Deliverable due date: M36 – November 2019

D5.2 Definition of the data sets and requirements

WP5, Task 5.2

Transition of EU cities towards a new concept of Smart Life and Economy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Project Acronym	mySMARTLife
Project Title	Transition of EU cities towards a new concept of Smart Life and Economy
Project Duration	1st December 2016 – 30th November 2021 (60 Months)
Deliverable	D5.2 Definition of the data sets and requirements
Diss. Level	PU
Status	Working
	Verified by other WPs
	■ Final version
Due date	30/11/2019
Work Package	WP5
Lead beneficiary	FVH
Contributing beneficiary(ies)	CAR, NAN, ENG, ARM, TSY, KON, ENH, HEN
Task description	Based on the KPI selection from Task 5.1, the present task will describe the data requirements for evaluating the Smart City Projects and the corresponding indicators on Smart City level. The identification of the data sets to be compiled will be based on the previous work by CITYkeys (D2.1 "Definition of data sets"). Available data sources, their reliability, formats, level of confidentiality and data access methods will be analysed by all the partners involved, considering the experience of FVH in the open data concept and for each specific domain. Privacy and security issues will be tackled. 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.

Date	Version	Author	Comment
03/10/2017	0.1	Timo Ruohomäki (FVH)	First draft
15/11/2017	0.2	Timo Ruohomäki (FVH)	Updated with city level indicator lists
22/11/2017	0.3	Timo Ruohomäki (FVH)	Updated with information of the latest revision of indicator list. Ready for review
27/11/2017	1.0	Timo Ruohomäki (FVH)	Final version, interim report ready for submission
30/11/2017	1.0	Jose Hernández (CAR)	Final version, ready for submission Interim version M12
28/10/2018	1.2	Timo Ruohomäki (FVH)	First draft of Interim version M24, minor additions
20/11/2018	1.3	Sebastian Schmitz (HAM)	Input on Hamburg datasets
23/10/2019	2.0	Timo Ruohomäki, Tiina Inki, Antti Saarinen (FVH)	First edition of final deliverable M36
12/11/2019	2.1	Guillaume Chanson (NAN), Pierre Nouaille (CER)	Input on Nantes datasets
12/11/2019	2.2	Mikko Virtanen (VTT)	Input on Helsinki datasets
19/11/2019	2.3	Jose Hernández, Ana Quijano (CAR)	Review of the document
28/11/2019	2.4	Timo Ruohomäki (FVH)	Updates on comments after CAR review
30/11/2019	3.0	Ana Quijano (CAR)	Final review

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Table of Content

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Table of Figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Table of Tables

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Abbreviations and Acronyms

Acronym	Description
mySMARTLife	Transition of EU cities towards a new concept of Smart Life and Economy
LH City	Lighthouse City
SECAP	Sustainable Energy and Climate Action Plan
LOD	Linked Open Data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

1. Executive Summary

This document is the final version of deliverable 5.2 that deals on list and analyze the data sets used to calculate indicators in the three LH cities of mySMARTLife: Nantes in France, Hamburg in Germany and Helsinki in Finland.

The data sets assessment has been mainly focused on the sources of datasets to calculate city level indicators through the application of 5-Star Scheme method. Analysis performed show how well the data are available by each indicator category in each city as well as some figures related to the availability of data sets by references used to define indicators.

The work has been heavily linked with Deliverable D5.1 and the city baseline assessments documented in D2.1, D3.1 and D4.1, where indicators were defined and calculated, respectively.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

2. Introduction

2.1 Purpose and target group

The purpose of this deliverable is to list and analyze the data sets used to calculate city level indicators in the three LH cities involved in the project that will be used to perform city audits. The target group of the document are all the participant partners who have been involved with indicator definitions or are otherwise interested in analysis of the data sets in relation to urban platform development.

2.2 Contributions of partners

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

Table 1: Contribution of partners

Participant short name	Contributions
FVH	Main author of the deliverable
CAR	Link with D5.1 and reviewer
VTT	List and analysis of the data sets for Helsinki
HAM	List and analysis of the data sets for Hamburg, description of the Hamburg open data services
CER	List and analysis of the data sets for Nantes
NAN	Description of the Nantes open data services

2.3 Relation to other activities in the project

The following table depicts the main relationship of this deliverable to other deliverables developed within the mySMARTLife project and that should be considered along with this document for further understanding of its contents.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Table 2: Relations with the other activities in the project

Deliverable Number	Contributions
D5.1	City and project level indicators are defined as main assessment tool to perform city audits and measure the success of the actions implemented in the three lighthouse cities which participate in mySMARTLife project. Conclusions on the availability of data that allow the quantification of city indicators will be considered as reference for the data sets analysis
D2.16	It is expected, that the data set analysis will provide requirements and recommendations in relation to the semantic data models of urban platform
D2.1/D3.1/D4.1	These deliverables provide the city audits developed in the three LH cities
D5.3	This deliverable develops a robust and complete monitoring program of the project actions that will allow the introduction of future data after the end of the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

3. Dataset Analysis

The project aims to improve the availability of the data sets by the means of urban platform and its data catalogue function. More of the data sets are expected to be machine readable and more up to date by the end of project. Ideally, the data sets would be real-time, allowing observation of the key city performance indicators without any manual preparations or calculations in the city dashboard.

On other hand, one of the core activities of the project is the evaluation of the impacts of the actions implemented in the cities through monitoring equipments as well as to work in methodologies for the definition of innovative urban transformation strategy of participant cities in the project. For both objectives, it has defined an evaluation framework that consists of city and project level indicators in Deliverable 5.1: Integrated evaluation procedure.

Whereas city level indicators have been used to perform city audits of LH and follower cities that help to identify the progress and challenges of these cities in the pillars analysed (environment, energy, mobility, economy, citizens and urban infrastructures), project level indicators will be employed to measure the impacts of the project actions in demoareas and at city level.

For the definition of such indicators, an exhaustive review of literature has been performed in order to identify the most suitable indicators. Additionally, other criteria have been taking into account such as the relevance of the indicator to measure the objective of evaluation and the expected availability of the required information. Whereas for the project level indicators, the availability is connected mainly to the monitoring equipment to be installed, the availability of city level indicators will be according to the current activities performed by the cities to collect and storage information. Consequently, before to propose the collection of the values of the city level indicators, an exercise was proposed to evaluate the availability of data and therefore to identify the existing data sources/data sets to calculate this type of indicators.

Now that the contexts of this deliverable has been introduced, the concept of data sets and the methods to analyze datasets are described. Thus, the data sets can be defined as the sources that provide the values for each of the indicator, existing the possibility that a single data set can be used in one or more indicator.

3.1 Methods

For the identification of the main characteristics that has to follow the method to evaluate data sets in the project mySMARTLife, CITYkeys (D2.1 “Definition of data sets”) has been used. This reference states that the analysis of data sets should include reliability, data access methods and existing data formats, as well as potential privacy and security issues. On other hand, the type of indicators defined in Deliverable 5.1 will be taking into account to focus the analysis of data sets. Thus, since the majority of the required data sets are quantitative, the quality or reliability of data sources are not assessed since only some indicators will be required the performance of surveys to collect the information. Consequently, the approach of the method

This project has received funding from the European Union‘s Horizon 2020 research and innovation programme under grant agreement No 731297.

will be the availability of data since is a well indicator on how well the cities' urban platforms provide data sets that are seen significant in analysing the city's performance. Hence, it is proposed 5-star assessment method.

3.2 5-Star Scheme method

In order to improve data interoperability and reusability, Linked Open Data (LOD) principles by Tim Berner-Lee provide many useful tools and schemes to assess and categorize data sources by how useful they are to other digital services. The goal was to judge the quality of data by its accessibility (open data access), by its format and structures and by its interoperability. There are two main types of data principles used to support this target, the FAIR -principle and the 5-star scheme. The first data principle has the acronym FAIR and it emphasizes that in order to data being interoperable, it should be Findable, Accessible, Interoperable and Reusable. In FAIR data is expected to have a stated license for access, thus emphasizing the license agreement in reusability. In FAIR scheme contextual information is also required to improve the reuse of data.

The 5-star scheme was introduced by Tim Berners-Lee in 2010 to encourage especially government data owners along the road to good linked data. It focuses less on the license than the FAIR principle and assumes the data is available with open license. Under the scheme, the first star is achieved by simply making the data public with an open license. The following stars are achieved when the data is machine-readable in a closed or open format and finally, the full five stars are given to linked open data (LOD), that fulfils all the requirements the semantic web has for data sources. However, having 5 stars in this scheme does not mean that the data is open data. The following figure illustrates the scheme:

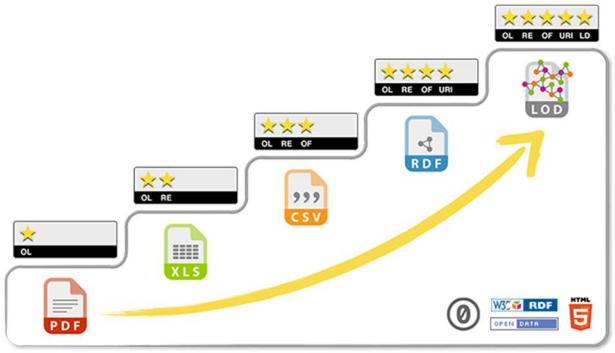

Figure 1: Five Star Scheme

Therefore, the 5-star assessment is a method to define the usability of a data source in the web context: the score is higher the more open, machine-readable and easily interpretable the data source is. While for the project purposes, it is reasonable to look up for certain indicator value from the printed statistical yearbook, this method does not allow creating new, online data services on top of such data set.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

4. City indicators related data-sets

Before the 5-Star Scheme method is applied to evaluate the performance of data-sets, this section describes the existing data sets in each LH city to calculate the city level indicators as well the categories defined for this type of indictors.

4.1 City level indicators defined

mySMARTLife evaluation framework at city level consist of 151 indicators and 8 categories that have been defined to have an integrated evaluation of the city. Additionally, it has to mention that these categories (named as Fields) have been used for the definition of project level indicators.

Table below shows the categories of city level indicators and the number of indicators defined by each one of these categories.

Table 3: City Level Indicators

Fields	City level indicators
City characterization	6
Economy	16
Environment	23
Energy	33
Mobility	22
Urban Infrastructures	20
Governance	15
Citizens	16
TOTAL	151

Bibliography used to identify these indicators correspond with the following initiatives: CITYkeys, SCIS, SECAP and other SCC projects previously to mySMARTLife such as REMOURBAN, SmartEnCity and REPLICATE. All the data sources employed to define the list of city level indicators for mySMARTLife are introduced and described in table below.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Table 4: Indicator Sources

Name	Description
CITYkeys	CITYKeys was a project supported by the Commission to define the evaluation framework of funded demonstration projects for Smart Cities and Communities projects (SCC). (Grant 646440)
SCIS	SCIS is a 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 focuses on the development of indicators to measure technical, social and economic aspects of energy related measures of SCC projects. This is a continuation of the CONCERTO program
TELEFONICA	Entity that work in ICT and that have elaborated a dosier which include a list of indicators to measure how to evaluate the development of a Smart City
EUROSTAT	Statistical office of the European Union
REMOURBAN	Regeneration Model for accelerating the Smart Urban transformation - (Grant agrement project 646511)
REPLICATE	Reneissance of Places with Innovative Citizenship and Technology (Grant agreement project 691735)
SmartEnCity	Towards Smart Zero CO2 Cities acrros Europe (Grant agreement project 691883)
SECAP	Monitoring program of the Covenant of Mayors for evaluate Sustainable Energy and Climate Action Plans

A detailed description of this data sources is performed in D5.1 but at this section has to emphasize that each that the combination of these sources allow to develop an integrated methodology since each of these sources is focused in some domains or complement some of them (e.g. ICT indicators are proposed by CITYKEYS and SCIS but they have some gaps that are covered with the dossier developed by Fundación Telefónica).

4.2 Summary of data sets

The following table provides a summary of all the indicators in each field and the data sets required to calculate them. In case an indicator has a formula (e.g. population/land area), there can be more than one dataset required for the calculation of indicator.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Table 5: Data sets for City Level Indicators

Fields	City level indicators	Data sets total
City characterization	6	8
Economy	16	23
Environment	23	30
Energy	33	35
Mobility	22	25
Urban Infrastructures	20	27
Governance	15	15
Citizens	16	19
TOTAL	151	182

This analysis is useful, since the effort put on WP5 actions will help the cities to process the monitoring reports like SECAP and also provide a way to create comparisons between the cities on key elements.

The following diagrams illustrate the differences on data requirements between each main domain of the project:

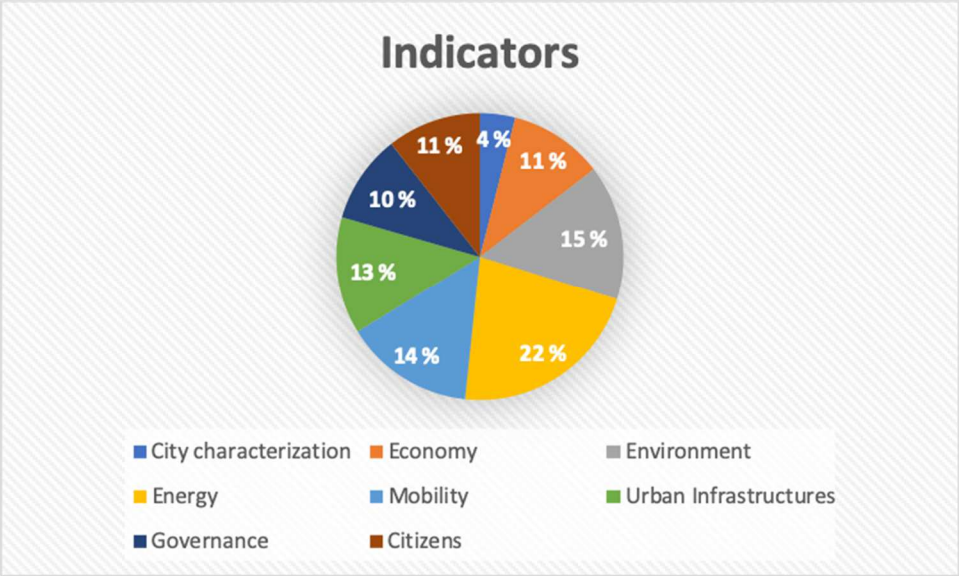

Figure 2: Indicators by Field

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

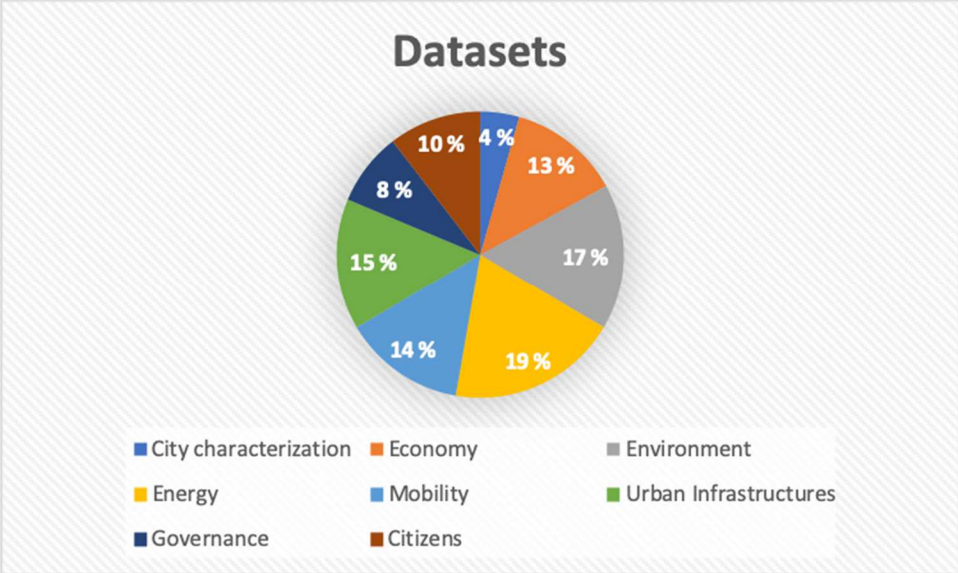

Figure 3: Datasets by Field

4.3 Nantes Datasets

4.3.1 Overview

A new open data portal https://data.nantesmetropole.fr was opened in 2018. The service was created to support the new data strategy and it was especially targeting the following goals:

- Encourage the opening of data to create conditions for the emergence of new services by developers
- Develop open data at the service of citizen dialogue and public life
- Enrich the portal by opening new datasets and supporting its partners, including all municipalities in the Metropolis

In addition, the Nantes Métropole participates in the national experiment https://transport.data.gouv led by the Inter-ministerial Direction for Digital Affairs and the State's Information and Communication System (DINSIC), the Ministry of Transport and Etalab, the French Agency for Public Open data. The platform is the French National Access Point (PAN) and aims to eventually reference all transport data. It significantly improves the links between producers and re-users of data and provides all the tools and information needed for their quality, interoperability and reuse.

The open data portal provides various datasets and APIs under the following domains: Citizenship (385 datasets), Mobility (129), Social and Health (225), Culture and Tourism (77), Environment (113), Sport (65), Urban Housing (66), Education, Youth and Training (33) and Territories (75). In most of the datasets the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

license is either Etalab or Open Database License (ODbL). As addition to datasets and API's, the service also hosts a large amount of information of events. The data can be illustrated in graphs or maps.

4.3.2 Statistics in Nantes Metropole

Nantes Metropole can leverage on plurality of actors to conduct and propose statistical data and analyses.In the first place, Nantes Metropole has a strong competency of data analysis. Notably, a whole department is dedicated to geographical data management, inheriting from a strong expertise of a more than 20 years of competency of this subject.

On specific subjects, Nantes Metropole teams up with local stakeholders:

- AURAN (Nantes' city planning agency), which produces analyses, deciphers trends and updates data for local authorities. AURAN works in partnership with Nantes Metropole on urban planning matters. AURAN works as decision-making tool and a resource for understanding and remembering territories. Notably, they've played a key role on the socio-economic characterisation of the Metropolitan area, the Territorial Coherence Scheme (SCoT), the Metropolitan Urbanism Plan (PLuM), to name a few.
- INSEE (French National Institute of Statistics and Economic Studies), which collects, analyses, and disseminates data and analysis on the French economy and society. INSEE contributed heavily to the demographical and socio-economical characterisation of the Metropolitan area.
- Air Pays-de-la-Loire, regional agency approved by the Ministry of Ecological and Solidarity Transition to ensure the monitoring of air quality in the Pays de la Loire region. Notably, Air Pays-de-la-Loire monitors air quality, informs the public and the competent authorities, supports decision-makers with actions assessments, and contributes to knowledge improvement of the subject in the region.

With the support of all its partners, of which only the most important contributors to the City-Level audit are mentioned above, Nantes Metropole can take advantage of a vast set of data & analyses to best implement its public policies.

In accordance with regulatory obligations, and the metropolitan strategy for data, Nantes Métropole strives to make Data as much Open as possible through its portal, and thanks to the urban data platform mySMARTLife has helped to develop.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

Figure 4: Nantes Open Data Portal

4.3.3 Dataset Examples

Table below includes an example of Data sets in Nantes and the result obtained when 5-star method is applied.

Table 6: Summary of Data sets (Nantes)

Dataset	Reference	5-star Assessment
Population	https://data.paysdelaloire.fr/explore/dataset/234400034_004-005_population_en_rpdl/	5
Unemployment Rate	https://data.paysdelaloire.fr/explore/dataset/234400034_chomage-zone-t1-2003-t4-2017/table/	5
GDP	https://www.collectors2020.eu/wcs-ppw/nantes-metropole-fr/	1
Land area of the city	https://www.insee.fr/fr/metadonnees/cog/epci/EPCI244400404-nantes-metropole	2
Final energy consumption per capita	https://data.paysdelaloire.fr/explore/dataset/244400404_bilan-renovation-energetique-coproprietes-nantes-metropole%40nantesmetropole	2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

4.4 Hamburg Datasets

4.4.1 Overview

In August 2017 a new organisational unit has been formed in the Free and Hanseatic City of Hamburg that is responsible for the central topic of urban data: The Urban Data Hub. The topic of urban data has been recognized as one of the pillars of the cities' digitalisation strategy "Digital City" (see also https://www.hamburg.de/contentblob/9260386/b9c01217ea0a8722e91b276d6906db0b/data/strategie-englisch.pdf). Among others, the Urban Data Hub is responsible for the operation and further development of the Urban Data Platform Hamburg (HH_UDP).

In mySMARTLife, the approach in Hamburg is that most datasets needed for the automated calculation of indicators and especially the indicators themselves should be integrated and accessible through the Urban Data Platform to allow broad re-use by other city or civic use cases. Thus, in the following the concept of the Urban Data Platform Hamburg and their two central data research portals are briefly described.

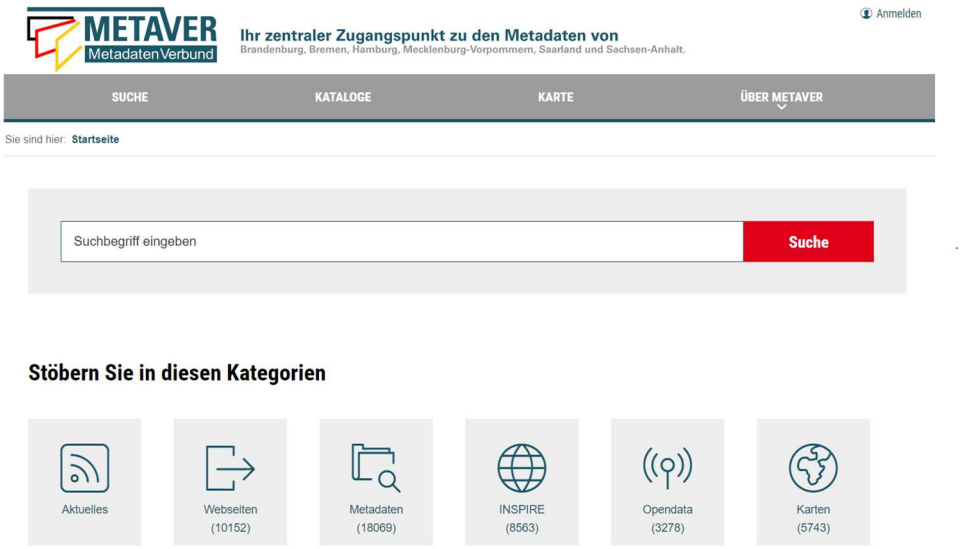

Figure 5: Metaver Data Catalogue

4.4.2 Urban Data Platform Hamburg

The Urban Data Platform Hamburg follows a system of systems approach and aims at making the cities' urban data discoverable, available and exchangeable between systems. At the time of writing, 35 IT-systems and more than 3000 datasets are integrated into the platform and made available via standardizes interfaces (APIs). The Urban Data Platform builds on the existing Spatial Data Infrastructure and extends it with additional components, e.g. for handling sensor data or conducting data analysis.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

4.4.3 Data research portals

There are two central data research portals available on the Web to search and find datasets from Hamburg. On the one hand, the central metadata catalogue (http://metaver.de) that offers metadata on services, applications and datasets in some way related to spatial data. On the other hand, there is the transparency portal (http://transparenz.hamburg.de/), which comprises most, but not all, datasets from the central metadata catalogue as well as other datasets from other city databases as well as many documents like contracts and reports. Both components have a strong legal basis and are heavily used for data discovery both by city employees and by many, anonymous users on the Web.

4.4.4 Dataset examples

Table below includes an example of Data sets in Hamburg and the result obtained when 5-star method is applied.

Table 7: Summary of Data sets (Hamburg)

Dataset	Reference	5-star Assessment
Population	Laut Tabelle der BWVI http://www.statistik-nord.de/zahlen-fakten/bevoelkerung/	2
Unemployment Rate	https://www.statistik-nord.de/fileadmin/Dokumente/Jahrb%c3%bccher/Hamburg/JB16HH_Gesamt_Internet_min.pdf	1
GDP	http://vgrdl.de/VGRdL/tbls/VGR_FB.pdf	1
Land area of the city	http://www.statistik-nord.de/zahlen-fakten/gebiet-flaeche/	2
Final energy consumption per capita	http://www.ugrdl.de/veroeffentlichungen.htm#gemein	1

4.5 Helsinki Datasets

4.5.1 Overview

As part of the organizational reform of the city of Helsinki, various statistics and research functions were merged into one central unit called Urban Research and Statistics Unit. The unit maintains key databases and organizes related research. The city-level statistical operations complement the national level work of Statistics Finland. The statistical and research projects are divided thematically into the following six

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

categories: Population, immigration, health and wellbeing, economy and competitiveness, housing, city structure and environment and urban culture and participation. For the definition of data sets for the mySMARTLife -indicators, this work provides a solid background, although the online availability of a single data set may be limited.

Because of practical reasons and the nature of city indicator lists being manually collected, some summary work such as the data set collected for the Helsinki's 2030 Climate Technologies report were useful and time-saving. Since many of the indicators are actually derived from the same data sets by the means of different filter parameters, even a quick view on the sources will a provide useful

4.5.2 Helsinki Region Infoshare

Helsinki Region Infoshare (HRI) is an online service for publishing data, based on the CKAN platform. It is currently hosting over 900 data sets that are provided by organizations from within the metropolitan area. An example query with the keyword "energy" will provide links to 15 datasets, including relevant excel files like "Energy Consumption in the Helsinki Metropolitan Area".

HRI is the default place for sharing open data. The actions of mySMARTLife will generate more data and HRI will be the part of the urban platform to host the aggregated and time series datasets from the monitoring system.

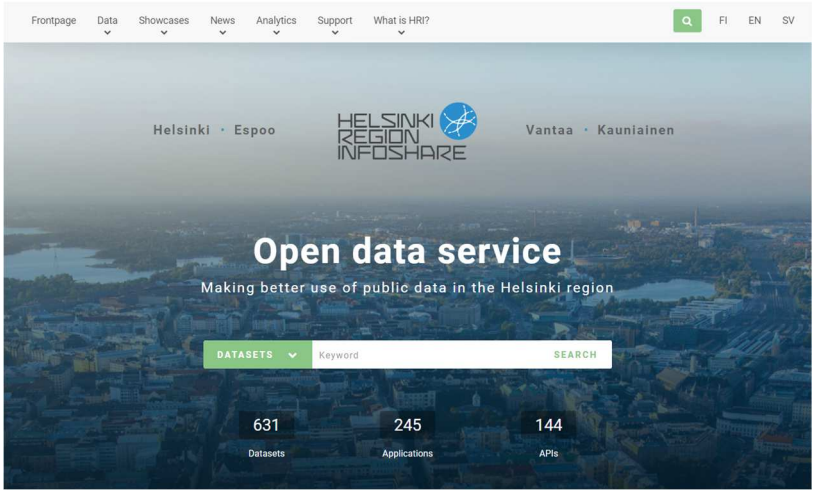

Figure 6: HRI Open Data Service

4.5.3 Statistics Finland

Statistics Finland is the only Finnish public authority specifically established for statistics. It produces the vast majority of Finnish official statistics. Currently, Statistics Finland maintains 153 statistics, containing 2.000 open data tables and 95 general government registries. The data is collected through inquiries and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

interviews. Due to the data protection requirements, the open data services are sometimes aggregated and may miss some key dimensions such as parameters to filter statistics based on the city. Researchers such as the personnel of the Urban Data and Statistics Unit of the city have access to more detailed information. Local data products such as the Statistical Yearbook of the City of Helsinki will then complement the work made on the national level.

The Statistics Finland has improved their online presence in the recent years. Most of the data products can be browsed on their website, although some filtering limitations may exist. Many of the data services are also made available with an API called PX-Web API. It is however mostly limited to the names and other metadata of the tables, not the data itself. It is expected that the API service will improve in the near future.

4.5.4 Operational Databases

For detailed information on built environment, the key operational system in Helsinki is called Facta. It is used as a building register and is maintained by the city survey services. The information maintained in the registries are aimed to be available on public API's as well, using the OGC WFS, WMS and WMTS API services. The registry contains information on all the buildings with the status of "completed". The registry also contains personal information according to the GDPR directive, thus limiting its public availability.

For more detailed information on energy usage, the project has been able to get datasets from the Helen Sävel+ -service. It is intended to provide smart meter information to the residents, so publicly one can only retrieve only information related to the energy consumption on her dwelling. The service requires authentication because of privacy requirements.

4.5.5 Dataset Examples

Table below includes an example of Data sets in Helsinki and the result obtained when 5-star method is applied.

Table 8: Summary of Data sets (Helsinki)

Dataset	Reference	5-star Assessment
Population	https://hri.fi/data/en_GB/dataset/helsingin-vtueittain-1962-alkaen	2
Unemployment Rate	https://hri.fi/data/fi/dataset/aluesarjat_a04s_hkis_tyottomat_sukupuoli	2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

GDP	https://www.tilastokeskus.fi/static/media/uploads/tup/suoluk/suomilukuina_tau_kan007.xlsx	2
Land area of the city	https://hri.fi/data/fi/dataset/helsinki-alue-ja-ymparisto	2
Final energy consumption per capita	https://hri.fi/data/fi/dataset/paakaupunkiseudun-energiankulutus/resource/c56b3e9f-c742-4d9b-aa51-34c90565e53a	2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

5. Dataset Assessment

5.1 Availability of Datasets by Cities

The following table illustrates how well the cities were able to provide datasets for the City Level indicators on each WP5 field. The reasons why a dataset was not available was mostly because such statistics is not collected or that no reliable or official figures were available. In City Level indicators, it is unlikely that privacy or security issues would prevent opening data because the data is generic and covers the whole city. As an example, none of the cities had a suitable dataset for the Economy -indicator "E-commerce" defined as the number of e-commerce transactions through electronic and mobility payment or Number of Internet connections per 100,000 inhabitants. Since for the first case is understood that the required information is not collected, it is more difficult to understand why it is not possible to collect information that allow measuring the extent of ICT in the cities.

Table 9: Availability of datasets

Name	Nantes	Hamburg	Helsinki
Main City Features	100%	100%	100%
Economy	50%	50%	69%
Environment	74%	52%	87%
Energy	58%	48%	100%
Mobility	73%	64%	100%
Urban Infrastructures	55%	75%	95%
Governance	60%	53%	100%
Citizens	31%	31%	88%

As a summary, there can be seen major differences between the cities in how well statistical data is available for the typical indicators. Especially in the case of Citizens -domain and Citizen Involvement to be precise, the differences were significant. Naturally this may not mean that the data and statistics is not collected but it may be collected for a purpose and not provided open.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

5.2 Availability of Data sets by Indicator Sources

Analysis of availability of data according to the data source used to identify the indicators is show in next figure.

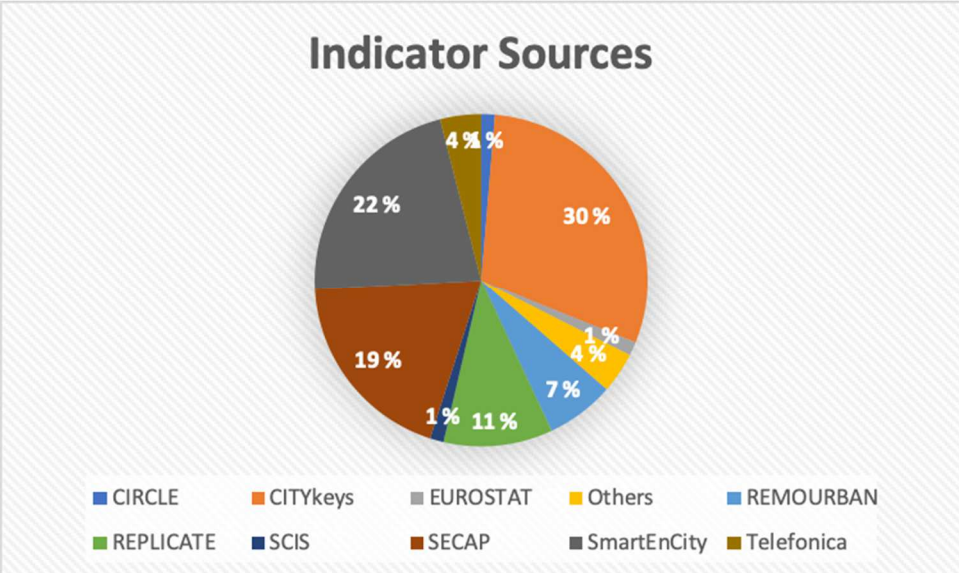

Figure 7: Indicator Sources

5.3 Requirements for project indicators related data-sets

In terms of project level indicators, data will be based on the actions to be carried out in mySMARTLife project throughout the monitoring phase. Therefore, these data sets still mostly do not exist. At this time, the intention is only to establish the high-level requirements for these data-sets which will be used for the monitoring programme definition.

- Available data sources, including both technical availability and possibly business, privacy or security restrictions
- Their reliability, mainly how accurately they will present the phenomenon and how the data gets updated in timely fashion
- Formats, with the attempt to harmonize formats and support the use of international standards instead of proprietary formats
- Level of confidentiality
- Data access methods, especially work on metadata and data APIs to ease use of data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

On data availability the project has provided an opportunity of improving the tools to locate suitable data with the emphasis on city-specific urban data platform, that also acts as a data catalogue. The project has also encouraged data owners to open data and provided facilities and services to support the effort.

A further analysis on these aspects will be included in D5.3 at M48.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

6. Conclusions

In the beginning of the project, the first impression on the collection of values for indicators was that the online availability of such data sets was raising concerns. Over the time, the situation has improved and now all the cities have online data portals with easy search functions for related datasets. The data portals provide data in various formats, not only as tables but also as charts and maps. In the case of Nantes, the new open data platform provides many of the datasets in JSON format in API. In case of Helsinki, the API mostly provides only metadata of the dataset, not the actual values. The data is quite well available for research purposes when tables can be looked up manually, but for automatic retrieval of a data sets there is still work to do on data catalogues. Improving the situation is mostly out of the scope of this project, but hopefully the work made on 5.1 and 5.2 and the feedback given to related parties will make online services like dynamic dashboards maintaining key indicators easier to provide in the future.

On other hand, it has to remark that the work made among T5.1 and T5.2 has been in parallel and in a coordinated way in order to find out the best working methods to manage the data set references with a reasonable effort. Thus, whereas D5.1 is focused in finding the most suitable indicators to perform city audits and find the indicators categories that cannot be evaluated due to the lack of available information, D5.2 has continued this work through the analysis of data sets availability by each of these indicator category that allow to take conclusions on the importance to have open data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.

7. References

Bosch, P., Jongeneel, S., Rovers, V., Neumann, H-M., Vielguth, S., Huovila, A., Pinto Seppä, I., Penttinen, T., van der Heijden, R., Kotakorpi, E., Pangerl, E., Malnar Neralic, S., Sarasa Funes, D. (2016) *D2.1 Definition of data sets.* CITYKeys Project.

Barners-Lee, T. Linked Data https://www.w3.org/DesignIssues/LinkedData.html

mySMARTLife D2.1 Baseline report of Nantes demonstration area

mySMARTLife D3.1 Baseline report of Hamburg demonstration area

mySMARTLife D4.1 Baseline report of Helsinki demonstration area

mySMARTLife D5.1 Integrated Evaluation Procedure

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731297.