C-SCALE

D1.2 Data Management Plan

Status:	Under EC Review	Planned due date:	30/06/2021
Version:	1.0	Submission date:	31/03/2022
Lead Participant:	EODC	Lead Author:	Charis Chatzikyriakou
Related WP:	WP1	Document Ref:	D1.2
Dissemination Level:	Public (PU)		
Document Link:	https://doi.org/10.5281/zenodo.6363615		

Deliverable Abstract

This document provides a report on the data types that are collected and generated per Work Package (WP), whether they will be shared with the public and how, whether they can be reused and how, and how they will be preserved during and after the end of the project. In this first version, the Data Management Plan is created based on the knowledge that the project members have so far, while a second – full – version will follow in the second half of the project.

COPYRIGHT NOTICE

DELIVERY SLIP

	Date	Name	Partner/Activity
Lead Author:		Charis Chatzikyriakou (CC)	EODC
Contributors:		Zdeněk Šustr (ZS)	CESNET
		Enol Fernández (EF)	EGI Foundation
		Björn Backeberg (BB)	Deltares
		Eleonora Testa (ET)	EGI Foundation
Moderated by:		Charis Chatzikyriakou (CC)	EODC
Reviewed by:		Małgorzata Krakowian (MK)	EGI Foundation
		Richard Kidd (RK)	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	04/06/2021	Deliverable document was created	CC, ZS, EF, BB, ET
V0.2	04/03/2022	First complete draft	CC
V0.3	08/03/2022	Updates to first draft from WP Leads	CC, ZS, EF, BB, ET
V0.4	11/03/2022	Further updates from WP Leads and first draft for external review	CC, ZS, EF, BB, ET
V0.5	29/03/2022	Integrated reviewers' feedback	CC, MK, RK
V1.0	31/03/2022	Version approved by AMB	C-SCALE AMB

Table of Contents

List of Acronyms

Acronym	Description
AMB	Activity Management Board
C-SCALE	Copernicus – eoSC AnaLytics Engine
DMP	Data Management Plan
EC	European Commission
FAIR	Findability, Accessibility, Interoperability, Reusability
ORDP	Open Research Data Pilot
WP	Work Package
FITS	Flexible Image Transport System
CIF	Crystallographic Information File
SRAM	SURF Research Access Management
VA	Virtual Access
EO	Earth Observation
MB	Megabyte
PB	Petabyte
TRL	Technology Readiness Level
CF	Climate and Forecast
EOSC	European Open Science Cloud

Executive summary

The C-SCALE (Copernicus – eoSC AnaLytics Engine) project brings together European commercial (e.g., Copernicus DIAS) and public data, computing and storage providers to deliver a federated infrastructure to support the Copernicus and Earth Observation (EO) user community. The project participates in European Commission's (EC) Open Research Data Pilot (ORDP) and is therefore required to maintain and deliver a Data Management Plan (DMP).

This document provides the first version of C-SCALE's DMP. In the sections below, the data types that are collected, generated and processed in the five Work Packages (WP) of the project from its start and until present (M15) are listed. Their characteristics as well as the FAIR principles that apply to them - and how – are detailed based on the DMP template provided by the EC and the Argos DMP platform.

1 Introduction

The Open Research Data Pilot is run by the EC with the main objective to improve and maximise the access and re-use of research data generated by Horizon 2020 projects. The ORDP promotes the FAIR principles, according to which the research data should be Findable, Accessible, Interoperable and Re-usable, and covers two major categories of data:

- the data needed to validate the project results presented in scientific publications, including the associated metadata,
- any other data, e.g., curated data not directly attributable to a publication, or raw data, including the associated metadata.

C-SCALE participates in this pilot, and with the present deliverable it is intended to provide a Data Management Plan for all data that are needed for the validation of the project results as presented in the scientific publications. More specifically, this DMP aims to provide information on the data management life cycle of all data generated, collected and processed within the project.

Therefore, this document provides information on the data types that are collected or generated per Work Package (WP), whether they will be shared with the public and how, whether they can be reused and how, and how they will be preserved during and after the end of the project. Initially, it was decided to use the open platform for Data Management Planning, Argos[1], created by OpenAIRE[2] and EUDAT[3]. Argos allows its users to create actionable DMPs that can be easily exchanged among infrastructures, making it a very useful tool for H2020 projects. However, since C-SCALE focuses on federating services rather than producing research data, it soon became obvious that the DMP template had to be simplified and adjusted to the project needs. The template used in this document has been created based on the one provided by the EC and the one used in Argos and adjusted to the characteristics of the data types used by C-SCALE. It is arranged in the following topics:

- High-level description
- Data summary
    - o Types of data, purpose of data collection/generation, data formats and standards, origin of data, expected data volume, data utility, scientific impact
- Data FAIRness
    - o Metadata, openly accessible data, repositories and preservation of openly accessible data, reusable data and their purpose, repositories and preservation of reusable data

After this first version of the DMP is submitted, it is intended to keep it as a living document. It will be updated every time there is a significant change or addition during the runtime of the project, and more details will be provided to the existing information. Later, a second version of it will be officially submitted. It shall be noted that the open access to scientific peer-reviewed publications is obligatory for all Horizon 2020 projects, and therefore they are not subject of the ORDP.

---

[1] https://argos.openaire.eu/splash/index.html
[2] https://www.openaire.eu/
[3] https://www.eudat.eu/

2 Data Management Plan per Work Package

2.1 WP1 Project Management

Main contact person	Charis Chatzikyriakou
Description	Data collected and generated to perform the Project Management activities according to the Grant Agreement and enable the communication and collaboration amongst the project members.
Data Summary	

Types of data	☐ Observational (e.g., sensor data, data from surveys)
	☐ Simulation (e.g., climate modelling data)
	☐ Derived or compiled (e.g., text mining, 3D models)
	☐ Reference or canonical (e.g., static, peer-reviewed data sets, likely published or curated, such as gene sequence databanks or chemical structures)
	☒ Other:
	• Project documentation
	  o In C- SCALE space in EGI Confluence:
	    ▪ Contracts and documents/guidelines related to the implementation of the project
	    ▪ Plans and procedures, KPIs and metrics, risk registry
	    ▪ Project and Board meeting material (agendas, Minutes of meeting, JIRA tickets)
	• Project deliverables stored in C- SCALE space in EGI Confluence
	• Personal data, i.e., names and e-mail addresses:
	  o For project communication, stored in EODC's mailing list platform and MS Teams administration
	  o For project collaboration, stored in EGI's Confluence
	• Effort and financial data of project partners collected in EU-Fin platform

Purpose of data collection/generation	☒ To obtain information ☒ To share information ☒ To keep on record ☒ To make informed decisions ☐ To develop a product ☐ To improve a product ☐ To combine with other data Comment: The data collection and generation under this WP intents to support the project management and the consortium during the implementation of the project.
Data formats and standards	☒ Text files - MS Word docs, .txt files, PDF, RTF, XML (Extensible Markup Language) ☒ Numerical - SPSS, Stata, Excel ☒ Multimedia - jpg / jpeg, gif, tiff, png, mpeg, mp4, QuickTime

	☐Models – 3D, statistical ☐ Software – Java, C, Python ☐ Discipline specific formats - Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) for crystallography
Origin of data[4]	☒ Primary data ☒ Secondary data Comment: All data are either collected from project partners or generated by the Project Management team. Exception to this are the plans and procedures that were reused from previous project and adjusted to the needs of C-SCALE.
Expected data volume	100 MB (megabyte)
Data utility	☐Researchers ☐Research communities ☐Decision makers ☐Education ☐Economy ☐Public ☐Industry ☒Other: Project Management team and consortium
Scientific impact	None.
Data FAIRness	
Metadata	N/A.
Openly accessible data	No.
Repositories of openly accessible data	N/A.
Preservation of openly accessible data	N/A.

Reusable data	• Project documentation
	  o In C- SCALE space in EGI Confluence:
	    ▪ Guidelines related to the implementation of the project
	    ▪ Plans and procedures
	• Project deliverables stored in C- SCALE space in EGI Confluence

Purpose	☐To reproduce and validate findings ☒To compare and combine with other data ☐To follow-up research on a specific area ☐To develop new products/services ☒Other:

[4] Primary data is data that have been collected for the first time and have not undergone through data processing and/or analysis, yet. Secondary data is data that have been cleaned up, analysed and shared by others (published or unpublished) and they are those that are being typically reused. Source: Argos (argos.openaire.eu).

	Comment: Implementation guidelines, plans and procedures can be adjusted accordingly and reused in future projects. Project deliverables can be reused both for scientific and Project Management purposes.
Repositories of reusable data	C- SCALE space in EGI Confluence
Preservation of reusable data	Available from the moment they are created and preserved for at least 5 years after the end of the project.

2.2 WP2 Copernicus Data Federation

Main contact person	Zdeněk Šustr
Description	Data Federation documentation: documentation of services and procedures in the C-SCALE Data federation.
Data Summary	
Types of data	☐ Observational (e.g., sensor data, data from surveys) ☐ Simulation (e.g., climate modelling data) ☐ Derived or compiled (e.g., text mining, 3D models) ☐ Reference or canonical (e.g., static, peer-reviewed data sets, likely published or curated, such as gene sequence databanks or chemical structures) ☒ Other: • User documentation, Admin documentation stored in the C-SCALE space in Confluence and README documents in GitHub repositories • Source code, Central service configuration stored in GitHub
Purpose of data collection/generation	☐ To obtain information ☒ To share information ☐ To keep on record ☐ To make informed decisions ☒ To develop a product ☐ To improve a product ☐ To combine with other data Comment: Documentation targeted at users and services administrators.
Data formats and standards	☒ Text files - MS Word docs, .txt files, PDF, RTF, XML (Extensible Markup Language) ☐ Numerical - SPSS, Stata, Excel ☐ Multimedia - jpg / jpeg, gif, tiff, png, mpeg, mp4, QuickTime ☐ Models – 3D, statistical ☒ Software – Java, C, Python ☐ Discipline specific formats - Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) for crystallography

	Comment: Plaint text (source code, configuration files), or widely established documentation text formats
Origin of data	☒ Primary data ☐ Secondary data
Expected data volume	MB (megabyte)
Data utility	☒Researchers ☒Research communities ☐Decision makers ☐Education ☐Economy ☐Public ☒Industry ☐Other:
Scientific impact	N/A.
Data FAIRness	
Metadata	N/A.
Openly accessible data	Open source code in public repositories, documentation publicly available on the Internet.
Repositories of openly accessible data	Open source code repositories (GitHub), open concurrent version system
Preservation of openly accessible data	Available immediately upon creation, preserved at least until July 2028.
Reusable data	• Source code reusable for developing similar SW components • Documentation reusable in community-specific extensions
Purpose	☐To reproduce and validate findings ☐To compare and combine with other data ☐To follow-up research on a specific area ☒To develop new products/services ☐Other:
Repositories of reusable data	GitHub
Preservation of reusable data	Reusable data (source code, documentation) will be preserved at least until July 2028

2.3 WP3 Copernicus Compute Federation

Main contact person	Enol Fernández
Description	Documentation related to the C-SCALE Compute federation and data collected and generated for its operations.
Data Summary	
Types of data	☐ Observational (e.g., sensor data, data from surveys) ☐ Simulation (e.g., climate modelling data) ☐ Derived or compiled (e.g., text mining, 3D models) ☐ Reference or canonical (e.g., static, peer-reviewed data sets, likely published or curated, such as gene sequence databanks or chemical structures)

	☒ Other:
	• User information for providing access to the providers in EGI Check-in and SRAM
	• In C- SCALE space in EGI Confluence:
	   ◦ VA installations' information (description, technology, resource amount, contacts)
	   ◦ VA metrics to monitor the usage of installations as defined in the project
	   ◦ OLAs/SLAs/AUPs for service delivery
	   ◦ Documentation and guidelines
	   ◦ Project meeting information, i.e., agendas, attendees, minutes of meeting
	• Configuration files in the C-SCALE Community in GitHub

Purpose of data collection/generation	☒ To obtain information ☒ To share information ☒ To keep on record ☒ To make informed decisions ☐ To develop a product ☐ To improve a product ☐ To combine with other data
Data formats and standards	☒ Text files - MS Word docs, .txt files, PDF, RTF, XML (Extensible Markup Language) ☒ Numerical - SPSS, Stata, Excel ☐ Multimedia - jpg / jpeg, gif, tiff, png, mpeg, mp4, QuickTime ☐ Models – 3D, statistical ☐ Software – Java, C, Python ☐ Discipline specific formats - Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) for crystallography
Origin of data	☒ Primary data ☒ Secondary data Comment: Created as a result of work done within the project.
Expected data volume	MB
Data utility	☐Researchers ☐Research communities ☐Decision makers ☐Education ☐Economy ☐Public ☐Industry ☒Other: Participants within the project and collaborators
Scientific impact	
Data FAIRness	
Metadata	N/A.
Openly accessible data	No.

Repositories of openly accessible data	N/A
Preservation of openly accessible data	N/A
Reusable data	• Configuration files • Documentation and guidelines
Purpose	☒To reproduce and validate findings ☐To compare and combine with other data ☐To follow-up research on a specific area ☒To develop new products/services ☐Other:
Repositories of reusable data	• C-SCALE Community in GitHub • C- SCALE space in EGI Confluence
Preservation of reusable data	Available from the moment they are created and preserved for at least 5 years after the end of the project.

2.4 WP4 User co-design and functional testing of Copernicus data and compute federation

Main contact person	Björn Backeberg
Description	Use case input and output data: WP4 is responsible for deploying use cases on the C-SCALE federation to test its usability and functional design and feedback via the User Forum to the C-SCALE federated infrastructure providers on how to improve its implementation. The use cases ingest Copernicus data to produce a result or output.
Data Summary	
Types of data	☒ Observational (e.g., sensor data, data from surveys) ☒ Simulation (e.g., climate modelling data) ☒ Derived or compiled (e.g., text mining, 3D models) ☒ Reference or canonical (e.g., static, peer-reviewed data sets, likely published or curated, such as gene sequence databanks or chemical structures) ☐ Other: Comment: Data from satellite remote sensing platforms, in situ observation networks and (data assimilative) numerical model simulations will be used and generated. The necessary data are sourced from within the federation when available, and if not, from the original source. The data are then stored at the providers supporting the use cases. The input and output data per use case are: <table><tr><td>**Use case**</td><td>**Input data**</td><td>**Output data**</td></tr></table>

		Land Surface Drought Analysis	• Hydro-MERIT • HydroLAKES • Global Reservoir and Dam Database (GRanD) • MODIS/Terra Leaf Area Index • CORINE Land Cover • SoilGrids • Global Land Ice Measurements from Space (GLIMS) • ERA5 reanalysis • SEAS5 seasonal forecast	• Model river discharge time series • Model discharge anomaly time series • Model soil moisture anomaly maps	
		Aquamonitor	• Sentinel-2 L1C	• Derived land surface change maps	
		WaterWatch	• Sentinel-2 L1C • JRC Water Occurrence	• Derived time series of reservoir surface water area • Derived maps of surface water extent for analysed reservoirs	
		HiSea	• Global Ocean Physics Reanalysis • Global ocean biogeochemistry hindcast • ERA5 • FES2012	• Model maps of ocean physics forecasts. • Model maps of ocean biogeochemistry forecasts.	
		RETURN	• Sentinel-1 ARD – Flattening Gamma • Sentinel-2 Surface Reflectance • Landsat-8 Surface Reflectance	• Derived maps of tropical forest recovery capacity • Derived statistics of tropical forest recovery capacity	

	Wetland Water Stresses	• Sentinel-1 γ_0^T • CORINE Land Cover • Sentinel-2 L2A • RAMSAR sites • Natura2000 areas	• Derived maps about the impact of designating protection zones on wetland behaviour	
Purpose of data collection/generation	☐ To obtain information ☐ To share information ☐ To keep on record ☒ To make informed decisions ☒ To develop a product ☒ To improve a product ☒ To combine with other data Comment: The use cases ingest a variety of EO and Copernicus data to generate products/results through analytics and/or modelling.			
Data formats and standards	☐ Text files - MS Word docs, .txt files, PDF, RTF, XML (Extensible Markup Language) ☐ Numerical - SPSS, Stata, Excel ☐ Multimedia - jpg / jpeg, gif, tiff, png, mpeg, mp4, QuickTime ☒ Models – 3D, statistical ☒ Software – Java, C, Python ☒ Discipline specific formats - Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) for crystallography Comment: Geospatial data; Image data; NetCDF is a commonly used data format.			
Origin of data	☒ Primary data ☒ Secondary data			
Expected data volume	PB (petabyte) Global scale analytics require PB-scale data volumes.			
Data utility	☒Researchers ☒Research communities ☒Decision makers ☒Education ☐Economy ☐Public ☒Industry ☐Other Comment: The Copernicus data delivered via the C-SCALE federation will facilitate downstream service providers, including research and industrial stakeholders.			

Scientific impact	It is not anticipated that the work in WP4 will result in scientific impact. The use cases deployed, have a relatively high TRL, and as such have already been published in the peer-reviewed literature.
Data FAIRness	
Metadata	Typically, the CF-compliant metadata standard is used
Openly accessible data	The use cases use a number of openly accessible datasets, which are listed, with their licenses, in the following:

Input Data	License
Hydro-MERIT	CC-BY-NC 4.0 license: Non-Commercial Use with less restriction. ODbL 1.0 license: Commercial Use is OK, but the derived data based on MERIT Hydro should be made publicly available under the same ODbL license
HydroLAKES	CC-BY-NC 4.0 license
Global Reservoir and Dam Database (GRanD)	Free for non-commercial use https://globaldamwatch.org/data/
MODIS/Terra Leaf Area Index	Public: This dataset is intended for public access and use. License: No license information was provided. If this work was prepared by an officer or employee of the United States government as part of that person's official duties it is considered a U.S. Government Work. https://catalog.data.gov/dataset/modis-terra-leaf-area-index-fpar-8-day-l4-global-500m-sin-grid-v006
CORINE Land Cover	Access to data is based on a principle of full, open and free access as established by the Copernicus data and information policy Regulation (EU) No 1159/2013 of 12 July 2013 https://land.copernicus.eu/faq/about-data-access
SoilGrids	CC-BY 4.0 License. https://www.isric.org/explore/soilgrids
Global Land Ice Measurements from Space (GLIMS)	Attribution 4.0 International license (CC BY 4.0). https://www.glims.org/RGI/
ERA5 reanalysis	https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-preliminary-back-extension?tab=overview
SEAS5 seasonal forecast	CEMS-FLOODS datasets licence https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-seasonal-reforecast?tab=overview
Sentinel-2 L1C	Sentinel Data Terms and Conditions

	JRC Water Occurrence	All data here is produced under the Copernicus Programme and is provided free of charge, without restriction of use. For the full license information see the Copernicus Regulation. https://www.copernicus.eu/en/about-copernicus/international-cooperation
	Global Ocean Physics Reanalysis	https://marine.copernicus.eu/user-corner/service-commitments-and-licence
	Global ocean biogeochemistry hindcast	https://marine.copernicus.eu/user-corner/service-commitments-and-licence
	FES2012	https://www.aviso.altimetry.fr/fileadmin/documents/data/License_Aviso.pdf
	Sentinel-1 ARD – Flattening Gamma	Sentinel Data Terms and Conditions
	Sentinel-2 Surface Reflectance	Sentinel Data Terms and Conditions
	Landsat-8 Surface Reflectance	https://www.usgs.gov/information-policies-and-instructions/copyrights-and-credits
	Sentinel-1 γ_0^T	Sentinel Data Terms and Conditions
	Sentinel-2 L2A	Sentinel Data Terms and Conditions
	RAMSAR sites	Creative Commons Attribution 4.0
	Natura2000 areas	EEA standard re-use policy: unless otherwise indicated, re-use of content on the EEA website for commercial or non-commercial purposes is permitted free of charge, provided that the source is acknowledged (https://www.eea.europa.eu/legal/copyright). Copyright holder: Directorate-General for Environment (DG ENV).

The accessibility to the use case output data is not fully determined yet, as the use cases are still being deployed. A detailed description of it will be included in the updated version of the document.

Repositories of openly accessible data		

Input Data	URL
Hydro-MERIT	http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/
HydroLAKES	https://www.hydrosheds.org/page/hydrolakes

	Global Reservoir and Dam Database (GRanD)	https://globaldamwatch.org/grand/
	MODIS/Terra Leaf Area Index	https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD15A2/
	CORINE Land Cover	https://land.copernicus.eu/pan-european/corine-land-cover
	SoilGrids	https://www.isric.org/explore/soilgrids
	Global Land Ice Measurements from Space (GLIMS)	https://www.glims.org/
	ERA5 reanalysis	https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5
	SEAS5 seasonal forecast	https://www.ecmwf.int/en/forecasts/documentation-and-support/long-range
	Sentinel-2 L1C	Sentinel Data Terms and Conditions
	JRC Water Occurrence	https://global-surface-water.appspot.com/download
	Global Ocean Physics Reanalysis	https://resources.marine.copernicus.eu/product-download/GLOBAL_REANALYSIS_PHY_001_030
	Global ocean biogeochemistry hindcast	https://resources.marine.copernicus.eu/product-download/GLOBAL_REANALYSIS_BIO_001_029
	FES2012	https://www.aviso.altimetry.fr/es/data/products/auxiliary-products/global-tide-fes/description-fes2012.html
	Sentinel-1 ARD – Flattening Gamma	https://scihub.copernicus.eu/
	Sentinel-2 Surface Reflectance	https://scihub.copernicus.eu/
	Landsat-8 Surface Reflectance	https://www.usgs.gov/landsat-missions/landsat-data-access
	Sentinel-1 γ_0^T	https://scihub.copernicus.eu/
	Sentinel-2 L2A	https://scihub.copernicus.eu/
	RAMSAR sites	https://rsis.ramsar.org/
	Natura2000 areas	https://www.eea.europa.eu/data-and-maps/data/natura-12
Preservation of openly accessible data	The input data needed will be preserved on the C-SCALE federation for 5 years.	
Reusable data	The openly accessible data listed above – used as input data for the use cases – are also reusable data.	

	The reusability of the use case output data is not fully determined yet, and similarly to their accessibility, a detailed description of it will be included in the updated version of the document.
Purpose	☒To reproduce and validate findings ☒To compare and combine with other data ☒To follow-up research on a specific area ☒To develop new products/services ☐Other:
Repositories of reusable data	See "Repositories of openly accessible data".
Preservation of reusable data	See "Preservation of openly accessible data".

2.5 WP5 Capacity building, dissemination, and exploitation

Main contact person	Eleonora Testa
Description	Outreach material: Public deliverables (publications), presentations, white papers and other reusable material to promote and document the project's outputs. Service Management Material: Service Catalogue, Capacity Plan, Use Case Approval Procedure
Data Summary	

Types of data	☐ Observational (e.g., sensor data, data from surveys)
	☐ Simulation (e.g., climate modelling data)
	☐ Derived or compiled (e.g., text mining, 3D models)
	☐ Reference or canonical (e.g., static, peer-reviewed data sets, likely published or curated, such as gene sequence databanks or chemical structures)
	☒ Other:
	• User training material and information in GitHub, C-SCALE website, C- SCALE space in EGI Confluence
	• In C- SCALE space in EGI Confluence:
	  o Service Catalogue
	  o Service description/registration Templates
	  o Capacity management tools
	  o Documentation and procedures
	  o Project meeting information, i.e., agendas, attendees, minutes of meeting
	  o Events schedule and registration
	  o Use Case applications
	  o Communication toolkit and branding
	• Published deliverables in Zenodo – in the C-SCALE Community
	• Data for the success of the project's communication and dissemination activities in C-SCALE website (Matomo add-on)

Purpose of data collection/generation	☐ To obtain information ☒ To share information ☒ To keep on record

	☒ To make informed decisions ☐ To develop a product ☐ To improve a product ☒ To combine with other data Comment: This dataset contains material that can be reused to present project's approaches and outcomes, service management implementation and knowledge base created.
Data formats and standards	☒ Text files - MS Word docs, .txt files, PDF, RTF, XML (Extensible Markup Language) ☒ Numerical - SPSS, Stata, Excel ☒ Multimedia - jpg / jpeg, gif, tiff, png, mpeg, mp4, QuickTime ☐ Models – 3D, statistical ☐ Software – Java, C, Python ☐ Discipline specific formats - Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) for crystallography
Origin of data	☒ Primary data ☒ Secondary data Comment: Created as a result of work done within the project.
Expected data volume	MB (megabyte) The maximum size of an individual document in the set is in the range of tens of MB, the total number of files is not expected to extend to thousands of files.
Data utility	☒Researchers ☒Research communities ☐Decision makers ☒Education ☐Economy ☒Public ☒Industry ☐Other Comment: Anyone interested in developing Copernicus or Earth Observation applications on an open, EOSC-compliant platform can benefit from the documents in this collection. Cross-Project initiatives and Knowledge base contribution in the area of Open Science and further.
Scientific impact	Facilitate the discovery and utilisation of Copernicus data through C-SCALE services.
Data FAIRness	
Metadata	N/A.
Openly accessible data	Yes. Dissemination and Knowledge purposes are ensured,

	Service catalogue, Training Material, Documentation and Procedures, Use Case Applications, Communication Toolkit& Branding, Published Deliverables.
Repositories of openly accessible data	C-SCALE Website, GitHub, Zenodo
Preservation of openly accessible data	Available from the moment they are created and preserved for at least 5 years after the end of the project.
Reusable data	Service Catalogue, Training Material, Published Deliverables, Communication Toolkit, Documentation and Procedures.
Purpose	☒To reproduce and validate findings ☒To compare and combine with other data ☒To follow-up research on a specific area ☒To develop new products/services ☐Other:
Repositories of reusable data	C-SCALE Community in GitHub C- SCALE space in EGI Confluence EOSC Portal
Preservation of reusable data	Available from the moment they are created and preserved for at least 5 years after the end of the project.

3 Conclusions and future work

With its participation in the Open Research Data Pilot, the C-SCALE project is required to create a DMP, make it publicly available and update it throughout the runtime of the project. Since the nature of the project is rather technical and focuses on federating services, not many research data are produced. Therefore, the original DMP template from the EC was simplified to provide the information in a more comprehensive way.

This document has provided the first version of the DMP by giving details on the types of data that are collected, generated and processed per WP as well as which FAIR principles apply to them and how, based on what the consortium has used until present (M15). A second version of it, which will be delivered towards the end of the project, will provide the complete DMP and more detailed information on the data used per WP.