

Caracterización del ecosistema de datos en Colombia: Análisis desde un enfoque de gobierno



**Caracterización del ecosistema de
datos en Colombia**

Alexander López Maya
Director General

Versión final

Mario Alejandro Valencia
**Subdirector General de Prospectiva y Desarrollo
Nacional**

Viviana Vanegas Barrero
Directora de Desarrollo Digital

Javier Alfonso Lesmes
**Coordinador Grupo Transformación y Economía
Digital**

Asesores:

Diana Paola Ramírez Roa

Agustín Jiménez Ospina

Nicolas Enrique Chibuque Pérez

Francisco Perea de Zubiria

Departamento Nacional de Planeación,
Calle 26 núm. 13-19
Bogotá, D. C., Colombia
PBX: 3815000
www.dnp.gpv.co
Julio de 2024

Contenido

Siglas.....	5
INTRODUCCIÓN.....	7
I PARTE: DESCRIPCIÓN TEÓRICA DE LOS ECOSISTEMAS DE DATOS	10
1. Definición general de un ecosistema de datos.....	10
2. Principios de los ecosistemas de datos.....	11
3. Características de los ecosistemas de datos	12
4. Taxonomía de los ecosistemas de datos.....	14
4.1. Meta-dimensión económica: dimensiones y características	15
4.2. Meta-dimensión técnica: dimensiones y características	16
4.3. Meta-dimensión de gobernanza: dimensiones y características.....	16
5. Componentes generales de los ecosistemas de datos.....	16
5.1. Actores.....	17
5.2. Roles.....	18
5.3. Recursos	21
5.4. Incentivos.....	21
6. Creación de valor social y económico.....	22
7. Ecosistemas de datos hacia el bien común.....	22
8. Metodologías de caracterización de ecosistemas de datos.....	24
II PARTE: PROPUESTA DE CARACTERIZACIÓN DEL ECOSISTEMA DE DATOS EN COLOMBIA DESDE UN ENFOQUE DE GOBIERNO.....	25
1. Contexto general del ecosistema de datos en Colombia desde un enfoque de gobierno.....	26
2. Marco de entendimiento del ecosistema de datos en Colombia desde un enfoque de gobierno.....	28
3. Características generales del ecosistema de datos en Colombia	32
4. Componentes del ecosistema de datos de Colombia.....	39
4.1.1 Actores y roles	39
4.1.2 Recursos del ecosistema de datos.....	44
4.1.3 Incentivos y barreras que tienen los actores para participar en el ecosistema de datos	47
5. Conclusiones y recomendaciones	48
BIBLIOGRAFÍA	49

Figuras

Figura 1. Marco de entendimiento ecosistema de datos de Colombia.....	29
Figura 2. Principios coincidentes entre infraestructuras y ecosistemas	32
Figura 3. Esquema de alineación de la gobernanza de los datos.....	38
Figura 4. Tipos de datos que consumen los actores del ecosistema de datos.....	47
Figura 5. Incentivos para participar dentro del ecosistema de datos	48

Tablas

Tabla 1. Taxonomía general del ecosistema.....	14
Tabla 2. Actores generales del ecosistema.....	17
Tabla 3 .Descripción de roles de datos en el ecosistema.....	18
Tabla 4 . Componentes de la infraestructura de datos.....	30
Tabla 5. Taxonomía del ecosistema de datos para Colombia	33
Tabla 6. Actores sector público ecosistema de datos Colombia	39
Tabla 7. Actores sociedad civil y organizaciones mixtas de datos de Colombia	40
Tabla 8. Actores organismos multilaterales ecosistema de datos de Colombia.....	41
Tabla 9. Definición aproximada de los nuevos roles identificados en el formulario.....	43
Tabla 10. Recursos del ecosistema de datos de Colombia	44

Anexos

[Anexo A. Matriz mapeo de actores.xlsx](#)

[Anexo B. Recursos del ecosistema](#)

Siglas.

AGATA: Agencia de Analítica de datos de Bogotá.

AMI: Infraestructura de Medición Avanzada.

AND: Agencia Nacional Digital.

API: Application Programming Interface.

BID: Banco Interamericano de Desarrollo.

C4RI: Centro para la Cuarta Revolución Industrial.

CAF: Banco de Desarrollo de América Latina.

CINTEL: Centro de Investigación y Desarrollo en Tecnologías de la Información y las Comunicaciones.

CONPES: Consejo Nacional de Política Económica y Social.

DAFP: Departamento Administrativo de la Función Pública.

DANE: Departamento Administrativo Nacional de Estadística.

DAPRE: Departamento Administrativo para la Presidencia de la República.

DDD: Dirección de Desarrollo Digital.

DNP: Departamento Nacional de Planeación.

FAIR: Principios de fácil búsqueda, accesibilidad, interoperabilidad y reutilización.

FURAG: Formulario Único de Reporte y Avance en la Gestión.

IaaS: Infrastructure as a Service.

ICDE: Infraestructura Colombiana de Datos Espaciales.

IGAC: Instituto Geográfico Agustín Codazzi.

MinCIT: Ministerio de Comercio, Industria y Turismo.

MinTIC: Ministerio de Tecnologías de la Información y las Comunicaciones.

MNO: Mobile Network Operator (Operador de Red Móvil).

MVNO: Mobile Virtual Network Operator (Operador Móvil Virtual).

OCDE : Organización para la Cooperación y el Desarrollo Económico.

ODI: Open Data Institute.

OLAP: On-Line Analytical Processing.

PaaS: Plataforms as a Service.

PNID: Plan Nacional de Infraestructura de Datos.

SaaS: Software as a Service.

SEN: Sistema Estadístico Nacional.

SIC: Superintendencia de Industria y Comercio.

SIGEP: Sistema de gestión y de Información Pública del Empleo.

SUIT: Sistema Único de Información de Trámites.

TI: Tecnologías de la información.

INTRODUCCIÓN

El presente documento es el resultado de los compromisos establecidos en la hoja de ruta del Plan Nacional de Infraestructura de Datos (PNID)¹, expedido por el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) a través de la Resolución 460 de 2022. El propósito del PNID es fomentar iniciativas para el uso e intercambio de datos con el fin de generar valor económico, social o público en el país. Por ello, dentro de las acciones establecidas para su cumplimiento se propuso la caracterización del ecosistema de datos en Colombia, con el objetivo de crear un marco de entendimiento común en torno a los actores, roles y componentes asociados a este ecosistema.

Esta acción del PNID se planteó teniendo en cuenta el contexto actual en el que los datos desempeñan un papel muy importante para la creación de valor social y económico. Por lo que, los actores interesados han establecido dinámicas y acciones para extraer valor de los datos y han generado relaciones o nexos que configuran los ecosistemas de datos. Dicha interacción se establece en el marco de políticas y tecnologías, con el propósito de intercambiar, producir y consumir datos para el cumplimiento de diferentes fines. Particularmente, la estructuración de ecosistemas de datos se destaca en la economía basada en datos, dado que conduce a la generación de conocimiento, oportunidades de negocio y a la promoción de la investigación.

El término “ecosistema de datos” se encuentra en distintos documentos estratégicos y en documentos técnicos de tipo científico, por lo cual, su abordaje conceptual es diverso. Este depende del contexto en el que se plantee, que puede ser a nivel de política pública global, a nivel del sector privado, o a nivel de nichos específicos. Por ejemplo, la Organización para la Cooperación y el Desarrollo Económico (OCDE) en su publicación *Data Driven Innovation* (2015) elaboró un marco de entendimiento del ecosistema de datos a nivel global, donde se define al ecosistema como la integración e interacción entre diferentes partes, incluidos los titulares de datos, los productores de datos, los intermediarios de datos y los interesados a todo el ciclo de vida² de los datos.

Por su parte, el Instituto de Datos Abiertos (ODI por sus siglas en inglés) entiende al ecosistema de datos, como un escenario que consta de una infraestructura de datos, en la que un grupo de personas, comunidades y organizaciones que recopilan y administran datos, crean productos y toman decisiones a partir de ellos (Open Data Institute, 2018); el ecosistema también incluye a todos los demás colaboradores que

¹ Disponible en el Portal de Infraestructura de datos del Estado: <https://infraestructuradatos.gov.co>

² De acuerdo con el PNID (2022), el ciclo de vida puede entenderse como el proceso que se requiere para poder aprovechar los datos y se constituye por: creación, procesamiento, almacenamiento, intercambio, uso y análisis de los datos, y archivo y preservación.

influyen en cualquiera de estas actividades o se ven afectados por ellas. En el caso del instituto McKinsey se comprende al ecosistema de datos como una plataforma que combina datos de numerosos proveedores y genera valor mediante el uso de datos procesados (McKinsey, 2021).

La caracterización de los ecosistemas de datos depende de su contexto, de su nivel de desarrollo y nivel de apertura de los datos. Por ejemplo, los ecosistemas de datos abiertos son ampliamente flexibles a la apertura e intercambio de datos, algunos de estos son el Portal de datos abiertos del Estado colombiano³, el Portal de datos europeo, y el Global Forest Watch⁴. Por su parte, también existen ecosistemas de datos para el desarrollo de aplicaciones específicas, que se crean y desarrollan en el marco de un proyecto. Así mismo, existen los ecosistemas de tipo sectorial, como el caso de Smart Agriculture ecosystem que es una plataforma de intercambio de datos basada en tecnologías de Interfaz de Programación y Aplicaciones (API por sus siglas en inglés), en el que participan alrededor de 30 asociados que integran el sector agrícola de la Unión Europea, o Skywise de Airbus que proporciona datos para la industria de la aviación y sus socios incluyen más de 100 aerolíneas de todo el mundo.

No obstante, en el marco de su diversidad, los ecosistemas de datos están constituidos por componentes similares y se enfrentan a desafíos comunes, que van más allá de su nivel de desarrollo. Algunos de estos desafíos son: 1) la preservación de la privacidad y seguridad de los datos de los actores del ecosistema; 2) la democratización de la participación de diversos actores en el ecosistema para reducir el poder de mercado de datos; 3) la definición de esquemas o modelos de valoración de los datos (Data economy, 2020). Así mismo, en los diversos tipos de ecosistemas se abordan asuntos comunes como la gobernanza del ecosistema, los asuntos técnicos en relación con la accesibilidad e intercambio de los datos, el compromiso de los actores partícipes en el ecosistema y su sostenibilidad (Unión Europea, 2021).

La comprensión de los ecosistemas de datos se torna cada vez más importante, en la medida en que los datos adquieren relevancia para la creación de bienes y servicios en la sociedad, y cada vez son más los actores que participan alrededor del ciclo de vida de los datos. Comprender las dinámicas e interacciones alrededor de los datos aporta a la identificación de falencias y beneficios para los distintos actores del ecosistema y permiten reconocer los roles que existen para lograr extraer el valor de los datos. Este tipo de ejercicios también son importantes para los gobiernos en términos de inclusión social, teniendo en cuenta que el diseño e implementación de estrategias nacionales de

³ Disponible en: <https://www.datos.gov.co/>

⁴ Es una plataforma abierta que tiene disponibles datos agregados de distintas fuentes para monitorear el estado de los bosques y en este ecosistema participan más de 100 organizaciones de todo el mundo.

datos⁵ requieren de la participación de diversos actores de la sociedad. En este contexto de política pública, es clave que los actores participantes de los ecosistemas, como empresas privadas, academia y sociedad civil, conozcan las dinámicas alrededor de los datos y el escenario de la creación de valor social y económico derivada de estos.

Dado lo anterior, el objetivo de este documento es proporcionar una visión del ecosistema de datos en Colombia, a través de la caracterización de los componentes e interacciones que lo definen, mediante la definición de un marco de entendimiento común, con un enfoque de gobierno. Lo anterior es relevante teniendo en consideración el marco de política pública que Colombia ha consolidado en los últimos años en materia de datos, en el que se hace necesaria la identificación de actores claves que participan en el ciclo de vida de los datos y la generación de valor social y económico del ecosistema de datos en el país. Este documento técnico está acompañado de una Guía de caracterización cualitativa, que contribuye a orientar a las entidades públicas u otros actores, a caracterizar otros ecosistemas de datos en el país, y se espera sea un aporte importante en la consolidación de la infraestructura de datos del Estado colombiano y su modelo de gobernanza⁶.

Además de esta introducción, el documento está compuesto por tres partes. La primera contiene una descripción teórica amplia de los ecosistemas de datos, a través de la cual se aborda la definición de este tipo de ecosistemas, sus componentes, características y principios. La segunda contiene la caracterización del ecosistema de datos en Colombia desde un enfoque gubernamental, la cual se construye a partir del marco teórico propuesto. Adicionalmente, este documento está acompañado de dos anexos: El anexo A incluye una matriz con el mapeo de actores del ecosistema de datos en Colombia; el anexo B que incluye una matriz de recursos disponibles en el ecosistema de datos.

⁵ Este es el caso del Reino Unido, en el que se implementó un ejercicio participativo para visualizar el ecosistema de datos del país, como herramienta para impulsar la Estrategia Nacional de Datos que se diseñó en el año 2020, y se implementó en el 2021. En el caso de la Unión Europea, en el año 2020 se publicó la estrategia de datos, y en este marco se han elaborado documentos para la caracterización de los ecosistemas de datos en la región.

⁶ El modelo de gobernanza de datos se define y se crea mediante el Decreto presidencial 1389 de 20222

I PARTE: DESCRIPCIÓN TEÓRICA DE LOS ECOSISTEMAS DE DATOS

El crecimiento de las tecnologías digitales, como la inteligencia artificial, ha generado que los datos sean activos cada vez más importantes para distintos actores de la economía. En el contexto actual, los datos no solamente son importantes para crear conocimiento y mejorar la toma de decisiones, sino también para crear bienes y servicios innovadores que respondan a diferentes propósitos. La oportunidad que brindan los datos para la creación de valor da lugar a la organización de distintos actores alrededor del ciclo de vida de los datos. Las empresas, las organizaciones sociales, las agencias gubernamentales, los centros de estudio e investigación se juntan para alcanzar el valor económico y social de los datos.

En los ecosistemas de datos los actores que participan contribuyen y se beneficia del mismo ecosistema. La organización en red permite abordar las distintas necesidades del ciclo de vida de los datos, que incluye la generación, procesamiento, almacenamiento, intercambio y análisis, así como también la generación de información y de conocimiento, lo cual garantiza una creación de valor conjunta y sostenible (Capello , Gal, Rehof, & Jarke, 2019).

1. Definición general de un ecosistema de datos.

Los ecosistemas de datos son la integración e interacción entre diferentes partes, incluidos los titulares de datos, los consumidores de datos, los intermediarios de datos y los interesados. Estos actores interactúan o se ven afectados por acuerdos de acceso e intercambio de datos, que dependen de sus diferentes roles y responsabilidades, y que se desenvuelven en torno a los derechos, las tecnologías y los modelos comerciales (OCDE, 2021).

Los ecosistemas de datos son por lo general escalables y sostenibles. Aunque se conforman alrededor del ciclo de vida de los datos, su dinámica está soportada en la interacción de las capacidades de los agentes para aumentar el aprovechamiento de los datos, en el marco de principios de calidad, seguridad, protección y accesibilidad de los datos (Capello , Gal, Rehof, & Jarke, 2019). Así mismo, emergen alrededor de una propuesta de valor compartido y, aunque no siempre se constituyen en torno a un actor principal, es muy común que su dinámica se fortalezca gracias a la gestión de un actor líder (Adner, 2017).

Dado lo anterior, el termino ecosistema está inspirado en el ámbito de la biología, por medio del cual se describe la interacción entre organismos de diferentes especies y su entorno, en el marco de un sistema integrado (Oliveira & Farias Loscio, 2018). Esta

analogía se ha aplicado a distintas áreas de investigación como ecosistemas de negocio (Moore, 1993,1996), ecosistemas digitales o ecosistemas de software.

Es importante aclarar que los ecosistemas de datos no son plataformas ni soluciones de intercambio de datos. Por el contrario, éstas últimas se definen como herramientas que facilitan la interacción entre los actores, les permiten producir y consumir datos, pero no reemplazan en ningún momento al ecosistema de datos. Esto se explicará más adelante en la sección de componentes del ecosistema.

Debido a lo anterior, las definiciones de ecosistemas de datos confluyen sobre la interacción de actores o stakeholders de diferentes roles y características que, directa o indirectamente, producen, recopilan, almacenan, analizan y consumen datos, apoyados en recursos como infraestructuras físicas o servicios de las tecnologías de la información (TI) (Pollock, Zubcoff, Ding et al). De acuerdo con la propuesta de innovación impulsada por datos (DDI por sus siglas en inglés) de la OCDE (2015), esta interacción debe promover el crecimiento económico y bienestar social, así como el mejoramiento de procesos en materia de innovación y competencia.

2. Principios de los ecosistemas de datos

Los ecosistemas de datos comparten principios comunes para promover el flujo de datos, la creación de dinámicas para el intercambio de datos, y la interacción, integración y participación de los actores que participan en este. La finalidad de dichos principios es garantizar que los datos puedan ser de utilidad para la toma de decisiones y para la generación de beneficios sociales y económicos, dependiendo del tipo de ecosistema. A continuación, se describen los principios generales en materia de ecosistemas de datos.

a) Fácil de encontrar, accesible, interoperable y reutilizable

El desarrollo y la creciente adopción de los principios de fácil búsqueda, accesibilidad, interoperabilidad y reutilización (FAIR) de los datos en los ecosistemas, permite fortalecer la capacidad de acción de quienes hacen uso de los recursos disponibles para el manejo, aumento en el volumen, complejidad y velocidad de creación de datos. La facilidad permite eliminar barreras de acceso a los datos, la accesibilidad permite un adecuado intercambio de los datos, la interoperabilidad promueve la compartición de los datos entre roles y su reutilización hace que los datos que se producen, consumen y analizan sean útiles para la toma de decisiones basadas en datos—de acuerdo con sus condiciones de acceso y uso—.

b) Confianza en el ecosistema

El intercambio de los datos se fortalece en la medida en que se promueve la confianza en su uso, de acuerdo con los límites derivados de las fuentes que lo autorizan. Para ello, se debe establecer un marco básico que contemple el intercambio de datos, la gestión de datos, las técnicas de protección de datos, el procesamiento de datos para proteger la privacidad, la rendición de cuentas y la trazabilidad de los procesos de intercambio.

c) Gobernanza de datos del ecosistema

El principio de gobernanza de datos es el conjunto de normas, políticas, estándares, roles y responsabilidades que permiten esquematizar el flujo de datos en el ecosistema para aumentar el uso y aprovechamiento de estos entre los actores conforme a sus roles. Este principio es clave para promover, en los contextos en que sea necesario, la articulación de acciones y responsabilidades entre los distintos actores, y generar un escenario de confianza para el intercambio y reutilización de datos.

d) Calidad de los datos

Los datos que hacen parte del ecosistema deben cumplir con los criterios de calidad entorno a: la exactitud, la completitud, la integridad, la actualización, la coherencia, la relevancia, la accesibilidad y la confiabilidad de los datos. Los datos de calidad no solo evitan incurrir en conclusiones o análisis erróneos — por causa de datos de mala calidad—, sino que también contribuyen a crear confianza en el ecosistema de datos y a fomentar su reutilización para la toma de decisiones.

3. Características de los ecosistemas de datos

De acuerdo con la revisión de literatura sobre las tipologías de ecosistemas de datos, en términos generales, los ecosistemas tienen características comunes que los describen. A continuación, se hace un breve recuento de las principales tipologías.

Los ecosistemas son una red socio-técnica: Los ecosistemas de datos se definen como una red socio-técnica, en la medida en que las relaciones entre actores están mediadas y soportadas por el uso de las tecnologías. Por lo cual, son redes complejas de individuos y organizaciones, que se interconectan en múltiples capas de redes heterogéneas de software, hardware, personas y procesos (Hussain Shah, Peristeras, & Magnisalis, 2020).

Organización alrededor del ciclo de vida de los datos: Los ecosistemas se constituyen alrededor del ciclo de vida de los datos en el que se incluyen las fases de: creación, procesamiento, almacenamiento, intercambio, uso y análisis de datos, y archivo y preservación. Los roles que surgen en el ecosistema de datos están

estrechamente relacionados con las etapas del ciclo de vida del dato, desde la generación hasta la extracción de valor.

Cooperación-Competencia (coopetención): Los ecosistemas de datos se caracterizan porque los agentes involucrados tienen una relación ambivalente, en la cual compiten y cooperan a la vez. Compiten en la medida en que buscan innovar sus productos o servicios a través del uso de los datos, pero cooperan en la medida en que comparten información y conocimientos (Gelhaar & Otto, *Challenges in the Emergence of Data Ecosystem*, 2020) (Gelhaar, Groß, & Otto, *A Taxonomy for Data Ecosystems*, 2021). La existencia de una relación ambivalente implica que los datos tienen la capacidad de generar valor y este valor puede ser económico (incentivo a la competencia) o social (incentivo a la cooperación).

Creación de valor de los datos: El ecosistema se constituye con el propósito de crear valor a partir de los datos y es el objetivo que comparten los actores que participan en él (Oliveira, Barros Lima, & Farias Lóscio, 2019). Este propósito diferirá según el tipo de ecosistema y su contexto, en el caso gubernamental se puede considerar la administración impulsada por datos, la prestación de servicios públicos y el gobierno abierto. Para estos casos, el potencial para generar valor agregado a partir los datos radica en la capacidad de extraer información significativa y procesable para el bienestar social.

Respecto a los beneficios económicos los ecosistemas de datos contribuyen a la generación de nuevos modelos de negocios, dado que posibilitan la innovación y la creación de bienes y servicios. Así mismo, los ecosistemas de datos se caracterizan porque ningún actor de manera individual estará en capacidad de crear igual valor sin ser parte del ecosistema (Capello, Gal, Rehof, & Jarke, 2019).

Coevolución: La mejora de los procesos relativos a las actividades de un actor contribuye al desarrollo de capacidades de los otros actores lo que, en consecuencia, permite desarrollo del mismo ecosistema. Esto es posible gracias a las dinámicas de cooperación entre actores y a las relaciones que se desarrollan entre ellos, donde un actor es codependiente de otro. Por ejemplo, puede existir un actor que no cuenta con las capacidades para analizar los datos que produce, pero al apoyarse en proveedores de servicios de datos, que tengan fuertes capacidades para este fin, logra extraer el valor de los datos (Oliveira & Farias Loscio, 2018).

Interconectados con otros ecosistemas: Los ecosistemas de datos interactúan con otros ecosistemas, incluso pueden comprenderse como parte de un ecosistema mucho más amplio. Este es el caso del ecosistema de datos y el ecosistema digital, que comparten recursos y habilitadores comunes que facilitan el desarrollo de ambos ecosistemas. O el caso del ecosistema de software que es considerado un subconjunto

del ecosistema digital (Jansen & Cusumano, 2013). Los ecosistemas de datos pueden ser habilitadores muy importantes para otros ecosistemas como el de innovación o el de negocios.

Dinámico: las dinámicas de los ecosistemas de datos se pueden asociar a las teorías de destrucción creativa propuestas por Schumpeter, donde los procesos de innovación permiten la evolución del sistema y de sus agentes que, en este caso, son impulsados por la creación y el uso de los datos. Característica que se relaciona intrínsecamente con las características de coevolución, de cooptación y creación de valor.

Dadas las características de los ecosistemas de datos, el desarrollo de éstos se puede dar en diferentes esferas de la sociedad, la economía o la academia, y pueden responder a estructuras particulares o generales. Por tanto, para modelar un ecosistema se requiere del análisis de experiencias particulares, dado que estos ecosistemas se forman dependiendo de los agentes que participan, sus intenciones (Olivera, Barros, & Farias, 2017) y del nivel de apertura de ecosistema (Gelhaar, Groß, & Otto, A Taxonomy for Data Ecosystems, 2021).

4. Taxonomía de los ecosistemas de datos

Pese a la amplitud y complejidad de los ecosistemas de datos, es posible describir y evaluar su estructura a través de una caracterización taxonómica. Para ello se hace una clasificación de sus estructuras con base en las propuestas de (Gelhaar, Groß, & Otto, A Taxonomy for Data Ecosystems, 2021).

Tabla 1. Taxonomía general del ecosistema

Meta-dimensión	Dimensión	Características			
Económica	Dominio	Científico	Gobierno		Industria
	Propósito	Innovación	Interacción		Transaccional
	Organización	Agente principal	Plataformas	Basado en el mercado	Descentralizado
Técnica	Infraestructura	Centralizado		Distribuido	
	Apertura	Abierto		Cerrado	
Gobernanza	Interdependencia	Vínculo estrecho		Vínculo débil	
	Control	Central		Descentralizado	

Fuente: (Gelhaar, Groß, & Otto, A Taxonomy for Data Ecosystems, 2021)

La Tabla 1 recoge las diferentes variantes taxonómicas que pueden constituir a un ecosistema de datos y, por tanto, reafirma la pluralidad de los ecosistemas. De acuerdo con la clasificación de (Gelhaar, Groß, & Otto, 2021), los ecosistemas de datos se pueden caracterizar por meta-dimensiones, dimensiones y características. A continuación, se explica el significado de cada uno de los componentes, a razón de su función dentro del desarrollo de los ecosistemas:

4.1. Meta-dimensión económica: dimensiones y características

En esta clasificación se define el tipo de modelo económico del ecosistema de datos, en términos de las dinámicas de competencia que caracterizaran las relaciones entre los agentes. Para evaluar dicha meta-dimensión se considera, en primera instancia, el dominio del ecosistema, que hace referencia al entorno donde se gesta y se desarrolla el ecosistema. Puede haber tres tipos de entornos o dominios: el científico, el gubernamental o el privado.

En segunda instancia, se considera el propósito del ecosistema, es decir, el fin que guiará a los agentes que participan y las estrategias que adoptarán par su desempeño. De acuerdo con los autores (Gelhaar, Groß, & Otto, 2021), existen tres tipos de propósitos: por un lado, la generación de una innovación, que puede dar como resultado una nueva forma de generar valor; por otro lado, la interacción de los agentes, donde los participantes compartan información y conocimientos; por último, el ecosistema de datos puede fundamentarse bajo el propósito de facilitar las transacciones entre los agentes, de tal forma que si no existiera el ecosistema los agentes serían incapaces de completar sus transacciones.

En tercera y última instancia, la meta-dimensión económica considera la dimensión organizacional, bajo la cual se estipula el tipo de relaciones, interacciones y organizaciones que se pueden dar entre los agentes. A razón de lo anterior, existen cuatro tipos de organizaciones posibles: el primero cuenta con un agente principal, que marca la pauta para el desarrollo de los demás, pues directa o indirectamente es el encargado de proporcionar la mayoría de los datos que funcionan en el ecosistema; el segundo es un sistema de plataformas, que supone la existencia de plataformas que proveen la infraestructura y los servicios necesarios para usar y compartir datos en el ecosistema, facilitando los procesos de interoperabilidad; el tercero es un modelo basado en el mercado, estructurado a través de una plataforma de mercado de datos; el cuarto parte de un modelo descentralizado, con el principio de que no existe un agente dominante, pero los agentes están relacionados por el propósito común de generar valor. Es importante aclarar que la estructura organizacional del ecosistema puede variar de forma constante, sin embargo, es imposible que dos estructuras funcionen al mismo tiempo (Gelhaar, Groß, & Otto, 2021).

4.2. Meta-dimensión técnica: dimensiones y características

La segunda meta-dimensión refiere a las características técnicas que modelan al ecosistema de datos. Esta meta-dimensión se compone de dos dimensiones: infraestructura y apertura. La primera, refiere a la infraestructura técnica necesaria para compartir datos, esta infraestructura puede presentarse de forma centralizada, es decir a través de un mecanismo preponderante, o distribuida, es decir mediante mecanismos distribuidos entre los agentes que contemplan medidas entre pares. La segunda dimensión, comprende el grado de apertura del ecosistema, donde, por un lado, puede ser abierto y, por ende, cualquier interesado puede acceder o, por otro lado, es cerrado e implica barreras de acceso. Las barreras para acceder se pueden presentar en diferentes formas, por ejemplo, pueden ser técnicas, tecnológicas o legales.

4.3. Meta-dimensión de gobernanza: dimensiones y características

La última de las meta-dimensiones alude a los aspectos y propiedades concernientes a la posesión de los datos y a la dependencia de los agentes dentro del ecosistema. Particularmente, está constituida por dos dimensiones. La primera es la interdependencia, que representa el grado de dependencia que existe entre los agentes que conforman el sistema, donde el vínculo, según los autores (Gelhaar, Groß, & Otto, A Taxonomy for Data Ecosystems, 2021), puede ser estrecho o débil. La segunda es el tipo de control sobre los recursos esenciales del ecosistema de datos, cuyo organigrama puede ser central, es decir controlado por un agente principal, o descentralizado, repartido entre los diferentes agentes relativos al ecosistema.

Como complemento a las meta-dimensiones económicas, técnicas y de gobernanza es posible integrar lo relacionado con recursos humanos y financieros como componentes transversales de los ecosistemas de datos. Atributos como el acceso a fuentes de financiamiento se consideran como factores propios al funcionamiento y sostenibilidad de los ecosistemas.

5. Componentes generales de los ecosistemas de datos

Descritas las características taxonómicas de los ecosistemas de datos, es relevante conceptualizar los componentes que interactúan en el ecosistema. Por tanto, se debe reconocer que los ecosistemas de datos, a nivel general, están constituidos por actores, roles -las actividades que desempeñan los agentes dentro del sistema- y recursos. También están influenciados por los incentivos de tipo organizacional, económico y

social, que motivan la participación de los actores y dinamizan la interacción entre los mismos. En esta sección se hace una descripción de cada componente.

5.1. Actores

Son los participantes del ecosistema, que pueden ser entidades de tipo gubernamental, privado o autónomo y que desempeñan un rol o diversos roles dentro del ecosistema de datos. El Banco Mundial, en su informe de *Data for better lives* de 2021, identifica cinco tipos de actores o stakeholders involucrados: entidades públicas, sociedad civil e individuos, academia, sector privado y organizaciones internacionales y regionales. Donde la interacción entre los actores se genera alrededor del ciclo de vida de los datos. De acuerdo con el Banco Mundial, los ecosistemas deben tener como máximo propósito el acceso a datos de calidad en un entorno seguro, para ser compartido y reutilizado para diferentes propósitos y por todas las partes interesadas. A continuación, se hace una descripción de los actores de acuerdo con lo definido por el Banco Mundial (2021) (ver Tabla 2).

Tabla 2. Actores generales del ecosistema

Actores	Descripción
Entidades públicas	Son entidades públicas productoras, reguladoras y administradoras de datos públicos. Establecen estándares de calidad para la protección, accesibilidad, uso y transacción segura de los datos.
Sociedad Civil e individuos	Son individuos u organizaciones sociales que participan como productores y usuarios de datos, en acción con las entidades públicas y privadas, con el propósito de reducir las brechas digitales y para promover los derechos humanos en el mundo digital.
Academia	Son instituciones académicas (universidades, centros de investigación, instituciones educativas, etc.) que producen estudios de investigación y capacitaciones para formar el conocimiento de las personas en el uso, manejo, producción y consumo de los datos. Adicionalmente puede comportarse como un consumidor de datos para fines investigativos y de actualización de informes y/o actividades de formación de acuerdo con lo que requiere el mercado.
Sector Privado	Son el grupo de empresas, gremios y emprendimientos que producen, consumen y proveen servicios analítica de datos, tecnológicos, infraestructura TI y datos a otros actores. Así mismo, son las empresas del ecosistema que permiten el mejoramiento de las decisiones y los procesos comerciales.
Organizaciones internacionales y regionales	Son aquellos actores que promueven e impulsan el desarrollo de políticas, proyectos e iniciativas de datos que contribuyen al desarrollo económico de los países.

Fuente: Elaboración propia

5.2. Roles

Los roles son las actividades y funciones que desempeñan los actores que constituyen el ecosistema de datos e interactúan en él. El rol de los agentes dentro del ecosistema no es excluyente, por lo que pueden desempeñar diferentes funciones a la vez. Es por esta razón que, dentro de los ecosistemas de datos, se pueden dar procesos de integración vertical, donde un agente puede tener varios roles dentro del ecosistema.

A nivel de roles, la (OCDE, 2015) estipula que el ecosistema de datos se estructura a partir del ciclo de vida de los datos. A partir de este proceso se reconocen cinco tipos de roles, que son: 1) los proveedores de servicio de Internet; 2) los proveedores de infraestructura TI; 3) proveedores, consumidores e intermediarios de datos; 4) proveedores de analítica, conformado por proveedores de hardware y software y consultores; 5) empresarios de datos, conformados por empresas emergentes (startups), empresarios y entidades públicas. Es importante reconocer que los proveedores de servicios de Internet e infraestructura TI posibilitan el desarrollo de todo el sistema, dado que proveen las condiciones básicas que estructuran el ecosistema y sus funciones, por lo cual su rol es completamente transversal en el sistema.

Como complemento a los roles propuestos por la OCDE (2015), se acude a la información recopilada en el artículo académico “Investigations into Data Ecosystems: a systematic mapping study” (Oliveira, Barros Lima, & Farias Lóscio, 2019) que propone roles complementarios. Principalmente, el rol más destacable es el encargado de las políticas, leyes y normas, que incentiva la participación de los agentes y puede llegar a regular las funciones de los actores en el ecosistema. Dependiendo de las particularidades y amplitud del ecosistema hay diferentes agentes que pueden fungir este rol: gobierno, institutos de estandarización o el agente central del ecosistema. Así mismo, es importante destacar que esta función normativa es transversal, dado que ayuda a regular y dinamizar el ecosistema.

En grandes rasgos existen seis roles principales. Sin embargo, cada uno de ellos contiene sub-roles. A continuación, se hace una descripción de los roles y sub-roles alrededor de las funciones que desempeñan (Ver Tabla 3 .)

Tabla 3 .Descripción de roles de datos en el ecosistema

Roles	Descripción
-------	-------------

Proveedores de servicios de Internet	Son los encargados de asegurar el despliegue y la conexión de Internet en los territorios y a nivel nacional. Por tanto, cumplen dos funciones específicas: el despliegue de infraestructura necesaria para la prestación del servicio y la comercialización y garantía de conexión a la red (OCDE, 2015).
	Este rol lo desempeñan los operadores de red móvil (MNO por su sigla en inglés), los operadores móviles virtuales (MVO por su sigla en inglés), los operadores de red fija, los operadores del servicio portador, los proveedores de tránsito IP, las compañías de infraestructura pasiva, las compañías torres, así como los proveedores de redes de distribución de contenidos y los proveedores de servicios de contenido y aplicaciones.
Proveedores infraestructura TI	Son los encargados de proveer los servicios necesarios para el uso y la explotación de los datos, en términos de manejo, seguridad, almacenamiento e intercambio. Por tanto, sus funciones específicas están asociadas a la provisión de hardware y software y, en consecuencia, servicios Software as a Service (SaaS), Infrastructure as a Service (IaaS) y Plataforms as a Service (PaaS) (OCDE, 2015; Oliveira, Barros Lima, & Farias Lóscio, 2019)
Proveedores de datos	Son los encargados de procesar y transformar los datos de tal forma que adquieran una estructura que les confiera valor. Dentro de esta categoría existen un número considerable de roles subyacentes, que consisten en:
	1. Corredores de datos, encargados de compilar información para su comercialización. Para desempeñar tal función, los corredores de datos emplean roles para almacenar, agregar datos de diferentes fuentes, armonizar datos, actualizar la información, publicar la información y mantener el control de los datos (Immonen, Palviainen, & Ovaska, 2014; Oliveira, Barros Lima, & Farias Lóscio, 2019)
	2. Proveedores de datos abiertos: Este rol se encarga de proveer datos y también de hacer uso de ellos (OCDE, 2015).
Consumidores de datos	Este rol refiere a los usuarios y consumidores de datos que, en cierta medida, también son proveedores de datos. Lo anterior teniendo en cuenta que, dentro de las actividades que realizan también generan datos de forma cotidiana. Por tanto, los consumidores se pueden considerar, a su vez, como titulares y productores de datos (OCDE, 2015; Oliveira, Barros Lima, & Farias Lóscio, 2019)
Beneficiarios de datos	Personas u organizaciones que se benefician del ecosistema de datos, tomando mejores decisiones usando productos y servicios.

Proveedores de analítica	Son los encargados de proveer servicios de analítica de datos para los consumidores. Este ejercicio consiste en transformar los datos en información útil para los consumidores que, precisamente, acuden a los proveedores de analítica porque no poseen los conocimientos o recursos suficientes para desarrollar dicho proceso. Por tanto, los roles que conforma a estos proveedores son: analista de datos, visualizador de datos, y extractor y transformador de datos [3]. (Lidman, Kinnari, & Rossi, 2016; (Oliveira, Barros Lima, & Farias Lóscio, 2019)
Emprendimientos impulsados por analítica de datos	Los empresarios de datos generan procesos de innovación en sus negocios a partir de los datos y de la analítica de datos, con el fin de mejorar funciones asociadas a su actividad de negocio. Este rol está dividido en dos funciones:
	<ol style="list-style-type: none"> <li data-bbox="641 808 1494 1092">1. Exploradores de datos: Cumplen una función semejante a los corredores de datos, con la diferencia que los datos que compilan y transforman tienen una característica bien definida, que se deriva de las necesidades de un consumidor o del objetivo de un negocio (OCDE, 2015). Sin embargo, su parecido con los corredores de datos implica que requiere de las mismas funciones para su funcionamiento. <li data-bbox="641 1092 1494 1486">2. Generadores de plataforma: Generan plataformas basadas en datos que ayudan al desenvolvimiento de sus actividades de negocio principales. Por tanto, los datos, bajo este escenario, son resultado de las actividades de negocio principal que desarrollan los agentes y su uso sirve para mejorar el desempeño de dichas actividades (OCDE, 2015). Para el desempeño de este rol se requieren las mismas funciones de los exploradores de datos y los corredores de datos, dado que internamente se debe recopilar y transformar los datos para generar información útil que contribuya al desempeño de las actividades de negocio.
Responsables de políticas, leyes y normas	Bajo este rol se construyen las políticas y normativas que incentivan el desarrollo del ecosistema, pero que, a su vez, se encargan de controlar las actividades de los agentes que lo constituyen (Oliveira, Barros Lima, & Farias Lóscio, 2019). Por tanto, bajo el desempeño de este rol se establece el marco regulatorio y normativo (OCDE, 2015).

Fuente: Elaboración propia con información de [(Gama & Bernadette, 2014); (Immonen, Palviainen, & Ovaska, 2014); (OCDE, 2015); (Oliveira, Barros Lima, & Farias Lóscio, 2019)]

Cómo se puede observar en la

Tabla 3., se resumen las actividades de los seis roles principales y, a su vez, los roles subyacentes a su desempeño. Las actividades que se realizan dentro del ecosistema requieren de diferentes habilidades y conocimientos, para el cumplimiento de los propósitos de los agentes que participan en el ecosistema. Así mismo, las características de los roles demuestran su codependencia para la contribución y desarrollo del ecosistema y la consecución de propósitos, reafirmando los principios de cooperación y competición.

5.3. Recursos

Son todos aquellos recursos necesarios para dinamizar al ecosistema de datos, y para apoyar el ciclo de vida de los datos. Entre estos se encuentran la infraestructura TI, infraestructura de conectividad y los recursos financieros. Los recursos pueden ser compartidos por los actores del ecosistema de datos, y se entienden como aquellos productos y capacidades que permiten producir, proveer, curar, y consumir datos.

En un ecosistema de datos los actores no necesitan tener sus propios recursos, y dada la connotación de red del ecosistema, pueden consumir, alquilar o contratar diversos recursos para explotar sus datos. La simbiosis de los recursos puede influenciar fuertemente la sostenibilidad del ecosistema por eso pueden abordarse desde el concepto de habilitadores. Algunos de los recursos utilizados e intercambiados en el ecosistema de datos, provienen de ecosistemas más amplios como el ecosistema digital, en el que se proveen recursos de infraestructura en conectividad a Internet.

La connotación de red también permite disponer de recursos compartidos y dinámicos que provean datos para generar beneficios al sector público, sector privado, la academia y la ciudadanía. En el caso del Gobierno, este es un actor que proporciona recursos para la sostenibilidad del ecosistema a través de la provisión de bienes públicos digitales, que permiten que el acceso a recursos por parte de otros actores del ecosistema sea menos costoso, pero también mediante la definición de políticas y de un marco normativo. Lo anterior, dado que, los denominados recursos compartidos (sistemas de información, arquitecturas, modelos, políticas públicas, entre otros) provistos por parte de las entidades públicas, facilitan el intercambio y promueven el uso y explotación de datos. (Departamento Nacional de Planeación, 2018).

5.4. Incentivos y barreras

Para estimular la participación de los agentes dentro de los ecosistemas de datos es importante generar incentivos. Por ejemplo, la producción, el intercambio y la explotación de los datos necesitan de inversiones que permitan sostener el ecosistema,

por lo cual se deben promover recursos y capacidades que generen beneficios a los actores del ecosistema.

En otras palabras, los incentivos deben motivar a los actores a participar en el ecosistema de datos a través de modelos de desarrollo sostenible, que permiten mantener los datos actualizados y accesibles y, dependiendo del tipo de ecosistema de datos, crear actividades comerciales y aplicaciones que sostengan las iniciativas del ecosistema de datos. (Oliveira, Barros Lima, & Farias Lóscio, 2019).

Por otra parte, los actores de los ecosistemas de datos se pueden enfrentar a barreras de tipo tecnológico, de capacidades o de gobernanza que podrían limitar o restringir su participación en el ecosistema, su interacción con otros actores y el intercambio y provisión de recursos. Por ejemplo, un entorno de baja confianza entre agentes podría convertirse en una barrera para el uso e intercambio de datos entre el sector público y privado.

6. Creación de valor social y económico

Los ecosistemas de datos son entornos en los cuales se crean sinergias de distintos actores alrededor del aprovechamiento de datos en las diferentes fases del ciclo de vida de los datos, por lo que es posible crear oportunidades entre organizaciones para la generación de conocimiento, investigación e innovación (Adner, 2006). Esta alianza entre actores potencializa la creación de valor a partir de los datos, tanto en el ámbito público como en el privado.

De acuerdo con la OCDE, la medición del valor de los datos es posible mediarla a través de cuatro enfoques. (OCDE, 2020). El primero, es un enfoque de costos, es decir que la medición se realiza a partir de los gastos en los que se incurre para almacenamiento y gestión de los datos. El segundo enfoque, mide los ingresos generados por emprendedores basados en datos y empresas de analítica de datos. El tercer enfoque mide las valoraciones de la empresa a partir de los datos. El cuarto y último mide los vínculos potenciales entre los flujos comerciales y los flujos de datos, comparando diferentes medidas del comercio de “productos entregables digitalmente”, como el volumen de datos/flujos de tráfico internacionales relacionados con el comercio de datos e intercambio de información.

7. Ecosistemas de datos hacia el bien común

Definir el propósito de los ecosistemas de datos implica un reto, en la medida en que supone la generalización de procesos que puedan ser aplicables a distintos contextos. Lo anterior, dado que, como se ha mencionado anteriormente, existen diferentes tipos

de ecosistemas de datos que involucran diferentes recursos, propósitos, agentes y roles. Independientemente del contexto en el que surjan los ecosistemas de datos, estos deberían responder a la necesidad de la creación de valor derivado del uso de los datos y al cumplimiento de los derechos humanos de los agentes involucrados, en torno a la seguridad de la información, la protección de datos personales y la gestión ética de los datos.

Con base en estas características, en la literatura se han identificado, de forma generalizada, la propensión por el desarrollo de ecosistemas de datos abiertos. Lo anterior, dado que habilitan la consecución de propósitos comunes entre agentes (OCDE, 2015; United Nations, 2020; Department for Digital, Culture, Media & Sport, 2020). La propuesta de un ecosistema de datos abiertos, que pueda funcionar de forma generalizada, se sustenta en la idea de que los datos se desempeñan como bienes de capital, puesto que permiten crear o innovar bienes y servicios. Por tanto, esta funcionalidad implica gran utilidad para todos los sectores de la sociedad y de la economía.

Dado que, por un lado, para los empresarios y el sector privado permite generar valor económico, y en esos términos el desarrollo competitivo de las actividades económicas. Por otro lado, es de gran utilidad para la sociedad y el sector público, dado que posibilita los procesos de creación de valor social, a través de la mejora e innovación de los bienes y servicios provistos por el Estado, y el aumento de la participación ciudadana.

Dado lo anterior, el uso y la explotación de los datos fomenta y permite la cooepetencia entre los sectores que componen las diferentes funciones de la sociedad. En este sentido, y dependiendo del contexto en el que surja el ecosistema, se debería propender por la creación de ecosistemas abiertos, sustentados en el desarrollo colaborativo y competitivo (cooepetencia) que permite el uso de los datos en torno a un propósito común. (World Economic Forum, 2021). En este contexto, se destaca la importancia de la "filantropía de datos", entendida como la manera mediante la cual el sector privado o la sociedad civil, comparte datos para obtener beneficios sociales. Este tipo de iniciativas pueden apoyar una acción política más oportuna y específica.

A pesar de esto, la filantropía de datos tiene un reto: buscar un método para que las empresas puedan compartir el análisis de sus datos con el sector público, protegiendo al mismo tiempo la privacidad de sus clientes y su ventaja sobre la competencia. Desde el punto de vista arquitectónico, a veces puede no ser claro qué forma puede adoptar la filantropía de datos. Lo que sí es claro, es que los propietarios de los recursos de Big Data del sector privado pueden poner este valioso recurso a disposición del público (UN Global Pulse, 2011).

Para las entidades del sector público, la apertura de sus datos también puede suponer un reto, cuando se trata de información sensible — que compartición afectaría la intimidad del Titular o cuyo uso indebido puede generar su discriminación—. Las entidades gubernamentales, normativamente tienen responsabilidades legales (que pueden llegar a ser civiles, penales, administrativas) sobre la protección que deben garantizar sobre la información que manejan de los ciudadanos. En ese sentido, para compartir este tipo de información se deben tener en cuenta las políticas de seguridad y privacidad, y los requerimientos para la anonimización del dato en su ciclo de vida.

A pesar de lo anterior, los procesos de apertura requieren de acciones que garanticen su correcto desarrollo, dado que los procesos de innovación basada en datos suponen cambios drásticos y a gran velocidad, que implican retos económicos y sociales que requieren de un enfoque participativo por parte de todo el gobierno. De acuerdo con (OCDE, 2015), los responsables de política pública encargados de promover un marco para el aprovechamiento de los datos en las distintas economías deben afrontar un gran reto principal asociado a lograr el balance adecuado entre los beneficios sociales de la "apertura" y las legítimas preocupaciones de los individuos y las organizaciones por dicha apertura.

8. Metodologías de caracterización de ecosistemas de datos

Para mejorar la comprensión de los ecosistemas de datos y llegar a un marco común, es necesario crear mecanismos que puedan caracterizarlos. Dado que, a partir de ello, se pueden diseñar lineamientos que permitan mejorar la toma de decisiones, de tal forma que garanticen el uso eficiente de la infraestructura de datos y, con ello, se maximice el valor que se deriva del uso, reúso e intercambio de datos. En esa medida, es importante mapear el ecosistema de datos en términos de los actores que participan en él, los componentes que lo integran y a su vez, caracterizar los procesos de intercambio y creación de valor.

Las metodologías de caracterización plantean un ejercicio participativo con actores de los ecosistemas de datos, que permita comprender las relaciones de interacción entre los actores y el intercambio y uso de datos y recursos que permite generar valor. Para ello, se debe identificar los administradores y usuarios de datos clave, las relaciones entre ellos y los diferentes roles que desempeñan. Las metodologías del Instituto de Datos Abiertos (ODI) y la Comisión Europea permiten identificar: componentes del ecosistema (actores, recursos, roles, activos de información), flujos de datos en el

ecosistema, intercambios de valor, gobernanza del ecosistema, barreras e incentivos de los actores para hacer parte del ecosistema⁷.

II PARTE: PROPUESTA DE CARACTERIZACIÓN DEL ECOSISTEMA DE DATOS EN COLOMBIA DESDE UN ENFOQUE DE GOBIERNO

Después de abordar la definición de los ecosistemas de datos, sus componentes, características y principios, en esta sección se presenta la propuesta de caracterización del ecosistema de datos para Colombia desde un enfoque de gobierno. Esta propuesta, además de la revisión de referencias bibliográficas y documentos técnicos, incluye insumos recolectados a través de la metodología presentada en la Guía de Caracterización de los ecosistemas de datos, que fue elaborada por el Departamento Nacional de Planeación (DNP) y acompaña este documento.

Es importante precisar que este abordaje se encamina hacia la comprensión del ecosistema de datos de Colombia desde un enfoque de análisis gubernamental a nivel país. Por lo cual, no tiene en consideración ecosistemas de datos de carácter sectorial, como los ecosistemas de datos para la investigación o de tipo industrial.

El enfoque de análisis gubernamental tiene como base la gobernanza del ecosistema de datos desde un marco de políticas públicas, en el que hay una actuación del gobierno para orientar, dinamizar e incentivar la cooperación y competición de diferentes actores en torno a los ecosistemas de datos y en favor de diversos propósitos. Aunque el análisis de caracterización se concentra en el rol del gobierno y su marco de análisis institucional, no desconoce los procesos de organización autónoma de tipo industrial e investigativo que han constituido ecosistemas de datos alrededor de propósitos comunes de interés del sector privado o la academia.

También es importante aclarar qué, aunque el análisis es de tipo gubernamental, no se orienta estrictamente a la comprensión de un ecosistema de datos abiertos, teniendo en consideración que el contexto actual de la política pública en Colombia para el aprovechamiento de datos no se concentra unívocamente en los datos generados por el sector público, sino que incluye los datos generados por los ciudadanos y el sector privado. Por lo anterior, esta caracterización comprenderá la interacción de actores para el aprovechamiento de datos del sector público, privado, la academia, los ciudadanos y organizaciones internacionales y regionales.

⁷ Para más información revisar la Guía para la Caracterización de ecosistemas de datos.

1. Contexto general del ecosistema de datos en Colombia desde un enfoque de gobierno

En Colombia, durante los últimos años se ha avanzado en la consolidación de un marco de política pública para el desarrollo de ecosistemas de datos públicos, como es el caso de datos abiertos, datos estadísticos y datos geoespaciales. Dado que el ecosistema de datos en Colombia está constituido por varios tipos de ecosistemas de datos, en esta sección se presenta un recuento de las iniciativas institucionales que han impulsado y dinamizado los ecosistemas de datos de carácter gubernamental.

El ecosistema de datos en Colombia inició su configuración a partir de la producción de información estadística orquestada por el Departamento Administrativo Nacional de Estadística (DANE) en el año 1953. Si bien los usuarios de esta información han hecho parte del sector privado, la sociedad civil y la academia, ha sido a través del diseño de políticas públicas desde el gobierno que se ha visibilizado la necesidad de contar con información estadística de calidad para tomar decisiones. Los datos de las encuestas y registros administrativos fueron los primeros en ser valorados como activo central de la planeación y el seguimiento de políticas públicas, y los primeros conjuntos de datos en ser empleados para la construcción de análisis en distintos campos y por diversos actores.

Aunque el inicio del ecosistema de datos se configuró principalmente alrededor de los datos estadísticos, con la Ley 1712 de 2014 “Transparencia y Acceso a la Información Pública” los datos abiertos⁸ por parte de las entidades públicas empezaron a tener mayor preponderancia en el país. Respecto a ello, el Ministerio de las Tecnologías de la Información y las Comunicaciones (MinTIC) ha sido el principal responsable del diseño, la implementación y la promoción de políticas de datos abiertos en el país, realizando un trabajo articulado con otras entidades públicas, la academia y la sociedad civil como, por ejemplo, la Secretaría de Transparencia de la República, el Departamento Administrativo de la Función Pública (DAFP) y el DNP.

En materia de datos estadísticos y datos abiertos, se han definido⁹ ecosistemas de datos que reflejan el flujo e interacción de los actores como consumidores y productores de datos. En el caso del DANE se destaca el ecosistema de datos del Sistema Estadístico

⁸ De acuerdo con la ley 1712 de 2014 sobre transparencia y acceso a la información pública nacional, se define que los datos abiertos son aquellos que datos primarios o sin procesar que están bajo custodia de las entidades públicas o privadas que cumplen fines públicos y están a disposición de cualquier ciudadano para que sean reutilizados y empleados para crear servicios.

⁹ Ver Sistema Estadístico Nacional (<https://www.dane.gov.co/index.php/plan-estadistico-nacional-pen>) y estrategia de datos abiertos ([Datos Abiertos - Gobierno Digital \(gobiernoenlinea.gov.co\)](https://gobiernoenlinea.gov.co))

Nacional, cuya articulación se refleja en el Plan Estadístico Nacional¹⁰ 2023-2027. Así mismo, se resalta el ecosistema de datos geográficos; por medio del Artículo 2.2.3.5.6. del Decreto 2404 de 2019 se creó la Infraestructura Colombiana de Datos Espaciales (ICDE) como la estrategia de gobierno diseñada para integrar y disponer información geográfica que promueva la creación de valor.

La creación de la ICDE fue una respuesta a un proceso regional de consolidación de la Infraestructura de datos Espaciales (IDE) en los países de América Latina, y una articulación entre estos que inicio a partir la Sexta Conferencia Cartográfica Regional de las Naciones Unidas para las Américas en 1996. La ICDE es conformada por todas las entidades productoras y usuarias de información geográfica o que poseen o son usuarias de registros administrativos que pueden ser utilizados para la producción de información geográfica; por medio de su modelo de gobernanza se consolidan los roles y las responsabilidades de los actores vinculados y se definen los procesos para la gobernanza de este tipo de información.

De acuerdo con lo mencionado anteriormente, en el país se han venido desarrollando ejercicios para la caracterización de ecosistemas de datos. Sin embargo, el cambio de paradigma de los datos y su valor social y económico en el marco de la digitalización genera la necesidad de plantear un marco de entendimiento del ecosistema de datos a nivel nacional, que no esté referido únicamente a datos abiertos o estadísticos.

La dinamización de este ecosistema de datos en materia de gobernanza institucional o gubernamental ha sido impulsada por el Gobierno nacional a través de la creación e implementación de políticas públicas que promuevan el aprovechamiento de datos como el CONPES 3920 de 2018 - Explotación de datos (Big data), el CONPES 3975 de 2019 - Transformación digital e Inteligencia Artificial y el CONPES 4023 de 2021 - Política para la Reactivación Económica. En el marco de la implementación del CONPES 4023 se estructuró la Resolución 460 de 2022 del MinTIC, que expide el Plan Nacional de Infraestructura de Datos (PNID) 2022 - 2025.

El PNID tiene como propósito fortalecer las condiciones para aprovechar los datos y definir un modelo de gobernanza de la infraestructura de datos para articular las acciones de los distintos actores del ecosistema de datos, que se formalizó mediante el Decreto presidencial 1389 de 2022. Este decreto definió el Comité Nacional de Datos como instancia de coordinación institucional para el desarrollo de la infraestructura de datos en Colombia, así como el Coordinador Nacional de Datos, y el rol de los administradores de datos en las entidades públicas.

¹⁰ Es la hoja de ruta para la producción estadística en los próximos años. Contiene estrategias y acciones orientadas a la implementación de proyectos de gran impacto estadístico y que están alineados con los objetivos de desarrollo nacional y territorial.

Por último, cabe mencionar que en el Plan Nacional de Desarrollo 2022-2026 “Colombia, potencia mundial de la vida” se incluyó el componente “Acceso, uso y aprovechamiento de datos para la transformación social”, que tiene como propósito impulsar iniciativas de política pública para desarrollar la interoperabilidad como bien público digital y definir estrategias de datos sectoriales que permitan aumentar la disponibilidad de datos en los sectores administrativos del país.

Los recuentos normativos y de política anteriormente presentados ilustran el avance del país hacia el desarrollo de ecosistemas de datos públicos, que a su vez han potenciado el surgimiento de ecosistemas de datos con fines científicos e industriales. Teniendo en cuenta que el alcance este documento es la caracterización del ecosistema de datos en Colombia desde un enfoque gubernamental, en la siguiente sección se presenta el marco de entendimiento del ecosistema de datos en Colombia desde una perspectiva de gobierno. Dicha propuesta se define con el objetivo principal de evidenciar la relación de la infraestructura de datos del Estado con los distintos ecosistemas de datos públicos, y así ayudar en la comprensión de las dinámicas de interacción de recursos, datos y tecnología.

Cabe mencionar que el marco de entendimiento que se presenta a continuación tiene vigencia en el contexto actual, por lo cual, las características generales del ecosistema de datos gubernamental pueden estar sujetas a cambios dado el dinamismo propio de los ecosistemas de datos. Así mismo, teniendo en cuenta la amplitud y complejidad de los ecosistemas, puede ser que existan agentes, recursos o características del ecosistema que no hayan sido incluidas en este documento.

2. Marco de entendimiento del ecosistema de datos en Colombia desde un enfoque de gobierno

El marco de entendimiento que es planteado, parte de la visión de ecosistema impulsado desde el sector público a través de políticas, iniciativas y lineamientos para aumentar el aprovechamiento de los datos en sus diferentes propósitos de tipo estadístico, geoespacial, abiertos, entre otros. A través de los recursos provistos por el ecosistema de datos y sus sub-ecosistemas se genera una sinergia entre los actores participantes, dado que se configuran escenarios de interacción necesarios para el intercambio y reutilización de los datos.

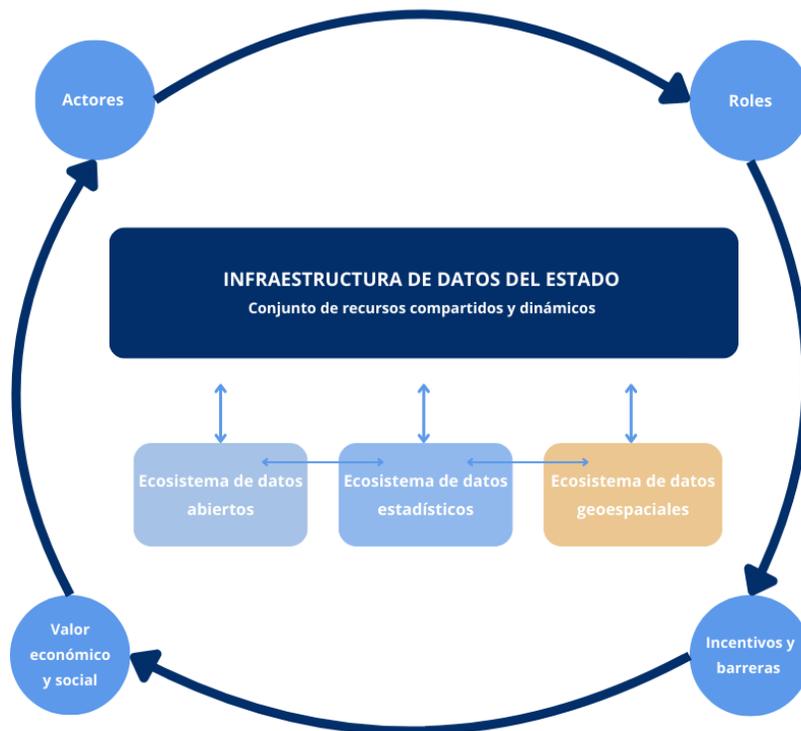
Lo anterior, pone en consideración que el ecosistema planteado comprende una aproximación desde la visión pública y la interacción que se genera entre diferentes sub-ecosistemas de carácter público. Este planteamiento también permite comprender que existen dinámicas bajo las cuales los agentes del ecosistema se relacionan con otros ecosistemas como, por ejemplo, el ecosistema de digital debido a que existen relaciones

simbióticas en el uso de recursos compartidos como la infraestructura de Tecnologías de Información (TI) y la conectividad a Internet, que son elementos claves para el desarrollo de plataformas de uso, almacenamiento e intercambio de datos.

Desde un enfoque de gobierno, es importante entender que el ecosistema de datos gubernamental se estructura a través de la intersección de la Infraestructura de datos del Estado colombiano, definida a través del PNID, con otros ecosistemas de datos públicos como lo son el de datos abiertos, espaciales y estadísticos. A su vez, este gran ecosistema se estructura a partir de sus actores, sus roles, sus incentivos y barreras, la creación de valor social y económico y, en general, de los principios y recursos compartidos que se intercambian en el flujo de datos que es generado.

En la Figura 1 se presenta el marco de entendimiento del ecosistema de datos para Colombia desde un enfoque gubernamental.

Figura 1. Marco de entendimiento ecosistema de datos de Colombia



Fuente: Elaboración propia

Ecosistema de datos de Colombia desde un enfoque de Gobierno

Como se observa en la Figura 1 el ecosistema de datos de Colombia se encuentra soportado por la infraestructura de datos del Estado, que actúa como articulador con otras infraestructuras y ecosistemas como lo es el ecosistema de datos estadísticos,

dinamizado a través del Sistema Nacional de Estadística (SEN), a cargo del DANE, la Infraestructura Colombiana de Datos Espaciales a cargo del IGAC y el ecosistema de datos abiertos abordado a través de la estrategia de datos abiertos y el portal de datos abiertos del MinTIC.

La propuesta plantea que la infraestructura de datos del Estado definida por una serie de recursos compartidos¹¹ (políticas, normativas, activos de información, arquitecturas, estándares y lineamientos, recursos tecnológicos y talento humano) potencia y habilita el acceso, disposición, reutilización y aprovechamiento de datos de los ecosistemas de datos públicos, utilizando como pilares esenciales la gobernanza y la interoperabilidad (Ver Tabla 4). En consiguiente, estos recursos compartidos como son, por ejemplo, las políticas de Big Data, las normativas relacionadas con la privacidad de los datos, o los registros de información de las entidades públicas, son a su vez aprovechados por ecosistemas públicos como es el caso del ecosistema de datos estadístico.

Tabla 4 . Componentes de la infraestructura de datos

Componentes de la infraestructura de datos	Descripción
Gobernanza	Conjunto de normas, políticas, estándares, roles y responsabilidades que permiten potenciar el uso y aprovechamiento de la infraestructura de datos.
Datos	Los datos son el activo más importante de la infraestructura de datos ¹² . Se incluyen datos maestros, datos de referencia, datos transaccionales, datos abiertos, datos estadísticos, datos geográficos.
Aprovechamiento de datos	El aprovechamiento de datos requiere no solamente su adecuada gestión durante su ciclo de vida, sino también la implementación de técnicas, herramientas y metodologías de analítica y explotación de datos, que permitan transformarlos en información valiosa para la toma de decisiones tanto en el gobierno, como en el sector privado, la academia y la sociedad civil.
Interoperabilidad	La interoperabilidad entre sistemas de información es el eje central de la infraestructura de datos, dado que habilita la reutilización e intercambio de datos entre distintos actores.

¹¹ Para precisar la explicación de recursos compartidos de la Infraestructura de datos en la tabla 5 se presentan los componentes que la constituyen. Estos han sido definidos mediante la Resolución MinTIC 460 de 2022, que establece el Plan Nacional de Infraestructura de datos y su hoja de ruta

¹² Como elemento que hace parte del componente en materia de datos se encuentran los metadatos que son generados por las entidades y las organizaciones.

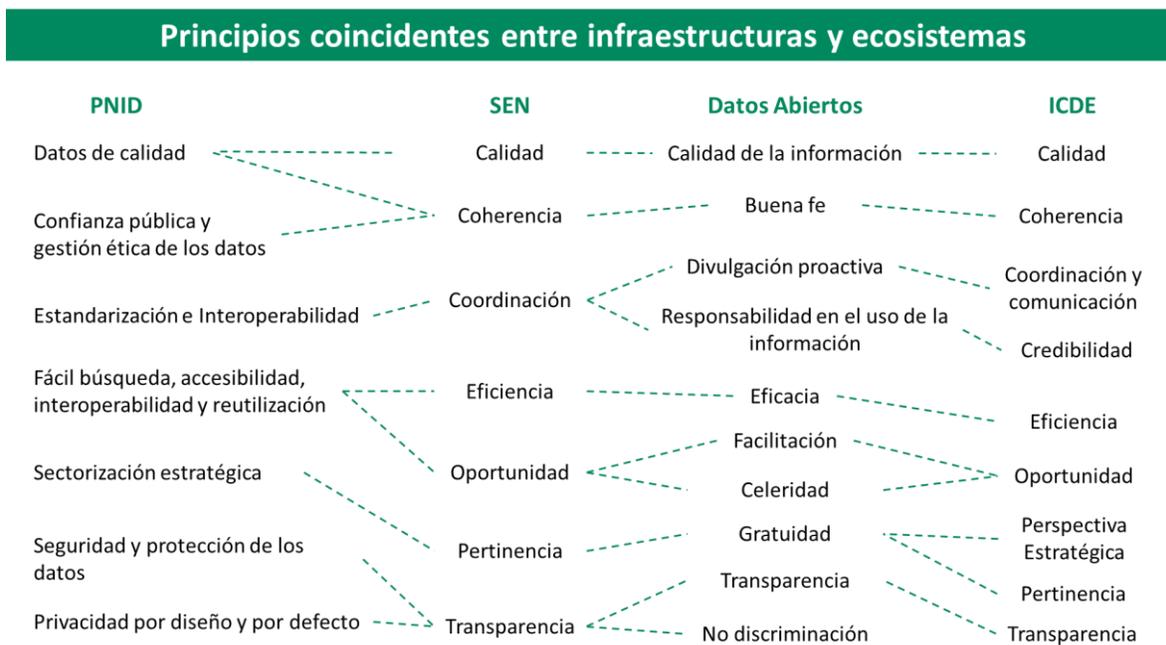
	Este componente está soportado por el desarrollo de políticas, normativas, infraestructura tecnológica y lineamientos técnicos que permiten eliminar las barreras técnicas, institucionales y operativas para el intercambio de datos.
Seguridad y privacidad de los datos	Las infraestructuras de datos están soportadas por procesos tecnológicos, normativos y de política en los que se debe garantizar la seguridad digital y privacidad de los datos.
Instrumentos técnicos y tecnológicos	Los instrumentos técnicos y tecnológicos se refieren aquellos recursos necesarios para desplegar la infraestructura de datos y habilitar su aprovechamiento por parte de distintos actores del ecosistema de datos. Por tratarse de una infraestructura de carácter público, los recursos que aquí se mencionan son de carácter compartido.

Fuente: Elaboración propia tomada del Plan Nacional de Infraestructura de Datos y de las relatorías adelantadas en el marco del Modelo de Gobernanza de la Infraestructura de Datos.

La implementación de la Infraestructura de datos del Estado se consolida en el PNID y su hoja de ruta, en el que se proponen acciones de articulación con el SEN y la ICDE. Esta articulación precisa un análisis de principios coincidentes en las normas que orientan el funcionamiento de estas infraestructuras y ecosistemas. A mencionar se destaca el principio de calidad de los datos y de la información que es generada, la promoción de la coherencia y gestión ética de los datos, la coordinación relacionada a promover la eficiencia y eficacia de los datos en las infraestructuras para fortalecer la interoperabilidad y estandarización de procesos en torno al ciclo de vida de los datos, la definición de estrategias de sectorización para la pertinencia en el desarrollo de proyectos e iniciativas basadas en datos, y la garantía de brindar información y datos con transparencia para su aprovechamiento y reutilización.

La Figura 2 presenta la identificación realizada a partir de los principios coincidentes entre los ecosistemas e infraestructuras que son soportados desde la infraestructura de datos del Estado.

Figura 2. Principios coincidentes entre infraestructuras y ecosistemas



Fuente: Elaboración propia con base en (DANE, 2022) & (ICDE, 2021)

Como ha sido mencionado existen relaciones simbióticas entre infraestructuras y ecosistemas, así como principios coincidentes como los que se presentan en la Figura 2. Lo anterior, permite identificar que los recursos, que son provistos desde la infraestructura de datos del Estado, permiten establecer relaciones a partir del intercambio entre los distintos actores. Es por ello, que la connotación de red del ecosistema de datos reconoce que la dinámica del valor del ciclo de los datos depende no solo de los recursos propios que son provistos por los actores, sino también por los recursos compartidos que son provistos para su uso. Por lo cual, el beneficio generado se distribuye entre el sector público, sector privado, la academia y la ciudadanía en la proporción con la cual han sido usados y compartidos dichos recursos¹³.

3. Características generales del ecosistema de datos en Colombia

De manera general, el ecosistema de datos de Colombia, desde un enfoque de gobierno, es un sistema sociotécnico en el que participa el Gobierno nacional y territorial, las empresas del sector privado, la academia y la sociedad civil que, en conjunto, aportan a la creación de valor económico y público. Es un ecosistema abierto en la medida que

¹³ Para tener un entendimiento sobre cómo estos recursos compartidos se relacionan entre los ecosistemas y las infraestructuras se sugiere dirigirse a la sección Recursos del ecosistema de datos del documento.

fomenta la competencia en todos los actores del ecosistema y a la vez fomenta las condiciones para la creación de nuevos bienes, servicios y modelos de negocio.

En esta sección se hace una breve descripción del ecosistema de datos en Colombia, teniendo en consideración la taxonomía presentada en la parte I del documento, relacionada con la definición de meta-dimensiones de economía, tecnología y gobernanza. Para esta descripción se realiza una adaptación de este marco de taxonomía, teniendo en cuenta las dinámicas propias del país y del desarrollo del ecosistema. La Tabla 5. presenta las características del ecosistema de datos en Colombia.

Tabla 5. Taxonomía del ecosistema de datos para Colombia

Meta-dimensión	Dimensión	Característica
Económica	Dominio	Gubernamental
	Propósito	Innovación Interacción Transaccional
	Estructura organizativa	Agente principal
Infraestructura tecnológica	Apertura	Abierto
	Infraestructura	Distribuido
Gobernanza	Control	Descentralizado
	Interdependencia	Vínculo estrecho

Fuente: Elaboración propia

Meta- dimensión: Económica

- **Propósito:**

Para el caso del ecosistema de datos en Colombia es importante destacar que la meta-dimensión económica evalúa los propósitos del ecosistema de datos en torno a la generación de valor y conforme a los propósitos e intereses del ecosistema y de los agentes. Al respecto, el Gobierno, como participante del ecosistema de datos, tiene el rol y el propósito de dinamizar la generación de valor público y económico y, con ello, garantizar las dinámicas de coopección entre los agentes, Por lo anterior, entre los propósitos del ecosistema de datos se incluye, la innovación basada en datos, la interacción entre los actores privados públicos y sociedad civil, y los fines

transaccionales que permiten el intercambio de datos y recursos entre los agentes del ecosistema.

- **Dominio del ecosistema de datos en Colombia – de tipo Gubernamental:**

El dominio o entrono hace referencia al sector de la sociedad donde se desarrolla y consolida el ecosistema de datos. En ese sentido, el dominio a caracterizar es gubernamental, donde el Gobierno nacional tiene un rol preponderante para ayudar a coordinar el ecosistema de datos por su propósito de dinamizar el ecosistema para generar valor económico y social. En el ecosistema de datos las entidades gubernamentales son actores centrales, que toman la iniciativa dentro de sistemas en red, organizados para lograr objetivos específicos relacionados con la innovación y el buen gobierno. La configuración del ecosistema de datos de Colombia, desde un enfoque gubernamental, visibiliza el rol que ha tenido el Gobierno nacional en la configuración del ecosistema de datos en el país, alrededor de i) la generación de políticas públicas en torno a los datos; ii) la generación de normatividad para la protección de datos personales, la apertura de información pública, y la protección de datos comerciales; iii) el fortalecimiento de capacidades y competencias para el aprovechamiento de los datos por diferentes actores de la sociedad; iv) la coordinación de los actores que hacen parte del ecosistema de datos; y v) la generación de condiciones en materia de infraestructura física como la conectividad digital y la plataforma de interoperabilidad para el intercambio de datos.

- **Estructura organizativa del ecosistema de datos:**

La estructura organizativa hace referencia al tipo de relaciones, interacciones y organizaciones que se pueden estructurar entre los agentes. En ese sentido, en el ecosistema de datos de Colombia el Gobierno funge un papel principal para la promoción de la gobernanza, dado que lidera la gobernanza de la infraestructura de datos del Estado, promueve la articulación con otros actores del ecosistema, y fomenta la cultura de datos, así como una sociedad y economía basada en datos. Si bien el Gobierno es un actor clave en el ecosistema, en la medida en que puede fungir como un agente dinamizador del ecosistema, no tiene un control completo de los recursos ni rige de manera autoritaria el comportamiento de otros actores, por lo cual, no se puede definir netamente como el agente principal.

En ese sentido, se encarga de proveer infraestructura común para que los demás actores del ecosistema de datos puedan disponer, acceder, usar y aprovechar los datos del ecosistema. Un ejemplo de esto es la provisión del Portal de datos abiertos “datos.gov.co” o el Portal único del Estado colombiano “gov.co”, en el que se disponen servicios de uso e intercambio de datos a diferentes actores del ecosistema.

Así mismo, el Gobierno cumple con la función de establecer lineamientos y políticas asociadas a la infraestructura de datos, el ecosistema de datos y el ecosistema digital. Entre ellas, podemos mencionar: el CONPES 3920 - Política Nacional de Explotación de Datos (Big Data), el CONPES 3975 - Política de Transformación digital e Inteligencia Artificial y la Resolución MinTIC 460 de 2022 - Plan Nacional de Infraestructura de Datos, y el Plan Estadístico Nacional 2023-2027, y lo definido en el Plan Nacional de Desarrollo 2022-2026 “Colombia, potencia mundial de la vida”. En ese sentido, el Gobierno también es el encargado de fomentar y generar condiciones para gestionar la ambivalencia del ecosistema (tanto para la competencia como para la cooperación), dado que, por un lado, provee las condiciones para generar la competencia de los actores y, por otro lado, fomenta la cooperación entre los mismos para garantizar la creación de valor social.

Por último, el Gobierno en la medida en que quiere generar cultura de datos, así como una economía y sociedad impulsada por los datos debe hacer un ejercicio de gobernanza en el sentido de que debe comprender a los agentes para fomentar el uso de los datos y, con ello, garantizar en el sector público la cooperación y en el sector privado la cooperación.

Meta- dimensión: Infraestructura tecnológica:

Infraestructura

La infraestructura tecnológica se refiere al espacio o plataforma que consta de componentes y servicios que habilitan el desarrollo del ecosistema de datos, con el fin de facilitar el acceso, la compartición e intercambio de datos entre los actores que participan en el ecosistema. Al respecto, es importante mencionar la definición de infraestructura tecnológica definida en el PNID) que comprende *“recursos como servidores, dispositivos de red, infraestructuras en nube, centros de datos que permiten la gestión del ciclo de vida de los datos para su captura, procesamiento, almacenamiento, compartición y análisis de los datos dispuestos en la infraestructura de datos del Estado”* (MinTIC, 2021). Así mismo, es importante aclarar que, conforme a esta definición, la infraestructura tecnológica es un recurso contextual y flexible, que depende de las características y necesidades de los actores del ecosistema.

En Colombia no existe la modalidad de un centro de datos que sirva como repositorio único de datos para todas las entidades públicas. Por tanto, se entiende que la infraestructura de datos a nivel nacional tiene un carácter distribuido, es decir que cada entidad desarrolla su infraestructura de datos conforme a sus necesidades y propósitos. Así mismo, la plataforma de intercambio de datos, facilitada por el Gobierno nacional es de carácter distribuido y sirve como una arquitectura técnica que apoya el intercambio de datos de manera confiable.

En el año 2019 se inició la implementación de *X-road*, adoptado del modelo de Estonia. Esta solución tecnológica de código abierto facilita el intercambio de datos entre diferentes sistemas de información de las entidades públicas y es un elemento central de la infraestructura de datos del Estado. Aunque los servicios de *X-road* son centralizados para las entidades públicas, esta plataforma es una capa de intercambio de datos de carácter distribuido, pues cada entidad sigue siendo custodia de su información. Es decir, es autonomía de las entidades públicas decidir quién accede al servicio, con qué permisos y a qué tipo de información. Aunque a la fecha esta plataforma de intercambio tiene mayor participación del sector público, tiene la posibilidad de habilitar servicios para otros actores como el sector privado.

La otra plataforma dispuesta para el consumo de datos provista por el Gobierno es la del portal de datos abiertos "*datos.gov.co*". Esta plataforma permite la publicación y reutilización de datos abiertos por parte de distintos actores del ecosistema de datos, además facilita su consumo a través de estándares y de herramientas como API. En este portal de datos abiertos, también se disponen los metadatos para la identificación de conjuntos de datos disponibles en el portal. En ese se proveen los datos de manera centralizada o federada, en la medida en que las entidades también disponen de portales donde publican sus datos y los ejercicios y análisis que hacen a partir de ellos.

En el país existen otros tipos de infraestructuras tecnológicas provistas por actores del sector privado o la academia para el consumo y reutilización de otros tipos de datos. Sin embargo, no se puede considerar a estas plataformas como infraestructura tecnológica centralizada. Por ejemplo, un elemento importante para considerar, a nivel de infraestructuras tecnológicas para el uso e intercambio de datos, son los modelos descentralizados como fábricas de datos¹⁴ que están siendo abordados a nivel local por organizaciones como AGATA¹⁵, y la infraestructura tecnológica provista para el desarrollo de modelos de gobernanza para el intercambio de datos que se están impulsando desde el PNID y el DANE

Apertura

El grado de apertura de la infraestructura tecnológica mencionada anteriormente (*X-road* y *datos.gov.co*) podría categorizarse como abierto, en la medida en que está diseñada para habilitar la participación de todos los actores del ecosistema, incluyendo al sector privado, la academia y la sociedad civil. No obstante, pese a la apertura del

¹⁴ Es un concepto de arquitectura basado en metadatos para conectar un conjunto de herramientas de datos que abordan puntos débiles en la gestión de los datos, que son claves en proyectos de big data. Una fábrica de datos brinda capacidades en las áreas de acceso a datos, descubrimiento, transformación, integración, seguridad, linaje y orquestación.

¹⁵ Agencia de Análítica de datos de Bogotá, cuyo propósito es impulsar la toma de decisiones basada en datos y la generación de soluciones de analítica para ayudar a la ciudad (<https://agatadata.com/index.html>).

ecosistema, pueden existir ciertas barreras de acceso para los diferentes actores, inclusive actores del Gobierno. Este tipo de barrera puede estar asociado a condiciones normativas, tecnológicas y técnicas, que deben cumplirse para la adopción de criterios y estándares que integren los servicios de uso e intercambio de datos y que, en ese sentido, permitan la participación de los actores.

Meta- dimensión: Gobernanza

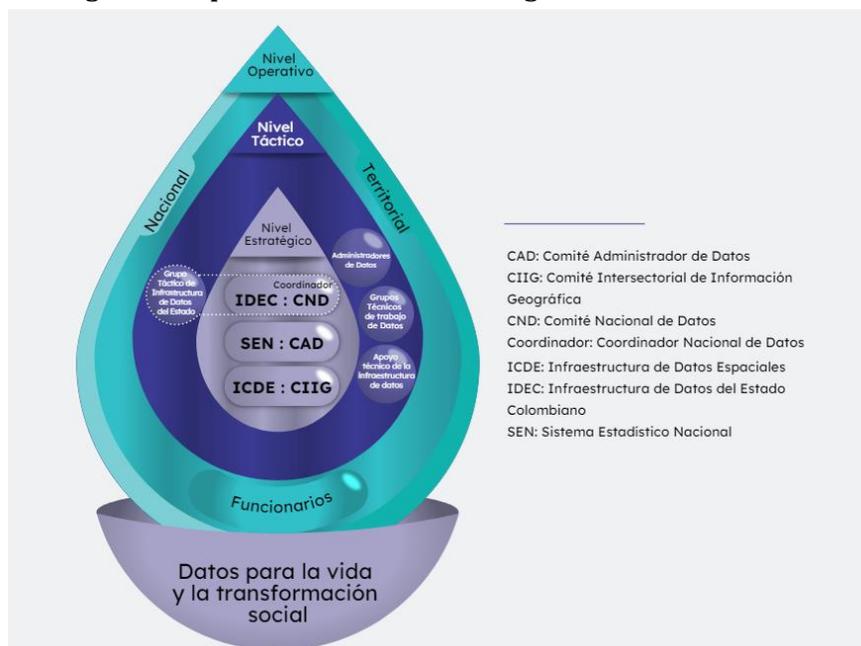
La gobernanza hace referencia a los aspectos y propiedades respecto a la posesión o dominio sobre los datos. Al respecto, el diseño e implementación de políticas públicas que han impulsado el aprovechamiento de datos en el país se ha enmarcado en un contexto gubernamental, que ha contado con la participación de entidades como el MinTIC, la Presidencia de la República, el DNP, el DANE, el Archivo General de la Nación y la Superintendencia de Industria y Comercio (SIC). En ese sentido, el marco de política ha tenido como propósito articular y orientar las acciones de distintos actores del país, con el fin de generar un entorno favorable para el aprovechamiento de datos por parte de los ciudadanos, el sector privado, la academia y el sector público, y, con ello, fomentar la creación de valor público y económico a partir de los datos.

- **Nivel de interdependencia: Vinculo estrecho**

Con el fin de desarrollar los diferentes ecosistema de datos en Colombia, se han definido instancias de gobernanza y articulación como el Consejo Asesor Técnico del Sistema Estadístico Nacional (CASEN) y el Comité de Administración de Datos (CAD) en el caso de los datos estadísticos, la Comisión Intersectorial de Información Geográfica (CIIG) en el caso de información geográfica y el Comité Nacional de Datos (CND) como instancia de articulación institucional para el desarrollo de la infraestructura de datos del Estado colombiano que es presidida por el Coordinador Nacional de Datos.

La alineación de la Gobernanza de datos en Colombia se plasma a través de la articulación del Modelo de Gobernanza de la Infraestructura de datos del Estado colombiano (IDEC) definido en el Decreto 1389, el Sistema Estadístico Nacional (SEN) y la Infraestructura Colombiana de Datos Espaciales (ICDE), como se puede observar en la siguiente figura:

Figura 3. Esquema de alineación de la gobernanza de los datos



Fuente: Plan Estadístico Nacional 2023-2027

Como se ve reflejado en el esquema de alineación de la gobernanza de los datos, el país ha alcanzado una articulación estrecha entre las entidades públicas que lideran la agenda de datos en el país, logrando proponer un trabajo conjunto entre las diversas estrategias institucionales para aumentar el aprovechamiento de datos en el país.

- **Gobernanza descentralizada y sin control de los recursos**

En términos de control de los recursos, el ecosistema de datos en Colombia es descentralizado en la medida en que no hay una sola entidad del Estado que controle o acapare todos los datos y los recursos esenciales para su aprovechamiento. Se aproxima a un escenario donde el control de los recursos se distribuye a través de los múltiples actores del ecosistema de datos. En este escenario, hay oportunidades para impulsar recursos tecnológicos compartidos e iniciativas para promover el fortalecimiento de capacidades para el aprovechamiento de los datos.

Un ejemplo de este enfoque es el proyecto de software público de Colombia¹⁶, cuyo objetivo es promover el uso de software libre para la solución de necesidades de la administración pública, con el fin de aumentar la innovación en las entidades públicas.

¹⁶ Este proyecto surgió de un acuerdo con el Banco Mundial cuyo objetivo fue "(...) dinamizar el entorno de la innovación, a través de la producción de insumos de conocimiento necesarios para conceptualizar un esquema de colaboración abierta, el diseño y la implementación de una iniciativa de datos abiertos y la preparación de una estrategia nacional para la creación de un modelo para compartir soluciones tecnológicas entre las instituciones gubernamentales" (MinTIC, 2020).

Otra iniciativa para destacar es el Data Sandbox diseñado en el año 2020, cuyo propósito fue crear un ambiente colaborativo para el desarrollo de proyectos de analítica de datos por parte de las entidades públicas, y el Sandbox regulatorio de la SIC, que es un espacio supervisado y de experimentación temporal, cuyo propósito es crear soluciones al cumplimiento colaborativo de las normas de protección de datos personales en proyectos de inteligencia artificial.

4. Componentes del ecosistema de datos de Colombia

Como complemento a la caracterización de los del ecosistema de datos en Colombia desde un enfoque gubernamental, en esta sección se hace una aproximación al entendimiento del ecosistema de datos a partir de los componentes generales de: actores, roles, recursos e incentivos. Para lo cual se revisaron documentos técnicos e información recolectada a través de la Metodología de caracterización cualitativa elaborada por el DNP para este estudio¹⁷.

4.1.1 Actores y roles

Como parte del ejercicio de plantear una metodología para la caracterización del ecosistema de datos colombiano, se resalta la importancia de identificar a los principales agentes que interactúan en el ecosistema y los roles que desempeñan. En ese sentido, a nivel del sector público se identifica que en Colombia las principales entidades responsables de la gobernanza, gestión, aprovechamiento y regulación de datos son:

Tabla 6. Actores sector público ecosistema de datos Colombia

Entidad pública	Descripción
Agencia Nacional Digital (AND)	La AND promueve, articula y administra la prestación de servicios ciudadanos digitales contribuyendo a la creación de un ecosistema de información pública en el marco de la transformación digital del país.
Archivo General de la Nación	El Archivo General de la Nación es el encargado de formular, orientar y controlar la Política Archivística, coordinar el Sistema Nacional de Archivos y garantizar la conservación del patrimonio documental, asegurando los derechos de los ciudadanos y el acceso a la información.

¹⁷ El formulario que hace parte de esta metodología fue diligenciado por 29 representantes del ecosistema de datos en Colombia, en el que se incluye un porcentaje de participación del sector gobierno (76%), academia (4%), privado (16%) y sociedad civil (4%).

Presidencia de la República	La Presidencia de la República ejerce la coordinación nacional de datos, rol definido en el Decreto 1389 de 2022 y, en ese sentido, tiene la función de articular a las diferentes entidades del gobierno en torno la infraestructura y gobernanza de datos.
Departamento Administrativo Nacional de Estadística (DANE)	El DANE en el marco de sus funciones promueve, planea, implementa y evalúa procesos rigurosos de producción y comunicación de información estadística a nivel nacional, que soporten la comprensión y solución de las problemáticas sociales, económicas y ambientales del país y sirvan de base para la toma de decisiones públicas y privadas.
Departamento Nacional de Planeación (DNP)	EL DNP, a través de la Dirección de Desarrollo Digital (DDD), tiene la responsabilidad de fomentar y promover el desarrollo de la infraestructura de datos en Colombia, y fomentar la articulación interinstitucional para el aprovechamiento de datos en el marco del diseño e implementación de políticas públicas. Adicionalmente, busca promover el fortalecimiento de las capacidades en las entidades públicas para el aprovechamiento de datos y gestión de metadatos.
Instituto Geográfico Agustín Codazzi (IGAC)	El IGAC es la máxima autoridad en regulación, producción y articulación, con altos estándares de calidad, de la información geográfica, catastral y agrológica del país, contribuyendo con su desarrollo para la toma de decisiones y definición políticas públicas.
Ministerio de Tecnologías de la Información y Comunicaciones (MinTIC)	El MinTIC lidera la iniciativa pública para impulsar la inversión en el sector TIC y para la transformación digital del Estado. Promueve aquellas iniciativas que permiten la compilación, procesamiento, almacenamiento y transmisión de información como los datos. Es la entidad encargada de liderar la implementación de la Política de Gobierno Digital y el Plan Nacional de Infraestructura de Datos
Ministerio de Comercio, Industria Y Turismo (MinCIT)	El MinCIT promueve, a partir del aprovechamiento de datos, el desarrollo económico y el crecimiento empresarial para mejorar la calidad de vida de los ciudadanos y empresarios.
Superintendencia de Industria y Comercio de Colombia (SIC)	La SIC es la autoridad nacional de protección de la competencia y los datos personales. Es la entidad que establece parámetros sobre intercambio/interoperabilidad de conjuntos de datos que contienen datos personales. También promueve la ética desde el diseño y por defecto en torno a este tipo de datos.

Fuente: Elaboración propia

En relación con el sector privado, academia y sociedad civil se destacan las organizaciones que abordan asuntos relacionados con la gobernanza de datos, así como la innovación pública y privada basada en datos, la protección y seguridad de los datos, la gestión ética de los datos, el desarrollo de nuevos modelos de negocio basados en datos, el desarrollo de talento humano en el país, estándares e interoperabilidad. Algunas organizaciones y entidades a destacar son:

Tabla 7. Actores sociedad civil y organizaciones mixtas de datos de Colombia

Organización	Descripción rol
Agencia de analítica de datos (AGATA)	La agencia busca impulsar la toma de decisiones basadas en datos mediante la generación de soluciones analíticas para transformar la ciudad de Bogotá y mejorar la calidad de vida de las personas.
Alianza CAOBA	Apoya el uso de las tecnologías de Big Data y Data Analytics a través de diferentes frentes que incluyen la formación del talento humano, la investigación aplicada y el desarrollo de productos a la medida.
Centro de Investigación y Desarrollo en Tecnologías de la Información y las Comunicaciones (CINTEL)	Realiza proyectos orientados a apoyar la transformación digital de las empresas públicas y privadas.
Centro para la Cuarta Revolución Industrial (C4RI)	Adelanta proyectos e iniciativas que apuntan a desarrollar pilotos, soluciones o marcos de análisis y regulación para el uso de datos e inteligencia artificial en propósitos comunes que signifiquen un impacto positivo para la sociedad.
INNpulsa	INNpulsa, mediante MiLAB, busca acelerar la transformación digital del Gobierno nacional conectándolo, a través de estrategias de colaboración e innovación abierta, con emprendedores y pymes que hacen uso de datos, tecnologías exponenciales y metodologías innovadoras, para promover mejoras en la eficiencia del Estado y robustecer su infraestructura digital.
Datasketch	Es una empresa de tecnología social que busca promover el uso responsable de las tecnologías de la información y de los datos. Buscan democratizar la ciencia de datos y fomentar la transparencia para combatir las desigualdades sociales.
Fundación Karisma	Busca proteger y promover los derechos humanos y la justicia social en el diseño y uso de las tecnologías digitales.
Instituto Colombiano de Normas técnicas y Certificación (ICONTEC)	Como Organismo Nacional de Normalización de Colombia participa en la definición y el desarrollo de normas internacionales y regionales para estar a la vanguardia en información y tecnología.

Fuente: Elaboración propia

Por último, el ecosistema de datos en Colombia cuenta con importantes aliados internacionales que promueven la discusión local y global en materia de gobernanza de datos y sus temáticas asociadas. Entre ellos se destaca:

Tabla 8. Actores organismos multilaterales ecosistema de datos de Colombia

Organismo	Descripción rol
Banco Mundial	El Banco Mundial ha elaborado recomendaciones para impulsar a los países y organizaciones internacionales en el desarrollo de políticas y estrategias de gobernanza de datos, con el fin de

	superar los desafíos relacionados con seguridad y privacidad de los datos, crecimiento y equidad. En el 2020, el Banco Mundial publicó el informe “Datos para una vida mejor” para orientar a los países en la definición de un contrato social en materia de datos.
Banco de Desarrollo de América Latina (CAF)	La CAF contribuye a la transformación digital del Estado impulsando iniciativas para que los Estados sean más ágiles, abiertos, innovadores y eficientes, apalancándonos en las nuevas tecnologías y la inteligencia de datos, con el objetivo de que con ello mejore la calidad de los servicios que se ofrecen a los ciudadanos de América Latina.
Banco Interamericano de Desarrollo (BID)	El BID trabaja junto a los países en el diseño e implementación de estrategias de transformación digital que impulsen la adopción y uso de tecnologías digitales (Internet de las cosas, la robótica, la inteligencia artificial, el aprendizaje de máquinas, el big data, la computación en la nube, entre otras).
Instituto Latinoamericano de Datos Abiertos (ILDA)	<p>ILDA se dedica a explorar los efectos de la datificación en América Latina y genera diálogos con múltiples sectores, buscando consolidarse como referentes en las áreas de: gobernanza de datos, transparencia del sector público, inclusión y diversidad en el uso de tecnologías emergentes.</p> <p>A partir del 2020, ILDA asumió el compromiso de implementar una edición regional del Barómetro de Datos Abiertos, que tiene como objetivo conocer el predominio e impacto de las iniciativas de datos abiertos en todo el mundo, analizar tendencias globales y proporcionar datos comparativos sobre gobiernos y regiones.</p>
Organización para la cooperación y el desarrollo económico (OCDE)	La OCDE ha apoyado a los responsables políticos en la búsqueda de una mejor comprensión de la transformación digital y los efectos de las tecnologías digitales en nuestras economías y sociedades, en un esfuerzo por dar forma a un futuro digital. La fase III del proyecto Going Digital de la OCDE relacionado con la gobernanza de los datos está ayudando a los países a diseñar mejores mecanismos de gobernanza de datos, informados por una mejor comprensión del impacto de los datos y la medición de su valor.

Fuente: Elaboración propia

En el Anexo A. Matriz mapeo de actores, se incluye una matriz con el mapeo de los actores que participan en el ecosistema de datos para el caso colombiano, vale la pena aclarar que pueden existir más actores involucrados al ecosistema. Así mismo, el mapa de actores no es estático y se puede ir complementando para futuras versiones del documento.

Adicionalmente, para la definición de roles del ecosistema de datos de Colombia se usó como referencia los roles definidos por la OCDE y por el Instituto de datos abiertos. Los roles definidos son consumidores de datos, emprendimientos impulsados por analítica

de datos, proveedores de analítica, proveedores de datos, proveedores de servicios de Internet, proveedores de infraestructura TI y responsables de política, leyes y normativas (Tabla 3). En relación con los roles, en el Anexo A también se presenta una propuesta de los roles que desempeñan los actores del ecosistema de datos. Es importante aclarar que un actor puede desempeñar varios roles como, por ejemplo, el MinTIC que es consumidor, proveedor de datos, y responsable de políticas, leyes y normas.

Actualmente, se han mapeado más de 200 actores, de las cuales el 46% son organizaciones del sector privado, el 32% son entidades gubernamentales, el 9% son organizaciones internacionales y regionales, 8% son de la academia, el 3% corresponde a actores de categoría mixta y el 2% representa a organizaciones de la sociedad civil e individuos. De estos actores el 37% desempeñan como rol principal el de consumidores de datos, el 16% son promotores de cultura de datos, el 11% son proveedores de datos, el 10% son proveedores de analítica, el 9% son proveedores de servicios de Internet, el 7% son responsables de política, el 6% son proveedores de servicios e infraestructura TI, el 2% son reguladores y el 1% son proveedores de estándares.

De acuerdo con el cuestionario diseñado para este estudio, y aplicado a varios actores del ecosistema de datos en Colombia, los principales roles que los actores desempeñan de forma preponderante son: Responsable de políticas y normativas (34%), proveedor de datos (21%) y proveedor de analítica de datos (25%). De forma complementaria, se identificó como rol secundario las actividades de: Emprendedor basado en datos (8%), proveedor de servicios de Internet (4%), otro (8%). Así mismo en la Tabla 9 se hace una descripción de otros roles identificados en el ecosistema de datos en Colombia.

Tabla 9. Definición aproximada de los nuevos roles identificados en el formulario

Roles	Descripción
Entrenadores en ciberseguridad	Son los encargados del entrenamiento acerca de la seguridad informática de las empresas y organizaciones. Por tanto, sus funciones están asociadas a diseñar y desarrollar proyectos, planes, programas y herramientas de seguridad.
Gestor para el gobierno de datos	Este rol se encarga de la definición, coordinación y vigilancia funcional de todos aquellos procesos, principios, políticas, estándares, herramientas y responsabilidades presentes que permitan gestionar los datos de manera eficiente.
Innovador de datos	Son los encargados de mejorar e innovar habilidades técnicas en ciencia de datos. Por tanto, sus funciones están asociadas al diseño

	centrado en usuarios, el uso de metodologías ágiles o la gestión del conocimiento y aprendizajes.
Articulador de cultura de datos	Son los encargados de gestionar y promover las creencias y los comportamientos colectivos de las personas que valoran, aprovechan y promueven el uso de datos para mejorar la toma de decisiones.

Fuente: Elaboración propia

4.1.2 Recursos del ecosistema de datos

En el entorno de ecosistemas de datos se generan e intercambian recursos provistos por distintos actores. Estos, vale la pena recordar, son recursos necesarios para dinamizar al ecosistema de datos y apoyar el ciclo de vida de los datos. Los recursos que se identifican para el ecosistema de datos en Colombia son los siguientes:

Tabla 10. Recursos del ecosistema de datos de Colombia

Categoría	Descripción	Ejemplos
Políticas y marco normativo	Es el marco normativo y de política pública que habilita y dinamiza la interacción de los actores que hacen parte del ecosistema de datos, el intercambio de datos y recursos para su aprovechamiento.	<ul style="list-style-type: none"> • Plan Nacional de Desarrollo • Planes sectoriales • Documentos de Política Pública • Leyes, decretos, resoluciones y circulares
Datos	Los datos son el recurso más importante del ecosistema. Se constituyen como datos estructurados, no estructurados y semiestructurados, adoptados en diferentes formatos y de distintos tipos.	<ul style="list-style-type: none"> • Estadísticos • Geoespaciales • Administrativos • Abiertos • Transaccionales • Maestros • Financieros • Médicos • Comerciales
Portales	Son enlaces digitales que centralizan recursos, servicios y/o conjuntos de datos para la publicación, intercambio, reutilización, comercialización o análisis de datos. Ofrecen a los actores del ecosistema de datos un mejor acceso a la información.	<ul style="list-style-type: none"> • Portal de datos abiertos del Estado colombiano (https://www.datos.gov.co/) • Portal del Estado colombiano (https://www.gov.co/) • Portal de la Infraestructura Colombiana de Datos Espaciales (https://www.icde.gov.co/)

		<ul style="list-style-type: none"> • Portal de Datos Estadísticos (https://www.dane.gov.co/)
Sistemas de información	Es un conjunto de recursos interconectados que permiten la administración y distribución de datos y de información. En los sistemas de información convergen elementos financieros, humanos y tecnológicos, materiales y administrativos.	<ul style="list-style-type: none"> • eKogui- Sistema Único de Gestión e Información Litigiosa del Estado • Sistema Electrónico para la Contratación Pública SECOP I • Sistema de Información de proyectos de Infraestructura (SIPI) • Sistema Único de Información de Trámites (SUIT)
Infraestructura tecnológica	Recursos como servidores, dispositivos de red, infraestructuras en nube, centros de datos que permiten la gestión del ciclo de vida de los datos para su captura, procesamiento, almacenamiento, compartición y análisis de los datos.	<ul style="list-style-type: none"> • Socrata-Portal de datos abiertos • X-road • Catálogo de productos y servicios - Datos espaciales
Lineamientos, guías y estándares	Recursos que contribuyen a la estandarización de procesos relacionados con el ciclo de vida de los datos y su aprovechamiento. Entre estos lineamientos y guías se encuentran temáticas sobre calidad de datos, anonimización, aprovechamiento y analítica de datos y Big data.	<ul style="list-style-type: none"> • Guía de herramientas de analítica para la explotación de datos • Guía para el uso y aprovechamiento de datos abiertos • Guía para la elaboración de metadatos de registros administrativos • Guía para la anonimización de bases de datos en el Sistema Estadístico Nacional • Guía de referencia de Blockchain para la adopción e implementación de proyectos del Estado Colombiano

Fuente: Elaboración propia

El mapeo completo de los recursos del ecosistema de datos en Colombia se puede consultar en el Anexo B “Recursos del ecosistema”. En este anexo, que no pretende ser exhaustivo se han mapeado más de 100 recursos disponibles para consulta de los actores del ecosistema, relacionados con lineamientos, guías y estándares, portales y recursos de infraestructura tecnológica.

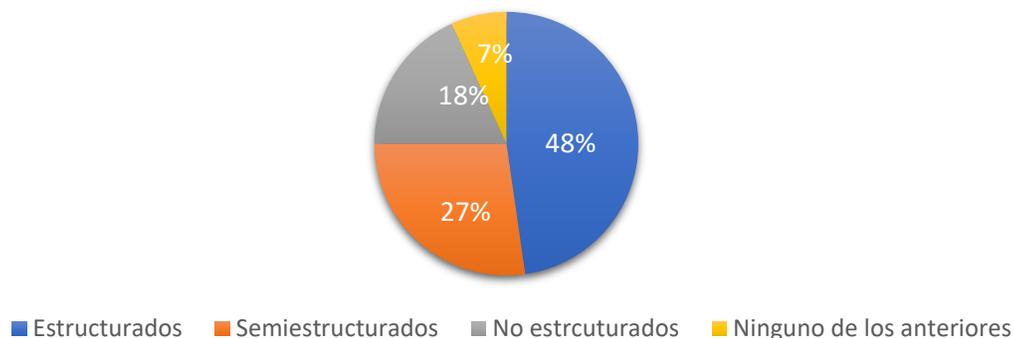
De acuerdo con el cuestionario que se aplicó a actores del ecosistema de datos, se tiene que los principales recursos que se provee en el ecosistema son datos, datos abiertos, datos estadísticos, sistemas de información como el Sistema de gestión y de Información Pública del Empleo (SIGEP), el Sistema Unico de Información de Trámites (SUIT), y el Formulario Unico de Reporte y Avance en la Gestión (FURAG). También se proveen recursos asociados a herramientas tecnológicas como: API para el consumo de datos, infraestructura de nube pública y privada (como el mecanismo de agregación de demanda de la Agencia Nacional de Contratación Pública - Colombia Compra Eficiente), plataformas como la del portal de datos abiertos del Estado colombiano, y cubos OLAP¹⁸ para consulta, servicios de analítica y consultas web. Así mismo, estudios y conocimientos académicos relacionados con datos, tecnologías de la cuarta revolución industrial, y estándares, normativas, semánticas y políticas públicas.

En relación con los principales recursos que se demandan por parte de los actores se destacan los datos abiertos, datos estadísticos, datos geoespaciales, y datos financieros, e infraestructura tecnológica, conectividad digital y recursos financieros. También se destacan los estándares y metodologías, las plataformas de intercambios de datos, repositorios de datos y servicios basados en datos. Se destaca la demanda de API de consumos de datos clusterizados por tipo de población o localizada.

A partir del instrumento aplicado a los actores del ecosistema de datos en Colombia, se identifica que los tipos de datos que más consumen los actores son los estadísticos (24%), seguidos de los geoespaciales (16%), financieros (12%), contables (11%), científicos de salud (10%), meteorológicos y ambientales (9%), ofimáticos (7%), otros (11%). Por su parte, los tipos de datos que más suministran los actores en el ecosistema de datos son los estadísticos (31%), seguidos de los geoespaciales (15%), financieros (13%), contables (10%), científicos de salud (9%), meteorológicos y ambientales (9%), ofimáticos (3%), otros (10%).

Así mismo se identifican otros tipos de datos que los actores consumen y suministran dentro del ecosistema de datos, entre ellos se encuentran los datos jurídicos, personales, legales, de seguridad/ciberseguridad, contractuales, relacionales, científicos y tecnológicos. Así mismo, de acuerdo con los resultados encontrados, la mayoría de los actores consumen o comparten datos estructurados (48%), los cuales por su estructura ordenada son los más sencillos de gestionar, tanto digital como manualmente (ver Figura 4).

¹⁸ Son los OnLine Analytical Processing cuyo propósito es generar una presentación del almacenamiento de datos en forma de cubos, de tal forma que sea más fácil organizar y resumir los datos para hacer más eficiente los procesos de consulta y analítica.

Figura 4. Tipos de datos que consumen los actores del ecosistema de datos

Fuente: Elaboración propia

Con el fin de establecer un marco básico para el intercambio de datos, los actores calificaron la importancia de los datos estructurados, semiestructurados y no estructurados para su entidad u organización. Siendo uno (1) nada importante y cinco (5) muy importante. La mayoría de los encuestados respondió que los datos estructurados son muy importantes (88%) para las funciones que su entidad u organización desempeña en el ecosistema de datos. Así mismo, ninguno de los actores desconoce la importancia que tienen los datos para su organización.

Preponderantemente se considera que los datos semiestructurados son muy importantes (46%) para las funciones que su entidad/organización desempeña, seguido de importante (29%), medianamente importante (21%) y poco importante (4%). Por otra parte, el 50% de los encuestados considera que los datos no estructurados son muy importantes para las funciones que su entidad/organización desempeña en el ecosistema de datos, seguido de importante (14%), medianamente importante (21%), poco importante (8%) y nada importante (8%).

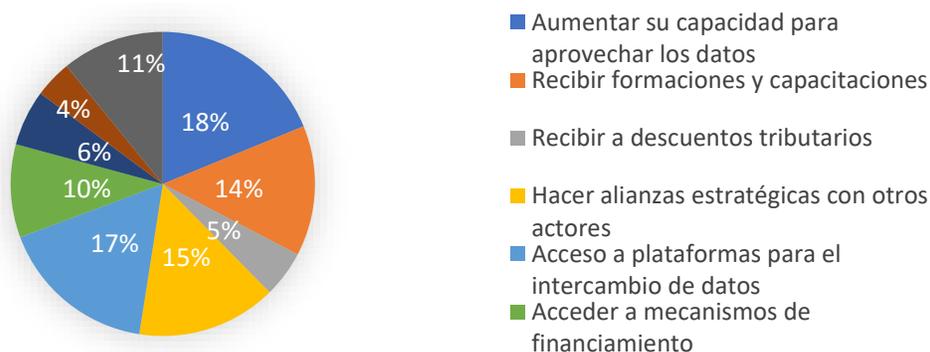
4.1.3 Incentivos que tienen los actores para participar en el ecosistema de datos

Los incentivos fomentan una mayor participación de los actores en el ecosistema de datos. Por consiguiente, en el cuestionario de caracterización se realizó una pregunta para identificar algunos de estos. La pregunta que se abordó fue: *“teniendo en cuenta su rol dentro del ecosistema de datos, identifique qué incentivos lo anima a participar dentro del ecosistema de datos”*.

Dado lo anterior, los principales incentivos que reconocen los actores promueven su participación en el ecosistema de datos son: mejorar su capacidad para aprovechar los datos (18%), acceder a plataformas para el uso e intercambio de datos (17%), hacer alianzas estratégicas con otros actores (15%), recibir formaciones y capacitaciones

(14%), y fomentar la protección de derechos de los ciudadanos en el marco del ecosistema digital (11%).

Figura 5. Incentivos para participar dentro del ecosistema de datos



Fuente: Elaboración propia

Así mismo, unas de las razones más fuertes que tienen los actores para interactuar con otros agentes del ecosistema son: el acceso a plataformas para el uso y el intercambio de datos, y la creación o compartición de conocimiento. A pesar de que el cuestionario está dirigido a diferentes sectores, de forma mayoritaria los actores que participaron hacen parte del sector público, por lo cual, algunos incentivos que pueden ser preponderantes para otros sectores tienen baja proporción de respuesta.

5. Conclusiones y recomendaciones

Como se ha demostrado a lo largo del documento los ecosistemas de datos son sistemas orgánicos, en las cuales interactúan diferentes agentes, conforme a sus propósitos, para generar valor entorno al ciclo de vida de los datos. Por ello, al ser dinámicos no es posible establecer una estructura unívoca, que englobe todas las características y los comportamientos de los ecosistemas.

A partir del ejercicio de caracterización y de la aproximación al ecosistema de datos en Colombia, que se estructuró conforme al marco de entendimiento propuesto en el documento, se identifica que los actores, roles, recursos, incentivos, barreras y el valor generado, son contextuales. Los componentes del ecosistema responden a las condiciones específicas de cada uno de los ecosistemas y de sus propósitos en torno al ciclo de vida de los datos. Sin embargo, como se menciona en el marco de entendimiento, las condiciones de cada ecosistema no impiden que se relacionen entre sí. Especialmente, los ecosistemas públicos, que están diseñados para nutrir y servir a otros ecosistemas, dado que intentan promover principios como el Gobierno Digital o el Gobierno abierto. Es por ello, que el análisis de la caracterización de Colombia se centra en la estructura pública, entendiendo que, en términos institucionales y de

gobernanza, este sistema puede relacionarse con todos los ecosistemas, a través de los recursos que provee, y sin la necesidad de convertirse en el actor principal, pero si en un promotor del desarrollo de los ecosistemas a nivel general.

Es necesario que el Estado colombiano pueda seguir fortaleciendo el desarrollo de los ecosistemas de datos. Para ello, los actores que representan al sector público, como responsables de políticas, leyes y normativas, deben generar procesos que vinculen de mejor manera a los gobiernos territoriales, el sector privado, la academia y la sociedad civil.

En ese sentido, se deben seguir avanzando respecto a las propuestas del CONPES 3920, el CONPES 4023, el PNID, la Política de Gobierno Digital y el fortalecimiento de la gobernanza de datos (definida mediante el Decreto 1389 de 2022), que comprenden la importancia de los datos para la sociedad, en términos del valor público y económico que se puede crear. Por el momento, los retos inmediatos siguen siendo el fortalecimiento de la cultura de datos; el desarrollo de habilidades digitales que permitan aprovechar, dentro del ecosistema digital, todos los recursos disponibles para generar valor; el fortalecimiento de ejercicio de gobierno abierto que permitan a la ciudadanía participar en ejercicios basados en datos relacionados con la transparencia y veeduría sobre la gestión pública; el fortalecimiento de la interoperabilidad y estándares de datos.

Para promover el desarrollo eficiente de los ecosistemas de datos, es importante seguir nutriendo la discusión teórica y la caracterización de diferentes tipos de ecosistemas. Lo anterior, entendiendo que existen innumerables ecosistemas de datos y el conocimiento de las características operantes de los ecosistemas son la mejor medida para garantizar el flujo de datos, la gobernanza, la coopetición y la generación de valor. En tanto, debe estar en la agenda del hacer público la consolidación de una cultura basada en datos donde los agentes tengan conciencia de la importancia de los datos para el desarrollo de sus funciones y, en ese sentido, también puedan comprender las estructuras operantes que configuran, dinamizan o impiden la compartición y uso de los datos.

BIBLIOGRAFÍA

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*. Obtenido de IEEE Jordan International 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*. Obtenido de https://www.accenture.com/_acnmedia/pdf-100/accenture-lideres-updated-marketing-report-up-integradores-big-data.pdf
- Adner, R. (2006). *Match your innovation strategy to your innovation ecosystem*. Obtenido de <https://hbr.org/2006/04/match-your-innovation-strategy-to-your-innovation-ecosystem>
- ANDI. (1 de Junio de 2021). *Transformación digital*. Obtenido de <http://www.andi.com.co/Home/alPagina/19-transformacion-digital>
- Armellino, M., & Paolino, M. (2019). *Modelo de madurez para gobernanza de Big Data*. Obtenido de <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*. Obtenido de <http://asip.org.ar/la-cadena-de-valor-publico-un-principio-ordenador-que-previene-la-colision-metodologica/>
- Banco Mundial. (2021). *Data for better lives*. Washington.
- Banco Mundial. (2021). *Data for better lives*. Washington: World Development Report.
- Barclays. (2019). *HMT Discussion Paper: The Economic Value of Data Barclays Response*. Obtenido de <https://home.barclays/content/dam/home-barclays/documents/citizenship/our-reporting-and-policy-positions/public-policy/Economic-Value-of-Data-Barclays-Perspective.pdf>
- Braun, H. (2015). *Evaluation of Big Data Maturity models*. Obtenido de <https://core.ac.uk/download/pdf/196555414.pdf>
- Capello, C., Gal, A., Rehof, J., & Jarke, M. (2019). *Data ecosystem: Sovereign data exchange among organizations*. Obtenido de https://www.researchgate.net/publication/339552585_Data_Ecosystems_Sovereign_Data_Exchange_among_Organizations_Report_from_Dagstuhl_Seminar_19391

- Colombia Vive. (12 de Julio de 2015). *Organismos Internacionales a los que pertenece Colombia*. Obtenido de <http://colombiavive2015.blogspot.com/2015/07/organismos-internacionales-los-que.html>
- Comisión de la verdad. (2008). *Transparencia*. Obtenido de Dato semiprivado: <https://comisiondelaverdad.co/transparencia/informacion-de-interes/glosario/dato-semiprivado#:~:text=Aquel%20que%20no%20tiene%20naturaleza,o%20a%20la%20sociedad%20en%20general>.
- Comisión de la verdad. (2008). *Transparencia*. Obtenido de Dato privado: <https://comisiondelaverdad.co/transparencia/informacion-de-interes/glosario/dato-privado>
- Comisión de Regulación de Comunicaciones. (2014). Capacidad de transmisión en las Redes de Fibra Óptica. En *Regulación de Infraestructura* (págs. 1-34). Bogotá.
- Comisión de Regulación de Comunicaciones. (2017). *Cadena de Valor y Cadenas de Acceso e Interconexión en Colombia*. Detecon consulting.
- Comisión Europea. (2020). *The European data market monitoring tool*.
- Conejero, E. (2014). *Valor público: Una aproximación conceptual*.
- CRC. (2021). *ANÁLISIS DEL SERVICIO DE ACCESO A INTERNET EN LA ISLA*. San Andres: Diseño regulatorio.
- DANE. (1 de Junio de 2021). *Organismos internacionales*. Obtenido de Dane.gov: <https://www.dane.gov.co/index.php/acerca-del-dane/informacion-institucional/organismos-internacionales>
- DANE. (07 de Julio de 2022). *Webinar SEN: calidad y aprovechamiento de Registros Administrativos*. Obtenido de <https://www.youtube.com/watch?v=XTdaeQnQ3dU&t=1767s>
- DANE. (s.f.). *Preguntas frecuentes*. Obtenido de <https://www.dane.gov.co/files/acerca/PF.pdf>
- Data economy. (2020). *Data economy*. Obtenido de <http://www.dataeconomy.eu/stakeholders/#page-content>

- De Bruin, S., Freeze, R., & Kulkarni, U. (2005). *Understanding the Main Phases of Developing a Maturity Assessment Model*. Obtenido de https://www.researchgate.net/publication/27482282_Understanding_the_Main_Phases_of_Developing_a_Maturity_Assessment_Model
- Departamento Nacional de Planeación. (2018). *Definición de la infraestructura de datos públicos*. Bogotá: Dirección de desarrollo digital.
- Departamento Nacional de Planeación. (18 de Julio de 2021). *CONPES 3975: Política nacional para la transformación digital e inteligencia artificial*. Obtenido de [Colaboracion.dnp.gov.co](https://colaboracion.dnp.gov.co): <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3975.pdf>
- Department for Digital, Culture, Media & Sport. (2020). *National Data Strategy*. Obtenido de <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>
- DNP. (2018). *Guía para la construcción y análisis de indicadores*. Bogotá: Departamento Nacional de Planeación.
- DNP. (2019). *4.11. Mercado de datos en Colombia*.
- DNP. (2020). *Aprovechamiento de datos para la toma de decisiones en el sector público*. Obtenido de <https://www.dnp.gov.co/DNPN/Paginas/Aprovechamiento-de-datos-para-la-toma-de-decisiones-en-el-sector-publico.aspx>
- Durans, F. (2017). *ENTIDADES GUBERNAMENTALES DE COLOMBIA*. Obtenido de [Colombiamania.com](http://www.colombiamania.com): <http://www.colombiamania.com/gobierno/index.htm>
- ESRI Colombia. (2019). *Datos abiertos*. Obtenido de ESRI Colombia: <https://esri.co/datos-abiertos/d-a/>
- Flórez, L. L., Numpaque, J. S., & Barreto Nieto, C. A. (s.f.). *Una aproximación al efecto de los proyectos de analítica de datos en la eficiencia y efectividad del Estado*. Bogotá.
- Foro Económico Mundial. (2014). *Rethinking Personal Data*. Obtenido de [Near-Term Priorities for Strengthening Trust](http://reports.weforum.org/rethinking-personal-data/near-term-priorities-for-strengthening-trust/?doing_wp_cron=1616508186.4629800319671630859375): http://reports.weforum.org/rethinking-personal-data/near-term-priorities-for-strengthening-trust/?doing_wp_cron=1616508186.4629800319671630859375
- Fraefel, M., Haller, S., & Gschwend, A. (2017). *Big Data in the public sector Linking city sensors*. Obtenido de

https://www.researchgate.net/publication/318879744_Big_Data_in_the_Public_Sector_Linking_Cities_to_Sensors

Función Pública. (2021). *Consulta: Directiva 03/21*. Obtenido de Directiva 03 de 2021 Presidencia de la Republica: https://www.funcionpublica.gov.co/eva/gestornormativo/norma_pdf.php?i=160326

Gama, K., & Bernadette, F. (2014). Towards Ecosystems based on Open Data as a Service Position Paper. *ReserchGate*.

Gelhaar, J., & Otto, B. (2020). Challenges in the Emergence of Data Ecosystem . *Pacific Asia Conference on Information Systems*.

Gelhaar, J., Groß, T., & Otto, B. (2021). A Taxonomy for Data Ecosystems. *54th Hawaii International Conference on System Sciences*, 6113-6122.

Ghavami. (2015). *Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics*.

Gobierno de la República de Colombia. (2015). *Open Data Readiness Assessment*. World Bank Group.

Gobierno Digital. (14 de Julio de 2021). *Software Público Colombia*. Obtenido de Gov.co: <https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Software-libre/>

González Guerrero, L. D. (2019). Control de nuestros datos personales en la era del big data: el caso del rastreo web de terceros. *Estudios Socio-Jurídicos*, 209-244. Obtenido de <http://dx.doi.org/10.12804/revistas.urosario.edu.co/sociojuridicos/a.6941>

GSMA. (Junio de 2018). *The Data Value Chain*. Obtenido de https://www.gsma.com/publicpolicy/wp-content/uploads/2018/06/GSMA_Data_Value_Chain_June_2018.pdf

Hajirahimova, M., & Aliveya, A. (2017). *About Big data measurement methodologies and indicators*.

Hussain Shah, S., Peristeras, V., & Magnisalis, I. (2020). Government (Big) Data Ecosystem: Definition, Classification of Actors, and Their Roles. *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, 14. Obtenido de <https://zenodo.org/record/3839626#.YFnjzK8zaUk>

- ICDE. (2021). *Objetivos y Principios Rectores de la ICDE*. Obtenido de Principios Rectores de la ICDE: <https://www.icde.gov.co/sobre-nosotros/objetivos-principios#:~:text=La%20ICDE%20se%20rige%20bajo,en%20la%20mejor%20informaci%C3%B3n%20posible>.
- Immonen, A., Palviainen, M., & Ovaska, E. (2014). Requirements of an Open Data Based Business Ecosystem. *IEEE Access*.
- Infotech. (2021). *¿Que es una empresa inteligente?* Obtenido de InfotechColombia: <http://infotechdecolombia.com/que-es-una-empresa-inteligente/>
- Intersoftware. (s.f.). *Veinte empresas del país cuentan con prototipos de inteligencia artificial gracias a Empresas 10X*. Obtenido de Intersoftware.org: <https://www.intersoftware.org.co/blog/veinte-empresas-del-pa%C3%ADs-cuentan-con-prototipos-de-inteligencia-artificial-gracias-empresas-10x>
- Kawalek, P., & Bayat, A. (2017). *Data a infrastructure*. Obtenido de <https://aura.abdn.ac.uk/bitstream/handle/2164/11906/DataAsInfrastructure.pdf;jsessionid=41EFF4A572CA2446825F967F14E48197?sequence=1>
- KbInformatika. (s.f.). *Data quality with data governance*. Obtenido de Confusing Data Quality with Data Governance: <https://kb.informatika.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*. Obtenido de <https://books.google.com.co/books?id=CpeAYWaTScYC>
- Letouzé, E., Pentland, A., Loaiza, I., Ricard, J., Clavijo, A., & Silva, D. (2019). *Recomendaciones e insumos para la definición de una estrategia nacional de Big Data*.
- Lidman, J., Kinnari, T., & Rossi, M. (2016). Business Roles in the Emerging Open-Data Ecosystem. *IEEE SOFTWARE*.
- López, L. (31 de Agosto de 2020). *Inteligencia Artificial: 4 ejemplos de su uso en Colombia*. Obtenido de Mundo: <https://folou.co/mundo/inteligencia-artificial-ejemplos-empresas-colombia/>
- Lotero Hernández, N. (s.f.). *Clasificación de los datos personales e implicaciones legales*. Medellín: Repositorio Universidad Pontificia Bolivariana.

- Martínez, S., Rodríguez, G., Ron, J., Triulzi, F., & Santana, P. (2017). *Hacia la utilización del Big Data en ciencias sociales: un enfoque histórico*. Obtenido de <https://es.scribd.com/document/443328023/Martinez-et-al-2017>
- Mckinsey. (2021). *data-ecosystems-made-simple*. Obtenido de <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/tech-forward/data-ecosystems-made-simple>
- Mertilos, A. (2015). *Development of a capability maturity model for bog data governance*. Obtenido de <https://datasciencebe.files.wordpress.com/2015/07/development-of-a-capability-maturity-model-for-big-data-governance.pdf>
- Mineducación . (13 de Febrero de 2017). *Cooperación Multilateral*. Obtenido de Gov.co: https://www.mineducacion.gov.co/1759/w3-article-150509.html?_noredirect=1
- MinEducación. (2017). *Introducción al MNC*. Bogotá.
- Ministerio de Ciencia Tecnología e Innovación. (s.f.). *Reconocimiento de actores*. Sistema de gestión de calidad colciencias.
- Ministerio de Tecnologías de la Información y Comunicaciones. (2020). *Datos abiertos*. Obtenido de <https://herramientas.datos.gov.co/es/content/preguntas-frecuentes>
- Ministerio de Tecnologías de la Información y las Comunicaciones. (23 de Abril de 2020). *Agremiaciones y Asociaciones*. Obtenido de Gov.co: <https://www.mintic.gov.co/portal/inicio/Atencion-y-Servicio-a-la-Ciudadania/Transparencia/135688:Agremiaciones-y-Asociaciones>
- MinTic. (2020). *CONVOCATORIA CIENCIA DE DATOS 2020 -I*. Bogotá.
- MinTIC. (2020). *Software Público Colombia: Gobierno Digital*. Obtenido de <https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Software-libre/>
- MinTic. (19 de Julio de 2021). *Identificación y medición de brechas de capital humano*. Obtenido de MinTic.gov.co: https://mintic.gov.co/portal/715/articles-159493_Infografia_con_los_resultados_mas_relevantes_del_sector.pdf
- MinTic. (18 de Julio de 2021). *Iniciativas* . Obtenido de MinTic.gov.co: <https://mintic.gov.co/portal/vivedigital/612/w3-propertyname-509.html>

- MinTIC. (2021). *Plan Nacional de Infraestructura de datos*.
- MinTIC. (2022). *Talento digital*. Obtenido de <https://talentodigital.mintic.gov.co/734/w3-propertyvalue-179944.html>
- MinTIC. (s.f.). *Guía para el uso y aprovechamiento de Datos Abiertos en Colombia*. El futuro digital es de todos.
- Moore, M. (1995). *Creating Public Value: Strategic Management in Government*. Obtenido de Cambridge, Harvard.
- Naciones Unidas. (2020). *Data Strategy of the Secretary-General for Action by Everyone, Everywhere*.
- OCDE / Eurostat. (2018). *Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities*.
- OCDE. (2015). *Data-Driven Innovation*. Paris: OCDE Publishing. Obtenido de https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page4
- 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). *Analytics for integrity: Data-Driven Approaches for Enhancing corruption*. Obtenido de Assessing the value of analytics: <https://www.oecd.org/gov/ethics/analytics-for-integrity.pdf>
- OCDE. (2019). *El camino para convertirse en un sector público basado en datos*. Obtenido de <https://www.oecd-ilibrary.org/sites/059814a7-en/index.html?itemId=/content/publication/059814a7-en>
- OCDE. (2019). *Enhancing Access to and Sharing of Data*. Obtenido de Reconciling Risks and Benefits for Data Re-use across Societies: <https://www.oecd.org/publications/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm>
- OCDE. (2020). *Going digital*. Obtenido de Data in the digital age: <https://www.oecd.org/going-digital/data-in-the-digital-age.pdf>

- OCDE. (Diciembre de 2020). *Perspectives on the value of data and data flows*. Obtenido de <https://www.oecd.org/publications/perspectives-on-the-value-of-data-and-data-flows-a2216bc1-en.htm>
- OCDE. (2021). *Recommendation of the Council on Enhancing Access to and Sharing data*. Obtenido de <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463>
- OCDEb. (2019). *OECD Open, Useful and Re-usable data (OURdata) Index: 2019*.
- Oliveira, M. L., Barros Lima, G. d., & Farias Lóscio, B. (2019). *Investigations into Data Ecosystems: a systematic mapping study*. Springer.
- Oliveira, M., & Barrios Lima, G. (2019). *Investigations into Data Ecosystems: a systematic mapping*. Obtenido de Knowledge and Information Systems.
- Oliveira, M., & Farias Loscio, B. (2018). *What is a data ecosystem*. Obtenido de <https://docplayer.net/185533503-What-is-a-data-ecosystem.html>.
- Olivera, M. I., Barros, G. d., & Farias, B. (2017). *Investigations into Data Ecosystems: a systematic mapping*.
- Open Data Insitute. (2018). *Mapping Data Ecosystems*. Obtenido de <https://theodi.org/article/mapping-data-ecosystems/>
- Open Data Watch. (2019). *The Data Value Chain: Moving from Production to Impact*. Obtenido de <https://opendatawatch.com/publications/the-data-value-chain-moving-from-production-to-impact/>
- Pérez Porto, J., & Gardey, A. (2021). *Definición de Organismo gubernamental*. Obtenido de Definición.de: <https://definicion.de/organismo-gubernamental/>
- Rockstart. (2020). *Top 100 Startups*. Obtenido de https://startupcol-recursos.s3.amazonaws.com/3mvTjdZSenv4jxyEstudio_TOP100_startups_%20%281%29.pdf
- Rodríguez Ospino, L. A., Gómez Gaviria, D., García Montaña, A., Ríos García, D. P., Durán Pabón, I. M., & Barrero Vanegas, V. (2021). *Modelo de explotación de datos para las entidades públicas*. Bogotá.
- Saltz, J., & Shamshurin, I. (2016). *Big data team process methodologies: A literature review and the identification of key factors for a project's success*. Obtenido de

- IEEE International conference:
<https://ieeexplore.ieee.org/document/7840936>
- The data administrator Newsletter. (2016). *non-invasive-framework-for-data-governance-implementation-details*. Obtenido de <https://tdan.com/non-invasive-framework-for-data-governance-implementation-details-part-2/20130>
- The ODI. (2018). *Report Mapping data ecosystems*. Obtenido de <https://theodi.org/wp-content/uploads/2018/03/374682869-ODI-report-Mapping-data-ecosystems.pdf>
- The ODI. (2022). *Data ecosystems to solve the world's biggest challenges*. Obtenido de <https://theodi.org/article/data-ecosystems-to-solve-the-worlds-biggest-challenges/>
- UN Global Pulse. (2011). *Data Philanthropy: Public & Private Sector Data Sharing For Global Resilience*. Obtenido de <https://www.unglobalpulse.org/2011/09/data-philanthropy-public-private-sector-data-sharing-for-global-resilience/>
- Unión Europea. (2021). *Establishment of Sustainable Data ecosystem*. Obtenido de Recommendation for the evolution of spatial data infrastructures.
- United Nations. (2020). *Data Strategy of the Secretary-General for Action by Everyone, Everywherewith Insight, Impact and Integrity*.
- Vieira, H. (2016). *The economic value of personal data for online platforms, firms and consumers*. Obtenido de <https://blogs.lse.ac.uk/businessreview/2016/01/19/the-economic-value-of-personal-data-for-online-platforms-firms-and-consumers/>
- World Economic Forum. (2021). *Data-driven Economies: Foundations for Our Common Future*.
- Yanseen, E., Bouzembrak, Y., Hendriksen, P., & Staats, M. (2017). *Big data in food safety: A overview, Critical Reviews in Food Science and Nutrition*. Obtenido de <https://www.tandfonline.com/doi/full/10.1080/10408398.2016.1257481>
- Zdonek, D., & Krol, K. (2020). *Analytics Maturity Models: An Overview*. Obtenido de https://www.researchgate.net/publication/339672162_Analytics_Maturity_Models_An_Overview