



UNIVERSIDAD
NACIONAL
DE COLOMBIA

GENERACIÓN Y PUBLICACIÓN DE LINKED DATA PARA EL MONITOREO DE LA CALIDAD AMBIENTAL DEL AGUA

Juan David Rondón Díaz

Universidad Nacional de Colombia

Facultad de Ciencias Agrarias

Escuela de Posgrados

Maestría en Geomática

Bogotá, Colombia

2018

GENERACIÓN Y PUBLICACIÓN DE LINKED DATA PARA EL MONITOREO DE LA CALIDAD AMBIENTAL DEL AGUA

Juan David Rondón Díaz

Tesis presentada como requisito parcial para optar al título de:

Magister en Geomática

Director:

Ph.D. Luis Manuel Vilches Blázquez

Codirectora:

M.Sc. Denisse Cangrejo Aljure

Línea de Investigación:

Tecnologías Geoespaciales

Universidad Nacional de Colombia

Facultad de Ciencias Agrarias

Bogotá, Colombia

2018

.

*A Viviana,
No soy un gran hombre, pero si tengo a mi
lado una gran mujer.*

Agradecimientos

En primer lugar, y en especial quiero manifestar mi más sincera gratitud a Luis, ya que con su apoyo durante este proceso no solo ha sido guía e inspiración en el campo académico, profesional y personal, sino un gran amigo sin el cual no hubiera podido llevar a cabo este proyecto.

A mis padres y mi hermano que siempre me llevan en sus pensamientos y me dan todo su apoyo incondicional, a Viviana que no solo me impulsó a terminar y me dio la fuerza que necesitaba, sino que ha estado conmigo haciendo más fáciles los momentos de adversidad y siendo mi más grande apoyo, gracias por su paciencia.

Por último, quiero agradecer a la Corporación Autónoma Regional de Cundinamarca, que con su labor constante sobre las cuencas de su jurisdicción generan información vigente sobre el estado del recurso hídrico. De esta institución quiero hacer una distinción especial a Edwin García, Director de la Dirección de Laboratorio e Innovación Ambiental, ya que no solo me brindó su colaboración mediante el suministro de información, también con sus palabras me ha motivado a mejorar y crecer de manera constante sin dejar que ningún aprieto se convierta en un obstáculo.

Resumen

El agua dulce resulta un recurso de vital importancia para la salud humana, la sostenibilidad del medio ambiente y la prosperidad económica. Por tanto, hace necesario el desarrollo de sistemas de información que permitan el monitoreo de este recurso, teniendo en cuenta aspectos como cantidad y calidad. Para facilitar su monitoreo se propone adoptar Linked Data, que consiste en un conjunto de buenas prácticas que no solo buscan la publicación de información estructurada en la Web, sino que también apuesta por la interconexión entre distintas fuentes de datos, estandarización e interoperabilidad. La adopción de estas buenas prácticas (Linked Data), también involucran el uso de vocabularios (ontologías), para permitir dar significado a la información contenida y que ésta se convierta en conocimiento una vez se ha puesto en contexto. De igual manera, la adopción de estas buenas prácticas brinda soporte para la interoperabilidad, especialmente semántica, de los sistemas con información geoespacial de tipo ambiental (contaminación hídrica).

Este trabajo propone una solución que permita poner dentro del contexto legal vigente información correspondiente a características físico-químicas y microbiológicas de cuerpos hídricos y, al mismo tiempo, facilitar la conexión de estos datos con otras fuentes de información mediante los principios de Linked Data. Asimismo, esta propuesta busca promover el aprovechamiento de ontologías existentes e, incluso, de recursos no ontológicos, conforme a las recomendaciones presentes en el estado del arte, para llevar a cabo la definición del tipo de agua, datos y usos asociados, lo que conducirá al desarrollo de un caso de estudio para la interpretación eficiente de los datos hídricos de la cuenca del río Bogotá, indicando la peligrosidad y el potencial del uso del agua de acuerdo a la legislación que le aplique en el contexto requerido y su uso.

Palabras clave: Linked Data, calidad hídrica, monitoreo, ontología y RDF.

Abstract

Fresh water is a resource of vital importance for human health, the sustainability of the environment and economic prosperity. Therefore, it is necessary to develop information systems that allow the monitoring of this resource, taking into account aspects such as quantity and quality. In order to facilitate its monitoring, we propose to adopt Linked Data, which consists of a set of good practices that not only seek the publication of structured information on the Web, but also achieves interconnection between different data sources, standardization and interoperability. The adoption of these good practices (Linked Data), also involves the use of vocabularies (ontologies), to allow giving meaning to the contained information and that this becomes knowledge once it has been put into context. In the same way, the adoption of these good practices provides support for the interoperability, especially semantics, of the systems with geospatial information of environmental type (water pollution).

This work proposes a solution that allows putting within the current legal context information corresponding to physical-chemical and microbiological characteristics of water bodies and, at the same time, facilitate the connection of these data with other sources of information through the guidelines of Linked Data. In addition, this proposal seeks to promote the use of existing ontologies and even non-ontological resources, in accordance with the recommendations present in the state of the art, to carry out the definition of the type of water, data and associated uses. This scenario is developed in a case study for the efficient interpretation of water data of the Bogotá river basin, indicating the danger and potential of water use according to the regulation that applies to it in the required context and its use.

Keywords: Linked Data, water quality, monitoring, ontology, and RDF.

Contenido

Pág.

Resumen	IX
Lista de figuras	13
Lista de tablas	15
Capítulo 1. Introducción.....	16
Capítulo 2. Marco de referencia.....	21
2.1. Web Semántica	21
2.2. Ontologías	24
2.3. Resource Description Framework.....	26
2.4. Linked Data	29
2.5. Calidad en Linked Data.....	30
Capítulo 3. Estado del arte.....	34
Capítulo 4. Planteamiento.....	40
4.1. Alcance.....	40
4.2. Objetivos.....	41
4.3. Planteamiento general	42
Capítulo 5. Contexto y metodología.....	44
5.1. Contexto del trabajo.....	44
5.1.1. Zona de estudio	44
5.1.2. Fuentes de información.....	45
5.2. Metodología.....	48
5.2.1. Especificación	51
5.2.2. Modelamiento	52
5.2.3. Generación de RDF	52
5.2.4. Generación de enlaces	53
5.2.5. Publicación.....	54
5.2.6. Explotación	54
Capítulo 6. Resultados.....	55
6.1. Establecimiento de patrones de nombrado	55
6.2. Desarrollo de la red de ontologías	57
6.2.1. NeOn methodology	57
6.2.2. Combinación de escenarios para el desarrollo de la red de ontologías.....	59

6.2.3.	Especificación de requerimientos.....	61
6.2.4.	Reutilización y reingeniería de recursos ontológicos.....	66
6.2.5.	Reutilización y reingeniería de recursos no ontológicos.....	69
6.2.6.	Reutilización, fusión y re-ingeniería de recursos ontológicos	73
6.2.7.	Red de ontologías.....	76
6.4.	Generación de RDF y enlaces.....	77
6.4.1.	Datos históricos	77
6.4.2.	Datos normativos.....	79
6.4.3.	Datos espaciales	81
6.4.4.	Generación de enlaces	86
6.5.	Publicación y uso	88
6.5.1.	Repositorio de datos semánticos	88
6.5.2.	Explotación de datos.....	92
6.5.3.	Aplicación de análisis espacial y contexto legal	95
Capítulo 7.	Discusión de resultados.....	100
7.1.	Red de ontologías desarrollada	100
7.2.	Transformación y enlazado de datos	101
7.3.	Publicación de Linked Data	103
7.4.	Explotación de Linked Data	104
7.5.	Calidad del Linked Data	105
Capítulo 8.	Conclusiones y trabajo futuro.....	109
8.1.	Conclusiones.....	109
8.2.	Trabajo futuro	111
Bibliografía		113
Anexo I.....		120
Anexo II.....		124

Lista de figuras

Pág.

Figura 2-1. Arquitectura de la Web Semántica. (Berners-Lee, 2000).....	23
Figura 2-2. Triplete base del RDF (Hess, 2012)	27
Figura 2-3. Lista de clases RDFS (Brickley & Guha, 2014).....	28
Figura 2-4. Lista de propiedades en RDFS (Brickley & Guha, 2014)	29
Figura 2-5. Linked Open Data (Abele & McCrae, 2017).....	30
Figura 5-1. Localización de puntos de monitoreo, <i>Google Maps</i>	46
Figura 5-2. Histórico de muestreo y análisis de laboratorio (vista en formato Excel).....	47
Figura 5-3. Actividades de la metodología para la generación y publicación de Linked Data (Vilches-Blázquez et al., 2014).....	49
Figura 5-4. Diagrama de flujo de proceso desarrollado	50
Figura 6-1. Ciclo de vida NeOn	57
Figura 6-2. <i>Grafo Semantic Sensor Network</i> , (Compton et al., 2005)	65
Figura 6-3. Jerarquía clase “ <i>Observation</i> ” (Izquierda: WaterML 2.0, Derecha: O&M)	67
Figura 6-4. Clase “ <i>Timeseries</i> ”	68
Figura 6-5. Esquema de equivalencias para la clase “ <i>Observation</i> ”.	68
Figura 6-6. Parámetros de CUAHSI-HIS tras su transformación a ontología	70
Figura 6-7. Clases generadas para contexto legal y propiedades.....	72
Figura 6-8. Izquierda: <i>SamplingPoint</i> , Derecha: <i>Observation</i> (asignación de geometría) 74	
Figura 6-9. Clase “ <i>TimeSeriesObservation</i> ” en UML. Open Geospatial Consortium, 2014	74
Figura 6-10. Clase “ <i>TimeSeriesObservation</i> ” en UML y Protégé	75
Figura 6-11. Visión de alto nivel de la red de ontologías desarrollada	76
Figura 6-12. Estructura de RDF para la clase <i>Observation</i> y asignación de geometría ...	78
Figura 6-13. Estructura RDF para Unidad de Medida de Conductividad.....	79
Figura 6-14. Marco Normativa en LODRefine.....	80
Figura 6-15. Estructura RDF de marco normativo	80
Figura 6-16. Columnas de Latitud y Longitud y su formato conforme a WKT	83
Figura 6-17. Estructura RDF <i>SamplingPoint</i>	83
Figura 6-18. Vista de las clases para la parte alta de la cuenca del río Bogotá	84
Figura 6-19. Estructura RDF de Unidades Hidrológicas	85
Figura 6-20. Grafo de RDF generado para observaciones y sus puntos de muestreo	86
Figura 6-21. Enlace de datos a DBpedia utilizando URIs de municipio (Suesca).....	87
Figura 6-22. Arquitectura Parliament	89

Figura 6-23. Consulta de valores asociados a conductividad y pH.....	90
Figura 6-24. Resultados de valores asociados a conductividad y pH	90
Figura 6-25. Consulta de valores de conductividad con metadatos.....	91
Figura 6-26. Valores de conductividad con metadatos.....	92
Figura 6-27. Consulta SPARQL para análisis estadístico.....	94
Figura 6-28. Análisis estadístico	94
Figura 6-29. Consulta espacial para aplicar contexto legal.....	96
Figura 6-30. Puntos de monitoreo que infringen la norma para DBO (Mapa)	97
Figura 6-31. Puntos de monitoreo que infringen la norma para DBO (Tabular)	98
Figura 6-32. Análisis temporal DBO en Q. Socotá	98
Figura 7-1. Encabezado de resultado en formato XML	106
Figura 7-2. Enlaces entre recursos desde el navegador (Clase: <i>ContaminantLevels</i>)...	106
Figura 7-3. Enlace de municipio de Mosquera a DBPedia	107
Figura 7-4. Resultados de desempeño para 10 consultas desde R en segundos	108
Figura 10-1. Resultados de query en SPARQL como data frame.....	123
Figura 11-1. Entorno web de consulta.....	127
Figura 11-2. Consulta de concentración en punto de interés.....	128

Lista de tablas

[Pág.](#)

Tabla 2-1. Dimensiones y métricas de la calidad en Linked Data. Adaptado de (Hitzler et al., 2012)	32
Tabla 3-1. Trabajos relacionados con la propuesta	38

Capítulo 1. Introducción

Conforme a lo establecido por los Objetivos de Desarrollo Sostenible en la Agenda 2030 de la Organización de las Naciones Unidas (ONU), el acceso al agua apta para el consumo y el acceso al saneamiento, así como la gestión racional de los ecosistemas de agua dulce son esenciales para la salud humana, la sostenibilidad del medio ambiente y la prosperidad económica. No obstante, en la actualidad, se presentan diversos factores que ejercen una importante presión sobre los recursos hídricos¹, tales como: la expansión de la actividad industrial y agrícola, el crecimiento de la población humana y la amenaza del cambio climático. Todos estos factores son causantes de gran impacto en la alteración del ciclo hidrológico a nivel mundial (ONU, 2016). Este escenario pone de manifiesto la necesidad de realizar un adecuado diagnóstico de los recursos hídricos, para lo que el monitoreo de su calidad resulta una valiosa herramienta para este fin.

El monitoreo de la calidad ambiental, entendido como el seguimiento al conjunto de características del medio ambiente asociadas a la disponibilidad y facilidad de acceso a los recursos naturales y la ausencia o presencia de agentes nocivos de cualquier tipo (Sánchez & Guiza, 1989), incluye la evaluación de la calidad del agua. En este trabajo se considera “calidad del agua” a las características químicas, físicas, biológicas y radiológicas del agua que condicionan su uso para un fin específico, por ejemplo: consumo humano, agricultura, etc. (Diersing Florida Keys National Marine Sanctuary, 2017).

¹ En este documento se utiliza “recurso hídrico” y “agua” de forma indistinta.

La evaluación de la calidad del agua resulta de vital importancia, ya que al ser este un recurso esencial para la vida se debe garantizar su acceso a todas las personas, mediante un suministro satisfactorio, suficiente, accesible e inocuo en cuanto a lo se refiere a su calidad (OMS, 2017). Para ello resulta necesario la implementación de programas que promuevan la calidad del agua que son cruciales en los procesos de toma de decisiones para entender, interpretar y usar dicha información con el fin de apoyar las actividades de gestión de los recursos hídricos que deriven en la protección de los recursos naturales (Behmel, Damour, Ludwig, & Rodriguez, 2016).

Desafortunadamente, el monitoreo de la calidad hídrica se enfrenta a importantes retos técnicos asociados con los métodos de colección de datos y las estrategias de análisis (espacial o temporal) de dichos datos para la caracterización significativa de la calidad del agua (Myers y Ludtke, 2017). Ante estos retos, la utilización de las Tecnologías de la Información Geográfica (TIG), como son el caso de los Sistemas de Información Geográfica (SIG), desempeñan un papel fundamental, ya que representan una herramienta adecuada para el almacenamiento, procesamiento y divulgación de información de tipo espacial (Neteler y Mitasova, 2004), facilitando el análisis de los datos que pueden apoyar la toma de decisiones y favorecer la calidad ambiental, lo que contribuye de forma directa al bienestar de la comunidad mediante la gestión eficiente de los recursos.

Asimismo, el monitoreo de la calidad hídrica, al atender problemas asociados a ciencias ambientales (calidad ambiental), posee un carácter multidisciplinar. Por tanto, tal como señalan Visser y Stuckenschmidt (2002), enfrenta problemas asociados a los datos y la necesidad de conocer el origen de la muestra y todo su contexto. Por ejemplo, ante un proyecto sobre vertidos ilegales sobre recursos hídricos se requieren datos asociados a los ríos de la zona de estudio, vertimientos cercanos, fuentes de agua subterránea, etc. Esta necesidad de múltiples fuentes de información, con frecuencia, involucra datos que no se encuentran disponibles en una sola fuente (base de datos) ni en un único formato.

Este escenario conlleva que los usuarios pasen mucho tiempo instalando y aprendiendo una variedad de software en máquinas locales, buscando y recopilando datos geoespaciales de una variedad de fuentes, y pre-procesando y analizando dichos datos en máquinas locales. Este paradigma, conocido como *everything-locally-owned-and-operated* (Zhao, Foerster, y Yue, 2012), hace que el análisis y la aplicación de los datos

geoespaciales sea muy costoso y tome mucho tiempo. Además, los datos, para un ejemplo como el mencionado con anterioridad, están encerrados en silos y no pueden compartirse e integrarse entre organizaciones y comunidades. Como resultado, el análisis de datos se convierte en un privilegio propiedad de algunos usuarios especialistas en el dominio, y es posible que muchos datos no se hayan analizado lo suficiente (Zhao et al., 2012).

Para hacer frente a estos problemas, el progreso en Linked Data, ontologías y semántica geoespacial (Goodchild *et al.* 1999; Kuhn, 2005; Zhao et al., 2012) se consideran en este trabajo, debido a que proporcionan el elemento decisivo para conectar modelos, datos y usuarios (Janowicz y Hitzler, 2012). La adopción de los principios de Linked Data conlleva la adopción de las mejores prácticas para exponer, compartir e integrar datos a través de URI sin referencias en la Web (Heath y Bizer, 2011). Estas mejores prácticas están siendo adoptadas por un número creciente de proveedores de datos, lo que lleva a la creación de un espacio de datos global que contiene miles de millones de recursos, en la conocida como Web de los Datos (Heath y Bizer, 2011).

Existen diversos trabajos que resaltan la importancia de integrar y combinar datos de múltiples y heterogéneas fuentes de información como lo sugiere la Web Semántica. En el campo de las ciencias ambientales se encuentran ejemplos como los mencionados en Hunter *et al.* (2011), dentro de los cuales se destaca CUAHSI-HIS (CUAHSI, 2010). Asimismo, el Departamento para el Ambiente, Alimentos y Asunto Rurales (*Department for Environment Food & Rural Affairs, DEFRA*) en el Reino Unido publica información histórica y actualizada de aguas de baño y de cuencas hidrológicas conforme a los principios de *Linked Data*². Estas propuestas conforman un referente de la pertinencia de las nuevas tendencias en el área de las Tecnologías de la Información Geográfica, constituyendo un aporte valioso al campo hidrológico mediante la extensión de las tecnologías asociadas a la Geomática.

Aunque los casos ya mencionados son prueba de la aplicación de esta tecnología para el monitoreo de recursos hídricos, este trabajo propone la generación y publicación de Linked Data geoespacial para el monitoreo de la calidad ambiental del agua. La novedad de esta propuesta se centra en la interpretación de información correspondiente a

² <https://environment.data.gov.uk>

propiedades físicas, químicas y biológicas de cuerpos hídricos en Colombia, así como su alineación con el marco legal vigente en el país, que incluye una serie de límites máximos permitidos de variables físico-químicas y microbiológicas del agua en función del uso al que sea destinado y su ubicación en un tramo, sin las cuales los datos podrían tener cualquier tipo de interpretación. De esta manera, este trabajo pretende mostrar la información de calidad asociada a los cuerpos hídricos e identificar la peligrosidad potencial de los mismos, haciendo accesible esta información y su respectivo análisis a cualquier tipo de usuario, sin necesidad de ser un experto en el tema.

Para el caso de estudio de este trabajo se busca dar una estructura a la información que permita hacer supuestos acerca de la interpretación de los términos asociados al estado de la **calidad del agua** y su contexto, el cual es requerido para la interpretación de la gran variedad de propiedades físico-químicas y microbiológicas. Para ello se tendrá en cuenta el uso contemplado para el agua a evaluar (por ejemplo: consumo humano, industrial, comercial) y la matriz a la que pertenece (agua superficial, residual industrial, subterránea, etc.), lo que permite determinar su potencial y/o peligrosidad de acuerdo al marco legal vigente que aplique según corresponda.

En conclusión, este trabajo pretende poner de manifiesto la pertinencia de la aplicación de las tecnologías asociadas a la Web Semántica en las ciencias ambientales en el estudio de cuerpos hídricos en la evaluación de su calidad y el grado de cumplimiento de acuerdo la legislación colombiana vigente, mediante la integración de distintas y heterogéneas fuentes de información. Asimismo, refleja el carácter innovador de este trabajo en el campo de la Geomática, no solo a nivel nacional sino también internacional, haciendo de este proyecto una propuesta que permitirá generar resultados relevantes para el área de las ciencias de la información geoespacial y la Geomática.

Este documento se estructura de la siguiente manera: En el capítulo 2 se recoge el marco de referencia asociado con este trabajo, seguido por el estado del arte en el capítulo 3. A continuación, se hace una descripción del alcance, objetivos y planteamiento general de este trabajo en el capítulo 4. En el capítulo 5 se presenta el contexto de la zona de estudio, los datos utilizados y la metodología adoptada para el desarrollo de este trabajo, mientras en el capítulo 6 se presentan los resultados obtenidos en este trabajo. Entre los resultados descritos en este capítulo se recogen: la red de ontologías desarrollada para modelar semánticamente las fuentes de información

consideradas se presenta en la sección 6.1. A continuación se describe el proceso de generación de RDF y enlaces, así como la publicación y explotación de los datos generados a partir de un caso práctico que permite mostrar la aplicabilidad de este trabajo, recogidos en las secciones 6.2. y 6.3. Por último, se presenta la discusión de resultados (capítulo 7) y las conclusiones y el trabajo futuro previsto en el capítulo 8.

Capítulo 2. Marco de referencia

Este capítulo presenta una breve descripción de los principales conceptos asociados con este proyecto, para dar a conocer y proporcionar claridad sobre las bases conceptuales que lo sustentan. Así, el capítulo describe las características de la Web Semántica, ontologías, *Resource Description Framework* (RDF) y Linked Data. Estos conceptos contribuyen a que el lector pueda comprender los requisitos y características requeridos por la información geoespacial para que ésta pueda ser publicada conforme a los principios de Linked Data.

2.1. Web Semántica

Tal como se presenta en (Berners-Lee, 1998), la Web ha sido diseñada como un espacio de información, cuyo objetivo no debería centrarse únicamente en ser útil para la comunicación entre personas, sino también debería permitir la participación y ayuda de las máquinas (computadores). Uno de los mayores obstáculos para la consecución de este objetivo ha sido el hecho de que la mayor parte de la información en la Web ha sido creada para ser consumida por humanos, incluso cuando esta información procede de una base de datos con significados bien definidos (al menos en parte), ya que su estructura de datos no resulta evidente para la búsqueda de información por parte de un robot en la Web. En este escenario, la Web Semántica busca desarrollar lenguajes para expresar información de manera que sea procesable por las máquinas.

La Web Semántica se puede definir como un conjunto de estándares y buenas prácticas para compartir datos y la semántica de los mismos en la web para su posterior uso por aplicaciones (DuCharme, 2011). El incorporar estándares en el escenario de la Web Semántica involucra la idea de incluir el modelo de datos RDF, el lenguaje de consulta SPARQL (*SPARQL Protocol and RDF Query Language*³) y el esquema del modelo de datos RDF (*RDF-Schema*⁴) junto con el estándar OWL (*Web Ontology Language*⁵) que hacen parte del *World Wide Web Consortium (W3C*⁶). La consideración de los estándares mencionados se justifica en el hecho de que a pesar de generarse un producto que incorpore la semántica, si esta carece de estándares, no puede conectarse y hacer parte de la Web Semántica. La importancia de esta Web se debe a la forma en que esta permite integrar datos heterogéneos una vez existe el enlace entre distintos agentes, servicios y aplicaciones para el intercambio de información, datos y conocimiento de manera significativa dentro y fuera de la Web. Esta integración se facilita una vez se tiene un mismo vocabulario compartido o se ha generado la debida correspondencia o mapeo entre los diferentes vocabularios (Noy, McGuinness, & Hayes, 2005). Este es uno de los objetivos para los que fueron diseñados el RDF y OWL al proporcionar mecanismos para la construcción de dichos vocabularios y mapeos.

En definitiva, para que la Web Semántica funcione se requiere que la computadora tenga acceso a colecciones de información estructurada, así como a conjuntos de reglas sobre las cuales pueda hacer inferencias y permitir un razonamiento automático (Berners-Lee, 2001). Esto es posible al brindar una estructura al contenido con significado que poseen las páginas Web, proporcionando un entorno para que agentes (entendidos como programas capaces de coleccionar contenido Web de distintas fuentes, procesar la información y presentar los resultados con otros programas) puedan llevar a cabo sofisticadas tareas al usuario durante la consulta.

A continuación, se describen brevemente las capas que conforman la Web Semántica y que se recogen en la Figura 2-1:

³ <https://www.w3.org/TR/rdf-sparql-query/>

⁴ <https://www.w3.org/TR/rdf-schema/>

⁵ <https://www.w3.org/OWL/>

⁶ <https://www.w3.org/>

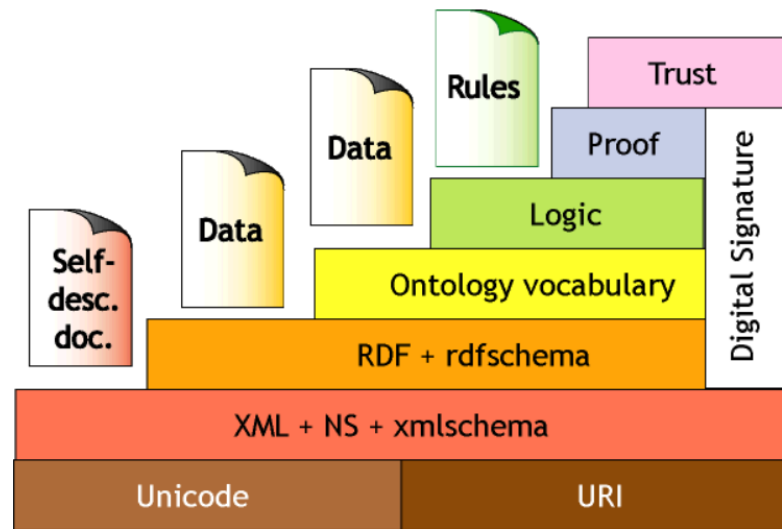


Figura 2-1. Arquitectura de la Web Semántica. (Berners-Lee, 2000)

- **Unicode:** Codificación para permitir símbolos de diferentes idiomas. De esta manera, se puede expresar información en la Web Semántica en cualquier idioma.
- **URI:** "Uniform Resource Identifier" o Identificador Uniforme de Recursos, permite la localización de un recurso de manera unívoca que puede ser accedido a través de Internet.
- **XML + NS + XML Schema:** En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan interactuar. XML es un lenguaje de marcado que ofrece un formato común para intercambio de documentos, NL (*namespaces*) sirve asociar elementos de XML con espacios de nombre y XML Schema ofrece un esquema para elaborar documentos estándar.
- **RDF + RDF Schema:** Esta capa define el lenguaje universal con el cual podemos expresar los diferentes conceptos en la Web Semántica. RDF, es un lenguaje simple mediante el cual definimos sentencias en el formato de una tripleta (sujeto: el recurso al que nos referimos; predicado: el recurso que indica qué es lo que estamos definiendo; y objeto: puede ser el recurso o un literal que podría considerarse el valor de lo que acabamos de definir). RDF Schema, provee un

esquema de metadatos definido sobre RDF que permite el modelado de objetos con una semántica claramente definida.

- **Lenguaje de Ontologías:** El uso de ontologías permite clasificar objetos y sus relaciones con otros objetos pertenecientes a un dominio concreto del conocimiento. Esta capa permite extender la funcionalidad de la Web Semántica, describiendo recursos a través de nuevas clases y propiedades.
- **Lógica:** Se requieren reglas de inferencia para aplicar “razonadores” sobre las ontologías para que de esta manera las máquinas puedan consultar y manipular términos de manera eficiente haciendo uso de la inteligibilidad humana.
- **Pruebas:** Intercambio de "pruebas" escritas en el lenguaje unificador de la Web Semántica con el fin de posibilitar las inferencias lógicas.
- **Confianza:** Suministran comprobación sobre el uso de fuentes de información en la Web Semántica. En este marco se utilizan las ontologías asociadas con *Web Of Trust RDF Ontology*⁷ (WOT) y FOAF⁸ (*Friend Of A Friend*).
- **Firma digital:** Método de encriptación para verificar que la información adjunta ha sido ofrecida por una fuente específica confiable. Se considera como referencia la propuesta de *XML Signature*⁹.

2.2. Ontologías

Dentro de las definiciones de ontología una de las más extendidas es aquella recogida en Gruber (1993), que afirma que *una ontología es una especificación explícita de una conceptualización compartida*. Donde por *conceptualización* se entiende un modelo abstracto de cómo normalmente la gente piensa acerca de un objeto del mundo real (Buccella *et al.*, 2009). Por *explícita* se refiere al tipo de conceptos utilizados que deben ser explícitamente definidos, donde si también puede describir otros conceptos del mismo tipo, se definen detalladamente. Por *formal* hace referencia a como la ontología debe ser legible para la máquina. Y, finalmente, la característica de *compartido* exige que la ontología no deba ser restringida para un solo individuo sino aceptada por una

⁷ <http://xmlns.com/wot/0.1/>

⁸ <http://xmlns.com/foaf/0.1/>

⁹ <http://www.w3.org/Signature/>

comunidad. Teniendo en cuenta lo anterior y como se explica en (Lozano, 2001), las ontologías definen conceptos y relaciones de algún dominio, de forma compartida y consensuada; y esta conceptualización debe ser representada de una manera formal, legible y utilizable por la máquina.

Asimismo, las ontologías se usan frecuentemente para describir de forma explícita las características de un conjunto de datos de una manera formal. Son una propuesta extendida para la solución de problemas asociados con la heterogeneidad de datos y representan una alternativa para permitir la homogenización e integración de Sistemas de Información Geográfica (SIG) e Infraestructuras de Datos Espaciales (IDE) necesarias para facilitar procesos de acceso, intercambio y consulta en el marco de la interoperabilidad semántica (Vilches-Blázquez, 2009). El desarrollo y uso de las ontologías hace posible realizar supuestos acerca del significado de los términos disponibles, correspondientes a una explicación del contexto en el cual el término es evaluado (Lenat, 1998).

Aunque el uso de ontologías tiene sus inicios en la inteligencia artificial, se han dado posteriores investigaciones en el campo de las ontologías dentro de la comunidad de las ciencias de la computación, en áreas como lingüística computacional y teoría de base de datos (Fonseca *et al.*, 2000). De esta manera, esta área de la Web Semántica abarca diversos campos que van desde la ingeniería del conocimiento, integración de información y análisis orientado a objetos en aplicaciones como medicina, ingeniería mecánica y/o sistemas de información geográfica.

Una ontología puede clasificarse de acuerdo al nivel de detalle y al grado de dependencia sobre una tarea o punto de vista particular (Guarino, 1997). Según el nivel de detalle se puede clasificar por la precisión ontológica que puede ir desde el catálogo hasta la teoría axiomatizada. Una de las clasificaciones más extendidas es la recogida a continuación (Fallahi *et al.*, 2008):

- **Ontología de nivel superior** (*Upper level*): Describen conceptos generales y son independientes de un problema o dominio específico.

- **Ontología de dominio o de tarea:** Responsable del vocabulario relacionado de un dominio genérico o una tarea o actividad general, por medio de la especialización de términos de la ontología de nivel superior.
- **Ontología de aplicación:** Describe conceptos de acuerdo al dominio particular y la tarea. De manera frecuente, corresponden a roles desempeñados por entidades del dominio mientras desarrollan cierta actividad.

Al clasificar las ontologías por su precisión a la hora de caracterizar la conceptualización a la que corresponde, existen dos mecanismos por medio de los cuales se acerca a la conceptualización (Guarino, 1998): por medio del desarrollo de una axiomatización más abundante y adoptando un dominio y conjunto de relaciones conceptuales relevantes más rico.

Cuando el dominio de conocimiento involucra varias aristas, por ejemplo, tiempo y espacio o la ontología puede dividirse por módulos, es común encontrar redes de ontologías, que son una colección de ontologías relacionadas entre sí, a través de diferentes tipos de relación tales como mapeo, modularización, versionamiento entre otras (Haase, Rudolph, & Wang, 2006). Dichos procesos se describen a continuación como se define en (Suárez-Figueroa, 2017):

- **Mapeo de ontologías:** Actividad de encontrar correspondencias entre dos o más ontologías y almacenar/explotarlas. También se conoce como *alineación de ontologías*.
- **Modularización de ontologías:** Actividad o identificación de uno o más módulos en una ontología con el propósito de apoyar su re-uso o mantenimiento.
- **Versionado de ontologías:** Actividad de manejo de cambios de la ontología mediante la creación y gestión de diferentes versiones de la ontología.

2.3. Resource Description Framework

Resource Description Framework (RDF) es un modelo de datos estándar para el intercambio de datos en la Web (Heath & Bizer, 2011). Este estándar del W3C ha sido

diseñado para de representar de manera integrada información proveniente de distintas fuentes que se encuentran heterogéneamente estructuradas y que son representadas mediante el uso de diferentes esquemas (Auer *et al.*, 2011). Las declaraciones RDF se conocen como triples o tripletas, lo que consiste en la representación de un sujeto (recurso), predicado (propiedad) y un objeto (valor) DuCharme (2011), (ver Figura 2-2), estos son identificados mediante el uso de URI con el fin de eliminar cualquier tipo de ambigüedad. En las *tripletas RDF*, el sujeto corresponde al elemento que se busca describir, mientras que el predicado o *propiedades* describe la relación entre el sujeto y el objeto, estas relaciones son formalizadas en lenguaje RDF (Hebeler, Fisher, Blace, & Perez-Lopez, 2009). Las tripletas constituyen una poderosa herramienta en la integración de información, ya que son colecciones de URI y literales donde estas dos tienen un alcance global de manera inherente. Es importante hacer uso de nombre globales (URI) y así poder realizar uniones sin necesidad de hacer traducción a algún idioma específico permitiendo que los grafos puedan ser transportados y combinados, facilitando así el intercambio de datos.

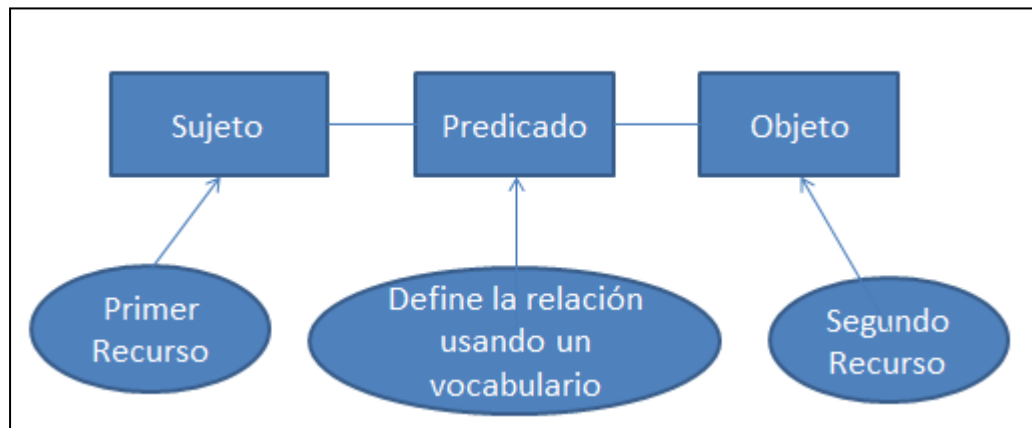


Figura 2-2. Tripletas base del RDF (Hess, 2012)

Gracias a sus características, el RDF facilita la unión de datos incluso si los esquemas de los mismos son diferentes, específicamente permite la evolución de los esquemas en el tiempo sin necesidad de realizar cambios sobre los datos a los que estos son aplicados (W3C, 2017). Asimismo, el RDF es un modelo de datos optimizado para el intercambio, este se facilita principalmente gracias a la estructura de la unidad básica del grafo del

RDF y espacio de nombre global proporcionado por el uso de URIs (cadenas de texto usada para identificar un recurso).

Como se mencionó con anterioridad, este es un modelo y no un formato de datos (Heath & Bizer, 2011). Esto quiere decir que el RDF no almacena los datos, por lo tanto, debe ser *serializado*, para ello, es necesario tomar las *tripletras* que conforman el RDF y plasmarlas en un archivo utilizando alguna sintaxis en particular, las más populares son: RDF/XML, *N-Triples*, N3 y *Turtle*.

Los elementos que componen al RDF se pueden dividir en grupos llamados clases y los miembros de una clase se conocen como instancia de una clase. Estas clases se pueden describir usando propiedades RDF, como, por ejemplo, *rdf:Type* para indicar que un recurso es una instancia de una clase.

Las clases se identifican como *rdfs:Class*, pero también puede ocurrir que una clase sea subclase de otra, si es así, se identifica como *rdfs:SubClassOf*. A continuación (Figura 2-3), se exponen algunas de las clases principales o más usadas en RDF-S (Brickley & Guha, 2014):

Class name	comment
rdfs:Resource	The class resource, everything.
rdfs:Literal	The class of literal values, e.g. textual strings and integers.
rdf:langString	The class of language-tagged string literal values.
rdf:HTML	The class of HTML literal values.
rdf:XMLLiteral	The class of XML literal values.
rdfs:Class	The class of classes.
rdf:Property	The class of RDF properties.
rdfs:Datatype	The class of RDF datatypes.
rdf:Statement	The class of RDF statements.
rdf:Bag	The class of unordered containers.
rdf:Seq	The class of ordered containers.
rdf:Alt	The class of containers of alternatives.
rdfs:Container	The class of RDF containers.
rdfs:ContainerMembershipProperty	The class of container membership properties, <i>rdf:_1</i> , <i>rdf:_2</i> , ..., all of which are sub-properties of 'member'.
rdf:List	The class of RDF Lists.

Figura 2-3. Lista de clases RDFS (Brickley & Guha, 2014)

Asimismo, las propiedades de RDF, las cuales permiten especificar restricciones de tipos de datos para los sujetos u objetos de las tripletras, forman parte del RDF. Estas propiedades permiten relacionar los recursos del sujeto y los recursos del objeto. Algunas de las propiedades más usadas se pueden observar en la Figura 2-4.

Property name	comment	domain	range
rdf:type	The subject is an instance of a class.	rdfs:Resource	rdfs:Class
rdfs:subClassOf	The subject is a subclass of a class.	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	The subject is a subproperty of a property.	rdf:Property	rdf:Property
rdfs:domain	A domain of the subject property.	rdf:Property	rdfs:Class
rdfs:range	A range of the subject property.	rdf:Property	rdfs:Class
rdfs:label	A human-readable name for the subject.	rdfs:Resource	rdfs:Literal
rdfs:comment	A description of the subject resource.	rdfs:Resource	rdfs:Literal
rdfs:member	A member of the subject resource.	rdfs:Resource	rdfs:Resource
rdf:first	The first item in the subject RDF list.	rdf:List	rdfs:Resource
rdf:rest	The rest of the subject RDF list after the first item.	rdf:List	rdf:List
rdfs:seeAlso	Further information about the subject resource.	rdfs:Resource	rdfs:Resource
rdfs:isDefinedBy	The definition of the subject resource.	rdfs:Resource	rdfs:Resource
rdf:value	Idiomatic property used for structured values.	rdfs:Resource	rdfs:Resource
rdf:subject	The subject of the subject RDF statement.	rdf:Statement	rdfs:Resource
rdf:predicate	The predicate of the subject RDF statement.	rdf:Statement	rdfs:Resource
rdf:object	The object of the subject RDF statement.	rdf:Statement	rdfs:Resource

Figura 2-4. Lista de propiedades en RDFS (Brickley & Guha, 2014)

2.4. Linked Data

Linked Data hace referencia al conjunto de buenas prácticas que permiten publicar y conectar datos estructurados en la Web (Bizer *et al.*, 2009), haciendo uso de la arquitectura de la web y del modelo de datos RDF. Esta iniciativa de datos enlazados (Linked Data) se basa en la aplicación de unos principios básicos para publicar y vincular datos estructurados en la Web Semántica. Estos principios de diseño fueron introducidos por Tim Berners-Lee y son los que se exponen a continuación:

- Usar URI (*Uniform Resource Identifier*) para identificar los recursos.
- Usar HTTP URIs para que los interesados puedan buscar estos recursos.
- Proporcionar información útil, usando los estándares (RDF, SPARQL), cuando se busque una URI.
- Incluir enlaces a otras URI, de manera que sea posible descubrir más cosas en esta información.

El objetivo del *Linked Data* es pasar de un modelo Web basado en documentos HTML a un esquema de datos estructurados enlazados (Vilches-Blázquez, 2011). Por tratarse de datos estructurados, estos se convierten en información legible que, mediante la conexión entre conceptos, que se da gracias a los principios ya mencionados, se

convierten además en datos enlazados para que los sistemas se encuentren en capacidad de explotar esta información de manera automática sin necesidad de un usuario que deba interpretar la información por encontrarse de manera no estructurada (por ejemplo, HTML, PDF).

En la Figura 2-5 se muestra el modelo de topología actual que conforma los múltiples conjuntos de datos publicados conforme a los principios de Linked Data.

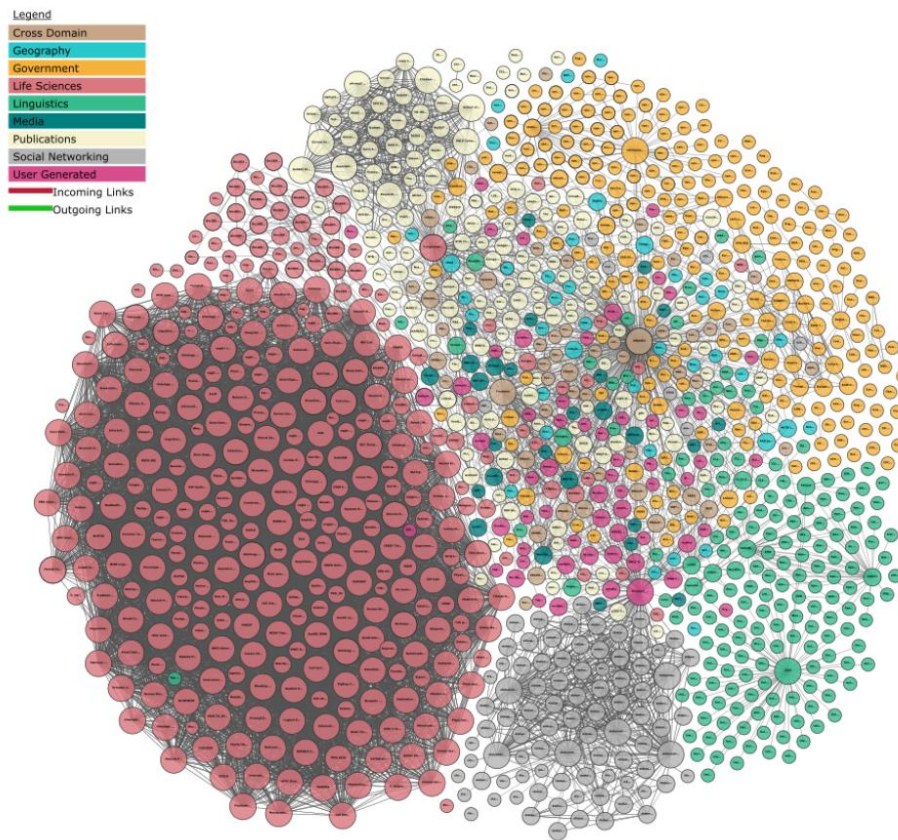


Figura 2-5. Linked Open Data (Abele & McCrae, 2017)

2.5. Calidad en Linked Data

El significado de calidad depende del contexto en el que se aplica (Longley, et al., 2005). No obstante, el término se usa comúnmente para indicar la superioridad de un producto manufacturado o atestiguar un alto grado de artesanía o arte. En el contexto de los datos, según (Longley, et al., 2005), la definición de calidad resulta más difícil de definir, ya que, a diferencia de los productos manufacturados, los datos no tienen

características físicas que permitan evaluar la calidad fácilmente. La calidad es, por lo tanto, una función de las propiedades intangibles, tales como "integridad" y "consistencia".

En el contexto de los datos geoespaciales, la proliferación de este tipo de datos y la democratización de su uso han convertido a la calidad en un tema de especial relevancia en los últimos tiempos. Esta situación, incluso, está derivando en una reevaluación de las responsabilidades de los productores de datos y los consumidores respecto de la calidad de los mismos (Longley, et al., 2005).

Ante este escenario, la calidad de un conjunto de datos puede tener un gran impacto en las aplicaciones que lo usan. Como consecuencia, conforme a las mejores prácticas asociadas a los datos en la Web¹⁰, propuestas con el W3C, resulta primordial incluir información de calidad de los datos en los procesos de publicación y consumo de los mismos. Así, por lo general, la evaluación de la calidad implica diferentes tipos de dimensiones de calidad, cada una representando grupos de características que son relevantes para los productores y los consumidores.

Desde la perspectiva de los datos publicados como Linked Data, estudios como los que se presentan en (Hogan, et al., 2012) han mostrado que la mayoría de los conjuntos de datos publicados como Linked Data poseen diversos problemas de calidad, tales como: representación, inconsistencia o asuntos asociados a la interoperabilidad. Así, con el objetivo de evaluar la calidad de datos en el contexto de Linked Data, entendiéndose calidad como *aptitud para el uso* (Rula & Zaveri, 2014), los trabajos actuales en esta área consideran diversas dimensiones y métricas. Algunas de las dimensiones y métricas más relevantes para la evaluación de la calidad del Linked Data se recogen en la Tabla 2-1.

¹⁰ <https://www.w3.org/TR/dwbp/>

Tabla 2-1. Dimensiones y métricas de la calidad en Linked Data.
Adaptado de (Hitzler et al., 2012)

Dimensión	Métrica	Descripción
Disponibilidad	Accesibilidad al Endpoint SPARQL y al servidor	Respuesta a consultas SPARQL del servidor
	Accesibilidad al volcado (<i>dump</i>) de RDF	Disponibilidad de volcado y descarga de RDF
	Diferenciabilidad de URI	Revisión de 1) Enlaces rotos. 2) Información útil (particularmente RDF) obtenida al consultar una URI. 3) Cambios en la URI.
	Ausencia de contenidos mal reportados	Respuesta HTTP con encabezado adecuado para el archivo requerido. Por ejemplo, <code>application/rdf+xml</code>
	Referencia de enlaces directos	URI local sea mencionada en el sujeto (p.e. descripción del recurso) para todas las tripletas.
Enlazado	Detección de enlaces de buena calidad	1) Presencia de cadenas sameAs en la red 2) Valor agregado añadido a la descripción del recurso mediante el uso de nodos sameAs

Dimensión	Métrica	Descripción
	Existencia de enlaces a fuentes de datos externas	Existencia y uso de URIS externas (p.e. usando enlaces owl:sameAs)
	Detección de los enlaces locales	Todas las tripletas de un conjunto de datos que usan la URI del recurso como objeto
	Uso de URIs-slash	Uso de URIs-slash cuando se proporciona grandes cantidades de información
	Baja latencia	(Mínima) Demora entre la petición de un usuario y la respuesta por parte del sistema
Desempeño	Alto desempeño	(Máximo) Número de respuesta a peticiones HTTP por segundo
	Escalabilidad de la fuente de datos	Detección del tiempo de respuesta que toman diez peticiones dividido entre diez que no sea mayor al de atender una

Capítulo 3. Estado del arte

Este capítulo presenta los trabajos relacionados con esta propuesta, con el fin de establecer la situación de partida del estado de la cuestión (alcances y limitaciones) previa al desarrollo generado en este proyecto.

En este sentido, la literatura presente en el estado del arte refleja que la Geomática no escapa al interés del *Linked Data*. De hecho, al involucrar la Geomática una gran cantidad de disciplinas enfrenta grandes retos relacionados con la interoperabilidad, siendo este el motivo por el cual la semántica geoespacial, las geo-ontologías y la interoperabilidad semántica en conjunto se han convertido en una importante área activa de investigación durante los últimos 20 años (Janowicz *et al.*, 2012).

Así, en la actualidad, hay varias iniciativas que están añadiendo datos enlazados (Linked Data) en el contexto de la información geográfica, por ejemplo: *GeoNames*¹¹, *LinkedGeoData*¹², *DBpedia*¹³, etc. Incluso organizaciones de gran importancia en el campo geoespacial, como es el caso del *Open Geospatial Consortium*¹⁴ (OGC), ha desarrollado un estándar (OGC, 2012b) capaz de desplegar y consultar información geoespacial en la Web Semántica. Este estándar, denominado *GeoSPARQL*¹⁵ desarrolla un vocabulario diseñado para representar datos geoespaciales en RDF y brinda una

¹¹ <http://www.geonames.org/>

¹² <http://linkedgeodata.org/About>

¹³ <http://dbpedia.org>

¹⁴ <http://www.opengeospatial.org/>

¹⁵ <http://www.opengeospatial.org/standards/geosparql>

extensión al lenguaje de consulta SPARQL para el procesamiento de datos geoespaciales.

Diversos autores han propuesto estrategias, metodologías y herramientas tecnológicas para solventar los problemas o exponer soluciones alternativas relacionadas al proceso de generación y publicación de Linked Data geoespacial (Varanka, 2008; Auer, Lehmann, y Hellmann, 2009; Goodwin, Dolbear, & Hart, 2009; Battle y Kolas, 2011; Usery y Varanka, 2012; GADM-RDF¹⁶; Vilches-Blázquez, Villazón-Terrazas, Corcho, y Gómez-Pérez, 2014; Koubarakis et al., 2014; Feliachi, Abadie, Hamdi, y Atemezing, 2013; Lehmann et al., 2015), datos de sensores (Patni, Henson, y Sheth, 2010; Barnaghi, Presser, y Moessner, 2010), y *geospatial provenance* (Yuan, Yue, Gong, y Zhang, 2013; Celino, 2013).

Por otro lado, entre los trabajos existentes en el estado del arte relacionados con metodologías para la generación y publicación de información geoespacial conforme a los principios de Linked Data, destaca el trabajo de (Bogdanovi, Stanimirovi, y Stoimenov, 2015), donde con el fin de mejorar el descubrimiento de fuentes de datos geoespaciales han propuesto una metodología para la integración de arquitecturas de geo-información basada en ontologías. En este trabajo se lleva a cabo la colección de geo-información mediante conexiones (mapeos) entre ontologías y fuentes de geo-información. Por otro lado, en (Vilches-Blázquez et al. 2014) se presenta el desarrollo de una propuesta metodológica para la generación, integración, desarrollo y publicación del Linked Data geoespacial a partir de distintas fuentes de información, y poder atender las limitantes que se presentan durante la integración de información geográfica dentro de la tierra digital (*Digital Earth*¹⁷).

Al igual que para el campo geoespacial, dentro del dominio hidrológico y ambiental se han formulado iniciativas como las presentadas en Hunter *et al.* (2011). En esta propuesta se destaca CUAHSI-HIS (CUAHSI, 2010) por tratarse de un claro aporte al campo de la hidrología al proveer un sistema para compartir información hidrológica. Para este fin, los autores han desarrollado un conjunto de reglas para describir y

¹⁶ GADM-RDF, <<http://gadm.geovocab.org/>> (accessed 02 November, 2017)

¹⁷ www.digitalearth-isde.org/userfiles/The_Digital_Earth_Understanding_our_planet_in_the_21st_Century.doc

representar el conocimiento (ontología) asociado a la hidrología. Estas reglas están destinadas a realizar un análisis de series de tiempo correspondientes a colecciones de datos definidos en un punto fijo. En el contexto de la ontología propuesta se incluyen propiedades físicas, químicas y biológicas. Por otro lado, existen otras soluciones como la propuesta por el *Department for Environment Food & Rural Affairs (DEFRA)*. Esta organización publica información histórica y actualizada del estado de calidad de aguas de baño e información de cuencas como *Linked Data*¹⁸, el cual se encuentra enmarcado dentro del *Defra Open Strategy* (Defra, 2014).

Además de estos trabajos, existen iniciativas dedicadas a la gestión integral del agua, donde se pone especial énfasis en el tratamiento de la información relacionada (Wilson, Reely, y Cox, 1997; Curry, Degeler, Clifford, Coakley, y Costa, 2014; Kämpgen, Riepl, y Klinger, 2014). En estos trabajos se han aprovechado métodos para la gestión del agua, formalizándose mediante el desarrollo de ontologías y aplicando los lineamientos del Linked Data. Además, junto a la utilización de tecnologías semánticas, estos trabajos proporcionan aportes en la generación de modelos dentro del dominio hídrico.

A nivel nacional el mecanismo que busca integrar todos los datos asociados al agua mediante la creación de una única base de datos es el Sistema de Información del Recurso Hídrico – SIRH¹⁹, dicha base de datos ha sido formulada teniendo en cuenta un diccionario de datos²⁰ creado para dicho fin. No obstante, esta propuesta no tiene ninguna relación con los principios de Linked Data.

Teniendo en cuenta las propuestas presentes en el estado del arte, existen diversas iniciativas relacionadas con este trabajo, tales como DEFRA (Defra, 2014), WATERCONOMICS (Curry et al., 2014) y SMART Project (Kämpgen et al., 2014). Sin embargo, entre las limitaciones de estas propuestas resulta destacable la falta de uso de estándares de amplio uso que incluyan el componente espacial dentro de la perspectiva de la Web Semántica. Además, en los casos de las propuestas mencionadas no se considera la utilización de un lenguaje común para la descripción de datos asociados al

¹⁸ <http://environment.data.gov.uk>

¹⁹ <http://sirh.ideam.gov.co:8230/Sirh/pages/inicio.html>

²⁰ <http://sirh.ideam.gov.co:8230/Sirh/pages/DiccionarioDeDatos.pdf>

dominio hidrológico. En la Tabla 3-1 se recogen las principales características de los trabajos relacionados con Linked Data y datos geoespaciales en el dominio hidrológico; estas características corresponden al origen de los datos, su naturaleza, formato en que estos son difundidos. Además se recogen los vocabularios/ontologías utilizados por cada iniciativa y su clasificación con respecto al dominio al que pertenecen (geoespacial e hidrológico).

Aunque los aportes recogidos en dicha tabla son claros exponentes de aportes en el campo de la Geomática mediante la incorporación del Linked Data como en el caso de GeoNames o DBpedia, estos aportes pueden ser ampliados mediante el uso del vocabulario GeoSPARQL, propuesto por Open Geospatial Consortium (OGC), que además de involucrar un diccionario geoespacial en RDF permite realizar consultas espaciales en el contexto de la Web Semántica.

En el campo de la hidrología existen propuestas que ofrecen un sistema de información para la gestión del recurso hídrico, como por ejemplo: SIRH, e iniciativas que utilizan los principios de la Web Semántica, como por ejemplo es el caso de DEFRA. No obstante, resultan escasas las propuestas que muestran prueba de hacer uso de vocabularios estandarizados que faciliten el análisis de esta información, como sucede en el caso de IWRM, que utiliza el vocabulario estándar propuesto por W3C en RDF Data Cube²¹.

Por otro lado, con respecto al tema específico de análisis de calidad hídrica, cabe destacar el proyecto desarrollado por DEFRA, el cual a partir de datos en RDF genera un visor con mapa web que no solo permite la consulta del estado del agua sino su ubicación geoespacial. Sin embargo, este trabajo no permite el uso de consultas geoespaciales, ya que no utiliza el vocabulario estándar GeoSPARQL, propuesto por OGC. Por tanto, la evaluación del estado del recurso hídrico no se encuentra ligada a su localización. Además, únicamente considera aguas dedicadas a actividades de ocio, tales como aguas para fines turísticos y playas.

²¹ <https://www.w3.org/TR/vocab-data-cube/>

Tabla 3-1. Trabajos relacionados con la propuesta

Iniciativa	Fuente de Información	Formato de Datos	Vocabulario	Conjunto de Datos Utilizado	Dominio
GeoNames	GeoNames	XML,JSON,RDF, CSV, TXT,RSS, KML	GeoNames Ontology, WGS84 Positioning	Base de datos geográfica con más de 10 millones de nombres geográficos	Geoespacial
LinkedGeoData	OpenStreetMap	RDF	LGD ontology, WGS84 Positioning	Base de datos de OpenStreetMap	
DBpedia	Wikipedia	RDF	DBpedia Ontology, WGS84 Positioning	Wikipedia	
GeoSPARQL	OGC	RDF	GeoSPARQL vocabulary	Vocabulario para la representación geoespacial de datos en RDF.	
Digital Earth (L. M. Vilches-Blázquez, Villazón-Terrazas, Corcho, & Gómez-Pérez, 2014)	Instituto Geográfico Nacional, Instituto Nacional de Estadística y Agencia Estatal de Meteorología	RDF	RDF Data Cube, hydrOntology, Time Ontology, SSN, PhenomenOntology, FAO geopolitical ontology, Geometry Ontology, Weather observation Ontology	Diversas fuentes de datos Españolas	
CUAHSI-HIS	Consortium of Universities for the Advancement of Hydrologic Science	XML	Observations Data Model Master Controlled Vocabulary	Conjuntos de datos hidrológicos de institutos que hacen parte de CUAHSI	Hidrológico
DEFRA (Department for Environment Food & Rural Affairs)	DEFRA	RDF	DEFRA bathing-water Ontology	Información asociada a calidad de agua así como al diagnóstico de la misma junto con su ubicación.	
Linked Water Data For Water Information	WATERNOMICS project	RDF	Waternomics Ontology	Datos para cuantificar el consumo y disponibilidad	

Iniciativa	Fuente de Información	Formato de Datos	Vocabulario	Conjunto de Datos Utilizado	Dominio
Management, (Curry, Degeler, Clifford, Coakley, y Costa, 2014)				para viviendas, compañías y ciudades.	
Integrated Water Resources Management (IWRM), Kämpgen, Riepl, y Klinger, 2014)	Sustainable Management of Available Water Resources - SMART Project	RDF	IWRM Ontology	Datos asociados a la Gestión Integral del Recurso Hídrico	

Capítulo 4. Planteamiento

En este capítulo se recoge una descripción del alcance del trabajo, así como de los objetivos perseguidos y el proceso general seguido para alcanzar dichos objetivos.

4.1. Alcance

Este trabajo se centra en generar y publicar Linked Data de información relacionada con aquellas propiedades fisicoquímicas y microbiológicas contempladas en la evaluación de la calidad de cuerpos hídricos, con el fin de brindar un contexto a los valores de las propiedades mencionadas para fuentes hídricas de acuerdo a la norma legal vigente en Colombia. Esto va a permitir identificar valores críticos de acuerdo a las regulaciones que correspondan según el contexto en que se desee evaluar el cuerpo de agua.

Por otro lado, este trabajo ha de proporcionar un conjunto de definiciones y vocabularios relacionados con el estado de la calidad del agua. Los conceptos que se consideran en este trabajo deben incorporar todos aquellos parámetros contemplados por la legislación colombiana utilizados para clasificar los distintos tipos de agua o los estándares que debe cumplir para asegurar la calidad de la misma de acuerdo al fin para el cual se encuentra destinado.

Así mismo, gracias a las capacidades de este trabajo para consignar información, tanto espacial como temporal, ha de permitir identificar tanto momentos como regiones donde

el aporte de contaminantes sea elevado para facilitar la gestión del recurso a la entidad que le corresponda. Sin embargo, al no tener en cuenta el monitoreo de puntos específicos de vertimiento no funciona como un mecanismo para identificar industrias aportantes de carga contaminante, aunque esta propuesta puede ser un apoyo para un desarrollo posterior que cumpla este fin.

Como resultado final se tiene un sistema que permite realizar consultas y visualizar los datos geoespaciales asociados que permiten conocer el estado de la calidad del agua en lugares y momentos específicos, así como las infracciones que se estén dando a la normal legal.

4.2. Objetivos

Objetivo general

Generar y publicar *Linked Data* geoespacial para el monitoreo de los recursos hídricos mediante evaluación de su uso potencial de acuerdo a las propiedades fisicoquímicas y microbiológicas definidas en el marco legal colombiano.

Objetivos específicos

- Desarrollar una red de ontologías que permitan incorporar el conocimiento asociado a datos espaciales para el monitoreo de la calidad hídrica en el contexto del marco legal vigente en Colombia tras el análisis y tratamiento inicial de la información.
- Transformar las observaciones asociadas a recursos hídricos en información estructurada e interoperable semánticamente.
- Publicar Linked Data sobre los recursos hídricos y su calidad ambiental en el contexto colombiano.
- Mostrar el potencial en la explotación de los datos una vez enlazados ya sea con fines estadísticos o aplicables al contexto legal.

4.3. Planteamiento general

El proceso de generación de Linked Data Geoespacial para el monitoreo de la calidad hídrica que se describe en el documento presente documento mediante un caso práctico y tal como se ha propuesto, por medio de la evaluación del uso al que puede asociarse teniendo en cuenta los límites para los parámetros fisicoquímicos y microbiológicos formulados en la legislación nacional vigente desarrollado en esta propuesta se lleva a cabo de la siguiente manera.

Para poder llevar a cabo el objetivo de este proyecto es necesario desarrollar una red de ontologías que permitan incorporar el conocimiento asociado a datos espaciales para el monitoreo de la calidad hídrica en el contexto del marco legal vigente en Colombia tras el análisis y tratamiento inicial de la información, por lo cual en la sección: Desarrollo de la red de ontologías se presentan todos los recursos utilizados en la conformación de la red, teniendo en cuenta los aspectos asociados a la muestra de agua, la legislación y el componente espacial y geográfico, así como la forma en que son desarrollados e integrados los modelos ontológicos en una única red. Teniendo en cuenta el alcance definido para el proyecto que permita definir el modelo que dé estructura a la información para cumplir la meta propuesta, tomando como insumo la información en la sección “Análisis de fuentes”.

Luego, con la red de ontologías es necesario transformar las observaciones asociadas a recursos hídricos en información estructurada e interoperable semánticamente, para lo cual en la sección Generación de RDF, se describe como para cada componente (serie histórica, legislación, unidades espaciales) la información ha sido transformada para satisfacer los requisitos del Linked Data mediante la transformación a este formato y la generación de enlaces entre fuentes de información.

Por último, es necesario publicar Linked Data sobre los recursos hídricos y su calidad ambiental en el contexto colombiano así como mostrar el potencial en la explotación de los datos una vez enlazados ya sea con fines estadísticos o aplicables al contexto legal, de manera que en la sección: Publicación y uso se muestra el proceso mediante el cual la información en RDF así como las ontologías han sido almacenadas en un repositorio

donde se permita su consulta para generar un caso de análisis mediante una aplicación aprovechando su componente espacial y legal, haciendo uso del lenguaje SPARQL (*SPARQL Protocol and RDF Query Language*), que luego puede ser aprovechado por cualquier herramienta que permita consultar datos en RDF como sucede en esta caso con el paquete estadístico R. De esta manera es posible hacer un análisis de la información generada en forma de conocimiento. Facilitando la evaluación del estado del recurso hídrico mediante el análisis espacial a partir de los valores máximos permitidos definidos a partir de la ubicación de toma de la muestra de agua, una vez se ha incorporado la legislación colombiana en el marco de la Web Semántica.

Capítulo 5. Contexto y metodología

En este capítulo se presenta el marco global del proyecto desarrollado, teniendo en cuenta sus componentes y actividades desarrolladas para la generación de Linked Data geoespacial que contribuya al monitoreo de la calidad ambiental del agua. Además, se tiene en cuenta el alcance general del proyecto y la metodología considerada para el desarrollo del mismo, que incluye actividades como: especificación, modelamiento, generación de RDF y de enlaces para su posterior publicación y aprovechamiento.

5.1. Contexto del trabajo

A continuación, se presenta el contexto del trabajo, donde se describe la zona de estudio seleccionada para el desarrollo de este trabajo y se muestran las características principales de las fuentes de datos a tratar. La importancia de conocer estas fuentes, además de brindar el contexto del estado de los datos tomados como referencia en este trabajo, radica en la forma en que dichas fuentes deben ser tratadas y transformadas para alinearse con los principios de Linked Data.

5.1.1. Zona de estudio

Este proyecto se desarrolla dentro de la cuenca del río Bogotá, la cual se encuentra en el Departamento de Cundinamarca y hace parte del grupo de corrientes de

segundo orden del mencionado Departamento, tal como se menciona en el Plan de Ordenamiento de Cuenca del Río Bogotá²².

La zona de estudio, corresponde a la cuenca del río Bogotá y en especial a su cauce principal (Figura 5-1). Esta zona limita por el Norte con el Departamento de Boyacá; en el extremo Sur con el Departamento del Tolima; al Occidente con los municipios de Bituima, Guayabal de Siquima, Albán, Sasaima, La Vega, San Francisco, Supatá y Pacho; y al Oriente con los municipios Nilo, Tibacuy, Sylvania, Chipaque, Ubaque y Choachí. Teniendo en cuenta que los datos de este trabajo corresponden a muestras tomadas en el cauce principal del río Bogotá o en puntos que se encuentran próximos a desembocar en el río Bogotá.

5.1.2. Fuentes de información

Para el desarrollo del proyecto, además de las fuentes asociadas a la legislación con los valores máximos permitidos, se tienen otras fuentes de información en distintos formatos, tales como: las coordenadas de los puntos de monitoreo que han sido consignadas en hojas de cálculo de Microsoft Excel o los valores de las pruebas fisicoquímicas y microbiológicas de las muestras de agua que han sido extraídas de un motor de base de datos *MySQL* y exportadas a hojas de cálculo de Excel.

La información tomada como referencia en el contexto de este proyecto ha sido generada por la Corporación Autónoma Regional de Cundinamarca²³ (CAR) como producto del análisis físico-químico y microbiológico de muestras de agua en el marco de su programa de monitoreo de la cuenca del río Bogotá. Este río se encuentra conformado por una red de 81 puntos de monitoreo a lo largo del cauce principal y de algunas de sus fuentes aportantes como se puede ver en la Figura 5-1.

²²

http://www.alcaldiabogota.gov.co/sisjur/adminverblobawa?tabla=T_NORMA_ARCHIVO&p_NORMFIL_ID=305&f_NORMFIL_FILE=X&inputfileext=NORMFIL_FILENAME

²³ <https://www.car.gov.co/>

La generación de la información por parte de la CAR tiene como objetivo el seguimiento de la calidad hídrica dentro de la cuenca. Por ello se realizan, siempre que sea posible, dos campañas de monitoreo al año generando en promedio dos datos por años en cada uno de los 81 puntos de la cuenca. Dicha información posee el sistema de referencia de coordenadas MAGNA SIRGAS – Bogotá (Código EPSG: 3116) y corresponde a una serie de datos por parámetro que han sido cuantificados brindando un aporte valioso del estado de la calidad del agua a lo largo del tiempo mediante un muestreo dos veces por año (para el periodo 2007-2013 considerado en este trabajo).

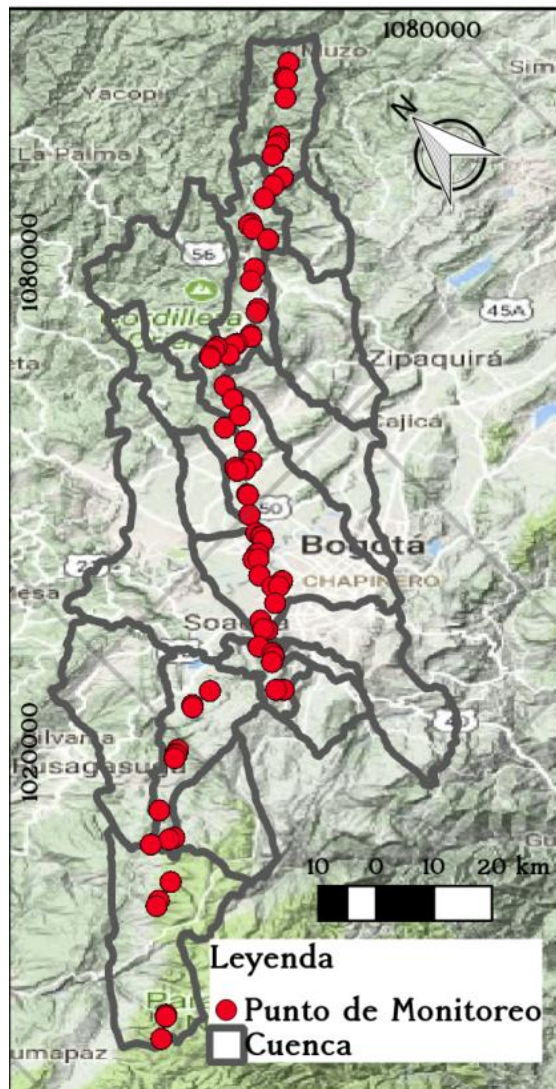


Figura 5-1. Localización de puntos de monitoreo, *Google Maps*

Todos los parámetros analizados son variados si se tiene en cuenta que describen distintas características, tanto físico-químicos como microbiológicos, y, por lo tanto, se deben identificar con distintas unidades y órdenes de magnitud. Además, aunque resulta obvio que por tratarse de distintos parámetros estos se obtienen por diferentes métodos de laboratorio (todos debidamente validados para garantizar la calidad de los datos); sin embargo en algunos casos para un mismo parámetro este puede ser calculado por distintos métodos dependiendo de las condiciones del laboratorio (según disponibilidad de reactivos o equipos de análisis) o del sitio donde este es cuantificado (puede ser analizado en campo) haciendo la base de datos más compleja teniendo en cuenta la diversidad de información consignada en los reportes de análisis de calidad de agua.

Estos reportes de laboratorio se encuentran consignados en múltiples hojas de cálculo en formato de Microsoft Excel en distintos libros y con diversos formatos, lo que hace necesaria una armonización de todos los datos facilitando su tratamiento y consulta, previos al proceso de transformación conforme a los principios de Linked Data. En la Figura 5-2 se muestra un ejemplo de las características de los datos considerados. En la mencionada figura se recogen datos históricos de muestreo y análisis de laboratorio.

A	B	C	D	E	F	I	K	L	V	Y	AC
Nº	Nombre punto	Municipio	No_muestra	Coor_este (Long Y)	Coor_norte (Latitud X)	Fecha_muest	Campaña	Q_lps	Cloruros*	Conductividad *	DBO *
1	Aguas arriba Villapinzón	VILLAPINZÓN	1334	1055836	1070203	2008-04-10T00:00:00Z	2008-1	190	1.8	13	2
2	Puente Villapinzón	VILLAPINZÓN	1335	1053523	1068988	2008-04-10T00:00:00Z	2008-1	192	5.1	31	3.1
3	Aguas arriba Q. Quincha	VILLAPINZÓN	1336	1053282	1068706	2008-04-10T00:00:00Z	2008-1	135	2.3	13	2.1
4	Q. Quincha	VILLAPINZÓN	2596	1053598	1068453	2008-07-10T00:00:00Z	2008-1	718	2.8	18.2	2
5	Estación LM Chingacio	VILLAPINZÓN	1338	1051139	1066295	2008-04-10T00:00:00Z	2008-1	479	50.1	277	25
6	Agregados Chocontá	CHOCONTÁ	1339	1045598	1062363	2008-04-10T00:00:00Z	2008-1	379	35.9	212	5.2
7	Río Tejar	CHOCONTÁ	1352	1044769	1061491	2008-04-11T00:00:00Z	2008-1	47	2.6	37.3	7.2
8	Puente Vía Telecom	CHOCONTÁ	1353	1044134	1061555	2008-04-11T00:00:00Z	2008-1	709	93.6	526	18.2
9	Descarga Mun. Chocontá	CHOCONTÁ	1354	1042676	1060785	2008-04-11T00:00:00Z	2008-1	1.4	57.4	572	46.7
10	Aguas abajo Mun Chocontá	CHOCONTÁ	1355	1042532	1060812	2008-04-11T00:00:00Z	2008-1	722	101	502	7.5
11	Estación LG Saucio	CHOCONTÁ	1356	1041097	1056901	2008-04-11T00:00:00Z	2008-1	734	88.2	526	9.9
12	Descarga embalse Sisga	CHOCONTÁ	1361	1038962	1057069	2008-04-14T00:00:00Z	2008-1	1500	3.2	25.9	3.1
13	Estación LM Santa Rosita	SUESCA	1362	1036255	1056612	2008-04-14T00:00:00Z	2008-1		20.7	124	4.9
14	Puente Santander	SUESCA	1363	1031591	1055100	2008-04-14T00:00:00Z	2008-1		13.7	80.9	<L.C.
15	Descarga Municipio Suesca	SUESCA	1357	1031102	1055197	2008-04-14T00:00:00Z	2008-1	12.2	72.9	621	97.5
16	Aguas abajo Mun Suesca	SUESCA	1358	1031155	1054269	2008-04-14T00:00:00Z	2008-1	3210	14.5	83.9	2
17	Descarga Embalse Tominé	SESQUILÉ	1359	1031692	1051077	2008-04-14T00:00:00Z	2008-1	6000	5.8	43.9	2
18	Aguas arriba descarga Papeles y Molinos	SESQUILÉ	1360	1026455	1049171	2008-04-14T00:00:00Z	2008-1	9735	12.6	113	<L.C.
19	Estación LG - Pte Florencia	GACHANCIPÁ	1365	1024543	1048112	2008-04-14T00:00:00Z	2008-1		11.6	70	<L.C.
20	Descarga Municipio Gachancipá	GACHANCIPÁ	1366	1021962	1043839	2008-04-14T00:00:00Z	2008-1	43.1	73.4	672	38.2
21	Aguas abajo Mun Gachancipá	GACHANCIPÁ	1367	1021344	1043552	2008-04-14T00:00:00Z	2008-1	8618	11.2	72	2.3
22	Estación LM - Tocancipá	TOCANCIPÁ	1368	1017904	1041499	2008-04-15T00:00:00Z	2008-1	7038	10.7	70.7	<L.C.
23	Descarga Mun Tocancipá	TOCANCIPÁ	1369	1017744	1041206	2008-04-15T00:00:00Z	2008-1	13	119.2	940	47.5
24	Aguas arriba Termozipa	TOCANCIPÁ	1370	1015132	1042109	2008-04-15T00:00:00Z	2008-1		14.9	137	3.3

Figura 5-2. Histórico de muestreo y análisis de laboratorio (vista en formato Excel)

Asimismo, en los datos considerados junto con la presentación de la cuantificación de los parámetros físico-químicos y microbiológicos, también se suministra información asociada a la muestra recogiendo el nombre del punto de muestra, sus coordenadas, el

municipio al que pertenece, fecha de toma de muestra, y se brindan metadatos relevantes para el análisis de los resultados obtenidos.

5.2. Metodología

Como se mencionaba en la sección Marco de Referencia de este trabajo, Linked Data hace referencia al conjunto de buenas prácticas que permiten publicar y conectar datos estructurados en la Web (Bizer et al., 2009), haciendo uso de la arquitectura de la web y de especificaciones del Consorcio de la Web (W3C), como es el caso del modelo de datos RDF. Esta iniciativa de datos enlazados (Linked Data) se basa en la aplicación de unos principios básicos para publicar y vincular datos estructurados en la Web Semántica. Estos principios de diseño fueron introducidos por Tim Berners-Lee²⁴ y son los que se exponen a continuación:

- Usar URI (*Uniform Resource Identifier*) para identificar los recursos.
- Usar HTTP URIs para que los interesados puedan buscar estos recursos.
- Proporcionar información útil, usando los estándares (RDF, SPARQL), cuando se busque una URI.
- Incluir enlaces a otras URI, de manera que sea posible descubrir más cosas en esta información.

Para la persecución de los principios mencionados y, en consecuencia, adoptar las recomendaciones y buenas prácticas del W3C y OGC para la publicación de datos geoespaciales en la Web (*Spatial Data on the Web Best Practices*²⁵) se adopta en este trabajo la metodología propuesta en (Vilches-Blázquez, et al., 2014), que a su vez toma como referencia el trabajo descrito en (Villazón-Terrazas, et al., 2011) y cuya secuencia de actividades se observa en la Figura 5-3. Además, es importante tener en cuenta que para el desarrollo de la red de ontologías propuesta en este trabajo también se considera la metodología NeOn (Suárez-Figueroa, Gómez-Pérez, Motta, y Gangemi, 2012).

²⁴ <https://www.w3.org/DesignIssues/LinkedData.html>

²⁵ <https://www.w3.org/TR/sdw-bp/>

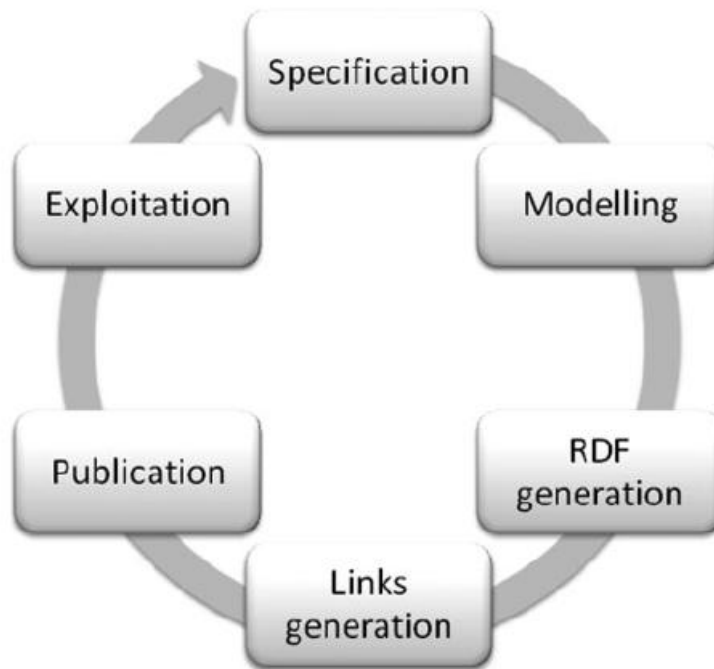


Figura 5-3. Actividades de la metodología para la generación y publicación de Linked Data (Vilches-Blázquez et al., 2014)

La metodología para la generación y publicación de Linked Data adoptada propone un ciclo de vida iterativo e incremental que consta de 6 fases (Figura 5-3):

1. Especificación
2. Modelado
3. Generación de RDF
4. Generación de enlaces
5. Publicación
6. Explotación

Cada uno de estos pasos de la metodología es dependiente del otro para poder avanzar, por lo tanto, queda patente la importancia en el cumplimiento de cada uno de los mismos. En el contexto de este trabajo las metodologías adoptadas se aplican a través de una serie de actividades desagregadas, reflejadas en el diagrama de flujo de la Figura 5-4, donde se parten de recursos ontológicos y no ontológicos como registros en archivos de Microsoft Excel y documentos normativos entre otros, con el fin de generar información

conforme a los principios de Linked Data que mediante el desarrollo de consultas semánticas permita realizar análisis expertos sobre la calidad del agua en la zona de estudio seleccionada.

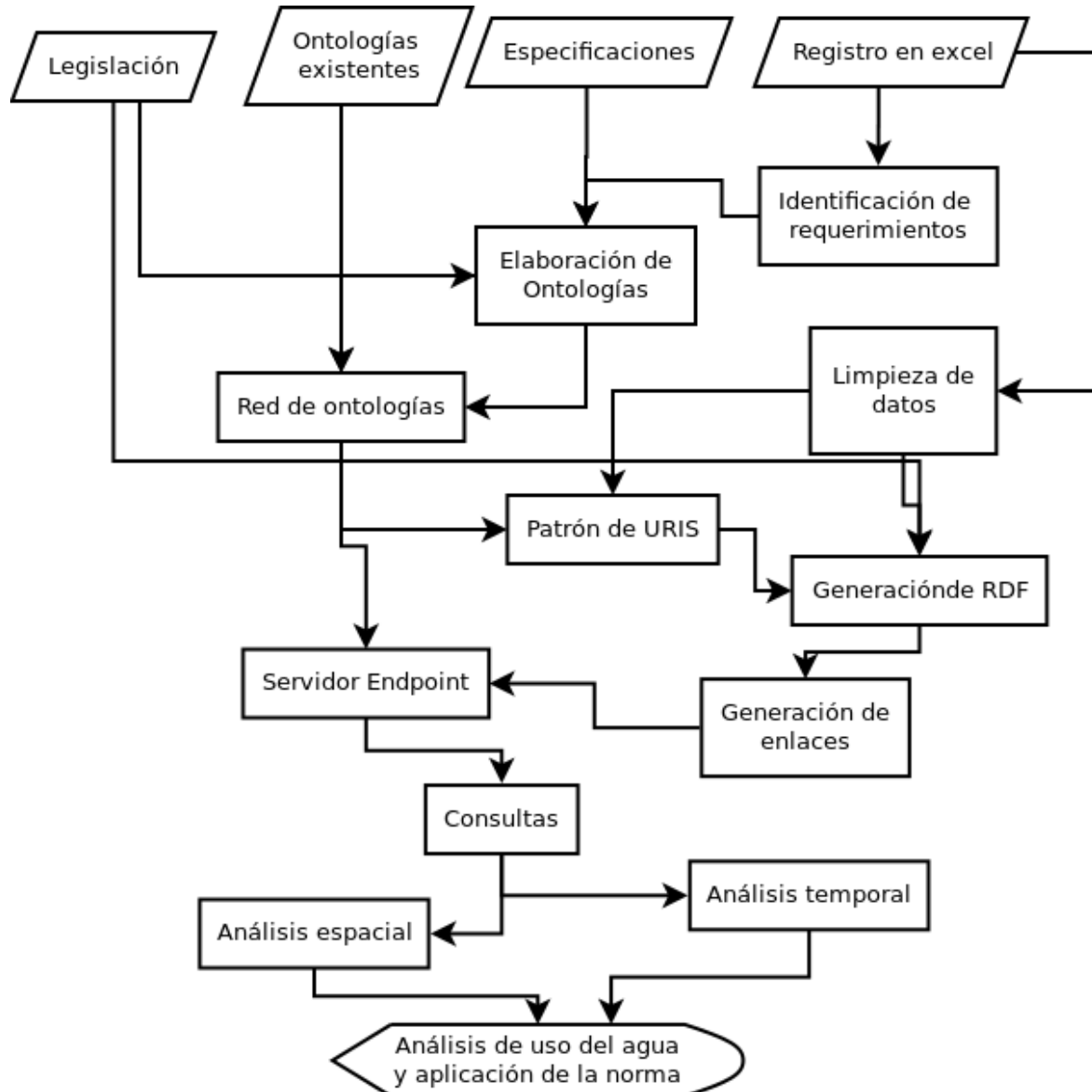


Figura 5-4. Diagrama de flujo de proceso desarrollado

Teniendo en cuenta esta metodología, a continuación, se describe, de manera breve, la forma en que cada actividad será desarrollada dentro del proyecto.

5.2.1. Especificación

Conforme a la metodología, esta actividad se centra en todo el proceso de levantamiento de información, donde se recopilan los datos con los que se va a trabajar, se especifican los requerimientos y se diseñan las URIs.

Identificación de fuentes de datos

Una vez se ha delimitado el alcance del trabajo a realizar se recopila toda la información que servirá de fuente de datos a utilizar en el desarrollo del trabajo. Esto incluye datos de todo tipo, aunque, principalmente, predominan las fuentes semi-estructuradas ya sea en formatos propietarios o libres (ficheros de Microsoft Excel y bases de datos espaciales). Además, se consideran fuentes de datos no estructurados (como, por ejemplo, el uso de documentos escaneados que contienen información no diligenciada en las bases de datos o que hayan sido alteradas por alguna fuente de ruido).

Junto con este proceso de identificación y recopilación de fuentes de información, se lleva a cabo un proceso de depuración de la información para identificar y depurar registros atípicos para su posterior tratamiento. Para ello se utilizarán herramientas especializadas que contribuyan a facilitar dicho proceso.

Una vez identificados, recopilados y depurados todos los recursos de información con los que se va a trabajar, es necesario identificarlos de manera unívoca para evitar cualquier tipo de ambigüedad, garantizando el uso de vocabularios consensuados. Con este fin, se ha de generar un patrón de nombrado (URI) para los datos del proyecto que serán transformados con posterioridad a RDF.

Generación del patrón de URI

Teniendo en cuenta los principios del Linked Data, es necesario tener una identificación apropiada de los recursos mediante URIs. Para ello, en el desarrollo de este trabajo se diseñarán unos patrones de nombrado (URI) para que los datos

considerados en este trabajo tengan asociados principios de claridad y que puedan ser nombrados de manera unívoca.

5.2.2. Modelamiento

Esta parte comprende el modelado del dominio que atañe al proyecto conforme a los requerimientos de la Web Semántica. Para abordar este trabajo se tomarán los lineamientos de la metodología NeOn (M. Suárez-Figueroa, 2010a), la cual se encuentra conformada por nueve escenarios que buscan la construcción de ontologías y redes de ontologías mediante la reutilización de recursos ontológicos y no ontológicos, la reingeniería y la fusión.

Para el desarrollo de esta actividad (modelamiento) se ha realizado una búsqueda de vocabularios (ontologías) sobre recursos hídricos y su calidad que puedan ser de utilidad para el desarrollo del trabajo. No obstante, este trabajo no solo incluye la recopilación de material ontológico, también se utilizan recursos de tipo no ontológico asociados con el dominio de trabajo para enriquecer el modelado, tales como documentos en formato PDF, hojas de cálculo de Microsoft Excel y todo recurso que no corresponde con una ontología propiamente dicha. En último lugar, cuando ha sido necesario, se han generado las ontologías necesarias para aquellos casos donde el material ontológico existente no ha podido satisfacer los requisitos del proyecto.

Mediante el uso de la metodología NeOn y las estrategias ya mencionadas se crea una red de ontologías de dominio que describe de forma adecuada el campo de los recursos hídricos, de acuerdo a la caracterización requerida para la evaluación de la calidad del cuerpo de agua en cuestión. Asimismo, esta red de ontologías incorpora el marco legal vigente a aplicar, que a su vez se encuentra sujeto a una tabla de reglas. Esta tabla contiene el conjunto de valores máximos y mínimos de las propiedades hídricas registradas para hacer posible su evaluación.

5.2.3. Generación de RDF

Como se mencionó con anterioridad, la información recopilada se encuentra en diversos formatos como hojas de cálculo, archivos con valores separados por coma, *shapefile*, etc. Por tanto, resulta necesario estructurar y armonizar dichas fuentes

conforme a los principios de *Linked Data*. Es por este motivo que toda la información considerada se transforma a RDF.

La generación de datos en formato RDF, como se ha mostrado en (Omitola et al., 2010) y (L. M. Vilches-Blázquez et al., 2013), permite la armonización de los datos objeto de publicación que proceden de distintas fuentes de datos (*shapefiles*, bases de datos, etc.), evitando hacer uso de formatos propietarios, y permitiendo así cumplir con los requisitos del *Linked Data*. RDF es uno de los estándares donde la información ha sido creada para estar disponible, de acuerdo a los principios del *Linked Data*, la razón de esto se debe a las ventajas que ofrece como, por ejemplo: proporcionar un esquema extensible, URIs diferenciados, y permitir la unión apropiada (enlaces) de diferentes fuentes de datos a través de la generación de enlaces RDF.

Para este proceso se trabaja con diferentes herramientas presentes en la comunidad, por ejemplo, Open-Refine²⁶ y su extensión para la generación de RDF, conocida como RDF-Refine²⁷. Esta herramienta ha sido diseñada para trabajar con datos desordenados, permitiendo limpiarlos y transformarlos a otros formatos extendiéndolos a través de servicios web y datos externos.

5.2.4. Generación de enlaces

Teniendo en cuenta el cuarto principio del *Linked Data* asociado con la *generación de enlaces a otras URIs* y que existe un incremento en la publicación de datos como tripletas RDF enlazadas a otras fuentes de datos al establecer equivalencias entre recursos de distinto origen (Feliachi et al., 2013), en esta actividad se lleva a cabo un proceso de enriquecimiento y conexión de los datos RDF transformados con otros conjuntos de datos ya publicados en la Web de *Linked Data*. De esta manera, se muestra que el valor de los datos y su utilidad se hace mayor en la medida en que estos se encuentren conectados con otros datos (Bizer et al., 2009).

²⁶ <http://openrefine.org/>

²⁷ <http://refine.deri.ie/>

5.2.5. Publicación

Este paso consiste en la publicación de la información por medio de un *triple store*, es decir, una base de datos para almacenamiento de tripletas RDF. El despliegue de este *triple store* permite que haya un punto de consulta (SPARQL *Endpoint*) de las tripletas RDF generadas en el contexto de este trabajo y que puedan ser consultadas a través del lenguaje de consulta SPARQL.

En el momento de elegir el *triple store* que se despliega en este trabajo se tiene en cuenta su capacidad para realizar consultas espaciales y, en especial, si es capaz de trabajar con el estándar *GeoSPARQL* del *Open Geospatial Consortium* (OGC).

5.2.6. Explotación

Antes de hacer uso de la información publicada se importa la información desde el SPARQL *Endpoint* al lenguaje estadístico en R (R Development Core Team, 2016). Esto se realiza utilizando la librería SPARQL (van Hage, 2013a) que permite consultar la información mediante lenguaje de consulta SPARQL para su análisis tanto estadístico como espacial y normativo.

Asimismo, junto al trabajo de análisis, en esta actividad de explotación se visualizan los datos resultantes a través de una aplicación que permite la generación de mapas web dinámicos, a partir del uso de las librerías *Leaflet* (Cheng, Xie, y Wickham, 2016) y *shiny* (Chang, Cheng, Allaire, Xie, y McPherso, 2016).

Capítulo 6. Resultados

En este capítulo se presentan los resultados obtenidos, discriminados a partir de cada uno de las actividades recogidas en las metodologías adoptadas para el desarrollo de este trabajo.

6.1. Establecimiento de patrones de nombrado

Para permitir la integración de los datos como Linked Data es necesaria la identificación de los mismos mediante URIs (*Universal Resource Identifiers*), las cuales son los identificadores, también conocidos como “direcciones” o “nombres” dentro de la Web, concebida como un espacio universal de información (Berners-Lee, 1996) y sirven para establecer la identidad de los recursos incorporados. Este establecimiento de patrones de nombrado mediante URIs se relaciona con la actividad de Especificación de la metodología de Linked Data seleccionada para el desarrollo de este trabajo.

Así, para la construcción de los patrones de nombrado (URIs) se han establecido dos clases:

- Nombres para términos hidrológicos: <http://WaterML.org>
- Nombres para términos Legales del agua: <http://www.waterregulations.gov.co>

El prefijo sugerido para WaterML: *wml*

El prefijo sugerido para términos Legales: *wr*

En el patrón diseñado se consideran diferentes caracteres para la separación de sus elementos. Así para las clases son definidas utilizando el carácter “ / ”, mientras que las propiedades se han definido con “ # ”. Por ejemplo, para la clase “*Clase I*” y la propiedad “*Límite de Nitratos*” los patrones de URI establecidos quedan de la siguiente manera:

- <http://www.waterregulations.gov.co/WaterClass/ClassI>
- <http://www.waterregulations.gov.co#hasOxygenDissolvedLimit>

Las URIs de tipo *wml* corresponden a información asociada a los recursos asociados a la muestra y sus características como los parámetros, lugares de toma, valores cuantificados y fecha entre otros. Así, por ejemplo, las clases definidas como la que se muestra a continuación definen para los puntos de muestreo:

- <http://WaterML.org/Geometry/Point/1>

Merece resaltarse que dentro de la URI se incluye la jerarquía de clases, es decir, en este caso el punto de muestreo “1”, corresponde a un objeto de la subclase *Point* y esta, a su vez, hace parte de la clase *Geometry*.

Por otro lado, las URIs de tipo <http://www.waterregulations.gov.co> definen aquellos componentes relacionados con la legislación y la norma aplicable a calidad del agua para este caso, donde al igual que en el caso anterior incluyen clases y subclases dentro de su estructura. En todos los casos y tal como se mencionó anteriormente a las propiedades les antecede el símbolo “ # ”, siguiendo la misma estructura en caso de hacer sub-propiedades, siempre siendo lo más auto-explicativa como en el siguiente ejemplo cuya propiedad asigna un límite de agua a una entidad.

- <http://www.waterregulations.gov.co#hasWaterLimit>

6.2. Desarrollo de la red de ontologías

A continuación, se presenta el modelo semántico que formaliza el dominio hidrológico asociado a calidad hídrica, así como su contexto legal y su componente espacial que permite dar una estructura pertinente a los datos considerados en este trabajo.

6.2.1. NeOn methodology

La metodología NeOn (Gómez-Pérez y Suárez-Figueroa, 2009), adoptada para el desarrollo de la red de ontologías de este trabajo, ofrece guías para el desarrollo de redes de ontologías como también para el desarrollo de ontologías individuales. Esta metodología está basada en escenarios y se apoya en los aspectos de colaboración de desarrollo de ontologías y la reutilización, así como en la evolución dinámica de las redes de ontologías en entornos distribuidos.

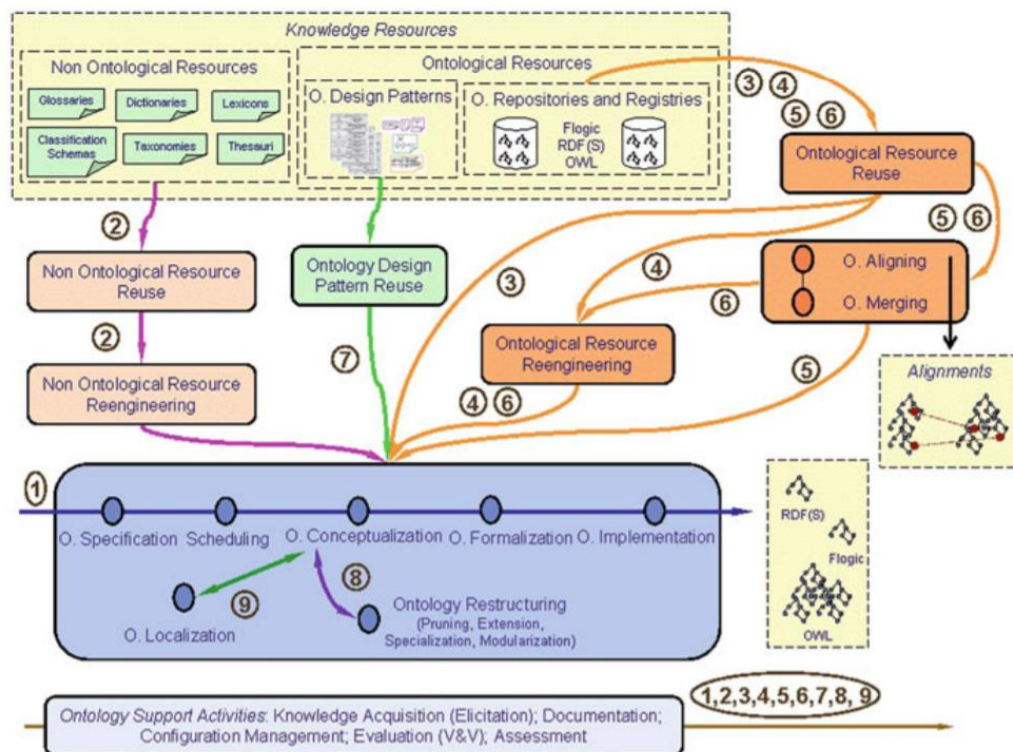


Figura 6-1. Ciclo de vida NeOn

Los escenarios propuestos por esta metodología, ilustrados en la Figura 6-1, para el desarrollo de ontologías y redes de ontologías son los siguientes:

- Escenario 1. Desde la especificación a la implementación: Para iniciar la construcción de una ontología es necesario especificar los requisitos; luego, desarrollar la búsqueda de los posibles recursos para reutilizar; y, proceder con la planificación del trabajo.
- Escenario 2. La reutilización y reingeniería de los recursos no ontológicos (NOR): En este escenario los desarrolladores deben ejecutar el proceso de reutilización de recursos no ontológicos, es decir, cualquier recurso de información que no sea una ontología, determinando los NORs a reutilizar para realizar la reingeniería.
- Escenario 3. La reutilización de los recursos ontológicos. Este escenario permite determinar los recursos ontológicos o declaraciones ontológicas que van a ser reutilizadas en el desarrollo de una red de ontologías.
- Escenario 4. La reutilización y reingeniería de los recursos ontológicos. Este escenario se centra en reutilizar y reorganizar los recursos ontológicos.
- Escenario 5. La reutilización y la fusión de los recursos ontológicos: En este escenario se aborda la unión de los recursos a reutilizar con nuevos recursos que se necesitan crear.
- Escenario 6. Reutilización, la fusión y reingeniería de los recursos ontológicos: En este escenario se realiza la reingeniería de los recursos ontológicos a reutilizar.
- Escenario 7. Reutilización de los patrones de diseño de ontologías (ODPs): Este escenario permite realizar el acceso y reutilización de patrones de diseño de ontologías para adoptar las buenas prácticas de modelado recomendadas por la comunidad.
- Escenario 8. Reestructuración de recursos ontológicos: En este escenario se reestructuran los recursos ontológicos para luego adaptarlos a la red de ontologías.
- Escenario 9. Localización de recursos ontológicos: Este escenario adapta los recursos ontológicos a otros idiomas, con el fin de obtener una ontología multilingüe.

La metodología NeOn detecta importantes limitaciones entre las metodologías más utilizadas para el desarrollo de ontologías, tales como: *METHONTOLOGY* (Fernández-López, Gómez-Pérez, y Juristo, 1997), *On-To-Knowledge* (Staab, Studer, Schnurr, y Sure, 2001) y *DILIGENT* (Pinto, Pinto, Staab, y Tempich, 2004). Estas carencias se centran en que ninguna de ellas trata simultáneamente el concepto de red de ontologías ni las dimensiones de colaboración, contexto y dinamismo. Además, estas no proporcionan guías metodológicas detalladas para la realización de los distintos procesos

y actividades involucrados en el desarrollo de ontologías, tampoco para la reutilización y la reingeniería de fuentes de conocimiento existentes que ya han alcanzado cierto grado de consenso en una determinada comunidad (Poveda, 2010). Por todo lo anterior, se considera pertinente la utilización de la metodología NeOn para el desarrollo de la red de ontologías de este trabajo.

6.2.2. Combinación de escenarios para el desarrollo de la red de ontologías

Uno de los elementos principales del presente proyecto consiste en el desarrollo de una red de ontologías que modele la información relacionada con los parámetros físico-químicos y microbiológicos del agua, y que considere factores que permitan contextualizar dicha información teniendo en cuenta elementos como los lugares de toma, unidad hidrológica y método por el cual se ha realizado la cuantificación entre otros aspectos. Adicionalmente, esta red de ontologías debe agregar un componente legislativo que no solo permita evaluar el grado de cumplimiento, sino también los posibles usos para los cuales pueda ser destinada el agua a evaluar de manera que permita a entidades competentes y tomadores de decisiones actuar en consecuencia conforme al recurso hídrico y la calidad del mismo teniendo en cuenta el lugar, ya que este permite definir el conjunto de normas que aplican a cada muestra en el tiempo.

El desarrollo de esta red de ontologías sigue el esquema metodológico de acuerdo a la metodología NeOn (Suárez-Figueroa, 2010b), que como se describía con anterioridad, se encuentra conformada por una serie de escenarios. De los escenarios propuestos en la mencionada metodología, para el desarrollo de la red de ontologías de este trabajo se adoptan aquellos que se describen a continuación:

- **Escenario 2: Reutilización y reingeniería de recursos no ontológicos.** Se acude a recursos no ontológicos como las hojas de cálculo que contienen la información asociada a la calidad hídrica. De igual manera, se incorporaron otros recursos como diccionarios para conformar el listado de parámetros físico-

químicos y microbiológicos definidos en la propuesta de *CUAHSI Hydrologic Information System (HIS)*²⁸.

- **Escenario 3: Reutilización de recursos ontológicos.** Se toma la ontología desarrollada para el estándar *Observations & Measurements (O&M)*²⁹, que brinda el punto de partida para crear una ontología para el estándar de OGC *WaterML 2.0*.³⁰ Así mismo se ha hecho uso de la ontología creada para el sistema *INWATERSENSE*³¹ (*INWS*), que a su vez ha tomado como base la propuesta del *Semantic Sensor Network*³² (*SSN*) para la creación de su ontología.
- **Escenario 4: Reutilización y re-ingeniería de recursos ontológicos.** Aunque se van a reutilizar las ontologías mencionadas en el caso anterior, no siempre se ajustan de manera adecuada a las necesidades del proyecto. Por lo tanto, en algunos casos es necesaria su modificación, es decir, la aplicación de procesos de reingeniería de los recursos ontológicos. Un ejemplo de esta reingeniería se da durante el uso del *INWS*, ya que a pesar tener un enfoque hidrológico asociado a la calidad hídrica posee un modelo asociado a sensores que no cumple las necesidades requeridas por el proyecto cuyos muestreos son manuales.
- **Escenario 5: Reutilización y fusión de recursos ontológicos.** La red de ontologías desarrollada en el contexto de este trabajo está conformada por la unión de varias ontologías. El caso más claro de esto es el vocabulario RDF *DataCube*³³, que ha sido incorporado a la red de ontologías para mejorar la forma en que la información puede ser presentada y analizada.
- **Escenario 6: Reutilización, fusión y re-ingeniería de recursos ontológicos.** No solo se han reutilizado diversas ontologías como base y se han fusionado para establecer relaciones que den lugar a una auténtica red de ontologías, sino que

²⁸ <http://hiscentral.cuahsi.org/>

²⁹ <http://www.opengeospatial.org/standards/om>

³⁰ <http://www.opengeospatial.org/standards/waterml>

³¹ <http://ceur-ws.org/Vol-1063/paper3.pdf>

^{32,32} <https://www.w3.org/TR/vocab-ssn/>

³³ <https://www.w3.org/TR/vocab-data-cube/>

también ha sido necesaria su modificación como sucede con la propuesta de O&M, que a pesar de ser tomada como base para la generación del *WaterML*, se ha modificado en clases como la geoespacial a la cual se le ha incorporado el estándar de OGC *GeoSPARQL*³⁴ permitiendo el enriquecimiento del modelo de datos.

- **Escenario 9: Localización de recursos ontológicos.** No se ha tenido en cuenta este escenario ya que el modelo de datos al igual que los recursos utilizados utilizan los mismos términos que se presentan en el idioma inglés.

Los detalles del trabajo asociado a cada escenario se recogen en las siguientes secciones de este capítulo.

6.2.3. Especificación de requerimientos

La red de ontologías a desarrollar en este trabajo, además de la integración de datos asociados a calidad hídrica de fuentes heterogéneas, debe permitir su análisis teniendo en cuenta el contexto espacio-temporal y legal facilitando la interpretación de usuarios que no necesariamente serán expertos en el tema.

Esta meta, así como las especificaciones de la red de ontologías se encuentran descritas mediante la plantilla presentada en el documento de especificación de requisitos de la ontología (*Ontology Requirements Specification Document - ORSD*) (Suárez-Figueroa, Gómez-Pérez, & Villazón-Terrazas, 2008), propuesto por la metodología NeOn, y que se encuentra conformado por los siguientes campos:

- Propósito: Define la función principal o el papel que ha de desempeñar.
- Alcance: Recoge el grado de cobertura y detalle.
- Lenguaje de implementación: Define el lenguaje formal que ha de utilizarse.
- Usuarios previstos: Presenta los usuarios a los que será destinada.
- Usos previstos: Define el uso para el que se ha definido.

³⁴ <http://www.opengeospatial.org/standards/geosparql>

- Preguntas de competencia: Recoge el contenido específico que debe satisfacer en forma de preguntas agrupadas por competencia.

6.2.3.1. **Propósito**

El propósito de la red de ontologías es permitir la interpretación de datos históricos con ubicación geográfica asociados a muestras de agua y sus respectivos análisis físico-químicos y microbiológicos teniendo en cuenta su contexto legal y los máximos niveles permitidos por la legislación vigente de acuerdo a su ubicación y uso. Para ello siempre se tendrá en cuenta el contexto de la muestra al incorporar información complementaria como unidades de medida, fecha o cuenca a la que pertenece.

6.2.3.2. **Alcance**

La red de ontologías para la publicación de Linked Data geoespacial asociada a calidad hídrica incluye los siguientes dominios:

- **Toma de muestra de agua:** Incluye información asociada a la muestra, como: mecanismo de captura, fecha, naturaleza de la muestra (natural, residual, etc.), identificación de la muestra y sus metadatos asociados.
- **Análisis de muestra de agua:** Se encuentra asociado a la información producto de los análisis de laboratorio como la cuantificación del analito y el sistema de unidades.
- **Punto de muestreo:** Ubica geográficamente el sitio al que pertenece la muestra y lo caracteriza teniendo en cuenta información como el municipio y las coordenadas del mismo.
- **Legislación:** Tiene en cuenta los límites máximos permitidos para cada parámetro, su ubicación geográfica y el uso de agua asociado.

6.2.3.3. **Lenguaje de implementación**

Teniendo en cuenta que se trata de un lenguaje de la Web Semántica creado para representar conocimiento complejo de manera enriquecida y hace parte de la

W3C³⁵, se ha decidido trabajar con el lenguaje OWL (*Web Ontology Language*). Además, también se trabaja con la sintaxis *RDF/XML* para permitir el buen funcionamiento de la red de ontologías a desarrollar con el repositorio de datos.

6.2.3.4. **Usuarios previstos**

Está dirigida principalmente a personas que por fines académicos o profesionales requieran de información asociada a calidad del agua, los cuales pueden o no ser expertos en el tema.

6.2.3.5. **Usos previstos**

La red de ontologías permite a los usuarios la consulta de información para el análisis de datos asociados a calidad del agua, así como su clasificación para uso y el grado de cumplimiento de la legislación vigente.

6.2.3.6. **Preguntas de competencia**

Esta red de ontologías debe encontrarse en capacidad de responder preguntas como las que se formulan a continuación agrupadas por tema:

Calidad hídrica

- ¿En dónde se da el mayor tipo de contaminación por parámetro fisicoquímico o microbiológico?
- ¿Cuál es la época del año donde se dan los mayores valores de contaminación?
- ¿Qué año ha presentado el mayor grado de contaminación?
- ¿Lugar y año en que la calidad del agua ha presentado los mayores valores de contaminación?
- ¿A partir de dónde se da una subida o caída de oxígeno disuelto?
- ¿Qué lugares son propensos a contaminación por contenido de sólidos?
- ¿Qué regiones son las que más aportan metales pesados al río a lo largo de su trayecto?

³⁵ <https://www.w3.org/OWL/>

Asociadas a la muestra

- ¿Quién tomó la muestra?
- ¿Dónde fue tomada la muestra?
- ¿Día y hora de la toma?
- ¿En qué cuencas se da la mayor frecuencia de monitoreo?
- ¿Qué cuencas o municipios poseen mayor número de monitoreos en calidad hídrica?
- ¿Qué muestras han sido tomadas en sistemas lénticos, como lagunas embalses o humedales?

Legislación

- ¿En qué lugar y momento se está incumpliendo la norma?
- ¿Qué parámetros son los que más se incumplen de acuerdo a la legislación?
- ¿Qué municipios poseen fuentes de agua aptas para algún consumo específico (por ejemplo, humano, agrícola, etc.)?
- ¿Qué ríos exceden las concentraciones máximas permitidas de plomo?
- ¿Qué lagunas, embalses o humedales incumplen la norma legal vigente?

Por otro lado, la red de ontologías a desarrollar debe servir a cualquier entidad relacionada con el estado de los recursos naturales que procure la conservación o realizar un inventario del estado de los mismos mediante la generación y ejecución de políticas que propendan el desarrollo sostenible. De igual manera, debe servir a cualquier institución o persona involucrada con el estudio del agua y su estado.

6.2.3.7. Reutilización de recursos ontológicos

La reutilización de recursos ontológicos se refiere al uso de parte o totalidad de ontologías ya existentes para el desarrollo de la nueva ontología. De esta manera, para aprovechar recursos ontológicos existentes en el proceso de desarrollo, por ejemplo, se reutilizan ontologías asociadas a monitoreo hídricos, como la generada para la iniciativa O&M (*Observations and Measurements*). También se hace uso de la ontología de la SSN (*Semantic Sensor Network*) en la descripción de sensores y observaciones, lo cual brinda un aporte valioso en el enriquecimiento de la descripción de los monitoreos.

A continuación se recoge una breve descripción de los recursos ontológicos mencionados con anterioridad:

- **om-lite y sam-lite:** Son dos ontologías OWL para observaciones y atributos de muestra basados en el modelo conceptual O&M, descrito en la norma ISO 19156:201136. Estas ontologías (*om-lite* y *sam-lite*) han sido formulada con el fin de proporcionar un vocabulario consensuado para las observaciones y sus propiedades (Cox, 2013).
- **Semantic Sensor Network (SSN):** Esta ontología fue desarrollada por la W3C *Semantic Sensor Network Incubator Group*³⁷ (SSN-XG) (W3C Semantic Sensor Network Incubator Group, 2009). SSN describe sensores y observaciones, así como conceptos relacionados. Sin embargo, no incluye conceptos de dominio como tiempo y ubicación, los cuales deben ser incorporados a la red que se desarrolla en este trabajo desde otros recursos. La Figura 6-2 muestra una representación de los principales elementos que componente la ontología de SSN.

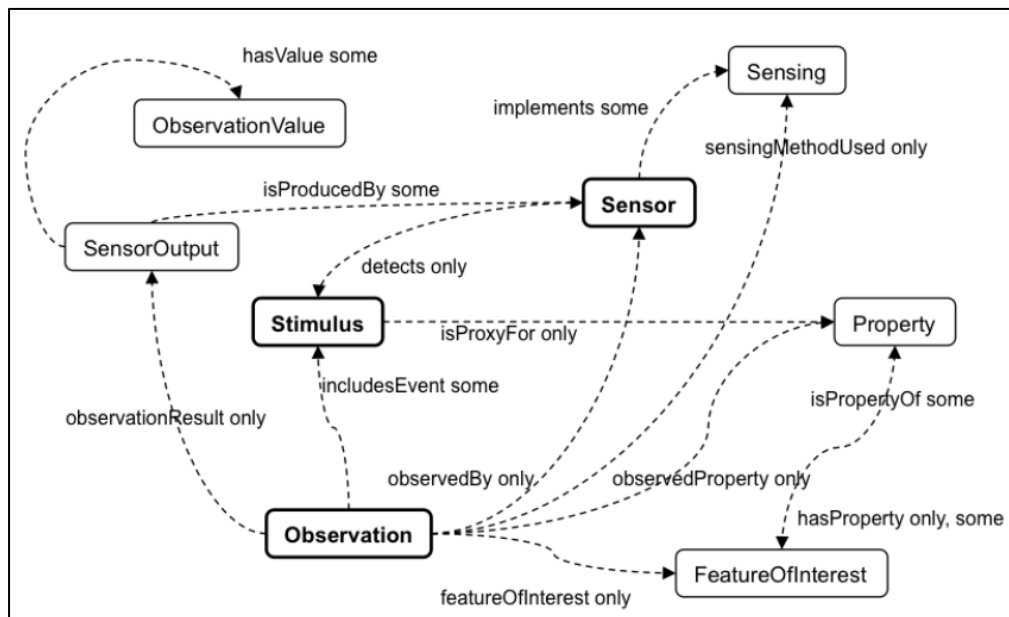


Figura 6-2. Grafo Semantic Sensor Network, (Compton et al., 2005)

³⁶ <https://www.iso.org/standard/32574.html>

³⁷ <https://www.w3.org/2005/Incubator/ssn/>

- **GeoSPARQL:** La ontología GeoSPARQL³⁸, propuesta como estándar por el Open Geospatial Consortium (OGC), permite y facilita la representación y consulta de datos geoespaciales en la Web Semántica. Esta propuesta define un vocabulario para representar datos geoespaciales en RDF y además proporciona una extensión al sistema de consulta SPARQL para el procesamiento de los datos geoespaciales.
- **INWATERSENSE – INWS:** Esta ontología (Ahmedi, Jajaga, y Ahmedi, 2013) está basada en la propuesta de SSN, pero se extiende para dar soporte en la clasificación del agua de acuerdo a las autoridades reguladoras como la Directiva Europea Marco del Agua.

En la reutilización de recursos ontológicos se ha tomado como referencia el modelo propuesto por INWS, ya que no solo utiliza la propuesta estándar de SSN, sino que también le aporta un uso asociado al monitoreo de calidad hídrica haciéndolo una excelente opción para el caso a utilizar en el presente trabajo.

6.2.4. Reutilización y reingeniería de recursos ontológicos

Mediante la reutilización y reingeniería de recursos ontológicos se lleva a cabo el desarrollo de la red de ontologías de manera eficiente al aprovechar modelos ontológicos ya existentes, no solo para evitar la inversión de esfuerzo en desarrollos ya existentes, también al reutilizar estos recursos se facilita la generación de un mismo lenguaje al utilizar una misma estructura con vocabularios consensuados.

A continuación, se presentan las modificaciones realizadas en los recursos ontológicos considerados, las cuales, principalmente, se encuentran asociadas al enriquecimiento de las ontologías y la formulación de equivalencias entre clases.

³⁸ <http://www.opengeospatial.org/standards/geosparql>

6.2.4.1. Modificaciones realizadas

Dentro de las modificaciones realizadas está la reasignación de jerarquías a clases ya existentes que permiten dar más detalle al modelo de datos o adaptarlas a las necesidades del proyecto. Un ejemplo de ello se presenta en la Figura 6-3, donde al asignar la subclase *Observation* de la ontología de O&M de *Feature Type* a *GeographicFeature* se enriquece al incorporar el contexto geográfico, ya que como subclase hereda todas las propiedades este contexto y resulta a fin con lo formulado en el estándar WaterML 2.0 de OGC.

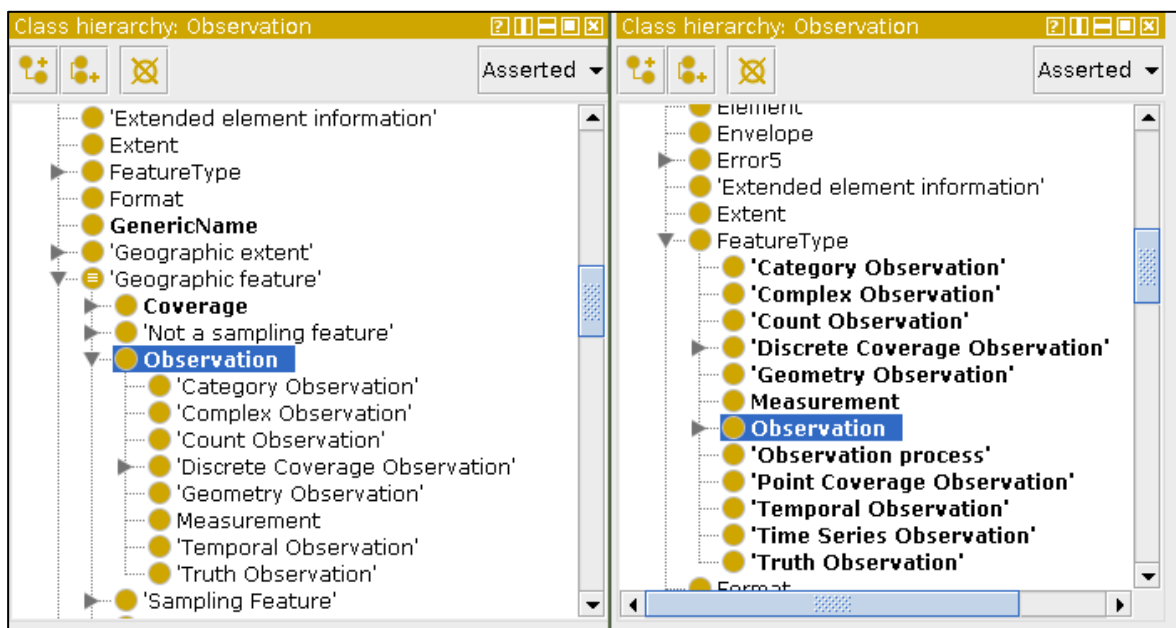


Figura 6-3. Jerarquía clase “*Observation*” (Izquierda: WaterML 2.0, Derecha: O&M)

También se han enriquecido las ontologías existentes mediante la generación de propiedades o nuevas subclases como la que se presenta en la Figura 6-4, donde se puede apreciar en UML (*Unified Modeling Language*) en color verde las clases del O&M y en azul la subclase *Timeseries* del WaterML 2.0. Esta última subclase al no encontrarse en la especificación y, por lo tanto, tampoco en la ontología ha sido incluida para cumplir con lo estipulado en el documento del estándar WaterML de OGC, añadiendo un mayor grado de especificidad en las clases.

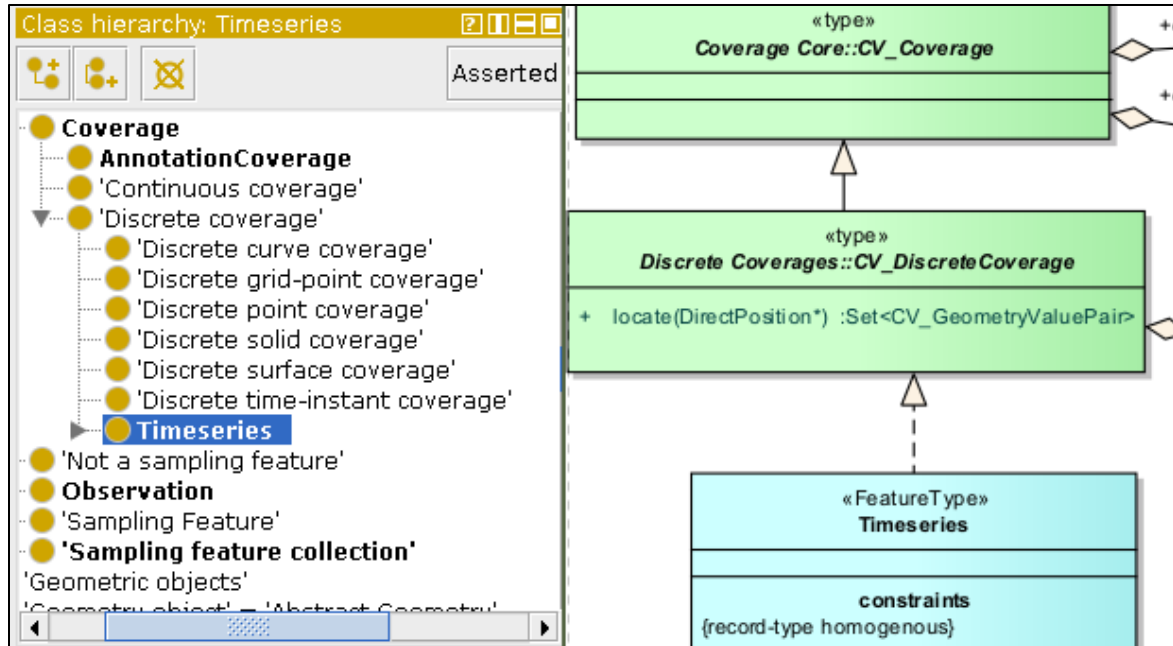


Figura 6-4. Clase “Timeseries”

Además, la reutilización de estos recursos ontológicos ha permitido establecer equivalencias entre las diversas ontologías que presentan clases semejantes. Esto permite establecer relaciones entre las diferentes ontologías y aprovechar las clases en común que puedan complementarse entre sí, como sucede para clase *Observation* como se puede ver en la Figura 6-5.

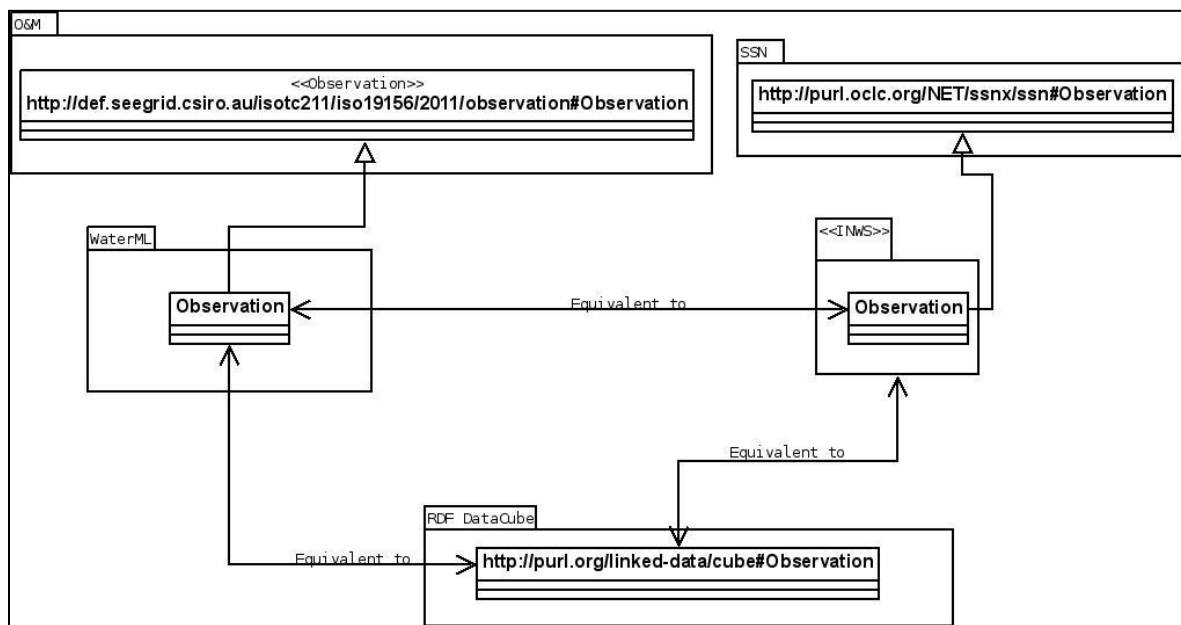


Figura 6-5. Esquema de equivalencias para la clase “Observation”.

6.2.5. Reutilización y reingeniería de recursos no ontológicos

En este trabajo se consideran otros recursos que no son ontologías, tales como documentos y especificaciones que brindan un aporte valioso al ofrecer lineamientos en la formulación de la red de ontologías. Por tanto, considerando su relevancia fueron reutilizados y se le aplicaron procesos de reingeniería para transformarlos en recursos ontológicos.

6.2.5.1. Dominio hidrológico

Dentro de contexto del dominio hidrológico se consideran los siguientes recursos no ontológicos, los cuales son transformados mediante procesos de reingeniería a ontologías y pasan a formar parte de la red de ontologías de este proyecto.

- **WaterML 2.0:** Es un estándar del OGC para la representación de datos de observaciones hidrológicas con enfoque en series de datos temporales. Este estándar fue creado con la intención de permitir el intercambio de estos conjuntos de datos entre sistemas de información. Su estructura en UML, en su *parte 1*³⁹, fue transformada a una ontología generando clases y subclasses (junto con sus propiedades) como se aprecia en la Figura 6-10.
- **CUAHSI-HIS:** El Sistema de Información Hidrológico CUAHSI es un sistema para compartir datos hidrológicos. Este sistema está conformado por bases de datos y servidores, conectados por medio de servicios web, a aplicaciones cliente, permitiendo la publicación, descubrimiento y acceso a los datos. El sistema CUAHSI posee un vocabulario controlado que permite definir recursos asociados con parámetros fisicoquímicos y microbiológicos que han sido incorporados a la red de ontologías, como se refleja en la Figura 6-6.

³⁹ https://portal.opengeospatial.org/files/?artifact_id=57222

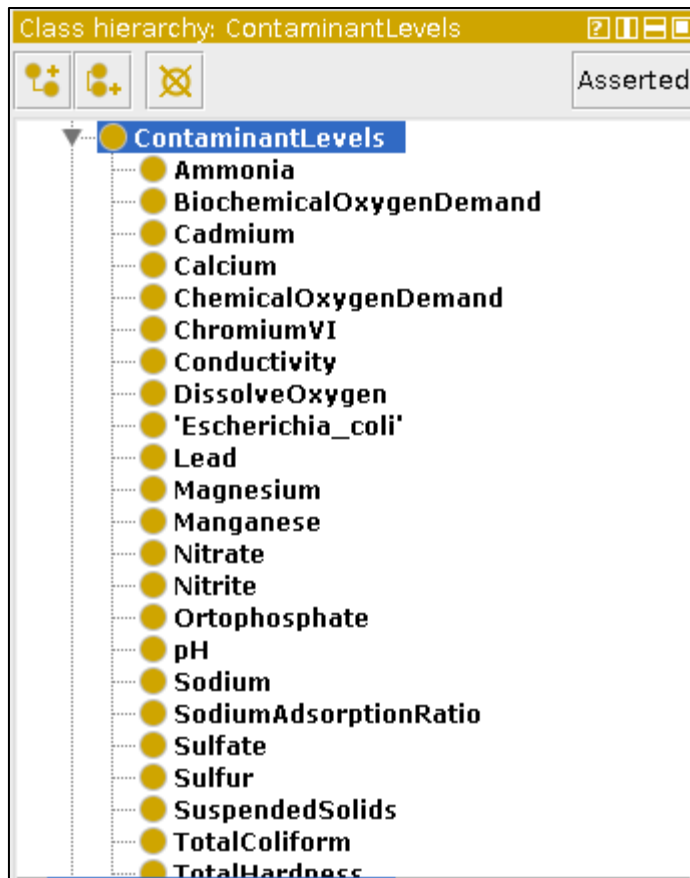


Figura 6-6. Parámetros de CUAHSI-HIS tras su transformación a ontología

6.2.5.2. Legislación Colombiana

Para poder definir los criterios y rangos aceptables del estado del recurso hídrico se necesita un referente que incluya los valores máximos permitidos. Para ello se toma como referencia la legislación actual que rige la cuenca de la zona de estudio seleccionada (río Bogotá).

El principal referente legal que articula los aspectos legislativos sobre la mencionada cuenca se recoge en el Acuerdo 43 del 2006 de la Corporación Autónoma Regional (CAR) de Cundinamarca donde *se establecen los objetivos de calidad del agua para la cuenca del río Bogotá a lograr en el año 2020*. En dicho documento se formulan cinco clases de uso de agua:

- Clase I: Agua para consumo humano y doméstico con tratamiento convencional, preservación de flora y fauna, uso agrícola y uso pecuario.
- Clase II: Agua para consumo humano y doméstico con tratamiento convencional, uso agrícola con restricciones y uso pecuario.
- Clase III: Calidad de los Embalses, Lagunas, humedales y demás cuerpos lénticos de agua.
- Clase IV: Usos agrícola con restricciones y pecuario.
- Clase V: Generación de energía y uso industrial.

Estas cinco clases poseen distintos límites de concentración y características físico-químicas y microbiológicas de acuerdo a los usos para los cuales pueda ser destinada el agua y han sido asignadas a distintas regiones de cada subcuenca (en total 19) del río Bogotá en función de las actividades y condiciones de cada tramo del río o de sus fuentes aportantes.

Los límites asociados a parámetros físico-químicos y microbiológicos del agua no son iguales a lo largo de la cuenca, ya que estos obedecen a características ambientales (por ejemplo, pendiente del cauce) y socioeconómicas (por ejemplo, presencia de centros poblados, instalación de parques industriales). Teniendo en cuenta esta diversidad, se han definido 5 clases de calidad del agua cada una con sus valores máximos permitidos para cada uno de sus parámetros las cuales a su vez son asignadas a distintos tramos dentro de la cuenca. En la Figura 6-7 se muestran las clases generadas para el contexto legal, así como sus propiedades (relaciones) asociadas, producto de su transformación a una ontología.

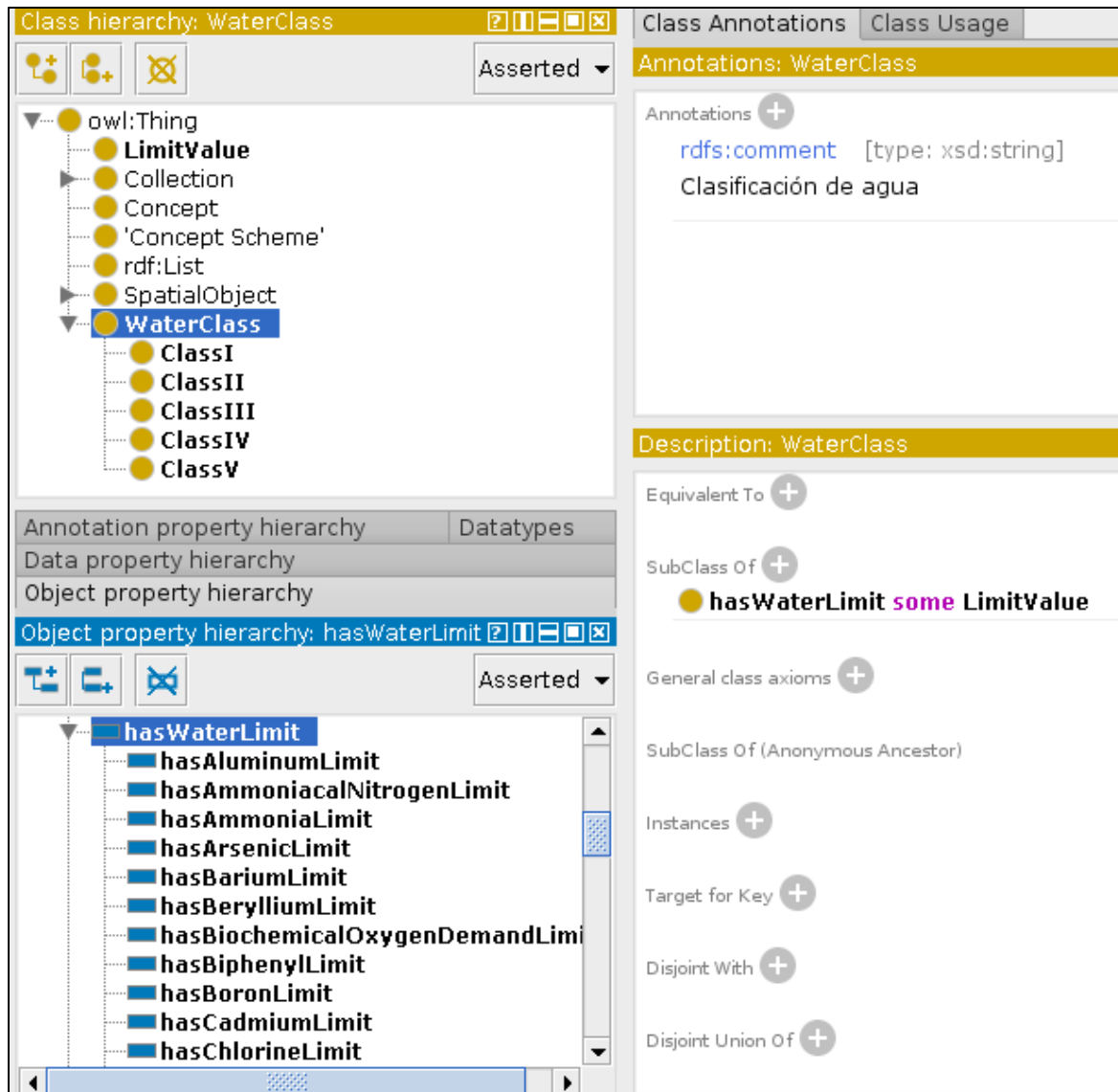


Figura 6-7. Clases generadas para contexto legal y propiedades

Dichas clases con requeridas en la asignación en la red de ontologías a cada unidad hidrológica, la cual posee su debida serialización como WKT (*Well-known text*) (International Organization for Standardization, 2015) y poder hacer consultas espaciales al momento de realizar la explotación de los datos, y así tener conocimiento de los límites máximos permitidos de acuerdo al contexto legal vigente.

6.2.6. Reutilización, fusión y re-ingeniería de recursos ontológicos

En relación a los recursos ontológicos, teniendo en cuenta que del estándar propuesto O&M (ISO 19156:2011) se deriva la especificación *WaterML* y para el cual ya ha sido generada una ontología (Cox, 2013), ésta se ha tomado como punto de partida para el proyecto. No obstante, esta ontología ha sido complementada de acuerdo a los lineamientos brindados por la especificación OGC sobre *WaterML*, ya que, aunque el O&M brinda buenas bases para el intercambio de datos provenientes de monitoreos, carece de elementos que permitan brindar un enfoque hidrológico a la información a gestionar.

Teniendo en cuenta esta restricción, ha sido necesario adaptar los recursos a la necesidad del proyecto. Este es el caso del *WaterML 2.0* (OGC et al., 2014) que conforma la base de esta red de ontologías y el cual consiste en una especificación del OGC cuyo principal objetivo es permitir el intercambio de datos asociados al agua entre sistemas de información. La propuesta de *WaterML* se basa en el modelo de información de Observaciones y Mediciones (O&M), versión 2.0. Sin embargo, ha sido creado para la incorporación de acuerdo a las reglas del GML (*Geography Markup Language*) (OGC, 2012), por lo cual ha sido necesario tomar como base los modelos UML (recursos no ontológicos) de dicho estándar y adaptarlo a una ontología.

Dichas modificaciones han sido realizadas para poder utilizar otros recursos que puedan enriquecer en mayor medida dicha red. Este es el caso del componente espacial al cual le ha sido incorporado otra especificación de OGC, como es el caso de *GeoSPARQL*, incluyendo a la clase *Observation* un atributo (*Object Property*) de clase *SamplingPoint* (Figura 6-8, Derecha) la cual es a su vez una subclase del objeto *Point* (Figura 6-8, Izquierda) que tiene sus atributos geométricos y espaciales definidos dentro de este estándar.

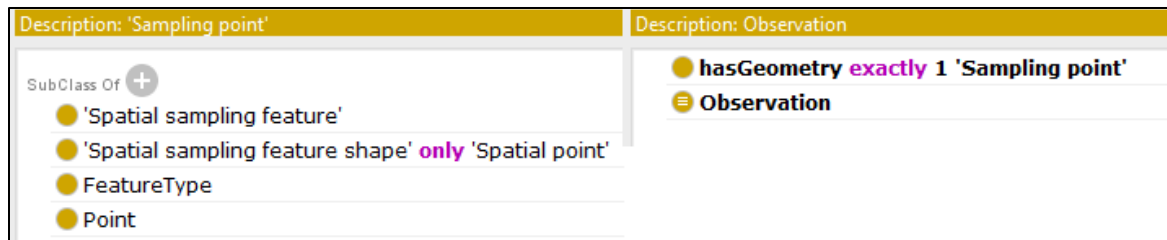


Figura 6-8. Izquierda: *SamplingPoint*, Derecha: *Observation* (asignación de geometría)

Los cambios restantes se encuentran asociados a recursos no ontológicos como los presentados en el estándar formulado por (Open Geospatial Consortium, 2014), donde se brindan los lineamientos del WaterML 2.0 en UML, motivo por el cual fue necesario transformarlo a OWL, como se mencionó con anterioridad. Para ello, se tomaron los mencionados modelos UML recogidos en la especificación WaterML (ver Figura 6-9) y se procede a generar los diferentes elementos que conforman la ontología (clases, relaciones y atributos). Un ejemplo de esta transformación es la clase *TimeSeriesObservation*, que puede verse en la Figura 6-10 junto con todas sus clases, subclasses y relaciones que también deben ser integradas en la ontología que conforma este módulo. Esta transformación (de UML a ontología) se realiza mediante el editor de ontología Protégé⁴⁰.

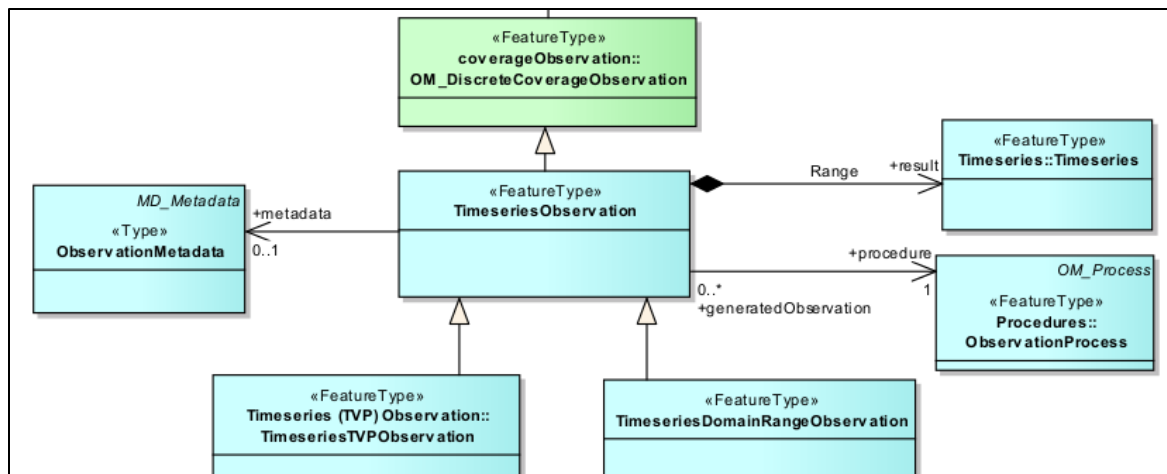


Figura 6-9. Clase “*TimeSeriesObservation*” en UML. Open Geospatial Consortium, 2014

⁴⁰ <https://protege.stanford.edu/>

Como puede observarse en la Figura 6-10, se han generado nuevas subclases (en negrita) de la clase *DiscreteCoverageObservation*, antes incluida en el O&M, dentro de las que se incluyen no solo la *TimeValuePaired* y *DomainRange*, también han sido incorporadas las subclases *Categorical* y *Measurement* de donde se heredan las características de las superclases a las cuales pertenecen.

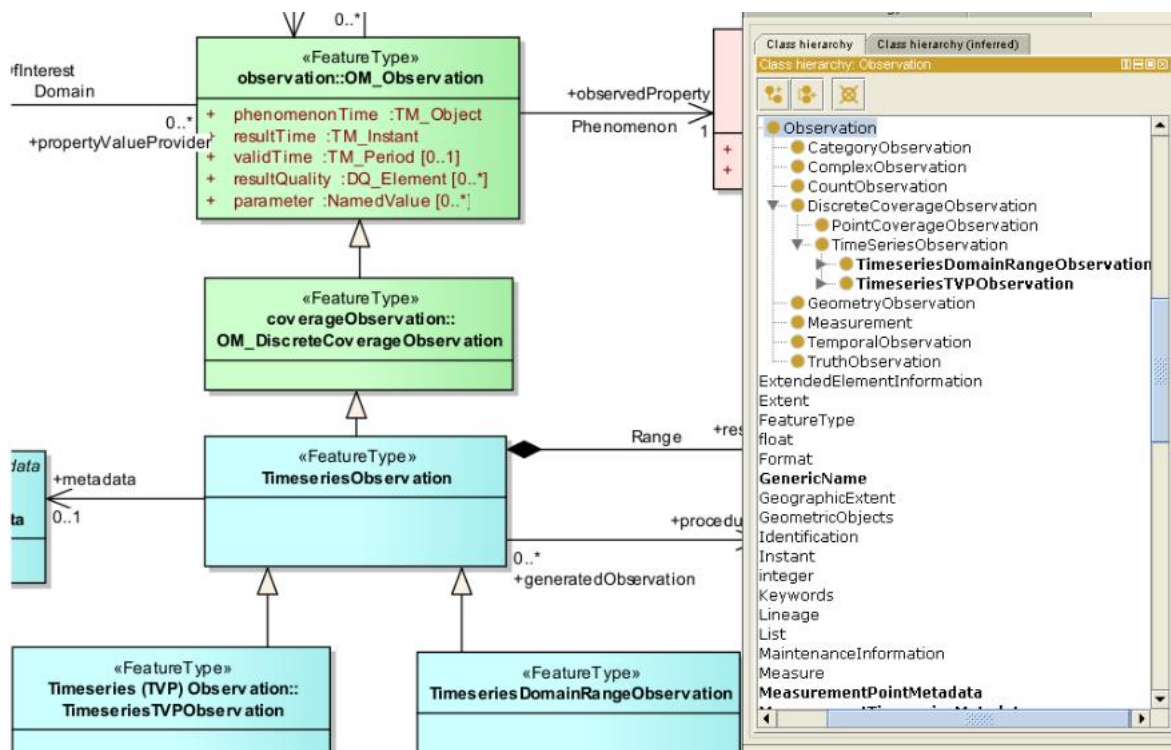


Figura 6-10. Clase “TimeSeriesObservation” en UML y Protégé

El uso del modelo UML formulado por la OGC no solo se limita al uso de clases y subclases, la asignación de atributos y relaciones se ha realizado mediante el uso de relaciones (también conocidas como *ObjectProperties* en el estándar OWL (OWL Working Group, 2012), indicando la cardinalidad y las restricciones asociadas a dichas indicaciones en los diagramas UML, siendo las *ObjectProperties* y las restricciones de subclase las que definen estas relaciones como se ha mostrado para el componente espacial de la clase *Observation*.

6.2.7. Red de ontologías

Una vez generado cada uno de los módulos y aplicados los escenarios descritos en la sección 6.2.2, se genera la red de ontologías cuya representación de alto nivel se muestra en la Figura 6-11, donde se puede apreciar la forma en que las ontologías se encuentran conectadas.

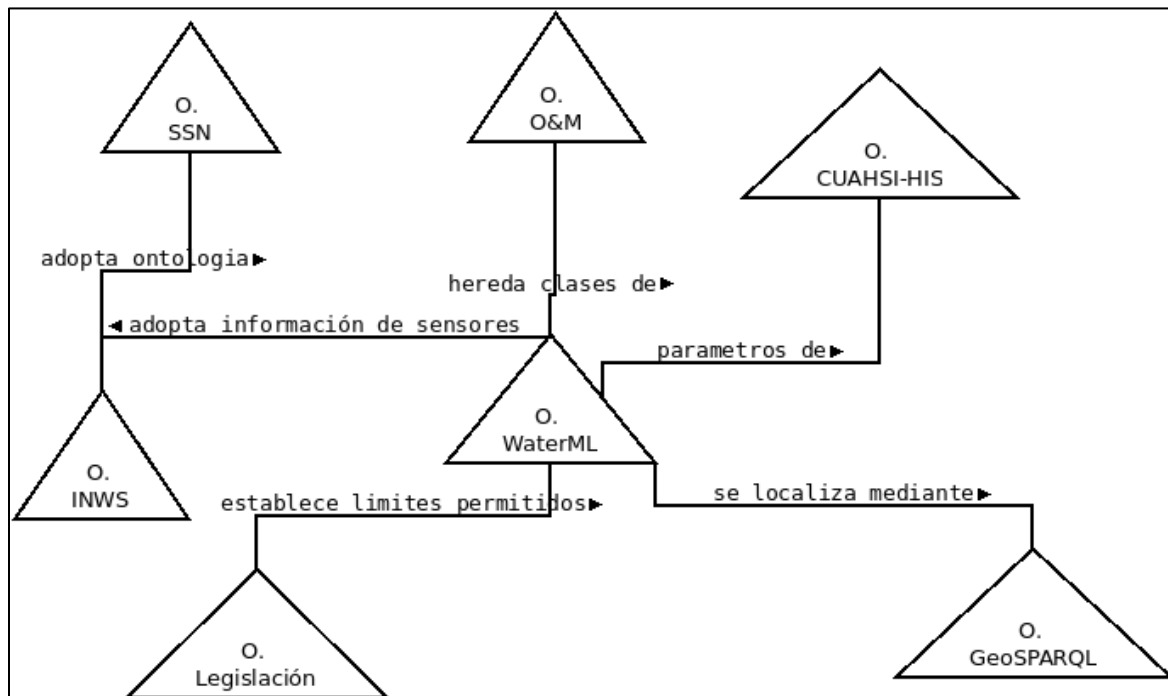


Figura 6-11. Visión de alto nivel de la red de ontologías desarrollada

En la figura mencionada se observa como la información estructurada mediante el WaterML 2.0 hereda las clases de la ontología propuesta para el O&M y se encuentra complementada por CUAHSI-HIS suministrando el listado de parámetros físico-químicos y microbiológicos a cuantificar para las muestras de agua y en temas asociados al muestreo mediante el SSN que ha sido incluido por medio del INWS. Por otro lado, el componente geográfico se incluye por medio del GeoSPARQL. Por último, la información asociada a concentraciones y magnitudes de los parámetros puede ser comparada con la norma vigente mediante la ontología generada sobre la legislación asociada con la calidad hídrica.

6.4. Generación de RDF y enlaces

En la siguiente sección se describe el proceso mediante el cual los datos considerados son transformados a RDF utilizando la red de ontologías desarrollada, conforme a lo propuesto por la metodología adoptada para la generación y publicación de Linked Data.

Teniendo en cuenta los diferentes formatos (PDF, *shapefile*, Microsoft Excel), naturaleza (tabular, espacial, alfanumérica, cadena de texto) y dominios (series históricas, límites de zonas hidrológicas, legislación vigente) que poseen los datos considerados, a continuación, se describe el proceso de transformación en cada caso teniendo en cuenta la naturaleza de cada uno.

6.4.1. Datos históricos

Los datos históricos hidrológicos, como se mencionaba con anterioridad, se encuentran almacenados en reportes de laboratorio en formato de Microsoft Excel, los cuales tienen asociados además de sus respectivos parámetros, la caracterización de la muestra y su contexto espacial y temporal (ver Figura 6-12). Dicha información requiere un tratamiento que consiste en un proceso de limpieza para eliminar inconsistencias realizar la asignación de fechas de toma de muestra mediante el establecimiento de un único formato, la estandarización de unidades de medida, eliminación de columnas y datos redundantes y generación de una base de datos en un formato tabular.

Este proceso de limpieza se ha llevado a cabo con la herramienta LODRefine (*Linked Open Data Refine*), una versión de *OpenRefine*⁴¹ que gracias a los *plugins* que incorpora facilita el proceso de generación de *Linked Data*.

La base de datos resultante es el insumo para la transformación de estos datos y generar RDF utilizando la red de ontologías desarrollada en este trabajo. Un ejemplo del proceso seguido para la generación de RDF de estos datos se muestra en la Figura 6-12, donde se indica la geometría y la información espacial a cada observación. En este caso, cada título en negrita corresponde a una columna en la base de datos creada, los cuales son definidos mediante URIs (por ejemplo, *No_Muestra* ha sido definida como un objeto de

⁴¹ <http://openrefine.org/>

clase *Observation*), y conectadas mediante URIs donde cada *Observation* tiene asignada una geometría (*geo:hasGeometry*) definida por los datos de la columna *Punto de Monitoreo*. Se debe tener en cuenta que todos los elementos se encuentran identificados siempre que así se requiera con las mismas URIs asignadas para las clases y propiedades de la red de ontología con el fin de establecer la conexión entre la red de ontologías desarrollada y los datos RDF.

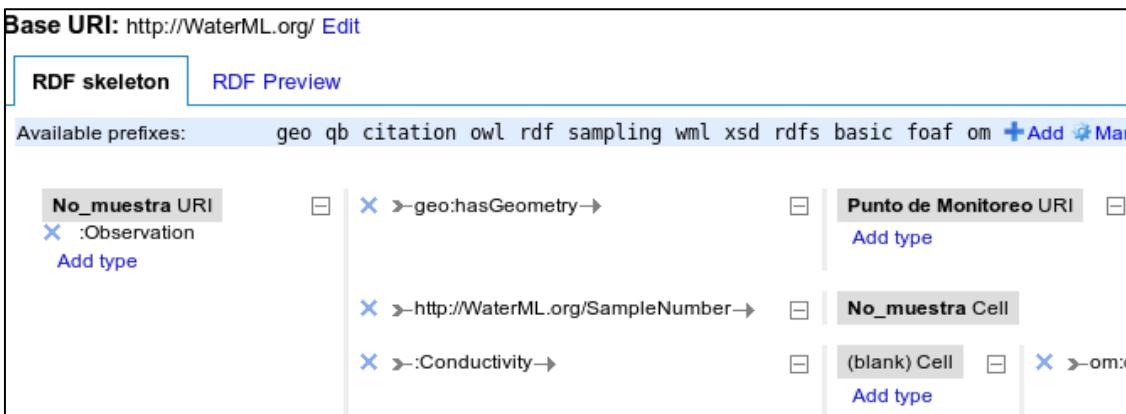


Figura 6-12. Estructura de RDF para la clase *Observation* y asignación de geometría

Por otro lado, la información asociada a la toma de la muestra, además de los valores cuantificados, contiene datos que caracterizan la muestra, tales como las unidades de medida (*unit of measure, uom*). En la Figura 6-13 se muestra un ejemplo de los grafos establecidos con la herramienta LODRefine que van a permitir la generación del RDF de los datos considerados.

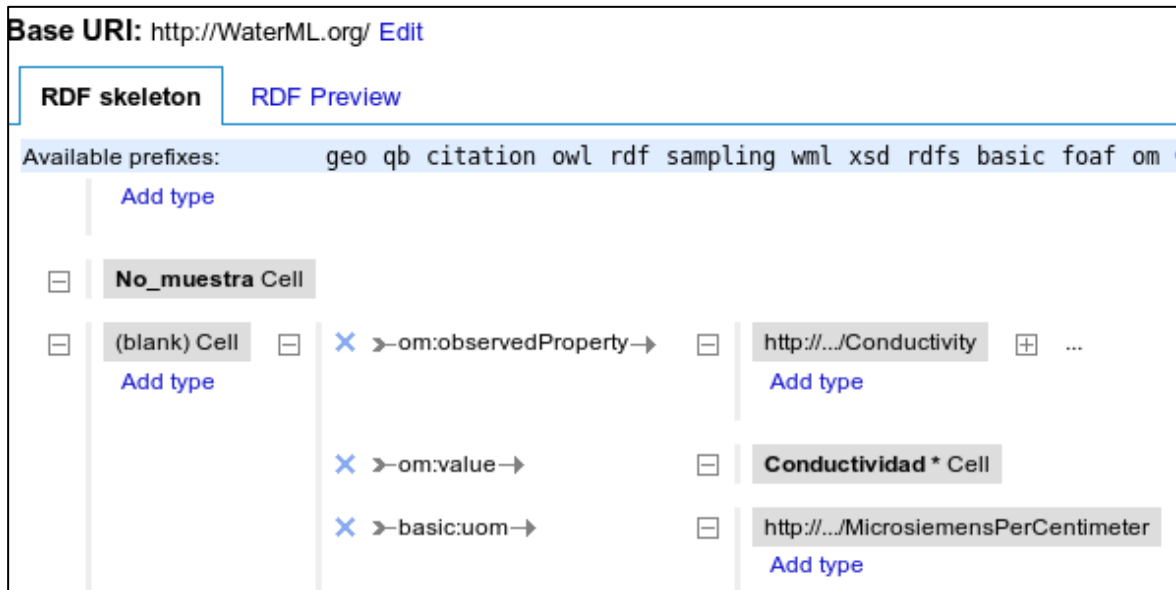


Figura 6-13. Estructura RDF para Unidad de Medida de Conductividad

6.4.2. Datos normativos

Siguiendo la estructura del modelo semántico recogido en la red de ontologías presentado en la sección 6.2.7, se toman los datos provenientes del documento normativo, el cual se encuentra en formato PDF, donde se describen los límites de cada parámetro en función de cada cuerpo hídrico. Considerando la situación inicial de los datos, se procede a compilarlos en un formato legible para la máquina, en este caso en formato CSV, para ser importado con posterioridad a la herramienta LODRefine.

Como se observa en la Figura 6-14, se ha generado una tabla donde se ha identificado con una URI cada clase recogida en el documento normativo. Estas clases corresponden con las modeladas en la red de ontologías desarrollada. Además, en esta tabla se recogen columnas con los valores máximos permitidos cuando así se haya indicado, de lo contrario el campo no registrará ningún valor, quedando en blanco.

Clase	DBO	OD	COLIFORMES T	NITROGENO AM	NITRATOS	NITRITOS
http://www.waterregulations.gov.co/WaterClass/ClassI	7	4	5000		10	10
http://www.waterregulations.gov.co/WaterClass/ClassII	7	4	20000		10	10
http://www.waterregulations.gov.co/WaterClass/ClassIII	20	4	5000	0.3	1	0.5
http://www.waterregulations.gov.co/WaterClass/ClassIV	50		20000			10
http://www.waterregulations.gov.co/WaterClass/ClassV	70					

Figura 6-14. Marco Normativa en LODRefine

A partir de esta tabla se procede a la utilización del *plugin RDF Refine*⁴² en la herramienta LOD Refine. Este plugin permite definir la estructura del RDF antes de ser exportada en este formato mediante la asignación de cada columna con los elementos recogidos en la red de ontologías desarrollada. En la Figura 6-15 se muestra un ejemplo de la estructura definida para la generación de RDF de los datos del marco normativo.

RDF skeleton

RDF Preview

Available prefixes: rdf rdfs wr foaf + Add Manage

Class URI		
Add type	✕ ➤ wr:hasBiochemicalOxygenDemandLimit ➔	DBO Cell
	✕ ➤ wr:hasOxygenDissolvedLimit ➔	OD Cell
	✕ ➤ wr:hasColiformTotalLimit ➔	COLIFORMES TOTALES Cell
	✕ ➤ wr:hasAmmoniacalNitrogenLimit ➔	NITROGENO AMONIAL Cell
	✕ ➤ wr:hasNitrateLimit ➔	NITRATOS Cell
	✕ ➤ wr:hasNitriteLimit ➔	NITRITOS Cell
	✕ ➤ wr:hasPhosphorusTotalLimit ➔	FOSFORO TOTAL Cell
	✕ ➤ wr:hasSolidsTotalSuspendedLimit ➔	SOLIDOS SUSPENDIDOS TOTALES Cell

Figura 6-15. Estructura RDF de marco normativo

En la mencionada figura se muestra como cada columna con la *Clase* de agua ha sido conectada a cada una de las clases para los parámetros físico-químicos y

⁴² <http://refine.deri.ie/>

microbiológicos mediante una propiedad, por ejemplo *wr:hasParameterLimit*, las cuales han sido previamente definidas en la red de ontologías desarrollada.

6.4.3. Datos espaciales

Los datos espaciales que se transforman a RDF corresponden a los puntos de monitoreo y a las unidades hidrológicas recogidas en los datos considerados. Cada uno de estos conjuntos de datos posee un origen y naturaleza diferente y, por lo tanto, su transformación a RDF se ha llevado a cabo de manera distinta en cada caso.

La estructura a asignar a los datos mediante la red de ontologías debe garantizar que la información pueda ser enlazada no solo mediante el uso de URIs, brindando una apropiada identificación de las entidades al dar un contexto adecuado a la información, también debe permitir satisfacer la necesidad de realizar análisis espaciales mediante los atributos geométricos de los lugares de monitoreo (punto), los límites de las zonas (polígono) y las clases asignadas a las mismas, junto con sus valores máximos. De esta manera, la generación de RDF, y su posterior explotación a través de análisis espacial, debe permitir la identificación de situaciones en el espacio y en el tiempo donde la calidad del agua supere el máximo permitido mediante la utilización de expresiones lógicas de valores numéricos y relaciones espaciales entre puntos y polígonos, incorporando conceptos descritos en el marco legal.

6.4.3.1. Puntos de monitoreo

Los puntos de monitoreo tienen su origen en una tabla contenida en un libro electrónico de Microsoft Excel, donde se recoge un listado de nombres, municipio y coordenadas en un sistema de referencia de coordenadas MAGNA-SIRGAS (EPSG: 3116). Considerando estos datos de partida, se realiza la reproyección al sistema WGS84 (EPSG:4326) y así habilitar consultas espaciales mediante GeoSPARQL, ya que en el momento en que se ha desarrollado este proyecto las tecnologías utilizadas no soportan sistemas de coordenadas proyectados.

Para ello se utiliza la herramienta OpenRefine, que permite la transformación de datos, haciendo efectiva la conversión de las columnas con los datos de latitud y longitud a formato *Well Known Text* (WKT). Aunque dentro de la red de ontologías desarrollada la

ontología GeoSPARQL contiene el tipo de dato WKT o GML, se ha escogido el primero para facilitar las consultas. La transformación a WKT se ha realizado mediante la siguiente instrucción GREL (*General Refine Expression Language*), un lenguaje creado para realizar transformaciones y manipulaciones específicas sobre los datos en OpenRefine. A continuación, se presenta el fragmento de la instrucción GREL presentada con formato JSON:

```
[
  {
    "op": "core/column-addition",
    "description": "Create column wkt at index 3 based on column Long using expression
    grel:'<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT ('+cells[\"Lat\"].value+'
    '+value+)'",
    "engineConfig": {
      "mode": "row-based",
      "facets": []
    },
    "newColumnName": "wkt",
    "columnInsertIndex": 3,
    "baseColumnName": "Long",
    "expression": "grel:'<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT
    ('+cells[\"Lat\"].value+' '+value+)'",
    "onError": "set-to-blank"
  }
]
```

Dicha codificación toma las columnas *Lat* y *Long* presentes en los datos espaciales y las une de la forma: (Lon, Lat), tras ello añade el encabezado para definir el sistema de referencia de coordenadas y el tipo de geometría (*POINT*) para generar los datos en la forma presentada en la Figura 6-16. Esta transformación posibilita, como se reflejará en la actividad de explotación, la realización de consultas espaciales mediante el lenguaje GeoSPARQL para evaluar la contención de los puntos frente a los datos normativos para aplicar el contexto espacial.

Lat	Long	wkt
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)
-73.573848	5.230846	<http://www.opengis.net/def/crs/OGC/1.3/CRS84> POINT (-73.573848 5.230846)

Figura 6-16. Columnas de Latitud y Longitud y su formato conforme a WKT

Tras la adición de esta nueva columna a los datos considerados, producto de la transformación aplicada, se procede a definir la estructura del RDF asignando URIs para cada *SamplingPoint*, respetando los elementos definidos en la red de ontologías desarrollada (Figura 6-17), para así garantizar la integridad del modelo de datos.

RDF skeleton
RDF Preview

Available prefixes: geo rdf sampling rdfs foaf ins [+ Add](#) [Manage](#)

Punto de Monitoreo URI

[X](#) [sampling:SamplingPoint](#)

[Add type](#)

[X](#) [ins:lat](#) → **Lat Cell**

[X](#) [ins:long](#) → **Long Cell**

[X](#) [geo:asWKT](#) → **wkt Cell**

[Add property](#)

Figura 6-17. Estructura RDF *SamplingPoint*

6.4.3.2. Unidades hidrológicas

Los datos de las unidades hidrológicas se encuentran descritos en el documento normativo y, aunque se trata de información espacial, no existe referente cartográfico de ningún tipo. Por este motivo, en el contexto de este trabajo se construyeron los límites formulados a partir de los cuerpos hídricos y las cuencas a las que pertenece, mencionados en el documento normativo. Para ello se crea un *shapefile* con sistema EPSG:3116 para garantizar la precisión de los datos usando una proyección local. Los polígonos construidos han sido generados indicando en la tabla de atributos la clase definida para cada caso, la cual aparece recogida en el documento normativo de la cuenca del Río Bogotá. En la Figura 6-18 se muestran las ubicaciones de las diferentes clases recogidas en el documento normativo de la cuenca del río Bogotá.

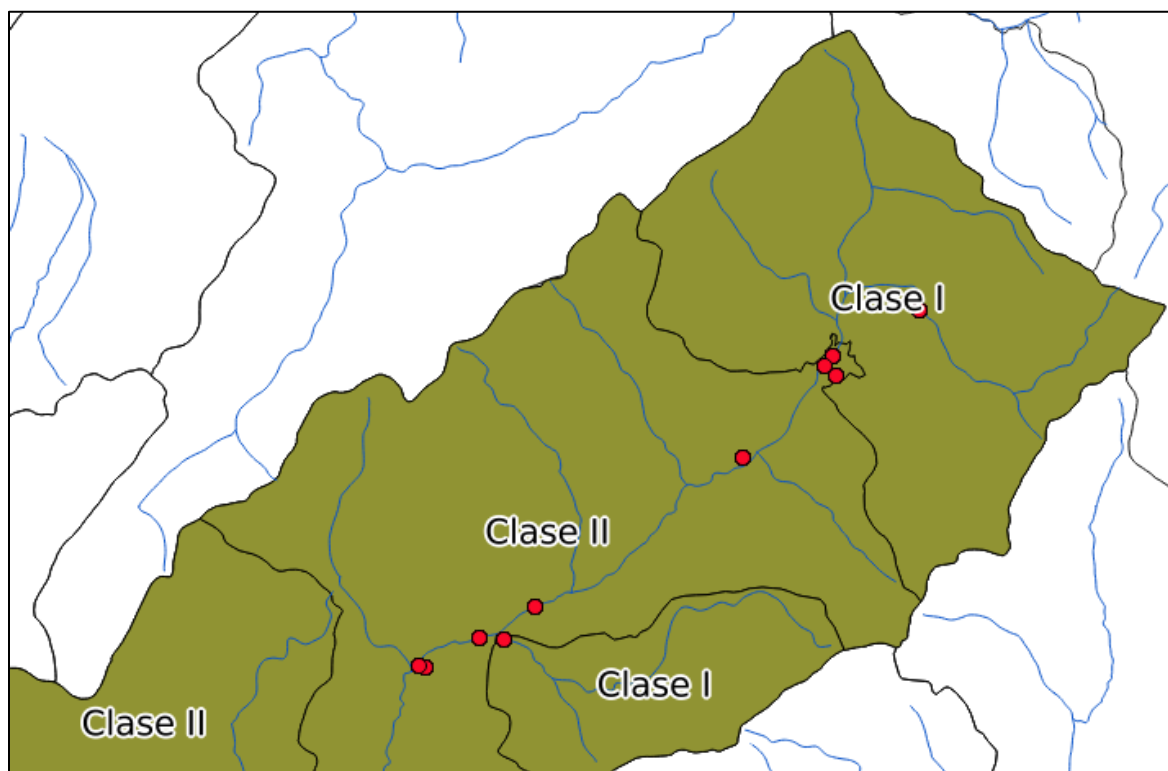


Figura 6-18. Vista de las clases para la parte alta de la cuenca del río Bogotá

Con la delimitación de las clases en un entorno espacial es posible ubicar cada punto de monitoreo y conocer los límites aplicables según la norma. Sin embargo, para hacer esto posible, mediante una consulta espacial utilizando GeoSPARQL, al igual que el caso

anterior, se han proyectado a un sistema de coordenadas geográfico y ha sido exportados a una tabla añadiendo una columna con la geometría en WKT y sus respectivos encabezados para definir la estructura del RDF, aplicando los mismos principios incorporados para los puntos de monitoreo como se aprecia en la Figura 6-19.

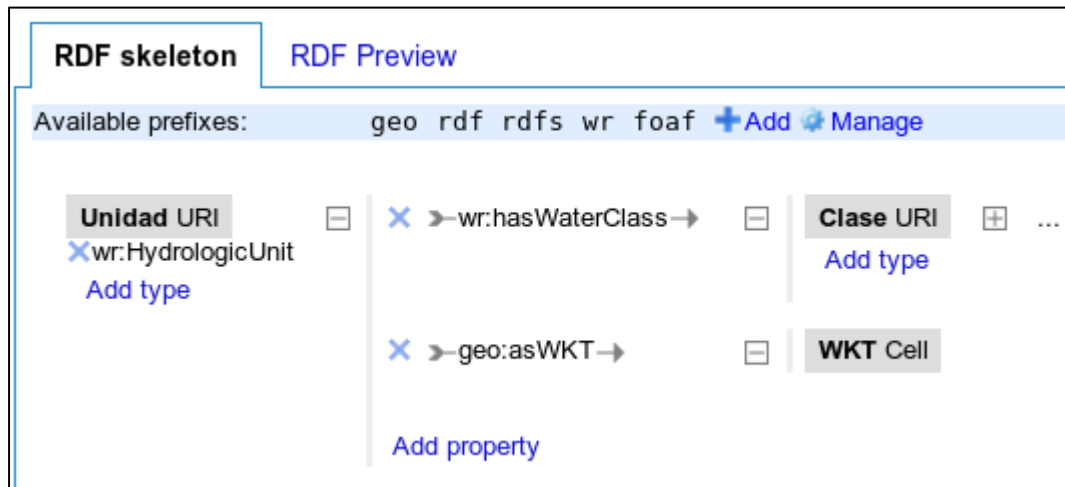


Figura 6-19. Estructura RDF de Unidades Hidrológicas

Una vez se tienen las diferentes fuentes de datos transformadas a RDF esto va a permitir la integración de los múltiples y heterogéneos datos considerados en este trabajo. Así, a partir de los puntos de monitoreo es posible integrar los datos de las series de datos histórica con cada uno de los parámetros que caracterizan la muestra de agua enlazada al componente espacial (Figura 6-19), donde a cada punto de monitoreo se les asignan las propiedades con la información de latitud (*ins:lat*), longitud (*ins:long*) y se serializa en WKT (*geo:asWKT*), con el fin de explotar este componente y darle uso mediante el análisis experto, tomando como referente no solo los valores cuantitativos sino su localización mediante un *join* espacial gracias a la integración de datos.

Esta integración de los diversos conjuntos de datos ha sido generada mediante la propiedad *geo:hasGeometry* que asigna a *Observation* una clase ya definida en la ontología como *SamplingPoint*, teniendo en cuenta que dichas clases a su vez poseen otras propiedades según se ha definido en la red de ontologías desarrollada, por tanto, pueden estar conectadas a valores con tipos de datos ya definidos (*entero-integer*,

decimal-*float*, *wktLiteral*) o a clases como las unidades de medida (*UnitsOfMeasure*) o la propiedad observada (*ObservedProperty*).

Con todos los atributos incorporados se genera una serie de datos en RDF con una estructura como la que se muestra en la Figura 6-20, donde se brinda todo el contexto de la muestra “1343” con una conductividad de 13 µS/cm.

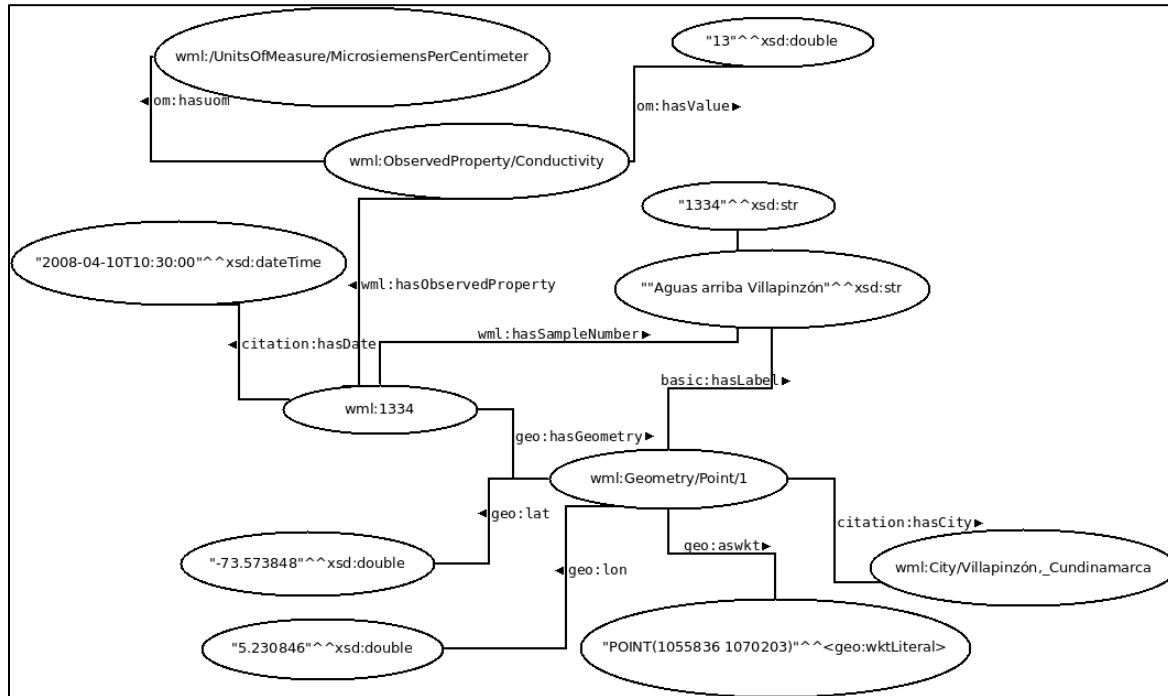


Figura 6-20. Grafo de RDF generado para observaciones y sus puntos de muestreo

6.4.4. Generación de enlaces

Para llevar a cabo la generación de enlaces ha sido necesario validar la información de cada registro asociada con el municipio donde se localiza la información de cada muestra asociada con la calidad hídrica. Para esto se han desarrollado diversos ejercicios de *join* espacial mediante la herramienta QGIS 2.16⁴³, tomado como referencia las coordenadas de cada punto de monitoreo y realizando el cruce con cartografía de división política, con el fin de verificar la veracidad del registro de datos o corregirlo de ser necesario.

⁴³ <http://www.qgis.org>

Esta información es de gran importancia no solo por el componente espacial que otorga a los datos, también porque información como el municipio permite conectar la información original con otras fuentes de datos presentes en la Web de Linked Data, como por ejemplo DBpedia⁴⁴, que es un proyecto que busca enriquecer la Web de Linked Data con contenido estructurado extraído de la información presente en Wikipedia⁴⁵.

Por tanto, la generación de enlaces entre la información RDF del proyecto y la información presente en DBpedia permite enriquecer considerablemente las descripciones originales de los municipios presentes en la información original, ya que a través de la generación de enlaces con la herramienta LOD-Refine se establece conexiones a la descripción de cada municipio presente en DBpedia. En la Figura 6-21 se muestra un ejemplo del enlace generado entre los datos del proyecto y la información de DBpedia para el caso de Suesca.



Figura 6-21. Enlace de datos a DBpedia utilizando URIs de municipio (Suesca)

⁴⁴ <http://wiki.dbpedia.org/>

⁴⁵ <https://www.wikipedia.org/>

6.5. Publicación y uso

En este capítulo se muestra la etapa final del proceso tal como se ha establecido en la metodología para las dos últimas etapas, dando lugar a la publicación que permita el posterior aprovechamiento, con el fin de mostrar la aplicabilidad que tiene esta tecnología sobre el dominio hidrológico. Esta publicación y uso permite generar un repositorio donde pueda ser importado todo el RDF geoespacial generado y la red de ontologías desarrollada para su posterior aprovechamiento a partir de los atributos espaciales y así mostrar el verdadero potencial de los datos como Linked Data geoespacial.

6.5.1. Repositorio de datos semánticos

La red de ontologías y los datos en formato RDF generados para poder ser explotados en el contexto de la Web de Linked Data se necesita crear un repositorio de datos semántico. Para ello se necesita desplegar un *triple store* que permita su almacenamiento, así como realizar consultas utilizando el lenguaje SPARQL (*SPARQL Protocol and RDF Query Language*).

Para el montaje de este servidor, se trabajó bajo un sistema operativo GNU/Linux y una distribución Ubuntu 16.04.1 LTS⁴⁶ (Xenial Xerus). Sobre dicha plataforma se ha hecho uso del *triple store* (repositorio de almacenamiento de los datos en RDF) y solución de gestión de datos Parliament⁴⁷. Este *triple store* diseñado para la Web Semántica fue originalmente desarrollado bajo el nombre DAML DB⁴⁸ y extendido por BBN Technologies⁴⁹, el cual es compatible con los estándares RDF, RDFS, OWL, SPARQL y GeoSPARQL. Este *triple store* incluye una serie de paquetes de código abierto como *Jetty*⁵⁰ (contenedor de servlets), *Jena*⁵¹ y *ARQ*⁵² (procesador de consultas), *Joseki* (implementación de SPARQL basada en *servlet*) y Berkeley DB (utilizada para

⁴⁶ <http://old-releases.ubuntu.com/releases/16.04.1/>

⁴⁷ <http://parliament.semwebcentral.org/>

⁴⁸ www.daml.org/2001/09/damldb

⁴⁹ <https://www.raytheon.com/ourcompany/bbn/>

⁵⁰ <https://www.eclipse.org/jetty/>

⁵¹ <https://jena.apache.org/>

⁵² <https://jena.apache.org/documentation/javadoc/arq>

implementar el diccionario de recursos). La estructura asociada al *triple store* Parliament se muestra en la Figura 6-22.

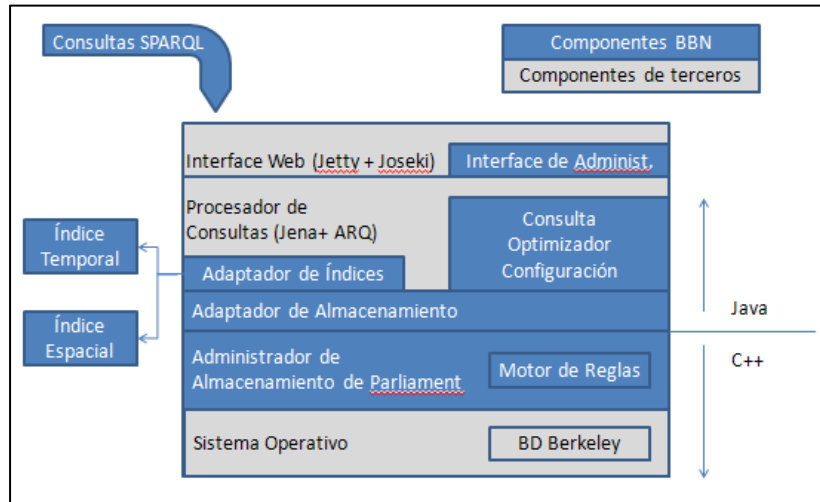


Figura 6-22. Arquitectura Parliament

Con la información asociada a las muestras y al contexto legal en el punto de consulta (SPARQL *Endpoint*) que ofrece el servidor Parliament desplegado resulta posible la consulta de la información mediante el lenguaje SPARQL. No obstante, es importante tener en cuenta que no sólo se tiene acceso a los datos asociados a calidad del agua que son objeto de este estudio, también a los metadatos de estos datos y que brindan el contexto de dicha información dando lugar a la semántica para permitir un nivel de abstracción más allá del que brinda el valor por sí solo.

Un ejemplo de consulta hecho desde el SPARQL *Endpoint* de Parliament se muestra en la Figura 6-23, en el cual se solicitan todos los valores de pH y conductividad debidamente identificados, estos se encuentran agrupados por número de muestra. En esta consulta se ha generado la variable “s” a la cual se le han asignado todos los elementos que corresponden a la clase *SampleNumber* (identificada con la URI de acuerdo al prefijo definido al principio de la consulta). Estos elementos se conectan a los nodos “*Conduct*” y “*phnode*” que corresponde a clases con información asociadas al valor y las unidades de la cuantificación, para este caso se consulta esta última solicitando la información asociada al recurso *value*.

```

SELECT or CONSTRUCT query

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX observation: <http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#>
PREFIX wml: <http://WaterML.org/>
PREFIX basic: <http://def.seegrid.csiro.au/isotc211/iso19103/2005/basic#>

SELECT ?Muestra ?Conductividad ?pH WHERE {
  ?s wml:SampleNumber ?Muestra ;
    wml:Conductivity ?Conduct ;
    wml:pH ?phnode.
  ?Conduct observation:value ?Conductividad.
  ?phnode observation:value ?pH.
}
    
```

Figura 6-23. Consulta de valores asociados a conductividad y pH

Al realizar la consulta anterior se obtienen los valores pH y conductividad eléctrica agrupados por número de muestra (Figura 6-24). Sin embargo, esta consulta puede no ser suficiente para la interpretación de resultados por parte del experto ya que es necesario añadir datos como el lugar de toma de muestra, municipio, unidades de trabajo y tipo de muestra con el fin de realizar y un adecuado análisis.

Count: 659

Muestra	Conductividad	pH
"4116-13"	"248"	"7.2" ^^<http://www.w3.org/2001/XMLSchema#double>
"4115-13"	"225"	"6.9" ^^<http://www.w3.org/2001/XMLSchema#double>
"4114-13"	"247"	"6.9" ^^<http://www.w3.org/2001/XMLSchema#double>
"4113-13"	"250"	"6.7" ^^<http://www.w3.org/2001/XMLSchema#double>
"4112-13"	"269"	"7.1" ^^<http://www.w3.org/2001/XMLSchema#double>
"4101-13"	"689"	"7.3" ^^<http://www.w3.org/2001/XMLSchema#double>
"4100-13"	"359"	"7.1" ^^<http://www.w3.org/2001/XMLSchema#double>
"4098-13"	"243"	"7" ^^<http://www.w3.org/2001/XMLSchema#double>
"4097-13"	"607"	"6.8" ^^<http://www.w3.org/2001/XMLSchema#double>
"4096-13"	"921"	"7.3" ^^<http://www.w3.org/2001/XMLSchema#double>
"4095-13"	"265"	"6.9" ^^<http://www.w3.org/2001/XMLSchema#double>

Figura 6-24. Resultados de valores asociados a conductividad y pH

Gracias a la semántica incorporada al conjunto de datos mediante el esquema brindado por documentos como el *WaterML* y el O&M o recursos ontológicos como el SSN, es

posible incluir en la consulta recogida en la Figura 6-25, donde a diferencia del caso anterior, no solo se piden los valores de pH de cada muestra, sino también los datos asociados a la fecha de toma, unidades de trabajo o ubicación como ciudad o geometría espacial. Esta información adicional (metadatos) corresponde a información como el lugar de toma de la muestra solicitando los recursos que correspondan a la propiedad *geo:hasGeometry*, que había sido previamente definida como se presentó en el capítulo 7, o la ciudad (*citation:city*), fecha de toma de la muestra (*citation:date*) o unidades de trabajo (*basic:uom*), en cuyo caso corresponde a microsiemens por centímetro. Esta última propiedad asociada al mismo nodo *phnode* el cual también contiene la propiedad *value* con la cuantificación del parámetro analizado, todas debidamente definidas con URIs, previamente definidas en la red de ontologías desarrollada. Los resultados obtenidos de la consulta, recogida en Figura 6-25, se muestran en la Figura 6-26.

```

SELECT or CONSTRUCT query

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX observation: <http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#>
PREFIX wml: <http://WaterML.org/>
PREFIX basic: <http://def.seegrid.csiro.au/isotc211/iso19103/2005/basic#>
PREFIX geo:<http://www.opengis.net/ont/geosparql#>
PREFIX citation:<http://def.seegrid.csiro.au/isotc211/iso19115/2003/citation#>

SELECT ?Muestra ?Ciudad ?Fecha ?Conductividad ?Unidades WHERE {
  ?s wml:SampleNumber ?Muestra ;
    wml:Conductivity ?Conduct ;
    wml:pH ?phnode;
    geo:hasGeometry ?Puntomuestreo;
    citation:date ?Fecha.
  ?Puntomuestreo citation:city ?Ciudad.
  ?Conduct observation:value ?Conductividad;
  basic:uom ?Unidades.
}
    
```

Figura 6-25. Consulta de valores de conductividad con metadatos

Count: 1429

Muestra	Ciudad	Fecha	Conductividad	Unidades
"5652"	http://dbpedia.org/resource/Ricaurte,_Cundinamarca	"2009-11-17T14:30:00" ^^< http://www.w3.org/2001/XMLSchema#dateTime >	"548"	http://www.WaterML.org/UnitsOfMeasure/MicrosiemensPerCentimeter
"415"	http://dbpedia.org/resource/Ricaurte,_Cundinamarca	"2009-02-05T11:00:00" ^^< http://www.w3.org/2001/XMLSchema#dateTime >	"592"	http://www.WaterML.org/UnitsOfMeasure/MicrosiemensPerCentimeter
"4035-13"	http://dbpedia.org/resource/Sopó	"2013-11-20T13:12:00" ^^< http://www.w3.org/2001/XMLSchema#dateTime >	"193"	http://www.WaterML.org/UnitsOfMeasure/MicrosiemensPerCentimeter
"2184"	http://dbpedia.org/resource/Sopó	"2013-07-11T13:15:00" ^^< http://www.w3.org/2001/XMLSchema#dateTime >	"109"	http://www.WaterML.org/UnitsOfMeasure/MicrosiemensPerCentimeter
"2438"	http://dbpedia.org/resource/Sopó	"2012-07-12T11:07:00" ^^< http://www.w3.org/2001/XMLSchema#dateTime >	"186"	http://www.WaterML.org/UnitsOfMeasure/MicrosiemensPerCentimeter

Figura 6-26. Valores de conductividad con metadatos

Como se puede observar en la Figura 6-26, tras realizar la consulta se obtienen los resultados de conductividad definidos mediante su unidad de medida (identificada con una URI), la fecha a la que pertenece la muestra junto con su número de identificación y la ciudad a la que pertenece definida con su URI de DBpedia, lo que permite reflejar el enlazado de datos los datos del trabajo con otros existentes en la Web de Linked Data.

6.5.2. Explotación de datos

En esta fase final de la metodología adoptada para la generación y publicación de Linked Data se explotan los datos previamente generados en RDF y almacenados en el *triple store* desplegado. Esto facilita la interpretación de los datos mediante análisis estadísticos y múltiples visualizaciones, aprovechando el potencial de este tipo de datos realizando diversos tipos de análisis.

Por tratarse de un potente paquete estadístico para el análisis de datos se ha hecho uso de R, dicho paquete posee la librería *SPARQL*⁵³ (van Hage, 2013b) para realizar consultas al *SPARQL Endpoint* desplegado como parte de este trabajo. Así, la mencionada librería permite realizar peticiones mediante el protocolo GET y solicita los datos en formato XML lo cual ocasiona una inadecuada asignación de clase de formato a datos de tipo numérico, ya que son almacenados como tipo cadena (*string*).

Considerando la situación expuesta, se utiliza el software *CURL*⁵⁴ (*Command line tool and library for transferring data with URLs*), donde para evitar este escenario se creó una nueva función, la cual toma una variable donde se almacena la URL del servicio y otra para la consulta que será codificada en formato URL para realizar la petición. Una vez se ha generado un mecanismo de consulta desde el paquete estadístico y gracias a las bondades proporcionadas por el *triple store* Parliament para exportar los datos en distintos formatos (CSV, XML, ttl), se ha planteado la función de manera que la información consultada pueda ser cargada mediante valores separados por comas (CSV), ya que de esta manera el paquete estadístico asigna los tipos de dato (texto, numérico, fecha) a las columnas teniendo en cuenta su contenido de manera más eficiente.

Mediante consultas como la mostrada en la Figura 6-27, donde se han obtenido los valores de parámetros como Conductividad, pH, Demanda Bioquímica de Oxígeno (DBO), Demanda Química de Oxígeno (DQO), Solidos Suspendidos, Solidos Totales y *E coli*; es posible realizar diversos análisis como los que se observan en la Figura 6-28. En esta figura se presentan parámetros estadísticos básicos (media, mediana, cuartiles), en la parte inferior izquierda, o diagramas de dispersión como se muestran en la sección inferior derecha de la mencionada figura, donde se muestra la relación presente entre DBO (Demanda Bioquímica de Oxígeno) y DQO (Demanda Química de Oxígeno), gracias al potencial que brinda el disponer de esta información debidamente estructurada por las ontologías y transformada a RDF, facilitando su consulta y proporcionando la interoperabilidad que brinda el sistema de información desarrollado bajo los estándares y

⁵³ <https://cran.r-project.org/web/packages/SPARQL/SPARQL.pdf>

⁵⁴ <https://curl.haxx.se/>

principios de Linked Data y que permite su conexión con otras herramientas como se ha mostrado en este caso para R.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX observation: <http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#>
PREFIX wml: <http://WaterML.org/>
PREFIX basic: <http://def.seegrid.csiro.au/isotc211/iso19103/2005/basic#>

SELECT DISTINCT ?Muestra ?Conductividad ?pH ?DBO ?DQO ?SolidosSupendidos ?SolidosTotales
?EColi WHERE {
?s wml:SampleNumber ?Muestra ;
wml:Conductivity ?Conduct ;
wml:pH ?phnode;
wml:DBO ?dbonode;
wml:DQO ?dqonode;
wml:ST ?stnode;
wml:SS ?ssnode;
wml:E.Coli ?ecolinode.
?Conduct observation:value ?Conductividad.
?phnode observation:value ?pH.
?dbonode observation:value ?DBO.
?dqonode observation:value ?DQO.
?ssnode observation:value ?SolidosSupendidos.
?stnode observation:value ?SolidosTotales.
?ecolinode observation:value ?EColi.
}
```

Figura 6-27. Consulta SPARQL para análisis estadístico

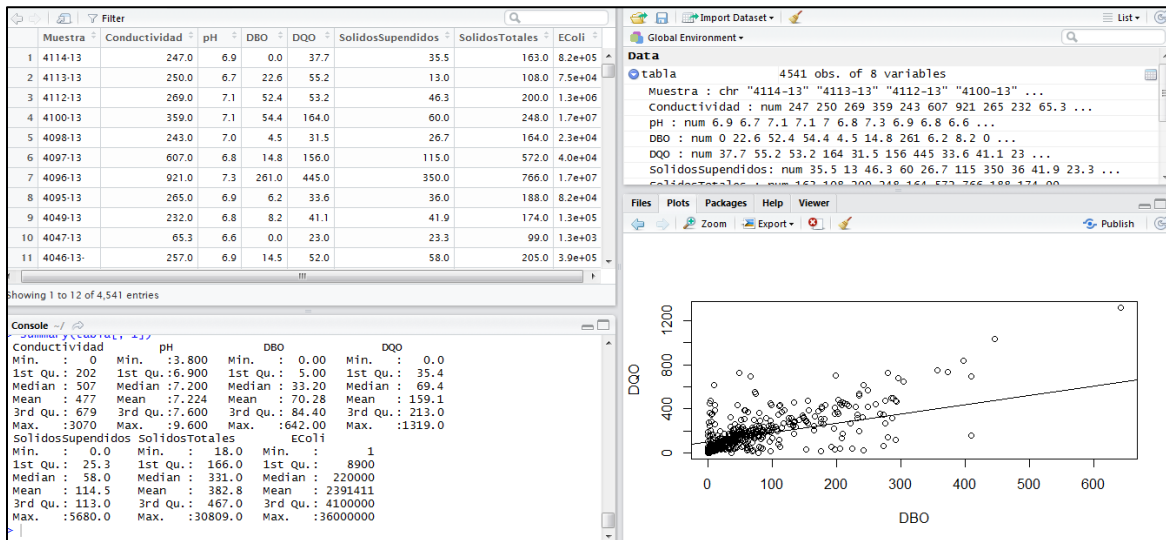


Figura 6-28. Análisis estadístico

Como resultado de la consulta y gracias a la estructura de los datos, además de los parámetros estadísticos de las variables, se ha generado un modelo lineal representado

en la línea del diagrama dispersión recogido en la Figura 6-28, donde se puede apreciar la relación que existe entre la DBO y la DQO especialmente para concentraciones altas.

6.5.3. Aplicación de análisis espacial y contexto legal

La incorporación del componente espacial a los datos de tipo legal y a los puntos de monitoreo permiten la posibilidad de realizar consultas utilizando GeoSPARQL como lenguaje de consulta. Esto da la posibilidad de realizar análisis para ubicar los puntos de monitoreo en un contexto legal, como consecuencia de que para este análisis ha sido necesario enlazar datos de distintas fuentes (series históricas con parámetros físico-químicos, localización de puntos de muestreo, unidades hidrológicas, clases hidrológicas).

Mediante una consulta como la que se puede apreciar en la Figura 6-29, es posible generar la unión entre la muestra y los resultados obtenidos y esta a su vez al punto de monitoreo, sirviendo de conexión a la información legal asignada a la unidad hidrológica correspondiente mediante un análisis espacial evaluando qué puntos de monitoreo se encuentran contenidos en cada unidad, la cual se le ha añadido los datos de clase de agua con sus respectivos límites para cada parámetro según corresponda. Con estos enlaces de datos, que no solo tabulares sino espaciales, es posible realizar el análisis de un experto evaluando en qué lugar se ha excedido la norma de acuerdo al uso para el que ha sido destinado el recurso hídrico, como se ha hecho en este caso para DBO. Con este fin, en primer lugar, se incluyen en la consulta la muestra, clase (tipo de agua según la legislación), concentración de DBO, Latitud y Longitud. De esta manera, con las coordenadas se ubican las muestras dentro de cada unidad haciendo uso de la geometría que cada objeto posee mediante un análisis espacial de contención (*sf:contains*), dichas unidades a su vez tienen una tabla de valores máximos permitidos que servirá de referencia para comprar, mediante una sentencia lógica (?DBO > ?DBOLimit), la identificación de aquellos lugares donde se infrinja la norma.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX observation: <http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#>
PREFIX wml: <http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX units: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX sampling: <http://def.seegrid.csiro.au/isotc211/iso19156/2011/sampling#>
PREFIX regulation: <http://www.waterregulations.gov.co/>

SELECT DISTINCT ?Muestra ?clase ?DBO ?DBOLimit ?lat ?long
WHERE {
?unidad a regulation:HydrologicUnit.
?unidad regulation:hasWaterClass ?clase.
?clase <http://www.waterregulations.gov.co#hasBiochemicalOxygenDemandLimit> ?DBOLimit.
?unidad geo:asWKT ?geounidad.
?punto a sampling:SamplingPoint;
<http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat;
<http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long.
?Muestra geo:hasGeometry ?punto.
?Muestra <http://WaterML.org/DBO> ?dbonode.
?dbonode observation:value ?DBO.
?punto rdfs:label ?nombre.
?punto geo:asWKT ?geopunto.
FILTER (geof:sfContains(?geounidad, ?geopunto) && ?DBO > ?DBOLimit)
}

```

Figura 6-29. Consulta espacial para aplicar contexto legal

Los resultados obtenidos que se presentan en forma tabular en la Figura 6-31 muestran el potencial de este tipo de consultas mostrando aquellos puntos que en algún momento han infringido la norma y han dejado de cumplir con los objetivos de calidad del agua de acuerdo a la ubicación de la muestra de agua, dando como ejemplo respuesta a una de las preguntas de la sección la sección 6.2.3.6 “¿En qué lugar y momento se está incumpliendo la norma?”. Con estos resultados se ha generado un aplicativo para el análisis espacio-temporal de la calidad del agua dentro del contexto legal aplicable a calidad del agua para este caso, gracias a las propiedades *lat* y *long* y haciendo uso de las librerías Shiny⁵⁵ y Leaflet⁵⁶ que posee R, estas librerías han sido desarrolladas para generar aplicativos web interactivos y mapas web mediante las respectivas librerías.

El aplicativo desarrollado consiste en un mapa en entorno web con la ubicación de aquellos sitios donde no se cumple con los objetivos formulados (Figura 6-30),

⁵⁵ <https://cran.r-project.org/web/packages/shiny/shiny.pdf>

⁵⁶ <https://cran.r-project.org/web/packages/leaflet/leaflet.pdf>

demostrando como estas tecnologías al incorporar el contexto (legal) en los datos gracias a la semántica, ofrecen una alternativa que no solo proporciona datos, también llega a generar conocimiento acerca del cumplimiento de la norma mediante la formulación de una nueva tecnología geoespacial para hacer análisis propio de un experto una vez se ha implementado la metodología expuesta en este trabajo.



Figura 6-30. Puntos de monitoreo que infringen la norma para DBO (Mapa)

El análisis realizado deja claro como en algún momento durante el periodo entre el 2007 y 2013, para el parámetro de DBO a lo largo de todo el cauce principal del río Bogotá se ha incumplido la norma. Por lo tanto, se puede decir que en distintos momentos para la demanda bioquímica de oxígeno no se han dado los lineamientos asociados a su uso, así como a los objetivos de calidad establecidos para la cuenca.

Count: 334

Muestra	clase	DBO	DBOLimit
http://WaterML.org/2432	http://www.waterregulations.gov.co/WaterClass/ClassV	"78.7" ^^< http://www.w3.org/2001/XMLSchema#double >	"70" ^^< http://www.w3.org/2001/XMLSchema#double >
http://WaterML.org/1570	http://www.waterregulations.gov.co/WaterClass/ClassV	"101" ^^< http://www.w3.org/2001/XMLSchema#double >	"70" ^^< http://www.w3.org/2001/XMLSchema#double >
http://WaterML.org/2431	http://www.waterregulations.gov.co/WaterClass/ClassV	"205" ^^< http://www.w3.org/2001/XMLSchema#double >	"70" ^^< http://www.w3.org/2001/XMLSchema#double >

Figura 6-31. Puntos de monitoreo que infringen la norma para DBO (Tabular)

Sin embargo, no solo son posibles análisis espaciales, también se pueden hacer análisis temporales como el recogido en la Figura 6-32, donde puede observarse como en la quebrada Socotá para DBO, durante todo el periodo de análisis se ha incumplido con el límite permitido (7 mg/L), allí no hay una tendencia aparente y las mayores concentraciones se dan para la primera parte del 2009, mitad del 2011 y principio del 2013.

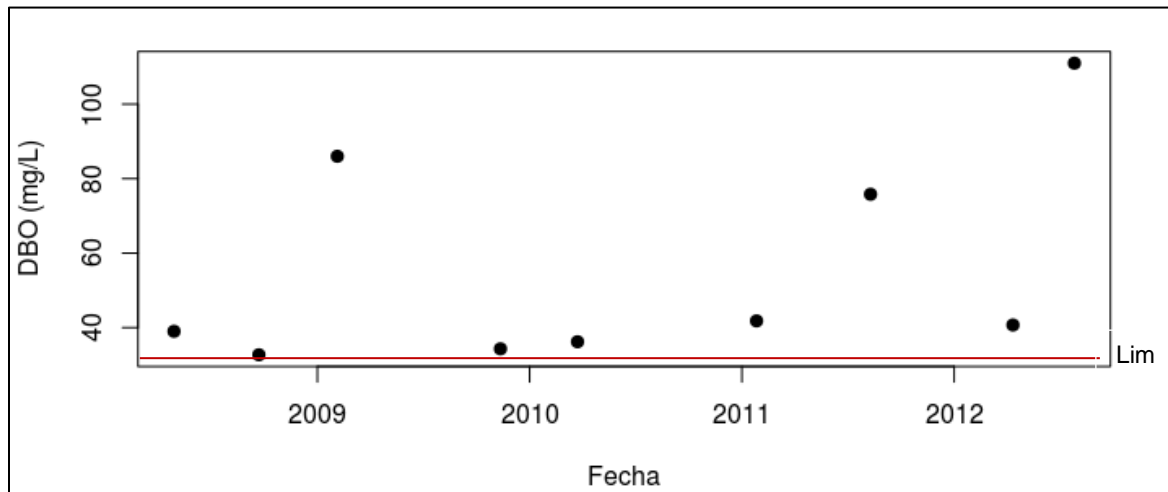


Figura 6-32. Análisis temporal DBO en Q. Socotá

En estos casos aplicados queda claro como al final mediante el uso de Linked Data y al unir conjuntos de datos de distintas fuentes correspondientes a muestras de agua,

puntos de monitoreo, clases de agua según el contexto legal y límites máximos permitidos mediante el uso de URIs y realizar consultas espaciales, es posible acercar al usuario final el análisis experto en el contexto hidrológico, específicamente para calidad del agua y su grado de cumplimiento con la norma aplicable, generando así un mecanismo para las tecnologías espaciales que permiten extender su uso al realizar análisis de manera más dinámica.

Capítulo 7. Discusión de resultados

Una vez presentados los resultados, en este capítulo se plantea el análisis de los productos obtenidos con el desarrollo del proyecto teniendo en cuenta el cumplimiento de los objetivos propuestos y los aportes generados al conocimiento en el marco de este trabajo.

7.1. Red de ontologías desarrollada

El desarrollo de la red de ontologías propuesta en este trabajo conlleva la construcción de un nuevo modelo semántico que combina la reutilización de diversos vocabularios estándares, tales como GeoSPARQL, SSN o RDF Data Cube, con el desarrollo de ontologías a partir de recursos no ontológicos, tales como WaterML o la legislación vigente en torno a los recursos hídricos en Colombia.

Además, esta red de ontologías comprende una nueva estructura para sistemas de información que no solo facilita la evaluación de la calidad del agua, de acuerdo a la legislación vigente, sino que gracias al uso de modelos de datos extendidos como WaterML, GeoSPARQL y O&M entre otros, permiten la integración de múltiples y heterogéneas fuentes de información relacionados con el dominio de información hidrológica.

Este desarrollo permite superar las limitaciones descritas en el capítulo del Estado del Arte en cuanto a que de este trabajo se deriva un nuevo modelo semántico que tiene en cuenta la utilización de vocabularios/ontologías estándar para el modelado del

componente espacial y de calidad de los recursos hídricos. Asimismo, aunque este modelo ha sido desarrollado para el contexto de los recursos y calidad hídrica en Colombia, fácilmente puede ser ampliable a otros contextos.

El desarrollo de la mencionada red de ontologías permite cumplir con lo establecido en el objetivo específico “Desarrollar una red de ontologías que permitan incorporar el conocimiento asociado a datos espaciales para el monitoreo de la calidad hídrica en el contexto del marco legal vigente en Colombia tras el análisis y tratamiento inicial de la información”.

Finalmente, merece destacarse que, gracias al análisis de información como series históricas de datos, legislación vigente asociada a la calidad del agua, especificaciones y estándares que brindan lineamientos sobre el modelo de datos desde el dominio hidrológico y espacial, ha sido posible generar un modelo de datos que describe el proceso de evaluación de calidad del agua y el grado de cumplimiento de la norma vigente para su materialización en una red de ontologías.

7.2. Transformación y enlazado de datos

El desarrollo de la mencionada red de ontologías permite establecer la estructura y los lineamientos a tener en cuenta para llevar a cabo la generación de RDF de las observaciones de los recursos hídricos y, de esta manera, lograr la consecución del objetivo específico relacionado con “Transformar las observaciones asociadas a recursos hídricos en información estructurada e interoperable semánticamente”.

Esta transformación de los datos originales a RDF permite incorporar un estándar que ayuda a integrar y etiquetar semánticamente los datos, de tal manera, que tanto una persona como una computadora pueden interpretar su significado, así como distinguir las relaciones entre las distintas fuentes de datos (Gonzales, 2014). En definitiva, junto a las ventajas y beneficios de utilizar RDF, descritas en (Bergman, 2009), entre las que destacan hechos como que es un estándar, abierto y expresivo; que es incrementable, ampliable y de gran adaptabilidad; el uso de RDF da a las máquinas la capacidad de

"entender" el significado semántico de las cosas en la web y la naturaleza de las relaciones entre ellas.

Junto a la transformación de los datos a RDF, y como parte de la adopción de los principios de Linked Data se han generado enlaces a otras fuentes de la nube de Linked Data, como es el caso de DBpedia. Esto ha supuesto la aplicación de las buenas prácticas para la generación de Linked Data puesto que se han identificado los recursos mediante URIs como se muestra en la sección 6.1, por medio del uso de clases incluidas en la red de ontologías desarrollada (sección 6.2), dando no solo una estructura semántica sino garantizando su interoperabilidad. Así, el RDF generado se enlazada a otras fuentes de información, como se puede ver en la sección 6.4.4, de manera que al final se tiene información espacial de dominio hidrológico en forma de Linked Data con su respectivo componente legal, siendo este el insumo necesario para el aprovechamiento y explotación posterior de los datos mediante consultas SPARQL.

Con la transformación de los datos a RDF y la generación de enlaces a otras fuentes de información de la nube de Linked Data, este trabajo se enmarca en el nuevo paradigma propuesto por OGC y W3C para la publicación de información espacial en la Web. Este paradigma ha iniciado su adopción en trabajos relacionados, como se recoge en el capítulo de Estado del Arte. Sin embargo, resultan escasas las propuestas que incluyen vocabularios extendidos a este dominio, como se presenta en (Vilches-Blázquez et al. 2014) para el caso de la ontología hydrOntology. Además, los trabajos relacionados no cuentan con la utilización de especificaciones de organismos como OGC, ya que los trabajos presentes en el estado de la cuestión no cuentan con una ontología que permitiera describir toda la información asociada a la toma de muestras y análisis físico-químico y microbiológico del agua con el fin de evaluar la calidad de la misma para definir su uso tomando como referente su componente espacial.

En definitiva, la generación de estos datos en RDF y sus correspondientes enlaces a la web de Linked Data permiten la integración entre conjuntos de datos. Además, permite el uso e incorporación de buenas prácticas y recomendaciones de organismos con amplio reconocimiento en la industria y en la academia, como son W3C y OGC. Además, esta integración de diversos y heterogéneos conjuntos de datos permite la posibilidad de

incorporar otros conjuntos de datos o servicios web, dando lugar a una legítima integración de datos.

7.3. Publicación de Linked Data

Tras la generación de RDF y enlaces, se necesitan almacenar y publicar dichos datos conforme a la metodología adoptada en este trabajo. Así, se procede a la configuración de un servidor para el almacenamiento y consulta de los datos transformados, donde cargan tanto la red de ontologías desarrollada como el RDF y los enlaces generados. Esto permite cumplir con el objetivo específico fijado para “Publicar Linked Data sobre los recursos hídricos y su calidad ambiental en el contexto colombiano”. Con estos recursos disponibles se tiene la herramienta base que proporciona el servicio para consulta de datos que facilita su explotación y aprovechamiento en el contexto legal durante la evaluación de la calidad del agua en el contexto colombiano, haciendo posible el análisis en la solución de preguntas como las presentadas en la sección 6.2.3.6, donde se presentan los usos que este sistema de información brinda.

La publicación del Linked Data generado, tal como se ha mostrado en este trabajo, permite que mediante una sola consulta se pueda hacer análisis experto, abarcando desde la simple consulta de datos históricos o parámetros analizados para una muestra, hasta la interpretación de resultados físico-químicos o microbiológicos que mediante el uso de la Web Semántica en el campo de tecnologías geoespaciales permiten conocer aquellos sitios donde un analito como la demanda bioquímica de oxígeno - DBO alcanza concentraciones que puedan afectar la salud pública o el medio ambiente, y siempre teniendo en cuenta el lugar y los límites que le aplican gracias a la aplicación de los principios de Linked Data y la Web Semántica.

Esta publicación de datos permite demostrar la generación de un nuevo aporte al conocimiento, ya que como se presentó en el estado del arte, ninguno de los trabajos antes realizados en el campo de ciencias geoespaciales aplicados a la gestión de recurso hídrico mediante Linked Data había hecho uso de vocabularios extendidos que permitieran hacer consultas espaciales desde la Web Semántica (haciendo uso de GeoSPARQL). Además, los trabajos existentes tampoco incorporan la aplicación del

contexto legal para la evaluación de la calidad del agua. Esto supone una ampliación del conocimiento en el campo de tecnologías geoespaciales al ofrecer un nuevo mecanismo para la gestión de datos que permiten evaluar el cumplimiento de una norma.

7.4. Explotación de Linked Data

Tal como se muestra en la sección 6.5, se ha generado RDF y enlazado distintas fuentes de información, tales como datos históricos, caracterización de muestreo, legislación y límites máximos permisibles, a través del uso de un lenguaje común por medio de la red de ontologías desarrolladas. Estos datos enlazados, además de enriquecer la información original, incluyendo el lugar de toma de la muestra para realizar análisis, permiten la explotación de estos datos a través de la interpretación de series históricas mediante su aprovechamiento en herramientas como paquetes estadísticos. A su vez, esto permite el análisis experto del cumplimiento de la norma, gracias a la incorporación del contexto legal, y medir el nivel de afectación de la calidad del agua que pueda tener según el uso para el cual dicho recurso pueda ser destinado.

La explotación de los datos enlazados generados en el marco de este trabajo, mediante la formulación de consultas semánticas, permite realizar análisis estadísticos que van desde la visualización de una serie de tiempo en un cuerpo hídrico a procesos más elaborados como la construcción de un diagrama de dispersión DBO vs DQO, e incluso la aplicación del contexto legal para conocer los lugares donde se ha infringido la norma para la concentración máxima de DBO, pudiendo así cumplir con el objetivo específico relacionado con “mostrar el potencial en la explotación de los datos una vez enlazados ya sea con fines estadísticos o aplicables al contexto legal”. Esta diversidad de posibilidades de explotación de los datos enlazados del proyecto permite que los resultados sean accesibles y usables por parte de un amplio perfil de usuarios. Así, usuarios con conocimientos en geomática, podrán acceder a estos resultados a través de la aplicación generada mediante la librería Leaflet, permitiendo que los resultados se visualicen a través de mapas interactivos, En este proyecto se recoge un ejemplo de dicha aplicación para la quebrada Socotá donde se visualiza el parámetro DBO (ver Figura 6-32). Asimismo, este trabajo genera resultados para usuarios con conocimientos en entornos de análisis estadístico, como es el caso de R. Este tipo de usuarios pueden obtener, por ejemplo, diagramas de dispersión DBO vs DQO, como se presenta en la Figura 6-28.

Finalmente, este trabajo genera resultados para usuarios con conocimientos en Web Semántica y Linked Data. Este tipo de usuarios pueden realizar múltiples consultas utilizando SPARQL, Entre estas consultas, por ejemplo, se encuentran las destinadas a descubrir los emplazamientos donde se han infringido la legalidad asociada a los recursos hídricos.

Considerando los diferentes perfiles de usuarios contemplados en este trabajo, se puede afirmar que se ha realizado un considerable esfuerzo para incrementar el grado de interoperabilidad, accesibilidad y usabilidad de los datos considerados en este trabajo. Además, la posibilidad de llevar a cabo estos procesos de análisis permite superar las limitaciones de los trabajos presentes en el estado de la cuestión, ya que este proceso de análisis y explotación de la calidad de los recursos hídricos utilizando Linked Data no había sido descrita en los mencionados trabajos.

7.5. Calidad del Linked Data

Teniendo en cuenta las dimensiones y métricas descritas en la sección 2.5 de este documento, se puede afirmar que se ha generado un trabajo con Linked Data de calidad si se resalta que se trata de datos que cumplen con las métricas propuestas para disponibilidad, enlazado y desempeño. Lo anterior se evidencia a lo largo de la sección 6.5, donde se muestra como en el servidor desplegado se encarga de gestionar, de manera apropiada, todos los recursos RDF generados (datos y enlaces). Este servidor se encarga de facilitar una adecuada respuesta a las consultas realizadas en distintos formatos, incluido RDF, y de proporcionar la información en una estructura adecuada, como se muestra en el archivo XML cuyo ejemplo se presenta en la Figura 7-1, resultado de la consulta expuesta en la Figura 6-27. De esta manera, se pone en evidencia el cumplimiento de lo recogido en las métricas de accesibilidad al SPARQL Endpoint y al servidor, accesibilidad al volcado de RDF y ausencia de contenidos mal reportados.

El servidor mencionado también evita la ocurrencia de enlaces rotos en los recursos RDF generados (datos y enlaces), ya que permite administrar todas las URIs de los recursos de manera que estas puedan ser consultadas desde el navegador, como se refleja en la Figura 7-2. De esta manera, con solo revisar la dirección de los mismos en las URIs donde se encuentran descritas, mediante la utilización de un navegador web, se puede

acceder a la información consignada en los datos RDF y en el OWL de la red de ontologías desarrollada (métricas de diferenciabilidad de URIs y referencia de enlaces directos).

```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
  <head>
    <variable name="Muestra"/>
    <variable name="Conductividad"/>
    <variable name="pH"/>
    <variable name="DBO"/>
    <variable name="DQO"/>
    <variable name="SolidosSupendidos"/>
    <variable name="SolidosTotales"/>
    <variable name="EColi"/>
  </head>
```

Figura 7-1. Encabezado de resultado en formato XML

subject:

predicate	object
w3.org/2000/01/rdf-schema#range	http://www.BasinRegulations.gov.co/ContaminantLevels

Showing statements for: <http://www.BasinRegulations.gov.co/ContaminantLevels>

Use resource labels in overview

Statements with this value as subject:

graph	subject	predicate	object
Default Graph	-	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Class

Statements with this value as predicate:

graph	subject	predicate	object
-- no statements found --			

Statements with this value as object:

graph	subject	predicate
	_:b0	http://www.w3.org/2002/07/owl#someValuesFrom
	_:b1	http://www.w3.org/2002/07/owl#someValuesFrom
	http://www.BasinRegulations.gov.co/ContaminantLevels/Escherichia_coli	http://www.w3.org/2000/01/rdf-schema#subClassOf
	http://www.BasinRegulations.gov.co/ContaminantLevels/pH	http://www.w3.org/2000/01/rdf-schema#subClassOf

Figura 7-2. Enlaces entre recursos desde el navegador (Clase: *ContaminantLevels*)

En cuanto al enlazado de los recursos, éstos se encuentran presentes en los datos al establecer equivalencias tanto en recursos internos; lo que permite validar la métrica relacionada con la detección de enlaces de buena calidad, reflejada en la sección 6.2.4.1 para la clase *observation*; como también se tienen enlaces a fuentes externas como DBpedia (sección 6.4.4). Estos enlaces externos permiten cumplir con la métrica de existencia de enlaces a fuentes de datos externas tal como se presenta en la Figura 7-3, donde se presenta en la parte superior el enlace desde el SPARQL Endpoint desplegado, y justo debajo el resultado de su consulta en DBpedia para el municipio de Mosquera. Así mismo es importante resaltar que todos los elementos de la red de ontologías desarrollada en este trabajo se encuentran definidas adecuadamente, lo que permite cumplir con la métrica de detección de enlaces locales, como se describe en la sección 6.2.

The image shows a screenshot of a SPARQL query result. At the top, there is a green header bar. Below it, a query result is displayed in a light blue box, showing a URL: `<gml:Point srsName='EPSG:3116' xmlns:gml='http://www.opengis.net/gml'>http://dbpedia.org/resource/Mosquera,_Cundinamarca`. An arrow points from this URL to the title of a web page: "About: Mosquera, Cundinamarca". Below the title, there is a description of Mosquera as a municipality in Colombia, and a table with properties and values.

Property	Value
<code>dbo:PopulatedPlace/areaTotal</code>	■ 107.0
<code>dbo:PopulatedPlace/areaUrban</code>	■ 7.67
<code>dbo:abstract</code>	■ Mosquera is a municipality of Colombia in the Western Savanna Province, part of the Bogotá metropolitan area, close to the capital Bogotá and is part of its metropolitan area. The urban centre is located at an altitude of 2,516 metres (8,255 ft) on the Bogotá savanna. The municipality borders the localities of Fontibón of Bogotá in the east, Soacha in the south, Madrid and Funza in the north and Bojacá in the west. (en)

Figura 7-3. Enlace de municipio de Mosquera a DBPedia

Por último, con respecto al desempeño, éste se ha garantizado mediante el uso de URIs-slash tal como se muestra en la sección 6.1, lo que permite cumplir con su métrica asociada. Por otro lado, tal como se expuso en la sección 6.5.1, se hizo uso del SPARQL Endpoint de *Parliament*, el cual tal como expone en su sitio web⁵⁷, incluye un motor de reglas de alto desempeño que sirve como medio eficiente de inferencia, permitiendo así realizar consultas a gran velocidad incluso para grandes volúmenes de datos como los mostrados en la sección 6.5.2 y 6.5.3. Esto permite cumplir con las métricas de baja latencia, alto desempeño y escalabilidad de la fuente de datos, evidenciadas en los resultados de tiempo se pueden observar en la Figura 7-4, donde para una serie de 10 consultas no se genera un retraso mayor a 0.37 segundos, con un promedio de 0.32.

```
> Performance<-c()
> for (i in c(1:10)){
+   start_time <- Sys.time()
+   consulta<-SPARQL(endpoint, query2, format = "xml")
+   end_time <- Sys.time()
+   Performance<-c(Performance,end_time - start_time)
+ }
> Performance<-round(Performance,digits = 2)
> Performance ###Duración de 10 consultas en segundos
[1] 0.31 0.32 0.30 0.31 0.32 0.30 0.33 0.33 0.37 0.29
> round(mean(Performance),digits = 2) ###Promedio en segundos de 10 consultas
[1] 0.32
```

Figura 7-4. Resultados de desempeño para 10 consultas desde R en segundos

Considerando el conjunto de resultados mostrados con anterioridad y la relación de los mismos conforme a las dimensiones y métricas más utilizadas para la evaluación de la calidad del Linked Data, nuevamente, se puede afirmar que se ha generado un trabajo con Linked Data de calidad.

⁵⁷ <http://parliament.semwebcentral.org/>

Capítulo 8. Conclusiones y trabajo futuro

Una vez culminado y desarrollado el trabajo al presentar una propuesta de Linked Data de datos asociados al agua y su contexto espacio-temporal y legal junto con su plataforma de aplicación, se presentan las conclusiones como producto de los resultados obtenidos junto con los trabajos futuros que permitan dar continuidad al trabajo desarrollado.

8.1. Conclusiones

En este trabajo se realizó una recopilación inicial de toda la información considerada en el contexto de este trabajo en sus múltiples formatos. Tras su recopilación se procedió a realizar un proceso de armonización y limpieza, donde se toman datos en hojas de Microsoft Excel y se procede a la remoción de columnas innecesarias, datos redundantes, etc. Además, para el componente legal de los datos considerados se realizaron capas espaciales mediante la interpretación de la norma incluyendo las magnitudes permitidas y sus fronteras de aplicación.

Asimismo, se procedió al desarrollo de una red de ontologías, tomando como referencia diversos escenarios propuestos en la metodología NeOn. Para el desarrollo de esta red se tomaron como insumo recurso no ontológicos (hojas de Excel, WaterML 2.0, legislación aplicable) y ontológicos (O&M, SSN, GeoSPARQL).

Una vez generada la red de ontologías de este trabajo, se transformó la información original a RDF, generando enlaces de los datos de los municipios donde se localizan las muestras a otras fuentes de datos de la Web de Linked Data, concretamente a DBpedia. Esto permite la integración de múltiples y heterogéneas fuentes de información, no solo desde el punto de vista sintáctico (XLS, PDF, etc.), sino también semántico (resultados de laboratorio, análisis hidrológico, legislación, estándares).

Asimismo, toda la información resultante (datos RDF y red de ontologías) fue almacenada en un *triple store* con soporte espacial, lo que ha permitido tener un nodo de consulta a los datos mediante el lenguaje SPARQL. Este punto de consulta facilita su explotación en un aplicativo que posibilita no solo realizar análisis histórico y espacial de los datos, sino también legal de acuerdo a los tipos de consumo definidos para región.

Teniendo en cuenta lo desarrollado se puede concluir que la Web Semántica resulta adecuada para describir el dominio asociado al recurso hídrico relacionado con la calidad del mismo, ya que permite incorporar recursos de distinta naturaleza (ontológicos y no ontológicos), facilitando la solución de problemas de tipo semántico mediante la construcción de una red de ontologías.

Gracias a la infraestructura que soportan la red de ontologías desarrollada así como el RDF generado sobre la información asociada a calidad del agua, resulta claro el beneficio de hacer uso de los principios del Linked Data al facilitar su interoperabilidad semántica, como queda demostrado al ver la forma en que el RDF puede conectarse con otras fuentes del mismo tipo o incluso la posibilidad de transformar esta información a otros formatos (HTML, CSV, XML), con el fin de ser aprovechado por otras herramientas como se hizo en este caso con R e incluso con visualizadores de información geoespacial como Leaflet.

Por otro lado, el desarrollo de este trabajo ha permitido generar una nueva tecnología que proporciona un uso más eficiente a los metadatos facilitando la explotación de los datos mediante el enlace de los mismos, ya sea a través de un análisis estadístico como curva de regresión DBO vs o DQO o del monitoreo de la demanda bioquímica de oxígeno y el grado incumplimiento a lo largo de la cuenca del río Bogotá.

De igual manera queda demostrado el beneficio de la interoperabilidad semántica al incorporar las clases definidas por URIs donde gracias a la adecuada documentación de dichas clases y la aplicación de estándares y especificaciones como el WaterML 2.0 y SSN entre otros, se identifica y describe semánticamente cada dato partiendo de información como el parámetro analizado, unidades de trabajo, fecha, lugar y tipo de muestra entre otras.

8.2. Trabajo futuro

Como trabajo futuro se abordará la modelación hidrológica que mediante la abstracción de los fenómenos descritos en la red de ontologías y consignados en los datos enlazados permitan conocer el estado del recurso hídrico en este caso en lugares o momentos donde no se dispone de información. De manera que con el sistema de información diseñado para la línea base del recurso hídrico que acá se presenta queda pendiente incorporar el contexto asociado a la modelación hidrológica espacio-temporal junto con las variables que así lo requieran según el modelo a desarrollar.

Por otro lado, considerando que la evaluación de la calidad del agua puede ser evaluada teniendo en cuenta indicadores hidrobiológicos, tales como conteo de macro-invertebrados, se prevé extender el presente trabajo y la red de ontologías desarrollada para incorporar en el modelo semántico estructuras que incluyan este tipo de información.

Así mismo durante el desarrollo del proyecto se trabajó con el único documento entonces disponible sobre WaterML 2.0 (Parte 1: Series de tiempo). Sin embargo, a la fecha se encuentran disponibles otros documentos asociados a instrumentación y agua subterránea que serán evaluados para ser incorporados y ampliar la red de ontologías desarrollada.

Bibliografía

- Abele, A., & McCrae, J. (2017). The Linking Open Data cloud diagram. Retrieved November 19, 2017, from <http://lod-cloud.net/>
- Ahmedi, L., Jajaga, E., & Ahmedi, F. (2013). An Ontology Framework for Water Quality Management. *Semantic Sensor Networks 2013 (SSN2013)*.
- Auer, S., Lehmann, J., & Hellmann, S. (2009). LinkedGeoData: Adding a Spatial Dimension to the Web of Data. In *ISWC 2009: The Semantic Web - ISWC 2009* (pp. 731–746). Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-642-04930-9_46
- Auer, S., Lehmann, J., & Ngomo, A. N. (2011). Introduction to Linked Data and Its Lifecycle on the Web. *Reasoning Web Summer School*, 1–75.
- Barnaghi, P., Presser, M., & Moessner, K. (2010). Publishing Linked Sensor Data. In *SSN'10 Proceedings of the 3rd International Conference on Semantic Sensor Networks*. Retrieved from <http://epubs.surrey.ac.uk/470673/1/sense2web.pdf>
- Battle, R., & Kolas, D. (2011). Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web – Interoperability, Usability, Applicability an IOS Press Journal*. Retrieved from http://www.semantic-web-journal.net/sites/default/files/swj176_3.pdf
- Behmel, S., Damour, M., Ludwig, R., & Rodriguez, M. J. (2016). Water quality monitoring strategies — A review and future perspectives. *Science of The Total Environment*, 571, 1312–1329. <http://doi.org/10.1016/j.scitotenv.2016.06.235>
- Bergman, M. (2009). Advantages and Myths of RDF - Advantages and Myths of RDF A 10th Birthday Salute to RDF's Role in Powering Data Interoperability. Retrieved from <http://www.mkbergman.com>
- Berners-Lee, T. (1996). Universal Resource Identifiers -- Axioms of Web architecture. Retrieved November 22, 2017, from <https://www.w3.org/DesignIssues/Axioms.html#uri>
- Berners-Lee, T. (1998). Semantic Web roadmap. Retrieved November 8, 2017, from <https://www.w3.org/DesignIssues/Semantic.html>
- Berners-Lee, T. (2000). Semantic Web - XML2000. Retrieved November 19, 2017, from <https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>
- Berners-Lee, T. (2001). The semantic web. *SCIENTIFIC AMERICAN*, (April). Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+semantic+web:+a+new+form+of+web+content+that+is+meaningful+to+computers+will+unleash+a+>

revolution+of+new+possibilities#0

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. Retrieved from <http://www.igi-global.com/article/linked-data-story-far/37496>
- Bogdanovi, M., Stanimirovi, A., & Stoimenov, L. (2015). Methodology for geospatial data source discovery in ontology-driven geo-information integration architectures. *Journal of Web Semantics*, 32, 1–15. <http://doi.org/10.1016/j.websem.2015.01.002>
- Brickley, D., & Guha, R. V. (2014). RDF Schema 1.1. Retrieved November 19, 2017, from <https://www.w3.org/TR/rdf-schema/>
- Buccella, A., Cechich, A., & Fillottrani, P. (2009). Ontology-driven geographic information integration: A survey of current approaches. *Computers & Geosciences*, 35(4), 710–723. <http://doi.org/10.1016/j.cageo.2008.02.033>
- Celino, I. (2013). Human Computation VGI Provenance: Semantic Web-Based Representation and Publishing. *IEEE Transactions on Geoscience and Remote Sensing*, 51(11), 5137–5144. <http://doi.org/10.1109/TGRS.2013.2252015>
- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherso, J. (2016). shiny: Web Application Framework for R. Comprehensive R Archive Network (CRAN).
- Cheng, J., Xie, Y., & Wickham, H. (2016). leaflet: Create Interactive Web Maps with the JavaScript “Leaflet” Library. Comprehensive R Archive Network (CRAN).
- Compton, M., Bamaghi, P., Bermudez, L., Castro, R. G., Corcho, O., Cox, S., ... Taylor, K. (2005). The SSN Ontology of the Semantic Sensor Networks Incubator Group. Retrieved from https://www.w3.org/2005/Incubator/ssn/wiki/images/f/f3/SSN-XG_SensorOntology.pdf
- Cox, S. J. D. (2013). An explicit OWL representation of ISO/OGC Observations and Measurements. *Proceedings of the 6th International Workshop on Semantic Sensor Networks Co-Located with the 12th International Semantic Web Conference (ISWC 2013)*, 1–18. Retrieved from <http://ceur-ws.org/Vol-1063/paper1.pdf>
- CUAHSI. (2010). CUAHSI-HIS. Retrieved May 12, 2013, from <http://his.cuahsi.org/ontologyfiles.html>
- Curry, E., Degeler, V., Clifford, E., Coakley, D., & Costa, A. (2014). Linked Water Data For Water Information Management Recommended Citation. *11th International Conference on Hydroinformatics*. Retrieved from http://academicworks.cuny.edu/cc_conf_hic
- Defra - Department for Environment, F. and R. A. (2014). Defra Open Data Strategy, (March).
- Diersing Florida Keys National Marine Sanctuary, N. (2017). Water Quality: Frequently Asked Questions. Retrieved from <https://nmsfloridakeys.blob.core.windows.net/floridakeys-prod/media/archive/scisummaries/wqfaq.pdf>
- DuCharme, B. (2011). *Learning Sparql*. (S. St. Laurent, Ed.). O'Reilly Media. Retrieved from http://books.google.com/books?hl=en&lr=&id=GVQcFBh4ebgC&oi=fnd&pg=PR5&dq=Learning+Sparql&ots=GbcvN_NHPn&sig=ogXdoemO00-GghWn9lg3XxQp4yw

- Fallahi, G. R., Frank, A. U., Mesgari, M. S., & Rajabifard, A. (2008). An ontological structure for semantic interoperability of GIS and environmental modeling. *International Journal of Applied Earth Observation and Geoinformation*, 10(3), 342–357. <http://doi.org/10.1016/j.jag.2008.01.001>
- Feliachi, A., Abadie, N., Hamdi, F., & Ateazing, G. A. (2013). Interlinking and Visualizing Linked Open Data with Geospatial Reference Data. Retrieved from <http://datalift.org/>
- Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*. Retrieved from <http://oa.upm.es/5484/>
- Fonseca, F., Egenhofer, M., Jr., C. A. D., & Borges, A.V., K. (2000). Ontologies and knowledge sharing in urban GIS. *Computer, Environment and Urban Systems*, 24(3), 232–251. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract>
- Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2009). NeOn Methodology for Building Ontology Networks: a Scenario-based Methodology. Retrieved from http://oa.upm.es/5475/1/INVE_MEM_2009_64399.pdf
- Gonzales, B. M. (2014). Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. *Information Technology and Libraries*, 33(4), 10–22. <http://doi.org/10.6017/ital.v33i4.5631>
- Goodwin, J., Dolbear, C., & Hart, G. (2009). Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. *Transactions in GIS*, 12(s1), 19–30. <http://doi.org/10.1111/j.1467-9671.2008.01133.x>
- Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, (April). Retrieved from <http://secs.ceas.uc.edu/~mazlack/ECE.716.Sp2011/Semantic.Web.Ontology.Papers/Gruber.93a.pdf>
- Guarino, N. (1997). Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. *Information Extraction A Multidisciplinary Approach to an Emerging Information Technology*, 1299, 139–170. Retrieved from http://link.springer.com/chapter/10.1007/3-540-63438-X_8
- Guarino, N. (1998). Formal Ontology and Information Systems. *Formal Ontology in Information Systems*, (June), 3–15. Retrieved from <http://fois08.dfki.de/proceedings/FOIS08PosterProceedings.pdf>
- Haase, P., Rudolph, S., & Wang, Y. (2006). NeOn: Lifecycle Support for Networked Ontologies. Retrieved from http://www.neon-project.org/deliverables/WP1/NeOn_2006_D1.1.1.pdf
- Heath, T., & Bizer, C. (2011). *Linked Data Evolving the Web into a Global Data Space*. Morgan & Claypool. <http://doi.org/10.4018/978-1-60960-593-3.ch008>
- Hebeler, J., Fisher, M., Blace, R., & Perez-Lopez, A. (2009). *Semantic Web Programming. Journal of Verbal Learning and Verbal Behavior* (Vol. 20). [http://doi.org/10.1016/S0022-5371\(81\)90569-7](http://doi.org/10.1016/S0022-5371(81)90569-7)
- Hess, C. (2012). Linked Data in Libraries. Retrieved September 27, 2017, from

- <https://www.slideshare.net/carlstephenhess/linked-data-in-libraries>
- Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., & Polleres, A. (2012). An empirical survey of Linked Data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14, 14–44. <http://doi.org/10.1016/J.WEBSEM.2012.02.001>
- Hunter, J., Becker, P., Alabri, A., van Ingen, C., & Abal, E. (2011). Using Ontologies to Relate Resource Management Actions to Environmental Monitoring Data in South East Queensland. *International Journal of Agricultural and Environmental Information Systems*, 2(1), 1–19. <http://doi.org/10.4018/jaeis.2011010101>
- International Organization for Standardization. (2015). ISO 19162:2015 - Geographic information -- Well-known text representation of coordinate reference systems. Retrieved October 21, 2017, from <https://www.iso.org/standard/63094.html>
- Janowicz, K., & Hitzler, P. (2012). The Digital Earth as Knowledge Engine. *Semantic Web – Interoperability, Usability, Applicability an IOS Press Journal*, 0, 1–0. Retrieved from https://geog.ucsb.edu/~jano/Semantics_Digital_Earth2012.pdf
- Janowicz, K., Scheider, S., Pehle, T., & Hart, G. (2012). Geospatial Semantics and Linked Spatiotemporal Data – Past, Present, and Future. *Semantic Web*, 0, 1–13. Retrieved from <http://iospress.metapress.com/index/M065W1130043W3P4.pdf>
- Kämpgen, B., Riepl, D., & Klinger, J. (2014). SMART Research using Linked Data – Sharing Research Data for Integrated Water Resources Management in the Lower Jordan Valley. *Semantic Publishing*. Retrieved from <http://ceur-ws.org/Vol-1155/paper-04.pdf>
- Koubarakis, M., Kyzirakos, K., Nikolaou, C., Garbis, G., Bereta, K., Smeros, P., ... Karpathiotaki, M. (2014). Linked Earth Observation Data: The Projects TELEIOS and LEO. In *In the Linking Geospatial Data Conference (LGD 2014)*. Retrieved from <http://cgi.di.uoa.gr/~koubarak/publications/lgd2014.pdf>
- Kuhn, W. (2005). Geospatial Semantics: Why, of What, and How? (pp. 1–24). Springer, Berlin, Heidelberg. http://doi.org/10.1007/11496168_1
- Lehmann, J., Athanasiou, S., Both, A., Rojas, A. G., Giannopoulos, G., Hladky, D., ... Zaslowski, V. (2015). Managing Geospatial Linked Data in the GeoKnow Project. *The Semantic Web in Earth and Space Science. Current Status and Future Directions*. Retrieved from http://jens-lehmann.org/files/2015/ios_geoknow_chapter.pdf
- Lenat, D. (1998). The dimensions of context-space. *Onto-Med Seminar SS2005*, (512). Retrieved from <http://www.onto-med.de/Archiv/ontomed2002/de/lehre/ont-eng-2005-ss/protectedFiles/lenat-d-1998--a.pdf>
- Longley, P. A., Goodchild, M. F., Maguire, J. D., & Rhind, D. W. (2005). *Data quality parameters. In: Geographical information systems: principles, techniques, management, and applications*. Wiley. Retrieved from <https://www.wiley.com/en-us/Geographical+Information+Systems%3A+Principles%2C+Techniques%2C+Management+and+Applications%2C+2nd+Edition%2C+Abridged-p-9780471735458>
- Lozano, A. (2001). I Jornadas de Ingeniería Web' 01 Ontologías en la Web Semántica. Retrieved from <http://eolo.cps.unizar.es/docencia/MasterUPV/Articulos/Ontologias en la Web Semantica.pdf>
- Myers, D. N., & Ludtke, A. S. (2017). Chapter Two – Progress and Lessons Learned from

- Water-Quality Monitoring Networks. In *Chemistry and Water* (pp. 23–120). <http://doi.org/10.1016/B978-0-12-809330-6.00002-7>
- Neteler, M., & Mitasova, H. (2004). *Open Source GIS: A GRASS GIS Approach* (Vol. 3). Springer. Retrieved from http://books.google.com/books?id=Qvp9iFg_WPEC&pgis=1
- Noy, N., McGuinness, D., & Hayes, P. J. (2005). Semantic Integration & Interoperability Using RDF and OWL. Retrieved April 15, 2017, from <https://www.w3.org/2001/sw/BestPractices/OEP/SemInt/>
- OGC. (2012a). OGC Geography Markup Language (GML) — Extended schemas and encoding rules. Retrieved from https://portal.opengeospatial.org/files/?artifact_id=46568
- OGC. (2012b). OGC GeoSPARQL - A Geographic Query Language for RDF Data. Retrieved from http://portal.opengeospatial.org/files/?artifact_id=44722
- OGC, Li, K.-J., Lee, J., Kolbe, T. H., Zlatanova, S., Morley, J., ... Becker, T. (2014). *WaterML 2.0: Part 1- Timeseries. OGC Member approved international standard* (Vol. Version 1). <http://doi.org/http://www.opengeospatial.org/>
- Omitola, T., Koumenides, C. L., Popov, I. O., Yang, Y., Salvadores, M., Szomszor, M., ... Shadbolt, N. (2010). Put in Your Postcode, Out Comes the Data: A Case Study. In *7th Extended Semantic Web Conference* (pp. 318–332). Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-642-13486-9_22
- OMS. (2017). *Guías de la OMS para la calidad del agua potable. WHO. World Health Organization.* Retrieved from http://www.who.int/water_sanitation_health/dwq/guidelines/es/
- ONU, O. de las N. U. (2015). Decenio Internacional para la Acción “El agua, fuente de vida” 2005-2015. Áreas temáticas: Calidad del agua.
- Open Geospatial Consortium. (2014). *WaterML 2.0: Part 1- Timeseries.*
- OWL Working Group. (2012). OWL - Semantic Web Standards. Retrieved October 21, 2017, from <https://www.w3.org/OWL/>
- Patni, H., Henson, C., & Sheth, A. (2010). Linked Sensor Data. In *Collaborative Technologies and Systems (CTS 2010)*. Retrieved from <http://knoesis.org/node/1705>
- Pinto, H. S., Pinto, H. S., Staab, S., & Tempich, C. (2004). DILIGENT: Towards a fine-grained methodology for Distributed, Loosely-controlled and evolving Engineering of ontologies. IN *PROCEEDINGS OF THE 16TH EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE (ECAI, 393--397.* Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.5.8078>
- Poveda, M. (2010). *METODOLOGÍA NEON APLICADA A LA REPRESENTACIÓN DEL CONTEXTO.* Retrieved from http://oa.upm.es/7936/1/TESIS_MASTER_MARIA_POVEDA_VILLALON.pdf
- R Development Core Team. (2016). R: A Language and Environment for. Statistical Computing. Vienna, Austria: The R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org>
- Rula, A., & Zaveri, A. (2014). Methodology for Assessment of Linked Data Quality. *CEUR Workshop Proceedings* . Retrieved from <http://ceur-ws.org/Vol-1215/paper-04.pdf>

- Sánchez, V., & Guiza, B. (1989). *Glosario de terminos sobre medio ambiente*. Unesco - PNUMA. Retrieved from <http://unesdoc.unesco.org/images/0008/000855/085533SB.pdf>
- Staab, S., Studer, R., Schnurr, H.-P., & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1), 26–34. <http://doi.org/10.1109/5254.912382>
- Suárez-Figueroa, M. (2010a). NeOn Methodology for building ontology networks: specification, scheduling and reuse, (June). Retrieved from <http://oa.upm.es/3879>
- Suárez-Figueroa, M. (2010b). NeOn Methodology for building ontology networks: specification, scheduling and reuse, (June).
- Suárez-Figueroa, M. C. (2017). The terms (and their definitions) included in the NeOn Glossary are. Retrieved November 8, 2017, from <http://mayor2.dia.fi.upm.es/oeg-upm/files/pdf/NeOnGlossary.pdf>
- Suárez-Figueroa, M. C., Gómez-Pérez, A., Motta, E., & Gangemi, A. (2012). Introduction: Ontology Engineering in a Networked World. In *Ontology Engineering in a Networked World* (pp. 1–6). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-24794-1_1
- Suárez-Figueroa, M. C., Gómez-Pérez, A., & Villazón-Terrazas, B. (2008). Ontology Requirements Specification. Retrieved from <http://www.neon-project.org/web-content/media/book-chapters/Chapter-05.pdf>
- Usery, E. L., & Varanka, D. (2012). Semantic Web - On linked spatiotemporal data and geo-ontologies. *Semantic Web*, 3(4), 371–384. Retrieved from <https://dl.acm.org/citation.cfm?id=2590212>
- van Hage, W. R. (2013a). SPARQL: SPARQL client. Retrieved from <https://cran.r-project.org/web/packages/SPARQL/index.html>
- van Hage, W. R. (2013b). SPARQL: SPARQL client.
- Varanka, D. (2008). National Topographic Modeling, Ontology-Driven Geographic Information in the Context of The National Map. In *First International Workshop on Information Semantics and its Implications for Geographical Analysis (ISGA '08) at GIScience 2008*. Retrieved from <https://portal.ikw.uni-osnabrueck.de/~isga08/Varanka.pdf>
- Vilches-Blázquez, L. (2009). An approach to comparing different ontologies in the context of hydrographical information. *Information Fusion and Geographic Information Systems*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-00304-2_13
- Vilches-Blázquez, L. (2011). *Metodología para la integración basada en ontologías de información de bases de datos heterogéneas en el dominio hidrográfico*. Retrieved from <http://oa.upm.es/7625/>
- Vilches-Blázquez, L. M., Villazón-Terrazas, B., Corcho, O., & Gómez-Pérez, A. (2014). Integrating geographical information in the Linked Digital Earth. *International Journal of Digital Earth*, 7(7), 554–575. <http://doi.org/10.1080/17538947.2013.783127>
- Villazón-Terrazas, B., Vilches-Blázquez, L. M., Corcho, O., & Gómez-Pérez, A. (2011). Methodological guidelines for publishing government linked data. *Linking*

- Government Data*, 1–23.
- Visser, U., & Stuckenschmidt, H. (2002). Ontologies for geographic information processing. *Computers & Geosciences*, 28, 103–117. Retrieved from <http://www.sciencedirect.com/science/article/pii/S009830040100019X>
- W3C. (2017). RDF - Semantic Web Standards. Retrieved September 24, 2017, from <https://www.w3.org/RDF/>
- W3C Semantic Sensor Network Incubator Group. (2009). Semantic Sensor Network Ontology. Retrieved October 21, 2017, from <https://www.w3.org/2005/Incubator/ssn/ssnx/ssn>
- Yaser Bishr, A., Pundt, H., Kuhn, W., & Radwan, M. (1999). *Interoperating Geographic Information Systems*. (M. Goodchild, M. Egenhofer, R. Fegeas, & C. Kottman, Eds.). Boston, MA: Springer US. <http://doi.org/10.1007/978-1-4615-5189-8>
- Yuan, J., Yue, P., Gong, J., & Zhang, M. (2013). A Linked Data Approach for Geospatial Data Provenance. *IEEE Transactions on Geoscience and Remote Sensing*, 51(11), 5105–5112. <http://doi.org/10.1109/TGRS.2013.2249523>
- Zhao, P., Foerster, T., & Yue, P. (2012). The Geoprocessing Web. *Computers & Geosciences*, 47, 3–12. <http://doi.org/10.1016/j.cageo.2012.04.021>

Anexo I

Mecanismo de consulta de datos mediante SPARQL en R

Las consultas mediante el lenguaje SPARQL, aunque pueden ser realizadas en el entorno web del SPARQL Endpoint generado, también tiene el potencial de ser consumido, mediante otras herramientas como R, que sirve de interfaz gracias a la librería *SPARQL*, cuyo uso se explica a continuación tomando de referencia la consulta utilizada en la Figura 6-27:

Las consultas se realizan con la función “*SPAQRL*” con los siguientes parámetros:

url: La URL del servidor Endpoint.

query: La consulta a realizar mediante el lenguaje SPARQL.

format: El formato en que se requiere la consulta (XML, CSV, TSV).

Aunque la función posee otros argumentos, estos han sido los requeridos para la petición de datos al servidor.

A continuación, se presenta un ejemplo tomado de la datos del proyecto con fines ilustrativos:

```

consulta <-
  SPARQL(url="http://127.0.0.1:8089/parliament/sparql",
  query="PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-
  ns#>

  PREFIX observation:
  <http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation
  #>

  PREFIX wml: <http://WaterML.org/>

  PREFIX basic:
  <http://def.seegrid.csiro.au/isotc211/iso19103/2005/basic#>

  SELECT DISTINCT ?Muestra ?Conductividad ?pH ?DBO ?DQO
  ?SolidosSuspendidos ?SolidosTotales ?EColi WHERE {

  ?s wml:SampleNumber ?Muestra ;

```

```

wml:Conductivity ?Conduct ;
wml:pH ?phnode;
wml:DBO ?dbonode;
wml:DQO ?dqonode;
wml:ST ?stnode;
wml:SS ?ssnode;
wml:E.Coli ?ecolinode.

?Conduct observation:value ?Conductividad.
?phnode observation:value ?pH.
?dbonode observation:value ?DBO.
?dqonode observation:value ?DQO.
?ssnode observation:value ?SolidosSuspendidos.
?stnode observation:value ?SolidosTotales.
?ecolinode observation:value ?EColi.

}"))

```

En la variable “*consulta*” se almacena una lista con dos objetos uno de los cuales tiene por nombre “*results*”, por tanto, solo resta consultar los datos en dicho objeto dentro de la lista:

```

resultados <- consulta$results

```

Estos resultados generan un data frame, el cual se presentan en la Figura 10-1. Resultados de query en SPARQL como data frame en la Figura 10-1.

	Muestra	Conductividad	pH	DBO	DQO	SolidosSuspendidos	SolidosTotales	EColi
1	4114-13	247.0	6.9	0.0	37.7	35.5	163.0	8.2e+05
2	4113-13	250.0	6.7	22.6	55.2	13.0	108.0	7.5e+04
3	4112-13	269.0	7.1	52.4	53.2	46.3	200.0	1.3e+06
4	4100-13	359.0	7.1	54.4	164.0	60.0	248.0	1.7e+07
5	4098-13	243.0	7.0	4.5	31.5	26.7	164.0	2.3e+04
6	4097-13	607.0	6.8	14.8	156.0	115.0	572.0	4.0e+04
7	4096-13	921.0	7.3	261.0	445.0	350.0	766.0	1.7e+07
8	4095-13	265.0	6.9	6.2	33.6	36.0	188.0	8.2e+04
9	4049-13	232.0	6.8	8.2	41.1	41.9	174.0	1.3e+05
10	4047-13	65.3	6.6	0.0	23.0	23.3	99.0	1.3e+03
11	4046-13	257.0	6.9	14.5	52.0	58.0	205.0	3.9e+05
12	4045-13	248.0	7.0	13.9	48.2	42.0	183.0	9.6e+04
13	4043-13	224.0	6.7	5.9	28.2	27.1	164.0	3.4e+04
14	4036-13	209.0	7.0	5.1	28.7	53.3	167.0	1.3e+04
15	4035-13	193.0	7.1	3.3	28.9	9.0	144.0	9.7e+02
16	4034-13	194.0	7.0	9.7	20.3	53.3	169.0	2.9e+04
17	4033-13	112.0	7.0	4.1	27.5	44.3	113.0	1.4e+04

Showing 1 to 17 of 4,541 entries

Figura 10-1. Resultados de query en SPARQL como data frame

Anexo II

Funciones de aplicativo de consulta de datos

Para realizar el adecuado análisis histórico y temporal, incorporando el componente legal asociado a los tipos de consumo definidos para cada región se debe realizar, en primer lugar, la siguiente consulta:

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX observation:
<http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#>
3 PREFIX wml: <http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#>
4 PREFIX geo: <http://www.opengis.net/ont/geosparql#>
5 PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
6 PREFIX units: <http://www.opengis.net/def/uom/OGC/1.0/>
7 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
8 PREFIX sampling:
<http://def.seegrid.csiro.au/isotc211/iso19156/2011/sampling#>
9 PREFIX regulation: <http://www.waterregulations.gov.co/>
10 PREFIX citation: <http://def.seegrid.csiro.au/isotc211/iso19115/2003/citation#>
11 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
12
13 SELECT DISTINCT ?Muestra ?Fecha ?clase ?DBO ?DBOLimit ?lat ?long
14 WHERE {
15 ?Muestra citation:date ?Fecha.
16 ?unidad a regulation:HydrologicUnit.
17 ?unidad regulation:hasWaterClass ?clase.
18 ?clase
<http://www.waterregulations.gov.co#hasBiochemicalOxygenDemandLimit>
?DBOLimit.
19 ?unidad geo:asWKT ?geounidad.
20 ?punto a sampling:SamplingPoint;
21 <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat;
22 <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long.
23 ?Muestra geo:hasGeometry ?punto.
```

```
24?Muestra <http://WaterML.org/DBO> ?dbonode.  
25?dbonode observation:value ?DBO.  
26?punto rdfs:label ?nombre.  
27?punto geo:asWKT ?geopunto.  
28FILTER ( geof:sfContains(?geounidad, ?geopunto) &&  
29?DBO > ?DBOLimit &&  
30?Fecha = "2013-01-01 00:00:00"^^xsd:dateTime)  
31}
```

De esta manera, se obtienen los datos asociados a DBO, resaltando que en las líneas 28-30 se hace un filtro donde se asocia cada punto a la unidad con límite máximo permitido (línea 28), utilizando GeoSPARQL con el fin de incorporar el componente espacial, para después establecer en la consulta condiciones adicionales que deben cumplir los datos para identificar “datos que exceden la norma según su uso establecido” (línea 29) y “datos con fecha anterior al 2013” (línea 30).

Una vez realizada la consulta y cargada la información de los puntos que presentan incumplimiento de la norma en un momento específico, con el fin de brindar el componente temporal, se puede generar una visualización utilizando el entorno R y Leaflet a través de un entorno web como el que se observa en la Figura 11-1, donde se ven marcados con puntos los lugares de muestreo que incumplen la norma y así proporcionar un análisis espacio-temporal y legal de los datos considerados.

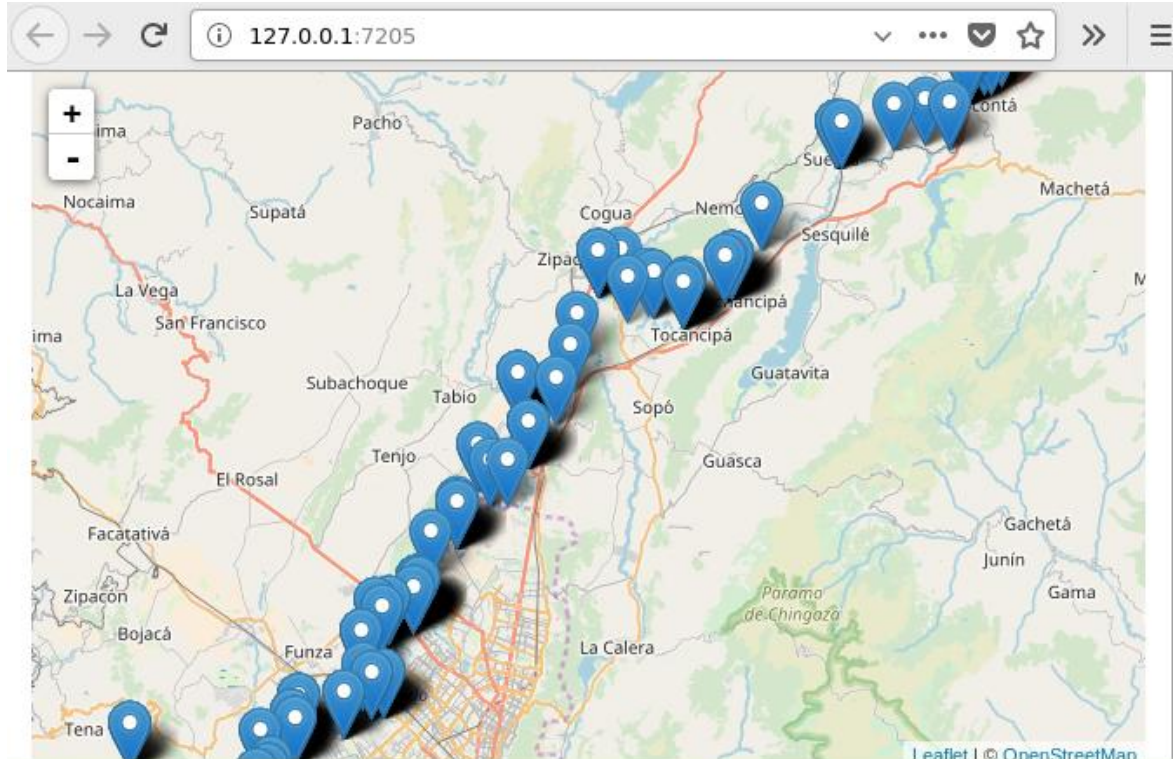



Figura 11-1. Entorno web de consulta

Como se puede observar en dicha figura, en el panel superior izquierdo se encuentra dos botones (+ y -) que permiten acercar o alejar el mapa respectivamente. Sin embargo, con las teclas “+” y “-” del teclado del usuario, así como con el rodillo del mouse se tiene el mismo efecto.

Por otro lado, con las flechas del teclado del usuario () , así como arrastrando con el botón derecho sostenido del mouse del computador del usuario es posible navegar mediante acciones de desplazamiento (horizontal/vertical) en el mapa. También si se tiene sostenida la tecla shift del teclado del usuario y se dibuja un recuadro con el cursor sostenido es posible realizar acercamientos a una determinada área de interés.

Por último, la aplicación desarrollada permite consultar la concentración obtenida en cada punto haciendo *click* con el botón derecho del mouse del computador del usuario sobre un sitio de interés, tal como se muestra en la Figura 11-2:

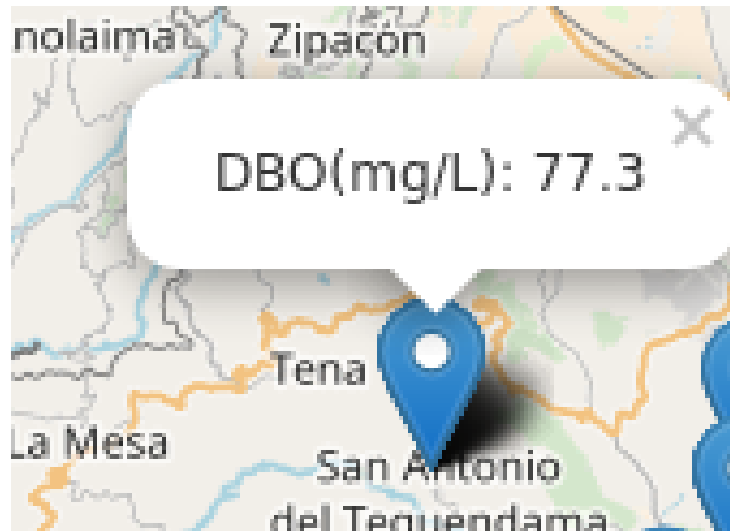


Figura 11-2. Consulta de concentración en punto de interés.