



*Centre adscrit a la*



**Grado en Ingeniería Informática de Gestión y Sistemas de Información**

**Reingeniería del modelo y de las fuentes de datos para un sistema de computación  
urbana  
Memoria final**

**Manuel Duro Santos**

**Tutor: Dr. Alfons Palacios González**

Curso 2021-22



## Agradecimientos

A mi tutor, Alfons, por su dedicación y sus consejos durante la realización de este trabajo. A Josep y a Miquel por darme la oportunidad de trabajar con ellos y aguantarme durante todo este tiempo.

Tamén agradecer a meus pais e á miña parella o seu apoio durante o desenvolvemento dos dous traballos de fin de Grado.

## Índice

Abstract .....	1
Resum .....	1
Resumen .....	1
1. Introducción.....	3
2. Objetivos del TFG.....	5
3. Marco teórico y estado del arte .....	7
3.1. Gestión urbana y análisis territorial.....	7
3.1.1. Computación urbana.....	7
3.1.2. Datos urbanos .....	8
3.1.3. Sistemas de Información Geográfica (GIS).....	11
3.1.4. QGIS.....	11
3.1.5. Plugins: Python y PostGIS.....	12
3.2. CCU.....	12
3.2.1. Descripción .....	12
3.2.2. Modelo actual de los datos .....	13
3.3. Origen de los datos .....	22
3.4. Escalabilidad del proyecto .....	23
3.4.1. ETL .....	24
4. Metodología y cronograma .....	27
5. Requerimientos .....	31
6. Desarrollo del proyecto .....	33
6.1. Problemas del modelo de datos antiguo.....	33
6.2. Estudio de las fuentes de datos .....	35
6.2.1. Marco municipal.....	35

6.2.2. Marco autonómico.....	37
6.2.3. Marco estatal .....	38
6.2.4. Marco europeo.....	39
6.3. Comparación de los datos actuales con INSPIRE.....	39
6.3.1. Parcelas .....	40
6.3.2. Direcciones .....	41
6.3.3. Manzanas .....	42
6.3.4. Edificios .....	44
6.4. Datos no extraíbles de INSPIRE.....	48
6.5. Propuesta del nuevo modelo de datos.....	49
6.5.1. Consideraciones previas.....	49
6.5.2. Parcel.....	51
6.5.3. Zone.....	53
6.5.4. Address.....	55
6.5.5. Building.....	57
6.5.6. Census .....	58
6.5.7. Place .....	60
6.5.8. Building_floor .....	61
6.5.9. Stretch .....	62
6.5.10. Otras agrupaciones de parcelas .....	63
6.5.11. Company .....	63
6.5.13. Thoroughfare.....	64
6.5.14. Epigraph .....	65
7.5. Implementación de la extracción y transformación.....	65
7.6. Plugin de configuración de las bases de datos de QGIS .....	70
7.7. Open Data .....	72

### III

8. Conclusiones.....	75
9. Posibles ampliaciones.....	77
10. Bibliografía.....	79

## Índice de figuras

Figura 1. Fases, tareas y deadlines. Fuente: elaboración propia .....	28
Figura 2. Calendario. Fuente: elaboración propia .....	29
Figura 3. Relación entre tabla de parcelas de Mataró y CadastralParcel de Inspire. ....	40
Figura 4. Relación entre tabla de direcciones de Mataró y Address de Inspire. Fuente: elaboración propia.....	41
Figura 5. Parcelas en Mataró. Fuente: elaboración propia .....	42
Figura 6. Manzanas (illes) en Mataró. Fuente: elaboración propia. ....	43
Figura 7. Manzanas en Inspire. Fuente: elaboración propia .....	43
Figura 8. Relación entre tabla de manzanas (illes) de Mataró y CadastralZoning de Inspire. Fuente: elaboración propia. ....	44
Figura 9. Ejemplo en FinquesUS en Mataró. Fuente: elaboración propia. ....	44
Figura 10. Ejemplo en Building (INSPIRE). Fuente: elaboración propia. ....	45
Figura 11. Usos en Mataró. Fuente: elaboración propia.....	45
Figura 12. Usos en INSPIRE. Fuente: elaboración propia. ....	46
Figura 13. Ejemplo en FinquesPlantes en Mataró. Fuente: elaboración propia.....	46
Figura 14. Ejemplo en BuildingPart (INSPIRE). Fuente: elaboración propia. ....	47
Figura 15. Ejemplo en FinquesAnyConstruccio en Mataró. Fuente: elaboración propia. .....	47
Figura 16. Ejemplo en Building (INSPIRE). Fuente: elaboración propia. ....	48
Figura 17. Modelo entidad relación. Fuente: elaboración propia .....	51
Figura 18. Islas de Mataró en el modelo de datos antiguo. Fuente: elaboración propia. .....	68
Figura 19. Islas de Mataró en el modelo de datos nuevo. Fuente: elaboración propia.	69
Figura 20. Plugin de configuración de bases de datos. Fuente: elaboración elaboración propia.....	71

## Abstract

This project consists in the study, design and development of a data model oriented to a decision-making system in the urban environment. This project is done in collaboration with the CCU (Centre de Coneixement Urbà), which has been developing plugins for geographic information systems used by city councils for years. Their data model is reviewed in order to improve it and to ensure that the services they provide are more scalable and ensure interoperability.

## Resum

Aquest treball tracta sobre l'estudi, disseny i desenvolupament d'un model de dades orientat a un sistema de presa de decisions en l'àmbit urbà. Aquest projecte es fa en col·laboració amb el CCU (Centri de Coneixement Urbà), el qual fa anys que desenvolupa plugins per a sistemes d'informació geogràfica utilitzats per ajuntaments. Es revisa el seu model de dades per a poder millorar-ho i poder aconseguir que els serveis que proporcionen siguin més escalables i assegurin la interoperabilitat.

## Resumen

Este trabajo trata sobre el estudio, diseño y desarrollo de un modelo de datos orientado a un sistema de toma de decisiones en el ámbito urbano. Este proyecto se hace en colaboración con el CCU (Centre de Coneixement Urbà), el cual lleva años desarrollando plugins para sistemas de información geográfica utilizados por ayuntamientos. Se revisa su modelo de datos para poder mejorarlo y poder conseguir que los servicios que proporcionan sean más escalables y aseguren la interoperabilidad.





## 1. Introducción

Este TFG, realizado en colaboración con el CCU (Centre de Coneixement Urbà), tiene como principal objetivo dar respuesta a las necesidades actuales del CCU. El proyecto inicial empezó trabajando sólo con el ayuntamiento de Mataró, pero con la ampliación de servicios a otros ayuntamientos, surge la necesidad de que las soluciones del CCU tengan un carácter escalable y estandarizable.

La obtención de los datos dependía, prácticamente en su totalidad, de que el ayuntamiento los facilitara con cierta periodicidad para poder introducirlos o actualizarlos en la base de datos. Por tanto, es necesario hacer un estudio de las posibles fuentes de datos para depender lo mínimo posible de los datos que pueda aportar el ayuntamiento y, de este modo, facilitar la automatización y generalización en los procesos de creación y actualización de tablas.

De este modo, en este TFG se trabaja en la estandarización de un modelo de datos para el análisis territorial basado en GIS (Sistemas de Información Geográfica). Este hecho implica, además, la consecuente refactorización de los plugin ya desarrollados con anterioridad por el CCU.



## 2. Objetivos del TFG

Los objetivos de este TFG se centran en dar respuesta a las necesidades actuales del CCU, planteando una reingeniería del modelo actual de datos, estudiar fuentes de datos alternativas y, en definitiva, contribuir a una correcta gestión de la escalabilidad del proyecto para poder ayudar a la toma de decisiones en el ámbito municipal.

- Analizar la situación actual del CCU.
- Proponer un nuevo modelo de datos estandarizado de la información urbana.
- Buscar fuentes de datos autónomas que permitan mantener la base de datos del GIS actualizada.
- Refactorizar los plugins actuales y propuesta de diseño al desarrollar nuevos plugins.
- Detectar y dar respuesta a las necesidades y problemas a las que se presentan los usuarios del servicio.
- Ayudar a la toma de decisiones en el ámbito de la gestión municipal.



### 3. Marco teórico y estado del arte

En este apartado se hace un estudio del estado del arte de la computación urbana y se relaciona con un análisis de situación del CCU.

#### 3.1. Gestión urbana y análisis territorial

Desde la primera revolución industrial, las ciudades han sufrido un proceso de crecimiento exponencial que ha provocado una serie de retos: la contaminación, el consumo energético, la congestión del tráfico y la organización territorial.

Tras siglos de una gestión urbana caótica, en el siglo XIX se empieza a tratar a la urbanización como una ciencia [1], destacando el ejemplo del plan Cerdà en Barcelona [2].

Con la llegada de la informática en el siglo XX (y su correspondiente desarrollo en el siglo actual) se empieza a utilizar en este campo al presentarse la oportunidad de construir ciudades inteligentes que utilicen los datos de los espacios urbanos para resolver los problemas intrínsecos de la urbanización [3] [4]. De esta forma nace la computación urbana.

##### 3.1.1. Computación urbana

La computación urbana (urban computing) es el proceso de adquisición, integración y análisis de un gran volumen de datos heterogéneos generados por las diversas fuentes de espacios urbanos, como sensores, dispositivos, vehículos, edificios y humanos para resolver los problemas a los que las ciudades se enfrentan (la contaminación, el consumo energético, la congestión del tráfico y la organización territorial) [5]. Se trata de un campo multidisciplinar que combina la ingeniería informática con campos más tradicionales como la ingeniería civil, economía, transporte, ecología y sociología en los espacios urbanos.

La computación urbana en la actualidad se enfrenta a tres grandes retos: los sensores urbanos y obtención de datos, la gestión de datos heterogéneos y sistemas híbridos que combinan el mundo físico con el virtual.

- Sensores urbanos y obtención de datos. Actualmente, las ciudades no poseen una infraestructura de sensores lo suficientemente sólida como para proveer datos de toda la ciudad. Por ejemplo, determinar con sensores el tráfico de un determinado segmento de una carretera es viable y asumible, pero es difícil escalarlo para todas las calles de una ciudad. La tendencia actual es recurrir a “humanos con sensores” para que, de este modo, cada ciudadano pueda proporcionar datos de distintos aspectos de la ciudad. Sin embargo, esta solución tiene implícitos una serie de problemas como su distribución temporal y no uniforme a lo largo de la ciudad, datos no estructurados con mucho ruido y, además, se debe buscar un equilibrio entre energía, privacidad y la utilidad de los datos compartidos por cada persona [6].
- Gestión de datos heterogéneos. La diversidad de datos y fuentes de datos involucradas en la computación urbana tienen un impacto directo a la hora de aplicar soluciones de minado de datos y machine learning, la velocidad de respuesta (que debe ser alta en algunas situaciones) y, además, dificulta su visualización.
- Sistemas híbridos que combinan el mundo físico con el virtual. La computación urbana integra datos del mundo real con los del mundo virtual, a diferencia de otros campos de la informática en los que todos los datos se generan y se utilizan en el mismo entorno.

### 3.1.2. Datos urbanos

A continuación, se describen los datos de más relevancia en la computación urbana para, de este modo, poder valorar la complejidad, variedad y heterogeneidad de todos los datos que pueden llegar a estar involucrados en los proyectos informáticos de esta materia. Esta clasificación está basada en las aportaciones de Zheng [5].

#### *3.1.2.1. Datos geográficos*

Los datos geográficos son los datos usados primordialmente en los Sistemas de Información Geográficos (GIS), cuya principal característica es que proporcionan información geométrica y geográfica sobre una determinada entidad.

A nivel geométrico, los datos geográficos pueden ser puntos, líneas o polígonos.

- Puntos. Sirven para determinar la posición en coordenadas de los POI (Point of Interest) como un colegio, un hospital o un centro cívico. Estas entidades, además de esta información geométrica, también están conformadas por otros datos genéricos y específicos como su nombre, dirección, número de plazas, etc.
- Líneas. Sirven, fundamentalmente, para describir el trazado de calles y carreteras, estando generalmente estructurado en un grafo en el que cada arista (línea) es una calle y, cada nodo (punto), es la intersección entre varias aristas. Esta red tiene especial interés para la planificación urbana, el ruteo, la monitorización de tráfico y el análisis de consumo de energía.
- Polígonos. Es la geometría utilizada para delimitar el área de las distintas parcelas de una ciudad, así como su agrupación en polígonos de mayor tamaño como barrios o secciones. Este tipo de información cobra relevancia a la hora de determinar el uso que se le da a una determinada parcela.

#### *3.1.2.2. Datos de tráfico*

Los datos de tráfico tienen como objetivo estudiar la movilidad en las ciudades a través de los datos obtenidos de diferentes sensores repartidos por la ciudad, cámaras de vigilancia e incluso los datos de geolocalización de los propios coches.

#### *3.1.2.3. Datos de señales de teléfonos móviles*

Gracias al registro de llamada detallada que produce un teléfono móvil (call detail record o CDR) se puede localizar y analizar el comportamiento de individuos o de grupos de individuos y, de este modo, estudiar la movilidad en la ciudad para detectar anomalías e incluso para mejorar la planificación urbanística.

#### *3.1.2.4. Datos de monitoreo ambiental*

En este grupo de datos se encuentran los datos de carácter meteorológico como la temperatura, la humedad, la presión atmosférica, la velocidad del viento y el tiempo. También tienen gran interés los datos que hacen referencia a la calidad del aire para, de este modo, activar protocolos anticontaminación cuando sea necesario. Además, en este apartado también se incluyen datos que hacen referencia a la contaminación acústica.

#### *3.1.2.5. Datos económicos*

Los datos económicos de una ciudad abarcan desde la distribución de la riqueza en una ciudad analizando los salarios y valores catastrales, hasta todas las transacciones económicas que se realizan en una ciudad, como la facturación de los distintos comercios y negocios o la compra-venta o alquiler de inmuebles.

#### *3.1.2.6. Datos de energía*

Estos datos tienen que ver con el consumo energético que se realiza en una ciudad, ya sea a nivel de electricidad, agua y gas o, por otra parte, el consumo de gasolina y diésel que se infiere de lo que facturan las gasolineras de una ciudad.

#### *3.1.2.7. Datos de salud*

Los datos de salud de una ciudad tienen que ver de forma directa con el bienestar de sus ciudadanos. Este tipo de datos adquieren un gran papel en la computación urbana cuando se agregan y relacionan con el hecho de vivir en una determinada zona.



Por ejemplo, se podrían apreciar correlaciones de que vivir en una determinada ciudad lleva consigo un mayor riesgo de padecer determinadas enfermedades respiratorias, ansiedad o depresión. Ser consciente de estas relaciones puede ayudar a realizar una planificación urbanística más eficaz que tenga en cuenta estas cuestiones.

### 3.1.3. Sistemas de Información Geográfica (GIS)

Un Sistema de Información Geográfica (Geographic Information System o GIS) es un sistema informático que captura, almacena, consulta, analiza y muestra datos geoespaciales; siendo la capacidad de gestionar este tipo de datos la que lo diferencia de otro tipo de sistemas [7].

Los distintos tipos de análisis espacial que se puede realizar con este tipo de sistemas comprenden el análisis de terreno, el análisis de proximidad, el análisis topológico y, en general, cualquier tipo de análisis de datos que tengan que ver con la computación urbana y los datos geoespaciales.

Los GIS se tratan de herramientas muy útiles para ayudar a la toma de decisiones de políticas en una determinada zona, además de permitir la realización de un profundo análisis sobre las características de un territorio. Tangencialmente, también tienen aplicación directa en la educación (sobre todo en materias relacionadas con geografía).

### 3.1.4. QGIS

QGIS es un Sistema de Información Geográfica (GIS) de código abierto y software libre de la Open Source Geospatial Foundation (OSGeo). Es multiplataforma, soportando Windows, Mac, Linux y Android y está desarrollado en C++.

QGIS permite trabajar con bases de datos PostgreSQL y, en concreto, con su extensión espacial, PostGIS, la cual permite trabajar con geometrías y funciones específicas que hacen uso de ellas.

Por otra parte, QGIS permite utilizar diversos archivos vectoriales (como Shapefiles) y soporta varios tipos de archivos de ráster.

Se trata de un proyecto que se mantiene gracias a la colaboración de voluntarios que contribuyen con código, reporte y corrección de errores, documentación, etc.

Está desarrollado en C++ y permite desarrollar plugins con Python y PostgreSQL. El desarrollo y mantenimiento de plugins son unas de las principales tareas del CCU.

### 3.1.5. Plugins: Python y PostGIS

Uno de los puntos más interesantes de QGIS es la posibilidad de programar y utilizar plugins que agreguen funciones nuevas y específicas al GIS para automatizar uno o varios procesos en concreto.

Estos plugins se construyen con Python y permiten una gran versatilidad al poder utilizar un lenguaje de programación para poder automatizar ciertos procesos, ya sea a través de algoritmos creados con el propio lenguaje, hacer llamadas a algoritmos propios de QGIS que funcionan de forma nativa y, además, de poder acceder a las bases de datos de PostgreSQL con todo lo que ello implica.

Por otra parte, los plugins pueden utilizar algoritmos internos del propio QGIS y, además, establecer conexiones con las bases de datos vinculadas a QGIS y hacer uso de funciones espaciales propias de PostGIS.

Los plugins desarrollados por el CCU permiten calcular índices que agregan diversos datos, calcular áreas de influencia de determinadas entidades, calcular las N entidades más cercanas a un punto, etc.

## 3.2. CCU

### 3.2.1. Descripción

El CCU (Centre de Coneixement Urbà) es un grupo de trabajo de la ESUPT (Escola Superior Politècnica del Tecnocampus) que trabaja conjuntamente con los Ayuntamientos de Mataró, Terrassa y Sabadell.

El CCU desarrolla herramientas para la realización de estudios sobre entidades geolocalizables que dan servicio a la población, así como el estudio de cualquier actividad económica que se desarrolle en el ámbito urbano.

Los dos aspectos más importantes en los que trabaja el CCU son la utilización de recursos gráficos tipo GIS (Sistemas de Información Geográfica) y, por otra parte, la integración de fuentes de diversos datos, destacando en todo momento el conocimiento asociado con la población.

Anteriormente, el CCU trabajaba con el GIS GeoMedia. Sin embargo, debido a que en los últimos años QGIS se ha convertido prácticamente en un software estándar en esta materia, el CCU se ha adaptado a esta tecnología, modificando todos los plugins que había desarrollado en GeoMedia para que pudieran funcionar en QGIS.

Hasta hace 3 años, el CCU trabajaba en exclusiva con el Ayuntamiento de Mataró. Este hecho ha traído consigo que muchas de las soluciones que se han desarrollado en el CCU han sido pensadas en exclusiva para la realidad de una sola ciudad. No obstante, el CCU ha empezado a trabajar con otros ayuntamientos como Terrassa o Sabadell, por lo que ha surgido la necesidad de buscar formas de que todo el software que se diseñe sea generalizable y escalable para cualquier ayuntamiento, teniendo en cuenta un contexto complejo y cambiante.

### 3.2.2. Modelo actual de los datos

A continuación, se detallan las principales tablas que conforman la base de datos, centrándose en la organización en particular que se ha dado en el Ayuntamiento de Mataró. Las tablas son las siguientes:

- Parcel. En esta tabla aparecen las distintas parcelas de la ciudad. Su geometría está basada en polígonos.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Geom	Geometry (MultiPolygon, 25831)
Planol	Varchar (254)
CODI_GIS	Varchar (254)
UTM	Varchar (254)
Utm_total	Varchar(254)
Refca_pol	Varchar (254)
Refca_par	Varchar (254)
Parcel_utm	Varchar (254)
Parcel_gis	Varchar (254)

- ILLES. Se tratan de agrupaciones de parcelas que conforman una misma manzana. Su geometría está basada en polígonos

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Geom	Geometry (MultiPolygon, 25831)
D_S_I	Varchar (254)
AV_COUNT	Float8
SUM_AREA	Float8
ID1	Int4

- Seccions. Son grandes agrupaciones que polígonos que dividen la ciudad en secciones.

<b>Campo</b>	<b>Tipo</b>
Id	Int8
Geom	Geometry (MultiPolygon, 25831)
Fid	Float8
Seccio	Varchar (254)
Districte	Varchar (10)
Dist_secc	Varchar (10)

- Barris. Polígonos que delimitan cada uno de los barrios de la ciudad.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Geom	Geometry (Polygon, 25831)
Numbarri	Int4
Nombarri	Varchar

- DistrictesPostals. Polígonos que determinan cada uno de los distritos postales de la ciudad.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Geom	Geometry (MultiPolygon, 25831)
Codi	Varchar
Cp	Int4

- Padro. En esta tabla figuran todos los habitantes de Mataró de forma anónima. Aparece la fecha de nacimiento, sexo, nivel de estudios y dirección. Esta tabla es facilitada por el Ayuntamiento de Mataró y se actualiza de forma

periódica y manual. La tabla presenta unos nombres de campo que no ayudan a saber qué tipo de información se está almacenando.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
HABDBOIDE	Varchar
HABCODIND	Varchar
HABDISTRI	Int
HABSECCIO	Int
HABCODMAN	Int
D_S_I	Varchar
HABNUMHOJ	Int
HABNUMORD	Int
BARCODBAR	Int
BARNOMBAR	varchar
TMOTIPOPE	Varchar
HABTIPMOV	Varchar
TMODESCRI	Varchar
HABFECOCU	Varchar
HABFECACT	Varchar
HABELSEXO	Int
SEXDESCRI	Varchar
HABNIVINS	Int
NINDESCRI	Varchar
HABTODDIR	Varchar
HABORDDIR	Varchar
HABFECNAC	date
HABCOPANA	Int
HABNOPANA	Varchar
HABCOPRNA	Int
HABNOPRNA	Varchar

HABCOMUNA	Int
HABNOMUNA	Varchar
COMCOCONA	Int
COMNOCONA	Varchar
HABLLEMUN	Int
HABFECBPR	Varchar
HABCOPAPR	Int
HABNOPAPR	Varchar
HABCOPRPR	Int
HABNOPRPR	Varchar
HABCOMUPR	Int
HABNOMUPR	Varchar
COMCOCOPR	Int
COMNOCOPR	Varchar
HABCOPADE	Int
HABNOPADE	Varchar
HABCOPRDE	Varchar
HABNOPRDE	Varchar
HABCOMUDE	Int
HABNOMUDE	Varchar
COMCOCODE	Int
COMNOCODE	Varchar
REFCAD	Varchar
CarrerNumBis	Varchar
HABNACION	Int
codigo	Varchar
REFCADAntic	Varchar

- Dintreilla. Se trata de una tabla en la que aparecen los números de policía con el punto geométrico que les corresponde. Su nombre se debe a que estos

puntos están dentro de cada una de las islas de la tabla ILLES para facilitar determinadas operaciones espaciales.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Geom	Geometry (Point, 25831)
Carrer_Num_Bis	Varchar
NUMERO	Varchar
EDIFICI	Varchar
CODI_GIS	Varchar
REF_CADAST	Varchar
UTM_x	Float8
UTM_y	Float8

- **BrossaComercial.** Esta tabla contiene informaciones de todos los negocios de Mataró, indicando su nombre, descripción de la actividad, dirección, etc. Esta tabla es facilitada por el Ayuntamiento de Mataró y se actualiza de forma periódica y manual.

<b>Campo</b>	<b>Tipo</b>
Id	Varchar
SECCIOIAE	Varchar
EPIGRAFIAE	Varchar
CALCODCAL	Int
DESCRIPCIOACTIVITAT	Varchar
FULLNAME	Varchar
DIRSIGLAS	Varchar
DIRNOMCAL	Varchar
DIRDUPINI	Varchar
DIRNUMINI	Int
NUM2ACTIVITAT	Int



DUPLI2ACTIVITAT	Varchar
DIRESCALE	varchar
AdreçaActivitat	Varchar
DIRPLANTA	varchar
DIRPUERTA	Varchar
ZIPCODE	Varchar
METRES2	Float
CADASREF	varchar
PLANREF	varchar
NumPol	varchar

- FinquesPlantes. Esta tabla muestra las parcelas (identificadas por su referencia catastral) e indica los pisos.

<b>Campo</b>	<b>Tipo</b>
Id	int
Ref_Cadastral	Varchar
Escala	Varchar
Pis	Varchar
Superficie_cons	Varchar
UTM	Varchar

- FinquesUS. Esta tabla muestra las parcelas (identificadas por su referencia catastral) e indica el uso que se le está dando a esa parcela.

<b>Campo</b>	<b>Tipo</b>
Id	int
Ref_Cadastral	Varchar
Quants	Varchar
Us	Varchar
Superficie_Cons	Varchar

UTM	Varchar
-----	---------

- **FinquesAnyConstruccion.** Esta tabla muestra las parcelas (identificadas por su referencia catastral) e indica el año de construcción de cada parcela.

<b>Campo</b>	<b>Tipo</b>
Id	int
Ref_Cadastral	Varchar
Any_constr	Varchar
UTM	Varchar

- **Continents.** Tabla con el listado de países y sus correspondientes códigos.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
CONCODPAI	Int4
CONNOMPAI	Varchar
NOMCAT	Varchar
CONCODCON	Int4
CONNOMCON	Varchar
CONZONCON	Varchar
UNIEUROPEA	Int4
UE25	Int4
UE27	Int4
UE77	Int4
UE78	Int4
Id1	Int4

- **Carrer.** Tabla con información de todas las calles de Mataró.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Tipus	Varchar
Categoría	Varchar
Nom	Varchar
TipusNom	Varchar
URBANITZACIONS	Varchar
Epigraf	Varchar
Codi	Varchar

- Seccio1. En esta tabla hay información sobre las secciones y epígrafes de los distintos tipos de empresas.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Seccio	Float8
Descripcio epigraf	Varchar
Epigraf	Varchar

- SegmentsXarxaCarrers. Tabla con el grafo de todos los tramos que conforman las calles y carreteras de Mataró.

<b>Campo</b>	<b>Tipo</b>
the_geom	Geometry (LineString, 25831)
Tid	Int4
Cost	Float8
Vel	Float8
Reverse_cost	Float8
Source	Int4
Target	Int4

Nombre_Semafor	Int4
Cost_Total_Semafor_Tram	Float8
Id	Int4
LONGITUD_SEGMENT	Float8
LENGTH	Float8
SENTIT	Float8
PENDENT_ABS	Float8
VELOCITAT_PS	Float8
VELOCITAT_PS_INV	Float8

- Entidades. Aquí se encuentran distintas tablas que tienen ciertos campos en común. Dentro de esta categoría se encuentran los CEIPs, CAPs, Farmacias, etc. Pueden tener más campos que los aquí descritos.

<b>Campo</b>	<b>Tipo</b>
Id	Int4
Geom	Geometry (Point, 25831)
Nom	Varchar
NPlaces	Int4

### 3.3. Origen de los datos

El modelo de datos actual tiene el problema de que se depende totalmente de que los ayuntamientos faciliten todas las tablas necesarias, por lo que es necesario hacer un estudio de las posibles fuentes de datos para poder tener más independencia con respecto a los ayuntamientos.

Antes de la realización de este TFG, los datos eran proporcionados por el propio ayuntamiento interesado. En el caso de tablas de carácter relativamente estático y constante (tal y como puede ser la geometría del mapa) podría no ser un problema. No

obstante, en datos que fluctúan constantemente a lo largo del tiempo, como el padrón, la actualización de los datos es necesaria y frecuente.

En el caso concreto de Mataró, los datos se recibían de múltiples formas: con archivos shape, csv, Access... Este método de actualización de datos es tremendamente ineficiente ya que, aunque estaba relativamente automatizado con instrucciones SQL, siempre era necesaria la supervisión de una persona para realizar este proceso debido a una baja calidad de los datos.

En los últimos años, muchos ayuntamientos están optando por crear y potenciar su propio portal de datos abiertos (open data) para que estén a disposición de la ciudadanía y de las empresas a las que les pueda interesar, tal y como ocurre en ayuntamientos como Mataró, Barcelona, Rubí, Terrassa, etc.

La propuesta del nuevo modelo de datos debe tener en cuenta esta tendencia para aprovechar esta fuente de datos.

Por otra parte, se deben tener en cuenta otras fuentes de datos de carácter oficial que pueden servir para crear todas las tablas que necesite el ayuntamiento. Estas posibles fuentes de datos son el catastro, el INE o INSPIRE, una directiva europea que pretende hacer que la información geográfica y espacial sea más accesible e interoperable. Además, son destacables iniciativas de código abierto como Open Street Map u OSRM.

#### 3.4. Escalabilidad del proyecto

Cuando un negocio amplía sus límites puede llegar a enfrentarse a un problema de escalabilidad. La escalabilidad se define como la propiedad de un sistema para soportar el crecimiento al añadir más recursos a un sistema [8].

En el caso del CCU, este problema ha aparecido en el momento en el que se ha pensado en ofrecer los servicios a más ayuntamientos. Anteriormente, el planteamiento de ciertas soluciones se ha hecho *ad hoc* para el ayuntamiento de Mataró, ya que era el

único ayuntamiento para el que se daba servicio y no había planteado escalar el proyecto.

Sin embargo, hace unos años ha surgido la posibilidad de ampliar al proyecto a otros ayuntamientos, con la consecuente necesidad de establecer una serie de pautas que permitan su correcto desarrollo.

El eje fundamental sobre el que gira todo el servicio es la base de datos, la cual está vinculada de forma directa con un GIS. Se ha detectado que cada ayuntamiento trabaja con sus bases de datos siguiendo un criterio propio y no unificado, por lo que este hecho dificulta el desarrollo de nuevos servicios para cualquier ayuntamiento.

Por este motivo, surge la necesidad de proponer un modelo de datos estándar que resuelva las necesidades más comunes del análisis territorial. Existen propuestas (como en [schema.org](http://schema.org)) de modelos de datos estándar para tablas muy comunes como “persona”, “organización” o “evento”; pero en lo que se refiere a tablas relacionadas con los GIS no hay ninguna estandarización clara.

#### 3.4.1. ETL

Debido a la naturaleza del problema, cobra especial interés utilizar procesos ETL para facilitar la utilización de diversas bases de datos. ETL (*extract, transform, load*) es un tipo de integración de datos que permite utilizar datos de fuentes distintas y se basa en las tres fases que definen sus siglas [9]:

- Extraer. Esta primera fase del proceso consiste en extraer los datos del sistema o de los sistemas de origen. Es posible extraer datos de sistemas muy distintos entre sí, aunque deberán de reunir los requisitos necesarios definidos por el proceso.
- Transformar. En esta fase el proceso realizará las acciones necesarias para poder tratar los datos de una forma homogénea independientemente de cuál

sea su origen. Estas transformaciones pueden consistir en separar una columna en varias, obtener nuevos valores derivados, reemplazar o modificar datos, etc.

- Cargar. Finalmente, los datos transformados se cargan en el sistema de destino.

Existen herramientas en el mercado que facilitan los procesos ETL. No obstante, debido a un estudio previo de las posibles fuentes de datos, se opta por buscar una solución que obtenga los datos de fuentes fiables (catastro, INSPIRE...) y que, a través de un programa propio, se haga una transformación de los datos que permita la correcta adecuación a las necesidades de los ayuntamientos.





## 4. Metodología y cronograma

Para realizar este proyecto se utiliza una metodología inspirada en “agile”. Esta metodología consiste en plantear un desarrollo iterativo e incremental que permite evaluar de forma rápida el progreso del proyecto y que, además, permite una respuesta al cambio ágil.

Las principales premisas de esta filosofía (definidas en “*The Manifesto for Agile Software Development*”) son las siguientes:

- Individuos e interacciones por encima de procesos y herramientas.
- Software funcional por encima de documentación demasiado extensa.
- Colaboración con el cliente por encima de la negociación del contrato.
- Responder al cambio por encima de seguir un plan.

Concretamente, se utiliza una metodología ágil, planteando una serie de sprints de duración corta (dos semanas) para poder ir evaluando el desarrollo del proyecto y hacer cambios donde sea necesario. Además, se han hecho reuniones semanales con el equipo para poder ir valorando el avance del proyecto.

El proyecto se divide en tres grandes fases: el anteproyecto, la memoria intermedia y la memoria final. Debido a ello, la metodología se adapta a estas fases de forma que el proyecto se vaya desarrollando de forma iterativa e incremental.

El cronograma inicial del proyecto (atendiendo a las fases, objetivos y deadlines) queda de la siguiente manera (Figura 1).

NAME	Start Date	Due Date
▼ ☰ 📁 Project Activities		
▼ ◆ 📅 Anteproyecto	Jan 10	Feb 11
■ 🔗 Documento anteproyecto	Jan 10	Feb 11
▼ ◆ 📅 Memoria intermedia	Feb 12	Apr 22
■ 🔗 Documento memoria inte...	Feb 12	Apr 22
■ 🔗 Diseño modelo de datos	Feb 12	Feb 28
■ 🔗 Implementación inicial ETL	Mar 1	Apr 1
▼ ◆ 📅 Memoria final	Apr 23	Jun 13
■ 🔗 Documento memoria final	Apr 23	Jun 13
■ 🔗 Propuesta de modelo de ...	Apr 23	May 20
■ 🔗 Implementación ETL	Apr 23	May 20
■ 🔗 Pluqin refactorizado	May 20	Jun 3

Figura 1. Fases, tareas y deadlines. Fuente: elaboración propia

Lo que en forma calendarizada adquiere la siguiente forma (Figura 2).

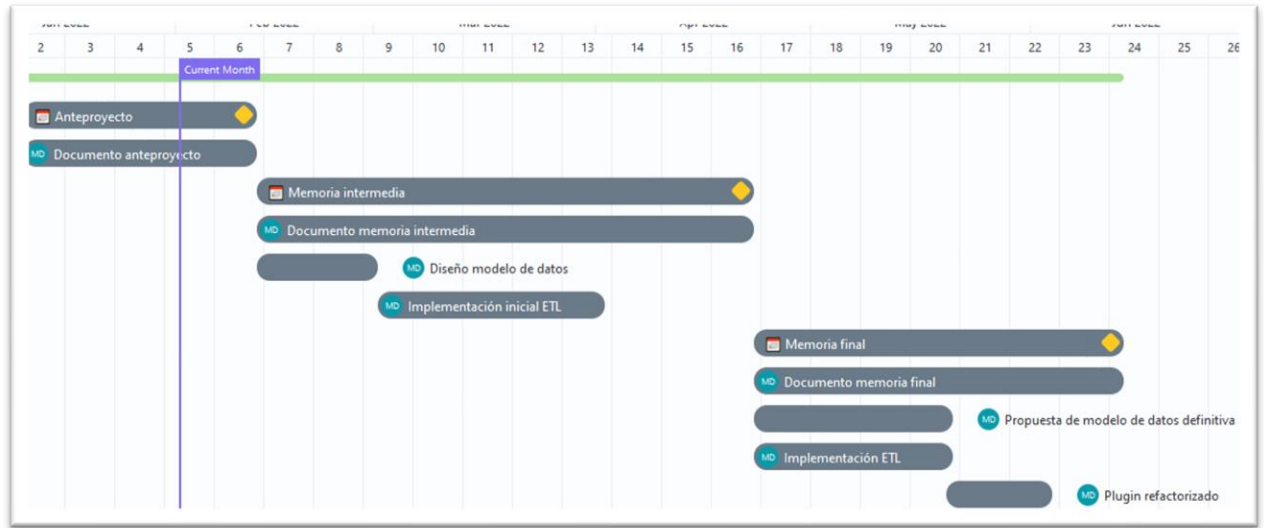


Figura 2. Calendario. Fuente: elaboración propia.



## 5. Requerimientos

Los requerimientos de este TFG se basan, en una parte, en un planteamiento teórico del modelo de datos más apropiado para el proyecto, así como un estudio de las posibles fuentes de datos oficiales de las que poder extraer los datos de forma autónoma e independiente; y, por otra parte, el desarrollo de un plugin que permita crear y actualizar las tablas de la base de datos que utilizan los plugins para QGIS del CCU.

- Definir un modelo de datos en el ámbito de la gestión territorial teniendo en cuenta las necesidades de entidades públicas.
- Desarrollar una herramienta que permita la creación y actualización de los datos de forma acorde al nuevo modelo de datos.
- Crear y actualizar la base de datos del GIS a partir de fuentes de datos autónomas (gubernamentales y/o abiertas).
- Hacer refactoring en los plugins del CCU para adecuarse al nuevo modelo de datos.



## 6. Desarrollo del proyecto

El proyecto comienza con un análisis pormenorizado de las distintas posibles fuentes de datos y su idoneidad en términos de escalabilidad y mantenimiento a lo largo del tiempo. En lo que se refiere a este último aspecto, en las fuentes de datos elegidas se debe tener constancia de que no sólo tienen utilidad a corto plazo en un espacio geográfico acotado (como una ciudad en concreto), sino que se debe valorar positivamente cualquier iniciativa que trate de establecer unos estándares que sirvan como marco de referencia para el análisis territorial.

### 6.1. Problemas del modelo de datos antiguo

Desde que el CCU comenzó a prestar sus servicios al ayuntamiento de Mataró, las tablas se han ido añadiendo de forma progresiva según las necesidades. Estas tablas provenían en su mayoría del propio ayuntamiento, generando una gran dependencia con respecto a ellos.

Si atendemos al diseño de la nomenclatura de las tablas, existen varios fallos de diseño destacables:

- Nombres que no aclaran de forma correcta qué datos hay en un campo. Tal y como se pudo comprobar en el punto “3.2.2. Modelo actual de los datos”, hay nombres de campos que no aclaran nada la información que guardan. Por ejemplo, en el caso del padrón encontramos nombres como HABNOMUNA, COMCOCON, NINDESCRI o COMCOCODE. Sin más contexto, resulta imposible saber qué datos puede haber en cualquier de estos campos.
- Nombres que no aclaran de forma correcta qué datos hay en una tabla. Debido a una cuestión de inercia, hay tablas que acaban quedándose con nombres que están lejos de ser auto explicativos. Es el caso de CONTINENTS, que en realidad se trata de un listado de países y no de continentes, o dintreilla, que es una tabla en la cual aparecen las direcciones de la ciudad con su localización geoespacial. En este último caso, el nombre de dintreilla viene de que estos puntos están

posicionados dentro de las islas, pero realmente la tabla tiene valor debido a que almacena todas las direcciones.

- Incoherencia al dar nombre a las tablas. Algunas tablas están en plural y otras en singular. Por ejemplo, existe la tabla parcel y la tabla ILLES.
- Incoherencia de nombres entre campos equivalentes en distintas tablas. En tablas que tienen campos que están referenciados a otras tablas (como puede ser la referencia catastral), se encuentran nomenclaturas radicalmente distintas. Por ejemplo, la referencia catastral puede aparecer como REFCAD, REF\_CADAST o UTM, dependiendo de a qué tabla nos refiramos.
- Mezcla de mayúsculas y minúsculas sin atender a ningún criterio común. Tanto en el nombre de campos como en el nombre de tabla, se mezclan mayúsculas y minúsculas indistintamente. Esto en PostgreSQL tiene importancia ya que cuando en una consulta se escribe un nombre de campo o de tabla, por defecto se interpreta como minúscula. Si se quiere hacer distinción entre mayúsculas y minúsculas es necesario utilizar comillas dobles en el nombre del campo o tabla. Esto acaba generando que las consultas SQL sean más difíciles de leer y escribir, al tener que haber comillas casi siempre que se utilice el nombre de un campo o tabla.
- Mezcla de idiomas. La mayor parte de los nombres de las tablas y campos están en catalán, pero, en algunos casos, algunas tablas están en inglés.
- Uso de caracteres especiales. Ocurre en “AdreçaActivitat”. Es una práctica poco aconsejable y que se debe evitar en la medida de lo posible.

En la mayor parte de los casos se trata de una cuestión continuista con respecto a los datos que se obtienen, pero si se quiere escalar el proyecto es necesario pararse a plantear un diseño y dejar de acumular deuda técnica.



A todas estas cuestiones se le suma el hecho de que el modelo de datos antiguo ha sido desarrollado ad hoc para Mataró, por lo que esto puede traer consigo consecuencias en el momento que se quiera prestar servicios a otros ayuntamientos.

## 6.2. Estudio de las fuentes de datos

En este apartado se desarrollan las posibles fuentes de datos en función de una clasificación político-territorial, empezando por las posibles fuentes de datos de carácter más local hasta acabar en un nivel internacional. Todas las fuentes de datos aquí presentadas gestionan sus propios datos o recopilan datos de otras fuentes y, además, los servicios que ofertan tienen una naturaleza institucional y/o gubernamental. Este hecho proporciona seguridad de cara a comprometerse con una determinada fuente de datos.

El foco de este estudio se hace eminentemente sobre Infraestructuras de Datos Espaciales (IDE), las cuales son un entorno en el que diversas instituciones publican y etiquetan información geoespacial mediante diversos servicios y formatos estándar. El objetivo de estas iniciativas es seguir unas pautas comunes para permitir su interoperabilidad y, si bien es cierto que en cierta medida sí que se publican datos con una cierta equivalencia entre ellos, en la actualidad las tablas involucradas no son necesariamente las mismas con respecto a otras fuentes de datos (por ejemplo, no todas disponen de una tabla en la que figuren todos los árboles de la ciudad) y tampoco tienen los mismos campos. Por tanto, aunque la meta es conseguir un modelo estándar, en la actualidad aún se presentan problemas para el diseño de servicios y soluciones extrapolables a cualquier lugar en el ámbito del análisis territorial municipal.

### 6.2.1. Marco municipal

En lo que se refiere a municipios, los ayuntamientos de un determinado tamaño disponen de servicios de análisis territorial que tratan con datos que obtienen de otras fuentes, como el propio catastro, e incluso tienen capacidad para generar nuevos datos al realizar determinadas gestiones en dicho ayuntamiento, como empadronarse o abrir un negocio en ese municipio. Este caso es, sin ir más lejos, el de Mataró.

En primer lugar, existe la posibilidad de acudir a un determinado ayuntamiento al que le puedan interesar los servicios del CCU y, consecuentemente, firmar un acuerdo de colaboración para poder acceder a determinados datos. Esta opción es la que actualmente se sigue en el CCU, teniendo como limitaciones un acoplamiento total a los datos proporcionados y una gran dependencia de tener un contacto estrecho con el ayuntamiento para añadir o modificar datos, además no disponer de los datos en tiempo real y depender de actualizaciones periódicas.

Por otra parte, cada vez más ayuntamientos ponen a disposición de las empresas potencialmente interesadas y de la ciudadanía diversos datos en portales propios de Open Data (datos abiertos). En el caso concreto de Mataró, su portal de Open Data ofrece datos acerca de los presupuestos del ayuntamiento, elecciones, personas (como el número de habitantes por manzana), cultura, ocio, territorio, la agenda del alcalde, servicios y transporte. Todos estos datos son consultables a través de archivos CSV o Web Services (además de las propias herramientas de visualización del portal web).

Esta última opción presenta ideas interesantes como poder desarrollar herramientas que se integren con este tipo de portales para poder extraer los datos de forma fácil y automática. No obstante, a día de hoy este tipo de portales tienen asociados varios problemas.

El primer problema de tomar como referencia los datos de los portales de Open Data, es que estos portales no están estandarizados por lo que una misma tabla que haga referencia, por ejemplo, a los colegios de una ciudad, podría ser radicalmente distinta con respecto a otro ayuntamiento en lo que concierne a número, nombre y tipos de campos (o incluso la existencia de esta tabla bajo el mismo criterio). Esta cuestión dificulta considerablemente la forma en la que se presta servicio a cada ayuntamiento, ya que implica realizar un estudio concienzudo de su portal para poder adecuarse a lo que actualmente ponen a disposición del público.

Otro de los problemas de tomar los portales de Open Data actuales como principal fuente de datos es que no hay garantías de que estos portales mantengan en el tiempo los nombres, formatos y tipos de las tablas que muestran en la actualidad. Dicho de otro modo, existe un riesgo potencial de fallos debido a cambios en los datos ofertados en el portal que no se comuniquen debidamente en tiempo y forma, además del impacto en términos de desarrollo de software al tener que modificar la forma en la que se extraen los datos.

#### 6.2.2. Marco autonómico

En cuanto a las potenciales fuentes de datos ofrecidas por las comunidades autónomas, en el caso particular de Catalunya existen diversos servicios con relación al análisis territorial como la “Infraestructura de Dades Espacials de Catalunya” (IDEC), el portal de “Dades Obertes de Catalunya” o el “Institut Cartogràfic i Geològic de Catalunya” (ICGC).

En el caso del ICGC, tiene competencias sobre cartografía y sobre la infraestructura de datos especiales de la comunidad. Más allá de ser una posible fuente de datos, el ICGC pone diversas herramientas a disposición de desarrolladores interesados en el ámbito del tratamiento de datos geoespaciales, entre otros servicios ofertados a empresas o a la ciudadanía en general.

En lo referente a datos que potencialmente puedan interesar al desarrollo de este proyecto, el ICGC aloja en su web datos relacionados con el territorio y el urbanismo como, por ejemplo, la altimetría, la planimetría o un referencial topográfico territorial en el que se incluyen diversas capas como el relieve, hidrografía, transporte, construcciones, coberturas de suelo y nombres geográficos.

En cuanto al portal de “Dades Obertes de Catalunya”, dispone de un catálogo de datos de ciencia, tecnología, comercio, cultura, ocio, demografía y economía, tomando algunos datos de otras fuentes como el Institut d'Estadística de Catalunya (IDESCAT). No obstante, el catálogo de datos ofertados en la actualidad no tiene una aplicación directa sobre los servicios del CCU debido a que tienen un carácter que se centra en cuestiones de toda la comunidad y, por tanto, dificulta el análisis territorial para una zona en

específico. Por ejemplo, en el caso del padrón disponen de tablas que muestran el número de habitantes por rangos de edad y sexo para cada municipio. Si bien es cierto que esta tabla puede tener interés para realizar un análisis a nivel de comunidad autónoma, en el caso de un ayuntamiento en concreto los únicos datos que se están conociendo son el número de personas que viven en dicho municipio, su edad y sexo y, por tanto, no permite analizar la distribución espacial de estas personas a lo largo del municipio.

Esta fuente de datos es muy sólida en cuanto a los datos y en la forma en la que se proporcionan, pero tiene como contrapartida la limitación que está delimitada por comunidades autónomas y, tal como ocurría en el caso de los municipios, dificulta la tarea de presentar un modelo de datos escalable y extrapolable a cualquier territorio, ya que se depende de forma directa de la existencia de instituciones como el ICGC y los datos que se oferten en dichos servicios, por lo que para cada comunidad autónoma sería necesario realizar un estudio en específico de su situación.

### 6.2.3. Marco estatal

En cuanto al marco estatal, las principales fuentes de datos que sirven como referente son la Infraestructura de Datos Espaciales de España (IDEE), la sede electrónica del catastro o el portal de datos abiertos del Gobierno de España.

La IDEE es un portal que ofrece multitud de servicios relacionados con datos espaciales, APIs para desarrolladores y un catálogo de datos que recopila y unifica datos de diversas fuentes. Este último apartado, el cual sería el de mayor interés para el CCU, necesita tiempo para convertirse en un catálogo de referencia para todo el territorio nacional, ya que en la actualidad sólo recopila información de ciertos ayuntamientos y ciertas comunidades autónomas, por lo que no es una fuente de datos que permita en la actualidad ser utilizada en todo el país para todos los tipos de datos que se publican. El portal de datos abiertos del Gobierno de España presenta características similares al IDEE en este aspecto, mostrando un catálogo de datos amplio pero que no llega a abarcar a todo el territorio de forma uniforme, siendo un servicio que aún se encuentra en un estado de recopilación fuentes de datos.

La sede electrónica del catastro dispone de datos y cartografía por municipio, como información alfanumérica, cartografía vectorial, datos estadísticos, entre otros. La contrapartida de esta fuente de datos es que acceder a los datos es más complicado que en cualquiera de los otros servicios que se han presentado en este trabajo, ya que es necesario acceder a título personal mediante certificado electrónico o cl@ve, además de que la propia web del catastro suele tener problemas en relación a la seguridad y sus certificados digitales. Por tanto, el empleo de esta fuente (a priori) implicaría tener que descargar los datos de forma manual de forma periódica.

#### 6.2.4. Marco europeo

La Unión Europea es una entidad que surge para promover la integración y gestión económico-política común entre las distintas naciones del continente. Es por ello que el análisis territorial, gestión y difusión de datos de carácter geoespacial son completamente susceptibles de ser legislados y regularizados en toda la UE. De esta necesidad surge la Directiva INSPIRE, que tiene como objetivos conseguir la interoperabilidad y reutilización de datos geográficos más allá de las fronteras a través de una infraestructura de datos espaciales que tome como referencia las infraestructuras previamente establecidas de los Estados Miembros. Además, INSPIRE establece diversas pautas en las que describe y define los distintos tipos de entidades que puede haber involucradas en los datos de tipo geoespacial.

Varios de los servicios previamente citados se adhieren a la Directiva INSPIRE, como el IDEE o la Dirección General del Catastro, por lo que utilizar estos servicios es una buena opción si cuentan con el sello de INSPIRE al asegurar su interoperabilidad y facilitar enormemente el desarrollo de software y las integraciones que necesiten este tipo de datos.

#### 6.3. Comparación de los datos actuales con INSPIRE

A continuación, se hace una comparativa de las tablas más relevantes que se utilizan en el CCU en la que se contraponen los datos actuales que tiene el CCU para el

ayuntamiento de Mataró, con los directamente extraídos de la Dirección General del Catastro siguiendo la Directiva INSPIRE.

### 6.3.1. Parcelas

Se define Parcela Catastral como un área individual de superficie de la tierra, sujeta a derechos reales de propiedad, homogéneos y de titularidad única [10].

Actualmente en Mataró hay 13349 parcelas y del catastro se obtienen 13724.

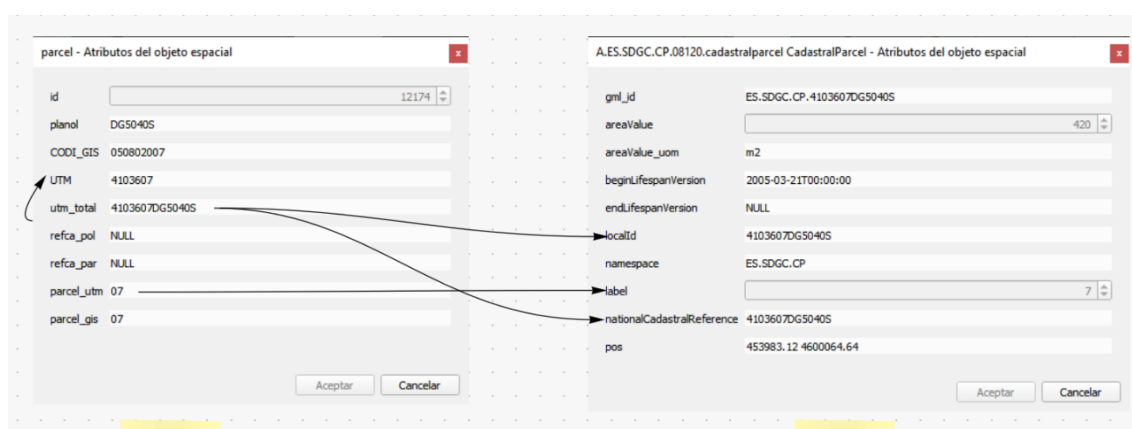


Figura 3. Relación entre tabla de parcelas de Mataró y CadastralParcel de Inspire.

El campo más importante de la tabla parcel para el CCU es el UTM, un campo de carácter único que identifica cada parcela. Este campo se deduce de otro campo, utm\_total, el cual tiene 7 caracteres más.

El utm\_total de parcel (Mataró) se corresponde con el localId o nationalCadastralReference de CadastralParcel (Inspire). El campo nationalCadastralReference es la referencia catastral de la parcela, mientras que en localId se alojan los 14 primeros caracteres de la referencia catastral, la cual tiene también una longitud de 14 caracteres (al menos en Mataró).

Se puede apreciar otra equivalencia entre parcel\_utm (o parcel\_gis) con label. Se trata del número de parcela que se corresponde con el número que se representa en la

cartografía. En parcelas urbanas son dos dígitos y en parcelas rústicas pueden ser hasta de 5 dígitos.

Por otra parte, ambas fuentes disponen de un campo de geometría. Cabe destacar que en Mataró está ligeramente desplazado con respecto a las coordenadas reales, pero no se trata de una diferencia que suponga algún problema para el tipo de análisis que se hace en el ayuntamiento.

### 6.3.2. Direcciones

Se define la dirección completa como una jerarquía que consta de varios componentes con un nivel de detalle incremental, por ejemplo: ciudad, nombre de la calle, la casa con su número o nombre [10]. También puede incluir un código postal u otros descriptores de postales.

Actualmente en Mataró hay 21660 direcciones y del catastro se obtienen 16625. Esta diferencia se debe a que en Mataró se proporciona un punto para cada dirección y en el catastro sólo se proporciona un punto para cada portal. Por ejemplo, si hay un portal con el número 3-5, en Mataró habrá un punto para el 3 y otro para el 5, mientras que en el catastro sólo se proporcionará un punto.

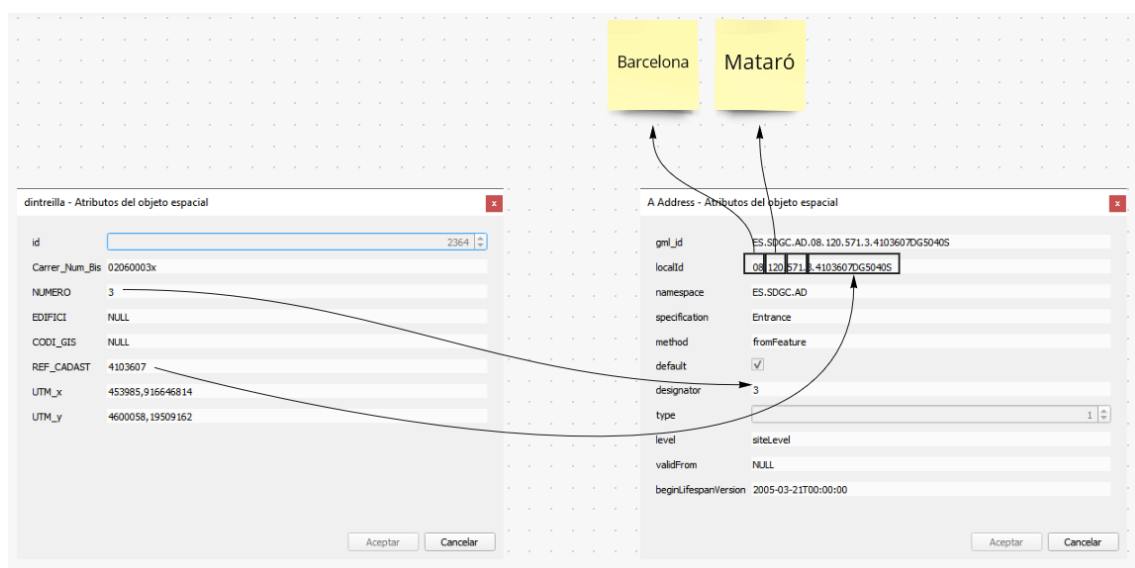


Figura 4. Relación entre tabla de direcciones de Mataró y Address de Inspire. Fuente: elaboración propia.

El campo más importante de la tabla *dintreilla* para el CCU es el *CarrerNumBis*, el cual es un campo generado por el propio ayuntamiento que contiene el código de calle, el número y la letra. Por tanto, no tiene una correspondencia directa con la tabla *Address* obtenida a través del servicio de Inspire.

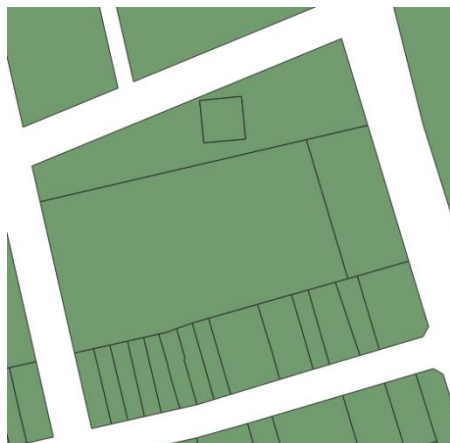
No obstante, sí que hay equivalencia a nivel de referencia catastral y del número.

Cabe destacar que en el *localId* de la tabla *Address* encontramos la referencia catastral, lo que permite vincularse con la tabla anterior de forma sencilla.

### 6.3.3. Manzanas

Se define como manzana al espacio urbano, edificado o destinado a la edificación, generalmente cuadrangular, delimitado por calles por todos sus lados [11].

Actualmente en Mataró hay 761 manzanas y del catastro se obtienen 1128. Se trata de una diferencia de criterio a la hora de definir lo que es una manzana. En la siguiente figura se muestra una captura de unas parcelas en concreto de Mataró (Figura 5).



*Figura 5. Parcelas en Mataró. Fuente: elaboración propia.*

En Mataró se determinan las manzanas a nivel geométrico, agrupando en una misma manzana todas las parcelas colindantes (Figura 6).



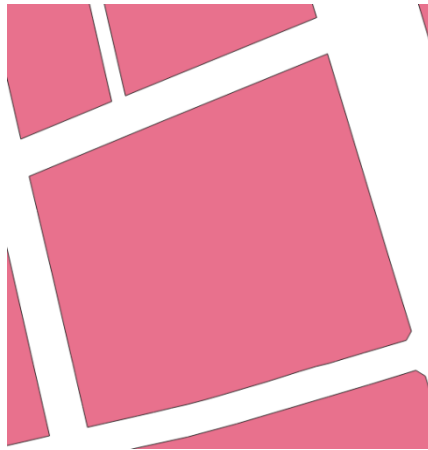


Figura 6. Manzanas (illes) en Mataró. Fuente: elaboración propia.

En el caso de INSPIRE, se considera que una parcela pertenece a aquella manzana en las que los primeros dígitos de la referencia catastral sean iguales al “label” de la manzana (Figura 5). Se considera que las parcelas que comparten los 5 primeros dígitos pertenecen a una misma manzana.

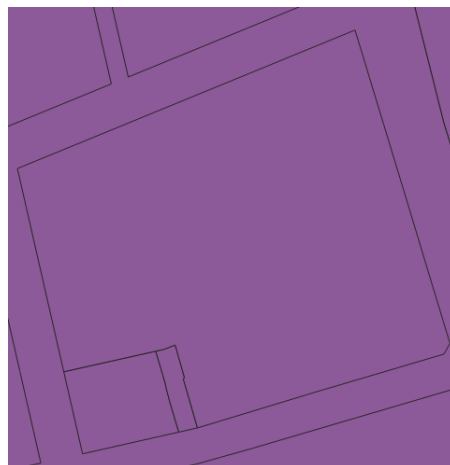


Figura 7. Manzanas en Inspire. Fuente: elaboración propia.

El campo más importante de la tabla ILLES para el CCU es el D\_S\_I, el cual es un campo generado por el propio ayuntamiento que contiene el código del distrito calle, la sección y la isla. Por tanto, no tiene una correspondencia directa con la tabla Address obtenida a través del servicio de INSPIRE.

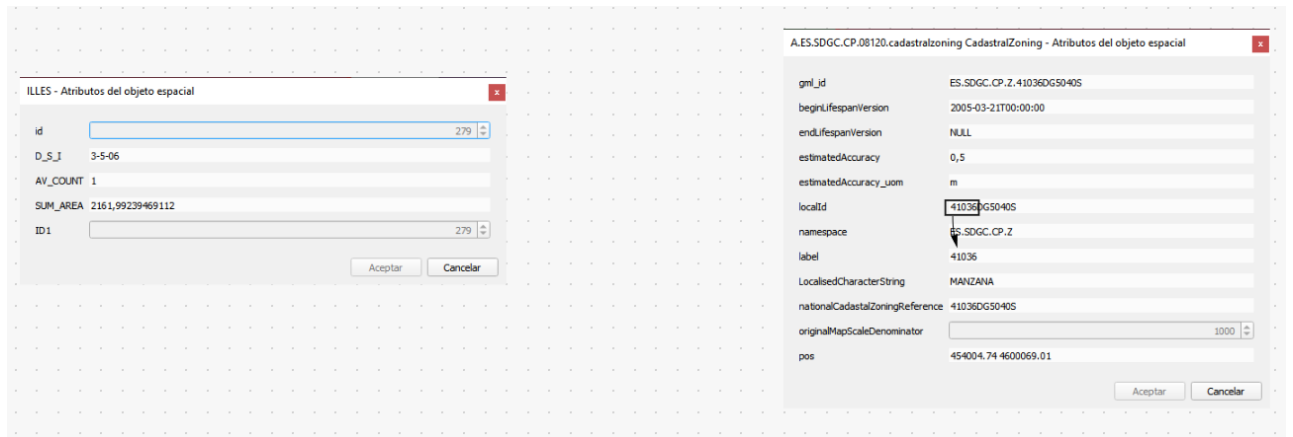


Figura 8. Relación entre tabla de manzanas (illes) de Mataró y CadastralZoning de Inspire. Fuente: elaboración propia.

Cabe destacar que en el label de la tabla Address encontramos los primeros dígitos de las referencias catastrales de las parcelas que pertenecen a esa manzana, lo que permite vincularse con las tablas anteriores de forma sencilla.

### 6.3.4. Edificios

#### 6.3.4.1. Usos de fincas

En Mataró, se obtenía una tabla descargando directamente de la sede del catastro en la que se detalla para cada finca, el uso que recibe como, por ejemplo, si esa parcela se trata de una vivienda. Esta tabla (que no tiene geometría) relaciona referencias catastrales de parcelas con su uso (Figura 9).

id	Ref_Cadastral	Quants	Us	Superficie_Cons	UTM
1	5741 4592713DF5949S	1	V	174	4592713

Figura 9. Ejemplo en FinquesUS en Mataró. Fuente: elaboración propia.

Para obtener este tipo de información de Inspire, hay una tabla de la que se puede obtener esta información: building. Esta tabla (Figura 10) tiene mucha otra información además del uso (incluyendo geometría y referencia catastral).

gml_id	ES.SDGC.BU.4592713DF5949S
beginLifespanVersion	2005-03-21T00:00:00
conditionOfConstruction	functional
beginning	1970-01-01T00:00:00
end	1970-01-01T00:00:00
endLifespanVersion	NULL
informationSystem	inmueble/OVListaBienes.aspx?rc1=45927138rc2=DF5949S
reference	4592713DF5949S
localId	4592713DF5949S
namespace	ES.SDGC.BU
horizontalGeometryEstimatedAccuracy	0,1
horizontalGeometryEstimatedAccuracy_uom	m
horizontalGeometryReference	footPrint
referenceGeometry	<input checked="" type="checkbox"/>
currentUse	1_residential
numberOfBuildingUnits	1
numberOfDwellings	1
numberOfFloorsAboveGround	NULL
documentLink	rarFotoFachadaGet?ReferenciaCatastral=4592713DF5949S
format	jpeg
sourceStatus	NotOfficial
officialAreaReference	grossFloorArea
value	174
value_uom	m2

Figura 10. Ejemplo en Building (INSPIRE). Fuente: elaboración propia.

Si bien es cierto que se puede establecer una cierta relación entre los distintos usos de las dos tablas, no hay una equivalencia directa entre ambas. En Mataró, el uso viene definido por un carácter que se corresponde con un determinado uso (Figura 11). En el caso de INSPIRE, encontramos una definición más auto explicativa del uso (Figura 12), aunque con menos variedad que en el caso de Mataró.

A	Magatzem
C	Comerç
E	Escola
G	Hotel
I	Indústria
K	Esportiu
M	Solar
O	Oficina
P	Administració
R	Religiós
T	Teatre, cinema
V	Vivenda
Y	Sense determinar

Figura 11. Usos en Mataró. Fuente: elaboración propia.

```

1_residential
2_agriculture
3_industrial
4_1_office
4_2_retail
4_3_publicServices

```

Figura 12. Usos en INSPIRE. Fuente: elaboración propia.

Actualmente el uso al que se le está dando más relevancia en los plugins de Mataró es V (Vivienda), por lo que probablemente la forma en la que se trata en Inspire (como “residential”) sea suficiente para estos casos.

Por otra parte, se debe resaltar que en INSPIRE también aparece un campo de la superficie construida (“value”).

#### 6.3.4.2. Plantas

En Mataró se obtenía una tabla descargando directamente de la sede del catastro en la que se detallan las distintas plantas que hay en una parcela. Cada planta muestra su superficie construida y comparte la misma referencia catastral con otras plantas si están en la misma parcela. Esta tabla (que no tiene geometría) relaciona referencias catastrales con las distintas plantas que hay en la parcela y su superficie construida.

id	Ref_Cadastral	Escala	Pis	Superficie_cons	UTM
18980	2296503DF5929N	1	00	260	2296503
18981	2296503DF5929N	1	01	176	2296503
18982	2296503DF5929N	1	02	176	2296503
18983	2296503DF5929N	1	03	176	2296503

Figura 13. Ejemplo en FinquesPlantas en Mataró. Fuente: elaboración propia.

Para obtener este tipo de información de Inspire, hay una tabla de la que se puede obtener esta información: BuildingPart. En esta tabla aparece el campo “numberOfFloorsAboveGround”, además de tener mucha otra información (incluyendo geometría y referencia catastral). En un edificio puede haber varias BuildingsParts, haciendo referencia a las distintas alturas que puede haber en un mismo edificio.

Atributos del objeto espacial

gml_id	ES.SDGC.BU.2296503DF5929N_part2
beginLifespanVersion	2005-03-21T00:00:00
conditionOfConstruction	NULL
localId	2296503DF5929N_part2
namespace	ES.SDGC.BU
horizontalGeometryEstimatedAccuracy	0,1
horizontalGeometryEstimatedAccuracy_uom	m
horizontalGeometryReference	footPrint
referenceGeometry	<input checked="" type="checkbox"/>
numberOfFloorsAboveGround	4
heightBelowGround	0
heightBelowGround_uom	m
numberOfFloorsBelowGround	0

Aceptar Cancelar

Figura 14. Ejemplo en BuildingPart (INSPIRE). Fuente: elaboración propia.

Se debe hacer mención a que en Inspire no aparece el detalle de la superficie construida para cada planta.

#### 6.3.4.3. Año de construcción

En Mataró, se obtenía una tabla descargando directamente de la sede del catastro en la que se detallan las distintas plantas que hay en una parcela. Cada planta mostrará su superficie construida y compartirán la misma referencia catastral si están en la misma parcela. Esta tabla (que no tiene geometría) relaciona referencias catastrales con las distintas plantas que hay en la parcela y su superficie construida.

id	Ref_Cadastral	Any_constr	UTM
1	3989175DF5938N	1900	3989175
2	3792821DF5939S	2011	3792821
3	3989173DF5938N	1860	3989173

Mostrar todos los objetos espaciales

Figura 15. Ejemplo en FinquesAnyConstruccio en Mataró. Fuente: elaboración propia.

Para obtener este tipo de información de Inspire, hay una tabla de la que se puede obtener esta información que también se estaba utilizando para saber los usos de una determinada parcela: building. Además de lo anteriormente comentado, esta tabla (Figura 16) tiene un campo denominado “beginning” en el que aparece el año de construcción.

Attribute	Value
gml_id	ES.SDGC.BU.3989175DF5938N
beginLifespanVersion	2020-11-11T00:00:00
conditionOfConstruction	functional
beginning	1900-01-01T00:00:00
end	1900-01-01T00:00:00
endLifespanVersion	NULL
informationSystem	nmueble/OVListaBienes.aspx?rc1=3989175&rc2=DF5938N
reference	3989175DF5938N
localId	3989175DF5938N
namespace	ES.SDGC.BU
horizontalGeometryEstimatedAccuracy	0,1
horizontalGeometryEstimatedAccuracy_uom	m
horizontalGeometryReference	footPrint
referenceGeometry	<input checked="" type="checkbox"/>
currentUse	1_residencial
numberOfBuildingUnits	1
numberOfDwellings	1
numberOfFloorsAboveGround	NULL
documentLink	rarFotoFachadaGet?ReferenciaCatastral=3989175DF5938N
format	jpeg
sourceStatus	NotOfficial
officialAreaReference	grossFloorArea
value	984
value_uom	m2

Figura 16. Ejemplo en Building (INSPIRE). Fuente: elaboración propia.

#### 6.4. Datos no extraíbles de INSPIRE

Los datos que tienen que ver más con el mapa (como las parcelas, manzanas, etc.) son fácilmente extraíbles de INSPIRE. No obstante, hay dos situaciones que ninguna fuente de datos basada en la Directiva INSPIRE soluciona actualmente: los datos con carácter sensible y datos con carácter específico.

En el caso de datos de carácter sensible (como el padrón o datos económicos) es difícil encontrar fuentes que faciliten este tipo de datos a la ciudadanía en general. En el caso de Mataró, el padrón que siempre han facilitado al CCU está anonimizado (sin nombres ni identificación) pero, sin embargo, se tienen datos de fecha de nacimiento, sexo, nivel de estudios y dirección de todas las personas empadronadas en el ayuntamiento por lo que siguen siendo datos sensibles que permiten identificar con relativa facilidad a una persona y su dirección exacta con sólo saber su fecha de nacimiento y cómo es su núcleo familiar. Por tanto, en este caso posiblemente sea necesario seguir dependiendo de forma directa de cada ayuntamiento, aunque se deberían establecer una serie de requisitos y recomendaciones a la hora de que faciliten estos datos al CCU para favorecer su integración con el servicio.

Por otra parte, aquellos datos que tienen un carácter más específico y localista (como la ubicación de farmacias, colegios, supermercados, etc.) aunque no se puedan extraer de fuentes bajo la Directiva INSPIRE, existe la posibilidad de acceder al portal de datos abiertos del ayuntamiento. Si en el ayuntamiento quieren conocer información detallada de este tipo de entidades, muy posiblemente ya la tengan registrada y colgada en su portal de datos abiertos. Esta solución está lejos de ser perfecta ya que implica buscar integraciones específicas para cada ayuntamiento y, además, depende de que el ayuntamiento disponga de un portal de datos abiertos o de algún servicio similar.

## 6.5. Propuesta del nuevo modelo de datos

### 6.5.1. Consideraciones previas

El nombre de todas las tablas y campos se plantea en minúscula debido a que en PostgreSQL hay distinción entre mayúsculas y minúsculas, ocasionando que sea necesario entrecomillar todos los nombres en las consultas SQL. Esta medida tiene como impacto que las consultas sean más limpias y fáciles de leer.

Los nombres planteados de las tablas y campos están en inglés para facilitar su internalización e interoperabilidad. No obstante, se pueden llegar a plantear soluciones que permitan multidioma en el caso de que así se requiera.

La propuesta aquí presentada establece una serie de campos mínimos necesarios para poder ofrecer un buen servicio en lo que se refiere al proceso de los datos que contribuyan a la toma de decisiones en el ámbito municipal. Por tanto, siempre se podrán añadir aquellos campos que se consideren precisos según las eventuales necesidades del ayuntamiento y, además, el modelo de datos se plantea desde una perspectiva iterativa que debe estar en constante revisión.

También cabe destacar la relevancia de la utilización de aquellos datos locales previos que sean de interés para el ayuntamiento como, por ejemplo, identificadores propios.

Algunos de los cambios más destacables con respecto al anterior son los siguientes.

El concepto de zona en el ámbito rural previamente mostraba una gran isla y ahora se tiene más resolución. Además, las zonas pasan a tener un identificador propio (cadastral\_zoning\_reference) que difiere del concepto anterior de D\_S\_I (distrito, sección, isla). No obstante, se mantiene este identificador siempre que sea posible en el campo local\_reference.

Algo parecido ocurre con address, ya que las anteriores direcciones contaban con la letra del portal y, en la nueva versión, no se tiene esta información. Por tanto, lo que anteriormente era el número de policía formado por el código de calle, número y letra, ahora se queda en código de calle y número. Cabe destacar que los códigos de calle proporcionados por la directiva INSPIRE difieren de los que había previamente, por lo que se ha de tener en cuenta este hecho.

En lo que respecta al uso, antigüedad y superficie construida por planta, anteriormente se utilizaban tres tablas mientras que ahora sólo son dos. Se conserva la tabla anteriormente llamada FinesPlantes (ahora building\_floor) y en el caso de la



antigüedad, uso y superficie construida de todo el edificio, se utiliza una única tabla que, a diferencia de las anteriores, tiene geometría.

A continuación, se muestra el diagrama de entidad relación del nuevo modelo (Figura 17).

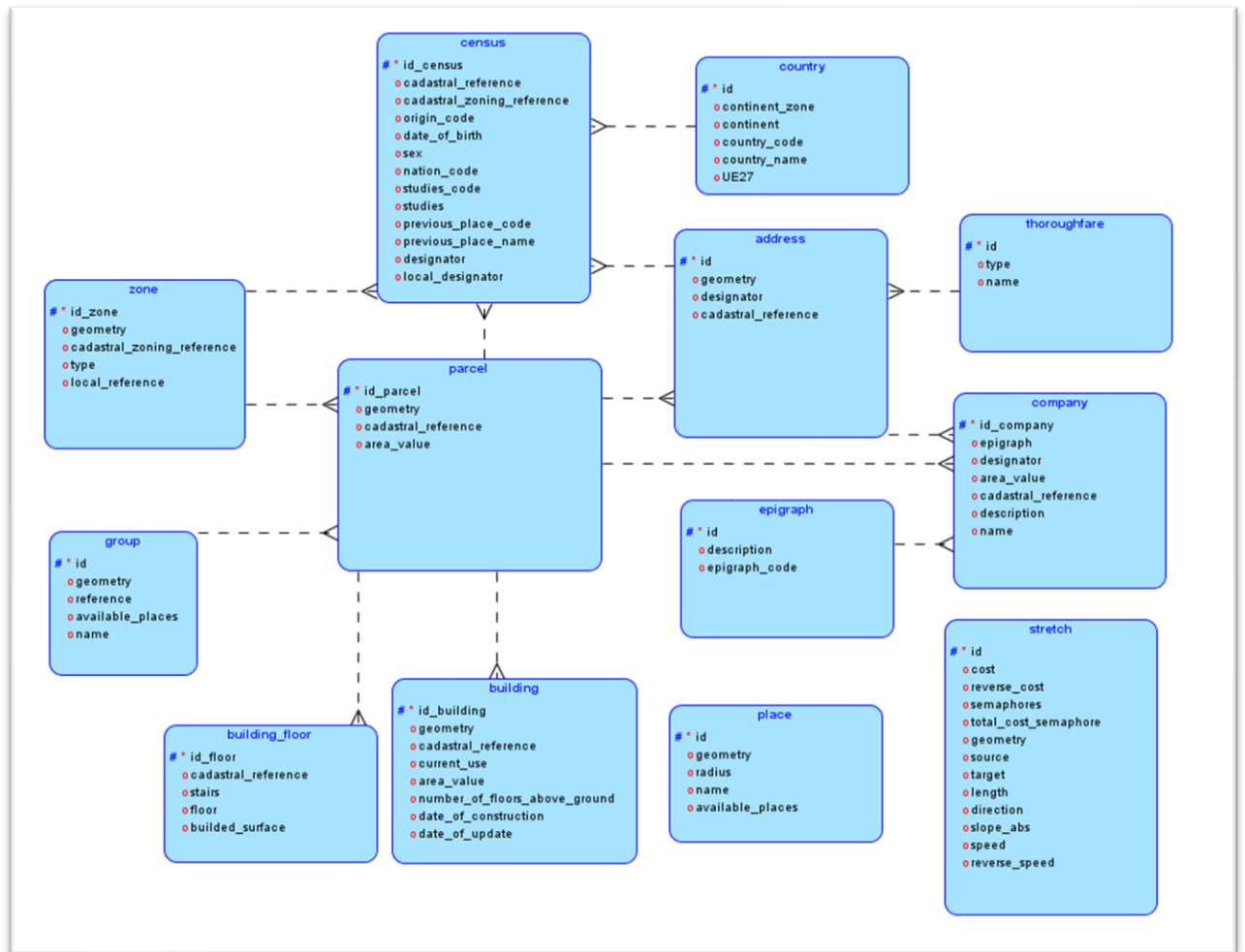


Figura 17. Modelo entidad relación. Fuente: elaboración propia

### 6.5.2. Parcel

La parcela catastral es el área individual de superficie de la tierra, sujeta a derechos reales de propiedad, homogéneos y de titularidad única (INSPIRE, 2016).

En esta tabla figuran las parcelas georreferenciadas de un determinado municipio.

Se obtiene de la sede del catastro, a través de un enlace que permite descargar los datos de “cadastral parcel” para un determinado municipio del Estado español siguiendo la directiva INSPIRE. En el archivo zip descargado hay una tabla en la que está un archivo de tipo gml (Geography Markup Language) denominado “cadastralparcel”. Para el caso de Mataró, sería el siguiente enlace:

<http://www.catastro.minhap.es/INSPIRE/CadastralParcels/08/08120-MATARO/A.ES.SDGC.CP.08120.zip>

En este caso, “08” es el código de provincia, “08120-MATARO” el código del municipio del catastro y “08120” el código del municipio.

A continuación, se plantean los campos mínimos que debería haber en “parcel”.

<b>parcel</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id_parcel	String	Identifier
geometry	Geometry	Geometry of the cadastral parcel. Polygon.
cadastral_reference	String	Thematic identifier at national level, generally the full national code of the cadastral parcel. In Spain there are two options of format. For urban parcels: 4394008 DF5949S, estate/parcel (7d), sheet (7c). For rural parcels: 08 120 A 008 00025, province (2d), municipality (3d), sector (1l), polygon (3d), parcel (5d).
area_value	Float	Registered area value giving quantification of the area projected on the horizontal plane of the cadastral parcel in square meters. It's a derived field from geometry

El campo id\_parcel es un identificador único autogenerado.

El campo geometry es del subtipo polígono y muestra la geometría de la parcela. Este tipo de dato está recogido en la extensión espacial de PostgreSQL denominada PostGIS.

El campo cadastral\_reference incluye la referencia catastral total de la parcela, siendo ésta distinta en cuanto a su formato dependiendo de si se trata de una parcela urbana

o rural. Este campo se obtiene directamente de “nationalCadastralReference” de la tabla CadastralParcel de INSPIRE.

El campo area\_value muestra el área de la parcela en metros cuadrados. Se trata de un campo derivado ya que esta información está implícita en el campo geometry, pero permite ahorrar el cálculo continuo de este valor.

### 6.5.3. Zone

En esta tabla se hace referencia a la zonificación de las parcelas en una agrupación superior la cual, generalmente, es cuadrangular y está delimitada por calles por todos sus lados.

Cabe resaltar que este concepto no aplica en todos los países que se adhieren a la directiva INSPIRE (como en Reino Unido, Finlandia o Noruega), pero se trata de un nivel al que aplican varios países con sus distintas particularidades (como en Francia, Dinamarca o en España). En el caso particular de España, se establecen las manzanas (o “blocks”) para el entorno urbano y los polígonos (o “polygons”) para el entorno rural.

En esta tabla figuran las manzanas y polígonos georreferenciados de un determinado municipio.

Se obtiene de la sede del catastro, a través de un enlace que permite descargar los datos de “cadastral zoning” para un determinado municipio del Estado español siguiendo la directiva INSPIRE. En el archivo zip descargado hay una tabla en la que está un archivo de tipo gml (Geography Markup Language) denominado “cadastralzoning”. Para el caso de Mataró, sería el siguiente enlace:

<http://www.catastro.minhap.es/INSPIRE/CadastralParcels/08/08120-MATARO/A.ES.SDGC.CP.08120.zip>

En este caso, “08” es el código de provincia, “08120-MATARO” el código del municipio del catastro y “08120” el código del municipio.

A continuación, se plantean los campos mínimos que debería haber en “zone”.

zone		
Property	Expect Type	Description
id_zone	String	Identifier
geometry	Geometry	Geometry of the cadastral zoning
cadastral_zoning_reference	String	Thematic identifier at national level, generally the full national code of the cadastral zoning. For urban zones: 43940 DF5949S, block (5d), sheet (7c). For rural parcels: 08 120 A 008, province (2d), municipality (3d), sector (1l), polygon (3d).
type	String	Descriptor of urban (BLOCK/MANZANA) or rural (POLYGON/POLIGONO) zones, in at least one official language of the European Union.
local_reference	String	Alternative thematic identifier provided by the city council (optional).

El campo id\_zone es un identificador único autogenerado.

El campo geometry es del subtipo polígono y muestra la geometría de la zona. Este tipo de dato está recogido en la extensión espacial de PostgreSQL denominada PostGIS.

El campo cadastral\_zoning\_reference incluye la referencia catastral de la zona, siendo ésta distinta en cuanto a su formato dependiendo de si se trata de una parcela urbana o rural. Cabe destacar que tiene en común con cadastral\_reference (de parcel) los 5 primeros dígitos de las parcelas que pertenecen a esa zona en el caso de las urbanas y los 9 primeros dígitos en el caso de las parcelas rurales. Con respecto a este campo de cadastral\_reference de parcel, se debe recalcar que cadastral\_zoning\_reference es prácticamente idéntico con la diferencia de que no aparece la parte que identifica a la parcela, debido a que es una unidad de medida más pequeña que la zona. Este campo se obtiene directamente de “nationalCadastralZoningReference” de la tabla CadastralZoning de INSPIRE.

El campo type hace referencia al tipo de zona de la que se trata (urbana o rural). Se debe resaltar que esta información se puede obtener también a través de

cadastral\_zoning\_reference, siendo urbana cuando tiene una longitud de 12 caracteres y rural cuando tiene una longitud de 9.

El campo local\_reference es de carácter opcional y su finalidad es preservar otros posibles identificadores propios que pudiera haber previamente en el ayuntamiento.

#### 6.5.4. Address

En esta tabla se hace referencia a los puntos de las distintas direcciones de un municipio, las cuales consisten en una jerarquía formada por varios componentes con distinto nivel de detalle: ciudad, nombre de la calle, la casa con su número o nombre

En esta tabla aparecen las direcciones georreferenciadas de un determinado municipio.

Se obtiene de la sede del catastro, a través de un enlace que permite descargar los datos de "address" para un determinado municipio del Estado español siguiendo la directiva INSPIRE. En el archivo zip descargado hay una tabla en la que está un archivo de tipo gml (Geography Markup Language) denominada "address". Para el caso de Mataró, sería el siguiente enlace:

<http://www.catastro.minhap.es/INSPIRE/Addresses/08/08120-MATARO/A.ES.SDGC.AD.08120.zip>

En este caso, "08" es el código de provincia, "08120-MATARO" el código del municipio del catastro y "08120" el código del municipio.

<b>Address</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id_address	String	Identifier
geometry	Geometry	Position of a characteristic point which represents the location of the address.
cadastral_reference	String	External, thematic identifier of the address spatial object, which relates address to parcel.
designator	String	Human readable designator or name that allows a user or application to reference and distinguish the address from neighbour addresses, within the scope of a thoroughfare name, address area name, administrative unit name or postal designator, in which the address is situated. Its format is like 00461 007, street (5d), number (3d).
local_designator	String	Designator of the address provided by the city council (optional).

El campo id\_address es un identificador único autogenerado.

El campo geometry es del subtipo punto y muestra la localización de la dirección. Este tipo de dato está recogido en la extensión espacial de PostgreSQL denominada PostGIS.

El campo cadastral\_reference es una clave ajena que hace referencia a la parcela en la que se encuentra la dirección.

El campo designator es un identificador único para cada uno de las direcciones. Este campo se obtiene del localId de la tabla Address de la sede del catastro, obteniendo los valores de código de calle y número. Por ejemplo, si obtenemos el localId 08.120.461.7.4191742DF5949S, se tomará el 461 (el código de calle) para añadir los ceros a la izquierda necesarios (hasta los 5 dígitos) y se tomará el 7 (el número de la dirección) para añadir los ceros a la izquierda necesarios (hasta alcanzar los 3 dígitos). De este modo, quedaría como designator el 00461007.

El campo local\_designator es de carácter opcional y su finalidad es preservar otros posibles identificadores propios que pudiera haber previamente en el ayuntamiento.

## 6.5.5. Building

En esta tabla figuran los edificios georreferenciados de un determinado municipio.

Se obtiene de la sede del catastro, a través de un enlace que permite descargar los datos de “building” para un determinado municipio del Estado español siguiendo la directiva INSPIRE. En el archivo zip descargado hay una tabla en la que está un archivo de tipo gml (Geography Markup Language) denominado “building”. Para el caso de Mataró, sería el siguiente enlace:

<http://www.catastro.minhap.es/INSPIRE/Buildings/08/08120-MATARO/A.ES.SDGC.BU.08120.zip>

En este caso, “08” es el código de provincia, “08120-MATARO” el código del municipio del catastro y “08120” el código del municipio.

<b>building</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id_buldng	String	Identifier
geometry	Geometry	Geometry of the building. Polygon.
cadastral_reference	String	External, thematic identifier of the address spatial object, which relates address to parcel.
current_use	String	Activity hosted within the building. This attribute addresses mainly the buildings hosting human activities.
area_value	Float	Registered area value of the building in square meters.
number_of_floors_above_ground	Integer	Number of floors above ground
date_of_construction	Date	Date of construction
date_of_last_update	Date	Date of last update

El campo id\_building es un identificador único autogenerado.

El campo geometry es del subtipo polígono y muestra la geometría del edificio. Este tipo de dato está recogido en la extensión espacial de PostgreSQL denominada PostGIS.

El campo `cadastral_reference` es una clave ajena que relaciona los edificios con la parcela a la que pertenecen.

El campo `current_use` muestra el uso que se le está dando a un determinado edificio. Los usos pueden ser residencial, agriculture, industrial, office, retail o public services (en el cual se incluyen edificios de educación, centros de salud, ocio o deportes).

El campo `area_value` muestra el área del edificio en metros cuadrados. Se trata de un campo derivado ya que esta información está implícita en el campo `geometry`, pero permite ahorrar el cálculo continuo de este valor.

El campo `number_of_floors_above_ground` hace referencia al número de plantas que hay por encima del nivel de suelo en ese edificio.

El campo `date_of_construction` muestra la fecha de construcción del edificio.

El campo `date_of_last_update` muestra la fecha de la última reforma registrada en el edificio.

#### 6.5.6. Census

En esta tabla aparece el padrón de habitantes anonimizados de un determinado municipio, indicando su sexo, nacionalidad, origen, nivel de estudios y dirección.

Esta tabla se obtiene a través de los datos facilitados por el ayuntamiento interesado.

census		
Property	Expect Type	Description
<code>id_census</code>	String	Identifier
<code>cadastral_reference</code>	String	External, thematic identifier of the address spatial object, which relates address to parcel. Related to parcel
<code>cadastral_zoning_reference</code>	String	External, thematic identifier of the address spatial object, which relates address to zoning. Related to zone.
<code>origin_code</code>	int	Country code of the origin of the person.



date_of_birth	Date	Date of birth
sex	int	1 = Man. 6 = Woman
nation_code	String	Country code of the nationality of the person.
studies_code	int	Code of the level of studies.
studies	String	Level of studies.
previous_place_code	int	Previous place code where the person was living. For people of the country, the code will be the previous city. For people that came for other countries will be the code of the country.
previous_place_name	String	Previous place where the person was living. For people of the country, the name will be the previous city. For people that came for other countries it will be the name of the country.
designator	String	External, human readable designator or name that allows a user or application to reference and distinguish the address from neighbour addresses, within the scope of a thoroughfare name, address area name, administrative unit name or postal designator, in which the address is situated. Its format is like 00461 007, street (5d), number (3d). Related to address
local_designator	String	Designator of the address provided by the city council (optional).

El campo id\_census es un identificador único autogenerated.

El campo cadastral\_reference es una clave ajena que relaciona los habitantes con la parcela a la que pertenecen.

El campo cadastral\_zoning\_reference es una clave ajena que relaciona los habitantes con la zona a la que pertenecen.

El campo origin\_code muestra el código del país de origen. Si el valor es 108, es España, si es 66 es extranjero. Se obtiene del campo HABCOPANA.

El campo date\_of\_birth muestra la fecha de nacimiento del habitante. Se obtiene del campo HABFECNAC.

El campo sex determina el sexo legal del habitante. Se obtiene del campo HABELSEXO.

El campo `nation_code` muestra el código del país de la nacionalidad del habitante. Se obtiene del campo `HABNACION`.

El campo `studies_code` muestra el código del nivel de estudios del habitante. Se obtiene del campo `HABNIVINS`.

El campo `studies` muestra el nivel de estudios del habitante. Se obtiene del campo `NINDESCRI`.

El campo `previous_place_code` muestra el código del lugar en el que estaba empadronado anteriormente el habitante. En el caso de ser nacional, será el código del municipio y, en caso de ser extranjero, será el código del país extranjero. Se obtiene del campo `HABCOMUNA`

El campo `previous_place_name` muestra el nombre del lugar en el que estaba empadronado anteriormente el habitante. Se obtiene del campo `HABNOMUNA`.

El campo `designator` es una clave ajena que relaciona los habitantes con la dirección a la que pertenecen.

El campo `local_designator` es de carácter opcional y tiene como objetivo preservar la forma en la que el ayuntamiento codificara los números de policía previamente.

#### 6.5.7. Place

Este concepto hace referencia a un conjunto de tablas que tienen una serie de características en común. Concretamente, se refiere a aquellos puntos que se pueden agrupar en función de uso o finalidad, como supermercados, colegios, farmacias, centro de salud, etc.

Este conjunto de tablas debe ser proporcionado por el propio ayuntamiento en función de sus necesidades, siendo idealmente obtenido a través de su propio portal de datos abiertos. Para extraer dichos datos, se puede hacer uso del plugin del CCU de Open Data.

A continuación, se proponen los campos necesarios para este tipo de tablas.

<b>place</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id	String	Identifier
geometry	Geometry	Location of the place. Point.
radius	int	Initial radius of influence of that place in meters.
available_places	int	Number of available places.
name	String	Name of the place.
type	String	Classification of the place.

El campo id es un identificador único autogenerated.

El campo geometry es del subtipo punto y muestra la localización del lugar. Este tipo de dato está recogido en la extensión espacial de PostgreSQL denominada PostGIS.

El campo radius hace referencia a la posible área de influencia inicial expresada en metros.

El campo available\_places hace referencia al número de plazas disponibles en dicho lugar. Por ejemplo, el número de plazas de un colegio.

El campo name hace referencia al nombre del lugar.

El campo type hace referencia al tipo de sitio que es. Los tipos posibles pueden ser supermercados, farmacias, CEIPs, etc.

#### 6.5.8. Building\_floor

Se trata de una tabla en la que se relaciona cada una de las plantas de un edificio con su correspondiente referencia catastral. Además, se proporciona la superficie construida para cada una de las plantas del edificio.

Esta tabla se obtiene de forma directa de la sede del catastro (a través de la firma electrónica) y no sigue la directiva INSPIRE.

<b>building_floor</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id_floor	String	Identifier
cadastral_reference	String	External, thematic identifier of the address spatial object, which relates address to parcel.
stairs	String	Code of stairs.
floor	String	Code of the floor.
buildded_surface	int	Buildded surface.

En esta tabla hay una fila para cada una de las plantas de una determinada referencia catastral.

#### 6.5.9. Stretch

Esta tabla contiene todos los tramos de calle que conforman el grafo de toda la red de calles de un determinado municipio.

<b>stretch</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id	String	Identifier
cost	float	Cost of going through this segment.
reverse_cost	float	Cost of going through this segment in the reverse way.
semaphores	int	Number of semaphores in the segment
total_cost_semaphore	float	Cost because of the presence of semaphores
geometry	Geometry	Geometry of the graph. Line.
source	int	Identifier of the source node.
target	int	Identifier of the target node.
length	float	Length of the segment.
direction	int	Direction of the segment. 1 if it goes up, 0 if it's straight, -1 if goes down.
slope_abs	float	Slope of the segment in absolute value.
speed	float	Speed going through this segment.
reverse_speed	float	Speed going through this segment in the reverse way.

## 6.5.10. Otras agrupaciones de parcelas

En este grupo podemos encontrar tablas correspondientes a una agrupación de parcelas que determina la extensión de los barrios de un municipio, los distritos postales, etc. Son tablas de carácter opcional que dependen de si el ayuntamiento en cuestión necesita trabajar a este nivel. A continuación se plantea qué estructura debería seguir cada una de estas tablas:

<b>group</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id	String	Identifier
geometry	Geometry	Geometry of the group. Polygon.
reference	String	Code or name of the group (name of the neighborhood, postal code, etc.).
name	String	Name of the place.

Estas tablas son facilitadas por el ayuntamiento en función de sus necesidades.

## 6.5.11. Company

Esta tabla contiene el listado de empresas de un determinado municipio

<b>company</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id_company	String	Identifier
epigraph	String	Epigraph of the activity of the company
designator	String	Human readable designator or name that allows a user or application to reference and distinguish the address from neighbour addresses, within the scope of a thoroughfare name, address area name, administrative unit name or postal designator, in which the address is situated. Its format is like 00461 007, street (5d), number (3d). Related to table address.
area_value	float	Registered area value giving quantification of the area projected on the horizontal plane of the parcel in square meters
cadastral_reference	String	Thematic identifier at national level, generally the full national code of the cadastral parcel. In Spain there are two options of format. For urban parcels: 4394008 DF5949S, estate/parcel (7d), sheet (7c). For rural parcels: 08 120 A 008 00025, province (2d), municipality (3d), sector (1), polygon (3d), parcel (5d). Related to table parcel.
description	String	Description of the activity.

name	String	Name of the company.
------	--------	----------------------

Esta tabla es facilitada por el ayuntamiento interesado.

#### 7.4.12. Country

En esta tabla aparecen todos los países con su correspondiente código, nombre, continente y zona continental a los que pertenecen.

country		
Property	Expect Type	Description
id	String	Identifier
country_code	String	Epigraph of the activity of the company
country_name	String	Human readable designator or name that allows a user or application to reference and distinguish the address from neighbour addresses, within the scope of a thoroughfare name, address area name, administrative unit name or postal designator, in which the address is situated. Its format is like 00461 007, street (5d), number (3d). Related to table address.
continent	String	Continent where is the country
continent_zone	String	Continent zone where is the country
UE27	int	1 if the country belongs to the European Union, 0 if not.

Esta tabla se relaciona con la tabla de census para obtener más información sobre el origen o nacionalidad del habitante.

#### 6.5.13. Thoroughfare

Esta tabla tiene un listado de todas las vías públicas de un determinado municipio, indicando su código, tipo y nombre.

Se obtiene de la sede del catastro, a través de un enlace que permite descargar los datos de “address” para un determinado municipio del Estado español siguiendo la directiva INSPIRE. En el archivo zip descargado hay una tabla en la que está un archivo de tipo gml (Geography Markup Language) denominada “ThoroughfareName”. Para el caso de Mataró, sería el siguiente enlace:

<http://www.catastro.minhap.es/INSPIRE/Addresses/08/08120-MATARO/A.ES.SDGC.AD.08120.zip>

En este caso, “08” es el código de provincia, “08120-MATARO” el código del municipio del catastro y “08120” el código del municipio.

<b>thoroughfare</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id	int	Code of street.
type	String	Type of thoroughfare (street, square, etc.).
name	String	Name of the thoroughfare.

Esta tabla se relaciona con la tabla address para poder obtener el nombre y tipo de vía a través del código de calle.

#### 6.5.14. Epigraph

Se trata de un listado de todos los epígrafes de la Agencia Tributaria en el que aparece su código y su descripción.

<b>epigraph</b>		
<b>Property</b>	<b>Expect Type</b>	<b>Description</b>
id	int	Identifier
description	String	Description of the epigraph
epigraph_code	int	Code of the epigraph

Esta tabla se relaciona con la tabla company para poder obtener la descripción del epígrafe a través del código del epígrafe.

#### 7.5. Implementación de la extracción y transformación

Para el desarrollo de la extracción y transformación de los datos se utiliza una base de datos alternativa en la que poder testear que la extracción y transformación de los datos ha sido exitosa y, además, poder realizar las modificaciones pertinentes en los plugins para que funcionen con el nuevo modelo de datos adecuadamente.

En esta parte del proyecto se han descargado las tablas provenientes de las nuevas fuentes de datos y se han estudiado los cambios que había con respecto a las fuentes anteriores.

Para poder realizar esta comparación, se ha tomado como referencia el padrón (denominado ahora como “census”) ya que esta tabla proviene del ayuntamiento y es común al modelo nuevo y al antiguo.

Las tablas a las que se les ha prestado especial atención son parcel, zone y address, ya que son las que se vinculan de forma directa con el padrón para poder realizar agregaciones y agrupaciones de forma rápida.

Estas comprobaciones se han hecho a través de las consultas que se muestran a continuación. Para la mayor parte de los habitantes no ha habido grandes inconvenientes, pero gracias a estas comprobaciones se han podido detectar aquellas situaciones problemáticas, que en todos los casos han sido anecdóticas.

```
-- parcel
SELECT distinct(c.national_cadastral_reference), total.id_census from
(select c.national_cadastral_reference from census c
EXCEPT
select p.national_cadastral_reference from census c join parcel p on
p.national_cadastral_reference = c.national_cadastral_reference) as c
join census total on c.national_cadastral_reference =
total.national_cadastral_reference order by 2;
```

Mediante la anterior consulta, se obtiene un listado de habitantes cuya referencia catastral no aparece en la tabla parcel. Hay varios motivos por los que ciertos habitantes pueden aparecer en este listado.

En algunos casos, se debía a que el antiguo utm\_total estaba incompleto en la tabla de parcelas antigua y, por tanto, se debía modificar el dato en la tabla de census para que fuese correcto.



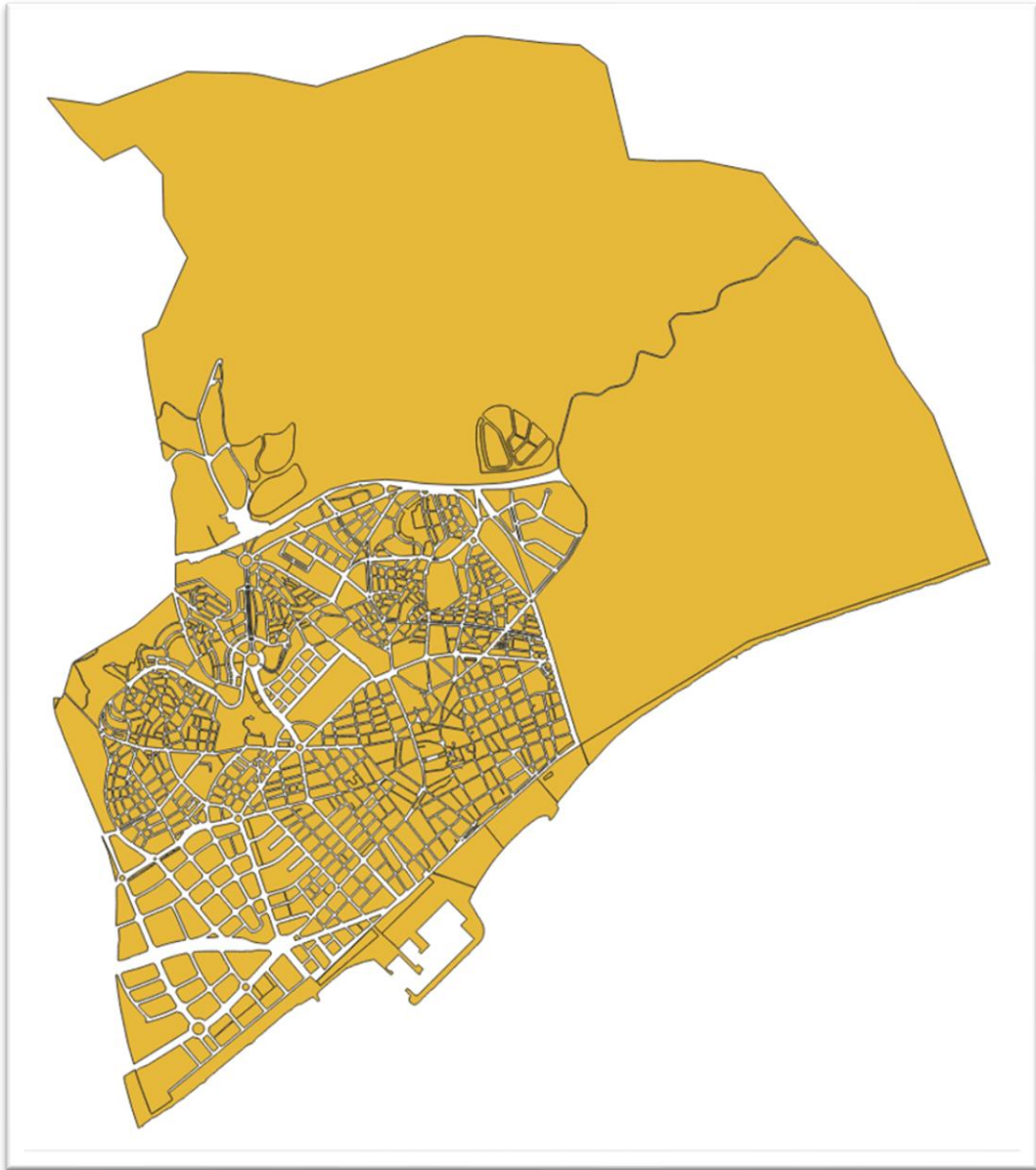
Existían otros casos en los que directamente la nueva referencia catastral no coincidía con la antigua debido a un cambio de criterio a la hora de formatear la referencia catastral como urbana o como rural.

También se han dado casos en los que la parcela no existía en la nueva tabla de parcelas, la parcela había pasado a ser más grande (lo que antes eran varias parcelas ha pasado a ser sólo una).

```
-- zone
SELECT distinct(c.national_cadastral_zoning_reference),
total.id_census from (select c.national_cadastral_zoning_reference
from census c
EXCEPT
select z.national_cadastral_zoning_reference from census c join zone z
on z.national_cadastral_zoning_reference =
c.national_cadastral_zoning_reference) as c join census total on
c.national_cadastral_zoning_reference =
total.national_cadastral_zoning_reference order by 2;
```

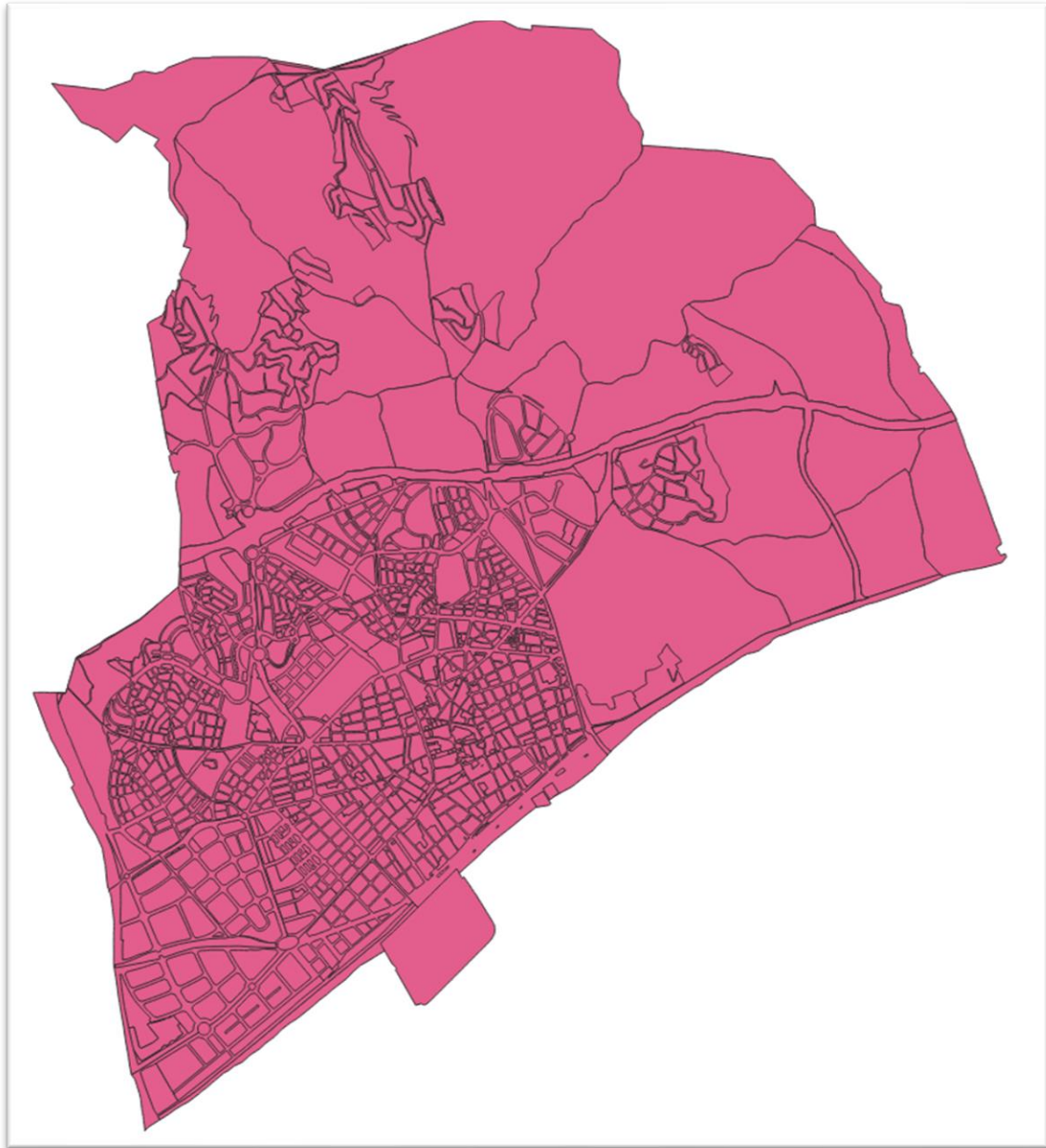
Con esta consulta, se obtiene un listado de habitantes cuya referencia de zona no aparece en la tabla zone.

El principal motivo por el cual no se encuentran las zonas es por el cambio de criterio que se ha establecido en las zonas rurales. En el modelo anterior, las parcelas rurales se relacionaban con islas de gran tamaño.



*Figura 18. Islas de Mataró en el modelo de datos antiguo. Fuente: elaboración propia.*

En la Figura 18 se puede apreciar cómo las zonas rurales de Mataró (al norte y este de la ciudad) están formadas por islas de gran tamaño, a pesar de haber una gran cantidad de parcelas, caminos y carreteras a lo largo de dichas islas.



*Figura 19. Islas de Mataró en el modelo de datos nuevo. Fuente: elaboración propia.*

En la Figura 19 se puede apreciar cómo las zonas rurales (los polígonos del nuevo modelo, al norte y este de la ciudad) ahora tienen un tamaño menor que se corresponde más con las parcelas, carreteras y caminos que hay en esas zonas.

Por tanto, debido a estas diferencias de criterio con respecto a qué es una zona o una isla en el ámbito rural, originaba que los habitantes de estas zonas no tuvieran asignada ninguna zona del nuevo modelo.

No obstante, debido a que las tablas de parcel y zone se relacionan de forma directa en el nuevo modelo, esta cuestión es fácilmente subsanable.

```
-- address
SELECT distinct(c.designator), total.id_census from (select
c.designator from census c
EXCEPT
select p.designator from census c join address a on a.designator =
c.designator) as c join census total on c.designator =
total.designator order by 2;
```

Finalmente, con esta consulta se obtiene un listado de habitantes cuya referencia de dirección no aparece en la tabla address.

En este caso se produce una diferencia significativa con el caso anterior. A través de la parcela se puede identificar la zona pero, a través de una parcela no se puede identificar una dirección, ya que en una parcela puede haber varias direcciones.

Por tanto, los problemas que han surgido con esta tabla tienen que ver con las diferencias que hay a la hora de definir qué es una dirección, así como que la codificación del nuevo modelo no tiene una equivalencia directa con el modelo anterior. Esto se debe a que los códigos de calle que se obtienen de INSPIRE no coinciden con los códigos de calle del modelo anterior (a la misma calle se le asigna un código distinto en cada fuente de datos). Esta cuestión tiene solución al establecer equivalencias entre los códigos de calle antiguos y los nuevos, pero provocan que el proceso sea más complejo.

#### 7.6. Plugin de configuración de las bases de datos de QGIS

Para poder automatizar las cuestiones vistas en el anterior apartado, se ha diseñado un plugin para QGIS que realiza el proceso de forma automática (Figura 20).

Figura 20. Plugin de configuración de bases de datos. Fuente: elaboración propia.

El funcionamiento de este plugin consiste en que el usuario puede fácilmente cargarlas tablas procedentes de INSPIRE de cualquier municipio del Estado español.

En los dos primeros desplegados, el usuario puede seleccionar su provincia y su municipio. El listado que se proporciona se incluyen los códigos de provincia y municipio, los cuales servirán para construir los enlaces que permitan acceder a los datos.

A continuación, el usuario debe indicar la base de datos y el esquema en el que se añadirán las tablas. Se recomienda utilizar un esquema distinto al de producción para evitar que se actualicen o se borren tablas por equivocación.

Después, el usuario puede seleccionar qué tablas quiere descargar. Adicionalmente, el usuario podrá indicar si existen tablas en la base de datos que se puedan utilizar como

referencia para la creación de las nuevas tablas, con el objetivo de preservar aquellos campos que pudieran ser de interés para el ayuntamiento en cuestión.

En cuanto el usuario haga click en el botón de “Iniciar”, comenzará el proceso de extracción (descargando los datos pertinentes), transformación (modificando las tablas y los campos para adecuarlos al modelo de datos) y, finalmente, se cargarán las tablas en la base de datos deseada.

Este plugin soluciona parcialmente la creación y la actualización del modelo de datos, ya que las tablas que se obtienen de otras fuentes de datos menos genéricas se deben tratar de forma manual o con otros métodos.

Cabe recalcar que se recomienda encarecidamente trabajar en un esquema distinto del esquema de producción (que en QGIS suele ser el esquema “public”) ya que este plugin puede reemplazar y borrar datos previos del esquema. Por otra parte, también es aconsejable realizar comprobaciones de que el resultado generado es correcto.

### 7.7. Open Data

Aquellas tablas de carácter específico y localista se deben obtener directamente del ayuntamiento ya que son datos que difícilmente se pueden conseguir de otra fuente. No obstante, cabe recalcar la tendencia de que cada ayuntamiento tenga su propio portal de datos abiertos y, consecuentemente, se trata de una oportunidad de cara a la obtención de datos para ofrecer servicios.

El problema principal de los portales de Open Data es que no están estandarizados aún y, por tanto, no existe un consenso entre los datos que se deben ofrecer, en qué formato, qué tipos, qué nombres, qué estructuras, etc.

Mientras no se desarrolle una normativa que establezca la forma en la que se ofrecen los datos, es necesario adecuarse a la realidad de cada ayuntamiento para transformar manualmente los datos y poder integrarlos en el modelo de datos.

No obstante, en uno de los anexos se incluyen unas pautas generales a la hora de diseñar un portal de datos abiertos. De este modo, los ayuntamientos pueden tener en cuenta estas recomendaciones para facilitar la colaboración con el CCU.

Esta propuesta dejará de ser útil cuando se legisle a nivel estatal o europeo la forma en la que los municipios deben ofrecer sus datos, pero hasta que ese momento llegue, es necesario que los portales de datos abiertos tengan unos mínimos criterios en común entre sí.





## 8. Conclusiones

La primera conclusión a destacar es la necesidad de establecer modelos de datos estandarizables en aquellos campos que hay un volumen de datos muy elevado y que proviene de multitud de fuentes distintas. Esta heterogeneidad de los datos y fuentes de datos son las que permiten que los sistemas de información geográfica puedan hacer análisis profundos en el ámbito de la gestión urbana, pero a la vez suponen un reto a la hora obtener los datos y normalizarlos para su posterior uso.

Es probable que muchas de las cuestiones que se abordan en este TFG cambien a medio y largo plazo, ya que muchos de los datos implicados dependen de entidades gubernamentales que tienen interés en el análisis de este tipo de datos y, además, existe una necesidad real de legislar para asegurar que haya una cierta estandarización y, de este modo, facilitar la interoperabilidad que permita la colaboración internacional.

En este último aspecto, es remarcable el trabajo propuesto por la directiva INSPIRE, la cual proporciona una serie de pautas que permiten asegurar que se sigan unos criterios comunes a nivel europeo. No obstante, se necesita más tiempo para que se implante de forma más generalizada en todos los países y que, además, permitan cubrir el mayor tipo de datos posible.

Con respecto al modelo de datos diseñado en este trabajo, se han solucionado los problemas más graves que presentaba el antiguo y, además, se ha reducido la deuda técnica del CCU. Sin embargo, los plugins del CCU aún pueden ser mejorados si se plantean arquitecturas más fáciles de mantener y si se hace un uso extensivo de tests, aunque estos propósitos excedían los límites de este TFG.

Durante el desarrollo del proyecto se ha invertido mucho tiempo en estudiar las fuentes de datos y su adecuación a un nuevo modelo de datos. Esto ha implicado buscar fuentes diversas y, en algunos casos, no obtener los resultados esperados. No obstante, ha sido un proceso necesario para poder diseñar el modelo de forma adecuada.

En el caso del padrón, no se ha conseguido obtener ninguna fuente alternativa a los ayuntamientos. Este hecho llama la atención debido a que los ayuntamientos no serían la fuente primaria del padrón, ya que estos datos provienen del Estado.

El motivo por el cual no se ha conseguido ninguna otra fuente de datos alternativa es que son datos de carácter sensible y, por tanto, su acceso está restringido. En el caso particular del CCU los datos que se conocen de los habitantes son su fecha de nacimiento, dirección, su nivel de estudios, su nacionalidad, su procedencia y su género. Los datos no contienen ninguna identificación de la persona (nombre o DNI), pero conociendo muy pocos datos de una persona es relativamente sencillo conseguir la dirección de una persona, por ejemplo. Por ello, para poder acceder a los datos del padrón de un municipio se requiere de firmar acuerdos de confidencialidad y protección de datos para asegurar que en todo momento se haga un uso ético y responsable de los mismos. En un futuro sería interesante que el CCU tratara de tener como interlocutor directo al INE para poder obtener estos datos de forma directa.

El proyecto ha concluido con la adecuación de los plugins del CCU al nuevo modelo, así como el diseño de un plugin que extrae, transforma y carga los datos teniendo en cuenta los nuevos requerimientos. Tras haber realizado el trabajo previo del diseño del modelo de datos, estas cuestiones se han podido resolver en poco tiempo.

## 9. Posibles ampliaciones

Con respecto a las posibles ampliaciones y nuevas líneas de trabajo que se pueden enmarcar en el CCU, a continuación, se mencionan las más destacables.

- Rediseñar la estructura de los plugins para que sean más mantenibles y escalables.
- Incorporar nuevas fuentes de datos al modelo de datos
- Añadir funcionalidades e indicadores relacionados con la gestión de riesgos urbanos.
- Incorporar datos sobre la gestión de residuos.
- Incorporar datos sobre el consumo de agua y electricidad.
- Incorporar datos obtenidos a través de la sensorización urbana.
- Propuesta de un portal de datos abiertos que integre el modelo de datos.



## 10. Bibliografía

- [1] I. Cerdà, Teoría general de la urbanización y aplicación de sus principios y doctrinas a la reforma y ensanche de Barcelona, Imprenta Española, 1867.
- [2] L. Permanyer, L'Eixample, 150 anys d'Història, Viena Edicions, 2008.
- [3] Y. Zheng, L. A. Tang, J. Yuan, J. Han, A. Leung, C. C. Hung y W. C. Peng, «Discovery of traveling companions from streaming trajectories.,» de *Proceedings of the 28th IEEE International Conference on Data Engineering*, 2012.
- [4] Y. Zheng, S. Ma y O. Wolfson, «A large-scale dynamic taxi ridesharing service,» de *Proceedings of the 29th IEEE International Conference on Data Engineering*, 2013.
- [5] Y. Zheng, L. Capra, O. Wolfson y H. Yang, «Urban Computing,» *ACM Transactions on Intelligent Systems and Technology*, vol. 5, nº 3, pp. 1-55, 2014.
- [6] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang y Z. Xu, «Destination prediction by sub-trajectory synthesis and privacy protection against such prediction,» de *Proceedings of the 29th IEEE International Conference on Data Engineering.*, 2013.
- [7] K.-t. Chang, Introduction to Geographic Information Systems (9th ed), McGraw-Hill, 2016.
- [8] A. B. Bondi, «Characteristics of scalability and their impact on performance,» de *Proceedings of the second international workshop on Software and performance*, 2000.
- [9] R. Kimball, The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Wiley, 2004.
- [10] Dirección General del Catastro, «Conjunto de Datos INSPIRE de la Dirección General del Catastro,» 7 2016. [En línea]. Available: <https://www.catastro.minhap.es/webinspire/documentos/Conjuntos%20de%20datos.pdf>.
- [11] RAE, «Diccionario de la lengua española. Definición de manzana.,» [En línea]. Available: <https://dle.rae.es/manzana>.
- [12] INSPIRE, «INSPIRE Data Specification on Addresses – Technical Guidelines,» 2014. [En línea]. Available: <https://inspire.ec.europa.eu/id/document/tg/ad>.
- [13] INSPIRE, «INSPIRE Data Specification on Cadastral Parcels – Technical Guidelines,» 2014. [En línea]. Available: <https://inspire.ec.europa.eu/id/document/tg/cp>.