

**GUÍA DE REFERENCIA PARA LA APERTURA DE DATOS ABIERTOS**

Gestión Tecnológica y Comunicaciones

22 de septiembre de 2017

Instituto nacional de formación técnica profesional - infotep

Av. Colombia, Barrio Sarie Bay, San Andrés Islas

**Datos de Contacto:**

|  |  |
| --- | --- |
| **Institución** | INSTITUTO NACIONAL DE FORMACIÓN TECNICA PROFESIONAL - INFOTEP |
| **NIT** | 892400461-5 |
| **Rector** | Silvia Montoya Duffis |
| **Documento preparado por** | * Ing. Juan Camilo Cárdenas Archbold “Versión 2017” |
| **Conmutador** | +57 8 5125770 - +57 8 5126607 |
| **Fax** | +57 8 5121350 |
| **Código DANE** | 88001 – nit: : 892000000-0 |
| **Correo Notificaciones Judiciales** | notificacion@infotepsai.edu.co |
| **Correo contacto y PQRD** | serviciocliente@infotepsai.edu.co |
| **Sitio Web** | www.infotepsai.edu.co |
| **Horario de Atención al Público** | lunes a viernes 8:00 am a 12:00 pm y de 3:00 pm a 7:00 pm |
| **Dirección** | Avenida Colombia, Barrio Sarie Bay. San Andrés Isla, Colombia |

CONTENIDO

[Introducción 4](#_Toc493841563)

[1. Licencia Abierta 6](#_Toc493841564)

[2. Formatos de intercambio de información utilizados. 7](#_Toc493841565)

[3. ¿Cómo uso un formato determinado? 11](#_Toc493841566)

[2.1. Servicios Web 11](#_Toc493841569)

[2.2. Base de datos 11](#_Toc493841570)

[GLOSARIO 12](#_Toc493841571)

[4. Bibliografía 13](#_Toc493841572)

[5. CONTROL DE CAMBIOS 14](#_Toc493841573)

# Introducción

Instituto de Nacional de Formación Técnica Profesional INFOTEP. Es una Institución de Educación Superior, que ofrece formación técnica profesional, programas académicos en extensión orientados bajo los principios y valores institucionales, hacia la formación integral para toda la población del departamento insular y el caribe, buscando el desarrollo social, económico, científico, cultural, tecnológico y ambiental a través de la investigación; generando proyectos en alianza con el estado, sector productivo, los gremios, y otras instituciones, con talento humano idóneo, dando como resultado profesionales integrales, pensantes, emprendedores y formadores de una mejor calidad de vida con el trilingüismo como identidad cultural y proyección social para el Archipiélago.

El compromiso de INFOTEP es generar las mejores condiciones en el ambiente tecnológico institucional, proyectando el cumplimiento de los objetivos trazados por la institución que desde el proceso de Gestión Tecnológica se visualizan como principios de garantía de éxito para la prestación de un servicio optimizado de cara a nuestros clientes internos y externos, con el propósito de proyectar la identidad institucional por medio de soluciones tecnológicas a la vanguardia.

En Colombia la Ley de Transparencia y Acceso a la Información, define los datos abiertos como “todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos”[[1]](#footnote-1)

Por tal motivo INFOTEP adopta esta guía de referencia para datos abiertos con base en las condiciones citadas en la guía de datos abiertos en Colombia en su versión 3, liberada en el junio del 2016. Donde se indica que el objetivo del documento es proporcionar orientaciones y buenas prácticas para el desarrollo de estrategias de apertura y reuso de datos abiertos, que estén orientadas a la generación de valor en lo económico, social, político, cultural, ambiental, y en general, en los distintos ámbitos de la sociedad.[[2]](#footnote-2)

# INFOTEP “PUBLICA” DATOS ABIERTOS

El Instituto Nacional de Formación Técnica Profesional publica Datos Abiertos en dos sitios:

1. Sitio Web: La Entidad desde el año 2015 publica datos abiertos, en el 2017 se rige bajo este documento y la guía de datos abiertos en Colombia:

Se publica en la url: http://infotepsai.edu.co/gestion/transparencia/item/108-datos-abiertos.html

1. Portal de Datos Abiertos: La entidad desde el año 2017 publica datos abiertos en este portal, donde está abierta al público en general y con licencia abierta.

https://www.datos.gov.co/, los datos de la entidad se pueden buscar y filtrar, por INFOTEP, San Andrés, educación

# Licencia Abierta

Es aquella que permite de forma libre y sin restricciones: El uso, Extracción, Adaptación, Transformación, Copia, Distribución, Redistribución, Reutilización, Compilación, Modificación y Difusión.

Tales principios se encuentran en concordancia con la Carta Internacional de Datos Abiertos, lanzada en la Cumbre de la Alianza por el Gobierno Abierto en Octubre de 2015. Puede consultar en el siguiente enlace, más información: http://opendatacharter.net . de los datos publicados.

# Formatos de intercambio de información utilizados.

* Hojas de Cálculo (XLS, XLSX)

Formatos usado de forma general y es generado desde el software Microsoft Excel, es una hoja de cálculo donde se pueden generar desde simples datos hasta potentes cálculos financieros, Muchas personas y empresas mantienen su información en hojas de cálculo, financieras, bases de datos, cálculos, gráficos, etc.

* XML

Extensible Markup Language – Lenguaje de Marcas extensibles, Es un formato ampliamente utilizado para el intercambio de datos debido a que ofrece buenas oportunidades de mantener la estructura de los datos y la forma en que los archivos son construidos, y permite a los desarrolladores escribir partes de la documentación con los datos sin interferir con la lectura de ellos.

* RDF

Es un modelo usado para el intercambio de datos en la Web. Posee características que facilitan la interacción de múltiples datos incluso si los esquemas difieren (lo cual genera interoperabilidad de formatos o esquemas), y apoya específicamente la evolución de esquemas a través del tiempo sin requerir que todos los consumidores de datos sean cambiados.

* Archivos separado por comas

Es un archivo de valores separados por comas. Cuando el texto y los números se guardan en un archivo CSV, es fácil moverlos de un programa a otro. Por ejemplo, puede exportar los contactos de Google al formato CSV y luego importarlos a Outlook.

Los archivos CSV se pueden usar con Excel o con cualquier otro programa de hoja de cálculo. Los archivos CSV se diferencian de otros tipos de archivo de hoja de cálculo en que no puede guardar ningún formato ni puede guardar las fórmulas.

* Documento de Texto

Son documentos de procesamiento de texto y los formatos más comunes son Open Office Writer, Libre Office Writer, Microsoft Word, AbiWord, KWord. Puede ser suficiente para mostrar ciertos tipos de datos, por ejemplo, listas de correo o equivalente. Puede que sea barato para exponerlos, siempre que ese sea el formato en que la data se haya creado. El formato no da soporte para mantener la estructura coherente, lo cual significa que a veces es difícil para introducir datos por medios automatizados. Asegúrese de utilizar las plantillas como base de documentos en las que se mostrarán los datos para su reutilización, por lo que es al menos posible extraer información de los documentos.

* Texto

Documentos de texto (. Txt) son muy fáciles para leer en los ordenadores. Por lo general, no incluyen metadatos estructurales, lo que significa que los desarrolladores necesitan para crear un programa de análisis que pueda interpretar cada documento tal y como aparece.

Algunos problemas pueden ser causados al cambiar archivos de texto plano entre sistemas operativos. MS Windows, Mac OSX y otras variantes de Unix tienen su propia forma de decirle a la computadora que se ha llegado al final de la línea.

* Imagen escaneada

Probablemente la forma menos adecuada para la mayoría de los datos, pero ambos TIFF y JPEG al menos puede marcarlos con la documentación de lo que está en la imagen - adecuada para marcar la imagen de un documento con contenido de texto completo del documento. Puede ser relevante a los datos que muestran como imágenes aquellos datos que no nacen por la vía electrónica -un ejemplo obvio es la antigua escuela de registros y el material de archivo- y una imagen es mejor que nada.

* Formatos propietarios

Algunos sistemas tienen sus propios formatos de datos que pueden guardar o exportar datos. A veces puede ser suficiente para exponer los datos en un formato - sobre todo si se espera que el uso de más estaría en un sistema similar del que vienen. Sobre donde obtener más información sobre estos formatos, debe estar siempre indicado, por ejemplo, al proporcionar un enlace a la página web del proveedor. En general se recomienda para mostrar los datos en formato libre, cuando sea factible.

* HTML

Hoy en día, muchos datos se encuentran disponibles en formato HTML en varios sitios. Esto sería suficiente, si se trata de datos estables y de alcance limitado. En algunos casos, sería preferible tener los datos de una forma fácil para descargar y manipular, pero como es barato y fácil de hacer referencia a una página en un sitio web, podría ser un buen punto de partida en la pantalla de datos.

Generalmente, sería más apropiado el uso de tablas en documentos HTML para almacenar datos y, es importante que los distintos campos de datos se muestren y sean dadas identificaciones, que hacen más fácil encontrar y manipular los datos. Yahoo ha desarrollado una herramienta (http://developer.yahoo.com/yql/) que puede extraer información estructurada de un sitio web, y ciertas herramientas que pueden hacer mucho más con los datos siempre que se encuentren cuidadosamente etiquetados.

* Archivos de formatos abiertos

Incluso si la información se proporciona en formato electrónico, formato legible por máquina, y en detalle, puede haber problemas relacionados con el formato del archivo en sí.

Los formatos en los cuales la información es publicada - en otras palabras, la base digital en la cual la información es almacenada - puede ser “abierta” o “cerrada”. Un formato abierto es aquel donde las especificaciones del software están disponibles para cualquier persona, de forma gratuita, así cualquiera puede usar dichas especificaciones en su propio software sin ninguna limitación en su reutilización que fuere impuesta por derechos de propiedad intelectual.

Si el formato del documento o el archivo es “cerrado”, el formato podría ser propietario y sus especificaciones no están disponibles públicamente, o podría ser porque el formato es propietario y aunque las especificaciones se han hecho públicas, su reutilización es limitada. Si la información es liberada en un formato de archivo cerrado, esto puede causar grandes obstáculos para reutilizar la información codificada en él, forzando a aquellos que deseen usar la información a comprar software innecesario.

La ventaja de los archivos de formatos abiertos, es que permiten a los desarrolladores producir varios paquetes de software y servicios utilizando esos formatos. Esto entonces reduce al mínimo los obstáculos para la reutilización de la información que contienen.

# ¿Cómo uso un formato determinado?

En INFOTEP cuando se muestren nuevos datos - datos que no han sido expuestos antes - se debe elegir el formato que proporciona el mejor equilibrio entre el costo y la idoneidad para el propósito. Para cada formato, se tienen en cuenta algunas estrategias o actividades a continuación explicadas.



## Servicios Web

Si la frecuencia de cambio en los datos es alta, y el tamaño de cada descarga es limitado, la importancia exponer la información a través de web services. Hay muchas maneras de crear web services, pero algunos de más usados son SOAP y REST. Generalmente SOAP antes que REST. Los servicios REST son fáciles de desarrollar y entender, por eso es un standard muy usado.

## Base de datos

Como los servicios web, las bases de datos permiten el acceso directo a los datos dinámicamente. Las bases de datos tienen la ventaja de extraer aquella información en la cual estén interesados.

Hay algunas preocupaciones de seguridad en cuanto a permitir extracciones remotas a bases de datos, y el acceso a las bases de datos es útil si su estructura y la importancia de las tablas y campos individuales están bien documentados. Generalmente es relativamente simple y no tiene costo crear web services que expongan datos de una base de datos, esta puede ser una manera fácil de lidiar con las problemáticas de seguridad.

# GLOSARIO

* **Base de Datos**: Es una herramienta para recopilar y organizar información. Las bases de datos pueden almacenar información sobre personas, productos, pedidos u otras cosas. Muchas bases de datos comienzan como una lista en una hoja de cálculo o en un programa de procesamiento de texto. A medida que la lista aumenta su tamaño, empiezan a aparecer redundancias e inconsistencias en los datos. Cada vez es más difícil comprender los datos en forma de lista y los métodos de búsqueda o extracción de subconjuntos de datos para revisión son limitados. Una vez que estos problemas comienzan a aparecer, una buena idea es transferir los datos a una base de datos creada con un sistema de administración de bases de datos robusta.
* **REST**: La Transferencia de Estado Representacional (REST - Representational State Transfer) fue ganando amplia adopción en toda la web como una alternativa más simple a SOAP y a los servicios web basados en el Lenguage de Descripción de Servicios Web (Web Services Description Language - WSDL). REST define un set de principios arquitectónicos por los cuales se diseñan servicios web haciendo foco en los recursos del sistema, incluyendo cómo se accede al estado de dichos recursos y cómo se transfieren por HTTP hacia clientes escritos en diversos lenguajes. REST emergió en los últimos años como el modelo predominante para el diseño de servicios. De hecho, REST logró un impacto tan grande en la web que prácticamente logró desplazar a SOAP y las interfaces basadas en WSDL por tener un estilo bastante más simple de usar.
* **SOAP**: El fundamento de todos los servicios web basados en SOAP, la especificación SOAP detalla el formato de los mensajes. También detalla la forma en que las aplicaciones deben tratar determinados aspectos del mensaje, tales como los elementos del “encabezado”, lo que le permitirá crear aplicaciones en las que un mensaje pasa entre múltiples intermediarios antes de llegar a su destino final. Este tutorial abarcará la especificación SOAP.
* **Web Services**: Están compuestos por un grupo de estándares que tratan de posibilitar que diversos sistemas puedan comunicarse, sin la necesidad de un middleware en particular, de lenguaje de programación, o incluso de un sistema operativo.

# BIBLIOGRAFÍA

Chase, N. (5 de agosto de 2011). *IBM*. Obtenido de IBM: https://www.ibm.com/developerworks/ssa/webservices/tutorials/ws-understand-web-services1/index.html

ICONTEC. (15 de 06 de 2011). NTC 5854. Bogotá, Colombia.

MINTIC. (06 de 2016). *Gobierno en Linea.* Obtenido de http://estrategia.gobiernoenlinea.gov.co/623/articles-8248\_Guia\_Apertura\_Datos.pdf

MINTIC. (s.f.). *Gobierno en Linea*. Obtenido de http://estrategia.gobiernoenlinea.gov.co/623/w3-channel.html

W3C. (2017). *Consorcio World Wide Web (W3C) .* Obtenido de http://www.w3c.es/

# CONTROL DE CAMBIOS

|  |  |  |
| --- | --- | --- |
| **FECHA** | **VERSIÓN** | **CAMBIOS** |
| septiembre de 2017 | 01 | Se define la estructura de la guía de referencia para datos abiertos. |

|  |  |  |
| --- | --- | --- |
| **ELABORADO POR:** | **REVISADO POR:** | **APROBADO POR:** |
| **Juan Camilo Cárdenas Archbold**  Profesional Universitario  Ingeniero de Sistemas |  |  |

1. Ley 1712 de 2014. Literal J, artículo 6. Definiciones. [↑](#footnote-ref-1)
2. Guía de Datos Abiertos en Colombia, MINTIC, junio de 2016, Intro. [↑](#footnote-ref-2)