Editorial IAI

PLANTEAMIENTO DE ARQUITECTURA TECNOLÓGICA PARA DATOS ABIERTOS GUBERNAMENTALES USANDO CKAN

Walter Hugo Arboleda M.

Edwin Nelson Montoya M.

Planteamiento de arquitectura tecnológica para datos abiertos gubernamentales usando CKAN

Walter Hugo Arboleda M.

Edwin Nelson Montoya M.

Editorial IAI
Instituto Antioqueño de Investigación

Medellín – Antioquia
2017

Arboleda, W. & Montoya, E.
Planteamiento de arquitectura tecnológica para datos abiertos gubernamentales usando CKAN
Editorial IAI, 2017
Medellín, Antioquia
Investigación Científica
ISBN: 978-958-59127-4-8

Planteamiento de arquitectura tecnológica para datos abiertos gubernamentales usando CKAN
Serie Ingeniería y Ciencia
© Instituto Antioqueño de Investigación

Edición: septiembre 2017
ISBN: 978-958-59127-4-8
Publicación electrónica gratuita

Diseño: Instituto Antioqueño de Investigación

Copyright © 2017 Instituto Antioqueño de Investigación (IAI)™. Except where otherwise noted, content in this publication is licensed under Creative Commons Licence CC BY-NC-SA 4.0.

Global Publisher: Instituto Antioqueño de Investigación (IAI)
Cover Designer: IAI, Medellín, Antioquia.
Editorial Instituto Antioqueño de Investigación is trademarks of *Instituto Antioqueño de Investigación*. All other trademarks are property of their respective owners.

The information, findings, views, and opinions contained in this publication are responsibility of the author and do not necessarily reflect the views of *Instituto Antioqueño de Investigación* (IAI), and does not guarantee the accuracy of any information provided herein.

Diseño, edición y publicación: Editorial IAI
Instituto Antioqueño de Investigación
http://fundacioniai.org
contacto(AT)fundacioniai.org

© 2017 Instituto Antioqueño de Investigación
Medellín, Antioquia

CONTENIDO

		Pág.
	RESUMEN	6
	INTRODUCCIÓN	7
1.	Consideraciones iniciales	7
2.	Organización del libro	8
	CAPÍTULO I	
	CONCEPTOS BÁSICOS ASOCIADOS A LAS ARQUITECTURAS DE DATOS ABIERTOS GUBERNAMENTALES	10
1.	Datos abiertos	10
2.	Los datos abiertos gubernamentales y el cambio	11
2.1	Metadatos usados en catálogos de datos abiertos gubernamentales	14
2.2	Formatos para datos usados en los datos abiertos gubernamentales	15
2.3	Problemas en la implementación de soluciones de datos abiertos gubernamentales	16
2.4	Cadena de valor de los datos abiertos gubernamentales	16
2.5	Arquitectura basada en datos	18
2.6	Arquitectura basada en servicios	19
2.7	Indexación de datos abiertos gubernamentales	20
2.8	Arquitecturas de datos abiertos gubernamentales y ciudades inteligentes	21
	CAPÍTULO II	
	CASOS DE IMPLEMENTACIÓN DE ARQUITECTURAS DE DATOS ABIERTOS GUBERNAMENTALES	26
1.	Estados Unidos	26
2.	Inglaterra	27
3.	Alemania	28
4.	Brasil	30
5.	Chile	30
6.	España	31
7.	Holanda	32
8.	Grecia	32
9.	Singapur	33
10.	Uganda	33
11.	Resultados	34
	CAPÍTULO III	
	DATOS ABIERTOS GUBERNAMENTALES EN COLOMBIA	36
1.	Descripción del modelo	36
1.1	Proceso de apertura de datos	37
1.2	Formatos	38
1.3	Escenarios de publicación	39
1.4	Instalación y uso del cargador de conjuntos de datos	40
1.5	Interfaces para el consumo de conjuntos de datos	41
1.6	Modelo de contexto y metadatos	42
1.7	Vista conceptual del catálogo	43
1.8	Vista lógica del catálogo	43
1.9	Vista de despliegue del catálogo	44
2.	Últimos cambios en el catálogo de datos de Colombia	46

CAPÍTULO IV
ARQUITECTURA TECNOLÓGICA PARA DATOS ABIERTOS GUBERNAMENTALES USANDO CKAN 47
1. Definición 47
1.1 Características técnicas de la arquitectura de datos 47
2. Implementación 47
2.1 Creación del prototipo de infraestructura 47
2.2 Interfaz de usuario del repositorio Harvester 51
2.3 Interfaz de usuario del repositorio Instancia 52
2.4 Aplicación cliente en Android 52
CAPÍTULO V
DEFINICIÓN DE ARQUITECTURA ANALÍTICA Y TECNOLÓGICA PARA DATOS ABIERTOS GUBERNAMENTALES USANDO CKAN 54
1. Importancia de la analítica de datos para el planteamiento de arquitectura tecnológica para datos abiertos gubernamentales 54
1.1 Características técnicas del prototipo para el análisis de datos 56
1.2 Software para la analítica de datos con SpagoBI Server Al-In-One 56
1.3 Software para análisis semántico son Virtuoso Open Source VOS 57
1.4 Características técnicas de la arquitectura básica de datos abiertos gubernamentales en CKAN 58
CAPÍTULO VI
CASO DE ESTUDIO PARA EL PLANTEAMIENTO DE ARQUITECTURA TECNOLÓGICA PARA DATOS ABIERTO GUBERNAMENTALES 59
CAPÍTULO VII
CONCLUSIONES 61
REFERENCIAS 62

RESUMEN

En este libro se hace un aporte a la perspectiva tecnológica de la estrategia de datos abiertos para Colombia, desarrollada por el Ministerio de Tecnologías de la Información y las Comunicaciones, para lo cual se plantea una arquitectura de datos abiertos gubernamentales en Comprehensive Knowledge Archive Network (CKAN), herramienta libre y usada por diversos gobiernos en el mundo para crear infraestructuras de datos abiertos. De esta manera se quiere aportar en la implementación de la estrategia en Colombia a nivel de entidades públicas municipales, departamentales y nacionales; se demuestra cómo crear instancias y centralizar los datos abiertos en un *harvester* en CKAN, tal como lo hace el Catálogo de datos abiertos del estado colombiano (http://www.datos.gov.co), para ofrecerlos a todos los interesados. Para esto se aborda la identificación de conceptos asociados a datos abiertos gubernamentales, los aspectos relacionados con los datos y su calidad, el análisis de arquitecturas implementadas en funcionamiento y sus componentes, las herramientas usadas, la descripción del modelo de datos abiertos gubernamentales en Colombia, el diseño de la arquitectura y la implementación de un prototipo. Al final se realiza el consumo de datos abiertos mediante una aplicación móvil en Android y el acceso a los conjuntos de datos abiertos desde un browser.

En este apartado se abordan los conceptos relacionados con la definición del problema de investigación, la pregunta de investigación, el objetivo general y los específicos, el tipo de metodología utilizada y sus etapas, el enfoque cualitativo de la investigación y la organización del trabajo, realizado con el fin de apoyar la Perspectiva tecnológica de la estrategia de datos abiertos en Colombia, liderada por el Ministerio de tecnologías de la información y las comunicaciones y Gobierno en línea. Se presenta la implementación de una arquitectura tecnológica para datos abiertos gubernamentales usando CKAN, con el objetivo de facilitar el aprendizaje y la creación de arquitecturas de datos abiertos gubernamentales a nivel municipal, departamental y nacional.

1. Consideraciones iniciales

Este libro describe un estudio cualitativo presentado como propuesta para apoyar la Perspectiva tecnológica de la estrategia de datos abiertos para Colombia. Se trata de una investigación empírica cuya fuente genérica es la realidad del país, cuyo objetivo es apoyar la generación de progreso económico y social mediante el acceso a la información pública por los ciudadanos, centros de investigación, empresas y periodistas de datos. Se busca generar valor agregado e incrementar los niveles de transparencia del Estado, aportando a la implementación que, desde el año 2000, el gobierno nacional inició de manera sistemática y coordinada en todas las entidades públicas mediante gobierno en línea (Directiva Presidencia 02 y decreto 1151 de 2008); y posteriormente con la incorporación de las nuevas temáticas en dicha estrategia en 2011, Gobierno abierto y Datos abiertos, con lo que se desarrolló el Modelo de datos abiertos para Colombia (Figura 1), a través de la evaluación de las mejores prácticas a nivel internacional en lo político-legal, tecnológica, cultural, organizacional y presupuestal (Artículo 230 del Plan Nacional de Desarrollo).

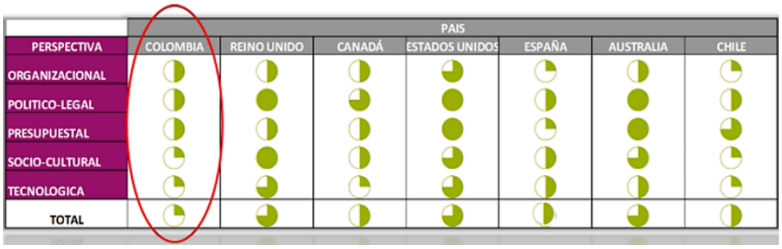


Figura 1. Implementación de datos abiertos en Colombia y en otros países (CINTEL, 2011)

Con base en la necesidad de implementar arquitecturas de datos abiertos gubernamentales en el país, se busca responder la pregunta de investigación: ¿Cómo implementar una arquitectura de datos abiertos gubernamentales, usando CKAN, que apoye la estrategia de Datos abiertos para Colombia?

En el Artículo 230 del Plan Nacional de Desarrollo 2010 - 2014 se presenta, como estrategia de buen gobierno, que todas las entidades de la administración pública deberán adelantar las acciones señaladas por el Gobierno Nacional a través de MINTIC (Figura 2) y la estrategia de Gobierno en línea, por medio de la cual se estimula el desarrollo de servicios en línea por parte de terceros basados en datos públicos, apoyándose en tecnologías de penetración tales como telefonía móvil y televisión digital terrestre, para la prestación de trámites y servicios en línea.

Figura 2. Implementación de la perspectiva tecnológica en Colombia (MINTIC, 2014d)

Además, se fomenta la participación y la democracia asociada al desarrollo de plataformas que permitan la sostenibilidad de los formatos de los datos abiertos, el inventario de los mismos, la oportunidad de actualización, el acceso, la administración y la madurez del modelo transformacional. Una herramienta que permite la apertura de datos y la generación de servicios de valor agregado por parte de terceros, usada para análisis y para ofrecer aplicaciones a los ciudadanos.

Con este trabajo se busca apoyar la perspectiva tecnológica desde lo organizacional, mediante el apoyo de procesos de publicación y acceso a los datos abiertos de las entidades gubernamentales; desde lo socio-cultural, mediante el acercamiento de información a los ciudadanos que les permita mejorar sus niveles de participación y democracia, y fomentando el desarrollo de aplicaciones móviles; y desde lo presupuestal, mediante la disminución de costos de implementación de catálogos de datos abiertos en las unidades de orden municipal, departamental y nacional, que deben ofrecer los datos abiertos de su gestión.

CKAN es una aplicación con la que se pueden resolver las necesidades de captura, alojamiento, visualización y despliegue de datos, a través de aplicaciones móviles y la centralización de conjuntos de datos, mediante la creación de *harvester* que permitan acceso a un mayor conjunto de datos. De esta manera se pretende aportar al gobierno colombiano y al Ministerio para la implementación del modelo de datos abiertos, mejorando la perspectiva tecnológica (MINTIC, 2014d), en cuanto a implementación de arquitecturas de datos abiertos gubernamentales en el país en entidades públicas de nivel municipal, departamental y nacional.

2. Organización del libro

El Capítulo I presenta los conceptos fundamentales asociados a las arquitecturas de datos abiertos gubernamentales, tales como los datos abiertos,

los datos y el cambio, los formatos usados, los problemas en implementación, la cadena de valor, las arquitecturas basadas en datos y basadas en servicios, la indexación, las arquitecturas y las ciudades inteligentes.

El Capítulo II presenta casos de implementación de arquitecturas de datos abiertos a nivel mundial, se analizan las arquitecturas implementadas y funcionales en Estados Unidos, Inglaterra, Alemania, Brasil, España, Grecia, Chile, Holanda, Uganda y Singapur.

El Capítulo III describe el Modelo de datos abiertos en Colombia, se abordan aspectos de orden legal y regulatorio, procesos y tecnológico, y se incluye el modelo conceptual y los requerimientos no-funcionales, así como las características del catálogo de datos abiertos en el país. Se aborda la Perspectiva tecnológica de la estrategia de datos abiertos en Colombia y se describen los aspectos asociados a la implementación de una arquitectura tecnológica para datos abiertos gubernamentales usando CKAN.

En el Capítulo IV se describe una implementación de la arquitectura tecnológica para datos abiertos gubernamentales usando CKAN, se aborda casos de catálogos funcionando en CKAN en Sur América y de prueba en Colombia; igualmente, se describe los componentes de la arquitectura y su uso, y el software para su implementación; se muestran la vista de procesos de la arquitectura y la vista de despliegue; además, el funcionamiento de la instancia y el *harvester* en CKAN y el acceso desde la aplicación en Android a los datos abiertos y el navegador.

En el Capítulo V se realiza una descripción de la importancia del uso del análisis semántico y la minería de datos abiertos gubernamentales con SpagoBI y Virtuoso Open Source VOS; se describan casos en los cuales se resalta la explotación de los datos abiertos para la generación de nuevo conocimiento, a partir de los conjuntos de datos; y se aborda algunos conceptos de capítulos anteriores y se establecen los componentes de la arquitectura básica en CKAN.

El Capítulo VI describe el aporte del presente trabajo al Catálogo de datos Abiertos en Colombia; se describe la tecnología utilizada y las ventajas que aporta a los datos abiertos en el país.

CONCEPTOS BÁSICOS ASOCIADOS A LAS ARQUITECTURAS DE DATOS ABIERTOS GUBERNAMENTALES

En este capítulo se abordan los conceptos sobre los que se sustenta el campo de estudio, se hace un acercamiento a las definiciones requeridas para entender los componentes y conceptos que usados en la creación de una arquitectura de datos abiertos gubernamentales; se orienta y centra la situación a investigar o problema de estudio; se hace una revisión y detección selectiva de fuentes directas en la revisión de la literatura; se abordan definiciones y temáticas tales como *open data*, formatos usados en los datos abiertos gubernamentales, cadena de valor de los datos abiertos gubernamentales y *open data* en *smart cities*, consumo de datos abiertos y uso de Application Programming Interface (API) en datos abiertos. Para la selección y revisión de los temas se consultó a especialistas en implementación de arquitecturas *open government data*.

1. Datos abiertos

El término dato se asocia frecuentemente con información, aunque existen marcadas diferencias entre ambos conceptos, principalmente cuando se refiere a datos abiertos y datos abiertos gubernamentales. Los datos son símbolos que representan características de objetos o los describen, y la información es el resultado del procesamiento de los datos, de forma que genere utilidad para la toma de decisiones (Currie, 2013). Open Knowledge Foundation (OKF) considera a los datos abiertos como el centro de una revolución en el manejo de datos gubernamentales, que pueden ser usados, reusados y redistribuidos libremente por cualquier persona pública, privada, natural o jurídica. Estos conjuntos de datos deben cumplir con características tales como disponibilidad, acceso, reúso, redistribución y participación universal. De esta manera se pueden usar, acceder de forma fácil, reusar y redistribuir por medio de APIS para que otros también los puedan utilizar (Peter, 2014). Los datos de este tipo pueden referirse a transporte, geografía, cultura, medio ambiente, clima, tráfico, consumo, estadística, finanzas, ciencia, salud, entre otros (Figura 3).

Figura 3. Tipos de Open Data (Peter, 2014)

Los datos abiertos garantizan transparencia, dan valor agregado en actividades sociales, comerciales y aumentan la participación de los ciudadanos. En UPF Handbook (2014) se aborda aspectos legales, sociales y de aspectos técnicos, y se establece que, para garantizar la apertura de los datos abiertos, se debe cumplir cuatro pasos: 1) Seleccionar el conjunto de datos que serán abiertos, 2) Identificar

los derechos de propiedad intelectual, 3) Aplicar una licencia de datos abiertos, pero si no se puede aplicar entonces se debe tomar otro conjunto de datos, y 4) Hacer disponible el conjunto de datos por medio de un formato adecuado para su acceso, y usar una plataforma o API que permita buscar, listar y acceder los conjuntos de datos. Además de tres reglas: Los conjuntos de datos deben ser sencillos y cortos para evitar daños en ellos, 2) Tener en cuenta que los datos tendrán usuarios potenciales que harán uso y reúso de los conjuntos de datos, y 3) Abordar los aspectos críticos de los conjuntos de datos, en especial si estos datos son gubernamentales.

2. Los datos abiertos gubernamentales y el cambio

La pobreza y la corrupción son problemas que a diario enfrentan los países en desarrollo, debido a que algunos políticos y líderes se apropian de recursos y propiedades del estado usando medios ilegales. Los datos abiertos gubernamentales son una forma de cambiar los problemas existentes y de disminuir la pobreza, y para lograr un cambio positivo en la sociedad garantizando transparencia, honestidad y participación ciudadana (Larsson, 2013).

En países con un buen nivel de vida, como Austria, los datos gubernamentales abiertos son un elemento primordial para sustentabilidad social, cultural, científica y progreso económico; y se usan como una herramienta para garantizar transparencia y acción administrativa, para mejorar la cooperación entre políticos, gobierno, industria, negocios, investigadores y ciudadanos, fortaleciendo la democracia. En este país se creó una estrategia de datos abiertos que garantiza la interoperabilidad de los datos y el correcto funcionamiento de las plataformas, desde la producción hasta la entrega de los datos abiertos para el consumo de los ciudadanos o de servicios creados para ello. De esta forma desarrollaron métodos para la concentración de datos abiertos gubernamentales, pero en el proceso encontraron riesgos tales como la falta de una adecuada interpretación de datos estadísticos por los ciudadanos, la falta de personal capacitado en proyectos para crear y desarrollar plataformas de datos abiertos gubernamentales, y la calidad de los datos cuando se mezclan conjuntos de baja calidad. En algunos casos la presentación de datos ha perjudicado a los sectores industrial, comercial y productivo, por ejemplo, el de los bienes raíces, que se ha visto afectado por los resultados publicados en los datos abiertos gubernamentales sobre la contaminación de los suelos y del aire, disminuyendo la construcción y venta de inmuebles en esas áreas (Eibl, 2013).

En Pakistán estos datos se definen como parte del gobierno electrónico y se ven como un elemento importante para la modernización y reforma del gobierno. Con esto se aumenta su rendimiento y resultados, mostrando fácilmente los indicadores de su gestión y entregando a los ciudadanos elementos de información valiosa para la toma de decisiones y la comprensión de la dinámica nacional;

además, sobre cómo esa dinámica se relaciona con todos los estamentos económicos, comerciales, políticos y sociales del país (Shahzad, 2007).

En Estados Unidos, en su primer día de gobierno, el presidente Obama hizo varias declaraciones sobre la importancia de los datos abiertos en su gobierno, y subrayó su utilidad para la recuperación de la economía, si se garantizaba un buen uso de los impuestos de los ciudadanos por medio de transparencia y control. Su idea fue permitir el acceso a los datos sobre la gestión del gasto público, haciendo que los ciudadanos fueran los ojos y los oídos de las decisiones gubernamentales (Lathrop y Laurel, 2010). De esta manera, investigadores y empresarios pueden conocer más de los datos abiertos gubernamentales, y en diversos países se presentan proyectos asociados a esta temática, tales como Estados Unidos, Inglaterra, Canadá y Australia, entre otros, y gracias al trabajo y liderazgo de especialistas como Tim Berners-Lee y Nigel Shadbolt (Mustafa, 2012).

Por otro lado, las crisis financieras y económicas les han permitido a muchos gobiernos pertenecientes a la Organization for Economic Cooperation and Development (OECD), implementar ajustes estructurales y estrategias a sus planes de gobierno para recuperar la salud financiera. Sin embargo, las ilusiones de los ciudadanos han decaído debido a las altas expectativas que se trazaron sus gobiernos, sabiendo que se tenían recursos limitados para esa gestión. Por eso es necesario entregar la información a los ciudadanos sobre el estado de los recursos y rubros disponibles, y sobre cómo se usan, porque de esta manera se logra transparencia y gestión gubernamental de los recursos disponibles; se logran cambios importantes en gasto y finanzas públicas; y se mejora los niveles de empleo y de exportaciones de bienes y servicios. Esto ha logrado que los datos abiertos gubernamentales ganen importancia como herramienta para el crecimiento regional y nacional, garantizando el uso de los recursos y creación de políticas mediante debates públicos. Lo que se logra porque los ciudadanos tienen acceso a datos sensibles, y se convierten en agentes gestores de mejoras que generan valor agregado al comercio mediante la creación de nuevos bienes y servicios a partir de su análisis (OECD, 2013).

Los datos abiertos gubernamentales se convirtieron en un método de publicación de datos, que se ponen a disposición y son utilizados para la toma de decisiones en el sector público. Además, con estas iniciativas las entidades gubernamentales crean conjuntos de datos, llamados *datasets*, que pueden ser descargados libremente, pero que deben cumplir con los ocho principios de los datos abiertos gubernamentales (Radl, Jäger, Mödritscher, y Komendera, 2013):

1. Estar completos
2. Estar en su formato y estado primario
3. Constantemente actualizados
4. Procesables por equipos de cómputo

5. Ser accesibles para los ciudadanos y las instituciones públicas y privadas que los requieran
6. Cumplir la característica de acceso abierto a todas las personas naturales y jurídicas
7. Utilizar formatos libres y licencias de acceso abierto para garantizar su gestión de forma abierta y fácil
8. Poseer un nivel de calidad que permita la aplicación de técnicas estadísticas y de visualización de la información para su análisis

De acuerdo con European Commission (EC), existen dos motivaciones principales para el uso de datos abiertos gubernamentales: 1) la democracia y la libertad de información, que sirven para garantizar la transparencia de los gobiernos; y 2) sus bases económicas, dado que los *datasets* generan información importante para los negocios, las empresas y las personas, para diseñar productos y servicios basados en la explotación de esos datos, lo que se convierte en una oportunidad para generar valor agregado e innovación (Chan, 2013).

En los últimos años la cultura del Free Open Source Software (FOSS) ha impulsado la colaboración y el compartir datos, asociándose al crecimiento del movimiento de los datos gubernamentales, que se asocian al concepto de abiertos y libres y donde los ciudadanos, además de tener acceso a la información, son proactivos y participativos en el gobierno. Las iniciativas de datos abiertos están enmarcadas en el trabajo en red, en el que aparecen varios valores producidos y co-creados, y que, finalmente, logran: 1) que las plataformas de datos abiertos gubernamentales estén disponibles para grupos de instituciones públicas y privadas, y 2) permitir la sincronización con otras de nivel superior, que involucren un mayor valor agregado, mediante la creación de nuevos servicios y negocios que garanticen una mejor eficiencia y eficacia, así como transparencia, participación y colaboración (Avital y Bjorn-andersen, 2012).

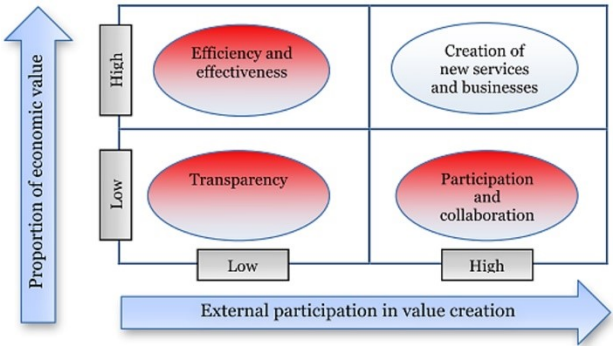


Figura 4. Elementos de valor de iniciativas estratégicas de datos abiertos (Avital y Bjorn, 2012)

Los datos abiertos gubernamentales los generan entidades públicas, con el objetivo de mostrar los resultados de su gestión como entidad estatal, y pueden

ser libres, reusados y redistribuidos por cualquier ciudadano. Además, garantizan la transparencia de una sociedad democrática mediante la liberación de información de su desempeño a los ciudadanos, de forma que se puedan analizar y visualizar para extraer información sobre el estado de los programas y proyectos gubernamentales. De esta manera se genera valor agregado que se puede utilizar con fines sociales y comerciales, maximizando el involucramiento de los ciudadanos en los aspectos de participación ciudadana. También se cumple con el uso de estándares que garanticen el acceso a la información de forma equitativa, usando formatos de datos de libre acceso, legibles y gestionables por medio de computadores y dispositivos móviles. Así se permite el acceso, consulta y análisis de datos en aspectos como agricultura, educación, migración, desarrollo, calidad de vida, exportaciones, empleo, pensiones, impuestos, proyecciones económicas, salud, pobreza, equidad y comercio, entre otros.

2.1 Metadatos usados en catálogos de datos abiertos gubernamentales

Data Catalog Vocabulary (DCAT) es una definición para conjuntos de datos abiertos definida por el World Wide Web Consortium (W3C), adoptada por Europa y usada a nivel internacional en diferentes catálogos de datos abiertos gubernamentales, tales como data.gov y data.gov.uk. DCAT un vocabulario Resource Description Framework (RDF) que facilita la interoperabilidad entre catálogos de datos abiertos gubernamentales, en el que se define los metadatos para identificación de conjuntos de datos mediante las clases (W3C, 2014):

1. *Class Catalog*: describe el catálogo y posee los metadatos *title*, *description*, *release date*, *update/modification date*, *language*, *homepage*, *publisher*, *spatial/geographic*, *themes*, *licence*, *rights*, *dataset* y *catalog record*.
2. *Class Catalog record*: describe un *dataset* mediante los metadatos *description*, *listing date*, *primary topic*, *title* y *update date*.
3. *Class Dataset*: describe un *dataset* con los metadatos *contact point, description, distribution, frequency, identifier, keyword, landing page, language, publisher, release date, spatial coverage, temporal coverage, theme, title* y *update date*.
4. *Class Distribution*: describe un archivo de un *dataset* mediante los metadatos *access URL, byte size, description, download URL, format, license, media type, release date, rights, title* y *update date*.
5. *Class Concept scheme*: describe los temas y las categorías de un conjunto de *datasets* usando los metadatos *catalog themes* y *dataset theme*.
6. *Class Concept*: describe los conceptos usados para identificar los temas y las categorías con los metadatos *catalog themes* y *dataset theme*.
7. *Class Organization/Person*: describe las organizaciones gubernamentales que suministran los *datasets* y a las personas que pueden gestionarlos.

2.2 Formatos para datos usados en los datos abiertos gubernamentales

La W3C publicó un documento de trabajo llamado Publishing Open Government Data (W3C, 2009), en el cual exhorta, en la implementación de proyectos de datos abiertos gubernamentales, a utilizar formatos de datos que permitan su utilización generalizada, que garanticen el acceso, uso y reúso de los datos. Por lo tanto, recomienda los formatos estructurados XML y RDF porque permiten manipulación y lectura a través de herramientas, bases de datos y lenguajes de consulta, tales como SPARQL, XQuery y JavaScrip. Esto los convierte en estándares abiertos para la producción y publicación de este tipo de datos y para el aprovechamiento de datos enlazados (W3C, 2009). En la Tabla 1 se describe los datos estructurados.

Tabla 1. Formatos utilizados en los datos abiertos gubernamentales

Formato	Descripción
JSON	JavaScript Object Notation. Es un formato de intercambio de datos entendible por computadores y personas, basado en JavaScript; es un formato de texto independiente del lenguaje de programación (JSON, 1999).
XML	Extensible Markup Language. Es un formato de texto flexible y sencillo derivado de SGML (ISO 8879); tiene un rol importante en el intercambio de datos en Internet (XML, 2008, 2013).
RDF	Resource Description Framework. Es el estándar para la codificación de metadatos y conocimiento en la Web Semántica, donde los computadores usan información estructurada de forma centralizada o distribuida; es un modelo abstracto que permite dividir el conocimiento en piezas discretas, de acuerdo con la semántica, usando una estructura jerárquica mediante clases, objetos, atributos y valor (W3C, 2014a).
XLS, XLSX	Excel Binary File Format: son hojas de cálculo en las que se organiza la información mediante filas y columnas con sus correspondientes encabezados (Microsoft, 2007, 2014) .
Documento de Texto	Open Document Format (ODF). Es una familia de estándares internacionales sucesora de los formatos .doc, .wpd, .xls y .rtf; son un método de almacenamiento y procesamiento de información multiplataforma usado por las soluciones de Apache Open Office, Libre Office, IBM Lotus Symphony, Calligra Suite, Euro Office and Microsoft Office ,Google Docs y Zoho Office (OASIS, 2014). OOXML o OpenXML. Es el estándar de los formatos de Microsoft Office (.docx, .pptx, .xlsx) basado en XML (ECMA, 2012; Microsoft, 2012). Portable Document Format (PDF). Es un estándar abierto, portable y muti-plataforma para el intercambio de documentos electrónicos; puede ser usado en archivos con texto, imágenes, elementos multimedia, enlaces de página web, entre otros elementos (Adobe, 2014) . DOC y DOCX. Es el formato de Microsoft Word para documentos de texto basado en el estándar abierto XML (Zetlin, 2010)(IEC, 2014).
TXT	Es un formato usado en una variedad de editores de texto plano, que genera una secuencia de caracteres de fácil lectura; no es un estándar; usa caracteres tipo ASCII y ANSI (Corporation, 2014)
Imágenes	Tag Image File Format (TIFF). Es un formato para gráficos de alta calidad, cuyos archivos se pueden convertir a LZW, JPEG, GIF o PDF (Adobe, 1992). Joint Photographic Experts Group (JPEG). Es un estándar para la codificación de imágenes fotográficas de calidad que se utiliza en una amplia variedad escenarios (JPEG, 2007).

2.3 Problemas en la implementación de soluciones de datos abiertos gubernamentales

Según European Public Sector Information Platform (PSI, 2013), los principales problemas (Figura 5) para la implementación de soluciones de datos abiertos gubernamentales son: 1) baja calidad, de datos que no son representativos para generar información; 2) censura, por lo que no son de acceso libre a los ciudadanos e interesados; 3) secretos, definición con la que algunos gobiernos amparan datos importantes perjudicando su transparencia; 4) presupuesto, en el que se dispone un bajo presupuesto para este tipo de proyectos y para capacitar a las personas; 5) islas de datos, por lo que no es fácil integrar los datos abiertos en un solo lugar; 6) Falta de prospectiva, porque no se visualiza la relación a interacción con los usuarios y consumidores de estos datos.

Figura 5. Problemas para la implementación de datos abiertos gubernamentales (ePSI, 2013)

2.4 Cadena de valor de los datos abiertos gubernamentales

Aprovechando las ventajas que ofrecen Internet y las nuevas tecnologías, los ciudadanos tienen acceso a gran cantidad de datos abiertos, ya sea desde su casa, trabajo o cualquier lugar del mundo. Una de esas herramientas es www.datacatalogs.org, donde se recolecta estos datos desde varios países, ofreciendo actualmente más de 300 catálogos de datos. Por su parte, Global Open Data Initiative (GODI) apoyó la implementación de este tipo de iniciativas en Estados unidos y Europa, logrando aumentar la cantidad de países participantes en el Open Government Partership a más de 57 miembros de varias partes del mundo (Corona, 2013).

Implícitamente los datos abiertos gubernamentales incluyen dos valores: 1) el de la inversión, utilizado para crear, almacenar y mantener los datos; 2) negocios y beneficios, que se puedan generar al usar los datos para crear aplicaciones de consumo. Este crecimiento hace pensar en la necesidad de crear una infraestructura que maneje la cadena de valor de los datos (Figura 6), desde la creación de los datos, su almacenamiento, el procesamiento y empacado de los datos, para finalmente entregarlos a los interesados (Van Grieken, 2011).

Figura 6. Cadena de valor de los datos abiertos gubernamentales (van Grieken, 2011)

La estructura de datos abiertos gubernamentales basados en la demanda de los ciudadanos, define las comunidades de generación de datos (Figura 7), tanto privados como gubernamentales, permitiendo que los usuarios tengan acceso a los que generan ambos sectores, para monitorear el gobierno y para conocer el comportamiento de la economía, el crecimiento de la industria, los nuevos sectores de negocios y los niveles de productividad por sector. Los datos abiertos gubernamentales finalmente son un subconjunto de los datos del gobierno, que pueden ser publicados y que no amenazan la seguridad nacional ni violentan el derecho a la privacidad en la información (Currie, 2013).

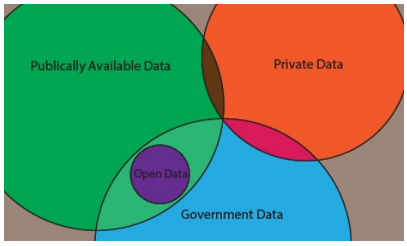


Figura 7. Esferas de los datos abiertos gubernamentales (Currie, 2013)

Para el consumo de datos abiertos existe actualmente un amplio número de aplicaciones, entre las que se encuentran los navegadores de propósito general para datos enlazados tales como Tabulator, Haystack, Longwell y IsaViz, y servicios como Sindice y DBpedia Mobile (Costabello, 2013). Por otro lado, los usuarios de Internet consumen datos de diversas maneras, por ejemplo, con Facebook, Foursquare, Youtube, Twitter e Instagram, que han sido masificadas por la comunidad y las redes sociales. Por eso es importante que las plataformas y los servicios de datos abiertos permitan el acceso a ellos de forma similar a las redes sociales, donde los usuarios comparten e intercambian información e ideas a través de foros y grupos de interés (Høgenhaven, 2013). En este mismo sentido, estos datos también deben garantizar la calidad, para lo cual se recomienda tener en cuenta los principios formulados por US Open Government y mejorados por Sunlight Foundation:

1. Completitud, los datos abiertos publicados deben estar completos
2. Fuente primaria, deben ser publicados tal como fueron procesados en su fuente original
3. Disponibilidad, estar disponibles por un tiempo considerable que garantice su consulta
4. Accesibilidad, fáciles de consultar, sin barreras técnicas o físicas (Eibl, 2013)
5. Facilidad de lectura electrónica, deben ser fáciles de procesar, abrir y acceder por los usuarios

6. Uso de estándares abiertos, deben ser independientes de formatos propietarios que puedan limitar el acceso, conservando las recomendaciones de W3C, BLSG y SAGA
7. Licenciamiento, deben ser publicados bajo licencia Creative Commons
8. Documentación, deben estar debidamente documentados
9. Durabilidad, deben ser correctamente identificados con metadatos adecuados y garantizar el control de versiones y de archivado
10. Costo de uso, no deben tener costo para ningún usuario

Para Yu (2012), desde que en 1950 se usará el termino gobierno abierto y aunque entonces pareciera difícil darles acceso a estos datos a los ciudadanos, el actual desarrollo de las Ciencias Computacionales ha convertido esta idea en una realidad mediante la creación de plataformas de datos abiertos. Además, se garantiza la libertad de información a través de la entrega de datos sensibles de los asuntos nacionales, que también pueden usarse para promover la innovación y el crecimiento económico. Por eso es importante que los gobiernos definan estrategias y adopten tecnologías que permitan la implementación de este tipo de plataformas, que garanticen el flujo de información entre gobierno y ciudadanos, y haciendo que la tecnología se convierta en un amplificador.

En la última década ha habido un alto crecimiento de instituciones, gobiernos y compañías (Costabello, 2013) que colocan sus datos de forma abierta en Internet, lo mismo que información relacionada con niveles de contaminación, tiempos y cronogramas del transporte público, catálogos de productos, entre otros. Estos datos están siendo consumidos por terceros de forma totalmente libre, para construir modelos de negocio y descubriendo información que aprovechan los medios para realizar análisis. Lo que se logra porque deben cumplir las leyes de Open Government Data: 1) si un dato no se puede indexar no existe, 2) si un dato no es legible por computador no se puede tener en cuenta, y 3) si no se puede reusar mediante un *framework* no agrega valor (Machado y Parente de Oliveira, 2011). Esto les ha brindado a los gobiernos oportunidades para transformar la manera cómo interactúan con los ciudadanos, usando TIC, ofreciendo los datos de forma electrónica, para que los usuarios conozcan de primera mano todo lo relacionado con la gestión de su país. Esto también es una ventaja para los ciudadanos, ya que pueden recuperar los datos abiertos y descubrir información instantáneamente sin ningún costo (Yu, 2012).

2.5 Arquitectura basada en datos

Actualmente, muchos gobiernos publican los datos de su gestión en formatos no-estructurados, lo que genera problemas para reusarlos y obtener información a partir de ellos. Para remediar estos problemas surge e-Government, lo que significa el uso TIC para interactuar con los ciudadanos y entregarles los datos que

requieran por medio de computadores, tabletas y teléfonos móviles. En este sentido surgen iniciativas como la del gobierno brasilero (Machado y Oliveira, 2011), que diseñó la arquitectura de la plataforma Delivering Information of Government (DIGO), y de esta manera definir cómo se puede recolectar los datos fuente en formatos estructurados, semiestructurados y no-estructurados. Además, permitiendo integrar datos desde diferentes fuentes, fusionándolos en una nube Linking Open Data (LOD) a través de una arquitectura por capas:

1. La capa de la base de conocimiento organizacional, que soporta datos estructurados y no-estructurados provenientes de diversas fuentes, incluyendo bases de datos y OLTP.
2. La capa de sintaxis, donde los datos estructurados, provenientes de bases de datos y sistemas OLTP, se almacenan en otra base de datos como estructurados, lo mismo sucede con los no-estructurados contenidos en correos electrónicos, mensajes, documentos, hojas de cálculo y páginas web.
3. La capa de semántica posee una capa de extracción con tres subcapas para extraer datos no-estructurados, semi-estructurados y estructurados; una capa de transformación y cargado donde se transforma o mapea los tipos de datos a datos semánticos, usando ontologías como vocabularios y taxonomías de tal manera que se pasa de ontología OWL a tripletas RDF, y una capa de persistencia, donde se almacenan datos semánticos RDF, índices, metadatos y ontologías en una base de datos semántica. La capa de manipulación provee datos abiertos mediante interface SparQL, procesador de consultas SparQL y un editor de datos semánticos, que puede realizar operaciones CRUD.
4. La capa de fusión de datos, que permite unir *datasets* de la plataforma con otros de otras plataformas LOD.
5. La capa de información, que le brinda al usuario final el acceso a los datos abiertos y manipularlos.

Para el consumo de datos abiertos por las aplicaciones se utiliza un modelo directo (Figura 8) por medio de una API, que entrega datos a los desarrolladores.

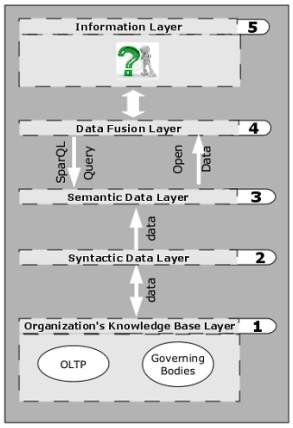


Figura 8. Modelo API de Integración Directa (Machado y Oliveira, 2011)

2.6 Arquitectura basada en servicios

Krantz (2014) propone una arquitectura (ver Figura 9) orientada a que las aplicaciones puedan consumir datos, que les permite a los desarrolladores descargar los datos abiertos, desde la plataforma gubernamental, a un servidor propio, en el que se ponen a disposición de sus aplicaciones.

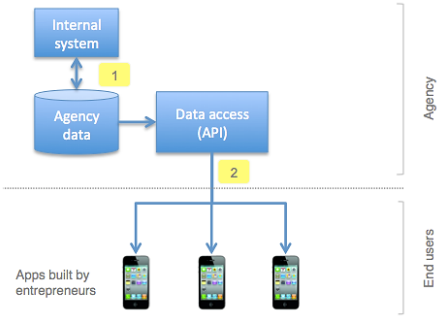


Figura 9. Modelo API de Integración Directa (Krantz, 2014)

Este modelo descarga responsabilidades a nivel de API para aplicaciones en las plataformas gubernamentales de datos abiertos, logrando un mejor desempeño en su funcionamiento y delegando responsabilidades a la infraestructura de las compañías y desarrolladores. Además, establece los roles a las agencias que ofrecen los datos abiertos, a los emprendedores que toman los datos y realizan las aplicaciones y los usuarios finales que usan las aplicaciones para tener acceso a información. Este modelo está basado en API que interconectan cada uno de los roles y ofrecen los datos (Figura 10).

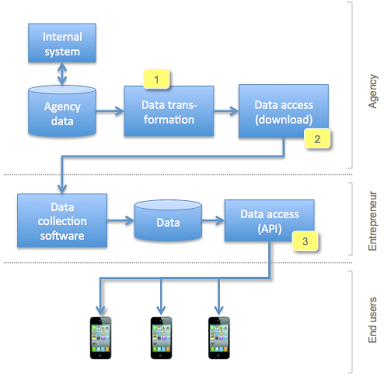


Figura 10. Modelo API de Integración Directa (Krantz, 2014)

El servidor en el que se almacenan los datos recibe las actualizaciones de conjuntos de datos abiertos desde las plataformas gubernamentales y los lleva a su base de datos, para ponerlos a disposición del consumo de las aplicaciones. Esto les permite a los gobiernos concentrarse en publicar sus datos en formatos estructurados, y así evitar el uso de interfaces para compartirlos y entregarlos a los ciudadanos o interesados (Yu, 2012). Entonces, de esta manera los datos se crean y almacenan en formatos que permitan su reúso por terceras personas. Además, al facilitar su explotación y consumo mediante una forma adecuada, se les agrega valor a los datos abiertos. Por eso es necesario desarrollar estrategias nacionales orientadas a capacitar en el proceso de recuperar la información de sitios web gubernamentales y a continuar publicando la información directamente en estas plataformas.

2.7 Indexación de datos abiertos gubernamentales

A partir de una iniciativa de Open Knowledge Foundation (OKF), UPF (2014) mantiene el listado de los gobiernos que ofrecen datos abiertos. Actualmente tiene 70 países con 700 conjuntos de datos, con 84 conjuntos de datos abiertos que representan el 12% de los existentes en el índice. Como se observa en la Figura 11, en él se clasifican las plataformas de datos abiertos gubernamentales teniendo en cuenta factores tales como: transporte, presupuesto público, gasto público, resultado de las elecciones, registro de compañías, datos del mapa nacional,

estadística nacional, legislación, códigos postales y emisiones de contaminantes. El valor total les permite a los usuarios conocer el comportamiento de sus gobiernos en todos los continentes, al mismo tiempo que de aquellos cuyo aporte es más restringido con los datos abiertos (Figura 12).

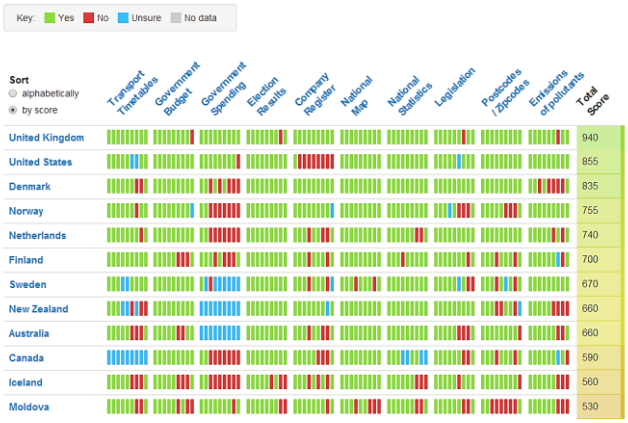


Figura 11. Primeros 12 países en The Open Data Index (UPF, 2014)

Figura 12. Últimos ocho países en The Open Data Index (UPF, 2014)

2.8 Arquitecturas de datos abiertos gubernamentales y ciudades inteligentes

Ren y Glissmann (2012) proponen una forma estructurada para identificar valor, sacar provecho y explotar las iniciativas de datos abiertos usadas para crear servicios para Ciudades Inteligentes, de tal manera que los gobiernos puedan formarse una idea clara de implementación de este tipo de soluciones, estableciendo el orden de prioridad y la calidad. Estos autores definen una arquitectura de negocio de administración municipal centrada en el funcionamiento de la ciudad en áreas tales como seguridad, transporte, salud y economía. Así crean una estrategia de gobierno (Figura 13) modelada mediante las

competencias que conforman el ecosistema de ciudad, a la vez que definen los elementos que deben planear, administrar, ejecutar y controlar, y sobre los cuales deberá trazar estrategias para alcanzar la visión de ciudad. A partir de estos se construye una iniciativa de datos gubernamentales abiertos para satisfacer las necesidades de información para diferentes actores interesados: ciudadanos residentes, empresas, visitantes y turistas, y gobiernos municipal y departamental.

	City Strategy & Governance	Public Safety	Transportatic	Citizen Health	Energy & Water	Environment Sustainabilit	Urban Planning & Building Managemen	Economic Developmen	Social Services	Education, Culture & Recreation	Municipal Administratic
Direct	City Vision and Strategy	Public Safety Strategy	City Transportatic Strategy	Citizen Health Strategy	City Utilities Strategy	Eco-City Strategy	Urban Planning Strategy	City Economic Policies	Social Services Strategy	Education and Culture Policies	Governmen wide Administratic Strategy
Control	City Performance Managemen	Crime, Fire and Emergency Managemen	Transportatic Service Managemer	Health Service Managemen	Utilities Service Managemen	Sustainability Programs Managemen	Developmen and Permit Managemen	Economic Programs Managemen	Social Programs Managemen	Education and Culture Programs Managemen	Administratic Services Managemer
Execute	City Governance Operations	Public Safety Operations	Transportatic Infrastructur Operations	Health Service Operations	Utilities Infrastructur Operations	Sustainability Programs Delivery	Land and Buildings Operations	Economic Developmen Operations	Social Services Delivery	Education and Culture Operations	Administratic Services Delivery

Figura 13. Modelo de negocios para una Ciudad Inteligente (Ren y Glissmann, 2012)

En 2011 el alcalde de Viena presentó la iniciativa Smart City Wien (2014), un concepto centrado en los datos que necesitaban los interesados, dentro y fuera de la ciudad. Para materializarla se realizaron grupos de trabajo que se concentraron en seis diferentes temas: desarrollo de la población, medio ambiente, administración, economía, energía y movilidad. Luego se realizaron tres foros para establecer la estrategia de ciudad, seleccionando el área de energía, con la que se creó la visión Smart Energy Vision 2050 con el objetivo de garantizar a largo plazo el suministro energético para la ciudad. Esto le permitió a Viena ocupar el primer lugar en el ranking mundial de Smart Cities en 2012, debido al manejo y control aspectos de la ciudad tales como emisiones de CO_2 por combustibles fósiles, calidad de vida, infraestructura e innovación. Con lo que se garantiza el crecimiento económico, ecológico y social, a la vez que estos datos gubernamentales abiertos aportan a otras soluciones de ciudad: Smart Business, Smart Energy y Smart Mobility.

Otro beneficio de este tipo de iniciativas es facilitar los datos gubernamentales para crear proyectos que, para Viena, fueron: aspern Vienna's Urban Lakeside para espacio público integrado; e-mobility on demand para incrementar los autos eléctricos; SMIEL para integrar información de todos los medios de transporte; EcoBuy Vienna para crear empresas ecológicamente sustentables; Citizens' Solar Power Plants para que los ciudadanos tengan en sus hogares su propia planta de energía solar.

Para (Clarke, 2013), Smart Cities es una realidad para la gestión de municipios en todo el mundo, porque la gestión inteligente les permite resolver problemas

actuales, abordar retos futuros y crear servicios sorprendentes. Pero sus alcaldes deberán emprender proyectos innovadores con socios del sector público y privado, para crear regulaciones y leyes para ciudades inteligentes, al mismo tiempo que clústeres innovadores para generar nuevas economías y puestos de trabajo. Estas tecnologías integran y analizan grandes cantidades de datos para anticipar, mitigar y prevenir los problemas y retos que enfrentan hoy las ciudades: reducción del crimen y accidentes, enrutamiento del tráfico, lugares de parqueo, tiendas, calidad del aire, entre otros. De esta manera se conectan los servicios con los ciudadanos, brindándoles más información y colaboración como respuesta a sus necesidades, además, se generan soluciones económica y ecológicamente sustentables. Como se observa en la Figura 14 esta arquitectura tiene cuatro capas:

1. Tecnologías habilitantes para el análisis de *open data* y *big data*, dispositivos móviles, redes sociales e internet con servicios en la nube.
2. Negocios imperativos con datos abiertos y transparencia para la gestión del riesgo, la prestación de un mejor servicio para los ciudadanos y el incremento de la productividad y la innovación.
3. Valor agregado en la satisfacción de necesidades de información de ciudadanos e interesados, predicción, nuevos productos y servicios y ventajas competitivas.
4. Las capas anteriores permitirán alcanzar los objetivos de estas ciudades: desarrollo económico, sustentabilidad y alta calidad de vida.

Figura 14. Arquitectura de una Smart City (Clarke, 2013)

La arquitectura de este tipo de soluciones para Smart City se relaciona directamente con los datos gubernamentales abiertos, porque se hace un censado y control que genera datos abiertos provenientes de las redes de sensores, vehículos inteligentes, calles, autopistas, puentes, quioscos multimedia y personas que usan dispositivos móviles. De esta manera los datos son transportados, almacenados, analizados y catalogados; luego son entregados mediante APIS a aplicaciones privadas y aplicaciones libres mediante servicios de datos abiertos gubernamentales y datos útiles para desplazarse en la ciudad (Figura 15).

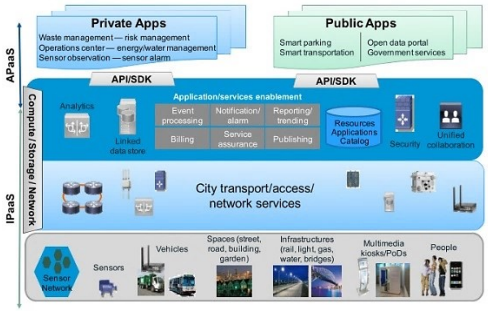


Figura 15. Arquitectura detallada de una plataforma inteligente (Clarke, 2013)

Según (Cohen, 2012) este modelo de rueda inteligente (Figura 16) ha sido inspirado en el trabajo del Centro Regional de Ciencia en la Universidad de Tecnología de Viena, Siemens y el modelo territorial de Buenos Aires, y permite identificar los componentes de una ciudad inteligente, los objetivos a alcanzar y los indicadores que ayudan a medir el alcance del objetivo. La urbanización de las ciudades, el aumento demográfico y el rápido avance en las nuevas tecnologías, está cambiando la forma como los alcaldes y gobernadores crean nuevos servicios digitales para los ciudadanos.

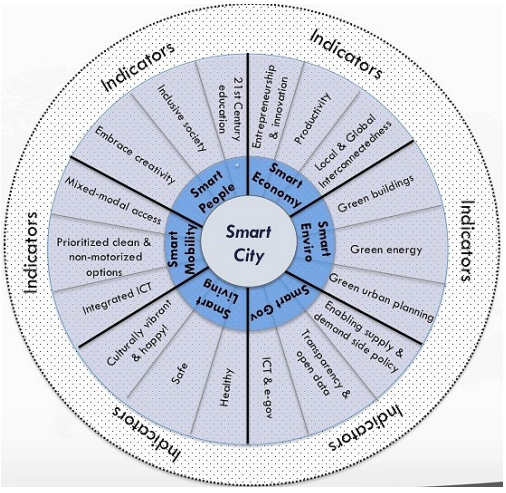


Figura 16. Rueda de las ciudades inteligentes (Cohen, 2012)

Este cambio genera un ecosistema de innovación que responde a las necesidades del mercado y de las personas en temas como hospitales, escuelas, universidades, sector comercial, seguridad, calidad del agua y el aire, temperatura, lluvia. De esta manera se ayuda a la realización de nuevos proyectos amigables con el medio ambiente, tales como edificios verdes, alumbrado público inteligente, redes eléctricas inteligentes, consumo de energía casera inteligente y el incremento en el uso del servicio público de transporte, reduciendo así el tráfico y la contaminación que genera, tal como se aprecia en la Figura 17.

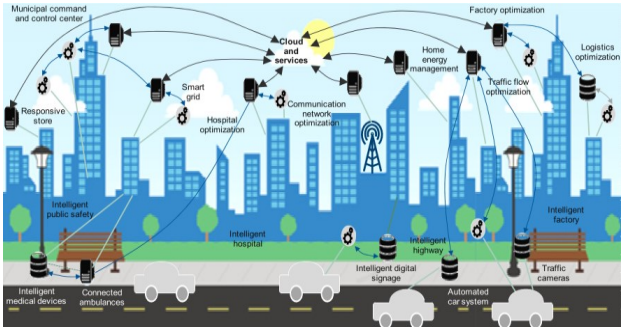


Figura 17. Ciudad con ecosistema digital (Clarke, 2013)

Por otro lado, cada vez más las personas desean acceder a los diferentes servicios desde sus dispositivos móviles, para informarse de los aspectos de la ciudad que llaman su interés y que pueden utilizar en su día a día. Las entidades de gobierno también se ven favorecidas al tener control mediante cámaras y sensores de lo que sucede en la ciudad, realizando análisis de datos para luego generar estrategias, ideas de negocio y nuevos empleos a partir de ellos (Clarke, 2013).

En este capítulo se describen algunos casos de implementación de plataformas de datos abiertos gubernamentales, que son un referente importante para tener en cuenta en el diseño y concepción de este tipo de plataformas, los cuales abordan aspectos normativos y técnicos para la puesta en marcha de estas soluciones. Se abordan los avances en la implementación de estas arquitecturas mediante el análisis a algunos casos funcionales en diversos países. El objetivo es comprender los componentes de las mismas y ayudar y la manera como se implementan. Se seleccionan fuentes de información primarias y relevantes que permitan tener un concepto integral del objeto de estudio, de forma que sean un insumo para una construcción crítica basada en estas experiencias.

1. Estados Unidos

En 2009 los gobiernos de Estados Unidos e Inglaterra realizaron acuerdos públicos para apoyar el establecimiento de los datos abiertos gubernamentales. En Estados Unidos se lanzó el portal www.data.gov y en Inglaterra el http://data.gov.uk diseñados con la arquitectura que se describe en la Figura 18.

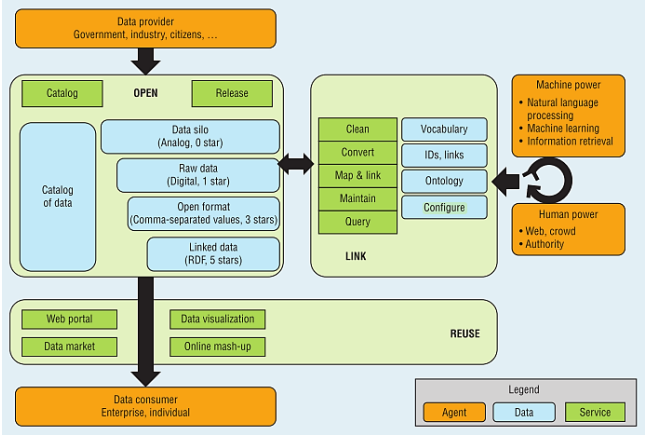


Figura 1. Arquitectura www.data.gov (Ding, 2012)

Esta arquitectura recibe datos gubernamentales abiertos y los cataloga conservando aspectos de calidad de *cinco estrellas*; luego los depura, convierte, mapea y mantiene de forma automática y computarizada o mediante la intervención de personas; posteriormente son publicados para su visualización en el portal web. También son enviados a almacenes de datos para reutilización y análisis por terceros a través de API, además, se envían a los usuarios mediante servicios en línea tipo *mash-ups*, permitiendo que los consumidores, tales como empresas y ciudadanos, los puedan recepcionar y utilizar.

A partir de esta iniciativa han aparecido varios movimientos proactivos que apoyan la directiva, debido a que permite el reúso, la aparición de nuevas oportunidades de negocio, mejora la transparencia y entrega la información oportuna de los ciudadanos. Por su parte la Comisión Europea abordó este tema desde 2003 a través de una directiva del sector público y en 2011 mediante el Open Data Package. Para enero de 2012 ya habían más de 700.000 conjuntos de datos gubernamentales abiertos de todo Estados Unidos en http://logd.tw.rpi.edu/, sitio del proyecto estadounidense TWC LOGD Linking Open Government Data (Ding, 2012).

2. Inglaterra

En 2013 inició el proyecto Maximising the Exploitation of Linked open Data in Enterprise and Science MELODIES (http://www.melodiesproject.eu/) que lidera la Universidad de Reading en asocio con 16 instituciones de ocho países. El objetivo del proyecto es maximizar el aprovechamiento de los datos abiertos gubernamentales y los enlazados por las empresas y la ciencia (ver Figura 19). Además, busca combinar múltiples fuentes de datos incluyendo los de observación de la tierra y gubernamentales para desarrollar servicios de beneficio social; aprovechar los datos enlazados para descubrir, interpretar, relacionar y generar nuevo conocimiento. Para lograrlo se garantiza la calidad en los datos abiertos, impulsando también la creación de e-infraestructuras que faciliten el reúso de los mismos y la demostración de su valor, para estimular la creación y el lanzamiento de nuevas plataformas gubernamentales que provean este tipo de datos.

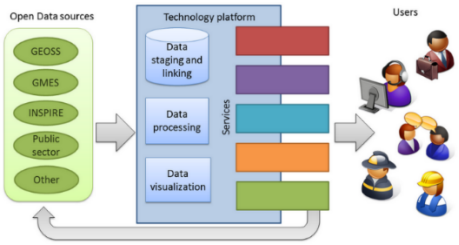


Figura 2. Arquitectura del proyecto MELODIES

Esta arquitectura la conforman: 1) las fuentes de datos de las diferentes entidades proveedoras del sector científico, el sector público y otras entidades; 2) la plataforma de almacenamiento de datos abiertos y enlazados, además de los servicios de procesamiento y visualización de datos; y 3) los servicios a usuarios para la consulta y utilización de los datos. El proyecto cuenta con un *workflow* (ver Figura 20) para la integración de otros elementos necesarios, tales como sustentabilidad (costo-beneficio de los datos abiertos, iniciativas y políticas europeas, evolución tecnológica y marco legal), diseminación (grupo líder del proyecto, usuarios de la comunidad, personal técnico y público en general), y de cuya integración se encarga la plataforma tecnológica.

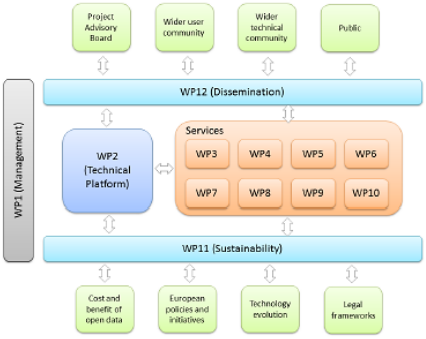


Figura 3. *Workflow* del proyecto MELODIES

Como dato adicional a la evolución y utilización de los datos abiertos en Inglaterra, Bates (2012) presenta un recorrido por la historia y aborda sus inicios desde 2009. Para este autor, las fuerzas sociales y los intereses de quienes han trabajado en el diseño de la iniciativa de datos abiertos gubernamentales en este país, los llevaron a ser adoptados como política principal por la coalición que llegó al poder en 2010, y que estaba integrada por demócratas y liberales. Además, involucraron esta directiva como herramienta para responder a la crisis económica y política del proyecto neoliberal, por lo que la domesticación de estos datos se convirtió en herramienta para el control de la sociedad civil y para mejorar el estado neoliberal. Esta es una muestra de cómo utilizar los datos abiertos para controlar la información de lo que sucede en el gobierno, pero con recursos del mismo país. Bates también considera que una infraestructura de datos abiertos gubernamentales deber estar orientada a la producción de recursos técnicos y de información, de tal forma que se garantice el su reusó. Porque el consumo y reutilización de datos se debería orientar al soporte de aplicaciones del medio ambiente (consumo de energía, emisiones de CO_2), ocio, servicios de información pública (academias, bibliotecas, escuelas), transporte y a la comprensión de los procesos políticos y económicos de la nación.

3. Alemania

El Instituto Fraunhofer desarrolló la arquitectura Fraunhofer Institute for Open Communication Systems (FOKUS) (http://open-data.fokus.fraunhofer.de/) (Figura 21) para recolectar metadatos de varios catálogos, entre los que se encuentra el Portal de Información Ambiental de Alemania, el portal de datos estadísticos Destatis, el portal de datos de Berlín y el portal de datos de Bremen. Además, la arquitectura publica y comparte los datos gubernamentales abiertos a través de portales tales como Content Management System Content Management System (CMS) Liferay (metadatos), Comprehensive Knowledge Archive Network (CKAN) (catálogo y recolector de metadatos), y mediante las aplicaciones móviles UmWelt Zone, Die EFRE App, Gesetze App y Bremen Nord 2GO entre otras además de los navegadores de internet.

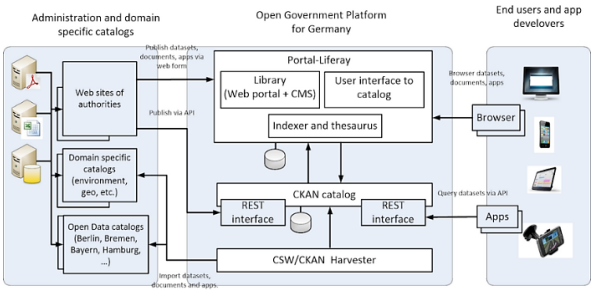


Figura 4. Arquitectura de la plataforma de datos abiertos gubernamentales de Alemania (FOKUS)

Para la estructura de metadatos utiliza el esquema Open Knowledge International (OFKN) que se usa como buena práctica en Europa y América, y que es el estándar de facto para CKAN; para el intercambio de datos utiliza JavaScript Object Notation (JSON), una estructura que soporta conjuntos de datos, documentos y aplicaciones. El portal ofrece los siguientes conjuntos de datos de información estática: leyes (30%), economía y demografía (12%), ambiente y geografía (11%), educación (6%), transporte urbano (5%) y otros tales como economía, construcción y hogar (36%); y de información en tiempo real: movilidad y transporte urbano (28%), turismo (3%), energía y utilidades (3%) y ambiente, geografía y meteorología (61%). En el ciclo de vida de los datos abiertos en esta arquitectura (Figura 22) intervienen dos actores: 1) los consumidores (descubren, enriquecen y consumen datos a través de Apps), y 2) los dueños de los datos (CMS gubernamentales y sitios web). El objetivo es ofrecer los servicios de refinamiento de datos, portal de datos, catálogo de datos y almacenamiento de datos enlazados, además de los conjuntos de datos que los propietarios le entregan.

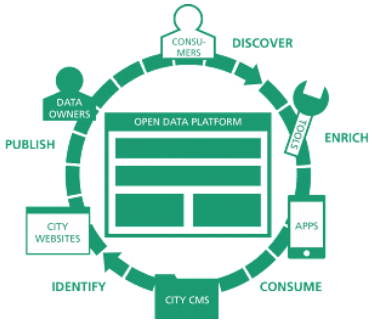
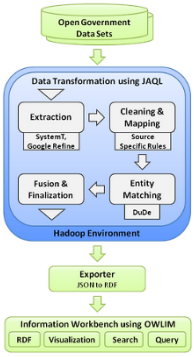


Figura 5. Ciclo de vida de los datos abiertos en ODP (FOKUS)

Por otro lado, para (Böhm et al., 2012) muchas organizaciones gubernamentales publican los datos en internet con el objetivo de garantizar la transparencia y cumplir sus obligaciones en relación con el derecho a la información que tienen los ciudadanos. Pero, frecuentemente, se encuentran diferencias entre ellos a nivel de contenido, formato, estructura y calidad, cuando son publicados usando *linked data*. Por eso surge la necesidad de publicar los datos usando Linked Open Data (LOD) y formato RDF, de tal manera que se facilite su reutilización.

Esto disminuye los problemas de integración y de acceso, y permite la publicación de datos limpios que pueden ser analizados, visualizados e



interconectados. Este es el objetivo del proyecto Government Web Data Integration for Linked Data (GovWILD) (Figura 23) desarrollado por Estados Unidos y Alemania, cuya función es enlazar los *datasets* que poseen datos y características comunes acerca de agencias, personas y compañías asociadas con el gasto público en estos países. La plataforma recolecta datos en los formatos CSV, TSV o XML, mediante *clusters hadoop*, luego realiza el mapeo semítico de fuentes de datos con esquemas diversos y los convierte a JSON, para luego analizar los datos usando lenguaje JAQL.

Figura 6. Arquitectura del proyecto GovWILD (GovWILD)

4. Brasil

El país ingreso como miembro del Open Government Partnership (http://www.opengovpartnership.org/) en 2011 (Breitman, Salas, Casanova, y Saraiva, 2012), con el objetivo de garantizar la transparencia pública; luego hizo el lanzamiento del portal de datos abiertos gubernamentales interno (www.dados.gov.br); y posteriormente conformó un comité encargado de la creación del catálogo centralizado de datos, una iniciativa que estaba disponible para el presidente y su grupo de consejeros. Dado el éxito que logró, fue puesta al servicio público en 2010 y los datos se clasificaron en dos dimensiones: 1) por territorio (estados y ciudades), y 2) por tiempo (año o mes), luego se crearon jerarquías más específicas tales como infraestructura, ciudadanía e inclusión social. Los datos se almacenan en una base de datos relacional y se publican usando los formatos XML y JSON, aunque ninguno soporta Linked Open Data para generar valor agregado. Esta iniciativa ha generado un trabajo colaborativo orientado a la construcción de la Infraestructura Nacional de Datos Abiertos, en la que el gobierno trabaja con representantes de todos los sectores con el objetivo de definir un vocabulario común para representar los datos. Se ha experimentado con Triplify para convertir la base de datos del portal a RDF y comprobándolo en DBPedia, GeoNames, DataCube. El reto con el que se han encontrado es promover internamente el uso de vocabularios RDF estándar, y a nivel internacional definir una estrategia para crear una plataforma central que permita recolectar metadatos para ser mapeados usando el esquema de metadatos para *datasets* Asset Description Metadata Schema (ADMS).

5. Chile

Walker (Walker, 2012) desarrolló una aplicación para socializar los datos abiertos gubernamentales del del Ministerio de Educación Chile, logrando crear una aplicación para visualizar todas las escuelas y sus puntajes. Esta iniciativa aporta en la búsqueda de transparencia en un gobierno de unidad con libertad de

información, que en el país se conoce como la Ley de Transparencia. La aplicación se desarrolló utilizando las librerías OpenLayers y OpenStreetMap de JavaScript, para resaltar los datos mediante marcadores coloridos.

6. España

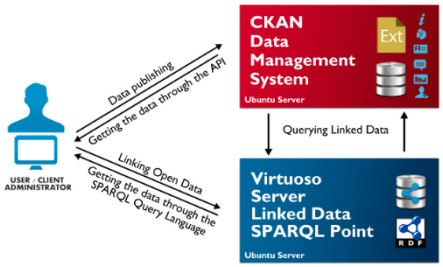


Figura 24. Arquitectura OGD de la UPF (UPF, 2014)

La Universitat Pompeu Fabra de Barcelona (UPF, 2014) desarrolló una infraestructura de OGD implementada bajo CKAN como plataforma para gestionar conjuntos de datos gubernamentales, además de publicar contenido en el portal de datos abiertos (Figura 24). También se utilizó Virtuoso como herramienta para la publicación de *linked data* y todo se instaló sobre Linux Ubuntu como sistema operativo. Esta arquitectura permite el consumo de conjuntos de datos desde CKAN y el envío de datos por medio de Virtuoso, estableciéndose una relación entre ambos, al tiempo que los usuarios pueden acceder por medio de la API de CKAN a los conjuntos de datos, e ingresar *linked data* hacia Virtuoso y hacer consultas en lenguaje SPARQL.

El objetivo de esta plataforma es mejorar la transparencia con los estudiantes, docentes y personal público en general, Su meta es crear finalmente una infraestructura que permita compartir conjuntos de datos comunes entre todas las universidades europeas. En la Figura 25 se observa el ecosistema de la plataforma en el que se involucran entes administrativos de la UPF, tales como generadores de conjuntos de datos y plataforma OGD usando CKAN y Virtuoso que, con el análisis de los datos, permiten generar ideas innovadoras, proyectos de investigación, desarrollo de aplicaciones, ideas de negocios y proyectos, sirviendo como posibilidad para desarrollar nuevas fuentes de ingreso para la universidad.

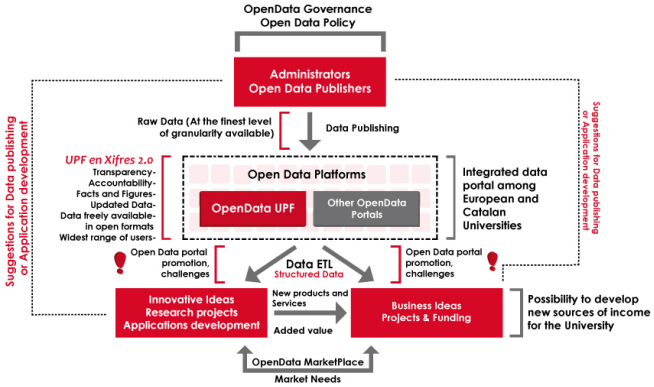


Figura 25. Ecosistema de la plataforma de OGD de la UPF (2014)

7. Holanda

van Grieken (2011) propone un modelo de negocio de datos abiertos gubernamentales en el que analiza el impacto en los diferentes interesados (Figura 26). La arquitectura del modelo está compuesta por entidades de gobierno que aportan sus fuentes de datos y los almacenan para uso interno. Entre las entidades se el Ministerio de Transporte, Catastro y los operadores de las redes de transporte público, y cuando otro usuario requiere estos datos los debe comprar para crear productos o servicios o para crear soluciones para Smart Cities, tal como sucede en ciudades como Nueva York y San Francisco con la creación de servicios de rutas, monitoreo y mapas. Este modelo requiere la definición de una estrategia de precios y una legislación sobre derechos y propiedad intelectual, de forma que los datos se puedan entregar fácilmente a los clientes, y lo mismo que la definición de una estructura que permita su entrega directamente o por medio de servicios web.

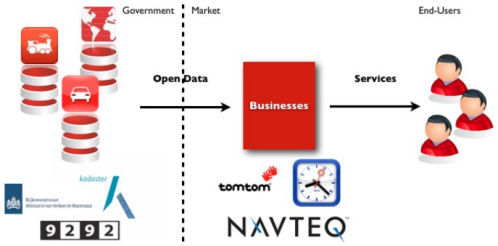


Figura 26. Arquitectura de modelo de negocios de datos gubernamentales abiertos (van Grieken, 2011)

8. Grecia

Strategic Planning Bureau del General Secretariat of the Prime Minister de Grecia adoptó e implementó CKAN como plataforma para datos abiertos gubernamentales (Alexopoulos, 2013; Data.gov.gr, 2014). Actualmente cuenta con 38 conjuntos de datos, 199 recursos, 9 aplicaciones y 20 organizaciones, tales como ministerios, prefecturas y municipalidades, que trabajan con el paradigma de datos abiertos enlazados y datos gubernamentales abiertos, manteniéndolos en plataformas que tienen funcionalidad semántica y tecnológica. Al igual que en los demás países, busca crear valor agregado social y comercial a partir del conocimiento del análisis de los datos, y crear aplicaciones, productos y servicios a partir del reúso de los datos por parte de empresas y ciudadanos.

Para recolectar sus datos, esta iniciativa realizó un análisis desde varias perspectivas a las fuentes de datos gubernamentales: 1) *técnicamente* analizaron las herramientas, plataformas y API existentes (Wordpress, Custom, Joomla, PostGIS, Altec CMS, QCMS, DotNet Nuke, Drupal y Liferay); 2) *semánticamente* analizaron aspectos tales como multilingüismo (alemán, francés, inglés y griego), estándares de metadatos, formatos de los datos (shp, open office, json, png, jpg,

kml, html, cvs, gml, ods, xml, pdf, xls y doc), licencias de los datos y datos enlazados; 3) *funcionalmente* hallaron que el 93% de las herramientas no ofrecían visualización de datos. Por otro lado, encontraron que, semánticamente, el 23% de los datos soportaba RDF, 23% SPARQL y que el 54% no soportaban ninguno de ellos. Para la interfaz, la mayoría de herramientas usaban PHP, query, ASP.NET, Mootools y SWFObject. En cuanto a servidores en los que se encontraron instaladas las herramientas, tienen Apache, IIS, D2R y Nginx, además, las interfaces utilizadas eran RESTful y Google Font API.

9. Singapur

El portal http://data.gov.sg es una plataforma centralizada para la búsqueda y acceso de *datasets*, con el compromiso de entregar datos gubernamentales de forma pública. Inicio en 2011 con más 5.000 *datasets* de 50 agencias del gobierno, que se ofrecen para para investigación y desarrollo de aplicaciones. Para finales de ese año número de *datasets* había sobrepasado los 6.000, algunos de acceso gratuito y otros bajo pago (Chan, 2013). El objetivo principal es hacer que los conjuntos de datos estén disponibles mediante una API para su lectura y modificación, de forma que los programadores puedan desarrollar servicios electrónicos como aplicaciones Web o aplicaciones móviles. La Figura 27 es un esquema conceptual de la infraestructura y la forma como se integra el consumo de datos gubernamentales mediante servicios electrónicos para facilitar el acceso a los usuarios.

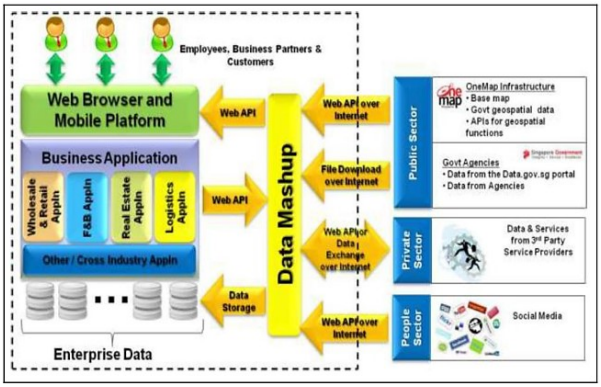


Figura 27. Arquitectura de la plataforma de OGD en Singapur (Chan, 2013)

10. Uganda

Inicialmente, Uganda rechazó la invitación para ingresar a Open Government Partership (OGP), a la que pertenecen Kenia, Ghana, Suráfrica, Liberia y Tanzania (Larsson, 2013). Posteriormente, desarrolló su plataforma de datos abiertos gubernamentales con el objetivo de promocionar la apertura de datos, y actualmente cuenta con conjuntos de datos en los formatos XLS, DOC y PDF, lo que

se ha convertido en desventaja porque la mayoría de información se ofrece en formatos propietarios. El proyecto básicamente está compuesto por una arquitectura con CKAN como gestor de conjuntos de datos y Open Aid como CMS.

La plataforma también posee otras fuentes de datos, tales como el Banco Mundial, el Ministerio de Agricultura de Uganda y la Oficina de Estadísticas de Uganda, las cuales aportan 49 conjuntos de datos en formato XLS, HTML, CSV, XLSX, SHP y PDF. La plataforma requiere mejoras operativas porque, cuando las visualizaciones de conjuntos de datos se hacen manualmente, se deben descargar en Microsoft Excel para luego graficarlos.

11. Resultados

El sector público es la más grande fuente de información en Europa y se espera que, en las próximas décadas, la explotación de datos masivos en infraestructuras públicas aporte para lograr una economía clara que favorezca el uso de datos gubernamentales por parte de científicos, empresas y ciudadanos. Por otro lado, el abordaje y la discusión sobre Open Government Data y el ofrecimiento de datos como política nacional, crece en Estados Unidos, Reino Unido y el resto de Europa. Desde 2003, los países miembros de la Unión Europea definieron una directiva para el reúso de información pública, que se conoce como PSI Directive. De esta manera eliminaron barreras y garantizaron transparencia y competencia limpia mediante el análisis de estos datos, lo que la deriva en conocimiento y análisis estadísticos. En todos estos países se ha identificado tres tipos básicos de fuentes de datos abiertos:

1. Catálogos de OGD, que permiten el acceso centralizado a los datos, a través de metadatos que describen el contenido ofrecido a los usuarios.
2. Plataformas web para compartir datos, que almacenan e indexan información estructurada, además de permitir el registro, búsqueda y envió de información a plataformas tales como CKAN, Microsoft OGDI y Freebase.
3. El sector público, al que pertenece la mayoría de fuentes en portales tales com Wordpress, Joomla, Liferay, Altec CMS, QCMS y Drupal.

Los datos almacenados en estas fuentes son:

1. Comerciales: registros de empresas patentes.
2. Geográficos: catastro, edificios, fotos aéreas, geodesia, geología, topografía e hidrografía.
3. Legales: de cortes nacionales e internacionales.
4. Sociales: de economía, gasto público, empleo, salud, población, administración pública y sociedad.

5. Meteorológicos: del clima y los pronósticos.
6. Transporte: de congestión del tráfico, trabajo en las vías, transporte público y registro de vehículos.

Para la creación de estas fuentes de datos se debe seguir una metodología que garantice que cada una puede proveer acceso a los *datasets*, lo que requiere diversos análisis:

1. Funcional: se aborda aspectos relacionados con las funcionalidades y utilidades que se ofrecerá a los usuarios en la plataforma: visualización y análisis estadístico de los datos.
2. Semántico: para conceptos de la representación y la estructura de datos abiertos, desde cómo mejorar estructuras de conocimiento simples, planas y enlazadas.
3. Tecnológico: aspectos relacionados con productos y plataformas utilizadas para la implementación de infraestructuras OGD, que incluyen utilidades para adquisición, entrega y visualización de datos, y observaciones de los usuarios.

Asimismo, la mayoría de proveedores de datos OGD, equivalente al 85%, tienen capacidades de almacenamiento, de listados y de descarga de *datasets*, y unos pocos tiene funcionalidades de visualización y catalogado de *datasets*.

A continuación, se presenta una descripción del modelo de datos abiertos gubernamentales en Colombia, que contiene apuntes desde las perspectivas tecnológica, sociocultural, organizacional, político-legal y presupuestal; y que se centra en el Modelo de Contexto del Soporte Tecnológico para Implementar el modelo de datos abiertos en el país.

1. Descripción del modelo

De acuerdo con MINTIC (2014a), la tendencia en el acceso a los datos abiertos, y la creación de proyectos que garanticen el acceso a la información, se encuentra organizada por sectores e iniciativas gubernamentales. Entre sus responsabilidades se encuentra la administración de los datos abiertos y la formación de personal de TI que genere valor e implemente infraestructuras de datos abiertos gubernamentales. En este orden de ideas se busca que las entidades sigan el marco de referencia para lograr un adecuado intercambio de datos, de acuerdo con lo estipulado en el artículo 230 de la Ley 1450 de 2011, que define al Ministerio de las Tecnologías de la Información y las Comunicaciones como la entidad autorizada a nivel nacional para el trámite de todo los relacionado con Open Data. Además, también legisla en cuanto a requisitos y órganos de control para la reutilización y transformación de datos abiertos.

Por otro lado, desde la Ley 57 de 1985 se expresan las sanciones para un empleado público que no proporcione la información requerida a quien la solicite. Los datos que se pueden publicar de forma abierta son: los relacionados con la seguridad nacional, investigaciones penales y datos personales. El Modelo de Datos Abiertos en Colombia involucra componentes estratégicos, tácticos, operativos y de soporte, tal como se observa en la Figura 28.

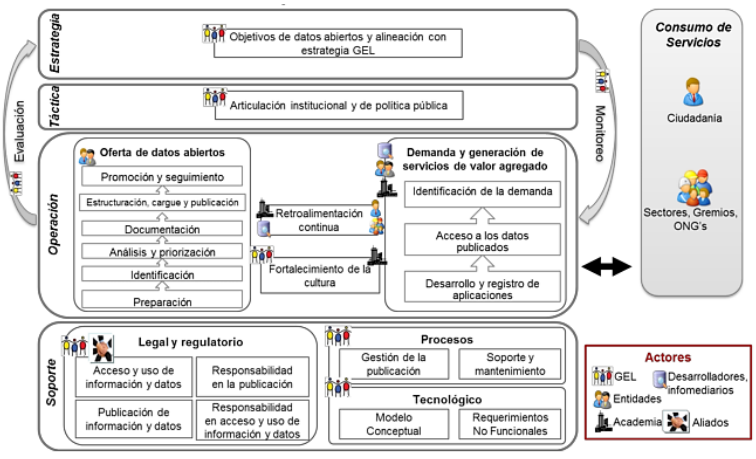


Figura 28. Modelo de Datos Abiertos en Colombia (CINTEL, 2011)

El Centro de Investigación de las Telecomunicaciones (CINTEL), que acompaña al Ministerio en la planeación e implementación del modelo, definió los lineamientos para Colombia mediante un estudio de mercado; además, definió actores, indicadores, metadatos, atributos de calidad, apoyo para el diseño, desarrollo e implementación de soluciones tecnológicas, capacitación, acompañamiento y gestión de la iniciativa en las entidades (CINTEL, 2011). Todo el proceso ofrece un catálogo de datos abiertos que responde a principios de transparencia, de acceso a la información pública y de participación; con parámetros de calidad tales como exactitud, totalidad, oportunidad, consistencia y formatos permitidos (Figura 29).

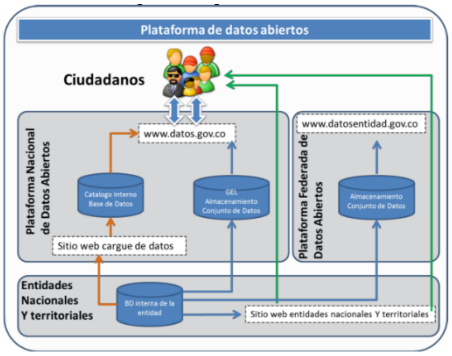


Figura 29. Catálogo de datos abiertos en Colombia (CINTEL, 2011)

1.1 Proceso de apertura de datos

Figura 30. Proceso de Apertura de Datos en Colombia (MINTIC, 2012)

- *Identificación de la información*. La información debe ser pública y ofrecer: nombre, descripción del contenido, área responsable, tipo (económica, comercial, ambiental, agrícola y pesquera, social, seguridad, política, científica, geográfica, transporte y trafico), ámbito geográfico, idioma, fuente primaria, demandante, salida (documento físico, medio digital, sistema, formato audiovisual), frecuencia de generación y formato.

- *Análisis de la información*. Se identifica y separa la que puede ser publicable y no tenga restricciones jurídicas, tales como datos publicables y no-publicables: actuaciones o expedientes políticos, defensa nacional y seguridad, política monetaria, investigación de delitos, material clasificado, información registral); información personal semiprivada: reservada y confidencial que afecte al país o a las personas (relaciones internacionales, vida, dignidad, seguridad o salud de las personas, propiedad industrial y derecho de competencia, descubrimientos científicos, tecnológicos o culturales, estrategias de negocio, competitividad, expansión, historias clínicas con ubicación geográfica, secreto profesional, entre otros valores).

- *Priorización de los datos*. Los que se publicarán primero según urgencia y cronograma de publicación a corto, mediano y largo plazo, con base en criterios de alto impacto y menor dificultad de publicación para la entidad.

- *Documentación*. Se realizad la descripción de los metadatos comunes y específicos de los conjuntos de datos a publicar, garantizando la interoperabilidad, extensibilidad, modularidad sintáctica y semántica, utilidad en la búsqueda y simplicidad.

- *Carga y publicación.* Se define el mecanismo de publicación de los datos en el Catálogo de Datos del Estado.

1.2 Formatos

La Plataforma Nacional del catálogo de datos abiertos en Colombia permite el manejo de los formatos XLS, ODF, CSV, XML, GEL-XML, JSON, TXT, RDF-XML, KML-KMZ, SHP, TMX, ZIP (MINTIC, 2012), que provienen de las entidades nacionales y territoriales. Conformando una plataforma federada de datos abiertos gubernamentales, que se ofrecen en el catálogo de datos abiertos Colombia. También se definieron los aspectos de los metadatos (Figuras 31, 32 y 33), calidad y exactitud de los datos, para realizar la correspondencia entre ellos y el sector que representan. De esta manera se garantiza su completitud y oportunidad, es decir, que los datos se actualicen constantemente, además, que sean consistentes porque guardan relación con sus antecesores.

Tipología	Metadato	Descripción	Tipos de datos
Metadato común	Título	Nombre concreto del *dataset*	Texto
	Identificador	Identificador único del *dataset* en la base de datos	Texto (URI), numérico (ISBN, DOI, número local)
	Descripción	Descripción/resumen del *dataset*	Texto
	Categoría	Nombre(s) de categoría(s) o grupo(s) al cual pertenece el *dataset*	Texto/lista de texto
	Autor	Entidad y dependencia que origina el *dataset*	Texto
	Idioma	Idioma de los datos del *dataset*	Texto
	Audiencia	Audiencia objetivo del *dataset*	Texto
	Palabras clave	Palabras clave para la ser identificados en la búsqueda	Texto separado por comas
	Ámbito Geográfico	Identifica la catalogación de la información a nivel: Nacional, Departamental, Municipal, Distrital.	Texto/lista de texto
Metadato específico	Fecha Publicación	Fecha de la carga del *dataset* en el catálogo de datos	Fecha (AAAA- MM-DD)
	Fecha Última Actualización	Fecha de última actualización del *dataset* en el catálogo de datos	Fecha (AAAA- MM-DD)
	Frecuencia Actualización	Frecuencia con la que se actualizan los datos del *dataset* en el catálogo de datos	Texto
	Fecha Inicio	Fecha desde la cual se tienen en cuenta datos para el *dataset*	Fecha (AAAA- MM-DD)
	Fecha Fin	Fecha hasta la cual se tienen en cuenta datos para el *dataset*	Fecha (AAAA- MM-DD)
	Número de Versión	Número de la última versión del *dataset* en el catálogo de datos	Texto/numérico
	URL Descarga	URL en el que el *dataset* puede ser descargado	Texto
	Formato Descarga	Formatos en los que se encuentra disponible el *dataset*	Texto/lista de texto
	Tamaño Descarga	Tamaño del dataset en Kb	Numérico
	URL Esquema	URL de acceso al esquema, diccionario de datos o descripción de los datos y la estructura del *dataset.*	Texto
	URL Documentación	URL a documentación adicional sobre el *dataset*	Texto
	URL Programa	URL a sitio web de la entidad que describe los programas o aplicaciones (hechos o por hacer) relacionados con el *dataset*	Texto
	Etiquetas	Palabras clave del contenido del *dataset*	Lista de texto
	Licencia	Tipo de licencia	Texto
	URL Licencia	URL de enlace a la descripción y detalles de la licencia	Texto
	Nombre Responsable	Nombre de la persona que administra el *dataset*	Texto
	Email Responsable	Email de la persona que administra el *dataset*	Texto
	Adicionar si es un inventario geográfico		
	Cobertura de *DataSet*	Delimitador de coordenadas según la NTC 4611	Numérico
	Metadatos adicionales	Describir metadatos adicionales según la NTC 4611	Texto

Figura 31. Estructura de Metadatos Específicos (MINTIC, 2012)

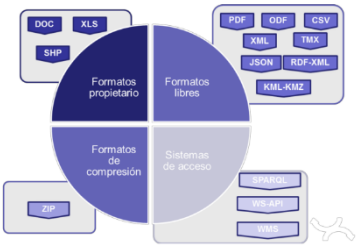


Figura 832. Formatos de conjuntos de datos (MINTIC,

Figura 33.7 Formatos de apertura de datos (MINTIC, 2012)

1.3 Escenarios de publicación

La plataforma de arquitectura de datos abiertos del Catálogo Colombiano incluye a las entidades nacionales y territoriales, las cuales podrán tener plataformas federadas de datos abiertos y elegir entre estos medios de publicación, seleccionado entre cuatro escenarios posibles: 1) carga de archivo, 2) referencia, 3) transferencia, y 4) servicio de información en línea, como se observa en la Figura 34.

Figura 34. Escenarios de Publicación de Datos Abiertos (MINTIC, 2012)

En la carga y si la entidad no tiene servidor para alojar los conjuntos de datos, puede usar el mecanismo del Catálogo de Datos (Figura 35).

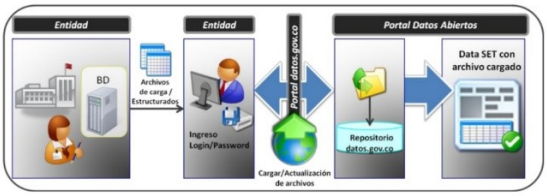


Figura 35. Carga de conjunto de datos MINTIC, 2012)

Para la referencia y debido a que la entidad posee un servidor que aloja los conjuntos de datos, el Catálogo de Datos Abiertos ofrece una URL pública en el servidor (Figura 36).

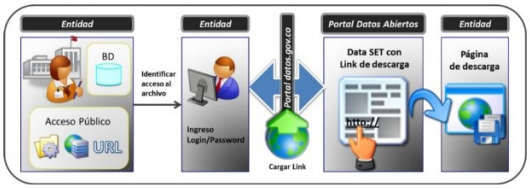


Figura 36. Referencia de archivo de set de datos (MINTIC, 2012)

En la transferencia la entidad realiza una transferencia automática de los conjuntos de datos al Catálogo de Datos (Figura 37).

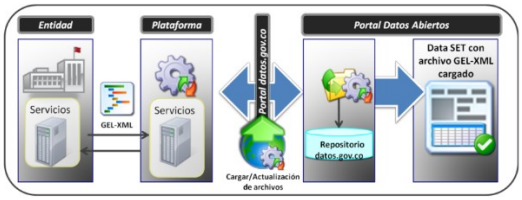


Figura 37. Transferencia de datos (MINTIC, 2012)

Por su parte, para el servicio de información en línea la entidad tiene un servidor o página web, donde ofrece los conjuntos de datos a la vez que los publica en el Catálogo de Datos Abiertos (Figura 38).

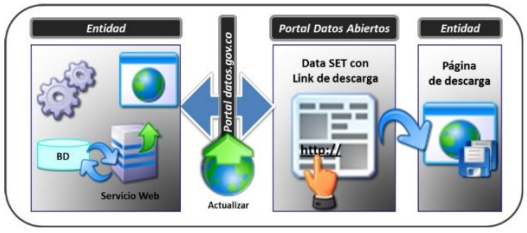


Figura 38. Servicio de Información en Línea (MINTIC, 2012)

1.4 Instalación y uso del cargador de conjuntos de datos

El cargador les permite a las entidades, que no tiene servidor de datos abiertos, subir los conjuntos de datos al Catálogo de Datos, esto incluye el conocimiento de los conceptos de conjunto de datos, propiedades, ficha técnica, conjunto histórico de datos, conjunto no-histórico, columnas y clave de agrupación y de fila, tal como se observa en la Figura 39.

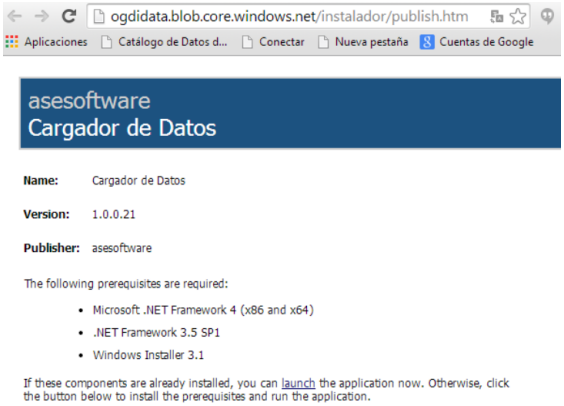


Figura 39. Instalación del cargador de datos (MINTIC, 2011)

Los requerimientos de instalación son los siguientes:

- Windows Installer 3.1
- Microsoft.NET Framework 4 (x86 y x64)
- .NET Framework 3.5 Service Pack 1
- Internet Explorer 8 o superior
- el conjunto de datos debe estar en formato CSV.
- La primera fila debe contener los nombres de las columnas.
- Se debe tener abierto el puerto 1433 (SQL Server).
- Se debe solicitar el usuario y contraseña para la entidad.

Una vez instalado se puede cargar los conjuntos de datos o su URL al Catálogo de Datos (Figura 40), diligenciando las opciones de ficha técnica, propiedades y columnas.

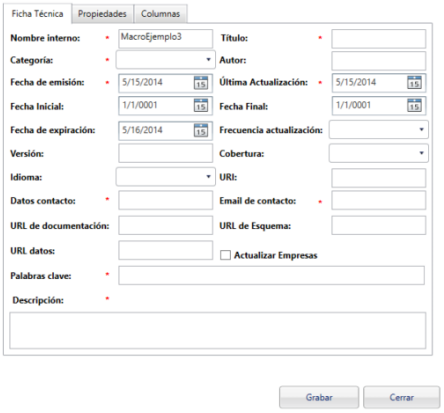


Figura 40. Ficha técnica del cargador (MINTIC, 2011)

1.5 Interfaces para el consumo de conjuntos de datos

MINTIC ofrece un código (Figura 41) para utilizar el servicio de consulta de conjuntos de datos para que los desarrolladores lo utilicen para consultar datos abiertos, encontrar contenedores, encontrar conjuntos de datos publicados, consultar los datos de un conjunto de datos y aplicar filtros en ATOM o JSON (MINTIC, 2013).

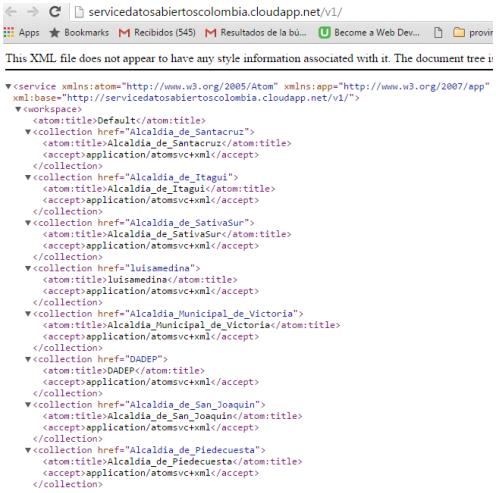


Figura 41. Colecciones de entidades Nacionales y Regionales (MINTIC, 2014a)

Otro código (Figura 42) para consumir datos desde el Catálogo de Datos y además consultar colecciones y conjuntos de datos.

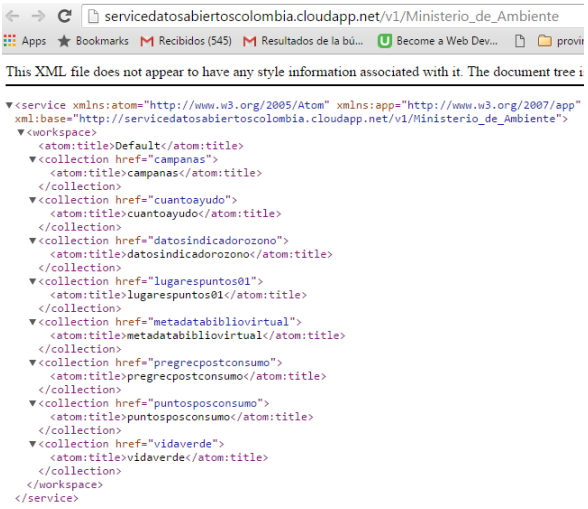


Figura 42. Conjuntos de datos Ministerio de Ambiente (MINTIC, 2014a)

1.6 Modelo de contexto y metadatos

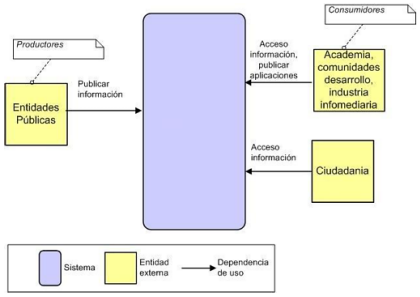


Figura 43. Modelo de Contexto del Soporte Tecnológico (CINTEL, 2011)

El diseño del modelo de datos abiertos en Colombia se basa en la adopción de principios socio-culturales, político-legales, tecnológicos, organizacionales y presupuestales, cuya creación se plasma en un modelo de contexto de soporte tecnológico (Figura 43), para implementarlo en las entidades públicas, como productoras y publicadoras de datos, y la ciudadanía, la academia, las comunidades de desarrolladores de software y la industria, como consumidoras de datos abiertos, mediante aplicaciones móviles y catálogos de datos abiertos.

El modelo depende del personal de TI para la implementación, garantizando el uso de los datos abiertos gubernamentales por parte de los ciudadanos, ofreciendo inclusión a las entidades y centralizando todos los datos del gobierno en un solo catálogo. La iniciativa está orientada hacia un mapa de ruta de cuatro niveles: inicial, básico, avanzado y mejorado, que van desde la puesta en funcionamiento de la infraestructura básica, hasta el uso de herramientas de la web semántica, permitiendo inteligencia de negocios en los datos y dándole valor agregado a los datos publicados. Para lograr esto se construirá una plataforma centralizada,

debido a que las diferentes entidades nacionales y regionales tendrán catálogos propios. Además, se colectarán mediante una plataforma federada por medio de sus metadatos (MINTIC, 2012).

1.7 Vista conceptual del catálogo

En esta vista (Figura 44) se definen las funcionalidades que debe soportar: presentación y descubrimiento, administración de datos y aplicaciones, mantenimiento y administración de la plataforma, consulta SPARQL, calendario de eventos, comentarios y votación de datos, comentarios y votación de aplicaciones, PQR, notificaciones, *blogs* y foros, transformación inter-formatos (XML a RDF, RDF a XML, CSV a RDF, XLS a CSV, CSV a XLS y TXT a CSV), integración de aplicaciones mediante API, integración con Redes Sociales y generación de reportes en PDF,HTML y Excel.

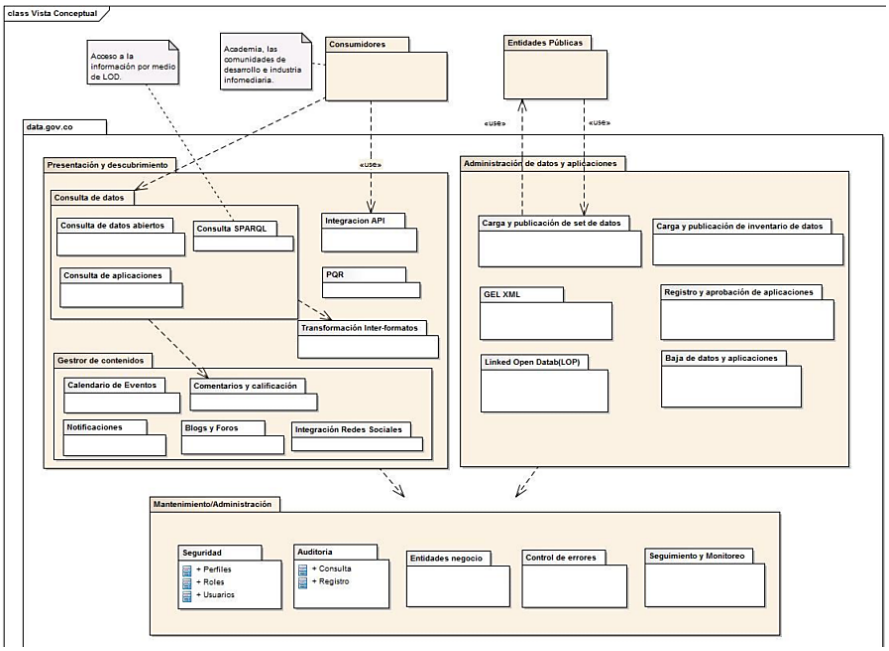


Figura 44. Vista Conceptual del Catálogo (CINTEL, 2011)

1.8 Vista lógica del catálogo

La vista lógica permite definir la plataforma mediante capas: presentación, lógica de negocio (interfaz de consulta, interfaz de gestión datos y aplicaciones, interfaz de redes con redes sociales, comentarios, notificaciones), integración para transferencia de datos XLS, ODF, CSV, XML, JSON, TXT, RDF, KML-kMZ (metadatos, RDBMS, ETL, GEL, batch), servicios de soporte seguridad, auditoria, control de errores, seguimiento y monitoreo, tal como se observa en la Figura 45.

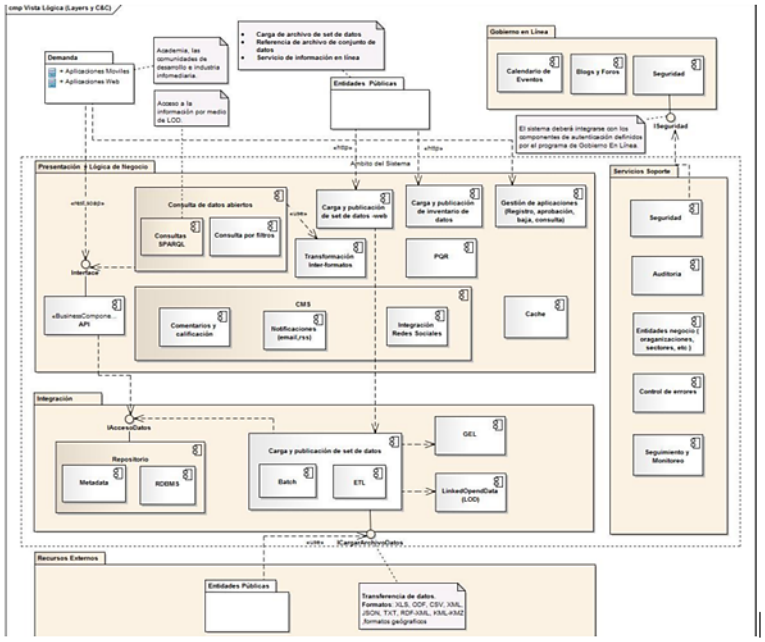


Figura 45. Vista Lógica del Catálogo (CINTEL, 2011)

1.9 Vista de despliegue del catálogo

Esta vista muestra la distribución física de los componentes técnicos del Catálogo, de forma que interactúe por medio de internet con las entidades públicas regionales y departamentales, otras aplicaciones de entidades de Gobierno en Línea y la ciudadanía.

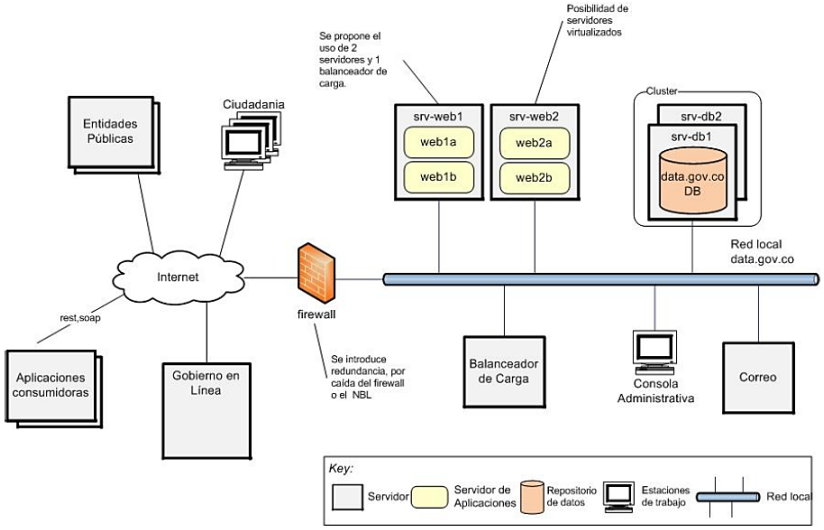


Figura 46. Vista de Despliegue del Catálogo (CINTEL, 2011)

En la Tabla 2 se muestra las características técnicas del Catálogo de Datos Abiertos en Colombia (MINTIC, 2012; 2014c) y CKAN.

Tabla 2. Características técnicas del Catálogo

Característica	**Catálogo Datos**	**CKAN**
Formatos de datos	XLS, ODF, CSV, XML, JSON, TXT, RDF, KML- KMZ	CSV, JSON, RDF
Metadatos para conjuntos de datos		CKAN (Formato nativo para metadatos) DCAT (Estándar para metadatos en un catálogo de datos abiertos)
Interfaz de búsqueda	Sí	Sí
Formato de datos de acceso	JSON, JSONP y AtomPub	JSON
Nombre metadatos conjuntos de datos	Título, Descripción, Entidad, Categoría, Autor, Idioma, URI, Fechas de Emisión, Inicial y Final, Frecuencia, Actualización, Versión, Cobertura, Última Actualización, Expiración, URL Documentación, URL Esquema, URL Datos, Datos de Contacto, Mail Contacto, Def. Columnas	Según definición deseada
Interfaz de publicación	Cargador	Sí
Interfaz de almacenamiento	Cargador	Sí
Gestión de metadatos	Sí	Sí
Interfaz de administración	No	Sí
Federación otros catálogos	No	Sí
Recolección de metadatos de otros catálogos	No	Sí
Uso de API para consumo de datos abiertos y accesibilidad	Open Data Protocol usando interfaces JSON, JSONP y AtomPub. Servicios web REST con el protocolo HTTP, usando GET, POST, PUT, DELETE.	JSON API
Integración con Google	No	Si
Integración con Twitter	No	Si
Integración con Facebook	No	Si
Reporte de nuevos datos	RSS	RSS/Atom
Comentarios datos		Sí
Consulta con SPARQL	No (MINTIC, 2014)	Sí
Indexado conjuntos de datos	Sí	full-text search

La implementación de la estrategia de datos abiertos en Colombia les ha permitido a los actores interesados acceder a los datos fácilmente, lo que ha llevado a que el país mejore sus niveles de transparencia y lograr el ingreso al Open Government Partnership OGP (Figura 47).

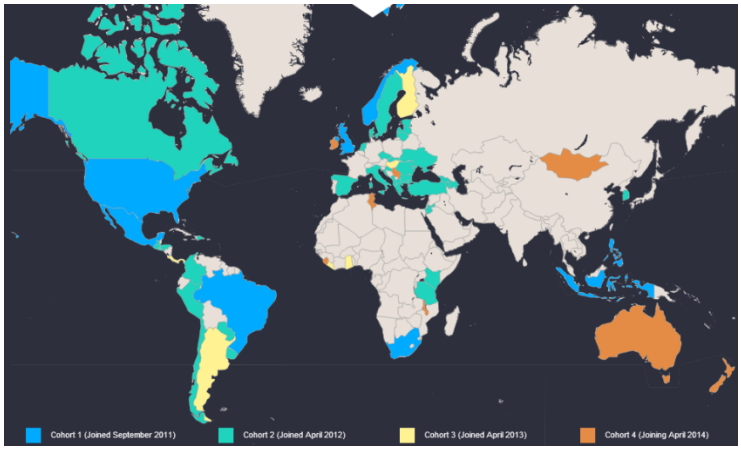


Figura 47. Países pertenecientes al OGP (OGP, 2014)

2. Últimos cambios en el catálogo de datos de Colombia

En la actualidad, el catálogo de datos abiertos gubernamentales fue cambiado a Socrata (Figura 49) según contrato de MINTIC con la empresa ASESOFTWARE, ofreciendo a los consumidores y usuarios un catálogo especializado en datos abiertos. El catálogo tiene restricciones de licenciamiento, lo que representa costos adicionales para su implementación en las entidades públicas del país. Aunque la empresa está desarrollando una versión abierta, no ofrecerá la funcionalidad total del producto comercial, ni la documentación apropiada para su implementación, lo que genera dependencia de los servicios en la nube.

Figura 49. Interface de Socrata en Datos Abiertos Colombia

1. Definición

1.1 Características técnicas de la arquitectura de datos

Primero, debe ser una arquitectura para almacenar, recolectar y publicar datos abiertos (Figura 48) usando un vocabulario que permita la interoperabilidad entre catálogos de datos, a la vez que el consumo de recursos por aplicaciones, de forma que se maximice el valor de la información por medio de la entrega a los ciudadanos. Esto define una instancia privada para la publicación interna de datos abiertos, de tal manera que el recolector los pueda utilizar, exponer y entregar como catalogo público y permitir el consumo de recursos. Además, que permita entender el funcionamiento de una arquitectura multicapa, como la de publicdata.eu, que recolecta data.gov.ro, data.kk.dk, govdata.de, etc., a la vez que recolectan conjuntos de datos de catálogos de entidades de cada país y, finalmente, los expone en el catálogo público de publicdata.eu. Este catálogo agrega los demás, funcionando como manejador de una federación que recolecta de varias instancias, utilizando el estándar de vocabulario para catálogos de datos abiertos DCAT y el formato nativo de CKAN. Por otro lado, en la definición de los formatos de datos se debe tener en cuenta la naturaleza de los análisis de los datos, para facilitar la lectura, generar cálculos y gráficos, y analizar la información desde diferentes equipos de cómputo y dispositivos móviles sin costo. Para esto se prefieren formatos que almacenen los datos mediante filas y columnas de forma estructurada y en formato no-propietario, tales como JSON, CSV y XML.

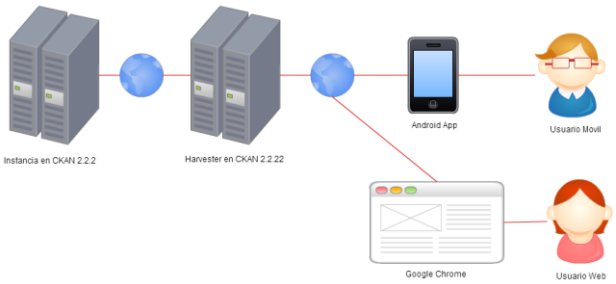


Figura 48. Definición de arquitectura de datos abiertos gubernamentales

2. Implementación

2.1 Creación del prototipo de infraestructura

CKAN cumple con los requisitos funcionales y no-funcionales definidos en la implementación de datos abiertos Colombia, en relación con la implementación de catálogos en las Entidades Nacionales y Territoriales, de conformidad con la

iniciativa INI-01, Diseño, desarrollo e implementación de soluciones tecnológicas y la INI-06, decreto de la implementación de datos abiertos en cuanto al catálogo de datos abiertos. Para esto se selecciona CKAN como plataforma de datos abiertos y por sus características técnicas: interfaz web amigable; API para federación de catálogos e integración con aplicaciones para ofrecer consumo de datos e interoperabilidad con otros catálogos de datos abiertos propietarios, tales como Socrata y Junar, y open source como DKAN; módulo de visualización y análisis de datos; flujo para gestión de grupos; conjuntos de datos para publicación; además, porque es *open source* y utiliza catálogos de datos abiertos tales como data.gov (Estados Unidos), publicdata.eu (Unión Europea) y data.gov.uk (Canadá). Actualmente, en el mundo existen más de 140 catálogos de datos abiertos implementados en CKAN y, de los 431 que aparecen registrados en dataportal.org, en la Tabla 3 se muestran los que existen en Sur América.

Tabla 3. Catálogos de datos abiertos gubernamentales en CKAN en Sur América

Catálogos	URL	País
Alcaldía Mayor de Bogotá	http://datosabiertos.bogota.gov.co/	Colombia
Archivo General de la Nación	http://lida2.archivogeneral.gov.co http://lida.archivogeneral.gov.co/ckan/	Colombia
Red de Ciudades "Cómo vamos"	http://www.n-gage.it	Colombia
Ministerio del Trabajo	http://demo.ckan.org/organization/ministerio-del-trabajo	Colombia
Catálogo de datos abierto de Brasil	http://dados.gov.br/dataset	Brasil
Estado de Alagoas	http://dados.al.gov.br/dataset	Brasil
Senado de Brasil	http://dadosabertos.senado.gov.br/	Brasil
Gobierno municipal de Recife	http://dados.recife.pe.gov.br/	Brasil
Rio de Janeiro	http://data.rio/dataset	Brasil
Porto Alegre	http://datapoa.com.br/dataset	Brasil
Gobierno Nacional	http://datos.argentina.gob.ar/dataset	Argentina
Buenos Aires	http://data.buenosaires.gob.ar/dataset	Argentina
Misiones Argentina	http://www.datos.misiones.gov.ar/dataset	Argentina
Ecuador	http://datosabiertos.ec/	Ecuador
Uruguay	https://catalogodatos.gub.uy/dataset	Uruguay
Paraguay	http://datos.gov.py/	Paraguay

En la Tabla 4 se detalla el software utilizado para la implementación de la arquitectura de datos sobre una distribución Linux Ubuntu Server de 64 bits, versión 12.01, con CKAN 2.2.2 como recolector de metadatos y conjuntos de datos. Estas funcionalidades son soportadas por el *datastorer*, para el almacenamiento de metadatos y por el *filestorer* para el almacenamiento físico de conjuntos de datos. Para esto se debe crear bases de datos por separado para la gestión de metadatos y para gestión de conjuntos de datos en postgreSQL. También se instaló en otro

servidor Linux Ubuntu Server con la misma configuración, que funcionará como instancia del recolector, ubicado en otro servidor, y que ofrece datos a una aplicación en Android llamada FarmaciasGuardia, que permite la lectura de un conjunto de datos XML a la vez que muestra las farmacias que están de servicio.

Tabla 4. Software usado para implementar la instancia y el harvester en CKAN

Software	Versión	Uso en la arquitectura
Linux Ubuntu Server	Description 12.04.5 LTS Release 12.04 Codename precise 64 bits	Instalar los servidores de CKAN para el harvester y la instancia
Comprehensive Knowledge Archive Network (CKAN)	CKAN versión 2.2.2 from source	Instalado desde los archivos fuentes bajo licencia Affero GNU GPL v3.0 y desarrollado en Python
Servidor Web Apache	Apache/2.2.22 Linux Ubuntu	Servidor de aplicaciones CKAN
PostgreSQL	PostgreSQL 9.1.15	Manejador de la bases de datastorer para gestión de metadatos y para el filestorer de CKAN
Servidor Web Jetty	Jetty 6.1.3	Como servidor para instalar Solr
Plataforma de indexado Solr	Solr 1.4.1	Backend para indexar con Lucene 2.9.4 los metadatos de los conjuntos de datos
Servidor REDIS	Redis	Usado por el harvester como memoria cache

Para realizar la recolección de metadatos y archivos entre el recolector y la instancia, se activan *plugins* en la configuración de CKAN, llamados ckan_harvester (para recolectar metadatos CKAN), harvest (para activar el recolector), dcat_rdf_harvester, dcat_json_harvester, dcat_json_interface (para recolectar archivos en formatos .rdf y .json, y activar la interface .json), datastorer (para almacenar metadatos), filestorer (para almacenamiento de conjuntos de datos) y datapusher (para pasar datos del datastorer al filestorer y para almacenarlos en el recolector). Para la indexación de metadatos se usa la plataforma Solr, que funciona sobre Jetty como servidor de aplicaciones web, y para la recolección de metadatos y conjuntos de datos de la instancia se utiliza Redis como servidor de memoria cache, para encolar las respuestas de metadatos o conjuntos de datos de la instancia y pasarlos a la base de datos del datastorer o el filestorer, según sea el caso de recolección y almacenamiento.

Usando el enfoque arquitectónico de Phillipe Krutchen con sus vistas 4+1 (http://philippe.kruchten.com/architecture/), en la Figura 49 se muestran la vista de procesos y la vista de despliegue de la arquitectura, que describen cómo interactúan los componentes de la infraestructura. En la vista de procesos de la Figura 50 se observa los componentes y *plugins* activados en CKAN 2.2.2 para permitir el funcionamiento del catálogo de datos abiertos gubernamentales (UI, DCAT/RDF, Harvester, Datapusher, Datastorer, Filestorer), y que interactúan con

PostgreSQL 9.1.15 (Base de datos de CKAN, Base de datos del Datastorer, Base de datos del Filestorer) para guardar las bases de datos de metadatos, conjuntos de datos y la base de datos por defecto para la instalación de CKAN. El servidor web Jetty 6.1.3 se utiliza para instalar la plataforma de indexado Solr 1.4.1, la cual indexa los metadatos de los conjuntos de datos para agilizar la búsqueda y acceso de los usuarios; harvester utiliza el servidor REDIS como memoria cache, en la que encola en línea los metadatos o archivos que se están recolectando, para luego pasarlos a la base de datos del datastorer en PostgreSQL 9.1.15 o a filestorer; a su vez, el datapusher pasa los conjuntos de datos al filestorer, por medio de los metadatos de la base de datos datastorer.

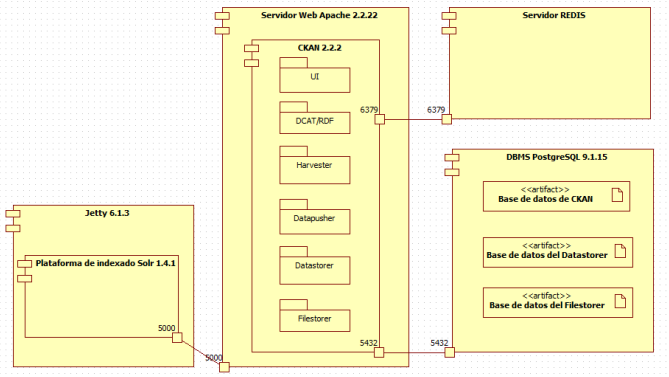


Figura 49. Vista de procesos de la arquitectura de alto nivel para el recolector y la instancia

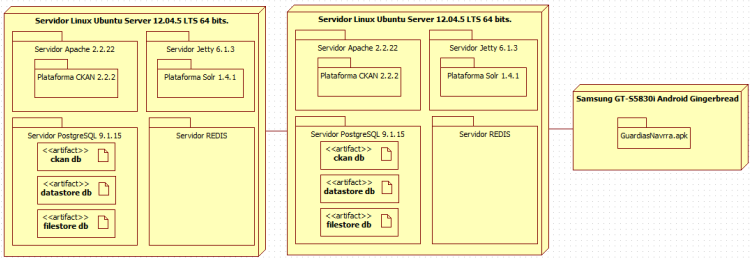


Figura 50. Vista de despliegue de la arquitectura de alto nivel para instancia, recolector y aplicación móvil

La vista de despliegue de la Figura 51 muestra cómo están distribuidos físicamente en la red cada uno de los nodos que interactúan en la arquitectura. La instancia de CKAN, que corre sobre un servidor físico con arquitectura de 64 bits y plataforma Linux Ubuntu Server 12.0.4.5 LTS 64 bits, es recolectada por el harvester y, finalmente, el harvester, por medio de las API, ofrece el recurso a una aplicación Android, que corre en un Samsug GT-5830i con Android Gingerbread 2.3.6 y que consume el archivo GuardiasNavarra.xml. Igualmente, en la vista de despliegue con clientes móvil y web se aprecia que, mediante API, la

infraestructura les permite a los clientes el consumo de recursos de los conjuntos de datos, a la vez que a clientes browsers, como Google Chrome, el acceso a los conjuntos de datos.

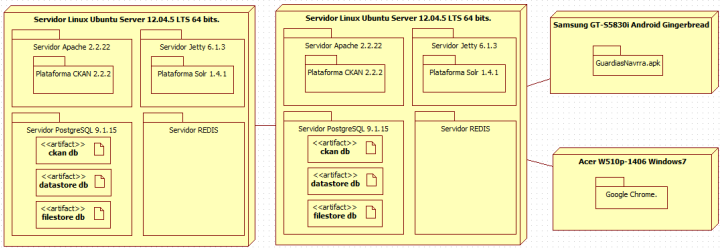


Figura 51. Vista de despliegue de la arquitectura de alto nivel con clientes móvil y web

2.2 Interfaz de usuario del repositorio Harvester

El harvester de la arquitectura de datos abiertos (Figura 52) se configuró para permitir la recolección de metadatos en formato nativo CKAN y Data Catalog Vocabulary (DCAT), un vocabulario diseñado para facilitar la interoperabilidad entre catálogos de datos en Internet, diseñado en DERI, refinado por eGov Interest Group y estandarizado por Government Linked Data (GLD), además, es una recomendación de W3C. En la Figura 53 se observa la Organización Navarra, creada con el conjunto de datos Guardias, que tiene a GuardiasNavarra.xml, un recurso que es consumido por la aplicación en Android.

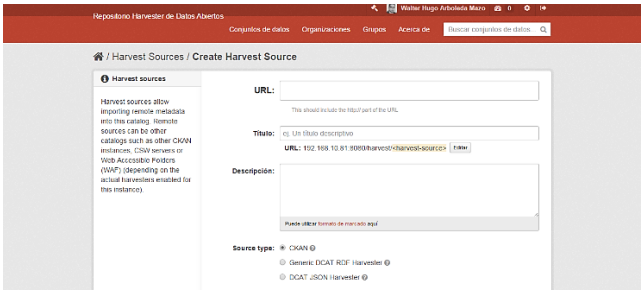


Figura 52. Interfaz de usuario del repositorio Harvester

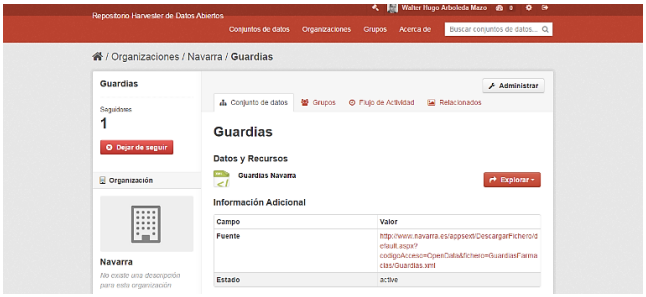


Figura 53. Conjunto de datos recolectado en la arquitectura

2.3 Interfaz de usuario del repositorio Instancia

La instancia de la arquitectura (Figura 54) se configuró para recolectar metadatos en formato nativo CKAN y Data Catalog Vocabulary (DCAT), un recurso en MS Excel recolectado en la instancia y que su vez es recolectado por el Repositorio Harvester de la arquitectura implementada. Ambos repositorios permiten ver el funcionamiento de una arquitectura de datos abiertos, que utiliza los vocabularios nativos de CKAN y DCAT para interoperación de los conjuntos de datos que se pueden ofrecer a las aplicaciones web y móviles en formato JSON, CVS y XLS. También permite el almacenamiento de conjuntos de datos en el Repositorio Harvester y los publica para darle acceso a los usuarios y otras herramientas, dejando al Repositorio Instancia como catálogo de datos privado y creando una arquitectura de catálogos abiertos multinivel.

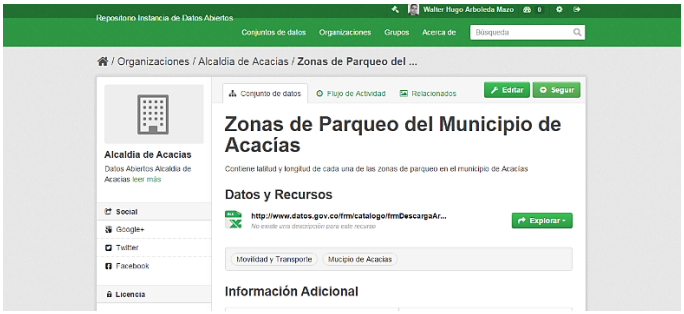


Figura 54. Recurso existente en el conjunto de datos de la instancia

2.4 Aplicación cliente en Android

Es un localizador de Farmacias de Guardia en Navarra (Figura 55), desarrollado en el concurso Open Data Navarra, que puede ser usado para fines académicos en actividades de datos abiertos bajo licencia GNU GPL v3. Los datos se toman del catálogo de datos abiertos del gobierno de Navarra.

La lectura del recurso Guardias.xml se realiza usando las instrucciones de la Figura 56 y la clase FarmaciaGuardiaNavarraData.java

```
try {

    //Uso de API de CKAN para obtener el recurso guardias.xml en el harvester http://192.168.10.81:8080
    String url = "http://192.168.10.81:8080/dataset/9c8d1690-3b3f-43f3-b264-1c9d193b283c/resource/edd6ffeb-37eb-442c-9efa-34
    Log.d(TAG, url);
    InputStream is = openHttpConnection(url);
    DocumentBuilderFactory docBuilderFactory = DocumentBuilderFactory
    .newInstance();
    DocumentBuilder docBuilder = docBuilderFactory.newDocumentBuilder();
    Document doc = docBuilder.parse(is);
    Log.d(TAG, doc.toString());
    processResponseString(doc);

} catch (Exception e) {
    Log.e("JSON", "Error parseando Archivo", e);
    error = true;
```

Figura 56. Lectura del recurso Guardias.xml en la aplicación móvil

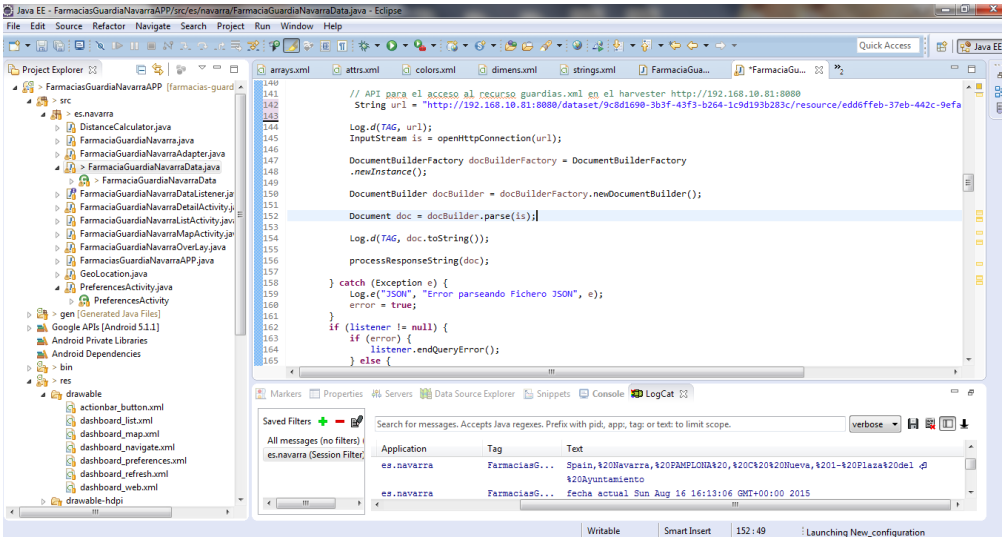


Figura 55. Código fuente en Eclipse de la aplicación FarmaciasGuardiaNavarraAPP

Cuando la aplicación .apk se instala y ejecuta desde el dispositivo móvil, se lee el recurso Guardias.xml desde el catálogo de datos abiertos y se genera la interfaz de la Figura 57.

Al finalizarse la carga de datos desde el catálogo, se visualiza el listado de farmacias que se encuentra en servicio, con la información necesaria para que el usuario seleccione (Figura 58).

Figura 57. Interfaz de la aplicación .apk

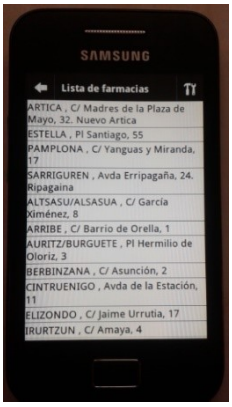


Figura 58. Lista de farmacias de guardia e información complementaria

DEFINICIÓN DE ARQUITECTURA ANALÍTICA Y TECNOLÓGICA PARA DATOS ABIERTOS GUBERNAMENTALES USANDO CKAN

A continuación, se presenta una descripción de la importancia del uso del análisis semántico y la minería de datos abiertos gubernamentales con SpagoBI y Virtuoso Open Source VOS; además, se describan casos en los cuales se resalta la explotación de los datos abiertos para la generación de nuevo conocimiento, a partir de los conjuntos de datos. También se aborda algunos conceptos de capítulos anteriores y se establecen los componentes de la arquitectura básica en CKAN, con un catálogo de datos principal (harvester) y un catálogo de datos secundario, que permite el consumo de datos desde un *browser* por defecto y una aplicación móvil en Android.

1. Importancia de la analítica de datos para el planteamiento de arquitectura tecnológica para datos abiertos gubernamentales

Las entidades gubernamentales deben entender que los datos abiertos son procesos que involucra creación, apertura, búsqueda, acceso, análisis y procesamiento de datos, además de adaptación, visualización, combinación, enlazado y retroalimentación (Mouromtsev et al., 2011), que se obtiene en el uso de los datos con base en su valor y utilidad a esas mismas. Estos aspectos están más allá de las barreras potenciales de aspectos tales como gobierno, costos, legales, características y formatos, forma de acceso y habilidades del personal encargado (Brugger et al., 2016), lo que convierte a los sistemas de datos abiertos en un servicio que no puede monopolizar el gobierno, porque es la forma en que se estructura y homogeniza las interacciones con los ciudadanos. De esta manera se ofrece no solamente catálogos de datos abiertos (Dessi et al., 2016), sino también servicios a los ciudadanos que buscan conocimiento para minería de datos y conjuntos datos enlazados (Shadbolt et al., 2012).

Otro tema que se contempla es el futuro de las infraestructuras de datos basadas en modelos de valor (Priyadarshy, 2015), estableciéndose una segunda generación orientada a una alto aprovisionamiento, búsqueda y descargado de datos, y donde los usuarios retroalimentan sobre la calidad y uso de esos datos. De esta manera se logra facilidad de uso, rendimiento, capacidades de microprocesamiento y cargado de datos desde diversos catálogos (Charalabidis, Loukis y Alexopoulos, 2014), para la explotación de múltiples y diversas fuentes de datos, en la que se integra la analítica (Dos Santos et al., 2014). De esa manera se puede replicar esta arquitectura desde el orden municipal, departamental y nacional (Erickson et al., 2013), con lo que se mejora la comunicación de los ciudadanos con las entidades gubernamentales a diferentes niveles (Zhang et al., 2012), logrando que el gobierno cree estrategias para que los ciudadanos y demás interesados hagan uso de los nuevos servicios soportados por los datos abiertos gubernamentales (Friberger y Togelius, 2012) y garantizando que estos datos sean

realmente abiertos, que se pueden buscar y acceder, que se tengan las herramientas adecuadas. Así, será posible mantener los datos actualizados en el tiempo, asegurando privacidad e integrando cada vez más fuentes de datos (Cowan, Alencar y McGarry, 2014).

Para autores como JoshuaTauberer, los datos abiertos pueden hacer del mundo un mejor lugar (Concha y Naser, 2012), pero para lograrlo es importante buscar su madurez, comenzando desde la producción en formatos accesibles, abiertos y estructurados con identificadores globales, garantizando el acceso con APIs hasta generar nuevo conocimiento mediante el análisis de varios conjuntos de datos y la relación entre ellos (Aryan et al., 2014). Esto permitirá construir una arquitectura que mejore el valor, basada en la generación de conocimiento y la usabilidad y mediada por la visualización, el análisis y el procesamiento (Lin et al., 2011). El descubrimiento de conocimiento a partir de diversas fuentes, tales como conjuntos de datos, es una variable importante cuando se aborda la toma de decisiones en cualquier entidad estatal, donde se implementan soluciones de datos abiertos gubernamentales. Esto permite ir más allá de los datos abiertos almacenados y ofrecidos desde el catálogo de datos mediante interfaces, pero exige implementar una plataforma de descubrimiento de información que involucre no solamente los procesos de publicación, sino también la comprensión del dominio de los datos, la creación de la base de datos de trabajo, la limpieza y pre-procesamiento, que se reduzcan los datos, que se haga una proyección de estos y, finalmente, que se elija un algoritmo de minería para su análisis (Cardona et al., 2014).

En este sentido han aparecido propuestas de arquitecturas orientadas al enlazado, la consulta y la minería de datos abiertos (Jamadagni y Simmhan, 2016). Por ejemplo, la de Koupidis et al., (2016), en la que se define una infraestructura para el análisis del presupuesto de las municipalidades de Atenas y Tesalonika, que lleva datos enlazados de fuentes de datos abiertos a Virtuoso Server mediante almacenamiento, cargado, preparación, análisis y visualización de conjuntos de datos (Liu, Yang y Tung, 2005). Estos autores proponen arquitecturas que consumen datos de catálogo de datos abiertos en CVS mediante un *endpoint*, que luego convierten a RDF para generar conocimiento con SPARQL (Corradi, Foschini y Ianniello, 2014), y que, además, son calificadas con cinco estrellas por Government Linked Data Working Group (GLDWG), que define la importancia de la implementación de una plataforma Linked Open Government Data (Shadbolt y O'Hara, 2013).

Por su parte, Yeh y sus colegas (2016) proponen un sistema de preservación de datos abiertos usando Hadoop, que cuenta arquitectónicamente con los módulos de preservado, computación distribuida, almacenamiento de datos y colección de datos, con lo que se genera cómputo y análisis de datos de alto volumen y variedad a alta velocidad. Esto permite visualizar cómo los datos abiertos gubernamentales son el insumo inicial de una serie de servicios que se pueden implementar para

explotarlos y obtener nuevo conocimiento, lo que genera u alto impacto (Vracic, Varga y Curko, 2016). Por otro lado, Alencar et al., (2015) definen la estructura de capas de una plataforma para facilitar el análisis web y móvil del impacto ambiental. Esta propuesta cuenta con las capas de datos abiertos y otras fuentes de datos, la de ingeniería y ciencia (algoritmos de análisis) y la de interface humana (acceso web y móvil). Neves, Souza y Silva (2015) consumieron datos abiertos relacionados con la ubicación de las comunidades, precipitaciones y el nivel de los ríos en formato DAT, CVS, TXT y XML, que convirtieron a RDF usando la plataforma D2RQ y analizaron con SPARQL, para luego visualizarlos con una implementación de JavaScript. En esta labor fue importante el abordaje de la visualización avanzada de datos abiertos gubernamentales para la generación de mejoras y debates con fundamentos (Kukimoto, 2014). La creación de soluciones con analítica de datos permite mejorar el nivel de acción mediante la creación de planes que eviten problemáticas en la ciudades o regiones (Lai et al., 2012) y haciendo uso del análisis semántico (Spoiala, Rinciog y Posea, 2015).

1.1 Características técnicas del prototipo para el análisis de datos

A continuación, se describe las características técnicas de los servicios implementados para obtener características de análisis de datos usando CKAN, SpagoBI Server Al-In-One y Virtuoso Open Source VOS, para crear una arquitectura compuesta por dos catálogos en CKAN, una plataforma de minería de datos (SpagoBI Server Al-In-One) y una plataforma de análisis de datos enlazados (Virtuoso Open Source, VOS).

1.2 Software para la analítica de datos con SpagoBI Server Al-In-One

Para realizar el análisis de los datos se utilizó SpagoBI, una plataforma de código libre usada para la implementación de análisis de datos, que permite funcionalidades de minería de datos y análisis funcional de los conjuntos de datos abiertos y que se integra desde los catálogos de datos abiertos, facilitando el análisis y la gestión de datos, previamente extraídos del catálogo, así como la transformación y la carga. En la tabla 5 se describe el software utilizado.

Tabla 5. Software utilizado para implementar minería de datos con SpagoBI

Software	Versión	Uso en la arquitectura
SpagoBI Server All-In-One	5.1	Plataforma para el análisis de datos abiertos
Linux Ubuntu Server	Description 12.04.5 LTS Release 12.04 Codename precise 64 bits	Plataforma utilizada para instalar los servidores de CKAN para el harvester y la instancia
Manejador de base de datos PostgreSQL	Postgres-dbscripts-5.2.0	Usado como manejador de base de datos

Mediante la interfaz de gestión de SpagoBI Server (Figura 50) se seleccionan los conjuntos de datos de los catálogos de datos abiertos, para realizar su análisis y visualización.

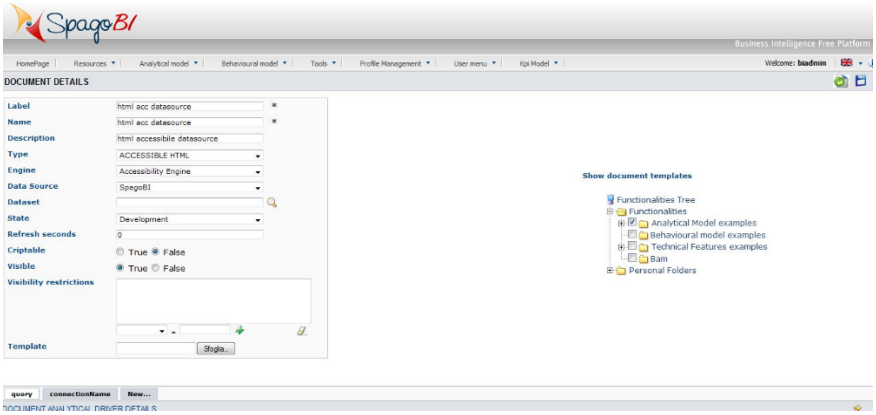


Figura 50. Interfaz de gestión de SpagoBI

1.3 Software para análisis semántico son Virtuoso Open Source VOS

Para la realización del análisis semántico se instaló el servidor semántico Virtuoso Open Source, que permitió el análisis desde los dos catálogos de datos abiertos en CKAN, integrados, mediante la extensión ckanext-semantic, en un archivo de configuración de CKAN. En la Tabla 6 se describe el software usado para implementar el análisis semántico con VOS.

Tabla 6. Software usado para implementar análisis semántico con VOS

Software	Versión	Uso en la arquitectura
Virtuoso Open Source Server	Virtuoso-opensource-6.1.3	Plataforma para el análisis de datos abiertos usando LOD
Linux Ubuntu Server	Description 12.04.5 LTS Release 12.04 Codename precise 64 bits	Plataforma utilizada para instalar los catálogos de CKAN
Manejador de base de datos PostgreSQL	postgres-dbscripts-5.2.0	Usado como manejador de base de datos

La interfaz de Virtuoso Open Source VOS (Figura 51) permite realizar la gestión para el consumo, tratamiento y análisis de los conjuntos de datos usando SPARQL.

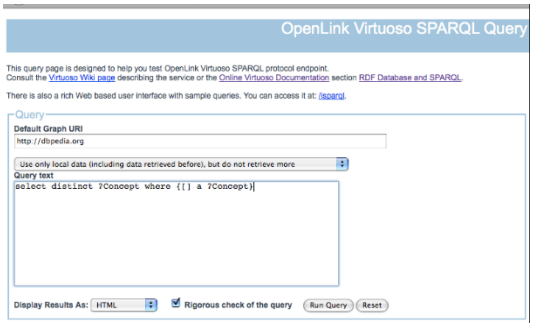


Figura 51. Interfaz de consulta de Virtuoso Open Source

Este grupo de plataformas fue integrado en la arquitectura de la Figura 52, la cual mejora la calidad e innova en los servicios prestados por los catálogos de datos abiertos en Colombia a entidades de orden regional y nacional.

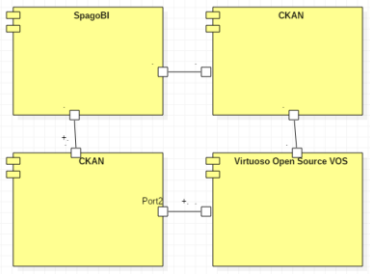


Figura 52. Vista de procesos de la arquitectura de alto nivel para análisis y minería

La vista de procesos permite ver cómo se integraron los servicios SpagoBI con dos catálogos de CKAN para minería de datos abiertos y Virtuoso Open Source VOS, con los catálogos de CKAN para generar un análisis semántico de datos abiertos. Estas funcionalidades enriquecen más los servicios prestados por infraestructuras de datos abiertos mediante la generación de nuevo conocimiento.

1.4 Características técnicas de la arquitectura básica de datos abiertos gubernamentales en CKAN

Debe ser una arquitectura para almacenar, recolectar y publicar datos abiertos, usando un vocabulario que permita la interoperabilidad entre catálogos de datos y el consumo de recursos por aplicaciones. En este caso, una aplicación móvil que maximice el valor de la información por medio de la entrega a los ciudadanos, además de definir una instancia privada para la publicación interna de datos abiertos, que puede ser utilizada por recolectores que permitan la exposición y entrega de datos como catalogo público y que permita el consumo de recursos. Otras características son: que permita entender el funcionamiento de una arquitectura multicapa como publicdata.eu, que recolecta data.gov.ro, data.kk.dk, govdata.de, etc. y que a la vez recolectan conjuntos de datos de catálogos de entidades de cada uno de esos países.

Finalmente, esos datos se exponen en el catálogo público de publicdata.eu, como agregador de los demás catálogos que permite manejar una federación que recolecte desde varias instancias; para esto utiliza el estándar de vocabulario para catálogos de datos abiertos DCAT y el formato nativo de CKAN. A su vez, la definición de los formatos datos abiertos, que se usen en la arquitectura de datos abiertos, debe tener en cuenta la naturaleza de los análisis de los datos, de forma que sea fácil leer, generar cálculos, gráficos y analizar la información desde diferentes equipos de cómputo y dispositivos móviles y sin costo alguno, prefiriéndose formatos que almacenen los datos mediante filas y columnas de forma estructurada y en formato no propietario, tales como JSON, CSV y XML.

CASO DE ESTUDIO PARA EL PLANTEAMIENTO DE ARQUITECTURA TECNOLÓGICA PARA DATOS ABIERTO GUBERNAMENTALES

En este capítulo se describe el aporte realizado para el Catálogo de datos Abiertos en Colombia del presente caso de estudio, se describe la tecnología utilizada y las ventajas que aporta a los datos abiertos en el país. Para ello se toma como región ejemplo de aplicación a la conformada por los municipios del Valle de Aburra en el departamento de Antioquia (Figura 66).

Figura 66. Municipios del Valle de Aburrá

Este Valle cuenta con 3.213.000 habitantes y 10 municipios (Tabla 7) y es la zona con mayor crecimiento económico y social del país, además, como todos los Municipios del país y de conformidad con la Ley 1712 del 6 de marzo de 2014 de Transparencia y acceso a la información pública, deben entregar a los ciudadanos datos de su gestión.

Tabla 7. Municipios del Valle de Aburra

Municipio	Habitantes
Barbosa	50.832
Bello	464.560
Caldas	78.762
Copacabana	71.033
Envigado	227.599
Girardota	55.477
Itagüí	270.920
La Estrella	63.332
Medellín	2.457.680
Sabaneta	52.559

En este caso, el Área Metropolitana del Valle de Aburra como ente público administrativo realiza actividades de gestión en integración económica, proyección y planeación de desarrollo, además de un Sistema de Alerta Temprana de Medellín y el Valle de Aburra SIATA como proyecto de ciencia y tecnología, que sirve para monitorear precipitaciones extremas, inundaciones, movimientos de tierra, episodios críticos de calidad del aire, entre otras variables.

A nivel municipal se puede crear en cada uno un catálogo de datos abiertos gubernamentales que ponga a disposición de los ciudadanos datos de las secretarias, con lo que se brinda transparencia y acceso a la información pública, además de estadísticas de peticiones, quejas, reclamos, denuncias y contrataciones, entre otros.

Para este caso de estudio se tomaron los conjuntos de datos de puntos vive digital de la ciudad de Medellín y se realizó un análisis a la cantidad del acceso gratuito a internet en relación con su ubicación geográfica y población en la ciudad de Medellín. Se encontró que en los lugares de mayor afluencia de personas existen mayor cantidad de puntos. Las plataformas usadas para la realización del caso de estudio se detallan en la Tabla 8. Igualmente, se utilizaron los datos abiertos de Guardias (disponibilidad de droguerías abiertas en Navarra), ingresados al catálogo de datos abiertos y consumidos desde una aplicación móvil en Android Guardias Navarra, elaborada en Open Data Navarra y facilitada por la Universidad de Navarra para este proyecto.

Tabla 8. Plataformas usadas en el caso de estudio

Plataforma	Funcionalidad
CKAN	Plataforma para la implementación del catálogo de datos abiertos gubernamentales en el caso de estudio, en la que se subieron los datos abiertos de puntos de internet gratuito de la ciudad de Medellín
SpagoBI Server All-In-One	Plataforma para minería de datos usando los conjuntos de datos cargados desde el catálogo de datos abiertos gubernamentales
Virtuoso Open Source VOS	Plataforma para análisis de datos enlazados, tomados del catálogo de datos abiertos y convertidos a RDF para analizarlos con SPARQL

En la actualidad se trabaja en la implementación sobre Microsoft Azure de la infraestructura conformada por CKAN, SpagoBI All-In-One y Virtuoso Open Source VOS (Figura 67), para realizar transferencia de conocimiento adquirido en este proyecto.

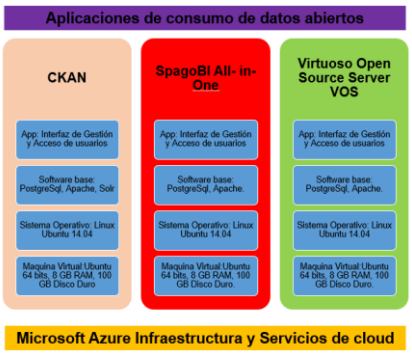


Figura 67. Esquema de la infraestructura tecnológica en Microsoft Azure

CAPÍTULO VII

CONCLUSIONES

En este texto se describe y analiza la importancia que tienen los datos abiertos gubernamentales para la transparencia y como herramienta para disminuir la corrupción, asimismo se muestra el crecimiento del ambiente en general de los datos abiertos, cómo se implementa una infraestructura de este tipo y cómo consumir los datos contenidos en ella.

Desde la iniciativa Perspectiva Tecnológica de la Estrategia de Datos Abiertos para Colombia, desarrollada por el Ministerio de Tecnologías de la Información y la Comunicación, este trabajo se convierte en apoyo para la implementación de arquitecturas de datos abiertos gubernamentales, aprovechando herramientas *open source* como CKAN. Este software fue desarrollado por OKFN y actualmente se ha convertido casi en un estándar para crear catálogos de datos abiertos en entidades municipales, departamentales y nacionales, También permite centralizar los datos en un harvester para ofrecerlos a los interesados, tal y como lo hace el Catalogo de Datos Abiertos del Estado Colombiano.

En cuanto a trabajos futuros se puede profundizar en el área de los datos gubernamentales abiertos a través del desarrollo de e-Services, aplicaciones, *mashups* y soluciones Open Cities, Smart Cities y datos para el trabajo de periodistas de datos.

Desde el entorno comercial y de negocios, es importante que el sector productivo de bienes y servicios visualice cómo aprovechar los datos abiertos gubernamentales como fuente para realizar inteligencia de negocios, además de explotarlos para crear nuevos productos y servicios que respondan a las necesidades ciudadanas, del comercio y del gobierno mismo, de tal manera que se garantice el éxito al utilizar datos reales.

Se debe apoyar más este tipo de proyectos y generar nuevas iniciativas a nivel nacional e internacional, tales como The International Open Data Hackathon, que reúne ciudadanos en varias partes del mundo para crear aplicaciones, liberar datos abiertos gubernamentales, creación, visualización y la adopción los lineamientos para implementar rápidamente arquitecturas a nivel municipal, departamental y nacional.

Si bien el apoyo de entidades líderes en datos abiertos gubernamentales, tales como la Organización de las Naciones Unidas y el Banco Mundial, es importante para el crecimiento de estas iniciativas en todo el mundo, pero también se necesita que se involucren más organizaciones internacionales públicas y privadas, que promuevan procesos regionales de datos abiertos con intereses internacionales.

REFERENCIAS

Adobe Developers Association (1992). TIFF Revision 6.0, 121.

Adobe Systems Incorporated. (2014). PDF files, Adobe Portable Document Format | Adobe Acrobat XI. Online [Jun 2015].

Alencar, P. et al. (2015). An Open and Big Data Platform for Cumulative Environmental Analysis and Management. Proceedings 3rd International Conference on Future Internet of Things and Cloud (pp. 412-417). Rome, Italy.

Alexopoulos, C. (2013). Open Data Movement in Greece: A Case Study on Open Government Data Sources. Proceedings of the 17th Panhellenic Conference on Informatics (pp. 279-286). Thessaloniki, Greece.

Aryan, P. et al. (2014). Fostering government transparency and public participation through linked open government data: Case study: Indonesian public information service. Proceedings International Conference on Data and Software Engineering. Bandung, Indonesia.

Avital, M. & Bjorn, N. (2012). The Value of Open Government Data: A Strategic Analysis Framework. Proceedings of the 18th Annual International Conference on Digital Government Research (pp. 2-9). Staten Island, USA.

Böhm, C. et al. (2012). GovWILD: Integrating Open Government Data for Transparency. Proceedings of the 21st World Wide Web Conference (321-324). Lyon, France.

Breitman, K. et al. (2012). Open Government Data in Brazil. IEEE Intelligent Systems 27(3), pp. 45-49.

Brugger, J. et al. (2016). Current barriers to open government data use and visualization by political intermediaries. Proceedings 6th International Conference for E-Democracy and Open Government (pp. 219-229). Krems, Austria.

Cardona, J. et al. (2014). Knowledge discovery process in the open government Colombian model. Proceedings 9th Computing Colombian Conference (pp. 96-102). Pereira, Colombia.

Chan, C. (2013). From Open Data to Open Innovation Strategies: Creating E-Services Using Open Government Data. Proceedings 46th Hawaii International Conference on System Sciences (pp. 1890-1899). Wailea, USA.

Charalabidis, Y., Loukis, E. & Alexopoulos, C. (2014). Evaluating Second Generation Open Government Data Infrastructures Using Value Models. proceedings 47th Hawaii International Conference on System Sciences (pp. 2114-2126). Waikoloa, USA.

CINTEL (2011). Entregable No 2: Lineamientos para Implementación de Datos Abiertos en Colombia. Bogotá: MINTIC.

Clarke, R. (2013). Smart Cities and the Internet of Everything: The Foundation for Delivering Next-Generation Citizen Services. IDC Government Insights. White Paper #GI243955.

Cohen, B. (2012). What Exactly Is a Smart City? Online [Feb 2015].

Concha, G. & Naser, A. (2012). Datos abiertos: Un nuevo desafío para los gobiernos de la región. CEPAL - Serie Gestión pública 74. Santiago, Chile.

Corona, A. (2013). Social Meets Civic. Master's Thesis. Faculty of Humanities, Tilburg University. The Netherlands.

Corradi, A., Foschini, L. & Ianniello, R. (2014). Linked data for Open Government: The case of Bologna. Proceedings IEEE Symposium on Computers and Communications (pp. 1-7). Funchal, Portugal.

Costabello, L. (2013). Context Aware Access Control and Presentation of Linked Data. PhD Thesis. Universite de Nice-Sophia Antipolis.

Cowan, D., Alencar, P. & McGarry, F. (2014). Perspectives on open data: Issues and opportunities.

Proceedings IEEE International Conference on Software Science, Technology and Engineering (pp. 24-33). Ramat Gan, Israel.

Currie, L. (2013). The Role of Canadian Municipal Open Data Initiatives: A Multi-City Evaluation. Master's Thesis. Queen's University.

Data.gov.gr. (2014). Online [Feb 2015].

Dessi, N. et al. (2016). Increasing open government data transparency with spatial dimension. Proceedings 25th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises, (pp. 247-249. Paris, France.

Ding, L. (2012). Linked Open Government Data. Online [May 2015].

Dos Santos, K. et al. (2014). Experiences integrating heterogeneous Government open data sources to deliver services and promote transparency in Brazil. Proceedings International Computer Software and Applications Conference (pp. 606-607). Vasteras, Sweden.

ECMA (2012). Standard ECMA-376. Online [Jun 2015].

Eibl, G. et al. (2013). Framework for Open Government Data platforms. Open Government Data. White paper.

Erickson, J. et al. (2013). Open government data: A data analytics approach. IEEE Intelligent Systems 28(5), pp. 19-23.

Friberger, M. & Togelius, J. (2012). Generating interesting Monopoly boards from open data. Proceedings IEEE Conference on Computational Intelligence and Games (pp. 288-295).

Høgenhaven, T. (2013). Open Government Communities Does Design Affect Participation? PhD Thesis. Copenhagen Business School.

IEC (2014). Standards Development Word file formats. Online [Jun 2015].

Jamadagni, N. & Simmhan, Y. (2016). GoDB: From Batch Processing to Distributed Querying over Property Graphs. Proceedings 16th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, CCGrid (pp. 281-290). Cartagena, Colombia.

JPEG (2007). JPEG committee. Online [Jun 2015].

JSON (1999). Online [Jun 2015].

Koupidis, K. et al. (2016). Fiscal Knowledge Discovery in Municipalities of Athens and Thessaloniki via Linked Open Data. Proceedings 11th International Workshop on Semantic and Social Media Adaptation and Personalization (pp. 1-6). Thessaloniki, Greece.

Kukimoto, N. (2014). Open government data visualization system to facilitate evidence-based debate using a large-scale interactive display. Proceedings International Conference on Advanced Information Networking and Applications (pp. 955-960). Victoria, Canada.

Lai, Y. et al. (2012). Virtual disaster management information repository and applications based on linked open data. Proceedings Fifth IEEE International Conference on Service-Oriented Computing and Applications (pp. 1-5). Taipei, Taiwan.

Larsson, R. (2013). Uganda Open Development Partnership Platform How can the platform be improved? Independent thesis Basic leve. Karlstad University.

Lathrop, R. & Laurel, D. (2010). Open government: [Collaboration, transparency, and participation in practice]. Sebastopol: O'Reilly.

Lin, J. et al. (2011). GeoKSGrid: A geographical knowledge grid with functions of spatial data mining and spatial decision. Proceedings IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, (pp. 121-126). Fuzhou, China.

Liu, L., Yang, J. & Tung, A. (2005). Data mining techniques for microarray datasets. Proceedings International Conference on Data Engineering. Tokoyo, Japan.

Machado, A. & Parente, J. (2011). DIGO: An Open Data Architecture for e-Government. Proceedings 15th International Enterprise Distributed Object Computing Conference Workshops (pp. 448-456). Helsinki, Finland.

Microsoft (2007). Microsoft Office Excel 97-2007 Binary File Format Specification. Microsoft Corporation.

Microsoft (2012). Office Open XML - What is OOXML? Online [Aug 2015].

Microsoft (2014). Import or export text (.txt or .csv) files. Online [Jun 2015].

MINTIC (2011). Anexo 11: Manual Usuario del Prototipo Tecnológico. Bogotá: Ministerio de Tecnologías de la Información y las Comunicaciones.

MINTIC (2012). Guía para la Apertura de Datos en Colombia. Bogotá: Ministerio de Tecnologías de la Información y las Comunicaciones.

MINTIC (2013). Instructivo para Uso de Servicio de Consulta de Conjuntos de Datos. Bogotá: Ministerio de Tecnologías de la Información y las Comunicaciones.

MINTIC (2014a). Datos Abiertos en Colombia. Online [Jun 2015].

MINTIC (2014b). Lineamientos de Datos Abiertos en Colombia. Bogotá: Ministerio de Tecnologías de la Información y las Comunicaciones.

MINTIC (2014c). Lineamientos para la Implementación de Datos Abiertos en Colombia 2010-2014. Bogotá: Ministerio de Tecnologías de la Información y las Comunicaciones.

Mouromtsev, D. et al. (2011). Development of the St. Petersburg's Linked Open Data Site Using Information Workbench. Proceedings 14th Conference of Open Innovations Association. Espoo, Finland.

Mustafa, M. (2012). Corporate Use of Open Government Data. Master Thesis. MODUL University Vienna.

Neves, C., Sousa, G. & Silva, F. (2015). A Linked Open Data Approach for Visualizing Flood Information - A Case Study of the Rio Doce Basin in Brazil. Proceedings 1st International Conference on Geographical Information Systems Theory, Applications and Management (pp. 227-232). Barcelona, Spain.

OASIS (2014). About Open Document Format. Online [Jun 2015].

Peter, K. (2014). What is Open. Knowledge Foundation. Online [Jun 2015].

Priyadarshy, S. (2015). Big data, smart data, dark data and open data: egovernment of the future. Proceedings Second International Conference on eDemocracy & eGovernment (pp. 16-16). Quito, Ecuador.

PSI (2013). European Public-Sector Information Platform. Online [Feb 2015].

Radl, W. et al. (2013). And Data for All: On the Validity and Usefulness of Open Government Data. Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies (pp. 6-9). Graz, Austria.

Ren, G. & Glissmann, S. (2012). Identifying Information Assets for Open Data: The Role of Business Architecture and Information Quality. Proceedings 14th International Conference on Commerce and Enterprise Computing (pp. 94-100). Hangzhou, China.

Shadbolt, N. & O'Hara, K. (2013). Linked data in government. IEEE Internet Computing 17(4), pp. 72-77.

Shadbolt, N. et al. (2012). Linked open government data: Lessons from data.gov.uk. IEEE Intelligent Systems 27(3), pp. 16-24.

Shahzad, H. (2007). e-Government Services in Pakistan. Master Thesis. Lulea University of Technology.

Spoiala, M., Rinciog, O. & Posea, V. (2015). The Semantic Representation of Open Data Regarding the Romanian Companies. Proceedings 19th International Conference on Information Visualisation (pp. 1-5). Bucharest, Romania.

UPF (2014). Open data. Universitat Pompeu Fabra Retrieved. Online [Feb 2015].

van Grieken, J. (2011). Open Data: A design for the provisioning of Dutch government public and geo-spatial transport data. Bachelor Thesis. University of Groningen.

Vracic, T., Varga, M. & Curko, K. (2016). Effects and evaluation of open government data initiative in Croatia. Proceedings 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (pp. 1521-1526). Opatija, Croatia.

W3C (2009). Publishing Open Government Data. Online [Sep 2015].

W3C (2014a). Data Catalog Vocabulary (DCAT). Online [Sep 2015].

W3C (2014b). RDF Current Status. Online [Sep 2015].

Walker, R. (2012). Open Government Data in Chile. Master Thesis. Pontificia Universidad Católica de Chile.

XML (2013). Extensible Markup Language. Online [Jun 2015].

Yeh, C. et al. (2016). A Scalable Privacy Preserving System for Open Data. Proceedings International Computer Symposium a Scalable Privacy Preserving System for Open Data (pp. 312-317). Chiayi, Taiwan.

Yu, H. (2012). Designing Software to Shape Open Government Policy. PhD Dissertation. Princeton University.

Zetlin, M. (2010). Doc or Docx? Which Office Format to Use. Online [Jun 2015].

Zhang, C. et al. (2012). An endmember extraction framework based on abundance constraint. Proceedings 4th Workshop on Hyperspectral Image and Signal Processing. Shanghai, China.

PLANTEAMIENTO DE ARQUITECTURA TECNOLÓGICA PARA DATOS ABIERTOS GUBERNAMENTALES USANDO CKAN

Walter Hugo Arboleda M.

Magister en Ingeniería de la Universidad EAFIT, Especialista en Redes Corporativas e Integración de Tecnologías de la Universidad de San Buenaventura, Ingeniero de Sistemas de la Universidad de San Buenaventura, investigador en Ingeniería del Software, Big Data, Open Government Data, Open Health Data, Supercomputo, Bibliotecas Digitales, Telemedicina y Redes y Telecomunicaciones.

Integrante de los grupos de investigación SISCO y GORAS de la Universidad Católica Luis Amigo y grupo de investigación en Ingeniería Aplicada gi2a de la Corporación Universitaria Adventista.

Edwin Nelson Montoya M.

Doctor Ingeniero en Telecomunicación de la Universidad Politécnica de Valencia, Ingeniero de Sistemas de la Universidad EAFIT, investigador en Big Data y Analytics, Redes y Servicios Avanzados, Bibliotecas Digitales, Sistemas Distribuidos y Sistemas Multimedia Distribuidos.

Coordinador de la línea de investigación en redes y sistemas distribuidos del grupo de investigación GIDITIC y Jefe del Programa de Ingeniería de Sistemas de la Universidad EAFIT.

Editorial IAI
Instituto Antioqueño de Investigación