

# CompBioMed

<b>Short description</b>	CompBioMed is a center of excellence (COEs) in computational Biomedicine. The data-intensive workflows and international partners involved in the CompBioMed community urge the use of proper data management solutions for handling large data over distributed sites. We aim to support FAIR research data life cycle, by practicing and implementing data publication workflows, data provenance and curation, and making reproducible data.
<b>Type of community</b>	Others
<b>Community contact</b>	Marco Verdicchio (marco.verdicchio@surfsara.nl), CompbioMed Technical Manager
<b>Interviewer</b>	Narges Zarrabi
<b>Date of interview</b>	
<b>Meetings</b>	
<b>Supporters</b>	Narges Zarrabi

- User stories
- Use cases
- Requirements
  - Technical Requirements
  - Capacity Requirements

## User stories

### Instruction

Requirements are based on a user story, which is an informal, natural language description of one or more features of a software system. User stories are often written from the perspective of an end user or user of a system. Depending on the community, user stories may be written by various stakeholders including clients, users, managers or development team members. They facilitate sensemaking and communication, that is, they help software teams organize their understanding of the system and its context. Please do not confuse user story with system requirements. A user story is an informal description of a feature; a requirement is a formal description of need (See section later).

User stories may follow one of several formats or templates. The most common would be:

"As a <role>, I want <capability> so that <receive benefit>"

"In order to <receive benefit> as a <role>, I want <goal/desire>"

"As <persona>, I want <what?> so that <why?>" where a persona is a fictional stakeholder (e.g. user). A persona may include a name, picture; characteristics, behaviours, attitudes, and a goal which the product should help them achieve.

Example:

"As provider of the Climate gateway I want to empower researchers from academia to interact with datasets stored in the Climate Catalogue, and bring their own applications to analyse this data on remote cloud servers offered via EGI."

No.	User stories
US1	As a researcher (data manager), I need safe data replication and large data transfer between international HPC sites
US2	As a clinician, I need direct access to HPC workflows without exposing the complexity of the underlying HPC environment
...	

## Use cases



### Instruction

A use case is a list of actions or event steps typically defining the interactions between a role (known in the Unified Modeling Language as an actor) and a system to achieve a goal.

Include in this section any diagrams that could facilitate the understanding of the use cases and their relationships.

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	<p>The workflow consists of the following steps:</p> <p><b>Step 1: Data creation and transfer:</b> The raw data is collected at ESRF (European Synchrotron Radiation Facility) in France. The data is being stored locally on tapes. Currently, a copy of the data is transferred to BSC.</p> <p><b>Step 2: Data pre-processing:</b> In BSC, researchers pre-process the data which includes manual and automated steps for image stitching, segmentation and meshing</p> <p><b>Step 3: Data replication:</b> The preprocessed data needs to be replicated from BSC to SURFsara and EPCC. The replicated data will then be used to run simulations with the Alya software which is installed on the supercomputers in these sites (e.g., Cartesius in SURFsara).</p> <p>The HPC centers involved are: BSC (Barcelona supercomputing centre), SURFsara (Netherlands) and EPCC (UK). Therefore, we needed at least 24 TB storage allocated at each of the HPC centers involved.</p>	<p>The diagram illustrates the data replication process. On the left, a box labeled 'BSC Barcelona Supercomputing Center' contains a server icon and the text 'Preprocessing data &amp; Compute'. Two blue arrows labeled 'Replicating data' point from this box to two boxes on the right. The top box is labeled 'SURF SARA' and contains a server icon and the text 'Archiving and Compute'. The bottom box is labeled 'epcc' and contains a server icon and the text 'Compute'.</p>

UC2	<p>In the CompBioMed community, there is a need to provide direct access to HPC workflows to clinical end-users without exposing the complexity of the underlying HPC environment. The Computed Tomography to Strength (CT2S) use case uses HPC approaches to provide quantitative metrics of bone strength based on CT images. In order to deliver the workflow directly to the clinical end-users, a web-service approach is used to send anonymized data and meta-data to an HPC cluster for pre-processing and computations on data. Based on the analysis a report will be sent back to the clinicians.</p> <p>As a follow-up, we would like to make a detailed description of the data transfer requirements and metadata scheme that will provide a solid-basis towards the re-creation of the same workflow in other non-UK settings (for example using EUDAT or EOSC services).</p> <p><b>User group:</b> Clinical users</p> <p><b>Data type:</b> Clinical data (sensitive data that is being gathered or used by clinicians in hospitals for clinical trials)</p>	
...		


## Requirements

### Technical Requirements



#### Instruction

- Requirement number: Use numbers RQ1, RQ2, RQ3, ...
- Requirement title: Use a short but descriptive title. Use the same title in the Jira ticket 'Summary' field
- Link to requirement JIRA ticket: Open a ticket in <this JIRA queue <https://jira.eosc-hub.eu/projects/EOSCWP10/issues/EOSCWP10-4?filter=allopenissues>> (click on 'CREATE' button in the middle-top of JIRA)
- Source use case: Refer back to the use cases above (UC1, 2, ...)

Requirement number	Requirement title	Link to Requirement JIRA ticket	Source Use Case
Example	EOSC-hub to provide an FTS data transfer service	 EOSCWP 10-21 - Jira .	UC1
RQ1			
RQ2			

## Capacity Requirements

EOSC-hub services	Amount of requested resources	Time period