



# Big Data: una visión general

## RESUMEN

- El volumen de información electrónica en forma de datos que se genera, recopila y almacena está creciendo exponencialmente. El análisis de toda esta información representa grandes oportunidades, pero también plantea retos que deben superarse.
- La acumulación y manejo computacional de estos grandes volúmenes de información se ha denominado Big Data, aunque no hay una definición universalmente aceptada de este término.
- Para obtener conocimiento y tomar decisiones usando Big Data, se requiere aplicar herramientas sofisticadas de análisis estadístico, así como técnicas innovadoras de procesamiento computacional, entre las que se encuentra la conocida como “Big Data Analytics”.
- Esta técnica se está aplicando en ámbitos tan diversos como los negocios, el transporte, los servicios públicos, el deporte, la atención médica y la investigación.
- Big Data Analytics permite analizar grandes volúmenes de información, para personalizar productos y servicios y optimizar su eficiencia, lo que se traduce en una mayor competitividad para las organizaciones y mayor desarrollo económico para las regiones que adoptan estas tecnologías.
- Las principales preocupaciones que presenta Big Data son: fallas en la seguridad de la información, violación de la privacidad, discriminación y pérdida de empleos debido a la creciente automatización de los sistemas.
- En México, el 5 de julio de 2016, se cumplieron 6 años de la publicación de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP).

Imágen: Erwin Rodríguez / FCCyT.

El término *Big Data*, macro-datos o datos masivos en español, se refiere a la acumulación y procesamiento computacional de información o “datos” en volúmenes tan grandes o de tal complejidad que utilizarlos e interpretarlos representa un reto. Esta nota examina su uso y regulación, así como las preocupaciones públicas que genera.

## Antecedentes

El volumen de información en formato electrónico que actualmente se recopila, almacena, copia y analiza a partir de diversas fuentes no tiene precedente. Se calcula que en 2013 había alrededor de 4.4 trillones ( $4.4 \times 10^{18}$ ) de gigabytes (GB) de datos almacenados a nivel global<sup>1</sup>, equivalentes a aproximadamente 120 películas en formato DVD por persona en el planeta (1 DVD almacena poco menos de 5 GB de video digital). Se estima que el volumen total de información a nivel mundial irá aumentando a una tasa de alrededor de 40% anual durante la próxima década. Este acelerado incremento de la disponibilidad y complejidad de información ha dado lugar al término “*Big Data*”, sin embargo no hay una definición universalmente aceptada (ver recuadro 1).<sup>2</sup>

En términos generales, *Big Data* se refiere a datos cuya recopilación, procesamiento, análisis o interpretación constituye un reto que con frecuencia se supera con el uso de técnicas innovadoras. Uno de los procesos para obtener nuevos conocimientos a partir de *Big Data* se conoce como *Big Data Analytics*.<sup>3</sup> Esta técnica surgió en gran medida como respuesta a los desafíos que presentan los volúmenes enormes de datos (ver recuadro 2).

No obstante, hay que mencionar las dudas que se tienen sobre hasta qué punto estas oportunidades se concretarán.

### Recuadro 1. ¿Qué es el Big Data?

Las principales características de *Big Data* son:<sup>2</sup>

- **Gran tamaño.** Acumulación y procesamiento de enormes volúmenes de información que pueden ir de los terabytes (1,000 GB) hasta los petabytes (1 millón de GB). El tamaño de los datos, por ejemplo, de las listas del buscador Google que se usan para recomendar páginas web, es sumamente grande para una hoja de cálculo o para una computadora común y los datos suelen contener muchas variables.

**Recuadro 1 (Continúa). ¿Qué es el Big Data?**

- **Gran velocidad.** Para que la información obtenida sea útil, los datos se deben recopilar, procesar y analizar rápidamente, a menudo a la misma velocidad a la que se recopilan en “tiempo real”. Por ejemplo, las aplicaciones que analizan datos del mercado de valores o de la vigilancia satelital de nuestros mares o de navegación y tráfico vía el sistema de geo-posicionamiento global (GPS).
- **Gran variedad.** Los datos se recopilan a partir de diversas fuentes y en formatos muy distintos. Por ejemplo, algunos registros médicos contienen no sólo los diagnósticos y recetas del médico de cabecera, sino también los resultados de estudios de laboratorio e imágenes de tomografías o de rayos X.

Entre otras características de *Big Data* para tomar en cuenta están la veracidad (si los datos son poco confiables o precisos) y la variabilidad (datos con cualidades que cambian, por ejemplo, a lo largo del tiempo). A partir de *Big Data* se puede trabajar con datos recientemente recopilados o usar de manera distinta los datos ya existentes, aplicando otros métodos de análisis o vinculando conjuntos de datos para proporcionar un panorama más amplio y que puede hacer surgir preguntas antes impensables.

**Recuadro 2. Algunas oportunidades que ofrece Big Data<sup>22</sup>**

- **Comercio electrónico e inteligencia de mercado.** Compañías como Amazon, Netflix y eBay, tienen plataformas web que almacenan grandes volúmenes de información de sus clientes; con base en ellas se puede recomendar preferencias de compra o de elección. Además, Google, Facebook y Amazon monitorean información del comportamiento de los usuarios de Internet y detectan hábitos y necesidades para diseñar productos que los satisfagan.
- **Gobierno electrónico y política.** Hay plataformas que concentran información y documentos de diversas dependencias de gobierno, agilizando y facilitando trámites y procesos burocráticos. El manejo de grandes bases de datos compartidas entre diversas dependencias evita que se duplique la información o los errores en los datos. También, al monitorear y analizar los datos en redes sociales, es posible detectar tendencias en la percepción de la población de los actores políticos e identificar errores o aciertos.
- **Ciencia y tecnología.** Se están creando bibliotecas virtuales gigantescas donde investigadores y tecnólogos almacenan desde libros, artículos y reportes técnicos, hasta material didáctico multimedia accesible para todo aquel con conexión a Internet. Esto da la posibilidad, además, de establecer redes de colaboración entre investigadores con intereses de investigación similares.
- **Salud digital y bienestar.** Se pretende generar bases de datos con toda la historia clínica de los pacientes, incluyendo estudios de laboratorio, radiografías, recetas, médicos, etc. Serían bases de datos sumamente grandes que permitirían alcanzar un mayor conocimiento de muchas enfermedades y posiblemente mejorar la atención médica a la población.
- **Seguridad pública y prevención.** La información electrónica que manejan bancos, empresas y gobierno es sumamente sensible e importante. Procesar y analizar la información concerniente a la seguridad de los datos se está volviendo un problema cada vez más complejo. Se necesita procesar información de muchas fuentes, en distintos formatos y en volúmenes cada vez más grandes. Por lo tanto, la investigación y desarrollo en esta área presenta muchas oportunidades y retos.

Hay quien sostiene que *Big Data* no es nuevo, sino que bien se trata de fronteras que se han desplazado gracias a las mejoras en el almacenamiento y procesamiento computacional de datos.<sup>4</sup>

**Impulso de Big Data**

El crecimiento vertiginoso de la adquisición, producción y uso de datos se debe a factores tecnológicos, sociales y económicos. Los factores tecnológicos incluyen la creación de nuevas fuentes de información, como los teléfonos inteligentes, y el aumento en la capacidad para almacenar y analizar datos (ver recuadro 3). Entre los factores sociales que impulsan *Big Data* sobresale la adopción generalizada de nuevas formas de comunicación a través de las redes sociales (como Facebook, YouTube y Twitter, entre otros)

Las iniciativas de datos abiertos en el sector público son otro factor que impulsa el *Big Data*. Éstas incluyen desde compartir recursos, como software y datos de investigación en la academia, hasta el aumento en el uso de datos recabados por organismos gubernamentales para mejorar la rendición de cuentas y generar beneficios económicos.<sup>6</sup>

Según la consultora IDC (Corporación Internacional de Datos) México, el valor del *Big Data Analytics* en 2014 para la economía mexicana fue de 190 millones de dólares, lo que representó 28.7% del total en América Latina (México y Brasil son los países de la región con mayor ritmo de crecimiento).<sup>7,8</sup> Entre los beneficios del manejo y análisis de *Big*

**Recuadro 3. Factores tecnológicos que impulsan el crecimiento del Big Data**

Nuevas fuentes de datos:

- Digitalización de los procesos y servicios existentes; por ejemplo, banca en línea, e-mail y registros médicos.
- Generación automática de datos, por ejemplo, muchas plataformas web guardan el número de visitas a sus páginas, el tiempo que los usuarios navegan en una determinada página, los elementos más consultados, etc.
- Reducción del costo y tamaño de los sensores en que se usan en autopistas, aviones, accesos a medios de transporte y en muchos otros lugares.
- Producción de nuevos dispositivos que recopilan y transmiten datos; por ejemplo, información de localización vía GPS obtenida de teléfonos inteligentes.

Entre las capacidades de cómputo mejoradas que impulsan *Big Data* destacan:

- Mejor almacenamiento de datos en grandes volúmenes y por un costo menor.
- Mayor poder computacional para realizar cálculos complejos rápidamente.
- La computación en la “nube” (acceso remoto a recursos de cómputo compartido a través de un dispositivo conectado a una red) posibilita el acceso más económico a poder computacional, software y almacenamiento de datos, así como a otros servicios.
- Avances recientes en técnicas estadísticas y computacionales que pueden ser usadas para analizar *Big Data*.
- Desarrollo de nuevas herramientas, como *Apache Hadoop* (que permite procesar grandes volúmenes de datos mediante el uso de muchas computadoras al mismo tiempo) y extensión del software existente, como Excel de Microsoft.

*Data* en el sector público se encuentran los siguientes: datos compartidos entre departamentos para disminuir la duplicación de la información; recopilación y análisis de datos más completos de la administración pública para mejorar el desempeño; identificación automática de errores potenciales y de fraude en los formatos para devolución de impuestos; y el uso de información sobre usuarios para adaptar y orientar servicios y productos.<sup>4</sup>

## Uso de Big Data

Hacer uso de cualquier clase de datos requiere su recopilación, procesamiento y análisis, para después interpretar los resultados. Para hacer frente a los retos que *Big Data* plantea se necesitan habilidades especiales para recopilar, procesar, analizar e interpretar la información de los datos.

### Recopilación

Los conjuntos de datos de *Big Data* se pueden adquirir en un sinnúmero de formatos a partir de un vasto y creciente número de fuentes. Éstas pueden incluir imágenes, grabaciones de sonido, tendencias de búsquedas y clics de los usuarios para medir la actividad en Internet, así como datos generados por simulaciones de computadoras (como las que se emplean para el pronóstico del clima). Los metadatos —datos que describen otros datos— son fundamentales para el manejo de datos. Un e-mail por ejemplo, genera automáticamente metadatos que contienen la dirección del emisor y del receptor, así como la fecha y hora en que el mensaje se envió. Esta información sirve para administrar el almacenamiento de los archivos de e-mail. La sola producción de metadatos para el *Big Data* representa un reto.

### Procesamiento

Los datos deben someterse a numerosos procesos para mejorar la calidad y la facilidad de su uso antes del análisis, entre los que se incluyen:

- **Extracción:** obtener la información requerida de los datos iniciales y expresarla en una forma estructurada.
- **Limpieza:** detectar y después corregir o remover registros mal capturados o inexactos.
- **Estandarización:** darle a los datos el formato adecuado para ayudar a la interoperabilidad.
- **Vinculación:** vincular registros de diferentes fuentes.

### Análisis

El *Big Data Analytics* se usa para obtener ciertos conocimientos o información a partir de datos. Por lo general esto implica aplicar un algoritmo (una sucesión de operaciones) a los datos para encontrar patrones que se pueden usar para hacer predicciones o proyecciones. Esto engloba además varias técnicas interrelacionadas, por ejemplo.

- La minería de datos identifica patrones examinando cuidadosamente los datos. Puede aplicarse a las secuencias de clics de los usuarios en una página web; esto permite a los diseñadores de estas páginas cono-

cer la manera en que los usuarios interactúan con ellas.

- El aprendizaje automático (machine learning) describe sistemas que aprenden de los datos. Por ejemplo, un sistema que compara documentos en dos idiomas diferentes puede inferir reglas de traducción y así la corrección humana de cualquier error posibilita que el sistema aprenda cómo mejorar el software.
- La simulación y modelación se pueden usar para describir el comportamiento de sistemas complejos. Por ejemplo, una simulación de la compra y venta de acciones puede ayudar a evaluar la eficacia de medidas orientadas a reducir el uso de información privilegiada para la compra o venta de acciones.

### Interpretación

Para que los resultados del análisis sean útiles, necesitan ser interpretados y comunicados. La interpretación del análisis de *Big Data* requiere tomar en cuenta el contexto, esto incluye la manera en la que los datos son recolectados, su calidad y cualquier suposición que se haya hecho. Son varios los motivos por los cuales hay que tener cuidado con la interpretación:

- A pesar de su gran tamaño, un conjunto de datos puede contener todavía sesgos y anomalías, o excluir conductas no capturadas en los datos.
- Puede haber limitaciones en la utilidad del *Big Data Analytics*; es posible identificar correlaciones (patrones consistentes entre variables), pero no necesariamente la causa u origen. Las asociaciones entre distintas variables pueden ser sumamente útiles para hacer predicciones o medir conductas inadvertidas con anterioridad. Sin embargo, también pueden ser engañosas. Por ejemplo, saber que muchas personas en un área determinada están buscando información en Internet acerca de la influenza, puede resultar útil para orientar la estrategia de ventas de medicamentos, pero esto no indica necesariamente el brote de una epidemia de influenza.
- Las técnicas pueden simplificar mucho el fenómeno y ser poco adecuadas a todos los contextos. Por ejemplo, algunos investigadores sostienen que el análisis de los resultados de *Big Data* está desconectado del contexto social y, por lo tanto, no es posible capturar las experiencias subjetivas de los individuos.<sup>9</sup> Las opiniones difieren sobre qué tan bien se puede emplear el modelado de *Big Data Analytics* para predecir sistemas sociales dinámicos.<sup>10,11</sup>

Comunicar los resultados del análisis de *Big Data* puede ser difícil. Para facilitar ésto se están desarrollando nuevas herramientas. Por ejemplo, algunas empresas están recurriendo a software de visualización, como *Spotfire* y *QlikView*, para ayudar a sus empleados a interpretar los datos.

### Infraestructura y recursos humanos

Para manejar y lograr que *Big Data* tenga sentido, se requiere una combinación de habilidades y conocimientos especializados —como técnicas matemáticas, estadísticas y de

programación por computadora—, así como otras habilidades de comunicación en general y conocimiento del contexto del que proviene la información. Dado que encontrar todo esto en una sola persona puede resultar difícil, la estrategia común es formar equipos de trabajo multidisciplinarios. De hecho, las organizaciones reportan una escasez de recursos humanos con las habilidades necesarias para la tarea. Un estudio de la consultora *McKinsey & Company* indica que para 2018 los Estados Unidos podrían enfrentar una escasez de entre 140 y 190 mil empleados con capacidades para implementar *Big Data Analytics*, y de hasta 1.5 millones de directivos con capacidad de tomar decisiones con base en estos análisis.<sup>12</sup> Además, la competencia del sector privado por personal calificado puede inhibir la implementación gubernamental del *Big Data Analytics*.<sup>13</sup>

Es cierto que *Big Data* tiene el potencial de generar nuevos empleos, sin embargo, al combinarlo con otras tecnologías como la robótica móvil, también puede reducir la necesidad de algunos puestos de trabajo. Por ejemplo, una investigación ha estimado que, en los próximos 20 años será tecnológicamente posible automatizar cerca del 50% de todos los empleos en los Estados Unidos. Transporte, logística, apoyo administrativo, servicios, ventas y construcción son los sectores considerados de mayor riesgo.<sup>14</sup>

## Aplicaciones de *Big Data*

En el ámbito científico es larga la historia de proyectos de investigación que han llevado al límite el manejo y análisis de datos. A principios de la década 2000-2009, los investigadores involucrados en el Proyecto del Genoma Humano superaron enormes retos computacionales y de manejo de información para producir una de las primeras bases de datos en el mundo que alcanzaba los gigabytes (GB). Hoy los investigadores que analizan datos de experimentos físicos con partículas en el Gran Colisionador de Hadrones de la Organización Europea para la Investigación Nuclear (CERN, por sus siglas en francés), examinan cada año 15 petabytes (cerca de 16 millones de GB) de datos. Se prevé que el radiotelescopio Square Kilometre Array, que entrará en operación a mediados de la década 2020-2029, generará muchos petabytes de datos por segundo, dando lugar a avances sustanciales en computación y en técnicas de procesamiento de datos.

Además de proyectos de investigación, el *Big Data Analytics* se emplea en muchos otros campos: negocios, educación, delincuencia y seguridad, medición inteligente de servicios, transporte y gobierno. A medida que se vayan desarrollando nuevas aplicaciones, jugará papeles importantes en otros sectores como la industria de seguros y de servicios. La información de diversas fuentes puede organizarse, reunirse y clasificarse en categorías (transacciones, documentos de texto, publicaciones en redes sociales, registros de llamadas, e-mails, imagen, audio, video, etc). Por ejemplo, en México el Instituto del Fondo Nacional de la Vivienda para los Trabajadores (INFONAVIT) gestiona con éxito más de 40 millones de registros. Mediante herramientas que hacen uso de *Big Data Analytics*, se ha logrado homoligar catálogos del Instituto Nacional de Estadística y Geo-

grafía (INEGI), el Servicio Postal Mexicano (SEPOMEX) y el Servicio de Administración Tributaria (SAT) de la Secretaría de Hacienda y Crédito Público (SHCP) para ordenar y agilizar el acceso a la información de sus acreditados.<sup>15</sup> En el sector educativo, se está usando los datos para conocer la forma en que los alumnos interactúan con los materiales didácticos en línea y entender cómo aprenden y por ende, personalizar los cursos por Internet. En el caso del transporte existen aplicaciones (comúnmente conocidas como apps) que sugieren rutas con base en información de tráfico tomada de los teléfonos inteligentes de los usuarios.

## Legislación y percepción pública

La recopilación, almacenamiento y procesamiento de información personal es regulada por la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) y el acceso a información pública gubernamental por la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental (ver recuadro 4). *Big Data* puede contener información personal y privada, como los registros médicos, información financiera, etc., y también puede contener información sensible para las instituciones del gobierno.

### Recuadro 4. Legislación para la protección de datos personales y acceso a la información pública gubernamental

#### Protección a datos personales

- **Constitución Política de los Estados Unidos Mexicanos**<sup>19</sup>. Las fracciones II y III del artículo 6 son las primeras referencias expresas que hacen un reconocimiento del derecho a la protección de datos personales y se constituyen como limitantes al ejercicio del derecho de acceso a la información. En tanto, el segundo párrafo del artículo 16 establece que cualquier persona tiene derecho a la protección de sus datos personales. En el artículo 73, fracción XXIX-O, se establece que el Congreso tiene la facultad para legislar en materia de protección de datos personales en posesión de particulares.
- **Ley Federal de Protección de Datos Personales en Posesión de los Particulares**<sup>20</sup> (LFPDPPP). Regula derecho a la privacidad, que se concreta en la facultad de toda persona para ejercer control sobre la información personal que le concierne, contenida en registros públicos o privados, especialmente —pero no exclusivamente— los almacenados en medios informáticos.
  - De acuerdo al artículo 3, fracción V de la LFPDPPP un dato personal es toda aquella información que permita identificar a una persona.

#### Acceso a la información pública

- **Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental**<sup>21</sup>. Reconoce y regula el derecho individual al acceso a la información de las instituciones y organismos del Estado.

Ambas Leyes (datos personales en posesión de los particulares e información pública gubernamental) dotan al Instituto Federal de Acceso a la Información y Protección de Datos (IFAI) de atribuciones para garantizar el ejercicio de estos dos derechos.

Investigaciones sobre las actitudes y percepciones del público sobre el uso de datos personales han identificado la privacidad, la seguridad y la discriminación como las principales preocupaciones.<sup>16</sup> Por ejemplo, en talleres organizados para conocer el punto de vista del público, los participantes indicaron que la seguridad de sus datos personales era muy importante, pero que tenían poco control sobre ella. Otro estudio reveló preocupación sobre la posibilidad de discriminación laboral, por ejemplo, si los patrones tuvieran acceso a datos sobre la salud mental o VIH de sus empleados. Una buena gestión pública y la debida regulación de *Big Data* deberían mejorar la calidad de los datos, así como la confiabilidad de los resultados; asegurar que se cumplan los requerimientos legales relacionados con la privacidad de los datos; y propiciar el interés e involucramiento del público en proyectos de este tipo.

## Privacidad

El incremento en el número de dispositivos capaces de registrar información significa que gran parte de los datos que se están obteniendo son de aspectos de la vida que anteriormente eran privados. Por ejemplo, las tecnologías para el reconocimiento facial pueden identificar a la gente a partir de fotografías en Internet, en tanto que los celulares con GPS pueden rastrear la ubicación de una persona. Es posible vulnerar la privacidad si los datos se usan para propósitos distintos de aquellos para los cuales se recabaron en un principio; por ejemplo, si se obtuvieron de sensores sin el conocimiento de la gente, si distintos conjuntos de datos se vinculan o si éstos se venden de improviso a terceros.

Aplicar herramientas para la protección de datos antes de su análisis puede ayudar a proteger la privacidad de las personas. Estas herramientas incluyen varios procesos para eliminar partes de los datos que permitan identificar a las personas. Sin embargo, existe el riesgo de que un individuo pueda volver a ser identificado si los datos con segmentos eliminados se vinculan con otro conjunto que contenga datos de identificación.<sup>17</sup> En tal caso, debe ponerse énfasis en la solidez del proceso para eliminar los datos de identificación. Estudios para probar la efectividad de procesos sólidos de este tipo revelaron que era posible volver a identificar a las personas, aunque resultaba difícil y consumía mucho tiempo.<sup>18</sup>

## Seguridad

Al igual que con todos los datos, la seguridad de los conjuntos de *Big Data* puede estar comprometida si personas sin autorización logran acceder a ellos, o bien si son usados de manera inadecuada. Existen herramientas y procedimientos para reducir el riesgo de que alguien acceda a los datos sin autorización (ver recuadro 5). No obstante, es imposible garantizar que los datos estén completamente seguros. La técnica de *Big Data Analytics* puede ayudar también a mejorar la seguridad del sistema al analizar, por ejemplo, grandes archivos de registros para rastrear patrones que puedan evidenciar ataques y acceso no autorizado a los datos.

## Discriminación

El uso de *Big Data* puede dar lugar a discriminación no intencional.<sup>17</sup> Por ejemplo, un informe reciente en los Estados Unidos reportó que una evaluación realizada en 2009 a una base de datos, utilizada por patrones estadounidenses para confirmar la elegibilidad de los nuevos empleados, reveló que tomaba más tiempo verificar ciudadanos con nacionalidad distinta a la estadounidense. El mismo informe hizo notar que cierto manejo de *Big Data* puede propiciar precios diferenciados: un mismo producto en Internet se puede ofrecer a distintos precios dependiendo de qué tan “solvente” aparenta ser el cliente. Sin embargo, las tecnologías de *Big Data* pueden ayudar también a los grupos a hacer efectivos sus derechos identificando y confirmando casos de discriminación.<sup>17</sup>



### Recuadro 5. Ejemplos de técnicas para la seguridad de los datos

Existen varias herramientas y procedimientos que pueden ayudar a aumentar la seguridad de los datos.

- La implementación de procesos eficaces para la gestión pública en una organización (y asociados externos) puede ayudar a controlar el acceso a los datos, así como su uso. Puede significar también que los individuos se responsabilicen de la seguridad de los datos, brindar capacitación relevante, minimizar el número de personas con acceso y eliminar los datos cuando sea conveniente.
- La encriptación puede aumentar la seguridad de los datos. Por ejemplo, la traducción de un mensaje a un código ilegible para su transmisión y luego su conversión a una forma legible utilizando una “llave” para deshacer la encriptación.
- Para una actividad comunitaria, las distintas partes pueden colaborar con información sin tener que intercambiar directamente datos entre sí. Por ejemplo, los países que desean calcular el riesgo de que choquen sus satélites sin revelar las posiciones exactas, pueden transmitir datos a un algoritmo seguro que calcula automáticamente dicho riesgo.
- El acceso a datos sensibles puede ser proporcionado a través de un “refugio seguro”, es decir, un sitio o conjunto de arreglos administrativos para ayudar a la comunicación segura de información confidencial.

## Referencias

1. The digital universe of opportunities, 2014, IDC, [bit.ly/1jvQFck](http://bit.ly/1jvQFck)
2. Schroeck, M. et al., Analytics: the real-world use of big data, 2012, IBM
3. Yiu, C., The big data opportunity, 2012, Policy Exchange
4. Manyika, J. et al., Big data: the next frontier, 2011, McKinsey
5. Social media data & real time analytics, HoC Sci. & Tech. Com. [bit.ly/1eMJcEK](http://bit.ly/1eMJcEK)
6. Improving the transparency & accountability of gov., 10/07/14, [bit.ly/1hDCHG2](http://bit.ly/1hDCHG2)
7. Tecnología Big Data & Analytics se expandirá 200% en América Latina en los próximos cinco años, consultado el 17 de septiembre de 2015 en <http://mx.idclatin.com/releases/news.aspx?id=1643>
8. Mercado de Big Data y Analytics se expande en México tres veces más que en Latinoamérica, consultado el 17 de septiembre de 2015 en <http://mx.idclatin.com/releases/news.aspx?id=1753>
9. Big data, RCUK, consultado el 24 de junio de 2014, [bit.ly/W1CHbF](http://bit.ly/W1CHbF)
10. Hetherington, S. et al., The thin data revolution, 2014
11. Uprichard, E., Big data, little questions? 2013
12. Big data: The next frontier for innovation, competition, and productivity. Consultado el de septiembre de 2015 en [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)
13. Market assessment for public sector information, 2013, Deloitte, BIS
14. Frey, C. B. et al., The future of employment, 2013, Universidad de Oxford
15. Revista Information Week México. 2012. Infonavit gestiona con éxito más de 40 millones de datos.
16. Big data, public views, 2014, Sciencewise, [bit.ly/1rZyXV7](http://bit.ly/1rZyXV7)
17. Big data: seizing opportunities, preserving values, 2014, Casa Blanca
18. Cavoukian, A. et al., Big data and innovation, junio de 2014, [bit.ly/1uyQ9ha](http://bit.ly/1uyQ9ha)
19. Constitución Política de los Estados Unidos Mexicanos texto vigente, consultado en <http://www.diputados.gob.mx/LeyesBiblio/htm/1.htm>
20. Ley Federal de Protección de Datos Personales en Posesión de los Particulares, consultada en <http://inicio.ifai.org.mx/SitePages/Consultar-la-LFPDPPP.aspx>
21. Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental, consultado en [http://www.dof.gob.mx/nota\\_detalle.php?codigo=727870&fecha=11/06/2002](http://www.dof.gob.mx/nota_detalle.php?codigo=727870&fecha=11/06/2002)
22. Chen, H., Chiang R. H. y Storey, V. C. Business intelligence and analytics: from big data to big impact, 2012