



POLITÉCNICA
"Ingeniamos el futuro"

CAMPUS
DE EXCELENCIA
INTERNACIONAL



Grado en Ingeniería Informática

Universidad Politécnica de Madrid

Escuela Técnica Superior de
Ingenieros Informáticos

TRABAJO FIN DE GRADO

TransportDCAT-AP Validator: web para la
validación de perfiles de metadatos en el dominio del
transporte público

Autor: Mahdi Alaoui Sossai

Director: Oscar Corcho

MADRID, ENERO 2018

ÍNDICE

ÍNDICE DE FIGURAS	4
AGRADECIMIENTOS	5
RESUMEN DEL TRABAJO REALIZADO.....	7
ABSTRACT	8
INTRODUCCIÓN	9
Metadatos	9
¿Qué son los metadatos?	9
Portales de datos abiertos en el transporte público	11
Proyecto Europeo: CEF-OASIS	12
Motivación	12
ESTADO DEL ARTE	14
WEB SEMÁNTICA Y DATOS ENLAZADOS	14
FORMATOS DE LA WEB SEMÁNTICA	14
Ontologías	14
Datos Enlazados	15
DCAT	16
Namespaces (Espacio de nombres)	16
Resumen Vocabularios.....	17
DCAT-AP: Data Catalog Vocabulary – Application Profile	20
¿Qué es DCAT-AP?	20
Principales y características	21
Clases y Propiedades	22
Herramienta de validación de vocabularios de metadatos	23
DCAT-AP Validator	23
Conclusión	24
TransportDCAT-AP.....	25
Introducción	25
Descripción	25
Clases y propiedades.....	27
Dcat-Catalog.....	27
Dcat-Dataset	28
Foaf-Agent	29
Dcat-Distribution.....	30
Dcat:CatalogRecord – Recommended Class.....	31
Spdx:Checksum.....	31
Dct:PeriodOfTime	31
OBJETIVOS	32
DESARROLLO	33
Tecnologías de Desarrollo	33



HTML5.....	33
CSS3:.....	33
Typescript:.....	33
Superset:	33
Angular.....	34
Git.....	34
NPM (Node Package Manager)	34
Tecnologías de metadatos.....	34
XML	34
RDF (Resource Description Framework)	34
Librería utilizada	35
rdflib.js	35
Aplicación web	36
Estudio Funcionalidad.....	36
Fases del Proyecto	36
Diseño de la Aplicación	36
Implementación y Visualización Final.....	36
Pruebas Finales y Correcciones.....	39
CONCLUSIONES	42
BIBLIOGRAFÍA	43



ÍNDICE DE FIGURAS

Figura 1: Logo de TransportDCAT-AP validator	7
Figura 2: Fragmento del diagrama Datos enlazados	15
Figura 3: Modelo DCAT	17
Figura 4: Catalog & Dataset	25
Figura 5: Representacion Grafica de u triplete	35
Figura 6: Componente Upload File Transport-DCAT-AP	38
Figura 7: Componente Input TransportDcat-AP	38
Figura 8: Error parser TransportDCAT-AP	39
Figura 9: Falta de Clase Obligatoria dcat:Catalog ,	40
Figura 10: Parseado con Exito TransportDCAT-AP	40
Figura 11: Error rdf /xml	41



AGRADECIMIENTOS

Después de un intenso tiempo de trabajo sobre este trabajo fin de grado, llega el momento de agradecimientos a todas aquellas personas que me han apoyado, no solo a nivel académico, pero también a nivel personal.

En primer lugar quiero agradecer a mis padres por haber sido mis compañeros en la vida, por apoyarme y luchar conmigo en el camino que he elegido. Por haber sido mi fuente de inspiración y de motivación, por los consejos y la paciencia. Muchas gracias por haberme ofrecido todo lo que era necesario para seguir adelante.

Quiero agradecer a mis hermanos y hermanas por respaldarme y estar siempre a mi lado, gracias a ellos he tenido muchísima protección y me hicieron sentir feliz siempre.

Quiero agradecer a mi tutor Oscar Corcho por haberme dado la oportunidad de realizar este proyecto tan interesante. Quiero agradecerle también el apoyo que me ha empujado a terminar el trabajo de la mejor forma posible.

En particular quiero agradecer muchísimo a David Chaves que ha sido mi guía en este proyecto, gracias por el esfuerzo, por los consejos y las recomendaciones que me ha dado, y que han sido determinantes para la finalización de este proyecto.



A María, Belrhite, Rachid, Simo, Sakina, Afafe, Oscar y David.

RESUMEN DEL TRABAJO REALIZADO

En el presente documento se describe el trabajo de análisis y el diseño de la aplicación TransportDCAT-AP Validator, una aplicación web para la validación de perfiles de metadatos basados en DCAT-AP y orientados al sector de transporte público.

TransportDCAT-AP es un perfil de metadatos desarrollado en el contexto del proyecto europeo CEF-OASIS¹ y su principal objetivo es proporcionar un vocabulario común para la representación de metadatos en el dominio del transporte público a nivel europeo. Se basa en el vocabulario estándar europeo para la creación de perfiles de metadatos, DCAT-AP.

El objetivo principal de este trabajo es proporcionar una interfaz sencilla para que las empresas de transporte y organismos públicos puedan utilizar la herramienta como validador de sus metadatos e integrarlos en sus portales de datos abiertos. Para ello se cree que la mejor manera de ofrecer este servicio y que sea accesible por cualquier institución o empresa es a través de un servicio web que de forma online proporcione los errores y recomiende la forma de solucionarlos dado un vocabulario creado.



Figura 1: Logo de TransportDCAT-AP validator

Para realizar el trabajo se ha dividido el proyecto en varias tareas que se especifican a continuación:

- Realización de un estudio profundo sobre que son los metadatos y porque son necesarios. El análisis se realizará, primero de forma general y a continuación de forma más específica centrándose en vocabularios semánticos estándares para la representación de estos datos.
- Debido a que la herramienta web desarrollada tratará de validar datos representados a través de vocabularios semánticos, se deberá realizar un estudio sobre qué es la Web Semántica, las Ontologías y las diversas tecnologías que se han desarrollado en estos campos para representar vocabularios.
- Hacer un análisis del diseño y del desarrollo de la aplicación, decidiendo que tecnologías se van a utilizar con que forma los vamos a hacer. Se ha hecho una investigación sobre las librerías disponibles y que sean compatibles al objetivo del proyecto.
- Dividir las funcionalidades que se requieren para la herramienta en tareas, que se irán cumpliendo según se vaya desarrollando el proyecto.
- Una vez se terminado el proyecto, se procederá a hacer las pruebas, arreglar los errores y publicar el servicio en una URL pública.

¹ <https://oasis.team/>



ABSTRACT

This document describes the analysis and design of the TransportDcat-AP web application for the validation of metadata profiles based on DCAT-AP but focused on the public transport sector.

It is fundamental first of all to make a definition of TransportDcat -AP which is a profile based on the standard proposed by the European Union DCAT-AP that uses RDF/XML Vocabulary for the presentation of metadata.

The main objective of this project is to provide a simple interface for companies and institutions that can validate their metadata to integrate them into their open data portals. The main service developed for this tool consists of parsing the metadata presented in RDF/XML and then making a TransportDcat-AP parser.

For the follow-up of the work, the work has been divided into several tasks to be specified below:

- To start this project has had to do a thorough analysis of all the concepts and subjects related to the main topic of metadata and understand perfectly the use of it, understand what is the semantic web, its framework, syntax of RDF/XML and what official standards must be used to have a global idea of the subject as and try to have a good start so that you can advance in the project and meet all the requirements and objectives determined in the project.
- An analysis has been made of how the interface is going to be and with which technologies, language and programming environment the application would be developed, it was necessary to look deeply for libraries that are available to use them and that could match the objective of our project.
- It has been decided how the design of the tool and its main components would be. When making a first design of the application, the functionalities required for the tool are divided into tasks that will be carried out as we develop our project.
- Once the project is finished, we will proceed to test and fix the errors and ultimately get the desired results.

INTRODUCCIÓN

A lo largo de esta sección se realizará una introducción de los temas más relevantes que se han tenido en cuenta a la hora de desarrollar este proyecto. Se realizará una descripción de qué son los metadatos y porque son necesarios, cuál es el actual estado de portales de datos abiertos con respecto al tema de los metadatos, centrados en el portal de datos abiertos del Consorcio Regional de Transportes de Madrid² (CRTM) y, por último, una breve descripción del objetivo principal del proyecto europeo CEF-OASIS que ha dado soporte al desarrollo de este proyecto.

Metadatos

¿Qué son los metadatos?

Hoy en día la cantidad de información y datos que existen en internet se hizo necesario establecer una serie de mecanismos para etiquetar, catalogar, describir, clasificar y gestionar los recursos. El objetivo principal de estos mecanismos es facilitar la posterior búsqueda y recuperación de la información de una forma común y estándar y se han formalizado en los llamados metadatos.

Los metadatos son datos que describen a otros datos. Es una información estructurada que describe a otra información y que permite mejorar el descubrimiento de esta, gestionarla, controlarla, entenderla y preservarla en el tiempo. Además, este tipo de datos es capaz de caracterizar otros datos, describir su contenido, su calidad, sus condiciones, su historia, su disponibilidad y otra serie de características.

La primera aplicación de los metadatos fue en el mundo de las bibliotecas para gestionar libros y archivos. Sin embargo, con el nacimiento de la informática y la World Wide Web se vio la necesidad de adaptarlos a este mundo para gestionar la inmensa cantidad de información que se genera diariamente. De esta forma, a través de los metadatos es posible, por ejemplo, realizar descubrimientos de datos relevantes de un dominio a través de consultas estándares en portales de datos abiertos. En el caso de la ciudad de Madrid, el portal de datos abiertos del ayuntamiento³ representa todos sus conjuntos de datos siguiendo el estándar DCAT⁴.

La generación de los metadatos no solo es aplicable a la información digital, también a cualquier conjunto de datos independientemente del soporte en el cual se encuentren. Ya que ello puede facilitar su localización y se consigue así, añadir valor a la información histórica con la que cuenta una entidad.

A continuación, se describirán cuáles son los tipos de metadatos existentes y sus principales funciones:

² <http://crtm.es/>

³ <http://datos.madrid.es/portal/site/egob/>

⁴ <https://www.w3.org/TR/vocab-dcat/>

Tipos de Metadatos

En general, los metadatos pueden clasificarse en tres amplias categorías:

- **Metadatos descriptivos:** son aquellos que sirven para la descripción e identificación de los recursos de información. Permiten la búsqueda y recuperación de los datos, así como distinguir recursos y entender sus contenidos. Se realizan mediante estándares como microdatos de HTML [1].
- **Metadatos estructurales:** son los que más influyen en la recuperación de la información, facilitando la navegación y presentación de los recursos. Ofrecen información sobre su estructura interna y permiten establecer relaciones entre ellos, de manera que pueden incluso unir los archivos de imagen y textos que están relacionados. Los estándares más difundidos para en este sentido son XML, RDF.
- **Metadatos administrativos:** incluyen datos para la gestión de derechos, firma digital, auditorías de calidad, control de acceso mediante contraseñas... etc. También se podrían enmarcar en esta categoría los metadatos sobre preservación a largo y corto plazo.

Para que sirven los metadatos

Las funciones principales de los metadatos son:

- **Búsqueda:** los metadatos deben proporcionar suficiente información, bien para descubrir si existen datos de interés dentro de la colección de datos disponibles, o simplemente, para saber que existen.
- **Recuperación:** los metadatos deben proporcionar información a los usuarios para que puedan adquirir la información que sea de su interés. La analogía con una biblioteca consistiría en el procedimiento a seguir para sacar un libro. El componente que recupera los datos desde el metadato puede ser tan simple como proporcionar un URL que identifique la localización de un conjunto de datos digitales, o tan complejo como para cubrir cuestiones de seguridad o realizar una transacción financiera para poder acceder a la información (compra en línea). En este sentido, también se considera la "función recuperación" a aquella información que describe cómo localizar fuera de línea los datos, la persona de contacto, los formatos de distribución de los datos o cualquier restricción de acceso a los datos, así como la información sobre los costes.
- **Transferencia:** los metadatos deben facilitar la información necesaria para que los usuarios hagan uso de los archivos recuperados en sus máquinas. Este componente incluirá información sobre el tamaño del conjunto de datos (y sus metadatos), la estructura tanto lógica como física de los datos y metadatos.

- **Evaluación:** los metadatos deben considerar información que asista a los usuarios a determinar si los datos van a ser útiles para una aplicación.
- **Interoperabilidad:** los metadatos facilitan la interoperabilidad, puesto que se han definido estándares de metadatos y existen protocolos compartidos para el intercambio de esta información.

Portales de datos abiertos en el transporte público

Para comprobar el estado actual en el que se encuentran los portales de datos abiertos de transporte público respecto al uso y creación de metadatos se ha hecho un análisis del uno de los portales de datos abiertos más relevantes a nivel nacional, el del CRTM⁵.

Este portal provee datos de horarios, rutas o paradas de todos los medios de transporte que existen en la comunidad autónoma de Madrid: metro, cercanías, metro ligero y autobuses. En la Tabla 1 se indica la información más relevante en este portal de datos abiertos:

Nº Conjunto de datos	Tipo Transporte	Ficheros de Metadatos
4	Bus Interurbano	1
4	EMT Bus	1
4	Madrid Urbano	1
102	Metro	2
30	Metro Ligero	2
77	Cercanías	2

Tabla 1 - Resumen del portal de datos abiertos del CRTM

Respecto a los metadatos se observa en la tabla que hay 9 archivos diferentes de metadatos representando la información relacionado con cada tipo de transporte en el portal. Cada fichero se basa en la norma ISO19139 y contienen la siguiente información:

⁵ <http://datos.crtm.es/>

- CreaDate: Fecha de la creación del Catálogo.
- idAbs: Descripción del Catálogo.
- Keywords: Colección de las palabras importantes
- idPur: descripción simple del catálogo
- idCredit: CRTM
- useLimit: Licencia
- Data: de tipo Binario.

Después del análisis de los metadatos se ha notado que la actual información disponible no es suficiente para un portal que ofrece todos esos datos del sector del transporte público. Es por ello que se llegó a la conclusión de que existe una necesidad y motivación para el desarrollo de un del modelo de metadatos del dominio basado en el estándar europeo DCAT-AP.

Además, los conjuntos de datos del portal se pueden descargar desde el portal de datos abiertos en diferentes formatos (GTFS, SHP, KML, SIG, and CSV), también existe una API para realizar consultas y una opción de visualización de las rutas y las paradas de forma online.

Proyecto Europeo: CEF-OASIS

El proyecto CEF-OASIS, que ha dado soporte para la implementación del validador de TransportDCAT-AP, es un proyecto financiado por el programa CEF-Telecom y en el que participan instituciones de Bélgica (Open Knowledge Belgium, Stad Gent, imec y V-ICT-OR) y España (Ontology Engineering Group - UPM - y Consorcio Regional de Transportes de Madrid - CRTM).

El proyecto se centra en la apertura y reutilización de datos en el ámbito de los servicios públicos ofrecidos por entidades locales y en el sector del transporte público. Para ello, se trabajará en el uso y extensión de estándares europeos como el DCAT-AP (que se va a definir en los siguientes capítulos con más detalle) y resultados de iniciativas nacionales, regionales o locales, en ambos sectores.

En mayo de 2017, este proyecto proporciona de forma abierta la documentación asociada al perfil TransportDCAT-AP⁶, que como se ha comentado anteriormente, se centra en ofrecer un vocabulario común para la representación de metadatos en el dominio del transporte público a nivel europeo.

Motivación

Hoy en día existen muchos datos abiertos en el dominio del transporte en toda Europa, ya sea en portales de datos de instituciones u organismos públicos que mantienen la gestión de estos servicios o en portales de datos de organizaciones de transporte. Uno de los grandes objetivos de las llamadas Smart Cities es el desarrollo de aplicaciones que mejoren la movilidad en las grandes ciudades con el fin de proporcionan información

⁶<https://oasis.team/storage/app/media/O1.2%20TransportDCAT-AP%20and%20Controlled%20Vocs.pdf>



fiable y de calidad a los usuarios y promover el uso del transporte público para reducir la contaminación.

Uno de los grandes problemas con los que tienen que lidiar los desarrolladores de estas aplicaciones es descubrir dónde se encuentran estos datos y como obtenerlos y usarlos. Es por ello que dotar a estos desarrolladores de un canal para encontrar estos datos de forma estándar es esencial. Y es esta, la principal motivación del desarrollo de TransportDCAT-AP.

Una vez desarrollado el perfil, se ha visto la necesidad de hacer un validador para el mismo. Primero, porque la creación de estos vocabularios no es un tarea trivial y puede ser compleja para las organizaciones o instituciones que no hayan trabajado nunca con este tipo de tecnologías y segundo, como un método fiable para que la calidad de los metadatos sea lo más alta posible y coincida con lo especificado en la documentación de TransportDCAT-AP.

ESTADO DEL ARTE

WEB SEMÁNTICA Y DATOS ENLAZADOS

En general la web semántica es la web de los metadatos que consiste en añadir metadatos en las páginas web que describen el contenido, el significado y la relación entre los datos.

Es una extensión de World Wide Web (www), dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar resultados de sus consultas de forma más rápida y sencilla gracias a una información mejor definida.

La idea principal de la web semántica es mediante los metadatos ampliar la interoperabilidad entre las máquinas sin que el usuario se vea obligado a intervenir de forma activa.

FORMATOS DE LA WEB SEMÁNTICA

El W3C [2] lleva años trabajando en varias tecnologías y, básicamente, en todas ellas se está utilizando sintaxis XML que tiene la capacidad de representar ontologías y vocabularios. Una de estas tecnologías, la más destacada en nuestro Proyecto, es RDF que será definida y descrita más adelante en la sección de desarrollo. Entre otras existen RDFa⁷, RDF [3], RDFS⁸ (una extensión de RDF) y OWL.

El hecho de presentar los metadatos y los datos en RDF permite aumentar la visibilidad de los recursos en internet y en especial su presencia en la web semántica.

Ontologías

Según Gruber [3] la definición más concreta es: *“Una ontología es la especificación explícita de una conceptualización.”*

Y en el contexto de la web semántica, se considera la definición presentada en la W3C:

“Una ontología define formalmente un conjunto común de términos, los cuales son usados para describir y representar un dominio.... Una ontología define los términos usados para describir y representar un área de conocimiento.”

Como referencia a esta definición, las ontologías se enfocan en un dominio específico, este dominio simplemente es un área de conocimiento por ejemplo área de transporte, educación.... Etc. Las ontologías contienen términos y relaciones en estos términos.

Se pueden definir las ontologías de forma más explicativa indicando los componentes que las forman, proporcionan un vocabulario común de un área y definen el significado de los términos y relaciones entre ellos.

⁷ <https://www.w3.org/TR/rdfa-primer/>

⁸ <https://www.w3.org/TR/rdf-schema/>

Las ontologías tienen los siguientes componentes:

- **Conceptos:** son las ideas básicas que se intentan formalizar, los conceptos pueden ser clases de objetos o métodos por ejemplo.
- **Relaciones:** representan la interacción entre los conceptos de algún dominio, suelen formar la taxonomía del dominio, por ejemplo: una clase es una subclase-de, parte-de, etc.
- **Funciones:** son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función, que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como: asignar-fecha, categorizar-clase, etc.
- **Instancias:** Son objetos, miembros de una clase que podrían estar agrupados en clases.
- **Axiomas:** son expresiones que son siempre ciertas. Por ejemplo: Son expresiones que son siempre ciertas.

Las ontologías como se puede ver, son un conjunto de conceptos-definiciones que se pueden ordenar en jerarquías de taxonomías y tener propiedades asociados. Dichos conceptos se pueden presentar en lenguajes como el RDF/XML y se pueden compartir por medio de un dominio.

Datos Enlazados

Método con el que se pueden mostrar, intercambiar y conectar datos a través de URI Desreferenciables en la Web [4].

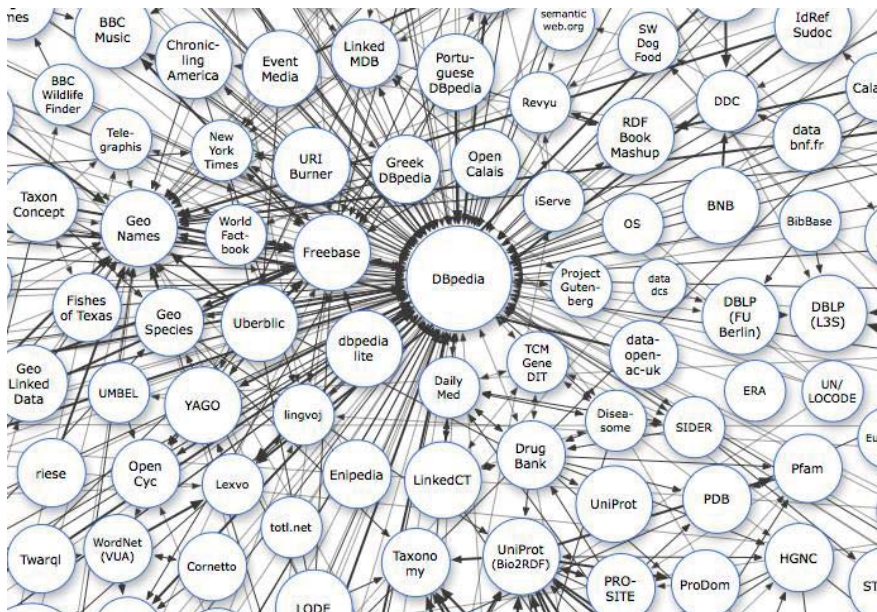


Figura 2: Fragmento del diagrama Datos enlazados

Utilizando estándares y el protocolo HTTP, se dará la posibilidad de acceder a conjuntos de datos y analizarlos para descubrir conexiones con otros conjuntos de datos

a partir de los vínculos establecidos entre ellos. Linked Data (Datos Enlazados) se basa en una serie de principios definidos por Tim Berners-Lee (Berners-Lee, 2006) [5]:

- Utilizar URIs como nombres para objetos y recursos.
- Que se pueda acceder a esos datos a través de la web a través del protocolo HTTP. La técnica es utilizar URIs http para que los usuarios puedan ver esos nombres.
- Cuando alguien busca una URI, proporcionar información útil usando los estándares apropiados (RDF, SPARQL⁹). SPARQL es un lenguaje estandarizado para la realización de consultas sobre RDF.
- Incluir enlaces a otras URIs, de tal forma que se pueda recuperar más información.

La figura 2 representa la gran cantidad de datos conectados existentes en la Web como datos enlazados.

DCAT

Según W3C, DCAT [6] es un vocabulario RDF, diseñado para interoperar entre catálogos de datos publicados en la web.

Fundamentalmente las siglas RDF significan Resource Definition Framework (marco de definición de recursos). Un vocabulario RDF es un vocabulario que define un recurso web, en general se está hablando de cualquier tipo de datos de la web. Todo esto está impulsado por la web semántica.

Namespaces (Espacio de nombres)

El espacio de nombres para DCAT es <http://www.w3.org/ns/dcat#>. Dcat utiliza términos de otros vocabularios en particular Dublin Core¹⁰, pero tiene definidas algunas propias clases y propiedades.

Se utilizarán los siguientes espacios de nombres asociados con los prefijos:

⁹ SPARQL (Protocol and RDF Query Language) <http://www.w3.org/TR/rdf-Sparql-query>

¹⁰ dublincore.org

Prefix	Namespace
dcat	http://www.w3.org/ns/dcat#
dct	http://purl.org/dc/terms/
dctype	http://purl.org/dc/dcmitype/
foaf	http://xmlns.com/foaf/0.1/
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
skos	http://www.w3.org/2004/02/skos/core#
vcard	http://www.w3.org/2006/vcard/ns#
xsd	http://www.w3.org/2001/XMLSchema#

Tabla2: Espacios de nombres DCAT

Resumen Vocabularios

La arquitectura de un catálogo de datos es la de la figura siguiente:

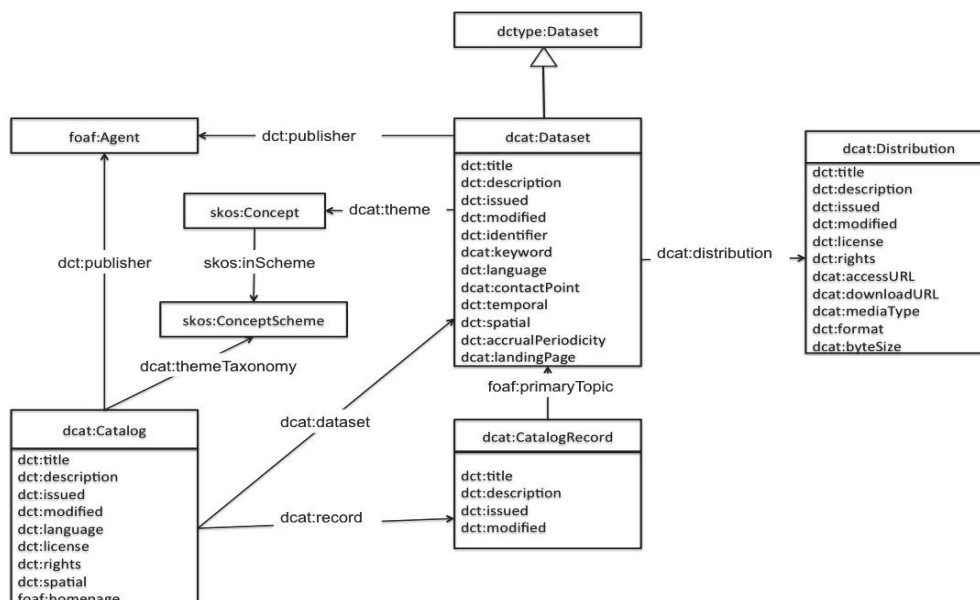


Figura 3: Modelo DCAT

Desde la figura insertada, vemos el modelo DCAT, representando un catálogo de datos.

Se va a definir las 3 clases importantes de DCAT, Catalogo, Dataset o conjunto de datos, y la distribución.

Dct:Catalog

La clase dcat:Catalog funciona como contenedor donde se encuentran los recursos denominados dcat:dataset, que son distribuidos a través de distribuciones dcat:Distribution.

Esta clase permite identificar al catálogo que proporciona los recursos.

Esta clase está compuesta de las siguientes propiedades

Propiedad RDF	Descripción
dct:title	Breve título o nombre dado al catálogo de datos.
dct:description	Resumen descriptivo del catálogo de datos.
dct:issued	Fecha de publicación inicial del catálogo.
dct:modified	Fecha en la que se modificó por última vez el catálogo.
dct:language	Idioma en el que se proporciona la información del catálogo.
foaf:homepage	Dirección web de acceso al catálogo de datos (acceso para el público).
dct:publisher	Entidad que publica el catálogo.
dct:spatial	Ámbito geográfico cubierto por el catálogo.
dcat:themeTaxonomy	Totalidad de materias incluidas en el catálogo.
dct:license	Referencia a los términos de uso generales del catálogo.
dct:rights	Describe los derechos de la utilización o la reutilización del catálogo.
dcat:dataset	Lista de cada uno de los recursos de información del catálogo.
dcat:record	Es una parte del catálogo.

Dcat-Dataset

Es el ‘conjunto de datos’ que son publicados y mantenidos por un sólo agente y están disponibles en uno o más formatos. Al final es un conjunto coherente de recursos disponibles, y es muy importante, que la responsabilidad de su actualización y publicación recaiga sobre el mismo agente.

Esta clase está compuesta de las siguientes propiedades:

Propiedad RDF	Descripción
dct:title	Nombre o título del recurso de información.
dct:description	Descripción detallada del conjunto de datos.
dct:issued	Fecha de creación del dataset.
dct:modified	Última fecha conocida en la que se modificó o se actualizó el dataset.
dct:language	Idioma en el que se encuentra la información del dataset.
dct:publisher	Organismo o entidad que publica el dataset.
dct:accrualPeriodicity	Periodo de tiempo aproximado entre actualizaciones del dataset, si hubiera.
dct:identifier	URI que identifica el dataset.
dct:spatial	Ámbito geográfico cubierto por el dataset.
dct:temporal	Fecha de inicio, fin y la duración del periodo cubierto por el dataset.
dcat:theme	Temática o materia primaria del dataset.
dcat:keyword	Etiqueta(s) textual(es) que permiten categorizar libremente dataset.
dcat:contactPoint	Nombre de contacto que atenderá dudas y comentarios sobre el conjunto de datos
dcat:distribution	Referencia a los recursos que identifican los volcados del dataset en sus posibles formatos.
dcat:landingPage	Una página Web que se puede navegar en un navegador para acceder a la base de datos, sus distribuciones y / o información adicional
dct:rights	Describe los derechos de la utilización o la reutilización del dataset.

Dcat-Distribution

La distribución es, en definitiva, un recurso con los mismos datos publicados de forma diferente. Por ejemplo los datos se pueden presentar en varios formatos como PDF, Excel, CSV...etc. Cada distribución está en un formato dado.

Esta clase está compuesta de las siguientes propiedades:

Propiedad RDF	Descripción
dct:title	Breve título o nombre dado a la distribución.
dct: description	Descripción detallada de la distribución.
dct:issued	Fecha de creación de la distribución
dct:modified	Última fecha conocida en la que se modificó o actualizó el dataset.
dct:license	Referencia a los términos de uso generales del catálogo.
dct:rights	Describe los derechos de la distribución.
dcat:accessURL	URL de acceso al fichero con los datos del conjunto de datos de esta distribución.
dcat:downloadURL	Una URL específica para descarga, para descargar el fichero de la distribución.
dcat:byteSize	Tamaño en bytes del fichero de la distribución.
dcat:mediaType	Formato en que se encuentra representada la distribución
dct:format	Formato de la distribución.

DCAT-AP: Data Catalog Vocabulary – Application Profile

Una de las tendencias que se están utilizando en el mundo de datos abiertos es el DCAT-AP.

¿Qué es DCAT-AP?

LA Comisión Europea da un paso importante al terminar la especificación DCAT-AP elaborada conjuntamente por el programa ISA, la oficina de publicaciones y DG Connect. Esta especificación es una extensión del vocabulario W3C Data Catalogue Vocabulary (DCAT), y la definición de una política normativa para su aplicación en la descripción de los conjuntos de datos públicos de Europa.

AP (application profile) es una especificación que reutiliza términos de uno o más de estándares pero en nuestro caso es DCAT, añadiendo especificidades por la identificación de elementos obligatorios recomendables y opcionales para ser usados en una aplicación específica.

DCAT-AP describe catálogos de conjuntos de datos lo que se llaman datasets del sector público en Europa, y por la tanto busca resolver la problemática de la conexión de datos entre portales que publican datos de la UE y sus principales objetivos son:

- Permitir la búsqueda de conjuntos de datos a lo largo de varios portales.
- Fomentar la reutilización efectiva de la información.
- Mantener la interoperabilidad entre las distintas fuentes.

En Octubre 2015 se ha publicado la versión actual 1.1, muchos portales Europeos han adoptado el DCAT-AP como el portal PanEuropeo y el Semantic Interoperability Community (SEMIC)... etc.

En general DCAT-AP utiliza vocabulario RDF destinado a facilitar la interoperabilidad entre catálogos de datos publicados en la web.

Principales y características

Desde el documento oficial publicado de DCAT-AP se han identificado los siguientes objetivos:

- **Mejorar el descubrimiento de los conjuntos de datos:** Se dio prioridad a aquellos aspectos que pueden utilizarse en la búsqueda y la navegación.
- **Asegurar compatibilidad con DCAT:** se van a utilizar las propiedades de DCAT, pero se han agregado elementos que puedan ser reutilizados desde un namespace (Lo voy a indicar en la siguiente parte).
- **Define una política completa para el uso de DCAT-AP:** Especificando qué clases y propiedades son obligatorias, recomendadas u opcionales en la aplicación del vocabulario dentro de la Unión Europea.

El DCAT-AP reutiliza términos existentes, se ha tenido en cuenta estos nombres para definir las clases y propiedades que se van a definir en la siguiente sección de este mismo capítulo.

Prefijo	URI
adms	http://www.w3.org/ns/adms#
dcat	http://www.w3.org/ns/dcat#
dct	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
owl	http://www.w3.org/2002/07/owl#
rdfs	http://www.w3.org/2000/01/rdf-schema#

schema	http://schema.org/
skos	http://www.w3.org/2004/02/skos/core#
spdx	http://spdx.org/rdf/terms#
xsd	http://www.w3.org/2001/XMLSchema#
vcard	http://www.w3.org/2006/vcard/ns#

Clases y Propiedades

Como se ha mencionado antes en DCAT-AP se han especificado las clases obligatorias y recomendadas u opcionales.

✓ Clases Obligatorias

Nombre De clase	Nota de uso	URI	REFERENCE
AGENT	Es una entidad que está asociada a catalogos y datasets , el agent es una organización, esta recomendado hay que usar la antología organization	FOAF:AGENT	http://xmlns.com/foaf/spec/#term_Agent http://www.w3.org/TR/vocab-org/
Catalogue	Se describe el catalogo que incluye el datasets descritos	Dcat:Catalog	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog
Dataset	Es una entidad que representa la información publicada.	Dcat:Dataset	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-dataset
Literal	El valor literal puede ser un string por ejemplo podría ser una fecha.	dfs:Literal	http://www.w3.org/TR/rdf-concepts/#section-Literals
Resource	Cualquier cosa que se pueda describir en RDF	rdfs:Resource	http://www.w3.org/TR/rdf-schema/#ch_resource

✓ Clases Recomendadas

Nombre De clase	Nota de uso	URI	REFERENCIA
Category	Es el sujeto de dataset	Skos:Concept	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-category-and-category-scheme
Category Scheme	Ese el concepto de una colección de datos	Skos:ConceptScheme	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-category-and-category-scheme
Distribution	Indica en que formato se puede descargar el dataset.	Dcat:Distribution	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-distribution
Licence document	Es un documento legal que da un permiso oficial para hacer alguna modificación	dct:LicenseDocument	http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#LicenseDocument

✓ Clases opcionales

Nota: en las clases opcionales solo voy a definir las clases más destacadas, ya que todas están definidas en el manual oficial y es de acceso público. Está disponible en el siguiente enlace:

https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-ap-v11

Nombre De clase	Nota de uso	URI	REFERENCIA
Catalogue Record	Es una descripción de la entrada de dataset en el catalogo	dcat:CatalogRecord	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog-record
Checksum	Es un valor que permite la autenticación para ver el contenido de los ficheros	spdx:Checksum	http://spdx.org/rdf/terms#Checksum
Document	Es un recurso textual que contiene y está destinado al usuario	foaf:Document	http://xmlns.com/foaf/spec/#term_Document
Frequency	Es como un índice que indica con qué proporción se está repitiendo algo.	dct:Frequency	http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#LicenseDocument
Identifier	Es identificador de un contexto en particular, el identificador puede ser un string por ej.	adms:Identifier	http://www.w3.org/TR/vocab-adms/#identifier

Hay muchas más propiedades opcionales y obligatorias, se ha descrito solo las más destacadas de las clases ya que en el perfil TransportDCAT-AP está basado en la reutilización de las clases y propiedades, entonces es imprescindible describir todo muy detallado en la sección exclusiva del perfil mencionado.

Herramienta de validación de vocabularios de metadatos

Existe una herramienta que se ha tomado como referencia para este trabajo de validación de perfiles basados en TransportDcat-AP. Dicha Herramienta es DCAT-AP Validator [7] es para validar metadatos basados en DCAT-AP se puede en utilizar a través de la web en este enlace¹¹.

DCAT-AP Validator

Es un proyecto de Open knowledge Belgium¹², esta herramienta ha sido creado por dos alumnos Stan Callewaert and Sébastien Henau. La función principal de esta herramienta

¹¹ <http://www.dcat.be/validator/>

¹² <https://www.openknowledge.be/>



es validar metadatos basados en DCAT-AP, demuestra errores y advertencias de las clases obligatorias y sus propiedades.

Dicha aplicación da la posibilidad de validar los metadatos de forma manual poniéndolos en una área de texto, a través de un fichero o desde una URI.

Se puede validar en 3 formatos, RDF/XML, JSON-LD y turtle.

Conclusión

Como conclusión del estudio y la investigación realizados, se ha podido obtener una serie de conceptos que han tenido mucha utilidad para este proyecto.

Era fundamental entender todos los conceptos y conocer todos los vocabularios relacionados con el DCAT y el DCAT-AP, porque son la base del TransportDCAT-AP.

Como idea que hay para el desarrollo del proyecto es ofrecer un servicio al usuario de poder utilizar el TransportDcat-AP, pero con un validador porque es muy difícil controlar y gestionar toda su sintaxis.

TransportDCAT-AP

Introducción

Es un perfil propuesto por el proyecto CEF-OASIS, con el objetivo de la presentación de metadatos en portales de datos abiertos del transporte público, dicho perfil está basado en el estándar propuesto por la unión europea, para la presentación de metadatos en portales abiertos DCAT-AP.

Se ha visto la importancia y la necesidad de proveer metadatos de forma simple y sencilla, para portales de datos abiertos en el sector de transporte público a nivel europeo, así sería posible a los agentes (pueden ser una empresa, organización, usuarios, desarrolladores..... etc.) descubrir dichos datos de transporte dentro de Europa.

Debido a lo dicho se ha trabajado en ofrecer la presentación de metadatos en sector de transporte público teniendo en cuenta sus principales características.

Descripción

Hablando del mismo perfil, y de su creación, su principal funcionamiento es, facilitar la interoperabilidad semántica entre datos de catálogos, publicados en la web, permitiendo la posibilidad de interacción a nivel de datos entre distintos sistemas, y garantizar el intercambio de información, hablando de catálogo hay que entender el concepto de metadatos, datasets y catálogos.

Metadatos son unos datos que describen otros datos, simplificando, metadatos describen informaciones dentro de unos datasets que son una colección de datos, y catalogs son una colección de Datasets, y cada uno tiene clases obligatorias (Mandatory), recomendadas (Recommended) y opcionales y dentro de cada clase existen también propiedades obligatorias y opcionales.

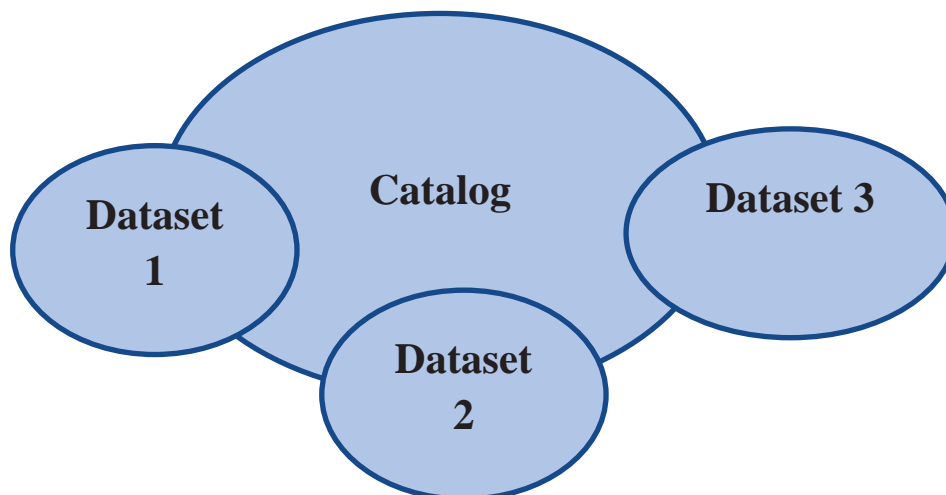


Figura 4: Catalog & Dataset

Esta figura refleja un catalog podría tener varios datasets.



Tal como se ha dicho anteriormente TransportDCAT-AP es un perfil de DCAT-AP , basado en el vocabulario DCAT para portales de datos abiertos en Europa , y DCAT es uno de los estándares de metadatos más usados.

TransportDCAT-AP ha tenido algunos cambios relevantes que se describirán a continuación basándose en los datos geoespaciales que son esenciales para presentar metadatos relacionados con el sector de transporte:

- La Relación (dcat:Catalog, dct:spatial, dct:Location) se ha cambiado de opcional a obligatoria .
- La Relación (dcat:Dataset, dct:spatial, dct:Location) se ha cambiado de opcional a obligatoria .
- La Relación (dcat:Dataset, dct:keyWord, rdfs:Literal) se ha cambiado de opcional a obligatoria .

El rango de la relación (dcat:Dataset, dct:keyWord, rdfs:Literal) se cambia de un string a una lista de palabras claves (keywords) controladas siguiendo el siguiente formato :

- Dct:keyword A , A es el tipo de transporte , los tipos de transporte permitidos son una extensión de la propiedad **route_type** in GTFD. Tram , Metro , Bus Ferry y otros . se puede repetir varias veces si el dataset incluye varios tipos de transporte
- DCT:Keyword B , B es el tipo de dataset que representa A , los tipos validos de Datasets siguen el **MODELO GTFS** y son : Calendar , Fare , Frequencies , Roues , Shapes , STOPS , transfers , trips o otros . Esta propiedad puede repetirse si el dataset incluye múltiples configuraciones.
- DCT:Keyword C , C es una característica INSPIRE para direcciones , el AdminUnitName , donde el nivel debería ser especificado conforme con la documentación oficial del framework (AdminUniLevel3(provincia): Madrid , AdminUniLevel4(Municipio): Madrid).

Se han implementado estos cambios porque en el sector de transporte es fundamental la información geográfica. Se ha considerado que dichos cambios deberían ser Obligatorias, para proveer esta información al generar metadatos para la colección de datos de los catálogos en los portales de datos abiertos.

Con tantos Datos que se van a recolectar en el sector de transporte, hoy en día en Madrid hay muchos medios de transporte que se desplazan por fuera de Madrid y por el centro de la ciudad. Con la cantidad de personas que utilizan el transporte público para desplazarse, había que innovar algún perfil que sea adecuado, al mismo tiempo que cumple con los requisitos basados en la necesidad cotidiana del ciudadano, y de las empresas que van a necesitar dichos datos.

Debido a estas razones mencionadas, se ha visto la obligación y la necesidad de intervenir y crear el TransportDCAT-AP, y un validador que controle y gestiona su sintaxis.

Clases y propiedades

En esta sección se facilita una guía, para generar metadatos cumpliendo los cambios propuestos en el TransportDCAT-AP. Esta guía está dividida en varias tablas, cada tabla contiene propiedades obligatorias o recomendadas u opcionales.

Dcat-Catalog

dcat:Catalog - Mandatory class		
Property	Type	Example
dcat:dataset	Mandatory	dcat:dataset <#dataset-01>
dct:description	Mandatory	dct:description "Catálogo de transporte urbano del CRTM"
dct:publisher	Mandatory	dct:publisher <#agent-01>
dct:spatial	Mandatory	dct:spatial <http://sws.geonames.org/3117735/>
dct:title	Mandatory	dct:title "Autobuses urbanos de la EMT"
foaf:homepage	Recommended	foaf:homepage <http://datos.crtm.es/>
dct:language	Recommended	dct:language <http://id.loc.gov/vocabulary/iso639-1/es.html>
dct:license	Recommended	dct:license <http://www.crtm.es/licencia-de-uso>
dct:issued	Recommended	dct:issued "2011-12-05"^^xsd:date
dcat:themeTaxonomy	Recommended	dcat:themeTaxonomy <http://datos.crtm.es/kos/vocabmetadata>
dct:modified	Recommended	dct:modified "2011-12-05"^^xsd:date
dct:hasPart	Optional	dct:hasPartOf <#catalog-02>
dct:isPartOf	Optional	dct:hasPartOf <#catalog-07>
dcat:record	Optional	dcat:record <#catalogRecord-01>
dct:rights	Optional	dct:rights "debe ser publico "

Dcat-Dataset

dcat:Dataset - Mandatory class		
Property	Type	Example
dct:description	Mandatory	dct:description "Dataset de las paradas de autobuses interurbanos"
dct:title	Mandatory	dct:title "Paradas autobuses interurbanos del CRTM"
dct:spatial	Mandatory	dct:spatial <http://sws.geonames.org/3117735/>
dcat:keyWord	Mandatory	dcat:keyword "Bus"; dcat:keyword "Stops" dcat:keyword "AdminUnitLevel3 (Province): Madrid, Spain";
dcat:contactPoint	Recommended	dcat:contactPoint "<mailto: contacto@crtm.es>"
dcat:distribution	Recommended	dcat:distribution <#distribution-01>
dct:publisher	Recommended	dct:publisher <#agent-01>
dcat:theme	Recommended	dcat:theme <http://datos.crtm.es/kos/vocabmetadata/interurbanbus>
dct:accessRights	Optional	dct:accessRights "public"
dct:conformsTo	Optional	dct:conformsTo "GTFS"
foaf:page	Optional	foaf:page <http://data.crtm.com/datasets/dataset01>
dct:accrualPeriodicity	Optional	dct:accrualPeriodicity <http://purl.org/cld/freq/monthly>
dct:hasVersion	Optional	dct:hasVersion <#dataset-0101>
dct:identifier	Optional	dct:identifier "M8Estaciones"
dct:isVersionOf	Optional	dct:isVersionOf <#dataset-0102>
dcat:landingPage	Optional	dcat:landingPage <https://services5.arcgis.com/UxADft6QPcvFyDU1/arcgis/rest/services/M8_Red/FeatureServer/0>
dct:language	Optional	dct:language <http://id.loc.gov/vocabulary/iso639-1/es.html>
adms:identifier	Optional	adms:identifier"InterUrbanBusesM8"
dct:provenance	Optional	dct:provenance "Este dataset pertenece al conjunto de todos los datos abiertos sobre autobuses interurbanos mostrados en datos.crtm.es"

dct:relation	Optional	dct:relation <http://data-crtm.opendata.arcgis.com/datasets?q=M8+red&sort_by=relevance>
dct:issued	Optional	dct:issued "2015-01-01"^^xsd:date
adms:sample	Optional	adms:sample <#distribution-01>
dct:source	Optional	dct:source <#dataset-05>
dct:temporal	Optional	dct:temporal <#periodOfTime-01>
dct:type	Optional	dct:type <http://inspire.ec.europa.eu/metadata-codelist/ResourceType/series>
dct:modified	Optional	dct:modified "2015-12-05"^^xsd:date
owl:versionInfo	Optional	owl:versionInfo "1.0"
adms:versionNotes	Optional	adms:versionNotes "final version for 2016"

Foaf-Agent

foaf:Agent - Mandatory class		
Property	Type	Example
foaf:name	Mandatory	foaf:name "Consorcio Regional de Transporte de Madrid"
dct:type	Recommended	dct:type <http://eurovoc.europa.eu/3060>

Dcat-Distribution

dcat:Distribution - Recommended class		
Property	Type	Example
dcat:accessURL	Mandatory	dcat:accessURL <http://data-crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d80f7858_0.csv>
dct:description	Recommended	dct:description "Distribucion del dataset de las paradas autobuses urbanos de la comuidad de madrid en CSV"
dct:format	Recommended	dct:format "CSV"
dct:license	Recommended	dct:license <http://www.crtm.es/licencia-de-uso>
dcat:byteSize	Optional	dcat:byteSize 1.0
spdx:checksum	Optional	spdx:checksum <#checksum-01>
foaf:page	Optional	foaf:page <http://data-crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d80f7858_0>
dcat:downloadURL	Optional	dcat:downloadURL <http://crtm.maps.arcgis.com/home/item.html?id=885399f83408473c8d815e40c5e702b7>
dct:language	Optional	dct:language <http://id.loc.gov/vocabulary/iso639-1/es.html>
dct:conformsTo	Optional	dct:conformsTo "CSV"
dcat:mediaType	Optional	dcat:mediaType "text/csv"
dct:issued	Optional	dct:issued "2015-12-05"^^xsd:date
dct:rights	Optional	dct:rights "public"
adms:status	Optional	adms:status "updated"
dct:title	Optional	dct:title "Paradas Autobuses Urbanos CRTM en CSV"
dct:modified	Optional	dct:modified "2015-12-05"^^xsd:date

Dcat:CatalogRecord – Recommended Class

dcat:CatalogRecord - Recommended class		
Property	Type	Example
foaf:primaryTopic	Mandatory	foaf:primaryTopic <#dataset-01>
dct:modified	Mandatory	dct:modified "2015-12-05"^^xsd:date
dct:conformsTo	Recommended	dct:conformsTo "TransportDCAT-AP"
adms:status	Recommended	adms:status "updated"
dct:issued	Recommended	dct:issued "2015-12-05"^^xsd:date
dct:description	Optional	dct:description "Incorporación del dataset de paradas de autobuses interurbanos del CRTM al catalogo de buses interurbanos"
dct:language	Optional	dct:language <http://id.loc.gov/vocabulary/iso639-1/es.html>
dct:source	Optional	dct:source <#catalogRecord-011>
dct:title	Optional	dct:title "Paradas autobuses interurbanos - Incorporación"

Spdx:Checksum

spdx:Checksum - Optional class		
Property	Type	Example
spdx:algorithm	Mandatory	spdx:algorithm <http://spdx.org/rdf/terms#checksumAlgorithm_sha256>
spdx:checksumValue	Mandatory	spdx:checksumValue "11542312"

Dct:PeriodOfTime

dct:PeriodOfTime - Optional class		
Property	Type	Example
schema:startDate	Optional	schema:startDate "2011-05-12"^^xsd:date
schema:endDate	Optional	schema:startDate "2011-05-12"^^xsd:date

OBJETIVOS

El principal objetivo es ofrecer un servicio que permita validar perfiles TransportDCAT-AP a cualquier usuario que sea una organización o cualquier entidad.

Se ha hecho una lista de objetivos que se deben cumplir, consisten en:

- Hacer un análisis de los temas importantes para el mejor desarrollo de la aplicación.
- Análisis del diseño de la aplicación, y que componentes tendrá.
- Análisis de las tecnologías que se van a utilizar.
- Implementación de la aplicación.
- Pruebas de la aplicación y detectar los fallos.
- Mejoras en la aplicación y corrección de errores
- Asegurar el buen funcionamiento de la aplicación.

DESARROLLO

En este apartado se describen todas las herramientas librerías y el entorno de desarrollo que se han utilizado para desarrollar la aplicación TransportDCA-AP.

Tecnologías de Desarrollo

HTML5

Es la última versión de HTML, es un lenguaje markup (de hecho, las siglas de HTML significan Hyper Text Markup Language) usado para estructurar y presentar el contenido para la web. Es uno de los aspectos fundamentales para el funcionamiento de los sitios web.

CSS3:

Cascading Style Sheets son las siglas de las hojas de estilo externas en las que se almacenan los estilos introducidos en HTML4. CSS3 es su última versión, y está aún pendiente de ser aprobado como recomendación oficial del W3C, sin embargo, al igual que HTML5, muchos navegadores ya implementan funcionalidades de esta versión.

Mediante el uso de hojas de estilo externas podemos ahorrar mucho trabajo, puesto que editando un único archivo nos permite cambiar el estilo de un documento HTML. De otra manera, habría que aplicar a mano los estilos, en cada zona que deseáramos, ofuscando el código HTML y dejando un documento poco mantenible.

Typescript:

TypeScript es un lenguaje de programación de código abierto desarrollado por Microsoft, el cual cuenta con herramientas de programación orientada a objetos, muy favorable si se tienen proyectos grandes. Anders Hejlsberg, arquitecto principal del desarrollo del lenguaje de programación C#, es el principal participante en el desarrollo de este lenguaje.

TypeScript convierte su código en Javascript común. Es llamado también Superset de Javascript, lo que significa que si el navegador está basado en Javascript, este nunca llegará a saber que el código original fue realizado con TypeScript y ejecutará el Javascript como lenguaje original.

Superset:

Se trata de un lenguaje escrito sobre otro lenguaje. En este caso Typescript, es un lenguaje basado en el original, ofreciendo grandes beneficios como el descrito anteriormente, aunque existen otros beneficios. Por ejemplo, mientras otros superset de JavaScript nos alejan del código original, Typescript, por el contrario, es muy similar a Javascript y a C# gracias a que su creador posee conocimientos de ambos lenguajes.

Actualmente Angular 4 y 5, uno de los frameworks más famosos de JavaScript, está siendo desarrollado en TypeScript.

Angular

Hoy en día, se construyen aplicaciones web impresionantes, sin embargo, la complejidad de su construcción no lo es menos. En Google, crearon Angular después de haber sufrido con la creación de aplicaciones web colosales como Gmail o Google Maps.

Angular, no es sino un framework Modelo Vista Controlador (MVC) que proporciona una capa de abstracción al programador sobre el Document Object Model (DOM), y además permite organizar el código del Front-End de una forma estructurada, facilitando la mantenibilidad, y logrando que las decisiones de diseño de la lógica de la aplicación sean más naturales.

Angular4

Es una evolución e Angular 2 donde se han introducido cambios para conseguir grandes beneficios como, por ejemplo, que las aplicaciones en Angular sean mucho más rápidas y pequeñas, con ahorros de alrededor del 60% del código generado en la mayoría de los casos (cientos de kilobytes en muchos casos). También se ha adoptado el proyecto de Angular Universal, que permite renderizar las páginas desde el servidor, y se ha agregado compatibilidad con nuevas versiones de TypeScript. Otra novedad muy significativa es que ya se pueden generar módulos de ES2016, lo cual es especialmente interesante ahora que todos los navegadores están a las puertas de soportarlos de forma nativa (Safari ya lo soporta desde la v10.1, Edge, Firefox y Chrome lo soportan mediante cambios de configuración manual).

Git

Gracias a esta herramienta se ha gestionado el control de versiones, integrando de forma muy frecuente el código desarrollado en nuestra rama de funcionalidad.

NPM (Node Package Manager)

Este programa nos permite mediante el mantenimiento de un fichero package.json, gestionar un proyecto de Node.js (Nombre del proyecto, estructura del árbol de directorios del proyecto, sus librerías, repositorios asociados de control de versiones, entre otras funciones).

Tecnologías de metadatos

XML

XML (eXtensive Markup Language), permite la codificación para la distribución de documentos complejos por Internet.

RDF (Resource Description Framework)

ofrece una superestructura para la descripción de recursos web de forma sencilla. Es un lenguaje capaz de representar metadatos, es decir, datos sobre datos, que ayudan a la identificación, clasificación o recuperación de documentos en línea. Cuenta con un

conjunto claro de reglas, que permite la codificación, intercambio y procesamiento automático de los metadatos normalizados así como la construcción de una estructura semántica no ambigua, gracias a la utilización de identificadores permanentes de recursos URI. Todo ello lo convierte en el estándar más popular y extendido en la comunidad web hoy en día. El elemento de construcción básica de RDF es la sentencia o “tripleto”, que consta de:

- **Sujeto.** Es el recurso, es decir, todo aquello que puede ser descrito.
- **Predicado.** Introduce la propiedad o atributo que va a detallarse sobre el recurso.
- **Objeto.** Es el valor de dicha propiedad.

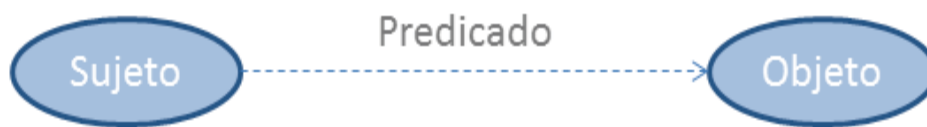


Figura 5: Representación Gráfica de un tripleto

Librería utilizada

Una de las librerías fundamentales que he utilizado para hacer un análisis de RDF, es RDFLIB.js [9].

rdflib.js

Es una Librería de Javascript para buscar, analizar y presentar datos RDF en (Turtle, RDF/XML, RDFa, etc).

el sistema usa objetos de javascripts para símbolos RDF , generándolos como predicados mediante espacio de nombres, convirtiéndolos en prefijos .

- var RDF = Namespace ("http://www.w3.org/1999/02/22-rdf-syntax-ns#").
- var RDFS = Namespace ("http://www.w3.org/2000/01/rdf-schema#").
- var FOAF = Namespace ("http://xmlns.com/foaf/0.1/").
- var XSD = Namespace ("http://www.w3.org/2001/XMLSchema#").

Aplicación web

Estudio Funcionalidad

En la aplicación, el objetivo principal es validar los perfiles de metadatos basados en el TransportDcat-AP.

El usuario que acceda a la aplicación podrá:

- Validar los metadatos insertándolos en un área de texto.
- Validar los metadatos insertándolos importando un fichero con extensión RDF/XML.
- Ver los errores de sintaxis de TransportDcat-AP.

Fases del Proyecto

Para tener una organización de las tareas que tenemos que realizar durante el proyecto dividimos el desarrollo en fases:

1. Análisis y definición de requisitos.
2. Análisis del diseño de la herramienta.
3. Concretar los componentes de la aplicación.
4. Analizar los errores que tienen que salir a la hora validar un perfil con errores.
5. Analizar los warnings también.
6. Pruebas.

Diseño de la Aplicación

A la hora de diseñar la aplicación se ha decidido que componentes son necesarios.

- Componente1 llamado Input es donde el usuario hará una validación del perfil TransportDcat-AP insertándolos en un texto de Área.
- Componente2 llamado uploadFile, es donde se validan los perfiles TransprtDcat-AP, importando un archivo desde el local del usuario.

Seguido a eso se ha decidido donde aparecerán los errores de la validación.

Implementación y Visualización Final

Parte de implementación código

Para la implementación de la aplicación, se he hecho un análisis de como empezar el desarrollo, tal como se ha dicho que la aplicación ha sido implementada con Typescript con el framework Angular4. Primero se ha visto la necesidad de hacer un parser de RDF/XML utilizando la librería anteriormente mencionada **rdflib.js**.

Esta librería tiene una función `graph()` que almacena los datos introducidos en una variable `store` y se guardan en forma de objetos JSON que más adelante hay recórrerlo para pasear siguiendo las normas contadas en el `TransportDCAT-AP`.

```
var store = $rdf.graph()
```

Una vez tenemos dicho Dato como objeto, hay que guardar los datos en buffer, porque, se va a utilizar como un primer parámetro de la función `parse()` de la librería `RDFLIB.JS`.

```
try {  
  $rdf.parse(body, store, uri, mimeType)  
} catch (err) {  
  console.log(err)  
}
```

Como se ha visto la función `parse` tiene 4 parámetros:

1. **body** - Buffer RDF que hay que parsear.
2. **store** - el objeto `graph/store` donde RDF debe ser analizado.
3. **uri** - el URI del recurso ('`http://IoFTriples.com`')
4. **mimeType** - el mime type corresponde al tipo de data que vamos a parsear en nuestro Caso '`application/rdf+xml`'.

Y así se parsea con éxito los metadatos RDF introducidos.

Después se hace un parse del propio `TRANSPORTDCAT-AP` siguiendo los requisitos del perfil.

Se han implementado las siguientes clases:

- Función **Catalog** para analizar las propiedades obligatorias recomendadas y opcionales de `Catalog`.
- Función **DATASET** para analizar las propiedades obligatorias recomendadas y opcionales de `DATASET`.
- Función **Organization** para analizar las propiedades obligatorias recomendadas y opcionales de `Organization`.
- Función **RecordCatalog** para analizar las propiedades obligatorias recomendadas y opcionales de `Organization`.

Visualización Final

Portada

En la portada existen dos componentes Input Validator y Upload Validator

Componentes

Upload Validator : Esta imagen refleja como el usuario pueda validar el perfil de transportDcat-AP importando un fichero.

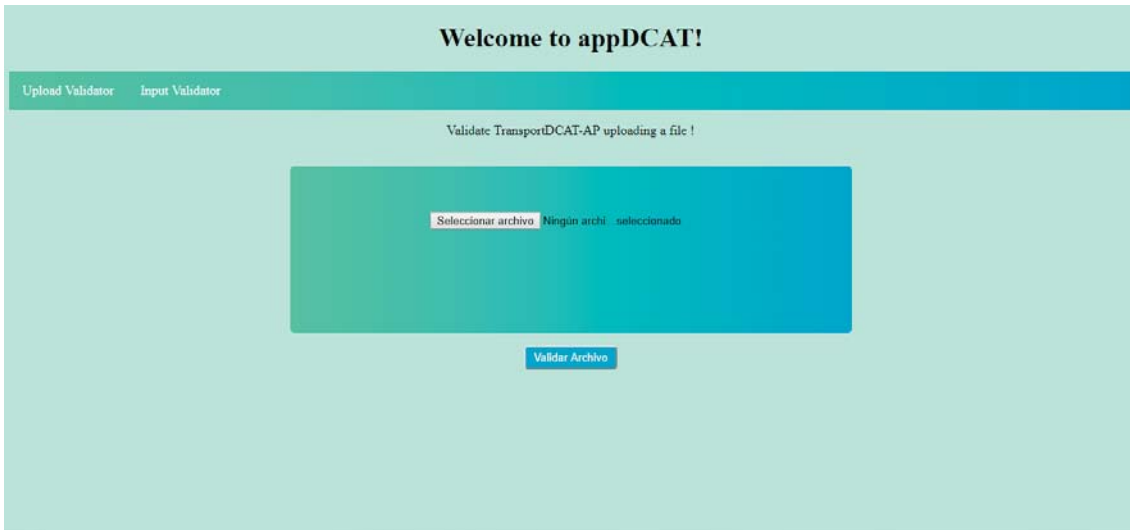


Figura 6: Componente Upload File Transport-DCAT-AP

Input Validator: esta parte refleja como el usuario pueda validar TransportDcat-AP desde un área de texto

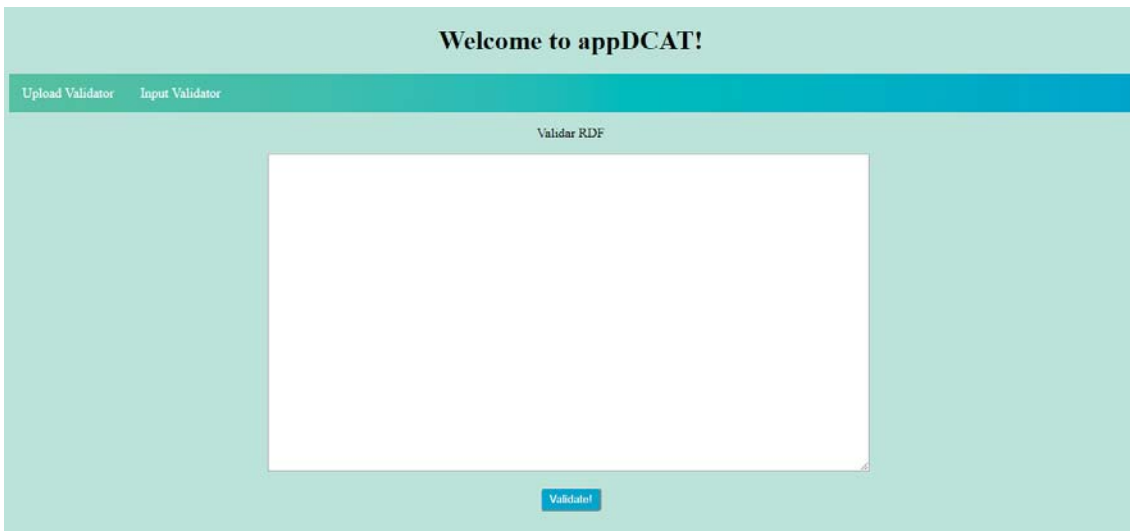


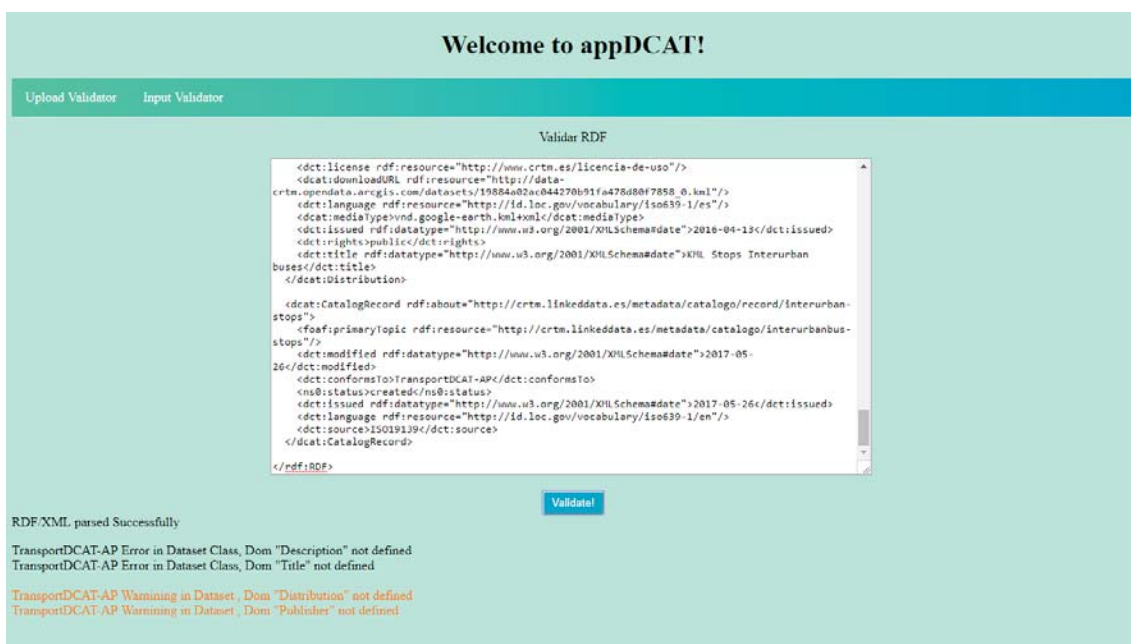
Figura 7: Componente Input TransportDcat-AP

Pruebas Finales y Correcciones

El parser comienza primero haciendo validación de los metadatos en RDF/XML, seguidamente se hace una validación del Perfil TransportDCAT-AP

Esta foto, demuestra algunos de los errores que puedan ser mostrados desde la aplicación si se insertan vocabularios erróneos. Los del Negro son errores que demuestran falta las propiedades `dct:description` y `dct:title` en la clase `Dataset`, las líneas de color naranja describen la falta de propiedades recomendadas, `dcat:distribution` y `dct:publisher` en la misma Clase `Dataset`

Pero antes de comenzar con el parser del TransportDCAT-AP, se hace un parse RDF/XML de los metadatos insertados.



The screenshot shows the 'Welcome to appDCAT!' interface. It features a 'Validator RDF' section with a text area containing RDF/XML code. Below the text area is a 'Validar' button. The output shows 'RDF/XML parsed Successfully' followed by several error and warning messages in red and orange text.

```
<dct:license rdf:resource="http://www.crtm.es/licencia-de-uso"/>
<dct:downloadURL rdf:resource="http://data-
crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d80f7858_0.kml"/>
<dct:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/es"/>
<dct:mediaType>nd.google-earth.kml+xml</dct:mediaType>
<dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2018-04-13</dct:issued>
<dct:rights>public</dct:rights>
<dct:title rdf:datatype="http://www.w3.org/2001/XMLSchema#date">XHL Stops Interurban
buses</dct:title>
</dct:Distribution>

<dcat:CatalogRecord rdf:about="http://crtm.linkeddata.es/metadata/catalogo/record/interurban-
stops">
  <foaf:primaryTopic rdf:resource="http://crtm.linkeddata.es/metadata/catalogo/interurbanbus-
stops"/>
  <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-
26</dct:modified>
  <dct:conformsTo>TransportDCAT-AP</dct:conformsTo>
  <ns0:status>created</ns0:status>
  <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-26</dct:issued>
  <dct:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/en"/>
  <dct:source>ISO19139</dct:source>
</dcat:CatalogRecord>
</rdf:RDF>
```

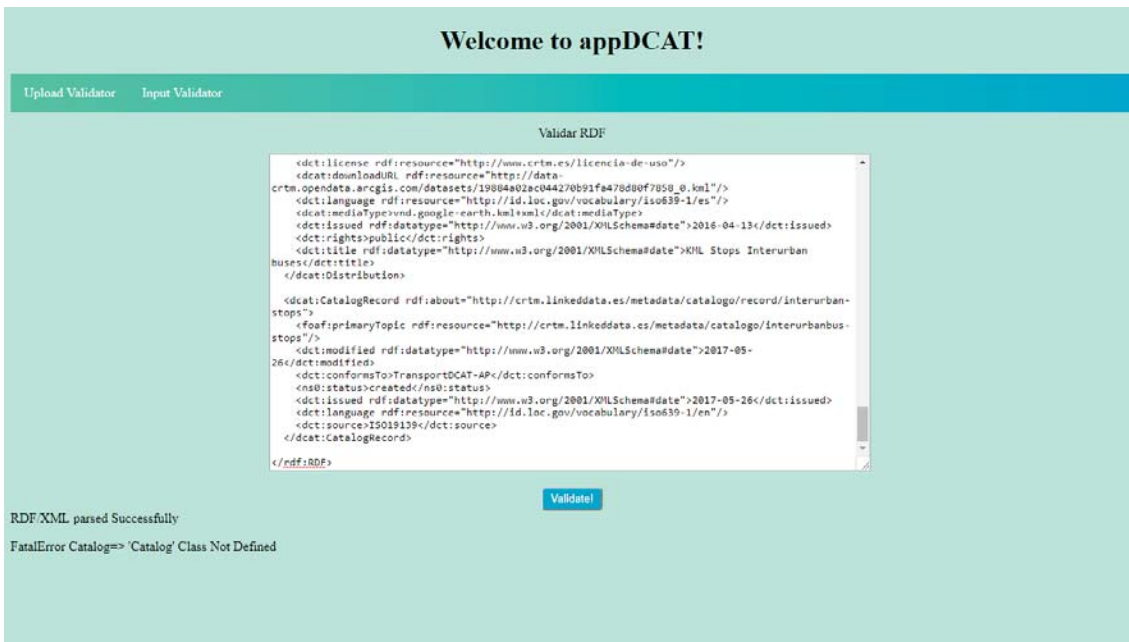
RDF/XML parsed Successfully

TransportDCAT-AP Error in Dataset Class, Dom "Description" not defined
TransportDCAT-AP Error in Dataset Class, Dom "Title" not defined

TransportDCAT-AP Warning in Dataset, Dom "Distribution" not defined
TransportDCAT-AP Warning in Dataset, Dom "Publisher" not defined

Figura 8: Error parser TransportDCAT-AP

Otro Ejemplo es cuando se insertan metadatos sin definir la clase Catalog



Welcome to appDCAT!

Upload Validator Input Validator

Validar RDF

```
<dc:license rdf:resource="http://www.crtm.es/licencia-de-uso"/>
<dcat:downloadURL rdf:resource="http://data-
crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d00f7858_0_kml"/>
<dc:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/es"/>
<dc:mediaType>vnd.google-earth.kml+xml</dc:mediaType>
<dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2016-04-13</dc:issued>
<dc:rights>public</dc:rights>
<dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#date">XHL Stops Interurban
buses</dc:title>
</dc:Distribution>

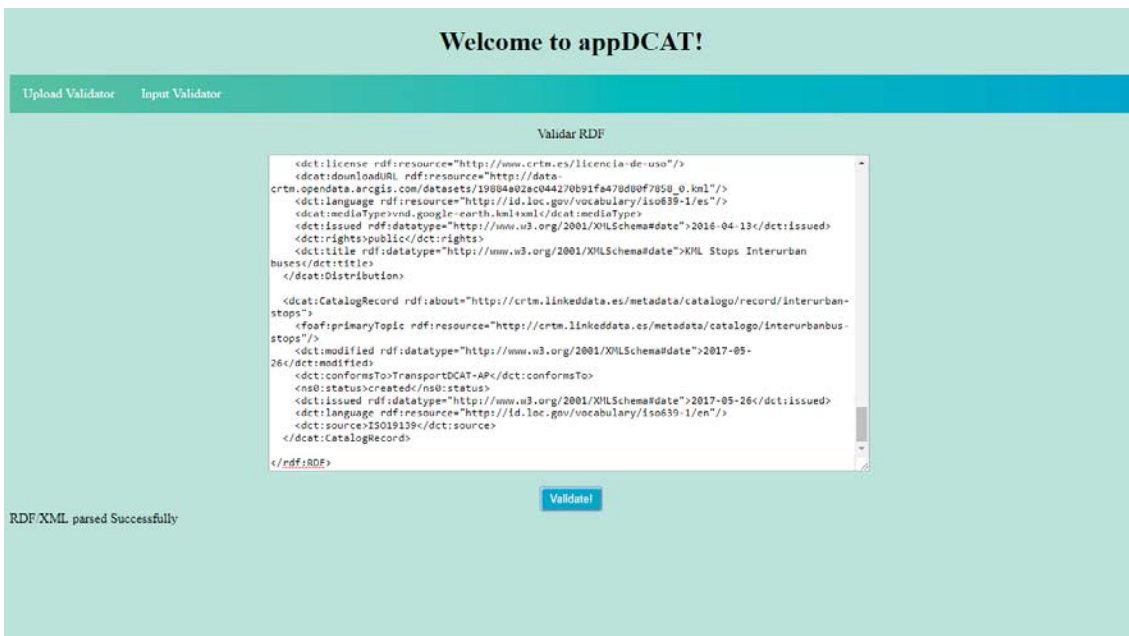
<dc:CatalogRecord rdf:about="http://crtm.linkeddata.es/metadata/catalogo/record/interurban-
stops">
  <foaf:primaryTopic rdf:resource="http://crtm.linkeddata.es/metadata/catalogo/interurbanbus-
stops"/>
  <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-
26</dc:modified>
  <dc:conformsTo>TransportDCAT-AP</dc:conformsTo>
  <ns0:status>created</ns0:status>
  <dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-26</dc:issued>
  <dc:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/en"/>
  <dc:source>I5019139</dc:source>
</dc:CatalogRecord>
</rdf:RDF>
```

Validar

RDF/XML parsed Successfully
FatalError Catalog=> 'Catalog' Class Not Defined

Figura 9: Falta de Clase Obligatoria dcat:Catalog ,

Si en el caso de que la sintaxis de RDF/XML y de TransportDCAT-AP no tenga ningún error y que se haya validado bien se muestra un mensaje “RDF/XML Parsed Successfully” sin ningún otro mensaje de error.



Welcome to appDCAT!

Upload Validator Input Validator

Validar RDF

```
<dc:license rdf:resource="http://www.crtm.es/licencia-de-uso"/>
<dcat:downloadURL rdf:resource="http://data-
crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d00f7858_0_kml"/>
<dc:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/es"/>
<dc:mediaType>vnd.google-earth.kml+xml</dc:mediaType>
<dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2016-04-13</dc:issued>
<dc:rights>public</dc:rights>
<dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#date">XHL Stops Interurban
buses</dc:title>
</dc:Distribution>

<dc:CatalogRecord rdf:about="http://crtm.linkeddata.es/metadata/catalogo/record/interurban-
stops">
  <foaf:primaryTopic rdf:resource="http://crtm.linkeddata.es/metadata/catalogo/interurbanbus-
stops"/>
  <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-
26</dc:modified>
  <dc:conformsTo>TransportDCAT-AP</dc:conformsTo>
  <ns0:status>created</ns0:status>
  <dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-26</dc:issued>
  <dc:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/en"/>
  <dc:source>I5019139</dc:source>
</dc:CatalogRecord>
</rdf:RDF>
```

Validar

RDF/XML parsed Successfully

Figura 10: Parseado con Exito TransportDCAT-AP

Otro ejemplo es cuando hay algún error en RDF/XML de los metadatos introducidos

Welcome to appDCAT!

Upload Validator Input Validator

Validar RDF

```
<dct:license rdf:resource="http://www.crtm.es/licencia-de-uso"/>
<dcats:downloadURL rdf:resource="http://data-
crtm.opendata.arcgis.com/datasets/19884402ac044270b91fa478d00f7858_0.kml"/>
<dct:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/es"/>
<dcats:mediaType>vnd.google-earth.kml+xml</dcats:mediaType>
<dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2016-04-13</dct:issued>
<dct:rights:public/>
<dct:title rdf:datatype="http://www.w3.org/2001/XMLSchema#date">KML Stops Interurban
buses</dct:title>
</dcats:Distribution>

<dcats:CatalogRecord rdf:about="http://crtm.linkeddata.es/metadatos/catalogo/record/interurban-
stops">
  <foaf:primaryTopic rdf:resource="http://crtm.linkeddata.es/metadatos/catalogo/interurbanbus-
stops"/>
  <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-
26</dct:modified>
  <dct:conformsTo>TransportDCAT-AP</dct:conformsTo>
  <ns0:status>created</ns0:status>
  <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-05-26</dct:issued>
  <dct:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/en"/>
  <dct:source>ISO19119</dct:source>
</dcats:CatalogRecord>
</rdf:RDF>
```

Error: end tag name: rdf:RDF is not complete:dct:rights
@[line:156,col:24]

Figura 11: Error rdf/xml



CONCLUSIONES

Este trabajo se ha implementado con el objetivo de crear una aplicación web que da un servicio de analizar sintácticamente perfiles basados en TransportDCAT-AP para el proyecto CEF-OASIS donde participa el grupo de ingeniería de ontología¹³.

Para la realización de 'TransportDCAT-AP Validator', se han abordado varios fundamentos y tecnologías relacionadas con los metadatos y la programación de los cuales se tenía un conocimiento muy básico o incluso nulo. Cursar la asignatura 'PROCESADORES DE LENGUAJES' y otras asignaturas de programación, han ayudado bastante en formar una idea básica para arrancar y avanzar en el proyecto, y realizar un plan de trabajo adecuado.

En cuanto a desarrollo personal, el hecho de empezar una aplicación desde cero utilizando tantas tecnologías nuevas y conceptos nuevos en el ámbito de web semántica, y metadatos, he podido aumentar mis conocimientos y mejorar mis habilidades de programación, análisis del trabajo que se ha dividido en partes, para que al final se haya podido llegar a los objetivos de esta herramienta.


Gracias al desarrollo de esta aplicación se me han ocurrido muchas ideas para seguir trabajando en este ámbito, pensar en reutilizar las tecnologías aprendidas para desarrollar otra aplicación que tengo pendiente implementarla.

¹³ <http://www.oeg-upm.net/>

BIBLIOGRAFÍA

- [1] [https://en.wikipedia.org/wiki/Microdata_\(HTML\)](https://en.wikipedia.org/wiki/Microdata_(HTML)).
- [2] W. W. Consortium, «<http://www.w3.org/>,» W3C. [En línea].
- [3] R. w. org, «<https://www.w3.org/RDF/>,» [En línea].
- [4] t. Gruber, What is ontology? <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
- [5] Wikipedia, <datos Enlazados> https://es.wikipedia.org/wiki/Datos_enlazados, Wikipedia .
- [6] T. Berners-Lee, «<https://www.w3.org/People/Berners-Lee/>,» El inventor de la web. [En línea].
- [7] W3C, «<https://www.w3.org/TR/vocab-dcat/>,» Documento Oficial DCAT. [En línea].
- [8] D.-A. Validator, «<https://joinup.ec.europa.eu/release/dcat-ap-validation-service-110>,» [En línea].
- [9] Github, «<https://github.com/linkedata/rdfliib.js/>,» Tutorial rdfliib.js. [En línea].

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=Facultad de Informatica - UPM, C=ES
Fecha/Hora	Thu Feb 01 15:10:12 CET 2018
Emisor del Certificado	EMAILADDRESS=camanager@fi.upm.es, CN=CA Facultad de Informatica, O=Facultad de Informatica - UPM, C=ES
Numero de Serie	630
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)