

▶ Guía de Transformación de
Datos Abiertos a Datos Abiertos
Enlazados

OD a LOD



Guía de Transformación de Datos Abiertos a Datos Abiertos Enlazados

OD a LOD

1.	INTRODUCCIÓN	1
2.	ACLARACIONES Y CONCEPTUALIZACIÓN	2
	DATOS ABIERTOS	2
	DATOS ABIERTOS ENLAZADOS (LOD)	3
	URI - UNIFORM RESOURCE IDENTIFIER (IDENTIFICADOR UNIFORME DE RECURSOS)	4
	ONTOLOGÍA	6
	METAVOCABULARIO	7
3.	TRANSFORMACIÓN DE OD A LOD	8
3.1.	IDENTIFICACIÓN DE CONJUNTOS DE DATOS (DATASETS)	10
3.1.1.	IDENTIFICACIÓN Y ANÁLISIS DE FUENTES DE DATOS	10
3.1.2.	ANÁLISIS DE DATOS INICIALES	11
3.1.3.	IDENTIFICAR COLABORADORES E INTERESADOS	12
3.1.4.	DEFINICIÓN DE LICENCIA	12
3.1.5.	DISEÑAR EL IDENTIFICADOR UNIFORME DE RECURSOS (URI) PARA LA IDENTIFICACIÓN Y LOCALIZACIÓN DE LOS DATOS I3	13
3.2.	LIMPIEZA DE DATOS	14
3.3.	MODELADO DE DATOS	15
3.3.1.	BUSCAR VOCABULARIOS	16
3.3.2.	CREAR VOCABULARIOS	19
3.3.3.	CREAR ONTOLOGÍA	21
	CICLO DE VIDA PARA LA GENERACIÓN DE ONTOLOGÍAS	22
3.4.	ENRIQUECIMIENTO DE DATOS	24
3.5.	VINCULACIÓN DE LOS DATOS	25
3.5.1.	REPRESENTACIÓN DE DATOS	25
3.5.2.	BUSCAR VOCABULARIOS	26
3.5.3.	CREAR TRIPLETAS	26
3.5.4.	ENRIQUECER REPRESENTACIÓN	26
3.6.	PUBLICACIÓN	27
3.6.1.	PREPARACIÓN DE LOS DATOS	27
3.6.2.	PUBLICACIÓN DEL CONJUNTO DE DATOS	28
3.6.3.	PUBLICACIÓN DE METADATOS	28
3.6.4.	HABILITAR DESCUBRIMIENTO EFECTIVO	29
3.6.5.	CONSTRUCCIÓN DE SENTENCIAS SPARQL	30
3.7.	VALIDACIÓN	30



4. GLOSARIO.....	32
REFERENCIAS	34



Listado de Tablas

Tabla 1. Matriz de trazabilidad de fuentes – Control presupuestal Secretaría de Educación.....	11
Tabla 2. Plantilla requerimientos	12
Tabla 3. Ejemplo de Requerimiento.....	12
Tabla 4. Vocabularios y repositorios más relevantes para reutilizar recursos.....	16
Tabla 5. Relación de Vocabularios existentes	17
Tabla 6. Vocabulario de contratación pública. Fuente (Comisión europea)	20
Tabla 7. Resumen de Metadatos.....	28
Tabla 8. Criterios de cuantificación para repositorios LOD (Fuente propia, 2017).....	31



Listado de Figuras

Figura 1. Estructura de una URI	4
Figura 2. Detalle de una URI	4
Figura 3. Estructura de una URI	5
Figura 4. Relación entre Identificador, Recurso y Representación (URI)	6
Figura 5. Metodología para Transformación OD a LOD.....	8
Figura 6. Detalle para Identificación de conjuntos de datos (DataSets).....	10
Figura 7. Ciclo de vida de datos abiertos. (Mintic, 2016).....	10
Figura 8. Mecanismos URI.....	14
Figura 9. Detalle para Limpieza de Datos	14
Figura 10. Detalle para Modelado de Datos.....	15
Figura 11. Búsqueda en Swoogle	18
Figura 12 Cantidad de resultados generados por Swoogle	18
Figura 13 Resultados iniciales de Swoogle	18
Figura 14. Visualización de datos del primer resultado Swoogle.....	18
Figura 15. Componentes núcleo encontrados en http://xmlns.com/foaf/spec/	19
Figura 16. Descripción del vocabulario Project encontrado en: http://xmlns.com/foaf/spec/#term_Project	19
Figura 17. Metavocabulario de conceptos gubernamentales.....	21
Figura 18. Vista de Alto Nivel Propuesta Metodológica Ontológica híbrida	22
Figura 19. Fases Metodología Propuesta (Fuente propia, 2017)	22
Figura 20. Fases Especificación y formalización (Fuente propia, 2017)	22
Figura 21. Detalle para Enriquecimiento de datos	24
Figura 22. Detalle para Vinculación de Datos	25
Figura 23. Indicadores de la ciudad de Barranquilla expuestos a través de datos abiertos	26
Figura 24. Triplete Contrato - EstadoProceso - Ejecución.....	26
Figura 25. Proceso de transformación de datos abiertos a datos abiertos vinculados	27
Figura 26. Detalle para Publicación de Datos	27
Figura 27. Propuesta Marco de Transformación de Datos Estructurados a Publicación de LOD (Fuente propia, 2017).....	30
Figura 28. Detalle para Validación de los Datos.....	30



1. INTRODUCCIÓN

Los avances tecnológicos actuales, han permitido mejorar los procesos de recopilación, análisis y difusión de la información, generando la necesidad de buscar y desarrollar alternativas que permitan dar utilidad a los datos obtenidos, de manera que se puedan gestionar, inspeccionar y mejorar a medida que se refinan los procesos de comprensión y análisis, de tal forma que se pueda extraer una base de conocimiento para ser aplicado en diferentes campos de interés.

El sector gubernamental no ha sido indiferente a este fenómeno, y ha investigado alternativas para dar utilidad a todos los conjuntos de datos que se genera diariamente, con el objetivo de mejorar los procesos de gobierno, descubrir conocimiento, encontrar patrones anómalos, y mejorar la relación estado – ciudadano con el fin de proporcionar la información y el conocimiento obtenido, haciendo uso de diferentes medios de difusión. Este conjunto de acciones ha creado un nuevo paradigma llamado gobierno abierto (o-Gov) que permite aproximar a los ciudadanos con las entidades estatales, basado en la premisa “*La información debe ser compartida al ciudadano, para tener una mayor cercanía, fortaleciendo la confianza y participación mutua*” (SALAZAR, MEJIA, & JOYANES, 2015) Este efecto ofrece mayor participación a los ciudadanos de una nación, proporcionándoles acceso libre a los diferentes datos generados en el sector público (*open data*), para que ellos puedan consultarlos y puedan emitir sus propios juicios respecto a los resultados que se estén dando desde el gobierno.

La publicación de datos por si solos no posibilita descubrimiento de conocimiento sobre ellos, es necesario aplicar enfoques y tecnologías específicas sobre los datos abiertos publicados por el sector gubernamental, para explotar todo el potencial que ofrecen. Es aquí donde entra el concepto tecnológico de Datos Abiertos Enlazados LOD por su sigla en inglés (*Linked Open Data*).

La tecnología LOD permite determinar un mecanismo de convergencia de conjuntos de datos abiertos y en un contexto determinado, generando nuevo conocimiento y permitiendo determinar nuevos servicios y aplicaciones con base en ellos. Lo anterior es posible, debido a que LOD facilita la interrelación de datos de diferentes fuentes, permitiendo gestionarlos e integrarlos. Adicionalmente LOD está siendo utilizada para la reutilización de la información del sector público, con el fin de mejorar procesos administrativos gubernamentales, integrando e interconectando datos oficiales y no oficiales.

Esta guía contiene un proceso metodológico para transformar datos abiertos -OD- en datos abiertos enlazados -LOD-, con base en las mejores prácticas de la *World Wide Web Consortium (W3C)* y en los lineamientos para su implementación en Colombia. La metodología se compone de siete fases así: 1) Identificación de conjuntos de datos; 2) Limpieza de los datos; 3) Modelado de los datos; 4) Enriquecimiento de los datos; 5) Vinculación de los datos; 6) Publicación de los datos; 7) Validación de los datos publicados. Cada una de las fases se describe de manera detallada, indicando sus pasos y aclarando conceptos en cada paso.



El propósito es promover en el nivel público y privado al incremento del conjunto de datos en formato LOD, obteniéndose altos estándares en el nivel de madurez en su uso, con la consecuente mejora en su potencial para la generación de conocimiento, transparencia, riqueza e interacción en la sociedad colombiana.

2. ACLARACIONES Y CONCEPTUALIZACIÓN

Datos Abiertos

¿Qué son datos abiertos?: En diferentes escenarios públicos y privados que inician sus procesos de transformación digital se habla de conjuntos de datos, información, archivos, bases de datos, entre otros elementos que permiten conservar la llamada *data* de la organización, que es necesario analizar y categorizar de manera que se seleccione lo publicable y lo privado. De los bloques que se consideran publicables (léase “públicos”), no todos están disponibles de forma gratuita o de acceso libre y abierto, es decir utilizables por cualquiera. Los datos abiertos rompen esa restricción, el (Open Data Charter, 2010) define los datos abiertos de la siguiente manera: *“Datos abiertos son datos digitales que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar.”* Así pues *“los datos abiertos pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera*

en que aparecen” (OpenKnowledge International, 2008).

¿Cómo están los datos en la WEB?: La *World Wide Web* está llena de datos que se publican en diferentes formatos como PDF, TIFF, CSV, hojas de cálculo, tablas incrustadas en documentos de Word y muchas formas de texto sin formato. Son datos a los que se puede acceder con una limitación: están formateados para el consumo humano, requiriéndose a menudo de una utilidad especializada para leerlos y no es fácil buscar, acceder o reutilizar estos datos mediante procesos automatizados.

Se presenta así la posibilidad de disponer de una forma universal para generalizar los procesos de lectura, uso y reutilización de datos en la Web.

¿Qué son datos enlazados?: Una nueva forma de presentar y publicar datos en la Web para que puedan ser consumidos y reutilizados en cualquier lugar y mediante procesos automatizados. La forma de hacerlo se llama Datos Vinculados - *Linked Data*-. El término se refiere a un conjunto de mejores prácticas en la Web, para publicar y conectar datos estructurados utilizando los estándares internacionales de la W3C. (WOOD, ZAIDMAN, RUTH, & HAUSENBLAS, 2013)

La iniciativa *Linked Data* promueve una visión de la WEB como una base de datos global, al interconectar datos de la misma forma que lo hacen los documentos WEB. *Linked Data* toma la información de repositorios existentes en diversos formatos (csv, hojas electrónicas, xml, etc.) y derriba las barreras entre diversas fuentes de información. Los datos vinculados, debido a los estándares a los que se adhieren, facilitan la integración de datos y la navegación a través de datos complejos, lo que también



permite fáciles actualizaciones y extensiones de los modelos de datos.

Datos Abiertos Enlazados (LOD)

¿Qué es Linked Open Data?: Citando a WOOD: *La World Wide Web está llena de datos que se publican en formatos como PDF, TIFF, CSV, hojas de cálculo, tablas incrustadas en documentos de Word y muchas formas de texto sin formato; estos archivos están vinculados desde HTML y otros documentos, en cierto sentido, son datos a los que puede acceder. Para llevar a cabo procesos automatizados no es fácil acceder, buscar o reutilizar estos datos; por lo general, se requiere un procesamiento posterior para que estos datos se incorporen en nuevos proyectos.*

Se abre entonces la posibilidad de tener una forma universal para que cualquiera pueda leer y reutilizar datos en la Web. El deseo es tener datos relacionados y que se fomente la reutilización entre personas.

Esta es una nueva forma de consumir, reutilizar y publicar datos en la Web para que puedan ser reutilizados en procesos automatizados en cualquier lugar. La forma de hacer esto se llama Linked Data - Datos Vinculados. El término Datos Vinculados se refiere a un conjunto de mejores prácticas para publicar y conectar datos estructurados en la Web utilizando los estándares internacionales del World Wide Web Consortium. (WOOD, ZAIDMAN, RUTH, & HAUSENBLAS, 2013)

LOD se apoya en la web semántica, la cual es una web que permite a las máquinas leer y compartir información, donde cada fuente de datos puede ser usada en diferentes aplicaciones

(KARMANOVSKIY, MOUROMTSEV, & et all, Linked Open Data in University, 2016)

¿Qué es WEB Semántica o WEB 3.0?

“La idea básica de la WEB 3.0 es definir estructuras de datos y vínculos que ordenados puedan llevar a: Descubrimientos más efectivos, automatización, integración y reutilización, a través de diferentes aplicaciones. La WEB 3.0 se esfuerza en vincular, integrar y analizar datos desde diferentes fuentes o conjuntos para obtener nueva información; posee la habilidad para mejorar la gestión de los datos, dar soporte a la accesibilidad a internet de dispositivos móviles, estimular la creatividad e innovación, elementos estos que van de la mano con el fenómeno de la globalización, además de mejorar la satisfacción de los usuarios ayuda a organizar los procesos de colaboración en una WEB Social” (AGHAEI, NEMARBAKHS, & KHOSRAVI, 2012)

¿Por qué pasar de datos abiertos a datos abiertos enlazados?: La transformación de Open Data a Linked Open Data -OD/LOD-, genera ventajas cualitativas y cuantitativas en la utilización y el análisis de datos. En lo cualitativo, LOD mejora la descripción conceptual de los datos, al dotarlo de contexto y relacionarlo con otros datos de una forma estructurada; en lo cuantitativo, permite a las máquinas realizar consultas que generan conocimiento de una forma automática y sobre grandes volúmenes de datos, donde cada fuente de datos puede ser usada en diferentes aplicaciones (KARMANOVSKIY, MOUROMTSEV, & et all, Linked Open Data in University, 2016).



URI - Uniform Resource Identifier (identificador uniforme de recursos)

“Este elemento sirve para identificar recursos en Internet. Dicho identificador de recursos tiene un formato estándar definido y su propósito es permitir interacción entre recursos disponibles en Internet -o en alguna red de cómputo-, recursos como lo son páginas, servicios, imágenes, vídeos, entre otros.

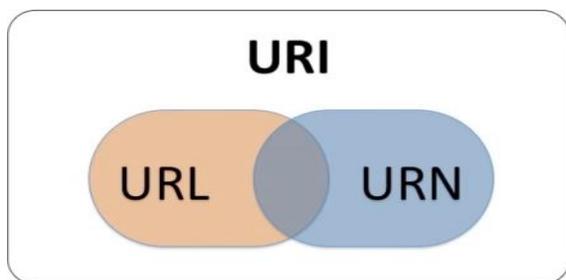


Figura 1. Estructura de una URI

El URI no se debe confundir con el URL (aunque es muy común que se haga referencia a ellos como la misma cosa), ya que el URI, incluye al URL. La diferencia fundamental es que los URI identifican y los URL localizan y, como las localizaciones también sirven para identificar, se puede decir que todos los URL son también URI y que, por otro lado, hay URI que no son URL.

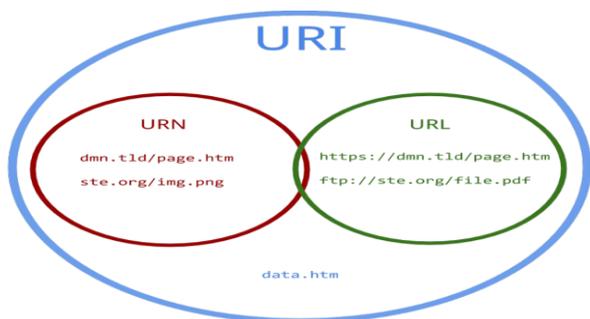


Figura 2. Detalle de una URI

Ejemplo - Mi nombre, Luis Castro, es una identificación (sería como un URI), pero no dice nada de cómo localizarme, es más, sirve para identificar a mucha más gente con el mismo nombre; en cambio, mi dirección dice exactamente cómo localizarme (sería como un URL) y así es posible diferenciarme de los otros Luis Castro en el mundo y establecer contacto conmigo.” (CASTRO, 2016)

URN. Son las siglas de *Uniform Resource Name*, que sirve también para identificar a un recurso en una red pero usando un nombre. Los nombres usados son menos amigables que un nombre propio, pero sí deben identificar de forma única un recurso.

De manera detallada Berners- Lee hace una caracterización detallada de las URI de la siguiente forma:

Uniformidad. La uniformidad proporciona varios beneficios:

- Permite utilizar diferentes tipos de identificadores de recursos en el mismo contexto, incluso cuando los mecanismos utilizados para acceder a esos recursos pueden diferir.
- Permite la interpretación semántica uniforme de una convención sintáctica común, a través de diferentes tipos de identificadores de recursos.
- Permite la introducción de nuevos tipos de identificadores de recursos, sin interferir con la manera en que se utilizan los identificadores existentes.
- Permite que los identificadores puedan ser reutilizados en muchos contextos diferentes.



Recursos. Esta especificación no limita el alcance de lo que podría ser un recurso; más bien, el término "recurso" se usa en un sentido general para cualquier cosa que pueda identificarse mediante un URI. Los ejemplos típicos incluyen un documento electrónico, una imagen, una fuente de información con un propósito consistente (por ejemplo, "Informe meteorológico actual para Los Ángeles"), un servicio (por ejemplo, una puerta de enlace HTTP a SMS) y una colección de otros recursos. Un recurso no es necesariamente accesible a través de Internet; por ejemplo, seres humanos, corporaciones, libros encuadernados en una biblioteca también pueden ser recursos. Del mismo modo, los conceptos abstractos pueden ser recursos, como los operadores y operandos de una ecuación matemática, los tipos de una relación (por ejemplo, "padre" o "empleado") o valores numéricos (por ejemplo, cero, uno o menos dos).

Identificador. Un identificador incorpora la información requerida para distinguir lo que se está identificando de todas las otras cosas dentro de su alcance de identificación. Es decir, los URI tienen un alcance global y se interpretan de forma coherente independientemente del contexto, aunque el resultado de esa interpretación puede estar en relación con el contexto del usuario final. Ejemplo Carro, es una identificación, el contexto varía si estamos en una fábrica de automóviles, en una carrera de la fórmula uno, o si estamos en un concesionario de vehículos. El identificador le dará condición única para los diferentes casos.

El uso de los términos "identificar" e "identificación" se refieren al propósito de

distinguir un recurso del resto de los recursos. El identificador no incorpora identidad de lo que se hace referencia. Tampoco debe suponerse que un sistema que use URI acceda al recurso identificado: en muchos casos, los URI se utilizan para denotar recursos sin ninguna intención de que se acceda a ellos. Del mismo modo, el recurso "uno" identificado podría no ser de naturaleza singular (por ejemplo, un recurso podría ser un conjunto con nombre o un mapeo que varía con el tiempo).

Un URI es un identificador que consiste en una secuencia de caracteres, en las que se describe el método de acceso, la ubicación del recurso y el nombre del recurso. Un ejemplo se muestra en la Figura 3.

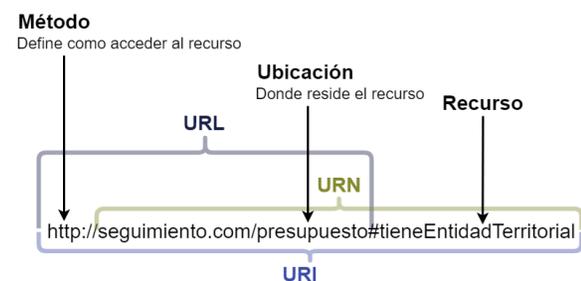


Figura 3. Estructura de una URI

Para una mirada más detallada se analiza un ejemplo tomado de (JACOBS Ian - W3C, 2004):

Al planear un viaje a México, Nadia lee "información meteorológica de Oaxaca: 'http://weather.example.com/oaxaca'" en una brillante revista de viajes. Nadia tiene suficiente experiencia con la Web para reconocer que "http://weather.example.com/oaxaca" es una URI y que es probable que pueda recuperar la información asociada con su navegador web. Cuando Nadia ingresa al URI en su navegador:



1. El navegador reconoce que lo que Nadia tipeó es un URI.
2. El navegador realiza una acción de recuperación de información de acuerdo con su comportamiento configurado para los recursos identificados a través del esquema de URI "http".
3. La autoridad responsable de "weather.example.com" proporciona información en respuesta a la solicitud de recuperación.
4. El navegador interpreta la respuesta, identificada como XHTML por el servidor, y realiza acciones de recuperación adicionales para gráficos en línea y otro contenido según sea necesario.
5. El navegador muestra la información recuperada, que incluye enlaces de hipertexto a otra información. Nadia puede seguir estos enlaces de hipertexto para recuperar información adicional.

Ontología

Gruber define **ontología** como “a formal explicit specification of a shared conceptualization” (GRUBER, 1993). Una ontología es una jerarquía de conceptos con atributos y relaciones. Una ontología proporciona un vocabulario de clases y relaciones para describir un dominio, haciendo énfasis en compartir conocimiento y el consenso en la representación de éste. Por ejemplo, una ontología sobre arte podría incluir clases como Pintor, Cuadro, Estilo o Museo, y relaciones como autor de un cuadro, pintores pertenecientes a un estilo artístico u obras localizadas en un museo. (CASTELLS, 2009)

La ontología permite caracterizar el conocimiento de un experto en el área de estudio a través términos (conceptos) y sus relaciones, utilizando reglas que definen las restricciones entre conceptos.

Una parte crucial del LOD es el modelo que permite a partir de ontologías, verificar los vocabularios, efectuar una definición de reglas, axiomas, instancias, de manera que se pueda inferir conocimiento. Ontología es una antigua disciplina que en sentido filosófico, se define como un esquema específico de categorías que refleja una visión específica del mundo. Desde el punto de vista informático las ontologías son teorías que especifican un vocabulario relativo a un cierto dominio – tema de estudio – área de conocimiento. Este vocabulario define entidades, clases, propiedades, predicados, funciones y las relaciones entre estos componentes. Las ontologías toman un papel clave en la resolución de interoperabilidad semántica entre sistemas de información y su



Figura 4. Relación entre Identificador, Recurso y Representación (URI)



uso. (GUZMAN LUNA, LÓPEZ BONILLA, & DURELY TORRES, 2012)

Una ontología permite tener un vocabulario en común para investigadores de diferentes ramas, quienes necesitan compartir información y datos que pueden apoyar a otros investigadores de la misma área de conocimiento. El compartir información digital con otras personas u organizaciones permite que no se repitan procesos o que se dé la redundancia de información almacenada en repositorios de datos y así el análisis de ellos se realice de mejor manera, ahorrando tiempo y evitando conflictos entre diferentes conceptos.

Metavocabulario

En la caracterización de información y componentes de ella se encuentra como base el dato, sobre él aparece el metadato (Corresponde a la descripción detallada de las características de ese dato en particular) Ejm Dato – NumeroCedula. El metadato describe las características de esta variable. Tipo de dato, Cantidad de campos usados, Condiciones de operación, Nombre usado, Valores predeterminados, Como se valida, Quien lo genera, entre otros elementos que darían completitud al dato. De igual manera al hablar de un vocabulario estamos frente a una estructura que permite la constitución y construcción de una ontología (Tal como se describe en el aparte anterior), un **metavocabulario** se relaciona directamente con la generación de un vocabulario propio que permite la comunicación con otros vocabularios; es entonces una construcción semántica que permite crear las condiciones

para los vocabularios de un sector o área de conocimiento, es decir, crea las condiciones de conectividad para las ontologías. Los metavocabularios ofrecen las condiciones para combinar los datos y paquetes RDF con los esquemas de los proyectos y ontologías particulares. Este se logra a través de componentes XML, los cuales han de tener las características particulares deseables o necesarias para un área específica, esto se presenta dado que los vocabularios disponibles son genéricos y se entregan de manera abierta para que cada persona – usuario, los use en base a sus necesidades. El metavocabulario permite segmentar o crear un espacio para tener agrupadas los vocabularios requeridos y que se adapten a mis necesidades.

3. Transformación de OD a LOD

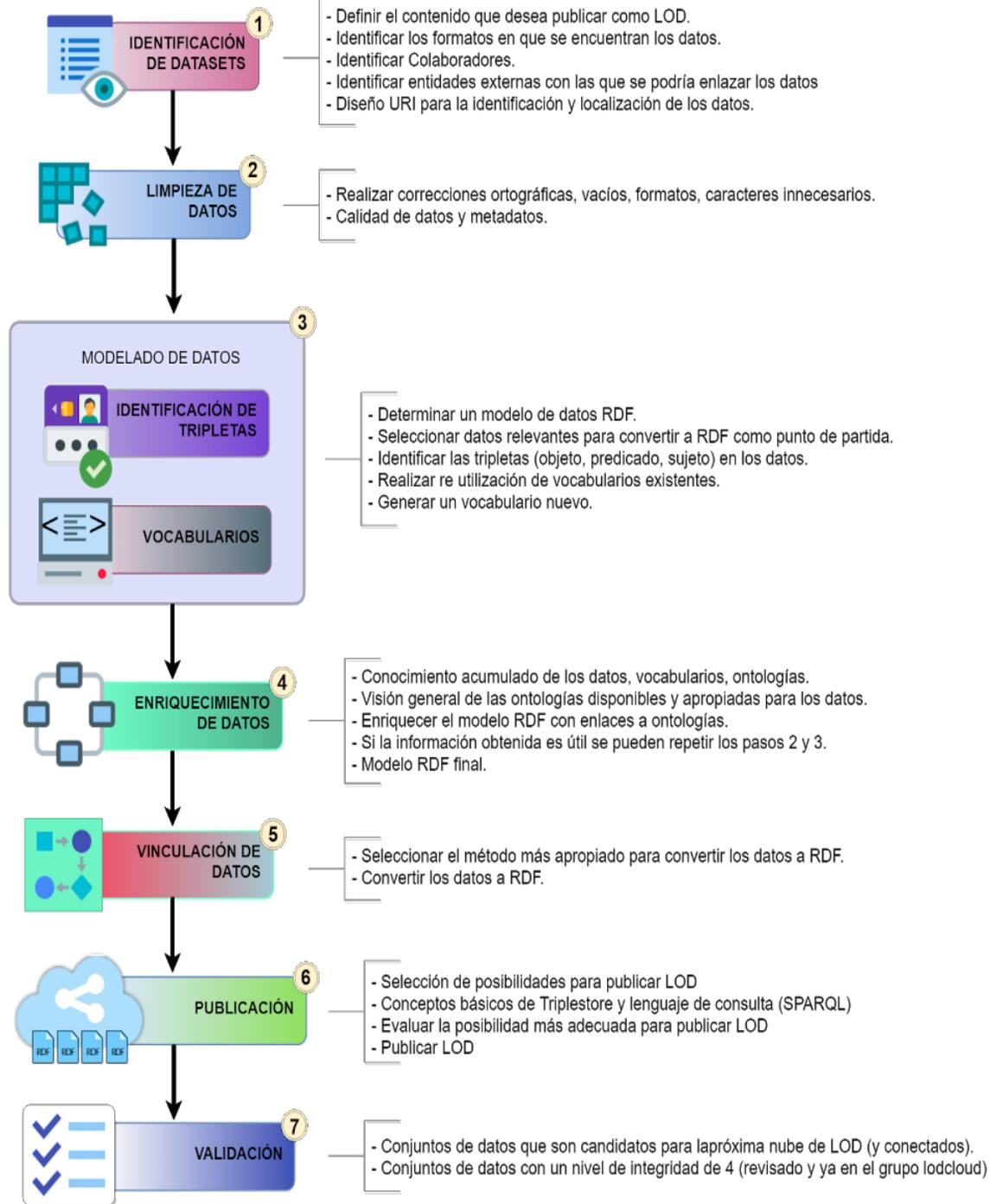


Figura 5. Metodología para Transformación OD a LOD.



En las responsabilidades asignadas por los gobiernos a las unidades de TI, los CIO o quien haga sus veces tienen el reto de proyectar procesos, formas y condiciones de manera organizada; solo es necesario analizar la cantidad de datos e información que se produce día a día en las entidades del estado. La ventaja del hoy, es la capacidad de llevar los avances informáticos a las entidades Gubernamentales, además de lo básico en las TIC nos encontramos con las implantaciones de nuevas tecnologías allegadas por la WEB.

Asociado a ello se encuentra la extensión del concepto de red, para las entidades estatales, Cardona y otros citando a (Ruiz 1998), comenta “*las instituciones de desarrollo han reenfocado sus esfuerzos, cambiando el sujeto de atención, hacia el concepto de redes empresariales, bajo la consideración de que es mejor apoyar a empresas en redes de producción, que a empresas individuales*”. (CARDONA, CANO, & RAMÍREZ, 2002). Esta concepción se lleva análogamente a las entidades del estado, en donde las entidades agrupadas y en redes generan un proceso de aprendizaje que no se da cuando operan de forma aislada, allí se genera un estado en red, un estado interconectado, un Gobierno en Línea con datos enlazados.

Para lograr este noble objetivo es necesario estar organizados y tener métodos, técnicas y estrategias que permitan avanzar en este recorrido.

A continuación se detalla la metodología que apoya este proceso ajustada a las necesidades del estado Colombiano.

Para desarrollar la propuesta metodológica se plantea un caso de estudio que permite

presentar el desarrollo y resultado en cada uno de los pasos.

Objetivo:

Analizar, diseñar e implementar una herramienta de control ciudadano para hacer seguimiento a la ejecución presupuestal en entes territoriales, usando un conjunto de datos enlazados del banco de proyectos de la secretaría de planeación con el POAI (Plan anual de inversiones), decretos de liquidación de presupuesto general de rentas y gastos, y el SECOP.

Descripción:

Herramienta online que permite comprender el proceso de ejecución de presupuesto de entes territoriales desde la etapa de planeación, pasando por la estructura jerárquica de sectores, programas, subprogramas y proyectos hasta llegar a contratos con desembolsos y estado de los mismos; así como revisar el estado de ejecución presupuestal en proyectos puntuales seleccionando una entidad territorial, sector, programa y subprograma de interés.

La herramienta pretende facilitar la comprensión del procedimiento de ejecución de presupuesto público, fortaleciendo la transparencia en las administraciones de los entes territoriales y sirviendo de medio para veeduría y control ciudadano a través de la revisión objetiva de los proyectos planteados en el plan de desarrollo de una administración.



3.1. Identificación de Conjuntos de Datos (DataSets)

Este primer componente por desarrollar, corresponde a la identificación de requisitos de datos vinculados seleccionados por el organismo estatal o privado.



Figura 6. Detalle para Identificación de conjuntos de datos (DataSets)

3.1.1. Identificación y análisis de fuentes de datos

Consiste en la identificación de los conjuntos de datos a ser transformados. Se debe tener en cuenta que la información puede encontrarse estructurada (bases de datos relacionales), semiestructurada (Formato XML, JSON, CSV, Tablas, Hojas electrónicas, entre otros.) o no estructurada (Videos, imágenes, scanner de documentos, documentos de texto, etc.). Se propone segmentar la identificación de los datos de los dos grupos: 1) Datos abiertos que aún no

han sido publicados, 2) Datos ya publicados y en uso.

Para el primer caso – Datos no publicados - es necesario usar el ciclo de vida de los datos, además, agregar otras acciones como determinar el propietario del dato, especificar el acceso y determinar su conectividad.

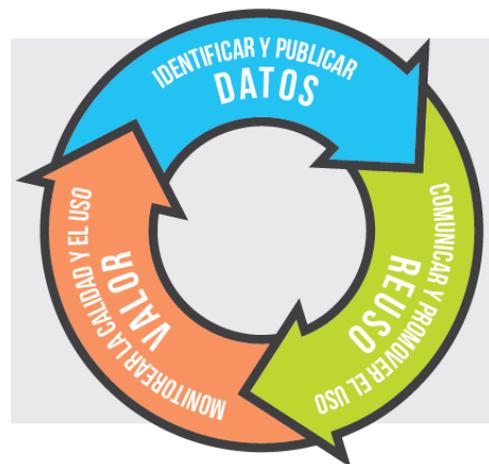


Figura 7. Ciclo de vida de datos abiertos. (Mintic, 2016)

Para el segundo caso – Datos publicados - Es necesario seleccionar los datos que la entidad recopila o crea de forma exclusiva; la idea es determinar las combinaciones que, con otros datos abiertos, proporcionan mayor valor. Las entidades poseen la capacidad para recopilar y seleccionar los conjuntos de datos más valiosos, los criterios que se pueden considerar son: reutilización, popularidad, e interés general.

Para el caso de seguimiento a la ejecución presupuestal de una gobernación, los datos seleccionados para transformar y enlazar son:

- **POAI** (Plan Operativo Anual de inversiones). Contiene la información de inversión de una entidad pública en fase de planeación.



- **Decretos de liquidación del presupuesto general de rentas y gastos.** Contiene la información correspondiente a la distribución de los rubros para un ente territorial en fase de ejecución.

- **Banco de proyectos.** El banco de proyectos es un Conjunto de datos que maneja la Secretaría de planeación con el objetivo de identificar y aprobar los proyectos que van a ser ejecutados.

- **SECOP I.** Los Conjunto de datos manejados por el SECOP I contienen información correspondiente a los contratos en fase precontractual. Los contratos están asociados a un subprograma a través de la ficha BPID¹.

3.1.2. Análisis de datos iniciales

Para los datos seleccionados se hace necesario conceptualizar y establecer la calidad de los datos existentes, para lo que se sugiere:

- Recopilar y analizar la documentación disponible sobre los conjuntos de datos seleccionados, incluido el propósito, los componentes, el modelo de datos y los detalles de la implementación.
- Identificar las estructuras de los componentes conceptuales (sentido y significado) y sus relaciones.

El objetivo de este análisis es seleccionar las categorías conceptuales que permitirán establecer el vocabulario para la vinculación de los datos.

Una forma de relacionar los datos con sus fuentes de origen se presenta en la matriz de trazabilidad de fuentes –Tabla I.

Tabla I. Matriz de trazabilidad de fuentes – Control presupuestal Secretaría de Educación

Dataset	Origen	Formato Inicial
POAI	Sitio web de la Gobernación.	PDF
	http://xurl.es/robzb	
Decreto de liquidación de presupuesto general de rentas y gastos.	Repositorio de datos de la Secretaría de Educación.	PDF
	http://xurl.es/hph30	
Banco de proyectos	Repositorio de datos de la Secretaría de Planeación.	XLS
SECOP I	Consolidados en la plataforma de datos abiertos	CSV
Relación de contratos para ejecución presupuestal	Repositorio de datos de la Secretaría de Educación.	XLS

Es posible encontrar baches en la información razón por la cual es de responsabilidad de los CIO hacer seguimiento efectivo de los datos, garantizando los lineamientos; si es del caso llegar a las fuentes documentales y/o grupos de trabajo para definir y solicitar los datos necesarios para complementar los paquetes que llegaran al ciudadano y lo que se desea proyectar. Es importante que el proyecto este socializado y exista voluntad y motivación de todos los actores.

¹ Banco de Proyectos de Inversión Departamental



3.1.3. Identificar Colaboradores e interesados

Corresponde a la identificación de requerimientos globales y de requisitos de datos vinculados que requiere el organismo, sea a manera general o para responder a casos puntuales.

Lo primero es convocar a los interesados iniciales que serán los beneficiados por los conjuntos de datos a ser generados y de ellos se obtiene la concepción inicial de requerimientos. En segunda instancia se considera a los gestores de los datos, que es el grupo de trabajo comprometido con la generación de los datos e información y quienes finalmente proveen el insumo de datos abiertos requerido.

En cada uno de los casos, es indispensable el uso de una plantilla o notación que documente las necesidades de los participantes, es discrecional del responsable. La intencionalidad es determinar los posibles resultados esperados con base en los datos y las posibilidades al escalar a LOD.

Tabla 2. Plantilla requerimientos

#	Requerimiento	Necesidad
		<Esencial / Condional />

En la Tabla 2, aparece una propuesta de plantilla para este paso, donde se clasifican las necesidades de la siguiente forma: Esencial (E)- cuando el requerimiento es indispensable para perfeccionar las necesidades de otros sistemas o usuarios; Condicional (C) cuando el requerimiento se considera deseable, mas no es indispensable, por su posible proyección,

A modo de ejemplo, consideremos un caso en donde se desean enlazar datos sobre la ejecución presupuestal de una gobernación; en este caso se considera la secretaría de educación (Interesado inicial). La entidad proporciona claridad sobre sus necesidades durante el proceso que se debe seguir para controlar la ejecución presupuestal. La Tabla 3, muestran los requerimientos y las necesidades consideradas.

Tabla 3. Ejemplo de Requerimiento

#	Requerimiento	Necesidad
01	Poseer medios de verificación a los avances de proyectos, para garantizar el seguimiento a la ejecución presupuestal de la secretaria de educación de la Gobernación de Caldas	E
02	Seguimiento a la ejecución presupuestal de a secretaría de educación	E

3.1.4. Definición de licencia

Dentro del contexto gubernamental, es importante definir la licencia de los datos que se publican; actualmente se encuentran diversas licencias que se pueden usar para datos gubernamentales.

Podemos considerar como buena práctica aquella que cumpla con el siguiente postulado: Es importante especificar a quién pertenecen los datos publicados en la Web y conectar explícitamente esa licencia con los datos en sí, se recomienda que las autoridades gubernamentales que publiquen datos abiertos reconozcan la guía relevante para licencias abiertas y derechos de autor (Mintic, 2016), de igual forma, se debería consultar la Guía de Apertura de Datos Abiertos en Colombia, que



recomienda usar licenciamientos tipo *Creative Commons*. La publicación de datos abiertos enlazados facilita la asociación de una licencia que viaja con los datos a la par; es más probable que las personas reutilicen los datos cuando hay asociada una licencia clara y aceptable.

A continuación, se enumeran algunos modelos de licenciamiento:

- El UK Open Government License, fue creado para permitir a cualquier titular de información del sector público hacer que su información esté disponible para su uso y reutilización según sus términos².
- El Open Database License (ODbL) es una licencia abierta para datos y bases de datos, que incluye requisitos explícitos de atribución y compartición³.
- Public Domain Dedication and License (PDDL) es un documento destinado a permitir, compartir, modificar y usar libremente una información particular para cualquier propósito y sin ninguna restricción⁴.
- Open Data Commons Attribution License, es una licencia específica para bases de datos⁵.
- The Creative Commons Licenses, son varias licencias de derechos de autor que permiten la distribución de obras protegidas por derechos de autor⁶.

3.1.5. Diseñar el identificador uniforme de recursos (URI) para la identificación y localización de los datos

La utilización de las URIs tiene como propósito poder identificar y localizar, datos y conceptos que se desean enlazar, se presentan como una solución a la necesidad de plantear una forma directa y única de direccionamiento, para acceder a los recursos que se desean enlazar.

Un URI es un identificador que consiste en una secuencia de caracteres, en las que se describe el método de acceso, la ubicación del recurso y el nombre del recurso.

La W3C⁷ recomienda algunas pautas para nombrar un URI bajo un criterio estandarizado y normalizado. A continuación se realiza una breve descripción.

- **Usar URI HTTP.** Para beneficiarse y aumentar el valor de la World Wide Web, el gobierno y las entidades centralizadas y descentralizadas deberían proporcionar URI HTTP como identificadores para sus recursos.
- **Proporcionar representaciones legibles.** Dado un URI, las personas y máquinas deberían poder recuperar una descripción sobre el recurso identificado por un URI, desde la Web, tal mecanismo de búsqueda es importante para establecer una comprensión compartida de lo que identifica un URI. Las máquinas deberían obtener datos RDF y los humanos deberían obtener una representación legible, como la que muestra la Figura 8 para HTML.

² <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

³ <https://opendatacommons.org/licenses/odbl/>

⁴ <https://opendatacommons.org/licenses/pddl/>

⁵ <https://opendatacommons.org/licenses/by/>

⁶ <https://creativecommons.org/>

⁷ <https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector>

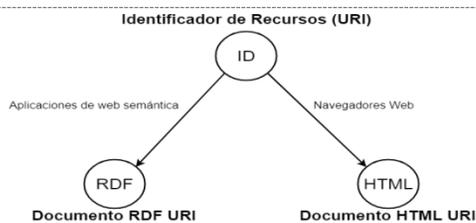


Figura 8. Mecanismos URI

- El URI no contendrá nada que se pueda cambiar: Una vez que se configura un URI para identificar un determinado recurso, debe permanecer así el mayor tiempo posible, como los próximos diez años. Los URI deben ser estables y confiables para maximizar las posibilidades de reutilización que Linked Data brinda a los usuarios. Debe haber un equilibrio entre hacer que los URI sean legibles y mantenerlos más estables, eliminando información descriptiva que probablemente cambie⁸.
- Opacidad del URI. Los agentes que utilizan URI no deben intentar inferir propiedades del recurso al que se hace referencia. Los clientes web que acceden a tales URI no deben analizar ni leer el significado de los URI. Se debe poner en el URI tanta información como sea posible.
- Usar en la medida de lo posible, el lenguaje oficial de la entidad, en algunos casos se puede tratar con algunos caracteres especiales que dependan del lenguaje.
-

⁸ El término "persistencia URI" se utiliza para describir la propiedad deseable que, una vez asociado a un recurso, un URI debe continuar indefinidamente para referirse a ese recurso. **Buena práctica:** La Representación consistente Un propietario de URI DEBERÍA proporcionar representaciones del recurso identificado de manera consistente y predecible. La persistencia de URI es una cuestión de política y compromiso por parte del propietario de URI. La elección de un esquema de URI particular no garantiza que esos URI sean persistentes o que no sean persistentes.

3.2. Limpieza de Datos



- Realizar correcciones ortográficas, vacíos, formatos, caracteres innecesarios.
- Calidad de datos y metadatos.

- Correcciones ortográficas realizadas
- Vacíos de datos documentados y completados
- Verificación y validación de formatos, eliminando los caracteres innecesarios.
- Mejoramiento de la calidad de los datos y metadatos

Figura 9. Detalle para Limpieza de Datos

La limpieza de datos es necesaria dado que provienen de múltiples fuentes y en diferentes formatos; por ejemplo, bases de datos, XML, CSV, datos geográficos, entre otros, requieren de esfuerzos adicionales para asegurar eficiencia en el proceso de modelamiento, esto incluye liberar la información de elementos adicionales que no se incluyen en los conjuntos de datos a ser publicados. (Kaltenböck, Bauer, & Blumauer, 2012)

Para lograrlo, las directrices recomiendan llevar a cabo un proceso de depuración de datos aumentando la calidad de los mismos; por lo general, los datos pueden (i) estar incompletos, (ii) estar disponibles en diferentes formatos, (iii) exhiben ambigüedades e inconsistencias; por lo tanto, "el proceso de limpieza de datos debe planificarse cuidadosamente, teniendo en cuenta las operaciones futuras y las transformaciones hacia el modelo RDF y la alineación con las mejores prácticas de la Web Semántica", (SPC shared infrastructures: semantic interoperability,



2013) en particular para obedecer a los principios de homogeneidad, integridad e interoperabilidad.

En algunos casos es importante determinar el costo de la labor de limpieza, como parte de la estrategia de gestión y control del proyecto; de igual manera, es importante examinar el tiempo que se tomará desarrollar las tareas asociadas a la limpieza:

- Seleccionar y ordenar los datos que serán objeto de limpieza.
- Analizar y determinar los datos corruptos.
- Valorar la necesidad de aplicar algoritmos de corrección.
- En caso de no disponer de herramientas que apoyen el proceso, puede ser necesario desarrollar aplicaciones básicas de limpieza de datos.
- Verificar y validar los archivos y contenedores de datos (si es necesario).
- Establecer cuáles son los datos correctos

Para el proyecto desarrollado, y con base en los conjuntos de datos cargados que se iban a utilizar, se procedió a eliminar los registros corruptos utilizando las funciones de filtro avanzado en excel. La idea básica en esta fase es crear filtros por cada columna del conjunto de datos para identificar datos extraños, campos incoherentes o espacios en blanco inadecuados. Dependiendo del tipo de error se procede a corregirse o se toma la decisión de eliminar todo el registro.

3.3. Modelado de Datos

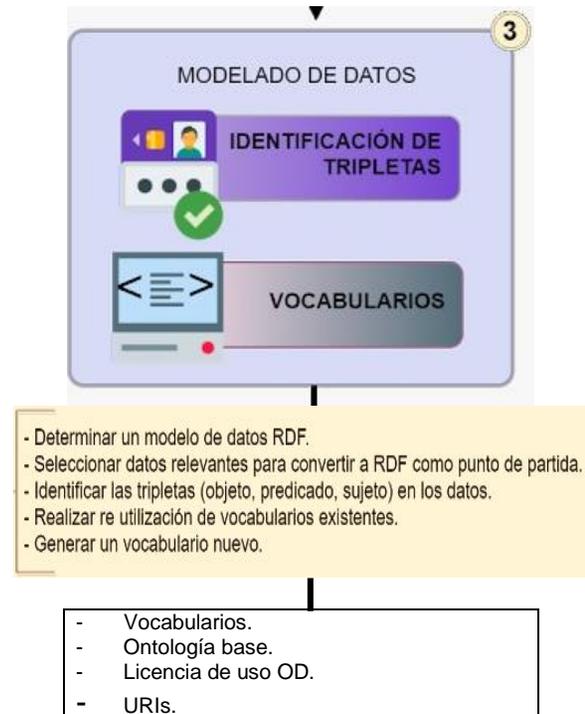


Figura 10. Detalle para Modelado de Datos

En esta etapa se busca establecer un lenguaje y una categorización que represente la información que contienen los datos que se van a transformar. Las acciones para lograr esta representación se dividen en tres escenarios.

- Buscar los vocabularios disponibles para ser reutilizados.
- Crear el vocabulario en caso de no encontrar ninguno disponible.
- Crear la ontología desde cero en caso de no lograr las acciones (a) y (b).

Cada una de estas acciones será detallada y acompañada por buenas prácticas que permitan un desarrollo adecuado y que garanticen la



estandarización y normalización de los vocabularios

3.3.1. Buscar Vocabularios

Actualmente existen algunos repositorios útiles para encontrar vocabularios disponibles, como SchemaWeb, SchemaCache, Swoogle y LOV. Para el caso de un vocabulario estándar, la W3C ha creado un componente que permite a los desarrolladores generar relaciones básicas o complejas para describir catálogos de datos generales, catálogos para organizaciones y catálogos para datos multidimensionales.

Buena práctica 1 - Usar herramientas de búsqueda que recopilan, analizan e indexan vocabularios y datos semánticos disponibles en línea para un acceso eficiente. Dentro de este tipo de herramientas se incluyen:

Watson⁹

Semantic Web Search Engine Swoogle¹⁰

W3C¹¹

Adicionalmente, existen repositorios donde se pueden explorar vocabularios y ontologías (Tabla 4), y la representación de estos en la (Tabla 5).

Tabla 4. Vocabularios y repositorios más relevantes para reutilizar recursos

Describir conjuntos de datos vinculados con el Vocabulario VoID	https://www.w3.org/TR/void/
---	---

⁹ <http://watson.kmi.open.ac.uk/WatsonWUI/>

¹⁰ <http://swoogle.umbc.edu/2006/>

¹¹ https://www.w3.org/wiki/Search_engines

Usado por organizaciones y autoridades gubernamentales, admite la publicación de información de la organización en varios dominios, como Datos Vinculados.	http://www.w3.org/TR/vocab-org/
Orientado a organizaciones gubernamentales que publican información estadística en la web. Utilizando el estándar W3C y RDF	http://www.w3.org/TR/vocab-data-cube/
DCAT es un vocabulario RDF diseñado para facilitar la interoperabilidad entre los catálogos de datos publicados en la Web.	http://www.w3.org/TR/vocab-dcat/
El Directorio LOV presenta categorías de vocabularios y un motor de búsqueda por patrones de uso.	http://lov.okfn.org/dataset/lov/
DBpedia es un recurso que trasciende Wikipedia hacia un modelo semántico, para el caso del idioma español, genera más de 100 millones de conexiones tipo RDF con tripletas completas.	http://es.dbpedia.org/

Buena práctica 2. Usar o ampliar un vocabulario existente antes de crear uno nuevo. Para ello se realiza una lista de chequeo básica previa; la W3C, recomienda:

- Asegurarse de que los vocabularios que utilizan son publicados por un grupo u organización de confianza.
- Asegurarse de que los vocabularios tengan URIs permanentes.
- Confirmar la política de control de versiones.



Del vocabulario seleccionado es necesario garantizar las siguientes características:

- **Los vocabularios deben estar documentados:** Esto incluye el uso liberal de etiquetas y comentarios, así como las etiquetas de idioma apropiadas. El editor debe proporcionar páginas legibles por humanos, que describan el vocabulario, junto con sus clases y propiedades constituyentes.
- **Los vocabularios deben ser auto descriptivos:** Cada propiedad o término en un vocabulario debe tener una etiqueta, definición y comentario definido. La capacidad de los datos vinculados para describirse a sí mismos y colocarse en contexto, contribuye a la utilidad de los datos subyacentes.
- **Los vocabularios deben escribirse en más de un idioma:** El multilingüismo debe ser apoyado por el vocabulario, es decir, todos los elementos del vocabulario deben tener etiquetas, definiciones y comentarios disponibles en el idioma oficial del gobierno, por ejemplo, español y al menos en inglés.
- **Los vocabularios deberían ser utilizados por otros conjuntos de datos:** Si el vocabulario es utilizado por otros conjuntos de datos abiertos enlazados, es un mejor candidato para la adopción y reutilización; es en la reutilización de vocabularios que conseguimos los beneficios de LOD.
- **Los vocabularios deben ser accesibles por un largo período:** El vocabulario seleccionado debe proporcionar alguna garantía de mantenimiento durante un

período específico, lo ideal es encontrarlo indefinida.

- **Los vocabularios deben ser publicados por un grupo u organización de confianza:** Aunque cualquiera puede crear un vocabulario, siempre es mejor verificar si se trata de una persona, grupo u organización autorizada, que es responsable de publicar y mantener el vocabulario.
- **Los vocabularios deberían tener URL persistentes:** El acceso persistente al servidor que aloja el vocabulario, necesariamente debería tener la facilidad de reutilización.
- **Los vocabularios deberían proporcionar una política de control de versiones:** El editor idealmente abordará la compatibilidad de las versiones a lo largo del tiempo. Los principales cambios en los vocabularios deben reflejarse en la documentación.

Continuando con el ejemplo asociado al seguimiento de ejecución presupuestal, proceso relacionado con la planeación de la contratación, se procede con la reutilización de vocabularios existentes, como el que se presenta en Tabla 5

Tabla 5. Relación de Vocabularios existentes

Section	Object	Description
Planning: Information from the planning phase of the contracting process.	Budger	The Budget objet has the following sub-elements: Soure, id, description, amount, Project, Project ID and URI20
	Rationale	The rationale for the procurement provided in free



		text.
	Documents	A list of documents related to the planning process.
Tender: The activities undertaken to enter into a contract.	ID	An identifier for this tender process
	Title	Tender title
	Description	Tender description
	Status	Status on of the tender (value from code list)
	Ítems	The goods and services to be purchased, broken into line items wherever possible
	Minvalue	The minimum estimated value of the procurement
	Value	Specify tendering

Para buscar más vocabularios asociados se realiza un ejemplo usando Swoogle. El término a ser localizado es proyecto (Project), como parte del ejercicio, es un componente de los elementos de subprogramas de la Gobernación.



[término del documento de](#) [ontología](#) [más >>](#)

Búsqueda Swoogle

Buscando más de 10,000 ontologías

Figura 11. Búsqueda en Swoogle

Esta búsqueda genera 4.370 resultados

¿Quieres más resultados? [Iniciar sesión](#)

[Versión RDF](#)

1 - 10 del total de 4.370 resultsfor proyecto en 0.098 segundos

ordenar por | [fecha](#) | [triple](#) |

Figura 12 Cantidad de resultados generados por Swoogle

<http://xmlns.com/foaf/0.1/index.rdf>
 [DEF] , Perfil, Proyecto , Propiedad, Servicio, Espaci
 SemanticWebDocument, RDFXML, 2010-08-08, 43K

http://rdf.geospecies.org/ont/gsonology_
 [DEF] , Proyecto , RDFtaxon, Rango, Referencia, Re
 SemanticWebDocument, RDFXML, 2009-04-06, 74K

http://rdf.geospecies.org/ont/geospecies_
 [DEF] , Proyecto , Prov, Provincia, RDFtaxon, Clasifi
 SemanticWebDocument, RDFXML, 2010-01-17, 171K

<http://swrc.ontoware.org/ontology>
 [DEF] , PhDThesis, Price, Proceedings, Product, Profe
 SemanticWebDocument, RDFXML, 2006-01-20, 43K

Figura 13 Resultados iniciales de Swoogle

El primer resultado demarcado corresponde al sitio: <http://xmlns.com/foaf/spec/> , indicando al inicio la sigla FOAF (*Friend of a Friend*) (Amigo de un Amigo) que corresponde a ontologías interconectadas legibles por maquinas.

FOAF Vocabulary Specification 0.99
 Namespace Document 14 January 2014 - Pa

This version:
<http://xmlns.com/foaf/spec/20140114.html> (rdf)

Latest version:
<http://xmlns.com/foaf/spec/> (rdf)

Previous version:
<http://xmlns.com/foaf/spec/20100809.html> (rdf)

Authors:
[Dan Brickley](#), [Libby Miller](#)

Contributors:
 Members of the FOAF mailing list (foaf-dev@lists.foaf-pr)

Copyright © 2000-2014 Dan Brickley and Libby Miller

Figura 14. Visualización de datos del primer resultado Swoogle



Al examinar la página se encuentran los componentes núcleo.



Figura 15. Componentes núcleo encontrados en <http://xmlns.com/foaf/spec/>

Al seleccionar el elemento esperado de búsqueda, carga la información correspondiente al vocabulario.

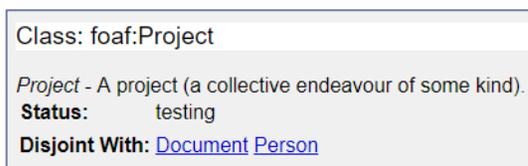


Figura 16. Descripción del vocabulario Project encontrado en: http://xmlns.com/foaf/spec/#term_Project

El texto asociado a la descripción del elemento es:

*The **Project** class represents the class of things that are 'projects'. These may be formal or informal,*

*collective or individual. It is often useful to indicate the **homepage** of a **Project**.*

Este ejemplo permite representar una definición que puede ser aplicable al caso y de uso general, creando relaciones a otros elementos en este caso a documentos y personas, permitiendo navegar entre esos componentes.

3.3.2. Crear Vocabularios

En caso de no descubrirse ningún vocabulario adecuado para los propósitos, se deberían crear este o estos, tratándose de reutilizar la mayor cantidad posible de recursos existentes; por ejemplo, catálogos gubernamentales y/o vocabularios disponibles¹². Para este caso (VILLAZÓN-TERRAZAS, WILCHES-VÁZQUEZ, CORCHO, & GÓMEZ PÉREZ, 2010) recomiendan:

- Buscar recursos gubernamentales desde sitios web altamente confiables, sitios relacionados con dominios y catálogos gubernamentales.
- Seleccionar el más apropiado recurso del gobierno.
- Transformarlos en Ontologías.

Desde las W3C, proporcionan un conjunto de consideraciones dirigidas al desarrollo de vocabularios.

- **El URI que identifica su vocabulario debe estar definido:** Esto está fuertemente relacionado con las mejores prácticas en la construcción de URIs.

¹² <http://semic.eu/>.



- **URI para propiedades con rangos no literales:** Consiste en asignar nombre a todas las propiedades, dando sentido al verbo, de modo que la generación de la tripleta se puede leer en el sentido que se construye. Ejemplo: elem1 (esPropietario).
- **Los vocabularios deben ser auto descriptivos:** Significa que cada propiedad o término en un vocabulario debe tener una etiqueta, definición y comentario definido.
- **Los vocabularios deberían proporcionar una política de control de versiones:** Se refiere al mecanismo establecido por el editor, para cuidar siempre las compatibilidades hacia atrás de las versiones, la forma en que esos cambios afectaron a las versiones anteriores. Los cambios importantes de los vocabularios deben reflejarse en la documentación, tanto en formato de máquina como de lectura humana.
- **Los vocabularios deberían proporcionar documentación:** Un vocabulario debe estar bien documentado para lectura mecánica (uso de etiquetas y comentarios, etiquetas para el lenguaje utilizado). También para lectura humana, el editor debe proporcionar una documentación adicional para comprender mejor las clases y propiedades, y si es posible con algunos casos de ejemplo.
- **Los vocabularios deben publicarse siguiendo las mejores prácticas disponibles:** Publique su vocabulario en la Web en un URI estable, usando una licencia abierta. Uno de los objetivos es contribuir a la comunidad compartiendo el nuevo vocabulario.¹³

Tabla 6. Vocabulario de contratación pública. Fuente (Comisión europea)

Nombre	Sinónimos	Acrónimos	Descripción	Tipo
Banco de proyecto de inversión departamental		BPID	http://schema.org/identifier	Atributo instancia
Transferencias		--	http://schema.org/MonetaryAmount	Atributo instancia
Sector Educación			http://schema.org/category	Instancia
Sector salud			http://schema.org/category	Instancia

Es importante para el sector gobierno, generar un metavocabulario (Vocabulario propio universal) que defina los conceptos utilizados en las instituciones del estado, ver **¡Error! No se encuentra el origen de la referencia..** La estandarización de los conceptos permite la liberación de ambigüedades para definir las relaciones utilizadas por las instituciones gubernamentales. Si el concepto “contrato” es claro y universal, las entidades pueden utilizarlo para definir un acuerdo legal, inclusive el concepto puede asociarse con la definición brindada por DBpedia.

¹³ Para la publicación se recomienda seguir los pasos detallados y estrategias para publicar RDF de la W3. <https://www.w3.org/TR/swbp-vocab-pub/>

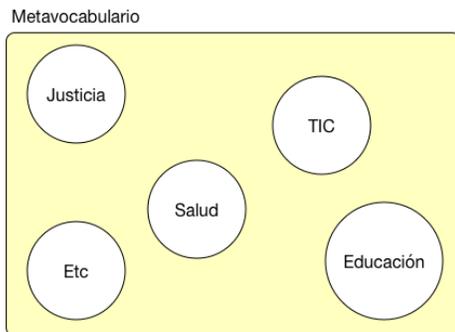


Figura 17. Metavocabulario de conceptos gubernamentales

El Metavocabulario serviría a los gerentes de TI de las dependencias generadoras de datos abiertos, para que lo apliquen en la transformación de datos abiertos a datos abiertos enlazados.

3.3.3. Crear ontología

Finalmente, si no se encuentran vocabularios disponibles ni recursos para construir la ontología, se debe crear la ontología desde cero. Para este fin, y con base en los documentos de Lineamientos se aplicará una versión híbrida entre MethOntology y NeOn; METHONTOLOGY (Fernández, Gómez-Pérez, & Juristo, 1997). Esta metodología brinda unos pasos para la representación de un dominio de conocimiento, mediante este proceso es posible crear las relaciones y cardinalidades que se suelen presentar entre los conceptos. Así mismo, se enfoca en la creación de reglas y axiomas que extienden las capacidades de inferencia de conocimiento a través de motores que se aplican sobre la ontología ya definida. Las metodologías resultantes del proceso

METHONTOLOGY son suficientes cuando se trata de agregar tan solo un dominio de conocimiento por sí solo, es decir, que no se deba involucrar con otros dominios o con ontologías existentes, en ese sentido, este tipo de metodologías se apoyan en otras que son más precisas en tales aspectos (reutilización de fuentes de datos y vocabularios), como es el caso de NeOn (Suárez-Figueroa, Gómez-Pérez, & Fernández-López, 2015), una metodología madura para la creación de ontologías con un alto grado de reuso.

La Metodología NeOn se aplica para “la construcción de redes de ontologías, basada en escenarios que se apoyan 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” (OEG, 2015).

En este sentido, la sugerencia explícita es basarse en un modelo híbrido de ingeniería ontológica donde la especificación está basada en MethOntology y las fases de reuso en NeOn.; la representación gráfica de estas metodologías se presentan en la Figura 18.

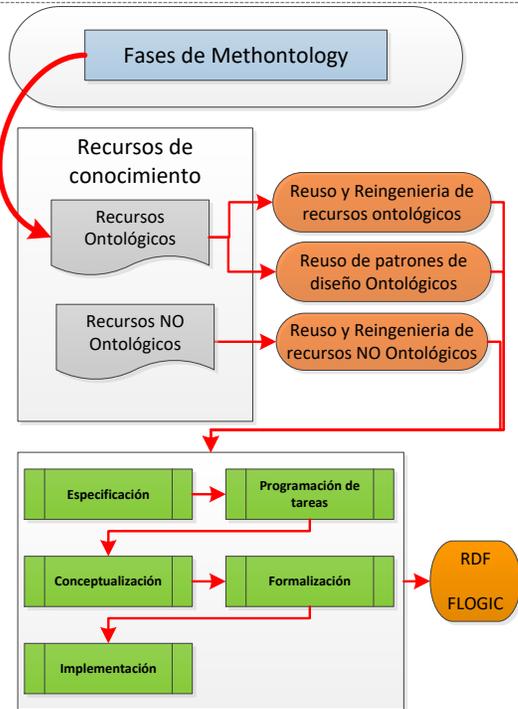


Figura 18. Vista de Alto Nivel Propuesta Metodológica Ontológica híbrida

requerida, fuentes necesarias, expertos, entre otras.

2. Establecer un Dominio de Expertos. La construcción de un repositorio Ontológico demanda la participación de grupos interdisciplinarios, que permitan definir un dominio de conocimiento, conjuntos de tareas e interacciones con suficiente rigurosidad, exhaustivas, suficientemente expresivas, semánticas y -muy importante- que no excluyan conceptos y relaciones relevantes o conduzcan a contradicciones entre vocabularios ya existentes.

3. Especificación. En la especificación se precisan las fases necesarias para considerar las diferentes variables, conceptos, elementos, relaciones y expresiones semánticas. A continuación, se especifica cada tarea (fase) a realizar:

Ciclo de Vida para la generación de Ontologías



Figura 19. Fases Metodología Propuesta (Fuente propia, 2017)



Figura 20. Fases Especificación y formalización (Fuente propia, 2017)

I. Planeación. En esta fase se determinarán los recursos necesarios para poder diseñar e implementar la Ontología vinculada a LOD, determinando tiempos, infraestructura

FI. Identificar Conceptos: Desde una interpretación abstracta del dominio a modelar, podría ser cualquier objeto o entidad; en subsiguientes fases se refinan dichos roles.



F2. Identificar Relaciones C2C (Concept to Concept):

Se pretende en esta sección lograr definir las primeras relaciones (propiedades) que vinculan un par de conceptos, teniendo en cuenta la simetría o no de los predicados.

F3. Construir taxonomía: Con estas relaciones básicas es posible aproximar una taxonomía de los términos; esta primera versión permite tener una visión global del grafo semántico. Se pueden modelar algunas tripletas identificadas desde las relaciones entre conceptos y la taxonomía básica.

F4. Incluir Modelo de Reutilización de Vocabularios Existentes: En el marco de la hibridación metodológica. Para esta fase se sugiere utilizar el modelo de escenarios propuestos en NeOn. (Suárez-Figueroa, Gómez-Pérez, & Fernández-López, 2015) Y (OEG, 2015).

Primer escenario: Cuando se debe diseñar la ontología desde el inicio. Los constructores deben determinar los requerimientos de la ontología; definir -con expertos- la exploración de datos sensibles de reúso.

Segundo escenario: El reúso de recursos no tiene representación ontológica (NOR). Para tal caso, se debe definir -y concertar- el tipo de transformación sintáctica que se puede aplicar a las fuentes originales, dependiendo del esquema requerido para el modelamiento de la ontología.

Tercer escenario: La reutilización de un LOD ontológico. En la medida en que se fortalezca este escenario, se puede obtener un verdadero grafo ontológico, semántico y vinculado.

Cuarto Escenario: Reutilización, fusión y reingeniería de los recursos ontológicos. Los desarrolladores de ontologías reutilizan, combinan y reorganizan los recursos-ontológicos, en caso de ser necesario “extienden” o “mejoran” la definición del recurso.

F5. Definir Axiomas, Reglas: Los axiomas son expresiones lógicas asertivas, que se definen en las relaciones, predicados y restricciones. Las reglas (de morfología implicativa Antecedente-> Consecuente tipo *if ... then*) son una extensión de los axiomas que escalan la semántica, proporcionando capacidades de inferencia y razonamiento. El estado de la cuestión de LOD en el contexto mundial, no define aún como mandatorios estos lineamientos, no obstante, iniciativas como LOD2, exhortan a la comunidad a escalar en lenguajes de representación ontológica con estas propiedades extendidas. Para estas tareas se recomienda seguir el formato definido por MethOntology para la especificación de axiomas y reglas. En la fase de implementación hay dos tipos de representaciones que marcan estas tendencias (SWRL y JENA Rules). SWRL (Semantic Web Rule Language), es un lenguaje basado en lógica descriptiva, del tipo antecedente->consecuente, donde se definen un conjunto de condiciones atómicas -asertivas- que, de cumplirse, conducen a un valor de instancia, propiedad o atributo. Un ejemplo de representación de regla para modelar sería: Si *x1* tiene como padre a *x2* y *x2* es hermano de *x3*, entonces *x1* tiene como tío a *x3*; en este caso, el ejemplo se modelaría de la siguiente manera:

TienePadre
 $(?x1,?x2)\wedge\text{TieneHermano}(?x2,?x3)\Rightarrow\text{TieneTío}(?x1,?x3)$

En la sintaxis abstracta, la regla se escribiría como:

Implica(Antecedente(TienePadre(I-variable(x1)
I-variable(x2))
TieneHermano(I-variable(x2) I-variable(x3)))
Consecuente(TieneTio(I-variable(x1) I-
variable(x3))))

La representación en JENA es un poco más simple, pero menos enriquecida en operadores que conectan y relacionan los predicados



axiomáticos del cuerpo/cabeza de una regla; no obstante, han resuelto -por lo pronto- las limitaciones de descubrimiento de información que tiene SWRL, por su alta dependencia de software privativo de costo como JESS para estas tareas de razonamiento e inferencia. Una representación de regla en JENA para determinar que: si un estudiante $?s$ ¹⁴ tiene clase $?c$ y un profesor $?p$ enseña en la clase $?c$, entonces el profesor $?p$ tiene al estudiante $?s$, se modelaría cómo:

[Regla_Tiene_Estudiante: (?s :tiene Clase ?c) (?p :Enseña ?c) -> (?p :tiene Estudiante ?s)]

F6. Validación La validación de la Ontología involucra procesos de razonamiento, consistencia y verificación con expertos que permitan medir la completitud del dominio representado. Los lineamientos presentados en el capítulo anterior para la fase de validación se aplican igualmente para nuevos repositorios ontológicos.

F7 Detalles de Implementación Se presentan a partir de los procesos de formalización en Protégé, generando archivos RDF y OWL, pruebas de consistencia, pruebas de razonamiento, validación e inferencia de reglas, que provee los plugins de este framework.

3.4. Enriquecimiento de Datos

¹⁴ El uso del carácter comodín significa cualquier valor de la variable asociada

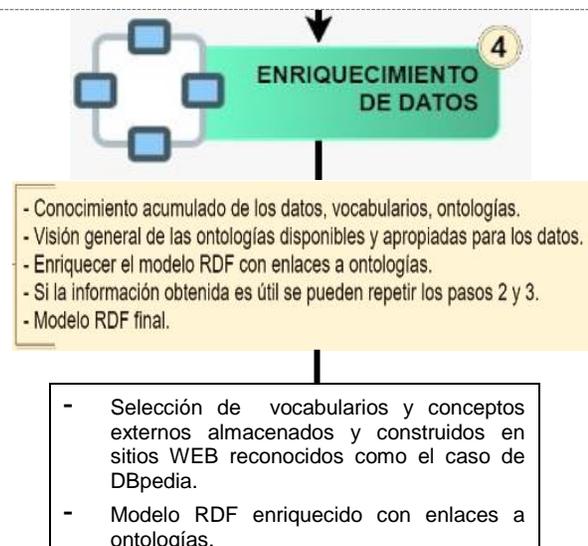


Figura 21. Detalle para Enriquecimiento de datos

El enriquecimiento de datos se genera construyendo relaciones entre conceptos, hecho que extiende la definición semántica de los mismos, a partir de vocabularios ya estandarizados y validados que se seleccionaron en el paso anterior. Para ejecutar el enriquecimiento, sólo se requiere conocer la URI del recurso externo y algunos aspectos de la descripción del modelo.

Para el enriquecimiento del conjunto de dato se deben analizar las fuentes de datos y vocabularios externos que tengan relación con el conjunto de datos; una vez identificados, se debe determinar su ubicación para establecer la URI que permita enlazarlo, enriqueciendo así la capacidad semántica del conjunto de datos que se va a transformar. Lo adecuado es reutilizar tanta cantidad de vocabulario disponible en la WEB como sea posible.

Esta reutilización acelera el desarrollo de la ontología (grafo/modelo) y- por lo tanto, los gobiernos ahorrarán tiempo, esfuerzo y recursos.



En el siguiente paso se muestra un ejemplo de cómo un dato de información asociado a una ciudad como barranquilla se puede enriquecer, asociándolo con el concepto que posee sobre esta ciudad la plataforma DBpedia.

3.5. Vinculación de los Datos

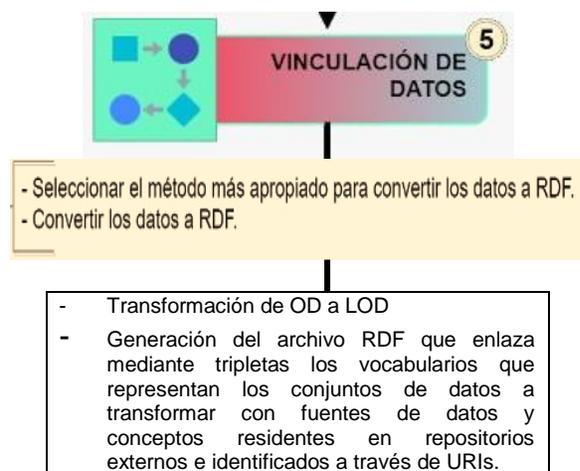


Figura 22. Detalle para Vinculación de Datos

El propósito de la vinculación de datos es convertir un conjunto de datos (Sean propios de la entidad o pertenecientes a algún paquete de datos abiertos) en una representación de Datos Abiertos Vinculados (LOD). Existen diferentes formas de convertir datos, entre las opciones se consideran: Scripts RDFicadores (convertidores de múltiples representaciones a RDF); Lenguajes de asignación declarativa; Lenguajes que realizan la traducción de consultas en lugar de la traducción de datos (por ejemplo, R2RML15).

¹⁵RDB to RDF Mapping Language, R2RML es un lenguaje para expresar mapeos personalizados de bases de datos relacionales a conjuntos de datos RDF.

Independientemente del enfoque utilizado, la conversión de datos implica mapear los datos fuente en un conjunto de declaraciones RDF. A medida que los datos se convierten, los datos se serializan en las declaraciones RDF.

La vinculación se da a través de la definición de enlaces entre diferentes repositorios o esquemas semánticos, lo cual permite navegar en redes semánticas y reutilizar las definiciones que sean definidas por el usuario. Así, a través de reglas y axiomas, se brinda la posibilidad de realizar la inferencia de un nuevo conocimiento.

La vinculación de los datos se realiza a través de las tripletas, las cuales proporcionan un método general y flexible para descomponer cualquier conocimiento en piezas pequeñas, representando así la semántica de cómo se relacionan los datos, el propósito de este tipo de representación es generar un lenguaje que sea interpretado por una máquina.

Para realizar la vinculación de los datos se debe contar con:

1. La representación de cada tipo de datos de la fuente (csv, xml, sql, tsv, xls, entre otros)
2. El conjunto de vocabularios nuevos y existentes que representan los conceptos y relación existentes en el conjunto de datos que se desea transformar.
3. Transformar la representación basada en duplas a un modelo semántico basado en tripletas Sujeto-Predicado-Objeto, que cumpla con los estándares de LOD.
4. Valorar y enriquecer la representación, ampliado sus posibilidades de aplicación.

3.5.1. Representación de datos



Una vez se definen los conjuntos de datos, se limpian y seleccionan los vocabularios que representan los datos, se procede a enlazar los conjuntos definidos; en este caso, se utilizarán datos abiertos del portal **ciudatos** (**¡Error! No se encuentra el origen de la referencia.**), el cual muestra información sobre la situación de las diferentes ciudades de Colombia

Ciudad del dato	Año del dato	Tasa de cobertura neta en preescolar	Tasa de cobertura neta en Primaria	Tasa de cobertura neta en Secundaria	Tasa de cobertura neta en Media
Barranquilla	2008	0.56	0.900	0.7500	0.5100
Barranquilla	2009	0.46	0.980	0.8000	0.5000
Barranquilla	2010	0.62	1.090	0.8700	0.5200
Barranquilla	2011	0.48	1.160	0.9517	0.6146
Barranquilla	2012	0.53	0.994	0.8771	0.5482

Figura 23. Indicadores de la ciudad de Barranquilla expuestos a través de datos abiertos

3.5.2. Buscar vocabularios

En el Metavocabulario (vocabulario propio universal) se encontrarían los conceptos de ciudad del dato, fecha (año), Tasa de cobertura neta en Preescolar, tasa de cobertura neta en Primaria, tasa de cobertura neta en Secundaria, etc.; estos conceptos serían transversales, únicos y generales a todas las dependencias del gobierno, además, podrían estar asociados a otros para su enriquecimiento, como lo puede ser a los indicadores demográficos del país.

3.5.3. Crear Tripletas

La tripleta busca dividir el conocimiento en un grafo dirigido y etiquetado. Cada arista en el

grafo representa un hecho, o relación entre dos cosas.



Figura 24. Tripleta Contrato - EstadoProceso - Ejecución

La arista en la Figura 24 del nodo etiquetado con Contrato se relaciona mediante EstadoProceso al nodo Ejecución. Un hecho representado de esta manera tiene tres partes: Un sujeto, un predicado y un objeto. El sujeto es lo que está al inicio de la arista, el predicado es el tipo de arista (La etiqueta) y el objeto es lo que está al final de la arista.

La transformación de datos abiertos a datos vinculados es un avance en la expresividad de la información, una forma de enriquecer los conjuntos de datos de *ciudatos* es vincularlos a otros repositorios que proporcionen significado a los conceptos. Para nuestro ejemplo el campo de ciudad (Barranquilla) se puede enlazar con el concepto definido en DBpedia - <http://dbpedia.org/page/Barranquilla> -, los indicadores se pueden vincular con el repositorio que poseen el DANE y el Ministerio de Salud y Proyección Social.

3.5.4. Enriquecer Representación

En la **¡Error! No se encuentra el origen de la referencia.**, se muestra como convergen los pasos 2, 3 y 4 en la presente guía, resultados que se consolidan en la estructuración del archivo RDF, el cual contiene las tripletas que vinculan los conceptos y atributos que



representan el conjunto de datos que se desea transformar con vocabularios y conceptos externos.

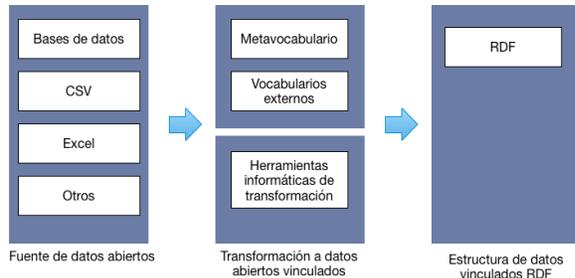


Figura 25. Proceso de transformación de datos abiertos a datos abiertos vinculados

El resultado de la etapa de transformación es un archivo con definición semántica a través de conceptos (predicados) en formato RDFs. Dentro de las herramientas que existen para la formación de los archivos RDF, se sugiere JENA, dotNet RDF, JSON, RDFa, JSON-LD, Turtle y N-Triples, RDF / XML¹⁶.

3.6. Publicación

¹⁶ En <https://www.w3.org/TR/ld-bp/#CONVERT> se describe las características de estas herramientas.



Figura 26. Detalle para Publicación de Datos

3.6.1. Preparación de los datos

Para poder reutilizar y permitir el procesamiento automático de grandes volúmenes de datos es necesario que la información esté publicada en los formatos adecuados y con el suficiente nivel de calidad y detalle, para que los datos sean aprovechables.

Algunos criterios a la hora de seleccionar los formatos adecuados son:

1. Formatos estructurados que siguen modelos de datos estandarizados.
2. Formatos directamente legibles por las máquinas.
3. Seleccionar los formatos más demandados entre los potenciales usuarios.
4. Formatos abiertos que permitan un uso sin restricciones para ampliar el público potencial.
5. Formatos que permitan publicar metainformación incorporada.



En muchos casos, es importante validar los datos con alguna herramienta que permita analizar la sintaxis mínima en el lenguaje de representación ontológico. La W3C¹⁷ provee un recurso en línea que facilita estas tareas, particularmente para evaluar errores mínimos de codificación y representación; asimismo, provee visualizaciones del grafo y conversión de formatos. Igualmente, se provee un conjunto de scripts que facilitan estas tareas de validación¹⁸.

El recurso (VisualDataWeb.org, S.F)¹⁹ apoya la visualización para verificar consistencia entre los elementos que conectan los conceptos (relaciones, predicados).

3.6.2. Publicación del conjunto de datos

Discutir las opciones para desarrollar una plataforma WEB, pensada para intercambiar, probar y vincular sus datos con otras fuentes pertenecientes a otras organizaciones, y quizás combinarlos conjuntamente para crear nuevos servicios.

La manera más fácil de publicar es utilizar servidores de RDF, como Virtuoso o Fuseki que ya vienen con un punto final (end point) de consulta SPARQL. El punto final permite a cualquier persona escribir consultas SPARQL contra datos RDF y no es necesario configurar herramientas y sistemas de consulta adicionales.

3.6.3. Publicación de MetaDatos

Una vez publicado el conjunto de datos, se debe incluir la información de metadatos relacionados. Para este propósito hay vocabularios ya mencionados y se reiteran recursos como:

VoID <https://www.w3.org/TR/void/>
DCAT <http://www.w3.org/TR/vocab-dcat/>
VOCAB ORG <http://www.w3.org/TR/vocab-org/>
DATA CUBE <http://www.w3.org/TR/vocab-data-cube/>

Esta tarea permite expresar metadatos sobre los conjuntos de datos RDF, y cubre metadatos generales, metadatos de acceso, metadatos y descripción de enlaces entre conjuntos de datos.

La procedencia de los conjuntos de datos gubernamentales juega un rol importante cuando se navega y explora los recursos del gobierno.

A su vez los metadatos servirán para describir los conjuntos de datos y los diferentes recursos (o archivos) que los componen. A continuación se muestra una tabla resumen de los metadatos que podemos proporcionar:

Tabla 7. Resumen de Metadatos

CONJUNTOS DE DATOS		
OBLIGATORIOS	RECOMENDADOS	OPCIONALES
Nombre	Etiqueta(s)	Fecha de creación
Descripción	Condiciones de uso / licencia	Fecha de la última actualización
Temática (s)		Frecuencia de actualización
Organismo		Idioma(s)

¹⁷ <https://www.w3.org/RDF/Validator/>

¹⁸ <https://github.com/lod-cloud/ckan-lod-validator>

¹⁹ <http://visualdataweb.de/webvowl/>



	Cobertura geográfica	
	Cobertura temporal	
	Vigencia	
	Recursos relacionados	
	Normativa	
DISTRIBUCIÓN (archivos o enlaces con los datos)		
OBLIGATORIOS	RECOMENDADOS	OPCIONALES
URL de acceso	Nombre	Tamaño
Formato	Información sobre formato	

LOD²⁰, para ello, debemos agregar una entrada de conjunto de datos en el repositorio CKAN²¹. El Grupo de trabajo de vinculación de datos abiertos proporciona algunas pautas para recopilar metadatos en conjuntos de datos vinculados en CKAN en su sitio <http://validator.lod-cloud.net/>.

Si resulta imposible mantener y admitir una tripleta y un punto final SPARQL en la infraestructura propia de TI de la institución, los datos RDF también se pueden compartir como simples archivos XML descargables en el sitio web.

Sin importar el método elegido para publicar los datos RDF (configuración de un punto final SPARQL, descarga simple o alguna otra forma), es crucial asegurarse de que otros puedan obtener información al respecto. Esto se puede hacer mediante la descripción de sus datos en los catálogos centrales de Linked Data. Estos existen principalmente a nivel de país, por ejemplo, <http://opendata.riik.ee/> para Estonia, <http://opendata.se/> para Suecia, <http://data.gov.uk/> para el Reino Unido, etc. y deben utilizarse si el alcance previsto de la reutilización de datos es nacional. A nivel internacional, la primera parada debería ser el catálogo de datos abiertos de Europa.

3.6.4. Habilitar descubrimiento efectivo

Esta actividad de publicación se relaciona con permitir la fidelidad de descubrimiento y sincronización del conjunto de datos que se desea publicar. Esta tarea consiste en definir *sitemaps* proporcionando indicadores de tiempo utilizando los campos *lastmod*, *changefreq* y *priority*, que se pueden usar para incorporar semántica en los motores de búsqueda web que solo descargan datos nuevos y páginas cambiadas.

Estos resultados permiten que los motores de búsqueda web (semánticos) descubran lo que es nuevo en el conjunto de datos, de manera eficiente y oportuna.

Posteriormente se podría incluir el conjunto de datos del gobierno en el diagrama de nubes

²⁰ Ver la nube [http://lod-\(cloud.net/\)](http://lod-(cloud.net/))

²¹ CKAN es un sistema de administración de datos, orientado a la publicación, intercambio y búsqueda de datos, con tecnología para interpretar LOD. El desarrollo es realizado por una asociación que reside en la Open Knowledge Foundation.

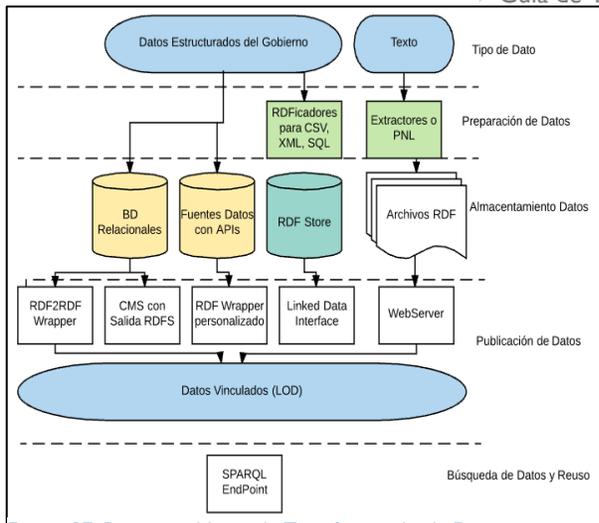


Figura 27. Propuesta Marco de Transformación de Datos Estructurados a Publicación de LOD (Fuente propia, 2017)

Publicar datos como Open Data es un buen comienzo, pero probablemente requiera actualizaciones en la medida que se agregue nueva información al conjunto de datos.

En la **¡Error! No se encuentra el origen de la referencia.**, se describe una arquitectura funcional que representa los datos y los procesos de transformación para publicar datos abiertos enlazados.

3.6.5. Construcción de sentencias SPARQL

El siguiente ejemplo muestra una consulta SPARQL para encontrar el título de un libro en el grafo de datos dado. La consulta consta de dos partes: la cláusula SELECT identifica las variables que aparecen en los resultados de la consulta, y la cláusula WHERE proporciona el patrón de grafo básico para la concordancia con el grafo de datos. El patrón de grafo básico de este ejemplo consiste en un único patrón de triplete con una sola variable (?title) en la posición del objeto.

```
<http://example.org/book/book1>
<http://purl.org/dc/elements/1.1/title>
"SPARQL Tutorial" .
```

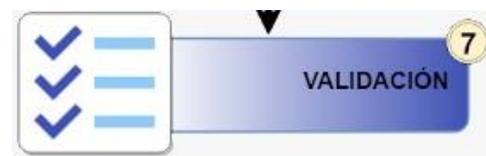
```
Consulta:
SELECT ?title
WHERE
{
  <http://example.org/book/book1>
  <http://purl.org/dc/elements/1.1/title> ?title .
}
```

Esta consulta, efectuada sobre los datos indicados anteriormente, tiene una solución:

Resultado de la consulta:

title
"SPARQL Tutorial"

3.7. Validación



- Conjuntos de datos que son candidatos para la próxima nube de LOD (y conectados).
- Conjuntos de datos con un nivel de integridad de 4 (revisado y ya en el grupo lodcloud)

Figura 28. Detalle para Validación de los Datos

El método CKAN²² ²³ (HPI, 2017), propone una metodología de validación para cumplir con unos requerimientos mínimos que faciliten la

²² Modelo de Código Abierto para el almacenamiento y distribución de datos abiertos

²³ <http://validator.lod-cloud.net/>



publicación de los datos en el LOD-Cloud²⁴. Esta tarea le proporcionaría un alto grado de visibilidad mundial a las entidades que se sometan a este riguroso pero importante proceso.

Los datos por presentar requieren de selección previa y la priorización debería adecuarse a las particularidades de cada caso. Los responsables de la información serán los mejores conocedores de los datos y su posible audiencia, pero algunos criterios que pueden servir de guía general son:

- Importancia de los datos para la sociedad, teniendo en cuenta su posible impacto y utilidad.
- Importancia de los datos para la economía, valorando el potencial de los datos para generar nuevas líneas de investigación o negocio.
- Actualidad de los datos, ya que en general, cuanto más actualizados tienen mayor potencial, aunque los datos históricos también pueden tener un gran valor.
- Cantidad de datos: cuanto mayor sea el volumen más valor añadido se podría aportar mediante el tratamiento posterior de la información.
- Nivel de detalle: cuanto más detallados sean los datos, mayores serán también las posibilidades de explotación.

La ponderación de los repositorios sensibles de publicarse en la nube LOD, se deben clasificar así:

Elementos Por Mejorar	Condiciones Cumplibles
Conjuntos de datos con un nivel de integridad <4	Conjuntos de datos que son candidatos para la próxima nube de LOD (y conectados). Conjuntos de datos con un nivel de integridad de 4 (revisado ya en el grupo <i>lodcloud</i>)
Simples Colecciones de conjuntos de datos	
Conjuntos de datos que necesitan más información del editor de los datos	
Conjuntos de datos que deben corregirse	
Conjuntos de datos que no tienen enlaces RDF externos o internos	

Tabla 8. Criterios de cuantificación para repositorios LOD (Fuente propia, 2017)

²⁴ <http://lod-cloud.net/>



4. GLOSARIO

AngularJS: Librería de JavaScript para aplicaciones MVC del lado del cliente.

API: Es una interfaz de programación que permite el acceso a servicios de datos.

APLICACIÓN WEB: Software que se ha desarrollado sobre tecnologías que pueden ser ejecutadas sobre un navegador.

ARQUITECTURA DE SOFTWARE: Define la estructura y los componentes de un sistema de información o aplicación.

AXIOMA: Proposiciones que deben ser asumidas por un modelo semántico.

BACK-END: Es la parte del software que se encarga de la interacción con los datos y en general de componentes del lado del servidor.

D3.js: Librería para el desarrollo de visualizaciones de datos.

DATASET: Conjunto de datos sobre una temática.

DATOS ABIERTOS (OPEN DATA): Son datos digitales puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar.²⁵

DATOS ABIERTOS ENLAZADOS (LINKED OPEN DATA): Es una poderosa combinación de Datos Vinculados y Datos Abiertos: está vinculado y utiliza fuentes de datos abiertas.²⁶

ENTE TERRITORIAL: Son entidades territoriales los departamentos, los distritos, los municipios y los territorios indígenas.²⁷

FRONT-END: Es la parte del software que se encarga de la capa de presentación.

HOVER: Acción de pasar el mouse sobre un objeto en una interfaz sin hacerle click.

HTTP: Es el protocolo de comunicación para transferencia de información en la web.

IP: Identificación única de un equipo de computo

JAVA: Lenguaje de programación bajo el paradigma de orientación a objetos.

JavaScript: Lenguaje de scripts utilizado por los navegadores.

JENA: software para la creación de modelos semánticos.

JSON: Formato de comunicación de datos.

MOCKUP: Bosquejo preliminar de una interface.

PERSISTENCIA DE DATOS: Almacenamiento de la información en memoria permanente.

PROTÉGÉ: Aplicación para la creación de ontologías.

RDF: Método para descripción de información.

SERVICIO WEB: Rutina que permite el acceso a la información a través de una aplicación.

SISTEMA OPERATIVO: Software base de un equipo de cómputo.

SPRING: *Framework* para el desarrollo de aplicaciones web en Java.

SQL: Lenguaje para hacer consultas a bases de datos

TAXONOMÍA: Hace referencia al término clasificación.

²⁵ <https://opendatacharter.net/principles-es/>

²⁶ (Gotttron & Staab. Linked Open Data, 2014).

²⁷ <http://www.mineducacion.gov.co/1621/article-82749.html>



TOMCAT: Implementación de software para servir aplicaciones web desarrolladas en Java.

URL: Dirección virtual utilizada en los navegadores para acceder a recursos disponibles en la web.



Referencias

- GINSBERG, W. (2011). *The Obama Administration's Open Government Initiative: Issues for Congress*. Washington DC: Congressional Research Service.
- Kaltenböck, M., Bauer, F., & Blumauer, A. (2012). Retrieved from <https://www.reeep.org>: <https://www.reeep.org/LOD-the-Essentials.pdf>
- Sentencia T 406 (Corte Constitucional Junio 5, 1992).
- Sentencia C-021 (Corte Constitucional Enero 1996).
- (2008). Retrieved from OpenKnowledge International: <http://opendatahandbook.org/guide/es/what-is-open-data/>
- Sentencia C-748 (Corte Constitucional Octubre 6, 2011).
- Sentencia C-274 (Corte Constitucional Mayo 9, 2013).
- Abedjan, Z. (2014). *Improving RDF Data with Data Mining*. Tesis Doctoral, Universität Potsdam - Hasso-Platter-Institut Potsdam, Alemania.
- Abele, A., McCrae, J. P., Buitelaar, P., Jentzsch, A., & Cyganiak, R. (2017, 08 22). <http://lod-cloud.net/>. Retrieved from <http://lod-cloud.net/>: http://lod-cloud.net
- AGHAEI, S., NEMARBAKSH, A., & KHOSRAVI, H. (2012, Enero). *Journals of AIRCC*. Retrieved from AIRCC Publishing Company: <http://airccse.org/journal/ijwest/papers/3112ijwest01.pdf>
- Banco Mundial. (2015, Septiembre). Retrieved Diciembre 4, 2017, from sitio web de la Estrategia de Gobierno en Linea: <http://estrategia.gobiernoenlinea.gov.co>
- Banco Mundial. (2015, Septiembre). *Estrategia de Gobierno en Linea*. Retrieved Diciembre 04, 2017, from <http://estrategia.gobiernoenlinea.gov.co>
- Barbosa, O., Pereira dos Santos, R., Alves, C., Werner, C., & Jansen, S. (2013). Systematic Mapping Study on Software Ecosystems from a Three-Dimensional Perspective. *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry, 1*, 59-81.
- BERNERS LEE, T. (2001). *The semantic WEB*. Scientific American.
- Berners Lee, T. (2005, Enero). *Uniform Resource Identifier (URI): Generic Syntax*. Retrieved from The Internet Engineering Task Force: <http://www.ietf.org/rfc/rfc3986.txt>
- BERNERS-LEE, T. (2005, Enero). *Uniform Resource Identifier (URI): Generic Syntax*. Retrieved from The Internet Engineering Task Force: <http://www.ietf.org/rfc/rfc3986.txt>
- BERNERS-LEE, T. (2009, Septiembre 18). *Linked Data*. Retrieved Diciembre 2017, from W3 organization: <https://www.w3.org/DesignIssues/LinkedData.html>
- Bloem, P., & K. D. de Vries, G. (2015, 25 25). Machine Learning on Linked Data, a Position Paper. *CEUR Workshop Proceedings (CEUR-WS.org), 1232*.
- Bühmann, L., L. J., & Westphal, P. (2016, 08). DL-Learner—A framework for inductive learning on the Semantic Web. *Elsevier, 39*, 15-24.
- Bustamante Donas, J. (2010, octubre-diciembre). Segundos pensamientos. La cuarta



- generación de derechos humanos en las redes digitales. *Revista Telos (Cuadernos de Comunicación e Innovación)*, 1-10.
- Calderon, A., Carolan, L., Palmer, R., & Vaughan, F. (2015, 10 28). *Open Data Charter*. Retrieved from Open Data Charter: <https://opendatacharter.net/principles-es/>
- CARDONA, M., CANO, C., & RAMÍREZ, C. (2002). *Tecnologías, organización y políticas: Mundos de producción en las PyMEs en Colombia en el periodo 1990 - 2002*. Medellín: Universidad de Medellín.
- CASTELLS, P. (2009). *Universidad Autónoma de Madrid*. Retrieved 11 15, 2017, from <http://ir.ii.uam.es/castells/>: <http://arantxa.ii.uam.es/~castells/publications/castells-uclm03.pdf>
- Castro, L. (2016, Marzo). *¿Qué es URI?* Retrieved from ABOUT Español: <https://www.aboutespanol.com/que-es-uri-157626>
- CASTRO, L. (2016, Marzo). *¿Qué es URI?* Retrieved from ABOUT Español: <https://www.aboutespanol.com/que-es-uri-157626>
- CODINA, L. (2009). *¿Web 2.0, Web 3.0 o Web Semántica?* España: Bilbao.
- CORCHO, O., FERNÁNDEZ-LÓPEZ, M., GÓMEZ-PÉREZ, A., & LÓPEZ-CIMA, Á. (n.d.). *Construcción de ontologías legales con la metodología Methontology*. Madrid: Universidad Politécnica de Madrid.
- CÓRDOBA TRIVIÑO, J. (2008). *corteconstitucional.gov.co*. Retrieved from <http://www.corteconstitucional.gov.co/relatoria/2008/C-1011-08.htm>
- Cotino Hueso, L. (2013). Derecho y «Gobierno Abierto». La regulación de la transparencia y la participación y su ejercicio a través del uso de las nuevas tecnologías y las redes sociales por las Administraciones públicas. Propuestas concretas. *Monografías de la Revista Aragonesa de Administración Pública*, 51-92.
- Datos Abiertos Gobierno Digital Colombia. (2017). *datos.gov.co*. Retrieved 12 12, 2017, from datos.gov.co: <https://datos.gov.co/>
- Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016, Enero). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33, 15-27.
- DBpedia del español. (2014, noviembre). *DBpedia*. Retrieved from <http://es.dbpedia.org/>
- Departamento Nacional de Planeación. (2000, Febrero 09). Retrieved Diciembre 10, 2017, from <http://www.mintic.gov.co>: http://www.mintic.gov.co/portal/604/articles-3498_documento.pdf
- Domingos, P. (2003). Prospects and challenges for multirelational data mining. *SIGKDD Explorations*.
- ECLAC Naciones Unidas. (2014). *Big data and Open data as sustainability tools*. Santiago de Chile: United Nations.
- EUROPEAN COMMISSION. (2004). *European interoperability framework for pan-european eGovernment services*. Belgium: European communities.
- Fensel, D. (2001). *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag.



- FERNÁNDEZ BREIS, J. (2003). *Un Entorno de Integración de Ontologías para el desarrollo de Sistemas de Gestión del Conocimiento*. Murcia - España: Tesis doctoral - Univesidad de Murcia.
- Fernández, M., Gómez-Pérez, A., & Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*, 33-40.
- Franco-Bedoya, O., Ameller, D., Costal, D., & Franch, X. (2014). QuESo: A Quality Model for Open Source Software Ecosystems. In *Proceedings of the 9th International Conference on Software Engineering and Applications ICSOFT-EA 2014* (pp. 209–218). *Software Engineering and Applications (ICSOFT-EA)*, 209-221.
- Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. United of States Of America: Cambridge university press.
- Getoor, L. (2015). Link Mining: A New Data Mining Challenge. *SIGKDD Explorations*, 4, 1-6.
- GIRALDO RENDON, J. P. (2015). *Metamodelo para implantación del proyecto Gobierno en Línea en la Republica de Colombia*. Madrid: UPSAM.
- Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2009). *Research at Sofia University*.
- Gómez-Pérez, A., Villazón-Terrazas, B. C., & Suárez-Figueroa, M. (2010). A Pattern-Based Method for Re-Engineering Non-Ontological Resources into Ontologies. *International Journal on Semantic Web & Information Systems*, 27-63 .
- Gottron, T., & Staab, S. (2014). *Encyclopedia of Social Network Analysis and Mining*. Retrieved from https://link.springer.com/referenceworkentry/10.1007%2F978-1-4614-6170-8_111
- GRUBER, T. (1993, Abril). *A Translation Approach to Portable Ontology Specifications*. Stanford: Stanford Unviersity.
- Guyon, I. (2006). *Feature Extracton - Foundations and Applications*. Springer.
- GUZMAN LUNA, J., LÓPEZ BONILLA, M., & DURELY TORRES, I. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia et Technica*, 133-141.
- Harris, S., & Seaborne, A. (2013, 03 21). W3C. Retrieved from W3C: <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>
- Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space* (Vol. 1).
- Herrera Vergara, H. (1994). [corteconstitucional.gov.co](http://www.corteconstitucional.gov.co). Retrieved from <http://www.corteconstitucional.gov.co/relatoria/1994/C-180-94.htm>
- Horridge, M., & Patel-Schneider, P. F. (2008, 10 05). [w3.org](http://www.w3.org). Retrieved from [w3.org](https://www.w3.org/2007/OWL/wiki/ManchesterSyntax): <https://www.w3.org/2007/OWL/wiki/ManchesterSyntax>
- HPI. (2017, 12 18). *Data Hub LOD Datasets*. Retrieved from *Data Hub LOD Datasets*: <http://validator.lod-cloud.net/>
- Inokuchi, A., Washio, T., & Motoda, H. (2000). An Apriori-based Algorithm for Mining Frequent Substructures from Graph Data. *Principles of Data Mining and knowledge Discovery*, 13-23.
- Izasa Echeverry, G., Arango, J., Vélez, J. I., Franco, O., & Henao, A. (2017). *LINEAMIENTOS PARA LA IMPLEMENTACIÓN DE LINKED*



- OPEN DATA (LOD) EN COLOMBIA*. Bogota: Mintic.
- JACOBS Ian - W3C. (2004, Diciembre). *Architecture of the World Wide Web, Volume One*. Retrieved from W3C Recomendations: <https://www.w3.org/TR/webarch/#id-resources>
- Jansen, S., Brinkkemper, S., Souer, J., & Luinenburg, L. (2012, Julio). Shades of gray: Opening up a software producing organization with the open software enterprise model. *Journal of Systems and Software*, 85, 1495-1510.
- Jansen, S., Finkelstein, A., & Brinkkemper, S. (2009, Mayo 16-24). A sense of community: A research agenda for software ecosystems. *Software Engineering - Companion*.
- Jensen, D. (2000). Statistical Challenges to Inductive Inference in Linked Data. *AAAI Technical Report FS-98-01*, 59-62.
- K.D. de Vries, G. (2013). A Fast Approximation of the Weisfeiler-Lehman Graph Kernel for RDF Data. *springer*, 606-621.
- KARMANOVSKIY, N., MOUROMTSEV, D., & et all. (2016). Linked Open Data in University. In F. I. Transformation, *Digital Transformation and Global Society* (pp. 400-420). San Petersburgo: Springer.
- KARMANOVSKIY, N., MOUROMTSEV, D., & et all. (2016). Linked Open Data in University. In F. I. Transformation, *Digital Transformation and Global Society* (pp. 400-420). San Petersburgo: Springer.
- Karmanovskiy, N., Mouromtsev, D., Navrotskiy, M., Dmitry, P., & Radchenko, I. (2016, Noviembre 24). A Case Study of Open Science Concept: Linked Open Data in University. 400-403.
- Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Retrieved from <https://dl.acm.org/citation.cfm?doid=1134285.1134500>
- Lee, D. (2014, Octubre 27-30). Building an open data ecosystem: an Irish experience. *ACM. Digital Library*, 351-360.
- Lehmann, J. (2006). *DL-Learner Manual*.
- LOPEZ, M. (2010). *Ciudadanía Digital, un modelo de implantación en la región de Manizales y Caldas, Colombia*. Madrid: UPSAM - Tesis Doctoral.
- Lüosch, U., Bloehdorn, S., & Rettinger, A. (2012). Graph kernels for RDF data. *Springer*.
- Manikas, K., & Hansen, K. M. (2013). Reviewing the health of software ecosystems-a conceptual framework proposal. 33-44.
- Marketplace. (2011). *marketplace.rapidminer.com*. Retrieved from [marketplace.rapidminer.com: https://marketplace.rapidminer.com/UpdateServer/faces/product_details.xhtml?productId=rmx_semweb](https://marketplace.rapidminer.com/UpdateServer/faces/product_details.xhtml?productId=rmx_semweb)
- Massal, J., & Sandoval, C. G. (2010). Gobierno electrónico. ¿Estado, Ciudadanía y democracia en Internet? *Análisis Político*(68), 3-25.
- Michalski, R. S., Carbonell, J. G., & Mitchell, T. (1983). *Machine Learning*. Morgan Kaufmann.
- Ministerio de Tecnologías de la Información y de las Comunicaciones. (2015). Retrieved Febrero 24, 2015, from <http://programa.gobiernoenlinea.gov.co/ques.html>: <http://programa.gobiernoenlinea.gov.co/ques.html>



- Ministerio de Tecnologías de la Información y las Comunicaciones (MINTIC). Programa Gobierno en línea. (2015, 08 03). *MINTIC.GOV*. Retrieved from MINTIC.GOV: http://www.mintic.gov.co/portal/604/articles-7244_recurso_1.pdf
- Ministerio de Tecnologías de la Información y las Comunicaciones. (2011, Diciembre). Retrieved Diciembre 07, 2017, from http://programa.gobiernoenlinea.gov.co/apc-aa-files/da4567033d075590cd3050598756222c/Resumen_Ejecutivo_Datos_Abiertos.pdf
- MinTIC & Gobierno en Linea Colombia. (2011). *LINEAMIENTOS PARA LA IMPLEMENTACION DE DATOS ABIERTOS EN COLOMBIA*. Retrieved Diciembre 07, 2017, from http://programa.gobiernoenlinea.gov.co/apc-aa-files/da4567033d075590cd3050598756222c/Resumen_Ejecutivo_Datos_Abiertos.pdf
- MinTIC. (2015). *gobiernoenlinea*. Retrieved from Ministerio de Tecnologías de la Información y las Comunicaciones: <http://vive.gobiernoenlinea.gov.co/>
- Mintic. (2016). *Guia de Datos Abiertos en Colombia*. Bogotá: Mintic.
- Mitleton-Kelly, E. (2003). en principles of complexity & enabling infrastructures. *Complex Systems and Evolutionary Perspectives of Organisations: the Application of Complexity Theory to Organisations*. Elsevier., 23-50.
- Nickel, M., Tresp, V., & Kriegel, H. (2011). A Three-Way model for collective learning on Multi-Relational data. *Proceedings of the 28th International Conference on Machine Learning*, 809 - 816.
- ÑUÑO, R. (2005). Conocimiento organizativo y ciencia de la complejidad. . *Aprende RH: la revista de los recursos humanos y del e-learning*, 92-96.
- OCDE. (2017, Julio 02). *oecd.org*. Retrieved from <http://www.oecd.org/gov/digital-government/digital-government-review-colombia-key-finding-es.pdf>
- OEG. (2015). *Ontology Engineering Group*. Retrieved from Ontology Engineering Group: <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/methodologies/59-neon-methodology/>
- Open Data Charter. (2010). *Principios Datos Abiertos*. Retrieved Julio 2018, from WEB Open Data Charter: <https://opendatacharter.net/principles-es/>
- OPEN GOVERNMENT PARTNERSHIP. (2017). *opengovpartnership.org*. Retrieved from opengovpartnership.org: <https://www.opengovpartnership.org/declaracion-de-gobierno-abierto>
- OPEN KNOWLEDGE FOUNDATION. (2017, Noviembre). *Guia para datos abiertos*. Retrieved from Open Knowledge: <http://opendatahandbook.org/guide/es/what-is-open-data/>
- Orduña, P., Almeida, A., Aguilera, U., Laiseca, X., Gómez-Goiri, & Aitor. (2010). *morelab*. Retrieved from morelab: http://www.morelab.deusto.es/code_injection/
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2017, Octubre). Retrieved Diciembre 7, 2017, from sitio web de la OECD: <http://www.oecd.org>
- Ortiz de Zárate, A., de la Fuente, C., & de los Ríos, N. (2015, 10 28). Open Data Readiness Assessment (ODRA). República de Colombia. Colombia.



- Paulheim, H. (2010). *Exploiting Linked Open Data as Background Knowledge in Data Mining*. University of Mannheim. Germany: Research Group Data and Web Science.
- Peralta Duque, B. (2010, Enero - Diciembre). La forma en que se ejerce la ciudadanía en la democracia colombiana -socialización política y cultura ciudadana-. *ELEUTHERA*(4), 253-290.
- Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. *Proceedings of the 12th EASE*, 68-77.
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for Conducting Systematic Mapping Studies. *Software Engineering*(64), 1-18.
- Protégé. (2016). Protégé. Retrieved from <https://protege.stanford.edu/>
- Puron-Cid, G., & Garcia Diaz, S. (2014, abril 1). *Análisis Comparativo e Inter-temporal de la Adopción de Gobierno Electrónico: Los Casos de México y Colombia*. Retrieved 12 10, 2017, from www.researchgate.net: https://www.researchgate.net/profile/Gabriel_Puron-Cid/publication/295853650_Analisis_Comparativo_e_Inter-temporal_de_la_Adopcion_de_Gobierno_Electronico_Los_Casos_de_Mexico_y_Colombia/links/56cdf4c408ae4d8d64998b68/Analisis-Comparativo-e-Inter-temporal
- Ramirez-Alujas, A. V. (2010, Julio-diciembre). Innovación en la Gestión Pública y Open Government (Gobierno Abierto): Una Vieja Nueva Idea (Innovation in Public Management and Open Government: An Old New Idea). *Buen Gobierno*(9), 1-35.
- Riascos Erazo, S. C., Martínez Giordano, G., & Solano Rodríguez, O. J. (2008). El Gobierno Electrónico como estrategia de participación ciudadana en la Administración pública a nivel de Suramérica -Casos Colombia y Uruguay. *COLLECTeR Iberoamérica*, 1-10.
- Russell, S. (2001). Identity uncertainty. *Proc. of IFSA-01*.
- SALAZAR, J., MEJIA, M., & JOYANES, L. (2015, Julio). *Gobierno Electronico*. Retrieved Diciembre 1, 2017, from http://yogob.org/wp-content/uploads/2015/12/Gobierno-electronico-a-gobierno-abierto-Mirada-global-en-la-ausencia-de-procesos-descubrimiento-de-conocimiento_Johnny-Salazar.docx
- SAURAZ PÉREZ, P. (2012). *Diseño de una ontología para aplicaciones en el dominio de movilidad sostenible*. Valladolid: Master - Univesidad Valladolid.
- Schema. (s.f). *Schema.org*. Retrieved from <http://schema.org/>
- Sen, P., Namata, G., Bilgic, L., & Getoor, M. (2008). Collective classification in network data. *AI Magazine*.
- Siegel, D. (2010). *Semantic Focus*. Retrieved from Semantic Focus: <http://www.semanticfocus.com/media/insets/semantic-web-layer-cake-2.png>
- SPC shared infrastructures: semantic interoperability. (2013, Febrero 19). *agid.gov.it*. Retrieved from http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/semanticinteroperabilitylod_en_3.pdf
- SUÁREZ DE FIGUEROA BOANZA, M. (2010, Junio). *Tesis doctoral: NeON Methodology for Building Ontology Networks*. Retrieved



- Dicembre 2017, from Universidad Politecnica de Madrid:
http://oa.upm.es/3879/2/MARIA_DEL_CARMEN_SUAREZ_DE_FIGUEROA_BOANZA.pdf
- Suárez de Figueroa Boanza, M. d. (2010, Junio). *Tesis doctoral: NeON Methodology for Building Ontology Networks*. Retrieved Dicembre 2017, from Universidad Politecnica de Madrid:
http://oa.upm.es/3879/2/MARIA_DEL_CARMEN_SUAREZ_DE_FIGUEROA_BOANZA.pdf
- Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. (2015). The NeON Methodology framework: A scenario-based methodology for ontology development. *Applied Ontology*, 10(2), 107–145.
- Tauberer, J. (2014). *Open Government Data: The Book*.
- Tresp, V., Bundschus, M., Rettinger, A., & Huang, Y. (2008). Towards machine learning on the semantic web. *Springer*, 282-314.
- Trist, E., & Murray, H. (1997). *The Social Engagement of Social Science, Volume I: A Tavistock Anthology: The Socio-Psychological Perspective (Vol. III)*. (Vol. 1 y 2).
- Vasquez, T. (2006, Julio Recuperado el 27 de Marzo de 2014, de <http://www.institut-gouvernance.org/es/analyse/fiche-analyse-238.html#1207>). *Instituto de investigacion y debate sobre la gobernanza*. Retrieved Marzo 27, 2014, from irg: <http://www.institut-gouvernance.org/es/analyse/fiche-analyse-238.html#12>
- Villazón Terrazas, B., Wilches Vázquez, L., Corcho, O., & Gómez Pérez, A. (2010, Diciembre). *Lecturas sugeridas*. Retrieved from Laboratorio de investigación en informatica:
https://www.lri.fr/~hamdi/datalift/tuto_inspire_2012/Suggestedreadings/egovld.pdf
- VILLAZÓN-TERRAZAS, B., WILCHES-VÁZQUEZ, L., CORCHO, O., & GÓMEZ PÉREZ, A. (2010, Diciembre). *Lecturas sugeridas*. Retrieved from Laboratorio de investigación en informatica:
https://www.lri.fr/~hamdi/datalift/tuto_inspire_2012/Suggestedreadings/egovld.pdf
- VisualDataWeb.org. (S.F). *Visual Data Web*. Retrieved from Visual Data Web:
<http://visualdataweb.de/webvowl/>
- W3C. (2012, 08 29). *w3.org*. Retrieved from w3.org:
https://www.w3.org/2011/gld/wiki/GLD_Life_cycle
- W3C. (2014). Best Practices for Publishing Linked Data. *W3C Working Group Note*.
- WOOD, D., ZAIDMAN, M., RUTH, L., & HAUSENBLAS, M. (2013). *Linked Data*. Manning.
- World Bank Group. (2015). *herramientas.datos.gov.co*. Retrieved from herramientas.datos.gov.co:
<https://herramientas.datos.gov.co/sites/default/files/Comparativo%20Plataformas%20Datos%20Abiertos%20Banco%20Mundial2015.pdf>
- World Wide Web FOUNDATION and Open Data Barometer. (2008-2017). *Open Data Barometer. org*. Retrieved from Open Data Barometer. org:
<http://opendatabarometer.org/4thedition/report/>
- YOUNES MORENO, D. (2009). DERECHO CONSTITUCIONAL COLOMBIANO. *Octava Edición Actualizada*, 545. Bogota D.C.: Grupo Editorial IBÁÑEZ.