C-SCALE

D3.4 Periodical assessment of the services V1

Status:	Final	**Planned due date:**	31/12/2021
Version:	1.0	**Submission date:**	28/02/2022
Lead Participant:	EGI Foundation	**Lead Author:**	Enol Fernández
Related WP:	WP3	**Document Ref:**	D3.4
Dissemination Level:	Public (PU)		
Document Link:	https://doi.org/10.5281/zenodo.6223264		

Deliverable Abstract
This report presents the first assessment of the C-SCALE services provided through the Virtual Access (VA) mechanism to the users. These include computing and storage resources. This first version of the deliverable presents the existing VA installations, their management within the project and the initial usage delivered from the start of the project (January 2021) and until November 2021.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017529.

COPYRIGHT NOTICE

C-SCALE receives funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017529.

DELIVERY SLIP

Date	Name	Partner/Activity
Lead Author:	Enol Fernández (EF)	EGI Foundation
Contributors:	Mario David (MD)	INCD
	Raymond Oonk (RO)	SURF
	Nikos Triantafyllis (NT)	GRNET
	Charis Chatzikyriakou (CC)	EODC
Moderated by:	Enol Fernández (EF)	EGI Foundation
Reviewed by:	Andrea Manzi (AM)	EGI Foundation
	Benjamin Schumacher (BS)	EODC
Approved by:	C-SCALE Activity Management Board (AMB): Christian Briese (CB), Diego Scardaci (DS), Charis Chatzikyriakou (CC), Zdeněk Šustr (ZS), Enol Fernández (EF), Björn Backeberg (BB), Eleonora Testa (ET)	EODC, CESNET, EGI Foundation, Deltares

DOCUMENT LOG

Issue	Date	Comment	Author(s)
V0.1	2021-12-23	Initial Version	EF
V0.2	2022-01-10	Added INCD Usage	MD
V0.3	2022-01-26	Added EODC, SURF and GRNET usage	CC, NT, RO
V0.4	2022-02-09	Added changes to handle reviewers' comments	EF
V0.5	2022-02-18	Second revision after reviewers' comments	EF
V0.6	2022-02-22	Final revision from reviewers	EF
V1.0	2022-02-28	Final version approved by AMB	C-SCALE AMB

Table of Contents

List of Tables

List of Acronyms

Acronym	Description
AAI	Authentication and Authorisation Infrastructure
API	Application Programming Interface
CO	Collaborative Organisation
C-SCALE	Copernicus - eoSC AnaLytics Engine
DIAS	Data and Information Access Services
EO	Earth Observation
EODC	Earth Observation Data Centre (C-SCALE provider)
EOSC	European Open Science Cloud
GPU	Graphics Processing Unit
GRNET	Greek Research and Technology Network (C-SCALE provider)
GUI	Graphical User Interface
HPC	High Performance Computing
HTC	High Throughput Computing
HTDP	High Throughput Data Processing
IaaS	Infrastructure as a Service
INCD-NCG	Infrastructura Nacional de Computação Distribuída-National Computing Grid
OLA	Operational Level Agreement
PBS	Portable Batch System
PaaS	Platform as a Service
SLA	Service Level Agreement
SRAM	SURF Research Access Management
STAC	SpatioTemporal Asset Catalog
TB	Terabyte
TRL	Technology Readiness Level
VA	Virtual Access
VM	Virtual Machine
WP	Work Package

Executive Summary

C-SCALE brings together European commercial (e.g., Copernicus DIAS) and public computing and storage providers to deliver a federated infrastructure to support Copernicus and Earth Observation (EO) use cases that deal with data- and compute-intensive workloads. The federation enables the access to a set of computing and storage resources, called *installations*, operated by the partners of the project and deployed via a Virtual Access (VA) mechanism. A considerable part of the C-SCALE total budget (almost 20%) is devoted to cover the costs of these installations that the project will deliver to EOSC users as services. While these C-SCALE services are still under development, the providers of the project were ready to offer their resources via Virtual Access from the start of the project.

This deliverable is the first of a series of status updates on how the C-SCALE infrastructure is used by the existing use cases. The document covers the first year of the project. During this initial phase of the project, C-SCALE providers tested the newly available federation access channels and integrated use cases into the project as foreseen by the project plan. Usage of the C-SCALE infrastructure is now starting to increase as the use cases of the project are deployed on the providers. Further increase in usage is expected when new use cases from the project's Open Call will start to engage with the project.

1 Introduction

This deliverable provides the initial assessment of the C-SCALE services based on the Virtual Access (VA) installations of the project, which provide the computing and storage resources that power the C-SCALE services. This first version of the deliverable will be focused on presenting the existing VA installations, their management within the project and the initial usage delivered until November 2021. Deliverables D3.5, D3.6 and D3.7, expected in M18 (June 2022), M24 (December 2022) and M30 (June 2023) respectively, will provide an updated assessment of the usage of the C-SCALE service covering further usage of the services from the project use cases and those received and accepted from the project's Open Call.

C-SCALE includes 23 different installations from 8 providers, all of which were ready to be used from the start of the project, and are classified into four main categories:

- Cloud service: delivering access to compute resources as virtual machines or containers. Usage of these installations is measured in the form of CPU core per unit of time.
- High Performance Computing (HPC)/High Throughput Computing (HTC) service: delivering access to large, shared computing systems for generic batch processing via a jobs queue that is managed by a workload scheduler. As with the cloud service above, usage is measured in CPU core per unit of time.
- GPU service: delivering access to compute resources where GPU hardware is available. In C-SCALE, GPUs are offered via cloud IaaS interfaces. Usage is measured in GPU node per unit of time.
- Storage service: delivering access to storage resources where users can deposit their input and output data. Usage is measured in TB per unit of time.

The list of installations is given in Table 1. The unit of time for each installation in C-SCALE varies due to the diversity of the accounting practices of providers. Moreover, differences between the computational platforms (e.g. types of CPU available at each provider) and details of the storage systems prevent realistic interchangeability or comparison of the units of access between providers.

Table 1. C-SCALE Installations per provider and category

Provider	VA installation	Category	Mode of access	Unit of access
EODC	EODC EO storage	Storage	Secure Shell (ssh)	TB/year
	EODC EO Cloud	Cloud IaaS	OpenStack[1] APIs	CPU/year
	EODC HPC	HPC/HTC	ssh	CPU/hour
CESNET	MetaCentrumCloud - CPU	Cloud IaaS	OpenStack API/GUI + PBS[2]	CPU core/month
INFN	INFN Bari Storage	Storage	OpenStack APIs	TB/month
	INFN Bari CPU	Cloud IaaS	OpenStack APIs	CPU core/month

[1] OpenStack is a Open Source Cloud Software, see https://www.openstack.org/
[2] Portable Batch System, a batch scheduling system, see https://www.openpbs.org/

SURF	SURFsara dCache Front end storage	Storage	ssh	TiByte*Year
	SURFsara Spider (HTDP) Storage	Storage	ssh	TB*Year
	SURFsara Data Processing Compute	HPC/HTC	ssh	CPU*Hour
	SURFsara MS4 Storage	Storage	ssh	TB*Year
	SURFsara MS4 SSD Storage	Storage	ssh	TB*Month
	SURFsara MS4 Compute	Cloud IaaS	ssh	CPU*Hour
VITO	VITO CVB (Storage)	Storage	via Notebooks, VMs, openEO[3] API	TB/Month
	VITO CVB (CPU)	Cloud IaaS	SSH & openEO API	CPU/hour
GRNET	GRNET KNS - Storage Cloud	Storage	OpenStack APIs	TB*Hour
	GRNET KNS - Cloud	Cloud IaaS	OpenStack APIs	Core*Hour
	GRNET ARIS - Storage HPC	Storage	ssh	TB*Hour
	GRNET ARIS - CPU HPC	HPC/HTC	ssh	Core*Hour
CloudFerro	CREODIAS - Storage	Storage	OpenStack APIs	TB/month
	CREODIAS - Compute	Cloud IaaS	OpenStack APIs	vCore/hour
	CREODIAS - GPU	Cloud IaaS	OpenStack APIs	VMGPU/hour
INCD	INCD Lisbon (NCG) (Storage)	Storage	OpenStack APIs	TB*month
	INCD Lisbon (NCG) (Compute)	Cloud IaaS	OpenStack APIs	CPU*hour

Although the C-SCALE services are still under development, installations can be accessed and have started to deliver computing and storage resources to use cases. The installations have been gradually incorporated into the C-SCALE federation as defined in the initial design documents for the Compute[4] and Data Federation[5].

[3] openEO: an open API to connect to Earth observation cloud back-ends, see https://openeo.org/
[4] D3.1 Initial Design of the Compute Federation https://doi.org/10.5281/zenodo.5084884
[5] D2.1 C-SCALE Copernicus Data Access and Querying Design https://doi.org/10.5281/zenodo.5045317

2 Virtual Access Metrics and current status

Each installation of C-SCALE defines a set of metrics to report VA usage to the project, with clear description of the metric and how the measurement is done by the provider. These metrics are reported on a 3-monthly basis to the central project Confluence by the providers. In this way the capacity planning group can keep track of the usage trends and react accordingly and plan the capacity allocation for the use cases to be supported.

In order to set clear expectations on the delivery of resources through the VA mechanism, C-SCALE implemented an SLA (Service Level Agreement)/OLA (Operational Level Agreement) framework in C-SCALE's internal Confluence. For every use case, a document describes in detail the agreed number of resources made available on every provider supporting the use case and the exact conditions (e.g. how to access the resources, start and end dates, usage quotas set, etc.). A template for the agreements was developed to create the SLA/OLA documents (available as Appendix A of this deliverable). When a use case is onboarded into C-SCALE project, the SLA/OLA document is created and agreed, and then reviewed periodically (every 6 months initially) with a satisfaction interview where the use case representatives can provide feedback and improvement suggestions for the services delivered by C-SCALE. Currently, the Aquamonitor use case has its agreement finalised while the rest of the use cases have their agreements under discussion.

The Operations of the Virtual Access installation task (T3.4) and Provider onboarding support and the VA coordination task (T5.1) form the capacity planning group to jointly coordinate the delivery of resources. The planning group established a capacity plan, which keeps track of the current allocation of resources to use cases. The allocation of the remaining capacity is then planned to include the consideration of the future needs of the project, the current VA consumption and the expected delivery of resources as specified in the existing SLAs.

Table 2 shows the cumulative metric values from the start of the project until M11 (November 2021), which is the last metric collection before the writing of this deliverable. Following deliverables covering the assessment of the services (D3.5, D3.6 and D3.7) will present the updated values for the metrics. The baseline is the initial value for the metric, which is 0 as the installations were not available for EOSC users before.

Table 2. VA Usage

VA Installation	Metric type	Description	Baseline	M01 - M11
EODC EO storage	Number of users	Number of user accounts	0	0
	Usage	Allocated TB/year	0	0
	Number of countries reached	Number of countries from the user accounts	0	0
	Names of countries reached	Names of countries from the user accounts	-	-
EODC EO Cloud	Number of users	Number of user accounts	0	1
	Usage	Time of VM usage	0	1.33 CPU/Year
	Number of countries reached	Number of countries from the user accounts	0	1

	Names of countries reached	Names of countries from the user accounts	0	Austria
EODC HPC	Number of users	Number of user accounts	0	0
	Usage	CPU hours	0	0
	Number of countries reached	Number of countries from the user accounts	0	0
	Names of countries reached	Names of countries from the user accounts	-	-
MetaCentrumCloud - CPU	Number of users	Number of user consuming CPU resources	0	0
	Usage	CPU core months consumed (with typically 4GB of RAM per core)	0	0
	Number of countries reached	Number of countries from the user accounts	0	0
	Names of countries reached	Names of countries from the user accounts	-	-
INFN Bari Storage	Number of user communities	Communities supported by the installation	0	0
	Usage	TB used / month	0	0
INFN Bari CPU	Number of user communities	Communities supported by the installation	0	0
	Usage	CPU hours	0	0
SURFsara dCache Front end storage	Number of users	Number of registered C-SCALE users	0	0
	Usage	Number of TB per year consumed by C-SCALE users	0	0
	Satisfaction	Degree of C-SCALE user satisfaction	0	4
SURFsara Spider (HTDP) Storage	Number of users	Number of registered C-SCALE users	0	14
	Usage	Number of TB per year consumed by C-SCALE users	0	19.15 TB*Year
	Satisfaction[6]	Degree of C-SCALE user satisfaction	0	4
SURFsara Data Processing Compute	Number of users	Number of registered C-SCALE users	0	14
	Usage	Wall time CPU core hours consumed by C-SCALE users	0	143599 CPU*hours
	Satisfaction	Degree of C-SCALE user satisfaction	0	4
SURFsara MS4 Storage	Number of users	Number of registered C-SCALE users	0	0
	Usage	Number of TB per year consumed by C-SCALE users	0	0
	Satisfaction	Degree of C-SCALE user satisfaction	0	0
SURFsara MS4 SSD Storage	Number of users	Number of registered C-SCALE users	0	0

[6] Measures customer satisfaction based on a feedback survey that grades services from 1-5 (being very dissatisfied and 5 being very satisfied)

	Usage	Number of TB per year consumed by C-SCALE users	0	0
	Satisfaction	Degree of C-SCALE user satisfaction	0	0
SURFsara MS4 Compute	Number of users	Number of registered C-SCALE users	0	0
	Usage	Wall time CPU core hours consumed by C-SCALE users	0	0
	Satisfaction	Degree of C-SCALE user satisfaction	0	0
VITO CVB (Storage)	Number of users	User that created an account and has actively logged into a VM or used an API service	0	0
	Usage	Time in months for VM usage and in hours for Hadoop usage	0	0
	Number of countries reached	Number of countries reached using our datawarehousing system	0	0
	Names of the countries reached		-	-
VITO CVB (openEO - CPU)	Number of users	User that logged in to openEO	0	0
	Usage	Time in months for VM usage and in hours for Hadoop usage	0	0
VITO CVB (CPU)	Number of countries reached	Amount of countries reached using our datawarehousing system	0	0
	Names of the countries reached		-	-
GRNET KNS - Storage Cloud	Number of users	Number of users accessing the service	0	11
	Usage	Total number of TB Hours for the duration of the project	0	Disk GB-Hours - 372120.86
GRNET KNS - Cloud	Number of users	Number of users accessing the service	0	11
	Usage	Total number of virtual CPU hours with 2 GB of RAM consumed for the duration of the project	0	CPU Hours - 93137.71
GRNET ARIS - Storage HPC	Number of users	Number of users accessing the service	0	5
	Usage	Total number of TB hours for the duration of the project	0	1.364 TB*Hours
GRNET ARIS - CPU HPC	Number of users	Number of users accessing the service	0	5
	Usage	Total number of CPU core hours with 2 GB of RAM consumed for the duration of the project	0	170 Core*Hours
CREODIAS - Storage	Number of users	Users that create an account and logged into OpenStack dashboard or used API services	0	0
	Usage	Number of credits in pay-per-use model	0	0

CREODIAS - Compute	Number of users	Users that create an account and logged into OpenStack dashboard or used API services	0	0
	usage	Number of credits in pay-per-use model	0	0
CREODIAS - GPU	Number of users	Users that create an account and logged into OpenStack dashboard or used API services	0	0
	Usage	Number of credits in pay-per-use model	0	0
INCD Lisbon (NCG) (Storage)	Number of users	Total number of new users	0	4
	Usage	Total storage (unit TB*month) allocated to the project	0	62015
	Number of countries reached	Total number of new countries	0	2
	Names of the countries reached		-	Portugal, Netherlands
INCD Lisbon (NCG) (Compute)	Number of users	Total number of new users	0	4
	Usage	Total number of virtual VCPU*hour with 2 GB of RAM consumed for the duration of the project	-	310832
	Number of countries reached	Total number of new countries	0	2
	Names of the countries reached		-	Portugal, Netherlands

Table 3 shows the percentage of consumption of the available units for all the installations. As planned in the project, the VA usage during the initial months of the project has been mostly related to test activities to make the new federation access channels fully operational and to integrate the existing use cases. Those providers without any usage so far are already engaged with the use cases but were not yet delivering resources at the time of the metric collection.

Table 3. VA Consumption

Provider	VA installation	Available units	Consumed M1 - M11
EODC	EODC EO storage	150 TB/year	0.00%
	EODC EO Cloud	620 CPU/year	0.21%
	EODC HPC	600000 CPU/hour	0.00%
CESNET	MetaCentrumCloud - CPU	2850 CPU core/month	0.00%
INFN	INFN Bari Storage	1500 TB/month	0.00%
	INFN Bari CPU	3840CPU core/month	0.00%
SURF	SURFsara dCache Front end storage	154 TiByte*Year	0.00%
	SURFsara Spider (HTDP) Storage	160 TB*Year	11.97%
	SURFsara Data Processing Compute	1500000 CPU*Hour	9.57%
	SURFsara MS4 Storage	10 TB*Year	0.00%
	SURFsara MS4 SSD Storage	96 TB*Month	0.00%
	SURFsara MS4 Compute	147600 CPU*Hour	0.00%

VITO	VITO CVB (Storage)	1200 TB/Month	0.00%
	VITO CVB (CPU)	876000 CPU/hour	0.00%
GRNET	GRNET KNS - Storage Cloud	250000 TB*Hour	0.15%
	GRNET KNS - Cloud	1650000 Core*Hour	5.64%
	GRNET ARIS - Storage HPC	219000 TB*Hour	0.00%
	GRNET ARIS - CPU HPC	1000000 Core*Hour	0.02%
CloudFerro	CREODIAS - Storage	2762 TB/month	0.00%
	CREODIAS - Compute	1500000 vCore/hour	0.00%
	CREODIAS - GPU	6000 VMGPU/hour	0.00%
INCD	INCD Lisbon (NCG) (Storage)	450 TB*month	13.78%
	INCD Lisbon (NCG) (Compute)	3240000 CPU*hour	9.59%

The consumption of EODC's resources though VA started with the onboarding of Use Case 6: Wetland Water Stresses to the EODC EO Cloud. The onboarding of this use case took place in November 2021, which is the last month of the reporting period for this document. The EODC EO Cloud installation has been used by TU Wien, which is based in Austria, in M11. This explains the usage of 1.33 CPU/Year. For Use Case 5: RETURN, where EODC is providing access to higher level data, there has been no VA usage recorded.

For the period M01-M11 the consumption of SURF's resources is divided into three components: (i) the setup and testing of a dedicated HTC cluster (Delena) for the HPC and HTC federation; this cluster is a clone of Spider7 that has been specially setup for and dedicated to C-SCALE in order to allow for integration with the SURF Research Access Management (SRAM) AAI solution8; (ii) onboarding and support of the High resolution land surface drought analysis (LSDA) use case on generic Spider HTC cluster; and (ii) onboarding and support of the RETURN use case.

GRNET has been actively supporting - in terms of resources and technical support - the HiSea9 use case, for the period M01-M11. Two major facilities are provided: Cloud and HPC, both delivering CPU and storage resources. On the cloud, one VM with 16 cores is used for the pre- and post-processing in the use case and a second VM is deployed as an NFS-Server with an additional of 300 TB storage for data caching. A Kubernetes cluster is currently being planned with the assistance of INFN. Moreover, in the HPC infrastructure, access to run jobs on the "fat nodes" partition of the HPC infrastructure10 (40 cores per node) has been granted. The initial VA consumption in this HPC infrastructure is attributable to the setup and testing of the system, such as calculating the scalability and the performance efficiency of the production code. Four jobs were submitted consuming 172 core hours. After this initial testing phase, 5 users from HiSea have started to run production code using Singularity11 containers. Up to now, 19 jobs were submitted, and 326 core hours were consumed.

[7] http://doc.spider.surfsara.nl/en/latest/Pages/about.html
[8] See D3.1 D3.1 Initial Design of the Compute Federation - https://zenodo.org/record/5084884
[9] https://hiseaproject.com/
[10] See technical description of nodes at https://doc.aris.grnet.gr/system/hardware/#fat-nodes
[11] https://sylabs.io/singularity

INCD has been actively supporting the Aquamonitor use case with a dedicated deployment of openEO service12 and a local STAC Catalogue13. Initially a Kubernetes cluster was deployed on top of the INCD OpenStack IaaS. This Kubernetes cluster is then used to support all needed services and platforms for the use case. Furthermore, a storage volume of 12TB was instantiated to host a local cache of 2 years (2000-2002) of Landsat data for Spain and Portugal. The images for the cache have been transferred from the CREODIAS service. The services are now in pre-production stage, with service endpoints with properly registered domains and encrypted with X.509 certificates. As such the services are mostly ready for the users to perform image processing analytics.

[12] https://openeo.org/
[13] https://stacspec.org/

3 Conclusions

C-SCALE Virtual Access installations are available to deliver resources for the use cases of the project. A set of Service Level Agreements specifies the conditions and the amount of resources to be delivered to each use case. These are documents agreed between customers and providers that help to set expectations and support the capacity planning of the project.

The 23 installations of C-SCALE report usage to the project periodically (every 3 months) starting from month 3. During the first year of the project the usage of these installations has been focused on testing integration of the providers (following the Compute[14] and Data federation[15] specs of the project) and supporting those use cases of the projects in a more mature state. For this deliverable, a special report for month 11 was produced so the metrics could be included and reported in this document. The project expects the usage to grow in the upcoming months as all the use cases of the project are now getting ready to consume resources and use cases from the open call will start being onboarded.

[14] D3.1 Initial Design of the Compute Federation https://doi.org/10.5281/zenodo.5084884
[15] D2.1 C-SCALE Copernicus Data Access and Querying Design https://doi.org/10.5281/zenodo.5045317

A. Appendix: C-SCALE Agreement template

C-Scale VA: [Use case name]

Use Case Information

Owner	
Customer contact[s]	
Customer organization	
C-SCALE User Forum Project	
Github / Website	
Customer status	*INACTIVE - customer became inactive* *NEW - an opportunity, a new entry. No engagement happened yet.* *EVALUATION - assessing the pilot set up* *PRODUCTION - production set up (agreement confirmed by all parties)*
News items	
Related projects	
First day of service delivery	
Last day of service delivery	
SLA approval date	
Providers	
Last review	
Next review	

Community roles

Role	Contact	Extra information
Data and access Manager		
User management		
Security manager		

AAI Configuration

AAI System	Name	Manager	Enrolment URL	Extra information
Check-in VO				
SRAM CO				

Providers/Installations

Provider	Contact	Installation	Allocated Units	Quotas	AUP	Start date	End date	Allowed IPs	Status	Operational check
									APPROVED approved by provider *DRAFT draft has been prepared*	*STARTED JIRA ticket or GitHub issue to the provider has been created* *FINALIZED C-SCALE verified configuration*

Service hours and exceptions

The Services operate during the following hours: twenty-four (24) hours a day, seven (7) days a week, three hundred sixty-five (365) days a year.

The following exceptions apply:

- Planned maintenance windows or service interruptions (“scheduled downtimes”) will be notified via email in a timely manner i.e. 24 hours before the start of the outage.

- Downtime periods exceeding 24 hours need justification.

Support

Support is provided via C-SCALE User Forum. Access requires a GitHub account. Support is available between:

- Monday to Friday.
- From 9:00 to 17:00 in the time zone of the relevant Resource Centres.

Service times always apply with the exception of public holidays in the country of the supporting Resource Centres.

Agreement reviews and customer satisfaction

The following table is updated after every review of the agreement.

Date	Review by	Service Satisfaction	Improvement Suggestions	Relevant Achievements

Actions log

Date	Action	Who	Comments