



# RECOMENDACIONES E INSUMOS PARA LA DEFINICIÓN DE UNA ESTRATEGIA NACIONAL DE BIG DATA

*Propuesta de una Estrategia  
Nacional de Big Data para el  
gobierno colombiano*

2019

4

**Producido por un equipo compuesto por:**

Emmanuel Letouzé, Alex Pentland, Isabella Loaiza, Julie Ricard, Andrés Clavijo, Diego Silva

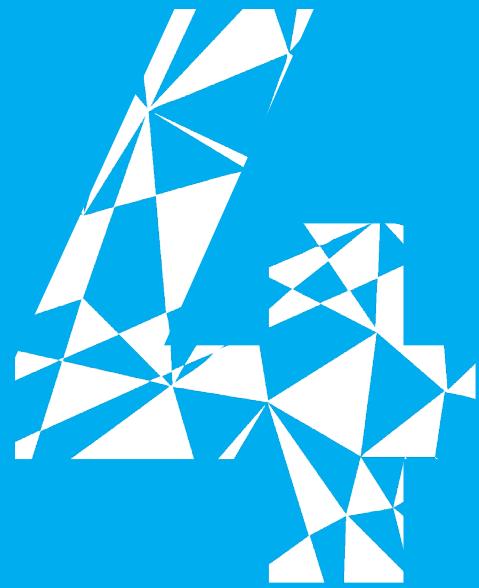
Coordinado por Emmanuel Letouzé

Con insumos de María Antonia Bravo, Valentina Casasbuenas Bajo la supervisión general de:  
Alex Pentland.

*Enero de 2019*

*Versión revisada y ajustada*

**PROPUESTA DE UNA  
ESTRATEGIA NACIONAL  
DE BIG DATA PARA EL  
GOBIERNO COLOMBIANO**





# ÍNDICE

---

<b>1. Introducción</b>	<b>8</b>
<b>2. Aproximación a una estrategia integral de datos en colombia</b>	<b>9</b>
<b>3. Fundamentos para una estrategia nacional</b>	<b>13</b>
<b>a. Lecciones Aprendidas del Contexto Internacional.</b>	<b>13</b>
<b>b. Diagnóstico del caso colombiano.</b>	<b>14</b>
<b>4. Acciones</b>	<b>16</b>
<b>a. Corto plazo.</b>	<b>16</b>
<b>b. Mediano plazo.</b>	<b>21</b>
<b>c. Largo plazo.</b>	<b>26</b>
<b>5. Actores</b>	<b>27</b>
<b>a. Instituciones responsables.</b>	<b>28</b>
<b>b. Comité de decisión y seguimiento.</b>	<b>29</b>
<b>c. Grupos de trabajo.</b>	<b>29</b>
<b>6. Conclusiones</b>	<b>30</b>
<b>7. Anexos</b>	<b>31</b>
<b>8. Bibliografía</b>	<b>34</b>

## RESUMEN EJECUTIVO

El objetivo de este documento es delinear las recomendaciones e insumos para la definición de la estrategia de Big Data para el Gobierno colombiano. De acuerdo al diagnóstico realizado para evaluar la situación actual de los datos y Big Data en el Estado Colombiano, se identifican tres tipos de retos: técnicos, sociales y de habilidades.

### **Retos Técnicos (de datos)**

1. Ausencia de datos públicos de alta calidad
2. Falta de conocimiento en cuanto a la ubicación, características, estructura y disponibilidad de fuentes de datos del sector público
3. Comprensión inadecuada de las necesidades en cuanto a datos entre y en las entidades del gobierno, y en general, en el ecosistema de datos del país
4. Necesidad de especificar estrategias de integración de datos, de arquitectura de datos, y protocolos de acceso a datos.
5. Existencia de un marco jurídico desarticulado con directrices confusas

### **Retos Sociales**

1. Desconocimiento del valor potencial de los datos
2. Baja implementación de herramientas habilitadoras para la explotación de datos por parte de funcionarios públicos

### **Retos de Habilidades**

1. Capacidad técnicas escasas para el procesamiento y análisis de datos
2. Ausencia de una cultura basada o “movida” por datos para la toma de decisiones

De acuerdo a estos obstáculos, se recomiendan las siguientes acciones para la estrategia de Big Data:

### **Acciones a corto plazo (implementación temprana)**

1. Mapeo y delineamiento de la arquitectura de datos nacional.
  - a. Identificar la ubicación, características, estructura, disponibilidad y calidad de las fuentes de datos de las entidades del gobierno Colombiano.
  - b. Mapear el flujo de información estructurada y no-estructurada y caracterizar deficiencias o avances en la arquitectura de datos actual.
  - c. Determinar necesidades en cuanto datos e información para la integración de la arquitectura de datos nacional
2. Repotenciamiento de los equipos de computo
  - a. Crear una agenda clara de adquisición, implementación y uso de equipos y servicios que permitan y faciliten la explotación de datos desde las entidades del gobierno
  - b. Garantizar la transformación y actualización de la manera como se almacena, procesa, usa y difunde información en el país
  - c. Implementar política de actualización de los equipos de cómputo para asegurar funcionalidad y relevancia de estos

### **Acciones a mediano plazo (implementación de alrededor tres años)**

1. Formación de capital humano y promoción de cultura de datos
  - a. Capacitar a funcionarios de gobierno en el uso y procesamiento de información y datos
  - b. Lanzar campaña que priorice habilidades y conocimiento sobre datos en currículos empresariales, organizacionales y educativos
  - c. Ampliar oferta educativa en materia de datos desde la educación primaria, a educación continua
  - d. Priorizar inversión para la innovación y el emprendimiento en asuntos relaciones a datos e información
2. Desarrollo de un marco legal articulado para la explotación de datos
  - a. Articular el cuerpo normativo y jurídico
  - b. Implementar protocolos de anonimización para la adecuadas publicación y difusión de datos
  - c. Crear una mesa de revisión para evaluar la situación actual de todos los reglamentos, normas y leyes que conciernen a la explotación de datos
  - d. Generar un marco ético para el tratamiento de datos — esto puede incluir un Comisionado de Ética de Datos.

3. Generación de alianzas institucionales para la explotación de datos
  - a. Apoyar iniciativas multisectoriales para el desarrollo del ecosistema de Big Data en el país, estas pueden ser alianzas privadas, público-privadas o entre el sector público.
  - b. Discusión nacional multisectorial para unificar y generar líneas de acción que construyan sobre aquellas propuestas por el gobierno

**Acciones a largo plazo (implementación de alrededor seis años)**

1. Desarrollar productos inteligentes en red y sistemas de identificación
  - a. Establecer una identidad digital segura
2. Mejoramiento y cualificación de la demanda por información
  - a. Expandir las posibilidad de uso y consumo de datos por parte de la ciudadanía
  - b. Vincular nuevas tecnologías y procesos arraigados en datos para la toma de decisiones
  - c. Establecer canales de mejoramiento continuo para la demanda de información
  - d. Crear cooperativas de datos en pro de una cultura ciudadana de datos

## 1. INTRODUCCIÓN

### Pensamiento en Datos, Big Data, Inteligencia Artificial y Desarrollo Social

La llegada de la denominada “Revolución de los Datos” durante la última década, especialmente la relacionada con Big Data, y con la reciente emergencia de la Inteligencia Artificial, ha creado algunos temores sobre sus promesas y peligros. Muchos gobiernos alrededor del mundo, principalmente países que pertenecen a la OCDE, han venido encarando este nuevo reto. Sin embargo, aún continúa la pregunta abierta sobre qué se debe hacer ante estos cambios acelerados.

Un primer paso es comprender la esencia de lo que está aconteciendo. Tanto la “Revolución de los Datos” como la Inteligencia Artificial funcionan a partir de datos – datos emitidos por acciones e interacciones entre personas con dispositivos y servicios digitales. Por ejemplo, firmas digitales en forma de registros de teléfonos celulares, contenido web y de redes sociales, datos de geolocalización, entre otros. Estos flujos de datos permiten ser analizados por algoritmos que pueden hacer inferencias, predicciones, sugerencias y prescripciones. En este contexto, día a día recibimos múltiples incentivos para decidir qué ruta manejar, qué libros o camisas comprar, qué artículos leer o con qué personas conectarnos. Así, las formas actuales de la Inteligencia Artificial es lo que hace algunos años era llamado simplemente Big Data. Estos son modelos analíticos computacionales alimentados y entrenados por grandes cantidades de datos para alcanzar un objetivo: predecir, inferir, sugerir y reconocer.

En este sentido, cuando un automóvil autónomo está sobre la vía, el computador que lo maneja observa lo que lo rodea y se pregunta si lo que está enfrente es pavimento o personas, y actúa de acuerdo a ello, ya que de alguna forma estuvo entrenado y aprende a hacerlo por sí mismo, basado en millones de casos previos. Antes de hacerlo correcto, la máquina usualmente “entiende” su entorno de manera inadecuada

– se le dice esto, y se corrige-. Otra máquina ve una imagen de un gato y dice que es un perro. Se le penaliza por esto y se le dice que lo intente nuevamente; y lo hace. Muchas, muchas veces lo repite. A veces lo hace bien, a veces aún lo hace mal. Con el paso del tiempo y a través de múltiples iteraciones, aprende qué atributos y combinaciones de atributos de aquello que ve, están asociados con un gato versus un perro. El algoritmo encargado de concluir si es un gato o un perro reescribe las reglas para decir “hocico largo, mucho más probable que sea un perro”. En este sentido, ha estado aprendiendo; la máquina estaba aprendiendo. La esencia de Big Data y de la Inteligencia Artificial contemporánea es este aprendizaje.

Por supuesto, hay más precauciones y complejidades que estas, pero para la mayoría de los propósitos es suficiente entender que la Inteligencia Artificial actual se trata principalmente de esto: de obtener muchos datos como entradas (*big data*) y de aprender a predecir un resultado gracias al entrenamiento y a las pruebas realizadas sobre casos anteriores. Dada la ubicuidad e importancia de esta tecnología, y esta nueva forma de relacionamiento con los datos, es importante comprender cómo funcionan estos, como hacen lo que hacen y lo que podríamos aprender de las máquinas y de la tecnología para impulsar un cambio social positivo. De manera crítica, como ha planteado Alex ‘Sandy’ Pentland, la ‘buena magia’ del Big Data y de las Inteligencias Artificiales actuales, , es su función de asignación de crédito. Esto se define como la capacidad de asignar crédito para lo que “funciona”; es decir, qué características y combinación de características le permiten a la máquina (precisamente a los algoritmos que impulsa) obtener las respuestas correctas.

La computadora que trata de distinguir a un perro de un gato extraerá millones de características de la imagen que ve, las ensamblará en millones de formas y aprenderá, como en un laberinto, qué rutas (combinaciones de características) y combinaciones de caminos permiten obtener la respuesta correcta. Por lo tanto, los caminos que llevan a la respuesta incorrecta se considerarán menos valiosos, porque condujeron a una penalización (porque se equivocó). La máquina creará un incentivo para no usarlos. Por el contrario, generará un incentivo para usar aquellos que lo ayuden a obtener la respuesta correcta. Ellos llevaron a una recompensa. Con el tiempo, aprenderá qué caminos le permiten llegar hasta 99.9999...% de las veces; es decir, es capaz de “asignar crédito” a los caminos que funcionan para la tarea en cuestión.

Ahora, si aplicamos este principio general a la forma en que funcionan las sociedades, lo cual se puede denominar como una Inteligencia Artificial Humana alimentada por Big Data, lograríamos que las políticas, comportamientos, y acciones que “funcionan” sean recompensados y luego reforzados; y aquellos comportamientos que “no funcionan” sean penalizados y luego debilitados. Estos procesos también serían facilitados a través de los datos. Con el tiempo, se desarrollarían sociedades con un sentido robusto de lo que “funciona”; qué tipos de políticas, programas, comportamientos, y acciones dieron los resultados esperados. Es decir, los algoritmos usados en este proceso arrojarían lo que “funciona”, lo que debería ser recompensado y reforzado, frente a lo que no se debería - esta es la importancia para impulsar un gobierno fundamentado en datos.

La visión aquí es claramente una donde los datos alimenten sistemas sociales más eficientes y racionales al aplicar la “buena magia” de los sistemas de IA, basados en lo que “funciona” para la mayoría. Se haría entonces sólo, o principalmente, lo que contribuya a estos fines. Con el tiempo, “lo que funciona” será más frecuente; tal vez omnipresente, y las sociedades estarán mejor: más seguras, más justas, más sostenibles. Lo anterior, porque lo opuesto no “funciona” para la mayoría de las personas. No se trata solo de proponer que la información necesaria para tomar buenas decisiones llegue a quienes toman las decisiones. Más bien, se trata de proponer una sociedad arraigada en el potencial de la IA Humana y del Big Data, al basar y evaluar las decisiones, políticas, metas y otras acciones, partiendo de los datos. De esta manera, no debe haber un

consenso en qué niveles de, por ejemplo, desigualdad son adecuados para un país como Colombia, sino basar la discusión en la evidencia que arrojan los datos. Esto es posible partiendo del principio de que las políticas, acciones, resultados, y metas, deben discutirse y evaluarse sobre la base de hechos o datos concretos - dándole crédito a lo que “funciona”.

Para lograr esto se necesita desarrollar una cultura de datos fuerte y resiliente, generando interés y confianza entre el común al desarrollar capacidades necesarias en la país, tal como es la alfabetización de datos. También requiere la construcción de un sistema de gobernanza público de estos sistemas, incluyendo información y sistemas del sector privado, para así permitir el acceso ético y seguro a la información. Para esto, será necesario crear otros incentivos y medios para que organizaciones de la sociedad civil, investigadores, reguladores, entre otros; exijan que las políticas públicas y los programas del gobierno se evalúen sistemáticamente con base en los mejores datos y metodologías disponibles para ajustar iteraciones futuras y contribuir a un conjunto de pruebas sobre qué funciona y cómo. También requerirá reconstruir la confianza (y la empatía), a lo largo del tiempo, entre los ciudadanos, empresas públicas y el gobierno colombiano para que este sistema funcione.

Aprovechar la “buena magia” de la inteligencia artificial permite utilizar una herramienta que se autocorrege, basada en hechos - no es una *tecnoutopía* - es una visión humana ambiciosa que coloca buenas fuentes de datos y sistemas en el centro de un nuevo contrato social entre humanos, y entre humanos y máquinas, en las sociedades del siglo XXI. Esta es una visión de Big Data y de Inteligencia Artificial Humana que funciona para y con los humanos, que es muy diferente de una en la que las máquinas se apoderan de los trabajos o del mundo. Más bien, es dejar que la “buena magia” de la Inteligencia Artificial y el poder de los datos desafíen y mejoren los antiguos sistemas de toma de decisiones y las estructuras de poder para mejorar el estado de la sociedad, con humanos en el asiento del conductor.

El diseño de insumos para la implementación de una estrategia de Big Data para el gobierno, así como el diseño de una hoja de ruta para el fortalecimiento de la industria de Big Data en todo Colombia, tiene como objetivo puntual hacer esto, potencializar los datos – tanto el campo de Big Data como de Inteligencia Artificial –desde el sector público, para mejorar la capacidad del gobierno de promover aquello que “funciona” para el país.

#### a. Estructura del informe

El desarrollo de este informe está estructurado alrededor de la visión y el marco conceptual para la estrategia nacional de Big Data que define y explica los insumos y consideraciones importantes en la generación de líneas de acción aplicables, relevantes y coherentes para Colombia. Se propone una aproximación a los datos al clasificarlos y generar protocolos a partir de estos; a su vez, se explican las barreras y problemas en el ecosistema de datos actual, y se propone una línea de tiempo para el desarrollo de la estrategia. Adicionalmente, se exponen las lecciones aprendidas del contexto internacional para después enmarcar las acciones a corto, mediano y largo plazo. Por último, brevemente se describen los actores a cargo de la estrategia y una propuesta general de los recursos de inversión.

## 2. APROXIMACIÓN A UNA ESTRATEGIA INTEGRAL DE DATOS EN COLOMBIA

#### a. Visión

La “Estrategia Nacional de Big Data de Colombia” pretende generar condiciones favorables para el desarrollo de un ecosistema digital sólido para el crecimiento de la economía y la consolidación del gobierno digital en Colombia. La estrategia promueve el desarrollo de nuevas capacidades analíticas para el mayor uso y producción de información, y el aumento de las posibilidades de acceso a la información y datos. Esto con el propósito de incentivar la innovación y la productividad tanto del sector público como del sector privado. Al hacerlo, el objetivo es mejorar la eficiencia, efectividad y transparencia de los servicios del sector público de cara al ciudadano. A largo plazo, la meta es transformar el gobierno digital en Colombia a un gobierno movido por los datos<sup>1</sup> (*Data-driven Government*), donde todas las decisiones y políticas se evalúen y basen en información organizada, de calidad, oportuna y coherente.

En las próximas dos décadas Colombia será el país líder de América Latina en digitalización y el más avanzado en asuntos de economía y gobierno digital al expandir el impacto de la digitalización a la totalidad de las esferas de incidencia social. Además, Colombia estará a la vanguardia en la producción, uso y diseminación de información y datos, potencializando las capacidades de innovación y mejoramiento de procesos, tanto económicos como de gobierno. En ese sentido, Colombia contará con un ecosistema digital maduro, generador de beneficios colectivos en materia de eficiencia, transparencia, control y promoción de procesos de modernización e innovación, que genere nuevas oportunidades de crecimiento y cambio económico.

---

<sup>1</sup> Una traducción más aceptada para Data-driven Government es la de “Gobierno basado en Datos”. Sin embargo, consideramos que el verbo drive implica dinámica mientras que “basado” se acerca más a una idea de estática. Consideramos por ende que un “Gobierno movido por los datos” se acerca más a lo que queremos enfatizar.

Colombia, será en el curso de las próximas dos décadas, un país que genera valor estratégico a partir de sus datos e información al (1) fortalecer el ecosistema de Big Data en el país, (2) desarrollar nuevas capacidades analíticas para el mayor uso y producción de información y de mayor calidad, (3) fomentar la economía digital, y (4) promover la adopción de soluciones de Inteligencia Artificial en la administración pública, que permitan mejorar la eficiencia y eficacia en la prestación de los servicios públicos de cara al ciudadano.

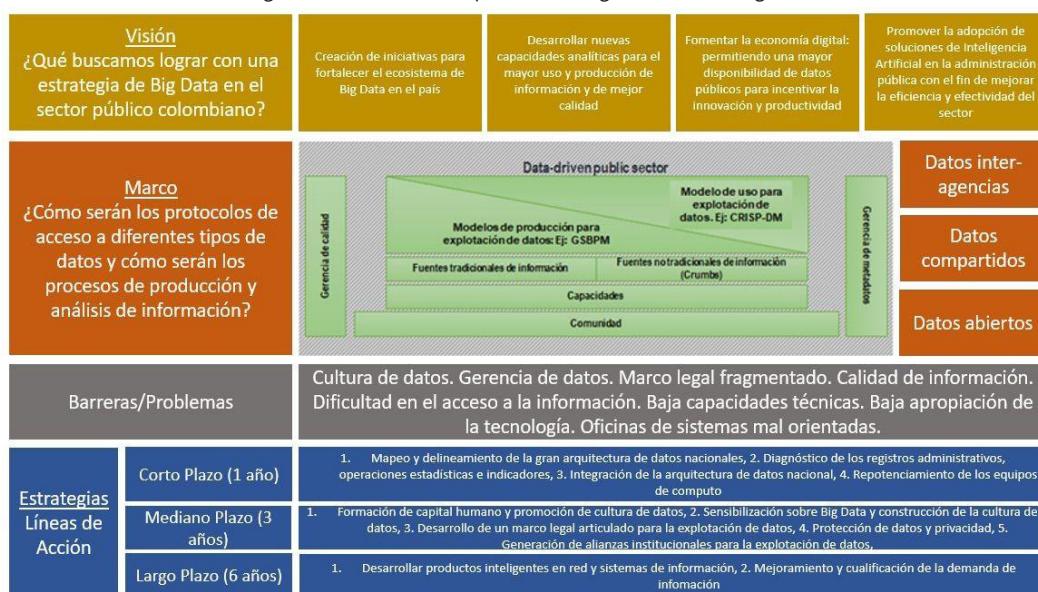
### b. Marco Conceptual

Partiendo del diagnóstico realizado en el Documento 3, estas recomendaciones e insumos para la definición de una estrategia de Big Data para el estado colombiano, buscan responder a las problemáticas identificadas, convirtiendo estos problemas en soluciones expresadas en forma de oportunidades.

El eje sobre el que se fundamenta la “Estrategia Nacional de Big Data de Colombia” es la pretensión de consolidar en el mediano y largo plazo, con un alcance duradero y estable, un “Gobierno movido por los datos”. En términos generales, esto significa que el uso más eficiente de la información se convierte en la estructura funcional del Estado, donde este ofrece de manera permanente mejores servicios a sus ciudadanos. El evitar convertirse en obstáculo, y facilitar el desarrollo de iniciativas de emprendimiento económico, son pilares de este rol, para así transformar las bases sociales y mejorar la calidad de vida de los ciudadanos.

La digitalización de la economía no es una etapa del proceso de prestación del servicio, ni un requisito de los procesos históricamente heredados, sino el punto de apoyo para la formulación, implementación, seguimiento y evaluación de las políticas públicas, de los sistemas de control y monitoreo del estado tanto en las dimensiones territoriales y sociales como digitales. La información se convierte así en el punto de apoyo para la administración territorial, de población y del relacionamiento digital que se da entre los ciudadanos y los diferentes actores de la dinámica pública.

Figura 1: Consideraciones para la Estrategia Nacional de Big Data



Fuente: Data-Pop Alliance

Con estos objetivos en cuenta, se debe priorizar la construcción de un modelo analítico que estructure las acciones para lograr estos objetivos, para luego diseñar un plan de implementación detallado y factible - las propuestas estratégicas incluidas en la siguiente sección, así como los actores responsables y la inversión necesaria, han sido diseñados como insumos, recomendaciones y líneas de acción para la estrategia nacional de Big Data

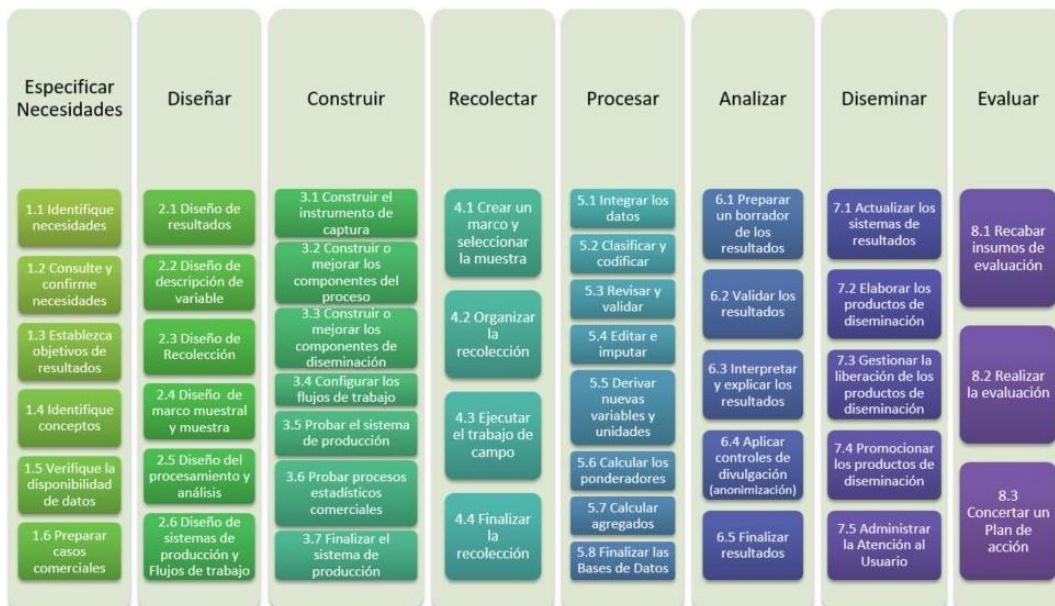
#### I. Diferencia entre los procesos de producción de información y explotación de información

Es importante diferenciar entre modelos de producción de información oficial y la generación de nuevos resultados o indicadores a partir de la explotación de datos. En ambos casos se puede aplicar Big Data, pensado como una herramienta

<sup>2</sup> El Modelo Genérico del Proceso de Producción Estadística (GSBPM) es un medio para describir la producción de estadísticas de una manera general y orientada a procesos. Se utiliza tanto dentro como entre las oficinas de estadística como una base común para trabajar con la producción estadística de diferentes maneras, como la calidad, la eficiencia, la estandarización y la orientación a procesos. <http://support.paris21.org/node/101?language=es>

y no un ecosistema. Las Comisiones de Estadísticas de Naciones Unidas o Eurostat actualmente trabajan en marcos de calidad estadística para poder integrar información no tradicional o Big Data a los procesos de generación de información estadística oficial.

Figura 2: Modelo Genérico del Proceso de Producción Estadística



Fuente: Tomado de: PARIS21 <http://support.paris21.org>

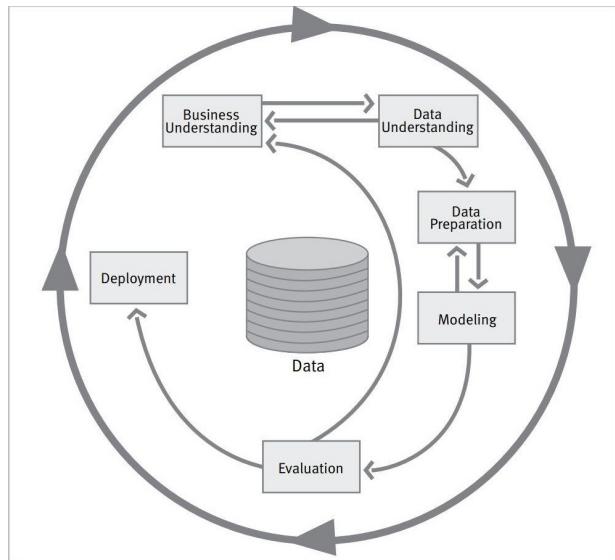
Para la producción de información oficial, existe en el mundo un modelo estándar reconocido y avalado por Naciones Unidas. Este modelo se conoce como el *Generic Statistical Business Process Model*<sup>2</sup> (GSBPM) este modelo contempla todas las etapas que debe surtir un proceso para la generación de nueva información de calidad. Además, el GSBPM es una herramienta articuladora de los procesos fundamentales que promueven y organizan no solo la digitalización, accesibilidad e interoperabilidad sino la aproximación transversal e integral a todo el ciclo de la información. Alrededor del modelo GSBPM es necesario definir las fuentes que soportan cualquier proceso de aprovechamiento de información que pueden por ahora pensarse como fuentes tradicionales de información (base de datos oficiales, registros administrativos, sistemas de información) y fuentes no tradicionales o *crumbs*, como nuevas formas de producción de información particularmente en los escenarios de Big Data.

El modelo permite identificar, desde una perspectiva de proceso, (donde no necesariamente la numeración implica requisitos de ordinalidad) las esferas fundamentales del proceso de producción de información, que proponemos debe convertirse en el eje estructurante de la estrategia nacional de datos en Colombia. El modelo propone (1) Especificar las necesidades y (2) Diseñar los procesos que hacen parte de la etapa de identificación de necesidades de información, así como de la existencia de fuentes que puedan ser esenciales para la estrategia nacional. Posteriormente viene la etapa de la producción estadística donde las dimensiones del (3) Construir, (4) Recolectar y (5) Procesar se configuran en las tareas centrales del proceso de generación de información y tendrá un impacto fundamental en la organización de grandes bases de datos, resultantes de muestras, registros administrativos y grandes cúmulos de datos. La tercera fase que incluye las dimensiones de (6) Analizar y (7) Diseminar son el punto estructurante de las nuevas tendencias donde la analítica, los datos abiertos, la innovación y el uso masivo de información - una vez se eliminan las barreras - se convierten en el objetivo principal tanto de la producción de información como la de la administración, regulación y gestión. Para finalizar, la etapa de (8) Evaluación es un constante llamado al monitoreo que permite comprender el alcance de las acciones ejecutadas y se convierta en el criterio para redefinir nuevas metas, a partir de ajustes y cambios que se derivan de los resultados de la revisión de impactos y resultados.

Otra manera de generar valor agregado a los datos, es aplicando el modelo para explotación de datos, Cross-industry Standard Process for Data Mining (CRISP-DM). Este modelo no busca en primera medida generar información oficial, sino ser un marco de referencia para desarrollar modelos analíticos o minería de datos. Actualmente es uno de los modelos más usados en el mundo.

En la Figura 3 se puede ver el flujo del modelo CRISP-DM, que consta de seis fases. Se puede observar que no es una estructura rígida y que permite interactuar entre sus diferentes fases. Las flechas indican las más importantes y frecuentes dependencias entre fases.

Figura 3: Fases del modelo CRISP-DM para explotación de datos



Fuente: <https://www.the-modeling-agency.com/crisp-dm.pdf>

El modelo tiene una naturaleza cíclica pues es un proceso constante de aprendizaje y más cuando se requiere que la explotación de datos arroje resultados estables y replicables. El CRISP-DM se divide en 6 fases principales: entender el problema, entender los datos, preparar los datos, modelamiento, evaluación e implementación.

## II. Ecosistema Nacional de Big Data

Como el resto del mundo, América Latina está en el proceso de adaptarse a la revolución digital, encontrando maneras para utilizar nuevas tecnologías y nuevas fuentes de datos para mejorar y complementar procesos estadísticos; para generar nuevos servicios; y para crear nuevas formas de relacionamiento entre el gobierno y sus ciudadanos, entre otros. El ecosistema de Big Data en Colombia, aún cuando está en constante desarrollo, se puede evaluar basado en los niveles de digitalización y penetración del internet, así como en las iniciativas y alianzas creadas para impulsarlo. En Colombia, alrededor de 61.2% de la población utilizan el internet, por debajo del promedio de la región.<sup>3</sup> El ecosistema de Big Data en el país se ha venido desarrollando desde principios de esta década, siendo incluido como una prioridad política e integrado en planes de desarrollo. Es importante que la política pública orientada a la explotación de datos esté pensada como ecosistema y no como unidades independientes que en algún punto se deben complementar.

### c. Barreras y problemas

La identificación de las barreras y problemas en cuanto al ecosistema de datos en el país fue analizado en el documento 3, “Diagnóstico de la situación actual de Colombia”, estructurado alrededor de las seis dimensiones propuestas para esta consultoría. Estos son expuestos en detalle en la próxima sección, pero incluyen la baja cultura de datos en el país, bajos niveles de datos públicos, la inadecuada gerencia de datos, un marco legal fragmentado, dificultad en el acceso a la información y la baja calidad de esta, bajas capacidades técnicas, baja apropiación de la tecnología y la ausencia de lineamientos y guías en la gestión de los datos. Esta son líneas bases para el monitoreo y seguimiento de los impactos de las acciones propuestas en los años venideros. Los problemas requieren atención expedita con el fin de reparar los asuntos que afectan la funcionalidad del ecosistema de datos para garantizar que las nuevas acciones se construyan sobre un escenario adecuado, evitando precariedad y debilidad en el futuro.

### d. Línea de tiempo

Para estructurar la Estrategia Nacional de Datos se mencionan los momentos temporales considerados relevantes para la implementación de la agenda. En primer lugar, el Estado colombiano debe tomarse un tiempo de preparación (*Setting up*) que le permita acomodar equipos, alcanzar consensos, identificar necesidades, conocer el estado detallado de la infraestructura nacional de datos e implementar acciones a corto plazo. Para este primer período consideramos que un año es el tiempo adecuado para la preparación de los equipos y la implementación de acciones estructurales para el ecosistema de Big Data - estas estrategias son a corto plazo. Un segundo tiempo es el de implementación de acciones a mediano plazo

<sup>3</sup> Para comparar regionalmente la penetración de internet, visitar

que requieren atención prioritaria. Se considera que en un espacio de 3 años se pueden alcanzar primeras victorias (*early wins*), donde se pueden identificar el alcance de las acciones y tomar decisiones sobre continuidad, ajustes o suspensión de la agenda de implementación para ciertas acciones. Finalmente, hay unas tareas que se deben programar para un espacio de seis años, a largo plazo, que implica pensar en la longevidad de las mismas y la supervivencia del ciclo electoral, con el fin de garantizar que no se detengan en el cambio de gobierno siguiente. Estas acciones también se construyen sobre las acciones propuestas a corto y plazo. En ese sentido las acciones que presentamos a continuación como recomendaciones e insumos las proponemos en el marco temporal anteriormente descrito.

### **3. FUNDAMENTOS PARA UNA ESTRATEGIA NACIONAL DE BIG DATA**

Para proponer una estrategia nacional de datos para Colombia se requiere de un conocimiento detallado de los avances en materia de datos en el mundo, así como las hojas de rutas existentes en cuanto a estrategias digitales en el país. Durante esta asesoría se han producido tres documentos que sirven como insumos fundamentales para el objetivo de la estrategia nacional. En primer lugar, un (1) “Estado del arte y análisis comparativo de estrategias nacionales de Big Data”; posteriormente un análisis a profundidad de (2) “Buenas Prácticas sobre cómo los Gobiernos pueden implementar una estrategia para estimular el ecosistema de negocios Big Data”; y finalmente, un trabajo específico para el caso colombiano, (3) “Diagnóstico de la situación actual del Estado Colombiano estructurado alrededor de las seis dimensiones”.

A continuación se presentan algunas de las lecciones aprendidas y estrategias utilizadas en el contexto internacional, seguido de los puntos claves del diagnóstico de la situación actual del caso colombiano frente a la explotación de datos desde el sector público. Estos sirven como modelos e insumos claves para guiar las acciones estratégicas propuestas para la estrategia nacional de Big Data. Comprender qué ha funcionado, cómo se pueden adoptar en Colombia propuestas exitosas de otros países e identificar prioridades que permitan desarrollar un gobierno digital en Colombia, y a posterior uno basado en datos, son consideraciones esenciales para producir una hoja de ruta tangible, sostenible y adecuada para el país.

#### **a. Lecciones Aprendidas del Contexto Internacional**

Partiendo del análisis comparativo de los casos de estudio, se reconocen varias estrategias aplicadas por los países estudiados<sup>4</sup> que han resultado en avances importantes para la consolidación de su infraestructura de Big Data. En general, el desarrollo de la industria de datos en estos países ha iniciado del objetivo de asegurar la eficiencia del gobierno y de ajustarse a la revolución digital, utilizando servicios digitales y procesos de gobierno electrónico para abordar desafíos económicos y sociales. Al enfocarse en eficiencia gubernamental, estos países han desarrollado estándares de gobernanza de datos, marcos legales de protección de datos y privacidad, así como propuestas de innovación e integración multisectorial. Entre las estrategias y programas implementados por estos países se destacan:

##### **i. Propuesta de identidad digital de Estonia**

La propuesta de identidad digital del gobierno de Estonia es un ejemplo de la utilización de canales digitales arraigados en Big Data para ahorrar tiempo y recursos financieros. Esta iniciativa se basa en un sistema para autenticar a las personas y a las empresas, sin necesidad de contacto físico, con el fin de proporcionarles un acceso digital seguro que permita acceder a innumerables servicios - este sistema se llama X-Road. Esta infraestructura de información digital permite que los ciudadanos, instituciones públicas y compañías privadas intercambien información de manera segura. Algunos de los servicios que esto ha permitido es declarar impuestos online, poder revisar historias clínicas, empezar una compañía en tan solo 18 minutos, e incluso votar en línea (Jaffe 2015).

En el caso de Estonia, el tener una identidad digital implica tener una base de datos de registros nacionales, así como establecer una red segura de intercambio de datos - elemento esencial en la estructura de Big Data. Establecer una identidad digital segura, según el Profesor Alex ‘Sandy’ Pentland, es ir más allá de la simple noción de una forma de identificación (única individual) digital, sino es proteger y habilitar a los ciudadanos. Una identidad digital segura elimina las posibilidades de fraude y robo de identidad, tiene la posibilidad de reducir la corrupción, incrementa los ingresos tributarios, y en general, genera un ambiente propicio para desarrollar el comercio y la economía digital. En este caso, el país fue capaz de desarrollar esta identidad digital al priorizar su implementación a corto plazo con empresas grandes, a mediano plazo con pequeñas y medianas empresas, y a largo plazo extenderlo a todos los ciudadanos.

##### **ii. El Reglamento General de Datos de la Unión Europea (GDPR)**

Entre los casos estudiados, solo el Reino Unido y Estonia cuentan con órganos independientes para garantizar la aplicación de la ley en cuanto a la protección de datos personales. En el Reino Unido, la Oficina del Comisionado de Información (ICO) es el organismo a cargo de ayudar a las personas y ante todo, a las empresas entender y adaptar a nuevos cambios normativos, como lo fue el Reglamento General de Datos de la Unión Europea (GDPR), vigente desde mayo de 2018. Este es considerado como uno de los conjuntos de normas más completos en cuanto a la protección de datos, basado en el principio fundamental

<sup>4</sup> Estos incluyen Reino Unido, Singapur, México, Corea del Sur, Estados Unidos y Estonia

<sup>5</sup> Para más información, ver “Definición de la estrategia de Big Data para el estado colombiano y para el desarrollo de la industria de Big Data en Colombia” pg. 85

de “devolver a los ciudadanos el control sobre sus datos personales” al incluir derechos en cuanto al acceso, rectificación, enmienda, restricción, portabilidad, objeción y el “derecho a ser olvidado”, entre otros<sup>5</sup>. Este reglamento introduce sanciones a empresas por incumplimiento con sus normas, y en general, pretende garantizar la privacidad y promover prácticas éticas en cuanto al manejo de datos.

### **iii. Promoción de datos abiertos**

En los países estudiados, los datos abiertos son una plataforma para impulsar el emprendimiento de datos, generar nuevas oportunidades económicas y crear herramientas de innovación. Corea del Sur, específicamente, se ha centrado en el aprovechamiento de los datos abiertos al poner a disposición del público cantidades masivas de datos del gobierno; esto ha permitido crear aplicaciones digitales que vuelven más eficientes los servicios del mismo. El derecho de los ciudadanos a utilizar datos públicos abiertos está dictado por la Ley de Provisión y Uso Activo de Datos Públicos que establece la obligación a las agencias del sector público a proporcionar datos. Ofrecer grandes cantidades de datos abiertos tiene la posibilidad de promover una apertura económica, fomentar el empleo, así como mejorar servicios de bienestar social.

### **IV. Construcción de capacidades humanas y de capital**

Una de las lecciones aprendidas más importantes del estudio es la importancia de impulsar las capacidades humanas y de capital, eslabón fundamental en la infraestructura de Big Data. Singapur, por ejemplo, ha invertido en programas de alfabetización de datos para sus ciudadanos, y específicamente en capacitar a los funcionarios del Estado; el hacerlo no solo garantiza *buy-in* por parte del sector público, sino una inversión en capacidades a nivel gobierno. Además de esto, Singapur ha desarrollado actividades específicas en el sector de Big Data, incluyendo *hackatones*, *innovation hubs*, *incubadores*, plataformas y redes tecnológicas, e incluso provee incentivos para que empresas emergentes creen aplicaciones que los ciudadanos puedan utilizar con datos extraídos del gobierno. La construcción de capacidades es esencial para generar demanda de datos y consecuentemente, para generar valor a partir de estos.

### **V. Inteligencia Artificial**

En cuanto a Inteligencia Artificial, este tema no fue incluido en los documentos 1 y 2, al ser un campo en desarrollo. A partir de un análisis de las iniciativas y estrategias de AI en 22 países alrededor del mundo<sup>6</sup> se pueden notar avances en cuanto a la incorporación de estrategias nacionales que buscan desarrollar la industria de AI. Entre las estrategias propuestas, se destacan aquellas que potencializan los datos abiertos, la creación de un marco regulatorio y normas de ética, inversiones en investigación y desarrollo, y el desarrollo de capacidades - estrategias estructuralmente parecidas a las de Big Data. El interés de estos países en ser líderes mundiales en cuanto a esta tecnología aún emergente, demuestra el potencial de utilizar datos como un activo nacional. Sin embargo, es importante tener en cuenta que el desarrollo de AI y capacidades relacionadas, depende enteramente de un ecosistema de datos saludable y productivo.

Las propuestas detalladas en esta sección han sido utilizadas exitosamente para el desarrollo de la industria de datos en cada uno de los países descritos. El adaptar modelos similares, o basados en estos éxitos es una de las propuestas de esta estrategia, detallado en la sección de Acciones. Estimular el sector naciente de Big Data, usando la experiencia de otros países que ya han implementado estrategias exitosas, es por un lado un reto, al considerar las diferencias en cuanto al contexto, presupuesto, y cronograma de estas estrategias. Sin embargo, proporciona unos fundamentos para el desarrollo de la estrategia nacional.

#### **b. Diagnóstico del caso colombiano**

Las recomendaciones e insumos para una nueva estrategia nacional de datos no pueden construirse sin la inclusión de los avances realizados en el “CONPES 3920: Política Nacional de explotación de datos” o con los resultados del diagnóstico realizado en el documento 3 de esta consultoría. Estos dos insumos diagnosticaron **bajos niveles de datos públicos**, lo que implica tareas importantes en el fomento a la producción de información bajo arquitecturas contemporáneas que faciliten el intercambio y la disponibilidad de la información tanto para el gobierno como para el público. Adicionalmente, se identificaron la **falta de confianza e incertidumbre sobre el uso de los datos** como barreras notables para el adecuado aprovechamiento de los datos. Por lo tanto, se han priorizado tareas en materia de normatividad y seguridad dentro de los lineamientos de ese documento. Se consideró también, que las **bajas capacidades técnicas** se convierten en un reto para el establecimiento de una hoja de ruta que fortalezca, a partir de recursos físicos y de talento humano, las condiciones para el aprovechamiento de datos en el país. Por último, la **ausencia de una cultura de datos** se identificó como una barrera que inhibe la institucionalización del uso de datos en la toma de decisiones.

Adicionalmente a lo largo del trabajo realizado se han identificado seis dimensiones fundamentales para la acción inmediata y de largo plazo que garantice sostenibilidad en la política nacional de Big Data. Las dimensiones, sin ningún orden jerárquico, son:

**Dimensión 1: Marco legal y ético Dimensión 2: Organizacional e institucional**

**Dimensión 3: Infraestructura física (tecnología) orientada a datos Dimensión 4: Infraestructura de datos**

**Dimensión 5: Capital humano y Cultura alrededor de los datos Dimensión 6: Gasto e inversión asociada a la explotación de datos**

<sup>6</sup> Publicado como Anexo A al documento dos de esta consultoría.

Durante la fase de diagnóstico, que se presentó en el documento 3, se analizó para cada una de las dimensiones los elementos que requieren atención para ser solucionados por cualquier estrategia que se proponga en los próximos años. De manera sintética se identificaron las carencias que existen en cada una de las dimensiones, las dificultades que explican dichas carencias y los obstáculos que hasta el momento han enfrentado los actores relevantes en cada una de las dimensiones. La estrategia que se empieza a esbozar en este documento pretende impactar los problemas de más relevancia que se identificaron en cada una de las dimensiones estudiadas.

La estrategia de Big Data debe seguir un camino específico basado en las dimensiones analizadas para el caso del Estado colombiano. En primer lugar, aparece la **infraestructura física** como el requerimiento básico de confluencia de los datos, las capacidades y las comunidades. Precisamente, es sobre esta infraestructura que opera cualquier iniciativa de datos que proponga desarrollar el gobierno colombiano. En este punto vale la pena resaltar que no se trata solamente de un rol de provisión de infraestructura física (que es muy necesaria al interior de las entidades del Estado y en algunos casos de interoperabilidad e intercambio), sino de un rol de regulación y coordinación para que se generen incentivos para la existencia y funcionamiento adecuado de una infraestructura que responda a las necesidades del país y, a su vez, esté en capacidad de conectarse con los requerimientos técnicos globales.

Por lo tanto, la infraestructura física es inocua sin la apropiada existencia y sostenibilidad de una **arquitectura de datos** del Estado colombiano que logre influir de manera positiva sobre los otros sectores de la economía. En ese orden de ideas, las migajas (*crumbs*) de información que se encuentran distribuidas a lo largo y ancho del aparato estatal colombiano deben organizarse de manera estructural con el objetivo de funcionar de manera articulada y coordinada para soportar y mejorar la capacidad del estado de responder a sus funciones. Cualquier esfuerzo por consolidar una arquitectura de datos nacional impactará de manera positiva no solo los escenarios descentralizados del Estado sino los sectores no gubernamentales, como lo son el sector privado u de organizaciones no gubernamentales, dado que tendrán a su disponibilidad un referente de operación para el manejo fundamental de datos.

Para el debido manejo de la infraestructura física y de una arquitectura de datos específica es requisito una **organización e institucionalidad** que priorice la implementación de acciones encaminadas al cumplimiento de una estrategia de Big Data. Solo a través de la capacidad organizativa y de un aparato institucionalizado se puede garantizar sostenibilidad de la estrategia de Big Data para el país. Ambas iniciativas deben estar acompañadas de un **marco normativo y ético** sólido que permita el debido funcionamiento de las acciones en el marco de lo legal y que incentive, promueva y genere externalidades positivas a partir del comportamiento ético de la sociedad. Lo anterior con el objetivo de promover el uso de datos y el uso de información para plantear soluciones adecuadas a sus condiciones.

Asimismo, comprender que una estrategia de datos requiere de la participación activa y amplia de los actores de la sociedad es fundamental para su éxito. En ese sentido, es necesario que se establezcan **alianzas** con el sector privado, academia y organizaciones no gubernamentales, con el objetivo de articular, cooperar, coordinar y establecer mecanismos de unión que lleven a fortalecer alianzas que profundicen en las externalidades colectivas de producir un bien público, como lo es la información. Además, las alianzas contribuyen a la sostenibilidad de la estrategia en general en el mediano plazo, a partir de la observación, la transparencia y el seguimiento a los procesos que implementan los diferentes actores.

Adicionalmente, la **capacidad comunitaria y el alfabetismo de datos** se convierten en un requisito fundamental para el buen desempeño de una estrategia de datos. En cierto sentido, se puede afirmar que no podrán existir ciudades inteligentes sin ciudadanía inteligente, lo cual implica un esfuerzo muy grande de formación de capacidades en las comunidades, un conocimiento de los alcances, las herramientas y los instrumentos a través de los cuales se puede hacer uso de la información. En un proceso de mayor profundización, son los usuarios los llamados a identificar los procesos, algoritmos y acciones que lleven a nuevas formas de abordaje de la información, de manera que se fomente una mentalidad de constante **innovación** y se mejoren los procesos de acceso, recolección, resguardo, procesamiento y uso de la información.

El Departamento Nacional de Planeación ha propuesto el trabajo de definición de una política nacional en materia de explotación de datos a partir de los avances, que, en materia de datos se ha llevado a cabo en el país en los últimos años. En primer lugar, vale la pena destacar el avance que se ha hecho en gobierno electrónico, el cual ha alcanzado niveles de digitalización y sistematización de gran parte de la información que se administra en el Estado. Lo anterior a partir de sistemas que recogen información de las múltiples entidades públicas y que facilitan el mejoramiento en las capacidades básicas informáticas de las unidades administrativas del mismo. Aun cuando es necesario mejorar la calidad de esta información, al enfatizar requerimientos contemporáneos como la conectividad, interoperabilidad, apertura y confiabilidad de la información – entre otros – se logran delinear las formas requeridas para el acceso y utilización de la información.

Adicionalmente, los retos en materia de gestión documental han implicado nuevas arenas de interacción entre el Estado y las nuevas tecnologías, lo que ha permitido el desarrollo de capacidades institucionales alrededor del reto de administración de los documentos de gobierno. Así, las tecnologías se han incorporado a algunos procesos administrativos existentes generando nuevas formas de interacción con documentos físicos y digitales. Las nuevas formas de manejo de la información en entidades públicas han significado un cambio en el manejo de correspondencia hasta en la administración de registros administrativos, convirtiéndose en una arena de incidencia e impacto de las nuevas tecnologías de información.

Finalmente, es importante anotar los avances logrados en materia de protección de datos personales, los cuales apuntan a mejoras en temas de confianza y certidumbre y aportan al pilar de normatividad y ética como requisito para una estrategia nacional sólida de datos. En este contexto, los últimos años han estado marcados por la promoción de la normatividad referente a protección de datos personales y la implementación de acciones específicas que buscan garantizar el cumplimiento de dichas normas. El proceso de implementación de la normatividad ha generado interesantes debates, promoviendo conversaciones sobre el adecuado uso de la información y las responsabilidades que se derivan del manejo de grandes grupos de datos en un contexto tecnológico y normativo contemporáneo. El reto es grande; por ende, a continuación presentamos las recomendaciones e insumos que consideramos fundamentales para la formulación de una estrategia nacional de Big Data para Colombia.

## 4. ACCIONES

### a. Corto plazo - construcción de confianza y del momento

#### i. Mapeo y delineamiento de la arquitectura de datos nacional.

El punto de partida para la definición de la estrategia nacional de Big Data consiste en un esfuerzo de gran envergadura para la identificación de los sistemas de información, registros administrativos y datos, operaciones estadísticas e indicadores existentes en el país. Es necesario conocer todas las fuentes de información públicas disponibles a nivel gubernamental, la forma como se produce la información, los metadatos de la información, los lugares de almacenamiento, el uso y difusión que se les está dando, y los futuros retos que enfrentan cada una de estos hasta el momento. Esta línea de acción es clave para poder determinar los procesos que se deben estructurar en un futuro cercano para lograr la integración y el eficiente relacionamiento de estos datos y sistemas. Para hacerlo, se necesita un mapeo del flujo de información estructurada y no-estructurada, desde la perspectiva de su administración para poder caracterizar deficiencias o avances y construir una arquitectura de datos líder. Este esfuerzo es transversal a toda la estrategia propuesta, y de no ser posible un mapeo exhaustivo y simultáneo, se debe iniciar esta acción con los datos, sistemas de información, registros administrativos y operaciones estadísticas que sean de alta prioridad. El Cuadro 1 expone dos casos de cómo otros países han abordado acciones similares.

#### Cuadro 1. Inventario de Datos ¿Por Dónde Empezar?

Tradicionalmente, los inventarios de datos son utilizados por instituciones para realizar un mapeo de los datos estructurados y no-estructurados que se manejan - saber qué tipo de información se recopila, dónde está almacenada, con quién se comparte y cómo se transfiere, es un componente central de cualquier estrategia de datos, particularmente en términos de privacidad y seguridad de los datos (Zetoony 2018). Sin embargo, también es un ejercicio esencial para reconocer cómo se puede generar valor agregado a partir de los datos existentes, y para reconocer los vacíos que se necesitan complementar para procesos de analítica avanzada, o para la toma de decisiones.

#### i. Inventario de Datos y Cumplimiento con el GDPR

El Reglamento General de Protección de Datos (para más información ver sección III.a.ii) requiere que las empresas conserven registros detallados de las actividades que involucren el procesamiento de datos personales. Por este motivo, las empresas deben realizar un ejercicio comprensivo de mapeo e inventario de datos para poder justificar el movimiento de estos datos desde el punto de recolección hasta cuando llegan a proveedores externos (Heimes 2018). El GDPR reglamenta que se conozca el propósito del procesamiento de datos personales, una descripción de la categoría de datos utilizada, información de los recipientes o individuos con quien se comparte los datos personales, y descripciones sobre la transferencia de estos datos, entre otros. Algunas de las consideraciones operacionales al realizar un inventario de datos, según la firma PwC, es entender cómo se va a utilizar el resultado del mapeo para la organización, realizar el inventario de una manera eficiente donde la proporcionalidad del esfuerzo corresponda a su uso, definir de quién es la responsabilidad de actualizar el inventario, y si se quiere limitar el inventario a datos personales, o aprovechar para mapear otro tipo de datos (Creegen 2018).

Según la Asociación Internacional de Profesionales de la Privacidad, las herramientas y los métodos para realizar un mapeo de datos depende del tamaño de la organización y la complejidad de los datos que se alberguen. Para organizaciones con grandes volúmenes de datos distribuidos, soluciones de datos maestros Master Data Management (MDM), donde se crea una única fuente de referencia maestro para todos los datos críticos, pueden ser útiles para reducir las redundancias y las fuentes de información (Informatica 2018).

Idealmente, el proceso de mapeo permite identificar la ubicación e información de los datos a una granularidad individual (i.e., ¿que datos tengo sobre la persona X? ¿puedo encontrar todos los datos de la persona X?). Sin embargo, al ser este inventario en respuesta al cumplimiento de la ley y los requerimientos del GDPR, se limita a datos personales, un punto de partida valioso si se prioriza la seguridad y privacidad de los datos.

## ii. Inventario de Datos Abiertos de Canadá

Construir un inventario de datos es requerido por la sección 6.3 de la Directiva de Datos Abiertos del gobierno canadiense. Este establece la necesidad de crear y mantener inventarios comprensivos sobre los datos o información que contengan business value o valor comercial de cada entidad para determinar su elegibilidad para ser incluida en el portal de datos abiertos, así como la prioridad de liberación de los datos. Este inventario de datos es publicado en el portal de datos abiertos del gobierno, para permitir que los ciudadanos conozcan qué datos son recolectados, y tengan la oportunidad de indicar cuales de estos datos son de más utilidad, asegurando que la liberación de los datos sea en función tanto de la demanda externa como de la capacidad interna (Government of Canada 2018).

El objetivo de este inventario es mantener un mapa de los datos del gobierno, reconociendo que no todos pueden ser liberados, pero dando acceso a la mayor cantidad posible. Cada departamento recibe una plantilla guía para realizar el inventario.

Recomendamos partir de los avances que hasta el momento ha tenido el **Sistema Nacional Estadístico (SEN)**, como el organismo encargado de la articulación y coordinación de información estadística oficial a nivel nacional. Este está compuesto por las entidades que producen y difunden estadísticas y son dueñas de registros administrativos con potencial uso estadístico. Reconociendo la capacidad de gerencia del SEN, se propone que éste coordine los esfuerzos de mapeo y evaluación como punto de partida de la estrategia de Big Data. La Figura 4 sugiere los aliados del SEN para realizar el mapeo del funcionamiento y los estándares de calidad para cada proceso/tipo/sistema de información y datos. Es esencial definir el estado del arte en cuanto al funcionamiento de estos, identificando oportunidades de mejora y prioridades para la consolidación de la infraestructura de datos en el país.

Por otro lado, el SEN tiene la fortaleza de que ha sido una iniciativa técnica por parte del DANE que ha permitido la estandarización, normalización y aseguramiento de la calidad de múltiples operaciones estadísticas en el país. Al articular lo anterior, el SEN debe ser un escenario de encuentro entre la estadística, la información, la política pública y las necesidades del país.

Figura 4: Coordinación SEN para mapeo y delineamiento



Fuente: Data-Pop Alliance.

El que el SEN esté a cargo del mapeo de información busca mejorar la coordinación y la eficiencia de la producción y difusión de datos, al centrar estas acciones en un solo actor; y encargar a un sólo organismo de evaluar la interoperabilidad y la calidad de los sistemas, bases de datos, y en general, fuentes de datos nacionales. Esto implicaría que el SEN, como en el caso de Estonia donde la gestión de la iniciativa de identidad digital está descentralizada pero supervisada por una identidad, tome este rol. De esta manera, el SEN coordinaría el intercambio de información necesario para llevar a cabo el diagnóstico sobre la situación actual de la arquitectura de datos nacional.

Generalmente, las oficinas nacionales de estadística son organismos menos politizados en los que la ciudadanía confía. Se propone entonces que el SEN recopile un inventario de los datos de otras entidades - estos siempre estarían alojados en los sistemas de la entidad correspondiente - en vez de ser “dueños” de esta información, para que este mandato se haga menos objetable.

El SEN se convierte en el mecanismo a través del cual se articulan los esfuerzos entre el sector público, privado, academia y organizaciones no gubernamentales en materia de datos. El mapeo de las bases de datos, registros administrativos, fuentes de información no estructuradas e indicadores del país, debe ser parte de un proceso continuo por parte de un equipo concentrado en actualizarlo de manera constante, y de estar haciendo seguimiento a los cambios que se van haciendo a lo largo del tiempo en términos de la generación de información nacional. El mapeo de la información del país sirve como base fundamental para la **construcción de una arquitectura nacional de datos**, que se debe delinear inicialmente para conocer elementos integradores, articuladores y atributos de interoperabilidad para exigir la construcción de espacios para el almacenamiento, intercambio, uso y difusión de la información.

En este orden de ideas, el Plan Nacional de Desarrollo debe incluir que en los próximos años (no más de dos), se debe contar con un equipo encargado del monitoreo constante de las necesidades de información del país - este debe contar con el SEN como ente coordinador. Este equipo debe haber entregado como resultado una primera evaluación del estado actual de los sistemas de información y bases de datos, las necesidades de mejora y los retos de cada una de estos. Este trabajo debe realizarse en conjunto con la priorización de acciones para el mejoramiento constante de la capacidad técnica y humana para el manejo de la información.

## **1. DIAGNÓSTICO DE LOS REGISTROS ADMINISTRATIVOS, OPERACIONES ESTADÍSTICAS E INDICADORES**

A la par con el mapeo de sistemas de información, registros administrativos y datos, operaciones estadísticas e indicadores existentes en el país, se debe realizar un diagnóstico sobre estos, para determinar no sólo qué tipos de datos y sistemas se están utilizando, si no la calidad de estos. Lo anterior, para sugerir mejoras, procesos de actualización, y poder resaltar las áreas de baja digitalización que requieren prioridad.

## **2. INTEGRACIÓN DE LA ARQUITECTURA DE DATOS NACIONAL**

Al final del proceso de mapeo y diagnóstico se debe definir cómo se van a integrar todos los tipos de datos, tanto estructurados como no-estructurados, relevantes para el desarrollo del ecosistema de Big Data en el país. La integración de estos requiere de estandarización, calidad e interoperabilidad, así como la inclusión de protocolos de anonimización. Se debe responder dónde van a residir los datos, si en un modelo descentralizado con una identidad rectora sobre estos, o con protocolos de acceso dependiendo de su tipología. De igual manera, se deben incluir las condiciones de cómo los datos nacionales se deben integrar y compartir con datos de entidades privadas, la academia o sociedad civil. De igual manera, se debe decidir sobre quien recae la gobernanza, mantenimiento y actualización de la infraestructura de datos nacional.

### **Cuadro 2. Infraestructura de Datos - Estándares Internacionales**

El mayor desafío de crear una estrategia de aprovechamiento de datos es encontrar la estructura para procesar datos de manera coherente y organizada, particularmente cuando el formato de los datos a explotar está atado al modelo operacional de una agencia o entidad, como usualmente es el caso de los gobiernos. Hay poca información sobre la infraestructura de datos que otros países han desarrollado; en general la inversión en esta área ha sido dirigida a construir plataformas de construcción de modelos, pero muy poca para herramientas de ingeniería de infraestructura de datos (Chamakkala 2018).

Aun cuando las necesidades de aprovechamiento de datos y de analítica de cada entidad son diferentes, los siguientes requerimientos son transversales para una iniciativa de Big Data (Lockner 2017, Klievink et al. 2017, Perret 2014, "Making Sense of Big Data" 2016)<sup>7</sup>:

#### **i. Almacenamiento:**

En la mayoría de los casos, las organizaciones ya cuentan con el almacenamiento necesario para respaldar una iniciativa de Big Data, sin embargo, es prioritario asegurarse que las soluciones de almacenamiento no solo cuenten con el espacio necesario, sino que también pero que estén optimizados para reducir la latencia en las transferencias de datos. Las organizaciones deben decidir entre almacenamiento físico o en la nube.

#### **ii. Procesamiento**

Los servidores utilizados deben tener suficiente poder de procesamiento para respaldar las actividades que se realizan sobre estas. Se puede escoger entre procesadores en la nube, con capacidades de procesar datos "on-demand" y sin necesidad de ser propietarios de estos, o con procesadores físicos. Es necesario evaluar si es necesario contar con una plataforma que pueda soportar procesos transaccionales y analíticos, en vez de contar con workload-specific systems (sistemas con capacidades de trabajo específicas).

### **iii. Software de analítica**

Estos se escogen con base en las funciones que se deben realizar, teniendo en cuenta factores como la seguridad y la facilidad de uso del software. Estos son particularmente importantes para entidades del gobierno en cuanto a su habilidad de realizar analítica predictiva para la detección de fraude, planeamiento estratégico, entre otros.

### **IV. Capacidad de red**

El alto volumen de datos que se transporta por una infraestructura de datos requiere de hardware robusto con capacidad de respaldar conexiones rápidas. De acuerdo a FedTech, muchas organizaciones (en Estados Unidos) ya cuentan con la conectividad y capacidad de red que facilitan conexiones de 10 gigabytes, y que requieren solo de modificaciones menores, al instalar nuevos puertos de conexión, requieren solo de modificaciones menores. Es esencial asegurar la privacidad y seguridad de la red, pues es el punto clave por donde se transporta información de carácter reservado.

Por otro lado, IBM reconoce la velocidad, acceso y disponibilidad como las tres variables claves a tener en cuenta para construir o actualizar las arquitecturas de datos. Se requiere tener en cuenta como los componentes de una arquitectura de datos afectan la velocidad con la que se producen los análisis o insights - minimizar el tiempo de demora entre la recopilación de datos en bruto, y la toma de decisiones basada en datos es uno de los objetivos principales de una infraestructura. De igual manera, se requiere garantizar el acceso a todos los tipos de datos relevantes, sin importar donde residen. Por esta razón, la infraestructura de datos debe ser flexible, escalable y diseñada para la integración de diferentes tecnologías. Por último, utilizar Big Data y analítica avanzada es solo útil cuando se puede proporcionar la información requerida a quienes están tomando decisiones en el momento adecuado (Perret 2014).

La infraestructura de datos es el bloque fundacional sobre el cual se construye un ecosistema de datos que habilita la explotación de datos hacia un data-driven government. Estos requerimientos coinciden con la cadena de valor de Big Data (para más información, ver Anexo C) pues la infraestructura es el punto de partida para recolectar, analizar, almacenar y explotar los datos.

### **Cuadro 3. Retos al implementar una Infraestructura de Datos Nacional (NDI)**

Según Kievinse et. al (2017), una infraestructura digital se refiere a los sistemas compartidos y heterogéneos que cambian y evolucionan gracias a la tecnología y a sus actores; estos incluyen colecciones de componentes técnicos interconectados e independientes, que operan a través de límites organizacionales. Los componentes claves para definir una infraestructura de datos son los “sistemas, redes, procesos de intercambio, modelos de datos, taxonomías, servicios web, instalaciones de autorización y autenticación, entre otras”.

El reto principal para construir una infraestructura para el gobierno, es que no se trata sólo de implementar nuevas tecnologías, pero de adoptar una visión integrada de todos los componentes (materiales e inmateriales) que han sido desarrollados en las últimas décadas para respaldar los procesos de transformación digital, tal como lo son las estrategias de gobierno electrónico y la interoperabilidad.

El reto, entonces es cómo integrar una estrategia ambiciosa a los múltiples elementos habilitadores que ya existen, pero que no han sido integrados a nivel nacional. Básicamente, como lidiar con la descentralización de los bloques básicos de la infraestructura de datos, así como con la fragmentación en control y propiedad de estos. Lograr un balance entre una infraestructura de datos que cumpla las necesidades y requerimientos del gobierno, pero que al mismo tiempo cumpla las de las entidades y organizaciones con quien se relaciona el gobierno, debe ser un objetivo fundamental.

En particular, se requiere esclarecer la gobernanza de la infraestructura, que puede ser entendida en tres componentes: quién tiene jurisdicción sobre las decisiones, cuáles son los mecanismos de control que se van a implementar, y sobre quien recae la propiedad de esta.

<sup>7</sup> La mayoría de la información relevante a la infraestructura de datos se concentra a nivel empresarial

### **3. PROTOCOLOS ASOCIADOS AL ACCESO A DATOS DE DIFERENTES TIPOS**

Con el fin de promover una explotación de datos ordenada, flexible, segura y dinámica en el país, es necesario clasificar o identificar protocolos de acceso a los datos según el grado de sensibilidad de la información que se maneja. En la medida en que los datos son más sensibles se requiere de tecnologías y protocolos más rigurosos para poder acceder a ellos, restringiendo incluso el tipo de usuario o entidad que requiera esta información según el nivel en el que se encuentre.

La creación de estos protocolos, permite generar la confianza a los ciudadanos con respecto a la gobernanza y protocolos de acceso a esta información. Es por eso que se recomienda su manejo por autoridades o entidades que gozan de tal confianza o respaldo jurídico para manejo de información a nivel personal. Por lo general, en el mundo, hay una fuerte confianza en los sistemas estadísticos nacionales, que si bien, no son una entidad como tal, si gozan de prestigio, rigurosidad y buenas prácticas en el manejo de información.

A continuación se encuentran descritas las acciones y protocolos necesarios dependiendo del tipo de dato manejado por el estado.

#### **1. Datos Abiertos**

La política de datos abiertos ya se encuentra definida por el Ministerio de las TIC en la Ley de Transparencia y Acceso a la Información (Ley 1712 de 2014). Estos datos se caracterizan como datos primarios o sin procesar “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”. Por lo tanto, aunque este protocolo ya se encuentra implementado, se hace necesario establecer una línea de acción para ejecutar estrategias de anonimización de datos en las entidades públicas, teniendo como objetivo aumentar el número de datos abiertos seguros con protocolos de anonimización en el país.

#### **2. Datos Compartidos**

El fin de compartir datos o registros administrativos es lograr un mayor aprovechamiento estadístico, generar valor a partir de la explotación de varios tipos de datos, mejorar los mismos registros o desarrollar un nuevo servicio que permita mejorar la prestación de los servicios dentro del sector público. Para lograrlo se propone crear protocolos de seguridad que permitan a un ente rector compartir los datos o registros, con previa identificación y aprobación de su finalidad. Estos protocolos deberán ser definidos por las autoridades competentes, garantizando el flujo permanente de la información y los accesos a los puntos focales de cada entidad.

Este tipo de datos compartidos contienen variables de identificación que son importantes para poder cruzar registros. Es por eso que no todas las entidades pueden tener acceso a este tipo de información. Por lo tanto, los protocolos sólo deberán brindar acceso a este tipo de información siempre y cuando la entidad interesada pueda demostrar su relevancia e importancia en su uso o explotación. Una vez esto se corrobore, esta información deberá interoperar bajo estándares de seguridad y privacidad definidos previamente.

#### **3. Datos Inter-agenciales**

Los datos inter-agenciales son los datos más sensibles de los tres tipos de acceso de datos presentados, pero son estrictamente necesarios para que el gobierno pueda prestar y atender mejor sus servicios. Para el acceso a estos datos, es necesario implementar estrictos protocolos de seguridad. Además, se recomienda que estos tengan como requisito la encriptación o el uso de tecnología segura para poder ser compartidos o consumidos entre entidades interesadas. En esta línea de acción, es importante mencionar que no necesariamente estos datos deben estar viajando por todas partes o sistemas. Lo ideal es que cada entidad dueña de este registro generé sistemas de validación para que otros sistemas de información puedan obtener algún tipo de validación e información adicional puntual que se requiera para una mejor prestación del servicio; siempre y cuando esté aprobado y verificado por un comité rector. Esta categorización de dato se debe tener en cuenta al momento de diseñar y actualizar la arquitectura de datos nacional.

Es en este grupo de acceso a datos donde se encuentra la idea de identidad digital; acción que se expone más adelante. Al tener cada ciudadano unas variables únicas que lo identifican y ciertos atributos privados que son de entera reserva y que no deben correr el riesgo de ser expuestos, se vuelve de vital importancia contar con protocolos de acceso que en primer lugar cumplan con la seguridad necesaria y que en un punto logren compartir información agregada sin que sea necesario que los datos salgan de las entidades propietarias.

#### **ii. Repotenciamiento de los equipos de cómputo**

En muchas de las entidades consultadas se evidencia la precariedad en el uso y manejo de la información, así como en algunos casos la obsolescencia de los equipos utilizados para albergar la información y ponerla al servicio de los ciudadanos. Consideramos que una estrategia de Big Data requiere una agenda clara de adquisición, implementación y uso de equipos y servicios que permitan y faciliten las tareas básicas identificadas a lo largo de los documentos presentados al DNP como son las requeridas en la nueva era de la revolución digital. La decisión de las formas como se van a adquirir los equipos, servicios o tecnologías, así como la identificación de requerimientos y prioridades no puede llevarse a cabo sin la clara identificación

del mencionado mapeo y diagnóstico de la arquitectura de información nacional.

Contando con requerimientos técnicos claramente definidos por un conjunto de actores relevantes para la estrategia nacional de datos, se pueden priorizar las acciones de acuerdo con el impacto, la necesidad y la capacidad de interoperabilidad; para definir un proceso de inversión en los próximos años. Estas acciones deben garantizar la transformación y actualización con estándares globales de la forma como hoy en día se almacena, procesa, usa y difunde información en el país. Para alcanzar este objetivo, se debe contar con recursos suficientes para la adquisición de equipos, servicios y tecnologías, capaces de prestar los servicios requeridos. A su vez, se hace necesario ser muy cuidadosos en la eficiencia de las compras evitando adquirir equipos, servicios o tecnologías que sobrepasen las necesidades de operación o que sean muy tempranamente adquiridos antes de tener la capacidad suficiente de uso, ya sea por volumen de información o capacidades de operación por parte de los técnicos, o por parte de los ciudadanos.

De igual manera, es esencial implementar una política de actualización de equipos de cómputo, para así cerciorarse de la funcionalidad y relevancia de los equipos sin dejar que estos se vuelvan obsoletos. En preparación de esta estrategia, no se recomiendan *clusters de hadoop*, un tipo especial de *cluster* diseñado para almacenar y analizar grandes volúmenes de datos no-estructurados. En varias instancias no se necesitan estos *clusters*, pues son costosos de construir, mantener y específicamente complejos en el manejo y en la capacitación por parte de funcionarios<sup>8</sup>. En vez, se propone que se priorice el *cloud computing*<sup>9</sup>, pues es la herramienta más rentable para múltiples tareas y permite la flexibilidad de incrementar o disminuir la capacidad de uso dependiendo de las necesidades de la(s) entidad(es), o pay-per-use. De ser necesario, esto se puede complementar con PCs con multiprocesadores de alta gama con capacidad de terabytes de RAM y petabytes en el disco duro.

#### b. Mediano plazo

##### iii. Formación de capital humano y promoción de cultura de datos

La infraestructura de datos no está exclusivamente relacionada con la adquisición de equipos físicos de cómputo, sino con la consolidación de equipos humanos capaces de

la operación de los mismos. En este orden de ideas es necesario y urgente contar con equipos altamente entrenados en manejo de Big Data, tanto en el escenario público como en el privado, de manera que se alcance niveles elevados de operatividad de los equipos adquiridos. Se requiere que todas las unidades del gobierno cuenten con individuos altamente capacitados en hablar lenguajes similares sobre el uso de la información y que cuenten con capacidades técnicas suficientes para dar respuestas a los retos que se enfrentan ante el necesario buen manejo de la información en el país.

Como se ha mencionado en varias oportunidades (en los documentos 1, 2 y 3) las capacidades no son exclusivas de los equipos de IT sino que se deben “transversalizar”. Lo que significa que un gran número de funcionarios y ciudadanos deben estar en capacidad de utilizar de manera adecuada la información en su cotidianidad, bien sea como parte del proceso de producción o como usuarios que esperan sacar provecho de la riqueza de la información adecuadamente organizada y accesible al público general. Se considera necesario revisar los cargos que se deben integrar a las funciones públicas, entre estos podría ser los de curador de datos, exploradores de datos, arquitectos de soluciones empresariales, científico de datos y expertos de campaña. Es necesario revisar la cadena de valor del dato en el gobierno para identificar los vacíos de habilidades en el proceso de explotación de estos.

#### Cuadro 4. Recursos humanos para la explotación de datos

La revista *Harvard Business Review* sostiene que tener un equipo ideal de Big Data no es sólo contar con capital humano altamente capacitado en matemáticas, computación y analítica de datos; sino también, , pero sí en alinear las habilidades de sus empleados con las necesidades institucionales. Estos identifican cinco roles esenciales en un equipo de Big Data:

1. Curador de datos: el curador de datos se asegura de que los datos que entran a los sistemas de información estén limpios y sean precisos, y de que se mantengan de esta forma durante el ciclo de vida de estos datos. Esto incluye revisar los metadatos de los datos para asegurarse de que los valores, por ejemplo, de fecha, estén en el mismo formato. Esto también involucra estandarizar información clave, por ejemplo, la manera como se mide

<sup>8</sup> Según el Prof. Pentland, en muchas ocasiones no se requiere de Hadoop - la gente usa estos clusters porque son reconocidos, y utilizados en varias instancias - sin embargo, es un sistema apto para organizaciones que están procesando un volumen de datos equivalente o superior a petabytes por hora (para escala, esto equivale a  $2^{50}$  bytes, a 1000 terabytes a la capacidad de almacenamiento de 61,000 computadoras de escritorio (Rouse, 2018)).

<sup>9</sup> Luxemburgo liberalizó su regulación de cloud computing, ofreciendo incentivos para sus proveedores, volviéndose el centro de computación en Europa. Este modelo sería uno que se podría imitar, pues significó una victoria para el gobierno en cuanto al desarrollo de tecnologías y a recaudación de impuestos. Para más información, ver: <https://www.ey.com/lu/en/services/advisory/ey-cloud-sourcing-are-you-familiar-with-luxembourg-s-revised-regulatory-environment>

- el un año (días en el calendario, días hábiles, o horas totales en el año). Esto también involucra estandarizar datos claves a través del sistema, por ejemplo, si se utiliza un categorización numérica para una variable como género, que sea estándar a través de los conjuntos de datos que femenino se categoriza como 01 y masculino como 02.
2. Exploradores de datos: son los encargados de descubrir los datos que verdaderamente se necesitan para realizar una tarea. Históricamente los datos no han sido estructurados ni almacenados para tareas de analítica de datos, por lo cual son de difícil acceso; este rol se encarga de crear el acceso a estos datos y de determinar si son relevantes o no.
  3. Arquitectos de soluciones empresariales: son los encargados de unir los datos descubiertos y de organizarlos para ser analizados. Estos estructuran los datos para garantizar que todos los usuarios puedan consultarlos en el plazo requerido para su uso. Este rol es el intermediario entre la organización y los ingenieros, científicos de datos, u otros.
  4. Científicos de datos: son los encargados de crear modelos de analítica de acuerdo a las necesidades previamente definidas - por ejemplo, predecir el comportamiento de los ciudadanos frente a una campaña mediática. Estos se aseguran de que los modelos estén actualizados y respondan a las necesidades empresariales/organizacionales. Generalmente, los científicos de datos son calificados en programación, analítica, estadística, matemáticas.
  5. Expertos de campaña: son los que traducen los modelos en resultados tangibles. Tienen un conocimiento profundo tanto de los sistemas técnicos que se utilizan para proveer soluciones, pero también de cómo utilizar estos resultados al convertirlos en acciones empresariales, priorizando líneas de acción de acuerdo al análisis de los datos. Por ejemplo, cómo estructurar una campaña de marketing, dependiendo del tema, para obtener los mejores resultados.

Se propone iniciar una campaña en donde las habilidades y el conocimiento sobre los datos se prioricen tanto en los planes de formación empresariales, organizacionales, como educativos. Se deben implementar reformas a nivel educacional, tanto básica como superior, dónde se invierta a partir de una edad temprana en el desarrollo de capacidades de datos. A mediano plazo se propone ampliar rápidamente las clases ofrecidas en universidades sobre ciencias de datos y **computer science** - en Estados Unidos, en la última década la mayoría de universidades han integrado estos programas en sus ofertas educativas.

De igual manera, se deben crear programas de educación continua, o extracurriculares que capaciten sobre estos temas, y que proporcionen incentivos en hacerlo, incluyendo programas de prácticas profesionales durante el verano o durante el semestre. A más largo plazo, es importante incluir clases de estadística y de computación en el currículum obligatorio para bachillerato.

A su vez, es importante también fomentar la formación profesional y promover el tiempo libre para capacitarse para desarrollar tanto las capacidades, como la cultura de datos en el país. Organizar eventos de comunicación y promoción de la cultura de datos es también importante para el proceso de transformación a una sociedad impulsada por los datos y orientada a resultados. Es importante impulsar alianzas con instituciones de educación superior para mejorar el acceso, tanto a nivel de pregrado como de posgrado, a programas de ciencias de datos, tecnología, **computer science**, mezclado con programas de política pública, economía y ética. Quizá el modelo que Estados Unidos tiene en este momento de **presidential management fellows**, un programa de desarrollo de liderazgo para estudiantes de posgrado que atrae y selecciona los mejores candidatos para desarrollar una base de líderes, es apropiado para el contexto de Colombia. Este programa ofrece capacitaciones y oportunidades laborales con el Estado - desarrollar posiciones con relevancia a Big Data y AI, en alianza con instituciones como el DNP, MinTic, DANE y otras agencias, impulsa las capacidades humanas y la cultura de datos en el país.

Una de las lecciones aprendidas más importantes del estudio comparativo es la importancia de impulsar las capacidades humanas y de capital, eslabón fundamental en la infraestructura de Big Data. Singapur, por ejemplo, ha invertido en programas de alfabetización de datos para sus ciudadanos, y específicamente en capacitar a los funcionarios del Estado. El hacerlo no solo garantiza *buy-in* por parte del sector público, sino una inversión en capacidades a nivel gobierno.

#### Cuadro 5. Alfabetización de datos

Singapur es uno de los países que se destaca en universalizar las habilidades básicas de datos a través de programas comprensivo de alfabetización de datos. Uno de los elementos claves del gobierno actual para estimular el sector de Big Data es la construcción de capacidades y competencias informáticas a partir de una serie de programas y proyectos, que forman parte de Smart Nation Initiative (SIN), que incluye: IMDA Playmaker Program que expone a niños de manera interactiva a la tecnología, y el movimiento Code@SG que introduce pensamiento informático y de programación en estudiantes. Dos programas que también forman parte de la Smart Nation Initiative son Skills Future, que desarrolla capacidades en TIC para el trabajo y el Smart Nation Fellowship Program que invita a científicos de datos e ingenieros a participar de la solución de desafíos nacionales o de otros países en diferentes áreas.

Otros países han desarrollado actividades específicas en el sector de Big Data, incluyendo *hackatones*, *innovation hubs*, *incubadores*, plataformas y redes tecnológicas; incluso, facilitando incentivos para que empresas emergentes creen aplicaciones que los ciudadanos puedan utilizar con datos extraídos del gobierno. La construcción de capacidades es esencial para generar demanda de datos y consecuentemente, para generar valor a partir de estos.

Consideramos que el Plan Nacional de Desarrollo, o políticas públicas que se desarrollen, debe tener un componente robusto de inversión para la innovación y el emprendimiento en asuntos relacionados con información. En el marco de los avances que se proponga en COLCIENCIAS consideramos que se deben impulsar proyectos de supercómputo, de Open Data, de Big Data y de AI, que redunden en el bienestar del país. Ubicar recursos públicos para el mejoramiento de la capacidad de innovación es crucial para garantizar que se puedan establecer acciones precisas que promuevan el desarrollo de la estrategia en el mediano y largo plazo.

### **1. Sensibilización sobre Big Data y construcción de la cultura de datos**

Una de las principales acciones, y que requiere una ejecución temprana, aunque se considere una acción transversal de la estrategia, es la de sensibilización temática en el país (es parte de la promoción de cultura de datos, pero es considerada como una estrategia que se puede implementar desde el primer año de la estrategia). Desde el gobierno nacional, en alianza con las unidades territoriales y en asociación con el sector privado, las organizaciones sociales y la academia, deben emanar espacios de sensibilización, capacitaciones, talleres, encuentros, entre otras, que promuevan el desarrollo de una cultura de datos nacional. Al igual que las políticas públicas de promoción, es fundamental que se articulen de manera adecuada y estructurada para magnificar el impacto y aumentar la eficiencia de los recursos invertidos. Adicionalmente el efectivo seguimiento al uso de datos a través de métricas y monitoreo a los sistemas de intercambio debe convertirse en uno de los pilares centrales de las evaluaciones al desarrollo del ecosistema nacional de datos.

### **IV. Desarrollo de un marco legal articulado para la explotación de datos**

Es necesario articular el cuerpo normativo y jurídico al servicio del ecosistema de información que garantice el funcionamiento, respete los derechos ciudadanos y promueva la inversión, innovación y desarrollo de negocios en torno a los asuntos relacionados con el campo. Se encontró en el diagnóstico que existe un cuerpo normativo desarticulado y disperso sobre asuntos de datos y por lo tanto es urgente que se establezcan acciones para el desarrollo de un cuerpo normativo articulado. De igual manera, se identifica que no existe confianza para el intercambio de información y esto se deriva de una débil capacidad de protección jurídica ante los abusos en el manejo de información; esto se convierte en un llamado urgente al establecimiento de formas y acciones para el uso y manejo de información pública, como lo es la tipología de datos que proponemos y los protocolos asociados al manejo de estos dependiendo de su clasificación. Abordar esto implica el establecimiento de métodos de anonimización para la adecuada publicación y difusión de datos. El Anexo A de este documento aborda los procesos de preanonimización, la anonimización de microdatos y anonimización de datos que se podrían implementar dependiendo del riesgo de identificación de acuerdo a los datos manejados.

Frente a las competencias difusas que las leyes actuales asignan, es necesario unificar y actualizar el marco jurídico y crear nuevas caracterizaciones éticas para el manejo de datos. Se propone una mesa de revisión para evaluar la situación actual de todos los reglamentos, normas y leyes que conciernen el acceso, difusión, producción, uso y gestión de datos. Esta mesa de revisión también debe estar en capacidad de realizar un compendio de la normatividad existente relativa a medios tecnológicos e intercambio de información, hacerle seguimiento constante a la aplicación de las normas de protección de datos y estar en la capacidad de sugerir cambios ante nuevas necesidades o requerimientos. La mesa de revisión debe estar compuesta por los actores principales del sector, entre los que se identifican el DNP, MinTic, InnPulsa u otros en el área de Ciencia y Tecnología. La intención es lograr identificar cuáles normas son relevantes para el desarrollo del ecosistema de Big Data, cuáles obstaculizan el desarrollo de estos, y cuáles son los vacíos de regulación que se deben complementar.

De igual manera, se debe pensar en incluir un Comisionado de Ética de Datos, Comité de Ética, o Oficina del Comisionado de la Información, siguiendo el ejemplo del Reino Unido, entre otros países que cuentan con esta unidad administrativa. El Comité de Ética gestiona auditorías a nivel local y emite sanciones penales a quienes en el sector privado infrinjan en el buen uso o explotación de datos. Este órgano institucional articula el marco ético para la explotación de los datos, incluyendo la privacidad de datos, pero más importante, se encarga de guiar a las entidades y personas con la realización de proyectos de ciencias de datos y Big Data, y a entender y a adaptarse a los nuevos cambios normativos.

Aunque el Plan Nacional de Desarrollo debe hacer un llamado contundente a la formulación de un cuerpo normativo, jurídico y legal para el manejo de información en Colombia, debe existir una iniciativa especial que permita llevar a cabo un esfuerzo integral y coordinado al estilo de las denominadas “Misión” que permita poner a un grupo de personas concentradas en sugerir las líneas para un bloque normativo propicio para la estrategia de información y grandes datos. En el nivel territorial solo se debe incluir adoptar la normatividad en la mayoría de las entidades pero no incidir directamente sobre la misma.

### **V. Generación de alianzas institucionales para la explotación de datos**

El reto más importante para alcanzar resultados satisfactorios a escala nacional como lo observamos en los casos analizados en los documentos 1 y 2 es la adecuada articulación entre el sector privado, gubernamental, academia y ONGs. Para esto

es necesario encontrar escenarios de intercambio de ideas y definición de líneas de acción que reduzcan los obstáculos y permitan la construcción de espacios de interacción.

El desarrollo de un ecosistema de información y datos estará claramente modulado por el sector público, pero el verdadero actor que permite profundizar los procesos es el sector privado. Este es el generador principal de la oferta de bienes y servicios, así como el consumidor, que significa una demanda de constante innovación y mejoramiento. Hasta el momento los escenarios de inversión en el país son escasos dado que no se percibe el sector de información y de Big Data como prioritario, y solo aquellos grupos empresariales que conocen a profundidad la riqueza del mismo se animan a proponer procesos de profundización en sus inversiones. Aunque no es tarea central del estado generar incentivos empresariales para la inversión, si es posible que apoye con la identificación de nichos de mercado a partir de las necesidades que encuentra en el proceso de seguimiento, así como en la generación de alianzas institucionales para la consolidación de un ecosistema de Big Data en Colombia.

El punto de partida debe ser el proceso de construcción conjunta de la arquitectura de información nacional. En un escenario donde se discute sobre el mapeo de necesidades, de fuentes de información, de calidad de información con estándares y normalización y una arquitectura conjunta; deben surgir iniciativas de trabajo conjunto para desarrollar el ecosistema de Big Data en el país. El elemento clave a solucionar es el llamado al fortalecimiento de la confianza que permita encontrar mecanismos de intercambio basados en iniciativas de interoperabilidad y diálogo continuo entre las fuentes de datos e información en el país. Iniciativas como OPAL son propuestas interesantes para la articulación de intereses entre múltiples actores. Se pueden plantear posibilidades adicionales que permitan solucionar fricciones existentes, que faciliten potencializar el trabajo conjunto entre actores y que lleve no solo a la visión pública de la información y a la generación de ambientes adecuados para la constante innovación y el emprendimiento en temas de Big Data.

Es importante esclarecer que la inversión y el desarrollo del ecosistema de datos no debe recaer únicamente en el gobierno, por lo cual las alianzas son claves para financiar el desarrollo de la estrategia. El punto de partida para generar estas puede ser el promover alianzas entre industrias (infraestructura, tecnología, servicios, banca privada), financiadas por estas, pero abiertas a la participación del sector público en función regulatoria.

### **Alianzas privadas**

El estado deberá apoyar la creación de alianzas entre actores privados para la creación de valor en cuanto a la explotación de datos en el país. Incluir en el marco jurídico el reconocimiento de estas alianzas y normativizar incentivos para alianzas entre entidades privadas; apoya el desarrollo del ecosistema de Big Data en Colombia. Las alianzas entre organizaciones privadas deben contar con el apoyo del gobierno, en cuanto a la promoción de proyectos pilotos, proporcionando seguridad ante el riesgo de experimentar por parte de estas industrias.

### **Alianzas público-privadas**

Es de vital importancia promover las alianzas público-privadas en el marco de la estrategia nacional de Big Data. Se debe tener en cuenta sin embargo, que este tipo de alianzas en Colombia son formalizadas a través de las Asociaciones Público Privadas (APP<sup>10</sup>) como un mecanismo de vinculación entre los sectores para proveer bienes y servicios. Estas son caracterizadas entre iniciativas públicas, donde los proyectos son estructurados por la entidad pública, o iniciativas privadas, donde la idea conceptual es del sector privado. Lo complejo de esta clasificación por parte del Estado colombiano, es garantizar que los recursos invertidos por el sector público sean reinvertidos a este tras finalizar los proyectos, entre otros.

El objeto de estas alianzas puede incluir el fortalecimiento de las plataformas de datos abiertos, tal como lo es datos.gov.co, al incluir datos de empresas privadas, o concentrarse en desarrollar servicios a partir de la explotación de estos datos<sup>11</sup>. De igual manera, es esencial promover *hackatones*, eventos de innovación, y en general, plataformas para que los ciudadanos, el Estado, y el sector privado promocionen y utilicen datos para generar valor económico y soluciones sociales. Puntualmente, se recomienda crear una alianza continua con el sector privado para establecer *innovation hubs* e incubadoras de empresas basadas en datos.

nacional que surgió a partir de alianzas entre líderes del sector privado y académico en Canadá en el 2010 y que cuenta con el gobierno nacional como aliado gubernamental. Esta ONG es líder en programas que promueven el emprendimiento, tanto en apoyo a fundadores de startups a través de *seed funding* y asesoramiento por parte de expertos, en el apoyo a iniciativas relacionadas con la inteligencia artificial, y en un programa de educación y apoyo a emprendedores.

---

<sup>10</sup> Abecé Asociaciones Público Privadas" (Bogotá D.C: Departamento Nacional de Planeación, 2013).

<sup>11</sup> Klievink et al. explican que en Holanda existen reuniones organizadas por el Ministerio del Interior, dónde se reúnen los usuarios de datos abiertos y oficiales del gobierno para actualizar a los consumidores de datos sobre los desarrollos en acceso y liberación de conjunto de datos, así como para intercambiar novedades y uso de los datos. Esta comunicación constante hace que la demanda y oferta de datos esté alineada.

#### Cuadro 6. Emprendimiento de datos

Definimos el “emprendimiento de datos” como emprendimientos de pequeño y mediano tamaño que hacen parte del ecosistema de Datos Abiertos y usan, generan o facilitan la generación de datos abiertos para construir negocios sostenibles que generen impacto económico, social y/o ambiental. De acuerdo con las últimas tendencias e investigaciones acerca de datos abiertos, se identifican dos buenas prácticas vinculadas a este tipo de datos que atienden, por un lado, la necesidad de crear o fortalecer empresas tecnológicas que desarrollen herramientas, plataformas, o servicios relacionados a Big Data (empresas que tradicionalmente entran en la categoría de TIC); y por el otro, fomentar a determinados sectores, industrias o empresas existentes para que usen los beneficios de los datos, en particular de los datos abiertos para la optimización de sus ventas, decisiones o procesos.

En Reino Unido, el gobierno se ha encargado de ofrecer datos abiertos gubernamentales de la más alta calidad a través de data.gov.uk. El Open Data Institute (ODI) en su investigación Open Data Means Business, ha analizado 270 empresas en el Reino Unido que usan, producen o invierten en datos abiertos como parte de su estrategia de negocio. Estas empresas (llamadas también “compañías de datos abiertos”) facturan más de 110 billones de dólares al año y emplean a más de 500 mil personas. La mayoría (70%) de las empresas analizadas usan datos abiertos proporcionados por el gobierno, aunque también usan datos abiertos de otras fuentes como Open Corporates. El sector privado está conformado por un gran número de pequeñas, medianas y grandes empresas, incluyendo grandes consorcios o nuevas empresas de tecnología y telecomunicaciones. Un análisis realizado por la firma de auditoría RSM encontró que para finales de febrero de 2017, el número de negocios de tecnología operando en el país había alcanzado los 5,995, con las concentraciones más altas de crecimiento de compañías tecnológicas en Londres.

Algunos ejemplos de estas empresas son: Mastodon C, especialistas en Big Data que desarrollaron una plataforma open source para toma de decisiones en ciudades; Carbón Culture, una plataforma comunitaria diseñada para ayudar a las personas a utilizar los recursos de manera más eficiente, realizaron un proyecto de datos abiertos para ahorrar carbono, energía, agua y dinero; y GeoLytix que ofrece servicios de geolocalización a retail, bancos y desarrolladores inmobiliarios utilizando datos abiertos.

Fuente: “Buenas Prácticas sobre Cómo los Gobiernos pueden implementar una estrategia para estimular el ecosistema de negocios de Big Data”, Data-Pop Alliance (2017-2018).

#### Alianzas público-público

Se deben fomentar alianzas entre ministerios, agencias, y otras divisiones administrativas del Estado, para impulsar proyectos de construcción del ecosistema de Big Data desde el sector público. Estos pueden incluir eventos de construcción de capacidades humanas.

#### 1. Discusión nacional multisectorial

La estrategia nacional de Big Data, aún cuando es estructurada por el gobierno y responde a las necesidades de este, requiere también de la participación del sector privado, de la academia y de la sociedad civil para impulsar de manera comprehensiva el ecosistema de datos nacional. El punto de partida para lograrlo debe ser una discusión multisectorial nacional con representantes de diferentes sectores. Esta participación se puede lograr a través una mesa redonda con representantes de gremios (i.e. Federación Nacional de Cafeteros), asociaciones industriales (i.e Asociación Nacional de Empresarios de Colombia), universidades, sociedad civil, entre otros. Aun cuando la estructura y organización de esta discusión debe ser definida por DNP, MinTIC, Innspulsa o quien se designe, se recomienda que cuente con uno o dos representantes de cada sector - y que estos ya hayan definidos sus necesidades en cuanto a Big Data y explotación de datos a través de un proceso de presentación de white-papers<sup>12</sup> (documentos técnicos) que comuniquen las necesidades concretas de estos sectores en cuanto a Big Data. Estos serían entonces propuestas cortas por cada sector. La discusión multisectorial tiene como objetivo unificar y priorizar líneas de acción que vayan de la mano de aquellas propuestas para el gobierno.

Lo que se debe lograr en la implementación de la estrategia, es que los diferentes sectores, particularmente aquellos en el sector privado participen, no sólo comunicando sus necesidades y puntos centrales, sino también, desarrollando oportunidades para la integración y el crecimiento del ecosistema de Big Data nacional. Si el objetivo es potenciar a Colombia como líder en cuanto a explotación de datos, y generar valor a partir de estos, es necesario contar con la participación de

<sup>12</sup> Un referente internacional de cooperativas de datos es MIDATA, una iniciativa que busca empoderar a los ciudadanos para que almacenen, manejen, y controlen de manera segura sus datos personales a través de cooperativas que actúan como fiduciarias de los datos de los individuos miembros (principalmente de datos médicos). Los miembros deciden si sus datos son utilizados para investigación o no, asegurando la transparencia y seguridad de los datos. El modelo MIDATA tiene el objetivo de ser replicable para permitir que los proyectos de investigación a través del mundo sean elaborados de una manera justa y democrática.

la industria; un punto de partida es esta discusión nacional. La estrategia propuesta con las acciones arriba es una general - depende de las necesidades actuales del país definir la líneas de acción específicas. De igual manera, esta discusión ayuda a educar a los actores relevantes en Colombia en cuanto a la estrategia y es una fuente potencial tanto de alianzas como de financiamiento.

### c. Largo plazo

#### **Vi. Desarrollar productos inteligentes en red y sistemas de identificación**

Establecer una identidad digital segura debe ser la prioridad a largo plazo de esta estrategia. La mayoría de países tienen alguna noción de una identidad digital - una cédula digital o un identificador tributario - pero son pocos los que han podido implementar una forma de identificación segura online, que al mismo tiempo proteja al ciudadano, y le abra las puertas a nuevos servicios y a más eficiencia. El no tener esta identidad digital segura en muchas ocasiones resulta en fraude, robos de identidad y corrupción.

Países que han adoptado este sistema han visto un incremento en el comercio, reducción de la corrupción y un aumento en los ingresos tributarios. Se propone entonces la creación de una identidad digital para Colombia, que construya sobre los Servicios Ciudadanos Digitales propuestos en el marco de la estrategia de Gobierno Digital, particularmente sobre el Servicio de Autenticación Electrónica. El punto es contar con una forma de autenticación digital segura que le permita al ciudadano acceder a una variedad de servicios tras ser identificado como tal. Para hacerlo, es necesario desarrollar la infraestructura que apoye este servicio — el precedente internacional para esto es la identidad digital de Estonia y X-Road, su sistema de datos distribuido (no centralizado), donde la información puede ser intercambiada instantáneamente. La propuesta no solo debe concentrarse en autenticación electrónica, sino también en la integración de servicios tanto gubernamentales como empresariales en el uso de esta identidad digital (impuestos, servicios financieros, firmas digitales, entre otros).

A corto plazo, se propone iniciar la integración a la identidad digital con grandes compañías en uno o dos años, para después moverse en un plazo de cinco años para medianas y pequeñas empresas. A largo plazo, en 10 años, todos los ciudadanos deben estar integrados en la estrategia de identidad nacional. Para el desarrollo de esta, consideramos que la dirección y gestión proveniente desde la Presidencia es crucial para articular la importancia de esta iniciativa.

En el centro del debate propuesto por el profesor Sandy Pentland está la concepción de la información como una de las tantas formas que puede tomar la propiedad y, por ende, el llamado es a encontrar mecanismos que garanticen la adecuada producción de información a partir de identidades confiables y seguras. Asegurar el uso de datos que garantice la incorruptibilidad de la operación y que permita establecer contratos digitales, y aumentar la eficiencia y facilidad de servicios digitales son piezas estructurales de la organización de la información nacional en función de una estrategia de datos clara.

#### **Vii. Mejoramiento y cualificación de la demanda por información**

Un gran número de las propuestas que se establecen apuntan a mejoras en la oferta de información y de los procesos de producción de información. De igual manera, es fundamental llevar a cabo acciones que apunten al mejoramiento y cualificación del uso de esta información. Para ello es necesario fortalecer los procesos de aprendizaje sobre el uso de la información para ciudadanos, organizaciones e instituciones para la consolidación de una sociedad movida por los datos, o una “*Data-driven society*”. En Colombia existen numerosos procesos de toma de decisión basados en información, donde se exige que existan buenas bases de datos para sustentar las decisiones. Los retos actuales que se presentan con el advenimiento del Big Data implican que se deben establecer canales de mejoramiento continuo que sean capaces de adaptarse a los cambios que normalmente acontecen de manera acelerada y súbita.

Una idea para lograr esto, es la creación de cooperativas de datos, manejadas por sindicatos, uniones, cooperativas, o grupos similares, que agreguen y gestionen datos de ciudadanos (*citizen data*) para obtener mejores servicios y más valor de la información . Esto genera *buy-in* de organizaciones ciudadanas, generalmente temerosas y resistentes al potencial del Big Data y de la Inteligencia Artificial.

Consideramos que es fundamental la inclusión en múltiples instrumentos de política pública la diversidad de acciones que posibiliten la profundización de procesos para el mejoramiento y cualificación de la demanda por información. Para ello se propone: (1) expansión de las posibilidades de uso y consumo de datos por parte de la ciudadanía (2) vinculación de nuevas tecnologías de analítica en los procesos de toma de decisión de emprendimiento, empresariales y de innovación, (3) fortalecimiento del bilingüismo como herramienta para mejorar los procesos de adquisición de habilidades para el manejo de nuevas tecnologías y (4) generación de incentivos para mejorar la remuneraciones a profesionales en el sector de información y Big Data. Este último se puede iniciar a través de fondos para mejoras en la formación académica y luego en incentivos laborales principalmente para empleo en el sector público.

#### **Consideraciones adicionales**

Los algoritmos y las capacidades analíticas que las entidades deben desarrollar para la explotación de datos desde el sector público dependen enteramente de las necesidades de cada entidad, pues es el tipo de datos que las entidades manejan y las necesidades de análisis que definen estas, corresponden a la naturaleza de cada una. Sin embargo, es necesario desarrollar habilidades dentro del sector público para realizar estas tareas, por lo cual se debe priorizar el desarrollo de cursos que

capaciten a individuos en la creación de algoritmos y en *data analytics* - cursos de educación continua serían ideales para desarrollar estas capacidades.

En cuanto a la implementación de la estrategia, la analítica avanzada es útil para medir el impacto de ésta - punto que será desarrollado en detalle en la propuesta de implementación de la estrategia-. Sin embargo, ésta se puede usar para hacer un estudio de efectividad de la entrega de datos, particularmente para entender los comportamientos de los usuarios de datos del gobierno (clicks, movimiento, entendimiento). La analítica avanzada es también esencial para estructurar varias de las estrategias, entre estas la de integrar la infraestructura de datos nacional. Una acción derivada de esta es automatizar y hacer más eficiente las consultas de queries comunes a través del estado, estudiando patrones para anticipar qué tipo de información es el que se solicita más a menudo.

A largo plazo, el valor agregado de una estructura de Big Data depende de dos factores, el movimiento de los datos y el valor agregado a los datos; pero estos dependen de los detalles de los datos, su aplicación y uso específico, y en general, de las operaciones que se están realizando con los datos. El transportar los datos de manera automática y sin necesidad de involucrar capital humano es uno de los elementos esenciales en esta cadena. Para medir el valor de los datos, como se mencionó en el documento tres, va a ser necesario desarrollar indicadores y formas de medir el mercado de datos en el país.

## 5. ACTORES

Los actores involucrados en la Estrategia Nacional de Datos deben ser entidades estatales con la capacidad técnica para liderar y coordinar la implementación de la política pública en materia de Big Data, tanto en sus dimensiones técnicas como en las administrativas y presupuestales. Asimismo, esta debe aparecer un conjunto amplio de organismos del Estado, específicamente aquellos que tienen un rol protagónico en la producción, procesamiento, almacenamiento o diseminación de información. Estos se convierten en pivotes fundamentales de *spillovers* y *spin-offs* en materia de profundización de la política, tanto en las dimensiones oficiales como en escenarios privados, académicos y de organizaciones no gubernamentales.

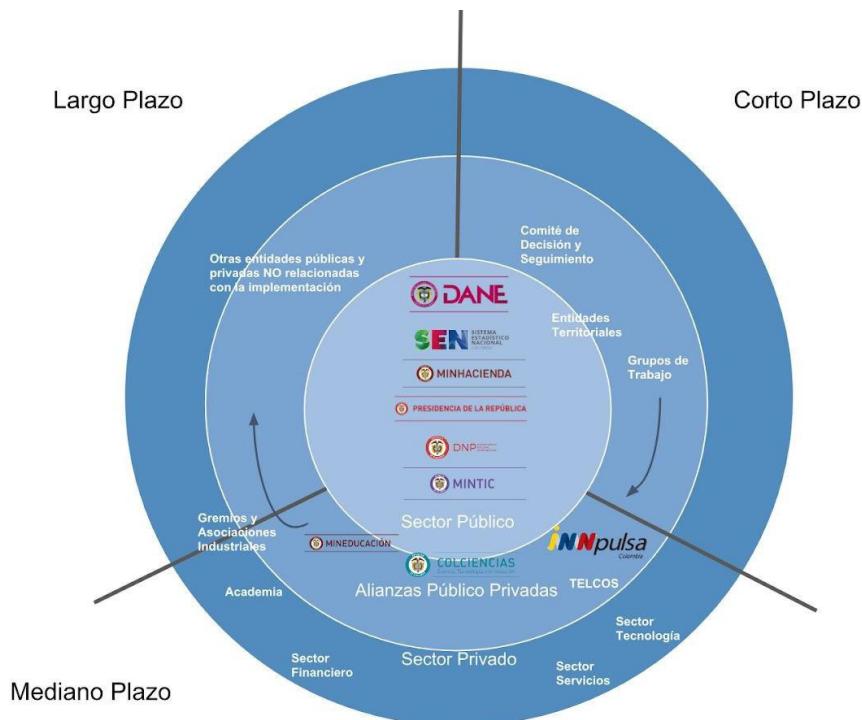
De manera complementaria, aparecen los actores del sector privado con alto nivel de liderazgo en el sector de información que pueden convertirse en líderes y coordinadores de las iniciativas nacionales en un escenario de concertación o mesa de trabajo conjunto. Ya en el largo plazo, están todas las entidades estatales y privadas que no cumplen un rol activo en la implementación de la política en materia de coordinación y liderazgo, pero son espacios donde debe tener alcance la iniciativa. Finalmente, en un grupo amplio aparece un conjunto múltiple y diverso de organizaciones privadas, academia e individuos que se convierte en nodos finales de la implementación de la política: los usuarios.

Como se puede apreciar en la siguiente figura y en la estrategia planteada, hay actores a nivel público que tienen un rol central (DANE, SEN, MinHacienda, MinTIC y Presidencia) a lo largo de toda la estrategia para que esta pueda ser correctamente concebida, financiada e implementada. Desde un principio, se deben crear grupos de trabajo y comités de decisión y seguimiento (compuesto por otras agencias del estado), para empezar a desarrollar un ecosistema y una cultura de datos en el estado colombiano. Además, es vital involucrar a las entidades territoriales en el diagnóstico e integración de la arquitectura de datos nacional, logrando coherencia entre las organizaciones y las regiones.

Continuando con el análisis de la figura están los actores involucrados necesarios para la implementación de las estrategias a mediano plazo. Contar con personal capacitado y crear una cultura de datos en Colombia sería imposible sin la participación de las carteras de educación y ciencia y la agencia de innovación. De igual forma, esta implementación no sería posible sin el sector privado, por lo que actores como las empresas del sector servicios, telecomunicaciones, financiero y la academia adquieren gran relevancia en este punto.

Ya en el largo plazo, se integran todos los actores involucrados tanto en el corto como en el mediano plazo involucrando cada vez más a otros actores del sector privado, gremios, asociaciones industriales que aunque no son participan directamente en la implementación, son parte vital de la estrategia como productores o beneficiarios de toda la estrategia de Big Data para el estado colombiano.

Figura 5. Mapeo de Actores para las tres etapas de la estrategia



Fuente: Elaboración propia

A continuación se listan las instituciones centrales para la estrategia:

#### a. Instituciones responsables

##### i. Equipo líder (coordinación, liderazgo y sostenibilidad)

Para la estructuración de la Estrategia Nacional de Datos es fundamental que exista un compromiso sólido por parte de cuatro entidades del gobierno nacional capaces de abarcar la complejidad inherente al ámbito de aplicación de la estrategia: Departamento Nacional de Planeación (DNP), Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), Departamento Administrativo Nacional de Estadística (DANE), y Ministerio de Hacienda y Crédito Público (MinHacienda). Cada uno aporta los elementos estructurantes necesarios para la solidez de la propuesta de implementación de la estrategia de datos.

Por un lado, el DNP aporta la posibilidad de articulación y coordinación con todas las dependencias del Estado colombiano. Tiene la aproximación sectorial, es decir, cuenta con el apoyo de un equipo con un nivel de especialización en asuntos de datos e información y, a su vez, tiene contacto con todos los otros sectores estructurantes de la planeación en el país, lo que aporta al proceso de articulación y coordinación. Adicionalmente, contribuye con la visión de mediano y largo plazo que le da su carácter centralizador de la planeación que se materializa en los dos de los instrumentos más importantes: el Plan Nacional de Desarrollo y los Documentos CONPES; esta visión de mediano y largo plazo se convierte en responsabilidad al darle al DNP el compromiso de convertirse en el garante de la sostenibilidad de la Estrategia Nacional de Datos. Finalmente, es fundamental la potencialidad que tiene el DNP para la coordinación territorial de la estrategia y la territorialización de esta a través de las acciones que se materializan en la sincronización con los planes de desarrollo territorial y en la vinculación con iniciativas del Sistema General de Regalías.

En la dimensión central y principal de la Estrategia Nacional de Datos se ha identificado que juegan un papel fundamental y estructurante el MinTIC y el DANE. Por un lado, el MinTIC tiene toda la capacidad pública de focalizar y priorizar las inversiones requeridas en infraestructura con un plan estratégico de corto plazo (choque) que complementa un programa de mediano y largo alcance. Asimismo, la modernización de la capacidad técnica vía infraestructura, equipo humano y capacidades institucionales requeridas para el éxito en la implementación de la Estrategia Nacional de Datos está en manos de las acciones de inversión pública, normatividad específica del sector y la gestión de alianzas sectoriales por parte del MinTIC. Por otro lado, el DANE se convierte en el aliado estratégico fundamental en las iniciativas que se ejecutan desde el MinTIC para el mejoramiento de la producción, procesamiento y disseminación de la información oficial tanto en ampliación de la cobertura de capacidades institucionales para el manejo contemporáneo de información como en el constante mejoramiento de la

calidad de la información disponible en el país. De esta forma, las acciones estructurantes de la Estrategia Nacional de Datos, los mecanismos de evaluación y revisión, así como los ajustes en el transcurso del proceso de implementación debe estar a la cabeza de estas dos instituciones.

Finalmente, el MinHacienda se convierte en el garante de la sostenibilidad de la iniciativa en el largo plazo. Los recursos presupuestales disponibles asignados a la Estrategia Nacional de Datos no pueden ser discretionales, se deben convertir en un continuo que garantice la aplicación coherente y sostenible de las acciones requeridas en las estrategias. En este punto, comprendemos que la decisión presupuestal está condicionada por las decisiones que se toman en la multiplicidad de proyectos que administra el Estado colombiano y, en función de esto, creemos que solo se debe delinear una Estrategia Nacional de Datos que se pueda sostener en el tiempo. Por ende, conocer el nivel de compromiso por parte de MinHacienda es fundamental para poder definir los alcances, temporalidad y profundización de la estrategia. En ese sentido, será preferible disminuir el alcance o la temporalidad de una acción estratégica a pretender alcanzar con recursos que no hacen posible obtener los resultados satisfactorios. Para finalizar, el DNP se debe encargar de una sincronización total y permanente con MinHacienda para la ejecución de las líneas estratégicas a lo largo de todo el proceso de implementación para lograr una maniobrabilidad clara en los momentos de inflexión del proceso, siempre con la asistencia y asesoría de MinTIC y DANE.

#### **b. Comité de decisión y seguimiento**

El Comité de decisión y seguimiento debe ser un conjunto ampliado de entidades del estado con rol protagónico en la Estrategia Nacional de Datos. Estas entidades son todas aquellas que tienen dentro de su función el manejo de información de alcance estratégico para el país y las que se requieren para la implementación de la política pública para la consolidación de un ecosistema de datos robusto en el país. En este orden de ideas, los Ministerios de Educación, Salud, Agricultura, Trabajo, Comercio, Ambiente, Vivienda, Hacienda, Defensa y Transporte son primordiales para las primeras acciones de trabajo conjunto. Otras entidades gubernamentales como la Alta Consejería para la Transformación Digital, la Registraduría, la Superintendencia de Notariado y Registro, la Función Pública, el Instituto Geográfico Agustín Codazzi, el Archivo General de la Nación, las Superintendencias, Banco de la República y múltiples institutos; son espacios fundamentales para el desarrollo de la estrategia en materia de datos enfocada en las acciones de captura, almacenamiento, análisis, uso y divulgación. Precisamente en estas entidades, los esfuerzos por estandarización, interoperabilidad y capacidades deben enfocarse para darle soporte y estructura al ecosistema de información nacional.

#### **c. Grupos de trabajo**

##### *Conjunto de organizaciones no gubernamentales protagonísticas en la Estrategia Nacional de Datos*

En este escenario se suman organizaciones de carácter privado relacionadas con el sector y las dinámicas de información en el país. Estas podrían ser empresas de información y computación (i.e, IBM), empresas de telecomunicaciones (i.e, Telefónica), empresas generadoras de información (medios de información y producción de datos), empresas de archivística y gestión documental, enfocadas en el manejo, producción, disseminación de la información; fundamentales para la implementación de la estrategia en el sentido en que se convierten en aliados claros de los procesos de ajustes y desarrollo de las dimensiones de la estrategia y, a su vez, demandan condiciones propias para el mejoramiento del relacionamiento en el sector y crecimiento de sus iniciativas.

De igual forma, aparece la academia donde se pueden identificar claramente acciones y actividades específicas relacionadas con la Estrategia Nacional de Datos. Grupos de investigación en Big Data, Supercomputación y Analítica son fundamentales para el constante trabajo en innovación y acompañamiento a los procesos de adopción.

## **6. RECURSOS DE INVERSIÓN**

Los recursos de inversión para esta estrategia se delinearon con base en el gasto general de la nación en áreas relevantes para la explotación de datos, teniendo en cuenta la inversión pública en ciencia y tecnología, información estadística, planeación, y otros sectores. De igual manera, se utilizó como referente lo planteado en el CONPES 3920, particularmente, lo relacionado con el plan de acción y seguimiento, así como las líneas de financiamiento consideradas factibles para la implementación de una estrategia de datos.

Sin embargo, para compensar la baja inversión que propone la Política Nacional de Explotación de Datos, los recursos de inversión delineados en la Tabla 1, se complementan con tres escenarios de inversión. El primero se acomoda a la inversión que el país ha destinado para la estrategia según el CONPES, el segundo es una aproximación a lo que se considera necesario para desarrollar la estrategia, y el tercer escenario se aproxima a lo que Colombia debería invertir en su estrategia de datos ajustado a su presupuesto anual y a su PIB<sup>13</sup>, de acuerdo a las comparaciones con respecto a inversiones de diferentes países

<sup>13</sup> Por ejemplo, Francia ha destinado 1.5 millones de euros para su estrategia de inteligencia artificial para los próximos cinco años. Sin embargo, el PIB de Francia es 9 veces el de Colombia - el país europeo también tiene un gasto presupuestal más alto, cerca al doble del presupuesto general colombiano. En términos relativos, si se ajustase este escenario de inversión a Colombia, correspondería un gasto de 80 millones de euros o 92 millones de dólares, gasto que no es factible considerando la inversión pública nacional. Para realizar este ajuste, se crean los multiplicadores de inversión para llegar al escenario alto de inversión.

en estrategias de Big Data e inteligencia artificial, (para más información, ver Anexo B). La inversión delineada también se ajusta a la inversión por sector de Colombia. Para lograr un impacto tangible, se necesita de una estrategia ambiciosa que cuente con la inversión pública necesaria para cumplir con las líneas de acción pactadas.

Aun cuando estos números pueden ser percibidos altos en términos absolutos - \$15 millones de dólares durante 6 años - son, en realidad, cifras conservadoras. Para este documento, una parte considerable de la inversión, se asignó para actualizar la infraestructura de datos nacional, la formación de capital humano y al desarrollo de alianzas. Para cumplir con los objetivos de esta estrategia, las entidades involucradas serán responsables de gestionar y priorizar los recursos para la financiación de las líneas de acción y de la inversión que estas requieren.

Tabla 1. Presupuesto de Inversión para Estrategia Nacional de Big Data

INVERSIÓN	ACTORES	DURACIÓN	1	*CIFRAS EN USD						TOTAL	
			2	3	4	5	6				
<b>ESTRATEGIAS A CORTO PLAZO (1 AÑO EN ADELANTE)</b>											
1	DANE, MINTIC, DNP, SEN	Mapear el flujo de información estructurada y no-estructurada y caracterizar deficiencias o avances en la arquitectura de datos actual.	\$180,000 \$360,000 \$540,000							\$180,000 \$360,000 \$540,000	
		Identificar la ubicación, características, estructura, disponibilidad y calidad de las fuentes de datos de las entidades del gobierno Colombiano.	\$300,000 \$600,000 \$900,000							\$300,000 \$600,000 \$900,000	
		Determinar necesidades en cuanto datos e información para la integración de la arquitectura de datos nacional	\$300,000 \$600,000 \$900,000							\$300,000 \$600,000 \$900,000	
2		Repotenciamiento de los equipos de computo	\$1,200,000 \$2,400,000 \$3,600,000							\$1,200,000 \$2,400,000 \$3,600,000	
			\$140,000 \$280,000 \$420,000							\$140,000 \$280,000 \$420,000	
<b>ESTRATEGIAS A MEDIANO PLAZO (3 AÑOS)</b>											
4	MINEDU, MINTIC, MINCIT	Formación de capital humano y promoción de cultura de datos	\$300,000 \$600,000 \$900,000	\$300,000 \$600,000 \$900,000	\$300,000 \$600,000 \$900,000					\$900,000 \$1,800,000 \$2,700,000	
		Desarrollo de un marco legal articulado para la explotación de datos	\$180,000 \$360,000 \$540,000	\$180,000 \$360,000 \$540,000	\$180,000 \$360,000 \$540,000					\$540,000 \$1,080,000 \$1,620,000	
5		Generación de alianzas institucionales para la explotación de datos	\$195,000 \$390,000 \$585,000	\$195,000 \$390,000 \$585,000	\$195,000 \$390,000 \$585,000					\$585,000 \$1,170,000 \$1,755,000	
			\$390,000 \$360,000 \$585,000	\$390,000 \$360,000 \$585,000	\$390,000 \$360,000 \$585,000					\$390,000 \$1,170,000 \$1,755,000	
<b>ESTRATEGIAS A LARGO PLAZO (6 AÑOS)</b>											
7		Desarrollar productos inteligentes en red y sistemas de identificación	\$36,000 \$72,000 \$108,000	\$36,000 \$72,000 \$108,000	\$36,000 \$72,000 \$108,000	\$36,000 \$72,000 \$108,000	\$36,000 \$72,000 \$108,000	\$36,000 \$72,000 \$108,000	\$36,000 \$72,000 \$108,000	\$216,000 \$432,000 \$648,000	
		Mejoramiento y cualificación de la demanda por información	\$120,000 \$240,000 \$360,000	\$120,000 \$240,000 \$360,000	\$120,000 \$240,000 \$360,000	\$120,000 \$240,000 \$360,000	\$120,000 \$240,000 \$360,000	\$120,000 \$240,000 \$360,000	\$120,000 \$240,000 \$360,000	\$720,000 \$1,440,000 \$2,160,000	
<b>TOTALES (USD)</b>			<b>AÑO 1</b>	<b>AÑO 2</b>	<b>AÑO 3</b>	<b>AÑO 4</b>	<b>AÑO 5</b>	<b>AÑO 6</b>	<b>TOTAL</b>	<b>TOTAL COP (miles millones)</b>	
			<b>BAJA</b>	<b>\$2,951,000</b>	<b>\$831,000</b>	<b>\$831,000</b>	<b>\$156,000</b>	<b>\$156,000</b>	<b>\$156,000</b>	<b>\$5,081,000</b>	
			<b>MEDIA</b>	<b>\$5,902,000</b>	<b>\$1,662,000</b>	<b>\$1,662,000</b>	<b>\$312,000</b>	<b>\$312,000</b>	<b>\$312,000</b>	<b>\$10,162,000</b>	
			<b>ALTA</b>	<b>\$8,853,000</b>	<b>\$2,493,000</b>	<b>\$2,493,000</b>	<b>\$468,000</b>	<b>\$468,000</b>	<b>\$468,000</b>	<b>\$15,243,000</b>	
										<b>\$ 15,243.0</b>	
										<b>\$ 30,486.0</b>	
										<b>\$ 45,729.0</b>	

## 7. CONCLUSIONES

A lo largo de este documento se han revisado los retos presentes y futuros que enfrenta la sociedad ante el proceso de profundización de la generación, uso y difusión de datos. Los temas de aprovechamiento de datos, *Big Data* e Inteligencia Artificial se convierten en prioridades para desarrollar nuevas capacidades analíticas en el país, para fomentar la economía digital y promover la adopción de un gobierno movido por los datos.

Es fundamental, prioritario y urgente que se expida un cuerpo normativo articulado que promueva la implementación de estrategias de Big Data. El cuerpo debe tener una pretensión de integralidad de manera que se armonicen las iniciativas ya existentes y al mismo tiempo se permita flexibilidad para el fácil ajuste de las necesidades que surjan en el mediano plazo, dado el cambio acelerado y constante en el campo de la tecnología. El cuerpo normativo debe desarrollarse en paralelo al ecosistema de Big Data nacional, por lo tanto deben existir revisiones periódicas de las limitaciones del marco normativo establecido, con la finalidad de promover ajustes y mejoramiento constante. No se trata de aumentar la regulación al sector, pues esto es costoso y en muchos casos impide el desarrollo óptimo de este, más bien, se trata de obtener un marco armonizado para que las regulaciones no obstaculicen, sino que habiliten el Big Data en el país.

La conformación de comités éticos en torno al aprovechamiento de datos, Big Data e Inteligencia Artificial es necesario para la promoción de buenas prácticas y la defensa de la protección de la información. El comportamiento de los actores que salvaguardan la información debe ser adecuado para evitar ruptura en la confianza del ecosistema y garantizar sostenibilidad. La promoción de buenas prácticas tiene como actor central el gobierno nacional, pero requiere de la participación comprometida de todos los actores sociales tanto en su promoción como en el compromiso de apoyo a la existencia de estos.

Aunque actualmente existen múltiples políticas, programas, iniciativas y acciones descentralizadas que promueven la producción de información, así como el intercambio y la generación de plataformas que faciliten el uso de éste, es vital para el fortalecimiento del ecosistema la aceleración de los procesos de articulación de estas. No es posible una consolidación integral y coherente de las relaciones entre actores del ecosistema sin lineamientos, estándares y procesos claramente definidos que faciliten la emergencia de armonía durante el desarrollo del ecosistema. Las políticas públicas de promoción no deben estar aisladas; esta tarea posiblemente recae en el Departamento Nacional de Planeación, o en una oficina o agencia especial que se le asigne la tarea específica de coordinación y articulación. Es urgente que se decida una línea de acción en esta materia dado que la implementación de políticas desarticuladas garantiza la fragilidad y debilidad del sistema, al mismo tiempo que entorpece su adecuado desarrollo.

Por otra parte, el CONPES 3920 ha identificado la interoperabilidad como una prioridad en materia de aprovechamiento de datos y nuestro trabajo lo ha hecho de igual manera. Una vez se alcancen los desarrollos deseados de las bases de datos y sistemas de información existentes en el país, se hace necesario que se profundicen los procesos de mejoramiento de la interoperabilidad dentro del ecosistema nacional. El diálogo constante y permanente entre los sistemas existentes es urgente y hemos identificado que el estado actual no lo permite, incluso lo dificulta. Por lo tanto es urgente la intervención y proponemos que se adopten métricas claras desde el diseño de las acciones a ejecutar para monitorear de manera constante el proceso de mejoramiento e integración de la capacidad de interoperabilidad nacional.

En la misma línea debemos insistir en que los estándares de calidad deben expandirse a grupos de datos e información transversales en el ecosistema. En este momento solo se puede garantizar rigurosidad en el proceso de captura y estructura de sistemas de información de un gran número de bases de datos del gobierno nacional, fundamentalmente las que han sido permeadas por los estándares internacionales de la producción estadística, y algunos sistemas de datos e información de organizaciones por fuera de la esfera gubernamental. Es prioritario expandir los criterios de calidad a todos los procesos de producción, procesamiento, analítica, uso y difusión de manera que se pueda contar con estándares equiparables para todos los actores que hacen parte del ecosistema.

Finalmente, el temor y miedo que se deriva de los nuevos retos que emergen con la llegada y adopción de nuevas tecnologías, discutidos en la introducción de este documento, deben ser mitigados a través de políticas que permitan generar mayor confianza y legitimidad sobre la tecnología. La adopción de las nuevas tecnologías tiene altísimos potenciales y beneficios al mismo tiempo que acarrea muchos riesgos que deben ser cuidadosamente discutidos y debatidos para alcanzar óptimos sociales necesarios para garantizar los mejores resultados posibles. La promoción del diálogo recae en la sociedad en general y cada uno de los actores debe ser responsable de hacer parte activa del mismo.

## 8. ANEXOS

---

### ANEXO A. ANONIMIZACIÓN

#### Preanonymización

*¿En qué consiste la preanonymización?* En esta etapa se debe determinar claramente cuáles son las variables y los identificadores que de manera directa o indirecta pueden aportar en la identificación de una persona. Para tal fin, es necesario clasificar cada una de las variables de información personal recolectadas para incorporarlas en el archivo o base de datos, teniendo en cuenta las siguientes categorías: a) variables de identificación personal; b) variables de identificación directa de personas o empresas; c) variables de carácter sensible o confidencial; d) variables con o sin restricción para el acceso público.

Estas variables se clasificarán como directas o indirectas según puedan llevar a la identificación de una persona. Asimismo, se debe procurar minimizar la cantidad de información personal a recolectar, especialmente con aquellos datos que contienen información sensible.

*¿En qué momento se hace la preanonymización?* Este proceso se debe realizar paralelo a la definición de las variables que se tendrán en cuenta en cada estudio, encuesta poblacional, o generalmente, en la etapa de recolección de información.

#### Anonymización de microdatos

*¿En qué consiste la anonymización del microdato?* Es la implementación de las actividades de protección de la privacidad de los datos sobre las características de la información una vez hagan parte de las bases de datos de una empresa o entidad. Su finalidad es filtrar la información estrictamente necesaria que se requiera para el tratamiento de los datos y, de esta forma, evitar que otros colaboradores de la compañía o entidad tengan acceso ilimitado a las variables de identificación. En esta forma, se entiende que la posibilidad de extraer datos a partir de una agregación permite la identificación de la fuente, de manera que la implementación de un proceso adecuado de anonymización evita la posibilidad de dar mal uso a la información disagregada (microdatos).

*¿En qué momento se hace la anonymización del microdato?* La anonymización del microdato se debe realizar en el proceso de recolección y registro en la base de datos y está encaminado a evitar las identificaciones directas o indirectas de individuos. Para ello, se deben tener en cuenta identificaciones indirectas, las cuales se puede dar por las características compartidas

por varios de los encuestados, y cuya combinación podría conducir a la reidentificación de uno de ellos. Por ejemplo, se podría identificar la combinación de variables tales como el lugar de residencia, edad, sexo y profesión, si sólo una persona de ese determinado sexo, edad y profesión vivió en ese lugar particular.

*¿Quiénes participan en esta anonimización del micro dato?* En primera instancia, los administradores de las bases de datos, quienes tienen a cargo de la custodia de la información; el equipo de sistemas se encargará de pseudoanonimizar los datos. Es decir, tienen la función de ocultar la identidad de las personas mediante seudónimos, ocultamiento y/o supresión de los datos personales según requerimiento.

### Anonimización

*¿En qué consiste la anonimización?* Consiste en la aplicación de una o varias técnicas tendientes a eliminar al máximo el riesgo de identificación de las personas o entidades, con el menor daño posible a los datos.

*¿En qué momento se hace la anonimización?* Una vez finalizada la anonimización de microdatos de uso interno se puede proceder con la anonimización.

*¿Quiénes participan en la anonimización?* Quien delegue la compañía para establecer la técnica de anonimización más apropiada para cada una de las variables, según los recursos más actuales y disponibles. Asimismo, la implementación de los algoritmos propuestos será efectuada por el área de tecnología quien se encargará de realizar las pruebas necesarias a fin de establecer la eficacia de la técnica.

De esta forma, las técnicas de anonimización les permitirá a las entidades públicas y compañías del Estado colombiano reducir los riesgos de identificación y re-identificación de los datos de una persona que puedan llevar a un perjuicio sobre la misma.

Tabla 2. Inversión en Estrategias de Big Data/AI por País

País	Estrategia de Datos/AI	Duración	Inversión	Gasto Anual en USD (Cifras 2018)	% de PIB Anual USD (%/"10^4) (Cifras 2017)	% de Presupuesto Anual USD (%/"10^4) (Cifras 2017)	Notas Adicionales
Alemania	IA Hecha en Alemania	7 años	3 billones EUR, 3.4 billones USD	426 millones	0.011510	115.8	0.026313
Australia	Australia 2030: Prosperidad a través de la innovación	4 años	29.9 millones AUD, 21.3 millones USD	7.6 millones	0.000574	5.7	0.002744
Canada	Estrategia Pan-canadiense de Inteligencia Artificial	4 años	125 millones CAD, 94 millones USD	31.25 millones	0.001891	19.9	0.009232
China	Parque Tecnológico para el desarrollo de AI	5 años	13.8 billones CNY, 2.1 billones USD	42 millones	0.000343	3.4	0.001269
Colombia	Política Nacional de Explotación de Datos (Big Data)	4 años	16.728 millones COP, 5.3 millones USD	1.176 millones	0.000380	3.8	0.001282
Corea del Sur	Inversión en AI	4 años	2.2 trillones KRW, 1.9 billones USD	475 millones	0.030844	0.141453	1414.5
Dinamarca	Estrategia de Dinamarca para el Crecimiento Digital	4 años	1.025 millones DKK, 157 millones USD	19.2 millones	0.005897	59	0.011368
Francia	Por Una Inteligencia Artificial Significativa: Avanzando Hacia una Estrategia Francesa y Europea	5 años	1.5 billones EUR, 1.7 billones USD	34 millones	0.001314	13.1	0.002330
Mexico	Estrategia Digital Nacional	5 años	(2019) 21.4 millones MXN, 1.1 millones USD	1.1 millones	0.000096	0.9	0.000402
Reino Unido	Acuerdo Sectorial de Inteligencia Artificial	3 años	0.95 billion EUR,	316.6 millones	0.000120		0.000293
Singapur	AI Singapur	5 años	150 millones de SGD, 110.6 millones USD	22.12 millones	0.006829	68	0.042645
							426.5

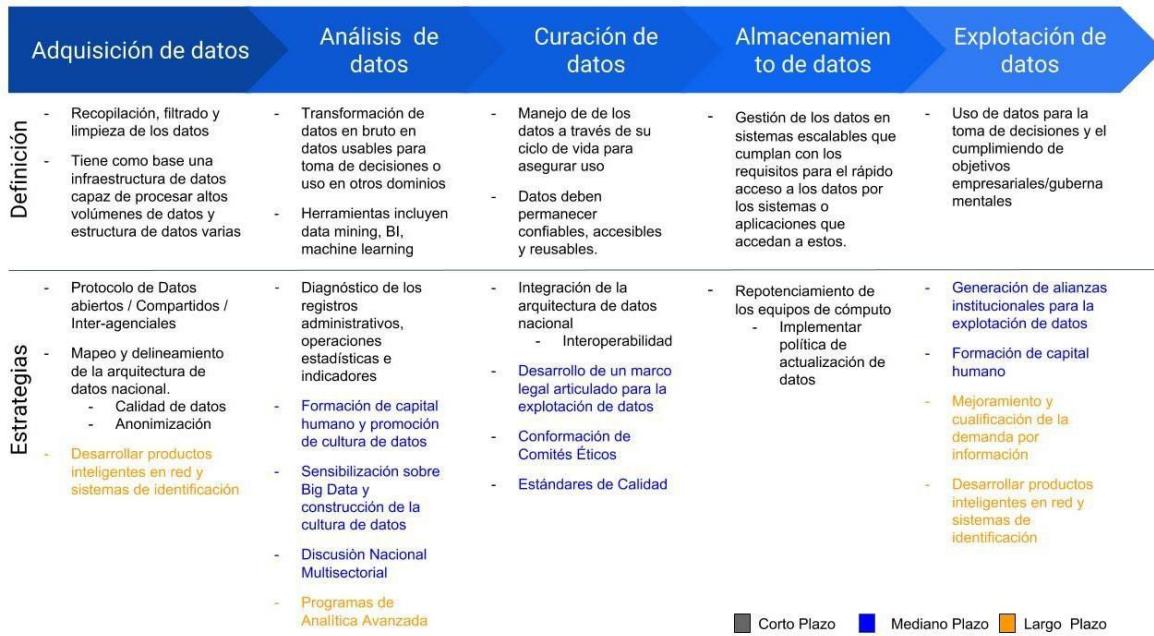
Fuente: Elaboración propia

## ANEXO B. INVERSIÓN ESTRATEGIAS BIG DATA/AI POR PAÍS

### ANEXO C. CADENA DE VALOR DE BIG DATA

Para estructurar la cadena de valor de esta estrategia, se tomó como base la cadena de valor de la Big Data desarrollada por Edward Curry y colaboradores en 2014. A continuación, puede apreciarse en primer lugar, cada etapa en la cadena de valor desde que el dato es recopilado y no se encuentra estructurado hasta que está listo para su explotación. En segundo lugar, se define cada etapa en líneas generales y por lo tanto se exponen los puntos clave usados en las líneas de acción propuestas anteriormente para lograr darle un valor añadido a la información. Una vez determinado esto, el tercer lugar identifica cada una de las acciones propuestas en la cadena, teniendo como premisa cubrir cada paso de la cadena de valor para lograr el objetivo final.

Figura 6. Cadena de Valor para la Estrategia Nacional de Big Data



Fuente: Elaboración propia

## BIBLIOGRAFÍA

---

- “Making Sense of Big Data”. 2016. White Paper. Federal Data Analytics. CDW. [https://fedtechmagazine.com/sites/fedtechmagazine.com/files/document\\_files/MKT1\\_4585\\_Federal-Data-Analytics-WP%281%29.pdf](https://fedtechmagazine.com/sites/fedtechmagazine.com/files/document_files/MKT1_4585_Federal-Data-Analytics-WP%281%29.pdf).
- “MIDATA.coop | My Data - Our Health”. s/f. Midata. Consultado el 15 de enero de 2019. <https://www.midata.coop>.
- “PROGRAMAS DE EDUCACIÓN SUPERIOR | Datos Abiertos Colombia”. 2017. el 16 de agosto de 2017. [https://www.datos.gov.co/Educaci-n/PROGRAMAS-DE-EDUCACI-N-SUPERIOR/up\\_r9-nkiz](https://www.datos.gov.co/Educaci-n/PROGRAMAS-DE-EDUCACI-N-SUPERIOR/up_r9-nkiz).
- Abarca, K., 2018. *Big data y las políticas públicas*. [En línea] Available at: <https://www.americaeconomia.com/analisis-opinion/big-data-y-las-politicas-publicas>
- Alianza CAOBA, 2018. *Alianza CAOBA*. [En línea]
- Available at: <http://alianzacaoba.co> [Último acceso: 3 Julio 2018].
- Alianza Gobierno Abierto Colombia, 2016. *Colombia, hacia un estado abierto*, Bogotá D.C.: Alianza Gobierno Abierto Colombia.
- Arquitectura TI Colombia, 2014. *Guía básica del dominio de Información*, Bogotá D.C.: Gobierno Nacional.
- Arquitectura TI Colombia, 2016. *Guía General de un Proceso de Arquitectura Empresarial*, Bogotá D.C.: Gobierno Nacional.
- Balbontín, R., Roeschmann, J. A. & Zahler, A., 2018. *Ciencia, Tecnología e Innovación en Chile: un análisis presupuestario*, Santiago de Chile: Gobierno de Chile.
- Banco Interamericano de Desarrollo, 2016. *Big Data in the Public Sector*, s.l.: Banco Interamericano de Desarrollo.
- Bhargava, R. y otros, 2015. *Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data*, s.l.: Data-Pop Alliance.
- Camargo, J. J., 2014. Conociendo Big Data. *Revista Facultad de Ingeniería*, pp. 63-77. CEPAL, 2012. *Datos abiertos: Un nuevo desafío para los gobiernos de la región*, Santiago de Chile: Naciones Unidas.
- CEPAL, 2014. *Plan de gobierno abierto Una hoja de ruta para los gobiernos de la región*, Santiago de Chile: Naciones Unidas.
- CEPAL, 2015. *Metodología del marco lógico para la planificación, el seguimiento y la evaluación de proyectos y programas*. Naciones Unidas ed. Nueva York: CEPAL.
- Colombia Compra Eficiente, 2018. *Colombia Compra Eficiente*. [En línea]
- Available at: <https://www.colombiacompra.gov.co> [Último acceso: 3 Junio 2018].
- Comisión Estadística de las Naciones Unidas, 2013. *Macrodatos y modernización de los sistemas estadísticos*, s.l.: Naciones Unidas: Consejo Económico y Social.
- Comisión Nacional de Moralización, 2017. *Declaración de Compromisos por un Estado Abierto*, Bogotá D.C.: Comisión Nacional de Moralización.
- Consejo Nacional de Ciencia y Tecnología, 2016. *Informe general del Estado de la Ciencia, la Tecnología y la Innovación*, México D.F.: Consejo Nacional de Ciencia y Tecnología.
- Coppola, A., Siegel, R., Sabet, E. & Calvo-Gonzalez, O., 2014. *Big Data in Action for Development Report*, s.l.: World Bank Group.
- Corte Constitucional, 2014. Artículo 15. Decreto 1649 del 2 de septiembre de 2014, Bogotá D.C.: Corte Constitucional.
- Cukier, K. & Mayer-Schönberger, V., 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, London: John Murray Publishers.
- Chamakkala, Vipin. 2018. “The Future of Data Infrastructure From Source to Insight”.

Medium (blog). el 28 de marzo de 2018.

<https://medium.com/work-bench/the-future-of-data-infrastructure-from-source-to-insight-b299c6b7cc30>.

Cregeen, Alison. 2018. "A Practical Guide to Data Mapping for GDPR Compliance".

PwC. el 6 de marzo de 2018.

<https://www.pwc.com/im/en/media-room/articles/a-practical-guide-to-data-mapping-guide.html>

Cupoli, P., 2014. *DAMA-DMBOK2 Framework*, Chicago: The DAMA Management Association.

D'Ignazio, C. & Bhargava, R., 2015. *Approaches to Building Big Data Literacy*, Nueva York: MIT.

Dans, E., 2011. *Enrique Dans*. [En línea] Available at:

<https://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html>

Data Revolution Group, 2014. *Data Revolution Group*. [En línea]

Available at: <http://www.udatarevolution.org/> [Último acceso: 17 Julio 2018].

Data Therapy, 2017. *You Don't Need a Data Scientist, You Need a Data Culture*, s.l.: Data Therapy.

Data-Pop Alliance, 2018. *Data-Pop Alliance*. [En línea] Available at:

<http://datapopalliance.org/resources/key-terms/#statistical-machine-learning> [Último acceso: 10 Julio 2018]

DATAVERSITY, 2018. *DATAVERSITY*. [En línea]

Available at: <http://www.dataversity.net/what-is-data-management/> [Último acceso: 26 Mayo 2018].

Departamento Administrativo Nacional de Estadística, 2014. *DANE participa en seminario de Big Data y Revolución de Datos*, Bogotá D.C.: DANE.

Departamento Nacional de Planeación, 2018. *Colombia primer país en Latinoamérica con una política pública para la explotación de datos Big Data*, Bogotá D.C.: Departamento Nacional de Planeación.

Departamento Nacional de Planeación, 2018. *Documento CONPES 3920*, Bogotá D.C.: Gobierno Nacional.

Digital Persevation Management, 2017. *Digital Persevation Management*. [En línea]

Available at: <http://lib-arc-1.mit.edu/dpm-eng/foundation/oais/data.html> [Último acceso: 3 Julio 2018].

Dirección de Impuestos y Aduanas Nacionales, 2017. *Dirección de Impuestos y Aduanas Nacionales*. [En línea] Available at: <https://www.dian.gov.co/impuestos/Reforma%20Tributaria%20Estructural/Concepto%20Unificado%20del%20Numeral%2024%20del%20Art%C3%ADculo%20187%20de%20la%20ley%201819%20de%202016.pdf> [Último acceso: 2 Julio 2018].

Estatuto Tributario Nacional, 2017. *Estatuto Tributario Nacional*. [En línea] Available at: <http://estatuto.co/?e=710> [Último acceso: 15 Julio 2018].

European Commission, 2017. *Final results of the European Data Market study measuring the size and trends of the EU data economy*, s.l.: European Commission.

Ferguson, R., Kiron, D. & Prentice, P., 2013. *Del valor a la visión: reimaginando lo posible con la Analítica de Datos*, s.l.: MIT Sloan Management Review.

FedTech. 2016. "4 Infrastructure Requirements for Any Big Data Initiative". Technology Solutions That Drive Government. el 22 de diciembre de 2016. <https://fedtechmagazine.com/article/2016/12/4-infrastructure-requirements-any-big-data-initiative>.

Financial Times, 2013. *Financial Times*. [En línea]

Available at: <https://ig.ft.com/how-much-is-your-personal-data-worth/> [Último acceso: 3 Julio 2018].

Función Pública, 2018. *Función Pública*. [En línea] Available at: [http://www.funcionpublica.gov.co/preguntas-frecuentes/-/asset\\_publisher/sqxafjubsrEu/content/-que-es-el-fura-1/28585938](http://www.funcionpublica.gov.co/preguntas-frecuentes/-/asset_publisher/sqxafjubsrEu/content/-que-es-el-fura-1/28585938) [Último acceso: 3 Julio 2018].

García, J., 2018. *Ciencia de Datos. Técnicas analíticas y aprendizaje estadístico*. Bogotá D.C.: Publicaciones Altaria.

Gobierno Nacional, 2006. *Plan Estratégico de Información Básica 2006 - 2010*, Bogotá D.C.: Gobierno Nacional.

- Gobierno Nacional, 2010. *Marco para la Interoperabilidad del Gobierno en Línea*, Bogotá D.C.: Gobierno Nacional.
- Gobierno Nacional, 2012. *Manual para la implementación de la Estrategia de Gobierno en Línea*, Bogotá D.C.: Gobierno Nacional.
- Gobierno Nacional, 2014. *Oferta Institucional de la Secretaría de Transparencia*, Bogotá D.C.: Gobierno Nacional.
- Gobierno Nacional, 2017. *Arquitectura TI en Colombia*. [En línea]
- Available at: <http://www.mintic.gov.co/arquitecturati/630/w3-propertyvalue-8117.html> [Último acceso: 3 Julio 2018].
- Harvard Business Review, 2013. Harvard Business Review. [En línea] Available at: <https://hbr.org/2013/07/five-roles-you-need-on-your-bi> [Último acceso: 1 Julio 2018].
- Heimes, Rita. 2018. "Top 10 operational responses to the GDPR: Data inventory and mapping". International Association of Privacy Professionals. el 1 de febrero de 2018.
- <https://iapp.org/news/a/top-10-operational-responses-to-the-gdpr-data-inventory-and-mapping/>.
- Informatica. s/f. "What is Master Data Management: Definition". Informatica. Consultado el 14 de enero de 2019.
- <https://www.informatica.com/services-and-training/glossary-of-terms/master-data-management-definition.html#fbid=-aYjsX39IV4>.
- International Organization for Standardization, 2017. International Organization for Standardization. [En línea]
- Available at: <https://www.iso.org/standard/61932.html> [Último acceso: 15 Junio 2018].
- ITU Data, 2017. ICU Development Index. [En línea] Available at: <http://www.itu.int/net4/ITU-D/idi/2017/index.html#idi2017economycard-tab&COL> [Último acceso: 3 Julio 2018].
- Jaffe, Eric, 2016. Medium. [En línea] Available at: <https://medium.com/sidewalk-talk/how-estonia-became-a-global-model-for-e-government-c12e5002d818> Jütting, J. & Letouzé, E., 2015. Official Statistics, Big Data, and Human Development, s.l.: Data-Pop Alliance.
- Kelleher, J. & Tierney, B., 2018. Data Science. The MIT Press essential knowledge series, s.l.: The MIT Press.
- Kitchin, R., 2014. The Data Revolution: Big Data, open Data, Data Infrastructures and Their Consequences, Los Ángeles: Sage.
- Klievink, B., Romijn, B.-J., Cunningham, S. & de Bruijn, H., 2017. Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers: a journal of research and innovation*. SpringerNature, 19(2), pp. 267-283.
- Klievink, B., A. Neuroni, M. Fraefel, y A. Zuiderwijk. 2017. "Digital Strategies in Action: A Comparative Analysis of National Data Infrastructure Development". En *Proceedings of the 18th Annual International Conference on Digital Government Research - Dg.o '17*, 129–38. Staten Island, NY, USA: ACM Press. <https://doi.org/10.1145/3085228.3085270>.
- LaValle, S. y otros, 2011. Big Data, Analytics and the Path From Insights to Value.
- MIT Sloan Management Review, 52(2).
- Lazer, D., Kennedy, R., King, G. & Vespiagnani, A., 2014. Science. [En línea] Available at: <http://science.sciencemag.org/content/343/6176/1203> [Último acceso: 17 Julio 2018].
- Letouzé, E., 2012. Big Data for Development: Challenges & Opportunities, s.l.: Global Pulse.
- Letouzé, E., 2015. Big Data & Development an overview, s.l.: Data-Pop Alliance.
- Letouzé, E. & Sangokoya, D., 2015. Leveraging Algorithms for Positive Disruption: On data, democracy, society and statistics, s.l.: Data-Pop Alliance.
- Letouzé, E., Vinck, P. & Kammourieh, L., 2015. The law, politics and ethics of cell phone data analytics, s.l.: Data-Pop Alliance.
- Linkedin, 2016. Linkedin. [En línea] Available at: <https://www.linkedin.com/pulse/cloud-vs-on-premises-hard-dollar-costs-greg-deckler> [Último acceso: 2 Julio 2018].
- Lockner, Julie. 2017. "Four Requirements of the Modern Enterprise Data Infrastructure". IT Pro Portal. el 31 de julio de 2017. <https://www.itproportal.com/features/four-requirements-of-the-modern-enterprise-data-infrastructure/>.
- Marcus, G. & Davis, E., 2014. The New York Times. [En línea] Available at: [https://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?\\_r=2](https://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?_r=2) [Último acceso: 17 Julio 2018].

McKinsey Global Institute, 2011. Big data: The next frontier for innovation, competition, and productivity, s.l.: McKinsey Global Institute.

Membrey, P., Plugge, E. & DUPTim, H., 2010. The Definitive Guide to MongoDB The NoSQL Database for Cloud and Desktop Computing. s.l.:Apress.

Ministry of Industry, Business and Financial Affairs. 2018. "Strategy for Denmark's Digital Growth". The Danish Government. [https://eng.em.dk/media/10566/digital-growth-strategy-report\\_uk\\_web-2.pdf](https://eng.em.dk/media/10566/digital-growth-strategy-report_uk_web-2.pdf).

Ministerio de Información de la Tecnología y las Comunicaciones, 2016. *Nace CAOBA, primera alianza público privada en Big Data*, Bogotá D.C.: Ministerio de Información de la Tecnología y las Comunicaciones.

Ministerio de Tecnología de la Información y las Comunicaciones, 2017. *Ministerio de Tecnología de la Información y las Comunicaciones*. [En línea]

Available at: [http://www.mintic.gov.co/portal/604/articles-54877\\_doc\\_pdf.pdf](http://www.mintic.gov.co/portal/604/articles-54877_doc_pdf.pdf) [Último acceso: 6 Julio 2018].

Ministerio de Tecnología de la Información y las Comunicaciones, 2018. Ministerio de Tecnología de la Información y las Comunicaciones. [En línea]

Available at: <https://www.mintic.gov.co/portal/vivedigital/612/w3-article-19487.html> [Último acceso: 6 Julio 2018].

Ministerio de Tecnologías de la Información y las Comunicaciones, 2010. Marco para la Interoperabilidad del Gobierno en Línea, Bogotá D.C.: Gobierno Nacional.

Ministerio de Tecnologías de la Información y las Comunicaciones, 2011. *Lenguaje común de intercambio de información: arquitectura de datos*, Bogotá D.C.: Gobierno Nacional.

Ministerio de Tecnologías de la información y las Comunicaciones, 2018. *Ministerio de Tecnologías de la información y las Comunicaciones*. [En línea]

Available at: <http://www.mintic.gov.co/portal/604/w3-propertyvalue-677.html> [Último acceso: 1 Julio 2018].

Ministerio de Tecnologías de la Información y las Comunicaciones, 2018. *Ministerio de Tecnologías de la Información y las Comunicaciones*. [En línea]

Available at: <http://lenguaje.mintic.gov.co/marco-de-interoperabilidad> [Último acceso: 5 Julio 2018].

Ministerio de Tecnologías de la Información y las Comunicaciones, 2018. *Ministerio de Tecnologías de la Información y las Comunicaciones*. [En línea]

Available at: <http://www.mintic.gov.co/arquitecturati/630/w3-propertyvalue-8117.html> [Último acceso: 5 Julio 2018].

Naciones Unidas, 2005. Global E-Government readiness report 2005, Nueva York: Naciones Unidas.

Observatorio de TI, 2017. Observatorio de TI. [En línea] Available at: <https://www.observatorioti.gov.co/> [Último acceso: 2 Julio 2018].

Observatorio Transparencia y Anticorrupción, 2018. Observatorio Transparencia y Anticorrupción. [En línea]

Available at: <http://www.anticorruption.gov.co/Paginas/index.aspx> [Último acceso: 23 Junio 2018].

OCDE, 2004. *Principios de Gobierno Corporativo de la OCDE*, Paris: OCDE. OCDE, 2015. Students, Computers and Learning, s.l.: OCDE.

OECD, 2003. *Implementing Good Governance*, Francia: OECD Publishing.

OECD, 2015. *Data-Driven Innovation*. Paris: OECD.

OECDiLibrary, 2017. *OECDiLibrary*. [En línea] Available at:

[https://www.oecd-ilibrary.org/science-and-technology/exploring-data-driven-innovation-as-a-new-source-of-growth\\_5k47zw3fc43-en](https://www.oecd-ilibrary.org/science-and-technology/exploring-data-driven-innovation-as-a-new-source-of-growth_5k47zw3fc43-en) [Último acceso: 21 Junio 2018].

Office of Science and Technology Policy, 2012. Obama Administration Unveils "Big Dat" initiative: announces \$200 million in new R&D investments, s.l.: Office of Science and Technology Policy.

Organización de las Naciones Unidas, 2013. Organización de las Naciones Unidas. [En línea]

Available at: [https://unstats.un.org/UNSD/statcom/statcom\\_2013/seminars/Big\\_Data/default.html](https://unstats.un.org/UNSD/statcom/statcom_2013/seminars/Big_Data/default.html) [Último acceso: 3 Junio 2018].

PARIS21, 2015. Road Map for a country-led, s.l.: Data Revolution.

Procuraduría General de la Nación, 2014. *Ley de Transparencia y del derecho de acceso a la información pública*, Bogotá D.C.: Procuraduría General de la Nación.

Perret, Rick. 2014. “3 Key IT Infrastructure Requirements for Big Data and Analytics”.

IBM IT Infrastructure Blog. el 6 de noviembre de 2014. <https://www.ibm.com/blogs/systems/3-key-it-infrastructure-requirements-for-big-data-and-analytics/>.

Schmarzo, B., 2015. Big Data: El poder de los datos, s.l.: Wiley.

Sistema Estadístico Nacional, 2017. *Plan Estadístico Nacional 2017 - 2022*, Bogotá D.C.: Departamento Administrativo Nacional de Estadística.

Sistema Estadístico Nacional, 2017. *Plan Estadístico NACIONAL 2017 - 2022*, Bogotá D.C.: Gobierno Nacional.

Stack OverFlow, 2018. Stack OverFlow. [En línea]

Available at: <https://insights.stackoverflow.com/survey/2018/#salary> [Último acceso: 3 Julio 2018].

Ubiquitous Web domain, 2007. Ubiquitous Web domain. [En línea]

Available at: <https://www.w3.org/XML/Schema> [Último acceso: 4 Julio 2018].

Unidad Administrativa Especial de Castastro Distrital, 2016. Unidad Administrativa Especial de Castastro Distrital. [En línea]

Available at: <https://www.ideca.gov.co/es/infraestructura-de-datos-espaciales> [Último acceso: 16 Mayo 2018].

Unión Europea, 1997. Directiva 95/46/CE Del Parlamento Europeo y del Consejo de 24 de octubre de 1995 relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, s.l.: Agencia de Protección de Datos.

USAID, 2012. The Logical Framework Approach, s.l.: USAID. VivelabBogotá, 2018. VivelabBogotá. [En línea]

Available at: <http://vivelabbogota.com/index.php/service/acerca-de-nosotros/>

World Economic Forum, 2015. World Economic Forum. [En línea] Available at:

<https://www.weforum.org/agenda/2015/12/how-is-big-data-going-to-change-the-world/> [Último acceso: 6 Julio 2018].

Zetony, David A. 2018. “Data Maps and Data Inventories”. Bryan Cava Leighton Paisner. el 10 de enero de 2018.

<https://www.bryancave.com/en/thought-leadership/data-maps-and-data-inventories.html>.

