H2020-EINFRA-2015-1

VI-SEEM

VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean

Deliverable D3.5

Final infrastructure overview and assessment report

Author(s): Dusan Vudragovic (editor)

Status –Version: Final

Date: September 25, 2018

Distribution - Type: Public

Abstract: Deliverable D3.5 – "Final infrastructure overview and assessment report," gives an overview of deployed VI-SEEM infrastructure resources and services, as well as details on usage and performance of the infrastructure.

The VI-SEEM Consortium consists of:

GRNET	Coordinating Contractor	Greece
CYI	Contractor	Cyprus
IICT-BAS	Contractor	Bulgaria
IPB	Contractor	Serbia
NIIF	Contractor	Hungary
UVT	Contractor	Romania
UPT	Contractor	Albania
UNI BL	Contractor	Bosnia-Herzegovina
UKIM	Contractor	FYR of Macedonia
UOM	Contractor	Montenegro
RENAM	Contractor	Moldova (Republic of)
IIAP-NAS-RA	Contractor	Armenia
GRENA	Contractor	Georgia
BA	Contractor	Egypt
IUCC	Contractor	Israel
SESAME	Contractor	Jordan

The VI-SEEM project is funded by the European Commission under the Horizon 2020 e-Infrastructures grant agreement no. 675121.

Document Revision History

Date	Issue	Author / Editor / Contributor	Summary of main changes
June 20th, 2018	a	Dusan Vudragovic	Initial version of ToC.
July 9th, 2018	b	Dusan Vudragovic	Final version of ToC.
August 25th, 2018	c	Anastas Mishev, Mihajlo Savic, Neki Frasheri, Stiliyan Stoyanov, Todor Gurov, Nicolas Liampotis, Athanasios Katsadas, Konstantinos Kangelidis, Kostas Merentitis, Antun Balaz, Petar Jovanovic, Dusan Vudragovic	Draft version with the input from the partners
September 3rd, 2018	d	Dusan Vudragovic, Antun Balaz	First draft
September 20th, 2018	e	Dusan Vudragovic, Antun Balaz	Final draft
September 25th, 2018	f	Dusan Vudragovic	Final

Table of contents

References

[1] Project VI-SEEM-675121 – Annex I – Description of the action

[2] VI-SEEM D3.1 – Infrastructure and services deployment plan

[3] VI-SEEM D3.3 – Infrastructure overview, assessment and refinement plan

[4] VI-SEEM D3.2 – Service registry, operational and service level monitoring

[5] VI-SEEM D3.4 – VRE AAI Model and compatibility with other e-Infrastructures

[6] TOP500 Supercomputer Sites
https://www.top500.org/

[7] European Grid Infrastructure
https://www.egi.eu/

[8] VI-SEEM Virtual Organization VOMS-admin interface
https://voms.ipb.ac.rs:8443/voms/vo.vi-seem.eu

[9] gLite User Guide
https://edms.cern.ch/document/722398/

[10] 1st Call for proposals for projects accessing VI-SEEM resources and services
https://vi-seem.eu/1st-call/

[11] 2nd Call for proposals for projects accessing VI-SEEM resources and services
https://vi-seem.eu/2nd-call/

[12] 3rd Call for proposals for projects accessing VI-SEEM resources and services
https://vi-seem.eu/3rd-call/

[13] Call for SME proposals for projects accessing VI-SEEM resources and services
https://vi-seem.eu/sme-call/

[14] ARGO Service Monitoring system
http://argoeu.github.io/

[15] A. Shrivastwa, S. Sarat, K. Jackson, et al. OpenStack: Building a Cloud Environment, Packt Publishing (2016).

[16] Official GitLab webpage
http://www.gitlab.com/

[17] Official MediaWiki webpage
https://www.mediawiki.org/

[18] FitSM – standards for lightweight IT service management
https://fitsm.itemo.org/

[19] Open Science Cloud EOSC-hub
https://www.eosc-hub.eu/

[20] H2020 project e-InfraCentral
http://einfracentral.eu/

[21] Webalizer, free web server log file analysis program
http://www.webalizer.org/

[22] eduGAIN federation
https://edugain.org/

[23] AARC Blueprint Architecture
https://aarc-project.eu/architecture/

List of Figures

List of Tables

Glossary

AAI	Authentication and Authorization Infrastructure
API	Application Programming Interface
BDII	Berkeley Database Information Index
CMDB	Configuration Management Database
CPU	Central Processing Unit
GOCDB	Grid Operations Centre Database
GPU	Graphics Processing Unit
HPC	High Performance Computing
HTTPS	HyperText Transfer Protocol Secure
IdP	Identity Provider
LB	Logging and Bookkeeping
LFC	LCG File Catalogue
PI	Programmatic Interface
PX	MyProxy software
REST	Representational State Transfer
SEEM	South East Europe and Eastern Mediterranean
SME	Small and medium-sized enterprise
SNMP	Simple Network Management Protocol
SP	Service Provider
UI	User Interface
VI-SEEM	VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean
VM	Virtual Machine
VOMS	Virtual Organization Membership Service
VRC	Virtual Research Community
VRE	Virtual Research Environment
WMS	Workload Management System
WP	Work Package
XML	Extensible Markup Language

Executive summary

What is the focus of this Deliverable?

The focus of this deliverable, as defined in the project's description of the action [1], is to provide a summary of integrated computing and storage resources, details on usage of these resources, as well as an overview of the developed and deployed services, which enable uniform management and access to these resources. The previous two versions of this document (deliverables D3.1 [2] and D3.3 [3]) were focused on implementation details, while this document is more oriented towards the usage and statistics on the provided support.

What is next in the process to deliver the VI-SEEM results?

This document describes the final shape of the deployed infrastructure, its usage, and statistics on the provided support. Such information is crucial for the efficient use of the VI-SEEM services, and essential for all aspects of the project. In particular, the outcomes of this deliverable will be used during the implementation of the upcoming pay-per-use offer, which includes the development of the bundle catalogue to attract customers coming from the private market/industrial sector.

What are the deliverable contents?

The deliverable contents include an overview of the infrastructure available to the VRE, and deployed infrastructure services that cover the needs of application services and users. In particular:

- Infrastructure resources – it gives a technical specification of the project's infrastructure constituent sites. For each type of resources, we present the total, dedicated, allocated, and used amount of resources, as well as support provided to end users. Several significant upgrades of the infrastructure were carried out during the project lifetime, and are reported in this section (Section 2).
- Usage and performance – details related to the usage and performance of the VI-SEEM services from the point of view of production calls are given (Section 3).
- Operations and resource management – gives an overview and describes usage of the deployed tools that facilitate interoperability of the infrastructure, and efficient management of the available computing and storage resources (Section 4).
- Service catalogue/portfolio and management – describes contents and usage of the system that supports the service portfolio management process and service discovery (Section 5).
- Authentication and Authorization Infrastructure (AAI) – presents management of user access to the services and related statistics (Section 6).

Conclusions and recommendations

In the last three years, the VI-SEEM consortium supported 66 production projects submitted by researchers and the SMEs from the region. The fact that 75% of these projects requested access to at least two different kinds of available resources (HPC, Cloud, Grid, Storage), goes in favour of the heterogeneous nature of our infrastructure, although it was a management challenge to the project's operations team. The main achievements and conclusions from this process are presented in this deliverable, and can be summarized as follows:

- Early access to the infrastructure was crucial for the realization of the 1st implementation phase, which happened in project month 5 (February 2016). By then, more than 80% of the projected size of the infrastructure was integrated into the VI-SEEM infrastructure, which enabled the development of the services offered a year later in the 1st call for production use of resources and services. The rest of the infrastructure resources were integrated gradually, and the final capacity includes 23,694 CPU-cores, 1,166,592 GPU-cores (516 GPU cards), 20,496 Xeon Phi-core (336 Xeon Phi cards), 3,112 Grid CPU-cores, 14,152 Cloud VM-cores, and 18 PB of storage space. The total size of the infrastructure matches the defined key performance indicators.
- The VI-SEEM consortium supported 66 projects, 23 of them related to climate research, 24 to life sciences, 16 to digital cultural heritage, and 3 of them were cross-disciplinary. In total, 41 projects requested access to the HPC CPU, 15 to HPC GPU, 4 to HPC Xeon Phi, 18 to Cloud, and 3 to Grid resources. The overall allocation of HPC resources was 22.3 million CPU-core hours, 2.6 million GPU-card hours, and 2.2 million Xeon Phi-core hours. Additionally, 70,000 CPU-core hours were used through the Grid interface, as well as 225 Cloud VMs. The total storage space consumption by all projects on HPC storage systems amounts to 95.2 TB. In addition to this, approximately 10 TB of storage space is occupied at the Simple storage service, 3 TB of storage space at the Archival service, and in total 3 TB of storage space at the application- and data-specific services, which include the Data repository service, the Live access server, and the Clowder.
- A set of operational and resource management tools were deployed for the efficient management of the available computing and storage resources. All of them are fully integrated with the project’s AAI. Although we created a centralized support system for the project's needs, majority of actual support cases were handled outside of this system. This is primarily caused by the fact that users can get much faster and more direct help by the support staff via commonly used instant communication systems. Therefore, in the corresponding section, we have collected and provided statistics that include various support channels, from direct email messages to holding a multipoint video conferences or a remote computer access, sidestepping the use of the classic support system.
- The service catalogue/portfolio management system contains information about the services offered to the customers. It is fully integrated with the project’s AAI, and could be used by other infrastructures as well. In total, 25 services are offered through the catalogue and grouped into six categories.
- The VI-SEEM Login service enables researchers to access the VI-SEEM infrastructure resources in a user-friendly and secure way. The service integrates 14 service providers, and 2,824 identity providers. In addition to this, we also integrated 3 social identity providers, namely Google, Facebook and LinkedIn. The VI-SEEM Login service was used for over 16,000 federated logins at the moment.

1. Introduction

The VI-SEEM WP3 team coordinated the development and deployment of the infrastructure and service in the project during the last three years. This work followed a detailed plan designed at the very beginning of the project (project month 4), which was described in the deliverable D3.1 [2].Working through the four primary tasks of WP3:

- Task 3.1 – Service registry and service level definition,
- Task 3.2 – Implementation of resources and service provisioning,
- Task 3.3 – Operations and resource management of the e-Infrastructure scientific communities,
- Task 3.4 – Authentication, authorization and access management,

the project integrated the available computing and storage resources, and developed and deployed a collection of the operational services for the uniform administration of those and users' access provisions.

From the management point of view, this can be seen as a set of horizontal and vertical activities. The horizontal ones ensured the provision of a unified underlying e-Infrastructure layer on top of the existing network links among the countries that participate in the project, while the vertical activities covered the development of a service catalogue/portfolio for the efficient management of project's services, as well as Authentication and Authorization Infrastructure to enable a uniform management and resource access provisioning. The additional bidirectional task supported site managers in operation of integrated services on top of the underlying infrastructures, the maintenance of common user services, and users in their day-to-day usage of the infrastructure.

The integrated computing and storage resources were heterogeneous: HPC resources with low-latency interconnection, Grid sites for high throughput calculations, Virtual Machines in the Cloud, and various types of storage resources. However, most of them (more than 80%) were made available in project month 4 (January 2016), while the rest of the infrastructure resources were integrated gradually, by project month 10 (July 2016). The early access to the infrastructure was crucial for the realization of the 1st WP5 implementation phase, which happened in project month 5 (February 2016). In collaboration with the WP5 team, ten applications started in February 2016 with the development of the services using the projects' infrastructure.

Chronologically, starting from February 2016, the WP3 focused its main effort on the development of operational services: GOCDB, source code repository, trouble ticketing system, and technical wiki, as well as the development of the service catalogue/portfolio information system. The aim was to build a production environment before the announcement of the 1st call for proposals for projects accessing the VI-SEEM resources and services in October 2016. Five (out of 6 planned) of project's operational and resource management tools became operational in project month 9 (June 2016), and the service portfolio management system was launched in project month 12 (September 2016). The deliverable D3.2 [4] documents the details of the service registry/portfolio, while the deliverable D3.3 [3] reports technical details related to the deployment of operational and resource management tools.

The project's second year (October 2016) started with an allocation of resources for the applications from the 1st call for production use of resources and services, and continued with support to the accepted applications. The first production use of the virtual research environment by these applications included further fine-tuning of the environment and deployment of several additional software modules. In parallel, within the task 3.4, the AAI team deployed previously designed AAI model - the VI-SEEM Login service. Established as a central identification service, the VI-SEEM Login identifies, authorizes and authenticates VI-SEEM users to allow their access to the offered project's services. The service became operational in project month 17 (March 2017), and is technically documented in the deliverable D3.4 [5].

During the last year of the project, the WP3 team performed a technical evaluation of proposals from the 2nd and 3rd call for production use of resource and services. The team also performed resource allocation and provided support to selected projects. Also, we have supported 6 SME proposals, finalized the project's monitoring system, and participated in the creation of a cost model for all services towards an open pay-per-use call. In particular, the WP3 team performed a price analysis of popular on-demand HPC, Cloud, and storage resources: Amazon EC2, Google Cloud, and Azure. Based on this, for each resource center within the VI-SEEM infrastructure, we have identified the corresponding on-demand general-purpose instance independently on EC2, Google, and Azure. This approach resulted in creation of an approximate resource center usage price to be offered through the upcoming pay-per-use call.

The deliverable D3.1 listed infrastructure available to the virtual research communities and the plan for services deployment, while the implementation details were reported in the deliverable D3.3. The deliverable D3.2 described project's service catalogue/portfolio, and the deliverable D3.3 AAI solution - the VI-SEEM Login service. The focus of this document is on the usage and provided support statistics.

2. Infrastructure resources

The infrastructure of the VI-SEEM project consists of resources of various types: HPC resources - clusters and supercomputers with different hardware architectures, Grid sites, Clouds with a possibility to launch virtual machines (VMs) for services and distributed computing, and storage resources with a potential for short- and long-term storage. Figure 1 illustrates the geographical distribution of these resources. The heterogeneous nature of the infrastructure presents several management challenges to the project's operations team, but is also an advantage for the users because of its ability to support different types of applications, or different phases of the same project. This is in alignment with the fact that 75% (49 out of 66) of all production projects that are granted access to the infrastructure asked for at least two different kinds of available resources.

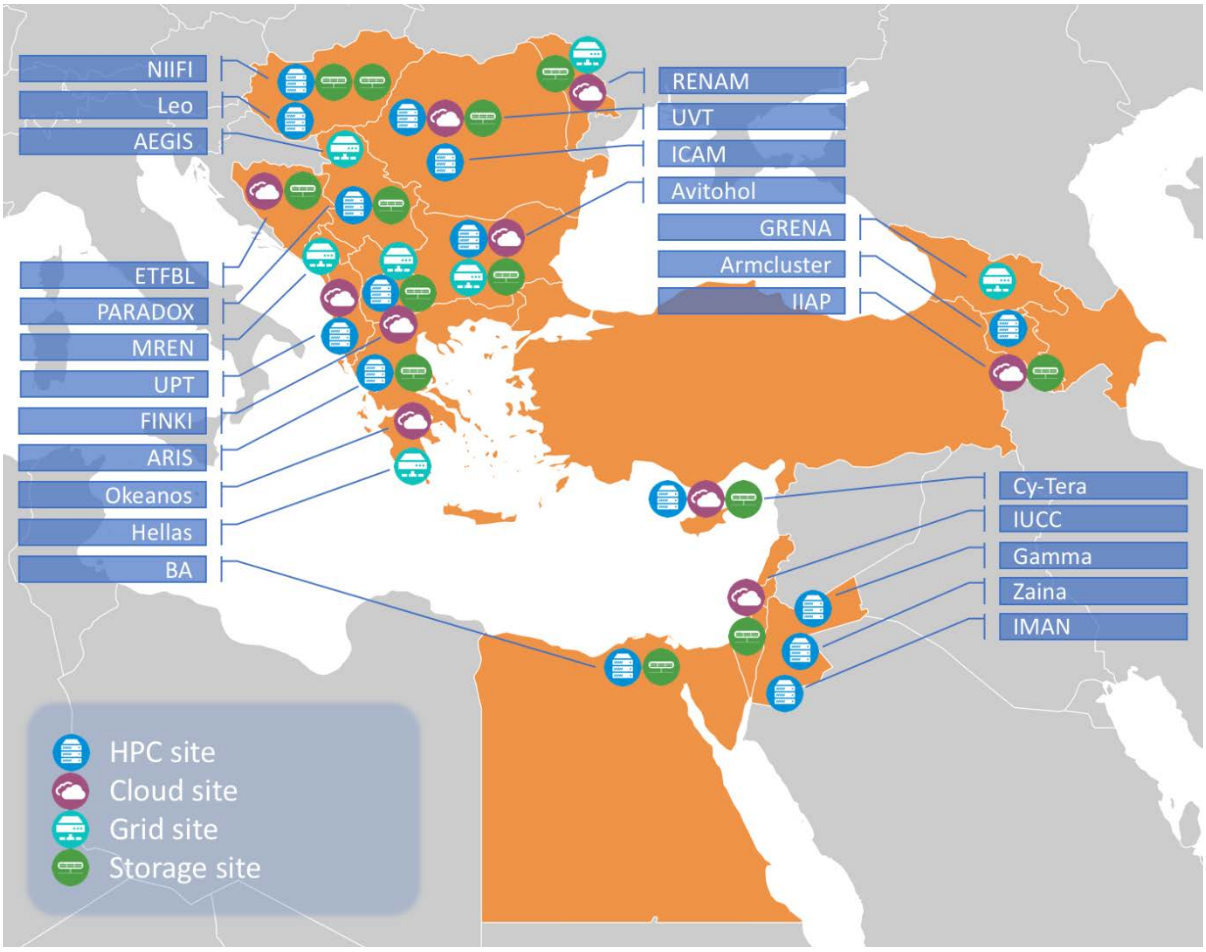


Figure 1: Geographical distribution of VI-SEEM resource centers.

The key performance indicators related to our e-Infrastructure, and defined in the project description, are related to the size of the infrastructure: 21,500 CPU-cores, 325,000 GPU-cores, and 18,500 Xeon Phi-cores of HPC, 2,900 Grid-cores, 10,500 Cloud VM- cores, and 11 PB of storage space. The performance indicators also define the infrastructure dedication to the project as 5-15% of HPC, 10-15% of Grid, 5% of Cloud, and 10% of storage resources. Up to this point, the VI-SEEM integrated 23,694 CPU-cores, 1,166,592 GPU-cores (516 GPU cards), 20,496 Xeon Phi-core (336 Xeon Phi cards), 3,112 Grid CPU-cores, 14,152 Cloud VM-cores, and 18 PB of storage space. The dedication to the project is as follows: 11% of CPU-cores, 7.5% of GPU-cores, 9.5% of Xeon Phi-core, 10% of Grid CPU-cores, 4% of Cloud VM-cores, and 3% of storage space. The total size of the infrastructure matches the defined key performance indicators. The dedicated amounts (percentages of dedicated resources) do not perfectly match the defined key performance indicators (in terms of Cloud and storage resources), but as it will be shown in the following sections, they are based on real resource demands by the projects from the calls for production use of the resources and services, and all user requests are satisfied.

The deliverable D3.1 - "Infrastructure and services deployment plan" gives a detailed technical specification of the project's infrastructure constituent sites. As the resource centers evolve, to keep their specs up-to-date, a set of wiki pages has been created – one page per resource center. In this section, we provide links to these pages. Also, for each type of resources, we presented total, dedicated, allocated, and used amount of resources, as well as provided support in a number of issues reported by users, the average number of communications per problem, and the average time spent on solving an issue.

2.1. HPC resources

The HPC resources of the project consist of clusters with low-latency interconnection and supercomputers listed in Table 1. Most of the systems are based on CPUs with the x86_64 instruction set, some of them are equipped with accelerators, but there is also a BlueGene/P system, as well as one based on the Cell processors (PS3 cluster IMAN1-Booster/King). Two of these resources, Avitohol from Bulgaria and Leo from Hungary, were present at the November 2016 Top500 list [6] of supercomputers (at 388th and 402nd place, respectively), while another one, ARIS from Greece, was present at the June 2015 Top500 list (at 468th place).

Resource	Country Code	Wiki URL
ARIS	GR	https://wiki.vi-seem.eu/index.php/HPC_Resources/ARIS
Cy-Tera	CY	https://wiki.vi-seem.eu/index.php/HPC_Resources/Cy-Tera
Avitohol	BG	https://wiki.vi-seem.eu/index.php/HPC_Resources/Avitohol
PARADOX	RS	https://wiki.vi-seem.eu/index.php/HPC_Resources/PARADOX
NIIFI SC	HU	https://wiki.vi-seem.eu/index.php/HPC_Resources/NIIFI_SC
Leo	HU	https://wiki.vi-seem.eu/index.php/HPC_Resources/Leo
InfraGRID	RO	https://wiki.vi-seem.eu/index.php/HPC_Resources/InfraGRID
ICAM BlueGene/P	RO	https://wiki.vi-seem.eu/index.php/HPC_Resources/ICAM_BlueGene/P
UPT-HPC	AL	https://wiki.vi-seem.eu/index.php/HPC_Resources/UPT-HPC
MK-03-FINKI	MK	https://wiki.vi-seem.eu/index.php/HPC_Resources/MK-03-FINKI
Armcluster	AR	https://wiki.vi-seem.eu/index.php/HPC_Resources/Armcluster
BA-HPC	EG	https://wiki.vi-seem.eu/index.php/HPC_Resources/BA-HPC
Gamma	JO	https://wiki.vi-seem.eu/index.php/HPC_Resources/Gamma
Zaina	JO	https://wiki.vi-seem.eu/index.php/HPC_Resources/Zaina

Table 1: Wiki pages with detailed description of HPC resources.

The largest number of resource centers were integrated into the project infrastructure at the project month 4 (January 2016), while the rest of them were integrated by the project month 9 (June 2016). To date, the aggregate size of the infrastructure is 23,694 CPU-cores, 516 GPU-cards (1,166,592 GPU-cores), and 336 Xeon Phi-cards (20,496 Xeon Phi-cores). Its peak performance measured in TFlops reaches 1.4 PFlops, with CPU-cores contributing 815 TFlops, GPU-cores 192 TFlops, and Xeon Phi-cores 405 TFlops.

A summary information about the available resources and their dedication to the project are shown in Table 2. The amount of resources dedicated to the project needs per year is 22,8 million CPU-core hours, 1.3 million GPU-card hours, and 0.3 million Xeon Phi-card hours. Based on these numbers, we distributed the dedicated resources in the calls for their production use. Having in mind the overlap between the calls, as well as their overlap with the integration phases, we managed to match the use of resources with the offered capabilities. Therefore, in the 1st call we allocated 10.2 million CPU-core hours, 1.2 million GPU-card hours, and 0.1 million Xeon Phi-card hours. In the 2nd call 2.2 million CPU-core hours and 1.2 million GPU-card hours were allocated. In the 3rd call we allocated 9.4 million CPU-core hours, 0.2 million GPU-card hours, 0.1 million Xeon Phi-card hours. Finally, in the SME call we allocated 0.5 CPU core-hours, 0.02 GPU card-hours, and 0.02 Xeon Phi card-hours. In total, the allocation of HPC resources for the production use was: 22.3 million CPU-core hours, 2.6 million GPU-core hours, and 0.2 million Xeon Phi-core hours. From all production calls 44 out of 66 projects requested access to the HPC resources, 41 of them requested the CPU time, 15 the GPU time, and 4 requested the Xeon Phi time.

Regarding the provided support, 221 HPC-related issues were reported in total by the users. The average number of communications per ticket was 6, and the average resolution time per ticket was 47 hours.

Resource	Country Code	Total			Dedicated (hours per year)		
		CPU-cores	GPU-cards	Phi-cards	CPU-hours	GPU-hours	Phi-hours
ARIS	GR	8,520	94	36	4,731,760	1,041,297	15,768
Cy-Tera	CY	1,392	24	-	1,829,088	31,536	-
Avitohol	BG	2,400	-	300	2,102,400	-	262,800
PARADOX	RS	1,696	106	-	742,848	46,428	-
NIIFI SC	HU	768	-	-	336,384	-	-
Leo	HU	1,344	252	-	588,672	110,376	-
InfraGRID	RO	4,096	7		798,912	12,264	-
ICAM	RO	456	-	-	7,176,192	-	-
UPT-HPC	AL	144	-	-	560,640	-	-
FINKI	MK	768	-	-	336,384	-	-
Armcluster	AR	128	-	-	112,128	-	-
BA-HPC	EG	1,960	32	-	3,433,920	56,064	-
Gamma	JO	8	1	-	70,080	8,760	-
Zaina	JO	14	-	-	36,792	-	-
	Total	23,694	516	336	22,856,200	1,306,725	278,568

Table 2: Available HPC resources and their dedication to the project.

2.2. Cloud resources

A list of available Cloud resources is given in Table 3, together with a link to the corresponding wiki page with the technical description of each particular Cloud center.

Resource	Country	Description URL
~Okeanos	GR	https://wiki.vi-seem.eu/index.php/Cloud_Resources/Okeanos
CyI Cloud	CY	https://wiki.vi-seem.eu/index.php/Cloud_Resources/CyI_Cloud_Facility
Avitohol	BG	https://wiki.vi-seem.eu/index.php/Cloud_Resources/Avitohol
InfraGRID Cloud	RO	https://wiki.vi-seem.eu/index.php/Cloud_Resources/InfraGRID_Cloud
UPT-Cloud	AL	https://wiki.vi-seem.eu/index.php/Cloud_Resources/UPT-Cloud
ETFBL-CC01	BA	https://wiki.vi-seem.eu/index.php/Cloud_Resources/ETFBL-CC01
MK-04-FINKI_CLOUD	MK	https://wiki.vi-seem.eu/index.php/Cloud_Resources/MK-04-
MD-RSC	MD	https://wiki.vi-seem.eu/index.php/Cloud_Resources/MD-Cloud
IIAP Cloud	AR	https://wiki.vi-seem.eu/index.php/Cloud_Resources/IIAP_Cloud
IUCC InfinityCloud	IL	https://wiki.vi-seem.eu/index.php/Cloud_Resources/IUCC_InfinityCloud

Table 3: Wiki pages with detailed description of Cloud resources.

The Cloud resources were integrated sequentially starting from the beginning of the project, and at the project month 10 (July 2016) the number of virtual machines within the infrastructure was 14,152. The number of virtual machines dedicated to the project needs is 484, which corresponds to 4,239,840 VM-core hours. Table 4 lists the Cloud infrastructure details.

In total, 18 projects (out of 66) from the calls for production use of resources required the Cloud infrastructure. In the 1st call, the consortium allocated 82 VMs for six projects, in the 2nd call 74 VMs for additional six projects, and in the 3rd call 69 VMs for another six projects. Participants in the SME call have requested no VMs. So, at the moment, 225 VMs are consumed by 18 projects.

Regarding the provided support, 120 Cloud-related issues in total were reported by the users. The average number of communications per ticket was 15, and the average resolution time per ticket was 23 hours.

Resource	Country Code	Total VMs	Dedicated VMs	Dedicated VM-hours
~Okeanos	GR	10,000	200	1,752,000
CyI Cloud	CY	176	18	157,680
Avitohol	BG	2,400	120	1,051,200
InfraGRID Cloud	RO	400	46	402,960
UPT-Cloud	AL	12	6	52,560
ETFBL-CC01	BA	60	13	113,880
MK-04-FINKI_CLOUD	MK	436	24	210,240
MD-Cloud	MD	12	19	166,440
IIAP Cloud	AR	96	10	87,600
IUCC InfinityCloud	IL	560	28	245,280
	Total	**14,152**	**484**	**4,239,840**

Table 4: Available Cloud resources and their dedication to the project.

2.3. Grid resources

The Grid resources within the project are provided mostly by smaller clusters that have already installed the Grid middleware and also share their resources within the European Grid Infrastructure [7]. Some of the clusters that are not certified at the level of the EGI were incorporated in the VI-SEEM Grid infrastructure by deploying the modified BDII services. Technical details of project's Grid infrastructure sites were given in the deliverable D3.1 - "Infrastructure and services deployment plan". The clusters evolve over time, and to keep their specification up-to-date a set of wiki pages has been created – one page per resource center. Grid site administrators maintain these pages, and Table 5 lists their URLs.

Resource	Country Code	Description URL
Hellas Grid	GR	https://wiki.vi-seem.eu/index.php/Grid_Resources/Hellas_Grid
BG01-IPP	BG	https://wiki.vi-seem.eu/index.php/Grid_Resources/BG01-IPP
AEGIS01-IPB-SCL	RS	https://wiki.vi-seem.eu/index.php/Grid_Resources:AEGIS01-IPB-SCL
MK-03-FINKI	MK	https://wiki.vi-seem.eu/index.php/Grid_Resources/MK-03-FINKI
MREN01CIS	ME	https://wiki.vi-seem.eu/index.php/Grid_Resources/MREN01CIS
MD-GRID	MD	https://wiki.vi-seem.eu/index.php/Grid_Resources/MD-GRID
ArmCluster	AR	https://wiki.vi-seem.eu/index.php/Grid_Resources/ArmCluster
GE-01-GRENA	GE	https://wiki.vi-seem.eu/index.php/Grid_Resources/GE-01-GRENA

Table 5: Wiki pages with detailed description of Grid resources.

A summary information about the available Grid resources and their dedication to the project is shown in Table 6. These resources were integrated by the project month 5 (February 2016). In total, 299 dedicated Grid CPU-cores are provided to scientific communities.

A project-dedicated Virtual Organization (vo.vi-seem.eu) has been deployed at the IPB, and its web interface, VOMS-admin is available at [8]. Besides the virtual organization management service, a set of Grid core services (VOMS, BDII, LFC, WMS, LB, PX) are deployed at IPB, IICT-BAS, and GRNET. A functional description of Grid core services from a user's point of view is given in the gLite User Guide [9].

Resource	Country Code	Total CPU-cores	Dedicated CPU-cores	Dedicated CPU-hours
Hellas Grid	GR	864	43	370,000
BG01-IPP	BG	640	120	1,051,200
AEGIS01-IPB-	RS	704	50	438,000
MK-03-FINKI	MK	768	38	336,384
MREN01CIS	ME	32	16	140,160
MD-GRID	MD	40	12	105,120
GE-01-GRENA	GE	64	20	175,200
Total		**3,112**	**299**	**2,616,064**

Table 6: Project's Grid resources and their dedication to the project.

During project lifetime, three projects used the VI-SEEM Grid resources in a production mode: the VINE project from Georgia in the 1st call, the OPERA-P project from Jordan in the 2d call, and the DESPHOTMAT project from Greece in the 3rd call. The utilization of the Grid infrastructure in all three projects is approximately 70,000 CPU-core hours.

Regarding the provided support, 9 Grid-related issues were reported in total by the users. The average number of communications per ticket was 4, and the average resolution time per ticket was 19 hours.

2.4. Storage resources

The storage resources are usually considered in relation to the respective type of computing resources, e.g., the HPC storage should be available to the HPC jobs and the Grid storage should be available to the Grid jobs. Such short-term storage resources are described on wiki pages listed in Table 7. The storage resources described here are provided for long-term or other special types of data storage.

Resource	Country Code	Description URL
ARIS	GR	https://wiki.vi-seem.eu/index.php/Storage_Resources/ARIS
Cyprus	CY	https://wiki.vi-seem.eu/index.php/Storage_Resources/ONYX,_Cy-Tera
Avitohol	BG	https://wiki.vi-seem.eu/index.php/Storage_Resources/Avitohol
PARADOX	RS	https://wiki.vi-seem.eu/index.php/Storage_Resources/PARADOX
NIIFI HSM	HU	https://wiki.vi-seem.eu/index.php/Storage_Resources/NIIFI_HSM_Service
NIIFI iSCSI	HU	https://wiki.vi-seem.eu/index.php/Storage_Resources/NIIFI_iSCSI_Service
UVT HPC GPFS	RO	https://wiki.vi-seem.eu/index.php/Storage_Resources/UVT_HPC_GPFS
ETFBL-CC01	BA	https://wiki.vi-seem.eu/index.php/Storage_Resources/ETFBL-CS01
MK-04-FINKI	MK	https://wiki.vi-seem.eu/index.php/Storage_Resources/MK-04-FINKI_CLOUD
RENAMstor	MD	https://wiki.vi-seem.eu/index.php/Storage_Resources/RENAMstor
IIAP Storage	AR	https://wiki.vi-seem.eu/index.php/Storage_Resources/IIAP
BA-IA	EG	https://wiki.vi-seem.eu/index.php/Storage_Resources/BA-IA
IUCC Storage	IL	https://wiki.vi-seem.eu/index.php/Storage_Resources/IUCC-InfinityCloud-

Table 7: Wiki pages with detailed description of Storage resources.

A summary of provided storage space is given in Table 8. Various storage resources were integrated sequentially starting from the beginning of the project, and starting from the project month 10 (July 2016) a full capacity, 334 TB of disk and 510 TB of tape dedicated storage space, is made available to the supported scientific communities.

All projects that were granted access to the VI-SEEM infrastructure are directly or indirectly also using the storage space. Directly, they store their data via a low-level interface to a particular storage site, and indirectly via high-level data services that mask the complexity of the infrastructure. The total storage space consumption by all projects is approximately 83 TB. In addition to this, about 10 TB of storage space is occupied at the Simple storage service, 3TB at the Archival service, and 3 TB for the application- and data-specific services, which include: the Data repository service, the Live access server, and the Clowder.

Resource	Country Code	Total		Dedicated	
		Disk [TB]	Tape [TB]	Disk [TB]	Tape [TB]
ARIS	GR	1,000	3,000	50	210
Cyprus	CY	500	-	100	-
Avitohol	BG	96	-	5	-
PARADOX	RS	250	-	10	-
NIIFI HSM	HU	87	6,800	3	300
NIIFI iSCSI	HU	1,000	-	50	-
UVT HPC GPFS	RO	50	-	5	-
ETFBL-CC01	BA	1		1	-
MK-04-FINKI	MK	36	-	2	-
RENAMstor	MD	4	-	1	-
IIAP Storage	AR	8	-	3	-
BA-IA	EG	5,200	-	100	
IUCC Storage	IL	40	-	5	-
Total		**8,272**	**9,800**	**334**	**510**

Table 8: Available Storage resources and their dedication to the project.

2.5. Major infrastructure upgrades

Several significant upgrades of the infrastructure happened during the project lifetime:

- Deployment of the PARADOX Hadoop cluster as a data analysis service,
- BA-HPC cluster Tesla K80 dual GPU upgrade,
- Upgrade of the PARADOX storage system.

IPB deployed the PARADOX Hadoop cluster in June 2016 as a VI-SEEM data analysis service. It consists of a single name node which runs a resource manager, and three additional data nodes. The name node is hosted on a machine with a 4-core Intel Xeon E3-1220v3 CPU at 3.1 GHz, with 4 GB of RAM, and 500 GB of local hard disk storage. Each of the data nodes, which perform the computation and storage, are on machines with a 24-core Intel Xeon E5-2620 CPUs at 2.4 GHz, with 64 GB of RAM and 2 TB of storage.

In total 1,826 jobs were submitted to the PARADOX Hadoop cluster. The average execution time for a single job is 37 hours. Figure 2 illustrates additional details regarding the use of this service. Blue bars represent the execution time of all tasks in a given month, while the red line gives the corresponding number of jobs. The current use of the HDFS storage space is 2.8 TB (53%). This percentage varies, and usually, when a particular type of jobs is finalized, a user transfers results to the external storage and clears the HDFS. The PARADOX Hadoop cluster is used by 8 projects from the calls for production use of services and resources, and its total utilization is 1.3 million CPU-core hours.

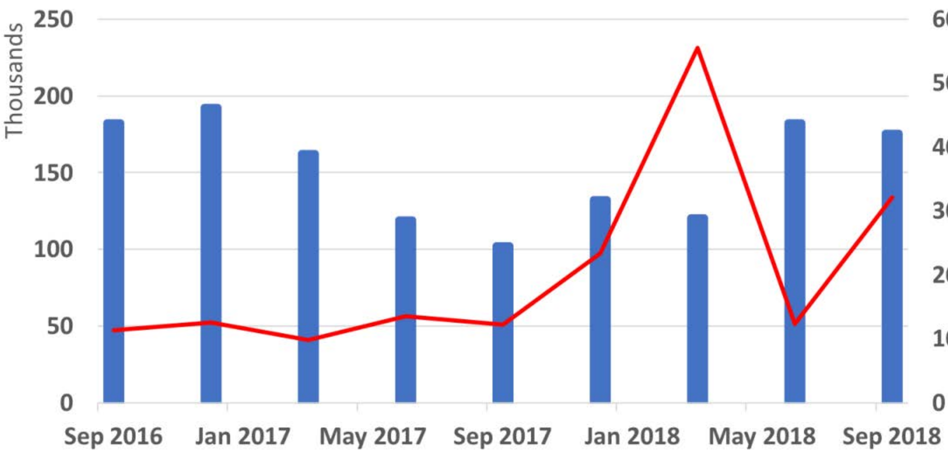


Figure 2: Total execution time (blue bars) and total number of jobs (red line) at the PARADOX Hadoop cluster for a given six month period.

The Bibliotheca Alexandrina (BA) has been operating an HPC cluster since August 2009. In late 2016, the cluster was upgraded to accommodate rising demands. The current cluster consists of 82 compute nodes, providing a total of 1,968 CPU-cores, with 5.33 GB of RAM per core. Additionally, the cluster has 16 accelerated compute nodes, each with a Tesla K80 dual GPU. A storage for input and output data is provided by a Lustre file system hosted on a storage hardware with a total raw capacity of 288 TB. The cluster is wired with 40 Gbps QDR InfiniBand.

In the summer of 2018, the PARADOX cluster at IPB received a substantial storage upgrade as a donation by the Raiffeisen Bank in Serbia. The current storage capacities, which include 96 TB of storage, will be extended by the Hitachi Adaptable Modular Storage (AMS) 2500 system, providing additional 250 TB. The installation illustrated in Figure 3 consists of two AMS 2500 Base units (each having 2 storage controllers) and 27 Expansion units, each equipped with 15x600GB SAS drives. The system interconnect technology is Fiber Channel supported by the Fujitsu Brocade DCX-4S switches, enabling 8 Gbps communication speed. The installation is housed in three server racks integrated with the previously installed liquid cooling system of the PARADOX cluster.

Figure 3: Upgrade of the PARADOX storage system.

Beside of these, several smaller upgrades happened during the project lifetime, such as upgrade of the storage system at UPT-HPC cluster to 13 TB.

3. Usage and performance

The provided infrastructure resources and services overall are, so far, mainly used through the development access, as well as through the calls for production use of resources and services. The VI-SEEM development access facilitates the development and integration of services by the selected collaborating user communities. In this process, applications are given access to the infrastructure and necessary computational resources for a six-month period, during which the application developers are expected to develop and integrate relevant services. The calls for production use of resources and services target specific communities and research groups that have already begun development of their projects. The VI-SEEM engagement with industry and external scientific communities is accelerated through the continuous SME Call. Launched in November 2017, the call supported partnerships involving SMEs and research institutions from the SEEM region, through the development of projects using the VI-SEEM services. These calls are intended for mature projects, which require significant resources and services to realize their workplans. Therefore, a considerable utilization of the VI-SEEM resources comes from the calls for production use of resources and services, as well as the SME call, and an order-of-magnitude smaller usage comes from the development access channel. Figure 4 illustrates a timeline of the project's integration phases and production calls.

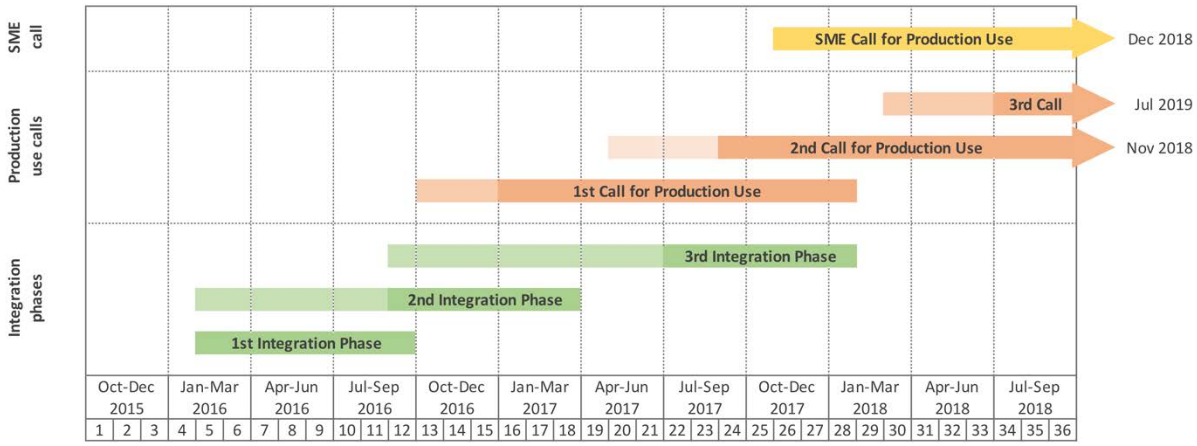


Figure 4: Timeline of the VI-SEEM integration phases and production calls.

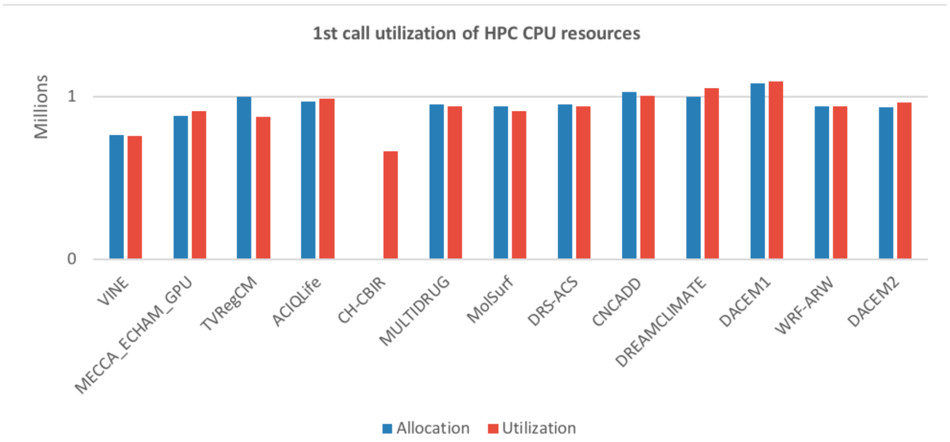
The VI-SEEM supported 66 projects, 23 of them related to climate research, 24 to life sciences, 16 to digital cultural heritage, and 3 of them were cross-disciplinary. In total, 41 projects had requested access to the HPC CPU, 15 to HPC GPU, 4 to HPC Xeon Phi, 18 to Cloud, and 3 to Grid resources. The overall allocation of HPC resources was 22.3 million CPU-core hours, 2.6 million GPU-card hours, 2.2 million Xeon Phi-core hours. Additionally, 70,000 CPU-core hours were used through the Grid interface, as well as 225 Cloud VMs. The access to storage resources was granted to all projects, and the total storage space consumption by all projects on HPC storage systems amounts to 95.2 TB. In addition to this, approximately 10 TB of storage space is occupied at the Simple storage service, 3TB of storage space at the Archival service, and in total 3 TB of storage space at the application- and data-specific services, which include the Data repository service, the Live access server, and the Clowder.

In this section, we present details related to the usage and performance of the VI-SEEM resources from the point of view of production calls.

3.1. Usage and performance by the 1st call projects

After a detailed technical and a brief scientific evaluation, the project's consortium accepted 21 applications (projects) in total in the 1st call for the production use of the VI-SEEM resources and services [10]. The majority of the projects (14 out of 21) requested access to the HPC, six of them requested access to the Cloud resources, one of them (VINE application from Georgia) needed the Grid infrastructure, and one project (AUTH_WRF371M_EUR0.44 application from Greece) was granted access to the VI-SEEM repository service only. During the access phase, the DIOPTRA application developers changed the scope of their project and decided to withdraw from the use of awarded resources. Therefore, 20 applications in total have used the allocated infrastructure resources.

Figure 5 illustrates the utilization of allocated HPC CPU and GPU resources. Blue bars represent the amounts of allocated CPU resources, while red bars show the corresponding usage per application. The applications accepted in the first call significantly utilized the allocated resources, up to 98% of the allocation. In absolute numbers, the projects consumed 10 out of the assigned 10.2 million CPU-core hours. During this call, the applicants overestimated their need for GPU resources, which resulted in the consumption of just about 52% (650,000 GPU-card hours) of the allocated resources. On the other hand, the requests for the Xeon Phi-card hours have turned out to be underestimated. With the consent of our colleagues from the IICT-BAS (Avitohol cluster), the initially allocated resources were significantly increased (6 times, from 1 to 6.3 million Xeon Phi-core hours) and utilized by the ACIQLife project. In case of the CH-CBIR application, although initially only the GPU resources were allocated, additional CPU resources became necessary to the project. In these and similar circumstances, the project's consortium approved additional, reasonable amounts of resources.



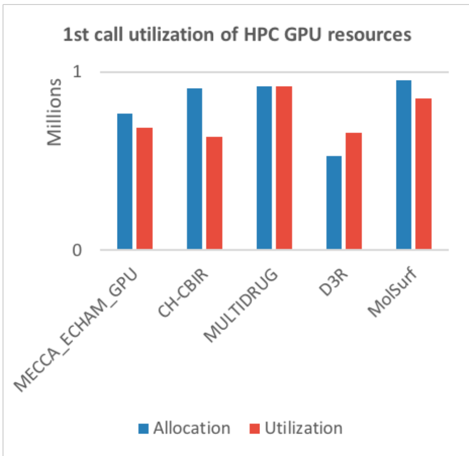


Figure 5: Utilization of the allocated HPC CPU (top) and GPU (bottom) resources by the 1st call projects.

The total utilization of the allocated Cloud resources stands at approximately 95%, and the first call projects consumed 78 out of 82 allocated VM-cores. Figure 6 illustrates the assigned (in blue) and used (in red) numbers of VM-cores per project.

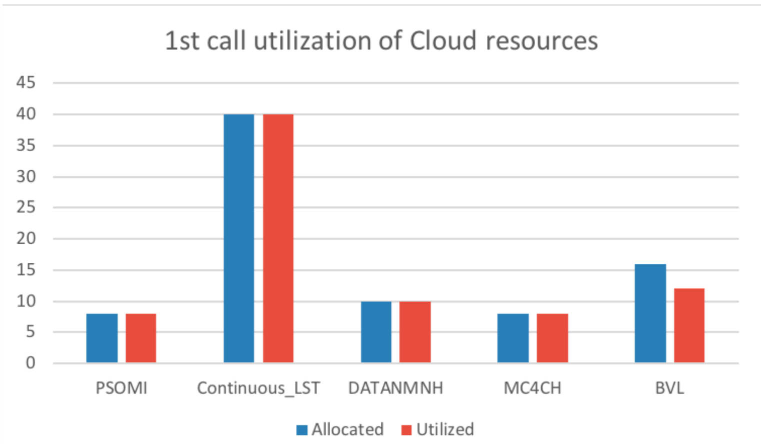


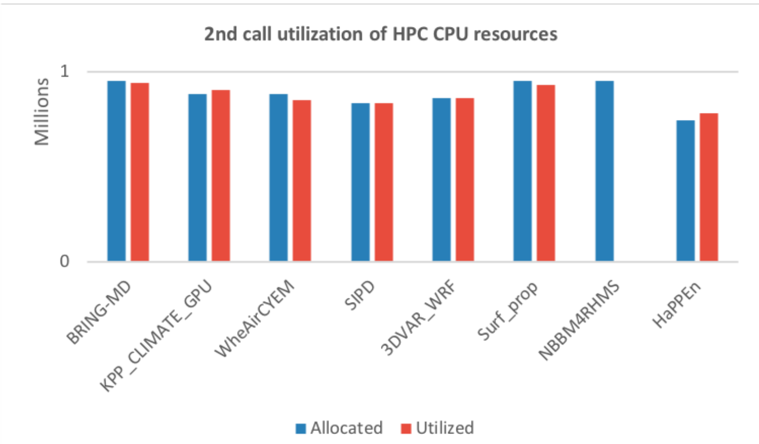
Figure 6: Utilization of the allocated Cloud resources by the 1st call projects.

Only one project from the first call for production use of the VI-SEEM resources and services used the Grid infrastructure (VINE from Georgia). Through this interface, the project consumed around 36,000 CPU-core hours.

3.2. Usage and performance by the 2nd call projects

The VI-SEEM consortium accepted 17 applications out of 19 submitted in the 2nd call for production use of the VI-SEEM resources and services [11]. As it was the case in the previous call, the consortium selected projects based on the results of their technical and scientific evaluation. Eight projects in total requested access to the HPC resources, six of them asked for access to the Cloud resources, while one of them (OPERA-P application from Jordan) needed the Grid infrastructure. Seven projects required access to the VI-SEEM Simple storage service, three to the Data analysis service, two to the Live access service, three to the Clouder, and one project to each of the Repository, Archival, ChemBioServer, and AFMM service.

In total, 2.2 million CPU-core hours, 1.1 million GPU-card hours, and 74 virtual machines were allocated for the 2nd call projects. Figure 7 illustrates utilization of the allocated HPC CPU and GPU resources, and Figure 8 the numbers of VM-cores per project. Blue bars represent the amounts of allocated resources, while red bars show the corresponding usage.



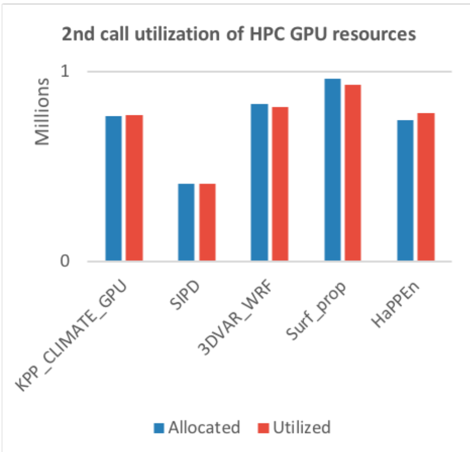


Figure 7: Utilization of the allocated HPC CPU (top) and GPU (bottom) resources by the 2nd call projects.

To date (August 2018), the total usage of the allocated HPC CPU resources is at 70%, for the HPC GPU resources it is at 75%, and for the Cloud VMs at 85%. In absolute numbers, the 2nd call projects consumed 1.5 million CPU-core hours, 0.8 million GPU-hours, and 63 virtual machines. The call is still active, and we expect additional utilization by November 2018, when it will end. One project (NBBM4RHMS from FYRoM) has faced problems with the model deployment, which required additional porting and development effort, and resulted in a delay of production use. The principal investigator of the NBBM4RHMS project informed us about this issue, which has been resolved by now, and in the upcoming three months we expect the full utilization of the allocated resources to this project.

One project from the second call for production use of the VI-SEEM resources and services used the Grid infrastructure (OPERA-P project from Jordan). The project consumed around 34,000 of CPU-core hours.

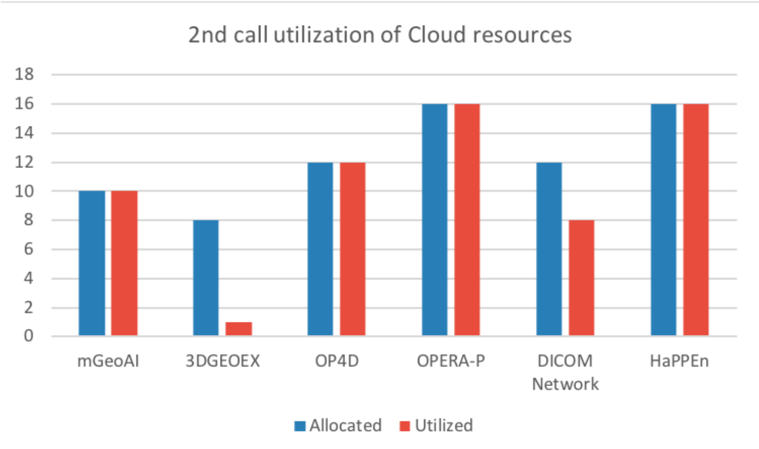


Figure 8: Utilization of the allocated Cloud resources by the 2nd call projects.

3.3. Usage and performance by the 3rd call projects

Following the two successful calls for project access to the VI-SEEM services and the associated infrastructure, the VI-SEEM opened its 3rd call [12] in February 2018 addressing scientists and researchers that work in academic and research institutions in the region in the fields of life sciences, digital cultural heritage, climate, as well as cross-disciplinary scientific fields, unlike in the previous calls. Also, to ensure high quality of the accepted applications in this call, the consortium organized a peer-review process that involved external reviewers.

The VI-SEEM consortium accepted 22 submitted applications in the 3rd call for production use of the VI-SEEM resources and services. The majority of the projects (16) requested access to HPC, six of them asked for access to the Cloud resources, while one of them (DESPHOTMAT project from Greece) demanded the Grid infrastructure access. The access to underlying computational resources is awarded for a maximum period of 12 months, while the access to the VI-SEEM data repository service is given for up to 2 years.

Since the allocation start date for the 3rd call projects was July 2018, no significant utilization has been recorded by these projects so far. Table 9 lists applications from the 3rd call for production use of resources and services, and their distribution over the available VI-SEEM HPC resources. For each application, the amount of allocated resources is given. In total, 9.3 million CPU-core hours, 0.2 million GPU-card hours, and 1.0 million Xeon Phi-core hours are allocated.

Application Acronym	Country Code	Virtual Community	HPC Resource	Resources allocated		
				CPU-hours	GPU card-hs	Phi-hours
RNA_LUPUS	GR	LS	ARIS	100,000	-	-
EuGenia	GR	LS	ARIS	1,000,000	-	-
BSI	GR	LS	ARIS	672,000	-	-
CAT-ICE	RS	CD	BA-HPC	750,000	-	-
EXMED	GR	CR	Cy-Tera	350,000	-	-
GQL	RS	LS	PARADOX	500,000	30,000	-
OPENFOAM	AL	CR	Avitohol	672,000	-	-
AT1R	GR	LS	ARIS	750,000	1,000	-
CORDEX-FPS Ph2	GR	CR	ARIS	2,000,000	-	-
RCM-MENA-CORDEX-II	CY	CR	Cy-Tera	400,000	-	-
DESPHOTMAT	GR	DCH	Avitohol	1,000,000	-	1,000,000
CDPrSc	GR	CD	ARIS	1,500,000	-	-
CDFOLK	MK	DCH	InfraGRID	100,000	100,000	-
ArmWRF	AM	CR	Armcluster	100,000	-	-
GAMMOS	AM	DCH	NIIFI SC	512,000		
SEMaCD	RS	LS	PARADOX	100,000	100,000	-
			Total	**9,384,000**	**231,000**	**1,000,000**

Table 9: The allocated HPC resources in the 3rd call for production use of resources and services.

Table 10 lists applications from the 3rd call for production use of resources and services, and their distribution over the available VI-SEEM Cloud resources. Overall, 69 virtual machines are allocated for this purpose.

Application Acronym	Country Code	Virtual Community	Cloud Resource	Resources allocated
				Used VMs
SARISI	AL	CD	Avitohol	1
CHERE	BA	DCH	MK-04-FINKI_CLOUD	12
DESPHOTMAT	GR	DCH	Okeanos	24
CMMC	GR	LS	Okeanos	18
CDFOLK	MK	DCH	MK-04-FINKI_CLOUD	18
ArmWRF	AM	CR	IIAP Cloud	16
			Total	**69**

Table 10: The allocated Cloud resources in the 3rd call for production use of resources and services.

3.4. Usage and performance by the SME call projects

Through the SME call [13], the VI-SEEM consortium opened up possibilities for SMEs and regional scientists from the selected scientific fields to have access, via joint projects, to the resources and services we offer. The following projects have been accepted and granted access to the infrastructure:

- MAGIA (Mutation cAlling algorithm IntegrAtion, LS area), submitted by the HybridStat Predictive Analytics company from Greece, is granted access to the ARIS VI-SEEM HPC facility from May to July 2018. The allocated resources for this purpose are 100,000 CPU-core hours and 4 TB of storage space. Additionally, the project is granted 50GB of storage space at the VI-SEEM Simple Storage service.
- IntelliRent (Assessing Impact of Air Quality on Rental Prices, CR area), submitted by the Rental 2000 company from Bulgaria, is granted access to the Avitohol VI-SEEM HPC facility from May to July 2018. The allocated resources for this purpose are 100,000 CPU-hours and 100,000 Xeon Phi-core hours and 1 TB of storage space. Access to the VI-SEEM Simple Storage service (50 GB of storage space) and the VI-SEEM Data Analysis Service is provided as well.
- AOGD (Assessment of Impact of Fleet Management on Air Quality, CR Area), submitted by the ICOM COMPUTERS company from Bulgaria, is granted access to the Avitohol VI-SEEM HPC facility from June to August 2018. In total, 20,000 CPU-core hours and 200 GB of storage space are allocated for this project.
- BRFAA-NGS (BRFAA_NextGenSequencing, LS area), submitted by the IBET company from Greece, is granted access to the ARES VI-SEEM HPC facility from July to September 2018. Allocated resources for this purpose are 10,000 CPU-core hours and 5 TB of storage space.
- NPssDNA (Atomistic simulations of immune stimulatory single stranded bacterial DNA, LS area), submitted by the Ingredio company from Greece, is granted access to the Leo VI-SEEM HPC facility from August to October 2018. Allocated resources for this project are 100,000 CPU-core hours, 6,500 GPU node hours, and 2 TB of storage space. Additionally, the project is granted access to the VI-SEEM Simple Storage service (50 GB of storage space), and to the VI-SEEM Life Sciences Application Specific Service - Nanocrystal.
- COMP-SS-COMP (Computational exploration of the solid-state compatibility between active pharmaceutical ingredients and excipients by quantum chemistry, multivariate statistics and machine learning techniques, LS area), submitted by the Research & Development Alkaloid AD company from FYRoM, is granted access to the Finki VI-SEEM HPC facility from September to November 2018. Allocated resources for this purpose are 100,000 CPU-core hours and 1 TB of storage space. Access to the VI-SEEM Simple Storage service (50 GB of storage space) is provided as well.

In summary, the total amount of resources used in the SME call is 430,000 CPU core-hours, 6,500 GPU node-hours, 100,000 Xeon Phi-core hours, and 13 TB of storage space. This is a continuous call and will be open until the end of the project, October 2018.

4. Operations and resource management

The efficient management of the available computing and storage resources, as well as interoperability of the infrastructure, is achieved by a set of operational tools. Static technical information, such as names, geographical locations, contacts and downtime information, list of service-endpoints provided by a particular resource center within the infrastructure etc., is manually entered and made available through the VI-SEEM GOCDB database. Based on this information, the project monitoring system can automatically trigger execution of monitoring service probes and enable efficient access to results of the probes via a customized monitoring web portal. Using standardized metrics, the VI-SEEM accounting system accumulates and reports utilization of different types of resources. User support and service-related problems are resolved mainly through the helpdesk system, but a technical mailing list is used for this purpose as well. The VI-SEEM source code repository contains all codes developed within the project, while the technical wiki collects technical documentation, know-how, best practices, guidelines, etc. All operations and resource management tools are fully integrated with the project's authentication and authorization infrastructure. This practically means that a user can access all these tools using the very same credentials as for the regular access to the granted resources.

The deployed VI-SEEM operational and resource management tools, responsible partners, and frontend URLs are listed in Table 11, while Figure 9 illustrates their geographical distribution. Technical details on the deployment of a particular tool are fully documented in the deliverables D3.1 and D3.3. A brief overview of main features and usage of the operations and resource management tools are given in this section.

System	Responsible Partner	Country Code	URL
GOCDB	UKIM	MK	https://gocdb.vi-seem.eu/portal
Monitoring	GRNET/UNI BL	GR/BA	https://mon.vi-seem.eu/
Accounting	IICT-BAS	BG	https://accounting.vi-seem.eu/
Helpdesk	UNI BL/UOM	BA/ME	https://support.vi-seem.eu/
Source code repository	UNI BL/UOM	BA/ME	https://code.vi-seem.eu/
Technical wiki	CYI	CY	https://wiki.vi-seem.eu/

Table 11: List of the deployed operational and resource management tools.

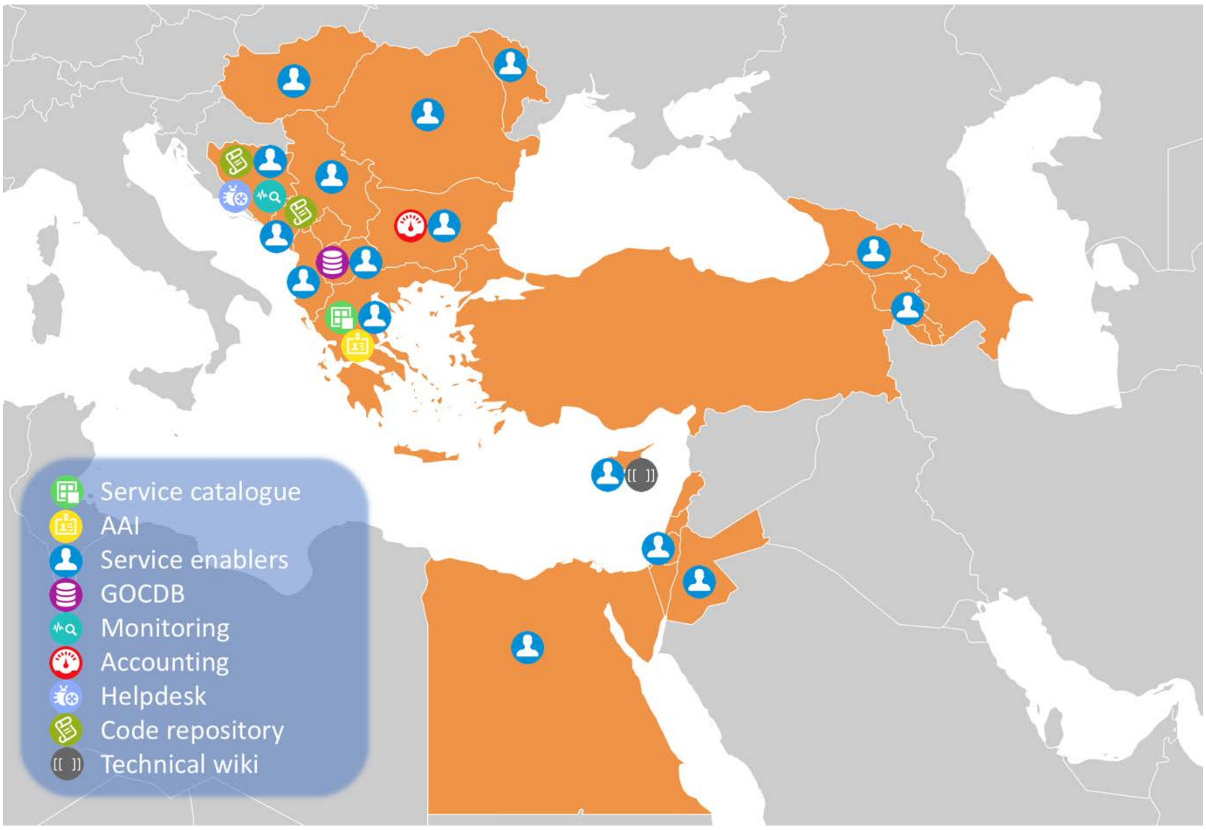


Figure 9: Geographical distribution of the VI-SEEM operations and resource management tools.

4.1. GOCDB

The GOCDB is a Configuration Management Database (CMDB) for recording and managing assets in federated environments. It defines a number of topology objects including admin-domains, sites, services, service-groups, service-endpoints, service-downtimes, users and roles. The tool provides a web portal for editing information and a REST style programmatic interface (PI) for querying data in XML. A comprehensive role-based permissions model controls user permissions. Remarkably, the project can self-manage their users: users make requests for roles over target objects and users that already hold the necessary role(s) can accept or reject those role requests.

Within the VI-SEEM, the GOCDB was deployed by the partner UKIM as a service registry, serving mainly the monitoring service of the VI-SEEM infrastructure. The instance is customized to reflect the actual needs and objects of the VI-SEEM project infrastructure and services. The VI-SEEM is defined at the level of project and at the level of NGI in the deployed instance. This allows the sites to be declared as part of the NGI, which in this case is the project itself. Project partners that are resource providers are defined as GOCDB sites. Since all the partners are also the providers, there are currently 16 sites in the database, one per each partner. Each partner is responsible for maintaining the information about its site.

The database is maintained by a total of 21 users, out of which 4 are the GOCDB admins. GOCDB registered services belong to 16 sites, one per each VI-SEEM partner. The geographical distribution of registered services is show in Figure 10 below.

Figure 10: Geographical distribution of the GOCDB registered services.

Currently, the GOCDB database contains a total of 92 production-level monitored services. The services are grouped into 37 service types, covering the grid, operations, storage and application specific services, as illustrated in Table 12.

Service area	Number of service types	Number of services
Grid computing	10	53
Application specific	12	12
HPC	1	12
Operational services	11	12
Repository and PIDs	3	3

Table 12: Production-level monitored services in the GOCDB.

4.2. Monitoring

The monitoring system used by the project consists of the three main layers: collectors, messaging and presentation. The first layer consists of collector agents that query the available data sources and transform the gathered data into standardized messages.

These messages are then asynchronously delivered via the messaging system to the presentation layer. The presentation layer can be an SNMP agent that enables the legacy protocol access to the monitoring data, an external database used for historical data collection (a time-series database), or a web-based application used for simple access to operations and historical data.

The system can collect the data from several sources including: the GOCDB service containing the project infrastructure topology, the ARGO monitoring system [14] that provides functional monitoring as well as the availability and reliability calculations, top- and site-level BDII services, OpenStack-compatible [15] IaaS sites, generic datacenter virtualization systems, etc. In order to facilitate simple inclusion of the project-level monitoring data into the local monitoring systems used by various partners, we have developed a MIB (management information base) and created the SNMP agents since practically every monitoring system has the support for SNMP. Due to the fact that all data is exchanged through standardized messages, the data can also be easily stored in external time-series database systems such as InfluxDB or OpenTSDB for further visualization and analysis with tools such as Grafana.

A proper functioning of various project services is maintained by the ARGO monitoring system. All monitored project services have been named using the eu.vi-seem prefix, which is followed by a component name designating the type of the service (app for applications, grid for Grid services, hpc for HPC services, and ops for operations services), and finally by the service name itself. For example, eu.vi-seem.app.dicom identifies the DICOM network application service. This approach enables the system to utilize both the generic test probes (reachability, HTTPS support, etc.) and the custom probes, where applicable.

The availabilities of service types for August 2018 are as follows: application-level services 96%, operations service 98%, Grid services 66% and HPC services 75%. It is worth noting that availabilities and reliabilities are calculated in such a manner that redundant system are penalized for each failing instance. For example, if there are four redundant instances of a single service and one instance is failing the tests, the service availability is calculated at 75%. This approach does produce the lower figures than end-users experience, but can help the infrastructure operators to identify problems before they become generally noticeable.

Accessing the monitoring data is possible via the Web UI available at https://mon.vi-seem.eu/. The dashboard of the monitoring system's Web UI is illustrated in Figure 11.

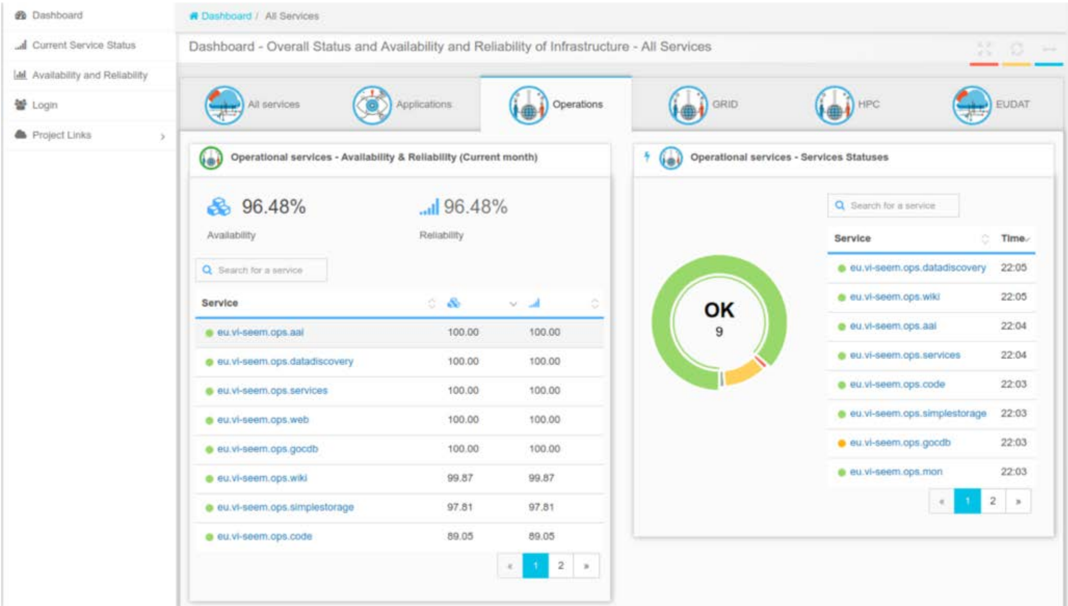


Figure 11: Dashboard of the VI-SEEM monitoring system.

4.3. Accounting

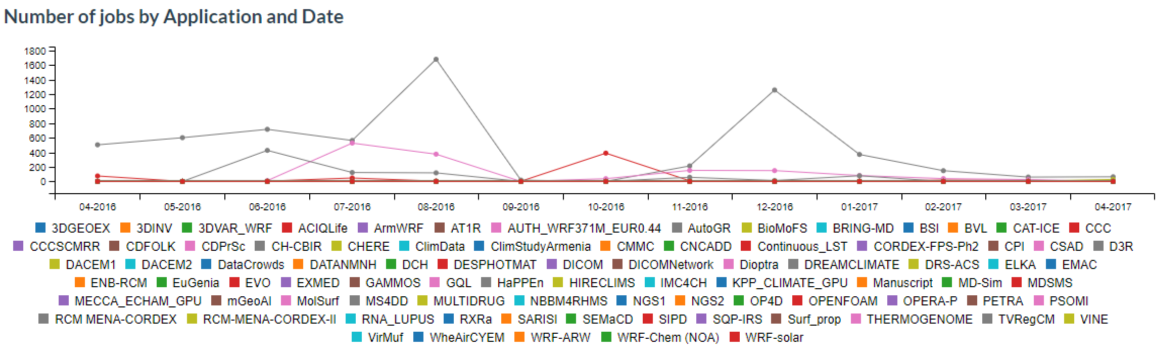
The accounting system accumulates and reports the utilization of underlying infrastructure, i.e., HPC, Grid, Cloud, and storage resources, using the standardized metrics. The service gathers and organizes accounting data in one central database. Reports from the database are made available to the community via the web interface available at https://accounting.vi-seem.eu/.

On the one hand, the accounting service stores and represents usage information on three main types of provided resources: computing, Cloud, and storage. On the other hand, all applications from the production calls are added to the database and are organized into three virtual communities, plus an additional one for cross-disciplinary projects.

The computing data includes the HPC and Grid accounting information gathered from the logged jobs by the workload managers deployed by each resource center available in the VI-SEEM project. The Cloud and storage accounting data are collected either manually, by resource administrators through the web interface accounting system, or by the accounting service API.

From the accounting system dashboard screen the users can query the accounting data about application/VRC/resource center/country for a specific period in the form of CPU time or jobs for computing data, as well as the total number of virtual machine hours, or number of virtual machines in the case of Cloud data. Results of the queries are returned as a table, but could be also visualized as two types interactive charts, line and pie chart, as illustrated in Figure 12.

Currently, there are a total of 36 users registered in the accounting system. Out of this, 25 are using their VI-SEEM login accounts to access the service, and 20 have been granted administrative privileges on at least one of the resources in the system. Users of the system are frequently administrators of resource centers who want to check their resource center statistics.



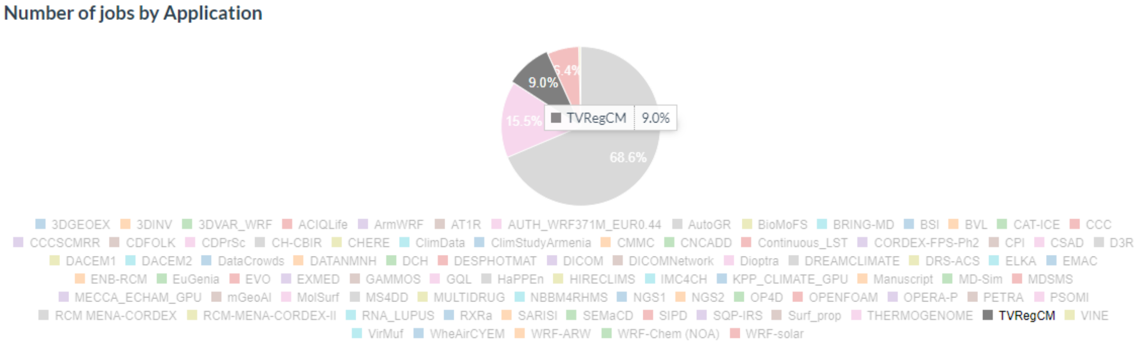


Figure 12: Example of generated charts from the accounting service for number of jobs per application.

4.4. Helpdesk

Although we have created a centralized support system for the project's needs (https://support.vi-seem.eu/), a vast majority of actual support cases has been handled outside of this system. This is primarily caused by the fact that users can get much faster and direct help by support staff utilizing any of the available instant communication systems, from direct email messages to holding a multipoint video conference or remote computer access, sidestepping the use of the classic support system.

We illustrate the provided support using the total number of tickets, the average number of conversations per ticket, and the average response time. The total number of recorded issues at the moment is 536. Statistically, the average number of conversations/replies per ticket is 6 with a standard deviation of 4, and the average response time is 21 hours with a standard deviation of 15. Figure 13 illustrates the distribution of the observed parameters per type of the problem.

Most of the reported problems were related to the usage of HPC (216 out of 536), Cloud (125), and Live access service (55). These are followed by the Simple storage service (27), the Data repository service (24), the Data analysis service (23), Clowder (20), etc. The obtained distribution is based on the usage of a particular service. Most of the projects are using HPC services, so a significant number of issues is related to this service. Similar applies to Cloud services, the Live access service, and the Simple storage service. Although the Simple storage service is often used, compared to the HPC, the total number of tickets is smaller since this service deals with a very concrete and simple technical problem. The HPC covers a wider variety of different problems, which is strongly reflected in the number of registered related issues.

The average number of conversations per ticket gives us information on how easily a supporter can detect the source of a problem and reproduce it. According to the obtained statistics, the Cloud-related issues require the most of interactions (15) with the end-user. These are followed by the Live access service (11), the archival service (8), HPC (8), etc.

The average response time per conversation gives us information on how easily a supporter can resolve the reported problem. As expected, HPC-related issues are usually the most complex, with an average response time of 63 hours, followed by the Cloud problems with 36 hours on average, the Data repository service with 32 hours, the Archival service with 21 hours, etc.

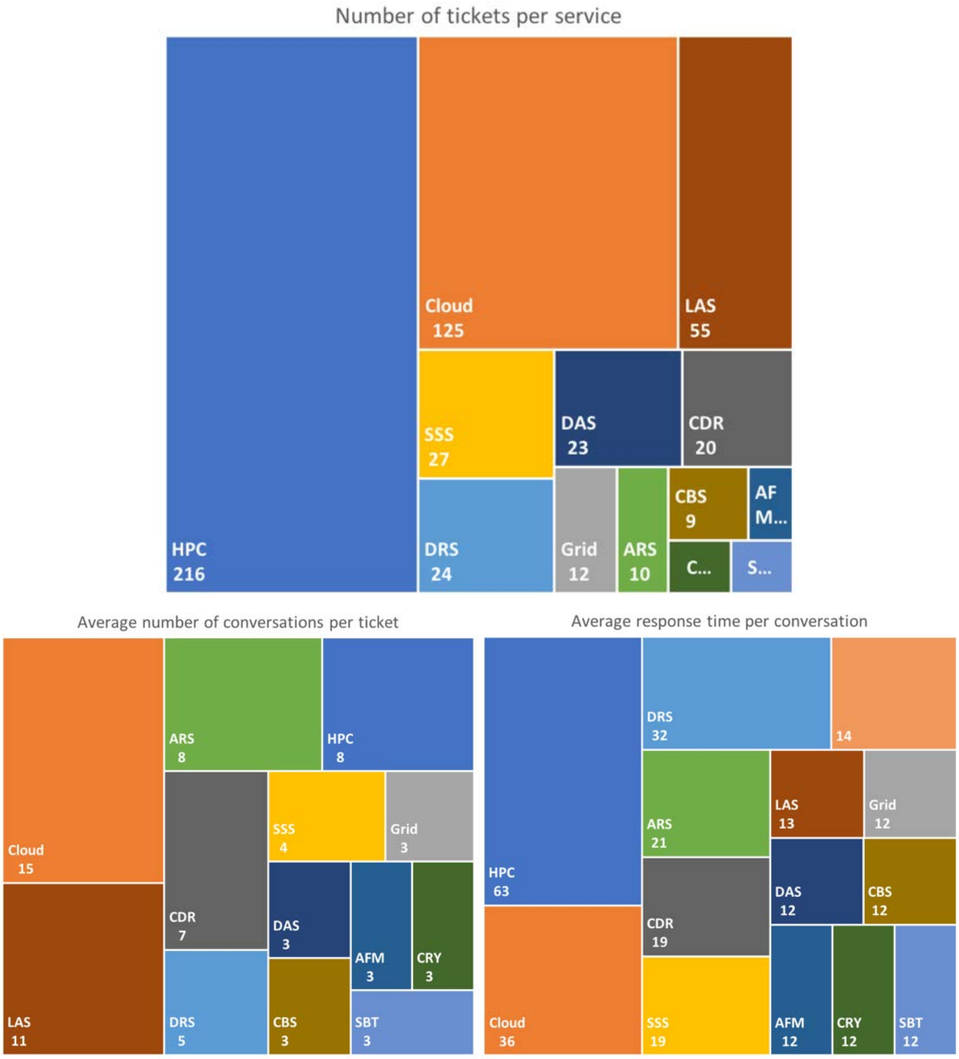


Figure 13: Total number of tickets per service, average number of conversations per ticket, and average response time per conversation.

4.5. Source code repository

In order to provide an environment for easy storing and exchanging of the codes used in the project, a code repository was created and installed at https://code.vi-seem.eu/. Due to the heterogeneous and decentralized nature of the project, we have opted for a code repository based on the Git software as it supports distributed and non-linear code development while providing support for data integrity, as well as full traceability of all relevant operations.

A web-based interface (illustrated in Figure 14) is served by the adapted GitLab installation [16] that supports not only source code version control, but also integrates tools for issue tracking, wiki, developer communication, code merging, etc. The code repository supports both local user accounts and logging in via the VI-SEEM authentication and authorization infrastructure. The system currently has 93 active users in five groups, working on 39 unique projects.

The repository also exports the data on public projects to the Data discovery service.

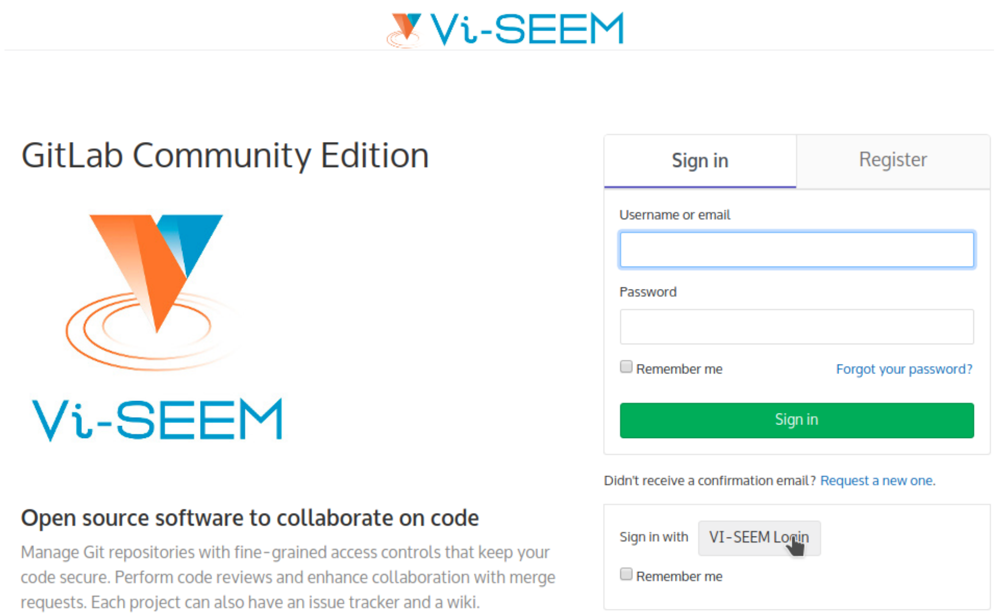


Figure 14: Main page of the VI-SEEM source code repository.

4.6. Technical wiki

The VI-SEEM technical wiki (http://wiki.vi-seem.eu/), which is hosted at the Cyprus Institute and is based on the MediaWiki platform [17], has been integrated with the VI-SEEM Authentication and Authorization Infrastructure (AAI). This integration provides single sign-on capabilities to the VI-SEEM wiki pages and therefore wiki users can now be authenticated through the central Vi-SEEM AAI without having special credentials for the wiki pages.

The VI-SEEM technical wiki has 280 pages and 180 uploaded files prepared by 46 registered users. At the moment, the wiki has 1,171,378 page views. The most viewed pages are: Access to resources, VI-SEEM Services Integration, Source Code Repository, VI-SEEM GOCDB, VI-SEEM GOCDB PI Technical Documentation, and Accounting System.

5. Service catalogue/portfolio and management

The Service Portfolio Management Tool (SPMT) is aimed at facilitating service management in IT service provision, including federated scenarios. The SPMT presents a complete list of the services managed by a service provider; some of these services are visible to the customers, while others are internal. The SPMT allows to maintain and manage the descriptions of services in the portfolio, and to manage the transition from the portfolio to the catalogue. The service management system has been designed to be compatible with the FitSM service portfolio management [18]. The SPMT implements the service portfolio and catalogue that has been described in the VI-SEEM deliverable D3.2 – Service registry, operational and service level monitoring [4].

An important aspect of the in-house developed SPMT is the integration possibility. The system supports access to the service catalogue (Figure 15) and portfolio using Open APIs, enabling seamless integration of the VI-SEEM service offering into the future pan-European scientific services catalogues, such as the European Open Science Cloud EOSC-hub [19]. Toward this goal, the VI-SEEM has already started an intensive collaboration with the H2020 project e-InfraCentral [20]. After receiving several highly favorable recommendations for its service catalogue in comparison with other EU e-Infra projects, the VI-SEEM started working on describing the service offering using the service definition template provided by e-InfraCentral, leading to the full-service integration of the VI-SEEM services into the e-InfraCentral central catalogue.

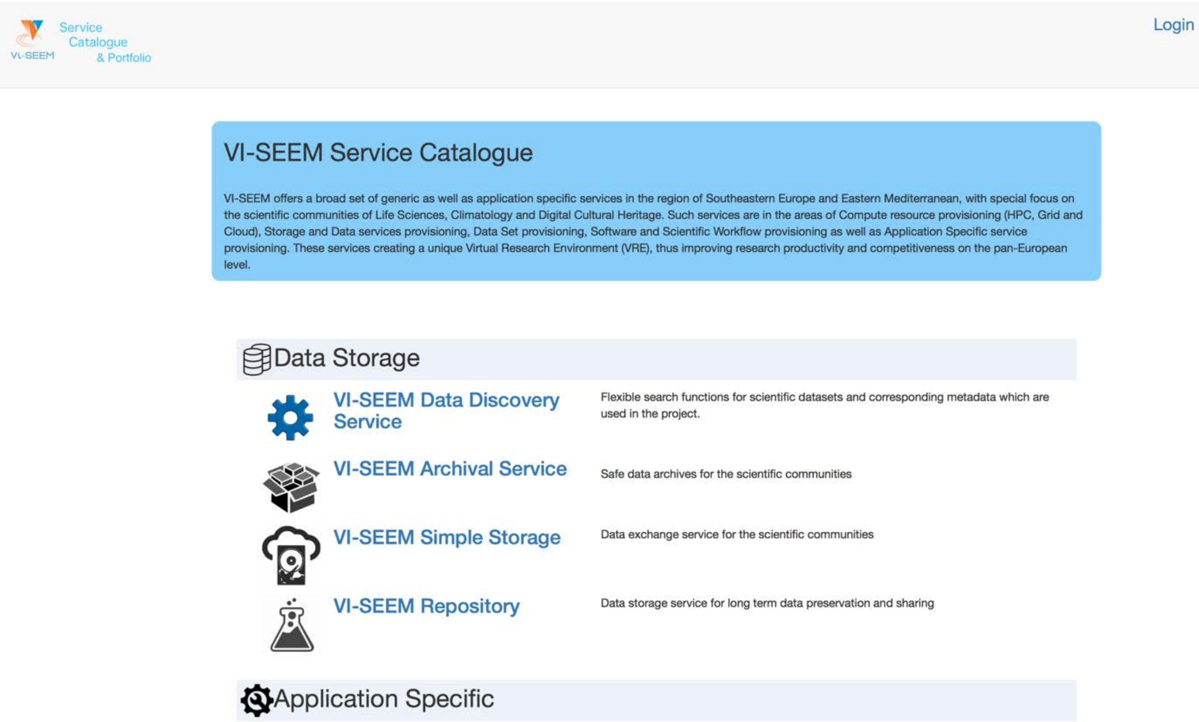


Figure 15: Frontpage of the VI-SEEM service catalogue/portfolio.

5.1. Usage and performance

The service catalogue contains information about the services offered to the potential customers. Usually, this is a subset of all the service information recorded in the service portfolio. In the case of the VI-SEEM service catalogue, only the latest versions of the production services are listed in the catalogue, while the portfolio contains multiple versions and different readiness statuses: concept, under development, beta, production, retired, etc.

The service catalogue/portfolio contains a description of each service and groups services into service areas. The following service areas exist in the VI-SEEM service catalogue:

- Data storage – services related to all aspects of the storage of data relevant to the project;
- Application-specific – services developed by the virtual research communities involved in the project;
- Training and education (VI-SEEM training portal);
- Computation – computing resources related services;
- Authentication and authorization (VI-SEEM Login service);
- Service provisioning (VI-SEEM Service Portfolio Management System).

In total, 25 services offered through the catalogue are grouped in these six categories, as Table 13 illustrates. A service owner (a person responsible for a specific service) classifies the service in one of the available groups and has the full responsibility for the content that is provided within the service catalogue and portfolio, which includes: name, description (both internal and external), service area and type, the procedure for requesting the service, funding, value for the customers, risks and competition, contact information, versions, internal and external dependencies, and users/customers roles.

Recently, to address the recommendations from the last review, the VI-SEEM consortium created a new working group dedicated to updating the service catalogue appearance and content. The working group agreed on the structure of services' descriptions, prepared practical guidelines on how to follow this structure, and requested updates by services' owners. This resulted in a more intuitive portal and user-friendly navigation for new visitors. In addition to this, a user quality control flag, i.e., a user rating per service was included in the catalogue.

The service catalogue/portfolio is fully integrated with the project's AAI, and therefore the AAI handles authentication requests sent to the service. Any user authenticated by the VI-SEEM Login service has privileges to register a new service. Also, the service is accessible via a RESTful API to accommodate 3rd party applications. The project's monitoring system currently uses this interface.

The production service catalogue/portfolio is available at https://services.vi-seem.eu/. It is running under the Apache web server, so the service usage data are generated by accumulating the Apache access log files. According to the Webalizer tool [21], the service catalogue/portfolio has on average 9,299 visits per month, out of which 1,222 are unique. On a daily basis, it has 619 visits, out of which 81 are unique. The project's monitoring system periodically visits the catalogue to create a list of services to be probed, and doing this creates the discrepancy between the total and the unique number of visits.

Service category	Number of services	Services
Data storage	4	Data Discovery Service
		Archival Service
		Simple Storage Service
		Repository Service
Application specific	15	Subtract
		CCC
		VirMuf
		DRS-ACS
		ChemBioServer
		Regional Community Datasets
		Live Access Server
		AFMM
		Scientific Application Environment
		Workflow, software tools repository
		NANO-Crystal
		MC4CH (ARCHES)
		DICOM
		CHERE
		Clowder
Training and education	1	Training
Computation	3	HPC
		Cloud
		Grid
Authentication and authorization	1	VI-SEEM Login Service
Service provisioning	1	VI-SEEM SPMT

Table 13: Services and service areas registered in the VI-SEEM service catalogue.

6. Authentication and Authorization Infrastructure

The VI-SEEM Login service enables researchers to access VI-SEEM infrastructure resources in a user-friendly and secure way. More specifically, VI-SEEM Login allows researchers whose home organizations participate in one of the eduGAIN federations [22] to access VI-SEEM infrastructure services and resources with the same credentials they are using at their home organization. Furthermore, VI-SEEM Login supports user authentication with social identities, enabling even those users who do not have an account at a university or research institute federated in eduGAIN to be able to seamlessly access the VI-SEEM services without compromising the security of the VI-SEEM infrastructure.

VI-SEEM Login serves as a central hub between federated Identity Providers (IdP) and Service Providers (SP). More specifically, it acts as a Service Provider towards the external Identity Providers and as an Identity Provider towards the VI-SEEM Service Providers. Common technical services, such as Identity Provider discovery and user registration, are provided centrally by VI-SEEM Login, and do not have to be implemented by each individual service. VI-SEEM Login is also responsible for aggregating user attributes originating from various authoritative sources (IdPs and attribute authorities) and delivering them to the connected Service Providers in a harmonized and transparent way. Service Providers can use the received attributes for authorization purposes.

It should be noted that the design of the VI-SEEM Login service follows the AARC Blueprint Architecture [23] which is addressing the growing need for research infrastructures and e-Infrastructures to use federated authentication and authorization mechanisms allowing interoperation in the context of international collaboration. More information about the architecture of the VI-SEEM Login service and its individual components can be found in the deliverable D3.4 – VRE AAI model and compatibility with other e-Infrastructures [5].

6.1. Usage and performance

The VI-SEEM Login service has been configured as a service provider proxy in order to allow multiple services within the VI-SEEM administrative domain to benefit from a single Identity Provider. The list of services that have been integrated follows:

- VI-SEEM Accounting,
- VI-SEEM Helpdesk,
- VI-SEEM Person Registry,
- VI-SEEM Repository,
- VI-SEEM Code Repository,
- VI-SEEM Monitoring,
- VI-SEEM Simple Storage Service,
- VI-SEEM Simple Storage Service instance at ETFBL,
- VI-SEEM Service Portfolio Management System,
- VI-SEEM Wiki.

VI-SEEM services that are connected to the Login service can become available to more than 2,800 Universities and Institutes from over 55 Identity Federations in eduGAIN with little or no administrative involvement. Compliance with the REFEDS Research and Scholarship (R&S) entity category ensure sufficient attribute release. Complementary to this, users without an account on a federated Identity Provider are still able to use social media or other external authentication providers, such as Google, Facebook and LinkedIn.

A total of 74 users have registered with the VI-SEEM Login service. Users registered with the VI-SEEM Login service have performed over 16,000 of federated logins as illustrated in Figure 16. The service integrates 14 services providers, and 2,824 identity providers. In addition to this, 3 social identity providers, namely Google, Facebook and LinkedIn are integrated as well.

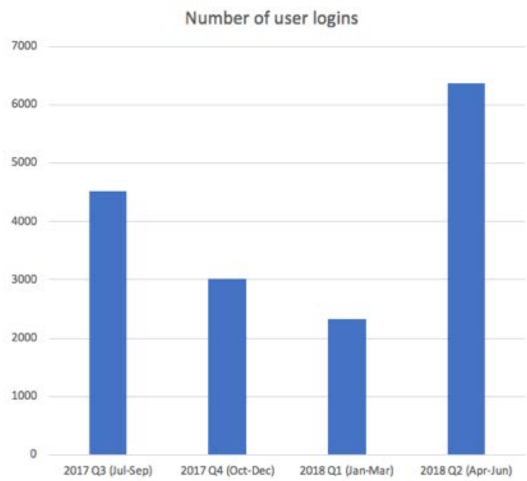


Figure 16: Number of user logins on a quarterly basis.

7. Conclusions

In last three years, the VI-SEEM consortium supported 66 production projects submitted by researchers and the SMEs from the region. The fact that 75% of these projects requested access to at least two different kinds of available resources (HPC, Cloud, Grid, Storage), goes in favour of the heterogeneous nature of our infrastructure, although it was a management challenge to the project's operations team. The main achievements and conclusions from this process are presented in this deliverable, and can be summarized as follows:

- Early access to the infrastructure was crucial for the realization of the 1st implementation phase, which happened in project month 5 (February 2016). By then, more than 80% of the projected size of the infrastructure was integrated into the VI-SEEM infrastructure, which enabled the development of the services offered a year later in the 1st call for production use of resources and services. The rest of the infrastructure resources were integrated gradually, and the final capacity includes 23,694 CPU-cores, 1,166,592 GPU-cores (516 GPU cards), 20,496 Xeon Phi-core (336 Xeon Phi cards), 3,112 Grid CPU-cores, 14,152 Cloud VM-cores, and 18 PB of storage space. The total size of the infrastructure matches the defined key performance indicators.
- The VI-SEEM consortium supported 66 projects, 23 of them related to climate research, 24 to life sciences, 16 to digital cultural heritage, and 3 of them were cross-disciplinary. In total, 41 projects requested access to the HPC CPU, 15 to HPC GPU, 4 to HPC Xeon Phi, 18 to Cloud, and 3 to Grid resources. The overall allocation of HPC resources was 22.3 million CPU-core hours, 2.6 million GPU-card hours, and 2.2 million Xeon Phi-core hours. Additionally, 70,000 CPU-core hours were used through the Grid interface, as well as 225 Cloud VMs. The total storage space consumption by all projects on HPC storage systems amounts to 95.2 TB. In addition to this, approximately 10 TB of storage space is occupied at the Simple storage service, 3TB of storage space at the Archival service, and in total 3 TB of storage space at the application- and data-specific services, which include the Data repository service, the Live access server, and the Clowder.
- A set of operational and resource management tools were deployed for the efficient management of the available computing and storage resources. All of them are fully integrated with the project's AAI. Although we created a centralized support system for the project's needs, a majority of actual support cases were handled outside of this system. This is primarily caused by the fact that users can get much faster and more direct help by the support staff via commonly used instant communication systems. Therefore, in the corresponding section, we have collected and provided statistics that include various support channels, from direct email messages to holding a multipoint video conferences or a remote computer access, sidestepping the use of the classic support system.
- The service catalogue/portfolio management system contains information about the services offered to the customers. It is fully integrated with the project's AAI, and could be used by other infrastructures as well. In total, 25 services are offered through the catalogue and grouped into six categories.
- The VI-SEEM Login service enables researchers to access the services in a user-friendly and secure way. The service integrates 14 service providers, and 2,824 identity providers. In addition to this, we also integrated 3 social identity

providers, namely Google, Facebook and LinkedIn. The VI-SEEM Login service was used for over 16,000 federated logins at the moment.