



E-CAM Software Platform II

E-CAM Deliverable 6.4

Deliverable Type: Report

Delivered in Month 18– March 2017



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 | D6.4 |
| Deliverable Nature | Report |
| Dissemination Level | Public |
| Contractual Date of Delivery | Project Month 18(March 2017) |
| Actual Date of Delivery | October 12, 2017 |
| Description of Deliverable | Update on E-CAM software tools and platforms jointly by CH and Juelich which includes: <ul style="list-style-type: none"> - the E-CAM library of software modules and interfaces; - end users portal (to access E-CAM's resources, make requests for software developments, register for events); - web infrastructure for teaching tools. |

Document Control Information

| | | |
|------------|-------------------|---|
| Document | Title: | E-CAM Software Platform II |
| | ID: | D6.4 |
| | Version: | As of October 12, 2017 |
| | Status: | Approved by WP Leader |
| | Available at: | https://www.e-cam2020.eu/deliverables |
| Review | Document history: | Internal Project Management Link |
| | Review Status: | Reviewed |
| | Action Requested: | Submit |
| Authorship | Written by: | Alan O'Cais(Juelich Supercomputing Centre) |
| | Contributors: | Catarina Mendonça (EPFL) |
| | Reviewed by: | Jony Castagna (STFC) |
| | Approved by: | A. O'Cais (JSC) |

Document Keywords

| | |
|-----------|-----------------------------------|
| Keywords: | E-CAM, HPC, CECAM, Materials, ... |
|-----------|-----------------------------------|

October 12, 2017

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.

Copyright notices: This deliverable was co-ordinated by Alan O'Cais¹ (Juelich Supercomputing Centre) on behalf of the E-CAM consortium with contributions from Catarina Mendonça (EPFL). This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>.



¹a.ocais@fz-juelich.de

Contents

| | |
|--|-----------|
| Executive Summary | 1 |
| 1 Introduction | 2 |
| 1.1 Scope of the update | 2 |
| 1.2 Target audience | 2 |
| 1.3 Infrastructure delivery and back-up | 2 |
| 2 Library of Software Modules | 3 |
| 2.1 What is a software module? | 3 |
| 2.2 How are software modules contributed to the library? | 4 |
| 3 End Users Portal | 5 |
| 3.1 Project website | 5 |
| 3.1.1 Pilot Projects | 5 |
| 3.2 GitLab | 5 |
| 3.2.1 Version Control | 6 |
| 3.2.2 Continuous Integration | 6 |
| 3.2.3 GitLab Pages | 6 |
| 3.2.4 Authentication | 7 |
| 3.3 Project Management | 7 |
| 3.3.1 Redmine | 7 |
| 3.3.2 Kanban | 7 |
| 3.3.3 EtherPad | 7 |
| 3.3.4 ShareLatex | 8 |
| 4 Web Infrastructure for Teaching Tools | 10 |
| 4.1 Lecture Capture | 10 |
| 4.1.1 Remote participation | 10 |
| 4.1.2 Example of captured content | 11 |
| 4.1.3 Capture of E-CAM events | 11 |
| 4.2 Content Management | 11 |
| 4.2.1 Subcontracting | 12 |
| 4.3 Future Development | 13 |
| 4.3.1 Collaboration with other projects | 13 |
| References | 14 |

List of Figures

| | | |
|----|--|----|
| 1 | An example of one of the four module repositories | 3 |
| 2 | A list of the E-CAM repositories on our GitLab service | 6 |
| 3 | A view of the E-CAM project within our Redmine service | 7 |
| 4 | An example use case for our Kanban service | 8 |
| 5 | An example use case for our Etherpad service | 8 |
| 6 | An example use case for our ShareLatex service | 9 |
| 7 | A subsection of the E-CAM training page | 10 |
| 8 | A captured lecture and YouTube playlist of an entire event | 11 |
| 9 | Captured presentation with navigation carousel. | 12 |
| 10 | Navigation carousel with individual slides for note-taking and comments | 12 |

List of Tables

| | | |
|---|--|---|
| 1 | E-CAM documentation repositories | 3 |
|---|--|---|

Executive Summary

This deliverable describes the provision of online services in the E-CAM project which together form the E-CAM web platform. The primary landing point for information about the resources of the project is the [E-CAM project website](#). This site alone covers the basic requirements of the E-CAM User Portal:

- *E-CAM library of software modules and interfaces*

The [software modules of E-CAM](#) are linked through the website to the E-CAM GitLab service where they are hosted.

- *Access to E-CAM's resources*

All of E-CAM resources are described and available through the E-CAM website. This includes our [upcoming E-CAM events](#) and the [E-CAM online services](#).

- *Make requests for software developments*

We deal with development requests directly on a case-by-case basis with the relevant Work Package (WP) leader and the Software Manager being in direct contact with the person making the request. For this reason we have created a very simple [technical first contact page](#) in order to channel users to the correct WP. We have also created a follow-up questionnaire that is modelled on the [Type C PRACE Preparatory Access](#) form.

- *Register for events*

All E-CAM events are funded and managed through CECAM with registration for events happening through them. On the E-CAM website we provide [detailed descriptions of the E-CAM events](#) and links to the registration process of CECAM.

With respect to the web infrastructure for teaching tools that E-CAM is developing, significant progress has been made. As regards content, we have a capture and remote participation methodology in place for our training events. This methodology has been successfully demonstrated for the 2nd EasyBuild User Meeting that E-CAM co-organised. Relevant portions of the Extended Software Development Workshop (ESDW) events of 2017 have also been captured.

The configuration, and some development, of our content management system has been outsourced with the final system expected to be in place by the end of 2017. The system upon which our service will be built is [Clowder](#), which is developed at [NCSA](#). Clowder is a research data management system designed to support any data format and multiple research domains.

E-CAM has delivered a number of additional online resources and capabilities that go beyond the original scope of this deliverable but are also included and described:

- Software modules are contributed to E-CAM through the documentation repository of the relevant research-related Work Package. The sources for the documentation are stored on the [E-CAM GitLab service](#) with rendered documentation available through [ReadTheDocs.org](#).
- A [Kanban service](#) has been made available to facilitate a lower setup overhead and direct interaction with the issue reporting features of GitLab.
- The [Redmine service](#) is used to manage larger software projects and track related issues. It allows users to manage multiple projects and associated sub-projects. It features project wikis and forums, issue tracking, time tracking, and flexible, role-based access control.
- An [Etherpad service](#) has been provided for a number of participants to simultaneously add to meeting notes and minutes during an online collaborative meeting.
- A [ShareLatex service](#) has been added to facilitate the collaborative production of publication-quality papers using \LaTeX .

Over the lifetime of the project these online services will mature and expand, particularly in the case of online learning where we will soon be able to evaluate the effort associated with content creation and the impact of the end product. With respect to content creation through collaboration with other projects, we are currently collaborating with Partnership for Advanced Computing in Europe (PRACE) on some of the training courses in its PRACE Advanced Training Centre courses. In particular, we collaborate with JSC in 2018 on an upcoming "High-performance scientific computing in C++" course. Furthermore we are contact with the Molecular Sciences Software Institute (MolSSI) which is a US-based project similar in scope to E-CAM (but significantly larger).

1 Introduction

In its role as an e-infrastructure for the wider E-CAM community, a number of services have been made available to facilitate the creation of, collaboration on and publication of software projects.

This report is an update to *E-CAM Software Platform I* [1] that highlights the services that are currently available to E-CAM users and outlines the purpose of each service. These services include

- Online software repositories and associated services,
- A number of project management services effective collaboration on software projects and to streamline the creation of associated publications,
- An online training platform populated with the training content delivered at the project's ESDW events

Over the lifetime of the project these online services will continue to mature and expand, particularly in the case of online learning which is undergoing a significant amount of both development effort and practical implementation.

1.1 Scope of the update

Each of the deliverables in this sequence is intended to be self-contained. In this respect a significant portion of what is included here has already been covered in previous iterations of the *E-CAM Software Platform* set of deliverables [1]. Apart from an overall update of the document, some noteworthy new inclusions in the current document are:

- Combining software module repositories that were previously separated by research WP into a single repository: the [E-CAM Software Library](#)
- Explanatory extensions to Section 2 (which covers the [E-CAM Software Library](#))
- The inclusion of the GitLab pages feature (see Section 3.2.3) in the E-CAM GitLab service

The primary addition in this deliverable is, however, the developments related to our online learning platform (Section 4), these include:

- Creation of lecture capture methodology (and purchase of required hardware)
- Trial run of methodology, including remote participation
- Capture of ESDW events of 2017
- Subcontracting of content management platform development

1.2 Target audience

This deliverable documents the status and plans for the software platform that E-CAM is creating for its target community. As a reporting and planning document, the primary intended audience is our funding body. This document is of limited interest to our wider target community since the services we currently have available are already thoroughly documented on the [E-CAM project website](#) (to which they are much more likely to be exposed).

1.3 Infrastructure delivery and back-up

Each E-CAM service is delivered within a Docker container² and is run on a special purpose server at Centre Européen de Calcul Atomique et Moléculaire (CECAM) headquarters in Lausanne. This approach makes the backup of E-CAM services efficient and straightforward. It also facilitates easy migration of services to either a new server or a new site in the future.

The disk which stores the services is mirrored (to mitigate the impact of a hardware failure) and the services themselves are remotely backed up (with daily/weekly/monthly versions available).

²Docker is an open-source project that automates the deployment of Linux applications inside software containers.

2 Library of 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.

Initially, software modules are contributed to E-CAM through the documentation repository of the relevant research-related WP (which are shown in Table 1, with an example of the landing page for the Classical MD WP shown in Fig. 1).

| | | |
|---------------------------------|--|--|
| Classical MD | WP1 Documentation Repository | WP1 Rendered Documentation |
| Electronic Structure | WP2 Documentation Repository | WP2 Rendered Documentation |
| Quantum Dynamics | WP3 Documentation Repository | WP3 Rendered Documentation |
| Meso- and Multi-scale Modelling | WP4 Documentation Repository | WP4 Rendered Documentation |
| Combined documentation | E-CAM Documentation Repository | E-CAM Rendered Documentation |

Table 1: E-CAM documentation repositories

The screenshot shows the landing page for the E-CAM Classical MD Module Library. At the top, there is a navigation bar with the E-CAM logo, the text 'Classical MD', and several menu items: '0.1', 'SITE', 'PAGE', and 'PATH DENSITY'. A search bar is located on the right side of the navigation bar. Below the navigation bar, the main heading is 'The E-CAM Classical MD Module Library'. To the left of the main text is a small image showing a molecular simulation. To the right is a 'General Information' sidebar containing a list of links: 'Rare events and path sampling', 'OpenPathSampling', 'Pilot Projects', 'Extended Software Development Workshops (ESDWs)', 'Contributing to this documentation', and 'Search Page'. The main text area contains several paragraphs explaining the library's purpose and the definition of a software module. A green heading 'Rare events and path sampling' is visible, followed by a paragraph of text. At the bottom right of the page, there is a small 'gitlab' logo.

Figure 1: An example of one of the four module repositories

After the second set of ESDW events it has been decided to merge these separate repositories into a single [E-CAM Software Library](#). This allows us to include (and cross-reference) software developments that do not map exclusively to a single research WP, as well as concentrate the development effort associated with the library infrastructure. This site is already feature-complete and will be advertised to the users and on the website after it has been approved by project management.

2.1 What is a software module?

In the context of E-CAM, the definition of a software 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.

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.

E-CAM modules necessarily form a heterogeneous collection but modules do however share the traditional computer science 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.

The E-CAM library is a *documentation* repository of software development efforts related to the project. Each module references the source code changes to which it applies, and provides detailed information on the relevant application for the changes as well as how to build and test the associated software.

Perhaps more important than exactly what a module is, is how it is written and used. A final E-CAM module adheres to current best-practice programming style conventions, is well documented and comes with either regression or unit tests (and any necessary associated data). E-CAM modules should be written in such a way that they can potentially take advantage of anticipated hardware developments in the near future (and this is one of the training objectives of E-CAM). Therefore, for a module to be accepted into the E-CAM repository the source code changes in the target application must pass a number of acceptance criteria:

- Style (use meaningful variable names, no global variables,...)
- Source code documentation (each function should be documented with each argument explained)
- Tests (everything you add should have either unit or regression tests)
- Performance (If what you introduce has a significant computational load you should make some performance optimisation effort using an appropriate tool. You should be able to verify that your changes have not introduced unexpected performance penalties, are threadsafe if needed,...)

2.2 How are software modules contributed to the library?

The library uses the [Sphinx documentation generator](#) and the documentation sources are stored on the [E-CAM GitLab service](#) (see Section 3.2). Contributions to the repositories are made through *Merge Requests*³. Each individual modification of the repository automatically causes the associated documentation on [ReadTheDocs.org](#) to be rebuilt.

Within the library itself, E-CAM maintains a set of [Module Contribution Guidelines](#) which explains the technical details of submitting a contribution.

³Merge (or pull) requests are created in a git management application and ask an assigned person to merge two branches. Tools such as GitHub and Bitbucket choose the name pull request since the first manual action would be to pull the feature branch. Tools such as GitLab and Gitorious choose the name merge request since that is the final action that is requested of the assignee.

3 End Users Portal

3.1 Project website

The primary landing point for access to different information about the project is the [E-CAM project website](#). This site alone covers the basic requirements of the End Users Portal:

- *Access E-CAM's resources*

All of E-CAM resources are described and available through the E-CAM website. This includes:

- [E-CAM events calendar](#)
- [E-CAM online services](#)
- [Submitted deliverables](#)
- [E-CAM publications](#) and open access information
- [Workshops scientific reports](#)
- [Software repositories](#) for each of the scientific WPs
- [Pilot projects with industry](#) (see section 3.1.1)
- [Job opportunities](#)
- [Latest news items](#)

- *Make requests for software developments*

Software development request are treated directly on a case-by-case basis with the relevant WP leader and the Software Manager being in direct contact with the person making the request. For this reason we have created a very simple [technical first contact page](#) in order to channel users to the correct WP. We have also created a follow-up questionnaire that is modelled on the [Type C PRACE Preparatory Access](#) form.

- *Register for events*

All E-CAM events are funded and managed through CECAM with registration for events happening through them. On the E-CAM website we provide [detailed descriptions of the E-CAM events](#) and links to the registration process of CECAM.

The scope of the services offered extend beyond the topics above and E-CAM has delivered a number of additional online resources and capabilities that are described in sections 3.2 and 3.3.

3.1.1 Pilot Projects

The project supports academic and industrial research via a set of pilot projects focused on industrially oriented problems, that are sustained by E-CAM funded Postdoctoral Research Associates (PDRAs) supervised by scientists in the team. Each of these pilot projects has a [webpage](#) dedicated to them on the E-CAM website, where the description of the work, modules produced and published results can be seen from people external to E-CAM. Outreach material produced with the results from these pilot projects, including use cases, success stories and interviews, will also be published on this page.

We believe that by doing so, we can increase public visibility towards these projects and E-CAM in particular, and most importantly increase interest by current and potential industrial partners.

The software developed by the PDRAs on the industrial pilot projects will be seen from this page, with a status, expected delivery date for those that are work in progress, and a link to the E-CAM repository where more information about the module documentation and testing can be found. By disseminating the PDRAs work, we also aim at raising employers interest by their professional profile and skills.

3.2 GitLab

We have chosen to create an [E-CAM GitLab service](#) as our online repository management system. This allows us the possibility of creating unlimited private repositories for our user community, something that is essential when considering either licensing or the Intellectual Property (IP) requirements of our userbase (particularly those in industry

but also those in academia who may wish to embargo the release of their software until associated results have been published).

As well as being our online Git repository manager, it also provide us with additional repository services such as a wiki, issue tracking, continuous integration, continuous deployment and the ability to create static websites for your GitLab projects. We outline some of these services in the following subsections (and they have also been mentioned in [D6.2: E-CAM software development tools](#)).

3.2.1 Version Control

The primary software development tool that E-CAM provides is a version-control system. Git is ubiquitous in modern software development and is the obvious choice as the main version control system.

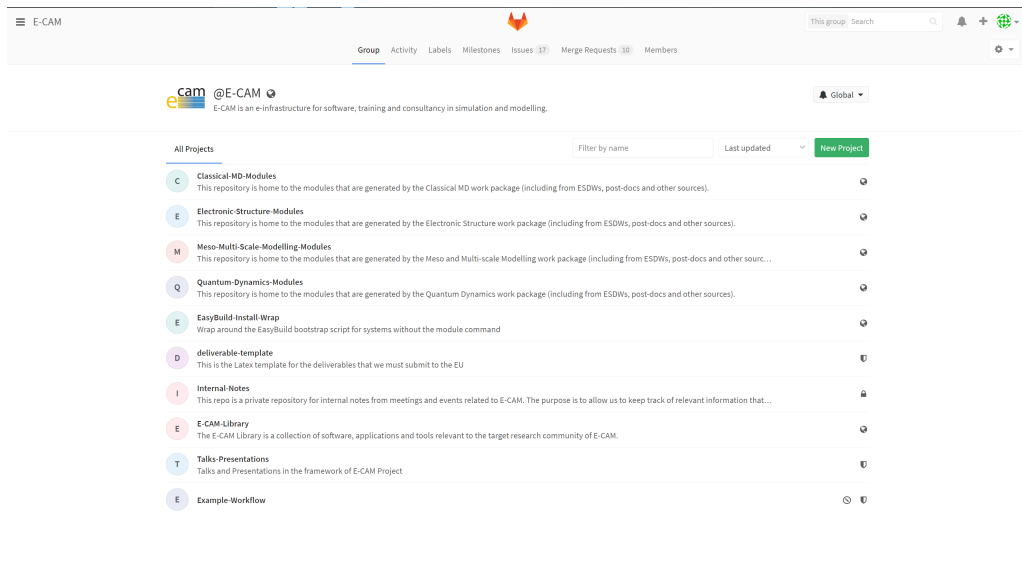


Figure 2: A list of the E-CAM repositories on our [GitLab service](#)

A screen shot of the direct repositories of the E-CAM software group can be seen in Fig. 2. Some of these repositories have already been discussed in Section 2. The actual number of software repositories is far greater than this: there are currently 96 users with a total of 177 projects (as of Sept. 2017). On our main project website we also provide a link to [all publicly accessible repositories of E-CAM](#).

3.2.2 Continuous Integration

Continuous Integration is a software development practice in which a developer builds and tests software every time that they create new code in an application. GitLab provides a method of implementing continuous integration and we in turn make it available to our community.

3.2.3 GitLab Pages

With GitLab Pages you can create static websites for your GitLab projects, groups, or user accounts. You can use any static website generator: Jekyll, Middleman, Hexo, Hugo, Pelican, ... The feature leverages the continuous integration feature of GitLab to ensure continuous delivery of the project documentation.

The final webpages are served under the domain `e-cam2020.io` and an example of such a project can be seen for a [clone of E-CAM documentation pages](#).

3.2.4 Authentication

Where possible we leverage the OAuth⁴ features of GitLab in order to have single-sign-on capabilities for all E-CAM services.

3.3 Project Management

Supporting and streamlining the management of software projects, including bug-tracking and feature requests, is another area where E-CAM can be of service to the community. Many of these capabilities already exist in our GitLab service, however we have enabled a number of additional services that seek to make remote collaboration on software projects as straightforward as possible.

3.3.1 Redmine

[Redmine](#) is a free and open source, web-based project management and issue tracking tool. It allows users to manage multiple projects and associated sub-projects. It features project wikis and forums, time tracking, and flexible, role-based access control. It includes a calendar and Gantt charts to aid visual representation of projects and their deadlines. Redmine integrates with various version control systems and includes a repository browser and a diff viewer.

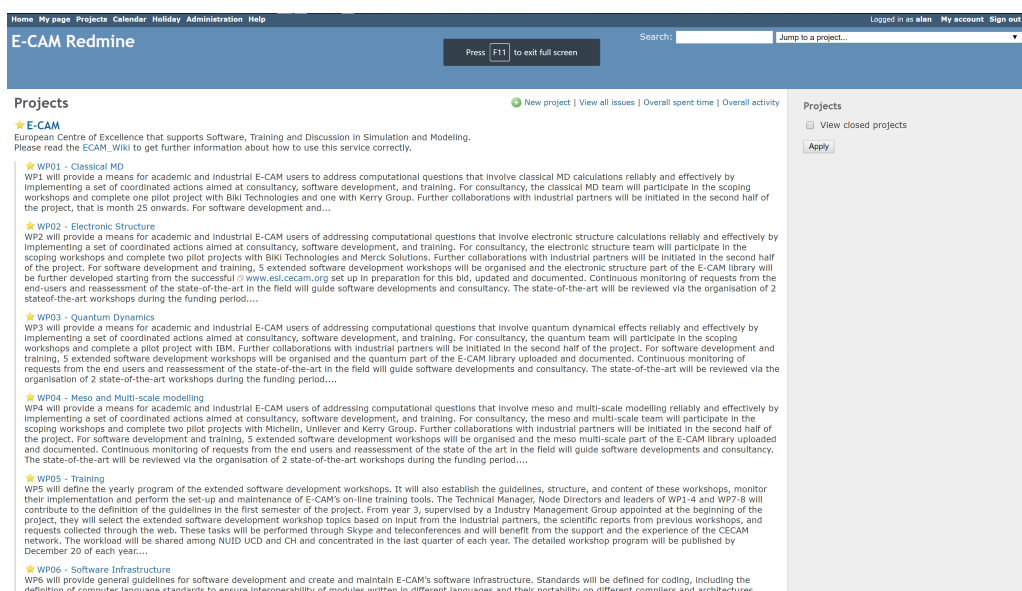


Figure 3: A view of the E-CAM project within our [Redmine](#) service

We also use the [Redmine](#) service to manage the overall E-CAM project and this example of Redmine usage can be seen in Fig. 3

3.3.2 Kanban

In order to facilitate a lower setup overhead and direct interaction with the issue reporting features of GitLab, we have also made a [Kanban](#)⁵ service available that connects directly to the GitLab repositories.

An example of its use in the case of one of the E-CAM repositories can be seen in Fig. 4

3.3.3 EtherPad

In geographically distributed collaborative projects, regular meetings are essential. To facilitate taking minutes in such meetings we have created an [Etherpad](#) service. Etherpad is a web-based collaborative real-time editor, allowing

⁴OAuth is an open standard for authorization, commonly used as a way for Internet users to authorize websites or applications to access their information on other websites but without giving them the passwords.

⁵[Kanban](#) in the context of software development provides a visual process-management system that aids decision-making about what to produce, when to produce it, and how much to produce.

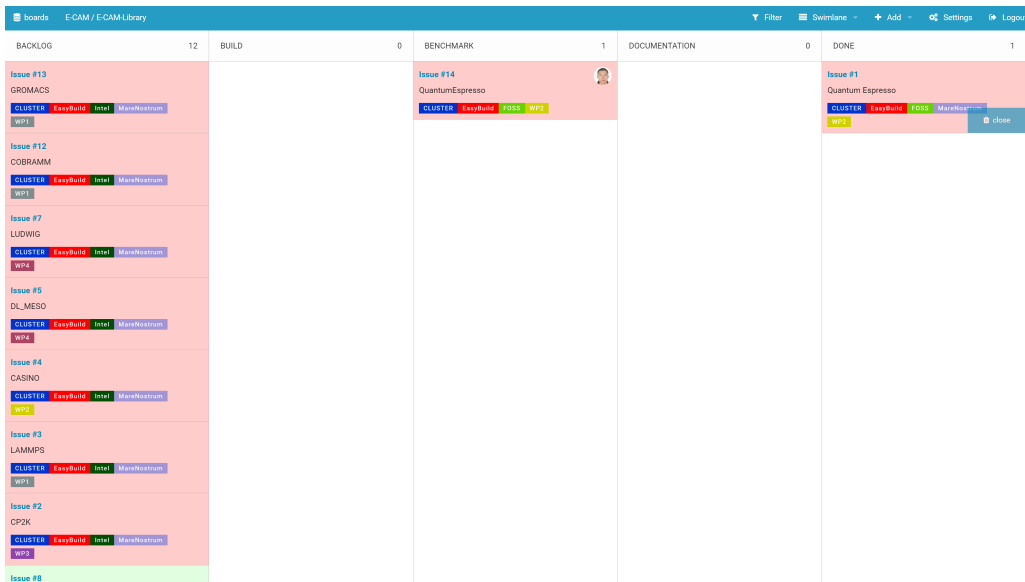


Figure 4: An example use case for our [Kanban service](#)

authors to simultaneously edit a text document, and see all of the participants' edits in real-time, with the ability to display each author's text in their own color.

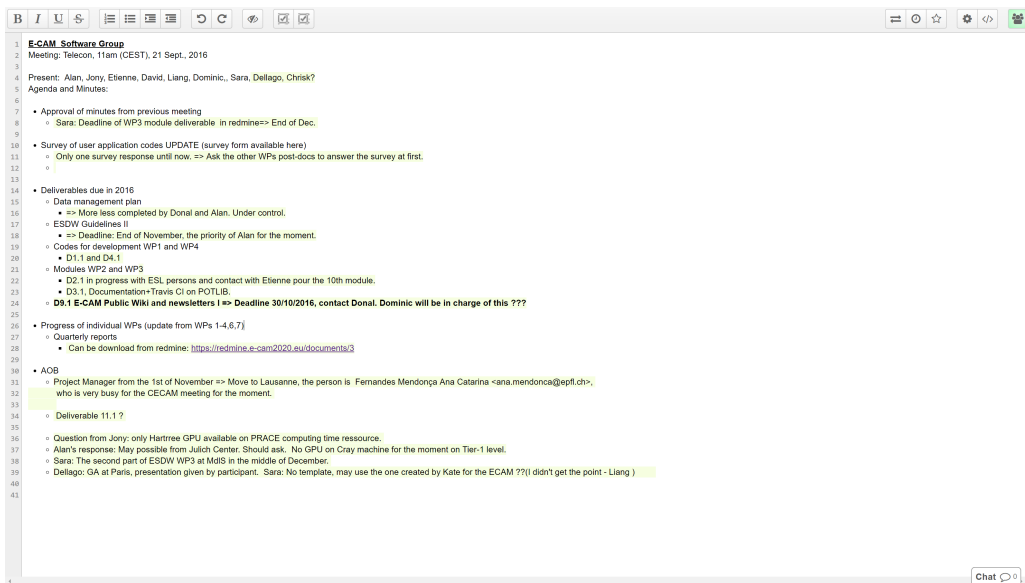


Figure 5: An example use case for our [Etherpad service](#)

An example use of an etherpad is given in Fig. 5.

3.3.4 ShareLatex

Many of the E-CAM beneficiaries are academic researchers, who use Latex to produce technical documents and papers. ShareLaTeX is an online LaTeX editor that allows real-time collaboration and online compiling of projects to PDF format. In comparison to other LaTeX editors, ShareLaTeX is a server-based application, which is accessed through a web browser. We have created a [ShareLatex service](#) for use by the E-CAM community to facilitate easier collaboration on project-related papers. This tool is also used in the production of all E-CAM deliverables. An example can be seen in Fig. 6

The image shows a LaTeX Beamer presentation slide titled "Deliverable 6.4". The slide is divided into three main sections: "Introduction", "What can the software do? What are its limitations?", and "ESPResSo+". The "Introduction" section describes the ESPResSo software as a flexible simulation tool for soft-matter systems with hydrodynamic interactions. The "What can the software do?" section lists various features like parallelization and coupling to other packages. The "ESPResSo+" section highlights its modularity and adaptability. The right side of the slide shows two 3D visualizations of polymer chains and a caption for Figure 1.

Figure 6: An example use case for our ShareLatex service

4 Web Infrastructure for Teaching Tools

E-CAM wishes to develop an appropriate online training infrastructure over the course of the project. To this end, it is establishing strong partnerships with PRACE and other similar initiatives in Europe and the US in order to provide appropriate training content that can bring the E-CAM user communities to extreme scale.

This infrastructure to support these efforts continues to be under development but significant progress has been made in the last 12 months. The extent of this effort is outlined in the following subsections.

List of online HPC training repositories

| | |
|--|---|
| NERSC Training Page | https://www.nersc.gov/users/training/ |
| University of Illinois College of Engineering Computational Science and Engineering | http://cse.illinois.edu/training |
| HPC University | hpcuniversity.org |
| OU Supercomputing Center for Education & Research (OSCER) | http://www.oscer.ou.edu/education.php |
| LinkSCEEM | http://supercomputing.cyi.ac.cy/index.php |
| Institute for Advanced Simulation (IAS) Jülich Supercomputing Centre (JSC) | http://www.fz-juelich.de/ias/jsc/EN/Expertise/Workshops/workshops_node.html |
| CINECA – SuperComputing Applications and Innovation | http://www.hpc.cineca.it/content/training-2015 |
| ARCHER – UK National Supercomputing Service | http://www.archer.ac.uk/training/ |

Figure 7: A subsection of the [E-CAM training page](#)

While we are not yet at the point where we are publishing our captured material online via our own service, we expect to reach that point by the end of 2017. Until we do, we are providing some information regarding existing (online) training opportunities relevant to our community through the [E-CAM training page](#). In Fig. 7 we highlight some of the training sites that are especially relevant to a High Performance Computing (HPC) environment.

4.1 Lecture Capture

For lecture capture, gamers have fantastic (and fantastically cheap) hardware and software options: [Live Gamer Portable](#) capture card for screen capture and the open source [Open Broadcast Software](#) that is capable of processing multiple sources, combining them and live streaming them. We have created a fully portable capture kit to capture training events made up of just a laptop, the [Live Gamer Portable](#) to capture the screen, an HD webcam and a wireless mic for speaker and audience.

4.1.1 Remote participation

The final combined video stream can also be passed as a screen share via, e.g., a Google Hangout (for public lectures) or a service such as [GoToMeeting](#) (for limited private participation). With this approach the remote participant can engage in realtime interaction with the speaker and/or local audience.

We also use the [E-CAM Etherpad service](#) to take live notes of the event. This tool has a chat window where any participant can ask questions without interrupting the flow of the speaker (which can be very useful if they may have, for example, missed a terminal command on a previous slide).

We intend to advertise remote participation to our [ESDW](#) events in our 2018 program of activities.

4.1.2 Example of captured content

We conducted a trial run of all of these facilities during the [2nd EasyBuild User Meeting](#) that was co-organised by E-CAM.

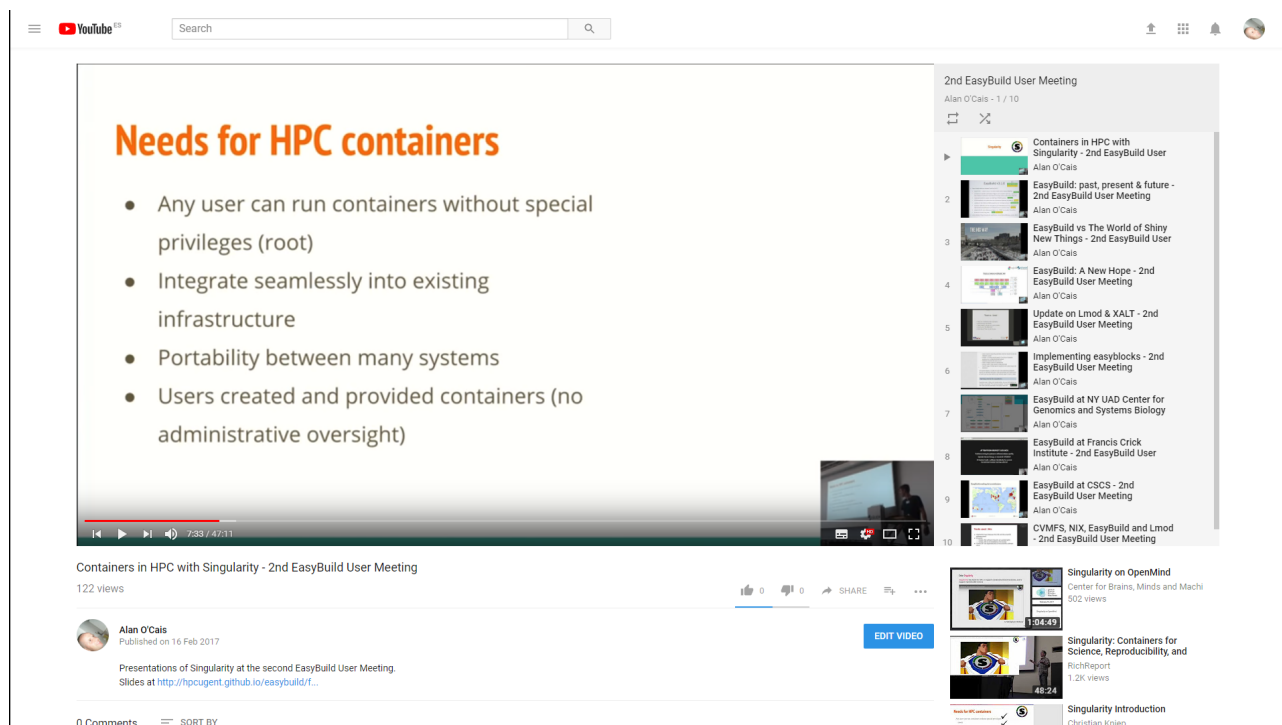


Figure 8: A captured lecture and YouTube playlist of an entire event.

We successfully captured all lectures with only minor technical issues with respect to MAC compatibility with the capture card. We even had a remote lecture with the lecturer able to discuss directly with the audience, the only technical issue was poor quality of the link at one point (which was outside of our control). The full program of the event with links to the captured content is available at the [2nd EasyBuild User Meeting website](#).

The [captured lectures are published on YouTube](#) and the playlist can be seen in Figure 8 (which includes an example of what a captured lecture looks like).

4.1.3 Capture of E-CAM events

High quality content that will translate well to our online training portal has been captured at all our ESDW events in 2017. After an initial evaluation of the effort involved and the measured impact, we expect to extend this to all our events in 2018.

4.2 Content Management

It is most definitely not purely the content generated by E-CAM which is of interest to the E-CAM community. In addition to managing our own content, we also want to be able to include material from other sources (whether it be video content, online material or PDF documents). We want to store and treat this content (video, document or URL) in a consistent manner. We also want the ability to associate additional material with each content type (source code for tutorials is an obvious example)

The content that is generated by E-CAM is captured with the permission of the speakers, in many cases the speakers are not comfortable with the material being publicly available but are willing to make the content available to the E-CAM user community. This ability to restrict who can access material places a further requirement on the selection of a content management system.

Once the video has been captured we would like the ability to extract additional information about the video, in particular we would like to automatically find the slide transitions and extract slide snapshots for these transitions that can also be used for navigating the presentation.

These requirements have brought us to [Clowder](#), which is developed at [NCSA](#). Clowder is a research data management system designed to support any data format and multiple research domains. Clowder contains three major extension points: preprocessing, processing and previewing. When new data is added to the system, whether it is via the web front-end, or through the RESTful web services, preprocessing is off-loaded to extraction services for extracting appropriate data and metadata. The extraction services attempt to extract information and run preprocessing steps based on the type of the data, for example to create previews. This raw metadata is presented to the user in the Clowder web interface.

The ability to add new file types and develop our own extraction services that act on them to create new metadata are seen as very valuable to our infrastructure. In the short term it has allowed to create a computer vision algorithm to extract and store slide transitions alongside video material, but we also foresee further use cases (for example, automated parsing of EasyBuild configuration files to create URL-based cross-referencing, syntax highlighting of uploaded source code, . . .). The automation aspect of the extraction services ensure that E-CAM should be able to expend a controlled amount of effort on adding new data even when the amount of new data is large.

The logical structure of Clowder is also of benefit to E-CAM. Single *files* are collected into *datasets*, which are a group of files that through some defined relationship or corresponding metadata are strongly tied together and not representable otherwise by the individual files. *Datasets* are what we expect our captured lectures to be (made up of the associated videos, lecture materials, reading materials, tutorial content and software requirements). *Collections* are a user defined group of *datasets* and other *collections*. In E-CAM, we foresee an [ESDW](#) being considered in this way. A *space* is a group of *collections*, *datasets*, and *files* with defined user access rights. Spaces are used to share data within datasets and collections with other users, in our case this will be how we will restrict access to (some of) our material.

4.2.1 Subcontracting

Much of the development work for our extractor services has been outsourced. Clowder is still under active development and an experienced computer scientist was required to understand and implement the requirements that E-CAM has outlined.

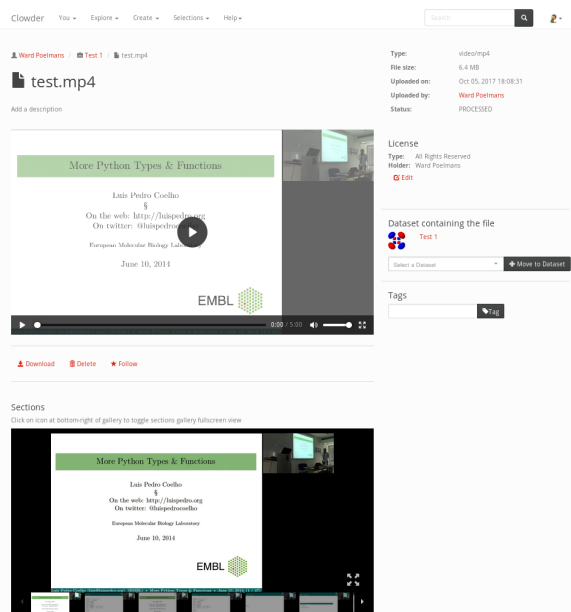


Figure 9: Captured presentation with navigation carousel.

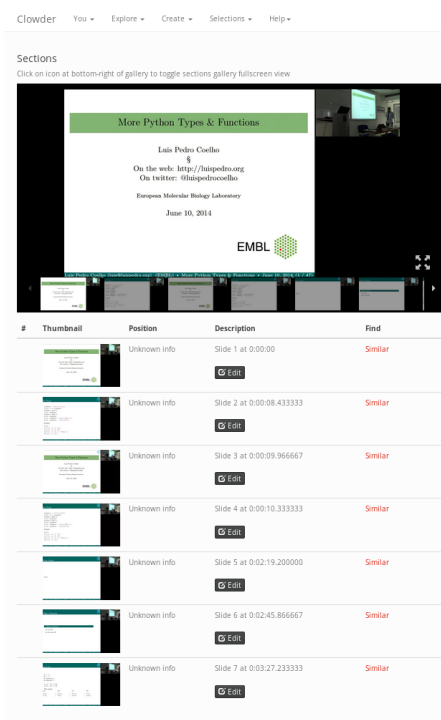


Figure 10: Navigation carousel with individual slides for note-taking and comments

As of September 2017 much of this technical infrastructure is now in place, as can be seen in Figures 9 and 10.

4.3 Future Development

Given the resources attributed to the online training component of E-CAM, the services that E-CAM can potentially provide are limited. Among the list of possibilities are:

- Clowder still requires further development configuration. We expect to release the service within 2017 when the content from the 2017 ESDW workshops has been integrated.
- HPC services - this service would provide an HPC environment to train users and to test modules before moving the testing onto larger PRACE machines. There is a possibility to create a pseudo-HPC-cluster using Docker containers and provide access to this over the internet. This would have significant development overhead as no such drop-in service currently exists, and we are carefully weighing the benefit of this service against the cost of provision.

In 2018, E-CAM will focus on streaming portions of the training events that it is associated. The goal is to allow remote participation for E-CAM members, and in particular for industrial partners who cannot commit to the allocated time for our ESDW events.

4.3.1 Collaboration with other projects

E-CAM has committed to collaborating with PRACE on some of the training courses in its PRACE Advanced Training Centre courses. In particular, we collaborate with JSC in 2018 on an upcoming "High-performance scientific computing in C++" course.

Furthermore we are contact with the MolSSI which is a US-based project that aims to serve as a nexus for science, education, and cooperation serving the worldwide community of computational molecular scientists – a broad field including of biomolecular simulation, quantum chemistry, and materials science. We are investigating the possibility of collaboration from a training perspective given the clear overlap of the scope of our projects.

References

Acronyms Used

CECAM Centre Européen de Calcul Atomique et Moléculaire

HPC High Performance Computing

PRACE Partnership for Advanced Computing in Europe

ESDW Extended Software Development Workshop

WP Work Package

IP Intellectual Property

PDRA Postdoctoral Research Associate

PATC PRACE Advanced Training Centre

MolSSI Molecular Sciences Software Institute

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/35>
a.ocais@fz-juelich.de ... <mailto:a.ocais@fz-juelich.de>
<http://creativecommons.org/licenses/by/4.0> ... <http://creativecommons.org/licenses/by/4.0>

Page iii

GitLab service ... <https://gitlab.e-cam2020.eu/>
Redmine service ... <https://redmine.e-cam2020.eu/>
Kanban service ... <https://kanban.e-cam2020.eu/>
Etherpad service ... <https://etherpad.e-cam2020.eu/>
ShareLatex service ... <https://sharelatex.e-cam2020.eu/>
E-CAM training page ... <https://www.e-cam2020.eu/resources/training/>
captured lecture and YouTube playlist of an entire event ... <https://www.youtube.com/playlist?list=PLVA9BuLC1j-yfxp2w-wraAGDCmhjb3o5Y>

Page 1

E-CAM project website ... <https://www.e-cam2020.eu/>
software modules of E-CAM ... <https://www.e-cam2020.eu/software-library/>
upcoming E-CAM events ... <https://www.e-cam2020.eu/events>
E-CAM online services ... <https://www.e-cam2020.eu/resources/>
technical first contact page ... <https://www.e-cam2020.eu/technical-questions/>
Type C PRACE Preparatory Access ... <http://www.prace-ri.eu/prace-preparatory-access/>
detailed descriptions of the E-CAM events ... https://www.e-cam2020.eu/events/list/?tribe_paged=1&tribe_event_display=past
Clowder ... <https://clowder.ncsa.illinois.edu/>
NCSA ... <http://www.ncsa.illinois.edu/>
E-CAM GitLab service ... <https://gitlab.e-cam2020.eu/>
ReadTheDocs.org ... <https://readthedocs.org/>
Kanban service ... <https://kanban.e-cam2020.eu/>
Redmine service ... <https://redmine.e-cam2020.eu/>
Etherpad service ... <https://etherpad.e-cam2020.eu/>
ShareLatex service ... <https://sharelatex.e-cam2020.eu/>

Page 2

E-CAM Software Library ... <http://e-cam.readthedocs.io/en/latest/>
E-CAM Software Library ... <http://e-cam.readthedocs.io/en/latest/index.html>
E-CAM project website ... <https://www.e-cam2020.eu/>

Page 3

WP1 Documentation Repository ... <https://gitlab.e-cam2020.eu/E-CAM/Classical-MD-Modules>
WP1 Rendered Documentation ... <http://e-cam-classical-md-modules.readthedocs.io/en/latest/index.html>
WP2 Documentation Repository ... <https://gitlab.e-cam2020.eu/E-CAM/Electronic-Structure-Modules>
WP2 Rendered Documentation ... <http://e-cam-electronic-structure-modules.readthedocs.io/en/latest/>
WP3 Documentation Repository ... <https://gitlab.e-cam2020.eu/E-CAM/Quantum-Dynamics-Modules>
WP3 Rendered Documentation ... <http://e-cam-quantum-dynamics-modules.readthedocs.io/en/latest/>
WP4 Documentation Repository ... <https://gitlab.e-cam2020.eu/E-CAM/Meso-Multi-Scale-Modelling-Modules>
WP4 Rendered Documentation ... <http://e-cam-meso-and-multi-scale-modelling-modules.readthedocs.io/en/latest/>
E-CAM Documentation Repository ... <https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library>
E-CAM Rendered Documentation ... <http://e-cam.readthedocs.io/en/latest/>
E-CAM Software Library ... <http://e-cam.readthedocs.io/en/latest/>

Page 4

Sphinx documentation generator ... <http://www.sphinx-doc.org/en/stable/index.html>
E-CAM GitLab service ... <https://gitlab.e-cam2020.eu/e-cam/E-CAM-Library>
ReadTheDocs.org ... <https://readthedocs.org/>
Module Contribution Guidelines ... <http://e-cam.readthedocs.io/en/latest/contributing.html#id2>

Page 5

E-CAM project website ... <https://www.e-cam2020.eu/>
E-CAM events calendar ... <https://www.e-cam2020.eu/events>
E-CAM online services ... <https://www.e-cam2020.eu/resources/>
Submitted deliverables ... <https://www.e-cam2020.eu/deliverables/>
E-CAM publications ... <https://www.e-cam2020.eu/publications/>
Workshops scientific reports ... <https://www.e-cam2020.eu/scientific-reports/>
Software repositories ... <https://www.e-cam2020.eu/software-library/>
Pilot projects with industry ... <https://www.e-cam2020.eu/pilot-projects-with-industry/>
Job opportunities ... <https://www.e-cam2020.eu/vacancies/>
Latest news items ... <https://www.e-cam2020.eu/category/news/>
technical first contact page ... <https://www.e-cam2020.eu/technical-questions/>
Type C PRACE Preparatory Access ... <http://www.prace-ri.eu/prace-preparatory-access/>
detailed descriptions of the E-CAM events ... https://www.e-cam2020.eu/events/list/?tribe_paged=1&tribe_event_display=past
webpage ... <https://www.e-cam2020.eu/pilot-projects-with-industry/>
E-CAM GitLab service ... <https://gitlab.e-cam2020.eu/>

Page 6

D6.2: E-CAM software development tools ... https://www.e-cam2020.eu/wp-content/uploads/2017/09/D6.2_Resubmit_04092017.pdf
GitLab service ... <https://gitlab.e-cam2020.eu/>
all publicly accessible repositories of E-CAM ... <http://gitlab.e-cam2020.eu/explore/projects>
clone of E-CAM documentation pages ... <http://e-cam.e-cam2020.io/E-CAM-Library>

Page 7

Redmine ... <http://www.redmine.org/>
Redmine service ... <https://redmine.e-cam2020.eu/>
Redmine service ... <https://redmine.e-cam2020.eu/>
Kanban ... <http://kanbanblog.com/explained/>

Page 8

Kanban service ... <https://kanban.e-cam2020.eu/>

Page 7

Etherpad service ... <https://etherpad.e-cam2020.eu/>

Page 8

Etherpad service ... <https://etherpad.e-cam2020.eu/>

Page 9

ShareLatex service ... <https://sharelatex.e-cam2020.eu/>

Page 8

ShareLatex service ... <https://sharelatex.e-cam2020.eu/>

Page 10

E-CAM training page ... <https://www.e-cam2020.eu/resources/training/>

E-CAM training page ... <https://www.e-cam2020.eu/resources/training/>

Live Gamer Portable ... http://gamerzone.avermedia.com/game_capture/live_gamer_portable

Open Broadcast Software ... <https://obsproject.com/>

Live Gamer Portable ... http://gamerzone.avermedia.com/game_capture/live_gamer_portable

GoToMeeting ... <https://www.gotomeeting.com/>

E-CAM Etherpad service ... <https://etherpad.e-cam2020.eu/>

Page 11

2nd EasyBuild User Meeting ... <https://github.com/easybuilders/easybuild/wiki/2nd-EasyBuild-User-Meeting>
captured lecture and YouTube playlist of an entire event ... <https://www.youtube.com/playlist?list=PLVA9BuLC1j-yfxp2w-wraAGDCmhjb3o5Y>

2nd EasyBuild User Meeting website ... <https://github.com/easybuilders/easybuild/wiki/2nd-EasyBuild-User-Meeting>
captured lectures are published on YouTube ... <https://www.youtube.com/playlist?list=PLVA9BuLC1j-yfxp2w-wraAGDCmhjb3o5Y>

Page 12

Clowder ... <https://clowder.ncsa.illinois.edu/>

NCSA ... <http://www.ncsa.illinois.edu/>

Citations

[1] A. O. Cais, "E-CAM Software Platform I," Nov. 2016. [Online]. Available: <https://doi.org/10.5281/zenodo.841751>