

Guía rápida para apertura de Datos Gubernamentales

I. PRELIMINARES

- 1.1 Introducción
- 1.2 Consideraciones
- 1.3 Del proceso de implementación para la apertura de datos

II. ACCIONES PREVIAS O PREPARATORIAS

- 2.1 Conformación del equipo de trabajo
- 2.2 Asignación de roles del equipo de trabajo

III. METODOLOGÍA DE APERTURA DE DATOS

- 3.1 Apertura por defecto
- 3.2 Apertura focalizada y gradual

FASE 1: Identificación de la información

FASE 2: Análisis de la información

FASE 3: Priorización de la información

FASE 4: Documentación de los datos (metadatos)

FASE 5: Preparación y publicación

GUIA RAPIDA PARA APERTURA DE DATOS GUBERNAMENTALES

I. PRELIMINARES

1.1 INTRODUCCIÓN

En el marco de los compromisos establecidos en el Plan de Acción de Gobierno Abierto aprobado mediante Resolución Ministerial N°176-2015-PCM, referido a la Mejora de la Prestación de los Servicios se promueve la apertura de los datos de las entidades públicas bajo el enfoque de una gestión del cambio considerado en la Política Nacional de Modernización de la Gestión Pública.

Las experiencias del nivel nacional y subnacional y su proceso de apertura de datos en el aspecto operativo, han sido contempladas en el presente instrumento para promover y fortalecer el proceso de apertura de datos gubernamentales en la administración pública.

La presente guía está dirigida a los servidores públicos que formarán parte del equipo de trabajo encargado de implementar el proceso de apertura de datos públicos en su Entidad, el cual constituye un nuevo servicio al ciudadano, que busca que todos aquellos datos generados por la administración pública, sean colocados en formatos estándar y abiertos para que los ciudadanos puedan acceder a ellos, de forma libre y sin restricciones con el fin de usarlos o reutilizarlos aprovechando el potencial de la innovación para mejorar los servicios, y aportar al desarrollo social y económico.

Los datos abiertos, en el marco de la política nacional de modernización y plan de acción de gobierno abierto, constituye una nueva forma de implicar a la ciudadanía en los asuntos públicos y ciclo de políticas públicas, incrementar la transparencia y rendición de cuentas, y conducir a mayor eficiencia gubernamental; impactando finalmente en el desarrollo social y económico del país y un mejor gobierno.

Metodológicamente esta guía constituye un instrumento de apoyo al proceso operativo para la apertura de los datos gubernamentales y con este propósito es presentada de manera estructurada en torno al ciclo de vida de los datos abiertos, abarcando las diferentes fases desde la identificación de la información con que cuenta la institución, el análisis y priorización, la elaboración de un catálogo de datos abiertos y su publicación en repositorios locales y en el Portal Nacional de Datos Abiertos.

Con la finalidad de organizar y gestionar un Catálogo Nacional de Datos Abiertos Gubernamentales de todo el Sector Público, la SGP y ONGEI han considerado la implementación de un Portal Nacional de Datos Abiertos (www.datosabiertos.gob.pe), el cual permitirá a todas las Entidades de la administración pública registrar la información (metadatos) de los conjuntos de datos, convirtiéndose en un punto de acceso para todos aquellos interesados en la reutilización de la información del sector público.

Cabe precisar que, cada Entidad Pública de acuerdo a la disponibilidad de recursos, condiciones técnicas externas, nivel de madurez tecnológico, podrán optar por implementar su propio portal de datos abiertos e incluirlo en su web institucional, considerando los lineamientos estratégicos y estándares establecidos por la Secretaría de Gestión Pública y la Oficina Nacional de Gobierno Electrónico referidos al Portal Nacional de Datos Abiertos.

1.2 CONSIDERACIONES

CONCEPTO DE DATOS ABIERTOS

Son todos aquellos datos generados por las Entidades de la Administración Pública que no han sido procesados y que se encuentran en formatos estándar e interoperables, de fácil acceso, que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que puedan reutilizarlos y crear información a partir de los mismos.

APERTURA DE INFORMACIÓN PÚBLICA

De acuerdo al principio de máxima publicidad establecido en la Ley de Transparencia y Acceso a la Información Pública (LTAIP), toda información que posea el Estado se presume pública, salvo las excepciones establecidas en la norma, debiéndose tomar en cuenta la Ley de protección de Datos Personales así como otras normas especiales complementarias relacionadas a la materia.

Toda aquella información contenida en las bases de datos de las Entidades de la Administración Pública es de carácter pública y como tal debe ser puesto a disposición de la ciudadanía como Datos Abiertos Gubernamentales.

DE LOS DATASETS O CONJUNTOS DE DATOS

El término dataset o conjunto de datos hace referencia a la categorización de los datos públicos en catálogos de datos, se organizan en datasets para ser fácilmente indexados y localizados. Por ello, se utilizan campos que definen el grupo de datos como la descripción, frecuencia de actualización, formato, licencia de uso, entre otros.

DE LOS METADATOS

Los metadatos son campos o “datos que describen otros datos” que ofrecen al usuario información suficiente para procesar y entender los datos. Estos campos pueden variar en detalle, desde descripciones muy básicas como explicar el tema general de la base de datos, hasta proveer detalle semántico para permitir un alto grado de legibilidad para máquinas; lograr esto último incrementa en gran manera la apertura y utilidad de los datos

PRINCIPIOS DE LOS DATOS ABIERTOS

- | | |
|-------------------------------------|---|
| 1. Abiertos por Defecto | 4. Comparable e interoperable |
| 2. Oportunos y comprensibles | 5. Mejora de la gobernanza y participación ciudadana |
| 3. Accesible y utilizable | 6. Datos abiertos para la innovación |

CARACTERÍSTICAS DE LOS DATOS ABIERTOS:

1. Completos

Todos los datos públicos están disponibles, sin restricciones de privacidad, seguridad o privilegio.

2. Primarios

Los datos son obtenidos en la fuente, con el mayor nivel posible de granularidad, sin ser modificados ni agrupados. Que utilicen estándares abiertos, procesables por máquinas y puedan ser descargables y operados con los requerimientos tecnológicos mínimos.

3. Actuales y permanentes

Los datos quedan disponibles tan pronto como sea necesario para preservar su valor.

4. Accesibles

Los datos quedan disponibles para la mayor cantidad posible de usuarios y propósitos.

5. Procesables

Los datos están razonablemente estructurados, de forma de permitir su procesamiento automático.

6. No discriminatorios

Los datos están disponibles para todos, sin requisitos de registro.

7. No propietarios

Los datos están disponibles en un formato sobre el que ninguna entidad tiene control exclusivo.

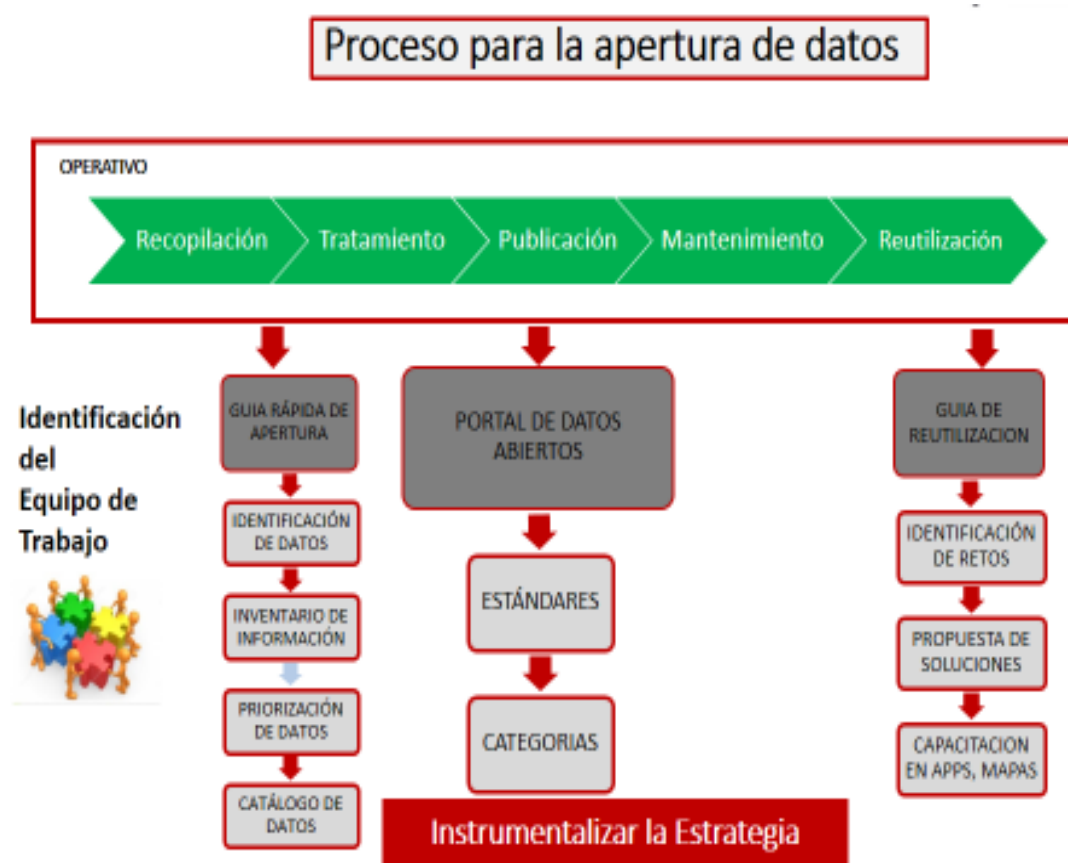
8. Sin licencia

Los datos no están sujetos a ningún tipo de regulación de derechos, patentes o registros de marca. Se podrán permitir restricciones razonables de privacidad, seguridad o privilegio. Se evitará costos de utilización para evitar barreras al uso por parte de la población

1.3 DEL PROCESO DE IMPLEMENTACIÓN PARA LA APERTURA DE DATOS

Bajo el enfoque de procesos se consideran cinco etapas con sus respectivos instrumentos del proceso de implementación para la apertura de datos, conforme al siguiente gráfico:

Gráfico 1: Proceso Operativo de Apertura de Datos



II. ACCIONES PREVIAS O PREPARATORIAS

Con el objetivo de garantizar la gestión del nuevo servicio de datos abiertos se ha considerado las siguientes acciones:

2.1 Conformación del equipo de trabajo

El Equipo de Trabajo debe estar integrado por funcionarios de la Alta Dirección, Tecnologías de la Información, Generadoras de Información, Especialistas en temas legales. Asimismo el equipo de trabajo puede estar integrado por personal del área de Sistemas y Comunicaciones, a fin de garantizar el éxito de la implementación, teniendo como propósito:

- **Contar con el patrocinio de la Alta Dirección, quien a su vez deberá designar a los funcionarios responsables que conformarán el equipo de trabajo.**
- **Garantizar que todo el equipo asuma un compromiso real sobre sus responsabilidades, que el proyecto sea conocido y comprendido por los participantes.**

- **Establecimiento de bases de colaboración;** el compromiso del personal de las áreas usuarias de la institución será un aspecto fundamental del proyecto ya que su éxito depende en gran medida del compromiso y participación activa de dicho personal.

2.2 Asignación de roles del Equipo de trabajo

Puesto	Rol	Integrante	Órgano
Coordinador del Proyecto de DAG	Supervisar e incentivar las acciones para cumplir con la implementación de Datos Abiertos	Asesor de la Alta Dirección o representante	Alta Dirección
Jefe o Gestor del Proyecto de DAG	Coordinar la creación del inventario de datos, priorización de datos, creación de datasets y publicación de los mismos. Implementar los procesos para que los datos se publiquen en el sitio institucional y en el portal nacional datos.gob.pe .	Jefe de TI o similar	Informática
Áreas generadoras de datos	Proveer la información, y la información acerca de ella (meta-información) Facilitar la información necesaria y la ejecución de actividades de inventario de datos, priorización y documentación de datasets. Participar activamente en las reuniones de trabajo.	Jefe de área o similar	Áreas institucionales
Especialista Legal	Velar por el cumplimiento de la normatividad vigente en materia de Transparencia, Acceso a la Información Pública, Protección de Datos Personales e institucional y normas conexas.	Jefe de área o similar	Asesoría Jurídica
Especialistas Informáticos	Efectuar la ejecución y aporte técnico especializado en la preparación de las bases de datos , la transformación en formatos abiertos y su publicación en un portal de datos abiertos (carga, portal, APIs)	Administrador de BD y Sistemas	Informático
Especialista en comunicación	Efectuar la ejecución y aporte especializado en estrategias y acciones de promoción y difusión del nuevo servicio de DAG.	Jefe de área o similar	Comunicaciones

III. Metodología de apertura de datos

3.1 Apertura por defecto

La implementación del servicio de Datos Abiertos Gubernamentales se realizará por defecto y de manera inmediata para la información gestionada directamente por la Oficina Nacional de Gobierno Electrónico (Portal de Transparencia Estándar, Portal de Servicios al Ciudadano y Empresas, Portal GEOIDEP, Portal del Estado Peruano, entre otros).

Gráfico 2: Apertura por defecto



3.2 Proceso de apertura focalizada y gradual

El resto de información de interés ciudadano e institucional debe ser abierta de forma focalizada teniendo en cuenta su potencial para generar valor social y económico, y de incrementar los niveles de transparencia; y de forma gradual, teniendo en cuenta las restricciones de índole técnico – legal que puedan afectar su pronta liberación.

Para tal efecto, se deberá aplicar los siguientes lineamientos de mejora continua, adoptando por ello la aplicación de una metodología del tipo PHVA (Planificar, Hacer, Verificar, Actuar) aplicada al proceso operativo de los datos abiertos.

La Guía rápida de apertura de datos aplica hasta la etapa de publicación y la metodología a emplearse se presenta estructurada en 5 fases:

- i. Identificación de la información pública.
- ii. Análisis de la información pública.
- iii. Priorización de la información pública.
- iv. Documentación de los datos.
- v. Preparación y publicación

Gráfico 3: Fases para la Apertura de Datos



FASE 1 IDENTIFICACIÓN DE LA INFORMACIÓN PÚBLICA

Objetivo: Identificar y caracterizar la información con que cuenta la Entidad.

Resultado: Inventario general de información institucional.

Actividades:

1) Identificación de la información bajo responsabilidad de la Entidad

Las áreas generadoras de datos elaborarán un listado de la información que producen y/o administran en cumplimiento de sus funciones (Formato Fase 1 de identificación de información) y lo reportarán al Gestor de datos abiertos.

2) Consolidación del Inventario de información institucional

El Gestor de datos abiertos consolidará la información en un solo documento para su posterior análisis.

Participantes: Gestor de datos abiertos, Áreas generadoras de datos.

Recursos: Formato de identificación de información pública, categorías generales.

Formato Fase 1: Identificación de información pública

Institución:

Área que reporta:

Responsable:

Contacto:

Item	Nombre	Descripción	Responsable	Medio	Formato	Frecuencia	Categoría

CATEGORÍAS GENERALES

CATEGORÍA	SUB CATEGORÍAS
Educación (promoviendo el acceso, información e inclusión)	<ul style="list-style-type: none"> • Acceso a la educación • Información y toma de decisiones en educación • Inclusión en educación
Salud (promoviendo el acceso, calidad y prevención de enfermedades)	<ul style="list-style-type: none"> • Acceso a la salud • Calidad en la provisión de servicios de salud • Prevención de enfermedades
Transporte (promoviendo el acceso, eficiencia y seguridad del transporte)	<ul style="list-style-type: none"> • Acceso al transporte • Eficiencia del transporte • Seguridad del transporte
Medio Ambiente (Enfrentando el cambio climático, la contaminación y sostenibilidad)	<ul style="list-style-type: none"> • Preservación del medio ambiente • Cambio climático y desastres naturales • Contaminación y sostenibilidad
Sanidad (Mejorando el acceso, gestión de desperdicios y reciclaje)	<ul style="list-style-type: none"> • Acceso a sanidad • Manejo de desperdicios • Reciclaje
Gobernabilidad (Promoviendo la transparencia y acceso a la información, seguridad y justicia)	<ul style="list-style-type: none"> • Transparencia y acceso a la información • Seguridad ciudadana • Justicia
Prosperidad económica (Promoviendo la innovación, inversión y empleo)	<ul style="list-style-type: none"> • Innovación • Inversión • Empleo
Sociedad de la Información y el Conocimiento (Promoviendo el acceso a las TIC, infraestructura de datos, y el desarrollo de capacidades)	<ul style="list-style-type: none"> • Acceso a las TIC • Infraestructura de datos espaciales • Desarrollo de capacidades
Otros	

El Inventario de información Institucional debe ser actualizado permanentemente. Los datos también constituyen un activo de información dentro del Sistema de Gestión de Seguridad de la Información de la Entidad.

FASE 2 ANÁLISIS DE LA INFORMACIÓN PÚBLICA

Objetivo: Clasificar la información susceptible de ser publicable y aquella que presenta algún tipo de restricción.

Resultado: Inventario de información institucional clasificada (información publicable, no publicable, y las restricciones identificadas).

Actividades:

1) Identificación de la información publicable y no publicable, restricciones existentes.

El Especialista legal evaluará la información para determinar de acuerdo con los criterios jurídicos si existe alguna restricción legal para ser considerada como no publicable, anotando su restricción en el Formato Fase 2.

De la misma forma, el personal de las Áreas generadoras debe identificar las restricciones de naturaleza técnica que pueden limitar la liberación de los datos, anotando su restricción en el Formato Fase 2.

2) Preparación del Inventario de información institucional.

El Gestor de datos abiertos consolidará la información resultado del análisis en un solo documento para su posterior priorización.

Participantes: Gestor de datos abiertos, Especialista legal, Áreas generadoras de datos.

Recursos: Formato ampliado para análisis de información; Normatividad vigente (Ley de transparencia, Ley de protección de datos personales, otras; Inventario de información pública.

Formato Fase 2: Análisis de información pública

Institución:
Área que analiza:
Responsable:
Contacto:

Item	Nombre	Descripción	Responsable	Publicable (S/N)	Restricción

Es imprescindible que la institución, antes de publicar o entregar los datos, efectúe el análisis y clasificación de la misma, para verificar que información puede ser publicada, y cual no. **La información puede ser transformada siguiendo algunas técnicas de anonimización que permitan respetar la privacidad, de ser el caso.**

FAS E 3 PRIORIZACIÓN DE LA INFORMACIÓN PÚBLICA

Objetivo: Definir los conjuntos de datos y el cronograma de publicación, teniendo en cuenta que se trata de un proceso permanente y gradual.

Resultado: Catálogo de datos a publicar según criterios de impacto y dificultad.

Actividades:

1) Revisión de los criterios de evaluación.

El equipo de trabajo debe conocer los criterios de evaluación de impacto y dificultad y el procedimiento a seguir.

2) Calificación de la información identificada

El equipo de trabajo reunido efectuará la calificación de la información identificada en el inventario de información institucional

3) Ordenamiento de los conjuntos de datos que conformarán el catálogo de datos

El equipo de trabajo procederá a ordenar los conjuntos de datos en grupos de corto, mediano y largo plazo.

Participantes: Todos los integrantes del equipo reunidos.

Recursos: Formato para priorización de los datos; criterios de evaluación y procedimiento para priorización de la información.

Formato Fase 3: Priorización de información pública

Item	Nombre	Impacto			Dificultad			Total Impacto	Total Dificultad	Plazo		
		i1	i2	i3	d1	d2	d3			Corto	Mediano	Largo

Cabe resaltar la importancia de iniciar la apertura de datos de forma focalizada y gradual, considerando aquellos conjuntos de datos que produzcan un mayor impacto e impliquen menor

dificultad, con ello se asegura evitar controversia y complejidad que podría retrasar el proyecto. Adicionalmente, puede existir variabilidad en cuanto a los criterios de priorización que puedan ser utilizados por las Entidades. Por ejemplo, en algunos casos se considera la viabilidad política, técnica e institucional.

Procedimiento de priorización

Paso 1: Aplique los siguientes criterios de evaluación a cada conjunto de datos.

De IMPACTO (i)	De DIFICULTAD (d)
i1: la información genera valor para la economía: <ul style="list-style-type: none"> • Crecimiento y competitividad (negocios, empleo, etc.) • Innovación 	d1: dificultad para extraer la información de la fuente (archivo, sistema)
i2: la información permite crear valor social <ul style="list-style-type: none"> • Empoderamiento ciudadano • Control y auditoria social • Participación y compromiso público 	d2: dificultad técnica u operativa para poder publicar la información en formato abierto
i3: la información permite crear valor para la gobernanza pública: <ul style="list-style-type: none"> • Rendición de cuentas • Transparencia • Eficiencia • Anti-corrupción 	d3: No se dan las condiciones en la información para cumplir con los 8 principios de los datos abiertos (en general su calidad)

Paso 2: Asigne un valor por cada criterio de evaluación al conjunto de datos.

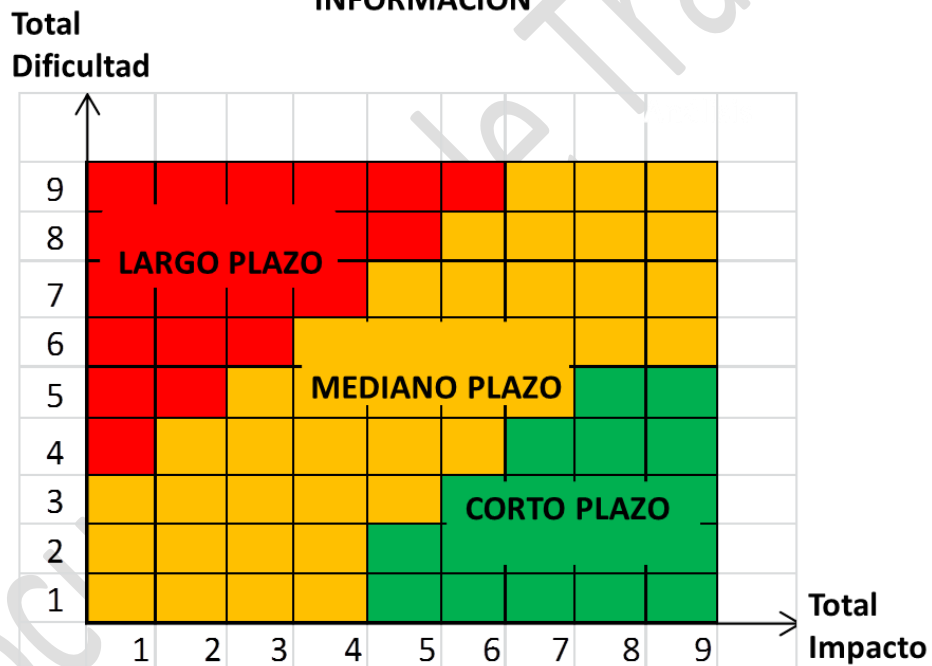
IMPACTO		DIFICULTAD	
Condición presentada	Puntaje	Condición presentada	Puntaje
Alto impacto	3	Mucha dificultad	3
Mediano impacto	2	Mediana dificultad	2
Bajo impacto	1	Poca dificultad	1

Paso 3: Totalice los valores asignados y grafique los totales para determinar el momento de la publicación del conjunto de datos.

$$X = \text{Total Impacto} = \sum i_1 + i_2 + i_3$$

$$Y = \text{Total Dificultad} = \sum d_1 + d_2 + d_3$$

Gráfico 4: PRIORIZACIÓN DE LA INFORMACIÓN



Objetivo: Documentar la información utilizando un estándar de metadatos para que los usuarios puedan entender y procesar los datos.

Resultado: Conjuntos de datos con metadatos definidos

Actividades:

1) Definición de los metadatos del conjunto de datos

Tomar conocimiento de la estructura y descripción **establecida por la ONGEI** (comunes y específicos) de los conjuntos de datos.

2) Consignar los metadatos de cada conjunto de datos.

Lleve un registro de los metadatos (comunes y específicos) definidos de cada conjunto de datos.

Participantes: Gestor de datos abiertos, especialista informático y áreas generadoras de datos.

Recursos: Formato en hoja electrónica: Descripción de metadatos.

El estándar de metadatos para los datos abiertos se basa en el ISO 15836, adicionalmente se debe tener en consideración otros estándares internacionales:

- Norma ISO 15836 – 2003 Iniciativa Dublin_Core_Metadata (DCMI), establece 15 elementos básicos y obligatorios para describir cualquier objeto de información, están referidos al contenido, propiedad intelectual e instanciación.
- Norma ISO 19139 – Geographic Information – Metadata – XML schema implementation
- Data Catalog Vocabulary (DCAT), vocabulario RDF para facilitar la interoperabilidad entre catálogos de datos publicados en la Web. W3C.
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), protocolo para la recolección de registros que contienen los metadatos de los repositorios de las entidades que producen datos abiertos.

Los **Metadatos son los datos de los datos que facilitan las búsquedas en el Catálogo de Datos**. Asimismo, el DCAT proporciona una arquitectura de catálogo de datos que facilita el acceso a otros catálogos en la Web.

Estructura y Descripción de Metadatos

Tipología	Metadato	Descripción	Tipos de datos
Metadatos e información de identificación	Título	Nombre concreto del dataset	Texto
	Descripción	Descripción/ resumen del dataset	Texto
	Palabras clave	Palabras clave para ser identificados en la búsqueda	Texto separado por comas
	Fecha creación		Fecha (AAA-MM_DD)
	Fecha Publicación	Fecha de la carga del dataset en el catálogo de datos	Fecha (AAA-MM_DD)
	Fecha Modificación	Fecha de última actualización del dataset en el catálogo de datos	Fecha (AAA-MM_DD)
	URL	URL en el que el dataset puede ser descargado	Texto
	Ruta del Diccionario de Datos	URL de acceso al esquema, diccionario de datos o descripción de los datos y la estructura del dataset.	Texto
	Ruta del XLS	URL en el que el dataset puede ser descargado	Texto
	Ruta del CSV	URL en el que el dataset puede ser descargado	Texto
	Ruta del XML	URL en el que el dataset puede ser descargado	Texto
	Ruta del GML	URL en el que el dataset puede ser descargado	Texto
	Ruta del JPG	URL en el que el dataset puede ser descargado	Texto
	Idioma		
Versión			
Punto de contacto	Nombre del área	Nombre del área que produce o genera el dataset	Texto
	Nombre del contacto	Nombre de la persona que administra el dataset	Texto
	Email del contacto	Email de la persona que administra el dataset	Texto

Objetivo: Preparar los conjuntos de datos y publicar en el Portal Nacional de Datos Abiertos.

Resultado: Portal Nacional de Datos Abiertos actualizado

Actividades:

1) Preparación de los conjuntos de datos

- Mantenga una copia de respaldo actualizada de la información o base de datos (backup) y trabaje sobre una copia de la información.
- Realice una anonimización de datos sensibles, de ser el caso.
- Elija el formato para la publicación de los datos (CSV, JSON, YAML, XML, otros)
- Extraiga la información o exporte a un archivo de formato abierto.

2) Selección del tipo de licencia abierta disponible

Las licencias que permiten el uso/reutilización liberando de responsabilidad a las entidades y personas que publican son:

- Public Domain Dedication and License (PDDL)
- Attribution License (ODC-By)
- Open Database License (ODC-ODbL)

3) Publicación de los conjuntos de datos en una herramienta de soporte o repositorio local.

La entidad de acuerdo a sus recursos puede optar por el uso de Infraestructura propia implementando alguna plataforma o herramienta disponible o contratando un servicio de terceros (en la nube).

4) Publicación de los conjuntos de datos en el portal Nacional DAG

El Portal Nacional (datos.gob.pe) a cargo de la ONGEI.

5) Difusión de los datos abiertos publicados

El especialista de comunicación.

6) Mantenimiento de los conjuntos de datos abiertos

Publicación de los Datos Abiertos Gubernamentales de la entidad en su plataforma de Datos Abiertos y en el Portal Nacional de Datos Abiertos Gubernamentales.

Participantes: Gestor de datos abiertos, Especialista Informático, Especialista comunicacional.

Recursos: Acceso al Portal Nacional de Datos Abiertos (datos.gob.pe) y/o portal institucional.

Esta fase corresponde a los procesos técnicos de preparación, carga y publicación de los datos abiertos en la plataforma tecnológica a la cual accederán los ciudadanos (infomediarios) vía web.

Anexo

Formato de Inventario de Información Pública – Catálogo de Datos Abiertos

Institución:

Área que reporta:

Responsable:

Contacto:

Item	Nombre	Descripción	Responsable	Medio	Formato	Frecuencia	Categoría	Publicable (S/N)	Restricción	Impacto			Dificultad			Total Impacto	Total Dificultad	Plazo
										i1	i2	i3	d1	d2	d3			