



Knowledge from the Global South

ALCANCES METODOLÓGICOS DE MAPEO DE FUENTES DE CONJUNTOS DE DATOS Y ALINEACIÓN CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE

Conecta - DataRepública



CONECTA

Conecta es un centro digital de fuentes de conjuntos de datos relacionados con los Objetivos de Desarrollo Sostenible (ODS) que han sido producidos por actores tradicionales y no tradicionales en los sectores público, privado, academia, filantropía y agencias multilaterales.

La información está organizada para que los usuarios rastreen cómo los países de América Latina avanzan hacia el cumplimiento de cada ODS. Asimismo, busca completar las brechas de conocimiento y proporcionar herramientas fáciles de usar para incorporar nuevas fuentes de conjuntos de datos. Para ello, Conecta pone a disposición:

—o **Un catálogo basado en web**, que permite encontrar de manera rápida datos sobre

países priorizados y cada ODS. Incluye información sobre los atributos de calidad de cada fuente de datos (metadatos).

—o **Visualizaciones amigables** que permiten a los usuarios descargar la información que necesitan y reutilizarla en otros estudios.

—o **Análisis y conocimientos**: informes anuales de investigación que incluyen una mirada a los perfiles de los países y sus políticas de datos abiertos, mecanismos de gobernanza de los ODS, sistemas estadísticos nacionales, disponibilidad de datos y progreso en países prioritarios (primer informe en 2021), análisis comparativo y tendencias entre países.

MAPEO DE FUENTES DE CONJUNTOS DE DATOS ASOCIADOS A LOS ODS

El sector público, (en especial las Oficinas Nacionales de Estadística), es el mayor productor de datos disponibles en línea. Esto no significa que los demás actores no generen datos, sino que estos no son de libre acceso por motivos de confidencialidad, competencia de mercado, actividades misionales, entre otros.

La amplia diversidad de temáticas según la experticia de los productores, el contexto y prioridades del país, el enfoque de responsabilidad social de algunos actores, las investigaciones

académicas y científicas, entre otros, producen una diversidad de datos sin precedentes que tienen una relación directa o indirecta con los ODS. Sin embargo, esta diversidad no se encuentra vinculada de forma explícita.

A través de un algoritmo de búsqueda, Cepei realiza el mapeo de fuentes de conjuntos de datos que se encuentran vinculadas a los ODS en los países participantes en Conecta. En la actualidad, este ejercicio ha sido replicado en Argentina, Colombia, Costa Rica, México y Paraguay.

En la figura 1, se presenta el proceso del mapeo de fuentes de conjuntos de datos conformado por siete pasos que estructuran una búsqueda robusta en los portales de datos abiertos, a través de una serie de criterios y validaciones tanto sistémicas como humanas. Una vez el catálogo se encuentra

diligenciado y gracias al mapeo efectuado por el algoritmo, se realiza una validación manual de la calidad de los datos, su alineación con los ODS y completitud de cada variable. En caso de ser necesario, se efectúa una búsqueda manual para ingresar las variables pendientes en la matriz.

Figura 1. Proceso de mapeo de fuentes de conjuntos de datos en Internet



Fuente: elaboración propia

01 Definición de palabras clave para cada Objetivo de Desarrollo Sostenible (ODS).

Para cada uno de los ODS se define una lista de palabras clave relacionadas con este a partir de las palabras relevantes que componen el metadato de los ODS propuestos en el [repositorio de metadatos de la ONU-DS](#). Por ejemplo, para el [ODS 1: Poner fin a la pobreza en todas sus formas en todo el mundo](#), algunas palabras clave son: ingreso per cápita, recursos económicos, recursos naturales, protección social, entre otros.

02 Identificación de portales de datos con Interfaz de Programación de Aplicaciones (API¹ por su sigla en inglés).

Para el desarrollo de este punto, deben tenerse en cuenta los siguientes pasos:

- a. Búsqueda del portal de los Institutos Nacionales de Estadística, portales gubernamentales u otro actor del país a mapear que cuente con datos.
- b. Revisión de la arquitectura del portal de datos para verificar si funciona con una API.
La elección de la API es priorizada por el hecho de que la búsqueda de datos en diferentes páginas web puede ser complicada y difícil de uniformizar.

03 Análisis de la definición y protocolo de la API.

Una vez identificado el portal de datos con API, es importante comprender la API. En general, cada API cuenta con un manual de uso que contiene una definición de las diferentes variables a las que puede accederse con la API y una explicación de los datos que contiene. Por lo tanto, debe familiarizarse tanto con la API como con su contenido. El análisis finaliza con la identificación del proceso de extracción de información con la API

en línea con las variables que mapeamos para la matriz de conecta.

04 Conexión al punto final de las fuentes de datos que describen los datos o metadatos.

A través de los distintos puntos finales de API, es posible acceder a diferentes tipos de información. Por lo tanto, se requiere consultar el punto final del API que proporciona los metadatos del portal de datos. Un metadato es un dato que describe otros conjuntos de datos.

05 Diligenciamiento de la matriz de metadatos de Conecta [Catálogo].

La matriz del catálogo de fuentes de conjuntos de datos de Conecta (ver Tabla 1) se completa según la información existente proporcionada por los metadatos obtenidos de la API. En caso de que falte algún dato o variable, se realiza un proceso de búsqueda manual.

06 Vinculación de las fuentes de conjuntos de datos a los ODS.

Para vincular el conjunto de datos a los ODS, se analizan las palabras clave definidas en el paso 1 con la pequeña descripción de los datos proporcionados por la matriz de metadatos de API. Un conjunto de datos está vinculado a un ODS o más si las palabras clave aparecen en la descripción.

07 Validación de vinculación de fuentes de conjuntos de datos.

A fin de validar la calidad de los anteriores pasos, se limpia la matriz diligenciada, al eliminar las filas vacías, corregir de forma manual los datos que no estaban vinculados con los ODS correspondientes y verificar que todos los vínculos asociados a los conjunto de datos se encuentren activos.

1. Conjunto de definiciones y protocolos que se utilizan para desarrollar e integrar el software de las aplicaciones. Esto permite que dos aplicaciones de software se comuniquen e interactúen a través de un conjunto de reglas.

CATÁLOGO DE FUENTES DE CONJUNTOS DE DATOS

El mapeo de datos se traduce en un catálogo de fuentes de conjuntos de datos compuesto por 15 variables que describen su fuente de origen, su relación con los ODS y sus atributos (desagregación, periodo de publicación, variables, formato). El catálogo se encuentra disponible para descarga en formato CSV o XLSX para el procesamiento de los usuarios en Conecta.

Tabla 1. Composición del catálogo de fuentes de conjuntos de datos – Conecta

VARIABLE	DESCRIPCIÓN
País	País al que hace referencia el conjunto de datos.
Subnacional	División política-administrativa a la que hace referencia el conjunto de datos según país de referencia. Por ejemplo, en el caso de Colombia, se identifican conjuntos de datos para la ciudad de Cali.
Objetivo de Desarrollo Sostenible	Nombre corto de los 17 Objetivos de Desarrollo Sostenible (ODS) asociados a cada conjunto de datos. Nota: Un conjunto de datos puede estar relacionado con más de un ODS.
Conjunto de datos	Nombre del conjunto de datos abreviado, identificando la temática principal.
Fuente	Nombre de la institución productora de datos.
Tipo de fuente	Categoría a la cual pertenece la institución productora de datos. Se identifican seis grupos de acuerdo con su actividad misional: academia, filantropía, organismos multilaterales, sector público, sector privado y sociedad civil.
Vínculo	Si el conjunto de datos se encuentra disponible en un sitio web, se incluye la URL correspondiente.

Objetivo del conjunto de datos	Descripción del propósito que busca el conjunto de datos identificado.
VARIABLES	VARIABLES que conforman el conjunto de datos.
Periodicidad	Periodicidad de publicación del conjunto de datos. Se identifican 10 periodos: decenal, quinquenal, bienal, anual, bianual, trimestral, bimestral, mensual, quincenal y semanal. Un conjunto de datos puede contar con más de un periodo de publicación.
Desagregación geográfica	Nivel de desagregación territorial que tiene disponible el conjunto de datos: regional, nacional, departamental, municipal.
Formato	Extensión o extensiones en que se encuentra disponible el conjunto de datos. Se identifican 26 clases de extensiones entre ellas CSV, xlsx, docx, pdf, HTML.
Métodos de recolección	Método a partir del cual se recolecta el conjunto de datos. Se identifican cinco tipos de métodos: censos, encuestas, mediciones ocasionales, múltiples fuentes y registros administrativos.
Fecha de última publicación	Fecha en que fue realizada la última publicación del conjunto de datos mapeado.
Fecha de consulta	Fecha en que el equipo técnico realiza el mapeo del conjunto de datos.

Fuente: elaboración propia

Los invitamos a consultar las diferentes herramientas del componente Conecta y a descargar el catálogo de fuentes de conjuntos de datos para explorar a detalle las diferentes variables que lo componen.

ACERCA DE DATAREPÚBLICA

DataRepública es un centro digital para la creación de conocimiento y narración basada en datos. El Centro de Pensamiento Estratégico Internacional (Cepei) desarrolló DataRepública en colaboración con Telefónica, para las organizaciones e individuos comprometidos e involucrados en el logro de los Objetivos de Desarrollo Sostenible (ODS) con enfoque en América Latina y el Caribe.

El Hub hace que el análisis de datos sea más accesible y revela las conexiones y relaciones entre los datos, las fuentes y el camino para cumplir con los ODS. Nuestros objetivos son:

1. Informar y fortalecer la toma de decisiones, el emprendimiento de datos y nuevas formas de comunicarse, ofreciendo datos oficiales calificados y no tradicionales.
2. Promover las asociaciones de múltiples partes interesadas y proporcionar las herramientas y el asesoramiento sobre los procesos necesarios para el intercambio creativo de conocimientos y la colaboración.

3. Desarrollar la capacidad de los actores del desarrollo sostenible para crear un sentido de apropiación al producir conocimiento.

DataRepública es una plataforma integral que incorpora cuatro componentes distintos e interconectados, cada uno de ellos con un programa independiente:

- **Conecta:** conecta fuentes de datos y rastrea sus relaciones con los ODS.
- **Aprende:** desarrolla las habilidades de los usuarios y fortalece sus capacidades para comunicar y transferir conocimientos.
- **Publica:** da vida a los datos con potentes visualizaciones.
- **Analiza:** permite a los actores del desarrollo transformar datos en conocimiento.



CONTACTO


Para mayor información referente a **Conecta** o cualquier componente de **DataRepública** por favor enviar un correo a m.vaca@cepei.org





Cepei

Knowledge from the Global South

..... www.Cepei.org

 [@infoCEPEI](https://twitter.com/infoCEPEI)

 [@infoCEPEI](https://www.facebook.com/infoCEPEI)

 [Info CEPEI](https://www.youtube.com/InfoCEPEI)

 [InfoCEPEI](https://www.linkedin.com/company/InfoCEPEI)

Contacto

Fredy Rodríguez

Coordinador del área de Datos

f.rodriguez@cepei.org