

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

Documento de investigación desarrollado para la Dirección de Gobierno en Línea del Ministerio TIC, en el marco del programa Talento Digital, a partir del trabajo de investigación entregado a la Universidad de Manchester para el grado de Maestría en Tecnologías de Información y Comunicaciones para el Desarrollo.

2014

Liliana Fernández Gómez

Beneficiaria Fondo Talento Digital

Table of Contents

Lista de tablas	4
Lista de figuras.....	4
1. Introducción	6
1.1. Objetivos de la investigación	6
1.2. Diseño de la investigación	6
1.2.1. Fase 1. Comprensión.....	6
1.2.2. Fase 2. Ideación y acciones propuestas para cierre de brechas.....	7
1.3. Métodos de recolección de datos	8
2. Revisión de literatura	9
2.1. Definición de big data.....	9
2.2. Beneficios potenciales del aprovechamiento de big data.....	16
2.2.1. Mejor toma de decisiones	16
2.2.2. Mejor segmentación de poblaciones para personalizar acciones	18
2.2.3. Innovación a través de nuevos modelos de negocio, productos y servicios.....	18
2.2.4. Eficiencia y efectividad.....	18
2.2.5. Mejores productos y servicios para los clientes	19
2.2.6. Reducción de amenazas de seguridad y crimen.....	19
2.3. Restricciones y riesgos.....	19
2.3.1. Recursos humanos.....	20
2.3.2. Disponibilidad, seguridad, privacidad y calidad de los datos	23
2.3.3. Aspectos legales con respecto al aprovechamiento de big data	24
2.3.4. Aspectos tecnológicos.....	24
2.3.5. Riesgos relacionados con el aprovechamiento de big data	25
2.4. Big Data y el sector público.....	28
2.4.1. Potenciales beneficios del aprovechamiento de big data en sector público	28
2.4.2. Restricciones y riesgos de uso de big data en el sector público.....	33
3. Condiciones ideales para aprovechamiento de big data.....	38
3.1. Información.....	38
3.2. Tecnologías	39
3.3. Procesos	39
3.4. Objetivos	40
3.5. Personal/Talento	40
3.6. Sistemas y estructuras de gestión.....	41

3.7. Otros recursos	41
4. Realidad y brechas	43
4.1. Realidad	43
4.1.1. Información	43
4.1.2. Tecnologías	45
4.1.3. Procesos	46
4.1.4. Objetivos	46
4.1.5. Personal/Talento.....	47
4.1.6. Sistemas y estructuras de gestión.....	51
4.1.7. Otros recursos	53
4.2. Brechas identificadas	55
5. Recomendaciones para el cierre de brechas.....	57
5.1. Ecosistema de datos	58
5.2. Innovación en el sector público	61
5.3. Talento humano	62
5.4. Cierre de brechas entre entidades públicas	64
5.5. Otras recomendaciones	65
Referencias.....	67
Anexo 1 Resumen de de los posibles usos de big data más mencionados en la literatura y sus potenciales beneficios.	73

Word count Text: 16.450

Word count appendixes: 596 (3 pages)

Lista de tablas

Tabla 1. Fuentes de variedad en big data 12

Tabla 2. Técnicas y tecnologías para el uso de big data 15

Tabla 3. Brechas identificadas 55

Lista de figuras

Figura 1. Design-Reality Gap Model 7

Figura 2. Ecosistema de datos propuesto por World Economic Forum 59

Resumen

Big data está generando una creciente atención como fuente de oportunidades para los negocios y el desarrollo social. Esta investigación busca contribuir a la evaluación de dichas oportunidades en Colombia con especial énfasis en el sector público.

La primera parte de la investigación ofrece elementos para una buena comprensión de big data, sus potenciales beneficios y las barreras y riesgos involucrados en su adopción. Esta fase también identifica las condiciones ideales para un mejor aprovechamiento de big data para compararlas con las condiciones reales de Colombia y su sector público.

La segunda fase de esta investigación presenta algunas recomendaciones sobre cómo puede el gobierno colombiano mejorar las condiciones para un mejor aprovechamiento de big data.

1. Introducción

1.1. Objetivos de la investigación

Como se mencionó anteriormente esta investigación contribuye a las fases de comprensión e ideación del proceso de diseño e innovación ejecutado por la Dirección de Gobierno en Línea. Big data fue priorizado, como tema de investigación, debido a la cantidad de autores que consideran que puede ofrecer grandes oportunidades para los gobiernos y la sociedad.

1.2. Diseño de la investigación

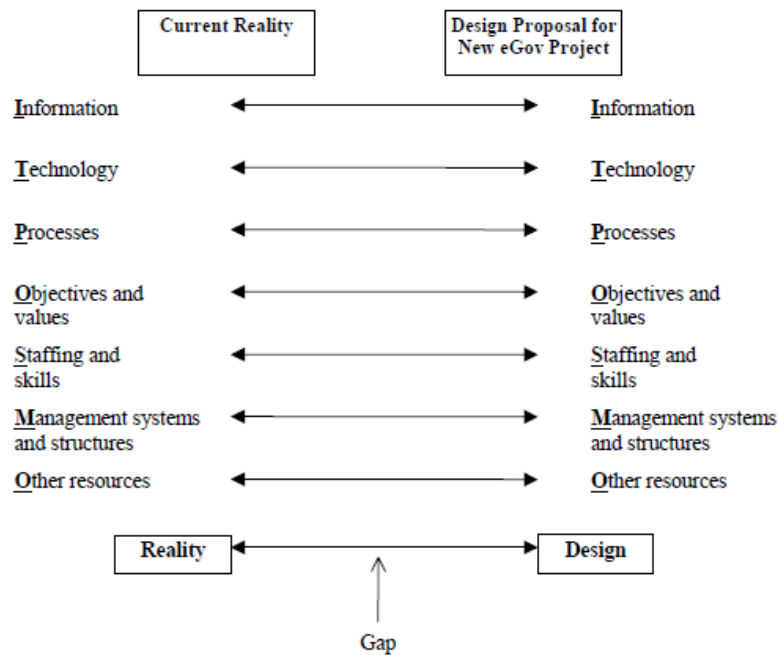
1.2.1. Fase 1. Comprensión

El primer componente de esta fase es presentado en la revisión de literatura e incluye la definición de big data y sus implicaciones en términos de beneficios, campos de aplicación, tecnología, barreras, riesgos y discusiones actuales relacionadas con este concepto.

El segundo componente de esta fase también se presenta en la revisión de literatura e involucra beneficios específicos, campos de aplicación, restricciones, riesgos y discusiones actuales relacionadas con el uso de big data en el sector público.

El tercer y cuarto componente de esta fase son desarrolladas con el uso del modelo Design-Reality Gap Model [Brechas entre diseño y realidad] de Heeks (2003) presentado en la Figura 1. Con este modelo, *Diseño* se usa para describir las condiciones ideales para el aprovechamiento de big data, *Realidad* se usa para describir las condiciones actuales de Colombia y las *Brechas* son usadas para identificar las diferencias entre las condiciones ideales y las reales.

Figura 1. Design-Reality Gap Model



Fuente: (Heeks 2003)

El análisis de diseño, realidad y brechas es guiado por la lista ITPOSMO (Heeks 2003) que usa un acrónimo para las dimensiones clave de los proyectos de gobierno electrónico:

- **I**nformation
- **T**echnology
- **P**rocesses
- **O**bjectives and values
- **S**taffing and skills
- **M**anagement systems and structures
- **O**ther resources

1.2.2. Fase 2. Ideación y acciones propuestas para cierre de brechas

Esta fase está orientada a proponer acciones y buenas prácticas que podrían contribuir al cierre de las brechas identificadas en la fase 1.

1.3.Métodos de recolección de datos

Dado que esta investigación explora un tema relativamente nuevo y la primera fase está enfocada en la comprensión; las fuentes de datos son en su mayoría documentos producidos por investigadores, compañías de TI, empresas de consultoría, organismos internacionales y entidades públicas en diferentes países.

Sin embargo, para evaluar la situación actual en Colombia y las brechas con respecto a las condiciones ideales, es necesario considerar algunos datos primarios e información de diferentes actores de interés en el contexto colombiano. Se hizo un trabajo de observación, durante 2 eventos organizados por MINTIC: la CIO Summit y Colombia 3.0. Además de eso se entrevistaron algunos actores que resultaron relevantes de acuerdo con criterios encontrados en la revisión de literatura.

2. Revisión de literatura

Esta sección contribuye a la fase de comprensión de esta investigación y presenta aspectos clave de big data, encontrados en la literatura disponible, incluyendo definiciones, beneficios potenciales, restricciones, riesgos y consideraciones específicas para el uso de big data en el sector público.

2.1. Definición de big data

El concepto de big data está siendo desarrollado a medida que su comprensión y adopción en las organizaciones está avanzando (United Nations 2012). Con respecto a la palabra “Big”[Grande], algunas definiciones están enfocadas en el tamaño de los archivos que pueden ser creados, almacenados y procesados (Manyika et al. 2011), otras consideran que el tamaño de los archivos no es relevante porque el continuo incremento en las capacidades de cómputo y almacenamiento produce un efecto por el cual el tamaño que hoy es considerado “grande” será considerado normal el próximo año (Beyer and Laney 2012; Russom 2011). United Nations (2012) plantea que esta palabra implica enorme alcance y poder; mientras Beyer and Laney (2012) mencionan que los datos no son necesariamente grandes sino que la necesidad de combinar nuevos tipos de información demanda mayores capacidades de procesamiento y gestión.

Aunque no hay una definición unánime de big data, la literatura disponible ofrece una entendimiento común de big data como un fenómeno producido por la abundancia de dispositivos electrónicos que están generando, almacenando, transmitiendo y procesando datos, así como por el creciente número de personas generando datos a través del uso de software y aplicaciones incluyendo servicios transaccionales, redes sociales, entretenimiento, dispositivos móviles y servicios basados en localización (Manyika et al. 2011, Helbing and Ballesteri 2011; United Nations 2012; TechAmerica Foundation 2012; Schulte et al. 2013) .

Usando este entendimiento común de big data como fenómeno, varios autores enfatizan tres aspectos que describen este fenómeno y relacionan estos aspectos con “3 Vs”, *Volumen, Variedad and Velocidad* (United Nations 2012). Adicionalmente, algunos autores agregan Veracidad (TechAmerica Foundation 2012), Valor, and Variabilidad (Fan and Bifet 2013) involucrando elementos para entender las implicaciones de big data. En esta sección, se presentan algunos elementos de comprensión del *Volumen, Variedad, Velocidad and Valor*.

Volumen está relacionado con la cantidad de datos a ser capturados, almacenados and procesados así como la enorme capacidad de cómputo que demandan (Beyer and Laney 2012). Adicionalmente, se menciona el hecho de que las decisiones tengan sustento no solo en datos internos sino también en grandes volúmenes de información externa (Buhl et al. 2013).

Velocidad está relacionada con altas tasas de generación y consumo de datos (Helbing and Balietti 2011; Beyer and Laney 2012). Algunas nuevas fuentes de datos, tales como dispositivos móviles, sensores, dispositivos de Identificación por Radio Frecuencia-RFID, videos de vigilancia o inclusive la actividad de redes sociales; son las principales generadores de velocidad y procesamiento de datos en tiempo real (Kart et al. 2013). La existencia de internet, una red que permite una rápida interconexión e intercambio de datos entre un creciente número de fuentes de datos también contribuye a la *Velocidad* en big data (Helbing and Balietti 2011). Otro factor es la volatilidad que implica que la información debe ser recolectada, procesada, verificada y usada en corto tiempo porque los datos cambian rápidamente (United Nations 2012).

Variedad está relacionada con diversidad de fuentes que están generando datos, big data no solo concierne a datos que están estructurados apropiadamente para ser almacenados y procesados por bases de datos o bodegas de datos sino que involucra también datos no estructurados, semiestructurados o complejos (volátiles) (TechAmerica Foundation 2012). Con respecto a datos volátiles que están en movimiento, es importante enfatizar que inclusive información que es transmitida pero

no almacenada es considerada parte of big data (Manyika et al. 2011). Adicionalmente, hay dos componentes de este fenómeno que contribuyen a la complejidad y variedad; fuentes de datos más descentralizadas y una cantidad de personas dejando rastros digitales en sus vidas diarias (Helbing and Baliaetti 2011).

La variedad de los datos también está relacionada con la manera en la cual son generados y recolectados, algunos datos son producidos digitalmente por dispositivos electrónicos (p.e. sensores); otros son producidos por seres humanos de manera pasiva como parte de su vida diaria (p.e. desplazamientos o compras); otros son recolectados automáticamente con propósitos específicos (p.e. sistemas transaccionales); otros son geográfica o temporalmente rastreables (p.e. localización de dispositivos móviles o duración de llamadas) y otros son monitoreados continuamente (p.e. datos recolectados por sismógrafos) (United Nations 2012).

Con el propósito de proporcionar una visión general sobre las fuentes de datos que son consideradas potenciadores del fenómeno de big data, la Tabla 1 presenta un compendio de las mismas. En resumen, la variedad es un aspecto de este fenómeno que involucra varias perspectivas, desde el nivel de estructuración de los datos hasta la manera en la cual son producidos los datos, los propietarios de los datos y el objetivo primario para el cual son generados.

Tabla 1. Fuentes de variedad en big data

Clasificación	Ejemplos
Registros públicos y privados	<p>Información recolectada en las actividades administrativas y operativas de entidades públicas y privadas incluyendo impuestos, salarios, salud, seguros, propiedades, negocios y compras entre otros.</p> <p>Conjuntos de datos abiertos</p>
Datos generados por servicios electrónicos	<p>Transacciones, comunicaciones, pagos, monitoreo de fluctuaciones en el consumo de electricidad de dispositivos electrónicos.</p> <p>Imágenes diagnósticas (p.e. tomografías, rayos X, resonancia magnética)</p>
Datos generados por actividades en internet	<p>Esto incluye:</p> <ul style="list-style-type: none"> • Logs de proveedores de servicios de internet (p.e direcciones IP y MAC) • Logs de acceso remoto de teléfonos y computadores • Historiales de navegación • Memoria caché de navegación • Archivos cookies • Términos de búsqueda • Rastreo de clics • Análisis físico de variaciones en el campo electromagnético de dispositivos inalámbricos.
Datos de dispositivos móviles	<ul style="list-style-type: none"> • Datos de localización • WLAN/Wi-Fi open hot-spots, • Dispositivos bluetooth • Datos de RFID • Carnés y tarjetas electrónicas (p.e. para acreditar participantes en conferencias, cuartos de hotel, etc.)
Contenido capturado desde diversos dispositivos multimedia	<ul style="list-style-type: none"> • Fotografías satelitales e imágenes de Google Street View • Webcams públicas
Contenidos generados por usuarios	<ul style="list-style-type: none"> • (Emails, foros web, blogs, chats, etc.), • Listas de correo o grupos de discusión en internet • Escaneo de palabras clave de las cuentas de correo gratuitas • Datos de redes sociales • Aplicaciones de crowdsourcing
Datos de seguridad	<ul style="list-style-type: none"> • Videos de vigilancia (CCTV) • Datos de reconocimiento de rostros • Datos biométricos • Grabaciones de audio, registros de micrófonos direccionales • Vigilancia de llamadas • Fotografías de radares de velocidad • Datos de objetos y personas escaneadas en aeropuertos • Formularios de seguridad diligenciados

Fuente: Adaptado de (Helbing and Baliaetti 2011; Manyika et al. 2011; Lohr 2012)

Con respecto al *Valor*, la generación de valor a partir de big data está relacionado con el análisis y la comprensión de los datos más que en el uso de tecnologías para su generación o almacenamiento. Manyika et al. (2011) plantean que el uso de los datos es la fuente de creación de valor en este fenómeno; de acuerdo con estos autores, el valor real y la ventaja competitiva para las organizaciones está relacionada con la cantidad de datos consumidos, comprendidos y aprovechados. El aprovechamiento de big data es la conversión de datos diversos, complejos y desestructurados en información aprovechable (United Nations 2012). En ese sentido, algunos autores presentan la analítica como la principal fuente de creación de valor en este fenómeno (TechAmerica Foundation 2012; Howard 2012a; Newman and Beyer 2012; Brown et al. 2013; Biesdorf et al. 2013; Heudecker 2013; Buhl et al. 2013; Baghdassarian et al. 2013; Schulte et al. 2013; Brown et al. 2014). Es importante resaltar que, aunque la analítica es reconocida por la literatura como la principal fuente de valor en big data, no es la única y por esa razón no se debe considerar big data como sinónimo de analítica o analítica avanzada.

La novedad ofrecida por el fenómeno de big data para el campo de la analítica merece atención especial. Los conceptos previos relacionados con la analítica, como inteligencia de negocios o minería de datos, han sido usados para proveer entendimiento sobre mercados, comportamiento de clientes y optimización de operaciones (Laney et al. 2013; Schulte et al. 2013). Sin embargo, big data habilita algunas mejoras en el campo de la analítica, la primera es la posibilidad de análisis en tiempo real que implica un cambio de paradigma como consecuencia de la velocidad (United Nations 2012; Schulte et al. 2013); el paradigma anterior era primero consolidar datos y después analizarlos (Helbing and Balmietti 2011); la segunda es consecuencia de la variedad de fuentes de datos lo cual lleva a múltiples oportunidades de análisis incluyendo:

- Análisis de comportamiento, análisis de sentimiento en redes sociales (Heudecker 2013) y análisis semántico (Laney et al. 2013).
- Un abanico de análisis descriptivo, diagnóstico, predictivo y prescriptivo (Schulte et al. 2013). La predicción y prescripción son habilitadas por el hecho

de que, a diferencia de la analítica en inteligencia de negocios, análisis estadístico y minería de datos (Newman and Beyer 2012), la analítica en el marco de big data no solo está soportada en información histórica y estructurada sino también en datos en tiempo real (Schulte et al. 2013) y en múltiples fuentes de información inclusive fuera de las organizaciones (Biesdorf et al. 2013).

- La abundancia de datos es un facilitador de procesos de aprendizaje de máquinas e inteligencia artificial (Lohr 2012).
- Descubrir es un verbo usado por varios autores en relación con la analítica en el marco de big data. Esto implica oportunidades para buscar elementos escondidos que pueden ser revelados por el uso de diversas fuentes de datos (Manyika et al. 2011; Lohr 2012; TechAmerica Foundation 2012; Beyer and Laney 2012; Schulte et al. 2013; Laney et al. 2013).
- Análisis de datos de ubicación (Manyika et al. 2011).
- Inferencias y razonamientos autónomos en el aprendizaje de máquinas, sin preguntas generadas por usuarios (Schulte et al. 2013).

Big data no es una tecnología por sí misma (TechAmerica Foundation 2012), hay un amplio rango de herramientas requeridas para el aprovechamiento de big data incluyendo captura, acceso, almacenamiento, transmisión, presentación/interpretación y análisis (Beyer and Laney 2012). De acuerdo con la literatura, la analítica y la visualización son dos grupos clave de tecnologías y técnicas requeridas para el aprovechamiento de big data.

Dichas técnicas y tecnologías no se originan solamente en el campo de las tecnologías de información, algunas de ellas están más relacionadas con estadística, matemáticas aplicadas y economía (Manyika et al. 2011). Así mismo, algunas de ellas fueron adaptadas para enfrentar los retos presentados por el fenómeno de big data mientras que otras fueron creadas específicamente para generar valor en contextos de big data (Manyika et al. 2011). También es importante entender que las tecnologías emergentes para big data no están en condiciones de reemplazar las bases de datos

relacionales ni las bodegas de datos sino que pueden ofrecer nuevas oportunidades para la gestión de información y la analítica (Kart et al. 2013).

La Tabla 2 presenta una visión general de los tipos de técnicas y tecnologías más directamente usadas para aprovechamiento de big data. Es importante decir que algunas de esas técnicas y tecnologías ya son parte de la vida diaria de los usuarios de computadores (p.e. el motor de búsqueda de Google usa enormes cantidades de datos para predecir los términos de búsqueda y facilitar la toma de decisiones de los usuarios en sus búsquedas) (Lohr 2012).

Table 2. Técnicas y tecnologías para el uso de big data

Técnica/tecnología	Propósito	Ejemplos
Técnicas para análisis de big data	Están siendo mejoradas para facilitar el análisis de nuevas combinaciones de datos y fueron principalmente originadas en los campos de la estadística y el aprendizaje de máquina.	<ul style="list-style-type: none"> • A/B testing • Association rule learning • Aprendizaje supervisado • Aprendizaje no supervisado • Cluster analysis • Crowdsourcing • Fusión e integración de datos (p.e. procesamiento de señales y procesamiento de lenguaje natural que es usado para análisis de sentimientos) • Data mining (involucra otras técnicas mencionadas acá) • Algoritmos genéticos (algunas aplicaciones incluyen optimización de horarios de trabajo y portafolios de inversión) • Redes neuronales (usadas para reconocimiento de patrones) • Modelamiento predictivo (p.e. regresiones) • Análisis espacial • Simulación • Análisis de series de tiempo
Visualización	Técnicas y tecnologías para crear imágenes, diagramas o animaciones para facilitar el consumo de información por parte de las personas.	<ul style="list-style-type: none"> • Tag cloud, clustergram, history flow, spatial information flow
Sistemas de bases de datos distribuidos /sistemas de bases de datos NoSQL /Bodegas de datos lógicas-Logical Data warehouse	Usadas para agregar, manipular, administrar y analizar big data	<ul style="list-style-type: none"> • Big Table, Cassandra (open source), Dynamo
Stream processing or event stream processing.	Grupo de tecnologías para procesar grandes cantidades de datos en tiempo real. Son usadas para la construcción de sistemas de información orientados por eventos.	<ul style="list-style-type: none"> • Algoritmos de negociación en servicios financieros • Aplicaciones de procesamiento de eventos RFID • Detección de fraudes • Monitoreo de procesos

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

Técnica/tecnología	Propósito	Ejemplos
		<ul style="list-style-type: none"> • Servicios basados en localización a partir de telecomunicaciones
R	Lenguaje estándar de programación, de código abierto, para cómputo de estadísticas y gráficas.	<ul style="list-style-type: none"> • Desarrollo de software estadístico y de análisis de datos.
Hadoop	Marco de trabajo de programación, de código abierto, para procesar grandes conjuntos de datos distribuidos. Es mencionado frecuentemente, en la literatura, como una de las principales herramientas para procesamiento de big data.	<ul style="list-style-type: none"> • Manipulación de grandes cantidades de texto • Construcción de índices de búsqueda en la web • Procesamiento de noticias/historial de contenidos y optimización de grupos objetivo • Clasificaciones y agrupaciones/clustering (p.e. detección de fraude y spam) • Análisis de léxico en redes sociales • Identificación de tendencias de palabras en redes sociales • Filtrado colaborativo • batch video/transcodificación de imágenes
Mashup	Aplicación que usa y combina presentación de datos desde dos o más fuentes (incluyendo fuentes de datos abiertos) para crear nuevos servicios que generalmente serán ofrecidos en la web.	

Fuente: Adaptado de (Manyika et al. 2011) (Helbing and Ballezzi 2011) (Newman and Beyer 2012)

2.2. Beneficios potenciales del aprovechamiento de big data

De acuerdo con la literatura, el aprovechamiento de big data ofrece múltiples beneficios a organizaciones e individuos y es por esto que es considerado una frontera de innovación (Manyika et al. 2011; Beyer and Laney 2012) así como un fenómeno revolucionario y transformador (United Nations 2012; Brown et al. 2014). Esta sección resume los beneficios potenciales del aprovechamiento de big data usando seis grupos de beneficios que son identificables en la literatura.

2.2.1. Mejor toma de decisiones

El uso de big data habilita un amplio rango de análisis; desde la descripción, al diagnóstico, predicción y prescripción en un gran número de áreas organizacionales incluyendo mercadeo, ventas, compras y contrataciones, gestión de la cadena de suministro, manufactura, ingeniería, gestión del riesgo, finanzas y recursos humanos. Este amplio rango de áreas análisis, y la velocidad involucrada en big data, ofrecen

oportunidades para mejorar la toma de decisiones en el nivel estratégico, tático y operativo (Schulte et al. 2013).

Un beneficio potencial del uso de big data para la toma de decisiones es el cambio de la toma de decisiones basada en la intuición (Kart et al. 2013) o “the Highest Paid Person’s Opinions”[la opinión de la persona mejor pagada de la organización] (Manyika et al. 2011) a la toma de decisiones guiada por datos (Manyika et al. 2011; Buhl et al. 2013), guiada por hechos/evidencia (United Nations 2012; Kart et al. 2013; Schulte et al. 2013) o liderada por datos (Kart et al. 2013) para mejorar el desempeño organizacional (Manyika et al. 2011; Schulte et al. 2013). Este nuevo enfoque para la toma de decisiones habilita la toma de decisiones inteligente (TechAmerica Foundation 2012) a través del descubrimiento de activos de información, relaciones y nivel de influencia entre nuevas fuentes de datos (Beyer and Laney 2012).

El uso de esas nuevas fuentes de datos podría reducir errores y atrasos en la toma de decisiones a través del aprovechamiento, por ejemplo, de sensores en tiempo real y la “wisdom of crowds”[sabiduría de las masas] (Helbing and Balmelli 2011). Las nuevas fuentes de datos pueden ser combinadas para llenar vacíos de información (Newman and Beyer 2012) y dar cabida al concepto de minería de la realidad (United Nations 2012).

Además del uso de nuevas fuentes de datos, las posibilidades de reemplazar o apoyar la toma de decisiones humana con algoritmos automatizados (TechAmerica Foundation 2012) es vista, por algunos autores, como otra manera de mejorar la toma de decisiones con big data. En este punto, la automatización podría ser uno de los objetivos y la facilitación de la toma de decisiones otro. Big data facilita la toma de decisiones a través del análisis de conjuntos de datos que, debido a su tamaño y complejidad, no son manejables para las personas con el uso de hojas de cálculo o herramientas similares (Manyika et al. 2011).

La complejidad puede ser manejada con el uso de técnicas y tecnologías de big data y esto puede llevar a reducir vacíos de conocimiento y comprensión de sistemas “tecnosocio-económico-ambientales” complejos (Helbing and Baliatti 2011). Big data podría favorecer tres escenarios para una mejor toma de decisiones: alertas tempranas, evaluación en tiempo real y retroalimentación en tiempo real (United Nations 2012).

2.2.2. Mejor segmentación de poblaciones para personalizar acciones

Una mejor segmentación puede ser habilitada por la variedad y granularidad espaciotemporal de los datos generando aplicaciones como microsegmentación en tiempo real de clientes y ciudadanos para servicios específicos, promociones y publicidad (Manyika et al. 2011).

2.2.3. Innovación a través de nuevos modelos de negocio, productos y servicios

Algunos autores mencionan que hay una amplia variedad de oportunidades de negocio que se pueden generar por el fenómeno de big data (Manyika et al. 2011; Beyer and Laney 2012; TechAmerica Foundation 2012; Laney et al. 2013; Helbing and Baliatti 2011). En ese sentido, Laney et al. (2013) menciona la existencia de un documento de Gartner Research que contiene 100 oportunidades e ideas de negocio que tienen gran potencial de generación de valor a través de la monetización directa o indirecta de datos.

2.2.4. Eficiencia y efectividad

Big data podría permitir a las organizaciones hacer más con menos y producir resultados de mejor calidad (Manyika et al. 2011). Esas mejoras en productividad se pueden generar a través de la reducción de insumos generados por personas y los retrasos en la producción, recolección y transmisión de información (United Nations 2012) así como a través de la eliminación de procesos de negocio sin valor agregado (Laney et al. 2013). Adicionalmente, el uso de big data podría facilitar la comprensión de las causas raíz con respecto a las diferencias en desempeño de unidades organizacionales, o inclusive proveedores externos, para mejorar el desempeño general de las organizaciones (Manyika et al. 2011).

2.2.5. Mejores productos y servicios para los clientes

Dado que las personas están produciendo datos, inclusive en el desarrollo de sus actividades diarias, ellas(os) son actores en un nuevo ecosistema de datos y podrían tener servicios mejorados así como mejores precios y ofertas (United Nations 2012; World Economic Forum 2012). En ese sentido, Laney et al. (2013) plantea que la nueva ola después de la analítica de negocios será la analítica personal, donde datos y analítica estarán orientados a atender las necesidades de análisis de los clientes (y ciudadanos) más que las de las organizaciones.

Las personas pueden obtener beneficios de este fenómeno debido a la aparición de nuevos modelos de negocio. World Economic Forum (2012) menciona un caso acerca de una empresa naciente en Estados Unidos que usa incentivos financieros para promover que los clientes en mercados emergentes compartan datos a través de sus celulares. Los datos son usados para investigaciones de mercado en más de 50 países y la empresa compensa a las personas con saldo para llamadas.

2.2.6. Reducción de amenazas de seguridad y crimen

El volumen, la velocidad y la variedad de datos podrían incrementar la capacidad para reducir el fraude y para identificar clientes, ciudadanos y empleados que abusan de beneficios o servicios (Manyika et al. 2011; TechAmerica Foundation 2012). Con respecto a la seguridad, big data podría facilitar la prevención de ataques usando una variedad de fuentes de datos como imágenes, identificación biométrica, reconocimiento de rostros, radares de velocidad, llamadas telefónicas y videos de vigilancia (Manyika et al. 2011).

2.3. Restricciones y riesgos

Aunque el aprovechamiento de big data es considerado revolucionario y hay muchos beneficios potenciales identificados, las organizaciones tienen que enfrentar importantes restricciones y una ruta relativamente larga para ser capaces de hacer realidad dichos beneficios (TechAmerica Foundation 2012). Adicionalmente, algunos

autores presentan posibles riesgos que surgen con este fenómeno. Esta sección resume dichas restricciones y riesgos.

2.3.1. Recursos humanos

Biesdorf et al. (2013) plantean que las iniciativas de big data pueden ser decepcionantes cuando las organizaciones carecen de las habilidades y capacidades adecuadas. Como se mencionó anteriormente, el valor de big data está muy relacionado con analítica y algunas técnicas y tecnologías que soportan la analítica son originarias de otros campos de estudio, diferentes a TIC, como estadística, matemáticas aplicadas, economía y genética. En este escenario, el talento humano surge como una restricción de gran relevancia para el aprovechamiento de big data y muchos autores identifican la presencia de esta restricción en diferentes niveles organizacionales.

En el nivel directivo, se requiere conocimiento práctico y habilidad para consumir información (Manyika et al. 2011) lo cual significa personas capaces de hacer las preguntas correctas, analizar, interpretar y cuestionar los resultados con una visión crítica para tomar las decisiones apropiadas (Manyika et al. 2011; United Nations 2012). Así mismo los gerentes requieren mejorar su habilidad para identificar qué información está haciendo falta y cómo se puede conseguir (Beyer and Laney 2012).

Sin embargo, la situación parece ser distinta en las organizaciones, una encuesta realizada por Gartner, entre 720 tomadores de decisiones de diferentes sectores (incluyendo gobierno) en Norte América, Europa, Medio Oriente, África, Asia Pacífico y Latinoamérica; muestra que 56% de ellos consideran un gran reto saber qué fuentes de datos analizar y qué problemas de negocio se pueden resolver con big data (Kart et al. 2013).

Además de la carencia de habilidades para el consumo de información, el nivel directivo podría presentar limitaciones que son más actitudinales y culturales. Dichas

limitaciones podrían resumirse como una mentalidad poco orientada hacia los datos, la mentalidad orientada hacia los datos es mencionada por Manyika et al. (2011) como un factor clave para el aprovechamiento de big data y podría ser entendida como la comprensión de los beneficios de la toma de decisiones basada en datos (Buhl et al. 2013). Esta restricción también está relacionada con la consideración que hacen Beyer and Laney (2012) en el sentido de que, el fenómeno de big data está cambiando la idea de "trabajar con lo que tenemos" y crea escenarios para "trabajar con lo que podamos conseguir/encontrar". En esos nuevos escenarios se requieren personas con pensamiento creativo y curiosidad intelectual (TechAmerica Foundation 2012). Con respecto a la carencia de esta mentalidad y del liderazgo necesario para iniciativas de big data, Kart et al. (2013) plantea que los ejecutivos están más acostumbrados a sustentar decisiones en su experiencia e instinto.

En el nivel analítico de las organizaciones, se requieren personas con capacidades técnicas en estadística y aprendizaje de máquina. Ellas (os) deberían estar en capacidad de analizar grandes volúmenes de datos para generar conocimiento de negocio (Manyika et al. 2011; Brown et al. 2013). En este nivel, también es importante contar con personas que tengan conocimiento práctico y que tengan la habilidad de consumir información que es requerida por el nivel directivo (Manyika et al. 2011).

En el nivel operativo y de soporte, se requieren profesionales dedicados a desarrollar, implementar y mantener hardware y software requerido para el uso de big data (Manyika et al. 2011). Aquí se esperan algunos cambios (p.e. los expertos en infraestructura TI necesitan cambiar su foco del volumen y la velocidad hacia la gestión e integración de gran variedad de fuentes de datos (Beyer and Laney 2012)). Adicionalmente, algunas técnicas y tecnologías requieren ser implementadas por expertos (Beyer and Laney 2012).

Brown et al. (2014) presenta un estudio en el cual los líderes de analítica en grandes compañías privadas exitosas contribuyen a la identificación de retos relacionados con el uso de datos. Ellos identifican la necesidad de contar con “traductores”, personas

cuyo talento haga puente entre disciplinas como TI, ciencias de los datos, analítica y el ambiente del negocio en el que trabajan. Los participantes de este estudio consideran que se requieren tres roles clave; el primero es el estratega de los datos que combina conocimiento en TI y experiencia en la toma de decisiones del negocio, este rol debería ser quien defina requerimientos de datos para generar valor con analítica. El segundo rol es el científico de datos, una combinación de vasta experiencia en analítica y conocimiento en TI; este rol debería ser el que desarrolle modelos y algoritmos de análisis de datos. El tercer rol es el consultor de analítica, una combinación de conocimiento práctico del negocio y experiencia en analítica; este rol debería contribuir a la formulación estratégica de iniciativas de big data. Además de estos, algunos autores consideran la necesidad de nuevos roles como el Chief Data Officer o el Chief Analytics Officer (TechAmerica Foundation 2012; Buhl et al. 2013; Brown et al. 2013; Laney et al. 2013) así como el estratega de la información o el gerente de productos de información (Laney et al. 2013).

En el mismo sentido, algunos autores consideran la interdisciplinariedad como un reto para la conformación de equipos de big data (Manyika et al. 2011; Lohr 2012; Brown et al. 2014). Con respecto a la interdisciplinariedad, Heudecker (2013) enfatiza que, además de los datos, este fenómeno involucra el uso creativo de varios tipos de conocimiento humano. Este autor menciona la psicología del comportamiento, la antropología social, la analítica del comportamiento y la lingüística cuando el objetivo es hacer análisis de sentimientos a través de datos de redes sociales. Esos conocimientos sumados a habilidades “suaves” como comunicación, colaboración, liderazgo y pasión por los datos son resaltadas por este autor.

Otro aspecto a ser considerado en la evaluación de restricciones relacionadas con el talento humano es que, tener personas con conocimiento profundo en estadística, matemáticas aplicadas y ciencia de los datos puede tomar mucho más tiempo que crear conciencia sobre el valor de los datos o mejorar el conocimiento en operación de tecnologías (Manyika et al. 2011). Debido a esto, se prevee una carencia de talento humano para iniciativas de big data a nivel global y para todos los sectores (Manyika et al. 2011; World Economic Forum 2012; United Nations 2014a). Adicionalmente,

World Economic Forum (2012) plantea que la gente con esas capacidades podría estar más interesada en usar su conocimiento para sus propias iniciativas de emprendimiento en vez de trabajar para organizaciones públicas o privadas.

2.3.2. Disponibilidad, seguridad, privacidad y calidad de los datos

Los datos valiosos no siempre están al interior de la organización ni son accesibles (TechAmerica Foundation 2012; Biesdorf et al. 2013). Aunque el acceso a diversas fuentes de información es necesario para su uso en nuevas iniciativas de analítica (Buhl et al. 2013), dicho acceso no está garantizado y algunas veces deben ser obtenidos a través de la compra o diferentes tipos de acuerdos (Manyika et al. 2011). Dichos acuerdos no siempre son fáciles de lograr porque no todas las organizaciones están interesadas en compartir. La falta de incentivos o la existencia de temores relacionados con la pérdida de ventaja competitiva podría obstaculizar la decisión de compartir datos (Manyika et al. 2011).

La falta de incentivos para compartir datos también es mencionada por United Nations 2012, como causa de la probable reticencia de algunas organizaciones para compartir datos así como aspectos legales, competitivos o reputacionales. Además de eso, incluso cuando las organizaciones estén interesadas en compartir datos hay algunas barreras como la ubicación de los mismos (p.e. backups que se encuentran en sitios y en un estado que dificulta que sean accedidos o transferidos) (United Nations 2012).

Algunos factores podrían reducir el impacto de esta restricción. TechAmerica Foundation (2012) menciona que muchos proyectos de big data no están basados en información regulada por la protección o la privacidad de datos y no necesitan información personal para generar valor. Además de eso, los datos abiertos del sector público tienen un rol clave para habilitar el aprovechamiento de big data a través de nuevos modelos de negocio y nuevos servicios para clientes y ciudadanos. El creciente número de conjuntos de datos publicados por los gobiernos son una fuente de innovación (Manyika et al. 2011; Lohr 2012; United Nations 2012; World Economic Forum 2012; TechAmerica Foundation 2012; Howard 2012a). La mayoría

de los autores resaltan los datos abiertos del gobierno como impulsores del aprovechamiento de big data; del mismo modo, Laney et al. 2013; Chui et al. 2014 también menciona los datos abiertos del sector privado como un creciente componente del fenómeno de big data.

Con respecto a la calidad de los datos, dado que los datos valiosos para el aprovechamiento de big data pertenecen a diferentes dueños, surge la necesidad de una gobernanza de datos que implica acuerdos sobre políticas de calidad, gestión y responsabilidades sobre los datos (Buhl et al. 2013). La importancia de la gobernanza de los datos también es planteada por Beyer & Laney (2012) y Kart et al. (2013) con respecto a asuntos como la seguridad y la privacidad. En términos de iniciativas que usan datos abiertos, Laney et al. (2013) mencionan las probables deficiencias en la calidad de los datos y la falta de continuidad en la generación de conjuntos de datos del sector público, que pueden afectar la sostenibilidad de procesos y servicios de big data.

2.3.3. Aspectos legales con respecto al aprovechamiento de big data

Generalmente las leyes no son claras sobre quién tiene derecho a recolectar, agregar, divulgar y aprovechar nuevas fuentes de datos tales como datos de ubicación generados por dispositivos móviles (Manyika et al. 2011). Además de eso, la privacidad, los límites sobre la información personal, los derechos de autor y la confiabilidad son temas para ser discutidos (Manyika et al. 2011; United Nations 2014a). El uso de fuentes de datos por parte de diferentes actores es uno de los puntos de discusión con respecto a derechos de autor y confiabilidad de la información producida con una variedad de recursos; una pregunta es quién es responsable cuando inexactitudes en los contenidos de conjuntos de datos producen consecuencias negativas (Manyika et al. 2011).

2.3.4. Aspectos tecnológicos

El principal aspecto tecnológico que puede obstaculizar el aprovechamiento de big data está relacionado con la inmadurez de este mercado lo cual implica carencia de

estándares y documentación sobre buenas prácticas (Beyer and Laney 2012; Kart et al. 2013; Schulte et al. 2013). Esto incluye combinaciones complejas de soluciones comerciales y de código abierto (Newman and Beyer 2012), escasez de soluciones que cumplan con todos los requerimientos y atiendan el amplio espectro de necesidades de analítica en las organizaciones (Schulte et al. 2013) y potenciales complicaciones y errores nuevos (Manyika et al. 2011).

Otras restricciones tecnológicas están relacionadas con intercomparabilidad e interoperabilidad de sistemas (United Nations 2012). Es necesario integrar procesamiento de datos ejecutado a través de internet, con procesamiento en tiempo real y procesamiento fuera de línea (Helbing and Baliatti 2011). Estos asuntos podrían ser más complejos considerando que, de acuerdo con algunas investigaciones, la mayoría de las organizaciones tienen un nivel bajo de madurez en la integración de datos (Buhl et al. 2013).

Con respecto a la infraestructura TI, (Helbing and Baliatti 2011; United Nations 2014a) plantean que se requiere infraestructura robusta para procesamiento intensivo de datos. Esto implica buscar soluciones de big data que se ajusten a las capacidades de cada organización y considerar opciones de infraestructura soportada en la nube (Schulte et al. 2013).

2.3.5. Riesgos relacionados con el aprovechamiento de big data

De acuerdo con la literatura, las principales fuentes de riesgo son la falta de confianza, la incertidumbre acerca del potencial real de big data para permitir el descubrimiento de información secreta, errores en analítica y probables brechas entre aquellos países y organizaciones que están en condiciones de aprovechar big data y los que no.

La falta de confianza es un factor que puede cambiar la situación actual en términos de acceso y uso de información, hay una creciente conciencia de las personas sobre los datos que son recolectados acerca de todos los aspectos de sus vidas y la gente

podría empezar a ser más sensible sobre su privacidad en el mundo electrónico (Manyika et al. 2011; Laney et al. 2013). Adicionalmente, las prácticas de algunas organizaciones, que han sido reveladas recientemente, como las de la Agencia Nacional de Seguridad [National Security Agency] de Estados Unidos y otros escándalos similares contribuyen a esta falta de confianza (Laney et al. 2013). Igualmente, algunos autores presentan sus preocupaciones sobre el potencial mal uso de big data incluyendo actividades criminales (United Nations 2012; Helbing and Balietti 2011).

La novedad de big data crea incertidumbre sobre el potencial real de sus aplicaciones (p.e. algunos conjuntos de datos son condiderados “anonimizados” porque no contienen datos personales pero no hay garantía de que la identidad de las personas no pueda ser revelada a partir de una combinación de varios conjuntos de datos “anonimizados” (Helbing and Balietti 2011; United Nations 2012)). Igualmente, dado que big data puede habilitar el descubrimiento de hechos o relaciones en diversidad de situaciones, no es claro hasta qué punto esto puede incluir, por ejemplo, el descubrimiento de información sensible para la seguridad nacional (Manyika et al. 2011) o la violación de la privacidad a través del uso de datos que no son compartidos por las personas de manera intencional; (i.e. Buhl et al. (2013) mencionan un sistema de entretenimiento doméstico diseñado para reconocer patrones de comportamiento en las parejas y enviar publicidad relacionada con sicoterapia de pareja o planes románticos de acuerdo con la situación actual de la relación).

La incertidumbre relacionada con la falta de confianza y los usos impredecibles de los datos generan un grupo de riesgos que pueden afectar la evolución del aprovechamiento de big data. Algunos supuestos de Gartner plantean que, para el año 2020, las empresas y gobiernos van a fracasar en la protección de información sensible y van a garantizar el acceso público a la misma (Laney et al. 2013).

El mal uso de la analítica, como origen de riesgo en la explotación de big data, no se limita al uso con intenciones criminales sino que también incluye interpretaciones y

decisiones erradas cuando no son considerados algunos elementos (p.e. datos generados de manera activa como el caso de crowdsourcing, blogs, foros en periódicos digitales y redes sociales pueden contener información falsa y pueden ser afectados por contenidos virales (Helbing and Balietti 2011; United Nations 2012)). En el mismo sentido, World Economic Forum (2012) menciona que los datos basados en teléfonos celulares no siempre son confiables para individualizar comportamientos porque algunas veces las SIM card son usadas por varias personas y una persona puede usar más de una SIM card.

La falta de talento humano capacitado podría llevar a errores de análisis surgidos de pasar por alto la diferencia entre percepciones y hechos o la selección inadecuada de muestras (United Nations 2012). Otro factor que podría producir errores de uso de big data es la falsa ilusión de que una enorme cantidad de datos es garantía para controlar sistemas complejos; en ese sentido, Helbing and Balietti (2011) sugieren que la falla en los modelos analíticos usados por los bancos y agencias calificadoras, que eran alimentados con grandes volúmenes de datos, fue una de las causas de la crisis financiera de 2008.

Por último, otra de las fuentes de riesgo mencionadas en la literatura es la brecha que se puede generar entre organizaciones y países capaces de aprovechar big data y aquellos que no tienen esta capacidad. Algunos “big players” podrían ganar a expensas de otras organizaciones (Manyika et al. 2011), en ese sentido, Buhl et al. (2013) menciona posibilidades de “a second wave of digital “colonialization” and domination” [una segunda ola de “colonización” y dominación digital] y United Nations (2012) plantea la probabilidad de que el valor de big data podría concentrarse en algunas pocas compañías (p.e. Facebook, Twitter, Google) a pesar de que esas compañías usan datos producidos por millones de personas y organizaciones alrededor del mundo.

2.4. Big Data y el sector público

Esta sección presenta los potenciales beneficios, restricciones y riesgos relacionados con el aprovechamiento de big data en el sector público. En general y de acuerdo con la literatura, el aprovechamiento de big data en el sector público tiene semejanzas con el sector privado; sin embargo, hay algunas consideraciones específicas que deben tenerse en cuenta para entender y evaluar su uso en el gobierno.

Dado que el sector público lidera la definición de políticas públicas y planes de desarrollo para países y regiones, la evaluación del impacto de big data para este sector no está limitado al nivel organizacional. Más allá de eso, deben ser evaluadas sus implicaciones para el desarrollo social, el crecimiento económico y la gobernabilidad.

2.4.1. Potenciales beneficios del aprovechamiento de big data en sector público

La mayoría de beneficios del uso de big data en el sector público son similares a los del privado y están relacionados con mejor toma de decisiones, mejor segmentación de poblaciones para focalizar acciones, innovación en modelos de negocio, eficiencia y efectividad, mejores servicios para los ciudadanos y reducción de amenazas de seguridad y crimen. Adicionalmente, big data ofrece oportunidades para mejorar en transparencia y participación ciudadana.

Manyika et al. (2011) presenta un índice usado para determinar el potencial de cada sector para crear valor a través de big data. De acuerdo con este índice, el sector público tiene un gran potencial debido a la cantidad de datos, en diversidad de temas, que son recolectados por los gobiernos en todas sus actividades. Más allá de eso, las instituciones que son intensivas en transacciones (p.e. impuestos) tienen mayor potencial.

Mejor toma de decisiones

Este es uno de los aspectos que podría generar beneficios más allá del nivel organizacional (TechAmerica Foundation 2012). De acuerdo con algunas opiniones, big data tiene el potencial de habilitar la existencia de un “humanity’s dashboard” [tablero de control de la humanidad] y contribuir a la lucha contra la pobreza, el crimen y el deterioro ambiental (Lohr 2012). Esas opiniones están fundamentadas en la idea de que este fenómeno puede proveer a los tomadores de decisiones en políticas públicas con evidencia en tiempo real en temas como precios, empleo, desempeño económico, desarrollo social y demografía (United Nations 2014a).

Un efecto de la disponibilidad de datos podría ser el conocimiento de la realidad en tiempo real y la retroalimentación en tiempo real sobre la efectividad de las acciones de gobierno (United Nations 2012). En ese sentido, están siendo utilizados nuevos conceptos incluyendo minería de la realidad y minería de opinión (United Nations 2012). Algunas aplicaciones específicas de estas ideas son los observatorios de crisis para áreas de interés como conflicto o atención de desastres (Helbing and Baliatti 2011), una mejor evaluación y asignación de recursos en programas y proyectos (Manyika et al. 2011; Howard 2012a) y la identificación de “puntos de dolor” en servicios públicos (Accenture 2014).

Con respecto a información para los tomadores de decisiones, el uso de big data para la generación de estadísticas oficiales merece atención especial (United Nations 2012). United Nations (2014a) encontró que la mayoría de instituciones de estadísticas oficiales alrededor del mundo así como organizaciones internacionales están considerando las fuentes de big data como alternativas complementarias para producir estadísticas rápidamente disponibles e indicadores beta. La velocidad, como aspecto de big data, también es considerada para situaciones en las cuales el tiempo es un factor crítico para la toma de decisiones como es el caso de la atención de desastres (United Nations 2012).

El uso de big data para la planeación urbana y la idea de gobierno inteligente (Howard 2012a; MIT 2014) también merece atención especial. Usando la diversidad de datos disponibles en las ciudades, los encargados de planeación urbana podrían estar mejor informados cuando toman decisiones en temas diversos desde el secuenciamiento de semáforos hasta la ubicación de espacios de parqueo (Manyika et al. 2011).

Mejor segmentación de poblaciones para focalizar acciones

Los beneficios potenciales relacionados con este aspecto están basados en la idea de que el gobierno debe saber lo que la gente necesita para servirles de manera efectiva (Howard 2012a). A pesar de algunas preocupaciones relacionadas con el principio de igualdad que debería ser aplicado a los servicios proveídos por los gobiernos, Manyika et al. (2011) plantea que el sector público puede obtener los mismos beneficios que la segmentación de clientes genera para el sector privado incluyendo efectividad, eficiencia y satisfacción del cliente/ciudadano; estos autores mencionan, como ejemplo, el uso de analítica avanzada aplicada a perfiles de pacientes para prevenir riesgos e incluir ciertos perfiles en programas de prevención (Manyika et al. 2011).

Como consecuencia del fenómeno de big data y debido al uso de servicios electrónicos, hay un rastro digital de los ciudadanos que puede crear un círculo virtuoso en el cual el conocimiento sobre los ciudadanos crece y habilita mejoramiento continuo en los servicios (Accenture 2012).

Innovación a través de nuevos modelos de negocio, productos y servicios

En este punto el rol del sector público va más allá del nivel organizacional y llega hasta la definición de visiones y políticas para el desarrollo nacional o local. Buhl et al. (2013) plantean que hay a amplio rango de oportunidades para pequeñas y medianas empresas en el aprovechamiento de big data y los gobiernos deben evaluar las acciones requeridas para promover dichas oportunidades.

De otro lado, United Nations (2014a) menciona la probabilidad de que los sistemas nacionales de estadística enfrenten competencia con la aparición de nuevos actores. Aunque este planteamiento es difícil de entender debido a las atribuciones legales que soportan el trabajo de las instituciones oficiales, es relevante evaluar esta clase de escenarios y sus efectos positivos y negativos.

Eficiencia y efectividad

Este es probablemente el grupo de beneficios más mencionados con respecto al aprovechamiento de big data en el sector público. Manyika et al. (2011) estiman que la administración pública podría tener enormes ahorros en eficiencia operativa (sin incluir beneficios potenciales relacionados con reducción de fraudes y errores). Varios autores coinciden en el enorme potencial de reducción de costos a partir de eficiencia operativa (Manyika et al. 2011; TechAmerica Foundation 2012; Kart et al. 2013; MIT 2014).

Dicha reducción de costos pueden ser concretadas con el uso de nuevas fuentes de datos para medir y comparar el desempeño de diferentes unidades dentro de una entidad para revelar brechas entre unidades o contratistas y situaciones de ineficiencia (Howard 2012a). Dado que el descubrimiento de elementos ocultos es una oportunidad clave generada por la analítica, big data podría ser usada para revelar variabilidad en el desempeño de entidades, empleados, proveedores y contratistas (Manyika et al. 2011; TechAmerica Foundation 2012). En ese sentido, la reducción de costos podría generarse en el monitoreo y evaluación de logro de objetivos, gestión de recursos humanos, compras y contrataciones, entre otros aspectos administrativos.

Otra fuente ahorro está relacionada con la reducción del tiempo de búsqueda y procesamiento de información que se podría habilitar por el relacionamiento de datos que actualmente están aislados en diferentes áreas o entidades (Manyika et al. 2011; Howard 2012a). Igualmente, la analítica puede contribuir a la identificación de errores e inconsistencias (Manyika et al. 2011).

Mejores servicios para los ciudadanos

Más allá de las mejoras en servicio que podría ser generadas por una mejor segmentación de ciudadanos, hay un amplio rango de oportunidades para mejorar la calidad de los servicios reduciendo costos y tiempos (Manyika et al. 2011). Más allá de eso, es posible la creación de nuevos servicios, por ejemplo en el sector salud, los gobiernos pueden ofrecer información sobre el desempeño de los proveedores de servicios de salud para permitir a los ciudadanos seleccionar el suyo a partir de dicha información (Manyika et al. 2011); otro ejemplo de un posible nuevo servicio es la oferta de información que ayude a los desempleados a encontrar un empleo (TechAmerica Foundation 2012).

Reducción de amenazas de seguridad y crimen

El uso de big data en el gobierno puede facilitar la identificación de inconsistencias, errores y fraudes en impuestos, sistema de salud y programas de bienestar social (Manyika et al. 2011; TechAmerica Foundation 2012; Accenture 2014). Con respecto a la seguridad, el uso de diferentes fuentes de datos incluyendo ubicación de personas, vehículos y lugares, direcciones ip así como información psicológica, política y cultural podría ayudar a identificar riesgos, motivos y organizaciones involucradas en terrorismo o ciberataques (Manyika et al. 2011; TechAmerica Foundation 2012). Adicionalmente, el análisis de big data podría contribuir en estudios forenses y protección de infraestructura crítica (TechAmerica Foundation 2012), aplicación de la ley (MIT 2014) y prevención de desastres (United Nations 2012; Helbing and Baliatti 2011).

Mejora en transparencia y participación ciudadana

El acceso público a más y mejor información es una fuente de beneficios favorecidos por el fenómeno de big data y pueden participación ciudadana de mejor calidad y enriquecer las discusiones sobre políticas públicas a través del análisis de datos del sector público y privado (Manyika et al. 2011). Este escenario y las oportunidades para revelar la variabilidad en el desempeño de unidades, empleados, proveedores y

contratistas podría mejorar la rendición de cuentas y la transparencia (TechAmerica Foundation 2012) así como la confianza en el gobierno (Manyika et al. 2011).

Del mismo modo, el uso de iniciativas de crowdsourcing presenta beneficios potenciales en términos de transparencia y participación ciudadana. De acuerdo con MIT (2014), dichas iniciativas podrían habilitar la recolección y procesamiento de datos de regiones con población vulnerable. Igualmente, aplicaciones móviles para teléfonos celulares y otros dispositivos móviles, podrían ser usadas por voluntarios para transmitir información en temas como ambiente o tráfico. El Anexo 1 presenta un resumen de los posibles usos de big data más mencionados en la literatura y sus potenciales beneficios.

2.4.2. Restricciones y riesgos de uso de big data en el sector público

La mayoría de restricciones y riesgos que presenta big data para el sector público son similares a los que enfrenta el sector privado y están relacionados con escasez de talento, disponibilidad y calidad de los datos, marco legal y asuntos tecnológicos. Con excepción de los aspectos tecnológicos, estas restricciones requieren consideraciones especiales en el ámbito público. Además de esto, algunos autores mencionan los aspectos financieros como otra restricción que involucra escasez de recursos, falta de evidencia para evaluar costo-beneficio (United Nations 2014a) y el hecho de que existan otras prioridades presupuestales (Fister 2013).

El efecto de todos estos riesgos y restricciones parece ser significativo; de acuerdo con Fister (2013) hay una brecha entre la percepción de los potenciales beneficios de big data y el número de proyectos e inversiones en big data que en realidad se ejecutan en las agencias federales de Estados Unidos. Esta sección presenta algunos elementos de comprensión de las restricciones y riesgos relacionados con big data en el gobierno.

Recursos humanos y talento

La carencia de talento humano con capacidades para el aprovechamiento de big data es presentado por varios autores como una restricción clave en el sector público, de acuerdo con Fister (2013) solo el 35% de las agencias de gobierno de Estados Unidos dicen tener las personas para dirigir proyectos de big data.

Otras capacidades requeridas están relacionadas con comunicación y gestión de actores de interés (Fister 2013). Esas capacidades se requieren debido al hecho de que algunas iniciativas de big data implican el uso de datos de varias entidades del sector público y privado (Howard 2012a) así como participación ciudadana (MIT 2014).

Aparte de las capacidades, aspectos culturales y actitudinales podrían ser barreras, Manyika et al. (2011) y Fister (2013) mencionan que el sector público tiene menor presión competitiva, y por eso, posiblemente menos motivación para generar valor con big data.

Con respecto al uso de big data para estadísticas oficiales, United Nations (2014a) plantea que la experiencia de las instituciones de estadística alrededor del mundo está basada en el conocimiento clásico de la estadística el cual no es totalmente pertinente para el uso de nuevas fuentes de datos.

Disponibilidad, seguridad, privacidad y calidad de los datos

Con respecto al acceso a información, algunos autores consideran que el sector público tiene una ventaja debido a la abundancia de registros administrativos en una variedad de temas (Manyika et al. 2011; Helbing and Balietti 2011; Accenture 2014). Sin embargo, el compartir datos entre diferentes entidades (MIT 2014) podría ser obstaculizado por situaciones en las cuales los conjuntos de datos relacionados con un asunto específico (p.e. licitaciones, permisos de construcción o tráfico y clima para la planeación urbana) están fragmentados a lo largo de diferentes fuentes y su integración, depuración y consolidación demanda grandes esfuerzos (Laney et al.

2013). De acuerdo con eso, la falta de interoperabilidad, estándares y acuerdos para trabajo intra e intergubernamental son restricciones que afectan el acceso a los datos (Accenture 2014).

Algunos otros factores pueden obstaculizar el acceso a información en el sector público, por ejemplo, posibles conflictos de interés. (Manyika et al. 2011) presenta una situación, relacionada con el sector salud, en la cual las entidades públicas responsables de pagos ganarían usando datos clínicos (reducción en el monto de los pagos). Sin embargo, los proveedores de servicios de salud, quienes deberían garantizar el acceso a los datos clínicos; tendrían algunos efectos negativos (menor facturación). En situaciones como esta, sería difícil obtener acceso a los datos porque el dueño no está interesado en el resultado esperado.

Con respecto a la calidad, (Laney et al. 2013) plantean que, como el gobierno publica sobre todo datos para ser usados por otros como un asunto de cumplimiento; las entidades no están comprometidas con el control, precisión, completitud y oportunidad. La fácil visualización es otro requerimiento para el aprovechamiento de big data y la mayoría de la información del sector público no cumple esta condición (MIT 2014).

Más allá de la calidad de los datos, en el caso de las estadísticas oficiales que se podrían producir usando big data, hay restricciones relacionadas con la confiabilidad y credibilidad en el uso de big data. United Nations (2014a) resalta que se requiere más investigación sobre metodologías nuevas o actualizadas para generación de estadísticas en el contexto de big data. Los temas de investigación incluyen comparabilidad, continuidad y coherencia que son dimensiones de la calidad en estadísticas y aspectos como la representatividad, volatilidad y continuidad de nuevas fuentes de datos pueden afectar dichas dimensiones.

Aspectos legales relacionados con el aprovechamiento de big data

Es importante considerar el rol de los gobiernos en la regulación para el uso de big data, la privacidad y la protección de datos. Buhl et al. (2013) analizan las aparentes desventajas que enfrentan las compañías establecidas en Europa y Asia que son más temerosas de tomar la iniciativa en big data porque enfrentan regulaciones más estrictas que las empresas establecidas en Estados Unidos. Adicionalmente, United Nations (2014a) plantean que la legislación diseñada para permitir el uso de big data para estadísticas oficiales no ha sido promulgada todavía en la mayoría de los países.

Con respecto al uso de datos del sector privado para propósitos oficiales, es importante considerar dos líneas de pensamiento que son mencionadas en la literatura; la primera es llamada “filantropía de los datos” y está orientada a tener acceso a datos “anonimizados” del sector privado para promover la innovación social y los avances relacionados con el el desarrollo social (United Nations 2012). La segunda línea es un campo de estudio emergente llamado infonomics y está orientado a determinar un valor económico para cada pieza de información (Laney et al. 2013).

Riesgos relacionados con el aprovechamiento de big data

La incertidumbre sobre el uso de datos afecta la confianza pública (United Nations 2014a) y es la fuente de riesgos relacionados con la ciberseguridad que deben ser manejados por cada país (MIT 2014). Sin embargo, MIT (2014) menciona un estudio desarrollado en 35 países cuya conclusión es que ningún país está preparado en términos de ciber seguridad [“No country is cyber ready”]. Por lo tanto, la ciberseguridad es un aspecto a ser evaluado con respecto a volumen, velocidad, variedad y valor de los datos en el contexto de este fenómeno. Una fuente de riesgo en el sector público es el uso malintencionado de los datos. Helbing and Baliatti (2011) presentan una situación en la cual un departamento de policía liberó información acerca de sus estadísticas operativas haciendo posible que grupos criminales infieran tácticas policiales, horarios de patrullaje y otra información valiosa.

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

Como resumen de esta revisión de literatura es posible decir que hay una comprensión común de big data como un fenómeno que puede ser aprovechado a través de la generación de valor tanto en el sector público como en el privado. Sin embargo, hay algunas restricciones y riesgos que deben ser manejados para favorecer el éxito de iniciativas de big data y el gobierno debe analizar su rol no solo para potencializar la explotación de big data en el nivel organizacional sino también a nivel social y económico.

3. Condiciones ideales para aprovechamiento de big data

Esta sección presenta las condiciones ideales para el aprovechamiento de big data en el sector público colombiano, dichas condiciones fueron identificadas en la revisión de literatura y se presentan usando la lista ITPOSMO (Heeks 2003) que abarca siete dimensiones: Información, tecnología, procesos, personal/talento, sistemas/estructuras de gestión y otros aspectos.

3.1. Información

Con respecto a la información, el aprovechamiento de big data requiere abundancia y disponibilidad de datos con especial énfasis en los aspectos de variedad y velocidad, que tienen mayor potencial que solo volumen (Beyer and Laney 2012).

- En términos de variedad, las condiciones ideales están relacionadas con la abundancia de datos del sector público y privado incluyendo registros administrativos, datos abiertos y transacciones.
- La existencia de nuevas fuentes de datos incluyendo búsquedas de internet, interacción en redes sociales, transacciones y servicios en línea, también contribuye al aspecto de variedad en big data.
- Con respecto a la velocidad, el uso de teléfonos móviles requiere especial atención porque los datos de ubicación personal tienen un gran potencial para generar valor (Manyika et al. 2011; World Economic Forum 2012) y son mencionados frecuentemente como un recurso clave en muchos casos de aprovechamiento de big data. Además de eso, el uso de RFID, satélites, sensores, conexiones máquina a máquina (M2M) y circuitos cerrados de televisión también contribuyen a la abundancia de datos que involucran el aspecto de velocidad.

3.2. Tecnologías

Con respecto a tecnologías, hay cuatro condiciones ideales a ser consideradas:

- La existencia de infraestructura para analítica avanzada hace posible combinar nuevas fuentes de datos internas y externas (Brown et al. 2013). Esto implica clusters de computadores capaces de analizar enormes cantidades de datos (Helbing and Baliaetti 2011),.
- La disponibilidad de conexiones de internet que permitan la transmisión de variedad de datos entre diferentes dispositivos incluyendo móviles y sensores (Helbing and Baliaetti 2011). Esto implica una amplia cobertura territorial de conexión a internet de banda ancha.
- La interoperabilidad es requerida para facilitar el intercambio de datos entre entidades públicas, de acuerdo con (Howard 2012a) es necesario implementar principios y prácticas de arquitectura empresarial y nuevos métodos para compartir datos.
- Ciberseguridad, la necesidad de proteger los datos de la nación toma mayor importancia debido al creciente número de servicios y actividades en línea (MIT 2014) así como la incertidumbre sobre el potencial mal uso de big data mencionado en la revisión de literatura.

3.3. Procesos

Con respecto a los procesos, el grupo de procesos más relevante está relacionado con el compartir datos que puede llevar a un uso exitoso de big data en el sector público:

- Procesos establecidos para compartir datos entre diferentes áreas y diferentes niveles de gobierno, así como entre el sector privado, el gobierno y los ciudadanos. En estos procesos, los componentes de aseguramiento de la calidad de los datos y la visualización de los datos tienen gran relevancia.

3.4. Objetivos

Con respecto a los objetivos, se requiere coherencia entre objetivos planteados en los planes y programas del gobierno colombiano y los beneficios que pueden ser generados con el aprovechamiento de big data. Por lo tanto, hay dos aspectos principales a ser considerados:

- La existencia de planes y programas alineados con la producción de volumen, variedad y velocidad de datos como aspectos generadores del fenómeno de big data.
- La existencia de planes y programas alineados con el aprovechamiento del fenómeno de big data en términos de analítica, visualización y toma de decisiones basada en datos.

3.5. Personal/Talento

La disponibilidad de talento humano, con conocimiento y actitudes específicas, es identificada por muchos autores como una condición clave para el aprovechamiento de big data, de acuerdo dichos autores, es posible identificar tres niveles de caracterización de este requerimiento:

- Para la operación de datos se requieren conocimiento en desarrollo, implementación y mantenimiento de herramientas de software y hardware usadas para big data (Manyika et al. 2011) incluyendo la gestión de clusters para manejo de volumen, velocidad y variedad de datos (Beyer and Laney 2012) así como bodegas de datos lógicas y herramientas como Hadoop.
- Para el análisis de datos se requiere conocimiento en estadística, aprendizaje de máquinas y ciencia de los datos; por ejemplo se necesitan personas que manejen la correlación como concepto estadístico usado para análisis predictivo (MayerSchönberger and Cukier 2013) y que tengan conocimientos actualizados que van más allá de técnicas de estadística clásica para asegurar un adecuado tratamiento de nuevas tipologías de datos (United Nations 2014a). Además de eso, dichas personas deben tener conocimiento en los temas misionales de cada entidad y una mentalidad orientada hacia los datos que es mencionada constantemente en la literatura.

- Para el uso de datos, se requiere también mentalidad orientada hacia los datos en los niveles de toma de decisiones lo cual implica conciencia del valor de los datos como un activo organizacional valioso (Beyer and Laney 2012; Laney et al. 2013) y curiosidad intelectual orientada a cerrar los vacíos de información con el uso de herramientas innovadoras (TechAmerica Foundation 2012; Beyer and Laney 2012).

3.6. Sistemas y estructuras de gestión

La literatura menciona dos aspectos esenciales de los sistemas y estructuras de gestión para el aprovechamiento exitoso de big data; la primera es la necesidad de sistemas de gobernanza y estrategia de los datos, la segunda un ambiente institucional que favorezca la innovación. De acuerdo con eso, es posible definir tres condiciones ideales:

- Existencia de áreas con responsabilidad sobre la estrategia de los datos y la gobernanza de los datos (p.e. Chief Information Officer-CIO, Chief Data Officer-CDO, estrategia de la información)(Laney et al. 2013).
- Existencia de áreas y roles que promuevan y lideren la innovación (Laney et al. 2013).
- Mecanismos sólidos de coordinación entre la estrategia de los datos, gobernanza de los datos, innovación y áreas misionales para asegurar que las iniciativas de big data contribuyan al cumplimiento de objetivos estratégicos (Beyer and Laney 2012).

3.7. Otros recursos

Otras condiciones clave para el aprovechamiento de big data fueron presentados en la revisión de literatura y tienen impacto más allá del nivel organizacional porque habilitan el uso de big data como fuente de valor para diferentes actores de interés incluyendo el sector privado y los ciudadanos. Dichas condiciones son:

- Existencia de marcos legales para la explotación de big data incluyendo asuntos relacionados con la privacidad de los datos, la protección de los datos, los derechos de autor y el acceso a los datos públicos.

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

- Existencia de alguna concepción de ecosistema de datos que habilite el compartir datos con incentivos para el sector privado y los individuos.

4. Realidad y brechas

Esta sección presenta algunos hallazgos de la investigación con respecto a las condiciones reales y las brechas entre la realidad y las condiciones ideales para el aprovechamiento de big data. Esta evaluación usa de nuevo la lista ITOSMO (Heeks 2003) con siete dimensiones: Información, tecnología, procesos, personal/talento, sistemas/estructuras de gestión y otros aspectos.

4.1. Realidad

4.1.1. Información

El uso actual de TIC en el sector público y privado en Colombia sugiere cierta abundancia de datos que no pueden ser evaluados de manera precisa porque no hay métodos estándares para realizar dicha evaluación. Sin embargo, usando la información disponible es posible resumir que hay condiciones favorables y desfavorables en términos de volumen, variedad y velocidad de datos. Las condiciones favorables están relacionadas con el creciente uso de TIC en gobierno electrónico, así como en comercio electrónico; de igual manera la tasa de penetración de internet, el creciente uso de redes sociales y servicios electrónicos, la tasa de penetración de telefonía móvil y el uso de teléfonos inteligentes en Bogotá y otras regiones. Con respecto a las condiciones desfavorables la disponibilidad de datos abiertos del sector público merece atención especial, el catálogo de conjunto de datos disponibles en el portal de datos abiertos fue evaluado, en el mes de agosto de 2014, usando la herramienta de búsqueda y clasificación que ofrece el portal (MINTIC 2014d). Una probable barrera para la generación de volumen, velocidad y variedad a través de datos abiertos podría ser algo de miedo a la presión del control social que experimentan algunos funcionarios públicos. Esto surge de la entrevista con Tamayo (2014) quien menciona alguna reticencia, en algunas entidades, para abrir algunos datos porque ellos prefieren esperar hasta que dichos datos presenten mejores indicadores. Esta podría ser una explicación parcial para los siguientes hallazgos:

- El volumen no fue evaluado en términos de cantidad de datos sino en número de conjuntos de datos; 620 publicados por todo el sector público colombiano que está conformado por más de 1300 entidades (CIO Summit 2014), podría considerarse bajo.

- La velocidad de datos es baja, la mayoría de los conjuntos de datos no son dinámicos (p.e. con respecto a medio ambiente regarding hay un conjunto de datos que lista 25 estaciones de monitoreo hidrometeorológico; sin embargo, no hay conjuntos de datos con datos recolectados por dichas estaciones y esn esta categoría solo 5 conjuntos de datos, de un total de 25, fueron publicados durante 2014). La falta de continuidad en conjuntos de datos es otra característica observable en este catálogo.
- La variedad de conjuntos de datos es escasa, hay una alta concentración de datos administrativos básicos incuyendo directorios de dependencias e indicaciones para los ciudadanos con respecto a procedimientos; adicionalmente, también se observó escacez de conjuntos de datos en categorías clave como trabajo y empleo (1 conjunto de datos) y estadísticas (2 conjuntos de datos). Se identificaron algunos elementos positivos con respecto a variedad: número de categorías disponibles, presencia de datos en diferentes formatos como video, mapas y algunos conjuntos de datos que parecen ser de gran interés para los usuarios (p.e. la infraestructura espacial de datos de Bogotá que muestra más de 4000 visitas).

Pasando ahora al potencial de uso de big data a través de nuevas fuentes de datos, la relativamente baja tasa de penetración de internet es reconocida por MinHacienda (2013) como una barrera para el análisis predictivo; sin embargo, este Ministerio muestra una experiencia exitosa, basada en el uso de Google Trends. En este caso la predominancia de este motor de búsqueda, que al final de 2012 era usado para el 96.1% de las búsquedas de internet en Colombia, permitió el desarrollo de dicho análisis predictivo. En otro caso reportado por El Espectador (2014a), la Policía Nacional usa videocámaras para reconocer las matrículas de los vehículos que pasan por una vía, después de eso, se verifica instantáneamente el cumplimiento y vigencia de requisitos como SOAT, pagos de comparendos e impuestos lo cual incrementa la eficiencia operativa de la Policía que puede detener a aquellos vehículos que según los datos no están cumpliendo con las regulaciones. Otro caso relevante es el proyecto “Tu Bogota” que usa sensores para proveer información actualizada acerca de 13

dimensiones de la vida en Bogotá (Tamayo 2014). Estos tres casos sugieren que actualmente ya existen fuentes de datos que hacen posibles algunas iniciativas de big data.

4.1.2. Tecnologías

La existencia o acceso a infraestructura varía entre diferentes sectores, tamaños de entidades y niveles de gobierno; esto fue mencionado por la Viceministra de TI durante la CIO Summit 2014 y también por Tamayo (2014) de la Infraestructura de Datos Espaciales de Bogotá-IDECA durante la entrevista para esta investigación. Tamayo (2014) mencionó que ellos están consolidando datos de más de 90 entidades del gobierno de la ciudad. Para tener acceso a datos de algunas de esas entidades es posible crear conexiones directas a conjuntos de datos, sin embargo, en otros casos es necesario recolectar dicha información en cada entidad de manera física.

Hay dos elementos favorables en términos de tecnología para aprovechamiento de big data, el primero es la existencia de un marco de arquitectura empresarial que promueve estándares, buenas prácticas e interoperabilidad para todo el sector público. Uno de los pilares de este marco, en términos de compras y contratación, es la preferencia por soluciones en la nube en vez de infraestructura propia (CIO Summit 2014; Bohorquez 2014). En adición a dicha preferencia, los acuerdos con el programa de Compra Eficiente (CIO Summit 2014; Bohorquez 2014) sugieren que, dejando a un lado el presupuesto, la infraestructura tecnológica para big data estaría disponible a través de servicios en la nube que serían ofrecidos bajo acuerdos de precios aplicable a todo el sector público y estandarización en tecnología, calidad y variedad de proveedores (CIO Summit 2014; Bohorquez 2014).

Con respecto a la disponibilidad de conexiones de internet de alta velocidad, cifras de World Economic Forum (2014) muestran la necesidad de que Colombia avance más en este aspecto, sin embargo, la creciente presencia de conexiones 4G podría ser

considerada como un aspecto positivo teniendo en cuenta que está creciendo la infraestructura para estas conexiones (EL TIEMPO 2014b).

La situación actual con respecto a la ciberseguridad fue presentada durante el CIO Summit 2014; es evidente el reconocimiento de la ciberseguridad como un asunto complejo que requiere involucrar diversos grupos de actores de interés y la consideración de aspectos institucionales, legales, tecnológicos y culturales.

4.1.3. Procesos

Los procesos para compartir y asegurar la calidad de los datos entre las instituciones públicas son promovidos y apoyados por MINTIC en varios documentos. MINTIC (2012a) define una ruta para las entidades públicas a través de etapas de maduración de gobierno electrónico y actividades obligatorias para crear interoperabilidad, creación de ventanillas únicas y publicación de conjuntos de datos abiertos en cada entidad pública. Igualmente, MINTIC (2012b) define lineamientos para la implementación de datos abiertos y MINTIC (2014e) ofrece una plataforma web y herramientas para la publicación estandarizada de conjuntos de datos. No hay ningún lineamiento con respecto a visualización.

4.1.4. Objetivos

El contenido del Plan Vive Digital 2014-2018, muestra que su ejecución puede contribuir al volumen, velocidad y variedad de datos y por lo tanto, reforzar el avance de big data como fenómeno en Colombia. En términos de infraestructura y servicios, hay metas relacionadas con banda ancha, conexiones 4G, Wi-fi en sitios públicos de más de 700 municipios e internet en zonas rurales. También las metas relacionadas con desarrollo de aplicaciones, servicios electrónicos del gobierno, apoyo al emprendimiento TIC, fortalecimiento de la industria TI e iniciativas de ciudades inteligentes pueden contribuir al fortalecimiento del fenómeno de big data.

En el plan hay un componente orientado a promover el talento humano con líneas de acción para el gobierno, la industria TI y la sociedad en general. Para el gobierno, la capacitación y el fortalecimiento del rol del CIO en cada institución pública y el trabajo conjunto con universidades son algunos de los más relevantes. Para la industria TI, hay una meta relacionada con el desarrollo de científicos de los datos en el mediano plazo, sin embargo, no se menciona nada en relación con la generación de talento con habilidades específicas para analítica en el corto plazo. La inclusión de TI en el curriculum para educación primaria, secundaria y superior, así como las recomendaciones para el mejoramiento de la enseñanza de matemáticas e idiomas son metas que podrían aportar al fortalecimiento del talento a largo plazo. De otro lado las metas relacionadas con alfabetización digital, teletrabajo, uso seguro de internet y soluciones TIC para personas con discapacidad también podrían contribuir al desarrollo del fenómeno de big data porque implican que cada vez más personas estarán usando más servicios electrónicos que generan datos al ser usados.

Aunque Vive Digital 2014-2018 contribuye al fortalecimiento de big data, como fenómeno, el plan no muestra evidencia clara de la intención de aprovechar dicho fenómeno. Los datos no están contemplados como componente en el ecosistema digital (a pesar de que los datos son inherentes a todos los ejes del ecosistema); una meta en el componente de aplicaciones y subcomponente de gobierno en línea es consolidar el ecosistema de datos abiertos, sin embargo, este ecosistema parece estar enfocado en la apertura de datos del sector público y no parece contemplar el valor relacionado con volumen, variedad y velocidad de datos producidos por diferentes fuentes incluyendo el sector privado y las personas (MINTIC 2012b).

4.1.5. Personal/Talento

No se encontró información específica sobre las capacidades actuales para la operación de herramientas de big data en Colombia (p.e. capacidades para operar Hadoop). Tamayo (2014) no reporta ninguna dificultad para contar con profesionales en los niveles operativos del equipo de big data de IDECA, sin embargo no se puede concluir que esta sea la situación para el sector público colombiano.

Con respecto a capacidades de análisis de datos, es importante resaltar la existencia de un caso relevante de análisis predictivo basado en el aprovechamiento de big data en MinHacienda (2013). De acuerdo con la descripción ofrecida por los autores, este es la primera aplicación en Colombia de una metodología propuesta por Choi and Varian (2012) para predicciones económicas de corto plazo. Esta metodología está basada en el uso de la herramienta estadística del motor de búsqueda de Google conocida como "Google Trends" la cual provee reportes semanales sobre el volumen de términos de búsqueda en tiempo real. Tomando estos datos como entrada, se constuyeron varios indicadores anticipados para analizar tendencias de corto plazo para nueve sectores del PIB. La descripción de este caso, elaborada por los profesionales involucrados en el proyecto, muestra un análisis riguroso de una variedad de elementos como investigaciones previas, metodología propuesta, validez de las correlaciones, escasez de literatura para el uso de Google Trends en Latinoamérica, potencial predictivo de Google Trends en el contexto colombiano, inexistencia de casos de uso similares en Colombia y el desempeño prescriptivo de los indicadores creados (MinHacienda 2013).

La descripción de este caso, sugiere conocimiento sólido y actualizado en economía, estadística y nuevas metodologías para análisis predictivo; es relevante notar que esta iniciativa fue ejecutada en el Dirección General de Políticas Macroeconómicas del Ministerio, un área en la cual es previsible encontrar profesionales de muy alto nivel en términos de formación y experiencia. Sin embargo, la situación podría ser diferente para el resto del sector público, Tamayo (2014) menciona brechas de conocimiento significativas entre las 90 entidades del gobierno de Bogotá; adicionalmente, las tendencias en educación y mercado laboral en Colombia podrían influenciar estas brechas.

Los hallazgos sugieren escasez de profesionales con conocimiento en matemáticas, estadística, aprendizaje de máquinas y ciencia de los datos. Hay dos aspectos que podrían estar relacionados con dicha; el primero es la oferta de programas en esas áreas, una búsqueda en el Sistema de Información sobre Educación Superior reporta 24 programas de postgrado sobre estadística aplicada, 39 sobre matemáticas aplicadas

y ninguno en analítica o ciencia de los datos (Ministerio de Educación 2014); además de eso, Colombia está ubicado en la posición 108 entre 148 países en la calidad de educación en ciencias y matemáticas (World Economic Forum 2014). El segundo aspecto a ser considerado es la demanda por esos programas, Business Software Alliance (2011) menciona que de acuerdo con UNESCO, el número de estudiantes de educación superior en Colombia está aumentando; sin embargo, el número de estudiantes en programas relacionados con ciencias ha estado disminuyendo durante los últimos años. Adicionalmente, el desempeño de los jóvenes colombianos en las pruebas PISA 2012 muestra que 73% de ellos tienen bajo desempeño en matemáticas y el puntaje en ciencias es también bajo en comparación con el puntaje promedio de los países de la OECD (OECD 2012). Todo esto crea una alerta sobre el alistamiento de los profesionales actuales y futuros para entrar en los campos de estudio requeridos para el aprovechamiento de big data; además de eso, estimaciones de (IADB 2014) muestran que, de acuerdo con las tendencias actuales, tomaría al menos 21 años que los países latinoamericanos logren los puntajes promedio actuales de los países de la OECD en matemáticas.

De acuerdo con la evidencia mencionada, no es probable un cambio significativo con respecto a la disponibilidad de dichos profesionales en el corto, mediano o largo plazo si no hay intervenciones específicas.

Pasando a las capacidades para el uso de datos y la mentalidad orientada hacia los datos; el reconocimiento del valor de los datos es evidente en muchas de las piezas de comunicación y los documentos técnicos de MINTIC y también es evidente el interés por diseminar dicha conciencia. Una evidencia específica es el documento de lineamientos para datos abiertos en el cual son explicados de manera clara los beneficios sociales, políticos y económicos para el sector público y privado (MINTIC 2012b). MINTIC también muestra su interés en este tema como organizador de uno de los eventos observados en esta investigación, 5 conferencias enfocadas en big data y sus oportunidades fueron desarrolladas en Colombia 3.0, en una de ellas el Vicepresidente de Internacionalización de Facebook presentó una sesión sobre el uso de analítica.

Más allá de esta conciencia acerca del valor de los datos en MINTIC, es posible identificar algún nivel de mentalidad orientada hacia los datos en varios funcionarios públicos observados o entrevistados en esta investigación. En caspo mencionado sobre uso de Google Trends en el Ministerio de Hacienda, ellos muestran una motivación generada por la necesidad de reducir atrasos en la información y cierto entendimiento de la creciente relevancia de la información en tiempo real producida por el uso masivo de TIC (MinHacienda 2013). Con respecto al proyecto “Tu Bogota”, Tamayo (2014) reporta que los datos consolidados fueron usados para la definición del POT y será usada para avanzar en asuntos prioritarios de movilidad y transporte. Otro caso en el cual es posible inferir que hay conciencia sobre los datos, como un activo valios para la toma de decisiones fue observada durante el evento CIO Summit a través de un comentario de un miembro de la Policía Nacional con respecto a su trabajo en ciberseguridad; ellos están creando herramientas analíticas para identificar patrones de ciberataques para mejorar la prevención y la respuesta a los mismos. La misma conciencia es evidente en un artículo de prensa que habla de la recuperación de más de 3000 historias clínicas de recién nacidos incluidos en el Plan Canguro desde 1978. En este artículo, un Subdirector del Archivo General de la Nación, sugiere que esas historias clínicas no son tan valiosas si se miran como documentos individuales pero podrían ser muy valiosas si son tratadas como un conjunto de datos (EL TIEMPO 2014a); con este comentario este funcionario también muestra cierta conciencia sobre el concepto de “dark data” que es mencionado por Beyer and Laney (2012) como datos que son conservados con el único propósito de cumplir con reglamentación pero que no generan valor agregado. Se requiere mayor investigación para determinar si la conciencia sobre el valor de los datos está presente en la mayoría de las entidades públicas y en los niveles directivos, sin embargo, es razonable pensar que la diseminación de estas ideas en los distintos escenarios mencionados está contribuyendo a que la mentalidad orientada hacia los datos se fortalezca.

Con respecto a las competencias de los líderes de TI del sector público, MINTIC (2012b) plantea que estos líderes están más enfocados en asuntos de infraestructura que en la generación de valor. Esta es la motivación de un programa orientado a

empoderar a un grupo de ellos, aunque el número de participantes aún es pequeño los resultados de este programa reporta alrededor de 30 iniciativas viables de generación de valor público (Universidad Nacional 2014).

También hay evidencia de la mentalidad orientada hacia los datos en la industria TI, Senseta es una compañía de servicios de big data mencionada, por el Ministerio de Industria y Comercio (2014), en un grupo de historias de innovación, esta historia de emprendimiento muestra el reconocimiento de las oportunidades de la analítica y el uso de nuevas fuentes de datos incluyendo los vehículos autónomos no tripulados (UAV). Aunque la existencia de este caso podría ser considerado positivo, se requiere mayor investigación para evaluar este aspecto en más de 1800 compañías de software y servicios TI que, según lo mencionado en Colombia 3.0, existen en Colombia.

4.1.6. Sistemas y estructuras de gestión

La Viceministra de TI presentó durante la CIO Summit algunos de sus resultados hasta ahora, incluidos los mencionados sobre seguridad y el marco de arquitectura empresarial. Aunque hay metas y acciones relacionadas con el empoderamiento de los CIO la Viceministra reconoce las dificultades que afrontan algunos líderes de TI quienes se ven forzados a desempeñar varios roles en sus entidades.

Con respecto al marco de arquitectura empresarial (MINTIC 2014c), es importante considerar que fue lanzado durante Julio de 2014, por lo tanto, no está implementado aún en las entidades. Este documento provee lineamientos generales para generar gobernanza de TI en términos de análisis estratégico, información, procesos y adopción de estándares lo cual implica trabajo de mediano y largo plazo en cada entidad. Debido a su contenido general, no hay ninguna mención de prácticas específicas con respecto a big data en el marco de arquitectura empresarial.

Con respecto a la existencia de áreas que lideren y promuevan la innovación, fueron identificados tres actores principales: la Coordinación de Diseño e Innovación, los nodos de innovación y el Centro de Innovación en Electrónico. La existencia de esas tres estructuras organizacionales, así como otras iniciativas en diferentes niveles de gobierno como el grupo de investigación sobre big data del gobierno de Bogotá mencionado por Tamayo (2014) y la convocatoria para la creación de un Centro de Excelencia en Big Data, sugieren que la innovación y las generación de capacidades institucionales son prioridades. Sin embargo, la documentación disponible no muestra una definición clara de las responsabilidades y el alcance de cada una de las áreas mencionadas.

Pasando a la coordinación requerida entre la estrategia de datos, la gobernanza de los datos, la innovación y las áreas misionales para lograr objetivos estratégicos con el uso de big data, la construcción colectiva y la innovación está definidos por ley como principios de la estrategia de gobierno electrónico (MINTIC 2012a). De acuerdo con el decreto vigente, MINTIC lidera la estrategia y las demás entidades deben cumplir actividades y metas. A pesar de que MINTIC tiene roles de liderazgo y apoyo, la estrategia es soportada principalmente por la ley que obliga la inclusión de la estrategia de gobierno en línea de cada entidad en el proceso de planeación que monitoreado por los organismos de control.

El monitoreo de los avances de la estrategia en cada entidad, que es ejecutado por los organismos de control es un buen propulsor de logros. Sin embargo, este modelo basado en el monitoreo del cumplimiento, que según el estudio sobre gobernabilidad en Colombia desarrollado por OECD (2013) es el centro de la gestión de recursos humanos en el sector público colombiano, podría promover la planeación de resultados fácilmente alcanzables menos complejos en vez de aquellos con mayor potencial de generación de beneficios para el desempeño organizacional o impacto social. Esta opinión resulta relevante considerando algunos comentarios de Guzman (2014) durante la CIO Summit con respecto a los avances del Índice Nacional de Gobierno Electrónico; él recomendó ver este índice como una herramienta para el

mejoramiento en vez de una evaluación externa impuesta; también recomendó reportar información real y consistente evitando mostrar resultados superiores a los reales. Esto parece ser coherente con una situación de puntajes de desempeño inflados, en la gestión de recursos humanos, que reporta el estudio de OECD (2013). Evaluados juntos estos hechos sugieren que los sistemas de coordinación y evaluación, están más enfocados en el cumplimiento que en la generación de valor y este enfoque es incompatible con la búsqueda de valor que por definición está asociado a big data.

En resumen, es posible decir que; el avance de gobierno electrónico en Colombia es concebido como un proceso de madurez que involucra estandarización, apertura y participación. En términos de coordinación y monitoreo, se requieren mejoras significativas y nuevos enfoques para asegurar que el uso de big data en todos los sectores y niveles de gobierno contribuya a la generación de valor.

4.1.7. Otros recursos

Aunque los aspectos más relevantes de protección de datos personales y acceso a la información pública están considerados en la legislación colombiana; una revisión general de dichas regulaciones y la respuestas de Suarez (2014) no son concluyentes para determinar si situaciones específicas relacionadas con el uso de big data están contempladas: posible descubrimiento de información personal a partir de conjuntos de datos anonimizados, derechos sobre el uso de nuevas fuentes de datos como la ubicación personal y las cámaras de vigilancia, los procedimientos para usar información privada cuando es requerida para propósitos estadísticos (este y otros usos de datos por parte de entidades públicas son mencionados como excepciones en la ley de protección de datos pero los procedimientos no son claros), comercialización de datos y el acceso continuo a información pública como aquella recolectada de manera permanente por sensores o bases de datos transaccionales.

Con respecto a la existencia de un ecosistema de datos, como se mencionó anteriormente en el plan Vive Digital 2014-2018 los datos no están considerados claramente como fuente de valor. De acuerdo con MINTIC (2012b) no hay ninguna

definición de tarifas por la reutilización de información pública, sin embargo, los lineamientos de datos abiertos los consideran una fuente de nuevos modelos de negocio y servicios que pueden ser desarrollados por el sector privado (MINTIC 2012b).

En relación con la voluntad de compartir información, una encuesta desarrollada por la Universidad de los Andes (2014) entre 573 colombianos muestra cierta conciencia acerca de la privacidad y los datos personales así como falta de confianza con respecto al uso de datos personales por parte del sector público. En términos de percepción la mayoría de las personas están clasificadas como Pragmáticas 54% (están de acuerdo con varios usos de sus datos personales dependiendo de qué recibirían por ello) y Fundamentalistas 39% (no aceptan el uso de sus datos personales). Con respecto a la confianza, 53% considera que ni el sector público ni el privado respetan la ley de protección de datos personales mientras que 79% considera que ambos usan datos personales para propósitos no autorizados por la ley.

4.2. Brechas identificadas

Usando un puntaje de 0 a 10 esta sección resume las brechas identificadas entre las condiciones ideales y las condiciones reales para el aprovechamiento de big data en el sector público colombiano. Se asignó puntaje 0 (cero) a aquellas dimensiones que no presentan brechas o ya tienen planes definidos para cerrarlas, se asignó puntaje 5 a las dimensiones que presentan brecha que requieren alguna intervención y se asignó 10 a dimensiones que presentan brechas más significativas y relevantes.

Tabla 3. Brechas identificadas

Dimensión	Puntaje	Descripción de la brecha
Información	10	Escasez de volumen, variedad y velocidad en el catálogo de datos abiertos del sector público. No hay evidencia de datos abiertos del sector privado.
Tecnología	5	Brechas en la infraestructura entre diferentes sectores y diferentes niveles de gobierno. La ciberseguridad es un asunto complejo; se han hecho las primeras definiciones, aunque algunas decisiones no están completamente implementadas; las amenazas están evolucionando constantemente (Business Software Alliance 2011) y hay cierta dependencia de acuerdos internacionales (CIO Summit 2014).
Procesos	5	Estándares de calidad y visualización podrían incrementar el valor de los datos compartidos entre entidades.
Objetivos	10	Los datos no están incluidos como fuente de valor en el ecosistema digital y no hay metas relacionadas con el aprovechamiento de big data en el Plan Vive Digital 2014-2018.
Personal/talento	10	Hay brechas de conocimiento entre diferentes sectores y niveles de gobierno. La evidencia muestra escasez de conocimiento y probable deficiencia de aptitudes para el aprovechamiento de big data.

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

Dimensión	Puntaje	Descripción de la brecha
Sistemas y estructuras de gestión	10	<p>Aunque se está promoviendo el rol del CIO en cada entidad, aun no está consolidado y los componentes de estrategia de datos para la generación de valor aún no han sido incorporados en el rol.</p> <p>Se deben definir los mecanismos de coordinación y el alcance de las áreas responsables de la innovación en el sector público.</p> <p>Se requieren nuevos enfoques para la coordinación y el monitoreo de los avances en gobierno en línea, que estén mejor alineado con la búsqueda de valor que está asociado con la definición de big data.</p>
Otros recursos	10	<p>Se requieren definiciones legales que consideren situaciones relacionadas con los posibles usos de big data.</p> <p>No se ha concebido un ecosistema de datos</p>

Fuente: autor

5. Recomendaciones para el cierre de brechas

Esta sección contribuye a la fase de *Ideación* del proceso de diseño e innovación de la Dirección de Gobierno en Línea a través de la consolidación de recomendaciones y buenas prácticas que podrían contribuir a cerrar las brechas entre las condiciones ideales y las condiciones reales para el aprovechamiento de big data por parte del gobierno colombiano. Dado que algunas brechas identificadas en las dimensiones de la lista ITPOSMO (Heeks 2013) tienen relaciones con otras dimensiones, esta sección una una visión sistémica para proponer acciones orientadas a mejorar la situación general.

La generación de valor, como aspecto clave de la definición de big data, debería ser guiar las acciones e iniciativas de big data en el gobierno colombiano, varios autores plantean que los esfuerzos deber apuntar a lograr objetivos estratégicos y no a seguir big data como una tendencia (Beyer and Laney 2012; Accenture 2012) . Esto podría significar cambios significativos en los enfoques usados en las estrategias de gobierno en línea considerando nuevos planteamientos de (United Nations 2014b), ellos proponen estrategias enfocadas en la generación de valor público y adaptación constante a los retos emergentes en vez de la visión de un modelo de maduración para el gobierno electrónico.

Igualmente, la generación de valor debería guiar las acciones para prevenir las potenciales brechas socio-económicas entre países y organizaciones con capacidades para aprovechar big data y aquellos que no las tienen (Manyika et al. 2011; Buhl et al. 2013; United Nations 2012). Esto debería orientar las posturas del gobierno colombiano sobre el rol de la revolución de los datos en la agenda internacional para el desarrollo Post 2015 (United Nations 2014a).

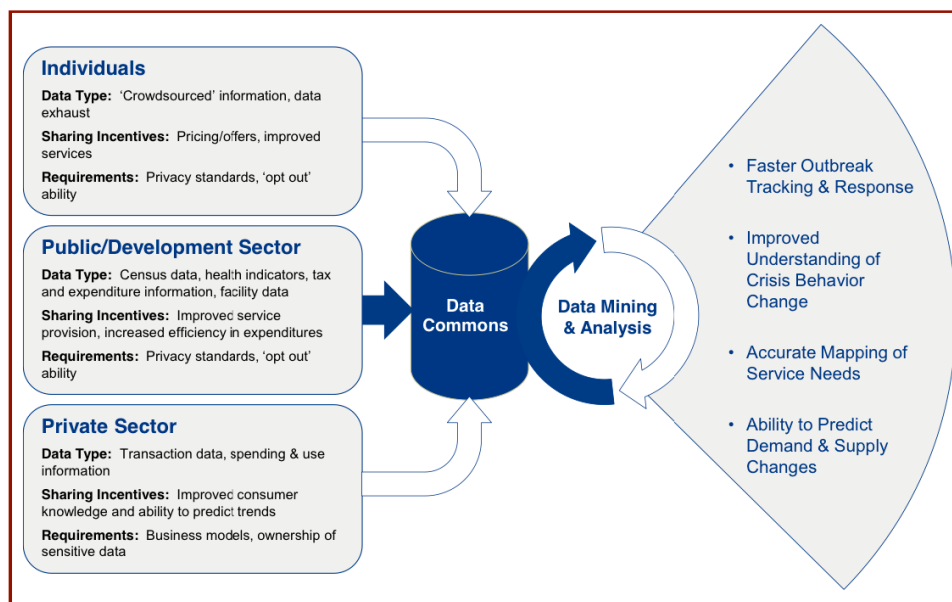
Como se mencionó en la revisión de literatura, el aprovechamiento de big data podría producir beneficios más allá del nivel organizacional; el presente trabajo ofrece

recomendaciones que empiezan por aquellas con mayor potencial para generar resultados a nivel social y posteriormente presenta aquellas que se centran en los beneficios organizacionales. Usando este enfoque, se presentan cinco líneas de acción: ecosistema de datos, innovación en el sector público, talento humano, cierre de brechas entre entidades públicas y otras recomendaciones.

5.1. Ecosistema de datos

La necesidad de concebir un ecosistema de datos con incentivos para el sector público y privado, así como para las personas es enfatizada por una variedad de autores desde compañías de consultoría hasta organismos internacionales. Un entendimiento común de este ecosistema implica, que compartir datos entre diferentes actores del ecosistemas sea fuente de valor para cada participante (World Economic Forum 2012; United Nations 2012; United Nations 2014b; TechAmerica Foundation 2012; Manyika et al. 2011; Shadbolt and Chui 2014). La Figura 2 muestra un modelo propuesto para este ecosistema por World Economic Forum (2012). Con respecto a cómo crearlo, el gobierno colombiano debería evaluar propuestas guiadas por la idea de “data philanthropy” [filantropía de los datos] (United Nations 2012) y datos compartidos para el bien común (United Nations 2012; Manyika et al. 2011) así como otras propuestas guiadas por una disciplina emergente conocida como infonomics que está enfocada en determinar el valor económico de los activos de información (Laney et al. 2013).

Figura 2. Ecosistema de datos propuesto por World Economic Forum (2012)



El ecosistema de datos podría incluir ambos enfoques aplicados de acuerdo con el uso planeado de los datos, por ejemplo la “filantropía de los datos” podría ser explorada como fuente de valor para minería de datos sociales (Helbing and Baliaetti 2011), modernización de los sistemas de estadística oficial (United Nations 2014a), gestión de riesgos, atención de desastres y prevención de conflictos (United Nations 2012). En este punto, los objetivos del plan Vive Digital 2014-2018 relacionados con el desarrollo de aplicaciones sociales (p.e. reducción de pobreza, agricultura) deberían considerar el hecho de que las personas están creando datos valiosos para el sector privado en sus actividades diarias (p.e. datos de ubicación generados por teléfonos celulares) y una reflexión que hace World Economic Forum (2012) sobre la necesidad de acciones que garanticen que los datos le ayuden a las personas y comunidades que los generan. Además de eso, Howard (2012b) recomienda establecer el valor económico de la mejora de desempeño del sector público que se habilita con el uso de big data para incluirlo como incentivo del ecosistema.

Con respecto a “infonomics”, el valor de los datos abiertos del sector público y su contribución a las oportunidades de negocio del sector privado deberían ser evaluados (Laney et al. 2013; World Economic Forum 2012), pueden surgir oportunidades de

generación de valor económico a partir de la planeación de la gobernanza de los datos de manera conjunta público-privada y de la participación de los usuarios en la identificación de conjuntos que deben ser abiertos (United Nations 2014b). Esto podría combatir la escasez en volumen, variedad, velocidad y valor encontrada en el inventario actual de datos abiertos y contribuir al cambio de mentalidad de algunos funcionarios públicos promoviendo la transparencia e innovación (MIT 2014).

Más y mejores conjuntos de datos abiertos favorecen el crecimiento de nuevos tipos de empresas como las que agregan y analizan datos abiertos (Manyika et al. 2011). Laney et al. (2013) comenta que los usuarios de datos estarán mucho más interesados en usar proveedores de datos procesados que en usar datos abiertos directamente. Sin embargo, la sostenibilidad de esos modelos de negocio requiere acuerdos acerca de la continuidad en el acceso a conjuntos de datos abiertos. Más allá de eso, los datos compartidos podrían producir beneficios comunes inclusive entre competidores quienes pueden compartir datos por ejemplo para ganar conjuntamente conocimiento sobre el comportamiento del cliente en un mercado específico (Chui et al. 2014).

El ecosistema de datos debe incluir reglas y transparencia para promover la confianza entre los participantes (United Nations 2014a; TechAmerica Foundation 2012). Además de los asuntos relacionados con datos personales, que ya se mencionaron, se requieren regulaciones para prevenir el uso desleal de datos compartidos (p.e. información privada revelada con el uso de analítica y usada para hacer benchmarking sin permiso del dueño de la información (Chui et al. 2014)). A pesar de que la regulación en Colombia está, hasta cierto punto, alineada con los principales asuntos de la privacidad de datos y el acceso a la información pública, se requiere mayor análisis legal de situaciones relacionadas con el aprovechamiento de big data. Se requiere especial atención a los asuntos de propiedad intelectual (Shadbolt and Chui 2014), el derecho de recolectar, agregar, diseminar y usar datos de ubicación personal con propósitos comerciales (Manyika et al. 2011) y los procedimientos para acceder a información pública de manera continua (p.e. si un servicio o un modelo de negocio requiere el uso permanente de conjuntos de datos que son dinámicos como transacciones o datos generados por sensores)

5.2. Innovación en el sector público

Se requiere una mejor definición de roles y procesos de coordinación para las estructuras que lideran la innovación (i.e. la Coordinación de Diseño e Innovación en la Dirección de Gobierno en Línea, el Centro de Innovación en Gobierno Electrónico, los nodos de innovación y el Centro de Excelencia y Apropiación de Big Data). Una vez definidos esos roles y relaciones, se recomienda acelerar la investigación y el desarrollo en áreas definidas (Manyika et al. 2011) y priorizar aquellas áreas con mayor potencial de generación de valor (el índice de valor potencial para big data usado por (Manyika et al. 2011) podría ser útil para hacer esta priorización). En ese sentido, Schulte et al. (2013) recomienda crear grupos de tarea para identificar beneficios tácticos e iniciativas "low-hanging fruit" [los frutos más fáciles de cosechar] y al mismo tiempo usar el enfoque de riesgo compartido con proveedores de tecnología para el cumplimiento de objetivos estratégicos o iniciativas de gran alcance.

Aunque la literatura disponible presenta big data como un fenómeno revolucionario para la creación de valor, algunos autores plantean que este potencial no implica que lograr capacidades para el aprovechamiento de big data sea simple o rápido (TechAmerica Foundation 2012; Brown et al. 2014). Por ese motivo, MINTIC debe evaluar varias rutas propuestas para la apropiación de big data, algunas de las más relevantes son presentados por Beyer and Laney (2012); TechAmerica Foundation (2012); Laney (2012) and Sicular (2013).

Aplicando el enfoque de "low-hanging fruit", hay algunas acciones que podrían ser consideradas por la Dirección de Gobierno en Línea, se debe explorar el concepto de "dark data" que es mencionado por Beyer and Laney (2012) como datos que son conservados con el único propósito de cumplir con reglamentación u operación pero que no generan valor agregado (en una conferencia sobre big data en Colombia desarrollada el 31 de octubre de 2014 en la Universidad Jorge Tadeo Lozano un expositor mencionó el enorme potencial que hay en más de 400 millones de registros del Ministerio de Salud que serían útiles para analizar aspectos económicos, poblacionales y científicos relacionados con este sector). Igualmente, Baghdassarian

et al. (2013) recomiendan que los desarrolladores de aplicaciones habiliten la recolección de datos y la analítica en sus aplicaciones para mejorar la generación de valor. También el lanzamiento de retos y la convocatoria a maratones de visualización como los que promueve Visualizing.org (2014) podrían producir resultados valiosos considerando que la Dirección de Gobierno en Línea ha consolidado prácticas de innovación abierta y que las herramientas de visualización son consideradas por varios autores como facilitadores del consumo de datos por parte de los niveles directivos y los ciudadanos (United Nations 2012; Brown et al. 2014; Biesdorf et al. 2013). En términos de gestión de conocimiento, debería ser implementado un monitoreo continuo de los avances de big data en el mundo y la documentación de casos de uso (OECD 2013; Buytendijk and Laney 2013). Con respecto a esto, Buytendijk and Laney (2013) recomienda buscar casos en industrias diferentes a la de la propia organización.

La gestión de conocimiento podría ser fortalecida a través de alianzas y la participación de instituciones colombianas en escenarios internacionales de discusión y trabajo sobre big data. Esto incluye el trabajo conjunto con instituciones de estadísticas oficiales para enfrentar retos comunes en este campo (United Nations (2014a) y el seguimiento al grupo global de trabajo sobre big data apoyado por Naciones Unidas (<http://www.unglobalpulse.org>), este grupo además de las actividades de divulgación de conocimiento, ejecutará proyectos piloto (United Nations 2014c). Algunas recomendaciones sobre el talento humano como componente clave del proceso de innovación se presentan en la siguiente sección.

5.3. Talento humano

Los cambios culturales requeridos para el aprovechamiento de big data son enfatizados por varios autores, hay dos cambios principales que se deben buscar en el sector público colombiano, el primero es la gestión guiada por datos desde la planeación hasta la evaluación, esta es una de las recomendaciones del estudio de OECD sobre gobernabilidad en Colombia y coincide con las recomendaciones de otros autores (Manyika et al. 2011; Laney 2012; Laney et al. 2013; OECD 2013).

El segundo cambio podría ser resumido como desempeño orientado hacia la generación de valor (OECD 2013; World Economic Forum 2012); aunque este cambio cultural es complejo y vá más allá del alcance de MINTIC, hay algunas recomendaciones que deberían ser consideradas en el dominio de la estrategia de gobierno en línea y los decretos y lineamientos que la soportan. Con respecto al enfoque actual de monitoreo basado en cumplimiento y la situación de puntajes inflados de desempeño que reporta el estudio sobre gobernabilidad en Colombia de la OECD (2013), esta organización da varias recomendaciones que pueden contribuir a que la gestión del desempeño organizacional esté alineada con la mentalidad de búsqueda y generación de valor que se requiere para el aprovechamiento de big data. Dichas recomendaciones incluyen una mejor evaluación del desempeño, el fortalecimiento del liderazgo y las capacidades de gestión y la mejora en los esquemas de compensación y reconocimiento. Con respecto a la compensación y el reconocimiento, World Economic Forum (2012) plantea la importancia de generar incentivos para aquellos individuos que contribuyen con su talento a la generación de valor público como una manera de mejorar el desempeño del talento humano actual mientras se fortalece el talento que se requerirá en el futuro. En este punto se podría considerar la promoción del intraemprendimiento como un esquema de incentivo para la mejora del desempeño.

Aunque el liderazgo y la capacidad de gestión se requieren en todas las entidades, en todas las áreas organizacionales y en todos los niveles de gestión; el rol de los CIO requiere atención especial dado que los planes y programas para su empoderamiento están siendo ejecutados. Considerando que no es razonable crear un Chief Data Officer cuando los esfuerzos en CIO apenas están comenzando, se deben incluir las implicaciones de estrategia y gestión de los datos en el rol de CIO siendo cautelosos con las limitaciones de contexto y las capacidades desiguales que tienen las personas que lo ejercen en las distintas entidades públicas. También se debe tener en cuenta la recomendación de TechAmerica Foundation (2012) acerca de que big data no debe ser visto como un tema del dominio exclusivo de las áreas de TI o de los CIO.

Con respecto a la creación de talento para big data, el programa Talento Digital, que usa un modelo similar al de servicio de la academia al gobierno propuesto por (TechAmerica Foundation (2012), debería crear convocatorias específicas para financiar personas interesadas en estudiar pregrados o postgrados en matemáticas aplicadas, estadística, analítica, aprendizaje de máquina y ciencia de los datos en Colombia o en el exterior. Considerando los resultados de largo plazo de este tipo de convocatorias, también debería considerarse la financiación de certificaciones internacionales en analítica como las mencionadas por Olavsrud (2014).

Adicionalmente, ante la dificultad de encontrar científicos de datos, que son entendidos como profesionales con conocimientos en modelos matemáticos y estadísticos, conocimiento profundo del sector para el que trabajan, creatividad e innovación en la formulación y resolución de problemas, habilidades para la comunicación de ideas a través de herramientas de visualización y gran capacidad de trabajo en equipo; se debe contemplar el enfoque de conformación de equipos multidisciplinarios para complementar capacidades (Newman and Beyer 2012; Buhl et al. 2013).

5.4. Cierre de brechas entre entidades públicas

La existencia de diferentes iniciativas de aprovechamiento de big data (lideradas de manera autónoma por el Ministerio de Hacienda, el gobierno de Bogotá y la Policía Nacional) sugiere capacidad técnica y financiera en dichas entidades. Sin embargo, varios hallazgos de esta investigación muestran una brecha de conocimiento e infraestructura entre diferentes entidades. En este escenario los roles de liderazgo y apoyo que desempeña MINTIC podrían contribuir a cerrar dichas brechas.

En términos de infraestructura TI, además de los potenciales beneficios del programa Compra Eficiente que ya se mencionaron; las soluciones en la nube para proveer herramientas clave para big data (p.e. Hadoop (Newman and Beyer 2012) y herramientas de visualización) para todo el sector público podrían contribuir a crear condiciones más parejas para aquellas entidades con menor capacidad técnica o

financiera. Este esfuerzo, que coherente con la recomendación de usar soluciones en la nube para big data (Brown et al. 2013), podría concentrar volumen y variedad de datos de distintas entidades en una sola plataforma lo cual podría dar paso a una cadena de iniciativas. Esta propuesta también podría favorecer el involucramiento de proveedores externos con capacidades relevantes (Schulte et al. 2013) y la mitigación del riesgo relacionado con la inmadurez de las soluciones de big data.

En términos de conocimiento, se requiere apoyo a las entidades en el análisis de factibilidad y en la ejecución de iniciativas de big data así como en la prevención de errores relacionados con inadecuada selección de fuentes de datos, deficiente evaluación de correlaciones o interpretación errónea de resultados. La creación de Centros of Excelencia transversales a las organizaciones es recomendada por varios autores (Manyika et al. 2011; Brown et al. 2014; TechAmerica Foundation 2012). Más allá de eso, TechAmerica Foundation (2012) menciona el caso de Estados Unidos en donde hay un “cross-government Big Data Steering Group”. Como se mencionó anteriormente el rol del Centro de Excelencia y Apropiación de Big Data con respecto a la innovación en las entidades públicas debe ser definido.

5.5. Otras recomendaciones

Hay un grupo de temas para los cuales el trabajo actual y el planeado está alineado con las condiciones ideales para el aprovechamiento de big data; para esos temas se recomienda continuar el trabajo actual e incluir acciones específicas para enfrentar los retos y riesgos mencionados en la revisión de literatura, y los que emerjan en el futuro, en relación con el fenómeno de big data. Esto aplica para la infraestructura de TI, la ciberseguridad, las regulaciones (incluyendo los asuntos que se mencionaron en el ítem de ecosistema de datos de esta sección) y la definición de procesos para asegurar la calidad de los datos y facilitar la visualización; esto implica un monitoreo continuo de la investigación sobre estos temas y las buenas prácticas documentadas (TechAmerica Foundation 2012).

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

Por tratarse de un trabajo desarrollado en el marco de investigación para la acción, este documento podría ser usado por MINTIC, para iniciar un ciclo de evaluación, implementación y reevaluación con la participación de actores relevantes para encontrar y divulgar buenas prácticas para el aprovechamiento de big data.

Referencias

- Accenture. (2012). *Build It and They Will Come? The Accenture Digital Citizen Pulse Survey and the Future of Government Operations*.
- Accenture. (2014). *Digital Government Pathways to Delivering Public Services for the Future*.
- Baghdassarian, S. et al. (2013). Predicts 2014: Apps, Personal Cloud and Data Analytics Will Drive New Consumer Interactions. *Gartner Research*.
- Bell, J. (2005). *Doing your research project: a Guide for First Time Researchers in Education, Health and Social Science*. O. U. Press, ed. Maidenhead: Open University Press.
- Beyer, M.A. and Laney, D. (2012). The Importance of “Big Data”: A Definition. *Gartner Research*.
- Biesdorf, S., Court, D. and Willmott, P. (2013). Big data: What’s your plan? *McKinsey Quarterly*, (March).
- Bohorquez, E. (2014). Public entities will be more sincronised [Entidades del Estado estarán más sincronizadas]. *El Espectador*. [online]. Available from: <http://www.elespectador.com/noticias/economia/entidades-del-estado-estaran-mas-sincronizadas-articulo-498984>.
- Brown, B., Court, D. and McGuire, T. (2014). Views from the front lines of the data-analytics revolution. *McKinsey Quarterly*, (March).
- Brown, B., Court, D. and Willmott, P. (2013). Mobilizing your C-suite for big-data analytics. *McKinsey Quarterly*, (Nov).
- Brown, T. (2008). Design thinking. *Harvard business review*, 86(6), p.84.
- Buhl, H.U. et al. (2013). Big Data A Fashionable Topic with(out) Sustainable Relevance for Research and Practice? *Business & Information Systems Engineering*, 55(2), pp.63–68.
- Business Software Alliance. (2011). *IT industry competitiveness index 2011*. [online]. Available from: <http://globalindex11.bsa.org>.
- Buytendijk, F. and Laney, D. (2013). Drive Value From Big Data Through Six Emerging Best Practices. *Gartner Research*.
- CAN. (2014). *Mobile Telephony in the Andean Community of Nations 2004-2013 [Telefonía Móvil en la Comunidad Andina 2004-2013]*. [online]. Available from: http://estadisticas.comunidadandina.org/eportal/contenidos/2434_8.pdf.
- CCIT and Fedesarrollo. (2013). *ICT industry: the new engine for Colombian economy [El sector TIC: la nueva locomotora de la economía colombiana]*. Bogota. [online]. Available from: http://www.fedesarrollo.org.co/wp-content/uploads/2013/06/2_08_2013-TIC-Junio-2103.pdf.

Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends. *Economic record*, 88, pp.2–9.

Chui, M., Manyika, J. and Van Kuiken, S. (2014). What executives should know about open data. *McKinsey Quarterly*, (January).

CIO Summit. (2014). *CIO Summit 2014*. Bogota. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-propertyvalue-7464.html>.

Cresswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage, ed. London: Sage.

Drever, E. (1995). *Using Semi-Structured Interviews in Small-Scale Research. A Teacher's Guide*. S. C. for R. in Education, ed. Edinburgh: Scottish Council for Research in Education.

El Espectador. (2014a). A system will reinforce compliance of vehicles with the regulation [Sistema detectará carros que no tengan documentos en regla]. *El Espectador*. [online]. Available from: <http://www.elespectador.com/noticias/nacional/sistema-detectara-carros-no-tengan-documentos-regla-video-495872>.

El Espectador. (2014b). How are regions regarding smartphones use? [¿Cómo están las regiones en uso de smartphones?]. *El Espectador*. [online]. Available from: <http://www.elespectador.com/tecnologia/estan-regiones-uso-de-smartphones-articulo-509479>.

Fan, W. and Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM SIGKDD Explorations Newsletter*, 14(2), pp.1–5.

Fister, S. (2013). *Data doesn't lie*.

GSMA Intelligence. (2014a). *ANALYSIS Mobile access — the last mile*. London.

GSMA Intelligence. (2014b). *ANALYSIS From concept to delivery: the M2M market today*. London.

GSMA Intelligence. (2014c). *Definitive data and analysis for the mobile industry*. [online]. Available from: <https://gsmaintelligence.com/>.

Guzman, F. (2014). National eGovernment Index 2013 Results [Resultados del Índice GEL 2013]. In Bogota: MINTIC, 29th and 30th of July. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-article-6454.html>.

Heeks, R. (2003). Most eGovernment-for-development projects fail: how can risks be reduced? *Manchester: Institute for Development Policy and Management, University of Manchester*.

Helbing, D. and Baliotti, S. (2011). From social data mining to forecasting socio-economic crises. *The European Physical Journal-Special Topics*, 195(1), pp.3–68.

Heudecker, N. (2013). Hype Cycle for Big Data, 2013. *Gartner Research*.

Howard, R. (2012a). Big Data Forces Big Change in the Age of Smart Government. *Gartner Research*.

Howard, R. (2012b). Big Data Forces Big Change in the Age of Smart Government. *Gartner Research*.

IADB. (2014). *The importance of maths for life [La Importancia de las Matemáticas para la Vida]*. [online]. Available from: <http://www.iadb.org/es/temas/educacion/bid-ensenanza-de-ciencias-y-matematicas-en-america-latina,6494.html>.

Kart, L., Heudecker, N. and Buytendijk, F. (2013). Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype. *Gartner Research*.

Laney, D. (2012). Big Data Strategy Components: Business Essentials. *Gartner Research*.

Laney, D., Buytendijk, F. and Linden, A. (2013). Predicts 2014: Innovating With Information Will Demand New Data, Organizations and Ideas. *Gartner Research*.

Lohr, S. (2012). The age of big data. *New York Times*, 11. [online]. Available from: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all&_r=0.

Manyika, J. et al. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. [online]. Available from: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

MayerSchönberger, V. and Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. 1st ed. London: John Murray Publishers.

MinHacienda. (2013). ISAAC Indicators: Following sectoral economic activity through Google Trends [Indicadores ISAAC: Siguiendo la actividad sectorial a partir de Google Trends]. *Fiscal Notes [Notas fiscales]*, 1(22). [online]. Available from: http://www.minhacienda.gov.co/portal/page/portal/HomeMinhacienda/politicafiscal/reportes-macroeconomicos/NotasFiscales/22_Siguiendo_la_actividad_sectorial_a_partir_de_Google_Trends.pdf.

Ministerio de Educación. (2014). *National Information System on Higher Education [Sistema Nacional de Información de la Educación Superior]*. [online]. Available from: <http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-212301.html>.

Ministry of Commerce Industry and Tourism. (2014). *Innova Award 2014 [Premio Innova 2014]*. Bogota.

MINTIC. (2014a). *Colombia is one of the countries with higher number of social networks users in its region [Colombia es uno de los países con más usuarios en redes sociales en la región]*. Bogota. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-article-2713.html>.

MINTIC. (2013). *Design and innovation process at the Directorate of eGovernment [Proceso de diseño e innovación en la Dirección de Gobierno en Línea]*.

MINTIC. (2014b). *Digital Talent [Talento Digital]*. Bogota. [online]. Available from: <http://www.vivedigital.gov.co/logros-plan/logro.php?lg=29>.

MINTIC. (2011). *Documento on strategic planning for the innovation subsystem for use and apropiation of ICT in government [DOCUMENTO DE PLANEACIÓN ESTRATÉGICA DEL SUBSISTEMA DE INNOVACIÓN PARA EL USO Y APROPIACION DE TIC EN EL GOBIERNO]*. Bogota. [online]. Available from: http://www.mintic.gov.co/portal/604/articles-6116_recurso_2.pdf.

MINTIC. (2012a). *eGovernment Strategy 2012-2015 for national entities, 2012-2017 for departments and municipalities [Estrategia Gobierno en línea 2012 – 2015 Para el orden nacional 2012 – 2017 Para el orden territorial]*. Bogota. [online]. Available from: <http://programa.gobiernoenlinea.gov.co/apc-aa-files/eb0df10529195223c011ca6762bfe39e/manual-3.1.pdf>.

MINTIC. (2014c). *Framework for IT Architecture for Colombian State [MARCO DE REFERENCIA PARA LA GESTIÓN DE TI EN EL ESTADO COLOMBIANO]*. Bogota. [online]. Available from: <http://estrategiaticolombia.co/gestioniti/lineamientos/>.

MINTIC. (2012b). *Guidelines for implementation of open data in Colombia [LINEAMIENTOS PARA LA IMPLEMENTACION DE DATOS ABIERTOS EN COLOMBIA]*. Bogota. [online]. Available from: http://programa.gobiernoenlinea.gov.co/apc-aa-files/da4567033d075590cd3050598756222c/Resumen_Ejecutivo_Datos_Abiertos.pdf.

MINTIC. (2014d). *Inventory/Catalogue of Open Data [Catálogo de datos abiertos]*. Bogota. [online]. Available from: <http://datosabiertoscolombia.cloudapp.net/frm/buscador/frmBuscador.aspx>.

MINTIC. (2014e). Ministry of Information and Communication Technologies, About us [Ministerio de Tecnologías de Información y las comunicaciones, ¿quiénes somos?]. [online]. Available from: <http://www.mintic.gov.co> [Accessed May 20, 2014].

MINTIC. (2014f). *MINTIC and Colciencias held the event Experiences in Nodes of Innovation 2014 [MINTIC y Colciencias organizan el evento Experiencias Nodos de innovación 2014]*. Bogota. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-article-6307.html>.

MINTIC. (2014g). *News [Noticias]*. Bogota. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-article-6670.html>.

MINTIC. (2014h). *Plan Vive Digital 2014-2018*. Bogota. [online]. Available from: <http://www.vivedigital.gov.co/2014-2018/>.

MINTIC. (2014i). *Vice-minister of IT [Viceministerio TI]*. Bogota. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-propertyvalue-543.html>.

MINTIC. (2014j). *What are innovation nodes? [Qué son los Nodos de Innovación?]*. Bogota. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-article-6116.html>.

MINTIC and COLCIENCIAS. (2013). *Strategic vision of software and IT related services industry, regionalised marketing plan. Executive Summary 2013 [VISIÓN ESTRATÉGICA DEL SECTOR DE SOFTWARE Y SERVICIOS ASOCIADOS PLAN DE MERCADEO Y VENTAS REGIONALIZADO DEL SECTOR EN COLOMBIA Resumen ejecu]*. Bogota.

MIT. (2014). International Trends in ICT An Initiative of the ICT Ministry of Colombia. *MIT Technology Review*.

Newman, D. and Beyer, M.A. (2012). Spotlight on Big Data: Separating Fact From Fiction. *Gartner Research*.

OECD. (2013). COLOMBIA Implementing good governance. *OECD Public Governance Reviews*. [online]. Available from: <http://dx.doi.org/10.1787/9789264202177-en>.

OECD. (2012). *Programme for International Student Assessment (PISA) Results overview Colombia 2012*. [online]. Available from: <http://www.oecd.org/pisa/aboutpisa/colombia-pisa.htm>.

Olavsrud, T. (2014). *11 Big Data Certifications That Will Pay Off*. [online]. Available from: <http://www.cio.com/article/2457266/certifications/160612-11-Big-Data-Certifications-That-Will-Pay-Off.html>.

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*.

Santofimio, N. (2014). *Monitoring meetings through Skype*. Bogota.

Schulte, R. et al. (2013). Predicts 2014: Business Intelligence and Analytics Will Remain CIO's Top Technology Priority. *Gartner Research*.

Shadbolt, N. and Chui, M. (2014). *Big Data, Big New Businesses*. London. [online]. Available from: <https://www.project-syndicate.org/commentary/nigel-shadbolt-and-michael-chui-estimate-that-the-annual-value-of-open-data-could-reach--3-trillion>.

Sicular, S. (2013). The Road Map for Successful Big Data Adoption. *Gartner Research*.

Suarez, S. (2014). *Personal interview through e-mail*. Bogota, 8th July.

Tamayo, Y.A. (2014). *Personal interview*. Bogota 12th August.

TechAmerica Foundation. (2012). Demystifying big data. *TechAmerica Foundation: Federal Big Data Commission*.

EL TIEMPO. (2014a). Recovered medical clinical records of “Mamá Canguro” method [Rescatan historias clínicas del método Mamá Canguro]. *EL TIEMPO*. [online]. Available from: <http://www.eltiempo.com/bogota/programa-mama-canguro/14325418>.

EL TIEMPO. (2014b). Which cities already have 4G networks [Sepa cuáles ciudades ya cuentan con redes 4G]. *EL TIEMPO*. [online]. Available from: <http://www.eltiempo.com/tecnosfera/novedades-tecnologia/redes-4-g-en-colombia/14347195>.

United Nations. (2014a). *Big data and modernization of statistical systems*.

United Nations. (2012). *Big Data for Development: Challenges & Opportunities*. New York. [online]. Available from: <http://globalpulse.org>.

United Nations. (2014b). *E-GOVERNMENT SURVEY 2014*. New York. [online]. Available from: <http://unpan3.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2014>.

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

United Nations. (2014c). *Global Pulse*. [online]. Available from: <http://www.unglobalpulse.org/>.

Universidad de los Andes. (2014). *Results of survey on privacy and personal data (Colombia) [Resultados de la encuesta privacidad y datos personales (Colombia)]*. Bogota.

Universidad Nacional. (2014). Empowering bright IT minds for the public sector [Empoderando una mente brillante en TI para el sector público]. In Bogota: MINTIC, 29th and 30th of July. [online]. Available from: <http://www.mintic.gov.co/portal/604/w3-article-6454.html>.

Visualizing.org. (2014). *Visualizing.org*. [online]. Available from: <http://www.visualizing.org/>.

World Economic Forum. (2012). *Big Data, Big Impact: New Possibilities for International Development*.

World Economic Forum. (2014). *The Global Information Technology Report 2014 Rewards and Risks of Big Data*. Geneva.

Anexo 1 Resumen de de los posibles usos de big data más mencionados en la literatura y sus potenciales beneficios.

Tema/sector	Ejemplo	MEJORES DECISIONES	MICRO SEGMENTACIÓN	NUEVOS MODELOS DE	EFICIENCIA Y EECTIVIDAD	MEJORES SERVICIOS	SEGURIDAD Y PREVENCIÓN DEL CRIMEN	TRANSPARENCIA
Impuestos, finanzas y compras estatales	Diligenciamiento previo de formatos							
	Uso de motores automatizados de análisis de riesgo para identificar individuos o empresas que requieren auditoría o controles específicos.							
Planeación urbana y ciudades inteligentes	Creación de servicios con el uso combinado de datos oficiales y datos no oficiales (reportados por los ciudadanos) para ofrecer información pertinente de acuerdo a la ubicación de los ciudadanos.							
	Mapeo de las condiciones y situación de las calles							
	Creación de centros de control en las ciudades que usan grandes volúmenes de datos para generar capacidades de recuperación en desastres naturales y para mejorar las tareas de policía y seguridad.							
	Policía predictiva							
	Ahorro de energía en la prestación de servicios.							
Salud	Tecnologías “chip-on-a-pill” [procesador en una píldora] —medicinas que actúan como instrumentos para reportar datos cuando son ingeridos por los pacientes—alimentan datos casi en tiempo real a bases de datos							
	Pagos basados en la mejoría real de los pacientes para medicinas/tratamientos innovadores.							
	Nuevos modelos de negociación con farmacéuticas o proveedores de servicios de salud.							
	Apoyo a la toma de decisiones en políticas públicas.							
	Monitoreo de asuntos como la incidencia y costos de enfermedades inducidas por estilo de vida y comportamiento.							
	Minimizar distorsiones económicas en modelos de pago en los sistemas de salud.							
	Reducción de complejidad administrativa en pagos.							
	Uso de datos sobre antecedentes genéticos sobre los pacientes para personalizar tratamientos.							
Seguridad, aplicación de la ley, gestión de	Twitter es considerado especialmente útil para proveer información enriquecida sobre comportamientos, percepciones, preocupaciones y creencias relacionadas con la salud.							
	Uso de analítica para la monitorear la diseminación de enfermedades infecciosas.							
Seguridad, aplicación de la ley, gestión de	Prevención y respuesta a una variedad de ataques.							
	Protección de la niñez							

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

Tema/sector	Ejemplo	MEJORES DECISIONES	MICRO SEGMENTACIÓN	NUEVOS MODELOS DE	EFICIENCIA Y EFECTIVIDAD	MEJORES SERVICIOS	SEGURIDAD Y PREVENCIÓN DEL CRIMEN	TRANSPARENCIA	
riesgos y prevención del crimen	La creciente disponibilidad de datos de ubicación personal, tráfico en tiempo real y aplicaciones de GPS ofrecen oportunidades para respuestas más oportunas y efectivas por parte de la policía, bomberos y ambulancias. Estas tecnologías habilitan a los despachadores/controladores de servicios de emergencia para identificar la ubicación de quien llama, selección inteligente de las rutas de llegada para mejorar los tiempos de respuesta y monitoreo de la seguridad de quienes atienden emergencias en zonas peligrosas.								
	Uso de mapeo computarizado y análisis de variables diversas como patrones históricos de arrestos, días de pagos, eventos deportivos, clima y días festivos, entre otros, para predecir “puntos calientes” de seguridad y desplegar personal y equipo en dichas zonas de manera anticipada. (policía predictiva)								
	Predicciones de clima y prevención de desastres con el uso de sensores y nuevos modelos de análisis.								
	Uso de imágenes de satélite para medir y monitorear cultivos ilícitos y daños ambientales causados por actividades ilícitas.								
Transporte	Uso de teléfonos celulares con aplicaciones de GPS o dispositivos ubicados en vehículos para reducir costos en los sistemas de recaudo en peajes. La ubicación del teléfono celular puede generar el cobro al pasar un peaje y cargarlo a la factura de telefonía celular o cuenta bancaria.								
	Uso de predictores de demanda de largo plazo, basados en datos de ubicación personal, para planes de tránsito y transporte.								
	Uso de sistemas de GPS en buses pueden alertar a centros de control cuando un bus se acerca a un semáforo, el sistema puede darle prioridad al transporte público. (semaforización inteligente)								
	Uso de dispositivos RFID por parte de adultos mayores o personas con discapacidad. La conexión de dichos dispositivos con semáforos puede ofrecer más tiempo para cruzar de manera más segura.								
	Mejorar el manejo de personal de seguridad en aeropuertos a partir de la predicción de flujo de usuarios.								
Estadísticas oficiales	Uso de datos de compra de almacenes, técnicas de extracción de datos de la web o análisis de búsquedas para monitorear o inferir indicadores económicos como índice de precios al consumidor, PIB o tasa de desempleo.								
	<i>En este campo, se encontró un caso de un caso de uso en Colombia, liderado por el Ministerio de Hacienda (se detalla en este documento).</i>								
	Los datos de teléfonos móviles podrían ser usados para analizar comportamientos de viaje y turismo, confianza del consumidor a través de análisis de sentimientos en redes sociales y disponibilidad de vacantes de empleo.								
	Hay varios ejemplos del uso de Google Trends para analítica (p.e. economía, salud, comportamiento del consumidor)								
	Cooperación con operadores de telefonía móvil para calcular estadísticas de desplazamiento georeferenciadas.								
	Uso de imágenes satelitales y sensores de suelo para estimar rendimiento de cultivos. Uso de retenes y sensores de pavimento para generar estadísticas de tráfico.								

¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de big data?

Tema/sector	Ejemplo	MEJORES DECISIONES	MICRO SEGMENTACIÓN	NUEVOS MODELOS DE	EFICIENCIA Y EECTIVIDAD	MEJORES SERVICIOS	SEGURIDAD Y PREVENCIÓN DEL CRIMEN	TRANSPARENCIA
	Algunas experiencias han encontrado que el PIB de un país puede ser estimado en tiempo real analizando las emisiones de luz en la noche a través de sensores remotos.							
	Colombia ha usado imágenes satelitales como fuente de big data en varias áreas. Un ejemplo es una prueba piloto para complementar el trabajo de encuestas del censo nacional agrícola.							
Gobierno abierto y transparencia	Por ejemplo, Combined On-Line Information System (COINS) ofrece acceso abierto a todos los gastos del gobierno en el Reino Unido.							
	Organizaciones de la sociedad civil podrían crear análisis de los gastos públicos sectoriales o el recaudo de impuestos en grupos o industrias específicas.							