



Derecho y big data

Jairo Becerra (Editor) • Lorenzo Cotino Hueso

• Ivonne Patricia León • Marco Emilio Sánchez-Acevedo

• Jheison Torres Ávila • John Velandia Vega

PÚBLICO 25

JAIRO BECERRA. Abogado; diploma de Estudios Avanzados en Derecho Internacional de la Universidad de Barcelona, España; especialista en Ciencia Política y de la Administración de la Universidad de Barcelona, España; director de Investigaciones de la Facultad de Derecho de la Universidad Católica de Colombia; profesor de posgrado en Gobierno Electrónico; profesor en Derecho Internacional y Derecho Aeronáutico. Ha escrito cinco libros, siete capítulos de libro, siete artículos científicos, desarrollado dos *softwares* y realizado más de treinta y tres ponencias nacionales e internacionales.

LORENZO COTINO HUESO. Abogado; doctor en Derecho; licenciado y DEA en Ciencias Políticas de la Universidad de Valencia, España; magíster en Derecho de la Universidad Ramón Llull, España; profesor investigador de la Universidad Católica de Colombia; catedrático de Derecho Constitucional en la Universidad de Valencia; coordinador de la Red Derecho TICs; Premio INAP, Ministerio de Defensa, Ejército, Extraordinario de Doctorado; profesor invitado en Konstanz (Alemania); profesor Honorario de la Universidad Nacional de Colombia; magistrado del Tribunal Superior de Justicia C. Valenciana. Ha escrito o coordinado dieciséis libros, ciento ventitres artículos o capítulos y doscientos cincuenta ponencias o comunicaciones nacionales e internacionales.

IVONNE PATRICIA LEÓN. Politóloga; magíster en Derecho de la Universidad Nacional de Colombia; profesora investigadora de la Universidad Católica de Colombia; coordinadora de Formación para la Investigación del Centro de Investigaciones Socio-Jurídicas (CISJUC) de la Facultad de Derecho de la Universidad Católica de Colombia; docente de la Universidad Nacional de Colombia. Ha escrito dos libros, catorce capítulos de libro, siete artículos científicos, y realizado más de cinco ponencias nacionales.

MARCO EMILIO SÁNCHEZ-ACEVEDO. Abogado; doctor con mención Cum Lauden en Tecnologías y Servicios de la Sociedad de la Información de la Universidad de Valencia, España; magíster en Ciberseguridad y Ciberdefensa Nacional de la Escuela Superior de Guerra, Colombia; especialista en Derecho Administrativo y Constitucional, y en Gobierno Local Electrónico; diploma de Estudios Avanzados en Derecho Administrativo; profesor investigador de la Universidad Católica de Colombia; exdirector de la firma Mas Abogados & P.; actual director de la firma IUSATIC Abogados y Consultores. Ha escrito seis libros, cuatro capítulos de libro, dos artículos científicos, y realizado más de diez ponencias nacionales e internacionales.

JHEISON TORRES ÁVILA. Abogado; doctor con mención Cum Lauden en Derechos Fundamentales de la Universidad de Alicante, España; magíster en Derecho de la Universidad Nacional de Colombia; profesor investigador de la Universidad Católica de Colombia; profesor asociado de la Universidad Nacional de Colombia; director de la Maestría en Derechos Humanos de la misma universidad. Ha escrito siete libros, siete capítulos de libro, nueve artículos científicos, y realizado más de trece ponencias nacionales e internacionales.

JOHN VELANDIA VEGA. Ingeniero de Sistemas; magíster en Information Technology de la Universidad de Stuttgart, Alemania; profesor investigador de la Universidad Católica de Colombia; profesor de Arquitectura de *Software*; profesor de especialización en la Universidad Católica de Colombia; fue líder del semillero de investigación GINOSKO; investigador en Open Data, representación de datos y marcos de referencia de TI. Ha escrito dos libros y cinco artículos científicos, desarrollado cuatro *softwares* y realizado más de ocho ponencias nacionales e internacionales.

Jairo Becerra (Editor) • Lorenzo Cotino Hueso
• Ivonne Patricia León • Marco Emilio Sánchez-Acevedo
• Jheison Torres Ávila • John Velandia Vega

Derecho y big data

RS
PÚBLICO 25



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

Becerra, Jairo

Derecho y Big Data / Jairo Becerra, Lorenzo Cotino Hueso, Ivonne Patricia León, Marco Emilio Sánchez-Acevedo, Jheison Torres Ávila y John Velandia Vega.— Bogotá : Universidad Católica de Colombia, 2018

128 páginas; 17 x 24 cm. — (Colección Jus Público; no. 25)

ISBN: 978-958-5456-28-0 (impreso)

ISBN: 978-958-5456-29-7 (digital)

I. Título II. Serie III. Cotino Hueso, Lorenzo IV. León, Ivonne Patricia V. Sánchez-Acevedo, Marco Emilio VI Torres Ávila, Jheison VII. Velandia Vega, John

1. DERECHO INFORMÁTICO. 2. PROTECCIÓN DE DATOS 2. MACRODATOS-ASPECTOS JURÍDICOS

Dewey 340.0285 SCDD ed. 21

Proceso de arbitraje

1.º concepto de evaluación:

18 de abril 2018

2.º concepto de evaluación:

2 de mayo de 2018

© Universidad Católica de Colombia

© Jairo Becerra (Editor)

© Lorenzo Cotino Hueso

© Ivonne Patricia León

© Marco Emilio Sánchez-Acevedo

© Jheison Torres Ávila

© John Velandia Vega

Primera edición, Bogotá, D. C.

Junio de 2018

Dirección Editorial

Stella Valbuena García

Coordinación Editorial

María Paula Godoy Casasbuenas

Corrección de estilo

John Fredy Guzmán Vargas

Diseño de colección

Juanita Isaza

Diagramación

Andrés Mauricio Enciso Betancourt

Publicación digital

Hipertexto Ltda.

www.hipertexto.com.co

Bogotá, D. C., Colombia

Impresión

Xpress Estudio Gráfico y Digital S.A.

Bogotá, D. C., Colombia

Cómo citar esta obra

En APA:

Becerra, J., Cotino Hueso, L., León, I. P., Sánchez-Acevedo, M. E., Torres Ávila, J., & Velandia Vega, J. (2018). *Derecho y big data*. Bogotá: Editorial Universidad Católica de Colombia.

En Chicago:

Becerra, Jairo, Lorenzo Cotino Hueso, Ivonne Patricia León, Marco Emilio Sánchez-Acevedo, Jheison Torres Ávila y John Velandia Vega. *Derecho y big data*. Bogotá: Editorial Universidad Católica de Colombia, 2018.

Facultad de Derecho

Carrera 13 n.º 47-49

Bogotá, D. C.

derecho@ucatolica.edu.co

Editorial

Universidad Católica de Colombia

Av. Caracas 46-72, piso 5

Bogotá, D. C.

editorial@ucatolica.edu.co

www.ucatolica.edu.co

Todos los derechos reservados. Esta publicación no puede ser reproducida ni total ni parcialmente o transmitida por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sin el permiso previo del editor.

Hecho el depósito legal

© Derechos reservados



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

Proyecto de investigación
Derecho y big data

Facultad de Derecho

Grupo de investigación
Derecho público y TIC

Línea de investigación
Derecho TIC

Facultad de Ingeniería

Grupo de investigación
Software inteligente y convergencia tecnológica

Línea de investigación
Software inteligente y convergencia tecnológica

Resumen

El tránsito de lo analógico a lo digital permitió la emergencia de procesos autogestionables que desplazaron a los medios de comunicación convencionales e hicieron posible el acceso a la información de forma extendida, transformando la economía, el derecho y la administración del Estado. Los diferentes textos que se presentan en este libro analizan los impactos de las nuevas tecnologías de la información desde una perspectiva transdisciplinaria y multidisciplinaria, tomando en consideración fenómenos como *big data*, derecho, internet, ciberdefensa, gobierno electrónico, acceso a la información, brecha digital, protección de datos, inteligencia artificial y cibernética. A partir de disciplinas como la sociología, la ciencia política y el derecho, se avanza en una perspectiva global y compleja acerca de la influencia de las tecnologías de la información y la comunicación en el campo jurídico y en el Estado

Palabras clave

Acceso a la información, *big data*, brecha digital, ciberdefensa, cibernética, derecho, gobierno electrónico, inteligencia artificial, internet, protección de datos:

Abstract

The transition from analog to digital technology enabled the emergence of self-managed processes that displaced conventional media, and made it possible to access information in an extended manner, transforming economy, law, and State administration. The different studies presented in this book analyze the impacts of new information technologies from a transdisciplinary and multidisciplinary perspective, taking into account phenomena such as big data, law, internet, cyberdefense, electronic government, access to information, digital divide, data protection, artificial intelligence, and cybernetics. Different disciplines, such as sociology, political science, and law, contribute to a global and complex perspective about the influence of information and communication technologies in the legal field, as well as in the State.

Key words

Access to information, big data, digital divide, cyberdefense, cybernetics, law, electronic government, artificial intelligence, internet, data protection

CONTENIDO

Introducción	11
Capítulo 1. La cuarta revolución tecnológica: un nuevo paradigma de comprensión de la sociedad y el Estado más allá del <i>big data</i> e internet	15
Una revolución tecnológica	15
Un mundo en cambio: contextos de la crisis.....	17
Entre la tercera y cuarta revoluciones tecnológicas	20
La integración de los sistemas: más allá del <i>big data</i> y la internet	24
Las transformaciones del Estado: la influencia de la revolución tecnológica	28
El gobierno y la democracia en la era digital.....	33
Hacia una gobernanza digital.....	36
Capítulo 2. Marco teórico y aproximación jurídica al <i>big data</i> : algoritmos, inteligencia artificial y transformación digital.....	39
Aproximación general al <i>big data</i> : premisas y cautelas	39
¿De dónde proceden los datos masivos y todo lo imaginable que se puede hacer con ellos?	41
Algunas cautelas que plantea el <i>big data</i> y la necesidad de abordarlo jurídicamente	43
Aproximación a elementos de economía colaborativa y transformación digital afines a la investigación	44
Actores, actividades y sistemas de economía colaborativa.....	45
La necesaria aproximación jurídica y nuevos enfoques desde los derechos fundamentales	46

Datos masivos, discriminaciones y brecha masivas.....	48
De la personalización masiva de la información al fin de la esfera pública y la posverdad: libertad de expresión artificial	50
De la “oscuridad en el diseño” a la “transparencia algorítmica” y su fundamentación	52
La aprehensión jurídica del fenómeno del <i>big data</i> desde la privacidad y el nuevo <i>Reglamento europeo de protección de datos</i>	53
Los datos masivos rehúyen el régimen jurídico del derecho de protección de datos personales	55
La anonimización de los datos masivos y sus dificultades para escaparse del derecho de protección de datos personales	56
Graves dificultades de exigir el consentimiento para el tratamiento de datos masivos.....	57
Dificultades para la proyección de los principios del derecho de protección de datos personales	58
Dificultades para proteger los datos sensibles que se pueden generar a partir del <i>big data</i>	59
Una aproximación a los usos de la inteligencia artificial y el <i>big data</i> en la aplicación automatizada de la ley.....	60
Capítulo 3. La reconstrucción del sujeto y el Estado: <i>big data</i> y psicopolítica	65
Los datos y el <i>big data</i> : esfuerzo por la comprensión de lo inconmensurable.....	65
Transformación y destrucción creativa en el Estado del <i>big data</i>	71
Estado, poder y <i>big data</i>	76
Capítulo 4. El <i>big data</i> en la ciberdefensa y la ciberseguridad nacional versus el derecho a la privacidad del ciudadano colombiano.....	87
Aproximaciones al <i>big data</i>	88
Aplicación del <i>big data</i> en la defensa y la seguridad nacional	91
Objetivos del <i>big data</i> en la aplicación de seguridad y defensa	93
Criminología en la aplicación del <i>big data</i>	96
Uso fraudulento de la información en el contexto del <i>big data</i>	97
La importancia de la conciencia situacional del <i>big data</i> en el contexto de la seguridad y defensa nacional de los Estados.....	97
La creación de equipos multidisciplinares para crear un control efectivo en la aplicación de <i>big data</i>	99
La necesidad del Estado en el análisis de datos	100
Fuentes de imágenes y video.....	101

Procesamiento y uso de imágenes como herramienta para garantizar la defensa y seguridad de los Estados	102
Aspectos legales y éticos en el procesamiento de información <i>big data</i>	104
El derecho de seguridad y defensa versus el derecho de privacidad: una mirada desde la jurisprudencia de la Corte Constitucional colombiana	105
Conclusiones	109
Capítulo 5. <i>Open data</i> y <i>big data</i>. herramientas de <i>software</i> para ciudades inteligentes; (caso de estudio).....	111
Aporte de <i>big data</i> y <i>open data</i> a ciudades inteligentes.....	112
Contexto en Colombia	113
Contexto internacional	114
<i>Big data</i> y sus desafíos en las ciudades inteligentes.....	116
Tipos de datos	116
Evolución del termino.....	117
Características.....	117
Tecnológicas en <i>big data</i>	118
Aspectos legales.....	119
<i>Big data analytics</i>	120
Desafíos.....	123
<i>Open data</i> y sus desafíos en <i>smart cities</i>	124
Open data o datos abiertos	124
Características	126
Calidad de datos abiertos	126
Datos abiertos en Colombia	128
Datos abiertos en el mundo	131
Caso de estudio.....	134
Módulos de la herramienta	135
Módulo ‘Ingreso de estructura’	135
Análisis de resultados de la herramienta	139
Conjunto de datos acumulados	139
Conjunto de datos específico	140
Conclusiones	141
Conclusiones.....	143
Bibliografía	147

INTRODUCCIÓN

La globalización y la revolución de la información son dos fenómenos que han impactado profundamente en las estructuras social, política y jurídica. La computación, la tecnología digital, la robótica, la inteligencia artificial, la impresión 3D, la nanotecnología y la computación cuántica, entre otras, han posibilitado la intensificación de la comunicación y han puesto a la información en el centro de las principales discusiones y transformaciones sociales.

Las nuevas tecnologías de la información se han acoplado entre sí, produciendo una extensa red con gran cantidad de información que involucra procesadores de conocimiento en sistemas de alta complejidad cuya interacción se asimila a una red neuronal. Este sistema se apoya en la tecnología digital y, de este modo, permite la multiplicación exponencial de las posibilidades de comunicación e información.

La nueva revolución tecnológica implica avances en los medios de comunicación e innovaciones que adquieren su potencial de transformación social una vez son asimilados en el sector productivo. La industria que ha surgido a partir de las tecnologías de la información y la comunicación (TIC) funciona a través de estructuras altamente flexibles que pueden ser rápidamente modificadas o reordenadas para responder eficiente y rápidamente a problemas de diferente naturaleza y complejidad. Para ello, se basa en el uso y análisis de una ingente cantidad de datos que la sociedad, en todos los niveles, suministra por medio de las TIC y que

pueden ser interconectados y analizados para luego ser reutilizados. A esto se le conoce como el fenómeno del *big data*.

En este contexto, ha surgido un entorno de relacionamiento completamente nuevo en el que, por un lado, es posible conectar personas, información y estructuras institucionales en tiempo real, independientemente de los límites geográficos y temporales; y, por otro, conectar datos en sí mismos, que originan a su vez nuevos datos utilizados para infinidad de situaciones diferentes a aquellas en las que fueron generados. Así, el tránsito de lo analógico a lo digital ha hecho posible la emergencia de procesos autogestionables que han terminado por desplazar a los medios de comunicación convencionales, y con ello, facilitado el acceso a la información de forma extendida.

En esta nueva realidad, la centralidad de la información ha llevado a la emergencia de nuevos derechos; el acceso a la información ha representado riesgos para la seguridad de los Estados y, en simultáneo, un imperativo de transparencia y eficiencia para su funcionamiento. La posibilidad de acceder a las nuevas tecnologías se ha convertido en un indicador para evaluar el desarrollo de las naciones; pero esto significa tan solo la punta del *iceberg* de los usos que se pueden desarrollar. El derecho y la política no pueden abstraerse de estas discusiones, por lo cual han aportado diversidad de investigaciones acerca de los impactos jurídicos y sociales de las nuevas tecnologías; en particular, la problematización de la relación entre *big data* y derecho.

En las últimas décadas, académicos y universidades han centrado su atención en los alcances y las repercusiones de la información en la vida cotidiana y en la estructura social. Son abundantes los estudios acerca de lo que ha significado la cuarta revolución tecnológica en el mundo contemporáneo. El avance de la cuestión se ha producido a gran velocidad, tratando de seguir el paso de las innovaciones tecnológicas que aparecen y pronto son acopladas al sector empresarial, el Estado y la sociedad.

Este libro avanza en el marco de estas discusiones y es resultado de la primera parte de la investigación sobre derecho y *big data* realizada por el Grupo de Investigación en Derecho Público y TIC de la Facultad de Derecho, en cooperación con la Facultad de Ingeniería de la Universidad Católica de Colombia; investigación en la que, además, participa la Universidad de Texas at Arlington, Estados Unidos. Este proceso se viene desarrollando desde 2013 con investigaciones

previas —como la responsabilidad del Estado por el uso de las TIC y el análisis de la Ley de Transparencia y del Derecho de Acceso a la Información Pública—, con el propósito de ofrecer herramientas para el seguimiento, control y acceso a la información en Colombia, y para la contextualización de las diferentes problemáticas en torno al derecho y las TIC, que sirva no solo al contexto colombiano, sino también a otras naciones, para la adaptación de sus sociedades a esta nueva realidad.

Los diferentes capítulos de esta publicación reconstruyen y analizan los impactos de las nuevas tecnologías de la información desde una perspectiva transdisciplinar y multidisciplinar en torno al *big data* y su marco dogmático a nivel jurídico y político. El libro avanza en torno a categorías de análisis como *big data*, *gobierno abierto*, *internet* y *ciberdefensa*, a partir de disciplinas como la sociología, la ciencia política y el derecho, hasta llegar a la aplicación práctica de un *software* que sirva de herramienta para varias disciplinas —p. e., en el análisis de información pública—. De esta manera, se avanza en una perspectiva global y compleja acerca de la influencia de las nuevas TIC en el campo jurídico y en el Estado, analizando el desarrollo del *big data*.

El primer capítulo: “La cuarta revolución tecnológica: un nuevo paradigma de comprensión de la sociedad y el Estado más allá del *big data* y la *internet*”, presenta un breve recorrido por el desarrollo de la cuarta revolución industrial, evidenciando los principales cambios que se han producido en la sociedad derivados de la influencia de las TIC.

El segundo capítulo: “Marco teórico y aproximación jurídica al *big data*: algoritmos, inteligencia artificial y transformación digital”, continúa la línea argumental y profundiza en el examen minucioso de las características y los elementos constitutivos de las nuevas tecnologías de la información, evidenciando sus posibilidades en la economía, el derecho y la administración del Estado, así como las aplicaciones que en estos campos se vienen produciendo.

El tercer capítulo: “La reconstrucción del sujeto y el Estado: *big data* y psicopolítica”, hace una revisión de las características alrededor del concepto de *big data*, para luego problematizarlo en dos frentes: por un lado, en la perspectiva de la idea de *destrucción creativa* de Schumpeter; por otro, desde las ideas que conciben la psicopolítica como superación de la biopolítica, y los retos alrededor de la sociedad de la información.

El cuarto capítulo: “El *big data* en la ciberdefensa y la ciberseguridad nacional versus el derecho a la privacidad del ciudadano colombiano”, reseña el alcance de las TIC en la administración electrónica, a partir del análisis de las implicaciones éticas y legales en el tratamiento de la información. Luego avanza en el diseño de una propuesta desde la teoría fundamentada que posibilite la aplicación de la información en el conocimiento de los procesos sociales, así como en la defensa de la seguridad nacional, sin afectar la seguridad de los ciudadanos.

Finalmente, en “*Open data* y *big data*: una herramienta de *software* para ciudades inteligentes; caso de estudio” se presenta el aporte de ambos fenómenos a las ciudades inteligentes en un contexto nacional e internacional, así como los desafíos a los que estas se enfrentan. Este capítulo, además, describe el desarrollo de la herramienta para el análisis de datos del gobierno, al igual que su metodología y los módulos de la herramienta.

LA CUARTA REVOLUCIÓN TECNOLÓGICA: UN NUEVO PARADIGMA DE COMPRENSIÓN DE LA SOCIEDAD Y EL ESTADO MÁS ALLÁ DEL *BIG DATA* E INTERNET

Una revolución tecnológica

La cuarta revolución tecnológica, o la revolución de la información, produjo una serie de transformaciones que impactaron tanto la sociedad como la organización del Estado. Lo anterior, aunado a los procesos generados a partir de la globalización, presionó la reestructuración del Estado hacia estructuras flexibles de administración. Estas transformaciones se dan en el marco de una sociedad interconectada, donde se produce un permanente flujo de información. La posibilidad de que los ciudadanos interactúen entre sí y con las instituciones representa un reto para la estructura del Estado en el contexto de la globalización, en todos los aspectos, así como para todo el derecho en sus diferentes dimensiones (Sánchez, 2014, p. 27).

Este cambio se presenta, por ejemplo, en el *open gov* (conocido en Colombia como la política del Gobierno en línea) y afecta los diferentes actores involucrados en la relación entre el Estado y sus ciudadanos. Esto lleva a no tener clara

la responsabilidad de los proveedores de servicios de telecomunicaciones en sus actuaciones en este entorno (Becerra *et al.*, 2015, p. 97).

Otro aspecto grave que se ha reiterado en varios textos es el reclutamiento de niños y niñas en Colombia en diferentes contextos, como consecuencia del conflicto que vivió el país (Rodríguez, Chacón y Cubides, 2017, p. 34) y que se extendió a las redes sociales y otros medios de comunicación electrónica. Asimismo, un factor determinante del comercio ha sido el desarrollo de medios electrónicos para esta actividad, que generan problemáticas como las que expresa Flórez (2016): “El comercio electrónico puede implicar muchos tipos de transacciones en las que es imprescindible la definición de *documento original*, con el fin de resolver cualquier controversia legal” (p. 20).

Otro ejemplo de su aplicación en el mundo del derecho civil es lo concerniente a los derechos de autor y la problemática en la falta de control en la reproducción, el copiado o la citación. Como lo expresa Woolcott (2015, p. 35), se busca, tanto en los orígenes como en el desarrollo, proteger el espíritu de la obra y su conexión con el autor. Estos problemas de no fácil solución requieren una visión moderna y estrategias que hagan repensar el derecho, más allá de las concepciones y arquetipos formales:

El arquetipo formal de los sistemas jurídicos, distinto a la forma de positivismo jurídico arcaico visto en la tradición del formalismo, y que tampoco obedece a la noción estipulativa de la formalidad del derecho, se erigió bajo el ideal o la pretensión de completitud. (Agudelo, 2015, p. 53)

Al final, estos cambios quedan a voluntad de la propia sociedad, por cuanto, como lo expresan Otálora y Hernández (2018), “el derecho es una construcción social y es uno de los elementos constitutivos de la identidad de una población, la cual se desarrolla a partir de representaciones sociales y decisiones de los individuos que la habitan” (p. 223).

Estas características que se desarrollan a finales del siglo XX y durante la primera década del siglo XXI se entrelazan con la necesidad del Estado de articularse con ámbitos regionales y globales. De este modo, se produce un nuevo sentido en la administración del Estado conocido como *gobernanza*. La articulación entre los intereses del Estado, las economías globales y la posibilidad de que múltiples actores interactúen son los principales ejes de este modelo. En este

escenario es importante avanzar en el estudio de lo que significa la cuarta revolución tecnológica y los procesos históricos que han generado esta nueva etapa en las formas de comunicación. Comprender qué es y cómo actúa la revolución de las telecomunicaciones permite analizar las nuevas funciones del Estado y la forma en que las instituciones deben responder a presiones provenientes de una multiplicidad de actores sociales.

Asimismo, es fundamental avanzar en el estudio del impacto de las nuevas tecnologías de la información y la comunicación en la democracia y en el rol del ciudadano frente a la administración pública. El acceso a la información, la transparencia y la eficiencia son principios rectores sobre los cuales se produce la participación ciudadana y se posibilita el control del Estado, y por qué no, entender la virtualidad como un fenómeno que exige una nueva comprensión de la realidad (Carreño, 2016, p. 57).

Un mundo en cambio: contextos de la crisis

El final del siglo XX estuvo marcado por profundas transformaciones económicas, sociales y políticas. El crecimiento económico asociado a sectores como la construcción, cuyo mayor alcance se produjo en los noventa, mostró sus límites durante los primeros años del nuevo milenio, cuando el poder adquisitivo se dinamizó gracias a la masificación de la tarjeta de crédito, situación que llevará a la crisis hipotecaria del 2008. El *boom* de la construcción que se produjo entre 1996 y 2005 generó la baja en los precios de la vivienda y la expansión de los créditos a personas con poca solvencia económica, lo cual derivó en que estos fueran considerados de alto riesgo y, en consecuencia, se aumentara la tasa de interés. Entre 2005 y 2006 los deudores de baja solvencia económica empezaron a incumplir en sus pagos, llevando así a una de las mayores crisis económicas de los últimos años en Estados Unidos.

Desde entonces, la crisis económica ha sido un referente constante en las discusiones frente a la reconfiguración del Estado; y a estas se unen los debates sobre el cambio climático y los límites de la energía fósil. A finales de los ochenta, el Programa de las Naciones Unidas para el Medio Ambiente (PNUMA) y la Organización Meteorológica Mundial (OMM) crean el Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) con el propósito de contar con

información objetiva, sistemática y rigurosa sobre las repercusiones socioeconómicas y ambientales del cambio climático.

En 2007, en uno de sus informes, el IPCC anunció que, en promedio, la temperatura global entre 1906 y 2005 aumentó 0,74 °C, que la cubierta de hielos permanentes y de nieve ha decrecido a escala global y que el periodo transcurrido entre 1995 y 2006 se ubica entre los más calurosos en el registro instrumental desde 1850 (Greenpeace, 2009, p. 3). Lo anterior se produce en un contexto de desarrollo económico desigual en el que la globalización introduce una nueva preocupación sobre el carácter excluyente de las transformaciones tecnológicas.

La interconexión que implicó la globalización generó el acople de los mercados, y con ello, la preocupación por la distribución de recursos entre los diferentes Estados de acuerdo con su nivel de desarrollo. En este contexto, se planteó la necesidad de una especialización productiva a escala internacional cuya tesis principal era que “cada país se beneficiaría del comercio al especializarse en aquellas actividades en que es relativamente eficiente e intercambiar esas mercaderías por aquellas en que es relativamente ineficiente” (Kiljunen, 1986, p. 112).

No obstante, los debates sobre la división internacional del trabajo que se produjeron en los ochenta introdujeron el concepto de *heterogeneidad estructural* para indicar la profundización de la desigualdad producida por la especialización. Se resaltó de este modo que la dispersión del desarrollo tecnológico en las economías periféricas implica un amplio abanico de la productividad del trabajo y, por tanto, una desigual distribución del ingreso. De esta forma, en los noventa se aceptó que los países con aparatos productivos orientados tanto al mercado interno como al externo tenían mayores posibilidades de mitigar los efectos de la distribución desigual de los ingresos (Cortés, 2016, pp. 35-37).

Las últimas décadas del siglo XX y las primeras del siglo XXI avanzaron en torno a la idea de reorientar el modelo económico hacia procesos de desarrollo autosostenido a través de la industrialización y la definición de límites a la participación del Estado. La tesis sobre la cual se apoyó el nuevo modelo fue aquella según la cual altos índices de desigualdad reducen la elasticidad de la reducción de la pobreza y el crecimiento; esto es, que la pobreza es inversamente proporcional al crecimiento. Un estudio reciente de Berg, Ostry y Tsangarides (2014),

investigadores del Fondo Monetario Internacional (FMI), señaló que el crecimiento económico sostenido se da en mayor grado en países con distribuciones de ingresos menos desiguales (Azout, 2012). En esta línea, señalan los autores:

El hecho de que la igualdad parece impulsar un crecimiento más alto y más sostenible *no* respalda, en sí mismo, los esfuerzos de redistribución. En particular, la desigualdad puede impedir el crecimiento, al menos en parte, *porque* provoca esfuerzos de redistribución a través del sistema fiscal, que pueden socavar el crecimiento. En tal situación, aunque la desigualdad sea mala para el crecimiento, los impuestos y las transferencias pueden ser un mal remedio. Si bien la literatura sobre este punto es controversial, la idea de un *trade-off* entre redistribución y crecimiento parece estar profundamente arraigada en la conciencia de los encargados de política. (Berg, Ostry y Tsangarides, 2014, p. 54)

La formulación de políticas en países con altos niveles de desigualdad debe pasar por la implementación de estrategias para la reducción de la pobreza y el aumento de la redistribución, y a ello se aúna la consolidación de las herramientas para el crecimiento económico. Es así como los Estados han venido avanzando en torno a conceptos como la transparencia y la eficiencia en la administración pública. La gobernanza se convierte, entonces, en el nuevo pilar sobre el cual se fundamentan procesos democráticos y transparentes que buscan reducir problemas como la pobreza y la corrupción en el interior de los Estados. Ban Ki-moon (2009), Secretario General de las Naciones Unidas, expresa al respecto:

Infraestructura no es solo cuestión de carreteras, escuelas y redes de energía. Es, asimismo, el fortalecimiento de la gobernanza democrática y el Estado de derecho. Sin transparencia, no solo desde el gobierno hacia el pueblo, sino también entre el propio pueblo, no hay esperanzas de conseguir un Estado democrático viable.

Este proceso de transformación institucional lleva a la transición de un gobierno estatal comprendido como *government* a un gobierno ampliado conocido como *governance*, gracias al cual se produce una gestión moderna y eficiente de la autoridad y las políticas públicas. Se trata de considerar el papel del Estado tanto en su ámbito interno como en su sentido integrador en el escenario internacional, teniendo en cuenta multiplicidad de actores y redes de interacción (Reyes, 2017, pp. 142-161).

Entre la tercera y cuarta revoluciones tecnológicas

Durante los noventa, de la mano de las transformaciones sociales, económicas y del Estado, emergen los debates acerca del cambio tecnológico, algo que se difundirá como una nueva revolución industrial. De acuerdo con Rifkin (2011), “las grandes transformaciones económicas de la historia ocurren cuando una nueva tecnología en el campo de la comunicación converge con unos sistemas energéticos también novedosos” (p. 57). En este sentido, Castells (2006, p. 56) clarifica el sentido de la tecnología, al entenderla como el uso de conocimiento científico para especificar modos de hacer las cosas de maneras reproducibles. De esta manera, a partir de 1990 tiene lugar un proceso de transformación tecnológica que se expande rápidamente gracias a la globalización y a nuevas herramientas como internet y el lenguaje digital.

Asimismo, se ha producido un cambio en la actividad científica misma que se manifiesta en la tríada investigación-desarrollo-innovación. La estructura científica no se limita ya a la investigación, sino que trata de generar desarrollos tecnológicos que deriven en innovaciones de alcance práctico para las empresas, el mercado y la sociedad (Echeverría, 2005, p. 10). Por su parte, para Schumpeter (1961, pp. 132-136) la innovación se puede entender como la introducción comercial de un nuevo producto, mientras la invención es restringida al dominio de la ciencia y la tecnología. No obstante, en las últimas décadas se ha producido la convergencia de la aplicación tecnológica y la ciencia en un nuevo paradigma en el que se articulan la economía, la ciencia y el contexto sociocultural e institucional.

El carácter revolucionario en la aparición de nuevas tecnologías a finales del siglo XX y comienzos del siglo XXI se puede considerar a partir de seis puntos:

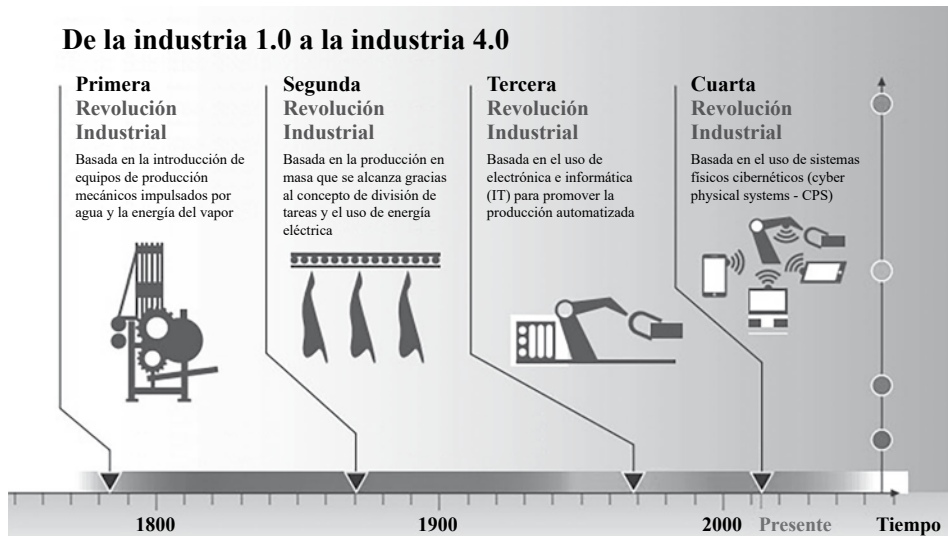
- a. Aparición de nuevos procesos tecnológicos y nuevas ramas de producción industrial.
- b. Disminución de la brecha entre los descubrimientos científicos y las innovaciones tecnológicas derivadas de estos.
- c. Impacto de la ciencia en la producción e implementación de laboratorios de investigación en las empresas.

- d. Intensificación del proceso de interacción entre diversas ciencias.
- e. La ciencia impacta en los ámbitos de la sociedad, el Estado y la economía.
- f. La ciencia orienta los procesos productivos, económicos y sociales a través de la formulación de sus bases teóricas (Schoijet, 1998, p. 136).

La tercera revolución tecnológica, o la revolución de la información, tiene como centro la aparición de microprocesadores, computadoras personales, *software*, equipos de telecomunicaciones e internet, cada uno con su propia trayectoria, pero en estrecha relación unos con otros y con un impacto significativo en los procesos productivos (Pérez, 2010, pp. 185-187). Estas tecnologías se venían desarrollando desde los cincuenta, pero fueron difundidas solo hasta que el sector productivo estuvo en capacidad de asimilarlas —lo cual ocurrió con la crisis económica de los setenta— y surgió la necesidad de adecuar el sistema económico a la generación de nuevas fuentes de energía. La tecnología que aparece en este contexto se caracteriza por emplear la menor energía posible y un mínimo de mano de obra; se trata de lograr un impacto generalizado en la vida cotidiana y en el Estado, y de asegurar una elevada productividad al emplear una menor cantidad de materias primas (Roel, 1998, p. 169).

Estas características son recogidas por Bell (2001, pp. 30-37) al describir lo que denomina la *sociedad posindustrial*, el advenimiento de una etapa histórica en la que tiene lugar la unión entre ciencia, tecnología y técnicas económicas, cristalizada en la fórmula de investigación y desarrollo (I+D), cuyo elemento diferenciador es su dependencia del trabajo teórico anterior a la producción, esto es, a la información necesaria (teoría) para llevar a cabo los procesos de creación de nuevas herramientas. La tercera revolución industrial generó el acoplamiento de los diferentes sectores en la empresa: se pasó de un sistema productivo basado en la segmentación o en la división del trabajo, a otro diferente basado en la integración de diferentes etapas, fases y tareas controladas por un solo punto encargado de todo el proceso. Para ello es indispensable reorganizar la estructura empresarial en función de las actividades por realizar (figura 1) (Roel, 1998, p. 181).

Figura 1. Revoluciones tecnológicas



Fuente: evaluandosoftware.com (s. f.).

Con la crisis de 2008, los procesos tecnológicos acaecidos durante la tercera Revolución Industrial se intensifican y generan una nueva ola de debates sobre el crecimiento económico y la dependencia de materiales fósiles como fuente de energía. Con la masificación de las nuevas tecnologías de la información y la comunicación (TIC), internet comienza a ser incluido entre los indicadores de desarrollo para los Estados. Como un reflejo de la primacía de la investigación y las nuevas tecnologías de información y comunicación como ejes dinamizadores de la economía, la Unión Europea (2014) plantea el impulso a la investigación y el desarrollo, así como la solución a problemas sociales como la pobreza y el desarrollo de nuevas tecnologías para el aumento de la competitividad, entre las cuales estarían la nanotecnología, la biotecnología y el desarrollo de las TIC.

Así, durante la primera década del siglo XXI se comienza a hablar de la cuarta Revolución Industrial, una nueva etapa caracterizada por la primacía de la información y de procesos asociados a las TIC: “La idea básica en que se fundamenta la Revolución 4.0 es la aplicación masiva de las nuevas tecnologías, con internet a la cabeza, a todos los procesos de una fábrica, de modo que el funcionamiento de la misma sea inteligente y absolutamente eficiente” (Castresana, 2016, p. 12):

En los primeros años del siglo XXI, para un creciente número de países —sobre todo los avanzados— la llamada Sociedad de la Información fue establecida como el siguiente punto de su itinerario. Con el impulso de un discurso que la presentaba como “el paso al progreso universal” gracias a la intervención de la tecnología, la nueva organización social se convirtió en la promesa del futuro a conseguir. (Alva de La Selva, 2015, p. 271)

Como característica central de la cuarta revolución industrial se puede señalar el carácter masivo que han adquirido las telecomunicaciones, llegando a posicionarse como dispositivos indispensables para la vida cotidiana. De igual manera, cabe señalar la lógica de interconexión compleja y en red de estas nuevas tecnologías, lo cual avanza de la mano de estructuras altamente flexibles que pueden ser rápidamente modificadas o reordenadas (Castells, 2006, pp. 87-90). No obstante, el eje principal del nuevo paradigma es la información, desde el *software* hasta el ADN, y su integración en prácticas en las que no es posible distinguir unos elementos de otros.

La información en cuanto centro de la innovación contemporánea produce un tipo de tecnología con dos ejes interconectados: el ordenador y la red; el primero aporta un componente individual, mientras el segundo se refiere a una característica relacional. Por esta vía se produce la ampliación de la gama de procesos posibles debido a la digitalización, lo que ha llevado a considerar a la revolución de la información como la era digital. El avance de la digitalización, la velocidad en la computación y la inteligencia artificial —cuyo principal exponente, pero no el único, es la personalización y la sugerencia basada en preferencias— han generado disminución en los costos de producción, aumento en la productividad y una mayor competitividad, por lo que los debates se han centrado en el avance del *machine learning* (Bericat, 1996, pp. 100-102).

En este contexto, se alerta sobre la posibilidad de que las máquinas sustituyan parte de la intervención humana mientras se minimizan los tiempos de producción y de desarrollo de productos. Asimismo, se hace cada vez más evidente la transformación del mercado ya no centrado en la oferta (vender lo que se fabrica), sino en la demanda (producir lo que el cliente va a comprar). Además del impulso por la eficiencia y la productividad gracias a la innovación tecnológica, también

es previsible que se produzca una reducción considerable en los costos de transporte, generando así una intercomunicación aun más profunda que materializa las expectativas de la globalización (De Pablos, 2016, p. 12).

La integración de los sistemas: más allá del *big data* y la internet

La primera década del siglo XXI ha estado marcada por el uso de la información descentralizada y la comunicación como centro del desarrollo industrial. En este escenario, el conocimiento es el eje del crecimiento económico que se manifiesta en la aplicación de nueve tecnologías características de la cuarta revolución industrial, a saber: integración horizontal y vertical de los sistemas, simulación, robots autónomos, análisis y *big data*, realidad virtual, fabricación adaptativa, la nube, ciberseguridad e industria del internet de las cosas (De Pablos, 2016, p. 13; Scalabre, 2018). Estas tecnologías comparten como característica primordial que en el centro de sus procesos se cuenta con núcleos complejos de información.

Una de las tecnologías características de la cuarta revolución industrial que permite a las empresas mejorar su productividad y conseguir mayor eficacia en sus procesos es la integración empresarial, que puede ser vertical u horizontal, según el propósito de la empresa. Pero, ¿qué es una integración?, ¿por qué la utilización de sistemas de información está relacionada con el concepto de *big data*? Para responder ambas preguntas, se expondrá una de las definiciones generales del concepto de *integración* y se relacionará con el de *big data*.

Según la Superintendencia de Industria y Comercio (s. f.), *integración* se refiere a “cualquier mecanismo utilizado para adquirir el control de una o varias empresas, el control de parte de ellas, o para crear una nueva empresa, con el objeto de desarrollar actividades conjuntamente”. Por ello, es válido afirmar que al utilizar un sistema de integración se requiere gestionar información de dos o más empresas; de ahí que la decisión de realizar una integración empresarial y el proceso de aplicación de este implique el manejo de un complejo de información que permita gestionar la empresa con miras a una mayor productividad.

Ahora bien, el concepto de *big data* nació como consecuencia de la proliferación y el uso de datos en el mundo; se ha concebido como “un fenómeno que da cuenta tanto de la gran cantidad de datos generados en ese ecosistema como de la complejidad de sus relaciones o de la velocidad en la que aparecen” (Puyol, 2015, p. 33).

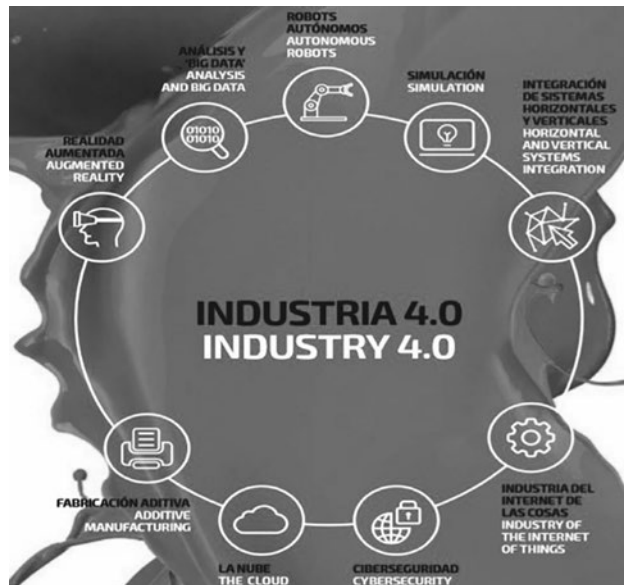
Asimismo, es un concepto que plantea la necesidad de sistemas y el uso de las tecnologías de la información para gestionar y analizar esta gran cantidad de datos; por esta razón, “el *big data* es una combinación de estas características que crea una oportunidad para que las empresas puedan obtener una ventaja competitiva en el actual mercado digitalizado” (p. 51).

Teniendo en cuenta que la utilización de sistemas de integración empresarial implica el manejo de un complejo de información que permita gestionar la empresa con miras a una mayor productividad, se puede afirmar que este concepto está íntimamente relacionado con el fenómeno del *big data*, y que ello no solo se basa en el análisis y la gestión de un complejo de datos, sino en las ventajas corporativas inmersas en el uso de los departamentos de tecnologías de la información para organizar datos sobre los negocios integrados.

De igual manera, y considerando que la globalización ha conllevado procesos de acoplamiento entre lo local y lo internacional, cada vez se exigen más procesos de comunicación eficientes y ágiles. La integración horizontal y vertical de los sistemas permite la automatización de los procesos a través de la integración de la información y la comunicación entre diversas unidades y funciones. Al mismo tiempo, internet de las cosas hace posible la innovación de materiales y herramientas que coadyuvan en la elaboración de análisis descentralizados, en tanto la nube posibilita el almacenamiento y la disposición de información más allá de los límites de la empresa, con el propósito de disminuir el tiempo de reacción y tomar decisiones con mayor agilidad.

Por otra parte, la simulación, los robots autónomos, la realidad virtual y la fabricación adaptativa permiten generar innovaciones en menos tiempo y con mayor calidad. Los procesos de producción transitan hacia dispositivos que permitan obtener mejores resultados aunque invirtiendo menos recursos; por ello, se promueven usos de la información que puedan generar sistemas de preferencias y automatización, con el propósito de eliminar los fallos del producto antes de que se puedan presentar. Estos avances posibilitan, además, la producción de materiales compactos y livianos que, por su proceso mismo de elaboración, tienden a privilegiar la demanda en el mercado, por cuanto se basan en el conocimiento sobre las preferencias de los clientes.

Figura 2. Nueve tecnologías de la industria 4.0



Fuente: The Boston Consulting Group (2018).

El carácter novedoso de estas tecnologías es que actúan en red y transportan micromateria (códigos y símbolos) a la velocidad de la luz. Los cuatro rasgos predominantes de estas redes son instantaneidad o comunicación en tiempo real, interactividad o comunicación bidireccional, virtualidad o amplitud comunicacional, y unicidad o integración comunicacional. Todo esto se encuentra asociado a la posibilidad de generar nuevos mecanismos de organización tanto en el ámbito político como en el económico, así como nuevas formas de relacionamiento que no dependen de una infraestructura física o de una territorialidad determinada:

La instantaneidad soporta la globalidad, en el sentido de que, en ausencia de distancia física y técnica, la comunicación se puede establecer, indiferentemente, con cualquier punto de la aldea global. Nuevas organizaciones de escala planetaria. Nuevas formas y fórmulas organizativas en el ámbito local y en el global. Por su parte, la interactividad contribuirá a la desmasificación de los medios. Frente a la unidireccionalidad de los medios de comunicación de masas, la bidireccionalidad de la red telefónica y telemática configura a todo elemento de la red como emisor/receptor de señales, no sólo receptor pasivo. La comunicación punto a punto fragmenta las audiencias masivas. (Bericat, 1996, p. 104)

Todo lo anterior ha generado una nueva economía en el marco de la globalización. Se han reducido las brechas entre espacio y tiempo, y con ello, la necesidad de aumentar la competitividad. El desarrollo de tecnologías como el *big data* y la nube ha posibilitado el vínculo entre la teoría formal y la acumulación de datos, impactando así las teorías econométricas modernas a través de la fórmula de investigación y desarrollo. La industria 4.0 emplea conocimiento previo al proceso de producción, por lo que privilegia conocimiento y habilidades como la inteligencia emocional y la solución de problemas complejos, situación que ha conllevado una creciente demanda en la cualificación de los trabajadores.

La nueva agenda económica incluye transacciones económicas a partir de servicios de información, formación y conocimiento; profesionales y técnicos ocupados y con disposición para aportar conocimientos; dinamización social por la innovación de conocimientos básicos; necesidad de controlar las consecuencias del conocimiento, y tecnología intelectual para tomar decisiones (Bericat, 1996, p. 107). De esta forma, el conocimiento se convierte en el principio axial de la sociedad y los centros de investigación en los ejes principales para el desarrollo de este conocimiento.

Se trata entonces de una tecnología intelectual desde la cual se generan normas para la solución de problemas a partir de algoritmos y cálculos matemáticos y estadísticos. Una situación determinada comporta límites y opciones contrapuestas, y en un esfuerzo por definir una acción racional e identificar los medios para llevarla a cabo, se pueden reconocer escenarios de seguridad, riesgo e incertidumbre. En este contexto, la educación también debe sufrir un proceso de ajuste con el propósito de generar el capital humano necesario para el mercado laboral de la tecnología intelectual:

Habrá que educar a las nuevas generaciones —y reeducar a las demás— en los nuevos valores que preconiza la Industria 4,0, como el aprendizaje y el uso de nuevas tecnologías, su aplicación práctica, el manejo de ingentes cantidades de información, la sostenibilidad de los recursos y la eficiencia máxima en su gestión. (Castresana, 2016, p. 46)

La seguridad se da cuando las limitaciones son fijas y conocidas; el riesgo, cuando se establece una serie de probabilidades en los resultados; y la incertidumbre, cuando las probabilidades son totalmente desconocidas. La tecnología intelectual hace posible prever las decisiones a partir de un sistema de preferencias

gracias a los desarrollos de la inteligencia artificial, disminuyendo la incertidumbre y el riesgo (Bell, 2001, pp. 48-53).

Se produce así una movilidad del capital que requiere la desregulación de los marcos normativos en los aspectos institucional y organizacional, y con ello, la transformación del Estado en su estructura y funciones. En este escenario emergen conceptos como aldea global y se pasa de la premisa del Estado mínimo a la demanda por un mejor Estado. Con estos términos se refieren cambios administrativos como la apertura de las economías nacionales, adición de precios a mercados internacionales, priorización de la economía de exportación, políticas fiscales y monetarias encaminadas a reducir la inflación y la deuda pública, vigilancia de la balanza de pago y la eficiencia en las finanzas del Estado, y la transparencia en el gasto público (Reyes, 2017, p. 50).

Las transformaciones del Estado: la influencia de la revolución tecnológica

El *big data* hace referencia a las “diferentes tecnologías asociadas a la administración de grandes volúmenes de datos provenientes de diferentes fuentes y que se generan con rapidez” (Kuan-Ching *et al.*, 2015). Como fenómeno, ha llevado a la innovación y producción de nuevas tecnologías en busca de la creación de sistemas que permitan la mejor gestión de la información. En concordancia con esta afirmación, UN Global Pulse (2012), una iniciativa de innovación de las Naciones Unidas, sugiere que hoy tiene lugar una “revolución industrial de los datos, precedida desde el comienzo del nuevo siglo por las innovaciones tecnológicas y los dispositivos digitales”.

Dentro de las innovaciones tecnológicas se encuentra el desarrollo tecnológico que busca maximizar la capacidad computacional, la precisión de los algoritmos (Boyd y Crawford, 2012) y facilitar el manejo de datos. Ejemplos de estos desarrollos son Hadoop, MapReduce y Cassandra. El primero fue “diseñado pensando en brindar poder de escalamiento desde un par de servidores hasta cientos de máquinas o nodos, las cuales manejan almacenamiento y procesamiento local” (Hernández, Duque y Moreno, 2017); el segundo, definido como “un modelo de programación que se ha asociado también a la implementación de estrategias de procesamiento de grandes conjuntos de datos que puede ser aplicado a una gran variedad de tareas del mundo real” (Dean y Ghemawat, 2008); y el tercero,

determinado como una base de datos que “brinda escalabilidad y alta disponibilidad sin comprometer el rendimiento, por lo que considera una plataforma ideal para tratar problemas de datos críticos” (Hernández, Duque y Moreno, 2017).

La globalización y la aparición de nuevas TIC generaron otras formas políticas cristalizadas en nuevas herramientas de gestión, la recomposición de la administración pública hacia modelos de gobernanza horizontal y cooperativa, y una ciudadanía orientada por la inclusión, la equidad y la participación eficaz. Todo esto lleva a la necesidad de plantear un Estado con las capacidades para concertar, coordinar y dirigir a la sociedad hacia sus metas de desarrollo (Rivera, 2010, p. 3).

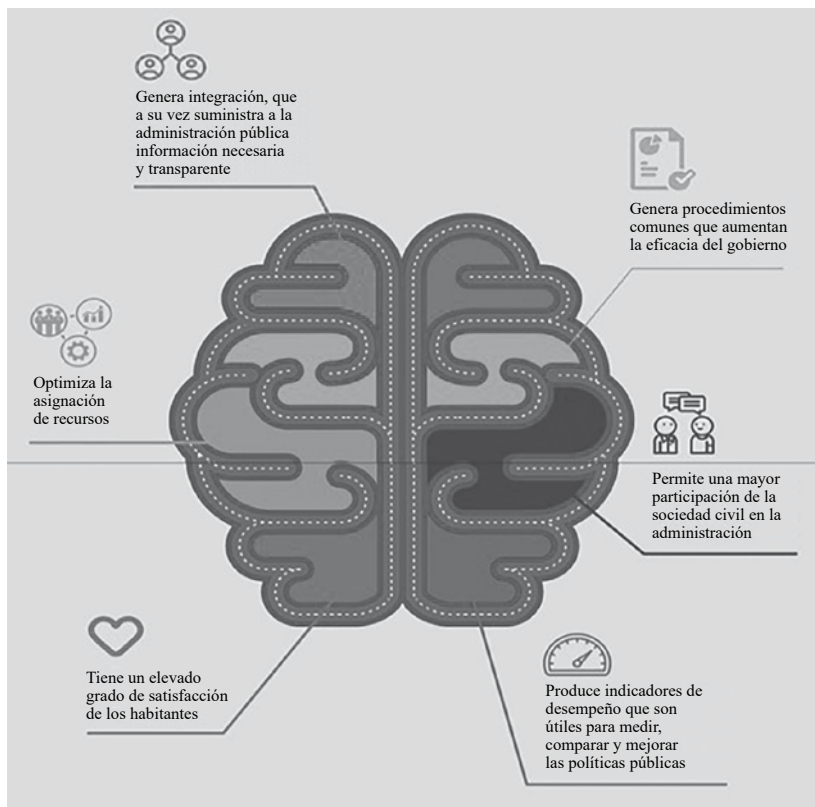
La gobernanza transforma no solo las estructuras organizacionales e institucionales, sino que afecta las prácticas sociales y los métodos de creación, acceso y divulgación del conocimiento. De esta manera, se puede definir la gobernanza como la organización de la acción o toma de decisiones colectivas, que incluye mecanismos formales e informales para el uso de las reglas, todo ello coordinado por actores estatales y no estatales (Reyes, 2017, p. 59).

Las nuevas TIC generaron el desdibujamiento de la frontera entre lo privado y lo público, toda vez que en el ámbito internacional no existe una autoridad única, sino un conjunto de actores que crean reglas para interactuar allí. La gobernanza actúa en este contexto para posibilitar la articulación entre lo local y lo global, por lo que se encuentra asociada a la capacidad de los gobiernos para conducirse a través de la participación de diversos sectores sociales (Reyes, 2017, pp. 59-79).

La cibernética y la inteligencia artificial aportan —desde una perspectiva similar a la de la gobernanza— elementos para generar el acople entre diversos niveles tanto de orden horizontal como vertical, con el propósito de consolidar sistemas flexibles con una capacidad de respuesta eficiente y ágil frente a problemas complejos. La inteligencia artificial permite pensar al Estado como una red neuronal de procesadores sociales de información y de conocimiento; esto es, como una red interconectada de agentes (Minsky, 1988, pp. 17-19). La cibernética mantiene el sistema en funcionamiento, lo cual permite realizar reajustes a los caminos iniciales o a los nuevos caminos donde ha habido desvíos u obstáculos, para adaptar el mecanismo a los objetivos establecidos (Bericat, 1996, p. 107; Bell, 1984, pp. 47-48).

Teniendo en cuenta los conceptos desarrollados a partir de la cuarta revolución tecnológica, se puede entender que el Estado reoriente su estructura hacia instituciones flexibles que buscan la resolución eficiente de conflictos y problemáticas sociales, políticas y económicas, a fin de avanzar en torno a objetivos concretos. En este esquema de funcionamiento es fundamental la retroalimentación, toda vez que permite generar los mecanismos de reajuste del sistema. El Estado, como la sociedad, es policéntrico, mientras que la información se presenta en mecanismos descentralizados (figura 3).

Figura 3. Gobierno digital



Fuente: Bouskela *et al.* (2016).

El *big data* resulta entonces esencial para el funcionamiento del Estado y para dar respuesta efectiva a las demandas sociales y presiones del contexto tanto en la

perspectiva local como internacional. Además, dispone de datos e información de forma tal que se facilite su procesamiento y análisis en tiempo real (Castresana, 2016, p. 32). Gracias a la innovación tecnológica que implica el uso de tecnológicas como la web 2.0, las redes sociales, el internet de las cosas o, simplemente, el correo electrónico, chats y blogs, se ha visto fortalecido el control ciudadano de los asuntos del Estado y la administración pública.

El gobierno electrónico, el gobierno abierto y el gobierno digital son avances en la estructura del Estado en cuanto a la posibilidad de la ciudadanía de acceder a la información. En este marco se produce una cultura “empresarial” en la que la eficacia y la capacidad de innovación se configuran como dos ejes esenciales a los que debe ajustarse el funcionamiento de todas las instituciones públicas y privadas (Estévez, 2009, p. 220).

Tanto el gobierno abierto como el gobierno digital cristalizan la posibilidad de que el Estado cuente con información permanente e ininterrumpida en función de que los ciudadanos tomen decisiones. Gracias a las nuevas tecnologías de la información, estas dimensiones implican, además, la posibilidad de que los gobernantes rindan cuentas sobre la administración pública. Lo anterior ha redundado en el fortalecimiento de la democracia y la gestión estatal desde concepciones colaborativas (Torres, 2016, p. 17). La transparencia se configura, por consiguiente, como uno de los valores más importantes de la acción desde la administración pública y una de las demandas más recurrentes desde la ciudadanía:

Se observan fenómenos similares con el derecho al debido proceso legal, que necesita una relectura de sus garantías, en muchos casos fijadas desde la tradición escrita, como soporte de legalidad y seguridad. La adecuación de mecanismos tecnológicos que permitan las mismas protecciones de contradicción, derecho a la defensa y respeto de las formas del juicio, pero en ambientes virtuales, son tareas que ya han iniciado un interesante camino de innovación. (Becerra *et al.*, 2015, p. 23)

Según lo anterior, es importante considerar las características de la información que es liberada y puesta a disposición de la ciudadanía, toda vez que esta puede referir datos e información general en tres niveles de profundidad, a saber: a) datos vinculados a las instituciones y su quehacer; b) grandes volúmenes de información en poder de la administración; c) información específica que se identifica como relevante frente a demandas sociales particulares o peticiones individuales de la ciudadanía (Villoria y Cruz, 2014, p. 13; Becerra *et al.*, 2015, p. 21).

El carácter de la información especifica un nivel de transparencia que evidenciará el desempeño de las entidades estatales: entre una transparencia clara que permite el acceso de la ciudadanía a la información de una forma clara y pertinente, y una transparencia opaca cuya difusión de información no aporta herramientas para el control ciudadano. De igual forma, desde la ciudadanía esta información puede generar una transparencia reactiva que se expresa en un bajo grado de participación en el ejercicio del derecho de acceso a la información, lo cual es evidencia de un sentimiento de apatía frente a los asuntos de carácter público.

La transparencia proactiva refiere un papel activo del Estado en la generación y difusión de información, lo cual contrasta con una baja comprensión y asimilación de estos datos por parte de la ciudadanía y, en consecuencia, en su latente dificultad para promover cambios en las políticas locales y nacionales. Finalmente, la transparencia colaborativa es el modelo hacia el que se orienta el Estado; implica recursos suficientes y necesarios para la difusión eficiente de la información, frente a una ciudadanía crítica y altamente participativa (Torres, 2016, pp. 19-20).

El Estado y el Gobierno encuentran límites a sus acciones desde la participación de la ciudadanía y la posibilidad de acceder a núcleos complejos de información. Así lo argumenta Estévez (2009):

Lo que se hace es instaurar una especie de mandato imperativo que vincula a los representantes políticos a las instrucciones de los grupos sociales dominantes. Este mandato imperativo se vehicula por medio del establecimiento de canales de participación de los grupos de interés en la elaboración de las normas generales y en el diseño e implementación de las políticas públicas. (p. 226)

La influencia de las TIC en la administración pública ha generado un marco normativo para posibilitar los procesos antes descritos. A través del uso de las nuevas herramientas de comunicación y el uso extendido de internet se ha hecho posible que los ciudadanos presenten solicitudes y peticiones a la administración, además de facilitar el acceso a datos que posibilitan el control y la veeduría ciudadana. Esto ha llevado a la necesidad de formular un marco jurídico que regule los plazos, términos y demás elementos que aseguran el adecuado uso de las TIC para el tratamiento de la información pública.

Así, el ordenamiento jurídico y los derechos en particular se han transformado para permitir la incorporación del uso de las tecnologías en la gestión gubernamental. De esta manera, se reconoce, por ejemplo, que el derecho de los ciudadanos a relacionarse con la administración pública a través de tecnologías de información debe ser tan amplio como lo permita la naturaleza del trámite y la pretensión de que se trate; esto es, de acuerdo con los marcos normativos establecidos para tal fin (Becerra *et al.*, 2015, p. 108).

En este sentido, cabe resaltar el papel del funcionario público como gestor de información encargado de los procesos de organización, conservación, uso y manejo de los documentos del Estado. Ello implica avanzar en una administración pública en la que el funcionario sea concebido como un agente en torno a un dispositivo articulado en clave de funciones. Asimismo, es importante considerar los retos que implican el manejo de la información y la posibilidad de la ciudadanía no solo de acceder a ella, sino de influir de manera efectiva en los asuntos del Estado.

El gobierno y la democracia en la era digital

Los cambios sociales y económicos que se han venido produciendo desde los primeros años del siglo XXI han presionado la consolidación de estructuras democráticas de participación. Las redes sociales, los chats y blogs han generado nuevas formas de comunicación que intensifican la comunicación social en una escala antes no vista. La manifestación de las demandas sociales ha alcanzado un eco importante en plataformas como Facebook y Twitter, donde además se ponen en circulación los comunicados más importantes de la administración pública, con el propósito de proporcionar información de manera ágil. Estas posibilidades de interconexión han generado un renovado interés por la democracia directa y la participación efectiva de la sociedad civil. Este es el sentido de la “aldea global”: la interacción comunicativa y la desmasificación de los medios de comunicación de masas, la dispersión del poder, una saludable desideologización y la revivificación de los regionalismos, en favor de una sociedad flexible con diversos grados de libertad (Bericat, 1996, p. 114).

Debido a la difusión de tecnologías digitales, el aumento en el número de usuarios de internet y el mayor uso de tabletas y teléfonos inteligentes que posibilitan el acceso a estas redes, se ha generado un flujo de información creciente y permanente. Este fenómeno ha modificado la participación de los ciudadanos, y con ello se ha transformado la democracia. La sociedad civil participa de forma eficiente y organizada en el control y monitoreo de la administración del Estado, identificando problemas e interactuando en diferentes niveles y por diversos medios para resolverlos. La integración de redes de comunicación, dispositivos inteligentes móviles, programas de análisis y sensores permite captar datos generados por diferentes agentes (personas o dispositivos); de este modo es posible procesar y generar informaciones, con miras a construir y aplicar ese conocimiento en la toma de decisiones, y así ofrecer mayor calidad de vida y beneficios a los ciudadanos (Bouskela *et al.*, 2016, p. 36).

El gobierno digital tiene cuatro etapas que fortalecen las relaciones entre el gobierno y los ciudadanos: la digitalización, la transformación, el compromiso y la contextualización. En la primera etapa, la información se digitaliza y los procesos se automatizan para facilitar su acceso; en la segunda fase se produce el rediseño de los procesos internos, las estructuras y las prácticas de trabajo de una organización de gobierno a través de la aplicación de tecnología digital; en la tercera etapa, el compromiso persigue la mejora en las relaciones entre los ciudadanos y el Estado. Como sostiene Estévez y Janowski (2016), esto último tiene propósitos concretos:

Aumentar el acceso, la conveniencia y la eficacia de los sistemas de prestación de servicios públicos, el involucramiento de los ciudadanos en los asuntos políticos y civiles, el aumento de la transparencia y de la rendición de cuentas de las operaciones del gobierno, y el desarrollo de ciudades inteligentes y sociedades de la información.
(p. 12)

En la última etapa, la contextualización, el gobierno digital traspasa los límites de lo local y se desplaza hacia lo global. En esta fase se busca la cooperación y el aprendizaje desde experiencias de desarrollo en otras regiones y Estados. De esta manera, la digitalización y la mejora en los procesos administrativos y las relaciones con los ciudadanos redundan en la posibilidad de que el Estado se inserte en un modelo de desarrollo económico de orden regional primero, y más tarde global.

•La cuarta revolución tecnológica•

Para avanzar en estos objetivos, el gobierno digital emplea plataformas colaborativas conectadas con la población por medio de aplicaciones móviles de fácil uso, en las que es posible interactuar con la administración pública y se tiene acceso a datos abiertos, *big data* y *analytics*, lo cual facilita tomar decisiones rápidas y eficientes, y extraer análisis predictivos (Bouskela *et al.*, 2016, p. 53). La democracia se plantea entonces desde el empleo de aplicaciones y redes de información que son posibles gracias a internet y al almacenamiento de datos en la nube. De esta manera, la democracia avanza hacia la digitalización, tal como lo hace el gobierno. La democracia digital se transforma, y con ello se profundiza la concepción de un ciudadano que participa activa y colectivamente en la toma de decisiones del Estado, a través de la integración en línea y el gobierno en línea. En este sentido, no solo se fortalecen los mecanismos de comunicación entre los ciudadanos y la administración, sino que es posible contar con una retroalimentación constante, fluida y ágil frente a las políticas públicas y las actuaciones del Estado:

[Mediante la democracia digital,] el Gobierno incentiva a la ciudadanía a contribuir en la construcción y seguimiento de políticas, planes, programas y temas legislativos, así como a participar en la toma de decisiones y, en general, está dispuesto a involucrar a la sociedad en un diálogo abierto de doble vía. En esta fase las entidades ofrecen específicamente información, servicios y trámites en línea para incluir a las poblaciones más vulnerables. (Ministerio de Tecnologías de la Información y las Comunicaciones, 2011, p. 7)

La inclusión digital se convierte en un derecho de vital importancia para que la población ejerza su ciudadanía y se inserte en la sociedad de la información, lo que implica el acceso real y efectivo a la información:

En el momento en el que la Administración tomó la decisión de usar la tecnología como medio de contacto con los ciudadanos, adquirió la obligación de capacitar y concientizar a la población para evitar la exclusión digital y lograr el desarrollo social. (Gómez y Pedreros, 2017, p. 176)

La participación de la sociedad civil constituye un elemento fundamental para el esquema de legitimación del gobierno digital y de las dinámicas del Estado en el contexto de las nuevas tecnologías de información. La apatía y la falta de confianza en las instituciones contrasta con la posibilidad de acceder a la información del Estado de manera confiable y poder ejercer control sobre los

recursos públicos, en lo que parece ser un nuevo impulso de la democracia ahora desde lo digital.

Hacia una gobernanza digital

La cuarta Revolución Industrial condujo a una serie de transformaciones en la estructura del Estado, la administración pública y la democracia. Las nuevas TIC generaron la posibilidad de acceder a información de manera permanente, y de este modo ejercer control sobre el gasto público, hacer seguimiento a las políticas públicas y, en general, interactuar de manera dinámica con las instituciones públicas. La gobernanza es un reflejo de la necesidad de conceptualizar al Estado en el marco de una red de relaciones en la que diversos actores tienen la posibilidad de influir en decisiones fundamentales para el gobierno de los ciudadanos. La articulación del Estado a la globalización y a las dinámicas económicas generadas a partir del siglo XXI se ha dado en sintonía con el desarrollo de nuevas tecnologías, las cuales incluyen en el modelo de desarrollo por el que habrá de orientarse la sociedad.

La democracia digital se ha convertido en un dispositivo mediante el cual se espera que la ciudadanía se empodere y tome decisiones de relevancia para los asuntos públicos, de manera eficiente y en un contexto de completa información. La transparencia se vuelve un elemento fundamental para evitar los fallos de la administración. La corrupción, el clientelismo y el manejo inapropiado de los recursos públicos son algunas de las problemáticas que, mediante el control ciudadano, podrían verse limitadas en favor de un mejor Estado.

Sin duda, el gobierno y la democracia digital son herramientas que favorecen la participación y representan un ejemplo claro de la cristalización de la influencia de los avances tecnológicos para el mejoramiento de la calidad de vida y la posibilidad de responder a problemas complejos. No obstante, es necesario profundizar en los efectos de estos avances en problemáticas sociales como la desigualdad, derivadas de la brecha tecnológica, con el propósito de consolidar los procesos democráticos que el gobierno digital pretende impulsar.

Finalmente, es evidente la influencia del fenómeno del *big data* en la revolución tecnológica, al ser un concepto que promueve el uso de las tecnologías de la información y la innovación de tecnología en el uso del manejo de datos;

•La cuarta revolución tecnológica•

por eso, “surge como una nueva fase del paradigma intensivo en información y comunicación que abarca no sólo su dimensión tecnológica, sino también una dimensión social, económica, política y cultural” (Facundo y Yoguel, 2016), como se expondrá en el siguiente capítulo.

MARCO TEÓRICO Y APROXIMACIÓN JURÍDICA AL *BIG DATA*: ALGORITMOS, INTELIGENCIA ARTIFICIAL Y TRANSFORMACIÓN DIGITAL*

Aproximación general al *big data*: premisas y cautelas

La expresión *big data* se debe al trabajo de Schönberger y Cukier (2013a). En español se utiliza también esta expresión inglesa. El Parlamento Europeo (2017) hace referencia a ‘macrodatos’, y no tanto a ‘datos masivos’:

Recopilación, análisis y acumulación constante de grandes cantidades de datos, incluidos datos personales, procedentes de diferentes fuentes y objeto de un tratamiento automatizado mediante algoritmos informáticos y avanzadas técnicas de tratamiento de datos, utilizando tanto datos almacenados como datos transmitidos en flujo continuo, con el fin de generar correlaciones, tendencias y patrones (analítica de macrodatos).

Se habla de las *V* (Gartner, 2012): volumen, variedad, velocidad y valor, a las que se añaden, entre otras, la veracidad (Puyol, 2014, p. 488). Se hace referencia al carácter *big* —‘grande’ o ‘masivo’— de los datos para hacer alusión al volumen, esto es, a cantidades de datos ingentes, de magnitud inabarcable a la

.....
* Apartes de este capítulo de libro fueron presentados y publicados por el Dr. Lorenzo Cotino en la sesión del Seminario de Derecho de la Universidad de Valencia, España; ponencia: “El Dret enfrontat las reptes del Big Data i l'automatització de les decisions”, del 26 de octubre de 2016, que puede consultarse en esta vínculo: <https://www.uv.es/seminaridret/sesiones2017/bigdata/ponenciaCotino2016.pdf>

mente humana, como la grandeza del espacio. Tales datos masivos pueden tener también la *V* por su variedad de fuentes y naturaleza; y se destaca asimismo la gran velocidad en la gestión y actualización de los datos. Los datos a los que se hace referencia pueden estar estructurados, no estructurados o parcialmente estructurados; no obstante, más allá de la cantidad o estructura, se destaca con el término *big data* elementos como la *V* del gran valor potencial presente y futuro; además, es nota diferencial de los datos masivos que no pueden ser manipulados, analizados, procesados con mecanismos o procesos tradicionales.

El gran reto de los datos masivos es la captación, gestión y tratamiento para agregar valor a grandes volúmenes de datos poco utilizados o inaccesibles hasta la fecha; todo ello, para aportar y descubrir un conocimiento hasta ahora oculto. Boyd y Crawford (2011, p. 6), entre otros autores, subrayan que el *big data* no solo se refiere a grandes conjuntos de datos y a las herramientas y procedimientos utilizados para manipularlos y analizarlos, sino a un giro en el pensamiento computacional y la investigación (Burkholder, 1992). Así como Ford cambió la forma en que se hicieron coches —y luego se transformó la forma de trabajo—, el *big data* va a cambiar toda la teoría social (Latour, 2009, p. 9) al automatizar tanto el objeto como el procedimiento del conocimiento.

El *big data*, en muy buena medida, está conectado con la inteligencia artificial cuando los sistemas computacionales son capaces de tratar, aprender, resolver problemas y tomar decisiones a partir de los grandes datos, desde un cambio de paradigma que automatiza tanto el objeto (los datos) como el procedimiento del conocimiento (Boyd y Crawford, 2011). Se aplican algoritmos, redes neuronales artificiales y patrones de razonamiento, en principio similares a los humanos (Nils, 1980). En esta dirección, el Parlamento Europeo (2017) recuerda que en algunos casos se capacitan “dispositivos de inteligencia artificial” como “redes neuronales y modelos estadísticos con el fin de predecir algunos acontecimientos y comportamientos”. Según el Parlamento Europeo (2017), los algoritmos pueden definirse como “conjuntos bien definidos de pasos para lograr una determinada meta” (Kroll *et al.*), o bien, como “código *software* que procesa un conjunto limitado de instrucciones”. Los algoritmos avanzados de aprendizaje automático (*advanced machine-learning algorithms*) destacan por su capacidad de adaptarse y aprender, y cabe recordar que los robots no precisan de una entidad física y,

•Marco teórico y aproximación jurídica al *big data*•

obviamente, tampoco de una forma pretendidamente humana, sino que cada vez más hay que percibirlos como algoritmos de autoaprendizaje.

No es difícil prever la importancia económica del *big data*; así, por ejemplo, para Europa algunos estudios (DemosEuropa, 2013) indican que en 2020 la combinación de *big data* y *open data* —especialmente el primero— implicará un crecimiento de 230.000 millones de euros, un 1,9% adicional al PIB. Ello implica un incremento adicional del PIB de un 23% en comercio, 22% en industria, 13% en finanzas y seguros, 13% en administración, 6% en sector TIC y 5% en sanidad y servicios sociales. Se afirma que un 50% de la economía europea queda afectada por el *big data* e incide en el crecimiento de un 5% o 6% de su eficiencia (Comisión Europea, 2013). Más recientemente se afirma que “el sector de los macrodatos está creciendo a un ritmo del 40% anual, siete veces más rápidamente que el del mercado de las tecnologías de la información” (Parlamento Europeo, 2017).

Dado el tratamiento económico y empresarial del *big data*, y en clave jurídica, cabe preguntarse sobre el equilibrio de contraprestaciones que se da entre los proveedores de servicios y usuarios a cambio del acceso y uso de datos que generan estos últimos y su empleo como datos masivos. Se trataría de especular sobre las consecuencias jurídicas que pueden darse y un posible reparto de la riqueza que se genera, al igual que sobre las exigencias y condiciones que los prestadores imponen a los usuarios.

¿De dónde proceden los datos masivos y todo lo imaginable que se puede hacer con ellos?

Los grandes datos son generados por humanos, también biométricamente, máquina a máquina, producto de grandes transacciones o del uso de la web y las redes sociales (Sunil, 2012). Billones de *whatsapps*, correos electrónicos, contenidos en Facebook y Twitter, búsquedas en Google y videos en Youtube se generan por la navegación en internet, las comunicaciones del internet de las cosas, las comunicaciones entre máquinas, industrias, estaciones meteorológicas, etc. Estas actividades, por lo general, están vinculadas a medidores y sensores de temperatura, luz, altura, presión, sonido, localización, GPS, así como en el entorno de tecnologías RFID, *wi-fi* o *bluetooth*, a lo que se suman los datos biométricos, normalmente vinculados al ámbito de seguridad, pero también de sanidad

(escáneres de retina, de huellas digitales, lectores de cadenas de ADN, monitoreos médicos de todo tipo, etc.).

Ahora bien, vale la pena cuestionar respecto a qué se hace con los datos; en palabras de Martínez (2014), a partir del clásico de Mayer-Schönberger, este “es un territorio abierto a la imaginación” (p. 3); se trata de una “estadística del todo”, porque el científico puede analizar todos los datos, eliminando el sesgo de la elección de una muestra. Además, según sean las posibilidades de tratamiento, se trata de combinar datos como el químico, que aleatoriamente va tomando muestras por doquier. Frente a la contrastación de una hipótesis a partir de los datos, se descubren correlaciones sin conocer previamente la causa. Así sucede al probar casi aleatoriamente la posible correlación entre datos en principio totalmente distantes (¿compra distintos alimentos la gente en función del estado del clima?, ¿cómo influye el embarazo en las decisiones de consumo?, ¿sería posible ofrecer seguros de salud en función de las búsquedas en Google o del análisis de las preferencias alimentarias manifestadas por los usuarios y por sus redes de amigos en espacios sociales de internet?).

Así las cosas, gracias al *big data* y la inteligencia artificial es posible generar patrones dinámicos de tendencias de futuro: la predictibilidad y el apoyo en la toma de decisiones, conocer mejor al cliente y el mercado, personalizar los productos y servicios, mejorar y agilizar la toma de decisiones, o prever el comportamiento (Agencia Española de Protección de Datos [AEPD] e ISMS Forum, 2017, pp. 6-7). El *big data* y la inteligencia artificial se proyectan en los sectores público y privado al ámbito empresarial, los recursos humanos, el marketing, el consumo, el comercio, el transporte, la sanidad, la educación, etc. Del lado de las ventajas, más allá de la perspectiva de intereses privados, sin duda pueden repercutir en ámbitos de la asistencia sanitaria, la lucha contra el cambio climático, la reducción del consumo energético, la mejora de la seguridad en el transporte y la posibilidad de establecer ciudades inteligentes.

De este modo, se busca la optimización y eficiencia de las empresas y contribuir a una mejora de las condiciones laborales, a la detección y la lucha contra el fraude, además de que los macrodatos pueden ofrecer una ventaja competitiva para los procesos de toma de decisiones (Parlamento Europeo, 2017). En cualquier caso, uno de los mayores retos es disponer de personas adecuadas y

•Marco teórico y aproximación jurídica al *big data*•

formadas para analizar y explotar los datos, esto es, convertir una gran cantidad de datos en decisiones, estrategias y mejores experiencias para los consumidores.

Algunas cautelas que plantea el big data y la necesidad de abordarlo jurídicamente

Como ya señalaron Boyd y Crawford (2011), los números no hablan por sí mismos; las afirmaciones de objetividad y precisión son engañosas dado que todos los investigadores son intérpretes de datos, y siempre hay un proceso de limpieza de datos inherentemente subjetivo. Apuntan que con los datos masivos también hay errores en internet, en razón de que interrupciones y pérdidas son a menudo poco fiables; así, los errores y lagunas se hacen también masivos. Se ha criticado igualmente que lo más grande no necesariamente es lo mejor.

El uso tecnológico genera datos vaporosos o *data fumes* (Thatcher, 2014), con todas sus limitaciones, sesgos y manipulaciones. Incluso se ha llegado a hablar de *Weapons of math destruction*, ‘armas de destrucción matemática’ (O’Neil, 2016). Como señala Surden (2017), lejos del “manto de objetividad de la tecnología” (p. 2), los sistemas tecnológicos pueden tener valores sociales incrustados o embebidos en su diseño, e incluso estos pueden ser contrarios a la igualdad, los principios constitucionales y los derechos humanos. En este sentido, la baja calidad de los datos o los procedimientos podría ocasionar algunos problemas:

Dar lugar a algoritmos sesgados, correlaciones falsas, errores, una subestimación de las repercusiones éticas, sociales y legales, el riesgo de utilización de los datos con fines discriminatorios o fraudulentos y la marginación del papel de los seres humanos en esos procesos, lo que puede traducirse en procedimientos deficientes de toma de decisiones con repercusiones negativas en las vidas y oportunidades de los ciudadanos, en particular los grupos marginalizados, así como generar un impacto negativo en las sociedades y empresas. [...] La información revelada por los análisis de los macrodatos no ofrece una visión general objetiva e imparcial de ninguna materia y que es tan fiable como lo permitan los datos subyacentes. (Parlamento Europeo, 2017)

Por su parte, Richards y King (2013) afirman importantes cautelas y una necesaria visión crítica frente a los grandes evangelistas del *big data*, esto es, los que prometen que los grandes datos pueden mejorar la toma de decisiones, por las mejores predicciones en áreas que van desde la admisión a la universidad, los servicios y las políticas médicas, hasta las políticas de seguridad nacional o prevención del delito. Los autores exponen la paradoja de la transparencia: mientras

que con los datos masivos se accede invasivamente a información privada, los resultados de estos tratamientos están casi completamente rodeados de secreto legal e industrial. Subrayan especialmente la paradoja que implica que un futuro casi milagroso y de transformación social sea al fin y al cabo un privilegio que queda en manos del gobierno y de grandes empresas a costa de los ciudadanos; todo ello, en el contraste de los riesgos de la identidad individual y colectiva. Lo anterior puede llevar a la “dictadura de los datos” (Cukier y Mayer-Schönberger, 2013b) o, como se ha señalado, a errores masivos y a una estupidez artificial, que sería, eso sí, muy humana.

Así las cosas, y como reacción frente a estos riesgos, se afirma la idea de una “responsabilidad algorítmica”, bajo la idea de que “normas científicas y éticas estrictas son fundamentales” (Parlamento Europeo, 2017); emerge así la necesidad de un “marco ético común sólido”, así como cuando se apela a “las normas éticas más elevadas”, especialmente aquellas que usen los macrodatos en la aplicación de la ley.

Aproximación a elementos de economía colaborativa y transformación digital afines a la investigación

Economía colaborativa es la expresión usualmente utilizada en castellano para designar lo que en inglés se conoce como *sharing economy*, *peer-to-peer economy*, *mesh*, *collaborative economy* o *collaborative consumption*; expresiones más o menos equivalentes, al margen de los particulares matices de significado. Con ellas se alude a los nuevos sistemas de producción y consumo de bienes y servicios que surgen a principios del siglo XXI y que aprovechan las posibilidades abiertas por los recientes avances de las tecnologías informáticas para intercambiar y compartir dichos bienes y servicios.

Por otro lado, de modo necesariamente vinculado, el proyecto de investigación conecta la economía colaborativa con la transformación digital de la economía. Ya no se trata de la mera informatización ni del uso de redes sociales o el *marketing* digital; tampoco consiste en la prestación de servicios a través de internet; las innovaciones y las TIC hacen reinventar la organización empresarial (y también las organizaciones públicas) para lograr los nuevos objetivos. La competencia es digital, lo que parte de la premisa misma de una estrategia

•Marco teórico y aproximación jurídica al *big data*•

digital de la organización, conjunta e integral: rediseño de procesos, de gestión de la información y documentación, etc. La orientación al cliente o usuario exige una constante retroalimentación y transformación continua, lo cual se proyecta en la cultura de la organización, y en su estructura y regulación.

Con la transformación digital emergen nuevas expectativas y experiencias de cara al consumidor o al administrado, las mejoras de eficiencia en la gestión y organización, la potencialidad de nuevos ingresos a la organización empresarial, el apoyo en la innovación constante para generar ventajas competitivas. En el caso de los poderes públicos, al mejorar los servicios se potencia la colaboración interna; de igual modo, la transformación digital impone el uso y la profundización del *big data*.

Ni la colaboración ni el uso de recursos compartidos son fenómenos en absoluto nuevos. La novedad de la llamada economía colaborativa y la transformación digital radica en el aprovechamiento de las posibilidades abiertas por las tecnologías informáticas que surgen a finales del siglo XX y principios del XXI; aprovechamiento que permite reducir considerablemente las asimetrías informativas y los costes de transacción que afectan a dichas actividades, incrementar la escala en la que se llevan a cabo y realizarlas de formas distintas a las hasta ahora conocidas. Los sistemas tradicionales de producción y consumo implican elevadísimos niveles de infrautilización de ciertos recursos y de sobreexplotación de otros. El caso de los autos familiares es paradigmático: se calcula que estos permanecen aparcados entre el 92% y el 96% de su vida útil y que un porcentaje sorprendentemente elevado del suelo urbano se dedica a este fin. Los nuevos sistemas de compartición de automóviles permiten aminorar esos porcentajes; tal compartición hace posible satisfacer las mismas necesidades de movilidad con un menor número de coches y de plazas de aparcamiento, lo cual mitiga la infrautilización de los coches y la sobreexplotación del suelo urbano.

Actores, actividades y sistemas de economía colaborativa

Los actores, las actividades y los sistemas de economía colaborativa pueden ser de diversos tipos. Las actividades pueden realizarse con ánimo de lucro (es el caso de Uber o Airbnb) o sin él. Algunas de las empresas que han florecido en este ámbito proporcionan a los consumidores acceso a determinados bienes o

servicios a través de una plataforma tecnológica que simplifica enormemente las correspondientes transacciones (Zipcar); otras se limitan a hacer de intermediarias, por ejemplo, a través de algún dispositivo tecnológico y entre sujetos que intercambian bienes o se prestan determinados servicios (Uber).

Algunas de estas actividades se desarrollan en ámbitos que hasta la fecha han carecido de una regulación jurídica específica (como los recados Taskrabbit); otras, en cambio, se llevan a cabo en mercados que han estado sometidos a una fuerte intervención pública (como los del taxi o los alojamientos), lo que ha restringido de manera muy notable la libertad de los particulares para realizarlas. De igual modo, no hay que desconocer el papel esencial de los poderes públicos y las administraciones: en primer lugar, dada la heterorregulación del sector y el desarrollo de políticas públicas que afectan la transformación digital; en segundo lugar, por cuanto el sector público es esencial en la economía y en el consumo, e incluso de producción de servicios en la economía digital; finalmente, por la gestión de las administraciones públicas, la redefinición de sus funciones y el ajuste competencial de estas en los distintos niveles de gobierno.

La necesaria aproximación jurídica y nuevos enfoques desde los derechos fundamentales

Los ámbitos sensibles cercanos al interés del jurista no son pocos, aunque por su novedad no siempre es sencillo delimitarlos. Antes de poner palos en las ruedas de la innovación y el avance; antes de, en palabras de Tene y Polonetsky (2013), vilipendiar el *big data* e imponer una regulación de mano dura; antes de eso, cabe establecer directrices, regulaciones legales y técnicas para limitar usos poco éticos, contrarios a derechos fundamentales y principios, en especial aquellos vinculados con la no discriminación y la privacidad, así como fortalecer el control y las garantías del individuo.

De igual modo, a partir de la nacionalidad y ubicación de los mayores generadores de datos masivos, cabe discutir las ventajas y los inconvenientes de los modelos de autorregulación o heterorregulación, incluso modelos de regulación borrosa o nebulosa, como se dan en ámbitos próximos como el de la nube (Cotino, 2015). Cabe preguntarse, en razón de los derechos e intereses en juego y la naturaleza de la relación jurídica, cuál sería la tipología de fuentes apropiada:

•Marco teórico y aproximación jurídica al *big data*•

autorregulación, heterorregulación, códigos, normas de conducta, etc. La eficacia real de las decisiones que se adopten depende de este tipo de elecciones, al igual que del papel del derecho nacional y del derecho supranacional e internacional ante un fenómeno transnacional. Incluso, sobre la base de existencia de normas, hay que abordar jurídicamente el tratamiento transnacional, ley aplicable y jurisdicción, así como negociabilidad del foro, de cara a otras empresas con las que contraten y respecto a los usuarios que generan los datos masivos. Debe tenerse en cuenta también el potencial del futuro papel del *big data* y la inteligencia artificial para la misma generación y creación de normas (*algorithmic regulation*) (Coglianese, 2017; Alarie, 2016).

Para ello, una de las premisas jurídicas es determinar y diferenciar el tratamiento jurídico de la actividad de *big data* cuando se realiza ya sea por poderes públicos o por el sector privado. El marco jurídico puede ser diferente a partir de la responsabilidad del Estado, el principio de legalidad y el interés público, frente a la libertad de empresa y los derechos en juego por el sector empresarial. Trátese del sector público o del privado que realicen acciones de *big data*, hay que plantearse la discrecionalidad o potestad para usar y tratar los datos masivos, la protección jurídica que tienen respecto a los métodos, las tecnologías y los resultados del *big data*; en especial debe tenerse en cuenta la propiedad industrial y la concurrencia de posibles obligaciones de transparencia y puesta a disposición de los datos abiertos para su reutilización.

Sin perjuicio de la necesidad de sólidos presupuestos éticos, el punto de partida jurídico deben ser los principios comunes del derecho constitucional y, en particular, los derechos y las libertades de las sociedades democráticas. Ellos no solo quedan afectados por el *big data* y la inteligencia artificial, sino que son principios básicos que han de orientar las respuestas futuras. Ahora bien, puede ser necesaria una readecuación de tales principios (Sánchez, 2016), incluso hay que plantearse el reconocimiento de nuevos derechos, como podría ser un derecho a la criptografía. Se trataría de una actualización o nueva versión que posibilite la privacidad como el derecho a dejar a la persona sola (Warren y Brandeis, 2010), o, en términos más castizos, el derecho a dejar a la persona tranquila o en paz.

Sin embargo, también es necesario un nuevo enfoque jurídico y dogmático en el tratamiento de los derechos fundamentales. El daño individual producido por el *big data* y la inteligencia artificial puede ser imperceptible para el derecho

fundamental desde la perspectiva del individuo titular del derecho, pero bien puede afectar masivamente a los derechos fundamentales de sectores o conjuntos de la sociedad de una manera relevante en esta dimensión colectiva. Y aunque puede ser necesario trabajar con una dimensión colectiva de los derechos —que no es lo habitual—, no hay que eludir la potencialidad jurídica de los elementos y las premisas de los derechos fundamentales concretos, como la dignidad y el libre desarrollo de la personalidad (art. 10.1 CE), pues cabe acudir a estos elementos esenciales para afrontar algunos retos jurídicos.

La doctrina no ha subrayado esta cuestión respecto al *big data*, solo ha conceptualizado alguna medida respecto a la privacidad (De Tullio, 2016); sin embargo, con una percepción práctica, el Parlamento Europeo (2017) acaba de indicar la necesidad de garantizar efectivamente —e incluso judicialmente a través de diversos derechos fundamentales— el uso del *big data*, siempre que repercuta de manera relevante en las personas.

Datos masivos, discriminaciones y brecha masivas

Cabe tener precaución por cuanto el *big data* crea nuevas brechas digitales. Tene y Polonetsky (2013) describen tres clases de personas en el ámbito de *big data*: los que generan los datos (consciente o inconscientemente), los que tienen los medios para recoger los datos y aquellos que tienen experiencia para analizarlos. Obviamente, estos últimos son los privilegiados de este nuevo mundo; son quienes fijarán las reglas reales de cómo se utilizarán y quiénes accederán al conocimiento; son quienes “pueden leer los datos”, esto es, quienes tienen el conocimiento y los medios para realizar el tratamiento masivo de datos, y por eso pueden imponer barreras de acceso o limitar selectivamente el acceso al conocimiento generado. Entre tanto, quienes no pueden acceder, los excluidos, no pueden evaluar la calidad y el valor de los datos masivos y los análisis. Se crean así nuevas jerarquías políticas, económicas y sociales. Puede hablarse, entonces, de un *big data* rico y un pobre *big data*, y ello no escapa al mundo universitario y de la investigación, por cuanto los que no tengan acceso a las cuotas no podrán utilizar el conocimiento, pero tampoco evaluar la calidad metodológica de los productos del *big data*.

Por último, como nuevos parias están quienes quedan en la periferia de los grandes datos, pues no los aportan. Como recuerda Lerman (2013), millones de personas en todo el mundo permanecen en la periferia de los grandes datos; así, sus preferencias y necesidades están en riesgo de ser ignoradas en las decisiones que se basen en el *big data* y la inteligencia artificial. El *big data* y los algoritmos pueden heredar o reflejar prejuicios y patrones de exclusión, o bien, ser resultado de quienes han tomado decisiones anteriores (Barocas y Selbst, 2016, p. 675). Más allá de la intencionalidad —que es muy posible que no se dé en muchas ocasiones—, se trata de un peligro objetivo que hay que prevenir.

Frente a estas situaciones, es necesario redescubrir nuevas proyecciones de la igualdad y el derecho a la no discriminación, pues las dificultades para actuar, vigilar, controlar y corregir son complejas. Selbst (2017) apunta para Estados Unidos la necesidad de introducir técnicas como la *discrimination impact assessments*, siguiendo el modelo de declaraciones de impacto ambiental. Por otro lado, otros autores afirman la necesidad de una ética en el diseño (Edwards, s. f., p. 30) o incluso una nueva “evaluación del impacto social” (Edwards, McAuley y Diver, 2016).

Precisamente, estas técnicas preventivas son bien conocidas en la UE en el ámbito de la discriminación, con numerosas directivas y legislación interna (Ley 4 de 2005); por otro lado, son relevantes los mecanismos *privacy impact assessment* (PIA) en el ámbito de la inteligencia artificial. El nuevo *Reglamento europeo de protección de datos* apuesta por mecanismos proactivos y preventivos en vez de reactivos. Así, impone la protección de datos desde el diseño y por defecto (art. 25), de modo que la privacidad se integre, desde el inicio, en la gestión y el ciclo de vida del tratamiento de datos. Lo mismo debe postularse respecto a la no discriminación.

En particular, es necesario integrar estas medidas preventivas de no discriminación en los casos en los que es exigible una evaluación de impacto (artículo 35 del Reglamento). Precisamente, los usos del *big data* abren la posibilidad de que sea obligatoria dicha evaluación, por cuanto suelen suponer la elaboración de perfiles, especialmente si sobre el resultado del tratamiento se basan decisiones que producen efectos jurídicos sobre el individuo o que lo pueden afectar de manera significativa (artículo 35, tercer Reglamento; AEPD e ISMS, 2017, p. 21). Hay que apostar por ir más allá de la protección de datos e integrar en estas

medidas la no discriminación; se hace referencia a la “discriminación y el sesgo algorítmico” (Parlamento Europeo, 2017).

Se entiende que este tipo de garantías va a tener que combinarse con el reconocimiento de fuertes poderes de acceso al conocimiento de algoritmos, datos y potestades de control sobre los agentes claves en el *big data* y la inteligencia artificial. Y ello debe quedar principalmente a cargo de autoridades independientes y cercanas al sector IT, como la Comisión Federal de Comunicaciones (FCC, por siglas en inglés) en Estados Unidos, o las autoridades de protección de datos y transparencia en la UE. En buena medida, así ya lo propuso la FCC en 2013 y recientemente lo han recordado AEPD e ISMS (2017, p. 26).

El Parlamento Europeo, además de un “marco ético común sólido” o de “máxima prudencia”, hace referencia a la necesidad de “evaluaciones periódicas sobre la representatividad de los conjuntos de datos [y de] examinar la exactitud e importancia de las predicciones” (Parlamento UE, 2017, cons. 20).

De la personalización masiva de la información al fin de la esfera pública y la posverdad: libertad de expresión artificial

La tendencia general a la personalización de servicios se da, singularmente, respecto al acceso a los contenidos de los medios de comunicación a través de internet. Desde el 2000 se discute si internet en general y los servicios de personalización masiva y filtrado selectivo de contenidos son positivos o no para la esfera pública y para la democracia deliberativa habermasiana (Cotino, 2013). En este sentido, del lado más pesimista —quizás el enfoque mayoritario—, Turow (1997) y especialmente Sunstein (2001) sostienen que la personalización de contenidos conlleva una limitación del mercado de las ideas que es esencial en una sociedad libre; se considera que la personalización refuerza las posiciones particulares, sin apertura ni compromiso con lo diferente; es más, se sostiene que la particularización de contenidos conlleva la desaparición del foro público. En la red y con los sistemas de personalización se crean enclaves deliberativos que no hacen sino reforzar las posiciones individuales, contribuyendo a extremar y polarizar la esfera pública (Sunstein, 2001, p. 67).

Por el contrario, y del lado más optimista, Blumler y Gurevitch (2001), Gimmler (2001) o Kellner (2004) consideran que internet amplía la esfera pública,

•Marco teórico y aproximación jurídica al *big data*•

propicia que los usuarios se encuentren y se relacionen con una gran diversidad de información y opiniones diferentes a las que habitualmente se encuentra en la vida *off line*. En esta dirección, internet propicia el hallazgo fortuito de ideas e información diferentes a través de sus enlaces.

Resulta extremadamente difícil mensurar objetivamente los efectos de la realidad de la selección y personalización de contenidos; sin embargo, no cabe duda de que el *big data* permite construir masivamente las realidades que nos circundan. Así las cosas, algunas cuestiones jurídicas que plantea el *big data* y la inteligencia artificial están relacionadas con la posverdad. Asimismo, el Parlamento Europeo (2017) ha mostrado su preocupación: el riesgo de manipulación de las personas en razón del uso del *big data* y la generación y empleo de patrones puede ser importante, más aún si se tiene en cuenta que se puede manipular el mundo virtual que nos rodea y conformar una realidad objetiva y subjetiva para el sujeto, en razón de los procesos de personalización masiva de la información que hoy se permiten con las redes sociales y el *big data*. Solove (2014), por ejemplo, demostró cómo se puede manipular el estado de ánimo de las personas a través de la personalización masiva de noticias e ingeniería semántica, transmitiendo noticias más o menos positivas, etc. Ello suscita cuestiones del todo interés desde parámetros de libertad y libre desarrollo de la personalidad.

En este sentido, Martín (2013) hace referencia a la inducción de comportamientos a través del uso de datos masivos. Al fin y al cabo, con el *big data* se puede inducir a un comportamiento: por quién votar, qué estudiar, dónde viajar, dónde vivir, qué comprar...; es más, se puede inclinar más claramente hacia la ilegalidad, la xenofobia, el odio, la discriminación, etc. Como recuerda este autor, hay una frontera muy fina entre, por un lado, influir en las decisiones a través de argumentos o información que permite valorar y tomar una decisión libremente, sobre la base de un juicio de valor formado por estímulos externos, y, por otro, generar o inducir tales decisiones. En este último caso, la libertad en la toma de una decisión puede estar viciada por el uso de técnicas poco transparentes que inducen al individuo a adoptar determinado comportamiento o decisión de manera inconsciente.

Así las cosas, resulta de interés preguntarse si los contenidos automatizados generados por máquinas están protegidos por la libertad de expresión (Wu, 2012), y cabe en principio dar una respuesta afirmativa desde la perspectiva

de las finalidades y funcionalidades del libre flujo de información en democracia (Massaro, Norton y Kaminski, 2016, 2017), si bien ello precisa de no pocas matizaciones para el futuro. También son muy relevantes las cuestiones que se suscitan por la responsabilidad ante los contenidos automatizados (Cheung, 2015). Igualmente hay que examinar con atención el uso de algoritmos de censura automatizada y la dificultad misma en su investigación (Narayanan y Zevenbergen, 2015).

De la "oscuridad en el diseño" a la "transparencia algorítmica" y su fundamentación

Stutzman y Hartzog (2013, p. 386) hacen referencia a la oscuridad por diseño (*obscurity by design*); la falta de transparencia resulta uno de los problemas claves para afrontar los usos del *big data* y la inteligencia artificial. Recientemente se aboga por la responsabilidad y la transparencia algorítmica (Parlamento Europeo, 2017); resulta casi un privilegio acceder a los algoritmos, los datos y el conocimiento generado, que incluso quedan fuera del acceso de los poderes públicos, pese a la transcendencia de su uso público y privado. Es más, los poderes públicos no siempre acceden, incluso en los casos en los que ellos mismos utilizan el *big data* y la inteligencia artificial en sus funciones públicas administrativas, policiales o judiciales. Han pasado de crear y controlar los instrumentos de acción a ser consumidores de los que les ofrece el sector privado (Joh, 2017). Los mismos científicos y académicos quedan en riesgo de exclusión; incluso para acceder al conocimiento se ven obligados a colaborar con el sector privado (Pasquale, 2015). Por ello, debe darse una respuesta jurídica y fundamentar cómo y en qué medida se puede acceder a tal información, algoritmos y conocimientos generados.

En la tradición jurídica angloamericana, el *due process* puede fundamentar dicha transparencia respecto a toda decisión basada en el *big data* y la inteligencia artificial que afecte a los derechos de las personas (Boyd y Balkin, 2017). En el ámbito europeo, el derecho de acceso que forma parte del derecho a la protección de datos también puede ser importante. En el ámbito anglosajón, desde una perspectiva amplia de los contenidos del debido proceso, cabe destacar el esfuerzo de Crawford y Schultz (2014) para vincular un derecho fundamental a las importantes implicaciones del *big data*; precisamente hacen el esfuerzo al partir de que la privacidad y los datos personales quedan en buena medida superados.

•Marco teórico y aproximación jurídica al *big data*•

Pues bien, la idea básica es que en la tradición jurídica angloamericana el debido proceso prohíbe al Gobierno privar los derechos de un individuo a la vida, libertad o propiedad sin brindar acceso a información sobre ciertos componentes básicos de procedimiento del proceso de adjudicación, así como los derechos para revisar y disputar la prueba en cuestión. La idea parte de que el uso del *big data* en la toma de importantes decisiones debe nutrirse de garantías. Así, sería desarrollar el alcance del debido proceso para los sistemas informáticos de la administración pública (Steinbock, 2005; Citron, 2010), de modo que cualquier persona evaluada o determinada en decisiones que le afecten por los datos masivos tuviera ocasión de acceder a la información al respecto. Se trataría de derechos similares a quienes son juzgados por los tribunales.

Pues bien, en sintonía con Miralles (2013), se propone el reconocimiento de un “derecho de acceso amplificado”, no limitado exclusivamente a los datos de que dispone un responsable de tratamiento, ni siquiera a los que les ha comunicado, sino sobre qué tratamientos concretos ha aplicado a los datos, qué información calculada se ha obtenido a partir de estos y cuáles han sido los usos concretos (incluso respecto a las operaciones de disociación de la información). De esta manera, se daría respuesta a los poderes derivados de la autodeterminación informativa, como un mecanismo preventivo ante posibles tratamientos opacos de los datos personales. Estas exigencias de transparencia y conocimiento se refuerzan jurídicamente con el derecho de acceso, que forma parte del derecho a la protección de datos y está especialmente reconocido en el ámbito de la UE.

La aprehensión jurídica del fenómeno del *big data* desde la privacidad y el nuevo Reglamento europeo de protección de datos

Sin perjuicio a los derechos fundamentales mencionados, se requiere abordar jurídicamente el *big data* desde la privacidad y, en concreto, desde el derecho de protección de datos personales, a modo de una *vis atractiva*. Más allá de la propia exigencia jurídica, se ha afirmado que el respeto de la privacidad ante el *big data* es una obligación moral de la sociedad (Allen, 2016). Krotoszynski (2015) subraya que el respeto de la privacidad ante el *big data* es una premisa esencial para que sea posible cualquier deliberación democrática y del ejercicio de la libertad de expresión vinculada a ella.

Se trata de un derecho joven y en plena expansión. Es muy difícil que algún ámbito de la realidad no quede vinculado a este derecho que afecta el tratamiento de cualquier información relativa a una persona física de modo más o menos estructurado:

[Este derecho confiere] poder de control sobre sus datos personales, sobre su uso y destino, con el propósito de impedir su tráfico ilícito y lesivo para la dignidad y derecho del afectado. [...] Atribuye a su titular un haz de facultades consistente en diversos poderes jurídicos cuyo ejercicio impone a terceros deberes jurídicos [...]: el derecho a que se requiera el previo consentimiento para la recogida y uso de los datos personales, el derecho a saber y ser informado sobre el destino y uso de esos datos y el derecho a acceder, rectificar y cancelar dichos datos. (Tribunal Constitucional Español, 2000)

El derecho de protección de datos personales, hasta la fecha, ha girado estructuralmente en el consentimiento del titular de los datos personales, pues teóricamente es él quien debe consentir el tratamiento de datos y tutelar bajo su voluntad sus derechos ARCO (acceso, rectificación, cancelación y oposición); no obstante, por diversos motivos, esta columna vertebral de la protección de datos quebró hace tiempo: en primer lugar, porque son legión las excepciones legales a la necesidad de consentimiento informado para la recogida o la cesión de datos; en segundo lugar, no hay que olvidar la STJUE, de 24 de noviembre de 2011, por cuanto es posible no exigir el consentimiento del afectado si el tratamiento se justifica en el interés legítimo del responsable del tratamiento. Es cierto que, como pone de manifiesto la Sentencia del derecho al olvido, de 13 de mayo de 2015, el tratamiento de datos personales masivos por el buscador Google no queda legitimado suficientemente por el mero interés económico.

En esta línea, Oliver y Muñoz (2014) señalan que es un hecho la banalización del consentimiento como mecanismo de tutela del derecho. No es realista, en modo alguno, creer que existe un efectivo control de la información personal a través del consentimiento y los derechos que lo complementan. El consentimiento se torna en una carta blanca al descontrol del flujo de los datos personales, y así acaba configurándose como un simbolismo que conlleva, a la postre, el fracaso de la privacidad pretendida y la inoperancia del sistema de protección. Ello debilita toda garantía jurídica que dependa del consentimiento de las personas, pues este se da *by default*. En cualquier caso, el consentimiento se hace casi inservible e

inoperante con la mayor complejidad del contexto tecnológico, la web 3.0 y el *big data* (Oliver y Muñoz, 2014). Y no es para menos: con los sistemas de inteligencia artificial y las decisiones automatizadas no es posible consentir, dado que es muy difícil prever las finalidades que se perseguirán y los algoritmos que utilizarán. Como apunta Martínez (2014), es una falacia afirmar que se pueda obtener para tratar una infinita cantidad de datos; según este autor, hay que arrinconar el consentimiento a un ámbito residual.

Pues bien, frente a las garantías subjetivas que en buena medida dependen del consentimiento y la acción del individuo, deben reforzarse las obligaciones legales preventivas de privacidad en el diseño por defecto y de la evaluación de impacto de protección de datos. Todo ello, en la línea con el reciente *Reglamento de protección de datos* de la UE; las garantías ya han sido mencionadas al abordar la no discriminación (arts. 25-34 y ss., Reglamento; ENISA, 2015; AEPD e ISMS, 2017).

Los datos masivos rehúyen el régimen jurídico del derecho de protección de datos personales

Crawford y Schultz (2014) afirman —coincidiendo con Rubinstein (2013) y Tene y Polonetsky (2013)— que mediante el análisis de los metadatos —entendidos como un conjunto de hallazgos predictivos o agregados sin mostrar o distribuir la procedencia de los datos— en la mayoría de los casos se estará al margen de las protecciones de privacidad usuales y actuales. Señalan estos autores que el análisis predictivo y la categorización pueden suponer una verdadera amenaza para las personas, sobre todo cuando se realiza sin su conocimiento o consentimiento.

Sin embargo, no se tratará necesariamente de un daño que cae dentro de la invasión de privacidad convencional, pese a que se trata de la relación de un individuo con datos sobre él; los datos masivos, aunque puedan crear una detallada imagen íntima del usuario con un alto grado de precisión, no surgen necesariamente de personas directamente, sino que pueden proceder de una combinación de técnicas de análisis de redes sociales y de la interpretación de los comportamientos en línea. Muchos de los usos sensibles o conflictivos del *big data* manejarán datos que no son personales, esto es, relativos a personas identificadas o identificables; por ello decaerán las garantías que implica este derecho. Como señala Martín (2013), simplemente el uso de metadatos sería

suficiente para los tratamientos que interesan de extraer patrones, perfiles, probabilidades, etc.

Otra vía de escape y de las claves jurídicas es que hay que centrar especialmente la atención en la premisa de que el *big data* implica un tratamiento de datos personales; de lo contrario, no procede a aplicar el régimen jurídico de este derecho. Si los macrodatos no son datos de personas concretas identificadas o identificables, no se aplica la legislación de protección de datos. Pues bien, una vía para escapar a la aplicación del exigente régimen de protección de datos es la anonimización, esto es, desvincular la información de su titular, por lo que deja de ser dato personal. La anonimización del *big data* presenta especiales problemas técnicos y jurídicos; es muy difícil hoy día garantizar que los datos no vuelvan a ser personales (AEPD e ISMS, 2017, p. 40; Stalla-Bourdillon y Knight, 2017). En la misma dirección, con los sistemas repartidos y el troceamiento de la información propios del *big data*, no es posible conocer la ubicación de los datos y su tratamiento efectivo, por lo que es ciertamente difícil predicar un control de estos por el usuario.

La anonimización de los datos masivos y sus dificultades para escaparse del derecho de protección de datos personales

Una barrera importante para la aplicación de la normativa de protección de datos a los datos masivos es lograr establecer patrones a partir de datos anonimizados, esto es, desligados completamente los datos de sus titulares, de modo que resulten imposibles de vincular. De este modo, no serían datos personales y, en consecuencia, no se proyectaría el régimen normativo. Como se puede prever, saber si se trata o no de datos personales es una cuestión esencial para conformar el régimen jurídico aplicable.

A este respecto hay que seguir especialmente el Dictamen 5/2014, de 10 de abril, del Grupo de Trabajo del artículo 29 (sobre anonimización), el cual sirve para ver la dificultad de que se produzca una efectiva anonimización que desvincule los datos del régimen de datos personales. Asimismo, como recuerda Martínez (2014), la anonimización constituye un tratamiento que exigirá consentimiento, una base legal o un interés legítimo. Especialmente se señala que con ella se ha de evitar que pueda ser establecido vínculo alguno entre el dato y su

•Marco teórico y aproximación jurídica al *big data*•

titular sin un esfuerzo desproporcionado, que sea irreversible y que en la práctica sea equivalente al de un borrado permanente. Se repara en que la finalidad de recogida inicial y la anonimización posterior, como tratamiento, sean compatibles. Cabe tener en cuenta también las expectativas del titular sobre usos posteriores, el impacto que puede tener la aplicación de patrones de *big data* en el titular de los datos y en sus derechos. De igual modo, se subrayan riesgos como la persistencia de datos que permitan reidentificar mediante inferencias, por vinculación o relación con otros paquetes de datos personales.

Graves dificultades de exigir el consentimiento para el tratamiento de datos masivos

El derecho de protección de datos personales ha girado estructuralmente en torno al consentimiento del titular de los datos personales; no obstante, por diversos motivos, esta columna vertebral de la protección de datos quebró hace tiempo. Rubinstein (2013) afirma que la gran ola del *big data* “arrollará” (*overwhelm*) los clásicos principios y consentimientos del derecho de protección de datos personales. La sociedad no está dispuesta a renunciar al uso de las IT y no tiene una fuerte cultura de la privacidad, lo que lleva a hacer casi irreal o inefectiva la garantía del consentimiento (Oliver y Muñoz, 2014; Martínez, 2014; Rubinstein, 2013; Heeger, 2015). En la práctica, el consentimiento viene masivamente por defecto; no es realista en modo alguno creer que existe un efectivo control de la información personal a través del consentimiento y los derechos que lo complementan.

Incluso una vez se considere que sí se trata de datos personales bajo su régimen general, no son pocos los problemas que hay que analizar respecto de la proyección de dicho régimen. Así sucede, por ejemplo, respecto al consentimiento, el cual, según Oliver y Muñoz (2014), se hace inservible e inoperante con la mayor complejidad del contexto tecnológico, la web 3.0 y el *big data*. Con los sistemas de ID y decisiones automatizadas, no es posible consentir dado que no siempre es posible prever la decisión automatizada que se adoptará. En la misma dirección, con los sistemas repartidos y el troceamiento de la información propios del *big data* no es posible conocer la ubicación de los datos y su tratamiento efectivo, por lo que es ciertamente difícil predicar un control de estos por parte del usuario.

Asimismo, como apunta Martínez (2014), es una falacia afirmar que se pueda obtener para tratar una infinita cantidad de datos; indica este autor que hay que

“arrinconar” el consentimiento a un ámbito “residual”. Y como se ha adelantado, habrá que perfilar los criterios para considerar que existe un interés legítimo suficiente para permitir tratamientos de datos del *big data* en razón de toda una serie de factores; de igual modo, no hay que excluir la oportunidad o necesidad de que sea la ley la que autorice los tratamientos del *big data* en sectores específicos (sanidad, educación, administración, justicia, etc.) o en razón de colectivos particulares.

Dificultades para la proyección de los principios del derecho de protección de datos personales

Más allá del consentimiento informado y de la seguridad en la protección de los datos personales, puede decirse que el núcleo duro de esta actividad está en los principios de calidad de los datos (por ejemplo, art. 4 LOPD española), a saber:

- Los datos de carácter personal solo se podrán recoger para su tratamiento —y sometidos a este— cuando sean adecuados, pertinentes y no excesivos en relación con el ámbito y las finalidades determinadas, explícitas y legítimas para las que se hayan obtenido.
- Los datos no podrán usarse para finalidades incompatibles.
- Los datos serán exactos y puestos al día, de forma que respondan con veracidad a la situación actual del afectado. Si los datos resultaran ser inexactos, serían cancelados y sustituidos de oficio por los correspondientes datos rectificados o completados.
- Los datos serán cancelados cuando hayan dejado de ser necesarios o pertinentes respecto a la finalidad para la cual hubieran sido recabados o registrados.
- Los datos no serán conservados en forma que permita la identificación del interesado durante un periodo superior al necesario para los fines con base en los cuales hubieran sido recabados o registrados.
- Los datos serán almacenados de forma que permitan el ejercicio del derecho de acceso.

Cabe remitirse a muy recientes estudios a este respecto, especialmente el realizado por AEPD e ISMS (2017); hay que analizar la proyección de estos principios

respecto del *big data* y su adecuación razonable. Como afirma Martínez (2014), a pesar de que los juristas disponen hoy de un sólido armazón de principios, se requiere del apoyo de una regulación que descienda al detalle. En este sentido, ya se ha mencionado el posible daño que puede proceder por un mal *big data*, pero también es cierto que, a pesar de la incorrección de los datos, este puede ser positivo y aportar valor a los datos. “Incluso lo ‘borroso’ nos ayuda en algo”. El autor en mención señala que hay que replantearse el principio de la finalidad: “Mi algoritmo se usó para estudiar patrones de consumo en un supermercado y, sin embargo, resultó que ofrecía elementos sustanciales en relación con las proyecciones de salud de mis clientes”. Así, es propio al *big data* que el concepto de finalidad pueda verse por completo alterado, no ya durante el uso o respecto a la decisión sobre este, sino ante resultados inesperados.

Dificultades para proteger los datos sensibles que se pueden generar a partir del big data

Como alertó la Comisión Federal de Comunicaciones de EE. UU. (FCC, 2014), los datos sensibles (raza, ideología, salud, orientación sexual, comisión de ilícitos, etc.) ya no son los que generarán los usos sensibles y problemáticos del *big data*; no en vano con ello es posible generar patrones y perfiles vinculables a estas categorías sensibles, pero no a partir de datos ajenos en principio a estos datos sensibles (por ejemplo, las pautas de fumar) y, por tanto, no contar con las especiales garantías de estos tratamientos.

Los datos de salud, ideología, religión, de orientación sexual o los relativos a delitos y sanciones suelen considerarse datos sensibles con un especial régimen jurídico y garantías. De nuevo con Martínez (2014) cabe señalar que las herramientas de *big data* ponen en cuestión esta categorización, pues se pueden establecer patrones o perfiles sensibles sin partir en modo alguno de datos de este tipo: deducir orientación sexual o de salud a partir del consumo televisivo, de contenidos o de compras, perfiles de criminalidad por pautas psicológicas de navegación, etc. De igual modo, con el *big data* es posible generar patrones y perfiles para usos de control o discriminación a partir de datos en principio ajenos al ámbito sensible. La geolocalización, las interacciones en redes sociales, el análisis semántico de expresiones emocionales, los hábitos sociales y el internet de

las cosas pueden aportar información relevante susceptible de ser usada con fines discriminatorios, o con fines de control social y policial.

Asimismo, cabe tener en cuenta los peligros del manejo de datos históricos, puestos de manifiesto por el TEDH (Gran Sala), de 4 diciembre 2008, Caso S. y Marper contra Reino Unido, o la STJUE (Gran Sala), de 8 de abril de 2014, que declara esta Directiva 2006/24/CE de retención de datos de las comunicaciones contraria a la intimidad, vida privada y protección de datos, restringiendo así los usos preventivos de los datos.

Cabe en este punto señalar el peligro que implica indexar las comunicaciones de la población con carácter puramente preventivo. Así pues, es necesario profundizar en la idea básica de Martínez respecto a que el modo tradicional de entender la sensibilidad de los datos debe abarcar ya no solo la naturaleza del dato o la finalidad del fichero, sino también el escenario que deriva de la predictibilidad.

Una aproximación a los usos de la inteligencia artificial y el *big data* en la aplicación automatizada de la ley

Desde 1972, si no antes, se detectan usos de automatización judicial y administrativa (Trazegnies, 2013, p. 123; Simmons, 2016, p. 957), y desde los noventa se hace en el ámbito policial (White House y NSTC, 2016). Buena parte de los usos más avanzados que se conocen hoy día son de apoyo a las decisiones públicas, y no tanto para determinar el contenido de tales decisiones. Es posible que se necesiten décadas para que el algoritmo tome las decisiones jurídicas complejas (Coglianese y Lehr, 2017, p. 36); no obstante, cada día más los algoritmos van guiando, apoyando o proponiendo decisiones públicas incluso de cierta complejidad, y en muchos casos la decisión automatizada propuesta pasa “automáticamente” a ser la decisión adoptada por el responsable humano.

Desde hace décadas se hacen esfuerzos para realizar ontologías jurídicas para los motores de búsqueda y la uniformización del lenguaje jurídico; se desarrollan también sistemas para la recuperación y el tratamiento de la documentación jurídica de asesoría jurídica, apoyar las decisiones jurídicas, la creación de documentos legislativos, la argumentación y la negociación jurídicas; así también, sistemas que actúan como consultores o asistentes inteligentes para encontrar solución legal a problemas (Cáceres, 2006; Zelezkikow, 1994; 2002). Incluso

•Marco teórico y aproximación jurídica al *big data*•

se prevé el uso de la inteligencia artificial para la misma generación y creación de normas (*algorithmic regulation*) (Coglianese y Lehr, 2017; Alarie, Niblett y Yoon, 2016).

En el ámbito administrativo, los algoritmos ya se usan en la contratación de bienes y servicios o en la selección de personal para asignar *medicaid*, *food stamp* u otros beneficios sociales (Coglianese y Lehr, 2017, p. 12; White House y NSTC, 2016, p. 30). Los algoritmos también seleccionan objetivos concretos para hacer inspecciones en el sector alimentario, en obras, en impuestos, etc. (Simmons, 2016, p. 957). El *big data* y la inteligencia artificial se utilizan y se van a utilizar más y más en la actuación pública, administrativa, policial o judicial, lo cual obviamente va a tener cada día más implicaciones en la vida y los derechos de la ciudadanía.

En principio, puede pensarse que el esfuerzo del jurista ha de ser el de proyectar el derecho tradicional o presente a las nuevas tecnologías y esforzarse especialmente en las dificultades de que el derecho se aplique en un nuevo marco tecnológico. Para internet, este modelo ha servido en general durante algún tiempo; no obstante, el derecho y las instituciones se ven superadas en no pocas ocasiones por el presente. Diversas ventajas pueden darse por el uso de la inteligencia artificial, los algoritmos y el *big data* en el ámbito específico de la aplicación del derecho. La inteligencia artificial puede aportar muchas opciones interpretativas e información al humano que debe decidir, o bien, puede poner en evidencia fallos o desvíos en la discrecionalidad administrativa o judicial, por la aplicación sesgada, intuitiva o errónea de la ley.

El uso de algoritmos y *big data* en la aplicación de la ley y la justicia puede favorecer la eficiencia y el ahorro de tiempo y costes tanto de los poderes públicos como de la ciudadanía; asimismo, puede permitir un sistema más eficaz y que gana en legitimidad. Esta última se incrementa cuando se adoptan soluciones que tienen en cuenta todas las variables, las cuales reducen la arbitrariedad, pueden ser más coherentes, más transparentes y comprensibles para la comunidad (Trazegnies, 2013). Además, la aplicación de la ley puede hacerse más previsible, y así gana la seguridad jurídica y se evita llegar a los tribunales en muchos casos (Zeleznikow, 2002). Como se verá, el uso de algoritmos y *big data* puede obligar a que garantías *due process* o la privacidad deban proyectarse en ámbitos policiales, administrativos o judiciales en los que no se exigían o donde se haga con

poca intensidad. En no pocos supuestos, los algoritmos van a poner en evidencia fallos o desvíos en la discrecionalidad administrativa o judicial en la aplicación sesgada, intuitiva o errónea de la ley.

Igualmente, tiene un gran potencial formativo: aparte de pilotos, se va a contar con simuladores estupendos de actuación administrativa, policial o judicial que pueden contribuir a la actuación y formación de muchos profesionales jurídicos. En el ámbito más concreto de la aplicación de la ley, cabe tener en cuenta:

- La aplicación escrupulosa de la ley con plena eficacia gracias a robots y algoritmos puede generar unos efectos nunca experimentados. Irónicamente, se puede llegar a unos niveles de eficacia, eficiencia y calidad administrativa y judicial que excedan lo asumible tanto por la sociedad democrática en su conjunto como por los individuos.
- Tradicionalmente, los poderes públicos han gozado de mayor o menor discrecionalidad o razonabilidad en la aplicación de la ley; ahora, en manos de algoritmos, pueden llevar a resultados insospechados (p. e., el positivo margen de actuación policial humana para resolver situaciones tensas, puesto en manos de robots).
- Incluso en actuaciones públicas llevadas a cabo exactamente igual por humanos o robots, la percepción humana será diferente, especialmente en las actuaciones tradicionalmente llevadas de cara al público; esta diferente percepción no necesariamente es negativa, puede ser una oportunidad para los objetivos de los poderes públicos (testigos, objetividad selección, etc.).
- Desde el punto de vista interno de las administraciones, no puede darse la misma exigencia a humanos que a robots; diferentes habrán de ser también los controles o las reacciones ante incumplimientos o errores.
- Los algoritmos tratan tan rutinaria y masivamente datos de los ciudadanos y en tal cantidad de ocasiones que el tratamiento jurídico y las garantías desde la privacidad o el proceso debido no pueden ser los mismos que los de cada afectación de modo individual y aislado. En relación con lo anterior, parece necesario trabajar con una dimensión colectiva de los derechos individuales que no es la habitual. El daño individual producido por el *big data* y la inteligencia artificial puede ser imperceptible

para un individuo, pero ser muy relevante para sectores o conjuntos de la sociedad.

- Desde el punto de vista interno de los poderes públicos, la exigencia a humanos y a robots no puede ser la misma, como diferentes habrán de ser los controles y mecanismos de estímulo de la productividad o reacción ante cumplimiento.
- Los humanos deben dirigir, controlar y, al mismo tiempo, aprender de la actuación automatizada. En no pocos supuestos el robot puede poner en evidencia fallos o desvíos en la actuación tradicional cargada de sesgos o erróneo entendimiento de la aplicación intuitiva de la ley.

El derecho es un plato que se sirve mejor frío; sin embargo, las nuevas tecnologías no dejan enfriar los problemas generados. No hay que dejar pasar el tiempo sin que el derecho guíe el desarrollo del *big data* y la inteligencia artificial en la aplicación de la ley y la actuación judicial. Este breve estudio se elabora desde la perspectiva jurídica y desde el marco de garantías constitucionales de Estados Unidos; no obstante, buena parte de los lineamientos que se incluyen pueden afirmarse en general para países democráticos constitucionales, por la afinidad de las garantías y los derechos constitucionales.

LA RECONSTRUCCIÓN DEL SUJETO Y EL ESTADO: *BIG DATA* Y PSICOPOLÍTICA

Los datos y el *big data*: esfuerzo por la comprensión de lo inconmensurable

Las sociedades han construido el concepto de lo inconmensurable. Se trata de una abstracción que significa aquello que no podemos asir, que sabemos que existe pero que es más grande que nosotros y nuestros conocimientos. Lo inconmensurable es generalmente un reflejo matemático, algo que solo otra abstracción puede explicar: los números. Ellos expresan estas relaciones de lo inasible, pues señalan algo que al parecer es tangible, pero que, al juntarse de forma geométrica, termina rompiendo las barreras de esa línea divisoria entre la certeza y la fe científica.

Los datos son cuantificables y por ello se expresan de forma abstracta; cuando estos datos aumentan su cantidad y complejidad, inician ese proceso místico de existencia basada en la fe científica, un salto de imaginación para el ciudadano promedio. No se trata de señalar que la ciencia que sostiene y estudia sus interrelaciones es inválida y poco consistente, pues seguramente lo es; se trata de pensar cómo las personas que no forman parte de la élite intelectual que ha creado los parámetros de comprensión de los datos se relacionan con ellos y construyen una epistemología asociada. La perplejidad que surge de la inconmensurabilidad se puede observar en la tabla 1.

Tabla 1. Unidades de capacidad de almacenamiento digital

Nombre	Cantidad de bytes	Equivalente
Bit	Unidad básica	--
Byte	8 bits	--
Kilobyte (KB)	1024	1024 bytes
Megabyte (GB)	1048576	1024 KB
Gigabyte (GB)	1073741824	1024 MB
Terabyte (TB)	1099511627776	1024 GB
Petabyte (PB)	1125899906842624	1024 TB
Exabyte (EB)	1152921504606846976	1024 PB
Zettabyte (ZB)	1180591620717411303424	1024 EB
Yottabyte (YB)	1208925819614629174706176	1024 ZB

Fuente: Wikipedia (s. f.).

¿Qué significa esto para nuestras vidas?, ¿qué podemos comprender de estas cifras? Es posible que muy poco o nada. Pero quizás el asunto no se trate de que todos podamos entender la inconmensurabilidad de los datos, sino que, en realidad, podamos usarlos de alguna manera. El *big data* no se refiere entonces a la existencia de los datos, pues, en efecto, desde que se contaban los sacos de trigo o de papa en diferentes civilizaciones, se comprendían y registraban datos a través de tablillas o quipus. La escritura y las matemáticas sirven naturalmente para el registro de datos. Lo novedoso es que las sociedades de la información están generando tantos datos que desbordan los métodos convencionales para su comprensión, lo cual pone en jaque la forma tradicional de almacenarlos y plantea retos para su uso:

De acuerdo con un estudio realizado por Cisco, entre el 2011 y el 2016 la cantidad de tráfico de datos móviles crecerá a una tasa anual de 78%, así como el número de dispositivos móviles conectados a internet excederá el número de habitantes en el planeta. Las Naciones Unidas proyectan que la población mundial alcanzará los 7,5 billones para el 2016, de tal modo que habrá cerca de 18,9 billones de dispositivos conectados a la red a escala mundial, esto conllevaría que el tráfico global de datos móviles alcance 10,8 exabytes mensuales o 130 exabytes anuales. Este volumen de tráfico previsto para 2016 equivale a 33 billones de DVDs anuales o 813 cuatrillones de mensajes de texto. (Paredes-Moreno, 2015, p. 41)

Según lo anterior, el *big data* es un gran banco de datos inteligente que no puede ser procesado con las herramientas de bases de datos creadas. En realidad, solo hasta ahora, en poco más de una década de estudio, estamos comprendiendo

•La reconstrucción del sujeto y el Estado: *big data* y psicopolítica•

lo que es y cómo puede transformar nuestras vidas (Klous y Nart, 2016). Por ello, las definiciones genéricas de *big data* apuntan a explicar dos ideas: el cúmulo de datos y la técnica de procesamiento. En una explicación más comprensiva, Hopkins (2011) da la siguiente definición de *big data*:

Técnicas y tecnologías que hacen que sea económico hacer frente a los datos a una escala extrema. Big data trata de tres cosas: 1) las técnicas y tecnología, lo que significa que su empresa necesita personas que sepan qué hacer con los datos para obtener un valor agregado; 2) la escala extrema de datos que supera a la tecnología actual debido a su volumen, velocidad y variedad; 3) el valor económico, haciendo que las soluciones sean asequibles y ayuden a la inversión de los negocios [traducción propia].

Estos datos resultan de las relaciones que se están produciendo con el uso cada vez más cotidiano de la tecnología, y suelen clasificarse como *estructurados* (datos de formato estándar), *semiestructurados* (datos que se pueden convertir a formato estándar) o *no estructurados* (Paredes-Moreno, 2015, p. 49). Esta idea de *información no estructurada* se refiere típicamente a aquellos datos que no están organizados bajo el modelo de datos relacional definido por Edgar Codd en 1970 (Puyol, 2014, p. 483). Ejemplo de ello son videos, fotos, textos Word, etc. Estos datos no estructurados, desde luego, son cada vez más complejos, pues implican dimensiones de nuestra vida y personalidad que no podían analizarse con los procesos anteriores.

De esta forma, cada vez que publicamos un tuit, utilizamos Facebook para subir fotos o videos, o para manifestar nuestro agrado o desagrado por algo o alguien, y desde luego cuando preguntamos dónde queda un lugar al usar Google Maps o tan solo encender el GPS, o cuando buscamos un restaurante o seguimos las noticias o un chiste en Youtube, estamos generando datos y, por ende, alimentando el *big data*. Esto en referencia solo a asuntos cotidianos de muchas personas en toda la tierra; pero también las instituciones, al prestar servicios, y las empresas, al vender productos, producen cantidades de datos.

Se ha acuñado la expresión *smart cities* ('ciudades inteligentes') (Townsend, 2013, p. 93) para referirse, por ejemplo, a las urbes que cuentan con sistemas que brindan información en tiempo real del tráfico, los servicios sociales o los servicios públicos, para reconocer alertas tempranas y ejecuciones gubernamentales que no serían posibles sin esta información permanente. También existen las fábricas inteligentes, que, en su afán de optimizar su operatividad, diariamente

producen datos de forma masiva. Ante esta diversidad, los datos que se producen pueden ser clasificados conforme a la tecnología existente:

- *Web and social media*: incluye contenido web e información que es obtenida de las redes sociales como Facebook, Twitter, LinkedIn, etc., blogs.
- *Machine-to-Machine (M2M)*: [...] se refiere a las tecnologías que permiten conectarse a otros dispositivos. M2M utiliza dispositivos como sensores o medidores que capturan algún evento en particular (velocidad, temperatura, presión, variables meteorológicas, variables químicas como la salinidad, etc.), los cuales transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa.
- *Big transaction data*: incluye registros de facturación, en telecomunicaciones, registros detallados de llamadas (CDR), etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados.
- *Biometrics*: información biométrica en la que se incluyen huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc. En el área de seguridad e inteligencia, los datos biométricos han sido información importante para las agencias de investigación.
- *Human generated*: las personas generamos diversas cantidades de datos como la información que genera un *call center* al establecer una llamada telefónica, notas de voz, correos electrónicos, documentos electrónicos, estudios médicos, etc. (Paredes-Moreno, 2015, p. 41).

Pero el concepto que parece convertirse en la clave del *big data*, para el presente y futuro, es el de *internet de las cosas*, que implica que cada objeto poco a poco está construyéndose para que se comuniquen con otros objetos y sistemas, y de esa manera pueda saberse en tiempo real las condiciones de su uso y situación. Esto elabora una visión en 3D digital de todo el mundo que nos rodea e implica cambios tan sencillos como alertas al teléfono celular —p. e., para indicar que los yogures van a caducar— e información precisa sobre hábitos —p. e., de aseo—. Lo anterior, sin duda, constituye un cambio que se debe reflexionar, pues no se trata de una simple novedad de Navidad: estamos asistiendo a modificaciones que podrían alterar el paradigma de comprensión de cosas que hoy damos por sentadas, como El Estado, el gobierno, los derechos, etc.

La dinámica en mención representa una mayor preocupación cuando los esfuerzos más importantes para manejar y dimensionar el *big data* se centran en

•La reconstrucción del sujeto y el Estado: *big data* y psicopolítica•

el lucro. Arend (2012), recogiendo lo señalado por la consultora IDC1, expresa: “Las tecnologías *big data* describen una nueva generación de tecnologías y arquitecturas, diseñadas para extraer el valor económico de grandes volúmenes de una amplia variedad de datos, al permitir la alta velocidad de captura, descubrimiento y/o análisis”.

En efecto, como parte de esta recreación se ha acuñado la expresión *minería de datos* (o *data mining*), que ilustra justamente la preocupación de las empresas ya no solo por almacenar estos datos, sino por procesarlos. Hoy se trata de saber cómo obtener valor agregado de estos datos y reconocer las oportunidades de cambiar la perspectiva del negocio. Encontrar las relaciones de los datos y su utilidad es similar a la búsqueda de metales preciosos en la minería, y desde luego puede ser una actividad muy rentable y prometedoras en el capitalismo digital:

Data mining, o minería de datos, es una tecnología que ayuda a las organizaciones a concentrarse en la información más importante existente en su base de información. A través de las herramientas de *data mining* es posible predecir futuras tendencias y comportamientos, permitiendo efectuar una toma de decisiones proactiva y conducida por un conocimiento extraído de la información (*knowledge-driven*). (Secades, 2015, p. 320)

Esta minería, siguiendo el símil, ha ido cambiando su proceso, que en un comienzo era de tipo artesanal, pero ahora se produce con maquinarias tecnológicas cada vez más potentes y que impulsan un acelerado desarrollo. Esto es posible, según Puyol (2014), en línea con lo señalado por Molina, por múltiples razones:

- a. El abaratamiento de los sistemas de almacenamiento tanto temporal como permanente.
- b. El incremento de las velocidades de cómputo en los procesadores.
- c. Las mejoras en la confiabilidad y aumento de la velocidad en la transmisión de datos.
- d. El desarrollo de sistemas administradores de bases de datos más poderosos. (p. 485)

Pero, al igual que la minería, ¿qué reglas existen para su extracción?, ¿cómo comprender la regulación de este campo? En realidad, la preocupación por la privacidad y la capacidad reguladora de las instituciones es una gran incógnita y apenas estamos profundizando en ella, pues la iniciativa privada y la tecnología

parecen ir más rápido que el desarrollo de estas mismas instituciones. El derecho y el Estado no son protagonistas centrales de estas disputas, y parecen no acomodarse a la misma velocidad de los cambios, lo que deja grandes interrogantes sobre el tipo de instituciones que deberían defender derechos (las cuales muchos consideran viejos valores) y que aún hoy tienen algo que decir en este tipo de capitalismo.

De esta manera, el *big data* se ha caracterizado a partir de las denominadas tres *V*, aunque varios autores han ido aumentando y perfilando nuevas características (Puyol, 2014, p. 486):

- *Volumen*. El *big data* se trata de un procesamiento intensivo y complejo de grandes cantidades de datos que tienen valor para una compañía o entidad. Este volumen de información plantea retos de almacenamiento y procesamiento, que además es siempre creciente y que se espera tenga un aumento exponencial conforme al desarrollo de nuevas tecnologías y la masificación de su uso.
- *Variedad*. Existen diversas fuentes generadoras de datos que se clasifican en estructuradas, semiestructuradas y no estructuradas. Esto implica la comprensión de que los datos no solo se refieren a los que una organización genera, sino a todos aquellos que se producen externamente. Esta variedad colige el desarrollo de técnicas para su procesamiento diverso, así como la comprensión de que el registro debe ser tan flexible como la cantidad de orígenes que pueda tener la información. Es un sistema complejo con varios canales abiertos a la vez.
- *Velocidad*. Se refiere a la capacidad de análisis de los datos que son producidos en tiempo real; es decir, implica la reducción de tiempos entre la producción de los datos, el momento en que se captan y su procesamiento para tomar decisiones organizacionales. Con arquitecturas de datos tradicionales y sistemas de procesamiento convencionales es imposible realizar esta tarea. La velocidad pretende que estas decisiones se puedan tomar no solo de forma rápida, sino que además sean eficientes y eficaces.
- *Valor*. Se trata de realzar que los datos seleccionados en el proceso de almacenamiento, procesamiento y decisión sean relevantes para la organización. No todos los datos cumplen estas características, pues pueden existir algunos que en realidad no presenten ventajas tanto en ahorro de

costos como en inversión estratégica de conocimiento. Ahora bien, este concepto de valor puede tener una connotación típicamente capitalista, que implica necesariamente la obtención de un beneficio generalmente de carácter económico. Sin embargo, desde una perspectiva humanizadora, el *big data* también puede plantear retos distintos sobre una teoría valorativa, construida sobre los fundamentos de la Constitución, en especial, de valores como la dignidad humana. En otras palabras, el *big data* no solo puede servir para obtener una plusvalía particular, sino para beneficiar a los perdedores tradicionales de esta ecuación.

- *Veracidad.* Se pretende que los datos que se almacenan y procesan gocen de una buena calidad, esto es, que tengan información de valor; pero para que este sea valor sea significativo, los datos deben ser veraces. Aunque las distorsiones siempre pueden darse, y la incertidumbre forma parte de este ejercicio —pues no es posible tener fiabilidad completa de todos los datos—, en este punto precisamente el *big data* muestra sus más interesantes aplicaciones: la predictibilidad de las decisiones. La combinación de todos los elementos descritos implica poder aumentar las posibilidades de vislumbrar el futuro, de anticipar la toma de decisiones de una organización.

Transformación y destrucción creativa en el Estado del *big data*

La evolución de las instituciones trae cambios que se presentan en formas diversas, en ocasiones de manera lenta y casi imperceptible, en tanto en otras de modo violento y rápido. Estos cambios institucionales a menudo se relacionan con el movimiento de los paradigmas que sostienen grandes arquetipos de comprensión del mundo perceptible. Asimismo, estos paradigmas recogen una interpretación de valores, ideas y hasta sentimientos, inciden en el comportamiento de las sociedades y modelan las instituciones; de este modo, perfilan sus formas de reproducirse y, sobre todo, de mantener su existencia (Resnik, 2004, p. 44). La autopoiesis, que se presenta como una característica repetida en todos los sistemas, se une a la idea de paradigma, la cual, aunque impulsa la existencia de cada institución que sostiene, termina cediendo en algún momento a nuevas formas paradigmáticas que modifican o crean nuevos sistemas.

El Estado como institución totalizante se comporta y estructura desde perfiles paradigmáticos y autopoieticos. El Estado moderno en particular plantea un sistema de valores y comportamientos que señalan lo correcto o incorrecto, y que se sitúan siempre como acciones que permiten no solo construir los límites de los ciudadanos, sino también alimentar su propia existencia. Además, como organización política ha permitido la ocurrencia de diversas propuestas dentro de su vasto espectro de definiciones del poder. Estas propuestas, que implican paradigmas dentro de paradigmas, están sujetas a cambios que se recrean en el concepto de *destrucción creativa* planteado por Schumpeter (1983); concepto que señala cómo los cambios institucionales siempre están acompañados de transformaciones coligadas a innovaciones en cada momento y que establecen modelos distintos de hacer las cosas.

Tales innovaciones se presentan no solo por contribuciones tecnológicas puntuales, sino también por reconceptualizaciones que ponen acentos diferenciados en otras ideas o fines. Esto ocurrió, por ejemplo, con transformaciones históricas como el paso del hombre nómada al sedentarismo y al desarrollo de la agricultura, o con la finalización del poder teológico del papado y el rediseño del concepto de poder y riqueza en el marco de la creación del Estado liberal, con el subsiguiente origen del capitalismo (Acemoglu y Robinson, 2014, p. 213). En el marco de la denominada *cuarta industrialización*, el Estado y la sociedad parecen registrar este momento de destrucción creativa, donde el *big data*, la minería de datos y el internet de las cosas se presentan como ejes de un nuevo planteamiento.

La destrucción creativa genera, por tanto, la pérdida de los valores, métodos y fines anteriores; en su lugar emergen otros nuevos que se establecen en el marco de un renovado aprendizaje cultural y siempre colectivo. Esta destrucción genera ganadores y perdedores, que a menudo se relacionan con la capacidad de los individuos y grupos sociales de adaptarse al cambio y con la pugna misma por el poder y sus formas de creación. Ahora bien, esta adaptación —con los estragos aparejados que pueda generar— termina siendo paradigmático y presenta procesos de reajuste, donde si bien no se eliminan los riesgos de los perdedores, sí pueden disminuir las tensiones producidas por estos cambios. Pero cuando las transformaciones se producen de forma rápida, o cuando los poderes existentes se resisten inelásticamente a ellos, es el momento en que se presenta una liberación

•La reconstrucción del sujeto y el Estado: *big data* y psicopolítica•

de tensión que genera efectos indeseados de desestabilización política e incluso violencia incontrolada.

Autores como Boaventura de Sousa han planteado que esta destrucción creativa no es siempre creativa: puede ocurrir una “destrucción destructiva” (Santos y Gandarilla, 2009, p. 70). Se llama la atención, entonces, sobre los perdedores y por qué estos existen en el vendaval de la destrucción del capitalismo innovador. No se trata simplemente de enaltecer a quienes con su ingenio y adaptabilidad subsisten dentro de las dinámicas del capitalismo global, sino de reconstruir la arqueología de quienes en el proceso son avasallados y puestos en el olvido, frente al “éxito” de quienes han obtenido un nuevo camino de generación de plusvalías.

El cambio de sistemas de producción y la generación de cadenas de valor cada vez más complejas en el marco del uso de las nuevas tecnologías —contexto en el que los datos devienen bienes que prometen nuevas formas de destrucción creativa del capitalismo— implican no solo ganancias importantes, sino la redefinición de límites y funciones estatales que rebasan las formas tradicionales de sujeción y control planteadas por el Estado liberal industrial. Un ejemplo de esto se observa más claramente si se piensa que cuando se produce un bien, no todos los involucrados en esta producción conocen cómo hacerlo. Si se está fabricando un computador, cada uno de sus componentes no procede de una misma parte; en realidad, se puede dar la vuelta al globo terráqueo para generar la trazabilidad de todas sus partes. Y, desde luego, donde son ensambladas no se sabe cómo se extrae el cobre, ni cómo se fabrican las pinturas de las teclas, ni tampoco de dónde se extrae el coltán, ni cómo se elaboran los microcircuitos y los discos duros.

La ignorancia de las cadenas de valor expone la necesidad de comprender que, no por el hecho de obtener valor, estas cadenas son benéficas para las partes que la componen; así también, que es posible que la obtención de valor pueda afectar de un momento a otro a los mineros del coltán, sin saber por qué o incluso quiénes han ocasionado afectaciones directas a su forma de vida (Ayestarán, 2011, p. 82). Aquí es donde la destrucción creativa es positiva para quienes obtienen este valor, pero destructiva para quienes fueron borrados del nuevo proceso, al ocultarles o minimizarles sus derechos y su reconocimiento como sujetos que reclaman no ser afectados.

En el caso del *big data* y, en especial, de los datos como bien generador de valor, se expone una cara de transformación creativa, al plantear la obtención de

beneficios económicos para las empresas que cada vez, de forma generalizada, están apostando por esta actividad (Stucke y Grunes, 2016, p. 36); sin embargo, poco se ha dicho sobre la dimensión de la destrucción, de los perdedores de minería de datos. Más preocupante es tal situación si se piensa que esta “extracción” se produce en términos muy similares a la del coltán (González, 2015): poca presencia de Estado como regulador de la actividad, poco control del proceso en general (extracción, almacenamiento, procesamiento y distribución) y pocas garantías para quienes pueden resultar afectados por dicho proceso. Esto es lo que se denomina la “cadena de ignorancia de los Friedman”:

Esto es, el desconocimiento de las implicaciones y extensiones de las innovaciones sociales en el mundo actual, donde una limitada visión de los acontecimientos y agentes involucrados acaba por generar una epistemología de la ceguera ante la “destrucción creativa” de los medios de producción, suministro y distribución de los bienes y servicios. El reduccionismo incorporado al evaluar semejantes innovaciones y producciones desde una sola variable axiológica y epistémica —ya sea el sistema de precios (Milton Friedman), ya sea la externalización o subcontratación internacional (Thomas L. Friedman)— provoca graves desviaciones y miopías en la interpretación de la información y el conocimiento que estos procesos significan. (Ayestarán, 2011, p. 82)

Según lo anterior, se hace necesario evidenciar todo el panorama frente a los procesos de innovación social, pues estos no pueden atribuirse al mercado y su mano invisible. Este tipo de explicación es fragmentaria y desconoce elementos axiológicos, ideológicos y prácticos muy importantes para recrear la comprensión y el papel de las instituciones, en especial las del Estado. Estos elementos pueden aplicarse desde luego a las reflexiones del *big data*, que muestran las mismas simetrías y proponen una nueva forma extractiva del capitalismo.

La innovación social puede caracterizarse, en línea con Ayestarán (2011, p. 84), como una novedad, es decir, no se trata de una reforma o renovación, sino de una propuesta nueva en el ámbito del conocimiento. Es axiológica porque se expresa en el marco de un sistema de valores. Posee un origen, es decir, es producto de un proceso interno o externo (o ambos). Es socializada en el espacio y en el tiempo, tiene un contexto; por tanto, es producto de un proceso social. Aunque sea impulsada por un individuo, es necesario que lo colectivo se exprese para lograr la dimensión innovadora; esto es, un invento puede llegar a ser una

innovación, pero no en todos los casos. Se hace a través del conocimiento y, por consiguiente, no se produce por eventos o procesos biológicos. Es producto de la capacidad de socialización y reflexión humana. Por último, se construye por medio de un complejo social, que pueden ser redes o sistemas.

Atendiendo estas características, el fenómeno del *big data* puede ser catalogado como una innovación, pues es claramente una novedad; si bien la tecnología se superpone a otra, las capacidades desarrolladas por el *big data* no tienen precedentes y plantean rupturas y discontinuidades con la forma en que la sociedad comprende fenómenos como el futuro, la voluntad o el poder mismo. Es axiológica no solo porque se presenta con raíces claras en las formas del capitalismo tecnológico, sino porque expresa en sí misma nuevas maneras de comprender valores y costumbres sociales, como los parámetros de lo correcto socialmente, la expresión de la afectividad y la construcción de relaciones sociales; esto último, especialmente ligado a aplicaciones como Twitter, Facebook y WhatsApp. Pero también redefine el concepto de lo público y de las fronteras de la humanidad misma, al generar una existencia global y local (Quintana, 2006, p. 76) expresada en el cuerpo físico y el avatar, un universo de *seconds lives*.

El origen del *big data* se sitúa en un lugar permanentemente endógeno y exógeno al individuo y los colectivos —algo propio de los fenómenos de globalización—, y se superpone persistentemente sin poder establecer su origen. Sin embargo, es necesario establecer que, si bien los datos y su proceso de análisis son interactivos y deslocalizados, sus beneficios sí pueden localizarse, pues no se plantean como verdaderos beneficios globales, sino solo de las empresas que pueden desplegar los recursos necesarios para extraer el “petróleo” de los datos. Sobre su dimensión innovadora, está claro que lo es rotundamente, por cuanto no es un invento aislado, dado que el *big data* consiste justamente en la posibilidad de aplicación de las *V*, y con ellas se construye su capacidad en la interacción colectiva.

El *big data*, por definición, no es un fenómeno particular o individual, pues implica la interoperabilidad como eje de su potencial. La variedad y el volumen se producen en un ambiente colectivo permanente. Además, no es un producto biológico, aunque la tecnología de la inteligencia artificial que puede optimizarse y desarrollarse en este contexto propone que la línea entre lo natural y lo artificial pueda ser difuminada. Y desde luego, el *big data* funciona en un complejo

con redes, sistemas y dispositivos, dado que se producen interconexiones entre nodos, y estos, además, intercambian información y reflexiones en tiempo real, lo que implica la posibilidad de analizar incluso relaciones de saber y poder en su interior.

Estado, poder y *big data*

El Estado en la modernidad se presenta como una institución cuyo basamento es el poder; es un sujeto de poder, una abstracción que termina siendo utilizada para organizar un campo de lucha entre quienes pretenden tomar el control. Anderson (2002), un clásico sobre la historia del Estado moderno, planteaba los elementos propios de su configuración: monopolio de la fuerza, control de la economía, establecimiento de un orden internacional, todo ejercido sobre un territorio y, desde luego, sobre “todas” las cosas que lo habitan. Pero para lograr que dicho “Leviatan” pueda existir, es indispensable que se soporte en un mito, en un esquema ideológico que implique que los gobernados crean en su existencia y su poder. Este discurso no es otro que el de la soberanía estatal, sustentado en la suposición de que el Estado se comporta como Dios, es decir, es omnipresente y omnisciente; por ello, su existencia no es realmente cuestionada, o al menos, solo pocos la ponen en duda. La legitimidad que desde Bodino se estructura a partir de la soberanía le permite al Estado crecer y fortalecerse. El modelo liberal y el constitucionalismo, por ejemplo, si bien discrepan de la versión estatal del absolutismo, no abandonan su existencia; por el contrario, la utilizan para desde allí elaborar el discurso liberal de la democracia y la Constitución, que persiste hasta el día de hoy (Porrás, 2001).

Ahora bien, el Estado, incluso desde la construcción más clásica liberal planteada por Locke (Crossman, 2004, p. 40), requiere de límites, pues su poder es tan grande que es necesaria la consolidación de mecanismos y tecnologías institucionales para impedir que unos pocos se apropien de sus capacidades y terminen generando grandes afectaciones a todos los ciudadanos. Instituciones como los derechos, los principios de división de poderes, frenos y contrapesos, garantías institucionales, principio de recambio político, etc., forman parte de estos desarrollos. Pero, aun con todo, el Estado sigue siendo una fuerza arrolladora cuyo

potencial dominador y destructor se encuentra presente en la vida contemporánea, y los gobiernos tienen este potencial a su disposición.

La literatura de la globalización advierte cambios en la forma de ejercicio del poder social, político y económico. Estos cambios se enmarcan en el problema de la deslocalización y, por tanto, en el deterioro de las formas tradicionales de control del Estado, como la ley, la burocracia y el aparato represor (Campillo, 2009, p. 253). En consecuencia, el poder se invisibiliza, se matiza entre la ignorancia del proceso productivo —como se revisó antes al reflexionar sobre la destrucción creativa y destructiva—, superando así de forma constante y persistente la noción de límites territoriales.

La ciudadanía plantea un esquema relacional con el Estado, y sus formas de poder se difuminan porque los sujetos se deslocalizan, se virtualizan, y en este proceso el Estado —al menos el tradicional— pierde la capacidad de control. Pero esto no significa que no existan otras formas de control, ni que los poderes económicos y sociales no aprovechen este intersticio para liberarse de los controles impuestos por el racionalismo del modelo liberal. Es una etapa diferente, un momento de elaboración de otra racionalidad: la configuración de un proceso de destrucción creativa; una propuesta que se mueve en la comunicación, en el discurso, en la digitalización de la existencia, pero que, al fin y al cabo, sirve para lo mismo que los modelos anteriores: para desarrollar el control, mantener el poder, contener nuevas formas de producción. A este respecto, Ragnedda (2011) plantea:

La televisión no es, por ahora, panóptica, porque no puede controlar (como sí lo hace en la metáfora de Orwell) la periferia o, lo que es lo mismo, a los espectadores. Internet, en cambio, es al mismo tiempo panóptico y synoptico, porque vigila y al mismo tiempo promueve valores modelos de comportamiento, controla y propone, vigila y seduce. Por todo ello puede ser analizado como un nuevo instrumento de control social que, como Jano Bifronte, tiene dos caras: vigila y seduce, observa sin poder ser observado y al mismo tiempo seduce manteniéndose como el centro de atención y bajo las miradas de todos. (p. 46)

El *big data* no es el modelo, sino que forma parte de un modelo nuevo que plantea la creación de otros modos de ejercicio del poder y de reelaboración de los antiguos, lo cual encierra una ideología, una intencionalidad. El poder estatal deberá reescribirse en el marco de las *V*: velocidad, volumen, variedad y valor;

características que se desarrollan a través de un modelo de un sujeto diferente: el sujeto virtual, pero con los retos de servir de estructura de control en asuntos como la justicia, la administración, la regulación económica, la seguridad ciudadana y la protección de derechos. El Estado se expande en la red, sin saber exactamente qué hace y cómo. La transparencia se convierte en el arma central del control (Freedom House, 2017).

Los tiempos de respuesta y la cantidad, diversidad y estructuración de información y decisiones, en el marco de una teoría de lo que es valioso para el sistema, son demandas nuevas que la estructura tradicional no podrá responder. Estas demandas han generado campos como el gobierno electrónico o el domicilio digital, pero en realidad aun no corresponden a respuestas que impliquen un acercamiento completo a un problema ciertamente complejo.

En este contexto surge el concepto de *psicopolítica* desarrollado por Byung Chung Han. La idea central se encuentra en la comprensión de otra forma de dominación acuñada en el proyecto neoliberal y que involucra una nueva fase del capitalismo: el capitalismo financiero. Las anteriores formas del capitalismo liberal se habían concentrado en el control del cuerpo, de lo físico. La libertad se define en el marco de este control y tiene como herramienta central el panóptico (Terán, 2007, p. 99). El panóptico estudiado por Foucault plantea una comprensión del dominio espacial, donde los sujetos son conducidos a espacios que implican control en sí mismos. Se disciplina el cuerpo, se regla el comportamiento a través del control y la fuerza. El Estado entonces se encarga de garantizar este control a través de los conceptos centrales del modelo liberal: la administración de justicia y la policía.

Pero dentro del panóptico quedan puntos ciegos, lugares que escapan al ojo vigilante del control. Estos lugares están precisamente ubicados en la subjetividad, en el espacio interior del controlado, en su posibilidad de tener el cuerpo encerrado y controlado, y aun así poder soñar y escapar con su mente a este control. Este modelo de lo físico se expresa en la posibilidad de dominar la naturaleza y comprenderla por medio de la razón; y como expresión de esta se encuentra *la estadística*: todo lo que pueda ser contado es controlado.

Sin embargo, en la actualidad el modelo de control ha cambiado de forma sostenida. Nos encontramos en otro paradigma, que sitúa el control en el espacio de la mente y de los sentimientos:

•La reconstrucción del sujeto y el Estado: *big data* y psicopolítica•

La biopolítica que se sirve de la estadística de la población no tiene ningún acceso a lo psíquico. No provee ningún material para el psicoprograma de la población. La demografía no es una psicografía. No explora la psique. En esto reside la diferencia entre la estadística y el *big data*. A partir del *big data* es posible construir no solo el psicoprograma individual, sino también el psicoprograma colectivo, quizás incluso el psicoprograma de lo inconsciente. De este modo sería posible iluminar y explotar a la psique hasta el inconsciente. (Han, 2014, p. 38)

Este psicoprograma se expresa de forma sutil, superando la idea de control por la fuerza; se redibuja en una comprensión de la perplejidad de lo no medible, como se señaló al inicio del capítulo. Si bien parece que los datos, eje de esta reflexión, son inconmensurables, en realidad es todo lo contrario: con ellos se pretende establecer un nuevo modelo que busca generar más control que antes, una nueva fase del poder político y económico:

A partir del uso de la *big data*, las empresas de tecnología, las empresas del capital y los gobiernos se adueñan de nuestros tiempos, deseos y libertades: toman posesión de nuestros ocios, nos definen lo que es libertad y entretenimiento, invaden nuestros momentos productivos y de deseos. El tiempo es una cosa que nos gestiona internet: el tiempo ya no nos pertenece, es de internet. Y el espacio ya no existe, habitamos la georreferenciación de la web: ella nos sigue, nos dice dónde estamos y qué hay ahí. Es más, ya existe Tinder, que me dice si en ese espacio-tiempo que habita nuestro wi-fi o celular hay alguien disponible para el sexo casual o compartir su soledad. Tener tecnologías es estar más conectado: y estar conectado es ser más súbdito de los gestores de datos. El sexo, los amigos, los gustos, los deseos son cosas que nos gestiona internet: somos humanos sujetados a los datos digitales: somos su gran cosa. (Rincón, 2016, p. 30)

En efecto, los datos cosifican lo que pretenden contener, el sujeto se diluye en un mar de datos. Como se explicó, una de las características del *big data* es que supera la comprensión de los ficheros, del registro, de la estadística a la que se refiere Han (2014). El *big data* plantea el uso de los datos semiestructurados y no estructurados, que en términos no técnicos implican el lugar donde se depositan nuestros sueños, sentimientos e impresiones más inmediatas. No se trata de grandes reflexiones, sino de lugares para ser reactivos en 280 caracteres. Allí se alojan las expresiones del odio o de lo más querido y valorado. El *big data* pretende gestionar nuestra vida entera, ya no hay lugares vedados al panóptico, todo es visible, “transparente”:

La apertura del futuro es constitutiva de la libertad de acción. Sin embargo, el *big data* permite hacer pronósticos sobre el comportamiento humano. De este modo, el futuro se convierte en predecible y controlable. La psicopolítica digital transforma la negatividad de la decisión libre en la positividad de un estado de cosas. La persona misma se positiviza en cosa, que es cuantificable, mensurable y controlable. Sin embargo, ninguna cosa es libre. Sin duda alguna, la cosa es más transparente que la persona. El *big data* anuncia el fin de la persona y de la voluntad libre. (Han, 2014, p. 26)

Ahora bien, una de las cosas que más llama la atención del dominio psicopolítico es la forma en que sucede: se realiza de forma voluntaria, el sujeto se sumerge en un mar de posibilidades expresadas en aplicaciones, páginas web y opciones allí disponibles. La web 2.0 permitió esta interacción que en principio parecía libertaria, y la web 3.0 la está llevando a un estadio de realización. La sensación de controlar las cosas es permanente, y ello se vive a través del fetiche del teléfono móvil. Ahora podemos saber todo, decir todo, recorrer todo; solo basta un clic. Pero en este ejercicio el sujeto no ve que está en un panóptico voluntario, donde el poder no se desgasta, pues se recrea desde la mente misma. Es el anhelo de estar conectados a la *matrix*, considerando que ello es la realización de la libertad, pues allí somos lo que queremos ser y donde se ha podido liberar todo lo que el espacio físico no permite. Turkle (1984), psicóloga norteamericana, ha investigado desde los ochenta esta idea de la identidad del sujeto en internet:

La computadora, al igual que el test Rorschach de manchas de tinta, es un poderoso medio proyectivo. [...] Lo que las personas hacen de la computadora habla de sus intereses más amplios, de su identidad como personalidades individuales. [...] La cualidad camaleónica de la computadora, el hecho de que cuando uno la programa se convierte en criatura propia, hacen de ella un medio ideal para la construcción de una amplia gama de mundos privados y, a través de ellos, para la exploración de uno mismo. (pp. 22-23)

Este aspecto camaleónico forma parte de la dominación en la reconstrucción de la ciberlibertad. Se trata de una autorreferencia permanente al *yo*, pero que en realidad no devela su naturaleza, sino que exhibe la posibilidad de su fraccionamiento, de su dislocación persistente. Es un juego, un espacio lúdico donde se buscan las sensaciones que el *yo* físico no puede realizar; es un camino de expresión de las frustraciones. Esta idea de lo lúdico es parte también del control. Como se trata de saturar los sentidos, de adormecer la crítica y la reflexión, si el

sujeto juega, entonces se siente pleno. La capacidad de las personalidades deteriora el sujeto revolucionario que antes predicaba la modernidad:

Ineficiente es el poder disciplinario que con gran esfuerzo encorseta a los hombres de forma violenta con preceptos y prohibiciones. Radicalmente más eficiente es la técnica de poder que cuida de que los hombres se sometan por sí mismos al entramado de dominación. (Han, 2014, p. 28)

Esta dominación que se cimienta en el placer, en la ingravidez de la red, genera distancia entre los individuos y produce distorsiones claras en el concepto de lo colectivo, de clase social. La pantalla es un maravilloso espejo donde todo es perfecto, *customizable* y expandible:

La pantalla nos daría una falsa sensación de compañía y de estar al mando, y esa placentera e ilusoria experiencia de la comunicación digital iría desplazando a la más verdadera, sagrada y compleja conversación cara a cara y del contacto físico. La tecnología nos seduce y se convierte con ello en “la arquitectura de nuestras intimidades”. (Turkle, 2012, p. 20)

Pero, en efecto, como señala Turkle (2012), la soledad es en realidad el estado permanente, dado que la interacción física, el silencio colectivo y la calidez del otro se hacen tan ausentes que ahonda la necesidad de ser aprobado y de sentir que siempre hay muchos a quienes el sujeto les importa en el muro, en el tuit, en el WhatsApp. Se aumentan las interacciones, pero en realidad no se construyen nuevas realidades, pues el sujeto sigue solo frente a la pantalla. De esta forma, el juego, el ocio de la red, en realidad revela una proyección de nuestras frustraciones, deseos y odios, que son muy reales y que cambian al sujeto, lo modifican en la vida real:

Aun cuando en los espacios mediados por computadoras podríamos transformarnos en seres fantásticos y nuevos, en la práctica seguimos patrones, valoraciones y conductas culturales a la hora de presentarnos y autodefinirnos, pues precisamente por más fantásticas las posibilidades de acción, en tales espacios son actores reales y concretos quienes deciden y responden a tales posibilidades. Actores que tienen diferentes razas, géneros, nacionalidades, costumbres y valoraciones respecto de sus realidades, de las cuales no se pueden desembarazar fácilmente. Por ello es que de manera más reciente Yee prefiera hablar más bien de la Paradoja de Prot. (Valderrama, 2017, p. 10)

Así las cosas, el sujeto, aunque parece experimentar una plenitud de libertad antes no expresada en la vida física, en realidad pierde su identidad, pues las distintas puestas en escena se convierten cada vez más en su “realidad”. El *big data* consume todas estas personalidades, las cuantifica, las reelabora y las pone de manifiesto en sus intereses. El dato como generador de plusvalía cuantifica todo —especialmente la persona—. Se acentúan así fuerzas normalizadoras que destrozan la identidad, lo anómalo, lo raro, y crean un estándar de *homo data* procesable y claramente impresionable. Así lo argumenta Valderrama (2017):

Navegar por internet o usar redes sociales, buscadores o dispositivos de *selftracking*, [...] deja una estela que permitiría conocernos de mejor manera según estos autores. Sin embargo, más que para experimentar nuevas identidades que no podríamos explorar en el mundo *offline*, los rastros digitales constituirían y fijarían nuestras identidades de manera fidedigna a la *offline*, al punto de llegar a enterarnos de gustos y elementos de nuestra personalidad gracias a las predicciones de algoritmos. (p. 13)

Esta presunta libertad virtual, en el fondo, reconfigura las diferencias ya existentes. Es claro que la versión clásica del conflicto capital/trabajo ya no encuentra los mismos referentes en el capitalismo financiero, globalizado y digital, pues el conflicto tradicional de la clase obrera pierde centralidad cuando precisamente ahora todos somos empresarios, aunque el único empleado sea uno mismo:

Como es conocido, la cuestión se ha planteado a partir del “oscurecimiento” de la clase trabajadora como sujeto histórico, que si dio lugar por parte del pensamiento conservador a declarar la victoria definitiva del capitalismo (el fin de la Historia), por parte del pensamiento crítico se procedió a buscarle “sustitutos” al trabajo, al haber perdido la “centralidad” que ocupó en el sistema, ya que la revolución y progreso tecnológico (el desarrollo de las fuerzas productivas) han hecho “desaparecer” la histórica “clase trabajadora”. (De Cabo, 2012, p. 134)

Se recrea entonces una dualidad: la de ser opresor y oprimido; se acaba la tesis y la antítesis, y el individuo se cree su propia síntesis, sin comprender claramente lo que pasa, sin un contexto sereno por analizar, ya que todo es repentino, es el ahora más que nunca. Contamos con todos los recursos, pero ahora no hay tiempo definido, el espacio es una fractura permanente. Las empresas hoy hablan de un día de más de 24 horas; esas horas eran físicas. Hoy una persona consume por pantallas, es decir, su consumo se mide en cuánto tiempo está en cada pantalla. Así se superponen los días, pues una persona puede consumir al mismo

•La reconstrucción del sujeto y el Estado: *big data* y psicopolítica•

tiempo en varios entornos digitales; se maximiza su consumo, se rompe la barrera de lo físico. No hay oposición, solo sensaciones; no hay contradicción, solo la recreación de una libertad sujeta a decir “me gusta”.

De forma mágica, ahora parecen todos saber más de mí; no se trata, como en el pasado, de vivir experiencias medidas y obligadas por otros, por la sujeción de lo correcto. Ahora es posible dislocar la personalidad y vivir varias experiencias éticas —e incluso sin ética—. Esa es la lógica del insulto en el tuit, o la oscuridad de la internet profunda, o el *second love* y las parejas. Han (2014) señala en este sentido:

El régimen neoliberal presupone las emociones como recursos para incrementar la productividad y el rendimiento. A partir de un determinado nivel de producción, la racionalidad, que representa el medio de la sociedad disciplinaria, topa con sus límites. La racionalidad se percibe como coacción, como obstáculo. De repente tiene efectos rígidos e inflexibles. En su lugar entra en escena la emocionalidad, que corre paralela al sentimiento de libertad, al libre despliegue de la personalidad. Ser libre significa incluso dejar paso libre a las emociones. El capitalismo de la emoción se sirve de la libertad. Se celebra la emoción como una expresión de la subjetividad libre. La técnica de poder neoliberal explota esta subjetividad libre. (p. 71)

Pero este mundo de posibilidades se desvanece cuando se observa que el *big data* no viene de la nada. Es un aparato ideológico y contextual. “Alguien” ha diseñado la forma de construir la correlación de datos, y “alguien” impulsa y gana con este diseño. El *big data* reproduce entonces las viejas desigualdades, la pobreza, la exclusión, la persecución a lo diferente; de ahí que el racismo, el sexismo y otros males de antaño no son extirpados de la data. Son un registro también, y pueden ser aprovechados; expresan emociones que sirven como vehículos para obtener un beneficio, el control del espíritu, la manipulación de la libertad. Valderrama (2017) sostiene:

El procesamiento de *big data* posibilitaría un manejo eficiente de los trazos digitales, individualizando el marketing de productos para realizar estudios de mercado y vigilar el comportamiento de clientes y contribuyentes por parte de conglomerados privados y gobiernos. Esto ha generado API más restrictivas y han proliferado compañías dedicadas a la venta de datos y servicios de minería. Las agencias de publicidad bajo sistemas como AdWords rastrear las *cookies* y aplican tácticas automatizadas según las páginas que hacemos clic. Incluso los gobiernos, después de las filtraciones de Snowden sobre la agencia NSA, se han revelado como entidades con altas

capacidades de vigilancia y rastreo. En suma, aquellos que tienen el financiamiento, los recursos, las redes y el poder necesario pueden acceder, procesar e interpretar esa información, manifestando nuevas asimetrías respecto al acceso de la data así como de su manipulación en la sociedad contemporánea. (p. 12)

La explotación se intensifica, y los datos como medio de producción permiten flexibilizar lo que antes no era explotable. La minería de datos expresa muy bien esta idea, pues es una actividad que avasalla, que rompe con las reglas permanentemente. La reflexión ecológica, la protección de la dignidad humana, el reparto equitativo de los recursos, etc., se desvanecen en la oscuridad de la mina. Las clases sociales en esta minería intensiva no desaparecen; por el contrario, se agrupan, se expresan de otra manera, menos dramática, más lúdica. El rostro de la pobreza desaparece por la animación y el aparente anonimato. Todo es estético en la red, se acaban las angustias y, por tanto, las contradicciones. “El *big data* da lugar a una *sociedad de clases digital*. Los individuos que son clasificados en la categoría ‘basura’ pertenecen a la clase más baja. A aquellos con una puntuación baja se les niega el crédito” (Han, 2014, p. 99).

La utopía de internet como lugar de las libertades se esfuma, dado que siempre estamos dejando trazos, datos, y este proceso promete profundizarse cada vez más, cuando el internet de las cosas cope nuestro espacio, se consolide la construcción del mundo virtual. Las personalidades performativas del entorno digital se consolidarán, y con el crecimiento de la capacidad de registro y procesamiento en alza, nuestro yo físico ya no tendrá la centralidad de ahora, el sujeto de derechos se verá transformado y moldeado por el yo virtual. El *big data* ve el futuro, sabe lo que quiero, manipula mis emociones, sabe a dónde quiero viajar, a quién quiero conocer, qué quiero que digan los políticos, qué ejercicio hacer, qué quiero aprender. El *big data* no vive en el pasado, y menos en el presente. Los tiempos de la estadística pasaron, pues es una mirada al pasado, la comprensión de lo que fue; ahora se trata de cómo saber a dónde voy, georreferenciando no solo mi cuerpo, sino además mi espíritu:

El *big data* sugiere un conocimiento absoluto. Todo es medible y cuantificable. Las cosas delatan sus correlaciones secretas que hasta ahora habían permanecido ocultas. Igual de predecible debe ser el comportamiento humano. Se anuncia una nueva era del conocimiento. Las correlaciones sustituyen a las causalidades. El ello

•La reconstrucción del sujeto y el Estado: *big data* y psicopolítica•

es así sustituye al por qué. La cuantificación de lo real en búsqueda de datos expulsa al espíritu del conocimiento. (Han, 2014, p. 102)

Pero en medio de este escenario, ¿no existe escapatoria? Si el *big data* es un instrumento ideológico, ¿no puede reinterpretarse, deconstruirse? Esta mirada del pesimismo antropológico debe buscar alternativas. En perspectiva de Gadamer (1996, p. 207), debe ser posible el reconocimiento de otros horizontes de sentido más allá del *imperium* del *big data*. Es necesario reconstruir las otras voces, reelaborar el sujeto, sacarlo del éxtasis de lo virtual, de la dependencia de las endorfinas, para ponerlo a pensar, a sentir de forma autónoma, en el silencio de la meditación y la dicha de su autorreconocimiento. Esto implica, entonces, tomar el discurso del *big data* y oír los textos, recrear caminos diferentes a la autosatisfacción inmediata y recuperar el sentido del ser más allá de la vigilancia digital, para entonces aplicar un nuevo sentido, una nueva interpretación de su proyecto de vida.

Las instituciones existentes deben estar acompañadas de esta reconstrucción. Está claro que el Estado y sus contornos de control físico son superados por el poder de los datos; y este Estado no puede procesar, al compás de las *V* (velocidad, volumen, variedad), todo lo que sucede alrededor. Las normas de protección de datos y su reflejo, las normas de acceso público de datos, dejan enormes intersticios regulatorios que muestran la inelasticidad del Estado como institución que contiene el poder, si lo que pretende es ser el jugador decisivo de antaño. Por ello, parece indispensable que reinventemos las instituciones de protección del sujeto que reivindican sus derechos, su dignidad y autonomía. Pero no en la lectura ius-naturalista liberal, sino una que recoja la sensibilidad “auténtica”, la producida sin las interferencias de los conglomerados del nuevo capitalismo, y a partir de allí domesticar el *big data*; en suma, que sus capacidades se reviertan en la posibilidad de profundizar la protección y la eficacia de los derechos, especialmente el reconocimiento de la diferencia, donde se encierra la clave de la identidad más allá de los datos, en una nueva forma de universalización:

Es la “otra Constitución”, la Constitución de los (sujetos) débiles o Constitución alternativa. A través suyo se han quebrado supuestos propios del constitucionalismo y el derecho liberal al romperse su formalismo característico mediante el reconocimiento de la realidad por el derecho que, así, adquiere una —inmediata— base material: frente al universalismo característico siempre del derecho burgués, el reco-

nocimiento y protección jurídica de la diferencia; frente al monopolio del individuo en el reconocimiento y ejercicio de los derechos, la posibilidad de reconocimiento del sujeto colectivo de los derechos y frente al predominio o exclusividad del aspecto subjetivo, el reconocimiento de lo objetivo (la situación o circunstancia) como elemento para la configuración y atribución del derecho. (De Cabo, 2012, p. 139)

De muchas maneras, el *big data* reconfigura ese elemento, subjetiva de forma negativa, generando la abstracción —que en el modelo liberal es coronación de las esencias— de un sujeto imaginado, pero lejos de su realidad. El *big data* lo hace con sus emociones, recrea no su esencia sino su éxtasis por una vida ficticia de soledades amistosas y juegos de tragedia. La idea es entonces recordar al sujeto y sus circunstancias, recrear su realidad no como hecho irrefutable, sino reconocer que existen hermenéuticas diferentes para la comprensión del mundo virtual y su mundo físico. Y a partir de esto, evitar que lo virtual oculte la construcción de un proyecto de vida colectivo y personal en la profundidad de su sentipensamiento como experiencia que se desprende de las veleidades del virtualismo y recupera su sentido en el ser mismo.

EL *BIG DATA* EN LA CIBERDEFENSA Y LA CIBERSEGURIDAD NACIONAL VERSUS EL DERECHO A LA PRIVACIDAD DEL CIUDADANO COLOMBIANO

El presente capítulo tiene por objeto realizar un estudio del *big data* en la ciberdefensa y ciberseguridad nacional vs. el derecho a la privacidad del ciudadano colombiano. Con ese propósito, se analizan y desarrollan los conceptos, los elementos, las fases y las técnicas del *big data* en el marco de la administración electrónica. Se realiza también un análisis del papel de los equipos multidisciplinarios y del uso del *big data*, los contextos de aplicación, la necesidad de análisis de información, el procesamiento de imágenes, voz y datos, el tratamiento de datos y determinados aspectos éticos y legales.

La metodología que se utiliza parte de explorar y comprobar los hechos desde una perspectiva jurídica, sigue los presupuestos de una investigación cualitativa y se diseña una teoría desde la interpretación de acciones y procesos sociales. En el marco de la utilización de las tecnologías de la información y las comunicaciones (TIC), en los últimos años han emergido herramientas de análisis masivo de información dentro de las propias administraciones. La combinación entre servicios, información, datos —muchos de estos privados—, herramientas y las nuevas técnicas de gestión de la información —como pueden ser las que utilizan *big data*— ocasionan situaciones que pueden poner en riesgo a gobiernos, administraciones,

ciudadanos, empresas, entre otros; por ende, su previsión debe ser un elemento en el marco del desarrollo de las competencias de cada administración.

Aproximaciones al *big data*

En Colombia, el Ministerio de Tecnología de la Información y las Comunicaciones (MinTIC, 2015a) ha señalado que “reconoce el valor de la información como herramienta para el fortalecimiento de sectores industriales, gubernamentales y académicos. La existencia de información disponible hace de su consecuente análisis un insumo cada vez más valioso en la toma de decisiones”. Desde esa misma línea, ha referido que las tendencias mundiales en TIC evidencian la convergencia de tecnologías necesarias para el análisis de datos. Por eso, en reconocimiento prospectivo de las tendencias y los avances tecnológicos, crea Centros de Excelencia y Apropriación (CEA) con el fin de capitalizar el análisis de datos en sectores estratégicos. Los CEA en *big data analytics* focalizan las competencias de los recursos existentes para el desarrollo de estrategias cuya diferenciación recae en el análisis de información. Entre los objetivos del MinTIC está la creación de valor a partir de *big data analytics* para ciberseguridad, internet de las cosas y formulación de política pública (MinTIC, s. f.).

El análisis de grandes cantidades de datos y la generación de información que da valor añadido a la toma de decisiones de las organizaciones traen la necesidad de estudiar, desde la nueva perspectiva de derechos —muchos de estos fundamentales—, la intimidad, la privacidad, los derechos de autor, el derecho de acceso, entre otros, así como sus límites, que son elementos que han de pensarse antes de implementar estrategias de *big data*. Un ejemplo palpable de ello son las leyes de protección de datos de carácter personal, que como regla general han fundamentado el tratamiento desde la autorización previa, expresa e informada que da el titular de los datos a partir del consentimiento, cuya esencia es la libertad (Díaz, 2016).

De acuerdo con IBM (citado por Carrillo *et al.*, 2013, p. 8), *hay big data* si el conjunto de datos supera el terabyte de información y es sensible al tiempo; lo es además cuando mezcla datos estructurados con no estructurados. Su enfoque trata de buscar la manera de aprovechar estos datos, su combinación, su gestión y la aplicación de algoritmos predictivos de comportamiento; todo lo anterior,

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

con la finalidad de permitir la toma de decisiones que añadan valor al negocio. Es precisamente esa combinación de elementos, unos técnicos, otros jurídicos, la que nos lleva a realizar un análisis desde una nueva perspectiva.

La implementación de técnicas de *big data* no se da por sí sola, no tiene lugar desde la simple incorporación de tecnologías que permitan la práctica material *per se*; para ello es necesario realizar un diagnóstico integral de la misión de la organización, ya sea en el ámbito privado o en el público; luego procede una determinación del cumplimiento de esa misión desde la utilización de TIC, el tipo de datos que se maneja en el cumplimiento de la misión, la categorización de los datos, las implicaciones jurídicas del tratamiento de esos datos, la finalidad de dicho tratamiento, los aportes potenciales en la toma de decisiones si se cuenta con la información y el análisis de esos datos, la herramientas técnicas de análisis, las implicaciones jurídicas que trae el resultado del análisis de datos y, por último, el límite en cuanto al tratamiento y el análisis efectuados. Carrasco (2013) refiere los pasos para la implementación de *big data*:

- a. Entender el negocio y los datos. Este primer paso exige un análisis detallado con las personas que hoy laboran y entienden los procesos y los datos que la empresa maneja.
- b. Determinar los problemas y cómo los datos pueden ayudar a resolverlos.
- c. Establecer expectativas razonables, es decir, definir metas alcanzables; esto se puede lograr si al implementar la solución de un problema, este no presenta alguna mejora y, por tanto, se debe buscar otra vía para hacerlo.
- d. Cuando se inicia un proyecto de *big data*, es necesario trabajar en paralelo con el sistema que hoy está funcionando.
- e. Al tratar de implementar un proyecto de *big data*, se debe ser flexible con la metodología y las herramientas; esto se debe a que ambas son recientes y pueden llegar a presentar problemas al ser implementadas.
- f. Es importante mantener el objetivo de *big data* en mente, porque el proceso es pesado, máxime cuando los métodos y las herramientas que usan *big data* para el análisis de datos aún pueden presentar problemas. La idea es que se mantenga en mente la meta final del proyecto sin desanimarse pronto.

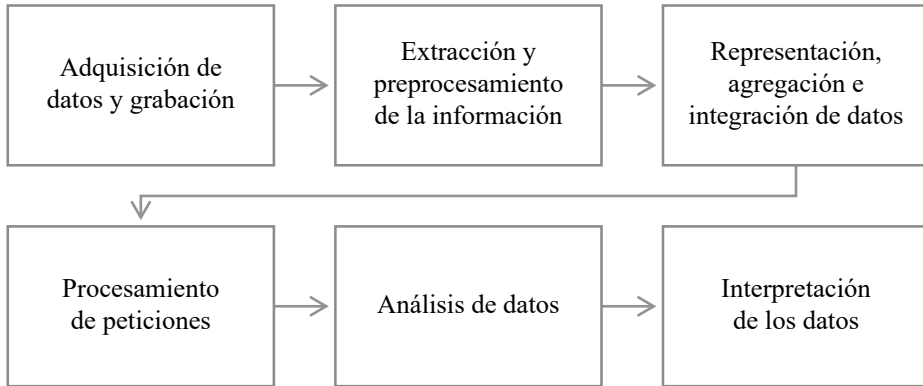
En relación con las dimensiones o los componentes del *big data*, en cuanto elementos esenciales y transversales a una gran cantidad de información, se

encuentran el volumen, la variedad, la velocidad y la veracidad de los datos. El volumen hace referencia a las cantidades masivas de datos que las organizaciones intentan aprovechar para mejorar la toma de decisiones en toda empresa, tanto privada como pública (Schroeck *et al.*, 2012, pp. 6-7). Asimismo, el volumen debe estar vinculado a las grandes cantidades de datos que llegan a las organizaciones (públicas o privadas) y que forman parte de la toma de decisiones estratégicas para el cumplimiento de su misión. De manera práctica, se puede afirmar que el volumen hace referencia al tamaño de la información (Grupo TRC, s f.).

Por su parte, la variedad tiene que ver con gestionar la complejidad de múltiples tipos de datos, incluidos los datos estructurados, semiestructurados y no estructurados. Entre los diferentes e innumerables tipos de datos se encuentran texto, datos web, tuits, datos de sensores, audio, video, secuencias de clic, archivos de registro y mucho más (Schroeck *et al.*, 2012, pp. 6-7). Uno de los grandes retos para el derecho será la delimitación de la finalidad para la cual los ciudadanos han autorizado el tratamiento de muchos de esos datos personales (algunos de ellos, de naturaleza sensible).

Los diferentes tipos de análisis siguen un conjunto de fases comunes que ayudarán en la toma de decisiones (Rayo, 2016). Las fases del *big data*, entendidas como aquellos pasos que deben seguir las empresas para la organización de la información, se muestran en la figura 1. Conviene resaltar que la fase final de interpretación de datos está directamente relacionada con el *data mining* o ‘minería de datos’, que se encarga de una multitud de tareas, como manipular, procesar, modelar, analizar y extraer la información que se necesite en función de resolver un problema determinado (Hop2croft, 2014).

Figura 1. Fases del *big data*



Fuente: Hop2croft (2014).

Aplicación del *big data* en la defensa y la seguridad nacional

Como se ha señalado, los datos (unos estructurados y otros no) de los que se pueden valer las organizaciones —incluido el Estado, desde los organismos de seguridad y defensa nacional— provienen de diversas fuentes y se encuentran en diferentes formatos. Aunque su uso en el sector de la seguridad nacional puede ser más que beneficioso, deberá estar limitado al cumplimiento de la finalidad legítima, pues su uso extralimitado puede ocasionar un gran perjuicio para los titulares de los datos. Desde la perspectiva de Carrillo *et al.* (2013, p. 44), el *big data* puede ser usado en diferentes entornos, pero ofrece grandes beneficios particularmente en el contexto de defensa y seguridad nacional, por lo cual será necesario precisar qué actividades presentes y futuras obtendrán resultados positivos en esta materia.

La defensa y la seguridad nacional tienen un enfoque preventivo, pues esto resulta menos costoso que invertir en daños que lesionen bienes jurídicos que se pretenden preservar; sin embargo, prevenir requiere decisiones de tiempo limitado y un alto nivel de síntesis en el número de datos y los factores involucrados. Por otro lado, las crisis actuales han visto aumentar su complejidad, todo ello debido a la globalización, los nuevos escenarios —con líneas difusas entre lo civil y lo militar—, los entornos intensivos en información con creciente mezcla de escenarios virtuales —p. e., ciberdefensa, económicos— y reales, etc. (Carrillo *et al.*, 2013, p. 44).

La Política de Defensa y Seguridad para la Nueva Colombia 2015-2018 prioriza el cumplimiento de varios objetivos:

Contribuir con capacidades, garantizar mayores y mejores niveles de seguridad ciudadana, contribuir a la modernización de la sociedad rural, combatir las nuevas y tempranas expresiones de crimen organizado que amenacen la seguridad y el funcionamiento transparente del Estado, garantizar la soberanía e integridad del territorio nacional, transformar y modernizar de forma continua el Sector Defensa, fortalecer la proyección internacional desde la cooperación bilateral, triangular y multilateral con los países aliados y estratégicos, poner al servicio del desarrollo nacional, comercial, industrial y agrícola las capacidades empresariales del Sector Defensa. Todo esto se podría materializar de mejor forma y de manera más eficaz desde la utilización de tecnologías, y de manera especial desde la utilización de técnicas de tratamiento y análisis de datos. (Ministerio de Defensa Nacional, 2015)

El *big data* ofrece alternativas en la solución de problemas existentes o emergentes. La aplicación del *big data* en materia de defensa y seguridad tiene por objeto captar y emplear grandes cantidades de datos con el fin de aunar sensores, percepción y decisión en sistemas autónomos, en clave de aumentar así el entendimiento de la situación y el contexto del analista y del agente del orden (Carrillo *et al.*, 2013, p. 44). La inteligencia militar lucha contra el fraude y busca garantizar la seguridad ciudadana, el planeamiento táctico de misiones, la vigilancia y seguridad de fronteras, la lucha contraterrorista y contra el crimen organizado. Y, claramente, el *big data* puede emplearse en el ámbito de la defensa y la seguridad nacional. Es evidente que un Estado elabora una estrategia de seguridad y defensa nacional desde diferentes enfoques: desde el análisis de sus riesgos, en los distintos espacios estratégicos (tierra, mar y aire), para dar cumplimiento a los objetivos estratégicos nacionales (Ministerio de Defensa e Instituto Español de Estudios Estratégicos, 2012, p. 42).

La doctrina especializada ha señalado que la ciberdefensa es el conjunto de acciones u operaciones activas o pasivas desarrolladas en el ámbito de las redes, sistemas, equipos, enlaces y personal de los recursos informáticos y teleinformáticos de la defensa, a fin de asegurar el cumplimiento de las misiones o los servicios para los que fueran concebidos, a la vez que para impedir que fuerzas enemigas los utilicen para cumplir los suyos. Se ha planteado iniciar el proceso de ciberdefensa a través de la inteligencia informática, con el ciberespacio como

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

ambiente, para poder obtener los elementos descriptores que conformen la identificación de los escenarios y, a la vez, parametrizar las amenazas. Con ello será posible dimensionar los riesgos y permitir así el diseño de los instrumentos de defensa (Consejo Argentino para las Relaciones Internacionales [CARI], 2013, p. 9).

En este contexto, la ciberdefensa se basará en la utilización de las TIC para el cumplimiento de su fin; por ello, sus agentes deberán contar con instrumentos jurídicos que les permitan la realización de actividades de inteligencia en el ciberespacio. Por ello, la Ley 1621 de 2013 tiene por objeto establecer el marco jurídico para los organismos que llevan a cabo actividades de inteligencia y contrainteligencia; allí se establecen los límites y fines de las actividades de inteligencia y contrainteligencia, los principios que las rigen, los mecanismos de control y supervisión, la regulación de las bases de datos, la protección de los agentes, la coordinación y cooperación entre los organismos, los deberes de colaboración de las entidades públicas y privadas, entre otras particularidades.

La ciberdefensa se efectúa en términos de *defensa activa* y *defensa pasiva* del centro de operaciones, de los medios de información que posee la institución, con el fin de repeler los ataques cibernéticos que esta sufra, y aquí su arma rectora por disposición son las comunicaciones militares, que coadyuvan a la protección cibernética de la infraestructura crítica del país. La primera de ellas se puede explicar como una estrategia para adquirir una capacidad de defensa del ciberespacio, que combina la protección interior de los sistemas, la vigilancia permanente de redes sensibles y la respuesta rápida en caso de ataque, contrarrestando de este modo las amenazas ciberespaciales y garantizando acceso al ciberespacio. La defensa pasiva, por su parte, es la estrategia para la protección de los activos relacionados con los sistemas de información, a través de controles detectivos, correctivos y disuasivos que contrarresten las posibles amenazas (Fuerzas Militares de Colombia, 2015).

Ahora bien, se hace necesario abordar las aplicaciones específicas del *big data* en cada uno de los contextos señalados. Entre las aplicaciones específicas identificadas se encuentran: detección de intrusión física en grandes espacios o infraestructuras abiertas, criminología computacional, computación sobre información encriptada, análisis automático de vulnerabilidades de red, uso fraudulento de recursos corporativos y/o sensibles, identificación de anomalías, patrones

y comportamientos en grandes volúmenes de datos, inteligencia visual en máquinas, análisis de texto (estructurado y no estructurado), apoyo a la toma de decisión en tiempo real en contextos intensivos en datos, conciencia situacional, predicción de eventos, etc. (Carrillo *et al.*, 2013, p. 46).

Las innovaciones de *big data* emergen en un momento en el que las organizaciones se enfrentan a nuevas problemáticas que se derivan de dos desafíos: a) disolución de los límites de las redes y b) extensión y apertura de las redes de datos de las organizaciones. Ello permite que socios, suministradores y clientes accedan a información de carácter corporativo a través de formas dinámicas para impulsar la innovación y la colaboración, lo cual, en último término, deviene en que las redes sean más vulnerables al mal uso y robo de información. Las aplicaciones y los datos corporativos son cada vez más accesibles por medio de, por ejemplo, servicios en la nube (Carrillo *et al.*, 2013, p. 50).

El siglo de la sociedad de la información —también llamado siglo de la revolución de la información— ha traído, y sigue trayendo, nuevos riesgos para los derechos de los ciudadanos y para la seguridad de las naciones, dadas las nuevas formas e instrumentos de criminalidad, así como las renovadas dinámicas del activismo, el espionaje y el terrorismo. Con frecuencia, los ataques o fraudes no son detectados, sino que solo se perciben cuando el daño se ha materializado. Por esta razón, como lo señalan Carrillo *et al.* (2013), es pertinente buscar soluciones más ágiles basadas en evaluaciones dinámicas de riesgo. Para ello serán esenciales las operaciones de seguridad en tiempo real, así como el análisis de grandes volúmenes de datos; todo ello con el fin de garantizar seguridad.

Objetivos del *big data* en la aplicación de seguridad y defensa

En la más reciente Directiva del Parlamento Europeo y del Consejo de la Unión Europea (UE), del 6 de julio de 2016, relativa a las medidas destinadas a garantizar un elevado nivel común de seguridad de las redes y sistemas de información de la Unión, a fin de mejorar el funcionamiento del mercado interior, se dan varias prerrogativas:

- a. Establece obligaciones para todos los Estados miembros de adoptar una estrategia nacional de seguridad de las redes y los sistemas de información.

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

- b. Crea un grupo de cooperación para apoyar y facilitar la cooperación estratégica y el intercambio de información entre los Estados miembros, y desarrollar la confianza y seguridad entre ellos.
- c. Crea una red de equipos de respuesta a incidentes de seguridad informática (en lo sucesivo, red de CSIRT, por sus siglas en inglés), con el fin de contribuir al desarrollo de la confianza y seguridad entre los Estados miembros y promover una cooperación operativa rápida y eficaz.
- d. Estipula requisitos en materia de seguridad y notificación para los operadores de servicios esenciales y para los proveedores de servicios digitales.
- e. Establece obligaciones para que los Estados miembros designen autoridades nacionales competentes, puntos de contacto únicos y CSIRT con funciones relacionadas con la seguridad de las redes y los sistemas de información.

La Ley de Inteligencia colombiana (Ley 1621 de 2013) establece de manera clara que la función de inteligencia y contrainteligencia es llevada a cabo por las dependencias de las Fuerzas Militares y la Policía Nacional —p. e., la Unidad de Información y Análisis Financiero (UIAF)— y por los demás organismos que faculte para ello la ley en mención (artículo 3). Por ello, el límite de la actividad de inteligencia está dado por el respeto a la Constitución, a la ley, a los derechos de los ciudadanos y a las normas del derecho internacional, para así asegurar los fines del Estado colombiano establecidos en el artículo 2 de la Carta Política: la vigencia del régimen democrático, la soberanía, la protección de las instituciones, hacer frente a amenazas como el terrorismo, el crimen organizado, el narcotráfico o el secuestro, y proteger los recursos naturales y los intereses económicos de la Nación.

Es precisamente sobre dichos contenidos que debe dársele sentido a lo que pretende la aplicación del *big data* en la ciberseguridad y defensa, a saber:

- a. Disminuir el tiempo que los analistas emplean descubriendo ciberataques, al reunir y relacionar las fuentes de redes de datos dispares.
- b. Aumentar la precisión, tasa y velocidad de detección de ciberamenazas a redes de computadores (Carrillo *et al.*, 2013, p. 49).

En este orden de ideas, uno de los elementos fundamentales será determinar cuáles son aquellas cuestiones, desde la aplicación de análisis de datos, que deberían ser consideradas para las actividades de seguridad y defensa nacional, y que

incluye esas actividades. Carillo *et al.* (2013) han señalado que en los próximos años, a largo plazo, se dimensiona que en el análisis de información de *big data* el rompimiento en la mayor parte de productos de seguridad en redes incluye la monitorización de estas, la autenticación y autorización de usuarios, la gestión de identidades, la detección de fraudes y la gestión de gobernanza, riesgos y conformidad. A mediano plazo, se espera que las herramientas mejoren, al punto de permitir un número considerable de posibilidades de predicción avanzadas y controles automatizados en tiempo real. Esto será aplicado en las áreas relacionadas con protección de redes de telecomunicaciones, ciberdefensa, ciberseguridad y protección de infraestructuras críticas (Carrillo *et al.*, 2013, p. 51).

Criminología en la aplicación del *big data*

La criminología computacional tiene un ámbito de aplicación directo en el *big data*, por cuanto analiza grandes cantidades de información relacionadas con actividades criminales. El uso de *big data* incrementa las probabilidades de neutralización de estas amenazas; en este sentido, se relaciona con diferentes técnicas como el cifrado de datos criminales, el análisis de agrupaciones, el aprendizaje de reglas de asociación para la predicción del crimen, el análisis de textos multilingües, opiniones y redes criminales, entre otras acciones (Carrillo *et al.*, 2013, p. 51).

Desde otra perspectiva, la aplicación del *big data* en la criminología computacional consiste en el seguimiento de actividades sospechosas en redes, internet o en las páginas web de las entidades; de ahí que sea fundamental identificar qué información es recogida como soporte para la actividad criminal. Por otro lado, *big data* puede ser implementado con el fin de recopilar, almacenar y documentar pruebas de actividades ilícitas o maliciosas realizadas dentro y fuera de internet o de las entidades corporativas; esto último, en un sentido preventivo, aprovechando la capacidad de analizar, procesar y almacenar volúmenes considerables de información. Las áreas de aplicación utilizadas son las siguientes: la seguridad general, la lucha contraterrorista y contra el crimen organizado, la lucha contra la usurpación de identidad, la lucha contra el fraude, la lucha contra la explotación infantil y la pedofilia, etc. (Carrillo *et al.*, 2013, p. 52).

Uso fraudulento de la información en el contexto del *big data*

Diferentes entidades y corporaciones han implementado el uso del *big data* en sus plataformas debido al gran número de usuarios y de recursos. Además del contenido de tipo sensible que puede tener esta información, se han dado unas pautas y reglas específicas, principalmente por el número considerable de sesiones concurrentes de acceso a estos recursos. De esta manera, gracias a las técnicas de *big data*, es posible analizar en tiempo real los datos de estas sesiones de acceso, identificando patrones de comportamiento y diferenciando usos permitidos de los recursos frente a otros, ya sean relacionados con el abuso de los recursos de información de la organización, o bien, con ataques cibernéticos (Carrillo *et al.*, 2013, p. 52).

Un ejemplo de lo anterior es que el país, con más de 180.000 empleados de la Policía Nacional, se enfrentó al reto de correlacionar toda la información procedente de diferentes fuentes: desde cámaras de videovigilancia hasta llamadas al 123 y flujos de trabajo del personal en las calles. Para dar respuesta a este requerimiento, la Policía está inmersa en un proyecto de *big data* que permite transformar esa información en conocimiento. Así lo explicó Jairo Gordillo, director de TI de la Policía Nacional, durante el evento Datacenter Dynamics Converged (Villarrubia, 2014).

Se debe señalar que los ataques al bien jurídico de la información, los datos y los sistemas de información se efectúan desde el interior de la misma entidad y corporación, incluso por parte de los empleados, que aprovechan las fallas. Las técnicas de *big data* favorecen el desarrollo de herramientas frente a seguridad informática y gestión de conocimiento de grandes organizaciones (Carrillo *et al.*, 2013, p. 52).

La importancia de la conciencia situacional del *big data* en el contexto de la seguridad y defensa nacional de los Estados

Para que el *big data* se vuelva una herramienta contundente en el ámbito de la seguridad y defensa, es fundamental que se analice el contexto de la situación respecto del analista o combatiente, agente de orden, para efectos de esta investigación. Es preciso resaltar que, dada la necesidad de identificar amenazas en

situaciones complejas, contradictorias e inciertas, en las que hay una cantidad considerable de datos disponibles de fuentes abiertas, es determinante realizar acciones para su neutralización y la detección de la amenaza (Carrillo *et al.*, 2013, p. 59).

La Organización de Estados Americanos (OEA) ha trabajado para fortalecer las capacidades de seguridad cibernética entre los Estados miembros desde principios de la década del 2000. El Programa de Seguridad Cibernética de la OEA apoya las iniciativas sobre la base de un análisis en profundidad y a partir de la comprensión de la magnitud de las amenazas; esto es, conocer el estado de la situación. La comprensión o el entendimiento contextual solo se pueden adquirir a través de la combinación de técnicas adecuadas —es decir, humano/máquina—, donde se aproveche la capacidad cognitiva de las personas para fundir y asimilar múltiples fuentes y tipos de información, en función de lograr nuevas perspectivas. Las técnicas de *big data* se aplican en dicho contexto, simplificando la exploración y el procesamiento de diversos datos; en específico, el *big data* puede aportar a través del procesamiento de sistemas de aprendizaje inmediato que procesen un lenguaje natural e inserten la representación semántica resultante en una base de conocimiento, en lugar de basarse en los actuales procesos costosos, que requieren de mucho tiempo para el aprendizaje de personas con diferentes áreas del conocimiento.

Hoy se ve posible unificar estas bases de datos con capacidades sensoriales como la “inteligencia visual” en máquinas, en las que se identifican no los objetos existentes en diferentes imágenes o videos, sino también las relaciones (acciones) entre ellos. De esta manera, se posibilita interpretarlos de manera inmediata y construir así una narrativa de información visual, que agrega resultados provenientes de la correlación en aspectos económicos, demográficos, patrones sociales, etc. (Carrillo *et al.*, 2013, p. 59).

En el futuro se espera obtener información en varios frentes: en el descubrimiento de sucesos específicos, tanto en su planeamiento como una vez hayan ocurrido; en el develamiento de temas y conceptos desarrollados colaborativamente; en valores y creencias que motivan ciertos comportamiento de interés; en el análisis semántico y de inclinación en el apoyo a grupos y personas; en áreas de aplicación relativas al planeamiento táctico de misiones; en la toma de decisiones en tiempo real para las operaciones concernientes a defensa y seguridad;

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

en traducciones automáticas en un considerable número de idiomas y volumen. Sin embargo, este tipo de actividades trae aparejados riesgos para la intimidad de los ciudadanos; por ello, es esencial la delimitación de la finalidad para la cual se realizan las actividades mismas.

Pese a que en la actualidad se han producido avances destacados en la recuperación y el tratamiento de datos, la realidad no es tan promisorias; un claro ejemplo son los motores de búsqueda inteligentes, donde el progreso en la comprensión de la información no ha avanzado con el mismo ritmo. Una de las razones para este desacuerdo es el considerable aumento de datos no organizados relacionados tanto con el teatro de operaciones y el entorno directo del combatiente como con los escenarios de seguridad (Carrillo *et al.*, 2013, pp. 59-60).

La creación de equipos multidisciplinarios para crear un control efectivo en la aplicación de *big data*

El *big data* no es un sistema *software* que arroja informes de manera automática; tampoco consiste en un conjunto de sistemas informáticos que, una vez instalado y configurado, empieza a generar soluciones y aplicaciones relativas a la seguridad informática (Carrillo *et al.*, 2013, p. 71). Por el contrario, se necesitan personas y profesionales para su operación; de ahí que sea fundamental la formación técnica y profesional en esta nueva área del conocimiento, que aborde las fuentes de datos, la tipología, la calidad, la naturaleza, la algoritmia estadística, la analítica, entre otros aspectos, desde la dimensión técnica, jurídica y práctica.

Carrillo *et al.* (2013) argumentan que este perfil de científico de datos implica conocer el contexto de complejidad del entorno de la aplicación. En las aplicaciones de seguridad y defensa es necesario señalar que la implementación de estos perfiles puede ser parte de la solución de la problemática ético-legal del *big data*. En suma, se busca incorporar el contexto de los datos de análisis, debido a que algunos de ellos son intrínsecamente inciertos —p. e., las señales GPS que rebotan entre los edificios—; también es necesario configurar un contexto de análisis que permita reducir la implementación de recursos (computación, almacenamiento, etc.), al hacer posible el planteamiento *big data* solo en aquellas fuentes que son relevantes para determinado contexto o en diferentes subconjuntos, de acuerdo con criterios de filtrado.

Actualmente se han creado diversas y numerosas herramientas informáticas; la mayoría de ellas forman parte de la evolución de plataformas *Business Intelligence*, un avance a la complejidad de visualizaciones. Entre los desafíos de la visualización de *big data* se encuentran la solución a la variedad de datos multidimensionales y aspectos espacio-temporales, el énfasis en análisis de información, la solución de problemas y la toma de decisiones.

La temporalidad, el análisis de escalas múltiples, la correlación y la causalidad son el gran reto; pero, más allá de esto, poder hacer frente al hecho mismo de que los criminales pueden igualmente identificar las correlaciones y anticiparse a las decisiones actuando de forma diferente a la planeada de forma inicial. Por ello, en el marco del equipo multidisciplinar, se deben incorporar expertos al respecto. Una de las características del *big data* es el análisis en tiempo real en *streams* de datos. Las capacidades existentes para procesar tal flujo de datos son tan intensivas en conocimiento y datos que necesitan razonamiento autorizado (Carrillo *et al.*, 2013, p. 74).

La necesidad del Estado en el análisis de datos

En el sector público colombiano se ha avanzado en algunos pilotos, como los realizados por el Departamento Administrativo Nacional de Estadística (DANE) con el uso de *big data* en estadísticas oficiales incorporadas en sus estrategias, en el monitoreo de los Objetivos de Desarrollo Sostenible y en el Censo Nacional de Población y Vivienda. De igual manera ha avanzado el Ministerio de Hacienda, junto con el Departamento Nacional de Planeación (DNP), en 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 adquiridos anteriormente con estadísticas tradicionales (Infraestructura Colombiana de Datos Espaciales [ICDE], 2017).

Actualmente ha surgido la necesidad, por parte de los Estados, de analizar datos y percibir acciones. Esta actividad se realiza de manera poco automática, más bien tiende a ser lenta y se realiza frecuentemente por parte de las administraciones públicas con el fin de garantizar la seguridad nacional y la defensa del Estado; sin embargo, ello puede llegar a violentar libertades individuales de los ciudadanos, como se verá más adelante. La consultora Govwin Networks señala

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

que el Departamento de Defensa de los Estados Unidos gasta aproximadamente el 58,4% de todo el gasto federal en almacenamiento de datos, y la mayor parte proviene de la necesidad de analizar videos. Esos programas tienen como finalidad captar una significativa cantidad de imágenes y videos en tiempo real para el reconocimiento y toma de decisiones en el contexto de seguridad y defensa nacional (Carrillo *et al.*, 2013, p. 100).

Otro avance significativo en el uso del *big data* tiene lugar con la creación de la Biblioteca Virtual en Salud para la Vigilancia en Salud Pública, impulsada por el Ministerio de Salud y Protección Social, con características de un Repositorio Digital Institucional, definido, según la Red Alfa Biblioteca de Babel, como “un archivo electrónico de la producción científica de una institución, almacenada en un formato digital, en el que se permite la búsqueda y la recuperación para su posterior uso nacional o internacional” (Palacios *et al.*, 2014, pp. 81-82).

Por último, el uso estratégico del *big data* surge en el marco de arquitectura empresarial que promueve estándares, buenas prácticas e interoperabilidad para todo el sector público. Según Bohórquez (citado en Fernández, 2014, pp. 48-49), 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. Así, de acuerdo con el Programa de Compra Eficiente, la infraestructura tecnológica para *big data* estaría disponible a través de servicios en la nube que serían ofrecidos bajo acuerdos de precios aplicables a todo el sector público y estandarización en tecnología, calidad y variedad de proveedores.

Fuentes de imágenes y video

Existen diferentes tipos de video que pueden tener diversas funciones para asegurar la defensa y seguridad nacional. Entre estos se encuentran diferentes dispositivos con variadas funciones, a saber: a) UAV o drones (cuya función es capturar una significativa cantidad de videos para el reconocimiento de imágenes en situaciones hostiles); b) videos procedentes de los vehículos de exploración y reconocimiento terrestre, tanto en el espectro visible como en el infrarrojo; c) imágenes tomadas de satélites de vigilancia; d) videos de cámaras de vigilancia en lugares públicos; e) ubicación de cámaras de vigilancia en espacios privados, como hospitales y empresas; f) publicación y difusión en diferentes redes sociales, como

Facebook, Twitter, Youtube, blogs y otros lugares de la red (Carrillo *et al.*, 2013, p. 101). Precisamente esas diversas fuentes nos llevan a entender que las actividades tendientes a garantizar la defensa y seguridad nacional deben estar amparadas en la Constitución y la ley, desde el respeto de los derechos de los ciudadanos —entre ellos, el derecho a la intimidad—, pero desde el entendimiento de la necesidad, idoneidad y proporcionalidad para garantizar la seguridad y defensa.

Procesamiento y uso de imágenes como herramienta para garantizar la defensa y seguridad de los Estados

La infraestructura *big data* frente al procesamiento de imágenes se ha aplicado progresivamente; pese a ello, esta herramienta le ha dado un aporte sustancial a la recolección de datos frente a la defensa y seguridad nacional, a través de la detección de movimientos y la detención en espacios no permitidos o en zonas de exclusión. Esta infraestructura implica el desarrollo de varias acciones: vigilancia de infraestructuras críticas, reconocimientos faciales en zonas determinadas, seguimiento y reconocimiento de objetivos, reconocimiento de comportamientos sospechosos en sitios públicos, identificación de actividades económicas o en zonas de conflicto, identificación de objetos abandonados, sospechosos, IED, etc. En el ámbito militar, estas herramientas tienen gran relevancia y utilidad, particularmente frente a los distintos sistemas existentes y desplegados. Entre los más importantes se pueden mencionar: herramientas de visualización, seguimiento de objetivos (ISTAR), herramientas de visualización del estado operacional de campo de batalla, herramientas de ayuda en torno a decisiones de carácter militar, etc. (Carrillo *et al.*, 2013, p. 101).

Por otro lado, la defensa requiere de ciertas habilidades y herramientas como son los programas I+D, que se pueden lograr a través de la implementación de diferentes tecnologías de procesamiento de imágenes e infraestructura *big data*. En el mercado se pueden encontrar diferentes alternativas, pues los fabricantes ofrecen productos de *big data*; por su parte, las empresas de ingenieros ofrecen sistemas de procesamiento cada vez más inteligentes. El uso de estas dos herramientas para lograr la capacidad requerida no es tarea fácil.

En 2000, la STO organizó un congreso en Canadá con el nombre de IST.020/rws-002: “Multimedia Visualization of Massive Military Data Dases”, que causó

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

gran interés frente a la visualización y presentación de datos, así como en relación con la vigilancia y función de grandes cantidades de datos y matemáticas de procesamiento. Vale señalar que desde hace casi dos décadas ya se plantea la necesidad de desarrollar actividades tanto investigativas como tecnológicas tendientes a la incorporación de proyectos de analítica de datos para las actividades de seguridad y defensa nacional. Posteriormente, la STO ha organizado, a través de su panel IST (Information Systems Technologies), diferentes grupos temáticos y de trabajo relacionados con la fusión de datos. Aunque el eje principal no es el procesamiento de imágenes, en el objeto de análisis se encuentran las fuentes heterogéneas que analizan estudios de la OTAN (Carrillo *et al.*, 2013, p. 103).

Uno de los programas implementados por la DARPA para garantizar la defensa y seguridad nacional es el Mind's Eye, utilizado en el gobierno del ex-mandatario Barack Obama. Este programa es empleado con el fin de “organizar, acceder y descubrir información útil de las grandes cantidades de datos de los que disponen su administración”. El resultado de los datos procesados permite anticipar el actuar de grupos que quieran realizar actos hostiles contra el pueblo norteamericano. El objetivo del programa Mind's Eye consiste en otorgarle a un sistema autónomo terrestre (UGV) una cámara inteligente que sea capaz de describir la escena que está visualizando. Esto se hace a través de verbos que puedan describir la escena. Además, podrá aprender nuevos conceptos a partir de su experiencia visual. La implementación de este programa ayudará sustancialmente en la minimización de riesgos, pues el combatiente no tendrá la necesidad de desplazarse (Carrillo *et al.*, 2013, p. 104).

Existe otro programa denominado Insight, cuyo objetivo es mitigar los costos de supervisar las oleadas de datos que pueden llegar a los combatientes en un momento determinado, con la finalidad de obtenerse en tiempo casi real. Para ello, este programa está trabajando un sistema de inteligencia, reconocimiento y seguimiento (ISR) en el que actúa el *big data* de la mano de imágenes que provienen de los sensores ISR ubicados en el campo de batalla (Carrillo *et al.*, 2013, p. 104).

En la defensa y seguridad nacional, el uso de imágenes es muy demandado debido a que estas se presentan como una herramienta que permite mejorar los procesamientos de imágenes en los siguientes aspectos: flexibilizar las infraestructuras para ejecutar soluciones a medida fácilmente; facilitar el procesamiento masivo en paralelo, dando potencia de cálculo a los sistemas; suministrar

información externa adicional con la cual poder fusionar información, mejorando su calidad y haciendo que esté compuesta de datos fidedignos y confiables (Carrillo *et al.*, 2013, p. 105).

Para el aprovechamiento del *big data* en el ámbito de defensa y seguridad nacional serán necesarias las infraestructuras respectivas, así como los grandes volúmenes por analizar, lo que puede significar en un principio que solo los países desarrollados, con los recursos necesarios, puedan aprovechar toda la amalgama de posibilidades que implica el *big data*. Sin embargo, cuando los costes se empiecen a hacer más accesibles, y se desarrollen más aplicaciones y herramientas, estas soluciones ganarán terreno en los sistemas que no puedan ser considerados como *big data*, y de igual manera serán implementados en países en vía de desarrollo (Carrillo *et al.*, 2013, p. 105). Asimismo, se puede llegar a aplicar el *big data* en la guerra electrónica, cuyo fin sería reducir o impedir el empleo del espectro electromagnético; la guerra electrónica se valdrá del espectro como campo de batalla.

En este punto es pertinente señalar las clases de acciones desplegables en el contexto de guerra electrónica, a saber: sincronización del momento de emisión con otros receptores situados en diferentes lugares para localizar el emisor (Angle of Arrival [AOA]); búsqueda y captación de cada frecuencia en la que no solo existe ruido; interceptación y mantenimiento de la frecuencia que emite para conocer sus parámetros; identificación y cotejo de dichos parámetros con emisores esperados o conocidos; localización; entre otros (Carrillo *et al.*, 2013, p. 109).

Aspectos legales y éticos en el procesamiento de información *big data*

Uno de los retos que enfrenta la aplicación del *big data* está relacionado con la utilización y explotación de datos, debido a que este procesamiento puede transgredir aspectos legales y éticos como son el derecho a la privacidad de los ciudadanos. Por ello, este aspecto es relevante en los ámbitos de la seguridad y la sanidad (Carrillo *et al.*, 2013, p. 71). La creciente utilización de tecnologías por las administraciones públicas trae consigo la necesidad de determinar si la infraestructura de tecnologías y la información propiamente dicha en que se funda la e-administración tienen una categoría especial de protección. Desde ya se debe decir que no existe norma colombiana que considere que el daño a la

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

infraestructura tecnológica de la administración pública sea considerado tan de alto nivel y, por ende, sea fundamental su protección.

Diferentes organizaciones de carácter gubernamental y no gubernamental implementan el *big data* como herramienta de procesamiento de información; no obstante, son las entidades públicas particularmente las que reciben y acceden a información de carácter sensible. Por esta razón, es necesario implementar nueva legislación que contemple este tipo de situaciones, ya que en la actualidad la legislación no cubre todo el tipo de posibilidades y usos de captación y análisis del *big data* (Carrillo *et al.*, 2013, p. 71).

El derecho de seguridad y defensa versus el derecho de privacidad: una mirada desde la jurisprudencia de la Corte Constitucional colombiana

El *big data* puede constituir una herramienta fundamental en la búsqueda de información, particularmente en lo concerniente a la defensa y seguridad de los Estados; sin embargo, su análisis y procesamiento puede llegar a vulnerar derechos fundamentales como la privacidad, la honra, el buen nombre y el debido proceso. Por esta razón, la Corte Constitucional ha analizado jurisprudencialmente dichas problemáticas, como se verá a continuación.

En la Sentencia T-277 de 2015 se analiza cómo con el *big data* las personas pueden acceder a todo tipo de información; sin embargo, se ha creado una figura denominada *Habeas data*, que tiene como pretensión crear la posibilidad de conocer, rectificar y actualizar las informaciones que se tienen sobre una persona en los bancos de datos públicos o privados. Las innovaciones de *big data* emergen en un momento en el que las organizaciones se enfrentan a nuevas problemáticas que se derivan de dos desafíos: disolución de los límites de las redes y extensión y apertura de las redes de datos de las organizaciones. Con ello se permitiría a socios, suministradores y clientes acceder a información de carácter corporativo a través de formas dinámicas para impulsar la innovación y la colaboración; y, a la final, ello podría derivar en que las redes sean más vulnerables al mal uso y el robo de información. Las aplicaciones y los datos corporativos son cada vez más accesibles por medio de servicios en la nube, por ejemplo.

El señor Roberto Eladio Espriella Fernández instaura una acción de tutela contra Ecopetrol S.A. con fundamento en lo indicado anteriormente, debido a que

la empresa lo señaló de pertenecer a grupos al margen de la ley (Sentencia T-022 de 2017). La empresa ejerce su derecho de defensa manifestando que la información publicada por ella el 19 de febrero de 2015 se refiere a hechos que ocurrían para ese momento. Por otro lado, señala que no puede hacerse responsable del sinnúmero de publicaciones periodísticas que surgieron a partir de esto.

De esta manera, el Juzgado Treinta Civil del Circuito concede las pretensiones del señor Espriella, pues evidentemente se vulneraban sus derechos a la honra, intimidad y buen nombre. Sin embargo, la decisión fue revocada por la Sala Primera Civil de Decisión del Tribunal Superior de Bogotá, razón por la cual la Corte entra a analizar el caso. Lo hace confirmando la decisión de primera instancia y señalando que la publicación, en efecto, violaba los derechos fundamentales del demandante anteriormente relacionados, ya que la información publicada por la Empresa no tenía un carácter público, solo le concernía a las partes, es decir, a la entidad y al trabajador. La decisión adoptada por la Corte se fundó, además de lo preceptuado en la Constitución, en la transgresión al Código de Ética, el Reglamento Interno de Trabajo y al Código Sustantivo del Trabajo.

Por otro lado, en la Sentencia T-277 de 2015 la Casa Editorial El Tiempo publicó una noticia en la cual se vinculaba a una persona con el delito de trata de personas, situación que vulneró los derechos de la accionante a la honra, el buen nombre y el debido proceso. No obstante, la editorial fue exonerada en la investigación por parte de la Fiscalía, pues había operado la prescripción de la acción. Pese a lo anterior, esta información seguía apareciendo en motores de búsqueda como Google y en las bases de datos de la editorial, situación que estaba afectando a la accionante y a su familia. Por esta razón, la accionante agotó la vía administrativa mediante derecho de petición ante la Casa Editorial El Tiempo; sin embargo, esta última se rehusaba a eliminar esta información de sus bases de datos argumentando que dicha información era veraz e imparcial. Por esta razón, la accionante acude a la acción de tutela, ya que para ella existe una doble discriminación: en principio se le adjudicaba un delito en el que ella no fue vencida en juicio y posteriormente esta editorial la señala públicamente como vinculada a estos delitos, sin rectificar o eliminar la información contenida, pese a que había operado la prescripción.

Para la Corte existe una contraposición de derechos relativa a la libertad de expresión vs. los derechos a la honra, el buen nombre y el debido proceso.

•El *big data* en la ciberdefensa y la ciberseguridad nacional.

Explica que el derecho a la libertad de expresión se refiere a la libertad de informar y recibir información veraz e imparcial; sin embargo, esto plantea unas problemáticas en torno a la aplicación de la red como herramienta que garantiza este derecho a la información. Así, establece que en la red —al ser descentralizada— los mensajes y los contenidos producidos se transmiten de forma tal que la revisión o censura previa de contenidos por parte de una autoridad central es una tarea compleja. En principio, esta parece ser una de las innovaciones y habilidades de internet, pues sin lugar a dudas lo hace un entorno libre, lo que posteriormente puede presentar ciertos retos en aspectos sensibles como los relativos al control de contenidos tendientes a afectar derechos como la intimidad, la honra, la imagen y el buen nombre de las personas.

La Corte señala en la Sentencia T-277 de 2015 el derecho de *Habeas data* del que gozan las personas, y refiere constitucionalmente el artículo 15 de la Carta Política, donde se consagra que las personas “tienen derecho a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bancos de datos y en archivos de entidades públicas y privadas. En la recolección, tratamiento y circulación de datos se respetarán la libertad y demás garantías consagradas en la Constitución”.

Lo anterior permite establecer que el derecho de *Habeas data* implica una doble vía: el conocimiento de información y la rectificación. Por eso, en el presente casi la Corte le ordena a la Casa Editorial El Tiempo la actualización de la información sobre los hechos que relacionan a la accionante con el delito de trata de personas, y que de igual manera se informe que esta no fue vencida en juicio. Por otro lado, le ordena a El Tiempo que por medio de la herramienta técnica “robots.txt”, “metatags” u otra similar neutralice el acceso a la noticia “Empresa de trata de blancas” a partir de la mera digitación del nombre de la accionante en los motores de búsqueda en la red (Sentencia T-277 de 2015).

Con el fin de evitar el uso indebido de datos tanto por entidades públicas como privadas, se han propuesto diferentes proyectos de ley. Un claro ejemplo de estas iniciativas fue el Proyecto de Ley 184 de 2010, ley estatutaria que pretendía la protección de datos personales en las bases de datos, con el fin de garantizar el derecho a la intimidad de los ciudadanos. Sin embargo, dicho proyecto de ley fue demandado en la Sentencia C-748 de 2011, ya que para el accionante tenía vicios en el procedimiento y vulneraba los principios a la unidad de materia.

Aunque dicho proyecto de ley cumplió con los requisitos constitucionales necesarios para ser Ley de la República, no fue aprobado por Senado y Cámara. Pese a esto, el proyecto de ley fue declarado executable en la mayoría de sus artículos y parecía un avance frente a la protección de datos, pues prohibía taxativamente la transferencia de datos personales de cualquier tipo a países que no proporcionaran niveles adecuados de protección de datos. Esto se presentaba como una gran evolución en materia de *big data*, pues imponía una obligación a los estándares fijados por la Superintendencia de Industria y Comercio sobre la materia; estándares que en ningún caso podían ser inferiores a los que la presente ley exigiera a sus destinatarios.

Otros aspectos significativos que vale la pena resaltar son las sanciones que contemplaba este proyecto de ley en su artículo 23, en el que se delegaba la facultad sancionatoria a la Superintendencia de Industria y Comercio:

[Podía] imponer diferentes tipos de sanciones, entre las que se encuentran: multas de carácter personal e institucional a favor de la Superintendencia de Industria y Comercio “hasta por el equivalente de dos mil (2000) salarios mínimos mensuales legales vigentes al momento de la imposición de la sanción. Las multas podrán ser sucesivas mientras subsista el incumplimiento que las originó. Suspensión de las actividades relacionadas con el tratamiento hasta por un término de seis (6) meses. En el acto de suspensión se indicarán los correctivos que se deberán adoptar. Cierre temporal de las operaciones relacionadas con el tratamiento una vez transcurrido el término de suspensión sin que se hubieren adoptado los correctivos ordenados por la Superintendencia de Industria y Comercio. Cierre inmediato y definitivo de la operación que involucre el tratamiento de datos sensibles”. (Corte Constitucional, Sentencia C-748 de 2011)

Es importante que se implemente una legislación que atienda a los diferentes cambios sociales, culturales y, especialmente, tecnológicos. Solo así se podrá mitigar la diferencia que existe frente a la posición dominante por parte de las empresas, corporaciones y el mismo Estado en la aplicación del *big data*; ello, atendiendo por supuesto a la Constitución Política y los avances legislativos que ya se generan sobre esta materia (Sentencia C-748 de 2011).

Conclusiones

A través de *big data* se pueden instrumentalizar actividades de inteligencia que servirán para la seguridad y defensa nacional, con el ánimo de prevenir la creciente incursión de cibercriminales y ciberterroristas en el ciberespacio. Existe una confluencia en las herramientas tecnológicas utilizadas para garantizar la seguridad y defensa nacional, que utilizan técnicas como el *big data*, en cuanto instrumento capaz de almacenar una gran cantidad de información, que en algunos casos confluye con información que afecta la intimidad de los ciudadanos.

La utilización de tecnologías de la información y las comunicaciones y, de forma específica, los desarrollos tendientes a dotar de capacidad tecnológica a los organismos de seguridad y defensa del Estado deben realizarse desde la ponderación y el respeto por los principios de idoneidad, necesidad y proporcionalidad, y desde el respeto de los derechos de los ciudadanos, apelando al cumplimiento de la finalidad legítima.

Es necesario contar con regulaciones claras que den seguridad jurídica a la práctica empleada, todo bajo el entendimiento de los cambios sociales, como en la implementación tecnológica. Debe señalarse que de forma tradicional la información era de las organizaciones, y estas la protegían; ahora todos los ciudadanos ponen, generan, producen información que está en redes sociales, en perfiles públicos, así como informaciones dispuestas por las administraciones públicas, lo que genera que esta pueda ser usada por cualquier tipo de persona y para fines diversos, incluso criminales. Esto se convierte en uno de los grandes retos del Estado: poder enfrentar esos nuevos fenómenos desde el análisis de los datos que circulan.

Los múltiples canales, los diversos tipos de datos y, sobre todo, las diversas formas en que estos se presentan se convierten en la esencia del tratamiento, pues vincular fotografías, datos de redes sociales, sonido, entre otros, deberá permitir la toma de decisiones que protejan la seguridad y defensa de los Estados. Estos últimos, desde el análisis de los datos, deben dotarse de capacidades que les permitan tener conciencia situacional, para de esa forma enfrentar las amenazas. No se podrá defender activa o pasivamente si no se cuenta con el conocimiento claro, real, exacto y permanente de la situación.

OPEN DATA Y BIG DATA: HERRAMIENTAS DE SOFTWARE PARA CIUDADES INTELIGENTES (CASO DE ESTUDIO)

Este caso de estudio forma parte del proyecto de investigación Derecho y Big Data, en el que participan dos grupos de investigación de la Universidad Católica de Colombia: a) Investigación en Derecho Público y TIC, perteneciente a la Facultad de Derecho; b) Software Inteligente y Convergencia Tecnológica (GISIC), adscrito a la Facultad de Ingeniería. Además, cuenta con la participación de dos colaboradores externos: la Universidad de Texas, en Estados Unidos, y el Departamento Nacional de Planeación (DNP); esto último, a partir de un convenio celebrado entre la Universidad y la entidad del Estado.

Las ciudades inteligentes o *smart cities* buscan optimizar sus procesos con el objetivo de mejorar la calidad de vida de sus ciudadanos, mediante la recopilación de datos y utilizando dispositivos de *software* y *hardware* —por ejemplo, cámaras de video, sensores y *smartphones*—. De esta manera, el Gobierno puede identificar los posibles problemas que se encuentran en alguna sociedad en particular e incluso adelantarse a situaciones que se pueden convertir en caos (Anaya, 2017).

El término *big data*, o ‘datos masivos’, indica grandes volúmenes de datos, de toda variedad, que se procesan a grandes velocidades con el fin de lograr un potencial de valor incalculable sobre ellos. La cantidad de datos generados pueden provenir del sector público o privado; es decir, conjuntos de datos pueden

convertirse en *big data* independientemente del sector macroeconómico, lo cual es una gran ventaja por cuanto el término es transversal a cualquier empresa y sociedad. Por su parte, *open data*, o ‘datos abiertos’, abarca todos los datos generados por entidades gubernamentales con el fin de ser utilizados, reutilizados y redistribuidos libremente por cualquier persona natural o jurídica, independientemente del propósito o beneficio que se tenga destinado (Vetrò *et al.*, 2016). *Open data* también engloba los datos que se almacenen en medios físicos —como los documentos impresos— y que deban ser mantenidos en esta forma por normatividad de la empresa o del Estado.

Aporte de *big data* y *open data* a ciudades inteligentes

Open data nace a partir de la ejecución de los procesos operativos y estratégicos en las entidades gubernamentales. El área de tecnología es normalmente la encargada de consolidar los datos que se van a compartir al público, ya que por su naturaleza misional debe gestionar datos para la elaboración de conjuntos de datos que son compartidos a diferentes instancias públicas y privadas. En este caso, debe publicar y compartir los datos en algún repositorio en internet destinado a datos abiertos.

Las áreas de tecnología deben tener la habilidad de identificar cuándo *open data* se torna *big data*, pues gestionar grandes volúmenes de información en varias empresas es engorroso para la elaboración de informes: el procesamiento tarda mucho tiempo y el almacenamiento es inmanejable, ya que alcanza los límites de los servidores. Por tal motivo, es crucial identificar mediante métricas cuándo el conjunto de datos forma parte de *big data* con el fin de garantizar una gestión adecuada de los datos. A continuación se presentan algunas características que ayudarán a las entidades gubernamentales a identificar cuándo *open data* se convierte en *big data*:

- Hasta 10 GB de datos en cualquier formato de archivo (p. e., Excel, Word, Power Point, etc.) se definen como *small data*.
- Entre 10 GB y 1024 GB (1 terabyte) de datos, normalmente almacenados en un gestor de datos, se les conoce como *base de datos*; por ejemplo, MySQL, SQL, ORACLE, entre otros.

• *Open data y big data*: herramientas de *software* para ciudades inteligentes.

- De 1 terabyte de datos en adelante se define como *big data*; para este tipo de volúmenes de datos normalmente se requieren bases de datos distribuidas y con gran capacidad de almacenamiento en los servidores.

Contexto en Colombia

En Colombia actualmente existen grandes volúmenes de información generados a partir de múltiples fuentes de datos y que normalmente evidencian una baja calidad en su contenido. Por ejemplo, una fuente de datos es el repositorio nacional de información www.datos.gov.co, utilizado por las entidades públicas para compartir en internet información financiera, contractual, logística y administrativa, entre otras, con el fin de dar cumplimiento a la Ley de Transparencia y Acceso a la Información Pública; sin embargo, este portal no suministra los datos con métricas de completitud, coherencia y exactitud, lo cual afecta directamente la calidad de la información, por ejemplo, al limitar la manera de identificar problemas de corrupción y otros.

En el sector privado hay un gran interés por el uso de datos de gobierno para diferentes propósitos; por ejemplo, para realizar seguimiento sobre los procesos contractuales que involucran dinero o para revisar el comportamiento de indicadores de educación. Sin embargo, en Colombia hay desafíos que aún se deben abordar antes de utilizar los datos abiertos. El primero de ellos es el aseguramiento de la calidad de los datos suministrados por las entidades gubernamentales, en términos de completitud, coherencia, oportunidad y consistencia. El segundo desafío es la claridad con la que se debe presentar la información, ya que, debido al volumen que se puede manipular, normalmente se cae en el error de querer mostrar olas de información, lo que, en consecuencia, colige desinformar a los ciudadanos. El tercer desafío es la importancia que deben darle las mismas entidades gubernamentales, ya que el adecuado uso de datos abiertos trae asociados valores agregados como la transparencia y optimización de procesos y recursos del Estado.

El tratamiento de los datos abiertos que se gestiona en las entidades gubernamentales está regido por la “Ley de Transparencia de Datos y del Derecho al Acceso de la Información Pública Nacional”, la cual exhorta a publicar y compartir los datos derivados de la gestión pública para que sean alcanzables, utilizados,

reutilizados y redistribuidos por la sociedad colombiana. Para el cumplimiento de la ley, las entidades gubernamentales publican y comparten sus datos en la página www.datos.gov.co o en cualquier portal en internet que cumpla los estándares necesarios para la aplicación de la ley. En estos portales se encuentran conjuntos de datos que son clasificados de acuerdo con los procesos derivados de la gestión pública; por ejemplo, datos relacionados con procesos financieros, contables, contractuales, a excepción de información que es confidencial o secreto de Estado.

Contexto internacional

La economía mundial se ha centrado en los datos; en consecuencia, quienes tengan las capacidades de extraer el máximo beneficio de sus datos tendrán el poder en términos políticos, sociales, culturales y, especialmente, económicos. En los últimos años, un número creciente de gobiernos ha comenzado a abrir sus datos. “Este movimiento llamado gobierno abierto ha resultado en el lanzamiento de numerosos portales de datos abiertos y de infraestructuras que tienen como objetivo proporcionar un punto único de acceso a datos del gobierno y explorar sus consecuencias” (Máchová y Lnénicka, 2017, p. 21).

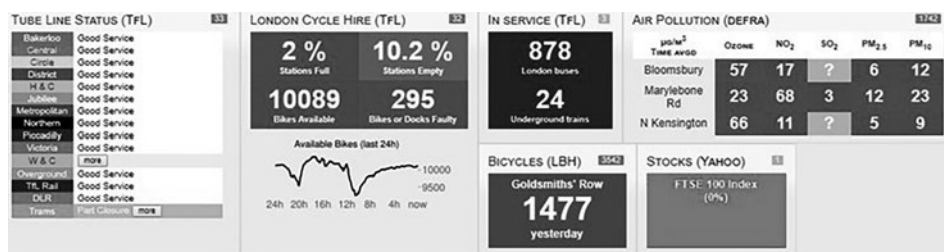
Los reportes en línea y tableros de indicadores en las ciudades miden el progreso de las ciudades utilizando métricas urbanas y mejorando sus estrategias de operación a medida que se disponga de nuevos datos, mediante la toma de decisiones informadas. Un ejemplo de ello es lo realizado en ciudades como Seúl, Chicago, Nueva York, Londres, entre otras, “donde se evidencia una mejora en la transparencia y la responsabilidad con la ciudad por parte del gobierno” (Martin y Begani, 2016).

Por otra parte, en los países desarrollados, el éxito de los proyectos relacionados con ciudades inteligentes está intrínsecamente relacionado con la existencia de grandes volúmenes de datos que podrían ser procesados para alcanzar sus objetivos. Ciudades o países como “Vancouver, Portland (Oregon), San Francisco (2009), Nueva York (2012) y Nueva Zelanda (2011) tienen sitios web donde publican datos sin requerir licencia facilitando su acceso y explotación por parte de terceros, generalmente empresas privadas y sin ánimo de lucro.

•Open data y big data: herramientas de software para ciudades inteligentes.

La figura 1 presenta un ejemplo de un tablero de indicadores que permite ver en tiempo real la información de Londres de manera resumida; así, se le permite a cualquier ciudadano hacer seguimiento de datos relacionados con el transporte y el aire, entre otros; consecuentemente, este tipo de información hace posible que cualquier ciudadano o empresa tome decisiones de acuerdo con su situación particular. En detalle, el tablero de indicadores presentado parte con los indicadores del estado de las líneas del metro, donde la mayor parte de ellas se encuentran en funcionamiento. Luego se disponen indicadores de la afluencia de bicicletas en las estaciones, incluyendo aquellas disponibles para utilizar. Posteriormente se registra la cantidad de buses y trenes disponibles para la prestación del servicio. Por último, se muestran indicadores de polución.

Figura 1. Tablero de indicadores de Londres, Inglaterra



Como caso de éxito, en la ciudad de Bandung, una de las más grandes de Indonesia, se propuso desarrollar una aplicación de monitoreo urbano a través de un tablero que permitiera resumir en tiempo real la condición climática, de transporte, económica, de salud y energía. El sistema de arquitectura utiliza sensores de red, que consiste en nodos con sensores que tienen la función de capturar las condiciones de la ciudad, como la temperatura, el nivel de polución del aire, el nivel de agentes contaminantes del agua y temas relacionados con el tráfico. La conclusión de este proyecto internacional se basa en el éxito alcanzado, por haber puesto en marcha el tablero de instrumentos de la ciudad inteligente y proporcionado información a los ciudadanos (Suakanto *et al.*, 2013, pp. 1-2).

***Big data* y sus desafíos en las ciudades inteligentes**

Big data hace referencia a la manipulación de grandes volúmenes de datos (Khan *et al.*, 2017; Kacfeh, Cullot y Nicolle, 2015); sin embargo, hasta el momento ni las ciencias de datos, ni la industria, ni la academia han logrado concluir cuál es la cantidad mínima de datos para otorgar tal nombre. Dos factores que podrían ayudar a determinar si el conjunto de datos forma parte de *big data* son la utilización de herramientas informáticas especiales para la administración de datos y la infraestructura física para procesarlos (Sivarajah *et al.*, 2017; Oussous *et al.*, 2017).

Según lo anterior, por ejemplo, un conjunto de datos que se almacenan en hojas de cálculo de Excel no podría ser considerado como *big data*, ya que el *software* utilizado forma parte de la *suite* de ofimática que un computador normalmente tiene instalado. En este escenario se evidencia capacidad limitada de almacenamiento y procesamiento. En específico, *big data* determina una utilización avanzada de recursos físicos, tales como servidores capaces de gestionar grandes volúmenes de datos de manera eficiente y por largos periodos (Khan *et al.*, 2017). Nuevos métodos matemáticos y tecnologías también son un factor fundamental para la limpieza, el análisis y el procesamiento de datos, con propósitos múltiples:

- Suministrar información a través de indicadores que facilitan la toma de decisiones en las empresas.
- Permitir un análisis detallado de los datos, llevando a las empresas a ser más competitivas en la personalización de servicios.
- Generar nuevas estrategias de mercadeo a partir del descubrimiento de nueva información.

Tipos de datos

Existe una clasificación de datos de acuerdo con su organización interna, definida con el fin de poder guardar, consultar, actualizar y eliminar información. De esta manera, las herramientas tecnológicas pueden gestionar adecuadamente los datos y cualquier operación que se desprenda de estos. A continuación se presenta la clasificación en mención:

• *Open data y big data: herramientas de software para ciudades inteligentes.*

- *Estructurados.* Hace referencia a los datos normalizados que se almacenan en bases de datos relacionales; por ejemplo, Oracle, MySQL o Sybase. Esto no aplica a datos que se encuentran en hojas de cálculo, ya que, por su volumen, no forman parte de *big data*.
- *No estructurados.* Son aquellos datos que carecen de organización estructural o que no se encuentran normalizados. Por ejemplo, datos en formato de audio, video y texto.
- *Semiestructurados.* Los datos semiestructurados no se ajustan a estándares estrictos; un ejemplo es el lenguaje XML (Extensible Markup Language), que contiene etiquetas que pueden ser legibles por cualquier máquina (Gandomi y Haider, 2015).

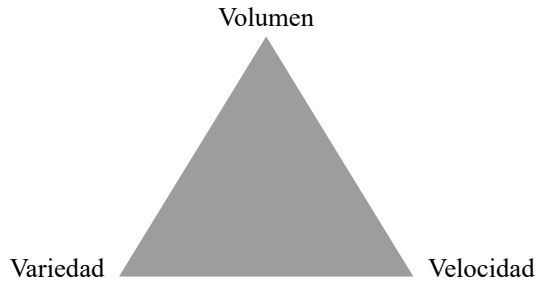
Evolución del termino

Aunque siempre han existido grandes volúmenes de datos de manera digital, el termino *big data* fue oficialmente utilizado en 1941, de acuerdo con el diccionario de inglés de *Oxford*. Posteriormente, fue consolidándose debido al crecimiento de datos que se generan en internet. Una evidencia de esto es la creación diaria de datos en fuentes tales como redes sociales, *smartphones*, carros y casos inteligentes, blogs, entre otros; 2,5 quintillones de bytes de datos son generados por estas fuentes diariamente, por lo cual *big data* es y seguirá siendo un concepto comúnmente utilizado (Actuaries, 2015).

Características

Las características de *big data* normalmente están orientadas a las tres *V*: volumen, variedad y velocidad; sin embargo, día tras día ha ido ampliando el número de características. En la figura 2 se presentan las comúnmente utilizadas de acuerdo con Actuaries (2015), Kacfeh *et al.* (2015) y Oussous *et al.* (2017).

Figura 2. Características del big data, o las 3 V



- a. *Volumen*. Esta característica está orientada a la utilización de grandes volúmenes de datos; por ejemplo, son los datos que el mismo internet genera diariamente (de acuerdo con la empresa International Data, 4.4 zetabytes son producidas anualmente).
- b. *Variedad*. Se focaliza en los diferentes formatos en que los datos se encuentran estructurados. Normalmente, los datos almacenados no poseen un orden definido ni tampoco se encuentran disponibles para procesar inmediatamente. Como se indicó, los datos se pueden encontrar de manera: a) estructurada (almacenada en bases de datos relacionales), b) semiestructurada (correo electrónico, sensores o redes sociales) y c) sin estructurar (video, imágenes o audios).
- c. *Velocidad*. Indica el flujo de información entre y hacia diferentes fuentes de datos. Esto involucra acciones tales como la creación, actualización y eliminación de datos. También incluye la oportunidad de acceder a la información, es decir, la disponibilidad y rapidez con que un usuario o sistema de información acceden a la información; por ejemplo, la rapidez con que se sube o descarga la información en una plataforma tecnológica.

Tecnológicas en big data

- a. *Apache Hadoop*. Es una herramienta de código abierto que se liberó en internet en 2011; su objetivo es el procesamiento de grandes volúmenes de datos a través de sistemas distribuidos, utilizando algoritmos computaciones simples (Actuaries, 2015). El procesamiento de datos se hace mediante un procesamiento en paralelo de hasta cientos de máquinas; incluso es capaz de contemplar errores causados por daños en el *hardware*

• *Open data y big data: herramientas de software para ciudades inteligentes.*

de las máquinas, garantizando así el correcto funcionamiento de esta herramienta. Empresas dedicadas a la recopilación de información en el ámbito mundial adoptaron Hadoop para incrementar sus ingresos y ser más competitivas (como Amazon, Facebook, Twitter y Google) (Oussous *et al.*, 2017).

- b. *Bases de datos no relacionales.* Son bases de datos que nacieron para gestionar *big data*, específicamente grandes volúmenes y variedad de datos. Las herramientas que se encuentran para almacenar datos son Oracle NoSQL, Apache Cassandra y Hive (Kacfeh *et al.*, 2015).
- c. *Data mining.* Un conjunto de métodos, algoritmos y técnicas utilizado para manipular datos; tiene como objetivo encontrar datos escondidos, los cuales no son visualmente perceptibles para el ser humano. En *big data*, la herramienta Map Reduce es comúnmente utilizada para practicar minería de datos.

Aspectos legales

Las empresas incluyen como procesos recopilar y procesar datos, con el fin de analizar los comportamientos de sus clientes; de esta manera, se pueden personalizar los productos, las ofertas y las promociones, ofreciendo aparentemente beneficios a una población específica de clientes fidelizados o potenciales. En este escenario, *big data* presenta oportunidades para las empresas, por cuanto les permite ser más creativas y competitivas a partir del análisis descriptivo y predictivo de datos, que ayudan a conocer el comportamiento de los clientes a partir de datos históricos que estos mismos han ido acumulando a partir de sus compras de bienes o servicios.

La protección de los datos personales es primordial durante el proceso de recopilación, procesamiento y análisis de información, debido a que la información básica de los clientes debe ser anónima, para prevenir discriminación y abuso durante el mercadeo digital (por ejemplo, enviar datos básicos tales como nombres, direcciones y números de documento a usuarios de manera errónea, poniendo en riesgo la privacidad). En el ámbito mundial se han establecido regulaciones, leyes y acuerdos que ayudan a garantizar el manejo de datos personales y la privacidad

de las personas. A continuación se presenta la evolución de los recursos legales que han concedido a las personas el derecho a la protección de sus datos:

- La Convención para la Protección de las Personas con Respecto al Procesamiento Automático de Datos Personales; adoptada por el Consejo Europeo en 1981.
- Actualmente, el instrumento legal para la Unión Europea es la Directiva de Protección de Datos (9546/EC), establecida en 1995, que regula la protección de las personas con respecto al procesamiento de datos y el libre movimiento de estos. Esta directiva es adoptada por cada uno de los países miembros de la Unión Europea.
- E-Privacy es otro instrumento legal (2012/58/EC) definido en la Unión Europea con el fin de garantizar la privacidad de los datos en el marco de servicios que se emplean a través de las comunicaciones electrónicas. Incluye el uso adecuado de dispositivos que se encargan de recopilar y almacenar información a través de *cookies*, las cuales deben ser informadas al usuario de internet, quien a su vez es libre de decidir si autoriza o no el manejo de sus datos.
- Regulación General para la Protección de Datos (GDPR), establecida en 2015 con el fin de aumentar la cobertura de la protección de datos y mejorar las oportunidades de negocio a partir del intercambio de información entre diferentes entidades.
- Directiva de Empleo Igualitario (2000/78/EC), orientada al manejo de datos evitando la discriminación de cualquier tipo: racial, étnica, religiosa, etc.
- Directiva de Servicios y Mercancías (2004/113/EC), la cual busca la utilización adecuada de los procesos inherentes al flujo de datos.
- En Colombia, la Ley 1581 de 2012 dicta un conjunto de disposiciones generales para la protección de datos personales. Con ella se busca garantizar el derecho a conocer, actualizar y rectificar cualquier información que se haya recopilado de los ciudadanos. Adicionalmente, la ley establece principios que garantizan la calidad, el tratamiento y la legalidad de los datos.

Big data analytics

El proceso de analítica de datos es utilizado para explorar, procesar y entender los datos y las relaciones que existen entre ellos. De esta manera, las empresas pueden extraer conocimiento invaluable para validar o establecer nuevas estrategias de mercadeo. Adicionalmente, se pueden determinar patrones que afectan positiva o negativamente los procesos de negocio (Oussous *et al.*, 2017). Dentro del proceso de analítica de datos existen técnicas tales como minería de datos, visualización de datos, análisis estadísticos y aprendizaje de máquina, que están basadas en algoritmos matemáticos que apoyan el procesamiento de datos para un fin específico; estos algoritmos están soportados en arquitecturas de *hardware* y *software*.

Por otra parte, también se han establecido métodos para el procesamiento de datos; su selección depende del conjunto de datos que se van a utilizar. A continuación se presentan los métodos comunes de acuerdo con el tipo de datos que se utilizan:

- *Extracción de información*. Consiste en extraer datos estructurados a partir de datos no estructurados. Posee dos actividades:
 - *Reconocimiento de entidades (ER)*. Toma el texto y lo clasifica de acuerdo con unas entidades (categorías) predefinidas.
 - *Extraer relaciones (RE)*. Se extraen relaciones semánticas entre entidades.
- *Resumen de textos*. Resume el contenido de uno o varios documentos. Tiene dos puntos de abordaje:
 - *Extracción de resumen*. El resumen es un subconjunto de todas las unidades de texto que componen el documento. La importancia de las unidades de textos depende de la frecuencia con que se repite una frase o palabra, así como su lugar, para luego construir el resumen (conjunto de unidades salientes). La ventaja es que no se requiere entender el texto analizado.
 - *Abstracción de resumen*. Es una técnica que consiste en analizar el texto semánticamente; luego se analizan y se incorporan algunas otras palabras a partir del estudio semántico utilizando técnicas de

lenguaje natural (NLP). Generalmente, este tipo de resúmenes son más coherentes que los resúmenes de extracción.

- *Respuesta a preguntas.* Suministra respuestas a preguntas utilizando lenguaje natural. Apples Siri o IBM Watson son un ejemplo.
 - Estos sistemas dependen de técnicas desarrolladas para lenguaje natural (NLP).
 - Existen tres categorías de preguntas a respuestas automáticas:
 - *Information retrieval (IR).* Se clasifica el tipo de pregunta, se analiza la información buscando en pasajes u oraciones que se precargan en el sistema y se determinan las respuestas candidatas.
 - *Basado en conocimiento.* Depende del contexto, la medicina, el turismo, etc.
 - *Híbrido.* Es la composición entre las dos categorías; es decir, se precargan las preguntas y respuestas, y al mismo tiempo se entrena al sistema con la información suministrada por el experto.
- *Análisis de sentimientos (opinion mining).* Consiste en definir qué tipo de opinión existe en una persona hacia una empresa específica, de acuerdo con su producto, servicio, etc. Existen tres tipos de análisis:
 - *Basado en documento.* Se lee todo un documento y se determina si tiene una posición positiva o negativa frente a la empresa.
 - *Basado en sentencias.* Se analiza una sentencia en particular y se determina el tipo de sentimiento.
 - *Basado en aspectos.* Se reconocen todos los sentimientos de una persona o conjunto de personas frente a una empresa, y se determina hacia qué parte de la empresa se tienen esos sentimientos; por ejemplo, un producto, servicio, área, etc.
- *Análítica de audio.* Consiste en analizar discursos, ligando sonidos con palabras. Normalmente, aplica a formatos de audio y video.
- *Análítica de video.* Consiste en analizar contenido de video; sin embargo, las técnicas aún no se encuentran en desarrollo, sobre todo cuando se quiere analizar video en tiempo real, debido a la gran cantidad de video. Por ejemplo, un segundo de alta definición de video equivale a 2000 páginas de texto. Si se quisiera analizar Youtube, sería una tarea titánica debido a que 100 horas de video son subidas cada minuto. Hay varias

• *Open data y big data*: herramientas de *software* para ciudades inteligentes.

ramas en las que se puede desarrollar análisis de video; por ejemplo, en la de mercadeo, específicamente en la situación en la que se requiera analizar cuántas personas ingresaron a una tienda, el género, la raza, la edad, el tiempo en que estuvieron, los patrones de comportamiento y el tiempo en una fila.

- *Análítica de redes sociales*. Se analizan datos estructurados y no estructurados producidos a partir de los eventos que generan los usuarios de redes sociales como Facebook, Twitter, LinkedIn, entre otras.

Desafíos

Desde sus múltiples perspectivas, *big data* ha venido desarrollándose debido a las necesidades y los beneficios que representa; sin embargo, aún existen desafíos por abordar y resolver desde diferentes áreas de conocimiento; los más sobresalientes son desafíos de datos, desafíos de procesos y desafíos de *software* y *hardware* (Sivarajah *et al.*, 2017; Bertot y Choi, 2013).

Desafíos de datos

Están orientados al aseguramiento de la calidad de los datos; específicamente seguridad, integridad, veracidad, velocidad, disponibilidad, volatilidad, gobernabilidad de los datos, privacidad, entre otros. El desarrollo de estándares que apoyen el aseguramiento de la calidad de los datos va a permitir que agencias, investigadores, científicos, compañías del sector privado y público, entre otros, gestionen datos de manera adecuada, con el fin de brindar análisis de resultados sin errores y con una alta calidad. De este modo, se impacta positivamente la toma de decisiones en las empresas (Bertot y Choi, 2013).

Desafíos de procesos

En este tipo de desafíos se busca investigar y solucionar técnicas para capturar datos, integrar datos a partir de sistemas de información, transformar datos, seleccionar el método adecuado para el análisis de información, etc. En línea con Sivarajah *et al.* (2017), los métodos comúnmente utilizados son:

- *Análisis descriptivo*. Describe la situación actual de una situación perteneciente al modelo de negocio analizado. Esta descripción permite abordar análisis de manera evidente, explicando patrones de comportamiento y excepciones.
- *Análisis predictivo*. Su objetivo es predecir comportamientos y posibilidades futuras del negocio a través de algoritmos y modelos estadísticos.
- *Análisis inquisitivo*. Se encarga de probar o certificar datos del negocio; por ejemplo, analizar factores de riesgo.
- *Análisis prescriptivo*. Su objetivo es optimizar y evaluar cómo el negocio puede mejorar sus niveles de servicio y al mismo tiempo disminuir sus gastos.

Desafíos de *software* y *hardware*

Las arquitecturas de *software* y *hardware* requeridas para obtener una plataforma que soporte *big data* son importantes con el fin de garantizar la recopilación, el procesamiento y el almacenamiento de información. Estos dos tipos de arquitecturas son dependientes, ya que si alguna de estas carece de las tecnologías mínimas requeridas para la gestión de *big data*, podría convertirse en una plataforma obsoleta. La capacidad computacional debe ser robusta, pues manipular *big data* así lo requiere. Esto es conocido como *super computing*, y es lo mínimo requerido para procesar largos conjuntos de datos, aplicaciones y visualizaciones para el análisis de datos. Adicionalmente, el almacenamiento de estos datos implica preservarlos por largo tiempo y aplicando políticas de gestión de datos.

Open data* y sus desafíos en *smart cities

Open data o datos abiertos

Open data es una tendencia mundial en la que los gobiernos se comprometen a publicar y compartir los datos derivados de la gestión pública, con el fin de contribuir a la transparencia como medio de confianza entre los ciudadanos y sus gobernantes. Esto también trae como consecuencia que empresas del sector privado y personas naturales puedan usar, reutilizar y distribuir los datos a través

• *Open data y big data: herramientas de software para ciudades inteligentes.*

de aplicaciones en línea (Datos.bcn.cl, 2017). La única condición en la mayoría de leyes que invitan a la aplicabilidad de los datos abiertos es que estos no sean alterados y se publiquen de la misma manera en la que se tomaron del repositorio de internet.

La definición del término *open data* puede variar entre autores y fuentes; sin embargo, todos coinciden en que son los datos originados de la gestión pública, publicados y compartidos a través de un repositorio de datos abiertos en internet, con el fin de que estos sean utilizados, reutilizados y redistribuidos libremente por cualquier persona natural o jurídica. El objetivo de *open data* es poner a disposición de la sociedad los datos del gobierno, para que estos se conviertan en información a través de *software* en internet que genere valores agregados para la toma de decisiones y la optimización de procesos, y de esta manera, se mejore la calidad de vida de los ciudadanos. Por tal razón, es crucial concientizar a los gobiernos mundiales y a las instituciones privadas acerca de los cambios positivos que esta tendencia trae para la humanidad, ya que transforma los esquemas tradicionales de los gobiernos, independientemente de la región en la que se encuentren ubicados (Repositorio.cepal.org, 2016).

El concepto de *open data* en el Gobierno colombiano se basa en que los datos generados por las entidades públicas pertenecen a la sociedad, dado que han sido financiados y recopilados con dinero público y, por tanto, deben estar disponibles para personas naturales o jurídicas, independientemente de su propósito e interés (Herramientas.datos.gov.co, 2016).

Una ventaja que brinda la implementación de *open data* en cualquier país es la posibilidad de que la ciudadanía se involucre en el análisis de datos, con el fin de mejorar indicadores de transparencia y eficiencia en la ejecución de procesos gubernamentales. A partir de lo anterior, también se obtiene:

- Generación de mayores ingresos para las empresas privadas y las personas naturales, ya que poseen información para analizar y generar estrategias que apoyen la creación de nuevos servicios o productos orientados a mejorar la calidad de vida de los ciudadanos. Por ejemplo, la ciudad de Seúl ha creado *startups* que ayudan a mejorar los procesos relacionados con el transporte público.

- Socialización de datos que permiten mantener informado a todo un país acerca de las decisiones que se toman y se ejecutan en las entidades gubernamentales.
- Implementar nuevas aplicaciones de *software* que complementen procesos ya creados, para de esta manera monetizar el nuevo conocimiento descubierto.

En conclusión, los datos abiertos buscan la interoperabilidad entre sistemas de información y personas, con lo cual se genera independencia entre fuentes de datos, repositorios de información, variedad de información y su destino.

Características

Según la Open Knowledge Foundation (s. f.), los datos abiertos son caracterizados a través de los siguientes aspectos:

- a. *Uso*. Indica la disponibilidad de los datos en tiempo real y cuando estos se requieran, utilizando internet como medio para su acceso. Los desafíos que se deben resolver para esta característica, en términos de estandarización, son los formatos, la ambigüedad, el descubrimiento y la representación de los datos.
- b. *Reutilización y modificación*. Consiste en la utilización y transformación de los datos obtenidos a partir de un portal web, el cual le permite a cualquier ciudadano acceder a los datos y generar un valor agregado a partir de ello, siempre y cuando los datos no se alteren.
- c. *Participación universal*. Los datos deben y pueden ser accedidos por cualquier ciudadano, es decir, no debe existir ninguna restricción de acceso a los datos a ciudadanos o grupos por su condición racial, étnica, etc.

Calidad de datos abiertos

A partir de las leyes que apoyan la gestión de datos abiertos en los países, las entidades publican y comparten la información; sin embargo, los formatos y las estructuras de los datos varían drásticamente entre las entidades del Estado, dado que la naturaleza de cada entidad es diferente. La calidad de datos surge como un proceso de control debido a estas múltiples fuentes de datos y la manera como se

comparten. El nivel de calidad de los conjuntos de datos es determinado por el grado de apoyo en diferentes escenarios para la toma de decisiones informadas. Este grado es calculado de acuerdo con la aplicación de métricas matemáticas y criterios sobre los conjuntos de datos por utilizar. A continuación se definen cinco métricas para la medición de la calidad de los datos:

- a. *Relevancia*. Los conjuntos de datos compartidos y seleccionados representan importancia para la toma de decisiones informadas. Las entidades públicas deben determinar si la información que se va a compartir ayudará a optimizar procesos en el estado, o bien, si simplemente son datos sin ningún valor para la ciudadanía.
- b. *Exactitud*. Los datos deben ser totalmente precisos, independientemente de que sean numéricos o no. Por ejemplo, en un proceso contractual, un ciudadano quiere evaluar la variable *valor del contrato*, pero este no es exacto, ya que la entidad únicamente compartió valores globales y aproximados. Esta situación en particular, seguramente, tendrá imprecisiones en la información que se genera para la toma de decisiones y, en consecuencia, conllevará la pérdida de credibilidad.
- c. *Oportunidad*. Los datos pueden y deben estar disponibles en el portal de internet, con el fin de que sean accesibles en el momento en que los ciudadanos los requieran. Por ejemplo, si la persona quiere acceder a un conjunto de datos, pero la entidad le manifiesta que esto no es posible y que, por tanto, debe esperar para que estos sean compartidos, se indica que no hay oportunidad en los datos, ya que estos no pueden ser accedidos.
- d. *Comparabilidad*. Indica la facilidad con la que diferentes variables se pueden contrastar, independientemente del conjunto de datos. Hay situaciones en las que se requieren comparar valores de gasto a través de diferentes conjuntos de datos; sin embargo, hay circunstancias en las que las variables, a pesar de que tienen el mismo nombre y objetivo, no cuentan con iguales valores, lo cual indica que no hay comparabilidad entre los conjuntos de datos.
- e. *Complejidad*. Esta característica evalúa si los valores de cada variable se encuentran definidos. Por ejemplo, en un conjunto de datos donde la variable *ciudad* es crucial para la toma de decisiones, la entidad, no

obstante, no comparte los datos con estos valores; para este ejemplo en particular, el conjunto de datos no está completo.

Datos abiertos en Colombia

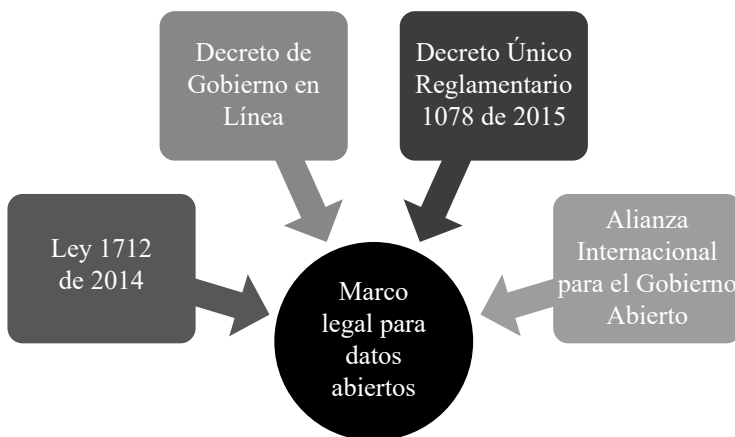
Datos Abiertos es un proyecto que inició en 2011 en el marco de la estrategia de Gobierno en Línea (GEL) y la Alianza Internacional para el Gobierno Abierto (AGA-OGP), cuyo objetivo principal fue compartir y publicar los datos derivados de la gestión administrativa pública. Los principales retos que se abordaron en el proyecto fueron brindar datos de forma oportuna y compartirlos en un formato que facilitara la utilización, reutilización y distribución por parte de los ciudadanos o empresas que quisieran su acceso.

Los objetivos del Proyecto de Datos Abiertos abarcaron diferentes puntos de vista: fomentar la utilización de datos abiertos, con el fin de generar servicios innovadores que suministren soluciones a problemas sociales en los ámbitos de la educación, la cultura, la salud, la seguridad, entre otros; involucrar a sectores como la academia, la industria, la sociedad civil y las organizaciones no gubernamentales, con miras a motivar su participación mediante la financiación y el desarrollo de proyectos que generen valor agregado a los datos abiertos; ser un ejemplo en la región por las buenas prácticas en la implementación de la estrategia de datos abiertos.

Los actores identificados que eventualmente podrían generar valor sobre los datos abiertos en Colombia son: academia, entidades públicas y privadas, emprendedores, ciudadanos inteligentes, medios de comunicación y cualquier veedor público. El proyecto definió cuatro etapas para su implementación: a) soporte a las entidades públicas para la publicación de datos abiertos en formatos estructurados que permitan su utilización, reutilización y distribución en el portal *www.datos.gov.co*; b) otorgamiento de incentivos a los actores que implementen soluciones o servicios basados en datos abiertos; c) mantenimiento y mejora continua del portal *www.datos.gov.co*, con el fin de garantizar el funcionamiento del portal destinado para publicar los datos abiertos; f) apoyar el uso de herramientas para la visualización de datos, las cuales ayudarán a los actores identificados anteriormente a interpretar y analizar datos de manera eficaz.

En el aspecto legal se han implementado leyes, decretos y estrategias que incentivan, motivan y apoyan a las entidades públicas a compartir los datos abiertos a través del portal web *www.datos.gov.co* (figura 3). La Ley 1712 de 2014 establece los procedimientos para garantizar la transparencia y el acceso a la información pública. El Decreto de Gobierno en Línea (Decreto 2573 de 2014) plantea las obligaciones para abrir, divulgar y promover la reutilización de los datos abiertos. El Decreto Único Reglamentario del Sector TIC - 1078 de 2015 suministra lineamientos para adoptar buenas prácticas en referencia a los estándares abiertos, con el fin de contribuir a la eficiencia y transparencia en el Estado colombiano.

Figura 3. Marco legal que apoya los datos abiertos en Colombia



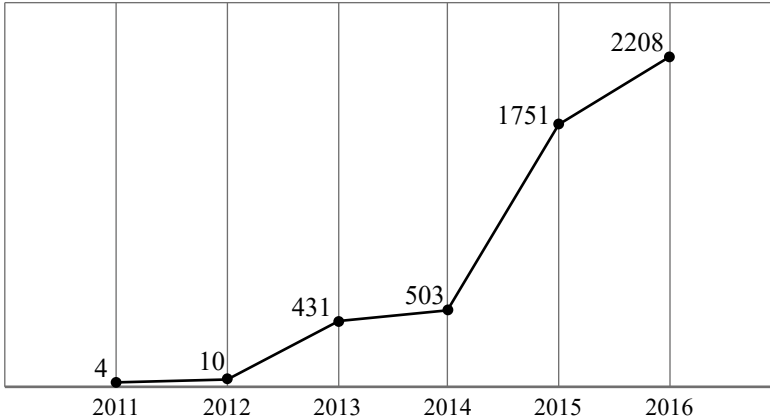
En 2015, el Banco Mundial realizó un diagnóstico al Proyecto de Datos a Abiertos en Colombia. El resultado de esta actividad arrojó recomendaciones para una adecuada implementación de una estrategia y política en esta materia. Entre las recomendaciones más destacadas se encuentran la definición de conjuntos de datos en formatos estándar que puedan ser interpretados por máquinas y usuarios finales, para de esta manera asegurar el intercambio de datos; por otro lado, la definición y el despliegue de un portal web público para compartir datos abiertos; finalmente, mantener los datos bajo una licencia abierta que permita dar cumplimiento a los principios de datos abiertos: uso, reutilización y redistribución.

De acuerdo con el Programa Gobierno en Línea (2016), los principios definidos bajo el marco del Proyecto de Datos Abiertos en Colombia son:

- a. *Primarios*. Los datos deben mantenerse en su forma y detalle original, es decir, no deben ser agregados ni transformados, con el fin de garantizar la exactitud.
- b. *Accesibles*. Los datos deben estar disponibles para los ciudadanos o las empresas que requieran usarlos, reutilizarlos y distribuirlos, independientemente de su propósito particular.
- c. *Completos*. Los datos deben estar en el mayor detalle posible, sin datos nulos, en clave de garantizar suficiencia y consistencia en el proceso de interpretación y análisis de información.
- d. *Procesables por máquinas*. Los formatos para publicar y compartir información deben ser estándar, con el fin de que máquinas y usuarios puedan intercambiar datos.
- e. *No propietarios*. Los datos publicados en el portal web no son exclusivos de ninguna persona o entidad pública.
- f. *Licenciados de forma abierta*. Los datos abiertos que se publican deben tener asociados términos y condiciones de utilización por parte de quienes van a consumirlos. Adicionalmente, los datos abiertos deben tener licenciamiento abierto.
- g. *No discriminados*. Los datos abiertos publicados en el portal web pueden ser consumidos o accedidos por cualquier usuario; no debe exigir ningún tipo de autenticación o registro para hacer uso de los datos.
- h. *Oportunos y actualizados*. Las entidades públicas deben mantener una frecuencia de actualización periódica, con el fin de garantizar que los datos se encuentren al día.

A continuación, en la figura 4, se presenta la evolución del número de conjuntos de datos entre 2011 y 2016.

Figura 4. Número de conjuntos de datos publicados en el portal www.datos.gov.co, periodo 2011-2016



Fuente: www.datos.gov.co (2016).

Datos abiertos en el mundo

Los países pioneros en la adopción de los principios y políticas de *open data* fueron Estados Unidos, Dinamarca, Noruega, Francia, Holanda y Gran Bretaña, que visualizaron las oportunidades de negocio para la misma sociedad, pues mediante la reutilización de datos podrían generarse nuevas innovaciones y cambios de paradigmas en los ciudadanos.

En América Latina, los países que ya han adoptado políticas y principios de *open data* son Colombia, Chile y México. Aunque existen otras naciones que están iniciando este proceso, en estos tres países ya existe una normatividad que permite y obliga a las entidades del Estado a compartir la información. En concordancia, las entidades gubernamentales están obligadas a publicar y compartir los datos con el fin de conseguir eficiencia y transparencia en sus procesos, mediante la participación de los ciudadanos y las empresas en el análisis y la interpretación de datos.

En general, los países que adoptan *open data* en sus políticas de gobierno coinciden en que esto les permitirá establecer una conversación constante entre el gobierno y la ciudadanía; tomar decisiones de administración pública basadas en las necesidades y preferencias de la comunidad; facilitar y promover la colaboración de la ciudadanía y las instituciones que comunican sus decisiones de

forma abierta; garantizar el acceso a la información. Este derecho es legitimado y defendido por el principio de transparencia (Mahecha, López y Velandia, 2017).

A pesar de que algunos países se encuentran más maduros que otros en los procesos inherentes a datos abiertos, aún se deben seguir atacando múltiples desafíos en función de asegurar el éxito en la implementación de políticas en esta materia. A continuación, de acuerdo con Charalabidis, Alexopoulos y Loukis (2016), se presenta la lista de desafíos que aún faltan por resolver en el contexto internacional:

- a. *Instrumentos de medición.* A través de las estrategias y los proyectos de datos abiertos se deben incorporar modelos de medición que permitan determinar indicadores tales como el grado de efectividad de la implementación de datos abiertos, el porcentaje de participación de los ciudadanos y empresas en la gestión de datos abiertos, el grado de innovación de proyectos a partir de la implementación de aplicaciones de *software* en la sociedad, etc.
- b. *Anonimizar datos abiertos.* Las entidades públicas deben velar y reservar conjuntos de datos que forman parte de la seguridad nacional y de tratados internacionales que solo les competen a ciertos actores en un país; sin embargo, una posible manera de publicar y compartir estos datos es a través de la anonimización, a fin de que nuevos servicios y aplicaciones se puedan desarrollar a partir de estos datos, salvaguardando los datos originales.
- c. *Limpieza de datos.* Los ciudadanos y las empresas interesados en desarrollar nuevos servicios o aplicaciones a partir de datos abiertos deben considerar métodos para la limpieza de datos, con el fin de extraer, transformar y cargar datos a otros repositorios, y de esta manera aplicar métodos de minería y visualización de datos. Por lo tanto, es crucial definir métodos estandarizados para realizar limpieza de datos abiertos.
- d. *Visualización de datos.* Los repositorios públicos de datos abiertos en el ámbito mundial están diseñados para almacenar, procesar y mostrar *big data*; ahora el objetivo es identificar, seleccionar y aplicar métodos adecuados para la fácil visualización, interpretación y análisis de datos por parte de cualquier ciudadano. De esta forma, las ciudades podrán

• *Open data y big data: herramientas de software para ciudades inteligentes.*

llamarse *ciudades inteligentes*, ya que la tecnología y las comunicaciones sirven como soporte para la toma de decisiones informadas.

- e. *Datos enlazados (linked data)*. El siguiente paso después de publicar y compartir la información es ofrecerles a los usuarios herramientas de *software* que permitan crear y mostrar datos relacionados entre diferentes conjuntos de datos, es decir, datos con una naturaleza diferente. Por ejemplo, crear relaciones entre un conjunto de datos financieros con un conjunto de datos de salud.
- f. *Proceso de publicación*. Si bien los procesos para publicar se encuentran definidos en documentos y guías para ejecutar correctamente las actividades asociadas, aún no se tiene automatizado un flujo que permita hacer más expedito el proceso de publicación; por consiguiente, la participación activa de flujos de procesos a través de herramientas de *software* simplificaría el proceso actual para la publicación de datos.
- g. *APIs y servicios*. La creación de estándares para el consumo de datos entre sistemas de información debe estandarizarse, por cuanto cada país es libre de implementar y utilizar librerías para exponer los datos; por eso, se complejiza la manera de extraer los datos de los portales que suministran datos abiertos.

Recientes estudios han demostrado que los datos abiertos han ayudado al desarrollo de servicios y aplicaciones en las siguientes áreas: innovación, analítica de datos, toma de decisiones, anticorrupción, ciudades inteligentes, entre otras. Ello trae como consecuencia la optimización de procesos y la generación de valor agregado en variables tales como economía, social y eficiencia de gobierno. Attard *et al.* (2015) presentan algunas aplicaciones que evidencian lo mencionado anteriormente:

- *Innovación*. La ganancia económica y el valor social no se encuentran directamente relacionados con esta área. La innovación está asociada a la manera en que un paradigma cambia disruptivamente a partir de un invento que implica el cambio en la manera de ejecutarse un proceso, normalmente con repercusión en la calidad de vida de las personas. *DontEat.at* es una aplicación que ayuda a clientes de restaurantes a validar la calidad del lugar que van a visitar (Global Open Data Index, s. f.).

- *Analítica de datos.* A través de herramientas de *software* que se enfocan en la aplicación de métodos para realizar análisis de *big data*, se han desarrollado nuevos procesos que involucran la optimización de actividades asociadas a medioambiente, transporte y educación. Un ejemplo es la aplicación *Street Bump*, que indica el estado de las calles en Estados Unidos.
- *Toma de decisiones y anticorrupción.* El alcalde de Seúl, Corea del Sur, tiene la posibilidad de integrar los datos abiertos de diferentes sectores que afectan a los ciudadanos: transporte, salud, precios de productos, entre otros. Toda la información se muestra a través de indicadores que le permiten a él tomar decisiones importantes. La información que se muestra en el *software* que el alcalde utiliza también está disponible para los ciudadanos, con lo cual se aplica el concepto de transparencia y como resultado se construye la confianza de los ciudadanos hacia el gobierno.
- *Ciudades inteligentes.* El indicador social enfocado a la seguridad también se ha visto beneficiado por la aplicación *Crime Finder*, que revela los delitos denunciados por las personas, y ayuda a la policía y a los ciudadanos a estar alerta, de acuerdo con la ubicación que se capture a través del celular (AppAdvice, s. f.).

Caso de estudio

El prototipo de *software* que se desarrolló para mejorar la calidad de vida de los ciudadanos, considerando los conceptos de *smart cities*, *open data* y *big data*, se encuentra publicado en internet con el fin de iniciar una democratización de la información generada por las entidades públicas del territorio colombiano. En específico, el caso de estudio se basa en el análisis de los datos suministrados por las entidades públicas al repositorio de datos www.datos.gov.co. El rango de datos que se tomó comprende las variables de los contratos celebrados en 2017, independientemente de su naturaleza. En la siguiente sección de análisis de los resultados se abordará el nivel de calidad que las entidades públicas comparten a la ciudadanía; la calidad se calculará a partir de las variables *trazabilidad*, *completitud* y *cumplimiento*.

•Open data y big data: herramientas de software para ciudades inteligentes.

A continuación se presentan los módulos que contiene la herramienta, para que de esta manera la organización o el ciudadano que requieran utilizarla puedan hacerlo con el conocimiento mínimo, de una manera eficaz y efectiva.

Módulos de la herramienta

Módulo 'Ingreso de estructura'

Este módulo tiene como objetivo almacenar los datos que se importan desde *www.datos.gov.co*. El módulo requiere dos variables para iniciar el flujo de importación de datos del repositorio en mención. La primera variable es llamada *ID del conjunto de datos*, que indica el bloque de datos que se va a descargar; por ejemplo, el conjunto de datos de los contratos celebrados en Bogotá. La segunda variable define el repositorio de datos del cual se quiere obtener los datos; para el caso de estudio únicamente se abordó el repositorio de *www.datos.gov.co*; no obstante, el prototipo puede descargar datos de cualquier otro repositorio (figura 5).

Figura 5. Módulo para ingresar estructuras de datos



The screenshot displays a web application interface. On the left, a dark sidebar menu is visible with the title 'Prototipo' and a gear icon. Under the 'GENERAL' section, there are four menu items: 'Ingreso datos', 'Ingreso estructuras' (which is highlighted), 'Calidad', and 'Madurez'. The main content area has a header 'Prototipo' and a sub-header 'Ingreso de estructura'. The main heading is 'Ingreso de estructura'. Below it, there is a prompt: 'Ingrese un ID para guardar su estructura en el prototipo para el análisis:'. This is followed by a text input field. Below that is another prompt: 'Inserte el dominio:', followed by another text input field. At the bottom left of the form area, there is a button labeled 'enviar'.

El valor de los dos campos mencionados son suministrados por el repositorio de datos abiertos *www.datos.gov.co*; sin embargo, el prototipo también es capaz de soportar la importación de datos de otros países, como Inglaterra y Estados Unidos. Por lo tanto, el prototipo valida la estructura de datos para cualquier repositorio que haga uso de los estándares mínimos para el intercambio de datos abiertos.

La figura 6 presenta la validación que realiza el *software* para garantizar que los datos importados se almacenen correctamente. De esta manera, se asegura que las estructuras de datos por importar funcionen correctamente de acuerdo con lo definido por el portal de datos en mención. Adicionalmente, el *software* realiza la validación de que la estructura que se va a crear no exista previamente, con el fin de evitar estructuras duplicadas (figura 7).

Figura 6. Validación exitosa del ingreso de datos

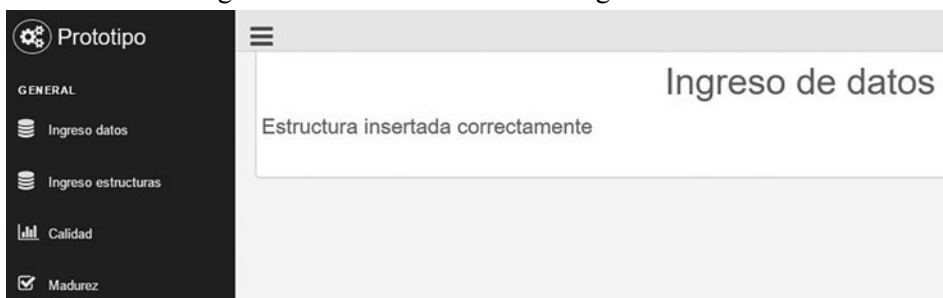


Figura 7. Validación del tipo de estructura utilizada



Módulo 'Ingreso de datos'

Este módulo tiene como función definir los conjuntos de datos por descargar, los cuales son identificados a partir de ID's definidos por el portal www.datos.gov.co (figura 8). Otra característica del *software* consiste en que el usuario final puede seleccionar varios conjuntos de datos. Un ejemplo de datos por importar es *r4jv-v36v*. Adicionalmente, se define el portal de datos abiertos por utilizar, lo cual se realiza a través del campo llamado "Ingrese el dominio", ya que el *software* tiene la capacidad de importar datos abiertos desde otros países (por ejemplo, el portal de datos abiertos de Estados Unidos). Una vez se envían, se define el conjunto de datos

•Open data y big data: herramientas de software para ciudades inteligentes.

y el repositorio sobre el cual se van a importar, el *software* corre un algoritmo que mide las siguientes métricas de calidad: completitud, trazabilidad y conformidad.

Figura 8. Módulo de ingreso de datos



La figura 9 muestra si el conjunto de datos se insertó correctamente. Esta operación se hace mediante la validación y el aseguramiento de que los datos se almacenen en la base de datos, con el fin de correr el proceso de calidad de datos. Por otra parte, y con el fin de garantizar que los datos no se dupliquen, el prototipo de *software* valida que los datos no se encuentren previamente guardados en la base de datos; esta es una manera de mantener la calidad de los datos (figura 10). Ahora bien, considerando las estructuras definidas por el portal *www.datos.gov.co* para el intercambio de datos abiertos, el prototipo de *software* se encarga de validar que la estructura de los datos se encuentre correctamente definida. Por ejemplo, si la estructura define que todos los datos tengan una columna ID, y la estructura por importar no la contiene, el *software* no ejecutará el proceso de importación del conjunto de datos (figura 11).

Figura 9. Validación exitosa de almacenamiento de datos

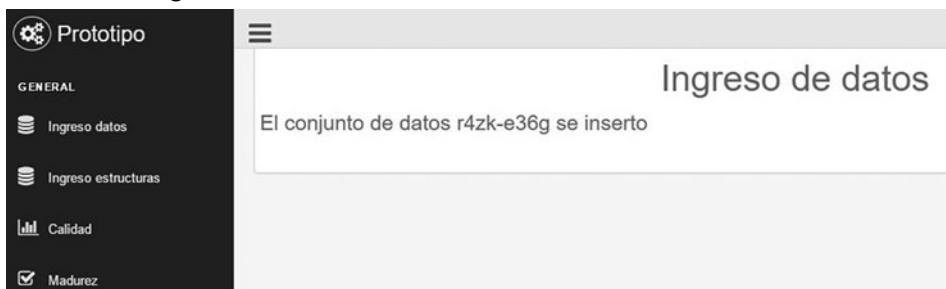


Figura 10. Mensaje de error que valida datos previamente guardados



Figura 11. Mensaje de error por estructura no válida



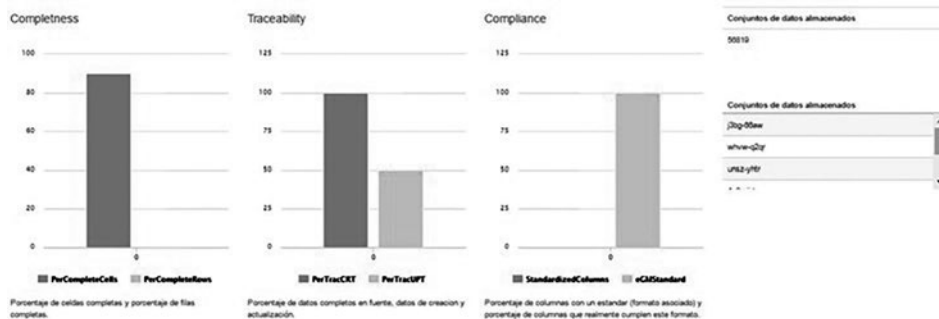
Módulo de reportes

El módulo de reportes fue construido con el fin de mostrar los resultados de la evaluación que el modelo propuesto ejecutó. Los resultados se muestran en tres categorías: completitud, trazabilidad y cumplimiento. Cada una de estas se definieron y explicaron en la sección de calidad de datos abiertos. La figura 12 es un resumen de los registros suministrados por *www.datos.gov.co* en lo referente a los contratos legalizados del 2017.

El módulo de reportes consta de dos partes: la presentación de resultados acumulada y la presentación de resultados para un conjunto de datos en particular. Los resultados acumulados son presentados a partir del análisis de todos los conjuntos de datos cargados al prototipo de *software*, independientemente de su naturaleza; es decir, el modelo de análisis podrá evaluar al mismo tiempo un conjunto de datos relacionado con educación y otro con contratación.

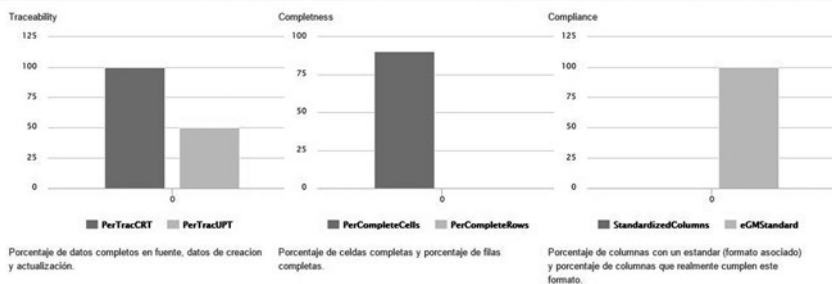
•Open data y big data: herramientas de software para ciudades inteligentes.

Figura 12. Módulo de resultados de calidad de datos acumulado



La figura 13 presenta un análisis de la calidad de datos para un conjunto de datos específico. Al igual que la anterior figura, existen las mismas categorías y se aplica el mismo análisis de datos; la única diferencia consiste en que el usuario final debe seleccionar el conjunto de datos por analizar.

Figura 13. Módulo de reporte de calidad de datos filtrado por conjunto de datos
Métricas por conjunto de datos



Análisis de resultados de la herramienta

Conjunto de datos acumulados

Con respecto a la métrica de *completitud*, el modelo planteado evalúa que celdas, filas y columnas se encuentren con algún dato, independientemente de su valor. La figura 13 evidencia que la muestra de conjuntos de datos seleccionada tiene un 90% de completitud en las celdas, mientras que existe un 5% de completitud en sus filas; es decir, las entidades públicas deben seguir trabajando en asegurar la amplitud y profundidad de los datos que compartan con la ciudadanía.

La métrica de *trazabilidad* evalúa si los metadatos (fuente de datos, frecuencia de actualización, periodo de datos actualizados, fecha de corte, entre otros) se encuentran definidos para cualquier conjunto de datos establecido durante el proceso de creación y actualización de dichos datos. Para el conjunto de datos evaluados, se puede concluir que el 100% de los conjuntos de datos contiene metadatos cuando se trata de la creación en *www.datos.gov.co*; sin embargo, cuando las entidades públicas actualizan datos, apenas un 50% contienen metadatos, dejando incompleta la información relacionada con los conjuntos de datos.

Con respecto al *cumplimiento*, el estándar define metadatos asociados a datos abiertos, como fuente, fecha de creación, categoría y título. Aunque existen metadatos opcionales, como descripción, publicación, cobertura, entre otros, estos no son considerados dentro de la evaluación. De acuerdo con la figura 13, el 99% del conjunto de datos analizados por la herramienta tiene metadatos definidos.

Conjunto de datos específico

De manera similar, el conjunto de datos (código de datos abiertos: *j3bg-66aw*) analizado corresponde a los contratos celebrados entre la Contraloría General de Antioquia y diferentes proveedores; para mayor detalle, la figura 14 muestra la ficha técnica del conjunto de datos. Puntualmente, 90% de los datos suministrados por la entidad se encuentran completos; el 100% de los datos creados contienen metadatos y el 48% de metadatos fueron definidos durante el proceso de actualización; por último, el 100% de los datos cumplen con el estándar de formato asociado a los datos.

Figura 14. Ficha técnica conjunto de datos específico

Información de la Entidad

Municipio	Medellín
Nombre de la Entidad	Contraloría General de la Antioquia
Orden	Territorial
Sector	Organismos de control y vigilancia
Área o dependencia	Contraloría General de Antioquia
Departamento	Antioquia

Información de Datos

Cobertura Geográfica	Departamental
Idioma	Español
Frecuencia de Actualización	Mensual
Fecha Emisión (aaaa-mm-dd)	2016-03-03

Temas

Categoría	Organismos de Control
Etiquetas	contratación, contratos, municipios, antioquia, contraloría, Mostrar más

Conclusiones

Big data es una tendencia que está permeando continentes, países, empresas y ciudadanos; aunque su origen es tecnológico, en este momento involucra cualquier área del conocimiento. Consiste en recopilar, procesar, almacenar y analizar información que apoye la toma de decisiones de una empresa o cualquier individuo que puedan gestionar grandes volúmenes de información. Aunque *big data* forma parte de la revolución digital, aún faltan muchos desafíos por resolver en todas sus aristas —por ejemplo, desafíos de calidad de datos e infraestructura tecnológica—. Las herramientas de *software* y *hardware* que actualmente se utilizan para gestionar *big data* permiten gestionar grandes volúmenes de datos, pero deben ser más robustas en términos de su rendimiento.

Open data es cualquier conjunto de datos que se genera a partir de la gestión propia de las entidades públicas en cualquier país. En el ámbito mundial ya existen países que implementaron una herramienta centralizada que permite consultar, reutilizar y distribuir los datos abiertos. El derecho y las regulaciones han contribuido a la utilización de herramientas tecnológicas que garanticen el acceso a la información pública.

Con respecto al caso de estudio, los datos proporcionados por Colombia Compra Eficiente y el análisis realizado por el prototipo de *software* permiten concluir que la calidad de la información es baja desde la perspectiva de la completitud, la conformidad y la trazabilidad, lo cual sesga el análisis y la identificación de problemas relacionados con la transparencia de los procesos de licitación. Los datos de contratación suministrados por la herramienta permiten concluir que en Colombia prevalece la contratación directa, a pesar de las diversas modalidades; según lo observado, en los datos de cada proceso hay un número reducido de participantes y una variedad de modalidades de contratación que agregan complejidad innecesaria.

Los desarrollos de políticas de gobierno para fortalecer los procesos de *open data* han ayudado a incluir a Colombia en el marco de ciudades inteligentes (*smart cities*); esto se evidencia en los datos que ya están compartiendo entidades del Estado, que permiten mejorar la calidad de vida de los ciudadanos a través de herramientas tecnológicas tales como los tableros de control.

CONCLUSIONES

La aparición y masificación de las tecnologías de la información y las comunicaciones (TIC), basadas en tecnología digital, ha generado nuevos retos para el Estado, la administración pública y la democracia, toda vez que este fenómeno posibilitó el acceso a la información en formas intensivas tanto en complejidad como en extensión. En este sentido, ha emergido la idea de una democracia digital basada en una cultura política cívica y participativa, comprometida con el control de los recursos públicos y la lucha contra la corrupción.

En el ámbito internacional, estas transformaciones han tenido un impulso importante en la gobernanza y la articulación del Estado con el ámbito global, llevando a modelos de desarrollo basados en la adquisición y el acceso a tecnologías digitales que posibilitan el acceso a la información. En el nivel local, la transparencia se ha convertido en el principio más relevante de la administración pública, pues ha generado un cambio cualitativo en la relación entre el Estado y los ciudadanos en favor de la consolidación de procesos democráticos de control sobre los datos más relevantes de la administración. Así, buena parte de los usos de aplicación de nuevas tecnologías como el *big data* se han llevado al campo de la administración pública con el propósito de orientar la toma de decisiones a partir de algoritmos complejos que pueden tener implicaciones también al derecho. Este proceso no puede separarse de los avances en la inteligencia artificial, ya que implica la posibilidad de crear documentos legislativos y sistemas de consultoría

jurídica que emplean la generación de respuestas y soluciones basadas en preferencias y sistemas automatizados.

El *big data* y la inteligencia artificial representan dos aplicaciones tecnológicas que ejemplifican las posibilidades del uso de la TIC en el derecho y el Estado. Las TIC ofrecen la posibilidad de aumentar la eficiencia en las decisiones judiciales y administrativas, de disminuir la corrupción a partir del control y la veeduría ciudadana y de alcanzar mayores umbrales de seguridad jurídica. De esta manera, la revolución tecnológica ha impactado profundamente en los derechos de la ciudadanía, a través de algoritmos que aumentan las garantías jurídicas de la sociedad.

No obstante, el *big data* puede afectar también la construcción del sujeto como un individuo apático, cuya frontera entre lo público y lo privado se hace tan débil que sistemas de dominación autoritaria no encuentran resistencia, toda vez que se ha configurado un espacio de destrucción de lo privado. En este contexto, así como *big data* puede representar una herramienta de control inconmensurable, así también la virtualidad amenaza la construcción de proyectos colectivos de creación y solidaridad. Por ello, se requiere un compromiso de las instituciones en el manejo responsable de la información y en el fomento de las habilidades sociales que posibilitan la consolidación de una ciudadanía crítica y comprometida con el avance del sistema político y social. La reconstrucción del sujeto debe pasar por la consideración de voces alternativas que van más allá de las nuevas tecnologías de información, en la reconstrucción de lazos sociales y proyectos de vida comprometidos con el bienestar común de la población.

El uso ético y respetuoso de los límites legales del *big data* y las demás TIC podrá generar la utilización responsable en aplicaciones como la ciberdefensa y la prevención de actuaciones contrarias a la seguridad nacional, tales como el terrorismo y el ciberterrorismo. El derecho se ha venido transformando para hacer frente al uso y la aplicación de la información que involucra datos personales; sin embargo, es necesario avanzar en la creación de regulaciones claras que generen garantías jurídicas y límites para el uso de la información a partir de la comprensión completa y sistemática de las transformaciones sociales y tecnológicas.

Las nuevas TCI en aplicaciones que abarcan desde la garantía de derechos individuales hasta el mantenimiento de la seguridad nacional deben pasar por el conocimiento profundo sobre las TIC, las posibilidades de aplicación y la calidad

•Conclusiones•

de la información. Resolver estos retos a través de acercamientos multidisciplinares que recogen desde la sociología y la ciencia política, hasta el derecho y la ingeniería de sistemas, les permitirá al sistema jurídico y al Estado consultar, reutilizar y distribuir datos abiertos, tal y como se viene haciendo en otras naciones.

BIBLIOGRAFÍA

- Acemoglu, D. y Robinson, J. (2014). *Por qué fracasan los países: los orígenes del poder, la prosperidad y la pobreza*. Barcelona: Deusto.
- AEPD e ISMS Forum (Eds.) (2017). *Código de buenas prácticas en protección de datos para proyectos de Big Data*. Madrid: Autor.
- Agencia Europea de Seguridad de las Redes y de la Información (ENISA) (2015). Privacy by design in Big Data: An overview of privacy enhancing technologies in the era of Big Data analytics. Recuperado de <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>
- Agudelo Giraldo, O. (2015). La enseñanza de textos jurídicos en sede indeterminista. En O. Agudelo Giraldo, J. E. León Molina y M. A. Prieto Salas (Eds.), *Teoría jurídica y enseñanza del derecho* (pp. 53-76). Bogotá: Universidad Católica de Colombia.
- Alarie, B., Niblett, A. y Yoon, A. (2016). *Law in the future*. Recuperado de <https://papers.ssrn.com/sol3/papers.cfm?abst>
- Allen, A. (2016). Protecting one's own privacy in a big data Economy. *Harvard Law Review Forum*, 130(71). Recuperado de <https://ssrn.com/abstract=2894545>
- Altman, N. S. (1992). An introduction to Kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185.
- Alva de la Selva, A. R. (2015). Los nuevos rostros de la desigualdad en el siglo XXI: la brecha digital. *Revista Mexicana de Ciencias Políticas y Sociales*, 55(223), 265-286.
- Anderson, P. (2002). *El Estado absolutista*. Ciudad de México: Siglo XXI.
- AppAdvice (s. f.). A crime Finder application form the maker so CrimeTime. Recuperado de <https://appadvice.com/app/california-crime-finder-iphone/447498678>

- Arend, C. (2012). *De qué forma Big Data transforma la protección y almacenamiento de datos*. Recuperado de <http://webdocs.commvault.com/assets/es/idc-big-data-transforms-data-protection-and-storage-analyst-report-es.pdf>
- Attard, J., Orlandi, f., Scerri, S. y Auer, A. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32, 399-418.
- Aucal Business School (2016). Ciberseguridad al día. Recuperado de <http://www.ciberseguridadparaempresas.com/ciberseguridad-en-la-union-europea/>
- Audea (2016). Diferencias entre ciberseguridad y seguridad de la información. Recuperado de <http://www.audea.com/diferencias-ciberseguridad-seguridad-la-informacion/>
- Ayestarán, I. (2011). Epistemología de la innovación social y de la destrucción creativa. *Utopía y Praxis Latinoamericana*, 54, 67-91.
- Azout, S. (2012). El crecimiento, la desigualdad y la pobreza. Recuperado de <http://www.portafolio.co/opinion/redaccion-portafolio/crecimiento-desigualdad-pobreza-108472>
- Balkin, J. M. (2017). The three laws of robotics in the age of Big Data. *Ohio State Law Journal*, 78(592). Recuperado de <https://ssrn.com/abstract=2890965>
- Banco Interamericano de Desarrollo (BID) y Organización de Estados Americanos (OEA) (2016). *Ciberseguridad. ¿Estamos preparados en América Latina y el Caribe? Informe Ciberseguridad 2016*. Recuperado de <https://publications.iadb.org/handle/11319/7449?locale-attribute=es>
- Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge: Cambridge University Press.
- Barocas, S. y Selbst, A. (2016). Big data's disparate impact. *104 California Law Review*, 671. Recuperado de <https://ssrn.com/abstract=2477899>
- Becerra, J. (2015). Revisión del marco teórico para establecer responsabilidades sobre el estado en sus relaciones con los proveedores de servicios de telecomunicaciones. En J. Becerra (Eds.), *El derecho y las tecnologías de la información y la comunicación (TIC)* (pp. 91-99). Bogotá: Universidad Católica de Colombia.
- Becerra, J., Torres Ávila, J., Cotino Hueso, L., García Vargas, C. B. y Sánchez Acevedo, M. E. (2015). *La responsabilidad del Estado por la utilización de las tecnologías de la información y la comunicación (TIC)*. Bogotá: Universidad Católica de Colombia.
- Bell, D. (1984). *Las ciencias sociales desde la Segunda Guerra Mundial*. Madrid: Alianza.
- Bell, D. (2001). *El advenimiento de la sociedad post-industrial: un intento de prognosis social*. Madrid: Alianza.
- Bericat Alastuey, E. (1996). La sociedad de la información. Tecnología, cultura, sociedad. *REIS, Revista Española de Investigaciones Sociológicas*, 76, 99-122.

•Bibliografía•

- Bertot, J. C. y Choi, H. (2013). Big data and e-government. *Proceedings of the 14th Annual International Conference on Digital Government Research*. Recuperado de <http://doi.org/10.1145/2479724.2479730>
- Blumler, J. G. y Gurevitch, M. (2001). The new media and our political communication discontents. Democratizing cyberspace. *Information, Communication and Society*, 4(1), 1-13.
- Bouskela, M., Casseb, M., Bassi, S. y Facchina, C. D. (2016). *La ruta hacia las smart cities: migrando de una gestión tradicional a la ciudad inteligente*. Washington: Banco Interamericano de Desarrollo (BID).
- Boyd, D. y Crawford, K. (2011). *Six provocations for Big Data*. Recuperado de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431
- Breiman, L. (1996). *Bias, variance, and arcing classifiers*. Berkeley: Universidad de California.
- Brodely, C. y Friedl, M. (1999). Identifying and eliminating mislabeled training instances. *Journal of Artificial Intelligence Research*, 11, 131-167.
- Brownlee, J. (2014, 26 de septiembre). Discover feature engineering, how to engineer features and how to get good at it. Recuperado de <http://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/>
- Burkholder, L. (Ed.) (1992). *Philosophy and the computer*. Boulder: Westview Press.
- Camacho, E. F. y Bordons, C. (2004). *Model predictive control* (2.^a ed.). Londres: Springer-Verlag.
- Camargo Vega, J. J., Camargo Ortega, J. F. y Joyanes Aguilar, L. (2014). Conociendo el Big Data. *Revista Facultad de Ingeniería*, 24(38), 63-77.
- Campillo, A. (2009). *El concepto de lo político en la sociedad global*. Barcelona: Herder.
- Carreño Dueñas, D. (2016). *Pensar el derecho como derecho virtual*. Bogotá: Universidad Católica de Colombia.
- Carrillo, J. et al. (2013). Big data en los entornos de defensa y seguridad nacional. Recuperado de http://www.ieee.es/Galerias/fichero/docs_investig/DIEEEINV03-2013_Big_Data_Entornos_DefensaSeguridad_CarrilloRuiz.pdf
- Castells, M. (1996). *La era de la información. Economía, sociedad y cultura*. Ciudad de México: Siglo XXI.
- Castells, M. (2006). *La sociedad red: una visión global*. Madrid: Alianza.
- Castresana Sáenz, C. (2016). *Industria 4.0*. (tesis de pregrado). La Rioja: Universidad de La Rioja.
- Charalabidis, Y, Alexopoulos, C. y Loukis, E. (2016). A taxonomy of open government data research areas and topics. Recuperado de <https://www.tandfonline.com/doi/abs/10.1080/10919392.2015.1124720>
- Clark, R. y Knake, R. (2011). *Guerra en la red. Los nuevos campos de batalla*. Madrid: Ariel.
- Coglianese, C. y Lehr, D. (2017). Regulating by robot: Administrative decision making in the machine-learning era. *Georgetown Law Journal*, 105.

- COMeIN (2014). Big Data se escribe con V. *Revista de los Estudios de Ciencias de la Información y de la Comunicación*, 37. Recuperado de <http://comein.uoc.edu/divulgacio/comein/es/numero37/articles/Article-Eva-Ortoll.html>
- Comisión Europea (2007). Programa Europeo para la Protección de Infraestructuras Críticas. Recuperado de <http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=URISERV%3A133260>
- Congreso de la República de Colombia (2010). Proyecto de Ley 184 de 2010, por la cual se dictan disposiciones generales para la protección de datos personales. Recuperado de https://www.huntonprivacyblog.com/wp-content/uploads/sites/18/2011/10/Colombia_Data_Protection_Law.pdf
- Congreso de la República de Colombia (2012, 17 de octubre). Ley Estatutaria 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales. *Diario Oficial* 48.587.
- Consejo Argentino para las Relaciones Internacionales (CARI) (2013). Ciberdefensa-ciberseguridad: riesgos y amenazas. Recuperado de http://www.cari.org.ar/pdf/ciberdefensa_riesgos_amenazas.pdf
- Corte Constitucional de Colombia (2011, 6 de octubre). Sentencia C-748/11. M.P. Jorge Ignacio Pretelt Chaljub.
- Corte Constitucional de Colombia (2015, 12 de mayo). Sentencia T-277/2015. M. P. María Victoria Calle Correa.
- Corte Constitucional de Colombia (2017, 23 de enero). Sentencia T-022/2017. M. P. Luis Guillermo Guerrero Pérez.
- Cortes, C. y Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Cortés, F. (2016). Discusiones sobre desigualdad y clases sociales en América Latina en los albores del siglo XXI. En L. Ortiz (Ed.), *Desigualdad y clases sociales* (pp. 23-60). Buenos Aires: CLACSO.
- Cotino Hueso, L. (2013). La selección y personalización de noticias por el usuario de nuevas tecnologías. En L. Corredoira y L. Cotino (Eds.), *Libertad de expresión e información en internet. Amenazas y protección de los derechos personales* (pp. 41-56). Madrid: Centro de Estudios Políticos y Constitucionales.
- Cotino Hueso, L. (2015). Algunas cuestiones clave de protección de datos en la nube. Hacia una “regulación nebulosa”. *Revista Catalana de Derecho Público*, 51, 85-103.
- Crawford, K. y Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review*, 55(93), 13-36.
- Crossman, R. (2004). *Biografía del Estado moderno*. Ciudad de México: Fondo de Cultura Económica.
- De Cabo Martín, C. (2012). *Dialéctica del sujeto, dialéctica de la Constitución*. Madrid: Trotta.
- De Pablos, E. (2016). Revolución Industrial 4.0. *CyC Prisma*, 21, 10-15.

•Bibliografía•

- De Tullio, M. F. (2016). La privacy e i big data verso una dimensione costituzionale collettiva. *Politica del diritto*, 47(4), 637-696.
- DemosEUROPA (2014). Big and open data in Europe. A growth engine or a missed opportunity? Recuperado de <https://news.microsoft.com/europe/2014/01/29/big-and-open-data-in-europe-a-growth-engine-or-a-missed-opportunity/>
- Departamento Nacional de Planeación (DNP) (2011, 14 de julio). *CONPES 3701: Lineamientos de política para ciberseguridad y ciberdefensa*. Bogotá: Autor.
- Díaz Granados, G. F. (2016). La comercialización del big data. *Universitas Estudiantes*, 14, 111-128.
- Drury, S. D. y Wagoner, N. J. (2014). Bargaining in the Shadow of Big Data. *Florida Law Review*, 66(5).
- Dumais, S. T. (2005). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38, 188-230.
- Echeverría, J. (2005). La revolución tecnocientífica. *Confines*, 1(2), 9-15.
- Edwards, L., McAuley, D. y Diver, L. (2016). From privacy impact assessment to social impact assessment. Recuperado de <http://eprints.nottingham.ac.uk/45237/8/20160315SIA-letter.pdf>
- El Derecho de Autor en la Era Digital (s. f.). Propiedad intelectual en la legislación colombiana [blog]. Recuperado de http://www.iered.org/miembros/ulises/representacion-ideas/Derechos-Autor/propiedad_intelectual_en_la_legislacion_colombiana.html
- Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. En M. R. Endsley y D. J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah: Lawrence Erlbaum Ass.
- Estévez Araújo, J. (2009). Gobernanza por democracia en la Unión Europea. *Novum Jus: Revista Especializada en Sociología Jurídica y Política*, 3(2), 213-236.
- Estévez, E. y Janowski, T. (2016). Gobierno digital, ciudadanos y ciudades inteligentes. *Revista Institucional de la Facultad de Informática de la UNLP*, 3, 11-13.
- Fahad, M. y Sikander, H. (2007). Classification of textual documents using learning vector quantization. *Information Technology Journal*, 6(1), 154-159.
- Fernández Gómez, L. (2014). ¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial Big Data? Recuperado de http://centrodeinnovacion.gobiernoonline.gov.co/sites/default/files/investigacion_big_data.pdf
- Flórez Acero, G. (2016). Algunas nociones básicas sobre evidencia digital en el contexto del comercio electrónico, desde la perspectiva anglosajona y su visión en Colombia. En G. Flórez Acero, D. Montenegro Reyes y D. Bernal Sánchez (Eds.), *Evidencia digital, distribución musical y derecho de consumo* (pp. 7-27). Bogotá: Universidad Católica de Colombia.

- Freedman, D. A. (2009). *Statistical models: Theory and practice*. Cambridge: Cambridge University Press.
- Freedom House (2014). *Informe de Freedom House, 2014*. Recuperado de <https://freedomhouse.org/publicaciones>
- Freedom House (2017, 14 de noviembre). Libertad en la Red 2017: manipulación de redes sociales para socavar la democracia. Recuperado de <https://freedomhouse.org/article/libertad-en-la-red-2017-manipulaci-n-de-redes-sociales-para-socavar-la-democracia>
- Fuerzas Militares de Colombia - Ejército Nacional (2015). Procedimiento de comunicaciones operacionales y ciberdefensa. Recuperado de <https://www.ejercito.mil.co/?idcategoria=357574&download=Y>
- Gadamer, H.-G. (1996). *Verdad y método I. Fundamentos de una hermenéutica filosófica*. Salamanca: Sígueme.
- Gartner (s. f.). Big Data. Recuperado de <http://www.gartner.com/it-glossary/big-data/>
- Geman, S., Bienenstock, E. y Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4, 1-58.
- Gimmler, A. (2001). Deliberative democracy, the public sphere and the internet. *Philosophy and Social Criticism*, 27(4), 21-39.
- Global Open Data Index (s. f.). Tracking the state of Open Government Data. Recuperado de <https://index.okfn.org>
- Gómez, P. A. y Pedreros Monrroy, V. (2017). Direito e questões tecnológicas, aplicados no desenvolvimento social. *Novum Jus*, 11(1), 175-177.
- González Garzón, H. (2015). Tránsito de derechos humanos a raíz del tráfico ilegal de coltán en el departamento del Guainía. Recuperado de <https://scinapse.io/papers/2417954039>
- Greenpeace. (2009). *Cambio climático: futuro negro para los páramos*. Bogotá: Greenpeace Colombia.
- Haenlein, M. y Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics*, 3(4), 283-297.
- Han, B.-C. (2014). *Psicopolítica: neoliberalismo y nuevas técnicas de poder*. Barcelona: Herder.
- Hartzog, W. y Stutzman, F. (2013). Obscurity by design. *Washington Law Review*, 88. Recuperado de <https://ssrn.com/abstract=2284583>
- Heeger, E. (2015). Controlling your online profile: reality or an illusion? A research into informed consent as a mechanism to regulate commercial profiling. Recuperado de <http://dx.doi.org/10.2139/ssrn.2658651>
- Ho, T. K. (1995). Random decision forests. Recuperado de <https://dl.acm.org/citation.cfm?id=844681>

•Bibliografía•

- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832-844.
- Hopkins, B. (2011, 28 de octubre). Beyond the hype of Big Data. Recuperado de http://www.cio.com/article/692724/Beyond_the_Hype_of_Big_Data
- Infraestructura Colombiana de Datos Espaciales (ICDE) (2017, 24 de abril). Avances de big data en el sector público colombiano. Recuperado de <http://www.icde.org.co/noticias/Avances-De-Big-Data-En-El-Sector-Publico-Colombiano>
- Institute of Actuaries (2015). *Big data: An actuarial perspective*. Londres: Autor.
- Internacional Telecommunication Union (2007). *Guía de ciberseguridad para los países en desarrollo*. Recuperado de <http://www.itu.int/ITU-D/cyb/cybersecurity/docs/ITUNationalCybersecurityStrategyGuide.pdf>
- James, G. (2003). Variance and bias for general loss functions. *Machine Learning*, 51, 115-135.
- Jefatura del Estado Español (2011). Ley 8/2011 de 28 de abril, por la que se establecen medidas para la protección de las infraestructuras críticas. *BOE 102*, 29 de abril de 2011.
- Joh, E. E. (2014). Policing by numbers: Big Data and the Fourth Amendment. Recuperado de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2403028
- Joh, E. E. (2016). Policing police robots. Recuperado de <https://ssrn.com/abstract=2817185>
- Joh, E. E. (2017). The undue influence of surveillance technology companies on policing. Recuperado de <https://ssrn.com/abstract=2924620>
- Jolliffe, I. (2002). *Principal component analysis* (2.ª ed). Nueva York: Springer.
- Kacfeh Emani, C., Cullot, N. y Nicolle, C. (2015). Understandable Big Data: A survey. *Computer Science Review*, 17, 70-81.
- Kanter, J. M. y Veeramachaneni, K. (2015). Deep feature synthesis: Towards automating data science endeavors. *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. París: IEEE.
- Kellner, D. (2004). The media and the crises of democracy in the age of Bush-2. *Communication and Critical Cultural Studies*, 1(1), 29-58.
- Khan, S., Liu, X., Shakil, K. y Alam, M. (2017). A survey on scholarly data: From big data perspective. *Information Processing & Management*, 53(4), 923-944.
- Kiljunen, K. (1986). La división internacional del trabajo industrial y el concepto de centro-periferia. *Revista de la Cepal*, 30, 103-124.
- Ki-moon, B. (2009). *Observaciones del Consejo de Seguridad sobre Timor-Leste*. Nueva York: Organización de las Naciones Unidas.
- Klous, S. y Nart, W. (2016). *We are Big Data: The future of the information society*. Ámsterdam: Atlantis Press.

- Kohonen, T. (1997). *Self-organizing maps*. Berlín: Springer.
- Kosutic, D. (2012). *Ciberseguridad en 9 pasos. Manual sobre seguridad de la información para el gerente*. Recuperado de https://advisera.com/wp-content/uploads/sites/9/2016/09/Ciberseguridad_en_9_pasos_ES.pdf
- Krotoszynski, J. y Ronald, J. (2015). Reconciling privacy and speech in the era of Big Data: A comparative legal analysis. *William & Mary Law Review*, 56(1279).
- Latour, B. (2009). Tarde's idea of quantification. En M. Candea (Ed.), *The social after Gabriel Tarde: Debates and assessments* (pp. 145-162). Londres: Routledge.
- Lee, J. A. y Verleysen, M. (2007). *Nonlinear dimensionality reduction*. Berlín: Springer.
- Lerman, J. (2013). Big Data and its exclusions. *Stanford Law Review Online*, 66.
- Liou, C., Cheng, C., Liou, J. y Liou, D. (2014). Autoencoder for words. *Neurocomputing*, 139, 84-96.
- Machin, N. y Gazapo, M. (2016). La ciberseguridad como factor crítico en la seguridad de la Unión Europea. *INISCI Journal*, 42. Recuperado de <https://www.ucm.es/data/cont/media/www/pag.../UNISCIDP42->
- Mahecha, J., López, N. y Velandia, J. (2017). Assessing data quality in Open Data : A case study. Recuperado de <https://doi.org/10.1109/CONIITI.2017.8273343>
- Manovich, L. (2012). Trending: The promises and the challenges of big social data. En M. Gold (Ed.), *Debates in the digital humanities*. Minnesota: The University of Minnesota Press.
- Martín Miralles, R. (2013). Big Data vs. Small low. Recuperado de <http://dialnet.unirioja.es/servlet/extart?codigo=4329765>
- Martin, E. G y Begany, G. M. (2016). Opening government health data to the public: Benefits, challenges, and lessons learned from early innovators. *Journal of the American Medical Informatics Association*, 24(2), 345-351.
- Martínez, A. y Kak, A. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228-233.
- Martínez, R. (2014). *Ética y privacidad de los datos*. Recuperado de http://sgfm.elcorteingles.es/SGFM/FRA/recursos/conferencias/ppt/1776180509_1472014102438.docx
- Massaro, T. M., Norton, H. L. y Kaminski, M. E. (2017). Siri-ously 2.0: What artificial intelligence reveals about the first amendment. *Ohio State Public Law Working Paper*, 374. Recuperado de <https://ssrn.com/abstract=2896174>
- Massaro, T. y Norton, H. L. y Kaminski, M. E. (2016). Siri-ously? Free speech rights and artificial intelligence. *110 Northwestern University Law Review*, 1169. Recuperado de <https://ssrn.com/abstract=2643043>

•Bibliografía•

- Mayer-Schönberger, V. y Cukier, K. (2013a). *Big Data: A revolution that will transform how we live, work, and think*. Recuperado de http://puntocritico.com/ausajpuntocritico/documentos/Big_Data.pdf
- Mayer-Schönberger, V. y Cukier, K. (2013b). *The dictatorship of data*. Recuperado de <https://www.technologyreview.com/s/514591/the-dictatorship-of-data>
- Mayne, D. Q., Rawlings, J. B., Rao, C. V. y Scokaert, P. (2000.) Constrained model predictive control: Stability and optimality. *Automatica*, 36(6), 789-814.
- McLachlan, G. J. (2004). *Discriminant analysis and statistical pattern recognition*. Nueva Jersey: Wiley Interscience.
- Ministerio de Defensa e Instituto Español de Estudios Estratégicos (2012). Ciberespacio: nuevo escenario de confrontación. Recuperado de <https://dialnet.unirioja.es/servlet/libro?codigo=547632>
- Ministerio de Tecnologías de la Información y las Comunicaciones (2015). *Estudio sobre el estado de apropiación de la seguridad de la información en entidades del Estado*. Bogotá: MinTIC.
- Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia (2015). *Panorama TIC. comportamiento macroeconómico del sector TIC en Colombia*. Recuperado de http://colombiatic.mintic.gov.co/602/articles-14305_panoranatic.pdf
- Ministerio TIC de Colombia (s. f.). Big data. Recuperado de <http://www.mintic.gov.co/portal/604/w3-article-6163.html>
- Minsky, M. (1988). *The society of mind*. Nueva York: Simon & Schust.
- Mohri, M., Rostamizadeh, A. y Talwalkar, A. (2012). *Foundations of machine learning*. Massachusetts: The MIT Press.
- Narayanan, A. y Zevenbergen, B. (2015). No encore for encore? Ethical questions for web-based censorship measurement. Recuperado de <http://dx.doi.org/10.2139/ssrn.2665148>
- National Standard Institute of Technology (NIST) (2017). *Framework for Improving Critical Infrastructure Cybersecurity Draft Version 1.1*. Gaithersburg: NIST.
- Neiva Santos, R. (2009). *Petrobras en la política exterior del gobierno de Lula: una mirada desde la Economía Política internacional* (tesis de maestría). Buenos Aires: Universidad de San Andrés.
- Nils, N. (1980). *Principles of artificial intelligence*. Palo Alto: Tioga Press.
- O'Neil, C. (2016). *Weapons of math destruction: How Big Data increases inequality and threatens democracy*. Nueva York: Crown.
- Observatorio de la Ciberseguridad en América Latina y el Caribe (2016). Ciberseguridad: estamos preparados en América Latina y el Caribe. Recuperado de <https://www.sites.oas.org/cyber/ES/Paginas/Documents.aspx>

- Oliver, D. y Muñoz, J. F. (2014). El mito del consentimiento, o por qué un sistema individualista de protección de datos (ya) no sirve para (casi) nada. En J. Valero Torrijos (Ed.), *La protección de los datos personales en Internet ante la innovación tecnológica*. Aranzadi: Cizur Menor.
- Organización de los Estados Americanos (OEA) (2013). Tendencias en la seguridad cibernética en América Latina y el Caribe y respuestas de los Gobiernos. Recuperado de https://www.symantec.com/content/es/mx/enterprise/other_resources/b-cyber-security-trends-report-lamc.pdf
- Organización de los Estados Americanos (OEA) (2017). Iniciativa de seguridad cibernética de la OEA. Recuperado de <https://www.sites.oas.org/cyber/ES/Paginas/Documents.aspx>
- Ostry, J., Berg, A. y Tsangarides, C. (2014). Redistribución, desigualdad y crecimiento. *Revista de Economía Institucional*, 16(30), 53-81.
- Ótalora, K. y Hernández, J. (2018). Arte y derecho: de la palabra del *pütchipü'üi* a la imagen. Resolución alternativa de conflictos. *Novum Jus*, 12(1), 221-244.
- Oussous, A., Benjelloun, F. Z., Ait Lahcen, A. y Belfkih, S. (2017). Big Data technologies: A survey. Recuperado de <http://doi.org/10.1016/j.jksuci.2017.06.001>
- Palacios, P., Delgado, E., León, E., Motaño, J. y Estupiñan A (2014). *Sistemas de Información Bibliográficos como estrategia para la disposición y acceso al conocimiento para la Salud en Colombia*. Recuperado de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/IA/SSA/Articulo%2011.pdf>
- Paredes-Moreno, A. (2015). Big Data: Estado de la cuestión. *International Journal of Information Systems and Software Engineering for Big Companies*, 2(1), 38-59.
- Parlamento Europeo (2017). Resolución de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)).
- Pasquale, F. A., (2015). Privacy, autonomy, and internet platforms. En M. Rotenberg, J. Horwitz y J. Scott (Ed.), *Privacy in the Modern Age: The Search for Solutions*. Maryland: Universidad d Maryland.
- Pérez, C. (2010). Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, 34(1), 185-202.
- Policía Nacional de Colombia (2014). Big data en la Policía Nacional de Colombia. Recuperado de <http://www.datacenterdynamics.es/video/big-data-en-la-polic%C3%ADa-nacional-de-colombia>
- Porras, J. D. (2001). *La lucha por la Constitución: las teorías del Fundamental Law en la Inglaterra del siglo XVII*. Madrid: Centro de Estudios Políticos y Constitucionales.
- Power Data (2013). El valor de la gestión de datos. ¿Cómo se relacionan Big Data y Hadoop? Recuperado de <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/328879/c-mo-se-relacionan-big-data-y-hadoop>

•Bibliografía•

- Press, W. H., Teukolsky, S. A., Vetterling, W. T. y Flannery, B. P. (2007). *Numerical recipes: The Art of Scientific Computing* (3.ª ed.). Nueva York: Cambridge University Press.
- Programa Gobierno en Línea (2011). *Evolución del Gobierno en línea en Colombia*. Bogotá: Ministerio de Tecnologías de la Información y las Comunicaciones.
- Puyol Moreno, J. (2014). Una aproximación a Big Data. *Revista de Derecho UNED*, 14, 471-505.
- Quintana, F. (2006). Ciudad, metrópolis y mundo global/local. *Athenea Digital, Revista de Pensamiento e Investigación Social*, 6, 66-79.
- Ragnedda, M. (2011). Internet y control social. Entre rizoma y gran hermano. *Perspectivas de la Comunicación*, 4(1), 42-52.
- Raschka, S. (2014). About feature scaling and normalization. Recuperado de http://sebastianraschka.com/Articles/2014_about_feature_scaling.html
- Recuperación de Información (2011). Los documentos estructurados, semiestructurado y no estructurados [blog]. Recuperado de <http://sisinfo-sri.blogspot.com/2011/10/los-documentos-estructurados.html>
- Rencher, A. y Christensen, W. (2012). *Methods of multivariate analysis* (3.ª ed.). Nueva York: John Wiley & Sons.
- Resnik, M. (2004). *Paradigmas en ciencia política*. Buenos Aires: Eudeba.
- Reyes Beltrán, P. (2017). *Derecho y globalización. Transformaciones del Estado contemporáneo*. Bogotá: Universidad Nacional de Colombia.
- Richards, N. M. y King, J. H. (2013). Three paradoxes of Big Data. *66 Stanford Law Review Online*, 41.
- Rifkin, J. (2011). *La Tercera Revolución Industrial: cómo el poder lateral está transformando la energía, la economía y el mundo*. Madrid: Paidós.
- Rincon, O. (2016). Los datos: la cancha donde se juega la democracia. *Chasqui, Revista Latinoamericana de Comunicación*, 131, 21-35.
- Rish, I. (2001). An empirical study of the Naïve Bayes classifier. Recuperado de https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier
- Rivera Méndez, R. G. (2010). *Gobernanza democrática. Concepto y perspectivas*. La Paz: Unidad de Gobernabilidad y Gobernanza.
- Rodríguez, C., Chacón, N. y Cubides, J. (2017). Reclutamiento ilícito de niños y niñas en Colombia: marcos de protección en el derecho internacional de los derechos humanos. En J. Cubides (Ed.), *Desafíos contemporáneos de los derechos humanos* (pp. 31-55). Bogotá: Universidad Católica de Colombia.
- Roel Pineda, V. (1998). *La tercera revolución industrial y la era del conocimiento*. Lima: Universidad Nacional Mayor de San Marcos.

- Rokach, L. y Maimon, O. (2008). *Data mining with decision trees: Theory and applications*. Singapur: World Scientific.
- Rubinstein, I. (2013). Big Data: The end of privacy or a new beginning? *International Data Privacy Law*, 3(2).
- Russell, S. y Norvig, P. (2007). *Artificial intelligence: A modern approach* (3.ª ed). Londres: Prentice Hall.
- Salter, S. y Thompson, D. (2017). Public-Centred Civil Justice Redesign: A Case Study of the British Columbia Civil Resolution Tribunal. *McGill Journal of Dispute Resolution*, 3, 2016-2017.
- Salvador, F. (2014). Big Data: ¿la ruta o el destino? *Tecnología y Crecimiento*, 3. Recuperado de https://www.ie.edu/fundacion_ie/Comun/Publicaciones/Publicaciones/Big%20Data%20ESP%207.pdf
- Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(401-402), 403-409.
- Sánchez Acevedo, M. E. (2014). *Eficacia y validez del acto administrativo electrónico*. Bogotá: Universidad Católica de Colombia.
- Sánchez Barrilao, J. F. (2016). El derecho constitucional ante la era de Ultrón: la informática y la inteligencia artificial como objeto constitucional. *Estudios de Deusto*, 64(2), 225-258.
- Santos, B. y Gandarilla Salgado, J. (2009). *Una epistemología del sur: la reinención del conocimiento y la emancipación social*. Ciudad de México: CLACSO y Siglo XXI.
- Scalabre, O. (2018). Embracing Industry 4.0—and Rediscovering Growth. Recuperado de <https://www.bcg.com/capabilities/operations/embracing-industry-4.0-rediscovering-growth.aspx>
- Schoijet, M. (1998). La revolución científica y tecnológica y la sociedad postindustrial. *Revista Mexicana de Ciencias Políticas y Sociales, Nueva Época*, 43(171), 127-154.
- Schölkopf, B., Smola, A. y Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), 1299-1319.
- Schroeck, M. et al. (2012). *Analytics: el uso de big data en el mundo real*. Oxford: Escuela de Negocios Saïd. Recuperado de http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf
- Schumpeter, J. (1961). *The theory of economic development*. Oxford: Oxford University Press.
- Schumpeter, J. (1983). *Capitalismo, socialismo y democracia*. Barcelona: Ediciones Orbis.
- Secades, V. A. (2015). Big Data: la eclosión de los datos. *Cuadernos Salmantinos de Filosofía*, 42, 315-330.
- Selbst, A. D. (2017). Disparate impact in Big Data policing. Recuperado de <https://ssrn.com/abstract=2819182>

•Bibliografía•

- Sivarajah, U., Kamal, M. M., Irani, Z. y Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
- SMART-Comisión Europea (2013). 2013/0063 - Study on a “European data market”. Recuperado de <http://ec.europa.eu/digital-agenda/en/news/smart-20130063-study-european-data-market-and-related-services>
- Smith, M. y Martínez, T. (2011). Improving classification accuracy by identifying and removing instances that should be misclassified. Recuperado de <http://axon.cs.byu.edu/papers/smith.ijcnn2011.pdf>
- Soares, S. (2012). *Big Data Governance: An emerging imperative*. Boise: MC Press.
- Solove, D. J. (2014). Facebook’s psych experiment: Consent, privacy, and manipulation. En M. Candeia (Ed.), *Debates and assessments* (pp. 145-162). Londres: Routledge.
- Stalla-Bourdillon, S. y Knight, A. (2017). *Anonymous data v. personal data - A false debate: An EU perspective on anonymization, pseudonymization and personal data*. Recuperado de <https://ssrn.com/abstract=2927945>
- Stone, J. V. (2004). *Independent component analysis: A tutorial introduction*. Cambridge: MIT Press.
- Stucke, M. y Grunes, A. (2016). *Big Data and competition policy*. Nueva York: Oxford University Press.
- Suakanto, S., Supangkat, S. H., Suhardi, A. y Saragih, R. (2013). Smart city dashboard for integrating various data of sensor networks. *International Conference on ICT for Smart Society*. Jakarta: IEEE.
- Sunstein, C. (2001). *Republic.com*. Princeton: Princeton University Press.
- Surden, Harry (2017). Values embedded in legal artificial intelligence. Recuperado de <https://ssrn.com/abstract=2932333>
- Suthaharan, S. (2016). Supervised learning algorithms. En *Machine learning models and algorithms for Big Data classification* (pp. 183-206). Nueva York: Springer.
- Tene, O. y Polonetsky, J. (2013). Judged by the Tin Man: Individual rights in the age of Big Data. Recuperado de <https://pdfs.semanticscholar.org/0b8e/847384e6ec615347a5806d80fab75784b342.pdf>
- Tenenbaum, J., de Silva, V. y Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2319-2323.
- Terán, T. M. (2007). *Filosofía y política en Michel Foucault*. Ciudad de México: Plaza y Valdés.
- Thatcher, J. (2014). Living on fumes: Digital footprints, data fumes, and the limitations of spatial Big Data. *International Journal of Communication* 8(19). Recuperado de <http://ijoc.org/index.php/ijoc/article/view/2174>

- Thomasen, K. (2016). Examining the constitutionality of robot-enhanced interrogation. Recuperado de <https://doi.org/10.4337/9781783476732.0002>
- Torres Ávila, J. (2016). *La transparencia y el buen gobierno: una perspectiva desde los derechos humanos y las obligaciones de los gobiernos locales*. Bogotá: Universidad Católica de Colombia.
- Townsend, A. (2013). *Smart cities: Big Data, civil hackers and quest for a new utopia*. Nueva York: Norton Company.
- Trazegnies Granda, F. (2013). ¿Seguirán existiendo jueces en el futuro?: el razonamiento judicial y la inteligencia artificial. *Ius et Veritas*, 47, 112-131
- Turkle, S. (1984). *El segundo yo. Las computadoras y el espíritu humano*. Buenos Aires: Galápagos.
- Turkle, S. (2012). *Alone Together. Why we expect more from technology and less from each*. Nueva York: Book News.
- Turow, J. (1993). *Breaking up America: Advertisers and the new media world*. Chicago: Universidad de Chicago.
- Unión Europea (2014). *Horizon 2020 en Breve. El Programa Marco de Investigación e Innovación de la UE*. Luxemburgo: Oficina de Publicaciones de la Unión Europea.
- Valderrama Barragan, M. (2017). El devenir de la identidad digital: del yo proteico al yo identificado. *Revista de Tecnología y Sociedad*, 6(11), 1-19.
- Vapnik, V. (2000). *The nature of statistical learning theory* (2.^a ed.). Londres: Springer Verlag.
- Vasilescu, M. y Terzopoulos, D. (2005). Multilinear independent component analysis. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)* (pp. 547-553). San Diego: IEEE.
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R. y Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government. Recuperado de <https://nexa.polito.it/nexacenterfiles/2016-giq-odq.compressed.pdf>
- Villoria Mendieta, M. y Cruz Rubio, C. N. (2014). *Curso Online: Transparencia y Gobierno Abierto. Módulo 2: Transparencia y buen gobierno: valores y herramientas del gobierno abierto*. Madrid: Banco de Conocimientos INAP.
- Walker, S. y Duncan, D. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54, 167-178.
- Wall, M. E., Rechtsteiner, A. y Rocha, L. M. (2003). Singular value decomposition and principal component analysis. En D. Berrar, W. Dubitzky y M. Granzow Eds.), *A practical approach to microarray data analysis* (pp. 91-109). Norwell: Kluwer.
- Warren, S. y Brandeis, L. (1890). *The right to privacy*. Recuperado de <http://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf>

•Bibliografía•

- Wegener, H. (2014). Ciberseguridad en la Unión Europea. Recuperado de http://www.ieee.es/en/Galerias/fichero/docs_opinion/2014/DIEEEO77bis-2014_CiberseguridadProteccionInformacion_H.Wegener.pdf
- Weinberger, K. y Saul, L. (2005). Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70(1), 77-90.
- Woolcott Oyague, O. (2015). El derecho de autor frente al copyright. En *Protección del derecho de autor* (pp. 5-48). Buenos Aires: Universidad Católica de Colombia.
- Wu, T. (2012, 19 de junio). Free speech for computers? *The New York Times*. Recuperado de <http://www.nytimes.com/2012/06/20/opinion/free-speech-for-compute>
- Zhi-Hua, Z. (2012). *Ensemble methods: Foundations and algorithms*. Recuperado de <http://www2.islab.ntua.gr/attachments/article/86/Ensemble%20methods%20-%20Zhou.pdf>



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

Editado por la Universidad Católica de Colombia en junio de 2018, impreso en papel propalibros de 75 g, en tipografía Times New Roman, tamaño 11 pts.

Publicación digital
Hipertexto Ltda.

Impreso por:
Xpress Estudio Gráfico y Digital S.A.

Sapientia aedificavit sibi domum

Bogotá, D. C., Colombia

El tránsito de lo analógico a lo digital permitió la emergencia de procesos autogestionables que desplazaron a los medios de comunicación convencionales e hicieron posible el acceso a la información de forma extendida, transformando la economía, el derecho y la administración del Estado. Los diferentes textos que se presentan en este libro analizan los impactos de las nuevas tecnologías de la información desde una perspectiva transdisciplinar y multidisciplinar, tomando en consideración fenómenos como *big data*, derecho, internet, ciberdefensa, gobierno electrónico, acceso a la información, brecha digital, protección de datos, inteligencia artificial y cibernética. A partir de disciplinas como la sociología, la ciencia política y el derecho, se avanza en una perspectiva global y compleja acerca de la influencia de las tecnologías de la información y la comunicación en el campo jurídico y en el Estado.

