



TRABAJOS FIN DE ESTUDIOS EN EL ÁMBITO DE LA ACTIVIDAD DE GENERACIÓN DE TALENTO DE LA CÁTEDRA LODCOREMADRID

1 ANTECEDENTES Y JUSTIFICACIÓN

La cátedra para I+D+i sobre dinamismo territorial en la Comunidad de Madrid con el uso de datos abiertos y enlazados (LoDCOREMadrid) tiene como objetivo constituirse como plataforma estable de I+D+i y generación de talento para el soporte en la evaluación de políticas favorecedoras de la competitividad territorial mediante la investigación, creación, agregación y desarrollo de servicios basados en la tecnología de los datos abiertos y enlazados.

Uno de los cometidos de la cátedra consiste en la generación de talento y conocimiento en el área de investigación considerado. Para ello, se han establecido los siguientes objetivos específicos:

- 1) Cooperación en programas de formación/experiencias de cooperación educativa relacionadas con el uso de datos semánticos, abiertos y enlazados para el dinamismo económico y la evaluación de políticas de desarrollo territorial en la Comunidad de Madrid.
- 2) Patrocinio de la realización de estudios (concesión de becas, apoyo a la realización de tesis doctorales y proyectos de fin de carrera, premios a tesis y proyectos de fin de grado, de máster, etc.).

El propósito de este documento consiste en definir la temática, objetivos y tipología de los trabajos fin de estudios que se podrán evaluar y premiar en el ámbito de la actividad de generación de talento de la cátedra LoDCOREMadrid.

1.1 Calidad de vida y cohesión territorial

La evaluación de la calidad de vida se basó en la definición de la ONU de 1954. Dice así: "El concepto de nivel de vida abarca el conjunto de las condiciones reales de vida de las personas y el grado de satisfacción material y cultural de sus necesidades mediante una corriente de bienes y servicios remunerados, así como con cargo a los fondos sociales". Asimismo, el nivel de vida se define como el grado de satisfacción de las necesidades materiales y culturales de los hogares y distingue siete tipos básicos de necesidades: alimentación, seguridad, protección de la salud, vivienda, comunicación y transporte, educación y cultura, y medio ambiente. También se indica que la "calidad de vida" es un concepto complejo condicionado por varias características sociales y económicas.



A través de la Estrategia Europa 2020, la UE formula tres objetivos adicionales con respecto al concepto de calidad de vida: "crecimiento inteligente, crecimiento sostenible y crecimiento integrador". La estrategia hace hincapié en la dimensión social de los territorios y se esfuerza por lograr un equilibrio entre los aspectos económicos, la política del mercado laboral y los aspectos sociales, y hace hincapié en los ámbitos de la "pobreza financiera" y las "condiciones de vida", el acceso al mercado laboral y la "educación". Aceptar el nivel de las condiciones de vida de la población como la función social como objetivo del desarrollo socioeconómico es, de hecho, sinónimo de tratar el desarrollo como el principal determinante de la calidad de vida.

El desarrollo socioeconómico afecta directamente al cambio de las condiciones de vida al proporcionar ingresos a la población. También lo hace indirectamente al estimular y activar la actividad económica, las infraestructuras sociales y técnicas y mejorar o deteriorar el estado del entorno natural.

En este marco, el gobierno de la Comunidad de Madrid tiene por objeto reforzar la cohesión económica, social y territorial, así como la igualdad de oportunidades de entre todos los municipios y zonas de la Comunidad de Madrid.

Esta línea de actuación hace hincapié en el problema de la cohesión en el contexto territorial. La esencia de la cohesión territorial es la necesidad de eliminar las desigualdades entre las condiciones de vida de la población. El concepto de calidad de vida es ambiguo, multidimensional e interdisciplinar.

La cohesión territorial consiste en igualar el nivel de vida socioeconómico de la población, integrando y configurando la justicia en el sistema territorial. Por una parte, está relacionada con la dimensión económica de la planificación y la gestión del desarrollo y, por otra, con el concepto de enfoque integrado del desarrollo.

En el contexto madrileño, se indica que la categoría de cohesión territorial no significa una simple igualación de las diferencias sociales y económicas en el espacio, sino el desarrollo coherente de la región de Madrid como un solo organismo.

El equilibrio y la cohesión territorial se sustenta en 3 pilares diferenciadores: la gestión pública, los servicios y una visión territorial global. Partimos de un elemento común: el territorio y población, a partir de aquí se construyen las dimensiones que dan respuesta al equilibrio, son: la economía, el medio ambiente, la persona y su calidad de vida, los servicios y la gestión de gobierno o administrativa.

La esencia de la cohesión territorial es la necesidad de eliminar las desigualdades entre las condiciones de vida de la población. En la Comunidad de Madrid, este proceso se ve estimulado por diversos instrumentos financieros, organizativos y jurídicos.

Las condiciones endógenas (geográficas y naturales) y exógenas (socioeconómicas y financieras) hacen que las distintas zonas presenten un desarrollo socioeconómico y



unas condiciones de vida y de calidad diferentes. El efecto del desarrollo desigual de las condiciones de vida en los sistemas territoriales es la polarización y la desproporción, que conducen a la despoblación y a la degradación social, económica y natural.

La competitividad territorial es un proceso que tiene lugar en el ámbito geográfico. Además, denota la capacidad de la economía y la sociedad para aumentar el nivel de vida de la población. En los sistemas internacionales, la competitividad significa la capacidad de los Estados para producir y distribuir bienes y prestar servicios en la economía internacional en competencia con los bienes producidos en otros países, de forma que se garantice un nivel de vida creciente. La localización es fuente de numerosas limitaciones y oportunidades que modifican las condiciones socioeconómicas del desarrollo en lo que respecta a las regiones fronterizas.

A pesar de la importancia que el concepto de cohesión territorial ha ido adquiriendo en el seno de la UE y de los avances realizados en el ámbito de diferentes disciplinas científicas para clarificar su significado, varios autores advierten de la imprecisión que aún existe en la definición de este concepto. Esta indefinición hace aún más difícil la medición de la cohesión territorial de un determinado territorio, pero despierta el interés científico por avanzar en el diseño de metodologías que permitan analizarla y aportar así nuevos conocimientos.

1.2 Evaluación y análisis de la cohesión territorial

En estas circunstancias -el contexto de desequilibrios territoriales, la necesidad de avanzar hacia la convergencia de los territorios y la dificultad de medir un concepto con un marcado componente abstracto- surge la necesidad de emprender investigaciones capaces de abordar el análisis y la medición de la cohesión territorial. Entre los pocos estudios que han abordado empíricamente la medición de la cohesión territorial, aparte de los realizados por la Red Europea de Observación de la Ordenación del Territorio (programa ESPON), destacan los realizados a nivel europeo y en América Latina.

Se trata de una serie de estudios en los que la cohesión territorial se mide en diferentes contextos y a distintas escalas territoriales. Asimismo, aunque todos los trabajos contemplan la diversidad de contextos en los que se desarrollan las unidades territoriales analizadas. Esto ha llevado a que, en muchas ocasiones, regiones y territorios muy diferentes hayan sido contemplados como un "todo" homogéneo. Por ello, a pesar de estos avances, que sin duda suponen una importante referencia metodológica, resulta de interés el desarrollo de estudios que contemplen el análisis de la cohesión territorial y la búsqueda de los factores que inciden en ella a través de territorios con características de homogeneidad entre sí, tanto en sus condiciones de partida como en su nivel de recursos territoriales.

A pesar de todo, la literatura científica muestra una falta de consenso sobre la existencia de una definición precisa y ampliamente aceptada del término "cohesión territorial" y, en consecuencia, desde un punto de vista operativo y práctico, reclama que se avance en la identificación de las principales dimensiones y elementos que la componen. Para



identificar estas dimensiones, algunos autores subrayan la necesidad de considerar en primer lugar las implicaciones del estudio de las partes analíticas del término "territorio".

Así, en el ámbito de disciplinas como la geografía humana o la economía del desarrollo, el territorio puede entenderse como un espacio topográfico métrico caracterizado por tres elementos o subsistemas: (i) los recursos territoriales específicos de los que dispone el territorio; (ii) los actores o agentes territoriales (el Estado, la sociedad civil y las asociaciones, y los actores privados); y (iii) los acuerdos institucionales que articulan los procesos de transformación y desarrollo orientados a resolver los problemas comunes o específicos a los que se enfrentan los territorios. La relación que se establece entre estos tres elementos es decisiva en los procesos de convergencia (o divergencia) territorial y, por tanto, debe tenerse en cuenta en los análisis de cohesión.

A partir de estos elementos y de los aportados por otros estudios, principalmente desarrollados por el programa europeo ESPON, pero también por la Comisión Europea y por investigadores del ámbito de la geografía y la economía, se pueden establecer cinco dimensiones básicas que proporcionan un soporte teórico para el diseño de una metodología de medición de la cohesión territorial:

1. Dimensión económica: Como base fundamental para abordar los principales retos a los que se enfrentan los territorios (es decir, el aumento constante de la competencia global) y contribuir así a los objetivos últimos de crecimiento y desarrollo económico. Algunos aspectos a tener en cuenta en esta dimensión incluyen: Agricultura, pesca y silvicultura; Agrupación y cooperación económica; Construcción y renovación; Mejora de las conexiones de transporte; Infraestructura Logística y transporte de mercancías; Energía tradicional; y Transporte y movilidad
2. Dimensión social: Fundamentada en las competencias y habilidades de los individuos y en su capacidad para establecer relaciones con otros actores territoriales a través de la formación de redes sociales. Estas redes actúan como nexo de unión para el despliegue de iniciativas económicas encaminadas a garantizar el bienestar de la población. Algunos aspectos a tener en cuenta en esta dimensión incluyen: Inclusión social e igualdad de oportunidades; Cooperación entre servicios de emergencia; Patrimonio cultural y artes; Cambio demográfico e inmigración; Educación y formación; Salud y servicios sociales; TIC y sociedad digital; Transferencia de conocimientos y tecnología; Mercado laboral y empleo; Seguridad; y Cooperación científica.
3. Dimensión medioambiental: Basada en el reconocimiento de que el medio ambiente es el soporte de la vida. Por ello, la integración medioambiental debe hacerse efectiva en los procesos de desarrollo, situando esta dimensión al mismo nivel de valor que las cuestiones económicas y sociales. Algunos aspectos a tener en cuenta en esta dimensión incluyen: Vías navegables, lagos y ríos; Cambio climático y biodiversidad; Eficiencia energética; Energías renovables; Calidad del suelo y del aire; Gestión sostenible de los recursos naturales; Residuos y contaminación; y Gestión del agua.



4. Dimensión institucional: Entendida como el marco normativo bajo el cual se establecen acuerdos formales e informales para garantizar el buen funcionamiento del sistema y el desarrollo de las otras dimensiones. El capital relacional y la gobernanza son elementos clave en esta esfera. Algunos aspectos a tener en cuenta en esta dimensión incluyen: Desarrollo urbano; Integración comunitaria e identidad común; Sistemas de evaluación y resultados; Gobernanza; Cooperación institucional; y Planificación y desarrollo regionales.

5. Dimensión de desarrollo espacial integrado: Basada en los principios del desarrollo territorial equilibrado orientado a promover una estructura territorial policéntrica, tanto en lo que se refiere a la morfología de los territorios como a las relaciones que se establecen entre los asentamientos o municipios que los componen.

6. Dimensión servicios, infraestructuras y equipamientos a disposición del ciudadano. En el artículo 3 de la Ley 1/1983, de 13 de diciembre, de Gobierno y Administración de la Comunidad de Madrid, se establece que "los servicios públicos de la Comunidad de Madrid son los medios técnicos, personales y materiales necesarios para la realización de las funciones y competencias atribuidas a la misma, y que se destinan a la satisfacción de las necesidades e intereses generales de la sociedad". Se vinculan con áreas tan diversas como la gestión de residuos, el abastecimiento de agua y saneamiento, el alumbrado, la educación, la sanidad, la justicia, la seguridad, el transporte, la vivienda, las tecnologías de la información y la comunicación o los servicios sociales, entre otras. En la Comunidad de Madrid, los servicios públicos son considerados como una herramienta fundamental para garantizar el bienestar y el progreso social, económico y cultural de la sociedad, y son responsabilidad de la Administración Pública de la Comunidad de Madrid.

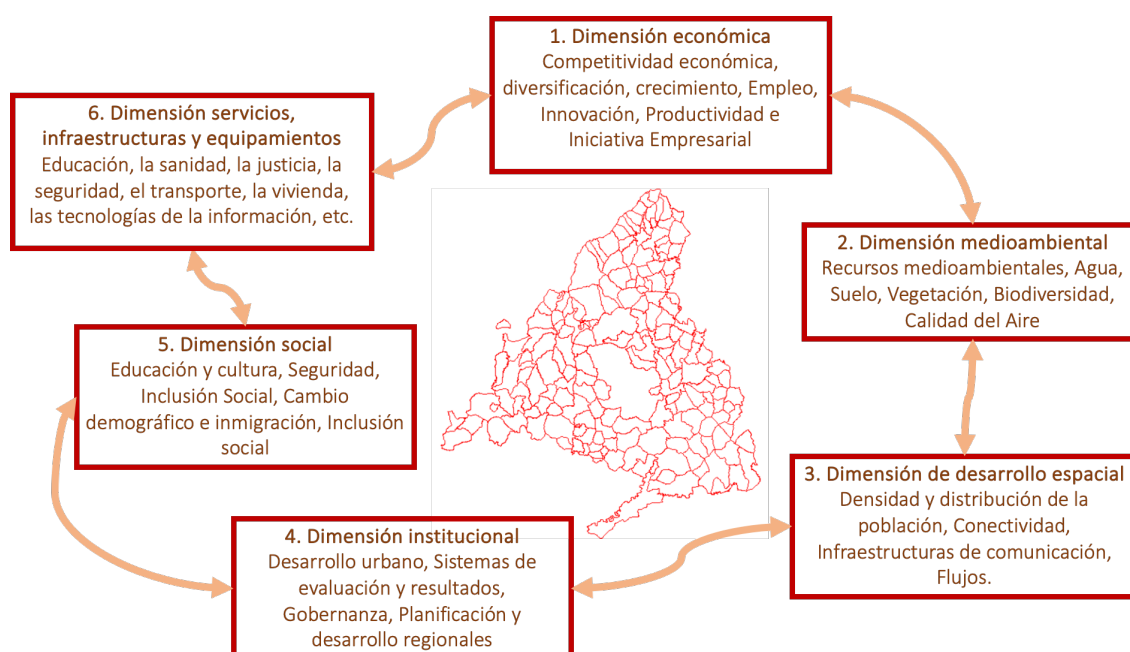


Figura 1: Dimensiones de la medición y análisis de la cohesión territorial



La identificación de estas dimensiones aclara el significado del concepto, pero, sobre todo, permite asociar a cada dimensión los componentes más adecuados para medir en consecuencia la cohesión territorial.

1.3 Análisis del dinamismo territorial a nivel económico.

En esta anualidad, LoDCOREMadrid ha decidido centrar el reto propuesto a los estudiantes la dimensión de dinamismo territorial a nivel económico.

Para la medición, evaluación y análisis del dinamismo territorial a nivel económico, la literatura propone diversos indicadores de interés (algunos ejemplos incluyen):

1. **Tasa de empleo:** Ratio de población activa/ocupada (100-tasa de paro) (%)
2. **Crecimiento:** Porcentaje de aumento de la renta del condado (%)
3. **Innovación e inversión:** Inversión en la creación de nuevas empresas (€/pers); centros de investigación, parques científicos y tecnológicos y su área de influencia, polígonos industriales, su tamaño y sectores.
4. **Estructura económica:** Se pueden tener en cuenta variaciones del Índice Herfindahl-Hirschman (IHH) que es una medida, empleada en economía, que informa sobre la concentración económica de un mercado. O, inversamente, la medida de falta de competencia en un sistema económico. Un índice alto expresa un mercado muy concentrado y poco competitivo.
Los datos corresponden al número de nuevas actividades empresariales y profesionales ponderadas por la población correspondiente a los sectores primarios, construcción, industria y servicios (adimensional).
5. **Grado de actividad económica:** Número total de nuevas actividades empresariales y profesionales ponderado por la población (adimensional)
6. **Variación de la población:** Tasa de variación de la población, expresada en porcentaje
7. **Conectividad digital:** Número de conexiones a Internet por cada mil habitantes (adimensional)
8. **Conectividad del transporte:** Proporción de superficie de autopistas, autovías y enlaces por carretera en relación con la superficie regional total (%)
9. **Grado de inversión:** Porcentaje de inversión empresarial per cápita frente a la renta per cápita (%)
10. **Grado de asociacionismo:** Porcentaje de miembros privados y empresas en las distintas asociaciones empresariales o sectoriales en el territorio (Grupos de Acción Local, Cámara de Comercio, etc.)

Para conocer el grado de dinamismo de un territorio es necesario generar datasets integrados que permitan generar estos (y otros) indicadores que permiten caracterizar el dinamismo económico de los territorios de la Comunidad de Madrid, poder establecer relaciones entre los mismos, poder determinar la cohesión del territorio en esta dimensión y poder relacionar estos con indicadores asociados a las otras dimensiones de la cohesión y competitividad territorial.



Es necesario considerar que los indicadores considerados deben contemplar las diferencias de población en los 179 municipios de la Comunidad de Madrid, que van desde municipios menores de 100 habitantes (con 144 municipios menores de 20.000 habitantes), hasta la ciudad de Madrid, lo que genera diferencias en la forma de acceder a ciertos servicios.

2 DATOS ABIERTOS Y ENLAZADOS

Linked Open Data es un conjunto de prácticas que se siguen para publicar datos estructurados y enlazados en la web, de forma que dichos datos se encuentren interconectados entre sí, y puedan ser utilizados y accedidos por millones de usuarios.

El objetivo que busca esta serie de principios es construir una web, la web semántica, que sea global y abierta, accesible por cualquier persona, donde los datos sean la unidad de acceso (en lugar de los documentos como hasta ahora) y que puedan ser enlazados desde otras datos para facilitar su reutilización y uso estandarizado.

En la web semántica, la publicación de datos no se centra en las páginas html sino en el uso de formatos distintos, tales como RDF u OWL, para la creación de aplicaciones y servicios web donde se apostaba fuertemente por la accesibilidad de los datos.

2.1 Open Data

Para que unos datos sean abiertos, se deben cumplir los siguientes requisitos:

- La información debe estar disponible y accesible en todo momento, a un costo que tiene que ser razonable, y, si fuera posible, debe ser descargable a través de internet.
- Todo usuario que quiera acceder a la información debe tener la posibilidad de utilizar o redistribuir dicha información. No puede haber restricciones ni ningún tipo de discriminación a personas o grupos que la utilicen.
- Los datos deben disponer de licencias o permisos que posibiliten y permitan tanto su redistribución como su reutilización, y, además, deben poder ser utilizados con otros conjuntos de datos.

2.2 Linked Data

Los datos abiertos no tienen por qué estar enlazados o interconectados a otros datos, y es ahí donde surge el término de Linked Data. Como bien se ha definido, es el conjunto de principios y prácticas para publicar datos enlazados en la web.

Cabe destacar que, de igual manera que en los datos abiertos, es posible que los datos no se encuentren enlazados o estructurados, lo mismo ocurre con los datos enlazados, pues no tienen por qué ser de libre acceso. La combinación de ambas vertientes da lugar a los datos abiertamente enlazados o Linked Open Data.



Los conjuntos de datos, que reciben el nombre de datasets, pueden enlazarse entre sí para formar conjuntos de conjuntos de datos, de manera que se forma una gran red global de datos. Esta red global de datos o red de datasets es lo que se conoce hoy en día como Web Semántica, y es un concepto muy importante en el ámbito de los datos abiertos enlazados. Estos datasets se caracterizan fundamentalmente porque pueden operar entre ellos de una manera muy efectiva y, además, presentan la capacidad de enlazarse a otros conjuntos de datos sencillamente.

Los datasets pueden provenir de distintas fuentes, pues no existe ningún tipo de restricción. Puede ocurrir que un conjunto de datos provenga de una biblioteca electrónica, mientras que otra provenga de una base de datos del ministerio de un país, y, a su vez, una tercera provenga de la base de datos de una empresa, lo que hace que la web sea muy robusta, pues “cualquiera” puede publicar su conjunto de datos en la web, enlazarlo con otros y crear una web más fuerte aún.

En los datos abiertos enlazados, existen conjuntos de datos muy famosos y grandes. Los más conocidos son: DBpedia, LinkedGeoData, MusicBrainz y LinkedMDB, entre muchos datasets enlazados que existen.

Los cuatro principios básicos sobre los que se centra el concepto de datos abiertos y enlazados establecidos por Tim Berners-Lee son:

1. Uso de URIs para referirse a cualquier cosa. Una URI, o Uniform Resource Identifier, es un sistema de identificación único que se utiliza para identificar cualquier cosa de manera única, desde conceptos abstractos o virtuales, hasta objetos físicos de la vida. Son muy importantes, ya que, gracias a ellos, datasets distintos pueden referirse a al mismo objeto de manera unívoca, sin ningún tipo de ambigüedad.
2. Uso del protocolo HTTP, puesto que proporciona un mecanismo muy simple para devolver URIs, y, por lo tanto, los recursos son mucho más fáciles de encontrar.
3. Uso del estándar RDF. Este estándar RDF o Resource Description Framework es un formato de datos utilizado para la publicación e interconexión de datos basándose en grafos.
4. Incluir enlaces a otras URIs, relacionadas con el contenido del dato del recurso, de manera que se mantiene enlazada la web y se facilita la navegación.

2.3 Linked Open Data y Web Semántica

La web tradicional está diseñada para que el usuario que accede a la información que quiere consultar pueda utilizarla, pero el acceso a esa información se realiza a través de un motor de búsqueda. La información en esta web está representada en documentos HTML que describen la semántica de los conceptos que ofrecen. Por otro lado, la web semántica es una extensión de la web tradicional, pues mantiene la estructura de



documentos HTML, pero, además, aporta un significado a los datos que presentan, facilitando la cooperación de distintas personas con la misma información.

El objetivo de la Web Semántica es proporcionar a los ordenadores la capacidad de manejar y estructurar la información en base al contenido semántico de sus recursos, de forma que no solo humanos, sino los propios ordenadores, procesen automáticamente los contenidos que alojan.

Ese objetivo se puede conseguir a través de la cooperación de la web actual y la web semántica. Esa cooperación se consigue dotando a la web de una nueva configuración de la información. Para modificar la forma en la que trabaja la web, se han investigado nuevos lenguajes y métodos que permiten ordenar la información que se encuentra por internet, de manera parecida a como se realizaría el análisis semántico en humanos. Esta semantización se puede lograr a través de XML (eXtensible Markup Language) y RDF.

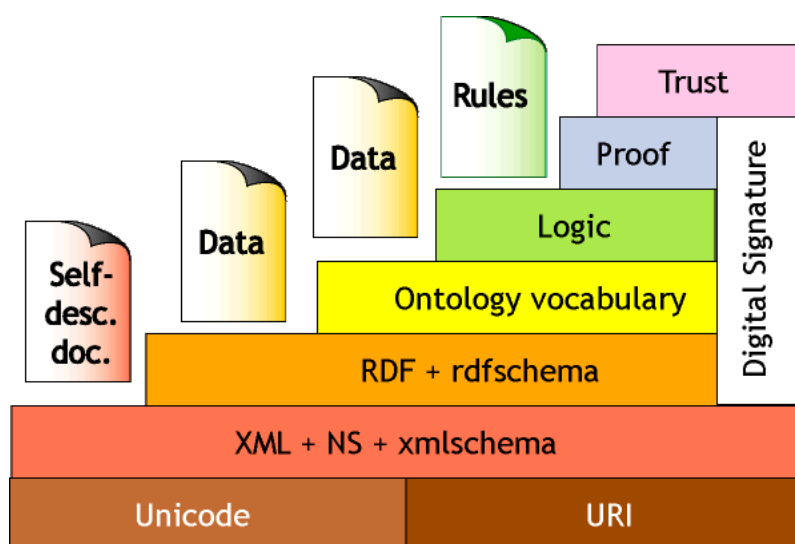


Figura 2: Lenguajes para la implementación de la web semántica

Las tecnologías y lenguajes que son necesarios para la implementación de la web semántica son:

a) RDF

El modelo de datos RDF es similar a los enfoques de modelado conceptual clásicos como entidad-relación o diagramas de clases, ya que se basa en la idea de hacer declaraciones sobre los recursos (en particular, recursos web) en forma de expresiones sujeto-predicado-objeto. Estas expresiones son conocidos como triples en terminología RDF.

El sujeto indica el recurso y el predicado denota rasgos o aspectos del recurso y expresa una relación entre el sujeto y el objeto. Por ejemplo, una forma de representar la idea de "El cielo tiene el color azul" en RDF es como el triplete de un sujeto que denota "el cielo", un predicado que denota "tiene el color" y un objeto que denota "azul". Por lo tanto RDF cambia objeto por sujeto que se utilizaría en la notación clásica de un modelo



entidad-atributo-valor2 en diseño orientado a objetos, objeto (el cielo), atributo (color) y el valor (azul). RDF es un modelo abstracto con varios formatos de serialización (es decir, los formatos de archivo), por lo que la forma particular en que se codifica un recurso o triple varía de un formato a otro.

El Marco de Descripción de Recursos (del inglés Resource Description Framework, RDF) es una familia de especificaciones de la World Wide Web Consortium (W3C) originalmente diseñado como un modelo de datos para metadatos. Ha llegado a ser usado como un método general para la descripción conceptual o modelado de la información que se implementa en los recursos web utilizando una variedad de notaciones de sintaxis y formatos de serialización de datos.

b) Ontologías

Una ontología es una descripción explícita, o un sistema de expresiones de conceptos o términos dentro de un dominio de conocimiento, que define tanto estos propios conceptos, como las relaciones entre ellos.

Las ontologías son imprescindibles, ya que las utilizan los programas informáticos para intercambiar información. Esto es muy importante, puesto que se sabe que la web semántica utiliza información única, donde no existen ambigüedades.

El objetivo de las ontologías dentro de la web semántica consiste en conseguir eliminar la ambigüedad del significado de los conceptos o términos, identificando el correcto significado de cada uno de ellos, y estableciendo una descripción formal de los mismos. De esta manera, las máquinas pueden entender el significado del concepto aun habiéndose utilizado palabras polisémicas, es decir, palabras distintas que se refieren al mismo concepto. Por tanto, se puede decir que las ontologías generan un vocabulario de conceptos, así como las relaciones entre dichos conceptos, estableciendo relaciones jerárquicas semánticas.

Si se quiere representar el dominio de conocimiento de árboles, por ejemplo, para el desarrollo de un proyecto de un parque, y se quiere modelar dicho conocimiento, es muy probable que se deba desarrollar una ontología para esquematizar los conceptos. Por ejemplo, se puede obtener un concepto árbol, el cual dispone de atributos o facetas. Es probable que se quieran modelar árboles más específicos, por lo que se deben crear nuevos conceptos, que serían clases que heredan los atributos del árbol principal. De esta forma, se pueden crear instancias más específicas del árbol, y se pueden establecer relaciones entre dichas instancias.

Además, la gran ventaja de todo esto, es que, en el caso de quererse realizar otro proyecto de un parque, y se quiera modelar el dominio de conocimiento de un árbol, el desarrollador podrá utilizar esta ontología para modelar sus datos, de manera que no utilice distintos términos para referirse al mismo concepto y evitando así posibles ambigüedades entre datos similares procedentes de distintos datasets.



RDFS o RDF Schema o Esquema RDF es una extensión semántica de RDF. Un lenguaje primitivo de ontologías que proporciona los elementos básicos para la descripción de vocabularios. La primera versión fue publicada en abril de 1998 por la W3C, la versión actual de la recomendación fue publicada en febrero de 2004 también por la W3C. Existen actualmente otros lenguajes de ontologías más potentes, como puede ser OWL.

c) Catálogo de datos (DCAT)

El Data Catalogue Application Profile (DCAT-AP), que traducido del inglés significa 'Perfil de Aplicación del Catálogo de Datos', hace referencia al vocabulario que describe los conjuntos de datos del sector público en Europa y que se integran en un catálogo (DCAT).

El objetivo es permitir la búsqueda de metadatos en portales y plataformas de datos abiertos y hacer que los datos del sector público se puedan buscar mejor a través de diferentes fronteras y sectores y sean interoperables. Esto se logra mediante el intercambio de descripciones de conjuntos de datos entre portales de datos y ya existen varias versiones, la última la DCAT-AP v2.1.0.

3 FUENTES DE DATOS INICIALES A SER CONSIDERADAS Y UTILIZADAS

En el ámbito del reto, algunas fuentes de utilidad que se pueden considerar para generar los datasets requeridos o enlazar algunos ya existentes en formato abierto y enlazado son:

A. DATOS ABIERTOS DE LA COMUNIDAD DE MADRID (<https://www.comunidad.madrid/gobierno/datos-abiertos>)

La Comunidad de Madrid trabaja para que los datos liberados en el contexto de su iniciativa de datos abiertos lo sean de una forma accesible, sencilla y libre. Lo que se persigue es que cualquier interesado pueda trabajar con ellos fácilmente y con las debidas garantías normativas y de calidad de la información.

El Catálogo de Datos Abiertos de la Comunidad de Madrid da acceso centralizado a los conjuntos de datos, o datasets, ofrecidos en formatos abiertos, para su reutilización por parte de la ciudadanía. Cada dataset dispone de una serie de metadatos conforme a la NTI y uno o más recursos. Los recursos son cada uno de los formatos en que se ofrece la información del dataset. En algunos casos un dataset puede ofrecer varios recursos en un mismo formato, con propósito histórico u otros criterios.

El catálogo dispone de un buscador libre, y de un buscador facetado que permite la navegación a través de diferentes criterios de clasificación:



* Grupo: clasificación temática establecida por la Norma técnica de interoperabilidad de reutilización de recursos de información.

* Formato de sus recursos: json, csv, wms, url, zip, etc.

* Etiquetas: transporte público, calidad del aire, género, Madrid, etc.

En función del formato del recurso se ofrecerán diferentes tipos de visualización. Para los datasets con componente geográfica se ofrece un mapa interactivo, que puede disponer de evolución histórica como el Registro de Uniones de Hecho.

Un ejemplo de reutilización de los datasets con componente geográfica es el Visor de Datasets interactivo.

La plataforma utilizada para el Catálogo es CKAN, software libre de uso extendido en el sector público, que ha sido diseñado por la Open Knowledge Foundation (OKF). Puedes encontrar más información sobre las funcionalidades de CKAN aquí.

B. NOME CALLES

NOME CALLES pretende ser un sistema de información gráfica tanto para usuarios expertos como para el ciudadano medio, donde se reúna por un lado la infraestructura cartográfica de delimitaciones geográficas, callejeros, fondos catastrales y fotografías aéreas y, por otro, la información puntual de interés georreferenciada y representada territorialmente.

Desde el comienzo, se diseñó NOME CALLES como una herramienta viva, con continuas actualizaciones y ampliaciones, donde se consideró esencial tener representados a todos los municipios de la Comunidad de Madrid con similar cobertura de capas y fondos.

Como fruto de estos procesos de mejora y de búsqueda de utilidad, hoy es posible poner a disposición de los usuarios el mismo Callejero y Nomenclátor Oficial de la Comunidad de Madrid, pero en versión en inglés, accesible desde la propia aplicación (pinchando sobre la bandera inglesa) o tecleando http://gestiona.madrid.orghttps://gestiona.comunidad.madrid/nomecalles_web/index.jsp?idioma=en desde cualquier aplicación. Esto permite ampliar enormemente el número de usuarios potenciales de esta poderosa herramienta y dar entrada a amplios colectivos de especial interés como el de las empresas extranjeras interesadas en conocer o invertir en la región o el universo del turismo.

C. ALMUDENA

Es el Banco de Datos Municipal del Instituto de Estadística, cuyo objetivo fundamental es el de recoger y facilitar agrupada la información estadística de carácter municipal de la Comunidad de Madrid, con un amplio número de series



organizadas por temas. Está integrado dentro del Banco de Datos Estructurales. Permite al usuario la realización de consultas dinámicas, así como la descarga de ficheros y mapas.

D. INSTITUTO NACIONAL DE ESTADÍSTICA

El Instituto Nacional de Estadística (INE) es un organismo autónomo adscrito al Ministerio de Economía y Hacienda. Su tarea principal y de mayor tradición es la de elaborar estadísticas públicas, que son estudios oficiales sobre la situación y evolución de la población, la economía y la sociedad de España.

E. EUROSTAT

Eurostat, la Oficina Estadística de la Unión Europea, se encarga de publicar estadísticas e indicadores de alta calidad a escala europea que permitan hacer comparaciones entre países y regiones. Sus principales responsabilidades son:

- desarrollar definiciones, clasificaciones y metodologías armonizadas para la elaboración de las estadísticas oficiales europeas, en colaboración con las autoridades estadísticas nacionales.
- calcular los datos agregados para la Unión Europea y la zona del euro a partir de la información recopilada por las autoridades estadísticas nacionales según normas armonizadas.
- poner estadísticas europeas a libre disposición de los responsables y los ciudadanos a través de la web de Eurostat y otros canales.

F. INSTITUTO GEOGRÁFICO NACIONAL

La Dirección General del Instituto Geográfico Nacional (IGN) tiene competencias en:

- Desarrollo de planes de observación del territorio y aplicación cartográfica.
- Producción y gestión de bases de datos geoespaciales.
- Programación y producción de cartografía nacional.
- Producción de cartografía temática para programas gubernamentales.
- Soporte técnico al Sistema Cartográfico Nacional, incluyendo la gestión del registro cartográfico y el mantenimiento del Nomenclátor Geográfico Nacional.

G. GOOGLE PLACES

La API de Google Places permite a los usuarios extraer el estado operativo de cada lugar identificado en la plataforma Google Places. Los lugares reciben el estado de "Abierto", "Cerrado temporalmente" o "Cerrado permanentemente". Esta información puede utilizarse para obtener varios indicadores útiles de la dinámica empresarial. Si suponemos que la plataforma Google Places tiene una cobertura casi censal de todos los lugares que operan en una región y que el número de reseñas es una buena indicación del tamaño relativo de un lugar con respecto a otro,



podemos utilizar esta información para medir el estado operativo de los lugares en cada zona geográfica.⁸ Dado que existe una fuerte relación entre el estado operativo del negocio y sus ingresos y empleo, estos indicadores podrían ser útiles para proporcionar una señal temprana de las tendencias en el mercado laboral o las tendencias en la actividad económica agregada para la región.

H. OTRAS BASES DE DATOS

- OpenStreetMap ofrece un mapa del mundo gratuito y de código abierto con el que puedes acceder a información sobre empresas, transportes y puntos de interés. Planet OSM es una función de OpenStreetMap que permite extraer millones de puntos de interés de forma gratuita.
- La API de Foursquare Places proporciona información de ubicación sobre locales, usuarios, fotos y check-ins. El formato de respuesta preferido es JSON. Puedes utilizar esta API para ayudar a los usuarios a encontrar lugares locales a través de elementos como recomendaciones, reseñas, fotos y comentarios.
- Si eres usuario de ArcGIS, puedes utilizar su servicio de geocodificación para buscar un lugar y devolver direcciones completas. También puede utilizarlo para buscar el nombre de un negocio situado a una distancia determinada de una ubicación o encontrar lugares por el nombre de una categoría. También puede utilizar la búsqueda de lugares dentro de ArcGIS para buscar ubicaciones geográficas y negocios o para mostrar lugares en un mapa.
- Con la API de contenido de Tripadvisor puede acceder a información sobre destinos turísticos para utilizarla en su sitio web o aplicación. Los datos se devuelven en formato JSON.
- Amazon Location Service le permite integrar datos geospaciales en sus aplicaciones procedentes de diversos proveedores de datos.



4 OBJETIVOS DEL RETO

El propósito del reto consiste en construir un dataset construido sobre los estándares de la web semántica y los datos abiertos y enlazados que permita caracterizar el grado de dinamismo económico de los municipios y áreas de interés estadístico de la Comunidad de Madrid que permita:

- a) Obtener nuevos datasets que incluyan indicadores relativos al dinamismo económico como los propuestos en la sección 1.3 (u otros que se puedan definir) en una serie temporal y localizados geográficamente que permita la evaluación y análisis de políticas relacionadas con el dinamismo económico en los municipios de la Comunidad de Madrid.
- b) Publicar los datasets en formato Linked Open Data de tal manera que sean accesibles para la comunidad de investigación y contribuyan al análisis de la cohesión territorial en la Comunidad de Madrid
- c) Visualizar este dataset sobre un visor GIS web de código abierto que acceda a los datasets generados en el trabajo junto con otros de interés que permitan enlazar los datos generados con otras áreas de evaluación de la cohesión territorial como: a) la dimensión social del territorio; b) medioambiente como soporte de la vida en los municipios; c) gobernanza, cooperación institucional, planificación y desarrollo regionales; y d) servicios, infraestructuras y equipamientos a disposición del ciudadano.

Los aspectos que se considerarán para la evaluación de propuestas son:

- Adherencia a los estándares de web semántica.
- Metodología y su automatización para la generación del dataset que recoja datos existentes de distintas bases de datos de interés, combinando información procedente de la Comunidad de Madrid y otras que no estén gestionadas directamente por entidades públicas.
- Fiabilidad de los datos generados de acuerdo con la calidad de la metodología de adquisición de la información de entrada para el dataset necesario.
- Disponibilidad de una aplicación web GIS que permita la navegación sobre los datos incluidos en los datasets generados.
- Ejemplos que permitan relacionar el dataset generados con otros disponibles en formato de la web semántica
- Facilidad para la integración con otros servicios de información semántica gestionados por la cátedra para I+D+i sobre dinamismo territorial en la Comunidad de Madrid con el uso de datos abiertos y enlazados (LoDCOREMadrid).