HORIZON 2020
The EU Framework Programme for Research and Innovation

# Data Management Plan
# 2nd Update

Deliverable D7.6



**DATE**
19 April 2023

**ISSUE**
1.0

**GRANT AGREEMENT**
No 870337

**DISSEMINATION LEVEL**
PU

**PROJECT WEB-SITE**
http://cure-copernicus.eu/

**LEAD AUTHOR**
Giorgos Somarakis (FORTH)

**CO-AUTHORS**
Dimitris Poursanidis (FORTH)
Zina Mitraka (FORTH)
Zaheer Khan (UWE)
Mario Dohr (GEOVILLE)
Tomas Soukup (GISAT)
Nektarios Chrysoulakis (FORTH)

# CONTENTS

# 1 INTRODUCTION

## 1.1 Purpose of the document

This document is the 2nd update of the Data Management Plan (DMP) of the CURE project (Copernicus for Urban Resilience in Europe). It contains information required for the management of all data and products to be collected and generated during the project. The document outlines how data are handled during the project implementation phase and after it, describing the kind of data to be collected, processed, or generated, as well as the standards to be followed and their compliance with the FAIR guiding principles (Findable, Accessible, Interoperable and Reusable) (Wilkinson et al., 2016).

A sound data management is pivotal for fully realising the CURE project, as well as for allowing the wider research and user communities to access and further assess the project achievements. Within the project consortium, well curated data stimulate and ensure smooth collaboration between the project partners. Regarding the researchers and users, sound data management allows easy evaluation and deployment of the CURE data products. For communication, dissemination, and exploitation purposes, open access to publications and research data generated by the project helps to underpin its credibility and stimulates the uptake of the CURE results.

The CURE data management evolves during the project implementation phase. Thus, this document constitutes an update of the two CURE DMP previous versions (CURE Deliverables D7.4 and D7.5) and presents the relevant management and planning in the end of the project (month 40).

## 1.2 Definitions and acronyms

*Definitions*

Data Management Plan (DMP): Document outlining how the research data collected or generated by a project will be handled during and after the life of the project. It sets out: the data to be collected/generated, the methodology and standards to be used, whether and how the data will be shared and/or made accessible, how data will be curated and preserved (EC, 2021).

Open access to publications: Practice of providing free online access to scientific information in 2 main categories: peer-reviewed research articles (published in academic journals), research data (data underlying publications and/or raw data) (EC, 2021).

Open access to research data: Practice of making results public by providing access to digital research data (such as statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images) and giving the possibility to

re-use it. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user (EC, 2021).

<u>The FAIR Guiding Principles</u> (Wilkinson et al., 2016)

- To be Findable: (meta)data are assigned a globally unique and persistent identifier; data are described with rich metadata; metadata clearly and explicitly include the identifier of the data it describes; (meta)data are registered or indexed in a searchable resource.
- To be Accessible: (meta)data are retrievable by their identifier using a standardised communications protocol (the protocol is open, free, and universally implementable - the protocol allows for an authentication and authorisation procedure, where necessary); metadata are accessible, even when the data are no longer available.
- To be Interoperable: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation; (meta)data use vocabularies that follow FAIR principles; (meta)data include qualified references to other (meta)data.
- To be Reusable: meta(data) are richly described with a plurality of accurate and relevant attributes ((meta)data are released with a clear and accessible data usage license - (meta)data are associated with detailed provenance); (meta)data meet domain-relevant community standards.

## *Acronyms*

| | |
|---|---|
| API | Application Programming Interface |
| C3S | Copernicus Climate Change Service |
| CAMS | Copernicus Atmosphere Monitoring Service |
| CLMS | Copernicus Land Monitoring Service |
| CURE | Copernicus for Urban Resilience in Europe |
| DIAS | Data and Information Access Services |
| DMP | Data Management Plan |
| DOI | Digital Object Identifier |
| EC | European Commission |
| EMS | Emergency Management Service |
| EO | Earth Observation |
| EU | European Union |
| FAIR | Findable, Accessible, Interoperable and Reusable |
| GEMET | General Multilingual Environmental Thesaurus |
| INSPIRE | Infrastructure for Spatial Information in the Europe |
| SDGs | Sustainable Development Goals |
| UN | United Nations |

## 2  PROJECT OVERVIEW

Urban resilience refers to the ability of an urban system to maintain or rapidly return to desired functions in the face of a disturbance, to adapt to change, and to quickly transform systems that limit current or future adaptive capacity (Meerow et al., 2016). Mitigation and adaptation actions that enhance the resilience of cities need to be based on a sound understanding and quantification of the drivers of urban transformation and settlement structures, human and urban vulnerability, as well as local and global climate change, as defined by United Nations (UN) in the Sustainable Development Goals (SDGs) and the New Urban Agenda (UN, 2015; 2017).

Copernicus, as the means for the establishment of a European capacity for Earth Observation (EO), is based on continuously evolving Core Services. The CURE project used information from the Copernicus Land Monitoring Service (CLMS), the Copernicus Atmosphere Monitoring Service (CAMS), the Copernicus Climate Change Service (C3S) and the Emergency Management Service (EMS) to deal with the multidimensional nature of urban resilience (Figure 1). Also, it deployed data and products from contemporary satellite missions in order to achieve spatially disaggregated environmental information at local (neighbourhood) and city scales.
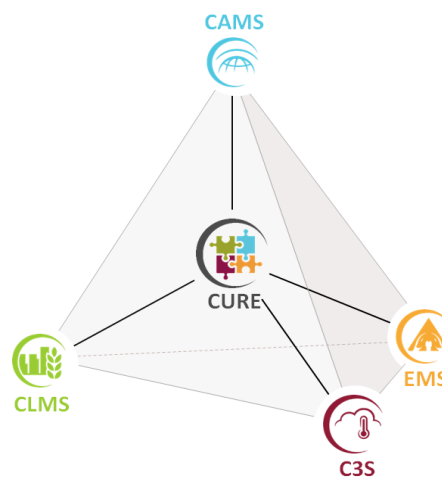


**Figure 1**. Conceptual illustration of the CURE system development as a cross-cutting structure, based on products derived from four different Copernicus Core Services: CLMS, CAMS, C3S and EMS. The CURE system is built on Copernicus Data and Information Access Services (DIAS) and contains cross-cutting applications related to urban resilience, reflecting urban sustainability dimensions.

Therefore, the CURE project addressed urban resilience, with the development of eleven cross-cutting applications among the above Copernicus Core Services, capable of coping with the required scale and granularity; by also integrating or exploiting third-party data, in-situ observations, and modelling. Moreover, it introduced novel ideas on how applications for climate change adaptation and mitigation, healthy cities, and social environments, as well as energy and economy can be developed across Copernicus Core Services.

# 3  DATA SUMMARY

The CURE consortium generated, collected, and handled geospatial data in the format of raster (e.g. GeoTIFF), vector (e.g. shapefiles, kml/kmz, etc.), and table (e.g. csv files with coordinates and environmental parameters such as temperature, precipitation, etc.). Also, data (i.e. email addresses of stakeholders) and products (i.e. reports, photos, videos, presentations, and documents of various formats) related to the communication, dissemination and reporting needs of the project were managed. Moreover, data from potential users were gathered through workshops and other ways of interaction. Indicatively, the total size of the CURE data and products is 2 TB.

The CURE data and products can be grouped as follows:
1. Copernicus Core Services data:
    a. CLMS
    b. CAMS
    c. C3S
    d. EMS
2. Copernicus Satellites data:
    a. Sentinel 1
    b. Sentinel 2
    c. Sentinel 3
3. Third-party data:
    a. Commercial satellite data from the European Commission (EC) data warehouse
    b. Meteorological data from in-situ measurement stations
4. Users' responses:
    a. Users' requirements for CURE applications
    b. User evaluation of CURE system and its applications
5. CURE applications data and products
    a. Sample data products
    b. CURE portal data products
6. Communication, dissemination, and reporting products:
    a. Internal project reports and confidential Deliverables
    b. Communication and dissemination material (articles, conference presentations and publications, posters, newsletters, leaflet, etc.)
    c. Public Deliverables
7. Personal contact information

The basic data of the CURE project correspond to Group 1, since it developed cross-cutting applications using large amount of data from the existing Copernicus Core Services (i.e. CLMS, CAMS, C3S and EMS). Information were also extracted from data of the Copernicus Satellite missions (Group 2). Furthermore, third-party data were used for algorithms calibration and

validation purposes (Group 3). Data were collected from potential users during the workshops, or with other interactive processes, for enhancing the CURE system and its applications based on users' perspectives (Group 4). The CURE applications data and products are related to the implementation of the CURE applications in various cities and concern sample data products (https://doi.org/10.5281/zenodo.6876179) and CURE portal data products (https://www.dropbox.com/sh/xivukxo3l2b1loq/AABXRKHvDaWYTSohZWIZWVETa?dl=0) (Group 5). Moreover, products related to the communication, dissemination, and reporting activities of the CURE project are considered in Group 6, while email addresses of the audience reached in the context of all communication and dissemination activities are included in Group 7. It has to be noted here that data from Groups 1, 2, 3, and 4 are data to be collected and used internally for the project implementation. Nevertheless, most of the data from Groups 1, 2, and 3 are also publicly available. Finally, the data and products from Groups 5, 6b, and 6c constitute the majority of the resulting products of the CURE project and are publicly available. Hence, this document mostly focuses on these last types of data and products.

Regarding CURE cross-cutting applications data and products (Group 5), local (neighbourhood) scale and city scale CURE products are provided from the implementation of these applications in various cities (https://portal.cure-copernicus.eu/city-demonstrations). Specifically, a range of cities were selected in the CURE project as case studies (Figure 2): highly urbanised cities (Berlin, Munich); typical Central (Basel), Western (Bristol, Vitoria-Gasteiz, San Sebastian), Northern (Copenhagen), and Eastern (Ostrava, Sofia) European medium size cities; as well as a low latitude Mediterranean city with dynamic urbanisation (Heraklion). Berlin, Copenhagen, Sofia and Heraklion had the role of front runner cities (i.e. CURE applications were built in these cities), whereas the remaining cities had the role of follower cities (i.e. CURE applications were transferred and tested to these cities).



**Figure 2**. The front runner (with blue) and the follower (with green) cities of the CURE project.

Moreover, publicly available communication, dissemination, and reporting products are specified in Deliverables D7.2 and D7.9, and mainly consist of: articles in scientific journals and conference proceedings (http://cure-copernicus.eu/journals.html); the CURE newsletter including 6 issues (http://cure-copernicus.eu/newsletters.html); publications for public consumption, dissemination and communication material including the CURE leaflet (http://cure-copernicus.eu/published_material.html); as well as Public Deliverables presented in Table 1 (http://cure-copernicus.eu/deliverables.html).

Table 1. The Public Deliverables of the CURE project.

| Deliverable (number) | Deliverable name | Delivery date (month) |
|---|---|---|
| D7.3 | CURE Web-site | 3 |
| D7.4 | Data Management Plan | 3 |
| D1.2 | Copernicus service review | 6 |
| D1.1 | Summary of user requirements | 7 |
| D1.3 | Methodology review and selection | 7 |
| D2.1 | Copernicus Core Services Interface and Relevant Data Portfolio Guide | 9 |
| D2.2 | Copernicus Core Services Interface | 12 |
| D3.1 | Urban Cross-cutting Applications Preparation | 12 |
| D6.1 | Scenarios for CURE integration to Copernicus | 12 |
| D4.1 | CURE System Requirements | 15 |
| D5.1 | Demonstration and Evaluation Methodology | 18 |
| D7.5 | Data Management Plan 1$^{st}$ Update | 18 |
| D3.2 | Urban Cross-cutting Applications Methods | 24 |
| D3.3 | Urban Cross-cutting Applications Sample Dataset | 24 |
| D5.2 | Users' Feedback on Demonstrations | 24 |
| D4.2 | CURE System Design | 25 |
| D4.3 | CURE System | 30 |
| D4.4 | CURE Application Solution Brochure | 30 |
| D6.2 | Benchmarking, Scenarios & Economic Feasibility Report | 30 |
| D4.5 | CURE Portal | 32 |
| D2.3 | Copernicus Core Services Interface Update | 36 |
| D3.4 | Urban Cross-cutting Applications Development Report | 36 |
| D5.3 | Demonstration and Evaluation Final Report | 39 |
| D7.6 | Data Management Plan 2$^{nd}$ Update | 40 |
| D7.8 | CURE Published Material | 40 |

# 4   FAIR DATA

All the CURE data and products from Groups 5, 6b, and 6c (publicly available project outputs) comply with the FAIR guiding principles (i.e. to be findable, to be accessible, to be interoperable, to be reusable), while the most of the other Groups data and products serve at least one guiding principle.

## 4.1   Making data findable, including provisions for metadata

The most of the CURE data and products can be characterised more or less as findable. However, these from Groups 5, 6b, and 6c serve this guiding principle the most, as each one of them is identified with a unique persistent identifier, i.e. a Digital Object Identifier (DOI), and described with rich standardised metadata that clearly include the persistent identifier. Towards increasing the findability of these data and products, the zenodo open access data repository is deployed and the CURE zenodo community (Figure 3) was created to host the most of the publicly available outputs of the CURE project, i.e. all products from Groups 6b and 6c as well as specific Group 5 data and products. Through zenodo communities, data and products can be collected, uploaded, curated and harvested, while each file upload can be combined with the following metadata information:

- Required: upload type, publication date, title, authors, description, access right, license.
- Recommended: communities, funding (grants), related/alternate identifiers.
- Optional: DOI, version, language, keywords, additional notes, contributors, references, journal, conference, book/report/chapter, thesis, subjects.



**Figure 3**. The zenodo community of the CURE project (https://zenodo.org/communities/cure-h2020).

Moreover, focusing on Group 5 data and products, they can be listed also in the official portal for European data (https://data.europa.eu), whereas the relevant data services can be findable through the DIAS platform. The Group 5 products were named following naming conventions of the existing Copernicus Services. For each of the products, proper keywords harvested from well-established and harmonised thesaurus, such as EnvThes (https://www.ecopotential-project.eu/images/ecopotential/documents/D5.6.pdf), GEMET (General Multilingual Environmental Thesaurus) (https://www.eionet.europa.eu/gemet/en/themes/) or other suitable vocabularies, were used. Also, clear version numbers, naming, keywords, and other metadata information indicate this distinction at the different steps. Additionally, Group 5 data are associated with metadata files compliant with the INSPIRE (Infrastructure for Spatial Information in the Europe) regulation and the metadata records can be indexed in a catalogue (e.g. within DIAS platform), linked with the related data.

## 4.2 Making data openly accessible

The accessibility status varies among the different CURE data and products types. Specifically, data from Groups 1 and 2 are already openly accessible. Group 3a data are not accessible, since they are commercial satellite data available to the CURE project from the Copernicus Space Component Data Warehouse. Group 3b data are openly accessible by the organisations owing the in-situ measurements equipment. Group 4 data are analysed and published through project Deliverables and other CURE publications. Data from Groups 6b and 6c are openly accessible through the CURE zenodo community and the CURE website (http://cure-copernicus.eu/). Group 6a data are confidential data shared within the CURE consortium and Group 7 data are accessible only to CURE partners responsible for communicating and disseminating CURE activities and outputs.

Finally, Group 5 data/products and their metadata are openly accessible through the CURE zenodo community and the CURE Portal (http://portal.cure-copernicus.eu/). The methods of the individual applications developed in the CURE project were deployed in DIAS. Therefore, the relevant data and products are accessible through the WEkEO DIAS service (https://www.wekeo.eu/), which was deployed for developing the CURE system; as it supports environmental data, virtual processing and skilled users as well as provides useful documentation along with the relevant software. Additionally, the CURE system development Git (Bitbucket) repository was used as a distributed version control system in order to store all code and documentation. The usage of the services is controlled via the user management system. Via an API (Application Programming Interface), there is a user creation with a bearer authentication access to the services. All services also need an additional license, which the user has to order. Thus, although everyone is able to create an account, the user is not allowed to order services without a respective license for them.

## 4.3 Making data interoperable

Almost all CURE data and products are interoperable and allow data exchange and re-use between researchers, institutions, organisations, countries, etc. Most data and products are in raster (e.g. GeoTIFF), vector (e.g. shapefiles), table (e.g. csv), text (e.g. pdf) and other formats, as they are described in Chapter 3. Such formats are fully compliant with openly available software (e.g. QGIS - https://qgis.org/) and allow re-combinations with data and products from different origins. Also, the CURE metadata support the interoperability of CURE data and products, as they accompany all publicly available CURE data and products and include useful information facilitating this principle. Especially, regarding Group 5 metadata, the metadata files include summary information on the data, indicative software to use for accessing, methodology information, references and licence information.

## 4.4 Increase data re-use (through clarifying licences)

The right to re-use is diverse among CURE data and products and depends on their types. On the one hand, data and products from Groups 3a, 4, 6a, and 7 are intended for CURE project internal usage and cannot be shared widely. On the other hand, Groups 1, 2, 3b, 5, 6b, and 6c include open access data and products associated with specific licenses, which are also defined in the corresponding metadata. Regarding data from Groups 1, 2, and 3b, their providers define the license types, while the license types for data and products from Groups 5, 6b, and 6c are defined by the CURE consortium or the respective publisher. Specifically, Creative Commons Attribution 4.0 International applies to products from Groups 6b and 6c published by the CURE project. Data from scientific publications are available according to the individual embargo time of the respective publisher, whereas Group 5 data and products are available for immediate re-use without embargo time. Group 5 data and products are provided under the Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license (https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode). This license allows to remix, adapt, and build upon provided work non-commercially, as long as the credit is provided and license of new creations is under the identical terms. Reproduction for non-commercial purposes is authorised, provided the source is acknowledged. Reports, articles, papers, scientific and non-scientific works of any form, including tables, maps, or any other kind of output, in printed or electronic form, based in whole or in part on the data supplied, must contain an acknowledgement of the source. Quality assurance processes were followed to ensure the quality of Group 5 data and products, described in D3.4.

# 5 ALLOCATION OF RESOURCES

All costs related to the CURE data management foreseen in the project were funded at least by its end. Regarding the CURE data storage, 1,134 € (without VAT) were spent for internal data (Data Groups 1, 2, 3, 4, and 6) and 6,521.76 € (without VAT) for data deposited on DIAS (Data Group 5). Costs regarding the DIAS platform are strongly connected with the

development of the services and the associated data and processing expenditure. Hence, the development of services is feasible under reasonable costs. Regarding the acquisition of commercial satellite data (Data Group 3a), very high-resolution imagery datasets covering 20,542 km$^2$ were requested and used.

# 6 DATA SECURITY

Apart from other repositories, all the CURE data and products are secured in FORTH internal data storage facilities with daily backup. Therefore, data recovery will be possible at any time and all the data and products are safely stored for long term preservation and curation. Moreover, FORTH repositories ensure secure storage and transfer of the sensitive data from Group 4, 6a, and 7. Raw data from Group 4 are stored on the UWE secure OneDrive and only processed data are transferred to FORTH repositories. Publically available data and products from Groups 5, 6b, and 6c are also deposited on the CURE zenodo community and DIAS. Thus, security of data and products in terms of recovery, storage and safety is ensured.

# 7 ETHICAL ASPECTS

No ethical or legal issues can have an impact on data sharing. Although data from Groups 4 and 7 are sensitive, since they contain personal data; they are not publicly available and openly shared. The processing and protection of personal data in the CURE project is subject to the rules set out in national, European Union (EU) and international law about the processing of personal data. Additionally, the CURE project does not collect any personal data falling within special categories of personal data, such as racial or ethnic origin, data related to religion, political opinion, etc. Finally, regarding Group 4 data, their analysis is utilised within the CURE project and processed results concerning the user requirements and evaluation regarding the CURE system and its application are provided in the public CURE Deliverables D1.1, D5.2, and D5.3. It has to be noted that in the context of gathering users' perspectives, templates of the informed consent forms and information sheets (in language and terms intelligible to the participants) were prepared and submitted as CURE Deliverable D8.1.

# 8 REFERENCES

CURE Deliverable D1.1: Summary of User Requirements. https://doi.org/10.5281/zenodo.4738613

CURE Deliverable D7.2: Dissemination and Exploitation Plan.

CURE Deliverable D7.4: Data Management Plan. https://doi.org/10.5281/zenodo.4739022

CURE Deliverable D8.1: POPD - Requirement No. 1.

EC "Funding & tender opportunities" Glossary. https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/glossary (accessed on 25 June 2021).

Meerow, S., Newell, J.P., Stults, M., 2016. Defining urban resilience: A review. *Landsc. Urban Plan.* 147, 38–49. https://doi.org/10.1016/j.landurbplan.2015.11.011

UN, 2015. Transforming our world: the 2030 Agenda for Sustainable Development (A/RES/70/1).

UN, 2017. New Urban Agenda (A/RES/71/256).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18