



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

TESIS PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SISTEMAS E INFORMÁTICA

**TEMA: EL IMPACTO DE LOS DATOS ABIERTOS ENLAZADOS EN LAS
PRINCIPALES INSTITUCIONES PÚBLICAS ENFOCADO EN LA
INNOVACIÓN DE SERVICIOS PÚBLICOS**

AUTOR: CUEVA PEÑARANDA, MARLON RICARDO

DIRECTOR: ING. CARLOS PROCEL

SANGOLQUÍ, OCTUBRE 2015

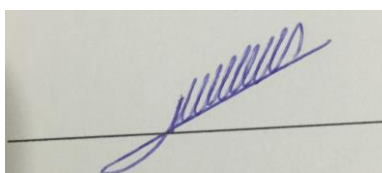
CERTIFICACIÓN

Ing. Carlos Procel

CERTIFICA

Que el trabajo titulado “EL IMPACTO DE LOS DATOS ABIERTOS ENLAZADOS EN LAS PRINCIPALES INSTITUCIONES PÚBLICAS ENFOCADO EN LA INNOVACIÓN DE SERVICIOS PÚBLICOS” realizado por el señor Marlon Ricardo Cueva Peñaranda, ha sido guiado y revisado periódicamente y cumple como requerimiento parcial a la obtención del título de INGENIERÍA EN SISTEMAS E INFORMÁTICA.

Sangolquí, Octubre, 01 del 2015



Ing. Carlos Procel

DECLARACIÓN DE RESPONSABILIDAD

Cueva Peñaranda Marlon Ricardo

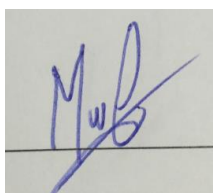
DECLARO QUE:

La presente tesis de grado titulada “EL IMPACTO DE LOS DATOS ABIERTOS ENLAZADOS EN LAS PRINCIPALES INSTITUCIONES PÚBLICAS ENFOCADO EN LA INNOVACIÓN DE SERVICIOS PÚBLICOS” ha sido desarrollada en base a una investigación exhaustiva, respetando derechos intelectuales de terceros, conforme las citas que constan en la bibliografía.

Consecuentemente este trabajo es de nuestra autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance científico del proyecto de grado en mención.

Sangolquí, Octubre, 01 2015



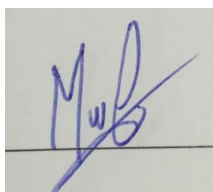
Marlon Cueva

AUTORIZACIÓN DE PUBLICACIÓN

Cueva Peñaranda Marlon Ricardo
AUTORIZO:

A la UNIVERSIDAD DE LAS FUERZAS ARMADAS de la publicación de la Biblioteca Virtual de la Institución, del trabajo titulado “EL IMPACTO DE LOS DATOS ABIERTOS ENLAZADOS EN LAS PRINCIPALES INSTITUCIONES PÚBLICAS ENFOCADO EN LA INNOVACIÓN DE SERVICIOS PÚBLICOS”, que es de mi propia autoría y responsabilidad.

Sangolquí, Octubre, 01 del 2015

A handwritten signature in blue ink, appearing to read 'M. C.', is written over a horizontal line.

Marlon Cueva

DEDICATORIA

El presente trabajo va dedicado a mis padres, Juan Carlos Cueva, Dania Peñaranda, y a mis tíos, Alicia Cueva, Luis Taípe, que me han brindado su apoyo incondicional a lo largo de la vida y en especial en esta etapa de la vida que va culminando, guiándome con sus reprimendas y concejos, gracias a los cuales han hecho posible que este trabajo sea una realidad.

A toda mi familia que han sabido preocuparse y acompañarme en su momento a lo largo de toda mi vida como estudiante universitario y crecimiento profesional.

Marlon Ricardo Cueva Peñaranda

AGRADECIMIENTOS

Agradezco a mi Dios por haberme sostenido todo este tiempo de vida universitaria, guiándome con su sabiduría e inteligencia para culminar exitosamente una etapa más de mi vida profesional.

Gracias a mi Director de Tesis el Ing. Carlos Procel por su paciencia y valioso conocimiento brindado, para desarrollar la investigación

Gracias mi Director de Carrera el Ing. Mauricio Campaña, por su preocupación, solidaridad y buenos consejos para con sus estudiantes.

A la prestigiosa Universidad de las Fuerzas Armadas y a todos sus maestros, por brindarme sus conocimientos que me supieron transmitir a lo largo de todo mi vida de estudiante.

A la Secretaria Nacional de la Administración Pública, por brindarme todo su apoyo y conocimiento para desarrollar la investigación.

A mi novia Katherine Granda y mis amigos por siempre estar preocupados y ayudarme con la finalización del trabajo.

Marlon Ricardo Cueva Peñaranda

ÍNDICE DE CONTENIDOS

CERTIFICACIÓN	ii
DECLARACIÓN DE RESPONSABILIDAD	iii
AUTORIZACIÓN DE PUBLICACIÓN	iv
DEDICATORIA	v
AGRADECIMIENTOS	vi
ÍNDICE DE CONTENIDOS	vii
ÍNDICE DE FIGURAS	x
ÍNDICE DE TABLAS	xi
ÍNDICE DE ANEXOS	xii
ÍNDICE DE TABLAS ANEXOS	xii
RESUMEN	xiii
ABSTRACT	xiv
1. CAPÍTULO 1	1
Introducción	1
1.1 INTRODUCCIÓN	1
1.2 JUSTIFICACIÓN E IMPORTANCIA	2
1.3 PLANTEAMIENTO DEL PROBLEMA	4
1.4 FORMULACIÓN DEL PROBLEMA	5
1.5 HIPÓTESIS	6
1.6 OBJETIVO GENERAL	6
1.7 OBJETIVOS ESPECÍFICOS	6
2. CAPÍTULO 2	7
ESTADO DEL ARTE.....	7
2.1 Terminología	7
2.1.1 Datos Abiertos	7
2.1.2 Las 5 estrellas de los datos abiertos	8
2.1.3 Datos Abiertos Enlazados	9
2.1.4 RDF	10
2.1.5 Grafos RDF	11
2.1.6 SPARQL	11
2.1.7 URI	11
2.1.8 Análisis y negociación del contenido	12

2.1.9 Representación de la información	13
2.2 Propuesta para la publicación de datos abiertos enlazados	14
2.2.1 Como publicar datos enlazados en la web	14
2.3 Estructuras para la publicación de datos abiertos enlazados	17
2.3.1 Publicaciones basadas en bases de datos relacionadas.....	17
2.3.2 Publicaciones basadas en APIs	18
2.3.3 Publicaciones basadas en datos estructuración.....	18
2.3.4 Publicaciones Basadas En Archivos RDF.....	18
2.3.5 Publicaciones basadas en documentos de texto	19
2.3.6 Publicaciones basadas en almacenes RDF	19
2.4 Herramientas	19
2.4.1 Almacenamiento de datos.....	20
2.4.2 Generación de grafos RDF sobre HTTP.....	21
2.4.3 Visualizadores de datos enlazados.....	23
2.4.3.1 Navegadores.....	23
2.4.3.2 Visualizadores.....	24
2.5 Casos de Éxito.....	24
2.5.1 data.gov.uk.....	25
2.5.2 data.gov	26
2.5.4.1 datos.bcn.cl	30
2.5.9 datos.gob.es	30
2.6 Los datos abiertos en Latinoamérica.....	32
2.6.1 El Ecuador y los datos abiertos.....	32
2.6.2 Nivel de madurez de los datos abiertos en Latinoamérica	36
3 CAPÍTULO 3.....	38
GUIA METODOLÓGICA.....	38
3.1 Descripción	38
3.2 Propuesta Metodológica.	38
3.3 Arquitectura de soporte para datos abiertos enlazados	39
3.4 Fases de implementación	41
3.4.1 Conceptualización	41
3.4.2 Diseño de Ontologías.....	42
3.4.3 Modelamiento del grafo RDF.....	43
3.4.4 Implementación del Endpoint SPARQL.....	45
3.4.5 Implementación de grafo RDF sobre Http	46
3.4.6 Actualización del grafo RDF	46

3.4.7 Web de documentación.....	47
3.4.8 Requerimientos no funcionales	48
3.4.9 Herramienta alternativa de visualización de datos.....	48
4 CAPÍTULO 4.....	50
DATOS ABIERTOS RELEVANTES A PUBLICAR	50
4.1 Antecedentes del Proyecto.....	50
4.2 Los ocho principios de los datos abiertos.....	52
4.3 Conceptualización	53
4.3.1 Identificación de las necesidades	53
4.3.2 Que datos se van a entregar	54
4.3.3 Quien va a consumir los datos	60
4.3.4 Formatos para la liberación de información	61
4.3.5 Como reutilizar los datos liberados.....	63
4.5 Perspectiva tecnológica.....	66
4.5.1 Acceso	67
4.5.2 Calidad de los datos	70
4.6 Aplicaciones a generar	72
5 CAPÍTULO 5.....	75
CONCLUSIONES Y RECOMENDACIONES.....	75
5.1 Conclusiones	75
5.2 Recomendaciones	76
BIBLIOGRAFÍA	78
ANEXOS	80
BIOGRAFÍA.....	86
HOJA DE LEGALIZACIÓN DE FIRMAS	87

ÍNDICE DE FIGURAS

Figura 1 Propuesta publicación de Datos Abiertos Enlazados.....	17
Figura 2 Estadística de Datos Abiertos en Reino Unido.....	25
Figura 3 Estadísticas Datos Abiertos en EEUU.....	27
Figura 4 Estadística Datos Abiertos Chile.....	29
Figura 5 Estadísticas Datos Abiertos España.....	31
Figura 6 Número portales de Datos Abiertos por País.....	33
Figura 7. Numero portales de Datos Abiertos y Enlazados por País.....	34
Figura 8. Instituciones Involucradas en los Datos Abiertos por País.....	34
Figura 9. Conjuntos de datos liberados por País.....	35
Figura 10. Aplicaciones elaboradas en base a Datos Abiertos por país.....	35
Figura 11. Participación en Iniciativas de Datos Abiertos.....	37
Figura 12. Propuesta de Arquitectónica de Infraestructura para publicar Datos Abiertos Enlazados.....	40
Figura 13. Proceso de liberación de información en las entidades publicas.....	58
Figura 14 diagrama de casos de uso.....	60
Figura 15 Elementos a incluir en una estrategia para la reutilización de datos.	65

ÍNDICE DE TABLAS

Tabla 1 formatos representación de información o recurso	14
Tabla 2 Herramientas de navegación para Datos Abiertos Enlazados.....	23
Tabla 3 Herramientas de visualización para Datos Abiertos Enlazados.....	24
Tabla 4 Aplicaciones Publicadas por Renio Unido.....	26
Tabla 5 Estadísticas Datos Abiertos en EEUU	27
Tabla 6 Aplicaciones publicadas por el gobierno de Chile.....	29
Tabla 7 Aplicaciones Publicadas en España	31
Tabla 8 Dominios importantes para liberar información	599
Tabla 9 define los formatos en que los datos abiertos se debían entregar	61
Tabla 10. Tipos de archivos según su uso.....	622
Tabla 11, Estrellas Datos Abiertos.....	633
Tabla 12. Licencias para publicación de datos.....	633

ÍNDICE DE ANEXOS

Anexos A-Glosario.....	80
Anexos B - Sistemas Informáticos y Portales Existentes.....	83

ÍNDICE DE TABLAS ANEXOS

Tabla de Anexos 1-Sistemas y Portales	833
Tabla de Anexos 2- Portales y sistemas en ejecución.....	844
Tabla de Anexos 3-Proyectos a Ejecutar	855

RESUMEN

En estos últimos 5 años se ha empezado a vivir una revolución en cuanto al acceso a la información pública por parte de la población. A esta iniciativa se le ha denominado “Open Data” o “Datos Abiertos”, la cual consiste en poner a disposición de la ciudadanía los datos de interés común para que, de una u otra forma, éstos actores puedan aportar con nuevas ideas o aplicación que entregue nuevos datos, conocimientos u otros servicios que el gobierno no es capaz de entregar, aportando a la innovación del Estado en sí. Esta revolución de la información ha creado un nuevo paradigma a través del cual las entidades públicas ya no se encargaran de generar todas las aplicaciones necesarias para la población sino que la sociedad misma a través de empresas privadas o personas naturales son quienes generan sus propios aplicativos para aprovechar de la mejor manera los datos que las entidades públicas dejan a disposición en general. El presente documento muestra el impacto y los beneficios que tienen los datos abiertos en el Ecuador, donde se indicara en que campos, sectores, industrias, instituciones u organizaciones son los que más requieren de los datos abiertos, junto a una propuesta metodológica para medir los resultados. En definitiva se pretende motivar a las autoridades competentes de las entidades afines a la SNAP acerca de los enormes cambios que este nuevo paradigma trae consigo para el ciudadano, y dar a conocer los nuevos desafíos que esta nueva ola está comenzando a tener en los países de la región, que rompe con la trayectoria que hasta ahora llevaba la evolución del denominado gobierno electrónico o gobierno digital ampliando su alcance más allá de la mera simplificación de trámites y de la mayor transparencia de información por parte de los gobiernos de la gestión.

Palabras clave

- Web Semántica
- Datos Abiertos Enlazados
- Metodología Web Semántica
- RDF
- SPARQL

ABSTRACT

In the last 5 years it has begun to live a revolution in terms of access to public information by the population. The initiative has been called "Open Data" or "Open Data" which is to make available to the public data of common interest so that one or the other form, these actors can contribute with new ideas or application to deliver new data, knowledge or other services that the government is not able to deliver, contributing to the innovation of the State itself. This information revolution has created a new paradigm through which public entities no longer be in charge of creating all the necessary applications for the population but society itself through deprive companies or individuals are those who create their own applications for best way to take advantage of data that public entities cease generally available. This document shows the impact and the benefits they have open data in Ecuador, which indicate that fields, sectors, industries, institutions or organizations are most require open data, along with a proposed methodology for measuring results . Ultimately it is intended to encourage the competent authorities related to the SNAP about enormous changes entities that this new paradigm entails for citizens, and to publicize the new challenges that this new wave is beginning to have in the countries of the region, which breaks with the trajectory so far carried the evolution of the so called e-government or e-government extending its reach beyond the mere simplification of procedures and greater transparency of information by governments in the management

Keywords

- Semantic Web
- Linked Open Data
- Methodology Semantic Web
- RDF
- SPARQL

1. CAPÍTULO 1

Introducción

1.1 INTRODUCCIÓN

En la actualidad, gobiernos como el Español, Americano e Inglés llevan desde el 2007, trabajando con los Datos Abiertos, han encontrado en este tema un recurso para generar nuevas tecnologías autosustentables, aportando a la innovación y desarrollo de los gobiernos locales y nacionales, la mayoría de sus iniciativas vienen por parte del sector público, trabajando conjuntamente con la empresa privada, donde el mayor beneficiado es el ciudadano.

En nuestro país, las instituciones que conforman el Estado ecuatoriano encargadas de la Administración Pública, Central y Dependiente, del Gobierno Central del Ecuador, procesan grandes cantidades de información referentes a diferentes sectores, la cual no está siendo publicada y re-utilizada debido a que el talento humano de los departamentos de sistemas de las diferentes instituciones públicas, no tienen conocimiento de cómo hacerlo además de que no saben sobre los Datos Abiertos. Esto ha causado un desaprovechamiento respecto al valor agregado que se puede generar, en el sector público y privado, en el mejoramiento de los servicios públicos que se brindan a los ciudadanos, a través de un ecosistema de emprendimiento y participación ciudadana. Una de las iniciativas claves del Gobierno hasta el momento, ha sido la creación del “Plan de Gobierno Electrónico” el cual en su parte medular contempla el uso de los Datos Abiertos para el mejoramiento de los servicios, principalmente basado en una tendencia que se ha generado a nivel mundial referente a este tema.

La publicación de Datos Abiertos, representa el eje principal que puede dinamizar este ecosistema de mejoramiento de servicios, y va ligado a temas de transparencia e innovación, con la colaboración de la ciudadanía, para agilizar el nivel de rapidez de creación y ejecución de ideas innovadoras. La Secretaría Nacional de la Administración Pública, en su afán de consolidar una gestión pública

de excelencia, cumple un papel fundamental en este aspecto, concentrando sus esfuerzos y trabajo cotidiano para coordinar, ejecutar e impulsar las acciones necesarias para mejorar los servicios públicos que se brindan a los ciudadanos. Sin embargo, carece de un estudio que le permita identificar qué información es prioritaria para poder cumplir con este propósito, orientado a las necesidades actuales de la ciudadanía. Además, carecen de una guía que permita orientar a las instituciones públicas, respecto al procedimiento adecuado para una correcta liberación de información y para el aprovechamiento de la misma. Los Datos Abiertos, pueden ser escalables a muchos niveles de apertura según se desee, pudiendo incluso ser útiles en ambientes complejos de tecnologías emergentes como la interconexión de objetos de uso cotidiano a la internet, web de las cosas o inteligencia artificial.

Para que las instituciones puedan conocer a que nivel de apertura de datos pueden llegar, es necesario que tengan una guía que les permita proporcionar la orientación técnica, legal y práctica dentro del proceso de apertura de datos.

Todos los datos potenciales que pueden ser liberados, generalmente son solo almacenados, lo cual generan un costo significativo de mantenimiento, almacenamiento e infraestructura. El retorno de este costo, corresponde a la iniciativa de poder identificar toda la información de valor, para la generación de nuevas ideas innovadoras, que incluso en un futuro puedan llegar a crear nuevas economías a través de la innovación de aplicaciones o servicios que se hayan pasado por alto, logrando de esta manera que el manejo de la información pueda ser auto sustentable.

1.2 JUSTIFICACIÓN E IMPORTANCIA

En el Ecuador hasta la actualidad no se han realizado Estudios sobre los Datos Abiertos por lo que considero se debe hacerlo ya que, esta iniciativa de Datos Abiertos se está haciendo de forma progresiva en todo el mundo. Comenzó con el proyecto data.gov en EEUU, el sitio donde se encuentra el catálogo de datos abiertos del Gobierno Federal, y paulatinamente se ha ido extendiendo a otros países. Cada vez son más los gobiernos que anuncian, comienzan y aplican iniciativas de transparencia y gobierno abierto y que publican sus catálogos de

datos. Actualmente existen proyectos de Datos Abiertos desarrollados en EEUU, además en países como Canadá, Reino Unido, Australia, Nueva Zelanda, Suecia o Noruega y también en España.

En el Ecuador los datos abiertos es un tema que todavía no se ha implementado, si se compara con otros gobiernos como el de Estados Unidos, ellos nos llevan un adelanto de más de 5 años, a nuestro país le falta por implementar normas, políticas, estándares y guías que permitan llevar a cabo de manera correcta la liberación de los datos abiertos, por lo cual para realizar toda esta investigación, hay que tener en claro los siguientes conceptos:

Datos Abiertos (Open Data): Open Data o Datos Abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona o entidad pública, privada, y que se encuentran vinculados, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.

Lenguaje de Ontologías Web (Web Ontology Language (OWL)): Está diseñado para el uso en aplicaciones que requieren procesar el contenido de la información en lugar de solo representar la información para los humanos. OWL facilita un mejor mecanismo para interpretar el contenido Web que los mecanismos admitidos por XML, RDF, proporcionando vocabulario adicional junto con una semántica formal.

RDF: Es toda una familia de especificaciones de la World Wide Web Consortium, originalmente diseñado como un modelo para metadatos, en la actualidad es usado como un método a implementar para la descripción conceptual o modelado de la información

XML: Especificación para diseñar lenguajes de marcado, el cual proporciona un formato para la descripción de datos estructurados, facilitando las declaraciones de contenido haciéndolas más precisas, obteniendo como resultado búsquedas más significativos en varias plataformas.

Conjunto de Datos (Datasets): conjunto de datos que corresponde a los contenidos de una única tabla de base de datos, o una única matriz de datos estadística.

Catálogos de Datos: son aquellos que hacen accesible y dan referencias exactas a los conjuntos de datos.

URI (Uniform Resource Identifiers): Son una serie de identificadores globales que hacen alusión a recursos que están en la web como documentos, imágenes, archivos descargables, servicios, buzones de correo electrónico u otros, que pueden ser creados y administrados en forma distribuida.

HTTP (Hyper Text Transfer Protocol): Protocolo para intercambiar datos en la web, cuya funcionalidad básica es la transferencia de ficheros hipertexto de un ordenador a otro

HTML (Hyper Text Markup Language): Es un lenguaje diseñado para estructurar textos y presentarlos en forma de hipertexto, el cual es un formato estándar de las páginas web. También representa información orientada a elementos visuales para presentarla visualmente a humanos.

1.3 PLANTEAMIENTO DEL PROBLEMA

Organizaciones como CTIC, Open Knowledge Foundation y World Wide Web Foundation, coinciden en sus estudios en nombran a países como Reino Unido, España y Estados Unidos como los más avanzados en el tema de desarrollar los Datos Abiertos en sus gobiernos, publicando datos que generan impacto.

En el Ecuador, actualmente no se ha podido identificar portales que estén brindando datos abiertos a excepción del portal del municipio de Quito y Fundapi, siendo estos esfuerzos individuales que ayudan a fomentar una cultura de Datos Abiertos, pero al no estar tan vinculados con el gobierno, la comunidad sabe muy poco de ellos, ahora bien el Gobierno a través de diferentes Instituciones ya está dando los primeros pasos en cuanto a liberación de información a través de los siguientes documentos:

- GUÍA DE POLÍTICA PÚBLICA DE DATOS ABIERTOS - GPP-DA-v01-2014

- LOTAIP Artículo 1 Principio de Publicidad de la Información Pública
- Resolución Defensoría del Pueblo Artículo 11, Artículo 12, Artículo 13

Como es de conocimiento general, la gran mayoría de instituciones públicas del Ecuador, basadas en la Ley de Transparencia, han publicado cierta información, a través de sus portales web, pero todos estos esfuerzos no son de gran aporte para los Datos Abiertos, ya que lo hacen en formatos que no son libres, un buen ejemplo es la página de Banco Central del Ecuador (<http://www.bce.fin.ec/index.php/2013-11-18-23-50-36/historico-transparencia>), que tiene publicada una gran cantidad de información e incluso deja realizar consultas de datos liberados años atrás, pero el problema es que se encuentran en formatos no compatibles con los estándares para ser libres, tal como lo propone el estudio.

1.4 FORMULACIÓN DEL PROBLEMA

A continuación se formula una serie de preguntas las cuales serán resueltas a lo largo de la presente investigación:

- ¿En el Estado ecuatoriano existe alguna Normativa para publicar los datos liberados de manera estándar?
- ¿Se quiere lograr el aprovechamiento y redistribución de la información pública?
- ¿Qué nivel de madurez tienen los Datos Abiertos en el Ecuador, para poder provocar algún tipo de beneficio e impacto para la sociedad civil y la empresa pública-privada?
- ¿Se logrará facilitar el libre acceso y disposición de toda la información para fomentar el mercado de la explotación de la información?
- ¿Se podrá desarrollar una cultura de generación y manejo de información en la sociedad civil y las empresas públicas y privadas del Ecuador?

1.5 HIPÓTESIS

Los Datos Abiertos en las instituciones públicas del Ecuador, fomentarán la innovación en servicios públicos dirigidos a la ciudadanía y la conexión de estos datos generará un valor agregado sobre toda la información liberada.

1.6 OBJETIVO GENERAL

Determinar si los Datos Abiertos en las instituciones públicas del Ecuador, fomentarán la innovación en servicios públicos dirigidos a la ciudadanía y si la conexión de estos datos generara un valor agregado sobre toda la información liberada.

1.7 OBJETIVOS ESPECÍFICOS.

- Investigar las normativas de publicación de información y el estado del arte de los Datos Abiertos a nivel mundial y en Ecuador
- Investigar cómo lograr el aprovechamiento y redistribución de la información pública basándonos en los Gobiernos de otros países, tomando en cuenta que datos han liberado y que aplicaciones han desarrollado.
- Definir el nivel de madurez de los Datos Abiertos en el Ecuador
- Definir una arquitectura estandarizada la cual permita el libre acceso a la información, con soporte para la implantación de los Datos Abiertos Enlazados en el contexto de la administración pública.
- Definir los lineamientos para la publicación de Datos Abiertos Enlazados, enfocado a las instituciones correspondientes a la Administración Pública, Central, Institucional y Dependiente, para generar una cultura de manejo de información.

2. CAPÍTULO 2

ESTADO DEL ARTE

2.1 Terminología

A continuación se presentan las definiciones necesarias para que nos guiarán en la presente investigación

2.1.1 Datos Abiertos

La definición de Datos abiertos, o en inglés Open Data, son todos aquellos datos que pueden ser utilizados, reutilizables y redistribuidos sin ninguna restricción por cualquier empresa o persona, siempre y cuando estén sujetos a los responsables de la liberación de información, además de compartirse de la misma manera en que aparecen.

Características más importantes:

- **Disponibilidad y Acceso:** Los datos siempre deben estar disponibles en su totalidad, preferible que se pueda descargar del internet. Además, la información debe estar disponible en una forma conveniente y modificable.
- **Reutilización y redistribución:** los datos deben ser aprovisionados bajo métodos que permitan generar una ganancia y redistribuirlos, e inclusive integrarlos con otros conjuntos de datos.
- **Participación Universal:** La información se debe utilizar, reutilizar y redistribuir. No debe haber exclusión alguna en términos de esfuerzo, personas o grupos. Restricciones “no comerciales” que prevendrían el uso comercial de los datos; o restricciones.

2.1.2 Las 5 estrellas de los datos abiertos

Sobre el mismo documento del creador de “World Wide Web“ Berners-Lee “Los datos abiertos enlazados”, desde el año 2010 se han venido agregando varios criterios, donde se han llegado a conformar cinco niveles sobre los Datos Abiertos enlazados, mismos que son presentados a continuación:

Una estrella: Se deben ofrecer datos e información en cualquier formato libre, el cual se pueda manipular sin importar su complejidad de manipulación, como un PDF o una imagen escaneada

Dos estrellas: Para cumplir con 2 estrellas, primero hay que cumplir con la primera, adicional, se debe presentar información estructura, en algún formato que pueda ser leído por maquinas, ejemplo documentos .xls, donde se aprecien tablas, en lugar de imágenes.

Tres estrellas: Para cumplir con las 3 estrellas, se debe cumplir con las 2 estrellas, pero además toda la información entregada será en formatos que no sean propietarios como .csv en vez de Excel.

Cuatro estrellas: Además de cumplir con las 3 estrellas anteriores, todos los datos presentados deben usar estándares abiertos como W3C, RDF para identificarlos y SPARQL para consultarlos, de esta manera quedan habilitados para que puedan ser usado por otros.

Cinco estrellas: Se debe cumplir con las 4 estrellas anteriores, además de enlazar los datos y conjunto de datos con otras fuentes, permitiendo de esta manera darles contexto a nuestros datos, además permitan generar información de manera estructurada.

Como consecuencia, a lo anterior mencionado aparecen los datos abiertos enlazados, ya que adicional a los principios ya establecidos, se considerará la publicación de los mismos, para dar libre referencia, consumo y uso de los mismos.

2.1.3 Datos Abiertos Enlazados

La definición de datos enlazados, o en inglés Link data, aparecieron a partir de la idea de dar crecimientos a los conjuntos de datos descritos en RDF, mediante conexiones con otras agrupaciones de datos descritos en RDF. Esta definición está basada en cuatro fundamentos descritos en 2006 por Berners-Lee, a continuación los 4 principios:

1. Utilización de URI para la identificación de cada recurso publicado en el internet (Web)
2. Los recursos publicados en URI basada en HTTP con el fin de que puedan ser consultados con rapidez
3. Publicar información útil, extra y detallada acerca del recurso que se accede URI basada en HTTP
4. Integrar las conexiones con otras URI relacionadas con los datos contenidos en el recurso, de forma que se logre potenciar el descubrimientos de la información en la internet (Web)

URI (Uniform Resource Identifiers): Son una serie de identificadores globales que hacen alusión a recursos que están en la web como documentos, imágenes, archivos descargables, servicios, buzones de correo electrónico u otros, que pueden ser creados y administrados en forma distribuida.

HTTP (Hyper Text Transfer Protocol): Protocolo para intercambiar datos en la web, cuya funcionalidad básica es la transferencia de ficheros hipertexto de un ordenador a otro

HTML (Hyper Text Markup Language): Es un lenguaje diseñado para estructurar textos y presentarlos en forma de hipertexto, el cual es un formato

estándar de las páginas web. También representa información orientada a elementos visuales para presentarla visualmente a humanos.

XML: Especificación para diseñar lenguajes de marcado, el cual proporciona un formato para la descripción de datos estructurados, facilitando las declaraciones de contenido haciéndolas más precisas, obteniendo como resultado búsquedas más significativos en varias plataformas.

2.1.4 RDF

Es toda una familia de especificaciones de la World Wide Web Consortium, originalmente diseñado como un modelo para metadatos; en la actualidad es usado como un método a implementar para la descripción conceptual o modelado de la información que se implementa en la web, bajo la utilización de una variedad de notaciones, sintaxis y formatos de serialización de datos.

RDF está basado en la forma básica de una oración, compuesta de sujeto, predicado y objeto, que pueden formar tripletas (s, p, o) para ser entendidas como fórmulas lógicas binarias del tipo $p(s, o)$.

Un conjunto de tripletas puede ser interpretado como una red semántica, es decir, como un grafo dirigido con nodos y arcos rotulados, donde para cada tripleta hay un arco rotulado con el predicado y los nodos inicial y final son rotulados con el sujeto y el objeto. RDF además de describir una estructura de grafos, también define los conceptos de clase e instancia. Es semejante al enfoque que poseen los modelos clásicos como entidad-relación o diagramas de clases, ya que está basado en realizar declaraciones sobre los recursos que en su mayoría son web

2.1.5 Grafos RDF

Los grafos RDF, más conocido como grafo de datos enlazados, corresponden a un conjunto de datos o recursos enlazados entre sí a través de URI, mediante una proposición que se conoce también como tripletas en RDF. Como cada triplete se define a partir nodos descritos por una URIs en forma de sujeto, predicado, objeto. El grafo RDF estará distribuido a lo largo de cada una de las URIs definidas en las tripletas RDF.

2.1.6 SPARQL

Es un lenguaje estandarizado de consultas para grafos RDF y actualmente es sugerido por Data Access Working Group (DAWG), Mediante su elocuencia permite realizar múltiples consultas a diferentes tipos de fuentes de datos o grafos, que se encuentran en lenguaje RDF. Toda la estructura en si es similar al lenguaje de base de datos SQL a pesar de que este orientado a tripletas y grafos RDF. Todas las consultas SPARQL resultan ser conjuntos de tripletas RDF, grafos RDF, URIs a entidades o simplemente valores.

2.1.7 URI

Es una serie de caracteres que identifican los recursos de una red en forma unívoca. Se diferencia de una URL (localizador de recursos uniforme), en que estos últimos hacen referencia a recursos que con el tiempo van cambiando.

Estos recursos pueden ser accesibles desde una red o sistema. Los URI pueden ser localizador de recursos uniforme (URL), Uniform Resource Name (URN), o ambos.

Un URI consta de las siguientes partes:

- **Esquema:** Hace referencia a una especificación para asignar los identificadores, En varias casos identifica el protocolo de acceso al recurso, como por ejemplo http, ftp. El esquema identifica el protocolo que va a usar. Hoy en día hay

variedad de esquemas registrados para diferentes funciones como por ejemplo ftp para transferencia de archivos.

- **Autoridad:** elemento jerárquico que identifica la autoridad de nombres
- **Ruta:** Información usualmente organizada en forma jerárquica, que identifica al recurso en el ámbito del esquema URI y la autoridad de nombres
- **Consulta:** Información con estructura no jerárquica, que identifica al recurso en el ámbito del esquema URI y la autoridad de nombres.
- **Fragmento:** Permite identificar una parte del recurso principal,

Por lo general se acostumbra llamar URL a todas las direcciones web, URI es mucho más que eso, es un identificador más completo y por se recomienda su uso en lugar de la expresión URL.

2.1.8 Análisis y negociación del contenido

La definición de negociación de contenido está relacionada con la información vinculada con la publicación de los recursos sobre las URI. Se sabe que los recursos se definen en las URI, sin embargo para un determinado recurso hay varias representaciones. Todo tiene sentido, tomando en cuenta que RDF permite varias sintaxis, tal como se ha visto anteriormente. Entonces se entiende que, un mismo recurso será descrito de forma similar por cualquier tipo de sintaxis perteneciente a RDF. Así de esta forma, se asigna una URI a un determinado recurso, el mismo será independiente de su representación.

Entonces la negociación de contenido, se ve relacionada en el contexto de los Datos Abiertos enlazados en el instante de ingresar a un recurso en una URI específica, cuando no se identifica la representación a la cual hay que acceder de forma concreta mediante la URI. Con esto se quiere decir que si se accede a la URI que define el recurso, y no se especifica la representación del recurso que se desea ver, iniciara el proceso de negociación de contenido.

En el mundo real. la negociación de contenidos se aplica desde la utilización del protocolo HTTP, cuando el usuario accede a un recurso publicado, este envía cabeceras indicando los tipos de documento que son los más seleccionados, como

respuesta de un conjunto de metadatos adicional, lo cual permitirá entregar toda la información relacionada de la forma más concisa posible. Ejemplo, un cliente envía los formatos aceptados de retorno (codificación de caracteres, idioma). La información es enviada mediante las cabeceras HTTP, una vez que accedió a la red LAN de la URI del recurso, llega la petición al servidor donde dicha solicitud indica el tipo de información preferida, automáticamente envía respuesta al usuario utilizando un código 303 HTTP que sabe ser denominado “See Other” o “Ver Otro” mostrando la URI, donde el usuario deberá acceder para visualizar el mejor recurso, de acuerdo a sus preferencias. De esta forma el usuario de nuevo accederá al recurso, ahora es cuando la URI especificada por el mensaje “Ver otro”, en donde obtendrá el recurso más apropiada, seguidamente se genera código en el servidor, identificado por un mensaje de respuesta de código 200 denominado OK.

2.1.9 Representación de la información

A continuación se hará referencia a ciertas representaciones y definiciones más utilizadas que un recurso informático puede tener en el contexto de los Datos Abiertos enlazados.

- **CSV (Comma-Separated Values):** información presentada en este formato, entrega datos en una hoja de cálculos.
- **Notation 3:** la información entregada en formato Notation 3, entrega tripletas RDF en un formato comprensible para los usuarios y simultáneamente capaces de ser procesados por analizadores RDF.
- **JSON (JavaScript Object Notation):** la información basada en este formato entrega los datos con sintaxis JavaScript
- **Notation Triples:** La información representada en este tipo de formato, entrega datos similares al formato Notation 3, con la diferencia que este formato no realiza análisis sintáctica del RDF.
- **RDF/XML:** La información en este formato entrega datos con estructura RDF mediante marcas, que pueden ser analizadas por cualquier procesador XML.

- **HTML (Hypertext Markup Language):** la información basada en este formato es procesable por cualquier navegador Web y legible para cualquier cliente, sin embargo la información entregada no es semánticamente útil para RDF.
- **HTML + RDF:** Es una representación que proporciona datos estructurados, los cuales pueden ser atributos en algún lenguaje de marcado. La información representada en este formato es capaz de ser procesada tanto para un navegador web como por aplicaciones que procesan código RDF, gracias a que la información semántica está incluida como metadatos de la página Web.

Tabla 1

Formatos representación de información o recurso

Formato	Contenido	Extensión
CSV	text/csv	.csv
Notation 3	text/rdf+n3	.n3
JSON	application/rdf+json	.json
Notation Triples	text/plain	.ntriples
HTML	text/html	.html
HTML + RDFa	application/xhtml+xml	.html

2.2 Propuesta para la publicación de datos abiertos enlazados

2.2.1 Como publicar datos enlazados en la web

En la siguiente propuesta, se presentará el proceso de publicación de Datos Abiertos Enlazados, a continuación se verán 7 etapas:

1. Definición de vocabulario
2. División del grafo RDF en páginas de datos
3. Asignación de las páginas de datos, para cada URI
4. Creación de variables para cada página en HTML
5. Asignación de las URI para cada entidad

6. Agregar metadatos en cada página
7. Agregar un mapa del sitio semántico

Definición de vocabulario

Es necesario darle importancia y énfasis a la reutilización de vocabularios existentes, con el propósito de estandarizar los datos, si no es posible la reutilización, se deberá construir su propio vocabulario, otro punto importante es la utilización de vocabularios que se han utilizado ampliamente:

- Simple Knowledge Organization System (SKOS) 6 o (SIOC) 7
- Dublin core (DC)
- Friend of a friend (FOAF)

Posteriormente se brindarán algunas guías para poder realizar una selección acertada de los vocabularios, de esta manera se podrá crear vocabularios auténticos y en una misma tendencia para así expandir los vocabularios ya utilizados.

División del grafo RDF en páginas de datos

En esta fase se pondrá en línea el grafo mediante documentos RDF. Si se tiene un grafo muy grande, se deberá dividir el grafo en varios documentos siguiendo los lineamientos.

Ejemplo: dividir las páginas de datos por entidades. Otro lineamiento es que si ya se tienen páginas HTML, usar la misma granularidad para las páginas de datos.

Asignación de las páginas de datos, para cada URI

En la fase de asignación se definirá una URI para cada documento RDF, que ya se definió en la etapa anterior. Se hace referencia a eliminar los detalles de implementación relacionados al despliegue de los documentos RDF, se podrá utilizar soluciones como Cool URIs o patrones de datos enlazados.

Creación de variables para cada página en HTML

En esta fase se recomienda crear páginas HTML por la compatibilidad que se debe tener con los navegadores Web. En este punto se realiza la negociación de contenido, definiendo que el usuario es quien dicta la preferencia por los formatos mediante la cabecera “Accept” de HTTP. También es válido HTML y RDF dentro de una página a través de RDFa.

Asignación de las URI para cada entidad

Aquí se plantearán las reglas para la definición de URIs a las entidades dentro del grafo. También se ve la parte del diseño que va relacionado con la utilización de los siguientes caracteres especiales: “/” “#” en la construcción de URIs.

Agregar metadatos en cada página

En esta etapa se agregan tripletas con metadatos en las páginas principales del grafo RDF, ayudando de esta manera al usuario a entender las páginas de datos. Para esto se utiliza propiedades existentes en vocabularios utilizados como Dublin core, o Friend of a Friend, y el uso de propiedades genéricas como `rdf:label` o `rdf:type`.

Agregar un mapa del sitio semántico

En esta fase, se agregarán elementos adicionales a los Datos Abiertos Enlazados para mejorar su visibilidad. En una primera etapa se propone definir metadatos orientados a los rastreadores de buscadores web, definiendo un archivo para el mapa del sitio de Datos Abiertos Enlazados, donde se tendrán los siguientes elementos:

- Nombre del conjunto de datos,
- Espacio de nombres,
- Ubicaciones del Endpoint SPARQL,
- Ubicación de un volcado de datos en el caso de existir.

Hay que mencionar que esta investigación, también se presentan como una motivación, para la apertura a conjunto de aplicaciones que utilizan los datos abiertos enlazados que se dividen en tres categorías:

- Navegadores de datos enlazados (Linked Data Browsers),
- Visualizaciones de datos enlazados (Linked Data Mashups)
- Máquinas de búsqueda o buscadores (Search Engines).

2.3 Estructuras para la publicación de datos abiertos enlazados

La siguiente gráfica, representa una agrupación de posibles soluciones, para la publicación de datos abiertos enlazados.

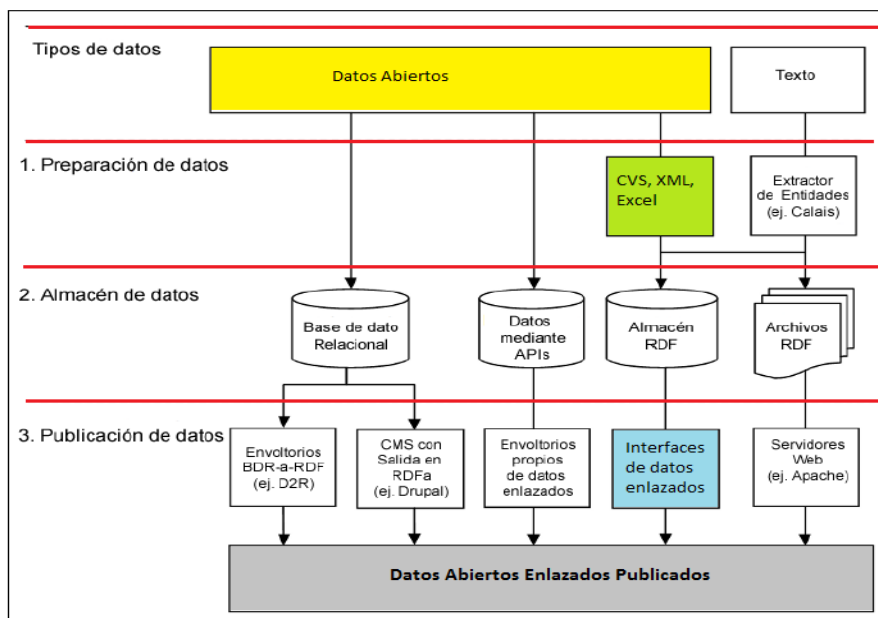


Figura 1 Propuesta publicación de Datos Abiertos Enlazados

2.3.1 Publicaciones basadas en bases de datos relacionadas.

Las bases de datos relacionadas pueden ser publicadas en forma de Datos Abiertos Enlazados, a través del uso de paquetes desde bases de datos relacionales hasta RDF. Dichas herramientas permiten a las personas que publican, definir mapeos desde estructuras en bases de datos relacionales hacia grafos RDF que estarán en el servidor de aplicaciones, regido a los principios de Datos Abiertos.

Algunas de estas herramientas, son: D2R Server, Virtuoso RDF Views Linked Data Wrapper 1 y Triplify. La desventaja de este tipo de soluciones es que se mantienen acopladas la base de datos corporativa a la base de Datos Abiertos Enlazados, por lo cual se agrega carga adicional de procesamiento al motor de datos relacional, producido por las nuevas consultas por el acceso a recursos en RDF.

2.3.2 Publicaciones basadas en APIs

La publicación de datos abiertos con estructura a través de APIs (Programmable Web, Flickr, Amazon). La implementación para una integración con RDF es más compleja ya que puede estar bajo muchos formatos (XML o JSON). Las soluciones de este tipo nos proporcionan varias ventajas como por ejemplo la utilización de fuentes externas de datos, pero estas a su vez generarán dependencias con estas fuentes externas sobre a las cuales no se las puede controlar.

2.3.3 Publicaciones basadas en datos estructuración.

La publicación de datos abiertos estructurados consiste en archivos CSV, hojas de cálculo Excel, archivos XML o dumps de bases de datos. Para que toda esta información pueda ser presentada como datos abiertos enlazados, deben pasar por un proceso para convertirse en RDF y seguidamente ser guardados en alguna base RDF. Varias herramientas se las puede encontrar en la pag. 17

2.3.4 Publicaciones Basadas En Archivos RDF

Lo más sencillo para la publicación en RDF es subir archivos a un servidor web y ponerlo disponible en una URI, pero esto es aplicable en estos casos

- Cuando un usuario crea y mantiene archivos muy pequeños, como por ejemplo la publicación de perfiles FOAF (Friend Of A Friend, literalmente "Amigo de un Amigo").

- Cuando una herramienta de software o algún proceso genera o exporta datos como RDF en archivos estáticos.

Normalmente en el primer caso la sintaxis N3 es la más utilizada por su legibilidad, pero cualquiera otra sintaxis puede ser utilizada.

2.3.5 Publicaciones basadas en documentos de texto

Si necesita archivos de texto en formato RDF como por ejemplo boletines de noticias o reportes, para obtener los datos estructurados se pueden consumir servicios de la web (Calais, Ontos o DBPedia Spotlight). El publicar documentación enlazada semánticamente, es una excelente ventaja, ya que nos permite incrementar la capacidad de descubrir los documentos, mejorando la recuperación de la información que aplican las aplicaciones web como los navegadores.

2.3.6 Publicaciones basadas en almacenes RDF

Para la implementación de tripletas RDF se debe crear un almacén para las mismas, consultando y sirviendo directamente. Este tipo de método de publicación presenta varias ventajas como rendimiento, escalabilidad y seguridad, hay que tomar en cuenta que para este tipo de publicaciones es necesario la instalación de una infraestructura adicional basada en almacén RDF. En cuanto a la publicación lo más sencillo es la utilización de herramientas (“Linked Data Frontends”, Pubby, WESO-DESH, Elda).

2.4 Herramientas

A continuación se presentan las herramientas útiles existentes en cada una de sus categorías, mencionando su definición, descripción, las mismas que prestan una gran ayuda en la implementación de infraestructura para datos abiertos enlazados.

2.4.1 Almacenamiento de datos

Herramientas dedicadas a sistemas para almacenamiento y gestión de información y recursos en RDF, comúnmente mencionada RDF Store o Triple Store. A continuación se presentan las herramientas más útiles relacionadas con el almacenamiento.

Openlink Virtuoso Universal Server

Es un tipo de servidor corporativo para bases de datos, plataforma de servicios e integrando funcionalidades de gestión de bases de datos RDF, XML, relacionales, de archivos, y texto; servidor de aplicaciones Web, proporcionando una gran cantidad de usos en trabajos corporativos. Existen dos versiones, una licencia privada para empresas y una versión de licencia abierto de libre comercio. Es multiplataforma y permite escalabilidad basado en clúster.

D2R Server

Herramienta para la publicación de datos abiertos enlazados, los cuales se pueden publicar directamente de la base de datos. Permite RDF-izar datos desde la base de datos, navegando por los mismos a través de páginas HTML de vistas en RDF, generadas por mapeos a la base de datos. Posee un Endpoint SPARQL el cual puede realizar consultas directamente sobre el RDF generado. Se construyó en Java 1.4, es de licencia abierta.

Sesame

Es un framework y almacén RDF que se utiliza con mucha regularidad en empresas privadas como por administraciones públicas. Sus principales características son que permite varios tipos de almacenamiento (como en memoria, archivos o en base de datos), además de múltiples lenguajes de consulta, razonadores

y protocolos cliente / servidor, es una de las soluciones más estables y flexibles. Han desarrollado varias extensiones para este almacén RDF, como módulos de consulta SQL, SPARQL, administración, adaptadores de bases de datos, exportación. Etc.

Redland

Conjunto de bibliotecas desarrolladas en lenguaje C de que son de soporte a RDF. Permite almacenamiento tanto en memoria como persistente, utilizando diferentes bases de datos como Oracle, MySQL, PostgreSQL, Openlink Virtuoso y SQLite. También las bibliotecas soportan el tratamiento de RDF con múltiples de sus sintaxis, permitir consultas en SPARQL y RDQL, integración nativa de las bibliotecas con Perl, PHP, Python y Ruby, es de distribución libre bajo licencia LGPL 2.1, GPL 2 y Apache

Bigdata

Base de datos de propósito general orientado con una escalabilidad horizontal, con la ventaja que se puede desplegar hasta en cientos de servidores. La base de datos RDF soporta razonamiento sobre reglas de RDF Schema y OWL Lite, consultas SPARQL e indexación mediante la utilización de Apache. Es de tipo GPL para uso no comercial, para uso comercial se debe pagar una licencia.

2.4.2 Generación de grafos RDF sobre HTTP

Las herramientas que se van a mencionar permiten desplegar una URI sobre HTTP un documento de texto en determinadas sintaxis RDF donde se describen datos abiertos enlazados. Están presentes las herramientas más utilizadas para la generación de grafos RDF sobre HTTP, también conocidas como Linked Data Frontend.

Pubby

Generador de grafos RDF sobre HTTP, es muy utilizado sobre un Endpoint SPARQL. Permite acceder tanto a servidores SPARQL locales y remotos, utiliza reescritura de URL, generando una salida en HTML, además realiza negociación de contenido utilizando 303 See Other y es compatible con contenedores de servlets, Apache Tomcat y Jetty. DBPedia es un buen ejemplo se construyó en base a Pubby. Se basa en expresiones regulares para definir los patrones de la URI. Una de sus desventajas es que solo permite utilizar consultas SPARQL, por lo cual solo es posible describir recursos sin realizar consultas paramétricas. Es de libre distribución bajo la licencia Apache.

Elda

Es una herramienta que genera grafos RDF sobre HTTP la cual utiliza URLs Restful para poder realizar consultas SPARQL. Una de sus desventajas es que no permite la definición de expresiones regulares para la descripción de las URIs. Fue elaborado en lenguaje Java, permitiendo que se pueda levantar en un servidor de aplicaciones Web en Apache Tomcat o como también de forma inmediata usando Jetty

Virtuoso – Conductor UI for URL Rewriter

Es un componente de Openlink Virtuoso denominado conductor. Su principal objetivo es publicar de forma automáticamente grafos RDF sobre HTTP, datos que son extraídos desde Virtuoso al ingresar a determinados patrones de URI. Es muy similar a Pubby y Elda, pero integrado a Virtuoso, lo que es una desventaja, posee una interfaz gráfica de usuario que facilita la gestión de datos abiertos enlazados..

2.4.3 Visualizadores de datos enlazados

Como su nombre lo indica permiten visualizar de manera gráfica, facetada o tabulados datos en RDF de tal manera que se haga legible por un usuario final, debido a esto se han creado una serie de métodos. A continuación se verá dos tipos de herramientas, primero herramientas para navegadores RDF y finalmente herramientas para su visualización.

2.4.3.1 Navegadores

A estas herramientas también se las conoce como “Linked Data Browsers”, permiten visualizar datos abiertos enlazados en RDF a través de páginas explicativas de recursos RDF, lo cual nos muestra legible a los recursos. La siguiente tabla se presenta varios navegadores de datos abiertos enlazados que se encuentran hábiles en la actualidad

Tabla 2

Herramientas de navegación para Datos Abiertos Enlazados

Nombre	Autor	Descripción
SemLens	DEI Interactive Systems	Permite analizar tendencias y correlaciones entre datos RDF utilizando un visualizador construido sobre Adobe Flex 3. → Continuacion
tFacet	DEI Interactive Systems	Permite visualizar datos RDF aplicando exploración facetada entre conceptos. Está construido sobre Adobe Flex 3.
Lodviz	Grupo WESO, Universidad de Oviedo	Permite visualizar datos de tipo RDF en forma de grafos. Para ello solo basta con referenciar una URL y entregará una representación gráfica utilizando HTML5 y Javascript.
RelFinder	DEI Interactive Systems	Permite visualizar las relaciones existentes entre diversos datos abiertos enlazados en formato RDF utilizando un visualizador construido sobre Adobe Flex 3.

2.4.3.2 Visualizadores

Permite visualizar datos abiertos enlazados en RDF mediante diferentes tipos de visualizaciones gráficas. La siguiente tabla nos presenta las herramientas más utilizadas de esta categoría.

Tabla 3

Herramientas de visualización para Datos Abiertos Enlazados

Nombre	Autor	Descripción
SemLens	DEI Interactive Systems	Permite analizar tendencias y correlaciones entre datos RDF utilizando un visualizador construido sobre Adobe Flex 3.
tFacet	DEI Interactive Systems	Permite visualizar datos RDF aplicando exploración facetada entre conceptos. Está construido sobre Adobe Flex 3. → Continuación
Lodviz	Grupo WESO, Universidad de Oviedo	Permite visualizar datos de tipo RDF en forma de grafos. Para ello solo basta con referenciar una URL y entregará una representación gráfica utilizando HTML5 y Javascript.
RelFinder	DEI Interactive Systems	Permite visualizar las relaciones existentes entre diversos datos abiertos enlazados en formato RDF utilizando un visualizador construido sobre Adobe Flex 3.

2.5 Casos de Éxito

A continuación se verán varios casos de éxito, de relevancia en el contexto de publicación de datos abiertos, datos abiertos enlazados y gobierno abierto.

2.5.1 data.gov.uk

El Reino Unido ha puesto a disposición de los ciudadanos, el portal data.gov.uk en la cual dan a conocer por medio de datos abiertos enlazados, toda la información pública de interés ciudadano. En la actualidad, se han publicado ya más de 5.400 conjuntos de datos disponibles con información acerca de todas las entidades públicas de gobierno y muchos otros entes concernientes a autoridades locales.

Toda esta iniciativa cuenta con la infraestructura de soporte a datos abiertos enlazados, dentro de ello: documentación desarrollada pensada para el consumo de los datos, un Endpoint SPARQL para consultas, un buscador semántico, y otras aplicaciones de visualización de datos enlazados específicas para su dominio.

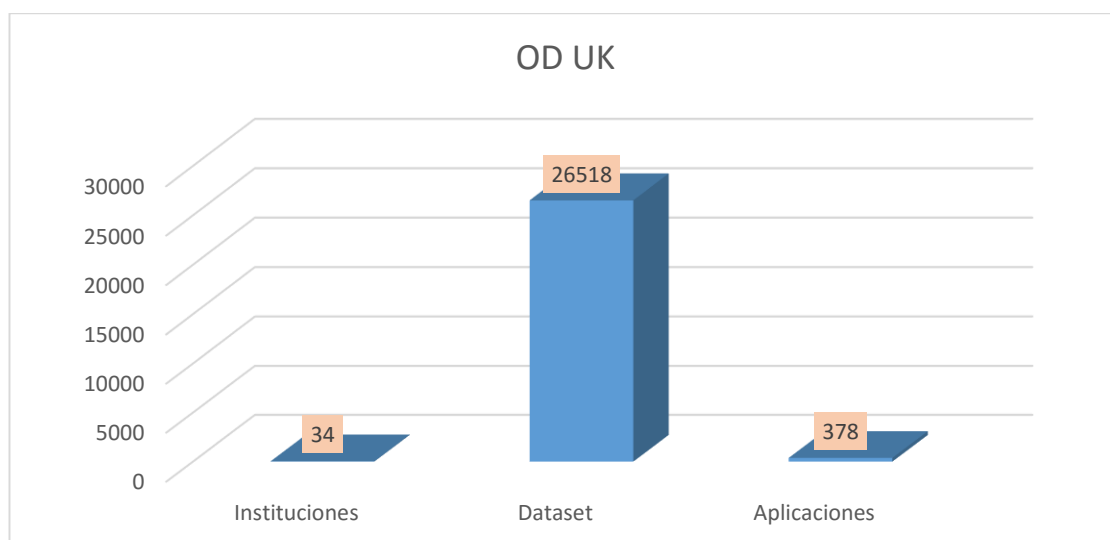


Figura 2 Estadística de Datos Abiertos en Reino Unido

Posee un total de 378 aplicaciones desarrolladas, para sectores públicos, privados y mixtos, donde 10 de esas aplicaciones son link open data:

Tabla 4

Aplicaciones Publicadas por Renio Unido

Aplicaciones			
Nombre	Descripción	Calificación	Descargas
Huxley	Comido abierto, para consumir API de manera fácil	5	
Route Risks	Analiza la seguridad de las carreteras	4,5	
Mersey Tunnels Fast Tag	Administración de Cuenta Bancaria	4,5	50-100
Flood Risk Finder	Busca riesgos de inundación para diferentes poblados	4,4	
Bona Vacantia	Muestra propiedades no reclamadas para la compra	4,7	
London Now	Clima del sector donde estas y el estado del transporte	4	
London Traffic	Proporciona imagen de 700 para ver tráfico de la ciudad	3,7	
Doorda	Permite utilizar mapas para la creación de páginas web	3,5	

2.5.2 data.gov

El gobierno de los Estados Unidos ha puesto a disposición de los ciudadanos el portal data.gov, en donde se publican alrededor de unos 6.400 millones de tripletas RDF sobre gobierno abierto. Para poder lograr esto, han tenido que definir índices de documentos RDF, los cuales están disponibles para ser consultados y consumidos, han implementado toda una infraestructura de soporte para los Datos Abiertos Enlazados dentro de la cual se incluyen visualizaciones de datos como elementos centrales para el uso de la población.

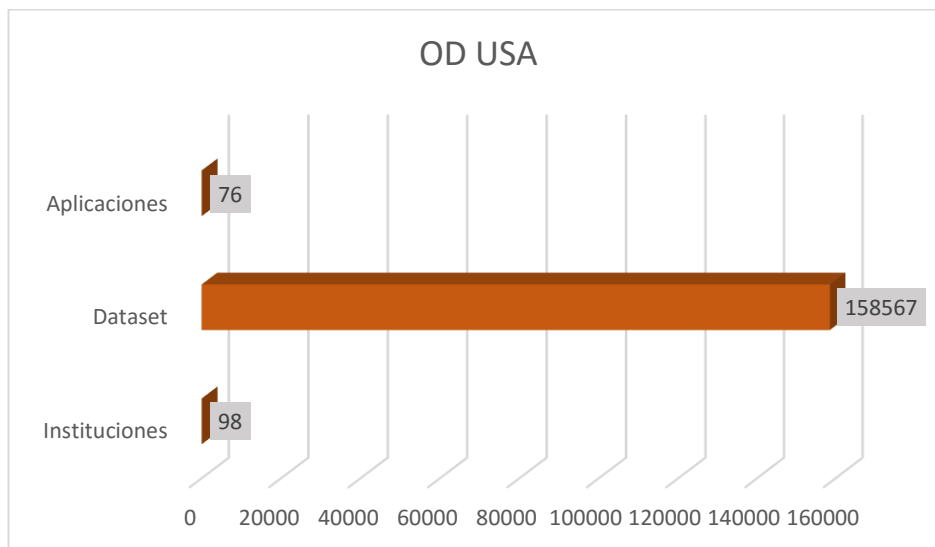


Figura 3 Estadísticas Datos Abiertos en EEUU

Tabla 5

Estadísticas Datos Abiertos en EEUU

Aplicaciones			
Nombre	Descripción	Calificación	Descargas
The AIRNow	proporciona información sobre la calidad del aire en tiempo real que se puede utilizar para proteger su salud	5	100-500
Citymapper	Permite planificación viajes entre dos puntos	4,5	1 a 5 millones
Climate	ayuda a optimizar la toma diaria de decisiones de los agricultores, proporcionando up-to-the-minute,	3,9	10 a 50 mil
HD Scores	Muestra los restaurantes que cumplen con normas de salud	3,2	500 a 1000

→Continuación

Healthy Out	Busca lugares donde pueden encontrar comida que se ajuste a tu dieta	3,9	50 a 100 mil
NAEP	resultados estadísticos de la educación en el país	4,7	100 - 500

2.5.4 datos.gob.cl

Datos.gob.cl es una iniciativa del gobierno chileno, orientado para que los ciudadanos tengan acceso a la información, en base de la cual el gobierno toma decisiones para las políticas públicas y que también la puedan utilizar para sus propias investigaciones e intereses. Por ejemplo, construir aplicaciones y el manejo de análisis.

Manifiestan que el acceso a la información y la transparencia benefician a la democracia y la rendición de cuentas por parte de las autoridades. En este portal web se encontrarán conjuntos de datos públicos por el gobierno de manera fácil. En varios casos, los datos están publicados en varios formatos.

El portal web contiene un buscador y catálogos de datos con varias categorías para ayudar a la búsqueda de la información. También se puede encontrar información geo referenciada y archivos de imágenes. Algunos de estos datos ya están disponibles en diversos sitios de gobierno pero www.datos.gob.cl los reúne en un solo sitio web donde se pueden realizar búsquedas.

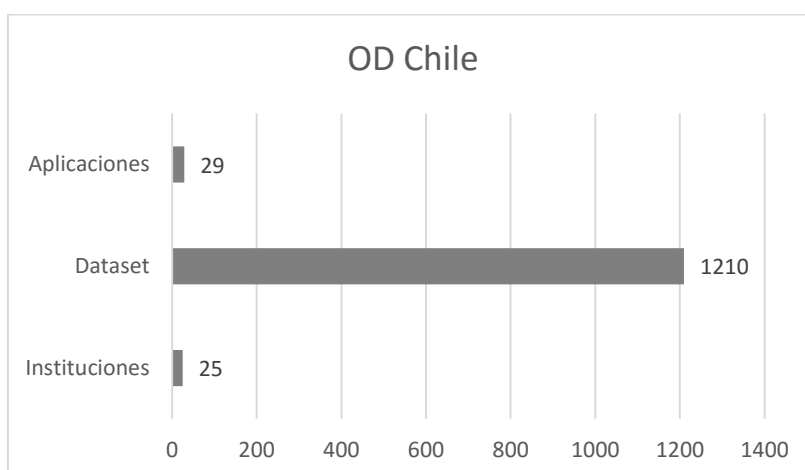


Figura 4 Estadística Datos Abiertos Chile

Tabla 6

Aplicaciones publicadas por el gobierno de Chile

Aplicaciones			
Nombre	Descripción	Calificación	Descargas
iTransantiago	aplicación que le permite encontrar fácilmente las mejores alternativas de desplazamiento en el sistema de transporte público	3,7	
Farmacias - Chile	Busca farmacias de turno en todo Chile	3,5	1000 a 5000 →Continuación
CorreosChile	Seguimiento de envío	1,7	1000 a 5000
Chilecompra	Revisa aquí las licitaciones más relevantes del portal compras públicas	1,3	1000 a 5000
Transparencia	Realice un seguimiento a los reclamos, presentados ante el Consejo para la Transparencia, por incumplimiento a la Ley	2,4	100 a 500
BauhausStgo	Permite encontrar museos y patrimonio cultural	4,2	100 a 500

2.5.4.1 datos.bcn.cl

En el gobierno Chile se encuentra la primera institución pública de América Latina que cumple con todas las normas y estudios para poder publicar Datos Abiertos Enlazados, avalado por organización internacional, Fundación CTIC e incluida en el diagrama oficial de Linking Open Data cloud. La aplicación es el portal web de la Biblioteca del Congreso Nacional de Chile.

En sus inicios el proyecto de Datos Abiertos Enlazados de la Biblioteca del Congreso Nacional de Chile, trabajó con el grupo de investigación WESO de la Universidad de Oviedo, obteniendo como resultado una implementación preliminar, un método de desarrollo y una arquitectura que puede ser replicada en otros espacios similares. Sobre el estado actual del proyecto, está en proceso de expandir los dominios de aplicación tanto en modelos (ontologías) como en datos publicados.

2.5.9 datos.gob.es

Este portal web inicia con el apoyo del Ministerio de Industria, Energía y Turismo como también con el apoyo del Ministerio de Hacienda y Administraciones Públicas del Gobierno de España. Su gestión directa corresponde a la Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (SETSI), adscrita al primero de estos departamentos.

Tiene como objetivo prioritario promover la publicación, mejorar el acceso y favorecer la reutilización de la información pública correspondiente al sector público. Tal esfuerzo pretende conseguir el diseño de una administración más transparente, eficiente y cercana, así como la creación de nuevas oportunidades –tanto sociales como económicas- a partir de la reutilización de la información del sector público.

Está abierto a profesionales, empresas, gestores institucionales y ciudadanos interesados en general. Se trata de los actores y beneficiarios principales de la reutilización de la información del sector público que hacen posible:

- El lanzamiento de nuevos modelos de negocio.
- La modernización de las administraciones.
- La creación de plataformas de transparencia, colaboración y participación ciudadanas.

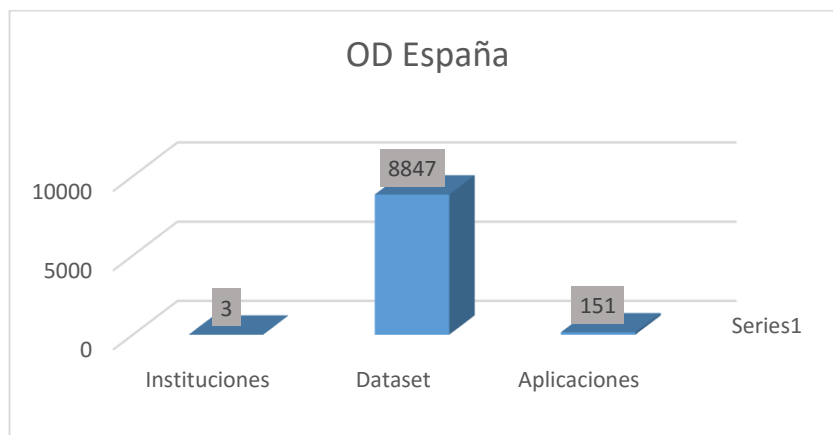


Figura 5 Estadísticas Datos Abiertos España

Tabla 7

Aplicaciones Publicadas en España

Aplicaciones			
Nombre	Descripción	Calificación	Descargas
BOE (Boletín Oficial de Estado)	Acceso a información de BOE	4	50 a 100 mil
Alcobendas	nuevo canal de difusión para la información del municipio madrileño	3,7	1000 a 5000
Agua y Mapas	Muestra a través de mapas, la situación real en España en materia de agua.		427

→Continuación

Citymapper	“Aplicaciones para encontrar rutas dentro de una gran ciudad	4,5	1 a 5 millones
Gasolineras Baratas	Encontrando la gasolinera más barata cercana a tu posición en España.	3,8	100 a 500 mil

2.6 Los datos abiertos en Latinoamérica

2.6.1 El Ecuador y los datos abiertos

Las primeras experiencias de los Datos Abiertos en el Ecuador, se las puede ver reflejado con la publicación de la “LEY ORGÁNICA DE TRANSPARENCIA Y ACCESO A LA INFORMACIÓN PÚBLICA” (Art.1-7), la cual rige para todas las instituciones públicas, de una u otra manera esto ha ayudado a liberar datos pero en formatos privados, lo cual no es un aporte para poder crecer en el ámbito de los Datos Abiertos.

Las primeras iniciativas formales por parte del Gobierno del Ecuador para incursionar en los datos abiertos se las ha realizado a través de la Secretaria Nacional de la Administración Pública, fomentando el uso de los datos públicos y datos abiertos, un claro ejemplo de esto es la “GUÍA DE POLÍTICA PÚBLICA DE DATOS ABIERTOS - GPP-DA-v01-2014 -” la cual permita regular y estandarizar la elaboración y publicación de información pública o de Datos Abiertos en las entidades de la Administración Pública Central, Institucional y Dependiente de la Función Ejecutiva (APCID), además de los artículos 11, 12 y 13 citados en la RESOLUCIÓN No. 007-DPE-CGAJ de la defensoría del pueblo para expedir los parámetros técnicos para el cumplimiento de las obligaciones de transparencia activa

establecidas en el art. 7 de la ley orgánica de transparencia y acceso a la información pública – lotaip.

También se tiene conocimiento de esfuerzos individuales que hasta el momento se puede ver mediante los siguientes portales. Uno de los primeros portales de datos abiertos en el Ecuador es datosabiertos.ec, el cual es una iniciativa de Fundapi con el apoyo de la Open Knowledge, el cual fue creado con el fin de fomentar el uso y adaptación de los Datos Abiertos en el país, además de facilitar el acceso de los ciudadanos a datos públicos en formatos y licencias que permitan su reutilización y la creación de aplicaciones y servicios.

- Actualmente solo el Ministerio de Finanzas y el SERCOP están ligadas al portal
- Solo posee 3 conjuntos de Datos publicados
- La última actualización fue realizada hace 3 meses y medio
- Tiene Datasets NO OFICIALES creados a partir de datos públicos del Ministerio de Finanzas de Ecuador y el SERCOP.

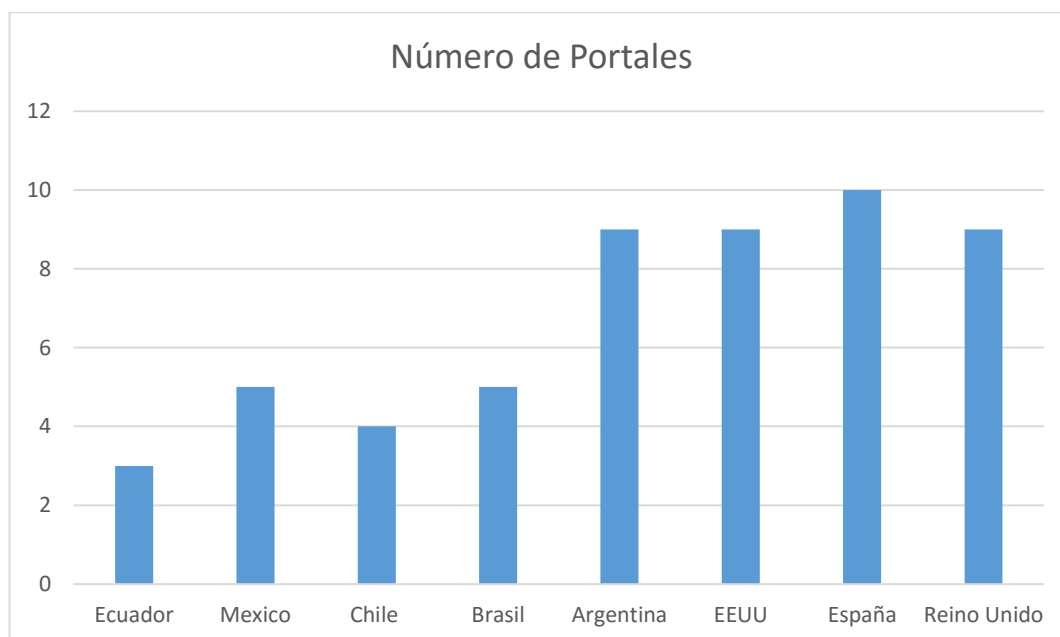


Figura 6 Número portales de Datos Abiertos por País

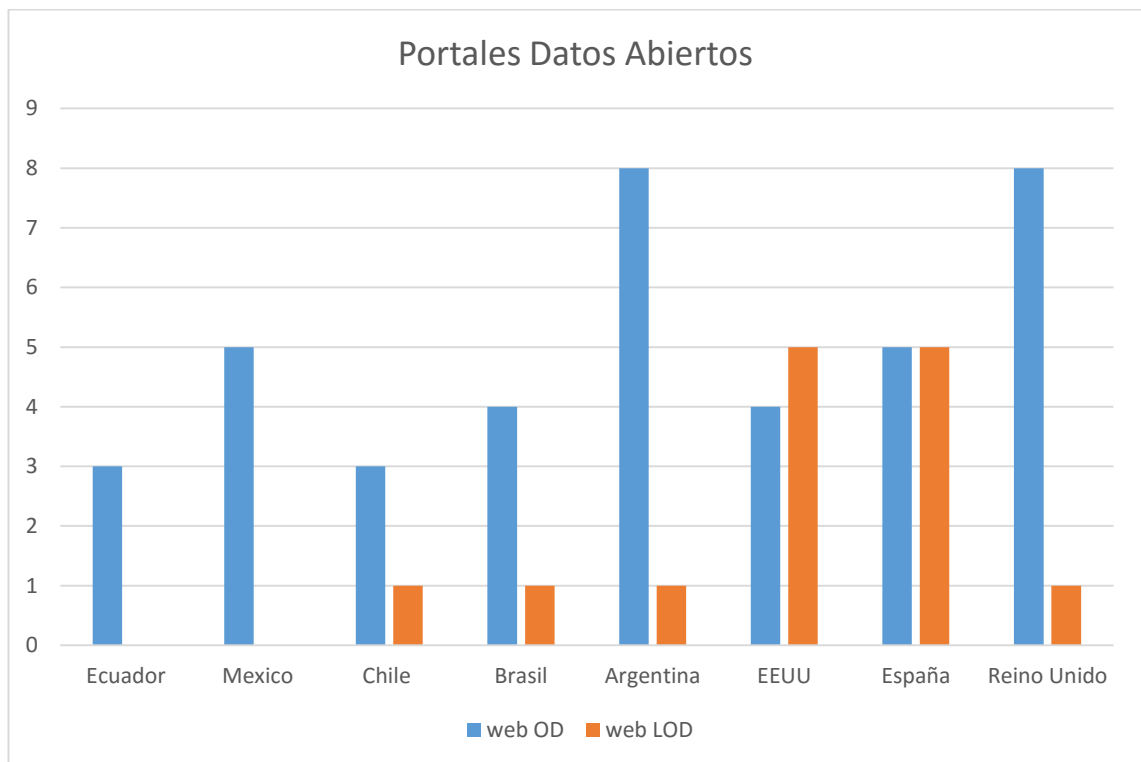


Figura 7. Numero portales de Datos Abiertos y Enlazados por País

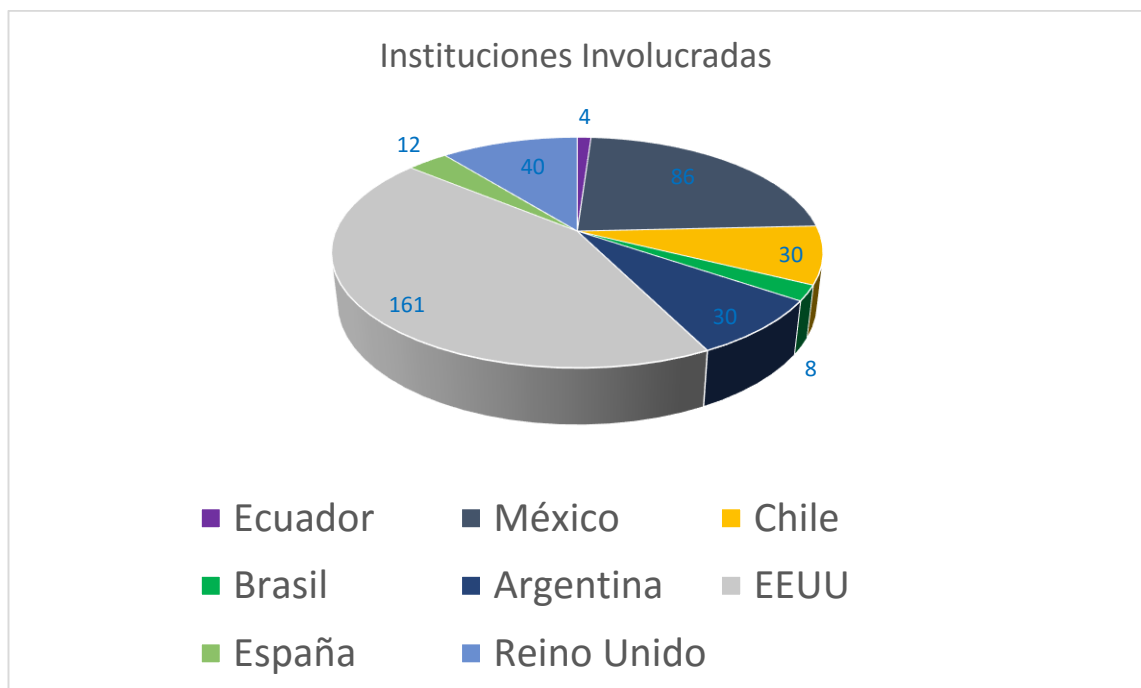


Figura 8. Instituciones Involucradas en los Datos Abiertos por País

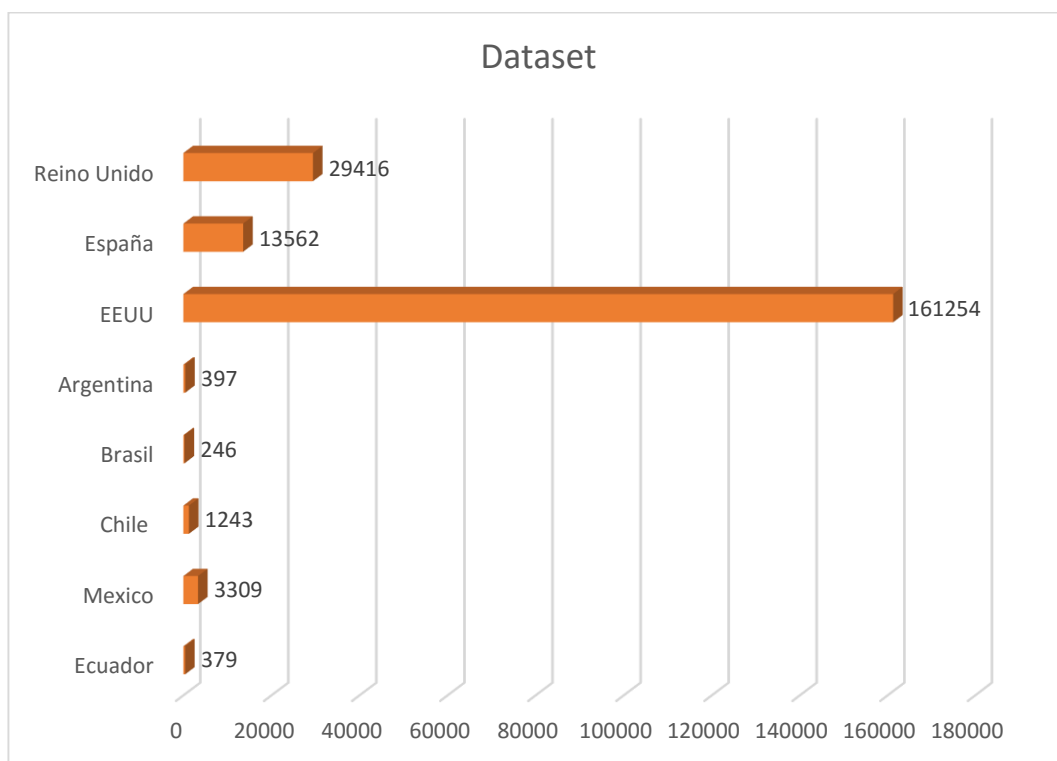


Figura 9. Conjuntos de datos liberados por País

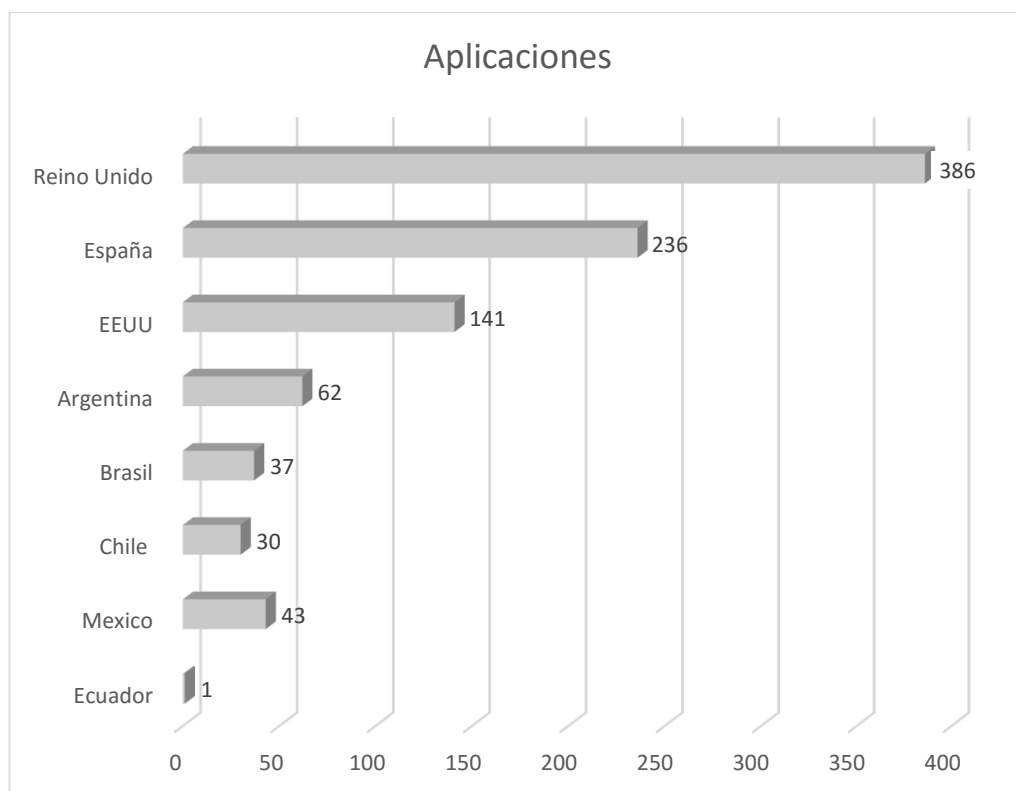


Figura 10. Aplicaciones elaboradas en base a Datos Abiertos por país

2.6.2 Nivel de madurez de los datos abiertos en Latinoamérica

Los Datos Abiertos en toda América Latina se han ido desarrollando en diferentes etapas o circunstancias. Uruguay fue el primer país en toda América Latina en desarrollar las primeras normas y estándares para el gobierno de Datos Abiertos, posterior a esto ciudades como San Pablo, México y Buenos Aires han hecho público que comenzaran con la apertura de sus respectivos datos para que puedan ser de uso público. El estudio “Opendata Latinoamérica” dio a conocer, que el 84% de los habitantes de toda Latino América están administrados por municipios que poseen gobierno de Datos Abiertos.

Países como Brasil, México, Uruguay, Chile, Costa Rica y Ecuador, entre otros países de la región de Latino América, están ubicados entre los 70 primeros países en tomar la iniciativa de los Datos Abiertos, los mismos que son monitoreados por el Censo de Datos Abiertos de Conocimiento Abierto, la cual es una entidad sin fines de lucro que está promoviendo los Datos Abiertos. El proyecto tiene por nombre “Opendata Latinoamérica” se basa en el modelo de un estudio similar desarrollado en África, su objetivo es consolidar todos los Datos Abiertos de la región y capacitar a la población en general, en la forma de utilizarlos.

Reino encabeza la lista de los países con más publicaciones de Datos Abiertos del 2014, manteniendo el primer lugar con una calificación general de 96%, seguido por Dinamarca. Francia sube al tercer lugar desde el décimo segundo que tuvo en el 2013. En cuanto a Latinoamérica, Colombia y Uruguay encabezan la lista en la posición #12, seguidos por Chile (#19), Brasil (#26), México (#28), Paraguay (#41), **Ecuador (#43)**, Argentina (#48), Costa Rica (#54), El Salvador (#57), Guatemala (#69), Panamá (#79) y Haití (#95).

Como información adicional en Brasil en la ciudad de Minas Gerais, se ha trabajado con el Laboratorio de Medios del MIT, han desarrollado el proyecto Data Viva, el cual es un sitio web en el cual los ciudadanos pueden visualizar varios tipos de datos económicos, donde están también los niveles salariales y las oportunidades de trabajo para diferentes tipos de plazas existentes. En el 2013, 25 entidades

realizaron 83 publicaciones de datos que permiten identificar las exportaciones que provienen del municipio, la deforestación y la ganadería ilegal en la selva amazónica.

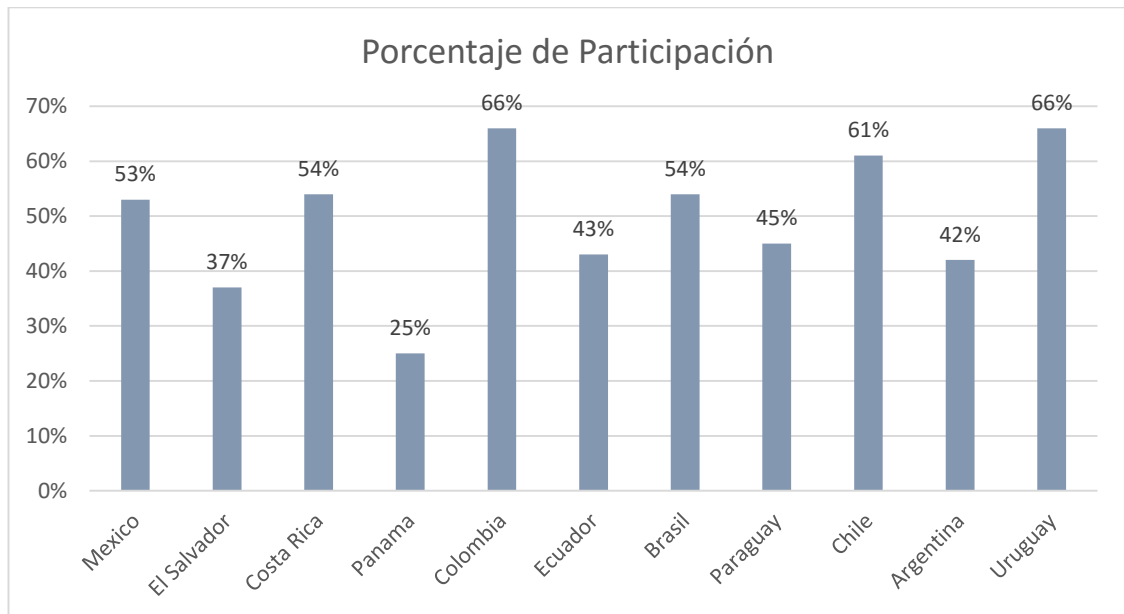


Figura 11. Participación en Iniciativas de Datos Abiertos

3 CAPÍTULO 3

GUIA METODOLÓGICA

3.1 Descripción

Para desarrollar esta investigación, se realizó un estado del arte basado en temas que respectan a las metodologías de la publicación de datos abiertos enlazados, sabiendo que es un tema nuevo en nuestro ámbito laboral, se ha considerado necesario también los requisitos más indispensables de infraestructura para el soporte de los Datos Abiertos enlazados, planteando un procedimiento y un modelo que permita dar cumplimiento a estos de forma concisa.

3.2 Propuesta Metodológica.

Se ha tenido algunas consideraciones y puntos, que han moldeado en gran manera el esquema de infraestructura, a continuación verán las características del contexto:

- Actualmente las instituciones públicas, no cuentan con un equipo totalmente capacitado en el área de web semántica, lo que dificulta un desenvolvimiento correcto desde el interior de las instituciones, por lo cual no se tiene una claridad total acerca de los requerimientos y sistemas necesarios para una infraestructura correcta que de un soporte a los Datos Abiertos Enlazados.
- Las entidades públicas en la actualidad ofrecen varios servicios a través de la Web, por lo que la implementación de nuevos proyectos no deberá interferir, con el desarrollo normal de las actividades de los diferentes proyectos ya sea que estos estén en ambiente de producción como en ambiente de pruebas.

- Las organizaciones donde se implementara esta metodología son las instituciones públicas vinculadas a la Secretaria de la Administración Publica, en estos escenarios, la información a publicar son de interés general para los ciudadanos, se tendrá un gran volumen de datos, los cuales serán consumidos de manera muy frecuente.

3.3 Arquitectura de soporte para datos abiertos enlazados

La arquitectura que se propone para implementación de los Datos Abiertos Enlazados, está definida por los siguientes elementos:

- **Base de datos relacional:** el motor de base de datos relacional, realiza tareas de consulta con varios objetivos, en la cual encontrarán la gestión de base de datos del portal de documentación, donde el control de la base de datos de cache para agilizar las consultas SPARQL, al sistema de almacenamiento RDF, para aquellas tareas de razonamiento y de mantenimiento del sistema.
- **Almacenamiento RDF:** El almacén RDF, es aquel que se encarga de gestionar las tripletas RDF, este componente será la base de trabajo para los componentes Endpoint SPARQL y grafo RDF.
- **Ontologías:** Conjunto de ontologías que permitan modelar el dominio sobre el cual se generaran los datos abiertos enlazados que sirve la infraestructura descrita.
- **Sitio web de documentación:** Este sitio web tiene estas principales funciones, albergar las ontologías del dominio del problema, almacenar la documentación para los desarrolladores, la cual está relacionada con las URIs y formatos del Grafo RDF y finalmente ser el acceso a todos los recursos de los datos abiertos enlazados.
- **Endpoint SPARQL:** herramienta la cual permite realizar consultas SPARQL, sobre un conjunto de tripletas RDF, definidas en un grafo RDF,

dando resultados aptos para el consumo de los clientes. Todas las consultas del Endpoint SPARQL son conjuntos de tripletas RDF, en diferentes formatos de sintaxis, sobre este componente se define los grafos RDF sobre HTTP, y bajo esta se definen el sistema de almacenamiento RDF y la base de datos relacional

- **Grafo RDF sobre Http:** Aplicación que genere un grafo RDF sobre URIs http. Esta herramienta es utilizada para la obtención de datos realizando consultas al Endpoint SPARQL, los datos son procesados y mostrados en URIs determinadas por el modelo del grafo. En el medio de habla ingles la herramienta es conocida como “Linked Data Frontend”.
- **Actualización del grafo RDF:** El objetivo de este componente es la actualización de la base de datos de la institución, hacia el almacén RDF, a partir de esto la creación de nuevas tripletas, ya que existen nuevos registros en la base de datos de la institución, o a su vez la eliminación de tripletas por obsolescencia.

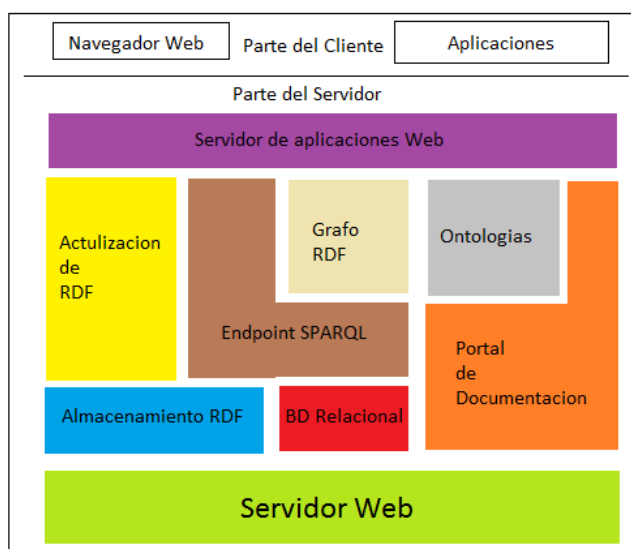


Figura 12. Propuesta de Arquitectura de Infraestructura para publicar Datos Abiertos Enlazados

3.4 Fases de implementación

Para llevar a cabo la implementación de la arquitectura planteada anteriormente, se ha propuesto una serie de fases que ayudarán con el proceso de adaptación, las cuales serán descritas posteriormente, Conceptualización, Diseño de Ontologías, Modelamiento del grafo RDF, Implementación del Endpoint SPARQL, Implementación de grafo RDF sobre http, actualización del grafo RDF, Requerimientos no funcionales, a continuación se explicara cada una de estas fases.

3.4.1 Conceptualización

En esta etapa se debe identificar el entorno de la aplicación y la condición de los datos que se van a liberar, para esto se puede utilizar un método de descripción de requerimientos como el UML, para la cual se planteó las siguientes preguntas.

- **Quienes los van a utilizar:** especifica quienes van a utilizar los datos, si serán un determinado grupo de usuarios, interfaces abiertos de consumo para el desarrollo de aplicaciones propias o de terceros y personas que visualicen los datos. El resultado de esta fase, es un acercamiento al sistema, la cual se describirá con exactitud, a través de un documento de análisis, que considera los apartados siguientes.
- **Formas y formatos, en que se van publicar los datos:** detallando los modelos de consulta y acceso a los datos, manteniendo como referente y prioridad los estándares libres como SOAP o REST. Determinar los formatos en que se entregaran los datos RDF, los posibles formatos de salida son HTML, RDF-XML, JSON, N3 o YAML. Se tomará en cuenta cómo se va a negociar el contenido, podría ser mediante HTTP 303 o mediante la negociación de contenido “transparente”.

- **Que datos se entregaran:** describir el modelo de dominio en forma de clases, atributos y relaciones, las cuales se deben expresar mediante notación UML, seguidamente el modelo se convertirá a un modelo RDF y descrito a través de una ontología que indique las clases como tales, los atributos como nodos, “DataType Properties” y las relaciones entre nodos como “Object Properties”

3.4.2 Diseño de Ontologías

En esta etapa complementa la definición de las ontologías respectivas al modelo de dominio determinado en la etapa anterior. Es una fase importancia para el sistema semántico ya que este modelamiento permitirá aplicar validaciones, reglas de consistencia e inferencias para los datos de tipo RDF. Por lo cual se consideró lo siguiente:

1. **Utilizar un modelo existente:** Si ya se tiene un modelo de dominio que ya ha sido elaborado y diseñado con anterioridad como una ontología y lo que se tiene disponible para usarlo, es buena opción utilizarlo en su totalidad o parcial. Para encontrar indicios de que si el modelo ya ha sido utilizado y modelado anteriormente, se podrá utilizar los siguientes buscadores de material semántico: Falcons, Swoogle, Watson o semejantes.
2. **Diseñar un modelo reutilizado:** Si se tiene un modelo de dominio que se necesita utilizar y no está modelado, se diseñará toda una ontología que represente el problema. Existen varias técnicas de ajuste de ontologías como para realizar el diseño, que permita la reutilización. En esta instancia los principios de reutilizar, no reinventar y mezclar libremente, se evidencian con mayor sentido.

Si se requiere implementar la ontología, la mejor opción es RDF Schema o OWL, teniendo en cuenta que RDF Schema posee una menor expresividad que OWL, con RDF Schema solo se puede expresar clases, propiedades y jerarquías, por otra parte OWL también posee definición de negaciones, cuantificadores, cardinalidades y atributos de propiedades.

En esta etapa es muy importante el establecimiento de una HTTP URI además del prefijo para la publicación de la ontología. La mejor solución es dejar todas las ontologías bajo un solo directorio “ontologies/”. Para el prefijo, este debe tener relación con el nombre o sigla de la organización los cuales deben ser únicos para su publicación. Si se encuentra modelando en la ontología solo parte de todo el dominio de la organización, lo más recomendable es realizar la implementación de la ontología de forma separada, haciendo referencia a la ontología raíz. De esta forma posible extender indefinidamente el diseño de la ontología y dominios de la aplicación a través de sub-ontologías. Para finalizar es recomendable que toda la ontología debería ser documentada en idioma inglés y en el idioma de contexto del proyecto.

3.4.3 Modelamiento del grafo RDF

En esta fase se diseñará las URIs sobre HTTP del grafo RDF al cual se accederá además, la estructura de datos que se visualizan en cada solicitud. Primero se definirán los patrones de URI, estos darán sentido al grafo RDF. Existen diversas aproximaciones que contribuyen con el modelamiento a través de patrones de diseño de URIs, el libro de Davis y Dodds son de gran ayuda, el cual trata de estudios sobre URIs como también Cool URIs creado o desarrollado por Berners-Lee. Se recomienda dar la importancia necesaria al diseño de las URIs abstrayendo los detalles de la implementación tecnológica.

Conjuntamente a diseñar cada patrón de URI en el grafo, también se necesita definir qué datos entregará cada acceso a la URI. En este aspecto se considera que los datos a entregar como resultado de cada consulta deben ser todos aquellos modelados en la ontología o vocabulario para la clase consultada. Para definir los

datos de salida, es un buen ejercicio detallar una URI de ejemplo describiendo el contenido RDF de salida. Para facilitar la realización del ejercicio se puede realizar sintaxis N3 que es más humanamente legible. Una vez que estén diseñados todos los patrones de URIs y salidas en RDF, el grafo RDF estará modelado casi por completo

Otro punto importante de esta fase es definir las URIs desde el punto de vista de la negociación de contenido. En este contexto se considera que si un usuario accede a cierto contenido publicado como datos enlazados a través del navegador Web y se pide mediante las cabeceras HTTP un cierto tipo de contenido, por ejemplo “text/html”, la aplicación tras el grafo debiera responder generando una representación correspondiente en HTML, mientras que si se está accediendo programáticamente y el tipo de contenido requerido es por ejemplo “application/rdf+xml”, la representación más apropiada debiera ser tripletas en RDF/XML. Sin embargo, en el caso que no exista tal parámetro en las cabeceras HTTP al momento de realizar la consulta, se considera que debe existir una representación por defecto, idealmente HTML.

Otro aspecto importante es el de la visibilidad de los datos para ayudar a rastreadores de los buscadores. Será muy útil la especificación de un mapa del sitio semántico en el archivo robots.txt del sitio Web, en donde se considere al menos la ubicación del Endpoint SPARQL, la URI base del sitio y la ubicación de un archivo de transferencia de datos si existiera.

Esta propuesta considera también la internacionalización de URIs cuando la contextualización del proyecto esté fuera del idioma inglés. Hay contextos de aplicación que exigen la publicación de los datos en el idioma de contexto del proyecto. Por ejemplo, si el gobierno Ecuatoriano quisiera publicar datos de cualquier institución educativa publica en la Web, perdería sentido el tener que acceder a URIs con partes como “school/” o “teacher” cuando los datos publicados estarían puestos a disposición del pueblo Ecuatoriano que no necesariamente sabe inglés. Sin embargo, si quienes quieren acceder a estos datos fueran de habla inglesa, por ejemplo para el desarrollo de algún mashing up comparativo de países, tiene sentido la publicación en inglés ya que es el idioma internacionalmente usado.

3.4.4 Implementación del Endpoint SPARQL

La siguiente fase contempla la implementación del denominado Endpoint SPARQL. Comúnmente las organizaciones mantienen y gestionan sus datos sobre bases de datos corporativas a las cuales acceden aplicaciones propias o de terceros. Para habilitar el uso de estos datos en el contexto de la Web Semántica, será necesario realizar un proceso de extracción, transformación y carga también denominado ETL en el contexto de la Inteligencia de Negocios, o al menos un proceso de transformación desde la base de datos corporativa hacia datos en RDF. En esta fase entonces se distinguen claramente dos tareas:

- **Extracción, transformación y carga de datos:** para realizar esta tarea que contempla el pasar los datos desde el uso en la organización hasta la publicación en la Web, se recomienda la utilización de herramientas ETL usadas en entornos de inteligencia de negocios o idealmente alguna herramienta de generación de tripletas como WESO-RUD,
- **Implementación del Endpoint SPARQL:** para realizar esta tarea se requieren básicamente tres elementos: un motor de base de datos para almacenar las tripletas RDF, una aplicación Web que conecte el motor de datos con la consulta SPARQL definida como entrada y si se requieren capacidades de razonamiento, un razonador que efectúe la operatoria lógica. Esta propuesta plantea el uso de un sistema de almacenamiento RDF basado en la arquitectura definida anteriormente, en donde las capacidades de razonamiento son delegadas a una capa de la arquitectura. En la práctica para resolver este componente, actualmente existen numerosas alternativas siendo algunas de ellas descritas en la sección 2.4.1 de este documento.

3.4.5 Implementación de grafo RDF sobre Http

En esta fase se realizará la implantación del componente generador del grafo RDF sobre las URIs HTTP, también denominado “Linked Data Frontend”. Para realizar esta tarea se requiere una aplicación capaz de recibir una petición por HTTP en una determinada URI y devolver un conjunto de datos en el formato indicado por la petición que normalmente será RDF. A nivel interno, la aplicación deberá comunicarse con el Endpoint SPARQL, realizando consultas SPARQL para cada petición HTTP, que pueden estar ya en la cache. Si bien esta aplicación puede ser desarrollada íntegramente a nivel programático, existen alternativas de software libre que proveen la funcionalidad como las descritas en la sección 2.4.2 del presente documento.

3.4.6 Actualización del grafo RDF

La implementación de fase permitirá mantener vigente el grafo RDF a través del tiempo. La idea principal es diseñar un servicio que de forma automatizada, sea capaz de mantener actualizado el grafo RDF con datos desde la base de datos corporativa de la organización. Para ello, se considera la utilización de alguna herramienta ETL como la definida anteriormente, pero cambiando la ejecución manual de los procesos de extracciones, transformaciones y cargas a programaciones periódicas, esto mediante utilidades como cron jobs o tareas programadas. Otra alternativa para resolver esta tarea, es la implementación de una aplicación externa basada en Web Services u otros métodos de transferencia como Sockets. Una consideración adicional en este contexto es mantener una comunicación segura a la hora de realizar el proceso de actualización

3.4.7 Web de documentación

Esta fase se propone la implementación de un portal web de documentación que sirva de soporte para el consumo de los datos. Los contenidos que deberá incorporar este portal como mínimo son los siguientes:

- Ontologías o Vocabularios diseñados.
- Una guía para el consumo de los datos publicados.
- Una guía de uso del Endpoint SPARQL que incluya consultas de ejemplo.
- Un índice de acceso a los servicios ofrecidos por nuestra infraestructura semántica.

Dentro de las características no funcionales del portal Web se consideran:

- **Internacionalización:** Si el idioma de contexto del proyecto no es el inglés, deberá existir una versión principal en el idioma de contexto, pero además una en inglés de manera que facilite la utilización y entendimiento de la documentación por usuario o entidades internacionales, ya que estas hacen sus estudios en idioma inglés.
- **Accesibilidad:** Este aspecto toma especial relevancia, ya que el sistema de datos enlazados está desarrollado en el contexto de las administraciones públicas, se considera que se deben cumplir los niveles de conformidad de accesibilidad web descritos en WACG, esto para asegurar la disponibilidad a toda la ciudadanía.
- **Estandarizado:** la implementación del portal web deberá estar en conformidad con los estándares W3C, asegurando la compatibilidad con cualquier navegador web

Por último una característica útil, pero no obligatoria, es implementar negociación de contenido en la entrega de la documentación, ya que de esta forma la documentación podrá ser entregada en distintos formatos dependiendo de las necesidades del cliente

3.4.8 Requerimientos no funcionales

Para la implementación de la infraestructura de Datos Abiertos Enlazados se considera la adopción de algunos requerimientos no funcionales que ofrecerán ventajas en producción. Algunos de los que se han considerado más relevantes son los siguientes:

- **Cache para el grafo RDF:** Con el propósito de mejorar el tiempo de respuesta de consultas que contemplen razonamiento o cálculos con grandes volúmenes de datos, se debe tener en cuenta la implementación de caches entre el motor de base de datos RDF y la aplicación de representación del grafo.
- **Seguridad:** Como ya se ha mencionado anteriormente, para el servicio de actualización del grafo, se consideran medidas de seguridad en la transferencia de los datos que generarán tripletas visibles sobre el grafo. Para ello se proponen soluciones como el uso de SSL en la transferencia o la utilización de VPN. Este aspecto toma vital importancia tratándose de administraciones públicas, en donde la información está relacionada totalmente con la realidad del país.
- **Monitoreo de rendimiento:** conocido como “profiling”, el monitoreo de rendimiento permitirá conocer las rutinas que se deban optimizar dentro del sistema de datos enlazados, habilitando un correcto funcionamiento a través del tiempo.

3.4.9 Herramienta alternativa de visualización de datos

Dentro de las aplicaciones más atractivas cuando se tienen disponibles datos en formato Linked Data, son las visualizaciones o mashups, más aun en entornos Web 2.0 como lo actuales. Considerando entonces que es posible inter operar con otras fuentes de forma limpia, sin necesidad de consumir datos mediante APIs o

transformaciones desde formatos específicos, es una buena idea diseñar algún mashup o visualización que permita representar los datos de una forma atractiva al usuario y al mismo tiempo dar mayor valor agregado al proyecto de implantación de Linked Data. Para la implantación de alguna herramienta de visualización de datos, existen múltiples alternativas libres de uso extendido, tanto como Linked Data Browsers o como Mashups. Algunos ejemplos de estos tipos de herramientas son descritas en la sección 2.4.3 de este documento. Otra alternativa es el diseño y desarrollo de un Mashup propio. En este sentido algunas pautas a considerar para el diseño son las siguientes:

- **Posibilidad de consultar:** una de las características que dan más valor a una herramienta de visualización es la capacidad de obtener resultados únicos. En este Sentido, es una buena idea el tener un pequeño conjunto de datos pre calculado, pero dando mayor potencial a la posibilidad de realizar consultas que generen información nueva en tiempo real proveniente de diversas fuentes.
- **Foco en la usabilidad:** la herramienta debe ser fácilmente entendible, tanto los controles que permiten interacción con el usuario como los resultados obtenidos.
- **Impacto visual:** la herramienta será utilizada cuanto más atractiva sea. En este aspecto es importante cuidar la interfaz gráfica de usuario, tanto en el diseño gráfico usado como en la interacción.
- **Utilizar diversas fuentes de datos:** uno de los atractivos mayores de las herramientas de visualización que utilizan Linked Open Data, es la posibilidad de cruzar diferente información proveniente de diferentes fuentes, aspecto que está en total concordancia con los principios de Linked Data.
- **Dimensionar la información:** la información puede ser dividida en diversas escalas tales como: temporal, económica o social.
- **Funcional sobre Web:** si la herramienta consume datos abiertos enlazados para su operación, lo ideal es que pueda ser utilizada por cualquiera desde cualquier parte. Por ello, a nivel general lo ideal es que la herramienta de visualización funcione sobre la Web.

4 CAPÍTULO 4

DATOS ABIERTOS RELEVANTES A PUBLICAR

4.1 Antecedentes del Proyecto

Como ya se ha mencionado, el estudio está en el contexto legislativo de la información. Como legislativo de la información se define al conjunto de normas y políticas que rigen el ordenamiento de las entidades públicas. La legislación de la información son todos los datos que provienen del sector público, que es producida por órganos del Estado y financiada con recursos públicos. Esta información es por naturaleza de interés público, ya que trata temas de interés general y particular que afectan la vida del ciudadano y por ende es de un alto valor público, porque genera experiencia en ciudadanos que consideran valiosa.

Por lo anterior se puede deducir que la legislación de la información, son datos públicos y debe ser de dominio público. A esto hay que agregar que la legislación de la información debe cumplir el denominado “mecanismo de concreción de principio de seguridad”, que es la fundamentada expectativa que tienen los ciudadanos de que la ley vigente se cumpla. Pero para cumplir con esto, son necesarios dos conceptos de seguridad, a saber: la “seguridad jurídica”, que es la certidumbre fundada y garantizada que la norma será cumplida y la “certeza jurídica”, que es la perceptibilidad de la norma jurídica y la certidumbre de su contenido.

Para asegurar estos principios, cada País posee un mecanismo de publicidad de la legislación la cual se hace pública por lo general a través de un boletín oficial y de los diferentes portales web de las entidades involucradas.

En el caso particular La Secretaria Nacional de la Administración Pública, las siguientes normas, políticas y guías ayudan con la publicación de datos:

- "Políticas Públicas de Datos Abiertos y la Norma Técnica de Interoperabilidad".
- la "GUÍA DE POLÍTICA PÚBLICA DE DATOS ABIERTOS - GPP-DA-v01-2014"

Actualmente se está analizando la Ley de Software Público y donde se debe tener en cuenta los siguientes artículos:

El artículo 66 de la Norma Fundamental, en su numeral 19, establece: "El derecho a la protección de datos de carácter personal, que incluye el acceso y la decisión sobre información y datos de este carácter, así como su correspondiente protección. La recolección, archivo, procesamiento, distribución o difusión de estos datos o información requerirán la autorización del titular o el mandato de la ley"; por su parte, en su numeral 25, prescribe: "El derecho a acceder a bienes y servicios públicos y privados de calidad, con eficiencia, eficacia y buen trato, así como a recibir información adecuada y veraz sobre su contenido y características"; y, en su numeral 28, dispone: "El derecho a la identidad personal y colectiva, que incluye tener nombre y apellido, debidamente registrados y libremente escogidos; y conservar, desarrollar y fortalecer las características materiales e inmateriales de la identidad, tales como la nacionalidad, la procedencia familiar, las manifestaciones espirituales, culturales, religiosas, lingüísticas, políticas y sociales";

La Ley Orgánica de Transparencia y Acceso a la Información Pública, publicada en el Registro Oficial, Suplemento 337, del 18 de mayo de 2004, señala en el artículo 4, los principios para el desarrollo del derecho de acceso a la información pública; y, en su literal a), establece que: "Se observará que la información pública pertenece a los ciudadanos y ciudadanas. El Estado y las instituciones privadas depositarias de archivos públicos, son sus administradores y están obligados a garantizar el acceso a la

Información"; por su parte, el literal b), prescribe: "El acceso a la información pública, será por regla general gratuito a excepción de los costos de reproducción y estará regulado por las normas de esta Ley"; y, su literal e), dispone: "Garantizar el manejo

Transparente de la información pública, de manera que se posibilite la participación ciudadana en la toma de decisiones de interés general y la rendición de cuentas de las diferentes autoridades que ejerzan el poder público”;

Mediante Decreto Ejecutivo No. 1384 del 13 de diciembre de 2012, publicado en el Registro Oficial Segundo Suplemento No. 86, del 2 de enero de 2013, en el artículo 1, dispone: “Establecer como política pública, el desarrollo de la interoperabilidad gubernamental, que consiste en el esfuerzo mancomunado y permanente de todas las entidades de la Administración Central, dependiente e institucional para compartir e intercambiar entre ellas, por medio de las tecnologías de la información y comunicación, datos e información electrónicos que son necesarios en la prestación de los trámites y servicios ciudadanos que prestan las entidades, así como en la gestión interna e interinstitucional”; por su parte, en su artículo 2, determina que: “La interoperabilidad gubernamental entre todas las Entidades de la Administración Pública Central, dependiente e institucional, será gestionada y normada por el conjunto de principios, políticas, procesos, procedimientos y estándares en los ámbitos operativo, conceptual y tecnológico dictado que para el efecto dicte la Secretaría Nacional de la Administración Pública”;

4.2 Los ocho principios de los datos abiertos

A continuación se presentarán los 8 principios que se han venido convertido en un excelente punto de partida, para llevar a cabo la evaluación de la apertura de los datos en las entidades públicas, los datos de las entidades públicas serán considerados abiertos si cumplen con los siguientes principios:

- **Completos:** todos los datos públicos están disponibles. Estos datos son aquellos que no están sujetos a restricciones de privacidad, seguridad o privilegio
- **Primarios:** los datos son obtenidos en la fuente, con el mayor nivel posible de granularidad, sin ser modificados ni agrupados.

- **Periódicos:** Los datos quedan disponibles tan pronto como sea necesario para preservar su valor.
- **Accesibles:** los datos quedan disponibles para la mayor cantidad posible de usuarios y propósitos.
- **Procesables:** Los datos están razonablemente estructurados, de forma de permitir su procesamiento automático.
- **No discriminatorios:** Los datos están disponibles para todos, sin requisitos de registro.
- **No propietarios:** Los datos están disponibles en un formato sobre el que ninguna entidad tiene control exclusivo.
- **Sin licencia:** Los datos no están sujetos a ningún tipo de regulación de derechos, patentes o registros de marca. Se podrán permitir restricciones razonables de privacidad, seguridad o privilegio.

Estos principios no indican cuales datos deberían ser públicos, sino que especifican las condiciones para que los datos públicos sean, además, abiertos.

4.3 Conceptualización

A continuación se describe el proceso de conceptualización realizado para el estudio.

4.3.1 Identificación de las necesidades

Las principales necesidades que se han logrado identificar en las entidades públicas con respecto a los Datos Abiertos van desde las más básicas hasta las complejas, a continuación se nombran:

- Sociabilizar el concepto “Los Datos Abiertos Enlazados” en las entidades publicas
- Elaboración de una guía estándar para la publicación liberación y consumo de Datos Abiertos Enlazados.
- Construcción de portales web que faciliten el consumo y al reutilización de la información a liberar.
- Generación de nuevas aplicaciones que impulsen la innovación en los gobiernos.

4.3.2 Que datos se van a entregar

Para tener claro que datos se van a entregar se debe tener en cuenta las normativas vigentes que están vinculas con la liberación de información, tal como se hace referencia en la Ley Orgánica de Transparencia y Acceso a la Información (LOTAIP), los conjuntos de datos referentes a la gestión pública, oferta y demanda de los servicios públicos, dichos datos deben ser útiles capaces de promover la creatividad emprendimiento e innovación.

Claro todo esto, cada entidad pública debe conformar un grupo de especialista, los cuales analizaran qué información es importante para la publicación, basándose en lo mencionado anteriormente y sin dejar de lado los siguientes artículos de la LOTAIP:

- Artículo 6: Toda información es libre para su publicación salvo información confidencial
- Artículo 17,18: Toda información es libre para su publicación salvo información reservada, la cual comprometa aspectos de la seguridad informática.
- Lo que afecte o comprometa aspectos de seguridad informática que pudieran resultar en un peligro de seguridad.

La SNAP actualmente solicita a las entidades de la Administración Pública Central Institucional y Dependiente de la Función Ejecutiva, los datos de uso de servicios públicos, atención al ciudadano. En el marco de Datos Abiertos, se están solicitando otras bases de datos, tal como lo señala el artículo 7 de la LOTAIP.

Art. 7.- Difusión de la Información Pública.-

Por la transparencia en la gestión administrativa que están obligadas a observar todas las instituciones del Estado que conforman el sector público, en los términos del artículo 118 de la Constitución Política de la República y demás entes señalados en el artículo 1 de la presente Ley, difundirán a través de un portal de información o página web, así como de los medios necesarios a disposición del público, implementados en la misma institución, la siguiente información mínima actualizada, que para efectos de esta Ley se la considera de naturaleza obligatoria:

- La estructura orgánica funcional de la institución, base legal que la rige, regulaciones y procedimientos internos aplicables a la entidad; las metas y objetivos de las unidades administrativas de conformidad con sus programas operativos.
- El directorio completo de la institución, así como su distributivo de personal;
- La remuneración mensual por puesto y todo ingreso adicional, incluso el sistema de compensación, según lo establezcan las disposiciones correspondientes;
- Los servicios que ofrece y las formas de acceder a ellos, horarios de atención y demás indicaciones necesarias, para que la ciudadanía pueda ejercer sus derechos y cumplir sus obligaciones;
- Texto íntegro de todos los contratos colectivos vigentes en la institución, así como sus anexos y reformas;

- Se publicarán los formularios o formatos de solicitudes que se requieran para los trámites inherentes a su campo de acción;
- Información total sobre el presupuesto anual que administra la institución, especificando ingresos, gastos, financiamiento y resultados operativos de conformidad con los clasificadores presupuestales, así como liquidación del presupuesto, especificando destinatarios de la entrega de recursos públicos;
- Los resultados de las auditorías internas y gubernamentales al ejercicio presupuestal;
- La información completa y detallada sobre los procesos precontractuales, contractuales, de adjudicación y liquidación, de las contrataciones de obras, adquisición de bienes, prestación de servicios, arrendamientos mercantiles, etc., celebrados por la institución con personas naturales o jurídicas, incluidos concesiones, permisos o autorizaciones;
- Un listado de las empresas y personas que han incumplido contratos con dicha institución;
- Planes y programas de la institución en ejecución;
- El detalle de los contratos de crédito externos o internos; se señalará la fuente de los fondos con los que se pagarán esos créditos. Cuando se trate de préstamos o contratos de financiamiento, se hará constar, como lo prevé la Ley Orgánica de Administración Financiera y Control, Ley Orgánica de la Contraloría General del Estado y la Ley Orgánica de Responsabilidad y Transparencia Fiscal, las operaciones y contratos de crédito, los montos, plazo, costos financieros o tipos de interés;
- Mecanismos de rendición de cuentas a la ciudadanía, tales como metas e informes de gestión e indicadores de desempeño;
- Los viáticos, informes de trabajo y justificativos de movilización nacional o internacional de las autoridades, dignatarios y funcionarios públicos;
- El nombre, dirección de la oficina, apartado postal y dirección electrónica del responsable de atender la información pública de que trata esta Ley;
- La Función Judicial y el Tribunal Constitucional, adicionalmente, publicarán el texto íntegro de las sentencias ejecutoriadas, producidas en todas sus jurisdicciones;

- Los organismos de control del Estado, adicionalmente, publicarán el texto íntegro de las resoluciones ejecutoriadas, así como sus informes, producidos en todas sus jurisdicciones;
- El Banco Central, adicionalmente, publicará los indicadores e información relevante de su competencia de modo asequible y de fácil comprensión para la población en general;
- Los organismos seccionales, informarán oportunamente a la ciudadanía de las resoluciones que adoptaren, mediante la publicación de las actas de las respectivas sesiones de estos cuerpos colegiados, así como sus planes de desarrollo local; y,
- El Tribunal de lo Contencioso Administrativo, adicionalmente, publicará el texto íntegro de sus sentencias ejecutoriadas, producidas en todas sus jurisdicciones. La información deberá ser publicada, organizándola por temas, ítems, orden secuencial o cronológico, etc., sin agrupar o generalizar, de tal manera que el ciudadano pueda ser informado correctamente y sin confusiones

Además de basarse en lo que exige la ley, la SNAP ha elaborado una hoja de ruta la cual es entregada a las instituciones públicas afines a la SNAP, en la se define una estructura para la identificación y liberación de información , lo cual facilita el trabajo en cuanto al proceso de identificar que datos debe liberar cada entidad, quedando en manos de cada institución, la cual debe tener claro el marco legal para hacer la liberación correcta, así de esta manera se tendrán los datos específicos a entregar por cada institución, a continuación se verá el proceso que se debe seguir para la liberación de la información:

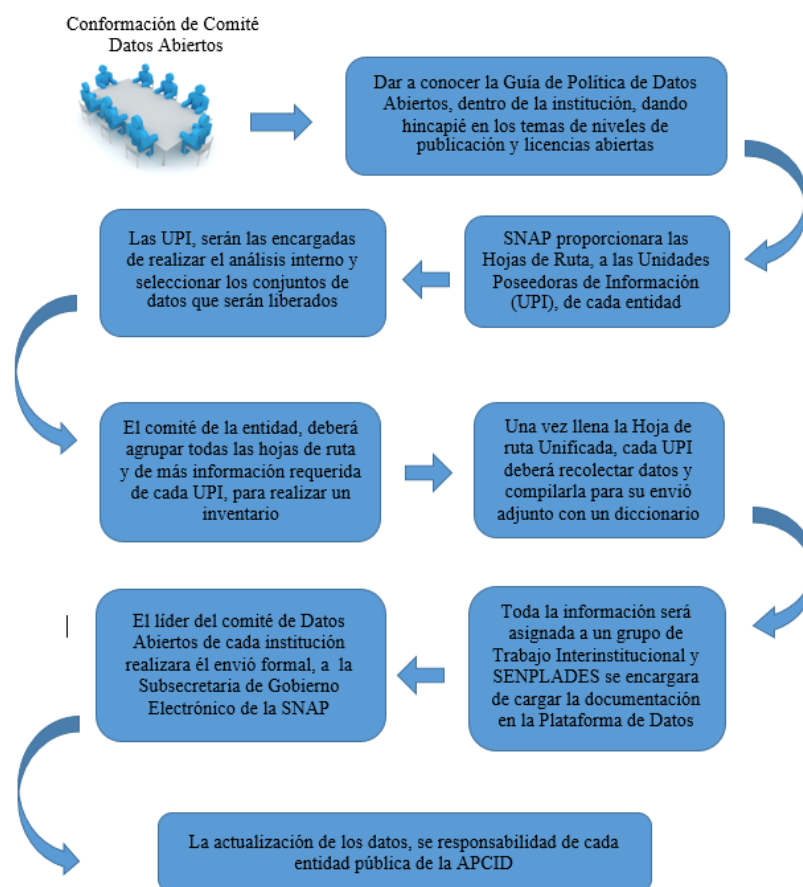


Figura 13. Proceso de liberación de información en las entidades publicas

Como referencia de que datos se tendrá para la liberación se basó en varios estudios que han intentado identificar aquellos conjuntos de datos considerados como de alto valor. Como referencia a este estudio se recopila e indica los siguientes dominios como relevantes y prioritarios:

Tabla 8

Dominios importantes para liberar información

Dominios de información de alto valor	Ejemplos
Información sobre negocios	Registro de empresas
Criminalidad y justicia	Estadísticas sobre delitos
Observaciones de la Tierra	Observaciones y pronósticos meteorológicos
Educación	Listas de centros educativos e índices de rendimiento en los mismos
Energía y medioambiente	Niveles de polución en el aire
Contratación pública	perfil del contratante, licitaciones
Datos financieros	Presupuestos y gastos
Información geo-espacial	Vartografía
Desarrollo global	Syuda internacional, seguridad alimentaria
Democracia y rendición de cuentas	Información sobre los representantes del gobierno
Salud	Datos sobre prescripciones médicas y gasto farmacéutico
Ciencia e investigación	Datos sobre el genoma, resultados de experimentos
Estadísticas	Estadísticas socio-demográficas
Bienestar social	Estadísticas sobre seguridad social, vivienda
Transporte e infraestructuras	Horarios de transporte público, penetración del ancho de banda

Según la guía de la política indique que en principio se dará prioridad solo a los datos contemplados en la ley de transparencia inicialmente serán publicados, ahora bien con los cambios y avances tecnológicos es prudente que se requiera liberar datos con alto impacto social y que puedan ser reutilizar para generar innovación social, principalmente en servicios públicos. Cabe recalcar que la SNAP ha facilitado.

4.3.3 Quien va a consumir los datos

Para dar respuesta a esta interrogante, se presentan los requerimientos a través de un diagrama UML de casos de uso representado por la figura 4.2:

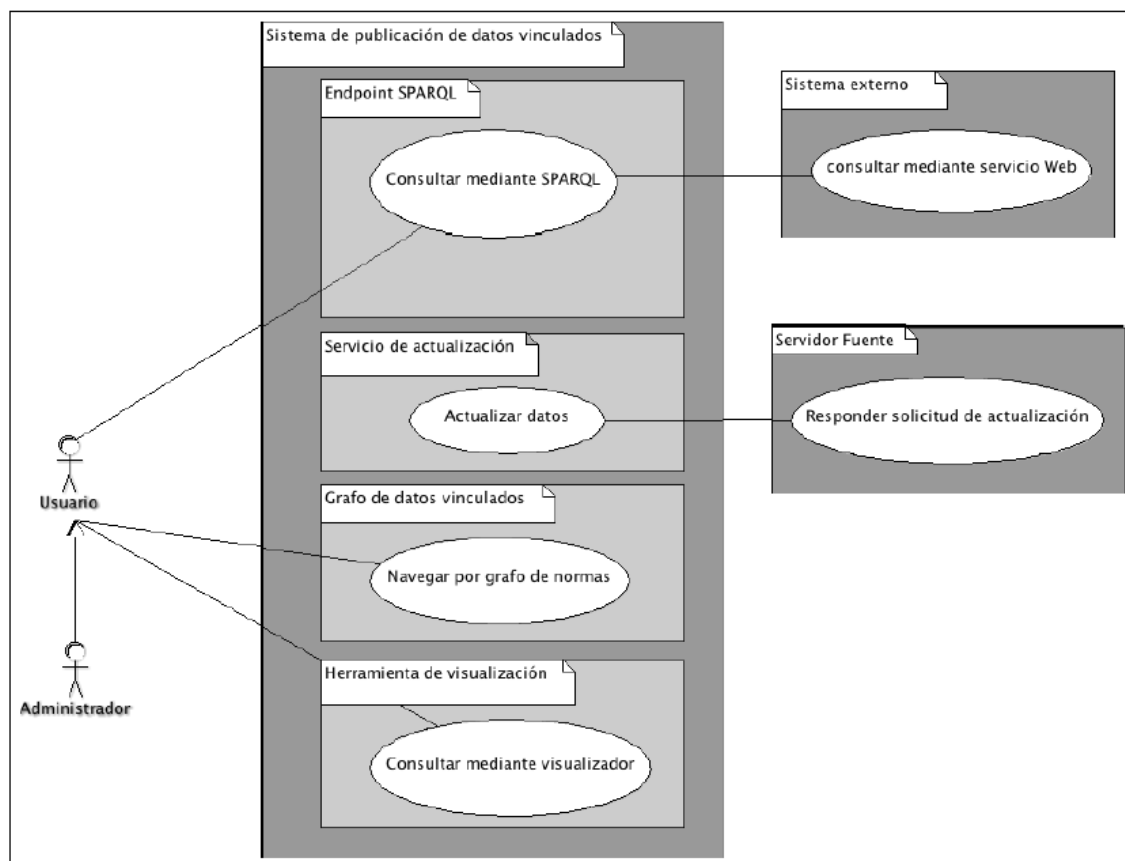


Figura 14 diagrama de casos de uso

En el diagrama se aprecian tres sistemas principales:

1. **Sistema de publicación de datos vinculados:** es el sistema web como tal, el cual estará alojado en diferentes servidores.
2. **Sistema externo:** Representa cualquier aplicación externa que consuma el servicio web disponible en el Endpoint SPARQL para consultas programáticas.
3. **Servidor fuente:** Representa al servidor que contiene la base de datos corporativa que son actualizadas diariamente, la herramienta de actualización

Actores:

1. **Usuario:** usuarios de este tipo podrán realizar todas las acciones de consulta y navegación dentro de los sistemas web.
2. **Administrador:** este tipo de usuario es una extensión al tipo Usuario, permite además de realizar todo tipo de consultas y navegación, gestionar y administrar tripletas a través de una interfaz de usuario de mantenimiento del almacén RDF.

4.3.4 Formatos para la liberación de información

Para llegar a definir qué formatos de liberación de información, se tendrá que tomar en cuenta la capacidad tecnológica y de conocimientos en que se encuentran las entidades públicas en el País, segundo lo investigado y en referencia a la guía de la política pública, se ha logrado identificar que las entidades públicas están en la capacidad de llegar al tercer nivel en cuanto a la definición de formatos.

Los datos deben estar estructurados de tal forma que permitan su procesamiento automático y procesables por máquinas y que puedan ser descargables y operados con los requerimientos tecnológicos mínimos.

Tabla 9

Define los formatos en que los datos abiertos se debían entregar

Niveles	Descripción	Formato
★	Los datos deben estar en la web en cualquier formato y bajo una licencia abierta, para que sean considerados Datos Abiertos.	PDF, DOC, JPG
★★	Los datos deben estar estructurados y en un formato que pueda ser interpretado o procesado por máquinas.	XLS
★★★	Los datos se deben entregar en un formato que no sea propietario.	CSV, XML

- Los datos deben ser publicados sin modificaciones, como fueron recolectados en su origen, y no separados o agregados
- Los datos deben estar disponibles tan pronto como sea necesario para preservar su valor
- Los datos deben estar disponibles para el mayor número de usuarios posible, para que puedan participar de la más amplia gama de propósitos.
- Los datos deben estar disponibles para cualquier ciudadano, sin necesidad de registro

Es importante tener en cuenta que los datos deben estar disponibles en formatos abiertos que permitan su reutilización, redistribución e integración a otros datos para facilitar el uso interactivo de la información.

Tabla 10.

Tipos de archivos según su uso.

Datos	Formato	Definición
Genéricos	CSV	Tipo de documento en formato abierto sencillo para representar datos en formato de tabla. Las columnas se separan por comas (o punto y coma) y las filas por saltos de línea.
	XML	Lenguaje de marcado desarrollado por la W3C para almacenar datos en forma legible y estructurada, que propone un estándar para el intercambio de información entre diferentes plataformas.
Geográficos	KML	Es un formato XML orientado a la visualización geográfica, que incluye anotaciones de mapas e imágenes.
	GML	Es un formato de intercambio abierto para transacciones de información geográfica a través de Internet.

Solo se ha considerado los formatos de liberación de información de hasta el tercer nivel, debido a que los datos abiertos en el país están dando sus primeros pasos y por el momento las instituciones públicas solo están en la capacidad de entregar ese tipo de formatos.

Para entregar Datos Abiertos Enlazados se tendrá que emprender un avance a los siguientes niveles con los siguientes formatos presente en la tabla 11 y 12

Tabla 11,

Estrellas Datos Abiertos

Niveles	Descripción	Formato
★★★★	Usar URIs para identificar cosas y propiedades, de manera que se pueda apuntar a los datos.	URI, RDF SPARQL
★★★★★	Enlazar los datos con otros datos externos.	Datos enlazados/Lin ked Open Data

Los formatos de liberación de información también van muy ligados con los tipos de licencia para sus publicación, por este motivo no se debe pasar por alto las licencias públicas que la Guía de Política Pública de Datos Abiertos utilizará, a continuación en el siguiente gráfico.

Tabla 122.

Licencias para publicación de datos

Licencia	Definición
Creative Commons 4.0 (CC BY)	Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de una obra, incluso con fines comerciales, siempre y cuando se den crédito por la creación original. Para poder aplicar esta licencia debe visitar el sitio web https://creativecommons.org/licenses/by/4.0/legalcode#languages y remitirse a las condiciones de uso de la licencia.
Open Database License (ODbL)	Esta licencia es orientada a bases de datos, la cual permite copiar, distribuir, cruzar, combinar, modificar y crear a partir de dicha obra, incluso con fines comerciales, siempre y cuando se den créditos a la fuente creadora y se mantenga la misma licencia para cualquier obra derivada. Para poder aplicar esta licencia debe visitar el sitio web http://opendatacommons.org/licenses/odbl/ y remitirse a las condiciones de uso de la licencia

4.3.5 Como reutilizar los datos liberados

Para la reutilización de la información hay que tener en claro lo siguiente la Reutilización de la Información del Sector Público (RISP) es el objetivo principal de la iniciativa de los Datos Abiertos. Consiste en poner la información del sector público disponible, en bruto y en formatos estándar abiertos, facilitando su acceso y

permitiendo su reutilización tanto a particulares como a empresas para fines comerciales o no.

El sector público produce una gran variedad de información que es potencialmente reutilizable, como puede ser la información social, económica, geográfica, estadística, etc. Esta información es muy atractiva para su reutilización ya que es completa, fiable y de calidad.

También hay que tomar en cuenta lo que plantea, la World Wide Web Consortium (W3C), la comunidad internacional que desarrolla estándares para el crecimiento de la web, recomienda que los datos que se liberen sean útiles y estén en formatos reutilizables. Cuanto mejor estructurados y enriquecidos estén los datos, más fácil será reutilizarlos y construir aplicaciones que puedan tratarlos automáticamente.

Se basará en lo planteado anteriormente se plantea el siguiente listado de buenas prácticas que debe seguir en cualquier administración pública que quiera llevar a cabo la reutilización de los datos:

- Publicar datos en formatos abiertos y estándares.
- Usar esquemas y vocabularios consensuados.
- Inventario de un catálogo de datos estructurado.
- Datos accesibles desde direcciones web persistentes, amigables.
- Exponer un mínimo de conjunto de datos relativos al nivel de competencias del organismo y un roadmap de publicación.
- Compromiso de servicio, actualización y calidad del dato, manteniendo un canal eficiente de comunicación re utilizador.
- Monitorizar y evaluar el uso y servicio mediante métricas.
- Datos bajo condiciones de uso no restrictivas y comunes.
- Evangelizar y educar en el uso de datos.

- Recopilar aplicaciones, herramientas y manuales para motivar y facilitar la reutilización.

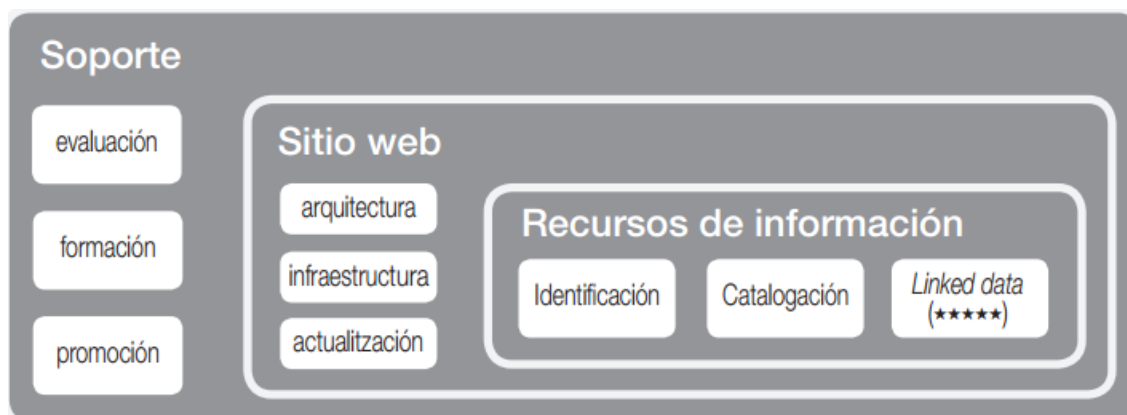


Figura 15 Elementos a incluir en una estrategia para la reutilización de datos.

La estrategia debería, al menos, seguir los siguientes pasos: Identificación y selección de la información potencialmente reutilizable, partiendo de los datos que ya ofrece el organismo. Inicialmente se optará por aquellos que puedan ser descargables de forma directa, desde sus portales web.

Diseño de un procedimiento para la preparación y generación de conjuntos de datos y documentos en el mejor formato posible incluso llegando a considerar un modelo de datos enlazados, teniendo en cuenta las limitaciones técnicas o legales (descartando información afectada por derechos de propiedad intelectual, datos de carácter personal, o por seguridad), mejorando la calidad de los mismos y priorizando los conjuntos de datos demandados por los agentes re utilizadores.

Establecimiento de las condiciones de reutilización aplicables dentro del marco legal que afecta al organismo público. Creación de un espacio web específico, donde se incluirá información divulgativa sobre el proyecto, información legal, elementos de comunicación con terceras partes y un inventario con acceso a los datos disponibles.

Publicación y mantenimiento de los datos y documentos, que serán catalogados y accesibles desde el sitio web. Los catálogos de dichos recursos deberán contener metadatos, información estructurada que describe, explica y localiza un

recurso de información actualizados para facilitar su descubrimiento y uso. Asimismo, es importante informar al público claramente sobre las condiciones de reutilización aplicables a los conjuntos de datos y documentos.

Definición de una estrategia de comunicación y formación que permita dar a conocer la base del proyecto, sus objetivos, intentando involucrar en este movimiento a otros colectivos.

Definición de una estrategia de evaluación y mejora de la calidad de la iniciativa, donde se estimen unos objetivos cuantificables, basándose en unos indicadores específicos (acceso a los conjuntos de datos, aplicaciones desarrolladas, etc.).

4.5 Perspectiva tecnológica

Esta perspectiva determina la capacidad tecnológica de la organización para articular una visión consistente y técnicamente viable de Datos Abiertos. Esto es, la existencia de una estrategia tecnológica alineada con la estrategia de la entidad pública que contemple explícitamente las buenas prácticas de *Datos Abiertos* en la organización.

Para asegurar los aspectos de disponibilidad, seguridad, acceso y calidad de los datos es necesario cautelar algunos elementos fundamentales por su carácter tecnológico-estratégico, a través de las siguientes acciones:

Desarrollo de un estudio de capacidad de la infraestructura TIC requerida

Con ello se quiere que los tomadores de decisión tomen conciencia de la necesidad de resguardar, que sus sistemas tengan suficiente capacidad para soportar la demanda y requerimientos de los ciudadanos y empresas Info mediarias. Por ejemplo, se debe evitar la saturación de las máquinas que soportan los servicios Web.

Incorporar progresivamente tecnologías semánticas

Este tipo de tecnologías están hoy día disponibles y valen la pena usarlas aunque al principio se usen sólo en forma experimental para capacitar al personal técnico de la institución. Con ello la institución puede alcanzar más fácilmente los niveles de 4 y 5 estrellas de Tim Berners-Lee. El empleo de estas tecnologías permite además manejar mejor una multiplicidad de catálogos de diferentes fuentes, tanto nacionales como de gobiernos locales, y de otros poderes del estado, fuentes privadas, etc., esto es, si se opta por un modelo distribuido de catálogos. En otras palabras, facilita la interoperabilidad y agregación de datos y catálogos junto a su actualización de fuentes externas.

4.5.1 Acceso

Hoy en día el internet es una gran herramienta para dar a conocer, todo tipo información, a nivel mundial, por lo cual la SNAP ha desarrollado dos portales web, a través de los cuales se dará a conocer toda la información debidamente liberada, será de fácil acceso sin restricciones de ningún tipo, a continuación los dominios:

- datosabiertos.gob.ec
- datosabiertos.gob.ec/catalogo

Los datos deben estar disponibles para el mayor número de usuarios posible, para que puedan participar de la más amplia gama de propósitos.

Por eso es importante la liberación inmediata de datos a través de las actuales páginas web y facilitar la generación de futuros portales concentradores interconectados, interoperables y accesibles de datos y de aplicaciones de utilidad ciudadana, sin alterar las actuales normas de accesibilidad de páginas web debiendo cumplirse toda la norma ya establecidas y para futuras modificaciones de las mismas.

Para determinar un correcto acceso hay que considerar las siguientes definiciones. Acceso engloba las actividades relacionadas con la habilitación de portales, la existencia de Data Sets, etc. Sus niveles de capacidad se determinan mediante las siguientes variables:

- Lectura Automatizada de los Datos (30%). La información es accesible a través de lenguaje de base de datos u otras formas de acceso.
- Metadatos (30%). Existencia de metadatos y tipo de los mismos.
- Facilidades de Categorización y Descubrimiento (20%). Existencia en el portal Web de facilidades de búsqueda y de creación y mantenimiento de taxonomías.
- Empleo de Tecnologías Semánticas (20%). El portal Web cuenta con soporte para el enriquecimiento semántico de los conjuntos de datos.

Los niveles de capacidad de las variables Lectura Automatizada de los Datos, Metadatos, Facilidades de Categorización, Descubrimiento, y Empleo de Tecnologías Semánticas, se identifican de acuerdo a la siguiente descripción:

Lectura Automatizada de los Datos

- Capacidades Inexistentes: no existen sistemas automatizados de lectura de información.
- Capacidades Incipientes (informal): existen iniciativas para crear sistemas automatizados de lectura de la información.
- Capacidades Existentes: un porcentaje de los datos pueden ser leídos de manera automatizada.
- Capacidades Avanzadas: todos los datos pueden ser leídos de manera automatizada.

Variable Metadatos

- Capacidades Inexistentes: se desconoce la existencia de metadatos o no dispone de los mismos.
- Capacidades Incipientes (informal): para algunos conjuntos de datos existen metadatos de tipo básico (tales como información del publicador, o fecha de actualización).
- Capacidades Existentes: todos los conjuntos de datos cuentan con metadatos de tipo básico. Algunos datos cuentan con metadatos de tipo complementario (tales como rango temporal de los datos).
- Capacidades Avanzadas: Todos los conjuntos de datos cuentan con metadatos de tipo básico, muchos datos cuentan con metadatos de tipo complementario, también algunos datos cuentan con metadatos de tipo extendido (tales como detalle temporal de los datos).

Variable Facilidades de Categorización y Descubrimiento

- Capacidades Inexistentes: No existen facilidades de búsqueda ni soporte de taxonomías.
- Capacidades Incipientes (informal): Buscador sintáctico simple.
- Capacidades Existentes: Buscador avanzado. Categorización mediante etiquetas
- Capacidades Avanzadas: Buscador semántico facetado. CMS incluye soporte nativo de taxonomías.

Variable Empleo de Tecnologías Semánticas

- Capacidades Inexistentes: no se emplea tecnología semántica alguna.
- Capacidades Incipientes (informal): soporte para publicar HTML enriquecido mediante RDF (atributos)/micro datos.
- Capacidades Existentes: uso de agente software para el volcado de datos estructurados a archivos RDF estáticos.
- Capacidades Avanzadas: Empleo de servidor semántico nativo Linked-Data + punto de consulta SPARQL.

4.5.2 Calidad de los datos

La calidad de los datos engloba las actividades relacionadas con mantener y gestionar la calidad de los datos de la organización, en términos de las definiciones existentes de Datos Abiertos y el grado de interoperabilidad de los datos de la organización.

Sus niveles de capacidad se determinan a través de estas variables:

- Formato de los Datos (30%): los datos deben estar en formatos que faciliten la interoperabilidad y no sean propietarios.
- Datos Libres (25%): los datos no están restringidos por propiedad intelectual u otras restricciones legales que puedan limitar su reutilización.
- Datos Primarios (25%): Los datos deben ser primarios, evitando datos secundarios procesados.
- Completitud de los Datos (20%): los datos abiertos deben estar completos (Todos los aspectos del dato están).

Los niveles de capacidad de las variables Formato de los Datos, Datos Libres, Datos Primarios, y Completitud de los Datos, se identifican de acuerdo a la siguiente descripción:

Variable Formato de los Datos

- Capacidades Inexistentes: la mayoría de datos se ofrecen en formatos propietarios y/o no estructurados (tipo JPEG, PDF imagen, etc.).
- Capacidades Incipientes (informal): muchos datos se ofrecen en formatos propietarios, si bien la mayoría son formatos estructurados (tipo CSV, HTML, etc.).
- Capacidades Existentes: La mayoría de datos se ofrecen en formatos no propietarios, además de formatos interoperables (tipo XML, JSON, RDF).
- Capacidades Avanzadas: los datos se ofrecen en formatos no propietarios, reconocidos por su capacidad de integración. Los formatos interoperables escogidos, están coordinados con los formatos escogidos por otras organizaciones.

Variable Datos Libres

- Capacidades Inexistentes: la mayoría de los datos contienen restricciones que limitan su uso y/o hay desconocimiento al respecto.
- Capacidades Incipientes (informal): la entidad tiene conciencia de las restricciones existentes en la mayoría de sus datos.
- Capacidades Existentes: existe un grupo de datos restringidos en su uso.
- Capacidades Avanzadas: los datos no tienen restricciones sobre su uso.

Variable Datos Primarios

- Capacidades Inexistentes: todos los datos son secundarios.
- Capacidades Incipientes (informal): existe un conjunto de datos primarios.
- Capacidades Existentes: la mayoría de los datos son primarios.
- Capacidades Avanzadas: todos los datos son primarios.

Variable Completitud de los Datos

- Capacidades Inexistentes: los datos no están completos.
- Capacidades Incipientes (informal): existen iniciativas que buscan completar los datos abiertos.
- Capacidades Existentes: los datos están completos en su mayoría.
- Capacidades Avanzadas: Los datos están completos en su totalidad.

4.6 Aplicaciones a generar

Se basará en toda la información que se puede liberar según la LOTAIP y las primeras hojas de rutas entregadas por las entidades públicas a la SNAP, se tendrá un

panorama más claro, de las posibles aplicaciones que podrían llegarse a generar, para aportar con la innovación y mejoramiento de los servicios públicos que se ofrece a la ciudadanía en general, a continuación una lista:

- **Visualización presupuesto general del estado:** la aplicación que se podrá generar "mis impuestos" se podría fundamentar sobre 2 pilas fundamentales: el primero sería hacer un parser que recorre las páginas de las Entidades Públicas relacionadas con el presupuesto del estado, extraiga la información que se requiera tratar y la convierta en reutilizable. Sobre esta etapa trabajara el segundo pilar del proyecto, creando una web que lee esos presupuestos en formato abierto y los visualice.
- **Visualización eventos culturales:** Sera una plataforma común, abierta y compartida con todos los actores de nuestra cultura, para la difusión y posible comercialización tanto de sus actividades como de aquellos contenidos de valor que se adecuen a los nuevos canales de interacción y comunicación. la aplicación visualización de eventos se la podría desarrollar en Google Maps, creando controles con Javascript y jQuery los cuales permitirán seleccionar un día para geolocalizar de eventos culturales en el sector. Por defecto, se cargara el mapa con todos los eventos del día de hoy presente. Cada vez que el usuario cambia el día se recarga el mapa.
- **Biblioteca Digital:** Este proyecto de bibliotecas digitales podrá ser capaz de generar una galería de documentos PDF con sólo indicar la ruta XML del documento descriptivo, para ser posible las bibliotecas del país tendrán que tener repositorios con los libros digitalizados, Este proyecto, se lo podrá realizar en php y mysql, para hacer uso del XML facilitado en la página web de datos abiertos de la SNAP sobre las bibliotecas digitales en el país.

- **Tramites en Línea:** La página web de trámites en line pondría a disposición del ciudadano un buscador de trámites de las diferentes administraciones públicas a escala nacional. El visitante puede buscar la información deseada introduciendo directamente el nombre del dato o filtrar a través de varios campos de búsqueda disponibles. Para poder acceder al listado de datos, el usuario deberá visitar la pestaña Trámites que se ubicara en el menú superior y entrar en la sección *Encuentra el trámite*. Una vez dentro, será posible clasificar la información gracias a las varias categorías existentes o segmentarla según la administración pública, ministerio, entidad Autónoma o según las etiquetas de información.
- **Plataforma de Estadísticas:** crear una plataforma web que recopila estadísticas de diferente información pública, y ponerla a disposición del público de manera homogénea, accesible y usable. ¿Alguna vez se ha querido localizar algún dato estadístico y no ha sabido donde buscarlo? ¿Se ha querido visualizar adecuadamente datos estadísticos pero el organismo que los provee no tiene herramientas de visualización adecuadas? Para poder dar respuesta a estas cuestiones, crear aplicación de tipo buscador desde el cual se pueden realizar consultas textuales básicas o avanzadas, gracias a un conjunto de filtros que permitan seleccionar características concretas y obtener, así, los subconjuntos de estadísticas más adecuados a cada usuario -con sus correspondientes metadatos-. Los resultados se visualizarían bien en tablas modulables, bien en gráficos de barras, secciones y líneas.

A medida que la información se siga liberando se irán, descubriendo nuevas necesidades que permitan, generar más aplicaciones, y de esta manera aportar con la innovación y automatización del País.

5 CAPÍTULO 5.

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

A partir del siguiente trabajo se han obtenido las siguientes conclusiones:

- Las Administraciones Públicas del Ecuador están sujetas a la ley de transparencia para la publicación de datos, las cuales obligan a la liberación de información pero de manera incorrecta, actualmente no hay normativas para la publicación de información que permitan liberar información de manera estándar, que permitan fomentar los Datos Abiertos en el país
- El nivel de madurez del Ecuador con respecto a los Datos Abiertos, se encuentra en una fase inicial donde las instituciones públicas están dando los primeros pasos, gracias a que la Secretaria de la Administración Pública, está publicando sobre la importancia de los Datos Abiertos, además de los esfuerzo individuales del municipio de Quito con su portal de Datos Abierto.
- En presente trabajo se planteó una metodología para la publicación de Datos Abiertos Enlazados que de soporte a todos los contextos, orientada a la problemática de las administraciones públicas que es posible replicar y extender a otros dominios.
- El talento humano del área informática y afines a tecnología, de las instituciones públicas están muy poco informados y en alguno de los casos no han oído hablar sobre tema, lo cual ha contribuido para que la cultura de los Datos Abiertos a nivel de las instituciones públicas no se desarrolle.
- Se proporcionó varias herramientas para visualizar los Datos Abiertos Enlazados como las mencionadas en esta investigación que aunque están hechas para propósitos específicos, permiten su uso en múltiples contextos.
- El desconocimiento de sobre los Datos Abiertos y su importancia a causado que sector público y privado aún no hayan visto el real potencial de este tipo de tecnologías, por lo que en ese sentido se tiene mucho trabajo por realizar.

- Las tecnologías de los Datos Abiertos son muy conocidas y tienen una antigüedad suficiente como para haber sido adoptados con anterioridad. En este sentido, se puede añadir que la adopción de estas tecnologías, además de factores económicos, se debe al desconocimiento del tema por parte de la mayor parte de quienes desarrollan en la Web. Desde un punto de vista, este trabajo viene a dar solución en parte a tal brecha.

5.2 Recomendaciones

En base a todos los estudios realizados, se obtenido las siguientes recomendaciones:

- Se recomienda implementar una serie de charlas y capacitaciones para todo el talento humano de las diferentes áreas que conforman la institución, sobre temas que abarquen que son los Datos Abiertos y su importancia
- Se recomienda que para las Universidades se deberían implementar materias que abarquen temas de web semántica y Datos Abiertos, además de que deberían incluir en sus seminarios de ciencia y tecnología charlas que den a conocer al estudiante sobre el uso de estas tecnologías y la vinculación con los Datos Abiertos.
- Se recomienda hacer un análisis de las tecnologías que se están utilizando en las instituciones del sector público para que tengan claro el estado actual en el que se encuentran y de esta manera facilitar el cambio que implica publicar Datos Abiertos Enlazados.
- Involucrar más a la ciudadanía en los temas de Datos Abiertos, se debe socializar más el tema ya estos serán los que consuman los servicios derivados que produce publicar datos, generando de esta manera un ingreso económico para las instituciones públicas.
- Apoyar e invertir en grupos de desarrollo de software en especial de aplicaciones móviles, los cuales están muy ligados con el tema de los Datos Abiertos.
- Impulsar los Hackaton, estos son eventos donde se reúnen a compartir científicos sociales y programadores para compartir y crear aplicaciones que

ayuden a mejorar la calidad institucional para transparentar y publicitar el gasto público y la gestión en los distintos niveles de gobierno

5.3 Trabajos futuros

Dada la presente investigación, algunos de los posibles trabajos futuros se describen a continuación:

Uno de los trabajos futuros es ampliar la propuesta metodológica descrita a un contexto más general, de forma que se consideren distintos escenarios de aplicación, mediante una arquitectura adaptable, distintos esquemas de componentes y fases de implantación dinámicas. También se podría mejorar incorporando nuevas características y estudios comparativos con herramientas similares existentes en el mercado.

Implementación de perfiles de usuario para consulta a grafos de Datos Abiertos Enlazados, con el fin de implementar seguridad a partes de un grafo RDF, un área no explorada por ninguna solución de software actual que soluciona problemáticas reales a las cuales se ven enfrentados los actuales desarrollos. Diseñar e implementar herramientas de monitoreo de servicios de soporte a Datos Abiertos Enlazados, ya que ocurre en muchos casos que producto de la “inmadurez” de las tecnologías semánticas, los servicios se encuentran caídos, impidiendo su uso.

BIBLIOGRAFÍA

Tim Berners-Lee. Universal resource identifiers – axioms of web architecture.

<http://www.w3.org/DesignIssues/Axioms.html>.

Tim Berners-Lee. Cool uris don't change.

<http://www.w3.org/Provider/Style/URI.html>.

Tim Berners-Lee. Notation3 (N3) a readable RDF syntax.

<http://www.w3.org/DesignIssues/Notation3>.

Chris Bizer, Richard Cyganiak, and Tom Heath. How to publish Linked Data on the Web

<http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

Tim Berners-Lee. Linked data - design issues.

<http://www.w3.org/DesignIssues/LinkedData.html>

Ben Adida and Mark Birbeck. RDFa primer.

<http://www.w3.org/TR/xhtml-rdfaprimers/>.

Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data - the story so far. International Journal on Semantic Web and Information Systems, paginas 1–26

Kendall Grant Clark Lee Feigenbaum and Elias Torres. SPARQL protocol for RDF.

<http://www.w3.org/TR/2008/REC-rdf-sparql-protocol-20080115/>

W3C HTML Working Group. XHTML 1.0: The extensible HyperText markup language

<http://www.w3.org/TR/html/>

W3C Working Group. Best practice recipes for publishing RDF vocabularies.

<http://www.w3.org/TR/swbp-vocab-pub/>,

Secretaria de la Administración Pública, Guía Política Pública de Datos Abiertos.

<http://www.gobiernoelectronico.gob.ec/GPP-DA-v01-20141128-SNAP-SGE.pdf>

Vicepresidencia del Ecuador, Ley Orgánica de Transparencia, paginas 3-7.

<http://www.vicepresidencia.gob.ec/wp-No.-337-del-18-de-mayo-de-2004.pdf>

ANEXOS

Anexos A – Glosario

API: Interfaz de programación de aplicaciones (IPA) o API (del inglés Application Programming Interface) es un canal estándar de comunicación entre diferentes componentes o sistemas de software. Referencia: http://es.wikipedia.org/wiki/Interfaz_de_programaci%C3%B3n_de_aplicaciones.

Creative Commons: Licencia que define el espacio entre la protección absoluta de los derechos de autor y el dominio público. Las licencias te ayudan a conservar tus derechos autorales invitando a usar tu obra bajo el esquema “Algunos derechos reservados”. Referencia: <http://www.creativecommons.mx/que/>.

CSV: Formato sencillo y abierto para representar datos en forma de tabla, en las que las columnas se separan por comas. En los casos en los que un valor contenga coma (como el separador decimal en algunos países), se recomienda encerrar entre comillas los valores o utilizar otro separador como “punto y coma” o “tabulador”. Referencia: http://es.wikipedia.org/wiki/Comma-separated_values.

DCAT: Vocabulario diseñado por la W3C para facilitar la interoperabilidad entre catálogos de datos publicados en la Web. Al usar DCAT para describir conjunto de datos en catálogos, los publicadores incrementan la facilidad de descubrimiento y manipulación de estos catálogos. <http://www.w3.org/TR/vocab-dcat/>.

GeoJSON: Formato nuevo basado en JSON para codificar una variedad de estructuras de datos geográficos. Bajo este formato se pueden representar objetos geométricos, una cualidad o una colección de éstas. Los tipos geométricos soportados son puntos, líneas, polígonos, multipunto, entre otros. Referencia: <http://geojson.org>.

ISO 8601: La norma ISO 8601 especifica la notación estándar utilizada para representar instantes, intervalos e intervalos recurrentes de tiempo evitando ambigüedades. Se sigue el criterio de especificar en orden primeramente los períodos de tiempo más largos y posteriormente los más cortos. Por ejemplo, para especificar la fecha 30 de mayo de 2008, escribiremos en esta notación: 2008-05-30. Referencia: <http://es.wikipedia.org/wiki/ISO8601>.

ISO 8859-1 (latin-1): Norma ISO que define la codificación de caracteres, cubriendo los caracteres especiales del español (<http://es.wikipedia.org/wiki/ISO8859-1>). Esta guía recomienda el estándar UTF-8 ya que, además de incluirlo los caracteres de esta norma (<http://es.wikipedia.org/wiki/ISO8859-1>), extiende a un mayor rango y se ha convertido en el estándar de mayor utilización en la Web (<http://googleblog.blogspot.ca/2012/02/unicode-over-60-percent-of-web.html>).

JSON: Formato ligero (y legible) para el intercambio de datos. Dada su simplicidad de estructura, se ha generalizado su uso en la Web. Referencia: <http://es.wikipedia.org/wiki/JSON>.

KML: Lenguaje de marcado basado en XML para representar datos geográficos en tres dimensiones, característico de la herramienta Google Earth. Un archivo KML describe elementos, como lugares, imágenes, o polígonos, que además contienen título, descripción, coordenadas geográficas y alguna otra información. Desde 2008 es un estándar en la industria y es aprobado por la OGC. Referencias: http://es.wikipedia.org/wiki/Keyhole_Markup_Language, https://developers.google.com/kml/documentation/kml_tut, <http://www.opengeospatial.org/standards/kml>.

Metadatos: (del griego *μετα*, meta, ‘después de, más allá de’ y latín datum, ‘lo que se da’, «dato»), literalmente «sobre datos», son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado recurso. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos. <http://es.wikipedia.org/wiki/Metadata>.

ODbL (Open Database License): Licencia orientada a bases de datos, la cual permite copiar, distribuir, remezclar, modificar y crear a partir de dicha obra, incluso con fines comerciales, siempre y cuando se dé crédito a la fuente creadora y se mantenga la misma licencia para cualquier obra derivada. <http://opendatacommons.org/licenses/odbl>.

ODI: El Open Data Institute es una compañía independiente, sin fines de lucro, fundada por Sir Tim Berners-Lee y el Profesor Nigel Shadbolt con el objetivo de catalizar la cultura de datos abiertos para crear valor económico y social. <http://theodi.org/about>.

Office Open XML: También llamado OOXML, es un formato de archivo abierto y estándar, cuyas extensiones más comunes son .docx, .xlsx, y .pptx. Se le utiliza para representar y almacenar hojas de cálculo, gráficas, presentaciones y documentos de texto. Un archivo Office Open XML contiene principalmente datos basados en el lenguaje de marcado XML, comprimidos en un contenedor .zip específico. http://es.wikipedia.org/wiki/Office_Open_XML.

OGC: El Open Geospatial Consortium es un consorcio internacional, con más de 400 compañías, agencias de gobierno y universidades, participando en los procesos para desarrollar estándares públicos, particularmente enfocados en la información geoespacial. <http://www.opengeospatial.org/ogc>.

PDF: El formato de documento portátil, inventado por Adobe Systems, es un estándar abierto para el intercambio de documentos electrónicos que mantiene la

Organización Internacional de Normalización (ISO).
<http://www.adobe.com/es/products/acrobat/adobepdf.html>.

RDF: Formato de la W3C diseñado para generar un modelo de datos que describa la información de un recurso web. Este mecanismo es un componente clave de la Web Semántica.
http://es.wikipedia.org/wiki/Resource_Description_Framework.

REST: Conjunto de principios de arquitectura de software para sistemas distribuidos como la World Wide Web. En la actualidad se utiliza el término para describir cualquier interfaz web simple sin las abstracciones adicionales de los protocolos basados en patrones de intercambio de mensajes, como el protocolo de servicios web “SOAP”.
http://es.wikipedia.org/wiki/Representational_State_Transfer.

Servicio web: Un servicio web (en inglés, Web Service o Web services) es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. http://es.wikipedia.org/wiki/Servicio_web.

SHP: Formato de datos geospaciales propietario de la empresa ESRI. Dada su popularidad, su especificación técnica es pública y hasta cierto punto abierto. Un shapefile (como se denominan a los archivos de este formato) es un formato vectorial que guarda la localización de elementos geográficos y los atributos asociados a ellos. El formato es multiarchivo, es decir, se genera en conjunto con archivos .shx y .dbf, los cuales definen (respectivamente) el índice de las entidades geométricas y la base de datos de los atributos de los objetos. <http://es.wikipedia.org/wiki/Shapefile>.

SOAP: SOAP (siglas de Simple Object Access Protocol) es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos XML. Es uno de los protocolos utilizados en los servicios Web. http://es.wikipedia.org/wiki/Simple_Object_Access_Protocol.

URL: Un localizador de recursos uniforme, más comúnmente denominado URL (sigla en inglés de uniform resource locator), es una secuencia de caracteres, de acuerdo a un formato modélico y estándar, que se usa para nombrar recursos en Internet para su localización o identificación, como por ejemplo documentos textuales, imágenes, vídeos, presentaciones digitales, etc. http://es.wikipedia.org/wiki/Localizador_uniforme_de_recursos.

UTF-8: UTF-8 (8-bit Unicode Transformation Format) es un formato de codificación de caracteres Unicode e ISO 10646 (<http://es.wikipedia.org/wiki/UTF-8>). Este formato es el más utilizado en la Web (http://w3techs.com/technologies/overview/character_encoding/all).

W3C: El World Wide Web Consortium, abreviado W3C, es un consorcio internacional que produce recomendaciones para la World Wide Web. Dirigido por Tim Berners-Lee, conocido como el padre de la Web. Referencia: http://es.wikipedia.org/wiki/World_Wide_Web_Consortium.

XML: Lenguaje de marcado desarrollado por la W3C para almacenar datos en forma legible y estructurada, que propone un estándar para el intercambio de información entre diferentes plataformas. Referencia: <http://es.wikipedia.org/wiki/XML>.

Anexos B - Sistemas Informáticos y Portales Existentes

Sistemas informáticos y Portales Existentes y su contribución a las estrategias de los Datos Abiertos. Se lista los sistemas informáticos y portales existentes que se identifican como servicios informáticos en los portales institucionales de las entidades de la Administración Pública Central Institucional y dependiente de la función ejecutiva, estos sistemas fueron formulados y ejecutados antes de existir la Guía para la liberación de información

Tabla de Anexos 1

Sistemas y Portales

Nombre Solución Existente	Acceso Centralizado	Documentos electrónicos	Autenticación única	Interoperable	Esquema de datos abiertos	Contenidos de capacitación	Derechos y Patentes del Estado	Disponibilidad en la Nube	Mecanismos de participación	Mecanismos de evaluación	Accesibilidad y Usabilidad	Cobertura de Estrategias
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	Total
Gestión Documental -Quipux		x	x			x	x	x				5
Gobierno por Resultados (GPR)				x		x		x				3
Portales Institucionales Homologados							x	x				2
Sistema de Inteligencia de Mercados				x				x				2
Sistema Único de Información Ambiental (SUIA)				x				x				2
Ventanilla única de Exportación (VUE)				x				x				2
Servicio de Turnos para Licencias de Conducción				x				x				2
Sistema de Consulta en Línea de Multas por infracciones de Tránsito (ANT)				x				x				2
Sistema de Emisión de licencias de conducir y consultas en línea para Discapacitados (ANT)				x				x				2
Sistema Integral de Información de Desarrollo Urbano y Vivienda (SIIDUVI)				x				x				2

Tabla de Anexos 2-

Portales y sistemas en ejecución

Nombre Solución Existente	Acceso Centralizado	Documentos electrónicos	Autenticación única	Interoperable	Esquema de datos abiertos	Contenidos de capacitación	Derechos y Patentes del Estado	Disponibilidad en la Nube	Mecanismos de participación	Mecanismos de evaluación	Accesibilidad y Usabilidad	Cobertura de Estrategias
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	Total
Sistema de Información del Ministerio de Educación (SIME)				x				x				2
Sistema de Incentivos para Vivienda para Personas Migrantes				x				x				2
Registro de comercializadores de bienes de producción nacional (SBPN)				x				x				2
Sistema de Registro Único del Ministerio de Industrias y Productividad.				x				x				2
Registro de producción Nacional de Bienes y Servicios (RPN-BS)				x				x				2
Sistema de Registro de Centros de Acopio y Recicladores (SIRCAR)				x				x				2
Sistema de Monitoreo de Importaciones del Ministerio de Industria y Productividad (SMIMIP)				x				x				2
Sistema de Registro de Reencauchadoras				x				x				2
Consulta de Antecedentes Penales				x				x				2
Sistema para Reclutamiento en Línea para aspirantes a Policía Nacional				x				x				2

Listado de los proyectos, que poseen las entidades públicas, para dar solución a la problemática de la publicación de información, solicitada por la ley de transparencia.

Tabla de Anexos 3

Proyectos a Ejecutar

Proyecto	Acceso Centralizado	Documentos electrónicos	Autenticación única	Interoperable	Esquema de datos abiertos	Contenidos de capacitación	Derechos y Patentes del Estado	Disponibilidad en la Nube	Mecanismos de participación	Mecanismos de evaluación	Accesibilidad y Usabilidad	Cobertura de Estrategias
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	Total
Si Salud (Segunda fase)	x	x	x	x	x	x		x		x	x	9
Nuevo E-Sigef	x	x	x	x	x	x		x		x	x	9
Nuevo sistema de Pasaportes	x		x	x	x	x		x	x	x	x	9
Portal Gestión de Comunidades	x		x		x	x	x	x	x	x	x	9
Portal Versionamiento de codificación	x		x	x		x	x	x	x	x	x	9
Ventanilla Única Virtual (ciudadanía, empresas)	x		x	x		x	x	x	x	x		8
Gestión Educativa	x	x	x	x		x		x		x	x	8
Software Policía y Migración	x	x	x	x		x		x		x	x	8
Gestión Documental	x		x		x	x	x	x		x	x	8
Levantamiento, diseño y automatización del proceso de reforma institucional de la administración pública central, institucional y dependiente	x	x	x	x		x	x	x			x	8
Consulado Virtual		x	x	x		x	x	x			x	7
Portal Ecuador	x			x		x	x	x	x		x	7
Bus Gubernamental de Datos: autenticación y servicios web (segunda fase)	x		x	x		x	x	x		x		7

BIOGRAFÍA



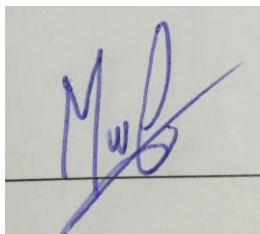
EL SR. Marlon Ricardo Cueva Peñaranda, nace en la ciudad de Quito, el 29 de Diciembre del 1989. Cursó sus estudios escolares en la Unidad Educativa Manuela Cañizares, inicio y culmino sus estudios de bachiller en el Unidad Educativa Josué graduándose en el año 2007 y obteniendo el título de Bachiller de Ciencias.

En el 2007 ingresa en la Carrera de Ingeniería en Sistemas e Informática de la Escuela Politécnica del Ejército y finaliza sus estudios superiores en Diciembre del 2013.

A principios del año 2013 empieza a trabajar brindando sus conocimientos adquiridos de la universidad en la empresa privada SISTELNET, especializada en redes, telecomunicaciones e Infraestructura, para desempeñar el trabajo de soporte y proyectos, Actual trabaja en ANF AC Ecuador donde está a cargo de la administración del área de Infraestructura.

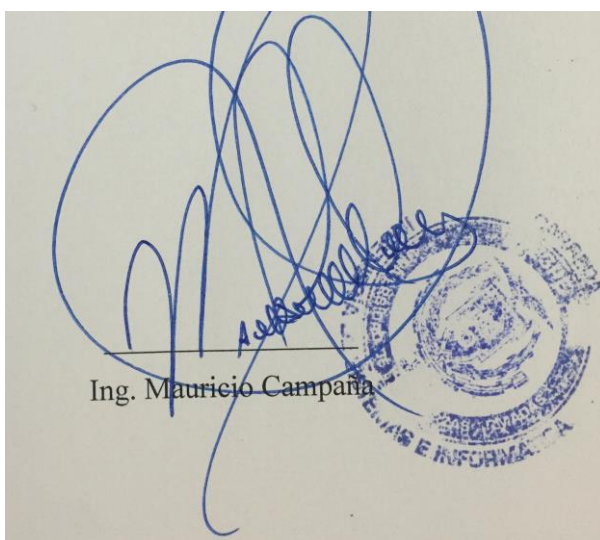
HOJA DE LEGALIZACIÓN DE FIRMAS

ELABORADA POR:



Marlon Cueva

DIRECTOR DE LA CARRERA



Ing. Mauricio Campana

Lugar y Fecha: Sangolquí, Octubre del 2015