Project ICT 287534
Start: 2011-09-01
Duration: 36 months
Co-funded by the European Commission within the 7th Framework Programme

SEMANCO Semantic Tools for Carbon Reduction in Urban Planning

SEMANCO

Deliverable 3.4
Ontology repository with migrated data

Revision: 5
Due date: 2013-02-29 (m18)
Submission date: 2013-04-30
Lead contractor: FUNITEC

Dissemination level		
PU	Public	x
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Deliverable Administration & Summary					
No & name	**D3.4 Ontology repository with migrated data**				
Status	Final	**Due**	m18	**Date**	2013-04-30
Author(s)	Álvaro Sicilia (FUNITEC)				
Editor (s)					
DoW	Deliverable 3.4 is based on the work done in Task 3.4 - *Ontology Repository and Data migration to OWL format*. In Task 3.4, An Ontology Repository will be implemented to manage the data sources offering an Ontology Access Layer which will abstract and encapsulate access to the ontologies stored in the repository and in the Linked Data cloud. The Ontology Repository will use web standards in communication protocols to expose previously modelled data stored. Translating the contents of databases to the standard energy model developed in T.3.2 and T3.3. Setting up an ontology repository to store the migrated data. Creation of automatic procedures to perform the migration, for each data source.				
Comments					

Document history

V	Date	Author	Description
1	2013-02-01	Álvaro Sicilia (FUNITEC)	First version of the document
2	2013-04-25	Álvaro Sicilia (FUNITEC)	Second version of the document
3	2013-04-26	Leandro Madrazo (FUNITEC)	Review by project coordinator. This version submitted to internal reviewers: Thomas Rønn (RAMBOLL) and Gonzalo Gamboa (CIMNE)
4	2013-04-29	Álvaro Sicilia (FUNITEC)	Inclusion of changes proposed by reviewers
5	2013-04-29	Michael Crilly (UoT)	Final proof-reading

Table of Contents

EXECUTIVE SUMMARY

Delivery 3.4 *Ontology Repository and Data migration to OWL format*, developed within Work Package 3 *Energy Data Modelling*, summarizes the work done and the results achieved in Task 3.4 *Ontology repository with migrated data*, whose goal is the implementation of the ontology repository to store the different data sources needed by the analysis tools of the SEMANCO platform.

Task 3.4 contributes to the project with the ontology repository which contains the data sources provided by the three case studies. The repository contains two kinds of data: building typologies based on building use or year of construction, and urban data including contextual information at city or neighbourhood scale. A semantic integration process –devised in the Task 4.1 *Environments for collaborative ontology mapping* has been carried out to transform the relational data sources into Resource Description Framework (RDF) according to the semantic energy model developed in the project. This process has applied the ontology mapping tools created in the project as well as the data sources mapping tables collected by the partners.

The report is structured in the following sections:

1. **Introduction**: Purpose of the deliverable, contributions of partners, and relationship between the work done in this task with other work packages.

2. **Data source repository**: Description of the architecture of the repository which includes three databases to store the data sources of each case study. For each database there has been a D2R server installed to translate its contents into RDF. Furthermore, a light-weight web interface has been provided to data owners and energy experts to visualise and modify the contents of the database. A list of the most relevant Uniform Resource Locators (URLs) of the repository has been provided.

   A description of the data sources included in the ontology repository has been included in the Section 2.2. The most important tables and columns have been detailed including diagrams of the databases structure.

3. **Semantic integration process**: Description of the semantic integration process carried out in this task. A brief explanation of how the ontology mappings tools have been applied in the process. The automatic generation of mappings has been illustrated including some examples, as well as, the customisation of mappings description has been provided.

4. **Outputs**: This section includes some examples of how to retrieve data from the repository using SPARQL Protocol and RDF Query Language (SPARQL) queries. This is useful to demonstrate that the data provided by partners can be queried by the tools which are being developed in Work Package 5. It also includes some statistics of the contents of the repository.

5. **Conclusions**: Contributions of the ontology repository to the project development, and the impact on other Work Packages, tasks, and demonstration. The repository is a key component of the demonstration since it contains the data sources needed to carry out the energy analysis performance with the tools of the SEMANCO platform.

6. **Appendices**: A list of tables which summarises the mappings carried out in the semantic integration process to transform the relational data into RDF.

1 INTRODUCTION

1.1 Purpose and target group

The purpose of this deliverable is to report on the work undertaken in Task 3.4 *Ontology Repository and Data migration to OWL format*.

An ontology repository has been created to store the data needed to carry out the implementation cycles of the three case studies. The repository consists of three databases with relational data provided by the three cases studies which are needed to carry out analysis by the analysis tools of the SEMANCO platform. Between the repository and the analysis tools, the Semantic Energy Information Framework (SEIF) acts as ‘glue’. It is composed of a semantic energy model, implemented as a global ontology, and a federation engine which queries the SPARQL endpoints of the ontology repository using the global ontology.

The data stored in the ontology repository can be queried by the analysis tools by means of the SEIF, thus the data owners and energy experts can visualise and modify the data through the web interface. The repository contains basically two kinds of data:

- Typologies based on building use or year of construction. The typologies include data like building properties (e.g. U values of Roofs, walls, and windows), energy demands (e.g. Cooling, heating, hot water, and electricity).
- Urban data of the areas including census (e.g. nationality and gender), cadastre (e.g. year of construction, building ground floor surface, and building use among others), Lower Layer Super Output Area (LSOA) data regarding fuel poverty and index of multiple deprivation.

The main goals of Task 3.4 are to set up the ontology repository and to carry out the semantic integration process incorporating data sources into the SEIF by means of the mappings tools developed in WP4.

The main target group of the work carried out in the Task 3.4 is the Work Package 5 developers since they are implementing the analysis tools which require the data stored in the ontology repository. Moreover, the target group of the Deliverable 3.4 is the ontology engineers and database experts since the document describe how to expose data stored in databases following Linked Data principles.

1.2 Contribution of partners

The work done in Task 3.4 has been led by FUNITEC in collaboration with HAS in the repository definition. The data has been provided by NEA, RAMBOLL, CIMNE and FORUM.

The mappings between the databases and the semantic energy model have been carried out by FUNITEC based on the data source mapping tables provided by NEA, RAMBOLL and CIMNE.

The report has been written by the FUNITEC and proofread by University of Teesside.

1.3 Relations to other activities in the project

The semantic integration process carried out in this task has been designed as part of Task 4.1. *Environments for collaborative ontology mapping* and relies on three prerequisites:

a) The semantic energy model developed in the Task 4.2 *Semantic Energy Model* which is the global ontology where the data sources will be mapped to.

b) The data sources mapping tables which describe how the data sources are aligned to the Standard Tables. Further information can be found in the Deliverable 3.2 *Guidelines for Structuring Energy Data*.

c) The ontology mapping tools developed to in the Task 4.1 which facilitates the codification of the mappings –written in D2RQ language– between the data sources and the ontology.

Figure 1 shows the relationship between the semantic integration process and the other lines of work of the project. The use case methodology described in the Deliverable 1.8 *Project Methodology* provides the data source mapping tables. Thus, the ontology design methodology described in the Deliverable 4.2 *Semantic Energy Model* provides the Semantic Energy Model which is needed to generate the mappings.

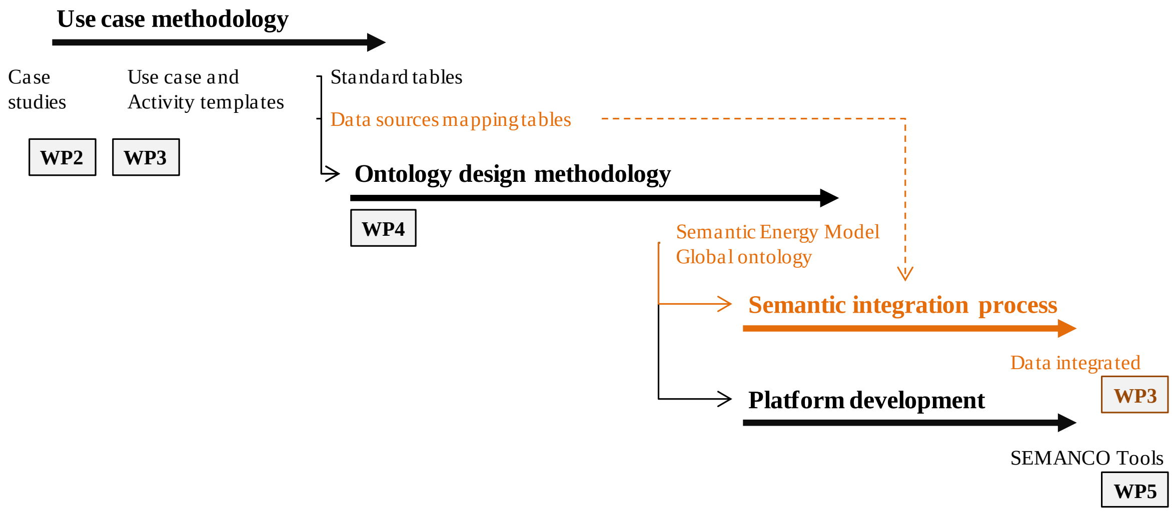


Figure 1. Relations of the semantic integration process with other lines of work

2 DATA SOURCES REPOSITORY

2.1 Repository architecture

The ontology repository is composed of three relational databases used to store the data sources of the three case studies (Figure 2). The D2R server (Bizier & Cyganiak, 2006) is a key component of the repository architecture since it translates the SPARQL queries formulated by the SEIF to SQL language needed by the relational databases and returns the relational data in RDF format. The translation between SPARQL-SQL and the transformation of the relational data into RDF format is based on the mappings coded in D2RQ language (Bizier & Cyganiak, 2007). This way, the repository is scalable to include future areas by adding new databases and a D2R server making the connection with the SEIF.

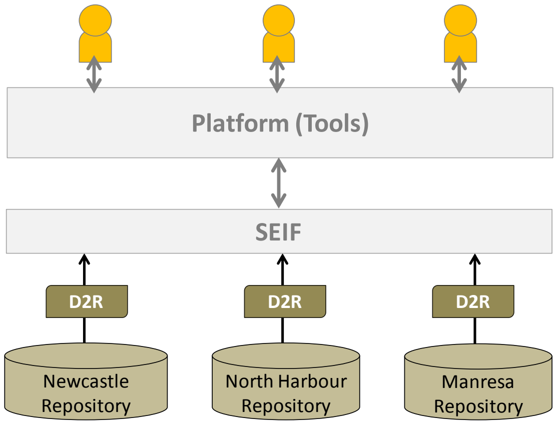


Figure 2. Relations between Data sources Repositories and SEMANCO's Platform and SEIF

The users –data owners, technicians, energy experts, decision makers and citizens– can access the data through the SEMANCO platform since its analysis tools query the data through the SEIF. It is a mediated access to the data because the tools retrieve, analyse, and present the data in the most appropriate format for the users' requirements. In addition, the data owners and energy experts can visualise and edit the raw data they have provided, through a light-weight web interface which has been developed in this task (Figure 3). Due to intellectual property rights of some of the data, the users have to login before they visualise or modify the data. For testing purposes, the North Harbour database can be accessed using the account login *Ramboll* and password *guest*.

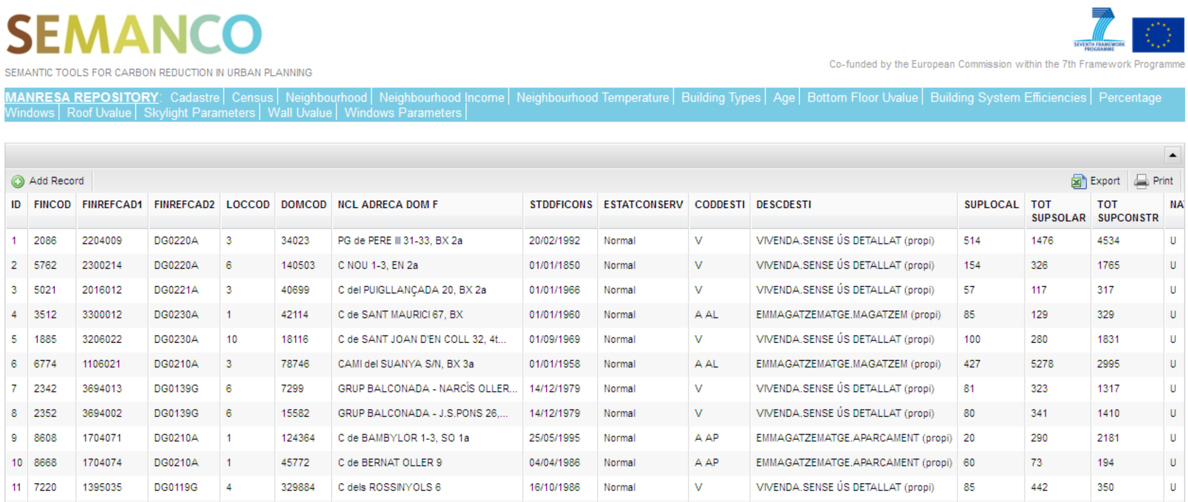


Figure 3. Web interface to modify and to visualize the raw data provided by the case studies

Table 1 contains the most relevant web addresses (Uniform Resource Locator, URL) to interact with the repository. For each case study is provided three web addresses: a SPARQL endpoint used by the SEIF to query the data, a SPARQL explorer which provides a graphical interface to create SPARQL queries and explore the data mainly used by the tools developers since they create the queries, and the user web interface address which let data owners and energy experts to modify and visualize their data.

Table 1 Repositories most relevant URLs

Area	Service	Web address (URL)
Newcastle	SPARQL Endpoint	http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/sparql
	SPARQL explorer	http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/snorql/
	User web interface	http://arcdev.housing.salle.url.edu/semanco/repository/newcastle
North Harbour	SPARQL Endpoint	http://arcdev.housing.salle.url.edu/semanco/repository/northharbourd2r/sparql
	SPARQL explorer	http://arcdev.housing.salle.url.edu/semanco/repository/northharbourd2r/snorql/
	User web interface	http://arcdev.housing.salle.url.edu/semanco/repository/northharbour
Manresa	SPARQL Endpoint	http://arcdev.housing.salle.url.edu/semanco/repository/manresad2r/sparql
	SPARQL explorer	http://arcdev.housing.salle.url.edu/semanco/repository/manresad2r/snorql/
	User web interface	http://arcdev.housing.salle.url.edu/semanco/repository/manresa

2.2 Data sources

The data stored in the ontology repository comes from different sources and domains. The data can be classified into two categories: a) data regarding building typologies (a simplified model of buildings) which provides statistical data about their physical properties and energy performance characteristics; and b) urban data which provides contextual information of the areas at city or neighbourhood scale.

2.2.1 Newcastle data sources

The Newcastle upon Tyne case study has provided two data sources which contain urban data, specifically tenure data and some LLSOA indicators both created and maintained by the national UK government. The building typology data is not present in this case study.

The tenure source is composed of a single table with the name of the street, the house number, the postal code and type of ownership as columns. Together the street name, house number and postal code are used to identify the buildings. In this way the analysis tools can correlate the Geographic Information System (GIS) data they have with this data to know the ownership of the buildings selected by the user in the 3DMaps environments.

The LLSOA indicators are relevant for Newcastle upon Tyne case study since are used to geographically target the most deprived areas where physical intervention is most beneficial. The LLSOA indicators are stored in a single table which have the following columns:

- LLSOA_CODE: The unique reference code of the LLSOA area.
- IMD_SCORE: Index of multiple deprivation which is calculated by combining seven domain scores such as Income, Employment, Health and Disability, Education, Skills and Training, Barriers to Housing and Services, Crime, and Living Environment.
- INCOME_SCORE: Income domain score of the indices of deprivation.
- Households: Total number of households within the LLSOA

- Fuel-Poor_Households: Number of households which are living in fuel poverty conditions within the LLSOA.

The structure of the Newcastle database is illustrated in the Figure 4.

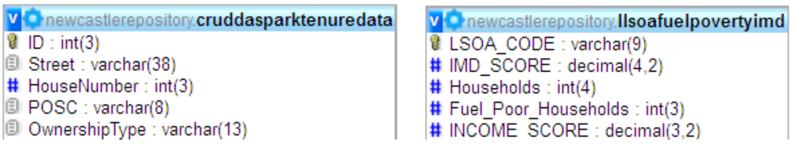


Figure 4. Newcastle database structure

2.2.2 North Harbour data sources

The data sources included in the North Harbour database are related to building typologies based on building use and year of construction. The typologies include energy demand details regarding space heating, cooling, hot water, and electricity demands. There are more sources which will be integrated in following versions, such as the energy supply technology sources and building geometry extracted from the 3DMaps environment.

The energy demand source contains data regarding the building typologies being used to plan the new area in the North Harbour. Since to data it remains undeveloped, the typologies are based on projections of construction into the future (e.g. 2015, 2020, 2030, and 2050). This data source has been provided by Ramboll as the project partner. The structure of the energy demand source comprises three interrelated tables (Figure 5). The age table contains the available age bands which are 2010-2015, 2016-2020, 2021-2030, and 2031-2050 that relate the potential phasing of future development. The building use table contains the Danish building uses which have been translated into English to match the building uses included in the Standard Tables. Finally, the energy demand table contains the building typologies which are based on specific age classes and building uses. Thus, this table has the following columns:

- HeatDemand: Energy demand for heating.
- HotWaterDemand: Energy demand for hot water.
- ElectrictyDemand: Energy demand of electricity.
- CoolingDemand: Energy demand for cooling.

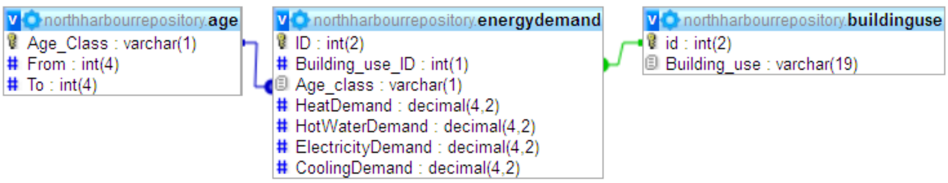


Figure 5. North Harbour database structure

2.2.3 Manresa data sources

The Manresa case study has provided several data sources for both categories: urban data and building typologies. These data sources are used as input for the URSOS tool –interfaced by the SEMANCO platform– to carry out the energy performance analyses.

There are three data sources which provide information at urban scale: the census, the cadastre, and the neighbourhood source. The census is important to extract the occupation of a building, the cadastre is relevant in obtaining building / property information such as the year of construction, the use, and the neighbourhood is a valuable source of socio economic parameters such as average income. Both census and cadastre sources are composed of single table, and the neighbourhood source is composed of three tables (Figure 6), their main columns are:

- Census:
    - DOMCOD: Code of the address where the household lives including postal code and house number.
    - ADRDESC: Name of the address where the household lives.
    - SEXE: Gender of the household.
    - TITULACIO: Education level of the household.
    - DESCPAISNAIX: Birth country of the household.
- Cadastre:
    - DOMCOD: Code of the address where the household lives including postal code and house number.
    - FINREFCAD1/FINREFCAD2: Cadastre reference.
    - STDFICONS: Year of construction of the building.
    - CODDESTI: Use of the building.
    - SUPLOCAL: Area of the building ground floor.
- Neighbourhoods:
    - Name: The name of the neighbourhood.
    - Income_percapita: The income per capita for each neighbourhood.
    - CS_Temperature_Heating_Mode: It is the internal average temperature as fixed by the control system in normal heating mode.

Regarding building typologies, the Manresa database encloses 12 tables where the main one is the *buildingtypes* which includes the building typologies based on year of construction and the relationship to the other tables. The data contained in these tables includes the U-values for ground floor, walls, roofs, windows, and skylights. Thus, the percentage of window areas and efficiency of the Heating, Ventilation, and Air Conditioning (HVAC) systems is incorporated. This data source contains typological values which are needed to carry out energy performance analyses and are specific to Manresa area, however, by defining new typologies the analyses can be carried out in other areas than Manresa.

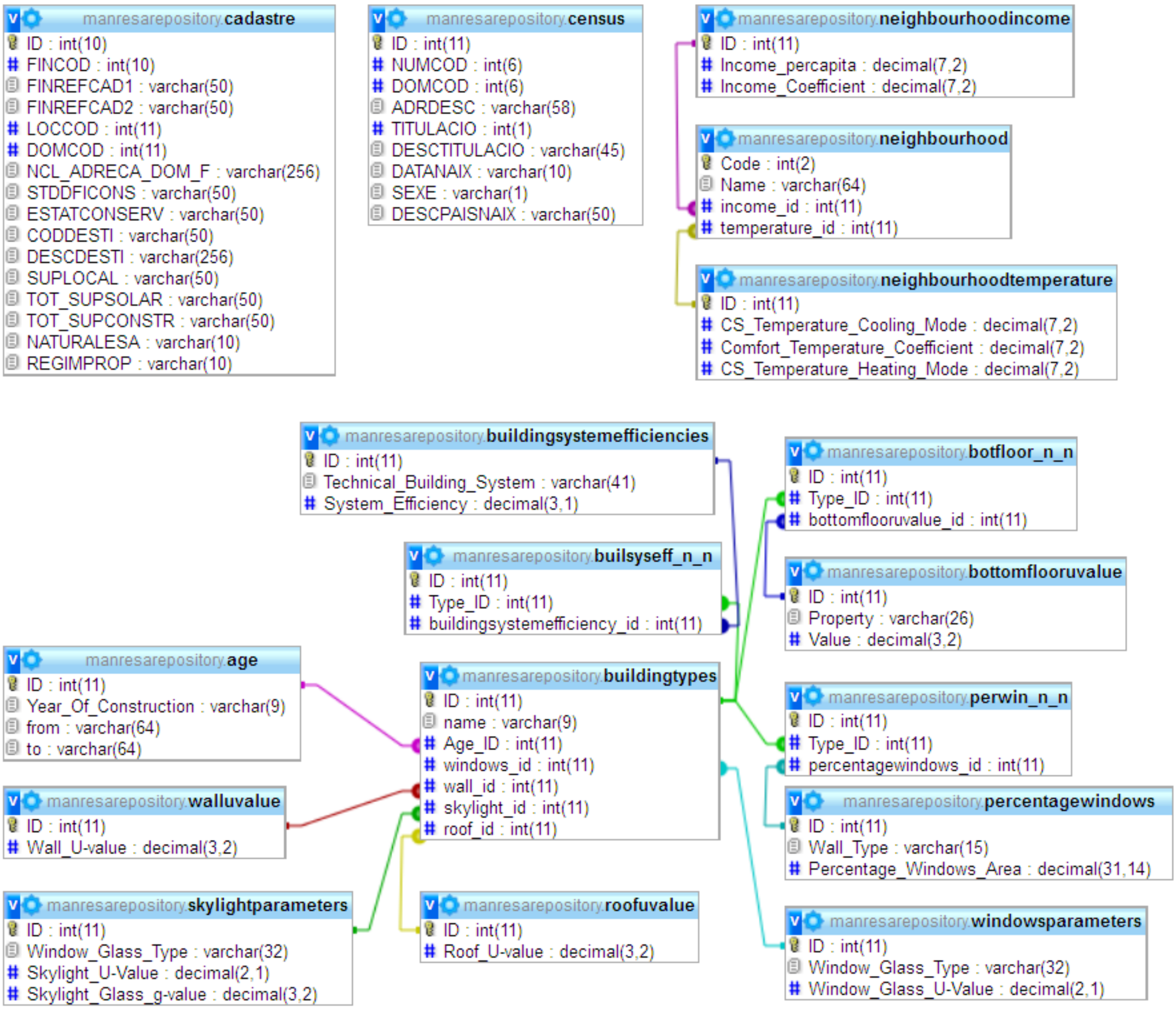


Figure 6. Manresa database structure

3 SEMANTIC INTEGRATION PROCESS

The work carried out in this task relies on the semantic integration process devised in Task 4.1 and the mapping tools also developed in Task 4.1. Furthermore, the semantic energy model created in the Task 4.2 is used. As stated in the Deliverable 4.1, the semantic integration process consists of three steps.

1. Creation of local ontologies for each data source
2. Integration of local ontologies within the energy model (global ontology)
3. Data source transformation to RDF

The semantic integration process is supported by the mappings tools developed in Task 4.1. As can be seen in Figure 7, the first step of the process –the creation of local ontologies– is fully automated by the OWL & mapping extractor tool. The second step, where the local ontology is integrated within the energy model (global ontology) is carried out in the ontology mapping collaborative web environment which receives the OWL ontology and the D2RQ mappings created by the extractor tool. These mappings have been established following the data sources mapping tables filled by the data providers of each case study. Finally, the third step of the integration process –data source transformation to Resource Description Framework (RDF) format– is fully automated by the D2R server which takes the D2RQ mappings generated by the ontology mapping collaborative web environment.

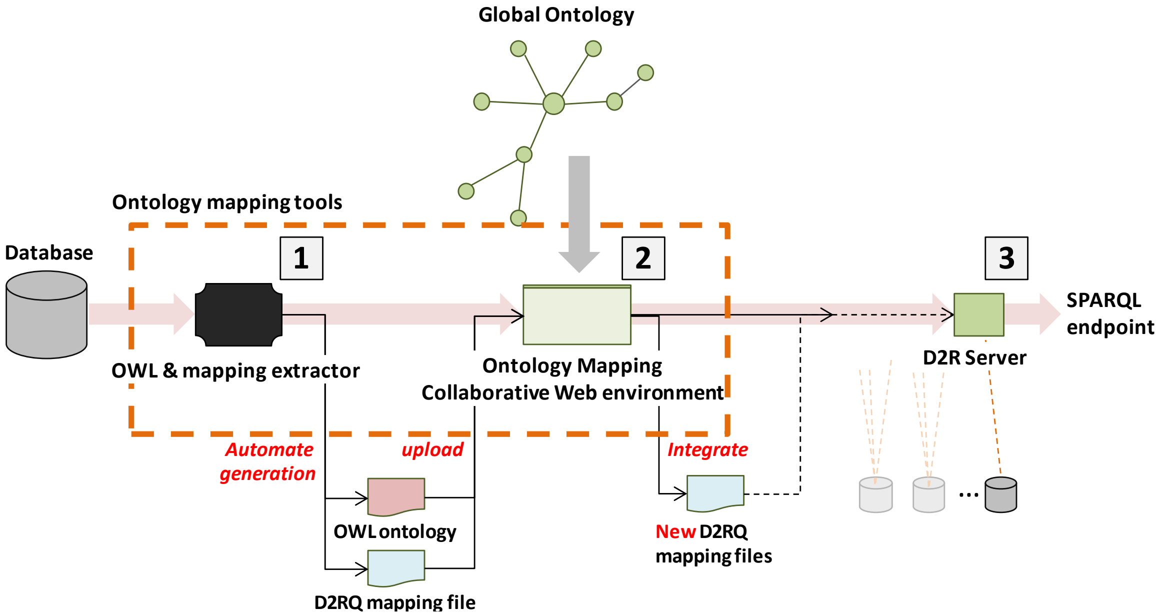


Figure 7.Ontology mapping tools. Inputs and outputs

For example, in the Manresa database the OWL and mapping extractor tool has been called with the following commands:

```
~/ java -jar owln3extractor.jar -generateconfig -db jdbc:mysql://localhost/manresarepository -driver
com.mysql.jdbc.Driver -u root -p "" -n manresarepository -o manresarepository

~/ java -jar owln3extractor.jar -extract -i manresarepository_config.n3
```

Following the same example, the Manresa local ontology and D2RQ mapping files have been uploaded to the web environment where the mappings have been established (Figure 8). In this environment, each concept of the local ontology –which is a column in the tables of the data sources– is mapped to a concept of the global ontology.

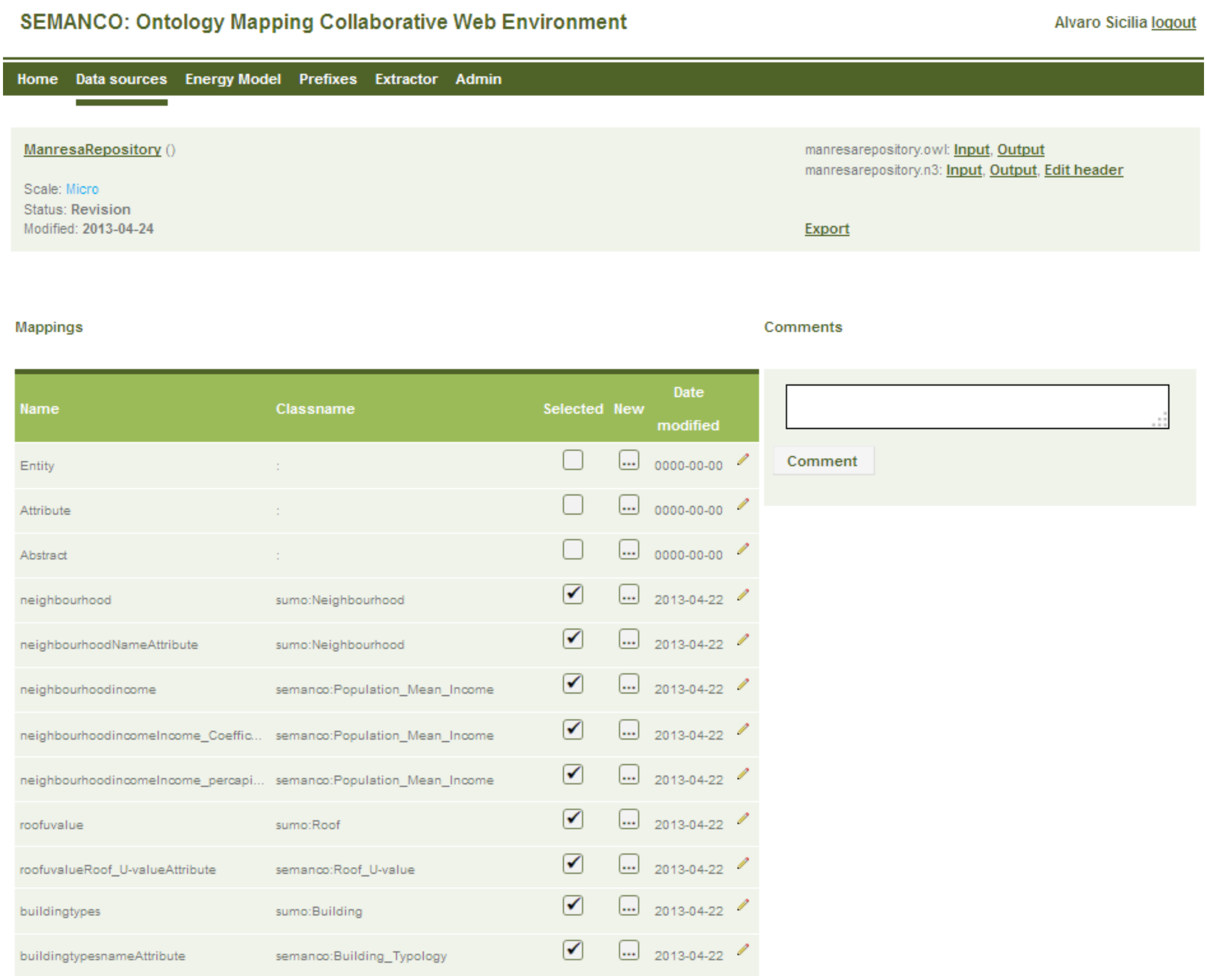


Figure 8.Ontology mapping environment with the mappings created for the Manresa database

The ontology mapping collaborative web environment generates the mappings between the local and global ontology in D2RQ language which is needed by the D2R server. Typically 90% of these mappings have been automatically generated by the environment. However the remaining 10% of the mappings have been coded manually because the environment could not cover their specialties and specifics. The following sections describe how the automatic mappings are generated and how the manual mappings have been coded. A complete list of mappings carried out in the semantic integration process can be found in the appendices.

3.1 Automated mappings

The main goal of the automatic generation of mappings is to ensure that data transformed into RDF follows the structure of the semantic energy model (global ontology). Most of the mappings have been automatically generated following the rules described in Deliverable 4.1. Given a table, the primary key has been used as the main concept, and then the columns of the table are mapped to other concepts. For example, the *ID* column of the *CruddasParkTenureData* table is mapped to *sumo:Building* class. Then, the column *Street* of the *CruddasParkTenureData* table is mapped to *semanco:Address*. In the ontology mapping collaborative web environment the URI patterns can be customised by selecting the most appropriate column of the table to define the URI. The example stated above generates the following mappings:

```
#Mapping for table cruddasparktenuredata
map:map_8065 a d2rq:ClassMap;
        d2rq:dataStorage map:database;
        d2rq:uriPattern "building/@@cruddasparktenuredata.ID|urlify@@";
        d2rq:class sumo:Building.

#Mapping for the object property connection sumo:Building    and semanco:Address
map:joining_8065_7_middle_11 a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:map_8065;
        d2rq:property semanco:hasAddress;
        d2rq:uriPattern "address/@@cruddasparktenuredata.ID|urlify@@".

#Mapping for the column Street of the table cruddasparktenuredata
map:map_8068 a d2rq:ClassMap;
        d2rq:dataStorage map:database;
        d2rq:uriPattern "address/@@cruddasparktenuredata.ID|urlify@@";
        d2rq:class semanco:Address.

#Mapping for the data property related to the column Street
map:datatype1676 a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:map_8068;
        d2rq:property semanco:addressValue;
        d2rq:column "cruddasparktenuredata.Street";
        d2rq:datatype xsd:string.
```

The ontology mapping collaborative web environment has automatically generated 90% of the mappings required to transform the data into RDF. The automatic generation process follows deterministic rules which have been designed to ensure that a database can be transformed following an ontology structure. Section 4 contains some query examples based on the semantic energy model (global ontology) to retrieve data from the repository demonstrating the mappings validity.

3.2 Custom mappings

There are ontology structures that cannot be covered by the automated generation of mappings. This is the case of the data sources where its columns are mapped to concepts which depend on the data value itself. For example, in the Census data source when the *SEXE* column takes the value 'H' then it should be mapped to *semanco:One_Person_Household_Male*, and when the value is 'D' then the mapped concept should be *semanco:One_Person_Household_Female*. In this case, in the ontology mapping collaborative web environment, the column is mapped to a more generic concept such as *semanco:Household_Type* which subsumes both concepts. Finally, the mapping is edited by adding a condition which takes into account the value of the column.

The mappings automatically generated for this example is the following:

```
# ClassMap for http://www.semanco-project.eu/2012/5/SEMANCO.owl#Household_Type
map:map_8041 a d2rq:ClassMap;
        d2rq:dataStorage map:database;
        d2rq:uriPattern "household_type/@@census.SEXE|urlify@@";
        d2rq:class semanco:Household_Type.
```

The customised mappings are as follows with changes highlighted bold:

```
# ClassMap for http://www.semanco-project.eu/2012/5/SEMANCO.owl#Household_Type
map:map_8041 a d2rq:ClassMap;
        d2rq:dataStorage map:database;
        d2rq:uriPattern "household_type/One_Person_Household_Male";
        d2rq:condition "census.SEXE = 'H'";
        d2rq:class semanco:One_Person_Household_Male.

# ClassMap for http://www.semanco-project.eu/2012/5/SEMANCO.owl#Household_Type
map:map_8041_b a d2rq:ClassMap;
        d2rq:dataStorage map:database;
        d2rq:uriPattern "household_type/One_Person_Household_Female";
        d2rq:condition "census.SEXE = 'D'";
        d2rq:class semanco:One_Person_Household_Female.
```

The mappings have been manually customised for the *SEXE* and *Titulacio* columns of the *Census* table, and *Ownership type* column of the *CruddasParkTenureData* table.

4 Outputs

4.1 SPARQL query examples

A number of examples are shown to demonstrate how the data can be queried by the tools. The examples are built on the SNORQL interface provided by the D2R server.

4.1.1 Searching for building typologies

Figure 9 contains a SPARQL query which retrieves the heat demand and the year of construction for a proposed Computer Centre (use typology) to be which use is constructed after 2018.

```
SELECT DISTINCT ?from ?heatdemand
WHERE {
?b semanco:hasBuilding_Use <http://arcdev.housing.salle.url.edu/semanco/repository
/northharbourd2r/resource/building_use/Computer%5Fcentre>.
?b semanco:hasSpace [semanco:hasHeat_Demand [semanco:heat_DemandValue ?heatdemand]].
?b semanco:hasAge_Class [semanco:hasFrom_Year [semanco:fromYearValue  ?from]].
        FILTER (?from > 2018).
}
```

SPARQL results:

from	heatdemand
2021	5.9
2031	5.9

Figure 9.SPARQL query to retrieve building typologies

4.1.2 Searching for deprived LLSOA

This example retrieves the LLSOA areas which have a relatively poor index of multiple deprivation score (IMD) and where their income score is low. Figure 10 contains a SPARQL query which retrieves the LLOSA areas (mapped to *semanco:Land* concept) which have the IMD data property below 0.05 value, and the Income data property below 3.

```
SELECT DISTINCT ?l ?IMD ?Income
WHERE {

?l a semanco:Land.
?l semanco:hasPopulation ?p.
?p semanco:hasPopulation_IMD_Score [semanco:population_IMD_ScoreValue ?IMD].
?p semanco:hasPopulation_Income_Score [semanco:population_Income_ScoreValue ?Income].
FILTER (?IMD < 0.05 && ?Income < 3).
}
```

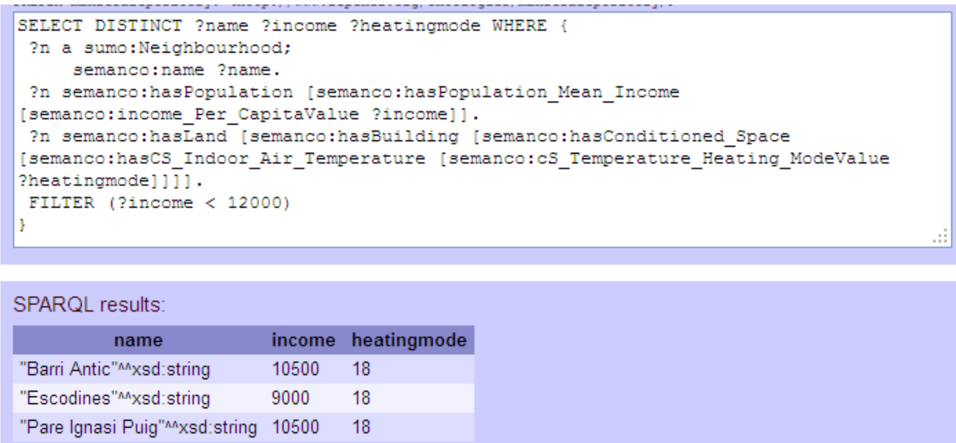
SPARQL results:

l	IMD	Income
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008309>	0.02	2.32
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008310>	0.02	2.32
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008366>	0.02	2.32
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008371>	0.02	2.32
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008418>	0.02	2.32
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008309>	0.02	2.86
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008310>	0.02	2.86
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008366>	0.02	2.86
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008371>	0.02	2.86
<http://arcdev.housing.salle.url.edu/semanco/repository/newcastled2r/resource/land/E01008418>	0.02	2.86

Figure 10.SPARQL query to retrieve LLSOA areas which are deprived

4.1.3 Searching for neighbourhood's heating mode and income per capita

This example searches for neighbourhoods –mapped to *sumo:Neighbourhood*– which have an annual income per capita below 10000 Euros. The data retrieved includes the neighbourhood name, the income (*semanco:income_Per_CapitaValue*) and the average temperature heating mode (*semanco:cS_Temperature_Heating_ModeValue*).



```
SELECT DISTINCT ?name ?income ?heatingmode WHERE {
 ?n a sumo:Neighbourhood;
     semanco:name ?name.
 ?n semanco:hasPopulation [semanco:hasPopulation_Mean_Income
[semanco:income_Per_CapitaValue ?income]].
 ?n semanco:hasLand [semanco:hasBuilding [semanco:hasConditioned_Space
[semanco:hasCS_Indoor_Air_Temperature [semanco:cS_Temperature_Heating_ModeValue
?heatingmode]]]].
 FILTER (?income < 12000)
}
```

SPARQL results:

name	income	heatingmode
"Barri Antic"^^xsd:string	10500	18
"Escodines"^^xsd:string	9000	18
"Pare Ignasi Puig"^^xsd:string	10500	18

Figure 11.SPARQL query to retrieve neighbourhoods with an income per capital below than 10000 Euros

4.2 Statistics

Table 2 provides a summary of the main statistics of the ontology repository. The 'number of mappings' figure includes the mappings to define the instances, object properties to connect instances between them, data properties to implement the literal values, and annotations properties to attach textual information to the instances. The 'number of distinct classes' figure considers the amount of concepts of the global ontology which are present in the repository. The 'number of triples' figure shows the amount of RDF triples (subject – predicate – object) stored in the repository.

Table 2. Contents of the repository

	Newcastle	North Harbour	Manresa
Number of mappings	40	41	260
Number of distinct classes	12	11	51
Number of triples	13005	564	3795940

The contents of the ontology repository reflect the heterogeneity of the three cases of study. While the Newcastle upon Tyne and North Harbour cases use a similar amount of diverse concepts, the Manresa case contains 4 times more concepts to model its data. This is a result of the Manresa case providing both building typology and urban scale data. As a result, the Manresa database is much bigger than the other two case studies since it includes census and cadastre data for the entire city.

5 Conclusions

5.1 Contribution to overall picture

The ontology repository stores the data sources needed by the analysis tools of the SEMANCO platform. The semantic integration process carried out in this task is an important step of the project development since put together different lines of work such as the use case methodology, ontology development process, and data modelling. By completing this work, it demonstrates that the data sources provided by the partners can be successfully integrated by means of the global ontology developed in the project and be accessed by the tools using SPARQL queries.

The processes described in this report can be applied whenever new case studies and projects wish to use the SEMANCO platform. In the first instance, the data sources mapping tables should be completed to map the new data sources structure to the semantic energy model. Then, if there are concepts missing in the semantic energy model this should be updated. Finally, a database is to be created for the new case study or project applying the semantic integration process described in D4.1.

5.2 Impact on other WPs and Tasks

The work undertaken in this task is related to previous Tasks 4.1 and 4.2 since the semantic energy model and the mapping tools have been applied to carry out the semantic integration process to transform the relational data into RDF.

The implementation of the repository is important to the platform development which is being carried out in WP 5. Furthermore, the data stored in the repository will be used in the second and third implementation iterations.

5.3 Contribution to demonstration

The repository is one of the components needed to implement the demonstration since it contains the data sources to carry out the analysis. For example, the North Harbour case study analyses the urban plans –by means of the SEMANCO platform– based on a list of building typologies which are stored in the ontology repository. Also, the LLSOA and tenure data of the Newcastle upon Tyne case study is required to enable stakeholders to graphically identify –through the 3DMaps integrated in the platform– the most deprived areas with a high social impact most suitable for intervention. Furthermore, the census, the cadastre, and the building typologies stored in the Manresa database are essential for carrying out the energy performance analysis in the URSOS tool interfaced by the platform.

5.4 Other conclusions and lessons learned

Some of the data sources provided by the partners had to be restructured to take advantage of the relational database structures such as n-to-n relationships. In this way, rewriting of queries was minimized.

One of the mains difficulties faced in the semantic integration process was dealing with diverse data sources related to the same area of interest. Often those sources use different names and properties to conceptualise the same concepts. A mediation action was required between the different partners involved, which relied on the sharing the data sources mapping tables and the Standard Tables. From a technical point of view, the automatic generation of mappings was very helpful since was coded more than 400 mappings. In the appendices can be found about 60 mappings established in the ontology mapping collaborative web environment, the other mappings were carried out automatically by the environment.

6 BIBLIOGRAPHY

Bizer, C. & Cyganiak, R. (2007) D2RQ – Lessons learned. *Position paper at the W3C Workshop on RDF Access to Relational Databases*, Cambridge, October 25-26, 2007.

Bizer C. & Cyganiak R. (2006) D2R Server – Publishing Relational Databases on the Semantic Web. *Poster at the 5th International Semantic Web Conference (ISWC)*, Springer, Nov. 2006.

7 APPENDICES

APPENDIX A. Newcastle mappings

Table A1 contains the mappings carried out for the Newcastle upon Tyne database.

Table A1. Mappings carried out for Newcastle data migration

Data source	**Data name (in the data source, usually a column)**	**Standard Tables name, in parentheses the related sheet.**	**Class of the SEMANCO Ontology**
CruddasParkTenureData	ID	Building (Building)	sumo:Building
CruddasParkTenureData	Street	Address (Building)	semanco:Address, addressValue data property
CruddasParkTenureData	House Number	Address (Building)	semanco:Address, houseNumber data property
CruddasParkTenureData	Posc	First_Part_Of_Postcode (Building)	Semanco:First_Part_Of_Postcode
CruddasParkTenureData	Ownership type	Housing_Tenure (Housing)	semanco:Housing_Tenure
LLSOAFuelPovertyIMD	LLSOA	Land (Territory)	semanco:Land
LLSOAFuelPovertyIMD	Households	Population_Number_Of_Households_Total (Population)	semanco:Population_Number_Of_Households_Total
LLSOAFuelPovertyIMD	Fuel Poor Households	Population_Number_Of_Households_In_Fuel_Poverty (Population)	semanco:Population_Number_Of_Households_In_Fuel_Poverty
LLSOAFuelPovertyIMD	INCOME SCORE	Population_Income_Score (Population) new	semanco:Population_Income_Score
LLSOAFuelPovertyIMD	IMD SCORE	Population_IMD_Score (Population) new	semanco:Population_IMD_Score

APPENDIX B. North Harbour mappings

Table B1contains the mappings carried out for the North Harbour database.

Table B1. Mappings carried out for North Harbour data migration

Data source	**Data name (in the data source, usually a column)**	**Standard Tables name, in parentheses the related sheet.**	**Class of the SEMANCO Ontology**
age	Age_Class	Age_Class (Building)	semanco:Age_Class
age	From	To_Year (Building)	semanco:To_Year
age	To	From_Year (Building)	semanco:From_Year
buildinguse	ID	Building_Use (b_use)	semanco:Building_Use
buildinguse	Building_use	Building_Use (b_use)	semanco:Building_Use
energydemand	ID	Building (Building)	sumo:Building
energydemand	HeatDemand	Energy_Need (Energy_quantities)	semanco:Energy_Need
energydemand	HotWaterDemand	Energy_Need (Energy_quantities)	semanco:Energy_Need
energydemand	ElectricityDemand	Delivered_Energy (Energy_quantities)	semanco:Delivered_Energy

energydemand	CoolingDemand	Delivered_Energy (Energy_quantities)	semanco:Delivered_Energy

APPENDIX C. Manresa mappings

Table C1 contains the mappings carried out for the North Harbour database.

Table C1. Mappings carried out for Manresa data migration

Data source	**Data name (in the data source, usually a column)**	**Standard Tables name, in parentheses the related sheet.**	**Class of the SEMANCO Ontology**
Census	ID	Building (Building)	sumo:Building
Census	NUMCOD	Address (Building)	semanco:Address, numcod data property
Census	DOMCOD	Address (Building)	semanco:Address, domcod data property
Census	ADRDESC	Address (Building)	semanco:Address, addressValue data property
Census	TITULACIO	Education_Level (Housing)	semanco:Education_Level
Census	SEXE	Household_Type (Housing)	semanco: Household_Type
Census	DESCPAISNAIX	Household_Origin_Country (Housing)	semanco:Household_Origin_Country
Data source	**Data name (in the data source, usually a column)**	**Standard Tables name, in parentheses the related sheet.**	**Class of the SEMANCO Ontology**
Cadastre	ID	Building (Building)	sumo:Building
Cadastre	FINREFCAD1	Cadastral_Reference	semanco:Cadastral_Reference, cadref1 data property
Cadastre	FINREFCAD2	Cadastral_Reference	semanco:Cadastral_Reference, cadref2 data property
Cadastre	LOCCOD	Address (Building)	semanco:Address, loccod data property
Cadastre	DOMCOD	Address (Building)	semanco:Address, domcod data property
Cadastre	NCL_ADRECA_DOM_F	Address (Building)	semanco:Address, addressValue data property
Cadastre	DESCDESTI	Ground_Floor_Use (Building) New	semanco:Ground_Floor_Use, ground_Floor_Usevalue data property
Cadastre	CODDESTI	Ground_Floor_Use (Building) New	semanco:Ground_Floor_Use, code data property
Cadastre	SUPLOCAL	Ground_Floor_Area (Building)	semanco:Ground_Floor_Area
Cadastre	STDDFICONS	Year_Of_Construction	semanco:Year_Of_Construction
Data source	**Data name (in the data source, usually a column)**	**Standard Tables name, in parentheses the related sheet.**	**Class of the SEMANCO Ontology**
Neighbourhood	Code	Neighbourhood (Territory)	sumo:Neighbourhood
Neighbourhood	Name	Neighbourhood (Territory)	sumo:Neighbourhood, name data property
Neighbourhoodincome	ID	Population_Mean_Income (Population)	semanco:Population_Mean_Income
Neighbourhoodincome	Income_percapita	Population_Mean_Income (Population)	semanco:Population_Mean_Income , income_CoefficientValue data property
Neighbourhoodincome	Income_Coefficient	Population_Mean_Income	semanco:Population_Mean_Income

		(Population)	, income_Per_CapitaValue data property
Neighbourhoodtemperature	ID	Conditioned_Space (Building)	semanco:Conditioned_Space
Neighbourhoodtemperature	Comfort_Temperature_Coefficient	Comfort_Temperature_Coefficient (cs_indoor_air_temperature)	semanco:Comfort_Temperature_Coefficient
Neighbourhoodtemperature	CS_Temperature_Heating_Mode	CS_Temperature_Heating_Mode (cs_indoor_air_temperature)	semanco:CS_Temperature_Heating_Mode
Neighbourhoodtemperature	CS_Temperature_Cooling_Mode	CS_Temperature_Cooling_Mode (cs_indoor_air_temperature)	semanco:CS_Temperature_Cooling_Mode
Data source	**Data name (in the data source, usually a column)**	**Standard Tables name, in parentheses the related sheet.**	**Class of the SEMANCO Ontology**
Bottomflooruvalue	ID	Bottom_Floor (CS_Envelope)	semanco:Bottom_Floor
Bottomflooruvalue	Property	Bottom_Floor_Type (CS_Envelope)	semanco:Bottom_Floor_Type
Bottomflooruvalue	Value	Bottom_Floor_U-value (CS_Envelope)	semanco:Bottom_Floor_U-value
Windowsparameters	ID	Window (CS_Envelope)	sumo:Window
Windowsparameters	Window_Glass_Type	Window_Glass_Type (CS_Envelope)	semanco:Window_Glass_Type
Windowsparameters	Window_Glass_U-Value	Window_Glass_U-value (CS_Envelope)	semanco:Window_Glass_U-value
Walluvalue	ID	Wall (CS_Envelope)	sumo:Wall
Walluvalue	Wall_U-value	Wall_U-value (CS_Envelope)	semanco:Wall_U-value
Buildingtypes	ID	Building (Building)	sumo:Building
Buildingtypes	Name	Building_Typology (Building)	semanco:Building_Typology
Skylightparameters	ID	Skylight (CS_Envelope)	sumo:Skylight
Skylightparameters	Window_Glass_Type	Skylight_Glass_Type (CS_Envelope)	semanco:Skylight_Glass_Type
Skylightparameters	Skylight_U-Value	Skylight_Glass_U-value (CS_Envelope)	semanco:Skylight_Glass_U-value
Skylightparameters	Skylight_Glass_g-value	Skylight_Glass_G-value (CS_Envelope)	semanco:Skylight_Glass_G-value
Percentagewindows	ID	Vertical_Enclosure (CS_Envelope)	semanco:Vertical_Enclosure
Percentagewindows	Wall_Type	Percentage_Of_Window (CS_Envelope)	semanco:Percentage_Of_Window
Percentagewindows	Percentage_Windows_Area	Type_Of_Wall (CS_Envelope)	semanco:Type_Of_Wall
Age	ID	Age_Class (Building)	semanco:Age_Class
Age	To	To_Year (Building)	semanco:To_Year
Age	from	From_Year (Building)	semanco:From_Year
Buildingsystemefficiencies	ID	Space_Heating_System (building_system)	semanco:Space_Heating_System
Buildingsystemefficiencies	Technical_Building_System	Space_Heating_System_Efficiency (building_system)	semanco:Space_Heating_System_Efficiency
Buildingsystemefficiencies	System_Efficiency	Space_Heating_System_Type (building_system)	semanco:Space_Heating_System_Type
Roofuvalue	ID	Roof (CS_Envelope)	sumo:Roof
Roofuvalue	Roof_U-value	Roof_U-value (CS_Envelope)	semanco:Roof_U-value