

El nuevo reto:
los datos abiertos en Colombia

El nuevo reto: los datos abiertos en Colombia

Giovanny Mauricio Tarazona Bermúdez
Luz Andrea Rodríguez Rojas





UD
Editorial

E2
ESPACIOS

© Universidad Distrital Francisco José de Caldas
© Centro de Investigaciones y Desarrollo Científico
© Giovanni Mauricio Tarazona Bermúdez
© Luz Andrea Rodríguez Rojas

Primera edición, octubre de 2020
ISBN: 978-958-787-236-1

Dirección Sección de Publicaciones
Rubén Eliécer Carvajalino C.

Coordinación editorial

Edwin Pardo Salazar

Corrección de estilo

Lilia Carvajal Ahumada

Diagramación y montaje de carátula

Astrid Prieto Castillo

Imagen de portada

Daniel Andrés Forero (Freelimages)

Propuesta de portada

Andrés M. Enciso

Editorial UD

Universidad Distrital Francisco José de Caldas

Carrera 24 No. 34-37

Teléfono: 3239300 ext. 6202

Correo electrónico: publicaciones@udistrital.edu.co

Catalogación en la publicación – Biblioteca Nacional de Colombia

Tarazona Bermúdez, Giovanni Mauricio

El nuevo reto : los datos abiertos en Colombia / Giovanni
Mauricio Tarazona, Luz Andrea Rodríguez Rojas. -- 1a. ed. --
Bogotá : Universidad Distrital Francisco José de Caldas, 2020.
p. -- (Espacios)

Incluye datos de los autores. -- Contiene referencias
bibliográficas.

ISBN 978-958-787-236-1

1. Acceso a la información - Investigaciones 2. Información
pública oficial - Investigaciones I. Rodríguez Rojas, Luz Andrea
II. Título III. Serie

CDD: 025.0635 ed. 23

CO-BoBN- a1057522

Todos los derechos reservados.

Esta obra no puede ser reproducida sin el permiso previo escrito de la
Sección de Publicaciones de la Universidad Distrital.

Hecho en Colombia

Contenido

Introducción	9
Objetivos específicos	10
Aportaciones y beneficios	10
Metodología y desarrollo de la investigación	11
Fase I. Recolección de información	12
Fase II. Estudio de plataformas y datasets	12
Fase III. Líneas futuras de investigación	13
Datos abiertos	15
Definición	15
Tipos de datos abiertos	17
Actores involucrados	17
Formatos de datos abiertos	19
Catálogos de datos abiertos	23
Beneficios de los datos abiertos	24
Barreras de los datos abiertos	27
Inteligencia de negocios y macrodatos	33
Arquitectura típica	33
Cuadrante mágico e inteligencia de negocios de Gartner 2019	35
<i>Microsoft</i>	37
<i>Tableau</i>	37
<i>Qlik</i>	38
<i>ThoughtSpot</i>	39
Beneficios de la inteligencia de negocios	39
Macrodatos (<i>big data</i>)	40
<i>Cadena de valor</i>	40
<i>Generación de datos</i>	41

<i>Adquisición de datos</i>	41
<i>Almacenamiento de datos</i>	41
<i>Análisis de datos</i>	42
Macrodatos y creación de valor	46
Retos del <i>big data</i>	47
Iniciativas de datos abiertos	49
Antecedentes	52
Latinoamérica y el Caribe	53
Conclusiones y trabajo futuro	63
Referencias bibliográficas	65
Autores	79

Introducción

Recientemente se ha iniciado una revolución en cuanto al acceso a la información pública por parte de los ciudadanos. Se le ha denominado *open data* o datos abiertos (Naser y Concha, 2012). Es una nueva industria en torno a la reutilización de la información.

Las entidades públicas en cumplimiento de su misión, almacenan y protegen grandes cantidades de datos propios de la actividad de gobierno. Liberar esos datos gana cada vez más atención debido a su potencial para estimular el crecimiento económico, apoyar el buen gobierno y facilitar la innovación social. Motivados por estos beneficios, varios Gobiernos del mundo han comenzado a incluir temas de datos abiertos en sus estrategias de administración electrónica, y han puesto en práctica iniciativas que han dado lugar a la aparición del llamado movimiento de datos abiertos gubernamentales (OGD) (Jetzek, Avital y Bjorn-Andersen, 2012).

Para la mayoría de los Gobiernos los objetivos están relacionados con cuatro aspectos: (i) el crecimiento económico (innovación, creación de empresas y puestos de trabajo); (ii) la participación ciudadana; (iii) el mejoramiento de la transparencia y la rendición de cuentas, y (iv) el aumento en la eficiencia de las operaciones y servicios (mejor toma de decisiones mediante el acceso a datos de otros organismos) (Stott, 2014).

Añadir valor a los datos originales proporcionados por los Gobiernos puede facilitar la creación de mercados, empresas y servicios personalizados (Comisión Europea, 2011). En los países pioneros en la apertura y utilización de datos abiertos ya hay evidencia de empresas que han explotado su potencial (Vickery, 2011).

Sin embargo, a pesar de los esfuerzos de los Gobiernos para promover el movimiento OGD, todavía hay una débil reutilización de los datos. Al mismo tiempo, se requiere que universidades, empresas y entidades con capacidad de promover la innovación se involucren activamente para aprovechar el potencial previsto (Gomes y Soares, 2014). Stephenson et al. (2017) encontraron cuatro obstáculos principales para el uso de la información: (i) disponibilidad de los datos, (ii) voluntad para su

uso, (iii) accesibilidad, usabilidad y calidad de los datos, y (iv) capacidad de recolección, manejo y uso de estos. También consideran clave mejorar la coordinación nacional e internacional y la colaboración intersectorial para la gestión de datos; producir y utilizar más productos derivados de los datos; lograr un trabajo conjunto entre los Gobiernos, la sociedad civil y los organismos académicos para romper los obstáculos a su acceso, y compartir e integrar la información para la toma de decisiones.

El objetivo general de esta investigación es mostrar la importancia de los datos abiertos en el mundo actual, los beneficios económicos y sociales de su aprovechamiento y las barreras y desafíos en su uso.

Objetivos específicos

En este documento se proponen los siguientes objetivos específicos:

1. Diagnosticar la situación actual de datos abiertos. Se identificaron las iniciativas gubernamentales y los avances en su implementación y uso.
2. Estudiar las plataformas web y los conjuntos de datos abiertos disponibles.
3. Proponer líneas futuras de investigación.

Aportaciones y beneficios

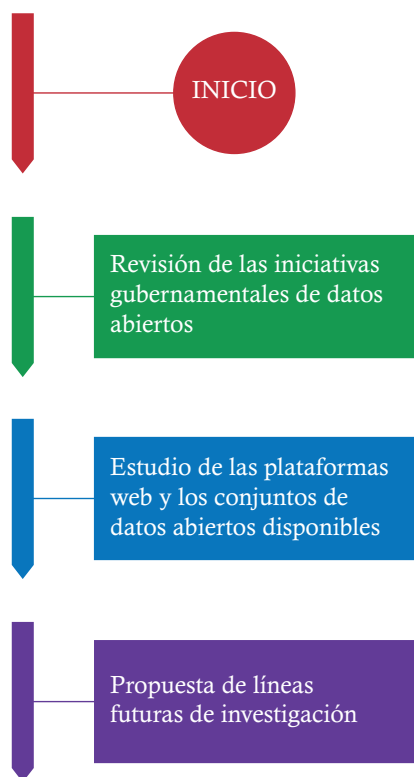
Se presentan las siguientes aportaciones y beneficios originales de esta investigación:

- Revisión de las iniciativas gubernamentales de datos abiertos, avances en su implementación y uso.
- Identificación de los beneficios y las barreras de las iniciativas de datos abiertos.
- Estudio de portales de datos abiertos y conjuntos de datos abiertos.

Metodología y desarrollo de la investigación

El proceso investigativo tuvo tres fases: la primera consistió en recoger información y elaborar un estado de la cuestión; en la segunda se estudiaron los portales de datos abiertos y los datasets disponibles, y la tercera estuvo orientada a proponer líneas de investigación futura. La metodología seguida para la elaboración de esta investigación se describe en la figura 1.

Figura 1. Metodología de la investigación



Fuente: elaboración propia

Fase I. Recolección de información

Se recogió información mediante la búsqueda de artículos de revistas, conferencias, libros, informes gubernamentales y no gubernamentales y de diversas bases de datos, incluyendo Science Direct, Web of Science, ACM Digital Library, IEEE Explore y Google Scholar. Las palabras clave utilizadas para ello fueron combinaciones de los términos “datos abiertos”, “datos públicos abiertos” e “información del sector público”. En la tabla 1 se muestran los resultados obtenidos.

Tabla 1. Búsqueda en bases de datos

Fuente	Encontrado	Relevante	No repetido
Science@direct	80	40	35
Web of Science	34	18	16
IEEE Explore	36	15	15
ACM Digital Library	70	35	33
Total	220	108	99

Fuente: elaboración propia

En total se encontraron 220 documentos. Estos fueron filtrados a partir de las palabras clave y los resúmenes, así se determinaron los relevantes para la investigación. Por último, se revisaron y eliminaron las publicaciones que aparecieron en más de una base de datos, de esta manera se seleccionaron 99 que fueron relevantes, y a partir de ellas se elaboró el estado de la cuestión.

Fase II. Estudio de plataformas y datasets

Se estudiaron los portales de datos abiertos, lo que incluyó los portales de Estados Unidos, Gran Bretaña y la Unión Europea por ser los referentes internacionales. Posteriormente, se delimitó la búsqueda a portales de países latinoamericanos y del Caribe, por mostrar estos un desempeño relativamente positivo comparado con el promedio global. Cinco países (México, Uruguay, Brasil, Colombia y Chile) se encuentran entre los treinta mejores del Barómetro de Datos Abiertos 2016, debido a que lideran y demuestran que las políticas de datos abiertos pueden tener un impacto positivo.

Fase III. Líneas futuras de investigación

A partir de este estudio, de la estructuración de proyectos de investigación y de la socialización de resultados, se proponen líneas y trabajos futuros de investigación con base en los retos identificados.

Datos abiertos

Definición

Desde principios de los años noventa los Gobiernos ofrecen distintos servicios vía web, para que el ciudadano pueda agilizar sus trámites y lograr una administración más eficiente. A esto se le denomina *e-government* o “administración electrónica”. Así, la información de los servicios y datos en poder de la Administración resultan de fácil acceso a los ciudadanos a través de páginas web, dispositivos móviles, etc., lo cual coadyuva a una administración más eficiente y objetiva.

El gobierno electrónico trae consigo múltiples beneficios, especialmente para la ciudadanía, dentro de los que cabe resaltar la facilidad en las comunicaciones, igualdad en el acceso a la información, mayor intercambio colaborativo e incremento en la producción de bienes y servicios de valor agregado. En general, el gobierno electrónico contribuye a la modernización del Estado, al control interno y externo, a la disminución de costos, a reducir la brecha entre Estado y ciudadanía, y a la participación de la ciudadanía en los procesos de toma de decisiones, entre muchas otras cosas (Naser y Concha, 2012).

Aquellos Gobiernos que divulgan la información del sector público reciben el nombre de *Open Government* o Gobiernos abiertos, los cuales dejan disponible información en internet para ser reutilizada por otros (Albano, 2013). Estos practican y promueven la transparencia en su gestión, potencian la colaboración y la participación en la puesta en marcha de políticas públicas, además de compartir recursos informativos y datos que han sido recabados en los procesos en áreas como salud, cartografía, meteorología, educación, datos bibliográficos, contratación pública, legislación, etc. (Manyika et al., 2013).

Algunas de las principales motivaciones son que el acceso a datos públicos proporciona mayores retornos de la inversión, puede generar riqueza a través del uso de los productos hacia abajo, proporciona a los responsables de la formulación de políticas los datos necesarios para abordar problemas complejos, y ayuda a involucrar a la ciudadanía en el análisis de grandes cantidades de conjuntos de datos.

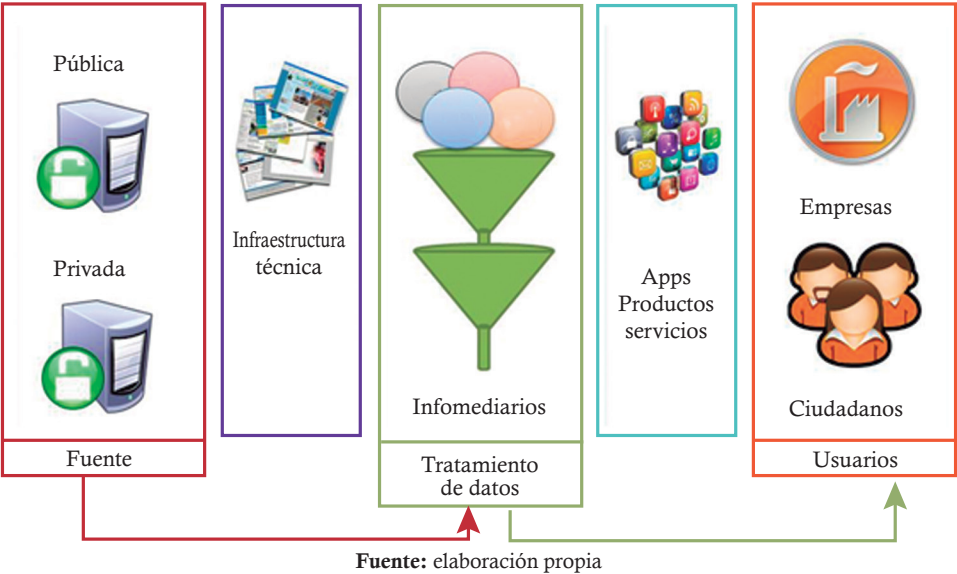
Los datos abiertos pueden ser utilizados o reutilizados por cualquier persona, y se encuentran sujetos al requerimiento de atribución y de compartirse de la misma manera en que aparecen (Open Knowledge Foundation, 2012).

La definición de apertura significa:

- *Disponibilidad y acceso:* la información debe estar disponible en una forma conveniente y modificable como un todo y a un costo razonable de reproducción.
- *Reutilización y redistribución:* los datos serán accesibles en términos que permitan reutilizarlos y redistribuirlos e integrarlos a otros conjuntos de datos.
- *Participación universal:* no debe haber discriminación alguna en cuanto a esfuerzo, personas o grupos.

La cadena de valor alrededor de los datos abiertos inicia con la fuente de datos (figura 2), que pueden ser originados en organizaciones públicas o privadas. Cuando los datos provienen del Gobierno son denominados *open government data*. El siguiente eslabón en la cadena son los mecanismos técnicos que permiten publicar los datos y ponerlos al alcance de todos. De esta manera los infomediarios crean productos y servicios a partir de dichos datos. De allí surgen aplicaciones para dispositivos móviles, portales, iniciativas que promueven transparencia, participación, desarrollo e innovación, entre otras cosas. El último eslabón son los usuarios, bien sea ciudadanos o empresas, que utilizan los productos o servicios ofrecidos por los infomediarios.

Figura 2. Cadena de valor de los datos abiertos



Tipos de datos abiertos

Entre los tipos de datos abiertos (figura 3) se destacan los datos gubernamentales (Yannoukakou y Araka, 2014) o la información del sector público (ISP), que incluye una amplia gama de datos recogidos o financiados por los Gobiernos nacionales, regionales y locales y agencias gubernamentales, recolectados a propósito (por ejemplo, estadísticas nacionales, meteorológicas u otros datos espaciales), o producto de la función gubernamental (registro de empresas, registros judiciales).

Otra área importante son los datos científicos, especialmente los que se derivan de la investigación financiada con fondos públicos. También hay una enorme gama de datos del sector privado que podrían abrirse para beneficio público y privado (por ejemplo, información de seguimiento a vehículos para control del tráfico y el diseño y desarrollo de la infraestructura, o datos de ventas de códigos de barras para la gestión económica, como la estimación del índice de precios al consumidor).

Figura 3. Tipos de datos abiertos



Fuente: Gurin, 2013, p. 35

Actores involucrados

Los Gobiernos, las empresas y los individuos necesitan comprender cómo aprovechar los datos abiertos. Todas las partes interesadas, Gobiernos, organizaciones sin fines de lucro, empresas, individuos (consumidores y ciudadanos), tienen un papel en maximizar los beneficios de los datos abiertos. Obtener información valiosa de

datos abiertos requerirá nuevas reglas y procedimientos y nuevas actitudes, así como inversiones en tecnología y capacidades.

Los Gobiernos pueden definir el éxito de las iniciativas de datos abiertos dentro de una sociedad, tanto mediante la liberación de datos como con la configuración del entorno normativo. Un primer paso importante es establecer prioridades para liberar datos con base en su valor potencial, en lugar de facilitar la “apertura” de los datos para compartirlos. Pueden establecer reglas claras que definan el tipo de datos que deben o no deben ser liberados, con especial énfasis en seguridad, seguridad nacional, privacidad, responsabilidad, derechos de propiedad intelectual y confidencialidad. Como fuentes primarias de datos abiertos, los Gobiernos de todos los niveles pueden ser líderes en el desarrollo de plataformas fáciles de usar para acceder a ellos. Los líderes también pueden buscar asociaciones público-privadas o colaboraciones para apoyar las actividades de datos abiertos.

El Gobierno puede dar forma al ambiente legal y económico que maximiza el valor social potencial del uso de datos abiertos, mientras que trata las preocupaciones legítimas de privacidad y propiedad intelectual de individuos y organizaciones. Esto puede incluir políticas sobre quién puede acceder a la información, y el tipo de información que se puede recopilar o utilizar en determinadas funciones y protocolos para notificar a las personas cuando se dispone de información sobre ellas. Los reguladores también pueden crear políticas para alentar a las empresas a recopilar y liberar datos (con las protecciones adecuadas).

Los datos abiertos brindan oportunidades para que los *empresarios* creen nuevas fuentes de ingresos y aumenten la productividad de sus operaciones. Otros tipos de datos abiertos disponibles por Gobiernos, empresas y terceros (a menudo combinados con otros datos) pueden utilizarse para idear productos y servicios innovadores. Un valor particular está en reunir información de sectores dispares de la economía y combinarla con datos de propiedad. Las empresas que aspiran a aprovechar las oportunidades de datos abiertos pueden buscar e inventariar fuentes de datos, desarrollar estrategias para influenciar a otros con el fin de que los datos valiosos estén más disponibles, invertir en el desarrollo de herramientas orientadas a transformar datos abiertos en productos y servicios, y dedicar personal a desarrollar ideas y diseñar productos utilizando datos abiertos.

Los datos abiertos pueden aportar nuevas ideas sobre cómo operan las *compañías* y ayudar a la Administración a identificar variaciones innecesarias y otras barreras a la productividad. Las empresas pueden optar por compartir datos de propiedad para crear puntos de referencia con la intención de mejorar el rendimiento general de la industria; así como refinar los requisitos del producto y crear nuevos productos y servicios. También tendrán que tener estrategias sobre cómo, cuándo

y en qué circunstancias abrir sus datos, teniendo en cuenta el impacto potencial que ello podría tener. Asimismo, pueden elegir plataformas para la liberación de datos, participar en la creación de un ecosistema de usuarios y considerar formas de monetizar su valor. Además, pueden participar en el establecimiento de normas, incluidos los metadatos, y aportar información sobre los marcos jurídicos emergentes al respecto.

Las empresas también deben evaluar los riesgos potenciales que plantean los datos abiertos, deben ser conscientes de cuáles podrían causar daño a la reputación, perjudicar su posición competitiva o perturbar la industria. Ante estos riesgos, deben participar en debates sobre las normas, el desarrollo de marcos legales y políticas, y la información que debe tener el público general acerca de temas como la propiedad intelectual, privacidad y confidencialidad.

Los *consumidores* pueden aprovechar gran parte del valor potencial de usar datos abiertos; buscar aplicaciones que los usen y proporcionar retroalimentación para mejorar estas herramientas. Si bien los datos abiertos tienen numerosos beneficios potenciales, los consumidores pueden ayudar a salvaguardar sus intereses mediante el monitoreo de políticas y prácticas de privacidad para contribuir a que no se usen con intención de causar daño social o financiero. Los ciudadanos pueden trabajar con el Gobierno para guiar la política de recolección y uso de datos.

Las organizaciones no gubernamentales (ONG) y otras sin fines de lucro pueden incorporar información como ubicación de los recursos —escuelas, hospitales, caminos— y la calidad de los sistemas de salud, educación y economía para identificar las áreas de mayor necesidad. El hecho de presionar a Gobiernos y organizaciones privadas para que divulguen información donde existan vacíos puede conducir a avances relevantes.

ONG y organizaciones sin fines de lucro también pueden organizar voluntarios con habilidades relacionadas con los datos, que pueden colaborar para crear herramientas útiles. OpenStreetMap, una fundación de carácter colaborativo, trazó un mapa en solo dos días para ayudar a las tareas de socorro luego del terremoto de 2010 que azotó Haití, gracias al trabajo de muchos voluntarios en varios continentes. Las organizaciones sin fines de lucro también pueden servir como una fuerza organizadora neutral para alinear las normas internacionales de datos, como la normalización de formatos para metadatos (datos sobre datos) y otros elementos de datos.

Formatos de datos abiertos

Los estándares de datos abiertos son el conjunto de especificaciones, modelos, formatos, protocolos, sintaxis e interfaces de programación usados para la publicación y reutilización automática de los datos.

Por lo general en los sitios web se encuentra información de interés, disponible en diversos formatos y catalogados en función de su naturaleza. Estos datos suelen ser dinámicos, producidos en tiempo real, o en su defecto un equipo de personas los genera periódicamente y los dispone para su descarga (datos estáticos).

En la medida en que mejor estructurados y enriquecidos estén los datos, será más fácil reutilizarlos y construir aplicaciones que los procesen de manera automática. Por esto hay formatos que se consideran de menor valor que otros. En la tabla 2 se presentan algunos de los formatos estándar usados para datos abiertos.

La información de la web puede presentarse en distintos formatos, los cuales permitirán o no reutilizar los datos con mayor facilidad para construir productos y servicios. Entre los más adecuados están XML (eXtensible Markup Language), JSON (JavaScript Object Notation) y el modelo RDF (Resource Description Framework), estos permiten crear servicios de alto valor añadido (Open Knowledge Foundation, 2012).

JSON es un formato ligero para el intercambio de datos basado en JavaScript, su sintaxis es muy simple lo que lo hace fácil de leer por cualquier lenguaje de programación, esto significa que los computadores lo pueden procesar más fácilmente que otros como el XML. Sin embargo, este último es muy usado en el intercambio de datos, ya que brinda grandes oportunidades para mantener la estructura de los datos y la forma como son contruidos, además de permitir a los desarrolladores escribir partes de la documentación sin interferir en su lectura (Open Knowledge Foundation, 2012). Para más información se puede consultar https://www.w3schools.com/js/js_json_intro.asp.

SGML (Standard Generalized Markup Language) o Estándar de Lenguaje de Marcado Generalizado, consiste en un sistema para organizar y etiquetar documentos, normalizado por la Organización Internacional de Normalización (ISO). No impone ningún conjunto de etiquetas en especial, y especifica las reglas de etiquetado de documentos. El lenguaje HTML está definido en términos del SGML (Diez y Millán, 2011). Para más información se puede consultar <https://www.w3.org/TR/WD-html40-970708/intro/sgmltut.html>.

El formato XML fue desarrollado por el World Wide Web Consortium (W3C), un consorcio internacional de reconocido prestigio donde las organizaciones, personal y público en general trabajan conjuntamente para desarrollar estándares web. Este formato XML resulta más sencillo de implementar, pues evita algunas características avanzadas de SGML. Se propone como un estándar para el intercambio de información estructurada entre diferentes plataformas (Open Knowledge Foundation, 2012; Diez y Millán, 2011). Más información en <https://www.w3schools.com/xml/>.

Really Simple Syndication (RSS) es un formato XML para la distribución de contenidos de páginas web. Su objetivo es facilitar la publicación de información actualizada a los usuarios suscritos, utilizando un *software* especializado en este formato. También se encuentran los formatos XLS o Microsoft Office Excel (Microsoft), que muestran la información en celdas organizadas en filas y columnas. Para más información se puede consultar https://www.w3schools.com/xml/xml_rss.asp.

SPARQL (Protocol and RDF Query Language), permite consultar información de diferentes orígenes a través de la web. Se compone de un lenguaje de consulta estandarizado y de un protocolo que realiza consultas a diversas fuentes de datos. Puede obtenerse más información en <https://www.w3.org/TR/rdf-sparql-query/>.

El formato RDF se usa para describir semánticamente recursos, es decir, dotar de sentido lo que representamos para que las máquinas lo comprendan, y hace posible representar datos que provengan de diversas fuentes; fomenta el uso de URL (localizadores uniformes de recursos) como identificadores de tripletas (sujeto-propiedad-objeto), lo cual provee una forma conveniente de interconectar directamente iniciativas de datos abiertos en la web, además, los datos de este tipo pueden ser almacenados en XML y JSON, entre otros (Open Knowledge Foundation, 2012). Para más información se puede consultar <https://www.w3.org/RDF/>.

Otro tipo de documentos en formato abierto son los llamados ficheros CVS o *Comma-separated values*, los cuales representan de manera sencilla datos en formato de tabla, separando las columnas por comas (o punto y coma). El PDF o Portable Document Format es otro tipo de especificación abierta y multiplataforma (se puede presentar en los principales sistemas operativos: Windows, Mac o Unix/Linux). Fue ideado para documentos susceptibles de ser impresos, determinando todos los detalles de cómo va a quedar, sin necesitar procesos anteriores de ajuste ni de maquetación; también se han desarrollado herramientas de *software* libre como OpenOffice.org y el procesador de textos de LaTeX, que permiten crear, visualizar o modificar documentos.

Con referencia a los formatos utilizados para presentar información geográfica, se encuentra SHP Shapefile, un formato propietario estándar de datos espaciales, desarrollado por la compañía ESRI. Los formatos .shp y .shx almacenan el índice espacial de las entidades en las aplicaciones SIG; los .dbf y .prj almacenan el índice espacial de las entidades para los shapefiles que son inalterables (solo lectura) en las aplicaciones SIG; los .prj guardan la información referida al sistema de coordenadas en las aplicaciones SIG, entre otros (Lasanta, 2011).

Para mapas interactivos se utilizan los servicios Web Map Service, el cual produce mapas a partir de información geográfica vectorial, presentando la información como imágenes digitales aptas para ser visualizadas en pantalla en formatos PNG (Portable

Network Graphics), GIF o JPEG. Web Computer Graphics Metafile (WebCGM) es óptimo para diagramas complejos en que se requiere hacer zum, visualizarlos por capas o sobreponerlos unos a otros.

Aunque oficialmente no hay una manera de clasificar los conjuntos de datos en la web, la propuesta formulada por Tim Berners-Lee (Matheus y Vaz, 2012) clasifica los conjuntos de datos puntuándolos con un número de estrellas entre 1 y 5. La tabla 2 muestra algunas de las ventajas y desventajas de estos formatos.

Tabla 2. Clasificación de los conjuntos de datos

Estrellas	Formato	Ventajas	Desventajas
★	PDF, texto, imágenes	No requiere ningún tipo de transformación previa, pueden ser publicados de manera sencilla.	Los formatos son muy difíciles de manipular y reutilizar.
★ ★	Ficheros Excel, XLS		Para acceder a los datos se requiere herramientas cerradas.
★ ★ ★	CSV, XML o RDF	El acceso y la manipulación de los datos es sencilla.	Requiere la transformación previa.
★ ★ ★ ★	URL	Facilita la labor de reutilización de los datos y su actualización.	Mayor esfuerzo de estructuración de los datos, su separación y asignación de las URL.
★ ★ ★ ★ ★	RDF, basado en XML Linked data	Enriquecimiento semántico de los conjuntos de datos y estandarización del vocabulario.	

Fuente: Matheus y Vaz, 2012, p. 24

La clave de los datos abiertos es mezclarlos con otro material, por ello la interoperabilidad es absolutamente fundamental. Al incrementar la posibilidad de combinar distintas bases de datos o conjuntos de datos, se podrán desarrollar más y mejores productos y servicios (Open Knowledge Foundation, 2012).

Catálogos de datos abiertos

Un catálogo de datos es la entrada organizada a los datasets o conjuntos de datos publicados. Está enfocado en facilitar el acceso sin restricciones de uso de los datos. La Unidad de Modernización y Gobierno Digital lo define como “la organización de los metadatos de los datos del repositorio, más metadatos y enlaces a otros datos relacionados relevantes para los usuarios” (2013, p. 10), entendiendo los metadatos como datos que describen otros datos, lo cual permite organizarlos, clasificarlos y relacionarlos.

Estos catálogos de datos se han elaborado en plataformas como Junar, CKAN y DKAN, entre otras, las cuales permiten a los desarrolladores crear sus páginas web donde los usuarios puedan buscar, publicar y reutilizar los datos que deseen, fomentando la creación de nuevas aplicaciones útiles para la sociedad.

La plataforma Junar ha sido utilizada por los Gobiernos de Chile, Sacramento y Palo Alto (Estados Unidos) y Lima (Perú), entre otros; y CKAN, por su parte, ha sido utilizada como plataforma de datos abiertos del Reino Unido (data.gov.uk), de la Unión Europea (<https://data.europa.eu/euodp/es/data/>) y del portal de Helsinki, Finlandia, entre otros.

CKAN es un *software* gratuito y de licencia abierta, que permite mantener y publicar colecciones de datos. Es empleado por Gobiernos, instituciones de investigación y otras organizaciones que coleccionan gran cantidad de datos (Open Knowledge Foundation, 2012). Con esta herramienta es posible buscar y encontrar datos y previusualizarlos usando mapas, gráficos y tablas.

Este *software* fue desarrollado por la Open Knowledge Foundation y ha sido muy útil para países que han desarrollado sus portales de datos abiertos como parte de los compromisos establecidos en la Alianza para el Gobierno Abierto (Open Government Partnership, OGP). Esta alianza hace esfuerzos para promover políticas de participación ciudadana, lucha contra la corrupción, empoderamiento ciudadano y gobierno electrónico para cada uno de sus Estados miembros. La OGP lidera varios proyectos aparte del CKAN como “The Open Definition”, que sienta las bases teóricas para los datos abiertos, o “Data Commons”, y ofrece soluciones legales (Alianza Regional por la Libre Expresión e Información, 2012; Peset, Ferrer-Sapena y Subirats-Coll, 2011). Para más información se puede consultar <https://ckan.org/>.

En cuanto a DKAN, es una plataforma basada en Drupal, un sistema para manejo de datos (Data Management System, DMS), que contiene un conjunto completo de funciones de catalogación de datos, edición y visualización; puede ser personalizada y extendida fácilmente por los usuarios, quienes pueden explorar datos, buscarlos y describirlos. El modelamiento de datos de DKAN es muy parecido al de CKAN; permite describir los datasets, las fuentes, crear grupos y etiquetas y es

totalmente abierta, lo que contribuye a acelerar el desarrollo, reducir costos y eliminar dependencia de proveedores y empresas (Assaf, Troncy y Senart, 2015). Puede encontrar más información en <https://www.drupal.org/>.

Beneficios de los datos abiertos

Los datos gubernamentales y científicos desempeñan un papel importante en las economías basadas en el conocimiento y, con muchos de los beneficios aún por alcanzar, podrían contribuir al crecimiento económico. El principal desafío es que los datos abiertos no tienen valor por sí mismos; solo lo adquieren cuando se utilizan. Uno de los principales beneficios de la apertura de un sistema es la capacidad de aprovechar la inteligencia colectiva del público. La idea clave es que, en las circunstancias adecuadas, los grupos pueden ofrecer alternativas y tomar mejores decisiones que las que podrían tomar las personas más inteligentes por sí solas.

Los beneficios se agruparon en: (1) políticos y sociales, (2) económicos, y (3) operacionales y técnicos (ver figura 4).

Figura 4. Beneficios de los datos abiertos



Fuente: BIS, 2013; Gruen, Houghton y Tooth, 2014; Janssen, Charalabidis y Zuiderwijk, 2012

Desde otro punto de vista, algunos estudios que exploran el valor económico de los datos abiertos se han centrado en otras dimensiones de valor, y han adoptado diferentes enfoques y métodos económicos. La mayoría apunta a la información del sector público, aunque algunos exploran datos privados y de investigación.

Adoptando un enfoque de arriba abajo, la Comisión Europea (2000) combinó medidas del costo de inversión (es decir, el monto gastado en la recolección/generación de la información del sector público, ISP) y el gasto por usuarios y reutilizadores; luego, para los usuarios finales, el valor estimado como gasto o, cuando la ISP estaba disponible libremente, como el costo de inversión de su recolección/generación; y para los intermediarios que agregan valor a los datos, atribuyó una proporción de su valor agregado a la ISP utilizada. De esta manera, se estimó la inversión en unos 9500 millones de euros al año en 1999, y el valor económico de la ISP, en la Unión Europea, en unos 68 000 millones de euros, aproximadamente el 1,4 % del PIB, lo cual representa un rendimiento de la inversión de siete veces.

Por otro lado, utilizando una gran encuesta de productores y usuarios de ISP, Dekkers et al. (2006) intentaron estimar el tamaño del mercado de información del sector público en Europa. En el estudio, denominado “Medición de los recursos de información del sector público europeo” (MEPSIR, por su sigla en inglés), la demanda y el desempeño económico se midieron en una extensa encuesta, que pidió directamente a los titulares de contenido público y a los reutilizadores datos económicos clave, como el volumen de negocios total, el número total de funcionarios frente al número del personal dedicado a la manipulación de la ISP, y las estimaciones del mercado interno para un tipo particular de ISP. Sobre la base de las estimaciones de los reutilizadores, ponen el mercado total de la ISP en la Unión Europea más Noruega en unos 27 000 millones de euros (aproximadamente el 0,25 % del PIB agregado).

DotEcon produjo un informe sobre el desarrollo de un enfoque de base para estimar el valor económico de los productos y servicios de la ISP en el Reino Unido, viendo el valor económico neto de la ISP como la disposición a pagar por ella, menos el costo de su suministro. También examinaron los costos (perjuicio) de los obstáculos al uso, entre ellos precios indebidamente altos, distorsión de la competencia y falta de explotación de la ISP (Dekkers et al., 2006). Los resultados indican que el valor neto de la ISP en el Reino Unido se situaba en torno a 590 millones de libras esterlinas al año en 2005. Los costes de los tres tipos de perjuicios se estimaron en 20 millones de libras esterlinas, 140 millones por restricción de la competencia y 360 millones por no explotar la ISP, lo que sugiere que el valor de esta podría duplicarse resolviendo los problemas identificados.

Al revisar la literatura sobre el tamaño y el crecimiento del mercado europeo de la ISP para la Unión Europea, Vickery (2011) estimó que el mercado relacionado con la ISP directa era del orden de 28 000 millones de euros en 2008. Todos los estudios muestran el crecimiento relativamente rápido de los mercados relacionados con la ISP.

Basándose en el enfoque ascendente desarrollado por DotEcon, el Department for Business Innovation and Skills (BIS) de Reino Unido estimó que el valor de la

información del sector público para los consumidores, las empresas y el sector público era aproximadamente de 1800 millones de libras esterlinas con precios de 2011, pero sugirió que el uso y la reutilización de la ISP tiene impactos mucho más profundos que afectan todas las áreas de la sociedad, más allá del cliente directo. Sobre la base de suposiciones conservadoras, estimaron que podría superar los 5000 millones de libras esterlinas a precios de 2011.

Manyika et al. (2013) sugieren que los datos abiertos tienen el potencial de desbloquear aproximadamente 3,2 billones de dólares en valor adicional anualmente, a través de los siete dominios de la economía que el estudio examinó (educación, transporte, electricidad, petróleo y gas, salud, productos financieros y de consumo). El valor puede surgir de varias maneras, incluyendo equipar a los trabajadores con las habilidades para aumentar la productividad, pues permite a los vendedores identificar poblaciones de microsegmentos con más éxito e impulsar el desempeño al compartir puntos de referencia, datos de mercado e información de mejores prácticas. Los consumidores pueden ahorrar dinero gracias a una mayor transparencia de precios y teniendo más información para tomar decisiones.

McBride et al. (2019) exploraron la relación entre datos abiertos gubernamentales y cocreación. Los autores señalan seis factores que parecen desempeñar un papel clave para permitir que tenga lugar la creación conjunta de servicios públicos impulsados por los datos abiertos: partes interesadas motivadas, líderes innovadores, comunicación adecuada, existencia de un portal de gobierno de datos abiertos, financiación externa y desarrollo ágil.

Stuermer y Dapp (2016) proponen un marco para evaluar el impacto de la liberación de datos abiertos mediante la aplicación del enfoque de retorno social de la inversión. Allí se evidencian posibles impactos sociales como mayor calidad educativa, más igualdad de oportunidades para los niños, más intercambio de conocimientos entre hospitales y médicos, más uso del transporte público, menos corrupción y mejor comprensión financiera de los ciudadanos, entre otras.

Específicamente, para el sector privado Heralda et al. (2016) y Lindman, Kinnari y Rossi (2015) encontraron como beneficios del uso de datos abiertos acciones colaborativas y mayor diálogo y participación de los clientes. Las empresas también pueden aprovechar el conocimiento de los actores privados y las comunidades, y aumentar la transparencia en su entorno de trabajo y en los procesos. Las pequeñas y medianas empresas (pymes), los empresarios y otros negocios no solo pueden encontrar un beneficio económico mediante el uso de datos gubernamentales, sino que gracias a servicios innovadores pueden repercutir indirectamente en la creación de empleo. Aunque solo el acceso a los datos no proporciona ventaja competitiva, las empresas pueden innovar sobre los datos disponibles para crear servicios de valor añadido (Attard, Orlandi y Auer, 2016).

El uso y reutilización de la información del sector público tiene impactos mucho más profundos que afectan a todas las áreas de la sociedad. La información del sector público y privado puede actuar como catalizador en el proceso de innovar, optimizar los costos y ayudar a los funcionarios, líderes empresariales y ciudadanos comunes a tomar mejores decisiones gracias a información de calidad.

Barreras de los datos abiertos

Cada vez más datos gubernamentales están a disposición del público, pero no siempre en un formato abierto dado los frecuentes problemas con las licencias, los formatos, las descargas a granel o la disponibilidad gratuita. Desafortunadamente, muchos Gobiernos con buenas intenciones piensan que están abriendo los datos, pero no cumplen los requisitos clave para su verdadera facilidad de acceso y reutilización (World Wide Web Foundation, 2015).

Si bien es cierto que gran parte de los resultados de los datos abiertos prometen ser benéficos en diversos escenarios, también es cierto que dependen en gran medida de la estructura política de cada Estado, es decir, es previsible que se presenten problemas o barreras frente a la apertura de datos. Esos obstáculos pueden clasificarse en institucionales, de complejidad de las tareras, de uso y participación, legales, de calidad y técnicos. La figura 5 muestra las barreras institucionales identificadas en la literatura.

Figura 5. Barreras ámbito institucional



Fuente: elaboración propia a partir de Naser y Concha, 2012; Janssen, Charalabidis y Zuiderwijk, 2012; Wang, Fang, Liu y Horn, 2018; Fundación COTEC, 2011; Michener, 2010; Zurada y Karwowski, 2011; Junior, Santos, Vidal y De Carvalho, 2012; Blakemore y Craglia, 2006; McLaren y Waters, 2011; Huijboom y Van den Broek, 2011; Zuiderwijk, Janssen y Choenni, 2013

La ISP se deposita en varias fuentes fragmentadas, por lo que los usuarios tienen dificultades para encontrar los datos. Se requiere transformar las organizaciones del sector público para garantizar que no abusen de su poder de monopolio como únicos productores de determinados tipos de datos, y que prioricen las necesidades de los ciudadanos y no las propias. Es necesario identificar un terreno común donde las partes implicadas puedan centrarse en la construcción de información de valor que beneficie a todos.

Figura 6. Barreras ámbito complejidad de las tareas



Fuente: elaboración propia a partir de Naser y Concha, 2012; Janssen, Charalabidis y Zuiderwijk, 2012; Fundación COTEC, 2011; Michener, 2010; Zurada y Karwowski, 2011; Junior, Santos, Vidal y De Carvalho, 2012; Blakemore y Craglia, 2006; McLaren y Waters, 2011; Huijboom y Van den Broek, 2011; Zuiderwijk, Janssen y Choenni, 2013

En cuanto al ámbito de la complejidad de las tareas (figura 6), un tema subestimado parece ser la disponibilidad de todo tipo de capacidades y niveles de conocimiento de los usuarios para el uso de datos complejos y sofisticados. No sirve de nada disponer de conjuntos de datos en portales si los usuarios que podrían beneficiarse de su reutilización no cuentan con las herramientas para sacarles provecho. Adicionalmente, en muchas ocasiones los datos no son fáciles de descubrir, están duplicados y son poco interoperables. Todo esto obliga a los usuarios a limpiar y tratar los conjuntos de datos antes de poder reutilizarlos.

La figura 7 presenta las barreras de uso y participación. Los usuarios de datos encuentran difícil interpretarlos, porque no están familiarizados con las definiciones y categorías que se utilizan, por lo tanto es complicado la comparación, el enlace y su reutilización. Como no todos los usuarios están en capacidad de usar los datos

abiertos, se corre el riesgo de limitarlos a ciertos grupos y, de esta manera, lejos de aprovechar su valor se contribuye a ampliar la brecha digital.

Figura 7. Barreras ámbito uso y participación

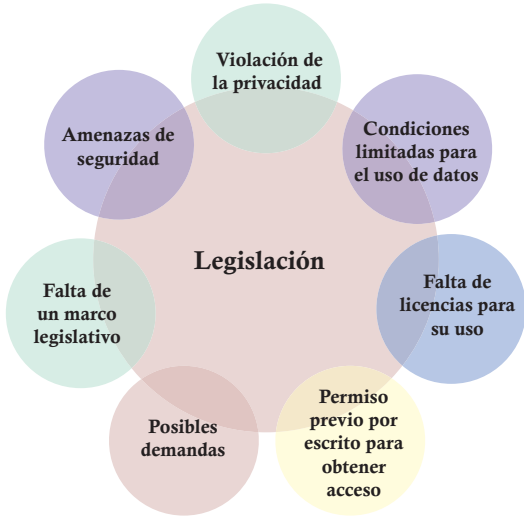


Fuente: elaboración propia a partir de Naser y Concha, 2012; Janssen, Charalabidis y Zuiderwijk, 2012; Fundación COTEC, 2011; Michener, 2010; Zurada y Karwowski, 2011; Junior, Santos, Vidal y De Carvalho, 2012; Blakemore y Craglia, 2006; McLaren y Waters, 2011; Huijboom y Van den Broek, 2011; Zuiderwijk, Janssen y Choenni, 2013

En muchos países ya existen iniciativas de datos abiertos, pero carecen de un marco legal que reglamente su uso. Por ello la confiabilidad de la fuente, la procedencia de los datos y los aspectos legales de su consumo son temas preocupantes que no han sido abordados. Las partes interesadas están expuestas a la violación de la privacidad y a ser legalmente responsables cuando los datos abiertos sean mal utilizados. Otras barreras legales se muestran en la figura 8.

En cuanto a la calidad de los datos (figura 9), se encuentra la carencia de metadatos que permitan conocer su calidad y cómo fueron recopilados. En muchos casos se requiere una carga de trabajo sustancial para limpiarlos y hacerlos comprensibles. Los portales están llenos de conjuntos de datos obsoletos o con información poco relevante. Esto es una gran barrera, pues disponer de demasiada información no necesariamente resulta en decisiones mejores, más democráticas o más racionales. Más información puede resultar en menos comprensión, más confusión y menos confianza (Strathern, 2000).

Figura 8. Barreras ámbito legislación



Fuente: elaboración propia a partir de Naser y Concha, 2012; Janssen, Charalabidis y Zuiderwijk, 2012; Fundación COTEC, 2011; Michener, 2010; Zurada y Karwowski, 2011; Junior, Santos, Vidal y De Carvalho, 2012; Blakemore y Craglia, 2006; McLaren y Waters, 2011; Huijboom y Van den Broek, 2011; Zuiderwijk, Janssen y Choenni, 2013

Figura 9. Barreras ámbito calidad de la información



Fuente: elaboración propia a partir de Naser y Concha, 2012; Janssen, Charalabidis y Zuiderwijk, 2012; Fundación COTEC, 2011; Michener, 2010; Zurada y Karwowski, 2011; Junior, Santos, Vidal y De Carvalho, 2012; Blakemore y Craglia, 2006; McLaren y Waters, 2011; Huijboom y Van den Broek, 2011; Zuiderwijk, Janssen y Choenni, 2013.

Figura 10. Barreras ámbito técnico

Fuente: elaboración propia a partir de Naser y Concha, 2012; Janssen, Charalabidis y Zuiderwijk, 2012; Fundación COTEC, 2011; Michener, 2010; Zurada y Karwowski, 2011; Junior, Santos, Vidal y De Carvalho, 2012; Blakemore y Craglia, 2006; McLaren y Waters, 2011; Huijboom y Van den Broek, 2011; Zuiderwijk, Janssen y Choenni, 2013

Por último, en el ámbito técnico (figura 10) hay barreras como la falta de soporte para los usuarios inexpertos, y falencias en infraestructura tecnológica para el almacenamiento, descarga, disposición y uso de los datos. Ser capaz de utilizarlos y encontrar patrones y tendencias en grandes cantidades de datos también sigue siendo un reto importante (Zurada y Karwowski, 2011), pues esta capacidad es necesaria, al igual que el contenido automatizado y el estudio de temas. Los algoritmos de aprendizaje automático a menudo pueden analizar un conjunto de datos e inferir reglas para clasificar y agrupar elementos de datos y de esta manera sacarles provecho. La automatización y clasificación de datos puede ayudar a simplificar búsquedas, establecer relaciones y extrapolar tendencias (Janssen, Charalabidis Zuiderwijk, 2012).

Los Gobiernos y la sociedad civil deberían trabajar juntos para identificar las prioridades sociales más apremiantes y las necesidades de datos relacionadas con estas. El éxito de los sistemas de datos abiertos implica la mejora de la calidad de la información gubernamental, la creación e institucionalización de una cultura de gobierno abierto, y la provisión de herramientas e instrumentos para utilizar los datos. Es necesaria una infraestructura que ayude a los usuarios a darles sentido a los

datos, así como medidas institucionales para garantizar la participación pública. En estas condiciones los datos abiertos pueden ir más allá del nivel actual de participación ciudadana y podrían dar lugar a un diálogo continuo entre los Gobiernos y sus electores, aprovechando la inteligencia colectiva del público (Janssen, Charalabidis Zuiderwijk, 2012).

El movimiento de datos abiertos está en un punto de inflexión. Los objetivos de desarrollo sostenible de las Naciones Unidas han impulsado una revolución de datos, y los compromisos e iniciativas de políticas de datos abiertos se han extendido rápidamente. Sin embargo, la implementación y el impacto están rezagados, haciendo que crezca el riesgo de que el movimiento de datos abiertos se desvanezca en una ciudad fantasma de portales abandonados y aplicaciones olvidadas (World Wide Web Foundation, 2015).

Además, es posible el rechazo y la resistencia de todos los niveles de la sociedad y en entes gubernamentales, debido a los cambios que implica la puesta en marcha de datos abiertos, la forma de pensar, cambios en los sistemas de información y, no menos importante, al brindar información que antes resultaba inaccesible para el ciudadano del común, Gobiernos menos versátiles pueden tener la falsa percepción de pérdida de poder (Naser y Concha, 2012).

La Comisión Europea (Unión Europea, 2017), para superar las barreras de los datos abiertos, recomienda: mejorar el desarrollo de los portales de datos abiertos, mantener el apoyo a su financiación, documentar su impacto, interactuar con los usuarios y ver su comportamiento, promover los datos abiertos en la transformación digital del sector público, aprovechar el uso de datos privados de interés público para aumentar la eficiencia y el valor de la formulación de las políticas, e impulsar la disponibilidad de datos en tiempo real con la ayuda de herramientas tecnológicas como sensores o con el internet de las cosas.

Las iniciativas de datos abiertos prometen un potencial con múltiples beneficios, por ello los entes públicos y privados han mostrado su interés en promover su disponibilidad al público, sin embargo, no se deben dejar de lado las barreras que limitan su uso y de esta manera garantizar una verdadera reutilización de la información.

Inteligencia de negocios y macrodatos _____

A pesar de que la inteligencia de negocios (*business intelligence*) y los macrodatos (*big data*) son tecnologías para el estudio de datos y la toma de decisiones, difieren en cómo lo hacen y el tipo de datos que analizan. A continuación, se presentan las generalidades de estas dos tecnologías.

La inteligencia de negocios es un “enfoque estratégico para orientar sistemáticamente el seguimiento, la comunicación y la transformación relacionada al débil conocimiento de la información procesable en la cual se basa la toma de decisiones” (Rouibah y Ould-ali, 2002, p. 135). Es una disciplina que integra información de diversos orígenes, facilitando el almacenamiento y tratamiento de datos históricos.

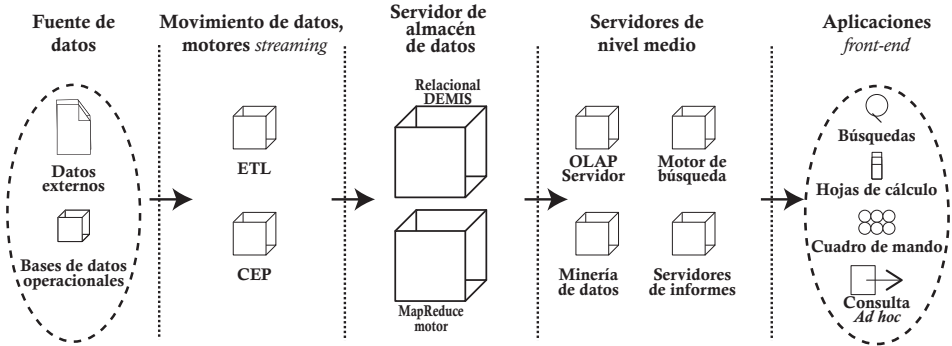
Con el fin de entregar la información útil para la toma de decisiones, la inteligencia de negocios es clave para que los usuarios extraigan eficientemente información valiosa de los océanos de datos. Garter Group fue el primero en introducir el concepto de inteligencia de negocios, y lo define como un término general que incluye aplicaciones, infraestructura, herramientas y mejores prácticas para mejorar y optimizar las decisiones y el rendimiento (Gartner, s. f.).

Arquitectura típica

Una arquitectura típica de inteligencia de negocios se muestra en la figura 11. Los datos a menudo provienen de diversas fuentes, normalmente de múltiples bases de datos operacionales entre departamentos dentro de la organización, así como de proveedores externos. Las fuentes contienen datos de calidad variable, representaciones inconsistentes, códigos y formatos diferentes. Por lo tanto, integrar, limpiar y estandarizar los datos puede ser bastante desafiante.

Estas tecnologías de soporte (*back-end*) para preparar los datos se denominan herramientas *extract-transform-load* (ETL). Como cada vez es más necesario trabajar en tiempo casi real, los motores de procesamiento de eventos complejos (CEP) han surgido para apoyar estos escenarios (Hermosillo, Seinturier y Duchien, 2010).

Figura 11. Arquitectura típica de inteligencia de negocios



Fuente: adaptado de Chaudhuri, Dayal y Narasayya, 2011

Los datos normalmente se cargan en un repositorio denominado almacén de datos (*data warehouse*), administrado por uno o más servidores. Normalmente, para almacenar y consultar datos de almacén se utilizan los sistemas de gestión de base de datos relacionales (RDBMS).

A medida que la cantidad de datos aumenta son necesarias plataformas de bajo costo, capaces de soportar mucho más volumen de datos que el tradicionalmente manejado por RDBMS (Duda, 2012). Por ello los motores basados en el paradigma MapReduce ahora se usan para el análisis de la empresa.

Los servidores de almacén de datos se complementan con un conjunto de servidores de nivel medio. Los servidores de procesamiento analítico en línea (OLAP) exponen eficientemente la vista multidimensional de los datos a aplicaciones o usuarios, y permiten las operaciones como el filtrado, la agregación, la clasificación y el pivoteo (Mansmann, Neumuth y Scholl, 2007). Los servidores de informes permiten la definición, la ejecución eficiente y la prestación de informes.

Los motores de búsqueda utilizan palabras clave sobre el texto y los datos estructurados en el almacén (por ejemplo, buscan mensajes de correo electrónico). Los motores de minería de datos sirven para un análisis en profundidad más allá de lo que ofrece OLAP o servidores de informes, y proporciona la capacidad de construir modelos predictivos para ayudar a responder a preguntas específicas (Vercellis, 2011).

Existen varias aplicaciones de *front-end* a través de las cuales los usuarios realizan tareas de inteligencia de negocios: hojas de cálculo, portales empresariales de búsqueda, aplicaciones de gestión de desempeño que permiten a los responsables de tomar decisiones identificar los indicadores clave del rendimiento de la empresa utilizando paneles visuales, y visores para modelos de minería de datos, entre otros.

Además, existen otras tecnologías como la analítica web (Chaudhuri, Dayal y Narasayya, 2011), que sirve para comprender cómo interactúan los visitantes del

sitio web de una empresa con las páginas, y la gestión de relaciones con el cliente (CRM). CRM puede proporcionar funcionalidad para segmentar a los clientes entre los más probables y los menos propensos a recomprar un producto en particular (Ranjan y Bhatnagar, 2011). Otra área incipiente, pero importante, es la inteligencia de negocios móvil (Airinei y Homocianu, 2010), que presenta oportunidades orientadas a nuevas y ricas aplicaciones para trabajadores del conocimiento en dispositivos móviles.

Cuadrante mágico e inteligencia de negocios de Gartner 2019

Un mercado apretado con nuevos competidores, la evolución rápida y la innovación constante crean un ambiente difícil para que los vendedores de soluciones de inteligencia de negocios diferencien sus ofertas de la competencia. La lista de proveedores se extiende desde los grandes actores de la tecnología, tanto los nuevos como los antiguos, que tratan de reinventarse para recuperar relevancia, a las empresas emergentes respaldadas por enormes cantidades de capital de riesgo privado.

El cuadrante mágico de Gartner (figura 12) ofrece un posicionamiento gráfico competitivo de cuatro tipos de proveedores de tecnología, en mercados donde el crecimiento es alto y la diferenciación de proveedores es distinta:

- *Los líderes* ejecutan bien su visión actual y están bien posicionados para mañana.
- *Los visionarios* entienden adónde va el mercado o tienen una visión para cambiar las reglas de este, pero aún no ejecutan bien.
- *Los jugadores de nicho* se enfocan con éxito en un segmento pequeño.
- *Los desafiantes* ejecutan bien hoy o pueden dominar un segmento grande, pero no demuestran una comprensión de la dirección del mercado.

Para 2019 los líderes tradicionales quedan relegados al cuadrante de “visionarios”. Los nuevos líderes son Tableau, Qlik, ThoughtSpot y Microsoft Power BI (Parenteau et al., 2020).

El informe considera que, en el futuro, las plataformas de Analytica e inteligencia de negocios se orientarán hacia:

- *Analítica aumentada*: cada vez se incluyen más capacidades de aprendizaje automático (*machine learning* —ML—) en las plataformas de inteligencia de negocios modernas.
- *NLP (Natural Language Processing)*: acceso a los datos a través del uso de voz o buscadores en lugar de interfaces *drag & drop*. Integración de chatbots y asistentes virtuales en el flujo de trabajo analítico.

- Precios por suscripción, a la baja y acuerdos corporativos: principalmente buscando competir con el modelo que propone Microsoft con PowerBI.
- *Storytelling* de datos: serán valoradas plataformas que permitan generar más fácilmente historias a partir de los resultados.
- Privacidad: se comenzarán a proveer mecanismos para clasificar y alertar sobre el uso de datos personales o elementos que deberían ser eliminados de reportes, tableros o resultados. Las plataformas deberán adaptarse para ayudar a las empresas a cumplir con regulaciones de protección de datos.
- Capacidades de escalabilidad: a pesar de estar avanzadas hoy, mejorarán en el procesamiento de mayores volúmenes de macrodatos y utilización de modelos de datos cada vez más complejos.
- Adopción de la nube: se da principalmente en nuevos despliegues. Amazon y Google con sus soluciones no cumplen con los criterios de inclusión en el cuadrante mágico, aunque muy probablemente serán protagonistas durante los próximos años.

Figura 12. Cuadrante mágico para plataformas analíticas e inteligencia de negocios



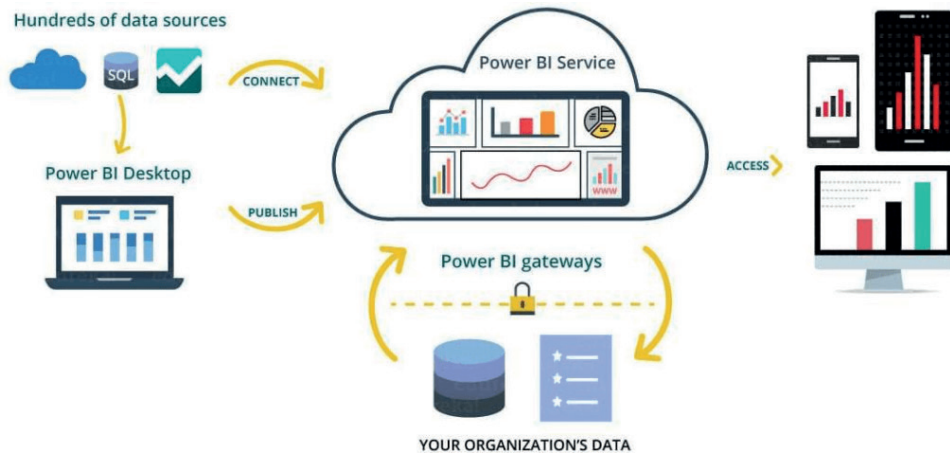
Fuente: Parenteau et al.,2020

Microsoft

Microsoft ofrece una amplia gama de capacidades tanto locales como en la nube de Microsoft Azure. Microsoft Power BI ofrece inteligencia de negocios basada en la nube con una nueva interfaz de escritorio. Power BI consta de una aplicación de escritorio de Windows denominada Power BI Desktop y aplicaciones móviles disponibles para teléfonos y tabletas Windows, así como para dispositivos iOS y Android (Microsoft Inc., 2016).

Microsoft ha bajado sustancialmente el precio de sus soluciones, lo que la convierte en una de las más competitivas en el mercado hoy en día. El flujo de trabajo habitual comienza en Power BI Desktop (figura 13), donde se integran los datos y se crea un informe, luego, ese informe se publica en el servicio Power BI donde se crean nuevas visualizaciones y después se comparte para que los usuarios de las aplicaciones de Power BI Mobile puedan compartir y usar la información.

Figura 13. Power BI service



Fuente: Microsoft, 2016

El modelo de Microsoft basado en la nube es destacado por los clientes gracias a sus bajos precios por usuario, la facilidad de uso para las empresas y su disponibilidad de recursos especializados.

Tableau

Tableau ofrece productos altamente interactivos e intuitivos, que permiten a los usuarios acceder fácilmente, preparar y analizar sus datos sin necesidad de codificación. Tableau se ha centrado en mejorar la experiencia analítica, y la facilidad de uso es el principal objetivo de gran parte de sus esfuerzos de desarrollo de productos. Una

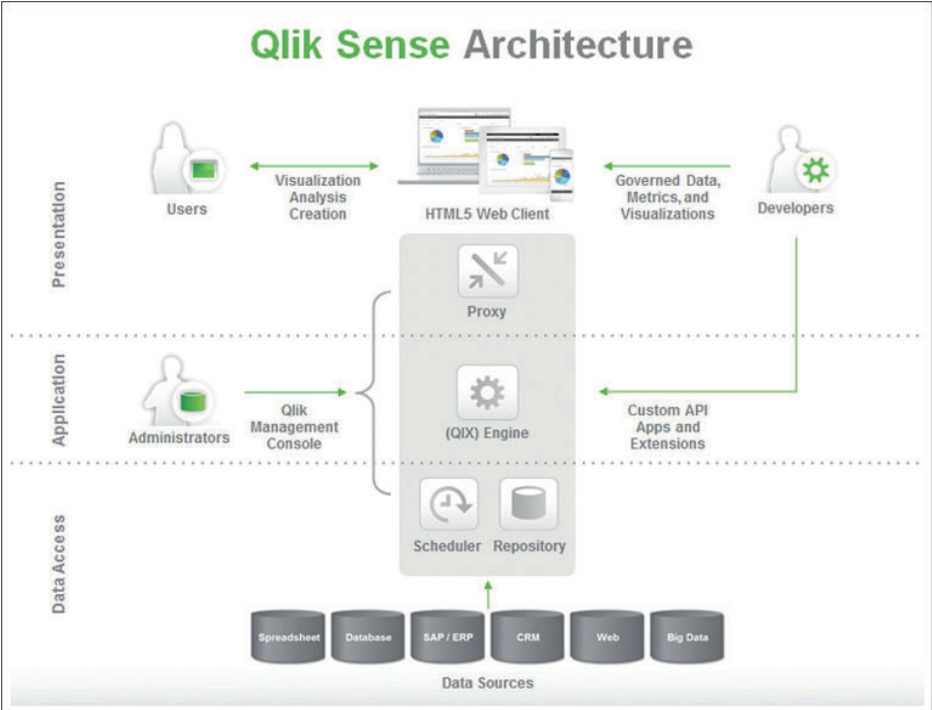
de sus principales virtudes es su versatilidad, gracias al despliegue local y a través de la nube.

Tableau ofrece una amplia gama de opciones de aprendizaje como tutoriales en línea, seminarios y capacitación práctica para educar y capacitar a sus usuarios (Tableau, 2017). Los puntos fuertes de Tableau son su diversa gama de conectividad de origen de datos, así como sus capacidades interactivas de visualización y exploración (Murillo et al., 2016). Así cumple la misión de ayudar a las personas a ver y comprender sus datos, con un acceso rápido a prácticamente cualquier fuente de datos, donde los usuarios no técnicos pueden comenzar a interactuar de inmediato con una interfaz visual intuitiva.

Qlik

Qlik ofrece información y análisis de datos gobernados a través de sus dos productos principales (figura 14): QlikView y Qlik Sense. Qlik está altamente calificado para la facilidad de uso, la complejidad del análisis y los beneficios del negocio (Ilacqua, Cronstrom y Richardson, 2015).

Figura 14. Arquitectura Qlik Sense



Fuente: Ilacqua, Cronstrom y Richardson, 2015

En comparación con sus principales competidores, Qlik obtiene puntuaciones más altas en la complejidad del análisis, debido a su mayor capacidad para soportar múltiples fuentes de datos, un motor de cálculo robusto y filtrado y búsqueda asociativa (Qlik, 2017).

ThoughtSpot

ThoughtSpot irrumpió en 2017 en un modesto puesto en el cuadrante de “jugadores de nicho”. En 2018 se posicionó en el cuadrante de los visionarios, sin duda gracias a su nuevo motor de búsqueda orientado al análisis numérico. En 2019 dio el salto a los líderes y es un nuevo competidor para tener en cuenta, junto a los tres grandes que durante los últimos años han ocupado este cuadrante.

La filosofía de este nuevo motor de búsqueda está orientada al usuario final con dos de sus soluciones: SpotIQ, que usa la potencia de la inteligencia artificial y promete responder miles de consultas en cuestión de segundos sobre billones de datos, así como devolver docenas de propuestas al usuario para analizar, y SearchIQ, que permite conversaciones analíticas con el motor de búsqueda sobre sus propios datos utilizando un lenguaje natural, además de proveer los resultados al usuario mediante un nuevo conjunto de herramientas analíticas.

Beneficios de la inteligencia de negocios

En la literatura se encuentran estudios que argumentan que el uso de los sistemas de inteligencia de negocios pueden traer numerosos beneficios. La inteligencia de negocios puede ofrecer ciertas ventajas competitivas a las empresas, ya que generalmente proporciona una mayor funcionalidad en cuanto a acceso y análisis en comparación con los sistemas de planificación de recursos empresariales (ERP) (Hočevár y Jaklič, 2008). Por otra parte, el procesamiento analítico de la información sobre los clientes y su comportamiento contribuye a optimizar la gestión de las relaciones con ellos, a maximizar su satisfacción y a aumentar su lealtad y rentabilidad.

Otros beneficios son la reducción de costos, el aumento de los ingresos, mejora en la satisfacción del cliente y en la comunicación dentro de la empresa (Moss y Atre, 2003; Carver y Ritacco, 2006; Watson, 2009; Işık, Jones y Sidorova, 2013). La información se considera a menudo como el segundo recurso más importante que una empresa tiene, junto con su gente. Así que cuando la empresa puede tomar decisiones basadas en información precisa, puede mejorar su rendimiento.

Ranjan (2009) menciona otros beneficios para las empresas con la adopción de soluciones de inteligencia de negocios:

- Identificar los clientes rentables y las razones de su lealtad, así como a clientes potenciales.

- Analizar datos de flujo de clics para mejorar las estrategias de comercio electrónico.
- Determinar qué combinaciones de productos y líneas de servicio comprar y cuándo.
- Detectar y disuadir conductas fraudulentas, como los picos de uso cuando las tarjetas de crédito son robadas.
- Mejorar el proceso de creación de informes para datos sofisticados sin necesidad de tecnologías analíticas. Además, el acceso remoto a cualquier dato deseado a través de un teléfono inteligente (Mohamadina et al., 2012).
- Hacer la detección temprana de riesgos para identificar oportunidades de negocio antes que los competidores (Azma y Mostafapour, 2012).

Las organizaciones saben muy bien que la información vital para la toma de decisiones está en sus bases de datos. Por lo tanto, la inteligencia de negocios permite ampliar la visión estratégica y reducir el riesgo y la incertidumbre para construir ventajas competitivas en un mercado cada vez más desafiante.

Macrodatos (*big data*)

El aumento en el volumen de datos ha creado la necesidad de generar nuevos modelos de almacenamiento y tratamiento de datos que compensen las falencias de los sistemas tradicionales. Así, los macrodatos son “conjuntos de datos cuyo tamaño está más allá de la capacidad de las herramientas típicas de software de base de datos para capturar, almacenar, administrar y analizar” (Manyika et al., 2011, p. 91).

Cadena de valor

La cadena de valor de los macrodatos se puede dividir en cuatro fases: generación, adquisición, almacenamiento y análisis de datos (Chen, Mao y Liu, 2014). La primera fase se entiende como un proceso en el cual se genera una cantidad inmensa de datos, estructurados o no, pero con características de gran valor. En la segunda, se recopilan, transmiten y preprocesan los datos, ya que en muchos casos hay redundancia o datos inútiles que afectan los resultados de la fase de análisis.

El almacenamiento de macrodatos ha creado la necesidad de nuevas estrategias como almacenamiento con conexión directa y en red, así como diferentes motores NoSQL. Finalmente, la fase de análisis tiene el objetivo de extraer rápidamente información para poder generar valor en las organizaciones y facilitar procesos de toma de decisiones.

Generación de datos

Los datos de internet pueden no tener valor individualmente, pero grandes cantidades pueden mostrar información útil. Las principales fuentes de grandes datos son la información operativa y comercial de las empresas, la generada por el internet de las cosas (IOT) y la investigación científica. El IOT es una fuente importante de datos. La adquisición de los datos consiste principalmente en redes de sensores, luego la información se transmite y procesa y, finalmente, llega a una capa de aplicación que soporta aplicaciones específicas de IOT.

Adquisición de datos

La recopilación de datos consiste en utilizar técnicas especiales para adquirir datos brutos de un entorno específico. Los archivos de registro, los sensores y el rastreador web son ejemplos de ello. Además, hay otros, como las herramientas especiales usadas en el ámbito científico para recolectar datos experimentales (v. gr. espectrómetros magnéticos y radiotelescopios).

Una vez completada la recopilación de datos en bruto, estos se transfieren a una infraestructura de almacenamiento para su procesamiento y análisis. Los datos grandes van principalmente a un centro de datos. El diseño de los datos debe ajustarse para mejorar la eficiencia de la computación o facilitar el mantenimiento del *hardware*.

Debido a la gran variedad de fuentes, los conjuntos de datos recolectados varían con respecto al ruido, la redundancia y la consistencia, por ello se requiere un preprocesamiento que incluya la integración, limpieza y eliminación de redundancia, pues es indudablemente un desperdicio almacenar datos sin sentido.

Almacenamiento de datos

Es el almacenamiento y manejo de conjuntos de datos a gran escala, al tiempo que se logra confiabilidad y disponibilidad para acceder a ellos. Hay de tres tipos: masivo, distribuido y de grandes datos. Las tecnologías de almacenamiento masivo pueden clasificarse como DAS (*Direct Attached Storage*) y almacenamiento en red, y a su vez el almacenamiento en red puede clasificarse en NAS (*Network Attached Storage*) y SAN (*Storage Area Network*). DAS se utiliza principalmente en computadoras personales y servidores de pequeño tamaño, y el almacenamiento en red consiste en utilizar la red para proporcionar a los usuarios una interfaz de unión para el acceso y el intercambio de datos.

Por otro lado, para utilizar un sistema distribuido para almacenar datos masivos deben tenerse en cuenta factores como la consistencia (asegurar que las copias múltiples de los mismos datos sean idénticas), la disponibilidad (que el sistema no se vea gravemente afectado para satisfacer las peticiones del cliente en términos de

lectura y escritura), y la tolerancia de partición (que el almacenamiento distribuido siga funcionando bien cuando la red esté segmentada). Dado que la coherencia, disponibilidad y tolerancia de las particiones no pueden lograrse simultáneamente, se puede tener un sistema de CA (*Consistency-Availability*) que ignore la tolerancia de partición, un sistema de CP (*Consistency-Partition*) que ignore la disponibilidad, y una AP (*Availability- Partition*) que ignore la consistencia de acuerdo con diferentes objetivos de diseño.

Finalmente, los mecanismos de almacenamiento para grandes datos pueden clasificarse en tres niveles: i) sistemas de archivos, ii) bases de datos, y iii) modelos de programación. Los sistemas de archivos son la base de las aplicaciones en los niveles superiores. Por ejemplo, Microsoft desarrolló Cosmos (Chaiken et al., 2008) para apoyar su negocio de búsqueda y publicidad, y Facebook utiliza Haystack (Beaver et al., 2010) para almacenar gran cantidad de fotos de pequeño tamaño. Por otro lado, las bases de datos NoSQL (bases de datos relacionales no tradicionales) se están convirtiendo en la tecnología principal de los grandes datos. Se destacan tres: de valores clave, orientadas a columnas, y enfocadas en documentos, cada una basada en ciertos modelos de datos. En las primeras encontramos Dynamo de Amazon (DeCandia, et al., 2007) y Voldemort, utilizado por LinkedIn. En las segundas están BigTable de Google, Cassandra (Lakshman y Malik, 2009) y HBase (George, 2011), y por último hay tres representantes importantes de sistemas de almacenamiento de documentos como MongoDB (Chodorow, 2013), SimpleDB (Murty, 2008) y CouchDB (Anderson, Lehnardt y Slater, 2010).

Algunos modelos de programación paralelos pueden mejorar el rendimiento de NoSQL y reducir la brecha de rendimiento a las bases de datos relacionales. Por lo tanto, estos modelos se han convertido en la piedra angular para el análisis de datos masivos. Entre ellos están MapReduce (Dean y Ghemawat, 2008), Dryad (Sard et al., 2007), All-Pairs (Moretti et al., 2008) y Pregel (Malewicz et al., 2010).

Análisis de datos

Muchos métodos tradicionales todavía se pueden utilizar para el análisis de datos de gran tamaño. A continuación presentamos algunas técnicas aplicables en una amplia gama de industrias.

- *Pruebas A/B*. Técnica en la que un grupo de control se compara con otros grupos de prueba con el fin de determinar qué cambios mejorarán una variable objetivo. Un ejemplo de aplicación es determinar diseños, imágenes o colores que podrían mejorar un página web (Unbounce, 2015). Los datos grandes permiten ejecutar y analizar un gran número de pruebas, así es posible detectar diferencias estadísticamente significativas entre los grupos control y tratamiento.

- *Fusión de datos e integración de datos.* Conjunto de técnicas que integran y analizan datos de múltiples fuentes, con el fin de desarrollar conocimientos de manera más eficiente y potencialmente más precisa que si se tomaran de una sola (Khaleghi et al., 2013). Por ejemplo, los datos de las redes sociales se pueden combinar con los de ventas para determinar qué efecto tiene una campaña de mercadeo en el cliente y el comportamiento de compra.
- *Minería de datos.* Proceso para extraer información o conocimientos ocultos o desconocidos, pero potencialmente útiles, de datos masivos, incompletos, ruidosos, difusos y aleatorios (Han, Pei y Kamber, 2011). Estas técnicas incluyen reglas de asociación, análisis de conglomerados, clasificación y regresión.
 - Reglas de asociación. Consisten en algoritmos para generar y probar posibles reglas. Una aplicación es analizar la cesta de mercado y determinar qué productos se compran juntos con frecuencia, para usar esa información en estrategias de comercialización.
 - Análisis de conglomerados. Es un método estadístico para agrupar objetos y, específicamente, clasificarlos según ciertas características. El análisis de clúster se utiliza para diferenciar objetos con características particulares y dividirlos en categorías (clústeres) de acuerdo con esas características, de manera que los objetos de la misma categoría tendrán una alta homogeneidad, mientras que diferentes categorías tendrán una alta heterogeneidad (Berkhin, 2006). El análisis de clúster es un método de estudio sin supervisión ni datos de entrenamiento.
- *Optimización.* Utilizada para rediseñar sistemas y procesos complejos con el fin de mejorar su desempeño de acuerdo con una o más medidas objetivas (por ejemplo costo, velocidad o confiabilidad). Los algoritmos genéticos son un ejemplo de una técnica de optimización.
 - Algoritmos genéticos. Se inspiran en el proceso de evolución natural o “supervivencia del más apto”. En esta técnica, las soluciones potenciales se codifican como “cromosomas” que pueden combinarse y mutarse. Esos cromosomas individuales se seleccionan para sobrevivir dentro de un “ambiente” modelado que determina la aptitud o el rendimiento de cada individuo en la población (Valencia, 1997). A menudo descrito como un tipo de “algoritmo evolutivo”, estos algoritmos son adecuados para resolver problemas no lineales.
- *Aprendizaje automático.* Una subespecialidad de la informática (dentro de un campo que históricamente se llama “inteligencia artificial”), se refiere al diseño y desarrollo de algoritmos que permiten a las computadoras evolucionar

comportamientos a partir de datos empíricos (Paluszek y Thomas, 2017). Un foco del aprendizaje de máquina es aprender a reconocer patrones y tomar decisiones inteligentes basadas en datos. El procesamiento del lenguaje natural es un ejemplo de aprendizaje automático.

- Procesamiento del lenguaje natural (PLN). Técnica dentro de un campo llamado históricamente “inteligencia artificial” y lingüística, que utiliza algoritmos informáticos para analizar el lenguaje humano (natural) (Augusto, Vega y Pariona, 2009). Muchas técnicas de PLN son tipos de aprendizaje automático. Una aplicación de PLN podría analizar los sentimientos en las redes sociales para determinar cómo reaccionan los clientes a una campaña de marca.
- *Redes neuronales*. Modelos computacionales inspirados en la estructura y el funcionamiento de las redes neuronales biológicas (es decir, las células y las conexiones dentro de un cerebro), que encuentran patrones en los datos (Bose y Liang, 1996). Son adecuadas para hallar patrones no lineales. Algunas aplicaciones de redes neuronales implican aprendizaje supervisado y otras aprendizaje no supervisado. Ejemplos de aplicaciones incluyen la identificación de reclamaciones fraudulentas de seguros.
- *Análisis de red*. Conjunto de técnicas utilizadas para caracterizar relaciones entre nodos discretos en un grafo o una red. En el análisis de redes sociales se analizan las conexiones, por ejemplo, cómo circula la información o quién tiene la mayor influencia sobre quién.
- *Modelado predictivo*. Modelo matemático para predecir mejor la probabilidad de un resultado. La regresión es un ejemplo de las muchas técnicas de modelización predictiva.
 - Regresión. Conjunto de técnicas estadísticas para determinar cómo cambia el valor de la variable dependiente cuando se modifica una o más variables independientes (Chatterjee y Hadi, 2015). A menudo se utiliza para pronóstico o predicción. Ejemplos de aplicaciones incluyen la previsión de volúmenes de ventas basados en diversas variables económicas y de mercado.
- *Análisis espacial*. Conjunto de técnicas que analizan las propiedades topológicas, geométricas o geográficas codificadas en un conjunto de datos. Con frecuencia los datos para el análisis espacial provienen de sistemas de información geográfica (SIG) que incluyen la localización, por ejemplo, direcciones o coordenadas de latitud y longitud (Fotheringham y Rogerson, 2013). Algunas aplicaciones incorporan simulaciones (*v. gr.*, ¿cómo funcionaría una red de cadena de suministro de fabricación con sedes en diferentes ubicaciones?).

- *Análisis de series de tiempo.* Conjunto de técnicas tanto de estadísticas como de procesamiento de señales para analizar secuencias de puntos de datos, que representan valores en momentos sucesivos para extraer características significativas de los datos (Nava, 2015). Ejemplos de análisis de series de tiempo incluyen predecir el número de personas que serán diagnosticadas con una enfermedad infecciosa.
- *Visualización.* Técnicas utilizadas para crear imágenes, diagramas o animaciones para comprender y mejorar los resultados de grandes análisis de datos.
- *Nube de etiquetas.* Es una visualización del texto en forma de una nube, es decir, una lista visual ponderada en la que las palabras más frecuentes son más grandes y las que aparecen con menor frecuencia son más pequeñas. Este tipo de visualización ayuda al lector a percibir rápidamente los conceptos más destacados en un gran cuerpo de texto.
- *Clustergram.* Técnica de visualización utilizada para el análisis de clústeres, que muestra cómo los miembros individuales de un conjunto de datos se asignan a los clústeres a medida que aumenta el número de estos. La elección del número de clústeres es un parámetro importante en el análisis (Schonlau, 2002). Esta técnica permite al analista comprender mejor cómo los resultados del agrupamiento varían con diferentes números de grupos.

En la actualidad, los principales métodos de procesamiento de grandes datos se muestran de la siguiente forma:

- *Bloom Filter.* Consiste en una serie de funciones Hash. El principio de Bloom Filter es almacenar valores de Hash de datos distintos de los propios datos mediante una matriz de bits, la cual en esencia es un índice de mapa de bits que utiliza funciones de Hash para llevar a cabo el almacenamiento con pérdida de compresión de datos (Broder y Mitzenmacher, 2011). Tiene ventajas tales como alta eficiencia espacial y alta velocidad de consulta, pero también tiene algunas desventajas en la falta de reconocimiento y eliminación.
- *Hashing.* Método que transforma esencialmente los datos en valores numéricos de longitud fija más cortos o valores de índice (Holt y Chung, 2002). Hashing tiene ventajas tales como lectura rápida, escritura y alta velocidad de consulta, pero es difícil encontrar una función de Hash de sonido.
- *Computación paralela.* Comparada con la computación en serie tradicional, la computación paralela se refiere a la utilización simultánea de varios recursos computacionales para completar una tarea de cálculo (Asanovic et al., 2009). Su idea básica es descomponer un problema y asignarlo a varios procesos sepa-

rados para ser completados independientemente y lograr el coprocesamiento. Actualmente, algunos modelos clásicos de computación paralela incluyen MPI (*Message Passing Interface*), MapReduce y Dryad.

Macrodatos y creación de valor

Hay cinco maneras ampliamente aplicables de aprovechar el potencial de los grandes datos para crear valor:

- *Crear transparencia.* Hacer que los datos sean más fácilmente accesibles puede reducir el tiempo de búsqueda y procesamiento, incrementar el descubrimiento y mejorar la rendición de cuentas (Kim, Trimi y Chung, 2014).
- *Habilitar la experimentación para descubrir las necesidades y mejorar el rendimiento.* Las tecnologías de la información (TI) permiten a las organizaciones instrumentar procesos y establecer experimentos controlados. El uso de datos para analizar la variabilidad surgida naturalmente o por experimentos controlados y para comprender sus causas fundamentales, les sirve a los líderes para administrar el rendimiento a niveles más altos (McAfee y Brynjolfsson, 2012; Yiu, 2012).
- *Segmentación de las poblaciones para personalizar las acciones.* El uso de los grandes datos contribuye a que las organizaciones adapten productos y servicios a las necesidades de los segmentos del mercado (Elgendy y Elragal, 2014). Familiarizado en mercadotecnia y gestión de riesgos, este enfoque puede ser revolucionario en otros lugares —por ejemplo, en el sector público—. Las empresas comienzan a desplegar en tiempo real la microsegmentación de los clientes para orientar las promociones.
- *Algoritmos para apoyar las decisiones.* La analítica sofisticada puede mejorar sustancialmente la toma de decisiones y desenterrar conocimientos valiosos que de otra manera permanecerían ocultos. Dichos análisis pueden desplegarse para triar los riesgos más altos de regulación y cumplimiento, así como para afinar los inventarios y los precios en respuesta a las ventas en tiempo real en las tiendas y en línea. Las decisiones pueden tomarse a partir del comportamiento de poblaciones enteras en lugar de basarse en muestras. La toma de decisiones nunca puede ser la misma; algunas organizaciones ya lo hacen (Raghupathi y Raghupathi, 2014; Demirkan y Denle, 2013; Provost y Fawcett, 2013) gracias al análisis de conjuntos de datos completos de clientes, empleados o incluso sensores incorporados en los productos.
- *Nuevos modelos, productos y servicios innovadores.* Los grandes datos permiten a las empresas crear productos y servicios (Groves et al., 2013; Heerschap et al., 2014), mejorar los ya existentes e inventar nuevos modelos de negocio. Los fa-

bricantes utilizan los datos para mejorar el desarrollo de la próxima generación de productos y crear servicios innovadores.

Retos del *big data*

El diluvio de datos trae consigo enormes desafíos en la adquisición, almacenamiento, administración y análisis de datos. Los sistemas tradicionales se basan en el sistema de gestión de bases de datos relacionales, y a su vez estos se aplican solo a datos estructurados, distintos de los datos semiestructurados o no estructurados. Algunos de los retos clave son:

- *Representación de los datos.* Muchos conjuntos de datos tienen ciertos niveles de heterogeneidad en el tipo, la estructura, la semántica, la organización, la granularidad y la accesibilidad (Chen y Zhang, 2014). La representación de datos tiene como objetivo hacer que estos sean más significativos para el análisis por computadora y la interpretación del usuario. Sin embargo, una representación inadecuada reducirá el valor de los datos originales.
- *Reducción de redundancia y compresión de datos.* Son eficaces para reducir el coste indirecto de todo el sistema sobre la base de que los valores potenciales de los datos no se ven afectados.
- *Gestión del ciclo de vida de los datos.* En comparación con los avances relativamente lentos de los sistemas de almacenamiento, la detección omnipresente y la informática produce datos a velocidades y escalas sin precedentes (Khan et al., 2014). Uno de los retos más apremiantes es que el sistema de almacenamiento actual no puede soportar datos masivos. Debe desarrollarse un principio de importancia de los datos relacionado con el valor analítico para decidir cuáles se almacenarán y cuáles se descartarán.
- *Mecanismo analítico.* El sistema analítico de grandes datos procesará masas de datos heterogéneos dentro de un tiempo limitado. Sin embargo, los RDBMS tradicionales están estrictamente diseñados con una falta de escalabilidad y capacidad de expansión, que no puede satisfacer los requisitos de rendimiento (Fan, Han y Liu, 2014). Bases de datos no relacionales han demostrado sus ventajas únicas en el procesamiento de datos no estructurados. Aun así, todavía existen algunos problemas de bases de datos no relacionales en su rendimiento y aplicaciones particulares (Chen, Member y Lin, 2014).
- *Confidencialidad de los datos.* En la actualidad, la mayoría de los grandes propietarios o proveedores de servicios de datos no pueden mantener ni analizar de manera efectiva enormes conjuntos de datos debido a su limitada capacidad. Deben contar con profesionales o herramientas para analizarlos, lo que aumenta

los riesgos potenciales de seguridad (Xu et al., 2014). Por ejemplo, el conjunto de datos transaccional generalmente incluye un conjunto de datos operativos completos para impulsar procesos empresariales clave. El análisis de grandes datos puede ser entregado a un tercero para su procesamiento solo cuando se tomen las medidas preventivas adecuadas para proteger información sensible.

- *Gestión de energía.* El consumo de energía de los sistemas informáticos principales ha atraído mucho la atención de los economistas y de los ambientalistas. El aumento del volumen, el análisis, el procesamiento, almacenamiento y la transmisión de grandes datos consumirá inevitablemente más energía eléctrica (Kaushik y Nahrstedt, 2012). Esta realidad constituye un reto, pues se requiere controlar el consumo de energía a escala de sistema, garantizando el mínimo de costos y de impacto al medio ambiente.
- *Cooperación.* El análisis de grandes datos es una investigación interdisciplinaria, que exige la cooperación de expertos en diferentes campos para aprovechar su potencial. Debe establecerse una arquitectura amplia de redes de datos para ayudar a científicos e ingenieros en diversos campos a acceder a diferentes tipos de datos y a utilizar plenamente sus conocimientos técnicos, a fin de cooperar para completar los objetivos analíticos (Ortoll, Canals y Cobarsí, 2014).

La revolución de los datos ha traído más ventajas y beneficios en sectores como los negocios, la salud, la ciencia y el gobierno, y gracias a ello se ha contribuido a la solución de problemáticas específicas. Sin embargo, también implica nuevos desafíos que no se contemplaban en los métodos tradicionales, que van desde la captura y el almacenamiento de los datos, hasta su análisis e interpretación. No se puede dejar de lado la cultura en las organizaciones, pues de esta depende el éxito o fracaso al emprender este tipo de iniciativas.

Iniciativas de datos abiertos

El Barómetro de Datos Abiertos (Open Data Barometer, ODB) es un observatorio promovido por la World Wide Web Foundation, cuyo objetivo es evaluar qué tan preparados están los Gobiernos para implementar políticas y prácticas de datos abiertos en sus países, y el impacto de esas iniciativas en la vida de sus habitantes. El ODB analiza las tendencias globales y proporciona datos comparativos de países y regiones utilizando una metodología que combina datos contextuales, evaluaciones técnicas e indicadores secundarios.

El ODB cubre 115 jurisdicciones, y clasifica a los Gobiernos según:

- *Disponibilidad* para iniciativas de datos abiertos.
- *Implementación* de programas de datos abiertos.
- *Impacto* de los datos abiertos en los negocios, la política y la sociedad civil.

La cuarta edición del ODB, realizada en 2016, está basada en tres tipos de datos (World Wide Web Foundation, 2018):

- Una encuesta de expertos con un rango de preguntas sobre contextos de datos abiertos, políticas, implementación e impactos, y una encuesta detallada del conjunto de datos completada para 15 tipos de datos en cada país, que incluyó temas de disponibilidad, formato, licencia, oportunidad y capacidad de descubrimiento.
- Una encuesta simplificada de autoevaluación del Gobierno con el mismo rango de contexto, implementación e impacto, como fuente adicional de información.
- Datos secundarios seleccionados para complementar las encuestas de expertos. Estos se toman del Foro Económico Mundial, del Banco Mundial, y de la Encuesta de Gobierno Electrónico de las Naciones Unidas y de la ONG Freedom House.

La Carta Internacional de los Datos Abiertos (que contiene los principios rectores para liberar datos abiertos de los Gobiernos) es un marco para incorporar la cultura y la práctica de la apertura en el Gobierno. La Carta propone seis principios para la publicación de datos:

- *Abiertos por defecto.* El Barómetro analiza la existencia y la calidad de 15 conjuntos de datos clave (como registros de tierras o presupuestos gubernamentales) en los 115 países. Estos conjuntos de datos se recopilan de alguna forma en el 97 % de los países. Sin embargo, el 29 % de esos conjuntos de datos aún no se publican en línea, y solo 7 % son realmente abiertos (World Wide Web Foundation, 2018).
- *Oportunos e integrales.* El 74 % de los datos analizados están actualizados, lo cual es prometedor, pero significa que una cuarta parte de todos los datos encuestados tiene un valor muy limitado (World Wide Web Foundation, 2018).
- *Accesibles y utilizables.* El 73 % de los conjuntos de datos fueron relativamente fáciles de encontrar. El 10 % no estaba disponible de forma gratuita. Solo una cuarta parte de los conjuntos de datos analizados contaba con una licencia abierta, lo que significa que las licencias siguen siendo una gran barrera para su uso (World Wide Web Foundation, 2018).
- *Comparables e interoperables.* Un poco más de la mitad de los datos (53 %) está disponible en un formato reutilizable y legible por máquina, pero de estos, solo el 24 % es accesible y se puede descargar en masa (World Wide Web Foundation, 2018).
- *Para mejorar la gobernanza y la participación ciudadana.* El impacto de los datos abiertos en el aumento de la eficiencia y la eficacia del gobierno sigue siendo muy bajo, con un puntaje promedio de solo 1,20 de 10 para todos los Gobiernos en el estudio (World Wide Web Foundation, 2018).
- *Para el desarrollo inclusivo y la innovación.* Cuando se trata de la disponibilidad de datos esenciales para la innovación (como los de mapas o los horarios de transporte público), solo el 8 % de los conjuntos de datos relevantes están realmente abiertos (World Wide Web Foundation, 2018).

Los líderes de cada región son Canadá, Israel, Kenia, Corea, México y Reino Unido. En general, estos hallazgos revelan que los campeones regionales han mejorado constantemente desde la última edición (figura 15).

Figura 15. Ranking y puntajes ODB

Fuente: World Wide Web Foundation, 2018, p. 1

Como conclusión del estudio se encontró que, para obtener los beneficios prometidos del movimiento de datos abiertos, la agenda actual de datos abiertos que persiguen los Gobiernos debe volver a centrarse en lo básico y en las personas. Los Gobiernos deben intensificar sus esfuerzos para garantizar que los datos abiertos sean para todos y que se divulgue realmente lo que las personas necesitan.

Los resultados muestran que los datos no están abiertos en la práctica. En muchos casos los datos abiertos se consideran una responsabilidad adicional. Las iniciativas de datos abiertos no duran más que los líderes o las Administraciones que las iniciaron. El retroceso en las puntuaciones del Barómetro, incluso entre los mejores, refleja esta realidad.

Con base en los hallazgos del Barómetro, la World Wide Web Foundation exige las siguientes acciones de los Gobiernos (World Wide Web Foundation, 2018):

- Los datos del Gobierno deben estar abiertos por defecto.
- Los Gobiernos deben integrar datos abiertos en todas las agencias y departamentos.
- Los Gobiernos deben adoptar la Carta Internacional de los Datos Abiertos para garantizar que las prácticas de datos abiertos se incorporen más allá de los mandatos políticos.
- Los Gobiernos deben consultar a los ciudadanos e intermediarios al priorizar qué datos abiertos se publicarán primero.
- Los Gobiernos deben invertir en el uso de datos abiertos para mejorar la vida de los grupos marginados.

Los beneficios potenciales de los datos abiertos para crear confianza y políticas mejores y más equitativas pueden ocurrir si se implementan estas recomendaciones. De lo contrario, los datos abiertos seguirán siendo una aspiración más que una realidad.

Antecedentes

A partir de 2009, países como Estados Unidos y Reino Unido empezaron a poner en marcha sus iniciativas en materia de datos abiertos y de reutilización de la información. Algunas de las iniciativas más importantes en datos abiertos a escala mundial son el portal Data.Gov donde el Gobierno estadounidense ha dispuesto para los ciudadanos cerca de 235 968 conjuntos de datos (Estados Unidos de América, 2019) para su reutilización. Mientras que la iniciativa de Reino Unido data.gov.uk dispone de 52 545 conjuntos de datos (Reino Unido, 2019).

Cada vez son más los países en donde el concepto de gobierno abierto toma fuerza, ya sea como efecto de cambiar los sistemas tradicionales de gobernanza, poner de manifiesto su intención en procesos, acciones o datos más transparentes, colaborar interior y exteriormente con actores que aporten nuevas ideas en aras de solucionar problemas de carácter social, o simplemente estimular la participación activa de ciudadanos en la formulación y ejecución de políticas.

El Reino Unido es uno de los pioneros en materia de gobiernos abiertos, una de sus motivaciones más relevantes es la de promover la toma de decisiones por parte de la ciudadanía, además de las sugerencias que pueda recibir. Allí la innovación se ha convertido en un pilar fundamental gracias a convocatorias que premian iniciativas relacionadas con datos abiertos gubernamentales, y al desarrollar varios sistemas de información producto de alianzas estratégicas entre el Gobierno y las compañías privadas locales. Hampshire Hub es un claro ejemplo de esta integración, consistente en un proyecto que asocia a proveedores de servicio público dentro y fuera de Hampshire, y que a través de los datos abiertos se suplen de lo que pueda ser ofrecido por el Estado, por personas naturales y hasta por la propia competencia (Ferrer-Sapena y Peset, 2011).

Por otro lado, Europa ha encontrado en la revolución de datos abiertos otra razón para integrarse internamente, muestra de ello son las iniciativas que han ocasionado que los Estados unifiquen en un solo portal web el contenido de los datos abiertos y su reutilización (<https://data.europa.eu/euodp/es/data>) (Giménez, 2014). Para 2019 este portal contaba con 13 852 datasets disponibles en veinte grupos diferentes, como trabajo y empleo, vida económica, industria, ciencia, medio ambiente, transporte, energía y Unión Europea, entre otros (Portal de Datos Abiertos de la UE, 2019).

En el continente americano, *Estados Unidos* es pionero en la reutilización de la información pública, y no solo porque posee portales especializados con una amplio catálogo de datos, sino porque también ostenta un robusto brazo legislativo que favorece la promulgación de información en el país; tal es el caso de la rendición de cuentas de fondos federales y la Ley Federal de Responsabilidad y Transparencia de los Fondos (FFATA, por sus siglas en inglés) promulgada en 2006, que dentro de sus múltiples objetivos busca empoderar a los estadounidenses para que responsabilicen al Gobierno del gasto público, y en este orden de ideas se haga seguimiento a los recursos públicos. De esta manera la FFATA se materializa a través de un único portal de búsqueda denominado *usaspending.gov* (Gobierno de los Estados Unidos de América, 2017).

Sin embargo, en la última versión del Barómetro de Datos Abiertos *Canadá* tomó el primer lugar por sus significativos avances en implementación de datos abiertos en su jurisdicción, muestra de ello es el portal centralizado *open.canada.ca*, el cual posee poco más de 82 806 datasets, y un portal de transparencia (<https://transparencycanada.ca/>), que divulga datos de diversa índole sobre su gestión financiera. Así la ciudadanía ha adoptado un papel activo desde cuando el Gobierno estatal decidiera adquirir compromisos ante la Alianza para el Gobierno Abierto (OGO), que hasta la fecha ha cumplido a cabalidad (Gobierno de Canadá, 2016).

Latinoamérica y el Caribe

Los resultados de Barómetro 2016 indican que Latinoamérica muestra un desempeño relativamente positivo comparado con el promedio global. Hay tres países de la región (México, Uruguay y Brasil) entre los primeros veinte, y cinco entre los treinta (se suman Colombia y Chile). En particular, los resultados destacan a México y Uruguay gracias políticas centradas en mejores servicios públicos que benefician a sus ciudadanos.

Uno de los países de la región que se esforzó por promulgar iniciativas de transparencia fiscal y rendición de cuentas, así como de acceso a la información pública fue *México*. La Ley Federal de Transparencia y Acceso a la Información Pública de 2002 constituyó una base lo suficientemente sólida como para soportar futuras iniciativas en materia de regulación, y preparó el país para la llegada de datos abiertos (Congreso Nacional de los Estados Unidos Mexicanos, 2002).

Otra ley crucial para que la apertura de datos en México se llevara a cabo fue la Ley Federal de Protección de Datos Personales firmada en 2011 (Congreso Nacional de los Estados Unidos Mexicanos, 2011). También en ese año se planteó un esquema inicial de interoperabilidad para publicar las bases, políticas, y los principios para compartir servicios digitales dentro de las múltiples dependencias de la nación,

esto facilitó el desarrollo de plataformas especializadas de datos abiertos en temas como rendición de cuentas (datamx.io) o participación ciudadana, como el sistema Infomex, a través del cual puede solicitarse información pública del Gobierno Federal (Secretaría de la Función Pública, 2011).

México actualmente cuenta con el Instituto Federal de Información y Protección de Datos (IFAI), dependencia encargada de garantizar el derecho de los ciudadanos a la información pública gubernamental y a la privacidad, la rendición de cuentas por parte del Estado y a fomentar el uso de la información pública, es decir, es una entidad promotora de la ley de transparencia en el Estado mexicano (INAI, 2015).

Uruguay por su parte, en su primer Plan de Acción de Gobierno Abierto se enfocó en tres grandes líneas: aumentar la integridad pública, gestionar de forma más eficiente los recursos públicos, y mejorar la prestación de servicios públicos al ciudadano (Presidencia de la República Oriental del Uruguay, 2014). Para gestionar de forma más eficiente los recursos públicos, se hicieron esfuerzos relacionados con la Agencia de Compras y Contrataciones del Estado (ACCE), que en 2013 inició la apertura electrónica. Además, se implementó el Registro Único de Proveedores del Estado (RUPE), que mejora y simplifica la relación entre el Estado y sus proveedores, y se creó un sistema profesional de reclutamiento y selección para la búsqueda de los mejores candidatos para trabajar con el Estado (www.uruguayconcurso.gub.uy) llamado Uruguay Concurso (Presidencia de la República Oriental del Uruguay, 2014).

Para mejorar la prestación de servicios públicos al ciudadano, se aprobaron los decretos 177/013 y 178/013 de simplificación de trámites, y la reglamentación de algunos artículos de la Ley 18719, que versa sobre el intercambio de información entre distintos organismos públicos (Presidencia de la República Oriental del Uruguay, 2014; Parlamento del Uruguay, 2008).

Para fortalecer los datos abiertos, desde diciembre 2012 se dispone del Catálogo Nacional de Datos Abiertos. Se realizó el DateIdea.uy, primer concurso de datos abiertos que cuenta con tres categorías: “aplicaciones existentes”, “prototipos” e “ideas”. Además, se aprobó el Decreto de Simplificación de Trámites, y Uruguay fue sede de la Primera Conferencia Regional de Datos Abiertos para América Latina y el Caribe, con la participación de periodistas, desarrolladores y expertos de Inglaterra, Francia, Alemania, España y Argentina (Presidencia de la República Oriental del Uruguay, 2014).

Brasil es uno de los países líderes en la región en la incorporación de la gestión pública a una era de transparencia, muestra de ello fue la cofundación de la Alianza para el Gobierno Abierto junto con Estados Unidos, además de las iniciativas de creación de repositorios de datos tanto para el Gobierno federal (dados.gov.br) como

para temas tan sensibles como la rendición de cuentas en una plataforma centralizada de transparencia (www.portaltransparencia.gov.br), la cual es objeto de mejoras permanentes por cuenta del Estado; lo anterior soportado en un marco legal de acceso a la información pública que contempla leyes como la 12527 de 2011 o la 9507 de 1997 (Fumega y Scrollini, 2013; Presidencia de la República de Brasil, 2011).

Colombia cuenta con Gov.co, el Portal del Estado colombiano (<https://estrategia.gobiernoenlinea.gov.co>) a cargo del Ministerio de Tecnologías de la Información y las Comunicaciones, punto de acceso a la información, trámites y servicios a través de medios electrónicos. En cuanto a los datos abiertos, el Gobierno nacional ha definido los lineamientos para que las entidades publiquen información de forma unificada en el portal de datos abiertos www.datos.gov.co, cuya finalidad es convertirse en el mecanismo de coordinación para el acceso y uso de los datos de las entidades (República de Colombia, 2013).

En Colombia la Constitución Política consagra el acceso a la información como derecho fundamental en los artículos 20 y 74, y lo reconoce como un derecho habilitante de otros derechos. En el 2014, se expidió la Ley 1712 o de Transparencia y del Derecho de Acceso a la Información Pública Nacional, que regula el acceso a la información pública. Su objetivo es poner la información bajo control de cualquier entidad pública, a disposición de todos los ciudadanos e interesados, de manera oportuna, veraz, completa, reutilizable, procesable y en formatos accesibles.

Así, la ley establece una serie de disposiciones para fortalecer algunas obligaciones, amplía las categorías obligadas a su cumplimiento y afianza el sistema de acceso a la información pública a partir de mecanismos de protección más eficientes e incluyentes.

Como instrumentos legales adicionales, el Decreto 1078 de 2015, que desarrolla la Política de Gobierno Digital, así como el Decreto 1081 de 2015, establecen pautas relacionadas con la forma en que las entidades públicas deben implementar lo definido en la ley. Asimismo, la Resolución 3564 de 2015 del Ministerio TIC establece, entre otros aspectos, los estándares para la publicación de información, incluyendo las condiciones técnicas para la apertura de los datos abiertos en Colombia.

Complementario a estas disposiciones, el Gobierno nacional aprobó el documento CONPES 3920 de 2018, Política Nacional de Explotación de Datos (Big Data), cuyo principal objetivo es aumentar el aprovechamiento de datos mediante el desarrollo de las condiciones de apertura y uso de activos que generen valor social y económico. Con la implementación de esta política se espera aumentar el promedio de activos públicos digitalizados y publicados, y que las entidades públicas desarrollen proyectos de aprovechamiento de datos para mejorar los servicios prestados a la ciudadanía.

En cuanto a *Chile*, el presidente Sebastián Piñera (2010-2014) fue el primero que se comprometió a impulsar el gobierno abierto con la Open Government Partnership (OGP), de la cual el país forma parte desde septiembre de 2011. Asimismo, desarrolló planes de gobierno abierto para mejorar el acceso a los servicios públicos, e intentó incorporar todas las herramientas digitales para que el Estado se acercara a la gente, fuera más amistoso, grato y cercano (Fundación CiGob, 2013).

El Gobierno chileno impulsa lo que se denomina la Agenda Digital Imagina Chile 2013-2020, la hoja de ruta gubernamental para impulsar el desarrollo digital y tecnológico del país. En ella se establecen cinco ejes estratégicos: conectividad e inclusión digital, innovación y emprendimiento, educación y capacitación, servicios y aplicaciones, y entorno para el desarrollo digital. Cada uno cuenta con un plan de acción para impulsar la competitividad y la economía digital del país (Gobierno de Chile, 2013).

También se aprobó la Ley 20285 sobre el acceso a la información pública, promulgada en agosto de 2008 por la presidenta Michelle Bachelet (2006-2010, 2014-2018), que entró en vigencia en abril de 2009. Su principal objetivo es abrir y facilitar el acceso a la información a todas las personas, promover un sistema estatal más cercano, con nuevos espacios para la participación, y mejorar las exigencias de rendición de cuentas (Fundación CiGob, 2013).

Además, se han desarrollados varios portales para fortalecer el gobierno abierto como Chile Atiende, que reúne distintos servicios y beneficios de las instituciones públicas en un solo lugar. El Portal de Gobierno Abierto <http://www.gobiernoabierto.cl/> es un espacio virtual que centraliza todas las iniciativas de transparencia, participación y datos públicos del país. El portal Acceso Inteligente busca facilitar el acceso a la información por medio de la transparencia pasiva. Otros portales son: Linked Data de la Biblioteca del Congreso Nacional (BCN), que permite utilizar tecnologías innovadoras para mejorar procesos y servicios al Congreso y a los ciudadanos; Open Data Consejo para la Transparencia, que es el catálogo de datos abiertos (*open data*) del Consejo para la Transparencia, y el Portal de Datos Abiertos www.datos.gob.cl (Gobierno de Chile, 2013).

En el plan de acción propuesto por *Argentina*, se busca fomentar el trabajo en conjunto y promover nuevas soluciones para la ciudadanía. El país cuenta con un portal web que promueve una relación interactiva con la ciudadanía y dar acceso a la información pública (República Argentina, 2015). Otro de los esfuerzos del Gobierno argentino fue la creación de la Ley 26653 de accesibilidad de la información en las páginas web, con el objeto de garantizarles la igualdad real de oportunidades y trato, evitando así todo tipo de discriminación, y el Decreto 1172/03 de Mejora de la Calidad de la Democracia y sus Instituciones, dictado en diciembre del 2003 y

aplicado en todo el ámbito del Poder Ejecutivo Nacional (Presidencia de la Nación Argentina, 2003; Feinman, 2006).

En cuanto al Plan de Acción de Gobierno Abierto (PAGA) de *Paraguay*, en 2012 se realizó el primer plan de acción. El segundo PAGA 2014-2016 tuvo entre los compromisos más importantes la promulgación de la Ley 5282 de 2014 “De libre acceso ciudadano a la información pública y transparencia gubernamental” y el fortalecimiento de los consejos de desarrollo municipal. El Tercer PAGA estuvo compuesto de diez compromisos que contenían 62 metas en total. El 90 % de estos compromisos cumplieron el 100 % de sus metas. Un logro importante en este periodo fue la creación y conformación de la Mesa Conjunta de Gobierno Abierto, foro multisectorial con participación de sector público y organizaciones de la sociedad civil para monitorear el cumplimiento de las metas en cada plan y supervisar la planificación participativa con la sociedad civil (República de Paraguay, 2018; Congreso de la Nación Paraguaya, 2014).

El Portal Único de Gobierno de Paraguay se presentó oficialmente en 2014 (<http://www.paraguay.gov.py/>). En él se encuentra el portal oficial de datos abiertos (<http://www.datos.gov.py/>), que se constituye como un único punto de acceso a los datos abiertos y presenta los conjuntos de datos y las licencias libres asociadas al uso, reuso y distribución de cada set de datos.

En *Perú* como marco normativo, la Constitución (art. 2.º, num. 5) reconoce el derecho fundamental de acceso a la información pública, y el Texto Único Ordenado de la Ley 27806 de 2002, Ley de Transparencia y Acceso a la Información Pública (DS 043-2003-PCM), desarrolla el derecho a la información (Gobierno del Perú, s. f.). En el Plan de Acción de Gobierno Abierto Perú 2012, se propusieron 47 compromisos, los cuales fueron revisados y reorganizados en 12 compromisos nuevos para el Plan de Acción de 2014-2016 (Casas, 2014).

En *Bolivia* no existe una cultura de acceso a la información, sin embargo, el Decreto Supremo 28168 de 2005 tiene por objeto garantizar el acceso a la información, y el Decreto 0214 de 2009 aprueba la Política Nacional de Transparencia y Lucha contra la corrupción (Alianza Regional por la Libre Expresión e Información, 2012; República de Bolivia, 2005). Actualmente, Bolivia no participa de la Alianza para el Gobierno Abierto, y cuenta con un portal de datos abiertos (<https://datos.gob.bo/>) donde apenas se dispone de 29 conjuntos de datos.

En *Venezuela* se identifica retraso, ya que el Gobierno aún se encuentra en fases de aprobación o adopción de leyes que permitan su inserción en la revolución de datos abiertos y, consecuentemente, no cuenta con catálogo de datos abiertos; tampoco forma parte de la Alianza para el Gobierno Abierto y no sitúa los datos abiertos entre sus prioridades (Michener, 2010).

En *Ecuador* la Ley Orgánica de Transparencia y Acceso a la Información Pública (Lotaip) es la norma que regula el ejercicio del derecho a acceder a la información pública. Además, la Constitución Política es una de las constituciones latinoamericanas que desarrolla con mayor precisión el derecho de acceso a la información pública. Ecuador no pertenece a la Alianza para el Gobierno Abierto, pero cuenta con un portal de datos abiertos (<http://catalogo.datosabiertos.gob.ec/>) donde se encuentran 107 documentos que pueden ser consultados por los ciudadanos.

Belice tiene una ley que consagra de manera más o menos amplia el derecho de acceso a la información pública: Freedom of Information Act, 2004. Aunque cuenta con el portal oficial <http://www.belize.gov.bz/>, no dispone de catálogos de datos abiertos ni pertenece a la Alianza para el Gobierno Abierto (Open Government Partnership) (Gobierno de Belice, 2000).

En *Cuba* no existe ninguna garantía constitucional o jurídica respecto al libre acceso a la información pública, el país no pertenece a la Alianza para el Gobierno Abierto y no cuenta con portal de datos abiertos o de gobierno electrónico, solo el portal <http://www.cuba.cu/>, pero no tiene información en formatos abiertos.

Con respecto a las Antillas, *Antigua y Barbuda* tiene una ley de libertad de información del 2004 y sus órganos de aplicación son el Comisionado para la Información, y el Comisionado para la Integridad; en *Bahamas* existe un proyecto de ley sobre la libertad de información; en *Barbados* hay una ley sobre la libertad de información; *Dominica* se encuentra en proceso de elaboración de la ley, y *Granada* tiene un proyecto de ley de libertad de información, aunque no cuenta con reglamentación sobre participación ciudadana. Ninguno de estos países tiene portales de datos abiertos (Cepal, 2013).

Haití tampoco cuenta con ley de acceso a la información pública ni con portal oficial de datos abiertos, aunque sí hay un portal sobre información geoespacial disponible en <http://haitidata.org/>.

Guyana solo cuenta con la ley de acceso a la información del 2011 (Guyana, 2011), mientras que en *Jamaica* el órgano de aplicación de la ley de acceso a la información, publicada en el 2003, es la Unidad de Acceso a la Información (Gobierno de Jamaica, 2002), sin embargo, ninguno de los dos países ha reglamentado la participación ciudadana o portales de datos abiertos.

Desde hace poco más de diez años *República Dominicana* ha mostrado claros esfuerzos por promover la participación ciudadana en su territorio, y con el ánimo de concretar esto, el Gobierno formuló un plan de acción orientado a adherirse a la Alianza para el Gobierno Abierto. En él se identifican compromisos como el de dar apertura a la información y crear plataformas únicas, entre ellas un sitio oficial de datos abiertos, entre otros. Esos compromisos en su mayoría fueron renovados en un

segundo plan de acción presentado ante la misma organización. Del mismo modo, el país cuenta con leyes orgánicas de acceso a la información pública y protección de datos personales (Observatorio Regional de Planificación para el Desarrollo de América Latina y el Caribe, 2016; Congreso Nacional, 2004).

Puerto Rico cuenta con un sitio web oficial de datos abiertos denominado data.pr.gov, el cual divulga datos de acceso libre provenientes del Estado, y herramientas que facilitan su apropiación, no obstante, el país no posee legislación pertinente al acceso a la información pública (está en curso de ser aprobada), lo que ha propiciado que el Gobierno no cuente con una plataforma unificada para la fiscalización de su gestión. Dicha labor ha sido adoptada por organizaciones sin fines de lucro como el Centro de Investigación y Política Pública, cuyo proyecto ABRE tiene como propósito eliminar barreras para acceder a la información pública a través de una página que publica permanentemente datos en formatos abiertos (Estado Libre Asociado de Puerto Rico, 2012).

Al igual que Brasil, *El Salvador* promulgó su ley de acceso a la información pública en 2011, así como la regulación de los servicios de información, de este modo, abrió el paso para que desarrolladores digitales con la dirección del Gobierno dieran apertura a su plataforma oficial datoselsalvador.org, lo cual significa uno de los primeros resultados en materia de desarrollo de plataformas de esta índole en el país. No obstante y a pesar de los esfuerzos, El Salvador se encuentra en una etapa preliminar en la cual debe centrar su empeño en concebir planes estratégicos más estructurados, así como vincular múltiples dependencias en dicho proceso a fin de dar cumplimiento a la legislación vigente en materia de transparencia (República de El Salvador, 2011).

Guatemala por su parte está en una etapa conceptual, la cual se ha formalizado a través de los planes de acción nacional de gobierno abierto. El primero, publicado en 2008, y el último, contemplan un horizonte de planeación hasta el 2016. Las estrategias que pretende emplear el Gobierno para implementar los datos abiertos se sintetizan en cinco compromisos y veintidós acciones concretas para ser ejecutadas por una comisión de transparencia a cargo de la Presidencia de la República. A pesar de lo anterior, Guatemala posee su propia ley de acceso a la información pública desde 2008, en donde además está consignado el derecho a la protección de datos personales, ya que no hay una ley independiente que desarrolle este derecho de forma más amplia e integral (República de Guatemala, 2014; Congreso de la República, 2008).

Honduras durante los últimos años ha focalizado sus esfuerzos en la lucha contra la corrupción, y en este orden de ideas por hacer que la gestión estatal sea lo más transparente posible y propenda a la rendición de cuentas, se creó el Instituto de Acceso a la Información Pública, cuyo objetivo es facilitar el acceso de los ciudadanos a la información pública y a supervisar que los procedimientos de las instituciones

obligadas se ciñan a la Ley de Transparencia y Acceso a la Información Pública firmada en 2006 (República de Honduras, 2014; Congreso Nacional, 2007).

El Gobierno de Honduras, adicionalmente, ha formulado catorce compromisos para los próximos cuatro años, dentro del Segundo Plan de Acción de Gobierno Abierto de Honduras 2014-2018. Este documento hace hincapié en estrategias anti-corrupción en su mayoría, así como en el fomento de la transparencia fiscal. A pesar de los notables esfuerzos para que el Estado ponga en marcha estrategias orientadas al acceso a la información, no hay acciones concretas en el desarrollo de plataformas de datos abiertos que las soporten y, en consecuencia, no existe una plataforma oficial de datos abiertos en el país (República de Honduras, 2014).

Nicaragua aprobó en 2007 la Ley de Acceso a la Información Pública, y en 2012 promulgó la Ley de Protección de Datos Personales. A pesar de tener una base sólida en lo que respecta a la normativa, se ha rezagado respecto a países de la región al no contar con planes bien estructurados en aras de dar cumplimiento a cada una de las leyes, igualmente, hay ausencia de portales de transparencia o rendición de cuentas y de participación ciudadana, y el Estado carece de una plataforma oficial de datos abiertos (República de Nicaragua, 2007; Asamblea Nacional, 2012).

En Centroamérica uno de los países con notorios avances en datos abiertos es *Costa Rica*, en donde a pesar de que no existe aún una ley exclusiva dedicada al tema de acceso a la información pública, el Estado ha tratado constitucionalmente de garantizar este derecho. De igual manera, se ha promulgado la Ley Contra la Corrupción y el Enriquecimiento Ilícito en la Función Pública, la cual establece pautas para prevenir, detectar y sancionar actos corruptos, que soporta el derecho constitucional de garantizar el libre acceso a los departamentos administrativos con el objeto de que la información sea más accesible para la ciudadanía (República de Costa Rica, 1949; Asamblea Legislativa, 2004).

Uno de los logros más trascendentales de Costa Rica, en cuanto a plataformas digitales en el marco del gobierno abierto, es contar en la actualidad con un sitio oficial de datos abiertos (datosabiertos.gob.go.cr) en donde puede encontrarse información vinculada a ministerios y diversas dependencias del país. Adicionalmente, cuenta con un plan de acción de gobierno abierto elaborado en colaboración con la sociedad civil, así como múltiples plataformas que sirven de canal entre el Estado y la ciudadanía, con la intención de mejorar los servicios públicos y facilitar la participación ciudadana y la colaboración (Aragón, s. f.; Costa Rica, s. f.).

Finalmente, *Panamá* en 2013 estructuró un plan de acción para el gobierno abierto, en el cual se consignan compromisos referentes a la transparencia y la lucha contra la corrupción. En él se menciona la importancia de crear plataformas oficiales tanto para garantizar el cumplimiento de la Ley 6 de 2002 (Ley de Transparencia),

como para promover la participación ciudadana. Asimismo, este plan de acción propone crear el sitio oficial de datos abiertos para el país, ya que carece de él en la actualidad. Por su parte, la Autoridad Nacional de Transparencia y Acceso a la Información ejerce la función de promover políticas de transparencia, y cuenta con un sitio web con normativa de conformidad con su misión, y un portal de datos abiertos con cuarenta conjuntos de datos abiertos (Gobierno de Panamá, Antai, 2015; República de Panamá, 2002).

El avance en iniciativas de datos abiertos a escala latinoamericana está concentrado en solo cinco países: México, Uruguay, Brasil, Colombia y Chile. Los resultados muestran políticas que a mediano y largo plazo no son sostenibles a pesar de los esfuerzos que se han venido haciendo. A corto plazo los datos abiertos dependen de la voluntad política de los Gobiernos.

Hay desconocimiento y rechazo sobre lo que significa implementar datos abiertos y sus beneficios, y falta de cooperación con intermediarios, usuarios y donantes de datos. Por el contrario, los Gobiernos que han tenido resultados positivos se destacan porque han incorporado políticas y acciones orientadas a generar impacto, además de fortalecer la cooperación con los múltiples actores involucrados.

Conclusiones y trabajo futuro

Esta investigación pretende contribuir al uso de los datos abiertos dispuestos por parte de entes públicos y privados, y proporcionar información valiosa para el apoyo de la toma de decisiones, sin embargo, un aporte real dependerá de aspectos técnicos, semánticos y organizacionales.

Desde el punto de vista técnico, es necesario establecer parámetros comunes respecto a la información publicada, la calidad, el formato, las licencias, el procesamiento y la carga en cada sitio. Respecto a la semántica, el reto consiste en tipificar los formatos y establecer un vocabulario de referencia estándar que las Administraciones puedan utilizar para facilitar la vinculación de los datos y ampliar el alcance a soluciones nacionales. Por último, en cuestiones organizativas se requiere un fuerte liderazgo de la Administración pública para identificar cuáles son los datos que crean valor añadido, definir indicadores para medir y evaluar iniciativas y reconocer la importancia de los datos abiertos como motor de generación de beneficios económicos, en la transparencia y en la mejora de la interoperabilidad vertical y horizontal de las Administraciones públicas.

La participación ciudadana real es indispensable; la frustración y la falta de confianza en las instituciones en muchos países son altas, por lo que no es ninguna sorpresa cuando la gente manifiesta su escepticismo respecto a la contribución de la apertura de los datos gubernamentales en la solución de los problemas. Quizá los problemas más difíciles de superar no son técnicos o semánticos, sino más bien aquellos relativos al entendimiento de que la información pertenece a todos y que el empoderamiento de las Administraciones públicas es esencial para compartir su información con otros. En este sentido, la investigación sobre nuevos mecanismos que contribuyan a facilitar la búsqueda, extracción, análisis y uso de datos abiertos es necesaria para aprovechar el potencial del Open Government Data.

En cuanto a la gobernabilidad, el Ministerio TIC es el principal responsable del diseño, la implementación y la promoción de la política de datos abiertos en Colombia, allí se han definido lineamientos que establecen directrices específicas

sobre la gestión y el uso de los datos. Es indispensable atender la hoja de ruta propuesta a partir de documentos como la *Guía de Estándares de Calidad e Interoperabilidad*, cuyo fin es proporcionar orientaciones y buenas prácticas para el desarrollo de estrategias de apertura de datos que cumplan con los principales estándares de calidad e interoperabilidad de datos abiertos.

Igualmente, el *Marco de Arquitectura Empresarial* es el principal instrumento para implementar la arquitectura de tecnologías de la información de Colombia, y habilitar la Política de Gobierno Digital junto al *Marco de Interoperabilidad*, que determina los lineamientos y las recomendaciones para el intercambio eficiente de información entre entidades del Estado.

No menos importantes son la *Guía de anonimización*, cuyo propósito es orientar a las entidades integrantes del Sistema Estadístico Nacional sobre el proceso de anonimización de bases de datos que provienen de registros administrativos y de operaciones estadísticas, y los *Lineamientos técnicos para la producción y gestión de información geográfica*, que contienen un marco general de normas técnicas encaminadas a consolidar bajo una estructura estándar la información geográfica de las entidades del Estado.

De esta manera, a partir del Marco de Gobernanza de Datos Abiertos y de acuerdo con el marco normativo vigente, es posible adelantar una estrategia orientada al proceso de apertura, identificar los actores que formarán parte del ecosistema, usar las herramientas disponibles, diseñar un esquema de incentivos y generar estrategias de monitoreo al valor de los datos abiertos.

Referencias bibliográficas

- Albano, C. (junio de 2013). Open government data: a value chain model proposal. *Proceedings of the 14th Annual International Conference on Digital Government Research*, pp. 285-286. Recuperado de <https://bit.ly/32OJQBO>.
- Alianza Regional por la Libre Expresión e Información (28 de septiembre de 2012). Saber más IV. Acceso a la información y la Alianza de Gobierno Abierto (Open Government Partnership). Recuperado de <https://bit.ly/2EgyxrT>.
- Airinei, D. y Homocianu, D. (enero de 2010). The Mobile Business Intelligence Challenge. *Economic Informatics*, 10(1), 5-12.
- Anderson, J., Lehnardt, J. y Slater, N. (2010). *CouchDB: the definitive guide*. California: O'Reilly Media.
- Aragón, I. (Inv.) (s. f.). Mecanismo de Revisión Independiente: Costa Rica. Informe de Avance 2013-2014. Washington: Open Government Partnership. Recuperado de <https://bit.ly/2F6yUW6>.
- Asanovic, K., Bodik, R., Demmel, J. [...] y Yelick, K. (2009). A view of the parallel computing landscape. *Communications of the ACM*, 5(10), 1-9.
- Assaf, A., Troncy, R. y Senart, A. (2015). HDL - Towards a Harmonized Dataset Model for Open Data Portals. Usewod-Profiles@ ESWC. Recuperado de <https://bit.ly/2ONMX4r>.
- Attard, J., Orlandi, F. y Auer, S. (marzo de 2016). Value Creation on Open Government Data. *International Conference on System Sciences*, pp. 2605-2614. Recuperado de <https://bit.ly/2CUJjDu>.
- Augusto, C., Vega, H. y Pariona, J. (2009). Procesamiento de lenguaje natural. *Revista de Ingeniería de Sistemas e Informática*, 6(2), 45-54.
- Azma, F. y Mostafapour, M. (enero de 2012). Business intelligence as a key strategy for development organizations. *Procedia Technology*, 1, 102-106.

- Beaver, D., Kumar, S., Li, H., Sobel, J. y Vajgel, P. (2010). Finding a Needle in Haystack: Facebook's Photo Storage. *OSDI*, 10, 1-8.
- Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. En Kogan, J., Nicholas, C. y Teboulle, M. (Eds.). *Grouping Multidimensional Data: Recent Advances in Clustering* (pp. 25–71). Berlín: Springer-Verlag Berlin Heidelberg.
- Blakemore, M. y Craglia M. (2006). Access to Public-Sector Information in Europe: Policy, Rights, and Obligations. *The Information Society. An International Journal*, 22(1), 13-24.
- Bose, N. y Liang, P. (1996). *Neural Network Fundamentals with graphs, algorithms and applications*. Nueva York: McGraw-Hill.
- Broder, A. y Mitzenmacher, M. (2011). Network Applications of Bloom Filters: A Survey. *Internet Mathematics*, 1(4), 485-509.
- Carver, A. y Ritacco, M. (2006). *The Business Value of Business Intelligence. A Framework for Measuring the Benefits of Business Intelligence*. California: Business Objects. Recuperado de <https://bit.ly/2Ejg9OQ>.
- Casas, J. (2014). Mecanismo de revisión independiente: el Perú. Informe de avance 2012-13. Recuperado de <https://bit.ly/3f8Wb62>.
- Chaiken, R., Jenkins, B., Larson, P. [...] y Zhou, J. (agosto de 2008). SCOPE: easy and efficient parallel processing of massive data sets. *Proceedings of the VLDB Endowment*, 1(2), 1265-1276.
- Chatterjee, S. y Hadi, A. (2015). *Regression analysis by example* (5.ª ed.). Nueva Jersey: Wiley.
- Chaudhuri, S., Dayal, U. y Narasayya, V. (agosto de 2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.
- Chen, C. y Zhang, C-Y. (agosto de 2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347.
- Chen, M., Mao, S. y Liu, Y. (enero de 2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), 171-209.
- Chen, X., Member, S. y Lin, X. (2014). Big Data Deep Learning: Challenges and Perspectives. *IEEE*, 2, 514-525.
- Chile, Unidad de Modernización y Gobierno Digital (febrero de 2013). Norma Técnica para Publicación de Datos Abiertos en Chile, pp. 1-28. Recuperado de <https://bit.ly/30Gv0uf>.

- Chodorow, K. (2013). *MongoDB: the definitive guide*. California: O'Reilly Media.
- Comisión Económica para América Latina y el Caribe [Cepal] (octubre de 2013). Acceso a la información, participación y justicia en temas ambientales en América Latina y el Caribe. Situación actual, perspectivas y ejemplos de buenas prácticas. Recuperado de <https://bit.ly/2BBm2G8>.
- Comisión Europea (20 de septiembre de 2000). *Commercial Exploitation of Europe's Public Sector Information*. Luxemburgo: Pira International-University of East Anglia y KnowledgeView.
- Comisión Europea (12 de diciembre de 2011). "Open data: An engine for innovation, growth and transparent governance". Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Bruselas: Comisión Europea. Recuperado de <https://bit.ly/2Zel0Ji>.
- Costa Rica, Presidencia de la República (s. f.). Alianza para el Gobierno Abierto. Plan de Acción Costa Rica, 2013-2014. Recuperado de <https://bit.ly/2XPWSLU>.
- Dean, J. y Ghemawat, S. (enero de 2008). MapReduce: simplified data processing on large clusters, *Communications of the ACM*, 51(1), 107-113.
- DeCandia, G., Hastorun, D., Jampani, M. [...] y Vogels, W. (octubre de 2007). Dynamo: amazon's highly available key-value store. *ACM SIGOPS Operating Systems Review*, 41(6), 205-220.
- Dekkers, M., Polman, F., Te Velde, R. y De Vries, M. (junio de 2006). Measuring European Public Sector Information Resources. Recuperado de <https://bit.ly/32AQc7p>.
- Demirkan, H. y Delen, D. (abril de 2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412-421.
- Department for Business Innovation and Skills [BIS] (mayo de 2013). Market Assessment of Public Sector Information. Londres: Deloitte. Recuperado de <https://bit.ly/3eMOPVV>.
- Diez, E. y Millán, R. (2011). *Open data y RISIP generando valor social y económico. Transparencia e innovación en la eAdministración*. Santiago de Compostela (España): Colexio Profesional de Enxeñaría en Informática de Galicia-Xunta de Galicia.
- DotEcon (diciembre de 2006). *The commercial use of public information (CUPI)*. Reino Unido: Office of Fair Trading.

- Duda, J. (2012). Business intelligence and NoSQL Databases. *Information Systems in Management*, 1(1), 25-37.
- Elgendy, N. y Elragal, A. (2014). Big Data Analytics: A Literature Review Paper. En Perner, P. (Ed.). *Advances in Data Mining. Applications and Theoretical Aspects: 14th Industrial Conference, ICDM 2014. St. Petersburg, Russia, July 16-20, 2014. Proceedings* (pp. 214-227). Cham: Springer International Publishing.
- Estados Unidos de América (2019). Data.Gov. Recuperado de <http://www.data.gov/>.
- Fan, J., Han, F. y Liu, H. (junio de 2014). Challenges of Big Data analysis. *National Science Review*, 1(2), 293-314.
- Feinman, V. (Ed.) (2006). *Guía para usar el Decreto 1172-03 de Acceso a la Información Pública*. Buenos Aires: Asociación por los Derechos Civiles [ADC].
- Ferrer-Sapena, A. y Peset, F. (2011). Acceso a los datos públicos y su reutilización: open data y open government. *Profesional de la Información*, 20(3) 260-269.
- Fotheringham, S. y Rogerson, P. (2013). *Spatial analysis and GIS*. Florida: CRC Press.
- Fumega, S. y Scrollini, F. (enero-junio de 2013). Primeros aportes para diseños de políticas de datos abiertos en América Latina. *Derecho Comparado de la Información*, 21, 3-37.
- Fundación CiGob (enero de 2013). Gobierno abierto: experiencias de implementación. Los casos de EE.UU., Reino Unido, País Vasco y Chile. Recuperado de <https://bit.ly/3f8SxcD>.
- Fundación COTEC para la Innovación Tecnológica (mayo de 2011). Reutilización de la información del sector público. Documento 28. Recuperado de <https://bit.ly/39rinaC>.
- Gartner (s. f.). Glossary. Recuperado de <https://www.gartner.com/en/information-technology/glossary>.
- Giménez, V. (2014). La reutilización de la información en el sector público: aplicación en la Unión Europea. *e-Ciencias la Información*, 4(1), 0-11.
- George, L. (2011). *HBase: The Definitive Guide: Random Access to Your Planet-Size Data*. California: O'Reilly Media.
- Gobierno de Belice (diciembre de 2000). Freedom of Information Act - Cap. 13. Recuperado de <https://bit.ly/3314sGP>.
- Gobierno de Canadá (2016). Canada's Action Plan on Open Government 2014-16. Recuperado de <https://bit.ly/3jOmRfU>.

- Gobierno de Chile (2013). Plan de Acción del Gobierno de Chile 2013-2014. Open Government Partnership. Recuperado de <https://bit.ly/2Da6c5Y>.
- Gobierno de Jamaica (2002). Access to Information Act. No. 21-2002. Recuperado de <https://bit.ly/2DSME6s>.
- Gobierno del Perú (s. f.). Plan de Acción Nacional de Gobierno Abierto del Perú, 2017-2019. Alianza para el Gobierno Abierto. Recuperado de <https://bit.ly/31VF4QX>.
- Gobierno de la República de Panamá, Autoridad Nacional de Transparencia y Acceso a la Información [Antai] (2015). Segundo Plan de Acción 2015-2017. Recuperado de <https://bit.ly/2DBkJIu>.
- Gobierno de los Estados Unidos de América (2017). USASPENDING.gov. Recuperado de <https://www.usaspending.gov/#/>.
- Gomes, Á. y Soares, D. (2014). Open government data initiatives in Europe: Northern versus southern countries analysis. *ICEGOV '14: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, pp. 342-350. Recuperado de <https://bit.ly/2Wn2GMn>.
- Groves, P., Kayyali, B., Knott, D. y Van Kuiken, S. (enero de 2013). The “big data” revolution in healthcare. Accelerating value and innovation. Center for US Health System Reform. Business Technology Office. Recuperado de <https://bit.ly/39uX08o>.
- Gruen, N., Houghton, J. y Tooth, R. (junio de 2014). Open for Business: How Open Data Can Help Achieve the G20 Growth. Lateral Economics-Omidyar Network. Recuperado de <https://bit.ly/3g66k50>.
- Gurin, J. (2013). *Open data now: the secret to hot startups, smart investing, savvy marketing, and fast innovation*. Nueva York: McGraw Hill.
- Guyana. Parlamento de la República (2011). Access to Information Act. Recuperado de <https://bit.ly/3ix1bnq>.
- Han, J., Pei, J. y Kamber, M. (2011). *Data mining: concepts and techniques*. Amsterdam: Elsevier.
- Herala, A., Vanhala, E., Porras, J. y Krri, T. (2016). Experiences About Opening Data In Private Sector: A Systematic Literature Review. *SAI Computing Conference 2016*, pp. 715-724. Recuperado de <https://bit.ly/3eN07JN>.

- Heerschap, N., Ortega, S., Priem, A. y Offermans, M. (2014). Innovation of tourism statistics through the use of new big data sources. *Statistics Netherlands*. Recuperado de <https://bit.ly/39FbXEU>.
- Hermosillo, G., Seinturier, L. y Duchien, L. (julio de 2010). Using Complex Event Processing for Dynamic Business Process Adaptation. *Services Computing (SCC), 2010 IEEE International Conference*, pp. 466-473. Recuperado de <https://bit.ly/30qCpxI>.
- Hočevar, B. y Jaklič, J. (2008). Assessing benefits of business intelligence systems. A case study. *Management*, 15(1), 87-119.
- Holt, J. y Chung, S. (2002). Mining association rules using inverted hashing and pruning. *Information Processing Letters*, 83(4), 211-220.
- Huijboom, N. y Van den Broek, T. (marzo-abril de 2011). Open data: an international comparison of strategies. *European Journal of ePractice*, 1, 1-13.
- Ilacqua, C., Cronstrom, H. y Richardson, J. (2015). *Learning Qlik Sense®: The Official Guide*. Birmingham (Reino Unido): Packt Publishing.
- Işık, Ö., Jones, M. y Sidorova, A. (enero de 2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*, 50(1), 13-23.
- Janssen, M., Charalabidis, Y. y Zuiderwijk, A. (septiembre de 2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems in Management*, 29(4), 258-268.
- Jetzek, T., Avital, M. y Bjorn-Andersen, N. (15 de diciembre de 2012). "The Value of Open Government Data: A Strategic Analysis Framework". Documento del taller previo al Open Data and Open Innovation in eGovernment. Orlando, Estados Unidos. Recuperado de <https://bit.ly/3fcqEkF>.
- Junior, M., Santos, M., Vidal, M. y De Carvalho, P. (enero de 2012). Overcoming the blame game to learn from major accidents: A systemic analysis of an Anhydrous Ammonia leakage accident. *Journal of Loss Prevention in the Process Industries*, 25(1), 33-39.
- Kaushik, R. y Nahrstedt, K. (2012). T*: A Data-Centric Cooling Energy Costs Reduction Approach for Big Data Analytics Cloud. *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, 1-11. Recuperado de <https://ieeexplore.ieee.org/document/6468476>.
- Khaleghi, B., Khamis, A., Karray, F. y Razavi, S. (enero de 2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1), 28-44.

- Khan, N., Yaqoob, I., Targio, I. [...] y Gani, A. (julio de 2014). Big Data: Survey, Technologies, Opportunities, and Challenges. *The Scientific World Journal*. Recuperado de <https://bit.ly/3jJrvvV>.
- Kim, B., Trimi, S. y Chung, J. (marzo de 2014). Big-Data Applications in the Government Sector. *Communications of the ACM*, 57(3), 78-85. Recuperado de <https://cacm.acm.org/magazines/2014/3/172509-big-data-applications-in-the-government-sector/abstract>.
- Lakshman, A. y Malik, P. (enero de 2009). Cassandra: structured storage system on a p2p network. *Proceedings of the 28th ACM Symposium on Principles of Distributed Computing*. Recuperado de <https://dl.acm.org/doi/10.1145/1582716.1582722>.
- Lasanta, M. (2011). *Reutilización de datos del sector público*. Madrid: Universidad Complutense de Madrid.
- Lindman, J., Kinnari, T. y Rossi, M. (enero de 2015). Business Roles in the Emerging Open-Data Ecosystem. *IEEE Software*, 33(5), 1-1.
- McAfee, A. y Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review*, 90(10), 61-67.
- McBride, K., Aavik, G., Toots, M., Kalvet, T. y Krimmer, R. (enero de 2019). How does open government data driven co-creation occur? Six factors and a “perfect storm”; insights from Chicago’s food inspection forecasting model. *Government Information Quarterly*, 36(1), 88-97.
- McLaren, R. y Waters, R. (agosto de 2011). Governing Location Information in the UK. *The Cartographic Journal*, 48(3), 172-178.
- Malewicz, G., Austern, M., Bik, A. [...] y Czajkowski G. (2010). Pregel: a system for large-scale graph processing. *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, pp. 135-146.
- Mansmann, S., Neumuth, T. y Scholl, M. (septiembre de 2007). OLAP Technology for Business Process Intelligence: Challenges and Solutions. International Conference on Data Warehousing and Knowledge Discovery, pp. 111-122. Recuperado de <https://bit.ly/2WFpJBV>.
- Manyika, J., Chui, M., Brown, B. [...] y Hung, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Recuperado de: <https://mck.co/2PFKLN7>.
- Manyika, J., Chui, M., Farrell, D., Van Kuiken, S., Groves, P. y Almasi, E. (1 de octubre de 2013). Open data: Unlocking innovation and performance with liquid information. McKinsey Digital. Recuperado de <https://mck.co/3exzOHd>.

- Matheus, R. y Vaz, J. (octubre de 2012). New Perspectives for Electronic Government in Brazil: The Adoption of Open Government Data in National and Subnational Governments of Brazil. *Conference: Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, pp. 22-29. Recuperado de <https://bit.ly/3hu0Y3B>.
- México, Congreso Nacional de los Estados Unidos Mexicanos (30 de abril de 2002). Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental. Recuperado de <https://bit.ly/39BWcyc>.
- México, Secretaría de la Función Pública (30 de agosto de 2011). Acuerdo que establece el Esquema de Interoperabilidad y de Datos Abiertos de la Administración Pública Federal. Recuperado de <https://bit.ly/3jMZu6F>.
- México, Congreso Nacional de los Estados Unidos Mexicanos (19 de diciembre de 2011). Reglamento de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares. Recuperado de <https://bit.ly/30TMyDl>.
- México, Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales [INAI] (2015). Visión, misión y objetivos estratégicos. Recuperado de <https://bit.ly/310EFMC>.
- Michener, R. (mayo de 2010). The Surrender of Secrecy: Explaining the Emergence of Strong Access to Information Laws In Latin America (tesis de doctorado). Austin: Universidad de Texas.
- Microsoft Inc. (2016). Power BI. Recuperado de <https://powerbi.microsoft.com/es-es/why-power-bi/>.
- Microsoft (2016). ¿Qué es Power BI? Recuperado de <https://bit.ly/39kVRjB>.
- Mohamadina, A., Ghazali, M., Ibrahim, M. y Harbawi, M. (noviembre de 2012). Business Intelligence: Concepts, Issues and Current Systems. *International Conference on Advanced Computer Science Applications and Technologies (ACSAT)*, 234-237.
- Moretti, C., Bulosan, J., Thain, D. y Flynn, P. (2008). All-pairs: An abstraction for data-intensive cloud computing. *Proceedings of the 22nd IEEE International Parallel and Distributed Processing Symposium, Program and CD-ROM*, pp. 1-11.
- Moss, L. y Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Boston: Addison-Wesley.
- Murillo, C., Ardao, A., Giraldo, M. y Garay, G. (2016). Estudio de las herramientas líderes de inteligencia de negocios basado en el cuadrante mágico de Gartner.

- XVI Convención y Feria Internacional Informática 2016. Recuperado de <http://www.informaticahabana.cu/es/node/740>.
- Murty, J. (2008). *Programming amazon web services: S3, EC2, SQS, FPS, and SimpleDB*. California: O'Reilly Media.
- Naser, A. y Concha, G. (marzo de 2012). *Datos abiertos: un nuevo desafío para los gobiernos de la región*. Serie Gestión Pública. Comisión Económica para América Latina y el Caribe-Naciones Unidas. Recuperado de <https://bit.ly/2PI5tfd>.
- Nava, F. (2015). *Procesamiento de series de tiempo*. Ciudad de México: Fondo de Cultura Económica.
- Observatorio Regional de Planificación para el Desarrollo de América Latina y el Caribe (2013). *Primer Plan de Acción Nacional de Gobierno Abierto de Colombia (2011-2013)*. Recuperado de <https://bit.ly/2QlQrMr>.
- Observatorio Regional de Planificación para el Desarrollo de América Latina y el Caribe. (2016). *Segundo Plan de Acción Nacional de Gobierno Abierto de República Dominicana (2014-2016)*. Recuperado de <https://observatorioplanificacion.cepal.org/es/planes/segundo-plan-de-accion-nacional-de-gobierno-abierto-de-republica-dominicana-2014-2016>.
- Open Knowledge Foundation (2012). *Open Data Handbook Documentation*. Recuperado de <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>.
- Open Knowledge Foundation (2012). *Manual de los datos abiertos*. Recuperado de <https://bit.ly/39nzKZV>.
- Ortoll, E., Canals, A. y Cobarsí, J. (2014). Principales parámetros para el estudio de la colaboración científica en big science. *Revista Española de Documentación Científica*, 37(4), 1-11.
- Paluszek, M. y Thomas, S. (2017). An Overview of Machine Learning. *MATLAB Machine Learning*, Berkeley: Apress, pp. 3-15.
- Parenteau, J., Sallam, R., Howson, C., Tapadinhas, J., Schlegel, K. y Oestreich, T. (4 de febrero de 2020). Magic Quadrant for Business Intelligence and Analytics Platforms. Gartner. Recuperado de <https://gtmr.it/30DtZTY>.
- Parlamento del Uruguay (7 de octubre de 2008). Ley 18381. Derecho de Acceso a la Información Pública. Recuperado de <https://bit.ly/32YtuGN>.
- Peset, F., Ferrer-Sapena, A. y Subirats-Coll, I. (marzo de 2011). Open Data y Linked Open Data: su impacto en el área de bibliotecas y documentación. *El Profesional de la Información*, 20(2), 165-174.

- Portal de Datos Abiertos de la UE (2019). Datos. Recuperado de <https://data.europa.eu/euodp/es/data/>.
- Presidencia de la Nación Argentina (2003). Reglamento General de Audiencias Públicas para el Poder Ejecutivo Nacional. Decreto 1172/2003. Recuperado de <https://bit.ly/30TrwF1>.
- Provost, F. y Fawcett, T. (2013). Data science and its relationship data-driven decision making. *Big Data*, 1(1), 51-59.
- Qlik (2017). Why Qlik® is different. Recuperado de <https://www.qlik.com/us/products/why-qlik-is-different>.
- Raghupathi, W. y Raghupathi, V. (febrero 7 de 2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 1-10.
- Ranjan, J. (enero de 2009). Business Intelligence: Concepts, components, techniques and benefits. *Journal of Theoretical and Applied Information Technology*, 9, 60-70.
- Ranjan, J. y Bhatnagar, V. (marzo de 2011). Role of knowledge management and analytical CRM in business: data mining based framework. *The Learning Organization*, 18(2), 131-148.
- Reino Unido (2019). GOV.UK. Recuperado de <https://www.gov.uk/>.
- República Argentina, Ministerio de Modernización (2015). II Plan de Acción Nacional de Gobierno Abierto de la República Argentina, 2015-2017. Recuperado de <https://bit.ly/3g95vIu>.
- República de Bolivia (17 de mayo de 2005). Decreto Supremo 28168. Transparencia en la Gestión Pública del Poder Ejecutivo. Recuperado de <https://bit.ly/304zNXq>.
- República de Brasil. Presidencia de la República (18 de noviembre de 2011). Ley 12.527. Regula el acceso a la información prevista en el artículo XXXIII del art. 5, en el ítem II del § 3 del art. 37 y en el § 2 del art. 216 de la Constitución Federal; modifica la Ley N° 8.112, de 11 de diciembre de 1990; deroga la Ley N° 11.111, de 5 de mayo de 2005, y las disposiciones de la Ley N° 8.159, de 8 de enero de 1991; y hace otros arreglos. Recuperado de <https://bit.ly/3g7KAW9>.
- República de Costa Rica, Asamblea Nacional Constituyente (7 de noviembre de 1949). Constitución Política de Costa Rica. Recuperado de <https://bit.ly/3g6qATW>.
- República de Costa Rica, Asamblea Legislativa (2004). Ley 8432 Contra la Corrupción y el Enriquecimiento Ilícito en la Función Pública. Recuperado de <https://bit.ly/2DgTX7n>.

- República Dominicana, Congreso Nacional (28 de julio de 2004). Ley General de Acceso a la Información Pública. Ley 200-04. Recuperado de <https://bit.ly/332IEeb>.
- República de El Salvador, Asamblea Legislativa (29 de abril de 2011). Ley de Regulación de los Servicios de Información Sobre el Historial de Crédito de las Personas. Recuperado de <https://bit.ly/30RWLQQ>.
- República de Guatemala, Comisión Presidencial de Transparencia y Gobierno Electrónico (2014). Segundo Plan de Acción Nacional de Gobierno Abierto. Recuperado de <https://bit.ly/3f82K91>.
- República de Honduras, Congreso Nacional (17 de julio de 2007). Ley de Transparencia y acceso a la información pública. Decreto 170/2006. Recuperado de <https://bit.ly/304ErEJ>.
- República de Honduras, Gobierno de la República (2014). II Plan de Acción de Gobierno Abierto Honduras 2014-2016. Recuperado de <https://bit.ly/3f9hLHR>.
- República de Nicaragua, Presidencia de la República (16 de mayo de 2007). Ley de Acceso a la Información Pública. Ley 621 del 16 de mayo del 2007. Recuperado de <https://bit.ly/39zu8M5>.
- República de Nicaragua, Asamblea Nacional (marzo de 2012). Ley de protección de datos personales. Ley No. 787 del 21 de marzo del 2012. Recuperado de <https://bit.ly/30bD73b>.
- República de Panamá, Asamblea Legislativa (22 de enero de 2002). Ley de Transparencia de la Gestión Pública. Ley 6 de 2002. Recuperado de <https://bit.ly/2X4xhhO>.
- República de Paraguay, Congreso de la Nación (18 de septiembre de 2014). Ley 5282 del Libre Acceso Ciudadano a la Información Pública y Transparencia Gubernamental. Recuperado de <https://bit.ly/2P5eHSh>.
- República de Paraguay (2018). Cuarto Plan de Acción Gobierno Abierto 2018-2020. Recuperado de <https://bit.ly/3faNqIK>.
- República Oriental del Uruguay, Presidencia de la República (2014). Segundo Plan de Acción Nacional de Gobierno Abierto 2014-2016. Recuperado de <https://bit.ly/3jQc0Cf>.
- Rouibah, K. y Ould-ali, S. (junio de 2002). PUZZLE: a concept and prototype for linking business intelligence to business strategy. *The Journal of Strategic Information Systems*, 11(2), 133-152.

- Sard, M., Budiu, M., Yu, Y., Birrell, A. y Fetterly, D. (2007). Dryad: distributed data-parallel programs from sequential building blocks. *ACM SIGOPS Operating Systems Review*, 41(3), 59-72.
- Schonlau, M. (2002). The clustergram: A graph for visualizing hierarchical and non-hierarchical cluster analyses. *The Stata Journal*, 2(4), 391-402.
- Stephenson, P., Bowles-Newark, N., Regan, E. y Stanwell-Smith, D. (enero de 2017). Unblocking the flow of biodiversity data for decision-making in Africa. *Biological Conservation*, 213, 335-340.
- Stott, A. (2014). Open data for economic growth. Whashington: Banco Mundial. Recuperado de <https://bit.ly/3egWThl>.
- Strathern, M. (junio de 2000). The tyranny of transparency. *British Educational Research Journal*, 26(3), 309-321.
- Stuerner, M. y Dapp, M. (2016). Measuring the promise of open data: Development of the Impact Monitoring Framework. *Conference for E-Democracy and Open Government*, pp. 197-203. Recuperado de <https://bit.ly/2CRzia5>.
- Tableau (2017). Tableau aprendizaje. Recuperado de <https://www.tableau.com/es-es/learn>.
- Unbounce (2015). *The Ultimate Guide to A / B Testing*. Recuperado de <https://get.unbounce.com/the-ultimate-guide-to-ab-testing/>.
- Unión Europea (15 de noviembre de 2017). Open Data Maturity in Europe 2017. Recuperado de <https://bit.ly/2CvLIEG>.
- Valencia, P. (1997). Optimización mediante algoritmos genéticos. *Anales del Instituto de Ingenieros de Chile*, 83-92.
- Vercellis, C. (2011). *Business intelligence. Data mining and optimization for decision making*. Nueva Jersey: John Wiley & Sons.
- Vickery, G. (2011). Review of recent studies on PSI re-use and related market developments. París: Comisión Europea. Recuperado de <https://bit.ly/2DlnXiz>.
- Wang, X., Fang, Y., Liu, Y. y Horn, B. (2018). A Survey on the Status of Open Data and Its Future. *4th International Conference on Universal Village, Boston*, pp. 1-4, doi: 10.1109/UVV.2018.8642128.
- Watson, H. (2009). Tutorial: Business Intelligence-Past, Present and Future. *Communications of the Association for Information Systems*, 25(39). Recuperado de <https://bit.ly/30Pr6iM>.

- World Wide Web Foundation (2015). *Open Data Barometer Global Report Third Edition*. Recuperado de <https://bit.ly/2WFB8Sl>.
- World Wide Web Foundation (2018). Open Data Barometer. Recuperado de <https://opendatabarometer.org/4thedition/report/>.
- Xu, L., Jiang, Ch., Wang, J., Yuan, J. y Ren, Y. (2014). Information Security in Big Data: Privacy and Data Mining. *IEEE Access*, 2, 1149-1176.
- Yannoukakou, A. y Araka, I. (25 de agosto de 2014). Access to Government Information: Right to Information and Open Government Data Synergy. *Procedia. Social and Behavioral Sciences*, 147, 332-340.
- Yiu, C. (25 de julio de 2012). The Big Data Opportunity. *Policy Exchange*. Recuperado de <https://policyexchange.org.uk/the-big-data-opportunity/>.
- Zuiderwijk, A., Janssen, M. y Choenni, S. (2013). Socio-technical Impediments of Open Data. *Electronic Journal of e-Government*, 10(2), 156-172.
- Zurada, J. y Karwowski, W. (junio de 2011). Knowledge discovery through experiential learning from business and other contemporary data sources: A review and reappraisal. *Information Systems in Management*, 28(3), 258-274.

Autores

Giovanny Mauricio Tarazona Bermúdez

Ingeniero Industrial con posgrados en Ingeniería de Software y Proyectos Informáticos de la Universidad Distrital Francisco José de Caldas. Magister en Diseño y Gestión de Proyectos Tecnológicos de la Universidad Internacional de la Rioja. Doctor en Sistemas Informáticos para Internet de la Universidad de Oviedo. Profesor universitario desde 1999, empresario y coinvestigador del grupo de investigación GICOECOL. Actualmente es director del Centro de Investigaciones y Desarrollo Científico de la Universidad Distrital Francisco José de Caldas.

Luz Andrea Rodríguez Rojas

Doctora e investigadora universitaria. Coautora de libros, artículos científicos, ponencias nacionales e internacionales y proyectos de investigación. Directora y codirectora de proyectos de investigación de pregrado y especialización, tesis de maestría y doctorado. Par evaluador. Con experiencia en publicaciones de artículos de alto impacto en Journal Citation Reports (JCR) y SJR Indicator (Scimago Journal and Country Rank). Coinvestigadora del grupo de investigación GICOECOL. Actualmente es docente de planta de la Universidad Distrital Francisco José de Caldas.

Este libro se
terminó de imprimir
en octubre de 2020
en la Editorial UD
Bogotá, Colombia