

OPTIMISED ENERGY EFFICIENT DESIGN PLATFORM FOR REFURBISHMENT

Optimised Energy Efficient Design Platform for Refurbishment at District Level H2020-WORK PROGRAMME 2014-2015 - 5. Leadership in enabling and industrial technologies H2020-EeB-05-2015: Innovative design tools for refurbishment at building and district level

	D1.3: Requirements and specification of geo-clustering data sets access module
	WP1, Task 1.3
	August 2016 (m12)
Deliverable version:	D1.3, v1.5
Dissemination level:	Public
Author(s):	Patricio Moreno ¹ , Giuliana Mangiapia ² , Ciro Caterino ² , Mariano Folla ² , Alvaro Sicilia ³ , Maxime Pousse ^{4,} Pedro Fernandez-Orellana ⁷ ,
	(¹ ACC, ² ES, ³ FUNITEC, ⁴ NBK, ⁵ TEC, ⁶ TUC, ⁷ UTRC-I)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 680676

Document History

Project Acronyr	n	OptEEmAL		
Project Title		Optimised Energy Efficient Design Platform for Refurbishment at District Level		
Project Coordir	ator	Miguel Á. GARCÍA-FUENTES (<u>miggar@cartif.es</u>) Fundación CARTIF		
Project Duratio	n	1 st 5	September 2015 – 28 th Fe	bruary 2019 (42 Months)
Deliverable No.		D1.3	3. Requirements and speci	fication of geo-clustering data sets access module
Dissemination	Level	PU		
			Working	
Status			Verified by other WPs	
			Final version	
Due date		31/	08/2016	
Work Package		WP1 proc		ent trough IPD design strategy and input data
Lead beneficia	ry	ACC	:	
Contributing beneficiary(ies))	ES, FUNITEC, NBK, TEC, TUC, UTRC-I		
DoA		Task 1.3 – Integration of existing external data sources: geo-clustering techniques		
Date	Version	Author Comment		Comment
12/01/2016	0.1	Patr	ricio Moreno (ACC)	Table of Contents
01/02/2016	0.2	EIH	assan Ridouane (UTRC-I)	UTRC contribution
04/02/2016	0.3	Giar	nluca Mameli (ES)	ES contribution
08/02/2016	0.4	Max	time Pousse (NBK)	NBK contribution
19/02/2016	0.5	Patr	ricio Moreno (ACC)	ACC contribution
29/02/2016	0.6	Patricio Moreno (ACC)		Final intermediate version including missing sections and modification related to internal reviews
06/05/2016	0.7	Giar	nluca Mameli (ES)	Updated table of contents
01/07/2016	0.8	Circ Caterino (FS) contents and content added to introduction,		Redefinition of the document, updated table of contents and content added to introduction, unstructured data service and unstructured data access module functional architecture
03/08/2016	0.9	Orel	ro Fernandez- Iana(UTRC-I) ricio Moreno Montero	UTRC-I contributions





		(ACC)	ACC contributions
		Maxime Pousse (NBK)	NBK contributions
		Giuliana Rosa Mangiapia, Ciro Caterino, Mariano Folla (ES)	ES contributions
09/08/2016	1.0	Giuliana Rosa Mangiapia, Mariano Folla (ES)	Unify contents – draft deliverable
18/08/2016	1.1	Miguel García (CAR)	Review
22/08/2016	1.2	Mariano Folla (ES), Patricio Moreno (ACC)	Document revision and re-writing
29/08/2016	1.3	Gema Hernández (CAR)	Review
29/08/2016	1.4	Giuliana Mangiapia, Mariano Folla (ES), Patricio Moreno (ACC)	Document revision and re-writing
30/08/2016	1.5	Patricio Moreno (ACC)	Minor editing

Copyright notices

©2016 OptEEmAL Consortium Partners. All rights reserved. OptEEmAL is a HORIZON2020 Project supported by the European Commission under contract No.680676. For more information of the project, its partners, and contributors please see OptEEmAL website https://www.opteemal-project.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 OptEEmAL 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 OptEEmAL members as of the date they are published. The OptEEmAL consortium does not guarantee that any information contained herein is error-free, or up to date, nor makes warranties, express, implied, or statutory, by publishing this document.





Table of Content

Execu	utive Su	ummary.		8
1	Introc	duction		9
	1.1	Purpos	e and target group	9
	1.2	Contrib	outions of partners	9
	1.3	Relatio	n to other activities in the project	
2	Struc	tured da	ta service	
	2.1	Objecti	ves	
	2.2	Externa	al data sources to be accessed	
		2.2.1	Energy Plus Weather	
		2.2.2	Eurostat	
	2.3	Conclu	sions for structured data gathering module	
3	Unstr	uctured	data service	
	3.1	Usage	examples	
4	Speci	fication	of Requirements	
	4.1	Data re	equirements coming from the simulation	
	4.2	Functio	onal requirements definition	
		4.2.1	Structured Data	
		4.2.2	Unstructured Data	
	4.3	Non-fu	nctional requirements definition	
		4.3.1	Structured Data	
		4.3.2	Unstructured Data	
5	High I	level des	ign of the geo-clustering connector	
	5.1	Integra	tion into the OptEEmAL platform	
	5.2	Functio	onal architecture of the Geo Connector component	
	5.3	Functio	onal architecture of the Unstructured data service	
6	Qualit	ty integra	ation test	
	6.1	Criteria	used to assess data quality	
	6.2	Minimu	ım requirement for data quality	
	6.3	Applica	tion to structured data sources	
		6.3.1	Energy Plus Weather	
		6.3.2	Eurostat	
7	Concl	usions		
8	Refer	ences		





List of Figures

Figure 1: Eurostat REST request	16
Figure 2: Capture of datasets list - Energy Statistics	16
Figure 3: Eurostat gas price response	17
Figure 4: Capture of datasets list - Energy Statistics	17
Figure 5: File obtained from Eurostat and codes	18
Figure 6: File obtained from Eurostat and codes	18
Figure 7: QOTF results for Cuatro de Marzo District	21
Figure 8: QOTF results for Sneinton District	21
Figure 9: Geo Connector integration	26
Figure 10: Unstructured Data Connector integration	27
Figure 11: Functional architecture of the Geo Connector component	27
Figure 12: Functional architecture of the Unstructured data service	28





List of Tables

Table 1: Contribution of partners	9
Table 2: Relation to other activities in the project	10
Table 3: External services to be used for the structured data gathering system (in bold those mandatory for the simulations)	12
Table 4: Relation between the data obtained by the EnergyPlus weather database and the simulation of the Energy Conservation Measures of the OptEEmAL ECM Catalogue	23
Table 5: Functional requirements for structured data	24
Table 6: Functional requirements for unstructured data module	24
Table 7: Non-functional requirements for structured data	25
Table 8: Non-functional requirements for unstructured data	25
Table 9: Definition of data quality assessment levels for Geographical Representativeness (GR).	31
Table 10: Definition of the data quality assessment levels for Time Representativeness (TiR)	31
Table 11: Definition of the data quality assessment levels for Accuracy (A)	32
Table 12: Definition of the data quality assessment levels for Completeness (C)	32
Table 13: Definition of the data quality assessment for Reliability/Credibility (R)	32
Table 14: Overall data quality level according to the achieved aggregated data quality indicator	33
Table 15: Data quality assessment for EnergyPlus Weather data	34
Table 16: Data quality assessment process for energy prices data from Eurostat	34
Table 17: Data quality assessment process for income level data from Eurostat	35





Abbreviations and Acronyms

Acronym	Description
DPI	District Performance Indicator
DQR	Data Quality Rate
ECM	Energy Conservation Measure
EPW	Energy Plus Weather (data format suitable for energy plus energetic simulation)
JSON	Java Script Object Notation
LCA	Life Cycle Assessment
LOD	Linked Open Data
NUTS	Nomenclature of Territorial Units for Statistics
OptEEmAL	Optimised Energy Efficient Design Platform for Refurbishment at District Level.
OWL	Web Ontology Language
QOTF	Query On The Fly
REST	REpresentational State Transfer
RDF	Resource Description Framework





Executive Summary

This document is the final version of the D1.3 "Requirements and specification of geo-clustering data sets access module". Since in the intermediate version high level requirements have been presented, in this deliverable more specific requirements related to the technical side have been developed. From this point of view the objective is to go further and deeper into the concepts explained in the intermediate version document in order to fully specify the requirement set. This deliverable provides detailed information about the technical integration of geo-clustering data sets access module with the rest of the platform, as the main objective of this document is to provide all the concepts needed to support the development of this module in D1.4. For this purpose, the deliverable describes how the platform can answer automatically to climate and socio-economic queries, focusing on the following three aspects:

- The gathering, processing and integration in the tool of contextual data and metadata.
- The definition of the functional architecture for the geo-clustering data access module.
- The definition of the criteria to be used to assess data quality and the minimum requirements for data quality.

The document has been structured to accomplish the above mentioned goals. Since the contextual data that have to be gathered by the geo-clustering data sets access module can be structured or unstructured data, the sections 2 and 3 introduce the services used for this purpose. In particular, section 2 presents the services that will be used to gather information from external databases related to two aspects that are going to be of paramount importance for the calculation of the DPIs: climate and socio-economic data, while section 3 presents the service which provides a search engine for gathering unstructured information from the web to support the platform users in their refurbishment projects. Section 4 specifies functional and non-functional requirements both for structured and unstructured data as also data requirements coming from the simulation. Section 5 is devoted to explain the high level design of the geo-clustering data sets access module named "Geo-clustering connector" which is going to be in charge of gathering the contextual data. The Geoclustering connector consists of two different components: Geo Connector that gathers data from well-structured data sources, and Unstructured Data Connector that acts as an interface between the OptEEmAL graphical user interface and the Unstructured data service. The section presents the functional architectures of both the components and their integration into the OptEEmAL platform. Finally, section 6 introduces the Quality integration test needed to make possible the calculation of energy (for climate data) and socio-economic (for energy prices and income levels) DPIs with a sufficient level of quality.





1 Introduction

1.1 Purpose and target group

The purpose of this document is to set the requirements that the external data gathering system is going to have. This system is going to be divided into two well differentiated parts; one for the structured data that is going to be called "Geo Connector" and other for the unstructured data that will be called "Unstructured Data Connector". The two parts are so differentiated that will not share any software part but they are described together since both of them are components that will gather information from external sources. The geo connector will be in charge of retrieving information related to climate, weather, energy prices and incomes, while the unstructured data service will provide the user with complementary information to be deployed at the objective-definition stage. The information of more adequate ECMs depending on the location; while the complementary information obtained by the second component will cover several topics that could influence the objective definition.

To make the description of both components understandable, this document starts with a description of how the information will be gathered using the Geo Connector for the structured part, describing the Urls that have to be formed to call the REST service. It continues with a description of the functionality that will be provided by the "Unstructured Data Service" since, in this case, a prototype of the software is already developed and thus no description of the architecture or the internal details of it is needed (nevertheless it was done in the first iteration) but it needs the whole picture to integrate the module into the rest of the platform.

Next, the specification of the requirements is done dividing them into "functional" (related to functionalities) and "non-functional" (related to performance and behaviour of the system) for both structured and unstructured data. An evaluation of the data that can be obtained using the methods described in this document is done to ensure the usability of the sets.

1.2 Contributions of partners

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

Participant short name	Contributions	
ES	Main contributor of Section 1 (Introduction), 3 (Unstructured data service), 4 (Specification of Requirements), 5 (Design of the geo-clustering connector) and 7 (Conclusions)	
ACC	Main contributor of Section 1 (Introduction), Section 2 (Structured data service), 4 (Specification of Requirements) and 7 (Conclusions)	
UTRC	Contributor of Section 2.2 (External data sources to be accessed)	
NBK	Main contributor of Section 6 (Quality integration test)	
FUNITEC	Contributor of Section 5 (Design of geo-clustering connector)	
TUC	Contributor of Section 6.3 (Application to structured data sources)	

Table 1: Contribution of partners





OPTIMISED ENERGY EFFICIENT DESI

PLATFORM FOR REFURBIS

OptEEm

Т	FC

Contributor of Section 6.3 (Application to structured data sources)

1.3 Relation to other activities in the project

The following Table 2 depicts the main relationship of this deliverable to other deliverables developed within the OptEEmAL 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 Number	Contributions
D1.1	E-Guide on stakeholders' involvement and IPD implementation for the design and execution
D1.2	Requirements and specification of input data process to evaluate users' objectives and current conditions
D1.4	Geo-clustering data sets access module
D2.1	Requirements and specification for the District Data Model
D2.3	Functional architecture of the data repository
D5.2	Functional architecture Interfaces Overall Platform Design



2 Structured data service

2.1 Objectives

The main purpose of this service is to gather information from external databases about two aspects that are going to be of paramount importance for the calculation of the DPIs. The two data categories are:

- Climate data to calculate energetic demands. Weather typical data, which is related to climate and doesn't vary along time, will be deployed since there is software in the platform that is going to make use of it such as Energy Plus. This tool uses hourly weather typical data so that it will be needed to use also weather typical data extracted from a reliable source. Weather and climate are different, while the climate is supposed almost static, weather varies over time (it is dynamic). This is the reason why the usual weather databases are not sufficient for the simulations, the average weather over a year is needed which is not what usual weather websites offer, to have a good idea of what environmental conditions the buildings will have to deal with.
- Energy price data. It will be used to calculate the costs associated with the energy consumption (and thus, also to calculate payback periods). This value varies over time so that a good update and reliable source is needed which differentiates between, at least, gas and electricity prices. If this information is not available, it will be asked to the users, through a graphical user interface.
- Income levels will be used to calculate the energetic poverty rate of the habitants of the district.

There are two possibilities to use structured information inside the platform; to retrieve it "completely on demand", that is, not to store the information in any internal repository, or to store the information to have it "at hand" when it is needed. The second option is preferred because of the speed in which it could be deployed (it is much faster to retrieve the information from an internal database than from an external website and this process will be done several times in the same project), so that the possibility of retrieving the information just one time per project is a desirable feature. The module will put the information in a RDF file format associated with an OWL file that will contain the ontologies.

The information gathered in this process will serve on the one hand for the simulation module in order for the external tools invoked by platform to deploy the necessary data for the DPI calculation. And also for the identification of ECMs that can be directly affected by the data obtained by the geoclustering service and that can be identified as adequate or not for a certain location.

2.2 External data sources to be accessed

In the table below it is possible to find the two sources that the structured data service is going to deal with. In bold are indicated those values that are mandatory top run the simulations needed in the DPI calculations. Those that are not in bold and are included in the same data source can be retrieved using the same procedure but it is envisioned not to be needed.





System name	URL	Application zone and other considerations	Measure(s) offered
European Commission Eurostat	http://ec.europa.eu/eurostat /web/gisco/geodata The method for extracting the	Eurozone	Administrative units (NUTS, Urban Audit, Countries, communes)
GISCO	values is depicted in: http://ec.europa.eu/eurostat /web/json-and-unicode-web-		Population Distribution /Demography
	services/about-this-service		Transport networks
			Land cover
			Digital Elevation Model(DD LAEA)
			Aspect Slope,
			Coloured relief
			Hill shade
			Hydrography
			Energy prices (euro)
			Income level (histogram)
Energy Plus Weather	<u>https://www.energyplus.net/</u> <u>weather</u>	The entire world	Dry Bulb Temperature (°C)
website			Relative Humidity (%)
			Barometric Pressure (Pa)
			Wind Speed (m/s)
			Wind Direction (degrees)
			Global Horizontal Solar Radiation (W/m ²)
			Cloud Cover (tens)
Ge20	http://www.geoclusters.eu/g e20/	Eurozone KML format	Heating Degree Days
			Cooling Degree Days
			Annual incident energy on a south orientation
			Avg ambient temperature over year
			Average Cooling Seasonal External Air Temperature
			Average Heating Seasonal

Table 3: External services to be used for the structured data gathering system (in bold those mandatory for the simulations)





AT DISTRICT LEVEL





			Gas prices for household consumers
			Electricity prices for household consumers
			Disposable income of households
			Electricity consumption of households
			National Energy Regulations
			Solar Cooling Incenties
NOAA / NWS's	http://weather.noaa.gov/pub	Northern	Wind speed 10m kts
Meteorological Development	<u>/SL.us008001/ST.opnl/DF.gr</u> <u>2/DC.ndfd/AR.nhemi/</u>	Hemisphere	Wind gust kts
Laboratory	http://weather.noaa.gov/pub /SL.us008001/ST.opnl/DF.gr	It is possible to get historical records	Pressure (MSL)
	2/DC.ndfd/AR.nhemi/		Temperature
	Method explained in:	Grib format (condensed one)	Dew Point (2m) Min
	http://www.nws.noaa.gov/m dl/degrib/dataloc.php		Dew Point (2m) Max
			Precipitation mm/h
			Cloud cover
			Relative humidity (2m)
			Isotherm 0°C
			Snow depth cm
			CAPE J/Kg
			CIN J/Kg
Dermap		Europe	Efficiency Index
	<u>http://www.dermap.com/en/</u> <u>social-gis.html#</u>		Innovation Index
			Budget Import-Export
			Density of Population
			Price of eggs
			18 years old Students in Europe
			Female Employment
			Population
		World	Public Debt





	Control of corruption (control score, 1 to 10 corruption scale
	Longevity difference in the world
	Life expectancy in the World
	Agricultural land in the World
	Rooms per capita in the World

The list provided in the first iteration of the document about the data sources that are going to be used has been modified because since then, an in-depth study carried out identified some wrong assumptions that were made in the first iteration. The wrong assumptions were:

- Data needed were thought to be weather, but they must be climate related (difference explained in section 2.1).
- Ge20 REST system is not installed neither updated.

The options taken are described in the following sections.

2.2.1 Energy Plus Weather

The Energy Plus website has a database in which it is included representative climate data sets (of each day of the year) for Europe. A regular REST call is required to transform the files where the needed information is contained to the proper format. The aggregation level that can be reached is NUTS 2. The url that is going to be used to do the REST call has to be made using the following steps:

1. There is a part of the url that is common to all cases, it is:

https://www.energyplus.net/weather_location/europe_wmo_region_6

2. It is needed to add one specific part for the country, it has to be done using the *ISO* 3166-1 *ALPHA* 3 code system:

https://www.energyplus.net/weather_location/europe_wmo_region_6/ESP

3. The next part of the url is the one related to the region. To obtain it, it will be needed to add a string that is specified in the webpage (just by clicking in the previous link it is possible to see the complete list).

https://www.energyplus.net/weather_location/europe_wmo_region_6/ESP/ESP_San_Seba stian.080270_SWEC

4. In the last step it is needed to add the format in which we want the data. The most useful for us will be the epw, which is a plain text file with a well stablished format.

https://www.energyplus.net/weather_location/europe_wmo_region_6/ESP/ESP_San_Seba stian.080270_SWEC.epw

Once the url is formed, the file can be retrieved using a simple REST call like:

wget

https://www.energyplus.net/weather_location/europe_wmo_region_6/ESP/ESP_San_Sebastian.080270_SWE C.epw

2.2.2 Eurostat

Based on D1.2 section 6.3 the minimum socio-economic data required is energy cost and income level per inhabitant. One of the services identified in section 3.1 is <u>http://ec.europa.eu/eurostat</u>





which provides energy price data as required in <u>KW/hour \in per country</u> and income level in <u>inhabitant income \in per NUTS-2</u>.

To retrieve data programmatically Eurostat provides several REST services that can be called following the format depicted in Figure 1:

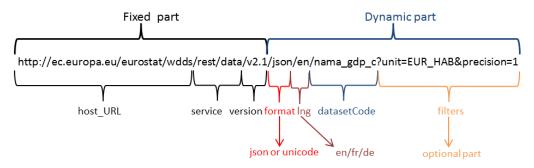


Figure 1: Eurostat REST request

A detailed explanation of this request can be found at link below:

http://ec.europa.eu/eurostat/web/json-and-unicode-web-services/getting-started/rest-request.

Sections "Retrieving Energy Prices" and "Retrieving Income Level" below explain in more detail which datasets will be used and how to infer all the parameters needed for the request. The list of datasets available in Eurostat is defined in http://ec.europa.eu/eurostat/data/database.

2.2.2.1 Retrieving Energy Prices

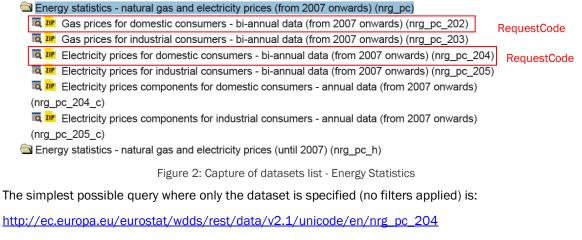
To retrieve energy prices is necessary to know the <u>country</u> where the design is made; this information should be defined already in OptEEmAL. The query format for countries requires the following ISO 3166-1 alpha-2 code representation. Country codes for OptEEmAL case studies are: Spain ES, Turkey TR, Unite Kingdom UK, Italy IT, and Sweden SE.

The datasets to query energy price are:

- 1. Gas Price for domestic consumers nrg_pc_202
- 2. Electricity Price for domestic consumers nrg_pc_204

🕰 🏧 Supply electricity - snort term monthly data (hrg_ind_342m)

Energy statistics - prices of natural gas and electricity (nrg_price)



http://ec.europa.eu/eurostat/wdds/rest/data/v2.1/unicode/en/nrg_pc_202

The response received follows the next format:





File Home Insert Page Layout Formulas	Data Revie	ew View	Team		-			nrg_pc	202.tsv - Micro	osoft Excel	-					_	1.1.1		
Arial - 11 - A a		· *	Wrap Text	G	neral			N	ormal	Bad	Goo	d	Neutral	Calcu	lation	- Hen (* 🖿	Σ AutoSum	· 🎢 🕯
Paste B I U - R - Sr - A -		(m (m	Merge &	Center + 🛄	- % ,		onditional F		heck Cell	Explanat	ny. Inpu		Linked Cell	Note		Insert D	lelete Format	Fill -	Sort & Fin
						Fi	ormatting * as	Table -			Styles	_		Location .		1		Clear -	Filter * Sele
	x	Alignme		14	Number	04					Styles						Cells		diting
🖬 🗅 🐸 😂 🗣 🔹 🖓 · 🗠 🖓 M · 🛄 🖽 · 🛄 🗇 Q.			• 🔐 🖌 🐜	XI 21 12	・日か同	옛 글 규	4												
A303 • fx 4100.414190	1.KWH.I TA	X.EUR.ES	S															12	
1 product,consom,unit,tax,currency,geoltime	2015S2	2015S1	2014S2	2014S1	2013S2	2013S1	2012S2	201251	2011S2	2011S1	2010S2	2010S1 2	00952 20	009S1 20	08S2 20	08S1 200	07S2 200	7S1	T
2/9 4100,4141901,GJ GCV X VAL PPS II	31.0217	22.1943	32.4978	23.9798	29.124	22.0013	29.9310	20.74	25.676	18:2401	23.0902	10.0893	15.8733	18:2204	21.3488	10.0845	17.1517	10.0859	
280 4100,4141901,GJ GCV,X VAT,PPS,LT	28.3952	28.7351	28.2669	31.4124	31.5965	33.2165	33.0377	28.757	3 28.4327	25.266	26.7708	24.4238	23.0612	26.2893	20.3953	19.0499	14,7362		
281 4100,4141901,GJ_GCV,X_VAT,PPS,LU	12.7534	12.7083	13.8827	15.1494	13.9659	17.4388	15.6976	12.78	15.926	14.8632	11.892	10.8575	15.7735	16.2213	14.8617	11.2411	12.173		
282 4100,4141901,GJ GCV,X VAT,PPS,LV	25.3934	25.4153	24.9195	25.3132	25.2888	25.5769	27.8286	26.078	9 24.918	23.4996	25.6474	21.3945	24.2097	29.6959	25.7776	12.1365	11.73		
283 4100,4141901,GJ_GCV,X_VAT,PPS,NL	25.7748	22.529	25.733	23.8289	27.9117	23.7564	26.9649	23.281	6 24.9736	21.4566	23.7757	19.9119	24.3608	24.3161	24.2989	22.2956	23.8282	24.4244	
284 4100,4141901,GJ_GCV,X_VAT,PPS,PL	23.758	25.0481	25.2918	24.3459	23.2441	26.0589	27.2281	23.679	5 23.5955	22.9267	22.7704	21.743	21.2569	21.3254	21.9571	18.5234	19.2518		
285 4100,4141901,GJ GCV,X VAT,PPS,PT	34.201	34.3871	36.4567	33.2173	33.6641	30.5862	30.6201	27.081	2 28.1571	25.1124	26.1066	23.5802	24.3975	24.7902	25.7584	25.8077	27.1055		
286 4100,4141901,GJ GCV,X VAT,PPS,RO	15.5291	13,7208	13,9309	13.5983	13.9275	12,7797	13,1393	12 593	2 12 6783	12.6973	12.84	12.8427	12,7085	13.8197	14.3153	13.9505	14.4843		
287 4100.4141901.GJ GCV.X VAT.PPS.RS	28.297	28,951	29.0468	26,1631	29.7785	25.7707													
288 4100,4141901,GJ GCV,X VAT,PPS,SE	32,5348	29.8254	30,1385	28.0649	28.0316	29.0468	33.8155	31.780	3 31.3211	29.5918	29,1157	27.6935	27.5126	24,9907	28.0417	20.0687	20.3748	24.2304	
289 4100,4141901,GJ GCV,X VAT,PPS,SI	19,7707	21,2429	18.8784	23.207	19.1896	23.8224	27,2735	27.66	9 24,7188	22 205	20.3961	19.5808	20.6945	21.78	25.5294	19.4087	18.6445		
290 4100.4141901.GJ GCV.X VAT.PPS.SK	36 3835	36 2167	36.0516	35,3311	35 5892	34,4853	33,7689	34.149	6 34.17	33 3842	33,7041	33 9225	34,1269	34,358	32,4808	32,7409	27,9928		
291 4100,4141901,GJ GCV,X VAT,PPS,TR	16.8536	16.1758	15.9426	14.8021	15.8085	15.4214	16,1865	13.55	5 13.3431	11,9509	12.5611	12.6751	12,7639	16,7733	18,9618	12.8443	12 6326		
292 4100,4141901,GJ GCV,X VAT,PPS,UK	19.8019	17.195	21.6619	18,4942	18.6913	14.6374	15.6944	14.119				11.6469	12,7914	12,9473	13,1925	10.9096	8.9242		
293 4100.4141901.KWH.I TAX.EUR.AT	0.0861	0.0879	0.0887	0.0902	0.0904	0.0918	0.0919					0.0756	0.0758	0.0785	0.0752	0.0729	0.0753		
294 4100,4141901,KWH,I TAX,EUR,BA	0.0514	0.0514	0.0512	0.0512	0.0512	0.0555	0.0555	0.055				-							
295 4100,4141901,KWH,I TAX,EUR,BE	0.0907	0.0794	0.0929	0.0862	0.0895	0.0924	0.1028	0.093		0.0873		0.0798	0.0789	0.088	0.1013	0.086	0.0773		
296 4100.4141901.KWH.I TAX.EUR.BG	0.0432	0.0492	0.0517	0.0483	0.0551	0.0497	0.0554	0.049		0.0427		0.0363	0.0338	0.0477	0.0417	0.0324	0.0306		
297 4100.4141901.KWH.I TAX.EUR.CZ	0.091	0.0895	0.0861	0.0855	0.0904	0.097	0.0945	0.094				0.0741	0.0706	0.0719	0.0757	0.0657	0.0544		
298 4100,4141901,KWH,I TAX,EUR,DE	0.106	0.1088	0.1063	0.1111	0.1112	0.1067	0.1052	0.10				0.0914	0.0945	0.1004	0.1087	0.0944	0.0935		
299 4100.4141901.KWH.I TAX.EUR.DK	0.0764	0.0802	0.0878	0.0908	0.0977	0.0993	0.0955	0.05			0.0965	0.0954	0.0851	0.0803	0.0979	0.0959	0.0885	0.0406	
300 4100,4141901,KWH,I TAX,EUR,EA	0.1208	0.1049	0.1233	0.1000	0.1204	0.1046	0.1174	0.00				0.0845	0.0916	0.0933	0.1015	0.0859	0.0867	0.0100	
301 4100,4141901,KWH,I TAX,EUR,EE	0.0414	0.0545	0.0548	0.0							4	0.0393	0.0375	0.0407	0.0408	0.0392	0.0386		
302 4100.4141901.KWH.I TAX.EUR.EL	0.116	98	0.1157	0.0	<u>14</u>	02	FL	(\\	/Ho	1.1.12	E I				0.0100		0.0000		
303 4100.4141901.KWH.I TAX.EUR.ES	0.1192	-NZO	0.1221	0.0	J. I	192	τr	\ V V	100	u	6	0.0681	0.0686	0.0737	0.078	0.0707	0.0695		
304 4100.4141901.KWH.I TAX.EUR.EU27	0.1084	0.0546	0.1098	0.0							i i i i i i i i i i i i i i i i i i i	0.0721	0.0769	0.0779	0.0841	0.0718	0.07		
305 4100.4141901.KWH.I TAX.EUR.EU28	0.108	0.0943	0.1095		mal	idi		-	an f	~ *	5	0.0719	0.0766	0.0777	0.0838	0.0716	0.0698		
306 4100 4141901 KWH TAX EUR FR	0.1487	0.1273	0.1512	0.1	IICI.	uai	na i	ах	es f	OL.	ă	0.0981	0.1148	0.0992	0.1124	0.0906	0.094		
307 4100,4141901,KWH,I TAX,EUR,HR	0.055	0.0553	0.0537	0.0			.9				8	0.0383	0.0328	0.0319	0.0277	0.0273	0.0274	0.0295	
308 4100,4141901,KWH,I TAX,EUR,HU	0.0372	0.0381	0.0377		-	-		1		- 5	4	0.0571	0.0515	0.0484	0.0494	0.0213	0.038	0.0200	
309 4100,4141901,KWH,I TAX,EUR,IE	0.0886	0.0748	0.09	0	ne	sec	con	n r	art	OT	6	0.0549	0.0632	0.0691	0.0709	0.0685	0.1085		
310 4100,4141901,KWH,I TAX,EUR,IT	0.131	0.0944	0.137	0.0		000		~ ~		0.	1	0.0664	0.0652	0.0748	0.0854	0.0668	0.0684	0.0666	
311 4100.4141901.KWH.I TAX.EUR.LI	0.1715	0.167	0.137		204	-					1	0.0004	0.0002	0.0740	0.0004	0.0000	0.0004	0.0000	
312 4100,4141901,KWH,I_TAX,EUR,LT	0.0746	0.0771	0.0755	0.0	201	5					0	0.0629	0.0619	0.0692	0.0546	0.051	0.036		
313 4100,4141901,KWH,I_TAX,EUR,LU	0.0599	0.0593	0.0636	0.0	201	0					3	0.0505	0.0734	0.0692	0.0661	0.051	0.0531		
314 4100,4141901,KWH,I_TAX,EUR,LU	0.0599	0.0593	0.0030	0.0750	0.0702	00075	0.0830	0.077	0.0733	0.0055	0.0003	0.0505	0.0654	0.0749	0.0696	0.05	0.0332		
315 4100,4141901,KWH,I TAX,EUR,MD	0.0747	0.0759	0.0/44	0.0750	0.0702	0.0113	0.0030	0.070	0.0755	0.0033	0.0003	0.0004	0.0034	0.0001	0.0090	0.0332	0.0332		

Figure 3: Eurostat gas price response

2.2.2.2 Retrieving Income Level

As per income level, Eurostat provides a wide range of data for income base on different metrics. The one that best fits OptEEmAL requirements is "income of households by NUTS 2 regions".

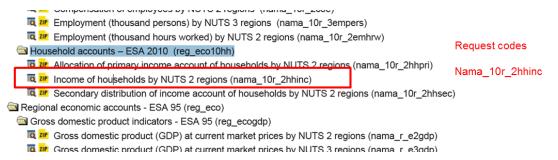


Figure 4: Capture of datasets list - Energy Statistics

To filter this data, the NUTS 2 identifier based on building location is necessary. Eurostat provides information that relates address with NUTS level per country. These data can be retrieved from http://ec.europa.eu/eurostat/tercet/flatfiles.do . The specific file per country can be retrieved programmatically being defined the country code. As an example, the following figure shows how the NUTS 2 code can be obtained for one of the case study (Historic city District, Santiago de Compostela, Spain).





Tile Home Insert PageLayout Formulas Data Review View Team	au a substanti - Microsoft DCB
💦 & Cut Anat - 11 - A' * = = 🚽 🗞 - 🚍 Wrap Text General -	Normal Bad Good Neutral Calculation
Pate Standberge & Center · · · · · · · · · · · · · · · · · · ·	Conditional Format Philod Coll Explanatory, Input Linked Cell Note Insert Delete Format
· · · · · · · · · · · · · · · · · · ·	Formatting + as Table + the second se
Clipboard is Font is Alignment is Number	Styles Cells
四口感効的な、6、2004、回知・回回の回・服務令・回想・5×202331日、コッロッコ	17.*
A715 • f 15078 Santiago de Compostela Santiago de Compostela ES111	
1 LAU_NAT_CODE:LAU_NAME_1:LAU_NAME_2_LAT:NUTS_3	M N O P
1 DAU_NAT_CODE,DAU_NAME_T,DAU_NAME_Z_DAT,NUTS_3	
689 15052 Mini-aMini-aES111	
690 15053 Muros Muros ES111	
691 15054,NarÅ'n,NarÅ'n,ES111	
692 15055.Neda,Neda,ES111	
693 15056/Negreira/Negreira/ES111	
594 15057,Noia,Noia,ES111 595 15058,Oteiros,Oteiros,ES111	
696 15059 Ordes Cordes ES111	
697 15050.0roso.0roso.ES111	
598 15051,Ot5gueira,Ot5gueira,ES111	
699 15052.Outes,Outes,ES111	
700 15063,0za dos RÃ-os;Oza dos RÃ-os;ES111 701 15064 Pademe:Pademe ES111	
701 15054 Pademe Pademe E 3111 702 15055 Padrám Padrám E 3111	
703 15066 Pino, O.Pino, O.ES111	
704 15057 Pobra do CaramiÁ±al, APobra do CaramiÁ±al, AES111	
705 15058,Ponteceso,Ponteceso,ES111	
705 15050 Pontedeume Pontedeume ES111 707 15070 Pontes de GarcÃ-a RodrÃ-guez, As Pontes de GarcÃ-a RodrÃ-guez, As ES111	
707 15070,Pontes de GarcA-a RodrA-guez, As,Pontes de GarcA-a RodrA-guez, As,ES111 708 15071 Ponto do Son Porto do Son ES111	
709 15072 Rianzo Rianzo ES111	
710 15073 Ribeira Ribeira ES111	Santiago de Compostela,
711 15074.Rois.Rois.ES111	
712 15075.Sada.Sada.E8111	Carriage de Compositiona,
713 15076 San SadumiÃeo San SadumiÃeo ES111 714 15077 Santa Comba Santa Comba ES111	
716 15077 Santa Compassanta Compastela Santa o de Compostela ES111	neetal and 15079
	postal code 15078
717 15080,Sobrado;Sobrado;ES111	
718 15081,Somozas, As,Somozas, As,ES111	•
719 15082.Teo.Teo.ES111	ES111 = NUTS 3
720 15083,Toques,Toques,ES111 721 15084,Tordola,Tordola,ES111	ESTTENUIS 3
722 15085 Touro, E0111	
723 15086 Trazo, Trazo, ES111	
724 15087,ValdovÅzo,ValdovÅzo,ES111	
725 15088,val do Dubra;Val do Dubra;ES111 726 15089 Vedra;Vedra;ES111	ES11 = NUTS 2
727 15090 Wilasantar Wilasantar ES111	
728 15091 Vilarmaior Vilarmaior ES111	
729 15092 Vimianzo (Vimianzo ES111	
730 15093,Zas,Zas,ES111	
731 15901.CariÃzo.CariÃzo.ES111 732 16001.Abia de la ObispalÃ-a.Abia de la ObispalÃ-a.ES423	
732 16001,Acia de la Ocispali-a,Acia de la Ocispali-a,ES423 733 16002,AcetriÁ'n, El AcetriÁ'n, El ES423	
734 16003.Alarok ¹ n.Eladok n.Eladok s	
735 15004 Albaladejo del Cuende Albaladejo del Cuende ES423	
736 16005,4lbalate de las Nogueras,4lbalate de las Nogueras,ES423	
737 16006 Albendea Albendea ES423	
738 16007 Alberca de Zâjncara, La Alberca de Zâjncara, La (ES423 739 16008 Alcalár de la Vega Alcalár de la Vega (ES423	
739 16008,Alcaliv, de la vega,Alcaliv, de la vega,ES423 740 16009,Alcantud,Alcantud,ES423	
741 16010 AloAgar del ReyAloAgar del ReyES423	
742 16011 Alcohuiate Alcohuiate ES423	
743 16012 Alconchel de la Estrella Alconchel de la Estrella ES423	

Figure 5: File obtained from Eurostat and codes

http://ec.europa.eu/eurostat/tercet/download.do?file=pc_es_NUTS-2013.txt

Further information about NUTS can be obtained from:

http://ec.europa.eu/eurostat/web/nuts/correspondence-tables/postcodes-and-nuts

The simplest possible query where only the dataset is specified (no filters applied) is:

http://ec.europa.eu/eurostat/wdds/rest/data/v2.1/unicode/en/nama_10r_2hhinc

The response received follows the next format:

Ne Home Insert Pag	e Layout Formulas Di	ata Review	View Team	-	-	-		tg	00036.tsv - Mich	soft Excel	-	_	-	_	-	-		
Arial Arial Arial Arial B I	· n · A* x* u · ⊞ · ゐ · ▲ ·	= = = *	Wrap		1.111	34.21	Conditional Formatting *	Format as Table -	Normal Check Cell	Bad Explana		Good Input	Neutral Linked Cell	Calculation			te Format	N IN Y
Ciphent 1	Fort		D Kar Sr &	*au 7,1 2,1	Numb Li- Li &		U v				50/	es.				. Cel	5	
A103 • 🤄	A PPCS_HAB.B	6N,ES11												N	0	P		ä
unit na item geo'time	А		2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	.14	0			w
NTA UND DAU ATU			10000	47000	10000	00000	00000	00000	00000	00000	04400	00400	01000					
PPCS_HAB_B5N_DK05			13200	13900	14400	15500	16200	16200	15500	16300	18500	17300	17400					
PPCS_HAB,B5N,EE00			5500	6100	6800	7700	9000	9400	8000	8200	9100		10700					
PPCS_HAB, B5N, EL11						12300	12200	12100	11700	10800	9600							
PPCS_HAB, B5N, EL12						12500	13500	13800	13100	12200	11200							
PPCS_HAB,B5N,EL13						12200	13200	13500	13200	12400	11600							
PPCS_HAB,B5N,EL14						12400	13200	14500	13200	11900	10700							
PPCS_HAB,B5N,EL21					_	A A A A A A A A A A A A A A A A A A A	50300			10.108	100	10800 :						
PPCS_HAB, B5N, EL22	2011 (10 10 10 10 10 10 10 10 10 10 10 10 10	and the second second	1.1.1	10.00	1.00	-	1000	1.00			100	11900						
PPCS_HAB, B5N, EL23	NUT	C	-0	11	10	\mathbf{n}	nn	C	1.10	0.0	500	9700						
PPCS_HAB,B5N,EL24		D				9		Ŧ	IVe	ar	100	9400						
PPCS_HAB,B5N,EL25			-0		1 4	-0		-	, , , ,	a	800	10600	14300					
PPCS_HAB,B5N,EL30 PPCS_HAB,B5N,EL41									-		800		14300					
PPCS_HAB,B5N,EL41 PPCS_HAB,B5N,EL42			6 - E			13800	14700	14900	14400	15500	14400		12500					
PPCS_HAB.B5N.EL42 PPCS_HAB.B5N.EL43						13800	14400	14900		11700	10400		9700					
PPCS HAB B5N EL51						12500	12400	12300		10900	9600		9000					
PPCS_HAB,B5N,EL52						12800	13800	13900	1330	12400	11100		10400					
PPCS HAB, B5N, EL53						12500	13400	13700		800	11500		11400					
PPCS HAB B5N EL54						11900	12800	13100	13000	120	11100		10300					
PPCS HAB.85N.EL61						12700	13400	14600	13400	12000	8700	10200	9900					
PPCS HAB B5N EL62						13300	14200	13900		13500		11800	11600					
PPCS_HAB.B5N.EL63						12100	12700	13300	12600	11500	10100	94	9100					
PPCS HAB, B5N, EL64						12800	13700	13600	13300	12500	11000		9000					
PPCS HAB B5N EL65						12800	13600	13600		12600	11000		10100					
PPCS HAB, B5N, ES11			6	6	:	-	:		-	13300	13300		12900					
PPCS_HAB,B5N,ES13				1					8	14500	14500	13900	13900					
PPCS_HAB, B5N, ES21										19900	19900	19600	19600					
PPCS_HAB, B5N, ES22										19600	19800	19200	19000					
PPCS_HAB,B5N,ES23										16000	16200	15700	15800					
PPCS_HAB,B5N,ES24										16800	16800		16300					
PPCS_HAB,B5N,ES30										21000	21300		20900					
PPCS_HAB,B5N,ES41										14600	14600		14200					
PPCS_HAB, B5N, ES42										12700	12500		12100					
PPCS_HAB,B5N,ES43										10900	10700		10500					
PPCS_HAB,B5N,ES51										18200	18100		17900					
PPCS_HAB,B5N,ES52										13300	13300		12800					
PPCS HAB, B5N, ES53										15900	16000	15600	15700					
PPCS HAB B5N ES61										11500	11500	10900	11000					

Figure 6: File obtained from Eurostat and codes





2.3 Conclusions for structured data gathering module

There is some confusion about what weather and climate data are, in the description of actions of this project too. This confusion is due to the fact that both concepts are related to the same phenomenon, the environmental conditions. To make a proper simulation, it will be needed to use data that is representative of a specific geographical location. Being representative does not mean to use real weather but to use data that is probabilistically the median of the measures, not the average. This concept is better called as "climate data" using the common definition of it but clearly it is on the border of the two concepts. Fortunately, there is a service set of representative weather data for each day of the year and with the data format needed for the simulation since it is provided by the site that hosts the software, <u>www.energyplus.net</u>.

All socio-economic data required can be retrieved from Eurostat programmatically as defined in this section. It is also possible to get energy prices that can be retrieved at country level as required in D1.2 by the DoA, income level per inhabitant is extracted at NUTS 2, what provides a better granularity than the required.





3 Unstructured data service

In order to support the end-users of the platform in their refurbishment projects, this service, named *Query On The Fly (QOTF)*, provides a search engine for unstructured data that allows them to find useful information in the web. This information will be deployed at the objective definition stage as an external service to aid the user in their decision-making process. The search engine helps users of the platform to identify the correct information in different contents that will be retrieved from the Web to support the evaluation of the current scenario also using a set of semantic metadata to filter the information. It will be possible, for example, for a user to insert in the search box the term "average U-value" and find 100 related documents then, using the semantic metadata as a filter, it will be possible to select a specific "location" reducing to 25 the number of related documents, then to add an "organization" and so on until the user finds exactly the 2-3 useful documents he was looking for.

The query on the fly search allows the user to experience a new way of surfing the web. By applying semantic analysis on the fly to search engines results (Google, Bing and Yahoo), the user is able to filter and explore the vast quantity of results near real-time, finding the information faster and with increased accuracy. With this tool it will be possible, for example, to find quickly the prices of the houses in a particular area or the average U-value in that area.

A district analysis can be performed with the following steps:

- 1. The user inserts a query in the search box, selects one or more search engine, the language of the information he would like to search and how many documents to consider in the analysis.
- 2. The QOTF acquires the results from the search engines, analyzes the snippets (produced by the search engines) and represents them semantically.
- 3. In the left column of the GUI the main semantic tags (topics, entities, relationships, emotions, etc.) have been published. Now, the user can explore the data of interest, select one of the tags to add a new constraint to the search or hide some results he is not interested in.
- 4. The user can also select one of the tabs at the bottom of the page (Topics, People, Domain Entities, etc.) for another view of the content that can offer a useful and alternative visualization of the information.
- 5. The user can extend the search to other semantic engines at any time by simply clicking on the specific check box in the search bar. In this scenario new data will be acquired and indexed.

3.1 Usage examples

To have a better understanding of the utility of the unstructured data service, some examples have been reported. Section 4 of D1.2 contains a basic description of the six case studies considered in the project, selected to validate the OptEEmAL platform. These six case studies reflect a decision-making scenario into which the OptEEmAL platform can be used to ease decision making. In the earlier phase of the project it was necessary to gather information about all the case studies regarding the initial situation of the district, the goal of future platform users, all available data, etc. In addition to these data, it can be useful to exploit the presented QOTF service, to increase knowledge about the district, gathering via web, news and topical information related to the district.

In the following figure is reported a usage example of the QOTF for the case study 1. "Cuatro de Marzo"; as shown, the QOTF results report information about "alternative energy" or "renewable energy" related to the analysed district.





Intelligence Taxonomy		Curro de marzo	۹
B Economy, Eucliness and Finance	0	🖉 🗟 Google 🖌 🐱 Bing 🖌 🖬 Yahon Language Spanish 🔹 Download. 50 Web	
Company Information	0	and the second sec	
El Energy and Resource	0	About 15 results Any Sources •	
Atternative Energy	0	About 15 results Any Sources •	
Economy (general)	0		
Environmental Issue	0	Q Search results for Custors de marzo X Valladolid X Reset	
Renewable Energy	0		
B Health	0	About search terms (Aprile version)	
B Labour	0	Custre de morzo	
Weather	0	Corre de mars	Partie of
T. Blur		CITIES	Pagetar may) 4
		Energy retrofitting in the district "Cuatro de	rly zero energy CIT
Places		The design of Course de Marzo's located in the south part of the Valladolid urban area. It was proverted	AND DESCRIPTION OF TAXABLE PARTY OF TAXA
			Addressing Provide
Valladolid	Ø	in 1955 as a residential area at the end	
Valladolid Campo Grande	0	in 1955 as a residential area at the end. @See full document: Effortis	Valuetated
		in 1955 as a residential area at the end 2 See full document: Efforties: Efforties: Efforts: Efforts: Efforts: Efforts: Efforts: Efforts: Efforts: Effort: Cuatro de Marzo', Valladolid	Valuested
Campo Grande Morava	0	In 1955 as a residential area at the end. In 1955 as a residential area at the end. In See half document, Ill Endlish Ill Topica III facts IIII facts III facts III facts III facts IIII facts IIIII facts IIII facts IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	Valuested
Campo Grande	0	In 1955 as a residential same at the end. Image: Start Artificity Country Count	Valuented Energy retroloting to the de "Control of Narra" Walked The scattering Energy Annual and service
Campo Grande Moraila Castilla León	0	In 1956 as a resolutional area at the end Size full document ill fontiles ill Tapics ill Facts ill Encotors ill Mesadeta Per Rent of Floors In Stroll Zorrita, Campo Grande, Four	Valuented Toroge of electronic colors "Control to Marcine" Solitocol The Society of
Campo Grande Morava Caetilia	0	In 1955 as a residential area at the end See full document: Efforties: Ef	Valuation Energy end descape of the de- Control de Marcall Valuation The facility of the descape Energy Annual and savings Valuations, faces
Carpos Grande Maraua Castilla Leden T-Pitter	0	In 1955 as a residential area at the end. Ber half document, E findles E Topics E facts E functions Material Rent of floors in Stroll Zornilla, Campo Grande, Four (March) (Sppt) Valiadolid See half document (Origonal language Spanne) E facts E Topics E facts E functions Materials Mat	Valiandar Bergerstenderung einen Ocean de Maria Dabate Forstender Maria Dabate Bergerstemmer der Gartige Halenner fam Maria Dabate Valiadate Rese Dabate Valiadate
Campo Grance Morsus León Y Filter Geography Taxonomy	0	In 1956 a a resolutional area as the end ≥ See full documents: Effortise III Topics III Facts III Encoders III Metadates III Facts of Honora Stroll Zorrilla, Campo Grande, Four, onor of Hart Stroll Zorrilla, Campo Grande, Four, onor of Hart Stroll Zorrilla, Campo Grande, Four, See full documents (Organis language: Spanne) III Fottises III Topics III Facts III Encoders IIII Metadate IIII Study Four of March 2007 S.I - Assor Study Four of March 2007 S.I - Assor St	Valiandad Barrag raitrofilming exities on "Care de Millouis" Valiand Tra-tearing Barrage Annuel en Lawing Valiantes familia Neuros traits Paras 1918 Barlagi barrage familia de annuel Barlagi barrage familia de annuel Barlagi barrage familia de annuel
Carego Grander Marsus León: T Filter: Geography Taxonomy Spain:	000000000000000000000000000000000000000	In 1958 as a residential tanks at the end Bise full document: Efforties El Tapica El Facta El Encottons El Metadetas Encort of Floors in Stroll Zorrilla, Campo Grande, Four- 	Valiandar Bergerstenderung einen Ocean de Maria Dabate Forstender Maria Dabate Bergerstemmer der Gartige Halenner fam Maria Dabate Valiadate Rese Dabate Valiadate
Campa Galaka Marasa Castila Leistia Leisti T Filter Geography Taxonomy Geography Taxonomy Autora	000000000000000000000000000000000000000	In 1955 as a residential area at the end Be found document: We find the weat the find the set of the set o	Voluence Registrations of the set of the set
Campo Grande Maraos Castila Lede : T-Fitur	000000000000000000000000000000000000000	In 1986 as a resolution large star the end Be fould document if fortises in Explose if facts if fa	Valuation The control of the contro
Campa Galaka Marasa Castila Leistia Leisti T Filter Geography Taxonomy Geography Taxonomy Autora	000000000000000000000000000000000000000	In 1958 as a residential area at the end Be full document: Efforties in Explore and Facts in Efforts in Efforts in Efforts in Explore and March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March (Capta) Validadid. Be full document: Stand Zornilla, Campo Grande, Four (March 2007), Sci Axesor Stand Be full document: Stand Zornilla, Campo Grande, Stand March 2007, Validadid. Be full document: Stand Zornilla, Campo Grande, Stand March 2007, Sci Axesor Stand Benetices (Stand Harrish), Efforties Efforts Efforts Efforts Efforts Efforts Stand Harrish (Stand Harrish), Capta Stand Harrish (Stand	Voluenced Registrationage states for the second second second second second The second second second second second The second second second second second Market Second Seco

Figure 7: QOTF results for Cuatro de Marzo District

In the figure below is reported a usage example of the QOTF for the case study 6. "Sneinton District – Nottingham"; the QOTF results report information about "critical infrastructures" and "civil infrastructures" related to the analysed district.

Highlights Topics People C	rganizations Domain Entities Geography Facts Tagging Relationships Emotions Quick Start Guide Sign	In
Facets	Search	
Geography Taxonomy	Sneirton	
United Kingdom	Sociel V Bing Vahoo Language English V Download So Web	
T Filter	n Badk Riften	
	5 results Any Sources •	
Places		
Nottingham	Q Search results for Sneinton X Nottingham X Economy, Business and Finance X Critical Infrastructures X Reset	
England	0	
Europe	About search terms (Alpha version)	
United Kingdom	3 Sheintan	
T Fiber		
Intelligence Taxonomy	🖬 Green's Mill, Sneinton - Wikipedia, the free	
Economy, Business and Finance	Il is a restored and working 19th century tower windmill in Sneinton, Nottingham	
Critical Infrastructures	③	
Transport (infrastructures)	0	
Industrial plants	In the Great Sneinton Bake Off Sneinton Nottingham	
Civil infrastructures	I.Sneinton, Green's Windmill & Science Centre, Windmill Lane, Sneinton, Nottingham	
Environmental Issue	② See full document III Entities III Topics III Facts III Emotions III Metadata	
Renewable Energy	0	
T Filter	Green's Mill, Sneinton - Wikipedia, the free encyclopedia	
	ll is a restored and working 19th century tower windmill in Sneinton, Nottingham	
Organizations	See full document Ⅲ Entities Ⅲ Topics Ⅲ Facts Ⅲ Emotions Ⅲ Metadata	
Sneinton		
T Filter	Sneinton Map United Kingdom Google Satellite Maps Les of Sneinton original name: Sneinton geographical location: City of Notlingham.	
	See full document III Entities III Topics III Facts III Entities III Metadata	
Buildings		
windmill	Sneinton Map United Kingdom Google	

Figure 8: QOTF results for Sneinton District



OptEEmAl

OPTIMISED ENERGY EFFICIENT DESIGN Platform for refurbishment

AT DISTRICT LEVEL



4 Specification of Requirements

4.1 Data requirements coming from the simulation

Since weather data will be retrieved exclusively from the EnergyPlus weather database, they will comply with the EnergyPlus Weather (.epw) format file, an ASCII, csv format file containing the hourly or sub-hourly weather data needed by the EnergyPlus simulation tool. However, these data are going to be used by other simulation tools as well, where the weather format file may differ. For instance, CitySim requires the weather data to be provided according to the Climate (.cli) format file, starting with a header that contains the city, and its geographical position, followed by the meteorological data for a year, organized by day, month and hour.

Hence, as mentioned above, in the attempt to store the data retrieved in a standard (unified) format, the weather data will be converted to an RDF format file, to be stored in the contextual repository through the corresponding connector of the communication logic layer (see figure 9). Once the information has been stored in the repository, the data management module maps the information in order to create one instance of the baseline scenario. The weather data retrieved are part of the simulation data models generated by the data integration component and stored in the project repository.

This information will be retrieved by the simulation model input generator module which will be in charge of configuring the simulation files and launching the simulation tools. Concerning the weather data retrieved, the simulation files configuration refers to the respective weather files generation, where certain transformation rules (refer to D4.4: Requirements and design of the simulation model input generator module, section 6.2.2) will be introduced to convert RDF format files to EPW, cli or other format files.

Table 4 below provides the relationships between data from EnergyPlus Weather and the ECM catalogue as an example of how the structured data will be used by the platform. Also, after the table, the relationship between EUROSTAT energy prices and the OptEEmAL economic assessment is provided.





Table 4: Relation between the data obtained by the EnergyPlus weather database and the simulation of the Energy Conservation Measures of the OptEEmAL ECM Catalogue

	Dry Bulb Temperature (°C)	Relative Humidity (%)	Barometric Pressure (Pa)	Wind Speed (m/s)	Wind Direction (degrees)	Global Horizontal Solar Radiation (W/m²)	Cloud Cover (tens)
A1.1-Ventilated Facade	Х	Х	Х	Х	Х		
A1.2-ETICS – External Thermal Insulation Composite System	х	х	х	х	х		
A2.1-Protected Wall – Internal lining	Х	Х	Х	Х	Х		
A3.1-Cavity wall – Air chamber insulation	Х	Х	Х	Х	Х		
B1-Roof	Х	Х	Х	Х	Х		
C1-Slab	Х	Х	Х	Х	Х		
D1 – Window replacement	Х	Х	Х	Х	Х	Х	Х
Biomass boiler							
ST-Flat collector						Х	Х
ST-Tube collector						Х	Х
PV-Mono-crystalline						Х	Х
PV-Multi-crystalline						Х	Х
Wind turbines					Х		
Geothermal-horizontal	Х	Х					
Geothermal-vertical	Х	Х					
High efficient boiler							
Condensation boiler							
Cogeneration							
High efficient chiller (electricity)							
High efficient heat Pump							
Energy storage-water tank	Х	Х					
Energy storage-Phase change materials units	Х	Х					
Reduction of distribution losses	Х	Х					
System Scheduling							
Optimal start-up shut down							
Weather compensation							
Load following							
Optimization-based control							





Regarding the economic assessment, the database of Eurostat is going to be used to define the Energy Price (EP) economic parameter. In this way, the following equation will make possible the calculation of the initial and refurbished operational stages economic impact:

$$B6_ec = \left(\frac{ED_b}{\rho} - RE_k\right) \cdot \frac{EP_y}{FU}$$

Where:

- B6_ec → Baseline or refurbished building operational energy use stages economic impact (€)
- $ED_b \rightarrow Baseline \text{ or refurbished building operational annual energy demand}$
- $\rho \rho \rightarrow$ Performance of the energy generation system (%)
- $RE_k \rightarrow$ Renewable energy generated by refurbishment strategies
- EPy \rightarrow Price of the energy source (\notin /kWh)
- FU → Functional Unit of the case study

The first factor refers to the energy demanded from the external net (without internal energy renewable sources) and the second to the price the energy has. Taking this into account, this equation can be identified with the one provided in D2.2 for the District performance indicator called ECO 01:

Op.energy cost = Energy consumption $(kWh/year) \times energy \cos t (\in /kWh)(perfuel)$

4.2 Functional requirements definition

4.2.1 Structured Data

Table 5: Functional requirements for structured data

REQUIREMENT ID	DESCRIPTION
FR1	The system has to be able to set up a url like it is described in the section Energy Plus Weather for the place in which it is located the district.
FR2	The system has to be able to set up a url like it is described in the section <u>Eurostat</u> for the place in which it is located the district.
FR3	The system has to be able to retrieve data from the urls formed using the $FR1$ and $FR2.$
FR4	The system has to be able to format data gathered using the FR3 for its utilization by the scenarios evaluation in a RDF format.

4.2.2 Unstructured Data

Table 6: Functional requirements for unstructured data module

Requirement ID	DESCRIPTION
FR1	The unstructured data system has to be able to extract semantic metadata from inside web sites.
FR2	The unstructured data module has to be able to be connected to the rest of the OptEEmAL platform in order to offer its results.





FR3	The unstructured data module should store the result of the searches into its repository.	
FR4	The platform should show the users the results of the searches.	

4.3 Non-functional requirements definition

4.3.1 Structured Data

Table 7: Non-functional requirements for structured data

REQUIREMENT ID	DESCRIPTION
NFR1	The system has to have high band connectivity to <u>http://ec.europa.eu/eurostat</u> as well as to <u>https://www.energyplus.net</u> .
NFR2	The system has to be able to process data gathered in a reasonable amount of time (less than 1 min).

4.3.2 Unstructured Data

Table 8: Non-functional requirements for unstructured data

REQUIREMENT ID	DESCRIPTION
NFR1	The system must be connected to Internet.
NFR2	The system must show the results of a search in few seconds (less than 1 minute).





5 High level design of the geo-clustering connector

The geo-clustering connector which is going to be in charge of gathering the contextual data consists of two different components:

- **Geo Connector.** This component gathers data from well-structured data sources, described in section 2.2.
- Unstructured Data Connector. This component acts as an interface between the OptEEmAL graphical user interface and the unstructured data service (*QOTF service*) described in the section 3. Through this component the users of the OptEEmAL platform can visualize the graphical user interface of the QOTF service and use its search tools.

As explained in the section 2.1, only the information coming from the structured part is going to be used in the optimization and simulation process performed by the platform, Therefore, these two components will be integrated into the platform separately as it is explained in the following section.

5.1 Integration into the OptEEmAL platform

The structured data retrieved from the **Geo Connector** have to be inserted and stored into the platform to be used in the simulations. Therefore, the **Geo Connector** component will be integrated into the Data Insertion Module of the platform, described in D5.2 section 3.1.1, as it is shown in the following diagram which explains the Data insertion and diagnosis process described in D5.2 section 3.1.

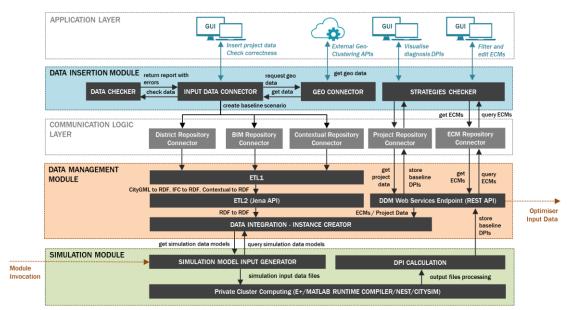


Figure 9: Geo Connector integration

The Unstructured data will not be used in the simulations, but it will be semantically analysed by a dedicated service to be shown to users of the platform in a useful way, as explained in section 3. This service will be developed as an external service (named QOTF service) and will provide users with a search engine through which they can search for useful information in the web. Therefore, the **Unstructured Data Connector** will be a component of the graphical user interface of OptEEmAL platform that will allow users to visualize the graphical user interface of the QOTF service, as described in D5.2 section 5.1.

The following diagram shows the integration of the Unstructured Data Connector into the graphical user interface of OptEEmAL platform and the process for gathering useful information from the web.





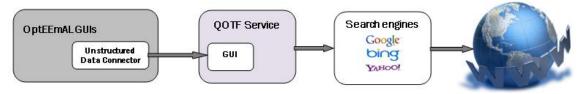


Figure 10: Unstructured Data Connector integration

The following sections explain the functional architecture designed for the Geo Connector component and the Unstructured data service (QOTF service), as the Unstructured Data connector will be a simple web page of the OptEEmAL GUIs (or even more simply a link/button in the project dashboard described in D5.2 section 5.3). Since the Unstructured data service will be an external service, it will be designed more deeply and developed in D1.4.

5.2 Functional architecture of the Geo Connector component

The Geo Connector component has the aim to gather the structured contextual data from external data sources, described in section 2.2. The data sets chosen as data sources allow requesting the available data through REST Services. Since the data retrieved from these services have to be stored in an RDF Store, that is the Context Repository described in D2.3 section 3.5.3, the Geo Connector has a specific sub-component (**Mapper**) that converts these data to RDF format, as it is illustrated in the following figure that shows the functional architecture of the Geo Connector.

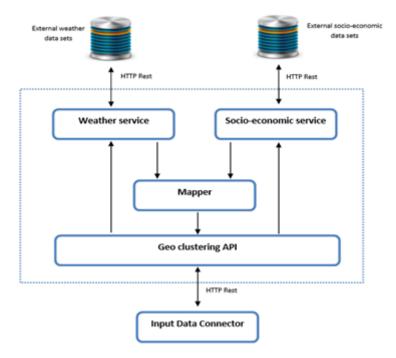


Figure 11: Functional architecture of the Geo Connector component

The Geo Connector component will technically be composed of the following sub-components:

- Weather service: this component allows requesting the weather data to the External weather data service (EnergyPlus) through its REST services. The REST calls to these services are explained in detail in section 2.2.1.
- Socio-economic service: this component allows requesting the socio economic data to the External socio-economic data service (Eurostat) through its REST services. The REST calls to these services are explained in detail in section 2.2.2.
- Mapper: this component parses the data retrieved from the external services in RDF format.





• **Geo clustering API:** this component provides the Application Programming Interface that will be used by the Input Data Connector component of the Data Insertion module, as it is depicted in the figure 9.

A class diagram for this component can be found in the D5.2 section 3.2.1.2, while its interaction with the others components of the Data insertion module of the OptEEmAL platform is described in the related sequence diagram in section 3.3.1 of the same document.

5.3 Functional architecture of the Unstructured data service

The following figure shows the functional architecture of the Unstructured data service.

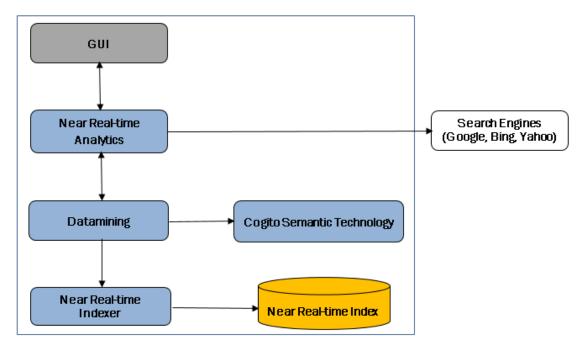


Figure 12: Functional architecture of the Unstructured data service

The component with grey background is the Graphical User Interface through which users can perform their searches and visualize results. The components with blue background are those that provide the functionalities of capture, analysis, manipulation, and indexation of unstructured information. In the meantime, the component with orange background stores some metadata of the retrieved information in order to support the near real time features.

The process is the following: The **Near Real-time Analytics** component is the entry point for user's requests; in a matter of seconds it acquires content, analyses it with **Cogito Semantic Technology** and joins together the many results (also from different search engines). To do this, it uses the **Datamining** component to elaborate the request of document analysis. The results are indexed and stored through the **Near Real-time Indexer** component.

The Unstructured data service is composed of the following components:

• **GUI**: the graphical user interface allows users to experience a new way of surfing the web. By applying semantic analysis on the fly to search engines results (Google, Bing and Yahoo), the user is able to filter and explore a vast quantity of results near real-time, finding the information faster and with increased accuracy. As a part of the faceted navigation capability, the user has the collected information grouped in different tabs (Topics, People, Domain Entities, etc.), allowing to have a different view of the content that can offer a useful and alternative visualization of the information.







OPTIMISED ENERGY EFFICIENT D

- Near Real-time Analytics: this component is designed to expose the Query-on-the-fly to users. In a matter of seconds, Query-on-the-fly acquires content, indexes it with Cogito Semantic Technology and joins together many results (also from different search engines).
- **Datamining**: this component elaborates the requests of document analysis and uses the Cogito semantic analysis features to discover useful knowledge in the search results.
- Cogito® Semantic Technology: this component represents the Expert System's software platform for document analysis. Cogito® via its ESSEX service exposes a set of simplified interfaces which allow for coordinated access to the base analysis capabilities, categorization and extraction. These capabilities are explained in detail in the first iteration of D1.3 section 3.2.
- Near Real-time Indexer: this component indexes the analysed results. It will be implemented with Apache Solr, that is an open source enterprise search platform, written in Java, from the Apache Lucene project. Its major features include full-text search, hit highlighting, faceted search, real-time indexing, dynamic clustering, database integration, NoSQL features and rich document (e.g., Word, PDF) handling. Providing distributed search and index replication, Solr is designed for scalability and Fault tolerance.
- Near Real-time index: This component represents a repository for storing the index generated by the indexer.



6 Quality integration test

Data quality is an essential parameter while using datasets from different origins. Indeed, an assessment of the quality of those external data is needed to ensure that the envisaged data does not reduce the quality of the analysis. This is well reflected in the ISO 14044 standard related to Life Cycle Assessment (LCA) which defines data quality as "characteristics of data that relate to their ability to satisfy stated requirements" [1]. In the frame of the OptEEmAL platform, as geo-clustering techniques will be used to complement the data model, it is essential to ensure that those data have the ability to satisfy stated requirements before their integration in the platform. In OptEEmAL, the requirement for the geo-clustered data is to make possible the calculation of energy (for weather data), economic (for energy prices) and social (for average incomes) DPIs with a sufficient level of quality.

Data quality is usually defined through a set of different indicators or criteria [2] [3] whatever the application domain is. For instance, in public health (where data quality is of crucial importance), a review of data quality assessment studies shows that up to 49 different criteria can be used to assess data quality [3]. In the frame of the OptEEmAL platform, it will be necessary to define the relevant criteria which will be used to assess data quality. Also, in order to make the data quality assessment procedure transparent and easier to apply in the platform, a quantitative assessment method for data quality is required.

As a consequence, the following overall methodology is proposed to ensure the data quality of geoclustering data:

- To define relevant criteria for the OptEEmAL platform and for each of this criterion, elaborate the relevant assessment matrix.
- To set a minimum requirement of data quality.
- To define a process to check the quality of the envisaged geo-clustering data sources (both for structured and unstructured geo-clustering data sources) and validate if they comply with the minimum requirement.

6.1 Criteria used to assess data quality

The number of criteria used to assess data quality is very different depending on the application sector as well as on the data origin. For instance, as indicated above, up to 49 criteria can be used in public health when assessing data quality. When investigating the specific case of Linked Open Data (LOD), it has been revealed that data quality can be represented with a maximum of 26 dimensions [4]. Finally, in LCA, where data quality related issues have been studied for several years and quantitative data assessment is commonly performed, it is widely accepted that input data quality can be evaluated using 5 different criteria [5]. It has to be mentioned that in the field of building energy simulations, data quality is a current area of concerns but is less advanced than in LCA for instance, without any data quality assessment procedure commonly agreed [6].

The analysis of the different criteria used in the different sectors reveals that some common criteria are used to define the quality of a given data or datasets. Those criteria are:

- Geographical Representativeness (GR)
- Time Representativeness (TiR)
- Technological Representativeness (TeR)
- Accuracy (also called "Precision/Uncertainty) (A)
- Completeness (C)
- Reliability/Credibility (R)





In the frame of the OptEEmAL project, two different and already defined geo-clustered data sources will be used: one for weather data and the other one for energy prices and income levels. Considering that these data sources are already known, it is proposed, for efficiency purpose, to perform an a priori manual assessment of these data sources in terms of data quality.

In case that the expected data sources are either not known or expected to change within the project, as it will be the case with unstructured data sources, it will be needed to implement an automatic data quality evaluation process. This is a more time consuming process and it implies the gathering of metadata associated with the data of interest. However, the methodology presented below will be exactly the same.

In addition, it is needed to establish a transparent data quality assessment methodology as it will make possible the comparison between different data sources and also the evaluation, in the future, of potential new data sources and the comparison of them with an already used one.

Considering the abovementioned requirements, the definition of a data quantitative assessment methodology in a transparent manner is needed. To do so, it is proposed to use as a basis the approach for semi-quantitative assessment of overall data quality proposed in the EC Product Environmental Footprint Guide [7]. This approach aims at defining for each criteria used in the data quality assessment process, a scoring and associated definition onto which each different data will be evaluated. Within the OptEEmAL project, it is necessary to adapt this approach for the selected criteria. This work is presented below for the different criteria to be used in OptEEmAL namely:

- Geographical Representativeness (Table 9)
- Time Representativeness (Table 10)
- Accuracy (also called "Precision/Uncertainty) (Table 11)
- Completeness (Table 12)
- Reliability/Credibility (Table 13)

Table 9: Definition of data quality assessment levels for Geographical Representativeness (GR).

Criteria	Definition	Score	Level	Associated values
Geographical representativeness Degree to which the dataset reflects the true population of interest regarding geography	1	Very good	Local (Municipal data)	
	Degree to which the	2	Good	Regional
		3	Fair	.ocal (Municipal data)
	· ·	4	Poor	Continental
		5 Very poor	Very poor	· ·

Table 10: Definition of the data quality assessment levels for Time Representativeness (TiR).

Criteria	Definition	Score	Level	Associated values
Time Degree to which the dataset reflects the specific conditions of the system being considered regarding the time/age of the data	1	Very good	2012 - 2015	
	dataset reflects the	2 Good 2010 - 20	2010 - 2012	
	•	3	Fair	2010 - 2012 2005 - 2010
		4	Poor	1995 - 2005
	5	Very poor	<1995; Unknown	





Criteria	Definition	Score	Level	Associated values
Accuracy Qualitative expert judgement	1	Very good	Very low uncertainty	
		2	2 Good Low uncerta	Low uncertainty
	•	3	Fair	Fair uncertainty
		4	Poor	High uncertainty
		5	Very poor	Very high uncertainty

Table 11: Definition of the data quality assessment levels for Accuracy (A)

Table 12: Definition of the data quality assessment levels for Completeness (C)

Criteria	Definition	Score	Level	Associated values
Completeness To be judged with respect to the coverage of each data source and in comparison, to a hypothetical ideal data quality	1	Very good	Very good completeness	
	to the coverage of each	2	Good	Good completeness
		3	Fair	Fair completeness
	51	4	Poor	Poor completeness
	quality	5	Very poor	Very poor or unknown completeness

Table 13: Definition of the data quality assessment for Reliability/Credibility (R)

Criteria	Definition	Score	Level	Associated values
Reliability/Credibil ity	Degree to which the	1	Very good	Widely acknowledged institution
		2	Good	Acknowledged institution
	data source is coming from a reliable/credible institution	3	Fair	Acknowledged institution Fairly acknowledged institution Poorly acknowledged
	Institution	4	Poor	Poorly acknowledged institution
		5	Very poor	Unknown institution





6.2 Minimum requirement for data quality

After the definition of such assessment tables for each criterion, it will be necessary to define the minimum requirement for data quality. This requirement will be the threshold that will be used to state whether a given data source can be used in the OptEEmAL platform or not.

To define this minimum requirement, it is proposed to set an overall minimum requirement. This overall minimum requirement is obtained through the aggregation of each criterion in a common data quality indicator. To do so, it is proposed to also use the methodology proposed in the EC Product Environmental Footprint Guide and presented in [8], which defines the following formula to calculate an aggregated data quality indicator called "Data Quality Rating (DQR)".

(1) $DQR = \frac{Criterion_1 + Criterion_2 + Criterion_n}{n}$ With n = number of criteria

The adaptation of this formula in OptEEmAL gives:

(2)
$$DQR = \frac{GR + TiR + A + C + R}{5}$$

Then, using a quality scale provided in [8], it is possible to assess the overall data quality level of a given dataset (Table 14).

Table 14: Overall data quality level according to the achieved aggregated data quality indicator

Data Quality Rating	Overall data quality level
=< 1.6	"Excellent quality"
1.6 to 2.0	"Very good quality"
2.0 to 3.0	"Good quality"
3.0 to 4.0	"Fair quality"
> 4.0	"Poor quality"

The final step consists in defining the minimum quality level that will be required for a given dataset to be included in the OptEEmAL platform. Considering the level of detail that the platform is expected to have and provide, we consider here the maximum acceptable value for the DQR to be 3.0. This value represents the limit between datasets of "good" quality and "fair" quality (according to the considered methodology).





6.3 Application to structured data sources

6.3.1 Energy Plus Weather

Data used considered in this section are available on the EnergyPlus Weather website mentioned before in this document (section 2.2.1). Their main characteristics with respect to data quality are indicated below (Table 15).

Data quality criteria	Value	Associated score (according to the OptEEmAL methodology)
G	Country specific	2
TiR	2015	3
А	Fair uncertainty	3
С	Very good completeness	1
R	Widely acknowledged institution	2

Table 15: Data quality assessment for EnergyPlus Weather data

Using equation (2) presented above, the aggregated data indicator for this example is:

(3)
$$DQR = \frac{2+3+3+1+2}{5} = 2.2$$

Consequently, it means that the considered dataset is of "good quality" and can be used in the calculations of the OptEEmAL platform.

6.3.2 Eurostat

Data considered in this section are available on the EUROSTAT portal as indicated before in this document.

6.3.2.1 Energy prices

Their main characteristics of energy prices data from EUROSTAT with respect to data quality are indicated below (Table 16).

Table 16: Data quality assessment process for energy prices data from Eurostat

Data quality criteria	Value	Associated score (according to the OptEEmAL methodology)
G	Country specific	3
TiR	2015	1
A	Fair uncertainty	3
С	Very good completeness	1
R	Widely acknowledged institution	1

Using equation (2) presented above, the aggregated data indicator for this example is:

(4)
$$DQR = \frac{3+1+3+1+1}{5} = 1.8$$





Consequently, it means that the considered dataset is of "very good quality" and can be used in the calculations of the OptEEmAL platform.

6.3.2.2 Income levels

Their main characteristics of income level data from EUROSTAT with respect to data quality are indicated below (Table 17).

Data quality criteria	Value	Associated score (according to the OptEEmAL methodology)
G	Regional	2
TiR	2014	1
А	Fair uncertainty	3
С	Very good completeness	1
R	Widely acknowledged institution	1

Using equation (2) presented above, the aggregated data indicator for this example is:

(5)
$$DQR = \frac{2+1+3+1+1}{5} = 1.6$$

Consequently, it means that the considered dataset is of "very good quality" and can be used in the calculations of the OptEEmAL platform.





7 Conclusions

The OptEEmAL platform needs a large amount of data as input to define the characteristics of the environment of the particular project that will work with it. This contextual data will be gathered using external services that will be accessed using two well differentiated sub-systems that have been denominated:

- Geo Connector.
- Unstructured data connector.

The external services to be used for gathering the structured contextual data have not been chosen among those described in the first iteration of the document. It was thought to use others that were chosen when, later on, it was done the research of the low level characteristics that data has to comply with, and we realized that the sources had to be different ones. This decision was motivated not only because of format issues but also because of aspects such the misunderstanding of the differences between weather and climate data (it is explained in section 2.1).

The aspect that goes beyond the actual state of the art is the use of data coming from external sources for simulation purposes. It is not easy (not to say impossible) to find references of a system that uses them as extensively as OptEEmAL will do. This fact makes of paramount importance to be strict when crossing the needs of the simulation software with the characteristics of the data that can be obtained from a particular data source. Despite the lack of specific requirements coming from other parts of the platform, it was only needed to know the software that will run the simulations at the end of a chain of several processes. Knowing what this software needs, in terms of data format and completeness, we can guarantee to fulfil the requirements imposed by it.

For the unstructured contextual data, it is explained how information from the external service will be extracted, performing queries on the fly regarding the district associated to the project. This explanation did not need to describe the internal aspects of its architecture since its prototype is already developed and just an adaptation will be needed. This adaptation will not have an important impact since data coming from this subsystem will not be stored in the database as data coming from the Geo Connector will be. It will be retrieved and shown to the user instantly on demand.

Once the description is done, the requirements specification follows. Two separated sets were stablished in which the integration with the rest of the platform is contemplated in a low enough level to develop the solution but not as low as to impose hard technical constraints that are not really needed in the development. The size and lack of complexity of the description of the software suggests not going too deep in describing the solution to allow freedom in its development, but it does not mean that the functionality that it has to comply with doesn't have to be described with enough level of detail as it was intended in the specific tables.

The last section, the one called "Quality Integration Test" performs a study of the data quality coming from the Geo Connector for ensuring that data gathered by the means described here will be useful for the platform purposes, mainly for the use of the information by the simulation engines. The score obtained was: "good quality" for climate data and "very good quality" for energy prices and income levels that are close to the top score.





OPTIMISED ENERGY EFFICIENT DES

8 References

- [01] ISO, "ISO 14044 Environmental Management Life Cycle Assessment Principles, frameworks and guidelines." 2006.
- [02] M. Bergdahl, M. Ehling, E. Elvers, E. Földesi, T. Körner, A. Kron, P. LohauB, K. Mag, V. Morais, A. Nimmergut, H. Viggo Saebo, U. Timm, and M. Joao Zilhao, "Handbook on Data Quality and Assessment Methods and Tools," 2007.
- [03] H. Chen, D. Hailey, N. Wang, and P. Yu, "A Review of Data Quality Assessment Methods for Public Health Information Systems," Int. J. Environ. Res. Public Heal., vol. 11, pp. 5170 – 5207, 2014.
- [04] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer, "Quality Assessment Methodologies for Linked Open Data: A Systematic Literature Review and Conceptual Framework," Semant. Web – Interoperability, Usability, Appl., vol. 1, p. 33, 2012.
- [05] B. Weidema, C. Bauer, R. Hischier, C. Mutel, T. Nemecek, J. Reinhard, C. Vadenbo, and G. Wernet, "Data quality guideline for the ecoinvent database version 3," *ecoinvent reports*, vol. 3, no. 1, 2013.
- [06] N. Jain, A. Ramallo Gonzalez, and S. Natarajan, "Understanding the Effect of Aggressive Energy Efficiency Regulation on an Unprepared Building Sector Using Uncertainty Analysis.," in Building Simulation International Conference, 2015.
- [07] S. Manfredi, K. Allacker, K. Chomkhamsri, N. Pelletier, and D. M. De Souza, "Product Environmental Footprint (PEF) Guide," 2012.
- [08] Official Journal of the European Commission, Commission Recommandation on the use of common methods to measure and communicate the life cycle environmental performance of products and organisations. Europe, 2013, pp. 1–210.
- [09] European Commission, "EUROSTAT," 2016. [Online]. Available: http://ec.europa.eu/eurostat/. [Accessed: 07-Jun-2016].
- [10] European Commission, "Electricity prices for domestic consumers bi annual data (from 2007 onwards)," 2016.

