# Metadata for the Energy Performance Certificates of Buildings in Smart Cities

Ana Alice Baptista
Centro ALGORITMI, Universidade do Minho, Guimarães, Portugal
analice@dsi.uminho.pt

Sara Catarina Silva
Centro de Computação Gráfica, Guimarães, Portugal
a69185@alunos.uminho.pt

## Abstract

SusCity is a MIT Portugal project that falls within the scope of smart cities. One of its tasks aims to research and develop metadata artefacts to be used in the scope of a Linked Open Data platform. In this article, we report the process and results associated with the development of the following metadata artefacts: an application profile, a metadata schema and four controlled vocabularies. The application field is the energy certification of buildings. For the development of the application profile, we inspired ourselves in the Me4MAP method although we did not use it thoroughly. The creation of the metadata schema and controlled vocabularies involved the use of Wikidata, so all new terms (RDFS classes and properties and SKOS concepts) are related to Wikidata terms. The results include the application profile, the metadata schema and the controlled vocabularies. The application profile has 13 properties, four of which are new. The controlled vocabulary on measures for energy performance has 22 new terms spread over four levels. The remaining controlled vocabularies just hold a few terms each. All the artefacts are open to the community for use and reuse.

**Keywords:** smart cities, energy performance certificate, metadata, linked open data, application profile.

## 1. Introduction

SusCity (Urban data-driven models for creative and resourceful urban transitions) is an MIT Portugal project in the scope of smart cities. The MIT Portugal program "is a strategic partnership between Portuguese Universities and Research Centres, the Massachusetts Institute of Technology and partners from industry and government (...) [and] its goal is to strengthen the country's knowledge base and international competitiveness through a strategic investment in people, knowledge and ideas in innovative technology sectors" ('MIT Portugal', 2017). The development of the concept of a "smart city" rises from a complex association between technology, society, economy, administration and politics. A smart city is a city that invests in human and social capital, traditional (transport) and modern (ICT) infrastructures to enable a sustainable economic development and ensure a high quality of life with the intelligent management of natural resources (Caragliu, Del Bo, & Nijkamp, 2011). SusCity "is focused on developing and integrating new tools and services to increase urban resource efficiency with minimum environmental impacts while contributing to promote economic development and preserving the actual levels of reliability" (SUSCITY, 2016).

The project is divided into 6 (six) work packages (WPs), from which WP1, WP3, WP4 and WP5 generate data that is fed into the data processing platform created in the scope of WP2 (see FIG. 1). The data processing platform has several constituents, two of which are the Analytics module and the Linked Open Data (LOD) platform. The internal analytics module provides mechanisms for analysing vast amounts of consolidated data. The LOD platform is where linked data is made openly available and for which we are currently developing several metadata application profiles (MAPs), controlled vocabularies and metadata schemas. This article reports on the development process and outcomes of a MAP, one metadata schema and four controlled vocabularies related to energy performance certificates (EPCs) of buildings.

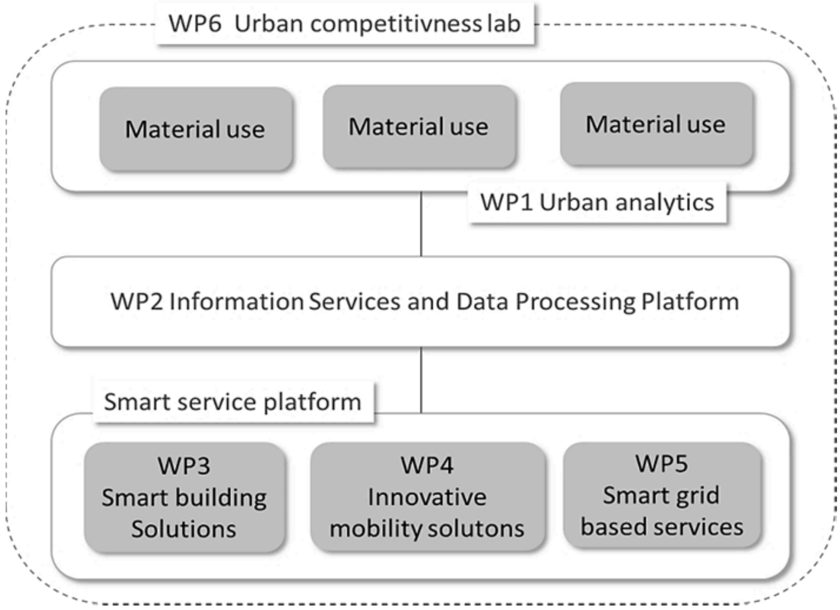


FIG. 1 - SusCity project structure (FCT Web site as cited by Aelenei et al., 2016).

In the past years, the world has seen a rapid growth in energy consumption. This growth has raised concerns related to exhaustion of energy resources and environmental impacts. Buildings are believed to be one of the main contributors to the increase in power consumption, exceeding major sectors such as industry and transportation (Pérez-Lombard, Ortiz, & Pout, 2008). Therefore, energy performance of buildings must be a top priority at a national and international level. Increase the energy performance of buildings can be achieved during the architectural design when choosing the building's mechanical system or with a smart management of the building utilities (Sozer, 2010). It can also be achieved by adopting maintenance measures specifically targeted to increase the energy performance of the building. Space heating is believed to be the most energy consuming end use (71%), followed by water heating (12%), cooking (4%) and lighting, air conditioning and other appliances (15%) (Odyssee-Mure, 2012).

In 2016, we and others (see acknowledgements) developed a research work on Linked Open Data, energy and energy efficiency of buildings. That work aimed at identifying, under these topics, 1) the main international and European standards; 2) metadata schemas and controlled vocabularies used in open datasets in the cities of Paris, London and Amsterdam; and 3) the metadata elements, encoding schemas and units of measure used in the datasets and catalogues identified in 2). The searches were made using the following services and websites: Google Scholar, Google, Linked Open Vocabularies (LOV), European Data Portal, and the Web open data portals of the cities mentioned above. As we were having difficulties to find the information we were looking for, in some cases we expanded the searches to other open data portals. Unsurprisingly, we were able to find legislation, international or European standards and guidelines regarding the energy efficiency of buildings and regarding Linked Open Data (LOD) separately, but almost nothing regarding the two combined. The little we found is limited to specific datasets encoded in RDF but none of them had the specific information we needed.

Because we could not find the artefacts we needed, we had to create them. We started by developing the MAP for the EPCs. The development of this MAP eventually required other developments: a metadata schema that would house new properties and classes, some of them specific to this domain and three small controlled vocabularies to constrain the range of three properties. Yet, another development was made as a request from our partner, ADENE - Agência para a energia (agency for the energy): the encoding in SKOS (Simple Knowledge Organization System) of a controlled vocabulary about measures to improve the energy performance of buildings. These developments required some design choices: whenever possible, we have given priority to simplicity and potential interoperability to the detriment of the wealth of semantic description. The work reported here has been done in close collaboration with ADENE, which is responsible for the management and operation of the Buildings' Energy Certification System, which has already 10 years of existence and has issued more than 1,250,000 energy certificates (ADENE, 2017).

This article is divided into four sections, starting with this introduction followed by a section on the data and methodological procedures, where we characterise the data that we had available and the methods and techniques used for the development of the metadata artefacts. Then we present the results in three different subsections: subsection 3.1 for the constraints matrix of the MAP, subsection 3.2 for the metadata schema and subsection 3.3 for the controlled vocabularies. In section 4 we present the final remarks and future work. The prefixes and URIs of the namespaces are made available before the references.

## 2. Data and methodological procedures

### 2.1. The data

We are opening two kinds of data: data sent by project partners and data generated within WP2 by the internal analytics module. In most of the cases, data sent by partners cannot be used in full because it can lead to privacy or security issues. For example, we cannot open data on energy consumption of each dwelling because besides being itself private data, it could allow the inference of other private information concerning those buildings or even concerning their residents (e.g., periods of time when those citizens are at or out of home). The same goes for security concerns: certain information, such as the architectural plans of buildings, or the building materials, could be used for purposes that could jeopardise the safety of their residents (e.g., terrorist attacks). For these reasons, when developing the MAP, it was necessary to eliminate some details about the data that came to our hands. On the contrary, the data generated within the WP2 internal analytics module already arrived anonymized and ready to be open, so there was no need to foresee changes when designing the MAP.

At their website, ADENE makes openly available for human consumption a fraction of the data related to individual EPCs (see http://www.adene.pt/sce/micro/certificados-energeticos). The data related to this article is part of that fraction as ADENE considered that because machine readable data is processable and easily relatable to other data, information about individual buildings should not be disclosed in a machine readable way. Therefore, door and floor numbers of apartments have not been entered into the MAP. Because of this and because EPCs already have parts of the address as standalone fields (e.g., Parish or Municipality), we decided to use the attribute "street" instead of "address" which was the one originally present at the EPCs. ADENE also considered that the name and number of the energy expert that analysed the building should not be opened in a machine readable way. Once that the data about the EPCs is to be shared globally, we decided to add the attribute Country. Therefore, the data to be opened for each EPC was the following:

- Certificate number - number that identifies the EPC;
- Document type - EPCs can be of two types: Regulatory Compliance Statement or Energy Certificate;
- Date of issue - the start date of the EPC;
- Validity date - the last day for which the EPC is valid for;
- Energy label - label that represents the energy performance;
- Country - country of the building;
- District - district of the building;
- County[1] - county of the building;
- Civil parish[2] - parish of the building;
- Street - street of the building;
- Land registry office - office that registers the ownership of land and property;
- Land registry entry number - number that identifies the property in the land registry office;
- Type of use of building - The type of use given to the building.

An EPC should clearly portray the energy performance situation of the building and should provide recommendations to the owner to improve the energy performance of the building in question (European Parliament and the Council of the European Union, 2003). One of the controlled vocabularies that were encoded in SKOS represents the improvement measures that ADENE uses in their recommendations to owners. In our case, as we are only dealing with a small fraction of the EPC's data, this data does not have a direct relationship with the controlled vocabulary of improvement measures, as they do not directly link to each other. Nevertheless, ADENE requested that we encoded it in SKOS, so that it could be made available for use and reuse both by them and by other organisations. The data and the structure of the controlled vocabulary were provided by ADENE in a word file.

### 2.2. Methods and techniques

When we started the development of the MAP we already knew what data we would release, so we did not need to go through the vast majority of steps that others perform and that are foreseen in the Singapore Framework (Nilsson, Baker, & Johnston, 2008) or in Me4MAP (Malta & Baptista, 2013). For example, we did not need to define the functional requirements as they had already been defined for the human interfaces (Costa & Santos, 2015; 2016) - our mission was just to make widely available in a machine readable way the data that was being used in human interfaces. Similarly, we could have designed our constraints matrix without having designed the domain model. However, a precise specification of the domain model is relevant in cases like this, since it helps us to keep in mind the context of the properties we intend to use. It is also appropriate for documentation purposes. In our case, we started the design of the MAP in the domain model (see FIG. 2).

---

[1] The administrative division of the territory varies from country to country. The word município in Portuguese refers to a place with a certain degree of administrative autonomy and run by a local government. It aggregates several freguesias. Several municípios are part of a district. For the sake of interoperability we will use the word "County" although we know the meaning may not be exactly the same.

[2] In Portugal, the smallest administrative unit is called freguesia and it may correspond with what in some countries is called civil parish, term that we will use.

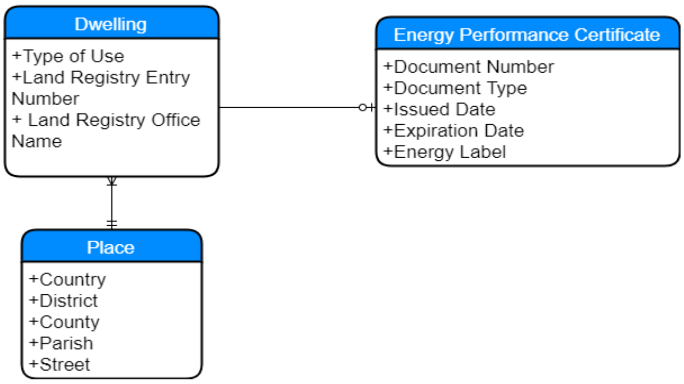


FIG. 2 - Domain model of the EPCs data opened in the scope of the SusCity project.

After designing the domain model, we looked for properties that matched its attributes. The majority of the searches for properties were done using LOV, which is a catalogue of the available vocabularies that allow the description of data on the Web (Vandenbussche, Atemezing, Poveda-Villalón, & Vatant, 2017). The identified schemas were also used to look for properties. One important aspect when choosing a property is to verify if it has restrictions (domains or ranges) that are incompatible with the intended use in our application. Although we are aware that this is not something taken into consideration by some developers, we think that it is an important aspect of the semantics of properties and we did not want to neglect it. Sometimes we have come to situations where we have had to do without something: either the semantic rigor of the definition of the property in the original schema, or the choice of a property of a widely used schema, or the respect for the rigor of the restrictions (compatibility between domain or the MAP range and the original domain or range of the property).

As an example, we have the attribute Freguesia (Civil Parish) for which we could have used the DBpedia property `parish`. However, this property has as domain the DBpedia class `PopulatedPlace` that is defined as "a place or area with clustered or scattered buildings and a permanent human population " (see http://dbpedia.org/ontology/PopulatedPlace). This domain makes all sense but not in our case, as the domain for this property should be left unspecified or it should be a document: the EPC. This apparent nonsense results from a design simplification: it is not an EPC that has a direct relationship with a given parish, but a building that is placed in a local that corresponds to that parish. We could have opted for having three classes, EPC, Dwelling and Place, which could be related by some properties such as Dublin Core `relation` or one or more of its sub-properties. If this were the case, the DBpedia property `parish` would apply well as its domain fits well with the class we could have defined (our Place would also be a populated place). However, we preferred to dispense with some semantic rigour in the description of resources and gain simplicity, mainly in the future processing of the resources' descriptions: it will not be necessary to execute complex queries to obtain all the required information as all the properties are constrained in the MAP to have as domain just one class (the Energy Performance Certificate), which implies that we have just one description per EPC. Summarizing, since we do not know the future users of our data, we endeavoured to opt for the simplicity of the descriptions and the possible semantic rigour in the use of the properties in what regards their constraints.

We had similar problems in what regards ranges of properties, but in some of these cases we dispensed with the respect for the original range, otherwise we would have a multiplication of properties and metadata schemas in our MAP. Our option, then, was to use just one property per attribute, even if we had to dispense with the rigour of ranges. For example, the `vcard:country-name` property has as range `xsd:string` and we are using it as a datatype property and also as an object property.

For each attribute that we wanted to use, we looked for a LOD property that corresponded to it. In cases where we found more than one candidate property that could be applied, we used the following priorities of choice: 1) property in a schema corresponding to some standard or some recommendation of some relevant entity in the field of application (energy certification) or in LOD; 2) property widely used by the metadata community; or, finally, 3) property defined in a formal schema and used by at least two datasets. Only in cases where we could not find properties that satisfied at least one of these criteria, we decided to create new properties. We then filled a table, the constraints matrix, with the following columns related to properties: attribute original name, label, namespace, property name, original domain, original range, domain in the MAP, range in the MAP, and cardinality (see Table 1 and Table 2). This constraints matrix is very different from those advocated by Coyle and Baker (2009) and by Malta and Baptista (2013). We have made the changes we deemed necessary to meet our needs for description and representation. Therefore, comparing do Malta and Baptista's table:

- We inserted a new column, "Attribute Original Name", that refers to the original name of the attribute in the domain model, so that the transition from the domain model to the constraint table is clear and evident.

- We kept the column "Label" (in English).

- The column "Property" was broken down in two columns: "Namespace" and "Property Name", so that we could grasp quickly which were the namespaces being used.

- We inserted two new columns, "Original Domain" and "Domain in the MAP", so that we could easily verify the compatibility of the properties' domains. The domain in the MAP should be a domain that is compatible with the original domain, i.e., it should be the same of the original domain or one of its sub-domains. For example, we can define as domain a specific type of document when the original domain is a general document.

- The column "Range" was broken down in two columns: "Original Range" and "Range in the MAP", so that we could easily verify the compatibility of the properties' ranges. The range in the MAP should be a range that is compatible with the original range, i.e., it should be the same of the original range or one of its sub-ranges (as you will see later, this was not always achieved). For example, we can define as range a controlled vocabulary in cases where the original range is unspecified.

- We removed the columns "Value String", "SES URI", "Value URI" and "VES URI" as their contents may be placed in the column "Range in the MAP".

- The columns "Min" and "Max" were replaced by the column "Cardinality", which means the number of times a given property may be used in a single description.

- We removed the columns "Type", "Usage" and "Related Description". We are using the "Range in the MAP" column to specify the type and we are specifying the usage at the level of the domain model. The "Related Description" is unnecessary in our case as we are only using one class. If we were using more, we could have inserted a new property for the relation and specify the related class in the "Range in the MAP" column.

The development of this MAP revealed the need to create a metadata schema with four new properties and a class. This schema was created using the Resource Description Framework Schema (RDFS). Equivalent terms for the properties and the class were created in Wikidata for two reasons: 1) to have them open for definition, evolution and control by the crowd (live properties); 2) expectation of having more stable URI's that assure long term existence of the terms; which result in 3) expectation of greater interoperability in the future. However, we created them with minimum structure so that they were not unnecessarily constrained. The structure was instead defined in the RDFS file.

For terms' names, Wikidata provides a character string mainly constituted of numbers, i.e., by reading the term name through the URL humans cannot immediately infer its semantics, even in a general way. Due to this fact, we created easily human readable term names in our RDFS file and then related them to the Wikidata terms through `owl:equivalentClass` and `owl:equivalentProperty` properties (OWL - Web Ontology Language). We also did not define domain and range restrictions to these new properties, leaving them open to different usages. The restrictions were instead defined at the MAP level. This facilitates the reuse of terms across different schemas while safeguarding minimal semantics.

Similarly, regarding the controlled vocabularies, all SKOS concepts have equivalent terms in Wikidata. The SKOS and Wikidata terms were related via `owl:sameAs` property. We also chose to establish in Wikidata the minimum relations between the concepts and to represent most of the vocabulary structure in the SKOS files. This facilitates the reuse of terms across controlled vocabularies while safeguarding minimal semantics.

## 3. Results

### 3.1. Constraints matrix

Screenshots of the constraints matrix are presented in Table 1 and Table 2. The constraints matrix is fully available online at http://hdl.handle.net/1822/46455.

TABLE 1. Constraints matrix for the EPC SusCity Metadata Application Profile (part 1).

| Attribute original name | Label | Namespace | Namespace prefix | Property name |
|---|---|---|---|---|
| Nº de documento (certificado) | Document number | http://dbpedia.org/ontology/ | dbpedia-owl | documentNumber |
| Tipo de documento | Document type | http://purl.org/dc/terms/ | dct | type |
| Data de emissão | Date of Issuance | http://purl.org/dc/terms/ | dct | issued |
| Data de validade | Date of Validity | http://purl.org/dc/terms/ | dct | valid |
| Classe energética | Energy Label | http://opendata.dsi.uminho.pt/terms/energy/EPCterms/ | ec | energyLabel |
| País | Country | http://www.w3.org/2006/vcard/ | vcard | country-name |
| Distrito | District | http://data.ordnancesurvey.co.uk/ontology/admingeo/ | osadm | district |
| Concelho | County | http://data.ordnancesurvey.co.uk/ontology/admingeo/ | osadm | county |
| Freguesia | Parish | http://data.ordnancesurvey.co.uk/ontology/admingeo/ | osadm | parish |
| Rua | Street | http://www.w3.org/2006/vcard/ | vcard | street-address |
| Conservatória do Registo Predial | Land Resgistry Office | http://opendata.dsi.uminho.pt/terms/energy/EPCterms/ | ec | landRegistryOffice |
| Artigo matricial nº | Land Registry Number | http://opendata.dsi.uminho.pt/terms/energy/EPCterms/ | ec | landRegistryNumber |
| Tipo de uso do edifício ou fracção | Type of use of building | http://opendata.dsi.uminho.pt/terms/energy/EPCterms/ | ec | useOfBuilding |

Some attributes of ADENE's EPCs (see http://www.adene.pt/sce/micro/certificados-energeticos) were not considered during the development of the MAP because of privacy and/or security reasons as reported above. These attributes are "address" and "building unit". We decided that the address would be left at the street level (buildings and dwellings are not identified). As we already had some components of the address as standalone attributes (Country, District, Municipality, Parish), we just needed to include the attribute Street instead. Also, as reported above, ADENE decided to leave out the name and number of the expert who issued the EPC. Some of the choices we had to make when creating this table are pretty straightforward and do not need further explanations.

TABLE 2. Constraints matrix for the EPC SusCity Metadata Application Profile (part 2).

| Namespace prefix | Property name | Original domain | Original range | Domain in the MAP | Range in the MAP | Cardinality |
|---|---|---|---|---|---|---|
| dbpedia-owl | documentNumber | dbo:Document | xsd:string | ec:EnergyPerformanceCertificate | xsd:string | 1 |
| dct | type | Unspecified | rdfs:Class | ec:EnergyPerformanceCertificate | string, EPC Types vocabulary (http://opendata.dsi.uminho.pt/terms/energy/EPCTypes.skos) | 1-2 |
| dct | issued | Unspecified | rdfs:Literal | ec:EnergyPerformanceCertificate | xsd:date | 1 |
| dct | valid | Unspecified | rdfs:Literal | ec:EnergyPerformanceCertificate | xsd:date | 1 |
| ec | energyLabel | Unspecified | Unspecified | ec:EnergyPerformanceCertificate | xsd:string + Energy Labels vocabulary (http://opendata.dsi.uminho.pt/terms/energy/energyLabels.skos) | 1 |
| vcard | country-name | Unspecified | xsd:string | ec:EnergyPerformanceCertificate | xsd:string, TGN, GeoNames | 1-3 |
| osadm | district | Unspecified | osadm:District & others | ec:EnergyPerformanceCertificate | xsd:string, TGN, GeoNames | 1-3 |
| osadm | county | Unspecified | osadm:County & others | ec:EnergyPerformanceCertificate | xsd:string, TGN, GeoNames | 1-3 |
| osadm | parish | Unspecified | osadm:CivilParish & Others | ec:EnergyPerformanceCertificate | xsd:string, TGN, GeoNames | 1-3 |
| vcard | street-address | Unspecified | xsd:string | ec:EnergyPerformanceCertificate | xsd:string | 1 |
| ec | landRegistryOffice | Unspecified | Unspecified | ec:EnergyPerformanceCertificate | xsd:string, New controlled vocabulary to be encoded in SKOS | 1-2 |
| ec | landRegistryNumber | Unspecified | Unspecified | ec:EnergyPerformanceCertificate | xsd:string | 1 |
| ec | useOfBuilding | Unspecified | Unspecified | ec:EnergyPerformanceCertificate | xsd:string, EPC Types of Buildings vocabulary (http://opendata.dsi.uminho.pt/terms/energy/EPCTypesOfBuildings.skos) | 1-2 |

However, the following cases need a clarification:

- The `dbpedia-owl:documentNumber` property for the Document Number attribute has as its original domain is `dbo:Document`. The new class `ec:EnergyPerformanceCertificate` is encoded in the metadata schema as subclass of (`rdfs:subClassOf`) `dbo:Document`. The domain of this property in our MAP is then restricted to `ec:EnergyPerformanceCertificate`. The original domain of the rest of the properties is unspecified, but in this MAP all of them are restricted to `ec:EnergyPerformanceCertificate`.
- The properties `dct:type`, `ec:energyLabel` and `ec:useOfBuilding` have as range the terms of controlled vocabularies created by us and encoded in SKOS (see next section).
- The OSADM schema does not provide properties for the Country and Street attributes, reason why we used another schema. VCard was chosen because it is widely known and used. Despite we tried, we could not find the five address related properties in a single metadata schema especially due to the peculiarity of attributes like County or Civil Parish.
- The properties `vcard:country-name`, `osadm:district`, `osadm:county` and `osadm:parish` will be used having as range in the MAP the terms of the Thesaurus of Geographic Names (TGN), GeoNames and/or a value string.
- The properties `ec:energyLabel`, `ec:landRegistryOffice`, `ec:landRegistryEntryNumber` and `ec:useOfBuilding` were created by us in RDFS.
- The property `ec:landRegistryOffice` has as range a value string and terms of a controlled vocabulary to be encoded in SKOS in the future.

### 3.2. Metadata schema

The new metadata schema has four new properties and a new class. The following properties have been created in RDFS:

- Land Registry Office (conservatória do registo predial), property: `landRegistryOffice` - office that registers the ownership of land and property;
- Land Registry Entry Number (Artigo matricial nº), property: `landRegistryEntryNumber` – number that identifies the property in the land registry office;
- Type of building or building part (Tipo de uso de edifício ou fracção), property: `UseOfLandOrProperty` – The type of use given to the LandOrProperty;
- Energy label (classe energética), property: `energyLabel` – label that represents the energy performance of some thing.

The following class has been created in RDFS:

- Energy Performance Certificate, class: `EnergyPerformanceCertificate`.

The domain and range of all properties were left unspecified. The new metadata schema is available at http://opendata.dsi.uminho.pt/terms/energy/ .

### 3.3. Controlled vocabularies

We have created and encoded four controlled vocabularies in SKOS. Three of them derive directly from the development of the MAP and have the following terms:

- For the type of document, the EPC Types vocabulary:
  - `Regulatory Compliance Statement`
  - `Energy Certificate`
- For the type of use of the building, the EPC Types of Buildings vocabulary:
  - `Domestic`
  - `Non-domestic (small commercial and service buildings and large small commercial and service buildings)`
- For the energy labels, the Energy Labels vocabulary:
  - Terms that coincide with the energy labels currently considered by ADENE, as shown in Figure 3.


FIG. 3 - Energy labels considered by ADENE (see http://www.adene.pt/sce/o-que-e-1).

The Energy Labels vocabulary currently only has the European energy labels. Other energy labels may be added in the future. It would be interesting if, at the MAP level, we could constrain the range of a property to a part of a controlled vocabulary. This would allow us to constrain the range of the `ec:energyLabel` property to the EU branch of the energy label vocabulary.

The controlled vocabulary on measures to increase the energy performance of buildings has 22 new terms spread over four levels. Details of the terms are still being worked out by our partner , ADENE. Figure 4 shows the tree structure of the vocabulary.

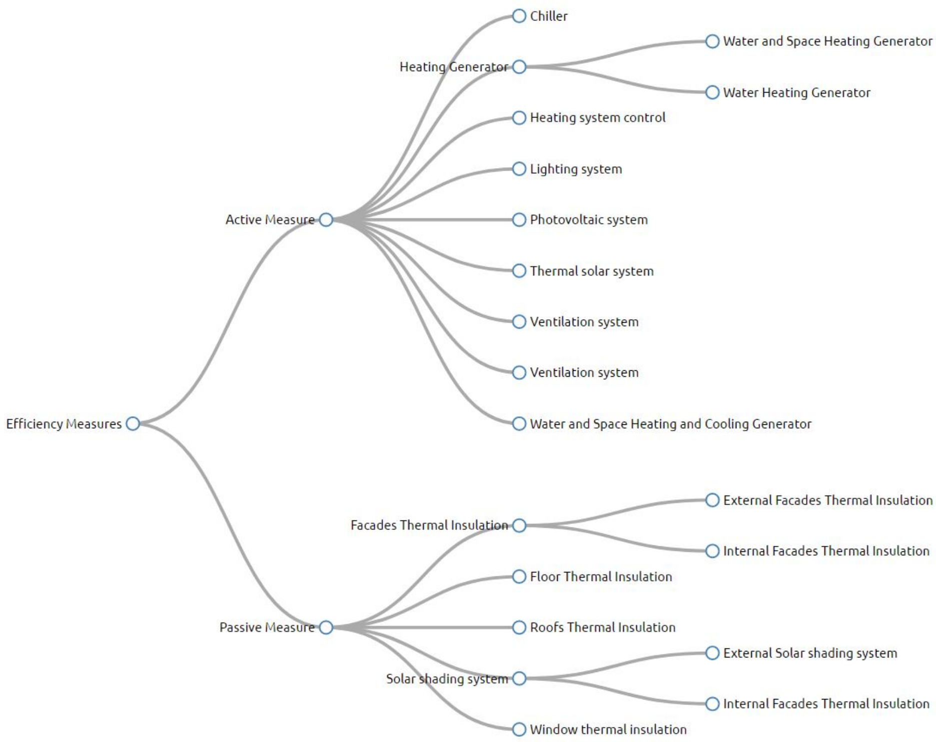


FIG. 4 - Tree structure of the vocabulary of measures for energy performance.

These controlled vocabularies are available at http://opendata.dsi.uminho.pt/terms/energy/.

## 4. Closing remarks and future work

The design of application profiles often involves several kinds of trade-offs. In many cases the choice is between levels of interoperability, and it is not uncommon to consider losing some semantic rigor in favour of the use of a property more used by the metadata community, potentially gaining interoperability. Or rather: to dispense with some interoperability in favour of semantic rigor. The option is not always easy and is frequently subject to controversy. In some cases we just have to give away something. The issue is: what can we afford to give away?

The SusCity project is an MIT Portugal project within the scope of smart cities. One of its tasks has to do with opening data generated in its scope, in particular data about energy consumption, energy performance of buildings and urban mobility. For this data to be opened as LOD and to maximise its interoperability, it is necessary to create MAPs. This article reported the process and the results of the development of a MAP for energy performance certificates of buildings that are issued by ADENE, a co-funder and partner of the project. The development of this MAP implied the development of a few more metadata artefacts: a metadata schema for the declaration of four new properties and a class, and four controlled vocabularies, three of which serve as constraints to the range of some properties. The fourth controlled vocabulary, on measures to improve the energy performance of buildings, was encoded in SKOS on the request of ADENE.

These developments have forced some design decisions that are relevant to the metadata community. We chose to have only one class (the `EnergyPerformanceCertificate`), which is the domain of all properties of the MAP. The use of one class results from a balance between power of description and simplicity: in order to facilitate the use and reuse of data, we have given up some power of description in exchange for more simplicity. Another decision we made has to do with rigour in the choice of properties. We consider imperative to make all efforts to respect the constraints defined in the schemas of the properties and this is what we did. However, there were cases where we had to leave something behind, as it happened with the compatibility of ranges. In some situations, our option was to dispense compatibility of ranges to be able to cope with the other criteria and use just one property per attribute. Another design option relates to the selection of properties when more that one respected our needs. In the face of two or more properties, which one should we choose? Our choice fell on the following priority criteria: 1) property in a schema corresponding to some standard or some recommendation of some relevant entity in the field of application (energy performance) or in LOD; 2) property widely used by the metadata community; or, finally, 3) property defined in a formal schema and used by at least two datasets. We only created new properties when none of these options were possible.

The creation of the metadata schema and controlled vocabularies involved the use of Wikidata, so all new terms (RDFS classes and properties and SKOS concepts) are related to Wikidata terms through specific OWL properties. The use of Wikidata to "declare" the terms was another design decision that we find relevant for evolution/scrutinity, availability and interoperability purposes. The RDFS and SKOS files also hold the vocabularies structure, leaving all created Wikidata terms free of any constraints. This facilitates the reuse of terms across different schemas while safeguarding minimal semantics.

Besides being available as RDFS and SKOS files, the vocabularies will also be stored in a triplestore together with other SusCity RDF data on energy performance and mobility. Future work includes the finalization of this process with the inclusion of the metadata schema in LOV and the design of new application profiles regarding energy consumption and mobility.

## Acknowledgements

We would like to thank the following MSc students who participated in the prior study referred in the Introduction of this article: Diana Postolaki, Henrique Santos Guimarães and Ivo Filipe Marques Menezes.

We would like to thank Joana Fernandes (ADENE) for her interest, for the fruitful discussions we had, and for making available the human-readable version of the controlled vocabulary for measures to improve the energy performance of buildings.

Lastly, we would like to thank the reviewers of this article who, with their interest and commitment, contributed to the clarification of several aspects of the article and improvement of others.

This work has been supported by FCT – Fundação para a Ciência e Tecnologia in the scope of the project FCT/MITP-TB/CS/0026/2013.

## Namespaces

dct - http://purl.org/dc/terms/

dbpedia-owl - http://dbpedia.org/ontology/

ec - http://opendata.dsi.uminho.pt/terms/energy/EPCterms/

osadm - http://data.ordnancesurvey.co.uk/ontology/admingeo/

owl - http://www.w3.org/2002/07/owl#

rdf - http://www.w3.org/1999/02/22-rdf-syntax-ns#

rdfs - http://www.w3.org/2000/01/rdf-schema#

skos - http://www.w3.org/2004/02/skos/core#

vcard - http://www.w3.org/2006/vcard/

xsd: http://www.w3.org/2001/XMLSchema#

## References

ADENE. (2017). Quem somos | ADENE - Agência para a Energia. Retrieved 14 June 2017, from http://www.adene.pt/quem-somos

Aelenei, L., Ferreira, A., Monteiro, C. S., Gomes, R., Gonçalves, H., Camelo, S., & Silva, C. (2016). Smart City: A Systematic Approach towards a Sustainable Urban Transformation. *Proceedings of the 4th International Conference on Solar Heating and Cooling for Buildings and Industry (SHC 2015)*, *91*(Supplement C), 970–979. https://doi.org/10.1016/j.egypro.2016.06.264

Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart Cities in Europe. *Journal of Urban Technology*, *18*(2), 65–82. https://doi.org/10.1080/10630732.2011.601117

Costa, C., & Santos, M. Y. (2015). Improving cities sustainability through the use of data mining in a context of big city data. In *The 2015 International Conference of Data Mining and Knowledge Engineering* (Vol. 1, pp. 320–325). IAENG.

Coyle, K., & Baker, T. (2009, May 18). Guidelines for Dublin Core Application Profiles. Retrieved 28 June 2010, from http://dublincore.org/documents/2009/05/18/profile-guidelines/

Curado Malta, M., & Baptista, A. A. (2013). A Method for the Development of Dublin Core Application Profiles (Me4DCAP V0.2): detailed description. In K. Eckert & M. Foulonneau (Eds.) (pp. 90–103). Lisbon, Portugal: Dublin Core Metadata Initiative. Retrieved from http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/178/81

European Parliament and the Council of the European Union. Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings, Official Journal L 001 , 04/01/2003 P. 0065 - 0071 § (2003).

MIT Portugal. (2017). Retrieved 14 June 2017, from https://www.mitportugal.org/

Odyssee-Mure. (2012). *Energy Efficiency Trends in Buildings in the EU Lessons from the ODYSSEE MURE project*. Retrieved from https://www.nahb.org/ /media/Sites/NAHB/LMA/FileUploads/35601-1-REPORT_local_20150318115955.ashx?la=en

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings*, *40*(3), 394–398. https://doi.org/10.1016/j.enbuild.2007.03.007

Sozer, H. (2010). Improving energy efficiency through the design of the building envelope. *Building and Environment*, *45*(12), 2581–2593. https://doi.org/10.1016/j.buildenv.2010.05.004

SUSCITY. (2016). SUSCITY – An MIT Portugal project. Retrieved 14 June 2017, from http://groups.ist.utl.pt/suscity-project/home/

Vandenbussche, P.-Y., Atemezing, G. A., Poveda-Villalón, M., & Vatant, B. (2017). Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. *Semantic Web*, *8*(3), 437–452. https://doi.org/10.3233/SW-160213