

DIAGNÓSTICO DE LA SITUACIÓN ACTUAL DE COLOMBIA ESTRUCTURADO ALREDEDOR DE LAS SEIS DIMENSIONES

*Un gobierno orientado al
desarrollo de un ecosistema
sostenible de explotación de
datos y Big Data*

2017-2019

3



DNP Departamento
Nacional
de Planeación



Producido por un equipo compuesto por:

Emmanuel Letouzé, Alex Pentland, Isabella
Loaiza, Julie Ricard, Andrés Clavijo, Nicolas de
Ligny, Daniel Rodríguez, Orlando Saavedra,
Luisa García y Mónica Moreno

Coordinado por Emmanuel Letouzé

Con insumos y apoyo de: Maria Antonia Bravo,
Valentina Casasbuenas Bajo la supervisión
general de: Alex Pentland

Asistencia de edición por: Laura Katalina Muñoz

Enero de 2019

Versión revisada y ajustada

**DIAGNÓSTICO DE LA SITUACIÓN
ACTUAL DE COLOMBIA
ESTRUCTURADO ALREDEDOR
DE LAS SEIS DIMENSIONES**



ÍNDICE

Glosario	8
1. Introducción	10
2. Metodología para el diagnóstico	10
3. Diagnóstico	12
3.1 Dimensión 1: Análisis normativo y ético	12
3.2 Dimensión 2: Organizacional e institucional	22
3.3 Dimensión 3: Infraestructura física (tecnología) orientada a datos	30
3.4. Dimensión 4: Arquitectura de datos	39
3.5 Dimensión 5: Capital humano y Cultura alrededor de los datos	46
3.6 Dimensión 6: Gasto e inversión asociada a la explotación de datos	52
Conclusiones	61
Anexos	62
Anexo A: Estudio Cualitativo	62
Anexo B. Guía de preguntas para las entrevistas realizadas a las entidades	68
Anexo C. Dimensión de Gasto e inversión asociada a la explotación de datos.	71
Anexo D. Dimensión 5, Porcentaje de Entidades Encuestadas	73
Anexo E. Madurez organizacional para la Adopción de Procesos Analíticos	73
Bibliografía	76

RESUMEN EJECUTIVO

El objetivo de este documento es presentar un diagnóstico sobre la situación actual del Estado colombiano frente a la explotación de datos y Big Data. El cuadro abajo resume los problemas más salientes a resolver, así como sus causas y efectos, elementos que fueron identificados tras realizar un estudio de corte cualitativo y cuantitativo analizando las normas, estructuras, institucionalidad e inversión, entre otras, que ha desarrollado el gobierno colombiano en su objetivo de consolidar la transformación digital del Estado. Se identificaron las brechas y elementos claves que el país debe abordar para desarrollar un ecosistema de datos avanzado, un objetivo cada vez más urgente a medida que los gobiernos de alrededor del mundo adoptan estrategias digitales ambiciosas y de alto impacto.

Problema	Causa	Efecto
Marco legal desarticulado	Ausencia de un principio armonizador para el tratamiento y divulgación de información y duplicidad de facultades normativas	Dispersión normativa y asignaciones confusas
Carencia de interoperabilidad	Bajo cumplimiento de lineamientos y guías de interoperabilidad, barreras técnicas, de habilidades y cultura para compartir información	Bajo aprovechamiento de datos por parte de las entidades
Adopción incompleta de estrategias de gobierno digital	Bajo cumplimiento de obligaciones legales, esfuerzos fragmentados para la integración de datos	Bajo aprovechamiento de lineamientos y regulaciones existentes que promueven el buen uso y la explotación de datos
Acceso limitado a datos de calidad del sector público	Ausencia de procesos de anonimización de datos, falta de cumplimiento con obligaciones jurídicas, ausencia de marcos de calidad de datos	Baja inscripción y publicación de datos, acceso a información personal contenida en registros públicos sujeta a justificación de uso y varias restricciones
Carencia de Códigos de Ética	No se han considerado necesarios para el desarrollo de estrategias digitales	No existen lineamientos éticos para el buen uso de datos, necesario según estándares internacionales
Ausencia de un marco de gobernanza de los datos	Ninguna organización estatal tiene el mandato o función de realizar procesos comprensivos de identificación de información al interior del gobierno.	Avance lento en la generación de valor añadido y explotación de los datos por parte del estado
Falta de conocimiento sobre el estado actual de la infraestructura de datos colombiana	Ausencia de inventario o de información detallada sobre producción de datos e información (almacenamiento, sistemas de información, tecnología)	Lineamientos que se proponen no construyen sobre las ventajas o desventajas de la infraestructura de datos actual
Funciones de entidades nacionales no articulan adecuadamente roles asociados a datos e información	No se reconocen estas funciones como prioritarias para el funcionamiento de la entidad	Funciones relacionadas a datos no son transversales a funciones de entidades, lo que permite se tomen como opcionales
Desconocimiento de demanda y oferta de información	No existen métricas que midan la demanda y oferta de información del sector público	Estrategias digitales y otras líneas de acción no se ajustan a la demanda u oferta
Guías para gestión de datos (estandarización, calidad, entre otros) incompletas	Gestión de datos orientada a la migración de documentos físicos a electrónicos sin priorizar interoperabilidad o explotación de estos	Gestión de datos no conduce a generar valor agregado a partir de estos

Problema	Causa	Efecto
Se desconoce la brecha de capacidades tecnológicas del sector público	Metodología para medir brecha en capital humano no se ha aplicado al sector público	No existe información detallada sobre el estado actual, la demanda u oferta de capacidades
Ausencia de educación tecnológica y bajo desempeño en habilidades digitales en el país	Baja alfabetización digital y pocos programas educativos enfocados en tecnología	Poca oferta de capital humano para la explotación de datos
Bajos incentivos salariales para trabajadores en el sector TI	No se considera necesario aumentar incentivos para retener talento	Baja motivación y retención de trabajadores de TI y obstáculos para desarrollar un capital humano capacitado
Baja inversión nacional en I+D	No se reconoce la importancia de inversión en ciencia, tecnología e innovación	Atraso en generación de conocimiento propio a partir de la I+D y bajo desarrollo del sector de datos

GLOSARIO

Las definiciones presentadas a continuación se relacionan directamente con el proyecto en curso y son conceptos emitidos a partir de múltiples investigaciones académicas llevadas a cabo por Data-Pop Alliance.

Big Data vs. big data: el ecosistema creado por la emergencia de 'las 3 C de Big Data'

Cuadro 1. Big Data versus big data

<p><i>Big Data</i> (con mayúsculas) en este documento (y otros) se refiere al ecosistema creado por la aparición concomitante de "las 3C de <i>Big Data</i>."</p> <ul style="list-style-type: none"> La primera C representa las migajas (<i>crumbs</i> en inglés), o pedazos de datos emitidos y recolectados de forma pasiva, subproducto de la interacción de las personas con y el uso de dispositivos digitales que proporciona una visión única sobre sus comportamientos y creencias; La segunda C representa las Capacidades de <i>Big Data</i>, lo que también se conoce como <i>Big Data Analytics</i>; es decir, el conjunto de herramientas y métodos, <i>hardware</i> y <i>software</i>, <i>know-how</i> y habilidades necesarios para procesar y analizar este nuevo tipo de datos, incluyendo técnicas de visualización, aprendizaje estadístico automatizado (<i>machine learning</i>), algoritmos, etc.; La tercera C representa las Comunidades de <i>Big Data</i>, y describe los diferentes actores involucrados en el ecosistema de <i>Big Data</i>, desde los generadores de datos hasta sus analistas y usuarios finales; es decir, potencialmente toda la población. <p>Este ecosistema puede ser descrito y analizado como un sistema complejo, es decir, uno donde existen bucles de retroalimentación entre sus diferentes partes. En los niveles más básicos, las nuevas empresas (por ejemplo Twitter o sus futuros competidores) ayudan a generar nuevos tipos de datos que a su vez conducen al desarrollo de nuevos tipos de instrumentos analíticos, dando lugar a nuevos tipos de datos, y luego a nuevos actores que toman ventaja de estos nuevos datos y herramientas. Es posible que este nuevo ecosistema pueda convertirse en o ser parte de un fenómeno social más amplio.</p> <p>Por el contrario, <i>big data</i> se refiere a la primera C mencionada anteriormente; es decir, los flujos y los conjuntos resultantes de las huellas digitales que dejan los seres humanos al usar teléfonos celulares (registros de llamadas), tarjetas de crédito (transacciones), el transporte (registros de metro o autobuses y de EZ Pass), medios sociales y motores de búsqueda, o cuando sus acciones son registradas por sensores, ya sean físicos (contadores eléctricos, sensores de pesaje en un camión) o remotos (satélites, cámaras).</p>
--

Fuente: Adaptado de Letouzé, E. "Big Data and Development Overview Primer". Data-Pop Alliance, SciDev.Net and the World Bank (2015)

Dato: Un objeto, variable o información que tiene la capacidad de ser identificable, recopilado y almacenado. Viene principalmente en dos formas: estructurado y no estructurado. Los datos estructurados son esencialmente respuestas a preguntas formuladas por el recopilador de datos; estos generalmente son fáciles de organizar e identificar y tienen una jerarquía estricta que no es de fácil manipulación. Por otro lado, los datos no estructurados no son fácilmente procesables en análisis automatizados y, a menudo, se utilizan de forma distinta a la finalidad prevista cuando se recopilan (como fotos, videos, tweets) y no necesitan seguir un método jerárquico de identificación. Los datos también se utilizan como un concepto de política y fenómenos sociales o como un atajo para los ecosistemas de datos, Big Data, etc.

Algoritmo: en matemáticas e informática, un algoritmo es una serie de instrucciones o reglas predefinidas, que por lo general, están escritas en un lenguaje de programación

destinado a ser utilizado por una computadora y diseñado para definir cómo resolver secuencialmente un problema recurrente a través de cálculos y procesamiento de datos. El uso de algoritmos para la toma de decisiones ha crecido en varios sectores y servicios, como en la policía y el sector financiero.

Tecnología cívica: es un tipo de tecnología que permite la participación ciudadana o hace que el gobierno sea más accesible, efectivo y eficiente para el bien económico y social de la sociedad. Este tipo específico de tecnología ayuda a conectar a las personas con recursos, ideas y otros individuos necesarios para mejorar sus sociedades o comunidades.

Ecosistemas de datos: sistemas adaptativos complejos que incluyen infraestructura de datos, herramientas, medios, productores, consumidores, curadores y personas que comparten datos. Los ecosistemas de datos son organizaciones complejas de relaciones sociales dinámicas a través de las cuales los datos/información se mueven y se transforman en flujos.

Inclusión de datos: es la capacidad universal de las personas para crear, controlar, acceder y usar datos.

Alfabetización de datos: el deseo y la capacidad de involucrarse constructivamente en la sociedad a través de datos y con datos.

Revolución de datos: este concepto se refiere a las aplicaciones e implicaciones de los datos como un fenómeno social. El término "revolución industrial de datos" fue acuñado por el informático Joseph Hellerstein en 2008.

Ciencia de datos: es un campo de investigación y práctica que se enfoca en resolver problemas del mundo real a través del uso de grandes cantidades de datos combinando habilidades de múltiples campos: matemáticas, informática, estadísticas, ciencias sociales e incluso narración de cuentos o arte.

Información abierta: datos de fácil acceso, legibles por máquina, accesibles de forma gratuita a un costo insignificante.

Gobierno electrónico (e-government): aplicación de las tecnologías de la información y la comunicación (TIC) para el correcto funcionamiento del sector público. Esto, con el objetivo de brindar mejores servicios al ciudadano e incrementar la eficiencia, la transparencia y la participación ciudadana (Departamento para la Gestión Pública Efectiva n.d.).

Gobierno en línea¹: Política pública de Colombia que se implementó en el país desde el 2000, y que se definió como Estrategia de Gobierno en Línea (GEL) mediante el Decreto 1151 de 2008. Esta política de gobierno electrónico estaba orientada hacia un modelo de gobernanza cobijado por la transparencia, el buen gobierno, la inclusión, y el acceso a la información y a los datos y se fundamentó en cuatro pilares: TIC para Gobierno Abierto, TIC Gestión, TIC para Servicios, y Seguridad y Privacidad de la Información, con el objetivo de empoderar al ciudadano para interactuar con el Estado mediante las TI. Las líneas de acción de la estrategia de Gobierno en línea incluían mecanismos de rendición de cuentas, la Ley de Transparencia y Acceso a la Información Pública, y canales de comunicación, interacción y de servicios con los ciudadanos. De acuerdo al observatorio TI, “los logros más importantes de la estrategia se relacionan con el número de trámites en línea de orden nacional y territorial incorporados al portal del Estado colombiano, creación de portales territoriales, servicios sectoriales en línea y consolidación de la plataforma de interoperabilidad para integrar la gestión de las distintas entidades” (MINTIC y Fedesoft, n.d.).

Gobierno Digital: política del Estado colombiano con enfoque en la transformación digital del país, vigente como la estrategia de gobierno electrónico a partir del Decreto 1008 del 14 de junio del 2018. El objetivo de la estrategia es “incentivar el uso y aprovechamiento de las TIC para consolidar un Estado y ciudadanos competitivos, proactivos e innovadores, que generen valor público en un entorno de confianza digital” (MINTIC 2018). Los componentes principales de la política de Gobierno Digital son TIC para el Estado y TIC para la Sociedad con tres habilitadores transversales a estos dos componentes: la seguridad de la información, la arquitectura y los servicios ciudadanos digitales. La política de Gobierno Digital busca responder a las tendencias y necesidades que se desarrollan a nivel nacional, con un nuevo enfoque al articular a diferentes actores de la sociedad con el estado.

CONPES 3920: Política Nacional de Explotación de Datos: Este documento CONPES presenta la política de aprovechamiento de datos del Estado colombiano, con énfasis en gestionar los datos como activos para generar valor tanto social como económico. Esta política cuenta con líneas de acción varias que apuntan al aprovechamiento de datos en el país, en un periodo de implementación de cinco años y con una inversión total de 16.728 millones de pesos (DNP, MINTIC, y SIC 2018).

¹ Para más información sobre los componentes de Gobierno en Línea, hacer click aquí.

1. INTRODUCCIÓN

En los últimos años, los gobiernos le han apostado a la implementación de estrategias basadas en el uso de datos para aumentar la eficiencia y transparencia de sus programas en la solución puntual de las necesidades de los ciudadanos. Con el desarrollo e innovación de nuevas tecnologías, los datos han tomado un papel protagónico, volviéndose fundamentales para la toma de decisiones gubernamentales y siendo los gobiernos tanto productores, como consumidores permanentes de grandes volúmenes de datos.

En este contexto, Big Data se volvió un proveedor de conocimiento para las organizaciones públicas y privadas. Esta innovación del campo de las tecnologías de la información y las comunicaciones, ha servido de impulso para nuevas oportunidades de negocio, nuevas profesiones, avances en salud, movilidad, entre otros. Además, se ha establecido en el sector público como un instrumento para el diseño e implementación de políticas para combatir la pobreza y mejorar la seguridad (World Economic Forum, 2015). El Big Data ha ganado tanta relevancia, que es usado con frecuencia en áreas gubernamentales, de ayuda humanitaria, educación, cambio social, salud pública, y otros campos de “bienestar social” (Coppola et al. 2014, D'Ignazio y Bhargava 2015). Este nuevo uso implica nuevos espacios de experimentación, innovación y adaptación.

De hecho, en el 2016 el Banco Interamericano de Desarrollo afirmó que el Big Data es un aspecto fundamental para la transparencia y eficiencia en el desarrollo de políticas públicas y en la evaluación y monitoreo de estas (Tomar et al. 2016). Al adquirir un tono social y político como herramienta, tecnología y a su vez fuente para el aprovechamiento de datos y la construcción de políticas públicas, es importante tener en cuenta los retos del sector público frente a la aparición del Big Data, así como los elementos necesarios para que este aproveche correctamente los datos a su disposición.

Teniendo clara su relevancia, este documento tiene como objetivo hacer un diagnóstico del estado actual de la captura, almacenamiento, procesamiento, acceso, análisis y visualización de grandes volúmenes de datos en las entidades públicas de orden nacional. Lo anterior servirá como insumo para el diseño de una estrategia de Big Data que pueda ser implementada en las diferentes entidades gubernamentales.

Para esto, se pretende analizar las siguientes dimensiones: (1) identificar el estado actual del marco normativo y ético aplicable a los datos, (2) análisis organizacional e institucional, (3) infraestructura física (tecnología) orientada a datos, (4) las arquitecturas para el almacenamiento y procesamiento de los datos, (5) el capital humano y la cultura de datos, y (5) gasto e inversión asociada a la explotación de datos en el país. Asimismo, este insumo define, entre otros elementos, los vacíos y necesidades del sector público en cuanto a materia de explotación de datos. En el caso de Colombia, a pesar de los avances en aprovechamiento de datos, producción estadística, transformación digital y múltiples estrategias de gobierno electrónico, todavía es necesaria la consolidación de un gobierno que priorice los datos como un insumo fundamental para la estructuración de políticas públicas.

En los documentos 1 y 2 de esta consultoría, (aquí y aquí) se elaboró un análisis a partir de seis estudios de caso - Reino Unido, Singapur, Estados Unidos, México, Estonia y Corea del Sur - para obtener lecciones claves y recomendaciones potenciales para el gobierno colombiano. En este documento se realiza el diagnóstico, estructurado alrededor de seis dimensiones mencionadas anteriormente.

De esta manera, el documento se organiza en 6 secciones. La primera contiene una lista de definiciones y términos que enmarcan conceptos importantes a tener en cuenta en las demás secciones. La segunda sección presenta la descripción de la metodología usada y el marco lógico utilizado para el análisis y desarrollo del diagnóstico. La tercera sección presenta los diagnósticos de las seis dimensiones, y la cuarta sección presenta las conclusiones principales de la situación actual del estado colombiano frente a la explotación de datos y Big Data. Finalmente, las últimas secciones presentan los anexos y la bibliografía utilizada.

2. METODOLOGÍA PARA EL DIAGNÓSTICO

La metodología de marco lógico es una herramienta que facilita el proceso de conceptualización, diseño, ejecución y evaluación de proyectos. El marco lógico hace énfasis en la orientación de la identificación de problemas y de una serie de objetivos; este método permite homogeneizar los conceptos y reduce posibles ambigüedades. En este contexto, la metodología del marco lógico contempla el análisis del problema, análisis de los principales actores involucrados y, además, ayuda a jerarquizar los objetivos para la selección de una estrategia de implementación óptima (“Technical Note: The Logical Framework” 2012).

Cabe resaltar que este documento está centrado en la primera etapa del proceso de marco lógico; donde se identifica el problema y se analiza a profundidad la situación actual con fortalezas y desafíos. Esto, permite crear una visión de los objetivos y metas a las cuales se quiere llegar, lo que permitirá plasmar en los siguientes documentos las líneas de acción y estrategias necesarias para llevar a Colombia a la situación deseada.

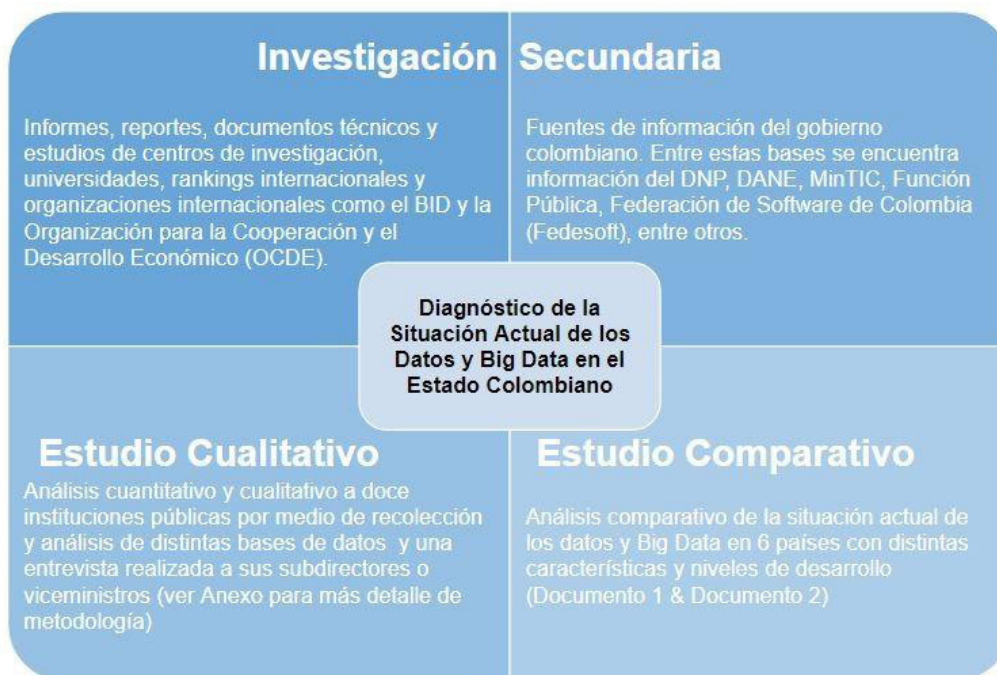
Por consiguiente, la identificación del problema nace de la necesidad de establecer parámetros donde se enmarque la estrategia nacional de Big Data. De acuerdo a este objetivo, es imprescindible llevar a cabo un análisis estructurado de la

situación existente, para lograr un marco lógico que incorpore los elementos necesarios para guiar el proceso de diagnóstico. Para esto, se realizó un análisis de múltiples variables con enfoque cuantitativo y cualitativo, lo que permitió entender la dinámica de las entidades del Estado colombiano en torno al manejo de información y aprovechamiento de este.

En el siguiente gráfico, se describe la metodología utilizada para el desarrollo del diagnóstico de la situación actual de los datos y Big Data en el Estado colombiano, estructurada alrededor de seis dimensiones: (1) análisis normativo y ético, (2) organizacional e institucional, (3) infraestructura física (tecnología) orientada a datos, (4) arquitectura de datos, (5) capital humano y cultura alrededor de los datos (6) gasto e inversión asociados a la explotación de datos.

Para el estudio de estas, se realizaron análisis de corte cuantitativo y cualitativo, complementados con referentes internacionales y documentos previos de esta consultoría. Con esto, se pretende entender a profundidad el proceso de captura, almacenamiento, procesamiento, acceso, análisis y publicación de grandes volúmenes de datos en cada una de las entidades entrevistadas.

Figura 1: Metodología para el desarrollo del diagnóstico de la situación actual de los datos y Big Data en Colombia.



Fuente: elaboración propia

Las entrevistas del estudio cualitativo se realizaron a 12 entidades públicas del orden nacional con un cuestionario de 54 preguntas diseñado por Data-Pop Alliance, el Departamento Nacional de Planeación (DNP) e InnPulsa. Dichas entrevistas fueron realizadas a sub-directores, viceministros o personas encargadas del manejo de datos. Esta fue estructurada principalmente en tres secciones: la primera sección consistió en preguntas para dar contexto sobre lo que las entidades entienden por explotación de datos y Big Data; la segunda sección se centró en el entendimiento del panorama actual de las diferentes entidades referente a la generación, análisis y difusión de los datos; y, por último, en la tercera sección, se hizo un enfoque sobre los proyectos futuros y estrategias con respecto a las Tecnologías de Información y fortalecimiento de cultura de datos en estas entidades.

Para la selección de las entidades, se tomó como base los resultados de la encuesta realizada por el DNP a 150 entidades para la construcción del Documento CONPES 3920 Política Nacional de Explotación de Datos (Big Data), donde se determinó el estado actual de preparación de cada una de estas para la explotación de datos. El criterio principal en la selección de las entidades fueron las capacidades tecnológicas de estas instituciones, especialmente en infraestructura y sistemas de información, la apropiación de cultura de datos para realizar analítica o toma de decisiones basada en datos, y las capacidades de capital humano para realizar análisis de datos.

A partir de estos insumos principales descritos anteriormente, fue posible identificar los problemas que afectan el funcionamiento de las entidades respecto a la explotación de datos y aprovechamiento del Big Data, los cuales serán expuestos a continuación.

Tabla1: Entidades seleccionadas para las entrevistas.

	Entidad
1	Colombia Compra Eficiente
2	Coldeportes
3	DIAN
4	DIMAR
5	Fondo de Adaptación
6	Función Pública
7	IDEAM
8	Ministerio de Agricultura
9	Ministerio de Hacienda
10	Ministerio de Trabajo
11	Ministerio de Salud
12	Policía Nacional

Fuente: Elaboración propia

3. DIAGNÓSTICO

3.1 DIMENSIÓN 1: ANÁLISIS NORMATIVO Y ÉTICO

En Colombia, existen directrices legales y judiciales que abordan temas relacionados con el manejo de datos e información. A continuación, se realiza inicialmente una descripción de la normativa existente aplicable a la captura, acceso, análisis y publicación de datos por parte de las entidades estatales, para después realizar un diagnóstico sobre el marco normativo en Colombia para el manejo y el uso de los datos. Esto implica analizar el conjunto de normas aplicables al manejo y uso de datos. Se considera esencial entender cómo las estrategias digitales que ha desarrollado Colombia a través de los años contribuyen al desarrollo de una política de Big Data.

Ahora bien, la obtención de información a través del procesamiento de datos no escapa, al necesario análisis ético. En primer lugar, por la sencilla razón de estar involucrado un componente humano dentro de la cadena de valor: la persona humana es la principal generadora de la materia prima, que son los datos. En segundo lugar, porque es el ser humano quien utiliza estos datos a su favor, sea de manera individual o colectiva. En sí, el hecho de aprovechar los datos no involucra mayor escollo, sin embargo, cuando generar utilidad a partir de los datos se torna nocivo o cuestionable, es necesario introducir nociones de si el uso de estos fue éticamente correcto o no. A nivel internacional, existen marcos y directrices éticas que establecen la correcta utilización de los datos, particularmente los personales.

Así pues, en esta sección, primero se analiza la normatividad desde un punto de vista jurídico, para luego abordar el análisis ético en cuanto a la explotación de datos en las entidades del Estado colombiano.

ANÁLISIS NORMATIVO

3.1.1 ANÁLISIS DEL MARCO JURÍDICO FRENTE AL ACCESO A LA INFORMACIÓN

Para trabajar con un enfoque hacia la normativa aplicable a aquellos aspectos relacionados con la información, los datos abiertos y los datos personales que son recolectados por las entidades públicas, se debe partir del análisis de dos leyes estatutarias²: por un lado, la Ley de transparencia y acceso a la información pública, Ley 1712 de 2014 y, por otro, la Ley

² El artículo 152 de la Constitución establece que la regulación de ciertas materias debe hacerse mediante ley estatutaria, como una categoría normativa sujeta a un procedimiento más exigente que el ordinario, dispuesto en el artículo 153 de la Carta. Así, el Congreso de la República debe tramitar mediante leyes estatutarias, las normas sobre: (i) los derechos y deberes fundamentales de las personas y los procedimientos y recursos para su protección; (ii) la administración de justicia; (iii) la organización y régimen de los partidos y movimientos políticos, el estatuto de la oposición y funciones electorales; (iv) las instituciones y mecanismos de participación ciudadana; (v) los estados de excepción; y (vi) la igualdad electoral entre los candidatos a la Presidencia de la República que reúnan los requisitos que determine la ley. Sentencia C-007 del 18 de enero de 2017, Magistrada Sustanciadora Dra. Gloria Stella Ortiz Delgado.

general de protección de datos personales, Ley 1581 de 2012. En este sentido, el manejo de datos se encuentra regido bajo la aplicación de leyes de igual valor y superioridad jerárquica en el ordenamiento jurídico colombiano, a partir de las cuales se imponen pautas de orden jurídico para la creación, desarrollo e interpretación de normas de menor jerarquía, tales como leyes ordinarias, decretos, resoluciones, circulares, etc. Lo anterior implica que, al momento de realizar algún tratamiento de datos que propicie la adquisición de nuevos conocimientos en beneficio del interés general y el desarrollo de los fines de cualquier entidad de carácter público, se debe respetar el derecho a la intimidad y habeas data de las personas cuya información pueda ser objeto de tratamiento.

Es imprescindible entender que el funcionamiento actual de las entidades públicas ha sido definido a partir de normas cuyo sustento ontológico y jurídico se encuentra en la Constitución Política de 1991, sirviendo como ejemplo la Ley 489 de 1998, la cual establece la organización y funcionamiento de las entidades y organismos pertenecientes a la Rama Ejecutiva, quien orienta sus actividades administrativas a la realización de los fines del Estado.

3.1.1.1 ANÁLISIS DE LOS MARCOS NORMATIVOS ESTATUTARIOS.

La ley 1266 de 2008 constituyó el primer paso hacia la protección del dato personal, a pesar de otorgarle un enfoque especializado: el dato financiero. En efecto, la ley 1266 cimentó las bases conceptuales del dato personal de manera general, para enfocarse en la protección de los datos que se almacenan en bancos y los deberes, responsabilidades y obligaciones de quienes administran tales bancos de datos. Si bien la regulación allí contenida es especializada, y en ciertos casos es de aplicación restringida, sirvió no sólo como sustento a la teoría del dato personal, sino al fomento de una cultura de protección a la información personal.

No fue sino hasta la promulgación de la ley 1581 de 2012 que se pudo hablar con propiedad de un régimen de protección de datos personales. Aparte de adoptar las definiciones que la ley 1266 de 2008 había implementado, la nueva norma desarrolló todo un marco y sistema normativo que permite a los particulares exigir al Estado y sus pares, la protección de sus datos personales sin necesidad de estar almacenados en un banco de información financiera. Ésta retomó los principios y la clasificación de dato público, semiprivado, privado y sensible, estableció el régimen del responsable y el encargado de los datos personales e impuso una serie de sanciones de alto rigor para quienes incumplieran las disposiciones normativas relativas al dato personal.

Adicionalmente, la ley 1581, debido a su característica de ser una norma general, requirió regulación gubernamental, con el fin de precisar algunos conceptos esenciales, adoptar otros mecanismos complementarios, así como adoptar un régimen frente a los datos obtenidos previa la promulgación de la ley. Así, se pudo definir el contexto de autorización del dato sensible, el aviso de privacidad, el contenido de la política de tratamiento de información personal y por último, se incluyó un principio de aplicación denominado la “responsabilidad demostrada”, mediante el cual los responsables y encargados del dato personal deberán dar evidencia del cumplimiento de los parámetros legales y reglamentarios de la norma, en razón a la naturaleza del dato personal, la naturaleza de responsable o encargado, el tipo de tratamiento, entre otros.

Al respecto, es importante precisar que el fundamento legal para negar el suministro de información personal a terceros, sin un interés legítimo, se encuentra en el artículo 13 de la Ley 1581 de 2012. Por cuanto, sólo habilita a las siguientes personas: a) a los Titulares, sus causahabientes o sus representantes legales; b) a las entidades públicas o administrativas en ejercicio de sus funciones legales o por orden judicial y; c) a los terceros autorizados por el Titular o por la ley. Tal como veremos más adelante, esta norma supone una posición normativa tendiente a la privacidad, contraponiéndose a la posición normativa que tiende al acceso y transparencia de la información, como veremos inmediatamente.

En efecto, la Ley 1712 de 2014, de Transparencia y del Derecho de Acceso a la Información Pública Nacional, tiene como objeto regular el derecho de acceso a la información pública, los procedimientos para el ejercicio y garantía del derecho y las excepciones a la publicidad de información. Esta Ley plantea las excepciones y limitaciones al acceso a todo tipo de información, incluida la personal. Corolario de lo anterior se encuentra el artículo 2º de dicha ley, en donde se aclara que toda información que repose en las entidades públicas es efectivamente de carácter público; no obstante se plantea igualmente una limitación para aquella información que esté limitada por reserva o disposición constitucional o legal, v. g.r. las disposiciones de la ley 1581 de 2012.

Dentro del mismo cuerpo normativo, el Literal (c) del artículo 6 de la Ley 1712 de 2014, define la información pública de carácter clasificada como:“(…) *aquella que estando en poder o custodia de un sujeto obligado en su calidad de tal, pertenece al ámbito propio, particular y privado o semiprivado de una persona natural o jurídica (...)*”, por lo cual es posible afirmar que su acceso podrá ser negado o exceptuado cuando se configuren circunstancias legítimas y necesarias en los términos de la misma norma. Así, el artículo 18 de la precitada Ley señala excepciones al tratamiento de información personal en caso de que el acceso pudiere causar un daño, al derecho a la intimidad, entre otros derechos, dicho acceso podrá ser rechazado o denegado de manera motivada y por escrito.

Al observar el artículo 18, se identifica la facultad adicional del responsable del tratamiento de los datos (entidad) de revelarlos cuando su titular lo haya autorizado. De esta forma, el artículo 9 de la Ley 1581 de 2012, relacionado con la autorización del Titular, señala que “*Sin perjuicio de las excepciones previstas en la ley, en el Tratamiento se requiere la autorización previa e informada del Titular, la cual deberá ser obtenida por cualquier medio que -ser objeto de consulta posterior-*”. Lo anterior

significa que deben obtenerse las autorizaciones correspondientes para el tratamiento de los datos con los requisitos que exige el artículo 12 de la Ley 1581 de 2012.

De modo que, si bien es cierto que existe una Ley que reglamenta el acceso a la información personal contenida en registros públicos (Ley 1712 de 2014), también existe una norma que impone restricciones al suministro de dicha información cuando genera afectación a otros derechos fundamentales de los Titulares de esta. Por lo tanto, ambas leyes deben ser sopesadas y tenidas en cuenta a la hora de considerar el acceso a la información pública.

Igualmente, es necesario tener en cuenta que la Ley 1581 de 2012, con el desarrollo del Registro Nacional de Bases de Datos, buscó la creación de un directorio público de las bases de datos personales sujetas a Tratamiento que operan en el país, lo que favoreció el acceso a la información pública. No obstante, de acuerdo con las consideraciones del Decreto 90 de 2018 “(...) pese a la divulgación y socialización de esta obligación realizada por la Superintendencia de Industria y Comercio dentro de los términos de extensión del referido plazo, el número de Responsables de Tratamiento que han inscrito sus bases de datos representa sólo el veinticinco por ciento (25%) del total de personas jurídicas y menos del uno por ciento (1%) de personas naturales”.

De esta forma, el propósito de la normatividad relativa a los datos personales del presente diagnóstico es buscar que la información que generen diferentes actores del sector público y privado, y que esté disponible para la sociedad, contenga un valor agregado que permita su utilización en la elaboración de proyectos, investigaciones y, en general, en la realización de consultas en diferentes temáticas, lo cual aporta avances para la sociedad. No obstante, la información que como individuos, empresas, organizaciones o Estado se decide compartir debe garantizar el respeto por los derechos de las personas que puedan estar relacionados con la misma para proteger su privacidad y sus datos personales. Por esto, una de las obligaciones de cualquier entidad es identificar qué información puede ser útil para la sociedad y depurarla de cualquier dato que pueda afectar la intimidad de las personas, para dar aplicación previa a la Ley Estatutaria 1581 de 2012.

3.1.1.2 ANÁLISIS DE LAS DEMÁS FUENTES DE DERECHO.

Por parte del legislador existía una preocupación por proteger los datos personales de manera general, previa la expedición de la ley 1581 de 2012, atendiendo la preocupación frente a la información que se comparte, la cual puede ser utilizada con fines inapropiados que conlleven a la comisión de hechos ilícitos o que perjudiquen ostensiblemente la privacidad de los individuos. Por ello, el legislador, a través de la Ley 1273 de 2009, creó nuevos tipos penales relacionados con delitos informáticos y seguridad de la información y de los datos, cumpliendo así con los fines de la pena de prevención general, prevención especial y retribución justa, y protegiendo los intereses jurídicos tutelados como la privacidad, protección del dato personal, entre otros de índole económico (protección del secreto profesional y la protección a la propiedad privada).³

También es necesario tener en cuenta el Decreto 3851 de 2006, expedido con el objetivo de estandarizar la calidad, el almacenamiento y consulta de información oficial básica en el territorio nacional. En él, se establece que a partir del procesamiento de bases de datos conformadas por registros, censos, encuestas y observaciones, se obtiene información básica de carácter estadístico, geográfico, de personas y territorial. Esta normatividad es relevante en cuanto podemos extraer de ella la diferencia fundamental entre información y dato.

3.1.1.3 PROFUNDIZACIÓN DE LOS TÉRMINOS LEGALES DE DATO PERSONAL E INFORMACIÓN.

Es posible afirmar que el dato es la materia prima a través de la cual se obtiene información. Este elemento, por tanto, merece un tratamiento jurídico por parte del Estado, para garantizar su correcto tratamiento y aprovechamiento. En ese orden de ideas, la Ley 1712 de 2014 enmarca, en términos generales, el concepto de información, describiéndolo “*como un conjunto organizado de datos*”. No obstante, esta ley no sólo se limita a definir información sino los subtipos en que la misma se puede clasificar, como: (i) información pública, referente a toda la información que cualquier entidad pública genere; (ii) información pública clasificada, entiende que dentro de la información que generen las entidades se puede encontrar información que afecte la privacidad de las personas naturales o jurídicas a las cuales se refiere, limitando su acceso si no se encuentra frente a circunstancias legítimas de acuerdo con la misma ley; (iii) información reservada, aquella información que genere una entidad, pero cuyo acceso puede ser negado cuando pueda afectarse el interés general, ya sea porque es sensible frente a temas de seguridad nacional, salud pública, relaciones internacionales, entre otras circunstancias indicadas por la Ley. Estas definiciones, se contemplan en el artículo 6 de la precitada Ley:

“(..)

- a) **Información.** *Se refiere a un conjunto organizado de datos contenido en cualquier documento que los sujetos obligados generen, obtengan, adquieran, transformen o controlen;*
- b) **Información pública.** *Es toda información que un sujeto obligado genere, obtenga, adquiera, o controle en su calidad de tal;*

³ Por ejemplo: el acceso abusivo a un sistema informático, la obstaculización ilegítima de un sistema informático, interceptación de datos informáticos, daño informático, violación de datos personales, entre otros.

- c) Información pública clasificada.** *Es aquella información que estando en poder o custodia de un sujeto obligado en su calidad de tal, pertenece al ámbito propio, particular y privado o semiprivado de una persona natural o jurídica por lo que su acceso podrá ser negado o exceptuado, siempre que se trate de las circunstancias legítimas y necesarias y los derechos particulares o privados consagrados en el artículo 18 de esta ley;*
- d) Información pública reservada.** *Es aquella información que estando en poder o custodia de un sujeto obligado en su calidad de tal, es exceptuada de acceso a la ciudadanía por daño a intereses públicos y bajo cumplimiento de la totalidad de los requisitos consagrados en el artículo 19 de esta ley (...)*

Los cuerpos normativos estatutarios previamente analizados frente a la protección de datos personales permiten una aproximación a la definición legal de dato personal. Tanto la ley 1266 de 2008 como la ley 1581 de 2012 contienen conceptos claros frente a la definición y clasificación legal de dato personal. De esta manera, cabe resaltar:

Artículo 3 de la Ley 1581 de 2012:

“(..)

- b) Base de Datos:** *Conjunto organizado de datos personales que sea objeto de tratamiento;*
- c) Dato personal:** *Cualquier información vinculada o que pueda asociarse a una o varias personas naturales determinadas o determinables;*
- 2. Dato público:** *Es el dato que no sea semiprivado, privado o sensible. Son considerados datos públicos, entre otros, los datos relativos al estado civil de las personas, a su profesión u oficio y a su calidad de comerciante o de servidor público. Por su naturaleza, los datos públicos pueden estar contenidos, entre otros, en registros públicos, documentos públicos, gacetas y boletines oficiales y sentencias judiciales debidamente ejecutoriadas que no estén sometidas a reserva.*
- 3. Datos sensibles:** *Se entiende por datos sensibles aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición, así como los datos relativos a la salud, a la vida sexual, y los datos biométricos.(..)*

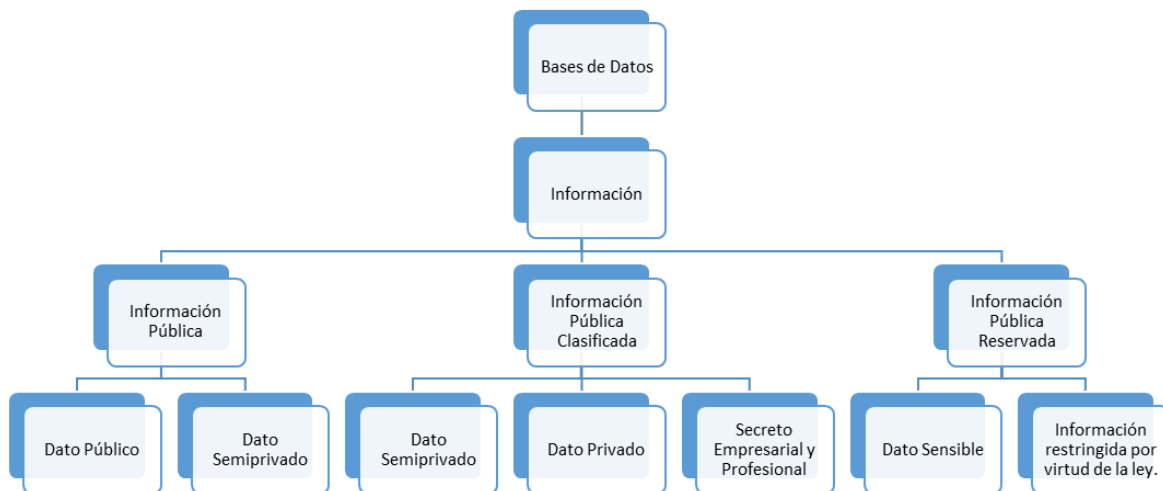
A. Artículo 3 de la Ley 1266 de 2008:

“(..)

- e) Dato personal.** *Es cualquier pieza de información vinculada a una o varias personas determinadas o determinables o que puedan asociarse con una persona natural o jurídica. Los datos impersonales no se sujetan al régimen de protección de datos de la presente ley. Cuando en la presente ley se haga referencia a un dato, se presume que se trata de uso personal. Los datos personales pueden ser públicos, semiprivados o privados;*
- f) Dato público.** *Es el dato calificado como tal según los mandatos de la ley o de la Constitución Política y todos aquellos que no sean semiprivados o privados, de conformidad con la presente ley. Son públicos, entre otros, los datos contenidos en documentos públicos, sentencias judiciales debidamente ejecutoriadas que no estén sometidos a reserva y los relativos al estado civil de las personas;*
- g) Dato semiprivado.** *Es semiprivado el dato que no tiene naturaleza íntima, reservada, ni pública y cuyo conocimiento o divulgación puede interesar no sólo a su titular sino a cierto sector o grupo de personas o a la sociedad en general, como el dato financiero y crediticio de actividad comercial o de servicios a que se refiere el Título IV de la presente ley.*
- h) Dato privado.** *Es el dato que por su naturaleza íntima o reservada sólo es relevante para el titular. (...)*

Como se puede observar, las definiciones de datos de las leyes 1581 de 2012 y 1266 de 2008 son complementarias entre sí, específicamente las referentes a dato público y dato personal. Para mejor ilustración de lo expuesto véase la siguiente figura:

Figura 2: Proceso normativo de los datos.



Fuente: Elaboración propia

Frente a lo indicado, surge la necesidad de implementar procesos de anonimización de datos personales, cuyo objetivo es controlar el riesgo de identificación de las personas; es decir, mitigar los riesgos que se presentan en la obtención y tratamiento masivo de datos de carácter personal. Para esto, se oculta la información privada o sensible del Titular de los datos⁴, lo que permite su divulgación sin que haya vulneración de los derechos a la protección de los datos personales.

Cuando se hace alusión al concepto de anonimizar datos personales, esto se refiere a la supresión, eliminación o transformación de toda aquella información específica que permita identificar a una persona en cualquier medio físico o automatizado. Lo anterior tiene como objetivo excluir de manera definitiva las posibilidades de identificación de los datos asociados a una persona en los motores de búsqueda y garantizar sus derechos conforme lo establece la Ley 1581 de 2012. Este concepto resulta relevante en los entornos que surgen con la evolución tecnológica y fenómenos como Big Data u Open Data, los cuales obligan a ser especialmente meticulosos en el tratamiento de los datos personales al interior de las compañías con el fin de garantizar de manera eficaz la protección de los derechos de sus Titulares.

En este contexto, es importante observar que la anonimización de los datos surge como un instrumento para minimizar los riesgos que se presentan en el tratamiento masivo de los datos de carácter personal. Por consiguiente, éste proceso permite adoptar procedimientos acordes a las normas y en beneficio de las empresas y el comercio, sin contravenir el ámbito privado de sus titulares. Esto implica la adopción de procesos que permitan identificar y ocultar la información que pueda generar cualquier afectación o violación a los derechos de los Titulares de ésta, lo que impide su divulgación.

Figura 3: Problema de anonimización: riesgo de re-identificación versus utilidad de la información



Fuente: Elaboración propia (calculos propios)

⁴ Literal f) del artículo 3 de la Ley 1581 de 2012. "Titular: Persona natural cuyos datos personales sean objeto de Tratamiento".

La Figura 3 muestra las dos dimensiones en las que se deberían enfrentar las entidades que quieran anonimizar información para su publicación. La utilidad de la información versus el riesgo o probabilidad de que se pueda re-identificar al individuo. A más información no anonimizada entregada, más es el riesgo de identificación; por eso, se debe buscar en los procesos de anonimización máximos riesgos tolerables que permitan maximizar la utilidad de la información, controlando a su vez la posibilidad de riesgo de identificación. Dichos procesos deben ser aplicados a la información contenida en los documentos físicos y digitales de cualquier compañía o entidad donde se registre información personal y que, por razones legales, no deba ser identificada o que sus titulares hayan solicitado su anonimato como condición para autorizar su tratamiento. Asimismo, una de las consideraciones más importantes para tener en cuenta en el proceso es realizar una prueba de “identificabilidad” a la técnica de anonimización aplicada para definir su efectividad y para ello se debe identificar con claridad el tipo de dato y el Titular de éste.

En Colombia, no se encuentran reglamentadas las condiciones para los procesos de anonimización de los datos personales, por lo que cada empresa o entidad debe evaluar la conveniencia y las técnicas más apropiadas según sus herramientas tecnológicas, los riesgos asociados a la información que trate, los tipos de datos y sus Titulares. En especial, aquellos que requieran una protección especial, tales como los datos relativos a los menores de edad o los datos sensibles. No obstante, existen algunas directrices aplicadas en entidades del ámbito público que señalan el procedimiento para la anonimización de datos; por ejemplo, las implementadas por el DANE y el Ministerio de Salud, entre otras entidades.

3.1.2. ESTRATEGIAS, INICIATIVAS Y CONSIDERACIONES NORMATIVAS RELEVANTES

3.1.2.1. NECESIDAD DE MANTENER ARCHIVOS DIGITALES

El Gobierno Nacional, en su propósito de hacer más eficiente la administración pública, ha desarrollado la estrategia Cero Papel — relevante en cuanto a la captura y almacenamiento de información — a través de diversas normas jurídicas, como la Ley 812 de 2003 y el Decreto 1151 de 2008 por el cual se establecen los lineamientos generales de la Estrategia de Gobierno en Línea. Cabe resaltar que es directamente la Directiva Presidencial 04 de 2012 la cual fija los lineamientos de la Política Cero Papel en la administración pública, en donde se incluyen las actividades que se deben abstener de realizar los órganos que hacen parte del presupuesto general de la Nación.

Dentro de los componentes de la Estrategia de Gobierno En Línea, el Plan de Cero Papel se enmarca en el componente de Transformación, el cual *“comprende todas las actividades para que las entidades realicen cambios en la manera de operar para eliminar límites entre sus dependencias y con otras entidades públicas, intercambiando información por medios electrónicos haciendo uso del lenguaje común de intercambio de información, liderando o participando en cadenas de trámites en línea. Asimismo, establece las pautas para que la entidad automatice sus procesos y procedimientos internos e incorpore la política de Cero Papel”* (Gobierno Nacional, 2012, p. 8)

Con esta estrategia se busca no solo reducir el consumo de papel en las entidades públicas, sino que también se propende el uso de medios electrónicos y la implementación del intercambio de toda la información relacionada con cadenas de trámites y los requerimientos de otras entidades para el desarrollo de sus funciones. Teniendo en cuenta que la política Cero papel se enmarca en la fase de transformación, las entidades deben realizar cambios operativos para eliminar los límites con otras entidades y organizar sus trámites y servicios alrededor de necesidades de ciudadanos y empresas, lo que implica el desarrollo de cadenas de trámites y el intercambio eficiente de información.

A este intercambio se le llama Interoperabilidad. La interoperabilidad se define como el *“ejercicio de colaboración entre organizaciones para intercambiar información y conocimiento, en el marco de sus procesos de negocio, con el propósito de facilitar la entrega de servicios a ciudadanos, empresas y a otras entidades”* (“Marco Para La Interoperabilidad Del Gobierno En Línea” 2010, p. 37). En Colombia, el modelo de Interoperabilidad vigente se diseñó para apoyar la implementación de la Estrategia de Gobierno en línea que desarrolló el Ministerio TIC, a partir de cuatro componentes: TIC para la gestión, TIC para servicios, TIC para el gobierno abierto y Seguridad y privacidad de la información (MINTIC 2015). Cabe resaltar que el Decreto 1413 de agosto del 2017 incorpora el Servicio de interoperabilidad como uno de los tres servicios digitales ciudadanos (autenticación electrónica y carpeta ciudadana), proponiendo la transformación de este servicio - sin embargo, este aún se encuentra en desarrollo de acuerdo al “Manual de Condiciones de Servicio de Interoperabilidad” del MINTIC y la recientemente creada Agencia Nacional Digital. El Decreto incorporó como deber de los operadores de servicios digitales el principio de responsabilidad demostrada y con éste, la necesidad de la creación de un programa integral de gestión de datos personales para garantizar que los operadores tomen medidas apropiadas, efectivas y verificables para evidenciar el cumplimiento de las normas de protección de datos personales.

Frente a lo anterior, el Ministerio de las Tecnologías de la Información y las Comunicaciones y el Archivo General de la Nación, desarrollaron la Guía No. 5 en el marco de la iniciativa Cero Papel en la Administración Pública, con el objetivo de brindar

⁵ “Digitalización: técnica que permite la reproducción de información que se encuentra guardada de manera analógica (soportes, papel, video, cassettes, cinta, película, microfilm y otros) en una que solo puede leerse o interpretarse por computador”. (ACUERDO 27 DE 2006 (octubre 31). Diario Oficial N.º 46.528 del 31 de enero de 2007. ARCHIVO GENERAL DE LA NACIÓN. Por el cual se modifica el acuerdo número 07 del 29 de junio de 1994. EL CONSEJO DIRECTIVO DEL ARCHIVO GENERAL DE LA NACIÓN DE COLOMBIA.)

orientaciones a las entidades, que les permitan avanzar en la implementación de estrategias que contribuyan a obtener una mayor eficiencia y reducir el consumo de papel, proponiendo la necesidad de implementar procesos de Digitalización⁵ certificada de documentos. En Colombia aún no se cuenta con normas de orden nacional que regulen el tema de digitalización certificada. Sin embargo, con la aplicación de la Ley 527 de 1999, se establece la validez de la información que se emita a través de mensajes de datos. Esta ley se expidió en el país después de las discusiones realizadas en la Comisión de Naciones Unidas para el Derecho Mercantil Internacional (CNUDMI), donde se aprobó la Ley Modelo sobre Comercio Electrónico (Resolución 51/162 de 1996). La ley 527 tiene el propósito de actualizar el ordenamiento jurídico colombiano con base en la realidad del comercio y el uso de nuevas comunicaciones. En este contexto, la firma digital y la firma electrónica fueron reconocidas mediante esta ley. En la actualidad, el artículo 17 de la Ley 527⁶, puede ser la clave para avalar el uso de nuevas tecnologías en los procesos comerciales, en particular la aplicación del *blockchain*. Igualmente, con la Ley 594 de 2007 (Ley General de Archivos), se comienza a desarrollar una serie de cuerpos normativos en busca de incentivar el uso de nuevas tecnologías para mejorar la conservación de archivos y buenas prácticas recogidas a través de la Guía No. 5 precitada.

En relación con lo indicado, para el Gobierno nacional es imprescindible contar con mecanismos idóneos para resguardar y almacenar la información. Por ello, los siguientes cuerpos normativos incentivan y validan el uso de medios electrónicos para el almacenamiento de documentos. En este sentido, en primer lugar, se identifica la Constitución Política de Colombia de 1991 en sus artículos 23 y 74, la cual consagra el derecho que tiene toda persona a presentar peticiones respetuosas a las autoridades por motivos de interés general o particular y a obtener pronta resolución; asimismo, a acceder a documentos públicos, salvo los casos que establezca la ley. En el mismo sentido, el artículo 209 señala que la función administrativa está al servicio de los intereses generales y se desarrolla con fundamento en los principios de igualdad, moralidad, eficacia, economía, celeridad, imparcialidad y publicidad, mediante la descentralización, delegación y desconcentración de funciones. En este marco, con la promulgación de La Ley 80 de 1989, se crea el Archivo General de la Nación como ente rector de la política archivística a nivel nacional que debe velar por la organización, conservación y difusión del patrimonio documental del país. Esta norma señala verdaderamente el primer paso al tratamiento responsable y estructurado a los documentos físicos en los cuales, naturalmente, se resguardan datos de forma oficial. Se crea el Archivo General de la Nación como establecimiento público de orden nacional, el cual, entre otras, tiene las funciones de seleccionar, organizar y conservar divulgar el acervo documental que integre el archivo de la Nación (Art 2, lit. c, *ibidem*).

Complementando la anterior normativa con la Ley 1437 de 2011, se autoriza la utilización de medios electrónicos en el procedimiento administrativo en particular en lo referente al documento público en medios electrónicos, el archivo electrónico de documentos, el expediente electrónico, la recepción de documentos electrónicos por parte de las autoridades y la prueba de recepción y envío de mensajes de datos. En el mismo sentido, el Acuerdo 006 de 2014 y el Acuerdo 003 de 2015 establecen los lineamientos generales para las entidades del Estado en cuanto a la gestión de documentos electrónicos generados como resultado del uso de medios electrónicos de conformidad con lo establecido en el capítulo IV de la ley 1437 de 2011. Finalmente, con la Circular externa 005 de 2012, se dan las recomendaciones para llevar a cabo procesos de digitalización y comunicaciones oficiales electrónicas en el marco de la iniciativa *Cero papel*.

Frente a lo anterior, cabe resaltar que tampoco se ha aprovechado la ventaja de contar con beneficios tributarios, definidos en la Ley 1819 de 2016, lineamientos del Artículo 45 de la Ley 1753, o el Marco de Referencia de Arquitectura Empresarial del Estado Colombiano, para computación y servicios de almacenamiento en nube; un elemento fundamental para el desarrollo de estrategias de Big Data. A pesar del ordenamiento jurídico que favorece el uso de estas tecnologías bajo la Arquitectura de TI de Colombia, se evidencia baja apropiación de esta tecnología.

Cabe resaltar que Colombia, como miembro de diversas organizaciones internacionales como la Unión Internacional de Telecomunicaciones y la Organización Mundial de Comercio, tiene la obligación de propender por un marco legal neutro como Estado miembro, frente al uso de nuevas tecnologías, en aplicación del principio de neutralidad tecnológica y el principio de trato nacional. Lo anterior referente al deber de otorgar a los servicios y a los proveedores de servicio de otros países, miembros de las organizaciones previamente indicadas, un trato no menos favorable que el otorgado a los servicios y a los proveedores de servicios nacionales. Esto al entender que el servicio de almacenamiento en nube es una propuesta tecnológica que permite ofrecer servicios de computación a través de internet; por lo tanto, no podría el Estado desincentivar el uso de estas herramientas tecnológicas (Concepto No.562992 del 10 de septiembre de 2012, MinTic).

Otro aspecto fundamental a tener en cuenta es que en Colombia, a través de la Ley 1581 de 2012, se establece un mecanismo para resguardar la transferencia de datos a terceros para la prestación de un servicio, como lo sería el servicio de almacenamiento en nube. Lo anterior como objetivo de resguardar dicha relación con la suscripción de un contrato de transmisión de datos, en atención al artículo 25 del Decreto 1377 de 2013, el cual fue compilado en el artículo 2.2.2.25.5.2 del Decreto 1074 de 2015, que indica lo siguiente:

⁶ “ARTÍCULO 17. Presunción del origen de un mensaje de datos. Se presume que un mensaje de datos ha sido enviado por el iniciador, cuando: “Haya aplicado en forma adecuada el procedimiento acordado previamente con el iniciador, para establecer que el mensaje de datos provenía efectivamente de éste, o El mensaje de datos que reciba el destinatario resulte de los actos de una persona cuya relación con el iniciador, o con algún mandatario suyo, le haya dado acceso a algún método utilizado por el iniciador para identificar un mensaje de datos como propio”.

“Artículo 2.2.2.25.5.2. Contrato de transmisión de datos personales. El contrato que suscriba el Responsable con los encargados para el tratamiento de datos personales bajo su control y responsabilidad señalará los alcances del tratamiento, las actividades que el encargado realizará por cuenta del responsable para el tratamiento de los datos personales y las obligaciones del Encargado para con el titular y el responsable.

Mediante dicho contrato el encargado se comprometerá a dar aplicación a las obligaciones del responsable bajo la política de Tratamiento de la información fijada por este y a realizar el Tratamiento de datos de acuerdo con la finalidad que los Titulares hayan autorizado y con las leyes aplicables.

Además de las obligaciones que impongan las normas aplicables dentro del citado contrato, deberán incluirse las siguientes obligaciones en cabeza del respectivo encargado:

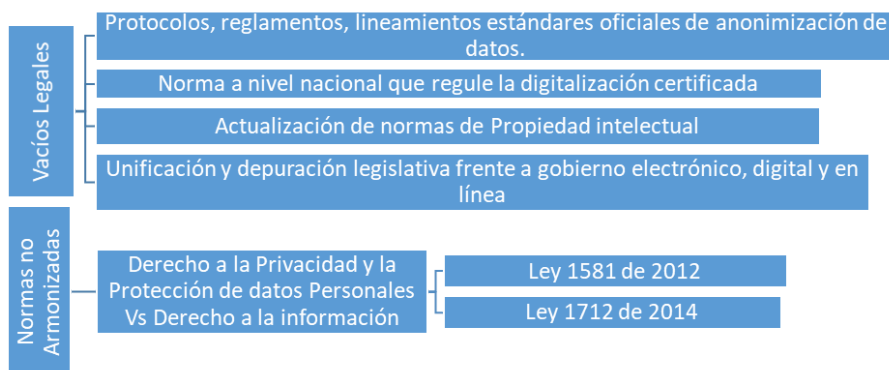
1. Dar Tratamiento, a nombre del Responsable, a los datos personales conforme a los principios que los tutelan.
2. Salvaguardar la seguridad de las bases de datos en los que se contengan datos personales.
3. Guardar confidencialidad respecto del tratamiento de los datos personales”

Como se observa, la suscripción del contrato de transmisión de datos no es facultativa sino deviene de una obligación legal que se debe cumplir. Adicional a ello, es deber del Responsable del tratamiento, según el literal i) del artículo 17 de la Ley 1581 de 2012 “Exigir al Encargado del Tratamiento en todo momento, el respeto a las condiciones de seguridad y privacidad de la información del Titular”, garantía que se cumple a través del Contrato de transmisión de datos.

Por todo lo indicado, se puede concluir que existe en Colombia una serie de regulaciones que favorecen la aplicación de estrategias para la explotación de datos y Big Data. Sin embargo, es necesario la articulación de un cuerpo normativo que supla todos los vacíos jurídicos previamente identificados como la falta de regulación de técnicas de anonimización de datos que favorezcan el acceso a la información de carácter público y/o que favorezca el uso de herramientas tecnológicas específicas, como el servicio de computación en nube. También consideramos necesario que la normativa existente se estandarice y depure respecto de los conceptos de gobierno electrónico, Gobierno en Línea y Gobierno Digital, con el fin de determinar las políticas que en tal materia se encuentran vigentes y aquellas que se encuentran derogadas.

El siguiente diagrama puede ilustrar mejor las conclusiones que anteceden:

Figura 4: Conclusiones sobre la dimensión normativa y ética.



Fuente: Elaboración propia.

3.1.2.1. DIAGNÓSTICO DE LOS FUNDAMENTOS JURÍDICOS DE LOS DATOS EN COLOMBIA:

Una vez conceptualizadas las normas en materia de información y la clasificación legal de datos en Colombia, se logran evidenciar oportunidades de mejoramiento desde la perspectiva jurídica de cara a dotar de seguridad jurídica al desarrollo adecuado de la explotación de datos por parte de las entidades públicas para generar valor económico y social. En esta medida, primeramente, se evidencian las siguientes barreras jurídicas para la explotación de la data en Colombia:

- a. Una dispersión normativa que no goza de un principio armonizador para el tratamiento y divulgación de información en cabeza del estado, lo cual genera un choque entre normas de igual jerarquía.
- b. Los principios de la protección de datos frente al Big data.
- c. La necesidad de actualizar las normas de Propiedad Intelectual para la Analítica Masiva de Datos
- d. Ausencia de lineamientos para la divulgación y anonimización de la información.

Bajo estas cuatro perspectivas se procede a desarrollar las barreras jurídicas evidenciadas así:

a. De las tensiones en las normas de información entre habeas data y acceso a la información pública

El derecho a la protección de datos personales y el derecho al acceso a la información pública son en conjunto la esencia de la teoría jurídica constitucional propia de la información. Son normas de rango estatutario que reglamentan la información pública y su acceso a través de la ley 1712 de 2014 y de otra parte el tratamiento de los datos personales con la ley 1581 de 2012 y sus decretos reglamentarios. Desde la Constitución de 1991, estos derechos fueron incluidos y salvaguardados con la protección y defensa por parte del Estado. Si bien es cierto, el legislador intenta, a través de la normatividad anteriormente descrita, dar un desarrollo completo al tratamiento de datos personales, en virtud de los principios de publicidad, legalidad, finalidad; libertad, veracidad, transparencia, acceso y circulación restringida; seguridad, confidencialidad, buena fe, eficacia, calidad de información, no discriminación; entre otros tantos, la legislación existente no es suficiente. A pesar de que ésta se encarga de definir conceptos importantes respecto al tema específico, enunciar diferentes clases de datos y de información, así como de especificar los deberes y las limitaciones de los responsables del tratamiento de datos y las políticas de su tratamiento; ésta no especifica lo relacionado a los criterios jurídicos que se deben tener en cuenta al momento de divulgar éstos, especialmente en la información pública, ni su relación con técnicas de tratamiento masivo de información tales como el data *analytics* y la estrategia de *Big Data*.

De igual manera, lo concerniente a los datos abiertos u *open data*, entendiendo estos como aquellos datos gestionados por la administración pública y que se encuentran a disposición de la sociedad, de modo que el ciudadano pueda analizar, reutilizar o redistribuir dicha información para generar un beneficio propio o en favor de terceros. Dentro de los principios que rigen la actuación de la administración encontramos como rectores aquellos relacionados con la publicidad y la transparencia y el acceso a la información. El primero indica que toda la información en posesión, bajo control o custodia de un sujeto obligado, es de carácter público y por consiguiente, no podrá ser reservada o limitada sino por disposición constitucional o legal (Ley 1712 de 2014, Art. 2). El segundo, está encaminado a señalar que toda la información en poder de los sujetos obligados se presume pública y por consiguiente éstos están obligados a proporcionar y facilitar el acceso a la misma en los términos más amplios y a través de los medios y procedimientos establecidos por la ley. Ahora, en contraposición, encontramos derechos fundamentales, como los relacionados al derecho a la intimidad personal y familiar, y el habeas data, indicando que el primero, está relacionado con la facultad que tienen las personas para manejar su propia existencia como a su bien les parezca con el mínimo de intromisiones exteriores; y que el segundo, permite a todos los ciudadanos conocer, actualizar y rectificar la información que tengan las diferentes entidades, públicas y privadas y bases de datos dentro de nuestro país (Constitución Política Artículos 15 y 20). El derecho de acceso a la información en poder de las autoridades es hoy en día reconocido como un derecho humano fundamental que debería aplicarse a nivel nacional con base en el principio de máxima divulgación. En éste, se establece la presunción de que toda información es accesible sujeta solamente a un sistema restringido de excepciones.

Sin embargo, el marco jurídico en Colombia se ha desarrollado de manera paralela y desarticulada con la regulación sobre el derecho a la protección de datos personales y el derecho a la información pública, lo que ha generado tensión entre las normas que desarrollan el derecho a la protección de datos, la privacidad e intimidad del titular de información en cabeza del estado aunado a la ausencia de desarrollo de normatividad sobre el tratamiento masivo de información, los procesos de anonimización y el papel del consentimiento del titular en la gestión, reutilización y procesamiento de la data personal. En este sentido, es menester en materia de tratamiento masivo de datos personales contar con lineamientos y directrices a nivel estatal que permitan solventar los conflictos de derechos de igual jerarquía, generando así una seguridad jurídica que evite la discrecionalidad de los funcionarios en las diferentes entidades, que para tales casos hoy en día realizan ejercicios de ponderación de derechos o similares para dirimir dichos conflictos.

b. Los principios de la protección de datos frente al Big data.

En materia de explotación de datos se evidencian barreras importantes a desafiar frente a los principios rectores para el tratamiento de información de carácter personal. Sobre el particular, una vez analizado la ley 1581 de 2012 se evidencian algunos principios relevantes para el presente diagnóstico los cuales se desarrollan a continuación:

- Principio de finalidad – El principio de finalidad se define en la norma como “*El Tratamiento debe obedecer a una finalidad legítima de acuerdo con la Constitución y la Ley, la cual debe ser informada al Titular*”. Este alcance conceptual implica la determinación anticipada y explícita de los fines legítimos para y por lo cual se han recolectado los datos. Este principio tal cual como ha sido interpretado por la sentencia de constitucionalidad 748 de 2011 y como ha sido aplicada por las autoridades establecen barreras y desafíos que cuestiona la posibilidad de reutilización de los datos y/o usos futuros.
- Principio de calidad de datos y el deber de recolección de Datos adecuados, pertinentes y no excesivos – Sobre los principios en comento encontramos en la recolección de datos pertinentes y no excesivos, una yuxtaposición con la necesidad de recabar todos los datos posibles para la consolidación de información suficiente que permita su procesamiento masivo y explotación.
- Principio de Temporalidad – El principio de temporalidad es entendido por la autoridad de la Superintendencia de Industria y Comercio como “*la necesidad de que el dato del titular no podrá ser suministrado a los usuarios cuando deje*

de servir para la finalidad del banco de datos” el cual incide con la necesidad de suprimir información que pierde su pertenencia según las finalidades para las cuales fue recabada lo cual coarta indirectamente la posibilidad de circulación con terceros y reutilización para propósitos afines o no afines en materia de explotación de datos.

EL CONSENTIMIENTO DEL TITULAR EN MATERIA DE PROTECCIÓN DE DATOS:

Adicional a los retos que implican los principios rectores de la protección de datos personales es menester verificar el enfoque, interpretación y aplicación del concepto de consentimiento libre, expreso y anticipado, por cuanto esté a la fecha tiene un protagonismo importante en materia de datos personales. Lo anterior, en tanto las tendencias mundiales se enfocan en ir más allá de un documento, que en la práctica no garantiza el consentimiento verdadero del titular y a que contrario sensu prepondera el *accountability* o la responsabilidad demostrada la cual si implica el verdadero esfuerzo de los responsables y encargados de la información, en una gestión segura, confidencial, leal y respetuosa de los derechos del titular de la información. En este sentido, para el presente diagnóstico vale la pena citar a la agencia española de protección de datos cuando señala que:

“Cuando hablamos de Big Data es especialmente relevante el hecho de que el consentimiento debe aplicarse a un contexto determinado, así como el hecho de que si el fin para el que se utilizarán los datos cambia, podría ser necesario volver a recabar el consentimiento.”

Esto a todas luces es una barrera importante para la explotación de la información,

“Y ello porque, precisamente, el valor del Big Data reside en que la nueva información que se crea permite dar nuevos usos a los datos. Es precisamente en estos usos secundarios donde reside el potencial del Big Data. Esta forma de concebir el consentimiento obligaría a que cada vez que se descubra un nuevo uso para los datos, el responsable debería volver a pedir el consentimiento a cada uno de los individuos cuyos datos estén siendo tratados por segunda vez. Esto, en muchas ocasiones, podrá ser técnicamente inviable, por no decir que las empresas no podrían asumir los costos.”⁷

En suma, a la fecha el ordenamiento jurídico colombiano enfocó la protección a la privacidad y el habeas data en el consentimiento del titular, lo cual debe evolucionar, buscando un enfoque al principio de transparencia y responsabilidad demostrada que centre la protección de los derechos de los individuos en el *“accountability”* del tratante de la información y no en el consentimiento. Lo anterior, permitirá un desarrollo individual tanto del Estado en sus políticas de gestión de la información, el compromiso con la seguridad de la información, la disponibilidad de la información y sobre la garantía de los derechos del titular; así como de los usuarios de la información frente a su debida diligencia en la recolección y procesamiento de la información que tiene carácter personal, así como su análisis y procesamiento masivo. En este sentido *“a través de normas robustas y adaptables que sean capaces, por una parte, de responder a la evolución social, tecnológica, económica y jurídica, lo que pasa por impulsar también la autorregulación, y, por otra parte, que faciliten también la innovación.”⁸*

Esta actualización y re-enfoque de la protección de datos llevará a dar un paso definitivo frente a esquemas actuales alineados a la *“privacidad por consentimiento”*, de manera que la privacidad y la protección de datos estarán basadas en la transparencia y la responsabilidad demostrada de los tratantes de la información.

c. Actualización de las normas de Propiedad Intelectual para la Analítica Masiva de Datos

Una vez analizadas las normas de derechos de autor en Colombia y la Comunidad Andina de Naciones se puede verificar que en Colombia no se cuenta con una legislación especial para las bases de datos, a diferencia de Europa, Reino Unido y Japón, entre otros que tienen un régimen de excepciones para la explotación masiva de datos. Sin embargo, se reconoce en el análisis legislativo que los derechos de autor en Colombia sí reconocen que las bases de datos que por la selección o disposición del contenido pueden llegar a constituir obras originales, pueden ser susceptibles de protección. Teniendo en cuenta lo anterior, no todas las bases de datos son susceptibles de ser protegidas por derechos de autor pues deben recoger el sello personal del autor y distinguirse de otras por la selección o la disposición de sus contenidos. En este sentido es importante aclarar que la información individualizada, y/o los datos que hacen parte de una base de datos desde luego no son susceptibles a ser protegidos por los derechos de autor, sino por la estructura. En este orden, es el resultado del Big Data el que es susceptible de ser protegido por la propiedad intelectual, e incluso de ser entendido como producto. Desde luego, son susceptibles de información tanto las herramientas para la explotación masiva de datos tales como software, invenciones y herramientas y por otra parte la estructura de las bases de datos.

Finalmente, en la divulgación de información en cabeza del estado, así como de los resultados de operaciones de Big Data es menester contar con criterios uniformes de orden estatal para las licencias que tendrán dichas producciones cuando así lo ameriten.

d. Ausencia de lineamientos para la divulgación y anonimización de la información.

⁷ Agencia Española de Protección de Datos, recuperado de <https://www.aepd.es/media/premios/big-data.pdf>

⁸ Recio Gayo Miguel, BIG DATA: HACIA LA PROTECCIÓN DE DATOS PERSONALES BASADA EN UNA TRANSPARENCIA Y RESPONSABILIDAD AUMENTADAS recuperado en https://derechoytics.uniandes.edu.co/components/com_revista/archivos/derechoytics/ytics260.pdf

No existen protocolos o marcos unificados que especifiquen el proceso y las técnicas de anonimización de datos, proceso esencial para garantizar la privacidad y el derecho a la intimidad de las personas, así como el buen uso de datos en cuanto a la explotación de estos. Si bien existen entidades como el DANE que han generado buenas prácticas frente a implementación de técnicas de atomización de datos, estas no han sido replicadas por otras entidades públicas, pues son protocolos definidos por los diferentes equipos temáticos de cada entidad, por ejemplo, acuerdo a funciones estadísticas específicas.

ANÁLISIS ÉTICO

3.1.3. CÓDIGOS DE ÉTICA

Respecto a los códigos de ética, la mayoría de las entidades entrevistadas en el estudio cualitativo de esta consultoría, manifiestan contar con unas normas de seguridad de la información, más no con Códigos de Ética (Coldeportes, DIMAR, Fondo de Adaptación, Ministerio de Agricultura, Dian, Función Pública y Ministerio de Salud). De esta manera, las entidades

Tabla 2: Panorama de las entidades respecto a los códigos de ética.

Entidad	Descripción
Coldeportes	Política de seguridad de la información; Interactúan con lo que establece la Función Pública en cuanto a códigos y lineamientos éticos del manejo de la información y de los datos generados.
Colombia Compra Eficiente	s
DIMAR	Política de seguridad informática.
Fondo de Adaptación	Política de seguridad para salvaguardar la información y el buen manejo de la misma.
Policía Nacional	
Ideam	
Ministerio de Agricultura	Política de seguridad establecida.
DIAN	Ética frente al manejo de la información que todos al interior de la entidad deben firmar; Reuniones para acuerdos de confidencialidad y protocolos para el uso de datos; Sin una orden judicial no entregan información de terceros.
Función Pública	Política de seguridad establecida.
Ministerio de Hacienda	
Ministerio de Trabajo	
Ministerio de Salud	La guía de Habeas Data; El comité de seguridad de la información.

Fuente: Elaboración propia.

buscan promover un sentido de responsabilidad en los funcionarios hacia el buen manejo de los datos. Al indagar por las áreas que deberían responsabilizarse por la gestión de datos se encontró que la mayoría considera que este debería ser un tema transversal de todas las áreas con soporte del área de TI. La tabla 2 resume las respuestas de las entidades encuestadas.

Aun cuando en Colombia existen Códigos de Ética y Buen Gobierno, estos no incluyen directrices sobre manejo de datos e información. Se han delineado normas de gobierno corporativo, pero hace falta delinear la ética para el sector privado y civil. De hecho, hay poco escrito sobre el manejo ético de datos e información en Colombia, por lo que es necesario desarrollar un marco ético para las entidades públicas o comités de ética, y códigos voluntarios de ética para el sector privado. Esto, con el objetivo de identificar las implicaciones e impacto que pueda generar el desarrollo avanzado de Big Data en la sociedad.

3.2. DIMENSIÓN 2: ORGANIZACIONAL E INSTITUCIONAL

La Dimensión Organizacional e Institucional busca entender los elementos existentes en el marco institucional colombiano que permiten el gerenciamiento y la explotación de datos en y desde el sector público. Muchos países alrededor del mundo han adoptado estrategias para la explotación de datos y la digitalización del estado, tal como ha sido el caso de Colombia, como se entiende a partir de la evaluación del marco normativo y ético. Sin embargo, el estado colombiano aún cuenta con una baja institucionalidad en cuanto a datos, si se compara con países como Estados Unidos, y su “Iniciativa de investigación y desarrollo de Big Data”, o con Estonia y Holanda. A continuación, se describe el marco institucional actual de Colombia en cuanto al manejo y gerencia de datos en el sector público. Esta sección cubre un análisis de actores relevantes y un diagnóstico sobre el volumen de datos existentes en las entidades colombianas, la gerencia de datos, el acceso a información pública y cultura de datos, e interoperabilidad.

3.2.1. ESTADO ACTUAL DE LAS ENTIDADES RESPECTO AL APROVECHAMIENTO DE DATOS

Para entender la situación actual de la institucionalidad en cuanto a la gestión y administración de datos en el Estado colombiano, se desarrolló un ejercicio de *web scraping* con datos de la página del Departamento Administrativo de la Función Pública. El propósito del ejercicio era evaluar el número de funciones asociadas al levantamiento, procesamiento y coordinación de información en las diferentes entidades del Estado colombiano. Al realizar este ejercicio, se obtuvieron todas las normas asociadas al funcionamiento de cada entidad de acuerdo a su respectivo manual de funciones. Luego, estas fueron clasificadas según el número de funciones identificadas asociadas al levantamiento, procesamiento y coordinación de información, para lograr identificar los actores principales en esta función. El criterio de la metodología para identificar las funciones relevantes a este análisis se centró en identificar palabras claves, entre estas información, sistemas de información, producción de información – entre otras – con el *web scraping*, para luego verificar de manera manual que los resultados del ejercicio verdaderamente reflejaban las normas relevantes en cuanto a manejo y explotación de datos. Este ejercicio da una primera mirada a la organización institucional del Estado colombiano en cuanto a la explotación de datos y Big Data.

RESULTADOS DEL EJERCICIO DE WEB SCRAPING

La Tabla 3 describe la estructura del Estado colombiano según niveles jerárquicos, una división preliminar que permite entender las diferentes maneras de clasificar a las entidades del Estado colombiano. Se puede observar que el Estado colombiano, según información del Departamento Administrativo de la Función Pública, comprende 33 sectores y 332 entidades, de las cuales el 57,83% son descentralizadas. Según tipo de entidad, el 32,54% son entidades adscritas y el 28,31% vinculadas. Si se hace un análisis de acuerdo con la naturaleza de la entidad, se encuentra que el 20,18% de las entidades son Corporaciones Autónomas Regionales, seguido de un 18,07% de establecimientos públicos.

Tabla 3: Estructura del Estado Colombiano - 2016

Criterio	N	Participación total
Sectores del Estado colombiano	33	100 %
Entidades	332	100 %
Según Organización del Estado	332	100 %
Centralizada	105	31,62 %
Descentralizada	192	57,83 %
Otros*	1	0,003 %
N/A	34	10,24 %
Según Entidad	332	100 %
Adscrita	118	35,54 %
Descentralizada Indirecta	4	1,2 %
Departamento Administrativo	5	1,5 %
Independiente	1	0,003 %
Ministerio	16	4,81 %
Órgano Autónomo	67	20,18 %
Vinculada	94	28,31 %
N/A	27	8,13 %
Según naturaleza de la entidad	332	100 %
Agencia Estatal de Naturaleza Especial	3	0,009 %
Corporación Autónoma Regional	67	20,18 %
Corporación de Investigación	1	0,003 %
Corporación e Institución de Investigación	5	1,5 %
Corporación de Partición Mixta	3	0,009 %
Ente Universitario Autónomo	19	5,72 %
Empresa de Servicios Públicos Domiciliarios	19	5,72 %
Empresa Industrial y Comercial del Estado	19	5,72 %
Empresa Social del Estado	4	1,2 %
Establecimiento público	60	18,07 %
Fondo Cuenta con PJ	5	1,5 %
Instituto	1	0,003 %
Naturaleza Especial	6	1,8 %
Organismo Principal	17	5,12 %
Organismo Sectorial de Asesoría y Coordinación	3	0,009 %
Sociedad de Economía Mixta	24	7,22 %
Sociedad de Naturaleza Especial	2	0,006 %
Sociedad Filial de Empresa Industrial y Comercial del Estado	1	0,003 %
Superintendencia con PJ	13	3,91 %
Superintendencia sin PJ	2	0,006 %
Unidad Administrativa Especial con PJ	25	7,53 %
Unidad Administrativa Especial sin PJ	8	2,4 %
N/A	25	7,53 %

*Entidad regulada por el código civil y la legislación de ciencia y tecnología

Fuente: Departamento Administrativo de la Función Pública (2016), cálculos propios

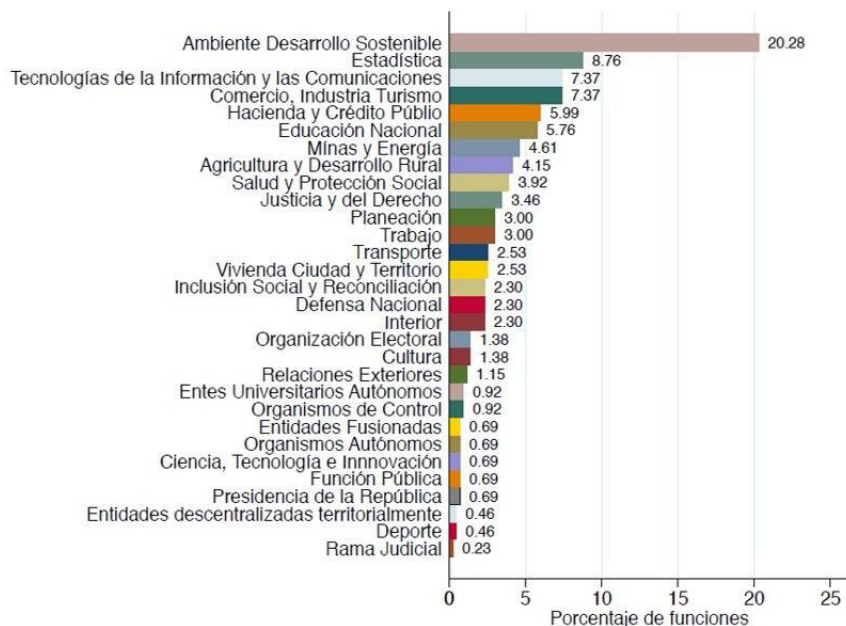
Por otro lado, la Tabla 4 describe las funciones asociadas al levantamiento, procesamiento y coordinación de información según el tipo de norma. Según los cálculos, existen 434 normas, donde 67 corresponden a leyes, 323 a decretos y el resto son otro tipo de normas.

Tabla 4: Funciones asociadas al levantamiento, procesamiento y coordinación de información, según tipo de norma

Variable	N
Dummy Ley	434
NO Ley (= 0)	367
Ley (= 1)	67
Dummy Decreto	434
NO Decreto (= 0)	111
Decreto (= 1)	323
Dummy Acuerdo	434
NO Acuerdo (= 0)	433
Acuerdo (= 1)	1
Dummy Norma orgánica Derogada	434
NO Derogado (= 0)	337
Derogado (= 1)	97

Fuente: Departamento Administrativo de la Función Pública, 2016. Cálculos propios.

Figura 5: Distribución del número de funciones asociadas al manejo de información según sector



Fuente: Departamento Administrativo de la Función Pública, cálculos propios

A partir de los resultados de este ejercicio, se pudo obtener también la distribución del número de funciones asociadas al manejo de información por cada sector. El sector de ambiente y desarrollo sostenible concentra el mayor número de normas asociadas al levantamiento, procesamiento y control de información, con una participación del 20,28%, seguido por el sector estadístico con 8,76% y el sector TIC con el 7,37%.

La Figura 6 presenta el valor absoluto del número de funciones asociadas a información por sector. Sin embargo, para identificar una relación clara, se normalizó por el número de funciones totales por sector, relación visualizada en la Figura 7.

Figura 6: Número de funciones asociadas de al manejo de información, según sector

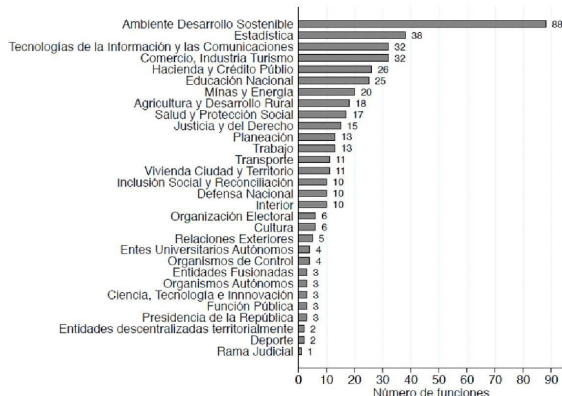
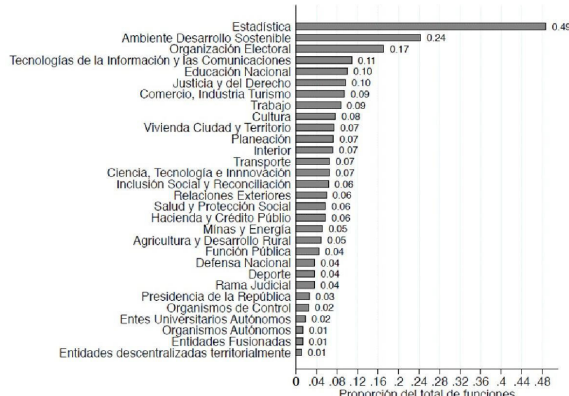


Figura 7: Proporción total funciones asociadas



Este análisis arroja entonces que el sector estadístico presenta la mayor cantidad de funciones en cuanto al manejo de información con un 49% de funciones asociadas. De hecho, el sector estadístico logra doblar el sector de ambiente y desarrollo sostenible. A nivel nacional, en promedio, de las funciones por sector, el 11% de estas según su concentración normativa, son asociadas a información. En las figuras arriba, por ejemplo, vale la pena destacar la baja proporción de la rama judicial – un sector que naturalmente produce altos volúmenes de datos. Cabe resaltar, sin embargo, que el hecho de que tengan una baja proporción, no quiere decir que no estén produciendo datos, simplemente que la gestión, producción y manejo de datos no se encuentra articulado en su manual de funciones. De acuerdo a esto, no se hace obligatorio priorizar la producción de información coordinada – quizá por tema – para que las diferentes entidades de este sector produzcan información replicable y de calidad.

De igual manera, si se divide el análisis entre entidades centralizadas o no, se observa que el 67.51% de funciones en cuanto al manejo de datos se concentra en entidades descentralizadas, que son aquellas con personería jurídica, autonomía administrativa y patrimonio propio. Esta descripción se hace relevante para identificar, por una parte, la oferta de información en el país y los sectores más relevantes para una estrategia de Big Data, de acuerdo a la concentración de funciones relacionadas con información. También sirve para identificar vacíos y oportunidades en los marcos regulatorios en cuanto a las funciones de recolección, producción, almacenamiento y difusión de información. De manera similar, el análisis según tipo de entidad arroja que las entidades adscritas concentran el 53,23% de las funciones asociadas a la información.

De este ejercicio de web scraping, se puede concluir, en primera instancia, la ausencia de funciones de datos en entidades que naturalmente levantan datos, así este no sea su enfoque principal, como podría ser el sector de salud o minas y energía. Es necesario revisar si verdaderamente se prioriza la gestión y levantamiento de información, de lo contrario, es importante asignarles estas funciones a las entidades y lograr una manera de armonizar estas funciones entre ministerios, entidades adscritas y otras.

3.2.2. RECONOCIMIENTO DE DATOS EXISTENTES EN LAS ENTIDADES COLOMBIANAS

Actualmente, Colombia no cuenta con un inventario detallado y centralizado de todos los datos que producen las instituciones y organizaciones del país. Una de las iniciativas ya planteadas, que funciona como inventario y repositorio de datos es el portal www.datos.gov.co del MinTIC. Esta plataforma permite que cualquier entidad pública suba datos al portal. En este momento, sin embargo, no existe un plan de actualización y mantenimiento, o un estudio de demanda por parte de los usuarios sobre los requerimientos de estos datos; tampoco existe una hoja de ruta sobre los nuevos datos publicados.

Aun cuando este portal tiene la posibilidad de ser el repositorio de datos para las entidades públicas, se observa que el ejercicio que más se aproxima a un inventario comprensivo de datos es el inventario de operaciones estadísticas desarrollado por el DANE. Una operación estadística es la “[...] aplicación de un proceso estadístico sobre un objeto de estudio que conduce a la producción de información estadística” (Plan Estadístico Nacional, 2017-2022). Si bien esta definición hace referencia a una operación estadística y no netamente a datos, el identificar la concentración de operaciones estadística del país permite una aproximación a quienes tienen mayores roles asociados al manejo de datos. La Tabla 5 presenta las operaciones estadísticas que las entidades de Colombia manejan. Naturalmente, el DANE concentra el mayor número de operaciones estadísticas del país, seguido por el Banco de la República.

Tabla 5. Número de operaciones estadísticas por entidad del Sistema Estadístico Nacional

ENTIDAD	OPERACIÓN ESTADÍSTICA
Departamento Administrativo Nacional de Estadística - DANE*	108
Banco de la República de Colombia	34
Instituto de Hidrología, Meteorología y Estudios Ambientales - Ideam	19
Unidad de Planeación Minero Energética - UPME	15
Dirección General Marítima - Dimar	14
Superintendencia de Servicios Públicos Domiciliarios	12
Departamento Administrativo de la Función Pública - DAFP	11
Ministerio de Salud y Protección Social*	11
Instituto Colombiano Agropecuario - ICA	10
Aeronáutica Civil – Aerocivil	9
Ministerio de Hacienda y Crédito Público	9
Ministerio de Hacienda y Crédito Público	9
Superintendencia Financiera de Colombia	9
Departamento Administrativo de Ciencia, Tecnología e Innovación – Colciencias	8
Ministerio de Comercio, Industria y Turismo	7
Ministerio de Defensa Nacional*	7
Ministerio de Educación Nacional	7
Observatorio Colombiano de Ciencia y Tecnología - OCYT	7
Otras entidades con 2 a 6 operaciones	178
Otras entidades con 1 Sola operación	26

Fuente: DANE- DIRPEN. Inventario de Oferta de Operaciones Estadísticas (actualizado a 31 de marzo de 2017)
Nota: * Entidades que realizan algunas de sus operaciones estadísticas en conjunto con otras entidades

Por otro lado, el Plan Estadístico Nacional (PEN) es la hoja de ruta del país para la producción estadística en el periodo 2017-2022. El PEN es un instrumento de política que establece los lineamientos estratégicos para la captura y oferta de información estadística de Colombia – y por extensión, de datos – elemento necesario para el desarrollo de una estrategia de Big Data. En este momento, las acciones del PEN atienden en mayor medida las necesidades de medición de indicadores provenientes de los Objetivos de Desarrollo Sostenible (ODS), los requerimientos de la Organización para la Cooperación y Desarrollo Económico (OCDE) y las necesidades puntuales del DANE, en materia de Cuentas Nacionales e indicadores oficiales.

Por lo tanto, ni datos.gov.co, ni el PEN cuentan con el mandato, función o capacidad para realizar procesos comprehensivos de identificación de información al interior del gobierno colombiano. No hay clara evidencia de que en Colombia actualmente exista un inventario detallado y centralizado de todos los datos, o si quiera, de las características de los datos, que producen las entidades del país, mucho menos del sector privado. En construcción de la estrategia de explotación de datos y Big Data es esencial ir más allá de ejercicios de oferta de información para avanzar en procesos de identificación de demanda de información, para poder generar valor agregado y permitir la explotación de datos, tanto por el sector privado, como por la ciudadanía en general.

3.2.3. GERENCIA DE DATOS EN COLOMBIA

El Plan Estadístico Nacional, descrito en la sección anterior, es relevante para el análisis sobre el estado de producción de información en Colombia y para la gerencia de datos. A través del Decreto 1743 de 2016, se define el Sistema Estadístico Nacional (SEN) como “el conjunto articulado de componentes que, de manera organizada y sistemática, garantiza la difusión de las estadísticas oficiales que requiere el país”. Este está compuesto por las entidades productoras y/o difusoras de datos estadísticos o responsables de registros administrativos. El DANE, como ente rector del SEN, es el encargado de generar, actualizar y presentar el PEN ante los miembros del SEN y al Consejo Asesor Nacional (CANE), con el objetivo de garantizar la producción y difusión de estadísticas oficiales.

Para el periodo del 2017 al 2022, como está consignado en el documento aprobado (Sistema Estadístico Nacional, 2017), el PEN busca “(...) definir un conjunto de objetivos relacionados con el fortalecimiento de la producción y difusión de estadísticas en el SEN (...)”. De igual forma, plantea que “(...) para alcanzar dichos objetivos se construyeron nueve estrategias que articulan la actividad estadística asociada a entidades encargadas de su implementación. Las acciones parten de acuerdos institucionales entre sus miembros, aprovechando recursos existentes e innovadores, mejorando continuamente los procesos de producción y difusión, capacitando el recurso humano del Sistema, y articulando las entidades de orden nacional con las entidades territoriales del país”.

El PEN propone como principales problemas para la producción de información estadística, los siguientes:

1. Falta de comunicación y conocimiento común sobre los proyectos que involucran actividades estadísticas.
2. Las entidades productoras, con sistemas de información, generalmente no cuentan con una comunicación efectiva entre ellos, lo que genera una duplicidad de esfuerzos en la producción estadística.
3. Falta de coordinación, desde las entidades nacionales, sobre los requerimientos de información para los entes territoriales, lo cual genera duplicidad de esfuerzos, recursos y tiempos.

En un reporte de la OCDE se confirman los problemas planteados en el PEN: “las estadísticas territoriales colombianas sufren de una falta de coherencia y coordinación. Aunque la variedad de sistemas de producción y difusión de las estadísticas nacionales y territoriales añade un elemento de riqueza al sistema nacional de información, (...) esto podría plantear algunas dudas sobre la coherencia de las estimaciones estadísticas, dada las posibles discrepancias entre las estimaciones generadas por diversos departamentos para el mismo indicador económico” (OECD 2014).

De acuerdo con los problemas identificados en el PEN, para la creación de un SEN coordinado y eficiente, y bajo los estándares de calidad las estadísticas requeridas en el país, es necesario el cumplimiento de 6 objetivos:

1. Mantener la producción de la información estadística del SEN.
2. Facilitar el acceso y disponibilidad de la información estadística nacional y territorial.
3. Elevar la calidad de las estadísticas producidas en el SEN.
4. Ampliar la oferta de información estadística nacional y territorial del SEN para dar respuesta a compromisos internacionales como ODS y OCDE.
5. Promover la gestión del conocimiento y la innovación para contribuir con el fortalecimiento de la capacidad estadística.
6. Fortalecer la articulación entre los miembros del SEN para la producción y difusión de la información estadística.”

Las estrategias del PEN que ayudan a generar valor sobre datos, son también pertinentes en el análisis de esta dimensión, debido a la labor en gestión y administración de datos, aunque el enfoque de estos sean en operaciones estadísticas. Estas estrategias son:

- Promoción de la difusión y del acceso a la información estadística: busca promover la implementación y aplicación de estándares de difusión y de acceso. De igual forma, unificar en los integrantes del SEN un lenguaje común que facilite el intercambio de la información, la apertura de datos, su uso y reutilización. El estándar propuesto es el denominado *Statistical Data and Metadata eXchange* (SDMX). Esta estrategia cuenta con 3 componentes principales:
 - Fomentar la difusión de las estadísticas oficiales y su documentación asociada.
 - Apoyar la interoperabilidad entre sistemas de información que requieran compartir estadísticas.
 - Fomentar el acceso e intercambio de microdatos.
- Identificación y promoción del aprovechamiento estadístico de registros administrativos: esta estrategia plantea usar los registros administrativos como fuentes de información, lo que impulsa la implementación de sistemas y políticas internas en entidades, vinculados o no al SEN, de diferentes sectores tanto públicos y privados. Como se plantea en sus componentes (identificación, fortalecimiento y gestión de datos), para cumplir con las metas es necesario una coordinación y cooperación entre entidades que permita la apropiación de una cultura de datos dentro de cada entidad.

Lo propuesto bajo estas estrategias representan avances también en los procesos de gerencia y producción de información para el país, aun cuando no existan procesos similares para datos no estadísticos. Ya que existen procesos para el fortalecimiento de la producción estadística, estrategias en cursos, y hojas de ruta, debe existir también una estructura similar para los datos en general, cementados en acuerdos institucionales y entre diversos actores.

3.2.4. ACCESO A LA INFORMACIÓN PÚBLICA Y CULTURA DE DATOS

En el 2011, se creó la Secretaría de Transparencia que cuenta con un Observatorio de Transparencia y Anticorrupción donde se recopilan, analizan y visibilizan los indicadores de transparencia y anticorrupción del país (Observatorio Transparencia y Anticorrupción, 2018). En el 2014, como se mencionó anteriormente, el Congreso de la República firmó la ley 1712 para “regular el derecho de acceso a la información pública que tienen todas las personas, los procedimientos para el ejercicio y la garantía del derecho fundamental, así como las excepciones a la publicidad de la información pública” (Procuraduría General de la Nación, 2014). Esta ley ayuda a la Secretaría de Transparencia a ejercer mejor sus funciones y a involucrar a todas las personas que pueden ejercer el Derecho de Acceso a la Información Pública, tornándose como una de las principales estrategias de la secretaría.

Esta estrategia ha permitido promover la creación de una cultura de datos y consumo de información pública, lo que a su vez ha establecido la necesidad en las entidades gubernamentales de implementar políticas internas para organizar y poner a disposición de los ciudadanos la información pública de sus ciudadanos (Secretaría de Transparencia n.d.). La ley 1712 también generó iniciativas multisectoriales para la apertura gubernamental y con ello, la Declaración de Compromisos por un Estado Abierto (Comisión Nacional de Moralización, 2017). Los avances en la apertura gubernamental para el 2017, se encuentran en el informe de Open Government Partnership (OGP):

“Colombia creó un plan de acción en cercana colaboración con la sociedad civil. Incluye cinco compromisos estelares en lo referente al acceso a servicios públicos para personas con discapacidad, acceso a la justicia, racionalización de trámites, transparencia y rendición de cuentas en el Consejo de Estado y memoria histórica del conflicto armado. El segundo año de implementación y el desarrollo del siguiente plan se enmarcan en un escenario de posconflicto. Los esfuerzos en gobierno abierto pueden afianzarse con la coordinación de lo nacional y subnacional, para promover el diálogo y participación en este contexto.”

Debido a las iniciativas de alto impacto para fomentar la transparencia, participación ciudadana, rendición de cuentas, uso de tecnologías e innovación; la OGP reconoció a Colombia por ser el único país, miembro del organismo internacional, con ocho compromisos transformadores en su plan de acción. Dentro de las iniciativas, las que más influyen en la apropiación de cultura de datos tanto para las entidades como para los ciudadanos, según la Alianza Gobierno Abierto Colombia (2016) son:

- **Centro de Relevo y Convertic:** herramienta tecnológica para generar acceso a la información pública y a los servicios que presta el Gobierno por parte de la población con discapacidad auditiva y visual.
- **Legalapp:** herramienta virtual para guiar a los ciudadanos en la realización de trámites y acceder a servicios de la justicia. En la actualidad, esta herramienta incluye nuevos contenidos, servicios y funcionalidades, de acuerdo con los requerimientos de los usuarios.
- **La identificación de trámites y servicios para la racionalización y automatización:** con participación ciudadana y la integración de la oferta de trámites y servicios en un mismo espacio físico, se logró que Colombia contará con cinco nuevos trámites accesibles y usables en línea de acuerdo con las necesidades y expectativas ciudadanas.
- **Archivo de Derechos Humanos y Memoria Histórica:** esta herramienta es una de las medidas para asegurar el derecho a la verdad, la justicia y la reparación a la población víctima del conflicto armado, el cual ha puesto al servicio más de 240.000 documentos de archivo o colecciones documentales que retratan el conflicto y las vivencias de la población víctima en todo el país.

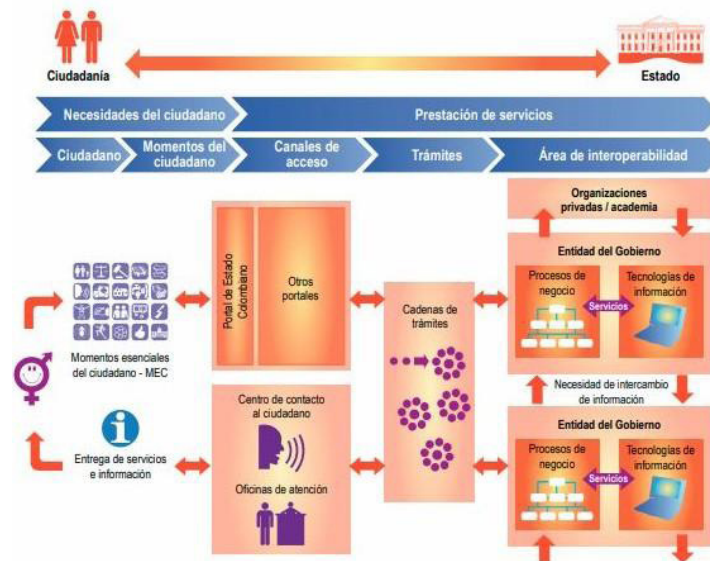
Todas estas iniciativas, junto con la Ley de Transparencia, muestran cómo las entidades están empezando a integrar políticas internas para la gerencia de datos y la creación de herramientas que permitan tomar decisiones basadas en datos, independiente si es directamente un resultado de las normas. Esto genera presión en las entidades para implementar políticas de gestión y administración de datos, al igual que un ambiente propicio para crear una sociedad con cultura de datos.

Cabe resaltar también el rol que en los últimos meses lleva liderando el MinTIC para empoderar la transformación digital en Colombia, en el marco de la nueva política de Gobierno Digital. Los esfuerzos institucionales para afianzar la cultura de datos y profundizar conocimientos que permitan implementar procesos arraigados en el buen uso de los datos se ven reflejados en propuestas como la aplicación “Soy CIO” y la ley de modernización del sector TIC, aun cuando estos estén en etapa temprana de implementación.

3.2.5. INTEROPERABILIDAD

Al ser una de las características transversales para la coordinación de una estrategia de datos, la interoperabilidad será analizada en varias dimensiones; bajo esta, se analiza desde el Marco de Interoperabilidad, lineamiento vigente desde el 2010, e inicialmente un componente de la estrategia de Gobierno En Línea del MinTIC. Este marco “nace como un conjunto de principios, políticas y recomendaciones que busca facilitar y optimizar el ejercicio de colaboración entre organizaciones para intercambiar información y conocimiento, en el marco de sus procesos de negocio, con el propósito de facilitar la entrega de

Figura 8: Modelo detallado para la entrega de servicios e información al ciudadano



Fuente:MinTIC, 2018

En este contexto, se define la interoperabilidad como “la acción, operación y colaboración de varias entidades para intercambiar información que permita brindar servicios en línea a los ciudadanos, empresas y otras entidades mediante una sola ventana de atención o un solo punto de contacto. Es decir, es la forma de ahorrarle a la gente los desplazamientos de un lugar a otro a la hora de realizar un trámite y de hacer el proceso menos engorroso” (MINTIC 2010). Aumentar la interoperabilidad tiene como objetivo crear un Estado funcional que opera como una sola entidad eficiente, convirtiendo la interoperabilidad en una consideración esencial en cuanto a la gestión y el manejo de la información (MINTIC 2018).

La Figura 8 detalla el modelo para la entrega de servicios e información al ciudadano, al igual que los elementos necesarios para la coordinación y el intercambio de información entre entidades.

El Marco de Interoperabilidad se puede entender en cinco dominios principales, estos son:

- Político-legal: Políticas y normas que permiten el intercambio de información.
- Sociocultural: Competencias requeridas en las entidades para el intercambio de información y cooperación entre entidades.
- Organizacional: Interacción requerida entre entidades reflejada en misiones, políticas, procesos de negocio y mecanismos de prestación de servicios de intercambio de información.
- Técnico: Aspectos técnicos requeridos para la interconexión de sistemas de intercambio de información.
- Semántico: Unifica el significado de la información en el momento de intercambiar de datos.

Este marco está basado en un modelo de madurez escalonado – donde se requiere haber alcanzado niveles anteriores para alcanzar otro nivel – que permite la implementación de forma progresiva de los servicios de intercambio de información. Estos niveles son:

- Nivel 1 (inicial): Identificación de los conjuntos de datos que requiere intercambiar y validar el lenguaje común de intercambio de información para el conjunto de datos.
- Nivel 2 (definido): Uso correcto de lenguaje común de intercambio de información, en al menos un servicio activo del catálogo de servicios.
- Nivel 3 (Optimizado): Contar con todos los servicios de intercambio de información del catálogo de servicios, utilizando el lenguaje común del conjunto de datos.

Como herramienta para facilitar el manejo y la gestión de los datos, el Marco de Interoperabilidad descrito permite evaluar los avances dentro de las diferentes entidades del estado. De igual manera facilita la creación de servicios de intercambio de información al proporcionar recomendaciones, protocolos, estándares y guías metodológicas.

3.2.5.1. INTEROPERABILIDAD SEGÚN ENTIDADES ENCUESTADAS

Las entidades encuestadas reconocen que la interoperabilidad agiliza los procesos y servicios a los ciudadanos, permite contar con información actualizada y real y permite unificar los datos y registros de varias entidades. Sin embargo, se reconoce también que en el momento hay limitaciones en cuanto a las capacidades técnicas y los recursos que facilitan la interoperabilidad. Por ejemplo, se encontró que algunas entidades no han podido desarrollar procesos interoperabilidad, como lo son el Fondo de Adaptación y la DIAN. En caso de hacerlo, cuentan todavía con importantes retos respecto al nivel de conocimiento de la plataforma de interoperabilidad del MinTic (DIMAR, Coldeportes, e IDEAM). Las entidades encuestadas coincidieron en que la falta de cultura de datos se evidencia frecuentemente en los obstáculos jurídicos que existen entre las diferentes entidades, la falta de calidad en los datos intercambiados, así como la renuencia a compartir información.

Las entrevistas indican también que aun cuando se realizan capacitaciones y talleres para la Interoperabilidad, no se hace un seguimiento al uso efectivo de los datos. El que en este momento no existan métricas ni información del uso de los sistemas de intercambio de datos dificulta el proceso de validar el impacto que iniciativas como el Marco de Interoperabilidad, o dichos talleres, tienen sobre el uso de datos en las entidades públicas. Cabe resaltar que la mayoría de los entrevistados consideran que la entidad que debería encargarse de controlar todo lo relacionado con el intercambio de información debe ser el Departamento Nacional de Planeación.

3.2.5.2. EVALUANDO LA INTEROPERABILIDAD EN COLOMBIA

De acuerdo a lo anterior, desde el 2010, Colombia ha trabajado en estrategias e implementación de sistemas de interoperabilidad — Gobierno En Línea, Gobierno Digital, Arquitectura de TI, Marco de Interoperabilidad, Interoperabilidad como Servicio Ciudadano Digital — sin embargo, es difícil identificar claramente el uso masivo de dichos sistemas. Desde el punto de vista de un ecosistema de datos, el tener sistemas interoperables sin que estos sean usados, como lo es el Marco de Interoperabilidad, no genera valor alguno. Una de las barreras identificadas en cuanto al bajo uso de datos, así estos hayan sido efectivamente compartidos, es la falta de atributos de calidad en los sistemas de intercambio de datos. Se puede observar en datos.gov.co, que una búsqueda simple arroja conjuntos de datos que carecen de descripción de variables o columnas, uno de los requerimientos básicos para asegurar la calidad de los datos y el uso de estos. Desde el 2010 no se ve un avance claro al respecto, ni métricas sobre los sistemas de interoperabilidad.

3.2.6. EVALUANDO LA INTEROPERABILIDAD EN COLOMBIA

El lograr un gobierno que tome decisiones basadas en datos, necesariamente requiere acceso a grandes volúmenes de datos de calidad de las entidades nacionales - entender tanto la oferta como la demanda de información es esencial para lograrlo. En el momento, con el marco institucional del país se ha avanzado en propuestas importantes para la gestión y administración de datos, en principio, al fomentar estándares de levantamiento y recolección de datos, iniciativas que impulsan una cultura de datos y propuestas para la coordinación de compartimiento de datos. Sin embargo, aun cuando existen estas iniciativas, se identifican barreras a la adopción transversal de estas.

Por un lado, no hay clara evidencia de que Colombia actualmente cuente con un inventario detallado y centralizado de todos los datos que producen las instituciones y organizaciones del país. El ejercicio más cercano a esto es el Plan Estadístico Nacional

- este inventario es vital para entender qué datos se producen en Colombia en este momento, y reconocer los vacíos en oferta de información, criterio vital para poder generar valor agregado y permitir la explotación de los datos, tanto por parte del estado, como del sector privado. Sin embargo, este también es un inventario de datos limitado, pues atiende principalmente a la demanda de información solicitada por la ODS, OECD y el DANE. El manejo de datos e información en Colombia es más orientado hacia el reporte o cumplimiento de una tarea, que hacia la toma de decisiones.

En cuanto a interoperabilidad, no se logra identificar claramente el uso masivo de dichos sistemas. Las entidades entrevistadas concuerdan que cuando hay intercambio de información entre algunas entidades, los atributos de calidad de los datos no están siempre presentes en todos los procesos de intercambio de datos, o que los datos no se presentan en tiempos relevantes para la toma de decisiones. Igualmente, se mencionó en las entrevistas que aun cuando se realizan capacitaciones y talleres para la interoperabilidad, no se hace seguimiento del uso efectivo de los datos. De hecho, las entidades explicaron que la tercerización de los sistemas de difusión o intercambio de información hace que los sistemas sean difíciles de integrar entre ellos, pues no se cuenta con el presupuesto ni personal para integrar los sistemas de interoperabilidad, o de lineamientos claros para la contratación de estos. Tampoco hay en este momento información ni métricas del uso de los sistemas de intercambio de datos, indicador básico para poder validar el impacto de estos esfuerzos, así como para poder ajustar los requerimientos en cuanto a la apropiación de tecnología y del ecosistema de datos.

Por otro lado, las entrevistas identificaron que muchas de las actividades típicas relacionadas con la explotación de datos recaen en contratistas o personal que tiene una alta rotación, lo que no permite un avance en cuanto a la apropiación y desarrollo de tecnología, en parte por las curvas de aprendizaje asociadas a estos roles. Tampoco hay una clara evidencia de que las entidades del Estado hayan implementado cargos enfocados a la gerencia de datos al interior de sus organizaciones, a pesar de normas como el decreto 415 de 2016, por el cual se señalan los lineamientos para el fortalecimiento institucional para la adopción de tecnología, y el literal B del artículo 45 de la Ley 1753 del 2015, por el cual se dispone que las entidades estatales tengan un Director de Tecnologías y Sistemas de la Información. El Anexo A, con los detalles del estudio cualitativo, contiene más información acerca de este punto.

Por último, dados los resultados del ejercicio de web scraping, se deben priorizar los esfuerzos institucionales y organizacionales en aquellas entidades cuyas funciones proporcionalmente son mayor en cuanto a la producción, manejo, y almacenamiento de información. Según los resultados, estas entidades pertenecen al sector estadístico, al de ambiente y desarrollo sostenible, y al electoral. De igual manera, se identifica la concentración de funciones relacionadas a información en las entidades descentralizadas y entidades adscritas.

3.3. DIMENSIÓN 3: INFRAESTRUCTURA FÍSICA (TECNOLOGÍA) ORIENTADA A DATOS

Al hablar de Big Data se debe hablar de los requerimientos de tecnología necesarios para poder mantener los procesos asociados a la explotación de datos. Una infraestructura típica para Big Data se conforma de 3 capas principales:

- Aplicaciones: esta capa contiene los sistemas de información para el manejo de datos, tales como aplicaciones personalizadas para usuarios, *business analytics*, proceso de *back-end*, entre otras.
- Almacenamiento de datos: en esta capa se mantienen las bases de datos de forma íntegra y fiable, para ponerlos a disposición de la capa de aplicaciones y se realizan los procesos de análisis de datos.
- Fuente de datos: esta capa se encarga de llenar las bases de datos bien sea de fuentes tradicionales, como aplicaciones de terceros, bases de datos, OLTP⁹; como de fuentes de Big Data, tales como sensores, datos no estructurados, entre otros.

Esta arquitectura es la estándar, pero varía según las necesidades de las empresas o entidades. Lo anterior quiere decir que el parque tecnológico necesario para un país está ligado a la cantidad de datos requeridos y generados y a la necesidad de procesamiento de estos datos.

En esta sección se evalúan los distintos componentes que conforman la infraestructura física y tecnológica en torno al aprovechamiento de datos y Big Data actualmente en Colombia. Para hacerlo, primero se analiza la infraestructura de datos en Colombia con la información disponible, esto incluye a los sistemas de información, la interoperabilidad desde el punto de vista técnico y las plataformas de datos abiertos. A pesar de los desarrollos tecnológicos, las actualizaciones en infraestructura y las estrategias orientadas a datos de las últimas administraciones; es difícil encontrar información relevante sobre el estado de la infraestructura tecnológica física del Estado colombiano, por lo cual determinar exactamente el parque tecnológico de todas las entidades, conlleva retos principalmente en el acceso y levantamiento de esta información. Determinar el verdadero aprovechamiento de la infraestructura física, por lo tanto, no es viable, sin utilizar indicadores internacionales que califican al país en cuanto a su infraestructura de conectividad, tecnología y desarrollo de sistemas de información. En este sentido, los rankings e índices de gobierno electrónico mencionados a lo largo de esta dimensión brindan una mirada breve al desarrollo de tecnología en el país, así como a la inversión en TIC que está realizando el gobierno nacional.

3.3.1. ESTADO ACTUAL DE LA INFRAESTRUCTURA TECNOLÓGICA EN EL SECTOR PÚBLICO, SEGÚN ESTUDIO CUALITATIVO

Según las entrevistas realizadas como parte de la investigación de este documento en cuanto a la infraestructura física orientada a datos, se encontró lo siguiente:

⁹ Siglas en inglés de procesamiento de transacciones en línea (On-Line Transactional Processing).

1. Servicios en la Nube: Se identificó una actitud positiva por parte de la mayoría de los entrevistados frente al uso de los servicios de la nube. Instituciones como Coldeportes, el Fondo de Adaptación, Ministerio de Agricultura, Función Pública y el Ministerio de Salud aseguran usarla; sin embargo enfatizan que son **nubes privadas**, las cuales les dan mayor seguridad y tranquilidad en la protección de los datos. Por otro lado, las entidades que no utilizan este servicio afirman que se están haciendo estudios con respecto a la posibilidad de adquirirlos. Solo en el caso de la Policía Nacional se evidencia una renuencia a hacerlo, pues hay una preocupación por la seguridad de la información. Los principales riesgos atribuidos a la nube son: a) pérdida de información; b) falta de seguridad y; c) falta de disponibilidad de la información. También, mencionan algunas de las soluciones para promover el uso de servicios en la nube, estas giran en torno a: a) dar garantías por parte del proveedor en cuanto a que la información va a estar salvaguardada; b) requerimientos técnicos y planes de contingencia en el centro de datos; c) *back-ups* mensuales; d) modernización tecnológica; y e) mayores recursos económicos. La Tabla 6 hace un resumen de las fortalezas y debilidades identificadas por las entidades entrevistadas en cuanto a servicios en la nube:

Tabla 6: Panorama de las entidades respecto a los servicios en la nube

Entidad	Fortalezas	Debilidades
Coldeportes	Nube privada, permite mayor eficiencia en el proceso.	
Colombia Compra Eficiente		
DIMAR		
Fondo de Adaptación	Política de gestión de información en la nube; el 90% de los sistemas de información están en nubes privadas.	
Policía Nacional		Restricción, sentir que la información no va a estar segura.
Ideam	Se está evaluando la posibilidad de contar con una nube.	
Ministerio de Agricultura	Nubes privadas que protegen los datos sensibles.	Las nubes privadas son más costosas.
DIAN	Se está evaluando la posibilidad de contar con servidores en Colombia; uso de una nube en Microsoft.	En la nube utilizada no está la información de los contribuyentes.
Función Pública	Nube que les permite reducir costos ya que no se utiliza infraestructura propia.	
Ministerio de Hacienda		
Ministerio de Trabajo		
Ministerio de Salud	Nube privada que garantiza mayor seguridad; uso de una nube pública para crear ambientes de desarrollo y de pruebas con datos anonimizados e inventados.	

Fuente: Elaboración propia

1. Datos Abiertos: Entre los entrevistados, existe una cultura por compartir información, en donde los principales motivadores para hacerlo son (i) la posibilidad de generar más datos e información para análisis e innovaciones para la toma de decisiones, y (ii) dar cumplimiento a la ley de transparencia. También se mencionó que la apertura de datos podría permitir una mayor visibilidad y posicionamiento. En cuanto a las barreras y riesgos, las entidades mencionan que hay datos sensibles y confidenciales que no se pueden publicar; por lo tanto, se identificó una falta liderazgo y compromiso con el tema o duplicidad de la información. De acuerdo con la situación actual de las entidades se puede afirmar que, en la mayoría de los casos, la decisión acerca de qué información publicar es el resultado del análisis que se hace internamente, en cuanto a lo que consideran puede ser requerido o relevante para el ciudadano, lo cual puede limitar la información al usuario final.

Lograr que las entidades públicas reconozcan el valor de los datos abiertos y que se logre compartir grandes volúmenes de datos y de calidad, es un paso importante para promover la infraestructura tecnológica que otras organizaciones y desarrolladores podrán utilizar como base para construir nuevas iniciativas.

- Interoperabilidad: En las entrevistas se definió la interoperabilidad como “el intercambio de información entre dos entidades estatales en pro de agilizar procesos”. Lo que motiva a las diferentes entidades a interoperar, es agilizar y facilitar los procesos, además de la posibilidad de contar con una información real y la unificación de los datos. A pesar de su importancia, entidades como el Fondo de Adaptación y la DIAN no han podido desarrollar un proceso de interoperabilidad y entidades como la DIMAR, Coldeportes y el IDEAM cuentan todavía con importantes retos con respecto al conocimiento de la plataforma de interoperabilidad del MinTic. Esto se debe a la existencia de algunas barreras como las limitaciones en cuanto a capacidades técnicas y recursos, la falta de cultura de datos por parte del personal, el temor frente a una calidad inadecuada del dato y obstáculos jurídicos en las diferentes entidades.
- Bases de Datos Internas y Bases de Datos Externas: Las tablas 7 y 8 identifican las bases de datos internas y externas que utilizan las entidades para el cumplimiento de sus funciones. En algunos casos hay información del número de sistemas de información que manejan estas entidades, aun cuando no se sabe el estado de estos.

Tabla 7: Panorama de las entidades respecto a las bases de datos internas

Entidad	Fortalezas	Debilidades
Coldeportes	22 sistemas de información.	
Colombia Compra Eficiente	La tienda virtual del Estado colombiano; Secop I; Secop II; VORTAL NEXT, plataforma transaccional para publicar adquisiciones y procesos de contratación.	No está el 100% de la compra pública en el SECOP II.
DIMAR	Datos básicos a cada una de las áreas, de la persona a cargo en cada subdirección, teléfono y correo electrónico.	
Fondo de Adaptación	Bases de datos de personas que han sido beneficiadas o damnificadas del fenómeno de la niña en el 2011; Datos de los diferentes proyectos.	
Policía Nacional		
Ideam	Un inventario de todos los procedimientos de trámites para los ciudadanos (SUIT); Sistema de Información Ambiental Colombiano en donde se consigna y organizada toda la información del medio ambiente.	
Ministerio de Agricultura	Consumidora de información; recolecta información de área, de producción y rendimiento de una región; Datos de empleo de cadenas productivas.	
Dian	Información tributaria de declaraciones de los contribuyentes; Información de comercio exterior.	
Función Pública	SIGEP, da cuenta de los funcionarios del Estado e incluye información de las diferentes entidades públicas; FURAG, evaluación continua de la gestión de las entidades.	

Fuente: Elaboración propia

Tabla 8. Panorama de las entidades respecto a las bases de datos externas

Entidad	Fortalezas	Debilidades
Coldeportes	Información de otros organismos del deporte; Sistema de Información Nacional del Deporte; Información de Registraduría.	No cuentan con información externa
Colombia Compra Eficiente	No requieren de bases de datos externas; interoperabilidad con el Sistema Integrado de Información Financiera y el Sistema Único Empresarial).	
DIMAR	Requieren información de centros médicos; información de la Registraduría para agilizar trámites.	
Fondo de Adaptación	Base de datos de la Universidad Nacional con información de damnificados de los diferentes fenómenos.	

Entidad	Fortalezas	Debilidades
Policía Nacional	Bases de datos de la Registraduría, Ministerio de Justicia, Contraloría y Procuraduría.	
Ideam	Imágenes de un satélite 'Land Sat'; Satélite del programa Copérnico.	
Ministerio de Agricultura	Bases de datos del DANE, gremios e Ideam; Información que procesan para sacar nuevos datos para publicación.	
DIAN	Bases del dato de Importaciones y Exportaciones del DANE; Información de agencias tributarias de otros países.	
Función Pública	Bases de datos del DANE y DNP.	
Ministerio de Hacienda	Bases de datos de La Gran Encuesta Integrada del DANE, La Encuesta Manufacturera y Las bases de Datos de Importaciones y Exportaciones de la DIAN.	
Ministerio de Trabajo	Bases de datos del Ministerio de Salud y Protección Social; RUES, RUT, Sisbén e información Prosperidad Social.	
Ministerio de Salud	Información de la Registraduría; Bases de datos de Migración Colombia, DNP con Sisbén y estadísticas vitales del DANE.	

Fuente: Elaboración propia

Lo anterior permite observar como los datos externos son un insumo fundamental para que las entidades puedan procesar datos relevantes a su función, lo cual permite confirmar la importancia de los fortalecer los sistemas de interoperabilidad, para generar nueva información al integrar estos datos

4. Tableros de Control: Con respecto a los tableros de control, la mayoría de las entidades asegura tenerlos (Coldeportes, Colombia Compra Eficiente, Fondo de Adaptación, Policía Nacional, IDEAM, Ministerio de Agricultura, DIAN, Ministerio de Hacienda, Ministerio de Trabajo) y, en las entidades que carecen de esta herramienta, se resalta el interés por hacerlo (Función Pública y Ministerio de Salud). Los tableros de control son importantes para la visualización de información y para comunicar *insights* sobre los datos alojados, especialmente pertinente para la planeación estratégica. Estos permiten explorar los datos según la necesidad del usuario, al proveer información actualizada y accesible al momento de toma de decisiones.

Tabla 9: Panorama de las entidades respecto a los tableros de control

Entidad	Fortalezas	Debilidades
Coldeportes	Tableros de control con seguimiento mensual y se alimentan con fuentes de cada área.	No hay un control de calidad.
Colombia Compra Eficiente	Sí cuentan con unos tableros que miden el comportamiento de la compra pública .	
DIMAR		Desconocen los tableros de control.
Fondo de Adaptación	Tableros dentro del sistema de gestión de proyectos.	
Policía Nacional	Tableros de tendencia criminal; Tableros estratégicos descriptivos de talento humano para comandancia.	
Ideam	Indicadores de reporte.	
Ministerio de Agricultura	Tableros (Agronet) del sector agropecuario.	
Dian	En desarrollo (Power BI).	
Función Pública		No cuentan con tableros de control
Ministerio de Hacienda	Cuentan con diferentes tableros de control que llevan un seguimiento de la economía.	
Ministerio de Trabajo	Tableros para la visualización (Tableau).	
Ministerio de Salud	Información de la Registraduría; Bases de datos de Migración Colombia, DNP con Sisbén y estadísticas vitales del DANE.	No cuentan con tableros de control.

Fuente: Elaboración propia

3.3.2. PROGRAMAS NACIONALES QUE FOMENTAN LA INFRAESTRUCTURA DE DATOS BOLETÍN TRIMESTRAL DE LAS TIC

Boletín Trimestral de las TIC

El MinTIC tiene un Boletín Oficial Trimestral que se enfoca en los indicadores de telecomunicaciones y mide la adopción de tecnología, principalmente en cuanto a conectividad en el país. Este reporta el acceso a internet fijo y móvil, conexión banda ancha, telefonía móvil, entre otros, para medir la penetración a departamentos y municipios en conectividad a internet. Para el segundo semestre de 2018, el boletín reporta 30.4 millones de conexiones a internet banda ancha, igual que la cifra para el primer trimestre.

Gobierno Digital

El Decreto 1004 de 2018 ratifica la transformación de la Estrategia de Gobierno en Línea a la Política de Gobierno Digital, la que se entiende como “el uso y aprovechamiento de las tecnologías de la información y las comunicaciones para consolidar un Estado y ciudadanos competitivos, proactivos, e innovadores, que generen valor público en un entorno de confianza digital”. Esta política continúa con el desarrollo y la masificación del gobierno electrónico, así como políticas, programas y planes de adopción y apropiación de TIC en las entidades del Estado. Específicamente prioriza habilitadores transversales de la política, entre estos los Servicios Ciudadanos Digitales¹⁰, lo que requiere del desarrollo de tecnologías en soporte de la autenticación electrónica, la carpeta ciudadana y la interoperabilidad.

Esta iniciativa busca seguir el ejemplo de políticas de gobierno electrónico desarrolladas por otros países para construir una nación transparente, eficiente y participativa para lograr mejores servicios online desde el gobierno; mejorar la administración de las TI, relacionarse con los ciudadanos y alcanzar los ODS.

Portal de Datos abiertos

El Gobierno Nacional, con el objetivo principal de aumentar la transparencia y la colaboración de la sociedad en la toma de decisiones, pone a disposición pública información para ser utilizada por la sociedad civil. Para esto, el Ministerio de Tecnologías de la Información y las Comunicaciones creó un portal web para compartir información y cubrir la estrategia de Gobierno Abierto en el marco de Gobierno en Línea (MinTIC, 2015).

En este portal se encuentran, de manera unificada, datos de las entidades públicas de Colombia en formato abierto, con el fin de que estos puedan ser usados por cualquier persona para desarrollar aplicaciones o servicios de valor agregado, hacer análisis e investigación, ejercer labores de control o para cualquier tipo de actividad comercial o no comercial. Este, basado en la nube, fue creado por la empresa Socrata y permite a todas las entidades gubernamentales compartir datos con el objetivo de contribuir a un gobierno transparente. Como se ha mencionado anteriormente, este portal cuenta con varias limitaciones y barreras, aun cuando es una característica importante de la infraestructura de tecnología de Colombia.

Iniciativa de Big Data de MinTIC

Promovido por el MinTIC a través de Centros de Excelencia y Apropiación (CEA) en Big Data y el internet de las cosas, los CEA promueven los beneficios sociales, culturales, económicos y políticos, relacionados con la implementación de Big Data, convirtiéndolos en un ‘data hub’. El principal objetivo de los CEA es enfocarse en el análisis de datos para los sectores productivos, la formulación de políticas públicas y el internet de las cosas. La Alianza Caoba, como una alianza público-privada y un CEA, es un ejemplo del continuo desarrollo de la infraestructura tecnológica en el país - aun así, es necesario continuar ampliando la cobertura de las CEA alrededor del país.

Con esta iniciativa, MinTIC espera crear sectores especializados que incentiven la creación de programas universitarios de Big Data, compañías de TI de Big Data y compañías de no TI que implementen Big Data. Además, busca fortalecer el diseño e implementación de políticas públicas basadas en datos y evidencia; aumentando la capacidad de infraestructura para datos del país.

3.3.3. INFRAESTRUCTURA ORIENTADA A DATOS EN EL SECTOR PÚBLICO EN COLOMBIA, SEGÚN EL CONPES 3920.

Dentro del Documento CONPES 3920: Política Nacional de Explotación de Datos (*Big Data*), se establecen varias estrategias para aumentar el aprovechamiento de datos en Colombia. Entre ellas, están temas de selección de tecnologías para compartir datos teniendo en cuenta las asimetrías en las capacidades técnicas, las cuales impactan directamente a la infraestructura física, bien sea en aumento, optimización o actualización del parque tecnológico.

¹⁰ <http://estrategia.gobiernoenlinea.gov.co/623/w3-propertyvalue-7650.html>

Las estrategias se centran en generar alineaciones, guías, estándares a implementar y regulación legal para crear un ecosistema de datos, pero la inversión para la implementación es responsabilidad de cada entidad. El MinTIC tiene una unidad Administrativa Especial denominada Agencia Nacional de Contratación Pública

–Colombia Compra Eficiente–, como una entidad descentralizada que acelera el proceso de compra y mejora los precios. Pese a estos esfuerzos por proporcionar herramientas que permiten a las entidades acceder a tecnología de forma fácil y a precios competitivos, la asimetría en las capacidades técnicas que se requiere para usar de manera adecuada dicha tecnología puede retardar la implementación de los sistemas necesarios para la explotación de datos. Esto puede obligar a migrar a servicios de infraestructura que no son óptimos para los requerimientos de cada entidad - en volumen, latencia, interoperabilidad u otros-.

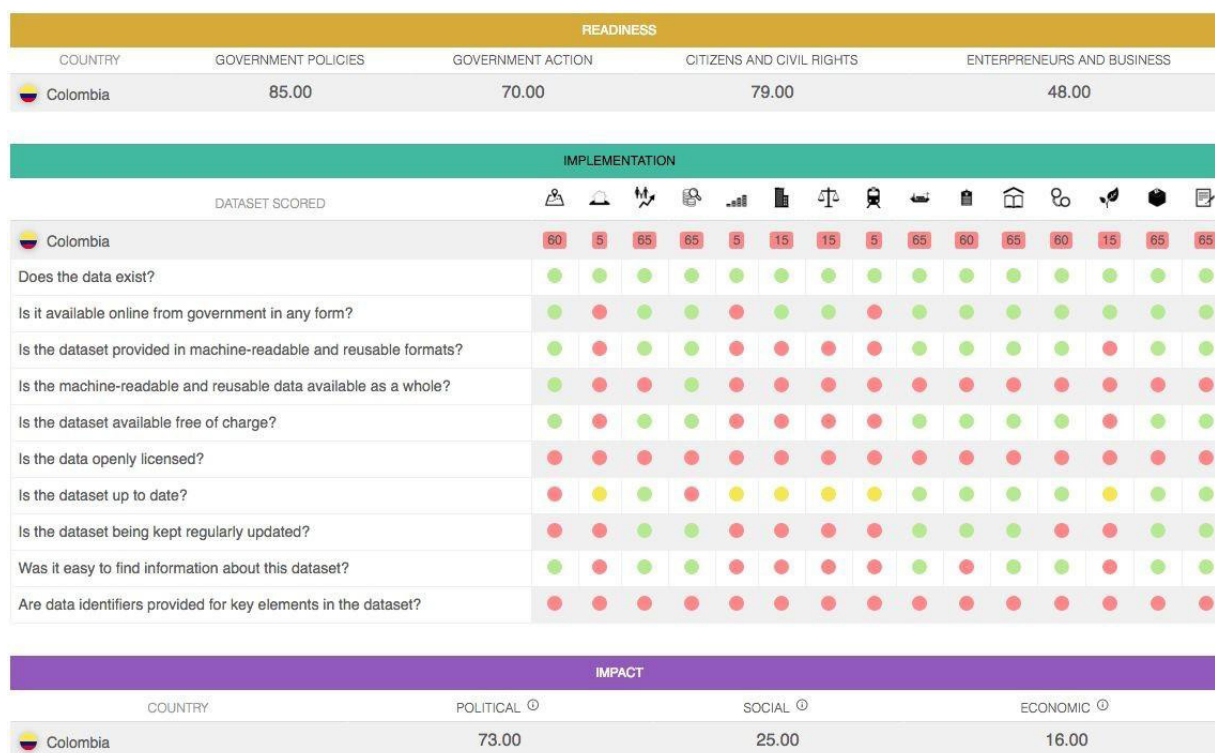
3.3.4. INFRAESTRUCTURA FÍSICA (TECNOLOGÍA) SEGÚN INDICADORES INTERNACIONALES

Ante la dificultad de obtener un inventario completo de los sistemas, bases de datos, servidores y aplicaciones existente en Colombia, debido a la complejidad de acceso a esta información y la tercerización de los servicios, se identificaron indicadores internacionales que son útiles para evaluar la infraestructura de datos en términos de trámites, optimización, disponibilidad de datos, entre otros.

El **Open Data Barometer** es un indicador producido por el World Wide Web Foundation que tiene como objetivo identificar la prevalencia e impacto de las iniciativas de datos abiertos a nivel mundial. Asimismo, analiza las tendencias y proporciona datos comparativos sobre países y regiones a través de una metodología que combina datos contextuales, evaluaciones técnicas e indicadores secundarios (Figura 9).

Para el 2016, Colombia ocupó el puesto número 24 a nivel mundial y 4 para la región de Latinoamérica y el Caribe, escalando 4 puestos desde el 2015 y logrando pasar del puesto 40 al puesto 24 en tan sólo tres años (2013-2016). Este ranking, que mide alrededor de 120 países año a año¹¹, , identifica que las barreras principales para datos abiertos en Colombia es la ausencia de estándares de datos, que hagan de la información fácil de usar e interpretar. De igual manera, se identifica que no existen descripciones sobre los datos, lo que confirma puntos elaborados anteriormente sobre datos abiertos. A su vez, este ranking mide también si los datos abiertos están en el formato apropiado para ser procesados.

Figura 9: Ranking de Open Data Barometer - Colombia



Fuente: Perfil de Colombia- Open Data Barometer, 2016

¹¹ Excluye países como Cuba, las Guayanas, algunos países de la Liga Árabe y de África Central

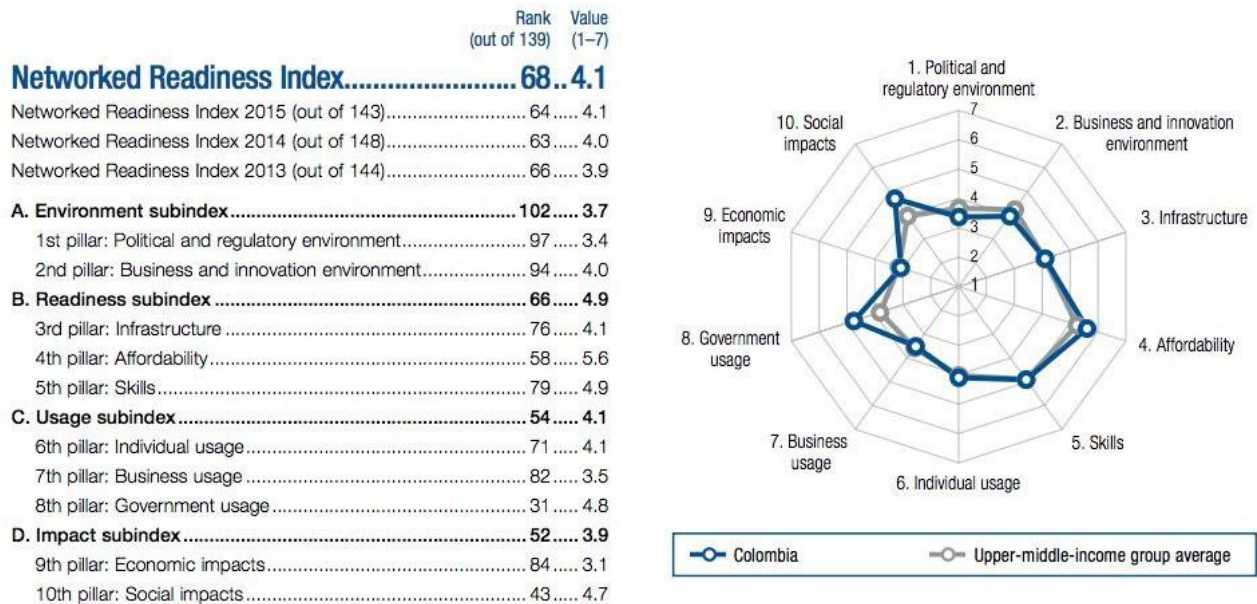
El Índice de Preparación a la Conectividad está incluido en el Reporte Global de Tecnologías en la Economía Digital en 139 países y “examina el crecimiento exponencial logrado con tecnologías digitales, la forma como se mide el impacto en innovación, la presión continua para los sectores tecnológicos y no tecnológicos para impulsar la innovación a través de medios digitales, y la necesidad de una gobernanza ágil y sistemas regulatorios que se adapten a la velocidad y a la escala de los cambios al tiempo que logran mitigar riesgos éticos, legales y regulatorios” (World Economic Forum 2016).

De acuerdo con la información presentada en la página de MinTIC se muestran 3 posiciones en este índice: en primer lugar, el país inicia en el 2013 en la posición 66 y en 2014 tiene un salto debido a los desarrollos en el acceso a infraestructura a la posición 63. Sin embargo, en 2016 baja a la posición 68, justificado según el MinTIC por las mejoras que tuvieron otros países en estas áreas de desarrollo. Asimismo, se puede observar que la adopción de las TIC, por parte de diferentes sectores de la población, se mantiene en incremento. Por otro lado, el índice de desarrollo en TIC de ITU establece que en 2017 había 45.5 suscripciones a conexión móvil de banda ancha por cada 100 personas, mientras que en 2013 esta cifra era de 25.0 y en 2011 de 3.7 (ITU Data, 2017). Internet World Stats reporta que el 63.2% de los colombianos cuentan con acceso a internet, cifra debajo de Perú, Ecuador, Chile, Brazil, Argentina, Paraguay y Uruguay. En términos comparativos, estos es una cifra baja a nivel regional, independiente de aumento en las suscripciones a conexión móvil de banda ancha en los últimos años.

También, según el Índice de Preparación a la Conectividad, tanto el uso extendido de las TIC para operaciones *b2b* y *b2c*, así como la creación de nuevos modelos de negocio, se ha estancado durante los últimos años. En general, se observa que el ambiente político y de negocios en el país sigue estableciéndose débilmente, con bajas tasas de efectividad por parte de las entidades encargadas de realizar las leyes (121) y un ineficiente sistema judicial donde se requieren 1,288 días para hacer cumplir un contrato, que dificulta la creación de una infraestructura eficiente de datos. Este ejemplo ubica al país en la posición 133 a nivel mundial. En cuanto a infraestructura, este índice mide la producción eléctrica, acceso a banda ancha, y acceso a servidores seguros – en estos Colombia está en los puestos 93, 67 y 66 respectivamente.

Este índice es útil para comparar el desarrollo de la infraestructura tecnológica vis a vis el ecosistema de negocios, o regulatorio. En comparación, la infraestructura es uno de los puntos más bajos reportados, aun cuando este se encuentre en el promedio de países de ingresos medios. Asimismo, este índice revela que el gobierno es el usuario principal de las tecnologías, lo que esclarece la necesidad de incrementar el uso de estas por parte del sector privado.

Figura 10: Índice de Preparación a la Conectividad, Colombia



Fuente: Perfil de Colombia 2016 - Network Readiness index- World Economic Forum, 2016

El **Ranking de Waseda E-Government** es una investigación sobre la evolución a nivel mundial del gobierno digital y también una encuesta sobre el gobierno digital ideal, con base en diferentes criterios. En el caso colombiano, el país ha estado en la muestra de la encuesta desde 2014. En 2015, Colombia se estableció en el puesto 47 del ranking, para 2016 bajó al 55 y en 2017 quedó en el puesto 57 de los 65 países encuestados; la mayoría pertenecientes a América, Europa y Asia.

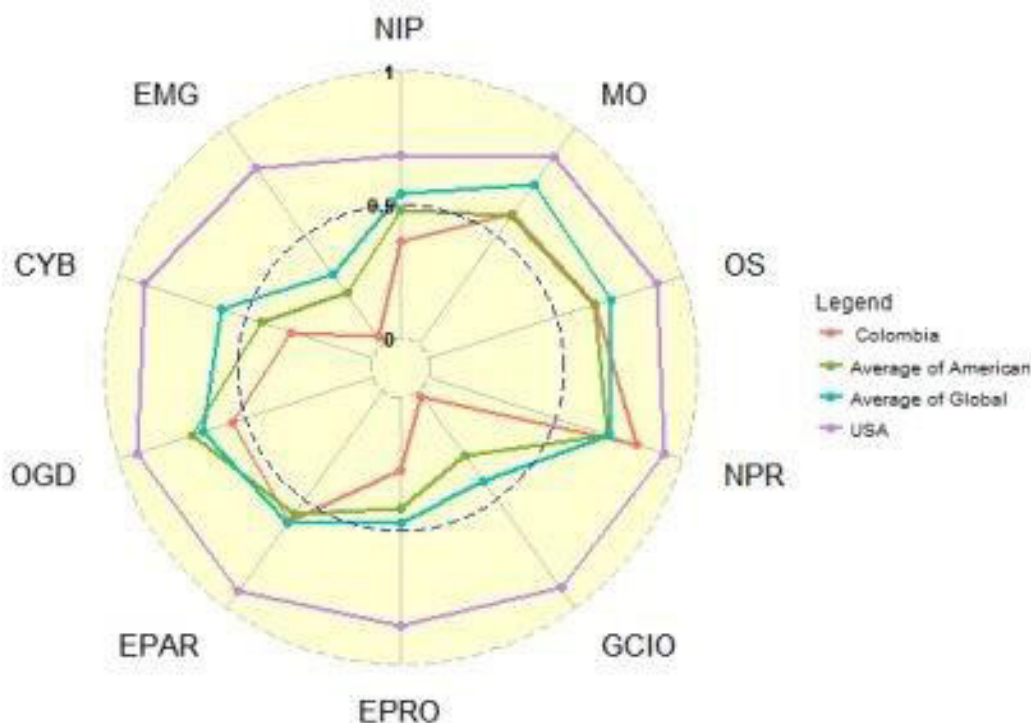
En este contexto, el último reporte indicaba que

“Colombia está realizando grandes cambios en algunas áreas, pero la tendencia general no ha sido positiva en los últimos años. En el Índice de Gobierno Digital de Naciones Unidas, Colombia pasó del puesto 31 al 50 del 2010 al 2014. Gran parte de

estos cambios pueden ser atribuidos a un entorno de gobierno mucho más competitivo y saturado de ambientes de Gobierno Digital a medida que más países desarrollan su infraestructura TIC. Sin embargo, recientes esfuerzos para disponer y promover plataformas de Gobierno Digital en el país indican que el gobierno ha reconocido su importancia. Además, hay que tener en cuenta que Colombia continúa en un proceso de adaptación para ser uno de los países del top 20 en el área de participación digital” (International Organization for Standardization, 2017, p. 101).

En este mismo reporte se menciona que “Colombia ha participado en diferentes iniciativas internacionales para mejorar los gobiernos digitales en las Américas. Por ejemplo, ha postulado muchos candidatos para los premios excelGOV y, por lo general, participa en grupos de trabajos regionales e internacionales en una tasa más alta que los países vecinos. Sin embargo, todavía hay mucho espacio para el mejoramiento en términos de rankings, Colombia está preparada para continuar en su trayectoria de crecimiento y realizar importantes avances en los próximos años” (International Organization for Standardization, 2017, p. 102).

Figura 11: Waseda E-Government Ranking, Colombia

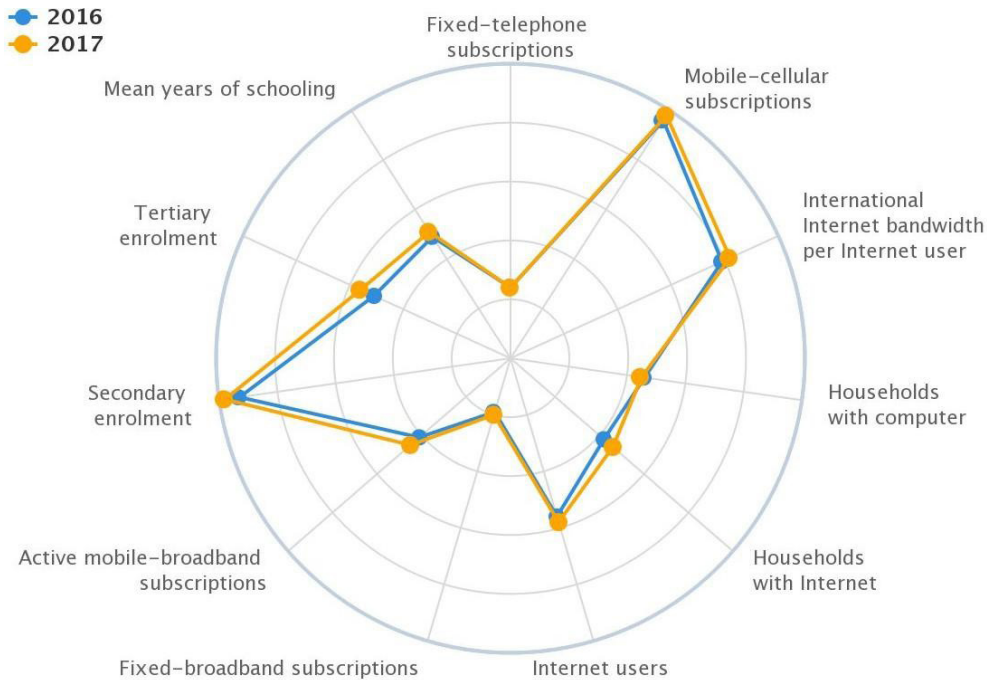


Fuente: Perfil de Colombia 2017 - Waseda E-Government Index, 2017

Es importante resaltar que en cuanto a la preparación de la infraestructura de red (NIP) Colombia se encuentra por debajo del promedio de América Latina.

El **Índice de la Unión Internacional de Telecomunicaciones**, organismo especializado de las Naciones Unidas para las TIC, mide el acceso a internet de banda ancha, uso de internet, y suscripción a telefonía móvil y fija para los casi 200 miembros de la ONU. Para 2016, la página web de MinTIC publicaba que “la tendencia de estos indicadores muestra un crecimiento a nivel global en las líneas móviles, que está impulsando el crecimiento de internet móvil desde el 2010. Para el caso colombiano, las estadísticas muestran un crecimiento en infraestructura, especialmente durante los últimos 5 años, donde las suscripciones de internet han aumentado en más de 2.7 millones, el número de usuarios alcanzó el 51.7% duplicando los valores para el 2008 y el cubrimiento de las líneas de telefonía móvil es de casi 100% desde 2012”. Esta información corresponde a las bases de datos de 2014. En 2017, Colombia se encuentra en la posición 84 a nivel mundial y en la 16 para la región de América, manteniéndose en el mismo lugar para en ambos rankings respecto al año anterior. Estas cifras confirman, que aun cuando hay crecimiento en cobertura nacional, a nivel regional aun el acceso a internet es bajo – la conectividad es un requerimiento esencial para la explotación de datos.

Figura 12: Índice de la Unión Internacional de Telecomunicaciones, Colombia



Fuente: Perfil de Colombia 2017 - ICT Development Index - International Telecommunication Union, 2017.

El **E-Government Development Index (EGDI)** es una encuesta realizada por el Departamento de Asuntos Económicos y Sociales de Naciones Unidas a los 193 países miembros y mide la efectividad en cuanto a gobierno digital y a la prestación de servicios económicos y sociales a las personas. Esta encuesta identifica patrones en el desarrollo y ejecución del gobierno electrónico, así como países y áreas en los que el potencial de las TIC y del gobierno digital no han sido explotados completamente y donde un apoyo al desarrollo de capacidades podría ser útil (Naciones Unidas, 2005).

Para este año (2017), Colombia se encuentra en el puesto 61 del ranking EGDI y en el puesto 23 del ranking E-Participation. Colombia, junto con Chile, México y Uruguay “han adoptado sistemas de gobierno electrónico para mejorar la prestación de servicios [...] La participación online de estos países presenta un diseño estratégico, aspectos de gobierno abierto, así como esfuerzos para mejorar la coordinación institucional, la transparencia y la facilidad para el acceso” (CEPAL, 2014). De los componentes del índice, el más bajo en Colombia es la Infraestructura Tecnológica (TII), comparado con Capital Humano (CI) y Servicios En-Línea (OSI).

Figura 13: E-Government Development Index, Colombia



Fuente: Perfil Colombia. E-Government Development Index - United Nations, 2018

Asimismo, en el perfil del país se encuentra que Colombia utiliza su portal nacional para involucrar socios en la toma de decisiones como parte de una política de gobierno, que tiene como objetivo implementar la transparencia en todos los niveles. Este enfoque busca generar un mayor acceso a la información, proveer servicios online, incentivar la participación de los ciudadanos en la generación de políticas públicas y establecer estrategias anticorrupción, lo que se enmarca en la política de buen gobierno.

3.3.5. DIAGNÓSTICO SOBRE LA DIMENSIÓN

Llegar a una conclusión calificativa sobre el estado de la infraestructura (física) en cuanto a datos en Colombia, sin explorar el estado de los sistemas de información, la capacidad de almacenamiento, y las aplicaciones u otras plataformas existentes desde y para el gobierno colombiano – tarea fuera del alcance de esta consultoría- es una tarea difícil. Sin embargo, las entrevistas realizadas, así como los indicadores internacionales permiten hacer las siguientes conclusiones:

- Las entidades encuestadas e indudablemente otras entidades públicas, se encuentran realizando estrategias para mejorar la sistematización e interoperatividad de los sistemas de información que permita la explotación de datos, pero no se encuentra mucha información de la articulación de estas entidades a nivel técnico. Aunque algunas entidades están avanzadas en la sistematización e interoperabilidad, no se encontró suficiente información sobre la implementación y transferencia de conocimiento en otras.
- De acuerdo al estudio cualitativo, no se cuenta con una política de escalabilidad dinámica de los sistemas de información según consumo y uso de recursos. Es decir, no están claros los parámetros de cambio o renovación cuando se llega a un mayor nivel de consumo de información que requieran recursos de forma inmediata, necesario para mantener un ecosistema de datos interconectado.
- Faltan políticas que permitan generar mayor confianza y legitimidad sobre la tecnología. Esta es una barrera identificada de las entrevistas realizadas y es un factor esencial en el desarrollo de la infraestructura tecnológica pertinente a la explotación de datos.
- Aún falta ampliar cobertura y conectividad en el país, como se evidencia en los indicadores internacionales, en donde a nivel región, Colombia se encuentra por debajo del promedio. Independiente del aumento en suscripciones a conexión móvil o banda ancha en los últimos años, el desarrollo del país es más lento que el del promedio de la región.
- Aunque existen varias iniciativas del gobierno que buscan fortalecer la tecnología en el país, principalmente del MinTIC, las oficinas de Tecnologías de la Información o Sistemas en las entidades públicas no son vistas como parte estratégica de las entidades gubernamentales, según las entrevistas realizadas. Pocas entidades han migrado los sistemas computacionales para un aprovechamiento o explotación de datos, a pesar de tener desde 2010 políticas para apropiarse o implementar sistemas interoperables que permitan la explotación de datos o implementación de proyectos de Big Data.
- De acuerdo a las entrevistas realizadas (para información detallada sobre entrevistas, ver Anexo A), en varias entidades del Estado y del sector privado, la inversión en tecnologías de la información es percibida como un gasto; reduciendo así las áreas de Innovación y Tecnologías de la Información a procesos para el manejo de soporte de sistemas. Por ejemplo, las políticas de interoperabilidad y de un Sistema Nacional de Estadística llevan varios años desde su aprobación y no se ve una cobertura amplia de las entidades en este aspecto; haciendo entrever que la inversión en Tecnologías de Información es más reactiva que proactiva. De igual forma, la inversión en proyectos de investigación y desarrollo dentro del sector público y empresas privadas es baja, lo que hace lento el proceso de apropiación de tecnología o de innovación. Un ecosistema de explotación de datos saludable necesita de inversión en TI, actualizaciones constantes en tecnologías para interoperar y de una cultura de I+D dentro de las entidades públicas y privadas.

3.4. DIMENSIÓN 4: ARQUITECTURA DE DATOS

A diferencia de la infraestructura física, la arquitectura de datos busca entender el estado actual de los esquemas que permiten compartir y consumir datos, las iniciativas de inteligencia de negocios y Big Data, gerencia y calidad de datos, entre otros aspectos relacionados con la producción, administración y mantenimiento de datos. La infraestructura de datos más nombrada, y una de las primeras en ser desarrollada, es la Infraestructura de Datos Espaciales (IDE), la cual se define como “un sistema informático conformado por estándares, políticas, tecnologías, recursos, catálogos, servidores, programas, aplicaciones, páginas web, entre otras, con un marco legal que permite la interoperabilidad con el fin de acceder y disponer de los datos espaciales que pueden ser compartidos por la comunidad que se encuentra en la red” (Unidad Administrativa Especial de Catastro Distrital 2016). Bajo esta definición inicial de una infraestructura de datos, se puede entender esta como todos los elementos o aspectos que permiten salvaguardar, proteger, preservar, procesar, almacenar, transferir y enviar datos basados en implementaciones tecnológicas para desarrollar este fin.

Desde el punto de vista tecnológico, estos sistemas deben tener las características denominadas RAS (por sus siglas en inglés) que son fiabilidad, disponibilidad y facilidad. El desarrollo o implementación, como se define para los IDE, debe estar acompañado de estándares, políticas y catálogos que permitan crear ecosistemas de datos para transformar los datos en información.

Para este documento, y para simplificar el análisis de esta dimensión, los aspectos a analizar para la infraestructura de datos son:

- Actual arquitectura de datos para el gobierno colombiano
- Esquemas de compartición de datos
- Iniciativas de Business Intelligence y Big Data
- Gerencia o administración de datos
- Calidad de los datos

3.4.1. ESTADO ACTUAL DE LA ARQUITECTURA DE DATOS EN EL ESTADO COLOMBIANO

Actualmente Colombia tiene varias iniciativas y una Plataforma de Interoperabilidad de Datos (PDI), que se ha venido gestando desde 2006 a través de varios planes, leyes y marcos promocionados por el gobierno, en especial por el Ministerio de Tecnologías de la Información y las Comunicaciones. A continuación, se presenta una explicación más detallada alrededor de estas iniciativas.

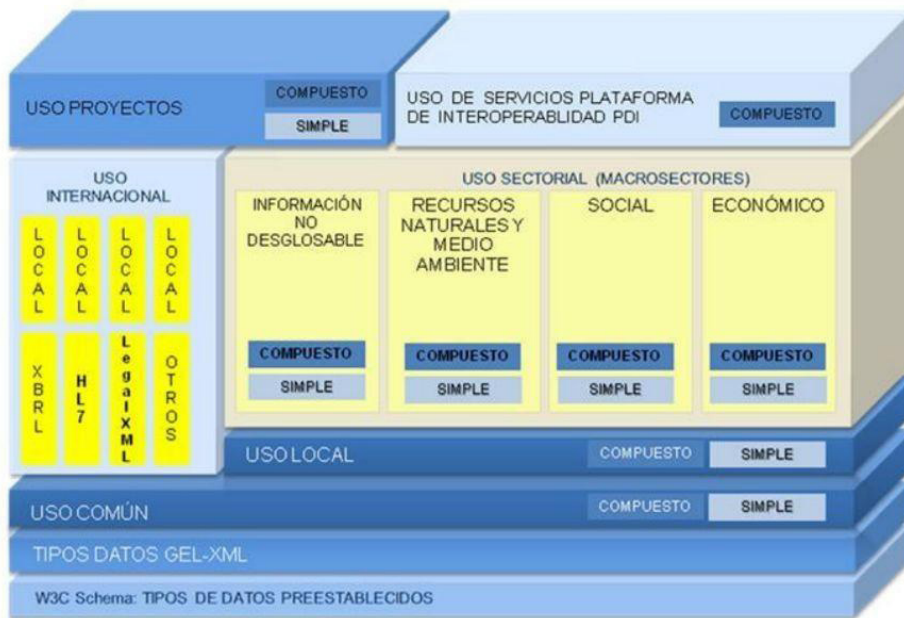
3.4.1.1. MODELO DE ARQUITECTURA DE DATOS SEGÚN MINTIC

El MinTIC, cuenta con una propuesta de arquitectura de datos de la plataforma de interoperabilidad de datos (PDI) que permite el intercambio de información entre entidades gubernamentales. Esta arquitectura de datos busca “presentar un conjunto de conceptos definidos (...) que soporta la identificación y definición de las unidades de información o elementos de dato que actúan como base en la definición estándar de conceptos del lenguaje”, lo cual está determinado en el documento oficial de la Arquitectura de datos (MinTIC, 2008, p. 2).

Este documento define el lenguaje común para el intercambio de información por los diferentes actores que prestan y/o demandan trámites y/o servicios con el objetivo de “facilitar la creación de interfaces estándar entre procesos y sistemas de información”, una iniciativa que hizo parte de la estrategia de Gobierno en Línea y aún se encuentra vigente (Ministerio de Tecnologías de la Información y las Comunicaciones, 2011, p. 13).

De acuerdo a estos lineamientos, la estructura de la Arquitectura de datos se divide por capas basadas en el uso, macro sector u origen.

Figura 14: Arquitectura de datos



Fuente: Ministerio de Tecnologías de la Información y las Comunicaciones, 2011

Esta arquitectura define una serie de capas de uso como “niveles de organización de los elementos de dato que representan”, las cuales se clasifican en 11 subconjuntos:

- Tipos de datos Preestablecidos: Tipos de datos de la recomendación W3C XML Schema (Ubiquitous Web domain, 2007), como por ejemplo date, DateTime, Floor, entre otros.
- Tipos de datos GEL-XML: Datos base para el del estándar y se definen con los tipos de datos preestablecidos, como porcentaje.

- Uso Común: Representaciones de conceptos de información que se usan a nivel nacional e internacional, como código de país, apellidos, entre otros
- Uso Local: Representaciones de conceptos de información de uso netamente interno del país, como divipola, puntaje SISBEN, entre otros.
- Uso Macro sector Económico: Información estadística relacionada con el sector económico.
- Uso Macro sector Social: Información estadística relacionada con el sector social
- Uso Macro sector Recursos Naturales y Medio Ambiente: Información estadística relacionada con el sector del medio ambiente
- Uso Información No Desglosable: Información estadística que no es desglosable por sector o tema.
- Uso Proyectos: Conceptos de información de uso particular a un determinado proyecto de una entidad.
- Uso Servicios de la Plataforma de Interoperabilidad – PDI: Información necesaria para el intercambio entre entidades.
- Uso Internacional: Información definida en estándares internacionales.

Cada una de estas capas está conformada por siete áreas de información:

- Temporal (ej. duración, fecha, año, hora, tipo período)
- Ubicación: (ej. coordenadas, dirección, piso, lugar, zona postal)
- Identificación (ej. número de cédula de ciudadanía, número acto administrativo, código único de identificación de entidad pública)
- Organización: (ej. código tipo persona jurídica, código tipo sociedad)
- Personal (ej. código sexo, código tipo estado conyugal, código pertenencia étnica)
- Documental (ej. código tipo documento, texto documento, número de páginas).
- General

De igual manera, el documento brinda los lineamientos para los tipos de elementos y la identificación de metadatos relacionados, con base en elementos de diferentes normas establecidas, como por ejemplo la ISO/IEC 11179-1 (International Organization for Standardization, 2017), y estándares internacionales para compartir datos. Además, establece que el desarrollo de un lenguaje común de intercambio de información es un proceso que se debe refinar a medida que es implementado y puesto en uso. Por último, propone un punto de partida consensuado para identificar y definir los elementos de datos a compartir de acuerdo con las necesidades de información del país.

3.4.1.2. MODELO DE ARQUITECTURA DE DATOS SEGÚN DANE

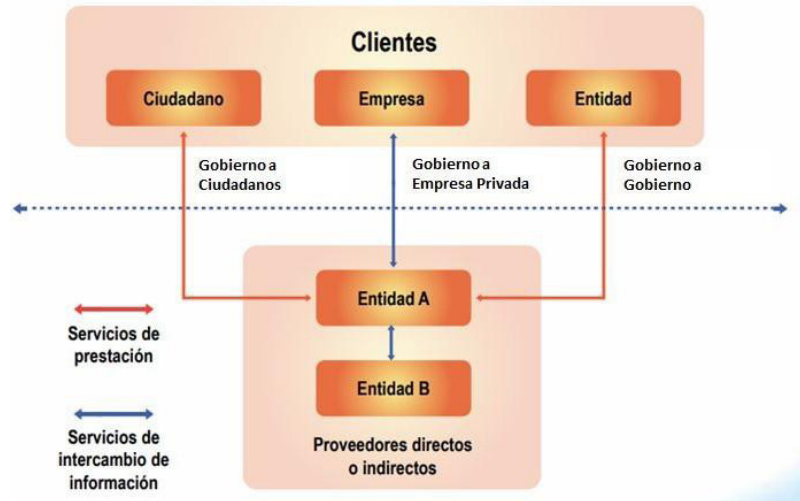
Previo a la existencia del Plan Estadístico Nacional, el DANE contaba con una arquitectura de datos propia, parte de su Plan Estratégico de Información Básica (2006-2010). Sin embargo, para el período 2017 – 2022 el PEN no determina exactamente una arquitectura de datos. En su lugar, plantea el estándar internacional de difusión de datos denominado Statistical Data and Metadata eXchange (SDMX). Este estándar busca garantizar la facilidad del acceso a los datos estadísticos y asegurar que los datos estén acompañados de sus correspondientes metadatos. El SDMX está basado en métodos de modelamiento y transformaciones simples de datos e incluye especificaciones en los formatos de intercambio, registros, web services, entre otros. Este estándar busca integrar bases de datos distribuidas bajo un mismo protocolo de exploración y solicitud, independiente de la ubicación física de estos. Es relevante por su énfasis en la calidad de los datos.

3.4.1.3. MODELOS DE COMPARTICIÓN DE DATOS ENTRE ENTIDADES

El Marco de Interoperabilidad, descrito en detalle en la Dimensión 2, es relevante también en esta dimensión, pues es el punto de partida para una arquitectura de datos integrada. El Marco facilita las necesidades de las entidades del Estado a través del Servicio de Intercambio de Información, el cual tiene como objetivo automatizar el proceso de solicitar y recibir información, evitando trámites innecesarios para el intercambio de información.

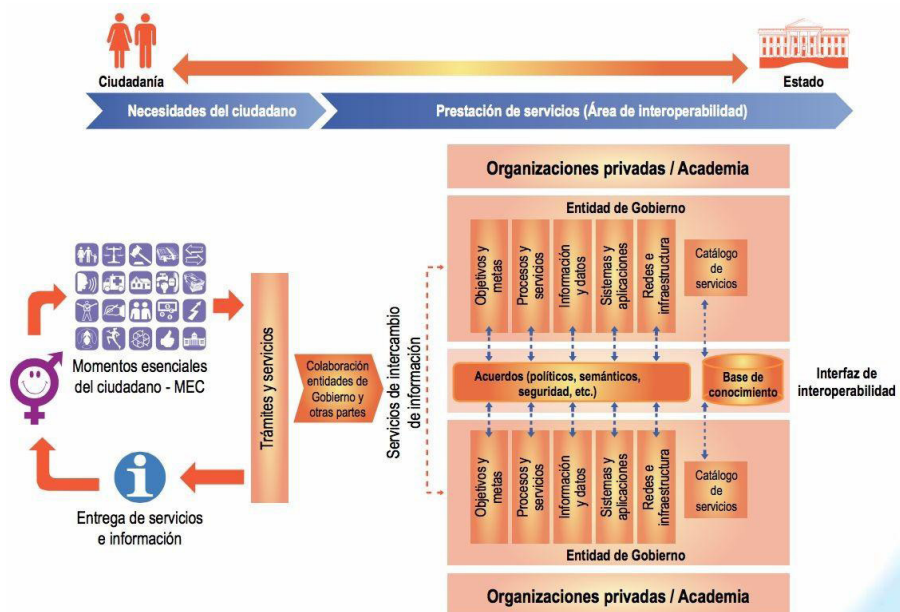
Este Marco define una arquitectura de intercambio entre entidades que garantiza transparencia y eficiencia, involucrando los niveles estratégicos, de negocio y tecnológicos de las entidades. Además, propone la centralización de los servicios de intercambio en un directorio agrupado por servicios, permitiendo encontrar de forma rápida la información disponible para entidades del Estado, ciudadanos y empresas.

Figura 15: Modelo de servicio de intercambio de información



Fuente: OECD, 2011

Figura 16: Modelo de servicios de intercambio de información entre entidades



Fuente: Ministerio de Tecnologías de la Información y las Comunicaciones, 2010

De manera complementaria, el MINTIC propone un Marco de Referencia para la arquitectura de datos empresariales, buscando maximizar el aporte a los objetivos institucionales y estatales por medio de los siguientes seis dominios (*Arquitectura TI Colombia, MinTIC, 2016*):

- Estrategia de TI
- Gobierno de TI
- Información
- Sistemas de Información
- Servicios Tecnológicos
- Uso y Apropiación

La Guía Técnica Básica de Información, guía parte del Marco de Referencia (*Arquitectura TI Colombia, 2014*) tiene como uno de sus principales objetivos "(...) orientar a la dirección de Tecnologías y Sistemas de la Información o quien haga sus veces,

durante la implementación del dominio de información” y, además, menciona modelos de compartición de información entre entidades. Estos modelos de compartición se vinculan a temas de interoperabilidad dentro del ámbito de Diseño de los componentes de información, como forma de implementación exitosa de los lineamientos.

Según el portal de datos abiertos, de diciembre de 2007 a diciembre de 2015¹², se notificaron del uso de Lenguaje Común de Intercambio de Información 725 entidades de los diferentes niveles, de las cuales según datos.gov.co, 524 son alcaldías municipales. De todas las entidades, sólo 73 están en el tercer nivel de cumplimiento, mientras que la mayoría (433) se encuentran en el primer nivel. La diferencia entre estos es de sólo diseñar los componentes de información según los lineamientos del estándar de lenguaje común, a publicar el servicio para su consumo o provisión teniendo pleno conocimiento del Marco de Interoperabilidad (MINTIC 2018). Que sólo el 10% de las entidades se encuentren en este nivel, indica la adopción baja del lenguaje común, elemento clave para la interoperabilidad.

Finalmente, el Documento CONPES 3920 (DNP, 2018), al realizar el análisis de la interoperabilidad, concluye que el principal objetivo de la Plataforma de Interoperabilidad (PDI) del Estado es permitir el intercambio de información entre entidades, pero se presentan muchos retos que limitan el uso de la PDI en las entidades. En palabras textuales, en este se afirma que “la intervención orientada a la exposición y consumo de datos en un lugar centralizado (como lo es la PDI) requiere de instrumentos para la articulación de las acciones de cada entidad que permitan alinear sus incentivos para materializar la interoperabilidad” (DNP, 2018).

Lo anterior, muestra la necesidad de cooperación y de mecanismos de apropiación de cultura de datos y habilidades digitales en las entidades para incentivar el uso de la información disponible en la PDI, compartir la información sin restricciones y la apropiación de tecnología para intercambio de información.

3.4.1.4. MODELOS DE GERENCIA DE DATOS EN EL SECTOR PÚBLICO

La organización Data Management Association (DAMA)¹³ define la gerencia de datos como “el desarrollo, ejecución y supervisión de planes, políticas, programas y prácticas de control, protección, entrega y mejoramiento del valor de los datos y activos de información” (Cupoli, 2014). En términos más sencillos, según el MIT, la gerencia de datos se puede entender como la que “permite juntar los sistemas de captura y todos los metadatos necesarios para operar un sistema” (Digital Preservation Management, 2017).

De acuerdo a esta definición, DAMA sostiene 11 áreas de conocimiento necesarios para poder gerenciar los datos de forma eficiente (Cupoli, 2014). Estos incluyen:

- **Gobernanza de datos:** planificación, supervisión y control sobre la gestión de datos y el uso de datos y recursos relacionados.
- **Arquitectura de datos:** estructura general de los datos y los recursos relacionados como parte integral de una arquitectura empresarial.
- **Modelamiento y diseño de datos:** análisis diseño, construcción, prueba y mantenimiento de los datos.
- **Almacenamiento y operaciones con datos:** activos físicos para el despliegue, difusión y administración de recursos relacionados con datos.
- **Seguridad de los datos:** aseguramiento de la privacidad, confidencialidad y acceso apropiado a los datos.
- **Integración e interoperabilidad de datos:** adquisición, extracción, transformación, entrega, replicación, virtualización y soporte operacional para los datos disponibles.
- **Documentos y contenidos:** almacenamiento, custodia, indexación y disponibilidad de acceso a datos en fuentes no-estructuradas (archivos electrónicos y físicos) y poner estos a datos disponibles para la integración e interoperabilidad.
- **Referencia y datos maestros:** administrar datos compartidos para reducir redundancia y garantizar una mejor calidad de los datos, mediante una definición estandarizada y el uso de los valores de los datos.
- **Almacenamiento de datos e inteligencia empresarial:** gestionar el proceso de análisis de datos y permitir el acceso a decisiones basadas en datos para reportes y análisis.
- **Metadatos:** recopilar, categorizar, mantener, integrar, controlar, administrar y entregar metadatos.
- **Calidad del dato:** definir, monitorear, mantener la integridad y mejorar la calidad de los datos.

Aunque no abarca todos los temas relacionados con gerencia de datos, el Archivo General de la Nación¹³ tiene una guía de implementación de un sistema de gestión de documentos electrónicos de archivo. Esta guía busca que las entidades

¹² Este es el rango de fecha disponible para la base de datos según datos.gov.co, para más información visitar <https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Niveles-de-Cumplimiento-del-Lenguaje-Com-n-de-Enti/ec-84-nmmg/data>

¹³ Entidad encargada de promover la conservación y consulta del legado documental nacional y ponerlo a disposición de los ciudadanos

tengan una visión clara de las necesidades requeridas para los procesos de gestión documental, que les permita saber las implicaciones y retos al momento de implementar una herramienta informática que apoye dichos procesos. Debido a la falta de cultura de datos, el contenido de esta guía se interpreta como los procesos de migración de la documentación física a una de tipo electrónica, dejando a un lado sistemas de compartición de datos. Es decir, los procesos de gestión de los datos publicados por las entidades quedan por fuera de los procesos de gestión documental, a menos que sean parte de una operación estadística que genere la entidad.

Por otro lado, en el PEN se determinan guías para que las entidades adjuntas al SEN implementen. Estas guías, incluyen planes de acción que contienen recomendaciones, estrategias y herramientas para la producción de datos. Además, se plantean estas nueve estrategias para alcanzar los objetivos de producción estadística del país:

- Definición de la oferta de operaciones estadísticas que se deben producir de manera continua.
- Promoción de la difusión y del acceso a la información estadística.
- Fomento al mejoramiento de la calidad de las estadísticas oficiales.
- Identificación y promoción del aprovechamiento estadístico de registros administrativos
- Promoción de la inclusión del enfoque diferencial e interseccional en la producción y difusión de las estadísticas del SEN.
- Promoción de procesos de innovación, aprendizaje y gestión del conocimiento para la producción estadística.
- Implementación de acuerdos sobre la producción de las estadísticas que requiere el país.
- Integración entre la información estadística y la información geoespacial.
- Fortalecimiento de la capacidad estadística territorial.

En cada una de estas estrategias se definen responsables y tareas a realizar para temas como: a) la actualización de la oferta y demanda de información; b) implementación de estándares de difusión y acceso; c) lenguajes de interoperabilidad; d) acceso e intercambio de microdatos; e) metodologías de anonimización de la información; f) calidad de la producción estadística; g) aprovechamiento de registros administrativos generados por entidades públicas y privadas; h) innovación en la producción y difusión estadística; i) fortalecimiento de la articulación entre los miembros del SEN.

Cabe aclarar que dicho PEN no es una política para todas las entidades del Estado colombiano, pero si presenta casi todos los elementos necesarios para realizar una gerencia de datos. En este sentido, aunque no se especifica al detalle cómo implementar dichas estrategias, se establecen períodos, metas, responsables y plazos para cumplir con los objetivos del plan.

Otra herramienta disponible por el gobierno colombiano para la gerencia de datos es la Guía para el Dominio de la Información (Arquitectura TI Colombia, 2014) del MinTIC. En esta, se establece la importancia y la definición del Dominio de Información:

En las organizaciones, la información se ha convertido en el principal generador de valor estratégico. La información se usa para responder a las necesidades de una institución, ya sea para tomar decisiones, para los procesos o los grupos de interés. Este dominio permite definir: el diseño de los servicios de información, la gestión del ciclo de vida del dato, al análisis de información y el desarrollo de capacidades para el uso estratégico de la misma.

Para este dominio se definen 4 ámbitos:

- Planeación y gobierno de los componentes de información: adecuada planeación y gobierno de datos, servicios y flujos de información.
- Diseño de los componentes de información: caracterización y estructuración de los componentes de información.
- Análisis y aprovechamiento de los componentes de información: orientar y estructurar procesos de análisis y toma de decisiones a partir de la información disponible.
- Calidad y seguridad de los componentes de información: definición y gestión de mecanismos y procesos para certificar niveles de seguridad requeridos.

Cada ámbito incluye la descripción de los lineamientos asociados, los roles necesarios para poder cumplir con los lineamientos y artefactos requeridos para que las entidades logren implementarlos exitosamente. Además, los ámbitos definen temas a implementar como gobierno y calidad de la información, gestión de ciclos de vida de la información, intercambio y trazabilidad de información, directorios maestros de datos y metadatos, mapas de sistemas de información y aplicaciones, análisis de demanda, interoperabilidad, entre otros. Todos estos temas tratan desde un punto de vista práctico, la documentación necesaria para cumplir con la implementación, y no como un tema de aprovechamiento de los datos para generar valor, que es el principal objetivo de la gerencia de datos.

3.4.1.5. MODELOS DE CALIDAD DE DATOS

Los criterios de calidad de datos, según el Open Data Barometer, cubren temas de licenciamiento, nivel de acceso (pagos/libres), formatos disponibles, período de actualización, facilidad de búsqueda y enlace con conjuntos de datos relacionados. Aun cuando algunos de estos criterios solo aplican a datos abiertos, da un marco de referencia para evaluar la calidad de datos disponible.

En Colombia, en el momento hay procesos de certificación de la calidad estadística, regidos bajo el Manual de Aseguramiento de Calidad de la Información de las Naciones Unidas (NQAF). Este proceso es liderado por el DANE y permite garantizar que los procesos de producción de la información estadística cumplan con altos estándares internacionales. Sin embargo, es necesario que todas las entidades desarrollen manuales de calidad de datos, pues el criterio preliminar para la explotación de estos son insumos de calidad. Para iniciar, se deben desarrollar protocolos asociados a los metadatos - asegurarse que los datos almacenados estén completos, con variables identificadas y estandarizadas para criterios que apliquen a todas las entidades. Las oficinas de planeación de cada entidad deben velar porque haya calidad y aseguramiento de los procesos de datos, pues la calidad de estos depende de la unidad de análisis con la que se trabaje.

3.4.2. INICIATIVAS DE INTELIGENCIA DE NEGOCIOS Y BIG DATA EN COLOMBIA

En marzo de 2016, se creó la Alianza CAOBA - un centro de excelencia en apoyo del uso de tecnología, conformado por actores del sector académico, público y privado - para posicionar a Colombia como líder en la región en cuanto a Big Data y Data Analytics. Esta Alianza define estos como:

- A. Big Data: es el uso de técnicas para capturar, procesar, analizar y visualizar grandes conjuntos de datos en un tiempo razonable.
- B. Data analytics y Data science: procesos que, junto con los principios y métodos que los soportan, permiten extraer información y conocimiento a partir de datos.

CAOBA busca dar soluciones a problemáticas sociales, industriales y académicas, por medio de 4 ejes estratégicos:

- **Investigación aplicada:** procesos de investigación para desarrollar soluciones aplicables en diversos contextos del país, tanto en el sector público como en el privado.
- **Transferencia del conocimiento:** procesos de vinculación de líderes e investigadores de Big Data y Data Analytics, para apoyo a diferentes sectores.
- **Consultoría tecnológica:** Procesos de acompañamiento a empresas colombianas para facilitar la apropiación de tecnologías, metodologías y herramientas de análisis de datos enfocadas a Big Data.
- **Apoyo al emprendimiento:** promover y apoyar iniciativas de emprendimiento en Big Data y Data Analytics.

Para construir sobre el éxito y desarrollo de CAOBA, es necesario ampliar los Centros de Excelencia y Apropiación alrededor del país, para así promover los ejes estratégicos, en particular la investigación y el emprendimiento, en construcción de una cultura de datos nacional.

3.4.3. ¿HACIA DÓNDE SE APUNTA?: INFRAESTRUCTURA DE DATOS EN LA POLÍTICA NACIONAL DE EXPLOTACIÓN DE DATOS

Dentro del Documento CONPES 3920 Política Nacional de Explotación de Datos (Big Data) se definen las siguientes acciones para impulsar el desarrollo de la infraestructura de datos:

- Definir y publicar un documento de la **infraestructura de datos públicos** y las condiciones requeridas para su implementación y aprovechamiento, con el fin de dar guías y estándares a las entidades. Dentro de esta definición se incluirán temas como necesidades de colaboración público-privada, acceso a la información pública, mantenimiento y manejo de propiedad intelectual y privada.
- Crear **mecanismos de gestión** en las entidades públicas para certificar la calidad del dato que permita validar los procesos de generación, recolección y difusión masiva de los datos, al igual que los responsables de la generación de datos.
- Un mecanismo de difusión para cuantificar la tasa de publicaciones de datos, con el fin de estar monitoreando constantemente la disponibilidad de datos en las entidades.
- Definir y/o crear procesos estándares de anonimización, como forma de tratamiento de los datos.
- De la mano con el documento de infraestructura de datos público, para asegurar la calidad, se plantea definir y adoptar estándares y lineamientos técnicos que garanticen la **calidad de los datos** dentro de las entidades. Esto implica definir el ciclo de vida y automatización de los procesos de calidad para su generación.

- Crear guías para definir el lenguaje común para la recolección y generación de datos, incluyendo formatos soportados para la descarga e interfaces de programación de sistemas de información. Dentro de la propuesta, los lineamientos deben incorporar las recomendaciones del Sistema Estadístico Nacional (SEN) para el uso estadístico de los registros administrativos.
- Apoyar la implementación de **interoperabilidad** que mejor cumpla con los criterios técnicos y económicos, facilitando así su implementación. Esta implementación debe considerar aspectos de capacidad de servicio, unidad semántica, disponibilidad de un sistema centralizado o el uso de tecnologías como Blockchain.

Todas estas iniciativas y estrategias buscan afianzar la arquitectura de datos, así como promover cultura de datos, desarrollar las necesidades dentro de las entidades de Business Intelligence y crear proyectos de Big Data, más allá de la obligación de cumplir los requerimientos mínimos de ley.

3.4.4. DIAGNÓSTICO SOBRE LA ARQUITECTURA DE DATOS

En general, se identifica que la arquitectura de datos actual del país aún no trata la información como un valor estratégico. Existen varias deficiencias en la gestión de proyectos asociados a la arquitectura de datos nacional, lo que conlleva a una duplicidad de esfuerzos y recursos innecesaria.

Aun cuando se encuentra información de políticas, manuales y guías para implementar sistemas de interoperabilidad, y se utilizan procesos de intercambio de información - como se evidencia también en las entrevistas realizadas - no todas las entidades han desarrollado manuales o wikis de los sistemas compartidos, que le permitan a las entidades con las que están interoperando tener disponible la información de cómo funcionan los sistemas dentro de esta entidad. En este sentido, sí hay digitalización de la información y posibilidades de intercambio, más no hay orden y medidas de calidad en esta transferencia. Además hay un cumplimiento mínimo en cuanto al lenguaje común de información, donde sólo una proporción baja de las entidades ha alcanzado la adopción completa de este lenguaje, y por lo tanto, de la interoperabilidad. Es por esto que se asume que los sistemas de intercambio de información no han sido pensados para la interoperabilidad completa con terceros, ya que es difícil poder desarrollar sistemas de consumo de esta información en la actualidad, sin tener que recurrir a contactar a las entidades directamente.

En cuanto a gestión de datos, el rol en este momento del Archivo General de la Nación en cuanto a explotación de datos, no propone una gestión comprehensiva como la que se requiere para esta. El PEN tiene un precedente importante, pues implementa los elementos necesarios para realizar una eficiente gerencia de datos, pero es una acción aislada, pues solo las entidades asociadas al SEN siguen estos lineamientos. Por otro lado, aún no existen protocolos estandarizados de calidad de datos no-estadísticos en el país, aun cuando este concepto ya haya sido abordado por varias entidades, entre estas el MinTIC y el DANE. La explotación de datos depende de una arquitectura de datos flexible, que se adapte a las necesidades de explotación de datos. Como los sistemas de información y las tecnología que usan las entidades no son las mismas, es necesario desarrollar protocolos internos a cada entidad, así como protocolos de calidad para interoperabilidad, y de esta manera, garantizar máxima calidad de los datos producidos a nivel gubernamental. para garantizar la compatibilidad de la información puesta a interoperar.

3.5. DIMENSIÓN 5: CAPITAL HUMANO Y CULTURA ALREDEDOR DE LOS DATOS

Para esta dimensión, el capital humano se analiza desde el punto de las habilidades o capacidades técnicas necesarias para poder generar valor a partir de los datos, al igual que la demanda de estos perfiles en sectores públicos y privados. Primero se define la cultura basada en datos o *data driven society*, y luego se realiza un mapeo general de la brecha de talento digital en Colombia teniendo en cuenta los roles y salarios relacionados a información y específicamente a la analítica de datos en el gobierno. Luego, se exponen los roles principales que tanto las empresas privadas como entidades públicas deben tener para la explotación de datos, con la intención de esclarecer el capital humano que se debe desarrollar en un futuro para una cultura alrededor de los datos.

3.5.1. DEFINICIÓN DE UNA CULTURA BASADA EN DATOS

Pese a tener una masa crítica de capital humano, los esfuerzos en talento pueden ser en vano si no hay una cultura orientada a los datos o ecosistema que aproveche los datos. Según la OCDE, la cultura de datos es una condición necesaria en sectores públicos y privados para que todo el personal y los tomadores de decisiones se concentren en la información transmitida por los datos existentes. Esto supone que todas las personas y entidades involucradas emplean consistentemente los datos para realizar procesos de toma de decisiones.

Data Therapy, un blog del proyecto MIT Center for Civic Media's Data Therapy, propone 4 definiciones de cultura de datos (Data Therapy, 2017):

- Los directivos priorizan e invierten en la recolección y gestión de datos para realizar análisis o producir conocimiento.
- Los directivos priorizan la creatividad en la alfabetización de datos para toda la organización, no solamente para las oficinas de tecnología o de control.

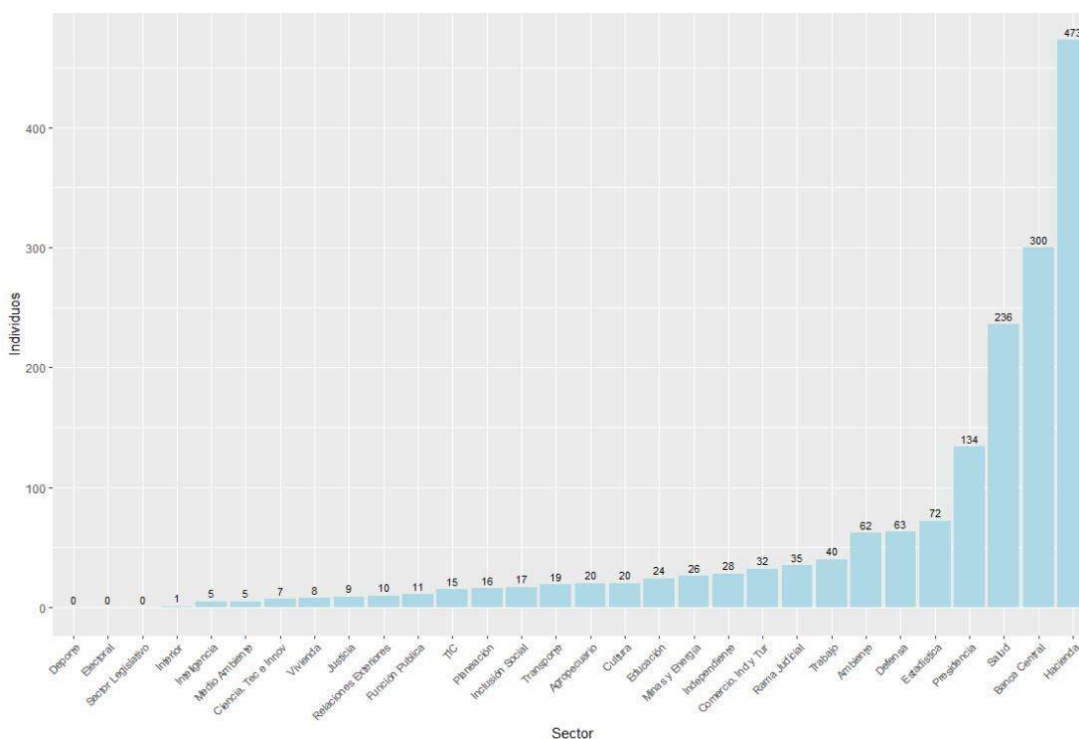
- El personal es motivado y apoyado a acceder, combinar y obtener información de los datos.
- El personal reconoce los datos cuando los ve, lo que ofrece formas creativas para su uso, solucionar problemas, tomar decisiones y contar historias.

Con base en lo anterior, se puede definir cultura de datos como una serie de comportamientos y actitudes enfocadas al consumo de datos para la toma de decisiones. Esta dimensión, se enfoca en revisar el estado actual del gobierno en cuanto a los recursos humanos que alimentan esta cultura, así como algunas de las estrategias actuales que buscan fortalecer, promover o crear una cultura de datos.

3.5.2. ESTADO ACTUAL DE LAS CAPACIDADES TÉCNICAS Y HABILIDADES ENTORNO A LOS DATOS EN COLOMBIA

De los 30 sectores y las 163 entidades analizados en la Encuesta de Explotación de Datos del DNP (2017), se encontró que la mayoría de sectores cuentan con individuos dedicados a la analítica de datos. Se puede observar que los sectores en los que más se concentran personas dedicadas a la analítica de datos son Hacienda, Salud y el Banco Central. En total, la encuesta arroja que hay 1688 personas dedicadas a los datos. Es importante esclarecer que estos resultados no son robustos, pues la encuesta no informa sobre el número de funcionarios en cada una de las entidades encuestadas, por lo cual no se puede hacer una aproximación correcta a la proporción de funcionarios dedicados a la analítica de datos por entidad, y tampoco es exhaustiva en cuanto a las entidades encuestadas, lo que quiere decir que este resultado puede ser producto del tipo de entidades que se encuestaron por sector, aun cuando más del 50% de entidades pertenecientes a cada sector fueron encuestadas. Sin embargo, estos resultados ayudan a identificar sectores en donde claramente se prioriza la analítica de datos como parte de su función, y otros donde se observa un vacío en esta función.

Figura 17: Individuos dedicados a la analítica de datos por sector¹⁴



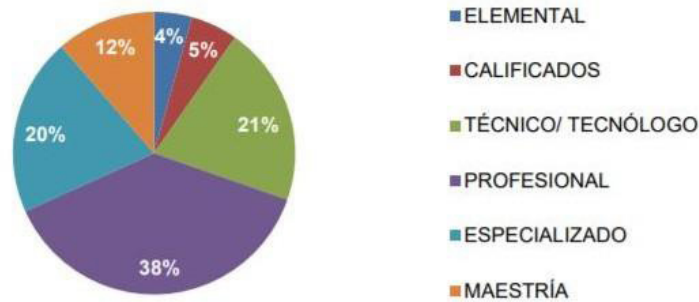
Fuente: Elaboración Propia - Encuesta de Explotación de Datos, DNP 2017

De los individuos reportados en cada entidad, el 21% se identifica como ingenieros, el 27% como economista, el 12% como estadístico, el 7% como profesional de las ciencias sociales, y el 32% restante, en otras profesiones varias. Estos resultados corresponden generalmente, a las profesiones que tradicionalmente se asocian al manejo de los datos.

Por otro lado, la encuesta realizada por el Observatorio TI en el 2015 “Estudio de Caracterización del Sector de Teleinformática, Software y TI en Colombia” identificó el nivel de educación de personas empleadas en el sector TI en empresas públicas. Al no estar estos datos disponibles para el sector público, se utilizan estas cifras como líneas guía para entender la distribución del nivel de formación de este tipo de empleados a nivel nacional.

¹⁴ Para ver ilustración que detalla porcentaje de entidades encuestadas por sector, ir al Anexo D

Figura 18: Nivel de Formación Sector Teleinformática, Software y TI 2015

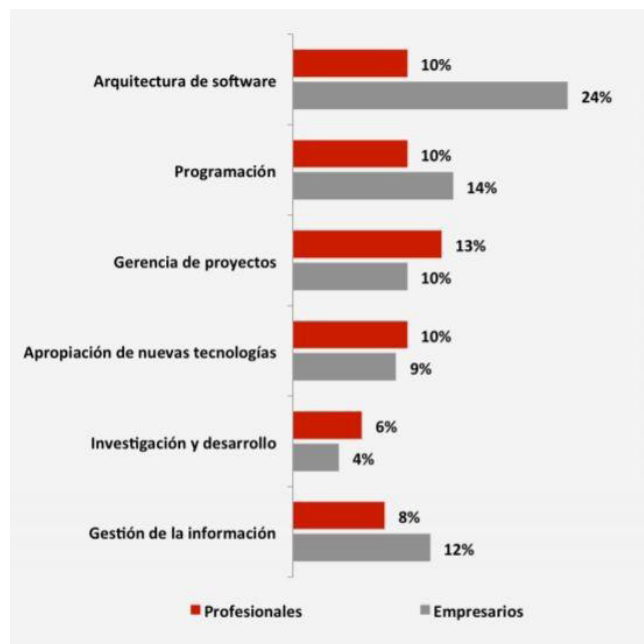


Fuente: Encuesta Estudio de caracterización ocupacional del sector de Teleinformática, Software y TI en Colombia, 2015

Vale la pena resaltar que la mayoría de trabajadores en este sector se han formado como profesionales o han accedido a educación especializada o maestrías según las empresas encuestadas. Generalmente, el nivel de formación profesional se relaciona con la proficiencia en el manejo y procesamiento de datos — esto sin obviar que gran parte de las capacidades digitales que se necesitan para realizar explotación de datos no requieren de educación formal. Sería de gran relevancia desarrollar este estudio en el sector público para evaluar qué tipo de perfiles académicos y niveles de formación tienen los funcionarios que están manejando los datos, para así lograr identificar la brecha en alfabetización digital.

El Estudio de Caracterización de la Brecha de Talento Digital en Colombia (2017) también del Observatorio TI ayuda a ahondar este análisis. Este estudio mide la brecha cualitativa de talento digital como el “conjunto de diferencias identificables entre los conocimientos, habilidades y competencias con que el sistema educativo forma el talento digital del país, y aquellos que realmente requiere la industria TI”. Estos resultados, en la Figura 19, muestran la diferencia porcentual de lo que es prioridad en conocimiento para los empresarios, versus lo empleados. Vale la pena rescatar las diferencias que existen entre la prioridad asignada a la arquitectura de software y programación, resultados que podrían ser extendidos al sector público si se asume que las capacidades necesarias son similares. Este estudio también concluye que la mayor brecha a nivel regional sobre las habilidades en TI son aquellas relacionadas a seguridad y riesgo, capacidades que aún faltan por desarrollar (Observatorio TI y Fedesoft 2017). Estos resultados son particularmente relevantes para el sector público, al ser la ciberseguridad una de las prioridades del manejo de datos para el gobierno.

Figura 19: Brecha de prioridades sobre conocimientos y habilidades durante entre profesionales y empresas del sector TI.



Fuente: Observatorio TI, 2017

3.5.2.1. ANÁLISIS DE SALARIOS

De acuerdo a los estudios del Observatorio TI (información detallada en la dimensión 6), el salario promedio para empleados del sector TI en Colombia varía según el cargo y tamaño de la empresa — la Tabla 11 detalla estos salarios. De acuerdo a estos salarios, se observa un desnivel entre los salarios que se ofrecen localmente frente al mercado laboral exterior para empresas privadas. La diferencia salarial promedio es observable comparando roles similares como Líder de Calidad, Administrador de Base de Datos y Gerente de Desarrollo con los salarios promedios de la encuesta hecha por el Observatorio de TI; incluso después de hacer una normalización de los salarios ajustándolos por el poder adquisitivo del salario (PPP 2012) y controlando por la diferencia promedio observada por la organización mundial del trabajo (ILO). Los resultados muestran que los equivalentes internacionales seguían ganando 2.0x y 2.5x más que sus homólogos locales.

En este contexto, es posible que en Colombia las necesidades en capital humano en general, no se reduzcan pues los trabajadores encuentran bajos incentivos salariales para trabajar en el sector. Esto mismo aplica al sector público, dónde los incentivos salariales no son sujetos a los multiplicadores ni cambios dinámicos del sector privado. Esto, tiene como efecto secundario la fuga del talento, afectando el desnivel de capacidad ya existente entre países.

De lo anterior, se puede concluir que la mayoría de las personas relacionadas a sectores de TI no obtienen una remuneración mayor a 5 SMLV en Colombia. Esto corresponde a una motivación monetaria muy baja para proyectos en TI en los que el capital humano debe estar altamente calificado.

3.5.3. ESTADO ACTUAL DE LAS ESTRATEGIAS EXISTENTES PARA PROMOVER Y FORTALECER LA CULTURA DE DATOS

En cabeza del MinTIC y Colciencias, el gobierno promueve la capacitación en TI desde el nivel técnico hasta el gerencial mediante becas, convocatorias y cursos de libre acceso. El Plan Vive Digital 2014-2018 del MinTIC contaba con una estrategia de Talento TIC que buscaba abastecer las necesidades de personal TI en el país. Al crear capital humano capacitado en tecnología e información se enfatiza el fortalecimiento de una cultura de datos, arraigada en un ecosistema con oferta y demanda de individuos especializados en esta industria. Para cumplir con las metas planteadas en la estrategia de Talento TIC se proponían las siguientes iniciativas:

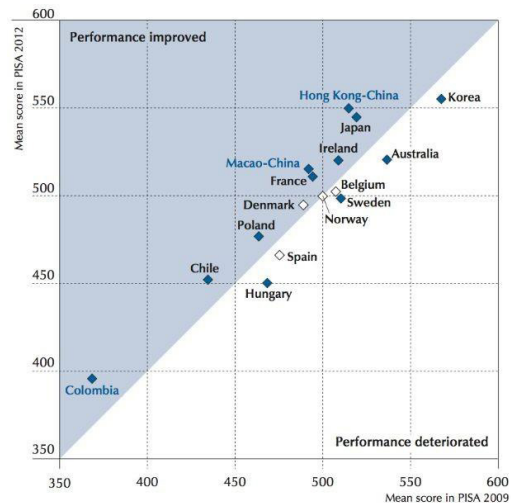
- Crear programas de incentivos para realizar estudios de alto nivel (maestría y doctorado) en el sector de TI.
- Estimular la coordinación entre la academia y la industria de TI para facilitar la empleabilidad de los egresados.
- Trabajar en conjunto con la industria para realizar recomendaciones en torno al fortalecimiento de la enseñanza de las áreas de matemáticas, ciencias, lenguaje e informática en educación básica y media.
- Desarrollar un plan nacional de formación para fortalecer la generación de talento TI en el país desde la educación básica primaria hasta la educación superior.
- Desarrollar competencias específicas para líderes TI y funcionarios.
- Certificar a funcionarios públicos y ampliar la oferta de programas para estos funcionarios.
- Implementar la función del GCIO (Government Chief Information Officer) o líder de tecnología para las entidades del Estado.
- Promover prácticas especializadas en el sector Gobierno para potenciar el aprovechamiento de las TI en beneficio de la gestión pública.

Sobre esta iniciativa de Talento TI, en el Informe de gestión al Congreso de la República para 2017 (MinTIC, 2017, p. 28), se reporta que “855 funcionarios culminaron sus programas de Educación no formal (cursos, diplomados, talleres y certificaciones). De otra parte, a la fecha, 210 funcionarios están cursando programas de Educación Formal (especialización o maestría), en temáticas de Gestión TI y Seguridad y Privacidad de la Información”. Esto muestra una continua capacitación a los funcionarios del Estado, mejorando la disponibilidad de talento en las instituciones públicas. Estas capacitaciones se complementan con becas proporcionadas por Colciencias para áreas de TI y temas de Big Data. La capacitación de 200 ciudadanos como *Citizen Data Scientists* realizada entre la Alianza CAOBA, Colciencias y MinTIC, es un ejemplo de estos programas.

Por otro lado, en el 2012 la OECD, bajo su programa PISA (Programme for International Student Assessment), planteó los siguientes problemas para el desarrollo de habilidades de TI en los estudiantes en Colombia (OCDE, 2015):

- El 37% de los estudiantes de 15 años no tenían acceso a computadores en la casa.
- Los estudiantes de 15 años se desempeñan por debajo del promedio de los países de la OECD en Lectura Digital (Digital Reading).
- Los estudiantes en Colombia tienden a tener peor desempeño en Lectura Digital que los estudiantes de otros países que tienen un desempeño similar en habilidades de lectura impresa.

Figura 20: Desempeño de Lectura Digital para 2009 y 2012



Fuente: PISA: Students, Computers and Learning,

Los resultados PISA, evidencian que Colombia cuenta con niveles de analfabetismo digital, que se ve reflejado en el capital humano que se encuentra en las empresas públicas y privadas. Este analfabetismo digital es un factor en la falta de apropiación tecnológica necesaria para crear una cultura de datos ya que, para este fin, asimilar rápido la tecnología es vital. Para contrarrestarlo, el gobierno colombiano realiza periódicamente capacitaciones, cursos, talleres y conferencias alrededor del tema de explotación de datos y cultura de datos. De igual forma se puede obtener, principalmente en el MinTIC, manuales y guías para la interoperabilidad y el consumo de datos. Según el Informe de Gestión al Congreso de la República para el 2017, el MinTIC dentro de la iniciativa de promover la apropiación de las TIC por parte de la ciudadanía, describe que “en el periodo que registra este informe se beneficiaron 156.610 personas con el desarrollo de procesos de formación en competencias y certificación en TIC, con el apoyo de la red y el equipo de Puntos Vive Digital y Kioscos Vive Digital de la Dirección de Conectividad.”

Adicionalmente, se entregaron 1547 certificados a 1159 personas en procesos de formación virtual en habilidades y competencias TIC a través del portal de Ciudadanía Digital. Además, desde marzo de 2017 se efectuó una alianza con la empresa Capacitate.co con el fin de multiplicar la oferta de contenidos. A través de esta alianza se han formado y certificado más de 500 personas, para un acumulado de 1.416.788 certificados. Estas capacitaciones buscan disminuir el analfabetismo digital y, aunque están menos enfocados en el uso y generación de datos, es un avance para poder crear un ecosistema de datos.

También existen iniciativas específicas para la promoción y uso de los datos, como DC Datos de ViveLab, la cual es una iniciativa para coordinar esfuerzos del sector público, privado, academia y ciudadanía en temas de tecnología. Como lo describen en su página web: “*En el ViveLab Bogotá unimos al gobierno nacional y distrital, empresas, academia y ciudadanos para promover la innovación, creando soluciones digitales de alto impacto social y económico. Trabajamos por transformar nuestra ciudad usando la tecnología, el empoderamiento ciudadano, el desarrollo sostenible, la modernización del gobierno y las actividades productivas, a la vanguardia de lo que nos demanda la 4ta revolución industrial. Los principios de la innovación ágil y colaborativa, con un enfoque centrado en los usuarios, definen las acciones que realizamos en nuestro laboratorio*” (VivelabBogotá, 2018). De igual manera organizan DataJams, los cuales buscan crear productos digitales basados en datos abiertos enfocados en ayudar a los ciudadanos a tomar decisiones a partir de la visualización de datos. Dentro de estos, también se realizan procesos de capacitación en tecnología y en temas relacionados al manejo y visualización de datos.

Por otro lado, el SEN tiene como una de sus estrategias, la “Promoción de procesos de innovación, aprendizaje y gestión del conocimiento para la producción estadística”. Para cumplirlas, promueve grupos de trabajo, tanto nacionales como internacionales, para realizar actividades relacionadas con la generación de capacidad estadística. Pese a que no se establece exactamente las áreas de dichas capacitaciones, se puede inferir que están directamente ligadas a la explotación de datos, tanto a producción como a estándares para compartir datos. Estas capacitaciones son herramientas que pueden ser aprovechadas por todas las entidades del estado para generar un ecosistema de datos, independientemente si pertenecen o no al SEN.

3.5.4. CAPITAL HUMANO ORIENTADO A LA EXPLOTACIÓN DE DATOS

En general, el recurso humano para la explotación de datos usualmente es entendido como los profesionales en tecnología que permiten la generación y la disponibilidad de los datos. Al hablar de datos, el primer recurso humano en los modelos empresariales es el Científico de Datos (*Data Scientist*), quien tiene el rol de elaborar modelos para entender problemas

o procesos relacionados con la empresa o entidad. Independiente de las definiciones o responsabilidades asignadas, los científicos de datos utilizan la estadística, algoritmos, minería de datos y técnicas de visualización de datos para mostrar sus resultados o brindar herramientas para analizar los datos.

Debido al fuerte uso de tecnologías de la información (TI), la explotación de datos también requiere de una variedad de ingenieros de sistemas, y científicos de datos, que puedan soportar los servicios de información, las bases de datos y la interoperabilidad. Para una estrategia exitosa de Big Data, es necesario involucrar diferentes roles multidisciplinarios para mejorar el aprovechamiento de los datos. La revista Harvard Business Review (2013) define 5 roles necesarios para crear un equipo de Big Data:

- **Curador de datos:** se encarga de asegurar que los datos que se obtienen estén limpios y precisos y se mantengan así durante todo su ciclo de vida.
- **Exploradores de datos:** se encarga de examinar cuidadosamente los datos para descubrir cuáles son los que se requieren para los proyectos o las necesidades de los usuarios.
- **Arquitectos de soluciones empresariales:** se encargan de agrupar y organizar los datos descubiertos para que estén listo para ser analizados.
- **Científicos de datos:** se encargan de manipular los datos organizados y crear modelos analíticos sofisticados que permitan obtener valor o información de estos datos.
- **Expertos de campaña:** se encargan de tomar los modelos y ponerlos en resultados tangibles, estos también pueden ser llamados *data translators*.

Aun cuando no existe un solo prototipo de un equipo de datos y analítica, el cubrir las tareas de estos roles cubre el ciclo productivo del dato, haciendo posible generar valor agregado a partir de la explotación de datos. Es importante entender que estos roles no pertenecen a una profesión o un perfil específico, sino que generalmente son profesionales de diferentes áreas que desarrollan las habilidades necesarias para el manejo de datos, o al menos entender los procesos, para poder obtener el mejor valor agregado de ellos.

En Colombia, la oferta de programas de educación superior relacionados a la producción y aprovechamiento de datos se concentra en sistemas de información y desarrollo de software. De acuerdo a las bases de datos del Ministerio de Educación (2017) solo 75 instituciones oficiales ofrecen programas relevantes a la explotación de datos, de las cuales cinco ofrecen maestrías y sólo una ofrece un doctorado¹⁵. La Tabla 10 resume los programas relacionados a datos por el nivel de especialización de estos. Entre estos, hace falta más énfasis en habilidades de analítica avanzada, ciencia de datos y programas dedicados a la computación y a la programación. De igual manera, es importante resaltar la ausencia de estandarización en los programas ofertados; por ejemplo, la oferta de programas de desarrollo de software e ingeniería de software sin la armonización de estos bajo un sólo título o currículo.

Tabla 10: Oferta de programas relacionados a explotación de datos en Colombia

NOMBRE CARRERA	DOCTOR ADO	MAESTRÍA	ESPECIALIZ ACIÓN	ESP. TECNOLÓGICA	TÉCNICA PROFESIONAL	# INSTITUCIONES
INGENIERIA DE SISTEMAS Y COMPUTACION		1	1	5		7
DESARROLLO DE SOFTWARE			3	12		15
INGENIERIA DE SOFTWARE			3			3
EN TECNOLOGIAS DE LA INFORMACION						
APLICADAS A LA EDUCACION			1			1
IMPLEMENTACION DE SISTEMAS DE INFORMACION GEOGRAFICA				4		4
DESARROLLO DE SOFTWARE				8		8
PRUEBAS DE SOFTWARE				2		2
MATEMATICAS Y COMPUTACION				1		1
BIOINFORMÁTICA Y BIOLOGÍA COMPUTACIONAL		1				1
INGENIERIA COMPUTACIONAL		1				1
PEDAGOGÍA DE LAS TECNOLOGIAS DE LA INFORMACION Y LA COMUNICACION		1				1
TECNOLOGIAS DE LA INFORMACION APLICADAS A LA EDUCACION		1	1			2
SISTEMAS Y COMPUTACION					1	1
PROGRAMACIÓN DE SISTEMAS DE INFORMACIÓN				1	2	3
SOPORTE DE SISTEMAS DE INFORMACIÓN					1	1
COMPUTACION				1		1
DESARROLLO						
(ADMINISTRACION/PROGRAMACIÓN/ANALISIS) DE SISTEMAS DE INFORMACION				3		3
GESTIÓN DE SISTEMAS DE INFORMACIÓN				1		1
SISTEMAS DE INFORMACION EN SALUD				2		2
SISTEMAS DE INFORMACION				17		17
TOTAL						75

Fuente: Ministerio de Educación 2017 (datos.gov.co) - Elaboración propia

¹⁵ Se refiere a habilidades técnicas relacionadas a la explotación de datos. El criterio utilizado en la búsqueda incluye programas que contengan las palabras: sistemas de información, programación, computación, analítica, datos, web, desarrollo tecnologías de la información.

Por otro lado, tras revisar las becas otorgadas por el Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el exterior (ICETEX 2018) se encontraron apenas 36 de 1.466 becas con estudios de posgrado para temas relaciones a datos, incluyendo temas de gobierno abierto, desarrollo de software e ingeniería de redes en internet.

3.5.5. ¿HACIA DÓNDE APUNTA?: CAPITAL HUMANO ORIENTADO AL ANÁLISIS Y GESTIÓN DE INFORMACIÓN

El Documento CONPES 3920 Política de Nacional de Explotación de Datos (*Big Data*), plantea como uno de los objetivos específicos disponer de capital humano para generar valor con los datos, lo que se alinea con la necesidad de crear cultura de datos y un ecosistema de explotación de estos. Dentro de este documento las acciones propuestas son:

- Armonización de la demanda y la oferta de capital humano para aumentar la fuerza laboral y los perfiles para manejo y análisis de datos.
 - Nivelar el alistamiento para la explotación de datos en las entidades públicas.
 - Medición de la brecha de capital humano y actualización de competencias.
 - Promover el emprendimiento de bienes y servicios basados en la explotación de datos.
- Creación de incentivos y condiciones óptimas que aumenten el aprovechamiento de los datos.
 - Generar los mecanismos para materializar el valor de los datos en las entidades de la administración pública.
 - Institucionalizar la explotación de datos en la toma de decisiones públicas.
 - Definir, implementar y dinamizar el funcionamiento del mercado de datos en Colombia.
 - Vincular la ciudadanía a la infraestructura de datos.

3.5.6. DIAGNÓSTICO DIMENSIÓN 5

En Colombia, no es muy clara la condición actual del capital humano en las entidades públicas, pero sí es claro que en general, existen una brecha tanto de oferta para la industria TI, como en incentivos y conocimientos. De acuerdo a la investigación realizada, hay varias entidades que cuentan con la capacidad de analítica de datos, pero no se conoce cómo se están utilizando los datos, o si verdaderamente el número de personas que se atribuyen a la analítica de datos están en capacidad y función de explotación de datos. Por ejemplo, el Banco de la República respondió que existen 300 individuos dedicados a la analítica de datos, pero puede que esta función sea muy diferente de la que se le atribuye a un analista de datos en otra entidad. Lo que sí es posible acertar, es la necesidad de desarrollar mecanismos de medición del capital humano en el sector público, quizá al replicar los estudios que ha realizado el Observatorio TI para empresas del sector privado. Por último, hace falta calificar de manera concreta la brecha cuantitativa de trabajadores de TI en Colombia, y en el sector público específicamente.

De igual manera, los salarios para posiciones dedicadas a TI son muy bajos en comparación con roles similares en países con industrias avanzadas de Big Data. La mayoría de trabajadores en estos cargos no ganan más de cinco salarios mínimos. En contraste, en Estados Unidos, un analista de datos en promedio se gana \$95,000 lo que corresponde a más de 60 veces el salario mínimo (sin embargo, este varía por estado). En el Reino Unido, el salario promedio de un científico de datos es de €50,000 euros. Los bajos salarios en Colombia pueden incidir en una fuga de talento en estas posiciones hacia el sector privado, y limitar los incentivos para el desarrollo de estas habilidades. Es importante resaltar también que el alto número de participación de contratistas en el sector de TI - problema identificado a través de las entrevistas realizadas - puede generar también falta de continuidad y momentum en el desarrollo de capital humano y una cultura alrededor de datos.

Por último, aun cuando existen iniciativas e inversiones en fortalecer el capital humano en Colombia, estas son limitadas y de corto alcance en cuanto a lo que requiere un ecosistema de datos avanzado. En Singapur, por ejemplo, se ofrece educación en programación a través de programas innovadores virtuales de alto alcance, como el SkillsFuture o Smart Nation Fellowship Programme. En México, el presupuesto general para proyectos digitales del Consejo Nacional de Ciencia y Tecnología es de 6 millones de dólares, y en Corea del Sur, aprender programación avanzada es obligatorio en la escuela.

La ausencia de habilidades digitales - desde las más básicas, a analítica de datos avanzada - en el gobierno, inhibe el progreso y desarrollo de la transformación de los servicios a los que las diferentes estrategias de gobierno electrónico del país le han apuntado. Oportunidades de capacitación para funcionarios del gobierno son esenciales, al ser este uno de los principales productores de datos a nivel nacional. Para lograr que el gobierno aproveche los datos, es necesario unir la brecha entre las posibilidades de análisis que brinda la tecnología y las capacidades analíticas de sus funcionarios.

3.6. DIMENSIÓN 6: GASTO E INVERSIÓN ASOCIADA A LA EXPLOTACIÓN DE DATOS

3.6.1. DEFINICIÓN DE MARCO ECONÓMICO PARA LA EXPLOTACIÓN DE DATOS

El Marco económico analiza las variables y factores esenciales para medir el impacto económico del sector de datos en Colombia. Aun cuando realizar esta medición no está en el alcance de este diagnóstico, identificar los componentes del Mercado de datos permite identificar cómo los datos y la información, desde el sector público, son de importancia para la economía nacional. La industria de los datos tiene el potencial de cambio y crecimiento que generó en el siglo pasado el

petróleo, ya que está generando nuevas infraestructuras, nuevos negocios, nuevas políticas públicas y nuevas economías. Sin embargo, por su naturaleza abstracta los países han tenido dificultades para medir la economía de datos y observar su evolución.

Para empezar a entender la manera en que el mercado de datos impacta la economía nacional, tomamos como ejemplo el ejercicio realizado en 2016 por la Comisión Europea, quienes con el objetivo de medir y evaluar las tendencias e impacto de los mercados de datos en las economías nacionales, publicaron una serie de herramientas para medir el impacto macroeconómico generado por la economía de datos (Comisión Europea 2016).

En este ejercicio, se partió de la observación que en las economías nacionales existen actores que generan ingresos sobre datos en bruto. Para este colectivo se propone usar la definición de “Mercado de Datos” y se proponen como guía las siguientes definiciones:

- **“Mercado de Datos”**: Este es el mercado donde los datos digitales son intercambiados como productos o servicios derivados de los datos en bruto. Para estimar la actividad que es relevante a los agentes, se propone atribuir al Mercado toda actividad que involucre la creación, colección, almacenamiento, procesamiento, distribución, análisis, elaboración, entrega y explotación de los datos permitida por las tecnologías digitales. Sin embargo, la Comisión Europea propone adicionar los impactos indirectos e inducidos por este mercado de datos para definir la totalidad de la economía de datos. Por lo tanto, el efecto de una economía de datos sobre una economía nacional estaría compuesta por tres elementos de análisis:
- **“Impacto Directo”**: los impactos generados por la industria del dato; esto representa la actividad total generada por los negocios que están activos en la producción de datos. La forma cuantitativa de medir el impacto directo sería atribuyendo los ingresos generados por las ventas directas de productos o servicios de datos. En esta definición se consideran actores directos de la industria de datos como Facebook, pero no incluiría los rendimientos económicos por el mejoramiento de procesos a través de la utilización de datos (Ejemplo: Uber incrementando sus ingresos basado en el análisis de tendencias). Por lo tanto, este término resulta equivalente/sinónimo al “*Mercado de Datos*”.
- **“Impacto Indirecto”**: se define como la suma de impactos que se originan en otras industrias asociadas a la industria de datos. En este caso, se analiza verticalmente este efecto; “Aguas arriba” o “hacia atrás”¹⁶, lo cual es el efecto indirecto que genera la economía del dato en los proveedores del mismo. Es decir, estos son los insumos requeridos, además de los datos en bruto, por el Mercado. La Comisión Europea decide medirlo por la variación en ventas adicionales a la industria de datos. Esta medición es la que puede capturar el mayor impacto del mercado de datos, pero también donde más dificultades metodológicas se identifican.
- **“Impacto Inducido”**: estos son los impactos generados por la actividad económica derivada de los salarios de la industria de los datos. Una porción de estos salarios es reincorporada en la economía en otros bienes y servicios. La Comisión Europea estima necesario incluirla dentro de la medida del valor total de la economía de datos considerando que representó casi el 26% del valor estimado de esta en 2016 (79.7 mil millones de Euros para EU28)

Esos diferentes tipos de impacto, incluyen los flujos de caja futuros generados por los agentes del sector (servicios de identificación de usuarios objetivos para mercadeo), el impacto económico generado a los usuarios (las eficiencias y reducciones de costos por la utilización de datos) y las ventas de productos y servicios requeridos por la industria del dato (infraestructura para el almacenamiento del dato).

Para entender el mercado de datos, es necesario entender la relación entre los costos y el precio de los datos. Esta relación, será particularmente útil para el análisis del impacto económico en el tercer punto señalado anteriormente: “Ingresos generados por los proveedores de la industria”.

Cómo es bien sabido, los costos de los datos, se dividen en los siguientes:

- Los costos de *Inversión*: definidos como la suma de costos que se incurren en la adquisición de los activos fijos necesarios para poner el proyecto en funcionamiento. Estos, incluyen activos tangibles (servidores, almacenamiento, espacios físicos, etc.) e intangibles como desarrollo de software, licencias, propiedad intelectual, entre otros. Estos, son costos capitalizados; por lo tanto, el impacto financiero del activo se reparte en el tiempo, lo cual tiene asociado un efecto de depreciación y amortización.
- Los costos de *Operación y Mantenimiento* son la suma total de gastos fijos y variables en materiales, recursos humanos, insumos y servicios necesarios para garantizar la operación, mantenimiento y soporte de la plataforma tecnológica, las aplicaciones, los sistemas de información y servicios informáticos.

Teniendo estos costos en cuenta, cabe resaltar que cada agente racional buscará maximizar sus ingresos, cubriendo sus costos operativos y generando ingresos rentables que permitan sostener en el tiempo la estructura operativa y obtener beneficios. El precio “fundamental” de los datos deriva de la totalidad de costos generados en la instalación y en el mantenimiento de

¹⁶ Del inglés “upstream” o “backwards”, se intenta capturar el valor de los insumos necesarios al procesamiento y almacenamiento de datos

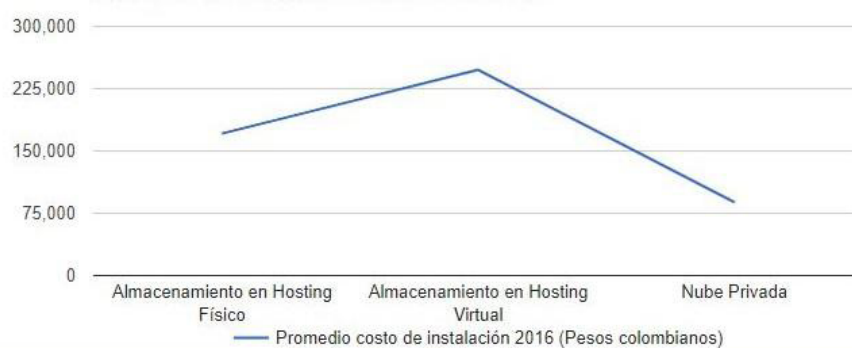
la infraestructura de datos, por la cantidad de unidades almacenadas. Al atribuirle un valor mínimo o intrínseco a los datos, todo agente racional buscando “aprovechar” estos datos, deberá mantener operacional el sistema que los contiene. Este factor precio del dato, será transferido a lo largo de la cadena hasta el usuario de los datos.

Sin embargo, esta afirmación tiene dos limitaciones exacerbadas por la economía de los datos. La primera, es que el análisis del costo del dato no permanece constante en el tiempo, pues depende de la infraestructura de almacenamiento. La segunda, es que el valor del dato no es producto únicamente de su almacenamiento, sino también del análisis y los beneficios que se puedan producir de su utilización.

La estructura de costos evoluciona atada a un elemento de costo variable. Ante un incremento en requerimientos computacionales se considera cómo minimizar los costos estructurales y esto se puede lograr aprovechando la variedad de opciones en términos de tipo de infraestructura para el almacenamiento y el procesamiento de datos. Esto, lleva a agentes a estar en una evolución constante para lograr mezclar infraestructura propia (o “tradicional”) e infraestructura en la nube. Las diferentes combinaciones en infraestructuras de datos generan estructuras de costos únicas que se traducen en el costo unitario de los datos. Teniendo esto en cuenta, un mismo dato podría encontrarse con dos valores por costo diferentes dependiendo del controlador y de los requerimientos de procesamiento de datos en el tiempo en el que se encuentra.

Los agentes económicos variarán con el tiempo, el tipo de infraestructura para almacenar y procesar los datos para hacerla más productiva. Al invertir en infraestructura propia la empresa adquiere una capacidad de procesamiento y almacenamiento de información fija, adicional a las capacidades que ya tiene, teniendo como resultado la disminución del costo operativo unitario. Esto, crea un desincentivo a invertir menos de lo necesario para cubrir la demanda excesiva con procesamiento en la nube, generando un costo de oportunidad, pues estos mismos recursos se habrían podido utilizar en otra actividad productiva que sí generará retornos. Esta estructura en la nube es capaz de responder dinámicamente a la demanda computacional. Por lo tanto, una inversión de capital en una estructura en la nube resultaría en un pérdida de oportunidad¹⁷, cuándo las necesidades sobrepasen la capacidad obtenida.

Figura 21: Costo de instalación de almacenamiento de datos



Fuente: Observatorio TI (Observatorio de TI, 2017)

Invertir sobre un tipo de almacenamiento, puede ser la primera forma de entender el impacto indirecto que tiene el mercado de los datos sobre la economía. Cómo se expone anteriormente, este impacto se relaciona con los proveedores y clientes de la industria. Al hablar de proveedores, es importante entender que estos quieren maximizar sus ingresos iniciales ya que considera que una vez den acceso a la información, no volverán a generar ingresos con ese cliente por esos datos.

A su vez el *procesador/cliente* de datos no puede comparar precios y puede correr el riesgo de que los datos no cumplan con las características requeridas por lo que podría estar incurriendo en un gasto innecesario. Esto, brinda al mercado de los datos importantes espacios de negociación del precio entre los incentivos e intereses del proveedor-titular de datos, el procesador-intermediario y/(o) cliente final. Se esperaría entonces que veamos emerger innovaciones de mercado que generen más transparencia en el proceso de negociación y que permitan compartir muestras de datos que certifiquen la calidad de estos.

En el caso colombiano, el Estado mediante el organismo de Colombia Compra Eficiente y a través de acuerdos marcos de precio, sirve como mecanismo de agregación de la demanda de las entidades estatales, teniendo acceso a una fuente que le permite contabilizar el impacto indirecto que con sus actividades está produciendo.

Estos marcos tienen por objetivo establecer condiciones y especificaciones técnicas uniformes de los servicios que son ofrecidos al estado. Se han hecho dos Acuerdos Marco de Precios (AMP), uno para servicios de Nube Pública (este fue renovado

¹⁷ Costo de oportunidad: el coste de la inversión de los recursos disponibles a costa de la mejor inversión alternativa disponible, o también el valor de la mejor opción no realizada

en agosto 2018 y vigente hasta agosto 2019), y un segundo acuerdo marco de servicios de centros de datos y de nube privada (entre septiembre 2014 y septiembre 2016). A través de la modalidad de Nube Pública las entidades disponen de una alta flexibilidad y elasticidad para abastecerse en recursos computacionales configurables (e.g. redes, servidores y aplicaciones).

Hay 263 servicios que son accesibles bajo este AMP, principalmente asociados a capacidad de procesamiento (servidores de diferentes tamaños), capacidad de almacenamiento, plataformas o software base (plataformas de bases de datos, servidores de aplicaciones y servicios sin licenciamiento), equipos o dispositivos de seguridad, expertos, medios, certificados, etc. Estas opciones entonces son ajustadas dependiendo del sistema operativo y por el nivel de elasticidad requerido. Los proveedores autorizados por Colombia Compra Eficiente para brindar estos servicios son; Avanzo Colombia, Empresa de Telecomunicaciones de Bogotá ETC S.A., IFX Networks Colombia S.A.S, O4IT Colombia S.A.S, Unión Temporal Nube Nacional, Unión Temporal Nube Pública NEC-GTS y Unión Temporal UNE.

Por otro lado, el impacto inducido, como se expuso antes, considera los salarios de los trabajadores de la industria como un factor para medir el *Mercado de Datos*. Estos hacen parte del costo variable de operación y funcionamiento y están incluidos en el precio de los datos.

Al estudiar el caso colombiano, el Observatorio TI estimó que el número de empleados en el sector de Tecnología para el año 2016 estaba alrededor de los 110.000, lo que representa en promedio 18 empleados por compañía y menos del 1% (~0.5%) de la población total ocupada. Dicha población está distribuida principalmente entre las Grandes empresas. Las Pymes concentran respectivamente 35% y 41% de la población empleada en el sector, y el 24% se ubicaba en las Microempresas. Esta cifra es mínima en comparación a la cifra estimada por la Comisión Europea donde el mercado laboral de datos generaba 6.2 millones de empleos en la Unión Europea, casi 3.1% del empleo total. Este mismo reporte estimó además que el número de empleados por compañía de datos estaba alrededor de 24 empleados por compañía.

Este observatorio también hizo un censo de los salarios promedios en este sector y las variabilidades observadas entre los tamaños de las empresas. De acuerdo con las dificultades y falta de capacidades identificadas por el mercado colombiano, se suma la falta de un multiplicador con el que las empresas locales valoran/recompensan esos talentos escasos y valiosos que desean capturar.

Especialmente, se ha identificado que para los cargos de gerente de desarrollo, gerente de operaciones, y gerente de infraestructura, las grandes empresas están dispuestas a pagar más del doble del promedio nacional. Estas capacidades, particularmente valoradas, reflejan la desigualdad de rendimiento y conocimientos que ha generado la brecha de capacidades tecnológicas.

Tabla 11: “Promedio de Salarios Mensuales en el Sector TI Colombia” (COP, multiplicador por tamaño empresa)

Cargo	Promedio Nacional	Micro	Pequeña	Mediana	Grande
Analista de Servicios / Soporte	1,103,699	0.95 x	1.06 x	1.23 x	1.30 x
Ingeniero de Servicios	1,584,723	0.95 x	1.02 x	1.18 x	1.53 x
Analista de Pruebas	1,670,958	0.96 x	1.07 x	1.09 x	1.36 x
Ingeniero de Desarrollo	2,032,887	0.92 x	1.12 x	1.24 x	1.55 x
Auditor TI	2,066,667	0.99 x	1.00 x	1.12 x	1.00 x
Analista Especificador Funcional	2,128,849	0.92 x	1.18 x	1.25 x	0.75 x
Líder Consultor Funcional	2,158,600	1.00 x	1.02 x	0.57 x	0.79 x
Líder de Desarrollo	2,408,267	0.95 x	1.10 x	1.25 x	1.64 x
Líder de Calidad	2,429,907	0.97 x	0.70 x	1.23 x	1.00 x
Gerente Comercial	2,572,099	0.91 x	0.95 x	1.60 x	2.04 x
Coordinador de Proyecto	2,604,853	0.94 x	1.20 x	1.00 x	1.00 x
Gerente de Servicio al Cliente	2,713,973	0.95 x	0.92 x	1.30 x	2.03 x
Gerente de Zona/Regional	2,729,075	0.95 x	1.16 x	1.00 x	1.00 x
Administrador de Base de Datos	2,782,344	1.00 x	0.96 x	1.00 x	1.24 x
Director Comercial	2,911,533	0.89 x	1.10 x	1.42 x	1.91 x
Gerente de Producto	2,922,769	0.82 x	1.19 x	1.89 x	1.96 x
Gerente de Desarrollo	3,148,758	0.84 x	1.19 x	1.58 x	2.09 x
Director de Proyecto	3,161,748	0.95 x	1.02 x	1.19 x	1.52 x
Gerente de Operaciones	3,331,957	0.81 x	1.26 x	1.61 x	2.16 x
Arquitecto de Software	3,471,150	0.88 x	1.29 x	1.32 x	0.43 x
Gerente de Infraestructura	4,545,579	0.93 x	1.07 x	1.00 x	2.04 x

Fuente: Observatorio de TI, 2017

Cabe resaltar algo que fue tocado en el capítulo sobre capital humano, y es que se observa un desnivel entre los salarios que se ofrecen localmente frente al mercado laboral exterior. La diferencia salarial promedio es observable comparando los salarios promedio para roles similares como Líder de Calidad, Administrador de Base de Datos y Gerente de Desarrollo con los salarios promedios de la encuesta hecha por el Observatorio de TI. Incluso después de hacer una normalización de los salarios ajustándolos por el poder adquisitivo del salario (PPP 2012) y controlando la diferencia promedio observada por la organización mundial del trabajo (ILO); los resultados muestran que los equivalentes internacionales seguían ganando 2.0x

y 2.5x más que sus homólogos locales. En este contexto, es posible que en Colombia las necesidades en capital humano no se reduzcan pues los trabajadores encuentran bajos incentivos salariales para trabajar en el sector. Esto, tiene como efecto secundario la fuga del talento, afectando el desnivel de capacidad ya existente entre países.

No podemos negar los límites inherentes en el análisis del valor del dato a través de sus costos. El costo de almacenar, procesar y distribuir la información es insuficiente para capturar el valor generado por esta y difícilmente serviría como indicador para justificar la inversión de capital, ya que estas actividades no son consideradas por el agente como un fin en sí mismo. Sabemos que el precio real del dato está atado al valor que puede generar a través de su análisis y uso. Un estudio realizado por el Departamento de Comercio de Estados Unidos en el 2014 estableció que los costos en recolección, almacenamiento, procesamiento y diseminación de los datos representaron casi 3,7 mil millones de dólares anualmente, no obstante esto solo representa una captura del 0.02% del valor generado por la industria¹⁸.

La dificultad para determinar el valor del dato está atada a su naturaleza intangible y es la razón por la que la economía del dato resulta ser un caso de estudio económico interesante. Aunque los datos pueden ser considerados bienes intangibles, su acceso inicial puede ser controlado. No obstante, una vez este es otorgado, es difícil limitarlo. En este sentido el dato compartido se comporta de forma similar a un bien público (ósea sin un valor de mercado) donde se observa a la vez, un uso no rival¹⁹ y una asimetría de información²⁰.

Varias estrategias han sido concebidas para establecer un valor del dato por encima del costo. Un análisis para determinar este precio sería a través de una aproximación por el mercado, determinando cuánto están dispuestos a pagar los usuarios por los datos o productos similares. Bajo este análisis, las características y la información contenidas por el dato son factores que ayudan a determinar el valor de este. El periódico Financial Times ilustra esta variabilidad en el precio de los datos a través de una herramienta virtual (Financial Times, 2013). Sus conclusiones son las siguientes:

“La información general sobre una persona, como su edad, sexo y ubicación, vale solo \$0,0005 por persona, o \$0,50 por cada 1,000 personas. Una persona que está comprando un automóvil, un producto financiero o unas vacaciones es más valioso para las empresas que desean vender esos productos. Los compradores de automóviles, por ejemplo, valen alrededor de \$0.0021 por persona, o \$2.11 por cada 1,000 personas.

Ciertos hitos en la vida de una persona provocan cambios importantes en los patrones de compra, ya sea convirtiéndose en un nuevo padre, mudarse de casa, involucrarse, comprar un automóvil o divorciarse. Los vendedores están dispuestos a pagar más para llegar a los consumidores en los principales eventos de la vida. Sabiendo que una mujer está esperando un bebé y está en su segundo trimestre de embarazo, por ejemplo, envía el precio para etiquetar esa información acerca de ella a \$0.11.

Cuanto más íntima es la información, más valiosa es. Algunos de los datos más personales y secretos son los más caros. Por \$0.26 por persona, los compradores pueden acceder a listas de personas con condiciones de salud específicas o que toman ciertas prescripciones.”

Formas alternativas de capturar un valor de mercado han sido exploradas, por ejemplo, por organizaciones nacionales de estadística de España y de México mediante métodos de preferencias reveladas/declaradas. Estas, intentaron capturar el valor atribuido por los medios a los datos cuando estos tienen que decidir entre publicar contenido estadístico y publicidad, lo que indicaría una equivalencia para el público general entre la información contenida y el ingreso renunciado.

En ecosistemas poco desarrollados, los intermediarios de datos tienen dificultad en encontrar clientes, lo que evidencia la falta de casos que muestren el valor de los bienes o servicios de la economía del dato. Por lo anterior, el proveedor de datos prefiere favorecer precios elevados contra el acceso a sus datos ya que tampoco quiere correr el riesgo de perder el máximo valor extraíble de su bien. Para salir de tal *status quo* en el que ningún actor se desviará de su comportamiento para alcanzar un estado óptimo, se podría justificar la presencia de un tercero que garantice el buen comportamiento entre actores.

Bajo el marco descrito anteriormente para medir el impacto económico de la economía de datos, queda claro que sin tener mayor información del sector privado se dificulta captar la totalidad del mercado de datos. Este es un trabajo que requerirá de mayor articulación del sector privado y un mayor desarrollo de la industria. Sin embargo, la descripción de las variables y factores a tener en cuenta para medir el mercado de datos en Colombia, es importante al momento de reconocer los ejes principales que le otorgan valor a esta industria a nivel nacional.

¹⁸ “Fostering innovation, Creating Jobs, Driving Better Decisions: The value of Government Data” Economics and Statistics Administration Office of the Chief Economist. July 2014

¹⁹ Los procesadores de datos tienen pocos precios de referencia o comparativos mientras que los proveedores saben lo que otros procesadores han estado dispuestos a pagar

²⁰ El uso de los datos por una persona o varias, no impide o limita el uso por otros agentes

²¹ Se utilizó principalmente dos series de datos que están a disposición: la plataforma de Seguimiento a Proyectos de Inversión (SPI) del DNP y los datos abiertos de contratación y licitaciones públicas mantenidas por Colombia Compra Eficiente en SECOP II. Por un lado, SPI permite capturar de forma amplia todas aquellas iniciativas que tienen un carácter tecnológico, además permite hacer una primera separación entre lo que tiene carácter de costo de inversión (gastos en Hardware, Infraestructura y Software) contra lo que se ha asignado propiamente a Servicios que corresponden a los gastos de operación y mantenimiento. Adicionalmente, la plataforma SECOP II permite estudiar las características de los contratos, particularmente aquellos firmados y asignarlos entre estos dos niveles de costos.

3.6.2. DESCRIPCIÓN DEL GASTO PÚBLICO ACTUAL

Para entender el mercado de los datos en su totalidad, es importante reconocer el involucramiento del Estado y los costos que este asume en el sector de tecnología, haciendo referencia a las inversiones que sus diferentes entidades han asumido en proyectos pasados. Con base en las variables y factores identificados anteriormente, se procede a evaluar los gastos del Estado colombiano, en cuanto a gastos de inversión, operación y/o mantenimiento²¹.

De esta manera, se puede hacer un análisis de los costos estructurales soportados por el Gobierno, que almacena y procesa un universo de datos que a su vez generan un impacto económico. Lo anterior, con el propósito de aproximarnos superficialmente a una cifra de inversión por parte del sector público, y a su participación en el mercado de los datos.

La tabla “Inversión Total Hardware/Infraestructura en Tecnologías de Información y Comunicaciones” es uno de los productos derivados de la extracción de cifras de la plataforma SPI del DNP. En este sentido, para que una parte del presupuesto pudiera entrar en las consideraciones del DNP se utilizó la siguiente definición de Las Tecnologías de la Información y las Comunicaciones (TIC) “el conjunto de recursos, herramientas, equipos, programas informáticos, aplicaciones, redes y medios que permiten la compilación, procesamiento, almacenamiento, transmisión de información como: voz, datos, texto, vídeo e imágenes” (DNP, 2010). En este, se observan tres tipos de actividades: a) Hardware o elementos de Infraestructura; b) Software, Sistemas de Información o Aplicaciones y; c) Servicios, cada una teniendo una lista detallada de elementos a considerar (ver anexo).

Estas agregaciones fueron las que, bajo observación del DNP, fueron implementadas para evaluar las dimensiones transversales de la plataforma SPI. En promedio, entre 2014 y 2018 el Gobierno presupuestó anualmente: 321,5 mil millones COP en Software y Aplicaciones, 452.1 mil millones COP en Hardware/Infraestructura. Estas dos líneas de gastos transversales en TIC representan casi el 51% de todos los gastos TIC (21% Software y Aplicaciones, 29% Infraestructura y Hardware). El remanente en Servicios representa casi 825.9 mil millones COP anuales.

Tabla 12: “Inversión Total Hardware/Infraestructura en Tecnologías de Información y Comunicaciones” (Millones COP)

	DIC 2014	(%)	DIC 2015	(%)	DIC 2016	(%)	DIC 2017	(%)	MAY 2018	(%)	Total (Constante e Precios 2014)	(%)
Agricultura y Desarrollo Rural	1,681,441	22.1%	2,106,114	39.8%	1,028,175	19.8%	1,018,734	30.1%	352,555	23.0%	5,969,711	28.8%
Comercio, Industria y Turismo	1,795,653	23.6%	920,904	17.4%	744,927	14.4%	622,851	18.4%	223,718	14.6%	4,186,878	20.2%
Defensa Y Policía	870,110	11.4%	525,928	9.9%	797,251	15.4%	465,939	13.8%	270,461	17.6%	2,830,354	13.7%
Ambiente y Desarrollo Sostenible	1,252,584	16.5%	0	0.0%	1,158,118	22.3%	133,807	4.0%	9,196	0.6%	2,490,951	12.0%
Ciencia, Tecnología e Innovación	305,644	4.0%	462,120	8.7%	159,715	3.1%	144,129	4.3%	248,412	16.2%	1,271,096	6.1%
Congreso de la República	346,820	4.6%	233,475	4.4%	233,818	4.5%	234,346	6.9%	51,990	3.4%	1,064,101	5.1%
Educación	141,400	1.9%	102,089	1.9%	326,978	6.3%	182,451	5.4%	111,267	7.3%	829,324	4.0%
Hacienda	241,194	3.2%	147,671	2.8%	112,658	2.2%	158,525	4.7%	58,557	3.8%	695,579	3.4%
Inclusión Social y Reconciliación	237,504	3.1%	131,234	2.5%	66,570	1.3%	0	0.0%	5,330	0.3%	430,841	2.1%
Cultura:	85,525	1.1%	101,076	1.9%	64,800	1.3%	81,448	2.4%	31,762	2.1%	351,150	1.7%
Fiscalía	94,890	1.2%	63,924	1.2%	80,414	1.6%	47,320	1.4%	25,100	1.6%	301,194	1.5%
Justicia y del Derecho	120,720	1.6%	73,138	1.4%	56,259	1.1%	20,200	0.6%	25,158	1.6%	287,046	1.4%
Minas y Energía	50,850	0.7%	40,323	0.8%	64,000	1.2%	44,255	1.3%	25,302	1.6%	216,344	1.0%
Información Estadística	33,862	0.4%	49,982	0.9%	70,880	1.4%	19,685	0.6%	12,250	0.8%	179,289	0.9%
Empleo Público	60,480	0.8%	110,351	2.1%	0	0.0%	0	0.0%	0	0.0%	165,509	0.8%
Organismos de Control	56,987	0.7%	27,267	0.5%	31,606	0.6%	33,695	1.0%	16,380	1.1%	160,680	0.8%
Inteligencia	29,124	0.4%	36,088	0.7%	19,686	0.4%	53,540	1.6%	13,155	0.9%	145,686	0.7%
Salud y Protección Social	10,618	0.1%	27,784	0.5%	34,435	0.7%	24,760	0.7%	17,600	1.1%	110,153	0.5%
Transporte	38,544	0.5%	35,408	0.7%	19,140	0.4%	0	0.0%	2,030	0.1%	92,393	0.4%
Tecnologías de la Información y las Comunicaciones	17,088	0.2%	17,058	0.3%	8,488	0.2%	32,700	1.0%	5,648	0.4%	77,900	0.4%
Trabajo	36,721	0.5%	15,900	0.3%	17,632	0.3%	3,106	0.1%	2,225	0.1%	73,710	0.4%
Interior	16,680	0.2%	21,209	0.4%	19,831	0.4%	10,363	0.3%	5,079	0.3%	70,438	0.3%
Registaduría	10,074	0.1%	3,203	0.1%	20,454	0.4%	19,582	0.6%	12,928	0.8%	63,532	0.3%
Relaciones Exteriores	37,148	0.5%	8,762	0.2%	11,235	0.2%	2,898	0.1%	0	0.0%	58,939	0.3%
Presidencia de la República	12,657	0.2%	10,444	0.2%	8,289	0.2%	15,844	0.5%	4,925	0.3%	50,254	0.2%
Planeación	19,776	0.3%	1,742	0.0%	10,802	0.2%	3,572	0.1%	3,555	0.2%	38,498	0.2%
Rama Judicial	7,156	0.1%	15,035	0.3%	6,567	0.1%	6,767	0.2%	0	0.0%	34,157	0.2%
Vivienda, Ciudad y Territorio	2,664	0.0%	3,334	0.1%	7,391	0.1%	1,499	0.0%	0	0.0%	14,298	0.1%
TOTALES (Corrientes)	7,613,914		5,291,563		5,180,119		3,382,016		1,534,583			
TOTAL (Constante 2014)	7,613,914		4,956,039		4,898,458		3,249,127		1,460,569		20,717,538	

Fuente: SPI (www.spi.dnp.gov.co), ver en detalle aquí, *las dos inversiones más altas resaltadas en verde

En promedio, el presupuesto asignado a las TIC ha ido disminuyendo anualmente a una tasa de 4.6%. De hecho, se asignaron 1.8 billones de pesos en proyectos TIC en el año 2014 mientras que en el 2018 solo fueron asignados 1.4 billones de pesos. Bajo esta situación presupuestal, las dimensiones más afectadas fueron Infraestructura y HardWare (HW), cuya inversión ha disminuido desde 2014, a una tasa anual de 14.4% (en el 2014 se invirtieron 560.5 mil millones de pesos en proyectos y en el 2018 se presupuestaron 317 mil millones para esta misma línea).

²² Solo se consideraron contratos “Firmados”, “Terminados” o “Cerrados”, se trabajó solo sobre contratos pertenecientes a los contratos de categoría nivel 1 “Servicios Basados en Ingeniería, Investigación y Tecnología” y “Difusión de Tecnologías de Información y Telecomunicaciones” y excluyendo la categoría de nivel 2 “Servicios profesionales de ingeniería”

²³ Para la adquisición de bienes tangibles e intangibles se seleccionó solo aquellos contratos que están clasificados como “Bienes” o “Servicios”

²⁴ Las categorías “Servicios de Aprovisionamiento”, “Compraventa”, “Consultoría” fueron asignadas a los gastos CAPEX, contratos en las

²⁵ Definido como un espacio presentando más de 5,000 servidores conectados y una superficie mayor a los 1,000 metros cuadrado, se habla de hiperscalas ya que esta industria ha observado problemas de escala al empezar a considerar esta magnitud de infraestructura

Para este análisis, también se extrajo información de la plataforma SECOP II de Colombia Compra Eficiente. Después de un tratamiento básico de esta base se pudo construir la tabla 13²², en la cual se observa que la mayoría de los contratos, hacia las tecnologías de la información y las comunicaciones, se concentran en 18 de 65 líneas de gasto. Estas líneas de gasto concentran 386.5 mil millones de pesos sobre 476 mil millones de pesos de los contratos firmados desde el 2015 hasta junio 2018. Asimismo, se identificó que dentro de las categorías de gasto era común observar una concentración del mercado hacia un solo actor capaz de representar más del 25% del valor total de contratos atribuidos.

Tabla 13: “Total Valor y Número de Contratos por Categoría de Gasto”/ “Principales Proveedores por Categoría de Gasto, Valor de Contrato, Concentración del Mercado”

Categoría Gasto	Total Valor de Contratos 2010-2018 (COP)	Cantidad de Contratos	Principal Proveedor	Tamaño de contratos	Concentración del mercado (%)
Servicios de sistemas y administración de componentes de sistemas	81,744,778,289	523	IMPRESA NACIONAL DE COLOMBIA	19,976,083,272.00	24%
Ingeniería de software o hardware	57,291,621,706	559	MANAGEMENT AND QUALITY	8,911,023,778.00	16%
Mantenimiento y soporte de software	44,537,325,691	335	VORTAL-COMERCIO ELECTRONICO	4,420,600,000.00	10%
Computadores	26,046,078,440	66	SUMIMAS S.A.S.	10,686,100,695.00	44%
Servicios de telecomunicaciones	25,835,124,668	36	NEWSAT S.A.S	5,866,055,040.00	23%
Servicios de alquiler o arrendamiento de equipos plataforma de voz y datos o multimedia	14,875,927,005	11	COLOMBIA TELECOMUNICACIONES S.A	13,514,949,286.00	91%
Software de seguridad y protección	14,657,431,566	155	SONDA DE COLOMBIA	2,799,000,000.00	19%
Cartografía	14,635,187,505	327	UNION TEMPORAL	3,215,760,000.00	22%
Software funcional específico de la empresa	11,959,544,897	104	LATINAMERICAN POLYGRAPH INSTITUTE Y/O LATINOAMERICAN	1,439,745,598.00	12%
Servicios de datos	11,694,325,650	150	EMTELCO S.A.S	1,829,067,759.00	16%
Mantenimiento y soporte de hardware	10,743,026,337	129	SOLUCIONES	3,703,633,779.00	32%
Sistemas de manejo de información MIS	9,600,337,864	66	UT SDG OPUK	1,685,132,360.00	18%
Servicios de alquiler o arrendamiento de licencias de software de computador	9,577,975,401	77	COUPA SOFTWARE INC	1,782,500,000.00	19%
Análisis económico	8,681,027,919	118	JUAN MANUEL ROLDAN	214,200,000.00	2%
Software de consultas y gestión de datos	8,675,565,301	51	CAUDALES Y MUESTREOS	1,451,086,000.00	17%
Dispositivos de almacenamiento	8,173,353,825	38	PROCIBERNÉTICA S.A.	1,400,000,000.00	17%
Metodología y Análisis	7,170,634,377	54	EMPRESA DE TELECOMUNICACIONES	2,899,016,932.00	37%
Otros	93,743,166,908	1045			
Total	459,645,433,039	3844			

(COP, contratos firmados entre 2015 y 2018)

Fuente: SECOP II

Para retomar el marco económico de costos que se expuso al principio de este segmento, se clasificaron los contratos publicados bajo Colombia Compra Eficiente para determinar si eran inversiones en activos fijos o gastos operacionales. Asimismo, se extrapoló a partir de los mismos criterios base usados en el análisis anterior, pero además se incorporaron algunos filtros de segmento de contrato²² y tipo de contrato²³.

Es importante resaltar que SECOP II captura solamente una pequeña parte del universo del presupuesto público, ya que la novedad de la herramienta y el nivel de adopción por parte de las entidades han distorsionado la información (2017 y 2018 reportan el 90% del valor de los contratos firmados).

Usando SECOP I observamos que los contratos en las categorías de gasto relevantes para los años 2015 a 2018 concentran el 98% del valor de los contratos liquidados y celebrados.

Tabla 14: “Total Valor y Número de Contratos por Categoría de Gasto”/ “Principales Proveedores por Categoría de Gasto, Valor de Contrato, Concentración del Mercado” (COP, contratos firmados entre 2010 y 2018)

Categoría Gasto	Total Valor de Contratos 2010-2018 (COP)	%	Cantidad de Contratos	Principal Proveedor	Tamaño de contratos	Concentración (%)
Ingeniería de software o hardware	1,298,806,049,725	14%	7,616	UNIÓN TEMPORAL DISTRIBUCIÓN PROCESOS ELECTORALES CONSULTA ANTICORRUPCIÓN 2018 UT DISPROEL CONSULTA	92,701,904,243.00	7.1%
Computadores	1,556,158,851,236	16%	9,563	SISTEMA TECNOLOGÍA E IMÁGENES LTDA	292,388,701,290.00	18.8%
Servicios de sistemas y administración de componentes de sistemas	1,153,126,238,525	12%	6,430	EMPRESA DE RECURSOS TECNOLOGICOS SA ESP	67,657,255,200.00	5.9%
Servicios de internet	911,246,544,882	10%	4,156	FONDO FINANCIERO DE PROYECTOS DE DESARROLLO FONADE	247,304,523,071.00	27.1%
Servicios de datos	839,662,973,351	9%	2,873	UNIÓN TEMPORAL SOLUCIONES INFORMÁTICAS ELECTORALES 2018. UT SIE 2018.	328,786,927,210.00	39.2%
Mantenimiento y soporte de software	700,926,512,838	7%	7,718	PRAGMATICA SOFTWARE LTDA.	83,716,833,200.00	11.9%
Servicios de alquiler o arrendamiento de hardware de computador	576,399,738,477	6%	986	LEASING DE OCCIDENTE	32,964,399,863.00	5.7%
Dispositivos informáticos de entrada de datos	263,769,787,028	3%	1,437	SUMIMAS S.A.S.	49,048,073,864.00	18.6%
Servicios de administración de acceso	257,738,935,697	3%	1,541	SOFTWARE AG S.A. DE C.V.	77,975,029,731.00	30.3%
Metodología y Análisis	293,524,753,544	3%	2,001	QRI INTERNATIONAL LLC	92,031,900,000.00	31.4%
	1,720,746,489,536	18%	19,943			

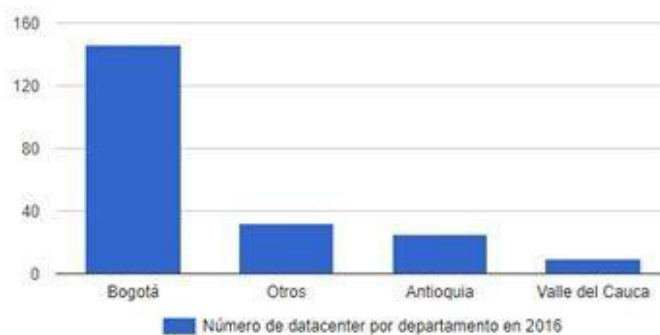
Fuente: SECOP I

²⁶ Colombia Compra eficiente, Coldeportes, DIAN, DIMAR, Fondo de Adaptación, Función Pública, IDEAM, Ministerio de Agricultura, Ministerio de Hacienda, Ministerio de Trabajo, Ministerio de Salud y Policía Nacional.

Aunque los datos de SECOP II no capturan la totalidad de la inversión pública en infraestructura de datos, esta plataforma representa un volumen y valores de contratos significativos para el erario por lo que su repartición entre costos de inversión y costos de mantenimiento y operación resultan relevantes (ver tabla en anexo). Como resultado de este análisis se identificó que una gran parte de los esfuerzos se han concentrado hacia gastos operacionales e intangibles. Estas dos categorías concentran casi el 86% del valor de los contratos mientras que la inversión en activos tangibles representa solo el 13%. Asimismo, las Categorías “Servicios de sistemas y administración de componentes de sistemas” e “Ingeniería de software o hardware” capturaron el 37% del valor de los contratos.

En términos de infraestructura de datos física, Colombia presenta un número relativamente bajo de *datacenters* locales, superando los 200 espacios en el 2016 (Ver Ilustración 37). Colombia todavía no ha llegado a una escala para justificar su permanencia en la lista de aquellos países que cuentan con *datacenters* de hiperescala²⁵. Es posible que Colombia se demore en llegar hasta ese nivel ya que los costos de importación de material tecnológico de alto nivel son restrictivos y no permiten procesar localmente los datos de forma costo eficiente, competitivo con los hipercentros al exterior.

Figura 22: Número de Datacenters establecidos en Colombia (Año 2016, # Total 213)



Fuente: Observatorio TI (Observatorio de TI, 2017)

3.6.3. ¿HACIA DÓNDE APUNTA?: GASTO Y PRESUPUESTO POLÍTICA NACIONAL DE EXPLOTACIÓN DE DATOS

El Documento CONPES también hace énfasis en que, en un mediano plazo, se hará un plan coordinado de gasto público que permitirá considerar el grado de involucramiento de otras organizaciones estatales. Al usar el anexo del Documento CONPES “Plan de Acción y Seguimiento”, como insumo para este diagnóstico, se puede analizar la distribución de los recursos propuestos e inferir aquellas líneas donde se requiere la mayor cantidad de esfuerzos (ver Tabla 15). Cabe resaltar preliminarmente, que casi 13.4 mil millones COP se concentran en sólo cinco de las trece líneas de acción propuestas por el CONPES.

- Línea de acción 1: “Diseñar e implementar la infraestructura de datos” - 9% del costo total
 - Una iniciativa principal dentro de esta acción busca generar un plan de gestión documental electrónica (Hoja de Ruta) para la automatización progresiva de procesos y procedimientos administrativos en la producción y gestión de documentos. Esta hoja implementará estándares que garanticen que los datos contenidos permanezcan descubribles, usables y fiables.
- Línea de acción 7: “Nivelar el alistamiento para la explotación de datos en las entidades públicas.” - 14% del costo total.
 - En esta línea de acción, el gobierno invertirá 2 mil millones en la creación de un espacio colaborativo para que los datos utilizados, los procesos y sus resultados se publiquen en un ambiente de trabajo para el aprovechamiento público. Esto ayudará en el desarrollo del ecosistema local, ya que contribuirá a crear estándares y a desarrollar vínculos de colaboración entre actores públicos, privados, la academia y la sociedad civil.
- Línea de acción 10: “Generar los mecanismos para materializar el valor de los datos en las entidades de la administración pública” - 21% del costo total.
 - Con este objetivo, el Gobierno invertirá alrededor de 3.3 mil millones de pesos enfocándose en las siguientes acciones: (i) creación de un índice de valor potencial de la explotación de datos; (ii) estimación del impacto económico y social de la implementación de la explotación de datos en el país; (iii) determinación de los requisitos para que las entidades públicas dispongan de las herramientas e infraestructura para capturar, almacenar, procesar y compartir datos; (iv) implementación de la explotación de datos en las entidades públicas, como herramienta para cuantificar y optimizar los recursos públicos; (v) disposición de un grupo de analítica de datos que incentive la generación de capital humano.
- Línea de acción 12: “Definir, implementar y dinamizar el funcionamiento del mercado de datos en Colombia” - 30% del costo total 2018-2021

- Dentro de este objetivo se invertirán 4.5 mil millones de pesos en la articulación de una Red Nacional de Servicios de Computación de Alto Rendimiento. Esta línea de acción debería tomar ideas de la implementación por la Comunidad Europea para medir el impacto macroeconómico que se está generando por el mercado y la economía de datos.
- Línea de acción 13: “Vincular la ciudadanía a la infraestructura de datos” - 7% del costo total
- Aunque se podría considerar este objetivo como amplio, una de las metas, atribuidas como responsabilidad al MinTIC, consiste en incentivar el uso del ecosistema de datos. Ello, requiere disponer de herramientas libres para construir análisis y resolver problemas por parte de los actores del ecosistema de datos; e incluir dentro de las acciones de apropiación línea específicas de intervención para fomentar el uso de la explotación de datos, y la apropiación de los productos que se generen.

Tabla 15: “Financiamiento de la política por año” (Millones COP)

Entidad	2018	2019	2020	2021	Total
Agencia nacional de contratación pública – Colombia Compra Eficiente	32	-	-	-	32
Agencia Nacional de Defensa Jurídica del Estado	110	130	-	-	240
Archivo General de la Nación	105	400	400	400	1.305
Departamento Administrativo de Función Pública					
Departamento Nacional de Planeación	891	1.485	476	-	2.852
Ministerio de Comercio, Industria y Turismo (iNNpulsa Colombia)	1.776	-	-	-	1.776
Ministerio de Educación Nacional	30	15	-	-	45
Ministerio de Tecnologías de la Información y las Comunicaciones	3.314	1.706	2.900	2.500	10.420
Departamento Administrativo de la Presidencia de la República [Secretaría de Transparencia]	16	9	4	-	29
Superintendencia de Industria y Comercio	10	19	-	-	29
Total	6.284	3.764	3.780	2.900	16.728

Fuente: DNP, 2018I

Tabla 16: “Financiamiento de la política por año” (Millones COP)

Líneas de Acción	Título de Línea de Acción	Importancia de la Acción	2018	2019	2020	2021	Total Costo Acciones (millones de pesos)	(%)
Línea de acción 1	Diseñar e implementar la infraestructura de datos	12%	358	300	400	400	1.458	9%
Línea de acción 2	Materializar la apertura por defecto	10%	505	160	0	0	665	4%
Línea de acción 3	Reglamentar y fortalecer técnicamente la habilitación general para el intercambio de información entre entidades públicas	8%	285	127	0	0	362	2%
Línea de acción 4	Reconocimiento jurídico de los datos como activo	12%	119	109	49	0	277	2%
Línea de acción 5	Solventar el déficit de protección legal y ético	6%	84	106	0	0	190	1%
Línea de acción 6	Definir condiciones para el intercambio de datos entre los sectores público y privado	2%	43	52	2	0	97	1%
Línea de acción 7	Nivelar el alistamiento para la explotación de datos en las entidades públicas	14%	100	200	1.000	1.100	2.400	14%
Línea de acción 8	Medición de la brecha de capital humano y actualización de competencias	12%	290	115	150	150	645	4%
Línea de acción 9	Promover el emprendimiento de bienes y servicios basados en la explotación de datos	4%	788	0	0	0	788	5%
Línea de acción 10	Generar los mecanismos para materializar el valor de los datos en las entidades de la administración pública	9%	1.349	1.231	929	0	3.509	21%
Línea de acción 11	Institucionalizar la explotación de datos en la toma de decisiones públicas	5%	123	142	0	0	265	2%
Línea de acción 12	Definir, implementar y dinamizar el funcionamiento del mercado de datos en Colombia	5%	1.952	1.000	1.000	1.000	4.952	30%
Línea de acción 13	Vincular la ciudadanía a la infraestructura de datos	2%	370	250	250	250	1.120	7%
TOTAL			6.256	3.792	3.780	2.900	16728	

Fuente: DNP, 2018I

3.6.4. DIAGNÓSTICO: GASTO E INVERSIÓN ASOCIADA A LA EXPLOTACIÓN DE DATOS

En el momento, no hay manera concreta de medir la economía de datos de Colombia, al ser este un sector emergente y en continuo desarrollo. Para guiar las políticas públicas orientadas a la explotación de datos, es necesario desarrollar herramientas y marcos para el análisis del mercado de datos, como lo hizo la Unión Europea en el 2017. Este indicador no solo mide el tamaño del mercado de datos, pero identifica patrones y actores relevantes para la economía de datos. El identificar estas variables, hace posible priorizar la inversión sectorial para investigación y desarrollo, así como monitorear el desarrollo de la economía de datos por medio de un seguimiento a los clientes / proveedores (impacto indirecto) y al mercado laboral fruto de la industria del dato (impacto inducido).

Por otro lado, el monto recaudado para la implementación de la Política Nacional de Explotación de Datos, en comparación con otras estrategias internacionales, es muy bajo, pues el presupuesto total de los próximos 4 años en esta política - de 16.728 millones de pesos - representa apenas el 11.1% del presupuesto asignado a proyectos TIC en 2017 (1.5 billones de pesos o 0.7% del Presupuesto General de la Nación). Esta cifra, de aproximadamente 5 millones de dólares, es significativamente menor a la inversión que otros países alrededor del mundo han hecho para iniciativas de Big Data e Inteligencia Artificial.

Adicional a esto, para el 2018 tanto el Fondo de Tecnologías de la Información y Comunicaciones (FONTIC) asignó 745 miles de millones de pesos y Colciencias presupuestó 75.5 miles de millones de pesos para mejorar la calidad y el impacto de la investigación y la transferencia de conocimiento y tecnología. También asignaron 3 mil millones de pesos a promover el desarrollo tecnológico y la innovación. En comparación con la cifra reportada por la administración pública de los 28 países de la Comunidad Europea, para el periodo 2013-2016, el 8.1% de los recursos invertidos a iniciativas TIC fueron asignados como gasto a la economía de datos. En el 2016, México reportó un gasto en compra de equipos de ciencia y tecnología por un valor de 426.8 mil millones de pesos COP; y este solo representa un 3.4% del gasto total del Gobierno Federal en Ciencia y Tecnología. El presupuesto en TIC, equivale a 2.01% del Presupuesto Federal Programable y 0.43% del PIB.

Aún cuando se asume que la inversión del país en materia de datos podría ser superior, el rol del Estado dentro de la economía de datos puede estar distorsionando esta percepción, pues la participación del sector público debe ser más de acompañamiento y de fortalecimiento de las economías de datos locales, en vez de un posicionamiento como ente central. Por ejemplo, la Comisión Europea reporta que la participación del sector público en la economía de datos representó solo el 5.5% del valor total de este en el 2017. El Estado debería tomar un rol principal en la economía de datos cuando los costos de entrada son excesivos, y quizá requieran de coordinación entre varios actores, o cuando los riesgos asociados a la explotación de datos dificultan el posicionamiento por parte del sector privado.

En Colombia es esencial proporcionar espacios de inversión para el sector privado en alianza con el sector público. Esto podría llevarse a cabo a través de alianzas público privadas, o la creación de fondos mixtos para inversión en tecnología con el sector público dirigiendo esta (como es el caso de México), entre otros.

De igual manera, se debe reconocer que la actual estructura fiscal en Colombia no está alineada en incentivar la instalación de capacidad de computación, a través de equipos locales. Las recientes modificaciones tributarias (a través del artículo 187 de la Ley 1819, que agrega el numeral 24 al artículo 476 del ET) incentivan el uso de servicios en línea. Los servicios que quedaron excluidos del IVA son: a) suministro de páginas web;

b) servidores (hosting); c) computación en la nube (cloud computing); d) mantenimiento a distancia de programas y equipos (DIAN, 2017). Sin embargo, lo único relevante a la instalación de servidores locales es la exclusión del IVA a computadores que tengan un valor inferior a 1.200.000 COP. Todos los demás equipos deberán pagar un valor del 19% por encima de su valor declarado. Por otro lado, las piezas de tecnología en el momento no están sujetos a un gravamen arancelario para la importación (Estatuto Tributario Nacional, 2017).

4. CONCLUSIONES

Al integrar el diagnóstico de las seis dimensiones, podemos llegar a conclusiones generales sobre la situación actual de Colombia en Big Data y los lineamientos que deben ser tomados en cuenta para el desarrollo de un plan de acción.

En primer lugar, es evidente la falta de conocimiento al interior del Estado sobre sus propias necesidades a nivel digital. Entidades como el MinTIC, Fedesoft y el DANE, han hecho innegables esfuerzos por cuantificar y conocer el sector privado en temas digitales y de datos. Sin embargo, el ejercicio que ha realizado el propio Estado por conocer sus deficiencias, retos y entender la brecha digital todavía requiere de más profundidad y un mayor análisis.

Además, aunque hay una gran cantidad de informes disponibles, con una gran cantidad de información abierta al público, estos son difusos e incluso pueden llegar a ser contradictorios. El Estado debe implementar estrategias que lleven a la unificación y uniformidad de la información entre sectores, regiones y tipo de entidades. En líneas generales, la información es difusa y poco estructurada entre entidades, lo que hace evidente una falta de coherencia, que sin duda afecta las estrategias que buscan llevar a la interoperabilidad; condición esencial para la explotación de datos.

En cuanto a fortalezas, es innegable el avance que Colombia ha tenido en la producción de datos estadísticos, y en general, en su transformación digital. Los datos han adquirido gran relevancia en el sector público y cada día hay una mayor necesidad de este tipo de conocimiento para el desarrollo de políticas públicas informadas. Esto, llevó al desarrollo de estrategias de gobierno electrónico que aunque han tenido retrasos en la implementación y no han sido correctamente adoptadas por todas las entidades, muestran cómo las Tecnologías de la Información tienen un papel prioritario en la visión de país que tiene el estado.

Ahora bien, en el 2011, el MIT Sloan Management Review y el Instituto IBM para el Valor de los Negocios, realizaron una encuesta con el fin de entender los retos y oportunidades asociadas al uso de la analítica de datos en los negocios. Para este ejercicio, se desarrollaron tres niveles de capacidad para la adopción de procesos analíticos (revisar Anexo E). Ajustando este ejercicio al caso colombiano, es claro como en todas las dimensiones analizadas, el país se encuentra en un nivel *aspiracional*, donde el análisis de los datos justifica las acciones gubernamentales, pero todavía no las guía.

Aunque ya hay una integración de los datos en el día a día de las organizaciones y ya son una prioridad para lograr el desarrollo estatal, Colombia al estar en un nivel aspiracional, enfrenta barreras relacionadas no tanto a la normativa y a la tecnología, sino a aspectos gerenciales, culturales y a una falta de comprensión sobre su utilidad.

Es claro cómo el desarrollo de una política de gobierno en línea va abonando el terreno para el aprovechamiento de datos y Big Data en el país. En particular, se observa que cada fase de crecimiento de gobierno electrónico implica un grado de madurez asociado a un avance en los procesos de digitalización, inteligencia de negocios, datos abiertos y big data. Sin embargo, Colombia debe pasar de la teoría a la implementación para crear un ecosistema de datos. Los vacíos a nivel de procesos, estándares, recursos humanos y el bajo nivel de inversión, entorpecen el esfuerzo hecho en los últimos años y limitan el potencial que el buen uso de los datos puede tener a nivel público y privado.

5. ANEXOS

ANEXO A: ESTUDIO CUALITATIVO

Con el objetivo de realizar un análisis de la situación actual en Colombia en relación con la apropiación de Big Data, se realizaron entrevistas a doce entidades públicas de orden nacional²⁶, las cuales fueron seleccionadas en una fase inicial de corte cuantitativo. Cabe resaltar que las entrevistas fueron hechas a sub-directores, viceministros y/o personas encargadas del manejo de datos en las entidades. Algunas entrevistas fueron realizadas en duplas/triadas e incluso sesiones de grupo, lo que permitió acceder a una información más completa sobre la entidad.

Esta sección presenta los resultados de las entrevistas realizadas a las entidades a partir de un análisis cualitativo a profundidad de doce aspectos seleccionados y presentados de la siguiente forma: 1) Explotación de datos; 2) Digitalización de la información y política de gestión de datos; 3) Códigos de Ética; 4) Bases de Datos Internas y Bases de Datos Externas; 5) Servicios en la Nube; 6) Apertura de datos; 7) Interoperabilidad; 8) Analítica de datos; 9) Tableros de control; 10) Expectativas y Sueños frente a Big Data. Por último, se presentan dos cuadros los cuales resumen las fortalezas, motivaciones y barreras, debilidades y riesgos de las entidades frente a los aspectos anteriormente mencionados.

Explotación de Datos

Para comenzar, es importante tener en cuenta que durante el trabajo de campo se encontraron dos aspectos de partida: por un lado, existen entidades del orden nacional que actualmente llevan un proceso avanzado, maduro y con una proyección clara en lo que a la explotación de datos se refiere. Por otro lado, hay entidades que todavía cuentan con una cultura de datos pobre en donde, si bien puede darse una concientización acerca de su relevancia, hay limitaciones que afectan los diferentes procesos.

En primer lugar, es fundamental abordar dos conceptos que actualmente han tomado fuerza en las diferentes entidades y, a pesar de la diferencia entre estos, cobran una importancia similar para estas instituciones: Big Data y explotación de datos. Aunque cada entidad le da su propio significado a estos conceptos, para todas es claro que hoy se convierten en un elemento primordial que permite, o permitiría, a las diferentes organizaciones tomar decisiones de una manera más estratégica y acertada.

Si bien algunos resaltan que Big Data es un concepto que tiene que ver más con un comportamiento y/o caracterización de los datos en cuanto a volumen, velocidad y variedad, se percibe también como un concepto articulado con la explotación de datos. En este sentido, algunas de las definiciones de Big Data apuntan a: un activo de información estratégica para la entidad; un mecanismo y estrategia para analizar los grandes volúmenes de información para tomar decisiones; explotación de datos para hacer analítica descriptiva o, en un uso más avanzado, hacer predicciones.

En las entrevistas se identificaron fortalezas de las entidades, especialmente relacionadas con la agilidad en los procesos, la posibilidad de dar respuestas oportunas a los ciudadanos y generar conocimiento. De manera similar, pero aislada,

las entidades mencionaron que la explotación de datos permite una mayor visibilidad nacional de la entidad y, además, la adquisición de un conocimiento más detallado de la entidad. No obstante, los entrevistados hacen alusión a algunas barreras y riesgos que, de una u otra manera, pueden afectar la explotación de los datos: por ejemplo, se menciona el desconocimiento de las herramientas tecnológicas por parte de la ciudadanía y por parte de los funcionarios, los recursos tecnológicos y presupuesto limitado, una cultura de datos pobre y la falta de recurso humano capacitado. Por último, los riesgos que identifican están relacionados con la calidad de la información que puede generar fallas al momento de las decisiones.

Tabla 17: Panorama de las entidades en la explotación de datos

Entidad	Fortalezas	Debilidades
Coldeportes	El paso de instituto a departamento representó un aumento en el presupuesto y mayor posicionamiento; Sistema de Información Nacional del Deporte : permite consolidar la información de los organismos del sistema nacional del deporte.	Cultura de datos pobre.
Colombia Compra Eficiente	Plataforma para una mejor información de la compra pública (2015); Bodega de datos interna en donde se llevan los datos importantes de las 3 plataformas que manejan en la entidad.	
DIMAR	Arquitectura Empresarial : diagnóstico de la situación actual para poder seguir con el proceso.	No hay un gobierno de datos y la gestión se realiza empíricamente.
Fondo de Adaptación	Concientización en el uso de datos; Mejoras al sistema de información de proyectos, a los tableros de control y la integración de información de proyectos, financiera y contractual.	No cuenta con un proceso maduro de explotación de datos.
Policía Nacional	Desarrollo de un proyecto para proveer de dispositivos móviles a la policía en la calle para tomar decisiones de manera más certera y ágil; Digitalización de la información y de todos los documentos y procesos asociados con la comandancia.	Cultura de datos es pobre.
Ideam	Construcción de un cubo de datos para almacenamiento y procesamiento de imágenes de satélite.	
Ministerio de Agricultura	Creación de 4 aplicaciones : a) AgroInsumos; b) AgroClima; c) AgroNegocios; d) AgroTeConecta.	
DIAN	Desarrollo de nuevos sistemas de inteligencia; Creación de una forma de pensar diferente a lo transaccional.	
Función Pública	Creación de una Dirección de gestión del conocimiento; Creación de un modelo matricial (grupo de asesoría y gestión; grupo de análisis y política para gestión del conocimiento).	
Ministerio de Hacienda	Creación del portal de transparencia; Creación del Marco Fiscal de Mediano Plazo; Desarrollo de nuevas herramientas e instrumentos para el 'forecasting'.	
Ministerio de Trabajo	Fuente de Información Laboral de Colombia a través del cual se busca hacer explotación de datos. Se hace un procesamiento y comienzan a generarse información para el público.	

Fuente: Elaboración propia

En general, en las entidades se resalta que durante los últimos tres años se han dado avances, representados en innovaciones que han generado, o se espera que generen, un impacto positivo dentro de la entidad.

Digitalización de la información y política de gestión de datos

La digitalización de la información es uno de los procesos más importantes y está relacionado con la captura y levantamiento de información. En la actualidad, esta etapa del proceso es altamente relevante para la mayoría de las entidades entrevistadas, quienes aseguran tener un gran avance en este sentido. Dicho nivel de relevancia se ancla en ventajas que apuntan a agilizar procesos; permitir conservar la documentación a largo plazo; posibilidad de interpretar y analizar la información; y exponer la información de una manera más sencilla.

En las entrevistas se evidencia que la mayoría de las entidades aseguran contar con casi el 100% de su información digitalizada. Entre estas entidades está Colombia Compra Eficiente, Policía Nacional, IDEAM, Ministerio de Agricultura y Ministerio de Salud. Por otra parte, cuando se explora por la existencia de políticas de gestión de datos dentro de las diferentes instituciones, se encuentra que hay varias entidades que aseguran no contar con ningún tipo de política como es

Tabla 18: Panorama de las entidades en la digitalización de la información y política de gestión de datos

Entidad	Fortalezas	Debilidades
Coldeportes	El 70% de la información está digitalizada	El 30% son documentos que se generan en juegos, planillas de resultados de eventos que no se han podido digitalizar; No cuentan con ningún tipo de política de gestión.
Colombia Compra Eficiente	El 100% de la información está digitalizada.	No cuentan con ningún tipo de política.
DIMAR		No logra calcular un porcentaje de información digitalizada; convenios a nivel internacional que demandan soportes físicos; No cuentan con ningún tipo de política.
Fondo de Adaptación	El 40% de la información está digitalizada.	Falta de presupuesto, cultura y recurso humano; Desconocimiento ; Limitaciones tecnológicas y de preparación del recurso humano; temor al cambio; El área financiera exige el papel en ciertos documentos; No cuentan con ningún tipo de política.
Policía Nacional	El 100% de la información está digitalizada.	La Fiscalía demanda documentos físicos firmados para poder iniciar investigaciones.
IDEAM	El 100% de la información está digitalizada; principios de transparencia.	
Ministerio de Agricultura	El 100% de la información está digitalizada.	No cuentan con ningún tipo de política.
DIAN	El 98% de la información está digitalizada; Proyecto de inversión sobre digitalización y gestión documental digitalizada; inventario de activos de información y política de seguridad.	La Contraloría y La Fiscalía exigen documentos físicos como evidencia.
Función Pública		
Ministerio de Hacienda	El 100% de la información está digitalizada.	Por normatividad, el banco fiscal debe radicar en el congreso la documentación tanto impreso como digital.
Ministerio de Trabajo	El 60% de la información está digitalizada.	El 40% no está digitalizada. Cuentan con una carga histórica que no han podido procesar.
Ministerio de Salud	El 100% de la información está digitalizada; desarrollan prácticas en gestión.	

Fuente: Elaboración propia

el caso de Colombia Compra Eficiente, Coldeportes, DIMAR, Fondo de Adaptación y Ministerio de Agricultura. Por el contrario, la Policía Nacional, la DIAN y el Ministerio de Hacienda hacen alusión a ciertos lineamientos o principios.

Códigos de Ética

Respecto a los códigos de ética, la mayoría de entidades manifiestan contar con unas normas de seguridad de la información (Coldeportes, DIMAR, Fondo de Adaptación, Ministerio de Agricultura, Dian, Función Pública y Ministerio de Salud), pero no con códigos de ética para el manejo de datos. En este sentido, las entidades buscan promover un sentido de responsabilidad en los funcionarios hacia el buen manejo de los datos. Por otro lado, al indagar por las áreas que deberían responsabilizarse por la gestión de datos se encontró que la mayoría considera que este debería ser un tema transversal de todas las áreas con soporte del área de TI.

Bases de Datos Internas y Bases de Datos Externas

En esta sección se identifican las bases de datos internas y externas que utilizan las entidades para el cumplimiento de sus

Tabla 19: Panorama de las entidades respecto a las bases de datos internas

Entidad	Bases de datos internas utilizadas por entidad
Coldeportes	22 sistemas de información.
Colombia Compra Eficiente	La tienda virtual del Estado colombiano; Secop I; Secop II; VORTAL NEXT , plataforma transaccional para publicar adquisiciones y procesos de contratación.
DIMAR	Datos básicos a cada una de las áreas , de la persona a cargo en cada subdirección, teléfono y correo electrónico.
Fondo de Adaptación	Bases de datos de personas que han sido beneficiadas o damnificadas del fenómeno de la niña en el 2011; Datos de los diferentes proyectos.
Policía Nacional	Son usuarios de datos.
IDEAM	Un inventario de todos los procedimientos de trámites para los ciudadanos; Sistema de Información Ambiental Colombiano) en donde se consigna y organizada toda la información del medio ambiente.
Ministerio de Agricultura	Consumidora de información; recolecta información de área, de producción y rendimiento de una región; Datos de empleo de cadenas productivas.
DIAN	Información tributaria de declaraciones de los contribuyentes; Información de comercio exterior.
Función Pública	SIGEP , da cuenta de los funcionarios del Estado e incluye información de las diferentes entidades públicas; FURAG , evaluación continua de la gestión de las entidades.
Ministerio de Hacienda	Son usuarios de datos.
Ministerio de Trabajo	
Ministerio de Salud	El 100% de la información está digitalizada; desarrollan prácticas en gestión.

Fuente: Elaboración propia

Tabla 20: Panorama de las entidades respecto a las bases de datos externas

Entidad	Bases de datos internas utilizadas por entidad
Coldeportes	Información de otros organismos del deporte; Sistema de Información Nacional del Deporte; Información de Registraduría.
Colombia Compra Eficiente	No requieren de bases de datos externas; interoperabilidad con el Sistema Integrado de Información Financiera y el Sistema Único Empresarial.
DIMAR	Requieren información de centros médicos; información de la Registraduría para agilizar trámites
Fondo de Adaptación	Base de datos de la Universidad Nacional con información de damnificados de los diferentes fenómenos.
Policía Nacional	Bases de datos de la Registraduría, Ministerio de Justicia, Contraloría y Procuraduría.
IDEAM	Imágenes de un satélite 'Land Sat'; Satélite del programa Copérnico.
Ministerio de Agricultura	Bases de datos del DANE, gremios e Ideam; Información que procesan para sacar nuevos datos para publicación.
DIAN	Bases del dato de Importaciones y Exportaciones del DANE; Información de agencias tributarias de otros países.
Función Pública	Bases de datos del DANE y DNP.
Ministerio de Hacienda	Bases de datos de La Gran Encuesta Integrada del DANE, La Encuesta Manufacturera y Las bases de Datos de Importaciones y Exportaciones de la DIAN.
Ministerio de Trabajo	Bases de datos del Ministerio de Salud y Protección Social; RUES, RUT, Sisbén e información Prosperidad Social.
Ministerio de Salud	Información de la Registraduría; Bases de datos de Migración Colombia, DNP con Sisbén y estadísticas vitales del DANE.

Fuente: Elaboración propia

funciones. A continuación, se presentan las bases de datos que tienen las entidades participantes y los datos externos que utilizan para sus procesos dentro de la organización.

Servicios en la Nube

Se identificó una actitud positiva por parte de la mayoría de los entrevistados frente al uso de los servicios en la nube. En algunos casos, se asegura que están utilizándose (Coldeportes, Fondo de Adaptación, Ministerio de Agricultura, Función Pública y Ministerio de Salud); sin embargo, enfatizan que son nubes privadas, las cuales les dan mayor seguridad y tranquilidad en la protección de los datos. Por otro lado, las entidades que no utilizan este servicio afirman que se están haciendo estudios con respecto a la posibilidad de adquirirlos. Solo en el caso de la Policía Nacional se evidencia una renuencia a hacerlo, pues hay preocupación por la seguridad de la información.

Los principales riesgos que le atribuyen a la nube son: a) pérdida de información; b) falta de seguridad y; c) falta de disponibilidad de la información. Asimismo, mencionan algunas de las soluciones a estos riesgos, que incluyen: a) dar garantías por parte del proveedor en cuanto a que la información va a estar salvaguardada; b) requerimientos técnicos y planes de contingencia en el centro de datos; c) Back-ups mensuales; d) modernización tecnológica; e) mayores recursos económicos.

Apertura de Datos

En general, entre los entrevistados, existe una cultura por compartir información, en donde los principales motivadores para hacerlo son: a) la posibilidad de generar más datos e información para análisis, innovaciones para la toma de decisiones; b) dar cumplimiento a la ley de transparencia, es decir, ayudar a reducir la corrupción en el sector público. Aisladamente, algunas entidades mencionaron que la apertura de datos podría permitir una mayor visibilidad y posicionamiento de su entidad.

Por otro lado, se hace alusión a una serie de barreras y riesgos que han obstaculizado el proceso de apertura para algunas de las entidades. Por ejemplo, las entidades mencionan que hay datos sensibles y confidenciales que no se pueden publicar, una falta de liderazgo y compromiso con el tema, y duplicidad de la información. En menor medida, se afirma que estos problemas están relacionados con la calidad de la fuente.

De acuerdo con la situación actual de las entidades se puede afirmar que, en la mayoría de los casos, la decisión acerca de qué información publicar es el resultado del análisis que se hace internamente, en cuanto a lo que consideran puede ser relevante e interesante para el ciudadano, lo cual puede limitar la información al usuario final. Por otro lado, también es el resultado de los requerimientos del usuario.

Interoperabilidad

En general, en las entrevistas se definió la interoperabilidad como “el intercambio de información entre dos entidades estatales en pro de agilizar procesos”. Lo que motiva a las diferentes entidades a interoperar es agilizar y facilitar los procesos, unificar datos, además de la posibilidad de contar con información a tiempo real.

Ahora, a pesar de lo anterior, se encontró que algunas entidades no han podido desarrollar un proceso de interoperabilidad (Fondo de Adaptación y DIAN) o, en caso de hacerlo, cuentan todavía con importantes retos respecto al nivel de conocimiento respecto a la plataforma de interoperabilidad del MinTic (DIMAR, Coldeportes, e Ideam). Esto se debe a la existencia de algunas barreras como las limitaciones en cuanto a capacidades técnicas y recursos, la falta de cultura de datos por parte del personal, el temor frente a una calidad inadecuada del dato y obstáculos jurídicos en las diferentes entidades. En menor medida, también se identificó una renuencia a compartir información por el temor hacia la calidad del dato, desconocimiento

Tabla 21: Panorama de las entidades respecto a la interoperabilidad

Entidad	Fortalezas	Debilidades
Coldeportes	Sistemas de interoperabilidad para consulta (SUPÉRATE, INDER ANTIOQUIA); Plataforma de historias clínicas.	No cuentan con Interoperabilidad integrada de operación.
Colombia Compra Eficiente	El SECOP II y La Tienda Virtual interoperan con el SIIF ; Interoperabilidades para traer información de sistemas relacionados con la compra pública.	
DIMAR	Se comparte información con Armada Nacional, Ministerio de Defensa, MinTic y la Fiscalía.	Bajo nivel de conocimiento con respecto a la plataforma de interoperabilidad de MinTic
Fondo de Adaptación		No cuentan con Interoperabilidad. Actualmente, la información llegar en un medio físico.

Entidad	Fortalezas	Debilidades
Policía Nacional	Sistema de interoperabilidad con la Registraduría y con el Ministerio de Defensa.	
IDEAM	Sistema de interoperabilidad a través del SIAC.	No hay claridad del uso de la plataforma de MinTic.
Ministerio de Agricultura	Interoperabilidad automática con el DANE.	
DIAN		El estándar es adecuado pero la infraestructura no es suficiente para la cantidad de información.
Función Pública	Interoperan con 10 entidades ; El SIGEP actual se hace por cargas masivas .	
Ministerio de Hacienda	Integran datos recopilando información de las diferentes áreas del Ministerio.	
Ministerio de Trabajo	Intercambian información con Ministerio de Salud, DANE, SENA, con el DPS y con la Unidad de Víctimas; Cruces de información y validación para que los beneficiados; Cruces de información de FILCO con el DANE y con el Ministerio de Salud.	
Ministerio de Salud	Actualización de cédulas nuevas con la Registraduría; Diferentes entidades consultan su sistema; Las pagadoras de pensión para garantizar que le están pagando a las personas, ingresan al sistema PISIS del Ministerio.	

Fuente: Elaboración propia

y falta de liderazgo. La mayoría de los entrevistados consideran que la entidad que debería encargarse de controlar todo lo relacionado al intercambio de información apunta al DNP.

Analítica de Datos

En términos generales, se encontró que pocas de las entidades cuentan con soluciones respecto a la analítica de datos (Policía Nacional, IDEAM, Dian, Función Pública y Ministerio de Hacienda). Aunque la mayoría reconoce su importancia y la necesidad de hacer este proceso, en la actualidad, están revisando y solucionando aspectos relacionados con las fases iniciales de este proceso (i.e, captura, procesamiento, almacenamiento, etc.).

Se identifica que los principales motivadores asociadas a este proceso son: (1) la necesidad de generar información para dar respuesta a las problemáticas que surgen en el sector y poder tomar decisiones más acertadas; (2) la iniciativa y liderazgo de los directores y coordinadores de las entidades. Entre las barreras identificadas para la analítica de datos, se observa la falta de recursos económicos y de un recurso humano capacitado. Adicionalmente, aunque en menor medida, algunas entidades reportan una falta de cultura y liderazgo en lo que a la gestión de datos se refiere (Colombia Compra Eficiente, Ministerio de Trabajo y Ministerio de Salud). De manera aislada, algunas entidades reportan que hacen uso de métodos predictivos e inferenciales y la mayoría coincide en que utiliza métodos descriptivos (Coldeportes, DIMAR y Ministerio de Agricultura).

También se identificó que los funcionarios que hacen el análisis de los datos son en su mayoría contratistas. Esto es un aspecto que, en algunas instancias, puede llegar a obstaculizar el análisis constante de los datos debido a los cambios del personal, lo que conlleva a una falta de continuidad y aprendizaje del conocimiento. Asimismo, la mayoría perfiles que responden a estas necesidades son: economistas, estadísticos, ingenieros industriales e ingenieros de sistemas. En segunda instancia, los perfiles corresponden a relaciones internacionales, administración y contaduría.

En este contexto, las entidades entrevistadas resaltan que es necesario un equipo multidisciplinario que incluya personas con conocimientos numéricos y gusto por la investigación. Finalmente, afirman que debería haber Científicos de datos que integren los equipos para la elaboración de la analítica - estos son aquellos usan tecnologías de analytics avanzada para proveer análisis que trasciende la estadística tradicional. No obstante, reconocen que es un recurso humano al cual es difícil acceder.

Tableros de Control

Con respecto a los tableros de control, la mayoría de las entidades asegura tenerlos (Coldeportes, Colombia Compra Eficiente, Fondo de Adaptación, Policía Nacional, Ideam, Ministerio de Agricultura, DIAN, Ministerio de Hacienda, Ministerio de Trabajo) y, en las entidades que carecen de esta herramienta, resaltan el interés por hacerlo (Función Pública y Ministerio de Salud).

Tabla 22: Panorama de las entidades respecto a los tableros de control

Entidad	Fortalezas	Debilidades
Coldeportes	Tableros de control con seguimiento mensual y se alimentan con fuentes de cada área.	No hay un control de calidad
Colombia Compra Eficiente	Sí cuentan con unos tableros que miden el comportamiento de la compra pública .	
DIMAR		Desconocen los tableros de control.
Fondo de Adaptación	Tableros dentro del sistema de gestión de proyectos.	
Policía Nacional	Tableros de tendencia criminal; Tableros estratégicos descriptivos de talento humano para comandancia.	
IDEAM	Indicadores de reporte.	
Ministerio de Agricultura	Tableros (Agronet) del sector agropecuario.	
DIAN	En desarrollo (Power BI).	
Función Pública		No cuentan con tableros de control.
Ministerio de Hacienda	Cuentan con diferentes tableros de control que llevan un seguimiento de la economía.	
Ministerio de Trabajo	Tableros para la visualización (Tableau).	
Ministerio de Salud		No cuentan con tableros de control.

Fuente: Elaboración propia

ANEXO B. GUÍA DE PREGUNTAS PARA LAS ENTREVISTAS REALIZADAS A LAS ENTIDADES

I. CONTEXTUALIZACIÓN Y CONCEPTUALIZACIÓN GENERAL

- **Entendimiento los conceptos** de “Explotación de datos” y “BIG DATA” à Para comenzar, me gustaría entender lo que significa para usted el concepto de “Explotación de datos” y “BIG DATA”... ¿A qué hace referencia? ¿Cuál diría usted que es su nivel de relevancia para su entidad?
- **Innovaciones desarrolladas** ¿En los últimos dos años se ha generado dentro de su entidad algún tipo de novedad relacionada con datos que haya producido algún impacto ya sea a nivel interno o externo? ¿Qué tipo de novedad fue? ¿Cómo definiría usted el impacto?

II. ENTENDIMIENTO DEL PANORAMA ACTUAL DE LAS DIFERENTES ENTIDADES EN LO QUE A LA GENERACIÓN, ANÁLISIS Y DIFUSIÓN DE LOS DATOS SE REFIERE

- Por ley o normatividad, ¿Qué bases de datos su entidad está obligada a levantar, procesar, o difundir?
- Puede decirnos ¿Cuáles son las principales bases de datos externas que su entidad maneja o requiere para algún análisis o proceso interno?
- **Digitalización de la información** ¿Qué entiende por digitalizar la información? Teniendo en cuenta el 100% de los datos e información que genera su entidad ¿Qué porcentaje está digitalizada? ¿Cómo lo hacen? ¿Con qué fin?
- ¿Y aquella que NO está digitalizada? ¿Qué tipo de información es? ¿Cuál es la razón de no hacerlo? ¿Considera que existen procesos que “nunca” podría prescindir del soporte de papel?
- **Planes de digitalización** En caso de no tenerla digitalizada ¿Tienen algún plan para hacerlo? En caso de no tenerlo, ¿por qué razón? En caso de tenerlo ¿Qué los llevó a eso?
- ¿En qué consiste este plan? ¿Cuándo han pensado implementarlo?
- **Toma de decisiones** Teniendo en cuenta el 100% de la información que tienen digitalizada ¿Qué porcentaje de esta se utiliza para la toma de decisiones dentro de la entidad? ¿Me puede por favor dar ejemplos del tipo de información y del tipo de decisiones?

- **Políticas actuales** ¿Tiene su entidad alguna política de gestión de datos establecida? ¿Cómo se estableció? ¿En qué consiste esa política? ¿Desde cuándo existe?
- ¿Cree que la gestión de datos es responsabilidad del área TI? ¿de la oficina de planeación? O ¿transversal?
- **Códigos de ética** ¿Siguen algún código de ética en lo que a la gestión de datos se refiere? ¿Su entidad sigue algún manual de buenas prácticas para el manejo y gestión de información? ¿Cuál?

INFRAESTRUCTURA DE DATOS

- ¿Cuál es la posición de su entidad con respecto a utilizar los servicios de la nube en lo que a los datos se refiere?
- ¿Cuáles son los principales riesgos de hacerlo? ¿Y las barreras? ¿En qué sentido?
- ¿Cómo contrarrestan o contrarrestarían ustedes estos riesgos y barreras? ¿Hay algún protocolo de seguridad? ¿Cuál?
- **Apertura de datos** ¿Qué entiende usted por datos abiertos? ¿Qué piensa de la publicación de datos abiertos? ¿Lo hacen en su entidad? ¿Que los motiva a hacerlo?
- ¿Qué otras estrategias o planes relacionados a planes abiertos desarrollan en su entidad?
- ¿Qué tipo de información o datos publican? ¿Cómo la publican? ¿Es una información que les piden sus usuarios o no? ¿Cuál ha sido la reacción de sus usuarios con respecto a esta información?
- ¿Cuáles son en su caso las limitantes para publicar datos abiertos? ¿Por qué razón las considera limitantes? ¿Cómo se pueden contrarrestar?
- Y por el contrario ¿Qué piensa de utilizar datos abiertos de otras entidades para el funcionamiento de la suya? ¿Razones asociadas, lo han hecho?
- ¿Con qué tipo de Entidades/ empresas comparten información a través de sistemas interoperables, APIs, etc.? (sector público/ privado).
- ¿Su entidad cuenta con prácticas frecuentes para evaluar internamente la calidad de los datos que se comparten?
- ¿Qué tipo de datos de otras entidades buscan y con qué fin? ¿Cuáles son las decisiones que toman con la información que buscan de otras entidades?
- A su criterio ¿Cómo es la proyección en cuanto a la apertura de los datos en su entidad a futuro?
- **Interoperabilidad** ¿Qué entiende usted por interoperabilidad? ¿Qué fortalezas le encuentra? y ¿Qué debilidades? ¿En qué sentido?
- ¿En su entidad han utilizado la plataforma de interoperabilidad? ¿Por qué razón?
- ¿En caso de usarla, que los motivó a hacerlo?
- ¿Cómo lo hacen? ¿Con otras entidades? ¿Al interior de la entidad? ¿De ambas formas?
- ¿Qué barreras o limitaciones tiene la plataforma de interoperabilidad para el uso por parte de su entidad? ¿Han hecho algo para solucionarlas? ¿Por qué razón? ¿Cómo cree que podrían contrarrestarse dichas barreras?
- En caso de no hacerlo hoy en día ¿Tienen algún plan para hacerlo a futuro? ¿Qué tipo de planes? ¿Cuándo están pensando hacerlo?

ANALÍTICA DE DATOS

- ¿Tienen implementada alguna solución de analítica de datos (analizar datos con herramientas automatizadas y técnicas científicas) actualmente? ¿Hace cuánto?
- ¿Qué tipo de analítica utilizan más? ¿Métodos descriptivos, exploratorios, inferenciales, predictivos o de causalidad?
- ¿Utilizan algoritmos? ¿Cómo ha sido esa experiencia aplicando algoritmos? ¿Cómo ha sido su impacto dentro de la entidad?
- ¿Cuál o cuáles son esas soluciones? ¿Me las podría describir por favor?
- Con adjetivos calificativos, ¿cómo describiría la analítica de datos que se lleva a cabo en su entidad? ¿Por qué razón?
- **Motivadores** ¿De dónde surgió la idea de implementar la solución? ¿Qué los motivó a implementarla?

- **Barreras/ Inhibidores** ¿Cuáles han sido las principales barreras para la implementación de estas soluciones?
- ¿Qué necesita una entidad pública para tomar la decisión de implementar estos proyectos y tener éxito?
- **Soluciones** ¿Qué han hecho para contrarrestar estas barreras? ¿Cuál es el resultado de esto hoy? ¿lo haría con contratación externa o crearía un grupo interno?
- ¿Quién es la persona o grupo responsable de hacer la analítica de datos? ¿Qué tipo de contratación tiene esa persona?

En caso de no tener un responsable

- ¿Por qué razón no tienen actualmente un equipo responsable para hacer este proceso? ¿Planean hacerlo en algún momento?

En caso de tenerlo

- ¿Quiénes y cuántas personas conforman el equipo? ¿Cuál es su perfil? ¿Qué competencias y capacidades deben tener? ¿Por qué razón? ¿Fue difícil conseguir el perfil?
- ¿Qué tipo de contratación tienen con la entidad? ¿Qué pasa cuando cambia el líder de la entidad? ¿Pierden relevancia o por el contrario se mantiene?
- En caso de no haberse mencionado anteriormente, ¿Qué profesiones tiene o tendría en cuenta para realizar la analítica de datos? Razones asociadas.

TABLEROS DE CONTROL

- ¿Qué entiende por tableros de control?
- ¿Su entidad tiene tableros de control?
- En caso de no tenerlos ¿Cuál es la razón por la cual no los tienen? ¿Han pensado en adquirirlos? ¿Cuándo tienen pensado hacerlo?
- En caso de tenerlos ¿Con qué fuentes de información alimentan estos tableros de control? ¿Con qué periodicidad? ¿Cómo aseguran la calidad de la información? ¿Qué protocolos tienen para asegurar la calidad de los datos? ¿Cómo hacen seguimiento a la información que los alimenta?

I. PROYECTOS FUTUROS Y ESTRATEGIAS CON RESPECTO A TI Y FORTALECIMIENTO DE CULTURA DE DATOS

PARA CONCLUIR

- **Fortalezas de explotación de datos Para concluir y viéndolo desde su cargo** ¿Cuáles cree que son las principales fortalezas de la explotación de datos? ¿En qué sentido?
- **Debilidades** ¿Y las debilidades? ¿Por qué razón?
- **Riesgos Para su entidad** ¿Cuáles son los principales riesgos en lo que a la explotación de datos se refiere?
- **Soluciones** ¿Cómo cree usted que el riesgo de mayor relevancia podría contrarrestarse? ¿En qué sentido?
- **Fortalecimiento de cultura de datos** ¿Qué entiende por cultura de datos? ¿Qué estrategias dentro de su entidad podrían desarrollarse para fortalecer la cultura de datos? ¿Qué tan viable es hacerlo?
- **Plan Estratégico de TI** ¿Me podría contar un poco sobre los planes o proyectos que tienen en cuanto a la tecnología de información en el corto, mediano y largo plazo? ¿Cómo se visualizan en este sentido?
- **Para terminar, a futuro** ¿Qué proyectos de Big Data considera que su entidad podría implementar? O continuar implementando en caso de que ya exista uno en marcha.

ANEXO C. DIMENSIÓN DE GASTO E INVERSIÓN ASOCIADA A LA EXPLOTACIÓN DE DATOS

(TABLAS 16-19)

Hardware o elementos de Infraestructura:

- Equipos de procesamiento
- Equipos de almacenamiento
- Equipos de redes y comunicaciones
- Equipos de impresión y otros
- Equipos de Redes
- Equipos de seguridad y ambientales para centros de cómputo

Software, Sistemas de Información o Aplicaciones

- Desarrollo y/o adquisición de aplicaciones o sistemas de Información
- Mantenimiento o modificaciones de aplicaciones o sistemas de Información
- Adquisición o renovación de licencias
- Sistemas operativos
- Software de ofimática,
- Aplicaciones comerciales,
- Herramientas de programación

Servicios

- Outsourcing de mesa de ayuda
- Call Center
- Contratación de consultorías o capacitaciones
- Arrendamiento de equipos
- Intranet Gubernamental (Infraestructura tecnológica)
- Centro de datos y servicios de base
- Centro de contacto al ciudadano
- Plataforma de interoperabilidad
- Estándares y políticas: GEL-XML
- Enrutador transaccional
- Gestión del cambio
- Divulgación y promoción informática y TIC
- Servicios de Conectividad a Internet
- Servicios electrónicos

ANEXO D. DIMENSIÓN 5, PORCENTAJE DE ENTIDADES ENCUESTADAS

Tabla 23: Contratos Públicos por tipo de Costo (COP)

	Costos de Inversión		Costo de Operación y Mantenimiento
	Activos Tangibles	Activos Intangibles	
Servicios de sistemas y administración de componentes de sistemas	0	42.174.163.355	39.399.869.054
Ingeniería de software o hardware	0	49.465.700.089	7.825.921.617
Mantenimiento y soporte de software	0	27.666.704.616	16.108.597.145
Computadores	16.795.373.675	2.500.000.000	4.789.523.340
Servicios de telecomunicaciones	0	19.827.306.461	6.007.818.207
Servicios de alquiler o arrendamiento de equipos o plataformas de voz y datos o multimedia	0	14.549.906.240	326.020.765
Software de seguridad y protección	7.680.664.573	1.607.607.745	1.227.401.030
Cartografía	0	14.635.187.505	0
Software funcional específico de la empresa	5.293.029.337	2.166.895.828	3.555.040.989
Servicios de datos	0	6.558.364.063	5.125.441.927
Mantenimiento y soporte de hardware de computador	0	5.598.750.365	5.144.276.172
Sistemas de manejo de información MIS	0	4.501.565.066	5.099.772.798
Servicios de alquiler o arrendamiento de licencias de software de computador	0	815.238.234	6.421.164.130
Análisis económico	0	8.683.027.919	0
Software de consultas y gestión de datos	1.524.508.313	1.907.325.479	4.504.397.598
Dispositivos de almacenamiento	6.510.959.933	228.032.892	1.434.361.000
Metodología y Análisis	0	2.998.706.306	4.171.928.071
Resto	19.393.038.902	38.778.929.217	19.392.966.844
Total	57.197.574.733	244.663.411.380	130.534.500.687

Fuente: SECOP II

Tabla 24: Gastos de inversión en TIC por Industria

Gastos de Inversión TIC por Industria y Proporción de Gasto TIC para el Mercado de Datos EUR Millones							
Industria	2013	2014	2015	2016	CAGR 2013-2016 (%)	Proporción de Gasto Total TIC para Mercado de Datos 2016	
						Inversión	(%)
Servicios profesionales	€ 86.497	€ 106.409	€ 117.151	€ 118.191	10,97%	€ 8.510	7,20%
Información y comunicaciones	€ 95.374	€ 104.271	€ 110.332	€ 111.763	5,43%	€ 10.170	9,10%
Servicios financieros	€ 93.290	€ 97.817	€ 104.754	€ 106.309	4,45%	€ 11.800	11,10%
Minorista y mayorista	€ 77.764	€ 63.701	€ 65.704	€ 66.287	-5,18%	€ 6.297	9,50%
Minería, fabricación	€ 47.072	€ 61.813	€ 65.084	€ 65.211	11,48%	€ 7.434	11,40%
Casa	€ 76.188	€ 45.274	€ 44.629	€ 44.400	-16,47%	€ 2.087	4,70%
Administración Pública	€ 43.448	€ 34.153	€ 35.331	€ 34.394	-7,49%	€ 3.336	9,70%
Utilidades	€ 13.539	€ 22.098	€ 23.553	€ 24.389	21,68%	€ 2.463	10,10%
Transporte y almacenamiento	€ 16.286	€ 17.393	€ 19.215	€ 20.011	7,11%	€ 2.681	13,40%
Cuidado de la salud	€ 15.947	€ 15.639	€ 16.116	€ 16.259	0,65%	€ 1.854	11,40%
Educación	€ 11.166	€ 12.154	€ 13.637	€ 14.383	8,81%	€ 1.568	10,90%
Construcción	€ 5.715	€ 4.637	€ 4.339	€ 3.752	-13,09%	€ 274	7,30%
Total EU28	€ 582.285	€ 585.358	€ 619.844	€ 625.347	2,41%	€ 59.408	9,50%

Fuente: SECOP II

Tabla 25: Valor del Mercado de Datos por Industria

Valor del Mercado de Datos por Industria EUR Millones							
Industria	2013	2014	2015	2016	2017	2017 (%)	CAGR 2013-2017 (%)
Servicios financieros	€ 9.407	€ 10.091	€ 10.760	€ 11.816	€ 13.030	20,04%	8,49%
Servicios profesionales	€ 6.982	€ 7.338	€ 7.855	€ 8.490	€ 9.117	14,02%	6,90%
Minorista y mayorista	€ 5.107	€ 5.451	€ 5.810	€ 6.319	€ 6.950	10,69%	8,01%
Información y comunicaciones	€ 4.457	€ 4.876	€ 5.231	€ 5.865	€ 6.491	9,98%	9,85%
Administración Pública	€ 2.628	€ 2.808	€ 3.011	€ 3.323	€ 3.607	5,55%	8,24%
Transporte y almacenamiento	€ 2.212	€ 2.307	€ 2.464	€ 2.690	€ 2.928	4,50%	7,26%
Utilidades	€ 1.892	€ 2.107	€ 2.240	€ 2.466	€ 2.710	4,17%	9,40%
Casa	€ 1.679	€ 1.771	€ 1.871	€ 2.071	€ 2.165	3,33%	6,56%
Cuidado de la salud	€ 1.454	€ 1.560	€ 1.676	€ 1.846	€ 2.018	3,10%	8,54%
Educación	€ 1.319	€ 1.348	€ 1.442	€ 1.562	€ 1.721	2,65%	6,88%
Construcción	€ 226	€ 239	€ 254	€ 275	€ 295	0,45%	6,89%
Total EU28	€ 47.420	€ 50.888	€ 54.351	€ 59.539	€ 65.032	100%	8,22%

Fuente: SECOP II

Tabla 26: Valor del Mercado de Datos por Industria

# Compañías de Datos (Proveedores, 000s)	2013	2014	2015	2016	2017	CAGR (2013-2017)
United States of America	273.535	277.821	283.340	289.556	302.810	2,57%
Brazil	34.346	34.840	35.456	35.979	36.387	1,45%
Japan	95.705	95.919	99.001	101.812	104.864	2,26%
EU28	239.845	243.600	249.100	251.450	276.450	3,81%

# Valor del Mercado de Datos (EUR Millones)	2013	2014	2015	2016	2017	CAGR (2013-2017)
United States of America	€ 97.237	€ 103.935	€ 115.519	€ 128.173	€ 145.546	10,81%
Brazil	€ 4.703	€ 5.289	€ 5.272	€ 6.049	€ 6.310	7,83%
Japan	€ 23.021	€ 22.228	€ 24.013	€ 25.513	€ 27.723	4,78%
EU28	€ 47.420	€ 50.888	€ 54.351	€ 59.486	€ 65.038	8,22%

# Valor de la Economía de Datos (Impacto Directo, EUR Millones)	2013	2014	2015	2016	2017	CAGR (2013-2017)
United States of America	€ 93.708	€ 99.398	€ 109.246	€ 108.521	€ 113.677	4,95%
Brazil	€ 4.521	€ 5.289	€ 5.477	€ 6.157	€ 6.395	9,08%
Japan	€ 22.250	€ 21.367	€ 22.928	€ 27.394	€ 29.849	7,71%
EU28	€ 44.113	€ 48.098	€ 52.087	€ 53.508	€ 65.038	10,19%

# Valor de la Economía de Datos (Impacto Indirecto, EUR Millones)	2013	2014	2015	2016	2017	CAGR (2013-2017)
United States of America	€ 3.528	€ 4.536	€ 6.273	€ 7.270	€ 7.788	21,81%
Brazil	€ 182	€ 217	€ 290	€ 290	€ 298	13,12%
Japan	€ 771	€ 880	€ 1.085	€ 1.189	€ 1.289	13,27%
EU28	€ 1.674	€ 1.992	€ 2.264	€ 2.780	€ 3.303	18,51%

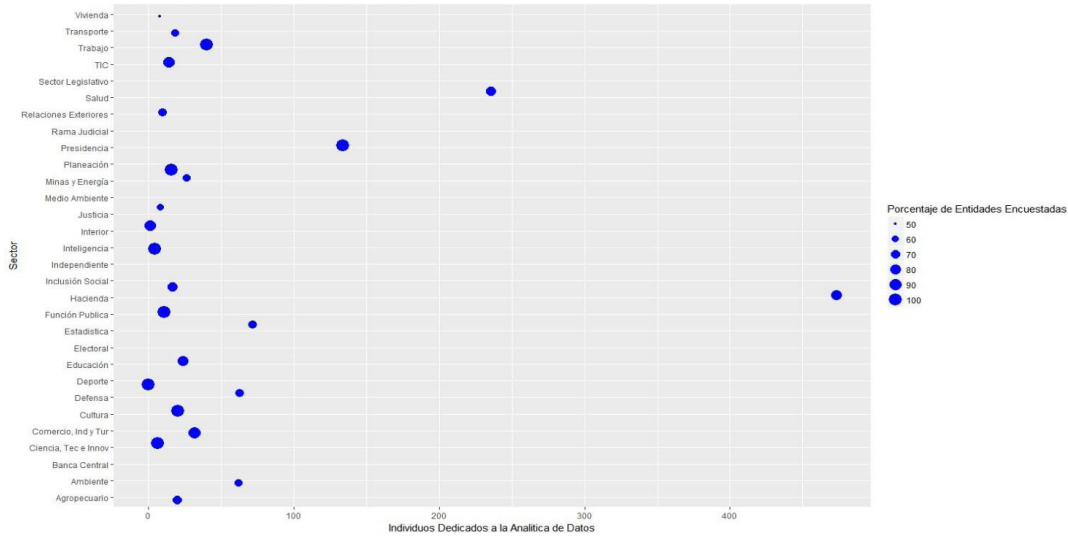
Incidencia de la Economía de Datos en el PIB	2013	2014	2015	2016	2017
United States of America	0,58%	0,60%	0,71%	0,78%	0,81%
Brazil	0,20%	0,23%	0,36%	0,16%	0,16%
Japan	0,47%	0,48%	0,65%	0,93%	0,96%
EU28		0,36%	0,37%	0,37%	0,52%

Valor de la economía de los Datos Principales 6 países UE (TOTAL IMPACTOS) Gastos totales TIC en Mercado de Datos										
EUR Millones										
						Share	CAGR	Gasto TIC	Inversión Mercado de Datos	Porcentaje gastos TIC para Mercado de Datos
	2013	2014	2015	2016	2017	2017	2013-2017	2016		
Alemania	€ 64.952	€ 67.360	€ 74.274	€ 77.028	€ 84.121	25%	9%	€ 130.774	€ 14.124	11%
Reino Unido	€ 43.833	€ 45.878	€ 56.649	€ 61.290	€ 68.611	20%	16%	€ 109.673	€ 9.871	9%
Francia	€ 34.373	€ 34.500	€ 36.352	€ 37.974	€ 42.405	13%	7%	€ 99.197	€ 7.725	8%
Italia	€ 25.291	€ 26.289	€ 27.514	€ 28.406	€ 29.688	9%	5%	€ 69.662	€ 5.991	9%
España	€ 18.922	€ 19.503	€ 20.335	€ 20.971	€ 23.920	7%	8%	€ 46.743	€ 3.693	8%
Holanda	€ 12.846	€ 13.532	€ 15.792	€ 16.719	€ 19.412	6%	15%	€ 28.931	€ 2.922	10%
Total EU28	€ 246.840	€ 257.589	€ 285.633	€ 299.989	€ 335.618	100%	11%	€ 625.347	€ 59.408	10%

Fuente: SECOP II

ANEXO E. MADUREZ ORGANIZACIONAL PARA LA ADOPCIÓN DE PROCESOS ANALÍTICOS

Figura 23: Individuos Dedicados a la Analítica de Datos / Porcentaje de Entidades Encuestadas



Fuente: Elaboración Propia

Figura 24: Individuos Dedicados a la Analítica de Datos / Porcentaje de Entidades Encuestadas

Aspectos	Aspiracional	Experimentado	Transformado
Motivación	Uso de analítica para justificar las acciones.	Uso de analítica para guiar las acciones.	Uso de la analítica para prescribir las acciones.
Obstáculos claves	Falta de comprensión sobre cómo aprovechar el análisis para obtener valor comercial Patrocinio ejecutivo. Falta de una cultura que fomente compartir información.	Falta de comprensión sobre cómo aprovechar el análisis para obtener valor comercial. Habilidades dentro de la línea de negocio. La propiedad de los datos no está clara o la gobernanza es ineficaz.	Falta de comprensión sobre cómo aprovechar el análisis para obtener valor comercial. Margen de acción debido a las prioridades competitivas. Accesibilidad a los datos.
Gerencia de datos	Habilidad limitada de captura, agregar, analizar o compartir información y conocimientos.	Capacidad moderada para capturar, agregar y analizar datos. Capacidad limitada para compartir información y conocimientos	Capacidad alta para capturar, agregar y analizar datos. Eficaz para compartir información y conocimientos.
Análisis para la acción	En pocas ocasiones se usan enfoques rigurosos para tomar decisiones. Uso limitado de información para guiar las estrategias futuras o las operaciones cotidianas.	Uso, no cotidiano, de enfoques rigurosos para tomar decisiones. Uso creciente de conocimientos para guiar las estrategias futuras. Sin embargo, hay un uso limitado de conocimientos para guiar las operaciones cotidianas.	La mayoría usa enfoques rigurosos para tomar decisiones. Casi todos usan información para guiar las estrategias futuras y la mayoría usa información para guiar las operaciones cotidianas.

Fuente: (LaValle, et al., 2011).

Tabla 28. Madurez organizacional. Características y etapas de crecimiento del gobierno en línea.

Etapa de crecimiento del gobierno electrónico	Actividades e intercambio de información	Instalaciones de TI	Sistemas de datos	Explicación
1. Organizaciones muy jerárquicas	La información y las actividades están limitadas dentro de departamentos separados dentro de la organización.	Todas las actividades son, cuando sea posible, soportadas con TI.	Digitalización de procesos.	Proporcionar herramientas digitales para procesos operativos en la entrada de datos y búsqueda de datos.
2. Organizaciones integradas	Las actividades trascienden los departamentos separados y la información se comparte en todas las organizaciones.	Infraestructura de TI de toda la organización desde la cual se puede acceder a la información en toda la organización.	Inteligencia de negocios y gerencia de datos.	Agrupamiento del almacenamiento de información en toda la organización y herramientas de análisis automatizadas que crean información a partir de los datos internos de la organización.
3. Portal nacional	Las actividades van más allá de los límites de la organización y la información es accesible desde fuera de la organización.	Infraestructura de TI adecuada para el acceso externo de las aplicaciones y para el acceso a la información dentro de la organización.	Inteligencia de negocios 2.0 y gerencia de datos 2.0.	Inteligencia de negocios y administración de datos con opciones para acceder y cambiar datos e información desde fuera de la organización.
4. Integración inter-organizacional	Las actividades y la información se comparten ampliamente con otras organizaciones.	Infraestructura de TI adecuada para el acceso completo de las aplicaciones y para el acceso a la información por parte de las organizaciones	Datos abiertos.	Suministro de tantos datos anónimos como sea posible de la organización como datos públicos, en un formato estandarizado.
5. Gobierno impulsado por la demanda y unido	Las actividades y la información se organizan centralmente y se ponen a disposición de todas las organizaciones y partes interesadas pertinentes.	Instalaciones de TI construidas centralmente que soportan toda la información y las aplicaciones, y totalmente accesibles para todas las partes interesadas.	Big data	Recolección, combinación y análisis de conjuntos de datos grandes y complejos con tecnologías no convencionales para crear nuevos conocimientos para la organización.

Fuente: Traducción propia (Klievink, et al., 2017).

BIBLIOGRAFÍA

- Abarca, Karelys. 2018. “Big data y las políticas públicas”. América Economía. el 28 de junio de 2018. <https://www.americaeconomia.com/analisis-opinion/big-data-y-las-politicas-publicas>.
- “Abecé Asociaciones Público Privadas”. 2013. Bogotá D.C: Departamento Nacional de Planeación.
- “About - Programs”. s/f. Next Canada. Consultado el 5 de diciembre de 2018. <https://nextcanada.com/program-explainer>.
- “Alianza CAOBA Centro de Excelencia Big Data y Data Analytics”. s/f. Alianza Caoba. Consultado el 16 de diciembre de 2018. <http://alianzacaoba.co/>.
- “Icetex Histórico de Becas | Datos Abiertos Colombia”. s/f. la plataforma de datos abiertos del gobierno colombiano. Consultado el 11 de enero de 2019. <https://www.datos.gov.co/Educaci-n/Icetex-Hist-rico-de-Becas/eb6r-dfgr/data>.
- “PROGRAMAS DE EDUCACIÓN SUPERIOR | Datos Abiertos Colombia”. s/f. Consultado el 11 de enero de 2019.
- AGN. s/f. “Archivo General de la Nación | Funciones del AGN”. Archivo General de la Nación. Consultado el 13 de enero de 2019. <http://www.archivogeneral.gov.co/Conozcanos/funciones-del-AGN>.
- Baker, L. R., W. R. Cattell, y M. Sensi. 1976. “Inosine in Experimental Acute Renal Failure”. *Lancet (London, England)* 2 (7994): 1091–92.
- Balbontín, Rodrigo, Jose Andrés Roeschmann, y Andrés Zahler. 2018. “Ciencia, Tecnología e Innovación en Chile: un análisis presupuestario”. Dirección de Presupuestos Gobierno de Chile. <http://ctie.economia.cl/wp-content/uploads/2018/04/Documento-Dipres-Presupuesto-CTIE-2018.pdf>.
- Bhargava, Rahul, Erica Deahl, Emmanuel Letouzé, Amanda Noonan, David Sangokoya, y Natalie Shoup. 2015. “Beyond Data: Reinventing Community Engagement and Empowerment in the Age of Data”. White Paper Series. Data-Pop Alliance.
- Buckley. 2015. “How is big data going to change the world?” World Economic Forum. el 1 de diciembre de 2015. <https://www.weforum.org/agenda/2015/12/how-is-big-data-going-to-change-the-world/>.
- Camargo-Vega, Juan José, Jonathan Felipe Camargo-Ortega, y Luis Joyanes-Aguilar. 2015. “Conociendo Big Data”. *Facultad de Ingeniería* 24 (38): 63–77.
- Certicámara. s/f. “e-Government en Colombia: llegó la hora de los servicios ciudadanos digitales”. Colombia Digital. Consultado el 12 de diciembre de 2018. <https://colombiadigital.net/opinion/columnistas/desde-afuera/item/9812-e-governme-nt-en-colombia-llego-la-hora-de-los-servicios-ciudadanos-digitales.html>.
- “Colombia”. s/f. ICT Development Index 2017. Consultado el 16 de diciembre de 2018. <http://www.itu.int/net4/ITU-D/idi/2017/index.html#idi2017economy-card-tab&COL>.
- “Colombia Compra Eficiente”. 2015. Colombia Compra Eficiente. el 18 de septiembre de 2015. <https://www.colombiacompra.gov.co/colombia-compra/colombia-compra-eficiente>.
- Comisión de Estadística. 2013. “Macrodatos y modernización de los sistemas estadísticos”. E/CN.3/2014/11. Informe del Secretario General. Consejo Económico y Social Naciones Unidas. <https://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-S.pdf>.
- Comisión Europea. 2016. “Comunicación de la Comisión al Parlamento Europeo, Al Consejo, Al Comité Económico y Social Europeo y al Comité de Regiones: Las Plataformas en Línea y el Mercado Único Digital, Retos y Oportunidades para Europa”. SWD 172. Bruselas: Comisión Europea. <https://ec.europa.eu/transparency/regdoc/rep/1/2016/ES/1-2016-288-ES-F1-1.PDF>.
- Coppola, Andrea, Oscar Calvo-Gonzalez, Elizabeth Sabet, Natalia Arjomand, Ryan Siegel, Carrie Freeman, y Neisan Massarrat. 2014. “Big data in Action for Development”. ACS11163. Washington DC: The World Bank Group.
- Corporación Colombia Digital. s/f. “Estudio: ¿cómo está Colombia en gobierno electrónico?” Colombia Digital. Consultado el 12 de diciembre de 2018a. <https://colombiadigital.net/actualidad/soluciones-tic/item/9728-estudio-como-esta-colombia-en-gobierno-electronico.html>.

- . s/f. “Un gobierno electrónico para Colombia”. Colombia Digital. Consultado el 12 de diciembre de 2018b. <https://colombiadigital.net/actualidad/soluciones-tic/item/9721-un-gobierno-electronico-para-colombia.html>.
- DANE. 2017. “Código Nacional de Buenas Prácticas del Sistema Estadístico Nacional”. Departamento Administrativo Nacional de Estadística. https://www.dane.gov.co/files/sen/bp/Codigo_nal_buenas_practicas.pdf.
- . s/f. “ABC del SEN”. Departamento Administrativo Nacional de Estadística. Consultado el 14 de diciembre de 2018. <https://www.dane.gov.co/files/sen/abc-del-SEN.pdf>.
- Dans, Enrique. 2011. “Big Data: una pequeña introducción”. *Enrique Dans* (blog), el 19 de octubre de 2011. <https://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html>.
- “Declaración por un Estado Abierto firmada por el Presidente Santos, será clave en posconflicto”. 2017. Presidencia de la Republica. el 23 de enero de 2017. <http://es.presidencia.gov.co/noticia/170123-Declaracion-por-un-Estado-Abierto-firmada-por-el-Presidente-Santos-sera-clave-en-posconflicto>.
- Departamento para la Gestión Pública Efectiva. s/f. “Sobre E-Gobierno”. Organización de los Estados Americanos. Consultado el 12 de diciembre de 2018. <http://portal.oas.org/portal/sector/sap/dptodemodernizaci%C3%B3ndelestadoygobernabilidad/npa/sobreprogramadeegobierno/tabid/811/default.aspx?language=es-co>.
- Department of Economic and Social Affairs. 2005. “Global E Government Readiness Report 2005”. New York: United Nations. <https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2005-Survey/Complete-survey.pdf>.
- D’Ignazio, Catherine, y Rahul Bhargava. 2015. “Approaches to Building Big Data Literacy”. En , 6. New York City. https://dam-prod.media.mit.edu/x/2016/10/20/Edu_D'Ignazio_52.pdf.
- DNP. s/f. “Colombia primer país en Latinoamérica con una política pública para la explotación de datos Big Data”. Departamento Nacional de Planeación. Consultado el 16 de diciembre de 2018. <https://www.dnp.gov.co/Paginas/Colombia-primer-pa%C3%ADs-en-Latinoam%C3%A9rica-con-una-pol%C3%ADtica-p%C3%BAblica-para-la-explotaci%C3%B3n-de-datos-Big-Data.aspx>.
- DNP, MINTIC, y SIC. 2018. “Documento CONPES 3920: Política Nacional de Explotación de Datos”. Departamento Nacional de Planeación. <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>.
- Fedesoft. 2015a. “Caracterización de la Brecha de Talento Digital en Colombia 2015”. Observatorio TI. https://www.observatorioti.gov.co/uploads/files/75_48460787-8efd-42ae-b762-5c3625480616.pdf.
- . 2015b. “Estudio de Caracterización del Sector Teleinformática, Software y TI en Colombia”. Observatorio TI. https://www.observatorioti.gov.co/uploads/files/67_ac63e77b-aacb-4974-8650-636889463838.pdf.
- . 2017. “Estudio de Salarios y Profesionales del Sector de Software y TI en Colombia”. Gobierno de Colombia. https://www.observatorioti.gov.co/uploads/files/82_fc5d000f-8477-42ac-ae90-798ee240479b.pdf.
- García, Jesus, Jose Molina, Antonio Berlanga, Miguel Patricia, Alvaro Bustamante, y Padilla Washington. 2018. *Ciencia de datos: Técnicas analíticas y aprendizaje estadístico en un enfoque práctico*. Alfaomega. <https://www.librosyeditores.com/tiendalemoine/ciencias-exactas/11359-ciencia-de-datos-tecnicas-analiticas-y-aprendizaje-estadistico-en-un-enfoque-practico--9789587784251.html>.
- Hellerstein, Joseph. 2008. “The Commoditization of Massive Data Analysis”. O’Reilly Radar. el 20 de noviembre de 2008. <http://radar.oreilly.com/2008/11/the-commoditization-of-massive.html>.
- Henke, Nicolaus, Jordan Levine, y Paul McInerney. 2018. “You Don’t Have to Be a Data Scientist to Fill This Must-Have Analytics Role”. Harvard Business Review, el 5 de febrero de 2018. <https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-a-analytics-role>.

- Järvisalo, J., y N. E. Saris. 1975. "Action of Propranolol on Mitochondrial Functions--Effects on Energized Ion Fluxes in the Presence of Valinomycin". *Biochemical Pharmacology* 24 (18): 1701-5.
- Kelleher, John D., y Brendan Tierney. 2018. *Data science. The MIT Press essential knowledge series*. Cambridge, Massachusetts: The MIT Press.
- Kiron, David, Pamela Kirk Prentice, y Boucher Ferguson. 2014. "The Analytics Mandate". *MIT Sloan Management Review*, mayo. <https://sloanreview.mit.edu/projects/analytics-mandate/>.
- Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.
<http://methods.sagepub.com/book/the-data-revolution>.
- Klievink, Bram, Bart-Jan Romijn, Scott Cunningham, y Hans de Bruijn. 2017. "Big Data in the Public Sector: Uncertainties and Readiness". *Information Systems Frontiers* 19 (2): 267-83. <https://doi.org/10.1007/s10796-016-9686-2>.
- Kruschwitz, Steve LaValle Eric Lesser, Rebecca Shockley, Michael S. Hopkins and Nina. 2010. "Big Data, Analytics and the Path From Insights to Value". *MIT Sloan Management Review*, diciembre.
<https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>.
- Lazer, D., R. Kennedy, G. King, y A. Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis". *Science* 343 (6176): 1203-5. <https://doi.org/10.1126/science.1248506>.
- Letouzé, Emmanuel. 2015a. "Big Data & Development: an Overview". *Primer Series*.
Data-Pop Alliance.
- . 2015b. "The Law, Politics, and Ethics of Cell Phone Data analytics". *Primer Series*. Data-Pop Alliance.
- Letouzé, Emmanuel, y Johannes Jutting. 2015. "Official Statistics, Big Data and Human Development". Data-Pop Alliance.
- Letouzé, Emmanuel, y David Sangokoya. 2016. "Leveraging Algorithms for Positive Disruption: On data, democracy, society and statistics". Data-Pop Alliance.
- Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, y Angela Hung Byers. 2011. "Big Data: The next Frontier for Innovation, Competition, and Productivity | McKinsey". *McKinsey Global Institute*. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>.
- Marcus, Gary, y Ernest Davis. 2017. "Opinion | Eight (No, Nine!) Problems With Big Data". *The New York Times*, el 20 de diciembre de 2017, sec. Opinion. <https://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html>.
- Matt Ariker, Tim McGuire. 2013. "Five Roles You Need on Your Big Data Team".
Harvard Business Review, el 22 de julio de 2013. <https://hbr.org/2013/07/five-roles-you-need-on-your-big-data-team/>.
- Mayer-Schönberger, Viktor, y Kenneth Cukier. 2014. *Big data: a revolution that will transform how we live, work, and think*. First Mariner Books edition. Boston: Mariner Books, Houghton Mifflin Harcourt.
- MINTIC. 2010. "Marco para la Interoperabilidad del Gobierno en línea". *Ministerio de Tecnología de la Información y las Comunicaciones*.
- . 2015. "Manual Estrategia de Gobierno en Línea". http://estrategia.gobiernoenlinea.gov.co/623/articles-7941_manualGEL.pdf.
- . 2016. "G.GEN.03 Guía General de un Proceso de Arquitectura Empresarial".
Ministerio de Tecnología de la Información y las Comunicaciones. https://www.mintic.gov.co/arquitecturati/630/articles-9435_Guia_Proceso.pdf.
- . 2017. "Así serán los Servicios Ciudadanos Digitales". *Ministerio de Tecnologías de la Información y las Comunicaciones*. el 29 de agosto de 2017. <https://www.mintic.gov.co/portal/604/w3-article-57706.html>.
- . 2018a. "Empoderar al ciudadano para interactuar con el Estado mediante las TI". *Ministerio de Tecnologías de la Información y las Comunicaciones*. 2018. <https://www.mintic.gov.co/portal/604/w3-propertyvalue-677.html>.

- . 2018b. “Manual de Gobierno Digital”. https://www.mintic.gov.co/portal/604/articles-81473_recurso_1.pdf.
- . 2018c. “Niveles de Lenguaje Común de Intercambio de Información”. Portal de Intercambio de Lenguaje Común. 2018. <http://lenguaje.mintic.gov.co/niveles-de-lenguaje-comun-de-intercambio-de-informacion>.
- . 2018d. “La nueva política de Gobierno Digital promueve la proactividad y la innovación ciudadana”. Ministerio de Tecnologías de la Información y las Comunicaciones. el 27 de junio de 2018. <https://www.mintic.gov.co/portal/604/w3-article-75180.html>.
- . 2018e. “Guía de Computación en la Nube”. Gobierno en Línea. el 29 de junio de 2018. <http://estrategia.gobiernoenlinea.gov.co/623/w3-article-75246.html>.
- . 2018f. “Manual de Condiciones Servicio de Interoperabilidad”. Gobierno de Colombia. http://micrositios.mintic.gov.co/servicios_ciudadanos_digitales/interoperabilidad/4_1_manual_condiciones_servicio_interoperabilidad_v270818.pdf.
- . s/f. “Autenticación”. Ministerio de Tecnologías de la Información y las Comunicaciones. Consultado el 12 de diciembre de 2018a. <https://www.mintic.gov.co/portal/604/w3-article-8232.html>.
- . s/f. “G.INF.01 Guía Técnica Básica de Información - Arquitectura TI”. Ministerio de Tecnología de la Información y las Comunicaciones. Consultado el 16 de diciembre de 2018b. <https://www.mintic.gov.co/arquitecturati/630/w3-article-9253.html>.
- . s/f. “Servicios Ciudadanos Digitales”. Consultado el 12 de diciembre de 2018c. http://micrositios.mintic.gov.co/servicios_ciudadanos_digitales/.
- MINTIC, y Fedesoft. s/f. “Informe de Demanda ‘Retos y Oportunidades Para las Empresas TI en el Sector Gobierno’”. Consultado el 12 de diciembre de 2018. https://www.observatorioti.gov.co/uploads/files/52_1aa441ce-3835-438c-b40c-ef9d1450db0d.pdf.
- Naser, Alejandra, y Gastón Concha. 2012a. “Datos abiertos: un nuevo desafío para los gobiernos de la región”, marzo. <https://repositorio.cepal.org/handle/11362/7331>.
- . 2012b. “Datos abiertos: un nuevo desafío para los gobiernos de la región”, marzo. <https://repositorio.cepal.org/handle/11362/7331>.
- Naser, Alejandra, y Álvaro Ramírez Alujas. 2017. “Plan de gobierno abierto: una hoja de ruta para los gobiernos de la región”, septiembre. <https://repositorio.cepal.org/handle/11362/36665>.
- Observatorio TI, y Fedesoft. 2017a. “Indicadores semanales: Brecha del talento digital”.
Semana 1. Observatorio TI. https://www.observatorioti.gov.co/uploads/files/58_3acaca09-cca4-4a2a-9a2c-f920efab9e4d.pdf.
- . 2017b. “Indicadores”. Observatorio TI. https://www.observatorioti.gov.co/uploads/files/49_f6ea06fd-1f30-4a7e-81d5-878042e3c1fd.pdf.
- OECD. 2006. Principios de Gobierno Corporativo de la OCDE 2004. Ministry of Economy and Finances, Spain. https://www.oecd-ilibrary.org/industry-and-services/principios-de-gobierno-corporativo-de-la-ocde-2004_9788485482726-es.
- . , ed. 2013. *Colombia: Implementing Good Governance*. OECD Public Governance Reviews. Paris: OECD.
- . 2014. *OECD Territorial Reviews: Colombia 2014*. OECD Territorial Reviews. OECD Publishing.
- http://www.oecd-ilibrary.org/urban-rural-and-regional-development/oecd-territorial-reviews-colombia-2014_9789264224551-en.
- . 2015a. *Students, Computers and Learning*. PISA. OECD Publishing. https://www.oecd-ilibrary.org/education/students-computers-and-learning_9789264239555-en.
- . 2015b. *Data-Driven Innovation*. OECD Publishing.

https://www.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en.

Organisation Internationales de Normalisation. 2015. "ISO/IEC 11179-1:2015". ISO. diciembre de 2015. <http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/19/61932.html>.

Ortegón, Edgar, Juan Francisco Pacheco, y Adriana Prieto. 2005. *Metodología del marco lógico para la planificación, el seguimiento y la evaluación de proyectos y programas*. CEPAL. <https://repositorio.cepal.org/handle/11362/5607>.

"Plan de Acción y Seguimiento (PAS) - Política nacional de explotación de datos (Big Data)". 2018. Departamento Nacional de Planeación.

"Preguntas Frecuentes - Gobierno Digital". s/f. Estrategia Gobierno en Línea.

Consultado el 7 de diciembre de 2018. <http://estrategia.gobiernoenlinea.gov.co/623/w3-propertyvalue-8164.html>.

Pulgarin Ayala, Gloria. s/f. "Respuesta a su derecho de petición. registro Mintic 497556 de 2012". 10/9/2012. Consultado el 6 de diciembre de 2018. https://www.mintic.gov.co/portal/604/articles-3292_documento.pdf.

Roe, Charles. 2017. "What Is Data Management?" DATAVERSITY. el 4 de diciembre de 2017. <http://www.dataversity.net/what-is-data-management/>.

Rouse, Margaret, Rodney Brown, y Erin Sullivan. 2018. "Petabyte". SearchStorage. junio de 2018. <https://searchstorage.techtarget.com/definition/petabyte>.

Secretaría de Transparencia. s/f. "Ley de Transparencia y del Derecho al Acceso a la Información Pública". Secretaría de Transparencia. Consultado el 14 de diciembre de 2018.

<http://www.secretariatransparencia.gov.co/estrategias/Paginas/ley-transparencia.aspx>.

"Sistema Estadístico Nacional (SEN)". s/f. Consultado el 14 de diciembre de 2018. <https://www.dane.gov.co/index.php/sistema-estadistico-nacional-sen#C3%B3digo-nacional-de-buenas-pr%C3%A1cticas>.

"Stack Overflow Developer Survey 2018". 2018. Stack Overflow. 2018. https://stackoverflow.com/insights/survey/2018/?utm_source=so-owned&utm_medium=social&utm_campaign=dev-survey-2018&utm_content=social-share.

Steel, Emily, Callum Locke, Emily Cadman, y Ben Freese. 2013. "How Much Is Your Personal Data Worth?" *Financial Times*, el 12 de junio de 2013, sec. Media. <https://ig.ft.com/how-much-is-your-personal-data-worth/>.

Suzuki, K., y K. Tauchi. 1975a. "Modification of Michael's Method for Determination of Serum Acetyl Choline Esterase Activity". *Bulletin of Environmental Contamination and Toxicology* 14 (6): 674-80.

———. 1975b. "Modification of Michael's Method for Determination of Serum Acetyl Choline Esterase Activity". *Bulletin of Environmental Contamination and Toxicology* 14 (6): 674-80.

"Technical Note: The Logical Framework". 2012. USAID. https://usaidealarninglab.org/sites/default/files/resource/files/2012_12_logical_framework_technical_note_final_2.pdf.

Tomar, Louisa, William Guicheney, Hope Kyarisiima, y Tinashe Zimani. 2016. "Big Data in the Public Sector: Selected Applications and Lessons Learned". IDB-DP-483.

Inter-American Development Bank.

Unidad Administrativa Especial de Catastro Distrital. 2016. "Infraestructura de Datos Espaciales". Unidad Administrativa Especial de Catastro Distrital. el 17 de junio de 2016.

[https://www.ideca.gov.co/es/noticias?field_fecha_de_publicacion_value\[value\]&page=2](https://www.ideca.gov.co/es/noticias?field_fecha_de_publicacion_value[value]&page=2).

World Economic Forum. 2016. "The Global Information Technology Report". World Economic Forum. http://www3.weforum.org/docs/GITR2016/WEF_GITR_Colombia_2016.pdf.

World Wide Web Foundation. s/f. "Country Detail | Open Data Barometer". Open Data Barometer. Consultado el 14 de diciembre de 2018. https://opendatabarometer.org/4thedition/detail-country/?_year=2016&indicator=ODB&detail=COL.

Zuckerman, Ethan, y Catherine D'Ignazio. 2017. "You Don't Need a Data Scientist, You Need a Data Culture". Data Therapy (blog). el 6 de diciembre de 2017. <https://datatherapy.org/2017/12/06/building-a-data-culture/>.

