D7.9 Data Management Plan v2

WP7 – Project Management

Deliverable Lead: CIS

Dissemination Level: Public

Deliverable due date: 31/01/2019

Actual submission date: 21/02/2019

Version 1.0

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730355

Document Control Page	
Title	D7.8 Data Management Plan v2
Creator	Pascal Dihé (CIS)
Description	The Data Management Plan describes what data the CLARITY project will generate, whether and how it will be exploited or made accessible for verification and re-use and how it will be curated and preserved after the end of the project.
Publisher	CLARITY Consortium
Contributors	Pascal Dihé (CIS), Jorge Humberto Amorim (SMHI), Lena Strömbäck (SMHI), Gerald Schimak (AIT), Romana Stollnberger (AIT), Wolfgang Loibl (AIT), Daniela de Gregorio (PLINIVS), Alessandra Capolupo (PLINIVS), Iñaki Torres Cobián (ATOS), Robert Goler (ZAMG), Astrid Kainz (ZAMG), Claudia Hahn (ZAMG)
Creation date	27/10/2018
Type	Text
Language	en-GB
Rights	copyright “CLARITY Consortium”
Audience	☒ Public ☐ Confidential ☐ Classified
Status	☐ In Progress ☐ For Review ☐ For Approval ☒ Approved

Disclaimer

Disclaimer

Table of Contents

List of figures

CLARITY Project Overview

Urban areas and traffic infrastructures that are linking such areas are highly vulnerable to climate change. Smart use of existing climate intelligence can increase urban resilience and generate benefits for businesses and society at large. Based on the results of FP7 climate change, future internet and crisis preparedness projects (SUDPLAN, ENVIROFI, CRISMA) with an average TRL of 4-5 and following an agile and user-centred design process, end-users, purveyors and providers of climate intelligence will co-create an integrated climate services Information System (CSIS) to integrate resilience into urban infrastructure.

As a result, CLARITY will provide an operational eco-system of cloud-based climate services to calculate and present the expected effects of CC-induced and -amplified hazards at the level of risk, vulnerability and impact functions. CLARITY will offer what-if decision support functions to investigate the effects of adaptation measures and risk reduction options in the specific project context and allow the comparison of alternative strategies. Four Demonstration Cases will showcase CLARITY climate services in different climatic, regional, infrastructure and hazard contexts in Italy, Sweden, Austria and Spain; focusing on the planning and implementation of urban infrastructure development projects.

CLARITY will provide the practical means to include the effects of CC hazards and possible adaptation and risk management strategies into planning and implementation of such projects, focusing on increasing CC resilience. Decision makers involved in these projects will be empowered to perform climate proof and adaptive planning of adaptation and risk reduction options.

Executive Summary

This report is the second deliverable of Task 7.3 “Data Management” and describes the updated Data Management Plan (DMP) for the CLARITY project, funded by the EU’s Horizon 2020 Programme under Grant Agreement number 730355. The purpose of the DMP is to provide an overview of all datasets collected and generated by the project and to define the CLARITY consortium’s data management policy that is used with regard to these datasets.

The first CLARITY DMP (deliverable D7.8 [1]) followed the structure of the Horizon 2020 DMP template [2] and reported on the datasets used and produced by the project in a dedicated annex. This initial version defined also the general policy and approach to data management in CLARITY that handles data management related issues on the administrative and technical level. This included for example topics like data and meta-data collection, publication and deposition of open data, the data repository infrastructure and compliance to the Open Access Infrastructure for Research in Europe (OpenAIRE).

The second CLARITY DMP (this deliverable) is implemented as a “living” DMP based on a dedicated CKAN catalogue (https://ckan.myclimateservice.eu/) that is continuously updated throughout the course of the project. This online meta-data catalogue reflects the status of the data that is collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

This report summarises the results of the data production activities in the project that are being carried out according to the data collection concept introduced in Task 2.2 “Data requirements definition, data collection concept, demonstration and result validation concept” and the guidelines on FAIR (Findable, Accessible, Interoperable and Reusable) data management [3] and that are described in detail in the CLARITY CKAN catalogue.

1 Introduction

The introduction chapter defines the purpose and scope of the CLARITY DMP as well as its relation to other deliverables and briefly explains the structure of the document.

1.1 Purpose of this document

This document is an accompanying report to CLARITY's "living" Data Management Plan (Figure 1) that is organisationally and technically represented by the CLARITY CKAN meta-data online catalogue at https://ckan.myclimateservice.eu/.

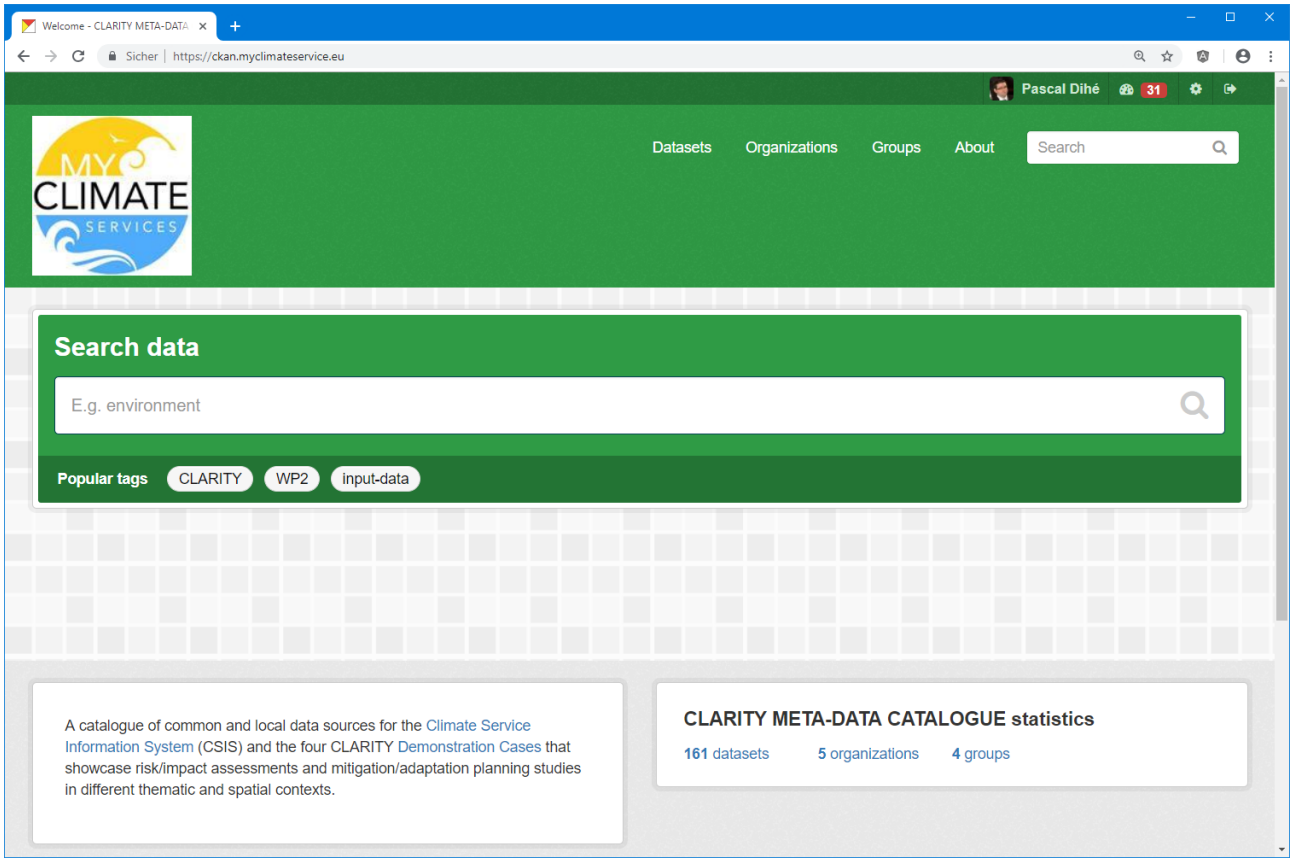

Figure 1: CLARITY CKAN meta-data catalogue

While the online catalogue provides detailed information (meta-information) on datasets both used and produced by the CLARITY project including links for accessing those data, the document at hand focuses on briefly summarising what (open) data has been produced by the project so far and what data is expected to be produced and released under an open license in the near future. The online catalogue will continuously be updated as soon as new data can be made available. A final of report (deliverable D7.10) summarising the overall (open) data-related outcome of the project will then be released in M36.

1.2 Relation to other deliverables

The fist CLARITY DMP (deliverable D7.8 [1]) defined CLARITY's general data management policy that has been developed in accordance to Horizon 2020 FAIR principles [2], open data requirements [4] and implementation guidelines [5]. It applies mainly to new results that are produced in CLARITY and that are to be made available by the project consortium as open source, open science and open data. Among others, it defines how data produced within the CLARITY project can be made findable and openly accessible.

It thereby relies on state of the art technical solutions and standards like Digital Object Identifiers, DataCite metadata, the OpenAIRE initiative and the Zenodo research data repository for the implementation of these procedures and their seamless technical integration into CLARITY's climate services Information System (https://csis.myclimateservice.eu/). The policy is a binding document for CLARITY data managers (e.g. WP leaders) and part of deliverable D7.8 [1].

Both the CLARITY DMP as well the WP2 "Demonstration" catalogue of CLARITY data and metadata share the same technical infrastructure. Deliverable D2.2 "Data Collection Report" [4] summarises the methodology set in place to facilitate the collection of information from the demonstration cases that are being carried out within the framework of the CLARITY project. It includes both the procedures defined and the support tools that allow for keeping a record of what data are used and facilitate their publication. While the D2.2 part of the catalogue (https://ckan.myclimateservice.eu/dataset?tags=WP2) focuses mainly on (local) data collected and produced for the preparation of the four CLARITY Demonstration Cases, the overall CLARITY DMP concentrates more on (open) data produced by the project and how this data can be preserved in the future. For this purpose, the CKAN catalogue has been adapted and configured to support four different groups (Figure 2) that can be used to categorise the datasets available in CKAN according to their role in the CLARITY DMP. The most interesting groups are thereby "**Non-open data *produced by CLARITY*"** (https://ckan.myclimateservice.eu/group/non-open data-produced-by-clarity) and "**Open data *produced* by CLARITY"** (https://ckan.myclimateservice.eu/group/open data-produced-by-clarity).

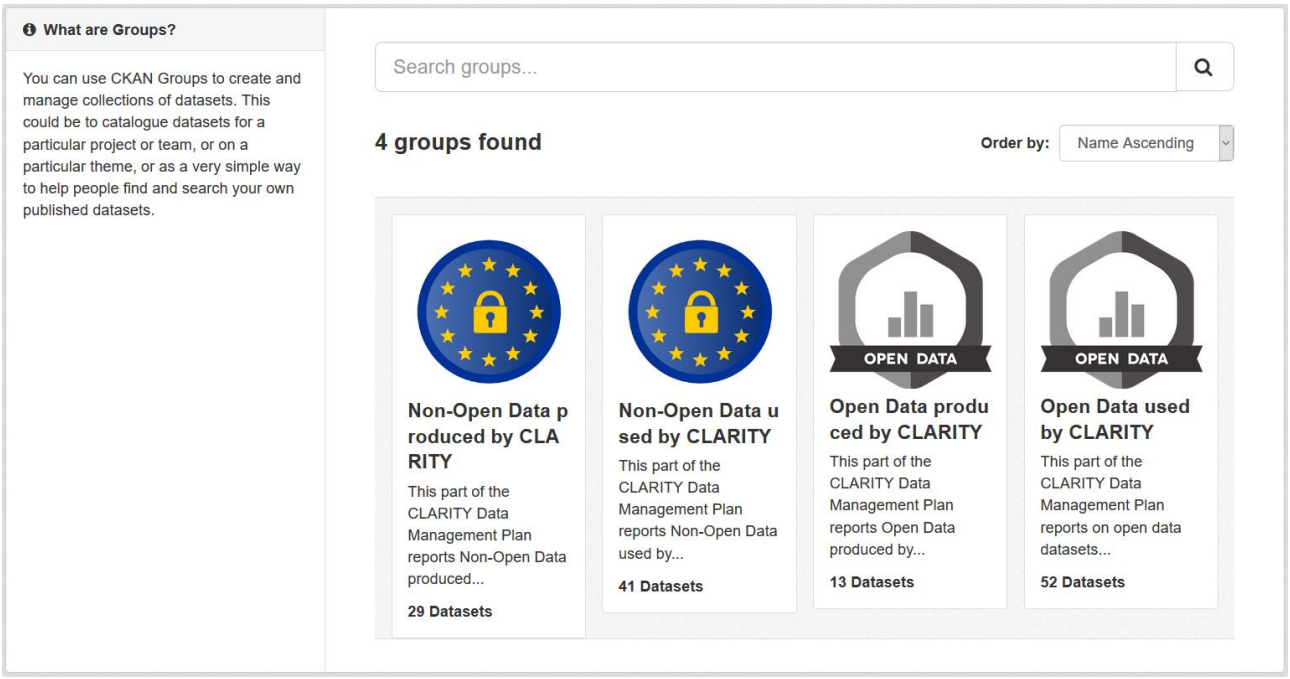

Figure 2: CKAN groups for CLARITY data management

The technical infrastructure in form of a containerised CKAN instance and the related containerised database and search engine services are provided and maintained by WP4 "Technology Support" in CLARITY's containerised service infrastructure (Figure 3) and the respective open source repositories at GitHub (https://github.com/clarity-h2020/ckan). The implementation and deployment details are reported in chapter 4.2 of deliverable D4.3 "Technology Support Report" [5]. In particular, the related building block "Catalogue of Data Sources and Simulation Models" is defined as "a meta-data catalogue that makes climate-related information accessible by providing functionalities to streamline publishing, sharing, finding and using data and models. The catalogue can be used for data discovery and meta-data storage by different climate services and building blocks, respectively." [6]

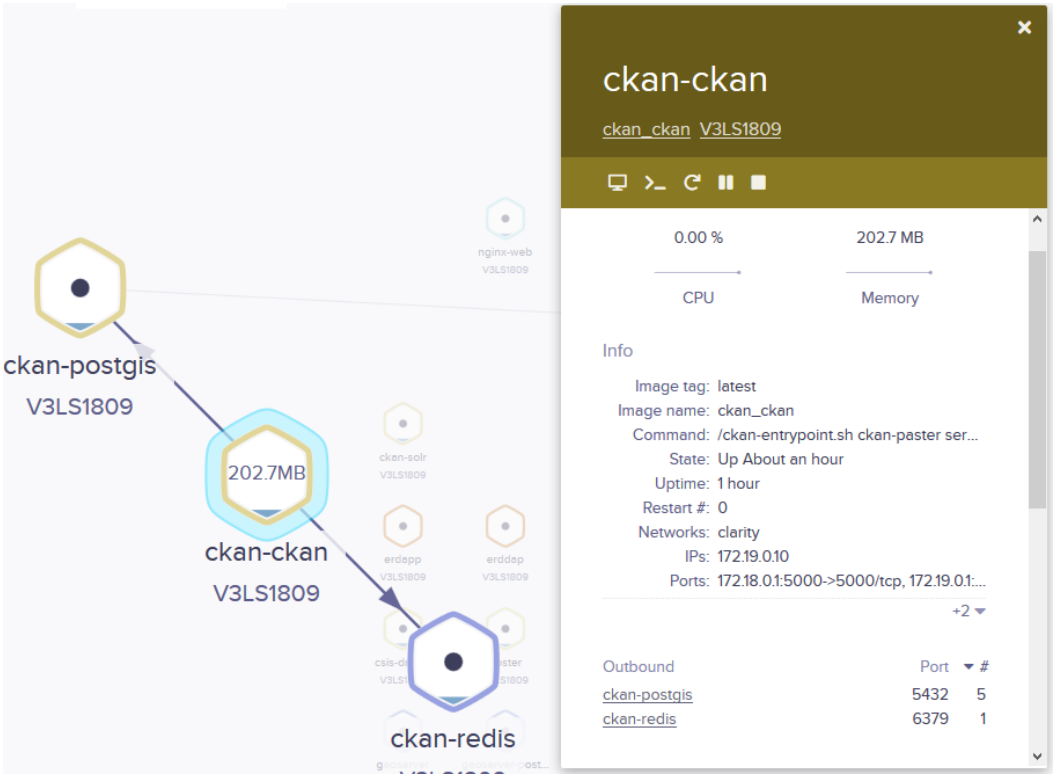

Figure 3: Containerised CKAN services

Task 7.3 "Data Management" follows CLARITY's agile project management approach and thus uses the overall CSIS product backlog (https://github.com/clarity-h2020/csis/issues) that is linked with task-specific project boards (Figure 4) representing a dynamic work plan for T7.3 (https://github.com/orgs/clarity-h2020/projects/14).

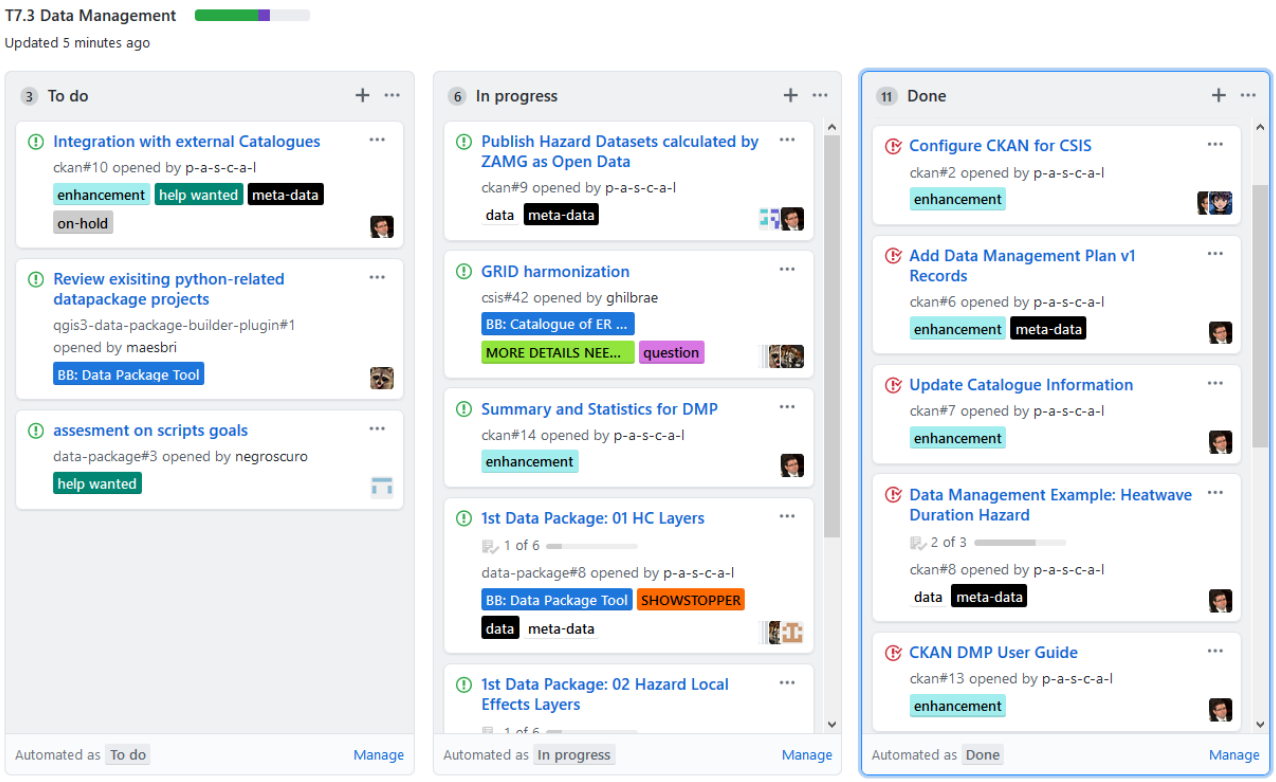

Figure 4: T7.3 project board (excerpt)

1.3 Intended audience

The target readers of this document are besides the members of the CLARITY consortium involved in the production and preservation of (open) data are all interested third parties that want to learn about the results of the CLARITY project in terms of (open) data.

1.4 Structure of the document

The structure of the document and the relationships between the different chapters is as follows:

Chapter 1 (this chapter) introduces the document and explains the overall purpose of this document and its relation to other work packages and deliverables.

Chapter 2 lists the documents that were used or referenced in the development of this report and provides a document-specific list of abbreviations.

Chapter 3 briefly summarises the procedures and responsibilities related to data management activities in CLARITY and explains the usage of the CKAN meta-data catalogue for the CLARITY DMP.

Chapter 4 summarises the results of the ongoing CLARITY data production activities that are currently reported in CLARITY's "living" DMP and thus provides a brief overview on the data that is generated following the methodology and standards initially set out in CLARITY's data management policy in deliverable D7.8 [1].

Chapter 5 provides the conclusions and a summary on follow-up activities.

2 Reference documents

The following documents were used or referenced in the development of this report:

- D2.2 “Data Collection Report”
- D4.1 “Technology Support Plan”
- D4.3 “Technology Support Report”
- D7.8 “Data Management Plan”
- CLARITY Grant Agreement and Description of the Actions
- CLARITY Consortium Agreement

2.1 Abbreviations and Glossary

A common glossary of terms for all CLARITY deliverables, as well as a list of abbreviations, can be found in the public document “CLARITY Glossary” available at http://cat.clarity-h2020.eu/glossary/main.

Abbreviation/Acronym	Definition
BB	Building Block
CC	Climate Change
CCA	Climate Change Adaptation
CCCA	Climate Change Centre Austria
CKAN	Comprehensive Kerbal Archive Network
CLARITY	Integrated Climate Adaptation Service Tools for Improving Resilience Measure
CS	Climate Service
CSIS	CLARITY Climate Services Information System
DC	Demonstration Case
DC	Dublin Core
DMP	Data Management Plan
DoA	Description of Action (Annex 1 to the Grant Agreement)
DOI	Digital Object Identifier
DRR	Disaster Risk Reduction
EC	European Commission
EU-GL	Non-paper Guidelines for Project Managers: Making vulnerable investments climate resilient (Document)
FAIR	Findable, Accessible, Interoperable and Reusable
GCM	Global Climate Model
GDAL	Geospatial Data Abstraction Library
GeoTIFF	Geographic Tagged Image File Format
GIS	Geospatial Information System
IPCC	Intergovernmental Panel on Climate Change
IPR	Intellectual Property Rights

MRT	Mean Radiant Temperature
NetCDF	Network Common Data Form
OGC	Open Geospatial Consortium
OSM	Open Street Map
PET	Potential Evaporation
PMV	Predicted Mean Vote
R&I	Research & Innovation
RCM	Regional Climate Model
RCP	Representative Concentration Pathway
WFS	Web Feature Service
WMS	Web Map Service
WP	Work Package

3 Data Management in CLARITY

3.1 About Data Management in CLARITY

CLARITY as H2020 project is obliged to provide a continuously updated Data Management Plan (DMP) that describes what data the project will use and produce, whether and how data produced will be exploited or made (openly) accessible for verification and re-use and how these data will be curated and preserved after the end of the project. Open data must be put into a public (research data) repository, e.g. Zenodo (Figure 5). The repository should be OpenAIRE-compliant to enable harvesting of metadata.

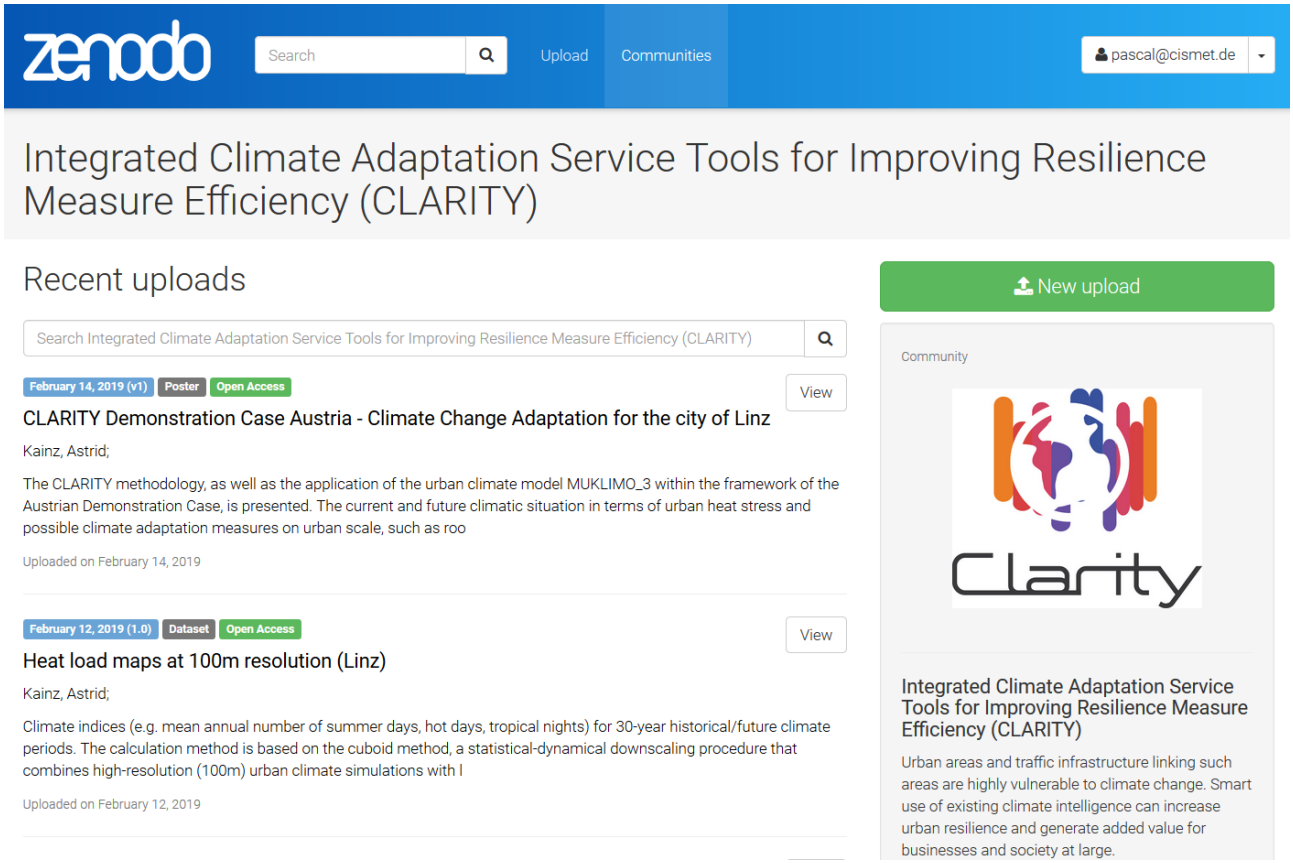

Figure 5: CLARITY community at Zenodo

A project-wide data management policy that handles these issues on administrative and technical level has been established and described in CLARITY deliverable D7.8 Data Management Plan v1 [1]. This includes for example topics like data and meta-data collection, personal data treatment, the data repository infrastructure and the mandatory compliance to OpenAIRE - the Open Access Infrastructure for Research in Europe (Figure 6).

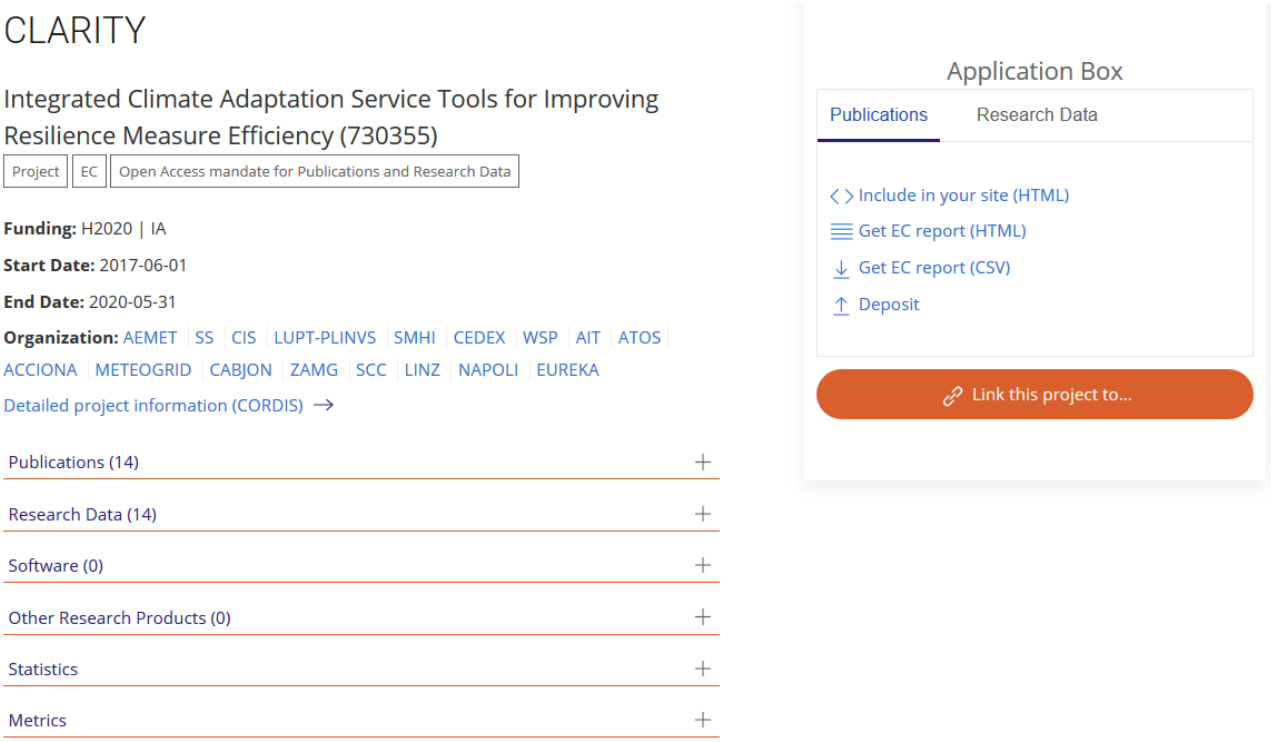

Figure 6: CLARITY Project in OpenAIRE

3.2 Roles and Responsibilities

Data management activities concern the whole project and need to be coordinated and monitored both at project and work package level. Data management is also linked to publication of project results and thus dissemination activities. Therefore, the following roles and responsibilities can be identified:

The **Project Data Manager (T7.3 task leader)** is responsible for

- developing the data management plan and policy in cooperation with the project management in WP7 and the technical partners
- coordinating the technical realisation of CLARITY's living DMP (https://ckan.myclimateservice.eu/) by means of a customised CKAN meta-data catalogue
- maintaining the technical and organisational infrastructure of the CKAN metadata catalogue (https://github.com/clarity-h2020/ckan)
- developing a user guide for the usage of CLARITY's living DMP
- monitoring data management activities (both collection and publication) and deadlines and sending reminders to WP data managers
- providing support to WP data managers
- coordinating the writing of the DMP deliverable documents (D7.x)
- providing solutions for specific issues in accordance with project management

The **Workpackage Data Managers** are responsible for

- the implementation of the data management policy in their respective WPs
- monitoring data management activities and deadlines and sending reminders to partners
- offering customized help and further guidance for using CLARITY's living DMP
- asking partners for missing information or clarifications
- providing input to the DMP deliverable documents (D7.x) by analysing and summarising the WP-specific datasets listed in CLARITY's living DMP
- offering customized help and further guidance for publishing open data and open source software

- monitoring that open results (data and software) are deposited in the default repository or a complementary OpenAIRE-compliant repository and sending reminders to partners
- monitoring that open results available in OpenAIRE are properly linked (https://www.openaire.eu/participate/claim)with CLARITY
- contacting the quality assurance and ethics committee in case of questions and ethical and privacy issues that may forbid a publication of the data
- ensuring that the meta-data of data used and produced at workpackage-level is made available in CLARITY's living DMP according to the CLARITY data management policy and guidelines in a timely manner.

The **Dissemination Manager** is responsible for

- offering assistance in choosing the right publication path (green or gold open access)
- offering customized help and further guidance for publishing scientific publications
- ensuring that the open access policy of the journal complies with the H2020 open data requirements (https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm) before the researcher submits a manuscript
- monitoring that green access (self-archiving) publications are deposited in repositories and sending reminders to partners
- monitoring that metadata about publications is made available in the R&I Participant Portal (preferably automatically through OpenAIRE) and on the CLARITY website (http://clarity-h2020.eu/)
- monitoring that research data related to a publication is made available in repositories and linked to respective publication
- monitoring possible embargo periods and sending reminders to partners
- monitoring that publications available in OpenAIRE are properly linked with CLARITY

3.3 The CLARITY Data Management Plan

The CLARITY Data Management Plan (DMP) follows the structure of the Horizon 2020 DMP template (http://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf). It reflects the status of the data that is collected, processed or generated by the project and, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

The initial Data Management Plan (DMP) was the first deliverable of Task 7.3 "Data Management" and as public document, it has been made available on Zenodo at https://zenodo.org/record/1491532.

For the second version of CLARITY's Data Management Plan, the consortium has decided to implement the DMP as a "living document" by means of CLARITY's CKAN Catalogue (https://ckan.myclimateservice.eu/). The actual deliverable documents (D7.9 and D7.10) are brief reports that summarise the contents of the CKAN catalogue regarding data produced by CLARITY.

3.4 CLARITY's CKAN Catalogue

CLARITY's CKAN Catalogue is available at https://ckan.myclimateservice.eu. It represents both the CLARITY deliverables D2.2 "Catalogue of local data sources and sample datasets" [4] and D7.x Data Management Plan v1 - v3. Thereby the datasets added to CKAN catalogue in the course of the preparation of D2.2 are a subset of the datasets that are added in the course of the preparation of the D7.x. deliverables.

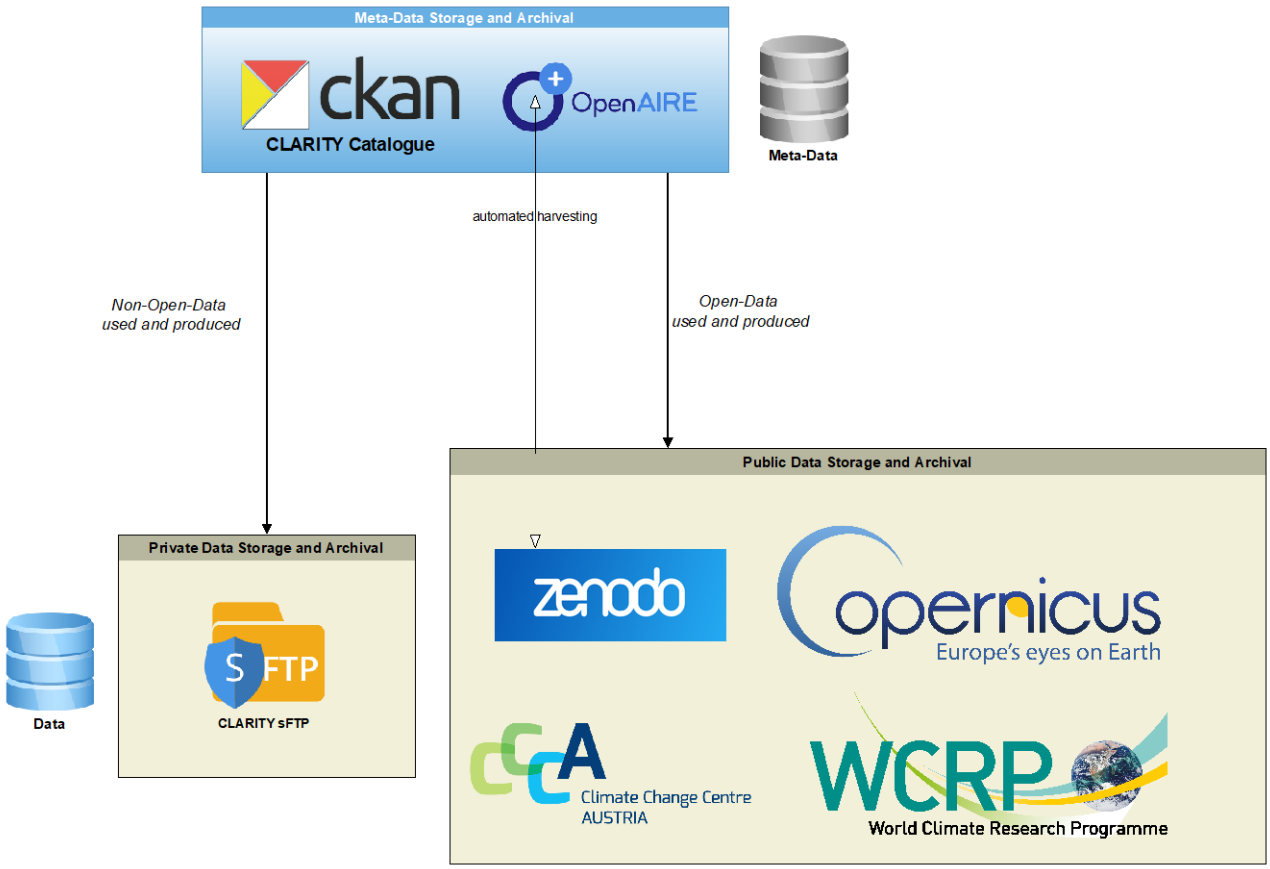

Figure 7: Role of the CKAN catalogue in CLARITY data management activities

The overall role of the CKAN catalogue in CLARITY data management activities (Figure 7) is that of a meta-data catalogue that serves as entry point to the data that is stored in many separate data repositories.

Especially for open data that is reused within the CLARITY project there is no need to replicate all datasets and store it on the (access restricted) CLARITY sFTP server (Figure 12). Instead, the meta-data entry ("Dataset[1]") in CLARITY CKAN links to the public storage location of the actual data ("Resource[2]) that is offered by the organisation (e.g. Copernicus) that initially created the data.

Unlike non-open data produced by CLARITY that is securely stored on the CLARITY sFTP server, open data produced by the project is stored in a public research data repository (e.g. Zenodo) and the respective meta-data is additionally replicated in OpenAIRE to support findability and long-term preservation of project results.

[1] „A dataset contains individual resources as well as metadata. Metadata is the "Who, What, When, Where, Why" of each dataset." (https://dkan.readthedocs.io/en/latest/introduction/catalog-basics.html)

[2] „Resources are the actual files, APIs or links that are being shared." (https://dkan.readthedocs.io/en/latest/introduction/catalog-basics.html)

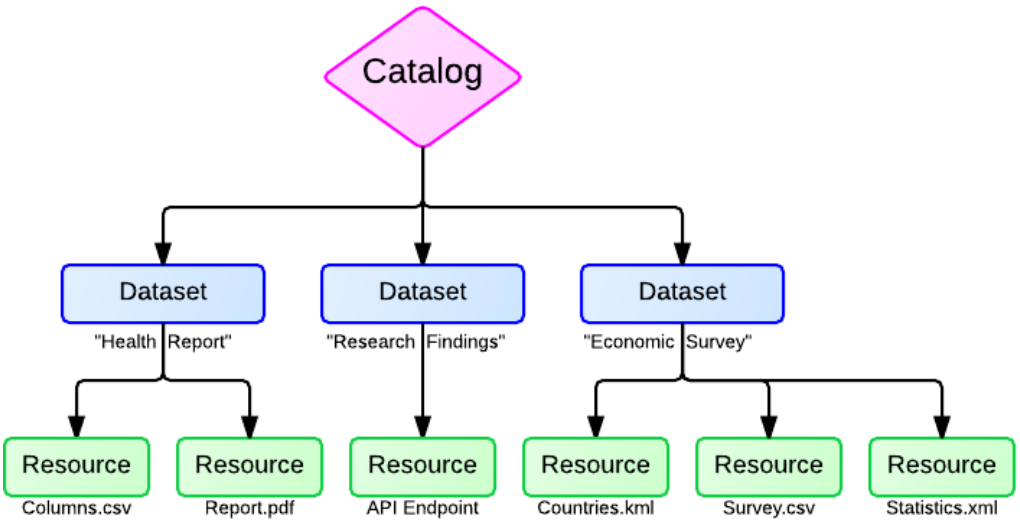

Figure 8: Datasets and resources within CKAN (https://dkan.readthedocs.io)

Figure 8 and Figure 9 display how datasets and resources are structured within CKAN.

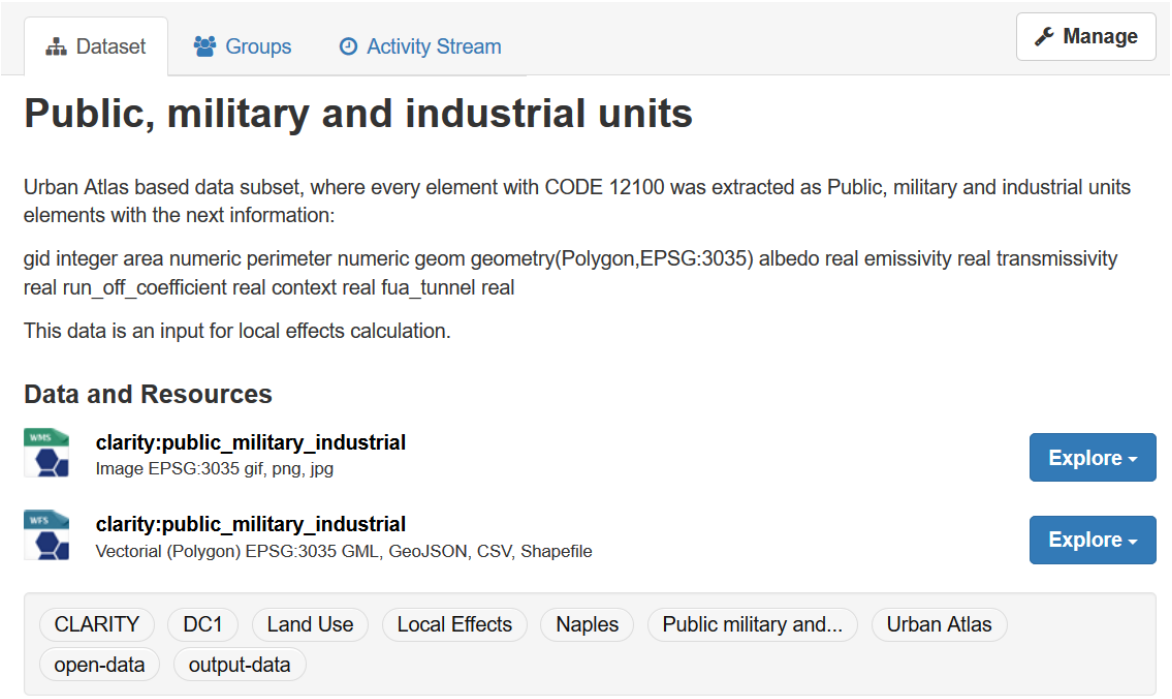

Figure 9: Example Dataset with two Resources in CLARITY's CKAN catalogue

Besides the instructions for transferring meta-data collected by Task T2.2 "Demonstrator-specific data collection" to CKAN (https://github.com/clarity-h2020/csis/wiki/T2.2-Data-Collection-Spreadsheet-to-CKAN-HOW-TO), there are a few more requirements to be considered when adding datasets to CLARITY's CKAN catalogue.

First, depending on whether the data is used or produced, or open data or non-open data and the WP or DC the dataset is relevant for, the following *tags* (Figure 10) must be assigned to the dataset:

- *WP1, WP2, WP3, WP4, WP5, WP6*
- *CLARITY, DC1, DC2, DC3, DC4*
- *input-data* **or** *output-data*
- *open data*

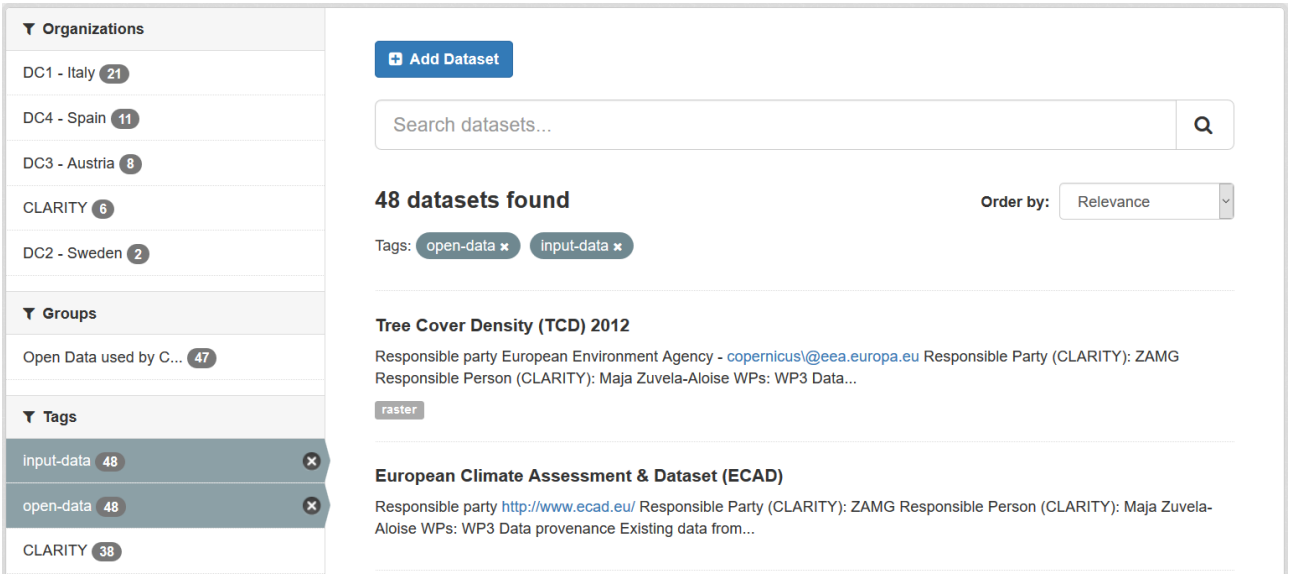

Figure 10: Datasets tagged with *input-data* and *open data* in CLARITY's CKAN catalogue

Second, the dataset has to be assigned to one of the four groups (Figure 2) defined in the CKAN Catalogue.

3.4.1 Non-open data used by CLARITY

This part of the CLARITY DMP reports on **non-open data** *used* by the CLARITY project. The respective datasets in CKAN are assigned the tag **input-data** (https://ckan.myclimateservice.eu/dataset?tags=input-data) (in addition to the respective DC and WP tags) and associated with the CKAN group "Non-open data used by the CLARITY" (https://ckan.myclimateservice.eu/group/non-open data-used-by-clarity) used by CLARITY (Figure 11).

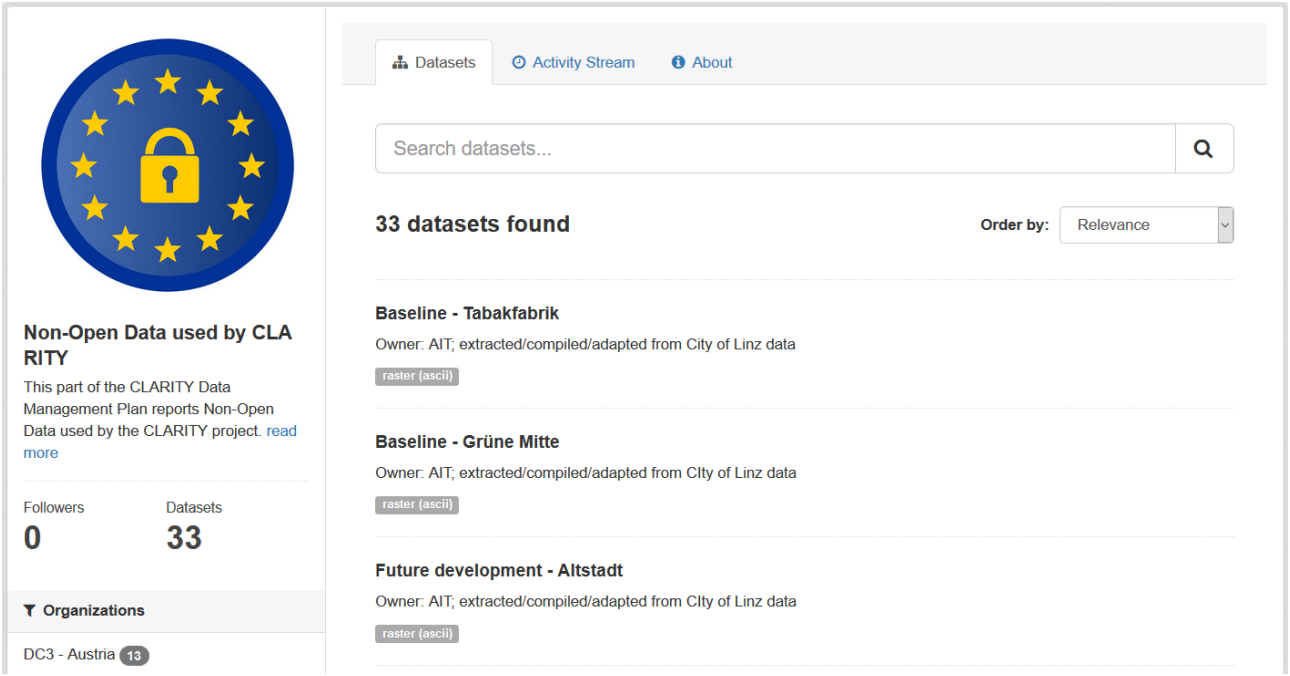

Figure 11: Group "Non-open data used by CLARITY" (excerpt)

Such data is made available or sold by the data provider under a restricted license. It may encompass for example data made available by local authorities (e.g. Comune di Napoli) just for the purpose to serve as input for (micro climate) models in the CSIS (e.g. confidential information on urban planning, population distribution, etc.).

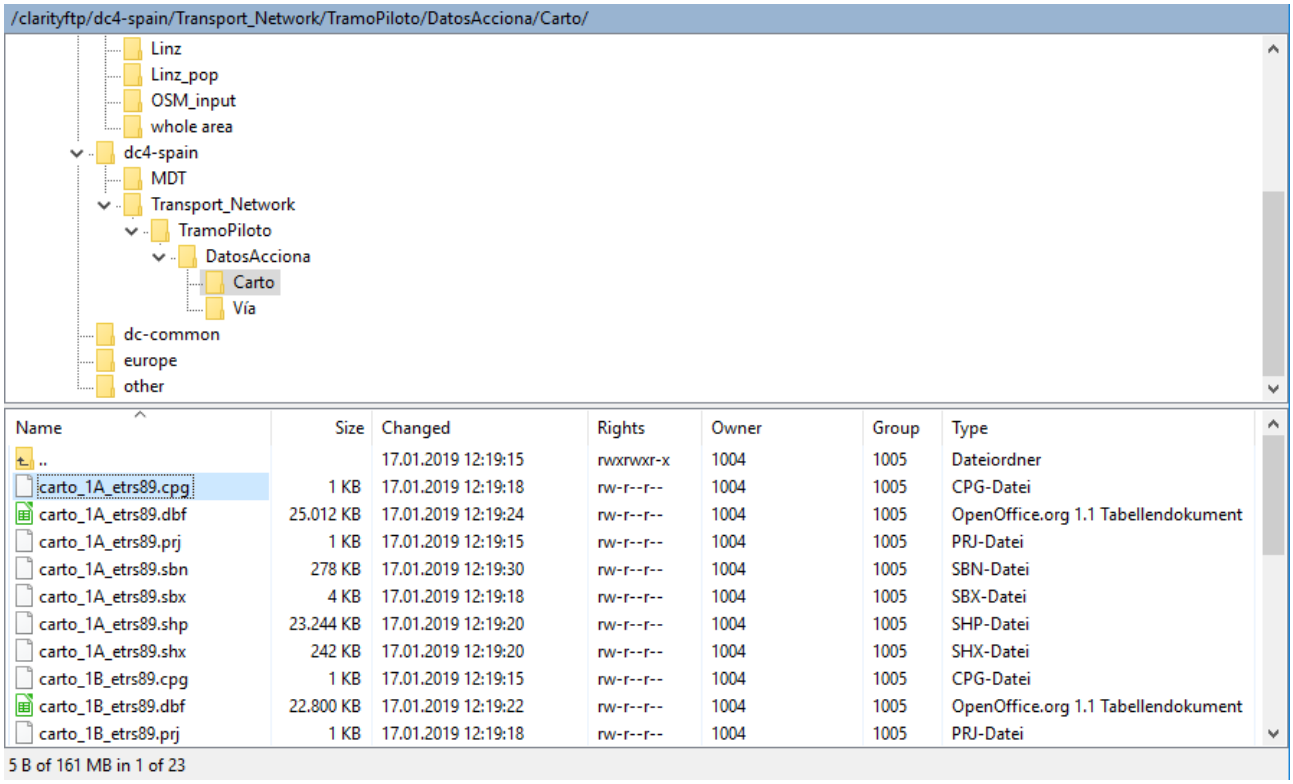

Figure 12: Non-open data is stored is stored securely on CLARITY sFTP server

While the respective meta-data can be published in CLARITY's CKAN, the actual data is stored on the internal CLARITY sFTP server (Figure 12) but not meant to be shared outside of the CLARITY consortium. Thus, the respective Resources (data) associated with the Dataset (meta-data) in CKAN are in most cases links to data files on CLARITY's **access-controlled** sFTP.

3.4.2 Open data used by CLARITY

This part of the CLARITY DMP reports **open data** *used* by the CLARITY project. The respective datasets in CKAN are assigned the tags **input-data** (https://ckan.myclimateservice.eu/dataset?tags=input-data) and **open data** (https://ckan.myclimateservice.eu/dataset?tags=output-data) (in addition to the respective DC and WP tags) and are added to the CKAN group "Open data used by CLARITY" (https://ckan.myclimateservice.eu/group/open data-used-by-clarity) in CKAN (Figure 13).

Such datasets may encompass datasets that are collected from public authorities and institutions like Eurostat (https://ec.europa.eu/eurostat/de/home) or Copernicus (https://www.copernicus.eu/en) and that are released under an open license that allows using the data for research, non-commercial or commercial purposes.

While CLARITY's CKAN Catalogue provides some meta-data related to the usage of the data in CLARITY, the actual data can in general be downloaded directly from the websites of respective organisations. Although some of the data is also stored on the CLARITY sFTP for the purpose of further processing and visualisation in the CSIS, the respective CKAN resource (data) associated with the CKAN dataset (meta-data) links to the original source of the dataset, e.g. the dataset "Digital Elevation Model over Europe" (https://ckan.myclimateservice.eu/dataset/digital-elevation-model-over-europe) links to eea.europa.eu (https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem).

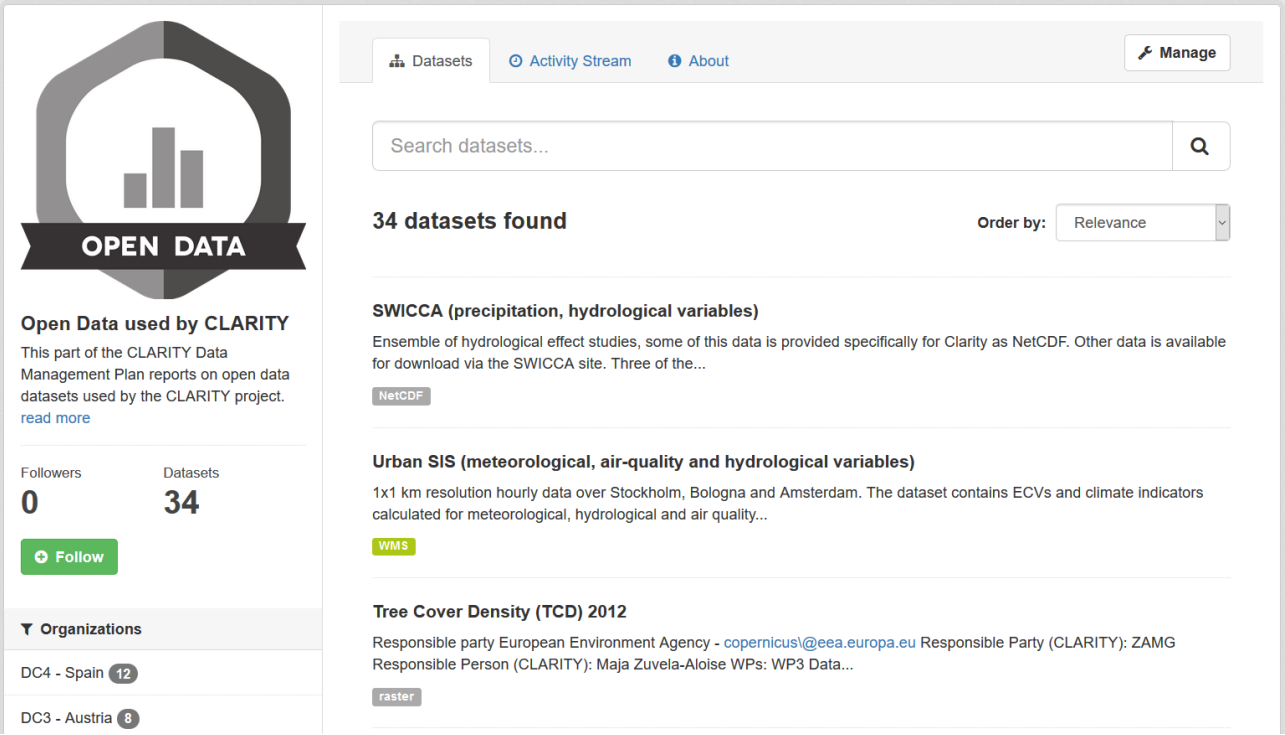

Figure 13: Group "Open data used by CLARITY"

3.4.3 Non-open data produced by CLARITY

This part of the CLARITY DMP reports **non-open data** *produced* by the CLARITY project. The respective datasets in CKAN are assigned the **output-data** tag (https://ckan.myclimateservice.eu/dataset?tags=output-data) (in addition to the respective DC and WP tags) and associated with the CKAN "group Non-open data produced by CLARTIY" (https://ckan.myclimateservice.eu/group/non-open_data-produced-by-clarity) by CLARITY (Figure 14).

According H2020 data management obligations, data produced by the project should be open by default. However, if one of the following general exceptions forbids open access to certain datasets produced by the project, the datasets can be released under a restricted license:

- **copyright and permissions for reusing third-party data sets**
  Processing and combining input data from many different sources may lead to unclear IPR situations regarding the generated output data, therefore such repurposed data (e.g. model output data) can only be made open if any of the underlying data (e.g. model input data) is open, too.
- **personal data treatment and confidentiality issues**
  Datasets referring to the quality and quantity of certain elements at risk, such as people and critical infrastructures, are not open by default as their publication may pose privacy, ethical or security risks.
- **data-driven business model**
  Data that is exploited commercially through the MyClimateService.eu marketplace will not be made open.
- **user-generated content**
  Data related to individual adaptation scenarios (e.g. adaptation options, performance indicators, criteria, etc.) that is generated by (external) end users during the usage of CLARITY CS, is only be made open with explicit permission from the end user.
- **other restrictions**

If such restrictions exist that prevent the provider / producer of the data to release it as open data, the reasons for not doing so have to be summarised in the DMP.

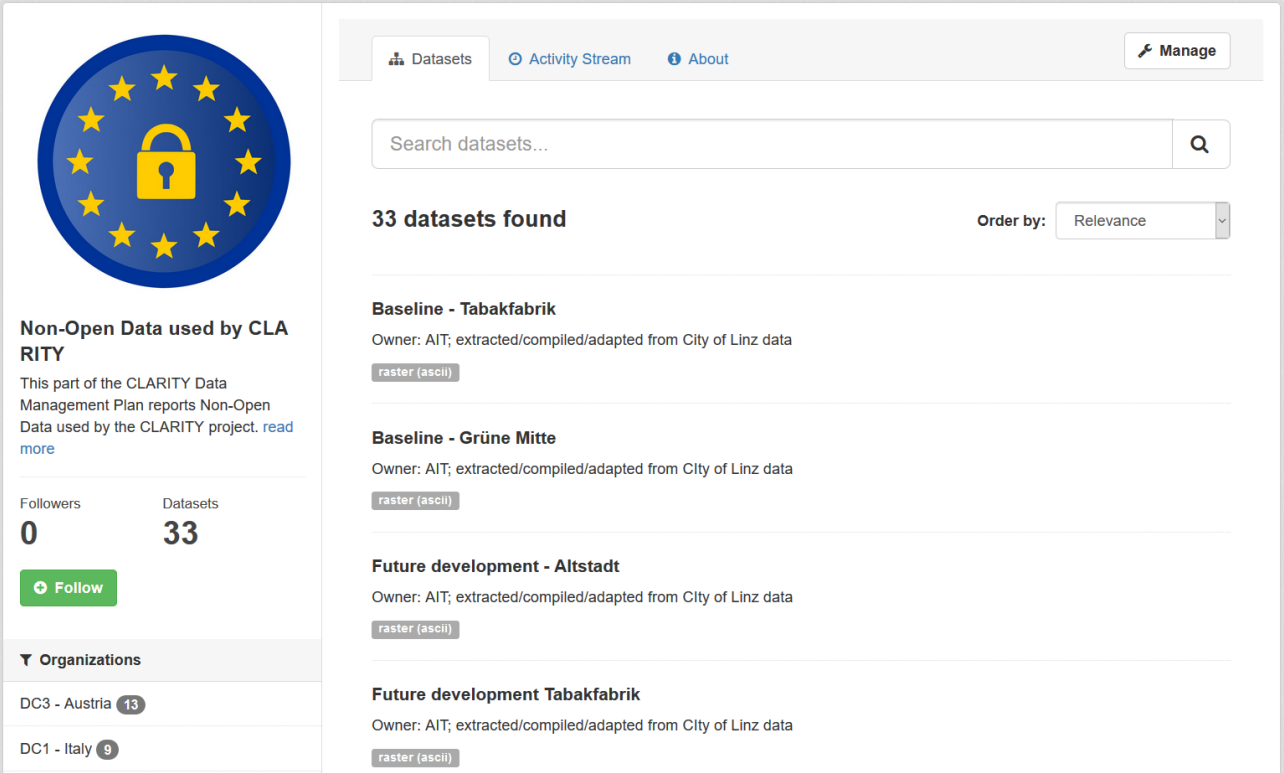

Figure 14: Group "Non-open data produced by CLARITY"

3.4.4 Open data produced by CLARITY

This part of the CLARITY DMP reports **open data** *produced* by the CLARITY project. The respective datasets in CKAN are assigned the tags **output-data** (https://ckan.myclimateservice.eu/dataset?tags=output-data) and **open data** (https://ckan.myclimateservice.eu/dataset?tags=open data) (in addition to the respective DC and WP tags) and associated with the CKAN group "Open data produced by CLARITY" (https://ckan.myclimateservice.eu/group/open data-produced-by-clarity) produced by CLARITY (Figure 15).

CLARITY open results are made accessible according to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. All open results (data, software, scientific publications) of the project have to be openly accessible at an appropriate Open Access repository. Specifically, research data needed to validate the results in the scientific publications has to be deposited in a data repository at the same time as a publication. The main intention data management plan is to ensure that such open data produced by EU-funded projects are deposited in a respective repository and thus are usable by third parties after the end of the project.

However, if confidentiality, security, personal data protection obligations or IPR issues forbid open access to certain data produced by the project, it is deposited in a restricted repository and access may be granted upon request and under the conditions of a restricted license. Such data produced by the project that cannot be released as open data is listed in the category Non-open data produced by CLARITY (see chapter 3.4.3) together with an explanation of the reasons that forbid open access.

Since the CLARITY sFTP is not publicly accessible nor meant for long term archival and preservation of data beyond the lifetime of the project, open data produced by CLARITY is uploaded to an OpenAIRE-compliant repository like Zenodo.

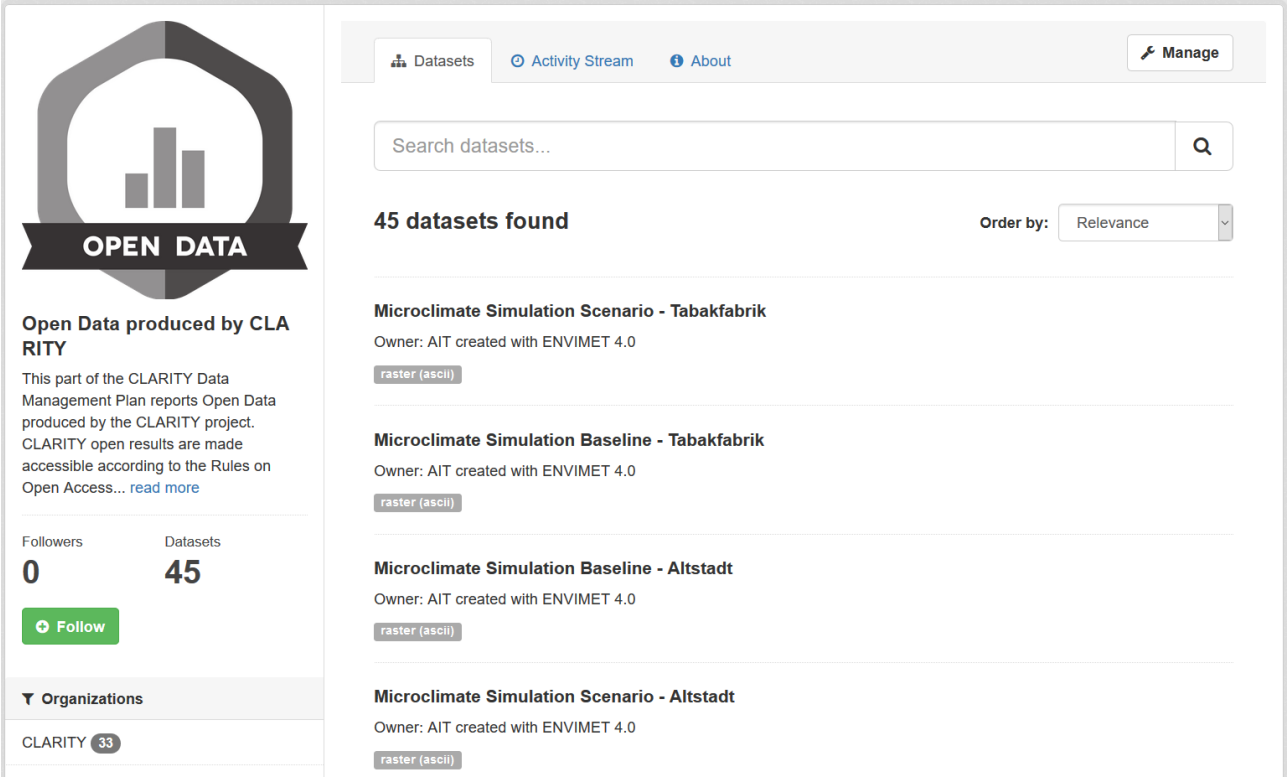

Figure 15: Group “Open data produced by CLARITY”

3.5 CKAN Workflow

The following diagrams describe the general data management workflow for adding datasets (meta-data) and linking resources (data) in CLARITY's CKAN.

3.5.1 Create Dataset

New users that want to register datasets in CLARITY’s CKAN catalogue have to create an account at https://ckan.myclimateservice.eu/user/register. The respective rights for adding datasets to the preconfigured CLARITY organisations (DCs) and data management groups will then be granted by the CKAN administrator. Users just interested in viewing the meta-data stored in the catalogue and downloading the linked and public datasets from the public data repositories do not need an account.

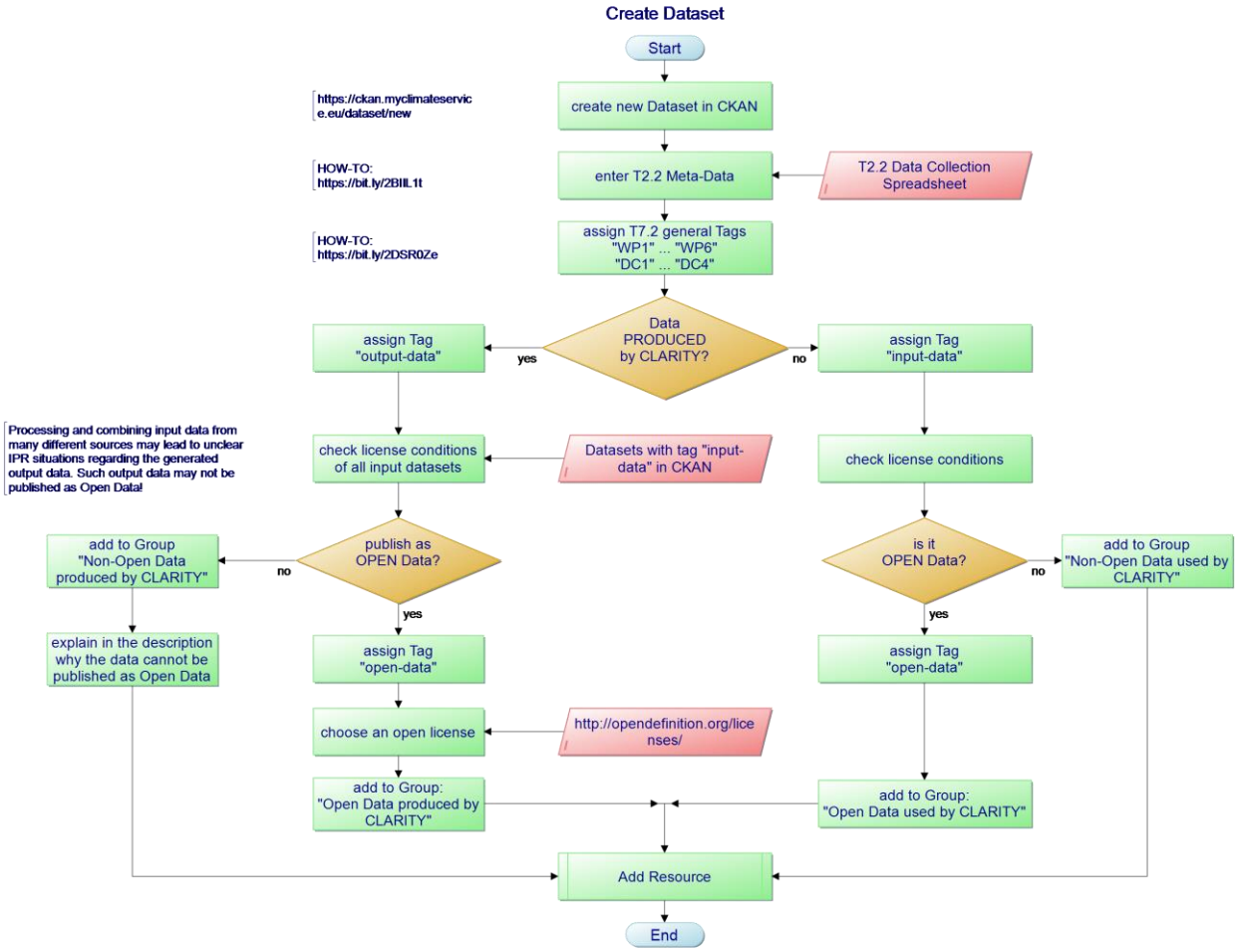

Figure 16: “Create Dataset” CKAN workflow

3.5.2 Add Resources

For data management purposes, it is not sufficient to link to (open) data stored on CLARITY's internal sFTP server as explained in chapter 3.4.4. In that case, a resource has to be added to the dataset that links to the public accessible data that is for example deposited in the Zenodo research data repository.

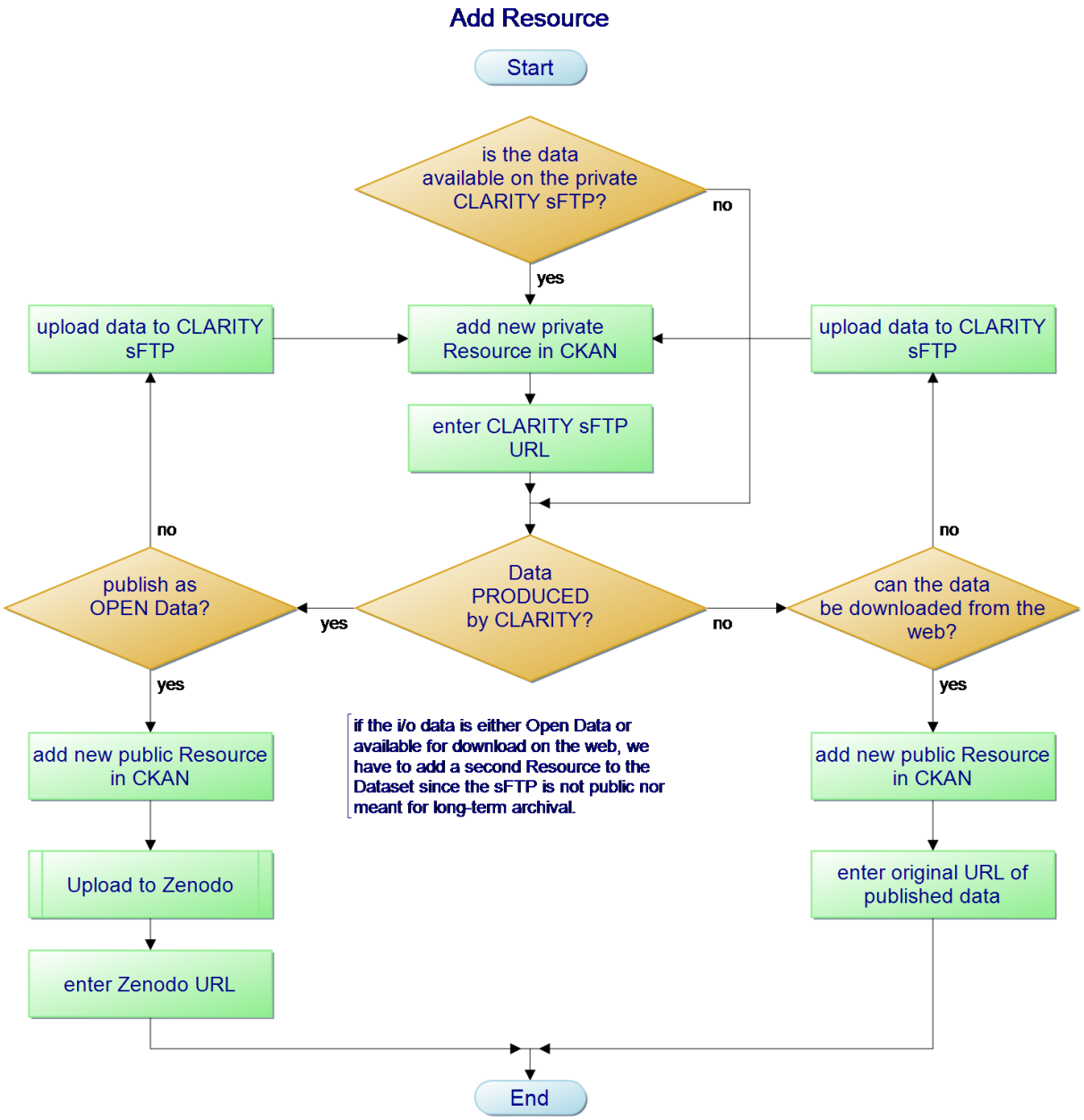

Figure 17: "Add Resource Dataset" CKAN workflow

4 Data Summary

The assessment of climate change adaptation measures requires the collection, processing and storage of a large variety of data sets related to the whole chain of climate change impact and adaptation assessment such as general environmental and meteorological variables, climate change projections based on relevant IPCC scenarios, downscaled climate projections, data on population, infrastructure, land use, etc.

While the first data summary of deliverable D7.8 [1] provided a good overview of the different types of datasets that are required as input of models and tools of WP2 “Demonstration & Validation” and WP3 “Science Support”, the second data summary of deliverable D7.9 reports on the data produced in the context of the project with special emphasis on data that can be made publicly available in open data repositories and registered at relevant catalogues.

Since this chapter provides only a summary of currently available open data produced by the project, for obtaining an up-to-date overview on all datasets used and produced within the project, it is advisable to consult the CLARITY CKAN catalogue at https://ckan.myclimateservice.eu/group/open_data-produced-by-clarity and the CLARITY Community at Zenodo (https://zenodo.org/communities/clarity/).

The catalogue lists currently (February 2019)

- 29 datasets in group “Non-open data produced by CLARITY”;
- 41 datasets in group “Non-open data used by CLARITY”;
- 20 datasets in group “Open data produced by CLARITY”; and
- 52 datasets in group “Open data used by CLARITY”.

The majority of (open data) datasets produced by CLARITY relates to climate change projections. It is based on regional climate model simulations for urban microclimate modelling both at pan-European level (EURO-CORDEX) for general screening studies and on urban microclimate model simulations at local level for tailored expert studies in the context of the four CLARITY DCs. While at pan-European level a simplified urban microclimate model that is feed by open data (see chapter 3.4.2) can be applied, at detailed urban level microclimate models requiring detailed local and partially access restricted data (see chapter 3.4.2) have to be used.

For the pan-European level datasets, CLARITY follows a novel and unique approach for downscaling high-resolution climate projections, in order to be able incorporate such urban microclimate features that strongly affect the risk conditions at “local” level also at European-level (Figure 18). This means, that thanks to the proposed CLARITY methodology that is outlined in details in deliverable D3.2 “Science Support Report v1” [7] and [8], even European-level screening studies can benefit from far better climate projections than such that are usually available “for free” today.

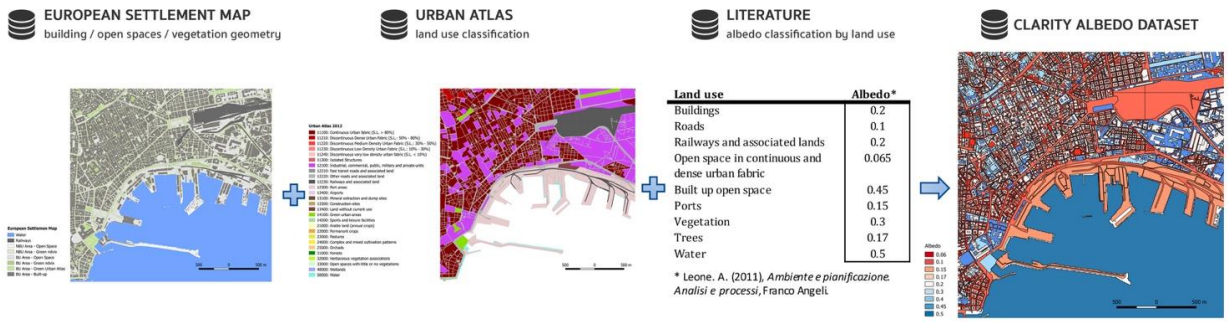

Figure 18: Example of information in the CLARITY urban microclimate simplified model

In fact, the proposed methodology allows to refine the information derived from climate models (see chapter 4.1), with a typical maximum resolution of 10-12km, such as EURO-CORDEX, at the level of a 250x250m mesh overlapped on European urban areas (see chapter 4.3), thanks to the high resolution of satellite data.

The wide amount of data generated by satellite images and made available at pan-European level by the Copernicus programme have been processed with specific algorithms and GIS spatial analysis tools to extract detailed information related to key parameters linked to urban morphology and surface type, such as albedo, emissivity, buildings shadows, green fraction and runoff coefficient. Indeed, the aim is to extend through a new generation of climate services the application of Copernicus data, as a harmonized, freely available and expanding data repository at EU.

Of course, adaptation measures need in fact to be identified and designed locally, since the specific settlement and microclimate conditions determined by the characteristics of built and natural environment play a crucial role in aggravating (or reducing) the intensity of extreme weather events, such as heat waves and flooding. In this sense, a major effort is devoted to embedding urban microclimate projections as additional refining step in the conventional Global Climate Model – Regional Climate Model (GCM-RCM) downscaling approach.

To this end, four different types of highly valuable open data datasets are produced by CLARITY. They are summarised in the subsequent chapters.

4.1 Pan-European Hazard Datasets

As input for the EU-GL steps “Characterise Hazard” and “Hazard Local Effects”, both for European-level screening studies and demonstration case specific expert studies, ZAMG is calculating several climate indices for the hazard characterization at European scale.

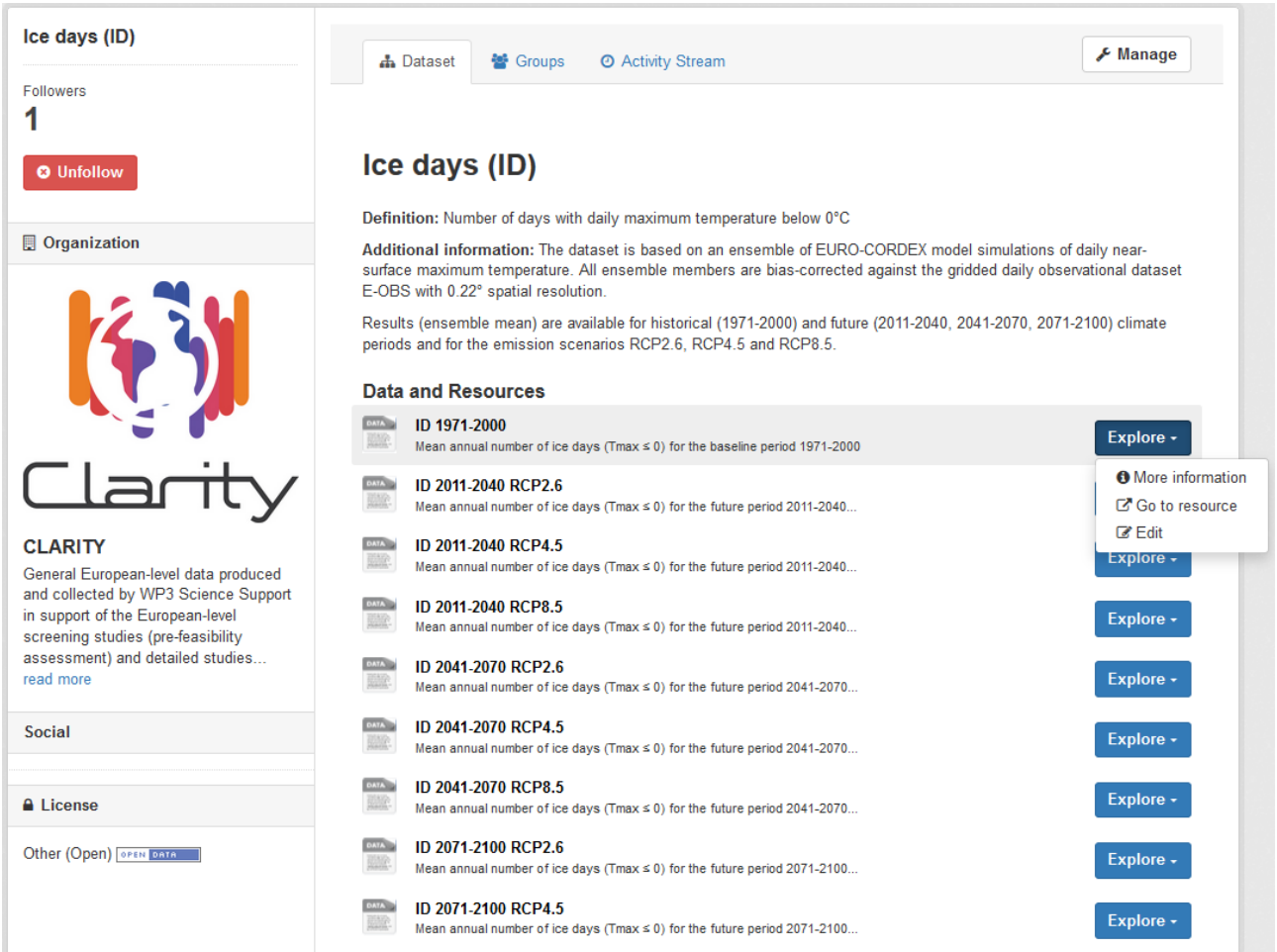

Figure 19: Example of an open data dataset

The indices are being calculated for several Global Climate Model – Regional Climate Model combinations from the EURO-CORDEX (https://ckan.myclimateservice.eu/dataset/euro-cordex-ensemble-climate-simulations) simulations at 0.11° resolution (EUR-11) to account for inter-model variability. However, not the individual indices will be available as open data on Zenodo, but for each climate index, the ensemble mean will be available as open data. Thus, for each climate index there will be an ensemble mean for each time period (1971-2000, 2011-2040, 2041-2070, 2071-2100) and each representative concentration pathway (RCP2.6, RCP4.5 and RCP8.5).

The final indices will be based on bias-corrected EURO-CORDEX data. Since bias correction is still ongoing, the indices calculated so far are based on the original EURO-CORDEX data. Therefore, meta-data about the data sets, which will be made available as open data, were entered in CKAN without adding a link to the actual file. Instead, the link to the CLARITY Zenodo community was entered (https://zenodo.org/communities/clarity/). The data will also be made available on a server of Climate Change Centre Austria (CCCA) at https://data.ccca.ac.at/organization/zamg. The meta-data of the datasets is already available in CLARITY CKAN at https://ckan.myclimateservice.eu/dataset?tags=open data&tags=output-data&tags=WP3, thus becoming part of CLARITY DMP.

In addition to the climate indices that are being produced for whole Europe, urban climate modelling results for the demonstration cases (e.g. heat load maps) will probably also be made available as open data (see also chapter 4.2). However, possible restrictions on usage of local data sets from city administrations need to be checked first.

The original datasets will be netCDF format. For internal usage within the CSIS the data is "rasterised" to GeoTIF 500km grid and made available via GeoServer and AIT EMIKAT. Then the local effects are taken into account to generate derived datasets. For the DMP it is mainly at interest that the original hazard datasets can be made publicly available for re-use by other interested parties.

4.2 Local Effects Input Datasets

Those are the local effects datasets that are applied to the pan-European hazard datasets (chapter 4.1) in order to derive the downscaled datasets (chapter 4.3). They are mainly based on open Copernicus data and encompass detailed information related to key parameters linked to urban morphology and surface type, such as albedo, emissivity, buildings shadows, etc.

Local effects datasets are planned to be calculated for the four pilot regions of the CLARITY DCs. Currently, datasets are available for the metropolitan region Naples (DC1). The related meta-data is made available at CKAN Catalogue at https://ckan.myclimateservice.eu/dataset?tags=Local+Effects and the data itself is stored in the Zenodo research data repository at https://zenodo.org/communities/clarity/. Thanks to generic scripts developed for extracting and combining the local effects parameters from original data sources, the process can be repeated to cover all major pan-European cities. These pan-European datasets will eventually be made available, too. Additionally, the scripts are available as open source at https://github.com/clarity-h2020/local-effects. This open source software, once fully tested and validated, will eventually be deposited in Zenodo to allow other scientist to use, correct or optimise the implementation of the modelling methodology [8].

Local effects datasets have been also published through OGC compliant web services at http://services.clarity-h2020.eu:8080/geoserver following WFS and WMS standards. They have been generated by using different data sources or some combinations of them:

- Urban Atlas (Water, Roads, Railways, Agricultural areas, Medium urban fabric, Low urban fabric, Dense urban fabric, Public, military and industrial units)
- European Settlement Map (Buildings, Built Open Spaces)
- Urban Atlas and Street Tree Layer (Vegetation)
- Urban Atlas and European Settlement Map (Trees)

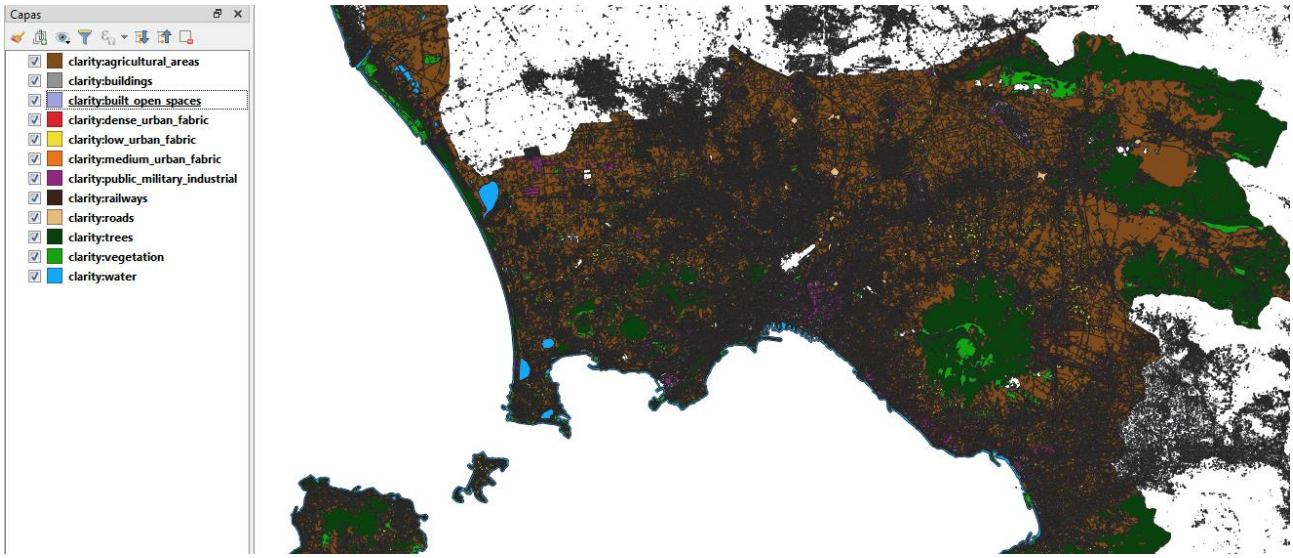

Figure 20: Local effects layers for the City of Naples

4.3 Downscaled Hazard Datasets

The data summarised in the previous chapter is used as the input of a simplified model for urban microclimate assessment developed by CLARITY project at European level [8], useful to perform a screening of the expected impacts of possible adaptation measures. The downscaling is currently performed for the four different study areas (Napoli, Italy; Linz, Austria; Stockholm and Jonkoping; Sweden and Spain). The respective meta-data and data are made available in CLARITY's CKAN catalogue and Zenodo.

Once the bias corrected pan-European Hazard Datasets and the local effects datasets for all major pan-European cities are available, the downscaling can be performed for the whole pan-European level, thus allowing end users to perform pan-European screening studies in CLARITY CSIS.

Downscaled hazard datasets that consider the local effects will be released as open data. These data will be produced during the development of the DCs. The process for combing local effects input datasets with pan-European datasets is performed using GDAL[3].

4.4 DC datasets

While downscaled datasets summarised in chapter 4.3 offer already a considerable advancement compared to the "coarse resolution" of EURO-CORDEX based data (see chapter 4.1) , detailed and dependable adaption planning that is performed in the course of the expert studies of the CLARITY Demonstration Cases requires even better data. Such high-resolution local data resulting from local climate modelling may possibly be restricted on usage from city administrations. It is therefore not guaranteed, that data specifically produced in the context of a DC can be made available as open data. The meta-data of the respective datasets is nevertheless published in CLARITY's CKAN catalogue at https://ckan.myclimateservice.eu/organization.

[3] GDAL is a translator library for raster and vector geospatial data formats that is released under an X/MIT style Open Source license by the Open Source Geospatial Foundation. It also comes with a variety of useful command line utilities for data translation and processing.

4.4.1 DC 1

The Demonstration Case related to the City of Naples (DC1) is intended to assess the benefits introduced by the application of adaptation measures in order to tackle heat waves, pluvial flooding and landslide hazards at local level. The models chosen to meet those purposes are characterized by the link capacity of climate, exposure, vulnerability, and impact data with potential adaptation/mitigation options across multiple scales of intervention. Therefore, the input and output data, both used and produced, regarding heat wave and pluvial flooding hazards, have been described in detail in CLARITY's CKAN catalogue at https://ckan.myclimateservice.eu/organization/dc1 and are stored on the CLARITY sFTP server. The input dataset related to the landslide hazard in the most relevant area of the Metropolitan City of Naples (Castellammare di Stabia), in terms of landslide, is currently being processed.

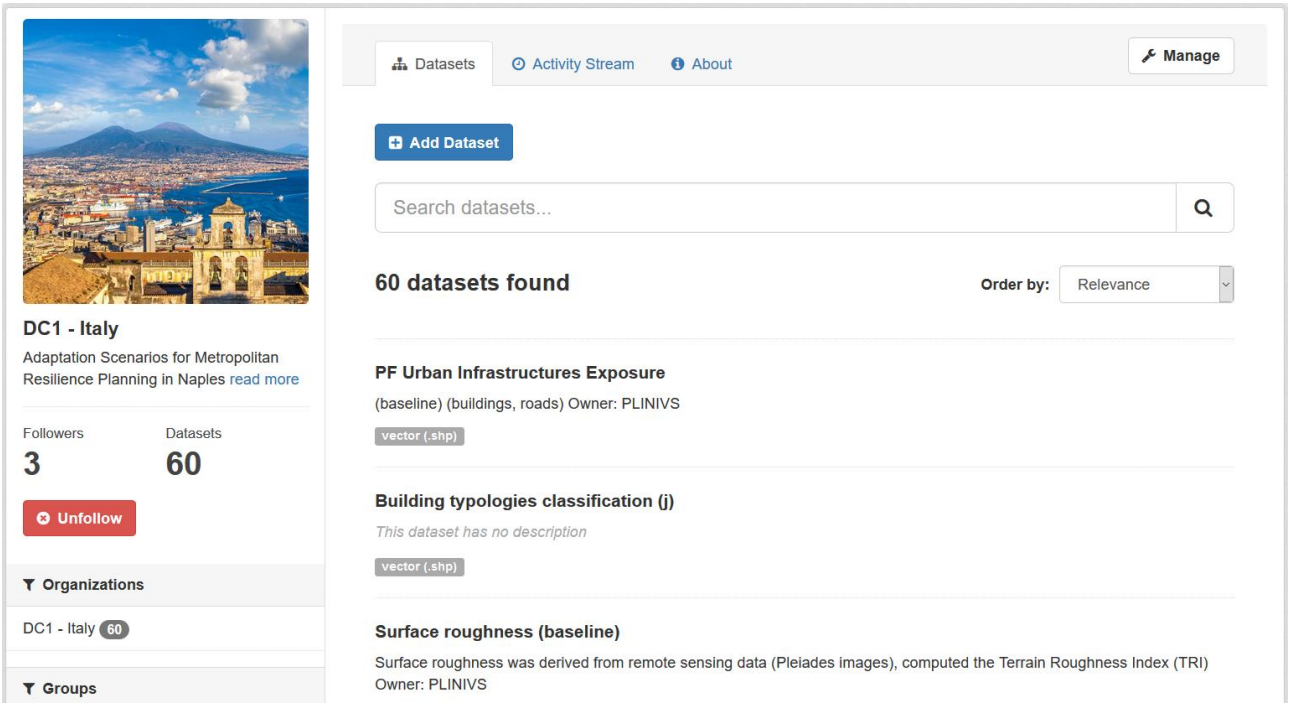

Figure 21: Datasets of DC1 in CKAN

Table 1 summaries input and output data for heat wave and pluvial flooding model. It reports also on the typology of input data (used and produced). The term "used" is applied for depicting the basic information processed for generating the "produced" layers, since the last ones are not directly available but are essential for implementing the models. The license of the "used" input data is set according to the original permission. They can be distributed or sold with an open or a restricted license depending on the possibility to be shared outside of the CLARITY consortium or not. Meanwhile, the "produced" input data will be included in the non-open data produced by CLARITY since they are the outcomes of PLINIVS processing step and are not useful for other users. The actual outputs of the models are also included in the non-open data produced by CLARITY group since they will be exploited commercially through the MyClimateService.eu marketplace. In addition, all the information has been distinguished in two categories: layers tagged with the term "baseline", able to describe the current (climatic) situation, and those tagged with the name "j", result from adaptation strategies application.

All the baseline input data, both for Heat Wave (HW) and Pluvial Flooding (PW), are already available. Either just for the whole city of Naples or for some sample areas of the municipality of Naples, which will be extended to the whole municipality by the end of month 24. First outcomes of HW and PF models have been already produced, even if the final results will be generated by the end of the project. Available datasets are stored on the CLARITY sFTP server and will be integrated into the CSIS.

Table 1 - Heat wave and pluvial flooding datasets

Models	Input data	Output data
Heat Wave (cuboid method implemented in MUKLIMO environment)	• Meteorological data (used) • Digital Surface Model and Digital Elevation Model (used) • Pleiades images (used) • Land Use (used) • Mean building height; Building typology classification; Wall area index; K-value of the building walls and roofs; Area heat capacity of the building walls and roofs (produced) • Fraction of impervious surface between buildings; Surface roughness of the non-built-up areas (produced) • Vegetation parameters (Tree height, Stem height, Leaf area density, Leaf area index, Vegetation height of the canopy layer, Tree cover, Vegetation cover) (produced) • Albedo of the walls, roofs and impervious parts of the canopy layer (produced)	• Heat Wave local effect map
Pluvial Flooding (PLINIVS simplified model)	• Precipitation data (used) • Digital Surface Model (DSM) and Digital Terrain Model (DTM) • Land Use (used)Basins; Flow direction; Flow accumulation; Run off (produced)	• Pluvial Effect local effect map
PLINIVS HW / PF Exposure model	• ISTAT census data (used) • Land use (used) • HW Population Exposure (produced) • PF Urban Infrastructures Exposure (produced)	• Heat Wave exposure map • Pluvial Flooding exposure map
PLINIVS HW / PF Vulnerability model	• Heat Wave Population Exposure (produced) • Pluvial Flooding Urban Infrastructures Exposure (produced)	• Heat Wave vulnerability map • Pluvial Flooding vulnerability map • Heat Wave vulnerability curves • Pluvial Flooding vulnerability curves
PLINIVS HW / PF Impact model	• Heat Wave vulnerability map (produced) • Pluvial Flooding vulnerability map (produced) • Heat Wave exposure map (produced) • Pluvial Flooding exposure map (produced)	• Heat Wave impact map • Pluvial Flooding impact map

4.4.2 DC 2

DC2 produces two datasets that are of general interest to release as open data: The scenarios for flooding in Sweden under a future climate and the urban heat scenarios over Stockholm. SHMI, as one of the data providers, intends to release these data as open by the end of the project. However, due to new regulations in Sweden SHMI might be restricted in what they are allowed to release. This issue will be clarified until the release of the next version of the DMP (deliverable D7.10).

Other data produced as result of expert studies, such as detailed flooding data over Stockholm, will not be openly released since it is based on data where restrictions in usage applies, due to commercial conditions or confidentiality.

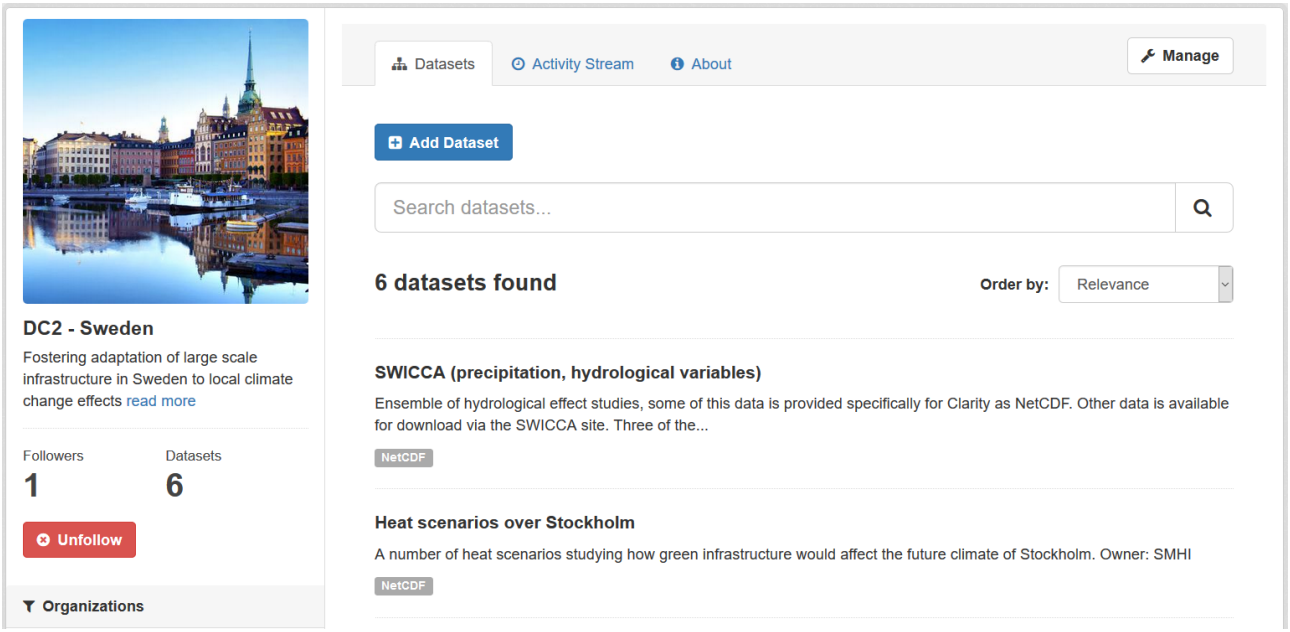

Figure 22: Datasets of DC2 in CKAN

Datasets used and produced in context of this Demonstration Case are described in detail in CLARITY's CKAN catalogue at https://ckan.myclimateservice.eu/organization/dc2.

4.4.3 DC 3

For entire City of Linz climate modelling various data sets describing the spatial characteristics of the study area have been compiled from existing open data and new ones generated by merging and adapting those data as well as from non-open data, property of City of Linz and Government of the province Upper Austria.

These data serve as spatial framework characteristics for climate and microclimate simulations: Digital terrain model, land use map (Urban Atlas, OSM), green area inventory and cadastre map tree distribution, etc. (City of Linz). The applied simple 3D city model and urban development scenarios (densification, new urban green) was generated by AIT.

These datasets are used to calculate different scenarios according to the user stories of this DC and considering the following measures: unsealing of land, roof greening and tree cover densification. Furthermore, the effects of new settlement areas on urban climate can be simulated and different adaptation scenarios can be calculated to make recommendations in terms of resilient urban planning.

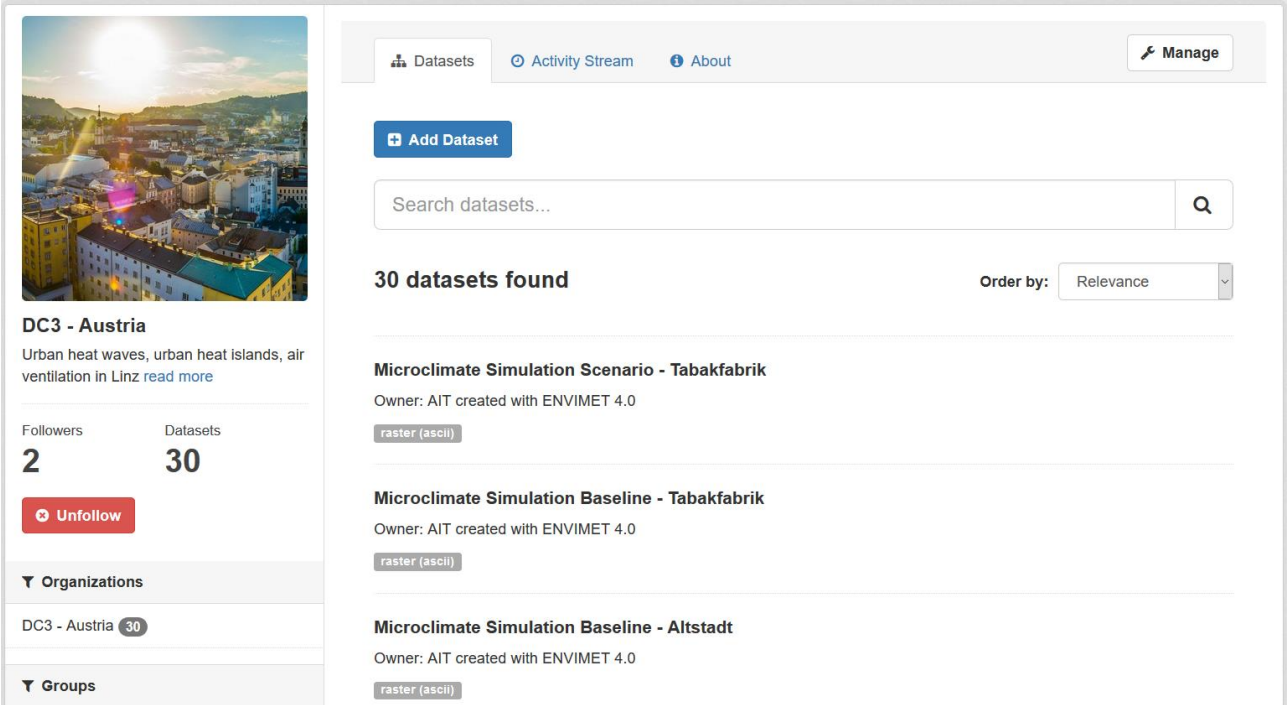

Figure 23: Datasets of DC3 in CKAN

Some of the produced datasets will – depending on the applied modelling tools and licenses – be published as open data. In particular, results obtained from the urban climate model MUKLIMO_3, e.g. climate indices at 100 m spatial resolution, are intended for non-commercial public use, after complying to the conditions of the MUKLIMO_3 modelling license for non-commercial tasks in research and teaching and with consent of all input data owners.

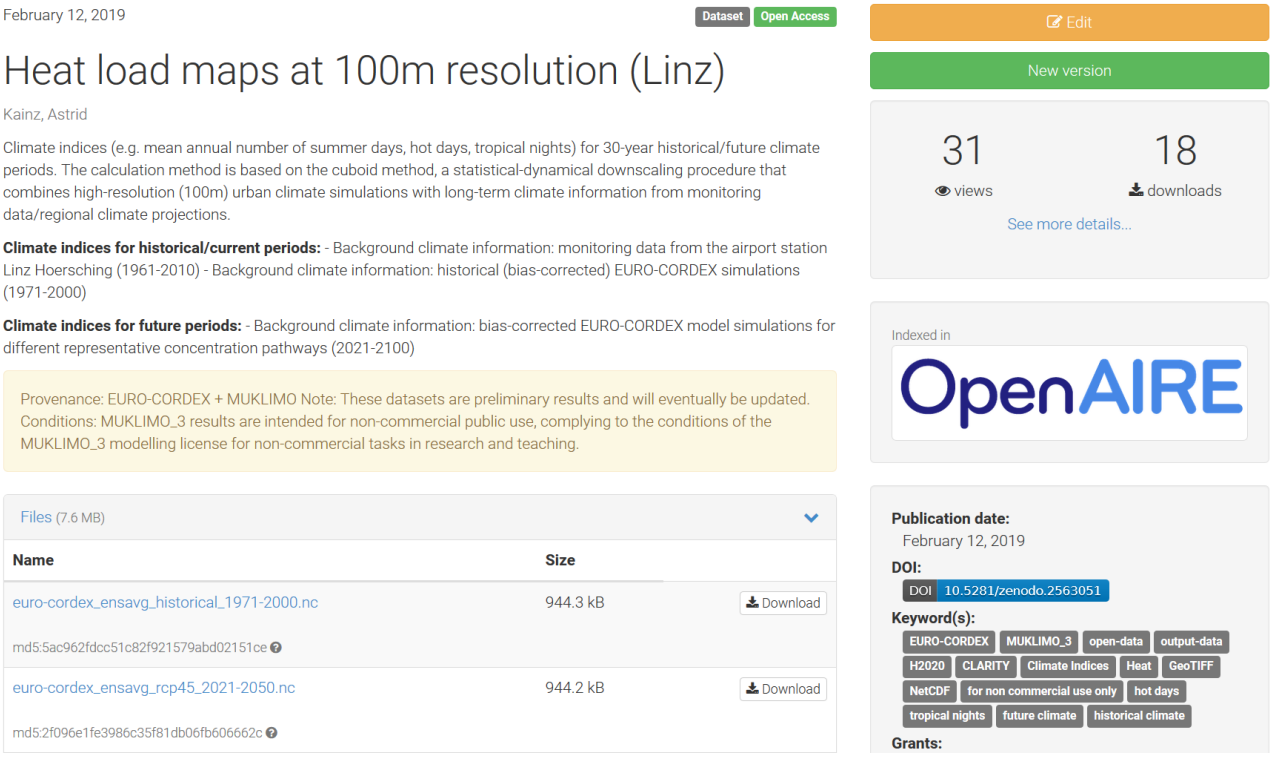

Figure 24: CLARITY open data in Zenodo research data repository

Some of the preliminary MUKLIMO_3 results (NetCDF format) have already been uploaded to Zenodo at https://doi.org/10.5281/zenodo.2563051 (Figure 24).

The main dataset related to MUKLIMO_3 output (Heat load maps at 100 m resolution; https://ckan.myclimateservice.eu/dataset/heat-load-maps) comprises urban climate indices (e.g. mean annual number of summer days, hot days, and tropical nights) for 30-year historical/future climate periods. These datasets are preliminary results and will eventually be updated using Zenodo's DOI versioning[4] feature.

The dataset will be extended by adaptation scenarios (e.g. roof greening, unsealing, increased albedo) resulting from a modification of input data with respect to the baseline (current climate conditions). After deciding which scenarios are the most relevant to be selected for the final implementation of DC3, the respective data will be published as open data and corresponding meta-data will be described in CKAN.

4.4.3.1 Linz demonstration cases - microclimate modelling

To generate microclimate simulations for "Tabakfabrik", "Grüne Mitte" and "Altstadt" case studies the relevant input layers have been extracted from the Linz-wide data sets as ENVIMET format as well as shape-data sets as input for RHINO/Grasshopper[5] and are now provided as open data. The case study areas of the most relevant results for both baseline and future development scenarios (e.g. roof greening/façade greening, implementation of trees and parks) will be published as open data with underlying meta-information on CKAN.

For the three case study areas, microclimate simulations have been conducted applying the ENVIMET 4.0 as well as the Rhino/Grasshopper as microclimate simulation tools. Various output parameters (air temperature, relative humidity, wind speed, wind direction, mean radiant temperature (MRT) and several heat comfort parameters (PET, PMV, etc.)) are provided as open data for those case study areas for current state and future urban development scenarios (urban density, additional urban greening)

Most relevant results (e.g. see https://ckan.myclimateservice.eu/dataset/microclimate-simulation-scenario-altstadt) are provided as metadata made available through CKAN at https://ckan.myclimateservice.eu/organization/dc3 and as open data for download at https://zenodo.org/communities/clarity/.

4.4.4 DC 4

Climate indices are calculated at Spanish level from CORDEX data in order to assess the vulnerability and risk of Spanish roads to climate change. For this purpose, the models available in EURO-CORDEX are used for two scenarios (RCP4.5 and RCP8.5) for three fixed periods (2011-2040, 2041-2070, 2071-2100) and a reference period (1971-2000). The results refer to different emission scenarios, different global climate models and different regional models. The spatial resolution of the EURO-CORDEX11 simulations is 12.5km.

Datasets used and produced in context of this DC are described in detail in CLARITY's CKAN catalogue at https://ckan.myclimateservice.eu/organization/dc4.

[4] This new feature enables users to update the record's files after they have been made public and researchers to easily cite either specific versions of a record or to cite, via a top-level DOI, all the versions of a record. (https://help.zenodo.org/#versioning)

[5]Grasshopper is a graphical algorithm editor tightly integrated with 3-D modelling tools for designers who are exploring new shapes using generative algorithms. (https://www.grasshopper3d.com)

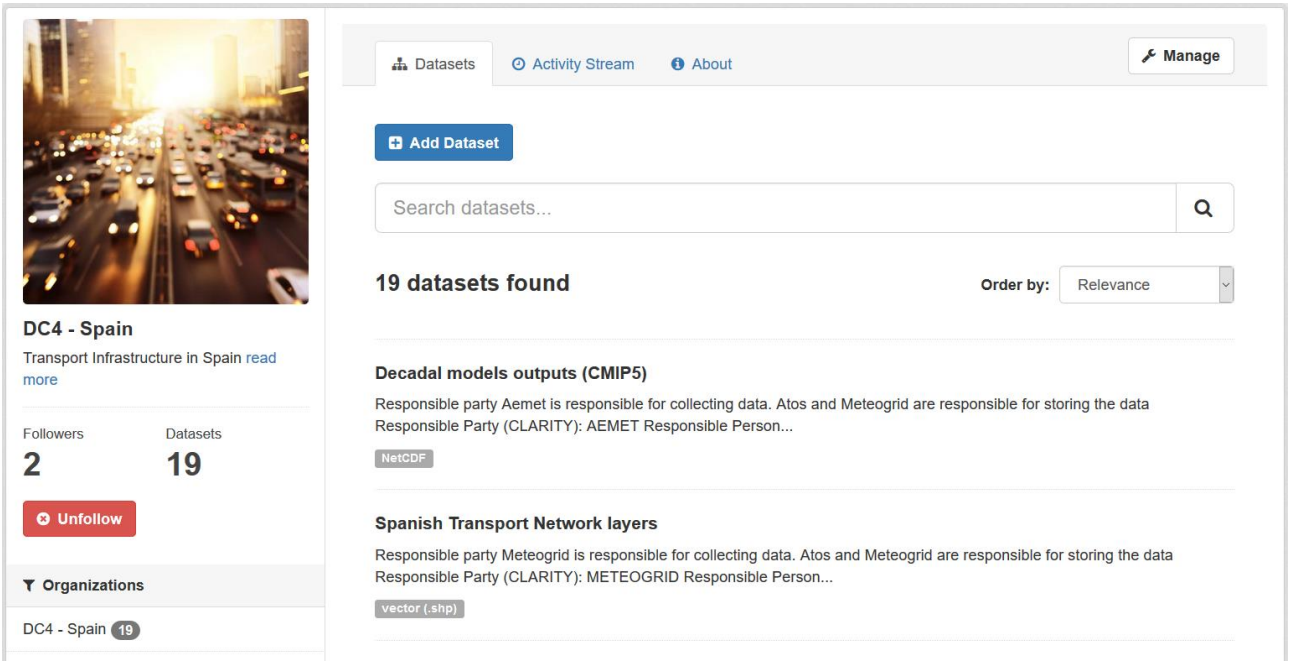

Figure 25: Datasets of DC4 in CKAN

5 Conclusion

The CLARITY project makes use of a large and diverse amount of open data offered by European agencies or EU-funded projects and initiatives (e.g. Copernicus, EU-CORDEX) as well as data from municipalities and institutions from the CLARITY Demonstration Cases. The CLARITY co-creation team combines, transforms and enriches these dataset by applying scientifically sound methods (e.g. hazard local effects methodology), open source software (e.g. Urban Atlas feature extraction scripts) and models (e.g. urban microclimate models) to produce new data (e.g. pan-European climate indices with and without local effects applied). The datasets are then assembled into so-called Data Packages that are used within the CLARITY CSIS to perform screening studies on pan-European level or detailed expert studies on the level of the CLARITY DC regions following the IPCC AR5/DRR/EU-GL methodology.

One of the main aim of data management procedures established in CLARITY is to ensure that data that is produced in the context of the project and that is not subject to commercial exploitation or access restrictions can be made available as open data following the FAIR (findable, accessible, interoperable and reusable) principle. In particular, open data that can be used by third parties, possibly in different contexts, to generate new beneficial results, including new open data. Accordingly, such open data is one of the sustainable results of the project and is/will be deposited for long term preservation in the Zenodo research data repository (https://zenodo.org/communities/clarity/), the DRMKC Risk Data Hub (https://drmkc.jrc.ec.europa.eu/risk-data-hub) and the Climate Change Centre Austria's (CCCA) Data Server (https://data.ccca.ac.at/organization/zamg).

The CLARITY DMP reports on data used and produced within the CLARITY, including open and non-open data. It is implemented as a "living" DMP that is continuously updated throughout the course of the project at https://ckan.myclimateservice.eu. A complete summary of what (open) data has been produced by the project and what data is released under an open license will be provided in the next report (deliverable D7.10) that accompanies the CLARITY DMP.

6 Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730355.

References

[1] P. Dihé, "D7.8 Data Management Plan v1," Deliverable D7.8 of the European Project H2020-730355 Integrated Climate Adaptation Service Tools for Improving Resilience Measure Efficiency (CLARITY), 2017. [Online].

[2] C. Ramjoue and O. Marganne, "TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)," 13 October 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx. [Accessed 3 November 2017].

[3] Directorate-General for Research & Innovation, "Guidelines on FAIR Data Management in Horizon 2020, Version 3.0," EUROPEAN COMMISSION, 26 July 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. [Accessed 28 October 2017].

[4] A. Rivera-Campos, L. Torres, M. Leone, F. Gallinella, L. Asensio, S. Rubio, A. Valverde, A. Kainz, R. Goler, M. Zuvela-Aloise, W. Loibl, L. Strömbäck and J. H. Amorin, "D2.2 Data Collection Report," Deliverable D2.2 of the European Project H2020-730355 Integrated Climate Adaptation Service Tools for Improving Resilience Measure Efficiency (CLARITY), 2019. [Online].

[5] P. Dihé, "D4.3 Technology Support Report," Deliverable D4.3 of the European Project H2020-730355 Integrated Climate Adaptation Service Tools for Improving Resilience Measure Efficiency (CLARITY), 2019. [Online].

[6] P. Dihé, "D4.2 CLARITY CSIS Architecture," Deliverable D4.2 of the European Project H2020-730355 Integrated Climate Adaptation Service Tools for Improving Resilience Measure Efficiency (CLARITY), 2018. [Online].

[7] M. Zuvela-Aloise, R. Goler, A. Kainz, W. Loibl, A. Compte, L. Torres, L. Parra, M. Iorio, M. Leone, G. Zuccaro, A. Capolupo and F. Gallinella, "D3.2 Science Support Report," Deliverable D3.2 of the European Project H2020-730355 Integrated Climate Adaptation Service Tools for Improving Resilience Measure Efficiency (CLARITY), 2019. [Online].

[8] G. Zuccaro, M. Leone, S. Nardone and C. Alessandra, "Urban microclimate modelling methodology at pan-European level to support adaptation planning and design in cities," in *European Climate Change Adaptation conference*, Lisbon, 2019.

