

JULIO 2021

MODELO DE EXPLOTACIÓN DE DATOS PARA LAS ENTIDADES PÚBLICAS



El futuro
es de todos

Consejería Presidencial
para asuntos económicos
y transformación digital



El futuro
es de todos

DNP
Departamento
Nacional de Planeación



El futuro digital
es de todos

MinTIC



Luis Alberto Rodríguez Ospino

Director General

Daniel Gómez Gaviria

Subdirector General Sectorial

Amparo García Montaña

Subdirectora General Territorial

Diana Patricia Ríos García

Secretaria General

Iván Mauricio Durán Pabón

Director de Desarrollo Digital

Viviana Vanegas Barrero

Subdirectora de Prospectiva Digital

Este estudio ha contado con el apoyo de los asesores:

Infométrika S.A.S

Carlos Alberto Barreto Nieto

Liliana Fernández Gómez

Diana Paola Ramírez Roa

Agustín Jiménez Ospina

Apoyo técnico, diagramación y diseño:

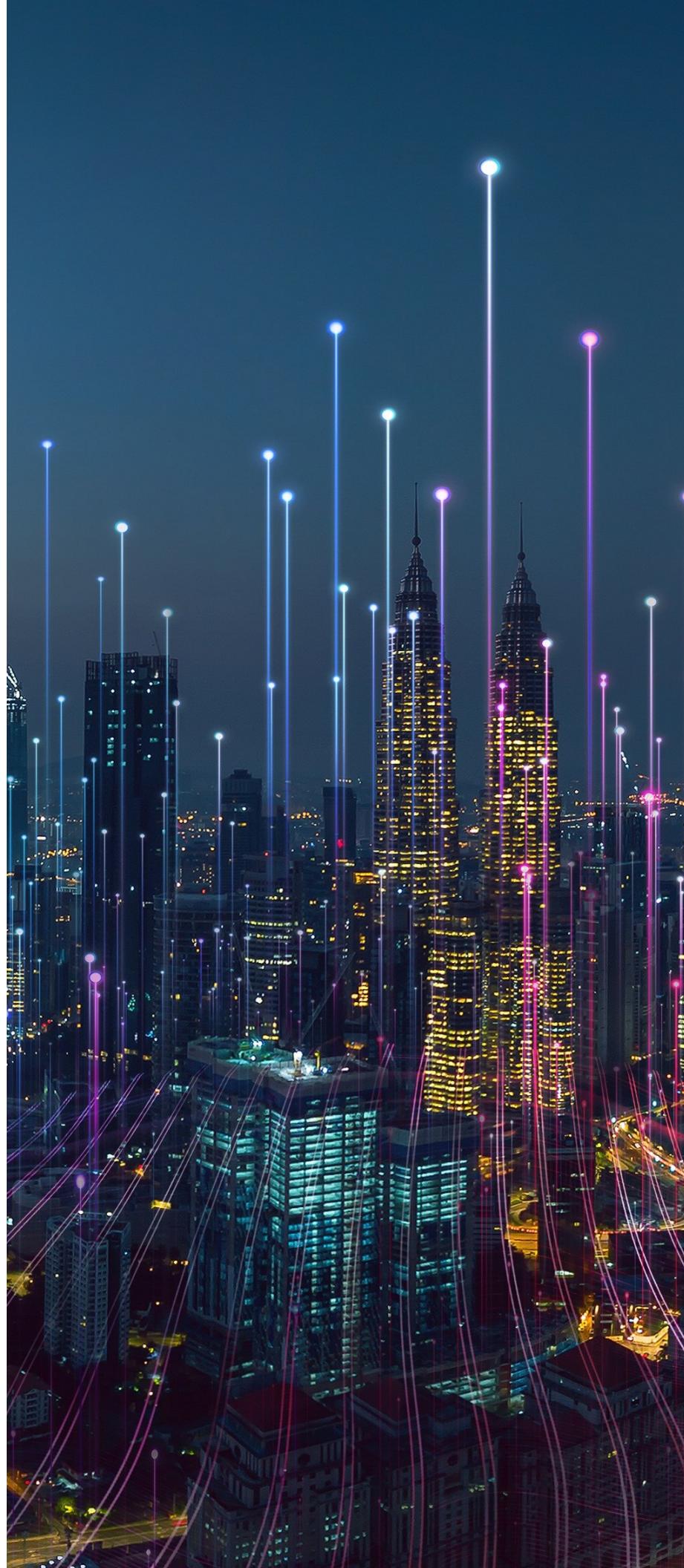
Consejería Presidencial para

Asuntos Económicos y

Transformación Digital

Aliados:

Ministerio de las Tecnologías de la Información y las Comunicaciones





MODELO DE EXPLOTACIÓN DE DATOS PARA LAS ENTIDADES PÚBLICAS

Documento definitivo

Grupo de Comunicaciones
y Relaciones Públicas

Luis Segundo Gamez Daza
Coordinador

Carmen Elisa Villamizar Camargo
Corrección de estilo

©Departamento Nacional de Planeación,
Calle 26 núm. 13-19 Bogotá, D. C.
PBX: 3815000
Diciembre de 2020



A minimalist design featuring a background of fine, wavy gray lines that curve from the top left towards the bottom right. In the top left corner, there is a small red plus sign (+). In the bottom left corner, there is a small red minus sign (-).

CONTENIDO

1 | MODELO DE IMPLEMENTACIÓN DE EXPLOTACIÓN DE DATOS

pág 10-17

2 | ECOSISTEMA DE EXPLOTACIÓN DE DATOS

pág 18-21

3 | DESCRIPCIÓN DE LOS COMPONENTES DEL MODELO DE EXPLOTACIÓN DE DATOS

pág 22-72

4 | RECOMENDACIONES FINALES A LAS ENTIDADES PÚBLICAS PARA LA APLICACIÓN DEL MODELO DE IMPLEMENTACIÓN DE EXPLOTACIÓN DE DATOS

pág 73-75

5 | BIBLIOGRAFÍA

pág 76-81



ANEXOS

Anexo 0.

Presentación Modelo implementación de explotación de datos

Anexo 1.

Formulario explotación de datos

Anexo 1.1

Guía de diligenciamiento del formulario

Anexo 2.

Matriz descripción de niveles por dimensión

Anexo 3.

Mapa de acciones

Anexo 4.

Metodologías de cálculo del modelo de explotación de datos

Anexo 5.

Ejemplo de cálculo

Anexo 6.

Arquitecturas de referencia para *big data*

Anexo 7.

Resultados piloto 60 entidades

• • • • • • • • • • • • • • • • • • •

+

INTRODUCCIÓN

.....



-



Los datos constituyen la columna vertebral de la transformación digital, y son los habilitadores de las tecnologías de la cuarta revolución industrial. La creciente interacción de datos, el desarrollo de la inteligencia artificial y el análisis de *big data* traen consigo gran cantidad de oportunidades para los gobiernos, los ciudadanos y el sector privado (OCDE, 2020). De acuerdo con la OCDE, en la actualidad, los datos se constituyen como un nuevo factor de producción, que impulsa la innovación de bienes y servicios para los ciudadanos, mejora la toma de decisiones basada en evidencia e impulsa la inteligencia del sector público (OCDE, 2019).

Entre los sectores de la economía, el sector público se destaca por ser uno de más intensivos en materia de generación, recolección y transformación de datos (Fraefel, Haller, & Gschwend, 2017). Esto presupone importantes retos para el sector público en materia de gobernanza de datos tanto en el orden nacional como en cada entidad pública, para consolidar sus capacidades en la gestión de los datos, y así garantizar su disponibilidad, accesibilidad, usabilidad, integridad y seguridad (OCDE, 2020).

En Colombia, en el marco de los documentos CONPES 3920: *Política Nacional de Exploración de Datos* y el CONPES 3975: *Política*



Nacional de Transformación Digital e Inteligencia Artificial, junto con la Política de Gobierno Digital y el Marco de Transformación Digital para el Estado, el país ha identificado los datos como activo para la generación de valor social y económico, a fin de mejorar la toma de decisiones y para impulsar la transformación digital en el país. No obstante, la materialización del valor de los datos en el sector público exige que las entidades públicas formen y fortalezcan sus capacidades en materia de talento humano, infraestructura tecnológica y una cultura organizacional basada en datos.

A partir de lo expuesto, el **Modelo de implementación de explotación de datos** para las entidades públicas se constituye en un instrumento conceptual estandarizado, cuyo objetivo consiste en fomentar la cultura de datos en las entidades públicas del país e impulsar el aprovechamiento de datos y *big data*. Lo anterior se consigue al proporcionar a las entidades herramientas para diagnosticar su nivel de madurez para la explotación de datos y *big data*, identificar el valor potencial y brindar recomendaciones para construir una hoja de ruta a fin de mejorar sus capacidades en explotación de datos.

Para hacerlo, en la primera sección de este documento se describe el modelo de implementación de explotación de datos, en la segunda se presenta el ecosistema de explotación de datos en Colombia, en la tercera se detallan los componentes del modelo de explotación de datos, en la cuarta se plantean recomendaciones a las entidades para la elaboración de la hoja de ruta en explotación de datos, y en la quinta y última, se exponen recomendaciones generales a las entidades públicas.

+

1

MODELO DE IMPLEMENTACIÓN DE EXPLOTACIÓN DE DATOS

.....

-



El **modelo de implementación de explotación de datos** es un instrumento conceptual estandarizado que busca fomentar la explotación de datos en las entidades públicas y fortalecer la cultura de datos como activo para impulsar la transformación digital y mejorar la toma de decisiones en el sector público. El modelo permite a las entidades identificar las capacidades que deben desarrollar para mejorar el aprovechamiento de datos y, a su vez, visibilizar el efecto que tiene la explotación y la analítica de datos para generar valor público a partir de la optimización en la prestación de bienes y servicios a los ciudadanos, mejorar la eficiencia en la gestión pública, y diseñar políticas públicas basadas en evidencia.

El diseño del modelo de explotación de datos se elaboró teniendo en cuenta las recomendaciones de la Organización para la Cooperación y el Desarrollo Económico (OCDE) a los países, para convertirse en un sector público basado en datos (OCDE, 2019)¹ y está soportado en una completa revisión bibliográfica. El modelo se construyó con base en la definición de *big data*² como un sistema sociotécnico³ en el que convergen (1) *tecnologías de punta* en constante evolución para enfrentar los retos del procesamiento masivo; (2) *procesos* que garantizan aproximaciones sistemáticas y repetibles; y (3) *personas y organizaciones* que ejecutan procesos y son los beneficiarios de los productos y servicios derivados. En la **figura 1-1** se presenta el esquema de los conceptos que definen el *big data* y la explotación de datos.

Figura 1-1

Definición de big data y explotación de datos en el contexto del modelo propuesto



Fuente: DNP - DDD

1.1 o

COMPONENTES DEL MODELO DE EXPLOTACIÓN DE DATOS

El modelo de explotación de datos está integrado por cuatro componentes que corresponden a (1) la preparación o madurez de las entidades para la explotación de datos, (2) el valor potencial generado por las actividades de explotación de datos y *big data*, (3) necesidades de inversión para avanzar hacia el nivel de madurez esperado y estimar un posible retorno de las actividades de explotación de datos y (4) la elaboración de hoja de ruta para avanzar al nivel de madurez esperado.



Figura 1-2

Componentes del modelo de explotación de datos para las entidades públicas

Madurez o preparación en explotación de datos	Estima el nivel de madurez en el que se encuentra una entidad para aprovechar y explotar los datos, al igual que el nivel de madurez al que la entidad aspira llegar de acuerdo con sus misionalidad.
Valor potencial en explotación de datos	Estima el valor potencial que tiene cada entidad para generar valor público a través de la implementación de procesos de explotación de datos, analítica de datos y <i>big data</i> .
Necesidades de inversión requerida en explotación de datos	Brinda recomendaciones para estructurar las necesidades de inversión en recursos humanos y tecnológicos de <i>big data</i> y estimar un posible retorno esperado.
Elaboración de hoja de ruta	Brinda recomendaciones para la construcción de una hoja de ruta que le permita a la entidad alcanzar su nivel de madurez esperado.



Los componentes de **madurez de explotación de datos** y **el de valor potencial** están acompañados del índice de madurez de *big data* y el índice de valor potencial, respectivamente (anexo 4: Metodologías de cálculo del modelo), que permiten a las entidades contar con una medida de referencia para indicar su progreso en el nivel de madurez para la explotación de datos, y para estimar el valor público que generaría la entidad a partir de la explotación de datos.

1.2 ◉

Hitos del modelo de explotación de datos y su interacción

El modelo de explotación para las entidades públicas de Colombia está soportado principalmente en un modelo de madurez que traza el puente entre el estado actual de la entidad y el estado que se desea alcanzar en relación con su nivel de preparación en explotación de datos. Sobre esto se abordará con más detalle en la próxima sección.

El modelo de explotación de datos, en general, permite a las entidades alcanzar los cuatro hitos referidos en la **figura 1-3**.

Figura 1-3.
Hitos del Modelo de Explotación de datos



Fuente: DNP, DDD



La interacción de los hitos se define en la **figura 1-4**, en donde las entidades identifican, a partir de un ejercicio de autodiagnóstico, tanto el nivel de madurez actual en explotación de datos como el que aspiran alcanzar. El modelo de explotación de datos, a su vez, proporciona a las entidades valores indicativos sobre el valor potencial de la explotación de datos, y recomendaciones para estructurar la inversión en tecnología y en recursos humanos y estimar el retorno esperado de la inversión.

Para llegar al nivel de madurez esperado, el modelo de explotación de datos brinda a las entidades recomendaciones específicas para la construcción de una hoja de ruta que incluya las acciones más estratégicas. El modelo de explotación de datos se operativiza a través un formulario autodiagnóstico que diligencian las entidades públicas (**figura 1-4**).

Figura 1-4.
Interacción hitos del Modelo de Explotación de datos



Fuente: DNP, DDD



1.3 o FORMULARIO DE AUTODIAGNÓSTICO

El formulario de autodiagnóstico acompaña el modelo de implementación de explotación de datos, dado que instrumentaliza en preguntas y opciones de respuesta, los componentes del modelo, e incluye preguntas para caracterizar a la entidad según su tamaño, sector y presupuesto. El formulario de autoevaluación (anexo 1) está compuesto en 6 dimensiones que corresponden a las contempladas del modelo de madurez de explotación de datos. Las entidades disponen de una *Guía de diligenciamiento del formulario* (anexo 1.1), en el que se plantean recomendaciones que deben tener en cuenta en el proceso del diligenciamiento. Lo anterior busca que el resultado de autodiagnóstico de las entidades responda a la realidad de la entidad.

El formulario se diseñó como una herramienta de autoevaluación neutra y homogénea, coherente con el marco conceptual del modelo de explotación de datos, así permite hacer una comparación objetiva entre entidades. Con esta premisa, se recomienda el diligenciamiento completo de la encuesta para evitar que se generen sesgos en el cálculo del puntaje.

+

2

ECOSISTEMA DE EXPLOTACIÓN DE DATOS

.....

-

El modelo de explotación de datos integra en el formulario autodiagnóstico el concepto de ecosistema de explotación de datos, en el que se describe una interacción constante entre los actores del sector público, del sector privado y de la academia para fomentar la explotación de datos en Colombia y promover la innovación de bienes y servicios. Para este modelo se considera que el fortalecimiento de capacidades de las entidades no se da en forma aislada, al igual que la interacción de iniciativas y estrategias favorece el desarrollo de los niveles de madurez en explotación de datos de las entidades públicas.

El ecosistema de explotación de datos refiere un entorno de colaboración con procesos estandarizados, aplicados por personas de diferentes entidades públicas y privadas a través de las tecnologías disponibles para todos los actores, entidades y organizaciones que lo conforman, con el fin de mantener el intercambio de datos, métodos, técnicas y resultados de la explotación de los datos, generando un fluido y vigoroso intercambio de información y conocimiento a partir de la explotación colectiva de los datos.





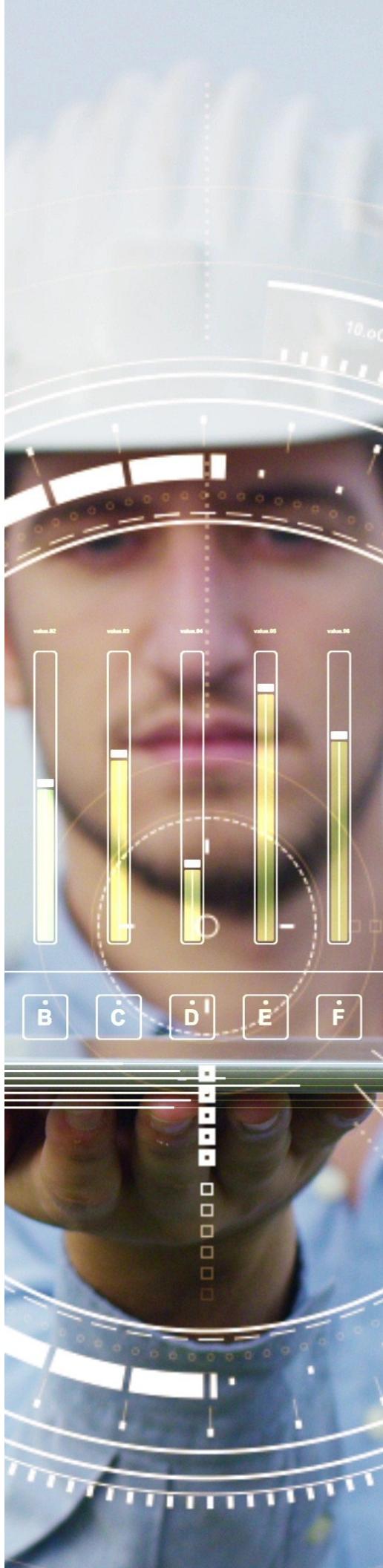
La estandarización de los procesos y los procedimientos para la explotación de datos debe considerar los actores que participan en este ecosistema, pues estos se instrumentalizan a través de metodologías documentadas, repositorios abiertos y licencias para el uso público de datos y tecnologías, destinados a la cocreación de productos de información y conocimiento que deben, de igual forma, estar protegidos por licencias de uso que faciliten su reutilización.

De otra parte, el uso de tecnologías libres facilita que estén disponibles para todos los actores que integran el ecosistema de explotación de datos en el país; además, deja abierta la posibilidad de la innovación al ser posible ajustar y modificar herramientas con ese tipo de licencias. Sin embargo, quizás lo más importante es que posibilita la replicación de cualquier ejercicio de analítica de datos, lo que permite avanzar en los frentes de Gobierno Abierto (Open Government). Las principales tecnologías de explotación de datos, como ya se explicó, son de uso libre y código abierto.

Figura 2-1.
Esquema de ecosistema de explotación de datos



Fuente: DNP, DDD



El ecosistema de explotación de datos se consolidará en la medida en que maduren las capacidades de las entidades públicas, los actores privados y la academia y se fortalezca la interacción entre los diferentes actores que lo componen (**figura 2-1**). A pesar de que en Colombia el modelo de ecosistema no está definido ni constituido en ninguna instancia o comité, el concepto se tuvo en cuenta en el modelo de explotación de datos, dada la apuesta del Gobierno nacional por promoverla en las entidades públicas aunada el trabajo colaborativo que se hace con el sector privado, la academia y las entidades de cooperación internacional.

+

3

DESCRIPCIÓN DE LOS COMPONENTES DEL MODELO DE EXPLOTACIÓN DE DATOS

.....

-



En esta sección se describe cada uno de los componentes del modelo de explotación de datos y de los indicadores que acompañan el modelo de madurez de explotación de datos y el valor potencial. La metodología de cálculo de cada componente se detalla en los anexos 4: Metodología de cálculo del modelo y 5: Ejemplo de cálculo.





3.1 ⚡

MODELO DE MADUREZ DE EXPLOTACIÓN DE DATOS

Por lo general la madurez describe un estado cabal, perfecto o listo en un dominio específico (Abdullah, Al-Sai, & Husin, 2019); así, para alcanzar un estado de madurez deseado es necesario avanzar en un camino de transformación desde una etapa inicial a una etapa objetivo (Zdonek & Krol, 2020). Los modelos de madurez son instrumentos aceptados para guiar el desarrollo y la transformación organizacional mediante la documentación sistemática para la adopción de actividades y mejores prácticas (Abdullah, Al-Sai, & Husin, 2019).

Los modelos de madurez, por lo común, presentan una progresión esperada del desarrollo en un tema, a través de niveles de madurez, o pasos que la entidad evaluada alcanza mediante la demostración de capacidades o el logro de objetivos. Los modelos de madurez permiten la comparación o *benchmarking* frente a otras entidades u organizaciones parecidas, y orientan la formulación de estrategias para aumentar la eficiencia en la utilización de recursos destinados a mejorar estas capacidades y alcanzar el éxito perfeccionado en el dominio (De Bruin, Freeze, & Kulkarni, 2005).

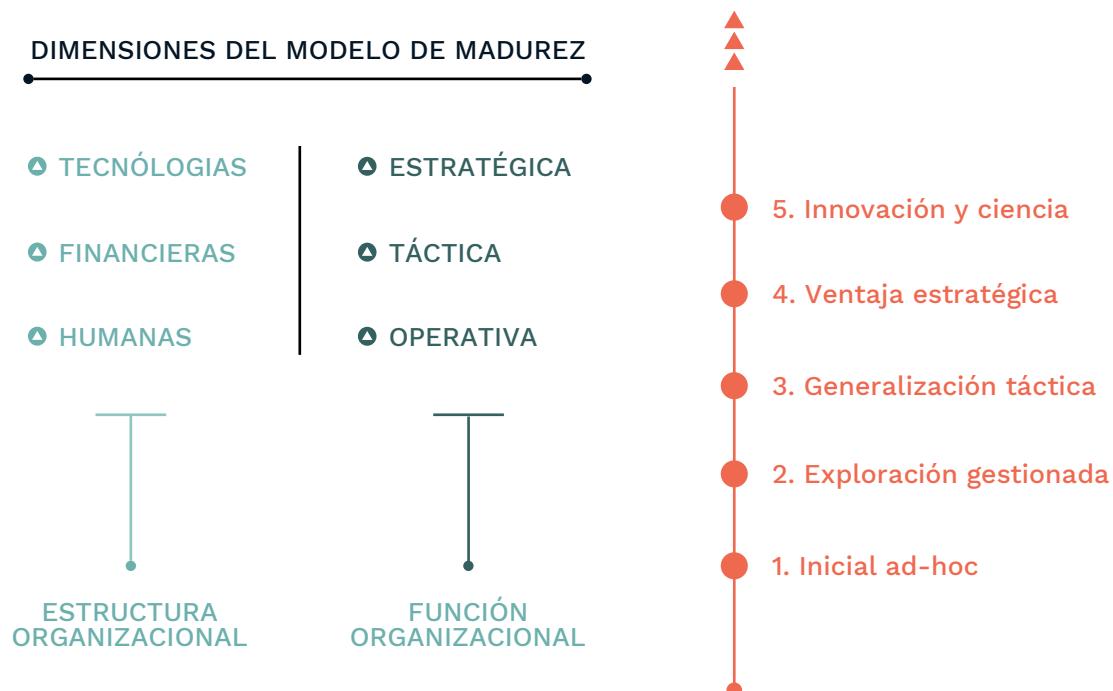


3.1.1 ◉

ARQUITECTURA DEL MODELO DE MADUREZ

El modelo de madurez en explotación de datos consta de seis dimensiones y de cinco niveles de madurez (**figura 3-1**), que se instrumentalizan en secciones distintas en el formulario diseñado para ser aplicado a las entidades públicas del país (anexo 1.1: Formulario de explotación de datos). Como ya se mencionó, el modelo de madurez de explotación de datos permite a las entidades identificar el nivel actual de desarrollo de sus capacidades para la explotación de datos, así como el nivel esperado que deberían alcanzar de acuerdo con su misionalidad y objetivos organizacionales.

Figura 3-1.
Arquitectura del Modelo de Madurez de explotación de datos



Fuente: DNP, DDD



DIMENSIONES DEL MODELO

Las dimensiones del modelo de madurez de explotación de datos se definieron sobre el entendimiento de que el *big data* es un desarrollo sociotécnico, en el cual los recursos tangibles y materiales se convierten en capital para el aprovechamiento de datos, una vez se alinean con la misión, visión y objetivos organizacionales de la entidad. La articulación integral alrededor de roles y procesos definen la función organizacional del modelo, mientras que el aprovechamiento de los recursos define la estructura organizacional. Las dimensiones del modelo se diseñaron con base en un esquema general de arquitectura empresarial, de modo que responden a la realidad de cualquier entidad pública o privada.

En Figura 3-1 se presenta un esquema de las dimensiones del modelo de madurez. Las dimensiones se organizan en dos grupos, correspondientes a la *estructura organizacional* (dimensiones de recurso humano, tecnológico y financiero) y a la *función organizacional* (dimensión operativa, táctica y estratégica).





DIMENSIONES DE LA ESTRUCTURA ORGANIZACIONAL

1. Recurso y capital humano

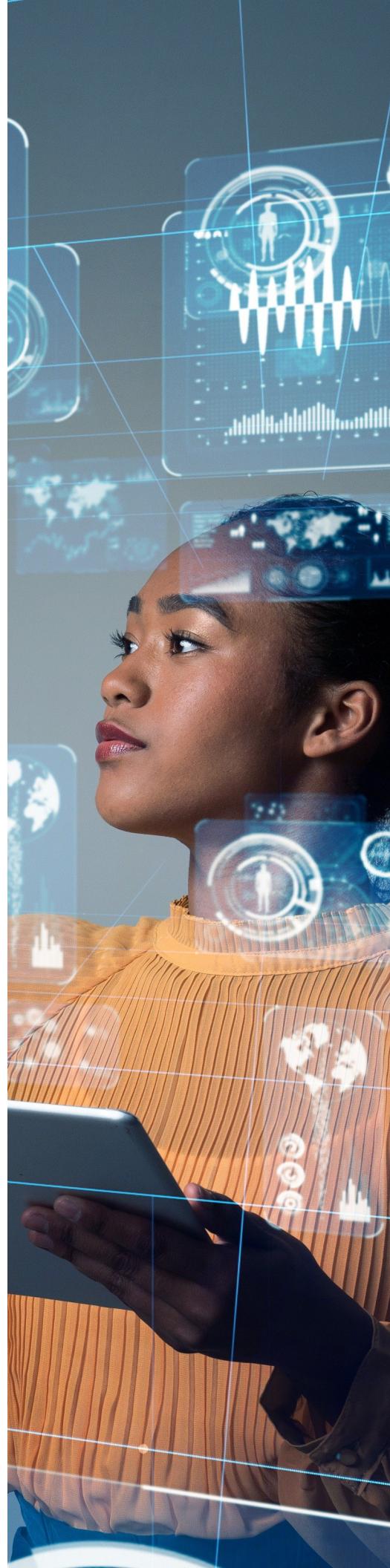
La explotación de *big data* requiere de capital humano con alto grado de capacitación y familiarizado con técnicas avanzadas de análisis y manejo de herramientas informáticas especializadas. No hay una única disciplina que abarque en su totalidad las competencias requeridas para el desarrollo e implementación de proyectos de analítica de datos, por lo que los equipos encargados del análisis de datos deben estar constituidos por personas que tengan diferentes habilidades y conocimientos en analítica de datos, programación, estadística y conocimiento de los objetivos misionales de la entidad.

En ese marco, se identifican cinco perfiles pertinentes para la configuración de equipos de analítica de datos, que permiten soportar las necesidades derivadas del diseño, desarrollo e implementación de proyectos de analítica. Los científicos de datos, el analista de datos, ingeniero de machine learning, ingeniero de software e ingeniero de datos. Estos perfiles se tomaron con base en el documento del DNP (2020): *Aprovechamiento de datos para la toma de decisiones el sector público*. A continuación, se describe cada uno de ellos.



Los equipos de explotación de datos y *big data* pueden tener un mayor o menor grado de madurez de acuerdo con la definición y operación de procesos de trabajo conjunto para obtener productos de complejidad creciente. En los siguientes apartados se describe una serie de factores críticos cuya omisión puede ser la causa de proyectos fallidos en *big data* (Saltz & Shamshurin, 2016):

- Desarrollo de un trabajo grupal, para la coordinación, análisis y priorización de requerimientos y hacer posibles las soluciones analógicas considerando el ciclo de vida de los datos de forma iterativa.
- La comunicación efectiva, dentro y fuera del equipo, con los directivos y los tomadores de decisiones, soportado también en un proceso y programas de rotación.
- Conocimientos en las implicaciones, éticas y legales del manejo de datos, la privacidad y protección de datos personales y de la propiedad intelectual.





2. Recurso y capital tecnológico

El desarrollo una arquitectura tecnológica de almacenamiento, procesamiento y análisis de datos diversos y en grandes volúmenes, es fundamental para la explotación de datos y *big data*. El componente tecnológico debe ser escalable y estar alineado con las necesidades actuales y futuras de las entidades. El acceso a los recursos tecnológicos es más importante que su adquisición, pues si bien hoy las dinámicas de computación de altos volúmenes de datos o de implementación de complejos algoritmos de computación pueden hacerse en casa —*in house* u *on premise*— también pueden hacerse a través de servicios en la red pública de internet o como comercialmente se denomina “en la nube” —*in cloud*—.



A continuación, se describen los esquemas de infraestructura empleados en mayor medida por las organizaciones:

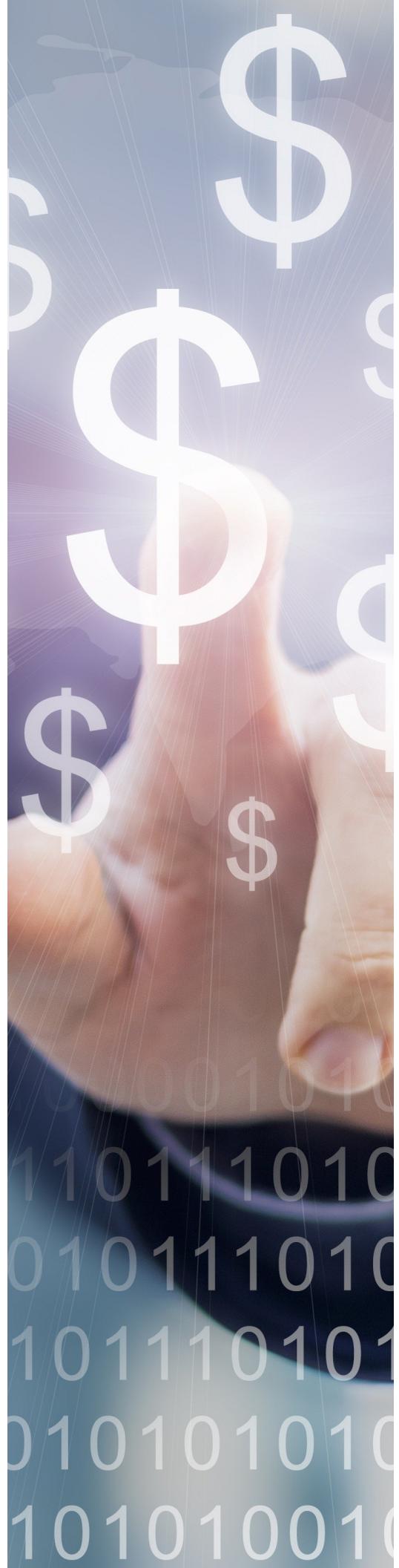
- **On-premise:** los servidores y software están ubicados en las instalaciones de la entidad y usualmente están conectadas a la red local de la empresa. La entidad tiene que asumir la gestión de la infraestructura incluyendo su seguridad (DNP, 2020).
- **Nube pública o privada:** los servidores que utiliza la entidad están disponibles en los servidores remotos de la propia entidad o de terceros y se accede a través de Internet. Usualmente la administración de la infraestructura se encuentra a cargo del proveedor del servicio en la nube y su modalidad de uso es escalable. La computación en la nube y el *big data* están fuertemente conectados, pues proporciona la infraestructura subyacente que puede servir como plataforma para abordar problemas relacionados con *big data*.⁴ Las modalidades de servicios en la nube son de tres tipos: infraestructura como servicio (IaaS), Plataforma como servicio (PaaS) y *Software* como servicio (SaaS)⁵. Actualmente a través de la nube es posible acceder a servicios de *big data* en la nube denominados *Big Data as a Service* que brindan infraestructura, *software* y servicios para *big data*, como limpieza de datos, aplicación de modelos estadísticos y matemáticos, lo mismo que a la ejecución y el mantenimiento de arquitecturas.
- **Híbridos:** este esquema lo integran la infraestructura en nube y la propia. En este caso, la solución tecnológica está disponible en el servidor de un proveedor y los datos que tienen un carácter más privado permanecen alojados en servidores de la entidad. (DNP, 2020).

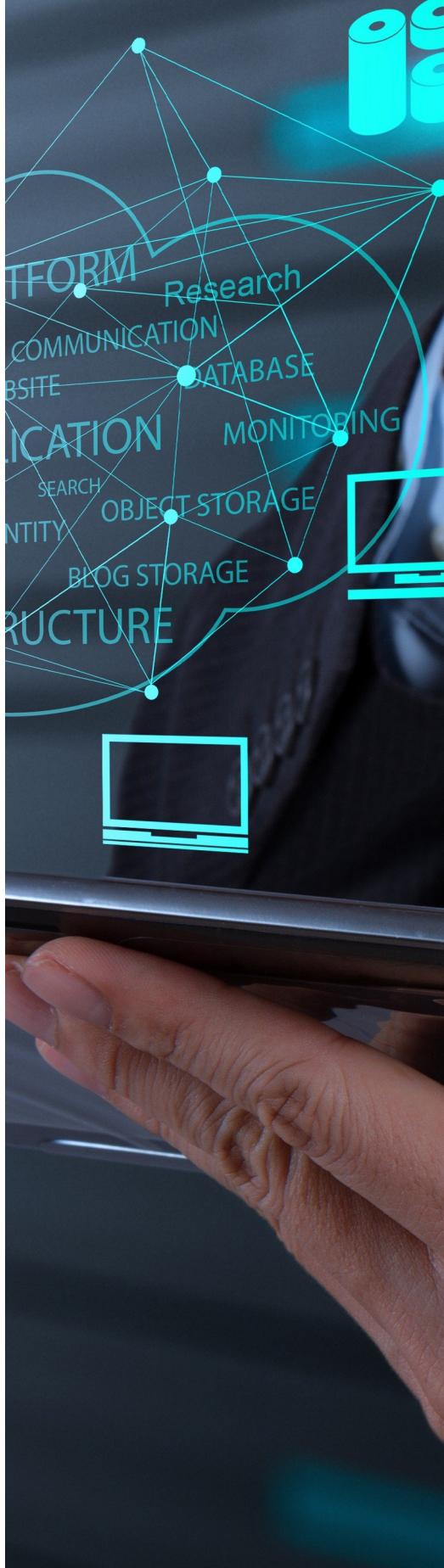


3. Recurso y capital financiero

La disposición de talento humano y recursos tecnológicos dependen de la asignación de recursos financieros. En este caso los recursos financieros representan la disposición y capacidad de la entidad para orientar recursos monetarios al despliegue de una estrategia de explotación de datos. El monto de inversión dependerá de la planeación de la entidad con base en sus necesidades de explotación de datos y con base en los proyectos de explotación de datos que se definan para ejecutar.

Un elemento importante en la priorización de inversiones consiste en reconocer los datos como activo intangible para la entidad y la contribución que tienen para la generación de valor público. Por ello, se hace necesario elaborar el plan de inversión en explotación de datos, identificar en qué medida la inversión en capital humano y en infraestructura tecnológica satisface la demanda requerida y contribuye a la generación de beneficios para la entidad.





DIMENSIONES DE LA FUNCIÓN ORGANIZACIONAL

4. Dimensión estratégica - Gobernanza de datos

La gobernanza de datos es el ejercicio de autoridad, control y toma compartida de decisiones - planeación, monitoreo en la gestión de activos de datos - (Ladley, 2012). Implica alinear a las personas y la tecnología para prevenir problemas costosos relacionados con los datos y asegurar que los datos de una entidad se manejen y usen de manera más eficiente (Ghavami, 2015). La gobernanza de datos involucra múltiples actores e interesados en una gestión progresiva y de largo plazo que controla de forma integral los activos de datos. Los procesos y subprocesos de la gobernanza de datos se sintetizan en la **tabla 3-1**.

Tabla 3-1.

Componentes y subcomponentes de la gobernanza de datos

Dominio	Subdominio
Roles, estructuras y políticas	Cultura y confianza
	Participación
	Políticas y estándares
	Inteligencia de negocio
	Administración de datos
Calidad de datos	Métricas de calidad
	Dimensiones de la calidad de datos
Gestión de metadatos	Metadatos de la entidad
	Repositorio de metadatos
Gestión de datos maestros	Administración de datos de referencia
	Integración de datos
	Almacenamientos de datos
Arquitectura de datos	Definición de la arquitectura de datos
	Catálogo de componentes de datos
Tecnología	Infraestructura
	Analítica
Seguridad y privacidad	Derechos de acceso a los datos
	Riesgos en el manejo de datos
	Cumplimiento de datos
Métricas de seguimiento y monitoreo	Esquemas de seguimiento
	Cuantificación de la creación de valor
	Identificación de beneficios

Fuente: DNP, DDD con base en Mertilos, 2015.



La gobernanza de datos se justifica cuando los datos, como activos institucionales, son objeto de múltiples usos por diversos cuerpos y personas en las entidades, en las que se asignan roles, responsabilidades, rendición de cuentas y derechos de decisión para asegurar que se gestionen adecuadamente, a fin de tener su acceso preciso y libre de riesgos (Ladley, 2012). Como muchos de los procesos en la gestión de información, la gobernanza de datos evoluciona rápidamente con la ampliación de sus fronteras, y, aunque parece trasladarse con otros procesos, en realidad interactúa con ellos; tal es el caso de la gestión de datos y la gestión del cambio.





5. Dimensión táctica

La capacidad táctica permite implementar de manera correcta el marco de políticas y normativas estipuladas en la gobernanza de datos y las orienta hacia procesos operativos en el ciclo de vida de los datos. En ocasiones se confunde la gobernanza de datos con la gestión del ciclo de vida de los datos; sin embargo, vale la pena resaltar que como esta última tiene que ver directamente con la administración del ciclo de vida de los datos, se orienta a que sean de calidad, precisos, y accesibles. El despliegue de la estrategia táctica para el uso y explotación de datos requiere del desarrollo de procesos de gestión de proyectos y datos, y de la participación del nivel directivo medio.

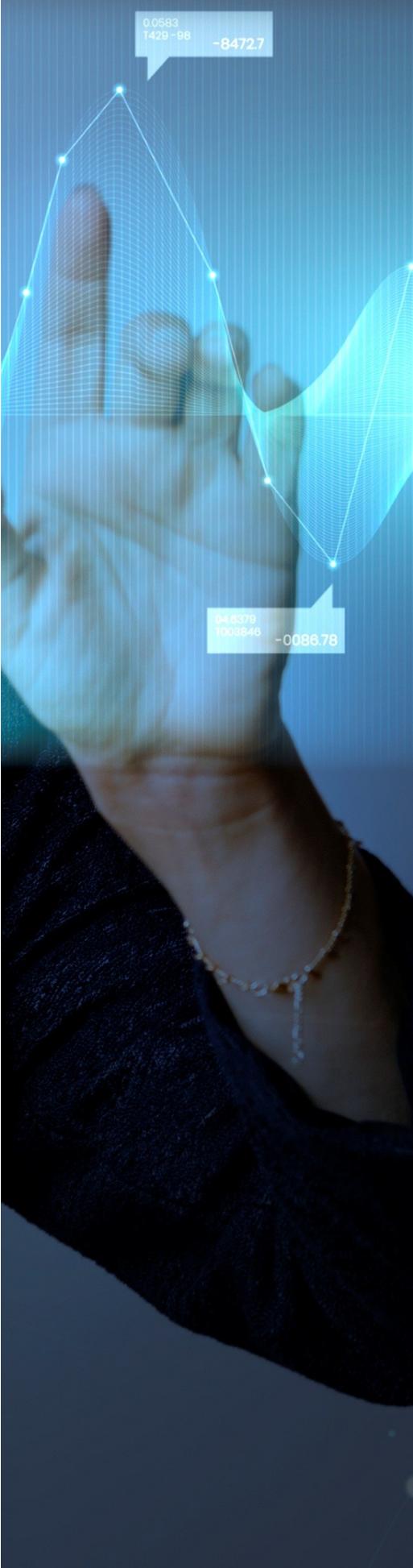


6. Dimensión operativa

- procesos y procesamientos en el ciclo de vida de los datos -

Se define a partir de procedimientos del ciclo de vida de los datos, entendidos como una sucesión de pasos sistemáticos que permiten un análisis inteligente.⁶ El carácter sistemático de la dimensión operativa apunta a que los resultados del análisis puedan ser repetibles y que sea eficiente el uso de los recursos de análisis. En esta dimensión se operativiza todo el ciclo de los datos en relación con su captura y generación, el almacenamiento, el procesamiento, la visualización de los datos, y la comunicación de la información para la toma de decisiones.





3.1.2 ⚠

NIVELES DE MADUREZ

El modelo de explotación de datos se compone de cinco niveles que se definen como *estados discretos ordinales* en los que uno es mayor que otro. Los niveles de madurez describen tanto el desarrollo de las seis dimensiones del modelo analizadas en la sección anterior, como el nivel de madurez general de la entidad. El enfoque del nivel de madurez de este modelo comprende dos hitos: (1) el nivel de madurez alcanzado y (2) el nivel de madurez esperado por la entidad. Estos hitos permiten crear escenarios de comparación entre un *estado actual* - representado por el nivel de madurez alcanzado - y un estado *objetivo/meta* - representado por el nivel de madurez esperado-, con el propósito de trazar rutas de acción para avanzar del nivel de madurez observado al nivel de madurez esperado, no solo en lo general sino también por dimensión.

Los niveles de madurez definen el estado temporal de una entidad y permiten evaluar su evolución en el tiempo. Las seis dimensiones del modelo cruzan de manera transversal con cada nivel del modelo. así, para cada dimensión se tiene una descripción de los niveles de madurez que la caracterizan, las cuales se pueden consultar en el anexo 2: Matriz de descripción de niveles de madurez por dimensión.



• • • • •

MATRIZ DE DESCRIPCIÓN DE NIVELES DE MADUREZ POR DIMENSIÓN

La matriz describe las características generales de todas las dimensiones y niveles del modelo propuesto. La matriz es de tamaño de 6 x 5 (seis dimensiones y cinco niveles) y permite comprender para cada una de las dimensiones, que significa estar en un nivel de madurez determinado. A continuación, se muestra un breve ejemplo de lo que se encuentra en el anexo 2: Matriz de descripción de niveles por dimensión (tabla 3-2).



Tabla 3-2.

Ejemplo de la descripción de los niveles 1,2 y 3, para las dimensiones de recursos humano y financiero

Dimensión	Nivel de madurez 1	Nivel de madurez 2	Nivel de madurez 3
Recurso humano	<p>El uso de datos se realiza para tareas puntuales de reporte en cada dependencia. No existe un equipo dedicado al uso y explotación de datos. Tampoco hay roles definidos para la gestión de datos.</p>	<p>Existen equipos en dependencias específicas desarrollando proyectos piloto y pruebas de concepto en explotación de datos y <i>big data</i>. Para el desarrollo de los proyectos se definen roles básicos de gestión de datos, principalmente relacionados con acceso a datos y preprocesamiento.</p>	<p>Existe una única dependencia que cuenta con un equipo dedicado a la explotación de datos que apoya a otras dependencias de la entidad a manera de centro de excelencia en proceso de conformación. Se han definido de manera formal algunos roles para la gestión de datos, principalmente relacionados con los procesos de operaciones.</p>
Recurso financiero	<p>La entidad no ejecuta inversión para fortalecer las capacidades de explotación de datos y <i>big data</i> para la vigencia en curso, ni se ha destinado un presupuesto para actividades de analítica o explotación de datos durante los últimos cinco años.</p>	<p>Se ha establecido un esquema de financiación bajo, inferior al 2 %, del total del presupuesto de la entidad. Se destina presupuesto para iniciativas, pruebas de concepto y proyectos de explotación de datos, pero la destinación del presupuesto no responde a un ejercicio de planeación organizacional.</p>	<p>En el último año, se ha incluido en el presupuesto una partida para atender las necesidades asociadas al uso de datos dentro de la entidad, en un esquema de financiación medio, que representa entre el 2 % y el 5 %, del presupuesto total de la entidad.</p>

Fuente: DNP, DDD.



DESCRIPCIÓN DE LOS NIVELES DE MADUREZ

- **Nivel de madurez 1. Inicial de concientización y procesos ad hoc**

Se identifica el potencial del uso y disponibilidad de datos masivos para impulsar la transformación digital de la entidad y alcanzar beneficios como institución. Más allá de esto, no existen iniciativas formales o de gestión institucional al respecto. En general, la capacidad analítica de la entidad es básica y se centra en la generación de reportes. Por iniciativas aisladas de algunos funcionarios o contratistas pueden desarrollarse algunos ejercicios de analítica avanzada, que al ser *ad hoc* no son sistemáticos y, por consiguiente, los procesos subyacentes no están controlados y son impredecibles. De esta manera, el recurso humano y los recursos informáticos no están dispuestos ni aliñeados con una iniciativa formal de *big data*. Pueden existir recursos que cumplan con los requerimientos de *big data* —científicos de datos, entre otros perfiles—, pero no se han capitalizado para este fin. No hay destinación de recursos financieros o estos son insuficientes para avanzar en la explotación de datos.

- **Nivel de madurez 2. Exploración gestionada en pruebas de concepto y piloto**

Por la convergencia de líderes tácticos, se emprenden, en algunas dependencias de la entidad, iniciativas para la formulación e implementación progresiva de explotación de datos y *big data*. Cuando dichos grupos enfrentan el reto del procesamiento de datos masivos y complejos en pruebas de concepto o proyectos piloto, se hacen evidentes varias necesidades.



Se inicia el reclutamiento de recurso humano capacitado de personas con formación y/o experiencia, o se considera recurrir al *outsourcing*. Aunque esto no es suficiente para garantizar que se desarrolla el capital humano en la entidad, permite dar respuesta rápida a los retos de esta segunda fase.

De igual forma, los recursos tecnológicos, financieros y físicos son contingentes. Se hace evidente la necesidad, también reactiva, de formular, ejecutar y documentar los diferentes procesos de gestión y operación de manera sistemática, ya que son los precursores de la automatización.

Los líderes tácticos deben tener el conocimiento técnico, o ser apoyados por quienes lo posean, para la gestión efectiva y eficiente de los recursos, lo mismo que para garantizar el éxito de los pilotos. El objetivo de los proyectos es explorar e identificar oportunidades de mejora y necesidades para impulsar la explotación de datos en la entidad. Por tratarse de proyectos de menor escala los riesgos pueden mitigarse con facilidad.

- **Nivel de madurez 3. Generalización táctica y optimización para producción en masa**

Por los aprendizajes de los proyectos piloto o pruebas de concepto de la fase anterior, se logra el apoyo generalizado del nivel táctico de la entidad, por lo que se emprende la definición de procesos estandarizados y su documentación regular, como base para la operación optimizada para la explotación de *big data*. La consecuencia puede ser el establecimiento de una dependencia, o área específica, con recursos humanos, financieros y físicos continuos, y la inversión tecnológica progresiva que incluya soluciones abiertas,



libres o no, las cuales pueden ajustarse según las necesidades particulares de la entidad. Al respecto, la identificación de tales necesidades debe ser también sistemática y cuantificable en de toda la entidad, por lo que se requiere implementar procesos de gobernanza y gestión de datos como base de la operación consistente, para el desarrollo de productos y servicios de utilidad para la toma de decisiones de las diversas dependencias.

Es posible que en un momento sea insuficiente la gestión centralizada en un único grupo o dependencia de la entidad, dadas todas las necesidades de productos, servicios de información y conocimiento resultantes de la explotación de datos y *big data*, por lo que deben promoverse procesos de capacitación para el desarrollo incremental de competencias en explotación de datos y *big data* para perfiles priorizados en la entidad. De igual manera, es necesario establecer el acceso e interacción distribuidos y remotos a los recursos tecnológicos para los usuarios de la entidad.

○ Nivel de madurez 4. Ventaja estratégica

El desempeño sostenido y consistente de los procesos analíticos de *big data* en de la entidad, desarrollado en el nivel anterior, provee el soporte suficiente para establecer confianza y obtener el apoyo de la dirección de la entidad que pueden usar sus resultados de productos y servicios en la toma de decisiones. Así, son más estables la gestión y los procesos operativos, su control y cuantificación, al igual que la continuidad de los recursos. Como el seguimiento y evaluación con retroalimentación son muy dinámicos, se requiere la incorporación de metodologías ágiles que den respuesta rápida a los hallazgos.



La visión estratégica permite identificar oportunidades de nuevos campos de aplicación de la explotación de datos. También cuando la explotación de datos y el *big data* llegan a posicionarse en la entidad, es posible abrir horizontes para el intercambio de experiencias y de identificación conjunta de casos de uso con entidades de este o de otros sectores. De esta forma se hace posible establecer alianzas para los procesos analíticos, el uso de tecnologías y de recursos financieros. La confianza puede ser estabilizada por medio de acuerdos interinstitucionales que permitan el procesamiento distribuido.

○ Nivel de madurez 5. Innovación y ciencia

Una vez se ha incorporado la explotación de datos y *big data* a través de toda la organización, integrando los niveles estratégico, táctico y operativo, gracias a la continuidad y capitalización de los recursos, es posible ir más allá en el uso de infraestructura y métodos para la producción de resultados con analítica y ciencia de datos. Como resultado puede ser posible la transformación institucional o su fortalecimiento con otras entidades para la conformación de una red de explotación de datos y *big data*. De estos procesos sinérgicos posiblemente surjan nuevos métodos y tecnologías, o el uso innovador de otros ya establecidos, así como nuevos perfiles especializados de recurso humano capitalizado para el *big data*.



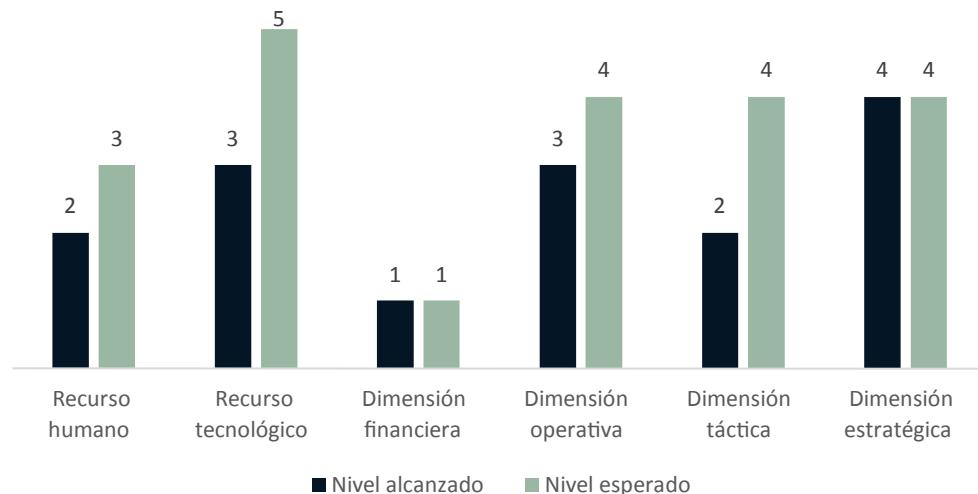
NIVEL DE MADUREZ ESPERADO POR LAS ENTIDADES

El modelo de madurez además de brindar a las entidades un diagnóstico actual de sus capacidades en explotación de datos, también permite que identifiquen el grado de madurez que se plantea como objetivo alcanzar a escala general y por cada dimensión. Lo anterior se determina a través de una pregunta incorporada en el formulario de autodiligenciamiento que indaga sobre el nivel de madurez que la entidad quiere alcanzar.

En la **figura 3-2**, se presentan el nivel alcanzado y esperado de una entidad ficticia para exemplificar lo descrito.

Figura 3-2.

Ejemplo del nivel de madurez alcanzado y esperado por una entidad ficticia, para cada dimensión del modelo de madurez



Fuente: DNP, DDD



La entidad debe seleccionar el nivel al que debería aspirar teniendo en cuenta: (1) su misionalidad, (2) sus objetivos organizacionales y (3) su caracterización en las dimensiones y niveles del modelo. Con base en esos tres criterios se espera que la entidad seleccione el nivel de madurez que mejor se ajuste a su realidad institucional. Lo anterior permite aclarar dos puntos fundamentales: (1) no todas las entidades deberían aspirar al nivel 5 y (2) las entidades no necesariamente tienen que aspirar al mismo nivel en las seis dimensiones.

Los niveles de madurez fueron diseñados de modo que no existiera uno mejor que otro, en términos de importancia o reconocimiento. En este modelo no se propone que todas las entidades llegan al nivel de madurez más alto, porque se desconocerían sus especificidades. Existen entidades que, de acuerdo con su misionalidad y a sus objetivos organizacionales, posiblemente no aspiren a los niveles más altos de madurez, lo que significa la generalización táctica y procesos de innovación y ciencia -niveles 4 y 5, respectivamente-. Sin embargo, para las entidades del orden nacional y en especial las que son cabeza de sector, dada su misionalidad de encargadas de coordinar las áreas estratégicas de gobierno, se espera que tengan unas capacidades equivalentes al del nivel 3, es decir, que al menos garanticen la producción y uso sistemático de datos, las capacidades para intercambiarlos y disponerlos en formatos estándar y abiertos.

Frente al segundo punto, los niveles de madurez esperados no tienen por qué ser los mismos en las seis dimensiones del modelo. Pueden darse casos de entidades que identifiquen una mayor importancia de algunas dimensiones y, por lo tanto, esperen un nivel más alto de madurez en ciertas dimensiones frente a otras. Las decisiones asociadas al nivel de madurez esperado deberían resultar del proceso mancomunado de autodiagnóstico entre diversos funcionarios, o grupos de trabajo, dentro de la institución. En dicho proceso se debe tener en cuenta la descripción de la intersección entre los niveles de madurez y las 6 dimensiones del modelo (anexo 2: Matriz de descripción de niveles de madurez por dimensión) de modo que la selección de nivel esperado se tome con el mayor grado de conocimiento y de comprensión del modelo.



3.2 ◉

VALOR POTENCIAL QUE PODRÍA GENERAR LA EXPLOTACIÓN DE DATOS

La aplicación de datos en el Gobierno tiene un potencial casi ilimitado para brindar servicios públicos más eficientes, efectivos y confiables (OCDE, 2019). Un punto de partida fundamental para aproximarse a la medición del valor de los datos es reconocerlos como un *activo⁷ intangible*, lo cual puede resultar valioso a la hora de establecer una cultura impulsada por datos en todo el sector público, porque plantea un desafío a los líderes de las entidades quienes deben comprender que el valor de los datos aumentará o disminuirá en relación con sus esfuerzos para administrarlos y utilizarlos (OECD, 2019b).

De acuerdo con la OCDE, la capacidad de las organizaciones para valorar los datos que poseen disminuye si no existe una metodología para verlos de manera equivalente como otros activos clave, por ejemplo, las personas y los recursos financieros (OCDE, 2019). Por lo expuesto, este modelo de explotación de datos plantea una estimación para que las entidades públicas reflexionen acerca de los beneficios de la explotación de datos y *big data* sobre sus actividades misionales.

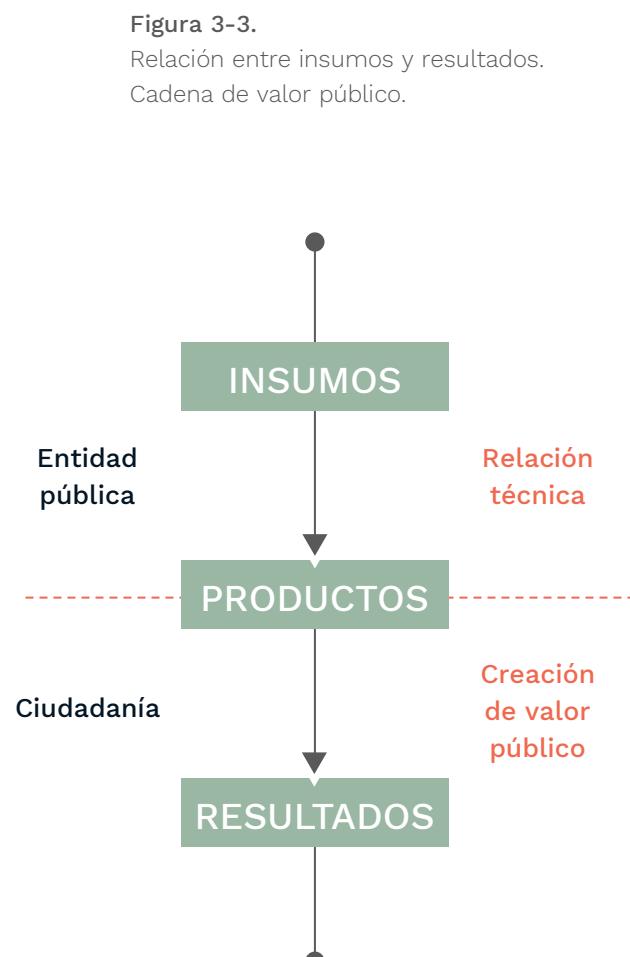


En el modelo de explotación de datos la estimación del valor potencial se aproxima al concepto de valor público, teniendo en cuenta que el propósito final de la gestión de los activos del sector público es generar beneficios sociales que sean percibidos por la ciudadanía (Conejero, 2014). De acuerdo con Moore, en el sector público el proceso de creación de valor está orientado al cambio social y no a resultados de rentabilidad económica, lo cual dificulta su formalización y conceptualización (Moore, 1995).

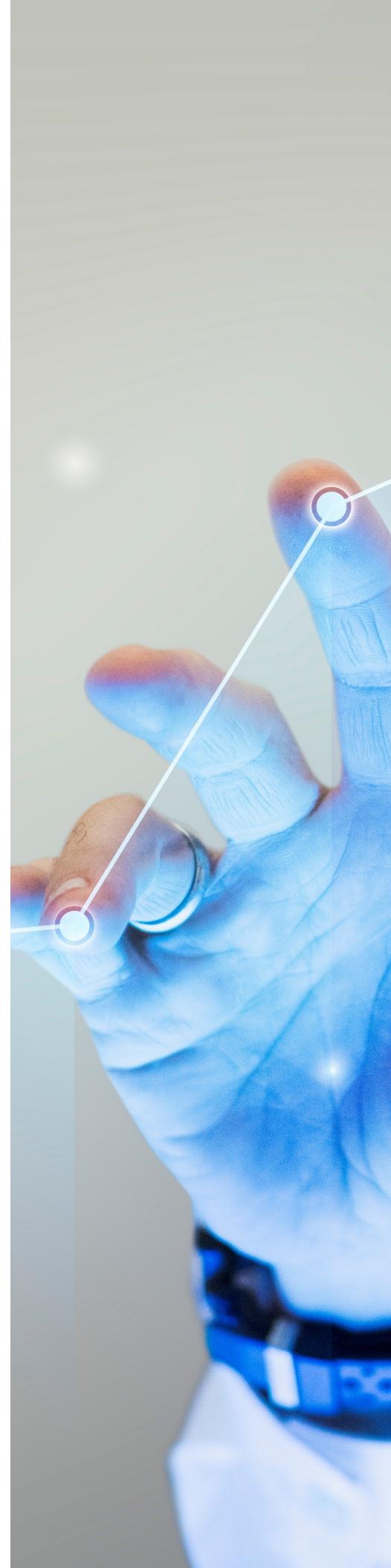
Para el caso del sector público en Colombia, el concepto de valor público se basa en la definición que plantea el Modelo Integrado de Planeación y Gestión (MIPG), que hace referencia a los resultados —observables y medibles— que un estado debe alcanzar para responder a las necesidades o demandas sociales, los cuales están asociados a los cambios producidos tanto por la acción gubernamental como por las actividades y productos entregados por cada entidad pública.

Para la generación de valor público las entidades orientan sus recursos organizacionales a forjar el máximo valor posible a los ciudadanos. Dentro de este marco, se surge el concepto de *cadena de valor público*, que es un modelo descriptivo (ASIP, 2014) en el cual se visibiliza el proceso por medio del cual las entidades del gobierno obtienen resultados para la creación de valor público, a partir de la producción de bienes y servicios.

En este caso, la aplicación de los datos para generar valor público se define a través de una relación entre los datos como *entrada* (OCDE, 2019), recurso humano, recurso tecnológico —insumo—, la generación de productos —bienes y servicios— y su efecto sobre la ciudadanía, comprendidos como resultados o *salidas* (figura 3-3).



Fuente: DNP, DDD, tomado de Sotelo Maciel (2014) “El modelo de la cadena de valor público en el análisis de las políticas”.



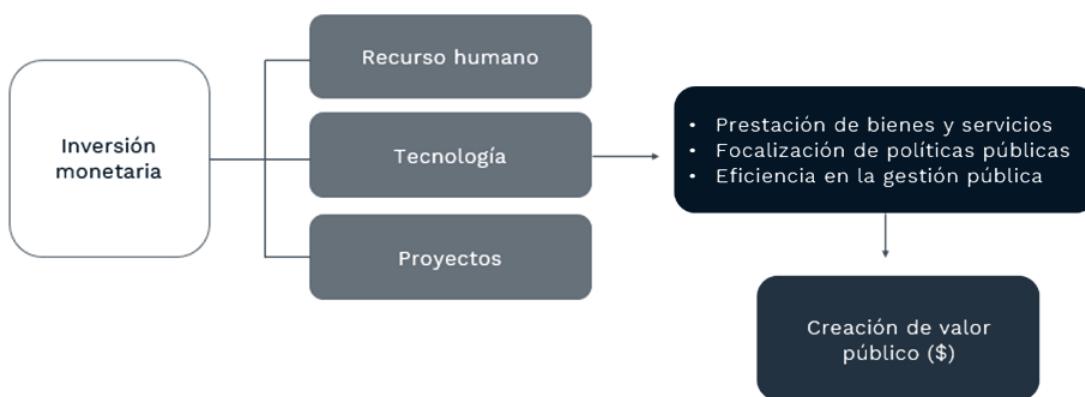
3.2.1 ◉

INSUMOS DE CREACIÓN DE VALOR A PARTIR DE LA EXPLOTACIÓN DE DATOS

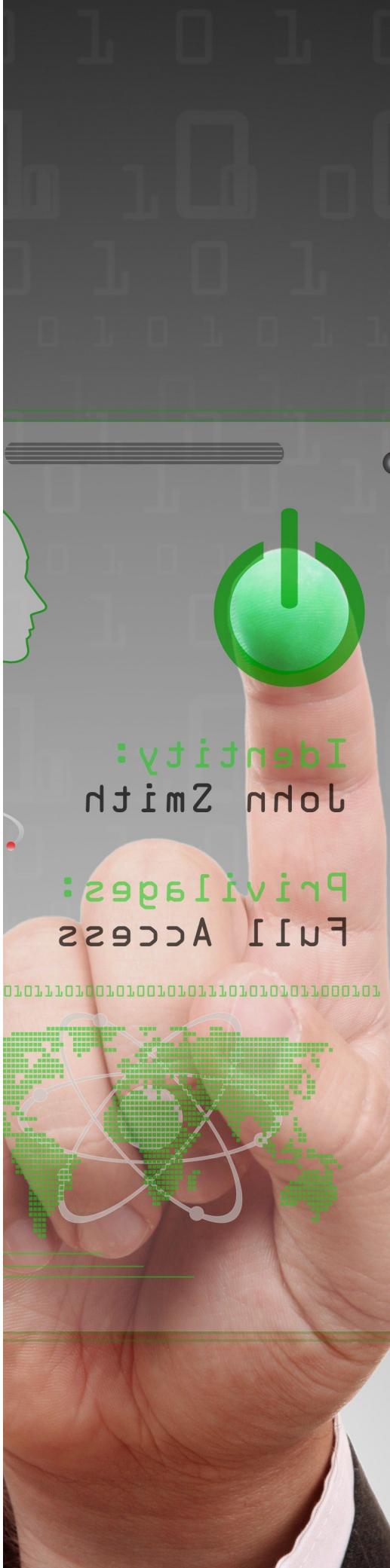
El modelo de explotación de datos identifica tres insumos de creación de valor público para materializar el valor de los datos (**figura 3-4**): (1) el fortalecimiento de recurso humano, (2) el desarrollo de recurso tecnológico y (3) el desarrollo de proyectos específicos de explotación de datos que responden a necesidades puntuales. La inversión en los insumos para la creación de valor basada en datos permite mejorar la prestación de bienes y servicios, optimizar la focalización de políticas públicas y mejorar la eficiencia en la gestión pública, lo que conlleva a la creación de valor público en las entidades.

Figura 3-4.

Generación de valor público a partir de las fuentes de creación de valor



Fuente: DNP, DDD



3.2.2 ◉

IDENTIFICACIÓN DEL VALOR POTENCIAL A PARTIR DE LA EXPLOTACIÓN DE DATOS EN LAS ENTIDADES PÚBLICAS

La aproximación que se planea en el modelo de explotación de datos parte de la base conceptual definida en el *Manual de Oslo* para medir los impactos sobre el valor social y económico de la innovación (OCDE / Eurostat, 2018). Desde un punto de vista conceptual, la *captura de valor derivado* de una innovación es incierta y solo puede evaluarse por completo cierto tiempo después de su implementación. Por ello, para estimar el valor que tiene la innovación se propone analizar por parte del sujeto expuesto al fenómeno -en este caso, las entidades y los funcionarios públicos de las entidades- sobre el fenómeno de la innovación -en este caso, la explotación de datos y *big data*-.

Los funcionarios públicos que diligencien el formulario del modelo de implementación de explotación de datos deben llevar a cabo un ejercicio retrospectivo y prospectivo que permita profundizar en qué medida las inversiones en actividades de explotación de datos y *big data* generan retribuciones para los grupos de interés de la entidad. Para ello, los funcionarios deben identificar hasta qué punto a la inversión en iniciativas o proyectos de explotación de datos y sus fuentes de creación de valor contribuye a crear valor público de la entidad, según su misionalidad.

3.2.3 ⚡

RELACIÓN DEL VALOR POTENCIAL Y EL NIVEL DE MADUREZ EN EXPLOTACIÓN DE DATOS

La percepción que tengan los funcionarios de las entidades públicas sobre el valor potencial que genera la adopción de modelos de explotación de datos dependerá de sus objetivos estratégicos (Ghoshal *et al.*, 2014), de sus capacidades tecnológicas, financieras y en recurso humano, al igual que de su capacidad organizacional para potenciar sus recursos (Wang & Hajli, 2017). En consecuencia, la estimación del valor potencial que genera la explotación de datos estará fuertemente relacionada con el grado de interacción y madurez que tenga la entidad en relación con el aprovechamiento de datos. Esto significa que si una entidad ha desarrollado antes proyectos de analítica de datos, es más factible que reconozca el valor que genera esta clase de proyectos y, en tal medida, pueda identificar contribución del aprovechamiento de datos para la generación de valor público.

Lo anterior implica que para muchas entidades el valor potencial de la *big data* se encuentra aún oculto y se revelará a medida que avancen en la implementación de procesos de explotación de datos.





3.3 ◉

RECOMENDACIONES PARA ESTRUCTURAR LAS NECESIDADES DE INVERSIÓN REQUERIDA Y ESTIMAR EL RETORNO ESPERADO DE LA EXPLOTACIÓN DE DATOS

La inversión en explotación de datos, y su retorno esperado, son dos conceptos importantes para las entidades, pues sirven referencia para impulsar proyectos e iniciativas de explotación de datos y *big data*. Al considerar que el objetivo del modelo de explotación de datos es convertirse en una herramienta que fomente en las entidades el aprovechamiento de datos, en este componente se abordan recomendaciones para estructurar las necesidades de inversión en tecnología y en recurso humano de acuerdo con el nivel de madurez que la entidad quiere alcanzar. Así mismo se brindan orientaciones técnicas para la estimación y comprensión de un retorno esperado de la inversión en explotación de datos.



3.3.1 ◉

INVERSIÓN REQUERIDA

Los recursos humanos y tecnológicos son factores de producción para el diseño e implementación de proyectos y estrategias de explotación de datos y exigen inversiones acordes con los requerimientos de las entidades. Este componente brinda lineamientos a las entidades para que estructuren sus requerimientos tecnológicos, y acudan en la medida de lo posible a las herramientas de agregación de demanda (nube pública III, nube privada III) que dispone el Estado colombiano a través de la Agencia Nacional de Contratación Pública- Colombia Compra Eficiente, para adquirir los servicios en tecnología y en recurso humano acordes con sus necesidades y características particulares⁸

Inversión en recursos tecnológicos

Al estimar la inversión en tecnología para el desarrollo de proyectos de analítica debe tenerse en cuenta los siguientes factores: 1) Características de los datos que gestiona la entidad: Variabilidad y volumen, 2) Necesidades alrededor del ciclo de vida de los datos en materia de generación y captura (tipo de ingesta batch o en streaming) y (Volumen de datos que ingresan a la plataforma kb/seg), almacenamiento (OLAP, OLTP, NoSQL, No estructurado), procesamiento, intercambio (cantidad de datos de ingreso y egreso, y número de solicitudes con otras entidades) y visualización (cantidad de datos que se van a publicar, consultas en promedio mes, tiempo de respuesta, cantidad de usuarios que acceden a los tableros), 3) Necesidades alrededor de los procesos organizacionales como gobernanza de datos, gestión del ciclo de vida de los datos, y seguridad y privacidad de la información.



Sobre estos factores se plantea para cada uno de los niveles de madurez un marco de necesidades de arquitecturas de referencia, que le permite a las entidades públicas orientar sus necesidades en tecnología (Consultar anexo 6. Arquitecturas de referencia para *big data*). Es relevante mencionar, que esta es una propuesta indicativa, dado que las necesidades de tecnología en materia de infraestructura, servicios y plataformas deberán estimarse de acuerdo con las necesidades propias de la entidad, considerando aspectos como su tamaño, su carácter nacional o territorial, entre otras.

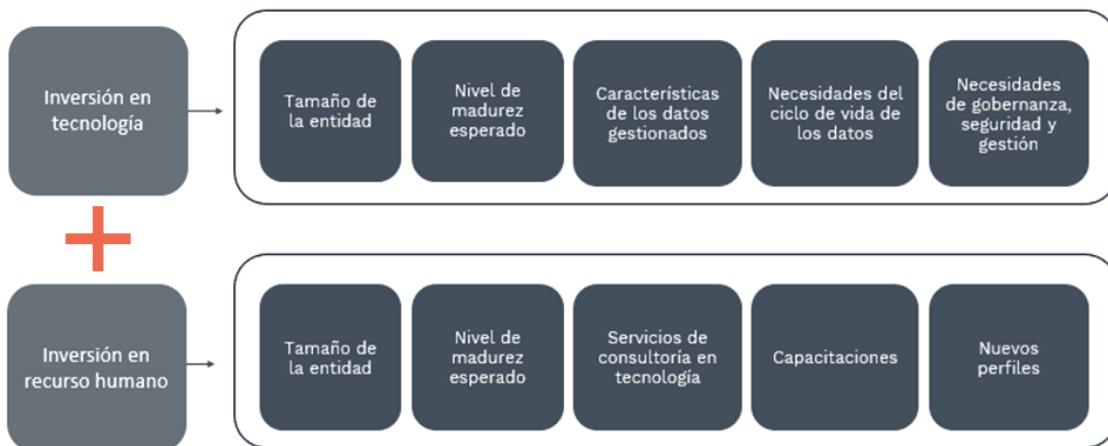
Para la estimación de precios en servicios de tecnología, las entidades públicas cuentan con las herramientas de compra agregada ofrecidas por la Agencia Colombiana de Contratación Pública - Colombia Compra Eficiente en las que se puede acceder a los precios de servicios de centros de datos, nube pública y nube privada. En caso de adquirir servicios en la nube, es importante que la entidad se guie por los procesos de migración de los datos de la entidad desde una arquitectura On-premise. Para ello es importante que la entidad tenga en cuenta la cantidad de datos a ser migrados, los sistemas de integración para la migración, y la cantidad de migraciones por día. Y, por último, el desarrollo de funcionalidades de centro de contacto con el ciudadano a través de chatbots para resolver solicitudes externas.

Inversión en recursos humanos

En lo concerniente con los recursos humanos, la definición del equipo que se requiere para la puesta en marcha de estrategias de explotación de datos debe considerar el tamaño de la entidad, los servicios de consultoría o asesoría en tecnología, las capacitaciones para fortalecer las capacidades del recurso humano actualmente vinculado a la entidad, y la contratación de nuevos perfiles en caso en que se considere. En la (figura 3-5) se incluye un resumen de lo mencionado en esta sección en materia de inversión tecnológica y en recurso humano.

Figura 3-5.

Estimación de la inversión tecnológica y en recurso humano



Fuente: DNP, DDD



3.3.2 ◈

RETORNO ESPERADO A PARTIR DE INVERSIÓN EN EXPLOTACIÓN DE DATOS

Este componente brinda orientaciones a las entidades públicas para comprender el concepto de retorno esperado de las inversiones en explotación de datos.

De manera general, calcular el retorno de inversión en explotación de datos requiere de un proceso gestionado de planeación en la entidad en el que: 1) se consignen las inversiones en tecnología (infraestructura, plataformas, software y mantenimiento) y en recursos humanos (capacitaciones, vinculación de nuevos perfiles) que realice la entidad en un periodo de tiempo. 2) Se identifique la variable de ganancia o de rendimiento que genera la entidad a partir de la explotación de datos, que para efectos de este modelo se concibe como la generación de valor. De acuerdo con la OCDE esta identificación de utilidades puede darse en torno a beneficios económicos directos (ahorro o reducción de gastos) o en términos de generar externalidades positivas para la sociedad.

La identificación del retorno esperado de la inversión en estrategias de explotación de datos, es recomendable iniciarla en proyectos tangibles que generen valor a la entidad alrededor de un objetivo medible. Para ello, se recomienda incluir en el proceso de planeación del proyecto de explotación de datos, cuáles son las ganancias potenciales en términos sociales y económicos que tendría para la entidad el desarrollo de ese proyecto y discriminar de manera precisa en qué costos estaría incurriendo la entidad a nivel de tecnología y recurso humano. Este ejercicio, también contribuye a impulsar las estrategias y proyectos de analítica de datos en los tomadores de decisiones de la entidad.



De acuerdo con Dataiku analytical y la OCDE (OCDE, 2019), la fórmula para calcular el retorno de la inversión de iniciativas, estrategias o proyectos de explotación de datos para un periodo de tiempo dado.

$$\text{Retorno esperado de inversión} = \frac{\text{Valor generado}}{\text{Costos de inversión}} \times 100$$

De esta forma los criterios para abordar el retorno esperado de acuerdo con el enfoque del modelo de explotación de datos sería la siguiente fórmula:

$$\text{Retorno esperado} = \frac{\text{Valor potencial de la explotación de datos}}{\text{Costos de inversión en tecnología y capital humano}} \times 100$$

En la siguiente tabla se encuentra un consolidado del valor potencial y de los costos de inversión en tecnología y recurso humano que se deberían tener en cuenta para la estimación del retorno.

Tabla 3-3. Estimación de la inversión tecnológica y en recurso humano

	Valor potencial	Costos
Identificadores	<ul style="list-style-type: none"> *Eficiencia en la prestación de bienes y servicios *Eficiencia en la toma de decisiones en política pública *Eficiencia en la gestión pública 	<ul style="list-style-type: none"> *Tecnología (Infraestructura, software, plataforma, servicios, soporte y mantenimiento) *Recurso humano (Capacitaciones, nuevos perfiles, consultorías)

Fuente: DNP, DDD



3.4 ◦

HOJA DE RUTA AJUSTADA PARA LA IMPLEMENTACIÓN DEL MODELO CON LAS ENTIDADES PÚBLICAS

La hoja de ruta como instrumento para avanzar hacia los niveles de madurez o preparación para la explotación de datos que espera alcanzar la entidad involucra la definición de objetivos de corto, mediano y largo plazo. Ese documento en el que constan las actividades para la implementación del modelo con las entidades públicas debe ser monitoreado de manera periódica, a fin de ajustarlo a las condiciones cambiantes que se presenten en el entorno y a las condiciones de la entidad; por ejemplo, las de tipo presupuestal y organizacional.

Para avanzar en el nivel de madurez que se espera alcanzar en cada una de las dimensiones, es recomendable la construcción colectiva en la entidad de una hoja de ruta, que se entiende como (Bonilla, 2018):

Un registro que permite identificar y planificar las diferentes operaciones que se deben realizar en un proyecto con miras al alcance de los objetivos establecidos para un fin, las cuales deben ser colocadas en secuencia, especificando los plazos y recursos necesarios para el incremento de las operaciones, el cual se puede convertir en un plan de acción a corto, mediano y largo plazo buscando servir como base a las entidades para saber cómo están establecidas y hacia dónde quieren llegar.

La finalidad de construir una hoja de ruta se enfoca en identificar elementos —proyectos, alianzas estratégicas, insumos propios de la entidad— que permitan alcanzar el nivel de madurez para la explotación de datos y *big data* que requiere la entidad según su misionalidad. Para la elaboración de la hoja de ruta se deben considerar los principios referenciados en la **figura 3-6**.

Figura 3-6.

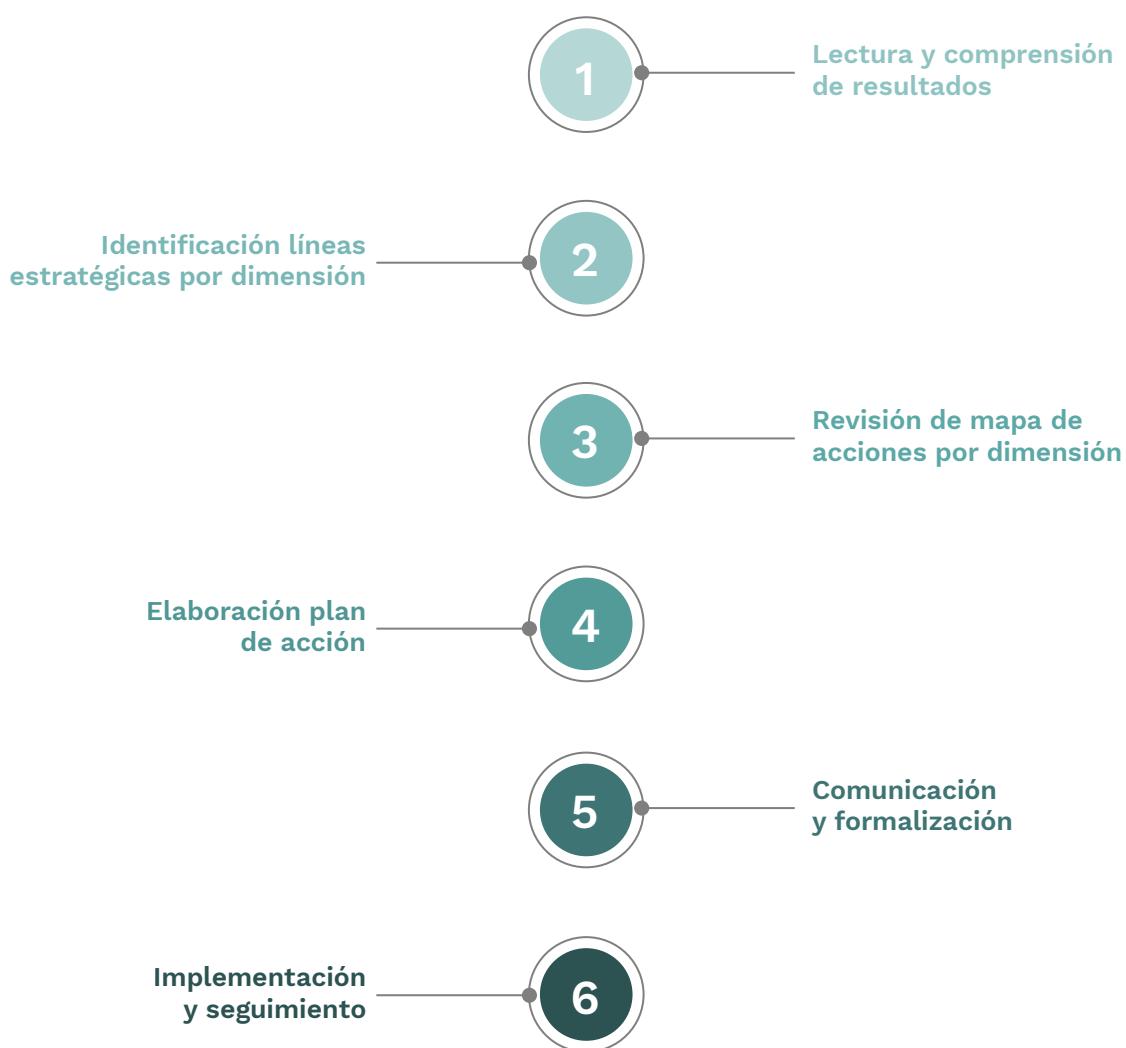
Principios para la definición de una hoja de ruta

- 1 ► Construcción colectiva en el que se incluya a las áreas estratégicas de la entidad
- 2 ► Tener el respaldo de los directivo de la entidad
- 3 ► Definir e implementar una secuencia de pasos lógicos
- 4 ► Considerar todas las dimensiones del Modelo de madurez que cubren la explotación de datos y los resultados obtenidos en el índice de Madurez
- 5 ► Tener en cuenta las condiciones específicas de la entidad, como presupuesto, tamaño, misionalidad y planeación estratégica

Fuente: DNP, DDD

Para el diseño de la hoja de ruta se establecen seis etapas, que orientan a las entidades en la definición de acciones para avanzar al nivel de madurez esperado en cada una de las dimensiones (**figura 3-7**).

Figura 3-7. Pasos para la definición de una hoja de ruta



Fuente: DNP, DDD



PASO 1. LECTURA Y COMPRENSIÓN DE RESULTADOS

Con los actores y dependencias internas que la entidad identifique pertinentes, se socializa el autodiagnóstico del modelo de madurez de explotación de datos, de acuerdo con los resultados obtenidos por la entidad.

La correcta interpretación de dichos resultados obtenidos en el modelo de explotación de datos es fundamental para la construcción de la hoja de ruta, pues permite identificar el punto de partida en el que se encuentra la entidad en cada una de las dimensiones del modelo de madurez. Los instrumentos de comunicación de resultados incluyen: (1) la caracterización de la entidad por nivel de madurez y dimensión, (2) la caracterización del nivel de madurez y dimensión de acuerdo con el nivel de madurez esperado y (3) los resultados sobre el valor potencial, la inversión requerida y el retorno esperado.



PASO 2: IDENTIFICACIÓN DE LÍNEAS ESTRATÉGICAS POR DIMENSIÓN

Por cada una de las dimensiones del modelo de madurez, se identifican líneas estratégicas (**figuras 3-8 y 3-9**) que facilitarán posteriormente la priorización de los hitos o acciones definidas en la hoja de ruta que elabore cada entidad.

Figura 3-8. Líneas estratégicas comunes a las dimensiones de la estrategia organizacional

Dimensión estratégica

Dimensión táctica

Dimensión operativa

Calidad de datos / Seguridad y protección de datos

Fuente: DNP, DDD



Las dimensiones estratégica, táctica y operativa comparten dos líneas estratégicas de manera transversal, calidad de datos y seguridad y protección de datos.

- **Calidad de datos.** La OCDE define la calidad de los datos como un concepto multifacético, en términos de exactitud y precisión, exhaustividad y claridad, coherencia e integridad, diseño de metadatos, integridad, singularidad y relevancia, puntualidad y validez (OCDE, 2019). La falta de datos de calidad conduce a tomar decisiones erradas para los grupos de interés, además produce impactos no deseados o negativos. Por lo anterior, la calidad de datos debe ser uno de los principales objetivos de la gobernanza que se define en el aspecto estratégico —roles, políticas y métricas—. También se debe fomentar, desde la alta dirección, una cultura enfocada en mantener la calidad de los datos para tomar buenas decisiones (Mertilos, 2015).

En el aspecto táctico se definen los procesos, protocolos y estándares para materializar la gobernanza de datos en materia de calidad. Aquí se deben monitorear y documentar la calidad de los conjuntos de datos, gestionar los riesgos y problemáticas que surjan, y escalarlos de forma adecuada a los directivos.

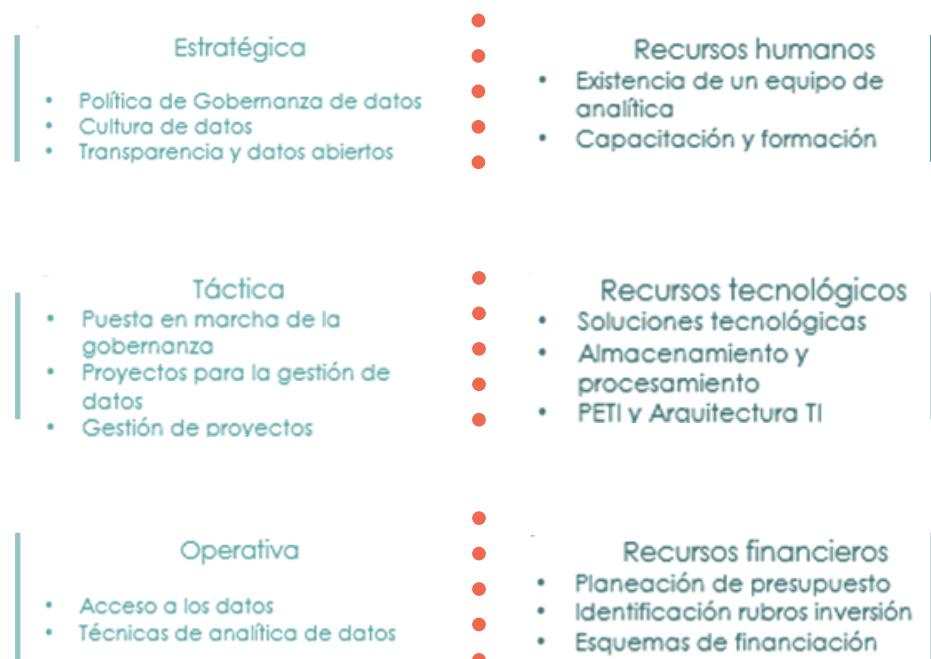
En el aspecto operativo, los perfiles encargados del tratamiento de los datos deben emplear técnicas para depurar datos internos y externos de forma tal que se garantice el cumplimiento de los estándares y se reporte el nivel de calidad de datos (KbInformatika, s.f.).

- **Seguridad y protección de datos.** La seguridad de los datos es uno de los objetivos principales que busca alcanzar una política de gobernanza de datos. Gartner define su gobernanza como un subconjunto de la gobernanza de datos que se ocupa específicamente de la protección y confidencialidad, tanto de los datos estructurados como de los no estructurados.

En el aspecto estratégico se definen las políticas para gestionar la protección de datos, los roles, los responsables y los niveles de acceso a los datos. En el aspecto táctico se definen los mecanismos para implementar las políticas de seguridad de datos de la entidad, la documentación de procesos y la gestión de riesgos de violación de la seguridad y privacidad de datos. El aspecto operativo se aplica a los estándares de seguridad de la información.

Figura 3-9.

Líneas estratégicas específicas a las dimensiones del Modelo de Madurez de explotación de datos



Fuente: DNP, DDD



A continuación, se describen las **líneas estratégicas por cada dimensión**:

Dimensión estratégica

- **Política de gobernanza de datos y big data.** La definición de una política de gobernanza de datos y *big data* asegura la accesibilidad y calidad de los datos de fuentes internas y externas durante todo su ciclo de vida, y la seguridad y calidad de la información en el marco de una estrategia estandarizada. Esta política debe contar con el apoyo de la alta dirección de la entidad y contar con un esquema de seguimiento compuesto de indicadores de medición. En el marco de los objetivos misionales de la entidad, la gobernanza debe abordar, como mínimo, la identificación de objetivos, estándares y métricas de medición sobre la calidad y seguridad de los datos, la gestión de datos maestros y metadatos.
- **Cultura basada en datos.** Una cultura basada en datos significa que los procesos, acciones, iniciativas, políticas y programas de la entidad están soportados en datos, a los cuales se considera como el activo central para la toma de decisiones en la entidad. Además, se cuenta con el respaldo de alta dirección para impulsar acciones que visibilicen la importancia de contar con datos de calidad, como la estructuración de una política de gobernanza de estos para articular procesos y recursos que permitan su aprovechamiento como un objetivo común.
- **Transparencia y datos abiertos.** La apertura de datos públicos se establece como principio de la gobernanza de datos. Para ello es importante integrar en la política de gobernanza de la entidad, las políticas, lineamientos y estándares de datos abiertos que ha elaborado el Gobierno, para garantizar la disponibilidad, calidad, acceso, reutilización y protección, e interoperabilidad. También conviene consolidar la cultura de datos abiertos, para impulsar la apertura de datos públicos digitales con el fin de ponerlos a disposición de los ciudadanos.



- ### Dimensión táctica
- **Puesta en marcha de la política de gobernanza.** En ella se definen y documentan estándares y procedimientos que respaldan la gobernanza de los datos y la gestión de riesgos sobre incumplimientos en las políticas definidas; al igual que se diseñan y ejecutan planes para coordinar las actividades operativas con los grupos de analítica de datos de la entidad. También en su implementación se define una estrategia de comunicación clara con el nivel operativo para lograr la ejecución de las políticas, las normativas y los estándares determinados en la gobernanza de datos, relacionados con su protección, calidad y gestión (The Data Administrator Newsletter, 2016).
 - **Proyectos y soluciones para la gestión de los datos.** La entidad estructura planes de gestión que pueden incluir temas relacionados con la arquitectura de datos, datos maestros, metadatos, calidad de datos (Armellino & Paolino, 2019).
 - **Gestión de proyectos e iniciativas de explotación de datos.** La entidad efectúa un seguimiento detallado a los procesos e iniciativas relacionados con la gestión del ciclo de vida de los datos. Los equipos de nivel medio comunican a los directivos, riesgos, barreras y obstáculos que se están presentando para el cumplimiento de las metas definidas en la política de gobernanza de datos.



Dimensión operativa

- **Acceso a los datos para su aprovechamiento.** Se ubican y administran repositorios centrales y distribuidos de datos, para el acceso a la información, sujetos tanto a los permisos de acceso como a las políticas de seguridad y privacidad definidas en la gobernanza de datos.
- **Técnicas de analítica de datos.** Los procesos de analítica de datos hacen uso de algoritmos con código en diversos lenguajes de *software* estadístico y de lenguajes de uso generalizado como *R* y *Python*, para el procesamiento automático de la explotación rutinaria de datos y su transformación, apoyados en soluciones tecnológicas escalables de acuerdo con las necesidades de análisis de la entidad.

Recursos humanos

- **Conformación de equipo(s) de analítica de datos.** La conformación del equipo de analítica de datos responde a las necesidades de información de la entidad y de los planes, programas y proyectos que se busque o solicita implementar. La vinculación de perfiles en la entidad se soporta en las necesidades en competencias y conocimientos que se demandan para conocer el modelo de negocio de la entidad, elaborar modelos de analítica de datos, así como para visualizar y presentar además de datos resultados que respondan a las necesidades directivas. No siempre se requiere la vinculación de personas externas, pues hay funcionarios y contratistas dentro de la entidad con los saberes necesarios para desarrollar proyectos de analítica de datos o conocimiento base para recibir capacitación.



- **Capacitación y formación en analítica de datos.** La capacitación en analítica y explotación de datos se incorpora en los planes de capacitación y formación de la entidad, identificando de manera previa cuáles son las necesidades de formación con base en los proyectos que se requiere ejecutar, se describen los perfiles y los equipos asignados para tomar los cursos y estrategias; por último, se construyen estrategias para capitalizar los conocimientos en la entidad.

Recursos tecnológicos

- **PETI y arquitectura TI:** El Plan Estratégico de las Tecnologías de la Información y el marco de arquitectura de TI de la entidad trazan la hoja de ruta para definir e implementar las estrategias de tecnologías de la información de la entidad. En tal sentido, el PETI debe estar alineado con las políticas de explotación de datos y transformación digital del Gobierno. Es fundamental para al aprovechamiento de datos en la entidad incluir en el Plan los dominios de gobernanza o gobierno de datos, la gestión del ciclo de vida de los datos, los planes de calidad de datos, lo mismo que los estándares para el intercambio y la interoperabilidad.

- **Adopción de soluciones tecnológicas.** La explotación de datos implica la adquisición de tecnologías para el almacenamiento y procesamiento de los datos. Es recomendable fijar la atención en pilotos o prototipos que visibilicen e incorporen soluciones de manera gradual a las necesidades de tecnología de tal manera que sean escalables. La nube pública es una solución que habilita la provisión de la infraestructura y los servicios para el almacenamiento y procesamiento de datos sin hacer una gran inversión financiera. Los integradores de soluciones de analítica de datos incluyen un portafolio completo de funcionalidades, como la integración con diferentes soluciones, aplicativos -proveedores *cloud*, Hadoop y otros- y plataformas de análisis y visualización de datos -Python, RStudio y otros- (Accenture, 2019).

Recursos financieros

- **Asignación de presupuesto en la planeación.** El desarrollo de iniciativas de analítica de datos ha de tener como soporte un presupuesto definido, que atienda a las necesidades de recurso humano y de infraestructura tecnológica de la entidad para el despliegue de proyectos estratégicos que visibilicen el valor de los datos. Dado que estos conforman un activo estratégico, las actividades y proyectos relacionados con su gestión mantenimiento y aprovechamiento para la toma de decisiones, deben tener un rubro asignado para cada vigencia. Este rubro está soportado por análisis técnicos que respalden la inversión financiera de la entidad en explotación de datos.



PASO 3.

REVISIÓN DEL MAPA DE ACCIONES POR DIMENSIÓN PARA AVANZAR EN EL NIVEL DE MADUREZ DE EXPLOTACIÓN DE DATOS

El mapa de ruta presenta por cada dimensión del modelo de madurez de explotación de datos, las acciones que la entidad podría llevar a cabo para avanzar al nivel esperado. Estas acciones aportarán a la elaboración de un plan de trabajo y su hoja de ruta para que la entidad lo ejecute (anexo 4: Mapa de acciones).



PASO 4:

ELABORACIÓN DEL PLAN DE ACCIÓN

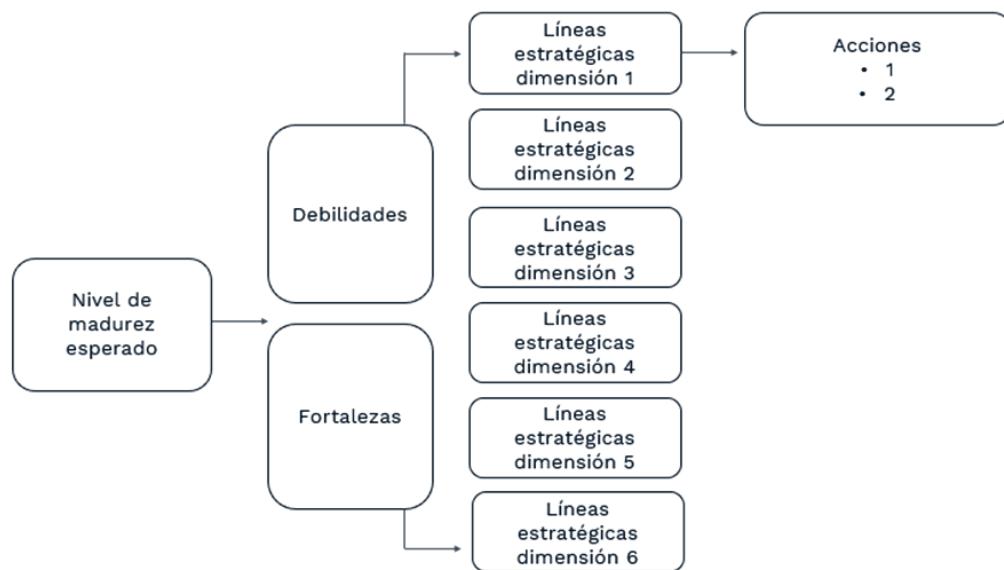
En el plan de acción se identifican las principales líneas estratégicas de la entidad y se definen metas de corto, mediano y largo plazo. Así, el nivel de madurez deseado será el marco general del plan para la identificación y posterior implementación de acciones.

Es recomendable practicar este ejercicio de manera colaborativa y garantizando que todas las áreas participen, para que sea desarrollada de manera integral cada una de las dimensiones que contempla el modelo de explotación de datos.

En el plan de acción se definen las líneas estratégicas priorizadas por la entidad y las acciones relacionadas (**figura 3-10**). Cada acción debe contemplar: (a) la descripción detallada de acción, (b) la definición de tiempos de ejecución, (c) la definición de recursos y (d) la asignación de líderes y equipos responsables y corresponsables (**figura 3-11**).

Figura 3-10.

Proceso de identificación de los elementos para la construcción de la hoja de ruta



Fuente: DNP, DDD

Figura 3-11.

propuesta de hoja de ruta

Líneas estratégicas	Acciones	Tiempo	Recursos	Responsables
Línea 1	a.			
	b.			
	c.			
Línea 2	a.			
	b.			

Fuente: DNP, DDD



○ Descripción detallada de la acción.

Son las actividades que refieren cómo se logrará el cumplimiento de los objetivos de cada línea estratégica, garantizan la integración, articulación y continuidad de esfuerzos, de manera ordenada, coherente y sistemática. Estas acciones se pueden establecer de acuerdo con lo establecido en el anexo 3: Mapa de acciones.

○ Establecer tiempos.

Aquí se señalan los lapsos para el cumplimiento de cada una de las acciones, que se organizan en corto, mediano o largo plazo.

○ Establecer recurso.

En los campos correspondientes se relacionan a los recursos necesarios para diseñar la estrategia que permitirá materializar las líneas de acción.

○ Asignación de líderes y equipos.

Esta parte integra la descripción del grupo o área de respaldo propuesto y aprobado para liderar cada una de las acciones.



PASO 5. COMUNICACIÓN Y FORMALIZACIÓN

Con el fin de cumplir lo planeado se requiere formar espacios de socialización en los que se involucre tanto el líder de la entidad como los de las dependencias más estratégicas para impulsar la hoja de ruta. Esto promueve que haya una visión holística del plan de trabajo y así se tenga el respaldo de la alta dirección de la entidad.



PASO 6: IMPLEMENTACIÓN Y SEGUIMIENTO

Es la ejecución, realización y gestión del plan de acción diseñado para materializar la hoja de ruta; para ello se requiere revisar la viabilidad del plan construido, integrar los procesos de la entidad y organizar pequeños equipos de trabajo que tendrán la responsabilidad de la implementación técnica del plan de acción. Aquí se requiere desarrollar un proceso de seguimiento del estado de avance de la puesta en marcha de la hoja de ruta en cada una de las dimensiones, para validar los logros y dificultades que se planearon. El seguimiento se efectúa a cada una de las acciones planteadas por las líneas estratégicas y al cumplimiento de los tiempos establecidos en el plan de acción. Este monitoreo debe realizarse de manera periódica con las dependencias responsables de la entidad.

La identificación de los tiempos para avanzar a un nivel de madurez depende de cada entidad: los tiempos de inversión, el presupuesto asignado para explotación de datos, la disposición de los líderes responsables en la entidad y el cumplimiento de la implementación del plan de trabajo.

+

4

RECOMENDACIONES FINALES

.....

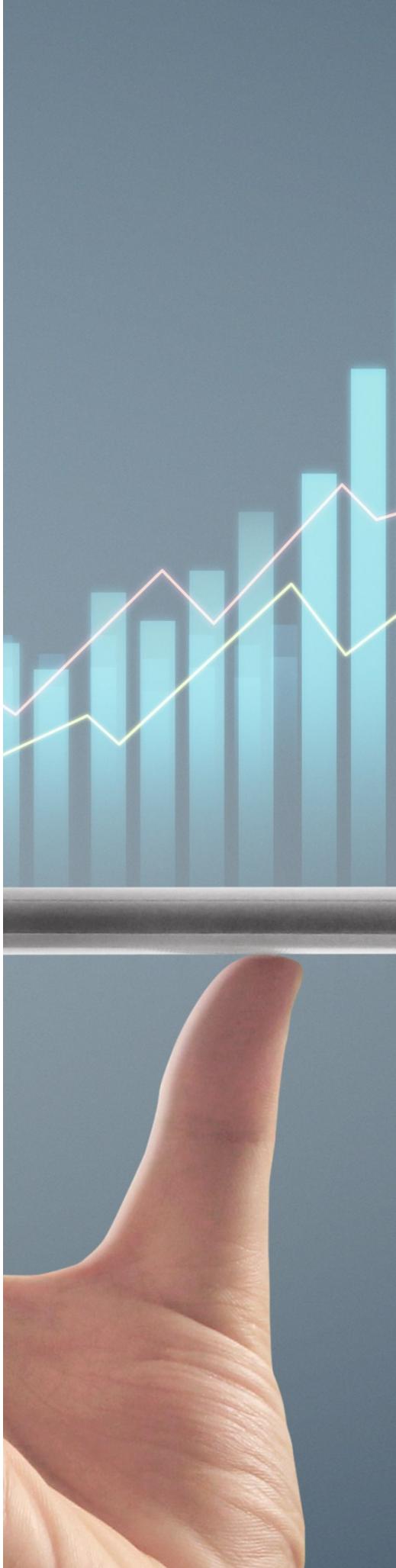
-

- ▲ Internalizar e institucionalizar el modelo de implementación para la explotación de datos, de tal forma que el autodiagnóstico sobre el nivel de madurez de la entidad, el valor público que genera la explotación de datos y las inversiones requeridas para avanzar hacia la consolidación de una entidad impulsada y guiada por datos se articulen con la implementación de la Política de Gobierno Digital dentro de la entidad.

- ▲ Definir indicadores claros sobre los proyectos de aprovechamiento de datos, con especial énfasis en los que permitan analizar los impactos de estas iniciativas sobre los grupos de interés de la entidad. Esos indicadores permitirán tener una idea más clara del valor público agregado que generan los proyectos y ayudarán a legitimar y fortalecer los esfuerzos relacionados con la explotación de datos (OCDE, 2019b).

- ▲ Involucrar a las distintas áreas en el proceso de autodiagnóstico sobre explotación de datos, para así garantizar la toma de decisiones y la definición de líneas estratégicas de acción desde un panorama más amplio e integral de la entidad.





- ▶ Optimizar procesos y posicionar la información como un activo intangible de infraestructura. Una opción para lograrlo es desarrollar o implementar métodos de valoración de los datos —como la estimación del valor público y el índice de valor potencial—, que sean fundamento para la toma de decisiones sobre inversión en explotación de datos.
- ▶ Buscar desde la misionalidad de cada entidad, la generación de valor a través del uso colaborativo de los datos de manera interinstitucional, de tal forma que para el desarrollo de iniciativas basadas en datos que busquen responder a una necesidad específica, se pueda contar con la mejor información de manera oportuna, aunque esta no se produzca dentro de la entidad líder de la iniciativa.
- ▶ Desarrollar nuevos modelos de financiación de iniciativas de explotación de datos que promuevan el desarrollo de ideas retadoras y escalables que respondan a preguntas o necesidades de la entidad. El desarrollo de hackatones, *data tasks*, incubadoras de ideas o programas de innovación abierta podrían constituirse como herramientas para la asignación de estas financiaciones.

+

5

BIBLIOGRAFÍA

.....

-



PIE DE PÁGINA

1. Estas recomendaciones se relacionan con implementar una gobernanza de datos en el sector público, y la aplicación de los datos en el sector público para generar valor público. En las consideraciones que debe abordar el Gobierno no solo se deben contemplar las asociadas a una base técnica y tecnológica para el aprovechamiento de datos, sino que también se debe incorporar el fortalecimiento de una cultura basada en datos, y una adecuada organización y gobernanza de los datos (OCDE, 2019).
2. El término de *big data* surgió inicialmente como un concepto para describir grandes volúmenes de datos que excede la capacidad de las bases de datos tradicionales para su almacenamiento y procesamiento; no obstante, el término tiene un alcance mayor que no se refiere solamente a los datos, sino a las tecnologías que se requieren para su almacenamiento y procesamiento, a procesos para ejecutar tareas y a las personas que ejecutan esas tareas y que aprovechan los datos para responder a problemáticas sociales. En ese sentido, *big data* se convierte en un sistema social y técnico, que requiere de reglas e instituciones para su consolidación (Martínez, Rodríguez, Ron, Triulzi, & Santana, 2017).
3. Este término, acuñado por el Tavistock Institute a mediados de los 1950 como lo señala Long (2013), se usa en la actualidad “para abarcar las complejas interacciones entre la tecnología y las personas, así como sus consecuencias psicológicas y culturales”. Las grandes cantidades de datos son resultado de la interacción entre personas, tecnologías y procesos.
4. Un buen ejemplo de desarrollar los procesos de *big data* en un ambiente de nube es *MapReduce*, el cual acelera el procesamiento de grandes volúmenes de datos en la nube (Braun, 2015).

PIE DE PÁGINA

5. En la infraestructura como servicio, el proveedor ofrece el espacio de servidores, almacenamiento, red, conexión a internet, y no son necesarias grandes inversiones. En las plataformas como servicio, el proveedor ofrece la posibilidad de desarrollar y adquirir aplicaciones con una función específica - servidor de correo, servidor de servidores- y presta la infraestructura necesaria para ello. En el *software* como servicio los clientes o usuarios se conectan al *software* con un propósito específico a través de un API (application programming interface, por sus siglas en inglés) del proveedor o por medio de la web (DNP, 2020).
6. Los ciclos de vida de los datos en el sector público son: (1) la recopilación y generación de datos; (2) el almacenamiento, protección y procesamiento de datos; (3) el intercambio, la conservación y la publicación de datos; y (4) el uso y reutilización de datos (OCDE, 2019).
7. En términos contables, las tres características esenciales de un *activo* se identifican mediante las siguientes preguntas: (1) ¿tiene el activo potencial de servicio o proporciona beneficios económicos futuros, cuando el beneficio proviene del uso del activo o de su venta; (2) ¿el activo está controlado por la organización?, es decir, ¿la organización tiene el poder de beneficiarse del activo y negar o regular cómo otros acceden a él?; (3) ¿es el activo el resultado de una actividad anterior?, ¿el activo ha sido recolectado a través de un proceso, adquirido de una fuente de alguna manera o realizando algún trabajo para desarrollarlo? (OCDE, 2019).
8. <https://www.colombiacompra.gov.co/tienda-virtual-del-estado-colombiano/tecnologia?page=1>

- Abdallah, Z., Du, L., & Webb, G. I. (2017). Data Preparation. En C. Sammut, & G. Webb, *Encyclopedia of Machine Learning and Data Mining*. Nueva York: Springer US.
- Abdullah, R., Al-Sai, Z., & Husin, M. (2019). *A Review on Big Data Maturity Models*. IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology: <https://doi.org/10.1109/JEEIT.2019.8717398>
- Accenture. (2019). *Accenture-lideres-updated-marketing-report-up-integradores-big-data*. https://www.accenture.com/_acnmedia/pdf-100/accenture-lideres-updated-marketing-report-up-integradores-big-data.pdf
- Armellino, M., & Paolino, M. (2019). *Modelo de madurez para gobernanza de Big Data*. <https://dspace.ort.edu.uy/bitstream/handle/20.500.11968/4070/Material%20completo.pdf?sequence=-1&isAllowed=y>
- ASIP. (2014). *La cadena de valor público: un principio ordenador que previene la colisión metodológica*. <http://asip.org.ar/la-cadena-de-valor-publico-un-principio-ordenador-que-previene-la-colision-metodologica/>
- Braun, H. (2015). *Evaluation og Big Data Maturity models*. <https://core.ac.uk/download/pdf/196555414.pdf>
- Conejero, E. (2014, mayo). Valor público: Una aproximación conceptual. *Empresa: investigación y pensamiento crítico* 3 (1), 31-40.
- De Bruin, S., Freeze, R., & Kulkarni, U. (2005). *Understanding the Main Phases of Developing a Maturity Assessment Model*. https://www.researchgate.net/publication/27482282_Understanding_the_Main_Phases_of_Developing_a_Maturity_Assessment_Model
- DNP. (2020). *Aprovechamiento de datos para la toma de decisiones en el sector público*. <https://www.dnp.gov.co/DNPN/Paginas/Aprovechamiento-de-datos-para-la-toma-de-decisiones-en-el-sector-publico.aspx>

- Fraefel, M., Haller, S., & Gschwend, A. (2017). *Big Data in the public sector Linking city sensors*. https://www.researchgate.net/publication/318879744_Big_Data_in_the_Public_Sector_Linkings_Cities_to_Sensors
- Ghavami. (2015). *Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics*.
- KbInformatika. (s.f.). *Data quality with data governance. Obtenido de Confusing Data Quality with Data Governance:* <https://kb.informatica.com/ipswhitepaper/1/Documents/Confusing%20Data%20Quality%20with%20Data%20Governance-2-23-15.pdf>
- Ladley, J. (2012). *Data Governance: How to design, deploy and sustain and effective data governance program.* <https://books.google.com.co/books?id=CpeAYWaTScYC>
- Martínez, S., Rodríguez, G., Ron, J., Triulzi, F., & Santana, P. (2017). *Hacia la utilización del Big Data en ciencias sociales: un enfoque histórico.* <https://es.scribd.com/document/443328023/Martinez-et-al-2017>
- Mertilos, A. (2015). *Development of a capability maturity model for big data governance.* <https://datasciencebe.files.wordpress.com/2015/07/development-of-a-capability-maturity-model-for-big-data-governance.pdf>
- Moore, M. (1995). *Creating Public Value: Strategic Management in Government*. Cambridge, Harvard.
- OCDE / Eurostat. (2018). *Guidelines for Collecting, Reporting and Using Data on Innovation*, 4th Edition, The Measurement of Scientific, Technological and Innovation Activitie.
- OCDE. (2019). *A data driven public sector. Obtenido de Enabling the strategic use of data for productive, inclusive and trustworthy governance:* <https://www.oecd-ilibrary.org/docserver/09ab162c-en.pdf?Expires=1602519478&id=id&accname=guest&checksum=59487CAA586D7326C250349F9EF857FA>



- OCDE. (2019). *El camino para convertirse en un sector público basado en datos.* <https://www.oecd-ilibrary.org/sites/059814a7-en/index.html?itemId=/content/publication/059814a7-en>
- OCDE. (2020). *Going digital. Obtenido de Data in the digital age:* <https://www.oecd.org/going-digital/data-in-the-digital-age.pdf>
- OCDE. (2019b). *OECD Open, Useful and Re-usable data (OURdata) Index: 2019.*
- Saltz, J., & Shamshurin, I. (2016). *Big data team process methodologies: A literature review and the identification of key factors for a project's success.* IEEE International conference: <https://ieeexplore.ieee.org/document/7840936>
- The data administrator Newsletter. (2016). *non-invasive-framework-for-data-governance-implementation-details.* <https://tdan.com/non-invasive-framework-for-data-governance-implementation-details-part-2/20130>
- Yanseen, E., Bouzembrak, Y., Hendriksen, P., & Staats, M. (2017). *Big data in food safety: A overview,* *Critical Reviews in Food Science and Nutrition.* <https://www.tandfonline.com/doi/full/10.1080/10408398.2016.1257481>
- Zdonek, D., & Krol, K. (2020). *Analytics Maturity Models: An Overview.* https://www.researchgate.net/publication/339672162_Analytics_Maturity_Models_An_Overview



El futuro
es de todos

Consejería Presidencial
para asuntos económicos
y transformación digital



El futuro
es de todos

DNP
Departamento
Nacional de Planeación



El futuro digital
es de todos

MinTIC