Cepei

El estado del ecosistema de datos en Colombia: análisis y recomendaciones para su desarrollo

Con el apoyo de:

Global Partnership for Sustainable Development Data

Reproducción

Esta publicación puede ser reproducida en su totalidad o en parte y en cualquier forma para fines educativos o no lucrativos sin un permiso especial del titular de los derechos de autor, siempre y cuando se cite la fuente. Cepei agradecería recibir un ejemplar de cualquier publicación que utilice la presente publicación como fuente.

Referencia sugerida

Cepei. (2022). *El estado de los ecosistema de datos en Colombia: análisis y recomendaciones para su desarrollo*. Cepei.

Autor principal
Andrés Clavijo

Investigador asociado
Hernán Muñoz

Edición
Alexandra Roldán

Corrección de estilo
Vivian González

Diseño y composición
Diana Paola Echeverría

Créditos
© Mapas, fotografías e ilustraciones, como se especifican en el texto.
- Creative Commons -

Tabla de contenido

1. Resumen ejecutivo 4

2. Apartado metodológico 9

2.1 Revisión de la literatura 9

2,2 Entrevistas grupales 11

3. Ecosistema de datos de Colombia 12

4. Principales innovaciones en los últimos años 18

5. Hallazgos y lecciones aprendidas 24

6. Conclusiones y recomendaciones para el desarrollo del ecosistema de datos 27

Anexos 32

1. Resumen ejecutivo

La cuarta revolución tecnológica ha generado un cambio en la forma como se consumen, producen y transfieren datos entre individuos, organizaciones y países. El abaratamiento de las tecnologías que permiten analizar comportamientos o fenómenos sociales, económicos y ambientales, ha llevado a que los datos se conviertan en un bien valorado y negociable.

El lugar donde estos procesos de producción, consumo y transferencia se generan, se asocia con el nombre de *ecosistemas de datos*. Aunque no existe una definición globalmente aceptada, los ecosistemas de datos comprenden procesos de innovación, desarrollo y crecimiento, que se encuentran conectados a su grado de madurez y desarrollo. En el caso de Colombia, el Departamento Administrativo Nacional de Estadística (DANE) propone una definición desde el Plan Estadístico Nacional 2020-2022, donde el ecosistema de datos para Colombia se define como la conformación de un conjunto de productores y usuarios de información estadística, registros administrativos e información proveniente de fuentes alternativas, así como por lineamientos, normas y estándares para la producción y difusión de estadísticas.

La Agenda 2030 para el Desarrollo Sostenible plantea una serie de desafíos en materia de política pública, que invita a reconocer la necesidad de contar con un marco de indicadores para el monitoreo y seguimiento de su implementación; lo anterior, aunado a la consolidación de otras agendas globales y regionales de desarrollo, ha llevado a exaltar la importancia de contar con ecosistemas de datos sólidos, donde exista un papel preponderante de los sistemas estadísticos nacionales y las capacidades estadísticas.

El Foro Mundial de Datos de las Naciones Unidas (UN-WDF) y el Plan de Acción Global de Ciudad del Cabo (CT-GAP), lanzado en 2017, son dos hitos en la agenda global de datos que exponen estos nuevos retos. En 2024, Colombia recibirá por primera vez el UN-WDF y será una oportunidad única para visibilizar esta agenda y construir un camino entre los actores nacionales y territoriales hacia ecosistemas de datos más sostenibles y maduros que contribuyan al cumplimiento de los Objetivos de Desarrollo Sostenible (ODS).

Con el apoyo de:
Global Partnership for Sustainable Development Data

Cepei, con el apoyo de la Alianza Global de Datos para el Desarrollo Sostenible (GPSDD, por sus siglas en inglés), realizaron el presente estudio sobre el mapeo de actores y prioridades de datos, con el objetivo de brindar recomendaciones frente a los avances, necesidades y futuras líneas de acción para el fortalecimiento de las capacidades de datos en Colombia. El objetivo de este trabajo es obtener una caracterización general del ecosistema de datos colombiano relacionado con las estadísticas oficiales, identificar sus principales actores e instrumentos regulatorios, y conocer las tendencias y prioridades de datos, explorando su conexión con los ODS y la agenda de desarrollo estadístico actual. Los hallazgos y recomendaciones serán una hoja de ruta para apoyar al gobierno y a los hacedores de política estadística, datos e información, en el fortalecimiento y consolidación de un ecosistema de datos maduro, y de ese modo canalizar los esfuerzos de cooperación internacional y el financiamiento para hacer frente a las demandas de la Agenda 2030.

Para cumplir con este propósito, se realizó una revisión de literatura sobre ecosistemas de datos y entrevistas[1] con actores nacionales y territoriales, organizaciones de la sociedad civil y el sector privado, con el fin de recopilar los distintos avances y perspectivas en relación con el funcionamiento del ecosistema de datos, identificando los avances, desafíos y oportunidades que enfrentan, así como sus prioridades para los próximos años.

Colombia se ha caracterizado por tener uno de los ecosistemas de datos más activos y dinámicos de la región. En las últimas dos décadas, la política de datos del país ha estado liderada por tres entidades: el Departamento Nacional de Planeación (DNP), el Ministerio de Tecnologías de la Información y Comunicaciones (MinTIC) y el Departamento Administrativo Nacional de Estadística (DANE). Recientemente, se ha evidenciado una voluntad por fortalecer la coordinación entre estos actores, que ha permitido avanzar hacia un ecosistema maduro. Otras entidades se han sumado a esta agenda, como la Superintendencia de Industria y Comercio, encargada de diseñar y reglamentar las políticas asociadas al manejo de datos personales en el país, y el Departamento Administrativo de la Presidencia de la República (DAPRE), quién asumió el liderazgo de varios temas de datos.

1. Cepei y GPSDD agradecen el tiempo, disposición e información proporcionada por las entidades y organizaciones participantes de este estudio, especialmente al DANE quien apoyó las convocatorias para realizar las entrevistas.

El avance hacia una fuerte consolidación de un sistema estadístico nacional[2], la puesta en marcha de una cultura orientada a datos abiertos, los adelantos en materia de interoperabilidad entre entidades del gobierno, la elaboración de una política de big data para el gobierno y el sector privado colombiano y, más recientemente, el desarrollo de marcos éticos para el uso de inteligencia artificial, así como nuevos marcos que promueven la creación de mercados y portales de datos, son claros ejemplos de un ecosistema de datos más maduro.

Hallazgos principales sobre el ecosistema de datos de Colombia

A continuación, se resumen los principales hallazgos identificados. En el cuerpo del documento se encuentran a mayor detalle, siendo estos los más relevantes:

Alcance

• Se evidencia que el Sistema Estadístico Nacional (SEN) es parte central del ecosistema de datos colombiano. Este brinda estándares, buenas prácticas y espacios de articulación a nivel temático, sectorial y territorial.

• El ecosistema de datos colombiano va más allá de las estadísticas oficiales para abarcar los datos producidos, intercambiados y usados por los sectores público y privado, organizaciones civiles y organismos multilaterales para una variedad de propósitos.

Marco legal y normativo

• El SEN cuenta con un marco normativo reciente, que le permite al DANE tener una mejor articulación y gobernanza de las estadísticas y datos, a partir de salas temáticas especializadas.

• Se ha venido avanzando en la integración y articulación de diferentes marcos normativos asociados a datos, información y estadísticas al interior del ecosistema de datos de Colombia. Sin embargo, no existe un único marco unificador y armonizador de las políticas de datos, información y estadísticas.

2. Actualmente el DANE lo denomina SEN 2.0.

Mapeo de actores

• En el nivel nacional, se identifican entidades que tradicionalmente han tenido un rol de formulador de política de datos, información o estadísticas: DANE, DNP, MinTIC, Superintendencia de Industria y Servicios e Instituto Geográfico Agustín Codazzi.

• En el nivel nacional, los principales productores de datos e información estadística son: DANE, entidades de gobierno nacional, Banco de la República y otras entidades independientes, organismos de control, entidades de la rama judicial y organización electoral, sector privado y organizaciones de la sociedad civil.

Operación y dinámica

• Aún se siguen presentando barreras en el acceso[3] a registros administrativos para aprovechamiento estadístico. A pesar de los marcos normativos existentes y de las salas y espacios para abordar los temas en el SEN, muchas entidades no entregan su información por desconfianza y desconocimiento del SEN.

• El SEN 2.0 le brinda al ecosistema de datos la directriz que las estadísticas oficiales[4] son de uso obligatorio para fines de política pública y, además, son las que se deben reportar a organismos internacionales. Sin embargo, en la práctica, no todas las políticas públicas se apoyan exclusivamente de estadísticas oficiales.

• A nivel de ecosistema de datos, se viene avanzando en estándares de interoperabilidad entre las entidades del sector público. Principalmente enfocado a mejorar la prestación de los servicios públicos a los ciudadanos. El modelo que se sigue es el de Estonia, que utiliza la tecnología X-Road.

• Rezagos sustantivos en producción y calidad de datos de servicios públicos, justicia y seguridad y defensa, y medio ambiente.

• **Falta una mayor articulación de la agenda ODS entre los actores del ecosistema de datos de Colombia.** En específico, se requiere una mayor difusión y coordinación de los avances, procedimientos y metodologías utilizadas para la medición y monitoreo y evaluación de estos indicadores en el país. Se evidencian diferencias significativas en los reportes de avance de cumplimiento de estos indicadores entre la OCDE y el DNP.

3. Se identifican principalmente cuatro: 1) desconocimiento de la normativa, 2) desconocimiento de técnicas y procesos de anonimización, 3) poca claridad en la articulación normativa frente a temas de hábeas data, protección de datos personales y reserva estadística, y 4) mayor entendimiento sobre los beneficios y costos de acceder a esta información por parte de los actores involucrados.
4. El DANE define estadística oficial como estadísticas que se encuentran identificadas y priorizadas en el plan estadístico nacional, y que han sido evaluadas y certificadas en procesos de calidad.

Con el apoyo de:
Global Partnership for Sustainable Development Data

Recomendaciones para la acción

De acuerdo con estos hallazgos, Cepei propone cinco recomendaciones para abordar de manera coordinada estos temas, las cuales se exploran con mayor detalle en el capítulo final de este documento.

- Promover una alineación estratégica entre las políticas de información, datos y estadística con la Agenda 2030 y las futuras iniciativas de desarrollo gubernamental, en el CAD del SEN.

- Impulsar una gobernanza de datos más efectiva, medible y con enfoque de resultados basada en las experiencias internacionales en la materia.

- Fomentar activamente la gestión, coordinación e integración de datos con roles y funciones más claros de los actores del ecosistema.

- Cuidar y avanzar en el mapeo de distintas fuentes de datos del país.

Este documento se encuentra organizado en 6 secciones de la siguiente manera: la introducción, a modo de resumen ejecutivo; un apartado metodológico de revisión de fuentes y entrevistas; una descripción del ecosistema de datos de Colombia; las principales innovaciones recientes del ecosistema de datos colombiano; los principales hallazgos que resultan del estudio, y, finalmente, las recomendaciones y conclusiones.

Con el apoyo de:
Global Partnership for Sustainable Development Data

2. Apartado metodológico

La metodología desarrollada consistió de una revisión exhaustiva de literatura sobre ecosistemas de datos, y el desarrollo de 5 entrevistas grupales e individuales, realizadas a distintos grupos de expertos en datos, caracterizados por su rol de productores, consumidores y reguladores. Con el fin de obtener un panorama completo, se convocó a entidades públicas nacionales y territoriales, así como a actores no gubernamentales.

2.1 Revisión de la literatura

Uno de los primeros resultados de la revisión bibliográfica evidencia que aún no existe una definición consensuada sobre lo que es un ecosistema de datos. La mayoría de autores explican su origen a partir de la definición del ecosistema biológico. Asimismo, algunos señalan que la conceptualización del ecosistema de datos abarca elementos de otros tres: el ecosistema empresarial, digital y de software. A pesar de compartir esta característica común, los autores seleccionados construyen diferentes aproximaciones de lo que entienden por ecosistema de datos, y algunas de estas son específicas para ciertos campos, como la salud o el big data.

A continuación, se destacan algunas de las definiciones más citadas en la literatura académica para conceptualizar y caracterizar los ecosistemas de datos (ver tabla 1).

Tabla 1. Principales definiciones y características de los ecosistemas de datos

Nombre del artículo	Autores	Año	Definición	Características clave
What is a data ecosystem?	Oliveira and Lóscio	**2018**	Un ecosistema de datos es un conjunto de **redes** compuestas por actores autónomos que directa o indirectamente consumen, producen o proporcionan datos y otros recursos relacionados (por ejemplo, software, servicios e infraestructura). Cada actor desempeña uno o más roles y está conectado a otros actores a través de relaciones, de tal manera que la colaboración y competencia promueve la autorregulación del ecosistema de datos.	• Actores • Roles • Relaciones • Recursos
"Real-time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems"	Syed Iftikhar Hussain Shah, Vasilis Peristeras, Ioannis Magnisalis	**2019**	Un ecosistema de datos es un sistema socio-técnico que permite extraer valor de las cadenas de valor de los datos, respaldadas por organizaciones e individuos que interactúan.	**Se identifican 4 tipos de ecosistemas de datos, que pueden formarse alrededor de sistemas inteligentes:** 1. Ecosistema de datos dirigidos 2. Ecosistema de datos reconocido 3. Ecosistema de datos colaborativos 4. Ecosistema de datos virtuales
National SDG Review: data challenges and opportunities	Paris 21	**2019**	Un ecosistema es un grupo de múltiples comunidades de datos que interactúan entre sí compartiendo datos, conocimientos para mejorar los procesos y el uso de datos, y otras actividades relacionadas con los datos a lo largo de la cadena de valor de los datos.	• **Productores de datos** • **Usuarios de datos** • **Brokers** • **Objetos de datos**
United Nations National Quality Assurance Frameworks Manual for Official Statistics	United Nations, Department of Economic and Social Affairs.	**2019**	Un sistema en el que un número de actores interactúan entre sí para intercambiar, producir y utilizar datos. Un sistema puede entenderse como un conjunto de partes conectadas que forman un todo complejo.	
Data Ecosystems: Sovereign Data Exchange among Organizations. Report from Dagstuhl Seminar 19391	Cappiello, Gal, Jarke, and Rehof	**2020**	Los ecosistemas de datos son un tipo especial de ecosistema digital. Como tales, son sistemas distribuidos, abiertos y adaptativos con las características de ser autoorganizados, escalables y sostenibles. Si bien se centran en los datos, la principal preocupación con los ecosistemas de datos es compartir y aumentar el conocimiento, que es al mismo tiempo un problema de aprendizaje de datos no estructurados y heterogéneos, construcción de nuevas abstracciones y mapeos, oferta de servicios, incluidas consultas, e integración y transformación de datos. Todo esto debe garantizarse de manera dinámica y escalable, manteniendo la coherencia, la evaluación de la calidad, la seguridad y la asequibilidad.	• **Conjuntos de datos** • **Operadores de datos** • **Metadatos** • **Mapeos**
"Government (Big) Data Ecosystem: Definition, Classification of Actors, and Their Roles"	Shah, Peristeras, and Magnisalis	**2021**	Un ecosistema de datos del gobierno es una red socio técnica de personas, procesos, tecnología, infraestructura, servicios de datos, estándares y políticas de registros básicos, procesos, organizaciones y recursos que trabajan conjuntamente para realizar funciones de datos (recopilación, integración, análisis, almacenamiento, intercambio, uso, seguridad y protección de datos y archivo) de los datos para garantizar una mejor prestación de servicios, apoyar a las empresas, promover la administración basada en datos y el gobierno abierto, impulsar la economía de datos y la creación de políticas para beneficiar a los ciudadanos, las empresas y los propios organismos gubernamentales.	**Se propone una clasificación holística de cuatro dimensiones de los actores del ecosistema de (big) data gubernamental y sus funciones:** 1. Editor de datos 2. Entidades comerciales 3. Proveedores de servicios de apoyo 4. Usuarios/ consumidores

Fuente: elaboración propia a partir de la revisión de literatura, 2022

Según las definiciones anteriores, los autores coinciden en que un ecosistema de datos está compuesto por diferentes tipos de intercambios de datos y de actores que entablan diversas relaciones y cumplen distintos roles. También coinciden en que estos actores desempeñan funciones de colaboración y competencia, que promueven y autorregulan el ecosistema. Además, en muchos casos son actores autónomos que trabajan de manera conjunta y espontánea, en un marco cuasi anárquico, que se regula gracias a procesos e interacciones, con el fin de obtener valor de los datos. En resumen, el intercambio de datos, los actores y sus roles, y las relaciones entre estos, son características comunes que comparten la mayoría de las definiciones de ecosistema de datos[5] .

2.2 Entrevistas grupales

Las entrevistas se realizaron en el marco de tres reuniones con los actores del ecosistema de datos clasificadoss por su pertenencia institucional de la siguiente forma: 1) entidades públicas de orden nacional, 2) entidades públicas subnacionales / territoriales, 3) entidades no gubernamentales (ver las preguntas en el anexo).

El mapeo de actores se encuentra también en el anexo de este documento. Este listado fue construido a partir de conversaciones con el DANE. El criterio empleado fue el grado de relevancia de la entidad dentro del ecosistema de datos.

Se consideró si la entidad es formuladora de políticas de datos, productora clave de información estadística o si ha participado activamente en la construcción de planes estadísticos sectoriales o territoriales.

5. Desde el punto de vista de sistemas complejos, aparecen otros atributos relevantes en los ecosistemas de datos como el valor asociado a los datos y las formas colaborativas y regulatorias para obtenerlo.

3. Ecosistema de datos de Colombia

Esta sección aborda la descripción del ecosistema de datos colombiano desde el punto de vista del avance regulatorio que ha tenido el país recientemente en materia de datos y estadísticas. Se aleja de la visión tradicional de listar productores y usuarios y describir sus tipos de relaciones existentes. En su lugar, se da un mayor valor a la forma como los distintos actores que lideran los temas de datos, estadísticas e información han venido apareciendo, interactuando y dejando huella para lograr obtener valor a partir de los datos. Además, se detalla el avance de la medición y reporte de los ODS, resultado clave, pero no único, en la madurez de un ecosistema de datos.

En Colombia, el marco de gobierno de los datos y la información está determinado principalmente por tres artículos de la Constitución. Por un lado, los artículos 20 y 74 se refieren al acceso a la información. El primero, garantiza la libertad de informar y recibir información veraz e imparcial y, el segundo, establece el derecho a acceder a documentos públicos, salvo en los casos establecidos por la ley. Por otro lado, el artículo 15 garantiza el derecho a la intimidad personal y familiar, al buen nombre, y a conocer, actualizar y rectificar la información que se haya recogido sobre uno mismo en los bancos de datos y en los ficheros de las entidades públicas y privadas. Este artículo también establece que en la recolección, tratamiento y circulación de los datos se respetará la libertad y demás garantías consagradas en la Constitución.

Con base en estos artículos, la Ley 1266 de 2008, conocida como "Ley de Hábeas Data", desarrolló por primera vez los derechos constitucionales consagrados en los artículos 15 y 20 de la Constitución. Su artículo 2 estableció que el ámbito de aplicación de la Ley se referirá específicamente a la información financiera y crediticia y excluyó de su alcance aquellos datos mantenidos en un ámbito exclusivamente personal o doméstico y los que circulen internamente, que no sean suministrados a otras personas naturales o jurídicas. Luego, en 2009 la Ley fue reglamentada por el Decreto Nacional 1727 de 2009, que determinó la forma en que los operadores de las bases de datos de información financiera, crediticia, comercial y de servicios deben presentar la información de los titulares.

Posteriormente, se aprobó la Ley 1581 de 2012, y luego se reglamentó parcialmente por el Decreto Nacional 1377 de 2013, que estableció los principios y disposiciones

aplicables al tratamiento de los datos personales que se registren en cualquier base de datos pública o privada. El literal k), del artículo 17, y el literal f), del artículo 18 de la Ley, establecen como deber de los responsables y encargados del tratamiento de datos, adoptar un manual interno de políticas y procedimientos para garantizar el adecuado cumplimiento de la Ley, especialmente para la atención de consultas y reclamos por parte de los titulares. Posteriormente, y para facilitar la implementación de la Ley y del manual de políticas para el tratamiento de datos, el Decreto Nacional 1074 de 2015 también reglamentó parcialmente la Ley, al establecer y adicionar requisitos para el adecuado tratamiento de los datos personales registrados en bases de datos públicas y privadas (Departamento Administrativo de la Función Pública, 2016, p. 5).

Luego, se aprobó la Ley 1712 de 2014 sobre transparencia y regulación del derecho de acceso a la información pública nacional, que tiene por objeto regular el derecho de acceso a la información pública, los procedimientos para el ejercicio y garantía de este derecho, y las excepciones a la publicidad de la información (Osorio-Sanabria, Amaya-Fernández & González-Zabala, 2021, p. 158). Esta nueva Ley, que posteriormente fue reglamentada por el Decreto Nacional 103 de 2015, estableció nuevas disposiciones, principios, conceptos y procedimientos para el ejercicio y garantía del mencionado derecho. Además, aclaró temas relacionados con la gestión de la información pública en cuanto a su divulgación, publicación, recepción, clasificación y reserva, así como el desarrollo de instrumentos de gestión y seguimiento de la información (Departamento Administrativo de la Función Pública, 2016, p. 6). En este sentido, la Ley reguló la política de datos abiertos del gobierno colombiano, definió lo que el Estado colombiano entiende por datos abiertos y estableció la obligación de las entidades públicas de divulgar los datos abiertos (Ministerio de Tecnologías de la Información y Comunicaciones [MinTIC], 2019, p. 5).

Este marco legal ha sido apoyado recientemente por una serie de políticas relacionadas con datos, las cuales buscan crear un marco de gobernanza de datos e información mucho más fuerte en el país. En Colombia, la importancia de los datos como un activo que permite la generación de valor social y económico no fue reconocida explícitamente hasta el Plan Nacional de Desarrollo (PND) 2014-2018 (Departamento Nacional de Planeación [DNP], 2018, p. 3). Y desde entonces se ha incrementado la creación de políticas de datos.

Una de estas políticas es la "Definición de la estrategia de Big Data para el Estado colombiano y el desarrollo de la industria de Big Data en Colombia" (2017), la cual fue dividida en 4 documentos y desarrollada con el apoyo de la Alianza Data-Pop y el MIT Media Lab. Estos documentos estudiaron el estado actual del gobierno en materia de datos (su marco legal, infraestructura, capital humano, inversiones, etc.), para luego identificar áreas prioritarias y acciones específicas que generen las sinergias necesarias en beneficio de cuatro objetivos estratégicos: 1) fortalecer el ecosistema de big data en el país, 2) desarrollar capacidades analíticas para un mayor y mejor uso de los datos, 3) promover la economía digital, y 4) promover la adopción de soluciones de big data e IA para la administración pública (Clavijo et al., 2017, p. 8).

También, se encuentra el Documento CONPES 3920 de 2018 sobre Política Nacional de Aprovechamiento de Datos (big data), un documento de política pública que se elabora para resolver problemas nacionales transversales, en cuya solución participan varios sectores y su formulación es coordinada por el Consejo Nacional de Política Económica y Social (Ministerio de Ciencia, Tecnología e Innovación, s.f.). El objetivo principal de esta política fue incrementar el uso de los datos en Colombia, desarrollando las condiciones para que sean gestionados como activos que generen valor social y económico. Para ello, la política contaba con cuatro objetivos específicos: 1) masificar la disponibilidad de datos de las entidades públicas accesibles digitalmente, usables y de calidad, 2) generar seguridad jurídica para la explotación de los datos, 3) contar con el capital humano para generar valor con los datos, y 4) generar cultura de datos en el país.

Otro elemento relevante a considerar es la Política de Gobierno Digital, que tiene como objetivo principal promover la utilización y aprovechamiento de las tecnologías de la información y comunicación para consolidar un Estado y ciudadanos competitivos, proactivos e innovadores, que generen valor público en un entorno de confianza digital.

La Política definió cinco propósitos: 1) habilitar y mejorar la prestación de servicios digitales confiables y de calidad, 2) lograr procesos internos, seguros y eficientes mediante el fortalecimiento de las capacidades de gestión de la información tecnológica, 3) tomar decisiones basadas en datos gracias al incremento en el uso y aprovechamiento de la información, 4) empoderar a los ciudadanos a través de la consolidación de un Estado Abierto, y 5) promover el desarrollo de territorios y ciudades inteligentes para resolver retos y problemas sociales, mediante el uso de las TIC (Gobierno Digital, s.f.).

El Documento CONPES 3975 de 2019 sobre Política Nacional de Transformación Digital e Inteligencia Artificial también es relevante. Esta política busca aumentar la generación de valor social y económico a través de la transformación digital de los sectores público y privado para que Colombia aproveche las oportunidades y enfrente los retos relacionados con la cuarta revolución industrial (Industria 4.0). Para lograr este objetivo, la política propone, primero, reducir las barreras que impiden la incorporación de las tecnologías digitales, tanto en el ámbito empresarial como en las entidades del Estado. Segundo, crear condiciones propicias para la innovación digital, privada y pública, como medio para incrementar la generación de valor económico y social a través de nuevos procesos y productos. En la tercera etapa, la política busca fortalecer las competencias del capital humano que faciliten la inserción de la sociedad colombiana en la Industria 4.0. Finalmente, se desarrollan condiciones habilitantes con el fin de preparar a Colombia para los cambios económicos y sociales que traerá la inteligencia artificial (IA), reconociendo esta tecnología como un acelerador de la transformación digital.

En el marco de esta política, el gobierno colombiano también desarrolló la Misión Colombiana de Expertos en IA y su asociado Marco Ético para la IA en Colombia (2021), apoyados por el Banco de Desarrollo de América Latina (CAF) y el Banco Interamericano de Desarrollo (BID). La Misión desarrolla la comprensión pública de la IA y aborda las percepciones negativas para impulsar la confianza en esta tecnología entre las empresas y la sociedad. Igualmente, busca aumentar las habilidades y la capacidad en IA, incluyendo la diversidad de personas que estudian y trabajan en ella, y explora cómo desarrollar y desplegar marcos de intercambio de datos seguros, justos, legales y éticos (Mujica et al., s.f., p. 20). En ese sentido, el Marco Ético proporciona una guía de derecho indicativo

de recomendaciones y sugerencias a las entidades públicas, para abordar la formulación y gestión de proyectos que incluyan el uso de la inteligencia artificial (IA). Esta guía reúne las buenas prácticas internacionales y no es obligatoria para las entidades (Departamento Administrativo de la Presidencia de la República, 2021, p. 21).

En relación con el funcionamiento del ecosistema de datos para el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS), el Documento CONPES 3918 de 2018, establece la estrategia para la implementación de los ODS en Colombia. El texto define un conjunto de indicadores y metas para el seguimiento a la implementación de los Objetivos, sus respectivos responsables, el plan de fortalecimiento estadístico necesario para robustecer los sistemas de información, la estrategia de territorialización de los ODS y los lineamientos para la interlocución con actores no gubernamentales. Además, dispone el esquema de seguimiento, reporte y rendición de cuentas, y fija las líneas base para cada uno de los objetivos.

En relación con el fortalecimiento estadístico, el CONPES señala que el DANE, junto con otras entidades, adelantó un diagnóstico preliminar en 2018 sobre la disponibilidad de información del país para responder a los indicadores globales definidos por el Grupo Interagencial y de Expertos sobre Indicadores ODS (IAEG-SDG, por sus siglas en inglés). De este ejercicio se estableció que para un 54% de indicadores se contaba con información, para el 30% se contaba con información parcial o se requerían mejoras metodológicas y para el 16% no se contaba con información o no existía aún una metodología definida para obtenerla. Con base en esta evaluación, se definieron tres líneas de acción para la implementación del marco de monitoreo global de los ODS en el país: 1) la articulación del Sistema Estadístico Nacional (SEN 2.0) para los ODS, 2) el trabajo articulado con el Sistema de las Naciones Unidas en Colombia y 3) el trabajo en el marco de la Global Partnership for Sustainable Development Data (GPSDD).

Respecto al avance en la consecución de las metas de los ODS, de acuerdo con el DNP, el porcentaje de avance general con corte a diciembre de 2020 es de 72,1%, en relación con las metas fijadas por el Gobierno nacional para 2030, en su gran mayoría establecidas en el documento CONPES 3918 (DNP, 2021, p. 25). Según la entidad, los objetivos con mayor avance son el ODS 15 sobre vida de ecosistemas

terrestres (100%), el ODS 17 sobre alianzas para lograr los objetivos (100%), el ODS 14 sobre vida submarina (99,5%), el ODS 12 sobre producción y consumo responsable (95,8%), y el ODS 9 sobre industria, innovación e infraestructura (94,6%). No obstante, hay otros objetivos como el ODS 3 sobre salud y bienestar, donde el avance ha sido bajo (55,4%), y otros como el ODS 2 sobre hambre cero del que no se tiene reporte.

Según la Organización para la Cooperación y el Desarrollo Económicos (OCDE), Colombia logró 11 de las 117 metas de los ODS para las que se dispone de datos comparables y, según las tendencias más recientes, se espera que para 2030 logre 12 metas más (OCDE, 2022, p. 1). De acuerdo con la OCDE el país está por alcanzar la mayoría de las metas relacionadas con la satisfacción de las necesidades básicas y la implantación de los instrumentos y marcos de política pública mencionados en la Agenda 2030. También señala que uno de los puntos fuertes de Colombia se encuentra en la categoría Prosperidad, en particular en el ODS 7 sobre energía asequible y no contamintante. Sin embargo, afirma que el desempeño entre los Objetivos es muy desigual y que la disponibilidad de datos continúa siendo un reto para alcanzar la totalidad de las metas, puesto que los datos disponibles para medir los indicadores aún dejan por fuera 52 de las 169 metas de la Agenda 2030.

Al comparar la información sobre el avance que reporta la OCDE con la presentada por el DNP, se evidencia que si bien para 2020 hubo progreso en la consecución de las metas nacionales de los ODS, aún hay un camino por recorrer respecto a las metas globales. De acuerdo con la OCDE, en 2030, año límite para cumplir con la Agenda 2030, Colombia solo alcanzaría 23 de las 117 metas de las que se dispone de datos comparables. Adicionalmente, el país no podría reportar sobre otras 52 metas debido a la falta de datos. Por lo tanto, es necesario robustecer y ampliar el alcance del ecosistema de datos colombiano y replantear las metas nacionales de manera que logren un mayor progreso en cada uno de los objetivos.

4. Principales innovaciones en los últimos años

En los últimos años, el DANE, el DNP, el MinTIC y el DAPRE han sumado esfuerzos, con el fin de avanzar en la formalización de un ecosistema de datos, desde cada una de las perspectivas institucionales de las organizaciones participantes, como puede ser el liderazgo del DANE en lo que refiere a las estadísticas oficiales.

El marco que permitió un proceso de innovación en el aprovechamiento de datos en el país fue la creación del Sistema Estadístico Nacional (SEN 2.0), a partir del artículo 155 de la Ley 1955 de 2019 y su Decreto reglamentario 2404 de 2019[6]. Este permitió fortalecer la producción de información estadística oficial y satisfacer las nuevas demandas de datos e información.

Para lograr las acciones de producción y coordinación propuestos en el SEN 2.0, el Plan Estadístico Nacional 2020 - 2022 (PEN) responde con nuevos lineamientos estratégicos y pasos a seguir en la ruta de la producción y difusión de información estadística. Este documento formula por primera vez el concepto de ecosistema de datos del SEN 2.0. El siguiente diagrama muestra las distintas capas de organización del sistema.

El ecosistema se compone de tres capas: una asociada al sistema estadístico y conformada por el DANE, entidades de gobierno, organismos de control, Banco de la República, sector privado y entidades de la rama judicial y organización electoral; otra conformada por el subsistema departamental y municipal, y una tercera constituida por los usuarios de la información estadística.

6. El SEN ya había sido planteado en la Ley del PND de 2014 y 2018, sin embargo, el SEN 2.0 establecido en el Art 155 de la Ley 1955, sí es una innovación en la medida que fortalece la gestión del SEN al establecer de manera precisa sus miembros, los enfoques que deben asegurarse en la producción y difusión de estadísticas, el establecimiento de la inoponibilidad de la reserva tributaria, la creación del CASEN y el establecimiento del Programa de Fortalecimiento Estadístico Territorial.

Con el apoyo de:
Global Partnership for Sustainable Development Data

Diagrama 1. Ecosistema de datos SEN 2.0

Fuente: DANE

Con el fin de mejorar la capacidad de producción de información en el nivel territorial (municipios y departamentos), el DANE desarrolló dos instrumentos que permiten identificar las fortalezas y debilidades en la producción y gestión de la información estadística.

Un primer elemento fue el Índice de Capacidad Estadística Territorial (ICET), con el cual por primera vez el DANE recreó indicadores de capacidad estadística, siguiendo algunas metodologías como las del Banco Mundial y PARIS21, para crear líneas de base y coordinar el sistema estadístico a partir de acciones basadas en datos o evidencia. El segundo elemento, con el objetivo de difundir la política estadística en diferentes canales institucionales, fue la implementación de una política de gestión de información estadística a través del Departamento de la Función Pública. Dicha política dispone de instrumentos, herramientas y documentación de carácter técnico para que las entidades públicas los acojan y ajusten de acuerdo con sus capacidades y contexto institucional, con el propósito de impulsar la planeación estadística, el fortalecimiento de los registros administrativos y la calidad estadística. La Función Pública tiene la capacidad de monitorear aspectos administrativos, gestión y planeación de las entidades públicas del gobierno colombiano.

Es de resaltar la manera como se ha consolidado la coordinación del SEN 2.0. Este modelo contiene un consejo asesor técnico del SEN 2.0 (CASEN), que tiene las funciones de asesorar y evaluar el desarrollo del sistema y la producción de estadísticas en el nivel nacional, a través de la emisión de recomendaciones y propuestas concretas frente a metodologías estadísticas, métodos y herramientas tecnológicas, enfocadas en mejorar la producción de estadísticas oficiales. Por otra parte, el CASEN se relaciona con el Comité Administrativo de Datos (CAD), el cual se encarga de fomentar acuerdos para lograr accesos a datos bajo ambientes seguros, responsables y éticos entre los integrantes del SEN 2.0. La instancia que hace el seguimiento a las estadísticas asociadas a crecimiento económico, inflación y empleo es el comité de seguimiento a estadísticas estratégicas (CSEE). Este se articula con las anteriores instancias a partir de la solicitud de asesorías al CASEN sobre análisis de estadísticas de desempeño económico y al CAD sobre temas de articulación e integración de datos. El CASEN también recibe propuestas de los comités estadísticos sectoriales (CES), los cuales formulan acciones asociadas a las necesidades de información estadística identificadas y discutidas en el nivel sectorial.

Esta estructura de articulación ha favorecido al DANE para avanzar en la gestión de un sistema de información e identificación de brechas de información en el país. Además, le ha permitido involucrar a otros actores del ecosistema que son

fuertes productores de información y, a su vez, potenciales multiplicadores de estándares y buenas prácticas estadísticas. Sin embargo, aún continúa la discusión por la aprobación de la Ley Estadística en la Cámara de Representantes, la cual, de ser aprobada, sería el primer marco legal constituido como norma jurídica, que establecería de manera tangible un mecanismo a través del SEN para que el ecosistema de datos del país se articule y gestione mejor.

Por otro lado, el MinTIC, el DNP y el DAPRE lanzaron a finales de 2021 el Plan Nacional de Infraestructura de Datos como estrategia nacional para avanzar en la definición e implementación de la infraestructura de datos de Colombia. Si bien el Plan no está terminado, continúa recibiendo insumos importantes para la discusión y formulación de estrategias. El marco legal que respalda este documento, también mencionado en el capítulo anterior, constituye el progreso del país en materia de protección de datos personales, acceso a la información pública, habeas data y protección comercial. El Plan de Infraestructura de Datos contempla una alineación estratégica con el SEN 2.0, para que los avances y acciones del Plan se puedan materializar de manera coordinada y blindada por las mejores prácticas en calidad, producción y gestión de datos para aprovechamiento estadístico. Además, contempla la articulación con la Infraestructura Nacional de Datos Espaciales (ICDE), liderada por el Instituto Geográfico de Colombia Agustín Codazzi.

El Plan de Infraestructura se fundamenta en siete principios: 1) calidad de los datos, 2) fácil búsqueda, accesibilidad, interoperabilidad y reutilización, 3) seguridad y protección de los datos, 4) privacidad por diseño y por defecto, 5) confianza pública y gestión ética de los datos, 6) estandarización e interoperabilidad, y 7) sectorización estratégica. Adicional a los principios, se suman los propósitos del Plan: 1) aumentar la reutilización de los datos que integran la infraestructura de datos, 2) consolidar un sector público basado en datos, 3) consolidar espacios de intercambio de datos que impulsen la innovación en el país, 4) promover el desarrollo e integración de tecnologías emergentes, 5) posicionar el modelo de gobernanza de los datos para la consolidación de una economía digital guiada por datos, 6) posicionar al país como un referente en el uso de los datos para el desarrollo de la economía digital, y 7) construir un entorno de confianza pública para el aprovechamiento y protección de los datos. La combinación de los principios y propósitos dejan ver de manera clara el tipo de madurez que el ecosistema de datos de Colombia pretende alcanzar.

La combinación de los principios y propósitos dejan ver de manera clara el tipo de madurez que el ecosistema de datos de Colombia pretende alcanzar. En este punto se debe resaltar el enfoque que Colombia le da a los datos como infraestructura (ver diagrama 2). En una lectura de arriba hacia abajo, primero se encuentra el **marco normativo o lega**l, el cual permite el desarrollo de políticas, estrategias o lineamientos de acción que fomentan la gestión y aprovechamiento de datos. Luego, **la clasificación de datos** en función de distintos atributos, es decir, si son de libre acceso (datos abiertos), maestros (claves o esenciales en alguna entidad), transaccionales (asociados a algún tipo de interacción entre el ciudadano y el estado), o de referencia. El siguiente es el **aprovechamiento de los datos,** que comprende los procesos de vida del dato, gobierno del dato y capacidades. Le siguen **la interoperabilidad y seguridad de los datos**, como elementos habilitadores y generadores de confianza y, por último, los insumos técnicos y tecnológicos que permiten materializar las estrategias y acciones de gestión y aprovechamiento de los datos.

Diagrama 2. Marco de entendimiento de la infraestructura de datos del Estado colombiano

Fuente: Plan Nacional de Infraestructura de Datos. Documento técnico y hoja de ruta. MinTic, DNP, DAPRE, 2021

El DNP se encuentra trabajando en un documento conceptual sobre caracterización de ecosistemas de datos para entender las dinámicas propias y evaluar cómo los actores evolucionan dentro del ecosistema. También busca generar taxonomías de ecosistemas para contar con un mejor entendimiento del ecosistema colombiano. Según el DNP, los ecosistemas de datos "se pueden entender como la interacción de diferentes agentes, que confluyen en un sistema de intercambio de datos con el propósito de generar valor económico o social de acuerdo con sus intereses en el sistema y a las funciones (roles) que desempeñan".

En función de estos avances, otro proceso de innovación para fortalecer el ecosistema de datos colombiano, es una plataforma digital impulsada por la CEPAL, con el apoyo de la Agencia Presidencial de Cooperación Internacional de Colombia, que busca mejorar la gestión de registros administrativos de entidades públicas y privadas para el aprovechamiento estadístico. Esta plataforma se lanzará a finales de 2022 en Colombia, Chile, Perú y Bolivia y permitirá a los proveedores de datos informar sobre la tenencia de registros administrativos (bases maestras), permitiéndoles abrir una hoja clínica de su registro. Igualmente, brindará información sobre su calidad o madurez, e informará posibles proyectos de innovación estadística que requieran su participación estratégica. Es una plataforma pensada para mejorar la gestión, identificación y caracterización del registro para su aprovechamiento estadístico.

5. Hallazgos y lecciones aprendidas

A partir de conversaciones y entrevistas con distintos actores del ecosistema de datos de Colombia, se listan los principales hallazgos y lecciones aprendidas, agrupados en cuatro categorías: alcance, marco legal y normativo, mapeo de actores y operación y dinámica:

Alcance

• Se evidencia que el Sistema Estadístico Nacional (SEN) es parte central del ecosistema de datos colombiano. Este brinda estándares, buenas prácticas y espacios de articulación a nivel temático, sectorial y territorial.

• El ecosistema de datos colombiano va más allá de las estadísticas oficiales para abarcar los datos producidos, intercambiados y usados por los sectores público y privado, organizaciones civiles y organismos multilaterales para una variedad de propósitos.

Marco legal y normativo

• El SEN cuenta con un marco normativo reciente que le permite al DANE tener una mejor articulación y gobernanza de las estadísticas y datos, a partir de salas temáticas especializadas.

• Actualmente se discute en el Congreso la Ley Estadística. Marco que permitirá al DANE actuar como coordinador del SEN.

• Colombia ha avanzando en la integración y articulación de diferentes marcos normativos asociados a datos, información y estadísticas al interior del ecosistema de datos de Colombia. Sin embargo, no existe un único marco unificador y armonizador de las políticas de datos, información y estadísticas.

Mapeo de actores

• En el nivel nacional, se han identificado entidades que tradicionalmente han tenido un rol de formulador de política de datos, información o estadísticas como el DANE, el DNP, el MinTIC, la Superintendencia de Industria y Servicios y el Instituto Geográfico Agustín Codazzi.

Con el apoyo de:
Global Partnership for Sustainable Development Data

• En el nivel nacional, los principales productores de datos e información estadística son: DANE, entidades de gobierno nacional, Banco de la República y otras entidades independientes, organismos de control, entidades de la rama judicial y organización electoral, sector privado y organizaciones de la sociedad civil.

• En el nivel territorial se encuentran las gobernaciones, regiones administrativas de planeación, órganos de control regional y departamental, academia, alcaldías, sector privado y organizaciones de la sociedad civil.

• Los usuarios del ecosistema de datos son principalmente: la ciudadanía, academia, medios de comunicación, sociedad civil, organismos internacionales, gobierno nacional y locales, sector privado y organismos de control.

Operación y dinámica

• A pesar de los esfuerzos aislados por generar y mejorar la producción de indicadores territoriales (MIPEG, Fortalecimiento territorial), aún se evidencia baja capacidad operativa del ecosistema de datos para atender la producción de indicadores en el nivel territorial.

• Aún se siguen presentando barreras en el acceso[7] a registros administrativos para aprovechamiento estadístico. A pesar de los marcos normativos existentes y de las salas y espacios para abordar los temas en el SEN, muchas entidades no entregan su información por desconfianza y desconocimiento del SEN.

• A nivel de ecosistema, la carga a proveedores de datos entre entidades es alta y muchas veces duplicada. Tal es el caso de las solicitudes a los territorios entre el Departamento Nacional de Planeación y el Departamento Administrativo de la Función Pública, entre otros.

• Con respecto a los estándares de calidad de datos que se manejan en el país, el DANE tiene metodologías de calidad estadística y el MinTIC guías de calidad y gobierno de datos, donde en algunas definiciones y alcances difieren. Por ejemplo, las definiciones y campos que deben contener los diccionarios de datos, metadatos y buenas prácticas de difusión de información.

7. Se identifican principalmente cuatro: 1) desconocimiento de la normativa, 2) desconocimiento de técnicas y procesos de anonimización, 3) poca claridad en la articulación normativa frente a temas de hábeas data, protección de datos personales y reserva estadística, y 4) mayor entendimiento sobre los beneficios y costos de acceder a esta información por parte de los actores involucrados.

Con el apoyo de:
Global Partnership for Sustainable Development Data

• El SEN 2.0 le brinda al ecosistema de datos la directriz que las estadísticas oficiales[8] son de uso obligatorio para fines de política pública y, además, son las que se deben reportar a organismos internacionales. Sin embargo, en la práctica, no todas las políticas públicas se apoyan exclusivamente de estadísticas oficiales.

• A nivel de ecosistema de datos, se viene avanzando en estándares de interoperabilidad entre las entidades del sector público. Principalmente enfocado a mejorar la prestación de los servicios públicos a los ciudadanos. El modelo que se sigue es el de Estonia, que utiliza la tecnología X-Road.

• A nivel del SEN 2.0, el DANE ha venido desarrollando distintos marcos de intercambio de datos e interoperabilidad con distintos actores del sistema. A manera de ejemplo se encuentran los servicios de acceso a geo-visores con estándares SDMX.

• Se identifican fallas de ciberseguridad y baja capacidad tecnológica en el DANE. Actualmente, el DANE se ha visto expuesto a ataques en su red que han comprometido gran parte del stock de datos del país.

• El DANE tiene el rol dual de producir información, crear marcos de calidad y evaluar la calidad a partir de terceras partes independientes.

• Rezagos sustantivos en producción y calidad de datos de servicios públicos, justicia y seguridad y defensa, y medio ambiente.

• Falta una mayor articulación de la agenda ODS entre los actores del ecosistema de datos de Colombia. En específico, se requiere una mayor difusión y coordinación de los avances, procedimientos y metodologías utilizadas para la medición y monitoreo y evaluación de estos indicadores en el país. Se evidencian diferencias significativas en los reportes de avance de cumplimiento de estos indicadores entre la OCDE y el DNP.

8. El DANE define estadística oficial como estadísticas que se encuentran identificadas y priorizadas en el plan estadístico nacional, y que han sido evaluadas y certificadas en procesos de calidad.

6. Conclusiones y recomendaciones para el desarrollo del ecosistema de datos

A continuación, se describen las principales recomendaciones y lineamientos para impulsar el desarrollo de un ecosistema de datos integral en Colombia. Las recomendaciones son abordadas desde los siguientes niveles de articulación de datos: **alineación estratégica, gobernanza de datos, gestión de datos, coordinación e integración de datos e inventario de fuentes de datos.**

Promover una alineación estratégica entre las políticas de información, datos y estadística con la Agenda 2030 y las nuevas iniciativas de desarrollo propuestas por el actual gobierno, en el CAD del SEN

- Se recomienda contar con un marco o documento regulatorio que establezca e impulse la toma de decisiones basadas en datos, articule distintos marcos nacionales en función de datos, estadísticas e información, y promueva la acción oportuna cuando la información o los datos oficiales así lo sugieran. Se sugiere que estos temas sean discutidos y analizados en el CAD del SEN.

- Se debe evitar traslapos en el desarrollo de la política de información del país. El SEN 2.0 y el Plan de Infraestructura han avanzado en esa articulación, sin embargo, se debe prestar mucha atención al desarrollo y puesta en operación de dichas articulaciones. Las salas del SEN 2.0 tienen la responsabilidad de brindar los espacios para que se sigan materializando políticas de información y datos de actores como el DNP y MinTIC.

- Se recomienda que el Plan Estadístico Nacional sea mandatorio como estrategia fundamental de datos. Es muy importante que este Plan Estadístico también articule los temas de datos en el país. Esto último se puede lograr con la aprobación de la Ley Estadística que fortalecerá al CASEN y permitirá impulsar el Plan Nacional de Infraestructura de Datos.

• Es recomendable que el ecosistema de datos se apoye en estructuras de gobernanza bien institucionalizadas que fomenten la cooperación entre actores del ecosistema y no duplicación de actividades y tareas.

• Las estrategias de gobierno y de datos deben estar alineadas. Es esencial impulsar, de manera ordenada, la formulación de políticas de desarrollo económico, social y ambiental, asociadas a una estrategia de datos, que blinde la toma de decisiones basada en evidencia. Además, fomentar el uso de los tableros de datos o indicadores estratégicos del país para responder de manera oportuna con políticas públicas. Es decir, generar una cultura para la medición, monitoreo y evaluación.

• Se recomienda que se evalúe la capacidad de operación de las mesas que comprenden el SEN 2.0 y se monitoreen compromisos y avances de manera permanente. Esto permitirá evaluar la gestión de los participantes de cada mesa y medir el progreso de las discusiones en materia de datos y estadísticas.

Impulsar una gobernanza de datos más efectiva, medible y con enfoque de resultados

• La gobernanza de datos integra personas, cultura, procesos y políticas. Es importante identificar los procesos y propósitos identificados en el Plan de Infraestructura de Datos para que no se queden las ideas en buenas intenciones.

• Los procesos permiten coordinar acciones y mostrar claramente cómo se pruduce, de manera controlada, una actividad asociada a datos. En la medida en que estos procesos perduran y se hacen más fuertes, pueden impactar y forjar la cultura de datos que se desea para el país. Sin embargo, una buena estrategia de comunicación de los principios sobre los que se construye el Plan de Infraestructura de Datos, daría un impulso fundamental para articular a más actores del ecosistema y motivar a la ciudadanía a propiciar rendición de cuentas bajo estos principios.

• El ecosistema de datos debe seguir liderando las políticas de datos abiertos y enfocar sus esfuerzos en los entes territoriales.

• Se debe seguir promoviendo las políticas y estrategias orientadas a impulsar los datos para el bien social en el país. No solo de cara al SEN 2.0, sino a todo el ecosistema de datos colombiano. Una forma es a partir de reconocimientos especiales a aquellas empresas privadas que hacen explícito el desarrollo de un programa o acceso a datos para el bien social.

• Se recomienda seguir la estructura de identificación de brechas de información planteada en el Plan Estadístico Nacional, complementada con estrategias de datos puntuales de acuerdo con los niveles de priorización del país. Muchos esfuerzos aislados no permiten ver avances rápidos en materia de datos. Por ello, la utilización de instrumentos existentes genera confianza en los actores del ecosistema respecto a su uso y aplicación.

• Es necesario que el DANE desarrolle programas de formación de evaluadores o auditores para que se alivie la presión de certificación de calidad de estadísticas en el país y se atienda la demanda existente en el nivel territorial.

• Se sugiere diseñar esquemas de responsabilidad de datos de principio a fin en las cadenas de valor del dato.

• Se recomienda implementar estrategias que impulsen el enfoque de demanda de datos sobre el de oferta de datos. Es decir, pasar de los datos existentes (oferta) a los problemas que se deben resolver o se pueden resolver con los datos (demanda).

Impulsar activamente la gestión, coordinación e integración de datos

• La gestión de datos aborda las actividades asociadas a datos maestros, bodegas de datos, inteligencia de negocio, big data, inteligencia artificial, calidad de datos y arquitectura de datos y modelamientos. Por su parte, la coordinación e integración de datos abarca temas como activos de datos y su integración y gestión de metadatos. Se recomienda que se apliquen actividades de gestión y coordinación de datos que permitan construir inventarios de datos, en especial bases maestras de distintas entidades con diccionarios y metadatos estandarizados, con el fin de construir modelos de entidad-relación entre entidades del sector público.

• Es recomendable que las actividades asociadas a calidad de datos sean formuladas por una sola entidad o comité del ecosistema de datos, con el objetivo de evitar descoordinación. En materia de información estadística y registros administrativos es el DANE el encargado de esta competencia.

• Se recomienda que se apliquen actividades de algoritmos abiertos, con el fin de fomentar la transparencia, evitar reprocesos e impulsar economías de escala dentro del ecosistema de datos. Para esto, se puede crear una wiki donde los actores del ecosistema accedan a los algoritmos ya desarrollados y verificados.

• A nivel de ecosistema de datos, es importante seguir avanzando en los procesos de interoperabilidad entre entidades del estado que viene impulsando el MinTIC. Con el propósito de garantizar que los fines no estadísticos y estadísticos gocen de igualdad y oportunidad para desarrollar sus objetivos.

• Es necesario articular las políticas o estrategias orientadas a los temas de big data e inteligencia artificial en el país con la gobernanza del ecosistema de datos. La conceptualización del ecosistema debe aproximar a los posibles habilitadores para que este tipo de políticas (big data, AI, etc.) generen valor agregado y no se queden en simples recomendaciones de políticas de datos. El SEN 2.0 es el espacio para su discusión y construcción colectiva entre las distintas entidades participantes.

• Se recomienda integrar prácticas de privacidad y responsabilidad por diseño a todo el ecosistema de datos. Es decir, que en el momento que se esté diseñando una operación estadística o proyecto de datos, se debe pensar en aspectos de privacidad, ética y equidad.

Cuidar y avanzar en el mapeo de distintas fuentes de datos del país

• Se sugiere crear mapas de datos en nivel territorial y sectorial que permitan identificar los distintos tipos de fuente tradicionales y no tradicionales de datos. Los mapas de datos deben contener modelos de entidad-relación. Realizar mapas de esta manera permite identificar visualmente dónde se encuentran las brechas de información. Estos mapas de datos deben complementarse con los inventarios de información estadística y registros administrativos ya adelantados por el DANE.

• Se recomienda realizar tableros de datos territoriales que muestren mapas de datos. Para el cumplimiento de los ODS, los instrumentos de contraste contribuyen a sumar experiencias e incentivan a los municipios a iniciar procesos de replicado de tableros.

• Es importante contar con un repositorio único sobre casos de uso de fuentes no tradicionales como datos de sensores remotos, datos de telefonía móvil, etc., para dimensionar los usos potenciales de esta información en los territorios.

Anexos

Preguntas formuladas en las entrevistas

- ¿Qué entiende usted o su organización por ecosistema de datos colombiano?
- ¿Cuáles cree que son las principales características de un ecosistema de datos?
- ¿Qué barreras o retos considera que se deben abordar con mayor detalle para avanzar en la consolidación de un ecosistema de datos más maduro?
- ¿Cuáles son esas entidades u organizaciones llamadas a coordinar un ecosistema de datos en Colombia? ¿Por qué?
- ¿Cuáles elementos habilitadores considera usted que son esenciales para un ecosistema exitoso?
- ¿Cuáles de esas podría aportar su entidad?
- ¿Qué iniciativas con datos ha desarrollado su entidad?
- ¿Cómo descubrieron esa necesidad?
- ¿Qué nivel de involucramiento ha tenido esta iniciativa con otros actores?

Listado de actores relevantes

Entidades

Banco de la República (BANREP)
Departamento Administrativo de la Función Pública (DAFP)
Superintendencia de Servicios Públicos Domiciliarios
Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC)
Instituto Geográfico Agustín Codazzi
Departamento Nacional de Planeación (DNP)

Con el apoyo de:
Global Partnership for Sustainable Development Data

Referencias

Aaser, M., Kanagasabai, K., Roth, M., & Tavakoli, A. (2020, November 23). Four ways to accelerate creating data ecosystems. McKinsey. Retrieved June 16, 2022, from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/four-ways-to-accelerate-the-creation-of-data-ecosystems

Abdulla, A., Janiszewska, E., & Podlesny, J. (2021, March 8). Data ecosystems made simple. McKinsey. Retrieved June 16, 2022, from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/tech-forward/data-ecosystems-made-simple

Azkan, C., Möller, F., Meisel, L., & Otto, B. (2020). Service Dominant Logic Perspective on Data Ecosystems - A Case Study based Morphology. In Conference: 28th European Conference on Information Systems At: Marrakesh, Morocco.

Cañón, V., Clavijo, A., Godoy, L., Letouzé, E., Pestre, G., & Ricard, J. (2017). Definición de la estrategia de Big Data para el Estado Colombiano y para el desarrollo de la industria de Big Data en Colombia. https://datapopalliance.org/wp-content/uploads/2020/11/Documento2_BuenasPracticasyDefiniciones_DPAMIT.pdf

Cappiello, C., Gal, A., Jarke, M., & Rehof, J. (2020). Data Ecosystems: Sovereign Data Exchange among Organizations. Report from Dagstuhl Seminar 19391

.

CLASIFICACIÓN DE LOS DATOS PERSONALES E IMPLICACIONES LEGALES 1 Clasificación de los datos personales e implicaciones legales. (n.d.). Universidad Pontificia Bolivariana. Retrieved May 31, 2022, from https://repository.upb.edu.co/bitstream/handle/20.500.11912/3595/Clasificaci%C3%B3n%20de%20los%20datos%20personales%20e%20implicaciones%20legales.pdf?sequence=1

Curry, E., & Sheth, A. (2018). Next-Generation Smart Environments: From System of Systems to Data Ecosystems. In Intelligent Systems, IEEE 33(3):69-76.

Curry, E., & Sheth, A. (2019). Real-time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems. SpringrOpen.

Departamento Administrativo de la Función Pública. (2016). Instuctivo de la Política para el Tratamiento de Datos Personales. Función Pública. https://www.funcionpublica.gov.co/documents/418537/1512450/InstructivoPolitica_Datos.pdf/f7d7cbe2-6739-46de-9f76-ee147cf1aa60

Departamento Administrativo Nacional de Estadística [DANE]. (2020). Plan Estadístico Nacional Actualización 2020-2022.

Departamento Nacional de Planeación. (2018). Documento CONPES 3920 de 2018 [Política Nacional de Explotación de Datos (Big Data)]. https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/3920.pdf

Departamento Nacional de Planeación. (2018, March 15). Documento CONPES 3918. Estrategia para la implementación de los Objetivos de Desarrollo Sostenible (ODS) en Colombia. Retrieved June 29, 2022, from https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3918.pdf

Departamento Nacional de Planeación. (2019, November 8). Documento CONPES 3975. Política Nacional para la Transformación Digital e Inteligencia Artificial. Retrieved May 31, 2022, from https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3975.pdf

Departamento Nacional de Planeación. (2021, junio). Acelerar la implementación para una recuperación sostenible: Reporte nacional voluntario 2021. Retrieved June 29, 2022, from https://downloads.ctfassets.net/27p7ivvbl4bs/5QHMJWk16oeBIbmAuhlkaO/346d94400d11453bc9523ee63be7cacd/VNR_2021_Colombia.pdf

Departamento Nacional de Planeación. (2022, May). https://media-exp1.licdn.com/dms/document/C4E1FAQEug1On4Qh4zg/feedshare-document-pdf-analyzed/0/1652140226779?e=1654732800&v=beta&t=xFjwTnRD9XoihbFBWaCHg21SuEqdbkCpZLDPAEUEi3k. Retrieved May 31, 2022, from https://media-exp1.licdn.com/dms/document/C4E1FAQEug1On4Qh4zg/feedshare-document-pdf-analyzed/0/1652140226779?e=1654732800&v=beta&t=xFjwTnRD9XoihbFBWaCHg21SuEqdbkCpZLDPAEUEi3k

Geisler, S., Vidal, M.-E., Cappiello, C., Lóscio, B. F., Gal, A., Jarke, M., Lenzerini, M., Missier, P., Otto, B., Paja, E., Pernici, B., & Rehof, J. (2021). Knowledge-driven Data Ecosystems Towards Data Transparency.

Gelhaar, J., Groß, T., & Otto, B. (2021). A Taxonomy for Data Ecosystems. In Conference: 54th Hawaii International Conference on System Sciences (HICSS)At: Hawaii, USA.

Gelhaar, J., & Otto, B. (2020). Challenges in the Emergence of Data Ecosystems. In Conference: Pacific Asia Conference on Information Systems (PACIS) 2020At: Dubai.

González García, H. G., & Villamizar Meneses, Ó. L. L. (2021). Informe de ponencia para segundo debate en la Plenaria de la Cámara de Representantes al Proyecto de Ley 222/21 Cámara "Por la cual se expiden disposiciones sobre las estadísticas oficiales en el país".

Grossman, R. L. (2018). Progress Toward Cancer Data Ecosystems. The Cancer Journal, 24(3). https://pubmed.ncbi.nlm.nih.gov/29794537/

Guío Español, A., Tamayo Uribe, E., Gómez Yarbe, P., & Mujica, M. P. (2021, Octubre). Marco Ético para la Inteligencia Artificial en Colombia. Departamento Administrativo de la Presidencia de la República. Retrieved May 31, 2022, from https://dapre.presidencia.gov.co/TD/MARCO-ETICO-PARA-LA-INTELIGENCIA-ARTIFICIAL-EN-COLOMBIA-2021.pdf

Heijlen, R., & Crompvoets, J. (2021). Open health data: Mapping the ecosystem. Digital Health, 7(1-16). https://journals.sagepub.com/doi/pdf/10.1177%2F20552076211050167

Ley 1266 de 2008 [Por la cual se dictan las disposiciones generales del hábeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan]. (2008). http://www.secretariasenado.gov.co/senado/basedoc/ley_1266_2008.html

Ley 1581 de 2012 [Por la cual se dictan disposiciones generales para la protección de datos personales]. (2012). http://www.secretariasenado.gov.co/senado/basedoc/ley_1581_2012.html

Ministerio de Ciencia Tecnología e Innovación. (n.d.). DOCUMENTOS CONPES. Minciencias. Retrieved May 31, 2022, from https://minciencias.gov.co/glosario/documentos-conpes

Ministerio de Tecnologías de la Información y Comunicaciones, Departamento Nacional de Planeación, & Departamento Administrativo de la Presidencia de la República. (2021). Plan Nacional de Infraestructura de Datos.

Ministerio de Tecnologías de la Información y las Comunicaciones. (2019). Guía para el uso y aprovechamiento de Datos Abiertos en Colombia. Datos Abiertos. Retrieved May 31, 2022, from https://herramientas.datos.gov.co/sites/default/files/Guia%20de%20Datos%20Abiertos%20de%20Colombia.pdf

Ministerio de Tecnologías de la Información y las Telecomunicaciones. (n.d.). Política de Gobierno Digital. Gobierno Digital. https://gobiernodigital.mintic.gov.co/portal/Politica-de-Gobierno-Digital/

Mujica Ramírez, M. P., & Guío Español, A. (n.d.). Misión de Expertos en IA de Colombia Misión de Expertos en IA de Colombia. Retrieved May 31, 2022, from https://inteligenciaartificial.gov.co/static/img/201021_Mision_de_Expertos_en_IA_de_Colombia.pdf

OCDE. (2022). Medición de la distancia para lograr las metas de los Objetivos de Desarrollo Sostenible (ODS) - Colombia. Retrieved June 29, 2022, from https://www.oecd.org/wise/medicion-de-la-distancia-para-lograr-las-metas-de-los-ODS-perfil-de-pais-Colombia.pdf

Oliveira, M. I., Barros Lima, G. d. F., & Lóscio, B. F. (2019). Investigations into Data Ecosystems: a systematic mapping study. Knowledge and Information Systems, 61(5).

Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem? In Proceedings of the 19th Annual International Conference on Digital Government Research, May 30-June 1, 2018, Delft, Netherlands.

ONU, UNFPA, FAO, & DANE. (2021). Guía para la medición y reporte de los indicadores globales en el Marco de Cooperación de las Naciones Unidas para el Desarrollo Sostenible. Retrieved June 29, 2022, from https://www.dane.gov.co/files/indicadores-ods/Guia-medicion-reporte-indicadores-ODS.pdf

Osorio-Sanabria, M. A., Amaya-Fernández, F., & González-Zabala, M. (2020). Exploring the components of open data ecosystems: a systematic mapping study. In " EATIS '20: Proceedings of the 10th Euro-American Conference on Telematics and Information Systems".

Osorio-Sanabria, M.-A., Amaya-Fernández, F. O., & González-Zabala, M. P. (2021). Políticas, normas y estrategias que fomentanlos datos abiertosen Colombia: un análisis de literatura. Revista Virtual Universidad Católica del Norte, (62), 155-188. https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/1250/1625

Otto, B., Lis, D., Jürjens, J., Cirullies, J., Howar, F., Meister, S., Spiekermann, M., Pettenpohl, H., Möller, F., Rehof, J., & Opriel, S. (2019). Data Ecosystems. Conceptual Foundations, Constituents and Recommendations for Action. In Affiliation: Fraunhofer Institute for Software and Systems Engineering ISST, Project: International Data Spaces.

Posada, M. (2020, June 23). ¿Qué ha hecho el Centro para la Cuarta Revolución Industrial? ELTIEMPO.COM. https://www.eltiempo.com/colombia/medellin/que-ha-hecho-el-centro-para-la-cuarta-revolucion-industrial-en-medellin-510004

Shah, S. I. H., Peristeras, V., & Magnisalis, I. (2020). Government (Big) Data Ecosystem: Definition, Classification of Actors, and Their Roles. Athens Greece, Apr 09-10(Part II).

Sistema Estadístico Nacional. (2022). Comité Estadístico Sectorial de Geografía, medio ambiente y ordenamiento territorial.

Stobierski, T. (2021, March 2). 5 Key Elements of a Data Ecosystem. HBS Online. Retrieved June 16, 2022, from https://online.hbs.edu/blog/post/data-ecosystem

van den Homberg, M. v. d. H. (2018). Characterizing Data Ecosystems to Support Official Statistics with Open Mapping Data for Reporting on Sustainable Development Goals [International Journal of Geo-Information 7(12):456].

Wiener, M., Sommer, F. T., Ives, Z. G., Poldrack, R. A., & Litt, B. (2016). Enabling an Open Data Ecosystem for the Neurosciences. Neuron, 92(3). https://reader.elsevier.com/reader/sd/pii/S0896627316307875?token=72634AA5250A32906A2F7B019C714539B23A377AAD1AB03A1FC95BA2B42C678C91C425EC9BF4636E69C40833E2945495&originRegion=us-east-1&originCreation=20220427182047

Wilson, L. (2019). Data-driven Marketing Content: A Practical Guide. Emerald Publishing Limited.

World Economic Forum. (2021, December). Data for Common Purpose: Enabling Colombia's Transition to a Data-Driven Economy. weforum.org. Retrieved May 31, 2022, from https://www3.weforum.org/docs/WEF_Enabling_Colombia%E2%80%99s_Transition_to_a_Data_Driven_Economy_2021.pdf

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. Information Polity, 19(1), 17-33.

Con el apoyo de:

Global Partnership for Sustainable Development Data