Funded by the European Union

**Call: HORIZON-CL5-2021-D5-01**

# Hyperconnected simulation ecosystem supporting probabilistic design and predictive manufacturing of next generation aircraft structures

**CAELESTIS**

## Deliverable D2.1

CAELESTIS Interoperable Simulation Ecosystem architecture and PRESS analysis

## Work Package 2

HPC digital ecosystem and extended enterprise context

| | | |
|---|---|---|
| **Document type** | : | Other |
| **Version** | : | 1.0 |
| **Date of issue** | : | 30/11/2023 |
| **Dissemination level** | : | PUBLIC |
| **Lead beneficiary** | : | BSC |

***Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN CLIMATE, INFRASTRUCTURE AND ENVIRONMENT EXECUTIVE AGENCY (CINEA)]. Neither the European Union nor the granting authority can be held responsible for them.***

Funded by the European Union

Funded by the European Union

# Executive Summary

| | |
|---|---|
| **Abstract** | This deliverable reports the results of the first phase of WP2. It includes the definition of the architecture for the CAELESTIS Interoperable Simulation Ecosystem (ISE) which is the main outcome of Task 2.1, the mechanism for integrating several simulation software and managing the digital thread which is the main outcome of Task 2.2. As part of Task 2.3, we also provide the implementations details of the first prototypes for Hybrid Twin Platform and the HPC Simulation Service as the main components of the CAELESTIS ISE. Finally, we include cybersecurity analysis and recommendations for the CAELESTIS ecosystem as result of Task 2.5. |
| **Keywords** | Interoperable, simulations, software |

Funded by the European Union

# Revision history

| Version | Author(s) | Changes | Date |
|---|---|---|---|
| 0.1 | Jorge Ejarque (BSC) | Table of Contents with assignments | 21/09/2023 |
| 0.2 | Santiago Montagud (ESI) | Sections 3 and 4 | 05/10/2023 |
| 0.3 | Giotta Lilli (EBOS) | Section 6 | 18/10/2023 |
| 0.4 | Jorge Ejarque (BSC) | Section 1, 2, 7 and 8 and summary figure of Section 6 | 26/10/2023 |
| 0.5 | Riccardo Cecco (BSC) | Section 5 | 26/10/2023 |
| 0.6 | Jorge Ejarque (BSC) and Santiago Montagud (ESI) | Format edits before internal review | 30/10/2023 |
| 0.7 | Jorge Ejarque (BSC) and Santiago Montagud (ESI) | Applying review comments | 21/11/2023 |
| 1.0 | Jorge Ejarque (BSC) | Final version | 28/11/2023 |

Funded by the European Union

# TABLE OF CONTENTS

Funded by the European Union

# 1 INTRODUCTION

CAELESTIS aims at developing a robust and secure digital simulation-driven ecosystem for designing manufacturing methodology that will efficiently support the transformation of the European aircraft industry. This transformation will be supported by enabling massive exploration of the design and manufacturing space to find new and innovative aerostructures and related systems. The proposed ecosystem focuses on the integration of the different actors involved in the design and manufacturing process with the High-Performance Computing (HPC) systems where these massive simulations are performed.

At design time, the different Distributed Engineering Teams (DET) across aircraft industry will be integrated with the HPC environments to enable complex product and process multiscale and multiphysics simulation workflows. These workflows will deliver fast and accurate predictive insights for many design and manufacturing scenarios, providing a wide range of outputs on mechanical performance, manufacturability, sensitivity analysis and uncertainty quantification. During this phase, and thanks to the massive simulations at HPC, reduced order models will be created to help as manufacturing decision support and transferred to the correspondent edge devices.

At manufacturing, data will be transferred from measurement tools to the edge devices. This data will feed the reduced order models enabling real-time predictions at the shopfloor based on eventual detected defects and their impact on product or process performance. Such knowledge will be used to feed a decision support system to detect if the obtained product will face non-conformance issues and manufacturing corrective actions are required.

In this deliverable, we present the results of the first part of the WP2 work. First, we present the architecture of the CAELESTIS Interoperable Simulation Ecosystem (ISE) which proposes a software solution to manage the execution of workflows for automating the simulation of the different design scenarios. It involves the integration of the different simulation software with the data processing algorithms to analyze the simulation results and create the desired reduced order models. This work is the continuation of task 1.2 where we produce an initial specification of the ecosystem. Apart from the architecture, we present in the implementation details of the main software components of the architecture: Hybrid Twin Platform (HTP) and

Funded by the European Union

HPC Simulation Service; and the mechanisms to integrate different simulation software and keeping track of the digital thread. Finally, we include the results of the analysis of the cybersecurity treats, requirements and recommendations to consider in the proposed ecosystem.

The rest of the deliverable is organized as follows: Section 2 describes the overall architecture of the CAELESTIS ISE; Section 3 presents the mechanism for managing the interoperability between simulation software interoperability and the digital thread; Section 4 and 5 provide the implementation details of the Hybrid Twin Platform and HPC Simulation Service; Section 6 presents the cybersecurity considerations analysis and recommendations; and finally, Section 7 provide the status of the implementation and links where the source code of the implemented Software Components is located.

# 2 CAELESTIS INTEROPERABLE SIMULATION ECOSYSTEM ARCHITECTURE

To evaluate the suitability of different design or manufacturing parameters, engineers require to perform several simulations where values of different parameters are changed to predict their effect in the final product. These simulations can require the usage of different software to simulate the different steps of the manufacturing process and some of these simulations require large computations which can take a lot of time in commodity hardware, but they can be accelerated using HPC clusters. However, the integration of the different software and the usage of HPC system require different specialized scripts and tools which are not common in aircraft manufacturing engineers and learning them can take time.

The CAELESTIS Interoperable Simulation Ecosystem is a software platform that aims at facilitating the massive explorations of the design and manufacturing space. In this ecosystem, the massive explorations are defined as workflows that combine the executions of one or several simulation software, depending on the manufacturing process to evaluate with data processing algorithm which analyses the simulation data according to the analysis to perform. Figure 1 provides an overview of the architecture of this software ecosystem which depicts the different user roles and components of the system and the main interactions to perform the execution of the massive simulation workflows. When engineers want to perform an analysis, they choose a predefined workflow to perform this analysis. The procedure consists

Funded by the European Union

of the selection of a workflow template which is suitable for the desired analysis and customizing it with the simulation software invocations and data processing actions required for the manufacturing process to evaluate. The implementations of the possible workflow templates are available in the CAELESTIS repositories. Once the workflows are defined, it is submitted into the system for execution. Since not all the software is suitable to be executed in the HPC environments, or does not require large computing resources, the workflow is divided into two levels, the DET level and the HPC level. The DET workflow, which includes the light invocations of external software or services is managed by the Hybrid Twin Platform (HTP) that will outsource to the HPC Simulation Service the part of the workflow which requires large computations. Data exchanges between the different components are performed by means of the Storage Service. The following paragraphs provide more details about user roles, components, and the required interactions to execute the massive simulation workflows.

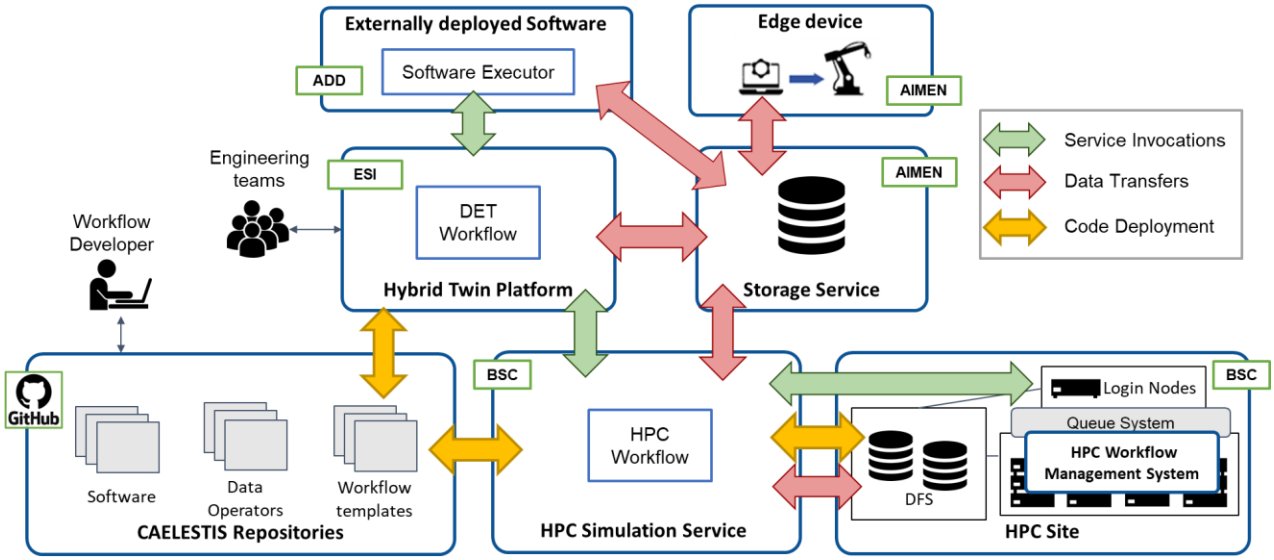


*Figure 1. Overview of the CAELESTIS Interoperable Simulation Ecosystem.*

## 2.1 User Roles

Two users' roles are expected to interact with the CAELESTIS ISE:

- **Workflow developer:** These users are in charge of managing the implementation of the workflow templates, data processing algorithms and software invocation required to perform the different analysis requested by the engineering teams. These users are

Funded by the European Union

mainly interacting with the CAELESTIS repositories where they store the different implementations.

- **Engineering team user**: These users are in charge of describing, submitting and supervising the workflows and their results. These users mainly interact with the Hybrid Twin Platform.

## 2.2 Main Components

The CAELESTIS Interoperable Simulation Ecosystem is composed by the following components:

- **CAELESTIS repositories:** This is a Git repository where workflow developers manage the different versions of the code of the workflow templates, the invocation of the different software and the data processing algorithms. These codes are downloaded by the HTP and the HPC Simulation Service to be executed during the workflow execution.
- **Hybrid Twin platform:** This is the main component for Distributed Engineering Teams to interact with the CAELESTIS ISE. It provides the functionalities to define the simulation workflows, orchestrates its execution at DET level invoking the external software and outsourcing the HPC part of the workflow to the HPC Simulation Service, and manages the workflow results to create Reduce Order Models to be used by the Edge Devices.
- **HPC Simulation Service:** This component is in charge of managing the execution of the workflow in the HPC Site. It will receive the workflow description from the HTP and according to this description, it will deploy the workflow code, transfer the input data to the HPC site, submit and monitor the workflow execution, and transfer the workflow results to the storage service.
- **Externally Deployed Software:** This component enables the execution of an external software or service from the HTP. This implements the required execution interface and manages the data staging with the Storage Service including downloading the input data for the software execution and the uploading the results.

Funded by the European Union

- **Storage Service:** It is used to store the input and output data of the workflows as well as other data that must be exchanged between the different components of the CAELESTIS ISE.
- **Edge Device:** This component is installed on the factory premises, and it implements two main features: providing the sensor's data required for the creation and validation of the reduced order models and making use of the generated models to take decisions during the manufacturing process**.**
- **HPC Site:** This component executes the massive simulations included in the workflow. It is composed by a login node, where users can deploy their codes and data, a set of compute nodes where the simulations and data processing algorithms are executed, a Distributed File System which will allow to access the data from the login nodes and all the computing nodes. The access to the computing nodes is done by means of a queue system which will schedule the executions of the different users in the available computing nodes. The orchestration of the workflow execution in the HPC site is performed by the Workflow Management System.

## 2.3 Workflow Definition

During the first month of the WP2, we have evaluated the different analysis and computations to be performed in CAELESTIS in collaboration with WP3 and WP4. We have seen that generating a workflow for each of these analyses will require a lot of effort for engineers and it is tedious and error prone. However, we have identified common patterns in the workflows that are related to the type of analysis that the engineers want to perform, and some parts that change depending on the manufacturing process to evaluate. Figure 2 shows an example of this pattern. In this figure, we can see a workflow template for a surrogate model creation, where we have several phases (in blue) that are mainly a sampling phase to generate the cases to simulate a simulation phase which includes the preparation, simulation and post process phases and a training and validation phases. The workflow structure and some of the phases will remain the same if we want to create a model for permeability, RTM or mechanical performance. However, some other parts (in green) are the ones that change depending on the process to evaluate.

Funded by the European Union

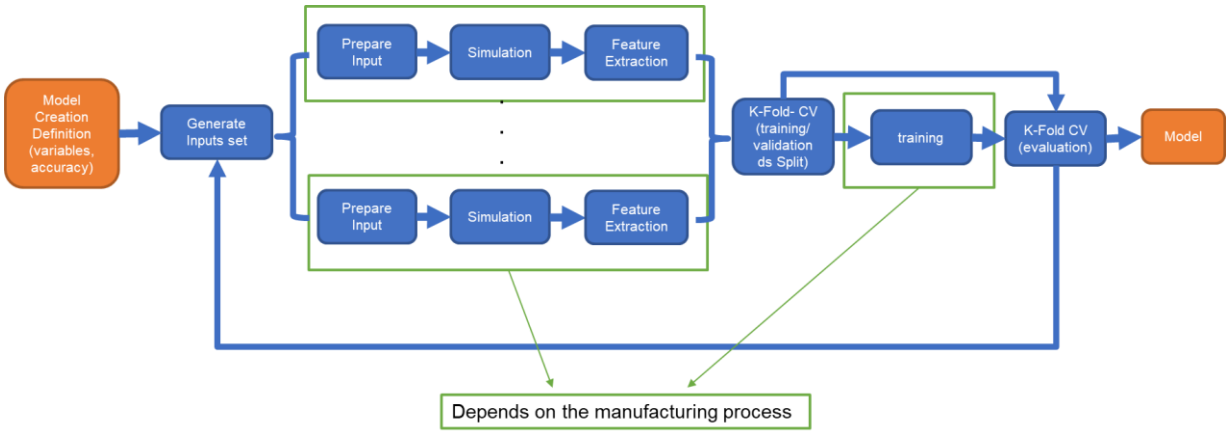


*Figure 2. Example of a Surrogate model creation workflow.*

Based on this evaluation, we propose to define the workflow in the CAELESTIS ISE as depicted in Figure 3. A workflow is defined as a template where the different phases are customized with different software and data processing implementations which will be selected according to the analysis to perform. These templates and phases will be linked to the implementation stored in the CAELESTIS Repositories. These repositories are Git repositories where they can store different versions of a workflow template or software code. For every code in a git repository, there is a default which indicates the latest stable version. Workflow Developers can implement new versions of these codes which will be stored in different branches of the Git repository. If users want to use a specific version of workflow template or software, they have to specify it when defining the workflow type and phases. In addition to the customized template, the workflow definition also includes the location of the necessary input files and expected location for the output files in the Storage Service. This is required to enable the automation of the data movements.

Funded by the European Union

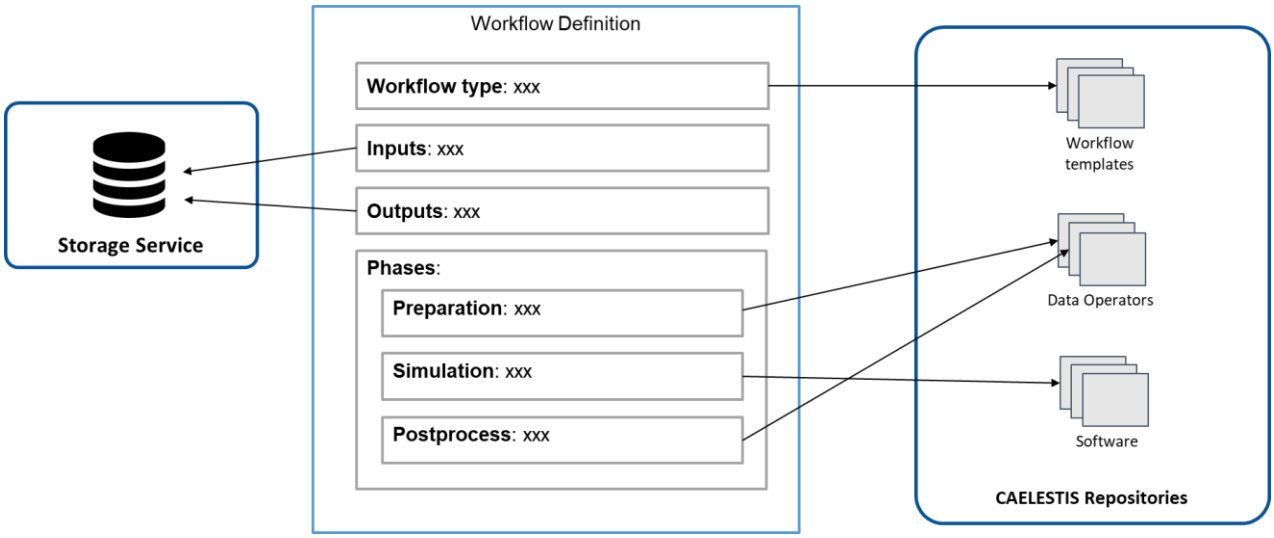


*Figure 3. Workflow definition overview.*

## 2.4 DET Workflow Management

This section provides more details about how the massive simulation workflow is managed at the DET level by the Hybrid Twin Platform. This management system is depicted in Figure 4. In this figure, we can see that DET engineers can define the workflow as explained above using the HTP features for workflow definition. Once it is defined, it is also orchestrated by the DET performing the invocation of the software at DET level and outsourcing the HPC part of the workflow to the HPC Simulation Service. To do the DET software invocation, the HTP performs a remote call to the external Software Executor with the details of the execution to do, including the location where the input data is stored in the Storage Service and the location where to store the outputs in the storage service once the execution is finished. With this information, the Software executor script downloads the data, performs the execution, and stores the results. In the case of the HPC workflow, it extracts the HPC part of the workflow description and calls the HPC Simulation Service passing this description. As we will see in the next section, this service will execute the HPC part of the workflow uploading its results to the Storage Service. Finally, DET Engineers can also use the ROM/ML tools from the HTP to create new models from the workflow results and sensors data. These models will also be stored in the Storage Service and will be used by the Edge devices to do fast inferences at manufacturing time or used in other workflows after being deployed as external software or at the HPC site.

Funded by the European Union

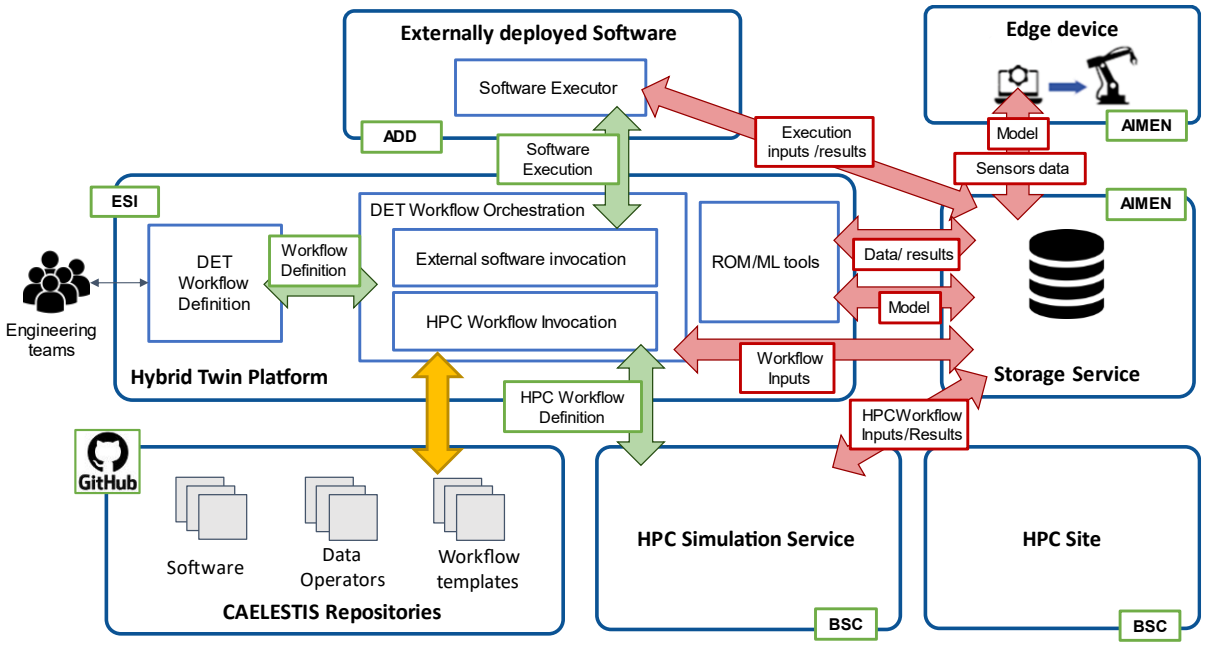


*Figure 4. Workflow Management at DET level.*

## 2.5 HPC Workflow Management

Figure 5 show the details about how the massive simulation workflow is managed by the Simulation Service at HPC level. The description of the workflow to execute at HPC level is submitted to the HPC Simulation Service by the HTP. It parses the description and creates the HPC workflow by downloading the workflow template associated to the workflow type and the required software and data processing algorithms which are defined in the workflow definition phases. All this code is deployed to the Distributed File System (DFS) of the HPC site. A similar procedure is done with the input data, the HPC Simulation Service gets the location of the input data from the definition and downloads it from the Storage service and stores it in the HPC site DFS. Once the deployment and data stage-in phases are complete the HPC Simulation Service connects to the HPC site login node to submit the execution of the workflow using the HPC Workflow Management System. This component orchestrates the execution of the different computations of the workflow inside the HPC Site, keeping track of the execution and doing periodical checkpoints to enable the restart of the computation in case of a failure. During the execution, the Simulation Service is monitoring the execution and

Funded by the European Union

detects if it has finished correctly, failed or the computation time is exhausted. If the execution finishes correctly, it uploads the results to the Storage Service and if in the other cases it allows the automatic resubmission to restart the workflow from a checkpoint.

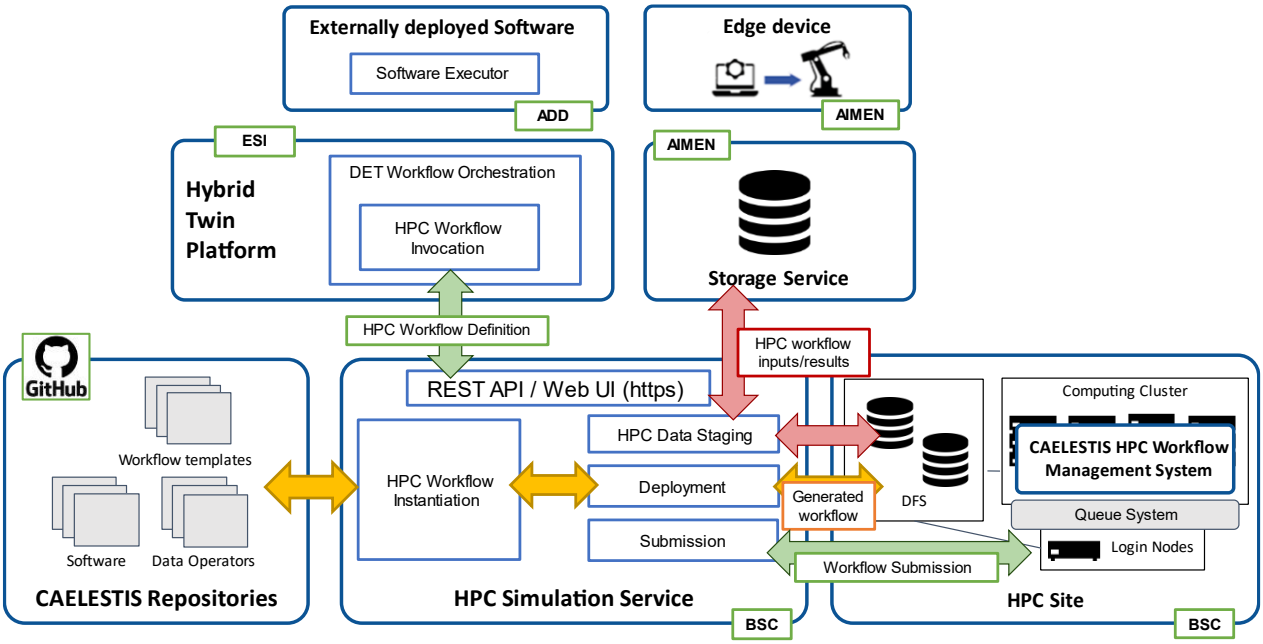


*Figure 5. Workflow execution at HPC.*

Funded by the European Union

# 3 SIMULATION SOFTWARE INTEROPERABILITY AND CAELESTIS DIGITAL THREAD

The manufacturing process involves different steps as shown in Figure 6. For each step in the process, there is at least one model presented in the simulations chain (Figure 7). To massively explore the design possibilities, manufacturing defects propagation, and their impact on the final properties, it is required to automate the transfer of data between models and simulations; otherwise, the time and cost to prepare the amount of simulations which are required would be unaffordable. The seamless data transfer between each model is achieved by appropriate translation mechanisms developed by the partners. A detailed explanation of each software and model is present on deliverable D1.2, as well as a detailed list of their correspondent inputs and outputs.

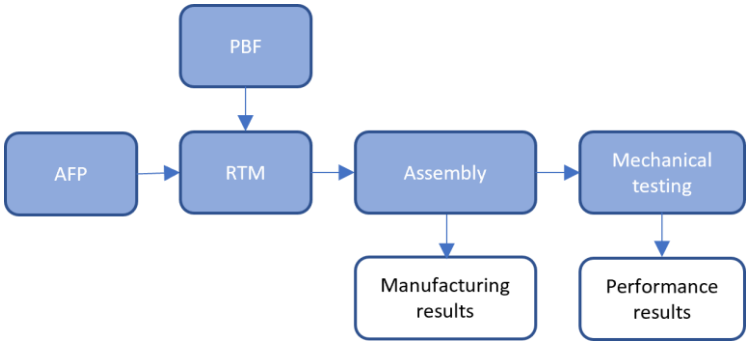


*Figure 6. Manufacturing Process*

Funded by the European Union

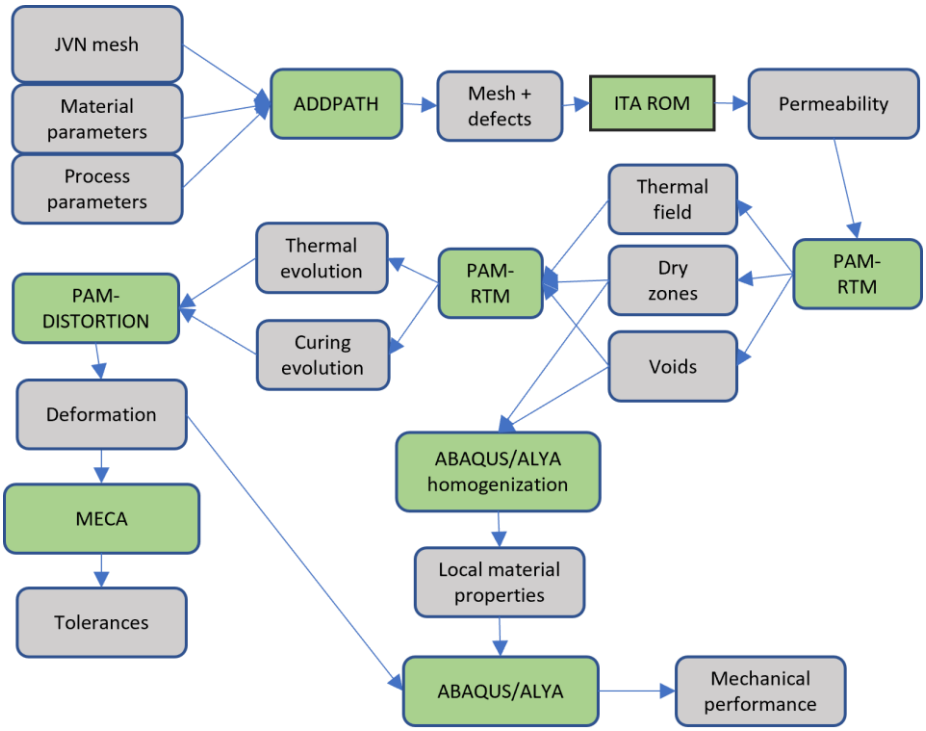


*Figure 7. Simulations chain*

A methodology has been designed to achieve the automatic transfer of data, which contains the following main phases:

- Consideration of simple models to simplify the development phase and its validation. Three models are considered, where the complexity of the model is increased at each stage.
- Identification of the required inputs and available outputs of each simulation or model.
- Discussion and development of the best method to translate the information between models.
- Testing of the solution.

## 1. Models

The models specification will be part of corresponding deliverables from Work Packages 3 and 4. Figure 8 shows only a brief description of their content that facilitates the comprehension of the methodology related to this section.

Funded by the European Union

1.a) POC1 considers a flat rectangular part with no defects.

1.b) POC2 considers a curved rectangular part with no defects.

1.c) OGV considers the industrial curved part with defects.

Variants of these three models could exist depending on the evolution of the developments. I.e., defects can be introduced one by one: gaps, overlaps, ply misalignment, etc.

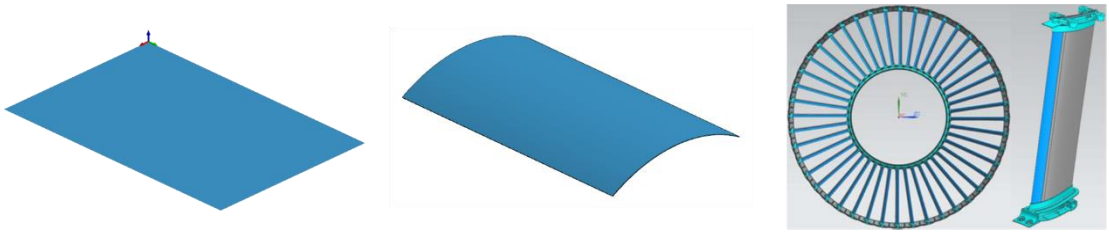

*Figure 8. Increase of Complexity of the part models.*

## 2. Data

Each model owner has listed the relevant inputs required to run the simulation and the available outputs. In addition, the available formats have been also provided. This detailed information has three main purposes:

a) Provide a complete understanding of the process data flow. This is an important aspect for complex projects as CAELESTIS where experts from different domains collaborate.

b) Identify the data that can be transferred and its format. Different physics and scales are mixed through the data flow along the simulation chain. Material properties need to be characterized to serve as inputs for the models. Some parameters could not be available on the laboratory and alternative options should be discussed. Early identification of this issues reduces future complications.

c) Prepare a list of parameters that will be available for design of experiments on the virtual domain, as well as unique identifiers. Simulation chain automatization is key for the success of the project. End user needs to know which parameters are exposed by the models to be varied for its desired purpose, and these parameters should be clearly distinguished. RTM process temperatures could refer to injection temperature or

Funded by the European Union

curing temperature. The objective is to avoid this kind of misunderstanding. In addition, software needs to modify the parameter values without human intervention. For that end, a list containing the unique identifier per parameter will be stored and used to refer to each of them.

The complete list of identified parameters considered as workflow variables has been collected and included in Table 1. The final list will detail the parameter, its unique identifier, the creator of the data, the receiver or receivers, the units and the format.

*Table 1. Parameters to be exchanged in the simulations chain.*

| Exchanged Parameters | |
|---|---|
| CAD Geometry | Filling factor |
| Tetra mesh | Resin injection volume |
| Hexa mesh | Temperature field |
| Mesh centroids | Heat flux |
| Number of Layers | Flow front velocity |
| Material per layer | Dried zones |
| Orientation per layer | Voids content field |
| Thickness per layer | Voids |
| Fiber fraction volume per layer | Temperature field after curing |
| Robot Trajectory | Degree of cure field |
| Robot Speed | Cure rate field |
| Preform properties | Stresses field during cure |
| Resin properties | Residual stresses field after demolding |
| Injection pressure | Deformation field during cure |
| Injection temperature | Deformation field after demolding |
| Injection flow rate | Shape distortion |
| Injection mold temperature | Temperature field distortion |
| Injection points location | Distorted mesh |
| Injection open/close sequence | Tolerance parameters |
| Injection vacuum pressure | Material properties with defects |

Funded by the European Union

| | |
|---|---|
| Injection vent open/close sequence | Failure test loading |
| Curing cycle | Fatigue loading |
| Laminate with defects | Max stress |
| Permeability field | Max strain |
| Resin pressure field | Evolution of the stress field |
| Resin velocity field | Evolution of the strain field |
| Filling time | |

Among the previous parameters, some of them will be exposed by the different software as variables, and some of them will be constant or not exposed to be modified by the user. Orientation of layers, number of layers or injection pressure are some of the parameters that will be exposed as variable parameters to be modified by the user. Other parameters, as distorted mesh, are not to be modified directly by the user, but they will be modified by the process as a consequence of other parameters variation. The exposure of each of the parameters is being analysed during the project based on key factors as its impact on other parameters or the complexity of the implementation algorithm. For instance, it has been decided to keep the mesh as constant. Automation of meshing is a complex task that leads in most cases to unacceptable meshes requiring human intervention to manually fix the errors. As a consequence of this approach, the impact of the size of the mesh will not be under analysis, and its influence on parameters like permeability or voids is at this moment outside the scope of the project.

## 3. Translation methods

Translation methods deal with the conversion of the data from one model to another. In CAELESTIS, for most of the data transfers that have been identified an automatic import/export method does not exist. To develop the translation methods, each exchange has been treated separately with dedicated sessions between the relevant partners. When a standard existed, it has been used to as the exchange data format. This is the case of the mesh between PAM-RTM and ABAQUS/ALYA, i.e. When there was not an existing format, the

Funded by the European Union

appropriate method has been discussed and developed. This is the case of the mesh with laminate information and defects provided by ADD to ITA in order to consider AFP manufacturing effects on the permeability. The use of the simplified models previously related will help on validating the developed translation methods.

There is remaining the transfer of some material parameters that remain complex in industrial environments and could increase the already complexity of the project. Distortion induced residual stress transfer between different software, for instance, is under analysis.

## 4. Testing of the solution

POC1 has already been used as a test for automatic model updating and data transfer. The process for each software for each exploring campaign can be resumed as:

- Modify the model according to eventual new values of the inputs included in the DoE.
- Import inputs from previous simulations.
- Run simulation.
- Export results to the correspondent format.

This has been tested at a small scale between partners, and once it is validated, it will be tested at a more complex scenario in the HPC.

Funded by the European Union

# 4 HYBRID TWIN PLATFORM

Industry 4.0 recent developments allow to integrate the physical and virtual worlds. Real data can be more easily considered as inputs for virtual models, leading to a large amount of work around the concept of Digital Twins. There is then a need for structures that allow both physical and virtual worlds to merge. The idea behind the ESI's Hybrid Twin Platform is to put in place a platform where data can be merged, formatted, analysed, and used to extract valuable information.

From design to production, several different teams contribute to the process, the so-called distributed engineering teams. Data is generated at each step of the process and shared between teams. Industry 4.0 deals with the seamless exchange of information between different teams or companies. This comprises the development of technologies around software architecture, security and standards for communication or information modelling. One of those standards for information modelling, AutomationML [1], is being tested in CAELESTIS. This provides means to keep the traceability of the data: who has generated each data and when, and by consequence improving the evaluation quality of the final products.

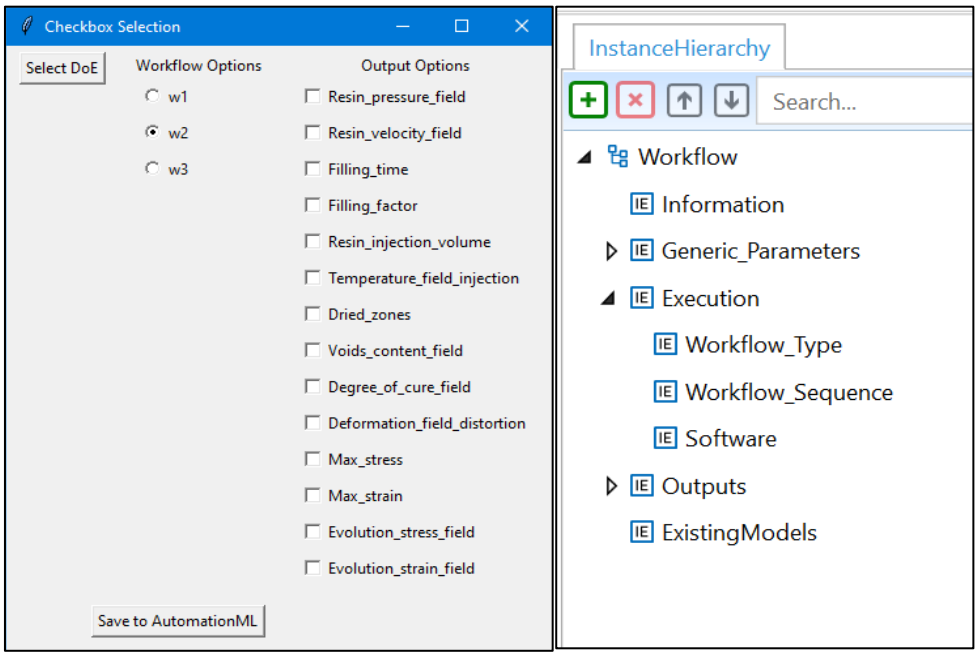


*Figure 9. Workflow Definition HTP interface and AutomationML template.*

Funded by the European Union

Hybrid Twin Platform is then aimed to be a platform for:

1. Designing a complete DoE, not only including parameters and its values, but also involved simulations, desired outputs, date, versions, etc. A graphical user interface has been developed to help the end user to define the workflow parameters. The user selects the DoE, a predefined workflow and the required outputs, and the information is saved to an AutomationML template file (Figure 9. Workflow Definition HTP interface and AutomationML template.Figure 9)
2. Orchestrating the workflow. Nowadays it is beyond demonstration that including simulations early in the design phase reduces time and cost in prototyping. In addition, chaining of simulations has become a common practice in the industry, allowing to propagate effects of each manufacturing step to the subsequent one. Hybrid Twin Platform provides an orchestrator of the designed workflow. This orchestrator is an extension for an existing ESI software called PAM-OPT. The required mechanisms to parse the AutomationML file are being created, as well as the required calls to external partners servers and software tools: ADDPATH, HPC and AIMEN's storage service. The history of launched workflows is stored in a master file.
3. Data analysis and model computation. The generated data during the simulation is captured and categorized using an Industry 4.0 [2] recommended standard: AutomationML. This technology allows to provide a structure to the data that is lacking in traditional methods like csv tables for example. Semantics and ontologies can be used to model the data providing unique definitions of their meaning, units, formats, etc., reducing human errors and facilitating translation methods (import/export between software). This provides an excellent structure to merge simulation and experimental data, which is the base to ESI Model Order Reduction tools like ADMORE or libraries like TensorFlow or machine learning techniques. The AutomationML master file will allow to keep track of existing workflows, compare versions, existing models, etc. and locate the desire data. Finally, ROM or surrogate models are related

Funded by
the European Union

to the data used to train those models in the correspondent field of the AutomationML file. Figure 10 shows the main elements of an AutomationML workflow template.

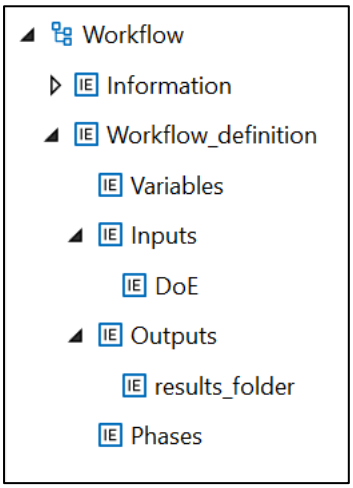


*Figure 10. AutomationML workflow template.*

Funded by the European Union

# 5 SIMULATION SERVICE FOR HPC WORKFLOW EXECUTIONS

The HPC Simulation service aims at facilitating the management of the execution of workflows in the HPC environments. It consists of a web service and an API which allow users and other CAELESTIS ISE components to automate the execution of workflows in the HPC sites. This service can be used by users in standalone mode interacting with the web GUI or used together with other CAELESTIS components interacting through the REST API.

## 5.1 HPC Simulation Service Implementation

The HPC Simulation service has been implemented using Django[3], Nginx[4], PostgreSQL[5], Gunicorn[6] and Paramiko[7] as depicted in Figure 11. Django is a high-level Python web framework that simplifies the development of web applications by providing a robust set of tools and libraries for building dynamic web services. Gunicorn is a Python WSGI (Web Server Gateway Interface[8]) HTTP server. It is responsible to serve Django applications by converting the HTTP request to the Django's Python code. Nginx is a high-performance web server and reverse proxy server. In the context of serving Django applications, Nginx is typically used as a reverse proxy. Nginx acts as an intermediary between external clients and Gunicorn. The status of the service is persistently stored in the PostgreSQL, so in case of a failure in the service happens, all the pending operations can be safely recovered. Finally, Paramiko is a Python library that offers an interface for handling SSH (Secure Shell) and SFTP (Secure FTP) operations. It facilitates secure interactions with remote servers, enabling tasks such as command execution and file transfers across secure connections. The choice of Paramiko for SSH connections is driven by its capacity to programmatically establish and oversee SSH sessions, simplifying the automation of tasks, remote command execution, and secure file transfers between systems. In the present implementation, Paramiko takes on the role of executing scripts and commands by interfacing with the HPC site.

Funded by the European Union

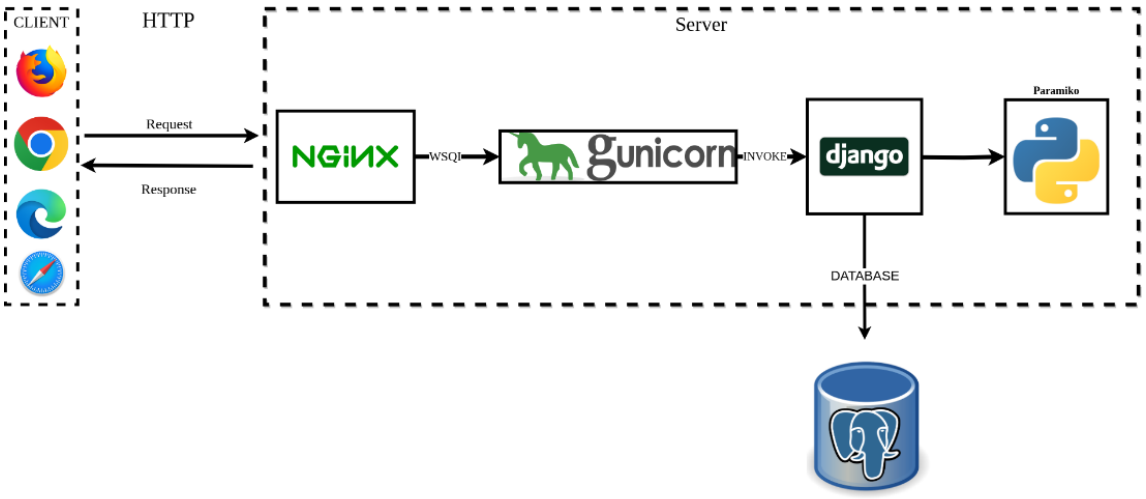


*Figure 11. Software framework used to implement the HPC Simulation Service.*

## 5.2 Integration of the HPC Authentication

One of the main functionalities of the HPC Simulation Service is the automation of the data transfers and execution of the workflow. To enable it, users must authorize the HPC Simulation Service to perform these actions using their HPC site account. This process is depicted in Figure 12 and outlined as follows:

1. **User Registration**: Users sign up for your service by providing the necessary details.
2. **Machine Definition and SSH Key Creation**:
   - After registration, when a user wants to add or define a new machine to be managed by the service, they initiate the process for that specific machine.
   - The service automatically generates a pair of SSH keys for that machine, consisting of a public key and a private key and a unique security token.
   - The public key and the token are sent back to the user.
3. **Private Key Storage**:
   - Before storing the private key in its database, the service encrypts it using the security token.
   - The encrypted private key is then securely stored in the service's database.
   - It ensures that even if someone gains access to the database, they cannot use the private key without the corresponding security token.

Funded by the European Union

4. **Public key Authorization in the HPC**
   - The user places the received public key on the “authorized_keys” in the defined machine. This public key acts as an identifier, allowing only those with the corresponding private key to establish a connection.

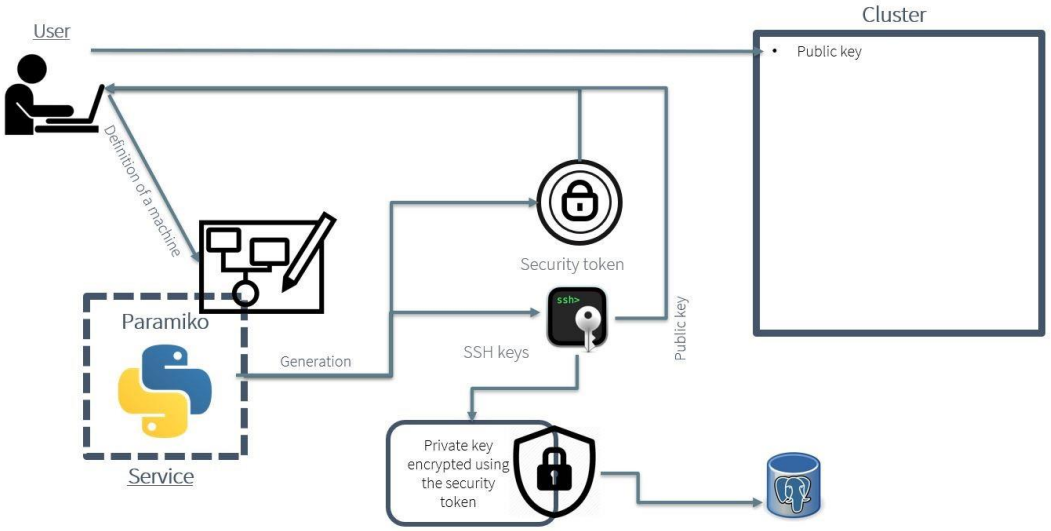


*Figure 12. SSH key generation and HPC authentication*

Once the user has setup the SSH key authorization, it can manage the execution of workflows providing the security token when accessing the HPC Simulation service. This token will be used to perform a secure connection to the HPC site as depicted in Figure 13.

1. **Token-Based service call:**
   - Every time the user wishes to use the service to connect to their machine via SSH, they must provide the security token.
   - This token is essential not only for authentication but also for decrypting the stored private key.
2. **Decryption and Connection:**
   - Upon receiving the correct security token, the service decrypts the user's private key.

Funded by the European Union

- Using the decrypted private key, the service can then establish an SSH connection to the user's machine.

This dual-layer security mechanism ensures that even if malicious actors compromise the service's database, they cannot misuse the stored private keys without the corresponding security tokens. Users retain control, as the service can only access their machines when provided with the correct security token. This system establishes a robust security framework, ensuring the protection of user machines. The combination of SSH key pairs and unique security tokens provides both convenience and enhanced security, safeguarding users from potential unauthorized access.

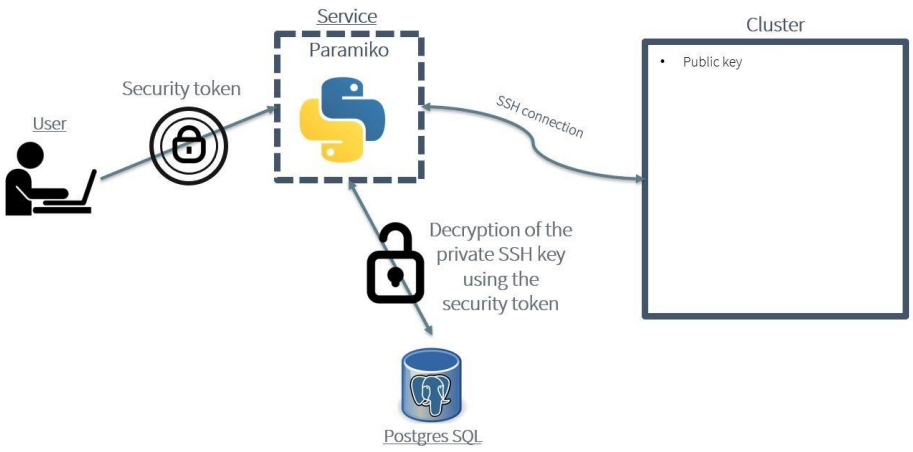


*Figure 13. Token- based authentication*

## 5.3 Workflow Execution in HPC

Figure 14 shows the sequence of steps performed by the HPC Simulation Service automate the different phases for executing workflows in HPC sites. The submission of the workflows as well as the monitoring of the execution can be performed using the web interface or the REST API. The workflow execution steps are executed as follows:

1. **Upload Inputs**: Users or the HTP component prepare the execution environment by uploading input files to the FTP server.
2. **Submit Workflow Description**: Users, or the HTP through the REST API, provide a file describing the workflow they wish to execute to the service as described in Section 2.

Funded by the European Union

Once the workflow is submitted, the status of its execution is displayed in the web interface or retrieved from the REST API.

3. **Download Inputs**: The service retrieves the user's input files from the FTP server, as specified in the workflow description.
4. **Download Workflow Code**: The service fetches the necessary workflow code from the GitHub repository[9] based on the workflow descriptions. It only downloads the code for the specified workflows, excluding others.
5. **Data Staging - Copy Input Files**: The service copies all required input files for simulations into the distributed file system.
6. **Data Staging - Copy Workflows**: The service transfers the downloaded workflow code into the installation directory within the distributed file system.
7. **Workflow Execution**: The service executes the specified workflows. This is done by starting the COMPSs workflow manager[10] in the HPC site.
8. **Copy Result Files**: After job completion, the service accesses the DFS to retrieve the execution results.
9. **Upload Result Files**: To ensure permanent storage of result files, the service uploads them to the FTP server.

This service aims to streamline the implementation and execution of dynamic workflows. By separating the definition of overall workflows from the specification of individual phases, it eliminates the previous static limitations, allowing for modular integration of various phases across workflows during setup.

Funded by the European Union

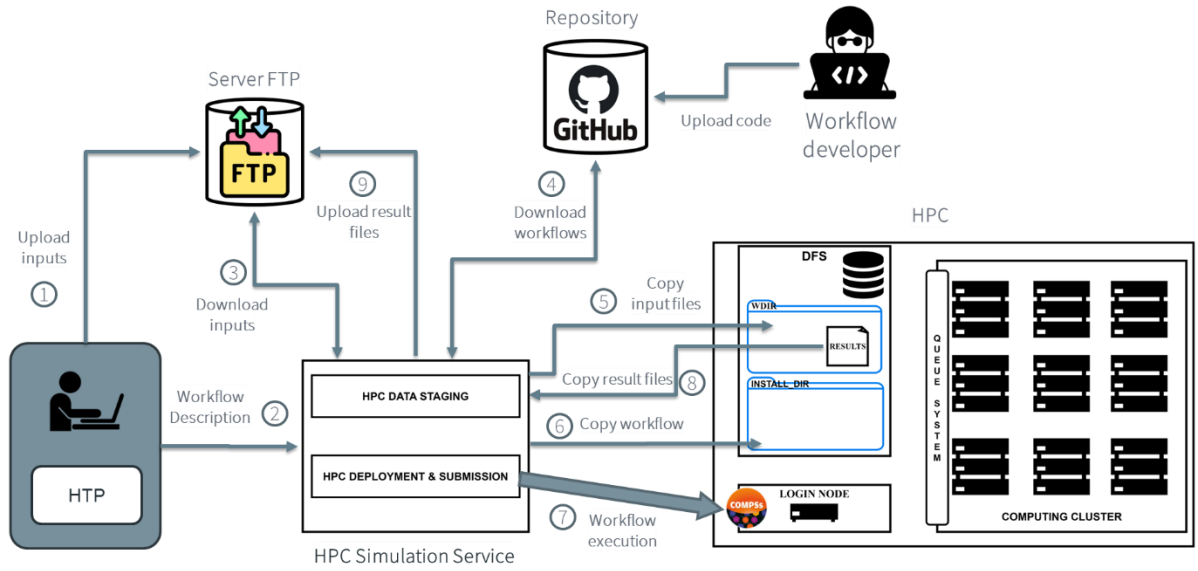


*Figure 14. Interactions for executing workflows in HPC.*

## 5.4 User's Step-by-Step Guide

In this section, we are going to show step-by-step how a user can use the graphical user interface to set up the access to an HPC machine and execute the workflows in this machine. After the completion of the registration and login processes, users will be greeted by the graphical user interface of our dashboard. Initially, the user is required to define a new machine as depicted in Figure 15 providing essential details such as their username, Fully Qualified Domain Name (FQDN), the path to the working directory, and the installation directory.

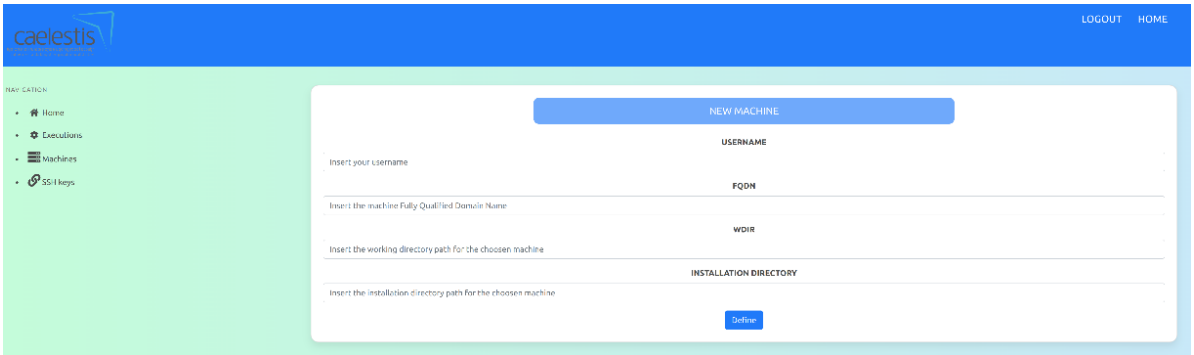


*Figure 15.HPC Machine definition view.*

Funded by the European Union

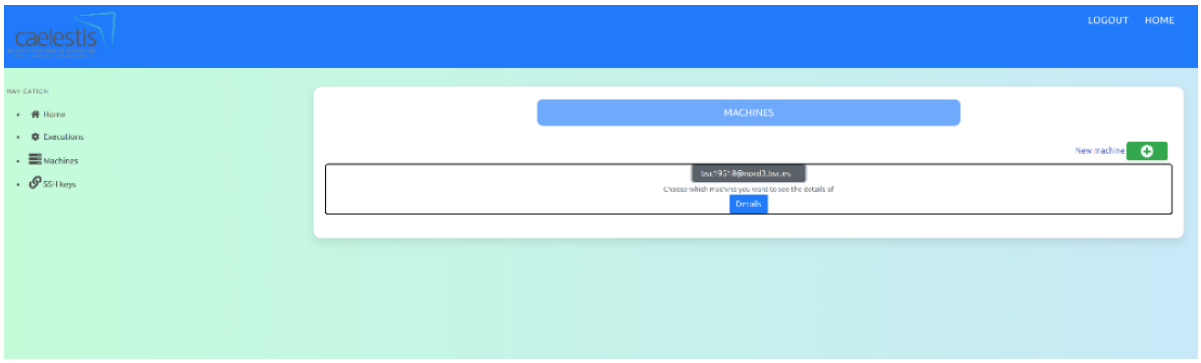


*Figure 16. HPC Machine created view.*

After successfully defining a new machine (Figure 16), the user proceeds to initiate the SSH key generation phase as shown in Figure 17. It is a crucial step to secure access and connect their machine with our service. At this point, our service automatically generates a security token and a pair of SSH keys for that machine: a public key and a private key.

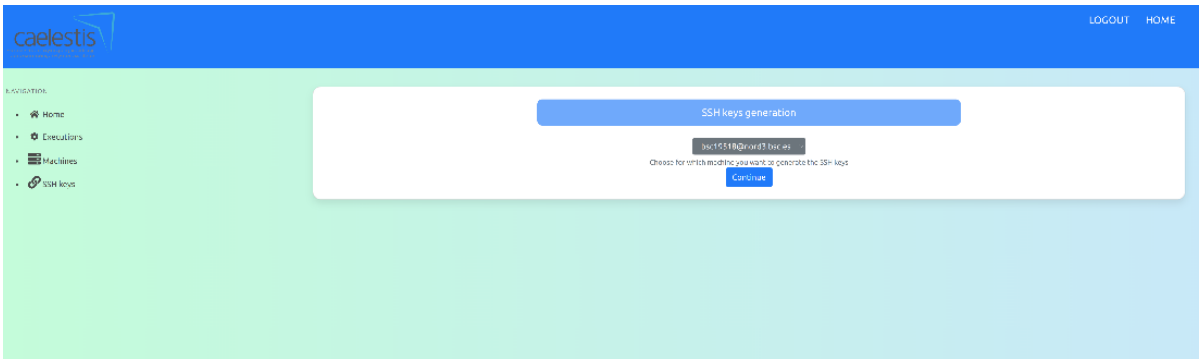


*Figure 17. SSH Key generation view.*

Our service gives back to the user the security token and the SSH public key as shown in Figure 18. The user writes the received public key on the defined machine inside the authorized keys file. This public key acts as an identifier, allowing only those who possess the corresponding private key to establish a connection. Before storing the private key in the service's database, our service encrypts it using the security token. This ensures that even if someone gains access to the database, they cannot use the private key without the corresponding security token.

Funded by the European Union

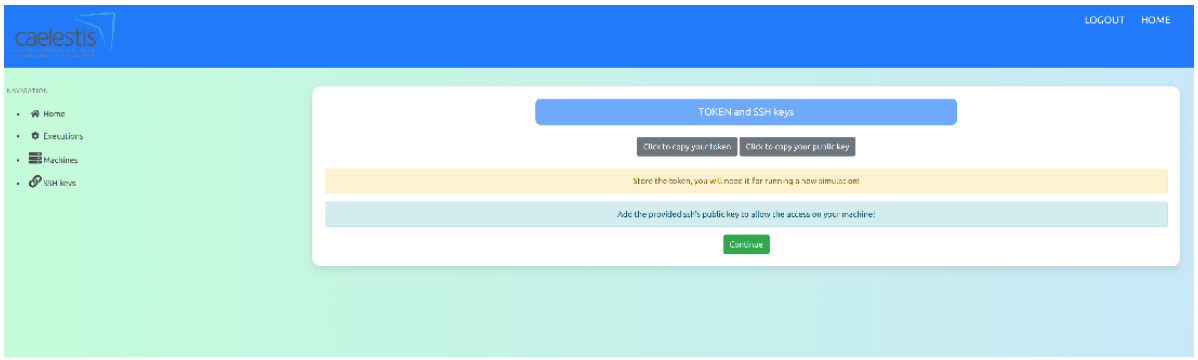


*Figure 18. Generated SSH Keys view.*

The security token will never be stored. Every time the user wishes to use the service to connect to their machine via SSH, he must provide the security token as indicated in Figure 19.

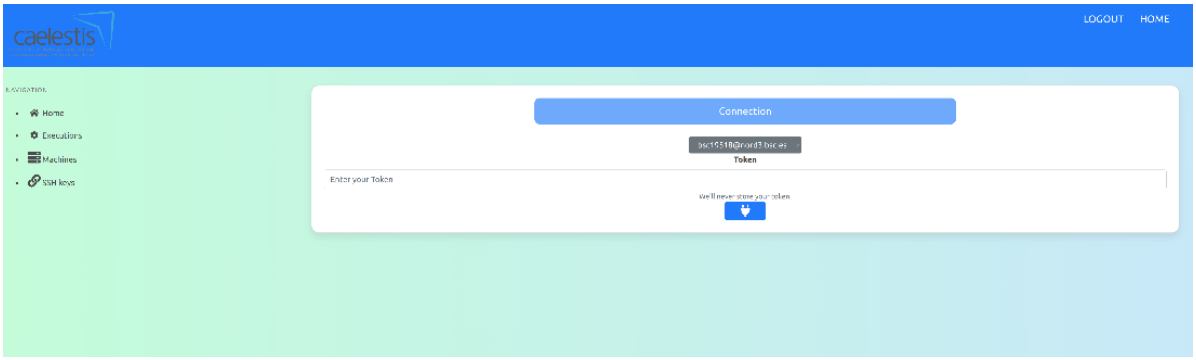


*Figure 19. Connecting to a HPC Machine using the Security token.*

Once the SSH connection is established, the user can see the executions' view (Figure 20) where it can manage the previous workflow execution and starts new ones.

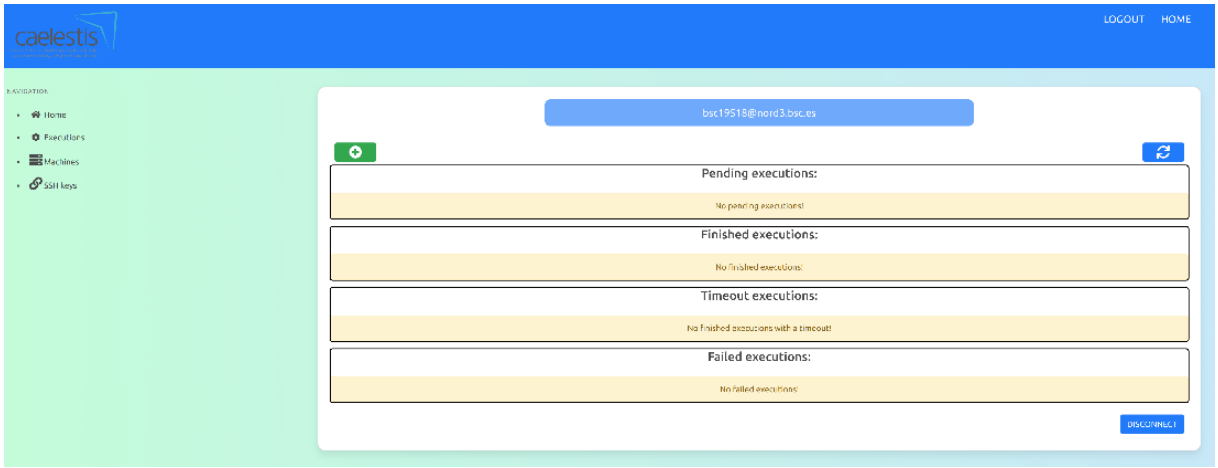


*Figure 20. Workflow Executions view.*

Funded by the European Union

To start new ones, the user needs to provide the workflow description file that describes the workflow to execute together with other options such as the maximum execution time, number of nodes to use in the computation, or enabling the checkpointing and auto-restart. This is depicted in Figure 21.

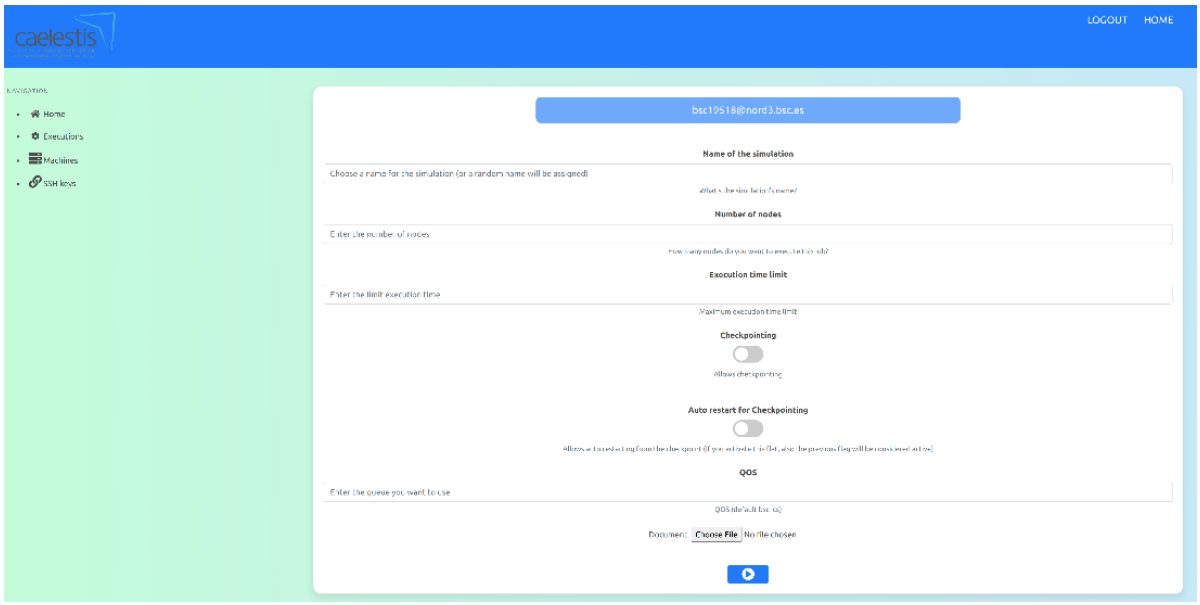


*Figure 21. New workflow execution view.*

After a successful execution, the new workflow execution will appear in the list of workflow executions as depicted in Figure 22.

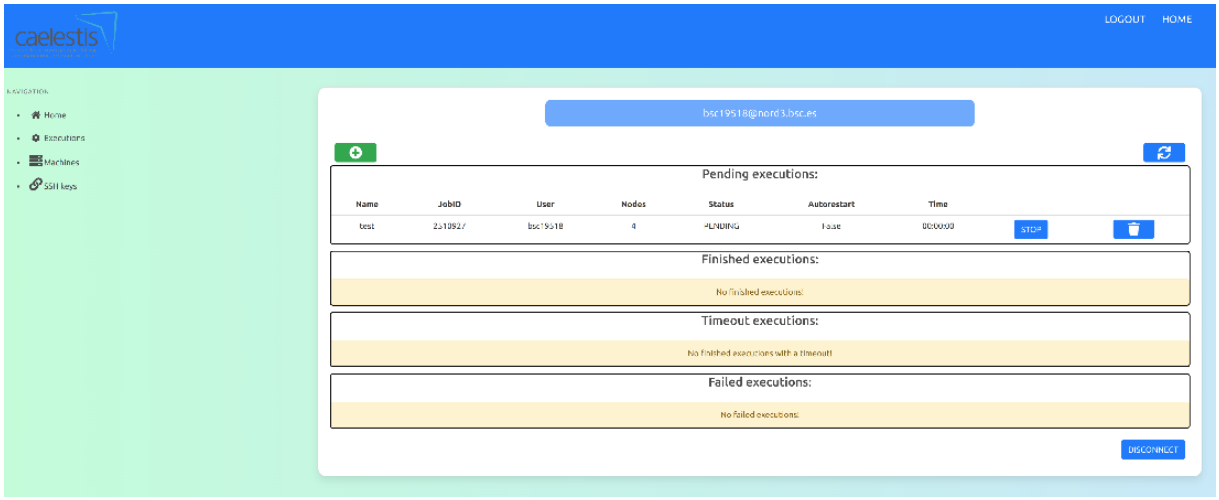


*Figure 22. Executions' view with the new workflow execution.*

Funded by the European Union

## 5.5 REST API for automating the HPC workflows execution.

In the previous section, we have presented the Web Graphical user interface where a user can interact with the HPC Simulation Service. In this section we present the REST API which is provided to enable the programmatic interaction of other components such as the Hybrid twin platform. There are several libraries and command tools to interact with REST API from different programming languages and systems. In the following lines we describe the REST API and provide an example of execution with the cURL command.

**API Authentication**

To secure the API transactions, we use JSON Web Token (JWT) through HTTPS. It is a common standard security method for REST APIs. Before doing any API interaction, the user must use the Web User Interface to generate the API security access token. It must be done once, and the generated token can be reused for every API invocation. This token will be travelling encrypted using the HTTPS protocol and the service will check the validation of the token compared to a hashed version of the token.

**API endpoints**

1. **Run a simulation workflow:**

- Method: POST
- Path: /simulations/
- Content-Type: multipart/form-data
- Content:
    - Mandatory:
        - MachineChoice: Cluster where to execute the workflow
        - SecToken: Token to identify the SSH keys used to submit the workflow
        - NumNodes: Number of nodes
        - ExecTime: Maximum execution time
    - Optional:
        - Branch: Version of the workflow templates used for the execution.
        - Checkpointing: Enabling the chekpointing in the workflow execution.

Funded by the European Union

    - Autorestart: Enabling the autorestart of the workflow if the execution is timed out.
    - QoS: Quality of Service of the execution according the HPC machine.
    - Name: Name for the workflow simulation execution
- Response:
    - message: Successful o error message
    - execution_id: Identifier of the execution

**2. Stop a simulation workflow:**

- Method: PUT
- Path: /simulations/execution/<id>/?status=stop
- Content-Type: multipart/form-data
- Content:
    - MachineChoice: Cluster where to execute the workflow
    - SecToken: Token to identify the SSH keys used to submit the workflow
- Response:
    - message: Successful o error message

**3. Get status of a simulation workflow:**

- Method: GET
- Path: /simulations/execution/<id>/
- Content-Type: multipart/form-data
- Content:
    - MachineChoice: Cluster where to execute the workflow
    - SecToken: Token to identify the SSH keys used to submit the workflow
- Response:
    - eID: Execution identifyer
    - Name: Execution Name
    - User: User used for submission

Funded by the European Union

- NumNodes: Number of Nodes used in the execution
- Status: INITIALIZING | RUNNING | TIME_OUT | FINISHED | FAILED

4. **Restart a simulation workflow:**

- Method: PUT
- Path: /simulations/execution/<id>/?status=restart
- Content-Type: multipart/form-data
- Content:
  - MachineChoice: Cluster where to execute the workflow
  - SecToken: Token to identify the SSH keys used to submit the workflow
- Response:
  - message: Successful o error message

5. **Delete a simulation workflow:**

- Method: DELETE
- Path: /simulations/execution/<id>
- Content-Type: multipart/form-data
- Content:
  - MachineChoice: Cluster where to execute the workflow
  - SecToken: Token to identify the SSH keys used to submit the workflow
- Response:
  - message: Successful o error message

Funded by the European Union

# 6 CYBERSECURITY CONSIDERATIONS ANALYSIS AND RECOMMENDATIONS

As CAELESTIS paves the way for a revolutionary hyperconnected simulation ecosystem, designed to empower the next generation of aircraft structures through innovative design and predictive manufacturing, the realm of cybersecurity stands as an ever-critical sentinel on the path to progress. This section delves into the vital domain of "Cybersecurity considerations analysis and recommendations" (T2.5) to ensure the integrity, confidentiality, and availability of the interconnected digital infrastructure that underpins CAELESTIS. In a landscape where DET and HPC systems converge to harness the immense potential of multiscale, multiphysics simulations, the digital environment becomes an invaluable asset. The CAELESTIS ecosystem, aimed at propelling the European aircraft industry forward, depends on a harmonious coexistence of cutting-edge technologies and the security measures that safeguard them.

In this exploration, the project consortium dissects the cybersecurity implications inherent in the design and manufacturing process, each phase intertwined with the digital underpinnings of CAELESTIS. Taking into account the complexities of DETs and HPC systems during design, to the real-time, edge-device-driven insights during manufacturing [11], the digital ecosystem forms the lifeblood of the project. However, this digital essence is not immune to vulnerabilities and threats, both known and emerging. In other words, T2.5 aims at ensuring that the CAELESTIS ecosystem, set in a hyperconnected simulation environment, federating a variety of actors of the European aircraft supply chain, adopts appropriate cybersecurity and defense mechanisms to safeguard the security of the data, Internet Protocol (IP), and end-users in its digital ecosystem. To this end, the task performs a Privacy, data pRotection, Ethics, cyberSecurity & Societal (PRESS) analysis applied to industrial design and manufacturing simulation applications with a focus on the specificities of HPC and cyber-physical systems. It reviews cybersecurity threats and risk factors in these environments following the ENISA standards applicable to the air transport sector as well as ECSO recommendations [12]. It also elaborates recommendations for ensuring cybersecurity and privacy in such settings via the implementation of digital security standards and technology.

This section is organized as follows. In the first part of the section, we present the PRESS Analysis applied to CAELESTIS project, and in the second part of the section we present how

Funded by the European Union

the cybersecurity requirements derived from the PRESS analysis are affecting the CAELESTIS Interoperable Simulation Ecosystem architecture.

## 6.1 PRESS Analysis

PRESS Analysis is a framework used to evaluate the potential impact of technology on privacy, data protection, ethics, cybersecurity, and the society as a whole [13]. The goal of PRESS analysis is to identify potential risks and benefits associated with the use of technology and to ensure that any negative impacts are minimised while maximising the positive effects. The analysis process typically involves assessing the data collection, storage, and usage practices of technology, as well as its potential impact on individual rights and societal values. Additionally, it also focuses on the ethics of the technology, including issues related to bias, fairness, and transparency. It also considers how the technology may affect cybersecurity and the ability to protect against malicious actors or cyber threats. PRESS Analysis is a comprehensive framework that aims to evaluate the potential impact of technology on various aspects of society. It is a holistic approach to assessing the risks and benefits of new technologies and ensuring that they are developed and used in a responsible and ethical manner. Overall, PRESS Analysis provides a structured and comprehensive approach to evaluating the potential impact of technology on various aspects of society and helps ensure that new technologies are developed and used responsibly and ethically [14].

Therefore, the PRESS Framework Analysis is separated into 3 directions, providing requirements along with potential concerns or threats impacting those requirements, as explained in the following paragraphs, and given tables:

1. **Privacy and Data Protection Requirements:**

Privacy Requirements: In this context, PRESS Analysis evaluates the technology's data collection and usage practices, such as what data is being collected, how it is being used, and who has access to it. It also assesses the technology's ability to protect personal information from unauthorised access or misuse. PRESS analysis of cybersecurity measures regarding privacy requirements includes developing policies that protect personal information, implementing appropriate data protection measures, being transparent about data collection

Funded by the European Union

and use practices, respecting data subject rights, and having an incident response plan in place for privacy breaches. By taking these measures, organizations can help ensure that they are protecting the privacy of personal information and maintaining the trust of customers and other stakeholders.

Data Protection Requirements: This aspect of PRESS Analysis examines how the technology handles and stores data, including data security, data integrity, and data retention policies. It also looks at how data is being shared, and with whom. PRESS analysis of cybersecurity measures regarding data protection includes developing policies that address the protection of data, implementing access controls and encryption, having backup and recovery measures in place, and monitoring and detecting unauthorised access and activity. By taking these measures, organizations can help ensure that their data is protected from unauthorised access or disclosure, and maintain the availability, confidentiality, and integrity of their data.

Table 2 provides a detailed descriptions of the requirements selected for privacy and data Protection.

Funded by the European Union

*Table 2. Privacy & Data Protection Requirements*

| No | Requirement | Description | Potential Concerns or Threats impacting those Requirements | Mitigation Actions | Connection with CAELESTIS |
|---|---|---|---|---|---|
| PDP1 | Data Creation | Design research, plan data management (i.e., formats, storage), plan consent for sharing, locate existing data, collect data (i.e., via survey, workshop) | (1) Data breaches<br>(2) Data loss<br>(3) Data corruption<br>(4) Data privacy violations<br>(5) Confidentiality<br>(6) Compliance with various regulations regarding data privacy, retention, and protection. | To minimise these risks, organisations should implement strong data protection policies and procedures, regularly back up their data, encrypt sensitive information, and ensure that only authorised personnel have access to sensitive data. | Data created in CAELESTIS:<br>• Simulation Data: Inputs and Output data required from simulations<br>• Access data: Passwords, ssh-key pairs<br>• Manufacturing Data: Data collected by sensors<br><br>All partners have signed confidentiality documents. SharePoint, HPC and HTP tools are accessed by credentials. |
| PDP2 | Re-using data | Re-using data focuses on ensuring that an organisation's cybersecurity efforts are aligned with the responsible and ethical reuse of data. | (1) Accuracy of the data<br>(2) Data Privacy<br>(3) Legal compliance<br>(4) Confidentiality<br>(5) Security | To minimise these risks, organisations should implement strong data protection policies and procedures, regularly review, and update the data that they have, ensure that data is used in accordance with the original purpose for which it was collected, and obtain proper consent before re-using data. Organisations should also take steps to secure their data and encrypt sensitive information to prevent unauthorised access. | Data collected from sensors, simulation results and models will be reused. For instance, generated models in HPC will be used in operation by the HTP.<br><br>HTP will include 'digital thread' procedures to track inputs and outputs, versions, etc. |
| PDP3 | Preserving | Migration to best format & medium, backup, and storage, metadata, creation, documenting, and archiving. | (1) Data loss or corruption<br>(2) Incomplete backup<br>(3) Data breaches | It is important to regularly backup data, store it in multiple locations, and use appropriate storage media to reduce the risk of data loss or corruption. | Each component will include its own backup storage, for example HPC, HTP and experimental facilities will each ensure storage and availability of their corresponding data. |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | | | | In HPC, data is maintained and backed-up during the period that users have access to the system. Then, there is 15-days period to move the data to another permanent storage, before the data is removed. In the case of HPC workflows, the final output data will be moved to the available storage at the HTP, and other intermediate data can be kept until the user who executed the workflow decides to remove them, or the access to the system has expired. Regarding the storage at HTP side, the data will be stored until the end of the project.<br><br>In the case of HPC, different storage zones are defined. Users' home directory and the shared spaces for different users of a project are periodically backed-up. Scratch spaces are faster storage but they are considered temporary storage, therefore no backup is done in this data space. The execution of the workflows will take into account these spaces. Scratch storage will be used for temporary data and final data will be moved to spaces with backup. |
| PDP4 | Analysing data | Data interpretation, research outputs, preparation for archiving, deletion, etc. | (1) Privacy violations<br>(2) Data breaches<br>(3) Lack of transparency | Implement appropriate privacy and security measures, ensure that data analysis algorithms are transparent and unbiased, and validate the results of data analysis to reduce the risk of | ESI commercial software used in CAELESTIS has large experience and validation. Data analytics tools used are known Python libraries well documented. Otherwise, required |

| | | | | | |
|---|---|---|---|---|---|
| | | | | privacy violations, bias, and incorrect decisions based on faulty data. | transparent information will be provided when possible.<br><br>The PyCOMPSS source code is publicly available, and it has also several tests to validate their functionalities. Workflows in PyCOMPSs are implemented as annotated Python scripts. They are executed in parallel, but the results of the computation are the same as it was executed in sequential order.<br><br>Alya code is present in the PRACE benchmark suite together with Code_Saturne in the field of Computational Fluid Dynamics and thus, considered as the reference in the European framework for supercomputing. The code is validated through a Test Suite (TS) with more than 50 unity tests, 345 regression tests and also the performance of relevant tests is evaluated through the Performance Benchmark Suite (PBS), the TS and the PBS are developed in-house.<br><br>OpenFOAM is a renowned open-source Computational Fluid Dynamics (CFD) software that undergoes several validation processes to ensure the accuracy and reliability of its results. These processes involve a diverse |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | | | | range of validation tests, including analytical solutions, experimental data, and benchmark cases from various domains of fluid dynamics and a wide spectrum of flow scenarios.<br><br>Twinkle is a library for building families of solvers to perform Canonical Polyadic Decomposition (CPD) of tensors[15]. Kfold cross-validation analysis is used to ensure the best approximation avoiding problems such as overfitting. |
| PDP5 | Storage limitation & Data erasure | The principle of storage limitation should ensure that personal data is stored only for a period necessary for the fulfilment of the data processing purpose for which data has been initially collected. After that period data will be erased. However, if there are requirements of mandatory law personal data might be kept for a longer period. Also, data might be kept for a longer period if there are reasons related to public interests, scientific or historical, and statistical purposes. | (1) Data breaches<br>(2) Unauthorised access<br>(3) Non-compliance with laws and regulations | • Establish and enforce data retention policies: Organisations should establish policies that outline what types of data should be stored, for how long, and how it should be deleted. This can help reduce the risk of unauthorised access and limit the potential for a data breach.<br>• Implement access controls: Organisations should implement access controls to ensure that only authorised personnel have access to sensitive data. This can include measures such as multi-factor authentication, role-based access controls, and privileged access management.<br>• Use data encryption: Organisations can use encryption to protect | All the mentioned data collected and used by CAELESTIS are stored in corresponding sites that will include a repository specific for this purpose with restricted access.<br><br>As commented in PDP4, several storage systems are available in the HPC system. Every user has a guaranteed but limited amount of space in the “home” storage. There is also a limited storage space in “projects” and “scratch” but this space is shared between all users of the project. Data in all these spaces will be removed once the user access grant is expired. |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | | | sensitive data both at rest and in transit. This can help ensure the confidentiality, integrity, and availability of the data.<br>• Regularly update security protocols: Organisations should regularly review and update their security protocols to ensure they are up to date with the latest threats and vulnerabilities.<br>• Use data erasure techniques: When data is no longer necessary, organisations should use appropriate data erasure techniques to ensure that the data is completely removed from their systems. | Nevertheless, via CAELESTIS, no personal data are foreseen to be collected.<br><br>After the project's completion, data used will be erased. |
| PDP6 | International transfers of data & data sharing | The project partners will exchange data for the purpose of the project realisation. All project partners are based in the European Union. If it would be necessary to share data outside the EU or to transfer data in countries that are not on the so-called 'adequacy list' of countries (decided by the European Commission) then data transferring will be carried out in accordance with appropriate safeguards regulated by GDPR. Whenever they transfer data to third parties that will process data on their behalf, not only | (1) Data breaches: Data transmitted over the internet or other networks is vulnerable to interception or hacking, leading to data breaches and theft of sensitive information.<br>(2) Unauthorised access: Data can be accessed by unauthorized individuals during transfer if proper security measures are not in place.<br>(3) Data loss or corruption: Data can be lost or become corrupted during transfer due to network failures, software bugs, or human error. | • Transfer of data:<br>○ Encryption: Encrypt data during transfer to prevent unauthorised access or interception of sensitive information.<br>○ Secure transmission: Use secure file transfer protocols (e.g., SFTP, HTTPS) or virtual private networks (VPNs) to transfer data to ensure that it is transmitted over a secure connection.<br>○ Verification of recipient: Verify the identity of the recipient before transferring sensitive | Simulation inputs and results will be transferred from the HTC and HPC sites as well as other sites, where the software tool can be only used locally. And this data can be shared with different users of the project.<br><br>For the data transfer, but also for the users to access HTP, secure channels as VPN and SSH will be utilised. |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | requirements imposed by GDPR, but also appropriate safeguards must be at place. Third parties must process data only on documented instructions from the project partner as well as in accordance with other requirements imposed by the agreement concluded by the partner and a third party. | (4) Data interception: Data can be intercepted by unauthorised individuals during transfer, leading to theft or misuse of sensitive information.<br>(5) Inadequate security: Data transmitted over public networks may not be adequately secured, leaving it vulnerable to interception or hacking.<br>(6) Data misuse: Shared data can be misused for unauthorised purposes, such as marketing or sales, leading to potential privacy violations.<br>(7) Data misuse: Shared data can be misused for unauthorised purposes, such as marketing or sales, leading to potential privacy violations.<br>(8) Data retention: Shared data may be retained for longer periods than necessary, increasing the risk of unauthorised access or breaches. | data to prevent unauthorised access.<br>• Data sharing:<br>○ Access control: Implement access control mechanisms to restrict access to sensitive data based on the user's role and responsibilities.<br>○ Data minimization: Only share the minimum amount of data necessary to accomplish the task at hand.<br>○ Contractual agreements: Establish a clear understanding of how the shared data will be used and protected through a written agreement or contract.<br><br>Encrypt data during transfer, use secure file transfer protocols, and verify the identity of the recipient before transferring sensitive data to reduce the risk of data breaches and unauthorised access. Also, implement access control mechanisms, establish clear agreements on the use and protection of shared data, and ensure that shared data is stored securely. | |

Funded by the European Union

**2. Ethical and Social Requirements:**

Ethical Requirements: This aspect of PRESS Analysis examines the technology's impact on societal values, such as fairness, transparency, and accountability. It also looks at issues related to bias, discrimination, and other ethical concerns. PRESS analysis of cybersecurity measures regarding ethical requirements includes developing policies that align with relevant laws and regulations, conducting risk assessments to identify potential ethical issues, being transparent about cybersecurity efforts and practices, ensuring inclusivity in cybersecurity efforts, and being accountable for cybersecurity incidents. By taking these measures, organizations can help ensure that their cybersecurity efforts are not only effective but also aligned with ethical considerations and individual rights.

Social Requirements: This aspect of PRESS Analysis examines the broader impact of technology on society, including economic, social, and environmental impacts. It considers how technology may affect different groups of people, such as vulnerable populations, and how it may impact the way we live and work. PRESS analysis of cybersecurity measures regarding societal requirements includes developing policies, following regulations, educating employees, adhering to industry standards, and implementing a range of security measures to protect against cybersecurity threats. By taking these measures, organisations can help ensure that they are protecting not only themselves but also society at large from the potentially devastating impact of a cybersecurity breach.

Table 3 provide the details about the requirements identified for ethics and social aspects.

Funded by the European Union

*Table 3: Ethics & Social Requirements*

| No | Requirement | Description | Potential Concerns or Threats impacting those Requirements | Mitigation Actions | Connection with CAELESTIS |
|---|---|---|---|---|---|
| ESR1 | Lawful data collection, protection of personal data, and processing. | The data processing shall originate from those personal data that have been collected with a lawful ground. This means that users should be informed about how their data is being collected, stored, and used, and must give their explicit consent for this to happen. In this respect, the CAELESTIS consortium should ensure that the personal data collected is accurate (i.e., data are correct and up to date in all details), and safeguarded from unauthorised access, theft, or misuse. | (1) An issue may arise in the case that the data subject is not aware of the data collected and shared, or the collection of data is made on the wrong legal basis, or in the absence of a legal basis.<br>(2) The lack of information among involved parties is the primary potential cause for inaccurate data in a system.<br>(3) It is also quite frequent the collection of unneeded (personal) data, i.e., data not relevant to the objectives of the system and for the agreed purposes of data processing.<br>(4) Unauthorised access or alteration of data, as well as reputational damage, legal liabilities, and financial losses, are additional threats that CAELESTIS consortium should bear in mind when it comes to lawful data collection, protection, and processing. | • CAELESTIS must have policies in place that address the collection, use, and storage of data in accordance with relevant laws and regulations, as well as industry standards and best practices for data protection;<br>• CAELESTIS must classify data according to its sensitivity and importance. This can help ensuring that appropriate measures are in place to protect data and that only authorised individuals have access to it;<br>• CAELESTIS must only collect and store data that is necessary for its business purposes. This can help reducing the risk of unauthorised access or disclosure of sensitive information;<br>• CAELESTIS must obtain consent from individuals before collecting their personal information. This can include providing clear and concise privacy notices that explain how personal information will be used to obtain the consent from | Each organization in the consortium is responsible to follow internal security procedures while working on CAELESTIS to safeguard any collection or use of personal information.<br><br>Sensors' data and simulations results will be collected and post-processed to generate models for predicting the impact of defects. Therefore, CAELESTIS will collect and store only the necessary for the project execution data. The security-by-design approach is an activity that shall be utilised by partners working on the implementation of software components to ensure that any future personal data will be handled accordingly. However, there is no foreseen collection or process of personal data. If this happened, the relevant consent will be obtained.<br><br>Documents should be prepared for HPC, that will follow usual procedures at all partners. They exist for SharePoint and HPC, at least for data protection. Access to HTP and its |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | | | individuals before collecting their personal information;<br>• CAELESTIS must have policies in place that address how long data should be retained and when it should be disposed of. This can help ensure that data is not retained for longer than necessary and is disposed of in a secure and appropriate manner;<br>• CAELESTIS must inform users about data processing practices, and that they can control how their data is used;<br>• CAELESTIS must implement adequate security measures such as encryption, access controls, and monitoring systems to detect and prevent data breaches, as well as ensure that data is stored and transmitted securely, and that data is kept up-to-date and accurate. | related data will follow legal requirements.<br><br>The data and sensitive information in the project will be kept confidential, during the action and for at least 5 years afterwards. |
| ESR2 | Transparency & Accountability | The purposes of the data processing should appear clear and intelligible for the data subject. This can be ensured by providing all the appropriate and necessary information to data subjects to exercise their rights to data controllers to evaluate their processors, and to Data Protection Authorities to monitor | (1) Lack of trust: A lack of transparency in data analysis can lead to a lack of trust in the results and the decisions made based on the results.<br>(2) Bias and discrimination: Data analysis algorithms that are not transparent or are biased can result in unfair treatment or discrimination against certain groups or individuals. | • Ensure transparency in the ethics and social requirements of data analysis to reduce the risk of privacy violations, bias and discrimination, and to increase trust and accountability in results.<br>• Provide clear explanations and interpretations of the results of data analysis to increase transparency and accountability. | CAELESTIS will collect and store only the necessary for the project execution data. There is no foreseen collection or process of personal data. If this happened, security procedures and the relevant consent will be obtained. Nevertheless, there is a clear view through the designed |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | according to responsibilities. The technology solutions, and their relative data models, thus should ensure that a data subject might get easy access to information at any time, also after the start of the data processing operations. | (3) Privacy violations: Lack of transparency in data analysis can result in privacy violations and the misusing of personal information.<br>(4) Irresponsible use: Data analysis that is not transparent can result in the results being used in an irresponsible or unethical manner, causing harm to individuals or the society.<br>(4) Informed decision-making: Lack of transparency in data analysis can prevent individuals from making informed decisions based on the results, leading to incorrect or misguided decisions. | • Document the methods and algorithms used in data analysis, including assumptions and limitations, to increase transparency and understanding.<br>• Use open-source software and algorithms for data analysis to ensure that the workings of the algorithms are transparent and can be reviewed by experts.<br>• Have independent experts validating the results of data analysis to ensure that they are accurate and unbiased.<br>• Provide users with control over the data used in analysis and the results generated, allowing them to verify the accuracy and fairness of the results. | architecture of how data are exchanged.<br><br>Data are stored on HPC with restricted access to the users, thus none of those data will be publicly available.<br><br>All utilised ESI tools are well documented.<br><br>Not all codes developed within CAELESTIS effort source code are open source. The background of commercial offerings is regulated by consortium agreement. Open-source software and algorithms might be used wisely and securely used. |

Funded by the European Union

### 3. Security Requirements:

Security Requirements: The process of identifying and defining the security measures and controls that are necessary to protect an organisation's information and information systems. By considering security requirements in the context of the broader PRESS analysis framework, organisations can ensure that their security measures and controls are aligned with their overall cybersecurity risk management strategy and can help to mitigate the risks of security incidents and breaches.

Table 4 provides the details for the requirements related to security.

Funded by the European Union

*Table 4: Security Requirements*

| No | Requirement | Description | Potential Concerns or Threats impacting those Requirements | Mitigation Actions | Connection with CAELESTIS |
|---|---|---|---|---|---|
| SR1 | Security measures | Information security addresses integrity, confidentiality, and availability concerns → Restricting access to personal data only to authorised people (i.e., permissions), and ensuring that the data is trustworthy and accurate (i.e., based on provenance information). The relevant CAELESTIS partner(s) should also: (i) regularly conduct a privacy risk assessment and audit processes; (ii) regularly run reviews of the security measures implemented; and (iii) design an ad hoc procedure to be followed in case of a data breach.<br><br>The IT infrastructure shall implement adequate and appropriate security measures able to protect the data to be ingested in the infrastructure as well as its functionalities. In this respect, such measures shall include both physical measures as well as technological ones, and in | Complexity: Complex security systems can be difficult to manage and maintain, leading to potential vulnerabilities.<br>False sense of security: A belief that the security measures in place are enough to prevent all attacks can lead to complacency and neglect of security best practices.<br>Interference with usability: Security measures that are too restrictive or complicated can interfere with the usability of systems and discourage their adoption.<br>Dependence on a single solution: Relying on a single security solution can create a single point of failure and leave the organisation vulnerable if that solution is compromised.<br>Resource drain: Implementing and maintaining security measures can be resource-intensive, requiring significant time, money, and personnel.<br>Legal and regulatory compliance: Security measures may need to be adjusted to comply with new laws and regulations, adding to the complexity of maintaining them. | • Conduct a security risk assessment.<br>• Develop a comprehensive security policy.<br>• Implement appropriate technical measures.<br>• Implement appropriate administrative controls.<br>• Regularly assess and update security measures<br>• Monitor and audit security measures<br>• Maintain compliance with relevant laws and regulations | No personal data is collected or used.<br><br>There is a restriction access for the users to the HPC, based on different roles (admin. Simple user, etc.).<br><br>In case of data breach, security partners and the project coordinator shall be informed.<br><br>Security by design approach will be followed during the software implementation and integration phases.<br><br>Any exchange of information via secured APIs (ensure API requests are authenticated, authorised, validated, and required HTTPS). |

| | | | | | |
|---|---|---|---|---|---|
| | | any case shall be designed applying a risk-based approach, which shall consider all the components and their interactions. | | | Consortium partners ensure they are compliant with relevant laws and regulations |
| SR2 | System Security | Comprehensive security policies are in place to outline the security measures that are required, as well as the consequences for failing to adhere to them. Policies should cover all areas of security, including physical security, network security, and data security.<br><br>There are a variety of industry standards that can be used to guide security efforts. The most well-known of these is the ISO/IEC 27001 standard, which outlines best practices for information security management systems. | (1) Malware<br>(2) Phishing<br>(3) Insider Threats<br>(4) Distributed Denial of Service (DDoS) attacks<br>(5) Physical Security Breaches<br>(6) Zero-day exploits<br>(7) Advanced Persistent Threats (APTs)<br>(8) Web-based attacks | • Firewalls, antivirus software, intrusion detection systems, and access controls.<br>• Regular testing and updates should be conducted to ensure that these measures are effective.<br>• Ongoing employee training and staying up-to-date on the latest security risks and trends. | HPC malware treatment to be treated with HPC experts.<br><br>System security of HPC site is managed by system administrators, CAELESTIS software that interact with this system have to be compatible with HPC system security. |
| SR3 | Secure design | Ensure security controls are applied at the system's design and that it is possible to implement them to meet security requirements (defined by the project). | (1) Lack of security expertise: One of the biggest threats to securing system design is a lack of security expertise among the designers and developers. If the people responsible for designing and building the system don't have a | • Organisations should ensure that their designers and developers have a strong understanding of security principles and best practices, and that they conduct thorough threat modelling and testing | The security-by-design approach is an activity that shall be utilised by partners working on the implementation of software components and integration activities (i.e. design of an API) to |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | | strong understanding of security principles and best practices, the resulting system is likely to be vulnerable to a wide range of security threats.<br>(2) Insufficient threat modelling<br>(3) Poorly designed security controls<br>(4) Failure to consider emerging threats<br>(5) Insufficient testing and validation: Finally, a lack of testing and validation can be a major threat to securing system design. Without proper testing and validation, it is difficult to know whether the system is secure and whether the security controls are working as intended. This can lead to vulnerabilities and security weaknesses that may be exploited by attackers. | throughout the design process. They should also regularly review and update their system designs to address emerging threats and vulnerabilities and should work with security experts to ensure that their security controls are properly designed and implemented. | ensure that any future personal data will be handled accordingly.<br><br>Testing and validation shall be performed after a release, ensuring the securing of CAELESTIS system design. Avoid lack of testing since without proper testing and validation, it is difficult to know whether the system is secure and whether the security controls are working as intended. |
| SR4 | Risk Assessment | Cybersecurity and risk assessment are closely related, as risk assessment is an essential part of developing and implementing effective cybersecurity measures. Risk assessment is the process of identifying, evaluating, and prioritising risks to an organisation's assets, including its information systems to determine | (1) Malware: Malware, or malicious software, includes viruses, worms, trojans, and other types of software designed to harm or compromise computer systems.<br>(2) Phishing: Phishing is a type of social engineering attack in which an attacker tries to trick a user into divulging sensitive information, such as login credentials or financial information, by impersonating a trustworthy entity. | • Identify assets: The first step in risk assessment is to identify the assets that need to be protected. This may include data, systems, hardware, software, and other resources.<br>• Identify threats: The next step is to identify potential threats to these assets. Threats may come from a variety of sources, including external attackers, | A risk assessment focus on the cybersecurity aspect is a process integrated into the overall risk assessment process of CAELESTIS.<br><br>As presented in the left column the threats need to be identified for each CAELESTIS components that it is implemented. Then for each risk identified, we need to assess the |

| | | | | | |
|---|---|---|---|---|---|
| | | the best way to mitigate those risks.<br><br>In the context of cybersecurity, risk assessment typically involves identifying potential threats to an organisation's information systems and data, evaluating the likelihood and potential impact of those threats, and developing strategies to reduce or eliminate the associated risks. This may include implementing security controls, training employees on security best practices, and establishing incident response plans. | (3) Distributed Denial of Service (DDoS): A DDoS attack involves overwhelming a network or server with traffic from multiple sources, causing it to become inaccessible to legitimate users.<br>(4) Insider threats: Insider threats involve employees, contractors, or other authorised users who intentionally or unintentionally compromise the security of the organisation's systems or data.<br>(5) Advanced persistent threats (APTs): APTs are sophisticated and targeted attacks in which an attacker gains persistent access to an organization's systems to steal data or cause other harm.<br>(6) Ransomware: Ransomware is a type of malware that encrypts an organisation's data and demands payment in exchange for the decryption key.<br>(7) Unpatched software: Unpatched software refers to software that has known vulnerabilities that have not been addressed through software updates or patches.<br>(8) Physical security breaches: Physical security breaches can involve theft of physical devices containing sensitive data, unauthorised access to server | internal threats, and natural disasters.<br>• Assess likelihood and impact: Once threats have been identified, the next step is to assess the likelihood and potential impact of each threat. This may involve considering factors such as the probability of a threat occurring, the potential financial or reputational impact, and the potential harm to individuals.<br>• Develop risk mitigation strategies: Based on the results of the risk assessment, the organisation can develop strategies to mitigate the identified risks. This may involve implementing additional security controls, training employees, and developing incident response plans.<br>• Review and update: Finally, the organization should regularly review and update its risk assessment process to ensure that it remains effective and up-to-date with emerging threats and vulnerabilities. | likelihood to happen, and the impact will have if happen. Finally, we need to provide a mitigation action and continue monitor/update. |

| | | | | | |
|---|---|---|---|---|---|
| | | | rooms or data centres, or other types of physical access. | | |
| SR5 | Accessing Control (user access) | (1) Access control refers to the practice of limiting access to resources or information based on a set of predefined rules or policies. It is an essential component of modern security practices and is critical in preventing unauthorised access to sensitive data and resources.<br>(2) Data access controls: Each project partner should allow access to data on a 'need-to-know' bases. Therefore, all reasonable measures should be taken to ensure that personal data is accessible and manageable only by authorised staff. Persons entitled to use a data processing system must have access to the personal data to which they have access privileges. Personal data cannot be read, copied, modified or removed, or otherwise processed without appropriate authorisation.<br>(3) Network access controls: Project partners are | (1) Brute force attacks: Attackers may use automated tools to try and guess login credentials, either through password guessing or by trying different combinations of usernames and passwords until they gain access.<br>(2) Social engineering attacks: These are attacks that involve tricking people into divulging their login credentials or other sensitive information. Examples include phishing attacks, where attackers send emails that look like they are from a trusted source and ask users to enter their login credentials on a fake login page.<br>(3) Insider threats: Insiders can pose a security threat to an organisation, as they may have access to sensitive data or resources that they can misuse or steal. This can be intentional, such as an employee stealing data to sell to a competitor, or unintentional, such as an employee accidentally sharing sensitive information.<br>(4) Malware attacks: Malware can infect a system, allowing an attacker to gain access to data or | • Access roles for users and defined the access rights of each role, on different functionality or information.<br>• Implementing a policy for protecting access to network and information systems, addressing for example roles, rights, responsibilities, and procedures for assigning and revoking access rights is a suggested standard by ENISA.<br>• User registration process & authentication/approval mechanism by an admin while a token shall be assigned to each user. This token will accompany any user request and by that, we will prevent anonymous access requests and control accessing activity.<br>• User access shall be based on authentication credentials of username and password. Authentication access can also be a key, based on a Secure Shell Protocol (SSH) protocol for access credentials. Its function is similar to usernames and passwords and generating a secured key can be | HPC systems are using their own access control. They are normally relying on UNIX access control mechanisms.<br><br>Access control is based on access roles for users and defined the access rights, while there is also a user registration process where credentials per user are defined and registered.<br><br>CAELESTIS software is compatible and integrated with this mechanism.<br><br>There is a utilisation of Hypertext Transfer Protocol Secure (HTTPS) for accessing requests not only on HPC but in general for components under the CAELESTIS ISE. |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | responsible for securing data while data is being processed on their infrastructure. Therefore, the use of firewalls or functionally equivalent technology, as well as authentication control measures may be required.<br>(4) Physical Access Controls: Each project partner keeps physical components of the network in its own facilities. Physical barrier controls must be used to prevent unauthorised entrance to the facilities. Accessing the facilities necessitates passing through physical barriers, which can be achieved through either electronic access control validation (such as card access systems) or validation by human security personnel. Visitors are obliged to register with designated personnel, present valid identification, and remain under the supervision of authorised employees throughout their visit to the facilities.<br>(5) Limited Access: Access to the facilities should be extended exclusively to staff and contractors with a genuine | resources that they should not have access to. Malware can be spread through various means, including email attachments, infected websites, or malicious downloads.<br>(5) Weak access controls: Poorly configured or weak access controls can leave systems vulnerable to attack. For example, if a system allows weak passwords or does not enforce multi-factor authentication, it can be easier for attackers to gain access. | used for authentication purposes. The keys are highly recommended for automated processes and for implementing single sign-on by system administrators and power users.<br>• Utilization of Hypertext Transfer Protocol Secure (HTTPS). The primary drivers for HTTPS are twofold: first, it ensures the authentication of the accessed web location, and second, it safeguards the privacy and integrity of data exchanged during transmission. These ENISA standards seamlessly align with HTTPS, offering protection against various threats, including man-in-the-middle attacks. Furthermore, HTTPS supports bidirectional communication encryption, effectively shielding the entirety of the communication from eavesdropping and tampering.<br>• Use of logging functionality. Access logs show unique identifiers for users and systems when granted or denied access.<br>• Conducting traffic management measures to preserve integrity and security of the network. This could basically consist of allowing | |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | business necessity. As soon as access privileges are no longer warranted for staff/contractors, these privileges should be promptly withdrawn. | | connectivity to and from specific endpoints only. | |
| SR6 | Availability | Availability is one of the key security requirements that refers to the ability of a system or service to be accessible and functional when needed. In other words, it means that the system or service should be available to users and operate properly without interruption or downtime. | (1) Denial of Service (DoS) attacks<br>(2) Distributed Denial of Service (DDoS) attacks<br>(3) Malware infections<br>(4) Hardware or software failures | • Redundancy and fault tolerance: This requirement ensures that the system or service has duplicate components and resources that can take over in case of failure of the primary components. This helps ensure that the system or service continues to function without interruption.<br>• Scalability: This requirement ensures that the system or service can handle an increasing number of users, requests, or data without becoming overwhelmed or crashing. This helps ensure that the system or service remains available and responsive even during peak usage periods.<br>• Monitoring and alerting: This requirement ensures that the system or service is constantly monitored for potential issues or failures, and alerts are sent to the appropriate personnel when necessary. This helps ensure that | HPC systems involves high number of resources, so is it very likely that a failure in a resource happen during the execution of a large computation.<br><br>Workflow management systems must include mechanism to checkpoint and recover executions.<br><br>Each service is constantly monitored for potential issues or failures, and alerts are sent to the appropriate personnel, notifying them when necessary. |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| | | | | problems are detected and addressed quickly, minimising downtime, and improving availability. | |
| SR7 | Authentication and Authorization | Authentication is a security requirement that ensures that users are whom they claim to be. Authentication is the process of verifying the identity of an individual, system, or device, and it is a critical component of security. Authentication helps prevent unauthorised access to systems and information and ensures that only authoried individuals have access to sensitive information.<br><br>Authorisation is a security requirement that ensures that users have appropriate access to resources or actions within a system. Authorisation determines what actions or data a user can access based on their role, permissions, and other factors. Proper authorisation helps prevent unauthorised access to sensitive resources or actions and helps maintain the security and integrity of a system. | (1) Password attacks<br>(2) Phishing<br>(3) Insider threats<br>(4) Weak authentication<br>(5) Misconfiguration | To mitigate these threats, it is important to implement strong authentication measures, such as multifactor authentication and strong password policies. Multifactor authentication requires users to provide two or more forms of authentication, such as a password and a fingerprint or a one-time code. Strong password policies can help prevent password attacks by requiring users to create complex passwords and change them regularly. Regular security assessments and vulnerability testing can also help identify and address vulnerabilities that could be exploited to compromise authentication. Additionally, user awareness training can help educate users on the importance of authentication and how to prevent unauthorised access to information. | HPC systems are using their own authentication and access control system.<br><br>Access control is based on access roles for users and defined the access rights, while there is also a user registration process where credentials per user are defined and registered. |

Funded by the European Union

| | | | | | |
|---|---|---|---|---|---|
| SR8 | Key Management | Key management is a security requirement that involves the secure generation, storage, distribution, and destruction of cryptographic keys used to protect sensitive information. Cryptographic keys are used to encrypt and decrypt data, as well as to authenticate users and secure communications. Proper key management is critical to maintaining the confidentiality, integrity, and availability of sensitive information. | (1) Loss or theft of keys: If a key is lost or stolen, it can be used to access sensitive information, compromising the security of the system.<br>(2) Weak key generation: If cryptographic keys are generated using weak algorithms or random number generators, they may be vulnerable to attacks that could compromise the security of the system.<br>(3) Inadequate key storage: Keys that are not properly protected can be vulnerable to theft or tampering.<br>(4) Inadequate key distribution: Keys that are distributed in an insecure manner can be intercepted or stolen, compromising the security of the system. | To mitigate these threats, it is important to implement strong key management practices, such as the use of strong algorithms and secure key storage. Keys should be generated using secure algorithms and random number generators to ensure that they are sufficiently complex and difficult to predict. Keys should also be stored in a secure manner, using techniques such as encryption or hardware security modules (HSMs). Additionally, key distribution should be carefully managed to ensure that keys are only distributed to authorised parties using secure channels. Regular key rotation and destruction can also help mitigate the risk of key loss or theft. Finally, regular security assessments and vulnerability testing can help identify and address any weaknesses in key management practices. | Some CAELESTIS software will manage user's keys and passwords to automate the execution of these workflows. |

Funded by the European Union

The PRESS requirements provided in depth in the above-mentioned tables, elsewhere known in the project as “PRESS Requirements Checklist”, are based on ENISA standards and ECSO recommendations. EBOS, as the cybersecurity expert in CAELESTIS, has conducted desk research to collect the relevant information and inform the project partners about the core requirements to bear in mind. The project partners have acknowledged and agreed to proceed with the proposed requirements, and they have made the connection with the CAELESTIS project. The aim is for CAELESTIS partners to collaborate to ensure throughout the project’s duration, on a 3-month basis, that those requirements are met and that no risks or threats occurred in the project. However, whenever a risk occurs, partners are obliged to take action internally, as proposed in the respective tables above, and inform EBOS and the Project Coordinator, AIMEN, accordingly.

The core mission is to secure the visionary digital ecosystem underpinning CAELESTIS, guaranteeing that its transformative potential is realised without compromising the integrity of the European aircraft industry, the privacy of its stakeholders, or the ethical and social principles that underpin our society. Through this comprehensive cybersecurity analysis, we endeavour to fortify the project, allowing it to soar to unprecedented heights while standing resilient against the turbulent winds of the digital age.

## 6.2 Cybersecurity implications in the CAELESTIS ISE Architecture

The requirements presented in the previous section has an impact on the design of the CAELESTIS ISE architecture which is summarised in Figure 23. Most of the components of the architecture are deployed in premises of the CAELESTIS partners. IT managers must control the security of the computing system where the components are deployed applying the system security requirements and recommendations presented above (e.g., firewalls, OS update, malware detection, …). Apart from this, all service invocations and data transfers must use secure protocols, such as HTTPS, SSH or SFTP or SCP, to ensure that service calls and messages are encrypted, and they must include Authorization, Authentication and other access control capabilities. These secure communications imply the usage of keys and tokens to automate the invocations and the data movements included in the workflow. These keys must be managed in a secured way and its storage must use specialized encryption mechanism to ensure that these

Funded by the European Union

credentials are not stolen even if a security issue is affecting the computer that is hosting the component. Finally, the Storage Service and the HPC file systems must incorporate the data protection recommendations to ensure that data are not lost or damaged.

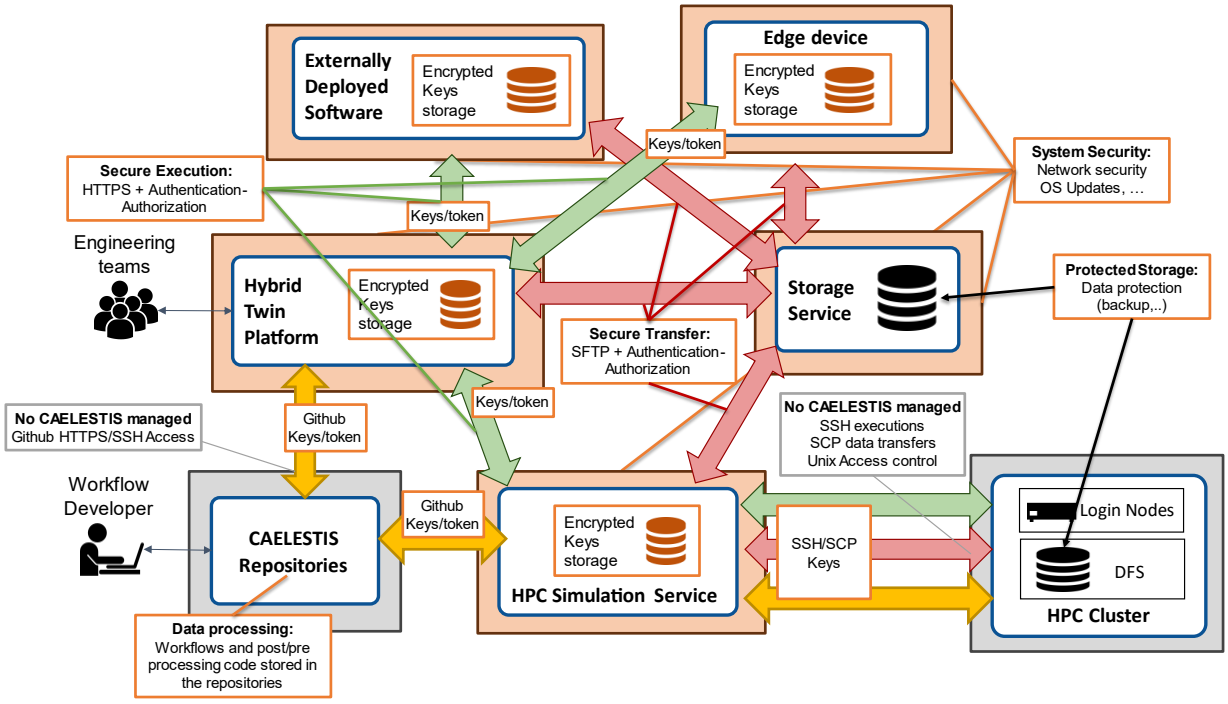


*Figure 23. Cybersecurity implications in the CAELESTIS ISE architecture.*

Funded by the European Union

# 7 IMPLEMENTATION STATUS

Some parts of the CAELESTIS ISE Components are covered by services which already exist. The HPC site is managed by the BSC HPC System Administrators. It is using the IBM's Global Parallel File System and Slurm as queue system. The access to this system is performed through login nodes using SSH key credentials. As Storage Service, AIMEN has deployed a Secure FTP service in their premises which can be accessed from the other component locations. Finally, the CAELESTIS Repositories are git repositories which are stored inside GitHub CAELESTIS organization space.

As part of this deliverable, we have released the first prototype of the main components of the CAELESTIS ISE. It includes the first version of the HTP and Simulation Service together with the codes of the workflow templates, software invocation and data processing code that we have used until M19. We have also included a prototype implementation of External Software Executor to enable the execution of software from the Hybrid Twin Platform. Table 5 provides the URLs of the repositories where the implemented CAELESTIS ISE components can be found. These repositories will be periodically updated with the implementation of new workflows, bug fixes and improvements.

*Table 5. Source code location for the implemented CAELESTIS ISE Components.*

| Component | Repository URL |
|---|---|
| Hybrid Twin Platform | https://github.com/CAELESTIS-Project-EU/AMLtool |
| HPC Simulation Service | https://github.com/CAELESTIS-Project-EU/Simulations_Service |
| External Software Executor | https://github.com/CAELESTIS-Project-EU/external_software_executor |
| Workflows repository | https://github.com/CAELESTIS-Project-EU/Workflows |

Funded by the European Union

# 8 CONCLUSION AND FUTURE WORK

In this deliverable, we have presented the architecture and implementation details of the CAELESTIS Interoperable Simulation Ecosystem (ISE) to facilitate the management of complex product and process multiscale and multiphysics simulation scenarios. The proposed Simulation Ecosystems provides a Software infrastructure to allow the integration of the different actors and environments involved in these complex simulations. Workflow developers can use this system using the mechanism to create workflows that combine different simulation software and execution environment. Engineers can easily execute these workflows in the HPC using the services used by the CAELESITIS ISE, where it automatically deploys the implemented workflow and required data to the HPC system, perform the computation and manage its results to create models that can be used during the manufacturing process. Moreover, the workflows and their executions are described using the AutomationML format to be compatible with Industry 4.0 standards and keeping track of the digital thread.

Regarding the cybersecurity aspects of the CAELESTIS ISE, we have conducted a PRESS Analysis for the scenario of the CAELESTIS project. In this analysis, we have identified a set of security requirements associated to the different threat and concerns. For each of these requirements, we have evaluated the connection with the CAELESTIS project and defined some mitigation actions. Apart from it, we also seen the impact of the PRESS Analysis in the CAELESTIS ISE architecture proposing the measures to make the CAELESTIS ISE a secure environment.

Finally, we have also provided the links to the source code of the first implementations of the CAELESTIS ISE components and the workflow templates, software invocations and data processing algorithms used until now.

The remaining task will be focused on the implementation of the workflows templates and phases required for the analysis required in the CAELESTIS project as well as maintaining and extending the features of the components of the CAELESTIS to support them. This deliverable

Funded by
the European Union

will be complemented with D2.2 where we will provide a more mature release of the CAELESTIS ISE components, workflow templates and implementations of the workflow phases.

Funded by the European Union

# 9 REFERENCES

[1] Rainer Drath, "*AutomationML, A Practical Guide"*, (2021), https://www.automationml.org

[2] Plattform Industry 4.0, *German Standardization Roadmap Industrie 4.0*, version 4, (2020), https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/standardization_roadmap_i40.htm

[3] The Django Framework Web site, https://www.djangoproject.com/

[4] NGIX Server Web Site, https://nginx.org/

[5] Postgre SQL Database Web Site, https://www.postgresql.org/

[6] Gunicorn Python WSGI HTTP Server Web Site, https://gunicorn.org/

[7] Paramiko Web Site, https://www.paramiko.org/

[8] Gardner, James. "The web server gateway interface (WSGI)." *The Definitive Guide to Pylons* (2009): 369-388.

[9] CAELESTIS GitHub Web site, https://github.com/CAELESTIS-Project-EU

[10] Badia, Rosa M., et al. "Comp superscalar, an interoperable programming framework." *SoftwareX* 3 (2015): 32-36.

[11] European Cyber Security Organisation, "ECSO Technical Paper on Cybersecurity scenarios and Digital Twins", version 1, (2023), https://ecs-org.eu/ecso-uploads/2023/07/ECSO_WG6_DigitalTwin-2.1.pdf

[12] European Union Agency for Networks and Information Security, "Privacy and Data Protection by Design – from policy to engineering", (2014), https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design

[13] IoT-NGIN project, "D4.1 Next Generation IoT PRESS Analysis & Confidentiality Requirements" (2021), https://iot-ngin.eu/wp-content/uploads/2022/09/IoT-NGIN_D4.1_v1.pdf

[14] PHOENIX project, "Deliverable D4.1 PRESS Framework Analysis", (2020), https://phoenix-h2020.eu/wp-content/uploads/PHOENIX_D4.1.pdf

[15] V.Zambrano, R.Rodriguez-Barrachina, S.Calvo, S.Izquierdo, "TWINKLE: A digital-twin-building kernel for real-time computer-aided engineering", *SoftwareX* 11, (2020): 100419.