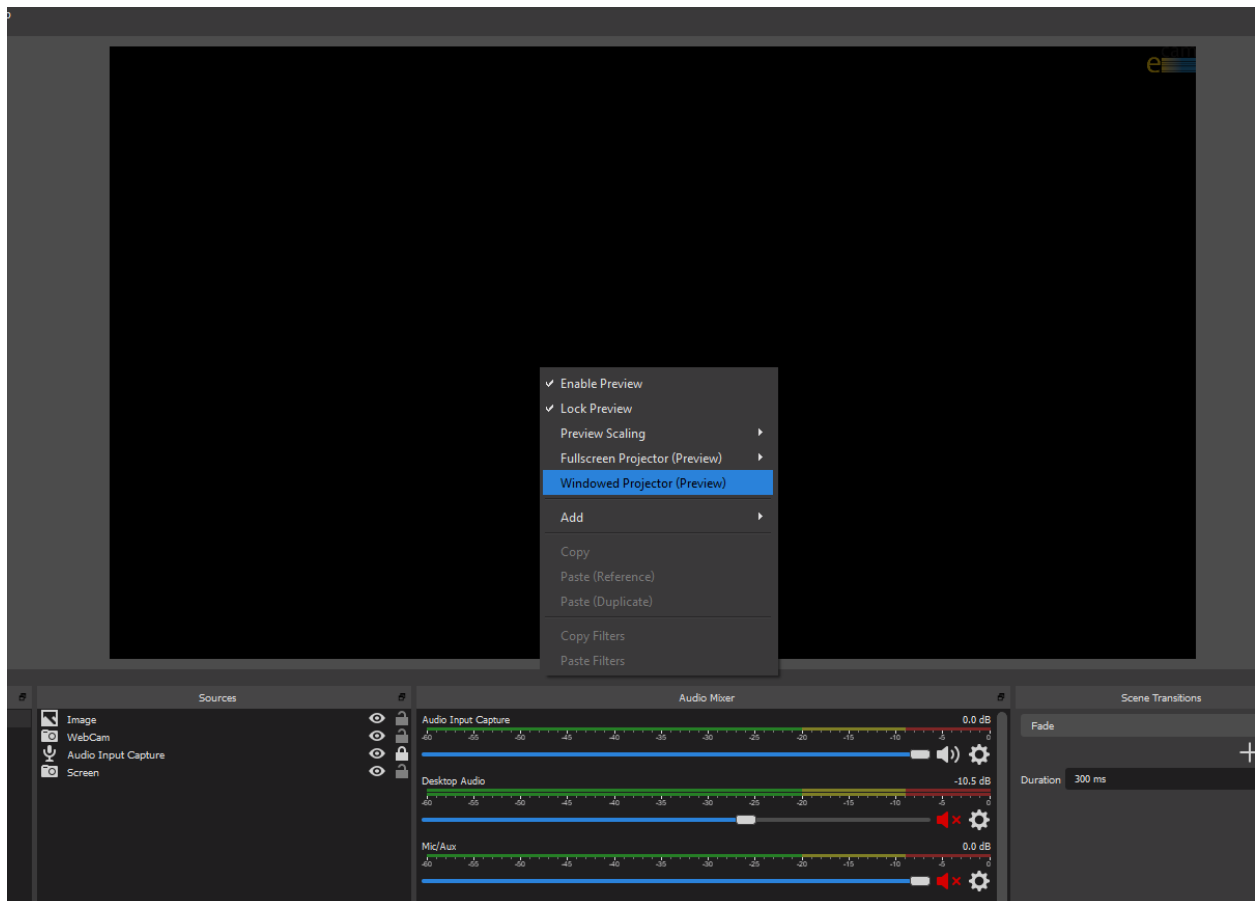
Make sure that the window is somewhat expanded, people on the call will only see it at the resolution it appears at on your screen.

For the audio in the video call I use the audio coming from the speaker. For the audio coming from the online participants, I enable `Desktop Audio` in my OBS `Audio Mixer` which appears beside `Sources`.

Editing the recordings

Hopefully you don't need to join different videos together, you just need to cut out parts of them (for example, where people took a coffee break or at the beginning or end of a clip). For this, there is another open source solution that doesn't involve re-encoding the video (which is a very expensive process), VidCutter. It have a really simple interface and, because it doesn't reencode, is super fast.

Compressing the final result

Once you have an edited result, you may want to compress it. For this I would suggest HandBrake or FFmpeg, both of which are open source. With videos of the type we have been discussing, you can actually be very aggressive in your compression because we are typically not really dealing with moving objects or complex images. The one thing you need to be careful of is to **maintain the same image resolution**, if you reduce it text may become unreadable due to averaging effects.

Closing Remarks

Hopefully this article is helpful in getting an easy to use, and portable, setup for capturing and broadcasting lectures or tutorials in place.

In terms of equipment you could get away with very little: a microphone and capture card could be sourced for less than €100 and the (free) OBS Studio software will paper over most cracks. A €250-€300 spend will get you more reliable equipment but beyond that there are probably diminishing returns.

The software solution is extremely powerful, however if you are doing everything (capturing, streaming, recording, webinar,…) all on the same laptop, it is best not to do any other intensive tasks on the same machine since this is likely to take a sizable amount of resources (CPU, RAM and I/O).

If you feel that this article glosses over important details, or is missing key information, please feel free to drop me an email at a.ocais@fz-juelich.de