**ESCUELA POLITÉCNICA DEL EJÉRCITO**

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN**

**CARRERA DE INGENIERÍA EN SISTEMAS**

****



**El Impacto de los Datos Abiertos Enlazados a Nivel del Ecuador Enfocado en la Innovación de Servicios Públicos**

**CASO DE ESTUDIO:** DIRECCION DE INNOVACION – SECRETARÍA NACIONAL DE LA ADMINISTRACIÓN PÚBLICA



MARLON CUEVA

**18 de junio de 2014**

Tabla de contenido

[I. ANTECEDENTES 2](#_Toc391908429)

[II. PLANTEAMIENTO DEL PROBLEMA 3](#_Toc391908430)

[III. JUSTIFICACIÓN 3](#_Toc391908431)

[IV. OBJETIVOS 4](#_Toc391908432)

[V. ALCANCE 5](#_Toc391908433)

[VI. MARCO TEÓRICO 5](#_Toc391908434)

[VII. METODOLOGÍA DE DESARROLLO DEL PROYECTO 5](#_Toc391908435)

[VIII. FACTIBILIDAD 5](#_Toc391908436)

[IX. TEMARIO 6](#_Toc391908437)

[X. CRONOGRAMA (10 meses) 6](#_Toc391908438)

[XI. Bibliografía 7](#_Toc391908439)

1. ANTECEDENTES

Actualmente en el Ecuador los ministerios y demás entidades públicas procesan grandes cantidades de información, la cual no está siendo utilizada, esta información es muy valiosa y de gran interés para el sector público y privado, hasta la fecha el gobierno solo ha puesto en marcha “El Plan de Gobierno Electrónico” el cual constituye una guía en relación a la automatización del Estado donde se encuentra involucrados diferentes actores (Ciudadanos, Gobierno, Sector Productivo, Servidores Públicos).

La Secretaría Nacional de la Administración Pública, en su afán de consolidar una gestión pública de excelencia, plasma sus esfuerzos y trabajo cotidiano para coordinar, ejecutar e impulsar las acciones de los distintos actores responsables, pero carece de un estudio que le permita identificar qué información es relevante, además de proporcionar un estándar a seguir por los instituciones públicas para una correcta liberación de información.

Existe mucha desorganización en ciertas instituciones públicas, ya que se conocen varios casos en los cuales han tratado de liberar información, sin regirse a las normas y estándares internacionales los cuales el Gobierno Electrónico debe implementar, causando que la información que no se esté liberando no posea la característica de ser reutilizada, por las demás instituciones públicas involucradas.

El sector público está generando grandes cantidades de información, la cual simplemente está siendo almacenada o incluso en el peor de los casos la información simplemente se pierde ya que generan un costo significativo de mantenimiento, almacenamiento e infraestructura, peor es aquí donde debemos tomar la iniciativa para poder identificar toda la información de valor para la generación de nuevas micro y macro 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. PLANTEAMIENTO DEL PROBLEMA

En las instituciones Publicas actualmente se está manejando todo tipo de información ya se está digital o física en grandes cantidades lo cual no es un impedimento en cuanto a procesamiento, ya que la tecnología de hoy en día permite que se pueda manejar todo tipo información, al nivel de complejidad que se esté presentando,

Uno de los problemas más grave que se presenta es la pérdida de información, que al no tener los recursos necesarios se la debe dejar perder, de igual manera al final terminan generando grandes pérdidas para el Estado.

La información en el Ecuador no ha sido tomada con la importancia que se merece, en mucho de los casos se la menosprecia, sin tomar en cuenta que esta es un gran puntal de desarrollo, crecimiento económico e intelectual, pero para poder aprovechar este gran recurso primero se debe investigar de los diferentes tipos de información que hay en el medio, los estándares y tecnologías correctos para tener una información fácilmente aprovechable por cualquier individuo o institución.

1. JUSTIFICACIÓN

1)         Conveniencia

La investigación pretende generar un impacto social que logre satisfacer las necesidades tanto para las instituciones públicas como a personas naturales y a la empresa privada que logre un aporte significativo, ayudando a construir las bases para la generación de nuevos proyectos o servicios, los cuales a su vez puedan generar una cultura de innovación a nivel de todo el Ecuador, aportando con el Plan de Desarrollo Nacional.

2)         Relevancia social

Este proyecto está enfocado en el desarrollo de una cultura de innovación e investigación, construyendo las bases para la generación de nuevos proyectos tecnológicos que pretendan dar a conocer que en el Ecuador se pueden generar nuevas tecnologías, brindando a la sociedad nuevos servicios más eficientes y rápidos.

3)         Implicaciones prácticas

El núcleo de la investigación y de su implementación es generar un impacto, motivando a las empresas públicas privadas, universidades y demás, a la generación de nuevas tecnologías al servicio de la comunidad y del Ecuador dando como resultado el surgimiento de nuevas micro y macro economías, aportando de manera considerable con la economía y desarrollo del país.

1. OBJETIVOS

**Objetivo General**

* Investigar cuales son los datos relevantes y necesarios que las instituciones pertenecientes al sector público debe liberar, para así fomentar la generación de nuevas aplicaciones o proyectos que puedan reutilizar dicha información, y que a su vez puedan generar un impacto social y económico en el Ecuador.

**Objetivos Específicos**

* Investigar cual es el estado del arte de los datos abiertos en el Ecuador, el cual nos permita identificar los pro y contras de las propuestas arquitectónicas actuales relacionadas a la publicación de datos enlazados abiertos.
* Definir una arquitectura estandarizada que de soporte en la implantación de datos enlazados en el contexto de la administración pública.
* Definir tecnologías y métodos a implementarse entre instituciones públicas para obtener un gobierno electrónico homogéneo.
* Dar a conocer los servicios y posibles aplicaciones que se podrán generar a partir de toda la información ya liberada correctamente.

1. ALCANCE

La investigación pretende elaborar un estudio que sirva como guía a la Secretaria Nacional de la Administración Publica en la identificación de los datos relevantes existentes en las instituciones públicas además de definir un método para publicar los datos liberados de manera estándar, fomentando un gobierno de libre información, basado en los valores de servicio, eficiencia y participación, donde se puedan involucrar los ciudadanos y las instituciones públicas generando un beneficio mutuo. Este estudio les permitirá considerar el aprovechamiento de la información pública, orientándola a la mejora de los servicios públicos potenciando y mejorando los sistemas existentes, migrando a tecnologías de multiplataforma para que de esta manera se fomente la interoperabilidad de los datos públicos entre los servicios del sector público.

Se quiere lograr el aprovechamiento y redistribución de la información pública, para así de esta manera dar a conocer las aplicaciones que se deberán generar para mejorar los servicios que ofrece el sector público o de igual manera implementar nuevos servicios que se hayan pasado por alto, obteniendo como resultado: nuevas economías, emprendimiento, innovación.

La investigación también promoverá la cultura del aprovechamiento de la información relacionada con las Instituciones Públicas, concientizando en la importancia y su respectiva valoración, para un uso continuo a futuro, principalmente enfocado a la generación de conocimiento. Además, se pretende facilitar el libre acceso y disposición de toda la información posible para fomentar el mercado de la explotación de la información de manera correcta regida a estándares y normas que lo controlen.

1. MARCO TEÓRICO

Se presenta a continuación los conceptos fundamentales sobre los datos abiertos enlazados.

**Conceptos relacionados con los datos abiertos enlazados**

**Open Data**

Open Data o datos abiertos son todos aquellos que pueden ser utilizados, redituables y redistribuidos libremente por cualquier empresa o persona los se encuentran sujetos, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.

Características más importantes:

* **Disponibilidad y Acceso:** Los datos siempre deben ser disponibles en su totalidad con un precio razonable de reproducción, 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 ser redituables 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 de uso para ciertos propósitos (por ejemplo sólo para educación) no son permitidos.

**URI (Universal Resource Identifiers):** conjunto de identificadores globales que pueden ser creados y administrados en forma distribuida para identificar documentos.

**HTTP (Hyper Text Transfer Protocol):** protocolo para intercambiar datos en la Web cuyas funcionalidades básicas son poner datos (put) y obtener datos (get) desde el espacio imaginario abstracto.

**HTML (Hyper Text Markup Language):** lenguaje para representar información orientada a elementos visuales para presentarla visualmente a humanos.

**XML (eXtensible Markup Language)**: es una segunda generación del lenguaje HTML, permite definir la estructura de los documentos con mayor precisión, representando los documentos como árboles y agregando reglas que permiten establecer restricciones en la anidación de los elementos.

**Datos Abiertos Enlazados**

Los datos abiertos están expresados a través de las normas de la web semántica. Se definen como datos abiertos enlazados o vinculados, con la expresión web semántica la cual fue creada e introducida por Tim Berners-Lee, se entiende la transformación de la web desde una red de documentos, comprensibles solo por el usuario final, a una red de datos, comprensibles por los dispositivos conectados, esta es una nueva web, que también es conocida Web of Data, es aquí donde surge la gran utilidad de los datos enlazados, ya que se explican las relaciones entre los datos dispersos en toda la red, esto permite armar una red totalmente funcional de datos donde los dispositivos además de interpretar de manera mecánica el significados de los datos, logran aportar un nuevo conocimiento, a partir de esa información correctamente interpretada

**Web Semántica**

Es la web de datos procesables por máquinas, que utiliza un modelo de datos que se conoce como Resource Description Framework (RDF) el cual fue diseñado para facilitar el tratamiento automático de la información en la web por medio de metadatos. Es de gran utilidad cuando la información necesita ser procesada por aplicaciones, en vez de sólo ser desplegada para el público en común.

**RDF (Marco de Descripción de Recursos)**

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

**Conjunto de Datos**

Un conjunto de datos corresponde a los contenidos de una única [tabla de base de datos](http://es.wikipedia.org/wiki/Tabla_(base_de_datos)), o una única matriz de datos estadística, donde cada [columna](http://es.wikipedia.org/wiki/Columna_(base_de_datos)) de la tabla representa una variable en particular, y cada [fila](http://es.wikipedia.org/wiki/Registro_(base_de_datos)) representa a un miembro determinado del conjunto de datos en cuestión. Un conjunto de datos contiene los valores para cada una de las variables, como podrían ser la altura y el peso de un objeto, que corresponden a cada miembro del conjunto de datos. Cada uno de estos valores se conoce con el nombre de [dato](http://es.wikipedia.org/wiki/Dato). El conjunto de datos puede incluir datos para uno o más miembros en función de su número de filas

**Catálogos de Datos**

Los catálogos de datos son aquellos que hacen accesible y dan referencias exactas a los conjuntos de datos, pero poseen inconvenientes a la hora de integrar datos, el modelo RDF proporciona las herramientas necesarias para integrar lógicamente los datos y consultas.

Los catálogos también pueden ser definidos como un portal web con una serie de entradas describiendo conjuntos de datos (datasets), todas estas descripciones suelen incluir metadatos (nombre, descripción, materias tratadas, origen, fecha de publicación, licencias de uso, etc.), para los cuales son fundamental todas las referencias, para poder acceder a los datos.

Los catálogos de datos abiertos son accesibles a través de la web por medio de documentos descargables (datos muertos) o aplicaciones que realizan consultas de datos en línea (datos vivos)

1. METODOLOGÍA DE DESARROLLO DEL PROYECTO

Para desarrollar esta investigación se va a realizar un estado del arte recabado en todo lo que respecta a metodologías de publicación de datos enlazados, teniendo en cuenta que se tratan de temas muy recientes.

Para esto vamos a tener en cuenta lo que recomienda el W3C que en su guía de publicación, manifiesta que se tienen que liberar datos útiles y en formatos reutilizables. Cualquier formato es útil, pero cuanto mejor estructurados y enriquecidos estén los datos, más redituables serán permitiendo construir aplicaciones que los traten automáticamente.

Según la clasificación propuesta por Tim Berners-Lee, los formatos más adecuados son el RDF, basado en XML y facilidad para pasar al siguiente nivel llamado “datos enlazados” (Linked data), que a su vez es la base técnica para alimentar la denominada web semántica, un estándar en el que cada dato contiene información asociada que lo relaciona automáticamente con otros.

También hay que tomar en cuenta que la W3C recomienda crear un catálogo con la descripción precisa de todos y cada uno de los conjuntos de datos (data sets) que libera un gobierno y a su vez libera su propio catálogo en formato RDF.

1. FACTIBILIDAD

* **TECNICA**
* **OPERATIVA**
* **ECONOMICA**

1. TEMARIO
2. CRONOGRAMA (4 meses)

1. Bibliografía