Project Title	Hybrid Eco Responsible Optimized European Solution
Project Acronym	HEROES
Grant Agreement No.	956874
Start Date of Project	01.03.2021
Duration of Project	24 Months
Project Website	heroes-project.eu

D5.4 – Sustainability and business model report

Work Package	**WP 5, Dissemination and Business Model**
Lead Author (Org)	**Philippe BRICARD (UCit)**
Contributing Author(s) (Org)	**Corentin LEFEVRE (Neovia), Clémentine FERRE (Neovia), Valentin JAY (Neovia)**
Reviewed by	**Benjamin DEPARDON (UCit)**
Approved by	**Management Board**
Due Date	**28/02/2023**
Date	**09/05/2023**
Version	**V1.0**

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

Versioning and contribution history

Version	Date	Author	Notes
0.1	01.02.2023	Corentin Lefevre (Neovia)	TOC
0.2	30.03.2023	Philippe Bricard (UCit), Clémentine Ferré (Neovia), Valentin Jay (Neovia), Corentin Lefevre (Neovia)	Revised TOC Draft plan of contents Complete draft of part 4
0.3	13.04.2023	Philippe Bricard (UCit), Corentin Lefevre (Neovia)	Complete first draft
0.4	02.05.2023	Philippe Bricard (UCit), Corentin Lefevre (Neovia)	First completed version before internal review
0.5	03.05.2023	Clémentine Ferré (Neovia), Valentin Jay (Neovia)	Internal reviewed version
0.6	05.05.2023	Benjamin Depardon (UCit)	Internal reviewed version
1.0	09.05.2023	Philippe Bricard (UCit), Corentin Lefevre (Neovia)	Final updates for submission

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

Table of Contents

List of Figures

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

TERMINOLOGY

Terminology/Acronym	Description
AI	Artificial Intelligence
AWS	Amazon Web Services
BSC	Barcelona Supercomputing Center
CAD	Computer-Aided Design
CAE	Computer-Aided Engineering
CC	Computing Continuum
CPU	Central Processing Unit
ETP4HPC	European Technology Platform for High Performance Computing
EuroHPC JU	European High Performance Computing Joint Undertaking
GPU	Graphics Processing Unit
HEROES	Hybrid Eco Responsible Optimized European Solution
HPC	High-Performance Computing
HPDA	High Performance Data Analytics track
ISC	International Supercomputing Conference
MFLOP/s	Mega Floating-point Operations Per Second
SaaS	Software as a Service
SME	Small and Medium-sized Enterprises
SRA	Strategic Research Agenda
VO	Virtual Organisation

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

Executive Summary

High Performance Computing addresses a wide variety of challenges for academic/research communities as well as Industrial users. They used to be addressed historically through large procurement of a dedicated system by a single organization and access organized through calls for proposals. This “one-to-one” relationship no longer exists, and the induced complexity of the “many-to-many” context drives the requirement for orchestration platforms aimed at masking to scientists and engineers the underlying infrastructure complexity when they want to submit HPC/AI workloads. Within the HEROES project, the prototype of such an orchestration platform has been defined for implementations that the project team often refer to as “marketplace”. This document describes the assumptions behind and the context of the need for a HPC hybrid marketplace as well as it develops the business model scenarios and their expected implementation scheme at the end of the project. Apart from the immediate exploitation paths, potential short and middle terms follow-ons have been identified by the HEROES consortium that are reported at the end of this report.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

1 Introduction

Central to the HEROES platform is the notion of Marketplace for computing resources used for HPC and AI workloads. In the introduction part of "The Anatomy of the Grid", Ian Foster, Carl Kesselman and Steve Tuecke defined the Grid concept as "…coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations…" and then "…This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a virtual organization (VO)"[1].

In this document, the HEROES consortium aims to describe the business models and opportunities for different envisaged deployments of the HEROES platform and related services. This report starts with a brief overview of the evolution of the hybrid HPC services, describing the current and expected needs of both HPC consumers and providers, as well as the overall market potential of services addressing those needs. Having described the "customer pain", this document then drafts the concepts of what a European marketplace for hybrid HPC access could be, first with an overarching description then detailed from the point of view of such a marketplace stakeholder.

Based on this description, the position of the HEROES platform is developed from a business perspective, e.g., drafting the scenarios of HEROES potential business models. The scenarios are completed from the stakeholders' perspective by the presentation of the HEROES business survey, finally leading to a description of the project follow-on for further exploitations and exploratory measures.

Several key drivers have been identified in the course of developing the vision and associated services for a European hybrid HPC marketplace, coming either from an EU-wide perspective led by EuroHPC and the installation of a network of world-class supercomputers in Europe, or from industrial stakeholders in the evolution of their HPC infrastructures and use.

Exploiting those key drivers in the rise of opportunities of hybrid HPC will be in the coming years one of the bottlenecks for the EU to operate and provide access HPC resources and therefore effectively realise the promised performance of HPC from an academic, industrial and more over societal point of view.

[1] From the introduction of HEROES D3.5 – Cost Service

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

2 The business context for hybrid HPC services

From the first MFLOP/s provided by CDC 6600 in the mid-60's, development and use of supercomputers have drastically evolved in parallel with its market share and value. In this part, we will provide a short description of the latest evolution and the current state of the business context for hybrid HPC services, first from the perspective of HPC users, then from the point of view of HPC providers.

2.1 A brief overview on HPC usage and needs

High Performance Computing is primarily designed to serve various User Communities by delivering the computing power they need to achieve their activities. At the edge of their respective disciplines, the user communities had to shape since the 1960's the development of the computing tools (hardware or software) that they needed, and as new tools emerged new usage surfaced. Today Scientific Research, Product Innovation, Decision Making or Artificial Intelligence use cases coexist independently and are even combined in complex workflows, using physics-based simulation, analysis of large data volumes and machine learning.

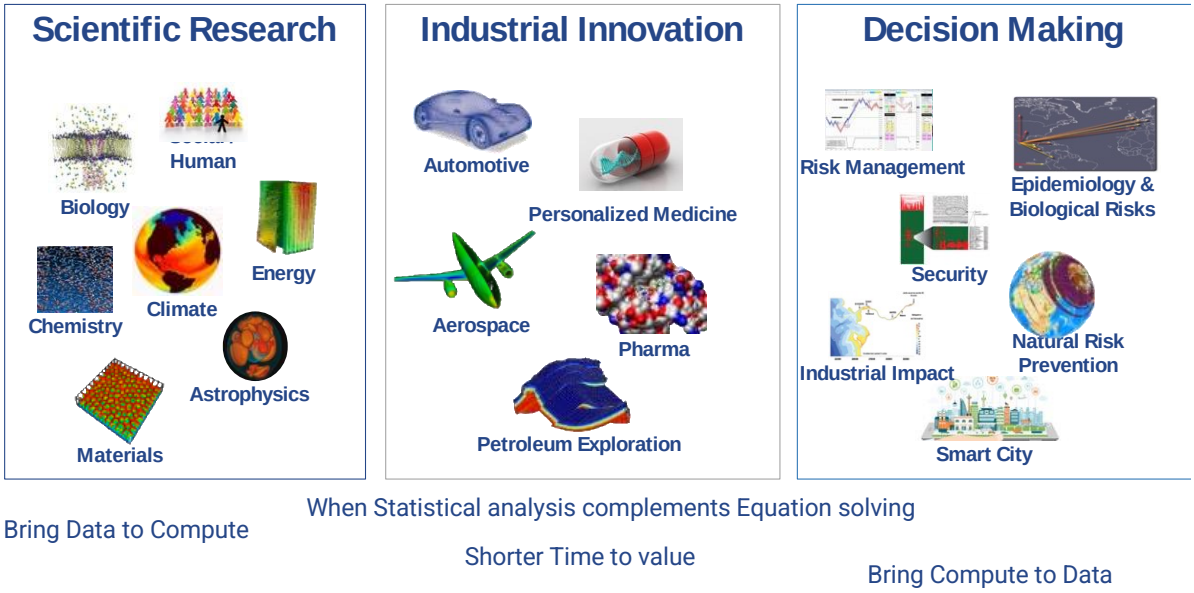

Figure 1 - HPC use cases

Physics-based simulations initially drove the development of Supercomputers measured by the number of floating-point operations they can perform in a given time and aimed at achieving the grand challenges required to support "big science". Essentially dependent on their compute capacity, their expensive resources are allocated carefully. From 1 MFlops in the 60's to a few hundred MFlops in the 80's they created an economy and business models which are still valid nowadays as Europe and others are on the road to Exascale.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

Supercomputers were centrally procured, and their core technologies evolved fast. They are expensive and mostly used in the public research community as most private companies can't afford them. As these systems became bigger and bigger, their environmental requirements including the energy to run and cool them, the space to host them or the network to interconnect their components is just making more critical the need to optimize them. Data life cycle management & Middleware Software are becoming an essential component along with the Services needed to operate them.

With the raise of workstations, CAD, CAE and workstations industries started to implement HPC on a lower and more modular scale with a bigger share of client's budgets devoted to application software. Time-to-Result varied depending on each industry time-to-market requirements while the limited skills to operate these environments pushed naturally towards efficient use. The balance between resource centralization strategies and distributed deployments lead to the creation of a software economy which still exist despite the raise of open-source software in the past 20 years. The pre-eminence in client budgets of application software drove a hardware standardization trend (to the benefit of x86 architecture) with the development of clusters built with commodity resources in the early stage extended recently to Cloud Computing and its various flavours.

The emergence of Artificial Intelligence and the move to HPC/HPDA and other urgent decision making is becoming the bigger usage. Characterized first with the 4 V's of Big Data (Volume, Velocity, Variety, Veracity), the consequences on infrastructure not only affects Storage and Data Life Cycle Management Software but also the need for specialized processors with ad hoc characteristics to process Data in HPC Centres or at the edge of the network where they are created and where energy is less abundant.

As a consequence of this coexistence, there is a variety of HPC architectures to address optimally the variety of HPC problems along with different priorities and goals from user communities.

2.2 The actual state of HPC infrastructure providers

The historical Infrastructure providers are the hardware and software vendors and the simple way for an organization to build a High-Performance Computer is to buy and own it to give access to it to its engineering and research employees. Public Labs or National Computing Centres followed the same model initially deployed in islands which are not affordable anymore. The trend for the On-Premises HPC infrastructure is to be mutualized (centralized in a reduced number of data centres and distributed in grids for the most advanced organizations).

Hosted Infrastructures are very commonly used as Data Centre and environmental requirements are hardly affordable at scale, various levels of managed services can be put in place. Resources are usually dedicated in co-located datacentre for security reasons to the exception of specific vertical applications which can be used as Services (Application Service Providers). Sharing of resources have always been limited for security reasons in both public and industrial market segments.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

A number of HPC Hardware providers (IBM, Dell, HPE…) started however to build HPC On-Demand Services in order to accelerate the deployment of their newest technologies and address the need for extra capacity on short period of times. These business models were usually not sustainable for HPC until enough capacity built outside of HPC market allowed AWS, Google, Microsoft and others to propose public clouds infrastructure progressively scaling high.

Lastly there are several services which can enter into play for execution of an HPC workflow, which are executed out of the Computing cluster and constitutes mandatory requirements (user authentication, software license check, data ingestion, pre-processing of data or post-processing of the results through visualization software…). Some of these services can be delivered locally to the user, locally to the organization or simply accessible as external services from the cluster.

2.3 When offers meets demands – the need for a hybrid HPC broker

The following picture represents a simplistic view of the challenge that an end user could face if having access to different resources to perform it's HPC/AI workflow/jobs.

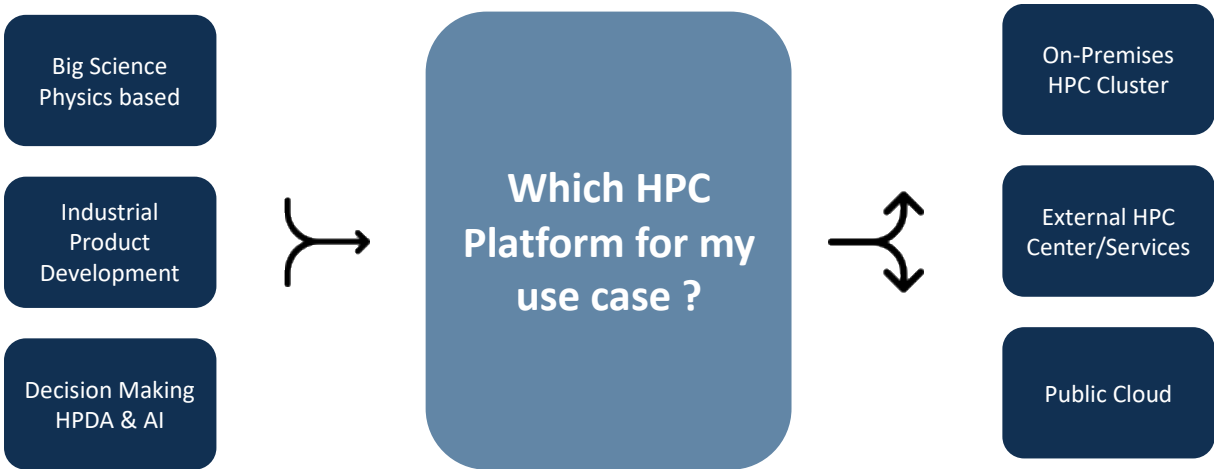

Figure 2 Which HPC platform for my use case?

From a more technical standpoint, we need to consider that there can be many technical configurations for each type of resources (an organization can have multiple on-premises resources and servers; can have access to multiple HPC Providers each one potentially able to propose various configurations and service level agreements while public cloud offers both a choice of hardware configurations and quantities).

As ChatGPT would say when asked "Which HPC platform should I use for my HPC workload?", the answer provided is "The choice of HPC platform to use for your workload will depend on several factors such as the nature of your workload, the resources you have available, and your budget…". Although we didn't rely on it - as it didn't exist – when thinking about the

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

HEROES project in 2019 and when looking at the specific workloads and contexts for HPC end users we firmly believed that there is a case for an orchestration platform acting as a broker between user demand and resource supply.

3 Description of generic HPC Marketplaces

In the previous section, we gave a general technical and market analysis of the current state of hybrid HPC both from academic/industrial and user/provider points of view. Realising hybrid HPC leads to the definition of an overarching HPC marketplace for which the HEROES platform could be one of the key components. This part aims to first describe such a marketplace and its usages for its stakeholders.

3.1 General elements of a marketplace concept

The overall concept of a marketplace needs three elements to be characterized:

- A marketplace first needs to be **organised with a defined set of rules** (when and where it will happen, how to access it for vendors, how to access it for clients, how should the information be organised – for instance on price transparency...). This set of rules could be more or less complex depending on the nature of the market. Generally, one body/organisation is in charge to define and apply such set of rules. By analogy to physical food marketplaces, we call it "**The mayor**".

- A marketplace then needs **vendors** who must agree with the above-mentioned set of rules, yet who are free to define specific terms or other conditions – typically fixing a price – for their product and/or service.

- Finally, a marketplace needs **clients** searching to acquire products or services matching their needs at the conditions set by the mayor and vendors that will suit them.

Figure 3 sums up such an overall concept illustrated by physical food market.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

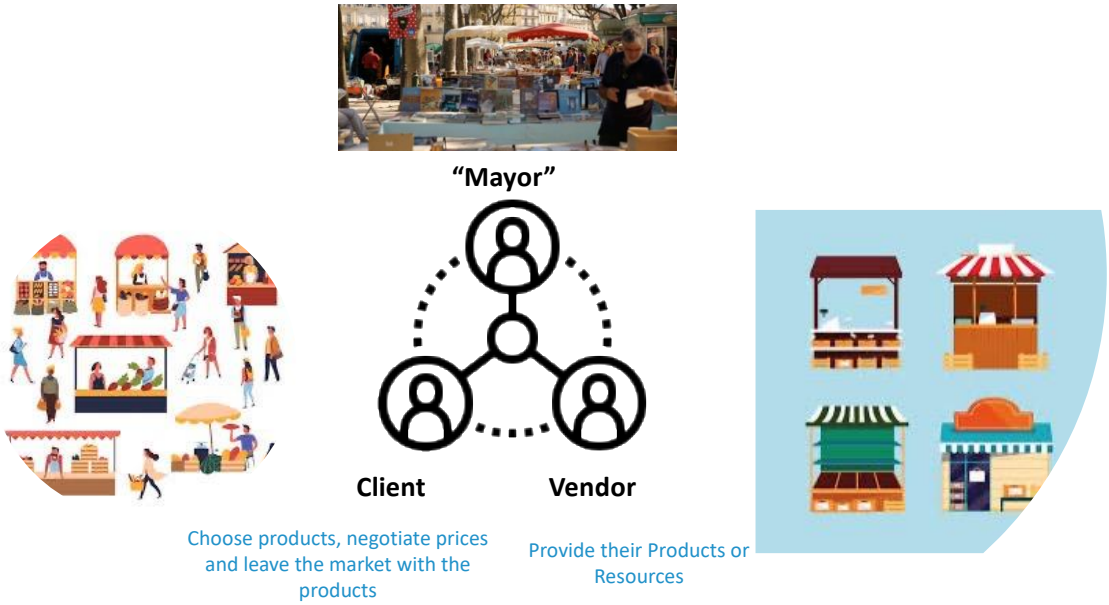

Figure 3 - Roles in a physical food marketplace[2]

3.2 Application of the marketplace concept to hybrid Cloud/HPC

Such an overall concept can be applied to provide a definition of a European hybrid cloud/HPC marketplace, with the following roles:

- "**The mayor**" is the **Administrator** of the marketplace:
  - Its main role is to set up the conditions under which vendors share their HPC resources and under which the clients execute their workflows.
  - **The HEROES platform** belongs to this block as it acts as an intermediary between vendors and clients with several key functionalities displaying the terms of services – performance, price and energy.

- "**The vendors**" are the **HPC resource providers**:
  - Such HPC resource providers wish to sell part of their resources and, upon the general set of rules, define their own terms and conditions which may vary over time. For instance, they may decide that queuing with GPU won't have the same availability and price than non-accelerated compute resources or that preemptible resources could be sold at a "spot" price much cheaper than "guaranteed"/on-demand resources...
  - There is a variety in the HPC resource providers typology. They can be public or private HPC centres, Cloud providers, or even being users that search to optimise the use of their HPC on-premises cluster.
  - It shall be noted that for a first definition and giving the perimeter of the HEROES project we concentrate on HPC resources. Our hybrid HPC marketplace concept could be extended to propose and sell predefined or packaged workflows.

[2] Schema for HEROES D3.5 – Cost Service

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

- "**The clients**" are the **HPC users**:
  - Such users want to execute their workflows using resources that fit their needs with several criteria: performance, price and energy – for instance: "I need my results within 12h with a 10k€ budget with my corporate social responsibility enforcing me to use the most energy efficient resources"...
  - There is a variety in the HPC users typology which we summarized in three main categories: SMEs, large enterprises and Universities and public research organisations.

Figure 4 presents the application of the basic marketplace concept to the case of hybrid HPC with the **Administrator** role in the centre playing as an intermediary between the **HPC users** on the left and the **HPC providers** on the right. The schema also provides a short description of the main interest for such a marketplace for each category of users and providers.

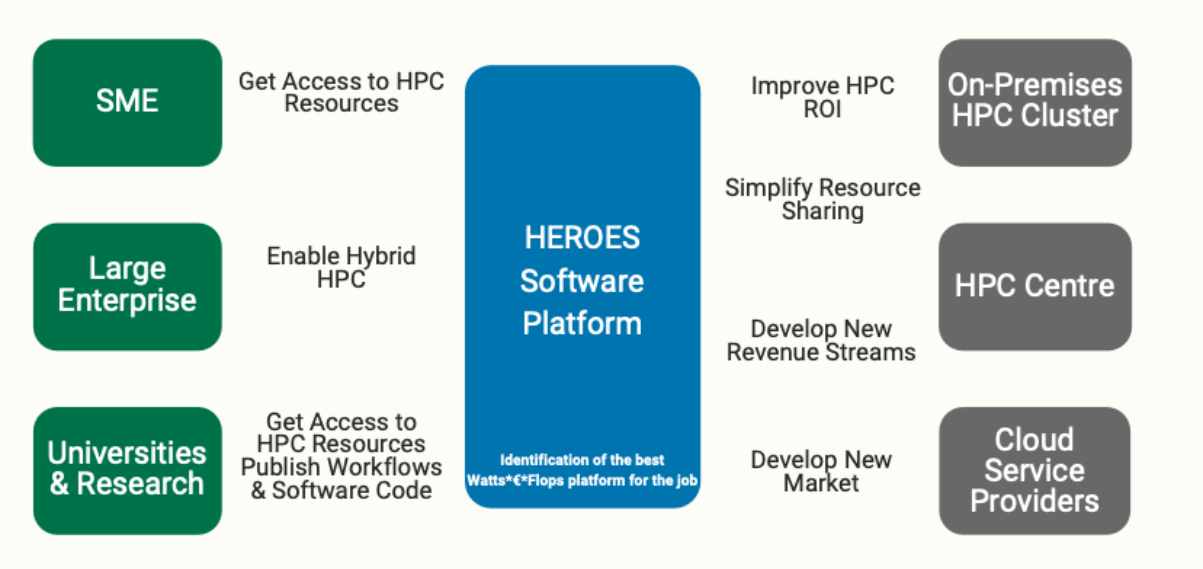

Figure 4 -The hybrid cloud/HPC marketplace concept with HEROES as a key component of the Administrator block

3.3 HEROES hybrid scenarios from demand and offer perspectives

Building on the latter static description, we now define the two main dynamic models of interactions between the users and vendors on our hybrid HPC marketplace by applying both our analysis of the hybrid HPC market developed in Section and the consortium experience in providing HPC services as well as its knowledge of the EU HPC environment.

From the HPC users' perspective: dynamically access to a multiplicity of HPC resources

On this first scenario a HPC user wants to access HPC resources that will cope with its needs, with three main parameters being identified as:

- The expected performance (what is the dimension of the infrastructure and the time to result)
- The price
- The energy consumption

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

As written above, HPC users are heterogeneous and so are their needs which may vary over time. Performance could be key for some workflows while optimizing the price could be another option that a same user may want to choose over time. Hence, apart from the set of rules defined by the HPC provider, the users will have internally defined their own policy on how they want to access the resource. Such policies would, for instance, defined who is in charge to establish the priorities when a workflow is submitted or to define a rule by default ”always maximise energy consumption, then performance, then price” as well as other subsets “prioritise submission on the organization on-premise cluster, then on the cloud”.

This **flexibility** is at the center of current and expected HPC usages in the near future and therefore a HPC access service shall address it both from on the technical and administration sides. The figure below exemplifies this case with a large client which relies both on on-premises cluster that is completed by a double access to a Public Cloud and a HPC Service Provider.

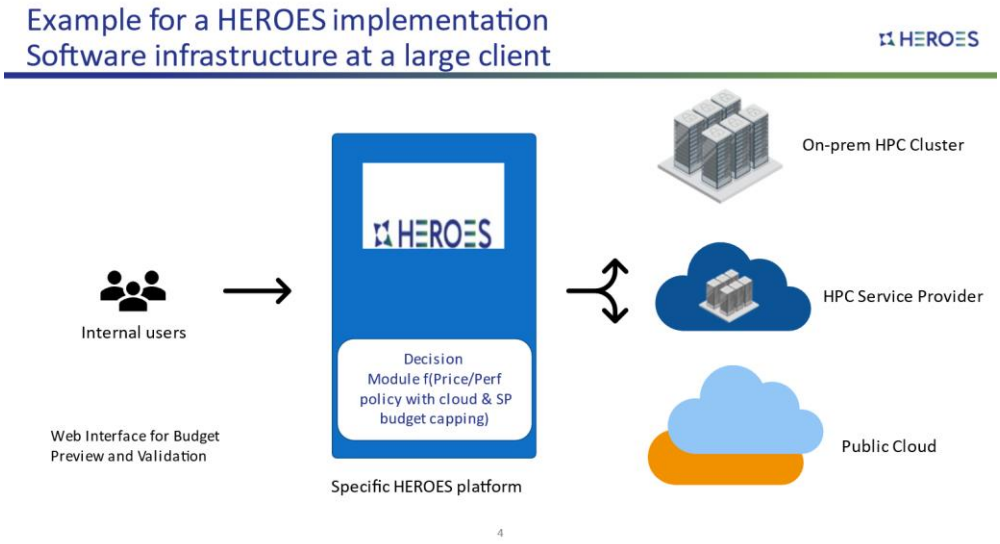

Figure 5 - Marketplace implementation 1 - HPC users perspective

From the HPC providers: dynamically provide access to HPC resources for a multiplicity of HPC users

On the other side, the HPC providers want to give access to their available resources and by doing so also at conditions that may vary over time and affecting the main three key points of performance, price and energy.

Variation of the conditions to access the resources is also a necessary flexibility for the infrastructures – for instance a HPC center may open or close computer hours depending on its run workplan, or it could allow emergency computing for a higher price…
Hence, to deal with such a variation and sells the access to its resources, the HPC provider must on one hand exchange information on the current state of its infrastructure and analyse information from a variety of users that want to access the infrastructure.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

Moreover, a HPC provider may not only propose a single cluster on the market but could perfectly be interested to package its offer with several machines, each of them having its own sub-set of constraints and characteristics.

The figure below exemplifies this case for a scenario where EuroHPC would give access to a part of its computing resources to any user where the goal would be to minimize energy consumption.

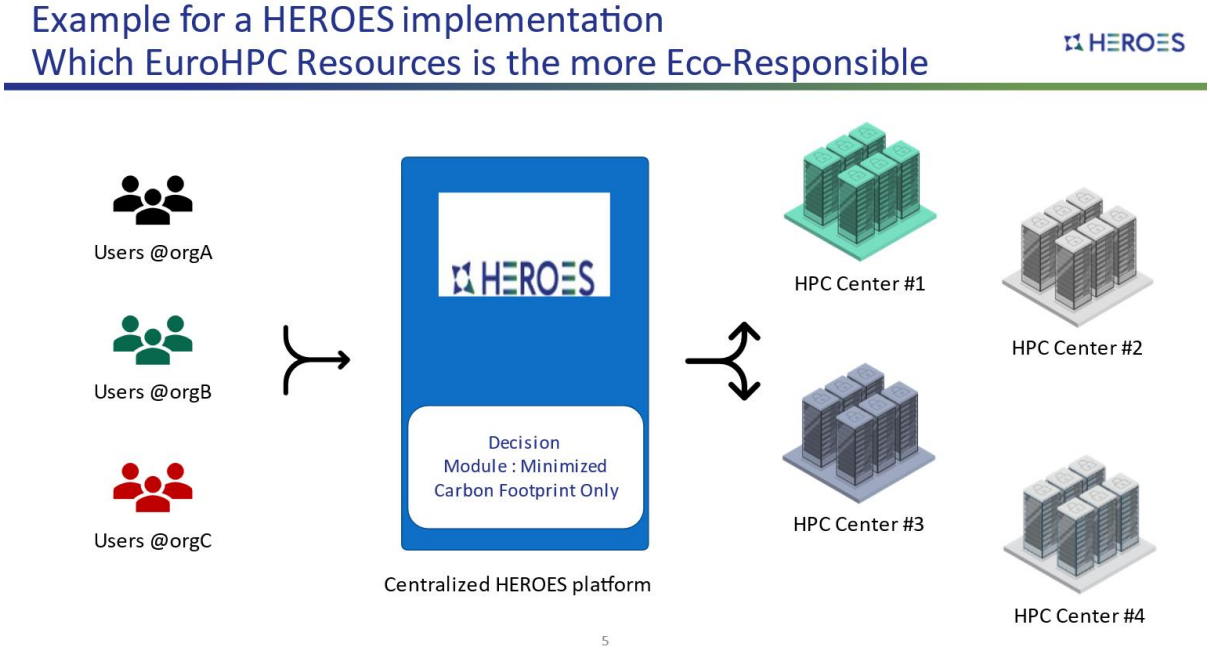

Figure 6 - Market place implementation 1 - HPC providers perspective

A (HEROES) broker at the center of dynamic interactions

In the HEROES vision, the flexibility is provided by a broker/intermediary put between the Users and the Provider. Such broker is able to describe in near real time the status of the infrastructure and the provisional conditions of a run which allow the users to make a choice (or to apply a predefined choice).

The HEROES platform is a prototype of such a broker aiming to provide an immediate assessment of the conditions under which a job could be submitted based on the current state of the infrastructure it covers.

As a conclusion of this chapter, the 2 dynamic models describe the key requirements of the hybrid HPC marketplace that led to the creation of the HEROES business survey – part 4 – and therefore to the creation of the HEROES business model scenarios – part 5 of this document.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

4 Feedbacks from the market/users and decision on a business model to exploit HEROES assets beyond the project

4.1 Description of the HEROES survey

The HEROES business survey aims to gain feedback from the market and users of a HPC hybrid marketplace based on the two scenarios described in part 3.

Collecting feedbacks from users was really important so that we are able to exploit HEROES assets beyond the project and nurture the follow-on of the initiative and the best options to develop the HEROES prototype into a viable commercial solution. The survey allowed the consortium to refine its market assumptions and prioritize the main features to be offered for the potential customers of the HEROES platform. It gives us information on the field based on real customer use cases and projections on their own access model that complete our description of the marketplace provided in part 3 of this document.

Moreover, gaining such feedback has also been one of the key point highlighted by the external experts during the project intermediate review.

The survey has been built on 9 blocks as follows:

- Affiliation information
- HPC Infrastructure Typology
- HPC Infrastructure Usage
- Energy Impact
- Cost
- HEROES functionalities
- Essential features
- Business
- Respondent information

Those 9 blocks aimed to gather information on the two scenarios and the critical features and usages of a HEROES broker.

The survey has been disseminated from November 2022 to March 2023 by several means: on key HPC events (Supercomputing, Euro HPC Summit, ISC), online dissemination and direct contacts. The communication and dissemination campaign on the survey is described in details in the HEROES deliverable 5.5 – Dissemination report.

In the end 277 answers were received but with a lot of them being automated spams and therefore not exploitable. The consortium was able to identify 16 contributions formally

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

identified as valuable with a wide variety of respondents (HPC provider such as BSC, large firm, universities, large research organisations and one SME).

4.2 Results of the HEROES survey

The results of the survey showed interesting tendencies that we described below for each bloc.

HPC Infrastructure typology

Regarding the bloc "HPC infrastructure typology", Most of the infrastructure for compute of are heterogenous (have both CPU+ GPU / accelerators). Only a few have homogenous infrastructure with mostly CPU and only one have an infrastructure composed of both.

HPC Infrastructure usage

We also asked respondents if their activities involved the use of the cloud for HPC workloads. A small majority declared that yes it did while a few organisations didn't.

Energy impact

The question of energy consumption of HPC activity is very interesting. A majority of users are aware of the energy consumption and/or carbon footprint of their HPC activity. Furthermore, most of them would be interested in finding out more about this subject.

Costs

Costs are also a central issue of the HPC activity. A strong majority of respondents is aware of their overall cost of HPC activity. The tendency keeps going at a job level where respondents are mostly aware of their costs. At project level a bit less people are aware of it even though it is still a majority. However almost all the respondents would be interested in finding out more on this topic.

HEROES functionalities

One of the survey's objectives was to determine the importance of some HEROES aspects. We translated the results for the two most important aspects (according to the survey) in the graphs that you can find below:

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

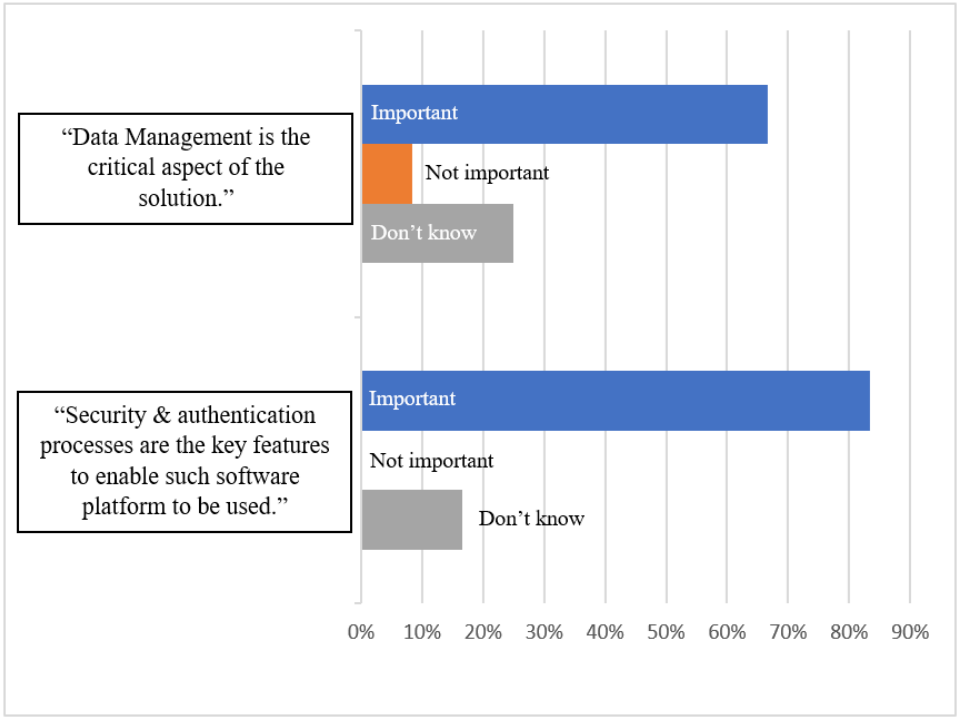

Figure 7- Results of the HEROES Survey- main expected functionalities

The survey also stated that a majority of respondents think that the ability to submit workflows easily through a web portal is important whereas approximately 20% consider that it is not very important. Regarding the following statement: "Energy is our most important challenge within European time frame", the results are quite alike with a bit more people considering it not really relevant. We also asked about the importance of the growing need of overall orchestration of HPC traffic in their organisation. Half of respondents declared it as important, 20% as not very important. However, almost 30% did not pronounced themselves. The approximatively same results can be constated when asked about the decision module as a critical piece to increase HPC efficiency.

Essential features

Moreover, all respondents think that the framework needs to be available in a "modular way" with most components available on open source and that the components need to be supported or interchangeable. Almost everyone that responded to the survey thinks that they need a skilled service organisation to integrate the components in their environment and build and support our HEROES platform including open-source components.

Business

Regarding the following statement "as a Computing Centre we are interested to open our infrastructure in a secure way to SMEs and/or HPC Research projects and/or Local startups needing HPC to address challenges related to global warming" the answers are quite divided with almost as many yes and no to the question whether it should apply or not. Finally, a majority of people agree with the fact that "we can't rely on the tools from a single vendor

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

(Cloud Vendor, Platform Vendor, Software Vendor) this is why the HEROES approach resonates to us".

4.3 Conclusions on the survey

As expected, the exploitable answers to our survey demonstrate both the heterogeneity of stakeholders needs as well as needs for flexibility in using a tools to access hybrid HPC. Moreover, the respondents demonstrated that there is a clear case to develop EU-based solutions on workflow orchestration on heterogeneous infrastructure and the need to orchestrate distant access on the infrastructure side.

The survey validated the market needs that were identified at the beginning of the project as it also copes with ongoing roadmaps for different stakeholders. The identification of the main features to be deployed give us insights on how an offer could be prepared and packaged to be sold. Based on those results, the next chapter will describe what kind of services and associated business models could be put in place for the exploitation of a HEROES platform.

5 HEROES business model scenarios

Among the deliverables from the HEROES project is the prototype of a Software Platform enabling the deployment of Hybrid HPC. Following the intermediate review, the decision was taken by the project partners for the prototype to target the use case of private clients building their Hybrid HPC and completing their on-premises clusters with Cloud capacity. During ISC22, we presented the HEROES poster to a few large private users of HPC and held a specific workshop with Airbus. Hybrid HPC don't come off-the-shelf as a product. This was understood from 2019 when the project proposal was defined and acknowledged during these workshops.

The following roadmap for implementation is a direct result from these sessions and its initial version validated through interviews conducted with the HEROES survey.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

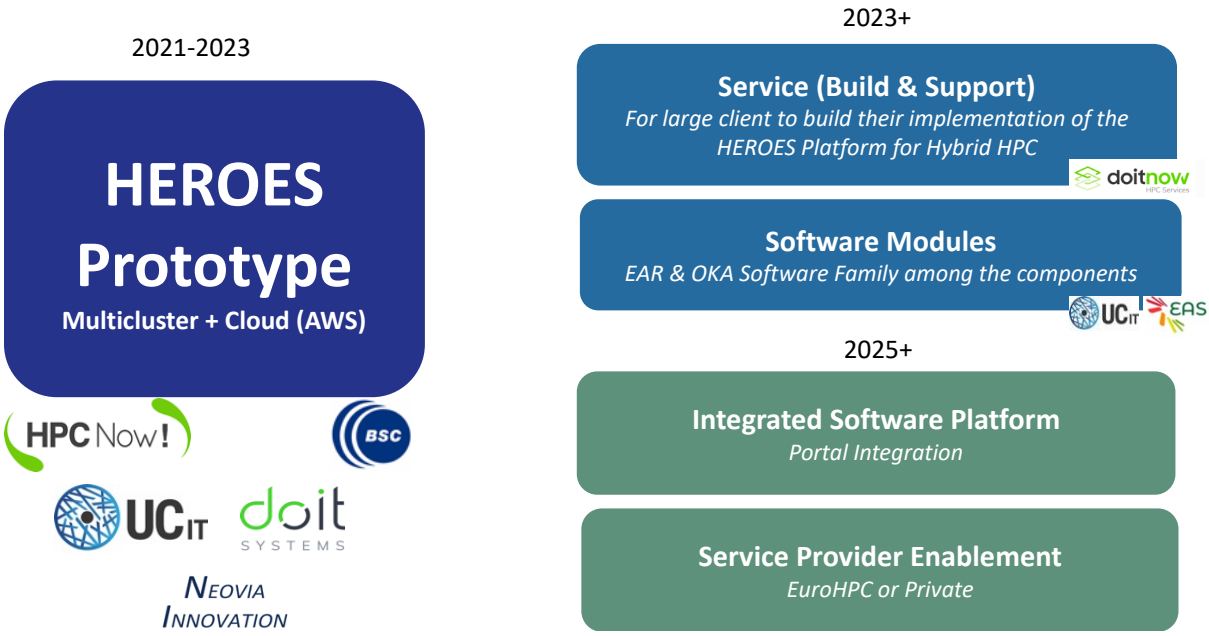

Figure 8 - HEROES exploitation plan expected roadmap

HEROES was conceived from the start as a modular architecture, defined with a set open APIs making it a framework with interchangeable components. In order to get an operable platform, the specific requirements of a client (workflow typology, internal standards, specific cloud platform to connect, specific software components to interconnect…) need to be considered driving the need for Services. In addition, a number of HEROES software components are available in the market along with Services needed to deploy, integrate and support them.

As the funding for the project was reduced, the integration with an HPC portal has been delayed. This is however a key element to deliver a replicable solution and a key piece for an evolution of the platform. In the following chapters in this section, we go through these different models.

5.1 Business model A: Service (Build, Support, Run)

The underlying concept of HEROES were documented by UCit in 2017 in its “HPC Wind Tunnel” R&D initiative. The software platform was then called “Hybridator” and embarked most of the function now part of HEROES. However, implementations are specific to each organization which builds its own path to Hybrid HPC Cloud driving the need for a broad HPC service expertise. This was reflected in the HEROES proposal (4Q2019) and lead 3 of the project partners to join forces one year later (creation of “Do IT Now” by Do IT, HPCNow! And UCit). When the project kicked-off a European HPC Service company already existed to assist the HPC ecosystem to customize, integrate, extend, deploy and support the project results and its software components (no exclusivity for any software components).

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

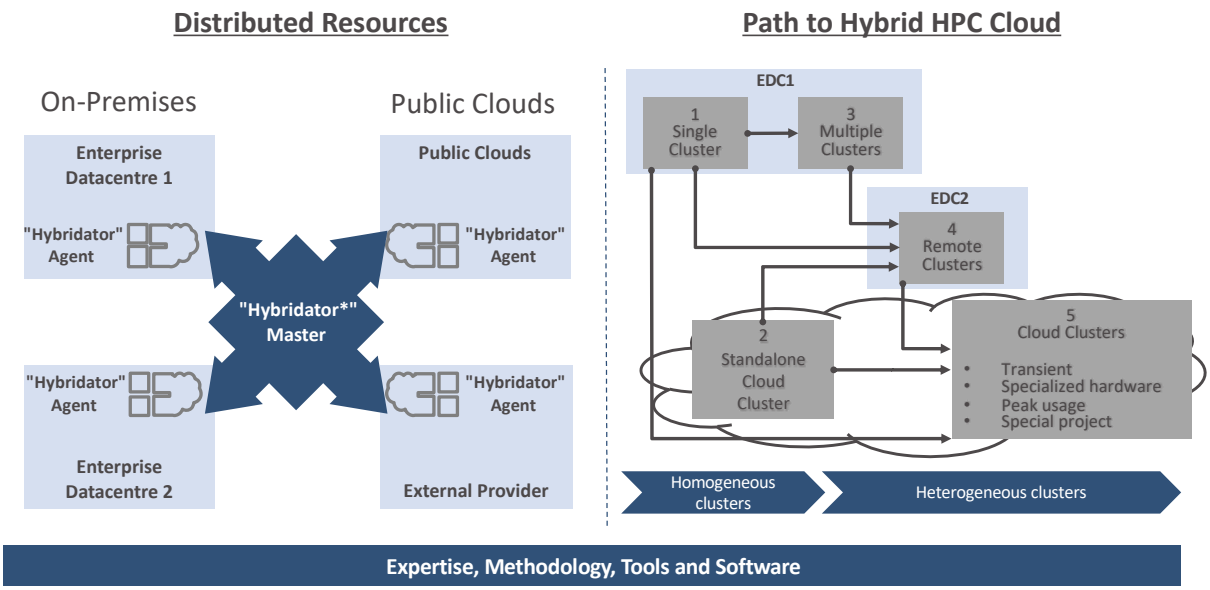

Figure 9 Expertise, methodology, tools and software

Most European HPC users are currently extending in production their on-premises clusters with public cloud access (mostly but not exclusively provided by AWS and Azure, the following abstract delivered by Hyperion at SC22[3] illustrates the main trends. Public Cloud continues to be mostly seen as a complementary resource and we have chosen for the HEROES prototype one such scenarios which represent a large portion of the alliance activity. Although there are some cloud-bursting activities, decisions are mostly taken per workload and Cloud Services mostly proposed as alternatives.
Most clients are building roadmaps towards HEROES-like solutions and have started or starts this journey with subsets of the solution and the HEROES conversation is a strong added value to support a Service Business. On top of the HEROES components, some of which are Open Source the Support of the software stack and the specific integration at one customer is a solid model.

5.2 Business model B: Software

Among HEROES modules, at least two technologies come from project partners and are readily available.:

- OKA (https://oka.how; https://doc.oka.how/index.html) is a Data Science platform for HPC environments and a licensed software from UCit (Perpetual Licenses, Software Subscriptions and Support Services). Some OKA modules will implement feature-based licensing in 2023 for Cloud-based licensing and distribution in market places.
- EAR (https://www.bsc.es/research-and-development/software-and-apps/software-list/ear-energy-management-framework-hpc) is an Open Source software framework from BSC optimizing the energy and efficiency of an HPC cluster. As most open-source software, the business model for the framework is based on support. EAS (Energy

[3] https://hyperionresearch.com/wp-content/uploads/2022/11/Hyperion-Research-SC22-HPC-Cloud-Update.pdf

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

Aware Solutions S.L.) is a spin-off of BSC and University Polytechnic of Catalonia providing installation, configuration training and support for the EAR software and partner of UCit.

As the HEROES platform evolves with deeper portal and workflow integrations, an Open-Core approach is currently evaluated (fully functional software for free with premium features to be defined).

5.3 Business model C: Service Provider

HEROES is in 2023 an orchestration platform initially available to be adapted to HPC users as a component of their Hybrid HPC environments in private contexts. There is a case to build services for an organization to use the platform in order to connect workflows with compute capacity. Specific terms and conditions will have to be defined with the service provider operating the HEROES platform as the platform will be critical to the service. The Service Provider will then potentially charge the users with a SaaS or a consumption-based model depending on the sourcing of the computing resources.

Applied to the EuroHPC context the 3 following contexts could lead to such implementations:

- Sharing EuroHPC resources (Public Researchers) with simplified allocation policies
- Giving access to EuroHPC centers to Industrial users (SMEs or Large clients)
- Creating a revenue stream for unused capacity

The implications of the latter business model is described more in-depth in the next chapter dedicated to HEROES follow-ons.

6 Description of other potential HEROES follow-ons

While business scenarios have been described in part 5, the HEROES consortium wanted to highlight other potential follow-on for the developed services that could take part of the deployment and use of the European HPC capacities in the coming years.

This vision is based on the HEROES platform playing the role of one of the building blocks in the European Continuum Computing.

As stated in the chapter 2, the cases for industrial and academic users are different and they answer to different constraints and needs and therefore would lead to the implementation of different mechanisms of “Infrastructure-As-A-Service”. That is why, in this vision chapter, we will distinguish on one part the general case of a hybrid access for the federation of European supercomputing service and on the other part what could be a dedicated industrial access to European supercomputing service.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

6.1 A hybrid access Marketplace for the federation of European supercomputing services

The vision and mission of the EuroHPC JU, as published in its multi-annual strategic plan 2021-2027[4] is, among others, to develop a connected federation of supercomputers "that is accessible to researchers from academia, industry (including SMEs) and the public sector".
Such an infrastructure "will provide a federated and coordinated access to all European supercomputers, data repositories, knowledge and will also be the place where users will get access to the latest future technology to support innovate solutions."

From the HEROES consortium perspective, such a vision is intertwined with the forecoming development of the HPC in the digital continuum – one of the main direction in HPC evolution as highlighted by the latest version of the European HPC Strategic Research Agenda (SRA) released by ETP4HPC in October 2022[5].
In this latest version of the SRA, ETP4HPC develops, among others, the technical needs that would support the deployment of the HPC in the digital continuum in which the HEROES project mostly belongs.

In particular, p48 and 49 of the SRA, the following path is desribed on achieving the Transcontinuum: "At the middleware level, a major challenge is to facilitate ***the development of tools for seamless deployment, orchestration, scheduling, and execution of complex workflows across hybrid, heterogeneous CC infrastructures***. This in turn calls for: Dynamic scheduling and orchestration of workflows which evolve at runtime, *optimizing time and energy to solution on a dynamically evolving system*. Support of a wide variety of processors, accelerators, storage devices and systems, and communication systems. A CC infra- structure will be deeply heterogeneous, and very likely so even within each of its layers."
By combining the two visions, we envisage that the deployment of European HPC federations will require building blocks such as HEROES – or a more advanced version of it – that would interface with other tools to manage seamless access to potentially both academia and industry.

The HEROES consortium has already identified ideas that could be integrated in such a system, in particular while dealing with so-called "smart contracts" that would allow the users to dynamically access available resources at flexible conditions while ensuring a clear access policy and rules. Such smart contracts could, for instance, be deployed through a token system that would also be used to define the prices of the resources at a given time and under changing conditions so that the "access fee" are constantly updated on the access marketplace.

[4] https://eurohpc-ju.europa.eu/system/files/2022-03/EuroHPC%20JU%20Decision%2024_2021-%20Approving%20the%20Multi-Annual%20Strategic%20Plan%202021-2027.pdf

[5] Downloadable at : https://www.etp4hpc.eu/pujades/files/ETP4HPC-SRA5_2022_web.pdf

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

As we believe HEROES could be part of the solution, we would work on integrating the future consortia that will be in charge to develop the EuroHPC federation – whether through calls for projects and dedicated tenders.

As an illustration of potential solutions under development that HEROES could interface with, we can cite the PUHURI solution[6] under development by the Nordic e-Infrastructure Collaboration that aims to deliver distant access to the LUMI HPC and data centres capacities. The PUHURI consortium has been approached by the HEROES partners during the 2023 EuroHPC Summit. We strongly believe that such emerging solutions could be leveraged by the JU to realise its vision of a EU HPC federation.

6.2 The HEROES platform for industrial users

The second case for a potential follow-on would be a "narrower" version of the EU HPC federation that would only target industrial users. This focus could be necessary to deal with the specificities and particular needs of academia on one side and industry access on the other.

In this focused version, part of the EuroHPC super computing capacities could be dedicated to industry usage through a distant and seamless access different from academic attribution that could follow the more traditional calls for proposals organisation.

Should this path be chosen, the HEROES technology would also remain one of the building blocks of a European solution to create the technical and commercial environment that would support the access to industrial users.

[6] https://puhuri.io/about

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.

7 Conclusion

At the end of the project, the HEROES consortium has assessed and reviewed the assumptions under which our envisaged exploitation plan was defined at the beginning of the project. The main trend for the development and need of flexible hybrid HPC access has been confirmed both from the institutional side (ETP4HPC SRA5) and potential consumers (HEROES Survey). Based on those conformation and the results attained at the end of the project, three main business model scenarios have been defined for short-term exploitation of the HEROES assets, each scenario corresponding to needs of specific stakeholder in the HPC value chain. Thanks to the joint venture Do It Now and the EAS startup, partners of the project are willing to define a commercial offer that will exploit those scenarios and effectively deliver HEROES services in the European market.

Finally, we identified two main follow-ons in the context of the implementation of EuroHPC JU vision of a HPC federation in which HEROES could be exploited and further developed.

The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.