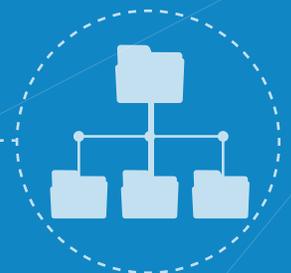


MANUAL DE DATOS ABIERTOS



Secretaría Nacional
de Ciencia y Tecnología





MANUAL DE DATOS ABIERTOS



DERECHOS DE AUTOR

Este documento fue creado por la Secretaría Nacional de Ciencia y Tecnología -SENACYT- y se reconoce su propiedad y autoría. Se puede reproducir gratuitamente en cualquier formato sin permiso expreso de la entidad con las siguientes condiciones.

- Puede ser distribuido total o parcialmente sin fines comerciales.
- El título de este documento debe ser incluido como referencia documental.
- La reproducción de este documento debe incluir la leyenda “Copiado/distribuido con permiso de la Secretaría Nacional de Ciencia y Tecnología. Todos los derechos reservados”

Si se desea copiar o distribuir el documento con otros propósitos, debe solicitar autorización expresa a las oficinas de la Secretaría Nacional de Ciencia y Tecnología.

CONTENIDO

DERECHOS DE AUTOR	4	BUENAS PRÁCTICAS Y CASOS DE ÉXITO	29
DE REPORTES ESTADÍSTICOS A DATOS ABIERTOS	8	BUENAS PRÁCTICAS	31
DATOS ABIERTOS	9	ESTANDARIZACIÓN DE DATOS	33
INTRODUCCIÓN	10	CONJUNTOS DE DATOS SUGERIDOS	34
Historia de los datos abiertos en Guatemala	11	A1. GLOSARIO	37
Política Nacional de Datos Abiertos	12	A2. TIPOS DE LICENCIAMIENTO	38
Ley de acceso a la información pública de oficio	13	A3. PROCESO DE ANONIMIZACIÓN	40
¿QUÉ SON DATOS ABIERTOS?	14	A4. FORMATO DE DATOS ABIERTOS	42
Principios que rigen los datos abiertos	14	A5. EJEMPLOS DE APLICACIÓN DE DATOS ABIERTOS	43
Características de los datos abiertos	14	A6. PLANTILLA PARA INVENTARIO DE DATOS	45
Nivel de apertura	15	A7. PLANTILLA PARA MATRIZ DE DATOS	46
¿POR QUÉ ABRIR MIS DATOS?	16	A8. PLANTILLA DE INFORMACIÓN DE PUBLICACIÓN	47
SELECCIÓN DEL EQUIPO DE TRABAJO	17		
¿QUIÉN SE HACE CARGO?	19		
Roles y perfiles	19		
Funciones principales	20		
ESTRATEGIA DE APERTURA	21		
PRESENTACIÓN DE LA ESTRATEGIA	23		
Identificar	23		
Obtener	23		
Procesar	23		
Publicar	23		
Actualizar	23		
ESTRUTURA DE DATOS	24		
ETAPAS DE APERTURA	25		
Antes	25		
Durante	26		
Después	27		

DE REPORTE ESTADÍSTICOS A DATOS ABIERTOS

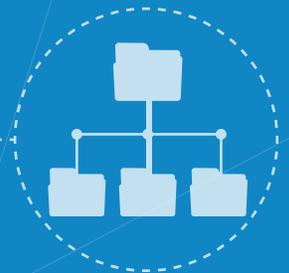
Los organismos públicos recogen, producen, reproducen y difunden datos para demostrar el cumplimiento de su función. Algunos de estos datos son publicados por diferentes vías, ya sea digital o impresa; en muchos casos no lo son. Integrar la publicación de los datos en formatos abiertos, abre la puerta a que los mismos sean reutilizados en nuevos proyectos, que al combinarse con otras fuentes de datos pueden generar nuevas soluciones en forma de aplicaciones desarrolladas tanto por el gobierno, por la sociedad civil, organizaciones públicas, privadas o académicas o ciudadanos en general.

Este manual brinda introducción al tema de datos abiertos. La misma reúne conceptos básicos de datos abiertos de gobierno, los pasos a seguir al momento de abrir los datos, ejemplos de reutilización de datos e información sobre metadato.



SECCIÓN 1

DATOS ABIERTOS



INTRODUCCIÓN

En la era de la innovación digital, se generan datos de toda índole. Estos datos en contadas ocasiones no son accesibles ya sea por la forma en la que se recolecta o porque tiene algún tipo de licenciamiento que no permite su publicación. En el despliegue de la economía digital, las tecnologías asociadas a ella están generando cambios trascendentales en la generación de valor mediante la innovación de servicios públicos.

Es por ello por lo que la relevancia de los datos en entidades del sector público se ha incrementado radicalmente. Por consiguiente, los datos abiertos de gobierno se están convirtiendo en una fuente de generación de nuevos servicios y aplicaciones.

Los Datos Abiertos, mediante su utilización, reutilización y redistribución, tienen el potencial para promover el desarrollo económico e innovación, mejorar y crear los servicios prestados por el gobierno, incrementar la transparencia, la rendición de cuentas y la eficiencia en las operaciones públicas, así como fomentar la creación de valor social. Todo esto con el fin de impulsar las oportunidades de emprendimiento e innovación en beneficio de la ciudadanía y mejorar el diseño de las políticas públicas.

Este manual constituye un instrumento de apoyo al proceso operativo de apertura de los datos gubernamentales y con este propósito es presentada considerando el ciclo de vida de los datos abiertos.

HISTORIA DE LOS DATOS ABIERTOS EN GUATEMALA

No se sabe a ciencia cierta en qué momento se inició el concepto de datos abiertos en el mundo, se tienen registros de ciencia abierta desde la década de los 50's como un medio de colaboración a la producción científica. Los datos abiertos en gobierno son relativamente nuevos en comparación a la ciencia, fue en el 2009 cuando el primer ministro Gordon Brown presentó el portal de datos abiertos del Reino Unido (data.gov.uk) como una iniciativa para generar oportunidades para las empresas, aumentar la transparencia y otorgar poder a los y las consumidoras.

En el año 2011, el 27 de julio específicamente, Guatemala se adhiere a la Alianza para el Gobierno Abierto -AGA-, Open Government Partnership -OGP- en inglés, posteriormente en 2012 ratificó su adhesión, habiendo implementado cuatro Planes de Acción Nacional de Gobierno Abierto al 2020; siendo dichos planes bianuales. El tercer Plan de Acción Nacional de Gobierno Abierto 2016-2018, fue creado bajo la idea fundamental de suma de voluntades y unificación de esfuerzos de todas las entidades públicas y organizaciones de sociedad civil que conforman la Mesta Técnica de Gobierno Abierto. En ese sentido, la Iniciativa de Gobierno Abierto se convierte en un mecanismo propicio para la prevención de

la corrupción y transformación de la gestión pública, mediante un espacio de discusión y diálogo de representantes de instituciones públicas y organizaciones de sociedad civil en la construcción de los 22 compromisos que forman parte del Plan de Acción Nacional de Gobierno Abierto 2016-2018, basados en 5 ejes de trabajo: a) Acceso a la Información Pública y Archivos Institucionales, b) Innovación Tecnológica, c) Participación Ciudadana, d) Rendición de Cuentas; y, e) Transparencia Fiscal. El eje de Innovación Tecnológica estuvo compuesto por 6 compromisos, siendo el compromiso no. 5 denominado "Creación e implementación del portal único de datos abiertos, creación y aprobación de la política nacional de datos abiertos", del cual el desarrollo e implementación del Portal de Datos Abiertos estuvo a cargo de la Secretaría Nacional de Ciencia y Tecnología -SENACYT- y la Política Nacional de Datos Abiertos a cargo del Ministerio de Gobernación -MINGOB-.

Es así como el 12 de junio del 2018 se lanza el Portal de Datos Abiertos bajo el dominio <http://www.datos.gob.gt/> teniendo seis instituciones como piloto, a su vez, la Política Nacional de Datos Abiertos fue lanzada el 26 de noviembre del 2018.

POLÍTICA NACIONAL DE DATOS ABIERTOS

La Política Nacional de Datos Abiertos, al igual que el Portal de Datos Abiertos, tiene sus orígenes en la Alianza para el Gobierno Abierto en Guatemala, durante su Tercer Plan de Acción, teniendo como resultado la publicación del Acuerdo Gubernativo 199-2018 que entró en vigor el 6 de diciembre del 2018.

Dicha política tiene como principios rectores los indicados en la Ley de Acceso a la Información Pública, Decreto número 57-2008 del Congreso de la República de Guatemala.

1. Máxima publicidad;
2. Transparencia en el manejo y ejecución de los recursos públicos y actos de la administración pública;
3. Gratuidad en el acceso a la información pública;
4. Sencillez y celeridad de procedimiento.

Para la implementación de la Política, esta establece cinco ejes:

- Eje I: Apertura de datos gubernamentales

- Eje II: Portal Nacional de Datos Abiertos
- Eje III: Promoción de la Innovación Pública y Privada
- Eje IV: Mejora de los índices internacionales de transparencia y adhesión a compromisos internacionales
- Eje V: Cultura de transparencia, rendición de cuentas y auditoría social

Teniendo como alcance las instituciones del Estado pertenecientes al Organismo Ejecutivo. La Política Nacional de Datos Abiertos define al Portal de datos abiertos como una plataforma digital (sitio web) que servirá para almacenar, compartir, conectar y visualizar conjuntos de datos. Es el punto de acceso único de los datos abiertos del Organismo Ejecutivo de la República de Guatemala. Este portal es de acceso libre para consulta y utilización de los datos.

Se puede acceder al portal de datos abiertos a través de <http://www.datos.gob.gt/>

LEY DE ACCESO A LA INFORMACIÓN PÚBLICA DE OFICIO

Conocida como LAIP (ley de acceso a la información pública) fue establecida bajo el decreto número 57-2008 tiene como objeto garantizar a toda persona interesada, sin discriminación alguna, el derecho a solicitar y a tener acceso a la información pública en posesión de las autoridades y sujetos obligados, así como garantizar la transparencia de la administración pública.

Establece dos mecanismos principales

- Información pública de oficio: definida en el artículo 10 y 11 de la misma ley, establece la información que deben publicar por defecto todos los sujetos obligados definidos en el artículo 6 de la LAIP. Esta información es actualizada mensualmente y no necesita solicitud para hacerlo.
- Unidad de información pública: es un mecanismo a través del cual cualquier ciudadano guatemalteco puede solicitar información específica, teniendo la institución 10 días máximo para dar respuesta al mismo.

La misma ley establece como órgano garante a la Procuraduría de los Derechos Humanos, ya que el acceso a la información pública es un derecho humano fundamental previsto en la Constitución Política de la República de Guatemala.

¿QUÉ SON DATOS ABIERTOS?

La definición de la Carta Internacional de Datos Abiertos establece como datos abiertos:

“Datos digitales que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar”.

A esta definición se le agregaría que los datos no vulneren la información privada de una persona, entidad o que representen peligro para una nación.

De esta cuenta, en Guatemala se definen datos abiertos a aquellos producidos por entidades públicas, empresas, academia o ciudadanos que en su forma básica resulta accesible para cualquier persona desde internet, sin restricción salvo las excepciones previstas en la normativa vigente, teniendo como características principales estar disponibles en formato que pueda ser fácilmente procesados, sin costos para su obtención, sin limitaciones para su reutilización y redistribución y que permita crear servicios derivados de los mismos.

Debemos tener claro que datos públicos de oficio no es lo mismo que datos abiertos, debido a la forma en que son presentados, actualizados y utilizados. Este documento ofrece una estrategia para que los datos públicos se conviertan en datos abiertos.

PRINCIPIOS QUE RIGEN LOS DATOS ABIERTOS

La Carta Internacional de Datos Abiertos¹ establece seis principios básicos que permiten sentar las bases para el acceso a los datos para su publicación y uso.

1. Abiertos por defecto :
2. Oportunos y exhaustivos
3. Accesibles y utilizables
4. Comparables e interoperables

5. Para mejorar la Gobernanza y la Participación Ciudadana
6. Para el Desarrollo Incluyente y la Innovación

CARACTERÍSTICAS DE LOS DATOS ABIERTOS

De igual manera la Opengovdata publicó ocho características que deben ser consideradas al momento de publicar datos. Estas características no indican qué datos abrir, sino que especifican las condiciones para que los datos públicos sean considerados abiertos. A continuación, se listan estas características:

1. **Completos:** Todos los datos públicos se ponen a disposición. Los datos públicos son datos que no están sujetos a limitaciones de privacidad, seguridad o privilegios válidos.
2. **Primarios:** Los datos se recogen en la fuente, con el más alto nivel de granularidad, no hay agregaciones o modificaciones.
3. **Oportunos:** Los datos se pondrán a disposición tan pronto como sea necesario para preservar el valor de los datos.
4. **Accesibles:** Los datos están disponibles para la gama más amplia de usuarios con la más amplia gama de propósitos.
5. **Procesables por máquina:** Los datos se estructuran razonablemente para permitir el procesamiento automatizado.
6. **No discriminatorios:** Los datos están disponibles a cualquier persona, sin necesidad de registro.
7. **No propietarios:** Los datos están disponibles en un formato sobre el

cual ninguna entidad tiene el control exclusivo.

- Libres de licencia:** Los datos no están sujetos a ningún derecho de autor, patentes, marcas o regulación. Se permiten restricciones razonables de privacidad y seguridad.

NIVEL DE APERTURA

Tim Berners-Lee propuso una categorización del grado de apertura de los datos basado en estrellas, respecto a qué tan abiertos y usables son los datos que puede ofrecer una institución.

1 estrella: Los datos deben estar disponibles en la web en cualquier formato y bajo una licencia abierta, para que sean considerados como datos abiertos. Entonces cualquier archivo disponible en sitios Web en formato PDF, DOC, ODT, TXT, entre otros; y que esté bajo una licencia abierta será clasificado como dato abierto de 1 estrella.

2 estrellas: Los datos deben estar estructurados y en un formato que pueda ser interpretado o procesado por máquinas. Para que un dato sea considerado de 2 estrellas debe cumplir todo lo que implica estar en 1 estrella, es decir, publicado y bajo licencia abierta, agregado los datos deben estar estructurados y en un formato que pueda ser procesado

por una máquina. Un ejemplo son los datos organizados en formato .XLS o los presentados en una tabla.

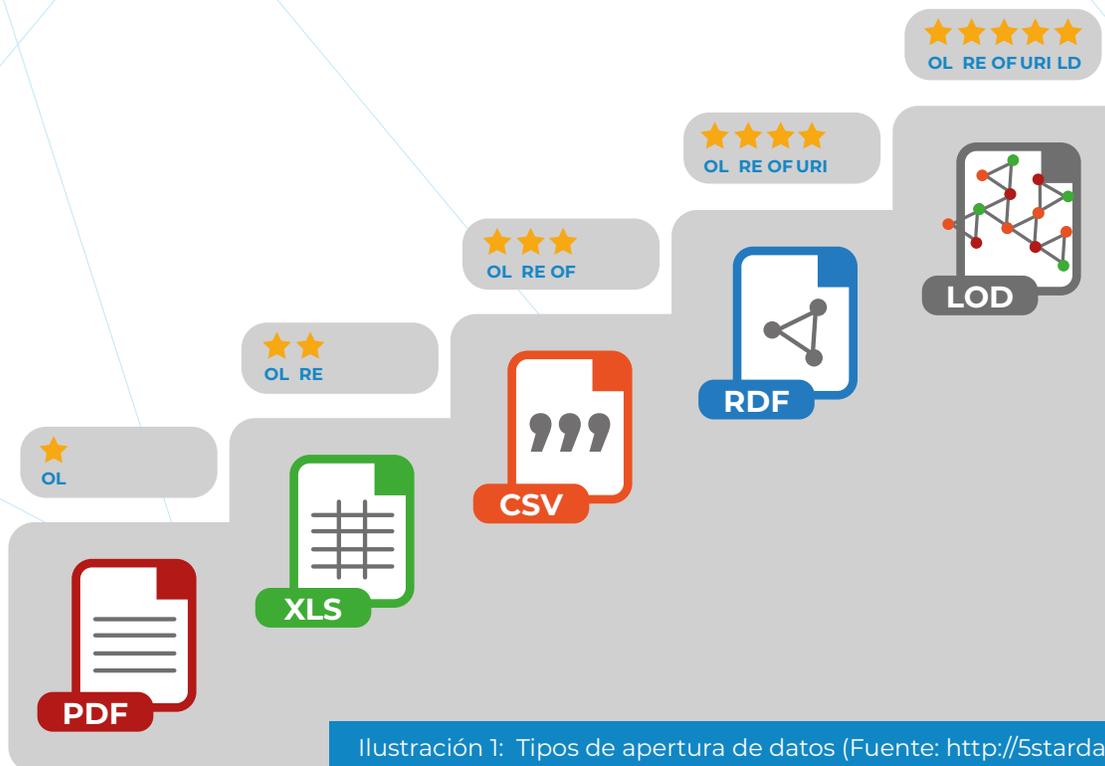
3 estrellas: En este nivel, el dato debe cumplir los niveles 1 y 2 y el formato debe ser no propietario, por ejemplo, datos en formato .CSV o .XML son considerados como 3 estrellas.

4 estrellas: Para que un dato sea considerado 4 estrellas debe cumplir lo establecido en los niveles anteriores y utilizar estándares recomendados por W3C, tales como RDF o SPARQL.

5 estrellas: Este es el nivel máximo de alcance, se considera que un dato es de 5 estrellas cuando cumple los 4 niveles anteriores y su información está enlazada entre sí, por ejemplo, siguiendo los principios de Linked Data y Web semántica.

No debemos confundir formato abierto con formato no propietario dado que el primero puede contener formatos propietarios hasta el nivel 2. A partir del nivel tres los datos deben ser no propietarios, esto es, no necesitan ninguna licencia para uso y manipulación.

En Guatemala, la cultura de apertura de datos está comenzando, por lo que se aspira que en las primeras etapas del proceso los datos abiertos alcancen niveles 1 al 3.



¿POR QUÉ ABRIR MIS DATOS?

Según la Open Data Handbook, “Datos Abiertos”, especialmente “datos abiertos de gobierno”, es un gran recurso, aunque todavía no ha sido explotado del todo. Muchos individuos y organizaciones reúnen una gran variedad de diferentes tipos de datos para llevar a cabo sus tareas. Pero ¿por qué esto es de interés para el ciudadano?, y sobre todo ¿por qué es importante para la institución abrir sus datos?

Existen muchas áreas donde se puede esperar que los datos sean valiosos y donde los mismos pueden ser utilizados para fomentar la innovación y emprendimiento a través de las tecnologías de información, de igual manera, los datos permiten proyectar a futuro y conocer dónde y cómo será creado el valor de estos.

Entre algunas de las acciones donde el gobierno abierto está creando valor se pueden mencionar:

- Permitir generar mayor transparencia y rendición de cuentas de parte de las instituciones de gobierno.
- Vincular al ciudadano en el proceso de fiscalización y poder crear mejoras de servicios públicos de manera alternativa.

- Crear una ‘comunidad’ de usuarios/ciudadanos que puedan intercambiar ideas y co-crear soluciones en base a datos abiertos (plataformas informativas, análisis de datos, comparativas, historias, etc.).
- Mejorar los procesos de participación ciudadana.

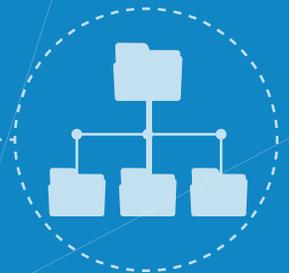
Se pueden mencionar ejemplos claros a nivel internacional donde el uso de datos abiertos a permitido grandes avances no solo en el tema de transparencia, sino en el sector económico, toma de decisiones, mejoramiento de servicios y vivienda, etc., tales como “¿where does my money go?” de Reino Unido, o Mapumental y Mapnificent, ambas aplicaciones enfocadas en mejorar la vivienda y el trabajo optimizando el tiempo de traslado entre ellos.

Este potencial sin explotar puede ser aprovechado si transformamos los datos públicos de Gobierno en datos abiertos. Esto solamente pasará, sin embargo, si se trata de datos verdaderamente abiertos, es decir si no hay restricciones (legales, financieras o tecnológicas) para su reutilización por terceros.



SECCIÓN 2

SELECCIÓN DEL EQUIPO DE TRABAJO



¿QUIÉN SE HACE CARGO?

Antes de iniciar con la apertura de los datos se debe tomar en cuenta quién lo hará, la conformación de un equipo de trabajo dinámico es fundamental para el desarrollo y avance de la apertura de datos. A continuación, se identifican los roles y perfiles para tener en cuenta para formar los equipos a nivel institucional, dependiendo de los objetivos, recursos y posibilidades de esta, algunos de los roles pueden concentrarse o dividirse en una o varias personas.

ROLES Y PERFILES

- **Coordinación gerencial:** Puede estar integrado por altos mandos que respalden y direccionen las acciones de una política de datos abiertos. Deben tener claro el valor de un proceso de apertura de datos, así como tener conocimientos de cómo utilizar esto en la gestión administrativa de la institución para alinearlos a los planes estratégicos y toma de decisión.
- **Coordinación operativa:** Tiene a su cargo la ejecución a nivel operativo de la Política de Datos Abiertos, así como la gestión de la estrategia de apertura de datos y monitoreo del material publicado en el portal de datos abiertos, es quien coordina a todo el equipo de trabajo y las acciones necesarias para fomentar la cultura de apertura de datos para validar el valor y calidad de estos.
- **Perfil técnico:** Tiene a su cargo el acompañamiento técnico para el equipo de trabajo enfocado en estándares de datos, conocimientos de programación, y gestión de la tecnología en el sector público. Su principal función estará enfocada a la transformación de datos crudos a formato abierto, validando que estos respeten las características que los conforman.
- **Equipo de soporte:** Este perfil debe tener conocimientos de la dinámica de publicación de información pública en diferentes medios. Estará enfocado en la publicación y monitoreo de los datos aperturados. De igual manera, debe tener claro el valor de un proceso de apertura y las acciones que conlleva.
- **Perfil analista:** La función de este perfil es hacer uso y explotación de los datos para facilitar y apoyar el uso de los datos abiertos por parte interna y externa a su institución, de igual manera validar la calidad de estos. Quien tome este rol necesita tener conocimiento en el manejo de herramientas de visualización de datos y una visión comunicacional sobre los productos a desarrollar.

FUNCIONES PRINCIPALES

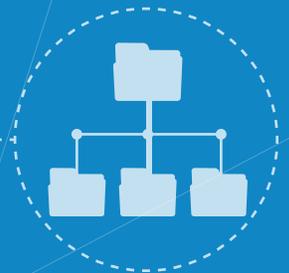
Existen múltiples funciones y acciones a realizar para la implementación de un proceso de apertura de datos institucional. Dependiendo de la naturaleza y tamaño de una organización, se necesita más o menos personal y coordinación. A continuación, se expone un listado mínimo de acciones que todo equipo de trabajo debe realizar para aperturar datos en una institución.

1. **Acuerdo para la apertura de datos:** Como paso inicial se sugiere la implementación de un acuerdo a nivel institucional para la definición, gestión, roles, alcances, nombramientos y otros lineamientos que se consideren oportunos en relación con la apertura por defecto. Esta función usualmente es realizada por la coordinación general y operativa.
2. **Gestión para la obtención y actualización de datos:** Partiendo de un inventario inicial de datos, se debe definir un criterio de apertura y actualización. Cada institución define sus prioridades de apertura, tiempos de actualización y generación de datos en formato abierto, durante esta acción se define un calendario de publicación y actualización. Esta función usualmente está a cargo de la coordinación gerencial y operativa, con el apoyo del equipo de soporte.
3. **Transformación de datos:** Cuando se define una lista priorizada o un calendario de apertura los datos son obtenidos en formato crudo, durante esta actividad, los datos pasan por un proceso de transformación para ser generados en formato abierto, en función de las directrices de la coordinación operativa. Esta acción es realizada por el perfil técnico.
4. **Publicación de datos:** Las personas a cargo de esta actividad publicarán los datos en formato abierto que el perfil técnico les brinde según el calendario establecido. Esta acción es realizada por el equipo de soporte.
5. **Uso y monitoreo:** Tiene a su cargo promover y monitorear los datos aperturados a través del uso y explotación de estos para público interno y externo a la institución. Esta acción es realizada por el perfil analista.



SECCIÓN 3

ESTRATEGIA DE APERTURA



PRESENTACIÓN DE LA ESTRATEGIA

¿Qué datos voy a abrir? Es la pregunta que más preocupa a los gestores de datos. Este manual no ofrece un listado de los conjuntos de datos a aperturar obligatorios, sino que presenta una estrategia para identificar, validar y priorizar los datos que causen más impacto al ciudadano y valor a la sociedad mediante el reúso de estos en emprendimientos y aplicaciones.

Si bien se comprende que no hay una forma correcta de hacerlo, ya que cada organización tiene su propia dinámica y naturaleza, se propone una estrategia que comprende cinco grandes pasos ejecutados en un antes, durante y después, los cuales se exponen a continuación.

IDENTIFICAR

A través de un inventario inicial se debe identificar todos los datos que la institución genere, definiendo el formato en el que se encuentran, para posteriormente priorizarlos según criterios que se consideren oportunos. Esta fase permite tener una visión general del estado de la institución en relación con el nivel de apertura que posee, de igual manera permite detectar actores claves tales como direcciones o departamentos involucrados.

La identificación oportuna de datos permitirá a la institución coordinar acciones y generar una programación efectiva para la apertura de datos a través de una calendarización definida.

OBTENER

Basados en una calendarización de apertura y publicación, el equipo de trabajo podrá gestionar la obtención de los datos a los departamentos o direcciones involucradas en el formato establecido. El nivel de dificultad para la obtención de datos depende de la naturaleza de la institución. El objetivo de esta fase es transportar los datos definidos de forma teórica a la siguiente fase. Se entiende que la obtención de datos puede estar en diferentes formatos, sea físico o digital.

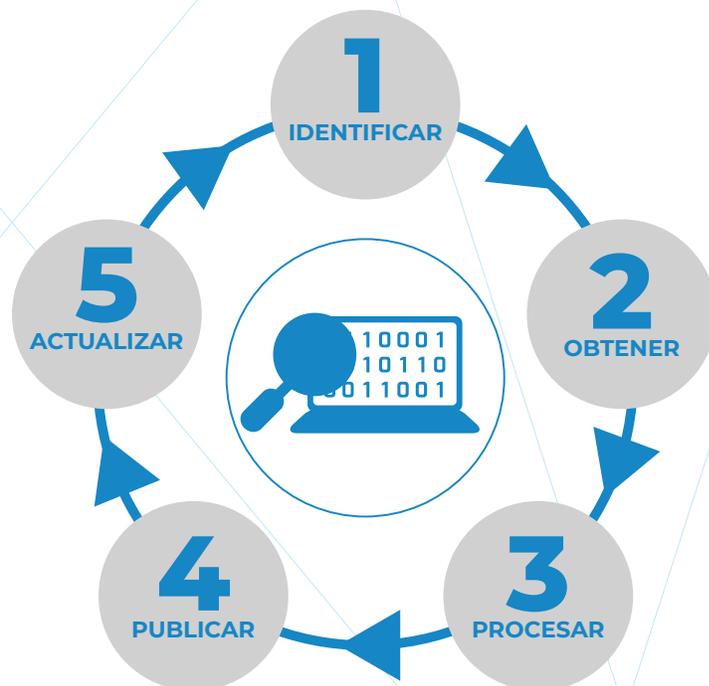


Ilustración 2: Ciclo de vida del dato

PROCESAR

Esta fase tiene como principal objetivo procesar los datos de entrada y transformarlos en datos en formato abierto. La acción de procesar datos puede ir desde la digitación de estos al provenir de una fuente física, para posteriormente pasar un proceso de transformación, que puede incluir tabulación, limpieza y anonimización de los mismos.

PUBLICAR

Tiene como objetivo la publicación de datos en formato abierto en el portal de Datos Abiertos de Guatemala, según la información que este solicite en la temporalidad definida por la coordinación del equipo de trabajo.

ACTUALIZAR

Esta fase del proceso tiene como objetivo actualizar periódicamente los conjuntos de datos institucionales. Se efectúa según la calendarización definida por la Coordinación.

ESTRUTURA DE DATOS

Previo a ejecutar la estrategia de apertura de datos es importante conocer la estructura de estos. Al hablar de apertura se debe tener presente los conceptos a continuación:

Conjunto: Almacena datos pertenecientes a una misma temática, en diferentes formatos y temporalidades. Un conjunto en sí no almacena datos en específico, almacena archivos.

Recurso: Se define como cualquier archivo que contiene datos útiles. Un conjunto está formado por múltiples recursos.

Campo: Es la unidad más pequeña de un conjunto, establece una estructura dentro de un recurso.

Metadato: Información de los datos. Es una descripción de cada campo para su mejor comprensión.

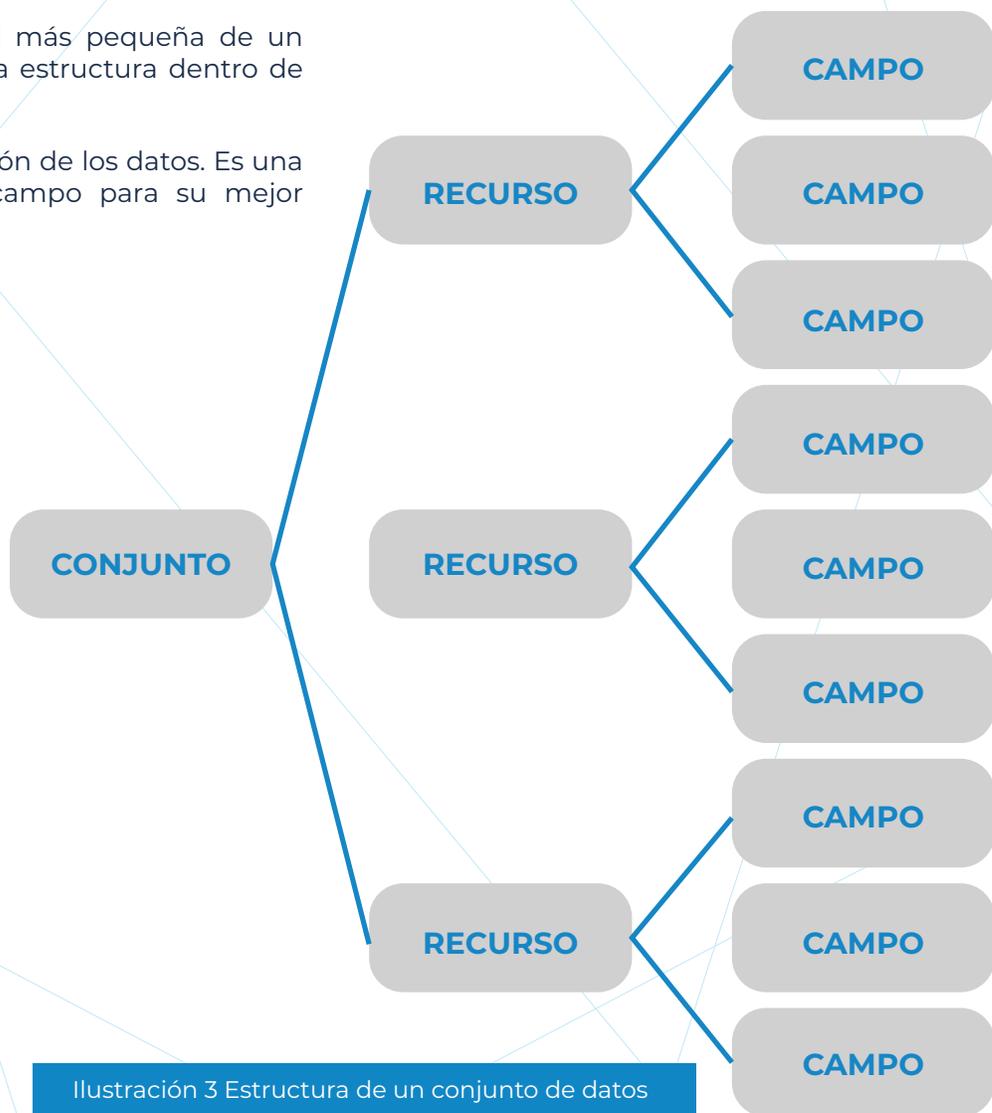


Ilustración 3 Estructura de un conjunto de datos

ETAPAS DE APERTURA



Antes

- Cree un inventario de datos
- Valide los datos
- Priorice la lista de datos
- Defina un calendario de publicación y actualización



Durante

- Prepare los datos en formato abierto
- Anonimice los datos críticos
- Defina la matriz de metadatos
- Defina qué licencia utilizará



Después

- Monitoree la actividad de sus datos
- Actualice sus datos
- Reutilice los datos publicados

ANTES

IDENTIFIQUE EL POSIBLE CONJUNTO DE DATOS A ABRIR:

Cuando hablamos de datos públicos podemos encontrarnos infinidad de estos. El primer paso para poder abrir dichos datos es identificarlos, debemos definir una lista de datos que producimos, esto se conoce como un inventario de datos, puede ver un ejemplo de este en el Anexo A6.

¿CÓMO DEFINO MI LISTA?

A continuación, mostramos una lista de criterios que pueden ser utilizados como referencia para identificar y definir los posibles datos a aperturar.

NORMATIVA

¿Existe alguna normativa que me pida datos? En Guatemala existe la Ley de Acceso a la Información Pública, en la cual toda institución del Estado está obligada por ley a publicar información establecida en sus artículos. Un paso inicial podría ser lo que la ley me exige. Son datos que ya se tienen disponibles.

IMPORTANCIA PARA LA SOCIEDAD

¿Estos datos son importante para la sociedad?

Tabla 3 Estrategia de apertura de datos

Debemos decidir si los datos que queremos publicar causarán impacto y son importantes para la sociedad misma, teniendo en cuenta el posible impacto y utilidad de los datos para los ciudadanos.

IMPORTANCIA PARA LA ECONOMÍA

¿Los datos que se abrirán generan oportunidades? Debemos valorar el potencial de los datos para generar nuevas líneas de investigación o negocios.

ACTUALIDAD DE LOS DATOS

¿Son datos actuales? Al momento de hacer el listado de datos potenciales debemos evaluar qué tan actuales son, sin despreciar los datos históricos que también pueden tener gran valor. Tome en cuenta que muchos datos públicos ya están publicados, pero no en formato abierto.

CANTIDAD DE DATOS.

¿Cuántos datos poseo? Recuerde, entre mayor sea el volumen más valor añadido se podría aportar con el tratamiento posterior de los datos.

NIVEL DE DETALLE

¿Mis datos están completos? Entre más detallados sean los datos, mayores serán también las posibilidades de explotación.

VALIDACIÓN Y PRIORIZACIÓN DE LOS DATOS

Una vez establecida la lista de datos a aperturar, debemos validar si esto es viable y si los mismos se encuentran disponibles, para ello este manual recomienda utilizar los siguientes criterios de selección:

CRITERIO

¿Qué es más importante abrir para mi institución? ¿Cuál necesita más procesamiento? ¿Tienen prioridad para mi institución estos datos? ¿Es una solicitud recurrente en la Oficina de Información Pública? Cada institución es un universo, y por ello se mueve a un ritmo diferente, el criterio bajo el cual un conjunto de datos tenga más peso que otro depende de la alta dirección. Puede ser enfocado en la naturaleza de la institución que apertura los datos, o en el recurso humano y/o tecnológico disponible.

DISPONIBILIDAD

¿Están sus datos en digital o en físico? ¿Cuenta con los datos a la mano o debe empezar a generarlos? Son preguntas que puede hacerse antes de escoger qué conjunto de datos abrir. Recuerde que al inicio es más sencillo trabajar con lo que se tiene que crear las cosas de cero.

RESTRICCIONES

¿Estos datos tienen derechos de autor? ¿Son de seguridad nacional? ¿Contiene información clasificada? Esto debe ser respondido antes de aperturar los datos. Puede encontrar tres posibles escenarios

- I. Los datos contienen información clasificada: Esto es información que pone en riesgo la intimidad, la salud, la seguridad o la vida de las personas, afecta los secretos comerciales, industriales o profesionales. En resumen, todos aquellos datos cuyo acceso libre puede causar daño a los derechos de personas naturales y jurídicas.
- II. Los datos contienen información reservada: Esto es información de defensa y seguridad nacional y pública, datos de relaciones internacionales, información de procesos judiciales, infantes y adolescentes, datos que afecten la estabilidad económica, financiera del país o que perjudique de alguna manera

la salud pública.

- III. Los datos contienen derecho de autor: Esto es información de patentes, información generada para usos comerciales, registro de propiedad, etc.

En el caso de las numerales I y II los datos pueden ser anonimizados (ver Anexo A3) y de esta manera eliminar cualquier vulnerabilidad que puedan provocar. En el caso de la numeral III los datos no pueden ser abiertos y deben ser descartados.

Por último, pero no menos importante, debe decidir el orden de prioridad de cada conjunto de datos con base a la capacidad tecnológica de su institución, así como de otros criterios que considere indicados según la naturaleza de su institución. Puede encontrar una plantilla para realizar una matriz de datos en el Anexo A7.

CALENDARIO DE PUBLICACIÓN Y ACTUALIZACIÓN

Para garantizar un proceso constante de apertura continua y actualización de datos se recomienda la programación anual de apertura, estableciendo, en función del nivel de prioridad, el orden de publicación. Para tal efecto se debe considerar el tiempo de procesamiento de datos previo a su publicación.

DURANTE

Una vez pasamos la fase de identificación, validación y priorización de los datos procedemos a crear el conjunto de datos que vamos a publicar. El primer paso es preparar los datos.

PREPARAR DEL CONJUNTO DE DATOS

Preparar los datos puede ser la fase más difícil que enfrente, todo depende de cómo se tienen almacenados los datos que quieren aperturar, o la legislación y/o tamaño de su institución. Esto también se conoce como obtención de datos. Veamos algunos escenarios:

- I. Si los datos se encuentran en formato físico el primer paso es digitalizar, para ello debe establecer la estructura que tendrán, así como el lugar donde los almacenará, esto puede ser un archivo de texto, una base de datos u otra estructura.
- II. Si sus datos se encuentran en formato

digital, verifique qué tipo de formato es y establezca su estrategia para convertirlos en formato abierto.

Es importante que cree un mapa de generación de datos donde indique la fuente (URL, base de datos, persona generadora). Esto facilitará la actualización de estos en las siguientes iteraciones.

Recuerde que la estructura de los datos es importante, la transformación de datos desde formatos no-textuales a formatos que permitan su manipulación y conexión con otros datos debe ser tratada con propiedad. Debe considerar que no importa la estructura que escoja, estos deben ser legibles para máquinas, que permita su entendimiento y uso. Esto le da más posibilidad al conjunto de datos de ser reutilizado.

DOCUMENTAR LOS DATOS

Una vez que se obtienen los datos estos pueden ser documentados, para ello se recomienda el uso de una matriz de datos (Ver Anexo A7), que permita definir cada recurso y sus respectivos campos que conforma el conjunto de datos completo, así como las acciones que se harán sobre ellos. Como se indicó en secciones anteriores, no todos los datos son aperturables por diferentes razones que comprenden desde la privacidad del ciudadano hasta la legislación que los rige. La matriz de datos será la guía para el equipo técnico al momento de transformar los datos crudos a datos abiertos, de igual manera con base a la matriz de datos se generará el metadato de cada recurso. En esta plantilla se definen los datos mínimos necesarios, puede agregar metadatos extras que considere oportunos. Recuerde que entre más información brinde de sus datos tendrá más oportunidad que los usuarios entiendan el contenido y las características de sus datos, así como permitirá identificar, integrar y/o interoperar con otras fuentes de datos.

ELIMINAR TODOS LOS DATOS VULNERABLES.

El siguiente paso en el proceso de apertura es eliminar datos vulnerables que fueron definidos en la documentación del conjunto. Se entiende como dato vulnerable los datos privados, clasificados y reservados. Si no está seguro acerca de qué datos se consideran clasificados o reservados, lo invitamos a visitar el Anexo A3 en donde encontrará la clasificación

y el proceso de anonimización de los datos.

GENERAR DATOS EN FORMATO ABIERTO.

A continuación, basados en la matriz de datos, se debe generar cada recurso (archivo) en formato abierto. Para ello se deja a discreción la técnica que considere apropiada, con base en los recursos informáticos que su institución posea. El resultado de este paso debe generar un conjunto de datos con recursos asociados que serán publicados en el Portal de Datos de Guatemala. De igual manera, se recomienda generar la documentación que dicho portal solicita, para ello puede utilizar de referencia la matriz de información de publicación en el Anexo A8. Entre los datos que solicita se encuentra el tipo de licencia. Para conocer los tipos de licencia que el portal único de datos abiertos de Guatemala le provee, consulte el Anexo A2.

PUBLICAR LOS DATOS

El último paso que necesita para aperturar sus datos es publicarlos, puede realizarlo siguiendo los pasos indicados en la Guía de usuario del portal único de datos, en donde encontrará toda la información necesaria, desde cómo crear su cuenta hasta cómo publicar, modificar y mantener los conjuntos de datos que publique.

DESPUÉS

Una vez publicados los datos el objetivo es que estos sean utilizados, para ello se hacen las siguientes recomendaciones.

ESTRATEGIA DE ACTUALIZACIÓN

La carga inicial de datos es solamente el primer paso en la apertura de estos. La incorporación de nuevos conjuntos de datos debe ser una constante a partir de ese momento y especialmente la actualización frecuente de los conjuntos que ya se han publicado.

Como la apertura de datos es un ciclo, se debe volver a la fase de identificación y basados en el inventario de datos, definir una estrategia de actualización organizada, de tal manera puedan abrir nuevos conjuntos y actualizar en la temporalidad establecida los conjuntos ya publicados.

FOMENTA LA OFERTA Y DEMANDA

Para que sus datos tengan mayor impacto, debe tener presente quienes serán sus usuarios, tanto internos como externos. La campaña de comunicación al exterior de la entidad es importante, pero también debe considerar empoderar a sus empleados acerca de los datos que está aperturando. De esta manera generará una cultura de datos abiertos que le facilitará en el futuro la apertura de nuevos conjuntos de datos.

APLICAR ESTRATEGIA DE REÚSO DE LOS DATOS

En esta parte debe establecer algunas actividades para que sus datos sean utilizados. El primer paso es darlos a conocer.

DAR A CONOCER LOS DATOS

Puede utilizar diferentes estrategias de comunicación, tanto internas como externas a la institución, con el fin de sensibilizar sobre el uso de datos abiertos y sus beneficios, de igual manera fomentar la transparencia a través de la mejora de servicios, periodismo de datos, desarrollo de aplicaciones, etc. Algunas actividades podrían ser:

- Publicaciones en redes sociales.
- Mensajes internos vía intranet o e-mailing
- Reuniones internas
- Campañas informativas con medios de comunicación

IDENTIFIQUE A SUS USUARIOS

El tener identificado quién utiliza sus datos le permite crear estrategias directas con ellos para involucrarlos en el uso de estos y así generar soluciones de mayor impacto.

Algunas actividades recomendadas:

- Desarrollar procesos de co-creación de herramientas tecnológicas y/o aplicaciones para impulsar el uso de los datos, participando desarrolladores y usuarios.
- Realizar talleres o encuentros con usuarios o grupos de interés para generar propuestas de usos creativos con datos abiertos.

- Generar planes o estrategias de incentivos para promover la participación de los usuarios en actividades que impulsen la identificación y uso de los datos abiertos de la institución.

MONITOREAR EL USO DE LOS DATOS

El ciclo de vida de los datos no termina cuando los mismos son publicados. Es por ello por lo que debemos tener presente darle el seguimiento necesario a los datos que publicamos mediante una estrategia de monitoreo que permita reportar hallazgos y hacer comentarios o sugerencias sobre ellos, y así permitir la retroalimentación en el proceso de identificación, promoción y uso por parte de los usuarios.

A continuación, indicamos algunos aspectos que debe tomar en cuenta:

COMENTARIOS DEL USUARIO:

Los usuarios son los mejores críticos, a través de ellos usted podrá conocer no solo el nivel de uso, sino enterarse de los posibles errores que tengan sus datos, así como las sugerencias para agregar más datos o ampliar los mismos para mejorar el uso de los datos publicados. Los comentarios, sugerencias u observaciones que los mismos usuarios hagan sobre el conjunto de datos abiertos, deben servir de referencia para mejorar sus catálogos.

CALIDAD DE LOS DATOS

De la misma forma, avanzar en la madurez de los datos abiertos es importante para aumentar la calidad, valor y utilidad de estos. Cuide la disponibilidad de los datos, cree estrategias de revisión para validar cualquier enlace roto.

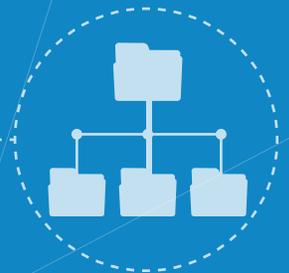
NIVEL DE IMPACTO

A través de herramientas estadísticas, validar si los datos que estamos aperturando está causando el impacto esperado en sus usuarios. Recuerde que el objetivo de abrir sus datos es que sean utilizados, valide y monitoree en qué áreas están siendo utilizados y cómo los están utilizando.



SECCIÓN 4

BUENAS PRÁCTICAS
Y CASOS DE ÉXITO



BUENAS PRÁCTICAS

Se entiende como buena práctica toda aquella acción que permita el mutuo entendimiento, es por ello por lo que acá se presentan algunas buenas prácticas al momento de abrir y publicar datos.

CON RELACIÓN A LOS NOMBRES Y DESCRIPCIONES DE CONJUNTOS Y RECURSOS

Una buena práctica al momento de publicar datos es definir los nombres tanto de un

conjunto como de su respectivo recurso, esto con el fin que cualquier persona que busque información, pueda encontrarla consolidada y de forma simple.

Un conjunto de datos es una agrupación de datos de una misma temática, atemporal y actualizable, es por ello por lo que se recomienda:

Recomendación	Uso incorrecto	Uso correcto
Incluir en el nombre palabras claves que describan la naturaleza del conjunto	Conjunto de datos para el presupuesto nacional de la republica hecho por el MINFIN del año 2020	Presupuesto Nacional de Guatemala
Evitar colocar fechas o temporalidades dentro del título del conjunto	Presupuesto Nacional de Guatemala 2020	Presupuesto Nacional de Guatemala
A menos que sea necesario, evitar colocar nombre de la institución que lo genera ya que esto puede incluirse en la descripción	Presupuesto Nacional de Guatemala elaborado por el MINFIN	Presupuesto Nacional de Guatemala
Si incluye siglas, se recomienda colocar su significado en la descripción	Clasificación de escuelas según el Ministerio de Educación de Guatemala -MINEDUC	Clasificación de escuelas según MINEDUC
Evitar colocar palabras de agrupación	Listado de concursos de Guatecompras	Concursos de Guatecompras
Evite crear conjuntos de una misma temática para diferentes versiones de esta	2020- Estudiantes Graduados 2020- graduaciones 2019- Estudiantes Graduados 2019- graduaciones	Estudiantes graduados 2020 - graduaciones 2019 - graduaciones

Se entiende como recurso un archivo con datos que pertenece a un conjunto. Al estar dentro de un conjunto es recomendable poder diferenciarse de otros recursos que representen datos similares en diferentes temporalidades. Algunas recomendaciones de conjuntos y recursos son las mismas.

Recomendación	Uso incorrecto	Uso correcto
Incluir en el nombre una nomenclatura que permita diferenciarlas de sus similares	Aportes municipales una parte Aportes municipales otra parte	2020- Aportes municipales 2019- Aportes municipales
Evitar colocar fechas o temporalidades dentro del título del conjunto. Si necesita diferenciar un mismo año y diferentes versiones, cree nomenclaturas	Presupuesto Nacional de Guatemala al 31 de marzo Presupuesto Nacional de Guatemala del 31 de marzo al 31 de julio	2019-Q1-Presupuesto Nacional de Guatemala 2019-Q2 Presupuesto Nacional de Guatemala
Cuando el recurso es un archivo de metadatos, identifíquelo adecuadamente	Estudiantes graduados información	[Metadatos] Estudiantes graduados

CONRELACIÓN A LA NORMALIZACIÓN DE LOS DATOS

Se entiende como normalización de datos proceso de organizar los datos en tablas o recursos de tal manera que al utilizar estos los resultados sean inequívocos y que den el resultado que se pretende, con el objetivo de organizar los datos en grupos lógicos y evitar la duplicidad de datos facilitando su gestión posterior. Parte de la normalización de datos se platica en la sección de Estandarización de datos.

NIVEL DE DETALLE

Al momento de publicar datos, valore el nivel de detalle, no es lo mismo indicar que en Guatemala existen 150 centros de educación que indicar en donde están localizados esos centros, incluyendo departamento, municipio, dirección e incluso georreferencia. Se debe evaluar hasta qué nivel abrir los datos para mejorar el impacto de estos.

ESTANDARIZACIÓN DE DATOS

La estandarización de datos se considera buena práctica ya que fomenta la interoperabilidad de datos, facilitando que diferentes instituciones puedan utilizar datos transversales o que respondan a una misma temática.

La interoperabilidad basada en estándares permite que diferentes conjuntos de datos puedan hablarse entre sí, de tal manera pueda existir el cruce de información, algo imposible cuando dos conjuntos de datos nombran de forma diferente a una misma entidad, por ejemplo, el nombre de un departamento o un país con diferente codificación.

Entre los casos de estandarización más recurrentes se pueden mencionar:

1. **Datos transversales:** Se considera un dato transversal cuando afecta a la mayoría de las áreas temáticas, ejemplo de ello son los siguientes:
 - a. **¿Dónde?:** Responden a una ubicación geográfica usualmente, tales como países, departamentos, municipios, localidades, direcciones, códigos postales, información georreferenciar.
 - b. **¿Quién?:** Responde a una identificación de personas ya sea físicas o jurídicas, tales como las entidades a niveles gubernamentales, organismos internacionales, otros países, sociedad civil, etc. En el caso de una persona física se debe considerar el respeto y resguardo a la información sensible.
 - c. **¿Qué?:** Aplica a objetivos o categorías de uso común, tales como categorías del presupuesto, clasificación métrica, etc.
2. **Datos específicos:** Se entiende como datos que afectan a temáticas o áreas específicas tales como actividades económicas, clasificación de enfermedades, terminología clínica, clasificación de unidades educativas, etc.

En caso no exista una clasificación oficial a nivel país, se recomienda utilizar estandarizaciones internacionales. A continuación, se mencionan algunos ejemplos.

ESTANDARIZACIÓN GEOGRÁFICA

PAÍSES:

Se conoce que el nombre del país puede variar según el idioma que corresponda, es por ello por lo que existe el estándar ISO 3166-1, que proporciona códigos para los nombres de países y otras dependencias administrativas. A través de esta codificación (sea alfa-2 o alfa-3) la clasificación de países puede ser utilizada sin equivocación a nivel mundial.

DEPARTAMENTOS Y MUNICIPIOS

Existe una norma internacional para la clasificación de departamentos en Guatemala, esta es la ISO 3166-2:GT. De igual manera el Estado de Guatemala está dividido según la estructura Departamento–Municipio, teniendo cada uno su respectivo código, de igual manera este puede ser extendido a entidades locales de ámbito territorial en que el municipio se divide, tales como: aldea, caserío, paraje, cantón, barrio, zona, colonia, lotificación, parcelamiento urbano o agrario, microregión, finca, y demás formas de ordenamiento territorial definidas localmente. Para fines de interoperabilidad el alcance de este queda sujeto a departamentos y municipios.

CÓDIGOS POSTALES

En Guatemala los códigos postales están formados por cinco dígitos, los dos primeros dígitos representan el departamento, la tercera cifra representa la ruta de distribución, los dos últimos dígitos representan la oficina de distribución

COMPRAS Y CONTRATACIONES

Puede considerarse un dato transversal al ser utilizado por toda institución del Estado, está dirigido por el Ministerio de Finanzas. Como parte de los datos estandarizados se encuentran:

- Renglones presupuestarios
- Catálogo de insumos
- Códigos de dependencias del Estado
- Actividades, programas y proyectos del Estado
- Entre otros.

CONJUNTOS DE DATOS SUGERIDOS

Según la Política Nacional de Datos todas las Dependencias que integran el Organismo Ejecutivo, deberán publicar y actualizar periódicamente sus conjuntos de datos (data sets) en el Portal de Datos Abiertos, por lo que surge la pregunta ¿qué datos se deben abrir? En la Estrategia de apertura de datos de la Sección 3 indica una serie de pasos para identificar y priorizar los datos que una institución puede abrir, basados en su naturaleza, tamaño y alcance legal, por otro lado, si es una institución del Estado, existe un compendio de datos genéricos que pueden incluirse.

TRANSPARENCIA ACTIVA

Se reconoce la Transparencia Activa, definida como: la obligación que tienen los organismos públicos de entregar cierta información relevante y actualizada cada mes de cómo están organizados; sus contratos y contrataciones, así como distintos modos de relación con la ciudadanía (transferencias, beneficios, mecanismos de participación) entre otros. Esta obligación se fundamenta en la entrega de

información pública relevante a la ciudadanía y al público en general; entrega que debe ser permanente, actualizada y de forma accesible y comprensible. De la mano van los datos indicados en la Ley de información Pública de Oficio tales como:

- Directorio de empleados y funcionarios, teniendo en consideración el resguardo de la información sensible.
- Remuneraciones
- Presupuesto de ingresos y egresos
- Ejecución presupuestaria
- Depósitos con fondos públicos
- Y demás datos contenidos en el artículo 10 de esta ley que puedan ser estructurados y procesados por máquina.

SEGÚN TEMÁTICA

Existen ciertos datos, según temática que pueden aperturarse, tales como

Ubicación geográfica

Si su institución cuenta con delegaciones o comprende diferentes ubicaciones ya sea propias o de servicios que preste, por ejemplo: Listados de escuelas, hospitales, oficinas institucionales, estaciones policiales, etc. Se sugiere incluir datos georreferenciales

Cobertura legal

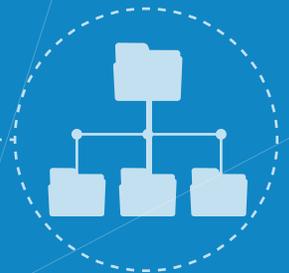
Como instituciones del Estado, estas están sujetas a disposiciones legales, artículos, decretos, y Todas las leyes y los estatutos aprobados.

Servicios y trámites

Cada una de las instituciones del Estado tiene una misión para con el ciudadano que se traduce en servicios que presta. Un buen catálogo de datos será el listado de servicios y la información (requisitos, datos de quien atiende, horarios, etc.) que este comprenda.



ANEXOS



A1. GLOSARIO

Anonimización: Proceso de remover los datos personales de una base de datos, buscando la publicación segura de datos para su reúso.

Data set: Conjunto de datos que forman una unidad significativa y que habitualmente pueden mostrarse con un fichero. También puede referirse a la unidad mínima de datos sujetos a carga, publicación, transformación y descarga.

CKAN: Una herramienta para publicar portales de datos abiertos que permite administrar y publicar colecciones de datos. CKAN es utilizado por gobiernos locales y nacionales, institutos de investigación y otras organizaciones que recopilan datos, a lo largo de todo el mundo

Creative Commons: Licencia que define el espacio entre la protección absoluta de los derechos de autor y el dominio público. Las licencias te ayudan a conservar tus derechos autorales invitando a usar tu obra bajo el esquema “Algunos derechos reservados”. Referencia: <http://www.creativecommons.mx/que/>.

CSV: Formato sencillo y abierto para representar datos en forma de tabla, en las que las columnas se separan por comas. En los casos en los que un valor contenga coma (como el separador decimal en algunos países), se recomienda encerrar entre comillas los valores o utilizar otro separador como “punto y coma” o “tabulador”. Referencia: http://es.wikipedia.org/wiki/Comma-separated_values.

- Formato de archivos separados por coma, donde cada valor es delimitado por una coma o punto y coma; es una forma sencilla de publicar datos, pero necesariamente se debe publicar también información sobre los datos que contiene (meta data). De otra manera los datos serán difícilmente interpretados.

JSON: JavaScript Object Notation es un lenguaje muy sencillo de automatizar su procesamiento. Si bien no está tan expandido como XML su filosofía similar y su fácil interpretación lo hacen un lenguaje muy apropiado para el intercambio de datos.

LINKED DATA (Datos Relacionados) – Forma parte de la web semántica, a diferencia de la web actual donde los documentos de

hipertexto se relacionan a diferentes puntos en el HTML, esta nueva forma enlaza los datos con cosas arbitrarias que se describen en formato RDF.

Metadatos: – Datos que describen otros datos.

ODS – Formato de archivo de Open Document, es un estándar abierto para documentos de hojas de cálculo.

PDF: El formato de documento portátil, inventado por Adobe Systems, es un estándar abierto para el intercambio de documentos electrónicos que mantiene la Organización Internacional de Normalización (ISO).

Servicio web: Un servicio web (en inglés, Web Service o Web services) es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones.

TXT – Formato de archivo de texto plano.

URI – Identificador único de recurso, permite identificar únicamente un recurso, página, documento, dirección de correo electrónico, etc.

URL: Un localizador de recursos uniforme, más comúnmente denominado URL (sigla en inglés de uniform resource locator), es una secuencia de caracteres, de acuerdo con un formato modélico y estándar, que se usa para nombrar recursos en Internet para su localización o identificación, como por ejemplo documentos textuales, imágenes, vídeos, presentaciones digitales, etc.

W3C: Consorcio internacional que produce recomendaciones para la WWW.

Web Semántica – La web semántica es una visión futura de la web donde los datos publicados deben tener un significado exacto, estos datos deben ser entendidos y procesados por computadoras. “Es una web de datos que pueden ser procesados directa o indirectamente por máquinas” según la definición de Tim Berners-Lee

XML: – Formato extensible de marcas, lenguaje que permite la definición de etiquetas o marcas derivado del SGML.

XLS – Formato de archivo creado por Microsoft para su herramienta de hoja de cálculo.

A2. TIPOS DE LICENCIAMIENTO

CREATIVE COMMONS

Creative Commons es una organización sin fines de lucro que permite el intercambio y uso de la creatividad y el conocimiento a través de herramientas legales gratuitos. Todas las licencias Creative Commons tienen

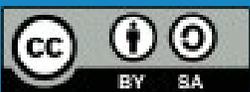
importantes características en común. Cada licencia ayuda a los creadores a mantener sus derechos de autor al mismo tiempo que permiten a otros copiar, distribuir, y hacer algunos usos de su obra, al menos de forma no comercial.

Existen cuatro condiciones:

	Atribución (BY) Attribution	El beneficiario de la licencia tiene el derecho de copiar, distribuir, exhibir y representar la obra y hacer obras derivadas siempre y cuando reconozca y cite la obra de la forma especificada por el autor o el licenciente.
	No Comercial (NC) Non-Commercial	El beneficiario de la licencia tiene el derecho de copiar, distribuir, exhibir y representar la obra y hacer obras derivadas para fines no comerciales.
	No derivadas (ND) No derivative works	El beneficiario de la licencia solamente tiene el derecho de copiar, distribuir, exhibir y representar copias literales de la obra y no tiene el derecho de producir obras derivadas.
	Compartir igual (SA) Share-alike	El beneficiario de la licencia tiene el derecho de distribuir obras derivadas bajo una licencia idéntica a la licencia que regula la obra original.

En el portal único de datos abiertos de Guatemala podrá encontrar las siguientes:

Acrónimo

	Creative commons attributions	BY
	Creative commons attributions Share-Alike	BY-SA
	Creative Common CCZERO	CC0
	Creative Common Non-Commercial (cualquiera)	BY-NC

GNU FREE DOCUMENTATIONS LICENCE

La Licencia de documentación libre de GNU o GFDL (GNU Free Documentation License) es una licencia copyleft para contenido libre, diseñada por la Fundación para el Software Libre (FSF) para el proyecto GNU.

Esta licencia, a diferencia de otras, asegura que el material licenciado bajo la misma esté disponible de forma completamente libre, pudiéndose copiar, redistribuir, modificar e incluso vender siempre y cuando el material se mantenga bajo los términos de esta misma licencia (GNU GFDL). En caso de venderse en una cantidad superior a 100 ejemplares, deberá distribuirse en un formato que garantice futuras ediciones (debiendo incluir para ello el texto o código fuente original).

Dicha licencia fue diseñada principalmente para manuales, libros de texto y otros materiales

de referencia e institucionales que acompañaran al software GNU. Sin embargo, puede ser usada en cualquier trabajo basado en texto, sin que importe cuál sea su contenido.

La licencia GNU free documentation está disponible en el portal único de datos abiertos de Guatemala.

OPEN DATA COMMONS LICENCE

Open Data Commons es un compendio de herramientas legales que le ayudan a proveer y usar los datos abiertos. Open Data Commons es un proyecto de Open Knowledge Foundation que funciona bajo el Advisory Council y tal como la Fundación es sin fines de lucro para el beneficio general de la comunidad del conocimiento abierto -general open knowledge community-. Cuenta con tres tipos de licencia que están disponibles en el portal único de datos abiertos de Guatemala:

ODC Public Domain Dedication and License (PDDL)

Public Domain Dedication & License is a document intended to allow you to freely share, modify, and use this work for any purpose and without any restrictions. This license is intended for use on databases or their contents ("data"), either together or individually

Open Data Commons Attribution License (ODC-By) v1.0

The Open Data Commons Attribution License is a license agreement intended to allow users to freely share, modify, and use this Database

Open Database License (ODbL)

The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Database while maintaining this same freedom for others.

A3. PROCESO DE ANONIMIZACIÓN

Los datos abiertos no son contrarios a la protección de la información, es por ello por lo que previo a abrir los datos debemos identificar qué datos podrían ser vulnerables. Para ello tenemos dos tipos de datos vulnerable.

DATO PÚBLICO CLASIFICADO

Corresponde a todo aquel dato cuyo acceso libre puede causar daños de derechos a personas naturales o jurídicas. Esto es un dato que:

- Pone en riesgo la intimidad de las personas, por ejemplo, nombre, DPI, dirección, teléfono, etc.
- Pone en riesgo la vida, salud o seguridad de la persona: Nombre y número de cuenta bancaria, DPI y tarjetas de crédito.
- Afecta los secretos comerciales, industriales y/o profesionales, por ejemplo, detalles de patente, listados de sociedades anónimas, etc.
- Todos aquellos establecidos por ley como privados.

DATO PÚBLICO RESERVADO

Corresponde a todo aquel dato cuyo acceso libre puede causar daño a los intereses públicos. Esto corresponde a:

- Datos de defensa y seguridad nacional
- Relaciones internacionales
- Datos relacionados al debido proceso en los procesos judiciales.
- Derechos de la infancia y la adolescencia.
- Estabilidad macroeconómica y financiera del país.
- Salud pública.

¿QUÉ HACER SI ENCUENTRA DATOS VULNERABLES?

El proceso de anonimización de datos es complejo y depende de la profundidad que quiera alcanzar. Este documento no brinda una guía paso a paso que permita crear datos anónimos perfectos, sino que ofrece un listado de buenas prácticas que puede aplicar a sus datos.

PRINCIPIOS DE LA ANONIMIZACIÓN

PRINCIPIO PROACTIVO

La protección de la privacidad es el primer objetivo de la anonimización y su gestión debe realizarse de forma proactiva y no reactiva. Desde el inicio conceptual del diseño del sistema de información o producto a utilizar en el proceso de anonimización se tomarán las medidas necesarias para garantizar la privacidad de las personas. La privacidad no puede garantizarse a posteriori como el resultado de la reparación de brechas existentes en el proceso de anonimización o perjuicios ocasionados a los interesados, por lo que es necesario asegurar la inexistencia de posibles cadenas de reidentificación de los interesados en los datos anonimizados.

PRINCIPIO DE PRIVACIDAD POR DEFECTO.

El primer requisito conceptual en el diseño de un sistema de datos será garantizar la confidencialidad de los interesados. Por lo tanto, conviene que desde el inicio se salvaguarde la privacidad teniendo en cuenta la granularidad o grado de detalle final que deben tener los datos anonimizados.

PRINCIPIO DE PLENA FUNCIONALIDAD.

Desde el inicio del diseño del sistema de datos se tendrá en cuenta la utilidad final de los datos anonimizados, garantizando en la medida de lo posible la inexistencia de distorsión con

relación a los datos no anonimizados. De esta forma, se garantizará la utilidad de los datos anonimizados.

PROCESO DE ANONIMIZACIÓN

Algunos aspectos que pueden ser tenidos en cuenta para abordar la eliminación o enmascarar las variables de identificación pueden ser los siguientes:

- Determinar la finalidad de los datos anonimizados: plazos de conservación, uso estadístico, científico, o cualquier uso posterior.
- Establecer las variables confidenciales necesarias para el tratamiento de los datos anonimizados e identificar las variables de confidencialidad que no vayan a ser necesarias en el tratamiento de los datos anonimizados. Como variable de confidencialidad se entenderá cualquier dato existente sobre una persona, tanto si permite su identificación como si a priori su identificación no es posible.
- Eliminación de datos identificativos directos o indirectos no necesarios: nombres, fecha de nacimiento, teléfono, DPI, email, dirección postal, número de cuentas bancarias, matrículas de vehículos, identificador dispositivo móvil, número de serie, dirección IP, identificadores biométricos, fotografía o imagen, etc.
- Utilización de rangos para enmascarar a las personas cuando existen microdatos concretos que permiten la identificación directa de personas o colectivos específicos.

Datos comunes que son eliminados:

- Nombres
- Todos los elementos de fecha de nacimiento, excepto el año.
- Números de teléfono y fax.
- Números de identificación: DPI, cédula, pasaporte, números asociados a la seguridad social, licencias de conducción, NIT, número de padrón, etc.
- Direcciones de correo electrónico.
- Números de cuentas bancarias.
- Identificadores del vehículo, placa, etc.
- Identificadores de dispositivos móviles y números de serie.

- Direcciones de IP.
- Identificadores biométricos.
- Fotografías e imágenes similares.
- Cualquier otro número único de identificación.
- Dirección de domicilio

TÉCNICAS DE ANONIMIZACIÓN

1. Algoritmo de Hash: Un algoritmo de hash6 es un mecanismo que, aplicado a un dato concreto, genera una clave única o casi única que puede utilizarse para representar un dato.
2. Algoritmo de cifrado: Un algoritmo de cifrado homomórfico permite realizar operaciones con datos cifrados de tal manera que el resultado de las operaciones es el mismo que si las operaciones se hubieran realizado con los datos sin cifrar.
3. Sello de tiempo: también hay que tener en cuenta la posibilidad de utilizar en el proceso de anonimización algoritmos de sello de tiempo con el fin de garantizar la fecha y hora en la que la anonimización ha sido realizada, o incluso algoritmos de firma electrónica que permiten garantizar la identidad electrónica de quien ha realizado la anonimización.
4. Capaz de anonimización: junto con estos procesos de enmascaramiento y anonimización podemos utilizar lo que podría denominarse la anonimización por capas.
5. Perturbación de datos: variación y supresión sistemática de datos que evita que las cifras resultantes faciliten información sobre casos específicos tales como: microagregación, intercambio aleatorio, permutación, redondeo, reajuste de pesos, ruido aleatorio, etc.
6. Reducción de datos: disminuyen el nivel de detalle de los datos originales evitando la presencia de datos únicos o atípicos sin relevancia para el resultado final.

La estrategia seleccionada será en base a las necesidades de la institución y al grado de detalle que quiera brindar. Recuerde que al realizar la anonimización esta es definitiva e irreversible.

A4. FORMATO DE DATOS ABIERTOS



FORMATO	FUNCIÓN
CSV (TSV)	Para representar datos en forma de tabla, donde las columnas se separan por comas (o tabulaciones) y las filas por saltos de línea.
DVI	(Digital Visual Interfaz). Interfaz estándar digital creado por DDWG (Digital Display Working Group) para convertir señales analógicas en señales digitales para satisfacer monitores analógicos y digitales.
ODS / XLS	Formatos (abierto en el caso de ODS y propietario de Microsoft en el caso de XLS) de hoja de cálculo que muestran los datos en celdas organizadas en filas y columnas. Cada celda contiene datos o una fórmula con referencias relativas, absolutas o mixtas a otras celdas.
GML	Vocabulario XML para el modelaje, transporte y almacenamiento de datos geográficos.

HTML	El texto se modela a partir del uso de etiquetas o tags. También se pueden agregar scripts al código fuente HTML (generalmente JavaScript, PHP, etc.). Por lo general los diseñadores utilizan herramientas gráficas WYSIWYG para la creación de páginas web, las cuales generan el código fuente HTML automáticamente
KML	Vocabulario XML para representar datos geográficos en tres dimensiones. Suelen distribuirse comprimidos en formato KMZ.
JSON	Notación muy popular para el intercambio de datos basada en una simplificación de la notación de JavaScript.
PDF	Es multiplataforma y está sumamente extendido especialmente en lo que se refiere a documentación de manuales y libros. No es el formato más recomendado para abrir los datos superiores a 1 estrellas.
RDF	Modelo de representación para la descripción de recursos siguiendo los fundamentos de la Web Semántica. No debe ser confundido con el texto enriquecido, que es una especificación completamente diferente.
RSS	Vocabulario XML para facilitar la distribución de contenidos y actualizaciones.
SHP	Formato propietario de datos espaciales que se ha convertido en un estándar de-facto y sirve para almacenar tanto la localización de los elementos geográficos como los atributos asociados a ellos.
TXT	Formato abierto para contener texto plano, o texto sin formato. Suele tener la extensión ".txt".
WMS	Formato de servicio web para producir mapas de datos referenciados espacialmente, de forma dinámica a partir de datos geográficos.
XML	Lenguaje general de etiquetado que define un conjunto de reglas para la codificación de documentos a través de modelos específicos legibles y procesables.

A5. EJEMPLOS DE APLICACIÓN DE DATOS ABIERTOS



A donde quieres ir ya lo sabes, MYBUS te dice como llegar.

MYBUS

MYBUS es una aplicación que resuelve consultas inteligentes basadas en la red de transporte público de Mar del Plata. Este es un sitio web y aplicación móvil desarrollado por programadores de la Ciudad de Mar del Plata Argentina.

GLOBAL FOREST WATCH

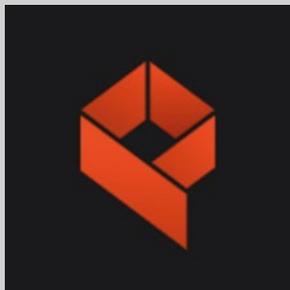
Global forest watch

El cambio climático y la protección del medio ambiente son una creciente preocupación de los gobiernos y la sociedad en general. Un caso interesante es el de la plataforma Global Forest Watch, que ofrece datos sobre la evolución de las masas forestales en el tiempo.



Open elections

Los datos abiertos pueden ser una pieza fundamental para revitalizar la democracia y hacerla más participativa. Podemos resaltar la página OpenElections, que pretende revisar y certificar los resultados de las elecciones en EE. UU. de los últimos años.



Poderopedia

Plataforma colaborativa de periodismo de datos que mapea mapea el periodismo de datos en los temas de negocio y política y ayuda a entender las relaciones entre las personas, empresas y organizaciones que se convierten en noticia e influyen en nuestra vida diaria.



Medicina

Es un plataforma que utiliza los datos nacionales de salud en Brasil para conectar los equipos de salud, pacientes, registros médicos y dispositivos en clínicas, hospitales y sistemas de salud.

A6. PLANTILLA PARA INVENTARIO DE DATOS

Tiene como objetivo generar un listado de datos que una institución posee. Da un vistazo general de los datos que la institución produce.

- **Nombre del conjunto:** Define un nombre con el cual pueda identificarse claramente el conjunto de datos
- **Descripción del conjunto:** Indica una descripción clara del conjunto, qué es, a qué corresponde, etc.
- **Dueño/generador del conjunto:** Indica quién es el que produce los datos, la dirección/unidad/departamento dueño de los datos de origen
- **¿Está digital?:** indica si los datos están en formato digital, físico, si son editables, y una ruta donde ubicarlos
- **¿Reglamento o normativa que le aplica?:** indica si los datos obedecen a alguna normativa existente, si son sujetos a arancel, o si existe razón legal que los afecte.
- **¿Es aperturable?:** Se define si los datos pueden aperturar (una vez procesados) o son de carácter privativo (por seguridad nacional, sujetos a derecho de autor o similar)
- **Prioridad:** Con base a las preguntas generadoras, se define el nivel de prioridad del conjunto para conocer en qué momento se publicará
- **Datos asociados:** Indica si los datos a aperturar tienen relación con otros, principalmente cuando no pertenecen al mismo conjunto o son estándar, esto aplica a, por ejemplo, departamento, países, identificadores únicos, etc.
- **Periodicidad:** Indica cada cuanto el conjunto será actualizado.

Nombre del Conjunto	Descripción del Conjunto	Dueño/generador del Conjunto	¿Está en digital?	¿Tiene datos sensibles?

Reglamento o normativas que le aplican	¿Es aperturable?	Prioridad	Datos asociados	Periodicidad

A7. PLANTILLA PARA MATRIZ DE DATOS

Se presenta como una propuesta para documentar conjuntos y recursos de datos. Se genera con base a los datos obtenidos y sirve de guía al equipo técnico para el proceso y transformación de datos crudos a datos abiertos.

- **Nombre del conjunto de datos:** Define un nombre, basado en el inventario de datos inicial.
- **Descripción del conjunto:** Define una descripción clara del conjunto, respetando a las buenas practicas definidas en la sección 4 de este manual.
- **Nombre del recurso:** Define un nombre para cada recurso que permita identificarlo del resto, tanto en temporalidad como en forma
- **Descripción del recurso:** Define una descripción clara del recurso, con

forme a las buenas prácticas definidas en la sección 4 de este manual

- **Periodicidad:** Con base en el inventario de datos, se define una periodicidad definitiva para cada recurso.
- **Nombre del campo:** Indica el nombre del campo que se va a procesar dentro de un recurso
- **Descripción:** Indica una descripción clara del campo que se va a procesar.
- **Tipo:** Define un tipo de dato, este puede ser cadena de texto, entero, booleano, fecha, etc.
- **Acción:** Define la acción que se hará sobre ese campo, esto puede ser anonimizar, convertir a una fecha en específico, no hacer nada, eliminar parte del campo, etc.

Nombre del conjunto de datos				
Descripción del conjunto				
RECURSO 1				
Nombre del recurso				
Descripción del recurso				
Periodicidad				
Nombre del campo	Descripción	Tipo	Acción	¿Es catálogo?

A8. PLANTILLA DE INFORMACIÓN DE PUBLICACIÓN

Esta herramienta tiene como objetivo servir de guía a la persona que publicará los datos en el Portal de Datos Abiertos. No se debe confundir con los metadatos propios de cada recurso

CONJUNTO DE DATOS

Título del conjunto	
Descripción	
Dependencia	
Etiquetas	
Tema	
Licencia	
Frecuencia	
¿Es privado?	

RECURSO 1

Título del recurso	
Descripción del recurso	
Inicio periodo	
Fin de periodo	
Área/Persona de contacto	
Correo electrónico de contacto	
URL	
¿Es privado?	



MANUAL DE DATOS ABIERTOS