



C-SCALE

Enabling Copernicus Big Data
Analytics through EOSC

Impact report

1.1.2021–30.6.2023

TABLE OF CONTENTS

03 Preface	●	●	●	●
04 About C-SCALE	●	●	●	●
06 Key results				
06 KER 1 FedEarthData	●	●	●	●
07 KER 2 EO-MQS				
09 KER 3 openEO Platform	●	●	●	●
10 KER 4 Workflow solutions				
12 KER 5 C-SCALE Support				
13 Impact on Science				
13 Creating high-quality new knowledge Strengthening human capital in Research & Innovation				
14 Fostering diffusion of knowledge and Open Science				
14 Impact on Society				
14 Addressing EU policy priorities & global challenges through R&I Delivering benefits & impact via R&I missions				
15 Strengthening the uptake of R&I in society				
16 Impact on Economy				
16 Generating innovation-based growth				
17 Leveraging investments in Research & Innovation Creating more and better jobs				
18 Success Stories				
18 SAR on the fly WaterWatch				
19 Aquamonitor HiSea LSDA				

PREFACE

The C-SCALE project idea emerged from the evident need of the European research communities for an integrated compute and storage infrastructure for processing and analysing Copernicus data. Throughout the course of 30 months, and despite the unprecedented challenges brought forth by the COVID-19 pandemic, the project's consortium has successfully fulfilled its objectives, revolutionising the processing and analysis of Copernicus data. C-SCALE created a Big Copernicus Data Analytics platform through federating European data and compute providers and enabled and supported numerous researchers across Europe to access and leverage Copernicus data.

With this publication, we aim to provide readers with a compelling glimpse into the key achievements and impacts of the C-SCALE project, from enhancing data accessibility and scalability to empowering cutting-edge research. The remarkable advancements made by the consortium have opened doors to new insights, accelerated scientific progress, and paved the way for a future where data-driven research can thrive.

We extend our deepest appreciation to our data and compute providers who have contributed their expertise and commitment to the success of the C-SCALE project, as well as to our researchers for their invaluable feedback that helped us co-design the C-SCALE federation in a way that can meet users' needs, and we are looking forward to the expanding the federation with new collaborations!

ABOUT C-SCALE

The H2020 C-SCALE project bridges the gap between the vast amount of Earth Observation data available through the Copernicus program and empowers European researchers with advanced computing and analytics capabilities to extract valuable insights into the Earth processes.

The project implemented a Big Data Analytics platform by federating pan-European computing and data resources for Copernicus, including resources from the Copernicus DIAS platforms and the Collaborative Ground Segment nodes.

C-SCALE enables researchers to seamlessly access, process and analyse Earth Observation and Copernicus data through the European Open Science Cloud, unlocking their full potential, fostering collaboration, promoting innovation and supporting informed decision-making.



Project Summary

Dates

January 2021 - June 2023

Budget

~2 million €

Coordinator

Earth Observation Data Centre (EODC)

Partners



C-SCALE in numbers

3

Services in
EOSC

4

Workflows

21

Use case
applications

11

Training
sessions

8

SMEs
supported

KEY RESULTS



KER 1 - FedEarthData

Federated Earth System Simulation and Data Processing Platform

The Federated Earth System Simulation and Data Processing Platform (FedEarthData) brings together providers of data and processing capacity so that Earth Observation products held in distributed archives across the federation can be easily discovered and seamlessly accessed and processed on the batch as well as interactive analytic platforms deployed on distributed computing resources anywhere across the federation. The target groups for FedEarthData are researchers and service providers.

Value Proposition

Infrastructure-as-a-Service clouds, High-Throughput Compute, High-Performance Compute and storage resources all available from a single federated infrastructure. Availability of platform services to manage user access, complex workloads, train AI models, deploy automated clusters, run containerised applications and perform interactive analysis with Notebooks. Main benefits:

- Streamline the creation and execution of Copernicus and/or Earth Observation data processing workflows.
- Easy access to European computing infrastructure and smaller reliance on proprietary, commercial platforms.
- Access to a large repository of Copernicus data collections.
- Flexibility in the choice of platforms and APIs to fetch and process the data when compared to proprietary alternatives where everything is predefined for the users.

Technological status

The FedEarthData service was created during the C-SCALE project. To achieve this: 1) cloud providers have been integrated into the EGI Federated Cloud (using EGI Check-in as the AAI solution); 2) PaaS Orchestrator has been configured for use cases to deploy virtual infrastructure on top of the cloud federation easily; 3) the HTC/HPC federation has been created from scratch with HTC/HPC providers configuring SRAM as the AAI solution; Documentation, tutorials and examples have been provided to both end users and service providers to exploit and join the FedEarthData, respectively.

A successful proof of concept integration using SRAM (as community AAI) and EGI Check-in (as Infrastructure proxy) took place in May 2023. It will move to production in the next few months to homogenise access to all FedEarthData components in the near future.

The **Technology Readiness Level is 8** (system complete and qualified).

Access on
C-SCALE
Portal



Access on
EOSC
Marketplace



KER 2 - EO-MQS

Earth Observation Metadata Query Service

The Earth Observation Metadata Query Service (EO-MQS) makes Copernicus data distributed across partners within the federation discoverable and searchable. The target groups are researchers and service providers.

Value Proposition

The EO-MQS provides a standardised interface (STAC) to

search for metadata across the data federation. The result fosters the discoverability of Copernicus data and enhances the interoperability between data providers.

Main benefits: Users have a central point to query all the available data in the federation easily (even without having to install software clients when using the browser). Providers have a standardised way of cataloguing the data available on their premises. Thanks to the rich ecosystem that has evolved around STAC and the growing list of tools that can interact with STAC APIs, working with the EO-MQS is straightforward.

Technological status

The EO-MQS has been developed in C-SCALE from scratch. Its development is based on STAC-FastAPI, which helps build STAC-conformant solutions. The EO-MQS has been deployed in its initial version on the C-SCALE federated infrastructure. It currently supports the core parameters of the STAC API specification and conforms to STAC version 1.0.0. Extended API functionalities have also been developed, e.g., improving the sorting and filtering of results and excluding unnecessary attributes from the response.

Documentation of the EO-MQS and search and access data tutorials have been made available for all interested parties to both exploit its functionality as an end user and to join the EO-MQS as a data provider. Actually, in addition to the data provider in the consortium, external data providers have also joined (e.g. [RELIANCE project](#)) or are in the process of joining the EO-MQS.

The Technology Readiness Level is 7 (system prototype demonstration in operational environment).

Access on
C-SCALE
Portal



Access on
EOSC
Marketplace





KER 3 - openEO Platform

openEO platform provides intuitive programming libraries alongside a large Earth Observation data repository to simplify processing and data management. This large-scale data access and computation is performed on multiple infrastructures using the common openEO API, allowing use cases from explorative research to large-scale production of EO-derived maps and information in an accelerated way. The target groups are both researchers and service providers.

Value Proposition

openEO Platform provides a common API that can be used on multiple backend infrastructures to avoid vendor lock-in and facilitate the upscaling of processing pipelines. The platform has built-in functionality to simplify big data management for researchers. openEO platform provides a service for the public, private and research sectors, which eases access to EO data and compute resources.

Technological status

C-SCALE use cases (i.e. Aquamonitor and WaterWatch) have helped with the co-design of openEO platform, improving its usability. The openEO back-end has also been deployed for the first time across new service providers in a federation, which increased its portability. Via C-SCALE, openEO platform has raised its **Technological Readiness Level to 8** (system complete and qualified).

Access on
C-SCALE
Portal



Access on
EOSC
Marketplace



KER 4 - Workflow solutions

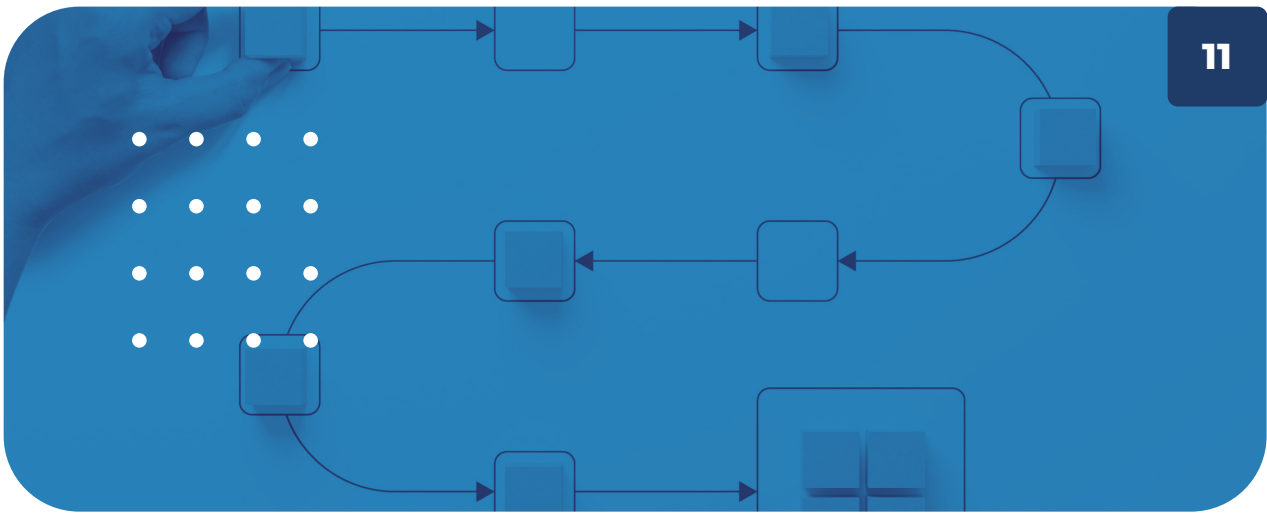
Solutions to easily deploy workflows supporting the Earth system's monitoring, modelling and forecasting. The target groups are both researchers and downstream application developers.

Value Proposition

The workflow solutions provide adaptable templates and examples, in the form of Jupyter Notebooks and workflow managers, of Earth Observation and Copernicus data and analysis workflows enabling users to more easily arrange a processing pipeline to create monitoring and forecasting applications on the FedEarthData service.

Technological status

Workflows have been refactored and deployed on FedEarthData, building on existing applications deployed elsewhere. Some of them have also been updated to



Access on
C-SCALE
Portal



Aquamo
nitor on
RoHub



include openEO as part of the analysis or refactored with Snakemake (a workflow management system).

Workflows have also been generalised (as far as reasonable) for reuse by other workflows and/or use cases. They have been redeployed across different service providers in the federation, improving their portability. Documentation with examples (e.g. using Jupyter Notebooks) has been provided for all of them in order to improve their uptake by external stakeholders. The workflows have also been published as Research Objects via <https://reliance.rohub.org> and they will be automatically ingested into the EOSC Marketplace to increase their visibility.

The **Technology Readiness Level is 7** (system prototype demonstration in operational environment).

HiSea on
RoHub

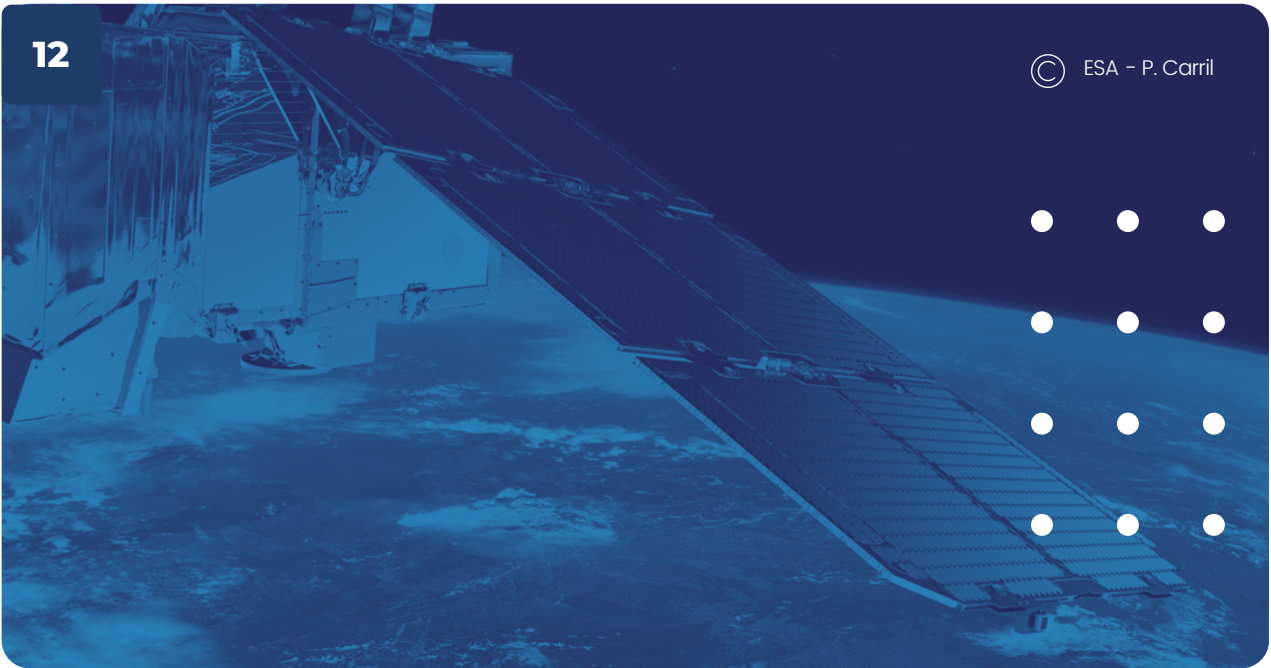


LSDA on
RoHub



WaterW
atch on
RoHub





KER 5 - C-SCALE Support

Set of activities and resources to engage with existing and new stakeholders, including both researchers and service providers in Earth Observation. It includes the platforms prepared to support the use cases (github, wiki, youtube, website), the material generated (documentation, publications, videos, tutorials, news articles), and the community of experts (both users and providers) interacting in the forum. The target groups are both researchers and service providers.

Value Proposition

Establish a two-way communication where service providers, users, experts and interested parties can exchange knowledge and best practices about optimising data analytics in Earth Observation.

C-SCALE
Documentation



C-SCALE
Community
Support



IMPACT ON SCIENCE

13



Creating high-quality new knowledge

Three publications in relevant peer-reviewed journals and more than 55 dissemination activities, including presentations in major relevant conferences in the sector (e.g. EGU, Living Planet Symposium, EOSC Symposium, EGI Conference, EODC Forum).

Follow-up activities ensure the publications get more citations and ensure the contribution to world-class science - impact in the longer term.



Strengthening human capital in Research & Innovation

Short-term: 11 workshops and tutorials shared on a Youtube channel and a project-specific wiki, including documentation and support material from all the relevant areas in the project, with an aggregated number of attendees of ca. 100 people and more than 500 online views (by June 2023).

In mid- and longer- terms: all partners, Academic (TU Wien), Research Organizations (INFN, Deltares, VITO) and Infrastructure providers (SURF, EGI, EODC, CESNET, GRNET, INCD and Cloudferro) can leverage the know-how generated in C-SCALE for their training activities and support to researchers.

C-SCALE
Youtube



C-SCALE
wiki





Fostering diffusion of knowledge and Open Science

Short-term: shared knowledge of more than 25 news publications on the [C-SCALE website](#).

The project has made its materials, including documentation, workflows, and software, openly available under permissive licenses (i.e. published under CC-BY when possible), and [FAIRNESS of data](#) has been a central point of the project.

In mid- and longer- terms: New collaborations / transdisciplinary collaborations: Knowledge generated within the use cases have opened new opportunities for follow up collaborations in cross-disciplinary research - such as the implementation of digital twins for science based on EO workflows developed and that potentially can enrich the large scale Earth digital twin ecosystem fostered under Destination Earth initiative (DestinE).

IMPACT ON SOCIETY



Addressing EU policy priorities & global challenges through Research & Innovation

8 SMEs have been using C-SCALE generated services (CloudFerro and EODC from C-SCALE consortium and Ubicube, Energy.Family, Tama Group, ITAINNOVA, KappaZetta, BioCarbon, Terradue, NivarIA via the open calls.

In mid- and longer- terms: it is expected that the EOSC Marketplace will create the necessary niche to incubate and consolidate the services and thereby contribute to the growth of the companies. The capabilities that C-SCALE will facilitate will enable scaling EO-based monitoring of key environmental aspects supporting the implementation of the European Green Deal policies, guiding Europe towards becoming carbon-neutral by 2050.



Delivering benefits & impact via Research & Innovation missions

C-SCALE has addressed EU policy priorities and global challenges through its research and innovation activities. By leveraging Earth Observation data and technologies, the project has contributed to various societal domains, including environmental monitoring, climate change mitigation, disaster management, and sustainable development. C-SCALE's Key Exploitable Results, such as the FedEarthData service, EO-MQS, openEO Platform, and Workflow Solutions, have the potential to empower policymakers, decision-makers, and stakeholders with valuable information for evidence-based decision-making and policy formulation.

As anticipated, new market opportunities (especially for European SMEs, including micro-enterprises) were fostered by bringing EO applications to EOSC. Thanks to the collaboration with EOSC DIH, a number of SMEs have been accessing the C-SCALE services on a test-before-invest approach enabling them to validate their solutions.



Strengthening the uptake of Research & Innovation in society

Co-creation activities: 6 initial Use Cases identified at the proposal stage helped to co-design C-SCALE services and make them tailored to the Domain Expert needs.

Further engagement: 15 additional Use Cases selected through the Open Call, which provided the experience to kick-start C-SCALE services, and create the momentum for their uptake.

In the mid-longer term, all use cases have a very significant impact in terms of societal uptake as they study the impact of climate change, environmental hazards, and agriculture, which are essential to ensure good living conditions & quality of life of European (and world-wide) citizens.

IMPACT ON ECONOMY



Generating innovation based growth

Innovative results: 5 Key Exploitable Results were generated - including the definition of 3 services and 4 main workflow solutions, and many others to be leveraged resulting from the 15 external use cases. Sustainability for

all the Key Exploitable Results have been granted, including a commitment from core organisations (KER owners) and infrastructure and technology suppliers.

In mid- and longer- terms: economic growth can be fostered by the expansion of the services among potential users, which opens the opportunity to follow up the research and development activities, but also opens the opportunity for SMEs and industry to make use of the platforms, and public authorities for evidence-based policymaking.



Leveraging investments in Research & Innovation

Co-Investment: By the end of C-SCALE, some of the use cases and activities are further developed under other follow-up projects & activities (such as the Horizon Europe interTwin project, the ESA-funded openEO platform and the INTERFACE project (enabled with national funding) and can potentially scale up in Destination Earth related digital twins. Altogether they represent an aggregated complementary direct funding of several 100k €.

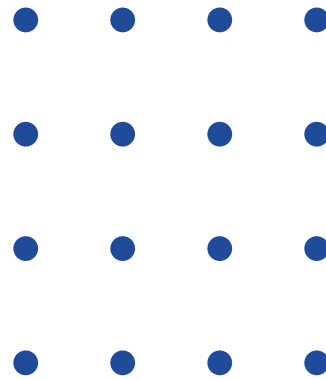


Creating more and better jobs

Around 80 persons among researchers, engineers and support teams have been involved in the project. C-SCALE's follow-up activities have been strategically designed to ensure the continued employment of these individuals within each participating organisation.

In the mid-longer term, take-up activities can foster the creation of new jobs as the usage of the services increases and further business and follow-up research opportunities emerge.

SUCCESS STORIES



Success story

Scientific impact

Societal impact

Economic impact



SAR on the fly

Scientific advancements in land subsidence monitoring, infrastructure stability assessment, and land cover classification.

Improved disaster preparedness, urban planning, and infrastructure management.

Identification of high-risk areas for land subsidence prevention and optimised land management decisions.



WaterWatch

Derive water surface dynamics from EO data for improved understanding of water resources and climate risks

Support response to climate risks, optimised water usage, and reduced socio-economic impacts of water-related disasters.

Reduced impact associated with water-related risks and improved water management strategies.

Success story

Scientific impact

Societal impact

Economic impact



Aquamonitor

Quantify land-surface changes due to man-made interventions, natural variability, and climate change.

Enhanced preparedness to climate change impacts

Support decision making and sustainable development



HiSea

High resolution model data for improved understanding of coastal ocean dynamics and water quality

Increased safety, efficiency, and sustainability in port operations and aquaculture activities.

Improved economic performance of ports and aquaculture industries through optimised operations and reduced risks.



LSDA

Ensemble river discharge forecasts for rivers for improved understanding of forecast accuracy

Improved river monitoring, flood forecasting, and water resource management.

Minimised flood damages, optimised water resource allocation, and increased efficiency in river-related industries.



Enabling Copernicus Big Data Analytics through EOSC



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

✉ contact@c-scale.eu

🌐 <https://c-scale.eu/>

🐦 [@C_SCALE_EU](https://twitter.com/C_SCALE_EU)

▶ [/@c-scale_project](https://www.youtube.com/@c-scale_project)

zenodo [/communities/c-scale](https://zenodo.org/communities/c-scale)

The content of this publication is correct as of June 2023.

The publication was developed by the following authors:

Xavier Salazar, EGI

Charis Chatzikyriakou, EODC

Björn Backeberg, DELTARES

Magdalena Brus, EGI

Sebastian Luna-Valero, EGI