European Commission

JRC TECHNICAL REPORT

Harmonisation of datasets of Energy Performance Certificates of buildings across Europe

ELISE Energy and Location Applications
Final Report

Serna-González, V., Hernández Moral, G., Miguel-Herrero, F., Valmaseda, C. (CARTIF)

Martirano, G., Pignatelli, F., Vinci, F. (JRC)

2021

ELISE

Enabling digital government through geospatial & location intelligence

Joint Research Centre

EUR 30795 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Francesco
Address: Pignatelli
Email: Francesco.PIGNATELLI@ec.europa.eu
Tel.: +39. 033278-6319

EU Science Hub
https://ec.europa.eu/jrc

JRC124887

EUR 30795 EN

PDF ISBN 978-92-76-40827-7 ISSN 1831-9424 doi:10.2760/500135

Luxembourg: Publications Office of the European Union, 2021

© European Union, 2021

The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2021, *(unless otherwise specified)*

How to cite this report: Serna-González, V., Hernandez Moral, G., Miguel-Herrero, F., Valmaseda, C., Martirano, G., Pignatelli, F., Vinci, F., *ELISE Energy & Location Applications: Use Case "Harmonisation of Energy Performance Certificates of buildings datasets across EU" – Final Report*, EUR 30795 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-40827-7, doi:10.2760/500135, JRC124887.

Contents

Acknowledgements

The results presented in this report were achieved thanks to the close collaboration with the Energy Saving and Efficiency Department of the energy agency of Castilla y León region - EREN (Ente Público Regional de la Energía de Castilla y León, Spain), acting as use case partner in the execution of the reported activities.

Special thanks to Lorena Hernández Quirós of JRC/B.6 for her thorough review and her very valuable comments and suggestions.

Authors

- Victor Iván Serna-González, Gema Hernández Moral, Francisco Javier Miguel-Herrero, César Valmaseda (CARTIF Foundation, Valladolid, ES)
- Giacomo Martirano, Francesco Pignatelli, Fabio Vinci (JRC B.6, Ispra, IT)

Abstract

Nowadays, datasets of Energy Performance Certificates (EPC) of buildings in Europe are highly heterogeneous: they follow different EPC schemes making their comparison challenging.

This report shows in the first place how a European-wide harmonised EPC scheme would be very beneficial to compare EPC datasets across countries and regions. Secondly, it showcases how Regional Energy Agencies can actively support energy efficiency policies.

The study proposes a European data model named EPC4EU, reusing a methodology developed in 2017 to harmonise EPC in Italy and testing it with real EPC datasets from Castilla y León (Spain). The methodology can be reused to generate new versions of the EPC4EU data model to harmonise EPC datasets from any Member State.

Furthermore, the study documents a series of difficulties encountered during the harmonisation process and the solutions adopted to overcome them.

Finally, the study also includes suggestions to use the resulting harmonised data in the QGIS software and documents the development of a web application designed to make available the harmonised dataset to non-GIS experts. The issues encountered in the web app adaptation and respective solutions adopted have been documented to ease possible future web app extensions.

The described activities have been executed in the frame of the Energy & Location Applications of the ELISE (European Location Interoperability Solutions for e-Government) action of the ISA2 (Interoperability solutions for public administrations, businesses and citizens) Programme.

Executive Summary

The Digital Economy Unit of the EC Joint Research Centre (JRC), in cooperation with other services of the European Commission, is coordinating the "European Location Interoperability Solutions for e-Government (ELISE)", Action 10 of the ISA2 (Interoperability Solutions for Public Administrations, Business and Citizens) Programme.

The ISA2 Programme supports long-standing efforts to create a European Union free from electronic barriers at national borders. It facilitates interaction between European public administrations, businesses and citizens, by enabling the delivery of interoperable cross-border and cross-sector public services and by ensuring the availability of common framework and solutions.

The ELISE Action is a package of solutions facilitating efficient and effective electronic cross-border or cross-sector interactions between European public administrations, citizens and businesses, in the domain of location information and services. Within it, the ELISE Energy & Location Applications consist of a series of use cases showing how location data can support different types of stakeholders engaged in energy policies' cycle at different geographical scales, from local up to EU level.

In particular, the use case named "Harmonisation of datasets of Energy Performance Certificates of buildings datasets across Europe" aims to demonstrate the benefits of harmonising EPC (Energy Performance Certificates) datasets across European countries and regions by reusing the methodology provided in the use case "INSPIRE Harmonisation of EPC of buildings datasets in Italy"1 and applying it in the Castilla y Leon region (Spain).

The challenge addressed by this report is that the EPC datasets in Europe are very heterogeneous across countries and regions, as they follow different schemes making its comparison very difficult.

Having available EPC datasets harmonised across Europe would be beneficial firstly at the European level, allowing the comparison of EPCs datasets from different regions and countries and secondly for Regional Energy Agencies, as valuable support to energy efficiency policies.

The use case has been executed into three steps:

- **Step 1,** illustrates the EPC4EU data model. EPC4EU has helped producing harmonised datasets from Italy and Spain. Most importantly, the same methodology can be easily reused by any Member State to generate harmonised EPC datasets.
- **Step 2**, shows how real EPC datasets in Castilla y León have harmonised using the EPC4EU model. Results, process, and difficulties encountered during the transformation process along with the workarounds found are described in detail. Practical suggestions to visualise the resulting harmonised datasets in QGIS are included.
- **Step 3**, documents the development of a web application capable of making accessible the harmonised dataset to non GIS experts. A first prototype produced in 2017 has been analysed, adapted, and reused to make accessible simultaneously EPC datasets sample from Spain and Italy. The issues encountered in the web app adaptation and the related solutions adopted, have been extensively documented to be re-used in the context of possible future web app extensions.

A series of resources developed during the use case (e.g. UML data models, mapping tables, XSD application schemas, examples of harmonised data) and referred to within this report, are available for download in the dedicated JoinUp page[2].

1 https://publications.jrc.ec.europa.eu/repository/handle/JRC104587
2 https://joinup.ec.europa.eu/node/704540

1 Introduction

The ELISE Energy & Location Applications consist of a series of use cases aiming to show how location data can support different types of stakeholders engaged in the energy policies' cycle at different geographical scales, from local to EU level.

In particular, one of the use cases, named "Harmonisation of datasets of Energy Performance Certificates of buildings datasets across Europe", aimed to show the benefits of harmonising EPC datasets across EU countries and regions, re-using the methodology applied in the use case "INSPIRE Harmonisation of EPC of buildings datasets in Italy"[3] and applying it in the Castilla y Leon region of Spain.

The problem addressed by the use case is that the EPC datasets from different countries and regions of the EU are very heterogeneous following different EPCs schemes, and then their comparison is very challenging.

The availability of harmonised EPC datasets across Europe would be beneficial firstly at the EU level, allowing the comparison of EPCs datasets from different regions and countries and secondly for Regional Energy Agencies, as a valuable support to energy efficiency policies.

The use case has been executed into three steps:

- **Step 1,** illustrates the EPC4EU data model. EPC4EU has helped producing harmonised datasets from Italy and Spain. Most importantly, the same methodology can be easily reused by any Member State to generate harmonised EPC datasets.
- **Step 2**, shows how a subset of real EPC datasets in Castilla y León has been harmonised using the data model EPC4EU.
- **Step 3**, documents the development of a web application capable of making accessible the harmonised dataset to non GIS experts.

The overall methodology used to create a new version of the EPC4EU data model capable of harmonising EPC datasets of Spain and Italy, starting from a previous version of the EPC4EU data model developed in another use case to harmonise EPC datasets adhering to the Italian data model for EPC, is described in section 2 of this report.

The methodology consists of nine steps described in details in sections from 4 to 9. In Annex further details on the national Spanish EPC data model are provided.

Once the mapping between the source data model and the EPC4EU target data model has been performed (in Step 1), the next step was to harmonise real data using the updated version of the "EPC4EU" as the target data model. The real data available for this purpose came from the EPC register of the Castilla y Leon (CyL) region, provided by EREN (the regional organism of Castilla y Leon for the Energy Management). A sample of the database with 2134 instances was made available for this aim[4].

The data transformation process is described in Section 10.

In Section 11, a series of difficulties encountered during the transformation process is described and the corresponding solutions adopted. This section has been divided into two main subsections. The first one explains the problem of filling in the information related to the *Building* and *BuildingPart* feature types. The second subsection identifies the problems encountered during the harmonisation process. Practical suggestions on properly visualising the harmonised data in QGIS (one of the most popular GIS desktop tools) is provided in Section 12.

The last activity executed during Step 2 of the use case was related to the visualisation of the GML file containing the harmonised EPC dataset in a GIS desktop tool. Subsequently, to make accessible the harmonised dataset also by non-GIS experts, a web application usable from a browser has been developed and the related activities described in this report.

[3] https://publications.jrc.ec.europa.eu/repository/handle/JRC104587

[4] It is important to note that the database contains not only the information stored in the EPC but also information added during the uploading phase and also during the validating and storing phase in the database.

In 2017, a first prototype of a similar web application was created for the use case "INSPIRE Harmonisation of EPC of buildings datasets in Italy", using two different web mapping libraries: OpenLayers[5] and Leaflet[6].

Starting from this prototype, an updated version of the web app[7] exposing harmonized EPC datasets coming from different sources, specifically from Castilla y León (ES) and from Trento (IT), representing the two EPC datasets harmonised within two use cases of the ELISE Energy & Location Applications, has been developed.

The datasets and the target data model used during the harmonisation are presented in section 13.

The web application is then described in section 14, starting from an analysis of the first prototype, then describing the workflow for the re-use of this prototype and finally presenting the final results to visualise all the harmonized datasets with the same tool.

Conclusions are outlined in section 15.

A series of resources developed during the use case (e.g. UML data models, mapping tables, XSD application schemas, examples of harmonised data) and referred to within this report, are available for download in the dedicated JoinUp page[8].

5 http://openlayers.org/
6 http://leafletjs.com/
7 https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/epc4eu/webapp/
8 https://joinup.ec.europa.eu/node/704540

2 Methodology for the analysis and the mapping exercise

This section explains the methodology for the analysis of the EPC-ES data model and the contextual mapping exercise toward the EPC4EU data model. The final objective of this analysis is to identify candidate elements of a new version of the EPC4EU data model, capable of mapping both EPC-IT and EPC-ES core datasets.

The first version of the EPC4EU data model (v2.1) has been developed, as an extension of the INSPIRE Buildings2D data model, to map all information present in the Italian data model for the Energy Performance Certificates.

Other Member States could follow this mapping process to update the EPC4EU data model to have a more complete data model considering the information available in the national Energy Performance Certificates.

In this section, firstly, the files that have been used for the needed analysis previous to the mapping processes are listed. After that, the tools used during all the process are also identified. Finally, the methodology followed for the mapping exercise (and for the harmonisation process) is described.

2.1 Files and tools used

2.1.1 Files used

For the whole process of mapping and harmonising the Spanish dataset following the EPC target data model, a wide variety of files has been used. In this section, the different types of files are listed, specifying the format of the files and the content of these. The relevant ones are available for download in the dedicated JoinUp page.

- Schemas of the data models and examples:
  - DatosEnergeticosDelEdificioSchema20.xsd: XSD application schema for Spanish EPC data model
  - epc4eu.xsd: XSD application schema for EPC4EU data model. Different versions of the file were used, mainly versions v2.1 (for the first analysis) and v3.
  - EjemploTerciario.xml[9]: virtual example of XML EPC file for the tertiary sector, provided by the Spanish Ministry of Development in collaboration with CSIS (Consejo Superior de Investigaciones Cientificas)
  - EjemploResidencial.xml[10]: virtual example of XML EPC file for the residential sector, provided by the Spanish Ministry of Development in collaboration with CSIS (Consejo Superior de Investigaciones Cientificas)
  - Certificación estado actual_completo_2.xml: a real example of XML EPC file, generated by an authorised certifier.

    These schemas can also be opened and modified with a normal text editor.
- EPC4EU UML project:
  - *EPC4EU_UML_Model_v2.1.eap:* Model of the EPC4EU in UML (Unified Modeling Language)
  - Spanish cadastre files[11]:
    - A.ES.SDGC.BU.47900.building.gml: example *Building* file from Spanish cadaster
    - A.ES.SDGC.BU.47900.buildingpart.gml: example *BuildingPart* file from Spanish cadaster
    - A.ES.SDGC.BU.47900.otherconstruction.gml: example *Otherconstruction* file from Spanish cadaster
- EPBD Documentation:

[9] https://visorxml.codigotecnico.org/static/docs/EjemploTerciario.xml
[10] https://visorxml.codigotecnico.org/static/docs/EjemploResidencial.xml
[11] The three files are included in a zip that could be downloaded from http://www.catastro.minhap.es/INSPIRE/Buildings/47/47900-VALLADOLID/A.ES.SDGC.BU.47900.zip

- *CELEX 32010L0031 ES TXT.pdf*[12]: DIRECTIVE 2010/31/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
- *CELEX 32018L0844 ES TXT.pdf*[13]: DIRECTIVE (EU) 2018/844 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

— Spanish EPCs documentation:

- 20150625 - Informe evaluación energética edificio en formato electrónico XML.pdf. A PDF file titled "Building energy assessment report in XML electronic format."[14]
- 052000NEIS100_ES.pdf: Standard ISO 52000-1 2017. Global evaluation of EPBD. Part 1: General framework and procedures[15]
- 052000NXCN100_ES.pdf: Standard ISO 52000-2 2017. Global evaluation of EPBD. Part 2: Explanation and justification of the standard[16]

— Italian EPC documentation:

- DM_Linee_guida_APE_allegato1.pdf
- DM_Linee_guida_APE_appendiceB_Template.pdf
- Manuale-Operativo-Redazione-APE_16-febbraio-2016.pdf
- Modulo C - Edificio gestione dati generali.pdf

— INSPIRE documentation:

- Inspire_dataspecification_bu_v3.0.pdf: INSPIRE Data Specification on Buildings[17]
- EPC4EU_MT_v2.1.xlsx: Mapping table, used for the documentation of the mapping process.

2.1.2 Tools used

The tools used for the mapping process are the following:

— *Enterprise Architecture Viewer 14.1. Build 1429*: The tool is a read-only version of EA intended for distribution with UML models, and free of charge, supporting all viewing functions. This tool was used for analysing the UML model.

— *hale studio: Version 3.5.0.release*: *hale studio* is an application for helping in the creation of schema mappings, allowing the transformation of data that conforms to the source schema to data that is compliant to the target schema. The tool was used for the creation of the mappings

— *Notepad++*: This tool is a text editor with some special characteristics that allow visualising the data more friendly. Also, it can install plug-ins to increment its functionalities. Notepad++ was used to visualise the XSD schemas, the EPCs in XML format and the GML files from cadaster.

— *XML Copy editor*: the tool is an editor able to edit and easily validate GML files. *XML Copy editor* was used to validate the XML using the XSD schemas.

2.2 Methodology

The methodology followed for the mapping process, and the harmonisation process is explained below. It is important to note that some of the steps were iterative because information already analysed must be re-analysed once knowledge of the problem increases in subsequent steps.

12 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0031&from=EN
13 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L0844&from=EN
14 https://visorxml.codigotecnico.org/static/docs/20150625%20-%20Informe%20de%20evaluaci%C3%B3n%20energ%C3%A9tica%20del%20edificio%20en%20formato%20electr%C3%B3nico%20(XML).pdf
15 https://www.iso.org/standard/65601.html
16 https://www.iso.org/standard/68232.html
17 https://inspire.ec.europa.eu/id/document/tg/bu

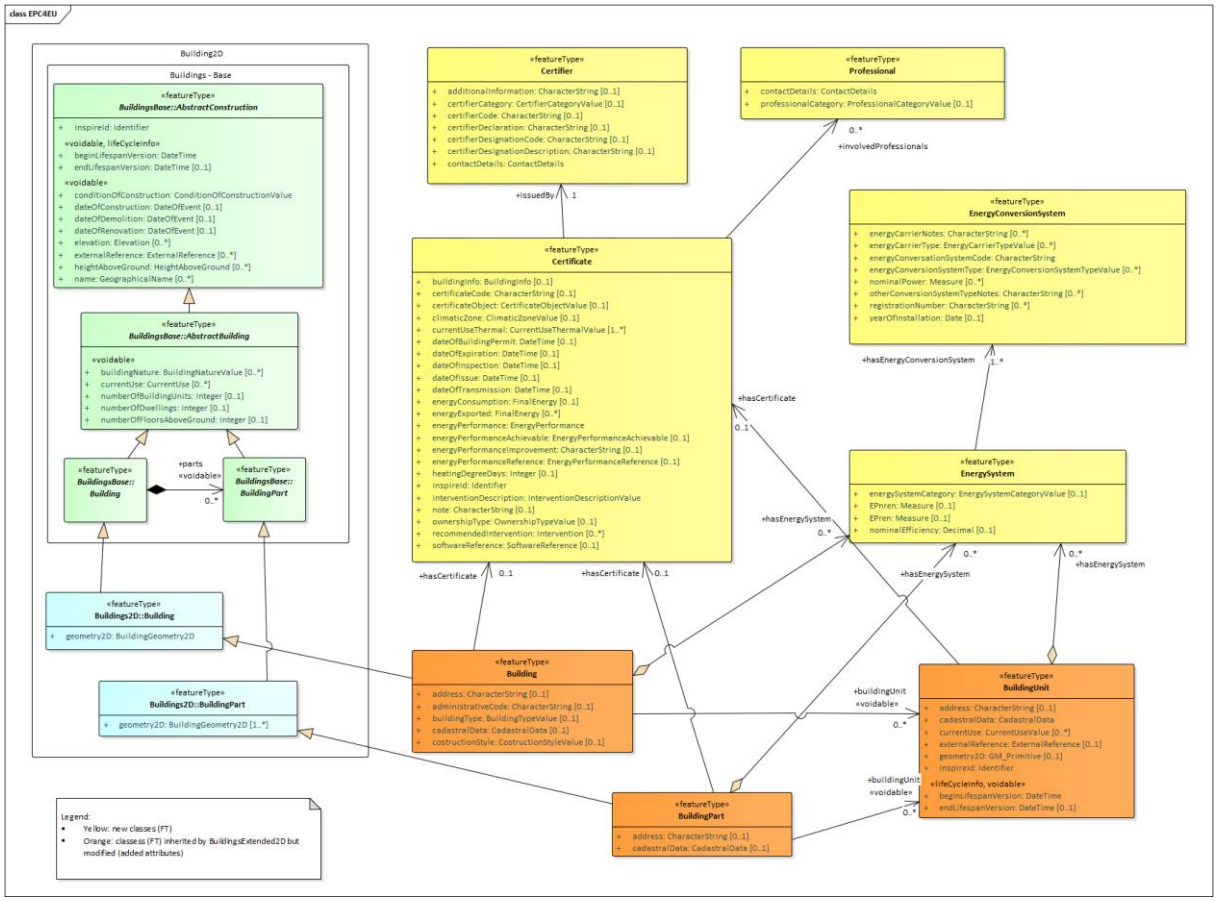
The steps related to the mapping process are:

1. Study of the INSPIRE Data Specification on *Buildings*. Specifically, the *Inspire_dataspecification_bu_v3.0.pdf* file was analysed to have a clearer vision of the INSPIRE directive in general and the Buildings data model in particular. This is the base of the EPC4EU data model.
2. Study the EPBD directives (2010 and 2018). The versions published in the Spanish language were analysed: *DIRECTIVE 2010/31/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL* and *DIRECTIVE (EU) 2018/844 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL.* The analysis of the information contained in this directive was important to understand the generation of the certificates (the meaning of the different terms, the acronyms, etc.)
3. Study of the Spanish normative: Documents Standard ISO 52000-1 2017 and Standard ISO 52000-2 2017 about energy efficiency in buildings.
4. Study of the Spanish EPC Data Model. Complementing the normative information with the real and virtual examples of the EPCs (.xml files).
5. Analysis of the XSD schema for the Spanish EPC data model using the XSD itself and the documentation related *20150625 - Informe evaluación energética edificio en formato elec XML.pdf.* This analysis is closely related to the previous step.
6. Analysis of the XSD schema for the EPC4EU data model and with the help of the EPC4EU UML project. Also Italian documentation was consulted to understand some of the parameters that mainly came from the Italian EPC data model.
7. Completing the *EPC4EU_MT_v2.1.xlsx* mapping table considers the target data model and the source data model, in this case, the Spanish data model. Step 7 finalises the **mapping** process. However, some steps are needed for the subsequent process, the **harmonisation**:
8. Using *the hale studio tool* to map the target data model and the source data model. In this step, it will be confirmed the suitability of the mapping indicated in the mapping table.
9. Analysis for the data model strictly related to the Building section (inherited by the INSPIRE Buildings2D data model). It is needed to study the information contained in the EPC and that contained in Spanish cadaster to know what information could be mapped to the building section of the target data model with more accuracy.

3 Target Data Model (v2.1)

The version of the target data model for the EPCs used for the mapping exercise has been developed, taking as a basis real EPC datasets found in Trento (Italy), already harmonised according to the Italian EPC data model. The target data model has been named EPC4EU. A UML representation of the data model is shown in Figure 1.

Figure 1. EPC4EU data model, version 2.1.



Source: own creation, JRC, 2019.

4 EPCs Management process

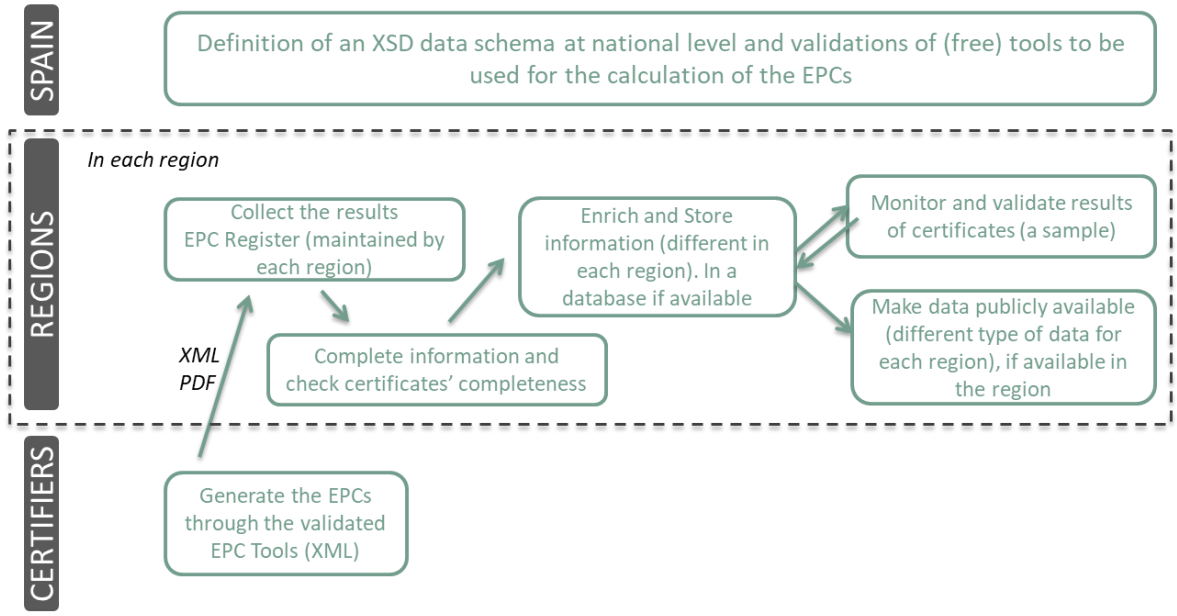
The Energy Performance of Buildings Directive establishes the obligation of Member States to issue Energy Performance Certificates and specifies some criteria that need to be followed. Each country must implement this Directive to ensure its compliance. In fact, in each Member State, the issuing process follows the same structure, but there are several differences between them. These differences are in the calculation methods, the input data required for these calculations, the way to manage the register (at the national or regional level, for example).

In this section, the management process for the EPCs in Spain is explained, emphasizing the description regarding the process in Castilla y Leon region.

4.1 Management process for the EPCs in Spain

In this section, the process followed in Spain for the management of the EPCs is explained. A summary can be seen in Figure 2.

Figure 2. Management process for the EPCs followed in Spain.



Source: own creation, CARTIF, 2019.

In the case of Spain, the data to be presented has been defined and also the way to present this data. There are two ways to present the data: in PDF format or XML format. Both ways can be presented to the administration in Spain. The XML file contains all of the information required in an EPC in Spain and also information used during the calculation process.

For the creation of both EPC files, different EPC calculation tools are available. Only those EPCs issued by official tools recognised by the Spanish Ministry for the Ecological Transition and the Demographic Challenge are admitted into the EPCs registers[18]. These tools (HULC, CERMA, CEX and CE3X), created by private entities, have been made available freely to the public to issue EPCs. These will be deployed by a professional certifier, who will be able to evaluate the performance of both existing and new buildings (depending on the tool) and also residential or tertiary buildings, either individual (individual assessment of a dwelling or commerce) or multiple (assessing a complete building block).

Consequently, the requirements and methodologies to calculate an EPC are set at the national level, having common EPCs (in PDF and XML versions) in the whole country and energy labels.

[18] https://energia.gob.es/desarrollo/EficienciaEnergetica/CertificacionEnergetica/DocumentosReconocidos/Paginas/procedimientos-certificacion-proyecto-terminados.aspx

Although the model has been defined at the country level in Spain, the jurisdiction to manage the EPCs has been established to the regional level. The PDF and XML file common nationwide has to be submitted to the regional EPC registers. Also, other complementary services related to the EPCs, for example, the possibility to provide a catalogue with public data from EPCs, are managed by each region independently.

The processes that are dependants on each region are:

- The collection of the EPCs files.
- The completion of the information and the verification of the completeness
- The storage of the EPCs files (organized in a register at the regional level)
- The monitoring and validation of the values in the certificates, typically a sample of the EPCs
- The publication of part of the EPCs values, if possible.

Depending on the region, these processes are present or not. Besides, some of these processes are digitally automated or not. In regions where the storage is done digitally and automatically, processes as verification and validation are more efficient and accurate.

The table below (Table 1) shows the processes related to the EPCs available online by each autonomous community. In Spain, there are 17 autonomous community and two autonomous cities, so 19 regions in total. The possibility to register the EPCs in an online way is available in 16 of these 19 regions.

Another service that can be provided is the publication of the registered EPCs (or some values of them). In the case of Spain, this service is available in 12 autonomous communities. However, in each region, the method to search the data and the data offered publicly is very heterogeneous. For example, only in 4 of these regions, the information is geo-located, and only in 3 of them, the information is more detailed than simply the value of the Energy label.

Another issue to be taken into account is that only in two regions is it possible to automatise collecting information from the EPCs register in an automatised and public way using Application Programming Interface (APIs). These regions are Cataluña and Castilla y Leon.

Table 1. EPCs processes addressed by each autonomous community in Spain (with links in some of the cells)

Autonomous community	EPCs Registry	Online registry address	Data query	Address for data query	Geolocated info	Provided data in the query	Automatable download
Andalucía	On-site and online	Yes	Form for selecting certificates	Yes	No	Only one label (not exact value)	No
Aragón	On-site and online	Yes	Form for selecting certificates	Yes	No	Only two labels (with value)	No
Asturias	On-site and online	Yes	Only on-site	No	N/A	N/A	No
Baleares	On-site and online	Yes	Only one label (not exact value)	Yes	No	Only one label (not exact value)	No
Canarias	On-site and online	Yes	Only on-site	No	N/A	N/A	No
Cantabria	On-site and online	Yes	Only on-site	No	N/A	N/A	No
Castilla la Mancha	On-site and online	Yes	Form for selecting certificates	Yes	No	Only two labels (without value)	No

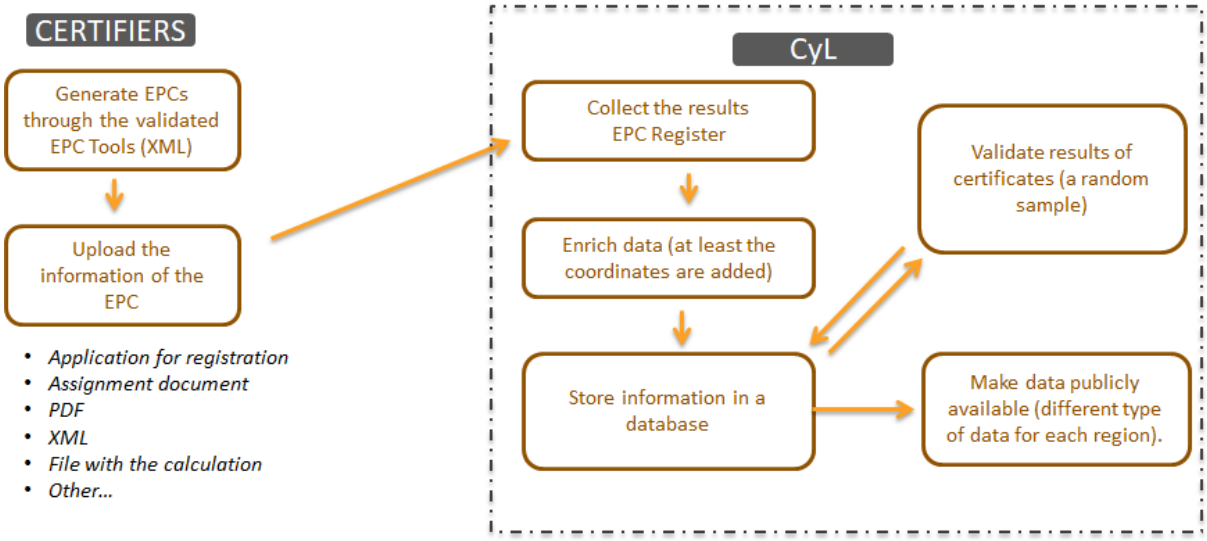
Autonomous community	EPCs Registry	Online registry address	Data query	Address for data query	Geolocated info	Provided data in the query	Automatable download
Castilla y León	On-site and online	Yes	Form for selecting certificates+API	Yes	Yes	Complete	Yes
Cataluña	On-site and online	Yes	Form for selecting certificates + API	Yes	Yes	Complete	Yes
Ceuta	On-site registration	No	Only on-site	No	N/A	N/A	No
Extremadura	On-site registration	No	Only on-site	No	N/A	N/A	No
Galicia	On-site and online	Yes	Form for selecting certificates	Yes	No	Only two labels (with value)	No
La Rioja	On-site and online	Yes	Form for selecting certificates	Yes	Yes	Only two labels (with value)	No
Madrid	On-site and online	Yes	Form for selecting certificates	Yes	Yes	Only label	No
Melilla	On-site registration	No	Only on-site	No	N/A	N/A	No
Murcia	On-site and online	Yes	Only on-site	No	N/A	N/A	No
Navarra	On-site and online	Yes	Form for selecting certificates	Yes	Indirectly	Complete	No
País Vasco	On-site and online	Yes	Form for selecting certificates	Yes	No	Only label	No
Valencia	On-site and online	Yes	Form for selecting certificates	Yes	No	Only label	No

Source: own creation, CARTIF, 2019.

4.2 Management process for the EPCs in Castilla y León region

In this section, the process followed in Castilla y Leon region to manage the EPCs is explained. A picture of this process can be seen in the following scheme shown in Figure 3.

Figure 3. Management process for the EPCs followed in Castilla y Leon region.



Source: own creation, CARTIF, 2019.

The processes are similar to those explained in the previous section, but it is highlighted that these processes follow a workflow with a digital infrastructure. In the Castilla y Leon case, the organism in charge of managing the EPCs registration is EREN (*Ente Regional de la Energía*).

The processes are the following:

1. The user (certifier) creates the certificate with the software. The documented generated are three:
    a. EPC in XML format. This file contains all the information mandatory for the Spanish EPC
    b. EPC in PDF format. The EPC in PDF format contains a part of the information (practically all the information) but in a more visual way
    c. File with the calculation. This file is exported by the software application and used to create the EPCs and contains all the information used to calculate this. Depending on the tool, the file would have a different format corresponding to the used tool.
2. The user uploads the EPC. The information added is:
    a. EPC in XML format
    b. EPC in PDF format
    c. File with the calculation
    d. Application for registration. This application has to be filled to upload the EPC. It is important because more data that are not present in the EPC are added here. At least the information added is:
        i. The use of the building (the use for the tertiary sector is specified)
        ii. Software used (more information than in the EPC could be added)
    e. Assignment document. Document where the owner grants permission to register it on his behalf. It has a standard format.
    f. Others. Other complementary files can be uploaded
3. EREN collects the information.

4. EREN enriches the data by adding additional information and classifying some of them. At least the coordinates are added (in a semiautomatic way), and some parameters of the energy conversion systems are mapped using *code lists.*
5. The information is stored in the database
6. EREN carries out validations of a random sample of the EPCs, doing a new certification process and comparing the results
7. Some data of the certificates are shown on the web page of the EREN[19]. The data are available online to be consulted. The data to be consulted are offered in two different ways:
    a. In a consultation service
    b. Geo-located in a map
    c. Through an API, by downloading the entire data of each province.

[19] https://servicios3.jcyl.es/seye/login.jsp

5 Target data model (v2.1) and EPC Spanish National data model

5.1 Analysis of the Spanish EPC structure

The Energy Performance Certificates issued in Spain share the same structure and information at the national level. In Spain, the EPCs currently have to be submitted in PDF version and XML version. While both are similarly structured, the PDF version is deployed as information means and holds the most important data, while the XML version can hold a higher amount of data, some of which is optional. The information contained in both documents is shown in the document “Informe de evaluación energética del edificio en formato XML” ([20]).

In this case, the matching between the EPC Spanish National data model and the target data model will be done using the XML format file, in which the information is more completed and organized in a way more digitally manageable.

5.1.1 Analysis of the Spanish EPC content

In this section, a short description of the sections encountered in the EPC in XML format is shown. To find out what specific data is included in which sections of the PDF version, please refer to Annex 1. Below, the different sections of the XML file are listed and explained.

It is worth mentioning that even though the amount of data presented in the XML version is much larger than the one presented in the PDF format, many attributes are identified as "optional", others refer only to residential or tertiary buildings and another express the compliance of the building with the national building code. In the information below, only the mandatory attributes have been mentioned, whereas in annex 1 a complete view of the XML attributes can be observed. So, the parts of the EPC in XML format are:

- Building identification (*Identificación del edificio*). The building is identified by its name, address, municipality, climatic zone etc. Also, the normative framework is expressed, as well as its cadastral reference. The type of building or part is defined: existing vs new building, residential vs tertiary and if it is referred to as a whole building or just a part of it. The dataset "Certification method deployed and version".
- Certifier's data (*Datos del técnico certificador*). Identification of the certifier: name, address, email, etc., and type of title that allows them to issue EPCs and the procedure deployed to issue the EPC.
- General data and building geometry (*Datos generales y geometría del edificio*). Information about the geometry of the building and also about the floor surface, building picture, building plan, DHW demand, number of floors above and below ground, volume, percentage of heated/cooled surface, window / opaque surface ratio, ventilation etc.
- Building envelope data (*Datos de la envolvente térmica*). Description of opaque surfaces (name, type, surface, U-value and how it has been calculated) and openings (name, type, surface, U-value, solar factor, and how they have been obtained). Extended and non-mandatory data on what is provided include information on layers of opaque surfaces, thermal bridges and the orientation of the listed elements.
- Building thermal energy systems' data (*Datos de instalaciones térmicas*). For every heating and/or cooling generator: name, type, nominal power, efficiency, type of energy and how they have been obtained. Daily hot water demand and per system: name, type, nominal power, efficiency, type of energy and how they have been obtained. Also, "nominal performance" is present. It is worth mentioning that the categories referring to auxiliary heating or cooling systems, cooling towers and ventilation and pumping only apply to tertiary buildings.
- Lighting systems' data (*Datos de instalaciones de iluminación*). Space, installed power, efficiency, average illuminance and how they have been obtained.
- Functioning and occupancy conditions (*Condiciones de funcionamiento y ocupación*). Space, surface and use type. All of these values refer only to tertiary buildings.
- Renewable energy (*Energías renovables*). Description of renewable energy sources in the building and percentage of the energy consumption covered. Besides, it includes two datasets that represent the savings in primary non-renewable energy and the emissions reduction. However, these are optional values.

[20] https://energia.gob.es/es-es/Participacion/Documents/propuesta-doc-reconocido-certificacion-energetica/Informe-evaluacion-energetica-edificio-formato-electronico.pdf

— Demand (*Demanda*). In this category, it can be observed that all the values are optional except for global energy demand (heating, cooling, DHW etc.). The only two datasets appearing in the PDF are the heating and cooling demand, are optional in the XML. Other values to be found in this section refer to demand of a reference building according to legislation or compliance with the Spanish building code, all of them optional.

— Consumption (*Consumo*). Except for a general correction factor, all of the values are optional in this category. The values present are global, heating, cooling DHW and lightning consumption. Even though they are optional, the additional datasets provided in the XML are related to energy conversion factors from final energy to non-renewable primary energy and from final energy to CO2 emissions in every type of fuel. Also some values related to compliance with the Spanish building code are included.

— Emissions (*Emisiones*). All of the datasets included in the PDF version correspond to the ones expressed in the XML version, which document the CO2 emissions of the building concerning heating, cooling, DH, electricity etc. However, all of them except for the global emissions are marked in the XML as optional.

— Building energy label (*Calificación energética del edificio*). The building energy label is described according to different criteria (such as demand, non-renewable primary energy, and CO2 emissions) both in the XML version and the PDF version. All of the values correspond in both versions; however, only global values and the description of the different energy label grading systems are marked as mandatory.

— Improvement measures (*Medidas de mejora*). All of the datasets are present in the PDF version. However, the ones marked as mandatory in the XML version correspond to global labels, that is, not disaggregated in heating, cooling, lighting etc.

— Tests, checks and building inspections performed by the certifier (*Pruebas, comprobaciones e inspecciones realizadas por el técnico certificador*). Section devoted to referring to the visits performed to the building/dwelling and any relevant comment that needs documenting.

— Customised data (*Datos personalizados*). Section devoted to the inclusion of another type of data not contemplated in the previous sections can be generated by the different software. This category is not contemplated in the PDF version and the XML version; it is marked as optional.

5.2 Matching table between EPC4EU v2.1 and EPC Spanish National data model

After analysing the EPC Spanish National data model, the next step compares it against the target data model, named EPC4EU v2.1[21]. The matching table between EPC4EU v2.1 and EPC Spanish National data model reflects this comparison, indicating the elements from the EPC4EU model that are present in the EPC Spanish National data model and their characteristics (multiplicity, format, description, etc.). A simplified matching table can be seen in Annex 2. The complete version, in the original excel format, is provided as an external annex to the report.

[21] https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/epc4eu/data-model/2.1/html/

6 Results of the analysis

The final objective of this analysis is to identify candidate elements of a new version of the EPC4EU data model, capable of mapping both EPC-IT and EPC-ES core datasets.

The final objective of the analysis is to adapt the target data model taking into account the Spanish model. For this reason, the results to be obtained are those related to the changes to be applied in the data model (that can be explicitly indicated in the excel mapping table) and the values that some parameters could get, especially for the code list values. Code lists are lists of possible values that a determined attribute could have.

6.1 Results of the analysis 1: Code list values added

In the study, some code lists have been updated, taking into account the values that the parameters of the Spanish EPCs could have. In the following tables, the different code lists updated can be seen and how the values are classified. Sometimes a code list has to be increased with new values, and sometimes the pre-existent values in the code list are directly mapped with the values of the Spanish EPC. The red values added in the "EPC4EU" column indicate values created to embrace the data belonging to the ES- EPC possible values.

Table 2. Code list for EnergySystemCategoryValue

EnergySystemCategoryValue	
EPC4EU	**ES**
spaceHeating	Calefaccion
spaceCooling	Refrigeracion
domesticHotWater	ACS
lighting	iluminacion

Table 3. Code list for EnergyCarrierTypeValue

EnergyCarrierTypeValue	
EPC4EU	**ES**
electricity	ElectricidadPeninsular ElectricidadBaleares ElectricidadCanarias ElectricidadCeutayMelilla
naturalGas	GasNatural
LPG	GLP
coal	Carbon
solidBiomass	BiomasaPellet
liquidBiomass	Biocarburante

gaseousBiomass	BiomasaOtros
diesel	GasoleoC

Table 4. Codelist for energyPerformanceLabelValue

energyPerformanceLabelValue	
EPC4EU	**ES**
B	B
C	C
D	D
E	E
F	F
G	G
A	A
notQualifiable	N.C

Table 5. Code list for ClimaticZoneValue

ClimaticZoneValue	
EPC4EU	**ES**
Spain/A3	A3
Spain/A4	A4
Spain/B3	B3
Spain/B4	B4
Spain/C1	C1
Spain/C2	C2
Spain/C3	C3
Spain/C4	C4
Spain/D1	D1
Spain/D2	D2
Spain/D3	D3

Spain/E1	E1

Table 6. Code list for ScopeValue

ScopeValue	
EPC4EU	**ES**
existentConstructionCertification	CertificacionExistente
existentConstructionVerification	VerificacionExistente
existentConstructionCertificationVerification	CertificacionVerificacionExistente
newConstructionCertification	CertificacionNuevo
newConstructionVerification	VerificacionNuevo
newConstructionCertificationVerification	CertificacionVerificacionNuevo

Table 7. Codelist for SurfaceTypeValue

SurfaceTypeValue	
EPC4EU	**ES**
conditionedSurface	SuperficieHabitable

Table 8. Codelist for VolumeTypeValue

VolumeTypeValue	
EPC4EU	**ES**
conditionedVolume	VolumenEspacioHabitable

Table 9. Codelist for QualityValue

QualityValue	
EPC4EU	**ES**
A	A
B	B
C	C
D	D

E	E
F	F

Table 10. Codelist for CertificateObjectValue

CertificateObjectValue	
EPC4EU	**ES**
wholeBuilding	BloqueDeViviendaCompleto ViviendaIndividualEnBloque EdificioUsoTerciario
realEstateUnit	ViviendaUnifamiliar LocalUsoTerciario

6.2 Results of the analysis 2: Proposed changes based on the analysis of the Spanish EPC model and the comparison with the EPC4EU

This analysis is focused on the EPC elements different from those related to building (*Buildings*, *BuildingParts* and *BuldingUnits*), so it is focused on the elements: *Certificate*, *Certifier*, *Professional*, *EnergySystem* and *EnergyConversionSystem*.

The analysis has been done taking into account the following aspects:

- Some important attributes not present in the EPC4EU target data model and present in the ES EPC data model could be incorporated.
- The name of the attributes has to be clear enough and should include the Spanish attributes.
- The multiplicity of all the attributes has been studied. Those attributes that are 1 or 1..* and they are not present in the Spanish data model; the multiplicity has been changed to 0 and 0..*, respectively. For those attributes with a multiplicity of more than one in the Spanish data model but are 1 or 0..1 in the target data model, the multiplicity of these last ones has been changed.
- The attributes that exist in the EPC4EU target data model but are considered unimportant, taking into account the analysis of the ES EPC data model, have been identified.

6.2.1 Proposals for adding or modifying

Studying the target data model and analysing the Spanish data model (checking the complete description of each element and all the information related), a set of proposals were made. These proposals were evaluated to decide if they are incorporated or not into the EPC4EU target data model.

1. Inclusion EPc,nd

Motivation: These values, *EPc,nd*, represent the cooling demand of a building, taking into account the characteristics of the envelope. In the EPC-EU, the value related to the heating (EPh,nd) is present but not that related to the cooling. In some regions, the cooling demand is as important as the heating demand.

Proposed action: EPc,nd could be included at the same level that EPh,nd

2. Inclusion CO_2 emission Label

Motivation: In Spanish EPC, there is a rating for the EPgl (energy performance for the primary energy non-renewable) and CO2 emissions, cooling demand, and heating demand. In the case of cooling and getting demand rating could be related to *EPWinterQuality* and *EPSummerQuality*. However, the rating of the CO_2 emissions is not present.

Proposed action: CO2EmissionLabel could be included at the same level that CO2Emission

3. Inclusion of estimated cost of the recommended intervention

Motivation: In Spanish EPC, there is information about the cost of each measure or intervention, so this information is more direct than the information contained in the *investmentReturnTime* that is processed information

Proposed action: estimatedCost could be included in the *recomendedIntervention complexType*

4. Inclusion nominal efficiency to the energy Conversion Systems

Motivation: In Spanish EPC, the nominal efficiency is associated with the energy conversion systems, not with the energy services. In fact, the most important parameters of the energy systems are the nominal power and the efficiency, parameters that characterise one system.

Proposed action: a *nominalEfficiency* field should be included in the EnergyConversionSystem element at the same level that *nominalPower*.

5. Changing multiplicity in some of the certifier fields and the format of one of them

Motivation: Some of the fields in the certifier part are presented with a multiplicity of 1. These fields are: certifierCategory, certifierDesignationCode and certifierDesignationDescription. The first is present in Spanish but not the rest. Besides, the *certifierCode* is declared as "*int*" format, but this value also has alphabetic values in the Spanish EPCs.

Proposed action: the multiplicity of these elements should be 0..1 (*certifierCategory*, *certifierDesignationCode* and certifierDesignationDescription). Besides, the value of *certifierCode* should be characterString.

6. Changing the name of the Intervention description

Motivation: The values that this parameter can take are very specific and related to the Italian model. In the Spanish data model, there is no information about the intervention, but there is information about the motivation of the certificate procedure

Proposed action: The name of the field should be changed to scope the real meaning of the parameter.

7. Modifying the *dateOfInspection*

Motivation: In Spain, the certifier can do more than one visit for the inspections. So in Spanish, EPCs could be more than one *dateOfInspection*, but in EPC-EU, the *dateOfInspection* multiplicity is set to 0..1. Besides, the Spanish EPCs has a text field about the aim of the visit, related to *dateOfInspection*.

Proposed action: the multiplicity of dateOfInspection should be changed from 0..1 to 0..*. For the second part, a field with the objective of the visit (string) could be added to the dateOnInspection (a complex type should be built)

8. Creation of InterventionType field for gathering all the parameters related to intervention

Motivation: A set of fields in the certificate element related to the intervention could be grouped in one *interventionType*. The fields are:

energyPerformanceAchievable

energyPerformanceImprovement

recommendedIntervention

Proposed action: The values could be shifted to one new category *interventionType* in the certificate element.

9. Shifting of ownershipType from certificate to another type more adequate

Motivation: The ownershipType is a characteristic of the building, not related directly to the certificate

Proposed action: The ownershipType could be shifted to buildingInfoType element or other elements more related to building

10. Changing the cardinality of the nominal Power

Motivation: For each EnergyConversionSystem we have only one nominal power.

Proposed action: The cardinality of the nominalPower could be changed from 0..* to 0..1

11. Inclusion of the EPc,nd,Lim value

Motivation: In the Spanish Certificate not only EPhndLim could be present, but also EPcndLim. In some regions could be more important than EPhndLim.

Proposed action: The EPcndLim could be included at the same level that EPhndLim.

12. Inclusion of more information about demand

Motivation: In Spanish EPCs are present not only the Heating and Cooling values regarding the demand, but also the global one (EPnd,gl) and the value for the DHW.

Proposed action: EPnd,gl and EPdhw,nd could be included at the same level that EPh,nd

13. Inclusion of information of CO2 disaggregated

Motivation: In ES EPC, there is information about CO_2 disaggregated:

CO2 emissions – Heating

CO2 emissions – Cooling

CO2 emissions – DHW

CO2 emissions – Lighting

CO2 emissions - Electric consumption

Proposed action: If it is considered important, some of the disaggregated fields could be included, for example, heating and cooling emissions.

14. Inclusion Label scale

Motivation: In Spanish EPC, the scale used for assigning the Label is available for each label value. The scale is indicated with the maximum limits values of each level (maximum of A value, maximum of B…). Not all the buildings have the same scale. For example, in Spain, the limits of the scale vary considering the climate zone and the use (residential or tertiary).

Proposed action: At least for the main value of the certificate (EPgl), the values of the scale used could be included (using, for example, EPLabelScaleType with energyPerformanceLabelValue used as code list in the "type" field and indicating in the "value" field the maximum limit value for the corresponding level).

15. Changing the name of two parameters: EPWinterQuality and EPSummerQuality

Motivation: In Spanish EPCs there are similar values to these, representing the label for the heating demand and the cooling demand. In the heating case, the *EPWinterQuality* and the Spanish Label for the heating demand represent the same value. In the cooling case, the *EPSummerQuality* could be considered the same Label for the cooling demand.

Proposed action: The names could be changed: *EPWinterQuality* to *EPhndLabel* and *EPSummerQuality* to *EPcndLabel*. The values of the Italian data sources could be assigned as levels also. If the decision is not to change the names, the names from Spanish levels should be included.

6.2.2 Attributes eligible for deletion

The main candidates for deletion are those attributes that are neither present in the Spanish data model nor the Italian data model (they are related to a previous version of the EPC4EU data model developed considering the Trento local EPC data model). After that, other attributes representing very specific concepts of the Italian data model have been identified as selectable for elimination.

Attributes as notes or free texts are candidates to remain in the model. Other attributes considered as important but not present in the Spanish data model are also candidates to remain.

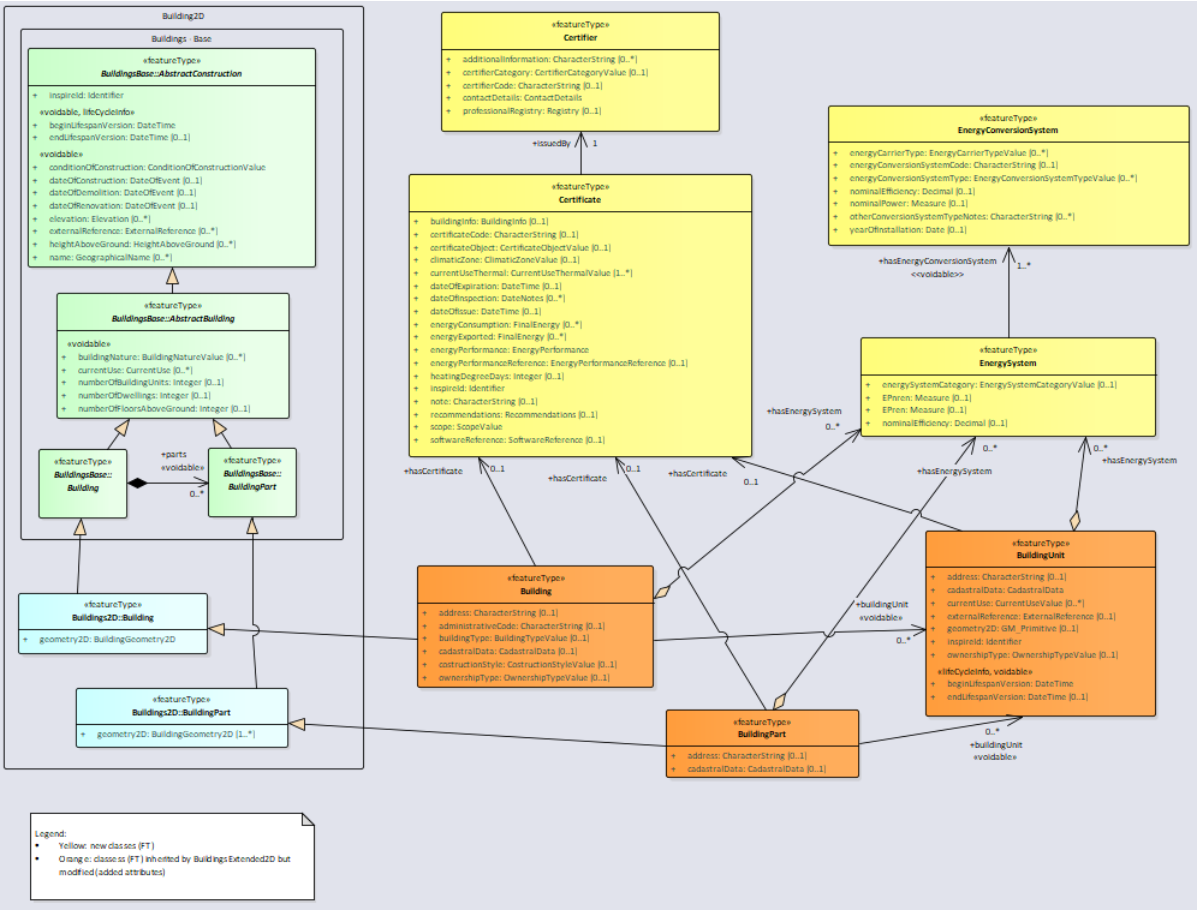
The candidates to be removed from the model:

- *dateOfBuildingPermit*. This value is neither in Spanish nor Italian certificate.
- *dateOfTransmission*. This value is neither in Spanish nor Italian certificate.
- *Profesional*. This value is neither in Spanish nor Italian certificate.
- *energyCarrierNotes*. This value is neither in Spanish nor Italian certificate.
- *AsolRatio* (Summer equivalent solar area per useful surface area). Parameter very specific for Italian data model
- *Yie*. Parameter very specific for Italian data model
- EPReferenceExisting. An epreference is already present
- EPReferenceExistingLabel.

7 EPC4EU target data model (v3.0)

The proposed changes were evaluated, and some of them were assimilated to the new version of the target data model, version v3.0[22]. So, this version can be used to harmonise datasets with different source data models, the Italian and the Spanish data model. A UML representation of the new data model is shown in Figure 4.

Figure 4. EPC4EU target data model, version 3.0.



Source: own creation, JRC, 2020.

22 https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/epc4eu/data-model/3.0/html/

8 EPC4EU target data model (v3.0) and EPC Spanish National data model

After updating the target data model to the new version (v3.0), a new analysis for comparing the EPC Spanish National data model with this new target data model (EPC4EU v3.0) has been made. The matching table between EPC4EU v3.0 and EPC Spanish National data model reflects this comparison, indicating the elements from the ECP4EU data model present in the EPC Spanish National data model and their characteristics (multiplicity, format, description, etc.). The complete version, in the original excel format, is provided as an annex to the report.

9 EPC4EU target data model and Castilla y Leon region EPC database

As explained in section 4, in Spain, there is only one data model at national level, but the management process for the EPCs registration, validation and storage is a competence of each region.

In the context of the project, the analysis has been performed with the region of Castilla y León due to the collaboration with EREN (*Ente Regional de la Energía*), the organism in charge of managing the EPCs registration.

In Castilla y León region, the EPCs are managed following the process described in section 4.2. For the storage of the ECP instances, a regional database is available. This database aims to implement the national data model, but it has peculiarities mainly related to the obligatory nature of some data and also in some of the relations between the data.

Due to the different structure of the source dataset, it has been necessary to perform the matching exercise using the database as the source model. In the next section, this analysis is shown.

The HTML viewer of the UML model of EPC4EU Data Model v3.0 is accessible at https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/epc4eu/data-model/3.0/html/, whilst the *.xsd* is accessible at https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/epc4eu/schemas/3.0/epc4eu.xsd.

9.1 Mapping the EPC database in the EPC4EU target data model

For the mapping of the EPC database, the process is similar to the one explained for mapping the Spanish EPC data model itself, but taking into account that most of the documents have been already studied in depth. So, in this step, the attention was focused on the following documents:

- Documentation of the EPCs database:
- Schemes of the database
- Script for the re-creation of the database (a sample)
- Re-creation of the database itself
- *EPC4EU_MT_v3.0.xlsx*: excel file for the documentation of the mapping
- *epc4eu_v3.0.xsd*: XSD application schema for EPC4EU data model

The results of the mapping were reflected in two documents:

— *EPC4EU_MT_v3.0:* mapping table between the EPC database and the EPC4EU target data model

— *DATABASE_MAPPING.hale*: file in “hale studio” format that contains the project with the mapping.

9.1.1 Differences between Castilla y Leon region EPC data model and the Spanish data model

It is important to highlight that there is information that is more organised in the database than in the Spanish data model (in the database, there are various code lists that can be used). Another important aspect is that the database contains information about the location of the building (coordinates in decimal: latitude and longitude). However, there is information present in the Spanish data model that could be mapped in the target data model, but it is missed in the database, so this information cannot be reflected.

The elements present in the Spanish *XML* but not in the Castilla y Leon region EPC database are indicated here:

— Year of construction

— Date of issue

— Volume

— Certifier (all the information relative to the Certifier)

— Energy consumption. The energy consumption for energy vector is missing in the database extraction

— For the thermal installation, the energy system related to each energy conversion system is missing, so the part of the energy conversion system cannot be mapped.

— Date of inspection

— *EPglren* (a global value is not provided)

— *EPglnrenRifStandard* (only included in the database for the tertiary sector)

— *EPhndLim* (only included in the database for the tertiary sector)

— *EPReferenceNew* (only included in the database for the tertiary sector)

The elements present in the Castilla y Leon region EPC database but not in the Spanish data model are indicated here:

— *Location* (coordinates in decimal degrees: latitude and longitude)

— dateofExpiration

— Software (name and version). The information is more complete than in the *XML* source

— *currenUse* of the building for the tertiary sector has more information

— the information for the energy conversion system type for the heating, cooling and DHW is more organised

Considering those mentioned above, using the data from the database has some problems that should be solved. The problems are:

— *Location*: it is only a Point (latitude, longitude). Besides, the coordinates are not always well placed and sometimes are not pointing to the right building

— Problems with lack of data. Some data that should be present in the database because they are mandatory in the Spanish data model are missing in the extract. The data is missing for different reasons:

  - some of the data are defined as not mandatory in the database, and these data have not been translated from the XML files.
  - Some data are confidential (mainly the data related to the certifier)

To overcome these problems, some strategies have been established to create mapping and posterior harmonisation.

Firstly, for the problem with the location of the building related to the certificate, the way to solve the problems is to use data of the cadaster (compliant with an old draft version of the INSPIRE Buildings data model) to cover the information of the geometry, that is a POLYGON for the building, and it is much more accurate. For this one *JOIN* function in *hale studio* has to be configured.

In the case of the lack of data due to confidentiality issues, a manner to overcome this topic would be to harmonise only public buildings. The information of these buildings should be public so all the information could be harmonised without problems. The confidentiality of the data is a major problem that will arise in each analysis of this type.

9.1.2 Classifications created

In the study, some code lists have been updated, taking into account the values that the parameters of the database could have. In the following tables, it can be seen the different code lists updated and how the values are classified.

Table 11. Code list for EPSummerQuality

Source values		Target value
1	CALIFICACIÓN ENERGÉTICA DE TIPO A	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/A*
2	CALIFICACIÓN ENERGÉTICA DE TIPO B	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/B*
3	CALIFICACIÓN ENERGÉTICA DE TIPO C	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/C*
4	CALIFICACIÓN ENERGÉTICA DE TIPO D	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/D*
5	CALIFICACIÓN ENERGÉTICA DE TIPO E	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/E*
6	CALIFICACIÓN ENERGÉTICA DE TIPO F	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/F*
7	CALIFICACIÓN ENERGÉTICA DE TIPO G	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/G*
15	NO CALIFICABLE	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/nonQualifiable*
16	SIN CALIFICACIÓN ENERGÉTICA	

**Not mapped source values will result in the following target value: null.*

Table 12. Code list for *currentUse*

Source values		Target value
9	COMERCIAL	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/trade*
11	VIVIENDA INDIVIDUAL EN BLOQUE	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/residential*
1	*BLOQUE DE VIVIENDAS COMPLETO*	
3	LOCAL	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/ancillary*
12	VIVIENDA UNIFAMILIAR AISLADA	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/individualResidence*
2	*VIVIENDAS UNIFAMILIARES*	
7	HOTELES Y RESIDENCIAS	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/residenceForCommunities*
13	VIVIENDA UNIFAMILIAR ADOSADA	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/moreThanTwoDwelling*

23	EDIFICIO DE OFICINAS	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/office*
4	*OFICINAS*	
15	ADMINISTRATIVO	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/publicServices*
16	*SANITARIO*	
5	*CENTRO DOCENTE*	
6	*HOSPITALES*	
8	*INSTALACIONES DEPORTIVAS*	
10	*OTROS USOS TERCIARIOS*	
14	VIVIENDA UNIFAMILIAR PAREADA	*http://inspire.ec.europa.eu/codelist/CurrentUseValue/twoDwellings*

*Not mapped source values will result in the following target value: null.

Table 13. Code list for *EPAchievableLabel*

Source values		Target value
1	CALIFICACIÓN ENERGÉTICA DE TIPO A	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/A*
2	CALIFICACIÓN ENERGÉTICA DE TIPO B	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/B*
3	CALIFICACIÓN ENERGÉTICA DE TIPO C	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/C*
4	CALIFICACIÓN ENERGÉTICA DE TIPO D	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/D*
5	CALIFICACIÓN ENERGÉTICA DE TIPO E	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/E*
6	CALIFICACIÓN ENERGÉTICA DE TIPO F	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/F*
7	CALIFICACIÓN ENERGÉTICA DE TIPO G	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/G*
15	NO CALIFICABLE	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/QualityValue/nonQualifiable*
16	SIN CALIFICACIÓN ENERGÉTICA	

Table 14. Code list for *climaticZone*

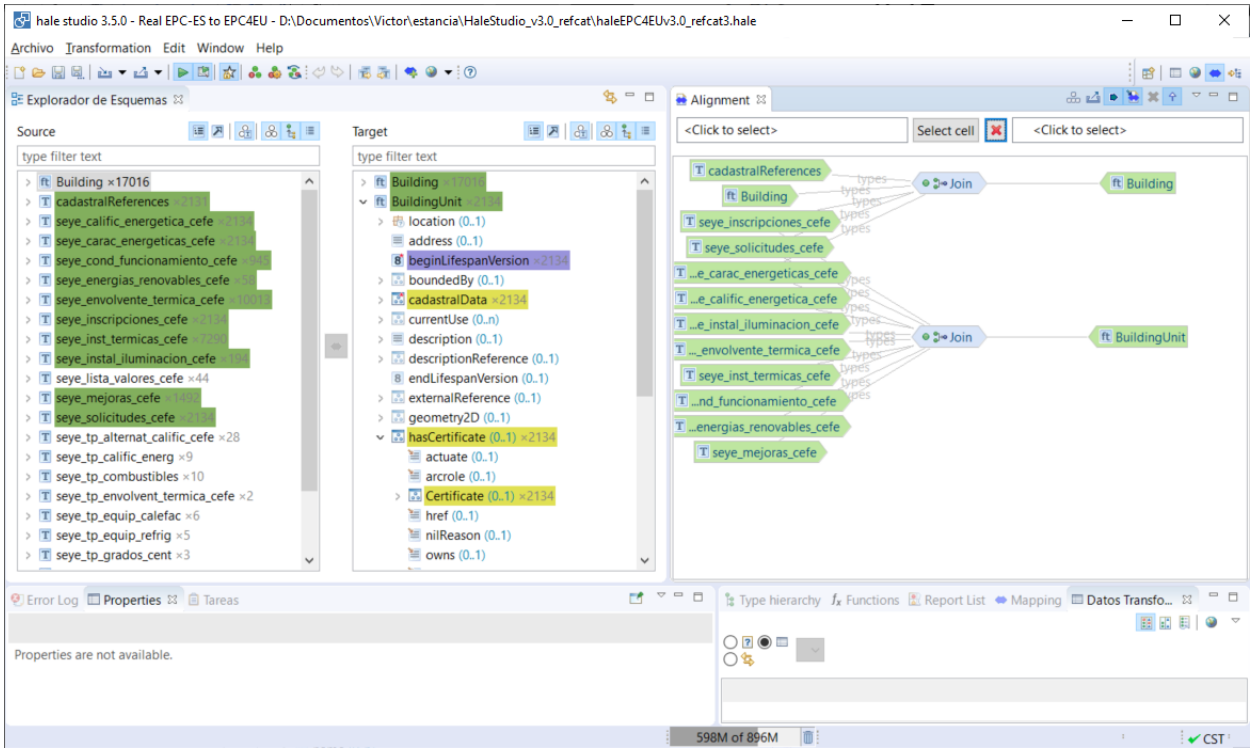
Source values		Target value
1	ZONA CLIMÁTICA D1	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/D1*
2	ZONA CLIMÁTICA D2	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/D2*
3	ZONA CLIMÁTICA E1	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/E1*
4	ZONA CLIMÁTICA A1	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/A1*
5	ZONA CLIMÁTICA A2	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/A2*
6	ZONA CLIMÁTICA A3	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/A3*
7	ZONA CLIMÁTICA A4	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/A4*
8	ZONA CLIMÁTICA B1	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/B1*
9	ZONA CLIMÁTICA B2	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/B2*
10	ZONA CLIMÁTICA B3	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/B3*
11	ZONA CLIMÁTICA B4	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/B4*
12	ZONA CLIMÁTICA C1	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/C1*
13	ZONA CLIMÁTICA C2	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/C2*
14	ZONA CLIMÁTICA C3	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/C3*
15	ZONA CLIMÁTICA C4	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/C4*
16	ZONA CLIMÁTICA D3	*https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/codelist/ClimaticZoneValue/D3*

10 Data transformation

Once the mapping between the Castilla y Leon region EPC data model and the EPC4EU target data model has been performed, the next step is to translate this mapping in transformation rules, using an appropriate tool to create a dataset compliant with the target data model.

For that, the "*hale studio*" tool[23] has been used. This tool allows creating relations between elements from one or several source data models with the target data model elements. Once all the relationships have been performed, the tool can generate the new dataset compliant with the target data model and export it in different formats. Figure 5 shows the tool with the project related to the work done in the task.

Figure 5. Harmonisation project in hale studio: transformation from Castilla y Leon region EPC database to EPC4EU target data model



Source: own creation, CARTIF, 2020.

Therefore, all the mappings were replicated in the tool, and the results can be seen in the following figures.

[23] hale studio form we transform: https://www.wetransform.to/products/halestudio/

Figure 6. Mappings performed in the harmonisation project using hale studio for the building feature type of the target data model

Source: own creation, CARTIF, 2020.

Figure 7. Mappings performed in the harmonisation project using hale studio for the BuildingUnit feature type of the target data model

Source: own creation, CARTIF, 2020.

11 Difficulties encountered

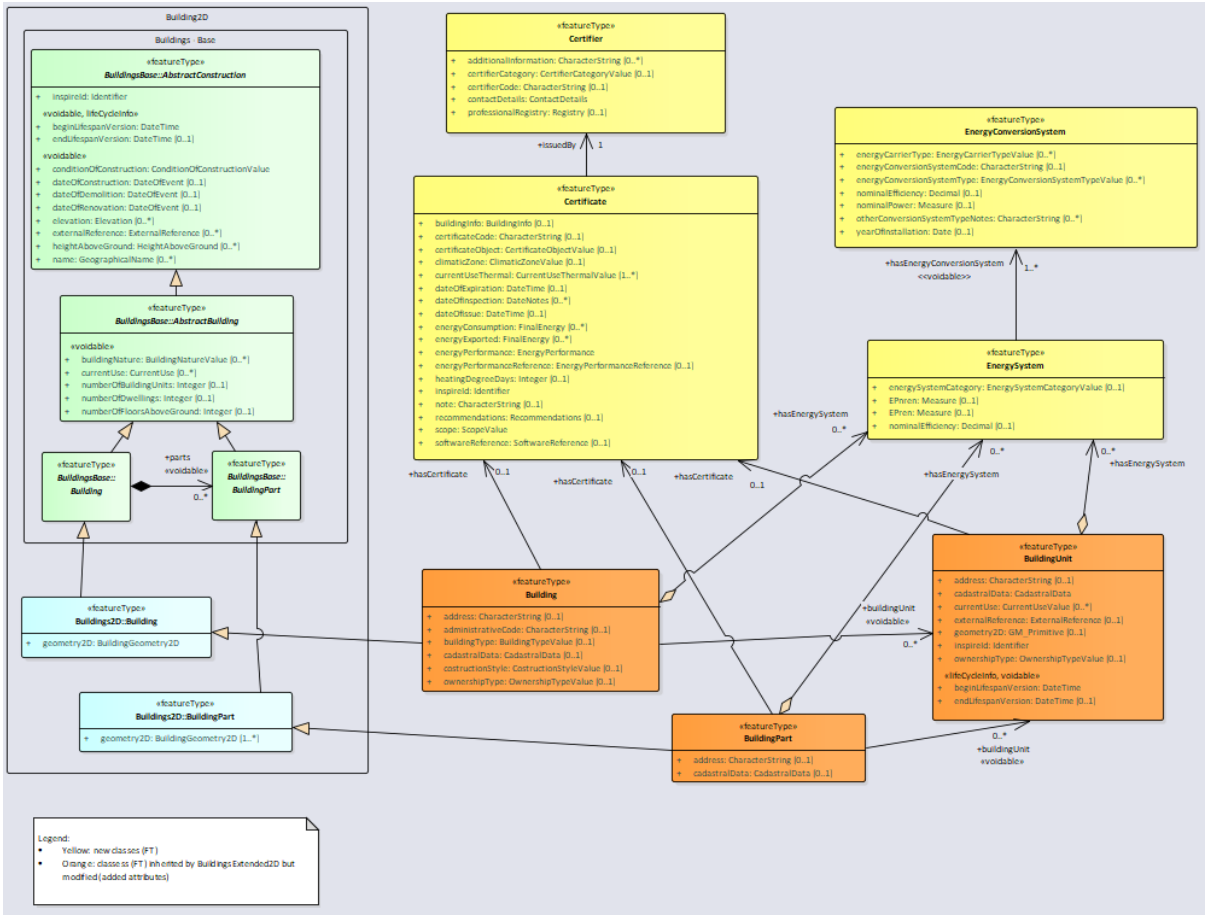
This section has been divided into two main subsections. The first subsection explains the problems encountered when filling the information related to the Building and BuildingPart feature types. The second subsection identifies the problems encountered during the transformation process.

11.1 Completion of building section in EPC4EU (Spanish case)

This section focuses on the analysis of the parts related to the main building information in the EPC4EU data model. The related feature types are *Building*, *BuildingPart*, *BuildingUnit*. In the Spanish case, two methods for completing this information have been studied: collecting this information from the cadaster or collecting information from Spanish EPC.

In the EPC4EU data model, the information of each EPCs is associated with one building (or one building part or one building unit). These Building and BuildingPart feature types are direct extensions of the related feature types defined by the *INSPIRE Building2D data model*. The relationship between these feature types and the Certificate feature type can be seen in the following scheme shown in Figure 8.

Figure 8. EPC Target data model, version 3.0.



Source: own creation based on JRC, 2020.

It is important to note that these building sections include the geolocated information as one *geometry2D* attribute. In the EPC information, the ES EPC data model does not include georeferenced information, so the building is identified by its cadastral reference and by its address. Spanish cadaster contains information about *Building* and *BuildingPart* for each building in Spain, also included the *geometry* element. It has to take into account that the dataset of the Spanish cadaster follows the INSPIRE directive, but not the last version of the

Building scheme, but a previous draft one, so the information does not match entirely with the information present in the building section for the EPC4EU target data model.

Because of this, the first plan was to use the information of the cadaster to fill the information of the building section. The alternative plan would be to use the EPC information directly, but taking into account that there is no explicit information for filling the geometrical part.

The information covered by each one of the approaches is indicated. in Table 15.

Table 15. Comparison between the information that could be covered from the Cadastral source and the EPC source

		Cadastral	**EPC**
Building	**dateOfConstruction - beginning**	beginning	
	dateOfConstruction - end	end	IdentificacionEdificio – AnoConstruccion
	externalReference-informationSystem	informationSystem	
	externalReference-reference	reference	IdentificacionEdificio – ReferenciaCatastral
	name		IdentificacionEdificio – NombreDelEdificio
	currentUse	currentUse	IdentificacionEdificio – TipoDeEdificio
	numberOfDwellings	numberOfDwellings	
	numberOfBuildingUnits	numberOfBuildingUnits	
	numberOfFloorsAboveGround	numberOfFloorsAboveGround	DatosGeneralesyGeometria - NumeroDePlantasSobreRasante
	address		IdentificacionEdificio - Direccion
	geometry	geometry	
	conditionOfConstruction	conditionOfConstruction	
Building Part	**conditionOfConstruction**	conditionOfConstruction	
	numberOfFloorsAboveGround		
	geometry	geometry	
Total elements covered in Building		9	6
Total elements covered in BuildingPart		3	0

In the current case, only the information of the *Building* feature type is relevant because the certificate cannot be related with the *BuildingPart* feature type based on the available information (in the EPC, there is no information about the cadastral reference of the dwelling, also the cadastral reference that identifies the whole building). So from cadastre, we can obtain nine elements and only six from the EPCs. The most important information is covered by both of the approaches, except for two fields:

- The geometry information is missing in the EPCs information
- The information of the address is missing in the cadastre

From the analysis of the above information, it is difficult to select the best option for completing the building section in the data model.

Besides, for the analysis, it is important to highlight that the example mapping has been done with an extract of the database in which the EPCs are stored. The database structure is slightly different from that indicated in the data model, and more information is included in this database (for example, the identifier id of the EPC and the coordinates of the building, crucial information). So, with the information contained in the EPC database, all the important fields of the target data model are covered.

However, some problems have to be solved before finishing the part for the data transformation. The most important problem is that the information about the geometry is not present in the certificates (not in the EPC in XML format, not in the EPCs Database). For this, a method for joining the certificates with the correspondent building in the Cadastre has been used, so the geometry coming from the cadastre can be used. For this purpose, the cadastral reference has been used as unique identifier.

In the harmonisation, the project performed the information of the 17,016 Building presented in the *Building* dataset has been used, although only some of them are finally related with certificates elements.

11.2 Other difficulties

Other problems were identified in the last part of the process. These problems were identified, and solutions were proposed and applied to obtain a final correct solution. In the following paragraphs, the problems encountered and the solutions applied are described.

11.2.1 Problem: Some elements are not correctly mapped

The elements not correctly mapped were the following:

1. *bu-base:referenceGeometry* and *bu-base:horizontalGeometryEstimatedAccuracy* in the Building feature types
2. *inspireId* in the following feature types:
    - Building
    - BuildingUnit
    - Certificate

Box 1. Solution for elements not correctly mapped

1. The mapping relationship for these elements was missing, so the solution was using the “rename” mapping function in hale studio from the Building source.

2. For the three elements:

- *inspireId* in Building can be mapped using rename from Building source

- *inspireId* in BuildingUnit can be created using one of the following identifiers:

1. c_ref_catastral (cadastral reference from database)

2. n_inscription (number of inscription of the EPC)

3. a field with the composition of the two elements discussed above

- *inspireId* in Certificate can be created using the identifier of the Certificate n_inscription (inscription number from database). Concretely, n_inscription is rename to *inspireId.Identifier.localId*

11.2.2 Problem: Incorrect encoding of the values assigned to the *gml:id* of the building units

The problem occurs because values assigned as id cannot start with a number.

Box 2. Solution for incorrect encoding of the values assigned to the gml:id of the building units

The solution is to start with the following pattern: "#ID"

11.2.3 Problem: Some identifiers (*gml:id*) have the same id, and they should be unique

The problem occurs because the EPCs corresponding to the same building have the same cadastral reference (from the different dwellings or new version of the EPCs for the same dwelling, for example).

Box 3. Solution for unique (gml:id) identifiers

The identifier *gml:id* has to be changed. A solution in *hale studio* is to find another attribute in the source dataset that can be used to differentiate the EPCs with the same cadastral reference. This attribute can be used in the "formatted string" function to build a unique gml:id for the BuildingUnit feature types. Probably this parameter is the c_inscripcion_id, but it has to be included in the table used to relate the Building and the BuildingUnit.

11.2.4 Problem: *gml:id* contains no valid characters

The field gml:id contains no valid characters (space, comma, etc. - e.g. *ID_5215611¿¿5151¿0179¿¿_471860005VI6044*). This is because the cadastral_reference value in the database could be only one cadastral reference or a set of cadastral reference separated by commas and spaces (in the Spanish data model, it is valid to have the same certificate for different dwellings at the same time).

Box 4. Solution for gml:id containing no valid characters

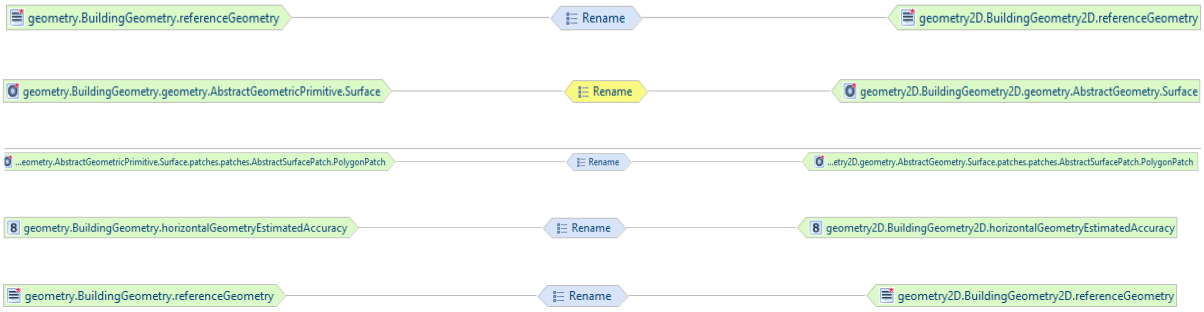
The problem was solved using only the number of inscriptions from the database to identify the building unit. Although the problem is easily solved, it is important to note the indirect problem of having the same certificate for different dwellings.

11.2.5 Problem: Missing geometry for some buildings

In the mapped file, some Geometries present problems, and they are not well mapped. This problem appears in 4745 of 17016 total buildings. It has to be taken into account that the original source (building dataset) does not follow exactly the last version of the Inspire Building data model. So, the inception of the problem could be the source and some problems in the mapping tool. It has to be analysed.

At first, the mappings done is shown in Figure 9.

Figure 9. Relationship performed in *hale studio* for the mapping of the geometry field



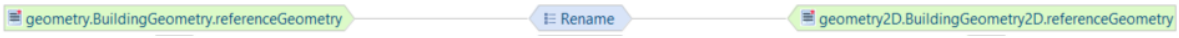
Source: own creation (CARTIF), 2020.

In all the "rename" function, the functional parameter related to *Structural rename* was set to false, i.e., the child properties are not copied in the transformed data.

Box 5. Solution for Missing geometry for some buildings

For solving this issue, it was needed a new remapping of the elements related to the Geometries. So all the mappings related were deleted and the mapping used is a rename from the surface (geometry.BuildingGeometry.geometry.AbstractGeometricPrimitive.Surface) to multiSurface (geometry2D.BuildingGeometry2D.geometry.AbstractGeometry.Multisurface). It is needed to put in "true" the parameter for coping child properties of equal names. With this mapping, the issue related to the geometries disappear.

Figure 10. Relationship performed in *hale studio* for solving the geometry error



Source: own creation (CARTIF), 2020.

11.2.6 Problem: Date of validation not valid

The date of validation contains values not accepted --01-01T00:00:00. The problem is the reconstructed database (from the extract provided by EREN). This database was not corrected recovered because of the format of the date used to reconstruct the database.

Box 6. Solution for Date of validation not valid

The table of the database that contains the date of validation was eliminated and reloaded. In this case, the function to transform the string to date was:

to_date('20/05/24','DD/MM/YY') instead of to_date('20/05/24','DD/MM/RR').

With this change, the file mapped with hale do not present errors in the date of validation

11.2.7 Problem: Date of construction not valid

The date of expiration contains values not accepted (--01-01T00:00:00). The problem is that the database extract has not right values for the value of the date of expiration. It is needed to analyse the values in the dump of the database provided by EREN to check if the problem is in the data source or in the process of recovering this dump.

Box 7. Solution for Date of construction not valid

This is a problem of the sources (Building file from cadastre), so it cannot be corrected. There are multiple solutions to eliminate the element in the mapped file:

1. Replacing with a notepad tool.

2. Other solution is to filter the source attribute in hale studio to exclude wrong dates, using "*add condition context (CQL)*" filter in hale studio with the following text: "*parent.beginning NOT LIKE '--01-01T00:00:00'*."

The problem was finally solved with the second option, more robust.

11.2.8 Problem due to wrong encoding made by hale

The numeric value of some "EPglnrenAchievable" attributes is encoded in a duplicated element in some cases. The value is encoded in one element, and the Unit is encoded in another element. It is interesting to highlight that this behaviour does not occur in all the EPglnrenAchievable instances, only in some of them.

Box 8. Solution for wrong encoding made by hale

A semi-automatic solution has been implemented. First, it has to be eliminated the mapping from the *uom* of the *EPglnrenAchievable*. After transforming the data, it is needed to "replace" in notepad with the following information: <epc4eu:EPglnrenAchievable> Replace with: <epc4eu:EPglnrenAchievable uom="kWh/m^2">

12 Results of the data transformation

Once all the mappings were performed and all the problems solved, the next step was the generation of the transformed dataset. In this case, the transformed data were exported in a GML format (GML as FeatureCollection option).

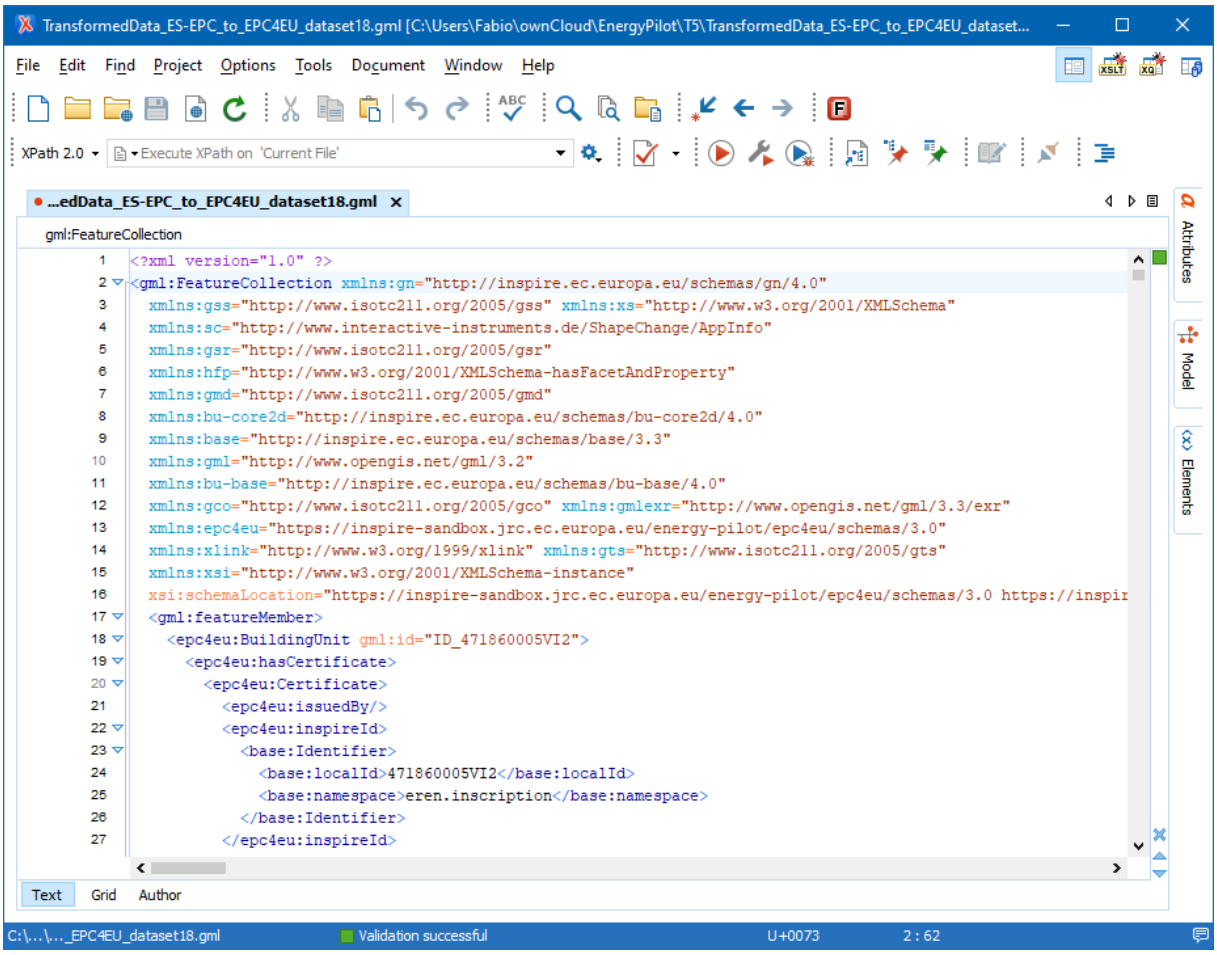
An extract of the GML transformed file is shown in Figure 11, whilst a screenshot of its successful validation is shown in Figure 12.

Figure 11. Results of the data transformation, in notepad++.

```
<gml:featureMember>
  <epc4eu:BuildingUnit gml:id="ID_471860005VI2">
    <epc4eu:hasCertificate>
      <epc4eu:Certificate>
        <epc4eu:issuedBy/>
        <epc4eu:inspireId>
          <base:Identifier>
            <base:localId>471860005VI2</base:localId>
            <base:namespace>eren.inscription</base:namespace>
          </base:Identifier>
        </epc4eu:inspireId>
        <epc4eu:certificateObject xlink:href="https://inspire-sandbox.jrc.ec.europa.eu/codelist/CertificateObjectValue/realEstateUnit"></epc4eu:certificateObject>
        <epc4eu:buildingInfo>
          <epc4eu:BuildingInfo>
            <epc4eu:surface>
              <epc4eu:SurfaceType>
                <epc4eu:type xlink:href="https://inspire-sandbox.jrc.ec.europa.eu/codelist/SurfaceTypeValue/conditionedSurface"></epc4eu:type>
                <epc4eu:value uom="m²">77.0</epc4eu:value>
              </epc4eu:SurfaceType>
            </epc4eu:surface>
            <epc4eu:surfaceVolumeRatio>2.5</epc4eu:surfaceVolumeRatio>
            <epc4eu:EPhnd uom="kWh/m²">145.35</epc4eu:EPhnd>
            <epc4eu:EPcnd uom="kWh/m²">16.27</epc4eu:EPcnd>
          </epc4eu:BuildingInfo>
        </epc4eu:buildingInfo>
        <epc4eu:climaticZone xlink:href="https://inspire-sandbox.jrc.ec.europa.eu/codelist/ClimaticZoneValue/D2"></epc4eu:climaticZone>
        <epc4eu:currentUseThermal/>
        <epc4eu:dateOfExpiration>2023-07-08T22:00:00Z</epc4eu:dateOfExpiration>
        <epc4eu:energyPerformance>
          <epc4eu:EnergyPerformance>
            <epc4eu:EPLabel xlink:href="https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/E"></epc4eu:EPLabel>
            <epc4eu:EPglnren uom="kWh/m²">307.07</epc4eu:EPglnren>
            <epc4eu:CO2Emission uom="KgCO2/m²*year">63.19</epc4eu:CO2Emission>
            <epc4eu:CO2EmissionLabel xlink:href="https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/E"></epc4eu:CO2EmissionLabel>
          </epc4eu:EnergyPerformance>
        </epc4eu:energyPerformance>
        <epc4eu:energyPerformanceReference>
          <epc4eu:EnergyPerformanceReference>
            <epc4eu:EPSummerQuality xlink:href="https://inspire-sandbox.jrc.ec.europa.eu/codelist/QualityValue/G"></epc4eu:EPSummerQuality>
            <epc4eu:EPWinterQuality xlink:href="https://inspire-sandbox.jrc.ec.europa.eu/codelist/QualityValue/E"></epc4eu:EPWinterQuality>
          </epc4eu:EnergyPerformanceReference>
        </epc4eu:energyPerformanceReference>
        <epc4eu:scope/>
        <epc4eu:softwareReference>
          <epc4eu:SoftwareReference>
            <epc4eu:name>CE3X</epc4eu:name>
          </epc4eu:SoftwareReference>
        </epc4eu:softwareReference>
      </epc4eu:Certificate>
    </epc4eu:hasCertificate>

<gml:featureMember>
  <epc4eu:Building gml:id="ES.SDGC.BU.9454510UM5195C">
    <bu-base:beginLifespanVersion>2020-05-08T05:59:38Z</bu-base:beginLifespanVersion>
    <bu-base:conditionOfConstruction xsi:nil="true"/>
    <bu-base:dateOfConstruction>
      <bu-base:DateOfEvent>
        <bu-base:beginning>1947-12-31T23:00:00Z</bu-base:beginning>
        <bu-base:end>1953-12-31T23:00:00Z</bu-base:end>
      </bu-base:DateOfEvent>
    </bu-base:dateOfConstruction>
    <bu-base:externalReference>
      <bu-base:ExternalReference>
        <bu-base:informationSystem>https://www1.sedecatastro.gob.es/CYCBienInmueble/OVCListaBienes.aspx?rc1=9454510&amp;rc2=UM5195C</bu-base:informationSystem>
        <bu-base:informationSystemName/>
        <bu-base:reference>9454510UM5195C</bu-base:reference>
      </bu-base:ExternalReference>
    </bu-base:externalReference>
    <bu-base:inspireId>
      <base:Identifier>
        <base:localId>9454510UM5195C</base:localId>
        <base:namespace>ES.SDGC.BU</base:namespace>
      </base:Identifier>
    </bu-base:inspireId>
    <bu-base:numberOfDwellings>0</bu-base:numberOfDwellings>
    <bu-base:numberOfBuildingUnits>1</bu-base:numberOfBuildingUnits>
    <bu-core2d:geometry2D>
      <bu-base:BuildingGeometry2D>
        <bu-base:geometry>
          <gml:MultiSurface srsName="urn:ogc:def:crs:EPSG::25830" srsDimension="2">
            <gml:surfaceMember>
              <gml:Polygon srsName="urn:ogc:def:crs:EPSG::25830" srsDimension="2">
                <gml:exterior>
                  <gml:LinearRing>
                    <gml:posList>359195.5435 4615203.141 359110.9695 4615045.721 359109.6495 4615043.271 359082.2185 4614992.1915 359082.1985 4614992.1515 35908
                    4614986.9015 359093.649 4614984.8515 359103.529 4614979.6815 359104.059 4614980.6715 359105.429 4614983.2215 359135.73 4615039.521 359137.75
                    359199.863 4615158.631 359217.5235 4615191.4305 359232.684 4615219.5905 359232.814 4615219.8405 359270.976 4615290.7205 359272.356 4615293.2
                    4615341.2805 359326.9885 4615394.57 359305.478 4615405.8405 359245.5255 4615296.161 359245.2555 4615295.661 359244.9855 4615295.161 359243.5
                  </gml:LinearRing>
                </gml:exterior>
              </gml:Polygon>
            </gml:surfaceMember>
          </gml:MultiSurface>
        </bu-base:geometry>
        <bu-base:referenceGeometry>true</bu-base:referenceGeometry>
        <bu-base:horizontalGeometryReference/>
        <bu-base:horizontalGeometryEstimatedAccuracy uom="m">0.1</bu-base:horizontalGeometryEstimatedAccuracy>
      </bu-base:BuildingGeometry2D>
    </bu-core2d:geometry2D>
  </epc4eu:Building>
</gml:featureMember>
```

Source: own creation (CARTIF), 2020.

Figure 12. Screenshot of the validation of the gml file with Oxygen.



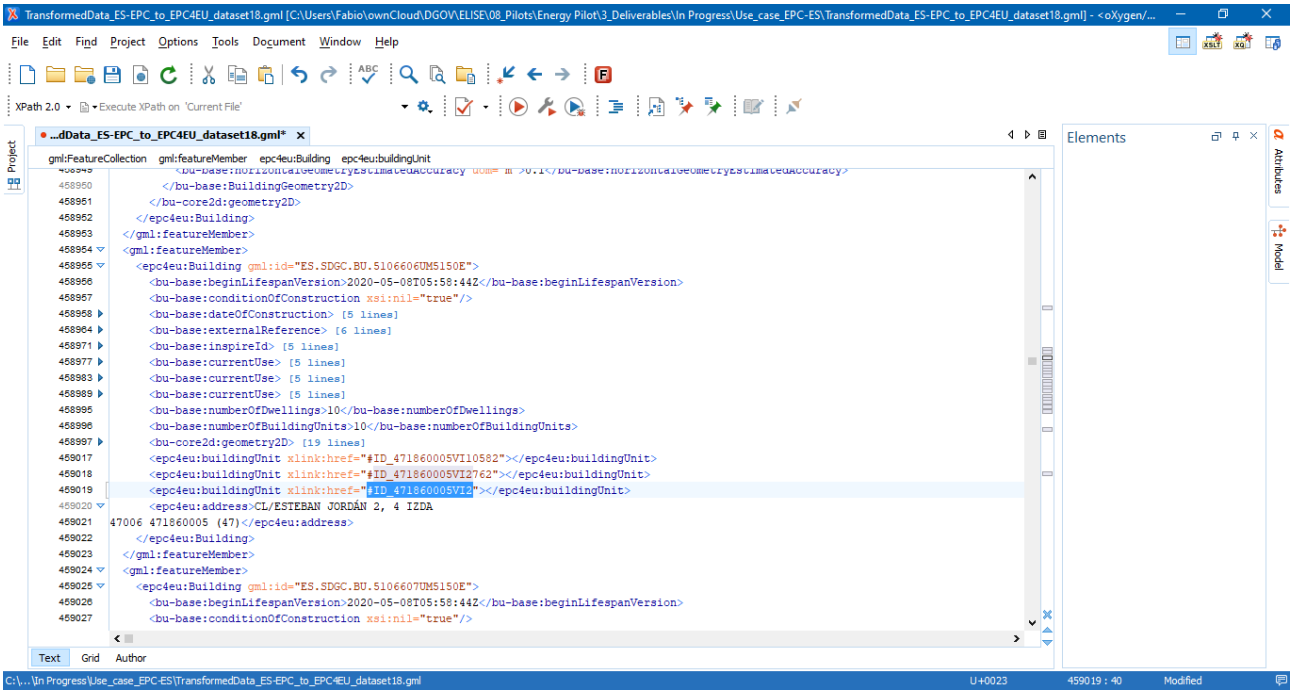
Source: own creation (JRC), 2020.

Regarding the visualization of this GML file in QGIS, it is necessary to establish some environment variables and generate a process for creating required functions. In the following sections, the steps needed are explained.

12.1 How to thematise the EPC4EU GML dataset in QGIS 3

In case of the association between the feature types "*Building*" and "*BuldingUnits*" used in the GML file, all the attributes related to the building units associated with the building are not visible using the default import functionalities of the building units QGIS.

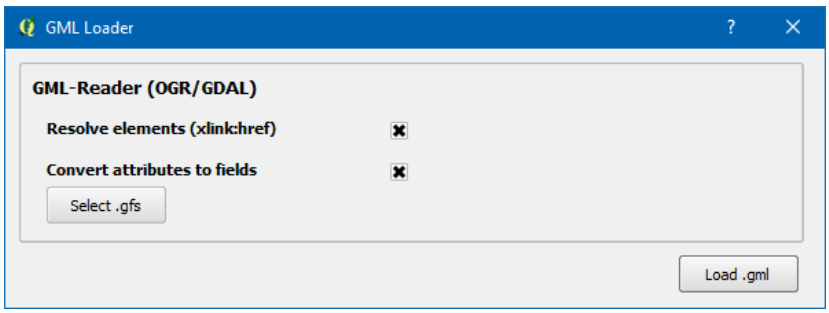
Figure 13. Association between “Building” and “BuldingUnits”



Source: own creation (CARTIF), 2020.

A solution is the use of the plug-in “GML Loader” in QGIS 2.18

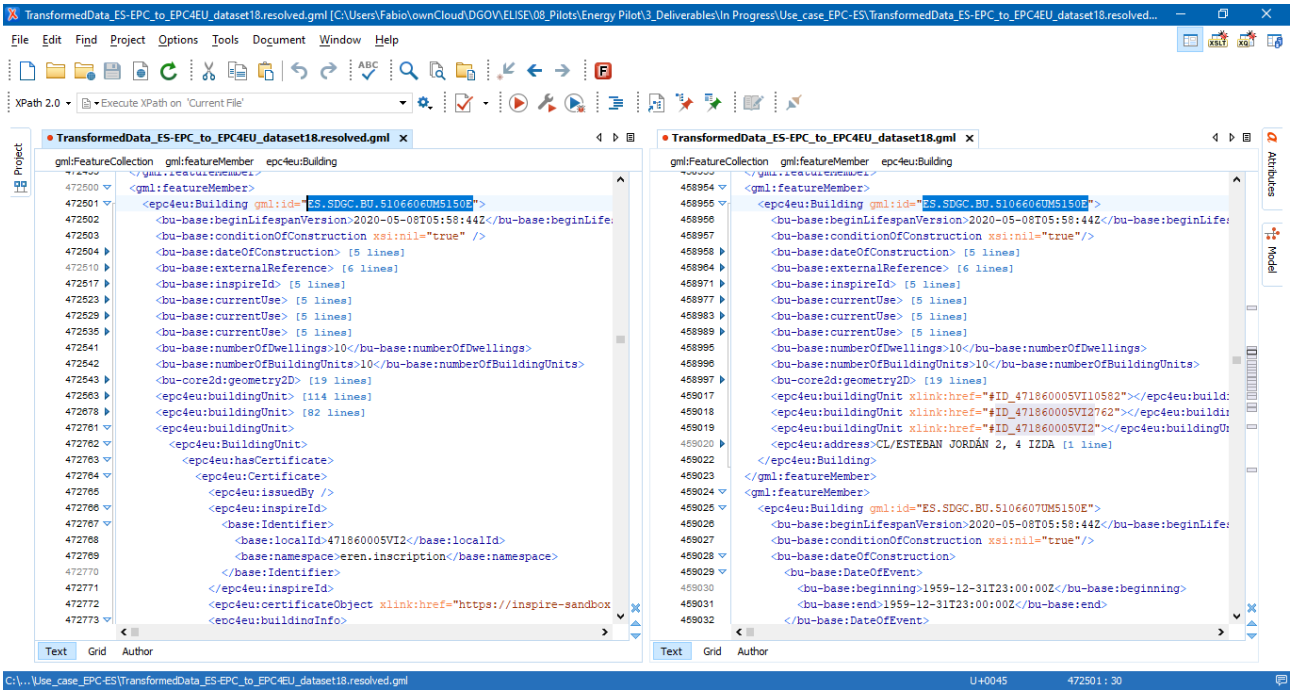
Figure 14. GML Loader plug-in available only on old QGIS version



Source: own creation (CARTIF), 2020.

Loading the GML file through the plug-in creates a new associated GML file named “**.resolved.gml*” in which all internal *xlink:href* (like #ID) are resolved. All the elements of the linked feature types (BuildingUnits) are embedded in the main feature type (Building).

Figure 15. Differences between the two GML files



Source: own creation (CARTIF), 2020.

The "GML Loader" plug-in is not available, at the moment, for version 3 of QGIS, but the same function can be implemented using an environment variable[24]:

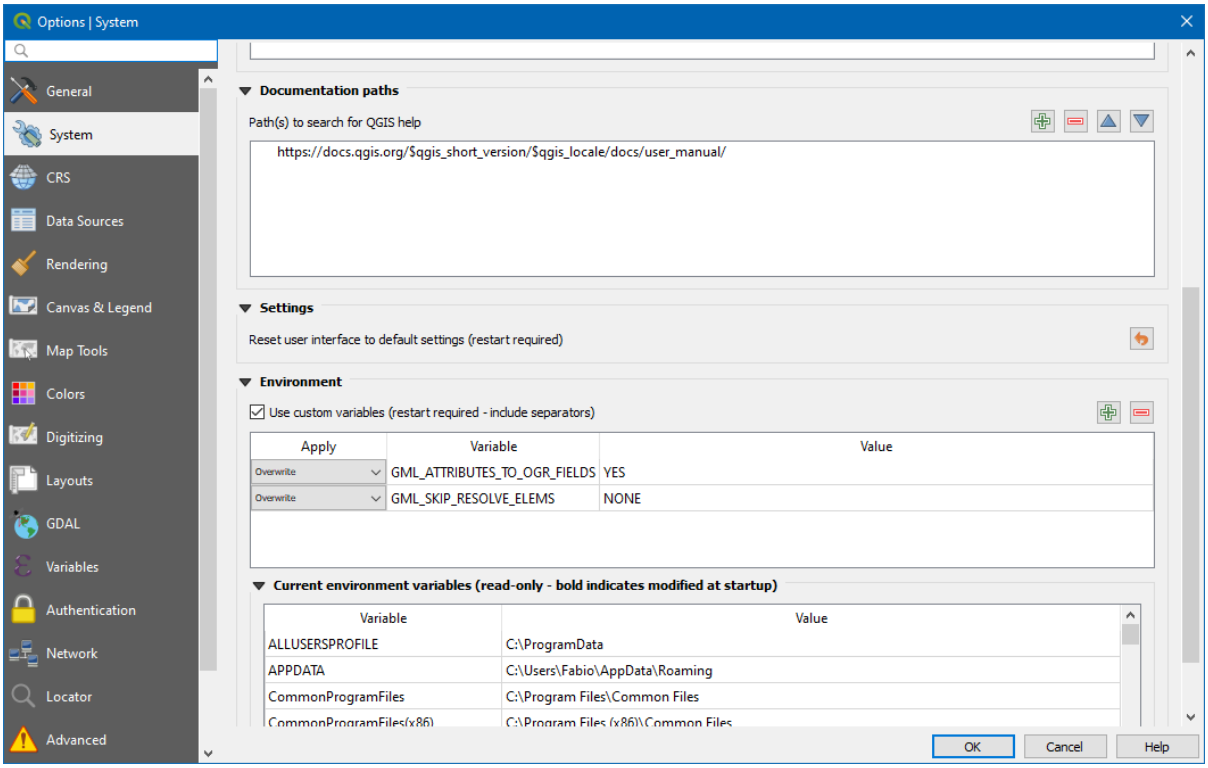
GML_SKIP_RESOLVE_ELEMS = NONE

To visualise the complete GML structure, the following steps can be followed:

1. Add the environmental variables:
    a. "**GML_ATTRIBUTES_TO_OGR_FIELDS"** with the value "**YES**". This allows QGIS to resolve the GML elements containing *xlink:href* attribute.
    b. "**GML_SKIP_RESOLVE_ELEMS**" with the value "**NONE**". This allows QGIS to resolve also the *xlink:href* pointing to an element with the related *gml:id* in the same GML file.

24 https://gdal.org/drivers/vector/gml.html#gml-xlink-resolving

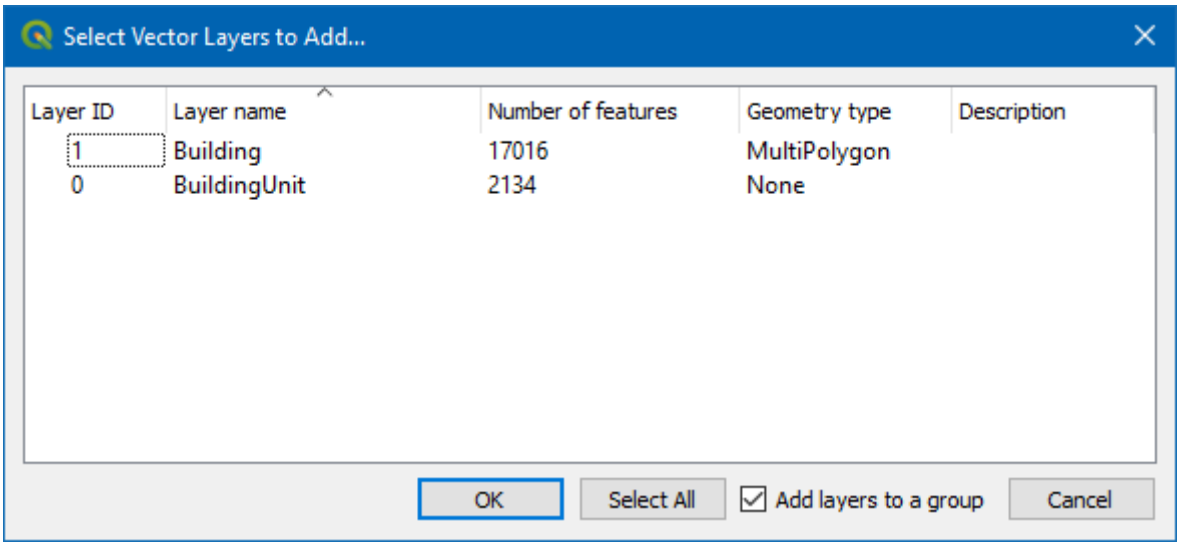
Figure 16. Setting of the QGIS environmental variables



Source: own creation (CARTIF), 2020.

2. Load the GML file (by drag-and-drop or through the related menu). QGIS will ask you to select the layers to add if there are more than one. In this case, select all.

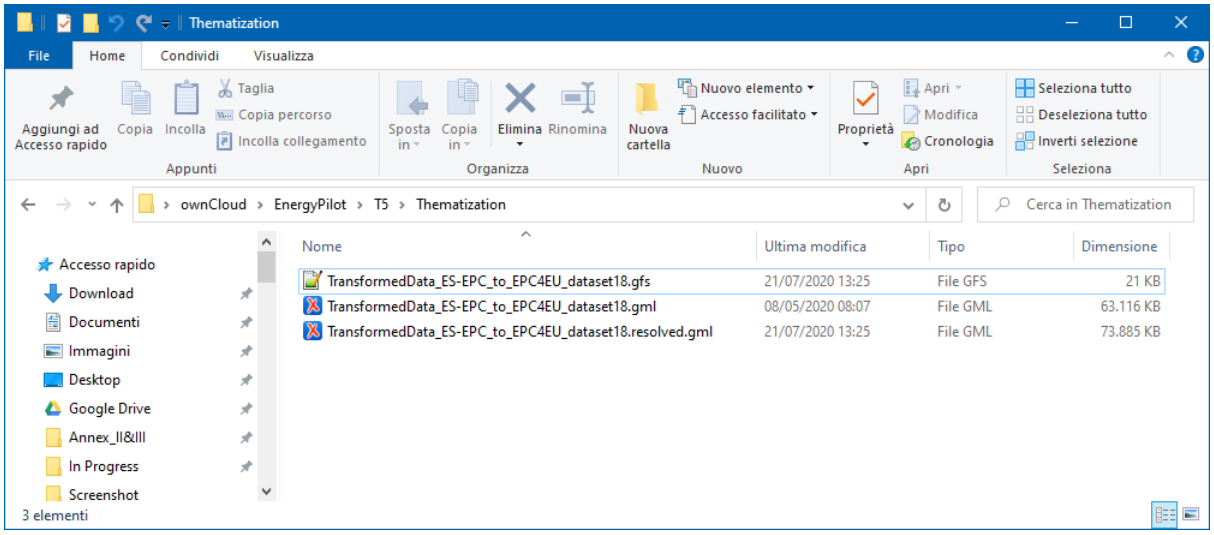
Figure 17. Loading the GML file



Source: own creation (CARTIF), 2020.

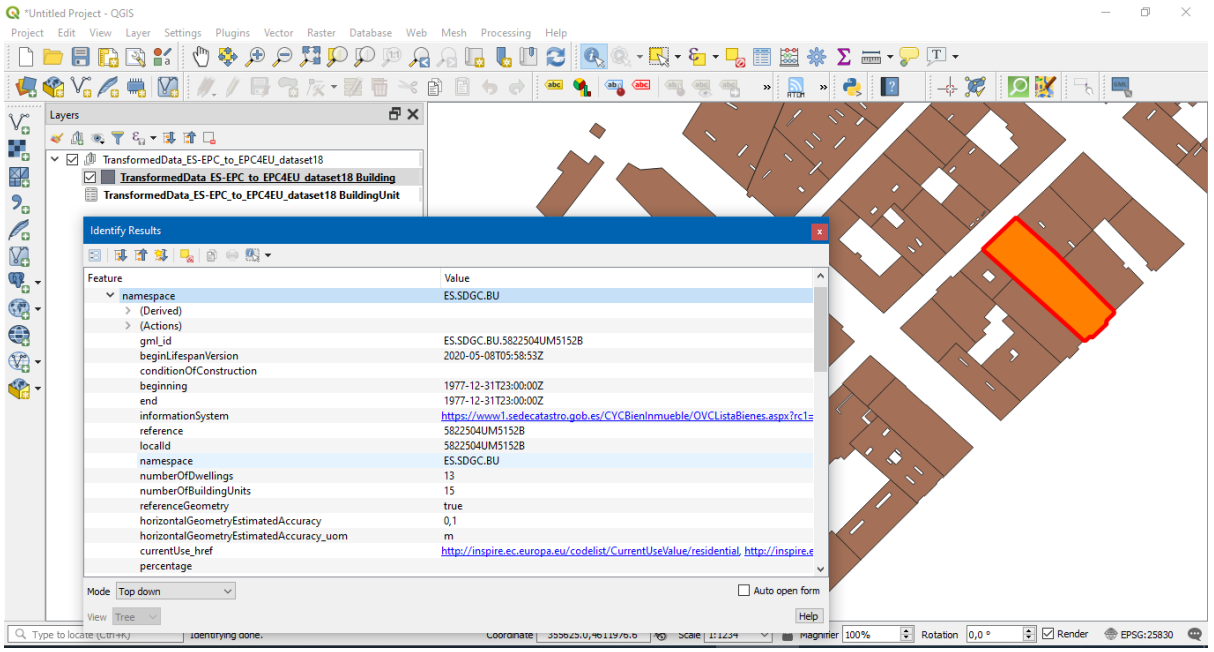
3. QGIS will create the associated GML file (the resolved one) and the related gfs file containing the map between the GML attributes (including those related to the associated feature types) and the elements of the attribute table.

Figure 18. Files created by QGIS during the import of the GML file



Source: own creation (CARTIF), 2020.

Figure 19. QGIS feature attributes



Source: own creation (CARTIF), 2020.

*it is important to note that this process will only work for the last versions of the QGIS 3, so it is necessary to update the application to the newest one, if possible.

12.2 How to visualise labels for each building

For this purpose, the best way is to create a function that allows manipulating the data of the different fields.

Taking into account that the label of interest is ***EPLabel_href*** and for some of the buildings there are multiple labels (array), the function we have to use is:

```
array_to_string("EPLabel_href")
```

Besides, new functions should be created:

— *extractBestLabel*: if the building has one only EPC, the colour should reflect the label (***EPLabel)*** of this EPC. If there is more than one EPC the colour should reflect the EPLabel more efficient. Buildings with no EPCs are marked in grey

— *extractWorstLabel*: if the building has one only EPCs, the colour should reflect the label (***EPLabel)*** of this EPC. If there is more than one EPC the colour should reflect the EPLabel less efficient. Buildings with no EPCs are marked in grey

— *extractLabel*: if the building has one only EPCs, the colour should reflect the label (***EPLabel)*** of this EPC. If there is more than one EPC the building should be marked in white with black lines. Buildings with no EPCs are marked in grey

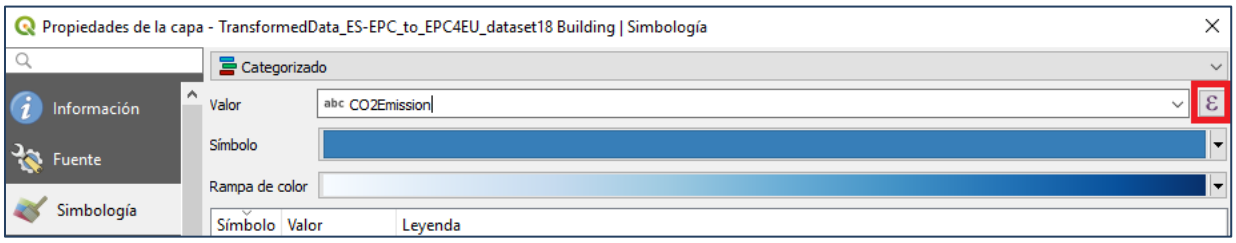
The following section explains how to create these functions and how to load them in QGIS.

12.2.1 Creation of the functions to visualise labels

The full process for the creation of the functions to visualise label is the following:

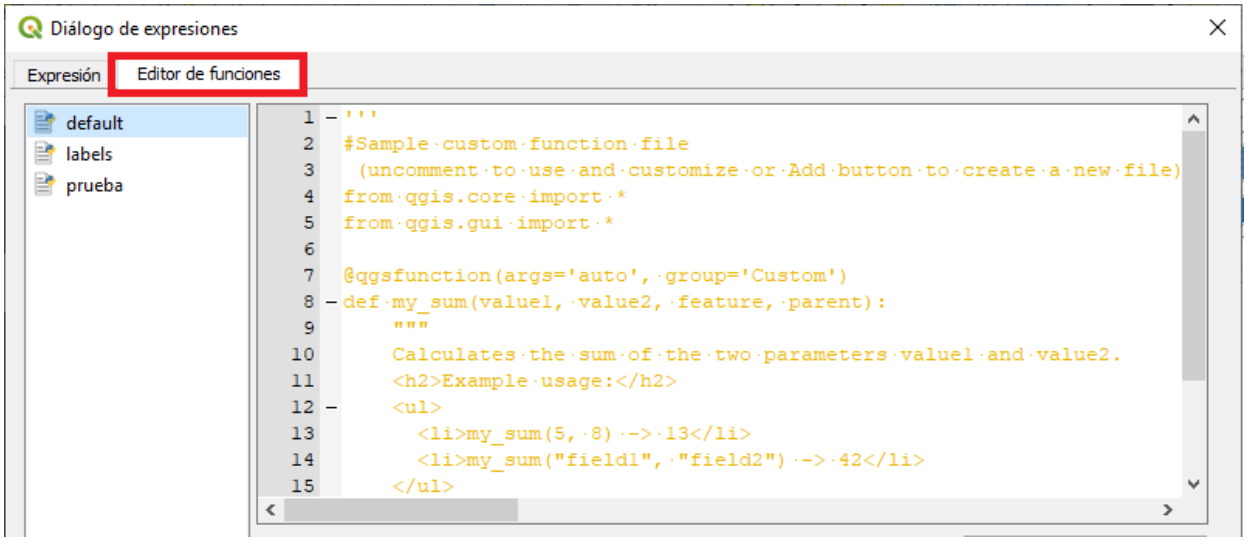
1. Open the function editor:

Figure 20. QGIS properties of the layer



Source: own creation (CARTIF), 2020.

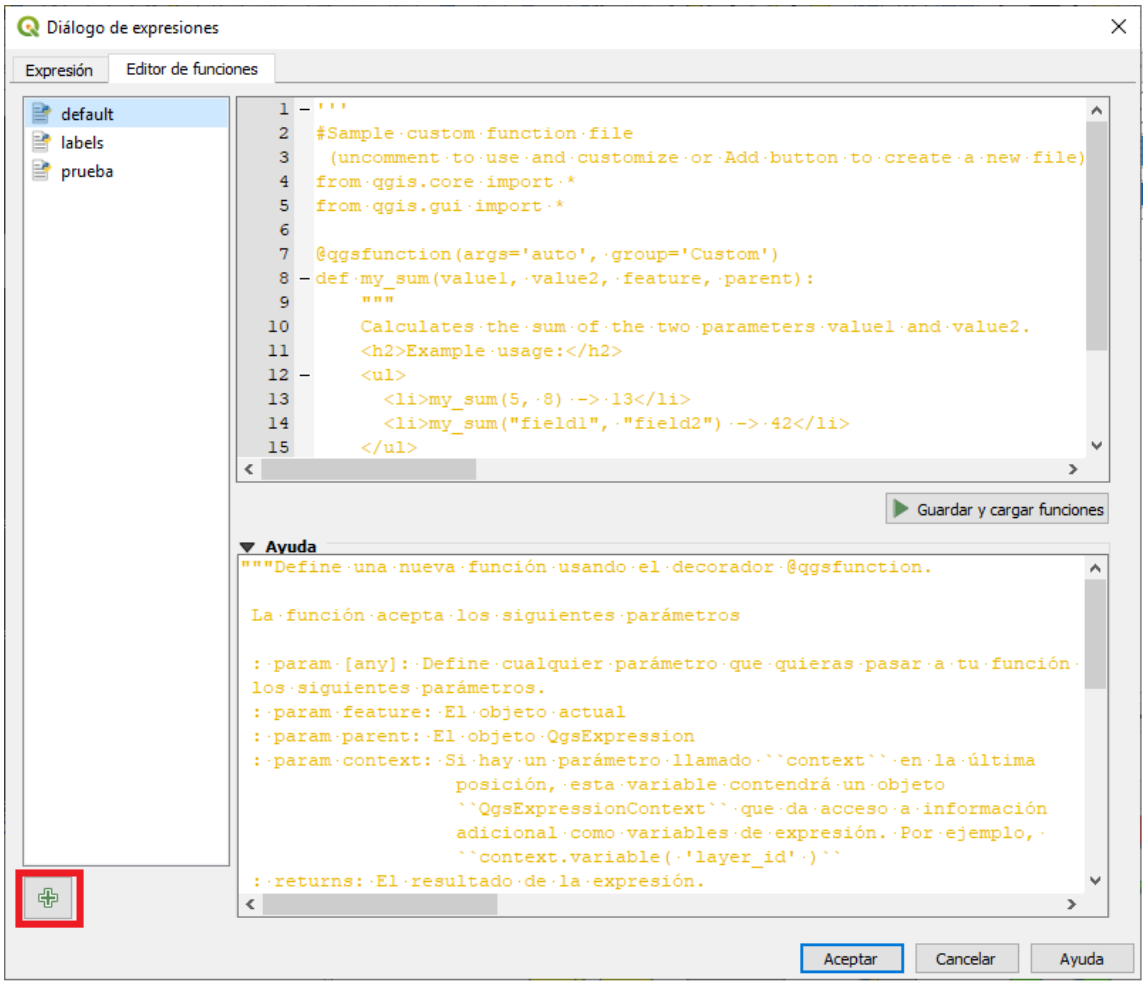
Figure 21. QGIS expressions dialog menu: functions editor



Source: own creation (CARTIF), 2020.

2. Create new function

Figure 22. QGIS functions editor: create new function



Source: own creation (CARTIF), 2020.

3. Write the functions in the box:

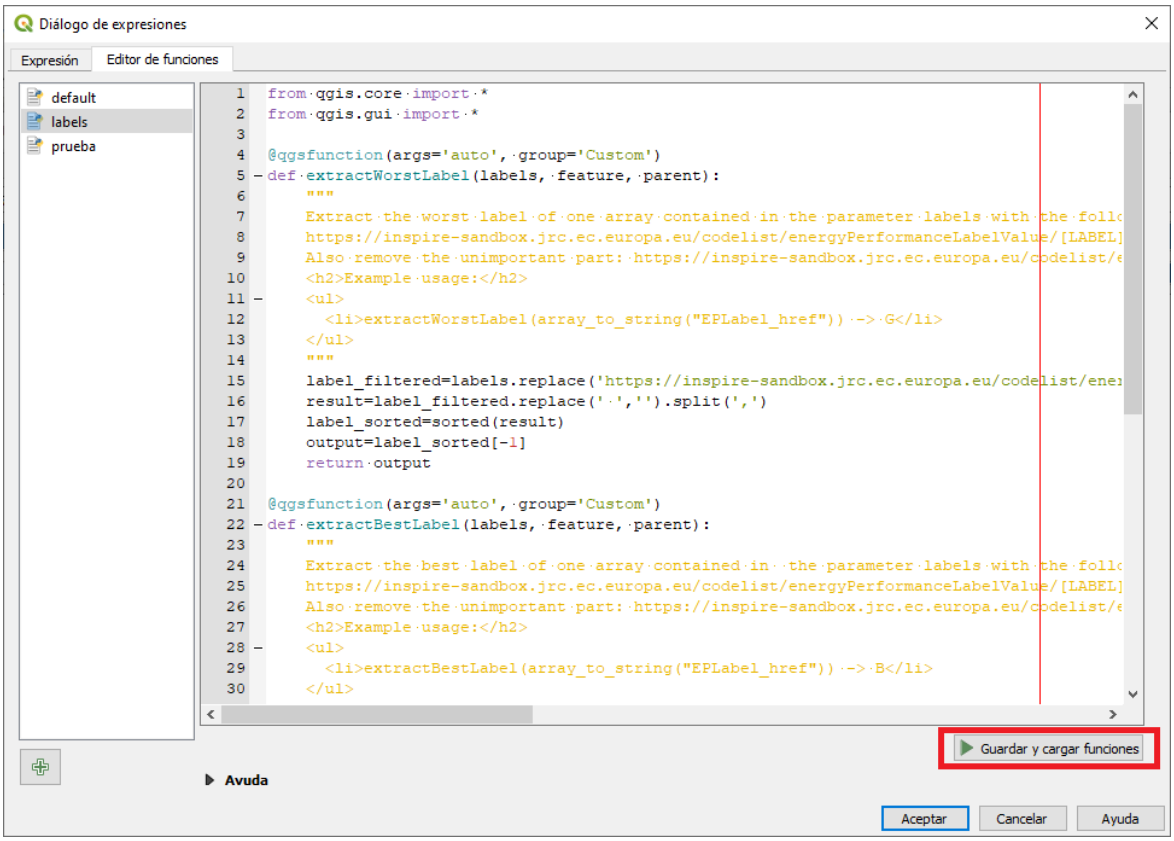
```
from qgis.core import *

from qgis.gui import *


@qgsfunction(args='auto', group='Custom')

def extractWorstLabel(labels, feature, parent):

    """

    Extract the worst label of one array contained in the parameter labels with the following format:

    https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/[LABEL].

    Also remove the unimportant part: https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/

    <h2>Example usage:</h2>

    <ul>

      <li>extractWorstLabel(array_to_string("EPLabel_href")) -> G</li>
```

```
    </ul>
    """
    label_filtered=labels.replace('https://inspire-
sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/', '')
    result=label_filtered.replace(' ','').split(',')
    label_sorted=sorted(result)
    output=label_sorted[-1]
    return output


@qgsfunction(args='auto', group='Custom')
def extractBestLabel(labels, feature, parent):
    """
    Extract the best label of one array contained in  the parameter labels with the following
format:
    https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/[LABEL].
    Also         remove         the         unimportant         part:         https://inspire-
sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/
    <h2>Example usage:</h2>
    <ul>
      <li>extractBestLabel(array_to_string("EPLabel_href")) -> B</li>
    </ul>
    """
    label_filtered=labels.replace('https://inspire-
sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/', '')
    result=label_filtered.replace(' ','').split(',')
    label_sorted=sorted(result)
    output=label_sorted[0]
    return output


@qgsfunction(args='auto', group='Custom')
def extractLabel(labels, feature, parent):
    """
    Extract the label of one array contained in the parameter labels with the following format:
    https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/[LABEL].
    Also       remove       the       unimportant       part:              https://inspire-
sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/
    For buildings with multiple labels the output is "Multiple"
    <h2>Example usage:</h2>
    <ul>
      <li>extractLabel(array_to_string("EPLabel_href")) -> B</li>
    </ul>
    """
    label_filtered=labels.replace('https://inspire-
sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/', '')
    output="None"
```

```
if len(label_filtered)==1:
    output=label_filtered
elif len(label_filtered)>1:
    output="Multiple"
return output
```

4. Save the new functions

Figure 23. QGIS functions editor: save and load function



Source: own creation (CARTIF), 2020.

5. Write the function in the "value" section in the symbology properties.
   - For the case of the best label the right function is

```
extractBestLabel(array_to_string("EPLabel_href"))
```

   - For the case of the worst label the right function is

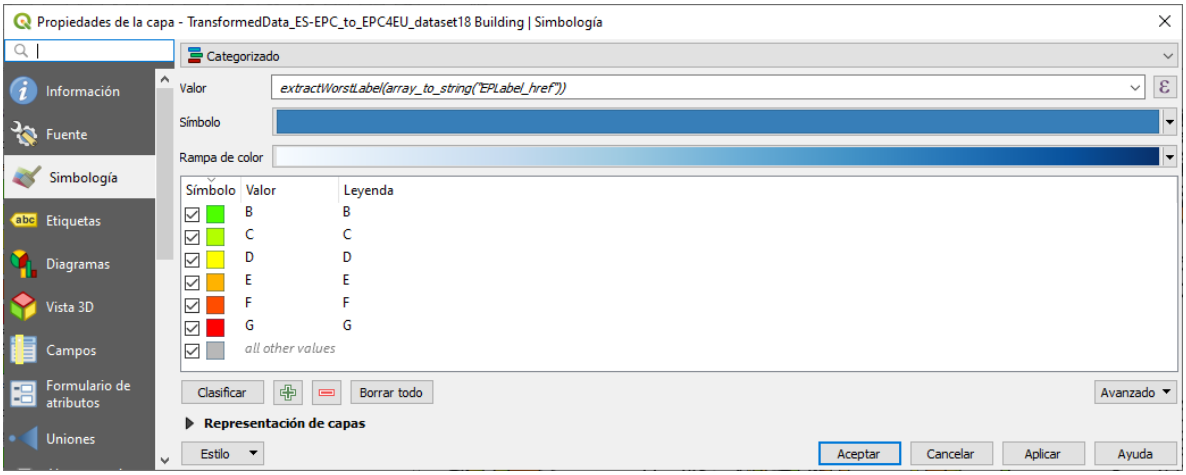
```
ExtractWorstLabel(array_to_string("EPLabel_href"))
```

   - For the case of the label indicating the buildings with the multiples EPCs the right function is

```
extractWorstLabel(array_to_string("EPLabel_href"))
```

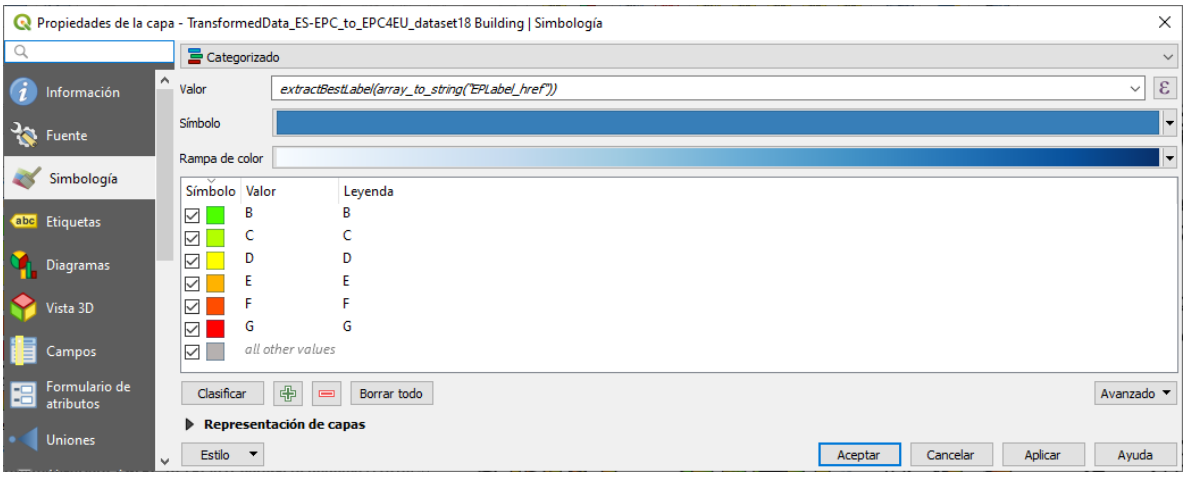
In the following pictures it can be seen how the function has to be introduced in the symbology box.

Figure 24. QGIS properties for extract worst label



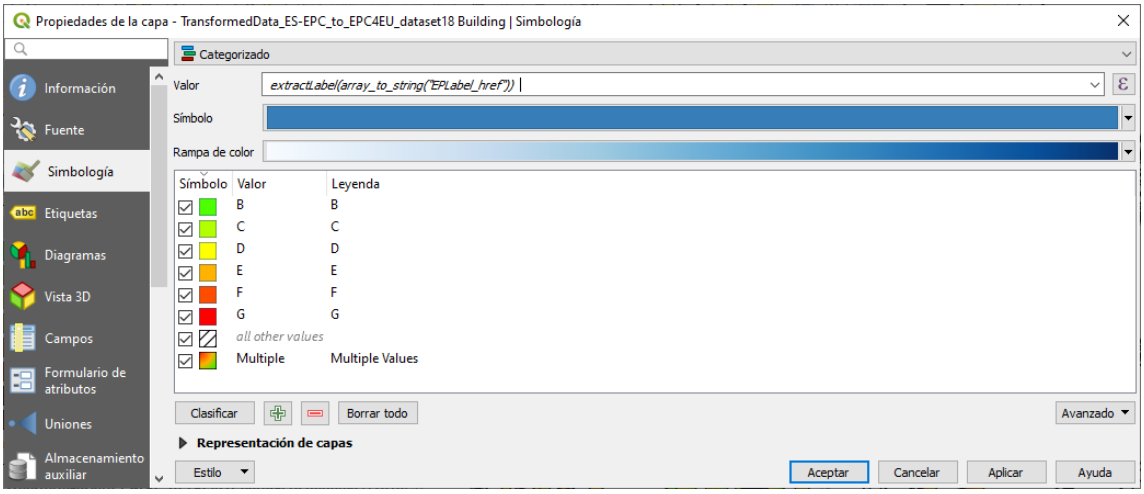
Source: own creation (CARTIF), 2020.

Figure 25. QGIS properties for extract best label



Source: own creation (CARTIF), 2020.

Figure 26. QGIS properties for extract multiple values label



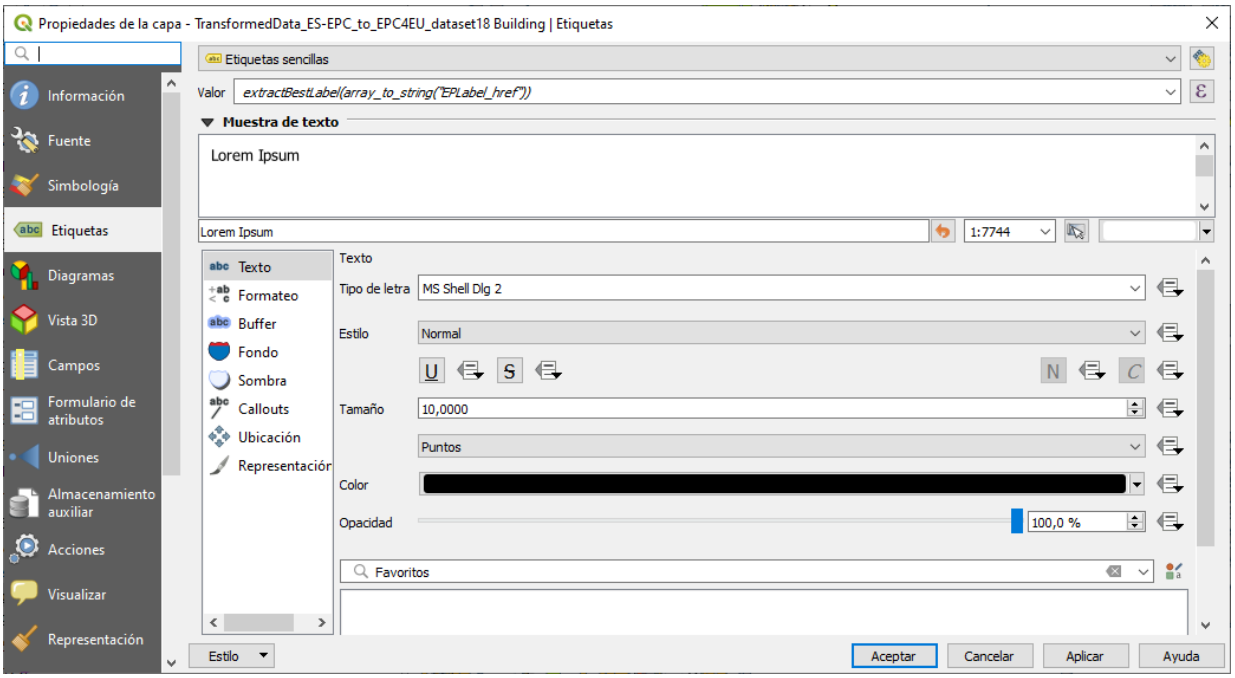
Source: own creation (CARTIF), 2020.

Besides it is needed to stablish the colour of the different labels. The values of the colour for the different labels in RGB are:

- A: #00ff00
- B: #4dff00
- C: #b3ff00
- D: #ffff00
- E: #ffb300
- F: #ff4d00
- G: #ff0000

Another characteristic that could help to allow visualising the tags for each dataset is to put the same function in the corresponding "tags" properties, as can be seen in Figure 27:

Figure 27. QGIS properties for "extract multiple values label" for the visualization of the tags with the label



Source: own creation (CARTIF), 2020.

12.2.2 Loading the function and the symbology

Instead of creating the functions in QGIS and generating the symbology, it is possible to storing the file directly in the corresponding directory and using them later and loading the symbology style.

So, the following text should be store in a python file (named as ***label.py***) in the directory of expressions in the configuration profile of QGIS. More likely, the directory would be here:

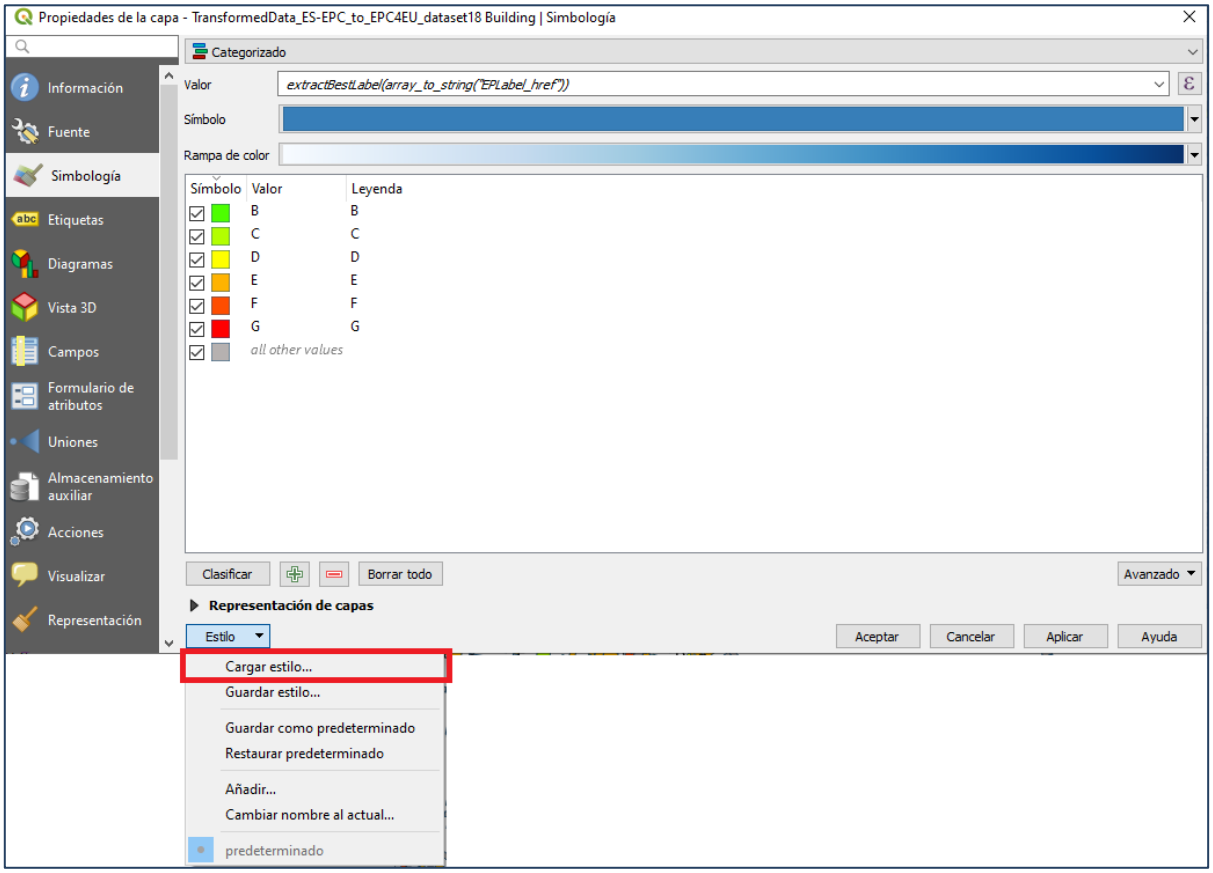
```
C:\Users\[USER_NAME]\AppData\Roaming\QGIS\QGIS3\profiles\default\python\expressions
```

If it is not the directory, this can be searched by opening the python console of QGIS3 and printing the path.

```
import sys; print(sys.path)
```

For loading the symbology and the properties of the tag, the style has to be loaded, as can be seen in Figure 28.

Figure 28. QGIS loading style



Source: own creation (CARTIF), 2020.

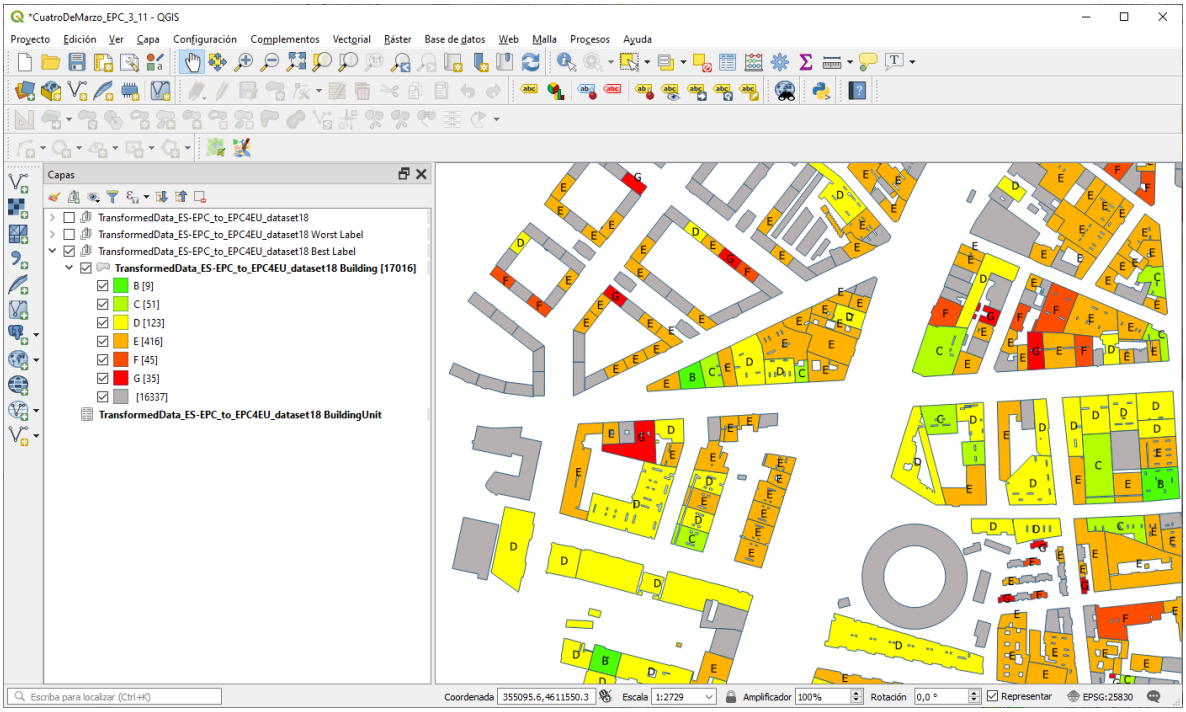
The styles for each type of visualization are:

- EPLabel_EPC4EU_extractBestLabel.qml
- EPLabel_EPC4EU_extractWorstLabel.qml
- EPLabel_EPC4EU_extractLabelMultivalues.qml

12.3 Visualisation of the results

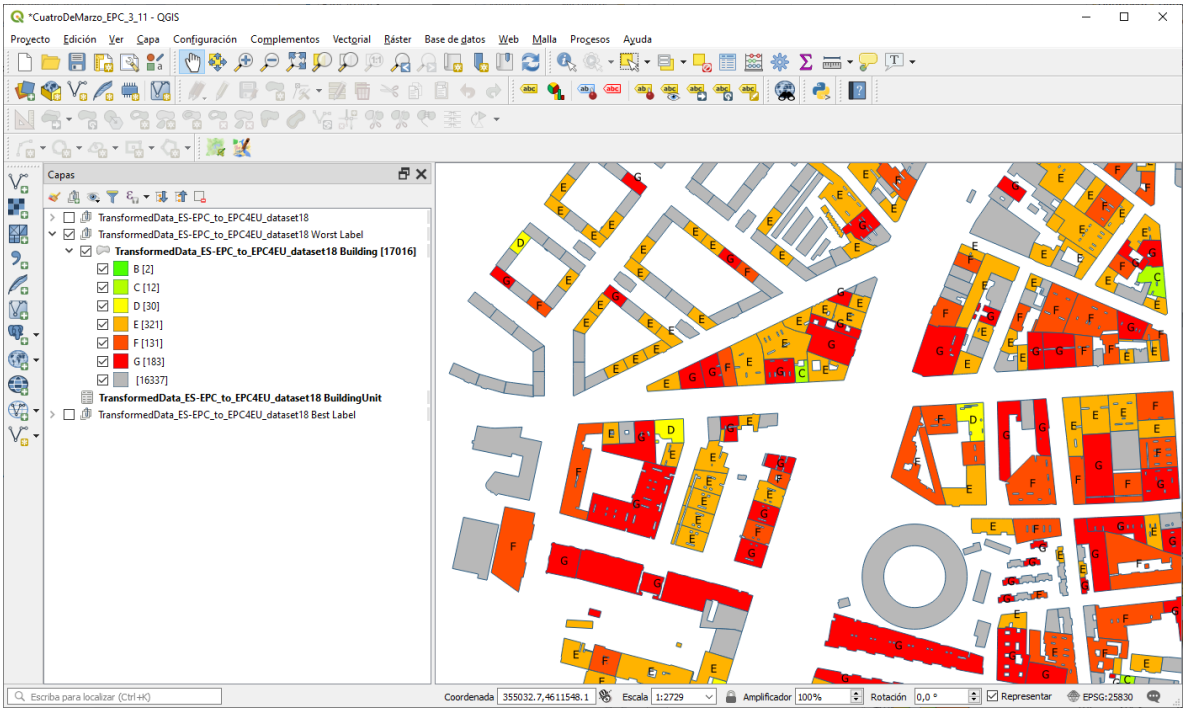
The results obtained (using any of the methods explained before) are shown in the following figures.

Figure 29. QGIS visualization of the best EPC



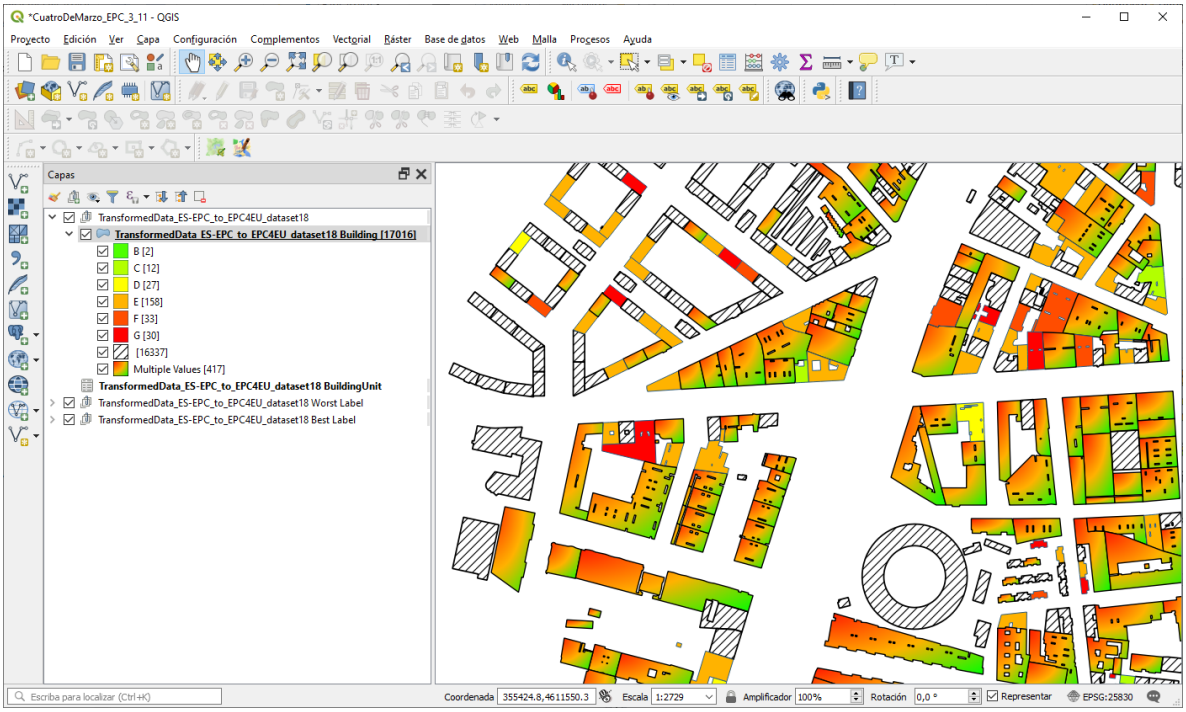
Source: own creation (CARTIF), 2020.

Figure 30. QGIS visualization of the worst EPC



Source: own creation (CARTIF), 2020.

Figure 31. QGIS visualization of the EPCs identifying the buildings with more than one EPC



Source: own creation (CARTIF), 2020.

13 Harmonized Energy Performance Certificates datasets

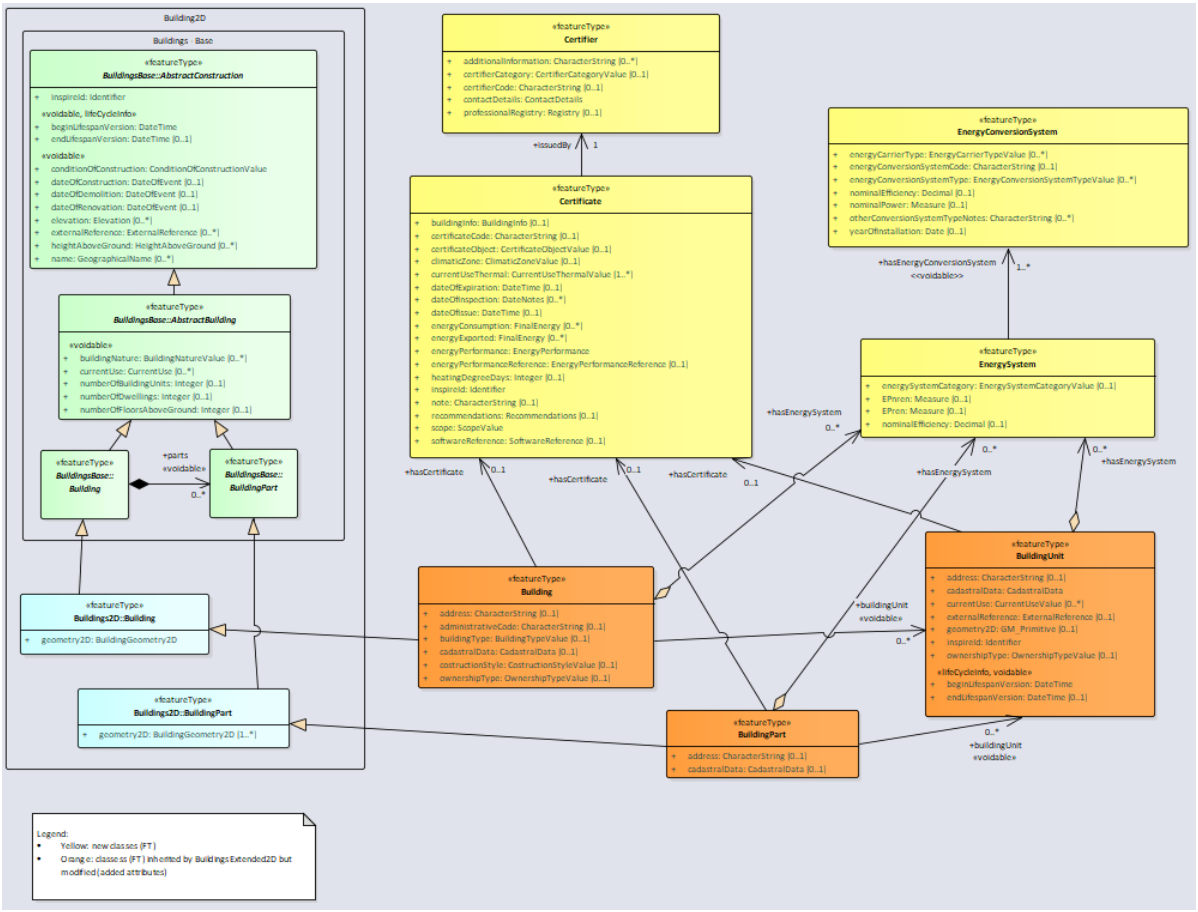
In this section, after referencing the target data model EPC4EU used in the harmonisation process, the harmonized Energy Performance Certificates datasets used as source datasets in the web application are presented: Trento EPC dataset and Castilla y León EPC dataset.

13.1 EPC4EU Target data model, version 3.0

The EPC4EU data model used as the target data model in the harmonisation process of the sample EPC datasets of Spain and Italy has been extensively described in the report of Step 1 of this use case.

The UML model is shown again in Figure 32 below.

Figure 32. EPC Target data model, version 3.0.



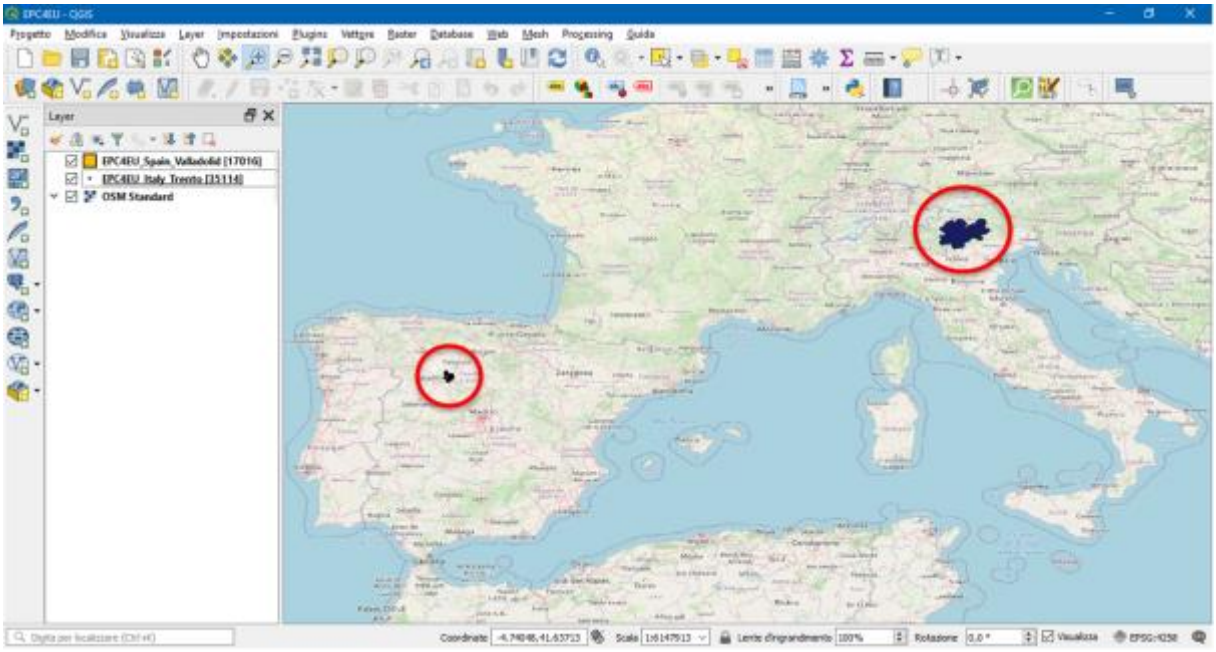
Source: own creation based on JRC, 2019.

13.2 EPC harmonised datasets

After the creation of the EPC4EU target data model and the mapping between the source (ES and IT) EPC data models and the EPC4EU target data model, the harmonisation of real EPC sample datasets of Spain and Italy has been performed.

The two harmonised datasets are shortly described in the following subsections, whilst Figure 33 shows as they are simultaneously visible using a GIS desktop (QGIS).

Figure 33. Harmonized EPC datasets visible in a GIS desktop (QGIS)



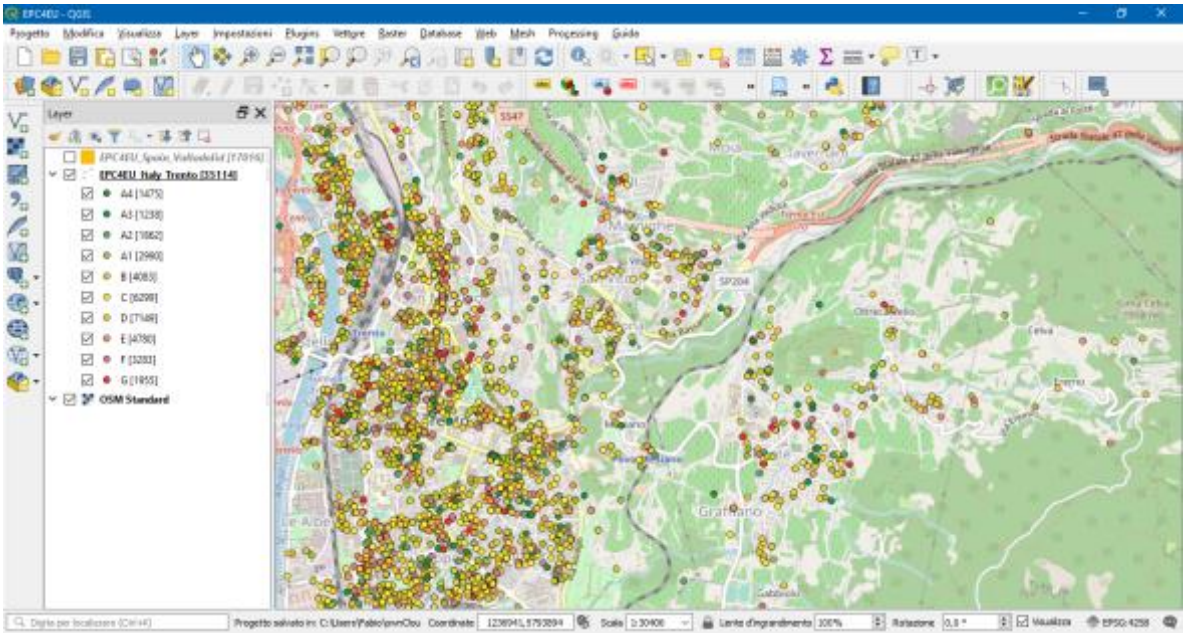
Source: own creation JRC, 2020.

13.2.1 Harmonised Trento EPC dataset

The dataset available for the harmonisation was provided by APRIE (Provincial Agency for Water and Energy) of the Autonomous Province of Trento (IT). The dataset contains all the certificates related to the private and public buildings of the Province of Trento, in total 35.114 records (see Figure 34).

Due to some restrictions about the publication of information related to private buildings, a subset has been extracted from the whole dataset, containing only the public buildings (schools, city hall, etc.). The subset used in the web app contains 340 certificates, and the related geometry of the buildings is a POINT (coordinates extracted from certificates).

Figure 34. Harmonised Trento EPC dataset in a GIS desktop (QGIS)



Source: own creation JRC, 2020.

13.2.2 Harmonised Castilla y León EPC dataset

The real data available for the harmonisation came from the EPC register in CyL region and was provided by EREN (the organism regional of Castilla y León for the Energy Management). A sample of the database with 2134 instances was available for this aim. The result of the harmonisation process was a valid GML file. Some relevant characteristics of the harmonised dataset are summarised below:

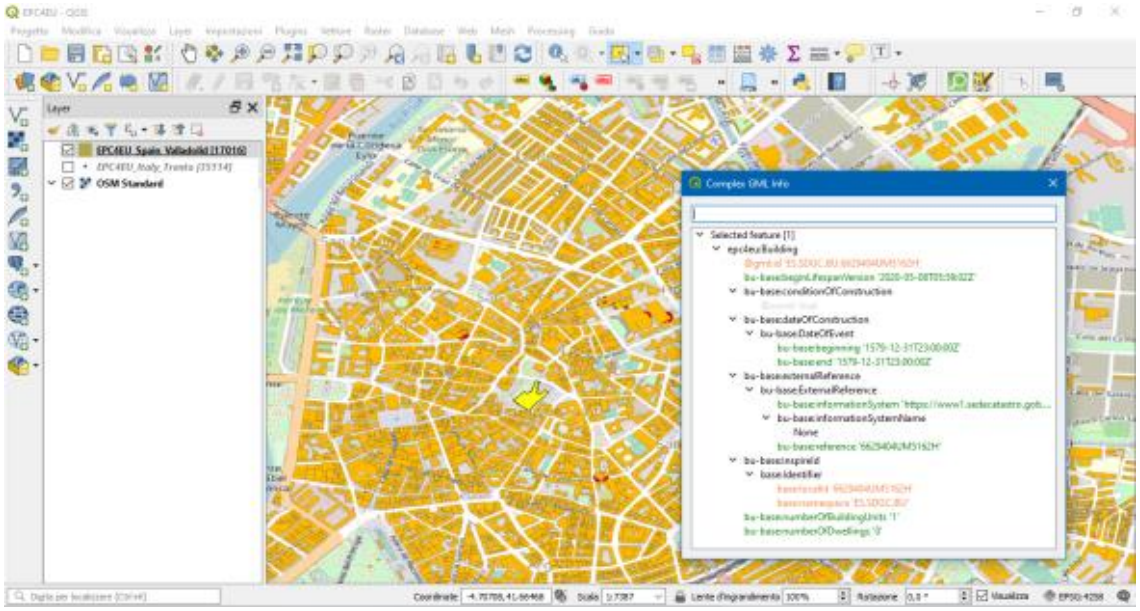
- There are buildings with more than one EPC (EPCs are directly related to BuildingUnits and buildings composed of more than one BuildingUnit).
- The buildings of the test dataset are located in the same city of Valladolid, and most of them in the same district.
- The geometry of the buildings is a POLYGON (extracted from the cadastre).

An extract of the GML dataset is shown in Figure 35, whilst Figure 36 shows how it appears when opened in QGIS. In this figure, the information for a building is shown.

Figure 35. Harmonized Castilla y León EPC dataset opened in Notepad.

Source: own creation CARTIF, 2020.

Figure 36. Harmonised Castilla y León EPC dataset in a GIS desktop (QGIS)



Source: own creation JRC, 2020.

14 Web application

To make accessible the harmonised datasets by non-GIS experts, a web application usable from a browser has been developed. It is described in the present section, starting from an analysis of a prototype developed in 2017, then describing the workflow for the re-use of this prototype and finally presenting the final results to visualise all the harmonised datasets with the same tool.

14.1 Web App prototype analysis

In this subsection, the first prototype of the web application, developed in 2017 in the frame of the ELISE Energy & Location Applications use case "INSPIRE Harmonisation of EPC of buildings datasets in Italy", is described.

Firstly the technical requirements for the tool were identified. The most relevant were the following:

- Tool with only read/visualisation functionalities
- Data format: INSPIRE GML 3.2 (produced using an INSPIRE extended schema)
- Data exposed by a WFS 2.0
- Web mapping library supporting WFS 2.0
- Web user interface HTML5
- The following development steps were planned:
- Prototyping of the client web application with static data
- Creation of the transformation proxy (live query to a WFS server)
- Deployment on a sandbox server for testing

The main planned functionality of the web app consisted in visualising the EPC harmonised dataset, adding an overlaying base layer on the map and visualising all buildings having an EPC, textured with different colours based on the different energy performance labels. Clicking on a building, a pop-up with additional information, consisting of pre-defined attributes of the harmonised EPC dataset, has to appear.

The analysis phase started with an investigation of the available web mapping libraries fulfilling the requirements. The most common open-source JavaScript libraries supporting complex GIS operations are OpenLayers and Leaflet. Although the differences between the two libraries are small, OpenLayers seemed the one is providing more functionalities. Therefore, the initial idea was to implement the prototype in both the libraries, mainly for testing purposes.

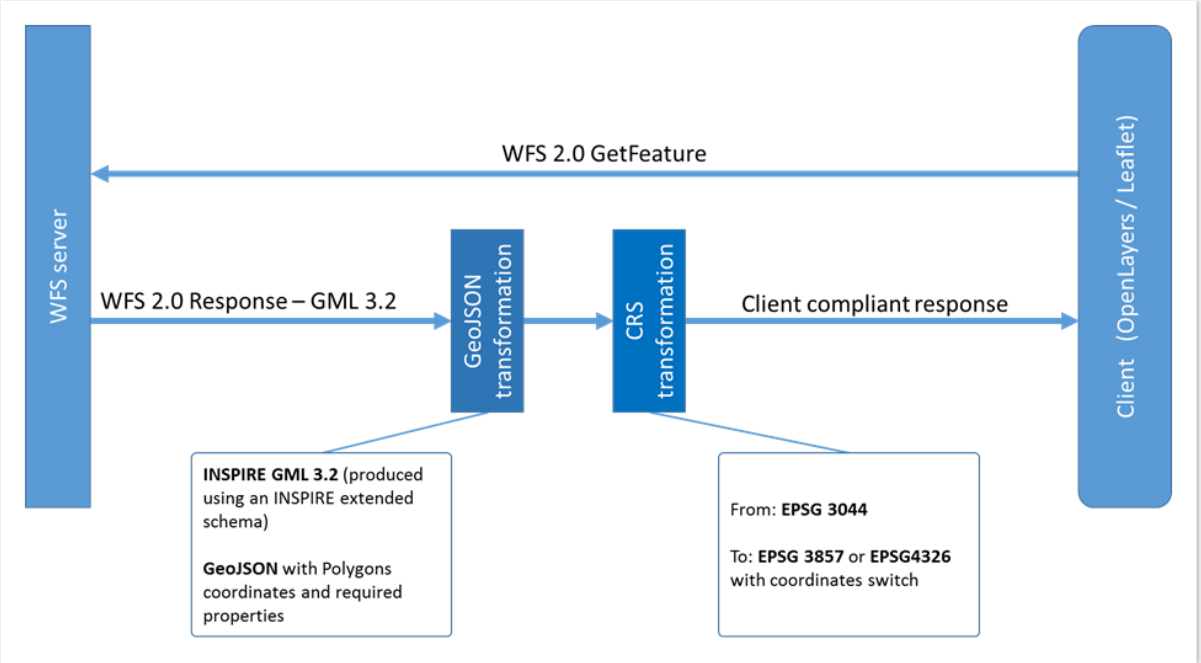
The input dataset came from a WFS 2.0.0 service and was encoded in INSPIRE GML 3.2 (produced using an INSPIRE extended schema). After some investigation, it appeared evident that the two mapping libraries previously identified could not handle WFS 2.0.0 and GML 3.2. Indeed, most of the implementations based on those libraries use GeoJSON as the input format for the dataset.

Another issue was related to the dataset CRS because the CRS natively supported by the web mapping libraries are EPSG:3857 or EPSG:4326, whilst the CRS of the input dataset was EPSG:3044 CRS. Therefore a CRS transformation was needed, as well as a swap of the coordinates.

The solution proposed to fulfil the requirements of the web mapping libraries was to use a broker/proxy approach to translate the data format (coming from the data provider) to the GeoJSON format.

The prototype of the transformation proxy was made in Node JS. The basic schema of the proxy is shown in Figure 37.

Figure 37. Basic schema of the transformation proxy



Source: own creation JRC, 2017.

The request coming from the client (web map library) is then forwarded to the WFS server. The answer from the WFS server is then transformed to GeoJSON format using an XSLT. This style sheet can be customized to fit the specific needs of each service.

In the example used to develop this prototype, the XSLT is configured to filter the complete WFS response and take only the features with the EPC dataset attributes shown in the pop-up. Furthermore, the XSLT creates the GeoJSON format with a subset of fields, taking only the ones required by this specific test case.

After the XSLT transformation, the output GeoJSON format is then processed by a CRS transformation system which transforms the original CRS to the one required by the client.

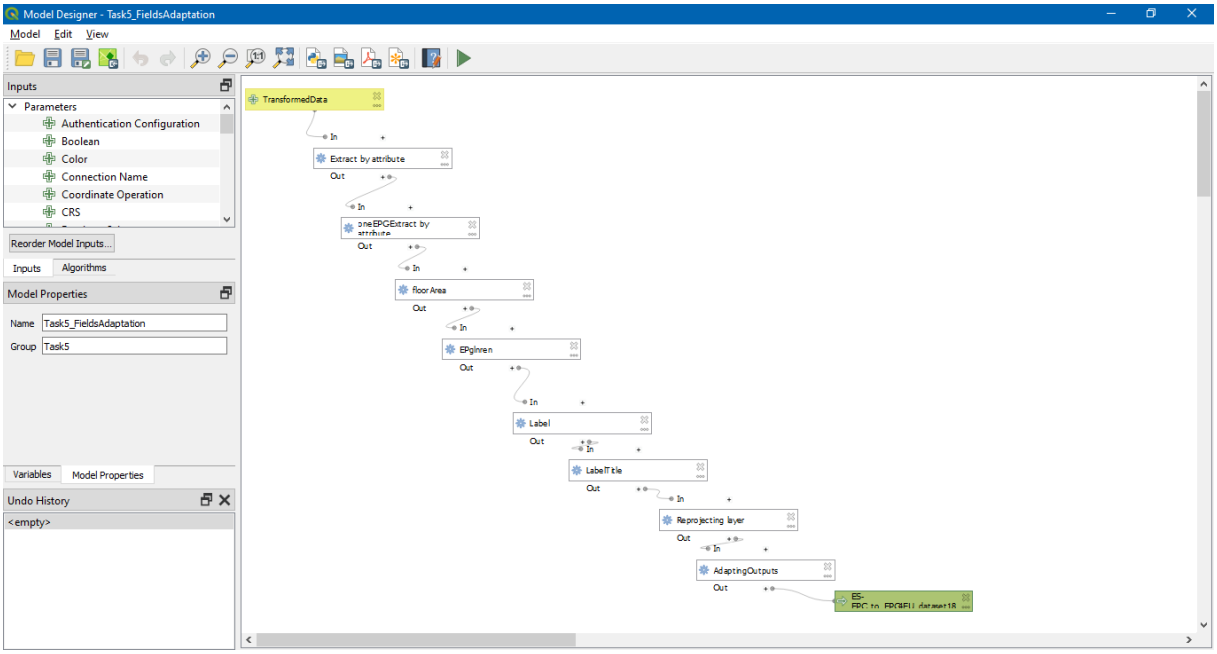
Finally, the response is sent back to the client application.

14.2 Adaptation of the Web App

After analysing the web prototype, the difficulties in adapting it to the present use case have been identified. A series of steps have been defined to adapt the prototype to the new datasets created. More details are provided in the next subsections.

It is important to highlight that in the adaptation of the web app, only Leaflet libraries have been used, and the process to generate the GeoJSON file (i.e. publication of the GML through a WFS and then transformation of the response in GeoJSON using an XSLT) has been simplified. To this end, *QGIS Graphical Modeler* (Figure 38) has been used, developing an ETL process for transforming the source GML files compliant to the EPC4EU data model into the target GeoJSON files, which have a simplified data model.

Figure 38. *QGIS Graphical Modeler*



Source: own creation, CARTIF, 2020.

14.2.1 Difficulties encountered for the adaptation of the web app

Analysing the web app prototype and the two datasets to be jointly visualised, some difficulties have been identified, due to some differences between the two datasets, described in more detail in the following.

1. **Heterogeneity in the geometry type of the two datasets**. Checking the two datasets to be jointly visualised, it can be seen that they have different geometry types. In the Trento EPC dataset, the geometry type used is *Point*, corresponding to one point inside the building the EPC belongs to. In Castilla y León EPC dataset, the used geometry type is *Polygon*, corresponding to the footprint of the building the EPC belongs to. This difference is important not only in terms of the building representation in the visualisation but also in terms of the approach used by the libraries to manage the datasets.
2. **Differences in the Energy Performance label taxonomy**. Although the methods used by the different countries for the energy performance assessments are based on the same European Directive, there are some differences in the EPC schemes used by different MS. In the case of the labels, the scales used for the two datasets are similar, from the letter A to the letter G, but in the Trento EPC dataset, it has a further subdivision for the letter A: A1, A2, A3 and A4. This difference has been taken into account for the common visualisation.
3. **More than one EPC for the same building**. Because the EPCs are related to the BuildingUnit and one Building can be composed by more or no BuildingUnits, the situation of one Building with more than one EPC can occur. In fact, in both datasets, we have several buildings with more than one EPC. The prototype did not consider this situation.

14.2.2 Adaptation of the prototype

To have the web app functioning, a set of adaptations have been done. Some of them have been done in the prototype web app code, but also the datasets have been modified. The changes done are described below:

Adaptation of the datasets:

1. **Eliminating buildings "out of scope".** In the Castilla y León dataset, the first adaptation has been eliminating the buildings without an EPC. In the harmonized dataset, all the buildings in the city were present, with and without an EPC. The buildings without EPCs have been considered "out of scope" concerning the web app and, therefore, eliminated.

2. **Adapting the buildings with more than one EPC.** The problem identified in the previous subsection about having more than one EPC for the same building has not been tackled. A first approach consisting of removing all but one EPC for the buildings with more than one EPC has been adopted.
3. **Eliminating attributes "out of scope".** The selected four attributes, whose values are shown in the pop-up, are EPLabel_href (identifying the EPC Label), end (identifying the date of construction), value (identifying the floor area), and EPglnren (identifying the energy global energy). The other attributes have been eliminated.
4. **Re-projecting the layers.** In the case of Leaflet, the CRS natively supported is EPSG:4326, so the datasets have been re-projected from the native CRS to EPSG 4326.

Adaptation of the web app:

1. **Adapting the codelists**. EPC label related codelists have been adapted taking into account the new INSPIRE codelists. The new values are:

```
codelistMatch:{

    epc_A:'https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/A',
    epc_B:'https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/B',
    epc_C:'https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/C',
    epc_D:'https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/D',
    epc_E:'https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/E',
    epc_F:'https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/F',
    epc_G:'https://inspire-sandbox.jrc.ec.europa.eu/codelist/energyPerformanceLabelValue/G'

}
```

2. **Eliminating parameters not present**. *typeBuilding* parameter has been eliminated from the pop-up because it is not present in the final dataset.
3. **Changing the properties of the map**. The map center as well as the zoom level have been changed in order to see the two datasets appearing simultaneously in the web app landing page:

```
// Map options

map:{

    center:{

        crs4326xy:[43.11,3.38]

    },

    zoom:6

},
```

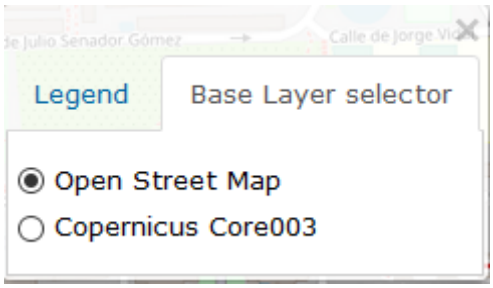
4. **Adapting the legend.** To have a common legend, the subdivisions existing for Label A in the Trento dataset (labels A1, A2, A3 and A4) have been merged in one label (Label A). The standard colours for the Energy Performance Labels have been selected and related to each Label.
5. **Adapting the app for Point geometries**. Considering that the prototype was developed to show the footprint of the buildings (i.e., for Polygon geometries), the web app has been adapted to deal with *Point* geometries. In this way, some style options for point markers have been added, and *Point* style has been managed.

14.3 Final web application

The application is a simple client-side web application that offers the visualization of the two datasets with two possible base layers and a legend to explain the colours of the elements.

For the selection of the base layer, a Base Layer selector has been implemented. With this selector, the user can select the base layer between *Open Street Map* and *Copernicus Core003*. The aspect of the Base Layer selector is shown in Figure 39.

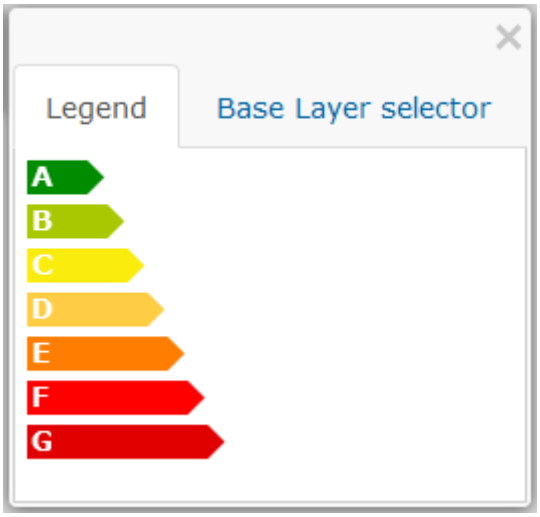
Figure 39. Base Layer selector



Source: own creation CARTIF, 2020.

The legend shows the colours used to fill in the *Polygons* and the *Points* elements in the map, according to the Energy Performance Label. The legend has the following aspect:

Figure 40. Legend



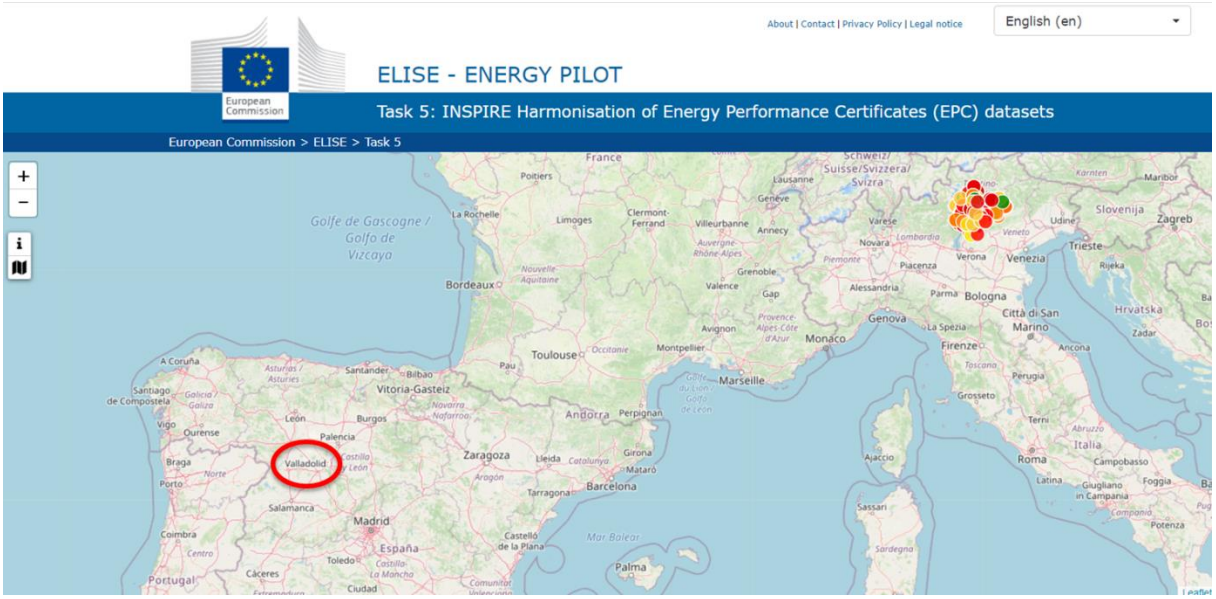
Source: own creation CARTIF, 2020.

The application can be accessed at the following link:

https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/epc4eu/webapp/

By clicking on the link, the aspect of the web page is shown in Figure 41. The zoom level of this map has been selected to access the Valladolid dataset or Trento dataset from the main page. The Trento dataset can be visualised at this level of the zoom because the *Points* always have the same size independently of the zoom level. In the case of the Valladolid dataset, the dataset is not visualised at this level because the geometry of the elements in this dataset is a *Polygon* and the sizes of the polygons are very small compared to the scale.

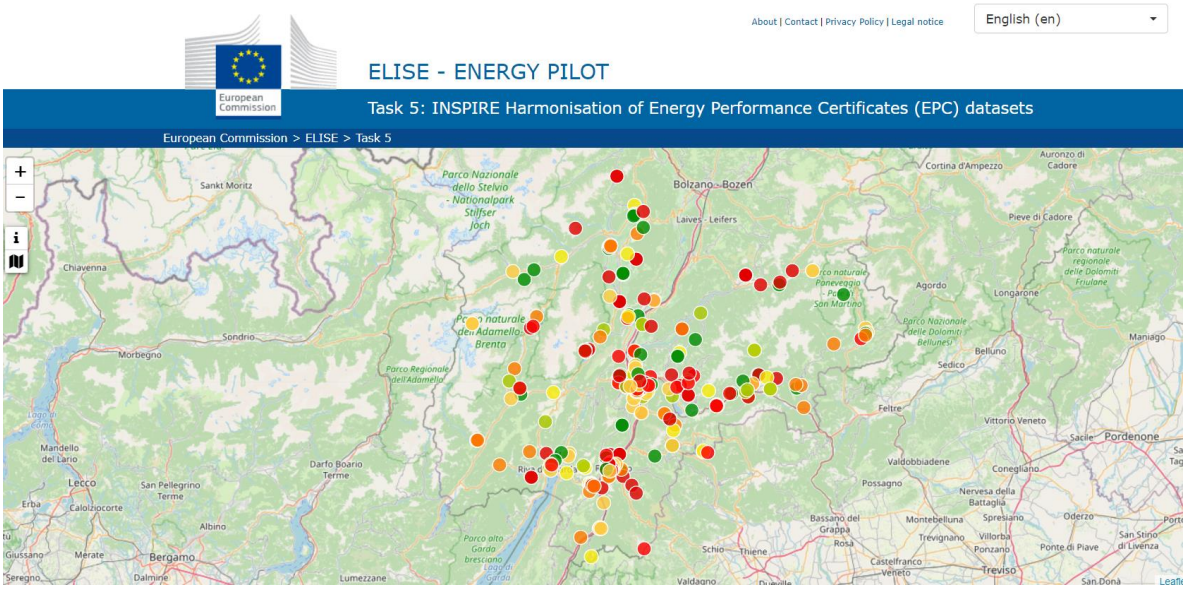
Figure 41. Lading page of web application



Source: own creation JRC, 2020.

Zooming on the Trento region, the web application shows the following:

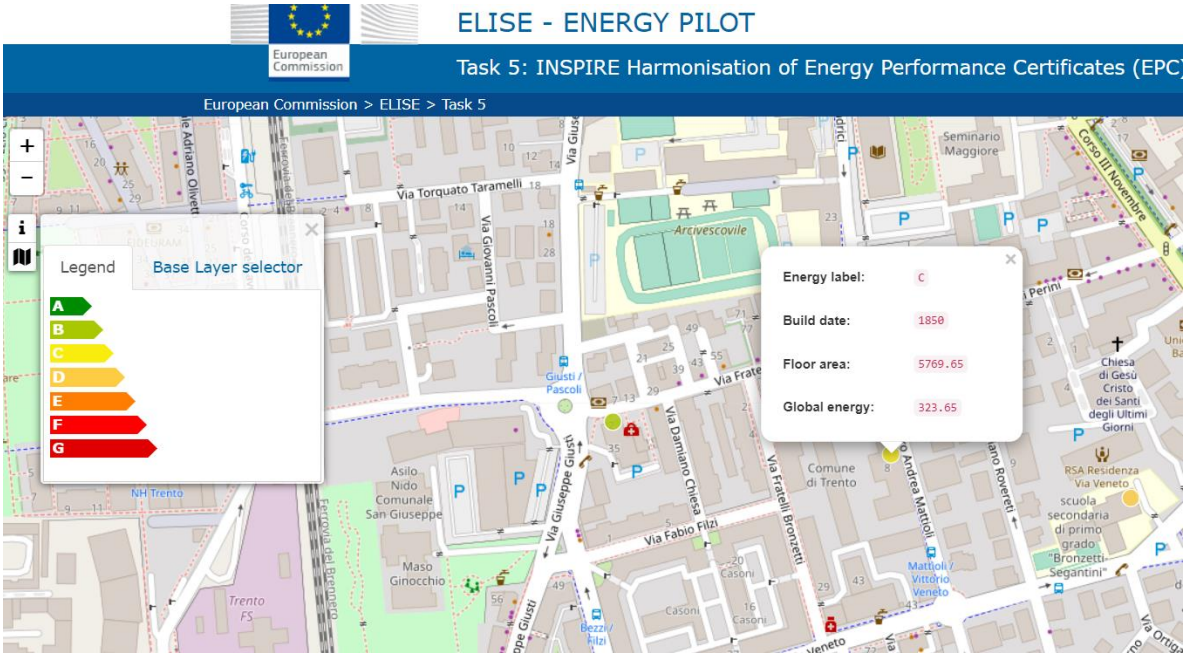
Figure 42. Web application, zoom in Trento



Source: own creation JRC, 2020.

Clicking on one of the points, the web application opens a pop-up with information about the year of construction, the floor area, the global energy and the energy label, as shown in Figure 43.

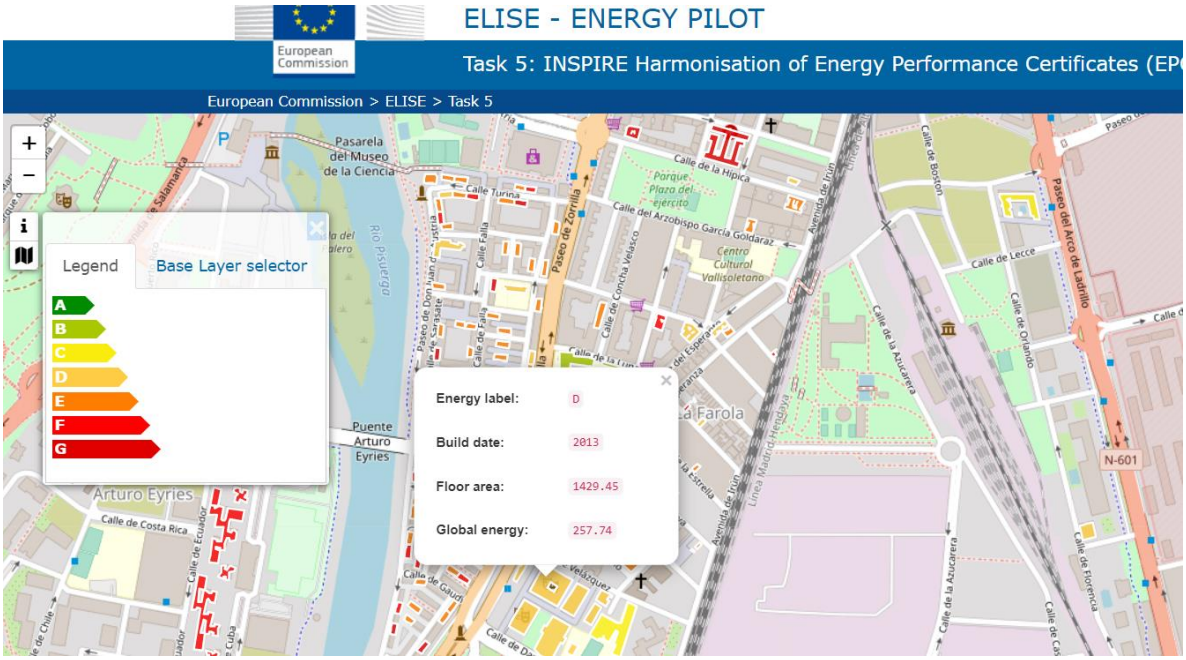
Figure 43. Web application, zoom on Trento Buildings



Source: own creation JRC, 2020.

The app has a similar aspect zooming in Castilla y León, but showing the buildings as polygons marking their footprints instead of showing points to indicate the location. In Figure 44, the aspect of the app in Valladolid city, with one example of building, is shown.

Figure 44. Web application, zoom in Valladolid, Castilla y León



Source: own creation JRC, 2020.

15 Conclusions

A methodology to create a new version of the EPC4EU data model capable of harmonising EPC datasets of Spain and Italy, starting from a previous version of the EPC4EU data model developed in another use case to harmonise EPC datasets adhering to the Italian data model for EPC, has been developed and applied.

The methodology has been designed to be easily re-used to generate further versions of the EPC4EU data model, capable of harmonising EPC datasets from other EU MS adhering to different EPC data models.

The data transformation process of a real EPC dataset consisting of a sample of the EPC register of CyL region, using the updated version of the "EPC4EU" target data model developed in Step 1 of the use case, has been successfully executed and a valid harmonised GML 3.2.1 file has been obtained.

A series of difficulties encountered during the transformation process has been described in detail, together with the corresponding solutions adopted, distinguishing between two types of issues: those encountered when filling the information related to the Building and BuildingPart feature types and those encountered during the transformation process itself.

Moreover, practical suggestions on how to properly visualise the harmonised data in QGIS have been provided, focusing on aspects of the thematisation of the EPC labels.

It is worth highlighting that both the solutions provided to overcome the obstacles encountered during the transformation process and the suggestions provided to properly visualise the harmonised data in QGIS, being fully re-usable in similar data transformation/visualisation contexts, can effectively support other actors facing similar issues.

Finally, the objective to make accessible by non-GIS experts heterogeneous EPC datasets coming from different sources and harmonised according to the common EPC4EU data model has been achieved through a web application accessible through a simple browser.

A first prototype of the web application realised in 2017 has been analysed, adapted, and re-used to make simultaneously accessible sample EPC datasets of Spain and Italy.

Clicking on a building for which an EPC is available, a pop-up with four main parameters and the corresponding values shows up, facilitating the EPC data usability by non-GIS experts.

The issues encountered in the web app adaptation, required by some differences present in the content and in the properties of the two datasets, and the related solutions adopted, have been extensively documented to be re-used in the context of possible future web app development extensions.

In conclusion, regarding the re-usability in other Member States of the EPC4EU methodology presented in this report, the following lessons have been learnt:

- One of the main problems encountered to adapt the EPC4EU target data model to the Spanish EPC scheme was the terminology used. Although the EPCs of both countries (ES and IT) are based on the EPBD, the names of the different parameters contained in the EPCs do not always match precisely or do not always refer exactly to the same element or the same units. In addition, the issue related to the different language issue has to be also taken into account. It was, therefore, necessary to study both models carefully to perform the adaptation. These particularities were also seen in the different code lists, which had to be adapted due to the particularities of the EPC in the different countries.
- Another important problem was the lack of geo-localized information present in the Spanish EPC. This problem was solved by combining the information from the EPCs with information from the Spanish cadastre, which, thanks to its adherence to INSPIRE, helped fill in other information in the target model not found in the original Spanish EPC.
- It is also worth highlighting the obligatory (or optional) nature of filling in some sections in the different EPCs. This obligation did not always refer to the same elements for the different models, so it was necessary to analyse it and change the cardinality of some elements accordingly.

List of abbreviations and definitions

API	Application Programming Interface
CRS	Coordinate reference system
CyL	Castilla y León
EA	Enterprise Architect
EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
EREN	Ente Público Regional de la Energía de Castilla y León
ETL	Extract, Transformation and Load
GIS	Geographic Information System
GML	Geographic Markup Language
JRC	Joint Research Centre
XML	eXtensible Markup Language
XSD	XML Schema Definition
XSLT	eXtensible Stylesheet Language Transformations
WFS	Web Feature Service

List of boxes

List of figures

List of tables

Annexes

Annex 1. Spanish EPC – XML version

This annex provides a list of the columns of a table[25] containing all the attributes of the XML version of the Energy Performance Certificates in Spain, which has been analysed in the context of the activities documented in this report. The basic list of attributes has been extracted from the official document *(*Informe de evaluación energética en formato digital (XML)[26]), with some additional processing to ease the quantification and further analysis performed.

- **Mandatory/optional:** Attributes are considered mandatory only if they always have to exist. For instance, inside the datatype "applicant", which is optional, the datasets "*fiscalCode*" and "name" are mandatory; however, since the data type is optional, they both will be considered not mandatory. All of the datasets that are not mandatory are considered optional.
- **Main cat. (Main category):** The main categories of attributes (in the Spanish data model) or feature type (in the target model) correspond to the main groups of attributes in each data model. They are shaded in grey. The main categories in the Spanish data model will be adapted to match the existing or new feature types in the target data model.
- **Sub-cat. (Subcategory):** Data types are elements containing other attributes inside them. In this sense, the term "sub-categories" is deployed in the case of the Spanish data model. These elements are not counted as attributes.
- **Code lists:** are lists of possible values that a determined attribute could have. These lists are commonly used in INSPIRE with the objective that each user can define their values to serve their needs. In the Spanish data model, all the attributes where a list of possible elements have been provided are considered code lists.
- **ID:** identification number of each attribute, which has been deployed in every table in the present document. The values "*na*." have been applied to those elements considered a main category (feature type) or sub-category (data type) since they are not counted as attributes.
- **Category:** Main categories are listed in this column. In the case of the Spanish data model, 17 categories are to be found.
- **#:** The mandatory attributes have been marked in blue and numbered to view them easily.
- **Field:** Name of the attribute, both in Spanish and English. The elements marked in grey correspond to main categories and subcategories.
- **Label:** Labels deployed in the XML are presented in this column. Also, the relationship among others is represented by depicting the different levels.
- **Section in PDF and ID:** This set of two columns indicates if the attribute or element is present in the PDF version of the Energy Performance Certificate. The ID number corresponds to the numbers marked in the Energy Performance Certificate example shown in Annex
- **Multiplicity:** these columns express if an attribute should appear one time (1), one or more (1...*), zero or more (0...*) or zero times (0). The different columns correspond with the level where the attribute is.
- **Const. (constraints):** some of the attributes have different constraints, which are expressed below.
- **T:** attributes that only refer to tertiary buildings.
- **R**: attributes that only refer to residential buildings.
- **CTE:** attributes related to the compliance with the Spanish Building Code (CTE – Código Técnico de la Edificación).
- **Admited values / format:** Type of admitted value or the format it may have. If an attribute has a set of possible values that are fixed, it is considered a code list, and the possible values are listed.

[25] https://joinup.ec.europa.eu/node/704540

[26] https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2021-07/20150625%20-%20Informe%20evaluaci%C3%B3n%20energ%C3%A9tica%20edificio%20en%20formato%20elec%20XML.pdf

— **Description:** Description in Spanish of the attribute.

— **Example:** An example of the possible value is provided in this column.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30795-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/500135

ISBN 978-92-76-40827-7