

EL CONTEXTO DE LA **CALIDAD DE LOS DATOS** EN LA MEDICIÓN DE LOS ODS



BIGDATA
FORDEVELOPMENT
Network



Cepei

Knowledge factory for the world we deserve



Autores

Jaime Gallego
Fredy Rodríguez

Edición

Alexandra Roldán

Diseño y diagramación

David Vargas

Dirección general

Philipp Schönrock

Bogotá D.C., Colombia

Primera versión

Octubre de 2019

Contenido

Introducción | 4

Algunas limitaciones en la generación de datos asociados a la medición de los ODS | 5

La importancia de la calidad de los datos para medir los ODS | 10

1. Desagregación de datos | 10

2. Fuentes de datos | 13

3. Precisión | 15

4. Pertinencia | 16

5. Oportunidad de la información | 17

Estándares de calidad en los datos e interoperabilidad | 20

Conclusiones y recomendaciones | 24

Figuras

Figura 1. Comparación entre la información reportada en el sector público y privado, en el contexto de los ODS | 7

Figura 2. Principales necesidades de fortalecimiento estadístico identificadas por los países de Latinoamérica y el Caribe en el marco de los VNR 2018 | 9

Figura 3. Escala de monitoreo de los ODS | 11

Figura 4. Desagregación de los indicadores para los ODS | 12

Figura 5. Estándares de calidad de los datos para el big data | 14

Figura 6. Clasificación TIERS para indicadores que involucran estadísticas ambientales | 15

Introducción

La época digital ha traído cambios significativos en la gestión de la información para la sociedad globalizada, en la que el uso de los datos es cada vez más extenso, y es la base fundamental para que los tomadores de decisiones asuman desafíos en materia de desarrollo, especialmente cuando se trata del cumplimiento de las agendas globales.

En este contexto, la revolución de los datos surge como un nuevo enfoque integral que busca dar respuesta a las necesidades de la producción estadística, con el objetivo de mejorar las mediciones sobre las diferentes dimensiones del desarrollo sostenible, y dar una mayor relevancia a la perspectiva de bienestar y sostenibilidad ambiental a la visión de los gobiernos.

Del mismo modo, la revolución de los datos promueve la integración de las fuentes de datos no tradicionales como el big data, para ayudar a solventar las falencias de información, especialmente cuando se hace seguimiento al cumplimiento de las metas de los Objetivos de Desarrollo Sostenible (ODS) en el marco de la Agenda 2030. Durante los últimos años, el uso del big data ha pasado de ser una iniciativa local fundamentada en las necesidades de intercambio de información (con fines de marketing) en el sector privado, a convertirse en uno de los principales ejes de acción en escalas regionales de la nueva gobernanza de datos, impulsado por políticas de datos abiertos entre sectores e instituciones.

Sin embargo, uno de los mayores retos que radican en el uso de las fuentes no tradicionales de datos es la calidad de la información, de manera que los datos producidos cumplan con determinados estándares que permitan su utilización y generen el impacto esperado en lo que se refiere al fortalecimiento de las iniciativas locales para el cumplimiento de los compromisos de la Agenda 2030.

A cuatro años del inicio de las presentaciones que realizan los países de sus Informes Nacionales Voluntarios de seguimiento y examen de implementación nacional de los ODS ante el Foro Político de Alto Nivel de las Naciones Unidas, resulta evidente la existencia de brechas entre los datos disponibles y los necesarios, pero también la insuficiente calidad de los datos existentes. En sus reportes, los propios países informantes asumen sus falencias en estas áreas, solicitan apoyo externo para progresar y adoptan planes estadísticos que prevén la utilización de nuevas fuentes de datos antes inexploradas.

La presente investigación busca ser un referente de consulta para instituciones que producen o utilizan datos. Brinda elementos sobre la importancia de la calidad de los datos, al tiempo que establece el punto de partida para futuras investigaciones, especialmente en temas de interoperabilidad de datos como un mecanismo de mejora a la calidad de las mediciones de los ODS.

El documento inicia con la presentación de un contexto general sobre las limitaciones evidenciadas en la medición de los ODS y la producción de datos de calidad, teniendo en cuenta algunos retos intrínsecos a la Agenda 2030 tales como la medición de un gran volumen de indicadores (242), que sumado a las limitaciones técnicas, dificulta su monitoreo y evaluación en algunos países. Posteriormente se analiza el concepto de la calidad de los datos, especialmente cuando se involucran mediciones asociadas al uso de big data y fuentes no tradicionales. Finalmente, se desarrolla una propuesta aplicable a los diferentes contextos de gestión de los Sistemas Estadísticos Nacionales (SEN), basada en los principios de interoperabilidad de datos.

Algunas limitaciones en la generación de datos asociados a la medición de los ODS

Múltiples barreras se han identificado en la implementación de la Agenda 2030, siendo algunas de ellas más complejas de resolver; al requerir la generación de alternativas tales como la definición de un sistema de responsabilidades compartidas, que en el marco del cumplimiento de metas globales demanda el diseño de iniciativas como asistencia financiera y cooperación técnica entre los países desarrollados con economías emergentes, o la identificación de mecanismos financieros que apoyen el cumplimiento de los ODS.

Sin embargo, la mayoría de las limitantes identificadas por los diversos grupos de expertos (principalmente el Grupo Interinstitucional y de Expertos sobre los Indicadores de los Objetivos de Desarrollo Sostenible o IAEG-SDG, por sus siglas en inglés), coinciden sobre **la importancia de mejorar la calidad de los datos** reportados por los Sistemas estadísticos Nacionales (SEN), y la necesidad imperante de fortalecer la capacidad técnica para la producción y procesamiento de datos, especialmente desde la escala subnacional.

En términos generales, la falta de capacidad técnica limita las competencias misionales de las oficinas estadísticas en los países y se traduce en la falta de **metodologías eficientes** que garanticen un adecuado proceso de

recolección, procesamiento, análisis y difusión de la información, para hacer seguimiento al cumplimiento de los ODS.

Actualmente la limitada capacidad de algunas oficinas estadísticas condiciona la medición de logros asociados a la erradicación de la pobreza, seguridad alimentaria, equidad de género, educación y reducción de las desigualdades, los cuales requieren de altos niveles de desagregación de datos, principalmente en aspectos como ingreso, sexo, edad, raza, etnicidad, estado migratorio, discapacidad, ubicación geográfica, entre otros. A lo que se une la falta de temporalidad y de información oficial en estos indicadores.



Cabe resaltar que el monitoreo de los ODS permite fortalecer los procesos de decisiones en materia de política pública y, sobre todo, el diseño de acciones encaminadas a la superación de la pobreza, la lucha contra el cambio climático y el logro de la prosperidad compartida, temas insignias de la Agenda.

La dificultad en la medición no solo debe atribuirse a las limitaciones en la capacidad técnica. También a las obligaciones que supone la implementación del sistema de indicadores de los ODS. Actualmente se conciben cerca de 242 indicadores¹, que incluyen tanto países desarrollados como de economías emergentes y cuya cuantificación supone un **costo elevado**, especialmente para aquellas naciones que no disponen de la infraestructura necesaria para la generación y procesamiento de estadísticas oficiales, y que deben formular nuevos sistemas para la construcción de estadísticas que contribuyan en la medición de los indicadores.

Del mismo modo, otras limitantes se generan desde el contexto de la gobernanza, en el cual para muchos países existen **políticas de datos abiertos** insuficientes, que tienen pocos estímulos para la participación de actores no estatales -como el sector privado- en el reporte de datos. Es claro que el potencial de generación de información proveniente del sector privado es alto en comparación con sectores como el público o la academia, dado que cuentan con suficiente capacidad técnica para

producir información a los niveles de desagregación y confiabilidad requeridos para un adecuado seguimiento a las metas de los ODS.

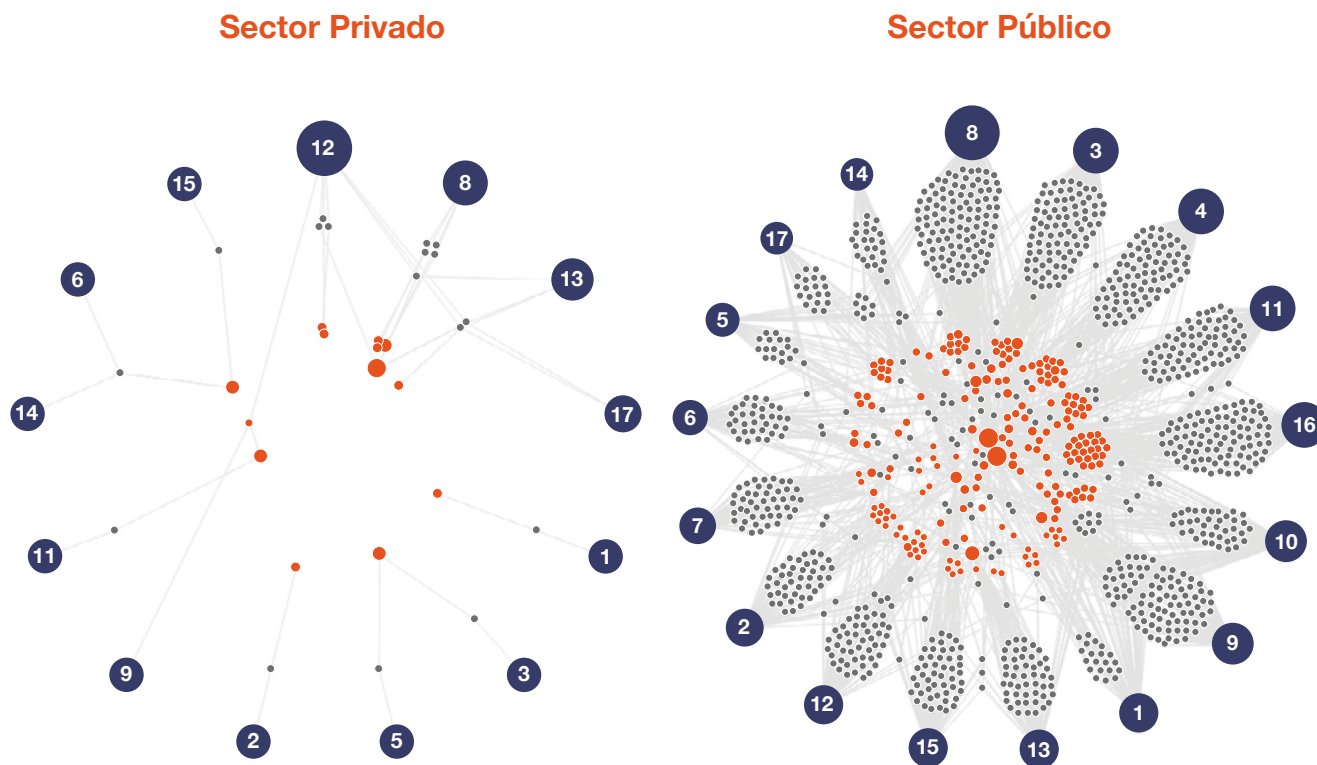
No obstante, la realidad es que existe un alto grado de incertidumbre en torno al interés de las empresas o gremios en aportar datos, lo que conlleva a una restricción de su aporte a la Agenda 2030 en cuanto a la entrega de datos a las oficinas estadísticas, que son las encargadas del reporte oficial de los indicadores ODS.

Un claro ejemplo de lo anterior es la disponibilidad de fuentes de datos presentada en la plataforma [DataRepública](#)², en la que se puede evidenciar la contribución de fuentes de datos del sector público y privado, y la diferencia en materia de aporte de datos (ver el contraste de aporte por sector en la figura 1):

1. Para mayor información sobre los indicadores se sugiere consultar el siguiente link: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework_A.RES.71.313%20Annex.Spanish.pdf

2. [DataRepública](#) es una iniciativa digital de Cepei y Telefónica que busca poner a disposición de los usuarios un laboratorio de datos para el desarrollo sostenible en Latinoamérica.

Figura 1. Comparación entre la información reportada en el sector público y privado, en el contexto de los ODS



Fuente: elaboración propia a partir de DataRepública

Existe una gran diferencia entre la disponibilidad de información proveniente del **sector público y privado**. Mientras que **el primero constituye** cerca del **63,5%** de las fuentes de datos disponibles en DataRepública, **el segundo representa** tan solo el **0,8%**. Sin embargo el problema real no radica en la disponibilidad de fuentes de datos, si no en las restricciones que tiene la información proveniente del sector privado, lo que dificulta su acceso y posterior mapeo.

Finalmente, se considera que otro de los aspectos que tiene gran incidencia en la capacidad de producción de datos en el contexto nacional y subnacional, es la falta de articulación y ausencia de mecanismos de cooperación institucional entre las diferentes organizaciones que hacen parte del SEN. Esta situación conlleva a que la producción de estadísticas sea un proceso

aislado y que se presenten conflictos por la comparabilidad y representatividad de los datos producidos por cada institución. En tal sentido, el deber ser de las entidades es coordinar esfuerzos interinstitucionales para establecer responsabilidades en la producción de datos, reconociendo la especialidad y experticia de cada una de ellas en sus sectores.

Los Reportes Nacionales Voluntarios como un mecanismo de seguimiento y monitoreo a los compromisos de la Agenda 2030 para los países

La implementación de la Agenda 2030 debe ser un proceso dinámico e integral, basado en un enfoque de planeación estratégica que permita dar trazabilidad a los compromisos asumidos por los gobiernos y, sobre todo, permita el diseño de estrategias de alto impacto que se traduzcan en acciones concretas, especialmente desde el contexto local. Para que esto sea efectivo, es necesario contar con mecanismos de seguimiento y monitoreo que documenten el estado y avance en el cumplimiento de los ODS, a su vez que aporten en la formulación de acciones para resolver los retos más apremiantes.

En ese sentido, el Foro Político de Alto Nivel o HLPF, por sus siglas en inglés (*High-level Political Forum*), al ser el principal escenario global para hacer seguimiento y revisión a los Objetivos de Desarrollo Sostenible (ODS) y la Agenda 2030, en el que participan todos los Estados Miembros de la ONU, incentiva a que los países presenten de manera voluntaria sus avances en materia de cumplimiento de los ODS, mediante la generación de Informes Nacionales Voluntarios o VNRs, por sus siglas en inglés (*Voluntary National Reviews*).

La presentación de los VNR se ha generado cada año desde 2016 durante el desarrollo de los respectivos HLPF. A nivel global, 142 países presentaron ya

como mínimo un informe entre 2016 y 2019. Catorce de ellos presentaron más de un reporte. Una mirada atenta a América Latina y el Caribe nos muestra que 21 de sus 33 Estados (64%) han presentado al menos un VNR entre 2016 y 2019 y cinco de estos han presentado dos informes.³

Una de las utilidades de los Informes Nacionales Voluntarios radica en que permiten compartir con el Sistema de Naciones Unidas y la comunidad internacional, los **avances**, **retos**, **oportunidades** y **expectativas** identificados por cada país en sus respectivos contextos de implementación de la Agenda, a la vez que pone a disposición las evidencias y datos a los diferentes públicos interesados en sumarse a las iniciativas nacionales para el cumplimiento de la Agenda 2030.

3. Los países de la región que han presentado segundos informes son Colombia, Chile, Guatemala, México y Uruguay. Por mayor detalle sobre el contexto regional del primer ciclo de trabajo del HLPF (2016-2019) se sugiere consultar el documento [¿De qué hablan los países de ALC cuando informan sobre sus procesos de implementación de los ODS?](#). Por información sobre segundos informes ver [Informes Nacionales Voluntarios de Segunda Generación](#). Ambos documentos han sido producidos por Cepei en 2019.

Figura 2. Principales necesidades de fortalecimiento estadístico identificadas por los países de Latinoamérica y el Caribe en el marco de los VNR 2018



Bahamas

Limitada capacidad técnica de las oficinas estadísticas en materia de compilación, procesamiento, análisis y producción de datos asociados a contribuciones y metas de los ODS, especialmente en lo concerniente a biodiversidad y cambio climático.



México

Necesidad de contar con un sistema estadístico que produzca datos con mayores niveles de desagregación y permita realizar análisis especializados en sectores estratégicos.



Colombia

Necesidad de fortalecer la producción estadística a nivel subnacional, que permita contar con datos confiables y actualizados en los municipios, sobre todo en temáticas como brechas de pobreza y desigualdades en áreas rurales.



Paraguay

Hasta el momento no se cuenta con una línea base de indicadores lo suficientemente robusta para establecer un diagnóstico sobre la contribución directa del país al cumplimiento de las metas de los ODS.



Ecuador

Mayor involucramiento de actores como la ciudadanía y las organizaciones de la sociedad civil.



República Dominicana

Insuficiencia de indicadores especialmente en la dimensión ambiental.



Jamaica

Insuficientes fuentes de datos, cuya confiabilidad, coherencia y calidad de la información suministrada no responde a los requisitos estipulados para la producción y análisis de información en materia de cumplimiento de los ODS.



Uruguay

Necesidad de implementar un modelo estándar para el reporte de información, que presente una única metodología y considere los requisitos mínimos para el reporte de datos asociados a los ODS, de común acuerdo con los marcos internacionales.

La importancia de la calidad de los datos para medir los ODS

1. Desagregación de datos

En lo que se refiere a la calidad de los datos, múltiples características deben ser consideradas por las instituciones y actores encargados de la producción estadística en el momento de generar datos confiables, que a su vez permitan determinar el impacto real en el desarrollo de programas de política pública o en la implementación de los proyectos de intervención con enfoque social. No obstante, en el contexto de la medición de los ODS existen algunos aspectos imperativos que deben ser abordados para el aseguramiento de la calidad y en cuyo detalle se centra el presente capítulo.

La desagregación de datos entendido como el proceso de desglose de los datos, permite generar observaciones más específicas sobre un fenómeno de estudio, a fin de establecer categorías de análisis con un mayor nivel de detalle⁴. En términos generales, este aspecto resulta fundamental para evaluar aspectos relacionados con temas socioeconómicos como: *el análisis de brechas de pobreza, igualdad de género, acceso a educación y equidad social*, cuyo énfasis se basa en la caracterización de los ciudadanos.

4. Para un mayor conocimiento sobre la importancia de la desagregación de los datos en el contexto del desarrollo, se sugiere revisar el artículo: [Información desagregada, elemento clave para la toma de decisiones](#), Espinosa, 2012

Figura 3. Escala de monitoreo de los ODS



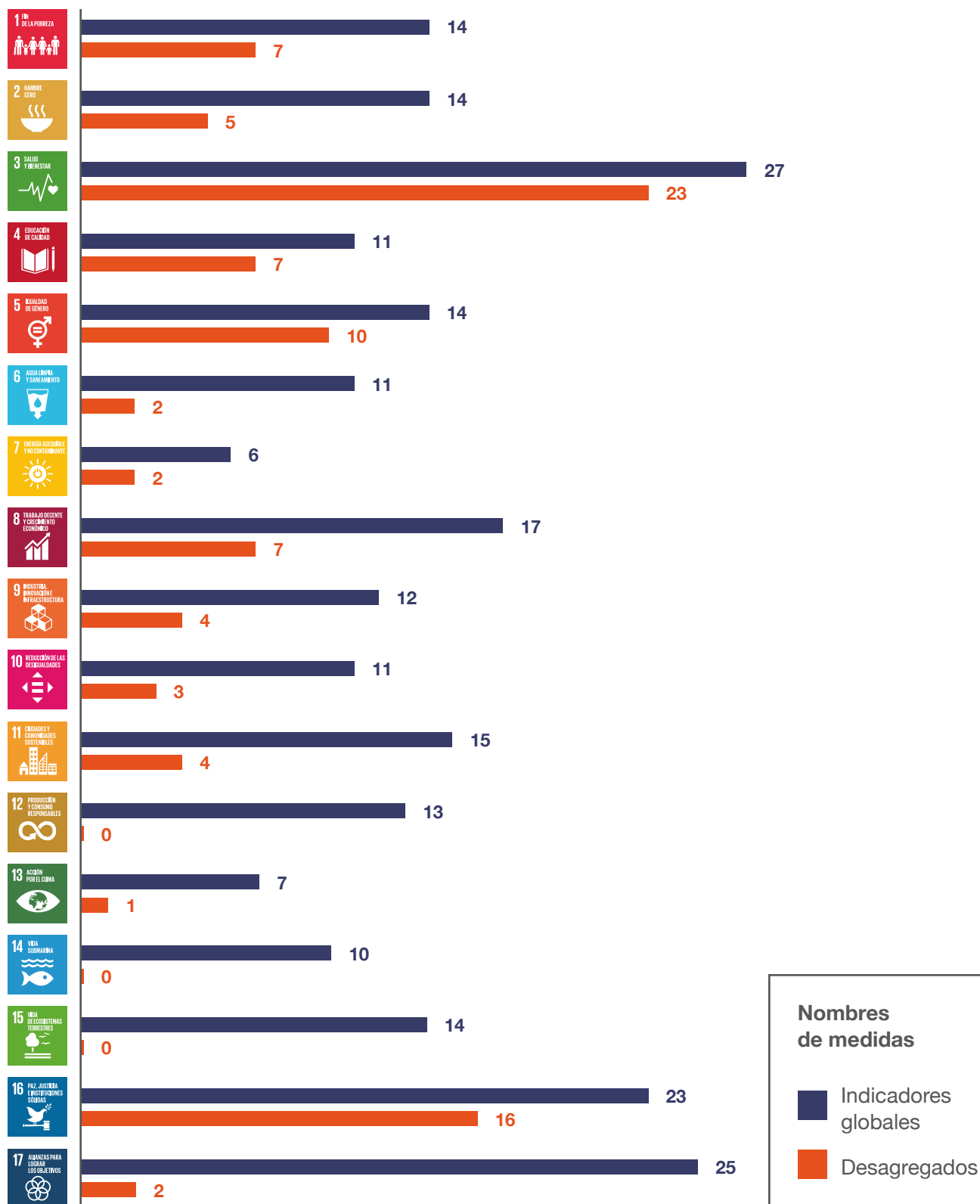
Fuente: elaboración propia

El Grupo de Alto Nivel de Colaboración, Coordinación y Fomento de la Capacidad, o por sus siglas en inglés HLG-PCCB de Naciones Unidas, sugiere que la correcta desagregación de los datos, en el marco de los procesos de monitoreo y seguimiento a los ODS, debe hacerse desde la escala nacional hasta la escala local (en la medida de lo posible). Sin embargo, esto supone un alto costo para los países, especialmente aquellos que no cuentan con la adecuada capacidad técnica, que garantice un eficiente proceso de recolección, procesamiento y difusión de datos.

En el contexto de los ODS, la desagregación debe hacerse en al menos 94 indicadores (CEPAL, 2018), sobre los cuales se pretenden generar mediciones en aspectos sociales tales como *ingreso, sexo, edad, raza, etnicidad, estado migratorio, discapacidad y ubicación geográfica*, o cualquier otra cualidad que estudie las características sociales, económicas y demográficas de los ciudadanos, de conformidad con los principios Fundamentales de las Estadísticas Oficiales (resolución 68/261 de la Asamblea General de las Naciones Unidas).

De acuerdo con la división de Estadísticas de la CEPAL, existen algunos ODS que requieren un mayor nivel de desagregación tales como los son el 3,4,5 y 16, cuyos niveles de desagregación en materia de indicadores corresponde a 85%, 64%, 71% y 70% respectivamente, y sobre los cuales se concentra cerca del 60% de los indicadores actualmente desagregados, tal y como se evidencia en la figura 4.

En contraste, aquellos indicadores relacionados con los ODS ambientales como lo son los ODS 6, 12, 13, 14 y 15 presentan los menores niveles de desagregación en comparación con todos los ODS en general (para los ODS 6 y 14, solo se considera que el 14% y 18% de los indicadores deben tener algún grado de desagregación). No obstante, **todos los ODS deben tener un alto grado de desagregación** si realmente se proyecta tomar decisiones de incidencia significativa con el cumplimiento de las metas de la Agenda 2030. Para los ODS ambientales es importante llegar al detalle de la información a nivel local, dado que los territorios tienen particularidades en materia de uso del suelo y de oferta de bienes y servicios ecosistémicos, los cuales desde el orden nacional o subnacional no es posible reconocer, lo que puede conllevar a la inadecuada formulación de políticas de conservación y uso de recursos naturales.

Figura 4. Desagregación de los indicadores para los ODS

Fuente: elaboración propia a partir de CEPAL, 2018

2. Fuentes de datos

La versión más reciente de los Principios Fundamentales de las Estadísticas (Comisión de Estadística de las Naciones Unidas, 2014), indica que para la construcción de las Estadísticas Oficiales se pueden usar diferentes tipos de fuentes, que incluyen desde mecanismos de recolección por medio de encuestas oficiales, registros administrativos y censos, hasta las fuentes no tradicionales asociadas principalmente al uso del big data, imágenes satelitales y cualquier otro tipo de información producida por instituciones (público-privadas) fuera del SEN.

Particularmente para las fuentes de datos no tradicionales, el mayor desafío radica en asegurar la calidad de los datos, ya que el SEN, en el marco de sus competencias misionales en cada país, ha implementado estándares para el reporte de la calidad de los datos, y cuya adopción no siempre es obligatoria para las instituciones por fuera de este. No obstante, cabe resaltar que la mayoría de la información generada por el sector privado (desde el contexto de la investigación de mercados y mercadeo) cuenta con muy buenos niveles de calidad (principalmente en términos de desagregación, actualización, consistencia, precisión, exactitud e integralidad), y suele ser de gran utilidad para medir aquellos indicadores ODS asociados a aspectos como: salarios e ingresos familiares (ODS 1), acceso y nivel de educación (ODS 4), equidad de género en el contexto laboral, empleabilidad y nivel de ingreso (ODS 5), empleabilidad formal e informal (ODS 8), empleabilidad por sectores, inversión en investigación y desarrollo (ODS 9), producción más limpia a nivel industrial (ODS 12) y alianzas productivas para el desarrollo sostenible (ODS 17)⁵.

En definitiva, el big data es una oportunidad **estratégica que orienta al sector empresarial en la incorporación de los ODS**, mediante el

diseño de modelos que permiten compartir datos de forma solidaria, y que contribuyen al bien social (filantropía de datos), tal como suele ocurrir con las empresas del sector de telecomunicaciones y servicios financieros, con cuyos datos es posible hacer monitoreos de indicadores de ODS con enfoque social (dimensión persona y prosperidad⁶), que requieren grandes niveles de desagregación.

Sin embargo, para que la aplicación del big data al proceso de seguimiento y monitoreo de los indicadores ODS sea efectiva, es necesario que cumpla con determinados estándares de información, siendo el principal referente el **Modelo de Calidad de los Datos** adoptado por la Norma ISO 25012, la cual indica algunas propiedades de los datos que son totalmente aplicables al contexto del big data para el desarrollo sostenible.

El primer aspecto relacionado con la **calidad de los datos para el big data** se asocia con la **coherencia** de los datos, en la que se evalúa la pertinencia en la aplicación de los datos para documentar algún tema en específico, es decir, si los datos pueden usarse aun cuando no tengan relación directa con la temática de interés (complementariedad).

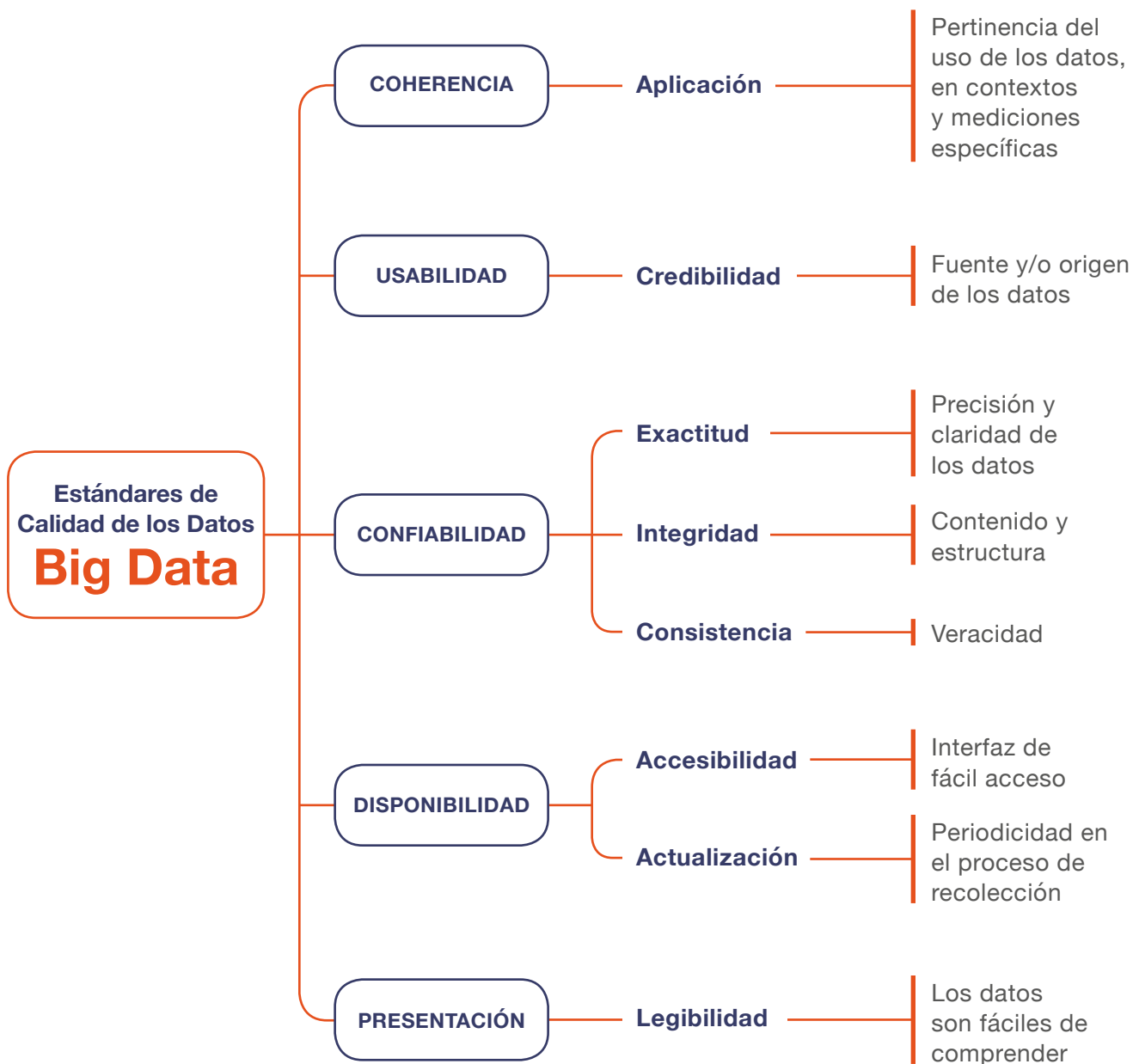
5. En 2017, La Cámara de Comercio de Bogotá (CCB), la Red Pacto Global Colombia, Fundación Bolívar Davivienda, Fundación Corona y Cepei desarrollaron la iniciativa “Unidos por los ODS: el aporte de nuestras empresas”, proyecto que midió la contribución de 21 empresas al cumplimiento de los ODS en Colombia, y sobre el cual se sustentan las presentes afirmaciones. Consulte el informe aquí http://cepei.org/wp-content/uploads/2018/11/Informe-general-Unidos-por-los-ODS.Vf_.pdf

6. <http://www.congdcar.org/es/sensibilizacion/ods.html>

El segundo se relaciona con la **credibilidad**, la cual depende, en gran medida, de los estándares implementados por la institución que produce los datos. Se enfoca especialmente en los criterios de calidad usados para el levantamiento de la información. El tercero es la **disponibilidad**, que se encuentra ligada a los mecanismos y canales a los cuales es

posible acceder a la información, por ejemplo: datos abiertos. El cuarto es la **confiabilidad**, en la cual se evalúan la precisión, la veracidad y la coherencia en la estructura y contenido de los datos. Finalmente, el último aspecto determina la **legibilidad**, es decir, la forma en la que se presentan los datos (formatos abiertos y editables, imágenes, pdf, etc).

Figura 5. Estándares de calidad de los datos para el big data



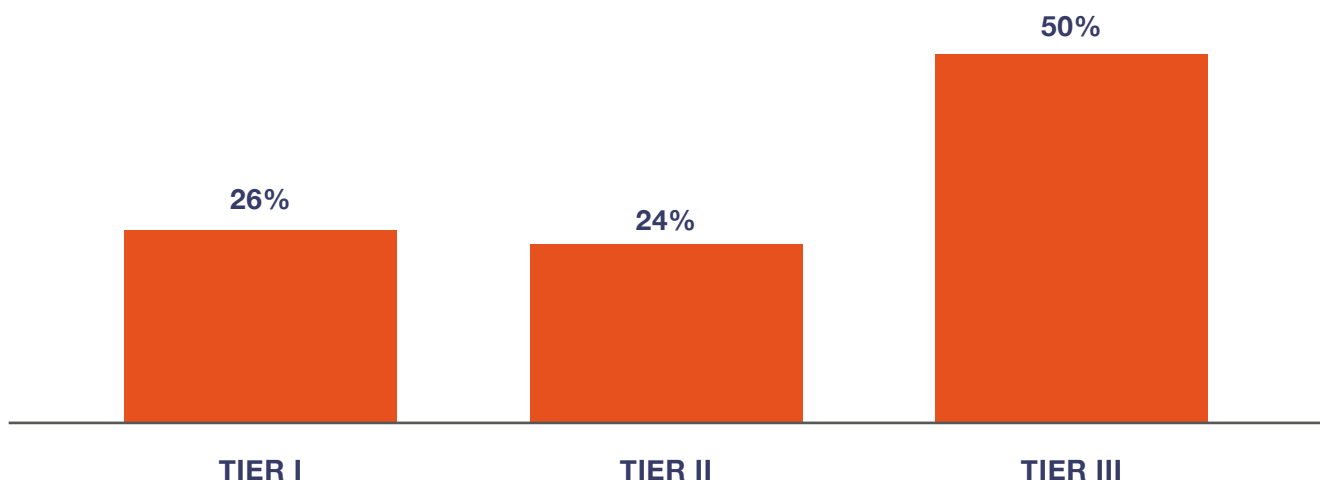
Fuente: elaboración propia a partir del Modelo de Calidad de los Datos (Norma ISO 25012)

3. Precisión

La precisión es la exactitud con la que los datos interpretan los fenómenos de observación, que para este caso concreto son las mediciones para el seguimiento y cumplimiento a las metas de los ODS. En el contexto de la calidad de los datos la precisión está determinada por los **parámetros** y **variables** usadas en el diseño de los indicadores ODS.

Las mayores dificultades para la precisión de los datos se encuentran en los indicadores de los ODS que involucran mediciones a través de estadísticas ambientales, (representando cerca del 51% de los indicadores totales). Y para los cuales el 50% todavía requiere el desarrollo de estándares y metodologías internacionalmente válidas. Mientras que tan solo un 26% cuenta con metodologías robustas y datos de suficiente cobertura para garantizar mediciones precisas (ver figura 6).

Figura 6. Clasificación TIERS para indicadores que involucran estadísticas ambientales



Fuente: CEPAL, 2017

Un claro ejemplo de la transversalidad de las estadísticas ambientales puede evidenciarse en el ODS 3, *Salud y Bienestar*, el cual incluye indicadores como tasa de mortalidad atribuida a contaminación ambiental (3.9.1 – TIER I), mortalidad atribuida al agua no apta para el consumo (3.9.2 – TIER II) y tasa de mortalidad atribuida a la intoxicación accidental (3.9.3 – TIER II).

No obstante, una de las principales limitaciones en materia de medición y precisión de los indicadores

que involucran estadísticas ambientales, es la dificultad que representa medir la biodiversidad a través de indicadores deterministas, para los cuales, si bien se han adaptado enfoques simplificados que permiten la identificación de cambios en el estado de conservación del capital natural, aún es necesario desarrollar metodologías más específicas que garanticen el monitoreo de procesos ecológicos más complejos a nivel general.

4. Pertinencia

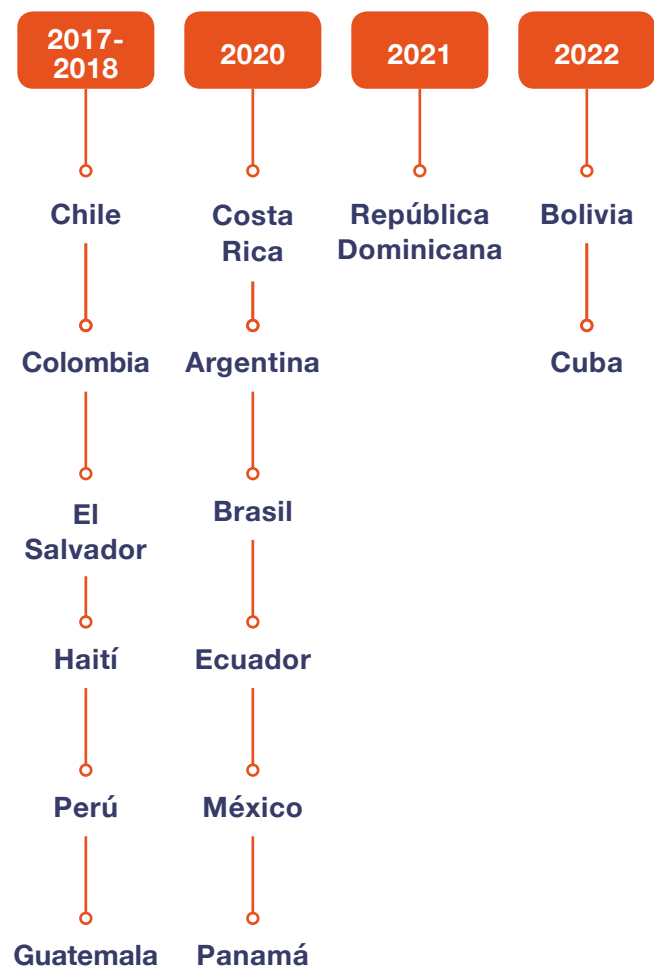
La definición de necesidades de información está fundamentada en los requerimientos formulados por los usuarios, que pueden ser dados por el gobierno, sectores específicos o la ciudadanía en general. A su vez, la información debe ser presentada con la calidad y periodicidad adecuados necesarios para tomar decisiones basadas en evidencia.

Un ejemplo de lo anterior, es la **aplicación de los censos** para determinar las dinámicas de crecimiento poblacional y la construcción de estadísticas sobre las condiciones sociales, económicas y demográficas que orienten la generación de políticas públicas. Actualmente, Naciones Unidas recomienda la realización de censos poblacionales cada 10 años, en los años terminados en cero, con el objetivo de que los programas de desarrollo de los gobiernos de turno tengan mayor coherencia con las realidades de los territorios (especialmente en lo concerniente al tamaño, crecimiento y composición de la población). Sin embargo, en países de la región como **Colombia**, han existido limitaciones de orden institucional que han retrasado la realización de los dos últimos censos poblacionales (entre 5 y 3 años, respectivamente), sin contar con que no estaban alineados a los años cero. Así como del Censo Nacional Agropecuario, que hasta su versión más reciente (2014) tuvo más de 40 años de desactualización. Otro ejemplo es el del **Atlas de Riesgo de la Ciudad de México**, el cual se encuentra desactualizado desde el año 2013, y cuya gestión está a cargo de las Comisiones Unidas de Desarrollo e Infraestructura Urbana y de Protección Civil.

Los últimos años han sido determinantes para la planeación y presentación de los Censos en los países de la región. Se conoce que actualmente seis (6) países presentaron sus respectivos censos

de población entre el periodo 2017 y 2018, mientras que nueve (9) ya tienen fechas para la elaboración de los mismos entre el 2020 y 2022.

Figura 7. Estado de los Censos en los países de la región de ALC



Fuente: elaboración propia a partir de Celade

5. Oportunidad de la Información

La oportunidad de los datos es un aspecto que se relaciona intrínsecamente con el **periodo de referencia** y el **tiempo de publicación de los datos**. Siendo el primero un factor determinante para asegurar la recolección de información que refleja la realidad del fenómeno en un periodo de tiempo determinado, que permite contar con información necesaria para la construcción de los indicadores ODS, y sobre el cual se configura una línea base que conlleve a la generación de evidencias para hacer un seguimiento efectivo al cumplimiento de las metas ODS.

Respecto al **tiempo de la publicación** de los datos, se deben considerar las necesidades de los usuarios para que la información les sea útil y con la actualización requerida para la toma de decisiones. Los aspectos técnicos (tratamiento, análisis, resultados) y las estrategias de difusión (diagramación y comunicación) que influyen en la entrega de los datos, permiten identificar aquellos puntos críticos que pueden implicar retrasos significativos.

Finalmente, uno de los mejores ejemplos que vincula el concepto de oportunidad de la información se

relación con los *SDGs Dashboard* usados para hacer monitoreo y seguimiento a los ODS en los países. Actualmente se constituyen como una herramienta interactiva que permite la consulta de las metas e indicadores a nivel nacional, a la vez que evidencia los principales retos y avances significativos en el cumplimiento de la Agenda 2030, y sobre los cuales es posible focalizar esfuerzos y recursos para generar estrategias de alto impacto que permitan superar las limitaciones evidenciadas en cada contexto de los países.

En términos generales el análisis de los informes voluntarios nacionales o VNR, permite identificar algunos retos que los países de la región deben afrontar, con el fin de consolidar mejores mecanismos para la medición de los ODS, en especial los Dashboards, siendo el más relevante el llamado a que los mismos reporten tanto el progreso y avance, como el estancamiento o retrocesos en el cumplimiento de los ODS, ya que esto realmente evidencia el estado real de la implementación de la agenda 2030 en los países.

Tabla 1. Contexto de los Dashboards ODS en los países de la región de ALC



Argentina



Belize



Bolivia



Brasil



Chile



Colombia



Costa Rica



Cuba



Ecuador



El Salvador



Guatemala



Guayana



Haití



Honduras



México



Nicaragua



Paraguay



Perú



República Dominicana



Trinidad y tobago



Venezuela

¿Qué dice el Código Regional de Buenas Prácticas Estadísticas para América Latina y el Caribe sobre el aseguramiento de la calidad estadística?

Con base en los principios establecidos por el código, se ratifica el deber de las instituciones productoras de estadísticas adscritas al SEN, de trabajar por el aseguramiento de la calidad estadística, conforme a las normas internacionales.

Desde el código, el Principio 2 hace un llamado a que las instituciones pertenecientes al SEN

trabajen articuladamente en un marco de cooperación conjunta, con el objetivo de mejorar la calidad, comparabilidad y coherencia de las estadísticas oficiales.

El principio 6 exhorta a los respectivos SEN a que consideren los siguientes aspectos de calidad:

- 1** Definir la política y modelo de calidad, que en consecuencia deben ser adoptados y ratificados por los miembros del SEN.
- 2** Implementar procesos de evaluación continuos y periódicos, con base en los estándares internacionales sugeridos por la Comisión Estadística de las Naciones Unidas.
- 3** Implementar procesos de control de calidad.

Fuente: CEPAL, 2012

En el principio 9, se estipula que la calidad de los datos es un aspecto que depende en gran medida de los métodos de recolección de información, especialmente cuando se trata de registros administrativos.

Por último, el Principio 10 recalca la necesidad de implementar procedimientos, metodologías y tecnologías que garanticen la calidad de las estadísticas durante las etapas de recolección, procesamiento y difusión de los datos.

Estándares de calidad en los datos e interoperabilidad

En términos generales, la interoperabilidad es definida como la *capacidad para unificar diferentes conjuntos de datos sin perder su esencia original o significado* (JUDS, 2016)⁷, de tal forma que contribuya a la generación de estadísticas de mayor calidad, y permita el mayor aprovechamiento posible de la información. La *interoperabilidad* abarca toda la cadena de valor de los datos e implica mejoras sustantivas en las etapas de recolección, procesamiento, análisis y difusión. Así mismo garantiza que la información responda a las necesidades identificadas, y aporte el impacto esperado en los diferentes contextos de aplicación.

Desde la perspectiva de datos para el desarrollo sostenible, el enfoque de la *interoperabilidad* resulta útil en la **construcción y adopción de un marco común**, a manera de protocolo, que facilite esquemas de cooperación coordinados y coherentes entre las instituciones estadísticas oficiales, y los sectores/entidades que producen datos o son usuarios de los mismos, para consolidar fuentes de datos que brinden un soporte efectivo al diseño de los indicadores ODS.

En consecuencia, dicho marco define los aspectos de mayor importancia que influyen de manera positiva en el mantenimiento de los ecosistemas de datos para el desarrollo sostenible, y cuya síntesis se basa en la propuesta de Goldstein (2018)⁸, la cual define una hoja de ruta trazable para el adecuado proceso de recolección, intercambio y uso de datos con enfoque de bien común. De esta manera, se

distinguen al menos cuatro (4) perspectivas que deben ser abordadas en cualquier sistema de planificación estadística con miras a la generación de datos para el desarrollo sostenible:

Tecnología

Canales y mecanismos que garantizan el acceso y la publicación de la información a través de interfaces estandarizadas en la red. En coherencia con los datos para el desarrollo sostenible, puede asociarse a las plataformas de monitoreo y seguimiento de los ODS (*SDG Dashboards*), las cuales han sido implementadas en al menos el 83,9% de los países que han adoptado la Agenda 2030 como un compromiso nacional.

La importancia de los *dashboards* está dada por poner a disposición evidencias sobre las cuales debe basarse el diseño de estrategias de los países para avanzar en el cumplimiento

7. The Joined up Data Standard es una iniciativa apoyada por la Publish What You Fund en conjunto con la Omidyar Network, la cual tiene por objetivo apoyar el uso de datos para fortalecer los procesos de toma de decisiones.

8. Data Commons Version 1.0: A Framework to Build Towards AI for Good, A roadmap for data from the 2018 AI for Good Summit; Berkman Klein Center.



de la Agenda 2030. Sin embargo, una lectura más detallada pone al descubierto el desequilibrio de capacidades en las regiones para medir las contribuciones de los ODS⁹.

Adicionalmente se identifican diferentes formas de implementar *dashboards*, no solo en la presentación y plataformas tecnológicas, también en los indicadores que publican y su periodicidad.

Datos y formatos

Definición de esquemas de modelado de datos que permitan una adecuada estructuración de los datos y metadatos. Esto implica la incorporación de un enfoque de transversalidad de la información en la cadena de valor, que incentive la unificación de un lenguaje de lectura de datos, para que las diferentes organizaciones (tanto aquellas que hacen parte o no del SEN) tengan una misma estructura para el reporte de datos en términos de atributos y variables.

Sin embargo, tener claridad sobre el alcance de cada una de las variables no es una tarea fácil, considerando que se pueden tener diferentes unidades de medida para una misma variable. Por ejemplo, en el caso de los censos, la variable población puede estar definida por el número de familias o desagregada a nivel de número de habitantes. Lo que en la práctica dificulta el establecimiento de un lenguaje común para el reporte de datos. La falta de coordinación en algo tan fundamental como lo es la definición de unidades de medidas, es producto de la planeación estadística individual de las instituciones, que en la ausencia de una directriz estándar para el reporte de datos, establece sistemas de medición para cada caso en particular y, en algunos casos, introduce nuevas métricas que complejizan los procesos de recepción y uso de la información.

Por lo tanto, el paso más importante para el establecimiento de un sistema de reporte y colección

de datos estandarizado en las instituciones, es la definición del conjunto **dimensiones** y **medidas**, aplicables a cada contexto de análisis (social, económico, ambiental), que simplifique el lenguaje de las variables y contenga los criterios mínimos para la asignación de métricas.

En la comunidad estadística, principalmente las oficinas de estadística, han implementado el estándar SDMX (Statistical Data and Metadata eXchange). Es así como el grupo de trabajo en SDMX¹⁰ del IAEG-SDGs viene creando modelos con estructuras de datos que permitan el intercambio de información para la medición de los ODS.

Medida

Corresponden a indicadores que pueden cuantificarse para determinar magnitudes de cantidad.

Dimensión

Hace referencia a aquellos atributos que permiten categorizar, segmentar y/o agrupar los detalles de los datos. Normalmente se asocian a aspectos cualitativos.

9. De acuerdo con el informe [SDG Index and Dashboards 2018](#), más del 60% de los países en Europa Occidental tienen un avance significativo en el cumplimiento de la Agenda 2030; en contraste con África, cuyo porcentaje de países han mostrado una mediana contribución a la Agenda, menor al 17%.

10. El grupo de trabajo de SDMX tiene como principal objetivo el desarrollo de definiciones globales de estructura de datos y estructura global de metadatos para los ODS.

En definitiva, esto garantiza la integralidad de la información y supone una mayor articulación en la producción de datos de las instituciones dado que evita duplicaciones de las mediciones para variables específicas.

Lenguaje

Entendimiento común entre los usuarios y productores de datos, especialmente en relación con la terminología para describir el contenido y uso apropiado de los mismos.

Gobernanza

Este aspecto determina la capacidad con la que las instituciones pueden gestionar y mejorar los datos que producen. Dentro del contexto del desarrollo sostenible, resulta fundamental definir esquemas de gobernanza integral, que promuevan la cooperación interinstitucional, especialmente con aquellas organizaciones que no hacen parte del SEN, pero que, como se explicó anteriormente, tienen un gran potencial de aportar al ecosistema de datos para el desarrollo.

Figura 8. Enfoques de gobernanza en el marco de la interoperabilidad



Fuente: elaboración propia

Actualmente, las instituciones deben centrarse en identificar cuál es el enfoque de gobernanza que más se adapta a sus necesidades. Para instituciones oficiales tales como las Oficinas Nacionales de Estadísticas, se asumen esquemas de gobernanza

centralizados, donde se lideran todos los procesos de recolección y gestión de los datos, a la vez que se encargan de verificar y validar los estándares de calidad bajo los cuales se reporta la información en la escala nacional.



En consecuencia, a instituciones y sectores que se consolidan como actores estratégicos de datos para el desarrollo por fuera del SEN (principalmente sector privado y sociedad civil), les conviene usar enfoques descentralizados o flexibles, que les permitan coordinarse para mantener la coherencia de sus estándares, procedimientos y políticas de gobernabilidad de datos, y que, aun con diferentes cronogramas y capacidad técnica, logran una sinergia efectiva que aproveche el potencial de fuentes no tradicionales como el big data o datos de la ciudadanía¹¹.

11. De acuerdo con la [Global Partnership for Sustainable Development Data \(GPSDD\)](#) los datos generados por la ciudadanía (DGC) son un tipo de datos centrados en problemas que pueden tomar muchas formas. A menudo se enmarcan en torno a personas que colaboran para recopilar datos que necesitan para comprender y abordar un problema que les afecta. Los DGC pueden contribuir a monitorear el desempeño de los gobiernos, rastrear problemas ambientales o recolectar datos espaciales. Para mayor información sobre los DGC consulta el siguiente documento: http://www.data4sdgs.org/sites/default/files/2019-06/Advancing%20Sustainability%20Together%20Summary%20Report_ES.pdf

Conclusiones y Recomendaciones

El análisis de los **Informes Nacionales Voluntarios** de los países de la región elaborado por Cepei evidencia la necesidad que tienen los sistemas estadísticos nacionales de fortalecer capacidades para el aseguramiento de la calidad de los datos. Elementos como la ausencia de procedimientos que produzcan los niveles de **desagregación** y **confiabilidad** requeridos, los pocos incentivos para vincular las **fuentes no tradicionales de datos** al SEN o la falta de estándares operativos que permitan una actuación coordinada de las instituciones en materia de reporte de datos, son las limitantes más comunes que presentan las oficinas estadísticas.

La **interoperabilidad** en el contexto de los datos para el desarrollo sostenible promueve el uso de estándares para el reporte de datos y metadatos (incluyendo aquellos que se generan a partir de métodos de recolección como encuestas y registros administrativos), proveniente de diferentes fuentes, con la finalidad de coordinar esfuerzos y generar acciones colaborativas entre las diferentes instituciones productoras de datos (incluyendo aquellas que no hacen parte del SEN). Y, así mismo, reducir las brechas de información existentes en algunos indicadores ODS, en aspectos como la medición de la pobreza, la desigualdad social y el enfoque de género.

La **calidad de los datos** debe ser un proceso dinámico, en el cual se haga un seguimiento a la trazabilidad en el proceso de generación de estadísticas, mediante la adopción de estándares para su reporte. En la actualidad, las oficinas estadísticas de países en la región como México, Ecuador y Costa Rica cuentan con modelos de aseguramiento de la calidad, en los que se fomenta desde el uso de manuales para la producción estadística hasta la aplicación de procesos de evaluación y certificación de la calidad de datos, basado en estándares internacionales.

Por otra parte, es deber de los gobiernos **promover una pedagogía más amplia en materia de ODS,**

dirigida al sector privado, la ciudadanía, la academia y la sociedad civil, con el objetivo de incentivar su participación en el fortalecimiento de los ecosistemas de datos para el desarrollo sostenible, y, de esta forma, aprovechar el gran potencial de información de las fuentes no tradicionales.

El enfoque de **gobernanza** propuesto para el abordaje de la **interoperabilidad** también es un llamado para que los institutos oficiales de estadística apoyen las fuentes no tradicionales en el mejoramiento de la calidad de sus estadísticas. Un claro ejemplo de lo anterior es la promoción del uso de estándares internacionales de datos, como el sistema para la clasificación de registros, que aplica en principio para fuentes oficiales.

Por último, se recomienda que las instituciones adopten un **modelo de gobernanza** que permita la colaboración efectiva entre los actores que hacen parte o no del SEN, y constituya una red de datos y metadatos descentralizada que facilite la gestión del intercambio de datos, a la vez que proporcione las condiciones para que los diferentes miembros de la misma interactúen entre sí. La principal característica que debe tener este enfoque es el aprovechamiento de los estándares y tecnología para permitir la colaboración y la implementación de mecanismos comunes de gestión de datos.



Bibliografía

Comisión Económica para América Latina y el Caribe. (2012). Código regional de buenas prácticas en estadísticas para América Latina y el Caribe [en línea], disponible en <https://www.cepal.org/es/publications/type/documentos-organos-subsidiarios>

Comisión Económica para América Latina y el Caribe. (2017). Los censos de la ronda 2020: desafíos ante la Agenda 2030 para el Desarrollo Sostenible, los Objetivos de Desarrollo Sostenible y el Consenso de Montevideo sobre Población y Desarrollo [en línea], disponible en https://repositorio.cepal.org/bitstream/handle/11362/42394/S1700849_es.pdf?sequence=1&isAllowed=y

Comisión Económica para América Latina y el Caribe. (2019). Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas [en línea], disponible en <https://www.cepal.org/es/publicaciones/44214-desagregacion-datos-encuestas-hogares-metodologias-estimacion-areas-pequenas>

Commission, U. N. (2017). Guidelines and Best Practices on Data Flows and Global Data [en línea], disponible en https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-06/20171108_Draft%20Guidelines%20and%20Best%20Practices%20for%20Global%20SDG%20Data%20Reporting.pdf

Franco, B., & Pérez, R. (2014). Directrices de Interoperabilidad [en línea], disponible en <http://mgd.redrta.org/mgd/site/artic/20150123/>

[asocfile/20150123112234/g_03_d01_g_directrices_interoperabilidad_adme.pdf](https://www.cepal.org/es/publicaciones/44214-desagregacion-datos-encuestas-hogares-metodologias-estimacion-areas-pequenas)

Goldstein, E., Gasser, U., & Ryan, B. (2018). Data Commons Versión 1.0: A Framework to Build Toward AI for Good [en línea], disponible en <https://medium.com/berkman-klein-center/data-commons-version-1-0-a-framework-to-build-toward-ai-for-good-73414d7e72be>

Luis González, M., & Tom, O. (2018). Data in interoperability: a practitioner's guide to joining up data in the development sector [en línea], disponible en <http://www.data4sdgs.org/resources/interoperability-practitioners-guide-joining-data-development-sector>

Naciones Unidas, Consejo Económico y Social. (2016). Informe del Grupo Interinstitucional y de Expertos sobre los Indicadores de los Objetivos de Desarrollo Sostenible [en línea], disponible en <https://unstats.un.org/unsd/statcom/47th-session/documents/2016-2-IAEG-SDGs-S.pdf>

Robert L., G. (2017). Supporting Open Data and Open Science With Data Commons: Some Suggested Guidelines for Funding Organizations [en línea], disponible en https://www.healthra.org/wp-content/uploads/2018/08/Data-Commons-Guidelines_Grossman_8_2017.pdf



BIGDATA
FOR **DEVELOPMENT**
Network



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

#BD4DGLOBALSOUTH

WWW.BD4D.NET