

Documento CONPES

**CONSEJO NACIONAL DE POLÍTICA ECONÓMICA Y SOCIAL
REPÚBLICA DE COLOMBIA
DEPARTAMENTO NACIONAL DE PLANEACIÓN**



3920

POLÍTICA NACIONAL DE EXPLOTACIÓN DE DATOS (BIG DATA)

Departamento Nacional de Planeación
Ministerio de Tecnologías de la Información y las Comunicaciones
Superintendencia de Industria y Comercio

Versión aprobada

Bogotá, D.C., 17 de abril de 2018

CONSEJO NACIONAL DE POLÍTICA ECONÓMICA Y SOCIAL CONPES

Juan Manuel Santos Calderón

Presidente de la República

General (R) Óscar Adolfo Naranjo Trujillo

Vicepresidente de la República

Guillermo Abel Rivera Flórez

Ministro del Interior

María Ángela Holguín Cuéllar

Ministra de Relaciones Exteriores

Mauricio Cárdenas Santamaría

Ministro de Hacienda y Crédito Público

Enrique de Jesús Gil Botero

Ministro de Justicia y del Derecho

Luis Carlos Villegas Echeverri

Ministro de Defensa Nacional

Juan Guillermo Zuluaga Cardona

Ministro de Agricultura y Desarrollo Rural

Alejandro Gaviria Uribe

Ministro de Salud y Protección Social

Griselda Janeth Restrepo Gallego

Ministra de Trabajo

Germán Arce Zapata

Ministro de Minas y Energía

María Lorena Gutiérrez Botero

Ministra de Comercio, Industria y Turismo

Yaneth Giha Tovar

Ministra de Educación Nacional

Luis Gilberto Murillo Urrutia

Ministro de Ambiente y Desarrollo Sostenible

Camilo Armando Sánchez Ortega

Ministro de Vivienda, Ciudad y Territorio

David Luna Sánchez

Ministro de Tecnologías de la Información y las Comunicaciones

Germán Cardona Gutiérrez

Ministro de Transporte

Mariana Garcés Córdoba

Ministra de Cultura

Luis Fernando Mejía Alzate

Director General del Departamento Nacional de Planeación

Alejandra Corchuelo Marmolejo

Subdirectora General Sectorial

Santiago Matallana Méndez

Subdirector General Territorial

Resumen ejecutivo

En los últimos años las principales economías del mundo han avanzado en el reconocimiento de los datos como elemento central para la adaptación a la cuarta revolución industrial y para el apalancamiento del crecimiento económico. En Colombia, hasta la adopción del Plan Nacional de Desarrollo 2014-2018 *Todos por un nuevo país*, no se había reconocido de manera explícita la importancia de los datos como activo para generar valor social y económico. Sin embargo, desde hace aproximadamente veinte años se identificó la necesidad de emplear las tecnologías de la información y las comunicaciones para aumentar la eficiencia en el desarrollo de los procesos y la gestión gubernamental.

La presente política tiene por objetivo aumentar el aprovechamiento de datos, mediante el desarrollo de las condiciones para que sean gestionados como activos para generar valor social y económico. En lo que se refiere a las actividades de las entidades públicas, esta generación de valor es entendida como la provisión de bienes públicos para brindar respuestas efectivas y útiles frente a las necesidades sociales.

Para lograrlo, la política busca superar los retos que hasta el momento han impedido la disponibilidad masiva de datos digitales de las entidades públicas. Esto, como condición esencial para su aprovechamiento, que hace necesario acelerar y fortalecer la digitalización, la apertura de datos y la interoperabilidad entre las entidades públicas, elementos que activan los procesos requeridos para la generación de valor social y económico.

Para adecuar la intervención pública y orientarla a la generación de valor con los datos se requiere del estímulo a las capacidades que dinamicen la oferta, esto es, el capital humano, así como la demanda de bienes y servicios basados en datos, que se manifiesta en una cultura de datos. Adicionalmente, para materializar este potencial, se propone la definición de un marco jurídico e institucional que maximice la obtención de beneficios del aprovechamiento de datos, permita la cooperación entre los sectores público y privado y, al mismo tiempo, refuerce la protección de los derechos de los ciudadanos en el contexto de la creciente transformación de la vida diaria en datos digitales cuantificables y procesables (llamada *datificación*). Para ello, se resalta la necesidad de establecer distinciones claras entre las categorías de datos: personales, del sector privado y públicos. Así mismo, se precisan límites que orienten a los actores hacia fines que aumenten el bienestar social.

El Departamento Nacional de Planeación y el Ministerio de Tecnologías de la Información y las Comunicaciones liderarán la implementación de la Política nacional de explotación de datos (*Big Data*), en un periodo de implementación de cinco años, con inversiones por 16.728 millones de pesos para la creación de las condiciones descritas.

Clasificación: O33.

Palabras clave: explotación de datos, *Big Data*, innovación basada en datos, economía digital.

TABLA DE CONTENIDO

1. INTRODUCCIÓN	9
2. ANTECEDENTES Y JUSTIFICACIÓN.....	11
2.1. Antecedentes de política y normativa para la explotación de datos en Colombia	12
2.1.1. Gobierno Electrónico	14
2.1.2. Eficiencia administrativa	17
2.1.3. Gestión documental	18
2.1.4. Protección de datos personales	19
2.2. Políticas internacionales.....	21
2.3. Justificación.....	23
3. MARCO CONCEPTUAL	24
3.1. Conceptos principales: datos, información y conocimiento	25
3.2. Explotación de datos (<i>Big Data</i>)	27
3.2.1. Gobernanza	28
3.2.2. Capital humano.....	30
3.2.3. Cultura de datos	31
3.2.4. Marco jurídico	31
3.2.5. Límites éticos	32
3.3. Datos como recurso de infraestructura.....	32
4. DIAGNÓSTICO	34
4.1. Bajos niveles de datos públicos digitales	36
4.1.1. Baja digitalización.....	37
4.1.2. Bajos niveles de datos abiertos	40
4.1.3. Baja interoperabilidad	47
4.2. Desconfianza e incertidumbre respecto de la explotación de datos	52
4.2.1. Dispersión y desarticulación del marco jurídico.....	53
4.2.2. Nuevos contextos generados por la explotación de datos	58
4.3. Bajo capital humano para la explotación de datos.....	61
4.4. Ausencia de cultura de datos	64
5. DEFINICIÓN DE LA POLÍTICA	69

5.1.	Objetivo general	69
5.2.	Objetivos específicos	69
5.3.	Plan de acción	69
5.3.1.	Masificar la disponibilidad de datos públicos digitales accesibles, usables y de calidad.....	70
5.3.2.	Generar seguridad jurídica para la explotación de datos	76
5.3.3.	Disponer de capital humano para generar valor con los datos	82
5.3.4.	Generar cultura de datos en el país	84
5.4.	Seguimiento	90
5.5.	Financiamiento.....	91
6.	RECOMENDACIONES.....	93
	GLOSARIO	96
	ANEXOS.....	102
	Anexo A. Plan de Acción y Seguimiento (PAS)	102
	Anexo B. Normas aplicables a la explotación de datos.....	103
	Anexo C. Antecedentes indirectos de la política de explotación de datos	104
	Anexo D. Experiencias de aprovechamiento de datos	105
	Anexo E. Capital humano requerido para la explotación de datos	106
	BIBLIOGRAFÍA	107

ÍNDICE DE FIGURAS

Figura 1. Marco jurídico aplicable a los datos	12
Figura 2. Generación de valor con datos.....	28
Figura 3. Ciclo de vida de la explotación de datos	30
Figura 4. Distribución de competencias	57

ÍNDICE DE TABLAS

Tabla 1. Marco internacional de políticas de explotación de datos - comparativo de estrategias.....	21
Tabla 2. Organización y almacenamiento de los datos digitales	25
Tabla 3. Tipologías de datos	26
Tabla 4. Los datos como bienes públicos, de capital y propósito general	33
Tabla 5. Beneficios de escala, alcance y redes de los datos	33
Tabla 6. Retos del marco jurídico.....	53
Tabla 7. Tipologías de datos	54
Tabla 8. Indicadores de resultado de la política de explotación de datos	90
Tabla 9. Cronograma de seguimiento implementación de la política.....	91
Tabla 10. Financiamiento de la política por año.....	91

ÍNDICE DE GRÁFICOS

Gráfico 1. Generación de datos estructurados y no estructurados proyectados a 2025 .	35
Gráfico 2. Propósito de la digitalización.....	38
Gráfico 3. Porcentaje de los archivos de las entidades que están en formato digital	39
Gráfico 4. Puntajes ODB Latinoamérica	42
Gráfico 5. Porcentaje de entidades con plan de apertura de datos.....	44
Gráfico 6. Porcentaje de entidades que realizan inventario de datos	44
Gráfico 7. Porcentaje de conjuntos de datos abiertos estratégicos publicados al cierre de la vigencia 2016	46
Gráfico 8. Avance en la implementación de cadenas de trámites por las entidades	48
Gráfico 9. Barreras que limitan el uso de la PDI por las entidades	49
Gráfico 10. Entidades con interoperabilidad.....	50
Gráfico 11. Principales barreras para interoperar entre entidades.....	51
Gráfico 12. Índice de redundancia de campos de registros	52
Gráfico 13. Mecanismo de adopción de la política de gestión de datos	55
Gráfico 14. Oferta y demanda de servicios de explotación de datos y analítica	63
Gráfico 15. Necesidades que se solucionarían con la apertura de datos de las entidades	65
Gráfico 16. Percepción de importancia de los datos de otras entidades.....	65
Gráfico 17. Método más común de compartición y percepción de barreras para la explotación de datos	66
Gráfico 18. Barreras para intercambiar datos con el sector privado	67
Gráfico 19. Mayores barreras para la transformación digital (2016)	68
Gráfico 20. Uso de tecnologías en el sector industrial	69

SIGLAS Y ABREVIACIONES

CANE	Consejo Asesor Nacional de Estadística
CONPES	Consejo Nacional de Política Económica y Social
DANE	Departamento Administrativo Nacional de Estadística
DNP	Departamento Nacional de Planeación
FURAG	Formulario Único de Registro de Avance a la Gestión
GODI	<i>Global Open Data Index</i> (Índice Global de Datos Abiertos)
IoT	<i>Internet of Things</i> (Internet de las cosas)
MGA	Metodología General Ajustada
OCDE	Organización para la Cooperación y el Desarrollo Económicos
ODB	<i>Open Data Barometer</i> (Barómetro de Datos Abiertos)
ONU	Organización de las Naciones Unidas
OURData Index	<i>Open-Useful-Reusable Government Data Index</i> (Índice gubernamental sobre datos abiertos, útiles y reutilizables)
PAS	Plan de Acción y Seguimiento
PDI	Plataforma de interoperabilidad
PEN	Plan Estadístico Nacional
PND	Plan Nacional de Desarrollo
SEN	Sistema Estadístico Nacional
SIC	Superintendencia de Industria y Comercio
SINAE	Sistema Integral Nacional de Archivos Electrónicos
TI	Tecnologías de la información
TIC	Tecnologías de la información y las comunicaciones
TRD	Tabla de Retención Documental
TRV	Tabla de Valoración Documental

1. INTRODUCCIÓN

En la década de 1950, el desarrollo de los sistemas digitales, las telecomunicaciones y el aumento exponencial de la capacidad de poder de procesamiento, permitieron nuevas formas de generar, procesar y compartir información, dando lugar a la tercera revolución industrial¹ (World Economic Forum, 2016), también denominada la primera revolución digital.

En los años ochenta se desarrolló la primera base de datos digital, para almacenar los datos recolectados de manera manual bajo una estructura previamente definida. Esto agilizó su procesamiento y la generación de información. Este proceso recaía sobre los mínimos necesarios para análisis concretos, que ofrecían una visión parcial de los fenómenos.

En la década del 2000, la digitalización y el despliegue de servicios basados en Internet dieron lugar a la generación de datos masivos, de manera consciente (*i.e.* el uso de sistemas de información, bases de datos, entre otros) e inconsciente (*i.e.* los rastros digitales derivados del uso de redes sociales, la navegación en Internet, el uso del teléfono celular y similares). Esto, unido al aumento en las capacidades de almacenamiento y la sofisticación de las capacidades técnicas para aprovechar datos, permitió identificar que estos son el activo más importante para el avance económico², dando lugar a la cuarta revolución industrial. La cual, partiendo de los avances de la digitalización, permitirá la transformación de los procesos productivos y científicos, mediante la convergencia de los ámbitos físicos, biológicos y digitales (World Economic Forum, 2016).

En general, el aprovechamiento de los datos permite: (i) la generación de valor social y económico, mediante la producción de información y conocimiento para mejorar o crear procesos, productos y servicios; (ii) la solución de problemas objeto de la política pública, de cualquier proceso empresarial, académico o de iniciativa ciudadana; y (iii) facilita la innovación porque evidencia patrones ocultos y correlaciones que no podrían obtenerse de otra manera. Esta labor, que antes dependía en gran medida de la creatividad y el ingenio humano, al ser mediada por la analítica, acelera y mejora las actividades socioeconómicas.

Por lo anterior, en los últimos seis años, ocho de las principales economías del mundo³ avanzaron en una política pública nacional de aprovechamiento de datos, para adecuar

¹ La primera revolución industrial (1784) fue marcada por la invención de la máquina de vapor. La segunda revolución surge con la invención de la electricidad y la profundización de la producción en serie. La tercera revolución (finales de la década del 50) inicia con la electrónica, los primeros computadores y robots.

² Dando lugar a procesos, bienes y servicios como los vehículos autónomos, drones, asistentes virtuales, herramientas de traducción, entre otros.

³ Estos son: Estados Unidos, Unión Europea, Australia, Corea del Sur, Reino Unido, Francia, Japón y China.

las condiciones de cada país con el objetivo de maximizar los beneficios sociales y económicos y minimizar los riesgos que se asocian a la explotación de datos.

En Colombia, hasta el Plan Nacional de Desarrollo (PND) 2014-2018 *Todos por un nuevo país* no se había reconocido de manera explícita la importancia de los datos como activo que permite generar valor social y económico. No obstante, desde hace aproximadamente 20 años se identificó la necesidad de emplear las tecnologías de la información y las comunicaciones (TIC) para aumentar la eficiencia en los procesos y la gestión gubernamental.

En este sentido, el país ha avanzado en la provisión de condiciones normativas e institucionales para que las entidades públicas intercambien información y publiquen los datos de su gestión. Adicionalmente, existe un régimen robusto de protección de datos personales con controles fuertes para su garantía.

A pesar de estos avances, aún no se alcanzan altos niveles de datos digitales de las entidades públicas, condición que resulta básica para el aprovechamiento de datos. Tampoco se han adelantado acciones para aumentar su explotación y generar valor social y económico, porque no se habían reconocido como el activo central de la economía digital.

Por lo expuesto, es necesario implementar una política pública de explotación de datos en Colombia que enfrente las fallas de articulación de intervenciones previas y la desalineación de los incentivos que determinan las acciones intra e inter-entidades. La política presentada en este documento busca superar la carencia de una visión de largo plazo y la tendencia al cumplimiento mínimo de mandatos legales, con los mecanismos y herramientas que procuren las transformaciones institucionales para la generación de valor, entendido como la provisión de bienes públicos para brindar respuestas efectivas y útiles, frente a las necesidades sociales⁴ (Banco Interamericano de Desarrollo, Centro Latinoamericano de Administración para el Desarrollo, 2007).

Para esto, se propone adecuar la intervención pública mediante la provisión de un marco jurídico⁵, ético e institucional que provea las condiciones habilitantes para aprovechar los beneficios de la explotación de datos⁶ y mitigar los riesgos asociados a

⁴ En esta medida, la explotación de datos se supedita al cumplimiento de los fines del Estado.

⁵ Bajo el reconocimiento de la protección de datos personales, sometida a los principios definidos en la Ley 1581 de 2012, cuyo tratamiento es una actividad que debe sujetarse a las normas, obedecer a finalidades legítimas y respetar la autodeterminación informativa de los titulares.

⁶ Cuando el presente documento menciona la explotación de datos o el aprovechamiento de datos para la generación de valor social y económico, debe entenderse que excluye aquellos cuyo tratamiento se encuentra proscrito. Concretamente se refiere a aquellos datos que cumplen las condiciones jurídicas para ser empleados

esta⁷. Como elementos habilitantes para la materialización de los beneficios señalados, se incorpora el estímulo a las capacidades que dinamicen la oferta y demanda de bienes y servicios basados en este activo. Lo anterior será llevado a cabo a través de cuatro estrategias propuestas en la presente política: (i) masificación de la disponibilidad de datos públicos digitales accesibles, usables y de calidad; (ii) generación de seguridad jurídica para la explotación de datos; (iii) mejorar la disponibilidad de capital humano para generar valor con los datos; y (iv) generación de cultura de datos en el país.

Este documento CONPES tiene seis secciones, incluyendo esta introducción. La segunda sección contiene el marco jurídico aplicable, los antecedentes, referentes internacionales y la justificación de la presente política. La tercera sección formula el marco conceptual con los conceptos básicos de la generación de valor con los datos. La cuarta sección describe el diagnóstico de las cuatro situaciones principales a superar para aumentar la explotación de datos. La quinta sección define los objetivos de la política, las acciones para alcanzarlos y los responsables de cada una de ellas. Adicionalmente, incluye el esquema de seguimiento de la política y su financiamiento. Finalmente, la sexta sección presenta las recomendaciones al Consejo Nacional de Política Económica y Social (CONPES).

2. ANTECEDENTES Y JUSTIFICACIÓN

Durante los años 90 y principios del 2000, los esfuerzos gubernamentales relacionados con el uso de las TIC se orientaron a maximizar la eficiencia en la gestión de las entidades públicas y del sector privado. En Colombia, las intervenciones públicas en mención se realizaron en contextos históricos disímiles, cuando no era evidente el valor social y económico de los datos. De acuerdo con las materias reguladas, estas intervenciones pueden agruparse en Gobierno Electrónico, eficiencia administrativa, gestión documental y protección de derechos. Adicionalmente, existen múltiples mandatos respecto del reporte de datos por entidades públicas y privadas⁸; normas que se agrupan con la denominación de reportes de información e impactan la generación y compartición de datos, a pesar de no ser un antecedente directo orientado a su aprovechamiento. La presente sección describe estos antecedentes y el marco jurídico para la explotación de

en dicho proceso (como datos impersonales, anonimizados, agregados, que impiden la individualización de personas, aquellos que la ley define como públicos, que no afecten la propiedad intelectual, el secreto comercial, así como los que hayan sido autorizados por el titular de manera expresa, libre, espontánea, sin vulnerar el principio de finalidad, entre otros).

⁷ En todo lo relacionado con la seguridad digital el presente documento se sujeta a lo dispuesto en el Documento CONPES 3854 *Política nacional de seguridad digital*, así como las acciones previstas en el mismo.

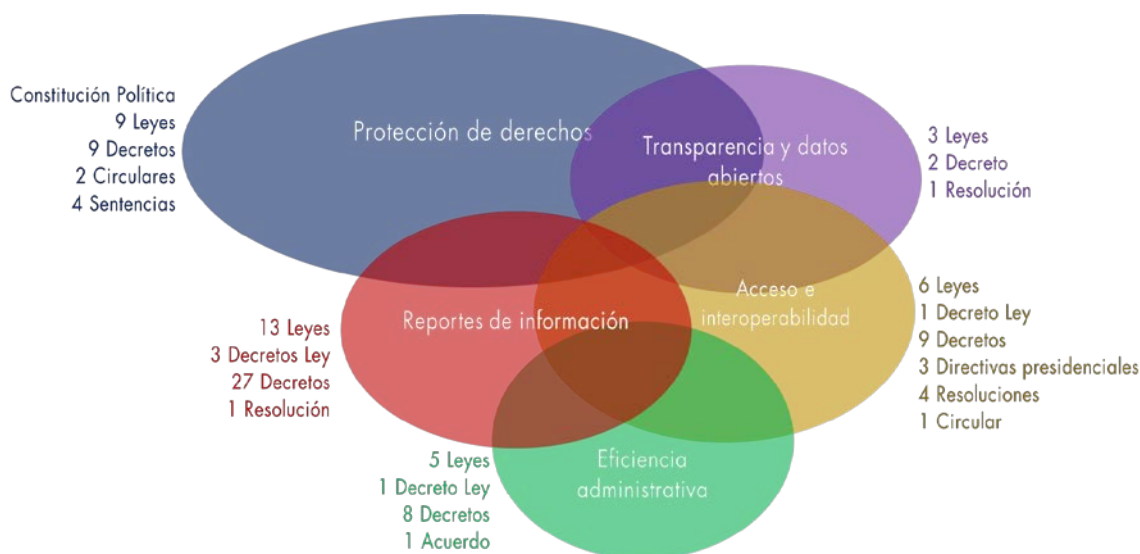
⁸ Que surgen con ocasión del desarrollo de la función de control y vigilancia de las actividades realizadas sobre particulares objeto de regulación, la prestación de servicios públicos y el cumplimiento de funciones públicas.

datos. Adicionalmente, refiere las políticas internacionales adoptadas en la materia y, finalmente, presenta la justificación de la política de explotación de datos.

2.1. Antecedentes de política y normativa para la explotación de datos en Colombia

La realización de los procesos de generación, recolección, agregación, compartición, explotación e innovación para aprovechar datos se encuentran mediados por diversas disposiciones jurídicas que conforman el marco relevante vigente para la explotación de datos, descrito en detalle en el Anexo B. Estas se agrupan en cinco regímenes de acuerdo con las materias reguladas⁹, según se ilustra en la Figura 1.

Figura 1. Marco jurídico aplicable a los datos



Fuente: DNP (2018).

Las disposiciones jurídicas con impacto en la explotación de datos señaladas en la Figura 1 se agrupan en cinco regímenes, según el tema regulado por cada una, así: (i) protección de derechos, esto es, límites y garantías que deben desplegarse para el tratamiento de datos personales y privados; (ii) transparencia y datos abiertos, que corresponde a la publicidad de las actuaciones de las entidades públicas, el derecho de acceso a la información pública y la necesidad de divulgación activa; (iii) acceso e

⁹ Se referencian normas únicas que usualmente hacen remisión a las “demás que sean aplicables”, sin generar una armonización real de la materia. Por ejemplo, en el Decreto 2693 de 2012, derogado por el Decreto 2573 de 2014, se indicaba que, para la incorporación de medios electrónicos, informáticos, telemáticos, ópticos o similares en la gestión de sus documentos, los sujetos obligados deberán seguir lo establecido en los artículos 19 y 21 de la Ley 594 de 2000, 12 de la Ley 527 de 1999, así como los elementos establecidos en la Ley 1437 de 2011 y las demás disposiciones legales vigentes aplicables.

interoperabilidad, que define condiciones mínimas para que las entidades públicas y privadas que desarrollan funciones públicas establezcan mecanismos para permitir el uso de datos entre sí; (iv) eficiencia administrativa, es decir, la gestión de los documentos generados por las entidades públicas y las privadas que prestan funciones públicas, así como aumento de la generación de documentos digitales para eliminar los soportes en papel y disminuir los trámites solicitados al ciudadano; y (v) reportes de información, que corresponde a los reportes de datos que las entidades privadas, la academia y los ciudadanos deben suministrar periódicamente a las entidades públicas en el marco de sus competencias, a los particulares que desarrollan funciones públicas, así como aquellos que las entidades públicas deben realizar.

Este marco jurídico se compone de disposiciones y precedentes judiciales de diferente jerarquía¹⁰, expedidas de manera independiente¹¹ para regular situaciones sociales concretas en momentos y sectores específicos. Por ello, carece de armonización para habilitar los procesos de explotación de datos para la generación de valor económico y social, y no contiene reglas para la mitigación de los riesgos asociados a ello. El ejercicio de compilación realizado por el Gobierno nacional entre 2015 y 2016 reunió en cuerpos normativos únicos los decretos propios de cada sector gubernamental. Esta compilación es un esfuerzo importante, que brinda seguridad jurídica, pero no corresponde a una labor de articulación, armonización o simplificación de la regulación que favorezca la explotación de datos.

Por otro lado, mediante el artículo 160 del PND 2014-2018 *Todos por un nuevo país* se creó el Sistema Estadístico Nacional (SEN)¹² para generar estadísticas oficiales a partir del aprovechamiento de registros administrativos y del intercambio de microdatos entre las entidades del SEN con altos estándares de calidad y de manera oportuna. De igual forma se definió al DANE como rector del SEN. Este Sistema es un conjunto articulado de entidades, organizaciones, usuarios, procesos e instrumentos técnicos, que garantiza la producción y difusión de las estadísticas oficiales nacionales y territoriales. Para ello, fue expedido el Plan Estadístico Nacional (PEN) que contiene la política estadística del país para los próximos cinco años.

¹⁰ Se recogen todos los mecanismos jurídicos que contengan algún mandato referido a captura (reportes), almacenamiento, procesamiento, análisis, publicación y creación de valor de datos o de información, en el sector público y privado.

¹¹ Se incluyen sentencias de control previo de constitucionalidad de leyes estatutarias, y la sentencia de tutela T-729/02, primera en referirse a la materia de protección de datos y los principios para su tratamiento, que posteriormente se incorporaron a las Leyes Estatutarias 1266 de 2008, 1581 de 2012, 1621 de 2013, 1712 de 2014 y 1755 de 2015.

¹² De acuerdo con el Decreto 1743 de 2016, el SEN es definido como “el conjunto articulado de componentes que, de manera organizada y sistemática, garantiza la difusión de las estadísticas oficiales que requiere el país”. El SEN está compuesto por las entidades que producen y difunden estadísticas o son responsables de registros administrativos.

Los antecedentes de política pública y el marco normativo relacionados con las ganancias en la eficiencia de la gestión pública y las mejoras en la prestación de servicios al ciudadano, a través del uso de TIC, son antecedentes indirectos del aprovechamiento de datos porque constituyen el marco inicial de su generación y compartición. Así mismo, Colombia ha avanzado de manera muy significativa en la creación de condiciones institucionales para la garantía de los derechos relacionados con los datos, entendidos en un contexto previo a la disponibilidad masiva de los mismos y de las capacidades para su uso como activo. Estas condiciones son, por un lado, las intervenciones relacionadas con la transparencia y acceso a la información pública que incluyen la posibilidad de aprovechar esta información por cualquier actor y, por otro lado, las disposiciones para la protección de datos personales que definen las reglas para el acceso, intercambio y uso de estos, por parte de personas públicas y privadas.

De acuerdo con la materia regulada, las intervenciones que se califican como antecedentes indirectos, en tanto no hacían referencia específica a una política de aprovechamiento de datos, se sintetizan así: Gobierno Electrónico, eficiencia administrativa, gestión documental y protección de derechos. La relación de documentos CONPES y normas que establecen lineamientos en estas materias se describen en detalle en el Anexo C. Es necesario resaltar que el PND 2014-2018 es el único antecedente directo que determina expresamente la necesidad de disponer de una política pública de explotación de datos. Así, el literal j) del artículo 45 del PND y el documento Bases del Plan Nacional de Desarrollo 2014-2018 *Todos por un nuevo país*, el cual constituye parte integral del mismo, de conformidad con lo dispuesto en el artículo 2 de la Ley 1753 de 2015¹³, disponen que el DNP liderará la formulación de la estrategia que permita el aprovechamiento de datos para el desarrollo de nuevos conocimientos, la creación de valor, así como la disponibilidad de información más amplia y de mayor calidad para la toma de decisiones públicas.

2.1.1. Gobierno Electrónico

Desde 1997 se reconoce en Colombia que las TIC¹⁴ aumentan la eficiencia de la administración pública y facilitan la provisión de servicios gubernamentales¹⁵. A mediados de la década del 2000, el cambio de paradigma de servicios centrados en el Gobierno a servicios centrados en el ciudadano da lugar al concepto de Gobierno Electrónico como estrategia que emplea las TIC para generar beneficios sociales (OCDE, 2009). Las

¹³ Por la cual se expide el Plan Nacional de Desarrollo 2014-2018 *Todos por un nuevo país*.

¹⁴ Inicialmente referidas como Telecomunicaciones.

¹⁵ PND 1998-2002 *Cambio para Construir la Paz*.

temáticas específicas de la estrategia de Gobierno Electrónico¹⁶ correspondientes a digitalización, datos abiertos, interoperabilidad e intercambio de datos, son el punto de partida para la explotación de datos, porque producen las condiciones iniciales de generación, recolección, agregación y compartición de datos digitales entre las entidades del sector público, de estas hacia el sector privado, los ciudadanos y la academia, y viceversa.

En 2003, mediante el Documento CONPES 3248 *Renovación de la Administración Pública*, se establecieron los lineamientos para adecuar la administración pública y fortalecer su capacidad de gestión y desempeño en el cumplimiento de sus funciones. Dentro de los componentes estructurales¹⁷ para la transformación pública requeridos para el adecuado funcionamiento institucional, el citado documento incluía el Gobierno Electrónico y los sistemas de información. Al respecto, se identificó la baja calidad y desactualización de la información de base y transaccional, así como la incompatibilidad tecnológica de los sistemas existentes, la ausencia de estándares que faciliten el intercambio y uso de la información, así como los retos en su usabilidad y calidad. Como resultado de lo anterior, se dispuso la creación del Sistema Nacional de Información Oficial y se generaron las bases para la política de Gobierno Electrónico para aumentar la eficiencia en la relación del ciudadano con las entidades públicas. Esta política fue establecida en la Ley 790 de 2000¹⁸ y adoptada de manera vinculante para su implementación en las entidades públicas mediante el Decreto 1151 de 2008¹⁹.

En 2010, mediante el Documento CONPES 3650 *Importancia estratégica de la Estrategia de Gobierno en Línea*, se resaltó la necesidad de que las entidades públicas asumieran la responsabilidad de cumplir los lineamientos previstos en el Decreto 1151 de 2008. Así mismo, se introdujo la definición de Gobierno Electrónico, entendido como el uso de las TIC para mejorar los servicios e información ofrecidos a los ciudadanos, aumentar la eficiencia y eficacia de la gestión pública e incrementar sustantivamente la transparencia del sector público y la participación ciudadana. En concordancia con lo anterior, mediante

¹⁶ La estrategia en Colombia es llamada Gobierno en Línea e implica temas más amplios y estructurados orientados a un Estado transparente, colaborativo y eficiente, que presta mejores servicios al ciudadano.

¹⁷ De acuerdo con el documento, son procesos que “inciden significativamente en la eficacia y la eficiencia de la administración pública” y corresponden a: empleo público, gestión de activos, estrategia jurídica de la Nación, sistema presupuestal, gestión por resultados, gobierno electrónico, sistemas de información, sistema de contratación pública, estrategia antitrámites, control interno y regulación y control.

¹⁸ Por la cual se expiden disposiciones para adelantar el programa de renovación de la administración pública y se otorgan unas facultades Extraordinarias al Presidente de la República.

¹⁹ Por el cual se establecen los lineamientos generales de la estrategia de Gobierno en Línea de la República de Colombia, se reglamenta parcialmente la Ley 962 de 2005, y se dictan otras disposiciones.

el Decreto 2693 de 2012²⁰ se actualizaron los componentes para la masificación del Gobierno en Línea dentro de los que se encuentran interoperabilidad, cadenas de trámites y ventanillas únicas virtuales, datos y Gobierno Abierto. Este Decreto fue derogado por el Decreto 2573 de 2014²¹, que amplió los plazos de implementación y actualizó las fases a alcanzar por las entidades públicas. El Manual de Gobierno en Línea y el Marco de Referencia de Arquitectura de tecnologías de la información (TI), son los instrumentos principales para la implementación de esta política.

De acuerdo con lo expuesto anteriormente y con lo presentado en detalle en el Anexo C, los esfuerzos de política pública y del marco normativo orientados al uso de las TIC en el funcionamiento del Gobierno colombiano buscan mejorar la prestación de servicios al ciudadano y el aumento de la eficiencia, coherentes con el enfoque internacional de Gobierno Electrónico. Se reconoce que las TIC tienen la capacidad de transformar la gestión pública y aportar al crecimiento económico. La incorporación en instrumentos jurídicos de declaraciones respecto de la necesidad de avanzar en la masificación del Gobierno Electrónico, así como la disposición de condiciones institucionales para la apertura de datos, la transparencia de la gestión pública y el acceso a la información, son un avance importante que refleja el compromiso del país respecto de la adopción de esta materia, situación que es objeto de reconocimiento internacional.

No obstante, en el momento histórico de definición de los instrumentos referidos no se había reconocido el valor social y económico que se obtiene con la generación de bienes, servicios y procesos mediante la explotación de datos²². Así, en el contexto actual surgen desafíos concretos para disponer de las condiciones iniciales para su aprovechamiento. Estos retos se sintetizan en la necesidad de disponer de una hoja de ruta clara que permita la materialización de los resultados de las acciones planteadas para la implementación de la política, dadas por componente y fases; las desarticulaciones intra e inter entidades, dadas por la ausencia de mecanismos que orienten la implementación de la estrategia de Gobierno Electrónico como una herramienta para el mejoramiento de la gestión institucional y la generación de valor; así como la ausencia de elementos mínimos de cumplimiento perentorio y exigible que fortalezcan las acciones implementadas hasta la fecha.

²⁰ Por el cual se establecen los lineamientos generales de la Estrategia de Gobierno en Línea de la República de Colombia, se reglamentan parcialmente las Leyes 1341 de 2009, 1450 de 2011, y se dictan otras disposiciones.

²¹ Por el cual se establecen los lineamientos generales de la Estrategia de Gobierno en Línea, se reglamenta parcialmente la Ley 1341 de 2009 y se dictan otras disposiciones.

²² La Ley 1714 de 2012 reconoce que el acceso a la información es un derecho fundamental que habilita el ejercicio de otros derechos. Esto ha servido para fortalecer la iniciativa de datos abiertos, orientada a aumentar su publicación y uso.

2.1.2. Eficiencia administrativa

Las normas que se ocupan específicamente de la disposición de cadenas de trámites, que favorecen la disminución de trámites exigidos al ciudadano, así como el intercambio de información entre entidades públicas son antecedentes relevantes para la explotación de datos porque complementan los esfuerzos de implementación del Gobierno Electrónico, orientados a contar con las condiciones iniciales de disponibilidad de datos e interoperabilidad, que permiten el aprovechamiento de estos.

Hasta la fecha se han proferido normas que definen expresamente la obligación de compartir datos en la gestión de las entidades públicas y disponer el intercambio sin necesidad de convenios. Mediante la Ley 962 de 2005²³ se autorizó la implementación de sistemas para integrar y compartir información. La Ley 1450 de 2011²⁴ estableció el deber de compartir entre entidades públicas las bases de datos requeridas para el cumplimiento de sus funciones. Los Decretos 235²⁵ y 2280²⁶ del 2010, definieron que las entidades pueden emplear el mecanismo que consideren idóneo para dicho intercambio. Con el fin de hacer efectivo el deber de compartir datos entre entidades públicas, el Decreto 19 de 2012²⁷ prohibió a las entidades solicitar al ciudadano información que ya se encuentra en otra entidad pública. Este es un avance importante porque brinda las condiciones jurídicas mínimas para habilitar el uso de información entre entidades públicas, para aumentar la eficiencia en la gestión, así como disminuir los trámites realizados por el ciudadano.

No obstante, estas normas no están acompañadas de un mecanismo que brinde claridad y alinee los incentivos de las entidades para su implementación. Por lo anterior, el PND 2014-2018, define la necesidad de expedir estándares, modelos, lineamientos y normas técnicas para la incorporación de las TIC, en la prestación de servicios²⁸. Así mismo, reitera el mandato de compartir información entre entidades públicas, aclarando que el mismo será gratuito, deberá solicitarse y realizarse respaldado en estándares que faciliten el proceso de intercambio y no en tecnologías específicas que impidan el acceso, y

²³ Por la cual se dictan disposiciones sobre racionalización de trámites y procedimientos administrativos de los organismos y entidades del Estado y de los particulares que ejercen funciones públicas o prestan servicios públicos.

²⁴ Por la cual se expide el Plan Nacional de Desarrollo, 2010-2014

²⁵ Por el cual se regula el intercambio de información entre entidades para el cumplimiento de funciones públicas.

²⁶ Por el cual se modifica el artículo 3 del Decreto 235 de 2010.

²⁷ Por el cual se dictan normas para suprimir o reformar regulaciones, procedimientos y trámites innecesarios existentes en la Administración Pública.

²⁸ Ley 1753 de 2015, artículo 45.

propende por la integración de los sistemas de información para el ejercicio eficiente y adecuado de la función pública²⁹.

La ausencia de herramientas claras para implementar los mandatos descritos propicia una tendencia de cumplimientos mínimos que obstaculiza las transformaciones organizacionales. Tampoco se ha determinado un mecanismo estandarizado para el acceso a los datos entre entidades públicas que brinde la seguridad jurídica que estas buscan al suscribir convenios. Adicionalmente, no se ha previsto la cuantificación del impacto de los aumentos en eficiencia, lo que afecta la disposición de invertir recursos en los cambios necesarios³⁰.

2.1.3. Gestión documental

Las disposiciones relacionadas con la gestión documental y la función archivística son un elemento transversal de toda la gestión pública aplicable a todas las actuaciones que impliquen un soporte documental. De acuerdo con la Ley 594 del 2000³¹, su objeto principal es preservar el registro histórico de la gestión en un conjunto de documentos conservados por su valor histórico o cultural. El documento es la unidad del sistema y se plantea el uso de las TIC como herramienta para aportar al objeto señalado.

Las normas que rigen la gestión documental de las entidades públicas reflejan un avance importante con la inclusión de la interoperabilidad como principio de esta, que debe ser adoptada en los procesos de archivo, de acuerdo con lo dispuesto en el Decreto 2609 de 2012, que reglamenta la Ley General de Archivos (594 de 2000), en la que se definió que las entidades podrían implementar nuevas tecnologías y soportes para la gestión documental. Así mismo, se ha sentado un precedente respecto a la importancia de transformar procesos organizacionales para adecuarse al Gobierno Electrónico. Al respecto, mediante el Decreto 2578 de 2012 se reglamentó el Sistema Integral Nacional de Archivos Electrónicos (SINAE) como un programa especial de coordinación de la política nacional de archivos electrónicos del estado colombiano, en el que se le encomienda al Archivo General de la Nación la tarea de emitir conceptos sobre el impacto de las TIC en la gestión documental para garantizar homogeneidad en la gestión archivística del Estado y la preservación de los documentos electrónicos que conforman el patrimonio documental digital.

²⁹ Ley 1753 de 2015 artículo 159.

³⁰ La OCDE (2014), encontró que en Colombia la evaluación de la implementación de las normas relacionadas con la eficiencia administrativa se ha centrado en la medición de los trámites simplificados por entidad y existe poca información sobre el impacto en la reducción de las cargas administrativas.

³¹ Por medio de la cual se dicta la Ley General de Archivos y se dictan otras disposiciones.

No obstante, estos lineamientos no han sido provistos de herramientas de articulación con los fines de aumento en la disponibilidad de datos, publicidad de la información, o la simplificación del intercambio de datos. Adicionalmente, el régimen de gestión documental ha sido menos ágil en la adopción de los lineamientos relacionados con la incorporación de las TIC en la gestión pública. Esto, porque las reglas para la gestión de documentos electrónicos y la conformación de expedientes fueron definidas recientemente por el Archivo General de la Nación mediante el Acuerdo 003 de 2015³². Sin embargo, aún no se han establecido elementos tendientes a la utilización y reutilización de los datos contenidos en los documentos públicos para generar valor en la gestión pública, sin perjuicio de la conservación de aquellos.

Al respecto, el documento Bases del Plan Nacional de Desarrollo 2014-2018 *Todos por un nuevo país* reconoce la necesidad de modernizar los archivos públicos para “avanzar en la simplificación de trámites y procesos, así como en la disminución del volumen de documentos, la reducción de costos y el incremento de la productividad de los servidores públicos”, mediante el proyecto Archivo Nacional Digital a cargo del Archivo General de la Nación. Por lo anterior, se requiere articular la transformación de los archivos públicos con la necesidad de aumentar la disponibilidad de datos digitales aprovechables, en condiciones que faciliten su explotación.

2.1.4. Protección de datos personales

Una de las principales fuentes de innovación en los últimos años ha sido la explotación de los datos personales, a partir de los cuales se han desarrollado nuevos productos como las redes sociales, las aplicaciones de salud y productividad personal, entre otras³³ (OCDE, 2013). Estos datos contienen información asociada a una persona, que permite su identificación e involucra los derechos de *habeas data*, privacidad e intimidad. Al mismo tiempo, este bien jurídico otorga al titular plenas facultades de acceso, inclusión, exclusión, corrección, adición, actualización y certificación, así como la limitación en las posibilidades de divulgación, publicación o cesión. En Colombia se han realizado importantes avances en materia de protección de estos derechos, desde el punto de vista de su tratamiento individual, en un contexto previo a su disponibilidad masiva y al aumento exponencial de las capacidades para procesarlos. Las normas relevantes se describen en el Anexo C.

Desde 1960, el Decreto 1633 reconoce en el país la reserva de los datos suministrados con fines estadísticos para proteger la intimidad, el secreto comercial y la información estratégica de las personas jurídicas (Consejo de Estado, Alcance de la reserva

³² Hasta la expedición del Decreto 2609 de 2012, que reglamenta la Ley 594 de 2000, no se habían considerado expresamente las reglas para la gestión de los documentos electrónicos en las entidades públicas.

³³ Así mismo, se evidencia la aparición de un modelo de negocio basado en la venta de datos, en la generación y segmentación de publicidad.

estadística, 1999). En 2008, debido al aumento de las actividades financieras y comerciales que implican el uso de datos personales, se expidió la Ley 1266 para gobernar la recolección de información que determine el comportamiento crediticio y financiero y los derechos de los titulares asociados a ello.

En 2012, ante la necesidad de un marco transversal para el *habeas data*, se dispuso el régimen general de protección de datos personales, en la Ley 1581. Se establecieron garantías robustas, y procedimientos especiales para su efectividad, para que el titular defina la recolección, el uso, la circulación, divulgación, actualización, rectificación y supresión de sus datos. Igualmente, se dispuso una autoridad de protección de datos personales con facultades sancionatorias amplias (la Delegatura de Protección de Datos Personales de la Superintendencia de Industria y Comercio). Esta norma se reglamentó a través del Decreto 1377 de 2013 que incorpora el principio de responsabilidad demostrada, cuya guía de implementación fue elaborada por la Superintendencia de Industria y Comercio. Este decreto también crea el Registro Único de Bases de Datos, administrado por la citada entidad, que consiste en un directorio público de bases de datos personales para identificar su finalidad, los canales de atención a los titulares, las políticas de privacidad adoptadas, el tipo de datos que contienen, las medidas implementadas para garantizar su seguridad, entre otros.

No obstante, el régimen de protección de datos descrito fue expedido siguiendo las mejores prácticas internacionales dispuestas en el Convenio 108 de 1981 del Consejo de Europa, la Directiva Europea 95/46 de 1995, la Resolución 45/95 de 1990 de la Organización de las Naciones Unidas (ONU) y la Resolución de Madrid de 2009³⁴. En esta medida, las normas contienen lineamientos previos al reconocimiento de los datos como un activo y no contemplan su explotación masiva, su condición de bien de capital y la posibilidad o límites a la realización de transacciones de mercado con los mismos. Por lo anterior, los aspectos éticos y jurídicos que se plantean con la explotación de datos, así como los riesgos potenciales que emergen en este contexto, requieren del diseño de un marco jurídico para maximizar los beneficios y mitigar los riesgos (Data & Society Research Institute, 2014).

La generación de valor con la explotación de datos, y la eventual materialización de los riesgos derivados de esta actividad, ocurren con la combinación de diversas fuentes y no únicamente con el uso del dato en sí mismo. Por ello, cuando el titular expresa su consentimiento para la recolección y uso, no le es posible determinar eventuales consecuencias de la agregación y análisis, prever otros datos que puedan inferirse cuando se aplican técnicas de analítica, o los efectos de su combinación con algunos que haya

³⁴ Gaceta del Congreso 488 del 4 de agosto de 2010, exposición de motivos proyecto de Ley estatutaria 046 de 2010 Cámara.

publicado previamente. Lo anterior plantea el reto de enfocar los esfuerzos de la intervención estatal en los usos permitidos y prohibidos de los datos, independientemente de que el titular haya consentido a su procesamiento (Mayer-Schönberger & Padova, 2016), de modo que se delimiten claramente las responsabilidades que permitan la protección del bien jurídico.

2.2. Políticas internacionales

Similar a lo que ocurre en el país, varios países cuentan con políticas de Gobierno digital y apertura de datos. No obstante, durante los últimos seis años, ocho de las principales economías del mundo comprobaron la urgencia de contar con una política pública de explotación de datos para el aprovechamiento de estos como un activo que tiene el potencial de impulsar el crecimiento económico, la productividad pública y privada e incidir en la preparación para la economía digital. Adicionalmente, para corregir las fallas de mercado y de gobierno que pueden obstaculizar el proceso de explotación de datos, anticipar los riesgos potenciales y garantizar la protección de los derechos. Esto, como condición para adaptarse a la economía digital, donde es fundamental que las empresas hagan parte de las redes digitales, porque la mayoría de los bienes de mercado (ideas, capital intelectual, acceso) pueden digitalizarse, permitiendo que se proliferen con facilidad, a un costo de escalabilidad cercano a cero, porque cada nodo adicional de una red digital incrementa el valor para los demás, de modo que la red genera su propio crecimiento (World Economic Forum, 2016).

En la Tabla 1 se presenta una síntesis del objetivo de cada una de las políticas adoptadas en el contexto internacional. Como factor común se encuentra la importancia de que el Gobierno lidere las acciones para hacer efectivos los beneficios sociales y económicos del aprovechamiento de datos, reconociéndolos como un activo y articulando esfuerzos del sector público, privado, la academia y los ciudadanos.

Tabla 1. Marco internacional de políticas de explotación de datos - comparativo de estrategias

Referente	Año y documento para adopción de la estrategia	Objetivos de la política pública
Estados Unidos	2012: Iniciativa de investigación y desarrollo en <i>Big Data</i>	• Promover habilidades para el aprovechamiento de datos en las agencias gubernamentales.
	2016: Plan Federal Estratégico de Investigación y Desarrollo en <i>Big Data</i>	• Explotar datos para acelerar descubrimientos científicos e innovación y promover el crecimiento económico.
Unión Europea	2013: Consejo Europeo 24/25 Conclusiones	• Promover la economía digital, la innovación y los servicios como motores del crecimiento y el empleo.
	2014: Comunicación de la Comisión al Parlamento	• Crear un mercado de datos único.

Referente	Año y documento para adopción de la estrategia	Objetivos de la política pública
	Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: Hacia una economía de los datos próspera	
Australia	2012: Estrategia australiana de TIC del Servicio Público 2012-2015 2013: La estrategia de <i>Big Data</i> para el servicio público de Australia: mejorar la comprensión mediante un aumento en el análisis de datos	<ul style="list-style-type: none"> • Aumentar la productividad nacional y del sector público, mediante la toma de mejores decisiones de política pública, que se basen en datos y en mejor información. • Los datos son reconocidos como un activo nacional para el Gobierno y los ciudadanos.
Corea del Sur	2013: Informatización nacional para alcanzar la sociedad del conocimiento y la información 2013: Gov 3.0 2013: Plan maestro de <i>Big Data</i>	<ul style="list-style-type: none"> • Liderar la economía creativa a través de la informatización, con un Gobierno proactivo, científico y capaz en el que convergen datos públicos y privados.
Reino Unido	2013: Aprovechando la oportunidad de los datos. Una estrategia para la capacidad del Reino Unido 2017: Estrategia de transformación del Gobierno: mejor uso de los datos	<ul style="list-style-type: none"> • Extraer conocimiento y valor de los datos para el beneficio de los ciudadanos, el sector privado, la academia y el sector público. • Desarrollar la capacidad de datos, es decir, la habilidad de consumidores, negocios y academia de acceder y compartir datos apropiadamente. Se compone de capital humano, infraestructura y datos. • Define las prioridades de Gobierno hasta 2020, centrado en la remoción de barreras para el uso de datos públicos, aumentar las capacidades en ciencia de datos al interior del Gobierno, definir herramientas para la seguridad en el uso de datos y aumentar la apertura.
Francia	2014: Industria del futuro	<ul style="list-style-type: none"> • Propone nueve soluciones industriales del país para modernizar la base de producción: nuevos recursos, ciudades inteligentes, eco-movilidad, transporte, medicina, economía de datos, dispositivos inteligentes, confianza digital, y elección inteligente de alimentos. • Mejor manejo y apropiación de los datos por las compañías y el sector público. • Crear un ecosistema de economía de datos.
Japón	2013: Declaración para ser la nación más avanzada en tecnologías de información (TI)	<ul style="list-style-type: none"> • Crear una sociedad que promueve la innovación y generación de nuevos servicios, industrias y crecimiento, mediante el uso de datos abiertos y <i>Big Data</i>.

Referente	Año y documento para adopción de la estrategia	Objetivos de la política pública
China	2015: Plataforma para la acción en la promoción del desarrollo de <i>Big Data</i>	<ul style="list-style-type: none"> • Unificar los datos gubernamentales para facilitar servicio al ciudadano. • Aumentar los datos abiertos para promover innovación industrial, nuevos negocios y reestructuración económica.

Fuente: DNP con base en cada uno de los documentos mencionados.

De acuerdo con el contexto nacional, cada estrategia presenta prioridades específicas para su implementación y desarrollo, las cuales no son excluyentes, pero determinan el enfoque de las acciones a seguir, según las características propias de los países. Las prioridades se agrupan en: aumento de innovación y capacidades científicas (Estados Unidos), mejorar la eficiencia administrativa (Australia), fortalecer las capacidades industriales (Francia, Corea, Japón, China) y promover el crecimiento económico en el contexto de la economía digital (Reino Unido, Unión Europea). Adicionalmente, las políticas descritas resaltan la necesidad de anticiparse a los riesgos potenciales que pueden surgir de usos indebidos del aprovechamiento de datos, principalmente: intromisiones en la privacidad y la intimidad, existencia de monopolios, prácticas discriminatorias y limitaciones a la libertad de elección y de acceso a la información (Greenwood, Stopczynski, Sweatt, Hardjono, & Pentland, 2014).

En línea con lo expuesto, la presente política busca disponer las condiciones para aumentar el aprovechamiento de datos como insumo central para el mejoramiento y creación de procesos, bienes y servicios en Colombia, que parte del impulso por parte del sector público de las transformaciones requeridas para masificar la explotación de datos.

2.3. Justificación

El proceso de *datificación*, es decir, transformar el mundo en datos procesables y cuantificables, da lugar a la generación de nuevos bienes, servicios y procesos, así como a la mejora de los existentes, aportando a la diversificación y sofisticación de la economía, que redundan en el crecimiento de la misma (OCDE, 2015).

Las intervenciones previas en materia de Gobierno Electrónico, eficiencia administrativa y gestión documental dan lugar a las condiciones iniciales para la explotación de datos. Estas políticas presentan retos que surgen de la desalineación de los incentivos que orientan las acciones al interior de las entidades y entre estas, la carencia de una visión de largo plazo para realizar las transformaciones institucionales requeridas y la tendencia al cumplimiento mínimo de mandatos legales. Adicionalmente, la generación de valor social y económico con los datos no fue incorporada en las intervenciones previas, porque su reconocimiento como activo es un fenómeno reciente. Esto impone retos de actualización de la política pública para permitir la obtención, utilización, y reutilización,

así como el estímulo a la oferta de los bienes y servicios basados en datos. Por otro lado, las intervenciones en materia de protección de datos personales se enfrentan al desafío de garantizar los derechos asociados, en un contexto de aumento exponencial en las formas de generación y recolección de estos datos en formato digital, unido a la sofisticación de las capacidades para aprovecharlos.

La intervención pública es especialmente necesaria en situaciones donde las presiones competitivas son bajas o inexistentes, como ocurre en el sector público, y en los particulares que prestan servicios públicos o de interés público, así como en aquellos segmentos del sector privado donde la inversión inicial para la adopción es percibida como costosa (McKinsey Global Institute, 2013). En otras palabras, es necesario que las condiciones para la explotación de datos sean impulsadas mediante la intervención pública, corrigiendo las fallas de gobierno que impiden el surgimiento de elementos habilitadores. Lo anterior, mediante el aprovechamiento de un activo público que es generado de manera rutinaria y masiva, que por su naturaleza no es creado por el mercado: los datos públicos. El aumento en su accesibilidad y usabilidad da lugar a oportunidades económicas y a líneas de investigación, aumenta la transparencia y el acceso a la información pública, entre otros. De este modo, se genera una externalidad positiva que sería demasiado costosa para un actor privado.

Colombia requiere de una política pública de explotación de datos que habilite la generación de nuevos bienes, productos y procesos basados en datos. Esto aportará a la diversificación y sofisticación de la economía. Adicionalmente, mejorará la transparencia e incrementará el acceso a la información pública, así como la eficiencia y eficacia en la intervención pública a través de la toma de decisiones mejor informadas. Al mismo tiempo, esta política aportará al cumplimiento de los Objetivos de Desarrollo Sostenible. Concretamente, a las metas relacionadas con el desarrollo de la innovación en los países con un entorno normativo propicio a la adición de valor a los productos básicos (9.b.); la promoción del acceso público a la información (16.10); y el aumento significativo de datos oportunos, fiables y de alta calidad (17.18).

Por lo expuesto, y siguiendo lo definido en el documento Bases del Plan Nacional del Desarrollo 2014-2018 y en el artículo 45 del PND, es necesario que el DNP lidere la estrategia de *Big Data* del Estado, mediante la definición de la política de explotación de datos para la generación de valor social y económico.

3. MARCO CONCEPTUAL

En esta sección se definen los conceptos centrales para la política propuesta y los componentes de la misma. Finalmente, se explican las características que permiten afirmar que los datos son un recurso de infraestructura.

3.1. Conceptos principales: datos, información y conocimiento

Los datos son la representación primaria de variables cualitativas y cuantitativas que son almacenables, transferibles, pueden ser visualizadas, controladas y entendidas (Etymonline, 2016). Los datos que se perciben por los sentidos humanos, de manera continua y sin interrupciones, se denominan analógicos. Cuando son interpretados mediante codificación binaria, se trata de datos digitales. Estos últimos pueden generarse de dos maneras: por la interacción de personas con sistemas, herramientas y servicios digitales o automáticamente por programas de *software* y dispositivos de *hardware* que los capturan.

La recolección, almacenamiento y procesamiento de datos da lugar a la información, de la cual es posible obtener conocimiento. La disponibilidad de inmensas cantidades de datos digitales, unido a la capacidad para aprovecharlos mediante técnicas cada vez más sofisticadas, es lo que permite la creación de bienes, servicios, procesos, el mejoramiento de los existentes y la toma de mejores decisiones. Esto último es lo que deriva en la generación de valor social y económico.

De acuerdo con la organización y almacenamiento de los datos digitales, estos pueden dividirse de acuerdo con lo señalado en la Tabla 2.

Tabla 2. Organización y almacenamiento de los datos digitales

Estructurado	Están organizados conforme a un modelo o esquema. Se almacenan en forma tabular y algunas veces su estructura también incluye la definición de las relaciones entre ellos. Típicamente están representados en bases de datos que hacen parte del funcionamiento de sistemas de información.
No estructurado	Su organización y presentación no está guiada por ningún modelo o esquema. En esta categoría se incluyen, por ejemplo, las imágenes, texto, audios, contenidos de redes sociales, videos.
Semiestructurado	Su organización y presentación tiene una estructura básica (etiquetas o marcadores), pero no tiene establecida una definición de relaciones en su contenido. En esta categoría se incluyen contenidos de e-mails, tweets, archivos XML.

Fuente: DNP (2017).

La vinculación de datos estructurados con otros datos complementarios de cualquier tipo, que son leíbles por máquinas, da lugar a los datos enlazados. Este concepto hace referencia al método con el que se pueden mostrar, intercambiar y conectar datos en Internet. Los conjuntos de datos disponibles en este que se encuentran enlazados conforman la web de datos, es decir, una gran base de datos entrelazados.

Por otro lado, todos los datos (independientemente de la forma como se originan, organizan y almacenan) se clasifican de acuerdo con los criterios orgánico y cualitativo, para determinar su tratamiento jurídico, que define las limitaciones a su acceso, uso y divulgación (Tabla 3).

Tabla 3. Tipologías de datos

Criterio orgánico: sujeto que genera el dato	Público	Datos producidos o recolectados por una entidad pública, que ejerce funciones públicas o presta un servicio público (p. ej. cantidad de contratos suscritos, listado de funcionarios de una entidad).
	Privado	Datos producidos o recolectados por un particular (empresa, ciudadano, academia). Como las notas de los estudiantes o la historia laboral de los empleados de una empresa.
Criterio cualitativo: variables representadas	Personal	Datos que describen características que permiten identificar a una persona (p. ej. fecha de nacimiento, estatura, etc.).
	Impersonal	Datos que no permiten la individualización de una persona o se refieren a fenómenos que acaecen (p. ej. temperatura en un día, cantidad de vehículos que transitan por una calle, etc.).

Fuente: DNP (2017).

Las condiciones específicas de almacenamiento, compartición y uso de cada tipología de datos dependen del respectivo ordenamiento jurídico. No obstante, los datos generados por entidades públicas suelen presumirse públicos, como es el caso de la legislación colombiana³⁵. En relación con los datos personales, el consentimiento previo del titular para su almacenamiento, compartición y uso constituye la regla general, a partir de la cual se ha desarrollado su régimen de protección.

Los criterios orgánico y cualitativo determinan el nivel de publicidad y apertura de los datos en aplicación de los derechos a la intimidad, privacidad, la garantía del secreto profesional, comercial e industrial. Así mismo, definen el tratamiento como clasificado o reservado de aquellos datos cuya publicidad pueda causar daño a otros derechos³⁶ o a los intereses públicos³⁷, respectivamente.

Los datos dan lugar a la información cuando son procesados (*i.e.* asociación, combinación, interpretación, agrupación, estructuración, visualización) e ingresados a un sistema y conforman un conjunto que tiene significado para un receptor o intérprete. La disponibilidad de mayores cantidades de datos, de diversas fuentes, debidamente

³⁵ Cuando los datos se encuentran en poder de una entidad pública o que cumple funciones públicas, se presume que estos son públicos y le corresponde a la entidad demostrar que tiene otra clasificación, de acuerdo con el criterio cualitativo. Para lo anterior, de acuerdo con el artículo 5 de la Ley 1712 de 2014, se incluyen como entidad pública las que cumplen funciones públicas, de autoridad pública, los servicios públicos, las empresas con participación pública, entidades que administren recursos públicos, así como partidos y movimientos políticos.

³⁶ Como la intimidad, la vida, la salud, la seguridad y los secretos comerciales, industriales y profesionales.

³⁷ Como la defensa y seguridad nacional, la estabilidad macroeconómica, la información tributaria, los datos de las personas en situación de desplazamiento, el desarrollo de procesos e investigaciones judiciales y disciplinarias, entre otros.

procesados, da lugar a mejor y más información. Finalmente, cuando la información es aprendida, comprendida y aplicada se obtiene el conocimiento.

3.2. Explotación de datos (*Big Data*)

A principios de la década del 2000, la inmensa cantidad de datos generados por diversas fuentes, y principalmente derivados de la masificación de los dispositivos móviles, hicieron que los medios tradicionales de almacenamiento y procesamiento, creados en los años ochenta para la recolección selectiva de datos estructurados, resultaran insuficientes. Por lo anterior, la primera referencia al término *Big Data* consistió en la identificación del reto tecnológico que implicaba la generación de datos cuyo volumen, velocidad y variedad, desafiaba los sistemas tradicionales (Laney, 2001).

En 2005 estuvo disponible al público el primer *software* diseñado específicamente para atender los retos de almacenar y explotar los datos digitales que se encontraban en audios, videos y textos no estructurados. La aparición de nuevas formas de almacenamiento, procesamiento, análisis y visualización significó la superación del reto tecnológico que dio origen al término *Big Data*. Esto, unido a la identificación del potencial de los datos para generar valor social y económico, los posicionó como un nuevo factor de producción (OCDE, 2013).

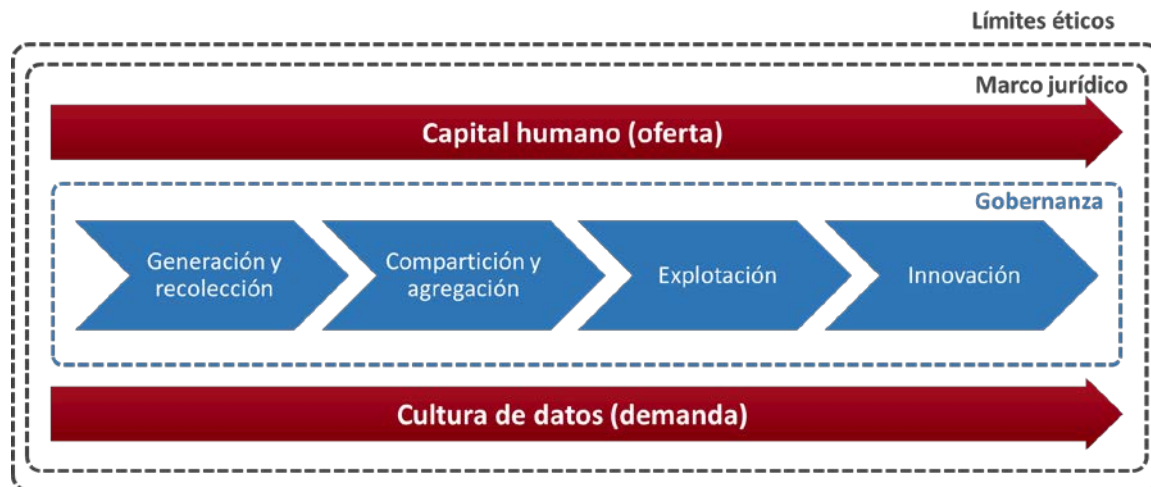
De este modo, la referencia técnica simple a las características de volumen, variedad y velocidad resulta insuficiente en tanto la generación exponencial de datos digitales significa que los datos considerados de gran volumen hace algunos años, hoy sean considerados “normales” (ITU, 2013). Por tanto, el desafío de los países en este contexto ahora va más allá de la tecnología y el almacenamiento digital. Ahora, el reto consiste en definir las condiciones para aprovechar los datos como insumo central de la economía digital, impulsado desde el sector público, así como mitigar los riesgos que puedan derivarse de la explotación de datos, para garantizar la protección de los ciudadanos en este contexto.

Por lo expuesto, actualmente la explotación de datos corresponde a la generación de valor social y económico mediante el aprovechamiento de estos para la creación de nuevos bienes, servicios, procesos, así como para el mejoramiento de los existentes³⁸. Este proceso³⁹ se representa en la Figura 2.

³⁸ Ejemplos de estos procesos de generación de valor se encuentran en el Anexo D.

³⁹ En todo lo relacionado con la seguridad digital el presente documento se sujeta a lo dispuesto en el Documento CONPES 3854 *Política nacional de seguridad digital*, así como las acciones previstas en el mismo.

Figura 2. Generación de valor con datos



Fuente: DNP (2018).

La generación, recolección, agregación y compartición dan lugar a las condiciones iniciales de disponibilidad de datos digitales, que constituyen el insumo básico para la generación de valor. Sin la disponibilidad de datos o la posibilidad de acceso a los mismos, no serían posibles las etapas posteriores que dan lugar a su aprovechamiento. Por ello, este documento resalta la necesidad de contar con datos digitales, como condición para alcanzar el objetivo de aumentar su aprovechamiento para la generación de valor social y económico. Adicionalmente, cuando la presente política menciona la explotación y el aprovechamiento de datos para la generación de valor social y económico se refiere a datos que cumplen las condiciones jurídicas para ser empleados en dicho proceso⁴⁰ y excluye aquellos cuyo tratamiento está proscrito. Los elementos señalados en la Figura 2 se explican a continuación.

3.2.1. Gobernanza

Es el conjunto de reglas que gobiernan el ciclo de vida y flujo de los datos de acuerdo con su tipología (*World Economic Forum*, 2015) con el fin de garantizar su calidad, uso, titularidad, compartición, seguridad y supresión, y que están orientadas a la generación de

⁴⁰ Datos impersonales, anonimizados, agregados, definidos como públicos por Ley, los autorizados por el titular de manera expresa, libre, espontánea, sin vulnerar el principio de finalidad, entre otros.

valor, minimizando los riesgos asociados a ello⁴¹. Esta se expresa a través de una política⁴² que articula los controles, las metas de negocio, objetivos estratégicos, los procesos involucrados e indicadores de gestión. La gobernanza enmarca los procesos de generación, recolección, compartición, agregación, explotación e innovación, que se describen a continuación.

Generación y recolección

Creación de los datos digitales por personas como resultado de la interacción con sistemas, herramientas y servicios digitales, o por máquinas con programas de *software* y dispositivos de *hardware* que capturan fenómenos.

Compartición y agregación

Corresponde a la disponibilidad de los datos para que sean conocidos, usados y reusados por terceros, mediante procesos y plataformas de combinación desde distintas fuentes (McKinsey Global Institute, 2016), incluyendo públicos y privados, diversos formatos y diversos tipos. Para ello se requiere eliminar barreras técnicas, procedimentales y jurídicas. Esto incluye la materialización del derecho de acceso a la información pública. Hacen parte de esta etapa los datos abiertos y la interoperabilidad.

Explotación

En esta etapa inicia la materialización del valor potencial de los datos, mediante la generación de productos de información útiles para la toma de decisiones. Lo anterior se obtiene empleando técnicas de analítica (OCDE, 2015). Esta es la disciplina orientada a analizar datos mediante técnicas científicas y herramientas automatizadas con énfasis en identificar hechos, relaciones, patrones ocultos de comportamiento de variables, correlaciones y tendencias⁴³, que brindan conocimiento respecto de los fenómenos de la realidad que antes permanecían ocultos debido a la complejidad de su medición y análisis por otros medios. De acuerdo con el objetivo que busca, la analítica puede ser descriptiva, diagnóstica, predictiva o prescriptiva.

⁴¹ Se distingue de la concepción tradicional de Gobierno de los datos porque incorpora consideraciones democráticas orientadas al aumento de la confianza, como los usos socialmente aceptados o límites reclamados por la ciudadanía.

⁴² Sus elementos son: administración de metadatos, calidad, integración de procesos de negocio, gestión e integración de datos maestros, ciclo de vida, procesos, actividades y medición continua. Adicionalmente, es necesario que se establezcan estructuras de Gobierno. Su implementación debe ser acorde con el modelo de madurez de cada organización y su arquitectura empresarial.

⁴³ Esto requiere medidas para evitar errores, sesgos y discriminaciones en las decisiones que se toman a partir de la explotación de datos.

El acceso cada vez mayor a herramientas y plataformas computacionales permite aprovechar datos que antes resultaban imposibles de procesar o combinar, porque su carencia de estructura o su volumen impedían obtener resultados en tiempos óptimos de manera asequible. El ciclo de vida de la explotación de datos se representa en la Figura 3.

Figura 3. Ciclo de vida de la explotación de datos



Fuente: DNP con información de Arcitura (2014).

Innovación

Ahora bien, la innovación basada en datos corresponde al aprovechamiento de los mismos mediante la aplicación de técnicas de analítica para mejorar o crear nuevos bienes, servicios o procesos, que aporten a la diversificación y sofisticación de la economía y a la generación de valor social, como una nueva fuente de crecimiento (OCDE, 2015). Para que la innovación efectivamente suceda se requiere la interacción de diversos actores e instituciones, que se orientan al desarrollo y la difusión de nueva tecnología y conocimiento. La explotación de datos facilita y acelera los procesos de innovación en tanto ofrece revelaciones que antes dependían exclusivamente del ingenio humano (McKinsey Global Institute, 2016). En este eslabón se obtiene el mayor valor de este activo.

3.2.2. Capital humano

Es el conjunto de conocimientos y pericias relacionadas con el aprovechamiento de los datos por parte de los ciudadanos, las empresas, el sector público y la academia. El capital humano hace posible la oferta de los bienes y servicios requeridos como elemento transversal para habilitar la explotación de datos, por ello es representado como una flecha en la Figura 2. A la fecha no existe una definición comúnmente aceptada de las condiciones que debe reunir el capital humano en esta materia. No obstante, el conocimiento requerido en todos los procesos involucrados está muy relacionado con la analítica, la ciencia de los datos, así como técnicas y tecnologías que soportan la analítica, provenientes de otros campos como la estadística, matemáticas aplicadas, economía, física y genética.

Adicionalmente, este nuevo contexto demanda personas con pensamiento creativo y curiosidad intelectual (TechAmerica Foundation, 2012).

Desde el punto de vista organizacional, es imperativo contar con personas suficientemente formadas para el aprovechamiento de los datos. La especialización en los conocimientos y habilidades requeridas depende del rol que desempeña cada uno, según se describe en el Anexo E. Como punto de partida para alcanzar las transformaciones institucionales requeridas para aumentar la disponibilidad de datos y aprovecharlos, debe impulsarse la explotación de datos desde el mayor nivel directivo y decisorio.

3.2.3. Cultura de datos

Hace referencia al reconocimiento del valor social y económico de los datos, y a su comprensión como un activo e insumo para la producción de bienes y servicios. Este elemento es habilitador de los cambios institucionales requeridos para el consumo de datos, así como de los bienes, servicios y procesos generados o mejorados con estos. Al ser un elemento transversal a la gobernanza, se representa como una flecha en la Figura 2.

La cultura de datos reúne a dos actores con diferentes capacidades para influir en el uso de los mismos como activo, estos son: los tomadores de decisiones del sector público y privado, y los ciudadanos.

Con respecto a los tomadores de decisiones en las organizaciones, además de las capacidades técnicas descritas, es imperativo que cuenten con una mentalidad orientada hacia los datos (Mckinsey Global Institute, 2011), que comprendan los beneficios de la toma de decisiones basada en este activo (Buhl, 2013) y la capacidad del mismo para generar innovación. En este contexto, las organizaciones cambian desde la alta dirección el paradigma de “trabajar con lo disponible” por el de “trabajar con lo que pueda obtenerse” (Gartner, 2012).

En relación con los ciudadanos, estos deben contar con alfabetización en datos⁴⁴, esto es, las capacidades para navegar en sus propios ecosistemas de datos para producirlos, apropiarlos, comunicarlos y usarlos (Data Pop Alliance, 2015). Dichas capacidades, implican habilidades para aprovecharlos, consumir servicios basados en estos y aportar a su construcción, mediante actividades como encontrar, determinar su credibilidad, producir e interpretar de manera crítica los datos⁴⁵.

3.2.4. Marco jurídico

⁴⁴ Concepto definido como el deseo y habilidad de participar en la construcción de la sociedad usando datos (Data-Pop Alliance, 2015).

⁴⁵ Incluye el uso de soluciones de procesamiento y análisis de datos basadas en computadores u otros dispositivos, y la búsqueda, compartición, uso y creación de contenido usando TIC e Internet.

Es el conjunto de elementos normativos (legales, reglamentarios y jurisprudenciales) que gobiernan y habilitan el aprovechamiento de datos. La generación de valor con los datos está supeditada a la garantía de los derechos y el cumplimiento de los fines del Estado, al igual que las demás actividades económicas y sociales que se someten al orden constitucional y legal vigente. Por tanto, según su tipología, los datos se someten a condiciones diversas de recolección, compartición y explotación, de acuerdo con las reglas definidas en el marco jurídico. En consecuencia, se entiende que los datos personales, sensibles, y aquellos protegidos por derechos de propiedad intelectual, secreto industrial, profesional, comercial, así como los que son objeto de reserva, gozan de protecciones jurídicas que imponen límites a su acceso y uso.

En el contexto contemporáneo de *datificación* de la vida diaria, en el que ocurren eventos no previstos por las normas es importante revisar los límites jurídicos actuales (o tradicionales) y su capacidad de responder a las situaciones generadas.

Por lo anterior, el marco jurídico abarca los procesos de generación, recolección, compartición, agregación, explotación e innovación, la gobernanza, así como las capacidades técnicas y la cultura de datos, los elementos habilitadores de oferta y demanda.

3.2.5. Límites éticos

Hace referencia a las pautas de comportamiento que orientan a los actores del ecosistema hacia los fines socialmente deseables, de modo que se tomen las decisiones que maximicen el bienestar social en cada circunstancia particular.

Los límites éticos generan confianza en el aprovechamiento de los datos y permiten resolver cuestionamientos frente a los cuales el ordenamiento jurídico resulta inadecuado o insuficiente (Data & Society Research Institute, 2014) porque estos temores se asocian con las aplicaciones imprevisibles e imperceptibles de la explotación de datos, que exigen el compromiso de los actores para respetar la dignidad humana y aumentar el bienestar social, mediante la definición de pautas o modelos de autorregulación.

3.3. Datos como recurso de infraestructura

Los datos son un insumo para la actividad investigativa, académica y productiva, para la toma de decisiones públicas y para la garantía de derechos. Son distintos a otros activos económicos en la medida en que su valor se materializa con el procesamiento y obtención de la información y el conocimiento intrínseco que se obtiene al combinarlos entre sí (European Data Protection Supervisor, 2014). Por tanto, el uso que se dé a los datos es el que, en últimas, materializa el valor potencial. Este activo reúne las características

económicas de bien público⁴⁶, de capital y de propósito general según se describe en la Tabla 4.

Tabla 4. Los datos como bienes públicos, de capital y propósito general

Tipo de bien	Características de los datos
Público (no rival, no excluyente)	Los datos no son rivales (pueden ser usados en múltiples ocasiones sin reducir su valor para otros usuarios) y no son excluyentes (el consumo de los datos por un actor no impide que otros actores los consuman, aunque no contribuyan a su mantenimiento y protección).
De capital	Los datos son empleados para generación de bienes y servicios, como un insumo de producción y su valor depende del contexto en el que sean utilizados.
De propósito general	En principio no hay límites a los usos que puedan darse a los datos, es decir, estos pueden utilizarse y reutilizarse en diferentes aplicaciones, incluso distintas a aquellas para las que fueron inicialmente recolectados.

Fuente: DNP con información de OCDE (2015).

Adicionalmente, el uso de los datos genera beneficios de escala, de alcance y efectos de redes, según se explica en la Tabla 5.

Tabla 5. Beneficios de escala, alcance y redes de los datos

Efectos	Descripción
Escala	Los servicios y productos creados o mejorados mediante la explotación de datos derivan en su mayor uso y, en consecuencia, en la generación de más datos.
Alcance	Enlazar diversos datos (que antes se encontraban en silos) permite obtener mejor información a partir de los mismos, de modo que se fortalece la contextualización de los datos cuyo valor en conjunto es mayor que la suma de las partes.
Redes	Los servicios y productos derivados de la explotación de datos y de la innovación basada en datos aumenta su utilidad (y valor) en la medida en que sean aprovechados por un mayor número de usuarios.

Fuente: DNP con información de OCDE (2015).

Las particularidades de los datos y su ubicación en el centro de la economía digital permiten afirmar que los mismos son un recurso de infraestructura⁴⁷. Este concepto debe ser considerado desde la perspectiva funcional en lugar de una puramente física u

⁴⁶ La referencia a bien público se realiza desde el punto de vista económico. No obstante, la clasificación jurídica de los datos implica que estos sean privados, semiprivados, públicos, clasificados, sensibles y reservados, conforme se describió.

⁴⁷ Esta consideración se realiza desde el punto de vista económico, sin que por ello se pretenda modificar o desconocer el tratamiento jurídico de los datos de acuerdo con la tipología a la que pertenezcan y, en todo caso, respetando las normas especiales en materia de datos personales, sensibles, intimidad, privacidad, reserva, secreto comercial, industrial, profesional, la propiedad privada, y demás que sean aplicables.

organizacional, es decir, que se entienda como las bases subyacentes o el marco de un sistema u organización, como medios compartidos para muchos fines (OCDE, 2015). De este modo, los datos son infraestructura en la medida en que en su conjunto soportan los productos, servicios y procesos generados en la economía digital. Son la base subyacente de la misma, al tiempo que aportan al desarrollo eficiente e innovador de las actividades económicas tradicionales. Igualmente, los datos satisfacen los criterios para ser considerados infraestructura⁴⁸ (OCDE, 2015).

Ahora bien, la infraestructura es provista por los gobiernos en tanto produce excedentes para la sociedad, o externalidades positivas, cuya generalidad impide que un único actor privado pueda encauzar y absorber los incentivos suficientes para incurrir en los costos para su implementación. En el caso de los datos, es posible transformar repositorios de datos desconectados en infraestructuras y mercados de datos (Suzuki, 2015). Adicionalmente, el mayor uso de los mismos implica mayor generación de valor⁴⁹ (OCDE, 2015). Por ello, las inversiones públicas realizadas para la disposición de los datos como infraestructura, que son esenciales para promover el crecimiento en la economía digital, darán lugar a más valor del que puede proyectarse, en tanto su aprovechamiento se caracteriza por generar servicios, productos y procesos que no fueron previstos al momento de la recolección y compartición de los datos.

4. DIAGNÓSTICO

En la actualidad, las actividades cotidianas de todo tipo son mediadas por desarrollos tecnológicos que disminuyen los costos de búsqueda y negociación de las transacciones económicas⁵⁰ y, al mismo tiempo, generan datos digitales. Por ejemplo, el uso de servicios de navegación para determinar la mejor ruta en los desplazamientos, el uso de plataformas de comercio electrónico, los servicios de calificación y recomendación de bienes y servicios,

⁴⁸ Esto es, un bien público (no rival, no excluyente), de capital y de propósito general (Tabla 4).

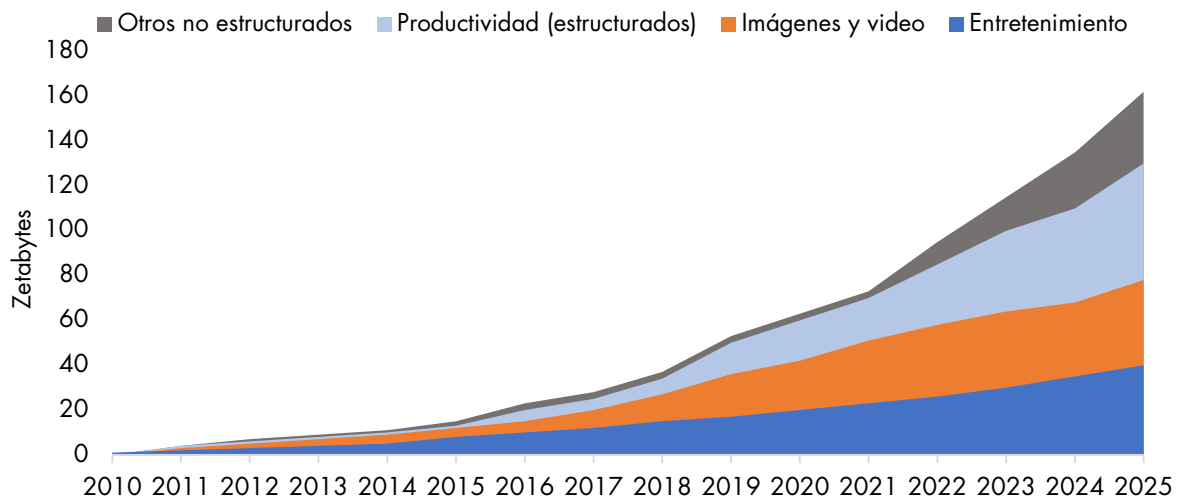
⁴⁹ En este caso, ocurre la comedia de los comunes, que en oposición a la tragedia de los comunes (en la que cada individuo buscando maximizar los beneficios de recursos compartidos deriva en el agotamiento del bien) (Hardin, 1968), implica la generación de mayor valor.

⁵⁰ Los costos de transacción se refieren a los esfuerzos que deben realizar los agentes económicos para perfeccionar una transacción, y se dividen en: i) búsqueda, ii) negociación y iii) vigilancia del contrato (Cooter, Ulen, 2012).

así como las redes sociales para compartir todo tipo de experiencias. Lo anterior, remueve los costos⁵¹, en tiempo y dinero, asociados a su recolección por medios tradicionales⁵².

Como se observa en el Gráfico 1, cada minuto del día millones de interacciones sociales dan lugar a datos digitales que reflejan el resultado de actividades humanas. La combinación y análisis de todos estos datos revela patrones que de otra manera no podrían conocerse o comprenderse y, al ser generados por eventos e interacciones con altos niveles de complejidad, permiten obtener un conocimiento mucho más preciso de la realidad⁵³. Esto da lugar a productos mejor adecuados a las necesidades, a modelos de analítica predictivos y prescriptivos para dar respuestas inmediatas y efectivas a diferentes problemáticas colectivas, entre otros.

Gráfico 1. Generación de datos estructurados y no estructurados proyectados a 2025



Fuente: DNP con datos de *International Data Corporation* (2017).

El uso de nuevas fuentes de datos e información como las redes de telefonía celular, los sistemas de localización o navegación, archivos digitales, como documentos PDF, o páginas web, ha permitido la identificación de fenómenos antes imposibles de medir debido

⁵¹ Para más información: https://www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2014/Discussion%20papers%20and%20presentations%20-%20GSR14/Session3_GSR14-DiscussionPaper-BigData.pdf

⁵² Observación de fenómenos para su posterior ordenación, mediante encuestas, estudios de campo, y demás formas manuales de captura de datos.

⁵³ Por ejemplo, la cantidad de esmog varía en una ciudad por factores como el clima, la humedad, la temperatura y el viento, haciendo extremadamente complejo predecir el comportamiento del mismo. Utilizando técnicas de analítica de datos de diversas fuentes (medioambientales, movilidad, crecimiento urbano, entre otros) se puede anticipar el comportamiento de este fenómeno e informar a las personas los mejores momentos para salir a la calle (Cooper, 2016).

a la complejidad, el costo y el tiempo que tomaba su observación y captura manual. Por ejemplo, mediante el análisis de texto de los planes de desarrollo territorial fue posible identificar sus tópicos principales y determinar su alineación con el PND 2014-2018 en 14 horas. La realización manual de este proceso habría requerido 3,8 años de trabajo humano (DNP, 2017).

Lo anterior, se constituye en un factor diferenciador en la economía digital, donde las organizaciones más avanzadas aprovechan los datos para aumentar los ingresos, entrar a nuevos mercados (e incluso crearlos), mejorar sus relaciones con los clientes y aumentar la eficiencia en las operaciones (McKinsey Global Institute, 2016). De este modo, las organizaciones que explotan datos para la toma de decisiones son entre 5 % y 6 % más productivas (European Political Strategy Centre, 2017). En general, la economía digital representa un cambio en el valor, que se centra menos en productos físicos y más en experiencias, acceso a información y personalización. Por ejemplo, entre 2008 y 2012 el comercio internacional de bienes y servicios creció 2,4 %, mientras que el de datos aumentó 49 % (European Political Strategy Centre, 2017).

Esto se evidencia, igualmente, en las compañías de mayor valor del mundo (Apple, Alphabet/Google, Amazon, Facebook, Microsoft, GE y Alibaba Group), donde la inversión en activos se centra en plataformas digitales, datos y talento analítico, mientras en el pasado estos recursos se invertían en fábricas y equipos (McKinsey Global Institute, 2016). Por lo anterior, el desarrollo de la explotación de datos para la generación de bienes, productos y servicios, así como para mejorar los existentes hizo evidente que los mismos son un bien de capital y un activo estratégico, cuyo aprovechamiento debe ser impulsado por los gobiernos nacionales (OCDE, 2015) para permitir el crecimiento económico en contextos donde los sectores tradicionales se encuentran rezagados.

Esto, igualmente, resalta la necesidad de desarrollar la economía digital en el país, concretamente, desde el aumento en la explotación de datos para la generación de valor social y económico, mediante la producción de bienes y servicios intensivos en conocimiento. En este contexto, se identifican cuatro situaciones principales que deben superarse para aumentar la explotación de datos: (i) bajos niveles de datos públicos digitales; (ii) desconfianza e incertidumbre respecto de la explotación de datos; (iii) bajo capital humano para la explotación de datos; y (iv) ausencia de cultura de datos. Estas situaciones se describen a continuación.

4.1. Bajos niveles de datos públicos digitales

En esta sección se describe la situación actual en cuanto a la baja disponibilidad de datos públicos digitales, la cual se divide en tres situaciones: baja digitalización, bajos niveles de datos abiertos y baja interoperabilidad. Se evidencia la ausencia de articulación de las políticas definidas e intervención insuficiente de la política pública, porque no se

dispone de herramientas y mecanismos concretos que permitan a las entidades cumplir los mandatos contenidos en las normas. Lo anterior, ha impedido la disponibilidad masiva de este activo público.

La digitalización, los datos abiertos y la interoperabilidad conforman las condiciones mínimas iniciales para la generación, recolección, compartición y agregación de datos. Es decir, permiten la disponibilidad y accesibilidad de estos para que posteriormente puedan ser aprovechados. Por ello, es esencial la remoción de barreras para permitir el acceso al insumo en que se sustenta la explotación.

Los antecedentes de la intervención pública para la implementación del Gobierno Electrónico, en articulación con el régimen de gestión documental, debían remover las barreras existentes y crear incentivos para la provisión de datos públicos digitales. No obstante, persisten retos para lograr niveles altos de digitalización, datos abiertos y de interoperabilidad.

4.1.1. Baja digitalización

La digitalización de documentos en Colombia está definida en el artículo primero del Acuerdo 27 de 2006⁵⁴, proferido por el Archivo General de la Nación, como la “técnica que permite la reproducción de información que se encuentra guardada de manera analógica (*i.e.* soportes, papel, video, cassettes, cinta, película, microfilm y otros) en una que solo puede leerse o interpretarse por computador”⁵⁵. En el contexto de la gestión archivística, este concepto está orientado a la conservación de documentos en alguno de los medios descritos⁵⁶.

De acuerdo con los resultados del Formulario Único de Registro de Avance a la Gestión (FURAG) publicado en 2016⁵⁷, solo el 28 % de las entidades del orden nacional había asignado presupuesto para llevar a cabo procesos de digitalización que, en promedio, correspondía al 1 % del total del presupuesto de las entidades. El 35,19 % de las entidades del orden nacional no realizó digitalización en 2015 y el 64,81 % de las entidades del orden nacional que sí lo hizo, indicó que mayormente lo realiza para facilitar

⁵⁴ En: <http://repositorio.archivogeneral.gov.co/repositorio/items/show/49>

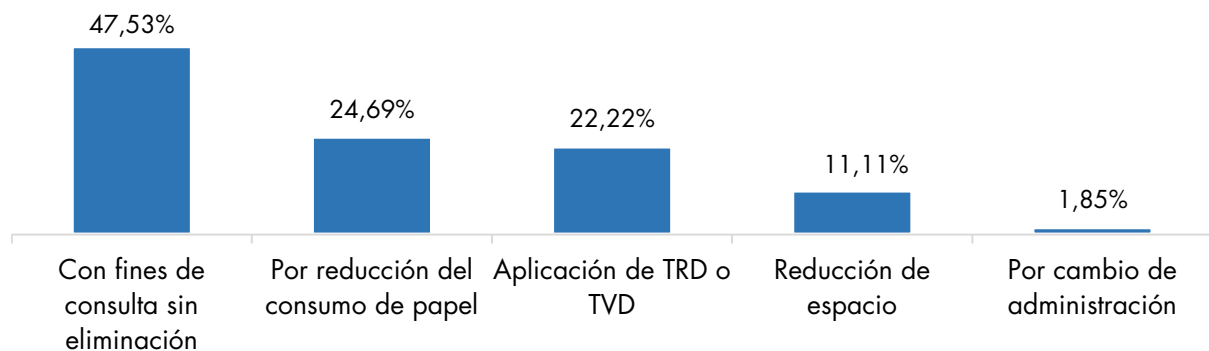
⁵⁵ Este concepto ha sido usado y difundido por el Archivo General de la Nación desde hace 25 años. Mientras que el término automatización ha sido empleado por esta Entidad para indicar las actividades que permiten agilizar el flujo de información mediante la utilización de sistemas informáticos (Archivo General de la Nación, 2017).

⁵⁶ La digitalización que no implica la eliminación del documento original análogo puede tener, al menos, tres tipos de aplicaciones: (i) digitalización con fines de control y trámite; (ii) digitalización con fines archivísticos; (iii) digitalización con fines de contingencia y continuidad del negocio.

⁵⁷ Evalúa la vigencia 2015.

la consulta de los documentos sin eliminar el original (47,53 %), así como para reducir el consumo de papel (24,69 %), situaciones que indican el uso de medios electrónicos principalmente para la gestión de los documentos de las entidades públicas (Gráfico 2). No obstante, la actual evaluación de la digitalización⁵⁸ no permite determinar si los datos contenidos en los documentos surten los procesos de generación y recolección, que son requeridos para que hagan parte de la cadena de valor de la explotación de datos. Esto, porque el proceso de transformación de un documento físico en digital no implica, por sí mismo, que se hayan incorporado las condiciones para generar datos aprovechables.

Gráfico 2. Propósito de la digitalización



Fuente: DNP con datos de DAFP (2016).

Lo anterior, sugiere la ausencia de articulación entre la intervención orientada a la gestión documental y el aumento del Gobierno Electrónico, porque no se evidencia un enfoque claro para alinear los propósitos de la conservación de documentos con la posibilidad de obtener el activo que está contenido en los mismos para su uso y compartición. Esto genera divergencias en las acciones implementadas por las entidades para uno y otro fin, que se refleja en la medición de las intervenciones.

Como se indica en el Gráfico 2, la aplicación de la Tabla de Retención Documental (TRD)⁵⁹ o la Tabla de Valoración Documental (TVD)⁶⁰ explica el propósito de digitalización en el 22,22 % de los casos en las entidades que la realizaron, ocupando el tercer lugar.

⁵⁸ A través del FURAG.

⁵⁹ Es el listado de series y tipos documentales, a las cuales se asigna el tiempo de permanencia en cada etapa del ciclo vital de los documentos. La TRD permite establecer cuáles son los documentos de la entidad, el tiempo de conservación y preservación, así como su disposición final (Archivo General de la Nación, 2017). Por otro lado, de acuerdo con el literal b) del artículo 18 del Acuerdo 5 de 2013 del Archivo General de la Nación, las series son el “conjunto de unidades documentales de estructura y contenido homogéneos, emanados de un mismo órgano o sujeto productor como consecuencia del ejercicio de sus funciones específicas”.

⁶⁰ De conformidad con el artículo segundo del Acuerdo 002 de 2004, la TVD es el listado de asuntos o series documentales a los cuales se asigna el tiempo de permanencia, así como su disposición final. Este instrumento es empleado para tratar los fondos acumulados, los cuales corresponden al “conjunto de documentos dispuestos sin ningún criterio de organización archivística, ni las mínimas condiciones de conservación y sin la posibilidad de ser fuente de información y consulta” según lo define el tercer inciso del mismo artículo.

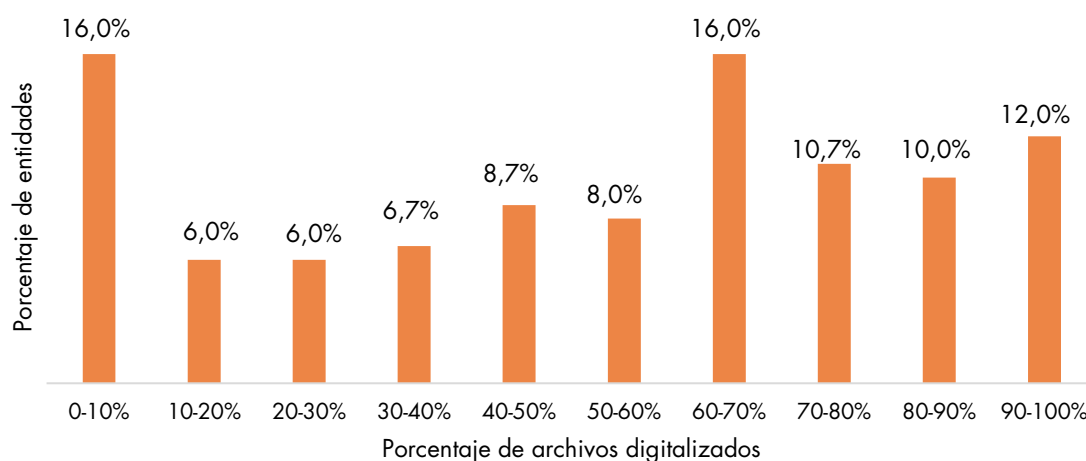
Dado que los instrumentos mencionados tienen la función de determinar, entre otras, la duración de los documentos en cada etapa de su ciclo vital, estos deberían ser el propósito principal de la digitalización en un 100 % de los casos.

De acuerdo con esto, y dados los resultados de FURAG, son evidentes las debilidades en la gestión de los archivos que contienen los datos públicos y la necesidad de que las TRD y TVD sean estandarizadas con el imperativo de aumentar la disponibilidad de los datos para que sean abiertos y accesibles por máquinas, y así remover barreras para su descubrimiento y aprovechamiento. Es decir, que incorporen la digitalización y automatización de modo que desde la generación de los documentos se favorezca la recolección y posterior aprovechamiento de los datos contenidos en estos.

Adicionalmente, de acuerdo con el estudio realizado por el Banco Mundial (2015) para Colombia, las entidades públicas del país suelen enfocarse en la implementación de tecnologías de digitalización como un fin y descuidan el proceso de gestión documental (captura, digitalización, tratamiento, gestión de metadatos, custodia, entre otros). El Archivo General de la Nación determinó que esto obedece a la falta de comprensión del valor y alcance de un adecuado tratamiento del documento administrativo⁶¹ (World Bank Group, 2015). Estos retos, persistentes en la gestión documental, afectan el acceso a los datos contenidos en los mismos y la posibilidad de su aprovechamiento.

Al respecto, el 67,3 % de las entidades tiene menos del 70 % de su información digitalizada (DNP, 2017) (Gráfico 3).

Gráfico 3. Porcentaje de los archivos de las entidades que están en formato digital



Fuente: DNP (2017).

⁶¹ Incluye documentos físicos y digitales.

A partir de la información que las entidades reportan como digitalizada se construyó el indicador que determina la desigualdad en los avances en la digitalización, teniendo en cuenta la proporción de información que aún reposa en papel en las organizaciones. Este indicador expresa los avances requeridos para masificar la digitalización en las entidades y se mide en un rango de 0 (0 % de la información se encuentra digitalizada por el 100 % de las entidades) a 1 (100 % de la información se encuentra digitalizada por el 100 % de las entidades). Actualmente, este corresponde a 0,55 lo que significa que, en promedio, el 55 % de la información de las entidades públicas se encuentra en un formato digital, en el que puedan extraerse los datos para su aprovechamiento, sugiriendo que hay un valor potencial significativo aún por explotar. Adicionalmente, el 33,33 % de las entidades manifiestan que no tienen planes para digitalizar la información que está en papel (DNP, 2017).

De este modo, es necesario que la modernización de los archivos públicos para aumentar la simplificación de trámites y procesos incorpore la digitalización como elemento transversal de la gestión de las entidades, teniendo en cuenta que la producción de datos digitales es la condición inicial para la generación de valor social y económico con este activo. En este sentido, el reto que debe superarse es la desarticulación entre las funciones de conservación archivística, que son propias de la política de gestión documental, con la automatización progresiva de procesos y procedimientos, que al mismo tiempo conlleve a la producción de datos públicos digitales, como elemento que hace parte de las acciones para el aumento de la eficiencia administrativa y el Gobierno Electrónico.

4.1.2. Bajos niveles de datos abiertos

En relación con la publicación de datos, existen igualmente retos importantes para promover desde el Gobierno la divulgación de la información pública y superar las barreras percibidas (no siempre reales) que manifiestan los encargados de la información.

Para la medición de los avances en materia de datos abiertos, existen dos abordajes metodológicos posicionados internacionalmente que establecen un ranking de países de acuerdo con el desempeño en materia de datos abiertos. La fundación *World Wide Web* realiza el análisis para 115 países, denominado Barómetro de Datos Abiertos (ODB, por sus siglas en inglés). De otro lado, la organización *Open Knowledge International* mide a 94 países mediante el Índice Global de Datos Abiertos (GODI, por sus siglas en inglés) (Open Knowledge International, 2017). En este ranking Colombia obtuvo un puntaje de 64 %, que lo ubica en décimo tercer lugar entre 94 países, mientras Taiwán, el mejor de la medición, alcanzó el 90 %. Brasil es el país de Latinoamérica con mayor puntaje (68 %) y es octavo en la medición global. Lo anterior, refleja los esfuerzos que se han realizado para aumentar la disponibilidad de conjuntos de datos específicos en Colombia.

Adicionalmente, en la última versión del informe denominado *Government at a Glance*⁶², realizado por la OCDE, Colombia fue incluido dentro de los países evaluados en el Índice gubernamental sobre datos abiertos, útiles y reutilizables (*OURData Index*, por sus siglas en inglés). Esta medición se realizó mediante una encuesta⁶³ respondida por los directores de sistemas de información⁶⁴ de cada país para determinar sus percepciones en cuanto a las prácticas y procedimientos relacionados con la apertura de datos (OCDE, 2017). Los avances del país en la formalización de acciones de Gobierno orientadas a publicar algunos datos de las entidades gubernamentales y la disponibilidad de un portal público en Internet que los centraliza se encuentran dentro de los principales esfuerzos que ubican al país en el cuarto lugar de la medición *OURData Index*, en la que había ocupado el décimo primer lugar en el año anterior de aplicación (2015).

Por otro lado, el ODB incorpora mediciones de varios dominios aplicados a 115 países, en procura de su objetividad⁶⁵, y se agrupa en tres variables: (i) preparación: ¿qué tan preparados están los gobiernos para las iniciativas de datos abiertos? ¿qué políticas existen?, (ii) implementación: ¿están los gobiernos llevando sus compromisos a la práctica?, (iii) impacto: ¿son los datos abiertos usados de forma que generen beneficios prácticos? (World Wide Web Foundation, 2017). Para su medición se emplean diferentes mecanismos de recolección de información que van desde encuestas a expertos, autoevaluación de cada uno de los gobiernos y estudios formales realizados por organismos internacionales como el Banco Mundial y Naciones Unidas, entre otros.

En este contexto, el ODB provee una visión integral respecto de los elementos que permiten corroborar la efectividad en la generación de valor con los datos públicos porque tiene en cuenta el impacto de las intervenciones y evidencias de la efectividad de las condiciones provistas, mientras que las mediciones GODI y *OURData Index*, se centran únicamente en la determinación de la existencia de condiciones consideradas como favorables. Los puntajes del ODB para Latinoamérica se presentan en el Gráfico 4.

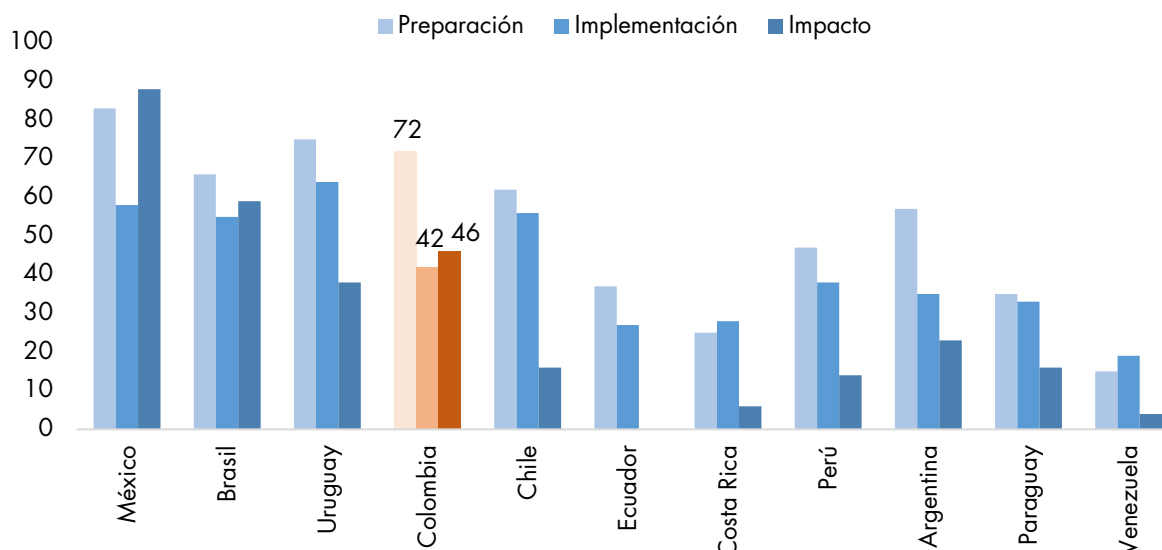
⁶² http://dx.doi.org/10.1787/gov_glance-2017-en

⁶³ Consta de 140 preguntas sobre apertura, uso y aprovechamiento de datos y es realizada por el grupo de *Open Digital Government and Open Data* de la OCDE.

⁶⁴ Conocidos como CIO por la sigla en inglés de *Chief Information Officers*.

⁶⁵ El proceso incluye tres componentes: una encuesta a expertos revisada por expertos en el área de datos abiertos, una autoevaluación del Gobierno y datos secundarios de organismos multilaterales que incluyen el Foro Económico Mundial, el Banco Mundial, la Encuesta de la administración electrónica de las Naciones Unidas y *Freedom House*.

Gráfico 4. Puntajes ODB Latinoamérica



Fuente: DNP con datos de la fundación *World Wide Web* (2017).

Colombia ocupa el cuarto lugar en la región y presenta brechas de avance entre cada uno de los aspectos medidos por el ODB. El puntaje en preparación (72) es el más alto de los tres dominios y permite identificar elementos favorables como la existencia de políticas y enfoques de gestión de los datos, presencia de acciones del Gobierno nacional y territorial, garantía de derechos y capacidad de las empresas y los emprendedores para aprovechar las oportunidades económicas ofrecidas por los datos abiertos. El puntaje de Colombia en implementación es significativamente más bajo (42) que el de preparación y se ve afectado por la disponibilidad real de datos abiertos del Gobierno, así como la calidad de los datos que sí están disponibles⁶⁶. Finalmente, es posible identificar una brecha considerable entre el impacto (46) y la preparación (72). El impacto económico, político y social, es medido por las evidencias de aprovechamiento real de los datos abiertos por parte del Gobierno, empresas y ciudadanía. El puntaje alcanzado por el país muestra la escasez de evidencias de dicho uso. En general, la medición del ODB sugiere que los esfuerzos realizados en materia de apertura de datos se han concentrado en proveer las políticas, protocolos y normas, orientadas a la publicación de estos.

De acuerdo con los indicadores internacionales, el desafío de la intervención pública es materializar la apertura por defecto y diseño de todos los datos públicos⁶⁷ que sean

⁶⁶ Los criterios de calidad que mide ODB incluyen que tengan el licenciamiento apropiado, que sean gratuitos, que estén disponibles en un formato usable y reusable, que estén actualizados, que sean fáciles de encontrar, que su publicación sea sostenible en el tiempo y que estén enlazados con conjuntos de datos relacionados.

⁶⁷ No solo de algunos conjuntos de datos que se consideren relevantes.

oportunos y de calidad. Para ello, además de las condiciones generales que ya han sido dispuestas, existe el reto de disponer de los mecanismos que hagan efectivas estas provisiones, alineando los incentivos de las entidades públicas, a efectos de reducir las brechas existentes entre la formulación de lineamientos, su implementación e impacto.

Por otro lado, en el marco del seguimiento a la implementación de la estrategia de Gobierno en Línea, el Ministerio de Tecnologías de la Información y las Comunicaciones hace seguimiento a los aspectos relacionados con la apertura de datos. Para las vigencias 2014⁶⁸, 2015⁶⁹ y 2016⁷⁰ esta medición ha sido modificada, teniendo en cuenta las actualizaciones de la estrategia, las recomendaciones de organismos multilaterales y los cambios normativos. Por lo anterior, no se ha guardado continuidad en las variables que son cuantificadas cada año⁷¹. Esta situación impide la realización de un comparativo histórico consistente.

Teniendo en cuenta lo anterior, en lo que se refiere a datos abiertos, a continuación, se describen los resultados relevantes.

Respecto de 2014 se indagó por la existencia del plan de apertura de datos y la realización del inventario de datos⁷² en las entidades del orden nacional. Se identificó que menos de la tercera parte de las entidades del orden nacional contaba con un plan de apertura de datos implementado (Gráfico 5) y solo el 37 % había realizado un inventario total de estos (Gráfico 6). Lo anterior, refleja las necesidades de mejora respecto de la gestión de los datos por parte de las entidades, que inicia con la identificación y cuantificación de aquellos a su cargo; esto es, la realización de un inventario que permita conocer todas las categorías de información de las entidades, los registros publicados y los registros disponibles para ser solicitados.

⁶⁸ En 2015 respondieron la encuesta 160 entidades del orden nacional, en la que se preguntó aspectos de la vigencia 2014.

⁶⁹ En 2016 respondieron la encuesta 155 entidades del orden nacional, en la que se indagaba por el desempeño de la vigencia 2015.

⁷⁰ En 2017 respondieron la encuesta 147 entidades del orden nacional, en la que se indagaba por el desempeño de la vigencia 2016.

⁷¹ A pesar de las modificaciones que se hagan a la estrategia. Por ejemplo, en 2014 se expidió el Decreto 2357, que modifica los componentes de la estrategia, amplía los plazos de implementación y deroga el Decreto 2693 de 2012.

⁷² Es el resultado de la identificación y caracterización de la información publicable, según la aplicabilidad de restricciones para ello y la priorización de los datos que se publicarán.

Gráfico 5. Porcentaje de entidades con plan de apertura de datos

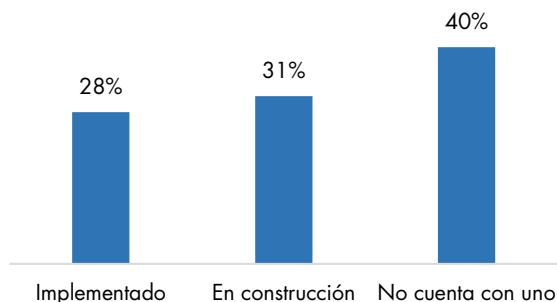
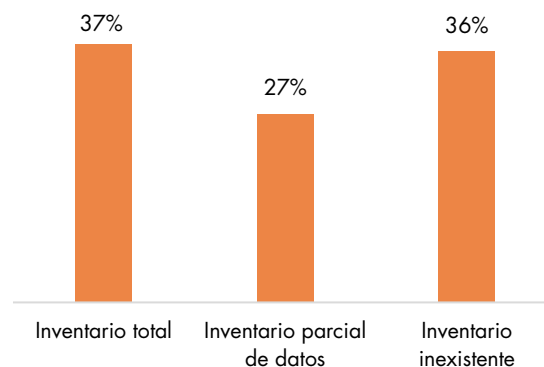


Gráfico 6. Porcentaje de entidades que realizan inventario de datos



Fuente: DNP con datos del Ministerio de Tecnologías de la Información y las Comunicaciones (2015).

Adicionalmente, de acuerdo con lo identificado por el Banco Mundial (2015), en las entidades del país no hay evidencia de un proceso sistemático de compilación y análisis para la elaboración de los registros de activos de información, por ello, faltan datos unificados, centralizados y disponibles para planificar el proceso de apertura y publicación.

En este sentido, el registro de activos de información dispuesto en el artículo 13 de la Ley 1712 de 2014⁷³, buscaba robustecer y reemplazar el inventario de datos que era propio de la estrategia de Gobierno en Línea. Este registro se estableció como un mecanismo de gestión de los datos buscando fortalecer las debilidades en su gobierno que limitan la apertura, en concordancia con la garantía de acceso a la información pública. A pesar de los avances en su adopción por parte de las entidades del orden nacional, solo el 58,37 % de estas han implementado el registro de activos de información (DNP, 2017).

Ahora bien, con respecto al año 2015, la medición de Gobierno en Línea indagó por la publicación de datos abiertos como parte de la estrategia de rendición de cuentas de las entidades. De acuerdo con lo anterior, apenas el 46 % de las entidades del orden nacional realiza dicha práctica. Esta pregunta empezó a ser relevante porque la reformulación de la estrategia en diciembre de 2014 incorporó la publicación de datos abiertos como un elemento del logro denominado transparencia. Lo anterior disminuye el reconocimiento de los datos como un activo para la generación de valor social y económico, porque limita su apertura principalmente a un asunto de transparencia y rendición de cuentas, reduciendo las posibilidades de masificar también la disponibilidad de datos públicos digitales que incentiven su aprovechamiento.

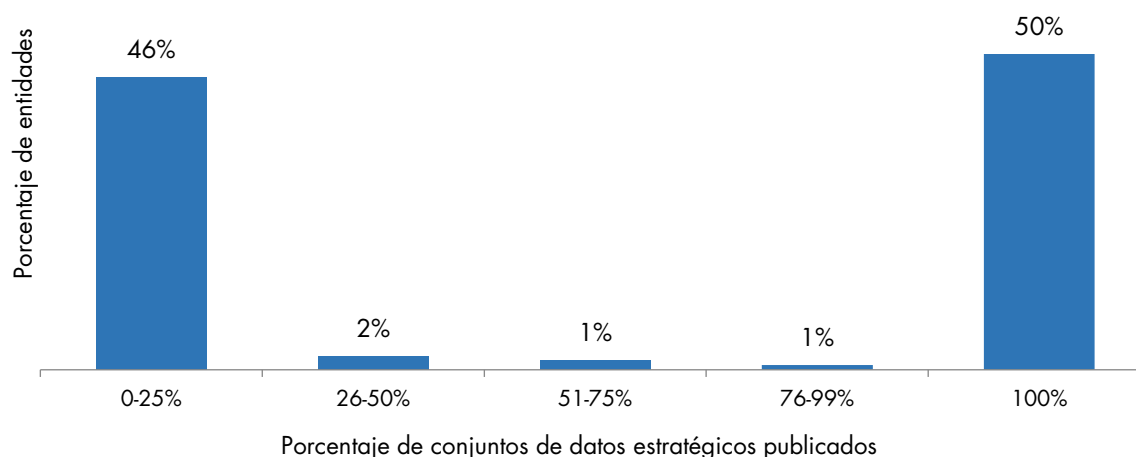
⁷³ Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones.

En 2016 se continuó con la medición del porcentaje de conjuntos de datos estratégicos publicados por cada entidad (Gráfico 7), esto es, aquellos que la entidad identifica como de mayor impacto en los usuarios, ciudadanos y grupos de interés⁷⁴. No obstante, la evaluación descrita no incluye la relación de los datos considerados como estratégicos o las razones concretas de su priorización por parte de cada entidad. Esta situación impide establecer con precisión el progreso en materia de apertura, en tanto se centra en el cumplimiento mínimo, pero no se dispone de información del total de datos susceptibles de publicación por entidad, o las discrepancias que estas puedan presentar con aquellos que se consideran estratégicos y priorizados.

De este modo, para el 2016 el 50 % de las entidades publicó la totalidad de conjuntos de datos considerados estratégicos, mientras que el 48 % había publicado menos de la mitad de estos conjuntos (Gráfico 7). No obstante, condicionar la apertura a que estos sean definidos como estratégicos contradice la definición de datos abiertos, en la medida en que se presumen públicos todos los datos gubernamentales, siendo responsabilidad de la entidad demostrar el fundamento de su reserva, clasificación o circulación restringida. Adicionalmente, el valor de los datos se materializa con su combinación y explotación, por lo que aspectos como su generación, recolección o el criterio del encargado de su apertura, son insuficientes para cuantificar su potencial o priorizar su publicación.

⁷⁴ Al respecto, la Guía de datos abiertos en Colombia (Ministerio de Tecnologías de la Información y las Comunicaciones, 2016), define los pasos a seguir para identificar datos que deben abrirse por las entidades públicas. Adicionalmente, define los temas priorizados mediante la Ruta de la Excelencia, correspondientes a: salud, educación, impuestos, movilidad, seguridad ciudadana, salud pública, atención y reparación a las víctimas y ordenamiento territorial. Estos temas se encuentran sujetos a actualizaciones de acuerdo con el avance de la estrategia. Por otro lado, el Banco Mundial (2015), determina que los datos representan mayor impacto y debe priorizarse su apertura, cuando aportan al cumplimiento de objetivos como el desarrollo económico, la mejora de los servicios públicos, el incremento de la transparencia y el Gobierno Abierto.

Gráfico 7. Porcentaje de conjuntos de datos abiertos estratégicos publicados al cierre de la vigencia 2016



Fuente: DNP con datos del Ministerio de Tecnologías de la Información y las Comunicaciones (2017).

Así mismo, de acuerdo con el análisis realizado en diciembre de 2017 al portal de datos abiertos de Colombia⁷⁵, se identificaron 2.368 conjuntos de datos publicados por 227 entidades del orden nacional. El 66 % de estas entidades ha publicado menos de 6 conjuntos de datos. Únicamente 10 entidades, correspondientes al 4 %, han cargado en el portal más de 50 conjuntos de datos⁷⁶. Lo anterior, sugiere que la apertura de estos se realiza para el cumplimiento de mínimos legales, en lugar de constituir una actividad rutinaria y proactiva de publicación para permitir el aprovechamiento de los datos para el público⁷⁷. Esto indica que las entidades aún no reconocen claramente que sus datos son un bien público que deben proveer a la sociedad.

En relación con la publicación de datos, el Banco Mundial (2015) determinó que en el país existe el riesgo de que los organismos públicos invoquen indebidamente la clasificación o reserva de la información, para evitar la apertura de datos, aun cuando la transparencia sea un bien superior al que se pretende proteger. Por esto, existe un reto de

⁷⁵ www.datos.gov.co

⁷⁶ Ministerio de Tecnologías de la Información y las Comunicaciones (273), Policía Nacional (236), DANE (168), Superintendencia de la Economía Solidaria (97), Ministerio de Agricultura y Desarrollo Rural (90), ICFES (79), Instituto Caro y Cuervo (63), INVIMA (61), Ministerio de Salud y Protección Social (61), Instituto Nacional de Medicina Legal y Ciencias Forenses (56).

⁷⁷ No obstante, existen portales de datos abiertos de iniciativa propia de algunas entidades públicas: Archivo Nacional de Datos del DANE (ANDA), Portal de datos abiertos geográficos del IGAC, Portal de datos abiertos de Biodiversidad Instituto Alexander Von Humboldt, y portales de datos abiertos de iniciativa local: Bogotá, Nariño, Tuluá, y Medellín.

apropiación del mandato expreso de apertura de datos y el principio de transparencia activa⁷⁸ dispuesto en la Constitución⁷⁹ y desarrollado en la Ley 1712 de 2014, en particular en su artículo tercero, de acuerdo con el cual la información se presume pública (por defecto y hasta que se demuestre lo contrario) y debe promoverse el acceso a la misma en los términos más amplios posibles.

4.1.3. Baja interoperabilidad

La interoperabilidad hace referencia a mecanismos y métodos que permiten a los sistemas de información interactuar e intercambiar datos de manera automática. En el marco del Gobierno digital, esta se orienta hacia la disminución de pasos en los trámites para la prestación de servicios al ciudadano y el aumento en la rapidez de la atención (implementación de cadenas de trámites). Para el aprovechamiento de datos, la interoperabilidad es imperativa porque agiliza y facilita la agregación y combinación de los mismos (Haav & Kungas, 2014). Esto requiere la alineación de las acciones de los responsables de los datos para facilitar el acceso y disminuir las gestiones aisladas.

Al respecto, el Banco Mundial (2015) detectó que las entidades públicas del país operan bajo el esquema de silos de información. Los sistemas desplegados funcionan de manera aislada y la información contenida en ellos es consultada solo por los encargados de la misma, sin relacionarse con la que se encuentra disponible en otra dependencia. Igualmente, existen debilidades en los sistemas de información misionales y su capacidad de interoperabilidad, debido al carácter general del marco de referencia definido para orientar estos aspectos en las entidades (World Bank Group, 2015), con escasa atención a la especificidad de cada sector. Por lo anterior, las entidades han generado desarrollos específicos para necesidades puntuales de intercambio de información que carecen de visión integral y sostenible (Criado, Gascó, & Jiménez, 2011). En este sentido, las entidades del orden nacional identifican como principal barrera para la interoperabilidad interna la alta complejidad para su implementación (37 %) (DNP, 2017), lo que refleja la persistencia en la baja capacidad para conectar los silos de información, que se propicia con la ausencia de estándares.

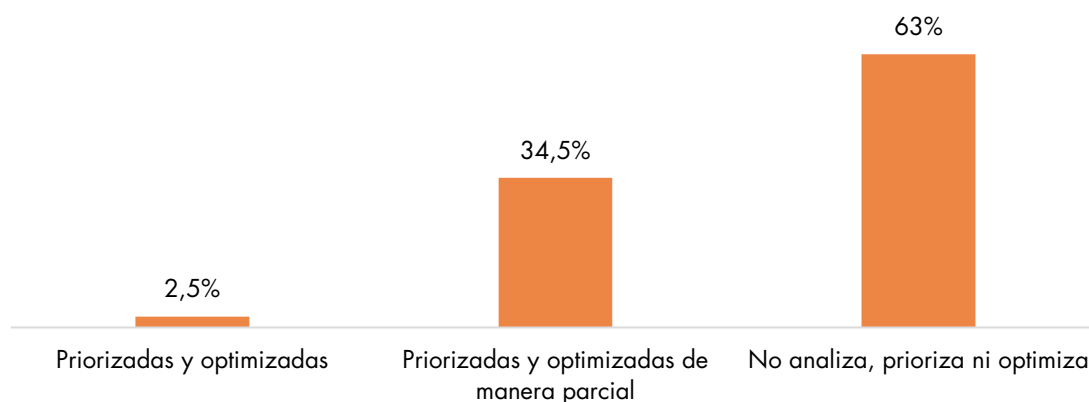
⁷⁸ Artículo 3 de la Ley 1712 de 2014: "(...) toda la información en poder de los sujetos obligados definidos en esta ley se presume pública, en consecuencia, de lo cual dichos sujetos están en el deber de proporcionar y facilitar el acceso a la misma en los términos más amplios posibles y a través de los medios y procedimientos que al efecto establezca la ley, excluyendo solo aquello que esté sujeto a las excepciones constitucionales y legales y bajo el cumplimiento de los requisitos establecidos en esta ley".

⁷⁹ Artículos 20, 23, 74 y 209, de la Constitución Política.

La interoperabilidad entre las entidades públicas es un mandato legal desde 2005⁸⁰, no obstante, para el cierre de la vigencia 2014 solo el 65 % de las entidades del orden nacional habían implementado el lenguaje común de intercambio (Ministerio de Tecnologías de la Información y las Comunicaciones, 2015). Este corresponde al estándar definido por el Estado colombiano para intercambiar información entre organizaciones, facilitando el entendimiento de los involucrados en los procesos respectivos.

Por otro lado, para 2014 solo el 2,5 % de las entidades había priorizado y optimizado las cadenas de trámites, es decir, implementaron la interoperabilidad para prestar un servicio al ciudadano (Gráfico 8). Lo anterior sugiere que, hasta el momento, las entidades conciben la interoperabilidad como el acceso a datos para la gestión institucional. Esto pone de presente el reto de implementar un enfoque de mayor impacto, orientado a la generación de valor mediante la optimización de los trámites públicos.

Gráfico 8. Avance en la implementación de cadenas de trámites por las entidades



Fuente: DNP con datos del Ministerio de Tecnologías de la Información y las Comunicaciones (2015).

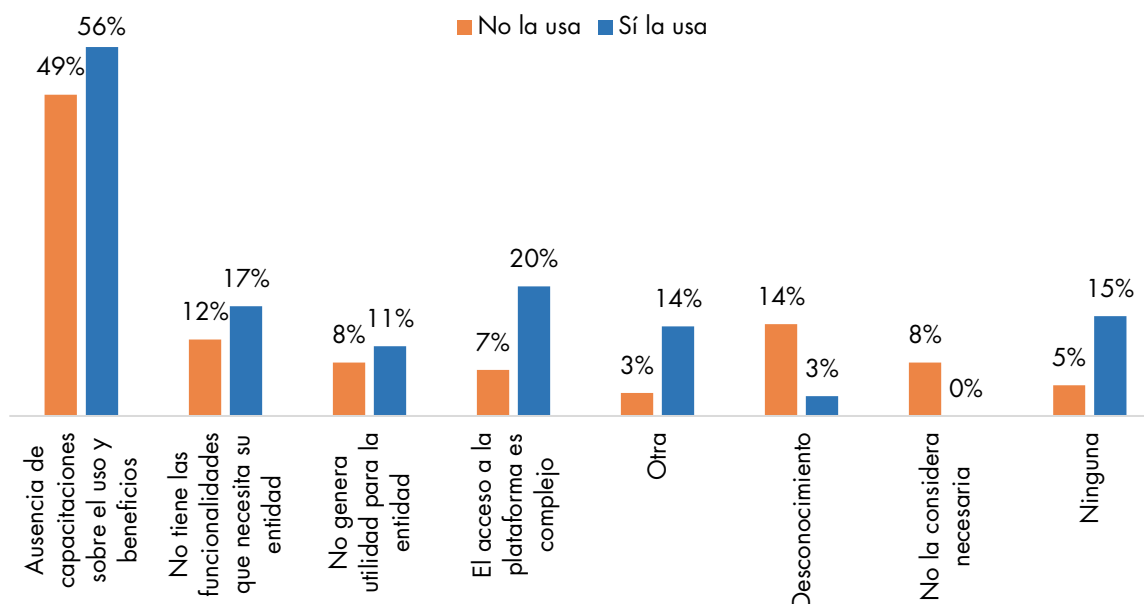
La plataforma de interoperabilidad del Estado (PDI) es un mecanismo que busca estandarizar los flujos de datos entre entidades, garantizando la calidad, la trazabilidad y el uso. A pesar de los esfuerzos para proveer este mecanismo y facilitar la interoperabilidad, en 2015 la PDI⁸¹ fue utilizada para suministrar o consumir información por el 44 % de las entidades públicas del orden nacional, el 50 % no empleó la PDI y el 6 % no respondió la pregunta (Ministerio de Tecnologías de la Información y las Comunicaciones, 2016). En la vigencia 2016, se identificó que el 57 % de las entidades había expuesto o consumido

⁸⁰ El artículo 14 de la Ley 962 de 2005 modifica el artículo 16 del Decreto Ley 2150 de 1995 y define que "(...) cuando una entidad pública requiera información de otra entidad de la Administración Pública, esta dará prioridad a la atención de dichas peticiones (...) para lo cual deben proceder a establecer sistemas telemáticos compatibles que permitan integrar y compartir información de uso frecuente por otras autoridades".

⁸¹ En: <http://www.mintic.gov.co/gestionti/615/w3-propertyvalue-6274.html>

datos en la PDI y el 43 % no la empleó (Ministerio de Tecnologías de la Información y las Comunicaciones, 2017). Ahora bien, el 15 % de las entidades usuarias de la PDI no encuentran barreras para ello, de lo que se deriva que el 85 % de las entidades restantes que han usado esta herramienta, identifican retos para su aprovechamiento, según se describe en el Gráfico 9.

Gráfico 9. Barreras que limitan el uso de la PDI por las entidades



Fuente: DNP (2017).

Lo anterior sugiere 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 incentivos⁸² para materializar la interoperabilidad⁸³. Esto es, acuerdos políticos, organizacionales, técnicos y presupuestales, para invertir los recursos requeridos.

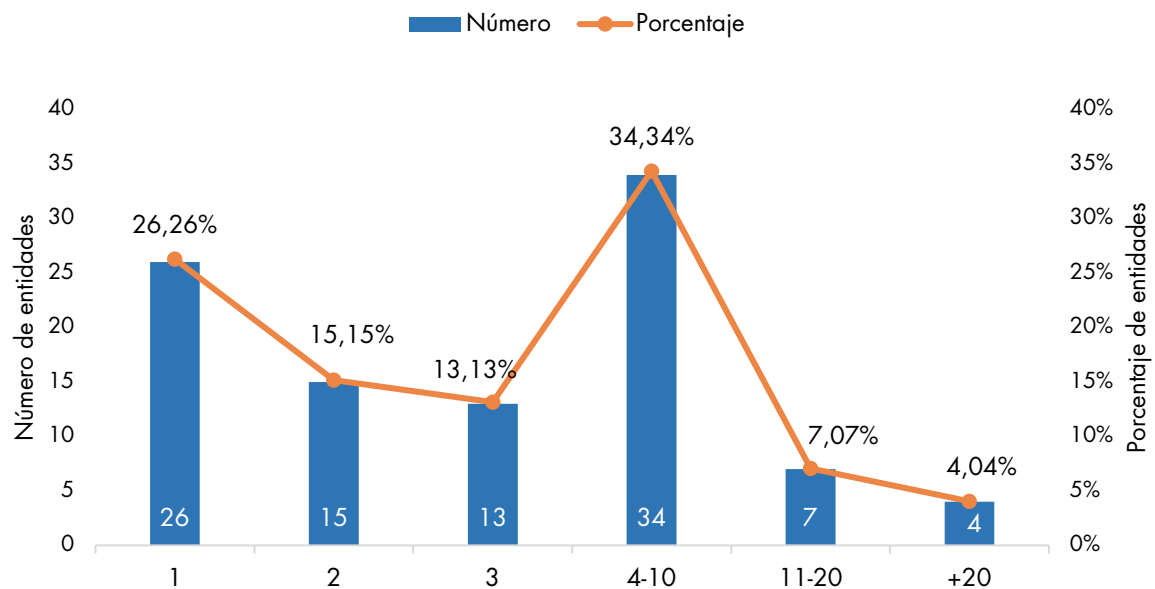
Por tanto, el mandato legal para racionalizar y simplificar trámites y el deber de compartir información entre entidades para aumentar la eficiencia administrativa es un avance importante desde el punto de vista normativo. No obstante, se requiere de esfuerzos

⁸² Ver, por ejemplo, el proyecto para la implementación del SIIF Nación en: <http://projects.worldbank.org/P106628/consolidation-national-public-management-information-systems?lang=en&tab=overview>

⁸³ Iniciativas adicionales en la materia son la Ventanilla Integral de Trámites Ambientales en Línea (VITAL), en la que se realizan trámites ante 35 autoridades ambientales del país. La Ventanilla Única de Comercio Exterior (VUCE), orientada a la integración de los trámites requeridos para operaciones de comercio exterior, en la que participan 21 entidades y se han realizado 3,8 millones de operaciones. La llave maestra de Prosperidad Social, que integra la información de las entidades del sector, con datos para la identificación de 20.000.000 de beneficiarios potenciales.

adicionales para adecuar las condiciones institucionales, de forma que estos mandatos se hagan efectivos y sea posible masificar la interoperabilidad. Al respecto, el 66 % de las entidades del orden nacional manifiesta que tiene interoperabilidad con otras entidades (DNP, 2017). De estas, el 54,5 % interopera con menos de 3 entidades (Gráfico 10).

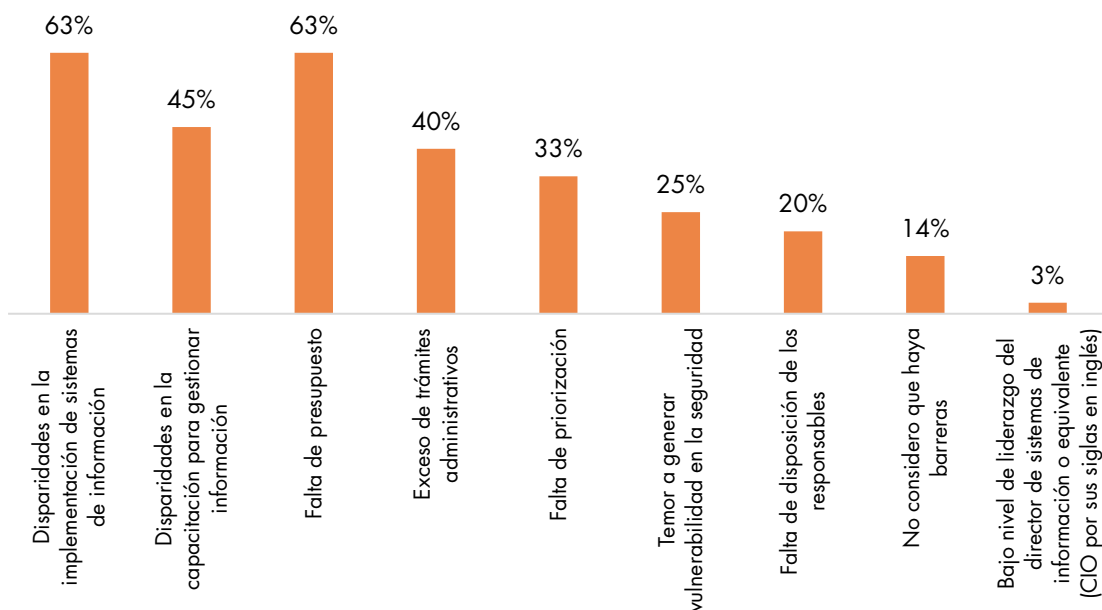
Gráfico 10. Entidades con interoperabilidad



Fuente: DNP (2017).

Adicionalmente, el 86 % de las entidades del orden nacional percibe limitaciones para implementar la interoperabilidad, de acuerdo con lo expresado en el Gráfico 11. Las disparidades en los avances y capacidades para la gestión de datos de las entidades son la principal limitación, que se explica porque no todas tienen el mismo avance en la implementación de sus sistemas de información, ni la misma capacitación para gestionar información.

Gráfico 11. Principales barreras para interoperar entre entidades



Fuente: DNP (2017).

Nota: respuesta múltiple.

La multiplicidad de registros administrativos⁸⁴ que se nutren con una proliferación de reportes, es otra manifestación de la necesidad de mecanismos que alineen las acciones de las entidades públicas en lo relacionado con la recolección, generación y compartición de los datos entre estas. Lo anterior, genera múltiples capturas, posibles errores y aumento del diligenciamiento manual para la recolección de datos⁸⁵. Así, los ciudadanos deben suministrar los mismos datos y documentos varias veces a las entidades públicas.

Igualmente, las necesidades de información de las entidades han sido resueltas de manera segmentada, dando lugar a la masificación de “registros únicos” mediante los que se recolectan y generan datos duplicados, de manera disímil.

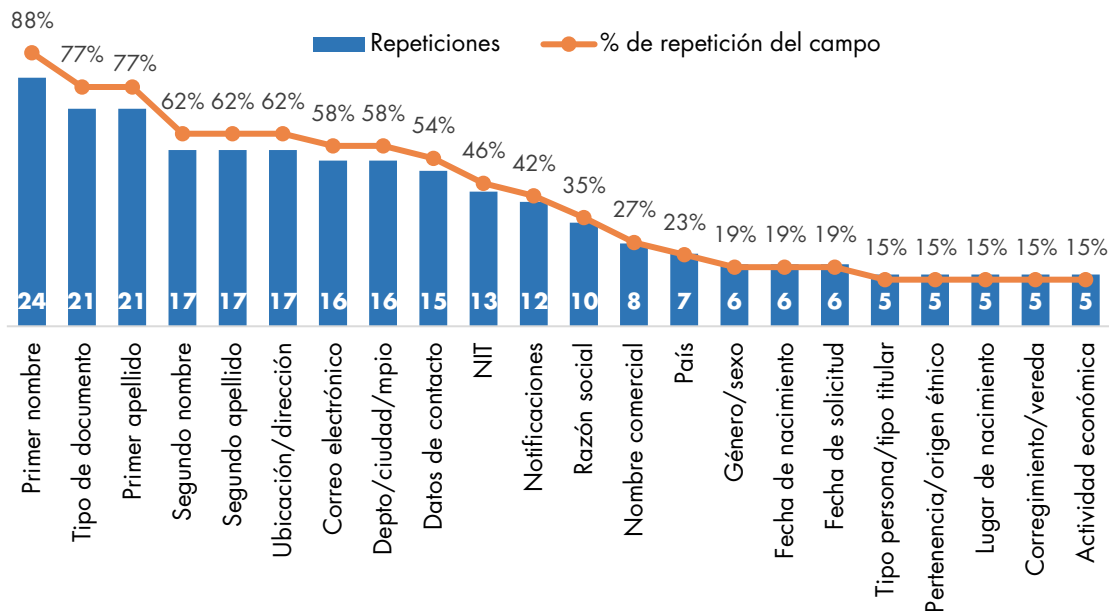
Al respecto, del análisis de 27 registros y sistemas de información que están disponibles para consulta pública en Internet (Gráfico 12), se identificó que estos agrupan un total de 115 campos y que, en promedio, cada sistema de información emplea el 10 % de estos campos (DNP, 2017). La identificación de estos 115 registros administrativos empleados en el funcionamiento de las entidades pone de presente la necesidad de superar los retos para la estandarización semántica y facilitar el aprovechamiento de los datos

⁸⁴ “Todo registro resultante de necesidad fiscales, tributarias u otras, creado con la finalidad de viabilizar la administración de los programas de Gobierno o para fiscalizar el cumplimiento de obligaciones legales de la sociedad” (Comisión Económica para América Latina y el Caribe (CEPAL), 2003).

⁸⁵ Aumentando la probabilidad de error en su recolección.

recolectados y generados por las entidades con mayor frecuencia, con el fin de permitir su acceso, intercambio y evitar múltiples capturas.

Gráfico 12. Índice de redundancia de campos de registros



Fuente: DNP (2017).

En general, el incumplimiento de estándares comunes para la gestión de los datos impacta su calidad, dificulta su interoperabilidad y genera la necesidad de esfuerzos adicionales en cada entidad para interpretarlos y emplearlos (Ministerio de Tecnologías de la Información y las Comunicaciones, 2013). Esto aumenta los costos de transacción para el uso de los datos, debido a los esfuerzos mencionados en los que deben incurrir las entidades.

4.2. Desconfianza e incertidumbre respecto de la explotación de datos

El marco normativo y de política pública relacionados con la explotación de datos fue proferido en un contexto anterior a la situación actual, donde existe una alta disponibilidad de este recurso y capacidades para su aprovechamiento masivo mediante técnicas de analítica y visualización. Por ello, este marco no tiene en cuenta el potencial de generar valor social y económico con los datos, y tampoco contempla la necesidad de mitigar los riesgos potenciales que se derivan del aumento de las capacidades técnicas y la creciente *datificación* de la vida diaria.

Dado lo anterior, se evidencia desconfianza e incertidumbre en relación con la explotación de datos, ocasionada por dos situaciones: (i) la dispersión y desarticulación del marco jurídico que se ocasiona con la producción normativa para la generación de normas en situaciones disímiles sin un principio armonizador, y (ii) los nuevos escenarios que surgen

con los avances en las técnicas para la explotación de datos, que requieren normas para mitigar los riesgos asociados. A continuación, se explican estas situaciones.

4.2.1. Dispersión y desarticulación del marco jurídico

En primer lugar, frente a la dispersión y desarticulación, el actual marco jurídico propicia situaciones que dificultan la explotación de datos y que hacen parte de las demás problemáticas identificadas. Esto, porque el marco jurídico impacta de manera transversal todas las acciones que adelantan los actores, para el caso concreto, en torno al aprovechamiento de datos. Estas situaciones se sintetizan en la Tabla 6.

Tabla 6. Retos del marco jurídico

Situación	Descripción
Multiplicidad de registros administrativos y reportes de información	Los datos son capturados por las entidades públicas de acuerdo con su necesidad de manera independiente, sin considerar que esta misma información puede ya reposar en otra entidad. A pesar de su naturaleza pública ^(a) , son de difícil acceso e intercambio entre las entidades ^(b) , y su publicación como datos abiertos es baja.
Debilidades archivísticas	La multiplicidad de registros administrativos involucra lineamientos de archivística, independientemente del medio en el que se generen, los cuales carecen de estandarización a la fecha. Adicionalmente, las normas de gestión documental han reaccionado de manera tardía a la incorporación de documentos electrónicos que fueron reconocidos legalmente desde 1999, a través de la Ley 527 ^(c) , pero reglamentados hasta 2012 con el Sistema Nacional de Archivos Electrónicos (SINAE).
Desaprovechamiento de datos actualmente disponibles	La baja digitalización y publicación de los datos públicos implica que la mayoría de estos sean oscuros.
Barreras de acceso a los datos públicos	Al no existir un mecanismo efectivo y sencillo para el intercambio expedito de información ^(d) , las entidades recurren a la figura del convenio interadministrativo, que para el caso concreto es un formalismo excesivo, supone una carga operativa que dificulta el intercambio dependiendo de la capacidad de gestión de cada una e implica la intervención de múltiples colaboradores en cada entidad para la definición de su contenido y perfeccionamiento. En este sentido, el 46 % de las entidades del orden nacional manifiestan suscribir convenios para el intercambio de información, mientras que solo el cumplimiento de un mandato legal ha sido empleado como mecanismo para facilitar el intercambio únicamente por el 27 % de estas (DNP, 2017).

Fuente: DNP (2017).

Notas: ^(a) En su mayoría, excepto lo sometido a reserva como información personal y sensible, así como aquello que afecte el secreto profesional y comercial; en cuyo caso debería realizarse una divulgación parcial, esto es, omitiendo los datos que tengan alguna restricción. ^(b) De acuerdo con la Directiva Presidencial 004 de 2012, a partir de 2013, las entidades públicas debían intercambiar correspondencia por medios electrónicos en aplicación de la política de eficiencia administrativa y cero papel. No obstante, el intercambio de correspondencia física persiste. ^(c) Por medio de la cual se define y reglamenta el acceso y uso de los mensajes de datos, del comercio electrónico y de las firmas digitales, y se establecen las entidades de certificación y se

dictan otras disposiciones. ^[d] Más allá del mandato de acceso contenido en los Decretos 235 y 2280 de 2010, y el artículo 159 de la Ley 1753 de 2015, que no hacen obligatoria la suscripción de convenios.

Por lo expuesto, se evidencia el reto de solventar los aspectos identificados (Tabla 6) bajo un criterio de simplificación y eliminación de barreras, que permitan superar las debilidades en la articulación propiciadas por la dispersión normativa.

Adicionalmente, la dispersión jurídica señalada da lugar a confusiones respecto de las tipologías de datos existentes y el tratamiento que debe darse a cada una, porque las normas expedidas para la protección de derechos y la transparencia definen el nivel de publicidad y circulación (Tabla 7).

Tabla 7. Tipologías de datos

Tipología	Descripción
Sensible	Afecta la intimidad y puede generar discriminación
Privado	Naturaleza íntima, solo es relevante para su titular
Semiprivado	Su conocimiento puede interesar a un grupo de personas o a la sociedad en general
Público	No pertenece a las tipologías anteriores o es definido por la ley
Abierto	Primario o sin procesar, en formatos estándar e interoperables, libre y sin restricciones. Puede ser usado y reutilizado
Clasificado	Puede generar daños a la intimidad, vida, salud o seguridad de una persona; o a los secretos comerciales, industriales y profesionales
Reservado	Acceso prohibido por la Constitución o la Ley

Fuente: DNP a partir de la Ley 1266 de 2008, Ley 1581 de 2012, Ley 1712 de 2014 y Decreto 1074 de 2015.

La pluralidad de criterios para definir a qué tipología corresponde un dato (Tabla 7) es una falla de articulación que constituye una barrera para la explotación de datos porque limita el acceso y uso de estos. Lo anterior disminuye la seguridad jurídica a los actores⁸⁶ porque las clasificaciones establecidas carecen de la sencillez y claridad necesaria para que los funcionarios tengan la certeza respecto de cuáles datos son reservados, clasificados, privados, semiprivados y públicos.

En este sentido, el Banco Mundial (2015), encontró que en el país es necesario fortalecer el concepto de que abrir datos para la reutilización, salvaguardando los principios reconocidos por la legislación en materia de transparencia y protección de datos personales, no es un delito. En concordancia con ello, solo el 31,58 % de las entidades del orden nacional manifiesta que ha aplicado alguna técnica de *anonimización* (DNP, 2017). Lo anterior confirma el requerimiento de fortalecer el conocimiento de los funcionarios respecto de las normas relacionadas con los deberes de apertura, publicación y

⁸⁶ A pesar de los esfuerzos en materia de divulgación y apropiación de las normas referidas, las dudas respecto de la clasificación de los datos persisten.

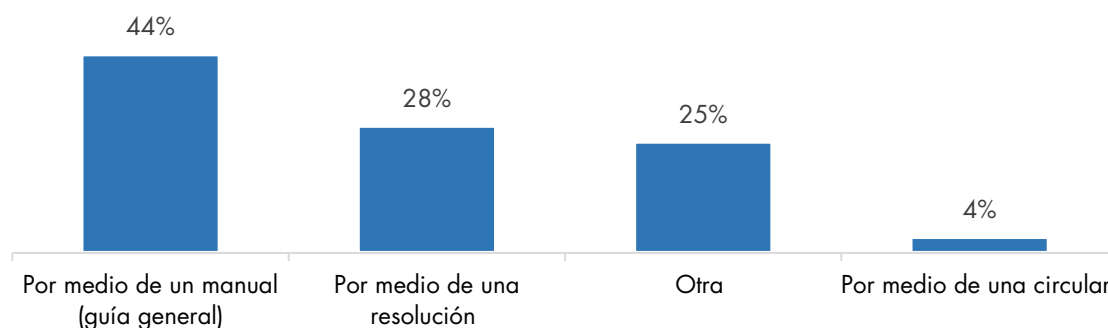
compartición de datos, en articulación con el régimen de protección de datos personales y de reserva de la información.

Adicionalmente, dentro de las preocupaciones de la industria para la adopción de desarrollos tecnológicos, el *habeas data* ocupa el tercer lugar (35,1 %) ⁸⁷ (ANDI, 2016). Esto evidencia que las complejidades en la definición del marco jurídico y la ausencia de un mecanismo que facilite la aplicación de las clasificaciones citadas en la Tabla 7 obstaculizan la explotación de datos e impactan la efectividad de las normas definidas.

Así mismo, la ausencia de claridad del marco jurídico puede afectar los derechos que se pretende proteger. Por ejemplo, si datos sensibles o privados son erróneamente identificados como públicos, como consecuencia de la complejidad de los criterios y clasificaciones referidos, los cuales generan dudas incluso entre los expertos en la materia ⁸⁸.

La adopción de una política de gestión de datos por parte de las entidades es un mecanismo que podría solventar la incertidumbre y aportar estandarización en el tratamiento de las situaciones descritas (Tabla 6). El 36 % de las entidades públicas la ha establecido. De estas, el 44 % empleó un mecanismo blando (*i.e.* instrumento con escasa capacidad vinculante y fácil modificación que limita la exigibilidad y efectividad en su cumplimiento) para su adopción (Gráfico 13).

Gráfico 13. Mecanismo de adopción de la política de gestión de datos



Fuente: DNP (2017).

Igualmente, el 34 % afirma tener definidas las reglas sobre la propiedad para la compartición y uso de datos (DNP, 2017). Es necesario precisar que las reglas sobre la propiedad para la compartición y uso de datos no se refieren a la necesidad de que cada

⁸⁷ El uso responsable de los desarrollos tecnológicos ocupa el primer lugar con 39,8%, lo que igualmente sugiere la necesidad de contar con un marco jurídico que ofrezca certeza respecto de las actuaciones permitidas.

⁸⁸ Ver la sentencia T-040/10 de la Corte Constitucional donde se indica que la errada categorización de la información como pública generó la equívoca aplicación del artículo 74 de la Constitución Política por parte del peticionario, del Defensor Regional del Pueblo, del plantel educativo y de los jueces de instancia, procediendo al suministro de información que era semiprivada y privada.

entidad adopte sus propios criterios para abrir y entregar datos públicos, de hecho, la adhesión a los criterios legalmente definidos (p.ej. en la Ley 1712 de 2014) es lo recomendable. No obstante, esta indefinición refleja la ausencia de claridad entre las entidades respecto de las reglas aplicables y la distinción entre la apertura por defecto y la identificación del responsable de los datos, que debe velar por su calidad y validez. Adicionalmente, el 53 % de las entidades confunden las licencias para la compartición de datos con los medios jurídicos y técnicos empleados para ello (DNP, 2017).

Los retos descritos implican que las entidades no implementan de manera proactiva políticas que definan las reglas aplicables a los procesos de generación, compartición, explotación e innovación. Esto, siguiendo estándares que se incorporen en el marco jurídico y se sometan a la vigilancia y control de una autoridad⁸⁹. Lo anterior ocurre porque no se ha definido normativamente la obligatoriedad de adoptar y garantizar el cumplimiento de políticas uniformes que gobiernen el aprovechamiento de datos. Así, el Banco Mundial (2015) estableció la necesidad de que el país incentive la adopción de mecanismos adicionales como Normas Corporativas Vinculantes u otros esquemas de autorregulación como códigos de conducta sectoriales, sellos de confianza y sellos de certificación, que garanticen un nivel adecuado de protección de datos, especialmente para los casos en los que debe garantizarse la privacidad.







La baja adopción de políticas de gestión de datos, que ocurre por la ausencia de obligaciones normativas al respecto, da lugar a la percepción de riesgos asociados a la explotación de datos. Así, el 69 % de las entidades del orden nacional percibe estos riesgos, correspondientes a posibles errores en los datos (48 %), violación de la privacidad (18 %) y de la reserva (1 %), es decir, el 73 % de los riesgos se relacionan con la debilidad en la aplicación de reglas y estándares en la gestión (DNP, 2017).

Ahora bien, de acuerdo con lo ilustrado en la Figura 4, actualmente las leyes que inciden en la explotación de datos asignan competencias difusas a varias entidades dentro de los regímenes señalados. Esto pone de presente la necesidad de articular el marco institucional para superar las fallas de coordinación, incertidumbres y duplicación de esfuerzos que se generan en este contexto. Por ejemplo, el Banco Mundial (2015) encontró que mientras el Archivo General de la Nación propuso incorporar un estándar semántico en el proceso de intercambio de datos, el Ministerio de Tecnologías de la Información y las

⁸⁹ Por ejemplo, en materia de protección de datos personales la reglamentación incorpora el principio de responsabilidad demostrada (artículo 2.2.2.25.6.1 del Decreto 1074 de 2015), de acuerdo con el cual las organizaciones deben ser capaces de probar, a petición de la Superintendencia de Industria y Comercio, que han implementado medidas apropiadas y efectivas para cumplir con las obligaciones establecidas en la Ley 1581 de 2012.

Comunicaciones planteó en 2011 el uso del lenguaje XML adaptado al Gobierno en Línea de Colombia (denominado GEL-XML). Esto indica que, si bien las entidades en mención han desarrollado esfuerzos para definir un modelo para la unificación semántica de los datos, cada una ha realizado avances independientes sin que se evidencie una gestión articulada⁹⁰.

Figura 4. Distribución de competencias

	Transparencia y acceso	Eficiencia administrativa	Gobierno electrónico	Gestión documental	Protección derechos	Política explotación de datos
 Industria y Comercio SUPERINTENDENCIA					●	●
 DNP Departamento Nacional de Planeación	●					●
 MINTIC MINISTERIO DE INTERIORES Y JUSTICIA	●				●	
 SECRETARÍA DE TRANSPARENCIA	●	●	●			●
 ARCHIVO GENERAL DE LA NACION COLOMBIA	●	●	●	●		●
 FUNCIÓN PÚBLICA	●	●	●	●		●

Fuente: DNP (2018).

La necesidad de articulación en materia de gestión de la información y adopción de herramientas para ello se ha evidenciado en oportunidades anteriores. Esta ha intentado solventarse mediante la conformación de comisiones intersectoriales compuestas por las entidades que tienen competencias afines. No obstante, este mecanismo no ha generado efectos concretos en el aumento de la coordinación. La OCDE (2013) analizó el rol de la Comisión Nacional Digital y de Información Estatal, creada mediante el Decreto 32 de 2013⁹¹ como instancia de coordinación y orientación superior del manejo de información pública. El citado estudio sugiere que este organismo podría ser insuficiente, en tanto

⁹⁰ Adicionalmente, en lo referido a la gestión de información por las entidades públicas se han dispuesto cuatro instrumentos: registro de activos de información, índice de información clasificada y reservada, esquema de publicación de información, las tablas de retención documental según el programa de gestión documental e inventarios de activos de información de acuerdo con el Sistema de Gestión de Seguridad de la Información. Estos fueron creados para fines diferentes, no han sido articulados e implican la duplicidad de esfuerzos en las entidades públicas para su gestión y administración.

⁹¹ Por recomendación del Documento CONPES 3701 *Lineamientos de política para ciberseguridad y ciberdefensa*, orientado a implementar instancias apropiadas para prevenir, coordinar, atender, controlar, generar recomendaciones y regular los incidentes o emergencias cibernéticas para afrontar las amenazas y los riesgos que atentan contra la ciberseguridad y ciberdefensa nacional.

arreglos institucionales similares en países OCDE incluyen una selección más amplia de ministerios y entidades⁹², que permiten la implementación efectiva de los asuntos requeridos⁹³. En este sentido, es importante señalar que una de las razones de supresión de la Comisión Intersectorial de Política y Gestión de Información en la Administración Pública en 2011 fue la falta de claridad en sus responsabilidades y débil capacidad de coordinación, generadas por sus recomendaciones no vinculantes, carencia de presupuesto, instrumentos y recursos.

4.2.2. Nuevos contextos generados por la explotación de datos

Las garantías específicas de los derechos asociados al tratamiento de datos personales conforman un régimen amplio y robusto. No obstante, están orientadas a la protección de estos en un contexto previo a la explotación masiva. En este sentido, el marco jurídico requiere de una actualización porque actualmente no incorpora los riesgos potenciales que surgen como consecuencia del aumento en las capacidades para el aprovechamiento de los datos y la creciente *datificación*.

Las transformaciones suscitadas por el aprovechamiento de los datos, por su naturaleza disruptiva, dan lugar a un nuevo contexto de intervención pública, producto del aumento en la digitalización de las actividades diarias y la evolución de las capacidades analíticas para explotar los datos. Lo anterior, hace imperativo un enfoque integral desde el Gobierno para maximizar los beneficios, mitigar los obstáculos y riesgos asociados (OCDE, 2015). Al igual que las demás actividades humanas, el aprovechamiento de datos tiene el potencial de generar riesgos. Estos se acentúan con el aumento en la disponibilidad de datos y de las capacidades técnicas para su explotación.

La mitigación de estos riesgos amerita la intervención pública, porque los titulares de los datos, es decir, las personas que los generan de manera consciente e inconsciente, constantemente se encuentran en una situación asimétrica respecto de las organizaciones privadas y públicas que los recolectan, agregan y explotan, las cuales tienen mayores conocimientos respecto de los usos posibles y sus consecuencias, usualmente imperceptibles

⁹² La Comisión está conformada por el Ministerio de Tecnologías de la Información y las Comunicaciones, el Departamento Nacional de Planeación, DANE, Departamento Administrativo de la Función Pública, la Dirección de Regulación Económica de la Seguridad Social del Ministerio de Hacienda y un representante del Presidente de la República.

⁹³ Mediante el Decreto 1499 de 2017, que reglamenta el artículo 133 de la Ley 1753 de 2015, se crea el Consejo para la Gestión y el Desempeño Institucional con el fin de integrar y articular de los sistemas de Gestión y Calidad y de Control Interno, para unificar y simplificar la conformación de comisiones intersectoriales. En virtud de lo anterior, mediante el Decreto 611 de 2018 se suprime la Comisión Nacional Digital y de Información Estatal.

para el titular. Por ejemplo, si los datos disponibles respecto de las preferencias de consumo de un ciudadano son usados para discriminar los precios de los productos que le son ofrecidos.

Igualmente, la explotación de datos para la generación de bienes y servicios es un proceso complejo⁹⁴ que requiere algún grado de conocimientos específicos para su comprensión. En este contexto, el titular de los datos suele otorgar autorización para su uso sin tener plena consciencia del alcance de esta aceptación⁹⁵, lo que presenta un reto para el alcance y aplicabilidad de los principios de libertad, finalidad y acceso, y circulación restringida, previstos en el marco jurídico actual⁹⁶.

El escaso tiempo de adopción de las políticas públicas para el aprovechamiento de datos en el contexto internacional⁹⁷ y el reconocimiento reciente del alto valor social y económico que estos aportan, implica que aún no se han solventado todos los retos relacionados con los potenciales riesgos identificados, del mismo modo que para otros avances tecnológicos y descubrimientos científicos⁹⁸. No obstante, se pone de presente la necesidad de construir marcos jurídicos y éticos pertinentes adecuados con lo anterior (Greenwood, Stopczynski, Sweatt, Hardjono, & Pentland, 2014). Estos aspectos deben hacer parte de la política pública para la explotación de datos, como ha sido el caso de las experiencias internacionales, particularmente de la Unión Europea y Estados Unidos.

De acuerdo con los referentes internacionales, dentro de los riesgos que deben ser tenidos en cuenta y examinados cuidadosamente, se encuentran, entre otros, el hecho de que el análisis de datos puede llevar a la toma de decisiones erróneas, debido a sesgos que no hayan sido contemplados o eliminados de los datos empleados⁹⁹. Por ejemplo, una herramienta usada para determinar la probabilidad de comisión de delitos y reincidencia

⁹⁴ Esta complejidad amerita la intervención pública, como suele suceder, por ejemplo, con los bienes y servicios financieros (seguros, pensiones, instrumentos de inversión).

⁹⁵ Por ejemplo, los datos entregados para el uso de aplicaciones y redes sociales.

⁹⁶ Artículo 4 de la Ley 1581 de 2012.

⁹⁷ El Reglamento general de protección de datos, expedido el 14 de abril de 2016 por el Parlamento Europeo, es la primera normatividad proferida con la intención de atender los riesgos que plantea la disponibilidad masiva de datos y el aumento en las capacidades científicas y tecnológicas para su aprovechamiento. El mismo entrará en vigor hasta el 25 de mayo de 2018.

⁹⁸ Como, por ejemplo, las armas, la energía nuclear, los medicamentos opioides, la conducción de medios de transporte, que en el momento de su introducción ameritaron la provisión de reglas específicas para maximizar los beneficios y minimizar los riesgos.

⁹⁹ Dentro de las condiciones mínimas que se proponen en la literatura internacional para prevenir errores, sesgos y discriminaciones en las decisiones que se toman a partir de la explotación de datos, se encuentran que las herramientas empleadas sean: responsables, explicables, precisas, auditables y justas (Diakopoulos & Friedler, 2016).

puede presentar resultados errados porque los datos contenían un sesgo racial que no fue identificado y eliminado¹⁰⁰.

Así mismo, los datos pueden ser empleados de manera indebida en contextos de transacciones privadas, por ejemplo, para imponer tasas más altas en la venta de seguros a partir de la revisión del historial de navegación en Internet¹⁰¹, limitar la capacidad de decisión mediante la exposición limitada de contenido¹⁰², o el aumento de precios de acuerdo con los patrones de consumo¹⁰³, entre otros.

En este contexto, donde los ciudadanos aceptan de manera rutinaria, y casi automática, los términos y condiciones de los bienes y servicios aparentemente gratuitos y que se lucran del uso de esos datos, libre y autónomamente entregados (Greenwood, Stopczynski, Sweatt, Hardjono, & Pentland, 2014), pone de presente la necesidad de evitar la entrega voluntaria de datos personales, pero sin plena conciencia de las consecuencias por parte de los ciudadanos.

El riesgo más comúnmente asociado al aprovechamiento de datos se refiere a las eventuales intromisiones a la privacidad de los individuos, concretamente el acceso a datos personales sin el consentimiento del titular. Al respecto, debe aclararse que el tratamiento de datos personales sin consentimiento es una vulneración al régimen de protección de derechos fundamentales, es decir, se trata de un uso ilegal que puede generarse en cualquier contexto¹⁰⁴ y no es propio o exclusivo de la explotación de datos masivos.

Para el caso concreto de Colombia, la actualización en los mecanismos de protección de los datos personales y los derechos asociados a estos (privacidad e intimidad) presentan las siguientes necesidades concretas:

- Protección del titular: la Ley 1581 de 2012 impone la carga de protección al titular del dato, quien debe consentir el uso del mismo. En el contexto de masificación de bienes y servicios que recolectan datos personales en todos los ámbitos de desenvolvimiento de una persona, la manifestación de este consentimiento se ha reducido a marcar una casilla manifestando haber leído y entendido los extensos

¹⁰⁰ <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

¹⁰¹ <http://www.bbc.com/news/business-24941415>

¹⁰² <http://thenextweb.com/insider/2016/03/14/962068/>

¹⁰³ <https://www.whitehouse.gov/blog/2015/02/06/economics-big-data-and-differential-pricing>

¹⁰⁴ Por ejemplo, cuando una empresa vende los datos recopilados a otra para fines de mercadeo, o cuando un empleado sustrae una base de datos con fines ilícitos, como el hurto de tarjetas de crédito.

términos y condiciones so pena de desistir del bien o servicio que solicita la autorización para su acceso y uso.

- Individualización a partir de datos impersonales: las normas fueron proferidas en un contexto donde no era técnicamente posible la individualización a partir de ciertos datos impersonales. Por tanto, no prevén claramente el tratamiento que debe darse a los datos anonimizados, los cuales, unidos con otros de naturaleza pública, pueden brindar información de una persona determinada o determinable. Tampoco se ha previsto el procedimiento que debe seguirse respecto a los datos impersonales, los cuales, luego de haber sido analizados de manera legal¹⁰⁵, se convierten en personales¹⁰⁶.

4.3. Bajo capital humano para la explotación de datos

Las debilidades en materia de capital humano dificultan la oferta de bienes y servicios generados a partir de los mismos, impidiendo la materialización de su valor potencial. Por ello, es un elemento transversal habilitador del aprovechamiento de datos que hasta el momento no han sido objeto de una intervención pública concreta para su fortalecimiento.

Actualmente es escasa la evidencia y cuantificación entre países del déficit o disparidad de habilidades para la explotación de datos. Sin embargo, se ha indicado que la demanda de destrezas de especialistas de datos superará su oferta en el mercado laboral (OCDE, 2015).

En cuanto a los profesionales en TI¹⁰⁷, desde hace varios años se identificó en el mundo un déficit que persiste hasta la fecha (Gartner, 2015). Dentro de este déficit suelen incluirse las necesidades de profesionales en materia de explotación de datos y las soluciones para su aprovechamiento.

Para el caso de Colombia, la industria ha evidenciado que existe una baja oferta de personal calificado en TIC y en competencias blandas (p.ej. liderazgo, creatividad, trabajo en equipo, comunicación, innovación y bilingüismo) que pueda desempeñarse con solvencia en la economía digital (Accenture; Cintel, 2016). Adicionalmente, la industria

¹⁰⁵ Recolectados con consentimiento informado, para finalidad legítima, almacenados de manera segura, posteriormente anonimizados y procesados o analizados con técnicas analíticas, para fines legítimos.

¹⁰⁶ No obstante, en principio podría sugerirse su tratamiento como dato personal, lo que implicaría, entre otros aspectos, la obtención de una autorización *ex post* de titular respecto de un dato que ya ha sido recolectado, almacenado y tratado (a pesar de la previsión legal de autorización como requisito previo). Ver: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006.

¹⁰⁷ Según la más reciente encuesta de *Manpower Group*, el talento en TI está clasificado de noveno en la lista de tipos de talentos más difíciles de conseguir. Talento en TI está definido como desarrolladores y programadores, administradores de bases de datos, y líderes de TI y gerentes. Consultado en: https://www.manpowergroup.com/wps/wcm/connect/408f7067-ba9c-4c98-b0ec-dca74403a802/2015_Talent_Shortage_Survey-lo_res.pdf?MOD=AJPERES&ContentCache=NONE

realiza un llamado a la academia para incorporar en sus programas las habilidades y competencias requeridas por el actual entorno económico tecnológico, desde la educación básica y a lo largo de la vida, con el fin de que el país cuente con el recurso humano necesario para aprovechar las oportunidades que brinda la economía basada en el conocimiento.

Por otro lado, el desempeño en matemáticas y ciencias impacta la disponibilidad de habilidades para el aprovechamiento de los datos. Por esta razón, es importante tener presente que los resultados obtenidos por los jóvenes del país en mediciones internacionales¹⁰⁸, indican la necesidad de realizar esfuerzos importantes para alcanzar en el medio plazo el promedio de la OCDE en las áreas mencionadas (BID, 2014).

Lo anterior, sugiere que las debilidades en las capacidades técnicas deben atenderse de manera integral, considerando todas las habilidades requeridas para el aprovechamiento de los datos. Entre otras, el conocimiento en matemáticas, estadística, aprendizaje de máquinas y ciencia de los datos. Así como los conocimientos complementarios en torno a los datos aplicables a todas las profesiones.

Al respecto se evidencia baja oferta de programas en esas áreas. Una búsqueda en el Sistema de Información sobre Educación Superior reporta en todo el país 24 programas de postgrado sobre estadística aplicada, 39 sobre matemáticas aplicadas y 7 de analítica o ciencia de los datos¹⁰⁹. Lo anterior sugiere un escaso alistamiento de los profesionales actuales y futuros para entrar en los campos de estudio requeridos para el aprovechamiento de los datos, así como para generar innovación basada en datos. Por ejemplo, el Banco Mundial (2015) identificó que en el país solo un 15 % de los periodistas ha escuchado hablar del periodismo de datos y se identifican solo 10 profesionales en todo el país que tengan conocimientos para poder ejercerlo. Igualmente, determinó que los profesionales contratados para interactuar en las redes sociales de las entidades públicas requieren desarrollar habilidades para localizar, interpretar y analizar datos, así como para realizar visualizaciones e informaciones con los mismos.

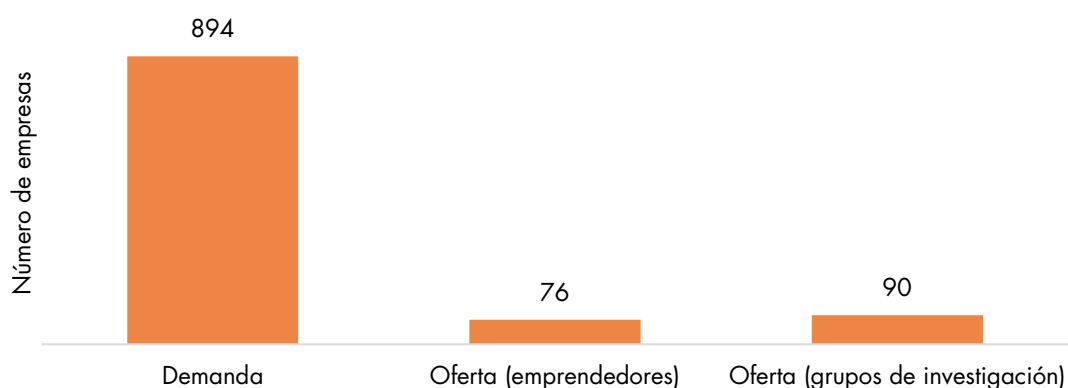
Actualmente no se dispone de una medición de la brecha de capital humano en materia de explotación de datos que permita, entre otros aspectos, revisar las necesidades del marco nacional de cualificaciones en esta materia. No obstante, se evidencian dificultades por parte de las organizaciones para ubicar y contratar servicios de explotación de datos y analítica, que sugieren la existencia de la brecha mencionada¹¹⁰ (Gráfico 14).

¹⁰⁸ Pruebas PISA 2012 (OCDE, 2012).

<http://snies.mineducacion.gov.co/consultasnies/programa#>

¹¹⁰ Información de Sunn, plataforma virtual que conecta la oferta (investigadores y *startups*) con la demanda (empresas e inversionistas) en materia de innovación tecnológica. Disponible en: <http://www.sunnlatam.co/>

Gráfico 14. Oferta y demanda de servicios de explotación de datos y analítica



Fuente: DNP con información de Ruta N (2018).

El capital humano requerido para la explotación de datos no se limita a los profesionales en TI, haciéndose necesarios perfiles cualificados en todo el ecosistema. En este sentido, el Banco Mundial (2015) encontró que, si bien mediante el programa para la excelencia se han capacitado miles de funcionarios en temas relacionados con Gobierno Electrónico, entre las materias impartidas no se encuentra capacitación en ciencia de datos, excepto aspectos incluidos en algunos cursos, como el de seguridad en el manejo de información.

Si bien el 63 % de entidades del orden nacional manifiesta que cuenta por lo menos con una persona encargada de realizar analítica, solo el 36 % de las entidades afirma que emplean algoritmos para llevar a cabo la analítica (DNP, 2017). Esto indica debilidades en las capacidades técnicas en tanto la explotación de datos que se realiza mediante la analítica implica, necesariamente, el uso de algoritmos.

Adicionalmente, al indagar por el tipo de algoritmo empleado, el 74 % de las entidades que indicó su utilización desconocía o confundía estos con herramientas para procesar información (DNP, 2017). Esto sugiere el desconocimiento existente entre las entidades en relación con las herramientas requeridas para el aprovechamiento de datos. Así mismo, refleja que actualmente existe poca adopción de estas, que incide en la baja explotación de datos para la generación de valor en la gestión pública.

Las capacidades técnicas inciden en la preparación de las entidades para materializar efectivamente el valor potencial de los datos. Al respecto, el 66 % de las entidades del orden nacional manifiesta que no ha empleado datos no estructurados para la implementación de una solución de analítica. De las que lo han hecho, el 88 % se ha concentrado en el análisis de texto (DNP, 2017). Esto sugiere la ausencia de capacidades para el aprovechamiento de los datos, porque la mayoría de las entidades no ha usado

aquellos que provienen de fuentes diferentes a la organización tradicional tabular (datos estructurados en filas y columnas) y las entidades que sí lo han hecho se concentran en un nivel inicial de explotación, dado por la baja complejidad que ello implica.

Igualmente, las entidades del orden nacional manifiestan que la falta de capacitación respecto de la analítica de datos y sus aplicaciones es la principal barrera (63 %) para el aprovechamiento de datos (DNP, 2017).

Lo anterior indica que las entidades públicas actualmente presentan una baja preparación para aprovechar datos, porque no ha existido entrenamiento específico en la materia y porque la evolución de la ciencia y la oferta académica en relación con la explotación de datos es reciente, lo que reduce la disponibilidad presente de profesionales que cuenten con estas capacidades.

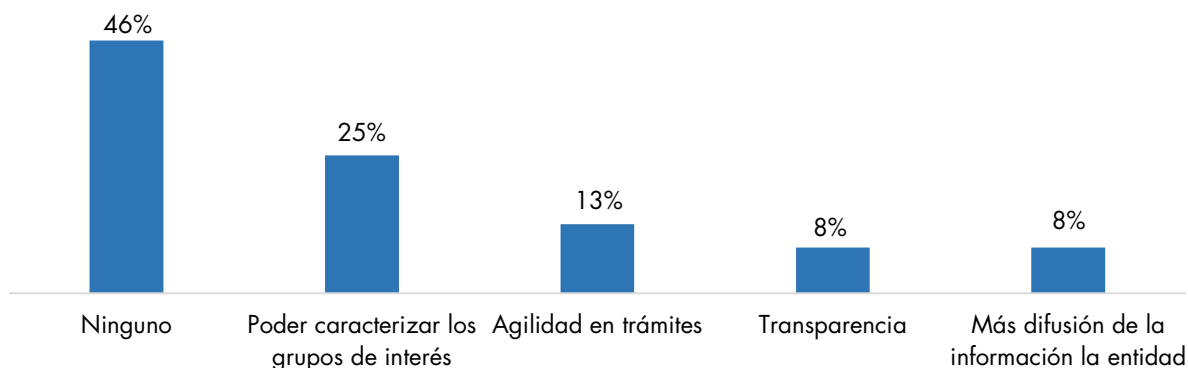
Por ello, es necesario aumentar el capital humano como elemento que permita hacer efectivo el potencial que representan los datos. Así, el 88 % de las entidades del orden nacional que no cuenta por lo menos con una persona encargada de realizar analítica tampoco ha desarrollado un servicio a la ciudadanía a partir del aprovechamiento de datos (DNP, 2017).

4.4. Ausencia de cultura de datos

El reconocimiento del valor económico y social de los datos por parte de las entidades públicas, privadas, la academia y la ciudadanía genera demanda de bienes y servicios basados en este activo. Así mismo, empodera a las personas para comprender, interpretar y usar los datos que encuentran, así como analizar los que producen.

La ausencia de cultura de datos, o en otras palabras la baja demanda, se refleja en aspectos como el desaprovechamiento de estos y los temores que se asocian con su explotación. Cerca de la mitad de las entidades públicas del orden nacional perciben que aumentar la disponibilidad de datos abiertos no solucionaría alguna necesidad (Gráfico 15).

Gráfico 15. Necesidades que se solucionarían con la apertura de datos de las entidades

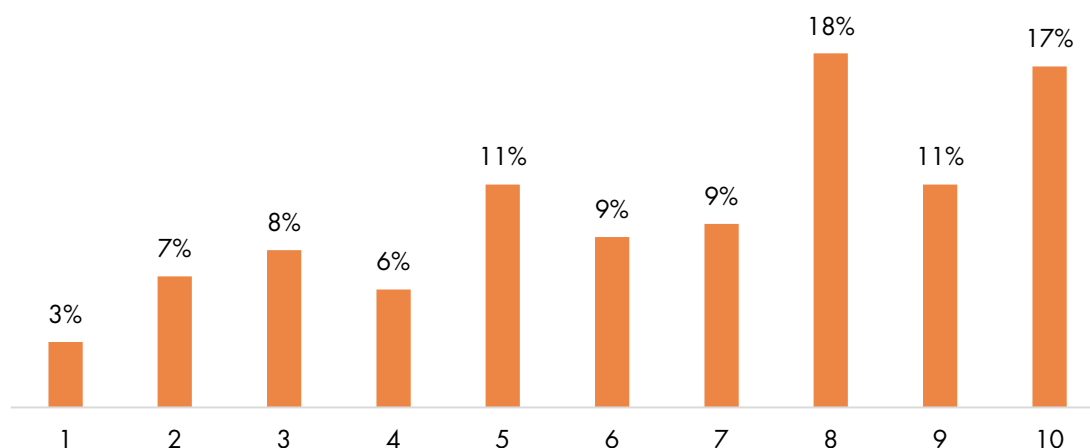


Fuente: DNP con datos del Ministerio de Tecnologías de la Información y las Comunicaciones (2015).

Adicionalmente, las soluciones basadas en datos que identifican las entidades se centran en aspectos relacionados con la disponibilidad de información (caracterización de los grupos de interés, transparencia y difusión). Esto sugiere la baja comprensión de los datos como un activo para la generación de valor y la producción de soluciones de mayor complejidad que el registro histórico o estadístico de eventos pasados.

Así mismo, el 44 % de las entidades del orden nacional expresó que los datos abiertos de otras entidades no son importantes para el funcionamiento propio, según se indica en el Gráfico 16.

Gráfico 16. Percepción de importancia de los datos de otras entidades

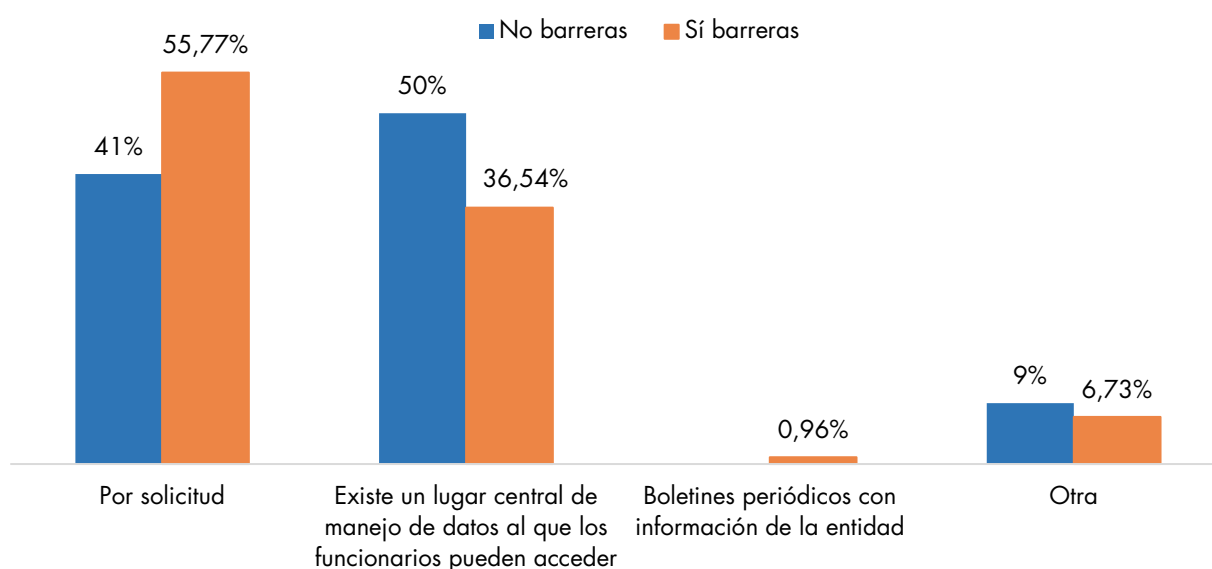


Fuente: DNP (2017).

Nota: por tratarse de una medición de percepción en una escala de 1 (nada) a 10 (completamente), se considera que las entidades que otorgaron una calificación inferior a 7 no perciben que los datos de otras entidades sean importantes.

En la actualidad, el 69 % de las entidades percibe barreras para compartir información internamente, mientras que el 31 % no considera que estas existan (DNP, 2017). El método más común para la compartición al interior de las entidades es la solicitud del interesado (52 %) (DNP, 2017), lo cual impacta la presencia de barreras para su explotación. Esto porque, como se expresa en el Gráfico 17, cuando la compartición es proactiva (existe un lugar central para el manejo de datos), disminuye la percepción de las barreras para acceder a los datos, en comparación a cuando se hace de forma reactiva (por solicitud).

Gráfico 17. Método más común de compartición y percepción de barreras para la explotación de datos

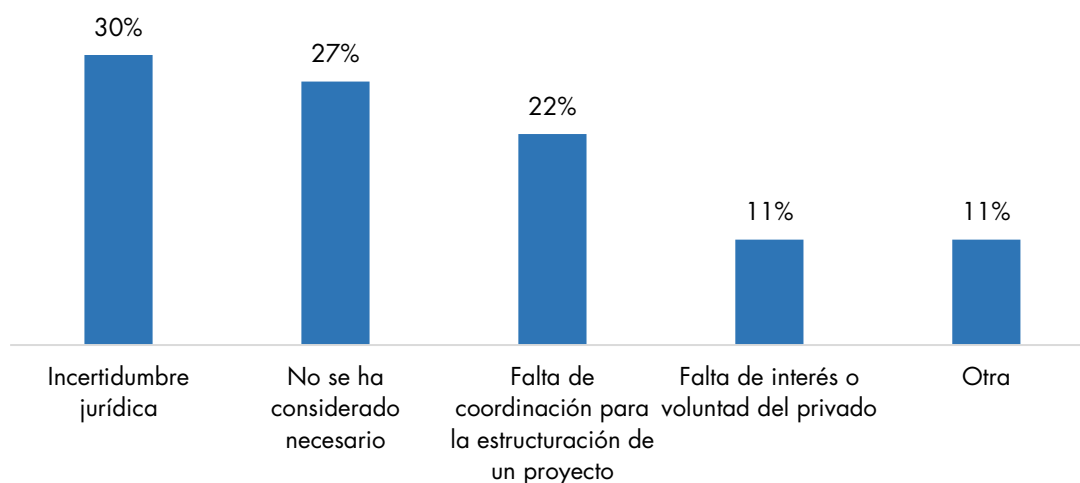


Fuente: DNP (2017).

Adicionalmente, el Banco Mundial (2015) encontró que en el país las obligaciones relacionadas con la publicación y compartición de datos se han desarrollado con “falta de interés y lentitud” por parte de las entidades, lo que afecta la disponibilidad oportuna de los mismos, disminuyendo la posibilidad de que sean aprovechados en la generación de valor. Esta situación persiste porque, aunque el 41 % de las entidades expresa que no existen barreras para la publicación de datos abiertos (DNP, 2017) y el 85 % manifiesta que sus sistemas de información tienen características para la apertura de datos (Departamento Administrativo de la Función Pública, 2017), los niveles de apertura y compartición siguen siendo bajos. Lo cual sugiere que la lentitud y falta de interés expresadas, no han sido superadas.

Por otro lado, el 31 % de las entidades del orden nacional aún percibe que compartir datos con el sector privado no es importante (DNP, 2017). Del total de entidades, el 35 % no lo ha hecho, por las razones expuestas en el Gráfico 18.

Gráfico 18. Barreras para intercambiar datos con el sector privado



Fuente: DNP (2017).

Lo anterior evidencia que los datos aún no se entienden ni gestionan como un activo que permite la generación de valor y persiste la incertidumbre sobre la viabilidad de su aprovechamiento, elementos que sugieren la ausencia de una cultura de datos establecida en las entidades.

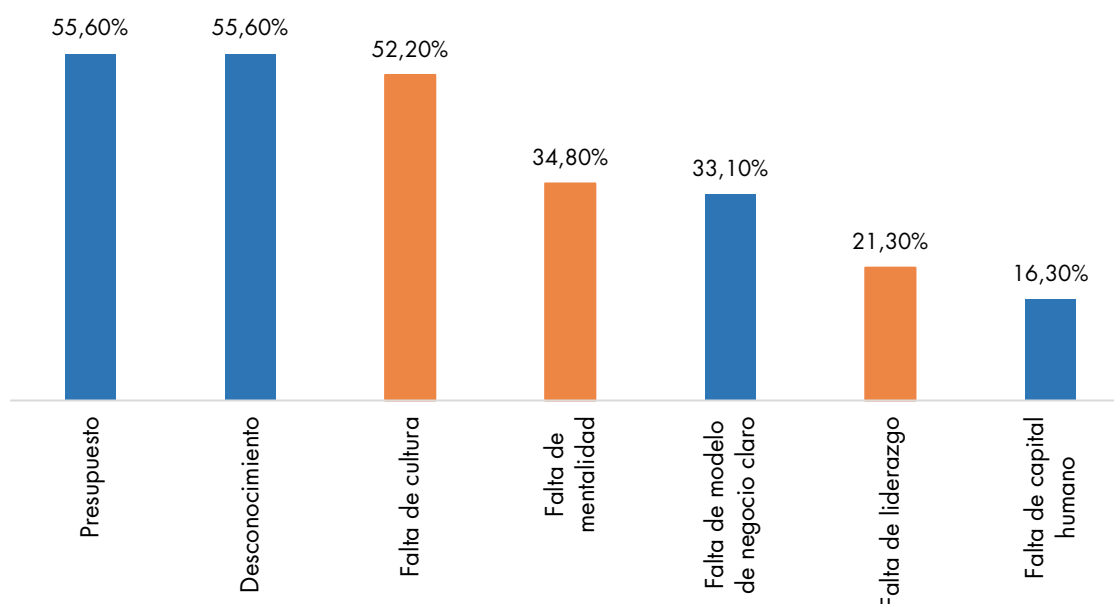
Esto redunda en la baja demanda de datos en las entidades públicas para la generación de valor en su gestión. Así, el 62 % de estas no ha explotado datos para ofrecer o mejorar los servicios¹¹¹ a la ciudadanía¹¹². Adicionalmente, la explotación de datos, en su mayoría, es usada para fines descriptivos (55 %). Solo el 37 % realiza analítica predictiva y el 33 % prescriptiva (DNP, 2017).

¹¹¹ En concordancia con lo dispuesto en los documentos CONPES 3649 *Política Nacional de Servicio al Ciudadano* y 3785 *Política Nacional de Eficiencia Administrativa al Servicio del Ciudadano*, se entiende por servicio a la ciudadanía los servicios y atención que prestan las entidades públicas y que implican un relacionamiento entre el Estado y el ciudadano.

¹¹² El 5,33 % de las entidades manifestó desconocer si se habían explotado datos para ofrecer servicios a la ciudadanía y el 32,67 % restante indicó haberlo hecho. No obstante, esta cifra se reduce a 5,3 % cuando se contrasta con el uso de algoritmos, herramienta necesaria para la explotación de datos.

Por otro lado, de acuerdo con la Encuesta de Opinión Industrial Conjunta¹¹³, los empresarios identifican dentro de las barreras para la transformación digital exitosa la falta de cultura (52,2 % de los empresarios encuestados), la falta de mentalidad (34,8 %) y la falta de liderazgo (21,3 %) (Gráfico 19).

Gráfico 19. Mayores barreras para la transformación digital (2016)



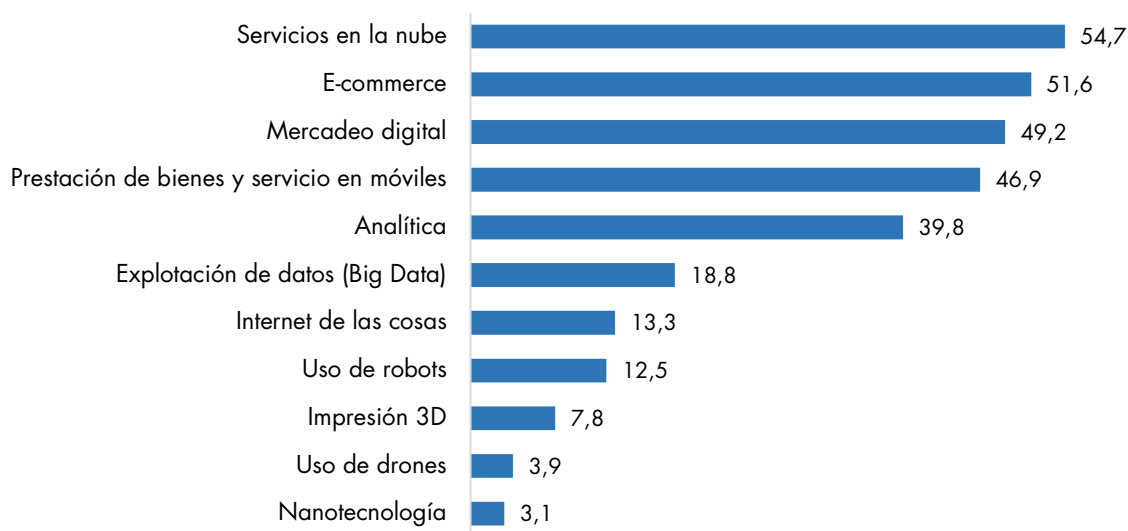
Fuente: DNP con datos de ANDI (2016).

Nota: pregunta con respuestas múltiples.

En concordancia con lo anterior, la encuesta citada incluyó una sección relacionada con la cuarta revolución industrial y la adopción de la digitalización en el país. De acuerdo con lo señalado en el Gráfico 20, la explotación de datos y la analítica aún no son utilizadas masivamente por la industria (ANDI, 2016).

¹¹³ Correspondiente al periodo enero a mayo de 2016. La encuesta de Opinión Industrial Conjunta es realizada por la Asociación Nacional de Empresarios de Colombia (ANDI) y recoge el comportamiento, las percepciones y los principales problemas que afrontan las empresas del sector manufacturero. La muestra se compone por las siguientes agremiaciones: ANDI, Acoplásticos, Andigraf, Acicam, Camacol y la Cámara Colombiana del Libro.

Gráfico 20. Uso de tecnologías en el sector industrial



Fuente: DNP con datos de ANDI (2016).

5. DEFINICIÓN DE LA POLÍTICA

5.1. Objetivo general

Aumentar el aprovechamiento de datos en Colombia, mediante el desarrollo de las condiciones para que sean gestionados como activos para generar valor social y económico.

5.2. Objetivos específicos

OE 1. Masificar la disponibilidad de datos de las entidades públicas que sean digitales accesibles, usables y de calidad.

OE 2. Generar seguridad jurídica para la explotación de datos.

OE 3. Disponer de capital humano para generar valor con los datos.

OE 4. Generar cultura de datos en el país.

5.3. Plan de acción

Las estrategias y acciones de la presente política se regirán bajo los principios que se describen a continuación.

1. Los datos son un activo estratégico que genera valor social y económico, por lo que su explotación debe cumplir una función social, orientada a la maximización de beneficios para todos los actores involucrados.

2. Los derechos de las personas (naturales y jurídicas) constituyen el límite de la explotación de datos, por lo que el marco jurídico y los instrumentos de gobernanza deben facilitar la generación de valor, garantizando todos los derechos. La dignidad humana orientará la definición de los usos socialmente permitidos de los datos para la generación de valor social y económico.
3. Siempre que sea posible, se emplearán mecanismos de colaboración entre el sector público, privado, la academia y la ciudadanía, para maximizar la cooperación entre todos los actores. La articulación entre todos estos guiará las acciones a implementar en la ejecución de la política.
4. Los datos públicos serán definidos como recurso de infraestructura pública, que soporta la economía digital y genera beneficios sociales.
5. Se deben potenciar los casos de éxito como mecanismo para acelerar las transformaciones requeridas. Los resultados de los proyectos ejecutados en la administración pública serán públicos hasta el máximo nivel permitido, de acuerdo con las normas que regulan el tratamiento de los datos según su tipología, para incentivar el aprovechamiento y evitar la duplicidad de esfuerzos.

Por otro lado, es importante mencionar que de acuerdo con la Comisión de Estadísticas Oficiales de la ONU (2017) y la OCDE (2017), la disponibilidad de nuevas y diversas fuentes de datos, unida a la capacidad para explotarlos, es una oportunidad para complementar y fortalecer, pero no para reemplazar, la producción de estadísticas oficiales. Por tanto, adicional a los cinco principios propuestos, la producción de estadísticas oficiales se rige por lo dispuesto por el Sistema Estadístico Nacional (SEN) y lo definido en el Plan Estadístico Nacional 2017 – 2022 (Consejo Asesor Nacional de Estadística, 2017), que se articulará, en lo pertinente, con las acciones que se dispongan en la presente política.

5.3.1. Masificar la disponibilidad de datos públicos digitales accesibles, usables y de calidad

Se enuncian las estrategias y acciones relacionadas con la digitalización, apertura de datos e interoperabilidad que permitirán dar cumplimiento al primer objetivo específico. La implementación de estas acciones generará las condiciones para el acceso, intercambio y uso de los datos dispuestos por las entidades públicas, el sector privado, la sociedad civil y la academia, permitiendo su explotación.

Acelerar la digitalización de servicios, procesos y toma de decisiones en el sector público

Línea de acción 1. Diseñar e implementar la infraestructura de datos

La comprensión de los datos como un activo que genera valor económico y social implica la definición, implementación, mantenimiento y explotación de la infraestructura de datos. Esta corresponde a la organización y disposición de los datos que garantiza su calidad y disponibilidad para el acceso, uso y reutilización por parte de las entidades públicas, privadas, la sociedad civil, la academia y la ciudadanía, de acuerdo con su tipología. Lo anterior, incluye la disposición de mecanismos de acceso ágil, el intercambio y la explotación de los datos. La infraestructura de datos será provista por el Estado, a través de las entidades públicas que generan y recolectan los datos estructurados y no estructurados que la integran, aplicando los estándares y lineamientos para la gestión de los datos, que sean definidos por la entidad competente. A continuación, se describen las acciones requeridas para disponer de la infraestructura de datos.

En primer lugar, el Departamento Nacional de Planeación, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones, publicará, en octubre de 2018, la definición de la infraestructura de datos públicos y las condiciones requeridas para su implementación y aprovechamiento. Esto incluye la garantía del derecho de acceso a la información pública, mantenimiento, necesidades de cooperación público-privada y garantía de no afectación de datos personales, reservados o clasificados, así como el respeto a la propiedad intelectual y la propiedad privada. Estas condiciones incluirán la determinación de mecanismos e incentivos de colaboración público-privada que permitan compartir datos y capacidades analíticas¹¹⁴ por parte de las empresas privadas, la ciudadanía, la academia y la sociedad civil.

En segundo lugar, para unificar las herramientas de gestión de datos en las entidades públicas se propone lo siguiente: dado que los datos generados, recolectados y gestionados por las entidades públicas son el elemento central de la infraestructura de datos, para que estos últimos se encuentren en abundancia y en formato digital es necesario contar con un mecanismo de gestión en las entidades públicas, que brinde certeza respecto de cuáles se generan y recolectan, el medio empleado, su responsable, nivel de publicidad e integración de un registro administrativo. Para ello, en diciembre de 2018, el Ministerio de Tecnologías de la Información y las Comunicaciones, en articulación con el Departamento Administrativo de la Función Pública, el Archivo General de la Nación, el Departamento Nacional de Planeación, la Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República, la Superintendencia de Industria y Comercio, y el Departamento Administrativo Nacional de Estadística, diseñará e implementará una

¹¹⁴ De acuerdo con la OCDE (2017) a pesar del creciente reconocimiento de la relevancia de los datos del sector privado, la ciudadanía y la academia para la solución de problemas públicos, así como del potencial de la colaboración público – privada, aún no se han definido mecanismos claros para la compartición de datos entre estos sectores que atienda a las preocupaciones de propiedad, titularidad, privacidad, entre otras. Esto, pone de presente la necesidad de encontrar soluciones mutuamente satisfactorias para aumentar las posibilidades de colaboración.

herramienta que facilite la administración de datos en las entidades públicas. Esta herramienta será un sistema de información disponible en el Portal de Datos Abiertos del Estado Colombiano para ser diligenciada y gestionada por todos los sujetos obligados y permitirá, a través de metadatos, la identificación de las categorías básicas de información, datos básicos del Estado, el responsable y el titular de los datos.

Esta acción se articulará con las demás políticas que imponen la generación de un instrumento de gestión de información, sea este denominado activos, datos, registros, entre otros¹¹⁵. Cuando se trate de información estadística, la herramienta se articulará con lo previsto en el SEN. De igual forma, esta acción se integrará con el programa de gestión documental de las entidades para facilitar el gobierno de los datos, en coherencia con el cumplimiento de los requisitos de autenticidad, integridad, fiabilidad y disponibilidad de los documentos públicos, así como de su planeación, producción y preservación a largo plazo. La transformación y robustecimiento del instrumento deberá orientarse a la disponibilidad de los datos, metadatos, registros y responsables de los mismos, para aumentar su uso. Lo anterior, removerá las barreras identificadas para la disponibilidad, apertura y compartición, disminuyendo las cargas de las entidades públicas en la gestión de los instrumentos previamente dispuestos.

En tercer lugar, para generar un plan de gestión documental electrónica para las entidades públicas, en el marco del proyecto de modernización de los archivos públicos, el Archivo General de la Nación pondrá a disposición de las entidades públicas, en diciembre de 2021, la hoja de ruta para que estas automaticen de manera progresiva los procesos y procedimientos administrativos en la producción y gestión de documentos nativos electrónicos¹¹⁶ y digitalizados. Esta hoja de ruta incluirá los estándares para que los datos contenidos en los documentos de las entidades públicas puedan descubrirse y usarse fácilmente. Lo anterior, sin detrimento de las normas de conservación y preservación documental. La digitalización de documentos históricos o de vigencias anteriores será priorizada y definida en articulación con las necesidades de las entidades.

Acelerar la apertura de datos públicos y el fortalecimiento de las políticas de datos abiertos del sector público

Línea de acción 2. Materializar la apertura por defecto

Para hacer efectivo el principio de máxima publicidad¹¹⁷ que hace parte del derecho de acceso a la información pública, es necesario materializar la apertura por defecto y

¹¹⁵ Actualmente no existe una definición única y estándar para estos términos que, en ocasiones, se usan como sinónimos o con alcances diferentes, según el contexto.

¹¹⁶ Documentos completamente electrónicos durante todo su ciclo de vida.

¹¹⁷ Según Ley 1712 de 2014, por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones.

diseño de los datos públicos. Esto, mediante la remoción de las barreras que han interferido en su aplicación, así como con la incorporación de la apertura como una actividad rutinaria de las entidades públicas, que incluye la disponibilidad para compartición de todo aquello que no tenga alguna restricción legal. Lo anterior, sin menoscabar las garantías y tratamientos dispuestos en el régimen de protección de datos personales. Para ello, deberán desarrollarse las acciones que se describen a continuación.

En primer lugar, se debe disponer de una vista pública para cuantificar la tasa de publicación de datos. Para tal fin, en diciembre de 2018, el Ministerio de Tecnologías de la Información y las Comunicaciones tendrá una opción de consulta en el Portal de Datos Abiertos del Estado que indique el avance de las entidades en la apertura de sus datos y cuantifique la tasa de crecimiento en publicación de datos públicos de las entidades. Se incluirán datos estructurados, no estructurados y semiestructurados. Lo anterior, permitirá realizar monitoreo constante a los conjuntos de datos disponibles en las entidades y aquellos que son efectivamente publicados, de modo que se superen los desafíos para la publicación de datos que se derivan de la consideración subjetiva de aquello que se decide publicar por las entidades y que ha desincentivado la materialización de la máxima publicidad.

En segundo lugar, se deben establecer estándares de *anonimización*. Para esto, el Archivo General de la Nación, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones, la Superintendencia de Industria y Comercio, el Departamento Administrativo de la Función Pública, el Departamento Nacional de Planeación y el Departamento Administrativo Nacional de Estadística, publicará, en diciembre de 2019, el documento que defina los estándares y lineamientos técnicos que orienten el proceso de *anonimización* de datos personales, así como los criterios para minimizar la recolección de datos personales, cuando estos no sean necesarios en casos concretos. Esto, de acuerdo con las alternativas técnicas disponibles que deben aplicarse a los conjuntos de datos para eliminar los identificadores personales directos e indirectos, de forma tal que sea posible la publicación de mayores conjuntos de datos abiertos y se facilite la compartición entre entidades, garantizando los derechos de los titulares. Lo anterior, incorporará lo definido por el SEN, en lo de su competencia, cuando ello corresponda.

En tercer lugar, se debe asegurar la calidad de los datos públicos. En este sentido el Ministerio de Tecnologías de la Información y las Comunicaciones definirá y adoptará los estándares y lineamientos técnicos, que deberán aplicar las entidades para garantizar la calidad de los datos durante todo su ciclo de vida, así como para verificar de manera automatizada la calidad de estos, siguiendo el criterio de publicidad por diseño y defecto de los datos públicos de las entidades públicas. Igualmente, se establecerá el procedimiento que deben seguir las entidades para su corrección o actualización, en caso de que se encuentren deficiencias en los mismos. El documento con los estándares y lineamientos

técnicos será publicado en junio de 2019 y se articulará con las acciones realizadas en el marco del SEN para garantizar la calidad de la información estadística.

En cuarto lugar, teniendo en cuenta la evolución tecnológica en la materia, para aumentar las condiciones que permiten el uso y la calidad de los datos, en diciembre de 2018, el Ministerio de Tecnologías de la Información y las Comunicaciones publicará la guía con la definición de los estándares que deben implementar las entidades públicas para disponer de datos públicos enlazados que permitan su aprovechamiento y faciliten el descubrimiento de una mayor cantidad de datos. Esta guía incluirá los plazos y condiciones de implementación de este estándar por parte de las entidades públicas.

Fortalecer la interoperabilidad en los sistemas de información del sector público para facilitar el intercambio y re-uso de datos

Línea de acción 3. Reglamentar y fortalecer técnicamente la habilitación general para el intercambio de información entre entidades públicas

Las intervenciones públicas previas han dejado claro el mandato de acceso e intercambio de datos entre entidades públicas, así como la necesidad de avanzar en la interoperabilidad entre estas. No obstante, es necesario disponer de reglas que precisen estos procedimientos, para que en la práctica se haga efectiva la habilitación general¹¹⁸, se suprima el uso de convenios para la compartición de datos y se fortalezca técnicamente el proceso de interoperabilidad. Para ello, deben adelantarse las siguientes actividades.

En primer lugar, para definir los estándares que deben cumplir los registros administrativos, el Ministerio de Tecnologías de la Información y las Comunicaciones, en articulación con el Departamento Administrativo Nacional de Estadística, el Departamento Administrativo de la Función Pública, el Departamento Nacional de Planeación y el Archivo General de la Nación, publicará en julio de 2019 la guía con los estándares que deben cumplir los registros administrativos para garantizar la publicidad por diseño y defecto de los datos públicos de las entidades públicas. La guía definirá el lenguaje usado para la recolección y generación de datos¹¹⁹, formatos estandarizados para su descarga, elementos para la generación de una interfaz de programación de aplicaciones que permita enlazar los registros administrativos y evitar su duplicidad. Los estándares deben incorporar los lineamientos definidos por el SEN para el uso estadístico de los registros administrativos. El uso y aplicación de los estándares en mención permitirá la eliminación de las barreras que han demorado la masificación de la interoperabilidad.

¹¹⁸ Dispuesta como una obligación de los funcionarios públicos en los términos del artículo 159 de la Ley 1753 de 2015, por la cual se expide el Plan Nacional de Desarrollo 2014 - 2018 *Todos por un nuevo país*.

¹¹⁹ Interoperabilidad semántica.

En segundo lugar, se debe normalizar el procedimiento para el acceso a la información entre entidades. Para esto, el Departamento Nacional de Planeación, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones, la Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República y el Departamento Administrativo de la Función Pública, propondrá, en junio de 2019, el texto de decreto reglamentario en el que se definirán las reglas claras que hagan expedito el intercambio de los datos al interior del sector público, a efectos de disponer de un procedimiento que remueva barreras de tipo operativo. Así, además del mandato de permitir el acceso a estos, se contará con la definición de las condiciones sencillas y precisas para remover los obstáculos que han afectado su efectividad; concretamente, la dependencia en la voluntad institucional, el establecimiento de cargas burocráticas, la ausencia de límites de tiempo para la entrega de los datos, la indefinición de los medios para su entrega y la tendencia a suscribir convenios y cláusulas de confidencialidad sin considerar la taxonomía de los datos.

En tercer lugar, para atender los retos identificados en materia de interoperabilidad, concretamente, la ausencia de estandarización y las asimetrías en las capacidades técnicas de cada entidad, el Ministerio de Tecnologías de la Información y las Comunicaciones, en articulación con el Departamento Administrativo Nacional de Estadística, establecerá, entre junio de 2018 y julio de 2019, la mejor alternativa de interoperabilidad que será implementada por las entidades públicas, de acuerdo con criterios técnicos y económicos que faciliten su adopción (considerando entre otras posibilidades la interoperabilidad como servicio, unidad semántica, disponibilidad de una plataforma centralizada o uso de tecnologías recientes como *blockchain*). Para ello, realizará un estudio que incluya la determinación de los tiempos de implementación de dicha alternativa por parte de las entidades públicas, en concordancia con lo dispuesto en el Decreto 1413 de 2017¹²⁰. Lo anterior, de modo que se superen los desafíos para la masificación de la interoperabilidad que han surgido de la diversidad tecnológica de las entidades públicas. La definición e implementación de la alternativa de interoperabilidad se articulará con los lineamientos que sean definidos en desarrollo de la mesa de trabajo de interoperabilidad liderada por el

¹²⁰ Por el cual se adiciona el Título 17 a la Parte 2 del Libro 2 del Decreto Único Reglamentario del sector de las TIC, Decreto 1078 de 2015, para reglamentarse parcialmente el Capítulo IV del Título III de la Ley 1437 de 2011 y el artículo 45 de la Ley 1753 de 2015, estableciendo lineamientos generales en el uso y operación de los servicios ciudadanos digitales. Dentro de los que se prevé la implementación del servicio de interoperabilidad.

Departamento Administrativo Nacional de Estadística en cumplimiento de lo definido en el Plan Estadístico Nacional (PEN)¹²¹ 2017-2022.

En cuarto lugar, el Ministerio de Tecnologías de la Información y las Comunicaciones, en articulación con el Departamento Nacional de Planeación, realizará, entre junio de 2018 y julio de 2019, un estudio que permita establecer el nivel de alistamiento y ajustes requeridos para implementar servicios basados en *Distributed Ledger Technology* (DLT)¹²² en el sector público, así como las necesidades de política pública para su aprovechamiento y posterior implementación. Lo anterior, teniendo en cuenta que esta tecnología está siendo adoptada por los gobiernos del mundo para mejorar la prestación de servicios y aumentar la eficiencia en las transacciones. Por ello, en el marco de la definición e implementación de la presente política, tiene el potencial de masificar la digitalización y el acceso a los datos entre entidades públicas, porque disminuye los costos asociados a su gestión y aumenta la seguridad y trazabilidad de las transacciones.

5.3.2. Generar seguridad jurídica para la explotación de datos

A continuación, se enuncian las estrategias y acciones para la consolidación de un entorno de confianza y seguridad, relacionadas con la explotación de datos, que permitan maximizar el bienestar económico y social, y proteger los derechos involucrados, para cumplir el segundo objetivo específico. La realización de estas acciones materializará el reconocimiento de los datos como un activo, porque eliminará las barreras jurídicas para su aprovechamiento. Adicionalmente, se adaptará el marco jurídico a los retos que plantea la constante evolución de las capacidades analíticas para el aprovechamiento de datos.

Consolidar y articular el marco jurídico relacionado con la generación, recolección, compartición, agregación y aprovechamiento de datos

Línea de acción 4. Reconocimiento jurídico de los datos como activo

Actualmente los datos son un factor de producción y esto los convierte en un activo desde el punto de vista económico. El reconocimiento jurídico de esta situación implica que se generen las condiciones normativas para habilitar y facilitar el aprovechamiento de los datos. Esto incluye la simplificación del marco jurídico aplicable al aprovechamiento de datos, la remoción de la incertidumbre respecto de la recolección, uso y compartición de este activo, así como el fortalecimiento de la protección de los derechos. La presente línea de acción describe las acciones requeridas para alcanzar dichos propósitos.

¹²¹ Hoja de ruta con las estrategias y acciones orientadas a fortalecer la producción y difusión de las estadísticas oficiales del país.

¹²² Conjunto de registros que se encuentran en una base de datos distribuida en red, los cuales son encriptados mediante algoritmos que garantizan la seguridad y verificabilidad de las transacciones. Estas se almacenan en bloques en orden cronológico.

En primer lugar, el reconocimiento de los datos como un activo implica que los mismos deban gestionarse, planearse y explotarse para generar valor. Para ello, el Departamento Nacional de Planeación, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones, publicará en junio de 2019, las condiciones para crear una instancia con capacidad de acción que materialice un nuevo marco de gobernanza para los datos, en articulación con instancias y autoridades existentes, como el Consejo Asesor Nacional de Estadística¹²³ y otras autoridades con competencias específicas sectoriales, sin afectar las competencias legales de cada una de las instancias y autoridades existentes. La instancia propuesta debe superar los esfuerzos anteriores orientados a la creación de comisiones que carecían de herramientas para su operación¹²⁴, así como trascender las actuales capacidades sectoriales que se focalizan, principalmente, en la generación de lineamientos centrados en la gestión de información. En este sentido, dentro de las condiciones de creación de esta instancia, se determinarán los requerimientos para su posterior transformación en un organismo con capacidades robustas para impulsar la transformación digital del sector público, desde una perspectiva transversal, que supere las limitaciones de la intervención sectorial focalizada y dependa del más alto nivel del Gobierno nacional. Las conclusiones del estudio que determine las condiciones de creación de esta instancia serán presentadas a la autoridad competente para lograr su materialización.

En segundo lugar, para simplificar las clasificaciones legales de los datos, la Delegatura para la protección de datos personales de la Superintendencia de Industria y Comercio, en articulación con el Departamento Nacional de Planeación y la Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República, en diciembre de 2019, diseñará e implementará un esquema que simplifique la comprensión y aplicación de las clasificaciones legales de los datos. Así mismo, la delegatura determinará la necesidad de realizar ajustes normativos para la unificación de conceptos y clasificaciones. Así, se removerá la limitante al aprovechamiento de datos que es generada por la confusión respecto del nivel de publicidad y tratamiento que debe darse a los datos y, al mismo tiempo, se fortalecerá la protección de los datos personales al evitar errores en su clasificación y tratamiento. Esto porque la producción y evolución normativa aplicable a los datos, generada en momentos específicos con objetivos particulares, ha dado lugar a la proliferación de clasificaciones de la información y los datos, los cuales dan lugar a inseguridad jurídica¹²⁵.

¹²³ Órgano de carácter consultivo, creado por el artículo 160, numeral 5, parágrafo tercero de la Ley 1753 de 2015 para promover y facilitar la coordinación del Sistema Estadístico Nacional (SEN).

¹²⁴ Presupuesto, capacidad de proferir mandatos vinculantes y disponibilidad de personal dedicado.

¹²⁵ Por ejemplo, la percepción de que todos los datos personales son privados, o toda la información del sector defensa es reservada, así como la confusión común respecto de la publicidad del número de identificación de un ciudadano.

En tercer lugar, el Departamento Nacional de Planeación, en articulación con la Superintendencia de Industria y Comercio, el Ministerio de Tecnologías de la Información y las Comunicaciones, la Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República y el Departamento Administrativo de la Función Pública, propondrá, en junio de 2020, el texto normativo que permita articular el marco jurídico aplicable a la explotación de datos, bajo un principio de simplificación y eliminación de barreras, siguiendo las recomendaciones respecto del análisis de impacto normativo previo a la generación de normas. Esto con el propósito de disminuir la dispersión normativa y armonizar la pluralidad de disposiciones jurídicas aplicables en la materia.

En cuarto lugar, el Departamento Administrativo de la Función Pública, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones, el Departamento Nacional de Planeación, la Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República y el Archivo General de la Nación, en junio de 2020, definirá una propuesta de articulación del modelo de gestión de datos¹²⁶ con las políticas de gestión y desempeño institucional, en el marco del modelo integrado de planeación y gestión, que será implementado por las entidades públicas. Dicha propuesta será presentada ante el Consejo para la Gestión y el Desempeño Institucional para su aprobación. De esta manera, se propiciarán cambios para la modernización de la gestión pública impulsada por datos. Los regímenes identificados como antecedentes del aprovechamiento de datos implican para las entidades públicas múltiples tareas y seguimientos mediante comités internos, que se simplifican y articulan de conformidad con lo dispuesto en el Decreto 1499 de 2017¹²⁷. Por lo anterior, los datos y su gestión son elemento común a estos regímenes y a las políticas de gestión y desempeño institucional, por tanto, impactan de manera transversal el cumplimiento de las funciones institucionales, enfocándolas a la generación de valor impulsado por datos.

En quinto lugar, la Agencia Nacional de Contratación Pública - Colombia Compra Eficiente actualizará los manuales, guías y documentos emitidos para que las entidades públicas incorporen los datos en los procesos de contratación. En diciembre de 2018 esta entidad publicará la guía de competencia en la compra pública, la guía de elaboración de estudios del sector y la circular única, con la inclusión de los criterios para que los datos se reconozcan explícitamente en la gestión contractual y para que la apertura de datos usables, accesibles y de calidad sea incorporada como criterio de selección y supervisión de la ejecución. De acuerdo con el valor que puede generarse con los datos producidos y recolectados en el desarrollo de las actividades que son mediadas por alguna forma de

¹²⁶ Que será definido por el Departamento Nacional de Planeación y el Ministerio de Tecnologías de la Información y las Comunicaciones en diciembre de 2019 (acción 4.6 del PAS (Anexo A)).

¹²⁷ Por medio del cual se modifica el Decreto 1083 de 2015, Decreto Único Reglamentario del Sector Función Pública, en lo relacionado con el Sistema de Gestión establecido en el artículo 133 de la Ley 1753 de 2015.

contratación financiada con recursos públicos, es necesario superar situaciones en las que este aspecto es omitido y se crean dificultades para el acceso a los datos, los cuales a pesar de su generación mediante el uso de recursos públicos reciben tratamiento de propiedad privada, desconociendo su potencial para generar valor social y económico por parte de las entidades públicas.

Armonizar el marco jurídico con el avance tecnológico y teniendo en cuenta los riesgos relacionados con el aprovechamiento de datos

Línea de acción 5. Solventar el déficit de protección legal y ético

La constante evolución de las capacidades analíticas para el aprovechamiento de datos da lugar a situaciones sociales novedosas que no habían sido previstas en el marco jurídico, por tanto, carecen de un consenso social respecto de las acciones que son exigibles en este contexto. A continuación, se describen las acciones para que se disponga de un marco jurídico acorde con las necesidades actuales, se fortalezca la protección de los datos personales de los ciudadanos y se establezcan límites éticos que orienten la conducta de las personas involucradas en las situaciones que surjan.

En primer lugar, se deben definir lineamientos que deben aplicar las entidades públicas para el adecuado tratamiento de los datos personales. Para este fin, la Delegatura para la protección de datos personales de la Superintendencia de Industria y Comercio, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones y el Departamento Nacional de Planeación, publicará, en diciembre de 2019, los lineamientos que deben aplicar las entidades públicas para el adecuado tratamiento de los datos personales de los ciudadanos, alineados con el principio de responsabilidad demostrada¹²⁸, desarrollado en la reglamentación e incorporando la privacidad por diseño y defecto de los datos personales. De esta manera, se busca superar el cumplimiento de mínimos y convertir la gestión de los datos en una actividad rutinaria que garantice la protección de los datos personales de los ciudadanos.

En segundo lugar, se deben determinar las necesidades de actualización del marco jurídico. Para esto, la Delegatura para la protección de datos personales de la Superintendencia de Industria y Comercio, en articulación con el Departamento Nacional de Planeación y el Ministerio de Tecnologías de la Información y las Comunicaciones, propondrá, en junio de 2019, las recomendaciones de actualización del marco jurídico para la protección de los ciudadanos, en las que se identifiquen los requerimientos y desarrollos requeridos en aspectos como los principios de autorización, información,

¹²⁸ Según el cual los responsables del tratamiento de datos deben ser capaces de demostrar, por solicitud de la autoridad competente, que han definido procedimientos y controles para garantizar el cumplimiento de las exigencias en materia de tratamiento de datos. Este principio es actualmente obligatorio para el caso de datos personales.

necesidad y finalidad, entre otros, que permitan mitigar los riesgos asociados a la explotación de datos.

Como resultado de lo anterior, la Delegatura para la protección de datos personales de la Superintendencia de Industria y Comercio propondrá un diagnóstico que definirá la instancia competente¹²⁹ a la que serán presentadas las propuestas de implementación de las actualizaciones requeridas. Esto, de acuerdo con el escenario actual de aumento de las capacidades técnicas para el aprovechamiento de datos y la constante *datificación* de la vida diaria, surgen riesgos que no han sido abordadas por el marco jurídico, o frente a las cuales este resulta insuficiente, como ocurre con la necesidad de replantear los mecanismos y alcances de la autorización otorgada por el titular para fortalecer su protección mediante herramientas jurídicas que se centren en el uso permitido de los datos.

Así mismo, deberá analizarse lo relacionado con la portabilidad y derechos patrimoniales de los titulares de datos personales recolectados por los particulares que ofrecen servicios.

En tercer lugar, el Departamento Nacional de Planeación, en articulación con el Departamento Administrativo de la Función Pública, liderará la realización de mesas de trabajo entre junio de 2018 y junio de 2019, en las que se identifiquen e incorporen buenas prácticas y esquemas de autorregulación exitosos para definir el marco ético para la explotación de datos en el país. Estas mesas de trabajo contarán con la participación de representantes del sector público, privado, la academia, la sociedad civil y los ciudadanos. El consenso ético se formalizará en un compromiso de todos los actores mencionados para la implementación de prácticas de autorregulación, de modo que, además de las previsiones jurídicas se cuente con un consenso social respecto del uso de los datos para la generación de valor económico y social, sin afectar las garantías de los ciudadanos o generar algún detrimento en los derechos reconocidos y protegidos por el marco jurídico.

En cuarto lugar, se definirán los criterios que deben cumplir los proyectos de explotación de datos de las entidades públicas para garantizar su transparencia. Para esto, el Departamento Nacional de Planeación, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones, propondrá, en junio de 2019, el documento con los elementos que deben incorporar los proyectos de explotación de datos implementados por las entidades públicas para garantizar que estos sean transparentes, explicables y auditables. Así mismo, analizará la pertinencia e impacto de elegir un estándar o modelo para la aplicación de estos criterios. Lo anterior, con el fin de garantizar que no se tomen decisiones que contengan sesgos, se revele información privada o, en general, se afecte algún derecho.

¹²⁹ De acuerdo con el tipo de norma o modificación al marco legal que se requiera.

Línea de acción 6. Definir condiciones para el intercambio de datos entre los sectores público y privado

La seguridad jurídica para la explotación de datos incluye la determinación de los mecanismos para que el sector público acceda a datos de propiedad privada, relevantes para el desarrollo de sus funciones, así como los eventos en los que es posible recurrir a colaboraciones público-privadas para el aprovechamiento de este activo en procura del bienestar económico y social. La presente línea de acción determina las actividades requeridas para el intercambio público-privado de datos.

En este sentido, el Departamento Nacional de Planeación, en articulación con la Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República, el Ministerio de Tecnologías de la Información y las Comunicaciones y el Departamento Administrativo Nacional de Estadística, propondrá, en diciembre de 2020, un mecanismo jurídico que remueva las barreras y habilite el acceso e intercambio de datos entre el sector público y privado. Este mecanismo deberá respetar las normas vigentes en materia de producción estadística, propiedad privada, propiedad intelectual, garantía de protección de los datos personales, la privacidad, la intimidad, así como la proporcionalidad de estos con el derecho de acceso a la información pública.

Para lo anterior, deberán explorarse alternativas como las licencias *freemium*¹³⁰, la generación de incentivos, los acuerdos de colaboración y los espacios de experimentación (*sandbox*) en los que se identifiquen mejores prácticas, entre otros. Lo anterior, sin perjuicio de la aplicación de las normas vigentes en materia de producción, acceso e intercambio de información estadística. De acuerdo con el resultado del estudio realizado para determinar el mecanismo jurídico que remueva las barreras y habilite el acceso e intercambio de datos entre el sector público y privado a aplicar, el Departamento Nacional de Planeación presentará las recomendaciones ante la entidad competente para su implementación.

Además, el Departamento Nacional de Planeación elaborará el documento con la identificación de las necesidades jurídicas requeridas para que las entidades interesadas lleven a cabo contratos de Asociación Público-Privada que permitan la explotación de datos, la promoción de la innovación basada en este recurso, así como el aprovechamiento de la infraestructura de datos y su mantenimiento. Este documento estará disponible en junio de 2019. Lo anterior, con el fin de determinar las condiciones y viabilidad para implementar proyectos de explotación de datos de mayor complejidad, en los que se requieren inversiones importantes que implican mayor riesgo porque dependen de la experimentación

¹³⁰ Tipo de licencia que ofrece servicios básicos de manera gratuita y requiere del pago para el acceso a funcionalidades adicionales o valor agregado respecto del mismo servicio. La palabra es una contracción de los términos gratis (*free*, en inglés) y *premium*.

(el ensayo y error) para la generación de los resultados esperados. Adicionalmente, en los proyectos de explotación de datos más intensivos en conocimiento, que implica mayor incertidumbre, el código desarrollado para generar productos con los datos representa el activo más valioso y requiere talento altamente especializado.

5.3.3. Disponer de capital humano para generar valor con los datos

El capital humano es un elemento habilitador de la explotación de datos que permite la oferta de bienes y servicios basados en datos, y es requerido para la materialización del potencial de este activo. Al respecto, se describen las estrategias y acciones para aumentar el capital humano a lo largo de la cadena de valor, orientadas a cumplir el tercer objetivo específico. Su implementación dará lugar a la oferta de talento humano con habilidades y conocimientos suficientes para crear bienes, servicios y productos basados en datos que aumenten la eficiencia en la gestión pública y dinamicen la economía.

Armonización de la demanda y la oferta de capital humano para aumentar la fuerza laboral y los perfiles para manejo y análisis de datos

Línea de acción 7. Nivelar el alistamiento para la explotación de datos en las entidades públicas

La generación de capacidades en las entidades públicas consiste en la provisión de las condiciones mínimas para fortalecer el capital humano requerido para el aprovechamiento de datos, mediante la generación de espacios, incentivos e instrumentos colaborativos que permitan contar con las capacidades requeridas en todos los niveles del empleo público para que estén acordes con los retos que plantea el avance en las capacidades para el aprovechamiento de datos y la creciente demanda de recursos humanos en la materia. Las estrategias y acciones descritas a continuación permitirán la implementación de intervenciones que fortalezcan el capital humano.

En primer lugar, el Ministerio de Tecnologías de la Información y las Comunicaciones, en articulación con el Departamento Nacional de Planeación, el Departamento Administrativo Nacional de Estadística, el Archivo General de la Nación y Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República, liderará y realizará, entre junio de 2018 y diciembre de 2021, espacios de entrenamiento en aprovechamiento de datos, para transferir conocimiento para actualizar y fortalecer el capital humano del ecosistema de datos. Lo anterior se articulará con las acciones del programa de innovación, aprendizaje y gestión del conocimiento definido por el SEN.

En segundo lugar, el Departamento Administrativo de la Función Pública realizará acompañamiento técnico, entre enero de 2019 y diciembre de 2020, para promover que, en el marco del Plan Nacional de Capacitación, las entidades públicas identifiquen e incluyan en sus componentes de capacitación las temáticas relacionadas con la explotación

de datos en las vigencias en mención, para garantizar su apropiación e incorporación en vigencias posteriores, de acuerdo con la evolución de las necesidades del capital humano.

En tercer lugar, el Ministerio de Tecnologías de la Información y las Comunicaciones, en articulación con el Departamento Nacional de Planeación, pondrá a disposición de las entidades públicas, en diciembre de 2021, un espacio colaborativo (*data sandbox*) en el que los conjuntos de datos utilizados y generados durante los proyectos adelantados, así como la documentación del proceso y sus resultados se publiquen en un ambiente de trabajo que permita su aprovechamiento público. Este espacio debe permitir, igualmente, la prueba de estándares y la colaboración de actores públicos, privados, la academia y la sociedad civil. Adicionalmente, este espacio facilitará la coordinación entre entidades para aunar esfuerzos en la realización de proyectos de analítica de acuerdo con sus capacidades y el interés en asuntos comunes, que permitan dinamizar la realización de proyectos de analítica e implantar capacidades institucionales.

Línea de acción 8. Medición de la brecha de capital humano y actualización de competencias

La masificación del aprovechamiento de datos en todos los sectores y la transición hacia una economía más intensiva en conocimiento aumenta la demanda de personal que pueda desempeñarse con solvencia en este contexto. Es decir, con educación y habilidades relacionadas con la explotación de este activo. Particularmente, matemáticas, estadística, aprendizaje de máquinas y ciencia de los datos. Al respecto, la presente línea de acción describe las actividades que se implementarán para cuantificar y cerrar la brecha de capital humano en explotación de datos.

En primer lugar, el Ministerio de Tecnologías de la Información y las Comunicaciones, entre junio de 2018 y diciembre de 2021, realizará un estudio que será actualizado en 2021 para identificar brechas de capital humano entorno a las competencias para la explotación de datos. El estudio permitirá identificar las necesidades del país en materia de preparación del talento humano para los cambios que impone el avance en el mediano plazo de la digitalización y la automatización creciente de los procesos y la toma de decisiones. Esto porque, a pesar de la evidencia cualitativa sobre las carencias y dificultades para la vinculación de personal con conocimientos y competencias específicas para el aprovechamiento de datos, actualmente no se dispone de una medición de las diferencias existentes entre la demanda laboral de perfiles relacionados con la explotación de datos y su disponibilidad.

En segundo lugar, el Ministerio de Tecnologías de la Información y las Comunicaciones, entre junio de 2018 y diciembre de 2021, diseñará e implementará acciones que activen la demanda de formación en materia de explotación de datos, entre las que se destacan: la realización de espacios de *networking*, ruedas de negocio, la

implementación de una red de especialistas en datos, entre otros. De este modo, se incentivará desde la demanda la generación de oferta en competencias para la captura, procesamiento, análisis y explotación de datos.

En tercer lugar, el Ministerio de Educación Nacional, en articulación con el Departamento Administrativo de la Función Pública, en septiembre de 2019 publicará el documento con las competencias necesarias para la explotación de datos que deben incorporarse en la metodología de diseño y desarrollo del Marco Nacional de Cualificaciones, como referente de la educación media y superior, para obtener el desarrollo de las habilidades, competencias y conocimientos que demandan el uso y comprensión de datos, así como la toma de decisiones basadas en estos.

Línea de acción 9. Promover el emprendimiento de bienes y servicios basados en la explotación de datos

La masificación del aprovechamiento de datos hace necesario que se impulsen los emprendimientos para la oferta de bienes y servicios basados en datos, los cuales materializan el potencial para el crecimiento económico que representa este activo, como centro de la economía digital. Para ello, el Ministerio de Comercio, Industria y Turismo, a través de la Unidad de Gestión de Crecimiento Empresarial -iNNpulsa Colombia de la Fiduciaria de Comercio Exterior S.A definirá, en julio de 2018, una estrategia de negocios que estimule la creación de nuevas empresas, o nuevas líneas de negocio en las existentes, orientada a incentivar el aprovechamiento de datos y la analítica, para su consolidación transversal en el aparato productivo. La definición de la estrategia incluirá la hoja de ruta para su implementación.

5.3.4. Generar cultura de datos en el país

La cultura de datos es un elemento habilitador del consumo de estos, así como de los productos, bienes, servicios y procesos derivados de este activo. Para el funcionamiento de la demanda de datos existen dos actores principales que influyen en la comprensión de estos como activo: los tomadores de decisiones en las organizaciones y los ciudadanos. La materialización del valor potencial que representan los datos requiere intervenciones específicas. A continuación, se describen las líneas de acción y actividades que permitirán el cumplimiento del cuarto objetivo específico, porque generarán las condiciones para masificar la demanda de soluciones basadas en datos.

Creación de incentivos y condiciones óptimas que aumenten el aprovechamiento de los datos

Línea de acción 10. Generar los mecanismos para materializar el valor de los datos en las entidades de la administración pública

Serán elementos centrales para incentivar la cultura de datos en las entidades públicas: (i) el índice de preparación para la explotación de datos para identificar las principales brechas y necesidades de intervención; (ii) el índice de valor potencial de la explotación de datos en las entidades públicas; y (iii) el modelo de estimación de retorno, que cuantificará los resultados esperados por cada entidad para incentivar la realización de los ajustes requeridos, esto es, disponer de un modelo de negocio que identifique las principales inversiones a realizar para el aprovechamiento de datos. Estos instrumentos, junto con los mecanismos para canalizar las intervenciones requeridas, permitirán enfocar efectivamente la gestión de las entidades en la generación de valor, es decir, en la optimización de recursos y la toma de mejores decisiones. La presente línea de acción busca disponer las herramientas para identificar los resultados que pueden obtenerse, hacer seguimiento a su consecución y, por lo tanto, alinear los incentivos de las entidades públicas hacia este propósito.

En primer lugar, para dinamizar el aprovechamiento de datos en las entidades públicas y obtener los beneficios sociales y económicos que esto supone, en junio de 2020, el Departamento Nacional de Planeación publicará la definición y metodología de aplicación del índice de valor potencial de la explotación de datos, que permitirá identificar: (i) aquellos problemas públicos que deban priorizarse para su atención con soluciones basadas en analítica de datos, y (ii) los servicios prestados por las entidades públicas cuya transformación con el uso de analítica de datos representa mayor impacto. Esta orientación a la obtención de resultados permite superar la tendencia al cumplimiento de mínimos, que obstaculiza la profundización de las transformaciones requeridas en la gestión pública.

En segundo lugar, el Ministerio de Comercio, Industria y Turismo, a través de la Unidad de Gestión de Crecimiento Empresarial -iNNpulsa Colombia de la Fiduciaria de Comercio Exterior S.A adelantará un estudio cuyo resultado, disponible en septiembre de 2018, permitirá estimar el impacto económico y social que podrá alcanzarse con la implementación de la política de explotación de datos en el país, así como el costo de oportunidad de no hacerlo. Lo anterior permitirá la toma de decisiones en materia de estímulo de la demanda de los bienes y servicios basados en datos.

En tercer lugar, en diciembre de 2018, el Departamento Nacional de Planeación publicará la metodología de formulación e implementación de proyectos de aprovechamiento de datos, que oriente a las entidades para responder oportunamente a los retos que estos proyectos imponen y permita la gestión del conocimiento al interior de cada entidad. Esta debe incluir, entre otros aspectos: identificación y uso de datos de terceros, compartición de datos con terceros, uso responsable de la explotación de datos. Lo anterior servirá de referencia en la definición y ejecución de los proyectos, teniendo en cuenta que las particularidades de la explotación de datos hacen insuficientes las herramientas existentes en materia de gestión de proyectos. En este sentido, la metodología

considerará la necesidad de generar espacios favorables a la innovación, es decir, tolerancia al ensayo y error, que permita a las entidades reorientar oportunamente los proyectos hacia la obtención de resultados¹³¹.

En cuarto lugar, el Departamento Nacional de Planeación, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones, publicará en diciembre de 2019 el documento con la definición y actualización de la metodología de aplicación del índice de preparación para la explotación de datos. Este índice permitirá cuantificar las necesidades institucionales para el aprovechamiento de datos (*i.e.* disponibilidad de datos, capital humano, cultura de datos y marco organizacional), medir los avances en su resolución e identificar casos de éxito.

En quinto lugar, el Ministerio de Tecnologías de la Información y las Comunicaciones definirá e implementará entre junio de 2018 y diciembre de 2020 un plan para que las entidades públicas dispongan de las herramientas e infraestructura para capturar, almacenar, procesar y compartir datos, así como las condiciones adecuadas para solventar los requerimientos de almacenamiento de manera sostenible. Este plan se construirá de acuerdo con lo identificado en el índice de preparación para la explotación de datos.

En sexto lugar, el Departamento Nacional de Planeación publicará, en diciembre de 2019, el modelo de implementación de la explotación de datos en las entidades públicas, que será la herramienta para cuantificar las alternativas que optimicen los recursos públicos, permitan estimar el retorno de la inversión (de acuerdo con factores como el tamaño, presupuesto, capacidades disponibles, cantidad y características de los datos recolectados, generados y usados, entre otros). Lo anterior permitirá superar las barreras que representan la baja voluntad o ausencia de priorización para la implementación de los cambios necesarios para la generación de valor con los datos, incluyendo aquellos que se derivan de mandatos previos a la presente política, que no han sido atendidos con suficiencia.

En séptimo lugar, entre junio de 2018 y diciembre de 2019, el Departamento Nacional de Planeación creará un grupo de analítica de datos e implementará proyectos piloto en colaboración armónica con las entidades públicas que lo requieran. Lo anterior incentivará la transferencia de conocimiento y permitirá la sostenibilidad mediante la apropiación interna de la explotación de datos en cada entidad pública. Este grupo especializado generará conocimiento que permitirá fortalecer el desarrollo de las soluciones basadas en datos y aumentar la eficiencia en la gestión pública.

Línea de acción 11. Institucionalizar la explotación de datos en la toma de decisiones públicas

¹³¹ Bajo el concepto de fallar a tiempo y barato, es decir, identificar prontamente las falencias o debilidades que puedan existir para ajustarlas oportunamente y obtener resultados.

Para la adopción del aprovechamiento de datos como herramienta de generación de política pública y gestión estatal, se definen a continuación las acciones que permitirán su uso en los procesos institucionales transversales, como mecanismo de transmisión de la analítica a la toma de decisiones en toda la administración pública.

En primer lugar, el Departamento Nacional de Planeación actualizará, en diciembre de 2018, la metodología para la formulación de proyectos, de modo que la identificación de situaciones que requieren de la intervención pública y que dan lugar a la formulación, aprobación, ejecución y seguimiento de proyectos financiados con recursos públicos, incluya el aprovechamiento de datos en los módulos de la Plataforma Integrada de Inversión Pública (PIIP). Mediante el aprovechamiento de los datos disponibles en esta herramienta se permitirá a las entidades tomar mejores decisiones para la orientación de la inversión pública, que logren su eficiencia.

En segundo lugar, la Secretaría de Transparencia del Departamento Administrativo de la Presidencia de la República, en articulación con el Departamento Administrativo de la Función Pública y el Departamento Nacional de Planeación, actualizará y publicará en julio de 2019 la metodología de la elaboración del Plan Anticorrupción y de Atención al Ciudadano para incluir el aprovechamiento de datos en la lucha contra la corrupción. Incorporar el aprovechamiento de datos dentro de la elaboración de esta herramienta permitirá su inclusión en los ejercicios de planeación, las estrategias encaminadas al fomento de la integridad, la participación ciudadana, la transparencia y eficiencia en el uso de los recursos físicos, financieros, tecnológicos y de talento humano. En este sentido, las decisiones públicas resultarán de la explotación de datos para generar valor a la ciudadanía.

En tercer lugar, entre mayo de 2018 y diciembre de 2019, la Agencia Nacional de Defensa Jurídica del Estado realizará acompañamiento a las entidades públicas, mediante la realización de talleres y la emisión de guías, que incentiven la explotación de los datos contenidos en eKogui¹³², herramienta informática dispuesta por la dicha entidad para gestionar la información de los casos y procesos judiciales de las entidades públicas y privadas que administren recursos públicos. Esta acción permitirá que los datos contenidos en esta herramienta se aprovechen en el fortalecimiento de las políticas y estrategias definidas en el marco del ciclo de defensa jurídica de cada entidad pública.

Línea de acción 12. Definir, implementar y dinamizar el funcionamiento del mercado de datos en Colombia

¹³² Sistema Único de Gestión e Información Litigiosa del Estado Colombiano. Ver: www.ekogui.gov.co

El reconocimiento de los datos como un activo no rival y de propósito general, permite afirmar que los mismos son un bien susceptible de transarse en el mercado¹³³, situación que ha llevado al planteamiento del término mercado de datos. No obstante, esas mismas características de los datos dificultan su valoración económica y, en general, su posibilidad de tratarlos como un bien fungible y líquido. Así mismo, debe determinarse si el mayor valor corresponde a los algoritmos para su explotación o a los datos en sí mismos. Por lo anterior, la presente línea de acción describe las actividades que aportarán a que se disponga de una definición del mercado de datos en el país, que facilite su aprovechamiento, permita impulsar la economía y aportar a su transformación.

En primer lugar, el Departamento Nacional de Planeación, en articulación con la Superintendencia de Industria y Comercio y el Ministerio de Tecnologías de la Información y las Comunicaciones, entre marzo y diciembre de 2018, adelantará un estudio sobre el mercado de datos en Colombia que permita su identificación, definición y caracterización. Así como, las tipologías de datos que actualmente son transadas y los análisis para la definición e incorporación de los diversos roles en el mercado de datos (*bróker* de datos, proveedor de datos, proveedor de bienes y servicios basados en la explotación de datos, y consumidor de estos bienes y servicios).

En segundo lugar, el Ministerio de Comercio, Industria y Turismo, a través de la Unidad de Gestión de Crecimiento Empresarial -iNNpulsa Colombia de la Fiduciaria de Comercio Exterior S.A. definirá e implementará, a partir de diciembre de 2018, un instrumento de apoyo directo (transversal) para asesorar a las empresas en el aprovechamiento y explotación de datos para el aumento de su productividad. Para ello, se diseñará e implementará el instrumento de acompañamiento que, de acuerdo con el perfil de las empresas seleccionadas, facilite la adopción de la explotación de datos, dentro de lo que se incluye la aplicación de la técnica “aprender-haciendo”. Lo anterior para facilitar la apropiación de la explotación de datos por las empresas, de modo que se incentive la demanda de bienes y servicios basados en datos.

En tercer lugar, con el fin de articular los recursos de computación de alto rendimiento que han sido dispuestos de manera previa por el Ministerio de Tecnologías de la Información y las Comunicaciones¹³⁴, esta entidad definirá e implementará, a partir de

¹³³ Respetando en todo caso lo dispuesto en el artículo 6 de la Ley 1581 de 2012 y el artículo 2.2.2.25.2.3. del Decreto 1074 de 2015 para el tratamiento de los datos personales y sensibles, según la voluntad del titular. Igualmente, se entienden excluidos del mercado los sometidos a reserva legal.

¹³⁴ Corresponsables al Centro de Bioinformática y Biología Computacional BIOS, alianza entre el Gobierno, la academia, la industria TI y la empresa privada. Se enfoca en ofrecer servicios de tecnología informática para supercomputación científica, así como hacer investigación científica, innovación y productividad que genere

diciembre de 2021, una plataforma que articule y potencie el aprovechamiento de estos recursos para facilitar el acceso a la academia, la sociedad civil, los emprendedores y las entidades públicas a una red nacional de servicios de computación de alto rendimiento.

En cuarto lugar, con el fin de disponer de un acuerdo marco de precios para la explotación de datos, la Agencia Nacional de Contratación Pública - Colombia Compra Eficiente, en articulación con el Ministerio de Tecnologías de la Información y las Comunicaciones y el Departamento Nacional de Planeación, incorporará en mayo de 2018 los bienes y servicios relacionados con el aprovechamiento y procesamiento de datos en los acuerdos marco de precios de nube pública y nube privada. Lo anterior, para facilitar la agregación de demanda y adquisición por parte de las entidades públicas de los bienes y servicios mencionados.

Línea de acción 13. Vincular la ciudadanía a la infraestructura de datos

Con el fin de incentivar el uso de los bienes y servicios basados en datos, así como permitir la toma de decisiones basadas en estos por parte de los ciudadanos, se describen a continuación las acciones para poner la analítica a su alcance, mediante casos de uso y capacitación. Lo anterior, permitirá avanzar en la alfabetización en datos.

En primer lugar, el Ministerio de Tecnologías de la Información y las Comunicaciones pondrá a disposición de la ciudadanía, en diciembre de 2018, una herramienta informática disponible en el portal de datos abiertos¹³⁵ para construir análisis y resolver problemas por parte de ciudadanos, estudiantes, emprendedores y otros actores del ecosistema de datos, orientada a la apropiación de la analítica y la desmitificación del concepto de *Big Data*. Esta deberá cumplir con los criterios de accesibilidad dispuestos en las normas. Lo anterior permitirá incentivar el desarrollo de las capacidades de los ciudadanos para producir, apropiar, comunicar y usar datos. Así mismo, se facilitará la comprensión del concepto de aprovechamiento de datos.

En segundo lugar, con el fin de incluir la analítica en la apropiación digital, durante el 2021, el Ministerio de Tecnologías de la Información y las Comunicaciones incluirá en las acciones de apropiación de TIC¹³⁶ líneas específicas de intervención para fomentar el

soluciones para las empresas y formación. También está la Red Nacional Académica de Tecnología Avanzada (RENATA), que es cofinanciada por el Ministerio de Tecnologías de la Información y las Comunicaciones y busca desarrollar proyectos colaborativos de ciencia, educación e innovación, que integren elementos de telepresencia, computación de alto rendimiento, procesamiento masivo y distribuido, interconexión de laboratorios, acceso a recursos remotos, simulación en entornos virtuales compartidos y educación virtual (e-Learning).

¹³⁵ <http://www.datos.gov.co/>

¹³⁶ Son proyectos que tienen como meta mejorar la calidad de vida de cada colombiano, reducir la brecha digital, disminuir la pobreza y contribuir al desarrollo del país con la apropiación de las TIC, que son implementados por el Ministerio de Tecnologías de la Información y las Comunicaciones en cumplimiento de lo dispuesto en el artículo 24 del Decreto 1414 de 2017.

uso de la explotación de datos, así como la apropiación de los productos que esta genera, orientado a dinamizar el ecosistema de datos en el país. Lo anterior permitirá que los ciudadanos incorporen en su vida diaria los beneficios que pueden obtener de los datos y de los bienes, servicios y productos basados en estos, así como tener la capacidad de encontrarlos, determinar su credibilidad, producirlos e interpretarlos de manera crítica.

5.4. Seguimiento

En primera medida se hará seguimiento a los indicadores que miden los cambios y resultados que se esperan observar una vez se cumplan los objetivos generales y específicos de la política de explotación de datos. La Tabla 8 muestra los resultados esperados con la implementación de la política y evidencia los efectos de la provisión de las condiciones para aumentar el aprovechamiento de datos en el país, particularmente por parte de las entidades públicas, proyectándose que el 50 % de estas avanzará al tercer nivel de preparación para la explotación de datos¹³⁷, con una realización promedio de cinco proyectos de explotación de datos para 2022. Lo anterior es especialmente relevante porque el objetivo de la presente política es generar los elementos para el aprovechamiento de datos mediante la corrección de las fallas que han impedido la transformación institucional.

Tabla 8. Indicadores de resultado de la política de explotación de datos

Indicadores de resultado	Línea base 2017	Meta 2020	Meta 2022
Porcentaje de entidades que tienen al menos un proyecto de aprovechamiento de datos	9,3 %	45 %	90 %
Entidades del orden nacional que avanzaron al tercer nivel del índice de preparación para la explotación de datos	8,7 %	20 %	50 %
Entidades del Gobierno que hacen proyectos de aprovechamiento de datos para mejorar servicios prestados a la ciudadanía	5,3 %	10 %	50 %
Promedio de activos públicos digitalizados y publicados	51 %	80 %	100 %

Fuente: DNP (2018).

La principal herramienta para hacer seguimiento a la ejecución física y presupuestal de las acciones de la presente política será el Plan de Acción y Seguimiento (PAS) (Anexo A). El reporte periódico al PAS se realizará por parte de todas las entidades involucradas en este documento y será consolidado por el DNP, de acuerdo con lo estipulado en la Tabla

¹³⁷ Se identifican cuatro niveles de preparación para la explotación de datos, primero a cuarto de mayor a menor. Actualmente, el 8,7 % de las entidades públicas se encuentra en el tercer nivel (avances no consistentes) (DNP, 2017).

9. Cabe aclarar que el cumplimiento de los indicadores contenidos en el PAS por parte de las diferentes entidades estará sujeto a la disponibilidad de recursos que se apropien para tal fin, sin perjuicio del seguimiento a los compromisos establecidos en este para la ejecución de la presente política.

Tabla 9. Cronograma de seguimiento implementación de la política

Corte	Fecha
Primer corte	30 de junio de 2018
Segundo corte	31 de diciembre de 2018
Tercer corte	30 de junio de 2019
Cuarto corte	31 de diciembre de 2019
Quinto corte	30 de junio de 2020
Sexto corte	31 de diciembre de 2020
Séptimo corte	30 de junio de 2021
Informe de cierre	31 de diciembre de 2021

Fuente: DNP (2018).

5.5. Financiamiento

Para efectos del cumplimiento de los objetivos de esta política, las entidades involucradas en su ejecución gestionarán y priorizarán, en el marco de sus competencias, los recursos para la financiación de las estrategias que se proponen, acorde con el Marco de Gasto de Mediano Plazo del respectivo sector.

La política tiene un costo total aproximado de 16.728 millones de pesos. En la

Tabla 10 se muestra el costo por año, así como los recursos de los cuales disponen las entidades ejecutoras de las acciones propuestas en el presente documento, que permiten la financiación completa de los costos estimados.

Tabla 10. Financiamiento de la política por año

Millones de pesos

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

Entidad	2018	2019	2020	2021	Total
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 (2018).

6. RECOMENDACIONES

El Departamento Nacional de Planeación y el Ministerio de Tecnologías de la Información y las Comunicaciones recomiendan al Consejo Nacional de Política Económica y Social:

1. Aprobar la política de explotación de datos descrita en el presente documento, incluyendo su PAS (Anexo A).
2. Solicitar a las entidades del Gobierno nacional involucradas en este documento priorizar los recursos para la puesta en marcha de las estrategias contenidas en el mismo, acorde con el Marco de Gasto de Mediano Plazo del respectivo sector.
3. Solicitar al Departamento Nacional de Planeación:
 - a. Consolidar y divulgar la información del avance de las acciones según lo planteado en el PAS (Anexo A). La información deberá ser proporcionada por las entidades involucradas en este documento de manera semestral durante todo el horizonte de implementación de la presente política, de acuerdo con lo establecido en la Tabla 9.
 - b. Definir la infraestructura de datos y las condiciones para su aprovechamiento y mantenimiento. Estas condiciones incluirán la implementación de mecanismos de colaboración público-privada que permitan compartir datos y capacidades analíticas.
 - c. Proponer el texto de Decreto reglamentario en el que se definirán reglas precisas que hagan efectivo y expedito el flujo de datos entre entidades, para materializar la habilitación general de compartición de datos al interior del sector público.
 - d. Coordinar el proceso de consolidación del marco jurídico aplicable a la explotación de datos, mediante la propuesta normativa que incorpore las pautas para hacer efectivos los principios de publicidad por defecto de los datos generados y recolectados por las entidades públicas, la privacidad por diseño de los datos personales, así como la garantía de mantenimiento, calidad y explotación de este activo.
 - e. Elaborar y actualizar periódicamente el índice de preparación para la explotación de datos.
4. Solicitar al Ministerio de Tecnologías de la Información y las Comunicaciones:
 - a. Diseñar e implementar una herramienta que facilite la administración de datos en las entidades públicas.
 - b. Definir y adoptar los estándares y lineamientos técnicos que deberán aplicar las entidades para garantizar la calidad de los datos durante todo su ciclo de vida,

así como para verificar de manera automatizada la calidad de estos, siguiendo el criterio de publicidad por diseño y defecto de los datos públicos de las entidades públicas.

- c. Definir los estándares que deben cumplir los registros administrativos para garantizar la publicidad por diseño y defecto de los datos públicos de las entidades públicas.
 - d. Liderar y realizar espacios de entrenamiento en aprovechamiento de datos que permitan transferir conocimiento para actualizar y fortalecer el capital humano.
 - e. Realizar un estudio que permita identificar las brechas de capital humano para la explotación de datos.
 - f. Diseñar e implementar acciones que activen la demanda de formación en materia de explotación de datos.
5. Solicitar al Departamento Administrativo de la Función Pública:
- a. Proponer un modelo de gestión de datos que se articule con las políticas de gestión y desempeño institucional en el marco del modelo integrado de planeación y gestión de las entidades públicas.
 - b. Incorporar temáticas relacionadas con la explotación de datos en los planes de capacitación de las entidades públicas para fortalecer el capital humano del ecosistema de datos.
6. Solicitar al Archivo General de la Nación:
- a. Generar un plan de gestión documental electrónica (hoja de ruta) que será aplicado en las entidades públicas para la automatización progresiva de procesos y procedimientos administrativos en la producción y gestión de documentos nativos electrónicos y digitalizados.
 - b. Definir y adoptar los estándares y lineamientos técnicos que orienten el proceso de *anonimización* de datos personales para garantizar la adecuada remoción de los identificadores personales y mitigar riesgos de reidentificación de datos personales.
7. Solicitar a la Secretaría de Transparencia de la Presidencia de la República actualizar la metodología para la elaboración del Plan Anticorrupción y de Atención al Ciudadano para incluir la analítica en proceso de su formulación.
8. Solicitar al Ministerio de Educación Nacional identificar las competencias necesarias para la explotación de datos que deben incorporarse en la metodología de diseño y desarrollo del marco nacional de cualificaciones.

9. Solicitar al Ministerio de Comercio, Industria y Turismo, a través de la Unidad de Gestión de Crecimiento Empresarial -iNNpulsa Colombia de la Fiduciaria de Comercio Exterior S.A proponer una estrategia de negocios que estimule la creación de nuevas empresas, o nuevas líneas de negocio en las existentes, orientada a incentivar el aprovechamiento de datos y la analítica, para su consolidación transversal en el aparato productivo, fortaleciendo especialmente a los emprendedores en la materia.
10. Solicitar a la Superintendencia de Industria y Comercio:
 - a. Proponer un esquema para simplificar la comprensión y aplicación de las clasificaciones legales de los datos y determinar la necesidad de realizar ajustes normativos para la unificación de conceptos y clasificaciones.
 - b. Definir los lineamientos que deben aplicar las entidades públicas para el adecuado tratamiento de los datos personales de los ciudadanos, alineado con el principio de responsabilidad demostrada, desarrollado en la reglamentación.
 - c. Determinar las necesidades de actualización del marco jurídico para la protección de los ciudadanos, en la que se identifiquen los requerimientos y desarrollos requeridos en aspectos como los principios de autorización, información, necesidad y finalidad, entre otros, que permitan mitigar los riesgos.
11. Solicitar a la Agencia Nacional de Contratación Pública -Colombia Compra Eficiente actualizar las guías, manuales y lineamientos, para que los datos se reconozcan explícitamente en la gestión contractual y para que la apertura de datos masivos, de calidad, sea incorporada como criterio de selección y supervisión de la ejecución.
12. Solicitar a la Agencia Nacional de Defensa Jurídica del Estado realizar acompañamiento a las entidades públicas para incentivar la explotación de los datos contenidos en eKogui para fortalecer el ciclo de defensa jurídica.

GLOSARIO

Analítica: disciplina orientada a analizar datos mediante técnicas científicas y herramientas automatizadas con énfasis en identificar hechos, relaciones, patrones ocultos de comportamiento de variables, correlaciones y tendencias.

Analítica descriptiva: responde preguntas sobre eventos que ya ocurrieron.

Analítica diagnóstica o exploratoria: determina la causa de un fenómeno ocurrido, mediante la identificación de las razones del comportamiento de variables específicas en un periodo definido.

Analítica predictiva: identifica la ocurrencia y comportamiento de un evento futuro a través de la predicción de variables específicas, partiendo de los registros históricos de las mismas, así como de otras adicionales.

Analítica prescriptiva: emplea los resultados de analítica predictiva para determinar acciones que deben tomarse en el futuro, así como el momento para su ejecución.

Anonimización: definida en el artículo 2.2.3.1.1 del Decreto 1170 de 2015 como el “proceso técnico que consiste en transformar los datos individuales de las unidades de observación, de tal modo que no sea posible identificar sujetos o características individuales de la fuente de información, preservando así las propiedades estadísticas en los resultados”.

Apertura por defecto: presume que toda la información será publicada a menos que haya una razón jurídica para mantenerla protegida (que se trate de datos personales, sensibles, reservados o que afecten la intimidad, privacidad, vida, salud, seguridad, secretos comerciales, industriales y profesionales).

Apertura por diseño: implica que la gestión de los datos es concebida, desde la captura o generación, para su publicación, es decir, esta no se agrega y se restringe solo por razones jurídicas.

Aprendizaje de máquinas: corresponde a los métodos para generar algoritmos que aprendan de los datos, esto es, que luego de analizarlos puedan identificar patrones que les permitan tomar decisiones o hacer predicciones basados en los mismos.

Aprendizaje profundo: conjunto de técnicas mediante las cuales los algoritmos realizan abstracciones de alto nivel para tomar decisiones respecto de datos nuevos a partir de otros conocidos. Las técnicas de aprendizaje profundo se caracterizan por su alto nivel de desempeño en la obtención de los resultados.

Ciclo de vida de los documentos: de acuerdo con el artículo 23 de la Ley 594 de 2000¹³⁸, corresponde al tiempo que los documentos permanecen como archivo de gestión (toda la documentación que es sometida a continua utilización y consulta administrativa por las oficinas productoras u otras que la soliciten. Su circulación o trámite se realiza para dar respuesta o solución a los asuntos iniciados), archivo central (agrupa documentos transferidos por los distintos archivos de gestión de la entidad respectiva, cuya consulta no es tan frecuente pero que siguen teniendo vigencia y son objeto de consulta por las propias oficinas y particulares en general) y archivo histórico (aquel al que se transfieren desde el archivo central los documentos de archivo de conservación permanente).

Computación cognitiva: conjunto de tecnologías que buscan replicar el funcionamiento del cerebro humano, es decir, la capacidad para percibir, razonar y responder a estímulos. Dentro de esta, una de las innovaciones más destacadas de los últimos años es la inteligencia artificial, que hace parte de las técnicas más avanzadas para el aprovechamiento de datos.

Conjunto de datos: unidad mínima de información sujeta a carga, publicación, transformación y descarga (Ministerio de Tecnologías de la Información y las Comunicaciones, 2016).

Datificación: La palabra es el resultado de combinar los conceptos de datos y cuantificación, *datification* en el texto original en inglés. Significa transformar el mundo en datos procesables y cuantificables (OCDE, 2015).

Dato personal: cualquier información vinculada o que pueda asociarse a una o varias personas naturales determinadas o determinables. De acuerdo con el tipo de información a la que se refieren, estos pueden ser sensibles, semiprivados, privados o públicos. Es decir, no todos los datos personales son privados.

Datos abiertos: son “datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos”¹³⁹. La definición de apertura involucra los siguientes elementos: (i) disponibilidad y acceso¹⁴⁰; (ii)

¹³⁸ Por medio de la cual se dicta la Ley General de Archivos y se dictan otras disposiciones.

¹³⁹ Literal j) del artículo 6 de la Ley 1712 de 2014.

¹⁴⁰ La información debe estar disponible como un todo y a un costo razonable de reproducción, preferiblemente descargándola de internet. Además, la información debe estar disponible en una forma conveniente y modificable.

reutilización y redistribución¹⁴¹; y (iii) participación universal¹⁴² (OKFN, 2016). La definición de apertura implica la exclusión de todos aquellos datos que, dada su tipología, presentan limitaciones jurídicas para su acceso, divulgación, compartición y tratamiento.

Datos digitales aprovechables: datos que cumplen con condiciones técnicas que facilitan su descubrimiento, uso y explotación.

Datos enlazados: *Linked data* en su nombre original en inglés. Es el método con el que se pueden mostrar, intercambiar y conectar datos en Internet. Los conjuntos de datos disponibles en este que se encuentran enlazados conforman la web de datos, es decir, una gran base de datos entrelazados. El término es atribuido a Tim Berners-Lee¹⁴³.

Datos oscuros: generados, almacenados, procesados o empleados durante las actividades humanas, que no son usados para otro propósito y cuyo contenido es desconocido (Gartner, 2012). Son entendidos como un activo que se desconoce y desaprovecha. La mayor parte de esta categoría corresponde a datos no estructurados, porque su descubrimiento y explotación implica mayor complejidad.

Director de sistemas de información: líder de la gestión estratégica de tecnologías de información, encargado de planificar, organizar, coordinar, gestionar y controlar la estrategia de uso y apropiación de TI, y todo lo que conlleva esta tarea¹⁴⁴.

Espacio colaborativo (*data sandbox*): plataforma escalable y de desarrollo que es empleada para explorar conjuntos de datos de manera colaborativa e interactiva entre varios usuarios, para determinar soluciones aplicables en el contexto real, sin afectar datos que no son objeto de la experimentación. Esta plataforma cuenta con altas capacidades de procesamiento y almacenamiento.

Explotación de datos: generación de valor social y económico mediante el aprovechamiento de datos para la creación de nuevos bienes, servicios, procesos, así como para el mejoramiento de los existentes

Gestión documental: de acuerdo con el artículo 3 de la Ley 594 de 2000, se define como el conjunto de actividades administrativas y técnicas tendientes a la planificación, manejo y organización de la documentación producida y recibida por las entidades, desde su origen hasta su destino final, con el objeto de facilitar su utilización y conservación.

¹⁴¹ Los datos deben ser provistos bajo términos que permitan reutilizarlos y redistribuirlos, e incluso integrarlos con otros conjuntos de datos.

¹⁴² Todos deben poder utilizar, reutilizar y redistribuir la información.

¹⁴³ Ver: <https://www.w3.org/DesignIssues/LinkedData.html>

¹⁴⁴ Ver: <http://www.mintic.gov.co/gestionti/615/w3-propertyvalue-6205.html>

Gobierno Abierto: componente del Gobierno Electrónico que busca la construcción de un Estado más transparente, participativo y colaborativo en los asuntos públicos mediante el uso de las TIC.

Gobierno Electrónico: aprovechamiento de las TIC para el funcionamiento del sector público con el fin de incrementar la eficiencia, la transparencia y la participación ciudadana.

Gobierno en Línea: nombre que recibe la estrategia de Gobierno Electrónico de Colombia. Su objetivo es construir un Estado más eficiente, más transparente y más participativo gracias a las TIC.

Habeas data: derecho reconocido en la Constitución para conocer, actualizar y rectificar las informaciones que se hayan recogido sobre las personas en bancos de datos y en archivos de entidades públicas y privadas. Este derecho tiene una naturaleza autónoma que lo diferencia de otras garantías con las que está en permanente relación, como los derechos a la intimidad y a la información.

Hardware: componentes físicos de un sistema informático o computador.

Innovación de mercadeo: la implementación de un nuevo método de mercadeo que involucre cambios significativos en el empaque, colocación del producto, promoción o precios.

Innovación de organizacional: la implementación de un nuevo método organizacional en las prácticas de negocio, la organización del sitio de trabajo o las relaciones externas.

Innovación de procesos: la implementación de un método de producción o servicio nuevo o significativamente mejorado. Esto incluye cambios significativos en técnicas, equipos o *software*.

Innovación de productos: la introducción de un bien o servicio que es nuevo o ha sido mejorado significativamente con respecto a sus características o usos. Esto incluye mejoras significativas en sus especificaciones técnicas, componentes y materiales, incorporación de *software*, facilidad de uso y otras características funcionales.

Inteligencia artificial: una rama de la ciencia de la computación que busca que las máquinas imiten el comportamiento inteligente para la resolución de problemas. Este campo ofrece los mayores avances en las capacidades técnicas para la explotación de datos. Se sirve de conocimientos de otras disciplinas, como la lingüística, psicología, biología, entre otras. Un paso previo para obtener la inteligencia artificial es el aprendizaje de máquinas.

Interfaz de programación de aplicaciones (*Application Programming Interface*): Conjunto de funciones y procedimientos que opera como una llave de acceso para usar datos entre aplicaciones y servicios web.

Internet de las cosas: hace referencia a dispositivos, diferentes a un computador o teléfono móvil, conectados directamente a Internet (v.g., electrodomésticos, ropa, relojes, zapatos, juguetes, edificaciones, dispositivos médicos, entre otros). Estos dispositivos tienen la capacidad de medir fenómenos y generar datos digitales derivados de esta actividad.

Interoperabilidad: habilidad de transferir y utilizar información de manera uniforme y eficiente entre varias organizaciones y sistemas de información (Ministerio de Tecnologías de la Información y las Comunicaciones, 2016)¹⁴⁵. Corresponde a la capacidad de dos o más sistemas (computadoras, medios de comunicación, redes, software y otros componentes de TI) de interactuar e intercambiar datos de acuerdo con un método definido¹⁴⁶.

Metadatos: información que describe los datos (ITU, 2015) y permite interpretar el contenido de un conjunto de estos, identificar características para evaluar su calidad y pertinencia frente al asunto bajo análisis, rastrear su origen y procesamiento previo e identificar reglas de validación. Por ejemplo: registro de autor, fecha de creación de un documento, atributos como tamaño de un archivo, entre otros.

Modelos de autorregulación: conjuntos de pautas que se orientan a guiar a los agentes para actuar de la mejor manera posible dentro de las circunstancias particulares, de modo que no hubiese sido factible otro resultado superior. Ejemplos de ello son las buenas prácticas implementadas para el uso de datos personales que no se conforman con simplemente obtener el consentimiento del titular, los consensos de los profesionales y la industria para actuar con lealtad incluso en situaciones de información asimétrica, el cuidado de los datos de los menores por parte de sus representantes, entre otros.

Privacidad por diseño y defecto: principio que guía el proceso organizacional para adoptar medidas preventivas de diversa naturaleza (tecnológica, organizacional, humana, procedimental) para evitar vulneraciones al derecho a la privacidad o a la confidencialidad de la información, así como fallas de seguridad o indebidos tratamientos de datos personales. Este proceso se define desde antes de que se recolecte información y durante todo el ciclo de vida de la misma. La privacidad y la seguridad deben hacer parte del diseño, arquitectura y configuración predeterminada del proceso de gestión de

¹⁴⁵ Ver: http://css.mintic.gov.co/ap/gel4/html/como_se_logra/interoperabilidad/introduccion_interoperabilidad.html

¹⁴⁶ Incluye 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 en línea a ciudadanos, empresas y a otras entidades.

información y de las infraestructuras que lo soportan (artículo 2.2.17.1.5 del Decreto 1078 de 2015).

Software: conjunto de programas, instrucciones y reglas informáticas para ejecutar ciertas tareas en una computadora.

Variable cualitativa: expresa cualidades, atributos, categorías o características.

Variable cuantitativa: expresa valores numéricos.

ANEXOS

Anexo A. Plan de Acción y Seguimiento (PAS)

Ver archivo Excel anexo.

Anexo B. Normas aplicables a la explotación de datos

Ver archivo en PDF anexo.

Anexo C. Antecedentes indirectos de la política de explotación de datos

Ver archivo en PDF anexo.

Anexo D. Experiencias de aprovechamiento de datos

Ver archivo en PDF anexo.

Anexo E. Capital humano requerido para la explotación de datos

Ver archivo en PDF anexo.

BIBLIOGRAFÍA

- Accenture; Cintel. (1 de septiembre de 2016). *Primer manifiesto sobre transformación digital en Colombia 2016*. Obtenido de cintel: <http://www.cintel.co/wp-content/uploads/2016/09/1er-MANIFIESTO-SOBRE-TRANSFORMACION-DIGITAL-EN-COLOMBIA-2016-V-FINAL.pdf>
- Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A., & Goodrich, V. (2014). Engagement vs performance: using electronic portfolios to predict first semester engineering student retention. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 103-112.
- ANDI. (2016). *Encuesta de Opinión Industrial Conjunta Enero - mayo 2016*. Bogotá: Asociación Nacional de Empresarios de Colombia.
- Archivo General de la Nación. (11 de septiembre de 2012). Circular Externa 005 de 2012. *Recomendaciones para llevar a cabo los procesos de digitalización y comunicaciones oficiales electrónicas en el marco de la iniciativa cero papel*. Bogotá, Colombia: Archivo General de la Nación.
- Archivo General de la Nación. (2013). *1-2013-4146 - 4434/2013/SGC - 420*. Bogotá.
- Archivo General de la Nación. (30 de noviembre de 2017). Alcance comunicación 2-2017-13925. *Radicado 2-2017-14155*. Bogotá: Subdirección de Tecnologías de la Información Archivística y Documento Electrónico.
- Archivo General de la Nación. (23 de febrero de 2017). *Archivo General de la Nación*. Obtenido de Tablas de Retención Documental - TRD: <http://www.archivogeneral.gov.co/tablas-de-retencion-documental>
- Arcitura. (2014). *Big Data Análisis & Technology Concepts. Module 2. Big Data Science School. Certified Big Data Approved Course*. Big Data Science School.
- Article 29 Data Protection Working Party. (13 de julio de 2010). *The European Commission*. Obtenido de http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp173_en.pdf
- Banco Interamericano de Desarrollo, Centro Latinoamericano de Administración para el Desarrollo. (julio de 2007). *Modelo abierto de Gestión Para Resultados en el sector público*. Washington, D.C.: Banco Interamericano de Desarrollo.
- BID. (2014). *BID*. Obtenido de La Importancia de las Matemáticas para la Vida: <http://www.iadb.org/es/temas/educacion/bid-ensenanza-de-ciencias-y-matematicas-en-america-latina,6494.html>
- Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 1073-1076.

- Bogomolov, A., Lepri, B., Larcher, R., Antonelli, F., Pianesi, F., & Pentland, A. (2016). Energy consumption prediction using people dynamics derived from cellular network data. *Springer Open*.
- Bogomolov, A., Lepri, B., Staiano, J., Letouzé, E., Oliver, N., Pianesi, F., & Pentland, A. (2015). Moves on the Street: Classifying Crime Hotspots Using Aggregated Anonymized Data on People Dynamics. *Big Data*, 148-158.
- Bolbol, A., Cheng, T., Tsapakis, I., & Haworth, J. (2012). Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems*, 526-537.
- Borgesius, F. Z., Gray, J., & van Eechoud, M. (2015). OPEN DATA, PRIVACY, AND FAIR INFORMATION PRINCIPLES: TOWARDS A BALANCING FRAMEWORK. *BERKELEY TECHNOLOGY LAW JOURNAL*, 2075 - 2131. Obtenido de <http://dx.doi.org/10.15779/Z389S18>
- Brown, B. C. (2014). Views from the front lines of the data-analytics revolution. *McKinsey Quarterly*.
- Buhl. (2013). Big Data A Fashionable Topic with(out) Sustainable Relevance for Research and Practice? *Business & Information Systems Engineering*, pp.63–68.
- Burgess, M., Giraudy, E., Katz-Samuels, J., Walsh, J., Willis, D., Haynes, L., & Ghani, R. (2016). The Legislative Influence Detector: Finding Text Reuse in State Legislation. *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 57-66.
- Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H., Mitra, P., . . . Jansen, B. J. (2011). Classifying text messages for the Haiti earthquake. *Proceedings of the 8th international conference on information systems for crisis response and management*.
- Cavallo, A., & Rigobon, R. (2016). The billion prices project: using online prices for measurement and research. *Journal of economics perspectives*, 151 -158.
- CEBR. (2012). *Data Equity: Unlocking the value of Big Data*.
- Centre for Information Policy Leadership. (2011). *Accountability: Data Governance for the Evolving Digital Marketplace*. Obtenido de Centre for Information Policy Leadership at Hunton & Williams LLP: http://www.huntonfiles.com/files/webupload/CIPL_Centre_Accountability_Data_Governance_Paper_2011.pdf
- Chui, M., Löffler, M., & Roberts, R. (2010). *The Internet of Things*. Obtenido de McKinsey Quarterly: <http://www.mckinsey.com/industries/high-tech/our-insights/the-internet-of-things>

- Cobo, A., Parra, D., & Navón, J. (2015). Identifying relevant messages in a twitter-based citizen channel for natural disaster situations. *Proceedings of the 24th International Conference on World Wide Web*, 1189-1194.
- Comisión de Estadísticas Oficiales de la ONU. (marzo de 2017). *Cape Town Global Action Plan for Sustainable Development Data*. Obtenido de Cape Town Global Action Plan for Sustainable Development Data: <http://unstats.un.org/sdgs/hlg/Cape-Town-Global-Action-Plan>
- Comisión Económica para América Latina y el Caribe (CEPAL). (2003). *Registros administrativos, calidad de los datos y credibilidad pública: presentación y debate de los temas sustantivos de la segunda reunión de la Conferencia Estadística de las Américas de la CEPAL*. Santiago de Chile: CEPAL.
- Comisión Europea. (2014). *Hacia una economía de los datos próspera*. Bruselas.
- Consejo Asesor Nacional de Estadística. (2017). *Plan Estadístico Nacional 2017 - 2022*. Bogotá: Consejo Asesor Nacional de Estadística.
- Consejo de Estado, Alcance de la reserva estadística, 1209 (Consejo de Estado 27 de septiembre de 1999).
- Cooper, L. (26 de agosto de 2016). *Air pollution in China and IBM green initiatives*. Obtenido de ibm: <https://www.ibm.com/blogs/internet-of-things/air-pollution-green-initiatives/>
- Criado, I. J., Gascó, M., & Jiménez, C. E. (08 de 2011). *Interoperabilidad de Gobierno electrónico en Iberoamérica. Estudio comparativo y recomendaciones de futuro*. Obtenido de Interoperabilidad de Gobierno electrónico en Iberoamérica.: <https://estratic.files.wordpress.com/2011/08/interoperabilidad-iberoamerica-ryd-501.pdf>
- Data & Society Research Institute. (2014). *Event Summary: The Social, Cultural, & Ethical Dimensions of "Big Data"*. New York: Data & Society Research Institute.
- Data Pop Alliance. (2015). *Beyond Data Literacy: Reinventing community engagement and empowerment in the age of data*.
- Data-Pop Alliance. (noviembre de 2015). *Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data*. Obtenido de <http://datapopalliance.org/>: <http://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf>
- Departamento Administrativo de la Función Pública. (2016). *Resultados de Formulario Único de Registro de avance de la Gestión*.

- Departamento Administrativo de la Función Pública. (2016). *Resultados Formulario Único de Registro de Avance en la Gestión*. Bogotá.
- Departamento Administrativo de la Función Pública. (2017). *Resultados Formulario Único de Registro de Avance en la Gestión*.
- Department for Business, Innovation and Skills. (2013). *Seizing the data opportunity. A strategy for UK data capability*.
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., . . . Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 15888-15893.
- Diakopoulos, N., & Friedler, S. (17 de noviembre de 2016). *How to Hold Algorithms Accountable*. Obtenido de MIT Technology Review: <https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/>
- DNP. (2016). *CONPES 3866 Política Nacional de Desarrollo Productivo*. Bogotá.
- DNP. (2017). Análisis de Registros Únicos.
- DNP. (2017). *Análisis de texto Planes de Desarrollo Territorial*.
- DNP. (Marzo de 2017). Encuesta explotación de datos entidades públicas del orden nacional.
- DNP. (2017). Government Exploratory Data Analysis. Bogotá.
- Engstrom, R., Hersh, J., & Newhouse, D. (2016). Poverty in HD: What Does High Resolution Satellite Imagery Reveal about Economic Welfare? *Working paper*.
- Etymonline. (16 de 10 de 2016). *etymonline*. Obtenido de etymonline: <http://www.etymonline.com/index.php?term=data>
- European Data Protection Supervisor. (2 de junio de 2014). *Report of workshop on Privacy, Consumers, Competition and Big Data*. Obtenido de European Data Protection Supervisor: https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Big%20data/14-07-11_EDPS_Report_Workshop_Big_data_EN.pdf
- European Political Strategy Centre. (21 de Enero de 2017). *Enter the Data Economy, EU Policies for a Thriving Data Ecosystem*. Obtenido de http://ec.europa.eu/epsc/home_en: https://ec.europa.eu/epsc/sites/epsc/files/strategic_note_issue_21.pdf
- Executive Office of the President, National Science and Technology Council. (2016). *The Federal Big Data Research And Development Strategic Plan*.

- Fernández Gómez, L. (2014). *¿How can Colombia's Government best exploit the potential of big data?* Manchester: Tesis de nivel de educación no publicada, University of Manchester, Faculty of Humanities.
- Gartner. (2012). *The Importance of "Big Data": A Definition*. Gartner Research.
- Gartner. (2013). *Hype Cycle for Big Data*.
- Gartner. (2015). Building the Digital Platform: Insights From the 2016 Gartner CIO Agenda Report. *Building the Digital Platform: Insights From the 2016 Gartner CIO Agenda Report*. Stamford, Estados Unidos: Gartner Corporate Headquarters.
- Gobierno de la República de Francia. (2015). *Rallying the «New Face of Industry in France»*.
- Goel, V. (29 de junio de 2014). *Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry*. Obtenido de The New York Times: https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html?_r=0
- Greenwood, D. ., Stopczynski, A., Sweatt, B., Hardjono, T., & Pentland, A. ". (2014). *Privacy, Big Data, and the Public Good, Chapter 9: The New Deal on Data: A Framework for Institutional Controls*. Obtenido de Civics: <http://civics.com/NewDealOnData/>
- Guidotti, R., Coscia, M., Pedreschi, D., & Pennacchioli, D. (2016). Going Beyond GDP to Nowcast Well-Being Using Retail Market Data. *International Conference and School on Network Science*, 29-42.
- Haav, H.-m., & Kungas, P. (4 de agosto de 2014). *Semantic data interoperability: the key problem of big data*. Obtenido de Research Gate: https://www.researchgate.net/publication/260706654_Semantic_data_interoperability_the_key_problem_of_big_data
- Hardin, G. (1968). The tragedy of the commons. *Science, New Series, Vol. 162, No. 3859*, 1243 - 1248. Obtenido de http://www.geo.mtu.edu/~asmayer/rural_sustain/governance/Hardin%201968.pdf
- International Data Corporation. (2017). *Data Age 2025: The Evolution of Data to Life-Critical. Don't Focus on Big Data; Focus on the Data That's Big*. Framingham: International Data Corporation (IDC). Obtenido de <http://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>
- IT Strategic Headquarters Japan. (2013). *Declaration to be the World's Most Advanced IT Nation*.

- ITU. (noviembre de 2013). *technology Watch Report (November 2013) Big data - big today, normal tomorrow*. Obtenido de International Telecommunication Union: http://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000220001PDFE.pdf
- ITU. (2015). *Big data – Cloud computing based requirements and capabilities*.
- Kang, J. S., Kuznetsova, P., Luca, M., & Choi, Y. (2013). Where not to eat? improving public policy by predicting hygiene inspections using online reviews. *EMNLP*, 1443-1448.
- Laney, D. (6 de febrero de 2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *3D Data Management: Controlling Data Volume, Velocity and Variety*. Stamford: META Group Inc.
- Lohr. (11 de Febrero de 2012). *New York Times*. Obtenido de <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>
- Mayer-Schönberger, V., & Padova, Y. (2016). Regime Change? Enabling Big Data Through Europe'S New Data Protection Regulation. *The Columbia Science & Technology Law Review*, 315-335.
- McKinsey Global Institute. (2011). Big data: The next frontier for innovation, competition and productivity.
- McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition, and productivity*.
- McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition, and productivity*.
- McKinsey Global Institute. (2013). *Game changers: Five opportunities for US growth and renewal*. McKinsey & Company.
- McKinsey Global Institute. (2015). The Internet of Things: mapping the value behind the hype. *McKinsey Global Institute*.
- McKinsey Global Institute. (2016). *THE AGE OF ANALYTICS*. McKinsey Global Institute.
- McKinsey Global Institute. (2016). The age of analytics: Competing in a data driven world.
- Ministerio de Tecnologías de la Información y las Comunicaciones. (Diciembre de 2015). Estudio de cultura de uso de TIC en los colombianos para relacionarse con el Estado. *Estudio de cultura de uso de TIC en los colombianos para relacionarse con el Estado*. Bogotá.

- Ministerio de Tecnologías de la Información y las Comunicaciones. (2015). *Estudio de evaluación de conceptos y levantamiento de la línea base de cuatro proyectos estratégicos para el gobierno en línea*. Bogotá.
- Ministerio de Tecnologías de la Información y las Comunicaciones. (2015). *Índice de Gobierno en Línea Entidades del Orden Nacional*.
- Ministerio de Tecnologías de la Información y las Comunicaciones. (23 de agosto de 2017). Estudio de acceso, uso y retos de las TIC en Colombia "Gran Encuesta TIC". *Estudio de acceso, uso y retos de las TIC en Colombia "Gran Encuesta TIC"*. Bogotá.
- Ministerio de Tecnologías de la Información y las Comunicaciones. (2017). *Índe de Gobierno en Línea Entidades del orden nacional*.
- Naciones Unidas. (2016). *E-Government in Support of Sustainable Development*. New York: Naciones Unidas.
- NASDAQ. (18 de julio de 2017). *NASDAQ*. Obtenido de Tesla, Inc. Cotización de valores y datos de resumen: <http://www.nasdaq.com/es/symbol/tsla>
- National Information Society Agency. (2013). *2013 National Informatización White Paper*.
- NYSE. (18 de julio de 2017). *New York Stock Exchange*. Obtenido de GENERAL MOTORS COMPANY (GM): <https://www.nyse.com/quote/XNYS:GM>
- OCDE. (2005). *Oslo Manual Guidelines for collecting and interpreting innovation data*. Paris: OECD Publishing.
- OCDE. (2005). *Oslo Manual Guidelines for collecting and interpreting innovation data*. Paris: OECD Publishing.
- OCDE. (2009). *Rethinking e-Government Services User-Centred Approaches*. Paris: OECD Publishing.
- OCDE. (2012). *Programme for International Student Assessment (PISA) Results overview Colombia 2012*.
- OCDE. (2013). *Exploring Data-Driven Innovation as a New Source of Growth*. Paris: OECD Publishing.
- OCDE. (2013). *Exploring the Economics of Personal Data*. Paris: OECD.
- OCDE. (2013). *Public Governace Reviews Colombia. Implementing good governance*.
- OCDE. (2014). *Estudio de la OCDE sobre la política regulatoria en Colombia: Más allá de la simplificación administrativa*. Paris: OECD Publishing.
- OCDE. (2015). *Data-Driven Innovation*. Paris: OECD Publishing.

- OCDE. (2017). *Development Co-operation Report 2017: Data for Development*. París: OECD Publishing. Obtenido de Development Co-operation Report 2017: Data for Development: <http://dx.doi.org/10.1787/dcr-2017-en>
- OCDE. (2017). Government at a Glance 2017. http://dx.doi.org/10.1787/gov_glance-2017-en. París: OECD Publishing.
- Ochoa, S., Rasmussen, J., Robson, C., & Salib, M. (2001). *Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study*. Boston: MIT. Obtenido de <http://web.mit.edu/sem083/www/assignments/reidentification.html>
- Open Knowledge International. (2017). *Global Open Data Index*.
- Rose, C. (1986). The Comedy of the Commons: Custom, Commerce, and Inherently Public Property. *University of Chicago Law Review*, 53(3), 711 - 781. Obtenido de <http://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=4473&context=uclrev>
- Ruta N Medellín. (enero de 2018). *Plataforma de innovación abierta SUNN*. Obtenido de Plataforma de innovación abierta SUNN: <https://www.rutanmedellin.org/es/oferta/empresarios-y-emprendedores/item/plataforma-de-innovacion-abierta-sunn>
- Sadilek, A., Kautz, H., & Silenzio, V. (2012). Predicting Disease Transmission from Geo-Tagged Micro-Blog Data. *Proceedings of the twenty-sixth AAAI Conference on Artificial Intelligence*.
- SAS. (2016). *The Value of Big Data and the Internet of Things to the UK Economy*.
- Schwab, K. (2016). *The Fourth Industrial Revolution*. Cologny /Geneva: World Economic Forum.
- Spence Stuart. (Junio de 2016). *Data leadership: Defining the expertise your organization needs*. Obtenido de Spence Stuart: <https://www.spencerstuart.com/~media/pdf%20files/research%20and%20insight%20pdfs/data-leadership-defining-the-expertise-your-organization-needs-june%202016.pdf>
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating Census Data for Population Mapping Using RandomForests with Remotely-Sensed and Ancillary Data. *PLoS ONE*, 1-22.
- Suzuki, L. R. (2015). *Data as Infrastructure for Smart Cities*", *PhD Thesis, University College London*. Obtenido de Larissa Romualdo Suzuki.: <http://www0.cs.ucl.ac.uk/staff/l.romualdo/DataInfraForSmartCities/>
- TechAmerica Foundation. (2012). *Demystifying big data*.

- The Australian Government Information Management Office. (2013). *Australian Public Service Big Data Strategy*.
- The Economist Intelligence Unit. (2013). *Fostering a data-driven culture*. Obtenido de http://www.tableau.com/sites/default/files/whitepapers/tableau_dataculture_130219.pdf:
http://www.tableau.com/sites/default/files/whitepapers/tableau_dataculture_130219.pdf
- The State Council People's Republic of China. (5 de septiembre de 2015). *The State Council People's Republic of China*. Obtenido de http://english.gov.cn/policies/latest_releases/2015/09/05/content_281475183253390.htm
- Timm, T. (19 de abril de 2016). *You may hate Donald Trump. But do you want Facebook to rig the election against him?* Obtenido de The Guardian: <https://www.theguardian.com/commentisfree/2016/apr/19/donald-trump-facebook-election-manipulate-behavior>
- United Nations. (2012). *Big Data for Development: Challenges & Opportunities*.
- United Nations. (2014). *Big data and modernization of statistical systems*.
- Uprimny Yepes, R. (2001). *Salvamento de voto sentencia T-1207/01 Corte Constitucional*. Bogotá.
- Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013). Learning to detect patterns of crime. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 515-530.
- Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., & M., A. (2012). Quantifying the Impact of Human Mobility on Malaria. *Science*, 267-270.
- Wesolowski, A., Prudhomme O'Meara, W., Tatem, A. J., Ndege, S., Eagle, N., & Buckee, C. O. (2015). Quantifying the Impact of Accessibility on Preventive Healthcare in Sub-Saharan Africa Using Mobile Phone Data. *Epidemiology*, 223-228.
- World Bank Group. (septiembre de 2015). *Open Data Readiness Assessment*. Obtenido de datos.gov.co: http://estrategia.gobiernoenlinea.gov.co/623/articles-9407_evaluacion_apertura_datos.pdf
- World Economic Forum. (2012). *Big Data, Big Impact: New Possibilities for International Development*.
- World Economic Forum. (enero de 2015). *Data-Driven Development Pathways for Progress*. Obtenido de World Economic Forum: http://digitalprinciples.org/wp-content/uploads/2015/05/WEFUSA_DataDrivenDevelopment_Report2015.pdf

World Economic Forum. (2016). *The 10 countries best prepared for the new digital economy*. Obtenido de weforum.org:
<https://www.weforum.org/agenda/2016/07/countries-best-prepared-for-the-new-digital-economy/>

World Economic Forum. (19 de enero de 2016). *World Economic Forum*. Obtenido de World Economic Forum Global Agenda:
<https://www.weforum.org/agenda/2016/01/what-is-the-fourth-industrial-revolution/>

World Wide Web Foundation. (2017). *Open Data Barometer*. Obtenido de <http://opendatabarometer.org/>:
<http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf>