

MÓDULO 2

TRADUCIENDO AGENDAS GLOBALES A CONTEXTOS LOCALES



PERIODISMO DE DATOS
PARA EL DESARROLLO SOSTENIBLE

ENTENDIENDO EL MUNDO DE LOS DATOS

Luego de conocer el contexto de los escenarios de evaluación de la agenda del desarrollo sostenible y del significado de “no dejar a nadie atrás”, es tiempo de pasar a la parte más técnica: la comprensión de los datos. A continuación encontrarás desde el concepto de dato, hasta lo que entendemos hoy por datos abiertos y su importancia para el seguimiento de los avances en la Agenda 2030.

1. ¿QUÉ ES UN DATO?

UN DATO:

Es la representación cuantitativa o cualitativa de un evento u objeto en particular.

Los datos por sí mismos carecen de sentido si no existe la intención de interpretarlos o analizarlos, ya que en su mínima expresión son meramente un número, un texto o un indicador sin contexto que habilite su entendimiento.

Los datos pueden usar distintos formatos para representar la información que almacenan, es importante familiarizarse con ellos ya que entenderlos facilita la comprensión sobre su uso y posteriormente, procesamiento y análisis.

TIPO DE FORMATO	USOS
Numérico	Cuantifica objetos, personas, distancias, peso, entre otros. También se usa como la fuente de información para realizar operaciones matemáticas.
Fecha	Determina la fecha de inicio o fin de un suceso. Se puede representar de distintas maneras dependiendo de la información a la que refiere. Por ejemplo: 12/05/2012 o 12/May/2012, entre otras.
Hora	Muestra la hora en que un evento acontece. Se puede representar en formato de 24 horas o 12 horas. Ejemplo: 15:29 PM o 3:29 PM.
Booleano	Representa un valor binario o también conocido como lógico, en la mayoría de los casos para afirmar o negar una característica. Se representa normalmente por indicadores como SI/No o 0/1
Texto	Son campos abiertos en los que se permite almacenar texto libre. En casos aplicados, se usan para guardar nombres de personas, direcciones y descripciones de objetos, entre otros.

Actualmente, para las personas, empresas, gobiernos, entre otros, la recolección de datos representa la posibilidad de almacenar evidencia inherente a distintas actividades de las cuales somos parte día a día, por ejemplo una agenda de contactos o una biblioteca de música representa diversos tipos de datos asociados al registro de una persona o artista. Actualmente, recolectar y almacenar datos es sencillo mediante el uso de diversas herramientas tecnológicas que facilitan esta tarea. En su mayoría, el uso intensivo de sistemas como gestores de bases de datos permiten el almacenamiento de grandes volúmenes de datos provenientes de todo tipo fuentes.

2. ¿QUÉ SON LOS DATOS ABIERTOS?

Los Datos Abiertos o también conocidos como "Open Data" tienen un significado relativamente nuevo que se refiere a la disponibilidad de información de toda índole en formatos manejables, sin costo de adquisición o limitantes para su uso.

Características de los datos abiertos

Para que los datos puedan considerarse abiertos, deben cumplir por lo menos con dos características esenciales en cuanto a su formato y el medio o forma de acceso.

a. **Formato:** Los datos abiertos deben ser entregados en formatos compatibles con todo tipo de software; es decir que no se requiera de la adquisición o configuración de software alguno para poder usarlos. Esto es importante ya que una de las principales limitantes, de cara a los usuarios finales, es el software necesario para abrir o usar una base de datos específica. En algunos casos la limitante de formato implica la compra de licencias o la adquisición de

complementos para computadora que limita el uso de la información.

EJEMPLO:

Estamos buscando datos sobre escuelas públicas en el sitio web de la dependencia encargada de su publicación. El archivo debería contener información como nombre, ubicación e información del mobiliario disponible en las escuelas por ciudad específica. El archivo se encuentra en formato .xlsx (extensión asociada al programa hoja de cálculo Excel). Pero, para un número de usuarios que descargaron el archivo de la misma página no fue posible abrirlo, ya que no cuentan con la licencia de software necesaria. Si el archivo fuera presentado en un formato abierto, como puede ser .CSV (Valores separados por comas por sus siglas en inglés), cualquier usuario (con software licenciado o no), tendría la posibilidad de usar el mismo archivo y de conocer la información contenida en él.

b. **Medio o la forma de acceso.** Mientras que en algunas empresas, instancias de gobierno o centros de investigación es necesario pagar un costo por la información solicitada; los Datos Abiertos, apoyándose en los principios de la cultura "abierto", proponen que el acceso a la información debe tener la menor cantidad de restricciones posibles, esto con el fin de promover un **acceso público, universal, permanente y libre de costo** para el usuario final.

EJEMPLO:

En algunos países, diversas dependencias, mayoritariamente los órganos de estadística, generan un gran volumen de información, pero para poder disponer de ella es necesario pagar un costo o hacer largos trámites. Si se desea presentar un

documento bajo la característica de Datos Abiertos, debe entonces eliminarse el costo o pago de derechos por la información (salvo los casos en que se pague el costo de reproducción de la obra) y deben minimizarse o eliminarse los trámites necesarios para adquirir la información requerida por parte de los usuarios.

Ambas características de los Datos Abiertos buscan habilitar al usuario para que haga uso de la información y en ese camino no existan restricciones técnicas, legales o económicas que lo impidan.

OTRAS DEFINICIONES DE DATOS ABIERTOS:

- La organización **Open Knowledge**¹ define a los Datos Abiertos como: "datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen".
- El **Open Data Institute**² indica que los Datos Abiertos son "los datos que las

organizaciones, empresas e individuos han puesto a disposición para que cualquier persona pueda acceder a ellos, usarlos y compartirlos".

Una vez hechos públicos los Datos Abiertos y que cumplan con estas dos características, hay una serie de atributos que dan certeza respecto a la información que se publica y la forma en que puede ser usada. Estos atributos son:

1. Open Knowledge Foundation (2015). Disponible en okfn.org/opendata

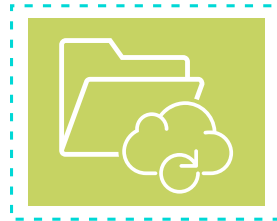
2. Open Data Institute (2015). Disponible en theodi.org/what-is-open-data

ATRIBUTOS DE LOS DATOS ABIERTOS

Que la **información** se encuentre **documentada**. Esto **facilita el entendimiento** de los tipos de datos que contiene, su origen, cuándo fueron actualizados, entre otros detalles. A todo esto también **se le conoce como Metadata**.



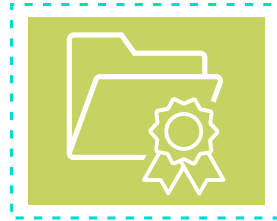
Que la **información** esté **actualizada**. Un conjunto de datos pierde valor si la información que contiene no está actualizada. **Sin importar que la información esté en formatos abiertos o sin costo**, los datos deben estar lo más actualizados posible para su uso.



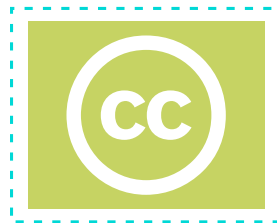
Que la **información** se muestre con el **máximo nivel granular**. Es decir **que se ofrezca el mayor nivel de detalle** que los datos puedan tener, de tal forma que el análisis de la información se enriquezca.



Que el **origen de la información** sea la **fuentes primaria**. Esto **garantiza la integridad de la información** no únicamente en el formato o la estructura de la misma, sino que también da **certeza de la calidad y veracidad** de los datos que contiene el archivo.



Que estén **libres de licencias que impidan su uso**, incluso que impidan su **adaptación o redistribución** para cualquier fin, por ejemplo, las licencias Creative Commons, licencia libre uso MX, Licencia Apache, entre otras, dan las garantías legales al usuario para utilizar la información.



Encuentra más información sobre atributos de los datos abiertos en opengovdata.org

En Latinoamérica existen actualmente diversos bancos de información que poco a poco están transitando a una cultura de publicación de conjuntos de datos basada en el concepto de Datos Abiertos. Estos

portales de datos presentan una variedad de fuentes de información y ejes temáticos los cuales se pueden aprovechar para diversos proyectos de investigación y análisis. Algunos ejemplos de estos portales son:

LISTADO DE PORTALES	
<p>México</p> <ul style="list-style-type: none">• Gobierno Nacional• INEGI• CDMX• LabCDMX <p>BID</p> <ul style="list-style-type: none">• Portal BID <p>Argentina</p> <ul style="list-style-type: none">• Gobierno• BAs• Bahía Blanca• PAMI <p>República Dominicana</p> <ul style="list-style-type: none">• Gobierno	<p>Colombia</p> <ul style="list-style-type: none">• Gobierno• Cámara de Comercio de Bogotá <p>El Salvador</p> <ul style="list-style-type: none">• Sociedad Civil <p>Costa Rica</p> <ul style="list-style-type: none">• Gobierno <p>Bolivia</p> <ul style="list-style-type: none">• Gobierno• Sociedad Civil <p>Paraguay</p> <ul style="list-style-type: none">• Gobierno• Funcionarios Gobierno

LAS CINCO ESTRELLAS DEL OPEN DATA

En la comunidad usuaria y consumidora de Datos Abiertos, un poco más ligada al área técnica, se comenzó a conversar sobre la calidad con la cual se presenta la información y bajo qué argumentos se podría considerar a un conjunto de datos dentro o no de la definición de Datos Abiertos. Esta conversación derivó en la creación de una escala que permitiera entender y catalogar a los conjuntos de

datos basados en sus características y así clarificar los límites entre cada nivel de apertura de información. Esta escala es conocida como la escala 5 estrellas de los Datos Abiertos.

La escala fue elaborada por Sir Tim Berners-Lee³ quien fue el inventor de la web. Las cinco estrellas de Lee nos pueden ayudar a entender las diferencias entre el mínimo nivel de apertura de información y

3. <http://www.w3.org/DesignIssues/LinkedData.html>

el más alto. A mayor número de estrellas, es más benéfico para el usuario final, ya que mejora la calidad del formato y facilita el procesamiento de los datos a gran escala.

A continuación una descripción de cada uno de los niveles (estrellas) de la escala:

LAS 5 ESTRELLAS DE LOS DATOS ABIERTOS



Entrega de información en **cualquier formato**, aunque no sea posible editar, desagregar o comparar.

Formatos y tipo de información

.PDF



Información en **formatos que requieren software especializado o de licencia** con costo para poderla utilizar.

Formatos y tipo de información

.DOC / .XLS / .KML



Disponible en la web, sin costo y en formato abierto con la información tabulada o estructurada que puede ser usada y analizada.

Formatos y tipo de información

.TXT Y ODT / .CSV / .XML
.DBF / .SHP



El formato de la información no es un problema, ya que en esta categoría **la información es almacenada bajo una *estructura RDF** (Resource Description Framework o marco de descripción de recursos por sus siglas en inglés) que permite conectar información bajo propiedades que la unen. De esta forma se pueden hacer **consultas y obtener respuestas inmediatas** desde el lugar que publica digitalmente la información **sin necesidad que el usuario realice tareas adicionales o manipule la información a través de otros programas.**

Formatos y tipo de información

.RDF / .XML



En este nivel **la información mantiene la estructura RDF** pero **su característica es que esta información se puede vincular con más fuentes de información a nivel global.** Esto implica que compartan la misma arquitectura. En este nivel diferentes bases de datos se conectan, veamos la versión más reciente de las diferentes bases de datos mundiales que están interconectadas publicada por la W3C
http://lod-cloud.net/versions/2014-08-30/lod-cloud_colored_1000px.png

Formatos y tipo de información

.RDF / .XML

Si está en México, Colombia o Costa Rica, pase por el ecosistema de datos de www.datapublica.org y explore un poco el panorama de la disponibilidad de datos para cada Objetivo de Desarrollo Sostenible (ODS) en estos países.