Deliverable D8.12

Data Management Plan (M36)

WP 8

Project Acronym & Number:	CityCLIM – GA 101036814
Project Title:	NEXT GENERATION CITY CLIMATE SERVICES USING ADVANCED WEATHER MODELS AND EMERGING DATA SOURCES
Funding Scheme:	Research and Innovation Actions (RIA)
Start date of the project:	01.10.2021
Duration:	36 months
Status:	Final
Dissemination Level:	Public
Author:	U. Koedel (Helmholtz Centre for Environmental Research GmbH – UFZ)
Contributors:	All Project Partners
Document Identifier:	D8.12_DMP_V100.docx
Date:	30.09.2024
Revision:	27.09.2024
Project website address:	https://cityclim.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036814

Project Summary

The strategic objective of CityCLIM is to significantly contribute to delivering the next-generation of City Climate Services based on advanced weather forecast models enhanced with data both from existing, but insufficiently used, sources and emerging data sources, such as satellite data (e.g., Copernicus data) or data generated by Citizens Science approaches for Urban Climate Monitoring etc. For City Climate Services, data products of interest related to land surface properties, atmospheric properties (e.g., aerosol optical thickness), geometry etc. For all of those, information of interest concerns e.g., Copernicus data products and services that are already existing (e.g., based on Sentinel-3/OLCI, PROBA-V, SPOT, Sentinel-1, MetopASCAT data), will exist in the near future (based on already flying satellites such as Sentinel-2), or will exist in the mid-term (based on satellites currently under development) and long-term (based on satellites soon starting concept phase) future. The project will establish; (i) an open platform allowing for efficient building of services based on access to diverse data; (ii) enhanced weather models based on data from diverse existing and emerging sources; (iii) a set of City Climate Services customizable to specific needs of users in cities; and (iv) a generic Framework for building next generation of Urban Climate Services. CityCLIM will be driven by 4 Pilots addressing diverse climate regions in Europe (Luxembourg, Thessaloniki, Valencia, Karlsruhe) which will define requirements upon the tools to be developed, support specification and testing of the services and serve as demonstrators of the selected approaches and the developed technologies. The Consortium will elaborate business plan to assure sustainability of the platform and services.

Project Consortium

	OHB-SYS	OHB System AG	DE
	OHB-DS	OHB Digital Services GmbH *(linked third party)*	DE
	OHB-DC	OHB Digital Connect GmbH *(linked third party)*	DE
	ATB	Institut für angewandte Systemtechnik Bremen GmbH	DE
	MTL	Meteologix AG	CH
	KAM	Kachelmann GmbH *(linked third party)*	CH
	OHB-DIG	OHB Digital Solutions GmbH	AT
	UFZ	Helmholtz-Zentrum für Umweltforschung GmbH	DE
	UVEG	Universitat de Valencia	ES
	CLT-UFA	CLT-UFA S.A. (RTL Luxembourg)	LU
	RCM	Region of Central Macedonia	GR
	VALENCIA	Ayuntamiento de Valencia	ES
	SK	Stadt Karlsruhe	DE

More Information

OHB System AG, OHB SYS (Coordinator)
Represented for the purposes of signing
the Agreement by: Dr. Stephan Holsten
E-Mail: stephan.holsten@ohb.de
Phone: +49 421 2020 9472

HRB 20411
Universitätsallee 27-29
28359 Bremen
Germany
Web: www.ohb-system.de

Dissemination Level

PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Change History

Version	Notes	Date
001	Creation of the document and sent to review	18.09.2024
100	Updated version after review	27.09.2024

Document Summary

The Deliverable D8.12-Final Data Management Plan is created on the basis of the D8.11. This D8.11 represents an initial version of the DMP, prepared during the first six months of the project lifetime and was regularly updated regularly during the project lifetime. This deliverable contains preliminary information about the data, the project will generate, whether and how it will be exploited or made accessible for verification and reuse, and how it will be curated and preserved. The DMP describes the standard formats, meaningful metadata, and open repositories to share data and enable other users to build on the knowledge gained during the project.

This deliverable also outlines how the research data collection or generation will be handled during and after the CityCLIM project, describes which data collection and generation methodology will be followed and whether and how data will be shared. It defines and establishes consistent practices among project partners to increase the efficiency and robustness of data handling during the project. The project will manage data in accordance with the principles of FAIR data management (Findable, Accessible, Interoperable and Reusable data) and this deliverable explains what information are needed to fulfil this tasks.

The proposed work in CityCLIM will fully comply with the regulations set out in the GDPR. In addition, CityCLIM comply with the principles of the European Charter for Researchers, the European Code of Conduct for Research Integrity, including ethical standards and guidelines, regardless country in which research is carried out (EC 2005, ALLEA 2017). Nothing in this project shall be deemed to require a party to breach any mandatory statutory law under which the party is operating. This includes any national or European regulations, rules and norms regarding ethics in conducting research. The ethical aspects impacting the CityCLIM project are described in detail in Ethics Requirements deliverable and responds to the findings of the ethics review performed by the European Commission in the proposal evaluation phase. However, since the ethic deliverables are confidential documents, a summary of the ethics and privacy strategy is included in this document.

A **data management plan in OpenAire** were established and published in ZENODO with following DOI https://doi.org/10.5281/zenodo.13777381. In addition, also other metadata regarding the data in the project were collected and summarized in this document.

Abbreviations

App	Software Application
D	Deliverable
DoA	Description of the Action
DMP	Data Management Plan
DoW	Description of Work
EC	European Commission
e.g.	Exempli gratia = for example
EU	European Union
GA	Grant Agreement
GDPR	General Data Protection Regulation
ICT	Information and Communication Technology
i.e.	id est = that is to say
IP	Intellectual Property
IPR	Intellectual Property Rights
KPI	Key Performance Indicator
M	Month
RTD	Research and Technological Development
SME	Small and Medium Sized Enterprise
T	Task
WP	Work Package

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Content and Purpose

Deliverable D8.12 - Data Management Plan is based on D8.11. The D8.11 is an initial version of the DMP prepared during the first six months of the project and has been updated regularly during the project as not all data or potential uses are well defined at this stage. This final version (D8.12) is released to the public in month 36 (project end).

The report contains information about the data the project will generate, whether and how it will be exploited or made accessible for verification and reuse, and how it will be curated and preserved. The purpose of the Data Management Plan is to provide an analysis of the main elements of the data management policy that the Consortium will use concerning all the datasets that the project will generate. The DMP describes the standard formats, meaningful metadata, and open repositories to share data and enable other users to build on the knowledge gained during the project. The first audience to which this report is addressed is the internal partners; there are thirteen partner organizations from Germany, Greece, Spain, Switzerland, Austria and Luxembourg.

This deliverable also outlines how the research data collection or generation will be handled during and after the CityCLIM project, describes which data collection and generation methodology will be followed and whether and how data will be shared.

Firstly, the DMP defines and establishes consistent practices among project partners to increase the efficiency and robustness of data handling during the project. The second audience of this report is the community of researchers, engineers, and biofuel producers interested in the data and the tools of the CityCLIM.

The DMP was continuously updated until the end of the project. Open access (OA) refers to the free provision of scientific information (data, findings, publications, etc.) to others through online means. In the question, why make my research open access; the justification lies on economic and ethical principles as:

- the undertaken research has been paid with public funds and
- there is no reason to pay again to gain access to this information when it is required for use by other researchers, industries, students, or citizens.

According to the "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020" (EC 2017), it is encouraged to provide open access to scientific publications and data, as this contributes to:

i) the quality improvement of research by building on a more substantial body of existing work,
ii) increase in efficiency of research by reducing duplication of effort,
iii) innovations to market quicker by reducing barriers to information flow,
iv) enhancement of the transparency of scientific progress.

The first step after each research is to decide how to disseminate its results. The two options are either to publish the research findings or to keep confidential some aspects for commercial exploitation.

2 Building a data management plan in the context of H2020

2.1 Purpose of the CityCLIM Data Management Plan

A Data Management Plan (DMP) is a crucial framework for managing data during and after a research project, ensuring efficiency, ethicality, and security. It outlines the types of data to be collected, methods for data collection, storage, and analysis, and procedures for data quality and integrity. The DMP also addresses data sharing and accessibility, specifying how and when data will be shared with others or the public and the formats for data availability. It is essential to comply with legal, ethical, and funder requirements, including data protection laws, obtaining necessary permissions, and handling sensitive data appropriately. The plan should also outline strategies for anonymizing or de-identifying personal data to protect privacy and confidentiality.

A DMP facilitates data preservation and future usability, outlining procedures for long-term storage, including the choice of repositories or archives and how data will be maintained over time. It promotes transparency and reproducibility in research by providing a clear record of data handling practices, allowing other researchers to understand and replicate the study, validating results, and contributing to the scientific community. The DMP ensures data is findable, accessible, interoperable, and reusable (FAIR) in accordance with the Horizon 2020 Open Research Data Pilot.

2.1.1 Requirements for Research Data

The European Union's Horizon 2020 framework includes specific provisions on Data Management Plans (DMPs) to ensure effective and responsible data management in research projects. Under Horizon 2020, all projects participating in the Open Research Data Pilot are required to develop and implement a comprehensive and dynamic DMP. The main aim of these rules is to promote open access to research data, ensuring that data produced with public funding is accessible, reusable and preserved for future use.

The regulations require a DMP to be submitted within the first six months of the project. This plan should provide detailed information on the data that will be collected, processed or generated. It must outline the methods for data collection and management, specifying how data quality will be maintained and how metadata will be created and managed. The DMP should also describe the procedures for data storage, including the physical and digital infrastructures that will be used to ensure data security and integrity.

One of the key components of the Horizon 2020 DMP rules is the emphasis on data sharing and open access. The plan must detail how data will be made available to other researchers and to the public, including any restrictions that may apply due to confidentiality, security or other ethical considerations. Researchers are required to deposit their data in trusted repositories, and the DMP should identify these repositories, along with the standards and formats that will be used to facilitate data interoperability and reuse.

The requirements of the participating projects regarding research data are listed in the following. The participating projects should, according to OpenAire (2022):

- manage all research publications, data, and other outputs following the FAIR principles,
- make research data openly accessible immediately (make data accessible at the latest within 30 days of generation) and in line with the FAIR principles,
- develop provisions for access to the data if open access is not possible because of exceptions as described in GA Article 29.3; principle "As open as possible, as closed as necessary.",
- deposit quality-controlled research data in a data repository as soon as possible and within the deadlines set out in the data management plan (DMP) ensures persistent and unique identifiers (PIDs) usages, keep in focus sustainability, Metadata, curation and quality assurance, access (e.g., free and easy access to reuse), security, privacy, common format, provenance (e.g., maintains a detailed log file of changes to datasets and metadata),
- provide essential information via the repository about any research output or any other tools and instruments needed to reuse or validate the data,
- include metadata of deposited data under a Creative Commons Public Domain Dedication (CC 0 1.0) or equivalent, in line with the FAIR principles (in particular machine-actionable).

As part of making research data findable, accessible, interoperable, and reusable (FAIR), a DMP should address the following essential aspects (OpenAire 2022):

- a description of the data generated/collected, including data types and an estimate of its size within the project.
- whether and how the data will be made accessible for verification and reuse, along with relevant security and privacy considerations.
- how the research data will be managed (organized, curated, accessed, shared, preserved)
- timelines of when generated data will be made open access.
- an estimation of curation and storage/preservation costs; person/team responsible for data management and quality assurance processes

2.1.2 DMP value

Data Management Plans are an essential mechanism for research groups to ensure their outputs are FAIR and ensure that data management becomes a core part of all research practice (EC 2018). The added value of a DMP can be summarized in the following points:

- To understand how to reuse and exploit data.
- To increase the quality of research by defining institutional and project-specific data management policies and plans in accordance with relevant standards and community best practice.
- To ensure that all research outputs are findable for everyone/available to people by using persistent identifier.
- To understand strength and weaknesses (e.g. what discipline/ unit produces more results? Are data described based on DMP requirements? etc.
- To ensure research integrity and excellence of researchers.

2.2 Legal Framework

General Data Protection Regulation (GDPR, since May 2018) is applicable in all EU Member States and the countries in the European Economic Area (EEA) to strengthen citizens' fundamental rights and guarantee their privacy in the digital age (Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, on the protection of natural persons concerning the processing of personal data and the free movement of such data, and repealing Directive 95/46/EC.

The individual rights include the right to (EC 2022):

- Obtain information about the processing of personal data.
- Obtain access to the personal data held about one person.
- Ask for correction of incorrect, inaccurate, or incomplete personal data.
- Request that personal data be erased when it is no longer needed or if processing it is unlawful.
- Object to the processing of personal data for marketing purposes or on grounds relating to a particular situation.
- Request the restriction of the processing of personal data in specific cases.
- Receive personal data in a machine-readable format and send it to another controller ("data portability").
- Request that decisions based on automated processing concerning one person or significantly affecting this person and based on this personal data are made by natural persons, not only by computers. In this case, there is the right to express the personal point of view and to contest the decision.

3 Data Management Plan Overview

3.1 CityCLIM Data Management Procedure

Figure 1 illustrates the Data Management Procedure. In the CityCLIM project, various data will be produced, used, or derived from satellite, airborne, in-situ, and historical data. Satellite data are accessed mainly by existing open-access hubs, e.g., Copernicus (https://scihub.copernicus.eu/). Copernicus provides complete, free and open access to different Sentinel products. If these products are processed, e.g., in a different format, the information should be included in the DMP. Airborne data include airborne sensors and unmanned aerial vehicle **(**UAV): LST and LSE maps from the brightness temperature data. Apart from this, measured data include in-situ data or data collected from citizen science (e.g., subjective perception of heat). Historical weather and model data will be used to assess the temporal variation in specific parameters to be included in forecast models or specific CityCLIM engines. A significant part of the data generated in the project is model data. Following exemplarily models are used in the project: UltraHD Weather Model Processor with the inclusion of CAMS model aerosol, Urban Heat Forecast Engine with the inclusion of soil model, Advanced UltraHD model to forecast conversion of gases (O3<→NO2, etc.) and generation, transport and sedimentation of dust or PM, several simulation scenarios. Other external data are required for model input and simulation, such as orography (land height), land use (urban, forest, agriculture etc.), albedo, and vegetation data (e.g., type, density, green fraction).

Within the data workflow, all these data will be classified into three groups given by available , derived , and measured data, to

- clarify the data provenience and the status of anonymization,
- define the level of openness,
- check if the data are in line with the FAIR principles, and,
- if the required metadata is complete, to store them accordingly in the repository or to provide licenses and to use this data in the CityCLIM Services, tools, and platforms.

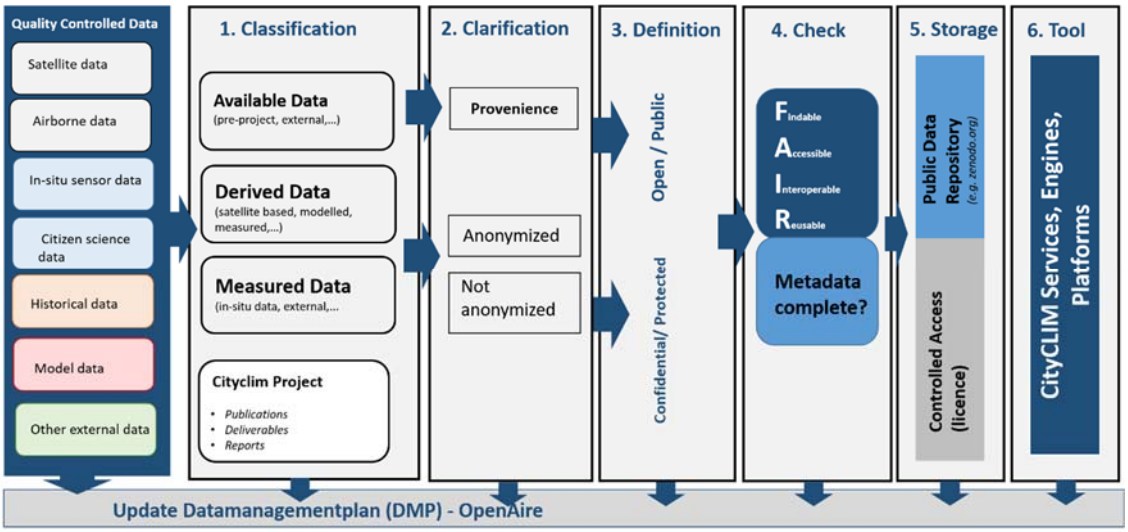

Figure 1: CityCLIM Data Management Procedure.

CityCLIM uses some data produced by citizen science. With regards to possible personal data, the Consortium ensures that

- data on individuals are transmitted and used in a secure environment,
- the use of the data complies with ethical and legal requirements (including signed informed consent and applicable data protection laws), and, that
- the use of both existing and new data is agreed with the data owner/data provider.

Data records containing personal data are managed in accordance with the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679).

3.2 Who will be responsible for data management in the project?

Data management activities concern the whole project and need to be coordinated and monitored at project and work package levels. Data management is also linked to the publication of project results and thus dissemination activities. Therefore, the following roles and responsibilities were identified and described in detail in D8.11. The responsibilities in CityCLIM were defined and the list is available within the Consortium in the document repository solution Sharepoint from Microsoft.

3.3 Data Management Plan in OpenAire

OpenAIRE is a pan-European research information system, which provides services to find, store, link and analyse research output from all disciplines. OpenAIRE aims at promoting and implementing the directives of the EC and the European Research Council on the promotion and funding of science and research. OpenAIRE supports the Open Access Mandate and the Open Research Data Pilot developed as part of the Horizon 2020 projects. OpenAIRE recommends to use the Creative Commons CC0 Waiver or CC-BY licence for open access to data. As a standard DMP tool, ARGOS provides DMP templates that meet the requirements and suggestions of the Horizon 2020 Data Management Guidelines.

CityCLIM used a collaborative online tool that helps project members plan their Research Data Management activities (*https://argos.openaire.eu*). A guideline on how to add parameter descriptions to the CityCLIM DMP was distributed to the data managers of each WP.

A DMP template was created for the CityCLIM project, data descriptions were filled in by the data managers during the course of the project and the DMP is finally published in Zenodo.

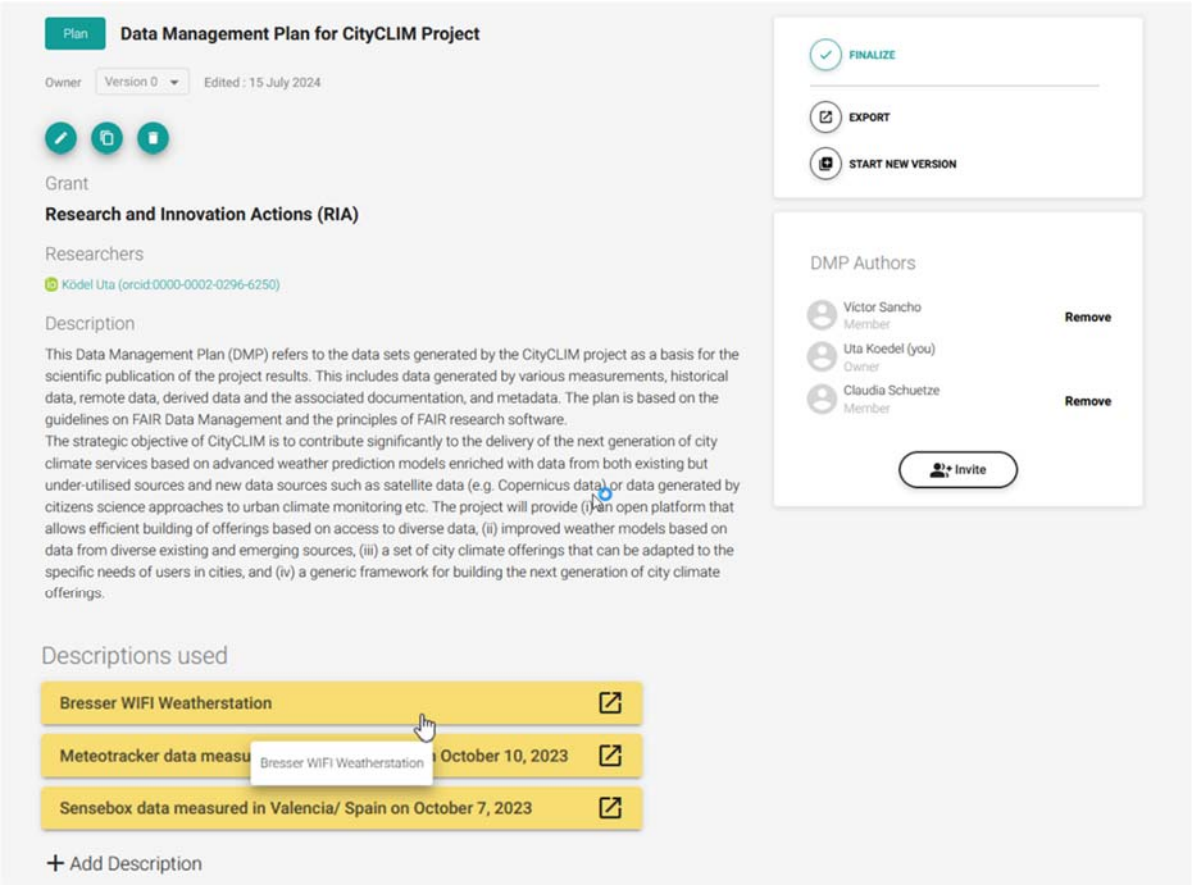

Figure 2. CityCLIM DMP in OpenAire

The DMP template for H2020 projects provided by the EC and used in OpenAire includes following points:

- A summary of your data

- How to make your data FAIR
  - Making data **f**indable, including provisions for metadata
  - Making data openly **a**ccessible
  - Making data **i**nteroperable
  - Increase data **r**e-use (through clarifying licences)
- Information about costs and resources
- Information about data security
- Ethical aspects
- Other issues

The CityCLIM DMP is published in ZENODO with following DOI: https://doi.org/10.5281/zenodo.13777381

3.4 CityCLIM Data in ZENODO Repository

If you want to make your research data open, you can deposit it in a special data repository. Some repositories accept both publications and datasets, such as Zenodo. Data repositories allow you to provide persistent links to your datasets. This allows them to be cited, linked and tracked. You can license your data to control the level of reuse you allow, just like other publications.

Zenodo is a "catch-all" open research data repository to gather research data across various scientific disciplines. It is for non-military purposes only. The repository is hosted and managed by CERN, and all data deposited to ZENODO is stored securely in the CERN Data Centre's cloud infrastructure. This multidisciplinary repository accepts multiple data types, publications, and software and assigns a Digital Object Identifier (DOI). Zenodo, an OpenAIRE and CERN collaboration, allow researchers to deposit both publications and data, while providing tools to link them.

Exemplary data sets were uploaded to the multidisciplinary open research data repository **Zenodo** (https://zenodo.org/) to comply with the H2020 Open Access Mandate. For that reason, a community CityCLIM were created and all data were uploaded within that community. Before, a guideline on how to upload datasets to the CityCLIM community was distributed to the data managers of each WP. The ZENODO repository and OpenAire have also been linked by including the DOI in the parameter description.

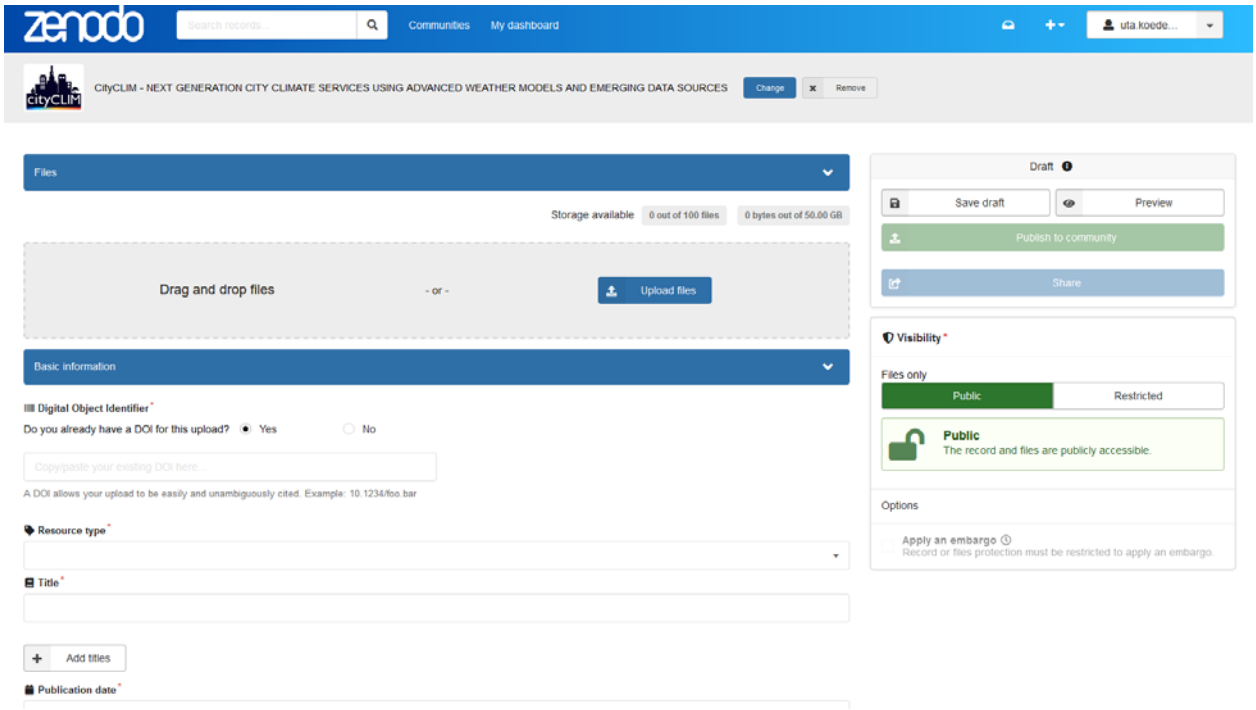

Figure 3. Template for uploading a dataset to the CityCLIM community in ZENODO

4 Data Summary

4.1 Data summary at a final state

Usually, an extended dataset description is required to describe different datasets or types of datasets that will be produced, derived, or reused during the project. Usually following important points are considered:

Table 1. Important Points and questions and their explanation regarding data summary

Points/Questions	Explanation
What is the purpose of the data collection/generation and its relation to the project's objectives?	Data collection is usually at the beginning stages of research data management lifecycles to set the background of what is needed (data generation), what is already there (data reuse), and how to best use it to fulfill the project's objectives (why).
What types of data will the project generate/collect?/ Will you reuse any existing data and how?	The main distinction between data types is between primary data and secondary data. Data collection may contain both primary and secondary data depending on the source where they have been derived from. Primary data has been collected for the first time and has not undergone data processing and/or analysis yet. Secondary data is data that has been cleaned up, analyzed, and shared by others (published or unpublished), and they are those that are typically reused. The expected data types include but are not limited to: – Primary/Secondary – Integers – Booleans – Characters – Floating-point numbers – Alphanumeric strings – Other (to be specified)
What formats of data will the project generate/collect?	Data may be found in different formats. Usually, outputs of specific instruments/ services/ tools where data are generated and/or processed have a default way of exposing their outputs following specific standards. The most common distinctions between formats are proprietary and nonproprietary formats. The expected data formats include but are not limited to: – ASCII text-formatted data (TXT) – CAD data (DWG) – Comma-separated values (CSV) – NetCDF – dBase (DBF) – eXtensible Mark-up Language (XML) – Tab-delimited file (TAB) – Geospatial open data based upon JavaScript Object Notation (GeoJSON) – Geo-referenced TIFF (TIF, TFW) – Hypertext Markup Language (HTML) – Keyhole Markup Language (KML) – MS Word (DOC/DOCX) – MS Excel (XLS/XLSX) – MS Access (MDB/ACCDB) – OpenDocument Spreadsheet (ODS) – Open Document Text (ODT) – Rich Text Format (RTF) – SPSS portable format (POR) – Other (to be specified)
What is the origin of the data?	It is essential to know the primary point where the data was generated or collected.
What is the expected size of the data (in Bytes)?	Provide measurable examples per type of output based on common practices in the field. These might be relevant to the volumes of data and how many bytes of storage they occupy, numbers of objects, files, rows, and columns.

Points/Questions	Explanation
To whom might it be useful ('data utility')?	Data generated or reused in the project can be useful for several stakeholders and third parties. Think about the target audience of your research, but also about possible third parties which could further exploit this data even after the project ends. ***→ Data Repository → ZENODO***

4.2 Naming Convention

Descriptive file names are essential for organizing, sharing, and keeping track of data files. Therefore, a naming convention needs to be set- up in each project considering the essential project aspects. Best practices in file naming consider the following facts:

- Consistent naming of all files and documents.
- File names should be short but descriptive (<25 characters).
- Special characters or spaces should be avoided.
- Use capitals and underscores instead of periods or spaces or slashes.
- Use date format ISO 8601: YYYYMMDD.
- Version number is obligatory.
- Describe naming convention in DMP.

Table 2 lists the proposed naming conventions for documents and research data. As stated in the D9.1 Quality Assurance Plan, each document in CityCLIM are identified with an appropriate ID equivalent to the document file name. We have applied the following naming convention also for research data.

Table 2. CityCLIM naming convention.

Specifications	Naming Convention		Reference
Document	**CityCLIM-ddmmyy-** **xxxxxxxx_v000_yyy.aaa**	• CityCLIM – Project indicator, • ddmmyy – optional – day, month and year when the file was prepared or used (especially relevant for presentation material) • xxxxxxxx – Short name of the document like deliverable name/ number or report number if necessary (e.g. D1.1, D3.2, R2, etc.). • V01 – version of document. • yyy – optional: the name or organisation like "V002_ATB". This will facilitate the integration of documents and help us to identify the latest file version, • aaa – file extension, e.g. » .docx » for Microsoft Word documents, etc.	D9.1 Quality Assurance Plan
Research data	**CityCLIM_WPx_X_xxxxxxx_v0x_yyy.aaa**	• CityCLIM – Project indicator, • WPx- Work Package • X- Pilot site identification codes are as follows: K:Karlsruhe, V: Valencia; M: Macedonia, L: Luxembourg • xxxxxxxxxx – Short name of the data set (maximum 8 characters) • V0x – version of document. (e.g. V01) • yyy –the name or organisation to be responsible for the dataset • aaa – file extension	

4.3 Versioning

Data versioning related to different versions of data storage that were created or changed at specific times. In the case of research data, a new version of a dataset may be created when an existing dataset is reprocessed, corrected or appended with additional auxiliary data. Versioning allows to track changes associated with 'dynamic' data that is not static over time.

Unique versioning allows for repeatability in data processing, enables data comparisons, and prevents confusion. It is a prerequisite for FAIR data handling and increases the trustworthiness of all kind of data. Numbering of data versions is a relatively well established in praxis. A consistent version numbering scheme enables data users to:

a) track changes and availability of new versions,
b) track which version is the previous version and with which version data processing was done,
c) allow tracking the differences of each version. Therefore, the CityCLIM naming convention requires the information on the version.

Zenodo provides DOI versioning of all datasets uploaded to their communities, which allows to edit and update the uploaded datasets after publishing. Therefore, it is also possible to cite specific versions of an upload and cite all versions of an upload.

4.4 Metadata provision

Metadata is simply information about data. It makes the data meaningful to machines. All users can understand and track important details of the work if the data is properly described and documented. When data is deposited in a data repository, metadata about the data not only describes the data, but also makes it easier to search and retrieve. This can include content such as contact information, geographic locations, details about units of measure, abbreviations or codes used in the dataset, instrument and protocol information, survey tool details, provenance and version information, and much more. Metadata formats and standards can take many different forms. They can range from free text to standardised, structured, machine-readable, extensible content. It is recommended that you use a standard metadata format that is used in your research area. Some specific examples of metadata standards are: Dublin Core (domain agnostic, basic and widely used metadata standard) and ISO 19115 and FGDC-CSDGM a Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata for describing geospatial information.

Figure 4 describes the required administrative Metadata for documents that needs to be provided for each dataset.

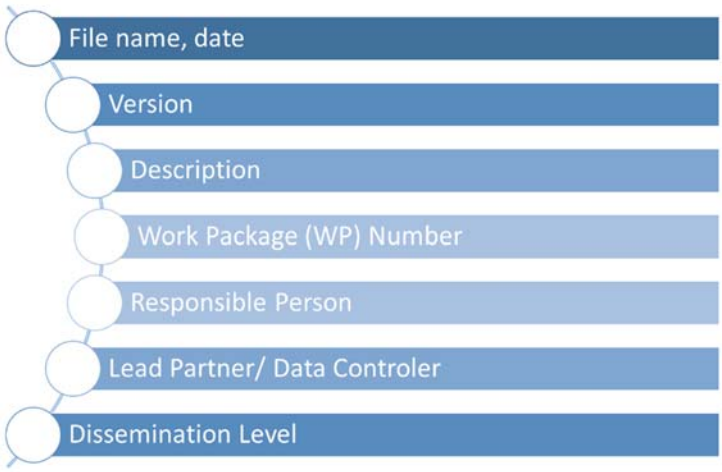

Figure 4. Required metadata information

Aside from administrative Metadata, every technical dataset must include Metadata describing the technical quantities, measurement- and/or processing procedure, unities, and optionally, if known, confidence parameters such as standard deviation. As an example, the INSPIRE platform of the European Commission provides guidelines for metadata provision and technical implementation for spatial data [EC INSPIRE 2013]. Technical Metadata can be understood as part of data handling.

4.5 Data Handling and Management

To provide a comprehensive overview of the data within CityCLIM, a questionnaire collecting the essential information of some exemplarily dataset was carried out by the responsible data managers.

Table 3. Metadata information on datasets

NR	*Number of Dataset*
WP	*To which Working Package the dataset relates to?*
Data Reference and Name	
Dataset description	*Please describe the dataset content: Parameter, units, temporal and spatial resolution?*
Purpose of these data	*What is the purpose of the data collection/generation?*
Relation to project objectives	*What is the relation to the objectives of the project?*
Data creator /Data owner	*Who is the data creator owner?*
Primary [measured]/Secondary dataset [derived with help of other data]	*Data collection may contain both primary and secondary data depending on the source where they have been derived from. Primary data is data that have been collected/modelled 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.*
If primary dataset; which sensor was used	*Please describe the sensor used for this data collection!*
Datatyp [datafiles, pictures,...]	*What kind of datatyp is it?*
Use standardized vocabularies	*Does the dataset use standardized vocabularies? If yes, please specify which standard is used?*
File Format [csv, jpg, json, tiff, ...]	*Please specify the file format of the data. Examples are txt, csv, netcdf, Json, tiff, kml, xls/xlsx*
Reuse of existing data; if yes please specify!	*Will you reuse any existing data (also auxiliary data) for processing and interpretation and how? Or if this dataset is a secondary dataset which existing data have been used to create this dataset?*
Temporal resultion [s, min, h, day,...]	*What is the temporal resolution of the dataset?*
Spatial resolution [m, km,...]	*What is the spatial resolution of the dataset?*
Boundary of spatial resolution(Latitude and Longitude in WGS64)	*Please provide the boundary of the spatial information of this dataset!*
Data production method	*Please describe the data production method or cite a reference if possible? (Manual/Semiautomatic)*
Used QA/QC method	*Do you use any quality assurance and quality control methods during the data processing?*
Data storage (where, size)	*Please describe the data utility (e.g. Servers, file systems, graph database management). Where are the data stored? What is the expected size of the data (in Bytes) for a Day/Month/Year?*
Describe potential for reuse	*Can this data provide an added value for others in the project and ouside? Please describe?*
Potential stakeholders for reuse	*Data generated or reused in the project can be useful for a number of stakeholders and third parties. Think about the target audience of your research, but also about possible third parties who could further exploit this data even after the project ends. To whom might it be useful?*
Usuage rights available	*Are this data Open Access (open data) or restricted data (closed data)? If not, please explain the reason? Are there any access conditions (licence, permits,...)? Does the data needs to be anonymized?*
Data retention	*Data retention is the storing of information for a specified period. Please describe this for the data?*
Follow FAIR principles [YES/NO]	*Does this dataset follow the FAIR principles?*
Curator (Contact person)	*Who is responsible for that dataset?*
Exemplary data available? If yes, please specify!	*Is an exemplary dataset uploaded to ZENODO? If yes, please provide DOI!*
Dataset included in OpenAire Data Management Plan	*Is this dataset fully described in the DMP in OpenAire?*

5 General Provisions - FAIR Data Management

The project aims to maximize access to and reuse of research data generated by the CityCLIM project. Hence, CityCLIM will manage data in accordance with the principles of FAIR data management– see also section 3.1. At the same time, there are datasets, or parts of datasets, generated in this project that cannot be shared due to, e.g., protection of the privacy of voluntary participants. The rules for FAIR data management and their questions that need to be answered are described in the following sections according to EC (2016). All these questions are implemented in the OpenAire Tool to create a DMP.

The following points needs to be considered:

- Making data findable, including provisions for metadata,

- Making data openly accessible,
- Making data interoperable,
- Increase data reuse (through clarifying licences),
- Allocation of resources,
- Ensuring Data security,
- Considering ethical aspect,
- Others.

6 Data generated in the CityCLIM Project

In Figure 3 all data sources are listed, which are required as input data for the generic city climate platform. This data can be divided into available, measured, derived (and historical data) and include remote sensing and airborne data, in-situ meteorological data, model and simulation data, and other auxiliary data.

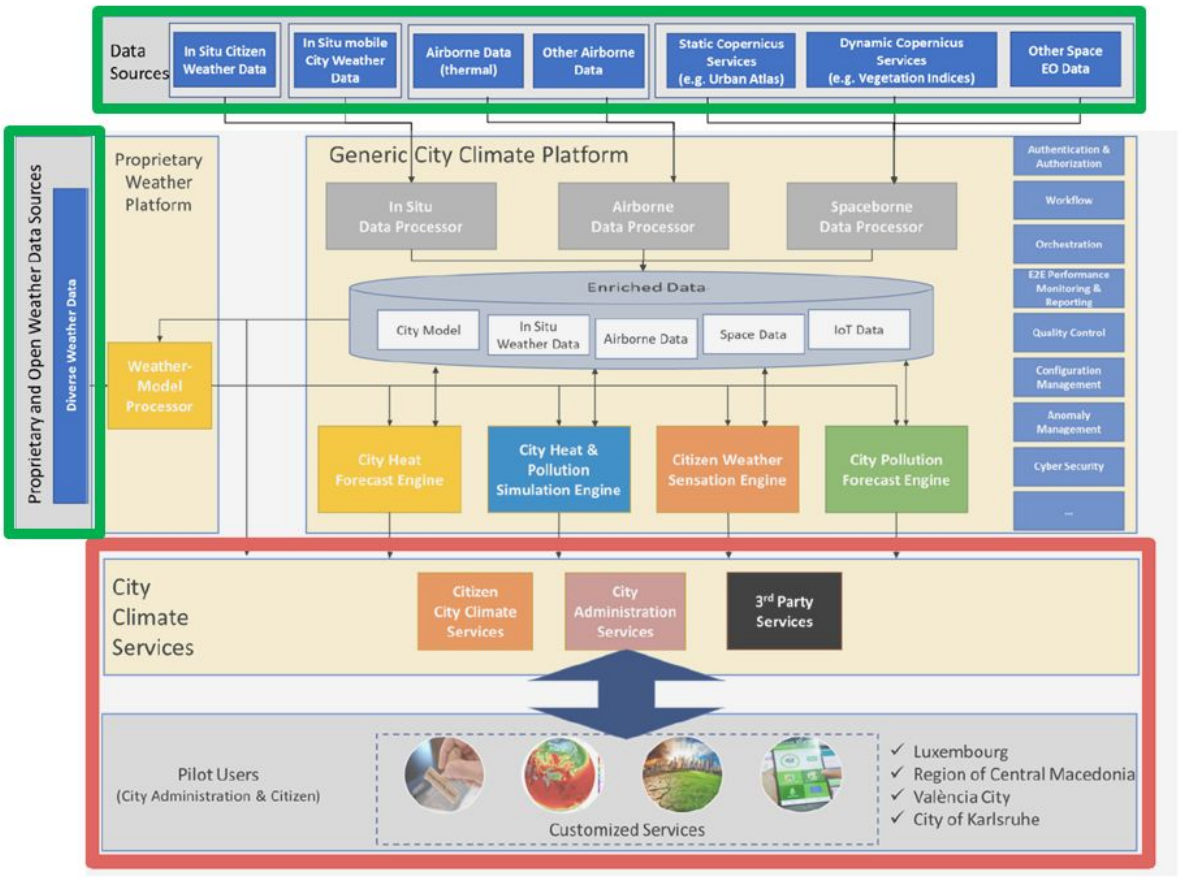

Figure 5. Data (green box) in the CityCLIM project according to proposal.

7 Ethics

7.1 Ethical Requirements

The ethical requirements for a Data Management Plan (DMP) in Horizon 2020 projects are strict and comprehensive. They aim to ensure that data are managed responsibly and ethically throughout the research lifecycle. A fundamental requirement is to ensure that personal data remains protected. Researchers must comply with the General Data Protection Regulation (GDPR). This requires that personal data is processed lawfully, fairly and transparently. The DMP must outline the specific measures that will be taken to protect personal data, including procedures for obtaining informed consent from participants, methods for anonymisation or pseudonymisation of data to protect identities, and protocols for secure storage and transmission of data.

Adherence to confidentiality agreements is another critical ethical requirement. Researchers need to ensure that sensitive information, whether it is personal or proprietary, is protected from unauthorised access and disclosure. The DMP should detail strategies for maintaining confidentiality, including the use of encryption, secure access controls, and restricted data sharing practices.

Ethical considerations related to data sharing and open access also need to be addressed in the DMP. While Horizon 2020 promotes the principle of open science, researchers must balance this with the need for protection of sensitive information. The DMP should identify any ethical or legal restrictions on sharing data, explain how these will be addressed, and justify any data that cannot be made openly accessible. In cases where data sharing is restricted, the plan should describe alternative mechanisms to ensure that data can be accessed in a controlled and ethical manner.

Measures to ensure data integrity and prevent scientific misconduct should also be included in the DMP. This includes the establishment of procedures for data verification, validation and documentation to maintain accuracy and reliability. The plan should outline the steps to be taken to detect and deal with any instances of data fabrication, falsification or plagiarism.

The ethical implications of data use and reuse must also be considered by researchers. The DMP should ensure that data are used in a way that respects the rights and dignity of participants and that any secondary use of data is consistent with the original consenting agreements. This includes providing clear information about how the data will be reused and ensuring that any new uses of the data do not compromise the privacy or welfare of participants.

Finally, provisions for ongoing ethical review and monitoring should be included in the DMP. This includes the regular updating of the plan to consider any changes in data management practices, the emergence of ethical issues, or new regulatory requirements. It also includes establishing mechanisms for continuous monitoring and evaluation of ethical compliance throughout the project.

7.2 Ethical Issues

Citizen science also raises ethical issues that should be addressed before projects begin and throughout the scientific investigation. Ethical issues include data quality and integrity, data sharing and intellectual property, conflicts of interest, and exploitation. The most important is all issues regarding collecting and handling personal data. Scientists and researchers must always adhere to a certain code of conduct when collecting data from people and should consider aspects such as protecting the rights of research participants, enhancing research validity, and maintaining scientific integrity. Also, legal structures are established and in place to manage Citizen Science data that regulate the evaluation and storage of data and their accessibility. There are also binding standards relating to the handling of data rights. Coordination and data information offices are set up as points of contact to offer guidance on handling data (copyright, administration, and use). Volunteers actively involved in scientific activities, for example, when collecting data, are covered by insurance through additional agreements. However, in many Citizen Science projects, aspects of copyright or intellectual property issues are not adequately considered.

All data collected from stakeholders in the project will be following applicable ethical standards and requirements in the respective countries of the data collection, as well processed and handled securely and in line with applicable rules and regulations on privacy and data protection. Since we do not know the exact scope and nature of the data at the beginning of the CityCLIM project, the ethical issues are under consideration during the project's progress.

7.3 European and national regulations

Germany, Spain and Greece are members of the European Union, and all countries are subject to EU law. Therefore, the most important data protection regulation is EU Regulation 2016/679 (the GDPR). In addition, there are several legislations in every country. This Law regulates among others the rules for processing personal data, formalities to obtain consent to process personal data, special rules when processing personal data about children and special rules when processing personal data about employees. Sensitive data are personal data consisting of racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, data concerning health or data concerning a natural person's sex life or sexual orientation. This kind of data can only be processed if a condition for processing special category data is satisfied but are not relevant in the CityCLIM Project.

Some critical issues according to this GDPR are (GDPR, 2022):

All processing of personal data must comply with all six general data quality principles. Personal data must follow six principles:

- ***lawfulness, fairness and transparency principle***: processed fairly, lawfully and transparently
- ***purpose limitation principle:*** collected for specific, explicit and legitimate purposes and not processed in a manner incompatible with those purposes
- ***data minimization principle***: adequate, relevant and not excessive
- ***accuracy principle***: accurate and, where necessary, up to date
- ***storage limitation principle:*** kept in an identifiable form for no longer than necessary
- ***integrity and confidentiality principle***: kept secure

The processing of personal data must also satisfy at least one condition for processing personal data. These conditions are that the processing is: (a) carried out with the data subject's consent; (b) necessary for the performance of a contract with the data subject; (c) necessary for compliance with a legal obligation; (d) necessary in order to protect the vital interests of the data subject; (e) necessary for the public interest or in the exercise of official authority; or (f) necessary for the controller's or recipient's legitimate interests, except where overridden by the interests of the data subject (DSGVO, 2022).

Table 4. Regulations in the partner countries (DLA Piper Intelligence (2022)).

Country	Regulations
Germany	• German Federal Data Protection Act (Bundesdatenschutzgesetz, or BDSG): in Sections 1–44, the BDSG contains supplementary regulations on the GDPR putting data controllers' rights and responsibilities in more concrete terms. Sections 45–85 of the BDSG implement the provisions of EU Directive 2016/680. • In addition to the BDSG, each federal state has its federal data protection law. These state laws only affect public bodies. • Germany does not have one central Data Protection Authority but a number of different Authorities for each of the 16 German states (*Länder*) that are responsible for making sure that data protection laws and regulations are complied with. • German Federal Commissioner for Data Protection and Freedom of Information (*Bundesbeauftragte für Datenschutz und Informationsfreiheit* – 'BfDI') is the Data Protection Authority for telecommunication service providers and represents Germany in the European Data Protection Board. • Germany's Federal Government set up the Data Ethics Commission (Datenethikkommission) on July 18, 2018. It was given a one-year mandate to develop ethical benchmarks and guidelines and specific recommendations for action, aiming at protecting the individual, preserving social cohesion, and safeguarding and promoting prosperity in the information age.
Spain	• The Spanish Data Protection and Digital Rights Act 3/2018 (the "**Data Protection Act**") helps implement the *GDPR* and creates a new digital charter of rights. • The Spanish Research Ethics Committee (Comité Español de Ética de la Investigación, EEI). Created by the Law of Science,32 under the Council for Science, Technology and Innovation Policy, as a collegiate body of an independent and consultative character, on matters related to professional ethics in scientific and technical research
Greece	• The Greek Law 4624/2019 "on the Hellenic Data Protection Authority, the implementation of Regulation 2016/679 and the transposition of Directive 2016/680" (hereinafter the "Law") (Government Gazette A/137/29.08.2019) was enacted and entered into force in August 28, 2019. The law regulates the operation of the Hellenic Data Protection Authority, introduces GDPR supplementary rules and transposes the Law Enforcement Directive into Greek Law. • lead supervisory authority :Hellenic Data Protection Authority (HDPA)

Country	Regulations
Luxembourg	• Two Luxembourg Data Protection Laws of August 1, 2018 have been enacted to implement the GDPR: (1) The Law on the organization of the National Data Protection Commission (CNPD) and the general data protection framework. It gives the framework for the CNPD's organization, composition, and powers under the GDPR and the applicable national law. (2) The law on the protection of individuals deals with the processing of personal data in criminal matters as well as regulates matters of national security. • Article L. 261-1(1) of the Labor Code provides specific regulations concerning employer workplace surveillance. • In addition, the amended Law of May 30, 2005 on data protection and electronic communications governs the protection of personal data in the field of telecommunications and electronic communications, implementing the Directive 2002/58/EC. • lead supervisory authority: Commission Nationale pour la Protection des Données (CNPD)
Austria	• The existing Data Protection Act 2000 (*Datenschutzgesetz 2000*) was amended by the Data Protection Amendment Act 2018 (*Datenschutz-Anpassungsgesetz 2018*) which constituted the first implementation of various regulations related to GDPR, and was intended to enter into force simultaneously with GDPR. • The 'Data Protection Act' (*Datenschutzgesetz, DSG)* has considerably amended the Data Protection Act 2000. In addition to the GDPR, it is now the central piece of legislation in Austria regulating data privacy. • The Privacy Deregulation Act 2018 (*Datenschutz-Deregulierungs-Gesetz 2018*) further amended the DSG. • Further amendsments were included in the General Data Protection Adjustment Act (*Materien-Datenschutz-Anpassungsgesetz 2018*) and the research-sector specific Data Protection Adjustment Act – Science and Research (*Datenschutz- Anpassungsgesetz 2018 – Wissenschaft und Forschung – WFDSAG 2018*)
Switzerland	• The Federal Act mainly regulates the processing of personal data on Data Protection of June 19, 1992 (DPA) and its ordinances, i.e., the Ordinance to the Federal Act on Data Protection (DPO) and the Ordinance on Data Protection Certification (ODPC). • In addition, the processing of personal data is further restricted by provisions in other laws, mainly concerning the public sector and regulated markets. • The revised DPA will still provide for certain deviations from the GDPR provisions, thus requiring specific "Swiss Add-Ons" in several areas.

The proposed work in CityCLIM will fully comply with the regulations set out in the GDPR. In addition, CityCLIM comply with the principles of the European Charter for Researchers, the European Code of Conduct for Research Integrity, including ethical standards and guidelines, regardless country in which research is carried out. Nothing in this project shall be deemed to require a party to breach any mandatory statutory law under which the party is operating. This includes any national or European regulations, rules and norms regarding ethics in conducting research. The ethical aspects impacting the CityCLIM project are described in detail in deliverable 10 Ethics Requirements and respond to the findings of the ethics review performed by the European Commission in the proposal evaluation phase. However, since this D10.1-3 is a confidential document, a summary of the ethics and privacy strategy is included in this document.

7.4 Ethical Measures in CityCLIM project

Following ethical measures were applied within the CityCLIM Project and described in D9.2:

1. The ethical guidelines are based on the vision of using science and technology to create a better society and are reviewed continuously to ensure they stay up to date with developments in society and the challenges of today. They generally fall into these categories: research ethics, business ethics, and ethics in interpersonal relationships (Hagendorff 2020)
2. Citizen Science data collection follows the six principles mentioned above, lawfulness, fairness, and transparency principle, purpose limitation principle, data minimization principle, accuracy principle, storage limitation principle and integrity and confidentiality principle.

3. All data collection activities (interviews, surveys, etc.) will be designed to maintain privacy. Personal data will not be requested unless this is absolutely necessary. Vulnerable groups like minors and individuals unable to freely provide an informed consent will be excluded.
4. All project partners are obliged by European and national law (GDPR) to protect personal data. To ensure compliance with all applicable ethical and privacy requirements, the privacy strategy encompasses all data assurance activities that will be performed in the context of the CityCLIM project. Under no circumstance will the deliverables or processes compromise the individual right to privacy and satisfactory handling of personal data.
5. All personal data will be collected only upon receiving informed consent from the participants, and any participant providing personal data can at any time withdraw their participation and related data from the project.
6. The participants will be given an information letter and a consent form (on paper or electronically). The information letter will provide information about:
    a. type of the collected data
    b. How the data will be collected (interview, automatic data collection, etc.)
    c. What the data will be used for. The information letter will explain the purpose of the project and the expected results. It will also be explained that published information always will be anonymous, and that no personally identifiable information will be published in any way.
    d. How the data collected will be handled. The information letter will explain that personal data will be treated in full confidentiality and will be registered and stored in a secure manner. The data will be de-identified before it is processed (name or other characteristics serving to identify a person will be replaced by a number, and the list of identifiers will be kept separate from the data).
    e. Who will have access to the data. The information letter will state that a very limited number of scientists will handle data and that confidentiality will be regulated by legal agreements. The data will be de-identified before discussing and processing within the project.
    f. The rights of the participants. The information letter will state that participation is voluntary and that participants have the right to see the data collected about them and that they can withdraw from the study at any time without any obligation to explain their reasons for doing so (contact information for such requests will be provided).
7. Informed consent is created in different languages and according to the data protection laws that each Citizen Science participant needs to sign.
8. Before any publications (e.g., scientific papers, public deliverables) with the potential of containing personal data is released to the public, it will go through an ethics and privacy screening to ensure that all data included is anonymized, aggregated and/or analyzed in such a way as to ensure that none of the content can be traced back to an individual participant or respondent
9. Data analysis (e.g. in evaluations) will be carried out on de-identified and anonymised data. At the end of the project, all personal data will be deleted, and the de-identified data will be anonymized entirely, meaning that the links to the lists of keys will be deleted.
10. CityCLIM does not collect any sensitive personal data as defined by either Directive 95/46/EC or by Article 19 of the new EU DP Regulation (GDPR Regulation (EU) 2016/679). The only personally identifiable data collected will be kept to a minimum. Furthermore, citizen scientists' names will be de-identified, anonymized, and not included in project reports or deliverables.
11. In CityCLIM we need to distinguish between data collected from people taking measurements, e.g., with a SenseBox, and data from people who are asked to assess subjective perception of heat with the help of an app. In the first case, only minimal data is collected. This includes contact details such as name, address, telephone number and e-mail. This data is only used for contact purposes and will be irreversibly deleted upon completion of the collaboration on the project.
12. The course of sensor data will be processed and displayed following the EU and national regulations. The possible tools include that the GPS data are shown with a location offset or only sections of the temporal-local course.
13. When using the app, users are asked for some important information (e.g., age group, health status, gender group, living environment (city, …) The answers are anonymized, as the answer is given by selecting a certain group e.g. age 12-18, 18-30, 30-45, 45-60, 60-70, from 70. The answers can thus no longer be assigned to individual persons.

8 Overview of Datasets

The list of expected datasets in the following constitutes the first version of dataset descriptions, and we recognise that it will be developed and increased as the project evolves. In addition, some information concerning the datasets remains unknown at this time, e.g., size of the datasets, potential reuse options, QA/QC methods etc. An updated version of this list will be provided based on the fully elaborated DMP at the end of the project.

Table 5. Data generated within CityCLIM.

No.	WP/ Region	Dataset reference and name	Main partners involved
1	2	Modified Urban Landcover	OHB
2	3	UltraHD binary output (1)	MTL
3	3	UltraHD LES images (1)	MTL
4	3	UltraHD LES movies (1)	MTL
5	4	Meteotracker data	UFZ
6	4	Sensebox data	UFZ
7	4	Weatherstation data	UFZ
8	5	ESA Worldcover datasets as global land cover products of last years	OHB
9	5	Multispectral data of Copernicus Sentinel-2 satellite	OHB
10	5	Multispectral and thermal data of USGS Landsat 8 and 9 satellites	OHB
11	5	Digital elevation model provided by city of Karlsruhe	*OHB*
12	6	Heat wave information and Warning service for Thessaloniki	MTL
13	Valencia	Valencia Monitoring Network	City Adminstration Valencia

All datasets described in Table 5 above are described in detail in the following.

(1) The boundary of spatial resolution (latitude and longitude in WGS64) for the data in WP3 are:

- **Karlsruhe_100m_Nest**
  - Upper Left (8.0341289, 49.2411853)
  - Lower Left (8.0341289, 48.7705472)
  - Upper Right (8.7489772, 49.2411853)
  - Lower Right (8.7489772, 48.7705472)
  - Center (8.3915531, 49.0058663)
- **Luxembourg_100m_Nest**
  - Upper Left (5.7381985, 49.8524966)
  - Lower Left (5.7381985, 49.3694293)
  - Upper Right (6.4815653, 49.8524966)
  - Lower Right (6.4815653, 49.3694293)
  - Center (6.1098819, 49.6109629)
- **Valencia_100m_Nest**
  - Upper Left (-0.7331131, 39.6940155)
  - Lower Left (-0.7331131, 39.2296477)
  - Upper Right (-0.1329591, 39.6940155)
  - Lower Right (-0.1329591, 39.2296477)
  - Center (-0.4330361, 39.4618316)

Figure 6. Example 1- WP2

NR	1
WP	2
Data Reference and Name	Modified Urban Landcover
Dataset description	see Argos
Purpose of these data	Used as input data for simulated UHD model runs
Relation to project objectives	necessary for service developments
Data creator /Data owner	OHB-DS
Primary [measured]/Secondary dataset [derived with help of other data]	ESA Worldcover
If primary dataset; which sensor was used	Sentinel-1 and Sentinel-2
Datatyp [datafiles, pictures,...]	geoTiff
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, ...]	tiff
Reuse of existing data; if yes please specify!	yes
Temporal resultion [s, min, h, day,...]	static
Spatial resolution [m, km,...]	10m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	same as for UHD model
Data production method	
Used QA/QC method	
Data storage (where, size)	azure cloud
Describe potential for reuse	
Potential stakeholders for reuse	researchers
Usuage rights available	
Data retention available	
Follow FAIR principles [YES/NO]	no
Curator (Contact person)	Ingo Schoolmann
Exemplary data available? If yes, please specify!	
Dataset included in OpenAire Data Management Plan	yes

Figure 7. Example 1- WP3

NR	2
WP	3
Data Reference and Name	UltraHD binary output
Dataset description	see Argos
Purpose of these data	numerical analysis, e.g. within Identification Services or for plotting of time series within Information Services
Relation to project objectives	necessary for service developments
Data creator /Data owner	MTL
Primary [measured]/Secondary dataset [derived with help of other data]	primary (LES Model)
If primary dataset; which sensor was used	
Datatyp [datafiles, pictures,...]	binary float32
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, ...]	dat
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,...]	5min
Spatial resolution [m, km,...]	100m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	
Data production method	UltraHD
Used QA/QC method	Comparison with measurements, RMS/Bias
Data storage (where, size)	azure cloud
Describe potential for reuse	high, numerical values, compareable
Potential stakeholders for reuse	researchers
Usuage rights available	
Data retention available	
Follow FAIR principles [YES/NO]	no
Curator (Contact person)	Stefan Horn
Exemplary data available? If yes, please specify!	https://zenodo.org/records/13373932
Dataset included in OpenAire Data Management Plan	yes

Figure 8. Example 2- WP3

NR	3
WP	3
Data Reference and Name	UltraHD LES images
Dataset description	see Argos
Purpose of these data	Map overlays for Model Viewer, movie generation
Relation to project objectives	necessary for service developments
Data creator /Data owner	MTL/OHB-DS
Primary [measured]/Secondary dataset [derived with help of other data]	primary (LES Model)
If primary dataset; which sensor was used	
Datatyp [datafiles, pictures,...]	png
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, ...]	png
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,...]	5min
Spatial resolution [m, km,...]	100m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	
Data production method	UltraHD
Used QA/QC method	Comparison with measurements, RMS/Bias
Data storage (where, size)	azure cloud
Describe potential for reuse	low, useful only within Service frontend
Potential stakeholders for reuse	media, citizen
Usuage rights available	
Data retention available	
Follow FAIR principles [YES/NO]	no
Curator (Contact person)	Stefan Horn
Exemplary data available? If yes, please specify!	https://zenodo.org/records/13638890
Dataset included in OpenAire Data Management Plan	yes

Figure 9. Example 3- WP3

NR	4
WP	3
Data Reference and Name	UltraHD LES movies
Dataset description	see Argos
Purpose of these data	Map overlays for Information Services
Relation to project objectives	necessary for service developments
Data creator /Data owner	MTL/OHB-DS
Primary [measured]/Secondary dataset [derived with help of other data]	primary (LES Model)
If primary dataset; which sensor was used	
Datatyp [datafiles, pictures,...]	mp4
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, ...]	mp4
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,...]	5min
Spatial resolution [m, km,...]	100m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	
Data production method	UltraHD
Used QA/QC method	Comparison with measurements, RMS/Bias
Data storage (where, size)	azure cloud
Describe potential for reuse	low, useful only within Service frontend
Potential stakeholders for reuse	media, citizen
Usuage rights available	
Data retention available	
Follow FAIR principles [YES/NO]	no
Curator (Contact person)	Stefan Horn
Exemplary data available? If yes, please specify!	https://zenodo.org/records/13627333
Dataset included in OpenAire Data Management Plan	yes

Figure 10. Example 1- WP4.

NR	5
WP	4
Data Reference and Name	Meteotracker data
Dataset description	The MeteoTracker is developed and marketed by a start-up company as a fully portable, low-cost mini weather station for mobile collection of meteorological and environmental data. The MeteoTracker is user friendly and simple to set up and operate. It can be mounted on a vehicle or bicycle and provides high accuracy temperature and humidity readings. This device measures air temperature air humidity and air pressure and provides GPS and altitude data. The MeteoTracker multi-sensor implements a patented system that enables high-resolution measurement of meteorological parameters. The sensor works at speeds above 7 km/h. It connects to a smartphone via Bluetooth and sends readings to an interactive map via the MeteoTracker app (free in the iOS and Android stores) and the data analysis of the measured data can be done using the MeteoTracker dashboard.
Purpose of these data	Dense spatial and temporal meterological data in a urban area
Relation to project objectives	Citizen Science, Urban heat issues, Heat Island identification service
Data creator /Data owner	UFZ
Primary [measured]/Secondary dataset [derived with help of other data]	Primary
If primary dataset; which sensor was used	Meteotracker
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, …]	csv
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	1 sec
Spatial resolution [m, km,…]	m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	depends on location of measurement
Data production method	Visualization, Interpolation
Used QA/QC method	outliers detection, gap identification, filtering data >7km/h
Data storage (where, size)	Meteotracker app from IoTopon
Describe potential for reuse	Assessing of local variability of meteorological parameters for health and planning issues
Potential stakeholders for reuse	adminstration in the pilot cities, health adminstration
Usuage rights available	Yes
Data retention available	No
Follow FAIR principles [YES/NO]	Yes
Curator (Contact person)	Uta Koedel
Exemplary data available? If yes, please specify!	YES, Zenodo (https://doi.org/10.5281/zenodo.10598570)
Dataset included in OpenAire Data Management Plan	Yes

Figure 11. Example 2- WP4.

NR	6
WP	4
Data Reference and Name	Sensebox data
Dataset description	The SenseBox is a toolkit for environmental monitoring especially designed for the flexible deployment in citizen science projects. Since its primary development focus is on educational purposes, the emphasis lies on the educational benefits derived from personally assembling and programming the measurement device. Due to the modular concept, various sensors are available to monitor physical and chemical properties of the atmosphere and the soil. The SenseBox system is based on an easy to program Arduino microcontroller unit and can be connected via WiFi or LoRaWAN with a web-based dashboard platform providing the data online in real time. A large active amateur community is constantly developing the platform and supporting the users. The Sensebox system we deployed consists of the multi-parameter sensor BOSCH BME680 for air temperature, humidity and air pressure measurements, the Sensirion SCD30 CO_2 concentration sensor, a GPS sensor by Watterott electronic for time and location measurements, and a Nova Fitness SDS011 dust sensor measuring particle concentration between 0.3 to 10µm in the air.
Purpose of these data	Dense spatial and temporal meterological data in a urban area
Relation to project objectives	Citizen Science, Urban heat issues, Heat Island identification service
Data creator /Data owner	UFZ
Primary [measured]/Secondary dataset [derived with help of other data]	Primary
If primary dataset; which sensor was used	Sensebox
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, …]	csv
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	10 sec
Spatial resolution [m, km,…]	m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	depends on location of measurement
Data production method	Visualization, Interpolation
Used QA/QC method	outliers detection, gap identification
Data storage (where, size)	Dashboard
Describe potential for reuse	Assessing of local variability of meteorological parameters for health and planning issues
Potential stakeholders for reuse	adminstration in the pilot cities, health adminstration
Usuage rights available	YES
Data retention available	No
Follow FAIR principles [YES/NO]	YES
Curator (Contact person)	Uta Koedel
Exemplary data available? If yes, please specify!	YES, Zenodo (https://doi.org/10.5281/zenodo.10598272)
Dataset included in OpenAire Data Management Plan	Yes

Figure 12. Example 3- WP4.

NR	7
WP	4
Data Reference and Name	Weatherstation data
Dataset description	The WIFI Colour Weather Station is a convenient weather station with sensors to measure wind speed, wind direction, humidity, temperature, rainfall, UV levels and light intensity. The WIFI feature allows users to share their data on weather platforms such as AWEKAS, Weather Underground or Weather Cloud. The sensor transmits 5 different weather readings to the base station's intuitive display over a frequency of 868 MHz at a range of up to 150 metres.
Purpose of these data	Dense spatial and temporal meterological data in a urban area
Relation to project objectives	Citizen Science, Urban heat issues, Heat Island identification service
Data creator /Data owner	UFZ
Primary [measured]/Secondary dataset [derived with help of other data]	Primary
If primary dataset; which sensor was used	Bresser WIFI Weatherstation
Datatyp [datafiles, pictures,...]	datafile
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, ...]	csv
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,...]	1 sec
Spatial resolution [m, km,...]	m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	depends on location of measurement
Data production method	Visualization, Interpolation
Used QA/QC method	outliers detection, gap identification
Data storage (where, size)	Dashboard
Describe potential for reuse	Assessing of local variability of meteorological parameters for health and planning issues
Potential stakeholders for reuse	adminstration in the pilot cities, health adminstration
Usuage rights available	YES
Data retention available	No
Follow FAIR principles [YES/NO]	YES
Curator (Contact person)	Uta Koedel
Exemplary data available? If yes, please specify!	YES, Zenodo (https://doi.org/10.5281/zenodo.12742931)
Dataset included in OpenAire Data Management Plan	Yes

Figure 13. Example 1- WP5

NR	8
WP	5
Data Reference and Name	Worldcover
Dataset description	ESA Worldcover datasets as global land cover products of last years
Purpose of these data	Reference for current land cover
Relation to project objectives	EO-HISMSS engine and service
Data creator /Data owner	OHB SYS
Primary [measured]/Secondary dataset [derived with help of other data]	secondary
If primary dataset; which sensor was used	
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	no
File Format [csv, jpg, json, tiff, …]	tiff
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	1 year
Spatial resolution [m, km,…]	10m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	bounding boxes of pilot cities
Data production method	raw data used
Used QA/QC method	none
Data storage (where, size)	GCCP data warehouse
Describe potential for reuse	Indication of current urban land cover
Potential stakeholders for reuse	city adminitstrations, planning departments
Usuage rights available	yes
Data retention available	no
Follow FAIR principles [YES/NO]	Yes
Curator (Contact person)	Anita Bayer
Exemplary data available? If yes, please specify!	https://viewer.esa-worldcover.org/worldcover/
Dataset included in OpenAire Data Management Plan	Yes

Figure 14. Example 2- WP5

NR	9
WP	5
Data Reference and Name	Sentinel-2 data
Dataset description	Multispectral data of Copernicus Sentinel-2 satellite
Purpose of these data	Calculation of vegetation and built-up indices
Relation to project objectives	EO-HISMSS engine and service
Data creator /Data owner	OHB SYS
Primary [measured]/Secondary dataset [derived with help of other data]	primary
If primary dataset; which sensor was used	Sentinel-2
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	yes
File Format [csv, jpg, json, tiff, …]	tiff
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	irregular
Spatial resolution [m, km,…]	10/20m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	bounding boxes of pilot cities
Data production method	index calculation
Used QA/QC method	none
Data storage (where, size)	GCCP data warehouse
Describe potential for reuse	optical satellite data, basis for calculation of vegetation and built-up indices
Potential stakeholders for reuse	city administrations
Usuage rights available	yes
Data retention available	no
Follow FAIR principles [YES/NO]	yes
Curator (Contact person)	Anita Bayer
Exemplary data available? If yes, please specify!	https://browser.dataspace.copernicus.eu
Dataset included in OpenAire Data Management Plan	Yes

Figure 15. Example 3- WP5

NR	10
WP	5
Data Reference and Name	Landsat 8/9 data
Dataset description	Multispectral and thermal data of USGS Landsat 8 and 9 satellites
Purpose of these data	provision of LST
Relation to project objectives	EO-HISMSS engine and service
Data creator /Data owner	OHB SYS
Primary [measured]/Secondary dataset [derived with help of other data]	primary
If primary dataset; which sensor was used	Landsat 8/9
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	yes
File Format [csv, jpg, json, tiff, …]	tiff
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	irregular
Spatial resolution [m, km,…]	30m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	bounding boxes of pilot cities
Data production method	raw data used
Used QA/QC method	none
Data storage (where, size)	GCCP data warehouse
Describe potential for reuse	LST shows heat in cities
Potential stakeholders for reuse	city administrations, climate departments
Usuage rights available	yes
Data retention available	no
Follow FAIR principles [YES/NO]	yes
Curator (Contact person)	Anita Bayer
Exemplary data available? If yes, please specify!	https://earthexplorer.usgs.gov/
Dataset included in OpenAire Data Management Plan	Yes

Figure 16. Example 4- WP5

NR	11
WP	5
Data Reference and Name	DEM Karlsruhe
Dataset description	Digital elevation model provided by city of Karlsruhe
Purpose of these data	derivation of 3D building geometry indicators
Relation to project objectives	EO-HISMSS engine and service
Data creator /Data owner	OHB SYS
Primary [measured]/Secondary dataset [derived with help of other data]	primary
If primary dataset; which sensor was used	Lidar data
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	yes
File Format [csv, jpg, json, tiff, …]	GeoTIFF, IMG, Gridfloat (. flt)
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	year of flight
Spatial resolution [m, km,…]	1m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	Karlsruhe city area
Data production method	raw data used
Used QA/QC method	none
Data storage (where, size)	GCCP data warehouse
Describe potential for reuse	ground and building height
Potential stakeholders for reuse	city adminitstrations, planning departments
Usuage rights available	No
Data retention available	no
Follow FAIR principles [YES/NO]	No
Curator (Contact person)	Anita Bayer
Exemplary data available? If yes, please specify!	No
Dataset included in OpenAire Data Management Plan	Yes

Figure 17. Example 5- WP5

NR	12
WP	5
Data Reference and Name	DEM Valencia
Dataset description	Digital elevation model provided by city of Valencia
Purpose of these data	derivation of 3D building geometry indicators
Relation to project objectives	EO-HISMSS engine and service
Data creator /Data owner	OHB SYS
Primary [measured]/Secondary dataset [derived with help of other data]	primary
If primary dataset; which sensor was used	Lidar data
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	yes
File Format [csv, jpg, json, tiff, …]	GeoTIFF, IMG, Gridfloat (. flt)
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	year of flight
Spatial resolution [m, km,…]	5m
Boundary of spatial resolution(Latitude and Longitude in WGS64)	Valencia vicinity
Data production method	raw data used
Used QA/QC method	none
Data storage (where, size)	GCCP data warehouse
Describe potential for reuse	ground and building height
Potential stakeholders for reuse	city adminitstrations, planning departments
Usuage rights available	No
Data retention available	no
Follow FAIR principles [YES/NO]	No
Curator (Contact person)	Anita Bayer
Exemplary data available? If yes, please specify!	No
Dataset included in OpenAire Data Management Plan	Yes

Figure 18. Example 1- WP6

NR	13
WP	6
Data Reference and Name	marcel1
Dataset description	Heat wave information and Warning service for Thessaloniki
Purpose of these data	This description describe the dataset used for the Heat wave information and Warning service for Thessaloniki city region. The dataset is generated and processed at Meteologix AG, transferred to the GCCP and used to fulfill the service.
Relation to project objectives	Develop a service for a Pilot city during the project. For the City of Central Mazedonia (Thessaloniki) a Heat wave Information and Warning Service is developed based on MOS model output. Enhanced forecast of 2m Air Temperatur and Humidity is used to get Heat Index forcast (maps and warning).
Data creator /Data owner	König, Marcel; Zimmer, Janek
Primary [measured]/Secondary dataset [derived with help of other data]	Primary simulation data
If primary dataset; which sensor was used	-
Datatyp [datafiles, pictures,…]	datafile
Use standardized vocabularies	-
File Format [csv, jpg, json, tiff, …]	.dat (binary file)
Reuse of existing data; if yes please specify!	no
Temporal resultion [s, min, h, day,…]	hourly
Spatial resolution [m, km,…]	4x4km
Boundary of spatial resolution(Latitude and Longitude in WGS64)	Lat: 40.35 - 40.878 22 Pixel; Lon: lon: 22.4 - 23.458 44 Pixel
Data production method	Numerical simulation + merging statistical Analyses
Used QA/QC method	-
Data storage (where, size)	GCCP, 600KB per run (4 runs a day)
Describe potential for reuse	-
Potential stakeholders for reuse	-
Usuage rights available	-
Data retention available	
Follow FAIR principles [YES/NO]	
Curator (Contact person)	König, Marcel
Exemplary data available? If yes, please specify!	https://doi.org/10.5281/zenodo.13378844
Dataset included in OpenAire Data Management Plan	Yes

Figure 19. Example 1- Valencia

NR	13
WP	Valencia
Data Reference and Name	VALENCIA MONITORING NETWORK
Dataset description	Data from the Air Pollution Monitoring Network, Weather Monitoring Network and Noise Level Monitoring Network
Purpose of these data	Monitoring enviromental variables
Relation to project objectives	Climate data related
Data creator /Data owner	Valencia city hall
Primary [measured]/Secondary dataset [derived with help of other data]	Primary
If primary dataset; which sensor was used	Valencia sensor monitoring
Datatyp [datafiles, pictures,…]	Text
Use standardized vocabularies	No
File Format [csv, jpg, json, tiff, …]	JSON
Reuse of existing data; if yes please specify!	No
Temporal resultion [s, min, h, day,…]	min
Spatial resolution [m, km,…]	none
Boundary of spatial resolution(Latitude and Longitude in WGS64)	none
Data production method	Through sensors and stations deployed at the city
Used QA/QC method	No
Data storage (where, size)	Valencia Smartcity Platform
Describe potential for reuse	Enviromental Research Projects
Potential stakeholders for reuse	Citizens, researchers
Usuage rights available	
Data retention available	1 month
Follow FAIR principles [YES/NO]	NO
Curator (Contact person)	Ana Viciano
Exemplary data available? If yes, please specify!	https://valencia.opendatasoft.com/pages/home/
Dataset included in OpenAire Data Management Plan	Yes

9 References

All European Academies (ALLEA) (2017). The European Code of Conduct for Research Integrity, Available at https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf [accessed March 21, 2022]

DLA Piper Intelligence (2022). Data Protection Laws of the World, available at https://www.dlapiperdataprotection.com/index.html?t=law&c=GR&c2= [accessed February 24, 2022]

DSGVO(2022) GDPR Chapter 2 - Art. 6, Lawfulness of processing, available at https://www.dsgvo-portal.de/gdpr_article_6.php [accessed March 1, 2022]

EC (2005). The European Charter for Researchers, The Code of Conduct for the Recruitment of Researchers, Available at https://euraxess.ec.europa.eu/sites/default/files/am509774cee_en_e4.pdf [accessed March 21, 2022]

EC (2017). H2020 Programme: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Available at https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf [accessed March 8, 2022]

EC (2016). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Version 3.0, July 26 2016, available at https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

EC (2018). TURNING FAIR INTO REALITY, Final Report and Action Plan from the European Commission Expert Group on FAIR Data, doi:10.2777/54599, available at https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf [accessed March 8, 2022]

EC (2022). What are my rights? available at https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights/what-are-my-rights_en [accessed February 24, 2022]

GDPR (2022). Art. 5 GDPR Principles relating to processing of personal data, available at https://gdpr-info.eu/art-5-gdpr/ [accessed March 1, 2022]

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds & Machines 30, 99–120, doi: 10.1007/s11023-020-09517-8

OpenAire (2022). Guidelines for research data. Available at https://www.openaire.eu/h2020-mandate-for-data/[accessed March 8, 2022]

Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC

EC INSPIRE (2013). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 at https://inspire.ec.europa.eu/documents/inspire-metadata-implementing-rules-technical-guidelines-based-en-iso-19115-and-en-iso-1


cityCLIM