

datos.gob.es

Vocabularios de representación de datos abiertos en Ciudades Digitales

Iniciativa
aporta



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

1º edición: enero 2018

AUTOR: Luis Polo Paredes

[Luis Polo es CEO de Triari Lab](#), empresa enfocada en ayudar a las organizaciones a extraer conocimiento de cada actividad y transformarlo en activos de negocio a través de la gestión y visualización de datos. Anteriormente desarrolló su actividad profesional en CTIC Centro Tecnológico, especializado en tecnologías de la información y la comunicación, y sede de la Oficina Española del W3C (www.w3c.es).

Este informe ha sido desarrollado en el marco de la Iniciativa Aporta desarrollada por el Ministerio de Energía, Turismo y Agenda Digital, a través de la Entidad Pública Empresarial Red.es, y en colaboración con el Ministerio de Hacienda y Función Pública. Los contenidos y los puntos de vista reflejados en esta publicación son responsabilidad exclusiva de su autor. El equipo Aporta no garantiza la exactitud de los datos incluidos en el estudio.

El uso de este documento implica la expresa y plena aceptación de las condiciones generales de reutilización referidas en el aviso legal que se muestra en: <http://datos.gob.es/es/aviso-legal>.

Iniciativa
aporta



red.es

Tabla de contenidos

1. Introducción	2
1.1. Norma UNE 178301:2015	4
1.2. Linked data como modelo de datos	5
2. Conjuntos de datos	8
2.1. Presupuestos municipales	9
2.2. Contratación pública	14
2.3. Catálogo de comercios	19
2.4. Agenda Cultural	21
2.5. Lugares de interés turístico	22
2.6. Estado del Tráfico	24
2.7. Transporte público: Autobuses regulares	28
2.8. Aparcamientos: Parkings públicos	30
2.9. Calidad del aire	32
2.10. Callejero	34
2.11. Población	36
3. Vocabularios genéricos	40
3.1. Vocabulario W3C: SKOS	40
3.2. Vocabulario W3C: RDF Data Cube	43
3.3. Ontología: Semantic Sensor Network	48
4. Discusión y conclusiones	51

1. INTRODUCCIÓN

La sociedad del siglo XXI ha evolucionado en paralelo al desarrollo de las tecnologías de la información. Las relaciones entre los ciudadanos y las instituciones se han visto afectadas de manera positiva por las oportunidades de comunicación gracias a la transformación de los medios digitales. Dentro de este contexto, se enmarca la iniciativa internacional “datos y gobierno abierto”, cuyo objetivo es promover la apertura de la información de las instituciones y administraciones y facilitar el acceso a los agentes sociales y económicos. En España, la apertura de información pública se articula a través del marco del ordenamiento jurídico correspondiente: la **Ley 37/2007**, de 16 de noviembre, sobre “reutilización de la información del sector público”, la **Ley 19/2013** de 9 de diciembre, de “transparencia, acceso a la información pública y buen gobierno”, así como los marcos legislativos autonómicos y ordenanzas municipales vigentes.

Por una parte, la apertura de información posibilita el establecimiento de mecanismos de transparencia, así como la creación de espacios para la participación y la colaboración de los ciudadanos, favoreciendo un sistema democrático más abierto. Por otra, la reutilización de la información producida por las Administraciones Públicas, potencia la oferta de nuevos productos y servicios digitales, dinamizando la actividad económica y empresarial, y generando en última instancia un valor para el ciudadano.

Los municipios (Ayuntamientos) son la entidad básica de la organización territorial del Estado y, como las administraciones más cercanas al ciudadano, son las responsables de la información del entorno local, sirviendo de cauce inmediato de participación de los ciudadanos en los asuntos públicos en las ciudades. Existen actualmente en España más de 8.000 municipios, variando entre ciudades con poblaciones que superan el millón de habitantes y municipios en entornos rurales que apenas llegan a 300 personas. La definición de políticas de datos abiertos uniformes dentro de la esfera municipal debe respetar esta heterogeneidad y dar una respuesta amplia al espectro de variedades de municipios en el territorio nacional.

Objeto del documento

El objeto del documento es el análisis de las recomendaciones de la Norma Técnica UNE 178301:2015 de AENOR en la que se establece una lista de conjunto de datos que se consideran prioritarios en las iniciativas de datos abiertos, acompañados de una serie de vocabularios recomendados para ser utilizados en su publicación.

La finalidad de este informe es realizar **un estado de la cuestión al respecto de los conjuntos de datos** identificados como prioritarios y evaluar **el grado de desarrollo y adecuación descriptiva de los vocabularios propuestos**.

Audiencia

La audiencia objetivo de este documento es:

- Responsables y agentes participantes a nivel nacional de la definición del marco normativo para la publicación de datos abiertos en los municipios.
- Responsables de la estrategia y las políticas de datos abiertos dentro de las ciudades.
- El personal técnico que forma parte de los equipos municipales de definición e implementación de estas políticas de apertura digital de datos.
- Cualquier persona interesada en conocer el estado de la cuestión en la apertura de datos en el contexto de las ciudades inteligentes (Smart Cities).

Nota para el lector

No se entrarán en detalles técnicos concretos sobre el paradigma publicación Linked Data y las tecnologías subyacentes, como la composición de URIs, los lenguajes OWL y RDF, etc. De este modo, por claridad expositiva, los conceptos de las ontologías y vocabularios que se recoge en este documento se escribirán simplemente en formato de letra cursiva, sin hacer referencia en la mayor parte de los casos a los espacios de nombres correspondientes.

Estructura del documento

El contenido de este documento se articula de la siguiente forma:

- En esta Sección 1, se introduce el objetivo de este documento y el alcance de la Norma Técnica UNE 178301:2015 para la representación y publicación de datos en el contexto de las ciudades inteligentes.
- En la Sección 2, se describen los conjuntos de datos identificados como estratégicos por esta Norma y se evalúan en diferentes aspectos los vocabularios recomendados.
- En la Sección 3, se describen los vocabularios genéricos del W3C: SKOS, RDF Data Cube y la ontología SSN, que son de aplicación para la publicación de datos en las ciudades digitales.
- En la Sección 4, se abordan las conclusiones del presente informe.

1.1. Norma UNE 178301:2015

La Norma Técnica de AENOR [UNE 178301:2015](#): “Ciudades Inteligentes. Datos Abiertos (Open Data)”, elaborada por el Comité Técnico AEN/CTN 178 Ciudades Inteligentes, es la referencia actual para la publicación de datos abiertos en el ámbito de las entidades municipales.

La norma proporciona directrices para facilitar la sistematización de los proyectos de datos abiertos y mejorar su gestión. Fundamentalmente se pretende que sea una referencia en manos de las Administraciones públicas que ayude a definir, documentar e implantar proyectos de datos abiertos. Además, y de manera fundamental, esta norma proporciona a las Administraciones Locales herramientas metodológicas para que puedan evaluar el grado de madurez de su proyecto de datos abiertos, facilitando la puesta y la mejora continua, teniendo en cuenta todos aquellos aspectos que repercuten en la sostenibilidad, calidad, eficiencia y eficacia del proyecto.

Como parte de esta Norma se han identificado un conjunto de datos estratégicos, de especial interés en el marco de las Administraciones locales, y que serán la base para la definición de las políticas municipales en todo el territorio nacional en los próximos años. Los conjuntos de datos son los siguientes:

- Catálogo de comercios
- Agenda Cultural
- Población (Padrón)
- Calidad del aire
- Contratos
- Presupuesto Inicial y ejecución
- Aparcamientos Parkings públicos
- Autobús regular: Horarios, Líneas, Paradas, Tarifas
- Estado Tráfico
- Lugares de Interés Turístico
- Callejero

El presente informe **analizará y describirá cada uno de estos conjuntos de datos** y **se evaluarán los vocabularios inicialmente propuestos** dentro del marco de esta norma UNE 178301:2015.

1.2. Linked data como modelo de datos

La publicación de datos en la Norma UNE 178301:2015 se enmarca en el paradigma Linked Data (o, como se conoce en español, “los datos enlazados”). En su propuesta conceptual más básica, este paradigma se refiere a construir una Web de datos, donde

la información tenga un nivel de representación suficiente que permita la construcción de relaciones y enlaces¹ entre recursos. Estos enlaces facilitan la navegación y la integración entre conjuntos de datos provenientes de diferentes fuentes.

El paradigma Linked Data es un conjunto de buenas prácticas para publicar y conectar datos estructurados en la Web². Esta propuesta se articula a través de un conjunto de tecnologías estándar del W3C:

- Las URIs, un mecanismo genérico para identificar entidades y conceptos, que funcionan de forma análoga a los nombres propios en el lenguaje natural. La principal ventaja con otros identificadores es que son globales (en la Web) y que son únicos (i.e., se refieren e identifican de forma unívoca y sin ambigüedad al mismo recurso).
- El protocolo HTTP, un mecanismo simple y universal que permite recuperar al recurso que identifica una URI. De esta forma, cuando “se resuelve” una URI es como si se estuviera llamando al recurso a través de su nombre propio, pudiendo acceder también a su descripción (su “significado”) construido a través de otros recurso, es decir, otras URIs.
- Modelo de datos RDF. Esta tecnología, piedra angular del stack tecnológico de la Web Semántica, es un framework para la descripción de recursos (entidades y objetos en la Web) que se apoya en los mecanismos de nombrado y acceso que hemos mencionado anteriormente (URIs y HTTP respectivamente). RDF no es más que una sintaxis para la expresión de tripletas: “sujeto - verbo/relación/enlace - objeto”, que nos permiten describir los recursos de forma sencilla y natural.

¹ Publicación de datos 5 estrellas.

² Para el lector con ganas de profundizar en el tema y conocer el estado actual de la tecnología Linked Data, se aconseja los siguientes documentos del W3C. Por un lado, la Recomendación Linked Data Platform 1.0 (2015). URL: <https://www.w3.org/TR/2015/REC-ldp-20150226/>; y además la Nota: Linked Data Platform Best Practices and Guidelines (2014). URL: <https://www.w3.org/TR/ldp-bp/>

RDF permite además un modelado abierto de los recursos, a través de esquemas y vocabularios compartidos que facilitan la integración de conjuntos de datos heterogéneos.

2. CONJUNTOS DE DATOS

En esta sección del informe, se describen y analizan los conjuntos de datos que se han identificado como estratégicos en la Norma Técnica UNE 178301:2015. Además, acompañando cada conjunto de datos, la Norma identifica inicialmente los vocabularios y ontologías recomendados para su uso en los procesos de publicación de datos.

El análisis de cada uno de estos conjuntos de datos se realizará con la estructura siguiente

1. **Descripción del conjunto de datos.** En esta parte, se describirá la naturaleza del conjunto de datos y los aspectos fundamentales para su publicación abierta. Se entrará en detalles de marco legislativo en aquellos casos en los que sea pertinente, ya que la normativa establece generalmente requisitos de modelado que deben ser cubiertos convenientemente por los vocabularios utilizados. De manera complementaria, y cuando sea posible, también se señalarán potenciales casos de uso.
2. **Formatos habituales de publicación.** Como parte del análisis del conjunto de datos, se identificarán los formatos habituales de publicación de cada tipo de información dentro de los portales web (y de datos) municipales.
3. **Vocabularios de representación.** Finalmente, se describirán y evaluarán los vocabularios recomendados por parte de la Norma UNE 178301:2015 para representar cada uno de los conjuntos de datos. En este análisis se estudiará la completitud y cobertura semántica del vocabulario respecto del conjunto de datos en cuestión. En algunos casos, por diversas razones, se indicarán otras alternativas de modelado que pueden ser tenidos en cuenta en desarrollos posteriores de proyectos municipales de datos abiertos.

2.1. Presupuestos municipales

2.1.1. Descripción del conjunto de datos

Los presupuestos municipales son, junto a la información relativa a la contratación pública, los conjuntos de datos más relevantes desde el punto de vista de la publicación de datos abiertos y la transparencia entre el gobierno municipal y los agentes locales, especialmente los ciudadanos.

El presupuesto municipal es el plan económico y financiero anual vinculado a la estrategia y políticas del gobierno del Ayuntamiento. Es una herramienta clave para la gestión pública, sirviendo a la planificación, programación, control y seguimiento y evaluación de las actuaciones municipales. Tiene carácter anual y contiene las previsiones de ingresos que se esperan obtener y los gastos planificados a realizar en el ejercicio correspondiente.

El presupuesto municipal está regulado por la legislación. En concreto, por lo establecido en:

- Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local.
- La Ley 39/1988, de 28 de diciembre, Reguladora de Haciendas Locales (modificada por el Real Decreto Legislativo 2/2004, de 5 de marzo).
 - El Real Decreto 500/1990 de 20 de abril, por el que se desarrolla el capítulo primero del título sexto de la Ley 39/1988, en materia de presupuestos.
- La Orden de 20 de septiembre de 1989 por la que se establece la estructura del Presupuesto de las Entidades Locales.

La presentación de los presupuestos municipales y, por tanto, su publicación en términos de datos abiertos, se realiza de varias formas:

1. **Por clasificación económica.** En este caso los ingresos y gastos estimados se describen según su naturaleza económica, siguiendo la jerarquía de elementos contables: **Capítulo > Artículo > Concepto (de gasto/ingreso)**. Esta clasificación económica es normativa y viene determinada por la legislación aplicable vigente.

Por ejemplo, la estructura de capítulos actual organiza los ingresos corrientes un Ayuntamiento de la siguiente forma:

- Capítulo I - Impuestos directos: IBI, IVTM, IAE, IIVTNU (o plusvalía).
 - Capítulo II - Impuestos indirectos.
 - Capítulo III - Tasas, precios públicos y otros ingresos: Licencias de apertura, ecotasa, licencias urbanísticas, precios públicos por acceso a instalaciones, sanciones por infracciones, etc.
 - Capítulo IV - Transferencias corrientes.
 - Capítulo V - Ingresos patrimoniales.
2. **Por clasificación funcional.** Esta presentación sólo aplica los gastos estipulados en los presupuestos municipales. La estructura funcional agrupa los gastos según los objetivos y las políticas que se pretenden financiar con los importes presupuestados. La jerarquía es la siguiente: **Área de gasto > Política de gasto > Grupo de Programas > Programa de gasto**.

Esta clasificación funcional también es normativa. Ejemplos de Área de Gasto son los siguientes:

- Área de gasto 0: “Deuda pública”. Comprende los gastos de intereses y amortización de la deuda pública y demás operaciones financieras similares.
- Área de gasto 1: “Servicios públicos básicos”. Incluye todos los gastos originados por los servicios públicos básicos. En esta área se incluyen **las cuatro políticas de gasto básicas**: seguridad y movilidad ciudadana, vivienda y urbanismo, bienestar comunitario y medio ambiente.

3. **Por clasificación orgánica.** En este caso, la estructura y clasificación muestra qué unidades orgánicas obtienen los ingresos y cuáles realizan los gastos. Por ejemplo, la Concejalía de Economía y Empleo. En este caso, la clasificación no es normativa, ni figura en su publicación oficial. Depende de la composición orgánica y organización interna de la Entidad Local.

2.1.2. Formatos habituales de publicación

El **Ministerio de Hacienda y Función Pública**, a través de la comunicación entre la Secretaría General de Financiación Autonómica y Local y las Entidades Locales, coordina todos los aspectos relacionados con la financiación y el intercambio de datos presupuestarios. Por tanto, los presupuestos de todas las Entidades Locales de España son accesibles a través del [portal web ministerial](#). Esta aplicación permite la consulta de los datos (a cierto nivel de agregación) en formato tabla HTML y su descarga como fichero PDF. Sin embargo **no permite su exportación a ficheros de datos**.

Por otro lado, los presupuestos también se publican en los **Boletines Oficiales** de la provincia o Comunidad Autónoma correspondiente. En este caso son documentos de texto con formato PDF, que también se pueden encontrar en formato HTML y, en algunos casos, ficheros XML. Sin embargo, en este último caso, aunque se trata de un formato estructurado, sólo hace referencia a la estructura del contenido, no a la estructura de los datos presupuestarios.

Desde el punto de vista de los **Ayuntamientos**, su publicación en las páginas web se realiza a través de:

- Contenido HTML.
- Documentos PDF.
- Ficheros CSV y Excel.

2.1.3. Vocabularios de representación

El vocabulario de referencia para la publicación de presupuestos municipales, señalado en la norma UNE 178301:2015, es "[Presupuesto](#)". Este vocabulario es una ontología OWL que facilita la descripción de los presupuestos anuales de una Entidad Local como Linked Data.

El vocabulario proporciona los mecanismos para estructurar un presupuesto municipal:

1. Descripción de los presupuestos a través los ingresos (concepto *Ingreso*) y gastos (concepto *Gasto*)
2. Descripción de los ingresos y gastos del presupuesto municipal según clasificación económica, proporcionando clases para cada uno de los elementos de la jerarquía de clasificación: *Capítulo* > *Artículo* > *Concepto* > *Subconcepto*.
3. Descripción de los gastos municipales según clasificación funcional, proporcionando clases para cada uno de los elementos de la jerarquía de clasificación correspondiente: *AreaGasto* > *PoliticaGasto* > *GrupoProgramas* > *ProgramaGasto*.
4. Conjunto de propiedades para asociar los diferentes elementos que componen el presupuesto, bien por organización funcional o económica, a la partida presupuestaria correspondiente.

Sin embargo, el alcance actual de la ontología, que se encuentra a fecha de redacción de este documento en fase inicial (versión 0.1), no cubre ciertos aspectos que es importante destacar para su evolución posterior, entre ellos:

1. **Presupuesto publicado vs. presupuesto ejecutado.** La ontología "Presupuesto" actualmente cubre la publicación de los presupuestos, es decir, la estimación de gastos e ingresos de un Ayuntamiento en una determinada anualidad. Sin embargo, no contempla la ejecución real del presupuesto. Esto es importante para:

- a. Conocer el grado de cumplimiento del presupuesto, según la planificación de ingresos y gastos proyectada.
 - b. Monitorizar y detectar desvíos a nivel presupuestario.
 - c. Comparar e identificar dónde se producen los desvíos entre presupuesto estimado y ejecutado (real).
 - d. Calcular indicadores de interés, que deben establecerse sobre datos de cumplimiento de gasto/ingreso, como nivel inversión por habitante, nivel de sostenibilidad financiera o
 - e. Comparar la evolución real de la ejecución presupuestaria en los últimos años.
2. **Mayor resolución temporal.** La especificación contable del ingreso y el gasto se realiza a nivel anual y trimestral. Esto es, desde luego, suficiente para una publicación de los presupuestos estimados. Sin embargo, este nivel de detalle no es suficiente en el caso de ejecución presupuestaria, es decir, de publicación de presupuestos reales, ya que limita la reutilización de los datos y su explotación posterior. Sería conveniente extender el vocabulario para publicar los registros contables de manera que los agentes reutilizadores de esta información puedan implementar los cálculos agregados correctamente según el nivel de resolución temporal deseado (mensual, trimestral, semestral y anual) y nivel de jerarquía dentro de los capítulos de ingresos y gastos.
3. **Incorporar los elementos correspondientes a las clasificaciones.** En el estado actual de la ontología se proporcionan las clases correspondientes a los niveles de la clasificación funcional y económica, pero no se incluyen los elementos de cada uno de estos niveles. Por ejemplo, no se recogen las nueve áreas de la clasificación funcional de gasto como recursos propios del vocabulario.

Con el objetivo de favorecer la reutilización y armonizar la publicación abierta de los presupuestos municipales, y dada su naturaleza cuantitativa, sería altamente favorable

alinearse el vocabulario “Presupuesto” con RDF Data Cube, la ontología recomendación del W3C para representar datos multidimensionales (véase Sección para una descripción más detallada de este vocabulario estándar).

Es importante destacar, de manera adicional, que el formato de **intercambio entre las Entidades Locales y el Ministerio de Hacienda** es **XBRL** (eXtensible Business Reporting Language), extendido con taxonomías específicas: PENLOC, LENLOC y TRIMLOC. XBRL es un formato XML para la representación de datos financieros y también se puede convertir en una alternativa real para la publicación de datos abiertos de los datos presupuestarios municipales.

2.2. Contratación pública

2.2.1. Descripción del conjunto de datos

Al igual que ocurre con los presupuestos municipales, la contratación pública de un Ayuntamiento es una de las materias de mayor interés para la ciudadanía y los agente socio-económicos del territorio. Por un lado, genera una parte muy importante del gasto público local mediante la contratación de obras, servicios y suministros. Por otro lado, la contratación implica la realización de un tercero de actividades que son responsabilidad del Ayuntamiento, lo que tiene repercusiones objetivas en el marco económico y el tejido empresarial local.

En este sentido, la publicación en formatos abiertos de los procedimientos de contratación pública municipal permite a los agentes locales:

- Conocer de forma transparente la competitividad de las empresas y los balances de contratación entre las diferentes empresas contratistas.
- Implementar procesos de matchmaking para encontrar contratos públicos que casen con las capacidades de las empresas, con el objetivo de maximizar la presentación de participantes y optimizar los procesos de concurrencia a las convocatorias de contratos.

- Detectar nichos de falta de oferta en el mercado, con el objeto de potenciar sectores de actividad económica en el marco de políticas locales.
- Proporcionar información a entidades como universidades, centros de estudios, medios de comunicación, y similares para el análisis y elaboración de estadísticas, la investigación de la relación entre lo público y lo privado, etc.

El análisis de la contratación pública es un campo tan amplio y complejo que no es posible tratarlo con el suficiente detalle dentro del alcance de este documento, de propósito introductorio y más genérico. Desde el punto de vista de la publicación abierta de datos, se señalan a continuación los aspectos principales que deben tener en cuenta los vocabularios para la representación de los contratos públicos:

1. Tipo y descripción del contrato
2. Descripción del objeto del contrato
3. Definición del procedimiento (administrativo) de contratación
 - Condiciones para la presentación de ofertas
 - Resolución del procedimiento de contratación

2.2.2. Formatos habituales de publicación

Los formatos habituales de publicación de los contratos públicos por las Entidades Locales dependen del tipo de contrato, ya que algunos procesos de contratación pública no están sujetos a publicación y convocatoria pública por exigencia legal.

De todas maneras, se pueden señalar los siguientes formatos de publicación más comunes:

- Boletines oficiales de provincias y Comunidades Autónomas, tanto como contenido no estructurado: formato PDF, como contenido semi-estructurado: XML.
- Páginas web: contenido HTML
- En el caso de contratos adjudicados, la información también se puede encontrar publicada en ficheros CSV y Excel.

2.2.3. Vocabularios de representación

Desde la norma UNE 178301:2015 se referencia el vocabulario “[Public Procurement Ontology \(PPROC\)](#)”, una ontología para la representación de contratos públicos. Este vocabulario cubre todo el amplio espectro de condiciones de representación para descripción de los procesos de contratación pública, de aplicación a las Entidades Locales. Se encuentra en versión 1.0 y está listo para ser usado con garantías en la publicación de contratos públicos como Linked Data.

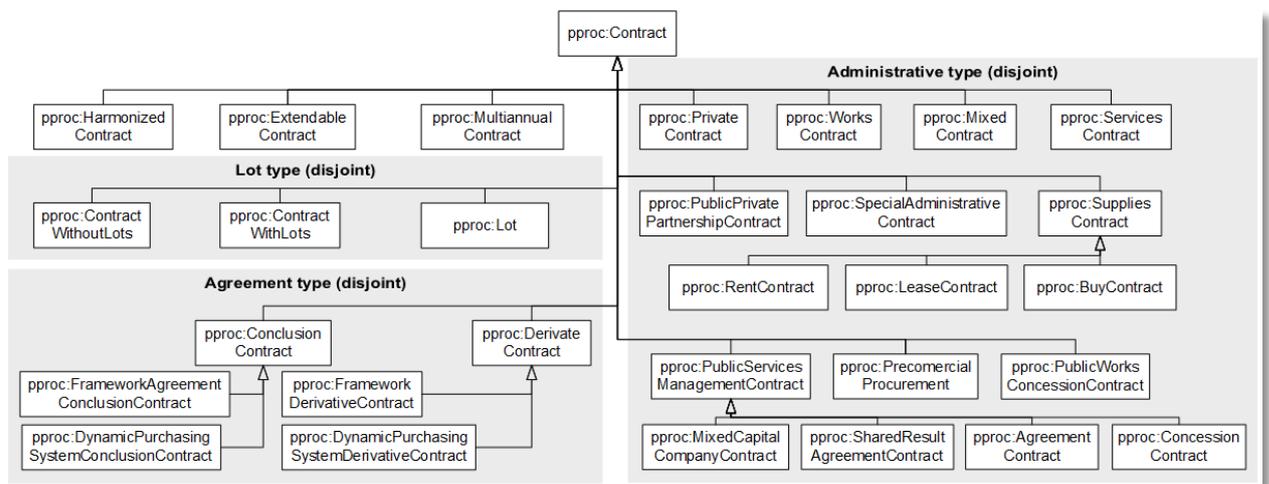


Figura 1. Tipos de contrato en PPROC

A continuación, se detalla brevemente el alcance del vocabulario PPROC para la descripción del contrato:

- Tipo de contrato, a través de una clasificación jerárquica de la clase *Contract*. El diagrama “Tipos de contrato en PPROC” ilustra los tipos contrato que recoge el vocabulario.
- Contratos de tipo administrativo.
- Lotes (subpartes de contratos, que tienen procedimientos de contratación independientes).
- Contratos de tipo Acuerdo.
- Descripción del objeto del contrato (*ObjectContract*)
- Condiciones económicas: precio (y posibles especificaciones adicionales).
- Condiciones de ejecución del contrato.
- Condiciones temporales del contrato.
- Especificaciones del procedimiento de contratación (*ContractProcedureSpecifications*)
- Condiciones de presentación de ofertas
- Requisitos de las ofertas
- Requisitos de los contratistas
- Órganos participantes
- Condiciones adicionales: financieras, legales, etc.

contratos públicos, incluyendo ciertas clases de PPROC para modelar la información relativa a los anuncios de ofertas dentro del proceso de contratación.

La eProcurement Ontology es de más reciente creación que la PPROC y está respaldada por la Comisión. Sin embargo, hay aspectos en los que PPROC profundiza y entra en más detalle que este nuevo vocabulario, como es el caso de los tipos de contrato. Estas cuestiones se abordarán en las siguientes fases de la eProcurement Ontology.

Dada esta disyuntiva actual de alternativas, se plantean posibilidades a explorar en el corto plazo.

- Armonizar y alinear PPROC dentro del marco que proporciona la eProcurement Ontology.
- Extender la eProcurement Ontology a partir del modelado ya existente en PPROC. Por ejemplo, en el caso concreto de los tipos de contrato.
- Revisar y ajustar el vocabulario conforme los cambios normativos en materia de contratos.

2.3. Catálogo de comercios

2.3.1. Descripción del conjunto de datos

La información sobre los comercios de un municipio es de interés para los agentes económicos y la ciudadanía. Por un lado, permite buscar y encontrar aquellos comercios que prestan servicios y venden productos (por ejemplo, tiendas de equipamiento electrónico o cafeterías). Por otro, también sirven para análisis de competencia y oferta de servicios en un territorio determinado, en el ámbito de estudios y análisis de geomarketing. Por tanto, es importante considerar dos aspectos en la categorización de los locales comerciales:

- Tipo de actividad económica del negocio
- Localización dentro del territorio municipal

Generalmente la información sobre los locales comerciales en una ciudad se encuentra almacenada en soporte electrónico en los sistemas municipales de registro administrativo y de gestión de licencias urbanísticas.

2.3.2. Formatos habituales de publicación

Los censos de locales y comercios municipales suelen estar generalmente disponibles a través de la página web del Ayuntamiento en los siguientes formatos:

- Contenido HTML, como parte de un sistema de búsqueda de comercios y servicios dentro de la ciudad integrado en el portal municipal.
- Ficheros CSV y Excel.
- Ficheros PDF.

2.3.3. Vocabularios de representación

El vocabulario propuesto por la norma UNE 178301:2015 es el "[Vocabulario para la representación de locales y zonas comerciales](#)" es una ontología sencilla para representar el tejido comercial de una ciudad. Básicamente consta de dos clases principales, una para identificar a los locales (*LocalComercial*) y otra a las zonas comerciales (*ZonaComercial*).

Respecto a las dos dimensiones importantes de representación de los locales comerciales:

- **Tipo de actividad económica.** El vocabulario ofrece la propiedad *tipoComercio* para describir el tipo de actividad que realiza el negocio. Aunque el vocabulario no proporciona una clasificación concreta, la norma UNE 178301:2015 señala la [Clasificación Nacional de Actividades Económicas \(CNAE\) en formato SKOS](#) para la descripción de la actividad económica.
- **Localización.** No se menciona esta parte en la versión actual del vocabulario. Sin embargo, se puede utilizar el vocabulario "Vocabulario para la representación de

direcciones postales en España” de la propia UNE 178301:2015 (Véase Sección 2.10 de este documento) con este propósito.

2.4. Agenda Cultural

2.4.1. Descripción del conjunto de datos

En las Agendas de los ayuntamientos se recoge información variada sobre los diferentes eventos y actividades que se realizan en la ciudad. Las agendas son de vital interés desde el punto de vista del ocio, la cultura y el turismo, ya que los programas culturales, espectáculos y otras actividades similares son accesibles desde estas agendas.

Además, estas agendas recogen otra información adicional de ámbito local sobre la organización y eventos que transcurren en la ciudad importante para la ciudadanía organizadas por el Ayuntamiento o asociaciones locales, como, ejemplo, talleres, charlas o presentaciones públicas de algún contenido.

2.4.2. Formatos habituales de publicación

El formato habitual de publicación de las agendas culturales en las páginas web de los ayuntamientos es el siguiente:

- Fichero PDF
- Página HTML
- Canal RSS
- Existen Ayuntamientos que publican las agendas en formatos orientados a su procesamiento y reutilización: CSV, XML y JSON.

2.4.3. Vocabularios de representación

La norma UNE 178301:2015 indica el siguiente vocabulario para la representación de estas Agendas: [“Vocabulario para la representación de eventos de una agenda cultural”](#).

Esta ontología se fundamenta en el concepto central *Evento*, entendiendo una agenda como una lista de eventos.

Un *Evento* tiene varias propiedades que permiten definir su contenido. A continuación se indican las más importantes:

- Propiedad *equipamiento*: Lugar en el que se celebra el evento
- Propiedad *lugarInscripcion*: Lugar de inscripción para la asistencia al evento (por ejemplo, para la compra de entradas).
- Propiedad *tipoAccesibilidad*: Criterios de accesibilidad relacionados con el evento
- Propiedad *tipoEvento*: Tipo de evento
- Propiedad *tipoPublico*: Indicaciones sobre el tipo de público recomendado (o permitido) para el evento. Por ejemplo, “infantil”, en el caso de que sea una obra de teatro para niños.
- Propiedad *fechaFin*: Fecha (estimada) de fin de la actividad
- Propiedad *fechaInicio*: Fecha de inicio de la actividad
- Propiedad *horaFin*: Hora de fin de la actividad
- Propiedad *horaInicio*: Hora de inicio de la actividad

2.5. Lugares de interés turístico

2.5.1. Descripción del conjunto de datos

El turismo es una de las industrias más importantes del mundo y uno de los motores fundamentales de la economía española. En las ciudades, especialmente aquellas con un patrimonio cultural elevado o aquellas que ofrecen amplias expectativas de ocio, el turismo es una de las grandes fuentes de ingresos, no sólo para las arcas municipales,

sino para todo el tejido empresarial relacionado, enriqueciendo la variedad de servicios de ciudad.

Además de su repercusión en el ámbito económico, los datos sobre los lugares de interés turístico (también conocidos como POIs en inglés⁴) son de importancia para la ciudadanía en general, ya que afectan a su vida diaria y a su relación con la ciudad. Algunos lugares turísticos están sometidos a políticas de protección artística y medioambiental. Esto puede afectar a las normas estéticas de edificios o a los criterios de recalificación del suelo, recogidas en los planes urbanos de ordenación del territorio. Además la afluencia y movilidad de los turistas según los lugares de interés de la ciudad tienen impacto sobre los servicios públicos municipales y sobre las infraestructuras. También influye en los precios de vivienda: adquisición y alquiler, según proximidad con polos de atracción de turistas. Conocer estos datos, relacionados con la actividad turística, es fundamental para encontrar el equilibrio y la convivencia entre todos los agentes sociales y económicos en el ámbito de la ciudad.

2.5.2. Formatos habituales de publicación

Los lugares de interés turístico, cuando se refieren a bienes patrimoniales o protegidos, forman parte de catálogos de bienes de interés cultural (BICs), que generalmente se recogen en documentos. De forma general podemos asegurar que la información relativa a la publicación de los lugares de interés turístico se realiza en los siguientes formatos:

- Documentos: ficheros PDF
- Contenido HTML, con las fichas de cada lugar de interés turístico.

2.5.3. Vocabularios de representación

⁴ POI: Point of Interest

En la norma UNE 178301:2015 se hace referencia al “[Vocabulario para la representación de lugares, zonas e itinerarios de interés turístico](#)” para la representación de los lugares de interés turístico en una entidad municipal.

El objetivo de este vocabulario es la representación de lugares de interés turístico utilizando tres conceptos centrales:

- Lugar de Interés Turístico (*LugarInteresTuristico*).
- Zona Turística (*ZonaTuristica*): zona geográfica de relevancia, y que típicamente incluye lugares de interés turístico
- Itinerario (*Itinerario*): ruta o recorrido, a través de lugares de interés turístico.

Actualmente la ontología se encuentra en versión 0.1 y sus autores advierten del estado inicial actual, así que se esperan nuevas versiones de la ontología para cubrir los casos de uso más habituales en las necesidades de representación de información turística.

Por otro lado, existen además otras ontologías y vocabularios en el ámbito turístico que pueden servir como complemento o alternativa a este vocabulario. Por ejemplo, el vocabulario: “[Turismo](#)” recoge un amplio conjunto de tipos de recursos turísticos: distinguiendo desde Espacios Naturales (como reservas naturales, paisajes protegidos, etc.), recursos hosteleros (restaurantes, hoteles, etc.), recursos culturales (monumentos, museos, palacios de congresos, etc.) hasta recursos generales de ocio (casinos, cines, eventos, etc.).

2.6. Estado del Tráfico

2.6.1. Descripción del conjunto de datos

Los datos de tráfico son la base para llevar a cabo las actividades de prognosis, planeamiento y mejora por parte de los Ayuntamientos en diferentes ámbitos de la planificación urbana: desde la proyección de infraestructura viaria, gestión del tráfico, definición de políticas de servicio de transporte público, etc. Del mismo modo, esta información es también útil para el ciudadano, con el fin de conocer, el uso y el estado

de las vías urbanas e inter-urbanas. Este interés no sólo aplica a cuestiones como la planificación de transporte personal o la búsqueda de mejores rutas, sino también a la hora de construir demanda social basada en datos sobre la inversión municipal en ciertas infraestructuras: por accidentes, estado del asfaltado en vías con alta densidad de vehículos, etc. También es importante para otros agentes socio-económicos, por ejemplo, a la hora de establecer ubicación de nuevos negocios y zonas comerciales.

Para conocer el comportamiento de la circulación, se utilizan **equipamientos y sistemas de medición aforo**. En la actualidad existe una gran variedad de dispositivos para recoger datos sobre el estado del tráfico, basados en diferentes tecnologías: lazo inductivo, visión artificial (análisis de imágenes), de radar de microondas, de infrarrojos, etc. A pesar de las particularidades de cada uno de estos tipos de dispositivo, la actividad de aforar, que es la acción de medir la cantidad y el valor de los vehículos circulantes para determinar, evaluar y calcular la capacidad de una vía, se realiza estudiando las (casi siempre) mismas variables:

1. El número de vehículos (intensidad).
2. La velocidad de circulación.
3. El tipo de vehículo (ligero o pesado).
4. La ocupación de la vía como porcentaje del tiempo de presencia.
5. Sentido de la circulación del vehículo.
6. Carril de la vía que ocupa el vehículo.

Otro componente importante de la información de tráfico es la identificación y georreferenciación de la vía de circulación.

2.6.2. Formatos habituales de publicación

La información de tráfico proviene generalmente de diferentes sistemas de medición de aforos. De manera independiente del sistema de captación utilizado, la mayor de los

dispositivos suelen permitir la exportación a ficheros CSV o un formato similar: de texto plano.

Su disponibilidad en los portales de datos de los Ayuntamientos se suele realizar con una frecuencia de actualización entre 5 minutos y varias horas, dependiendo de la entidad municipal.

- CSV o formato plano de datos similar (.txt)
- KML (formato mapa).

2.6.3. Descripción técnica del vocabulario/modelo

Desde el punto de vista de esquemas para la representación de datos, la norma UNE 178301:2015 recomienda el "[Vocabulario para la representación de datos sobre tráfico](#)". Esta ontología introduce las siguientes entidades para la representación del tráfico de manera agregada:

- Observaciones sobre el tráfico (*TrafficObservation*)
- Intensidad del tráfico (*TrafficIntensity*): para representar la magnitud que se está midiendo. En este caso, sólo intensidad de tráfico
- Tipo de dispositivo (*TrafficIntensitySensor*)
- Tramo de la vía (*TramoVia*): se refiere al elemento vial que se está midiendo, según se define en el vocabulario de callejero.

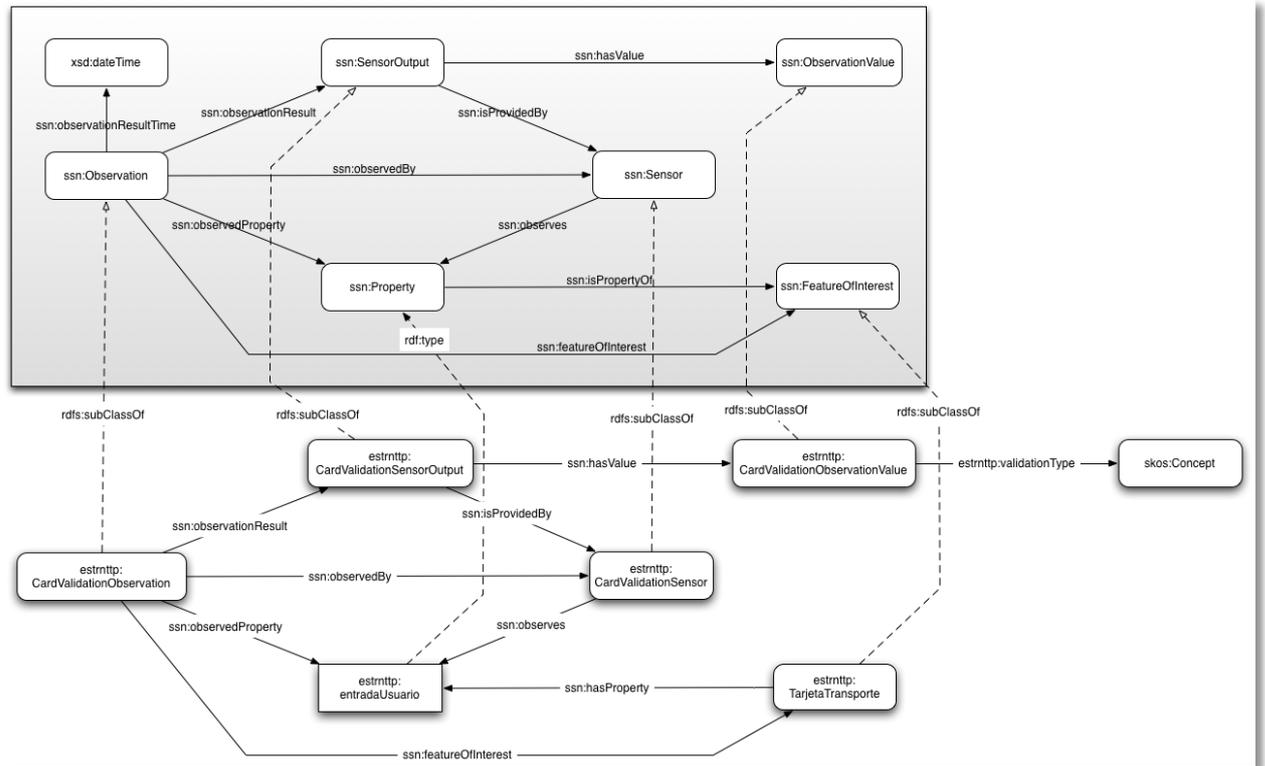


Figura 3.- Diagrama de clases de la ontología

Es importante señalar dos cuestiones al respecto de este vocabulario:

1. La versión actual de la ontología sólo contempla la intensidad. Es necesario extender la ontología para cubrir otros aspectos importantes del estado del tráfico, por ejemplo, el número de vehículos pesados en una zona viaria.
2. Este vocabulario es una extensión del vocabulario "[Semantic Sensor Network Ontology](#)" del W3C. Sin embargo, existe una versión más reciente de este vocabulario y es importante que las próximas ediciones del "Vocabulario para la representación de datos sobre tráfico" sean compatibles con la última versión de la ontología del W3C (véase Sección 3.3 de este documento para más información acerca de esta ontología).

2.7. Transporte público: Autobuses regulares

2.7.1. Descripción del conjunto de datos

El transporte público se refiere a los servicios que proporciona una administración municipal dentro de su territorio de competencia para garantizar la movilidad de las personas. El transporte facilita que las personas puedan desplazarse dentro del espacio municipal, sin necesidad de vehículo propio, a su trabajo, al hospital, espacios de ocio, etc.

Desde el punto de vista de la publicación de datos, la información que es necesario cubrir está ligada a las líneas regulares de los medios de transporte:

- Horarios
- Tarifas
- Líneas
- Paradas
- Accesibilidad

En la norma UNE 178301:2015, se establece como prioridad las líneas de autobuses regulares. Sin embargo, la lista de medios de transporte públicos inter-urbanos es más amplia y conviene tenerla presente para proporcionar una solución que se pueda extender a cualquier tipo de entidades municipales, especialmente a las ciudades de gran tamaño, donde la oferta de transporte público es mucho más variada, como se indica a continuación:

- Transporte terrestre
- Autobuses
- Alquiler de bicicletas
- Transporte por ferrocarril:

- Metro y tren ligero
- Tranvía
- Tren de cercanías
- Transporte marítimo
- Lanchas y transbordadoras

2.7.2. Formatos habituales de publicación

La información relacionada con el transporte público suele estar disponible en las páginas web de las Entidades Locales, con los siguientes formatos:

- Contenido HTML
- Infografía PDF con el plano de líneas regulares del tipo de transporte: líneas de autobuses, metro, etc.
- Ocasionalmente puede estar publicada en ficheros CSV y Excel para su explotación posterior.

2.7.3. Descripción técnica del vocabulario

El vocabulario de referencia para la norma UNE 178301:2015 es el “[Vocabulario para la representación de medios de transporte públicos](#)”. Este vocabulario es una extensión de la ontología del gobierno británico para la representación de datos relacionados con el transporte público: transport.data.gov.uk.

Esta ontología se ha usado para la publicación oficial por parte del Boletín Público Nacional (National Public Transport Gazetteer) y de la Red Nacional de Transporte Público (National Public Transport Access Network) de la siguiente información en el Reino Unido:

- Base de datos de NaPTAN (The National Public Transport Access Node): el sistema nacional británico para identificar y registrar todos los puntos de acceso existentes al transporte público: buses, metros, tranvías, etc.
- Lista de aeropuertos en Reino Unido.
- Lista de estaciones de tren en Reino Unido.
- Observaciones de tráfico.

En la versión actual de la ontología (versión 0.1), el vocabulario extiende la ontología para los siguientes casos de uso:

- Línea de bus (*LineaBus*)
- Línea de metro (*LineaMetro*)
- Línea de metro ligero (*LineaMetroLigero*)
- Línea de tranvía (*LineaTranvia*)
- Línea de tren (*LineaTren*)

2.8. Aparcamientos: Parkings públicos

2.8.1. Descripción del conjunto de datos

La información relativa a los espacios de estacionamiento públicos es de alto interés para la ciudadanía. Los aparcamientos municipales pueden ser de tres tipos:

- Aparcamientos **públicos**: A los que se puede acceder a través del pago de la tarifa correspondiente
- Aparcamiento para **residentes**: a los que pueden acceder las personas físicas que, disponiendo de vehículo, residan de hecho y estén

empadronadas en el área de influencia que se establece para cada aparcamiento.

- Aparcamientos **mixtos**: que contienen plazas de las dos categorías anteriores (residentes y público en general).

De manera adicional, con el fin de racionalizar el uso del estacionamiento de vehículos en las vías públicas, especialmente en zonas urbanas con alta densidad de población y vehículos, existen zonas públicas de aparcamiento reguladas. Estas zonas tienen varias denominaciones, según la ciudad, como: ORA (Ordenanza Reguladora de Aparcamiento), OLA (Ordenanza Limitadora del Aparcamiento), etc.

2.8.2. Formatos habituales de publicación

Los datos sobre estacionamientos municipales se publican habitualmente como:

- Ficheros CSV y Excel
- A través de mapas en los portales web municipales, para geo-posicionar los aparcamientos.

2.8.3. Vocabularios de representación

La norma UNE 178301:2015 señala para la representación de los aparcamientos municipales el “[Vocabulario para la representación de equipamientos](#)”. Esta ontología se compone de una sola clase para la representación de los estacionamientos públicos: *Aparcamiento* y un conjunto de propiedades para la descripción de estos aparcamientos:

- Relación entre el aparcamiento y el entorno urbano: *tipoEquipamiento*. Por ejemplo, si es un parking en un museo, centro cívico, etc.
- Tipo de aparcamiento: *titularidadAparcamiento*.
- Accesibilidad: *accesibilidad*.

Aunque la ontología está en un estado de desarrollo inicial, es importante remarcar que la clase *Aparcamiento* está definida como una subclase del concepto [ParkingFacility](#) del conjunto de vocabularios de Schema.org. En este vocabulario los mecanismos para la descripción de los estacionamientos tienen mayor alcance y son más ricos expresivamente. Estos mecanismos son re-utilizables dentro del vocabulario.

2.9. Calidad del aire

2.9.1. Descripción del conjunto de datos

El análisis de la calidad del aire es necesario para la calidad de vida, la salud de las personas y el medio ambiente. Es de gran importancia conocer qué contaminantes afectan y de dónde provienen, para poder avisar a la población y aplicar medidas preventivas y correctoras. A continuación, se indica la normativa actual de aplicación en España sobre calidad y evaluación del aire, que fija los principios esenciales en materia de prevención, vigilancia y reducción de la contaminación atmosférica, así como de control, inspección, vigilancia y seguimiento:

1. El Real Decreto 102/2011, de 28 de enero, **relativo a la mejora de la calidad del aire**, que transpone al ordenamiento jurídico español el contenido de la Directiva 2008/50/CE, de 21 de mayo de 2008, relativa a **la calidad del aire ambiente y a una atmósfera más limpia en Europa**.
2. Ley 34/2007, de 15 de noviembre, **de calidad del aire y protección de la atmósfera**, que establece el deber de las comunidades autónomas y en su caso, entidades locales (como los Ayuntamientos), de disponer de estaciones, redes y otros sistemas de evaluación de la calidad del aire suficientes para el cumplimiento de sus obligaciones.

La evaluación de la calidad del aire ambiente se realiza, dependiendo del nivel de los contaminantes con respecto a sus valores legislados, utilizando mediciones fijas, técnicas de modelización, campañas de mediciones representativas, o una combinación

de algunos de estos métodos. En las zonas definidas se evalúa la calidad del aire para los siguientes elementos:

- Dióxido de azufre (SO₂)
- Dióxido de nitrógeno y óxidos de nitrógeno (NO₂, NO_x)
- Partículas (PM₁₀ y PM_{2,5})
- Plomo (Pb)
- Benceno (C₆H₆)
- Monóxido de carbono (CO)
- Arsénico (As)
- Cadmio (Cd)
- Níquel (Ni)
- Benzo(a)pireno (B(a)P)
- Ozono (O₃)

2.9.2. Formatos habituales de publicación

Los Ayuntamientos que tienen desplegadas redes y estaciones para el control de la calidad del aire publican generalmente series históricas de datos según diferentes niveles de resolución temporal: horas, días, meses y años. Debido al volumen de esta información, especialmente en el caso más desagregado (por horas), la información calidad del aire se suele publicar en formatos planos de datos: como CSV o similares (por ejemplo, .txt).

2.9.3. Vocabularios de representación

El vocabulario que se recomienda en la norma UNE 178301:2015 es el vocabulario del W3C para la representación de datos provenientes de redes de sensores: “Semantic Sensor Ontology” (véase la Sección 3.3 de este documento para una introducción a este vocabulario).

Sin embargo, esta ontología es un marco conceptual genérico para la descripción de sistemas de captación de datos basados en sensores y otros dispositivos físicos. Es necesario extender este vocabulario para el caso de uso concreto de las redes de medición de calidad del aire:

- Tipos de dispositivo
- Descripción de las observaciones en base a los niveles de presencia de los contaminantes en el aire.

2.10. Callejero

2.10.1. Descripción del conjunto de datos

Los callejeros municipales, como su propio nombre indica, son el compendio de las calles que comprende el territorio de una administración municipal. Son instrumentos de gran importancia porque sirven el posicionamiento de todo tipo de entidades dentro del entorno de la ciudad: bienes inmuebles residenciales, fincas, negocios, edificios administrativos, etc.

2.10.2. Formatos habituales de publicación

Los callejeros son elementos auxiliares para la generación de contenido de interés (por ejemplo, directorio de empresas y negocios de la ciudad) y del desarrollo de aplicaciones para los agentes de la ciudad.

Como elemento de consulta, suelen ser publicados habitualmente como mapas, atendiendo a los siguientes formatos:

- Mapa callejero de la ciudad: fichero PDF.
- Aplicación web dentro del portal del Ayuntamiento.

2.10.3. Vocabularios de representación

El vocabulario de referencia para la publicación del callejero, según la norma UNE 178301:2015, es el “[Vocabulario para la representación de direcciones postales en España](#)”.

Este sencillo vocabulario permite describir las calles de la ciudad a partir de las siguientes entidades centrales:

- Tipo de Vía (*Vía*): que engloba todos los tipos de vías urbanas dentro del territorio municipal.

Actualmente el vocabulario no recoge, como subclases, los diferentes tipos de vía. Esta mejora está contemplada en el roadmap de desarrollo del vocabulario. Existen varias clasificaciones oficiales de las vías, como las utilizadas por el Catastro o el Instituto Nacional de Estadística (INE). Es recomendable que la extensión del vocabulario sea compatible con alguna de las clasificaciones normativas, o al menos de amplio uso en el territorio nacional, ya que esto facilitará las labores de compatibilidad y de implementación de funciones de búsqueda.

- Tramo de Vía (*TramoVía*): Tramo parcial o total de una vía
- Portal (*Portal*): Acceso exterior a una finca o construcción.

Además el vocabulario ofrece dos tipos de mecanismos de descripción de las vías urbanas:

- Propiedades composición y relación entre las vías, tramos de vía y portales. Por ejemplo, para expresar: “este es el portal XX de la calle YYYY”.
- Propiedades de nombrado (i.e., literales) y numeración de los elementos anteriores.

Existe, en paralelo a este vocabulario, la ontología: "[Vocabulario de localizaciones](#)". Este vocabulario también facilita la descripción de vías, aunque tiene un ámbito más general que el meramente municipal.

2.11. Población

2.11.1. Descripción del conjunto de datos

El estudio poblacional y demográfico de una unidad administrativa se realiza a partir de estadísticas conocidas como censos de población. En el caso de las Entidades Locales estos censos se elaboran principalmente a partir de los padrones municipales: registro administrativo donde constan los habitantes de un municipio. En España estos padrones constituyen la única fuente oficial para el cálculo de estadísticas poblacionales. La coordinación de estos censos y los procesos de cálculo estadístico agregados los realiza el Instituto Nacional de Estadística (INE). Los padrones municipales son de registro continuo y, por tanto, son una fuente de datos actualizada, siempre según la información que proporcionan los habitantes a las administraciones acerca de movilidad y empadronamiento.

La información demográfica tiene una transcendencia mucho mayor que el simple conteo de personas que habitan en una región (en este caso, un municipio) determinada. Desde el punto de vista gubernamental, el censo poblacional es necesario para el diseño de políticas de planificación económica, social, urbanística y provisión de servicios por parte del Ayuntamiento.

Estos estudios de población municipales suelen incluir más datos en relación con los puramente demográficos, como económicos y sociales, para entender correctamente la estructura de la población: natalidad, la mortalidad, la familia, la fecundidad, el sexo, la edad, la educación, el divorcio, el envejecimiento, la migración, el trabajo, la emigración y la inmigración. A continuación se indican variables estudiadas habitualmente en relación con el padrón y que son de interés para la **publicación de datos sobre la población**:

1. Personales

- a. Sexo
- b. Edad
- c. Estado Civil
- d. Minusvalías

2. Geográficas y culturales

- a. Nacionalidad
- b. Lengua
- c. Lugar residencia
- d. Migraciones
- e. Lugar estudio
- f. Lugar trabajo

3. Educativas

- a. Estudios básicos
- b. Estudios superiores
- c. Escolaridad

4. Matrimonio y Fecundidad

- a. Número hijos
- b. Número matrimonios
- c. Número de divorcios

- d. Años matrimonio
- e. Edad de los esposos

5. Socioeconómicas

- a. Actividad económica
- b. Nivel de ingresos
- c. Profesión
- d. Rama actividad
- e. Situación profesional

6. Familias

- a. Formas convivencia
- b. Composición familiar
- c. Tamaño familia
- d. Personas dependientes

2.11.2. Formatos habituales de publicación

La información de padrón se publica generalmente utilizando los siguientes formatos:

- Documentos e informes que describen la estructura demográfica del municipio. Formato: PDF y MS Word.
- Ficheros CSV y Excel.

2.11.3. Vocabularios de representación

El vocabulario que se recomienda en la norma UNE 178301:2015 es el vocabulario estándar del W3C para la representación de datos multidimensionales: [RDF Data Cube](#)

(véase la Sección 3.2 de este documento para una explicación más detallada de este vocabulario).

Sin embargo, es importante recordar que RDF Data Cube es un vocabulario genérico para la representación de información cuantitativa, pero no proporciona conceptos y propiedades para la publicación de conjuntos de datos de naturaleza específica (como es el caso de la población). Es necesario, por tanto, extender el vocabulario Data Cube en dos niveles:

- Estructuras conceptuales que describen las dimensiones de los datos, según criterios de clasificación estadística. Por ejemplo: sexo, edad, nivel de ingresos, nivel de estudios, etc. Estos conceptos y relaciones entre conceptos se expresan utilizando el vocabulario estándar del W3C para la representación de sistemas terminológicos: SKOS (véase la Sección 3.1 de este documento para una explicación más detallada de este vocabulario)
- Propiedades para cuantificar y medir la población según las dimensiones estadísticas correspondientes: *qb:MeasurementProperty*

3. VOCABULARIOS GENÉRICOS

En esta sección, se explicarán ciertos vocabularios genéricos que se mencionan en el ámbito de la Norma UNE 178301:2015 como vocabularios recomendados para la representación de ciertos conjuntos de datos. Dada además la importancia y el rol principal que desempeñan estos vocabularios en las iniciativas de datos abiertos a nivel internacional es conveniente ofrecer una breve introducción al objetivo de cada uno de ellos y a sus mecanismos básicos de descripción de datos.

Los vocabularios son los siguientes:

- Vocabulario: “SKOS”
- Vocabulario: “RDF Data Cube”
- Ontología: “Semantic Sensor Network”

3.1. Vocabulario W3C: SKOS

[SKOS](#) (Simple Knowledge Organization System) forma parte de la familia de estándares del W3C, contruidos sobre la base de RDF y OWL para la facilitar la publicación de información estructurada en la web dentro del paradigma Linked Data. SKOS permite la representación de sistemas de organización de conocimiento, como pueden ser:

- Tesoros
- Esquemas de clasificación (como los sistemas de clasificación estadística)
- Taxonomías
- Vocabularios controlados
- Glosarios
- Sistemas de tags, como el caso de las folksonomías

El vocabulario SKOS proporciona las clases y propiedades necesarias para representar las características principales que presenta un tesoro o un vocabulario controlado regular. La visión concepto céntrica suministrada por SKOS permite un modelado donde los objetos primitivos no son términos, sino conceptos, esto es, nociones abstractas representadas por términos.

A continuación se presentan los puntos importantes de esta ontología que es necesario conocer para entender su alcance en la publicación de recursos terminológicos:

- Esquemas conceptuales y conceptos:

La clase *skos:Concept* recoge el conjunto de conceptos SKOS, donde un concepto es cualquier significado que puede ser nombrado a través del lenguaje. Estos conceptos se agrupan en esquemas (*skos:ConceptScheme*), que representan cualquier sistema de organización de conocimiento, como los listados anteriormente.

- Labels y multilingüismo:

Las etiquetas léxicas (es decir, los términos, o “labels” en un lenguaje más técnico) son cadenas de caracteres en una lengua natural, como el inglés, el quechua o el ruso. SKOS facilita la distinción entre etiquetas preferentes y alternativas para cada concepto. Esto es sumamente pertinente para generar representaciones de sistemas de organización del conocimiento en los que son necesarios establecer prioridad de presentación o tratamiento del conocimiento léxico.

SKOS proporciona mecanismos adicionales para describir estas etiquetas de manera más compleja si es necesario a través de la extensión conocida como SKOS+XL.

- Relaciones entre conceptos: jerarquías y composicionalidad

Las relaciones semánticas en SKOS son enlaces entre conceptos. SKOS distingue entre dos categorías básicas de relación semántica: jerárquicas y asociativas. Por un lado, una conexión jerárquica entre dos conceptos implica que uno de ellos es más general (“broader”) que el otro (“narrower”). Por el otro, una conexión asociativa entre dos conceptos indica que los dos están relacionados inherentemente, pero no de manera jerárquica, sino de igual a igual.

Las propiedades *skos:broader* y *skos:narrower* se usan por tanto para trazar un enlace jerárquico directo entre dos conceptos. La propiedad *skos:related*, en cambio, se usa para establecer un enlace asociativo entre dos conceptos.

- Equivalencias semánticas

SKOS también permite establecer equivalencias de significado entre conceptos dentro del mismo esquema o conceptos de esquemas diferentes. Las principales son las siguientes. Por un lado, la propiedad *skos:closeMatch* se usa para vincular dos conceptos que son suficientemente similares como para se pueden intercambiar en algunas contextos, mientras que la propiedad *skos:exactMatch* se usa para vincular dos conceptos plenamente usables de manera equivalente en cualquier contexto.

El diagrama⁵ a continuación ilustra el vocabulario SKOS en acción para la representación de conceptos.

⁵ Fuente: W3C (<https://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>)

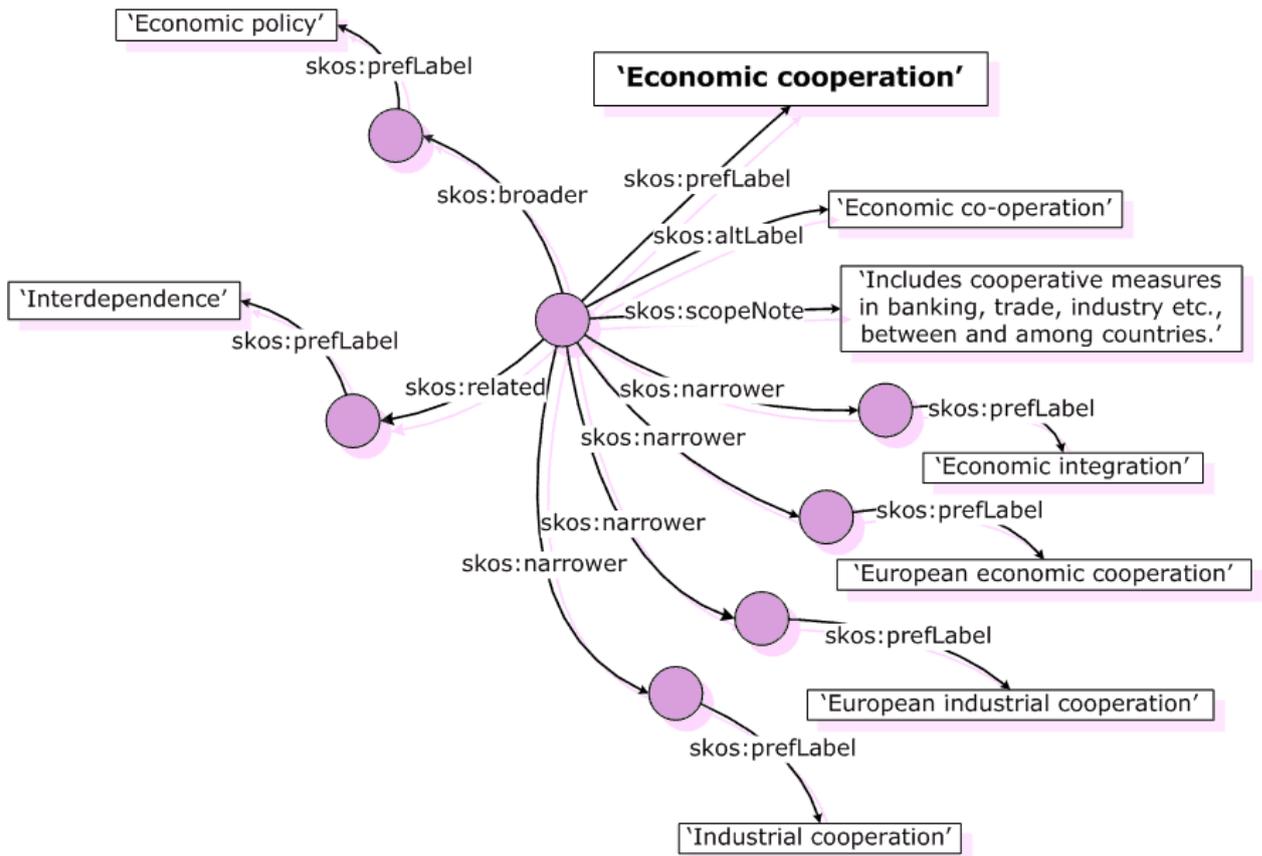


Figura 4.- Ejemplo de uso del vocabulario SKOS (fuente: W3C)

3.2. Vocabulario W3C: RDF Data Cube

El vocabulario estándar, recomendación W3C, para la representación de información cuantitativa se conoce como [RDF Data Cube](#). Este vocabulario, que nació inicialmente con el objetivo de representar información estadística y como la alternativa RDF para el intercambio de datos estadísticos a partir del estándar SDMX, está fundamentado en el modelo multidimensional de representación de datos. Esto proporciona a RDF Data Cube un alcance mucho más amplio y un potencial de aplicación en múltiples casos de uso de publicación de datos, más allá del propósito original con el que se diseñó.

3.2.1. Modelo Multidimensional

Para comprender correctamente la aplicación de RDF Data Cube es importante realizar una breve introducción explicativa a la multidimensionalidad. Este modelo se refiere a

una forma determinada de expresar los datos, ampliamente utilizada por herramientas de analítica y explotación:

- Analítica y minería de datos
- Análisis estadístico
- “Business intelligence” (o inteligencia de negocio)
- Visualización de datos

Un **conjunto de datos multidimensional**, también conocido como “cubo” se organiza de acuerdo a un conjunto de **dimensiones, medidas y atributos**.

Para explicar estos conceptos, usaremos como ayuda la siguiente tabla hipotética de datos de población:

Tabla de datos	2003		2004		2005	
	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre
Macondo	3.000	2.891	3.079	3.060	3.050	2.989
Xanadú	10.500	11.030	10.753	10.647	11.467	11.543
Barataria	301	289	352	276	333	297

Tabla 1.- Conjunto hipotético de datos de población

Dimensiones (o variables nominales):

- Son las variables que describen el conjunto de datos y que siempre toman valores discretos. Ejemplos de dimensiones son: región geográfica, sexo, edad, etc.
- Las dimensiones se pueden organizar en jerarquías, como el caso de las fechas: año > mes > semana > día. Los valores se van agregando a desagregando según el nivel de la jerarquía.
- En la Tabla 1 las **dimensiones** para representar los datos son las siguientes:

- *Sexo*: {Mujer, Hombre}
- *Año*: {2003, 2004, 2005}
- *Ciudad*: {Macondo, Xanadú, Barataria}

Medidas (o variables numéricas):

- Son las variables que miden/representan el fenómeno observado. Ejemplos de población son: temperatura, peso, densidad, ingreso/habitante, etc.
- En la Tabla 1 la **medida** que se representa es *Población*.

Atributos

- Los atributos sirven para calificar e interpretar el valor observado. Por ejemplo, para especificar las unidades de medida, o si se tratan de valores estimados.

A partir de esta estructura conceptual, los datos se organizan en lo que se conoce como observaciones o registros. Por ejemplo, podemos expresar la información de la Tabla 1: “Macondo tenía en 2003 3.079 mujeres” de la siguiente forma:

<Ciudad: Macondo, Sexo: Mujer, Año: 2004, Población: "3.079">.

Gracias a esta capacidad multidimensional, el modelo semántico de RDF Data Cube es muy general y se puede utilizar para la publicación de y datos que se encuentran actualmente en:

- Hojas de cálculo
- Ficheros de datos planos, que respeten el formato tabular. Por ejemplo: CSV
- Ficheros estadísticos
- Cubos de datos OLAP

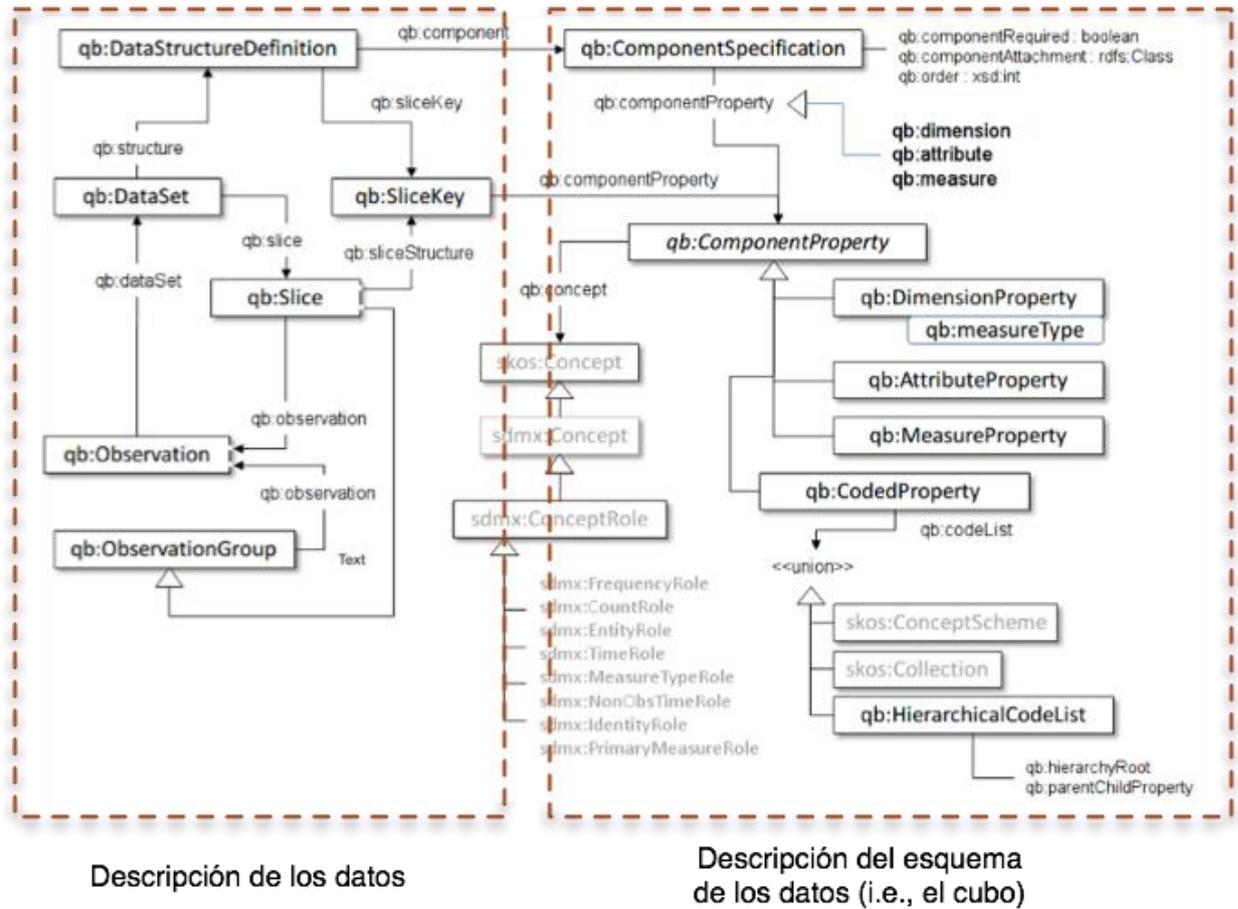
- Información cuantitativa almacenada en bases de datos relacionales

3.2.2. Descripción del vocabulario

El estándar RDF Data Cube implementa este modelo multidimensional como un vocabulario OWL. A continuación se explica el funcionamiento básico del vocabulario para la publicación de datos cuantitativos como Linked Data.

- Los datos se organizan y agrupan en datasets (*qb:DataSet*)
- Es necesario especificar el esquema multidimensional del dataset, es decir, las dimensiones, atributos y medidas, lo que en RDF Data Cube también se denominan componentes (*qb:ComponentProperty*). Estos componentes se expresan como propiedades RDF, instancias de las correspondientes clases del vocabulario: *qb:DimensionProperty*, *qb:AttributeProperty* y *qb:MeasureProperty*.
- Para especificar el significado de los componentes se usan esquemas conceptuales en el estándar SKOS (véase Sección 3.1 de este documento), lo que permite proporcionar una semántica reutilizable al esquema multidimensional de los datos (*skos:Concept*).
- El vocabulario RDF Data Cube también permite definir grupos de subconjuntos del dataset, evitando tener que trabajar siempre con todo el dataset. Cada uno de estos subconjuntos se denomina técnicamente “slice” (*qb:Slice*).

El esquema completo del vocabulario RDF Data Cube es el siguiente:



A modo exclusivamente de ilustración y sin entrar en detalles técnicos, una representación en RDF Data Cube del ejemplo anterior de la Tabla 1 sería el siguiente:

```
# Definición del dataset
eg:dataset1 a qb:DataSet ;

# Definición de los compontes
ex:ciudad a rdf:Property, qb:DimensionProperty ;
ex:año a rdf:Property, qb:DimensionProperty ;
ex:sexo a rdf:Property, qb:DimensionProperty ;
ex:poblacion a rdf:Property, qb:MeasureProperty ;
```

Observación: <Ciudad: Macondo, Sexo: Mujer, Año: 2004, Población: "3.079">

eg:obs1a a qb:Observation ;

qb:dataSet eg:dataset1 ;

ex:ciudad ex:Macondo ;

ex:sexo ex:Mujer ;

ex:año "2004" ;

ex:poblacion "3079" .

3.3. Ontología: Semantic Sensor Network

La [Semantic Sensor Network Ontology \(SSN\)](#) es una ontología para describir sensores y sus observaciones, es decir, los datos físicos que miden los dispositivos. Este vocabulario es fundamental para las ciudades inteligentes en el desarrollo de sus políticas de apertura de información pública en cuestiones que hemos visto en este informe como el control medioambiental (calidad del aire y del agua) y la gestión del tráfico.

La ontología SSN es un vocabulario OWL elaborado por el W3C, en colaboración con Open Geospatial Consortium (OGC), para definir un estándar general de descripción de redes de sensores alineado con las tecnologías de la Web Semántica y el paradigma Linked Data.

Dada la multitud de escenarios y casos de uso de esta ontología, su diseño está basado en una arquitectura modular (horizontal y vertical) que permita su extensión de forma natural a partir de un vocabulario mínimo común. Este vocabulario se ha denominado **SOSA**, de acuerdo a sus conceptos principales:

- **Sensores** (*ssn:Sensor*): dispositivos que miden una o variables magnitudes físicas: por ejemplo, acelerómetros, giroscopios, barómetros, etc.

- **Observaciones** (*ssn:Observation*): el valor real o estimado de una magnitud física.
- **Muestras** (*ssn:Sample*): ejemplos representativos de una determinada magnitud.
- **Actuadores** (*ssn:Actuator*): dispositivo que modifican magnitudes físicas a partir de instrucciones: por ejemplo, un ventilador o el automatismo de una puerta.

El siguiente diagrama ilustra este diseño modular de la ontología SSN:

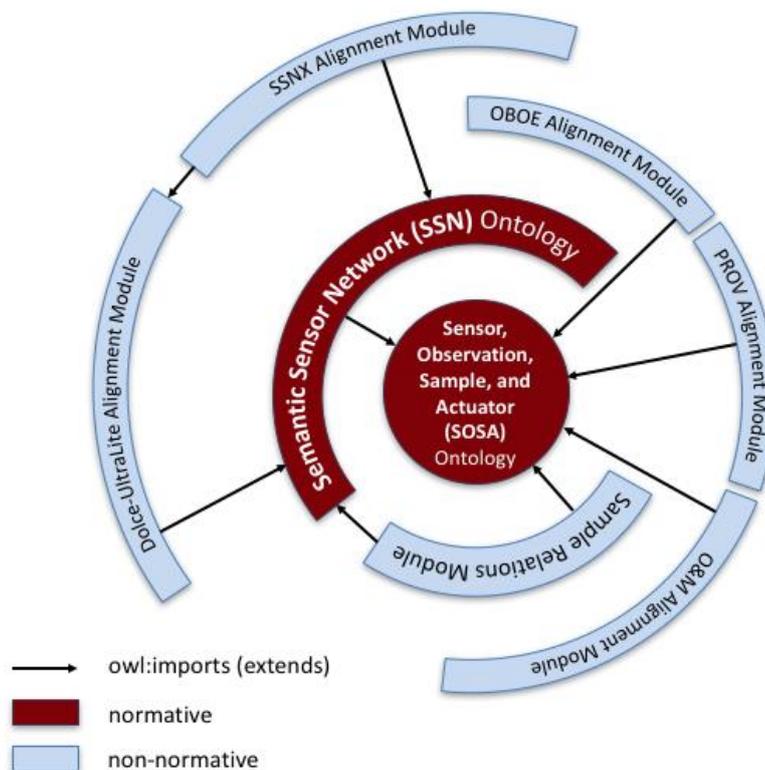


Figura 5.- Organización modular de la SSN (Fuente: W3C)

Uno de los aspectos más destacables de esta ontología SSN es, como se ha comentado, la participación del OGC como parte integrante del proceso. Esto significa que es un estándar conjunto de las dos organizaciones: W3C y OGC. Por otro lado, se ha trabajado

en la compatibilidad de SSN con los **estándares OGC existentes para la anotación de sensores y observaciones**. En concreto con:

- [“OGC Observations and Measurements standard”](#). Vocabulario XML para la descripción del modelo conceptual de “Observations and Measurements (O&M)” para la publicación de observaciones ambientales, publicado conjuntamente por el OGC y la organización ISO (en este caso como estándar ISO/DIS 19156).
- [“OGC Sensor Model Language”](#) (*SensorML*). Vocabulario XML para describir sensores, actuadores, así como los procesos de medición y transformación de observaciones.

4. DISCUSIÓN Y CONCLUSIONES

Este informe se ha centrado en el análisis del estado actual de los conjuntos de datos identificados como prioritarios y los vocabularios de representación recomendados por la Norma UNE 178301:2015 para las estrategias de publicación de datos abiertos de las Ciudades Digitales. Las conclusiones principales que se extraen sobre el estado actual del arte, que se pueden utilizar como directrices para continuar este trabajo de estandarización de las políticas municipales, son las siguientes:

Sobre los conjuntos de datos prioritarios:

1. En la Norma UNE 178301:2015, se identifican 11 conjuntos de datos estratégicos, que abarcan desde la publicación de los presupuestarios hasta el callejero municipal. Si bien se trata de una primera selección de interés, requiere de una evolución atendiendo, entre otras, a las sugerencias enumeradas a continuación:
 - a. Estos 11 conjuntos de datos son un buen punto de partida, si bien insuficiente en relación al volumen de datos que maneja un municipio y el gran potencial que tienen por su cercanía a los ciudadanos. Por ejemplo, hay conjuntos de datos municipales que no se han incluido y que, sin embargo, representan la mayor parte de las competencias y de la actividad administrativa de los Ayuntamientos en la gestión y organización de la vida en la ciudad. Es el caso de los Expedientes Administrativos, que abarcan desde licencias urbanísticas hasta las sanciones de multas de tráfico.
 - b. Sería deseable avanzar hacia una descripción de la naturaleza de cada conjunto de datos y de los requisitos mínimos de modelado que se necesitan. En este sentido, este informe es un primer paso en esta dirección que sin duda puede ser continuado por un grupo de trabajo coordinado de expertos.

- c. Sería de gran valor el que se incluyeran casos de uso. Esto es importante para comprender el potencial de aplicación de estos conjuntos de datos y el grado de interés estratégico para priorizarlos en las estrategias municipales de publicación de datos abiertos.

Sobre los vocabularios recomendados:

1. En la Norma UNE 178301:2015, se proponen ciertos vocabularios para la representación de los conjuntos de datos identificados como prioritarios. La mayor parte de estos vocabularios han sido creados *ex profeso* para representar estos conjuntos de datos. En este sentido, todos los vocabularios están alineados con los objetivos de la Norma y el nivel de cobertura semántica es adecuado de manera general.

Sin embargo, se han identificado varios aspectos que es importante tener en cuenta en el desarrollo de estrategias de estandarización para las Ciudades Digitales:

- a. La mayoría de los vocabularios todavía están en fase inicial (versión 0.1) y sus últimas fechas de actualización datan de 2014 y 2015. Es importante retomar el trabajo desde esta fase inicial para proporcionar especificaciones de calidad y robustas que puedan ser utilizadas en proyectos de producción.
- b. La mayor parte de los vocabularios deberían ser enriquecidos con información adicional y con documentación explicativa. Esto facilitar el a conocer y entender correctamente el vocabulario correspondiente.
- c. Tal y como indicábamos anteriormente, sería adecuado proporcionar ejemplos de uso y buenas prácticas en la aplicación de los vocabularios, además de reglas de validación para la construcción de datasets que asegurarían que el mismo vocabulario es interpretado de igual forma

tanto por los agentes de publicación, como por los agentes consumidores.

2. Debido al momento en que han sido elaborados y a la existencia de diferentes líneas de trabajo en este ámbito, se han identificado ciertos conflictos entre los vocabularios recomendados por la Norma, de ámbito nacional en su mayoría, y vocabularios recientes promovidos por la Unión Europea. El caso más significativo es la ontología PPROC y la eProcurement Ontology. Es necesario, dentro del trabajo futuro, tomar una decisión sobre la política a seguir con los vocabularios en estas situaciones de conflicto.
3. En el caso de propuesta de vocabularios genéricos para ciertos conjuntos de datos (concretamente, como RDF Data Cube y la ontología SSN para la información de Población Calidad del Aire respectivamente), sería conveniente un trabajo y un desarrollo adicional para proporcionar los mecanismos de descripción específicos. De esta forma, estos vocabularios que son de propósito genérico, podrían ser extendidos para cubrir las necesidades de estos casos de uso.

Más allá del marco de trabajo de la Norma UNE 178301:2015, los conjuntos de datos identificados, y los vocabularios propuestos por esta, han servido como punto de partida para otros grupos de trabajo en pro de la publicación de datos abiertos de municipios y entidades locales. Así, en Julio de 2017, la Federación Española de Municipios y Provincias (FEMP) presentó la [“Datos Abiertos – Guía estratégica para su puesta en marca - Conjuntos de datos mínimos a publicar”](#) que condensa acciones, pautas y recomendaciones para que ayuntamientos y entidades locales publiquen sus datos de forma útil y eficaz para su uso y reutilización por parte de ciudadanos, organizaciones, empresas u otras administraciones. [Esta guía](#) contiene una selección de 20 conjuntos de datos que los autores consideran que son los más interesantes, posibles y fácil de publicar por las Entidades Locales. Estos 20 conjuntos de datos incluyen los 11 propuestos por la Norma UNE 178301:2015 y que se han analizado en este informe.

Siguiendo esta línea de trabajo, el proyecto “[Plataforma de Gobierno Abierto, Colaborativa e Interoperable](#)” que están desarrollando los Ayuntamientos de Madrid, Zaragoza, A Coruña y Santiago de Compostela junto con Red.es, tiene previsto la publicación consensuada de 27 conjuntos de datos, sumando 7 más a los seleccionados por el grupo de trabajo de la FEMP.

Este proyecto tiene previsto el desarrollo de los vocabularios y las estructuras de datos necesarios para que todos los conjuntos de datos seleccionados tengan un vocabulario de referencia para su publicación. En las [actuaciones previstas para este proyecto](#), se incluye la creación de aquellos vocabularios que no existan en la actualidad. Tanto los vocabularios ya existentes, como los nuevos vocabularios que se creen en la marco de este proyecto, se publicarán y se podrán a disposición del resto de municipios para su uso generalizado y, de esta forma, contribuir a la interoperabilidad semántica que tan relevante resulta para asegurar el éxito en la reutilización de datos generados por el sector público.