

Generalidades del big data para el desarrollo sostenible en Colombia

Autores:
Margarita Vaca
Juan Camilo Vega

Junio, 2019



BIGDATA
FORDEVELOPMENT
Network



Cepei

Knowledge factory for the world we deserve



BIGDATA
FORDEVELOPMENT
Network

Generalidades del big data para el desarrollo sostenible en Colombia

Dirección general: *Philipp Schönrock*

Autores: *Margarita Vaca y Juan Camilo Vega*

Edición y revisión técnica: *Fredy Rodríguez*

Edición de apoyo: *Alexandra Roldán y Carolina Vegas*

Diseño y diagramación: *Alba Pabón*

Bogotá D.C., Colombia

Junio de 2019

Contenido

Introducción.....	4
Contexto del desarrollo sostenible	5
¿Qué es el BIG DATA?	8
Contexto del big data en Colombia	9
Fuentes de datos.....	13
La importancia de las fuentes de datos	16
Conclusiones	21
Bibliografía	22
Acerca de Cepei	25

Introducción

La necesidad de aprovechar los datos que son producidos día a día es cada vez más evidente. En especial, cuando son un insumo principal para la toma de decisiones a nivel regional, nacional y subnacional. Sin embargo, uno de los principales problemas que enfrentan los países es la alta inversión de recursos que representa la recolección, el procesamiento, el análisis y la difusión de esta información.

Esporeo, que el uso de nuevas herramientas de tecnología y la implementación de fuentes no oficiales de información, como el big data, ofrecen alternativas económicas, eficientes e innovadoras. Estas permiten cerrar brechas de información, así como generar evidencia para la toma de decisiones y la construcción de nuevas estrategias para el desarrollo sostenible.

En el caso colombiano, por ejemplo, la deforestación es un problema grave. Solo para el 2015, habían sido deforestadas 56.962 hectáreas (IDEAM, 2018). Esta área equivale en tamaño a ciudades capitales como Popayán, Cali, o, aproximadamente, 5 veces Mocoa. Con el fin de encontrar soluciones a estas y otras problemáticas, surge la necesidad de obtener datos que permitan identificar las zonas críticas, y otra

información, que ayude a determinar las causas de este problema. Algunos métodos tradicionales para la recolección y análisis de datos han sido, por ejemplo, sobrevolar diferentes zonas del país e identificar las áreas donde se presenta el fenómeno para calcular la superficie afectada. En el caso de la deforestación, el big data provee el uso de imágenes satelitales para identificar patrones de cambio en los bosques del territorio colombiano. Lo cual permite la implementación de políticas orientadas a controlar esta situación.

El presente documento expone el contexto del desarrollo sostenible y habla sobre los Objetivos de Desarrollo del Milenio (ODM) y los Objetivos de Desarrollo Sostenible (ODS). Luego aborda el tema de la presencia de los tipos de fuentes de información dentro del contexto colombiano, y cómo el big data puede contribuir a la implementación de los ODS, así como la importancia de usar todas las fuentes de información disponibles en la consecución de la Agenda 2030. También se presentan iniciativas en el uso de fuentes no oficiales de información, las cuales pueden guiar futuros esfuerzos para el uso de este tipo de datos, con el fin de motivar a distintos actores a tenerlas en cuenta como nuevas herramientas de medición.

Contexto del desarrollo sostenible

Los Objetivos de Desarrollo del Milenio (ODM) se consolidaron como un concepto revolucionario que logró establecer sinergias entre distintos países y actores, al definir un discurso común enfocado a mejorar el bienestar general de la población. Para este fin, se fijaron metas para la erradicación de la pobreza, la promoción de la igualdad de género, el mejoramiento de la salud materna y la sostenibilidad del medio ambiente.

Aunque la implementación de esta hoja de ruta presentó avances en cada uno de sus objetivos y alcanzó una reducción significativa de la pobreza extrema¹ al 14% en 2015; un aumento del número de estudiantes matriculados en instituciones de educación primaria en 8 puntos porcentuales; una reducción en la mortalidad infantil² que pasó de 90 a 43 niños muertos por cada 1000; el alcance de los objetivos fue desigual y no permitió la inclusión de todos los países.

Luego, en 2015, se identificó una nueva hoja de ruta con un enfoque holístico

del desarrollo sostenible, que plantea la importancia de una articulación sólida de las diferentes esferas de la sociedad: medio ambiente, inclusión social y desarrollo económico. Su objetivo es eliminar la desigualdad y proteger el medio ambiente a través de la construcción de alianzas entre diferentes sectores y actores, resaltando la responsabilidad compartida que deben asumir todos los países y personas. Esta hoja de ruta fue denominada como la Agenda 2030 y está compuesta por los Objetivos de Desarrollo Sostenible (ODS), los cuales incorporan las lecciones aprendidas de los ODM y portan el lema de "No dejar a nadie atrás". Para lograr el propósito de la Agenda 2030, 189 países acordaron 17 ODS, compuestos por 169 metas y 241 indicadores, que buscan fortalecer los procesos de medición y monitoreo de los avances de la misma.

Una de las principales características de los ODS es que están interconectados, es decir que los logros en cualquiera de

1 El índice de pobreza extrema es calculado como el porcentaje de la población que vive con menos de 1,90 dólares al día, corregido por Paridad de Poder adquisitivo (PPP en inglés).

2 La mortalidad infantil es presentada como el número de menores de un año muertos por cada 1.000.

ellos impacta en menor o mayor medida el cumplimiento de los demás. Asimismo, conciben un enfoque articulado que tiene en cuenta componentes innovadores como: Ciudades y Comunidades Sostenibles, Producción y Consumo Responsables, y Paz e Instituciones Sólidas, presentes en los ODS 11, 12 y 16, respectivamente. Además brinda un mayor detalle de atención al medio ambiente a través de la Acción por el Clima, Vida Submarina, y Vida y Ecosistemas Terrestres, considerados en los ODS 13, 14 y 15, respectivamente.

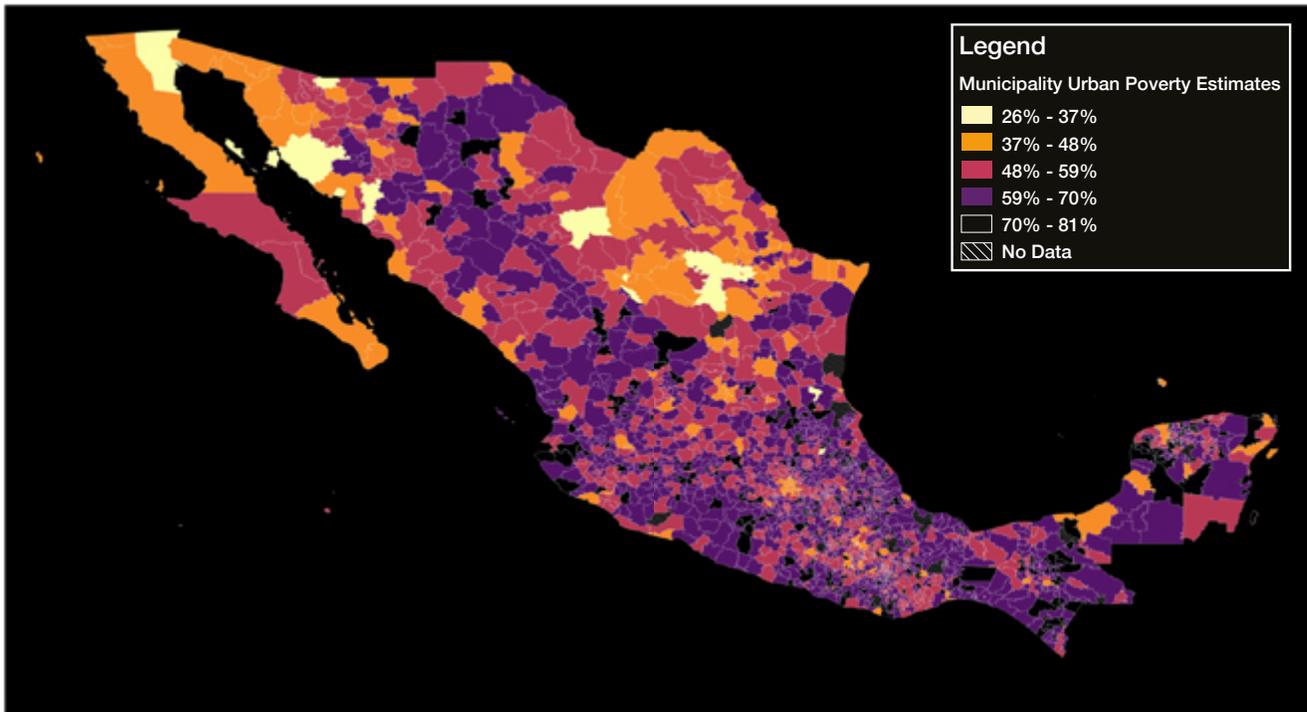
Los ODS han representado un desafío para todos los países. No sólo por la escasez de recursos en términos de dinero, capital físico y humano, sino también por la necesidad de información que permita enfocar estos recursos de forma eficiente y de medir la efectividad de los esfuerzos hechos.

A partir de esto, los países han trabajado a través de sus Oficinas de Estadística, Ministerios, Comisiones ODS, entre otros actores en fortalecer las estadísticas nacionales y subnacionales. Empero, esta nueva Agenda ha evidenciado la necesidad de incorporar nuevas fuentes de información que faciliten la identificación de datos a mayores niveles de desagregación.

Es así, que la información que se produce en el día a día, y que usualmente no es considerada por las oficinas de estadística u oficinas gubernamentales, ha tomado un rol significativo para determinar nuevas soluciones a problemáticas sociales y estadísticas. Por ejemplo, la pobreza generalmente se mide a través de encuestas presenciales en un número significativo de hogares del territorio a través de preguntas que buscan reflejar el estado socioeconómico de las personas que habitan en la vivienda. Hacer este tipo de mediciones requiere de una alta inversión, lo cual limita la continuidad y cobertura de la misma.

A causa de esta necesidad, el big data, como fuente de datos no tradicional, ofrece medidas con menores costos gracias a las que se pueden llevar a cabo este tipo de ejercicios, sin tener que recurrir a los medios tradicionales. En un estudio realizado por David Newhouse et al. (2017) se busca medir el porcentaje de personas que están bajo la línea de pobreza en México, usando imágenes satelitales de alta resolución entre 2014 y 2015. Estas imágenes identifican los materiales con los que están hechas las casas, los techos, las paredes, el número de pisos de las casas, si tienen un auto, y a partir de eso se estima el porcentaje de población que se encuentra en estado de pobreza, por municipio.

Figura 1. Estimaciones de pobreza: municipalidades urbanas México



Fuente: David Newhouse et al. (2017)

Un ejemplo adicional del uso de big data es la alianza entre Cepei y Telefónica, para medir el impacto migratorio de la avenida torrencial ocurrida en Mocoa en el 2017. A través de la posición georeferenciada de los teléfonos móviles, es posible explicar la

dinámica migratoria de la población hacia y fuera de Mocoa antes, durante y después de la tragedia. Este es uno de los varios casos de acción que ofrecen las fuentes no oficiales de información, que ayudan a cerrar las brechas de información.

¿Qué es el BIG DATA?

Especialistas en analítica de datos argumentan que:
“El término Big Data se aplica a la información que no se puede procesar o analizar mediante los procesos o herramientas tradicionales”

(Zikopoulos et al, 2011).

Partiendo de esta noción, big data puede definirse como un concepto evolutivo que describe la transformación de grandes conjuntos de datos (estructurados, semiestructurados y no estructurados) o la combinación de estos, que se generan de acuerdo con la dinámica de las sociedades y se traducen en diagnósticos, directrices o parámetros, para la correcta toma de decisiones en un tiempo y espacio determinado. Del mismo modo, big data es un insumo potencial para la extracción de correlaciones entre variables clave que permiten explicar las causas de un evento, anticipar tendencias o resultados, e inferir posibles soluciones a nivel global, nacional y subnacional.

Este concepto engloba las cinco características que comprenden al big data: variedad, volumen, velocidad, valor, y veracidad. Para una mayor comprensión, en el siguiente caso se ejemplifican las características del big data:

El mundial de fútbol de Rusia 2018 atrajo a miles de personas a distintas ciudades del país organizador. Esto se convirtió en una gran oportunidad para las cadenas hoteleras, al apoderarse de la mayor porción de demanda por habitaciones en las ciudades donde se celebraron los partidos. Para cumplir con su objetivo, era necesario analizar la información recolectada de sus clientes durante los

días del evento, para así poder ofrecer los paquetes de estadía adecuados a los clientes potenciales.

La información recolectada puede comprender la edad, nacionalidad, sexo, preferencias por servicios (paquetes de transporte o de sitios turísticos), entre otros. Con esta *variedad* y gran *volumen* de información, los hoteles crearon perfiles de clientes potenciales en los cuales podrían enfocar sus esfuerzos. Asimismo, cientos de clientes llegaban a los hoteles en esta época, por esta razón, la información que se generaba diariamente aumentaba a una *velocidad* impresionante. Debido a que la duración del evento era limitado

a fechas determinadas, el análisis de la información debía hacerse lo más rápido posible para tomar ventaja de esto y prever eventualidades desfavorables.

Por último, analizar esta información representó una oportunidad para la toma de decisiones basada en evidencia. El big data representó un gran *valor* para las empresas a la hora de unificar recursos a través de la identificación de oportunidades en el uso de los insumos. Por otra parte, la *veracidad* de los datos es un factor crítico, dado que no se tiene un control de la calidad de los mismos, ni un proceso estandarizado en la recolección de la información.

Contexto del big data en Colombia

Colombia ha sido uno de los primeros países en promover políticas públicas para el desarrollo sostenible. Esto se ve reflejado en la creación de la Comisión ODS, estructurada por las entidades responsables de impulsar acciones que vayan en línea con el cumplimiento de las metas de la Agenda 2030. Adicionalmente, existe una ley y una política de datos abiertos que apoya la publicación de datos de interés general,

con el fin de investigar o visualizar esta información y promover la publicación de nuevos conjuntos de datos. En esta política se comparte información bajo la protección del Habeas Data, con la finalidad de no vulnerar los derechos de los ciudadanos durante el proceso.

Para ilustrar al lector acerca de la realidad colombiana en materia de planeación, el

Consejo Nacional de Política Económica y Social (CONPES) -que es el principal organismo en materia de desarrollo económico y social- se encarga de brindar asesoría al gobierno nacional en estos temas y visibiliza su labor a través de documentos denominados Conpes.

En el contexto de América Latina, Colombia ha sido uno de los países más comprometidos en promover espacios donde se generan y comparten datos para el desarrollo sostenible. Esto se evidencia en la inclusión

de esfuerzos dentro de sus políticas públicas. Por ejemplo, el gobierno nacional junto con el Departamento Nacional de Planeación (DNP) presentó una de las herramientas para la puesta en marcha de la Agenda 2030 en Colombia: el Documento Conpes 3918, el cual comprende a las entidades encargadas de cada una de las 147 metas³ de los ODS, la definición de indicadores nacionales y las metas del país para el año 2030, incluyendo la regionalización de las rutas de acción y la definición del esquema de reporte de información para los indicadores.

Organizaciones a cargo de los ODS en Colombia

En Colombia existen 30 entidades encargadas de liderar el cumplimiento de la Agenda 2030. La principal entidad es el DNP, que además de encargarse de la medición de los indicadores ODS, formula las políticas públicas para cada período presidencial. Estas luego constituyen el Plan Nacional de Desarrollo.

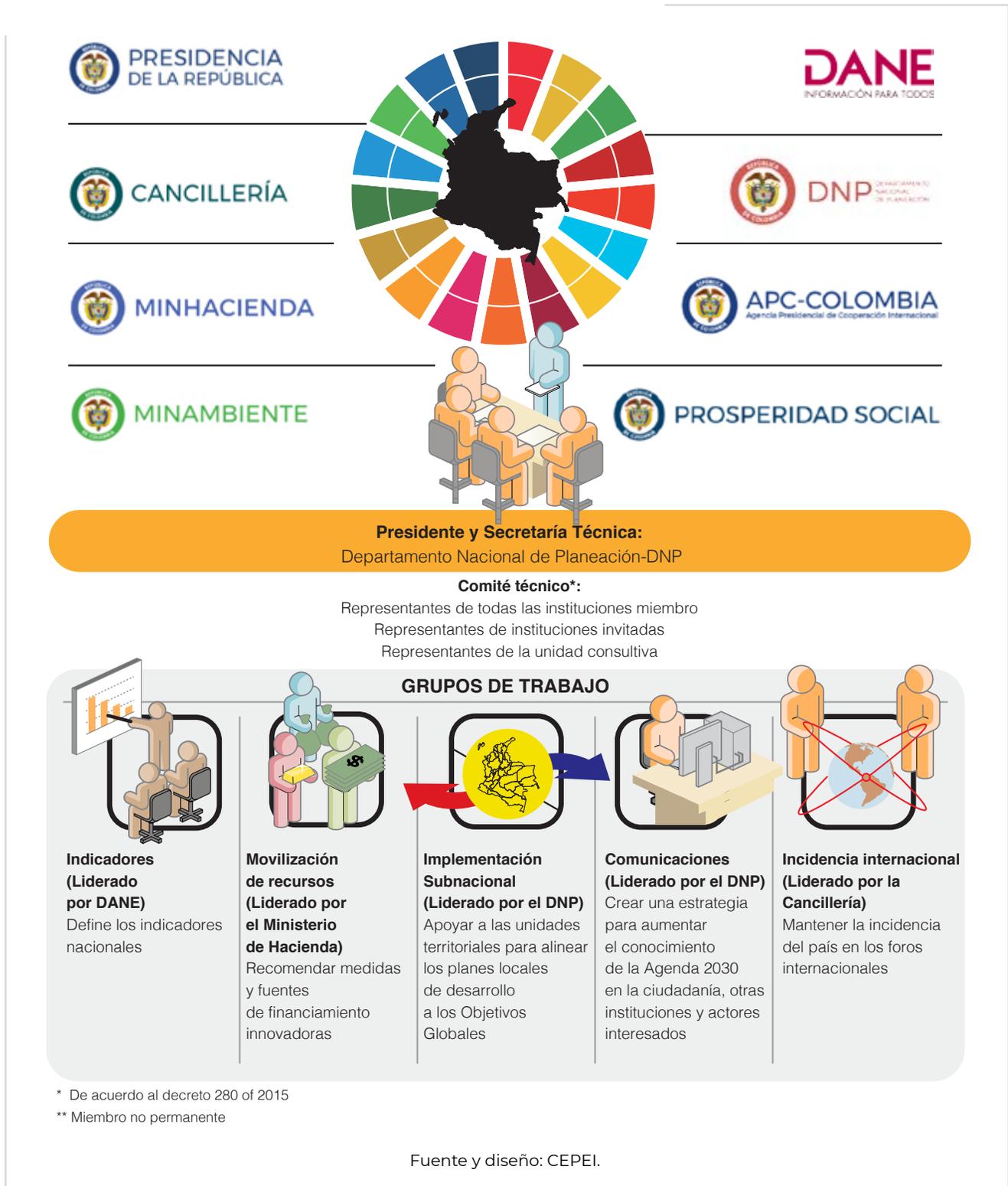
En el documento Conpes 3918, están definidas⁴ las instituciones a cargo de cada meta. Se definen como instituciones líderes

aquellas que tienen una responsabilidad directa, mientras las instituciones acompañantes son aquellas que tienen la misión de brindar apoyo a las instituciones líderes, pues pueden incidir de manera directa sobre esa meta. Por ejemplo, para la meta 1.1 (erradicar la pobreza extrema), la institución líder es el Departamento de Prosperidad Social y como entidades acompañantes están los ministerios de Agricultura, Minas y Energía, y Educación.

3 Dado que las 22 metas restantes requieren de esfuerzos conjuntos entre actores gubernamentales e internacionales.

4 Consultar Anexo E del documento Conpes 3918.

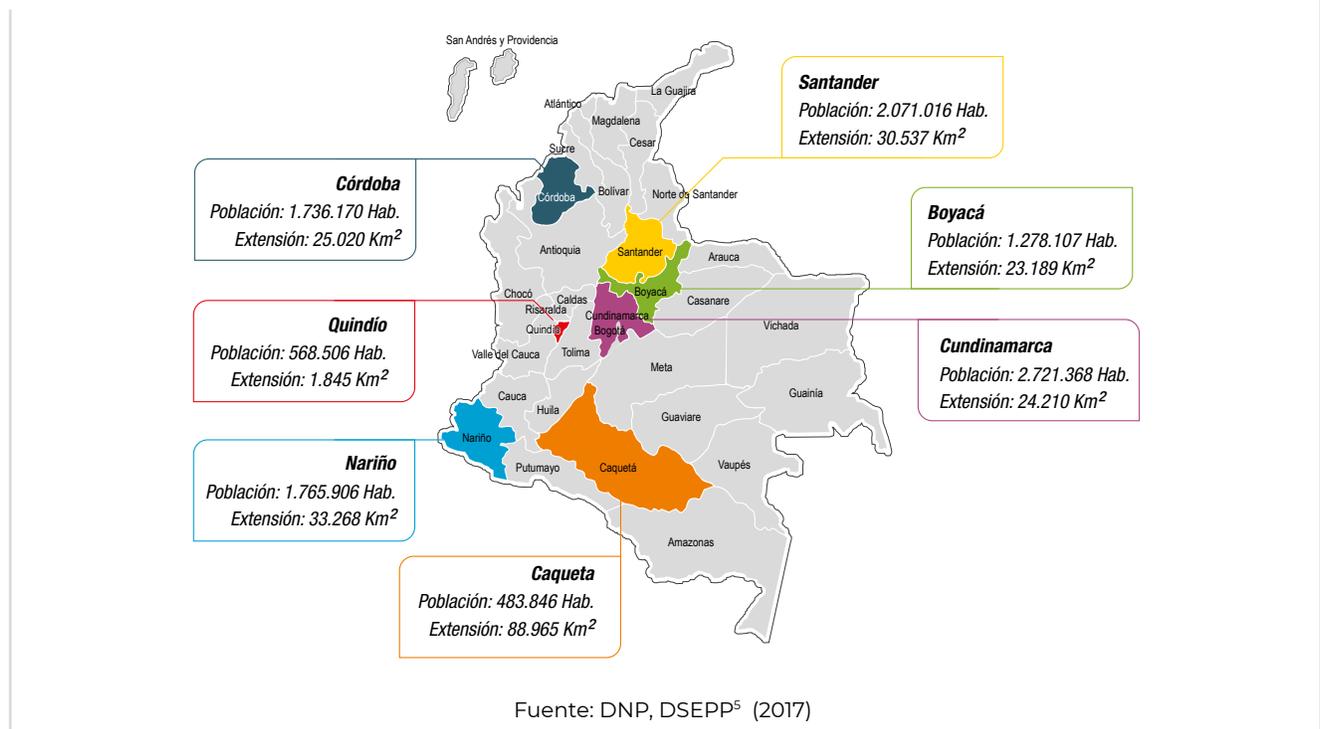
Figura 2. Comisión ODS en Colombia - Miembros y estructura



Con el fin de identificar fortalezas y debilidades del país en la implementación de la Agenda 2030, el Departamento Administrativo Nacional de Estadística (DANE)- oficina nacional de estadística-, tiene la misión de medir los indicadores ODS, así como generar y compartir información para realizar el adecuado

seguimiento al cumplimiento de la Agenda 2030. Para garantizar que las rutas de acción sean implementadas a nivel regional, la información de seguimiento debe tener un nivel de desagregación que permita identificar brechas sociales que orienten la toma de decisiones de los actores departamentales y municipales.

Figura 3. Buenas prácticas en la incorporación de los ODS en los PDD



En la figura 3 son presentados los departamentos de Santander, Boyacá, Cundinamarca, Córdoba, Quindío, Nariño y Caquetá, que son reconocidos por promover buenas prácticas en incluir los ODS dentro de sus Planes de Desarrollo

Departamental (PDD). Tanto en la alineación con componentes estratégicos⁶, como con componentes de medición⁷ (inclusión de los Objetivos de Desarrollo Sostenible en los planes de desarrollo territoriales, 2016-2019).

5 Dirección de Seguimiento y Evaluación de Políticas Públicas

6 Los componentes estratégicos son los que tienen asociación con dimensiones, ejes, líneas, objetivos, programas o proyectos como lo evidencian los PDD de Boyacá, Córdoba, Cundinamarca y Santander.

7 Los componentes de medición están enfocados a indicadores como en los casos de Caquetá, Quindío y Nariño.

Fuentes de datos

Las fuentes de datos pueden ser clasificadas en dos grandes categorías. De acuerdo a su origen y articulación con las estadísticas oficiales, y según la estructura de los datos. En el primer caso, las fuentes de datos estarán definidas por el nivel de rigurosidad metodológica de medición dado por la Oficina Nacional de Estadística, y se dividen

entre oficiales o tradicionales, y no oficiales o no tradicionales.

En la segunda categoría, las fuentes de datos se caracterizan como estructuradas, semiestructuradas y no estructuradas, lo cual indicará el nivel, el tipo de formato de la información, y su nivel procesamiento para obtener resultados.

Fuentes de datos según su origen

Las fuentes oficiales de información son aquellas producidas o supervisadas por las oficinas de estadística nacional de cada país. En el contexto colombiano, el Sistema Estadístico Nacional (SEN) es el conjunto articulado de componentes encargados de producir y difundir las estadísticas oficiales nacionales y territoriales. El DANE es ente rector del sistema (DANE, s.f.).

En la figura 4 se presentan el número de entidades⁸ que constituyen el SEN, las cuales incluyen tanto entidades estatales como personas de los sectores público, privado y mixto. Asimismo, el SEN garantiza que cada proceso estadístico tenga un código de ética, que asegura la rigurosidad con la que se recopila y difunde la información, y protege la confidencialidad que proveen los dueños de los datos (Naciones Unidas, 2013).

8 Para mayor información, el lector puede consultar el siguiente [link](#), donde puede encontrar el listado de las entidades que conforman el SEN.

Figura 4. Integrantes del SEN



Las fuentes de datos no oficiales se entienden como aquellas que no cuentan con el aval de la oficina nacional de estadística y, por lo tanto, no son consideradas o incluidas en las estadísticas oficiales. Un ejemplo son las investigaciones, reportes o estudios que realizan algunas instituciones del sector privado o la sociedad civil.

Si se considera que las fuentes no oficiales de información tienen un alto potencial de uso, por el nivel de detalle y temáticas específicas de trabajo, esto refleja una oportunidad de trabajar nuevos campos de medición o complementar las estadísticas oficiales.

Fuentes de datos según su estructura

El gran conjunto de fuentes de información que está disponible para el análisis de datos se clasifica en tres tipos: estructurados, semi estructurados y no estructurados.

Las fuentes de datos estructuradas son aquellas que se presentan en un formato predefinido (por ejemplo las hojas de cálculo) y se caracterizan por tener un orden claro que

hace posible un análisis, sin la necesidad de procesamiento previo. Estas son las fuentes de datos que usualmente genera el DANE para calcular los indicadores ODS.

También existen las fuentes de datos semi estructurados. Estas generalmente tienen un formato definido, pero son de difícil comprensión para el usuario y requieren

un procesamiento para facilitar su interpretación. Un ejemplo de estas fuentes de datos sería el texto de etiquetas en lenguajes XML y XHTML.

Finalmente, existen las fuentes de datos no estructuradas. Estas carecen de un formato de presentación claro, pues consisten de información proveniente de correos electrónicos, fotografías, mensajes de texto o archivos de audio y vídeo.

Es pertinente resaltar que las fuentes de datos semiestructurados y no estructurados, que actualmente no son consideradas como parte de las estadísticas oficiales, pueden llegar a contribuir de forma

significativa al cálculo de indicadores ODS. En especial si se consideran las falencias de información que se presentan actualmente en el país, pues según la información del portal ODS de Colombia⁹, 61 de las 169 metas carecen de información que permita realizar un adecuado seguimiento. El DNP en su informe *Avances y Desafíos para el Seguimiento de los ODS en Colombia y el documento Conpes 3918*, elaborado por el DNP, para el 2017, sólo el 54% de la información para los indicadores ODS estaba disponible. De esta manera, 30% era información parcial o requería mejoras, y no se contaban con datos o metodologías para el 16% restante.

9 Consultar <https://www.ods.gov.co/tools/data-availability>

La importancia de las fuentes de datos

Colombia ha mostrado gran interés en los datos no estructurados. Por esta razón, se desarrolló el Documento Conpes 3920, en el cual se propone: "la definición de un marco jurídico e institucional que maximice la obtención de beneficios del aprovechamiento de datos, permita la cooperación entre los sectores público y privado y, al mismo tiempo, refuerce

la protección de los derechos de los ciudadanos en el contexto de la creciente transformación de la vida diaria en datos digitales cuantificables y procesables" (Conpes 3920). De esta forma, se promueven esfuerzos en la generación y divulgación de fuentes de datos no oficiales para el desarrollo sostenible.

Datos abiertos en Colombia

Otro de los grandes avances en el país es su política de datos abiertos. Entre estos adelantos se puede citar el portal de datos abiertos datos.gov.co, creado por Ministerio de las Tecnologías de la Información y las Comunicaciones (MINTIC), en el cual se publican bases de datos que incluyen big data e investigaciones de diferentes organizaciones gubernamentales para libre consulta. Esta información es revisada previamente por el MINTIC. El portal está compuesto por tres módulos: *Descubre*, *Publica* y *Conoce*. *Descubre* sirve para consultar bases de datos publicadas para investigación, crear visualizaciones o desarrollar historias. *Publica* funciona desde la perspectiva de la oferta, ya que

en este módulo los usuarios pueden compartir información con el portal. *Conoce* está a disposición de los usuarios que quieran enterarse sobre investigaciones y visualizaciones hechas con datos abiertos.

Otro ejemplo fue la aprobación de la ley 1712, "por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones." (Ley 1712, 2014). Esta ley obliga a todas las entidades a publicar sus datos, siempre y cuando no se encuentren dentro de dos categorías: información pública clasificada e información pública reservada. (Ver figura 5)

Con el fin de entender la política de datos abiertos, el MINTIC publicó una guía de datos abiertos. Allí se explica, que la publicación de datos abiertos tiene un proceso en el cual se acompaña a la empresa en la divulgación de la información, se hace una revisión de los

datos para garantizar que no pertenezcan a algunas de las dos clasificaciones exentas, y, se hacen recomendaciones para publicar la información en el portal de la empresa, con el link de sus datos publicados en la página de datos abiertos.

Figura 5. Información exenta de publicación masiva



Iniciativas de big data en Colombia

Colombia ha trabajado en la puesta en marcha de diferentes iniciativas enfocadas en el big data, que parten de la identificación de problemáticas o necesidades de la sociedad, o el sector empresarial. A continuación se presentan algunas iniciativas enfocadas en salud, agricultura o seguridad en el territorio, que han utilizado el big data como herramienta para tomar decisiones que mitiguen o erradiquen problemas relacionados.

- **El grupo ODS¹⁰** del DANE junto con la **NASA** articularon esfuerzos para usar imágenes satelitales de alta definición, con el fin de calcular el indicador 11.3.1 que mide el cociente entre la tasa de consumo de tierra y la tasa de crecimiento de población. El estudio realizado en el año 2014 resultó en el cálculo de este indicador para 138 ciudades del país, en las cuales no se tenía medición previa.
- **Alianza Caoba:** Es un ecosistema de capacidades analíticas creado en el 2016, que busca contribuir a la competitividad, la investigación aplicada y el posicionamiento de la cultura de la innovación. Está conformada por el Ministerio de Tecnologías de la Información y las Telecomunicaciones,

el DNP, las universidades Javeriana, Eafit y Andes; IBM, EMC, SAS, Clúster Creativ; y los grupos Nutresa y Bancolombia. Esta alianza se ha enfocado en ofrecer soluciones innovadoras a las problemáticas que se presentan en diferentes contextos del país.

El primer proyecto que realizó esta iniciativa fue una investigación de mercadeo para el grupo Nutresa, en la cual analizaron las tendencias alimenticias en la red social Twitter. Identificaron qué tipo de alimentos buscaban los usuarios y cómo se sentían. Con esta información, crearon perfiles específicos de consumidores para poder enviar campañas de marketing y así aumentar su impacto.

- **Open Algorithm (OPAL):** DataPop Alliance, Imperial College London, el MIT Media Lab, Orange y el World Economic Forum en conjunto con Telefónica Colombia y el gobierno de Colombia crearon esta iniciativa en el año 2016, con el fin de usar la información que recopilan empresas del sector privado y analizarla para realizar políticas públicas. Las organizaciones interesadas se comunican con OPAL, exponiendo los

¹⁰ Consultar en la bibliografía: Experiencias del DANE en medición de indicadores para ODS fueron presentadas en el marco de la Asamblea General de la ONU

datos que tienen y el análisis que quieren lograr con esta información para mejorar el bienestar de los habitantes. Por su parte, OPAL desarrolla algoritmos que procesan esta información para ofrecer las soluciones requeridas.

- **Cultivando Futuro:** Esta iniciativa, impulsada por los emprendedores Darío González, Carlos Castellanos y Edinson Ballesteros en 2014, busca apuntarle a la meta 2.3, *duplicar la productividad agrícola y los ingresos de los productores de alimentos a pequeña escala*. Lo hace a través de la identificación del nivel de asistencia técnica que tienen los agricultores en el país, al considerar variables como el acceso a servicios bancarios, la conectividad a internet o la capacitación recibida para la mejora de su actividad. Con esta información, el Ministerio de Agricultura puede enfocar sus recursos en brindar asistencia a los agricultores, e identificar las zonas donde se concentra la mayor necesidad de capacitación.
- **Can Mobile Phone Traces Help Shed Light on the Spread of Zika in Colombia?:** Daniela Perrotta, investigadora de ISI¹¹ Foundation describe el proceso para identificar las personas contagiadas de zika entre 2015 y 2016 a través del rastreo

de celulares. En este proceso participan GlobalPulse, una iniciativa de la Secretaría General de Naciones Unidas sobre Big Data, en conjunto con Telefónica España, una empresa de telecomunicaciones. Esta iniciativa apunta a la meta 3.3. que es *poner fin a las enfermedades tropicales desatendidas*.

En su investigación explica que puede asociarse el contagio del zika, rastreando la posición geográfica de las personas a partir de las actividades realizadas con el celular. Por ejemplo, una persona que es rastreada por la compañía telefónica después de hacer una llamada telefónica en Bogotá, tiene una probabilidad muy baja de estar contagiada, porque la altitud de la ciudad no permite que los mosquitos portadores de la enfermedad se reproduzcan. Pero, una persona en la ciudad de Cúcuta, que cuenta con un clima tropical, tiene una mayor probabilidad de estar contagiada.

- **Terra-i, una colaboración entre CIAT-DAPA, FTA, TNC, HEIG-VD, y KCL,** actores¹² interesados en la vegetación, cuentan con un proyecto desde 2004 que tiene la finalidad de identificar zonas de deforestación, a partir del uso de imágenes satelitales tomadas cada 16 días en las cuales se monitorean los

11 Institute for Scientific Interchange

12 CIAT-DAPA: the International Center for Tropical Agriculture - FTA: the program of Forestry, Trees and Agroforestry - TNC: the Nature Conservancy - HEIG-VD the School of Business and Engineering. KCL: King's College London.

cambios repentinos en la vegetación. Estos cambios repentinos son vistos como una ganancia o una pérdida en la vegetación, tal como se daría en una jornada de plantación de árboles o un ejercicio de deforestación. Su trabajo se enfoca en varios países de América Latina como Bolivia, Brasil, Colombia, Perú y Paraguay.

En el caso colombiano, explica la dinámica del cambio en la vegetación por la construcción de la vía Pasto-Mocoa. Los resultados muestran que la pérdida de vegetación no fue tan alta como se estimaba. Por otra parte, la carretera facilitó la migración de la población rural e indígena desde zonas montañosas hacia zonas más planas.

Conclusiones

Colombia ha tenido un fuerte impacto en la generación de espacios que promueven el uso de big data para el desarrollo sostenible. También se ha evidenciado el compromiso con la generación de un ambiente propicio para el uso de los datos abiertos y el big data, así como su alineación con los ODS. Sin embargo, aún existe un amplio campo de acción para aprovechar el potencial de las fuentes de datos no oficiales y contribuir a la reducción de brechas de información.

Si bien es cierto que las fuentes de datos no oficiales, como el big data, han sido objeto de crítica por la veracidad de los datos (pues usualmente no cuentan con metodologías como las implementadas por las oficinas de estadística), no se han explorado posibilidades de analizar las metodologías realizadas por el sector privado, la academia y la sociedad civil, ni de ofrecer recomendaciones para su mejora

y articulación futura con las estadísticas oficiales.

Por otra parte, además de señalar la importancia de crear los espacios de convergencia y uso del big data, es pertinente incorporar mecanismos rigurosos que permitan garantizar la eficiencia y sostenibilidad de las iniciativas que se emprendan en este marco, para asegurar resultados en el corto y mediano plazo. Así mismo, facilitar la construcción de alianzas entre los diferentes actores que conlleven a la creación de una red de implementación del big data para el desarrollo sostenible.

El big data debe ser considerado, tanto por el gobierno como por los demás actores, como una herramienta para fortalecer los sistemas de medición de datos y monitoreo de ODS, pues amplía los niveles de detalle, obtiene resultados en tiempo real y complementa las fuentes de datos oficiales.

Bibliografía

[About Terra-i]. sin fecha. Recuperado de: <http://www.terra-i.org/terra-i/about.html>

Babenko, B. Hersh, J. Newhouse, D. Ramakrishnan, A. Swartz, T. (2017). Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico. Cornell University. Recuperado de: <https://arxiv.org/abs/1711.06323>

Castro, F. (2017). Avances y Desafíos para el Seguimiento de los ODS en Colombia. [ebook] Bogotá. Recuperado de: <https://www.dane.gov.co/files/images/eventos/ods/presentaciones/Alianzas/avances-desafios-seguimiento-ODS-colombia.pdf> [Consultado 16 Apr. 2019].

CONPES. (2018). Anexos Documento Conpes 3918. Recuperado de: https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3918_anexos.pdf

CONPES. (2018). Documento Conpes 3918: Estrategia para la Implementación de los Objetivos de Desarrollo Sostenible en Colombia (ODS) . Recuperado de: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3918.pdf>

CONPES. (2018). Documento Conpes 3920: Política Nacional de Explotación de Datos (Big Data). Recuperado de: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>

[Cultivando Futuro]. sin fecha. Disponible en: <https://cultivandofuturo.com/>

Datos Abiertos. Disponible en: <https://datos.gov.co/>

D. Deutsch, P. Eaton, Lapis, deRoos, C. T.G Zikopoulos. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. New York. United States. The McGraw-Hill Companies, Inc.

El Papel que Desempeña el CONPES. (2004). Diario El Tiempo. Recuperado de: <https://www.eltiempo.com/archivo/documento/MAM-1594532>

[Ésta herramienta muestra la disponibilidad de información para cada una de las 169 metas]. sin fecha. Recuperado de: <https://www.ods.gov.co/tools/data-availability>

[Etiqueta: Alianza Caoba]. (2018). Recuperado de:
<https://www.javeriana.edu.co/pesquisa/tag/alianza-caoba/>

[Experiencias del DANE en medición de indicadores para ODS fueron presentadas en el marco de la Asamblea General de la ONU]. sin fecha. Recuperado de: <https://www.dane.gov.co/index.php/actualidad-dane/4328-experiencias-del-dane-en-medicion-de-indicadores-para-ods-fueron-presentadas-en-el-marco-de-la-asamblea-general-de-la-onu>

Guzmán, W. Riaga, J. (2016). Análisis de Políticas de Datos para el Desarrollo de un Compendio de Buenas Prácticas en Entidades Públicas Colombianas. Tesis de pregrado. Universidad Católica de Colombia. Bogotá D.C. Recuperado de: <https://repository.ucatolica.edu.co/bitstream/10983/14156/1/TRABAJO%20DE%20GRADO%20-%20DATOS%20ABIERTOS.pdf>

Inclusión de los Objetivos de Desarrollo Sostenible en los planes de desarrollo territoriales, 2016-2019. (2017). Departamento Nacional de Planeación (DNP). Recuperado de: https://colaboracion.dnp.gov.co/CDT/Sinergia/Documentos/ODS_en_los_PDT.PDF

Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). (2018). Caracterización de las Principales Causas y Agentes de la Deforestación a Nivel Nacional. Recuperado de: <http://documentacion.ideam.gov.co/openbiblio/bvirtual/023780/Caracterizacion.pdf>

[Integrantes del SEN]. sin fecha. Sistema Estadístico Nacional Colombia (SEN) y Departamento Administrativo Nacional de Estadística (DANE). Recuperado de: <https://www.sen.gov.co/page/SEN>

Ley 1712. Sistema Único de Información Normativa, Bogotá D.C. 06 de Marzo de 2014. Recuperado de: <http://suin.gov.co/viewDocument.asp?ruta=Leyes/1687091>

Ministerio de las Tecnologías de la Información y las Telecomunicaciones. Guía de Datos Abiertos. (2016). Recuperado de: http://estrategia.gobiernoenlinea.gov.co/623/articles-9407_Guia_Apertura.pdf

Naciones Unidas. (2015). Objetivos de Desarrollo del Milenio Informe de 2015. Recuperado de: https://www.un.org/es/millenniumgoals/pdf/2015/mdg-report-2015_spanish.pdf

Perrotta, Daniela. (2018). "Can Mobile Phone Traces Help She Light on the Spread of Zika in Colombia?" Global Pulse. Recuperado de:

https://www.unglobalpulse.org/news/using_mobile_traces_curve_Zika_spread_Colombia

[¿Qué es el Sistema Estadístico Nacional - SEN?]. sin fecha. Departamento Administrativo Nacional de Estadística (DANE). Recuperado de: <https://www.dane.gov.co/index.php/sistema-estadistico-nacional-sen>

Sistema Estadístico Nacional Colombia (SEN). sin fecha. Big Data para el cálculo de la relación de la tasa de consumo de suelo urbano y la tasa de crecimiento de población, mediante el uso de imágenes de sensores remotos.

[Sobre los datos]. sin fecha. Recuperado de: <https://www.ods.gov.co/tools/data-availability>

[Sobre OPAL]. sin fecha. Recuperado de: <https://www.opalproject.org/sobre-opal>

United Nations (2013) Fundamental principles of official statistics. Recuperado de: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx> (consultado 21 Mayo 2019).



Acerca de Cepei

Cepei es un think tank independiente y sin ánimo de lucro, que trabaja a través de la exploración de datos, el análisis de campo, los proyectos de investigación y la incidencia de alto nivel; para ampliar y promover el diálogo, el conocimiento y la participación multi-sectorial en las agendas globales de desarrollo, y hacer una contribución significativa y sostenida al desarrollo sostenible.



BIGDATA
FORDEVELOPMENT
Network



Cepei

Knowledge factory for the world we deserve

Contacto

Fredy Rodríguez

Coordinador del área de Datos

f.rodriguez@cepei.org

Dirección:

Calle 30A # 6-22, bloque B, oficina 2504

Bogotá D.C., Colombia

Teléfono: +57-1-3001051

Con el apoyo de:



IDRC | CRDI

International Development Research Centre
Centre de recherches pour le développement international

Canada

#BD4DGLOBALSOUTH

WWW.BD4D.NET