



BIG DATA:

UN APORTE PARA LA DISCUSIÓN
DE LA POLÍTICA PÚBLICA EN COLOMBIA

Con el apoyo de:

**PRIVACY
INTERNATIONAL**



www.karisma.org.co

Noviembre de 2016

“En un esfuerzo para que todas las personas tengan acceso al conocimiento, Fundación Karisma está trabajando para que sus documentos sean accesible, es decir, tienen un formato electrónico diseñado para que su contenido pueda ser leído por el mayor número de personas posible, incluidas las que tienen algún tipo de discapacidad o de dificultad para la lectura y comprensión. Más información sobre el tema: <http://www.documentoaccesible.com/#que-es>

En la elaboración de este informe intervinieron:

Juan Diego Castañeda
Laura Mora
Carolina Botero
Amalia Toledo
Stéphane Labarthe

Consulta esta investigación en línea en

<https://karisma.org.co/descargar/big-data-un-aporte-para-la-discusion-de-la-politica-publica-en-colombia/>



Big Data: un aporte para la discusión de la política pública en Colombia está disponible bajo Licencia Creative Commons Reconocimiento compartir igual 4.0

“Usted puede remezclar, retocar, y crear a partir de esta obra, incluso con fines comerciales, siempre y cuando le de crédito al autor y licencien nuevas creaciones bajo las mismas condiciones. Para ver una copia de esta licencia visite: https://creativecommons.org/licenses/by-sa/4.0/deed.es_ES

Indice

INTRODUCCIÓN	5
<i>BIG DATA</i> EN EL SECTOR OFICIAL EN COLOMBIA.....	6
<i>BIG DATA</i> , LO QUE COLOMBIA APRENDE DE EEUU, REINO UNIDO, AUSTRALIA Y COREA	8
Justificaciones de las políticas sobre <i>big data</i>	9
Riesgos del <i>big data</i>	12
Intimidad y protección de datos	12
Anonimización de datos	14
Discriminación y transparencia	16
CONCLUSIONES Y RECOMENDACIONES	19
NOTAS	20

Introducción

En marzo de 2016, el Departamento de Planeación Nacional (DNP) anunció que Colombia entraría en “la revolución de los datos” de la mano de una política pública que desarrollaría esta entidad y que sería presentada a principios de 2017.¹

Con este documento pretendemos dar una mirada a las justificaciones y razones que llevan a los gobiernos a implementar políticas de *big data* y a los beneficios que ven en esta herramienta y repasamos algunos riesgos que se ha identificado en su uso. Esto con el objetivo específico de servir de insumo para el desarrollo de la política anunciada y de propiciar una discusión entre el sector público, la sociedad civil y las demás partes interesadas en el tema.

En la primera parte, se revisa el origen normativo de la política pública de *big data* en Colombia, así como algunos avances institucionales que ya hay en la materia. Seguidamente, repasamos y analizamos algunos aspectos de las políticas públicas de *big data* implementadas en Estados Unidos, Reino Unido, Australia y Corea, que son las que han inspirado a nuestras entidades públicas para encontrar en ellas (i) qué entienden por *big data*, (ii) cómo lo justificaron, y (iii) qué riesgos identificaron en su implementación. Al finalizar el análisis ofrecemos algunas conclusiones y recomendaciones para alimentar el proceso de construcción de la política nacional.

Big data en el sector oficial en Colombia

El artículo 45 del Plan Nacional de Desarrollo (PND) es la única norma que menciona *big data*.² Dice que el Ministerio de las Tecnologías de la Información y las Comunicaciones (Ministerio TIC) definirá estándares, modelos, lineamientos y normas técnicas para la incorporación de la TIC, que serán aplicados a la

Interoperabilidad de datos como base para la estructuración de la estrategia que sobre la captura, almacenamiento, procesamiento, análisis y publicación de grandes volúmenes de datos (*big data*) formule el Departamento Nacional de Planeación.

Es decir, mientras el Ministerio TIC dará lineamientos para poner en marcha la interoperabilidad necesaria al interior del Estado que permita implementar la estrategia, el DNP planteará la política de *big data*.

El Ministerio TIC va avanzando en sus obligaciones y ya publicó una primera versión de lo que será la política derivada del mencionado artículo 45 en lo que se refiere a autenticación electrónica, carpeta ciudadana e interoperabilidad.³ El DNP está trabajando en su política de *big data* sin que se conozca aún ningún documento público al momento de escribir este análisis.

A pesar de que no hay una política pública, el sector público ha hecho algunos pilotos a partir del *big data*. El Departamento Administrativo Nacional de Estadística (DANE) ha venido realizando varios pilotos sobre uso de *big data* y estadísticas oficiales que ha incorporado en sus estrategias. Concretamente, el DANE planea utilizar el *big data*, principalmente, en dos líneas de trabajo: el monitoreo de los Objetivos de Desarrollo Sostenible y el Censo Nacional de Población y Vivienda.

El DANE también ha realizado ejercicios exploratorios de fuentes no tradicionales. Esto hace referencia al uso de datos no estadísticos como CDR, información de redes sociales, etc. Dentro de estos ejercicios se cuentan: (i) CeluPop, un proyecto que generaría estadísticas sociodemográficas a partir del uso de celulares, y (ii) la utilización de *big data* para “el cálculo de la relación de la tasa de consumo de suelo urbano y la tasa de crecimiento de población, mediante el uso de imágenes de sensores remotos”. Vale señalar que este último proyecto fue presentado al grupo de trabajo global de la Organización de las Naciones Unidas (ONU) en *big data* para estadísticas oficiales.⁴ Por su parte, el Ministerio de Hacienda junto con el DNP desarrollaron una metodología a partir de datos de *Google Trends*, que analiza la frecuencia de términos de búsqueda para inferir actividad económica en ciertos sectores y obtener indicadores más rápido que a través de las estadísticas tradicionales.⁵

Estos desarrollos están relacionados con y alentados por la visión de *big data* que impulsa la ONU, el Banco Mundial y la Comisión Económica para América Latina, a la que haremos referencia también en este documento y que se identifica con la *Revolución de los Datos* y el mejoramiento del monitoreo de las sociedades para la consecución de los Objetivos de Desarrollo Sostenible⁶, Big Data para el Desarrollo⁷, y el papel destacado de las agencias de estadística en el tema.⁸

En términos institucionales y con el fin de conseguir la creación de esa política pública, el Gobierno Nacional proyecta el impulso de los llamados Centros de Excelencia y Apropiación (CEA), formados por una alianza entre el gobierno, el sector privado y la academia. El objetivo de estos CEA es el apoyo en la creación de una política pública sobre *big data*, obtener soluciones problemáticas de interés público y privado, contribuir a la competitividad y al posicionamiento de una cultura de innovación e investigación aplicada y capitalizar el análisis de datos en sectores estratégicos.

A la fecha, se cuenta con un Centro de Excelencia y Apropiación, llamado CAO-BA, que cuenta con el apoyo del Ministerio TIC y de Colciencias, aunado a la participación del Grupo Nutresa, Bancolombia, Universidad ICESI, Universidad EAFIT, Universidad de los Andes, IBM, EMC, SAS, el Clúster Creativ como aliado estratégico, y la Universidad Javeriana, que es la entidad ejecutora del CEA.

***Big data*, lo que Colombia aprende de EEUU, Reino Unido, Australia y Corea**

En una reunión que tuvo la Fundación Karisma con el DNP, se mencionó que para la elaboración de la política colombiana se estaban observando los ejemplos de EEUU, Australia, Reino Unido y Corea. Cada uno de estos países ha desarrollado una estrategia, en las que se pueden identificar desde qué concepto de *big data* parten, qué razones hay para realizar una política dedicada al tema y cuáles son los riesgos que supone el uso de esta herramienta.

Aunque no haya un consenso sobre el significado y límites del concepto, en los documentos que recogen la política pública se define de una u otra forma *big data*. En EEUU, por ejemplo, el reporte *Big Data: Seizing Opportunities, Preserving Values* menciona que se trata de la “[c]reciente habilidad tecnológica de capturar, combinar, y procesar grandes volúmenes, y variedades de datos”.⁹ En Reino Unido se hace referencia a la ubicuidad de los datos, el volumen de las bases que los contienen, el crecimiento en la producción y los nuevos orígenes de los datos como características de *big data*.¹⁰ Australia¹¹ y Corea¹², por su parte, además de estas características, relacionan las especiales capacidades de análisis y procesamiento de los datos. Estas técnicas y metodologías diferencian el *big data* de los métodos tradicionales, pues enfrenta la dificultad de analizar el gran volumen y variedad de datos.

En el artículo 45 del PND se define el *big data* como “la captura, almacenamiento, procesamiento, análisis y publicación de grandes volúmenes de datos”. De esta forma, recoge los elementos más generales de la forma como se ha limitado el concepto.

El *big data* normalmente se identifica con el manejo de grandes volúmenes de datos, que se producen rápidamente y que, a diferencia de los datos usados tra-

dicionalmente, permiten trabajarlos en diferentes formatos y con estructuras variadas.¹³ En relación con el *big data* para el desarrollo, se ha entendido que el concepto debe pasar de considerar la definición clásica de volumen, velocidad y variedad como constituyentes del concepto de *big data* a hablar de *crumbs* (migajas), capacidades y comunidad.¹⁴ *Crumbs* o migajas hace referencia a que el *big data* hace uso de datos producidos por la interacción con plataformas o dispositivos digitales que no tenían como propósito principal generar esos datos. Son rastros individuales e interconectados de acciones humanas que recogen los dispositivos digitales.

Capacidades o *big data analytics* son las habilidades y métodos para producir y analizar los nuevos tipos de datos, por ejemplo, visualización, aprendizaje automático estadístico (*statistical machine learning*), o algoritmos. Se trata de las capacidades necesarias para buscar nuevas perspectivas (*insights*) en los datos. Finalmente, se señala que hay una comunidad variada de actores relacionados con la producción y análisis de los datos, o las migajas que potencialmente involucra a todo el mundo.

Justificaciones de las políticas sobre *big data*

Las declaraciones públicas por parte de representantes del Estado y del gobierno colombiano dejan ver un conjunto amplio de justificaciones en favor del *big data*. Por ejemplo, el director del DNP ha dicho que:

El objetivo de usar *big data* en el Estado es hacer un mejor uso de un activo gigantesco que ya tiene el Estado y que está subutilizado: sus datos. La estrategia de *big data* debe construirse sobre una buena infraestructura de datos de calidad, abiertos, y en donde los diferentes entes estatales tengan acceso a datos producidos por otras entidades.¹⁵

En ese mismo anuncio se confía en que el *big data* ayude a mejorar las bases para la toma de decisiones que permitan la creación de estrategias públicas más sólidas, crear valor para la ciberseguridad, la internet de las cosas, mejorar la creación de negocios, optimizar los negocios existentes, permitir la formulación de políticas públicas más asertivas, hacer presupuestos informados, poder conocer la opinión de la ciudadanía con precisión y en cuestión de horas, etc.

Otras supuesta aplicaciones benéficas del *big data* según el gobierno, incluyen la gestión del agua, pues permitirá optimizar el recaudo y generar esquemas de ahorro. También la movilidad en el análisis de los patrones de desplazamiento de la población con el fin de mejorar la movilidad y la seguridad ciudadana. En el sector salud, se menciona que la analítica permitirá una mejor utilización de los datos existentes, además de que optimizaría el uso de los recursos del

Estado. En cuanto a educación, se habla de evaluación de la rentabilidad de la inversión por análisis en cadena. Finalmente, en el tema de la gestión de residuos sólidos se observan posibles beneficios en el análisis del volumen y tipo de desechos por zonas, y en el aprovechamiento que se les estaría dando.

Con base en lo anterior, se deduce que la posición oficial sobre los beneficios de crear una política de *big data* es, en el mejor de los casos, muy amplia todavía. La característica que mejor la describe hasta ahora es *confusa* y el reto del DNP será superar esta situación para evitar que esto afecte negativamente su resultado. Sin conocer documentos públicos que muestren la ruta de la futura política de *big data* en Colombia, podemos visualizarla a través de los países que se han mencionado como referencia: Reino Unido, Australia, EEUU y Corea,

En los países referencia, los gobiernos justifican en forma muy diversa las políticas de *big data* en el sector público. Las principales son:

1. *Mejorar los procesos de toma de decisiones*

Se considera que *big data* permite usar la enorme cantidad de datos que actualmente se manejan casi en tiempo real e incluso de manera automática.

En Australia, por ejemplo, se buscan mejorar las políticas públicas bajo la idea de que entre más información se maneje, mejores decisiones se pueden adoptar y más pronto se podrán predecir los resultados de una determinada política incluso antes de que ella sea puesta en marcha. Adicionalmente, se resalta que permitiría el análisis en tiempo real, lo que impactaría en forma importante campos como la salud, el medio ambiente y los servicios sociales.¹⁶

2. *Mejorar la eficiencia y productividad en la prestación de servicios.*

Este tema se relaciona con el gasto público y tiene como trasfondo la presión que hay sobre los distintos gobiernos a nivel mundial para reducir costos de funcionamiento. Según las justificaciones de estas políticas, con el uso de *big data* se puede ahorrar, destinar y aprovechar mejor los recursos.

En EEUU se menciona como propósito

[C]onseguir decisiones eficientes en las actuaciones policia-les para: (i) destinar mejor los recursos: identificar patrones en la actividad delictiva, y luego distribuir los recursos económicos y humanos; (ii) identificar agentes que necesitan entrenamiento adicional para prevenir el exceso en el uso de la fuerza.¹⁷

También se considera que su despliegue

Incrementará la confianza de la comunidad en las fuerzas policiales, [pues] con la implementación de la *Police Data Initiative* se pretende identificar las políticas que se pueden mejorar [y] se constituye en un canal de comunicación para que la ciudadanía pueda manifestar las inconformidades que tengan con el servicio.¹⁸

En Australia se busca mejorar en algunos sectores donde se da una relación directa con el incremento de la eficacia en la prestación de servicios del Estado y con la mejora en la relación costo-beneficio. Se menciona que con el *big data* se busca personalizar y focalizar los servicios para evitar *over-servicing*. Esa adaptación se basa en las necesidades y los intereses específicos que tengan tanto la comunidad como los individuos¹⁹. Este mejoramiento tiene para ellos aplicación en varios campos: salud, seguridad social (que sería personalizada), y reducción de crímenes en el sector público y privado.

3. *Fomentar la transparencia*

La creación de algún portal para la consulta y descarga de bases de datos oficiales en formato estándar es uno de los primeros pasos que dan los gobiernos que buscan desarrollar *big data*.²⁰ Por ejemplo, existen portales oficiales en Chile, Kenia, Estados Unidos, Reino Unido, Australia, Canadá y Singapur.

Precisamente, en Corea, el *big data* se concibió como una herramienta para la implementación de datos abiertos en el gobierno. Quizá por eso hay un importante desarrollo en la idea de fortalecer las pequeñas y medianas empresas, así como la generación de nuevos empleos.²¹ En todo caso, entre las justificaciones están también mejorar los procesos de toma de decisiones en política pública, la prestación de servicios en el sector público y el desarrollo de capacidades para el uso de datos en ciencia.²²

4. *Mejorar los procesos estadísticos*

En Reino Unido se consideró que, si se tienen mejores tecnologías, el uso del *big data* puede impactar positivamente en los métodos estadísticos tradicionales, sobre todo, teniendo en cuenta el importante incremento en el volumen de los datos que alimentan las bases de datos.²³ Australia también menciona esta tema como uno de los propósitos de su estrategia nacional.²⁴

Parte de la estrategia de instituciones como la ONU, la Organización para la Cooperación y el Desarrollo Económico (OCDE), el Fondo Monetario Internacional y el Banco Mundial consiste en reforzar el lugar de las oficinas nacionales de estadística en el desarrollo de *big data*. Para Letouzé y Jütting, los roles de las instituciones u oficinas de estadística oficial en el desarrollo de *big data* son principalmente dos: (i) proveer a la sociedad conocimiento sobre sí misma para que sirva en la toma de decisiones y la evaluación de sus políticas; y (ii) proponer un espacio de deliberación sobre qué vale la pena medir, cómo y por qué.²⁵ En este sentido, el *big data* se combinaría con las formas de medición tradicional con el objetivo de hacer un mejor monitoreo del avance hacia los Objetivos de Desarrollo Sostenible.²⁶

Riesgos del *big data*

Los beneficios y opciones de desarrollo que trae el *big data* son importantes. Sin embargo, hablar de la recolección, almacenamiento y gestión de datos masivos no puede hacerse en forma ligera, es necesario analizar los riesgos. Algunos riesgos que nos interesa resaltar son los riesgos para el derecho a la intimidad, para la seguridad de los datos y para la dignidad.

Intimidad y protección de datos

El derecho fundamental a la intimidad y a la protección de datos puede verse afectado por el *big data* de diversas formas, pues los resultados del análisis de datos pueden revelar información personal que corresponde a su espacio íntimo, sin necesidad ni proporcionalidad alguna.

En este sentido, el riesgo más directo en este punto es el hecho de que la recopilación y análisis de toda clase de datos pueda dar a conocer aspectos de la esfera íntima de una persona como sus hábitos, preferencias o su círculo social y familiar. Para explicar mejor este punto, repasaremos el marco legal colombiano sobre protección de datos y lo contrastaremos con el tipo de datos que se analizan en *big data*.

La legislación reconoce la existencia de un tipo particular de información o datos llamados personales. Este tipo de datos son los que están vinculados o que puedan asociarse a una persona natural, determinada o determinable.²⁷ El tratamiento de este tipo de datos puede hacerse contando con el consentimiento de la persona a la que se refieren o sin autorización en los siguientes casos:

- Cuando la ley permite el tratamiento sin autorización;
- Cuando media una solicitud por parte de entidad pública o administrativa en ejercicio de sus funciones;

- Cuando se cuenta con una orden judicial;
- Cuando el dato sea público, esto es, cuando la ley lo determine así o cuando no está calificado como privado²⁸ o semiprivado²⁹;
- Cuando el dato se requiera para fines científicos, estadísticos o históricos; o
- Cuando los datos estén relacionados con el registro civil.

Una categoría especial de datos personales es la de datos sensibles.³⁰ Este tipo de datos está definido como aquellos que “afectan la intimidad” o cuyo uso pueda generar discriminación de cualquier tipo. La ley enuncia, a modo de ejemplo, los datos que revelan información racial, política, religiosa o de pertenencia a grupos como organizaciones sociales o partidos políticos. También son datos sensibles los relativos a la salud, la vida sexual y biométricos, como huellas digitales, iris, voz, etc. El tratamiento de estos datos está prohibido excepto cuando haya autorización, salvo que una ley determine que no se necesite o el tratamiento tiene finalidad histórica, estadística o científica con “supresión de la identidad de los titulares”, entre otros.³¹

Como se ha mencionado anteriormente, el *big data* tiene que ver con la recopilación y análisis de grandes volúmenes de datos que tienen orígenes diversos. Una tipología de datos para el *big data* los clasifica así:³²

- Datos subproducto de las comunicaciones móviles (metadatos de las comunicaciones), transacciones financieras (uso de tarjetas de crédito), transporte (GPS) o actividades en línea (*cookies* o direcciones IP).
- Contenidos digitales que aparecen en redes sociales, por ejemplo.
- Datos de sensores físicos (medidores “inteligentes”) o remotos (imágenes satelitales).

Los problemas que plantea el *big data* para la intimidad, en primer lugar, pasan por el hecho de que muchos de estos datos, como los metadatos de las comunicaciones, son considerados privados por ser de igual valor que el contenido de la comunicación.³³ Por tanto, acceder y utilizar estos requiere de una solicitud de un fiscal dentro de una investigación criminal y orden judicial.³⁴

El *big data* también plantea riesgos para la protección de datos. Por ejemplo, los principios de finalidad y libertad del tratamiento de datos se pueden ver amenazados. Según el principio de finalidad, “el tratamiento debe obedecer a una finalidad legítima de acuerdo con la Constitución y la Ley, la cual debe ser informada al titular”.³⁵ Para la Corte Constitucional la finalidad debe ser constitucionalmente legítima, definida previamente y sin ambigüedad. El uso o divulgación de datos para una finalidad diferente a la inicial está prohibida.³⁶

Esto conecta con el principio de libertad en la medida en que los datos no pueden ser tratados para una finalidad que la persona titular no autorizó y de que tiene derecho a decidir si sus datos pueden estar en una base.

Ya que una característica del *big data* es que usa datos que no fueron recogidos para un propósito específico, o que fueron entregados con un propósito pero se analizan para otro, se puede concluir que es imposible pensar que una persona autorizó el uso de sus datos para las infinitas posibilidades que teóricamente ofrece el *big data*. En otras palabras, el consentimiento para el tratamiento de datos es inefectivo porque no se sabe para qué se puedan o van a usar los datos.

Puede alegarse que, como se anotó anteriormente, el tratamiento de datos personales, inclusive los sensibles, puede ocurrir por autorización de la ley y para fines estadísticos, históricos o científicos. Consideramos que en ambos casos la limitación del derecho a la intimidad que implica el uso de datos personales sin autorización de quienes están relacionados con esos datos debe pasar por un juicio de necesidad y proporcionalidad, de manera que no cualquier ley permita el tratamiento sin autorización. En ese sentido, la Corte Constitucional sostuvo que la posibilidad de tratamiento de datos sensibles sin autorización de la persona titular debe “estar contenida en una ley, sea conforme a las garantías que otorga el *habeas data*, por ejemplo en materia de finalidad, y cumpla las exigencias del principio de proporcionalidad”.³⁷

En cuanto al *big data* para fines estadísticos, la legislación colombiana impone el deber de darle uso puramente estadístico a la información que recibe el DANE cuando hace censos y encuestas. Esto significa que dicha información solo puede darse a conocer a través de “resúmenes numéricos” de los que no pueda extraerse ninguna información personal que sirva para fines comerciales, tributarios o de investigación judicial.³⁸ Esto representa una garantía para quienes suministran información al Estado para fines estadísticos y, por tanto, crea un silo de datos que debe ser respetado, aunque choque contra la pretensión del *big data* de analizar datos provenientes de distintas fuentes.

Anonimización de datos

Las herramientas para garantizar que los datos analizados en el *big data* no estén directamente relacionados con una persona son insuficientes. Hasta la fecha, no existen mecanismos infalibles para la anonimización de la información.

Para algunas aplicaciones de *big data*, recolectar información personal de forma desproporcionada parece ser esencial. Por ejemplo, según la ONU, los metadatos de las comunicaciones móviles pueden ser útiles para la respuesta de

emergencia en situaciones de desastre, investigación en salud o el levantamiento de datos socioeconómicos.³⁹ Sin embargo, cuando se anonimizan o agregan datos puede perderse parte de su utilidad, por lo cual parece haber una dicotomía entre intimidad y utilidad de los datos.⁴⁰

En relación con el problema de discriminación que puede generar el *big data*, la importancia de los datos y la confianza en la realidad que ellos reflejan provocan una tensión entre el derecho a la intimidad y el acceso a servicios públicos para la protección de otros derechos fundamentales. Por ejemplo, la agregación de datos de distintos grupos étnicos oculta desigualdades entre las diferentes comunidades, lo que puede llevar a representar a dichas comunidades de manera equivocada (p. ej. como más saludables).⁴¹ Desagregar datos sirve entonces para identificar, monitorear y atender las necesidades de servicios sociales de estos grupos particulares, sirve para asegurar la prestación de servicios y programas vitales para atender estas poblaciones y para que los gobiernos centrales tomen mejores decisiones.

Si la existencia de datos sobre una persona o comunidad es un requisito para el disfrute de un derecho o la creación y mantenimiento de políticas públicas dirigidas a sectores o poblaciones particulares, la anonimización va a ser más un obstáculo para alcanzar esos servicios públicos que una garantía del derecho a la intimidad. Por tanto, los métodos estadísticos tradicionales deben mantenerse vigentes paralelamente al uso de datos no convencionales propios del *big data*, de forma tal que la anonimización y la agregación puedan cumplir su función de protección del derecho a la intimidad y los datos conserven cierta utilidad auxiliar.

En EEUU el riesgo a la intimidad forma parte de lo que llaman “exceso de confianza”, pues admiten que existen métodos tecnológicos para proteger la privacidad pero que están muy lejos de la perfección. Han reconocido que es necesario instaurar normas sociales y políticas sensibles al tema, además de un marco legal acorde con esa necesidad de protección.⁴² Con una aproximación muy práctica, EEUU ve como riesgos a la intimidad: (i) la imposibilidad de limitar la cantidad de datos que se recolecta dado que provienen de fuentes variadas y, frecuentemente, de tecnologías con las que no se puede optar no producir los datos (celulares, internet inalámbrico en el hogar, etc.)⁴³; y (ii) el alto grado de dificultad para anonimizar la información. Si bien hay esfuerzos para desarrollar mejores técnicas para este fin, también es cierto que el trabajo por conseguir técnicas de reidentificación son aún más fuertes.⁴⁴

En el Reino Unido también se reconoce que el riesgo a la intimidad está unido al cifrado, la anonimización y la vigilancia. Se identifican como problemas que

se van a intensificar en el futuro, especialmente relacionados con la tecnología autónoma y la inteligencia artificial.⁴⁵

La política australiana, por su parte, reconoce el riesgo constante a la intimidad. La dimensión del problema supuso que a mayo de 2016 se lanzará una consulta sobre la guía de principios para *big data* y privacidad, encaminada a obtener comentarios del público. En concreto, una forma de abordar la intimidad se relaciona con las normas de protección de datos. Si bien los riesgos que corren los datos siguen siendo los mismos que en el mundo analógico, el *big data* obliga a pensar en unos nuevos. Por ejemplo, los riesgos que se generan en el proceso de recolección de datos, o los que surgen tratándose de los sistemas de transporte y entrega de datos entre entidades o entre las dependencia de esas entidades.⁴⁶

Discriminación y transparencia

Aunque está estrechamente relacionada con la violación al derecho a la intimidad y la protección de datos, el análisis y la toma de decisiones basada en esa metodología puede traer como consecuencia la discriminación de grupos particulares.

Este riesgo proviene de dos factores: (i) decisiones automatizadas; y (ii) falta de transparencia en los algoritmos. Significa que mientras se siga sin saber la forma en que funcionan y la información que los alimenta, los casos de discriminación se seguirán presentando, pues no hay quién pueda hacer una auditoría de los algoritmos para encontrar fallas o mejorarlos.

En EEUU este riesgo se relaciona con el de discriminación pero está vinculado directamente con el uso de algoritmos, toda vez que no se sabe ni cómo funcionan ni cuál es la información que los alimenta. No obstante, son los que definen los parámetros del análisis de datos y pueden llegar a tomar las decisiones. Es decir, dado que suelen ser cajas cerradas, no se puede hacer un control sobre ellos ni tampoco es posible rebatir las decisiones que toman.⁴⁷ Las decisiones tomadas con algoritmos, que se relacionan con acceso a créditos, vivienda, empleo, etc., son entendidas como de especial importancia para aplicar la transparencia.⁴⁸ Por otra parte, así como se reconoce la existencia de errores en algunos sistemas, se concluye que otra función de la transparencia es permitir mejorar su funcionamiento por medio de revisiones periódicas que deben exigirles.⁴⁹

Según la política del Reino Unido, la discriminación es un problema derivado del funcionamiento de los algoritmos y del hecho de que en ellos se encuentren inmersos prejuicios personales, de donde surge el riesgo de que en la provisión de servicios la discriminación sea sistemática. Sobre los algoritmos, las ideas

de que son perfectos o casi perfectos, y que, por lo menos, son mejores que el criterio humano puede resultar peligrosas.⁵⁰

Los sistemas de toma de decisión automáticos o semiautomáticos, donde el algoritmo propone una decisión o da una puntuación (*scoring*) pero la decisión final es humana, no es nuevo. Los bancos, por ejemplo, usan un sistema de puntuación nombrado FICO que fue creado en 1956. Sin embargo, la aplicación de estos sistemas algorítmicos a todas las áreas de nuestras vidas desde la publicidad en línea hasta la policía predictiva y los drones, su alimentación con volúmenes gigantes de datos (*big data*) e inteligencia artificial (*machine learning*) es reciente.

Algunos han identificado ciertos riesgos de la toma de decisiones automatizadas, que han sido retomados por publicaciones estatales.⁵¹ La Unión Europea enfocó su análisis en los problemas de privacidad (*privacy by design*)⁵², mientras que en EEUU se centraron en los problemas de discriminación (*equal opportunity by design*). En su último informe sobre este asunto, la Casa Blanca asume que bien concebidos e implementados, los algoritmos podrían permitir ir hacia más igualdad mientras no presenten riesgos de discriminación muy altos.⁵³ Esta afirmación es llamativa, pues así como para algunos el riesgo de discriminación es muy alto, para otros es posible que una implementación correcta de estos métodos pueda reforzar la igualdad, lo que en últimas demuestra que no hay consenso sobre el tema.

Crear una política de *big data* implica asumir este riesgo e impone la obligación de crear métodos para exigir transparencia, concepción ética y mecanismos de control que hoy no existen. El tema ha sido discutido también por la sociedad civil y en este terreno la organización estadounidense *Electronic Privacy Information Center* lanzó este año una campaña para la transparencia de los algoritmos.⁵⁴

La necesidad de ética y transparencia busca evitar que los derechos de una persona se vean disminuidos como producto de una decisión que no puede conocer ni auditar. Dicho de otra manera, es posible que muchas puertas se cierren en virtud de los resultados de estos algoritmos y la persona ni siquiera sepa que estas oportunidades le fueron negadas. Es muy difícil defenderse contra “un enemigo invisible”.⁵⁵

Para solucionar estos problemas, la legislación francesa, por ejemplo, obliga a que las decisiones tomadas a partir de algoritmos mencionen esta circunstancia y que la persona sepa las reglas que definen ese tratamiento.⁵⁶ Otros instrumentos legales como la Ley Suplementaria sobre Protección de Datos

Personales de la Comunidad Económica de Estados de África Occidental que incluso prohíben que las decisiones con efectos legales sobre una persona se originen solamente en el procesamiento automático de datos para perfilar y evaluar ciertos aspectos de su personalidad.⁵⁷

Teniendo en cuenta que un algoritmo puede definirse como una sucesión de reglas y de operaciones aplicadas a datos para resolver un problema, hay tres puntos esenciales en los que se necesita transparencia:

1. *La existencia de un proceso de decisión basado sobre un algoritmo y una puntuación (scoring)*

La transparencia es necesaria en caso, pues muchas veces tanto la recolección de datos como su tratamiento por algoritmos resultan invisibles para la persona, que solo ve el resultado y a veces ni siquiera eso.

2. *Tipos de datos usados para tomar una decisión*

Hoy, la gran mayoría de los datos no son llenados por las personas en un formulario de manera activa y consciente, sino que son entregados de manera pasiva e inconsciente: dirección IP, localización, navegación en internet, etc. Además, estos datos se intercambian, se venden y se producen.

3. *Procedimientos (técnicos y humanos) de toma de decisiones, presentes en particular en los algoritmos, el “corazón” de la caja negra*

El análisis de los algoritmos puede ser un trabajo titánico ya que algunos de ellos pueden tener un nivel de complejidad muy alto (p. ej. uso de teorías matemáticas de alto nivel) o pueden contener hasta miles de reglas diferentes. Por esto, además de exigir reglas de autocontrol (*accountability*), conviene preguntarse en función de los casos a qué nivel se requiere la transparencia.

En este sentido, se puede pedir transparencia en (i) el proceso de toma de decisión: lógica, grandes lineamientos; (ii) los algoritmos: reglas preestablecidas y construidas automáticamente (*machine learning*); y (iii) el código fuente, la base de datos, etc.

También puede tener una parte humana, puesto que muchos procesos de decisión son semiautomáticos. Un algoritmo puede entregar una puntuación o elementos de análisis de datos y hasta una propuesta de decisión que en últimas un humano es el que la toma. En este caso, hay procedimientos, formaciones, etc., que son elementos en los que la persona se basa para tomar esta decisión.

Igualmente, debe haber transparencia en estos “procedimientos humanos”. Unas de las recomendaciones del informe de la Casa Blanca es que los algoritmos deben incluir mecanismos que faciliten la transparencia, el autocontrol y la autocorrección, además de la necesidad de desarrollar algoritmos de auditoría.⁵⁸ Ahora bien, considerando el impacto que tiene en la vida de las personas las decisiones que toman los privados, el informe señala que estas reglas también pueden aplicarse a ese sector.

Conclusiones y recomendaciones

Pensar en la construcción de una política pública sobre *big data* no es una tarea sencilla. Por eso, en primer lugar, el Gobierno debe propiciar espacios de participación de otros sectores y partes interesadas en este tema. De esta forma, podrán tenerse en cuenta las distintas preocupaciones para que, en últimas, la política sea el producto de una discusión democrática y abierta.

La escasa información pública sobre el desarrollo de la política a cargo del DNP solamente deja una lista heterogénea de las posibles aplicaciones del *big data* en Colombia. Por eso, es recomendable empezar por determinar unos objetivos claros para el desarrollo de esta política y un campo de aplicación restringido que permita obtener experiencia en la aplicación de esta técnica. Así, se podrá evitar poner injustificadamente en riesgo el derecho a la intimidad y a la protección de datos.

En ese mismo sentido, es necesario tener en cuenta que en el sector TIC periódicamente aparecen y abundan temas de moda, como sería el de *big data*, que suelen llegar acompañados de promesas de progreso, bienestar, desarrollo y democracia y que se ven frustrados por muchos factores que es mejor reconocer desde el principio. En otras palabras, el *big data* puede ofrecer muchos beneficios cuya realización no está garantizada sin que se evalúen los nuevos problemas que se enfrentan. Por esta razón, es recomendable no desmontar los desarrollos actuales en favor del derecho a la intimidad y de la protección de datos con el único propósito de permitir la ejecución de la política de *big data*. Por el contrario, los riesgos que hemos referido antes tienen que servir para que estas protecciones se mantengan, refuercen y actualicen.

NOTAS

1. Departamento de Planeación Nacional. (2016, 22 de marzo). *Big Data: Colombia entra en la revolución de los datos*. Disponible en <https://www.dnp.gov.co/Paginas/Big-Data-Colombia-entra-en-la-revoluci%C3%B3n-de-los-datos-.aspx>.
2. Ley No. 1753 de 9 de junio de 2015. Disponible en <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=61933>.
3. Fundación Karisma. (2016). *Comentarios de Fundación Karisma al documento de Servicios Digitales Básicos*. Disponible en <https://karisma.org.co/comentarios-de-fundacion-karisma-al-documento-de-servicios-digitales-basicos/>.
4. DANE. (s.f.). *Informe de rendición de cuentas DANE, 2014-2015*. Disponible en https://www.dane.gov.co/files/rendicion_cuentas/Informe%20Rendicion%20de%20cuentas%202015.pdf.
5. Mejía, L.F. et al. (2013). Indicadores ISAAC: siguiendo la actividad sectorial a partir de Google Trends. *Notas Fiscales*, 20.. Disponible en http://www.minhacienda.gov.co/HomeMinhacienda/ShowProperty?nodeId=%2FOCS%2FMIG_15924604.PDF%2F%2FidcPrimaryFile&revision=latestreleased.
6. Véase UN Economic and Social Council. (s.f.) *Millennium Development Goals and post-2015 Development Agenda*. Disponible en <http://www.un.org/en/ecosoc/about/mdg.shtml>; y IEAG on a Data Revolution for Sustainable Development. (2014). *A World That Counts*. Disponible en <http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>.
7. ONU Global Pulse. (2012). *Big Data for Development: Challenges & Opportunities*. Disponible en <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>.
8. Data-Pop Alliance. (2016). *Oportunidades y requerimientos para aprovechar el uso de big data para las estadísticas oficiales y los Objetivos de Desarrollo Sostenible en América Latina*. Disponible en <http://datapopalliance.org/wp-content/uploads/2016/05/Data-Pop-Alliance-LAC-NSO-ES.pdf>.

9. Executive Office of the President. (2014). *Big data: seizing opportunities, preserving values*. Disponible en https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.
10. Royal Statistical Society. (2015). *The opportunities and ethics of big data*. Disponible en <http://www.rss.org.uk/Images/PDF/influencing-change/2016/rss-report-ops-and-ethics-of-big-data-feb-2016.pdf>.
11. Office of the Australian Information Commissioner. (2016). *Consultation draft: Guide to big data and the Australian Privacy Principles*. Disponible en <https://www.oaic.gov.au/engage-with-us/consultations/guide-to-big-data-and-the-australian-privacy-principles/consultation-draft-guide-to-big-data-and-the-australian-privacy-principles>.
12. Seung-ah, L. (2013, 12 noviembre). *El uso de datos públicos estimula la economía creativa*. Disponible en <http://spanish.korea.net/NewsFocus/Policies/view?articleId=114888>.
13. Laney, D. (2001, 6 de febrero). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group Research*, 6. Disponible en <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
14. Véase Data-Pop Alliance, *op. cit.*, nota 8; y Letouzé, E & Jütting, J. (2015). Official Statistics, Big Data and Human Development. *Data-Pop Alliance White Paper Series*. Disponible en <http://datapopalliance.org/wp-content/uploads/2015/11/Big-Data-Official-Statistics-FINAL.pdf>.
15. Departamento de Planeación Nacional. (2016, 22 de marzo). *Big data: Colombia entra en la revolución de los datos*. Disponible en <https://www.dnp.gov.co/Paginas/Big-Data-Colombia-entra-en-la-revoluci%C3%B3n-de-los-datos-.aspx>.
16. Australian Government Information Management Office. (2013). *The Australian Public Service Big Data Strategy*. Disponible en <https://www.finance.gov.au/sites/default/files/Big-Data-Strategy.pdf>.
17. Executive Office of the President. (2016). *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. Disponible en https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.
18. Ibid.
19. Australian Government Information Management Office, *op. cit.* (nota 16).
20. Kim, G.H., Trimi, S., & Chung, J.H. (2014). Big-data applications in the government sector. *Communications of the ACM*, 57(3), pp. 78–85. Disponible en <https://doi.org/10.1145/2500873>.
21. Desai, A. & Sánchez-Andrade, B (2015). Korea shows how to use Big Data for development [blog]. *Voices: Perspective on Development*. Disponible en <http://blogs.worldbank.org/voices/korea-shows-how-use-big-data-development>.
22. Seung-ah, *op. cit.* (nota 12).
23. Royal Statistical Society, *op. cit.* (nota 10).

24. Australian Government Information Management Office, *op. cit.* (nota 16).
25. Letouzé, *op. cit.* (nota 14).
26. IEAG on a Data Revolution for Sustainable Development, *op. cit.* (nota 6).
27. Véase Ley Estatutaria No. 1581 de 17 de octubre de 2012, artículo 3, literal (c). Disponible en <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=49981>; y Ley Estatutaria No. 1266 de 31 de diciembre 2008, artículo 3, literal (e). Disponible en <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=34488>.
28. El dato privado está definido como el que, por su naturaleza íntima o reservada, solo es relevante para la persona titular. Ley Estatutaria No. 1266 de 31 de diciembre 2008, artículo 3, literal (h).
29. Un dato semiprivado es el que no es íntimo, reservado o público, pero cuyo conocimiento o divulgación interesa no solo a la persona titular sino a un “cierto sector o grupo de personas o a la sociedad en general”. Ejemplos de este dato son el dato financiero y crediticio de actividad comercial. Ley Estatutaria No. 1266 de 31 de diciembre 2008, artículo 3, literal (g).
30. Ley Estatutaria No. 1581 de 17 de octubre de 2012, artículo 5.
31. *Ibid.*, artículo 6.
32. Data-Pop Alliance. (2015). *Big data for climate change and disaster resilience: realising the benefits for developing countries*. Disponible en https://assets.publishing.service.gov.uk/media/57a0897eed915d622c00023d/DFID_Synthesis_Report_Big_Data.pdf.
33. Corte Interamericana de Derechos Humanos (2009). Caso Escher et al. vs. Brasil. Decisión del 6 de julio de 2009. Disponible en: http://www.corteidh.or.cr/docs/casos/articulos/seriec_200_esp1.pdf.
34. Constitución Política de Colombia, artículos 15 y 235.
35. Ley Estatutaria No. 1581 de 17 de octubre de 2012, artículo, literal (b).
36. Corte Constitucional. Sentencia T-729 de 2002. Disponible en <http://www.corteconstitucional.gov.co/relatoria/2002/t-729-02.htm>.
37. Corte Constitucional. Sentencia C-748 de 2011. Disponible en <http://www.corteconstitucional.gov.co/relatoria/2011/c-748-11.htm>.
38. Ley No. 79 de 20 de octubre de 1993, artículo 5(2). Disponible en <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=14376>.
39. Global Pulse. (2013). *Mobile Phone Network Data for Development*. Disponible en http://www.unglobalpulse.org/sites/default/files/Mobile%20Data%20for%20Development%20Primer_Oct2013.pdf.
40. Yakowitz, J. (2011). Tragedy for the Data Commons. *Harvard Journal of Law & Technology*, 25(1). Disponible en <http://jolt.law.harvard.edu/articles/pdf/v25/25HarvJLTech1.pdf>.
41. Asian and Pacific Islander American Health Forum. (2016). *Public-Private Partnerships for Data Equity*. Disponible en http://www.apiahf.org/sites/default/files/2016.04.28_PUBLIC-PRIVATE%20PARTNERSHIPS%20FOR%20DATA%20EQUITY_Report_0.pdf.

42. Executive Office of the President, *op. cit.* (nota 17), p. 2.
43. Executive Office of the President, *op. cit.* (nota 9), pp. 53-54.
44. *Ibid*, p. 54.
45. Royal Statistical Society, *op. cit.* (nota 10).
46. Australian Government Information Management Office, *op. cit.* (nota 16).
47. Executive Office of the President, *op. cit.* (nota 9).
48. Véase *Ibid.*, p. 53, & Executive Office of the President, *op. cit.* (nota 17), pp. 8-9.
49. Véase *Ibid*, pp. 52-53 (nota 9) & *Ibid*, p. 9 (nota 17).
50. Royal Statistical Society, *op. cit.* (nota 10).
51. Citron, D.K. & Pasquale, F.A. (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, 89, p. 1. Disponible en <https://ssrn.com/abstract=2376209>.
52. D' Acquisto, G. et al. (2015). *Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics*. Europe: European Union Agency for Network and Information Security. Disponible en <https://www.enisa.europa.eu/news/enisa-news/privacy-by-design-in-big-data-an-overview-of-privacy-enhancing-technologies-in-the-era-of-big-data-analytics>.
53. Executive Office of the President, *op. cit.* (nota 17).
54. Electronic Privacy Information Center. (s.f.) *Algorithmic transparency: end secret profiling. Disclose the basis of automated decision making*. Disponible en <https://epic.org/algorithmic-transparency/>.
55. Executive Office of the President, *op. cit.* (nota 17).
56. Ley No 2016-1321 pour une République Numérique, artículo 4.
57. Economic Community of West African States. (2010, 16 de febrero). Supplementary Act A/SA.2/10/10 on Personal Data Protection, artículo 35(2). Disponible en <http://www.statewatch.org/news/2013/mar/ecowas-dp-act.pdf>.
58. Executive Office of the President, *op. cit.* (nota 17).