



**SEVENTH FRAMEWORK PROGRAMME**

**Research Infrastructure**

**FP7-INFRASTRUCTURES-2010-2 – INFRA-2010-1.2.3:**

**Virtual Research Communities**

**Combination of Collaborative Project and Coordination and Support  
Actions (CPCSA)**



**LinkSCEEM-2**

**Linking Scientific Computing in Europe and the Eastern  
Mediterranean – Phase 2**

**Grant Agreement Number: RI-261600**

**D4.6**

**Report on available training portal content**

*Final*

Version: 1.0  
Author(s): Alan O’Cais, JSC  
Date: 26/02/2015

## Project and Deliverable Information Sheet

<b>LinkSCEEM Project</b>	<b>Project Ref. №:</b> RI-261600	
	<b>Project Title:</b> LinkSCEEM-2	
	<b>Project Web Site:</b> <a href="http://www.linksceem.eu/ls2/">http://www.linksceem.eu/ls2/</a>	
	<b>Deliverable ID:</b> < D4.6>	
	<b>Deliverable Nature:</b> <DOC_TYPE: Report>	
	<b>Deliverable Level:</b> PU *	<b>Contractual Date of Delivery:</b> 28 / 02 / 2015
		<b>Actual Date of Delivery:</b> 17 / 03 / 2015
<b>EC Project Officer: Leonardo Flores Anover</b>		

\* - The dissemination level are indicated as follows: **PU** – Public, **PP** – Restricted to other participants (including the Commission Services), **RE** – Restricted to a group specified by the consortium (including the Commission Services). **CO** – Confidential, only for members of the consortium (including the Commission Services).

## Document Control Sheet

<b>Document</b>	<b>Title:</b> <Report on available training portal content>	
	<b>ID:</b> <D4.6>	
	<b>Version:</b> <1.0 >	<b>Status:</b> Final
	<b>Available at:</b> <a href="http://www.eniac.cyi.ac">http://www.eniac.cyi.ac</a>	
	<b>Software Tool:</b> Microsoft Word 2007	
	<b>File(s):</b> LinkSCEEM-2-D4.6 Report on available training portal content.docx	
<b>Authorship</b>	<b>Written by:</b>	Alan O’Cais, JSC
	<b>Contributors:</b>	WP 4 participants
	<b>Reviewed by:</b>	
	<b>Approved by:</b>	PMO

**Document Status Sheet**

<b>Version</b>	<b>Date</b>	<b>Status</b>	<b>Comments</b>
1.0	26/02/2015	Final version	
1.0	03/03/2015	approved	

**Document Keywords**

<b>Keywords:</b>	LinkSCEEM-2, Computational Science, HPC, e-Infrastructure, Eastern Mediterranean
------------------	--

© 2010 LinkSCEEM-2 Consortium Partners. All rights reserved.

## Table of Contents

<b>Project and Deliverable Information Sheet .....</b>	<b>ii</b>
<b>Document Control Sheet.....</b>	<b>ii</b>
<b>Document Status Sheet .....</b>	<b>iii</b>
<b>Document Keywords .....</b>	<b>iii</b>
<b>Table of Contents .....</b>	<b>iv</b>
<b>Executive summary .....</b>	<b>iv</b>
<b>1 Introduction.....</b>	<b>5</b>
<b>2 Portal Content.....</b>	<b>5</b>
<b>2.1 Capturing Content .....</b>	<b>6</b>
<b>2.2 Hands-on Material .....</b>	<b>6</b>
<b>3 Portal Technologies.....</b>	<b>6</b>
<b>3.1 Web Portal.....</b>	<b>7</b>
<b>3.2 Content Repository .....</b>	<b>7</b>
<b>3.3 Terminal Emulator .....</b>	<b>8</b>
<b>3.4 HPC Resource.....</b>	<b>9</b>
<b>4 Available Material .....</b>	<b>10</b>
<b>5 Outlook.....</b>	<b>11</b>

### Executive summary

The dynamism and breadth of HPC technologies and application development poses a challenge when attempting to create comprehensive training content for new or existing user communities. We have developed a low-overhead training portal with a number of innovative features: a database of sources of training material (collected from web-pages, documents or captured events), automated encapsulation of this content within an on-line training platform, and the possibility to group relevant content together to form training courses for specific topics. A crucial component of the platform is that it leverages HTML5 technologies to enable in-browser access to HPC resources, providing the possibility of performing practical exercises on-line and in real time, directly through a web-browser.

This report only intends to give a brief overview of the portal in it's state at the time of writing. The reader is encouraged to visit the portal under <http://supercomputing.cyi.ac.cy/> to fully enjoy the available portal content.

## 1 Introduction

The purpose of the [Supercomputing Training Portal](http://supercomputing.cyi.ac.cy/)<sup>1</sup>, is to effectively address the educational requirements of a developing HPC community and provide them with the means to independently further their training on-line.

With respect to in-person learning opportunities and the capacity for distance education, the portal provides facilities for the better preparation on the part of the student and more targeted content from the instructor.

The design of the portal separates the collection and cataloguing of training content from the presentation of that content to the end user. In addition, it provides a means of performing hands-on exercises through a web-integrated terminal emulator that connects directly to a true HPC resource.

## 2 Portal Content

Cataloguing available training content is an important first step in creating an effective training resource and many training providers already do this:

- PRACE (Partnership for Advanced Computing in Europe)
  - <http://www.training.prace-ri.eu/>
- XSEDE (Extreme Science and Engineering Discovery Environment)
  - <https://portal.xsede.org/online-training>
- HPC University
  - <http://www.hpcuniversity.org/trainingMaterials>

In our case, typical training content for HPC comes in a variety of forms which is initially classified on the portal as:

<sup>1</sup> <http://supercomputing.cyi.ac.cy/>

- Web-based material, where the content is a (series of) web-page(s) or a web-application
- Captured material, where a lecture (or screen) capture has taken place
- Documents, such as a PDF of a presentation or perhaps a detailed training document

Within the Training Portal, such content is collected, categorized and given a set of meta-data that can be used to manipulate the content when presenting it to the end user.

## **2.1 Capturing Content**

Due to the bleeding-edge nature of HPC technologies and applications, many opportunities for education come in the form of presentations or workshops targeting new developments. In such cases there may not be much additional on-line material. Making this content available to a larger audience requires audio, video and/or screen capture of training events.

The target region of LinkSCEEM suffers from severe bandwidth limitations and so the focus when recording presentations is on high quality audio and high resolution screen capture, since both of these can be heavily compressed. Standard high quality audio equipment is used for audio capture. Capture hardware (of the type typically used for video game capture) is inserted between the presenter laptop and projector, and used to record the presentation screen in high resolution. Even when including video capture of the presenter, such equipment is highly portable and can easily fit in carry-on luggage.

## **2.2 Hands-on Material**

Integrating hands-on material is considered to be an essential feature of effective training. Currently hands-on material relating to training content, when available, is integrated into the HPC resource of the portal and how to find and use it is a component of the associated meta-data in the database.

## **3 Portal Technologies**

Given the dynamism and breadth of HPC technologies and application development, effectively addressing the educational requirements of a developing HPC community is a difficult task. For on-line services, separating the collection and cataloguing of training content from the presentation of that content to the end user can allow independent progress in both areas.

Indeed many web content management systems implement this approach, using a logical storage facility (content repository) for content, metadata, and other information assets that might be needed by the system.

We adopt this approach for the training portal and, from a technical standpoint, decompose features into 4 areas:

- The front-end web portal that gathers information from the database and presents it to the end user
- A database that acts as a content repository for training material
- The in-browser terminal emulator that interfaces with the HPC resource
- The HPC resource itself, where hands-on training activities can be done

### 3.1 Web Portal

[Joomla](#)<sup>2</sup> is currently used as the content management system of the website, primarily controlling the design, presentation, formatting and user-authentication features. A number of additional plug-ins are used to build the desired functionality of the frontend.

The content of the database is formatted in such a way that the front-end makes no distinction between the type of content contained therein. This is achieved by staging the target material through an intermediate formatting page. This formatting page also serves to insert the javascript required for the terminal emulator. By using iframes, the formatting page provides a live connection to any target websites coming from the database.

### 3.2 Content Repository

Currently the content repository database on the live portal is implemented as a back-end Joomla plug-in, with the entries stored in an SQL database. This had allowed for the creation of a user-friendly interface for adding new entries and the associated meta-data.

Entries in the database were categorized and ranked. The categorizations are created and described via Joomla, with the concept of a training course mapping to these categorizations. A PHP script combs the database and automatically creates and/or updates content on the website corresponding to the requested categorizations.

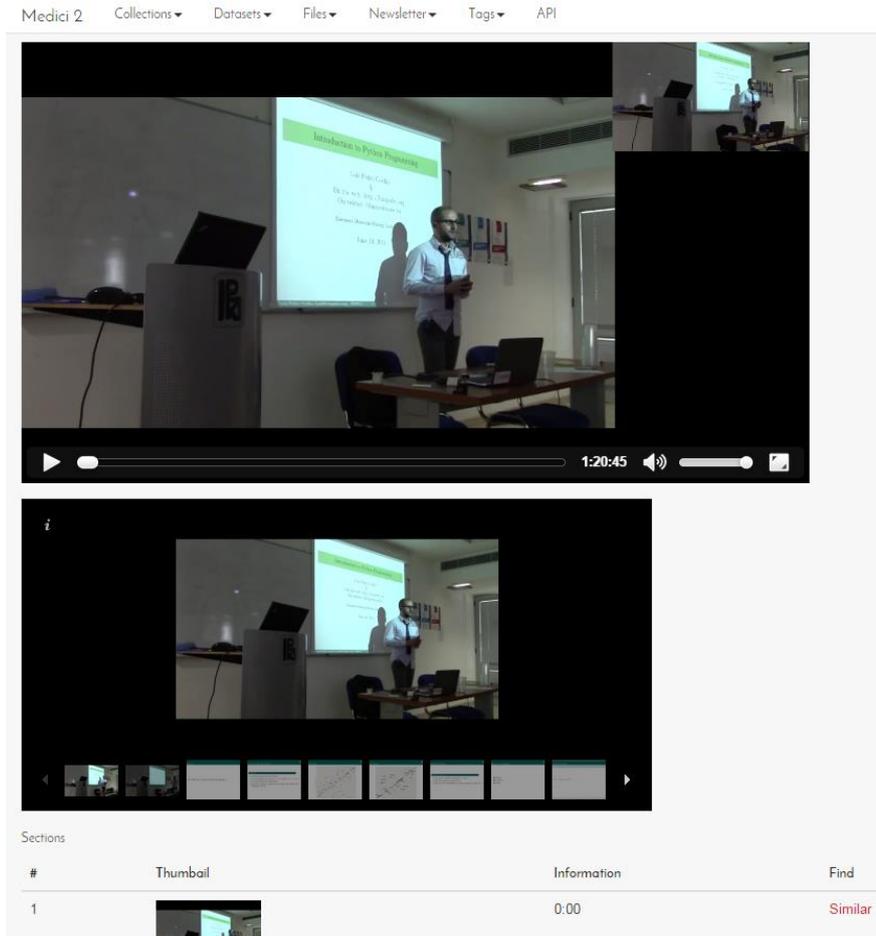
In the ongoing second iteration of this effort, the multimedia content repository, [Medici](#)<sup>3</sup>, provides a feature-rich playground for improving this approach.

Medici is designed to support any data format and multiple research domains and contains three major extension points: pre-processing, processing and previewing. When new data is added to the system, pre-processing is off-loaded to extraction services in charge of extracting appropriate data and metadata. The extraction services attempt to extract information and run pre-processing steps based on the type of the data, for example to create previews. This raw metadata is presented to the user in the Medici web client.

---

<sup>2</sup> <http://www.joomla.org/>

<sup>3</sup> <https://opensource.ncsa.illinois.edu/projects/MMDB>



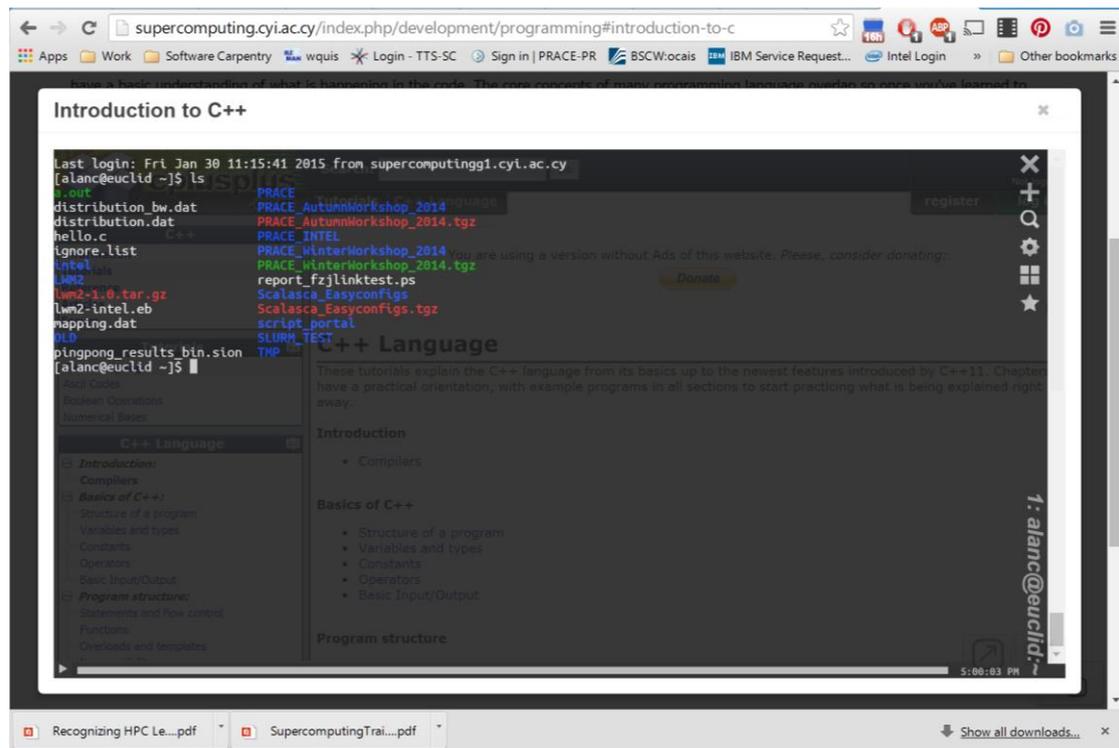
In the image above you can see an example of the result of extraction data on a captured video. Medici is given the screen capture, video and audio as input and combines them to provide the image above. It analyses the screen capture and uses computer vision algorithms to identify slide changes. Once it has the slide changes it creates navigation thumbnails for the video.

The RESTful API that Medici provides also creates an ideal opportunity for collaboration in this cataloguing and curating effort.

### 3.3 Terminal Emulator

The web-based terminal emulator functionality is provided by Gate One<sup>4</sup>, an example screen of which can be seen below.

<sup>4</sup> <https://github.com/liftoff/GateOne>



Its functionality is implemented entirely on the server side and requires no browser plug-ins or client. The SSH capabilities of Gate One are used to directly connect the front-end to the HPC resource without additional authentication on the part of the user.

### 3.4 HPC Resource

The ability to perform integrated hands-on material has been seen to be an essential feature of effective training material. Without this ability, the impact of any content repository will be limited. Given the complex computing environment typically required for HPC training purposes, this can however be very difficult to provide in practice. Adopting common tools and implementation approaches when it comes to practical examples could potentially lead to further fruitful collaborations.

For this reason, a dedicated educational cluster with accelerators is provided by the Cyprus Institute. The cluster uses [EasyBuild](#)<sup>5</sup>, an HPC software build and installation framework which can be leveraged to provide a homogenized execution environment for practical material. It provides the possibility of supporting over 500 software packages (from MPI implementations like OpenMPI to applications such as Gromacs and OpenFOAM). Using the EasyBuild system allows for the possibility of the end-user easily creating exactly the same environment on any other resource.

Another open-source tools that can be leveraged to reduce the barriers to learning caused by HPC-related technical requirements includes Eclipse PTP. It provides an integrated development environment to support the development of parallel applications and includes support for a wide range of batch systems and runtime systems. Its use of a Java runtime environment makes it OS-independent and its support of remote development and execution would make it an ideal hands-on training tool.

<sup>5</sup> <https://github.com/hpcugent/easybuild>

## 4 Available Material

In this section we provide a short description of the material which is currently accessible via the Supercomputing Training Portal. Since the design stage of the portal, it has always been the intention to showcase high quality material, initially this has been done through incorporating content that has been primarily sourced through the training material contained in [XSEDE](#)<sup>6</sup>, supplemented by additional resources from around the internet as well as material from captured LinkSCEEM training events.

The material currently accessed via the portal falls under 4 different categories – most of which are further categorised into more specific subject areas. These are identified in the table below:

Category	Practical HPC	Development	Improvement	Visualisation
Sub-Catogories	Supercomputing Basics	Programming	Performance Analysis	
	Scripting	OpenMP	Code Improvement	
	Version Control	MPI		
	Parallel Tools Platform	GPU		

Within each of these sub-categories a variety of training material can be found. Examples of these include “Python for High Performance” which can be found under Scripting, “MPI Collective Communications” under MPI, “Periscope” and “Scalasca” under Performance Analysis.

Externally sourced content accessible through the Supercomputing Training Portal is linked to the actual source via the user of iframes within the user’s browser. Other content is directly hosted by the Supercomputing Training Portal – such as the various Performance Analysis videos created by VI-HPS.

---

<sup>6</sup> <https://portal.xsede.org/online-training>

## 5 Outlook

The current iteration of the Supercomputing Training Portal serves as a functional proof-of-concept for this methodology of delivering HPC training content. This approach also provides the freedom to create very specific and targeted training programs for particular communities (developers, application users, students, etc.).

Even at this early stage of maturity it has already been used as the main training platform of the LinkSCEEM-2 project and gathered 100 users in the 3 months since its initial public release.

Future development will be done in collaboration with a number of other projects such as the cataloging initiative of the [HPC University](http://www.hpcuniversity.org/)<sup>7</sup>, the [Medici content repository system](https://opensource.ncsa.illinois.edu/projects/MMDB)<sup>8</sup> and the [Parallel Tools Platform IDE](http://eclipse.org/ptp/)<sup>9</sup>. These collaborations will greatly increase the scope, potential, functionality and impact of the portal.

---

<sup>7</sup> <http://www.hpcuniversity.org/>

<sup>8</sup> <https://opensource.ncsa.illinois.edu/projects/MMDB>

<sup>9</sup> <http://eclipse.org/ptp/>