

Grant agreement no. 675451

## CompBioMed

*Research and Innovation Action*

H2020-EINFRA-2015-1

Topic: Centres of Excellence for Computing Applications

# D5.4 Report on Best Practices for e-Infrastructure Application Usage

Work Package: 5

Due date of deliverable: Month 24

Actual submission date: 28 / September / 2018

Start date of project: October, 01 2016

Duration: 36 months

Lead beneficiary for this deliverable: SARA

Contributors: UEDIN, USFD, UNIGE, BSC, UOXF

Project co-funded by the European Commission within the H2020 Programme (2014-2020)		
<b>Dissemination Level</b>		
<b>PU</b>	Public	YES
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	
<b>CI</b>	Classified, as referred to in Commission Decision 2001/844/EC	

## Disclaimer

The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s).

## Table of Contents

1	Version Log .....	3
2	Contributors.....	4
3	Definition and Acronyms .....	5
4	Introduction .....	6
5	The CompBioMed Software Hub .....	8
	5.1.1 The Current Structure of the Software Hub.....	8
	5.1.2 Application Best Practice Guidelines in the Software-Hub .....	9
6	Applications Best Practice Usage.....	10
	6.1.1 Software Description .....	10
	6.1.2 Download and Installation .....	11
	6.1.3 Running the Example .....	11
	6.1.4 Developments and Expert Usage.....	12
7	Conclusions .....	13
8	References .....	14

## List of Tables and Figures

Figure 1 CompBioMed Software Hub applications list as presented on the CompBioMed web site. ....	9
--	---

## 1 Version Log

Version	Date	Released by	Nature of Change
V0.1	5/07/2018	Marco Verdicchio (SARA)	First Draft
V0.5	29/08/2018	Jazmin Aguado-Sierra (BSC), Andrew Narracott (USFD), Charaka Palansuriya (UEDIN), Narges Zarrabi (SARA).	Integrated contributions from co-author.
V1.0	3/9/2018	Marco Verdicchio (SARA).	First draft submitted to internal review.
V1.1	17/9/2018	Gavin Pringle (UEDIN) Alexander Patronis (UCL)	Internal review.
v1.2	28/9/2018	Emily Lumley (UCL) Peter Coveney (UCL) Marco Verdicchio (SARA)	Executive Board review and final version.

## 2 Contributors

---

Name	Institution	Role
Marco Verdicchio	SURFsara	Author
Narges Zarrabi	SURFsara	Author
Charaka Palansuriya	UEDIN	Co-Author
Jonas Latt	UNIGE	Co-Author
Jazmin Aguado Sierra	BSC	Co-Author
Andrew Narracott	USFD	Co-Author
Gavin Pringle	UEDIN	Reviewer
Alexander Patronis	UCL	Reviewer
Emily Lumley	UCL	Reviewer
Peter Coveney	UCL	Reviewer

### 3 Definition and Acronyms

---

Acronyms	Definitions
CMS	Content Management System
CoE	Centre of Excellence
DT	Deep Track
EU	European Union
FT	Fast Track
HPC	High Performance Computing
MS	Milestone
WP	Work Package

## 4 Introduction

---

One of the main foci of the CompBioMed Centre of Excellence (CoE) is to promote computational tools and techniques adopted in biomedicine and develop existing community codes to the extent that they become sustainable and widely used. These software packages are mainly used throughout computational biomedicine research and its applications in healthcare, industry and medical practice. The CompBioMed software stack is composed of a heterogeneous and extended set of applications and computational services which span CompBioMed's three biomedical exemplars: cardiovascular, molecular and neuro-musculoskeletal medicine. The project as structured has developed around two main concepts: the fast track (FT) and the deep track (DT), which apply to all three selected exemplars. The *Fast Track* activities focus on the existing capabilities of the partners in the use of high performance computing for modelling and simulation. The *Deep Track* extends and provides new capabilities to the applications and tools from the FT, addressing more complex requirements that involve a deeper and extended period of development.

During the course of the project, we have worked extensively within the work packages on "Biomedical Research Activities" (WP2), "Resource and Infrastructure Support" (WP5) and "Empowering Biomedical Applications" (WP6) to identify computational services across the biomedical community in order to help users to engage with these solutions. To date, we have produced several reports which are a useful resource of information for an End User who would like to start using HPC-enabled CompBioMed services:

### **D5.2: Report on Computing and Data Needs of the Biomedical Community**

Public report which identifies the needs for infrastructure usage and access models for the biomedicine community

### **D2.1: Report on Application Software Readiness and Fast Track Exploitation**

Public report on the applications and research being undertaken within the CompBioMed environment based on the initial, FT deployment.

### **D6.1: Report on Existing Solutions in Support of Biomedical Applications**

Public document, which describes the existing solutions (tools, services, datasets) that can be of benefit to a range of new biomedical research applications.

These documents describe the computational characteristics (D5.2), usage within research (D2.1), the HPC motivations and the End Users (D6.1) of the main services provided by this CoE.

We have also set up the CompBioMed Software Hub (1), which is one of the main services offered by the CompBioMed project which is accessible through the project website. The Software Hub is the main access point to the resources developed, centralised, is coordinated by CompBioMed and a point of connection with our network of consortium partners for code development and applications deployment tasks.

The consortium is continuously working to promote the usage of these services and to efficiently exploit High Performance Computing (HPC) resources, for both supercomputers and cloud infrastructures. Users wishing to run modelling and simulation codes must, for example, first obtain access and computational time on an HPC or cloud system, familiarise themselves with the peculiarities of the hardware, operating system and environment, compile their codes and

learn how to use the application efficiently on the target resource, which often involves additional costs for specialized services.

In this document we describe the context and how we structure CompBioMed guidance to support best practice for e-infrastructure application usage. The content of the guidelines for different applications is provided through the CompBioMed Software Hub (1). These guidelines focus on the applications and services accessible through CompBioMed, with technical instructions on how to access and run the applications, together with examples coming from our network of developers and users within the consortium. To ensure the information is dynamic and up-to-date, the guidelines produced by the project are included in the CompBioMed Software Hub (1). This will not only guarantee enhanced accessibility but will also allow a better integration with the other resources offered on the website, such as training material and various divulgatory media.

Based also on the inputs from WP2 and WP6 on the CompBioMed applications pipelines and the existing solutions adopted in the CoE, this report contains information describing to End Users how to engage with CompBioMed's tools and services. These guidelines are user driven and based on real biomedical examples with emphasis on HPC-enabled applications and their optimisation and efficient deployment on both cloud and high-performance computing e-infrastructures.

We anticipate that these guidelines will grow and we will keep them up-to-date with the codes and services provided through the CoE. This document provides the basis to rationalise and organise information about the CompBioMed computational services and offer them to end users. The information contained here can be extended and adapted to a specific use case as well as to new use cases that will be added in the future. The maintenance of updated information and the inclusion of new software will be controlled and updated by WP5 in collaboration with the Sustainability Operation Procedure, D4.3 due by March 2019 from WP4, which will describe how CompBioMed monitors current computational biomedical research and employment of emergent technologies, and incorporate any relevant technologies into the Centre.

The definition of the CompBioMed best practices for e-infrastructure application usage is part of the Milestone MS16 for which this report is a mean of verification.

<b>Milestone number</b>	<b>Milestone title</b>	<b>Lead beneficiary</b>	<b>Due Date (in months)</b>	<b>Means of verification</b>
MS16	Definition of e-infrastructure best practices	SARA	24	D5.4

## 5 The CompBioMed Software Hub

As described above, the CompBioMed best practices for e-infrastructure application usage are hosted on the CompBioMed website, and directly link from and integrate with the CompBioMed Software Hub. In this section we describe the current structure and functionalities of the Software Hub and how the best practices for e-infrastructure application usage will be integrated and exposed through the website.

### 5.1.1 The Current Structure of the Software Hub

The CompBioMed Software Hub has been implemented within the CompBioMed website, based on the Wordpress CMS (2). It can be visited by selecting the menu item “CompBioMed Software Hub” under the dropdown menu “Service” in the top bar of the website (3).

The structure and the content of the Software Hub have evolved with multiple additions and new functionalities added during the first two years of our CoE. The current version of the hub includes 17 solutions and is continuously kept updated and extended with new software. Figure 1 shows a snapshot of the Software Hub and the different software on offer.

The CompBioMed Software Hub addresses the needs of two communities:

- the European and wider international computational biomedicine research community, which uses the Software Hub to access the resources developed, centralised, and coordinated by CompBioMed;
- an internal network of CompBioMed partners and selected users invoking the Software Hub to find assistance for software development and deployment tasks within CompBioMed.

The CompBioMed Software Hub does not host any software, but rather provides a link to the relevant software repositories. Popular software repositories like GitHub offer all the desired functionality to host software, therefore there is no need to redevelop a similar infrastructure. In addition, linking to the projects, instead of hosting a copy, guarantees up-to-date version control and avoids redundancy that would be confusing to potential users.

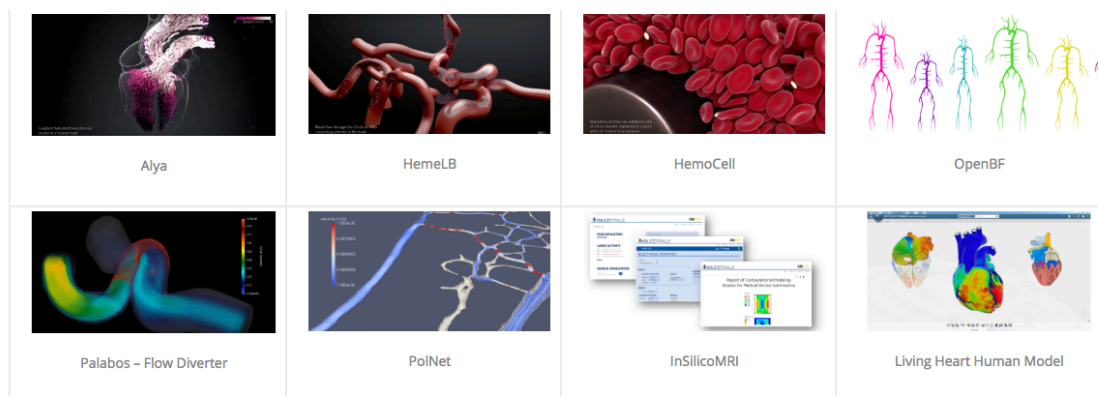
Currently for each of the applications included in the Software Hub we present:

- *Software Description*: high-level description of the software capabilities and its use in the project.
- *Use Scenario*: the type of use of the service (i.e. research, *in silico* trials, industrial, clinical, etc.)
- *HPC Motivations*: motivation for the need of HPC-enabled or Cloud computing solutions and resources.
- *External Links*: links to relevant external references and repositories (developers, website, git, tutorial, training material, etc.).

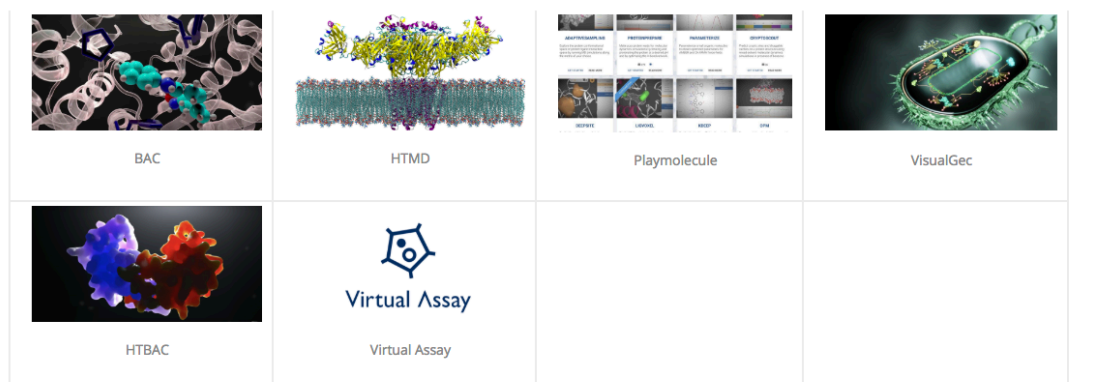
The CompBioMed software Hub has also been included in the list of services defined in the eInfraCentral catalogue (4), which is a central repository for EU digital e-infrastructure services in terms of networking, computing and data management. Synergy with the eInfraCentral EU-project will allow for better visibility and discoverability of the Software Hub and a way to unify and harmonise the exposed services under a common service catalogue framework.



CompBioMed Software: Cardiovascular



CompBioMed Software: Molecular Medicine



CompBioMed Software: Neuro-musculoskeletal Medicine



Figure 1 CompBioMed Software Hub applications list as presented on the CompBioMed web site.

### 5.1.2 Application Best Practice Guidelines in the Software-Hub

As already described above, the best practice guidelines described in this report are integrated with existing CompBioMed services and made available to external users through the Software Hub hosted in the CompBioMed website.

This set up allows us to connect each of the guidelines directly to the related application in the Software Hub and for them to be “live documents” which can be regularly monitored and updated as the codes evolve and/or as new applications are added. Moreover, with an electronic version of such guidelines we can include tutorials and visual help (e.g. asciinema (5)) to extend the document and help the users with connections to related training material and other continuous integration online tools and repositories.

The website permits the hosting of configuration files for containers (such as Singularity (6) and Docker (7)) or automated building scripts and input/output data examples which are essential for the best practice guides.

## 6 Applications Best Practice Usage

---

The CompBioMed Applications Best Practice Usage, provide information on how to obtain, install and run CompBioMed codes to non-expert users, and enable them to easily run on HPC machines and Cloud systems or to use the on-line services provided by the consortium partners.

In this Section we describe the structure of the guidelines, providing, for each section, a reference for the information which will be included. The guidelines comprise four main parts:

- Software Description
- Download and Installation
- Running the Example Provided by CompBioMed
- Developments and Expert Usage

However, not all four components may be suitable for each service/software; for example, download and installation instructions won't be available for online services.

Original installation instructions, along with any other relevant information, provided by developers of the service, are linked to and referenced in the Software Hub. The purpose of these guidelines is to provide a help for end users in adopting CompBioMed services providing biomedical use cases and the latest developments from the consortium partners.

We also provide links and integration of our documentation with other sets of guidelines and best practices as provided by PRACE (8) and other CoEs (e.g. E-CAM (9), BioExcel (10)). The main goal of these guidelines is not to define yet another best practice guide, but to provide insight and expertise from the biomedical domain to End Users and integrate already existing information around CompBioMed applications in a single, central location.

### 6.1.1 Software Description

The "Software Description" Section provides basic information to contextualise the application usage within the computational biomedical domain. This section also contains the application version, license and target e-infrastructures of software/service and a description of the examples provided.

#### 6.1.1.1 Application Information

Description of the software or service, with the following suggested fields:

- *Name*: specific name of the software/service for the associated guidelines.
- *Version*: version or ID number to identify the software/service.
- *License*: type of license for the software or service.
- *Target infrastructure*: infrastructure used by the code and/or type of service access (e.g. online service, on-demand service, HPC application, etc.).
- *Target users*: user or customer which the service targets.

#### 6.1.1.2 CompBioMed Example Description

Provides a brief description and biomedical characterisation of the example(s) provided with the guidelines.

### 6.1.2 Download and Installation

This Section contains technical information on how a user can download and install the software. We describe how to obtain the code and the installation procedure as well as the tools and dependencies required to run the provided CompBioMed example. Note that this section does not apply to online or on-demand services.

#### 6.1.2.1 How to Obtain the Code

- *Type of code*: description of how the code is distributed by the developers (e.g. source code, binary files, etc.).
- *Obtain the code*: description of what registration may be required.
- *Download*: instructions on where and how to get the code and the link to source hosting websites.

#### 6.1.2.2 Installation Instructions

Installation instructions and information on the application dependencies and prerequisites.

- *Installation*: step by step compilation and installation instructions.
- *Prerequisites*: any system prerequisite and OS dependencies.
- *Language*: programming language(s) in which the code is written.
- *External tools*: tools used to help with the compilation and installation.
- *Dependencies*: external dependencies not provided with the code.
- *Testing*: provides tests to check installation procedure.

### 6.1.3 Running the Example

This Section provides information on how to set up and run the example provided by CompBioMed. The section is divided in three parts which explain how to prepare, run and verify the example model.

#### 6.1.3.1 Input Preparation

- *Type of input*: description of file(s) type and input structure.
- *Pre-processing*: description of pre-processing need for the application.

#### 6.1.3.2 Runtime Information

Information on how set up the environment and run the code.

- *Basic run info*: information to launch the program and run a small example.
- *Advanced run info*: technical details for parallel or high-throughput execution.
- *HPC usage*: technical information on how to run on HPC machines or Cloud computing infrastructures.

#### 6.1.3.3 Output Analysis and Post-Processing

Description of the expected output and output files produced (when applicable actual data files to compare example's output will be included). Post-processing instructions with links to external codes needed, datasets or repositories.

#### **6.1.4 Developments and Expert Usage**

This Section focuses on the Application Expert usage and recent developments in three areas of interest for the CompBioMed project: namely data management, software deployment, and workflows.

##### **6.1.4.1 Data Management**

This part contains information on the steps needed for the integration of the compute service with data management tools and systems. We are continuously working to promote EUDAT (11) B2SHARE (12), B2SAFE (13) and B2STAGE (14) services and we will work with the EUDAT project to integrate their policies within CompBioMed best practice guides. In this report we will focus on B2SHARE, B2SAFE and B2STAGE services, which can potentially be integrated into the community codes and applications.

##### **6.1.4.2 Software Deployment**

Description of the efficient porting and deployment of CompBioMed applications on HPC systems and/or Cloud infrastructures. This section contains a technical description about work done within the project on the profiling and optimisation of the biomedical codes as well as on the usage of software build technology and installation frameworks (e.g. Easybuild (15)) and software containers technologies such as Singularity (6), Docker (7), etc.

##### **6.1.4.3 Workflows**

Description of the available technologies (e.g. Taverna (16), MUSCLE2 (17)) designed to facilitate execution and management of workflows, for the orchestration of two or more distinct codes, where the output of one code is the input to the other.

## 7 Conclusions

---

In this Report we have described the content and structure of the CompBioMed best practices for e-infrastructure application usage for the applications we present within our Software Hub. Clearly stated, these are a set of instructions and a repository of information about the main computational services offered by the CompBioMed consortium.

These guidelines are included and integrated with the CompBioMed Software Hub, the central access point for users who want to engage with the applications promoted and/or developed by the CoE. The main goals of the CompBioMed best practices for e-infrastructure application usage is to provide to end users a set of instructions and guidelines to engage and use the computational services offered by the consortium.

The guidelines are designed to inform end users and are heavily focused on examples and real-world use cases. These are heavily integrated with the CompBioMed Software Hub and are flexible enough to be adapted to the different computational services.

The CoE, and WP5 in particular, will continue to monitor and extend the guidelines when new or additional information about a specific service is made available. We also plan to engage in further work to enlarge the Software Hub with more applications coming also from other members of the computational biomedicine community who perform research on other disease cases, and to help developers to integrate their applications into the Software Hub and to align our work with the existing best practice guidelines.

## 8 References

---

1. [www.compbioimed.eu/services/software-hub](http://www.compbioimed.eu/services/software-hub)
2. [www.wordpress.org/ideas/topic/wordpress-cms](http://www.wordpress.org/ideas/topic/wordpress-cms)
3. [www.compbioimed.eu](http://www.compbioimed.eu)
4. [beta.einfracentral.eu/service/compbioimed.compbioimed\\_software\\_hub](http://beta.einfracentral.eu/service/compbioimed.compbioimed_software_hub)
5. [asciinema.org/](http://asciinema.org/)
6. [www.sylabs.io](http://www.sylabs.io)
7. [www.docker.com](http://www.docker.com)
8. [www.prace-ri.eu/best-practice-guides](http://www.prace-ri.eu/best-practice-guides)
9. [github.com/E-CAM/best-practices](https://github.com/E-CAM/best-practices)
10. [bioexcel.eu/software/code-repositories/](http://bioexcel.eu/software/code-repositories/)
11. [www.eudat.eu](http://www.eudat.eu)
12. [www.eudat.eu/services/b2share](http://www.eudat.eu/services/b2share)
13. [www.eudat.eu/services/b2safe](http://www.eudat.eu/services/b2safe)
14. [www.eudat.eu/b2stage](http://www.eudat.eu/b2stage)
15. [easybuild.readthedocs.io/en/latest/](http://easybuild.readthedocs.io/en/latest/)
16. [taverna.incubator.apache.org](http://taverna.incubator.apache.org)
17. [github.com/psnc-apps/muscle2](https://github.com/psnc-apps/muscle2)