



ClairCity - Citizen Led Air pollution Reduction in Cities

D1.9 ClairCity Data Management Plan

October 2016 – Updated in May 2018, August 2018

Confidential (only for members of the ClairCity Consortium including the Commission Services)

Contents

Contents	2
Document Details.....	3
Version History	3
Contributions and Acknowledgements.....	4
1. Introduction to the ClairCity Project	5
2. Purpose of Data Management Plan.....	5
3. The ClairCity Data Landscape	6
4. Dataset reference and name	7
4.1. Compliance with the FAIR Principles.....	7
5. Data set description.....	8
5.1. Type of external data	8
5.2. Data generation.....	9
5.3. Data formats	10
5.4. Quality of data.....	10
5.5. Usefulness of data and related data.....	11
6. Standards and metadata	12
7. Data sharing	12
7.1. Data access.....	12
7.2. Data sharing mechanisms.....	12
7.3. Data sharing limitations.....	14
7.4. Data security	14
8. Archiving and preservation of data.....	14
8.1. Data storage	14
8.2. Backing up of data	15
8.3. Maintaining and updating data	15
8.4. Costs of data storage	15
8.5. Archiving of data after project life	15
9. Flow and responsibilities	15
10. Data Management Plan Maintenance	16
Annex - Data collection and storage across the project, per partner	17

Document Details

Authors	Kris Vanherle (Transport & Mobility Leuven – TML) Olav Ten Bosch (Centraal Bureau voor de Statistiek – CBS)
Contact	kris.vanherle@tmleuven.be o.tenbosch@cbs.nl
Creation Date	06/09/2016
Date of Last Revision	27/08/2018
Version	V7
Description	The purpose of this Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used within the ClairCity project with regard to all the datasets that will be generated by the project. The DMP serves to summarise the way data is handled in the project, what kind of data is used, how it is collected, stored and shared among partners and stakeholders. The deliverable has some links with D8.1&D8.2 Ethical Framework. The ClairCity DMP has been updated in May and August 2018.

Version History

Version	Updated By	Date	Changes / Comments
v1.0	Kris Vanherle	06/09/2016	First draft
v1.1	Kris Vanherle	29/09/2016	Comments from Enda Hayes (UWE)
v1.2	Hans Bolscher Irati Artola Kris Vanherle	15/10/2016	Final comments included
v2.0	Enda Hayes	19/10/2016	Review for final for submission to Portal
v3.0	Olav Ten Bosch & Kris van Herle	08/05/2018	Redraft for internal review

v4.0	Enda Hayes	21/05/2018	Review for final for resubmission to Portal
V5	Laura Fogg-Rogers	16/07/2018	Review to integrate ethics related aspects
V6	Irati Artola	20/07/2018	Updated with data collection, processing and storage information collected from all consortium partners

Contributions and Acknowledgements

The authors would like to thank the following people for their important contributions used in the preparation for this final document.

Quality Assurance	Enda Hayes (UWE)
Native Language Check	Enda Hayes (UWE)
Project internal comments	Hans Bolscher (Trinomics) Irati Artola (Trinomics) Laura Fogg-Rogers (UWE)
Other contributions	All partners have provided inputs regarding their data collection, processing and storage

1. Introduction to the ClairCity Project

ClairCity is a Horizon 2020 funded project responding to the call 'Improving the Air Quality and Reducing the Carbon Footprint of European Cities' (SC5-4-2015). ClairCity will apportion air pollution emissions and concentrations, carbon footprints and health outcomes by city citizens' behaviour and day-to-day activities in order to make these challenges relevant to how people chose to live, behave and interact within their city environment. Through an innovative engagement and quantification toolkit, we will stimulate the public engagement necessary to allow citizens to define a range of future city scenarios for reducing their emissions to be used for supporting and informing the development of bespoke city policy packages out to 2050.

Using six pilot cities/regions (Amsterdam, NL; Bristol, UK; Aveiro, PT; Liguria, IT; Ljubljana, SI; and Sosnowiec, PO), ClairCity will source apportion current emissions/concentrations and carbon emissions not only by technology but by citizens' activities, behaviour and practices. ClairCity will explore and evaluate current local, national and international policy and governance structures to better understand the immediate policy horizon and how that may impact on citizens and their city's future. Then, working with the new methods of source apportionment to combine both baseline citizen and policy evidence, ClairCity will use innovative engagement methods such as games, apps and city days to inform and empower citizens to understand the current challenges and then subsequently define their own visions of their city's future based on how they want to live out to 2050. The impact of these citizen-led future city scenarios will be analysed, to develop city specific policy packages in which the clean-air, low-carbon, healthy future, as democratically defined by the city citizens, is described and quantified. The results of the ClairCity process will be evaluated to provide policy lessons at city, national and EU levels. Additionally, the toolkit structure will be developed for all EU cities with more than 50,000 citizens establishing a basis to roll out the ClairCity process across Europe.

For more information please visit the ClairCity website – www.claircity.eu

2. Purpose of Data Management Plan

The purpose of this Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used within the ClairCity project with regard to all the datasets that will be generated by the project.

In ClairCity different types and volumes of data will flow between project partners and stakeholders. The DMP serves to summarise the way data is handled in the project, what kind of data is used, how it is collected, stored and shared among partners and stakeholders. Such a DMP is important for the efficiency and integrity of the research by keeping data organised, accessible and secure, and in turn to avoid duplication, and enhancing validation and replicability when needed.

Given the scope and complexity of the ClairCity project, it is not possible to address each dataset individually at this stage; instead in this DMP we describe data policies rather than individual data. This document will be an evolving document that will grow over the lifetime of the ClairCity project and will be updated periodically.

This document, together with other referenced documents, defines the responsibilities and procedures to be adopted to ensure that data and information produced as part of the ClairCity

project are reliable, fit for purpose and consistent with documented objectives and deliverables and meets the requirement set out by Horizon 2020. It summarises the system of internal management that governs the decisions and instructions concerning project data management.

This DMP covers staff responsibilities, data collection policies, resourcing of data management, data quality, intellectual property right considerations, policies on access to data and data dissemination during the project and arrangements for long-term management of data beyond the project.

Since ClairCity is a complex research project where the right data architecture will gradually become clear during research phases, this DMP is not a fixed document. It will evolve during the lifespan of the project. This is a revised version of the DMP, describing the expected data policies as far as can be seen the end of year 2 for the project.

3. The ClairCity Data Landscape

In short, ClairCity will need data for developing advanced models as well as data to feed the products that were developed using these models. Both types of data will use as many existing data from already available data sources as possible, for example already available data from statistics offices on land-use, emission data from sensors and transportation data from mobility research. Only when necessary additional data collection will take place.

Even when using existing data, within the ClairCity project a considerable amount of data processing will take place as data is almost never in the right format, on the right aggregation level or using the right classification scheme. This will especially happen when combining data sources from multiple disciplines, as is the case in ClairCity. Both for administering the use of existing data, as well as for carefully managing the results of data modelling, data processing and analysis, it is essential to organize data well. For this a [ClairCity Data Portal](#) has been established. This Data Portal will play a central role in the organisation of data in the project and will follow the **FAIR principles for data management** in European projects as appropriate. In particular it supports symbolic data identifiers for indexing and finding ClairCity specific public datasets. The data portal is based on a widely used generic Data Portal software system (CKAN) supporting acceptable metadata standards, supplemented with ClairCity metadata and extensions.

In ClairCity we make a clear distinction between privacy-sensitive data that might be collected by cities for feeding the ClairCity products and data that is not privacy-sensitive and can be shared among the ClairCity partners and to the world. Privacy sensitive data will not be shared among project partners and will not be allowed to be uploaded to the ClairCity Data Portal; this is outlined further in Section 9.4 of this document. In the sequel of this document we describe the data management policies for non-privacy sensitive data to be managed in the Data Portal.

The ClairCity Data Portal is conceived as a “single point of access” for all (major) datasets in the project. The portal can give access to a database /-file developed in the project or link to an original source (e.g. SILC, Eurostat, external city database...) including meta-data. It must be generic because the different data sources will have different formats and different structure and cover a different geographical area or time span or time resolution. The Data Portal aims at offering the following primary services to the rest of the project.

1. Maintaining and issuing a **catalogue** and search facility which researchers and modellers can consult in order to discover data sets about ClairCity subjects
2. Providing best practices and **standards** for a uniform description for these dataset with **metadata** keywords, tags etc.
3. Offering an **upload facility** for data sets on this portal in accordance with best practices and standards
4. Maintaining a facility with which modellers /researchers can manage their **data**, provided they have the proper authorization
5. Supporting the modellers in **describing and storing the results of their modelling**

The Data Portal has been developed iteratively; starting from a generic Data Portal with support for some basic generic data formats, adding more specific ClairCity functionality on each iteration.

4. Dataset reference and name

In the ClairCity Data Portal each dataset can be either an upload of some data or a link towards another publicly available dataset. In the latter case, a reference to the original data will be administered. Each ClairCity dataset in the Data Portal can be referenced by a single unique identifier, has a name and a description and some other minimal set of meta fields.

Only organisations participating in the ClairCity project will be able to manage datasets in the ClairCity Data Portal. Datasets uploaded to the portal will be open to the consortium by default. In addition, datasets worth sharing widely can be opened to the general public. Details to follow in section 10.

4.1. Compliance with the FAIR Principles

Every dataset in the ClairCity Data Portal will have a number of ClairCity specific metadata fields which will be designed within the project. One of the mechanisms that will be offered is tagging with ClairCity specific tags. This is useful to facilitate data discovery.

Since ClairCity aims to develop a new models for air quality improvements throughout Europe, it is desirable to meet the **FAIR Principles** where appropriate. We do this as follows:

- The design of the data portal supports the FAIR principles through the use of **unique identifier** generated by the data portal system from the name of the dataset. This adds to the principle of findability as search engines build up their index of the internet from URLs. Meaningful URLs (such as URLs with a unique identifier) will be higher in an index than non-meaningful URLs but it needs to be recognized that only public datasets can be found by search engines.
- In addition, every dataset has a unique identifier that is stable for the lifetime of the dataset. This adds to the principle of findability and accessibility.
- Also every dataset has a number of predefined metadata fields (organisation, license, city, assignment to the conceptual ClairCity model, etc.) which adds to the principles of findability and interoperability.
- In addition, there is a free description field, which is to be used (among other things) for describing the method / algorithm being used to generate the data / model. This adds to the principle of reusability.
- Finally, the data portal supports the option of tagging. Dataset owners can add their own tags to datasets, which build up a knowledge base of tags describing dataset contents. Tagging supports findability for datasets.

5. Data set description

5.1. Type of external data

Due to the nature of the ClairCity project, data to be used in the project (secondary data) may come from various domains. When writing this deliverable, data collection for the first case city, Amsterdam, is already progressing. We list a few examples of datasets primarily from the Amsterdam case:

Typical examples are:

- Geocoded data on land-use within the city, in shape-file format or GeoJSON:
 - European scale
 - Urban morphological zones (UMZ) defined by Corine land cover classes considered to contribute to the urban tissue and function http://ftp.eea.europa.eu/www/umz/v4f0/RpD_UMZ_Methodology_f3.0.pdf (methodology) <http://www.eea.europa.eu/data-and-maps/data/urban-morphological-zones-2006-1> (data)
 - Amsterdam, Aveiro, Genova (Liguria), Bristol, Ljubljana
 - <http://maps.amsterdam.nl/> open Geo Data, specifically for Amsterdam
 - Urban Atlas
The Urban Atlas is providing pan-European comparable land use and land cover data for Large Urban Zones with more than 100.000 inhabitants as defined by the Urban Audit. The GIS data can be downloaded together with a map for each urban area covered and a report with the metadata <http://www.eea.europa.eu/data-and-maps/data/urban-atlas#tab-gis-data>
 - Liguria Land use sc. 1:10000 - ed. 2012 http://geoportale.regione.liguria.it/geoportal/catalog/search/resource/details.page?uuid=r_liguri:D.1577.2013-03-21
- Input and output and location of economic sectors within the city and within the region. Specifically, for Amsterdam a myriad of map-data exists via the open geo Data Portal <http://maps.amsterdam.nl/>. Note that for this specific example, obviously only a link to the original source will be included in the ClairCity Data Portal.
- Consumption: data on consumptions goods and services consumed (survey).
- Socio-economic data: Census data; population, employment, economy (e.g. sectors), household types etc
- Transport: use of transport modes (modal share), public transport data (e.g. use of PT), data from transport models, individual/household-level data on purpose and mode (surveys on travel commuting).
- Housing/energy: energy consumption data?, use of different energy technologies, residential energy consumption (survey).
- Eurostat urban audit data: <http://ec.europa.eu/eurostat/web/cities/data/database>
- Air quality: measurement data (if available). Access to the online data that the cities have on met and AQ. This is essential for the modelling and could be difficult to arrange because of data formats. Examples:
 - Aveiro: <http://gualar.apambiente.pt/index.php?page=2&zona=&year=2016&month=6&day=03>

- Liguria: http://www.cartografiarl.regione.liguria.it/SiraQualAria/script/Pub2AccessoDatiAria.asp?Tipo=CinqueAnni&_ga=1.30700854.847368869.1464961130
- Meteo data. Examples:
 - Aveiro: http://climetua.fis.ua.pt/legacy/main/current_monitor/cesamet.htm
 - Liguria: http://93.62.155.214/~omirl/WEB/mappa_sensori.html?parametro=dir&zoomStart=9&lat=44.22748846630169&lon=8.886315917968773&b=yes
- Emission inventory data (if available). Example for Liguria: regional emission inventory on municipal level disaggregated on grid 1km x 1km. Data by main point sources, the main roads, ports, etc. the inventory follow the SNAP classification from UNECE/EEA task force3 (about 250 elementary activities).
- (Intermediate) model output data: large data dump, unprocessed, high volumes. A key element of the ClairCity Data Portal is also to manage the flow of data between different modules of the modelling toolset. Description of these datasets is not possible at the early stage of the project. In any case, typical dataset are large data dump-files (.txt, .csv, .mdb,...). Partners involved need to upload and download these datasets easily as modules are managed by different project partners.
- Health data, e.g. from such sources as Eionet: The European environment information and observation network (Eionet) aims to provide timely and quality-assured data, information and expertise for assessing the state of the environment in Europe and the pressures acting upon it. This enables policy-makers to decide on appropriate measures for protecting the environment at national and European level and to monitor the effectiveness of policies and measures implemented. <https://cdr.eionet.europa.eu/>

This list of data sources will be expanded in the course of the project. We will explore opportunities for exchanging information of (open) data with the other Horizon 2020 projects iSCAPE and ICARUS.

5.2. Data generation

The previous section focuses on external data to be used in the project. The project itself will also generate data that will be managed (particularly under WP2, WP4 and WP6). The Data Portal will be the central “collection point” for these data also.

The ClairCity work programme is comprised of seven work packages designed to meet the project aims and objectives. WPs are demarcated for project management purposes but in practice there will be significant interaction between them.

The following provides a short description of the key data generation activities under each work package.

- WP1 – presentations from Consortium Meetings, proceedings of such meetings**
- WP2 – Evaluation of engagement activities**
- WP3 – Micro-data on behavior**
- WP4 – Survey outputs, game and app data**
- WP5 – Modeling output (emissions, air quality maps etc)**
- WP6 – Interview responses**
- WP7 - Quantitative scenario description**

Data generated in the ClairCity project will be collected in accordance the EU General Data Protection Regulation (GDPR) as described in the D8.1 and D8.2 ClairCity Ethics Framework.

5.3. Data formats

The format of the data might vary per domain. The ClairCity Data Portal will support generic (blob) data that can be accessed as a whole as well as more structured data formats which can be accessed via an API. The structured formats to be supported will be defined in the project, based on an analysis of formats to be expected and usefulness for the project. Some data will be stored in a cloud facility wherever useful and feasible, while for datasets available via third party open data access points, a link and dataset description is sufficient.

5.4. Quality of data

In general, quality control (QC) procedures include generic quality checks related to calculations, data processing, completeness, and documentation that are applicable to all data sources. This section lists the QC checks a compiler should use routinely throughout the preparation of the inventory. It is good practice that these checks are applied irrespective of the type of data used.

Figure 1 represents the QC approach from the EMEP emission inventory guidebook (2013). Although we will work with different types of data also, the approach as set out in the guidebook is applicable to the ClairCity project as well:

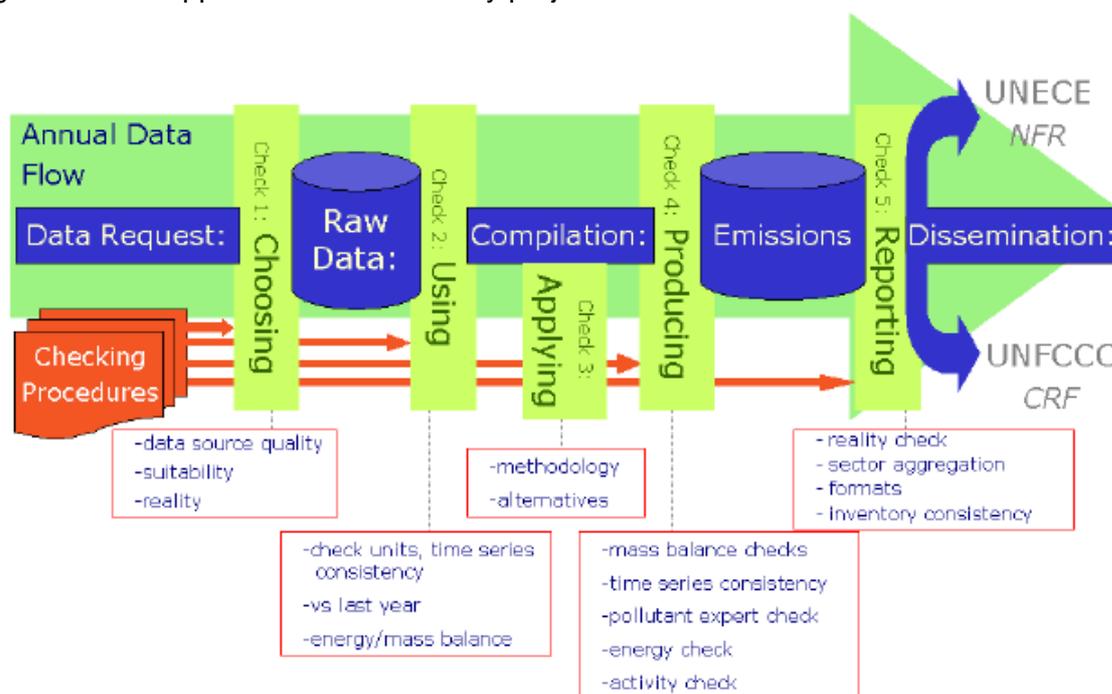


Figure 1: QC flow chart for emission inventories (EMEP)

It is good practice to discriminate between input data, the conversion algorithm of a calculation and the output. Not only does the output need to be recorded, but also the input, the conversion algorithm, and how this algorithm accesses the input. Such an approach allows for intrinsic documentation of the work, and for easy understanding of the calculation procedure. The QC checks therefore cover three areas:

- o choosing data: what information and data is used? Selection and import of data from data providers into a raw data set;

- using data: how is this information used in the modelling? Compilation of the emissions inventory: conversions and calculations using the raw data to build the emissions database;
- reporting outputs: how are the emissions data included in the inventory formats and reports? Have any errors or mistakes made during this process?

It is a prerequisite that all calculations leading to modeling estimates should be fully reproducible. Adequate documentation and archiving of the inventory compilation process is therefore crucial. Obviously, it is good practice that any errors found during the QC checks are repaired.

Specifically on QC with regards to the input selection:

In a typical modelling process the compiler searches for data on both activity rates, emission factors and other parameters for certain source categories, and then uses these in the model as input. Quality controlling the input data collection process aims at ensuring that the data used in the model are traceable and appropriate. While this is elaborated on in the EMEP guidebook specifically for emission inventories, it is also applicable to any data used in the ClairCity modeling datasets. As such, it is good practice to perform the following QC checks:

- Where do input data come from?
- check whether the input data are properly referenced,
- check the availability of the referenced material,
- confirm that bibliographical data references are properly cited;
- what input data are used?
- check that assumptions and criteria for the selection of activity data, emission factors, and other estimation parameters are documented,
- cross-check descriptions of input activity data, emission factors and other estimation parameters with information on categories and ensure that these are properly interpreted and used,
- check that parameters and units are correctly recorded and that appropriate conversion factors are used,
- check that units are properly labelled in calculation sheets,
- check for consistency in data between categories:
 - identify parameters (e.g., activity data, constants) that are common to multiple categories, and
 - confirm that there is consistency in the values used for these parameters in the emission calculations,
- Check time series consistency:
 - identify temporal inconsistency in time series input data for each category,
 - take into account the effects of mitigation.

The Data Portal will offer the ability to add some additional remarks on quality aspects of a dataset based on the description above. On data upload there will be some hints on how to use this mechanism. Since only partners in ClairCity project have the right to manage datasets, data spoofing is not an issue.

5.5. Usefulness of data and related data

The main use case for the Data Portal is not the dissemination of the data to the general public, but the organization of the data among project partners and the use of the data in the ClairCity products to be disseminated to the cities. However, by using the public mode of the Data Portal, data that deserves wider sharing can be used by other researchers, institutions and products as well.

6. Standards and metadata

The Data Portal will be based on the internationally recognized CKAN software (<http://ckan.org>) and its metadata management will be based on the DCAT standard (<https://www.w3.org/TR/vocab-dcat/>). The ClairCity specific extensions to metadata will follow internationally agreed best practices.

Metadata describes the dataset in detail, providing information such as tabular data headings and units, which are essential for using the data. Provide information on where detailed-level metadata will be stored and who will create it.

The following information (where applicable) needs to be presented for each dataset:

- Experimental design / sampling regime (in case of surveys)
- Collection / generation / transformation methods
- Fieldwork and / or laboratory instrumentation (in case of experiments)
- Calibration steps and values
- Nature and units of recorded values
- Analytical methods
- Quality control
- Format of stored data
- Miscellaneous

7. Data sharing

7.1. Data access

As mentioned earlier, data in the Data Portal will be open to the consortium by default. In addition to this the Data Portal will contain other levels of access on datasets, such as fully public, open to consortium (default), open to a single partner in the consortium. Further details will be described in consecutive versions of this DMP.

Compliant to Article 29.3 of the Grant Agreement, the Data portal, which will play a central role in the data management throughout the project, will be designed in such a way that third party access of research data generated may be enabled for validation purposes. In short, the Data Portal is setup so that that a customized level of access can be granted to any third party to project metadata and, external or newly generated data.

The Data Portal will contain a guideline how to refer to ClairCity data from scientific publications. On data upload, one of the available licensing forms have to be chosen. Datasets that rely on other publicly available data obviously have to refer to these base data and have to respect the license of these other datasets.

Data in the ClairCity Data Portal may be used for commercial exploitation if the license that applies to that particular dataset is respected and the ClairCity project is quoted the way the Data Portal advises.

There is no embargo period.

7.2. Data sharing mechanisms

The Data Portal has been designed to support the ClairCity way of working as close as possible, taking into consideration the technical and organizational possibilities. As previously discussed some datasets may not be made publically available due to licensing or commercial sensitivity but data and project results which can be made publically available. Over the final two years, ClairCity will systematically release data bundles linked to the completion dates for each case study city-region (this may include baseline emissions input data, emissions data, air quality, exposure and health data/maps, MLW/Delphi/Game data, scenarios etc).

Anticipated data release dates for each case study dataset includes:

- City 1 relevant data and results (anticipated release date April 2019)
- City 2 relevant data and results (anticipated release date Sept 2019)
- City 3 relevant data and results (anticipated release date Dec 2019)
- City 4 relevant data and results (anticipated release date Jan 2020)
- City 5 relevant data and results (anticipated release date Feb 2020)
- City 6 relevant data and results (anticipated release date March 2020)
- All remaining relevant EU and generic datasets (April 2020)

As exemplars, two typical use cases of the Data Portal in the ClairCity project are described:

Case 1:

Partner A is producing model output to be used as input for a model developed and operated by partner B. For a specific case study, the city partner (C) will need to provide some inputs to both partners. In this setup, use of Data Portal could be:

1. Partner C identifies, cleans and uploads a specific dataset for the case city on the Data Portal cloud repository OR links to the original open data platform where the data can be downloaded. This partner executes quality checks, tags and adds meta-data to the specific dataset.
2. Partner A downloads the dataset when needed to execute model runs. The model specifications are adapted to the specific data format the dataset is available from the Data Portal, in line with the meta-data as described in the previous step.
3. Partner A executes model runs (using own infrastructure) and produces a large data dump file as an intermedia model-output, to be used by other modules. The data dump is uploaded to the Data Portal, including meta-data (format, dimensions, version etc)
4. Partner B downloads the intermediate model output from the previous step. Subsequent module specifications are adapted to the specific data format, in line with the meta-data as described in the previous step.

Case 2:

Partner A is developing calibrated model algorithms, using survey data, including privacy sensitive data, collected in the ClairCity project by partner B

1. Partner B designs and executes the survey and processes the results on own infrastructure, adapted to storing privacy-sensitive data. This stage does not involve the use of the Data Portal
2. Partner B anonymizes the data and/or performs various data manipulations to reduce the privacy sensitivity and uploads the processed,
3. Partner 1 downloads the data from the portal and performs analysis and uses derived conclusions to calibrate (for example behavioral) model

These two cases serve as an example in what way the Data Portal will be used in the context of modeling in ClairCity.

Data and metadata from the ClairCity Data Portal can be accessed using a modern web browser. Depending on the format and size of a dataset, viewing or processing such data may require dedicated software tools.

These data sharing mechanisms focus on large data sets (external or generated in the project). Apart from datasets, all internal reports, templates, and useful documentation are stored internally using a password protected Sharepoint file storage system (hosted at UWE, Bristol) as outlined in the D2.1 ClairCity Communication Plan. This allows ClairCity consortium members to access relevant information and search through past meeting minutes, to aid Project Management, coordination and dissemination.

7.3. Data sharing limitations

There are no other limitations to sharing data from the ClairCity Data Portal other than what is mentioned at the section on Data Access above.

7.4. Data security

Privacy sensitive data such as survey results may be collected in the ClairCity project. Privacy sensitive data will be collected in accordance with the EU GDPR as discussed in the D8.1 and D8.2 ClairCity Ethics Framework. All research participants will be made aware of what data is being collected, and what will be done with it via the Participant Information Sheets/Notices and Consent Forms. All personal data will only be stored within the country within which it was collected, and any partner dealing with privacy sensitive data on their local research sites is expected to be in full compliance with Article 39 of the GA and GDPR on the protection of individuals with regard to the processing of personal data.

Data will be protected to ensure no sensitive data is released that can be linked to specific individuals or entities, the Statistical Disclosure Control (SDC) software T-ARGUS will be used on all collected and generated tables. Any data that can be identified to individuals or entities will be stored separately to their research responses to ensure confidentiality. Personal data will not be shared among project partners and will not be allowed to be uploaded to the ClairCity Data Portal; only fully anonymised research responses will be uploaded. These data are described in the Annex at the end of this document.

Partners dealing with privacy sensitive data are expected to have procedures and infrastructure in place to adhere to the following:

- Personal data identifying participants will be kept on password protected computers at the local research sites. Only researchers trained in data management will have access
- Personal data will not be shared publicly, or across national borders amongst the ClairCity Consortium, in accordance with the EU GDPR
- Data will be anonymised using a coding system before being uploaded to the Data Portal (i.e. each participant gets a number, and the number code file stays on the local site, while the actual data can be shared with the number)
- The Data Portal will only store anonymised data which cannot be linked to individual participants.

8. Archiving and preservation of data

8.1. Data storage

Across ClairCity WPs several data is being collected, processed and stored. An overview of partners involved in data collection and how everyone is locally storing such data is provided in the Annex. Overall city buddies and research institutions collect most of the data, as opposed to the City (and Regional) partners.

The total volume of data to be contained in the Data Portal is not yet known. The limits on size of a dataset to be uploaded and on the total amount of data in the Data Portal will be able to be calculated as the project progresses. They will be scaled to what is practically and financially feasible.

8.2. Backing up of data

The Data Portal will have a backup schedule according to the regular backup procedures of Statistics Netherlands (CBS).

8.3. Maintaining and updating data

The ClairCity project partners are responsible for maintaining the datasets they upload to the Data Portal and updating them if necessary.

8.4. Costs of data storage

The costs of setting up the Data Portal are contained in WP5. Costs for hosting the Data Portal may depend on the amount of data to be managed and the connectivity required. If these costs exceed what is foreseen within the project budget of the partner managing the Data Portal a solution will be found within the project.

8.5. Archiving of data after project life

For long-term preservation (archiving) the following procedure will be executed: at the end of the ClairCity project the Data Portal will be frozen. This means that no datasets can be uploaded or changed any more, but that they are **still available for inspection and download**. At the end of the project datasets that are sufficiently valuable for long-term preservation will be uploaded to one of the data archives from the EU. The Data Portal will remain in the frozen status after project life for a period, conforming to the regulations referred to in the DoA. After that it will be put offline, unless another financing project is found to keep it running as outlined in the D2.9 ClairCity Impact and Innovation Plan.

9. Flow and responsibilities

The flow of data for the data portal collected/generated in the project is as follows (data collected by other means are explained in Annex):

Table 1: overview of flow and responsibilities of data management in ClairCity

Activity	Who is responsible?	Protocol or documentation details
Identification of data from 3 rd parties	Any partner	/

Activity	Who is responsible?	Protocol or documentation details
Data capture from field / lab work / modelling output	Any partner	See Section 5.2
Verification of privacy sensitivity of the data set	WP lead with CBS: Data Portal manager	See Section 7.4
Data uploading to database / storage	Any partner / with CBS: Data Portal manager	Data Portal procedure (to be developed)
Back-up of data	CBS	(cloud storage)
Data processing	Any partner	/
Data validation / quality control	WP lead (depending on source, WP3, 4, 5)	See Section 5.4
Metadata creation	Any partner / verified with CBS: Data Portal manager	See Section 6
Documentation of datasets	CBS (Data Portal catalogue)	See Section 5
Establishment of access arrangements / IPR to the data	WP lead (depending on source, WP3, 4, 5) with CBS	See Section 7
Archiving and clear up of all remaining project data	WP lead (depending on source, WP3, 4, 5) with CBS	See Section 8
Update of data management plan	TML-CBS	See Section 10

10. Data Management Plan Maintenance

The DMP will be reviewed and updated, if necessary, by the Data Manager (CBS-TML) and the Quality Control Manager with the support of the WP-leads on the following dates:

- End of April 2017 (end of Year 1 – v3.0)
- End of May 2018 (end of Year 2 – v4.0)
- End of May 2019 (end of Year 3 – v5.0)
- End of April 2020 (end of Year 4 and project – v6.0)

Annex - Data collection and storage across the project, per partner

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
UWE Bristol (City buddy)	2.2. Evaluation – live events and social media (project lead)	General public (aged 18+) Social media in the public domain (including children over 13 years of age)	Personal opinions (no personal identifying characteristics)	Only anonymous opinions, or comments on social media in the public domain on Data Portal	N	N - as once submitted it is anonymous as per the Ethics Framework
	2.2 Evaluation – project staff (project lead)	Project staff members (aged 18+)	Personal data and opinions	Data is only stored locally by UWE and not uploaded to the Data Portal as it contains project sensitive data	Y – demographic information and contact details	Y – as per ethics framework
	4.1 Delphi surveys Delphi workshops (project lead)	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.3 Game (project lead)	Adults aged over 18 years Children aged over 13 years if they receive parental consent on social media	Personal data, game play and opinions	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information	Y – as per ethics framework
	4.4.3 Videos with the elderly	General public (18+) including active older adults	Publicly available videos showing participants' opinions	Videos are stored on YouTube	Y contact details are kept locally with the consent forms	Y – as per ethics framework
	4.4.1 Mutual Learning Workshop	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the	Y – demographic information and contact details	Y – as per ethics framework

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
				Data Portal and Sharepoint.		
	4.4.2 Schools competition	Children aged 13-16 years in schools, with their teachers	Online schools competition with no identifiable information given by the children.	The online tool holds the anonymized competition results and is stored by REC	Y – adult teacher contact details are kept locally with the consent forms	Y – as per ethics framework
	6.1 Interviews with policymakers	Policymakers (aged 18+)	Personal data and opinions	Data is only stored locally by UWE and not uploaded to the Data Portal as it contains project sensitive data. Data is anonymised and used for general analysis.	Y – demographic information and contact details	Y – as per ethics framework
Bristol City Council (City Partner)	NA	NA	NA	NA	NA	NA
Trinomics (City buddy)	4.1 Delphi surveys Delphi workshops	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.3 Videos with the elderly	General public (18+) including active older adults	Publicly available videos showing participants' opinions	Videos are stored on YouTube	Y – contact details are kept locally with the consent forms	Y – as per ethics framework
	4.4.1 Mutual Learning Workshop	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.2 Schools competition	Children aged 13-16 years in schools, with their teachers	Online schools competition with no identifiable	The online tool holds the anonymized competition	Y – adult teacher contact details are kept locally	Y – as per ethics framework

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
			information given by the children.	results and is stored by REC	with the consent forms	
	6.1 Interviews with policymakers (project lead)	Policymakers (aged 18+)	Personal data and opinions	Data is only stored locally by Trinomics and not uploaded to the Data Portal as it contains project sensitive data. Data is anonymised and used for general analysis.	Y – demographic information and contact details	Y – as per ethics framework
	6.5 Policy workshop (project lead)	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
GGD Amsterdam (City Partner)	4.1 Delphi surveys Delphi workshops	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	6.5 Policy workshop	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
Techne Consulting (City buddy)	2.2. Evaluation – live events and social media	General public (aged 18+) Social media in the public domain	Personal opinions (anonymised data)	Only anonymous opinions, or comments on social media	N	No as once submitted it is anonymous as per the Ethics Framework

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
	2.2 Evaluation – project staff	Project staff members (aged 18+)	Personal data and opinions	Data is only stored locally by Techne	Y – demographic information and contact details	Y – as per ethics framework
	4.1 Delphi surveys Delphi workshops	Adults aged over 16 years Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally The data is stored in anonymised way and uploaded to the Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4 Videos	General public (18+) including active older adults General public aged under 18 years if they receive parental consent		Videos are stored on YouTube	Y contact details are kept locally with the consent forms	Y – as per ethics framework
	4.4.1 Mutual Learning Workshop	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored locally. The data is stored in anonymised way and uploaded to the Sharepoint.	Y – demographic information and contact details with the consent forms	Y – as per ethics framework
	4.4.2 Schools competition	Children aged 13-16 years in schools, with their teachers	Online schools competition with no identifiable information given by the children	The online tool holds the anonymized competition results and is stored by REC	Y – adult teacher contact details are kept locally with the consent forms	Y – as per ethics framework
	6 Interviews with policymakers	Policymakers (aged 18+)	Personal data and opinions	Data is only stored locally	Y – demographic information and contact details	Y – as per ethics framework
Regione Liguria (City partner)	NA	NA	NA	NA	NA	NA
UAVR (City buddy)	4.1 Delphi surveys Delphi workshops	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.3 Videos with the elderly	General public (18+) including active older adults	Publicly available videos showing	Videos are stored on YouTube	Y contact details are kept locally with the consent forms	Y – as per ethics framework

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
			participants' opinions			
	4.4.1 Mutual Learning Workshop	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.2 Schools competition	Children aged 13-16 years in schools, with their teachers	Online schools competition with no identifiable information given by the children.	The online tool holds the anonymized competition results and is stored by REC	Y – adult teacher contact details are kept locally with the consent forms	Y – as per ethics framework
	6 Interviews with policymakers	Policymakers (aged 18+)	Personal data and opinions	Data is only stored locally by UAVR and not uploaded to the Data Portal as it contains project sensitive data	Y – demographic information and contact details	Y – as per ethics framework
	3.2 & 3.3	General public	Publicly available aggregated census data	Stored locally	N	N
	3.2 & 3.3	General public	Publicly available population data	Stored locally	N	N
	3.2 & 3.3	Private/ free dataset from national meteorological office + Physics Department of University of Aveiro	Time series dataset in an hourly basis	Stored locally and shared with WP5 partners	N	N
	3.2 & 3.3	Synthetic dataset (city buddy)	Synthetic dataset: population, household, dwelling, location, energy consumption and emissions	Stored locally and shared with WP5 partners	N	Y – as per ethics framework
	3.2 & 3.3	Region employment rate/ CIRA open public data	Annual share information	Stored locally and shared with WP4 and WP5 partners	N	N
	3.2 & 3.3	Region age groups	Annual share information	Stored locally and shared with	N	N

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
				WP4 and WP5 partners		
CIRA / Aveiro Region (City Partner)	4.1 Delphi surveys Delphi workshops	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored by the University of Aveiro. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.3 Game (project lead)	Adults aged over 18 years Children aged over 13 years if they receive parental consent on social media	Personal data, game play and opinions	Personal information is stored by the project lead The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information	Y – as per ethics framework
	4.4.3 Videos with the elderly	General public (18+) including active older adults	Publicly available videos showing participants' opinions	Videos are stored on YouTube	Y contact details are kept locally by the University of Aveiro with the consent forms	Y – as per ethics framework
	4.4.1 Mutual Learning Workshop	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored by the University of Aveiro. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.2 Schools competition	Children aged 13-16 years in schools, with their teachers	Online schools competition with no identifiable information given by the children.	The online tool holds the anonymized competition results and is stored by REC	Y – adult teacher contact details are kept locally with the consent forms	Y – as per ethics framework
	6 Interviews with policymakers (project lead)	Policymakers (aged 18+)	Personal data and opinions	Data is only stored locally by UWE and not uploaded to the Data Portal as it	Y – demographic information and contact details	Y – as per ethics framework

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
				contains project sensitive data		
REC (City buddy)	4.1 Delphi surveys Delphi workshops	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.1 Mutual Learning Workshop (project lead)	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.2 Schools competition (project lead)	Children aged 13-16 years in schools, with their teachers	Online schools competition with no identifiable information given by the children.	The online tool holds the anonymized competition results and is stored by REC	Y – adult teacher contact details are kept locally with the consent forms	Y – as per ethics framework
	4.4.3.Videos with elderly (city buddy)	General public (18+) including active older adults	Publicly available videos showing participants' opinions	Videos are stored on YouTube and locally	Y - contact details are kept locally with the consent forms	Y – as per ethics framework
Sosnowiec City Council (City partner)	4.1 Delphi surveys Delphi workshops	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.3 Videos with the elderly	General public (18+) including active older adults	Publicly available videos showing participants' opinions	Videos are stored on YouTube	Y contact details are kept locally with the consent forms	Y – as per ethics framework
	4.4.1 Mutual Learning Workshop	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the	Y – demographic information and contact details	Y – as per ethics framework

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
				Data Portal and Sharepoint.		
Ljubljana (City Partner)	4.1 Delphi surveys Delphi workshops	Adults aged over 18 years	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
	4.4.3 Videos with the elderly	General public (18+) including active older adults	Publicly available videos showing participants' opinions	Videos are stored on YouTube	Y contact details are kept locally with the consent forms	Y – as per ethics framework
	4.4.1 Mutual Learning Workshop	General public over 18 years old	Personal data and opinions Contact details	Personal information is stored locally. The data is resaved so that only non-identifiable data is uploaded to the Data Portal and Sharepoint.	Y – demographic information and contact details	Y – as per ethics framework
NILU	4.3 App (Project lead)	Adults aged +16 (consensus in app)	GPS position and time collected within user selected areas. User might choose to share nickname. Data is delayed minimum 6 hours before uploaded to server	No personal information is stored on server except gps position, time and nickname if the user chooses to share. User can delete all data with one click from app. Data stored on server at NILU. Only statistics and completely anonymous data will be available for research.	Y – gps position and time	Y – as per ethics framework
PBL	3.2 & 3.3	General public	Publicly available aggregate census data	Stored locally	N	N
	3.2 & 3.3	General public	Publicly available survey data, fully anonymised	Processed, stored locally and on Data Portal	N	N
	3.2 & 3.3	General public	Publicly available population forecasts	Stored locally	N	N

Partner	WP/Activities where data is collected	Whose data? (particular focus if vulnerable groups e.g. elderly or children)	Type of Data collected	Processing (e.g. organising/ combining/ storage/ erasing)	Is this sensitive data according to the definition of the GDPR? Y/N (if Y, explain)	Withdrawal procedure in place Y/N
DTU	3.2 & 3.3	Open dataset from national meteorology office: 1. Amsterdam: KNMI 2. Bristol: MetOffice UK	Time series dataset Hourly resolution	Cleansing Transformation Integration	N	N – open public data
	3.2 & 3.3	Open data	Discrete/categorical data	Integration	N	N – open public data
	3.2 & 3.3	Synthetic dataset Owner: ClairCity partner PBL	Panel synthetic dataset: 1. Population 2. Household 3. Dwelling 4. Location 5. Energy consumption 6. Emissions	Transformation Integration Reduction	N	Y – as per ethics framework
	3.2 & 3.3	Region/city's employment rate Bristol: Open public data Owner: City of Bristol Amsterdam: Open public data Owner: ClairCity partner CBS, The Netherlands	Annual share information	Integration	N	N – open public data
	3.2 & 3.3	Region/city's age groups	Annual share information	Integration	N	N – open public data
TML	5.1-5.2-5.3	Any data related to air quality calculations for the 5 case cities involved.	Publically available data	Clair city dataportal, either directly or links to external download centres	N	N/A
CBS	5	Statistics Netherlands manages datasets uploaded by the other partners to the data portal as described in the DMP				