Chapter 5: Validation of the BCH-ontology

In this chapter the BCH-ontology is validated through use cases where data provided by the stakeholders are translated into semantic means and are linked so that some basic queries can be executed [Section 5.1.1-5.1.3]. Further applications are then explored by workshops developed with the stakeholders [Section 5.1.4]. The general advantages of using ontologies are then discussed with examples from the BCH-ontology [Section 5.2]. The final part of the chapter focuses on the necessary considerations to use BCH-ontologies within the Cuenca stakeholders.

5.1 Introduction

In the following sections we illustrate how the steps of the knowledge process lead to the validation of the BCH-ontology. This use case's aim is to answer the competency questions set in the ontology requirements specification document (Table 4.1). Building block 0801007 (Fig. 5.1) in the San Roque neighborhood has been selected as study area; it is composed of 28 properties for which information is available from different stakeholders.

Fig. 5.1 Study area, building block 0801007 composed of 28 properties for which information is available from different stakeholders.

Table 5.1 summarizes the available data sources for the study area.

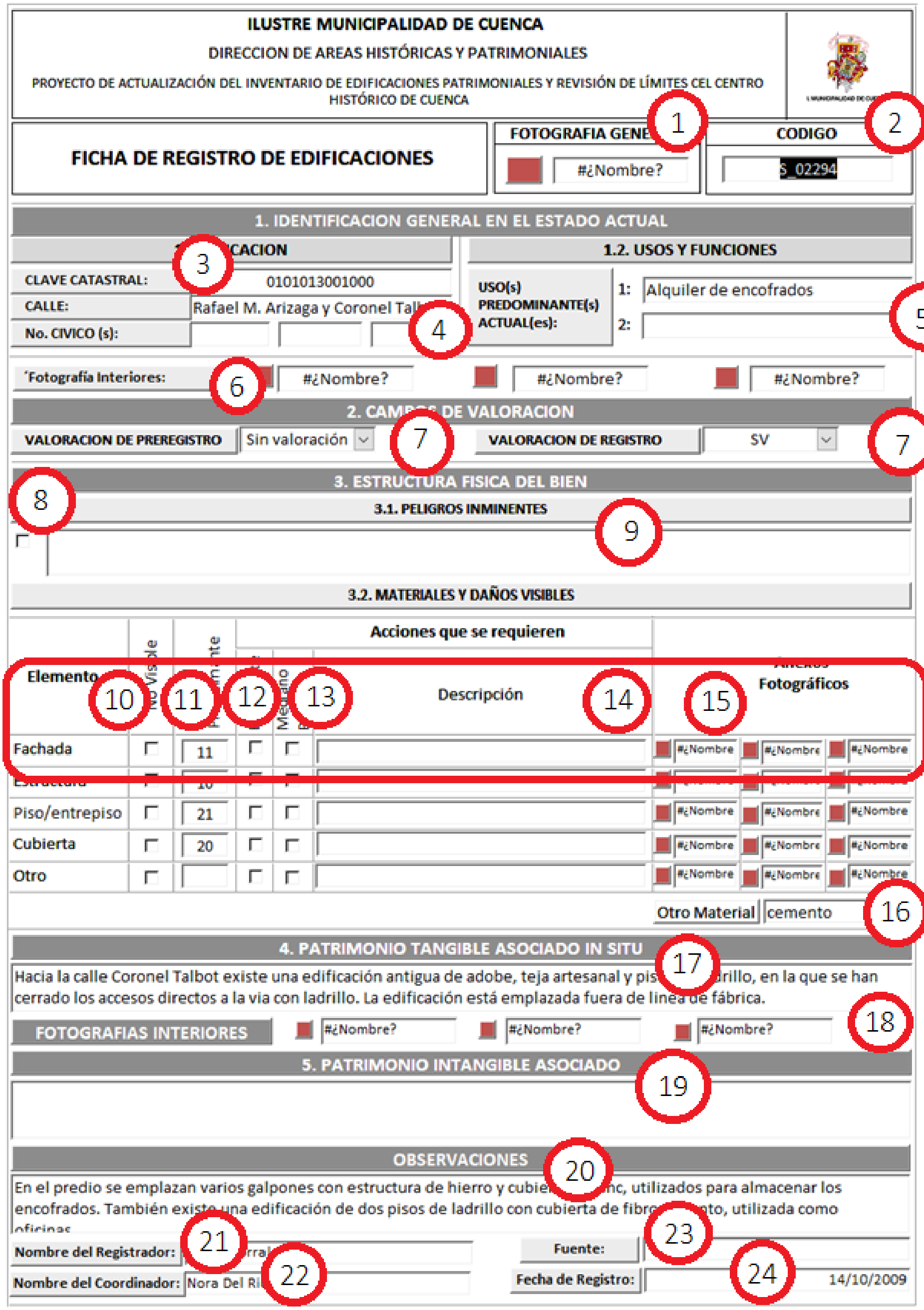
Stakeholder	Data	Description
Municipality of Cuenca	Inventory 2010; Microsoft Access database	This is the last official registry of historic buildings
Municipality of Cuenca	Building shapefiles (plots)	2d plots of the properties where the buildings are located
CPM research group - San Roque maintenance campaign	Workbook; Microsoft Access database	In 2014 a maintenance campaign was carried out in a block of the San Roque neighborhood so it is a good example to obtain data on the activities carried out during the campaign
CPM research group - San Roque maintenance campaign	Elements shapefiles	2d drawings of each of the floors
INPC	Abaco inmueble dataset; Microsoft Excel	It contains the location of the buildings by means of coordinates (points), architectural style, etc

5.2 Creation/import

In the creation or import phase, stakeholder information has to be migrated to a semantic environment. According to Table 5.1 there are 2 types of data: alphanumeric information (Microsoft Access, Excel) and spatial information (shapefiles).

5.2.1 Migration of alphanumeric data

Alphanumeric data are mapped using BCH-ontology classes and properties. As an example Fig. 5.2 and Table 5.2 shows this mapping for the inventory database of the municipality of Cuenca.



ILUSTRE MUNICIPALIDAD DE CUENCA
DIRECCION DE AREAS HISTÓRICAS Y PATRIMONIALES
PROYECTO DE ACTUALIZACIÓN DEL INVENTARIO DE EDIFICACIONES PATRIMONIALES Y REVISIÓN DE LÍMITES CEL CENTRO HISTÓRICO DE CUENCA

FICHA DE REGISTRO DE EDIFICACIONES

FOTOGRAFIA GENE[RAL] 1 #¿Nombre?
CODIGO 2 S_02294

1. IDENTIFICACION GENERAL EN EL ESTADO ACTUAL

1.1. [UBI]CACION 3

CLAVE CATASTRAL: 0101013001000
CALLE: Rafael M. Arizaga y Coronel Tal[bot] 4
No. CIVICO (s):

1.2. USOS Y FUNCIONES

USO(s) PREDOMINANTE(s) ACTUAL(es): 1: Alquiler de encofrados 5
2:

Fotografía Interiores: 6 #¿Nombre? #¿Nombre? #¿Nombre?

2. CAMP[OS] DE VALORACION 7

VALORACION DE PREREGISTRO: Sin valoración
VALORACION DE REGISTRO: SV 7

3. ESTRUCTURA FISICA DEL BIEN

8

3.1. PELIGROS INMINENTES 9

3.2. MATERIALES Y DAÑOS VISIBLES

Elemento 10	No Visible	Pr[edominante] 11	12	Mediano 13	Descripción 14	Anexos Fotográficos 15		
Fachada		11				#¿Nombre	#¿Nombre	#¿Nombre
Estructura		10						
Piso/entrepiso		21				#¿Nombre	#¿Nombre	#¿Nombre
Cubierta		20				#¿Nombre	#¿Nombre	#¿Nombre
Otro						#¿Nombre	#¿Nombre	#¿Nombre

Acciones que se requieren

Otro Material: cemento 16

4. PATRIMONIO TANGIBLE ASOCIADO IN SITU 17

Hacia la calle Coronel Talbot existe una edificación antigua de adobe, teja artesanal y pis[o de la]drillo, en la que se han cerrado los accesos directos a la via con ladrillo. La edificación está emplazada fuera de linea de fábrica.

FOTOGRAFIAS INTERIORES #¿Nombre? #¿Nombre? #¿Nombre? 18

5. PATRIMONIO INTANGIBLE ASOCIADO 19

OBSERVACIONES 20

En el predio se emplazan varios galpones con estructura de hierro y cubie[rta de zi]nc, utilizados para almacenar los encofrados. También existe una edificación de dos pisos de ladrillo con cubierta de fibro[cemen]to, utilizada como oficinas.

Nombre del Registrador: 21 [...]rra[...]
Nombre del Coordinador: Nora Del Ri[o] 22
Fuente: 23
Fecha de Registro: 24 14/10/2009

Fig. 5.2 2010 Inventory. Each section is enumerated to be further linked with BCH-ontology classes and properties.

Table 5.1 Inventory 2010; properties and classes of the BCH-ontology with which the sections in Fig. 5.2 can be represented.

Nro.	**Municipality Dataset**	**BCH-ontology property**	**BCH-ontology class**
1	Foto general	P138 represents	E38 Image
2	Codigo	P1 is identified by	E42 Identifier
3	Clave Catastral	P1 is identified by	E42 Identifier
4	calle num_civ_1 num_civ_2 num_civ_3	P1 is identified by	E45 Address
5	Uso 1 Uso 2	P140 assigned attribute to P141 assigned	HB58 Usage HB59 Usage Assignment
6	Foto interior	P138 represents	E38 Image
7	Valoracion preregistro Valoracion registro	P140 assigned attribute to P141 assigned	HB25 Heritage Value
8	Peligro	P140 assigned attribute to P141 assigned	M7 Hazard
9	Peligro descripción	rdfs:comment	String
10	Fachada visible	rdfs:comment	String
11	Fachada material	P45 consist of	E57 Material
12	Fachada emergente	rdf:type	HB50 ActionType
13	Fachada mediano plazo	rdf:type	HB50 ActionType
14	Fachada descrip	rdfs:comment	String
15	Foto fachada	P138 represents	E38 Image
16	Otro Material	P45 consist of	E57 Material
17	Tangible	P129 is about	E26 Physical Feature
18	Foto tangible	P138 represents	E38 Image
19	Intangible	P129 is about	HB29 Conceptual Feature
20	Observaciones	rdfs:comment	String
21	Registrador	P14 carried out by P14.1 in the role of P107 has current or former member	E7 Activity E21 Person E55 Type E40 Legal Body
22	Coordinador	P11 had participant P14.1 in the role of	E21 Person E55 Type
23	Fuente	P11 had participant P14.1 in the role of	E21 Person E55 Type
24	Fecha	P4 has time span	E52 Time Span

The address is stored as a composite attribute combining the street and the identification number. In case the building has several numbers, an equal number of addresses are registered. Descriptions can be documented with the 'P3 has note' property of the CIDOC-CRM ontology or in general with 'rdfs:comment.' In our case we selected the second option as it will be easier for those who are not familiar with CIDOC-CRM.

Photos taken with each inventory register were not provided to us so these fields have been ignored, however if they exist they can be documented as indicated in Table 5.2, row 1.

Rows 10-15 are repeated for each building element. The inventory form is poorly designed since other materials are recorded but are not always linked to any element, in which case they can be associated with the building in general. Additionally, other elements of interest are recorded, but the description of each one must be processed manually to know which element is being discussed. However, an element whose type is unknown can be created to associate this information to.

The mapping table (Table 5.2) is undoubtedly a good guide to start mapping. The mapping tool OntoRefine integrated in GraphDB has been selected to make the necessary transformations; it allows direct connections with PostgresSQL, MySql and MariaDB databases. Additionally the tool allows mappings on structured data such as TSV, CSV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. Other formats can be added with OpenRefine extensions. In our case it is simple to export the Microsoft Access database to Microsoft Excel.

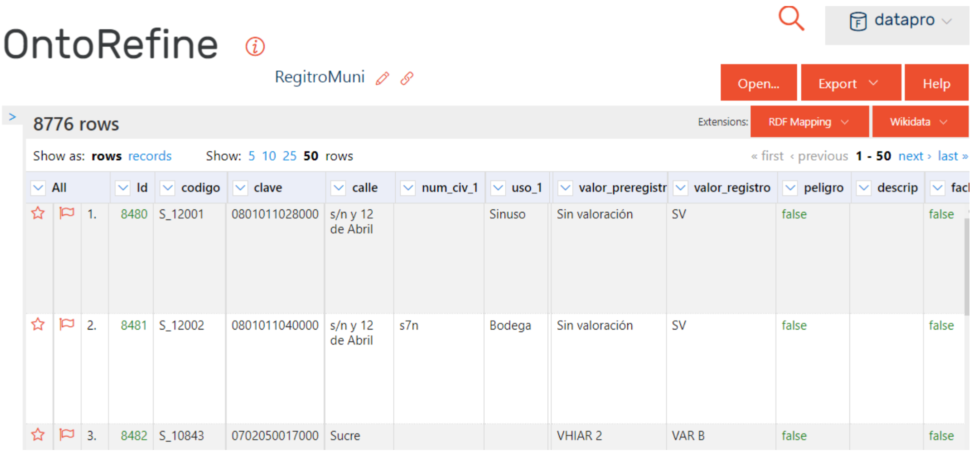


Fig. 5.3 Inventory table in Graph DB; where data can be transformed individually or in groups.

Although all the data can be mapped, in our experience it is important to do so in an incremental way to facilitate error control. We start with the creation of the conceptual building and the inventory activity to which all other information will be linked. The "RDF Mapping" option allows us to create mappings in a graphical way, as shown in Fig. 5.4.

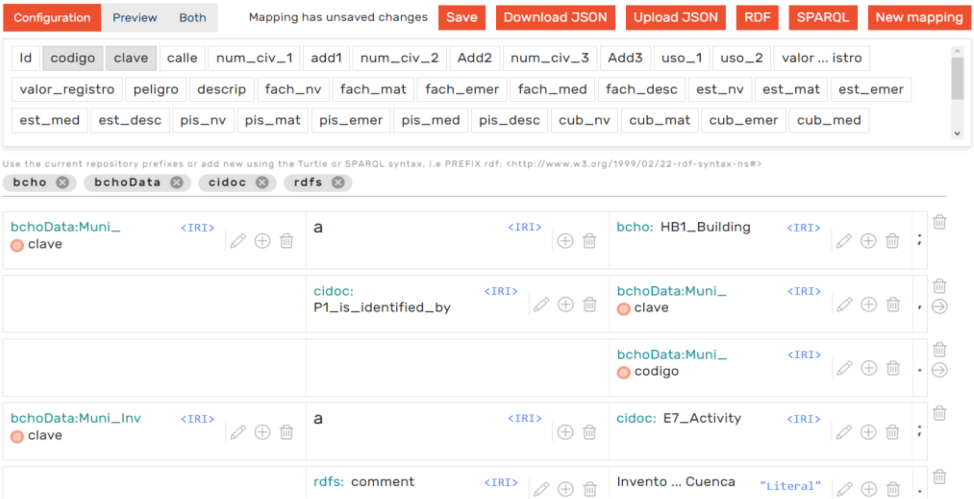


Fig. 5.4 Graph DB triplets of building, identifiers and activity.

In this case (Fig. 5.4), 5 triplets are created. The first one creates a 'building', the second and third ones assign identifiers (Table 5.2 lines 2,3), the fourth one creates an inventory activity (Table 5.2 line 24) and the fifth one describes this activity. The mapping can be stored in different formats, RDF is selected and a file is created (Fig. 5.13) with the 5 triplets described above for each row of the inventory table.

```
bchoData:Muni_0801011028000 a bcho:HB1_Building;
  cidoc:P1_is_identified_by bchoData:Muni_0801011028000, bchoData:Muni_S_12001 .

bchoData:Muni_Inv0801011028000 a cidoc:E7_Activity;
  rdfs:comment "Inventory 2010 by the Municipality of Cuenca" .

bchoData:Muni_0801011040000 a bcho:HB1_Building;
  cidoc:P1_is_identified_by bchoData:Muni_0801011040000, bchoData:Muni_S_12002 .

bchoData:Muni_Inv0801011040000 a cidoc:E7_Activity;
  rdfs:comment "Inventory 2010 by the Municipality of Cuenca" .

bchoData:Muni_0702050017000 a bcho:HB1_Building;
  cidoc:P1_is_identified_by bchoData:Muni_0702050017000, bchoData:Muni_S_10843 .

bchoData:Muni_Inv0702050017000 a cidoc:E7_Activity;
  rdfs:comment "Inventory 2010 by the Municipality of Cuenca" .
```

Fig. 5.5 RDF file generated after using Graph DB's OntoRefine mapping tool.

Additionally new fields can be created to perform more complex transformations, e.g. the address which is a compound field. In this case (Fig. 5.6), a new column is created to save the address and the same process is performed when it has more than one building number.

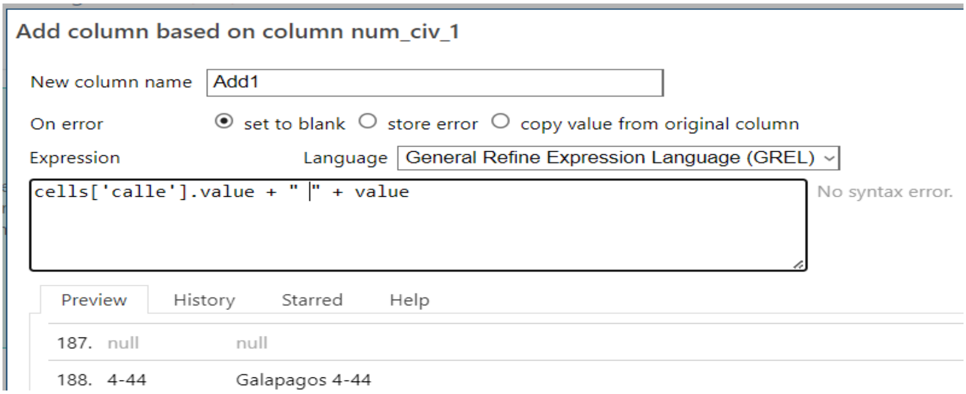


Fig. 5.6 Column transformations of Fig. 5.3 where the address is concatenated.

For some mappings it is also necessary to have a base list, such as for materials, heritage values, uses, registrants, coordinators, etc. Fig. 5.13 shows different variants of use 'Housing' that have been registered (left column), which have been normalized using a base list made by the CPM project (middle column) and transformed to RDF (right column). Double uses such as 'housing - carpentry' are registered as two uses 'housing' and 'carpentry'. In this way, several RDF files are created that can later be processed to perform queries on the data.

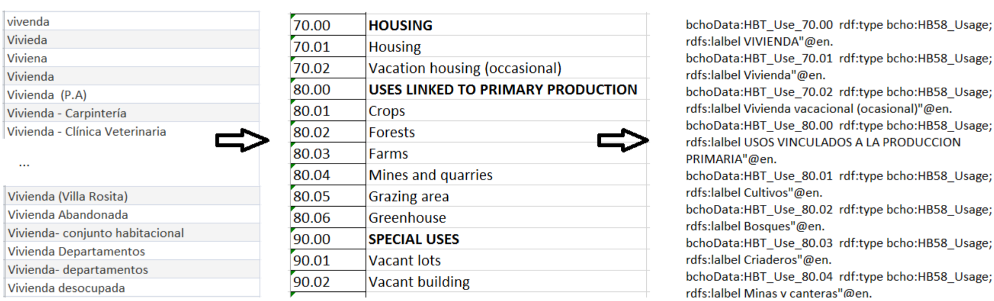
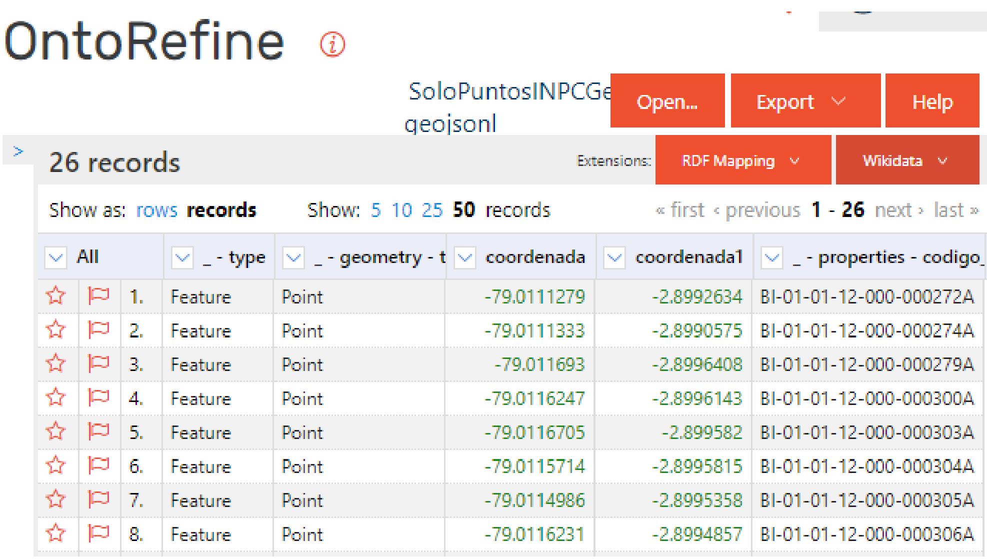


Fig. 5.7 Variants of ‘housing’ use. The column in the left shows different variants of use 'Housing' that have been registered, which have been normalized using a base list made by the CPM research group(middle) and transformed to RDF (right).

5.2.2 Migration of geospatial data

In addition to the alphanumeric data, there is geospatial information. In this case polygons of properties and elements in each floor are already available from the Municipality of Cuenca (Table 5.1, row 4).

The INPC provided an excel document with information about the buildings (Table 5.1, row 5). Most columns are mapped like the examples in figures 5.3-5.7, thus the focus here is on the geospatial information stored in the 'coordinate' and 'coordinate_1' columns which represent the coordinates of the point where the building is located using UTM - 17S base on WGS84. The Municipality of Cuenca uses a polygon to represent the building parcel. In order to export the polygons to OntoRefine, the shapefile is exported as GeoJson file which uses geographic coordinates WGS84 and the units are degrees (Fig. 5.8). For this reason the excel file is also opened in QGIS and exported to GeoJson, so that both spatial objects (points, polygons) employ the same units.


OntoRefine

SoloPuntosINPCGe geojsonl

Open... Export Help

26 records

Extensions: RDF Mapping Wikidata

Show as: rows **records** Show: 5 10 25 **50** records « first ‹ previous **1 - 26** next › last »

All	_ - type	_ - geometry - t	coordenada	coordenada1	_ - properties - codigo_
1.	Feature	Point	-79.0111279	-2.8992634	BI-01-01-12-000-000272A
2.	Feature	Point	-79.0111333	-2.8990575	BI-01-01-12-000-000274A
3.	Feature	Point	-79.011693	-2.8996408	BI-01-01-12-000-000279A
4.	Feature	Point	-79.0116247	-2.8996143	BI-01-01-12-000-000300A
5.	Feature	Point	-79.0116705	-2.899582	BI-01-01-12-000-000303A
6.	Feature	Point	-79.0115714	-2.8995815	BI-01-01-12-000-000304A
7.	Feature	Point	-79.0114986	-2.8995358	BI-01-01-12-000-000305A
8.	Feature	Point	-79.0116231	-2.8994857	BI-01-01-12-000-000306A

Fig. 5.8 GeoJson file with the information of the INPC buildings' location is opened with OntoRefine.

Fig. 5.9 shows the creation of a new column with the point representation in the required RDF format. Fig. 5.10 shows the RDF file produced after the mapping of this information is linked to the INPC building.

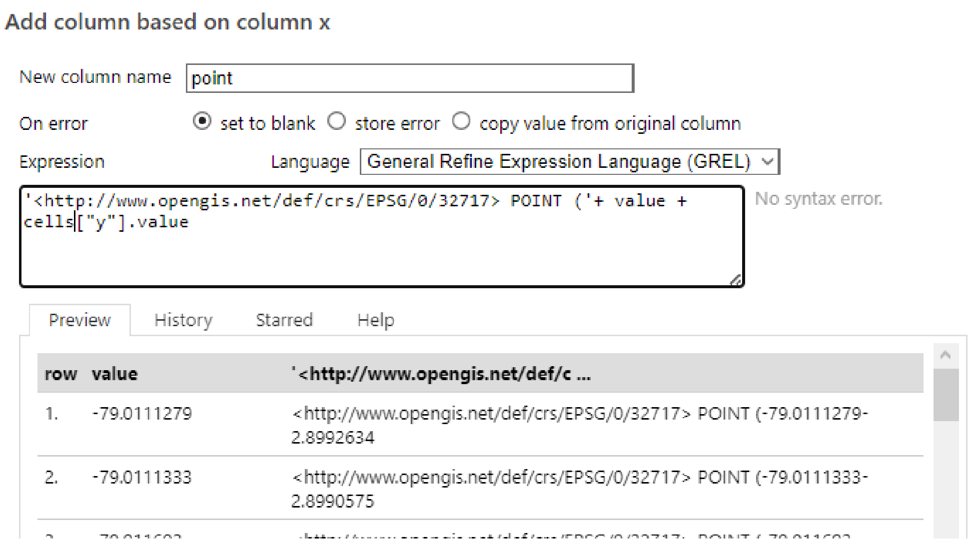


Fig. 5.9 Creation of a new column with the point representation in RDF format.

```
bchoData:Inpc_BI-01-01-12-000-000272A a bcho:HB1_Building;
  geo:hasGeometry bchoData:GeomInpc_BI-01-01-12-000-000272A .

bchoData:GeomInpc_BI-01-01-12-000-000272A geo:asWKT
"<http://www.opengis.net/def/crs/EPSG/0/32717> POINT (-79.0111279, -
2.8992634)"^^geo:wktLiteral .

bchoData:Inpc_BI-01-01-12-000-000274A a bcho:HB1_Building;
  geo:hasGeometry bchoData:GeomInpc_BI-01-01-12-000-000274A .

bchoData:GeomInpc_BI-01-01-12-000-000274A geo:asWKT
"<http://www.opengis.net/def/crs/EPSG/0/32717> POINT (-79.0111333, -
2.8990575)"^^geo:wktLiteral .
```

Fig. 5.10 RDF mapping of the mapping in Fig. 5.9.

Although OntoRefine provides compatibility with Json files, the spatial attributes of polygons are stored as separate rows which are of variable size, since some polygons can have more vertexes than others. OntoRefine offers the 'Join multi-valued cells' function (Fig. 5.11) which joins the coordinates with a specified character, in this case ','. However since vertexes are represented with two coordinates, some additional transformations are required (Fig. 5.12). Also the column was renamed to 'coordinates' to improve readability (Fig. 5.13).

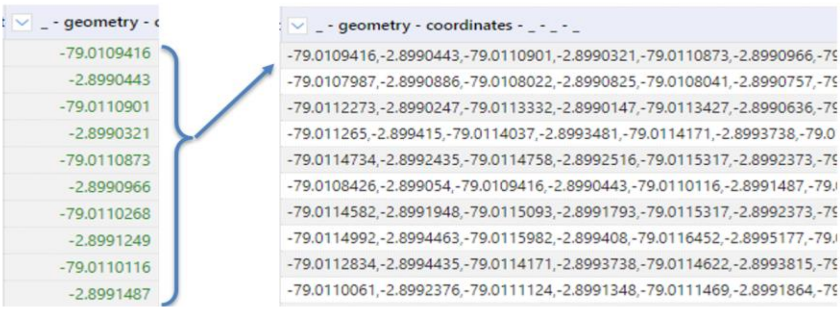


Fig. 5.11 OntoRefine transformation of multiple values.

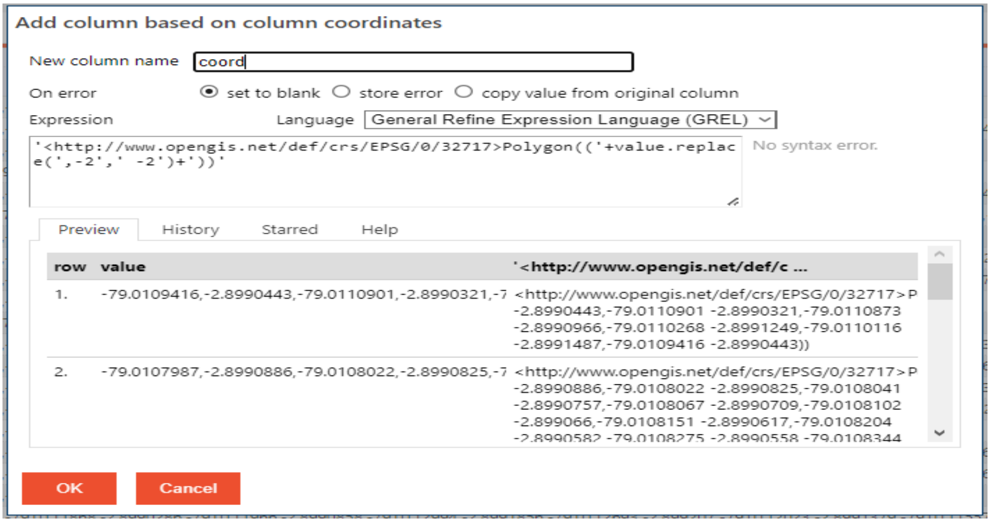


Fig. 5.12 New column with the final transformation to represents polygons in RDF.

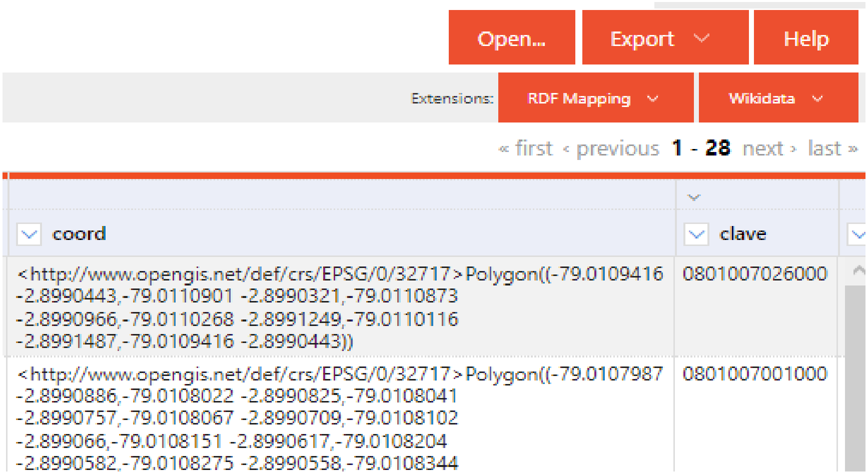


Fig. 5.13 Final file with the transformations in OntoRefine.

The column 'clave' contains the cadastral key and column 'cood' the coordinates of the polygon. Fig. 5.11 shows the mapping and the RDF file obtained.

Fig. 5.14 Mapping and the resulting RDF file of the Fig. 5.13.

Finally, several RDF files have been created containing both alphanumeric and geospatial information of the stakeholders, which is imported by means of the Import/RDF file option to a previously created repository (Annex IV).

5.3 Knowledge capture step

Once knowledge items are created, the next step is to capture their essential contents by linking the knowledge previously created. The data sets created above are not related to each other. From the Municipality there are 28 records, from the INPC 26 and from the CPM research group there are 28 more records, thus a query about the number of buildings will return 82 records. A visual query of the records information shows that they are not linked to each other (Fig. 5.15).

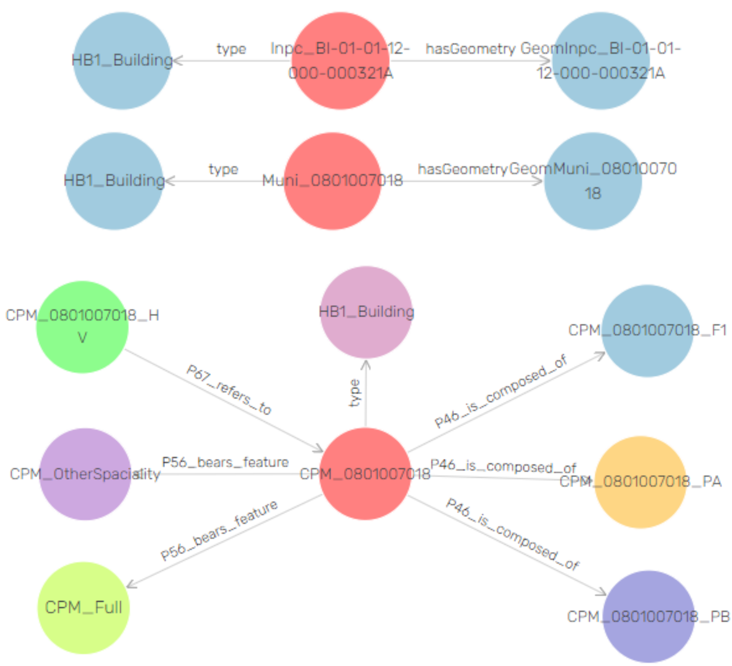


Fig. 5.15 Shows the 3 records of the building with cadastral key 0801007018 unlinked.

If the stakeholders use the same identifier, the records can be linked. The owl:sameAs clause allows us to link two records. For this use case the name bchoData:Muni_0801011028000 is used to identify a property of the municipality and bchoData:CPM_0801011028000 to identify a property of the CPM research group. The triplet "bchoData:CPM_080100707028000 owl:sameAs bchoData:Muni_080100707028000" indicates that it is the same building.

However, there are not always known identifiers that can be automatically linked. The Municipality of Cuenca manages a cadastral code while the INPC manages its own codes. In this case the geospatial properties are used to determine which records should be linked. Fig. 5.16 show the location of buildings of the INPC and the Municipality of Cuenca.

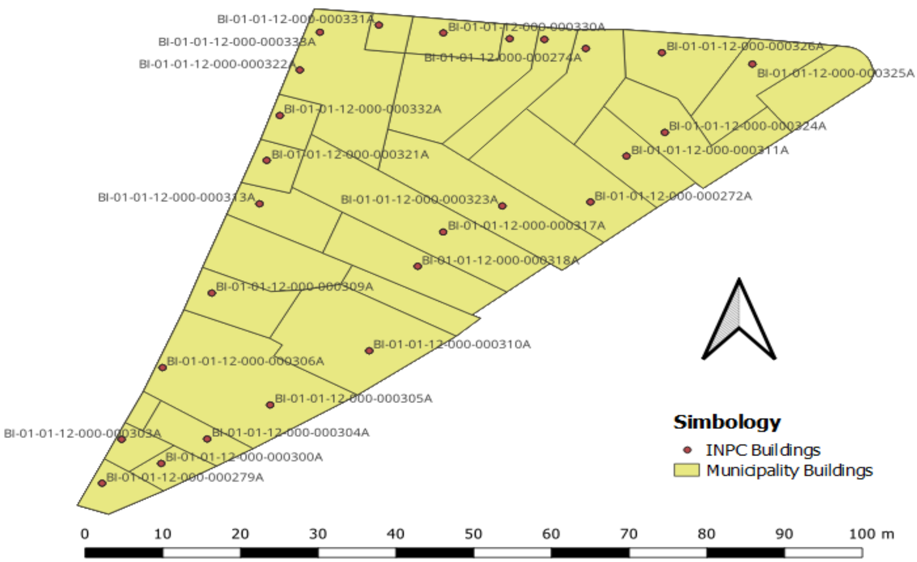


Fig. 5.16 Location of buildings of the INPC and the Municipality of Cuenca.

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
INSERT { ?m owl:sameAs ?n }
WHERE {
?m geo:hasGeometry ?mgeo .
?mgeo geo:asWKT ?p.
?n geo:hasGeometry ?ngeo .
?ngeo geo:asWKT ?q.
FILTER (geo:sfContains( ?m, ?n) )
}
```

Fig. 5.17 SPARQL query that integrates the records of the INPC and the Municipality of Cuenca.

Fig. 5.17 shows a SPARQL query that identifies the point contained in the polygons and linked them with the predicade ‘owl:sameAs’. After linking we can query the information of a particular building and we see that the information of the other stakeholders is also linked.

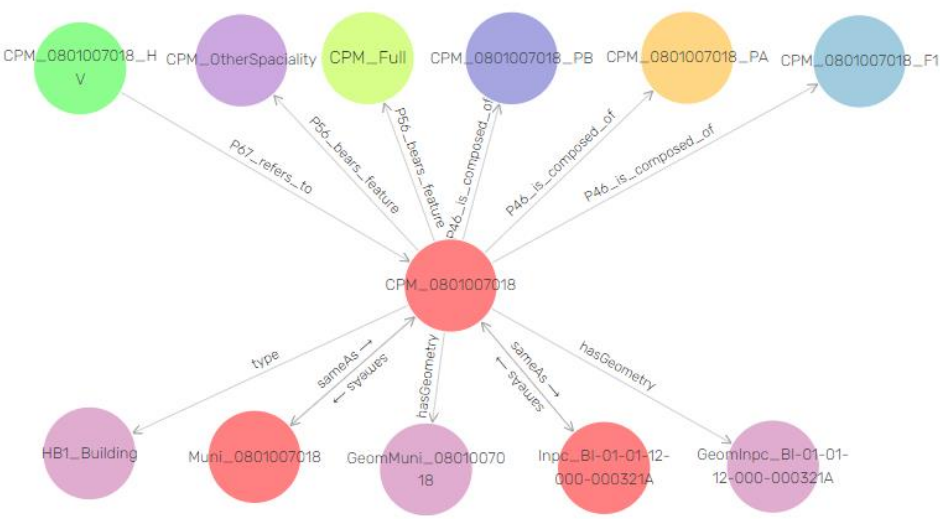


Fig. 5.18 Linked records, at the top the information of the CPM research group. In the lower part the information of the INPC and the Municipality of Cuenca.

The integration made in Fig. 5.18 is relevant because it allows connection with the complete datasets from the 3 stakeholders. On traditional systems as databases when a 'join' is performed just the information in the two tables that are being joined is available. However in this case all the information related to the building is available without restrictions of which data source comes from.

5.4 Knowledge retrieval and access step

Information stored in the ontology is accessed satisfying simple requests of the user requirements. A use case querying and searching general information regarding BCH assets was developed. In accordance with the requirements listed in Table 4.1, the following competency questions are defined and have been applied in the study area.

CQ1: What are the most common damages in historical buildings?

To answer this question the information from the maintenance campaign carried out by the CPM research group in 2014 (Table 5.1, row 3) is used. A damage record was created by means of a Microsoft Access file (Fig. 5.19).

Clave Pred	Nivel de edificació	Factor de	elemen	Código	Estado	Materia	Materia
0801007003	1ra P. Subsuelo	Estabilidad	Entrepiso	SO_03	Malo	27. Madera	
0801007003	P. Baja	Estabilidad	Muros portante:	PB_09	Regular	1. Adobe	16. Empañ
0801007003	2da P. Alta	Estabilidad	Estructura	2PA_03	Regular	27. Mader	17. Enchar

Daño 1	Daño 2	Causa 1	cod_ca	Magnitu	Calas de	Descrip	Fotogra	Observa
2.1.5		F1		Alto (67 - 1	FALSE	V003_02	F003_02 / F003_03	
2.1.2.		M19		Medio (34	FALSE	V003_03		
2.2.3.		F1	03	Medio (34	TRUE	V003_04	F003_05	

Fig. 5.19 Damage record of the San Roque maintenance campaign.

The damage record identifies the building, the floor, the room, and the element in which a damage is found, as well as the magnitude and possible causes that led to the occurrence of the damage. The first competency question can easily be answered with a query that counts the number of times a damage has been recorded in the DB. However, since there is more specific information such as magnitude, which is classified as High, Medium or Low, high magnitude damages are query. The results are presented in Fig. 5.20; at this point the ontology does not present any advantage in comparison with traditional systems since the same query can be performed in similar terms using SQL.

```
select ?DamLabel  (COUNT(*) AS ?count)
{
    ?s a mondisCore:ManifestationOfDamage.
    ?s cidoc:P46_is_composed_of ?DamType.
    ?x rdf:type cidoc:E13_Attribute_Assignment.
    ?x cidoc:p141_assigned bchoData:HBT_CPM_DamMagnitud1.
    ?x cidoc:P140_assigned_attribute_to ?s.
    ?DamType rdfs:label ?DamLabel.
}
group by ?DamLabel
order by desc (?count)
```

Run

	DamLabel	count
1	"Exploration craks"@en	"5"^^xsd:integer
2	"Detachment (fragments)"@en	"4"^^xsd:integer
3	"Stains"@en	"4"^^xsd:integer

Fig. 5.20 SPARQL query of damages of high magnitude recorded in the San Roque maintenance campaign.

CQ2: Where are the most common damages located (Diagnosis phase)?

The second competency question refers to the location of these data. This question could be answered in a similar way to the previous one, using only the textual fields of floor room and element. However, in the same maintenance campaign, drawings of the buildings were made by floor (Fig. 5.21), which were not linked to the damage.

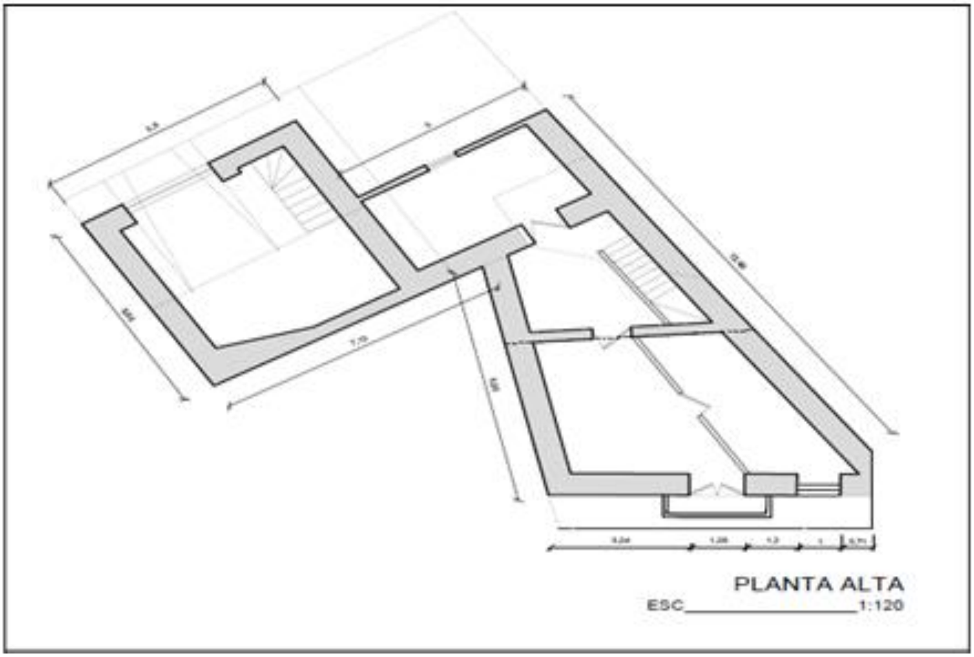


Fig. 5.21 Drawings of floor by building.

For this question the integration of the different datasets is exploited to link the damage to the geospatial element. The greatest difficulty was to link the elements of the damage card with the polygon shapefile elements because they were at different levels. In some cases the element in which the damage was recorded was not present in the shapefile, for example the ceilings. In this case the damage is linked to the next container element higher in hierarchy, in this case the room in which the damage is located. Finally, the damage can also be located by building. For all these queries the building is handled as a single entity and its attributes are queried in a general way regardless of the dataset it comes from. Fig. 5.22 shows the SPARQL query performed to locate the damages by building.

```
prefix cidoc: <http://www.cidoc-crm.org/cidoc-crm/#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX mondisCore: <http://kbss.felk.cvut.cz/ontologies/2011/monument-damage-core.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bchoData: <http://www.ciudadpatrimoniomundial.com/ontologias/BCH/Data/>
PREFIX bcho: <http://www.ciudadpatrimoniomundial.c1om/ontologias/BCH#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
select distinct ?BuildingRoom ?BuildingCompType ?DamageCode ?DamageName ?Geom  {
?BuildingDam a mondisCore:ManifestationOfDamage;
             cidoc:P46_is_composed_of ?DamageCode.
?DamageCode rdfs:label ?DamageName.
?x rdf:type cidoc:E13_Attribute_Assignment.
?x cidoc:p141_assigned bchoData:HBT_CPM_DamMagnitud1.
?x cidoc:P140_assigned_attribute_to ?BuildingDam.
?BuildingComp cidoc:P56_bears_feature ?BuildingDam.
?BuildingRoom cidoc:P46_is_composed_of ?BuildingComp.
?BuildingComp rdf:type mondisCore:Component; cidoc:P2_has_type ?BuildingCompType.
?BuildingFloor cidoc:P46_is_composed_of ?BuildingRoom.
?Building cidoc:P46_is_composed_of ?BuildingFloor.
?Building geo:hasGeometry ?GeomName.
?GeomName geo:asWKT ?Geom.
}
```

Fig. 5.22 SPARQL query performed to locate the damages by building.

The results obtained can be stored as a GeoJason file and opened directly in Qgis where the results can be visually explored. This is very useful information for the control technicians and shows the direct collaboration between academic institutions and municipality. Fig. 5.23 shows the results of damage location at different levels. Annexes V, VI and VII show the complete queries. Competency question 5 can be answered in a similar way and the complete query and results are shown in Annex (VIII).

Fig. 5.23 Location of damages by element, room and building.

This use case is the one that shows the advantages of the BCH-ontology. After the integration of Fig. 5.18 the information is treated indistinctly of where it comes from. Damages from the maintenance campaign (CPM-Microsoft Access) is shown in QGIS liked to the element, the room (CPM-shapefiles) or the building (Municipality of Cuenca, parcels shapefile).