



SIGAMOS *creando futuro*

POLÍTICA DE DATOS ABIERTOS

GUÍA DE IMPLEMENTACIÓN

**Secretaría Técnica y de Planificación y
Secretaría de Participación,
Transparencia y Anticorrupción
de la Presidencia**

Junio 2018



CONTENIDO

PRESENTACIÓN

1. DEFINIR ENLACE Y ADMINISTRADOR INSTITUCIONAL DE DATOS ABIERTOS	5
2. CREAR LA COMISIÓN INSTITUCIONAL DE DATOS ABIERTOS (CIDEA)	5
3. CREAR INVENTARIO INSTITUCIONAL	7
4. CREACIÓN DEL CATÁLOGO GUBERNAMENTAL DE DATOS ABIERTOS	9
4.1 Plan de Publicación	10
5. IDENTIFICAR Y RECOLECTAR DATOS	10
5.1. Normativa de acceso y uso de los datos entre las instituciones	11
5.2. La homologación y relación de procesos entre organizaciones	11
5.3. La equivalencia semántica de los campos que se comparten	12
5.4. El acoplamiento técnico entre distintos sistemas de información	12
6. PREPARAR LOS CONJUNTOS DE DATOS	12
6.1. Promover la interoperabilidad	15
6.2. Hacer entendible los datos a los usuarios	15
6.3. Habilitar la legibilidad por máquinas	15
6.4. Asegurar la consistencia y calidad de los datos	15
6.5. Asegurar la normalización de los datos	16
7. PUBLICAR DATOS	16
7.1 Documentar los conjuntos de datos de acuerdo al estándar DCAT	17
7.2 Publicar el Catálogo Gubernamental de Datos Abiertos en el Portal datos.gob.sv	18
8. COMUNICAR Y PROMOVER	20
9. GLOSARIO	21

ÍNDICE DE TABLAS

Tabla 1: Roles y funciones de la Comisión Institucional de Datos Abiertos (CIDEA)	6
Tabla 2: Pasos para elaborar el Inventario Institucional	8
Tabla 3: Referencia de formatos, su nivel de apertura y estructura	13
Tabla 4: Ejemplos de un conjunto de datos sin errores	16
Tabla 5: Ejemplos de conjuntos de datos con errores	16
Tabla 6: Metadatos (descriptores) del conjunto de datos (dcat:Dataset)	17
Tabla 7: Metadatos (descriptores) del recurso o descargable (dcat:Distribution)	18

PRESENTACIÓN

Los Datos Abiertos son una herramienta que facilita la toma de decisiones y la rendición de cuentas de múltiples sectores. Para poner los datos en acción y detonar la generación de valor, es importante socializar, publicar y promover el uso de los Datos Abiertos. Por esto, las instituciones deberán utilizar los canales de difusión pertinentes para dar a conocer los planes y acciones para dar cumplimiento a las disposiciones de la Política Nacional de Datos Abiertos (PNDA).

Para conocer sobre la importancia y los principios de Datos Abiertos, refiérase al documento de la PNDA ubicado en <https://datos.gob.sv/plantillas>

Este documento está orientado a facilitar la implementación de dicha Política, a través de los siguientes pasos:

Lineamientos de la PNDA	Pasos
Marco Administrativo	1. Nombrar al enlace y al administrador de Datos Abiertos institucionales
	2. Crear la Comisión Institucional de Datos Abiertos (CIDEA)
Inventario Institucional	3. Crear inventario Institucional de datos
Seguridad de los datos	4. Priorizar datos a publicar
Planificación de Recursos	5. Crear plan de publicación
Gestión del cambio	6. Identificar y recolectar datos
Alianzas para compartir datos	
Datos Interoperables	7. Preparar Datos
Apertura de los Datos	8. Publicar datos
Alianzas para entregar datos	
Aprovechar datos Abiertos	9. Comunicar y promover el uso
Divulgación y promoción	

1. DEFINIR ENLACE Y ADMINISTRADOR INSTITUCIONAL DE DATOS ABIERTOS

El titular de cada institución designará a dos servidoras o servidores públicos responsables de dar cumplimiento a la Política Nacional de Datos Abiertos: el Enlace Institucional y la persona Administradora de Datos Abiertos.

El "Enlace Institucional" debe ser una persona integrante de la alta gerencia y actuará en representación directa del titular de la institución, por lo que contará con poder de decisión en los temas relacionados a la PNDA y será responsable de coordinar las acciones para cumplir con la PNDA.

Para la persona "Administradora de Datos Abiertos" se recomienda delegar al Director o Directora de Tecnologías de la Información de la institución, o su equivalente. Quien será responsable de coordinar la creación de un Inventario Institucional de Datos, implementar los procesos tecnológicos para intercambiar datos con otras instituciones de forma segura, y que los datos abiertos se publiquen en el sitio web institucional y se vean reflejados en el Portal datos.gob.sv.

Sobre estas designaciones, el Titular de la institución deberá enviar un oficio con las designaciones al Secretario Técnico y de Planificación de la Presidencia, siguiendo la plantilla No. 1 que se encuentra en el Portal www.datos.gob.sv.

2. CREAR LA COMISIÓN INSTITUCIONAL DE DATOS ABIERTOS (CIDEA)

La Secretaría Técnica y de Planificación de la Presidencia (STPP) podrá hacer recomendaciones sobre la designación de estas áreas cuando las instituciones cuenten con áreas administrativas estratégicas encargadas específicamente de planeación y administración de datos.

La persona Enlace y la Administradora deberán coordinar la creación de la CIDEA bajo el lineamiento que manda a cumplir la Política de Datos Abiertos y la presente Guía, con las Unidades Administrativas pertinentes al interior de su institución, además del Enlace y Administrador de Datos, la CIDEA debe incluir al menos a:

- **Dirección de Planificación**
- **Oficina de Información de Respuesta** o Unidad de Acceso a Información Pública, según sea el caso.
- **Oficial de Archivo**
- **Unidad de Estadística Institucional**

El titular deberá facilitar los espacios y tiempos necesarios para que la CIDEA elabore el plan de implementación de la Política para toda la institución.

La asignación y roles del personal de la institución dentro de la CIDEA serán:

Tabla 1 : Roles y funciones de la Comisión Institucional de Datos Abiertos (CIDEA)

Funciones	Rol
<ul style="list-style-type: none"> Garantizar la elaboración del plan de implementación de la PNDA de su institución. Asegurarse de que las recomendaciones de esta Guía sean atendidas por las áreas responsables. Facilitar el proceso de apertura de datos, priorizar y divulgar su publicación. Facilitar alianzas con otras instituciones para intercambiar datos Ser el principal punto de contacto con la STPP, representando a su institución en todas las actividades a las que sea convocado. 	Enlace de Datos Abiertos
<p>Coordinar con Unidades Generadoras de datos para:</p> <ul style="list-style-type: none"> Crear el Inventario de Datos y asegurar la conversión de conjuntos de datos a formatos abiertos e interoperables. Asegurar la calidad de los datos, su consistencia, seguridad y disponibilidad para ser consultados desde su sitio web institucional y del Portal nacional de Datos Abiertos. Implementar los mecanismos técnicos para compartir datos con otras instituciones de un mismo Dominio de Integración. Proponer a la CIDEA el plan de publicación de datos y asegurar su cumplimiento. Coordinar con la OIR para que la información más solicitada sea publicada en formato abierto. 	Administradora de Datos Abiertos
<ul style="list-style-type: none"> Identificar los mecanismos institucionales adecuados para la recopilación y ordenamiento de los datos que se producen institucionalmente. Incorporar en el plan institucional el plan de publicación de acuerdo a las prioridades que se identifiquen, vinculándolo a las estrategias del Plan Quinquenal. 	Dirección de Planificación
<ul style="list-style-type: none"> Apoyar la priorización de datos a ser publicados. Preparar un reporte anual con los conjuntos de datos más solicitados por la ciudadanía para ser publicados como Datos Abiertos. Llevar estadísticas de la descarga de bases de datos y hacer encuestas o consultas sobre otros datos que puedan interesar al público. Velar por la aplicación de la normativa vigente en materia de datos personales, reservados o confidenciales en relación a los Datos Abiertos de la Institución. 	Oficial de Información y Respuesta
<ul style="list-style-type: none"> Apoyar en el ordenamiento de datos a ser publicado y en la revisión de los metadatos relacionados. Velar por la aplicación de la normativa vigente en materia de gestión documental física y electrónica 	Oficial de Archivo
<ul style="list-style-type: none"> Apoyar la priorización de datos a ser publicados con especial énfasis en los datos estadísticos misionales que tiene la institución. Velar porque los datos estadísticos estén actualizados, cumplan la normativa estadística nacional y normativa de igualdad (desagregando los datos como mínimo por sexo, en los casos que corresponda). Comunicar a la CIDEA y STPP cuando los datos estadísticos en la institución también sea recabada por otra institución para evitar duplicidad de datos y contradicción entre las mismas. 	Unidad de estadística institucional
PARTICIPACIONES DE APOYO A LA CIDEA:	
<ul style="list-style-type: none"> Participar eventualmente de las reuniones a las que sean convocados por la CIDEA, a fin de que aseguren la calidad mínima de los datos y prepararlos para su publicación siguiendo las recomendaciones de esta guía. Participar en la elaboración del plan de implementación de la PNDA y apoyar en la publicación de Datos Abiertos según las prioridades que se identifiquen y de los datos que vayan a ser abordados. 	Unidades Generadoras o Gestores de datos
<ul style="list-style-type: none"> Diseñar y coordinar la ejecución de la estrategia de promoción y difusión de los datos que están a disposición ciudadana. Publicitar el inventario institucional de los datos y del Portal de datos abiertos 	Dirección de comunicaciones

Consideración: Si la institución por su naturaleza o tamaño no cuenta con suficiente personal o las unidades enunciadas, el enlace institucional y el administrador de datos serán quienes asumirán la implementación de la política en su totalidad.

A partir de la aprobación de la Política Nacional de Datos, la definición de los referentes y de las CIDEA se realizará en un plazo de 10 días hábiles.

3. CREAR INVENTARIO INSTITUCIONAL

El Administrador de Datos deberá solicitar a todas las Unidades Administrativas el llenado del **Inventario institucional** y **Catálogo de datos reutilizable** para cada conjunto de datos que generan o administran. Cada unidad institucional deberá reportar al Administrador de Datos Abiertos los datos que generan, la prioridad en este caso son los datos primarios y misionales, es decir aquellos que por competencia produce la institución y que nadie más produce para lo cual la institución deberá analizar su ley de creación, servicios al ciudadano y Plan Estratégico. En dicho catálogo se incluye información sobre la clasificación de los datos de acuerdo a la Ley de Acceso a Información Pública.

El Administrador junto a la CIDAE será responsable de consolidar los datos y aplicar filtros para elaborar el **Inventario Institucional** y seleccionar aquellas que conformarán el Catálogo Gubernamental de Datos Abiertos en el Portal datos.gob.sv.

Este ejercicio también servirá para identificar duplicidades en la generación de datos. Si existiera duplicidad se deberán tomar acciones correctivas que garanticen una sola área responsable para la generación (o consolidación) de un conjunto de datos. Es posible además que se identifiquen vacíos en los datos, en cuyo caso la institución es responsable de evaluar si pueden generarse internamente o deben obtenerse de otra institución. Este inventario deberá ser revisado semestralmente y enviado al enlace de la STPP que haya sido designado para ello.

El **Inventario institucional**, que debe elaborarse en el plazo de un mes calendario, es la base para crear el **Catálogo Gubernamental Datos Abiertos** y para su llenado se realizarán los siguientes pasos:

Tabla 2 : Pasos para elaborar el Inventario Institucional

Actividad	Descripción
Identificar temas sustantivos	<p>La identificación de los temas sustantivos se elabora a partir de un ejercicio de revisión de los siguientes seis elementos:</p> <ol style="list-style-type: none"> 1. Información derivada de los procesos misionales de la Institución según su ley de creación. Esta debe estar incluida dentro de las principales funciones que desempeña la Unidad Generadora de Datos (marco jurídico de la institución, plan estratégico, manuales de procedimientos, entre otros). 2. Información utilizada para diseño y seguimiento de programas presupuestarios o políticas públicas ejecutadas por la institución. 3. Información para la atención a tratados, acuerdos o compromisos internacionales. 4. Información para la atención a solicitudes presidenciales, legislativas o sectoriales. 5. Información utilizada para la difusión periódica en publicaciones oficiales o académicas, e 6. Información recurrente en solicitudes ciudadanas.
Definir los conjuntos de datos	<p>Cada uno de los temas identificados en el paso anterior está relacionado con al menos un conjunto de datos, que a su vez está relacionado con el quehacer de la Unidad Generadora de Datos correspondiente.</p> <p>Estos pueden ser, entre otros, datos sobre la situación histórica, actual o proyectada:</p> <ul style="list-style-type: none"> • del sector donde se desempeña la función, • de la infraestructura utilizada, • del ejercicio presupuestal asociado, • del personal dedicado, • de indicadores de la operación, o • de la regulación aplicable (información reservada, confidencial, pública). <p>En el Inventario Institucional, estos Grupos de Datos se deben registrar bajo la Columna A, titulada 'nombre del grupo', preferentemente con 'unidades de medida' y 'tema'.</p> <p>Por ejemplo:</p> <ul style="list-style-type: none"> • número de viviendas con servicio de agua potable, • número de habitantes beneficiados por el programa escuela solidaria, • monto ejecutado en proyectos de infraestructura de telecomunicaciones <p>En este Inventario se debe identificar cuáles de estos pueden ser reutilizados por otras entidades y cuáles pueden ser dato abierto.</p>
Definir los recursos de Datos	<p>Cada Conjunto de Datos pueden ser publicado en diversos formatos para consumo del público ya sea por variaciones en su estructura (ej.: cartográfico, tabular), temporalidad (ej.: años, meses, días), medio (ej.: CSV, API), unidades de medida (población, viviendas), etc. Cada una de estas variantes es conocida como Recurso de Datos.</p> <p>Es importante asegurarse de que el Recurso de Datos considere un solo concepto, en un solo ámbito geográfico y dentro de un solo periodo de tiempo, sin mezclar datos de diversos tipos.</p> <p>Cada Recurso de Datos se debe registrar en la Columna C 'nombre del recurso' como 'formato', 'unidad', 'materia', 'detalle geográfico' y 'período'</p> <p>Por ejemplo:</p> <ul style="list-style-type: none"> • mapa de número de viviendas con cobertura de alcantarillado por municipio 2014 - 2016, • tabla de número de habitantes beneficiados por el programa escuela solidaria por departamento en julio 2013 • monto ejecutado en proyectos de infraestructura de telecomunicaciones por departamento, período 2009-2014 <p>Adicionalmente, se incluye una breve descripción aclaratoria del Recurso de Datos en la Columna X '¿de qué es?'</p>

Actividad	Descripción
<p>Identificar formato y responsables</p>	<p>Es necesario identificar el formato de los Recursos de Datos (base de datos, Excel, etc.) y se registra en la Columna G '¿en qué plataforma, tecnológica, programa o sistema se albergan?'</p> <p>De igual forma, se debe registrar el cargo del funcionario designado como Responsable para su actualización en la Columna D, 'responsable', siendo la persona que ostenta la jefatura de la unidad generadora.</p> <p>Una vez identificadas las fuentes de datos y sus responsables, cada Unidad Generadora registra su listado en la plantilla del Inventario de Datos para su entrega al Administrador de Datos.</p> <p>Partiendo de la definición de Información Pública prevista en la LAIP y de la regulación aplicable a la función específica, la Unidad Generadora verifica en la Columna B 'Puede estar disponible al público' de ser necesario, en la Columna C 'razón por la cual los datos son confidenciales o reservados' debe referir el fundamento jurídico, con base en la LAIP, que así lo acredita.</p> <p>Por último, en la Columna L 'fecha tentativa de publicación (en formato AAAA-MM)' la Unidad Generadora determina una fecha estimada de publicación de cada Recurso de Datos.</p> <p>Para realizar este paso se puede descargar el documento <i>Inventario Institucional</i> disponible en http://datos.gob.sv/plantillas/inventario.xls</p>
<p>Verificar y Aprobar Inventario</p>	<p>Una vez recibidos los Inventarios de Datos de cada Unidad Generadora, el Administrador de Datos de la dependencia verifica que esté completa cada plantilla e integra el Inventario Institucional de Datos, ordenando e identificando claramente los temas misionales a los que pertenecen los diferentes Conjuntos de Datos.</p> <p>Una vez integrado el inventario, la CIDEA deberá revisar que efectivamente el Inventario sea representativo y cumple con la intención fundamental de uso, que los temas identificados sean exhaustivos y sustantivos, en correspondencia a las funciones principales de la institución (según lo establecido en el paso 1). Esta sesión deberá ser documentada mediante una minuta debidamente firmada por los miembros de la CIDEA.</p>
<p>Presentar Inventario</p>	<p>Presentar el Inventario Institucional de Datos (ver formato de Hoja de Cálculo disponible en datos.gob.sv/plantilla/inventario.xls) junto con el acuerdo de la sesión de la CIDEA donde se formaliza el Inventario a la STPP.</p> <p>Es importante considerar que el Inventario deberá aprobarse dentro de los 70 días posteriores a la publicación de la Guía de Implementación de la Política de Datos Abiertos, por lo que se recomienda enviarlo junto al acuerdo para su aprobación con al menos 10 días antes de esta fecha.</p>

4. CREACIÓN DEL CATÁLOGO GUBERNAMENTAL DE DATOS ABIERTOS

La priorización de datos es importante para identificar qué datos tienen mayor potencial de generar impacto y por lo tanto deben ser publicados como Datos Abiertos. Esta priorización deberá hacerse con base en los datos que tengan más valor para la sociedad y los objetivos del Gobierno.

La priorización, que debe realizarse en 10 días hábiles, deberá incluir:

1. Todos los conjuntos de datos considerados como información de interés nacional.
2. Todos aquellos conjuntos de datos de carácter público que sean publicados en los sitios web y otras plataformas institucionales.
3. Los datos demandados por la ciudadanía mediante las solicitudes de acceso a la información más recurrentes.

4. Datos sobre los programas de la institución que ayuden a cumplir con los objetivos del Plan Quinquenal u otros compromisos internacionales que el Gobierno ha firmado; así como con las acciones relevantes de la institución.

Actividades a realizar por la institución

- a. Solicitar a la Oficina de Información y Respuesta un reporte de las solicitudes de acceso a la información pública más recurrentes para que sean incluidas en la primera publicación.
- b. Procurar la reunión de la CIDEA para priorizar los datos del Inventario Institucional de Datos según los objetivos de la institución y el Plan Quinquenal.

La STPP podrá hacer recomendaciones sobre la priorización de los datos cuando las entidades o dependencias cuenten con información que demuestre el cumplimiento de los objetivos del Plan Quinquenal de Desarrollo, inclusive cuando la misma esté relacionada con objetivos y metas de otras instituciones.

4.1 Plan de Publicación

Este Plan será publicado anualmente, en los primeros sesenta días hábiles del año, con el objetivo de dar a conocer las acciones que serán realizadas por cada institución en materia de datos abiertos. Como uno de sus componentes principales, el Plan de Publicación contendrá el calendario de publicación anual de los datos abiertos de la institución, esto con el objetivo de señalar al público las acciones y fechas propuestas para la publicación de Datos Abiertos.

Actividades a realizar por la institución:

- a. Preparar el calendario de publicación de acuerdo al formato disponible en <http://datos.gob.sv/plantillas/calendario.xls>
- b. Incluir el Plan de Publicación como parte del catálogo de datos de la Institución, siguiendo las instrucciones del Paso 8 de esta guía.

5. IDENTIFICAR Y RECOLECTAR DATOS

La recolección de datos debe incluir acciones que garanticen que los datos están completos. Estas acciones pueden requerir la colaboración entre distintas unidades de la misma institución o entre varias instituciones agrupadas según sus respectivas áreas de atención a la población, es decir, su Dominio de Integración.

El intercambio de datos entre entidades es necesario para que las instituciones de Gobierno obtengan el mejor provecho de ellos, en función de la mejora a la atención ciudadana, la producción de datos completos y su reutilización. Cuando una dependencia o entidad requiera datos como requisito para

satisfacer una petición de información de la población o para la prestación de un servicio, deberá solicitarlos en primera instancia a la dependencia o entidad encargada de generar ese dato. La transmisión electrónica de estos datos estará protegida usando firma simple en cumplimiento al artículo 33 de la Ley de Firma Electrónica.

El intercambio de datos administrativos está basado en una planificación coordinada entre las entidades de Gobierno. Esta planificación debe contemplar la:

1. Normativa de acceso y uso de los datos entre las instituciones;
2. La homologación y relación de procesos entre organizaciones;
3. La equivalencia semántica de los campos que se comparten y;
4. El acoplamiento técnico entre distintos sistemas de información.

La identificación y recolección de datos, que debe hacerse en un período de 45 días hábiles e incluye:

5.1. Normativa de acceso y uso de los datos entre las instituciones

Para generar la normativa de acceso y reutilización de los registros administrativos las dependencias y entidades a través de su CIDEA, en el ámbito de sus atribuciones, deberán coordinarse para:

- Definir y acordar, conjuntamente con las unidades administrativas responsables, los reglamentos que definan las reglas de acceso y el uso apropiado de los datos;
- Ajustar sus procedimientos y flujos de trabajo para reemplazar procesos manuales por procesos automatizados reutilizando datos de otras instituciones;
- Proponer e introducir las enmiendas a leyes y reglamentos que sean necesarios para aprovechar el intercambio de datos de forma electrónica y eliminar la documentación física en la entrega de servicios a la población.
- Desarrollar las competencias y habilidades de su personal para diseñar y utilizar los servicios de intercambio de datos entre entidades de gobierno.

5.2. La homologación y relación de procesos entre organizaciones

Las entidades a través de su CIDEA, en el ámbito de sus atribuciones, deberán:

- Adecuar los procesos relacionados con trámites y servicios que ofrecen a la población, previo análisis y simplificación de los procesos y procedimientos respectivos, para propiciar en el menor tiempo posible el intercambio de datos administrativos;
- Promover la digitalización de la información, la automatización de procesos y procedimientos, así como el establecimiento de servicios integrados;
- Revisar mejoras a los procedimientos y la captura de datos con la colaboración y participación de las demás entidades que usan esos datos;

- Compartir los flujos de intercambio de datos y acuerdos con todas las entidades dentro de un mismo dominio de integración.

5.3. La equivalencia semántica de los campos que se comparten

Las entidades a través de su CIDEA, en el ámbito de sus respectivas atribuciones, deberán:

- Establecer, con el apoyo de la STPP, definiciones semánticas homogéneas para facilitar el intercambio de datos, las cuales constituirán una base común de conocimiento dentro de un mismo dominio de integración de servicios.
- Otorgar un significado preciso y unívoco a los datos a ser compartidos de modo que puedan ser interpretados correctamente por cualquier sistema o aplicación;
- Establecer y mantener actualizado el modelo de relación y flujos de intercambio de datos.
- Establecer y publicar en el inventario de datos las normas de acceso y reglas de uso las cuales serán de aplicación obligatoria para las entidades de Gobierno que deseen consumir los datos;

5.4. El acoplamiento técnico entre distintos sistemas de información

El acoplamiento técnico posibilita el flujo seguro de datos entre instituciones para simplificar procesos y mejorar la calidad de los datos abiertos que se publiquen. Dicho acoplamiento técnico estará basado en recomendaciones técnicas emitidas por la STPP. Las entidades a través de su CIDEA, en el ámbito de sus atribuciones, deberán:

- Usar estándares abiertos o, a falta de éstos, estándares que sean de uso generalizado previa consulta con la STPP;
- Generar y poner a disposición de las demás entidades sus herramientas y soluciones para facilitar la reutilización de software aplicando la licencia de la PNDA (CC-BY-SA);
- Utilizar estándares internacionales y las mejores prácticas adoptadas por organismos internacionales para el intercambio seguro de datos.
- Establecer, cuando se trate de desarrollos contratados a terceros, los términos y condiciones que garanticen la portabilidad de las soluciones a través del tiempo.
- Implementar túneles cifrados TLS y otros mecanismos de seguridad requeridos por la CIDEA, con el apoyo de la STPP, para el acceso y protección de los datos administrativos.

6. PREPARAR LOS CONJUNTOS DE DATOS

Es importante estandarizar la manera en que los datos serán utilizados y consumidos por el público en general y el propio Gobierno. Al exportar los conjuntos de datos a formatos estructurados, se facilita el consumo e interpretación de la información por terceros, así como por máquinas.

Al utilizar formatos abiertos, se extiende la audiencia de usuarios que pueden aprovechar los datos, ya

que los formatos abiertos tienen su especificación o estándar abierto, que permite que sean interpretados por la mayoría de los sistemas computacionales.

Esta Guía pide que los datos se publiquen en formatos estructurados CSV. Sin embargo, para atender la diversidad de audiencias, sectores y casos de uso, se fomentará la publicación de los mismos conjuntos en diferentes **formatos estructurados de estándar abierto que resultan convenientes de acuerdo a la naturaleza de cada conjunto de datos**, como lo son XML, JSON, RDF, GeoJSON, KML, DBF, e inclusive propietarios como SHP y XLSX cuando la OIR establezca que este es el formato más solicitado por la ciudadanía. En general, se recomienda la publicación de datos en formatos altamente estructurados y abiertos; ver tabla 3 "Referencia de formatos, su nivel de apertura y estructura".

Se contará con 15 días hábiles para preparar los conjuntos de datos.

Tabla 3: Referencia de formatos, su nivel de apertura y estructura

Ejemplo	Notas	Estructura	Apertura	Tipo de datos	Formato
Población total por cada departamento. Número total de beneficiarios de un programa social por departamento.	Es un formato simple y universal, que estructura los datos en filas y columnas separadas por un carácter definido.	Alta	Alta	Tabulares	CSV (Obligatorio)
Número total de accidentes de carretera.	La estructura es un XML con una especificación abierta, lo cual permite que cualquier software pueda interpretarlo.	Alta	Media	Tabulares	XLSX
Número total de doctores y enfermeras en cada hospital.	Es un archivo binario y de formato propietario que restringe la utilización de este archivo por software de terceros.	Alta	Baja	Tabulares	XLS
Número total de escuelas primarias y secundarias en cada municipio. Producción de petróleo y gas por entidad. Valor de las exportaciones de crudo por cada destino geográfico.	Es un archivo que estructura los datos de forma simple. Aunque es uno de los archivos que contiene un SHP, es comúnmente utilizado de forma individual para publicar datos tabulares.	Alta	Alta	Tabulares	DBF
Listado de médicos y enfermeras autorizados para ejercer la profesión según universidad en la que estudiaron (nacional o internacional) Listado de beneficiarios de	Lenguaje de marcado desarrollado por la W3C para almacenar datos en forma legible y estructurada, que propone un estándar para el intercambio de información entre diferentes plataformas.	Alta	Alta	Estructurados	XML

Ejemplo	Notas	Estructura	Apertura	Tipo de datos	Formato
becas según centro de estudios (universidad nacional o internacional)	Los archivos con esta terminación, generalmente refieren a RDF/XML, que es la sintaxis XML. Otra sintaxis común es JSON-LD, que serializa RDF Utilizando JSON.	Alta	Alta	Estructurados	RDF
Información geoespacial de los Municipios	Es un estándar abierto aprobado por la OGC y connotación XML. Comúnmente son agrupados y comprimidos en formato KMZ.	Alta	Alta	Espaciales(vectorial)	KML
Información geoespacial de zonas urbanas y rurales	Es una interface estándar aprobada por la OGC para realizar peticiones de entrega datos en notación GML(basada a su vez en XML).	Alta	Alta	Espaciales (vectorial)	WFS
Total de familias beneficiadas por localidad.	Formato con notación XML, del estándar abierto OpenDocument.	Baja	Alta	Texto	ODT
Volumen de las exportaciones por semestre.	Ver notas del formato XLSX.	Baja	Media	Texto	DOCX
Información sobre unidades médicas y recursos en salud.	Este formato contiene los datos de manera desestructurada.	Baja	Media	Texto	PDF
Número total de afiliados al sistema del Seguro Social.					

Para cada tipo de conjunto de datos; sea tabular, vectorial, estructurado o de texto, existen una serie de formatos adecuados que se podrán elegir para su publicación, esto con el objetivo de poner los conjuntos de datos a disposición del público en los formatos más útiles.

El formato CSV será obligatorio para todos los conjuntos de datos, se podrán publicar formatos adicionales para cada conjunto, según los usos del público en general; por ejemplo, en el caso de datos tabulares hacer disponible la versión XLS de archivo junto con el CSV. Esto para asegurar la mayor inclusión posible.

Es importante asegurar la calidad del archivo exportado, para lo anterior, las Instituciones encargadas de la publicación deberán considerar lo siguiente:

6.1. Promover la interoperabilidad

- 6.1.1.** Evitar el uso de caracteres fuera del rango de ASCII6 (como los acentos) en el nombre del archivo y URL.
- 6.1.2** Utilizar codificaciones universales, como UTF-87, y en caso de utilizar otra codificación como ISO 8859-18 (latin-1), informar a los usuarios en los metadatos que acompañan el conjunto de datos.
- 6.1.3** Estructurar los campos de tiempo (fechas, horas y rangos temporales) utilizando ISO 8601 (p. ej. 2015-02-22).
- 6.1.4** Apegarse a las recomendaciones de la Dirección General de Estadística y Censos cuando se hable de regiones geopolíticas del país. P. ej., evitar "SS", y utilizar "02", "San Salvador", o "SS" uniformemente.

6.2 Hacer entendible los datos a los usuarios y usuarias

- 6.2.1.** Describir claramente los nombres de columnas, p. ej. "Fecha de solicitud" en lugar de "fech_sol". Se deberá acompañar el conjunto de datos con un diccionario de datos (ver ACTIVIDAD C: Documentar los conjuntos de datos).

6.3. Habilitar la legibilidad por máquinas

- 6.3.1.** Eliminar cálculos adicionales en formatos tabulares, p. ej. Una celda "Total" con la suma de una de las columnas.
- 6.3.2.** Remover imágenes, logotipos, celdas unidas, y cualquier otra decoración de los archivos; para facilitar su procesamiento, por lo que deberán ser planos.
- 6.3.3.** Los campos numéricos, incluyendo los monetarios y magnitudes, deben permanecer en un formato numérico de tipo entero o flotante. Es decir, evitar agregar símbolos monetarios o de unidades de medición, p. ej. "200 m2" o "\$20", haciendo explícitas las unidades en el nombre de la columna, o en una segunda columna. P. ej. "Monto en Dólares", "Distancia en KM".

6.4. Asegurar la consistencia y calidad de los datos

- 6.4.1.** Evitar tener valores de diferentes tipos (como texto y número) para una columna en diferentes registros o filas.
- 6.4.2.** Evitar datos duplicados.
- 6.4.3.** Eliminar redundancia en datos.
- 6.4.4.** Asegurar la representación de variables (unificar datos).

6.4.5. Evitar errores sintácticos.

6.5. Asegurar la normalización de los datos

6.5.1. Apegarse a las normas y catálogos nacionales oficiales o en su defecto a mejores prácticas nacionales o internacionales para la estructura de datos estadísticos y geográficos.

6.5.2. En caso de que los datos representan dimensiones físicas, deberán corresponder al Sistema Internacional de Unidades.

Tabla 4 : Ejemplos de un conjunto de datos sin errores

Turistas Extranjeros	Turistas Nacionales	Código Oficial	Departamento	Fecha
6193	45729	01	Santa Ana	2013-01
73386	238338	02	San Miguel	2013-01

Tabla 5 : Ejemplos de conjuntos de datos con errores

Turistas		Departamento	Fecha
Extranjeros	Nacionales		
6,193	45,729	Santa Ana	Ene 2013
73,386	238,338	SM	01/2013
			* pie de página

Actividades a realizar por parte de las instituciones:

1. Exportar los datos a archivos estáticos tabulares.
2. Convertir los archivos a formatos abiertos CSV.
3. Identificar si se requiere estructurar algún formato adicional que facilite su lectura y uso de acuerdo al tipo de conjunto de datos.
4. Asegurar la validez del formato y estructura del archivo.

7. PUBLICAR DATOS

Los conjuntos de datos deben estar disponibles para descarga directa en un URL único y evitando cualquier barrera de acceso, como el uso de contraseñas o interfaces de usuario. Debemos considerar que estos datos serán consumidos por máquinas, no sólo por personas.

Cada conjunto de datos debe estar disponible para ser descargado individualmente mediante un hipervínculo en el sistema de la institución. Para lo anterior, se recomienda cargar los conjuntos de Datos Abiertos en un servidor web, ya sea en el mismo Portal en línea con el que cuente la institución o en otros servidores de archivos propios o mediante servicios en la nube contratados con tercero.

Para facilitar la descarga, se podrán segmentar y comprimir los archivos mayores a 100 MB (utilizando formatos abiertos de compresión como gzip).

Estos hipervínculos serán utilizados posteriormente para referenciar los conjuntos de datos en el Catálogo Gubernamental de Datos Abiertos en el Portal - www.datos.gob.sv.

Para publicar los datos se contará con 15 días hábiles.

7.1 Documentar los conjuntos de datos de acuerdo al estándar DCAT

Utilizar datos no debería ser confuso para el usuario. Es importante explicar de dónde vienen los datos y compartir referencias adicionales para entender a fondo el origen de éstos. Existen estándares internacionales que sugieren descripciones de uso común y facilitan la documentación de datos (Dublin Core, etc.).

Todas estas descripciones se capturan en los metadatos con formato DCAT, conforme a la plantilla que se encuentra en la sección correspondiente del Portal de internet datos.gob.sv, que permiten identificar los datos abiertos de una organización, describirlos y hacerlos públicos en el Portal datos.gob.sv. Esta tarea de documentación debe ser liderada por el Administrador de Datos Abiertos.

La Ficha descriptiva de conjunto de datos (disponible en datos.gob.sv/plantillas/ficha.xls) contiene la información que deberá ser llenada de acuerdo a las siguientes descripciones para cada metadato:

Tabla 6 : Metadatos (descriptores) del conjunto de datos (dcat:Dataset)

Identificador único del conjunto de datos, utilizado para agrupar recursos dentro de éste, p. ej. "rezago-social", "unidades_medicinas", "adquisiciones". Utilizar caracteres ASCII (p. ej. sin acentos).	ds:identifier
Título descriptivo del conjunto de datos, p. ej. "Programa de fomento a la agricultura", "Vuelos comerciales".	ds:title
Una explicación de los datos, con suficiente detalle para que los usuarios puedan entender si es de su interés, p. ej. "Apoyos otorgados a través del programa Opciones Productivas, desglosado a nivel de localidad".	ds:description
Lista de términos clave separados por coma, que facilitarán al usuario la búsqueda del conjunto de datos. Es importante considerar el uso de términos no técnicos. p.ej. "salud, medicinas, compras, agricultura".	ds:keyword
Fecha y hora de la última modificación del conjunto de datos; en formato ISO8601 14, p. ej. "2014-05-27T01:42:05-05:00"	ds:modified

Nombre de la persona de contacto que atenderá dudas y comentarios sobre el conjunto de datos.	ds:contactPoint
Correo electrónico de contacto para atender dudas y comentarios sobre el conjunto de datos.	ds:mbox
La fecha o fechas que cubren los datos, p. ej. "2013", "2010/2012", "2014-01/2014-04". Si es un rango de fechas, deberán ordenarse ascendentemente.	ds:temporal
El espacio geográfico que cubre el conjunto de datos. Puede ser una región, el nombre de un lugar, un polígono o un cuadro delimitador de coordenadas geográficas (bounding box) en GML15. P. ej. "San Miguel", 009, http://www.geonames.org/3583462/departamento-de-san-miguel.html , o "32.71,-112.32 27.99, -118.45".	ds:spatial
Dirección electrónica para obtener mayor documentación o información sobre el conjunto de datos, como lo puede ser un manual, un sitio web, o un diccionario de datos. Este documento sirve como guía adicional para que el usuario entienda con mayor detalle los datos.	ds:landingPage
Frecuencia con la cual el conjunto de datos será publicado o actualizado, p. ej. "mensualmente".	ds:accrualPeriodicity

Tabla 7: Metadatos (descriptores) del recurso o descargable (dcat:Distribution)

La clave que identifica al conjunto de datos al que pertenece (y bajo el que se agrupa) este recurso. Ver ds:identifier.	ds:identifier
Título descriptivo del recurso o descargable, p. ej. "Asignación Presupuestaria del 2013", "Apoyos por municipio", "Apoyos por localidad".	rs:title
Ver ds: description. Esta explicación es adicional a la que existe en el conjunto de datos.	rs:description
Dirección electrónica (enlace) para la descarga del recurso.	rs:downloadURL
Formato de archivo del recurso a descargar, p. ej. "text/csv", "application/rss+xml". Este campo permite al usuario buscar conjuntos de datos por formato en www.datos.gob.mx .	rs:mediaType
El tamaño en bytes del recurso o descargable, p. ej. 3145728	rs:byteSize
Ver ds:temporal	rs:temporal
Ver ds:spatial	rs:spatial

El producto final de este paso será un archivo CSV con la documentación de todos los conjuntos de datos a publicar.

Para documentar metadatos usando DCAT debe usarse la Ficha descriptiva de conjunto de datos que se encuentra en la sección correspondiente del portal de internet datos.gob.sv/plantillas/ficha.xls.

7.2 Publicar el Catálogo Gubernamental de Datos Abiertos en el Portal datos.gob.sv

La persona administradora de datos abiertos será el encargado de generar y publicar el catálogo

generado en el paso anterior.

Al cargar el catálogo de datos de la institución en formato CSV, se transformará y traducirá automáticamente al estándar DCAT en formato JSON y los conjuntos de datos aparecerán en el Portal datos.gob.sv. Además, se registrará un histórico de catálogos, y se proveerá de ayuda a los funcionarios responsables.

Las mejores prácticas de la comunidad internacional de Datos Abiertos apuntan a que un catálogo central de datos abiertos facilita la búsqueda y acceso para los ciudadanos.

Al publicar los datos abiertos en el Portal <http://datos.gob.sv>, éstos son asociados inmediatamente a los términos de la licencia CC-BY-SA, con los cuales se da la certidumbre de que cualquier persona puede usar los datos para cualquier fin legal que se proponga, con el único requisito de citar la fuente de origen y compartirlo con la misma licencia.

Estos términos promueven el uso, reutilización y redistribución de los conjuntos de datos abiertos, tomando en cuenta los siguientes derechos y deberes:

Derechos

- Copiar del conjunto de datos y poder distribuirlos libremente;
- Difundir el conjunto de datos;
- Publicar el conjunto de datos en otros Portales;
- Reutilizar el conjunto de datos;
- Extraer el contenido del conjunto de datos para su uso libre;
- Explotar comercialmente el conjunto de datos; y
- Crear conjuntos de datos derivados del conjunto de datos original.

Deberes

- Citar, de forma clara y fácil para el usuario, la fuente de origen de donde obtuvo el conjunto de datos de la siguiente forma: Nombre del conjunto de datos, Nombre de la entidad, Link de publicación del conjunto de datos y la fecha de la consulta [AAAA-MM-DD].
- No utilizar los datos de mala fe para confundir o estafar a los usuarios.
- No aparentar que el Portal particular representa una entidad gubernamental o que el mismo está avalado así.
- Compartir la información con la misma licencia que se encuentra en el Portal nacional de datos datos.gob.sv. Para esto deberá utilizar el logo o un equivalente. Para más información sobre la licencia puede visitar <http://datos.gob.sv/licencia>.

En el Portal datos.gob.sv se concentrarán una gran variedad de conjuntos de datos de diferentes instituciones, Gobiernos locales y municipales, así como de órganos autónomos. Al publicar en datos.gob.sv la institución se suma a la Política de Datos Abiertos oficialmente.

La CIDEA designará el servidor público encargado de cargar los datos en el Portal de datos abiertos.

8. COMUNICAR Y PROMOVER

Cada institución por medio de la CIDEA, y el Gobierno por medio de la Dirección de Gobierno Electrónico, deberán trabajar para la creación del ecosistema que genere valor agregado a la publicación de los datos.

Los objetivos de esta promoción son:

- 8.1 Fomentar la oferta por medio de la liberación de la mayor cantidad de datasets publicados.
- 8.2 Fomentar la demanda por medio de la reutilización y consumo de datos que otras instituciones gubernamentales han publicado.
- 8.3 Comunicar la publicación de nuevos conjuntos de datos, así como el uso que se estén dando a datos publicados por otras instituciones.
- 8.4 La estrategia de promoción deberá estar acompañada de un proceso de información, educación y capacitación a los servidores públicos orientado a promover una cultura de apertura por defecto y que los datos sean lo más rico y completos posibles.

De la misma forma se deberá promover el uso especializado de los datos por medio de:

- Festivales de creación colaborativa de soluciones a problemáticas públicas
- Premios a aplicaciones especializadas que se deriven del uso del Portal datos.gob.sv
- Comités interinstitucionales e intersectoriales de uso y transferencia de datos entre instituciones públicas.
- Convenios con aliados estratégicos para llevar los datos e información a la población que no tiene acceso a internet y recursos tecnológicos.

Una vez finalizado el proceso hasta el paso 7.2, se inicia la fase de comunicación y promoción, lo que significa no más de seis meses después de aprobada la PNDA.

9. GLOSARIO

ASCII: es un código de caracteres basado en el alfabeto latino, comúnmente utilizados por sistemas computacionales para representar un mínimo conjunto de caracteres derivados del inglés moderno. La mayoría de los sistemas informáticos utilizan y entienden este código.

Catálogo de información reutilizable: Listado de conjuntos de datos que son generados o administrados por una institución pública que puede ser reutilizada por otras unidades administrativas a su interior o por otra institución pública.

Catálogo gubernamental de Datos Abiertos: el inventario único de los conjuntos de datos que producen las entidades del Órgano Ejecutivo, consolidado por la STPP, puestos a disposición de la población en el Portal Nacional de Datos Abiertos (datos.gob.sv).

CIDEA: Comisión institucional de Datos Abiertos

Conjunto de datos: Serie de datos vinculados entre sí y agrupados dentro de una misma temática, para que puedan ser procesados apropiadamente para obtener información.

CSV: el tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas (o punto y coma en donde la coma es el separador decimal y las filas por saltos de línea).

Dato: Es una cifra, letra, palabra o valor que se almacena digitalmente y que representa un atributo o variable. Por su contenido, al verlo de forma aislada, no puede ser interpretado o analizado por las máquinas o seres humanos.

Datos Abiertos: datos digitales de carácter público, accesibles en línea que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona.

DCAT: el Vocabulario de Catálogos de Datos es un estándar internacional diseñado para facilitar la interoperabilidad entre los catálogos de datos publicados en la red; para mayor información consultar <http://www.w3.org/TR/vocab-dcat/>.

Dominio de Integración: La agrupación por sectores de acuerdo a áreas de atención a la población que permite identificar conjuntos de datos reutilizables para la creación de servicios integrados.

Entidad: Institución que forma parte del Órgano Ejecutivo, entendiéndose con ello a Ministerios, Secretarías de la Presidencia, Instituciones autónomas y Sociedades de economía mixta y personas jurídicas que manejan recursos o información pública para su funcionamiento.

Estándar abierto: Especificaciones cuya utilización está disponible de manera gratuita, de fácil acceso, y que su uso y aplicación no es condicionado al pago de un derecho de propiedad intelectual o industrial;

Firma electrónica simple: Mecanismo de seguridad para intercambiar datos entre una o más instituciones donde la misma institución que genera los datos certifica la validez de la firma.

Formato abierto: Características técnicas y de presentación que cumplen los datos y por las cuales se pueden almacenar en digital. Sus especificaciones siguen un estándar abierto y están disponibles públicamente.

Formatos Tabulares: Formatos que presentan los datos en forma de tabla con filas y columnas.

Fuente de origen: Entidad o unidad dentro de la entidad que produce, modifica y resguarda la información, de acuerdo a su competencia.

Gobierno Electrónico: Aplicación de las Tecnologías de la Información y Comunicación en el marco de la transformación del Estado para acercar el Gobierno a la sociedad aplicando buenas prácticas, políticas y estándares que aseguren una gestión pública eficaz y la entrega de servicios de calidad.

HTML (Lenguaje de marcas de hipertexto): el lenguaje de marcado para la elaboración de páginas web. Define una estructura básica y un código (denominado código HTML) para la definición de contenido de una página web, como texto, imágenes, entre otros.

Información: Los datos interrelacionados con otros datos e interpretados por personas la generación de informes o documentos (sin importar su formato) para que sean entendidos por otras personas.

Inventario institucional: Listado de todos los conjuntos de datos que ha sido generada, consolidada y filtrada por la institución pública según sus objetivos misionales, ley de creación, plan operativo y plan quinquenal de desarrollo.

JSON (JavaScript Object Notation): Formato ligero para el intercambio de datos que parte de un subconjunto de la notación literal de objetos de JavaScript. Dada su simplicidad, su uso se ha generalizado y popularizado en la Web.

Metadatos: Lista de información que conforma la ficha descriptiva de cada conjunto de datos a ser publicado como parte del Catálogo Gubernamental de Datos Abiertos.

Neutralidad tecnológica: La opción de elegir la alternativa tecnológica más adecuada a las necesidades y contexto de cada institución, favoreciendo la entrega de soluciones sobre el uso de tecnologías específicas y eliminando barreras de acceso, modificación y mejora de los sistemas.

Nivel de servicio: El conjunto de métricas que establecen los niveles aceptables de calidad, disponibilidad, velocidad, etc. con los que operará un sistema o servicio.

OIR: Oficina de Información y Respuesta, unidad encargada de la implementación y ejecución de la Ley de Acceso a la Información Pública.

PDF: (*portable document format*): Formato de almacenamiento de documentos digitales independiente de plataformas de software o hardware.

Redistribución: Los permisos que facilitan la transferencia, sin que desmejore su calidad al momento por parte de las personas que lo han utilizado para ponerlos nuevamente a disposición de todos.

Reutilización: Se refiere a que los datos deben ser provistos bajo términos y licencias que permitan su utilización, trabajo e incluso integración con otros conjuntos de datos.

Servicios integrados: Son aquéllos en que para su prestación intervienen dos o más instituciones utilizando servicios comunes de intercambio de datos.

Servicio de intercambio de datos: Conjunto de protocolos y estándares que permiten intercambiar datos entre sistemas de diferentes instituciones, de forma automática, con independencia de las tecnologías en la que fueron desarrollados;

TXT: Archivo informático compuesto únicamente por texto sin formato.

URL: Secuencia de caracteres, de acuerdo a un formato estándar, que se usa para nombrar recursos en Internet para su localización o identificación, como por ejemplo documentos textuales, imágenes, páginas Web. La URL del sitio de datos abiertos es <http://datos.gob.sv>

XLS: Extensión de archivo por defecto del formato propietario de Excel.

XLSX: Extensión de archivo de Excel, en su versión de formato abierto. Ésta cumple con la característica de seguir el estándar abierto Office Open XML.

XML: (lenguaje de marcas extensible): el lenguaje de marcas utilizado para almacenar datos en forma legible. Permite definir la gramática de lenguajes específicos (de la misma manera que HTML) para estructurar documentos grandes. Da soporte a bases de datos, siendo útil cuando varias aplicaciones deben comunicarse entre sí o integrar información.



GOBIERNO
DE EL SALVADOR