

# Datos Abiertos en Panamá

Boris Gómez<sup>1</sup> , Maritza Morales<sup>2</sup> , Nicholas Beliz<sup>3</sup> , Yazmina Villarreal<sup>2</sup> , Elida González<sup>2</sup> 

1,2,3 {Dirección de Innovación y Tecnología Educativa; Centro de Investigación, Desarrollo e Innovación en Tecnologías de la Información y las Comunicaciones; Facultad de Ingeniería de Sistemas Computacionales}, Universidad Tecnológica de Panamá

1-3 {boris.gomez, maritza.morales, nicholas.beliz, yazmina.villarreal, elida.gonzalez}@utp.ac.pa

DOI: 10.33412/pri.v12.1.2383



**Resumen:** *Hay una tendencia para que los gobiernos permitan el acceso público a sus datos y que estos sean oportunos y actualizados, de allí el concepto de “datos abiertos gubernamentales”. Sin embargo, es difícil dar seguimiento a la función pública cuando los datos están muy disgregados.*

*Para resolver esta dificultad, existe el concepto denominado Web Semántica, cuyo propósito es permitir que los programas informáticos asistan o reemplacen a las personas en tareas difíciles o inabordables para los humanos, tales como relacionar datos de múltiples fuentes.*

*Este artículo muestra los componentes básicos para implementar la Web Semántica, específicamente en el marco de los “Datos Abiertos Gubernamentales”, y se señalan, además, los pasos para alcanzar las cinco estrellas definidas en el esquema de Datos Abiertos propuesto por Tim Berners-Lee, utilizando un extracto de la división política de Panamá, como ejemplo.*

**Palabras claves:** Ontología, Web Semántica, grafo de conocimiento, RDF, tripleta, SPARQL, Datos Abiertos, Datos Enlazados.

**Title:** Linked Open Data in Panama

**Abstract:** There is a tendency for governments to allow public access to their data and that these be opportune and updated, hence the concept of Open Government Data. However, it is difficult to follow the public function when the data is very disaggregated.

To solve this difficulty, there is the concept called Semantic Web, whose purpose is to allow computer programs to assist or replace people in difficult or unapproachable tasks for humans, such as relating data from multiple sources.

This article shows the basic components for implementing the Semantic Web, specifically within the framework of “Government Open Data”, and also indicates the steps to achieve the five stars defined in the Open Data scheme proposed by Tim Berners-Lee, using an excerpt from the political division of Panama, as an example.

**Key words:** Ontology, Semantic Web, knowledge graph, RDF, triplet, SPARQL, Open Data, Linked Data.

Tipo de artículo: Estudio.

Fecha de recepción: 28 de octubre de 2019.

Fecha de aceptación: 5 de enero de 2021.

## 1. Introducción

Internet nació en la década de los 60 como una necesidad para que investigadores y científicos se comunicaran y compartieran información. Décadas después, en 1991, Tim Berners-Lee subió a internet la primera página web, dando origen a la World Wide Web (www o W3C).

Desde ese momento, la web cambió el modo en que las personas se comunicaban o realizaban negocios, al permitir el intercambio de información las 24 horas del día, los 365 días del año, con alcance mundial y a costos reducidos. Sin embargo, el incremento en el volumen de los datos y la heterogeneidad en los formatos y fuentes de información empezó a plantear un problema para la búsqueda de información.

Surgió, entonces, la necesidad de un nuevo paradigma para organizar el contenido en la web, la denominada Web Semántica, una evolución del entorno web cuyo objetivo era permitir que los datos almacenados en la web pudieran ser procesados por las máquinas de manera inteligente, facilitando a las personas la búsqueda, integración y análisis de la información disponible [1].

Pero la transición a la Web Semántica involucraba un alto costo si se consideraba el volumen de los datos, la disgregación de la información y la diversidad de formatos. Además, crear y publicar datos enlazados suponía un gran esfuerzo que podía resultar tedioso cuando se agregaban nuevos contenidos. Las estrategias más viables combinaban una pequeña parte de trabajo manual, sumado a la automatización del resto del proceso [2].

Tim Berners-Lee, el inventor de la World Wide Web, propuso entonces un esquema para clasificar los Datos Abiertos con la finalidad de que estos empezaran a adquirir características que permitieran convertirlos en datos plenamente conectados [3].

El concepto de Web Semántica está ligado a los conceptos de Datos Abiertos (Open Data) y de Datos Enlazados (Linked Data), en donde los Datos abiertos describen a aquellos que pueden ser usados, reutilizados y redistribuidos libremente por cualquier persona, mientras que los Datos Enlazados se refieren al conjunto de mejores prácticas para publicar datos estructurados en la web [4], [5], lo que los hace más útiles.

### 1.1 La tecnología detrás de los Datos Enlazados

Tim Berners-Lee definió cuatro principios básicos para publicar Datos Enlazados en la web:

- Utilizar Identificador de Recursos Uniforme (URI, por sus siglas en inglés) para nombrar a cada uno de los recursos en la web. Las URIs funcionan de forma similar a los nombres propios en nuestro lenguaje natural. Los indicadores son únicos a nivel global.
- Utilizar HTTP URIs (HTTP: protocolo de transferencia de hipertextos, por sus siglas en inglés) para localizar los recursos a través de internet.

- Proveer información útil utilizando estándares (Resource Description Framework o RDF) cuando se busca una URI.
- Incluir enlaces a otras URIs, que permitan el descubrimiento de más recursos en la web.

La tendencia actual para que los gobiernos permitan el acceso público a sus datos, sin restricciones y que estos sean oportunos y actualizados indujo a la génesis del término "Datos Abiertos Gubernamentales", entendiendo que el concepto implica "Datos Enlazados".

Para transformar datos públicos a Datos Abiertos es necesario realizar un proceso que tome en cuenta los principios establecidos en la Carta Internacional de Datos Abiertos:

- Abiertos por defecto
- Oportunos y exhaustivos
- Accesibles y utilizables
- Comparables e interoperables
- Para mejorar la gobernanza y la participación ciudadana
- Para el desarrollo incluyente y la innovación

En el año 2010 Tim Berners-Lee, sugirió un esquema de cinco estrellas para categorizar los Datos Abiertos, como se muestra en

la figura 1 [2]. En la tabla 1 se dan ejemplos que ilustran el esquema.

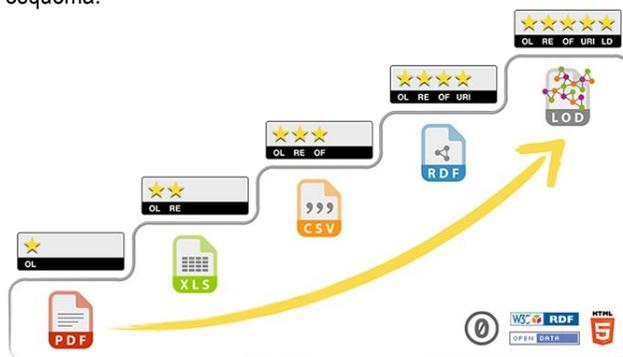


Figura 1. Las 5 estrellas de los Datos Enlazados Abiertos.

Llama la atención que con el solo hecho de colocar información de interés público en internet, bajo una licencia abierta, se logra obtener la primera estrella.

En Panamá, los esfuerzos se consolidaron con la formulación de la Ley 33 de 25 de abril de 2013, que crea la Autoridad Nacional de Transparencia y Acceso a la Información (ANTA).

Tabla 1. Esquema de las cinco estrellas de los Datos Abiertos.

Estrellas	Las instituciones públicas	Los ciudadanos
	No divulgan datos de interés público en internet.	Deben acercarse a las instituciones públicas para obtener información.
★	Publican datos en internet en formatos tipo TXT, DOC, ODT, PDF, PPT, JPG, por ejemplo, bajo una licencia abierta.	Deben leer cada uno de los documentos publicados para extraer la información que les interesa, compilarla y estructurarla, de ser necesario.
★★	Publican datos en internet en formatos tipo XLS.	Acceden a datos estructurados, fáciles de manejar, pero requieren de software propietario para acceder a la información.
★★★	Publican datos en internet en formatos tipo CSV, ODS.	Acceden a datos estructurados, fáciles de manejar y no requieren software de pago.
★★★★	Publican datos en internet en formatos legibles tanto por seres humanos como por máquinas, por ejemplo: RDF, JSON, RDF/XML.	- Utilizan aplicaciones inteligentes para extraer los datos que les interesa - Con una sola búsqueda acceden a datos de múltiples fuentes (los datos, de diferentes fuentes, se identifican de manera única, a través de las denominadas URIs) - No requieren software de pago.
★★★★★	Publican datos que están enlazados con información de otras organizaciones a nivel mundial, utilizando esquemas de datos consensuados internacionalmente.	- Utilizan aplicaciones inteligentes para extraer los datos que les interesa - Con una sola búsqueda acceden a datos de múltiples fuentes a nivel global - No requieren software de pago.

## 2. Datos Abiertos Gubernamentales en Panamá

La iniciativa de Datos Abiertos de la República de Panamá nace del proyecto "Fortalecimiento de las capacidades de los países en desarrollo para proporcionar Acceso a la Información para el Desarrollo Sostenible a través de los Datos Abiertos de Gobierno", implementado por la División de Administración Pública y Gestión del Desarrollo, del Departamento de Asuntos Económicos y Sociales de las Naciones Unidas, quien escogió a cuatro países del mundo para implementar dicho proyecto: Bangladesh, Nepal, Uruguay y Panamá [6].

Posteriormente, en el año 2018, la ANTA desarrolla la "Política Pública de Transparencia de Datos Abiertos de Gobierno", a través de la Resolución No. DS-3513-2018, en la

que establece la definición oficial de Datos Abiertos de Gobierno del país: "Los Datos Abiertos de Gobierno, son los datos recolectados y/o producidos por las instituciones públicas, puestos a disposición de los ciudadanos en formatos abiertos, para que puedan ser reutilizados y redistribuidos por cualquier persona, para cualquier propósito, incluyendo el re-uso comercial, libre de costo y sin restricciones, con el fin de posibilitar la lectura, el seguimiento y combinación con otras fuentes de información para generar nuevos servicios de valor".

Además, define al "Portal de Datos Abiertos" ([www.datosabiertos.gob.pa](http://www.datosabiertos.gob.pa)) como el inventario único nacional de los conjuntos de Datos Abiertos puestos a disposición de la

población por las instituciones públicas de la República de Panamá [7].

Se puede observar que la información en este Portal se presenta en tres formatos distintos: pdf, xls y csv, bajo licencia Creative Commons - CC Zero License, lo que significa que es de dominio público (se puede copiar, modificar, distribuir y reproducir) y cumple con tres estrellas (según el esquema de Tim Berners-Lee) al incluir archivos en formato csv.

### 3. Taller de Linked Data: Calidad de Datos

Con el objetivo de dar a conocer el concepto de Datos Abiertos y Enlazados a profesionales de informática del sector público y privado en nuestro país, en el año 2019 el Grupo de Investigación en Ontologías y Web Semántica (GIOWeS) del Centro de Investigación, Desarrollo e Innovación en Tecnologías de la Información y las Comunicaciones (CIDITIC) de la Universidad Tecnológica de Panamá (UTP), con el patrocinio de la Secretaría Nacional de Ciencia Tecnología e Innovación (SENACYT), organizó el "Taller de Linked Data: Calidad de Datos" y la Conferencia "Creación de Valor desde la Ciencia de las Redes y los Grafos de Conocimiento Semánticos". El taller se enmarcó en la Misión de la UTP "generar conocimiento apropiado para contribuir al desarrollo sostenible y responder a los requerimientos del entorno".

El expositor del evento fue el Dr. Nelson Piedra, docente e investigador en el Laboratorio de Ciencia de Datos y Sistemas Basados en el Conocimiento, en el Departamento de Ciencias de la Computación y Electrónica de la Universidad Técnica Particular de Loja (UTPL), Ecuador.

Además de estudiantes, docentes, personal administrativo e investigadores de la Universidad Tecnológica de Panamá, al taller asistieron profesionales del área informática de la Universidad de Panamá (UP), Universidad Católica Santa María La Antigua (USMA), Instituto Panameño de Derecho y Nuevas Tecnologías (IPANDETEC) y del Ministerio de Economía y Finanzas (MEF).

Los participantes tuvieron la oportunidad de conocer el concepto de Datos Abiertos y Enlazados, identificar las aplicaciones requeridas tanto en los servidores web como en las computadoras clientes y aplicar la metodología recomendada para alcanzar la quinta estrella del esquema de Tim Berners-Lee.

### 4. Grafos de conocimiento semántico

Algunos de los términos manejados en los niveles de la cuarta y quinta estrellas son: ontología, grafo de conocimiento, RDF, tripleta, SPARQL.

La ontología es uno de los modelos utilizados para describir conceptos del mundo real y su propósito es facilitar la compartición de información. Por otro lado, grafo de conocimiento es una forma de organizar el conocimiento para que una máquina pueda entenderlo y extraer información fácilmente, como se presenta a continuación.

Por otro lado, RDF es un formato para el intercambio de datos en la web, desarrollado por el W3C. Posee características que facilitan la integración de los datos aun cuando sus

estructuras difieran. La información se expresa como una lista de sentencias de la forma: sujeto-predicado-objeto, en donde el sujeto y el objeto se relacionan por medio del predicado. La sentencia sujeto-predicado-objeto se denomina tripleta [8] y se representa gráficamente como se observa en la figura 2.

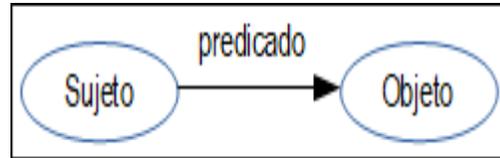


Figura 2. Tripleta.

Una manera estructurada de almacenar datos, ampliamente utilizada, son las tablas de bases de datos relacionales. La tabla 2 muestra un extracto de la división política de Panamá, mientras que las tablas 3, 4 y 5 se obtienen luego de normalizar dicha tabla.

Tabla 2. Extracto de la división política de Panamá.

País	Capital de país	Provincia	Distrito
Panamá	Panamá	Colón	Chagres
Panamá	Panamá	Colón	Portobelo
Panamá	Panamá	Coclé	Aguadulce
Panamá	Panamá	Coclé	Antón
Panamá	Panamá	Herrera	Parita

Tabla 3. país-capital-de-país.

País	Capital de país
Panamá	Panamá

Tabla 4. país-provincia.

País	Provincia
Panamá	Colón
Panamá	Coclé
Panamá	Herrera

Tabla 5. provincia-distrito.

Provincia	Distrito
Colón	Chagres
Colón	Portobelo
Coclé	Aguadulce
Coclé	Antón
Herrera	Parita

En contraste con el modelo de base de datos relacional, la Web Semántica utiliza tripletas. A continuación, se muestra un fragmento de la misma información convertida en tripletas:

- Panamá es-capital-de Panamá
- Colón es-provincia-de Panamá
- Coclé es-provincia-de Panamá
- Herrera es-provincia-de Panamá
- Chagres es-distrito-de Colón

La figura 3 muestra el grafo correspondiente.

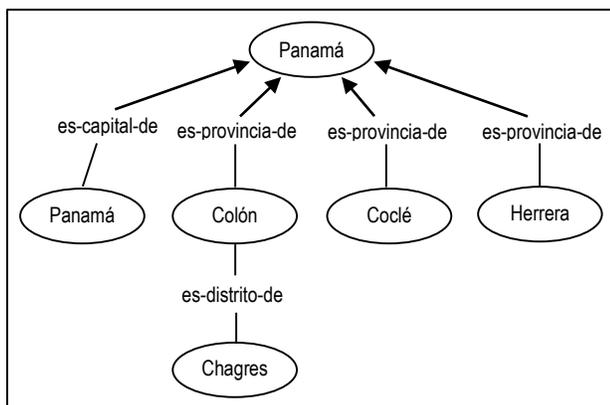


Figura 3. Ejemplo de grafo de la división política de Panamá.

SPARQL es el lenguaje utilizado para interrogar grafos en este modelo, similar a como funciona SQL al consultar datos en una base de datos.

En resumen, para evolucionar a la cuarta y quinta estrellas, los datos deben ser convertidos a tripletas (grafos).

#### 4.1 Requerimientos

Para agregar semántica a los datos publicados en la web se requiere de un equipo (servidor web) que se encargue de responder a las solicitudes de información de los usuarios (clientes), además de varias aplicaciones tanto en el servidor como en las computadoras personales de los diseñadores de la ontología. Por su parte, los usuarios acceden al servicio utilizando un navegador web en sus computadoras personales.

Para el Taller, el grupo de investigación GIOWeS utilizó exclusivamente software libre, de conformidad con los lineamientos del Centro CIDITIC.

##### 4.1.1 Requerimientos en el servidor

- **Debian:** el sistema operativo Linux utilizado en el proyecto fue Debian 9 Stretch [9].
- **OpenLink Virtuoso:** es un híbrido que combina la funcionalidad de un administrador de base de datos SQL y un servidor de aplicación web para el manejo de datos SQL, XML y RDF en un único sistema multiproceso. [10].
- **Apache Tomcat:** software desarrollado con Java (puede funcionar en cualquier sistema operativo, con su máquina virtual java correspondiente) que sirve como servidor web con soporte de servlets y JSPs [11].
- **Pubby:** aplicación web que permite, mediante consultas SPARQL, acceder a almacenes de tripletas. La figura 4 muestra cómo Pubby relaciona las aplicaciones para acceder a las tripletas y presentar la información requerida en el navegador web [12].

Los servlets son componentes de lenguajes de programación Java utilizados comúnmente para generar páginas web de forma dinámica, a partir de las peticiones que

envíe el navegador web [13]. Mientras que JSP (Java Server Pages) es una tecnología que permite crear contenido dinámico para aplicaciones web que se ejecuta del lado del servidor.

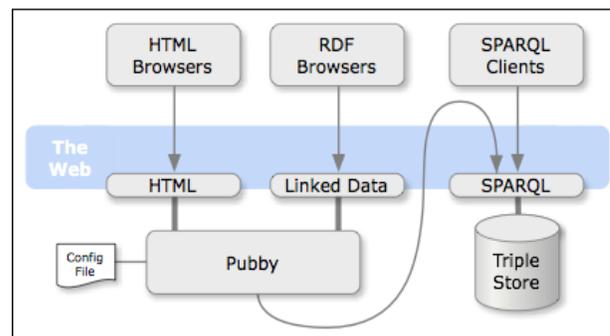


Figura 4. Esquema de Pubby.

#### 4.1.2 Requerimientos en las computadoras personales

**Protégé** [14]: es un editor de ontologías gratuito y de código abierto.

**Tabula** [15]: aplicación gratuita, utilizada para extraer datos de tablas contenidas en documentos PDF.

**EasyRdf** [16]: aplicación accesible desde la web para convertir archivos RDF de un formato a otro (por ejemplo, de N-Triples a RDF/XML).

**RDF Validator** [17]: aplicación accesible desde la web para validar modelos de datos en formato RDF/XML y desplegar las tripletas y grafos correspondientes.

**GraphDB** [18]: aplicación gratuita que permite vincular diversos datos, indexarlos para la búsqueda semántica y enriquecerlos mediante análisis de texto para construir grafos de conocimiento. Es una alternativa al OpenLink Virtuoso.

#### 4.2 Esquemas de datos consensuados internacionalmente

En el ejemplo de la división política de Panamá se utilizaron los términos "es-capital-de" y "es-provincia-de", con los que se pudo iniciar el diseño de la ontología del país. Al completar la ontología se logra la cuarta estrella, pero, para avanzar a la quinta estrella, se deben emplear identificadores consensuados internacionalmente, en lugar de los dos términos anteriores.

Para ello, la comunidad internacional ha desarrollado múltiples ontologías que pueden utilizarse libremente, en lugar de crear las propias. Estas pueden consultarse en el sitio web Linked Open Vocabularies (LOV) [19].

Un ejemplo de ontología es la que se utiliza para describir personas, denominada FOAF (*friend of a friend*, por sus siglas en inglés). Ésta es una ontología que describe a las personas, sus actividades y sus relaciones con otras personas y objetos [20].

#### 4.3 Información de Panamá en la Web Semántica

- **DBpedia** [21]: es un proyecto comunitario internacional para extraer información de Wikipedia y transformarla en tripletas RDF. Es interesante observar que este sitio web contiene información de Panamá, en los siguientes enlaces:

<http://dbpedia.org/page/Panamá> (página web accedida por los humanos, a través de navegadores web).  
<http://dbpedia.org/resource/Panamá> (página web accedida por las máquinas).

La figura 5 es un ejemplo de cómo se muestra la información general del país, en la web semántica.

Property	Value
dbo:PopulatedPlace/areaTotal	74177.3
dbo:PopulatedPlace/populationDensity	45.9
dbo:abstract	Panamá es un país ubicado en
dbo:areaTotal	74177300000.000000 (xsd:double)
dbo:capital	dbp:Panamá_City

Figura 5. Panamá en Dbpedia.

Es necesario aclarar que los navegadores web son accedidos por humanos, por lo que las dos páginas web, señaladas anteriormente, muestran igual contenido (“resource” redirecciona a “page” en los navegadores web). Además, la información que se observa en las páginas de la web semántica no es tan amigable para los seres humanos, como lo son las páginas web comunes, con textos e imágenes armoniosamente organizadas en las páginas. Más bien se observan conjuntos de datos (títulos y columnas con datos), teniendo en consideración que su objetivo es la utilización de comandos y aplicaciones web para extraer información de interés, de múltiples fuentes y a escala mundial.

DBpedia contiene, además, una página web para realizar consultas SPARQL a estos datos [22].

**DBpedia-LatAm** [23]: es la versión latinoamericana de DBpedia. Una de las páginas web sobre Panamá es: [http://es-la.dbpedia.org/page/resource/Categor%C3%ADa:Provincias\\_de\\_Panam%C3%A1](http://es-la.dbpedia.org/page/resource/Categor%C3%ADa:Provincias_de_Panam%C3%A1)

#### 4.4 Hackatones como estrategia

La compilación de los datos de un país para convertirlos en tripletes y la creación de aplicaciones que atiendan las necesidades de información de los ciudadanos es una labor que requiere de tiempo y esfuerzo. Un medio novedoso que se podría experimentar en Panamá, aplicado en otros países, es la organización de encuentros (*hackatones*) en universidades y organizaciones interesadas en el tema, en donde los organizadores establecen retos a superar por los participantes.

### 5. Conclusiones y recomendaciones

Los Datos Abiertos Gubernamentales surgen como consecuencia de la exigencia de los ciudadanos, a nivel mundial, para que las entidades públicas sean transparentes.

En Panamá se puede observar que el Portal de Datos Abiertos Gubernamentales contiene información en formato csv, entre otros formatos, bajo una licencia abierta, por lo que se ubica en el tercer nivel del esquema de las cinco estrellas propuesto por Tim Berners-Lee.

Es por ello que, en este artículo, utilizando un extracto de la división política del país, se muestran los pasos requeridos para adecuar este portal al último nivel de dicho esquema. Además, se señala la infraestructura (de hardware y de software) necesaria para su implantación.

Como trabajo futuro, es necesario definir una estrategia con miras a lograr que el Portal de Datos Abiertos Gubernamentales de Panamá obtenga la quinta estrella.

### Referencias

- [1] B. Vázquez, E. Morales, K. Nájera y S. Vieyra. Web Semántica: La evolución de una web con significado. [En línea]. Disponible en: <https://sg.com.mx/revista/44/web-semantica-la-evolucion-una-web-significado>
- [2] M. Abián, El futuro de la web: XML, RDF/RDFS, ontologías y la web semántica, 2005.
- [3] (2012) Cinco Estrellas Datos Abiertos. [En línea]. Disponible en: <https://5stardata.info/es/>
- [4] Principios. [En línea]. Disponible en: <https://opendatacharter.net/principles-es/>
- [5] (2016) LinkedData. [En línea]. Disponible en: <https://www.w3.org/wiki/LinkedData>
- [6] Datos Abiertos de Gobierno. [En línea]. Disponible en: <http://www.antai.gob.pa/datos-abiertos-de-gobierno/>
- [7] (2018) Gaceta Oficial Digital. [En línea]. Disponible en: [https://www.gacetaoficial.gob.pa/pdfTemp/28454/GacetaNo\\_28454\\_20180129.pdf](https://www.gacetaoficial.gob.pa/pdfTemp/28454/GacetaNo_28454_20180129.pdf)
- [8] (2014) RDF. [En línea]. Disponible en: <https://www.w3.org/RDF/>
- [9] Debian - The universal operating system. [En línea]. Disponible en: <https://www.debian.org/index.es.html>
- [10] (2012) OpenLink Virtuoso. [En línea]. Disponible en: [https://www.w3.org/2001/sw/wiki/OpenLink\\_Virtuoso](https://www.w3.org/2001/sw/wiki/OpenLink_Virtuoso)
- [11] Tomcat, Apache Tomcat, Jakarta Tomcat. [En línea]. Disponible en: <http://ajpdsoft.com/modules.php?name=Encyclopedia&op=content&tid=769>
- [12] R. Cyganiak, C. Bizer. Pubby A Linked Data Frontend for SPARQL Endpoints. [En línea]. Disponible en: <http://wifo5-03.informatik.uni-mannheim.de/pubby/>
- [13] (2018) Java Servlet. [En línea]. Disponible en: [https://es.wikipedia.org/wiki/Java\\_Servlet](https://es.wikipedia.org/wiki/Java_Servlet)
- [14] Protégé. [En línea]. Disponible en: <https://protege.stanford.edu/products.php>
- [15] Tabula. [En línea]. Disponible en: <https://tabula.technology/>
- [16] N. Humfrey. EasyRDF. [En línea]. Disponible en: <https://www.easyrdf.org/converter>
- [17] E. Prud'hommeaux. Validation Service. [En línea]. Disponible en: <https://www.w3.org/RDF/Validator/>
- [18] GraphDB Free Edition. [En línea]. Disponible en: <https://www.ontotext.com/products/graphdb/graphdb-free/>
- [19] Linked Open Vocabularies (LOV). [En línea]. Disponible en: <https://lov.linkeddata.es/dataset/lov/>
- [20] D. Brickley y Libby Miller. (2014) FOAF Vocabulary Specification 0.99. [En línea]. Disponible en: [http://xmlns.com/foaf/spec/#term\\_Person](http://xmlns.com/foaf/spec/#term_Person)
- [21] DBpedia. [En línea]. Disponible en: <https://wiki.dbpedia.org/>
- [22] Virtuoso SPARQL Query Editor. [En línea]. Disponible en: <http://dbpedia.org/sparql>
- [23] DBpedia LatAm. [En línea]. Disponible en: <http://es-la.dbpedia.org/home/>