**Grant agreement no. 823712**

# CompBioMed2

***Research and Innovation Action***
H2020-INFRAEDI-2018-1
Topic: Centres of Excellence in computing applications

# D5.1 Software Incubation Report

Work Package: 5

Due date of deliverable: Month 14

Actual submission date: 30th November 2020

Start date of project: 01 October 2019 Duration: 48 months

Lead beneficiary for this deliverable: *USFD*
Contributors: *USFD, BSC, UNIBO, UCL, UPF, CBK, UOXF*

**Disclaimer**



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



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

## Table of Contents





*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

# 1 Version Log

| Version | Date | Released by | Nature of Change |
|---|---|---|---|
| V0.1 | 09/09/2020 | Andrew Narracott | USFD Outline document, overall sections defined including contributions from BSC |
| V0.2 | 09/09/2020 | Andrew Narracott | Updated Draft, including contributions from partners |
| V0.3 | 20/10/2020 | Andrew Narracott | Updated Draft, highlighting issues for partner input |
| V0.4 | 29/10/2020 | Andrew Narracott | Partner input integrated and sections reworked |
| V0.5 | 15/11/2020 | Andrew Narracott | Version which addresses reviewer comments |
| V0.6 | 23/11/2020 | Andrew Narracott | Final WP inputs integrated to version for coordinator review |
| V1.0 | 30/11/2020 | Emily Lumley | Final Draft, submitted to the EC |

# 2 Contributors

| Name | Institution | Role |
|---|---|---|
| Andrew Narracott | USFD | Principal Author |
| Daniele Tartarini | USFD | Co-Author |
| Paul Richmond | USFD | Co-Author |
| Alberto Marzo | USFD | Co-Author |
| Mariano Vázquez | BSC | Co-Author |
| Marco Viceconti | UNIBO | Contributor |
| Jon McCullough | UCL | Contributor |
| Adrià Pérez | UPF | Contributor |
| Paul Best | CBK | Contributor |
| Hugh Martin | CBK | Contributor |
| Lei Wang | UOXF | Contributor |
| Blanca Rodriguez | UOXF | Contributor |
| Gavin Pringle | EPCC | Reviewer |



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

| | | |
|---|---|---|
| Gábor Závodszky | UvA | Reviewer |
| Peter Coveney | UCL | Reviewer |
| Emily Lumley | UCL | Reviewer |

# 3 Definition and Acronyms

| Acronyms | Definitions |
|---|---|
| ASME | American Society of Mechanical Engineering |
| BAC | Binding Affinity Calculator |
| BSC | Barcelona Supercomputing Centre |
| CBM | CompBioMed |
| CE Marking | Conformitè Europëenne Marking |
| CI/CD | Continuous Integration/Continuous Development |
| CoE | Centre of Excellence |
| CRO | Contract Research Organisation |
| EEAB | External Expert Advisory board |
| EMA | European Medicines Agency |
| ETP4HPC | European Technology Platform for HPC |
| FDA | Food and Drug Administration |
| GSP | Good Simulation Practice |
| HPC | High Performance Computing |
| HPDA | High Performance Data Analytics |
| HTA | Health Technology Assessment |
| HTC | High Throughput Computing |
| IC | Intellectual Capital |
| IP | Intellectual Property |
| IPR | Intellectual Property Rights |
| IST | *In Silico* Trials |
| KPI | Key Performance Indicator |
| IEEE | Institute of Electrical and Electronics Engineers |
| IEC | International Electrotechnical Commission |
| IST | In Silico Trial |



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

| | |
|---|---|
| ISO | International Organisation for Standardisation |
| MDR | Medical Device Regulation |
| PoP | European Centre of Excellence in HPC for Performance Optimisation and Productivity |
| SaaS | Software as a Service |
| SaMD | Software as Medical Device |
| SME | Small and Medium Enterprise |
| SSI | Software Sustainability Institute, UK |
| TRL | Technology Readiness Level |
| VVUQ | Validation, verification and uncertainty quantification |
| WP | Work Package |



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

# 4 Public Summary

CompBioMed is concerned with the development of computational techniques and research software applications in the computational biomedicine domain, with particular focus on the use of High Performance Computing (HPC) to support this research. The Centre of Excellence (CoE) acts as a hub of best practice for the community, facilitated by a programme of dissemination and engagement with a broad range of stakeholders from academia, clinical practice and industry.

The sustainability of the software and services developed by the CoE is influenced by the ability of end users to exploit the software generated by the CoE. Improving both access to applications and their usability can be achieved through a process of software incubation, informed by end-user needs. Dissemination of best practice in software incubation also has the potential to provide benefit to code developers beyond the CoE partners, for application to other community software tools.

This deliverable presents the structure of the software incubation plan for CompBioMed2, including an overview of activities already undertaken and an outline of activities planned for the remainder of the project.

# 5 Introduction

This document constitutes Deliverable 5.1 Software Incubation Report. Software Incubation has a strong impact on the sustainability of any Centre of Excellence. Incubation activity is focussed on broadening the impact of the computational solutions developed by the CoE and promoting best practices within the biomedical research community. In CompBioMed2, our strategy of Software Incubation is designed to enhance the potential of the CoE to provide services to stakeholders with interests in biomedical HPC applications.

The definition of business models and sustainability plans is a challenging topic for all Centres of Excellence, with a key issue arising from the apparent tension between exploitation of exascale computing infrastructures and commercial sustainability. Therefore, CompBioMed is developing a strategy, discussed primarily in WP4, which addresses a wide range of HPC applications, across a range of infrastructure scales, helping to achieve sustainability while paving the road to exascale. Whilst exascale is targeted as a major objective for CoEs and their deployment of computational codes, we aim to deliver services across a range of scales as computing architectures grow in power. The objectives of this deliverable are to define the incubation plan, describe its components and implementation strategy covering aspects including development of best practice for software documentation and evaluation (regression tests, benchmarking, etc.). This deliverable also reports on activities already undertaken in these areas and outlines future plans to extend this activity.

Finally, it should be noted that this work is informed by and builds upon the foundations established by two deliverables from the first phase of CompBioMed, namely D4.3 Sustainability Plan and D4.5 Report on Pre-commercialisation Activities.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

# 6 Definition of the Incubation Plan

The incubation plan is driven by the guiding objectives of WP5, as defined in the Description of Work, which are summarised in the table below

| OBJECTIVE | ACTIVITIES |
|---|---|
| Prepare CompBioMed2 applications for wider usage and potential commercialisation | Task 5.1: Code review: code stability, development, documentation. Provide guidelines so codes are prepared for "wider usage and potential commercialisation". Review and verify preparedness to a future certification according to technical standard IEC 62304 [1] – "Medical device software — Software life cycle processes". |
| Propose, develop, test, scale, validate and disseminate an incubation model for HPC/HPDA (High Performance Data Analytics) applications within the biomedical domain | Task 5.2: Establish the HPC/HPDA applications developed in CBM1 as a reference for best practices (see 7.3.3). Define the incubation model for new HPC/HPDA applications with specific requirements for development, scaling and validation. |
| Establish a set of commercial services that will underpin the long term sustainability of the centre | Task 5.3: We are currently exploring various revenue streams including consulting, software licensing, software-as-a-service, etc. |
| Establish predictive models based on HPDA techniques that support the development of digital patient technologies | Task 5.4: Assessment of the available models within WP2 applications, described in D2.1 [2]. Application of WP3 data storage, curation and analysis solutions to deploy the HPDA predictive models |
| Support the use of simulation data generated in the CoE by the wider community to be considered in combination with experimental and clinical datasets | All WP5 tasks: To promote exploitation: define policies for dataset interoperability, selecting the proper formats, sources, location, etc. This data could include inputs for simulation codes, output of simulations for clinical analysis, data assimilation and model improvement, anonymised clinical data, experimental measurements, etc. |
| Study containerisation strategies for rapid software deployment in HPC Clouds, assessing parallel efficiency and cost effectiveness. | Task 5.5: Review containerisation technologies maturity as sustainable solutions for reproducible software and deployment in Cloud-HPC.<br>Define recommendations to achieve best performance adopting containerisation solutions.<br>Identify a strategy to support a CoE financially sustainable model providing commercial access to the biomedical applications via a Cloud-HPC for healthcare professionals and SMEs. |

This deliverable describes activities planned to develop best practice in deploying computational biomedicine software tools on HPC platforms, either via traditional HPC or Cloud HPC, and to share the resulting expertise with all members of the community, both inside the CoE and beyond. The activity is also targeted at promoting translation of these tools to clinical practice and commercial exploitation through consideration of target deployment mechanisms during the incubation process.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

We consider software incubation as an integral and fundamental initial element of the exploitation journey that will bring the CoE's software solutions towards end-user fruition. This process can be decomposed to the following elements (software incubation steps in bold). Early stages (steps 1-4) of this process are supported by effort in WP2, WP3 and WP4. The aim of WP5 is to inform these stages from the perspective of future exploitation activities:

1. **Development:** Development of the predictive model from its scientific foundations, conceptualisation, through to its development into a *solution* targeting a well-defined *Context of Use*.

2. **Scalability:** Computational refinement which enables the computational scale required to address the *Context of Use to be reached*.

3. **Validation:** Technical and clinical assessment of the predictive accuracy of the solution. This step also includes assessment of the solution against the requirements of verification, validation and uncertainty quantification to demonstrate application reliability.

4. **Community Engagement:** Early engagement with scientific and end-user communities, for example, clinicians to validate scope of use of the solution within its *Context of Use*. This includes scientific dissemination, solution and data exposure via our webinar series and training courses all aimed at receiving early feedback to inform the development stage and building a community of users that will act as a springboard in the early stages of solution exploitation.

5. **Innovation Impact Assessment:** Assessment of the benefits and impact potential of the solution. This includes the exploitation of the existing IP Register created in phase 1, for the recording of any background/results intellectual property association, an incubator/accelerator register, listing a number of CompBioMed centres dedicated to the exploitation of HPC and related e-infrastructures, and the External Expert Advisory Board (EEAB), which is a group of industrialists, academics and clinical practitioners offering advice and support on the CoE's solutions.

6. **Enhancement of Software/Service or Optimisation:** Preparation of software documentation, user interface development, code hardening, co-design aspects, and cloud deployment to enhance access and engagement by end-users.

7. Accreditation: Depending on the *Context of Use* this might imply certification for SaMD (Software as Medical Device), for IST (In Silico Trial) according to ASME VV-40 [3], or qualification for use in drug trials with European Medical Agency (EMA) or Food and Drug Administration (FDA). NB Pre-clinical applications do not require a regulatory approval.

8. Exploitation: Clear definition of a business case including Business Model, Business Drivers, Financial Metrics, Cash Flow Statement, Costs, Benefits, and Risks.

The first six elements of the exploitation journey described above are described in Sections 7 and 8 of this document whilst Section 9 provides further detail of the last two elements of the process.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

# 7 Description of the Incubation Plan components

This section defines the specific components of the incubation plan and the scope of these activities. The progress to date and future plans in each of these domains is reported in Section 8 of this document.

## 7.1 Innovation Management

WP5, and in particular Task 5.3 ('Preparing Content for External Use and Commercialisation'), will assess the benefits and impact potential of CompBioMed2 solutions under incubation, along with the capacity to exploit them both internally and externally to the CoE. Specific plans for exploitation, commercial or otherwise, will be developed. Mechanisms of exploitation are formulated under the guidance framework defined in D1.4 Innovation Plan (developed by Task 1.7) and this process will be applied to the applications that are incubated in WP5.

WP1 manages CompBioMed's promotion and outreach, together with coordinating and advising application owners across CompBioMed2, via representatives who participate in each of the work packages' regular teleconferences. In WP5's case, the WP1 representative is Hugh Martin (CBK) who is responsible for dissemination activities coordinated by WP1. In terms of promotion and outreach, the application incubation activities from WP5 and the engagement, training and sustainability activities in WP6 will be channelled through WP1.

### 7.1.1 IP Register

In the first phase of CompBioMed, a formal Intellectual Property (IP) register was created, recording background, results (foreground), and related third party intellectual property across the project. The IP Register contains all information regarding the IC (Intellectual Capital)/IP components in the project, with each component defined and detailed within it. The IP Register is available centrally within the CompBioMed intranet. This register supports owners to record candidate results/applications/outputs for incubation and exploitation and promotes the consideration of the IP status as these results evolve. In CompBioMed2, WP5 Task 5.3 will maintain and update the register, where this will be done by UCL with the support of BSC, updating records for current applications, adding information for new solutions that are incubated in WP5 as well as any other new solutions that surface in the project. UCL coordinates the management of all Intellectual Property in CompBioMed2 from Task 5.3.

### 7.1.2 Incubator/Accelerator Register

In the first phase of CompBioMed, a register of incubator/accelerator centres that might be used for our purposes was set up and maintained, and is publicly available at https://www.compbiomed.eu/services/central-incubator-registry/. Incubators are where academic and industry partners collaborate to exploit HPC and associated e-infrastructure by raising awareness in industry, especially in SMEs, making available and providing support for the use of cutting edge HPC facilities. Our incubator register helps in the final stage of the innovation process in supporting the exploitation planning and subsequently, where appropriate, coordinating the entry of innovation candidates into incubator environments that are suited to



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

their specific exploitation needs. These resources will be harnessed as part of Task 5.3's activities.

### 7.1.3 External Expert Advisory Board (EEAB)

The External Expert Advisory Board (EEAB) is a group of individuals representing a selection of our industry partners, academic and clinical practitioners for the purpose of offering advice and support on a wide range of issues relevant to all activities in the project. This Board is chaired by the Project Coordinator with members appointed by the General Assembly. This Board is a reimagining of the Innovation Advisory Board (IAB), created during phase 1 of CompBioMed in order to broaden the scope to include clinicians and other relevant experts that joined during the first phase. One role of the EEAB that continues from phase 1 of CompBioMed is to be an advisory resource for issues that arise in innovation, collaboration, dissemination and exploitation. This board will be asked to advise on the planned incubation activities, offering valuable perspectives from the variety of industry sectors involved. Future members will be appointed as the project evolves ensuring a cohort of innovation focused members within the EEAB along with other expertise required throughout the Centre. Task 1.7 will oversee the nomination of potential members of the EEAB, who will then be formally voted onto the board by the General Assembly.

## 7.2 Stakeholder engagement

Incubation activity is focussed on broadening the impact of the computational solutions developed by the CoE and promoting best practices within the biomedical research community. CompBioMed2 facilitates these interactions through the innovation and community activities of WP6, along with other EU funded projects in this space, for a new initiative called In Silico World. In Silico World is a social network, based on Slack technology [4], and a twin website [5] for the sharing of public contents. The social network provides an online meeting place for the emerging community of practice on *In Silico* Medicine and *In Silico* Trials. The vision of In Silico World is to become the honest broker that enables the Community of Practice of academics, industries, regulators, software developers, CROs, and research hospitals to collaborate in fairness and mutual benefit, and its mission is to drive a wider adoption of *In Silico* Trials in the biomedical industry.

Additional community engagement is achieved through the continued dissemination of training materials through a variety of channels, including the CoE's Training Portal and Repository, the CompBioMed.eu website, our webinar series, and open-source repositories such as GitHub. This is further supported by the establishing of a Helpdesk, to track all support requests from potential clients, associate partners, etc., and the creation of general compbiomed.eu e-mail addresses, both of which are detailed in Deliverable D4.1 [6].

Task T2.4 will feed into the incubation activity of WP5 to facilitate the migration of new research applications and the development of existing research applications from outside the Core Partners to an HPC environment. This includes engagement of Associate Partners and other users identified through In Silico World.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

## 7.3 Enhancement of software/service offering

### 7.3.1 Software application review

Task 5.1 focusses on Code Usability and Documentation Assessment. This activity involves engagement between end-users and code developers to improve the usability of research codes and systematic evaluation of documentation provided to support end-users in deployment and use of the applications. This has been undertaken using internal CompBioMed applications as exemplars of best practice, with guidance on best practice made available to the community through the stakeholder engagement activities. For example, the HPC software Alya has been improved with more validation examples and new functions through a close collaboration between BSC and UOXF. As an end-user, UOXF deployed Alya and performed simulations of the 3D electromechanics of the human heart on different supercomputers, including LRZ SuperMUC-NG and CSCS Piz Daint, and provided feedback to BSC. Opportunities to identify exemplars that highlight innovative aspects of CompBioMed applications (e.g. HPDA techniques and application scaling) will be considered to link this task with other activity in the Work Package.

### 7.3.2 Application hardening

From the viewpoint of WP5, in Task 5.2, we will set guidelines to improve code robustness, showcasing what is done with the CompBioMed software stack (see Appendix A of Deliverable 2.1 First Report on Fast Track Application Readiness [2]). Based on our own extensive experience with our own software, we will define good practices to enhance code robustness and stability, such as CI/CD (Continuous Integration/Continuous Development) integration (from basic ideas to complex pipelines), documentation, regression tests, performance tests, etc. We will showcase software deployment in HPC infrastructures to facilitate code access to users, assessing availability of suitable resources, access mechanisms and performance. We will define policies for dataset interoperability, selecting the proper formats, sources, location, etc. This data could include inputs for simulation codes, output of simulations for clinical analysis, data assimilation and model improvement, anonymised clinical data, experimental measurements, etc.

### 7.3.3 High Performance Data Analytics (HPDA)

Software and services based on HPDA applications have their own unique requirements in terms of development, validation and optimisation, often revolving around data. In WP5, there is a specific task (T5.4) focused on supporting the combination of clinical, experimental and/or HPC generated data with HPDA techniques. Based on our previous experience in the first phase of CompBioMed with such applications, we will develop an incubation model specific for these types of applications, focussing on specific requirements for data management, pre-processing and curation that are critical for developing and testing HPDA applications.
During the first phase of CompBioMed UPF developed several machine learning methods which have been tested, published and made available through Playmolecule. Some methods have even been tested with internal databases from pharma companies like Jansen, Pfizer or Novartis. Further details of these applications are available at https://playmolecule.org/ including Deepsite, KDeep, Bindscope, Delta-Delta, Lig-Voxel, LigDream, LiGANN, and Pathway Map all of


*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

which are HPDA apps developed during the first phase of CompBioMed and being optimised and further disseminated to end-users during this second phase.
This task will leverage the data generated from WP2 related applications and the data management and analysis techniques developed within WP3 in order to help improve the performance and accessibility of HPDA applications, facilitating the access to these applications for both partners and external users.

### 7.3.4 Best practice for application scaling

Based on our experience, covering the full HPC range, CompBioMed will provide expertise on how to improve performance bottlenecks, where possible, so developers can move their codes from the "desktop" to HPC platforms as a first step, devoting development effort when required. WP5 will provide channels to make best practice available to the project's "users".

### 7.3.5 Containers for Cloud-HPC

Cloud-HPC systems decouple containerised applications from traditional HPC systems and provide a method to allocate HPC resources on-demand. Containerised applications can be migrated smoothly among traditional HPC platforms and (public/private) clouds. In this task a set of biomedical codes representing different computational requirements will be containerised and deployed in Cloud-HPC systems. Their computational performance, cost, scalability, efficiency, and adherence to user's quality of service requirements will be studied and reported in D5.5: *Final Report on Applications Stack Service*. A set of guidelines will be produced to improve incubation of future containerised applications, achievement of best performance and long term sustainability.

# 8 Description of implementation strategy

The components of the incubation plan will be monitored through the deliverables of WP5. Activity to month 13 of CompBioMed2 (M13) is reported in this document, D5.1; subsequent activity will be reported in D5.2 (M26), D5.3 (M38), D5.4 (M47) and D5.5 (M48). The sections below describe the progress to date and future plans in each of the domains outlined in Section 7.

## 8.1 Innovation Management

Task 1.7 has delivered the overarching Innovation Plan of phase 2 (D1.4) [7] to assist application owners and service owners across the Work Packages as they develop their assessment and implementation of exploitation plans. This also provides a framework for advice and guidance to application owners as Task 5.3, 'Preparing Content for External Use and Commercialisation', progresses. Newly generated methods/tools/services will be captured by review at monthly WP leader meetings. Where a new element of application software results from project activity, advice will be provided to the owners to support software incubation. Task 1.7 will also configure the composition of the EEAB to provide external advice in response to any unanticipated major issue that may arise from the activities in Task 5.3 that the wider project team cannot address alone.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

## 8.2 Stakeholder engagement

The In Silico World community, started only recently, already has 224 members who are all professionals working in *in silico* medicine in academia, industry, regulatory agencies, contract research organisations, etc.

The community hosts a long list of both public and private channels, each dedicated to the discussion and the exchange of best practices on specific topics. CompBioMed2 works with a number of private channels, used by the consortium, to discuss specific issues, but also operates a public channel called #Scalability where our experts provide advice to all members with problems of scalability of their computational biomedicine codes.

An Application Form has been produced by T2.4 to support users in accessing support from the CoE, with input from WP2 and WP4. This is planned to go live in December 2020. The form will also appear in a reduced format within a web form hosted on the project website.

Due to the coronavirus pandemic, it has been more difficult for CompBioMed members to attend and represent the CoE at scientific conferences. For example, the EuroHPC conference scheduled for March 2020 was postponed until 2021 subject to restrictions. Prior to the cancellation, two CompBioMed members were selected to deliver talks at the conference and these have been re-offered for participation in the 2021 event. The VPH2020 conference in August 2020 was converted to a wholly digital format due to restrictions resulting from the coronavirus pandemic. This meant that a booth was not stationed physically at this conference. Despite this, CBM2 had excellent representation at this conference with several participants contributing posters, oral presentations, a keynote speech and the chairing of sessions.
The CompBioMed webinar series started in phase 1 and, to date, CompBioMed has had 13 webinars, where the last three webinars were delivered within CBM2, on topics related to a high level overview of machine learning, and two examples of deep learning and drug discovery with the most recent webinar discussing some of our contributions towards researching the SARS-CoV-2 coronavirus. At the time of writing, the last three webinars alone have accumulated over 470 views on YouTube. In addition to these a CompBioMed webinar has also been delivered within the POP CoE webinar series discussing CompBioMed2's research on HPC facilities and identifying potential areas of collaboration between these centres. Our webinar series aims to provide two "academic" talks and two "industrial" talks per annum.

Future work will exploit *In Silico World* to disseminate developments of the CoE among domain specialists, demonstrating how the use of HPC can have a profound and positive impact on the sector. Webinar content will continue to be produced that makes our research efforts accessible to the wider public. In addition, Task 1.7 will continuously monitor the engagement with external organisations such as ETP4HPC (European Technology Platform for HPC), EuroHPC and FocusCoE and report on aspects of relevance to the software incubation plan. Finally, to further collaboration, Task 6.5 is creating a directory of external bodies, either EU Projects, user communities, or others, and describes if each Core Partner is an active partner, an unpaid collaborator, or if the body is neither but could offer potential synergies with CompBioMed.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

## 8.3 Enhancement of software/service offering

### 8.3.1 Implementing the software application review

As stated, Task 5.1 focusses on Code Usability and Documentation Assessment. Effort within the first 12 months has focussed on interactions between BSC and UOXF to support development of the Alya code. This has included deployment of the code through collaboration between UOXF and BSC to develop Alya (M1-M12) and interactions with LRZ (M1-M12) and Swiss National Supercomputing Centre (M9-M12) for code installation.

The application focus here is the development of HPC frameworks for multiscale/multiphysics modelling and simulation of human electromechanical activity based on clinical data for specific disease conditions such as post-myocardial infarction and hypertrophic cardiomyopathy. Given the tightly coupled character of the physics and their efficiency requirements towards exascale scenarios, this activity has begun to generate valuable information on best practice for all multiscale/multiphysics code developers and users in the biomedical domain. This involves: 1) collaboration between BSC and UOXF to develop the Alya software considering verification, validation and uncertainty quantification of the models and simulations, based on multimodal clinical datasets, 2) collaboration with supercomputer centres (LRZ and Swiss National Supercomputing Centre) to deploy software such as Alya on supercomputers and test accuracy, efficiency and scalability of this software.

The evaluation of Alya for wider use was implemented with the purpose of enabling simulations of the electromechanical activity of the human heart in healthy and diseased conditions. This was conducted through close collaboration between UOXF and BSC. Initially, we conducted the systematic evaluation of the Alya code, for the specific application considered, in terms of its numerical stability, the features and functionality (e.g. specific cell models) and also the ability of the simulations to reproduce the clinical measurements. Based on this evaluation, work continued to improve the code stability and also its functionality, and this was addressed through co-development between BSC and UOXF. It was critical to setup two meetings per week to discuss these issues and ensure fluid communication. This process is documented in the online discussion space, Basecamp [8]. Basecamp is well suited for collaboration between teams working remotely and allows the discussions relating to this application to be accessed by new collaborators if other core partners are interested. All code developed by UOXF was made available to the BSC team for its incorporation in the main Alya software, in order to make it available to other users, encapsulating when possible respective background IP from BSC and UOXF. The close collaboration, discussion and co-development between BSC and UOXF have contributed to producing a more robust software for human-based cardiac electromechanical simulations consistent with clinical measurements.

End-user requirements have been informed through two meetings per week between BSC and UOXF to discuss the gap between developers and users of Alya, and solve the issues together (M4-M12). The progress on these solutions is archived on our collaborative online space, Basecamp.

Alya has been tested under various conditions to evaluate convergence and efficiency (M7). This work has focussed on development of the latest model for the human cellular electrophysiology and contractility in Alya (M8) and image-based-computational models of the human heart for



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

electromechanical simulations. Calibration and validation of these models has been undertaken with clinical data (M11).

This effort will continue throughout the project and experience with Alya will be used to provide best practice guidance which will be applied to other internal applications (described in D2.1: *First Report on Fast Track Application Readiness* [2]) and made available to external stakeholders as described in Section 8.2.

### 8.3.2 Implementing application hardening

Task 5.2 will start in M13 and focus on application hardening. As stated in the DoA, this task will focus on improving the robustness of codes which aspire to be used either commercially or in the wider community. Activity in WP5 is three-fold. Firstly, we will check the infrastructure already created and maintained by WP4 from an incubation perspective. Secondly, we will assess the infrastructure access conditions and availability granted by the HPC providers. Finally, we will collect and organise code optimisation and VVUQ documents produced in WP2 Task 2.5. Priorities for focus of these activities and dissemination of outcomes will be achieved through the Stakeholder engagement activities described in Section 8.2. Best practice for application scaling will be informed through this activity with selected CompBioMed applications and used to provide more general guidance to external users.

### 8.3.3 Implementing High Performance Data Analytics (HPDA)

Task 5.4 starts in M25 and focusses on applications centred around the integration of HPDA techniques and simulation with clinical and/or experimental data. The first activities of this task will be focused on developing an incubation model specific for these types of applications and provide guidance to external users. We will gather information from the HPDA applications in the CompBioMed software stack and select exemplar cases, from both phases of CompBioMed, to develop the incubation model.
During the first part of the project, several molecular research applications were developed which involved the creation of machine learning based models using extensive experimental datasets. These applications presented unique challenges in their development, validation and deployment. This task will leverage that experience to guide newly developed applications based on HPDA techniques, such as deep learning-based models or gaussian processes. There are also activities currently ongoing in the project related to the development of HPDA applications. In WP2, under Molecularly-based Medicine Exemplar Research (Task 2.2) there is a specific subtask working on combining Molecular Dynamics simulation data with state-of-the-art deep learning models, looking to solve the sampling issues of protein simulation. Activity associated with the development of Alya, described in Section 8.3.1 above also involves collaboration with cardiologists (e.g. Erica Dall'Armellina in Leeds, Rina Ariga in Oxford) and medical image analysis team (e.g. Vicente Grau in UOXF as part of T3.4 and T3.5) to develop HPDA tools for analysing combination of data from clinics and our computer simulations.

These exemplar activities with internal applications will be used to inform best practice guidance for HPDA techniques centred around data availability, curation and pre-processing. A particular focus for Machine Learning applications will be on the assessment of validation and/or reproducibility which imposes specific requirements on the level of information provided to describe the datasets used for research applications.


*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

### 8.3.4 Implementing containers for Cloud-HPC

Task 5.5 starts in M13 and will be initiated through the organisation of a workshop on Cloud-HPC to identify biomedical users' requirements, building on activity undertaken within phase 1. Outcomes of the workshop will include the identification of a subset of applications to be used as exemplar cases to derive a set of guidelines. These guidelines will be formalised as a white paper describing best practice for deployment of biomedical applications to Cloud-HPC. Experience within this task will be used to assist in identification of long-term strategies to deploy and maintain Cloud-HPC applications (e.g. influence on business models for Cloud-enabled services).

# 9 Software incubation: towards sustainability

This Section outlines the role of the incubation process in supporting sustainable business models for future CompBioMed activity.

## 9.1 Selection of the business model

One of the biggest problems of exploiting commercial computational biomedicine solutions, in comparison to other sectors, is that this field is so young and immature that it is still unclear which will be the best business model to provide commercial access to these predictive models. If we look at the early start-ups, we see a range of different business models in play: companies that provide the software as Open Source combined with a consulting service, or companies that sell software as a service directly or via portals, such as insilicotrials.com, or companies that offer a combination of software and consulting, sometimes embedded in a membership fee model where you pay to be part of the club. Then there is the whole world of Original Equipment Manufacturer (OEM) business models where the software is provided to another company that commercialise it as part of a large suite, in combination with some hardware, or framed within specific services.

Thus, at this stage, the choice of the business model is the most critical aspect for the exploitation of a computational biomedicine solution. As part of the best practices to be developed in T5.3 we will propose, in liaison with WP6 where relevant business models will be defined and explored, a systematic approach to this critical initial analysis. This will build upon the business models already described in Deliverable 4.3 Sustainability Plan (not currently published) produced in phase 1 of CompBioMed.

It should be stressed that a possible "business" model for academic solutions is also to pursue a non-commercial route. However, in this case, it is even more critical to define the business model that should ensure the long-term sustainability of that solution, in absence of revenues from commercial operations. So even in this case the definition of a business model is mandatory.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

## 9.2 Identification of the regulatory requirements

Once the business model is chosen, it will be possible to analyse the regulatory requirements. Depending on the contexts of use, the software may require a full CE (Conformitè Europëenne) Marking, a qualification as new methodology with the EMA, or none of the above.

One element that is almost always present, and that is recommended even if imposed by the regulatory requirements, is to establish a quality assurance system for the maintenance and further development of the software. We recommend using as reference the technical standard IEC 62304 – "Medical device software — Software life cycle processes" [1]. The IEC 62304 has provisions more specific to healthcare use, when compared to more general standards like ISO/IEC 90003 [9] or ISO/IEC 15504 [10]. Also, IEC 62304 is required by some regulatory pathways. The most important step is the actual implementation of the IEC 62304 standard in all aspects of the software development process. Once this implementation is performed, the identification of the specific regulatory pathway will suggest if a formal certification according to this standard is required.

If the solution is to be used pre-clinically, it may be possible that no formal certification is required. Still, in most cases the availability of software quality assurance certification is an essential marketing need. A conversation with the prospective customers is indispensable at this stage, to better understand what their expectations in this regard are.

If the solution is to be used to support the diagnosis, prognosis or treatment of individual patients then the software is actually a medical device, and to be sold in Europe it needs the CE mark. The regulatory pathway for software as medical device (SaMD) is now well established, and also reasonably harmonised between Europe and USA, so the FDA certification would not impose completely different requirements from the EMA. The recent ASME VV-40 standard [3] provides a step-by-step guideline for the evaluation of the regulatory credibility of predictive software, and we recommend its use, even in Europe where an equivalent harmonised standard does not exist yet.

In Europe, the new Medical Device Regulation 2017/745 [11], or MDR, will enter in force in 2021. This poses an additional challenge to start-ups, since there is not yet any experience with the application of this new regulatory framework. For sure, the need of clinical trials for many categories of device, may increase considerably the cost of certification, and might impose the need for some venture capital to support this.

The qualification of a computational medicine solution as a new methodology to be used in the assessment of new drugs is the most challenging regulatory pathway (but also the most rewarding commercially). As for the certification under the new MDR the clinical validation is almost unavoidable, and thus the costs of qualification can be important.

## 9.3 Deployment of the solution

Once the business model and the regulatory pathway are defined, there should be enough information to decide what the specifications should be for the deployment of the solution in terms of documentation, user interfaces, testing and bug-tracking, licensing, liability limitation



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

and insurance, accessibility, quality assurance, etc. It may be possible that the deployment specifications for the period of clinical testing will be substantially different from those for the subsequent commercialisation; in this case a double deployment plan is required, which provides separate specifications for clinical testing and subsequent commercialisation.

## 9.4 Best practices for the commercialisation of computational medicine solutions

In order to develop best practices that guide the exploitation of the solutions developed in CompBioMed, we will first systematically analyse two solutions: Alya and BAC that are representative of the spectrum of solutions being developed because Alya is a high-quality example of a monolithic HPC simulation whereas BAC is a high-quality example of a HTC solution. Working with the development teams, and with the start-ups that were created to exploit such solutions, particularly ELEMBio, we will develop a set of draft guidelines to the commercialisation of computational medicine solutions.

These draft guidelines will then be used as a guide toward the exploitation of all other solutions being developed; while many are at a lower TRL than Alya and BAC, the guidelines should help to define the exploitation trajectories, the relative constraints, and guide early decisions about the development and positioning of the solutions. As we apply the draft guidelines to all other solutions, we recognise that some modifications to the guidelines themselves might be required. When we feel that the guidelines are sufficiently established, we will release them as best practices and disseminate them to the community at large, using the In Silico World community as a sounding board.

The Avicenna Alliance (https://avicenna-alliance.com/) is a not-for-profit international organisation that represent all the companies involved with computational medicine either as providers, users, consultants, etc. Recently the alliance established a task force with the mandate to draft a set of recommendations (Good Simulation Practices, GSP) regarding the model development plan as a function of its Context of Use; the model credibility assessment; the regulatory and HTA (Health Technology Assessment) pathways, the ethical review; the role of the modeller, the sponsor, and the protocol; the essential documentation and the reporting. While this effort is specific to *in silico* trials and does not cover digital twin solutions [12], many of the concepts and principles will be similar. Also, the scope of this alliance's task force and the goal of developing exploitation best practices in the CompBioMed project significantly overlap. Leveraging the fact that UNIBO PI, Prof Marco Viceconti, is also the co-chair of the GSP Task Force, we will avoid any effort duplication and promote all possible synergies between the two initiatives.

# 10 Risk Management

| Risk | Level | Mitigation |
|---|---|---|
| Poor engagement with end users with existing pool of solutions | Medium | The consortium will work in concert with the EEAB, to ensure proposed solutions are well aligned with end-users' needs. Requirements will |



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

| | | |
|---|---|---|
| | | be gathered from Core and Associate Partners as well as from external user engagement through the VPH Institute, the Avicenna Alliance and the *In Silico* World community of practice and the CompBioMed conference series. |
| Major unanticipated issues arise associated with preparation of content for external use and commercialisation that the wider project team cannot address. | Low | The EEAB will provide external advice to support the consortium in addressing these issues. |
| Viable business models cannot be identified to support ongoing development of specific codes/solutions. | Medium | Multiple contexts of use will be considered to provide redundancy in targeting of exploitation routes, along with multiple classes of end-users for each application. |
| Insufficient engagement in co-design strategies from computer hardware manufacturers | High | Work with computer hardware design academics, publishing joint papers and being involved in joint initiatives to rise the attention of manufacturers on these issues. |

# 11 Conclusions

This document has defined the nature of the activities that provide the basis for incubation of software applications within CompBioMed2. Progress to date has been reported for tasks which have already commenced, along with future plans for activity which will be reported in later WP5 deliverables. Links to activity in other Work Packages has been described, particularly in the context of the contribution of the incubation activities to the overall sustainability plan for the CoE.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

## 12 Bibliography/References

[1] https://www.iso.org/standard/38421.html

[2] https://www.compbiomed.eu/wp-content/uploads/2020/11/D2.1_FastTrackApplicationReadiness_v1.0.pdf

[3] https://www.asme.org/codes-standards/find-codes-standards/v-v-40-assessing-credibility-computational-modeling-verification-validation-application-medical-devices

[4] https://slack.com/intl/en-gb/

[5] http://insilico.world/

[6] https://www.compbiomed.eu/wp-content/uploads/2020/11/D4.1_CompBioMed2_Compute_and_Data_Services_Strategic_Plan_v1.0.pdf

[7] https://www.compbiomed.eu/wp-content/uploads/2020/06/D1.4_Innovation-Plan_v1.0.pdf

[8] https://basecamp.com/

[9] https://www.iso.org/standard/66240.html

[10] https://www.iso.org/standard/60555.html

[11] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745

[12] Bruynseels, Koen et al. “Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm.” *Frontiers in genetics* vol. 9 31. 13 Feb. 2018, doi:10.3389/fgene.2018.00031



*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **823712**“*