



Innovation Action
H2020-LC-SC3-SCC-1-2018

D5.5 - Data sets: Requirements, collection and protection

WP5; Task 5.2

November 2021 [M36]

Author(s): Peter Hut (CGI), Alvaro Corredera Cano (CAR)



www.makingcity.eu



[@MakingCity_EU](https://twitter.com/MakingCity_EU)



This project received funding from the Horizon 2020 programme under the grant agreement n°824418.

Disclaimer

The content of this deliverable reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Copyright notice

©2019 MAKING-CITY Consortium Partners. All rights reserved. MAKING-CITY is a HORIZON2020 Project supported by the European Commission under contract No. 824418. For more information on the project, its partners and contributors, please see the MAKING-CITY website (www.makingcity.eu/). You are permitted to copy and distribute verbatim copies of this document, containing this copyright notice, but modifying this document is not allowed. All contents are reserved by default and may not be disclosed to third parties without the written consent of the MAKING-CITY partners, except as mandated by the European Commission contract, for reviewing and dissemination purposes. All trademarks and other rights on third party products mentioned in this document are acknowledged and owned by the respective holders. The information contained in this document represents the views of MAKING-CITY members as of the date they are published. The MAKING-CITY consortium does not guarantee that any information contained herein is e-free, or up-to-date, nor makes warranties, express, implied, or statutory, by publishing this document.

Document Information

Grant agreement	824418
Project title	Energy efficient pathway for the city transformation: enabling a positive future
Project acronym	MAKING-CITY
Project coordinator	Cecilia Sanz-Montalvillo (cecsan@cartif.es)- Fundación CARTIF
Project duration	1 st December 2018 – 30 th November 2023 (60 Months)
Related work package	WP 5 – Evaluation framework and social innovation
Related task(s)	Task 5.2
Lead organisation	09-CGI & 20-VTT
Contributing partners	01-CAR, 03-GRO, 09-CGI, 10-SB, 13-OUK, 20-VTT
Due date of deliverable	30 th November 2021
Actual submission date	1 st December 2021
Dissemination level	Public

Date	Version	Submitted by	Reviewed by	Comments
05/10/2021	2.0.1	CGI	Peter Hut	Copy of submitted D5.18.
19/10/2021	2.1.0	CGI	Peter Hut	Added additional data sets and prepared for other contributors.
16/11/2021	2.2.0	CGI	Peter Hut	Updated dataset descriptions, ready for preliminary peer review.
30/11/2021	2.2.1	VTT	Jussi Rönty	Verify data set requirements of the Oulu monitoring program
01/12/2021	3.0.0	CAR	Cecilia Sanz-Montalvillo	Finalized

Table of content

Executive Summary.....	7
1 Introduction.....	8
1.1 Purpose and target group.....	8
1.2 Contribution partners.....	8
1.3 Relation to other activities in the project.....	9
2 Urban Data Platform.....	10
2.1 Evaluation objectives in WP5.....	11
3 Data Requirements.....	13
3.1 Data Governance.....	13
3.2 Open data, public data, confidential data.....	14
3.3 Data access.....	14
3.4 SCIS.....	14
3.5 Quality of Open Data.....	15
3.5.1 ★ One-Star Open Data.....	17
3.5.2 ★★ Two-Star Open Data.....	17
3.5.3 ★★★ Three-Star Open Data.....	17
3.5.4 ★★★★★ Four-Star Open Data.....	18
3.5.5 ★★★★★★ Five-Star Open Data.....	18
3.6 Data Set Definition.....	19
4 Internal Confidential Data.....	20
4.1 Data Set: Building Energy Consumption Measurements.....	20
4.2 Data Set: Building Energy Production Measurements.....	20
4.3 Data Set: Building Emissions.....	21
4.4 Data Set: Building Net Energy Consumption.....	21
4.5 Data Set: Primary Energy Balance.....	22
4.6 Data Set: Buildings.....	23
4.7 Data Set: EV Charging Logs.....	24
5 Published Open Data.....	25
5.1 Data Set: Project Indicators – Energy.....	25
5.2 Data Set: Project Indicators – EV Charging.....	26
5.3 Data Set: Project Indicators – Investment.....	26
5.4 Data Set: Project Indicators – Social.....	27
6 External Data Sets.....	28
6.1 Data Set: Building Location Information.....	28
6.2 Data Set: District Energy Consumption.....	28
6.3 Data Set: City Population Information.....	29
6.4 Data Set: Energy Mix.....	29
7 Conclusions.....	31
Bibliography.....	32

List of figures

Figure 1: Urban Data Platform context.....	11
Figure 2: Definition of KPIs, data requirements and monitoring in tasks 5.1 – 5.3.	12
Figure 3: Coordination among lighthouse cities and other initiatives to define useful and usable information as open data for the urban platforms.	12
Figure 4: Universe of Open Data.....	15
Figure 5: Linked Data	16
Figure 6: Open License.....	17
Figure 7: Machine readable	17
Figure 8: Non-proprietary	17
Figure 9: Linked data.....	18
Figure 10: Interlinked data.....	18

List of tables

Table 1: Contribution of partners	8
Table 2: Relation to other activities in the project.....	9
Table 3 Template used for data set description including data requirements.....	19

Abbreviations and Acronyms

Acronym	Description
CPMS	Charging Point Management System
GDPR	General Data Protection Regulation
KPI	Key Performance Indicator
PED	Positive Energy District
PII	Personal Identifying Information
POPD	Protection of Personal Data
SCIS	Smart Cities Information System

Executive Summary

MAKING-CITY project is going to develop a large scale demonstration of three Positive Energy Districts in two European cities, Groningen (Netherlands) and Oulu (Finland) where a rigorous monitoring and evaluation program will be deployed, with special attention to data collection, regulation (GDPR), evaluation framework and integration in a monitoring platform.

WP5 aims to monitor and evaluate the effectiveness of the project actions and interventions, compared to the initial situation, initial objectives and expected results. Robust monitoring and evaluation protocols will be developed and implemented, including a full methodology for the monitoring and evaluation of the project actions and interventions that will allow the introduction of future data after the end of the project. Based on the selection of KPIs in Task 5.1 the data requirements for those indicators are needed. The result is a description of all data sets that will be needed as input for the Urban Data Platform to calculate the KPIs.

1 Introduction

1.1 Purpose and target group

This report constitutes Deliverable *D5.18 Data sets: Requirements, collection and protection (Initial Version)*, which is based on the outcome of the “Task 5.2: Definition of the data sets and requirements (data collection, GDPR)”.

The objectives of the deliverable are:

- Defining requirements for data that needs to be collected
- Identifying existing common data sets that are used in the calculation of KPIs

The contents in this report are developed in parallel with the selection and definition of the smart city KPIs (project task 5.1). That results in two deliverables:

1. D5.1 – City level indicators
2. D5.2 – Project level indicators

They include the full description of all the indicators selected for MAKING-CITY, including their data requirements. This deliverable focuses on the analysis of the available data sources for the implementation of those indicators.

Chapter 2 describes the Urban Data Platform as the platform to store the internal data sets and perform the calculations resulting in the Project indicators. Chapter 3 describes the general data requirements for data used in MAKING-CITY and data governance concerns related to the stored and retrieved data. Finally, chapter 4, 5 and 6 describe the specific data required for the selected KPIs. Chapter 4 describes the requirements for data internally stored in the Urban Data Platform and chapter 5 describes data published as Open Data. Finally, chapter 6 describes other data used for calculation of KPIs, but not stored in the platform.

1.2 Contribution partners

The following Table 1 depicts the main contributions from participant partners in the development of this deliverable.

Table 1: Contribution of partners

Partner nº and short name	Contribution
01-CAR	Definition of the Quality Data Requirement.
03-GRO	Suggest relevant open data sources and provide the data governance requirements.
04-TNO	Requirements on data sets for the Groningen Monitoring Programme.
09-CGI	Leading contributor.
10-SB	Provide information about the data measured in Groningen PEDs.
13-OUK	Review

20-VTT	Requirements on data sets for the Oulu Monitoring Programme.
25-DEM	Review

1.3 Relation to other activities in the project

Based on the KPI selection from Task 5.1, this deliverable describes the data requirements for evaluating demonstrations according to the corresponding indicators. The identification of the data sets to be compiled will be based on the previous work by CITYkeys, D2.1 “Definition of data sets” (Bosch, 2015). Available data sources, their reliability, formats, level of confidentiality and data access methods will be analysed by all the partners involved. Privacy and security issues will be tackled considering the new General Data Protection Regulation (GDPR) that will be the main reference when developing (at the early stages of the project) the Data Management Plan of the project. The result will be the description of all data sets that will be needed as input for the Smart City data collection system and their aggregation levels. The data collection and KPI calculation will be done in WP2 (subtask 2.7.2) and WP3 (subtask 3.7.2).

The following Table 2 depicts the main relationship of this deliverable to other activities (or deliverables) developed within the MAKING-CITY Project and that should be considered along with this document for further understanding of its contents.

Table 2: Relation to other activities in the project

Deliverable/ Action n°	Relation
D2.8 / D3.8	<i>Adaptation of Oulu ICT platform and Adaptation of the Groningen ICT platform respectively.</i>
D5.1	City level indicators (T5.1)
D5.2	Project level indicators (T5.1)
D5.7	Oulu Monitoring Programme Main producer of the data sets in this deliverable.
D5.8	Groningen Monitoring Programme Main producer of the data sets in this deliverable.
D5.9	ICT-City Platforms: common open specifications
D5.10	Data collection and KPI calculation
D5.11	Evaluation (city level, project level)
D10.2	<i>POPD Requirement No. 2</i> Describes the methodology that will be followed by partners in charge of collecting personal data for the development of the tasks and interventions.

2 Urban Data Platform

The MAKING-CITY project is going to develop a large-scale demonstration of three Positive Energy Districts in two European cities, Groningen (Netherlands) and Oulu (Finland) where a rigorous monitoring and evaluation program will be deployed. This monitoring program requires an Urban Data Platform to securely store and share collected data about the city. Each city will use a separate Urban Data Platform with the data is accessible via the same open standards.

The impact of the project will be measured directly where possible and these measurements will be stored in the Urban Data Platform. The platform is responsible for securely storing these measurement and enables access for consortium partners to the high quality information they provide. For all evaluation indicators to accurately describe the impact of the project, the data quality is very important. In this report the measurement data sets can be found in chapters 4. Chapter 6 describes the data sets containing the project indicators, which are calculated based on the measurements or on externally obtained data, and published using the Urban Data Platform as Open Data.

Figure 1 gives an overview of the role of an Urban Data Platform and the interfaces it provides for access to the data. More information about the interfaces and open standards can be found in deliverable *D5.9 ICT-City Platforms: common open specifications*.

The indicators in the monitoring program will also require additional data sources that are not measured as part of the MAKING-CITY project. These external data sources are described in chapter 6 External Data Sets but not stored in the Urban Data Platform.

More information about the actual implementation of the platforms in the Lighthouse cities can be found in *D2.8 Adaptation of Oulu ICT platform* and *D3.8 Adaptation of the Groningen ICT platform*.

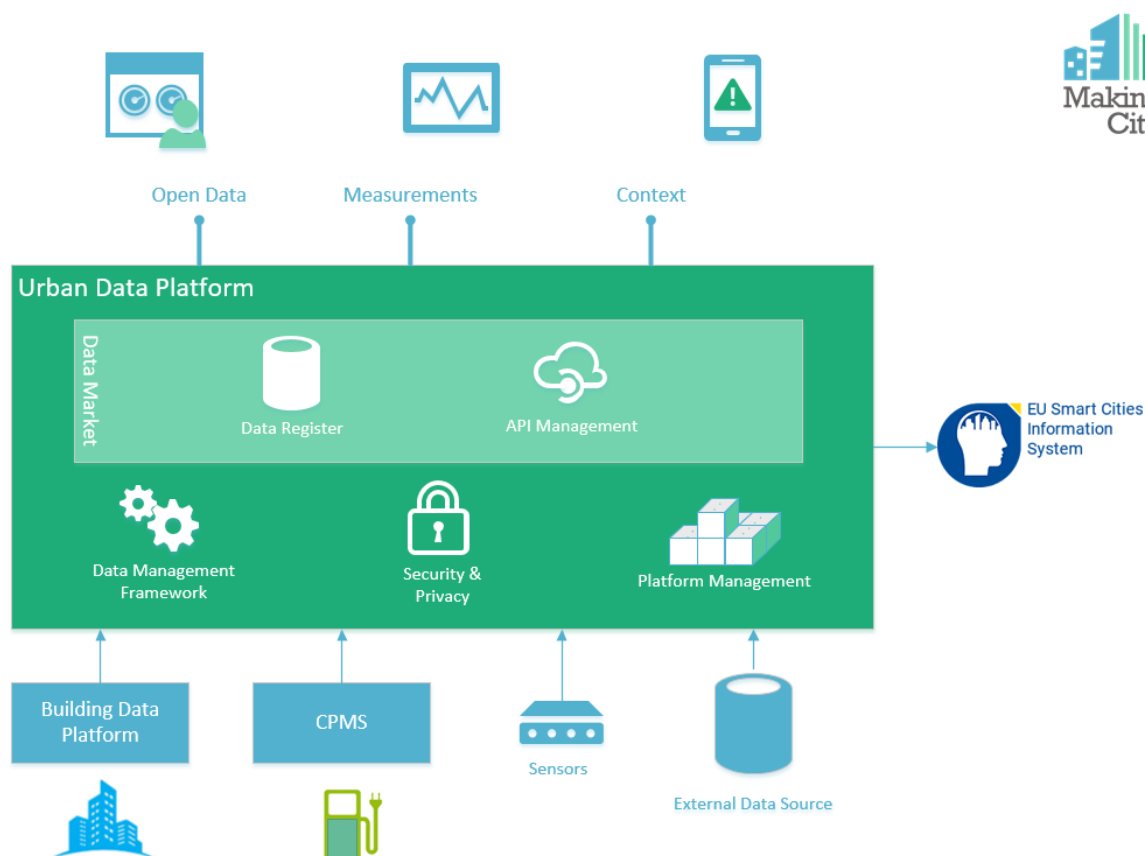


Figure 1: Urban Data Platform context

2.1 Evaluation objectives in WP5

Task 5.1 will result in a definition of the evaluation framework in order to measure and assess the project activities at PED level (demonstration areas) and at city level considering the indicator categories defined by CITYkeys (Smart City Indicators and related methodology), SCIS (Key Performance Indicators Guide) and other relevant reference frameworks.

Within Task 5.2 the data sets and requirements for evaluating demonstrations are defined based on the indicator selections from Task 5.1. The identification of the data sets will also be based on the previous work by CITYkeys and SCIS, and all the information related to these data sets will be included in this report. The actual data collection and KPI calculation will be carried out in WP2-Oulu (subtask 2.7.2) and WP3-Groningen (subtask 3.7.2). Finally, all relevant performance data (i.e. project level KPIs) will be incorporated into the SCIS database. The whole procedure is illustrated in Figure 2.



1. Selecting and defining the city level indicators
2. Defining the baseline situation in the city and calculating the indicator values at the beginning of the project (before the planned city level actions)
3. Monitoring the indicators during the course of the project (following the indicators for the evaluation of progress), and
4. Final calculation of the indicators at the end of the project for the final evaluation and impact assessment.

[illegible]

3 Data Requirements

Since most of the indicators are selected from the CITYkeys' results, the same data requirements are relevant for MAKING-CITY, as well.

First of all, the set of indicators consists of a mixture of quantitative and semi-quantitative indicators. Quantitative indicators are based on quantities obtained using a quantifiable measurement process. Semi-quantitative indicators are based on qualitative information that is then assessed according to a Likert Scale (e.g. 1 = not at all; 2 = poor; 3 = fair; 4 = good; 5 = excellent). For example, a semi-quantitative indicator may provide an assessment that the city government encourages a healthy lifestyle. For the semi-quantitative indicators, data needs to be collected using interviews or an analysis of policy documents. (Bosch, 2015)

Secondly, indicators are defined at two levels: city and project. The project indicators are meant to assess the success of the MAKING-CITY project. This data is collected directly (measured) or from the project office, the project leader and/or others closely involved in the project. Data for the majority of the city indicators can be retrieved from statistical sources within the city administration. Some have been made available as Open Data. However, some of the governance indicators also require a person to gather information. All performance data (i.e. KPIs) will be added to the SCIS database via the Self-Reporting Tool (SCIS-SRT).

The data sets are described based on the current information. The following details about the storage and retrieval of the (external) data sets, per city, are described in this deliverable:

- Data classification – Open Data, public data, confidential
- Data format – Open standard or proprietary format. Examples of open formats are CSV, JSON, or in case of spatial data: WFS, Shapefile (shp)
- Data model – For example FIWARE NGSI, OGC CityGML
- Platform – CKAN, custom, REST API, SCIS
- Available columns/properties/data items – Detailed specification of the exact format

3.1 Data Governance

Not all data is held by city authorities. Many different stakeholders can hold relevant data. The related institutions and bodies can be described by their activities as described in CITYkeys (Bosch, 2015):

- Municipality and governance bodies
 - local and city councils
 - municipal services
- Building and other asset managers
- Network managers, such as energy providers, traffic managers
- Other companies
- Scientific and other parties such as meteorological institutes
- Citizens

The consortium partners have agreed that data stored in the Urban Data Platform collected for the MAKING-CITY project is owned by the municipality. The owner of the data will ensure that Privacy Impact Analysis is performed. Results of the analysis and other considerations for specific datasets are described in this report. More specifically any privacy related data sets will have a limited lifetime and are collected and used only for a specific purpose. Finally, the Urban Data Platforms storing the data

should correctly handle PII data. Detailed procedures and agreements for data collected as part of the Monitoring Programmes is described in *D5.7 Oulu Monitoring Programme* and *D5.8 Groningen Monitoring Programme*. More general procedures of consortium partners handling personal data can be found in *D10.2 POPD Requirement No. 2*.

3.2 Open data, public data, confidential data

A data set is defined as Open Data, when it fulfils following three fundamentals:

1. data is available and is in readable form
2. data is published with a license which allows re-use and redistribution
3. data is published with equal terms for every user

(The Open Knowledge Foundation, 2019)

When data is openly accessible in a standardized format it will be used directly and not copied to the Urban Data Platform. Indicators are always aggregated and anonymized so they can be published as Open Data. Section 5 lists data sets that are published as Open Data and some of the external data sets listed in section 6 are Open Data as well.

Data that is public, but not in a standardized or even readable format, is considered public data and can still be used as a source of data for calculation of an indicator. Some of the data sets in section 6 of this deliverable fall in this category.

Finally, some data is not made open because of privacy protection and is considered confidential data, which is only available to consortium partners. These internal data sets are listed in section 4.

3.3 Data access

For data stored in the platform (the internal, confidential data sets) access is limited. Specifically the raw measurements are considered confidential. On the other hand, some data is published as Open Data as described in the previous section and is therefore accessible for everyone.

Another aspect of data access is the accessibility of external data sets that are used for the calculation of indicators. Open Data, as described in the previous section, can be accessed at the source for KPIs that are calculated in the platform. Indicators that are entered manually can also use non-Open Data, like the public or confidential data described in the previous section. This data will then be accessed when entering the KPI, but their source data is not stored in the platform.

Calculated indicators using public/confidential information requires copying the data into the platform.

3.4 SCIS

The Smart Cities Information System (SCIS) is a knowledge platform encouraging exchange of data, experience, know-how and collaboration on smart cities to ensure a high quality of life and a clean, energy efficient and climate friendly living environment for the citizens (SCIS, 2019). From the point of view of lighthouse projects, the most typical use of SCIS is its database because reporting of monitoring data to that is mandatory for all projects.

In SCIS, the current approach for data collection is through individual project data collection done by monitoring experts, and this information is periodically updated in the self-reporting tool (SCIS, 2018a). The aim of the data collection is to allow the comparison of results of the projects (SCIS, 2018a). In data collection, a distinction is made between new systems and renovations of existing systems. The

evaluation process uses a bottom-up approach, collecting data from small Energy Supply Units (ESU), buildings and ICT solutions at unit level. These are aggregated in cases where the objective is to evaluate the energy performance of a whole neighbourhood or city. Data quality in SCIS is ensured with:

- Compliance with SCIS data requirements
- Documentation on metadata (such as time of measurement, unit, application area...)
- Adjustments to apparently implausible data is discussed and checked with SCIS

3.5 Quality of Open Data

Although the more data are put on the Web, the richer an information space it would be, not all data have to be open. There are commercial, private and sensitive data that should be protected and exchanged in environments with limited access. Also, as good as the value proposition of Linked Open Data is, not all the Open Data across the Web will be linked. But when it is, it will be the 5-star data that will transform the Web and the way we navigate data.

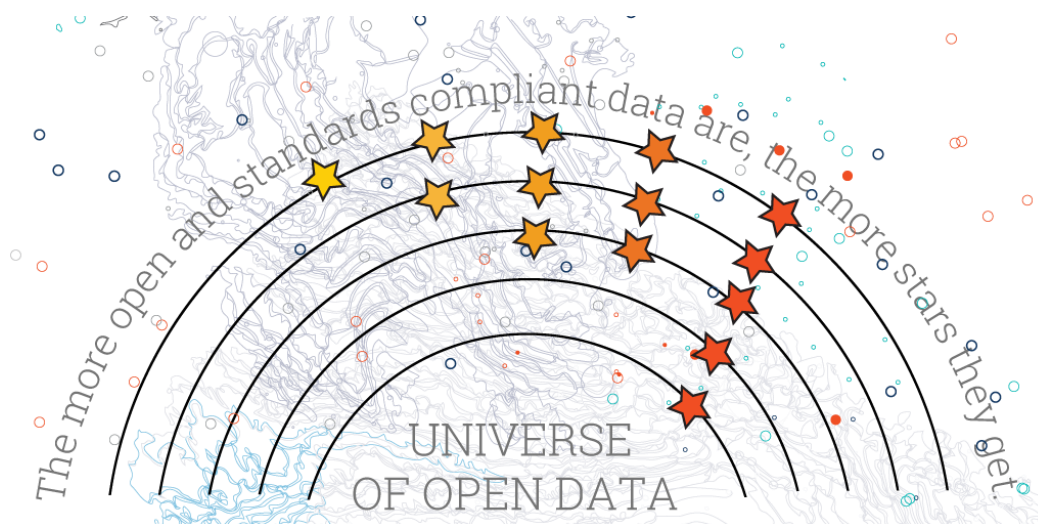


Figure 4: Universe of Open Data

The road to useful open data is paved with stars, each of them representing the level of openness and usability of data.

The more open and standards compliant data are, the more stars they get. The inventor of the World Wide Web (Sir Tim Berners-Lee) created this rating to encourage data providers to take the first step towards opening their data and, ultimately, publishing their data according to the Linked Data principles.

Linked Data is a set of design principles for sharing machine-readable interlinked data on the Web. Open Data, on the other hand, is data that can be freely used and distributed by anyone, subject just to the requirement to attribute and share-alike, at most. **Datasets that are both open and linked are Linked Open Data.**

Question: What Are Linked Data and Linked Open Data?

Linked Data is a set of design principles for sharing machine-readable interlinked data on the Web. When combined with Open Data (data that can be freely used and distributed), it is called Linked Open Data (LOD). An RDF database such as Ontotext's GraphDB is an example of LOD. It

is able to handle huge datasets coming from disparate sources and link them to Open Data, which boosts knowledge discovery and efficient data-driven analytics.

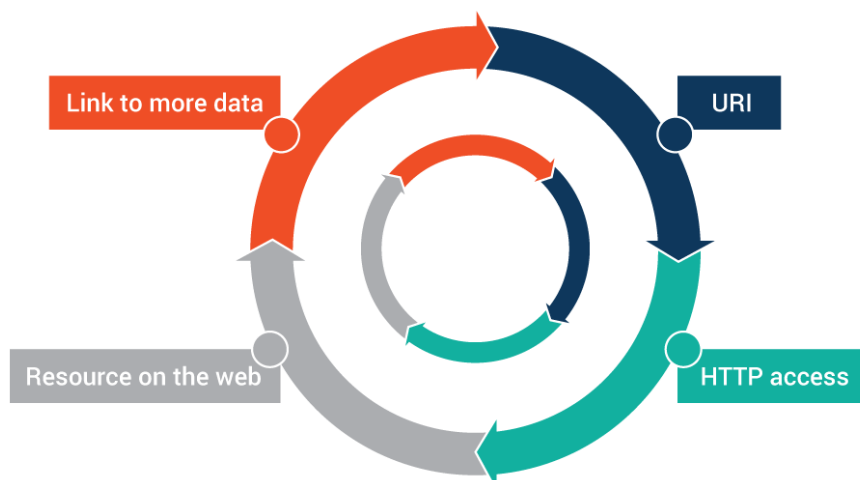


Figure 5: Linked Data

Linked Data is one of the core pillars of the Semantic Web, also known as the Web of Data. The Semantic Web is about making links between datasets that are understandable not only to humans, but also to machines, and Linked Data provides the best practices for making these links possible. In other words, Linked Data is a set of design principles for sharing machine-readable interlinked data on the Web. (Ontotext, 2020)

Similarly to the principles and standards for defining what linked data or open data is, we can also measure how much linked and open a set of data is.

Under the star scheme, you get one star if the information has been made public at all, if it has an open licence. The you get more stars as you make it progressively more powerful, easier for people to use.

- ★ Available on the web (whatever format) but with an open license, to be Open Data
- ★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)
- ★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)
- ★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
- ★★★★★ All the above, plus: Link your data to other people's data to provide context

3.5.1 ★ One-Star Open Data

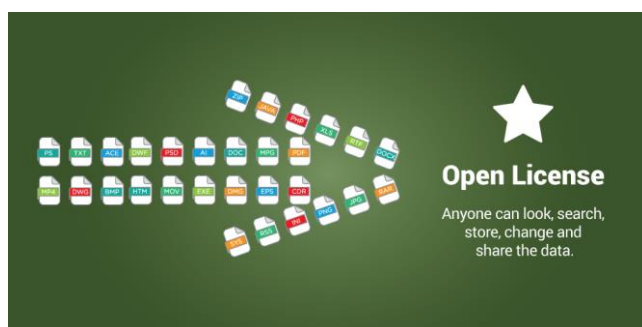


Figure 6: Open License

The one-star open data is defined as data available on the web, in whatever format, but with an open license, so as to be Open Data. Consumers can look, search, store, change data and share the data with anyone they like. As a data publisher, an organization knows that it's simple to publish and does not need to constantly explain to others that they can use the data.

3.5.2 ★★ Two-Star Open Data



Figure 7: Machine readable

In order to win a second star, the open data needs to be available as machine-readable structured data, for example, an excel spreadsheet instead of an image scan of a table. The users of 2-star open data can do anything they do with a 1-star data plus directly processing it with proprietary software and exporting it into another structured format. However, that type of data is still locked up because users depend on proprietary software to be able to get the data out of a document.

3.5.3 ★★★ Three-Star Open Data

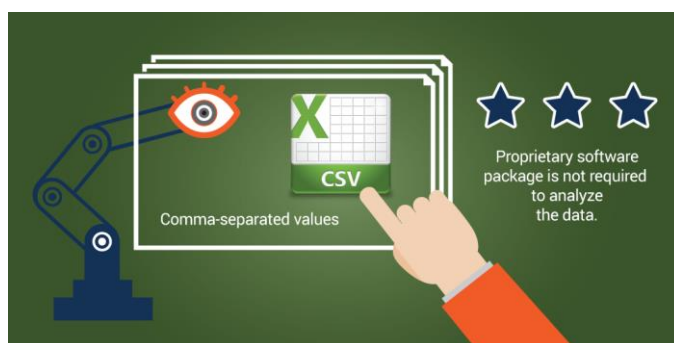


Figure 8: Non-proprietary

Therefore, the third star is awarded to data for which users don't require proprietary software package in order to analyze it. One example of this is the comma-separated values (CSV) format that stores tabular data in plain text.

3.5.4 ★★★★★ Four-Star Open Data

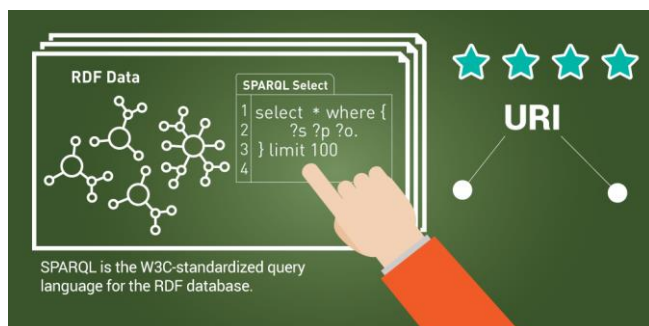


Figure 9: Linked data

Another star goes to data that uses open standards from W3C, such as RDF and SPARQL, to identify things. RDF, which stands for Resource Description Framework, is the standard used in a semantic graph database. This graph database, also called an RDF triplestore, is a type of semantic technology for storing and managing interlinked data and making sense of that interconnected data. Unlike the relational database, the triplestore maps the various relationships between entities in graph databases. SPARQL is the W3C-standardized query language for the RDF database.

The core concept of the triplestore and the underlying Linked Data principle is the Uniform Resource Identifier (URI), a unique ID for all things linked. By representing data in a graph database, the user can link to it from any other place or reuse parts of the data.

3.5.5 ★★★★★ Five-Star Open Data



Figure 10: Interlinked data

With the help of the W3C standards and Linked Data principles, data publishers link their data to other people's data to provide context. This is the prerequisite for getting the fifth star for Linked Open Data, according to Sir Berners-Lee.

The semantic graph database is capable of handling various datasets and maps links to linked open data sources such as DBpedia or GeoNames, for example.

Users of five-star data can discover more and more interlinked information while using the data. As the semantic graph database is capable of inferring new links out of existing facts, users can discover more relationships within their linked data.

3.6 Data Set Definition

The following sections will describe every data set using the same template. The template includes a description as well as the requirements for the data set.

Table 3 Template used for data set description including data requirements

<i>Name of the data set</i>	
Description	<i>Short description of the data set.</i>
Used by Indicators	<i>The indicators for which this data is obtained.</i>
Quality	<i>The Quality rating as defined in section 3.5. Only relevant for Open Data.</i>
Data requirements	
Expected data sources	<i>Where does the data come from?</i>
Expected availability	<i>How easily can the data be obtained?</i>
Collection interval	<i>How often is the data updated? For more real-time data sets: what is the resolution?</i>
Expected reliability	<i>Can the data be trusted to be an accurate reflection of reality?</i>
Expected accessibility	<i>Who has access to the data? Open means it is accessible for everyone.</i>
References	
<i>Any deliverables or other references with information about the data set.</i>	

4 Internal Confidential Data

This section will describe the data sets generated in the project. Most of these data sets are considered confidential data which is reflected in their accessibility being limited to consortium partners.

4.1 Data Set: Building Energy Consumption Measurements

Building Energy Consumption	
Description	The energy consumption per building per type of energy source. Types of energy sources: Thermal, Electricity
Used by Indicators	E1 Final energy consumption E2 Primary energy consumption E3 Energy Imported E6 PED energy balance E7 Energy savings in the PED
Data requirements	
Expected data sources	Measured for buildings that are part of the MAKING-CITY actions.
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Electricity: 15 minutes, Gas: 1 hour, Heat: 1 hour
Expected reliability	High
Expected accessibility	Limited to consortium partners. Could be personal data for residential buildings and sensitive for industrial buildings.
References	
<i>D5.2 Project level indicators</i>	
<i>D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme</i>	

4.2 Data Set: Building Energy Production Measurements

Building Energy Production	
Description	The energy production per building from RES. The energy production per building per type of energy source. Types of energy source: Photovoltaic, Solar thermal, Biomass, DH Geothermal, Hybrid heat
Used by Indicators	E4 Energy export E5 RES production

	E6 PED energy balance
Data requirements	
Expected data sources	Measured for buildings that are part of the MAKING-CITY actions.
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Electricity: 15 minutes
Expected reliability	High
Expected accessibility	Limited to consortium partners. Could be considered personal data for residential buildings and sensitive for industrial buildings.
References	
<i>D5.2 Project level indicators</i>	
<i>D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme</i>	

4.3 Data Set: Building Emissions

Building Emissions	
Description	The greenhouse gas emissions, derived from the energy consumption metrics.
Used by Indicators	E8 GHG emissions E9 CO2-eq. emission reduction
Data requirements	
Expected data sources	Calculated based on 4.1 Data Set: Building Energy Consumption Measurements
Expected availability	Highly dependent on the accuracy of the calculation and its assumptions.
Collection interval	Monthly
Expected reliability	Medium, highly dependent on the accuracy of the calculation and its assumptions.
Expected accessibility	Limited to consortium partners.
References	
<i>D5.2 Project level indicators</i>	
<i>D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme</i>	

4.4 Data Set: Building Net Energy Consumption

Building Net Energy Consumption	
--	--

Description	<p>The energy consumption per building per type of energy usage and per type of energy.</p> <p>Types of energy: Thermal, Electricity</p> <p>Types of energy use: Space heating + cooling + air conditioning, Hot water, Lighting</p> <p>The aggregation of these energy uses are the 'Net Energy Consumption' used as input for calculations of indicators and Primary Energy.</p> <p>Split is based on the Building Energy Specification Table (BEST) format in the Grant Agreement.</p>
Used by Indicators	<p>E1 Final energy consumption</p> <p>E2 Primary energy consumption</p> <p>E3 Energy Imported</p> <p>E6 PED energy balance</p> <p>E7 Energy savings in the PED</p>
Data requirements	
Expected data sources	Calculated based on 4.1 Data Set: Building Energy Consumption Measurements and 4.2 Data Set: Building Energy Production Measurements
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Daily
Expected reliability	Medium, highly depends on completeness of measurements and accuracy of the model for the building (example model is a fixed % of thermal energy allocated to hot water usage).
Expected accessibility	Limited to consortium partners. Could be personal data for residential buildings and sensitive for industrial buildings.
References	
MAKING-CITY Grant Agreement Part B	
<i>D5.2 Project level indicators</i>	
<i>D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme</i>	

4.5 Data Set: Primary Energy Balance

Primary Energy Balance	
Description	All net energy consumption and energy production, converted to Primary Energy enabling the calculation of the energy balance

Used by Indicators	E2 Primary energy consumption E3 Energy Imported E4 Energy export E6 PED energy balance
Data requirements	
Expected data sources	Calculated based on 4.1 Data Set: Building Energy Consumption Measurements, 4.2 Data Set: Building Energy Production Measurements and 6.4 Data Set: Energy Mix
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Daily
Expected reliability	Low to Medium, depends on the accuracy of the conversion from measurements to Primary Energy.
Expected accessibility	Open
References	
MAKING-CITY Grant Agreement Part B	
<i>D5.2 Project level indicators</i>	
<i>D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme</i>	

4.6 Data Set: Buildings

Building Properties	
Description	Information collected about the buildings from public sources or directly from building owners. Examples include floor area, location, number of PV panels etc. Where available linked to 6.1 Data Set: Building Location Information.
Used by Indicators	Input for calculation of Net Energy Consumption and derived energy indicators.
Data requirements	
Expected data sources	Building owners and public information (see also the External Data Sets section).
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Once
Expected reliability	High

Expected accessibility	Limited to consortium partners. Could be personal data for residential buildings and sensitive for industrial buildings.
References	
MAKING-CITY Grant Agreement	

4.7 Data Set: EV Charging Logs

Building Properties	
Description	Logs of EV charging stations specifying duration and energy consumption for each charging session at the station.
Used by Indicators	M2.1 Energy delivered for EV charging M2.2a Total EV charging time M2.2b Total number of EV charges
Data requirements	
Expected data sources	Exports from the CPMS (Charging Point Management System) the charging points are connected to.
Expected availability	Complete for participating charging stations in PEDs.
Collection interval	Yearly, containing all sessions of the past year.
Expected reliability	High
Expected accessibility	Limited to consortium partners.
References	
N/A	

5 Published Open Data

This section will describe the data sets generated in the project and published as Open Data. Most are derived from the internal data sets described in the previous section.

5.1 Data Set: Project Indicators - Energy

Project Indicators - Energy	
Description	<p>Energy related indicators for a building or district:</p> <ul style="list-style-type: none"> E1 Energy consumption E2 Primary energy consumption E3 Energy Imported E4 Energy export E5 RES production E6 PED energy balance E7 Energy savings in the PED E8 GHG emissions E9 CO2-eq. emission reduction <ul style="list-style-type: none"> F1 System flexibility for energy players F2 Energy storage usage F3 Peak load reduction Separate dataset - interval 15min/Day
Used by Indicators	N/A
Quality	★★★★
Data requirements	
Expected data sources	Calculated based on 4.1 Data Set: Building Energy Consumption Measurements, 4.2 Data Set: Building Energy Production Measurements and 4.5 Data Set: Primary Energy Balance
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Monthly
Expected reliability	Medium, depends on the accuracy of underlying data sets.
Expected accessibility	Open
References	
D5.2 Project level indicators	
D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme	

5.2 Data Set: Project Indicators - EV Charging

Project Indicators – EV Charging	
Description	EV charging related indicators for the district: M1.1 Number of public EV charging stations M2.1 Energy delivered for EV charging M2.2a Total EV charging time M2.2b Total number of EV charges
Used by Indicators	N/A
Quality	★★★★
Data requirements	
Expected data sources	City data and 4.7 Data Set: EV Charging Logs
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Month
Expected reliability	High
Expected accessibility	Open
References	
D5.2 Project level indicators	
D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme	

5.3 Data Set: Project Indicators - Investment

Project Indicators – Investment	
Description	C1 Total investments C2 Payback time C3 Economic value of savings
Used by Indicators	N/A
Quality	★★★
Data requirements	
Expected data sources	Grant agreement and 4.5 Data Set: Primary Energy Balance
Expected availability	Complete for participating buildings in PEDs.
Collection interval	Year
Expected reliability	Low
Expected accessibility	Open

References

MAKING-CITY Grant Agreement

D5.2 Project level indicators

D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme

5.4 Data Set: Project Indicators - Social

Project Indicators – Social	
Description	S1 Energy Poverty S2 Consciousness of residents S3 Resident engagement / empowerment to climate conscious actions
Used by Indicators	N/A
Quality	★★★
Data requirements	
Expected data sources	Surveys performed as part of the planned research actions.
Expected availability	Limited
Collection interval	Several times during the project.
Expected reliability	Medium
Expected accessibility	Open
References	
<i>D5.2 Project level indicators</i>	
<i>D5.7 Oulu Monitoring Programme and D5.8 Groningen Monitoring Programme</i>	
<i>D2.11 New citizens' engagement strategies in Oulu and D3.11 New citizens' engagement strategies in Groningen</i>	

6 External Data Sets

This section will describe the externally available data sets used for calculations of performance indicators or which are useful for interpreting/extending the internal data sets.

6.1 Data Set: Building Location Information

Building address information	
Description	Information about buildings including their address, location and usable area.
Used by Indicators	E1-8 (requiring per m2 calculations)
Quality	★★★★★
Data requirements	
Expected data sources	For example in the Netherlands: BAG (http://bag2.basisregistraties.overheid.nl/)
Expected availability	Complete
Collection interval	Once
Expected reliability	High
Expected accessibility	Open
References	
N/A	

6.2 Data Set: District Energy Consumption

District energy consumption	
Description	Energy consumption by district of the city.
Used by Indicators	Baseline for all energy consumption indicators.
Quality	★★★
Data requirements	
Expected data sources	For example in the Netherlands: CBS https://www.cbs.nl/
Expected availability	High
Collection interval	Once
Expected reliability	Medium, borders of districts in the data do not fully match PED borders leading to a small mismatch.
Expected accessibility	Open

References

N/A

6.3 Data Set: City Population Information

City population information	
Description	<p>The following information about the population of the city:</p> <ul style="list-style-type: none"> • Population numbers • Population living with housing cost overburden • Population living in dense areas of the city
Used by Indicators	<p>S1 Energy Poverty</p> <p>S2 Consciousness of residents</p> <p>S3 Resident engagement / empowerment to climate conscious actions</p>
Quality	★★★
Data requirements	
Expected data sources	For example in the Netherlands: CBS https://www.cbs.nl/
Expected availability	Variable
Collection interval	Before and at the end of the project
Expected reliability	High
Expected accessibility	Open
References	
N/A	

6.4 Data Set: Energy Mix

Energy mix	
Description	Information about the energy mix in the city that is used as input for the calculation of Primary Energy.
Used by Indicators	E2 Primary energy consumption
Quality	★★
Data requirements	
Expected data sources	<p>For example in the Netherlands:</p> <p>CBS</p> <p>https://www.cbs.nl/nl-nl/publicatie/2020/44/klimaat-en-energieverkenning-2020 (Tabel 5a Energieverbruik)</p>

Expected availability	Complete
Collection interval	Once
Expected reliability	Medium
Expected accessibility	Open
References	
N/A	

7 Conclusions

Based on the KPIs and Project Indicators selected in Task 5.1 as well as the descriptions of the monitoring programmes, the data sets to be used in calculations have been identified. The data set descriptions in this deliverable contain the requirements for the data sets, their reliability, data access methods, existing data formats and level of confidentiality.

Starting the monitoring programmes encountered challenges and some of the requirements described in this deliverable may not be fully met. The impact and resolution steps will be addressed in the upcoming deliverables about the monitoring programme and related data collection and evaluation deliverables (D5.10 and 5.11).

As the monitoring programmes have now started the data sets are being produced and analysed. This analysis will result in the calculation of Project Indicators and the Open Data described in this deliverable. It will also result in data sets not yet described and together this information will give new insights into the data requirements for the energy transition.

Bibliography

- Bosch, P. J. (2015). *Deliverable 2.1 Definition of data sets. CITYkeys report 2.1*. Retrieved from CITYkeys: <http://www.citykeys-project.eu/citykeys/resources/general/download/CITYkeys-D2-1-Definition-of-data-sets-WSWE-AJENTU>
- Ontotext. (2020, 11 25). *What Are Linked Data and Linked Open Data?* Retrieved from Ontotext Knowledge Hub: <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>
- SCIS. (2018a). *Monitoring KPI Guide D23.1*. Retrieved from Smartcities Information System: <https://smartcities-infosystem.eu/library/resources/scis-essential-monitoring-guides>
- SCIS. (2018b). *Smart Cities Information System Self-reporting guide*. Retrieved from Smartcities Information System: <https://smartcities-infosystem.eu/library/resources/scis-essential-monitoring-guides>
- SCIS. (2019). *About the Smart Cities Information System (SCIS)*. Retrieved from Smartcities Information System: <https://smartcities-infosystem.eu/content/about-smart-cities-information-system-scis>
- The Open Knowledge Foundation. (2019, 08 02). *What is Open Data?* Retrieved from <http://opendatahandbook.org/guide/en/what-is-open-data/>