



MINISTERIO
DE TELECOMUNICACIONES Y DE
LA SOCIEDAD DE LA INFORMACIÓN

Guía de Datos Abiertos

PROPUESTA

Agosto 2020

CONTENIDO

CONTENIDO	2
MARCO LEGAL	5
ASPECTOS GENERALES	9
Objeto de la guía	9
Fines de la guía	9
Ámbito de aplicación	9
Términos y definiciones	10
CICLO DE LOS DATOS ABIERTOS	11
I. CONOCE	11
Objetivo	11
¿Qué son los datos abiertos?	11
¿Para qué sirven los datos abiertos?	12
¿Qué es la protección de datos personales?	13
Consideraciones para la anonimización	14
Técnicas de anonimización	15
¿Quiénes participan en la apertura de datos?	17
¿Quién está a cargo en el país del proceso de apertura de datos?	17
¿Cómo organizarse dentro de la institución?	18
II. PLANEA	19
Objetivo	19
Selección y priorización de la demanda de datos abiertos	19
Estrategias para el levantamiento y actualización de un inventario de la información pública que es más solicitada por los ciudadanos	21
Metodologías de participación ciudadana para establecer demanda de datos abiertos	21
Planificación del portafolio de conjuntos de datos	22
Promover la innovación con datos de la institución	24
Datos abiertos para promover la investigación científica	26
III. PUBLICA	28
Objetivo	29
Consideraciones técnicas	29

FASE PREPARATORIA	29
Sistemas de origen	29
Aspectos de interoperabilidad	30
Estándares para los datos abiertos	30
Vocabulario estandarizado	31
Para nombres de archivos	31
Los formatos a publicar	32
NIVELES DE APERTURA DE LOS DATOS ABIERTOS	33
Primera Estrella	34
Segunda Estrella	34
Tercera estrella	35
Cuarta estrella	36
Quinta Estrella	37
FASE DE CONSTRUCCIÓN	38
Licencias de datos abiertos	38
Estructura de los metadatos	38
Estructura del Diccionario de datos	39
Aportes a la gestión de la calidad de los datos abiertos	40
Aplicación de criterios de calidad de los datos abiertos	40
Evaluación de la calidad de los datos previa a la publicación	42
FASE DE PUBLICACIÓN	43
Plataforma gubernamental de publicación de datos	43
Interfaz de Programación de Aplicaciones (API)	44
Conjuntos de datos	44
Recursos	44
Usuarios de la plataforma y roles	44
IV. PROMUEVE	45
Objetivo	45
Difusión para la promoción de datos abiertos	45
V. EVALÚA Y PERFECCIONA	47
Percepción de los usuarios de los datos publicados	47

La publicación versus la planificación del portafolio institucional de datos abiertos	48
Nivel de descarga de conjuntos de datos	48
Generación de informes anuales	48
ANEXOS	49
HERRAMIENTA DE PLANIFICACIÓN	50
Portafolio institucional de datos abiertos (ejemplo sujeto a ajustes)	50
ANEXO SOBRE METODOLOGÍAS DE PRIORIZACIÓN DE DATOS	51
HERRAMIENTA DE PREPARACIÓN	52
Referencia de licencias para publicación de datos abiertos	52
Metadatos (ejemplo sujeto a ajustes)	52
Diccionario de datos (ejemplo sujeto a ajustes)	52
HERRAMIENTA PARA LA PUBLICACIÓN	53
Categorías	53
REFERENCIAS	54
CO-CREACIÓN	55

MARCO LEGAL

La Constitución de la República, en su artículo 18, señala como derecho de todas las personas el acceder libremente a la información generada en entidades públicas.

La Ley Orgánica de Transparencia y Acceso a la Información Pública (LOTAIP), publicada en 2004, establece en su artículo 1: *“El acceso a la información pública es un derecho de las personas que garantiza el Estado”*. Por otro lado, en su artículo 7, lista la información mínima actualizada que las entidades de gobierno deben difundir desde sus sitios web, siendo este el principal insumo en la actualidad para la disponibilidad libre de información del sector público.

La Resolución No. 007-DPE-CGAJ de la Defensoría del Pueblo, que abarca los parámetros para cumplimiento del artículo 7 de esa ley, establece en su artículo 13: *“(…) los contenidos a publicar en el link de transparencia de los sitios web institucionales en las respectivas matrices homologadas en formato PDF, con los enlaces (hipervínculos) y los documentos para descargar la información que corresponda, sin perjuicio de que dicha información adicionalmente se publique en formato de dato abierto, para aportar a los procesos de transparencia, participación y colaboración ciudadana, innovación y emprendimiento para el desarrollo del país”*.

La Ley Orgánica del Sistema de Registro de Datos Públicos, indica en su artículo 3, que: *“Los datos públicos registrales deben ser completos, accesibles, en formatos libres, sin licencia alrededor de los mismos, no discriminatorios, veraces, verificables y pertinentes, en relación al ámbito y fines de su inscripción”*.

El Código Orgánico de la Economía Social de los Conocimientos, la Creatividad y la Innovación (COESCCI), incluye en su artículo 142, que *“Los estándares abiertos son formas de manejo y almacenamiento de los datos en los que se conoce su estructura y se permite su modificación y acceso no imponiéndose ninguna restricción para su uso; además que los datos almacenados en formatos de estándares abiertos no requieren de software propietario para ser utilizados”*.

El artículo 116, ibídem, indica que *“La información y el contenido de las bases de datos producto de las investigaciones financiadas con recursos públicos serán de acceso abierto. Las instituciones o entidades responsables de tales investigaciones deberán poner a disposición dicha información a través de las tecnologías de la información”*.

La Ley Orgánica de Participación Ciudadana obliga a las autoridades del Estado electas o de libre remoción, representantes legales de las empresas públicas o personas del sector privado que manejan fondos públicos, a rendir cuentas sobre sus actos de manera anual. La Ley de Presentación y Control de Declaraciones Patrimoniales Juradas, establece el acceso a la información patrimonial de todos los servidores públicos a través de la plataforma virtual de la Contraloría General del Estado.

Mediante Acuerdo No. 1184, de 20 de julio de 2015 se emite la Norma Técnica en la Gestión Pública, dentro de lo cual se estipula que *“La disponibilidad de los datos abiertos entre actores internos y externos de la organización, promueve las relaciones con otros grupos de interés relevantes, la generación de círculos de innovación, y favorece la participación de personas, embajadores de*

innovación y organizaciones de distinta índole, con lo cual se dinamiza la interacción de los tres sectores de un sistema nacional de innovación, es decir, público, privado y académico; entendiendo que el eje de los tres sectores es la ciudadanía en general”.

Mediante Acuerdo Ministerial No. 015-2019, de 18 de julio de 2019, publicado en Registro Oficial 69, de 28 de octubre de 2019, el Ministro de Telecomunicaciones y de la Sociedad de la Información aprobó la Política Ecuador Digital que tiene como objetivo transformar al país hacia una economía basada en tecnologías digitales, mediante la disminución de la brecha digital, el desarrollo de la Sociedad de la Información y del Conocimiento, el Gobierno Digital, la eficiencia de la administración pública y la adopción digital en los sectores sociales y económicos.

El artículo 4 de la Política Ecuador Digital establece: *“La implementación de la presente política se realizará con base en tres ejes de acción: Conectividad, Eficiencia y Seguridad de la Información, e Innovación y Competitividad. (...) 4.2. Eficiencia y Seguridad de la Información.- Garantizar la participación ciudadana, la democratización de los servicios públicos, la simplificación de trámites, la gestión estatal eficiente, la publicación y utilización de datos abiertos, la gestión de la seguridad de la información, y la protección de datos personales (...)”.*

El Decreto Ejecutivo No. 981, suscrito el 28 de enero de 2020, señala en su artículo 1: *“Del gobierno electrónico.- La implementación del gobierno electrónico en la Función Ejecutiva, consiste en el uso de las tecnologías de la información y comunicación por parte de las entidades para transformar las relaciones con los ciudadanos, entre entidades de gobierno y empresas privadas a fin de mejorar la calidad de los servicios gubernamentales a los ciudadanos, promover la interacción con las empresas privadas, fortalecer la participación ciudadana a través del acceso a la información y servicios gubernamentales eficientes y eficaces y coadyuvar con la transparencia, participación y colaboración ciudadana”.*

El artículo 2 del Decreto Ejecutivo No. 981, señala: *“El Ministerio de Telecomunicaciones y de la Sociedad de la Información será la entidad rectora en gobierno electrónico de la Función Ejecutiva. Para la correcta implementación del gobierno electrónico ejercerá las siguientes atribuciones y responsabilidades: (...) 4. Emitir políticas, directrices, acuerdos, convenios y desarrollar proyectos respecto a datos abiertos de la Función Ejecutiva. 5. Articular y coordinar con las demás instituciones de la Función Ejecutiva, así como con las otras Funciones del Estado y demás actores públicos y privados que directa o indirectamente coadyuvan a la aplicación del presente Decreto. 6. Ejercer la representación oficial del Estado ante organismos y entidades internacionales en temas de gobierno electrónico”.*

Otros instrumentos que hacen referencia a estándares y publicación de información para la transparencia son el Plan Nacional de Gobierno Electrónico 2018-2021, emitido con Acuerdo Ministerial No. 011-2018¹; el Plan Nacional de Desarrollo 2017-2021², que se organiza en tres ejes programáticos: Derechos para todos durante toda la vida; Economía al servicio de la sociedad; y Más sociedad, mejor Estado. El tercer eje, “Más sociedad, mejor Estado”, comprende tres objetivos: Objetivo 7: Incentivar una sociedad participativa, con un Estado cercano al servicio de la ciudadanía;

¹ Plan Nacional de Gobierno Electrónico, <https://bit.ly/2VvHCDG>

² Plan Nacional de Desarrollo 2017-2021 (2017), <http://www.planificacion.gob.ec/>

Objetivo 8: Promover la transparencia y la corresponsabilidad para una nueva ética social; Objetivo 9: Garantizar la soberanía y la paz, y posicionar estratégicamente al país en la región y el mundo.

El Decreto Ejecutivo No. 371, de 19 de abril de 2018, señala en su artículo 1. “Declarar como política pública del Gobierno Nacional, la adopción de la Agenda 2030 para el Desarrollo Sostenible, orientada al cumplimiento de sus objetivos y metas en el marco de su alineación a la planificación y desarrollo nacional”. El Objetivo 16 de la Agenda 2030, entre otras metas plantea: *“16.5 Reducir considerablemente la corrupción y el soborno en todas sus formas; 16.6 Crear a todos los niveles instituciones eficaces y transparentes que rindan cuentas.; 16.7 Garantizar la adopción en todos los niveles de decisiones inclusivas, participativas y representativas que respondan a las necesidades; 16.10 Garantizar el acceso público a la información y proteger las libertades fundamentales, de conformidad con las leyes nacionales y los acuerdos internacionales”*.

Con Acuerdo Ministerial No. 011-2020, publicado en el Registro Oficial No. 190 el 24 de abril de 2020, el Ministerio de Telecomunicaciones y de la Sociedad de la Información emitió la Política de Datos Abiertos, con el cual dispone la implementación de los datos abiertos en la Función Ejecutiva para fortalecer la participación ciudadana, la transparencia gubernamental, mejorar la eficiencia en la gestión pública, promover la investigación, el emprendimiento y la innovación en la sociedad. La Primera Disposición Transitoria señala que para implementación de la presente Política el ente rector de Gobierno Electrónico emitirá la Guía de Datos Abiertos, en el plazo de noventa (90) días contados desde la publicación del presente Acuerdo en el Registro Oficial.

En el marco internacional, Ecuador ha suscrito varios instrumentos internacionales que tienen relación con los datos abiertos:

- En 2007, suscribió Carta Iberoamericana de Gobierno Electrónico (CLAD), uno de sus objetivos consiste en: *“Incrementar la calidad de los servicios y productos públicos que el Estado tiene que suministrar a los ciudadanos al mejorar la eficiencia, la eficacia y una mayor transparencia de la gestión pública, aprovechando la utilización de las TIC en el Gobierno y en la Administración Pública”*. Asimismo, promueve los principios del gobierno electrónico entre los que está la *transparencia*.
- En 2016, suscribió la Carta Iberoamericana Gobierno Abierto CIGA, cuyo apartado C, destaca: *“Datos públicos para el desarrollo incluyente y sostenible. En materia de apertura de datos, los gobiernos deberían diseñar, implementar y desarrollar portales de datos abiertos y elaborar normativas y/o pautas metodológicas para su adecuada categorización, uso y reutilización por parte de la ciudadanía y otros actores del ecosistema del gobierno abierto. Para ello, los países podrán apoyarse en la Carta Internacional de Datos Abiertos”*.
- En abril de 2018, el Ecuador suscribió el Compromiso de Lima en la VIII Cumbre de las Américas, bajo el lema Gobernabilidad democrática frente la corrupción, los países de las Américas, entre otros aspectos, se comprometieron a: *“20. Impulsar el establecimiento de un Programa Interamericano de Datos Abiertos, en el marco de la OEA, con el objetivo de fortalecer las políticas de apertura de información, e incrementar la capacidad de los gobiernos y ciudadanos en la prevención y el combate a la corrupción, teniendo en cuenta los importantes trabajos*

realizados en el ámbito interamericano en esta materia y otras iniciativas regionales y mundiales”.

- El 18 de julio de 2018, Ecuador se integró a la Alianza para el Gobierno Abierto (OGP) y dio inicio a la co-creación de su Plan de Acción³ implementado un proceso colaborativo con la intervención de actores de sociedad civil, academia y gobierno. Este Plan de Acción incluye diez compromisos, uno de ellos es la “Elaboración colaborativa de la política de datos abiertos y la guía para su implementación”, el cual tiene como contraparte desde la sociedad civil, a la Red Ecuatoriana de Datos Abiertos y Metadatos (REDAM).

Con Memorando Nro. SENADI-DG-2020-0104-M (21 de julio de 2020), el Servicio Nacional de Derechos Intelectuales emitió su pronunciamiento respecto al uso de licencias:

Solo en virtud de la existencia de una obra protegible por propiedad intelectual, estaremos ante la posibilidad de establecer licencias de uso de esta, de lo contrario no se podrá establecer un régimen de licenciamiento de derechos de propiedad intelectual, puesto que no hay derechos de autor o conexos que licenciar frente a un objeto que no constituya obra original y por tanto no sea susceptible de apropiación mediante este régimen particular de Propiedad Intelectual.

Considerando el espectro de autorización de uso que otorgue el titular sobre la obra⁴, se contemplan varias posibilidades en el régimen de licenciamiento de obras, a saber: licencias abiertas y cerradas. Las licencias abiertas serán aquellas que se establezcan sobre obras que se encuentran bajo la protección de los derechos de propiedad intelectual de las que sin embargo su titular, ha autorizado una libertad amplia y casi sin restricciones en sus usos.

Adicionalmente, según lo que señala la Organización Mundial de la Propiedad Intelectual: “(...) los titulares de derechos de autor pueden optar por renunciar efectivamente al ejercicio de los derechos, ya sea total o parcialmente. Dicha renuncia efectiva se define también a veces como una licencia sin condiciones de utilización, o con condiciones de utilización limitadas. El titular puede, por ejemplo, proceder a publicar material protegido por derecho de autor en Internet y ponerlo a disposición de todo el que quiera utilizarlo o puede restringir esa renuncia a utilidades con fines no comerciales, con o sin determinados requisitos adicionales (...)”⁵.

En virtud del grado de libertad y las restricciones establecidas, las licencias libres pueden ser: puramente abiertas, esto es aquellas en las que el titular no se reserva ningún derecho y posibilita el uso libre y sin restricciones de la obra, permitiendo incluso al titular de las obras derivadas a establecer licencias cerradas sobre las mismas; y las denominadas licencias con efecto viral, como suelen ser las "creative commons", en las que el titular se reserva el derecho de impedir que las obras derivadas tengan licencias cerradas.

Existen diferentes tipos de licencias creative commons, en virtud de los derechos que se reserve el titular de la obra, situación que es de particular consideración, puesto que la Propiedad

³ <https://www.gobiernoabierto.ec/primer-plan-de-accion/>

⁴ Régimen de Derecho de Autor y Derechos Conexos

⁵ OMPI, 2016 Principios básicos del derecho de autor y los derechos conexos

Intelectual en sí misma es una herramienta jurídica que utilizada de manera estratégica y adecuada promueve la inversión económica y la generación de nuevo conocimiento.

“Para la aplicación de licencias de datos abiertos, se deberá en principio verificar la efectiva existencia de una obra protegida por propiedad intelectual, para en función de la obra y en atención de las circunstancias determinar el tipo de licencia que más convenga aplicar, considerando al efecto el justo equilibrio entre el bienestar colectivo, el reconocimiento al trabajo realizado por el creador y el incentivo a la inversión”.

ASPECTOS GENERALES

Objeto de la guía

Proporcionar criterios técnicos y metodológicos para planificar, abrir, publicar y promover la utilización de los datos abiertos gubernamentales.

Fines de la guía

- a. Orientar la apertura y publicación de datos de acuerdo con los principios de apertura establecidos en la Política de Datos Abiertos.
- b. Involucrar aliados estratégicos de los diferentes sectores de la sociedad, con criterios de corresponsabilidad en las diferentes etapas del ciclo de vida de los datos abiertos.

Ámbito de aplicación

La presente Guía de Datos Abiertos instrumenta la implementación de la Política de Datos Abiertos para la Administración Pública Central, según lo señalado en el Artículo 45 del Código Orgánico Administrativo, comprende:

1. La Presidencia y Vicepresidencia de la República
2. Los ministerios de Estado
3. Las entidades adscritas o dependientes
4. Las entidades del sector público cuyos órganos de dirección estén integrados, en la mitad o más, por delegados o representantes de organismos, autoridades, funcionarios o servidores de entidades que integran la administración pública central.

En el caso de otras entidades no dependientes de la Administración Pública Central, y otras funciones del Estado, podrán aplicar lo establecido en la presente Guía para la mejora de su gestión en materia de datos abiertos.

Términos y definiciones

Para efectos de aplicación de la Presente Política se definen los siguientes términos:

API: Siglas de Application Programming Interface Una manera en que los programas informáticos se hablan entre ellos. Puede ser entendido en términos de cómo un programador envía instrucciones entre los programas. (Open Data Handbook, <https://opendatahandbook.org/glossary/es>)

Anonimización: El proceso de adaptación de datos de manera que los individuos no pueden ser identificados a partir de ella. (Open Data Handbook, <https://opendatahandbook.org/glossary/es>)

Datos abiertos: Son 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. (Open Data Charter)

Digital: que crea, presenta, transporta o almacena información mediante la combinación de bits. (RAE)

Datos personales: Aquellos que permiten identificar o volver identificable a una persona natural (Reglamento a la Ley Orgánica de Gestión de la Identidad y Datos Civiles)

Información pública: Toda la información que emane o que esté en poder de las instituciones, organismos y entidades, personas jurídicas de derecho público o privado que, para el tema materia de la información tenga participación del Estado o sean concesionarios de éste. (LOTAIP)

Conjunto de datos: Es una colección de datos separados que son tratados como una unidad por un computador. (Diccionario Cambridge)

Reutilizar: Volver a utilizar algo, bien con la función que desempeñaba anteriormente o con otros fines. (RAE)

Metadatos: Se definen comúnmente como "datos acerca de los datos". Describen el contenido, la calidad, el formato y otras características que lleva asociadas un recurso, constituyendo un mecanismo para caracterizar datos y servicios de forma que usuarios y aplicaciones puedan localizarlos y acceder a ellos. (NORMA ISO de Información Geográfica - Metadata, <https://www.iso.org/standard/26020.html>)

Ontología: En ciencias de la comunicación y en inteligencia artificial, red o sistema de datos que define las relaciones existentes entre los conceptos de un dominio o área del conocimiento. (RAE)

CICLO DE LOS DATOS ABIERTOS

Se consideran cinco pasos básicos para cumplir con la implementación de la Política de Datos Abiertos en las entidades del Ejecutivo, buscando a través de este ciclo fomentar una cultura de datos abiertos en las entidades del ejecutivo y en la ciudadanía, centrándose en un ciclo que lleve a su utilización efectiva, y a mejorar de forma permanente; y se establece en los siguientes pasos; conoce, planea, publica, promueve, evalúa y perfecciona.

I. CONOCE

Objetivo

Conocer plantea el fundamento que posibilita la aplicación, partir del conocimiento es básico, esta sección responde preguntas básicas en torno a los datos abiertos, saber qué son, quiénes participan, por qué abrir datos, cuál es su potencialidad para impulsar procesos políticos, sociales y económicos.

¿Qué son los datos abiertos?

Datos abiertos son 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.

Esto significa que los datos abiertos pueden ser utilizados por todas las personas y con cualquier objetivo, ya sea informarse, crear un emprendimiento, generar una investigación científica o periodística, crear nuevos servicios o mejorar servicios existentes, etc.

Los datos que se abran y se publiquen deben responder a los principios internacionales de los datos abiertos.

- a. **Abiertos por defecto:** la información pública es de libre acceso, por lo que debe ser abierta por defecto, cumpliendo con las disposiciones legales en cuanto a información de carácter confidencial y acceso restringido, y lo relativo a la protección de los datos personales.
- b. **Oportunos y exhaustivos:** los datos abiertos deben corresponder a una demanda centrada en sectores y casos de uso específico; deben ser publicados de manera oportuna, tener una frecuencia de actualización que garantice la utilidad de los mismos; se debe cuidar que estos datos tengan el mayor detalle posible.
- c. **Accesibles y utilizables:** los datos abiertos deben ser fácilmente visibles y accesibles, estar disponibles para cualquier persona, sin necesidad de registro previo para su descarga.
- d. **Comparables e interoperables:** los datos abiertos deben ser presentados en formatos estructurados y estandarizados para facilitar la interoperabilidad, trazabilidad y reutilización efectiva.
- e. **Para mejorar la gobernanza y la participación ciudadana:** los datos abiertos fortalecerán la gobernanza y la confianza en las instituciones públicas.

- f. **Para el desarrollo incluyente y la innovación:** se buscará estimular la creatividad y la innovación promoviendo la utilización de los datos abiertos, esto es válido para usos gubernamentales, comerciales y no comerciales.

¿Para qué sirven los datos abiertos?

Para la sociedad, los datos abiertos son importantes porque con ellos se puede conocer las problemáticas actuales y cuáles son las acciones que se están llevando a cabo para solucionarlas como por ejemplo: conocer las condiciones ambientales de determinadas zonas geográficas, o la localización de centros educativos y culturales; para el sector privado, los datos abiertos permiten realizar análisis que les ayuden a ver qué pasa en el mercado, dónde pueden encontrar oportunidades de negocios y de nuevos emprendimientos; para la Academia los datos abiertos son la fuente para realizar investigaciones, para el desarrollo de aplicaciones; para el sector público y sociedad civil son un instrumento para promover la transparencia, la rendición de cuentas, la participación ciudadana.

Los datos abiertos tienen grandes potenciales, el valor radica en su utilización y reutilización, en la medida en que existan más infomediarios que analicen, organicen, relacionen unos datos con otros y presenten los resultados de estos procesos como información que sea fácilmente entendible y útil para la ciudadanía.

Para que una publicación de Datos Abiertos sea apropiada deberá cumplir las siguientes consideraciones básicas:

- Los datos deben ser completos. Deben contener la totalidad del tema y el mayor detalle posible, garantizando que la información suministrada sea suficiente y consistente y que no contenga datos nulos.
- Los datos deben ser primarios. Los datos se publican tal como se extraen de la fuente de origen, con el mayor nivel posible de granularidad.
- Los datos deben ser oportunos. Estar disponibles tan rápido como sea necesario para garantizar su valor y mantener una frecuencia de actualización que garantice la utilidad del dato.
- Los datos deben ser procesables por máquina. Los datos deben estar estructurados para permitir el procesamiento automático de los mismos y no estar en formatos propietarios o restrictivos a un software específico.
- El formato de los datos no debe ser propietario. Los datos estarán disponibles en un formato sobre el cual ninguna entidad tenga control exclusivo.
- Se debe establecer el nivel de utilización que puede darse a los conjuntos de datos, de manera que se promueva su utilización y reutilización sin restricciones. Como referencia, en los Anexos se incluye un listado de las licencias más utilizadas.

¿Qué es la protección de datos personales?

Los datos abiertos constituyen un recurso importante para aumentar el acceso de las personas a la información de interés generada por entidades del sector público, social y privado, y en consecuencia promover el desarrollo, la innovación, el crecimiento económico, entre otros aspectos de interés relevantes para la sociedad en general; sin embargo, no se puede dejar de lado, la preocupación de disponibilizar datos abiertos que pueden afectar derechos como el de la protección de datos personales.

La Dirección Nacional de Registro de Datos Públicos (Dinardap) hace énfasis en la Constitución de la República del Ecuador, que en su artículo 66, numeral 19 reconoce y garantiza el derecho a la protección de datos personales, en virtud de ello, debe considerarse que uno de los principios que lo rigen es el de la confidencialidad, es por eso que, el acceso a los mismos solo puede darse por aquellos que están debidamente legitimados para hacerlo, es decir por una persona natural o jurídica que cuente con consentimiento del titular, mandato de ley u orden judicial.

Existen formas en que los datos personales pueden convertirse en datos abiertos, como por ejemplo ser sometidos a procesos de anonimización. Acogiendo las experiencias internacionales, las recomendaciones a seguir para garantizar el derecho a la protección de datos personales en relación a la materia de datos abiertos son:

1. Analizar los casos de uso y el valor de los datos personales.
2. Identificar los actores involucrados en el proceso.
3. Analizar e identificar los posibles riesgos ante la apertura de datos personales.
4. Es recomendable contar con el consentimiento de los titulares previo a la apertura de datos personales.
5. Reunir a todos los actores interesados en la construcción de guías, políticas y paquetes.
6. Contar y cumplir con la normativa en protección de datos personales.
7. Observar la garantía de derechos como la honra, la intimidad, la protección de datos personales, entre otros.
8. Generar confianza pública.
9. Únicamente se podrán disponibilizar aquellos datos personales conforme a la normativa especializada en datos personales.
10. Anonimizar datos personales y monitorear permanente el anonimato de los datos.
11. Monitorear permanentemente la utilidad de los datos publicados y en caso de carecerla se sugiere eliminarlos.
12. Hacer seguimiento constante del contexto en que los datos abiertos están siendo utilizados.
13. Realizar un análisis de riesgo y de impacto que permita generar un plan de crisis a posibles casos en que se vulnere el derecho a la protección de datos personales.

Esta Guía promueve el cumplimiento de las disposiciones legales en lo relativo a la protección de datos personales. En caso de que la institución tenga solicitudes de información que incluyan datos personales, pero que a la vez se haya determinado, a través del Equipo de datos abiertos, que tiene gran potencial de que sean utilizados al ser publicados como datos abiertos, o que hagan eficiente la entrega de respuestas a pedidos ciudadanos que son reiterativos, o que hayan sido resultado del

establecimiento de la demanda de datos abiertos realizada por el ente rector de la Política de Datos Abiertos, se debe proteger los datos personales.

Consideraciones para la anonimización

La anonimización es un proceso de disociación a través del cual los datos se desvinculan de la persona a la que pertenecen, de forma que dejan de ser “datos personales”. La anonimización consiste en un proceso que permite identificar y ocultar la información sensible, permitiendo su divulgación sin que ello implique vulnerar los derechos a la protección de datos de personas y organizaciones que se puedan referenciar en los mismos. Citamos e incluimos aspectos relevantes de la anonimización de datos producida por la CEPAL⁶.

La anonimización de datos efectuada de manera adecuada, preserva la utilidad de los datos.

El análisis previo al proceso de disociación es esencial así también el personal implicado en el proceso de anonimización debe cumplir todos los requisitos de formación e información relativos al cumplimiento de la normativa de protección de datos personales.

Es aconsejable definir un protocolo de actuación. Algunas de las tareas o actividades que deben realizarse durante la fase de anonimización son:

- Determinar los recursos y equipo técnico necesarios para proceder a la anonimización de los datos.
- Determinar la técnica de anonimización que sea más apropiada en función de las variables que hubieran sido identificadas.
- Validar la técnica de anonimización por expertos.
- Aplicar la técnica seleccionada y ejecutar el proceso de anonimización.
- Incluir un sello de tiempo con el fin de garantizar la fecha y hora en la que la anonimización ha sido realizada, o la firma electrónica permiten garantizar la identidad electrónica de quien ha realizado la anonimización.
- Validar y aprobar los archivos anonimizados por expertos y por el equipo de evaluación.
- Incluir en la tabla de metadatos la característica de la anonimización realizada, para la publicación del conjunto de datos.

El equipo de trabajo para la anonimización debe estar integrado por:

- Responsable del tratamiento: decidir sobre la finalidad y objetivos a los que debe responder la información
- Responsable de protección de datos personales: velar por el cumplimiento de la normativa que precautele la privacidad.
- Responsable de seguridad de la información: evaluar riesgo residual de reidentificación, realizar pruebas de hacking ético encaminadas a valorar la fortaleza de los procedimientos que deben garantizar la irreversibilidad de la anonimización.

⁶ <https://biblioguias.cepal.org/c.php?g=495473&p=4961125>

- Responsable del proceso de anonimización: encargado de determinar las variables que se precisa anonimizar y proponer técnicas de anonimización que serán validadas por el responsable de seguridad de la información, y de su aplicación.

Durante el proceso de anonimización los datos biométricos, registros de voz o registros de imagen pueden presentar una complejidad específica que deberá abordarse en las fases iniciales del proceso de anonimización con carácter previo. Por ejemplo, en cuanto a los registros de voz cabe la posibilidad de realizar una transcripción previa con su respectiva eliminación de posibles identificadores (expresiones autóctonas, elementos epitécticos, identificadores retóricos, etc.) para posteriormente proceder a la reproducción de las transcripciones mediante dispositivos sintetizadores de voz en caso que fuera necesario mantener un registro sonoro.

Los registros de imágenes presentan su riesgo de reidentificación en el conjunto de la imagen, ya que en ocasiones puede reidentificarse a las personas por su entorno y no directamente por sus propios rasgos genéricos. Las variables de reidentificación de las personas mediante imágenes pueden ser múltiples, por lo que en ocasiones los datos de imagen requerirán un tratamiento específico para impedir la reidentificación de las personas.

Se debe tener en cuenta las garantías jurídicas necesarias para preservar los derechos de los interesados. Algunos de los aspectos que deben ser tenidos en cuenta son los acuerdos de confidencialidad que impliquen a los siguientes actores: responsable del proceso de anonimización, responsable del tratamiento de datos anonimizado, y otros actores que participen en el proceso.

Hay que tener en cuenta que la anonimización de la información siempre generará, independientemente de las buenas prácticas empleadas, cierto grado de distorsión entre la información anonimizada y la información no anonimizada. Esta distorsión es conocida como “diferencial de privacidad”. Así también se debe considerar el rápido avance tecnológico podría dejar obsoletas las técnicas de anonimización elegidas en un corto espacio temporal con la aparición de técnicas más seguras o la aparición de técnicas que permitan vulnerarlas. La selección de técnicas de anonimización debería someterse también a análisis de riesgos que ayudarán al establecimiento de pautas que permitan, al responsable del proceso de anonimización, optar o decidir por las técnicas más apropiadas en cada momento concreto.

Técnicas de anonimización

Existe una variedad de técnicas, así como de clasificaciones, como referencia tenemos algunos de los más conocidos y aplicados.

Algoritmos de Hash

Un algoritmo de hash es un mecanismo que, aplicado a un dato concreto, genera una clave que puede utilizarse para reemplazar el dato real.

El algoritmo de Hash permite que, partiendo de un mismo dato podamos generar siempre la misma huella digital, pero partiendo de una determinada huella digital nunca podremos obtener el dato original, garantizando la confidencialidad al tratarse de una operación matemática de un solo

sentido. Las claves resultantes de la aplicación de un algoritmo de hash son comúnmente conocidas como “huella digital”.

Algoritmos 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. Los resultados de las operaciones con datos cifrados dan por resultado valores igualmente cifrados que pueden ser descifrados posteriormente por el usuario que disponga de la clave para descifrar. El esquema de cifrado homomórfico abre la posibilidad del tratamiento de datos personales anonimizados garantizando la privacidad del tratamiento y que los resultados de los tratamientos van a ser accesibles únicamente al poseedor de la clave de descifrado.

Perturbación de datos

Consiste en la variación y supresión sistemática de datos que evita que las cifras resultantes faciliten información sobre casos específicos. Puede aplicarse como:

- Microagregación: técnica utilizada para anonimizar datos numéricos consistente en la sustitución de valores numéricos concretos por el valor medio calculado para un determinado grupo de datos mediante la agrupación, segregación, supresión o sustitución de registros independientes.
- Intercambio aleatorio de datos: introduce una distorsión aleatoria en un conjunto de microdatos manteniendo el detalle y estructura de la información original.
- Datos sintéticos: como
 - Distorsión de datos: se generan datos aleatorios que mantienen los resultados del conjunto de datos originales; o
 - Distorsión con microdatos híbridos: combinación de datos originales con datos sintéticos
- Permutación de registros: intercambio de valores de datos con valor clave que garantiza valores promedios y distribuciones estadísticas.
- Permutación temporal: movimiento aleatorio de rangos temporales que no genera distorsión sobre los resultados medios finales.
- Redondeo: sustitución de variables por valores redondeados de forma aleatoria.
- Reajuste de pesos: cuando se trabaja con muestras de datos conocidas, supone la distorsión de los valores de las muestras originales para evitar la reidentificación.
- Ruido aleatorio: inyecta ruido manteniendo la estructura de datos originales

Reducción de datos

Se reduce el número de datos originales sin alterar los mismos, disminuyen el nivel de detalle de los datos originales evitando la presencia de datos únicos o atípicos sin relevancia para el resultado final. Pueden aplicarse como:

- Eliminación de variables: eliminación de datos especialmente sensibles que pueden ser identificadores directos.
- Reducción de registros: cuando tras aplicar otra medida los sujetos sigan siendo identificables.
- Recodificación global: determinadas categorías de datos se agrupan en una nueva categoría reduciendo las posibilidades de reidentificación.
- Codificación superior o inferior: para casos en los que valores superiores o inferiores de un rango sean identificables, consiste en ampliar o reducir el rango mayor o menor.

¿Quiénes participan en la apertura de datos?

Usuarios: en general quienes consumen datos, como ciudadanos y demandantes de datos abiertos.

Academia y Comunidad científica: son demandantes de datos abiertos, y en la medida en que los utilizan para realizar investigaciones, generar información, conocimiento y obtener nuevos datos, también son generadores de datos abiertos.

Infomediarios: son personas o entidades que recopilan grandes cantidades de datos provenientes de diversas fuentes, los analizan y organizan de manera relevante para presentar de manera más comprensible o dar valor agregado, a usuarios que requieran esos datos.

Entidades públicas: son generadoras y demandantes de datos abiertos. Todas las entidades mencionadas en el ámbito de este documento implementan procesos de apertura de datos.

¿Quién está a cargo en el país del proceso de apertura de datos?



Imagen 1. Responsables de la apertura de datos desde el Estado

El ente rector de Gobierno Electrónico, por medio de la Subsecretaría de Gobierno Electrónico y Registro Civil, es quien coordina con las demás entidades del Ejecutivo y con organizaciones de la Sociedad Civil como la Red Ecuatoriana de Datos Abiertos y Metadatos (REDAM) y organizaciones aliadas; y realiza el rol de Administrador de datos abiertos.

El ente rector de la Planificación nacional gestiona la plataforma gubernamental de datos abiertos, facilita las herramientas necesarias para que las entidades del Ejecutivo publiquen los conjuntos de datos abiertos, considerando al portal www.datosabiertos.gob.ec, como punto de acceso a los datos abiertos gubernamentales.

Se considera la vinculación con los sectores público, privado, academia y sociedad civil, en todo el ciclo de vida los datos abiertos, y de manera permanente; considerando su involucramiento en los requerimientos como en la colaboración para la consecución de los fines que se busca en esta guía.

¿Cómo organizarse dentro de la institución?

De acuerdo con lo señalado en la política de Datos Abiertos, las entidades identificarán al Responsable de datos abiertos y al Gestor institucional de datos abiertos que es quien preside el Comité de transparencia; su conformación respetará lo estipulado en la LOTAIP.

El Gestor institucional de datos abiertos es quien organizará internamente un Equipo de Datos Abiertos que se encargará de cumplir el ciclo de los datos abiertos, según lo señalado en esta Guía.

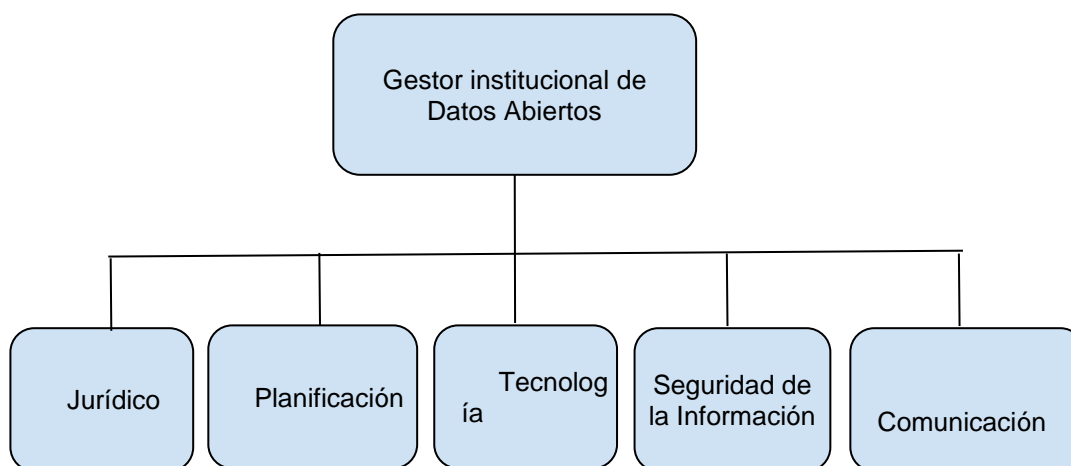


Imagen 2. Conformación del Equipo de datos abiertos

El Equipo de Datos Abiertos incluirá al Comité de Transparencia que está integrado por los titulares de las Unidades Poseedoras de la Información; o como mínimo por delegados de las áreas clave de la organización Comunicación, de Tecnología, el Oficial de Seguridad de la Información, o quienes hagan sus veces, dependiendo de la realidad de cada institución.

El Equipo de Datos Abiertos deberá reunirse como mínimo tres veces en el año manteniendo actas de reuniones, que pueden tener un formato libre, cuyo requisito mínimo es establecer las circunstancias de tiempo, modo, lugar, registro de asistencia, temas tratados (orden del día), los acuerdos alcanzados, y las tareas que se llevarán a cabo junto con los responsables de su ejecución y la fecha estimada de finalización.

El Equipo de Datos Abiertos analizará los requerimientos y sugerencias dadas por la ciudadanía, que sean recogidas a través del portal gubernamental de datos abiertos.

El Equipo de Datos Abiertos debe asegurar el cumplimiento de las acciones planteadas en todo el ciclo de los datos abiertos que promueve esta Guía: conoce, planea, publica, promueve, evalúa y perfecciona.

II. PLANEA

Objetivo

Planificar el proceso de apertura de datos desde los sujetos obligados o representantes del Estado de acuerdo con la demanda ciudadana, la oferta de datos estratégicos, sectoriales, alineados con los planes de desarrollo, planes estratégicos, cumplimiento de los Objetivos de Desarrollo Sostenible (ODS), otras agendas internacionales e índices internacionales para su uso y reúso de cualquier persona que así lo requiera, los mismos que deben estar actualizados de acuerdo con su período de generación.

Son objetivos específicos:

- Realizar un proceso de planificación para establecer el Portafolio institucional de datos abiertos en el último trimestre del año, previo a la publicación.
- Establecer un plan de actividades, en el último trimestre de cada año (a excepción del año 2020), para la promoción y utilización de los datos abiertos institucionales.

Selección y priorización de la demanda de datos abiertos

Identificar si los datos que se va a publicar son pertinentes y que no han sido ya publicado por entidades del orden nacional que consolidan estos datos

Los aspectos que se deben identificar en la información para publicar son:

- Dependencia del responsable del desarrollo y actualización de la información.
- Dependencia del responsable de la actualización de la información.
- Origen de la información (manual y/o automática).
- Periodicidad de la actualización de esta información.
- Si el conjunto de datos tiene diferentes orígenes de información
- Existencia de Metadatos, descripción y contexto claro de los datos.

Algunas de las consideraciones que se deben tener en cuenta para la priorización de publicación de datos abiertos son:

- Permitir el cumplimiento a las metas institucionales de transparencia, participación, control social y colaboración.
- Contribuir a un propósito de política pública específico y que hagan parte de las categorías establecidas por las mejores prácticas a nivel nacional e internacional en la materia.
- Estar alineados con los objetivos estratégicos trazados en el Plan Nacional de Desarrollo y los planes institucionales de los sujetos obligados del Estado, públicos y privados.
- Datos abiertos de gran tamaño, alta concurrencia de usuarios y de permanente actualización, esto con el fin de que se planee la construcción de procesos automáticos de extracción, transformación y carga de datos (ETL) desde los sistemas de información de la entidad al Portal de Datos del Estado.
- Datos que han sido solicitados por los ciudadanos en ejercicios virtuales o presenciales de consulta pública.
- Datos que han sido caracterizados como prioritarios por uno o varios actores de interés del ecosistema.
- Datos de derechos constitucionales e instrumentos internacionales.
- Objetivos de desarrollo sostenible.
- Índices internacionales de datos abiertos: Open Data Barometer, Open Data Index, Our Data Index, Procesos y directrices privacidad y uso de información delicada.
- El costo (recursos, tecnología, tiempo) en preparar los datos.
- Licencias para facilitar la reutilización.
- Demandas de conjuntos de datos.
- Articulación del sector público con la sociedad civil y la academia.
- Datos reutilizados para generar múltiples beneficios desde perspectivas económicas y sociales.
- Datos requeridos para monitorización y observatorios para aportar con los indicadores a fin de facilitar el análisis de la situación actual del país y contrastar con los de la región y el mundo.
- Publicar datos históricos que sean relevantes y factibles de publicar.
- Los problemas más comunes que se tienen en la planificación de aperturas de datos son:
- La falta de un estándar.
- Acceso.
- Calidad de datos: Agregación, actualización, completitud.
- Participación, confianza y colaboración de entidades.
- Persona responsable.
- Estrategia y políticas definidas y claras.

Estrategias para el levantamiento y actualización de un inventario de la información pública que es más solicitada por los ciudadanos

El Portafolio institucional de datos abiertos es un instrumento que los sujetos obligados del Estado deben realizar anualmente. El instrumento base del plan de apertura será el Portafolio de datos abiertos, proporcionado por la entidad competente, la cual debe incluir el levantamiento y actualización del inventario de la información pública, para lo cual se plantean los siguientes recursos:

- Los datos de otras entidades o de dependencias al interior de la misma entidad con el fin de evitar la duplicidad a la hora de publicar datos en el portal.
- Los datos solicitados por la ciudadanía de manera recurrente mediante solicitudes de acceso a la información o los que deriven de ellas; también aquellos que puedan resolver alguna problemática de la entidad.
- Los datos públicos generados o administrados por los Sistemas de Información de la entidad.
- Los datos públicos utilizados por herramientas digitales, aplicativos web o móviles desarrollados de la entidad.

Recursos adicionales a considerarse para actualizar el portafolio:

- Proyectos de impacto país.
- Reportes frecuentes de información estratégica o complementaria.
- Resuelvan problemáticas públicas y/o ciudadanas.
- Potenciar la oferta de datos a través de grupos focales con infomediarios, periodistas, investigadores para conocer qué tipo de datos y de qué sectores demandan el uso de datos.
- Estándares internacionales en materia de transparencia, acceso a la información y lucha contra la corrupción.
- Análisis ex post investigaciones realizadas por la academia, los sujetos obligados, públicos y privados y sociedad civil.
- Herramientas para analizar datos: Google analytics, BI y Big data.
- Sesiones con profesionales expertos por sectores para liberar datos.

Metodologías de participación ciudadana para establecer demanda de datos abiertos

La política aborda varios ejes para propiciar las condiciones adecuadas que garanticen la apertura de datos desde las instituciones públicas con el objetivo de generar el escenario mediante el cual parte de la demanda para la apertura de datos se establezca a partir de canales de participación ciudadana, es así que las instituciones como parte de su planificación deberán mantener habilitados estos canales de comunicación directa con los usuarios, para recibir propuestas y observaciones, que deberán ser contestados de manera oportuna y tener un seguimiento (trazabilidad) ya sea que

se incluyan, o no, en el portafolio institucional de datos cuya respuesta no supere un período de 30 días hábiles.

Los canales de comunicación con la ciudadanía deberán contener ítems específicos para que la solicitud de información sea concisa y clara, de tal manera que se impulse una cultura de datos. Se sugiere las siguientes metodologías para recolectar información:

- Formulario de enlace en la sección de transparencia de la web de la institución.
- Sección o subsección específica para colocar solicitudes de conjuntos de datos.
- Registros administrativos.
- Procesos participativos que pueden incluir, entre otros: Rendición de cuentas participativas, mesas de diálogo para tratar la temática de datos abiertos y plantear nuevos criterios y elementos de interés de la ciudadanía.
- Observatorios regionales y globales, así como convenios con otras entidades para identificar metodologías específicas de recopilación de información para contrastar con estándares internacionales.
- Otras metodologías innovadoras y ejercicios virtuales que identifique la institución, como por ejemplo el uso de redes sociales, chatbots, u otros elementos interactivos con los las personas usuarias.

Algunos ejemplos de metodologías y buenas prácticas de otros países se incluyen en los Anexos.

Planificación del portafolio de conjuntos de datos

El catálogo de datos abiertos tiene como objetivos:

- Conformar un listado público de los datos generados o administrados por las entidades que se van a abrir y publicar, los mismos que deben cumplir los principios de los datos abiertos.
- Indicar a las partes interesadas en un conjunto de datos publicado en el portal de datos abiertos de Ecuador, las fechas propuestas para la publicación de los datos que las entidades han determinado factibles de abrir y publicar.
- Llevar un control de la información y manejar de forma eficiente el flujo de información.
- Proporcionar enlaces de interés que faciliten el paso a otras plataformas con el detalle de la información requerida.
- Delimitar la estructura de la presentación y la publicación de los diferentes tipos de datos abiertos
- Los datos cumplen con el principio de oportunidad y utilidad para ciudadanos, academia y organizaciones sociales
- Contar con un catálogo de datos abiertos que permita contribuir a la usabilidad y reutilización de la información que se genera, produce o custodia en las entidades del Estado y las privadas que administran fondos públicos
- Ofrecer información clave a los usuarios para entender las características de los datos a publicarse.

El Equipo de datos abiertos trabajará en las propuestas de liberación de datos, para lo cual debe realizar actividades de identificar, analizar, priorizar datos que se encuentran en las unidades poseedoras de Información, y que responden a tres aspectos básicos:

- a. Información pública que es solicitada frecuentemente por los ciudadanos. El equipo de datos abiertos debe velar porque se mantenga actualizado el registro de la Información pública que es solicitada por los ciudadanos, de manera que pueda ser analizada y determinados cuáles datos cumplen las características y principios de los datos abiertos, y son factibles de incluir en el Portafolio de datos abiertos.
- b. Datos que la institución considere oportuno publicar como datos abiertos. En las hojas de ruta también se debe incluir los datos que la entidad considere contribuyen a un propósito institucional o pueden ser de beneficio para la sociedad.
- c. La que disponga el ente rector de la política de datos abiertos, como resultado del establecimiento anual de la demanda de datos abiertos. Se incluirá también en el Portafolio de datos abiertos los datos que sean el resultado del proceso realizado por el ente rector de la política de datos abiertos para establecer la demanda de datos abiertos, se comunicará a cada entidad, los datos que han sido requeridos, en concordancia con sus atribuciones.

El producto de la consolidación de las opciones mencionadas, más un proceso de priorización, dan como resultado las hojas de ruta para la publicación de datos abiertos, mismas que deben ser publicadas periódicamente según se vayan generando, facilitando el acceso desde el portal gubernamental de datos abiertos.

El Equipo de datos abiertos debe asegurarse que las propuestas de datos a publicar cumplen con los requerimientos técnicos y jurídicos que promueve la Guía de datos abiertos.

Dentro de este proceso se deberá definir el portafolio de conjuntos de datos, es decir, determinar los datos que deben publicarse y la fecha de publicación. Se realiza un análisis de compensación entre los requisitos de la ciudadanía y los conjuntos de datos disponibles para priorizar los datos que se van a abrir. La estructura del portafolio de datos abiertos se puede observar en los Anexos.

- Código del conjunto de datos
- Nombre del conjunto de datos a publicar
- Descripción: dirección que publica los datos, Ciudad, Región, Tipo de entidad publicadora (Nacional, Territorial)
- Organizaciones que generan la información de origen
- Fecha prevista de publicación
- Estado actual (iniciado, en construcción, previo a publicación)
- Descripción de la licencia para publicación
- Organización solicitante
- Las partes interesadas o potenciales usuarios de este conjunto de datos
- Categorías
- Nombre de la entidad publicadora
- Correo de contacto del publicador
- Formato del dataset: número de variables y de observaciones, porcentaje de datos perdidos por variable, forma de recolección de datos.

Los Portafolios de datos abiertos deberán ser actualizados de manera constante de acuerdo con la priorización, peticiones, uso y demanda de datos abiertos. Para su publicación se crearán versiones del mismo. Para nombrar al archivo se debe seguir la siguiente estructura:

SIGLAENTIDAD_PDA_Año_versión

Ejemplo: MINTEL_PDA_2021_v1

Promover la innovación con datos de la institución

La promoción de los datos de la institución como vía para promover la innovación es uno de los usos específicos que puede darse. En este sentido, se parte de la conceptualización de la innovación:

Innovación. Es la introducción de un nuevo o significativamente mejorado producto, es decir bienes y/o servicios con cambios en su calidad; así como la introducción de nuevos o significativamente mejorados procesos de producción, es decir la forma de generar y proveer los bienes y/o servicios; métodos organizativos, con cambios estructurales y de gestión a nivel micro y macro; a través de la conexión de ideas y conocimientos multidisciplinarios que permitirán impactos positivos y darán como resultado valor social. Innovación es una idea implementada efectivamente para resolver un problema⁷.

El uso de los datos de la institución se convierte en un insumo para la generación de valor por parte de los gobiernos, sociedad civil, academia, sector privado y ciudadanía en general. El desarrollo de la innovación basada en datos es fundamental para una transformación digital y que permita a toda la sociedad, beneficiarse de la cuarta revolución industrial. Las actividades de innovación incluyen todas las actuaciones científicas, tecnológicas, organizativas, financieras y comerciales que conducen a la innovación.

De este modo, para promover procesos de innovación en el país a través de la utilización de los datos abiertos gubernamentales se debe considerar lo siguiente:

- Difundir y promocionar los datos institucionales (referirse a la sección anterior).
- Realizar un mapeo de actores institucionales y potenciales usuarios de los datos.
 - Nivel interno: Actores que potencialmente usen/brinden los datos para la generación de herramientas, productos y servicios a nivel del sector público, y que orienten la transformación digital.
 - Nivel externo: Actores que potencialmente usen/brinden los datos para la construcción de innovación hacia afuera del sector público, mediante procesos de datos e innovación abierta.
- Generar y promover el desarrollo de habilidades de procesamiento y manejo de datos para que los actores tengan la oportunidad y los incentivos de utilizar la información de la institución.

⁷ PNUD, 2019

- Se sugiere generar alianzas con actores claves para brindar capacitaciones o webinars sobre procesos de innovación a partir de datos abiertos.
- Asegurar espacios de intercambio de ideas colectivas y mecanismos de participación de actores multisector en base al uso de datos de la institución en proyectos innovadores. Algunas vías son:
 - Realizar webinars que se centren en casos de éxito de proyectos de innovación con el uso de datos abiertos, y en el caso de haber, estudios que han utilizado los datos de la institución.
 - Organizar/participar de congresos o seminarios que tengan como objetivo presentar casos de éxito de proyectos de innovación con el uso de datos abiertos, y en el caso de haber, estudios que han utilizado los datos de la institución.
 - Compartir proyectos innovadores desarrollados internamente de la entidad con la utilización de sus datos.
 - Promover ejercicios de uso de datos tales como retos públicos de innovación, datatones, hackatones, concursos de visualización, entre otros similares.
 - Implementar/crear programas que apoyen a proyectos de innovación con el uso de los datos abiertos de la institución.

Por su parte, la entidad pública generará actividades y procesos de Gestión de Innovación internos como los siguientes:

- Investigación y Diagnóstico
 - Las instituciones deberán valerse de los estándares, métodos, técnicas y/o metodologías nacionales o internacionales para favorecer el continuo levantamiento, procesamiento y análisis de datos abiertos.
- Gestión de ideas
 - Las instituciones deberán implementar espacios para la divergencia y convergencia de ideas a fin de promover la creatividad y aprovechamiento del capital intelectual de los propios servidores públicos. Se deberá definir un proceso institucional para la recopilación, evaluación y selección de ideas, las cuales podrían ponerse en práctica a manera de testeos o prototipos.
- Gestión de Proyectos
 - Los proyectos que se gestionen a nivel institucional deben contar con métricas de inicio, de ejecución y de impacto. Estos datos serán abiertos a la ciudadanía de manera periódica de acuerdo a los procesos especificados en la presente Guía.
- Protección y Aprovechamiento de Resultados
 - Las instituciones deberán generar estrategias para la protección y el aprovechamiento de resultados a través de una adecuada gestión del conocimiento con agentes internos y externos.
- Difusión de la Innovación
 - Se deberán difundir las innovaciones a fin de expandir y aprovechar el conocimiento desde los distintos sectores de la sociedad.
 - Las instituciones deberán valorar y reconocer la gestión y toma de decisiones en base a datos abiertos en sus diferentes áreas y deberán promover procesos de gestión del conocimiento.

- Documentación:
 - Se deberá guardar registro, de preferencia virtual, de las acciones, seguimiento, control, monitoreo, evaluación, resultados, retroalimentaciones relacionados a los aciertos y desaciertos que se generen dentro de la gestión de innovación.

Datos abiertos para promover la investigación científica

Los datos abiertos para promover la investigación científica es uno de los usos específicos que puede darse. En este sentido, se parte de la conceptualización de la investigación científica:

La Investigación es la búsqueda sistemática y lógica de nueva información sobre un tópico particular. Cuando se aplican los pasos de método científico se tiene investigación científica. De forma general la investigación científica se divide en dos: (a) Investigación básica, cuando se buscan los principios básicos que expliquen la ocurrencia de un fenómeno o evento. (b) Investigación aplicada: cuando se busca resolver problemas aplicando o adaptando teorías o principios ya conocidos.

En este sentido, los datos son la base sobre la cual se construye la ciencia. "Un intercambio de datos más eficiente y un acceso más abierto a la información y los recursos facilitarán la confirmación de las observaciones, la replicación de los experimentos, el respaldo, el rechazo o el refinamiento de las hipótesis y, en última instancia, la respuesta a los desafíos de la sociedad."

Para promover procesos de investigación científica a través de la utilización de los datos institucionales, se debe identificar a los actores del proceso de intercambio de datos e internamente definir los responsables:

1. **Investigador:** Solicita datos para sus tareas de investigación.
 - La investigación la realiza el personal académico de las instituciones de educación superior públicas y privadas como parte de su carga académica, los investigadores nacionales o extranjeros registrados en el SENESCYT, y el personal de investigación de los institutos nacionales de investigación.
2. **Institución fuente:** Es la Institución que almacena en cualquier medio la información solicitada.
 - Institución pública custodia de los datos.
3. **Custodio de datos:** Funcionario de institución fuente, responsable de resolver las solicitudes de datos.
 - Identificar un equipo/persona interna para procesar las solicitudes.
4. **Contraparte de investigación:** En caso de existir, es la persona de la institución fuente que colabora en tareas de investigación con el investigador y/o le atañe el resultado de la investigación.
 - Identificar un equipo/persona interna para colaborar con la investigación.

Por otro lado, con el objetivo de lograr un verdadero y efectivo fomento a la investigación, se proponen los siguientes principios que, en ausencia de una norma expresa, deberán ser la guía para el accionar de los interesados.

1. Desconcentración: El intercambio de datos debe ser directo entre el investigador y el custodio de los datos.
2. Agilidad: El proceso de solicitud, autorización e intercambio de datos debe realizarse por vía electrónica utilizando las credenciales de las instituciones donde laboren.
3. Trazabilidad: Los custodios y receptores de datos deben registrar el intercambio de datos y la evidencia de los productos generados con ellos a fin de diseminar conocimiento generado.
4. Responsabilidad: En cada etapa del intercambio de datos alguien debe ser responsable de ejecutarla a fin de garantizar la marcha del proceso.
5. Retroalimentación: Estadísticas e información detallada de cada solicitud será de acceso público.

Cabe destacar que el mecanismo para solicitar y registrar el intercambio de datos entre el solicitante y la institución será el Sistema de Registro de Solicitudes de Datos para Investigación (RSD), el mismo que debe ser incorporado y aplicado por la institución pública. El RSD seguirá el siguiente proceso frente a una solicitud de información para una investigación científica:

- **Requisitos de acceso a datos**

Un investigador debe tener acceso a los datos de su interés mientras no se contravenga prohibición expresa en las leyes, sin mayor requisito que su solicitud registrada en el RSD.

De existir algún acuerdo u otro instrumento legal que ampare el acceso a datos entre el investigador (o la institución donde labore) y la institución que tenga en su posesión los datos solicitados, se utilizará el RSD con el fin de eliminar trámites administrativos y transparentar el intercambio de datos.

- **Información sensible**

Información sensible es aquella que permite identificar a los ciudadanos. Si los datos solicitados incluyen información sensible los campos relevantes deberán ser eliminados o sujetos a cualquier proceso automático de anonimización en la institución.

Si la institución fuente no tiene capacidad técnica o el conocimiento para manejar datos deberá coordinar con el investigador la extracción y procesamiento de los datos in-situ. La institución de origen deberá supervisar estos pasos con la finalidad de certificar que en los datos entregados al investigador no existe información sensible.

- **Manejo de solicitudes**

El RSD debe registrar el flujo de acciones entre los actores de la solicitud de datos desde su inicio hasta que el acceso a datos haya sido negado o los datos descargados y utilizados para investigación.

1. Los plazos de cada paso del proceso serán acordados entre las partes y registrados.
2. El estado de las solicitudes (pendiente, aceptada, ejecutada, en espera de evidencia) cambiará de forma automática en función de las acciones ejecutadas por las partes.
3. El investigador debe calificar la calidad de los datos entregados, si cumple con las especificaciones solicitadas y la calidad del servicio prestado por la institución fuente.
4. El investigador puede cancelar su solicitud en cualquier momento especificando la razón.

5. La negativa de acceso a datos puede dar lugar a solicitudes recurrentes, no es necesario hacer una nueva solicitud.
6. Estadísticas de cada investigador, custodia de datos, institución fuente serán publicadas para el conocimiento del público general.

- **Evidencia de investigación**

La investigación puede encontrarse en cualquiera de los siguientes estados:

1. Preparación: Es cuando se usan “pocos” datos para evaluar el potencial de una hipótesis de investigación.
2. Ejecución: Es cuando se usan “muchos” datos en un proyecto de investigación que quiere averiguar el comportamiento de una plaga a lo largo del tiempo en todo el país.
3. Cierre: Es cuando se han generado conocimiento en base a los datos.

Dependiendo del estado del proceso de investigación el investigador deberá incluir evidencia del uso de datos en forma de un reporte técnico, un artículo científico o cualquier otro medio que permita difundir el conocimiento generado incluso si los resultados son negativos o la investigación resulta inviable por cualquier motivo.

- **Gestión de Datos de Investigación**

La Gestión de Datos de Investigación (GDI) es el proceso activo de manejo de los datos, recopilados, procesados y generados en una investigación. La GDI se realiza de forma continua y cubre todas las decisiones relacionadas con la gestión de los datos a lo largo de su ciclo de vida, desde la etapa de planificación de la investigación, su ejecución, hasta la diseminación de sus resultados y la preservación de los conjuntos de datos de forma que estos sean precisos, completos, auténticos y fiables, y se mantengan accesibles y reutilizables a lo largo del tiempo.

La GDI es una responsabilidad compartida entre los diversos actores que están involucrados en el desarrollo y comunicación de una investigación. En consecuencia, es vital que todos los actores involucrados conozcan con claridad cuáles son sus roles y responsabilidades en el proceso de gestión.

III. PUBLICA

Es un proceso que va a permitir aplicar características técnicas de formato, estandarización y criterios de calidad para que los conjuntos de datos se encuentren disponibles abiertamente de modo que puedan ser utilizados por computadoras y personas.

Para la publicación, esta sección se encuentra organizada considerando: la fase preparatoria, los formatos, la fase de construcción, la fase de evaluación de la calidad y la fase de publicación.

Objetivo

Definir las actividades necesarias para la publicación de los conjuntos de datos en formatos abiertos priorizando la representación real de la información almacenada en la institución a través de las siguientes acciones:

- Estandarizar los formatos de publicación
- Estructurar los metadatos
- Estructurar los diccionarios de datos
- Determinar las licencias de publicación
- Verificar el cumplimiento de criterios como parte de evaluación de la calidad de los conjuntos de datos, previa a la publicación
- Publicar los conjuntos de datos

Consideraciones técnicas

- Presentamos una adaptación del concepto de la gestión del Archivo Maestro (Master Data Management) procedente del estándar ISO 8000, para considerar las características de la publicación de datos abiertos.
- Proponemos una estructura de metadatos basado en el estándar del Vocabulario del Catálogo de Datos (DCAT⁸) que favorece la comprensibilidad de los datos publicados en formato abierto y propone métricas para su evaluación.
- Consideramos al portal gubernamental www.datosabiertos.gob.ec como punto centralizado de acceso a los datos abiertos de todas las entidades, considerando que el proceso de preparación de los datos es descentralizado e incluso está considerada la publicación en sus propios portales institucionales.
- Utilizamos categorías (ver en el Anexo) que permiten agrupar los conjuntos de datos en el momento de su publicación en el portal, lo cual facilita la búsqueda.

FASE PREPARATORIA

Sistemas de origen

Una vez definido el conjunto de datos que se requiere publicar, se deberá analizar los sistemas de origen de los datos requeridos, por cuanto es importante tener el registro de la procedencia de la información que se va a publicar como datos abiertos, ya que muchas veces se encuentra dispersa, almacenada en fuentes (silos) diversas, en diferentes formatos, que no necesariamente están relacionados y al momento de publicarlos en formato abiertos, ocasiona que sean difíciles de entender por no contar con información contextual adicional, lo que limita el valor de los datos por no poder reutilizarlos. En este aspecto la gestión se basa en utilizar una herramienta para unificar

⁸ <https://www.w3.org/TR/vocab-dcat-2/>

los datos en una sola fuente incorporando normas y reglas, como eliminación de registros duplicados, completar campos vacíos, entre otros, que permitan preparar el conjunto de datos previo a su publicación.

Aspectos de interoperabilidad

La interoperabilidad consiste en estandarizar, comprende un conjunto de principios que orientan los esfuerzos políticos, legales y organizacionales de las entidades, con el fin de facilitar el intercambio seguro y eficiente de información.

La interoperabilidad en el **dominio semántico** permite garantizar que, en el momento de intercambiar datos, el significado de los datos sea el mismo para todos los actores involucrados.

La interoperabilidad en el **dominio técnico** hace referencia a las condiciones que se deben cumplir para conectar los sistemas de información con el propósito de intercambiarla.

Estándares para los datos abiertos

Se priorizará el uso de estándares internacionales para los datos, tomando como referencia las publicaciones: Datos enlazados de información geográfica del Ecuador⁹ y Datos ecuatorianos geoespaciales enlazados¹⁰, como se detalla a continuación:

Dato	Estándar a aplicar	Detalle
Unidades de medidas	Ontologías de cantidades, Unidades, Dimensiones y Tipos de Datos QUDT http://www.qudt.org/qudt/owl/1.0.0/unit/Instances.html	Unidades de medida
Moneda	ISO 4217-2015 https://www.iso.org/iso-4217-currency-codes.html	La moneda de cada monto debe especificarse usando el código de moneda de 3 letras mayúsculas del ISO 4217-2015
Tipo de archivos	IANA media types https://www.iana.org/assignments/media-types/media-types.xhtml	Estándares para gestionar tipos de archivos
Idioma	ISO 639-1 https://docs.oracle.com/cd/E13214_01/wli/docs92/xref/xqisocodes.html	Se usa el código de 2 dígitos estandarizado en la ISO 639-1
Países	ISO 3166 https://docs.oracle.com/cd/E13214_01/wli/docs92/xref/xqisocodes.html	Se usa el código de 2 dígitos estandarizado en la ISO 3166
Fecha corta	ISO 8601 https://www.w3.org/TR/NOTE-datetime	Se usa el estándar ISODATE con año, mes, y día YYYY-MM-DD

⁹ Datos enlazados de información geográfica del Ecuador (julio-diciembre 2018), <https://pdfs.semanticscholar.org/36d2/5f8f743aafb2eb80719bb9adc0787418c85a.pdf>

¹⁰ Ecuadorian Geospatial Linked Data, https://www.researchgate.net/publication/260983879_Ecuadorian_Geospatial_Linked_Data

Fecha larga	ISO 8601 https://www.w3.org/TR/NOTE-datetime	Se usa el estándar ISODATE con año, mes, día, horas, minutos, segundos y zona horaria. YYYY-MM-DDThh:mm:ssTZD
Codificación	ISO 10646 https://www.iso.org/standard/73773.html	El formato de codificación de caracteres es el UTF-8. Define todos los caracteres necesarios en la mayoría de los idiomas.
Catalogar los datos	DCAT-2 https://www.w3.org/TR/vocab-dcat-2	Para definir campos de los metadatos

Tabla 1. Estándares internacionales para los datos

Para los casos que no están detallados por ser aplicados únicamente en el ámbito ecuatoriano se considerarán los siguientes estándares locales:

Dato	Estándar a aplicar	Detalle
Provincia, cantón, parroquia	INEN http://apps.normalizacion.gob.ec/descarga/ https://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/C_PV_aplicativos/modulo_cpv/dpa_.pdf	Se usan los códigos normalizados por el INEN

Tabla 2. Estándares nacionales para los datos

Vocabulario estandarizado

Para nombres de archivos

- Formato para la denominación del nombre del conjunto de datos:
SIGLAS DE LA ENTIDAD_NombreConjuntoDatos_MesAño
Ejemplos:
MINTEL_RemuneracionPuesto_Mayo2020
MERNNR_ServicioOtorgamientoPersonalidadJuridica_Enero2020
- Formato para la denominación del nombre del diccionario de datos:
SIGLAS DE LA ENTIDAD_NombreConjuntoDatos_DD_MesAño
Ejemplos:
MINTEL_RemuneracionPuesto_DD_Mayo2020
MERNNR_ServicioOtorgamientoPersonalidadJuridica_DD_Enero2020
- Formato para la denominación del nombre del archivo de metadatos:
SIGLAS DE LA ENTIDAD_NombreConjuntoDatos_PM_MesAño
Ejemplos:
MINTEL_RemuneracionPuesto_PM_Mayo2020
MERNNR_ServicioOtorgamientoPersonalidadJuridica_PM_Enero2020

- No usar artículos ni preposiciones.
- Usar únicamente letras y números ASCII, comprendidos en el rango "a-z" y "0-9".
- Separar las palabras con guión bajo "_".
- No usar tildes ni caracteres especiales.
- Usar palabras en singular.
- Usar palabras en español.
- Incluir Mes (en letras con mayúscula inicial) y año (4 dígitos).

Los formatos a publicar

Los archivos a ser publicados deben poseer las siguientes características¹¹:

- Los datos deben ser claros para su reutilización, las imágenes deben tener una calidad media a alta, y en caso de ser audios deben ser entendibles.
- Los archivos deben garantizar una cómoda descarga de acuerdo con su volumen.
- Cuando se trata de mostrar gráficos estadísticos o geográficos, se debe incluir los datos que originaron las imágenes, caso contrario no se consideran datos abiertos.

Esta guía promueve la publicación en formatos estructurados, siendo los más utilizados:

Formato	Funcionalidades
CSV (Comma Separated Values)	Es un tipo de documento en formato abierto sencillo para representar datos en forma de tabla donde las columnas son separadas por comas (o punto y coma, donde la coma es el separador decimal), y las filas por saltos de línea.
ODS / XLS (Open Documento Spreadsheet / Archivos de Excel)	Documento abierto en el caso de ODS, y propietario en el caso de Microsoft XLS; muestra la información en celdas organizadas en filas y columnas.
XML (Extensible Markup Language)	Es un lenguaje general basado en etiquetas utilizado para representar información estructurada, define un conjunto de reglas de codificación a través de modelos específicos legibles y procesables. Es uno de los formatos más utilizados para la compartición de información entre programas, entre personas, entre programas y personas, tanto a nivel local como en red.
RDF (Resource Description Framework)	Es un modelo de representación para la descripción de recursos siguiendo los fundamentos de la web semántica. Se representa la información utilizando expresiones con la forma sujeto-predicado-objeto; el sujeto es el recurso que se describe, el predicado es la propiedad sobre la cual se quiere establecer el recurso y el objeto es el valor de la propiedad con el cual se establece la relación. A fin de que sea interpretable, se representa en formato XML
RSS (Really Simple Syndication)	Es un formato de datos basado en XML utilizado para redifundir contenidos a suscriptores de un sitio web. Se suelen utilizar para describir brevemente contenidos cuya actualización es bastante frecuente.

¹¹ <http://www.iim.csic.es/wp-content/uploads/2017/09/Datos-abiertos-2020.pdf>

JSON (JavaScript Object Notation)	Es un formato ligero para el intercambio de datos basado en una simplificación de la notación JavaScript.
SHP (ESRI Shapefile)	Formato propietario de datos espaciales, sirve para almacenar tanto la localización de los elementos geográficos como los atributos asociados a ellos.
VMS (Web Map Service)	Formato de servicio web para producir mapas de datos referenciados espacialmente, de forma dinámica a partir de información geográfica.
GML (Geography Markup Language)	Vocabulario XML para el modelaje, transporte y almacenamiento de información geográfica.
KML (Keyhole Markup Language)	Vocabulario XML para presentar datos geográficos en tres dimensiones. Suelen distribuirse comprimidos en formato KMZ.

Tabla 3. Formatos de datos abiertos

NIVELES DE APERTURA DE LOS DATOS ABIERTOS

A continuación, exponemos el modelo de cinco estrellas propuesto por Tim Berners-Lee¹², con las características a imprimir a los conjuntos de datos, cuya aplicación responderá a las capacidades técnicas de la institución en el manejo de la información.

★	Publicar datos en la web con cualquier formato y bajo una licencia abierta
★★	Publicar como datos estructurados
★★★	Usar formatos no propietarios
★★★★	Uso de estándares abiertos de W3C (Consortio World Wide Web) de manera que los datos puedan ser identificados mediante una URI (Identificador uniforme de recursos) y que así estén integrados en la web. Para este nivel una forma de representación de los datos es RDF.
★★★★★	Vinculando además los datos con los que otras personas o instituciones publican, de modo que se proporcione un contexto para ellos. El formato para este nivel es el linked RDF

Tabla 4. Esquema de desarrollo de 5 estrellas para datos abiertos, adaptado de <https://5stardata.info/en/>

¹² <https://5stardata.info/es/>

Características de metadatos

Los metadatos deben estar presentes en todos los niveles de apertura y se estructurarán considerando el estándar DCAT, priorizando los datos en origen previo a la publicación. En la sección Estructura de Metadatos se describe los campos a implementarse.

Primera Estrella

Exige que los datos estén disponibles en la web, sea en el formato que sea, con licencia abierta. Entre los tipos de archivos para cumplir este nivel se encuentra el PDF

Dominio semántico (vocabulario común)

El dominio semántico se caracteriza por una serie de rasgos distintivos, que son las unidades mínimas de significado que facilitan el entendimiento de los datos.

Las características de este nivel permiten lo siguiente:

- Acceso libre al archivo
- Posibilidad de modificación
- Capacidad de compartir
- Se requiere de la creación de software la extracción de los datos del documento o su copiado manual
- Facilita la publicación de los datos
- Dificulta la reutilización por máquina
- Los tipos de archivos que se incluyen en esta estrella son:

Tipo de archivo	Cerrado	Abierto	Libre
Documentos solo de lectura		PDF	EPUB, MOBI, DJVU, PS
Imágenes			PNG, JPEG-2000, TIFF
Archivos de texto		RTF	ODT
CAD		DXF, SAT	IGS, STP
Audios			OGA, FLAC
Videos	MPEG-4		OGV, WEBM
Data Geo		KML, KMZ	

Tabla 5. Tipos de archivos ★ estrella

Al preparar conjuntos de datos en este nivel, se debe tener las siguientes consideraciones:

- Verificar que las imágenes son claras, si se trata de letras que sean legibles.
- Definir claramente los metadatos que facilite la utilización del recurso.
- No utilizar varios formatos en mismo documento (A3, A4), lo cual dificulta la lectura.
- Verificar que los datos representen a los de origen.

Segunda Estrella

Los datos deben estar disponibles de manera estructurada, para que sea legible por máquinas. Los formatos que se manejan en esta estrella pueden ser XLS, DOC, MDB.

Dominio semántico (vocabulario común)

Las características de este nivel permiten lo siguiente:

- Exportar los datos, porque se encuentran en formato tabular.
- Se requiere un software propietario para extraer datos.
- Facilitan su procesado directo para realizar modificaciones, cálculos o visualizaciones.
- Los datos están accesibles en la web en forma estructurada
- El conjunto de datos contiene datos numéricos y alfanuméricos dispuestos en tablas. Cada tipo de unidad observacional es una entidad, donde cada variable es una columna y cada observación es una fila.

Al preparar conjuntos de datos en este nivel, se debe tener las siguientes consideraciones:

- No hacer uso de colores u otras señales estilísticas para codificar la información.
- El archivo no debe tener macros ni fórmulas que puedan alterar el contenido al momento de copiarla a otro archivo.
- Incluir dominios de datos estandarizados (por ejemplo, distribución política, tipo de institución, nombre de instituciones), emitido por INEC.

Tercera estrella

Al igual que la segunda estrella los datos deben estar disponibles en forma estructurada, pero en este caso los formatos deben ser no propietarios.

Dominio semántico (vocabulario común)

Las características de este nivel permiten lo siguiente:

- Se podrá hacer todas las cosas que puede hacer con el nivel dos estrellas, pero sin la limitación que impone el uso de un software en concreto.
- Para el gestor la preparación es sencilla previa a la publicación de datos, pero podrían surgir necesidades como la conversión de formatos o la necesidad del uso de plugins de exportación desde un formato propietario
- Datos disponibles en la web y dispuestos de manera que puede utilizarse sin restricciones.
- Formatos generales, los más habituales son CSV, XML y RDF, JSON, KML.

Al preparar conjuntos de datos en este nivel, se debe tener las siguientes consideraciones:

- Open Fiscal Data Package: Publicación de presupuesto (se podría sugerir a las Entidades Financieras)
- Open Contracting: Publicación de compras y contrataciones (actualmente empleado por SERCOP)

Cuarta estrella

Este nivel requiere el uso de estándares abiertos de W3C (Consortio World Wide Web) de manera que los datos puedan ser identificados mediante una URI (Identificador uniforme de recursos) y que así estén integrados en la web. Para este nivel una forma de representación de los datos es RDF.

Dominio semántico (vocabulario común)

Las características de este nivel permiten lo siguiente:

- Enlazar los datos desde otro sitio web, marcarlos como favoritos, reutilizar partes de los mismos, usar herramientas y librerías existentes incluso en el caso en que estas solo puedan manejar parte de los patrones de los datos, o incluso combinarlos con otros.
- Ofrece distintos formatos de serialización entre los más conocidos están: RDF/XML, RDF/JSON.
- Facilita la interpretabilidad de los contenidos con respecto al uso de XML o RDF por parte de aplicaciones o máquinas si se utiliza el Lenguaje de Ontologías Web (OWL).
- Las URIs seguirán una estructura uniforme con la finalidad de facilitar a los agentes reutilizadores su lectura e interpretación por ejemplo para reconocer la temática del conjunto de datos, a qué organismo pertenece, etc.

Al preparar conjuntos de datos en este nivel, se debe tener las siguientes consideraciones:

- Se requiere el análisis de los datos, la preparación para su representación, la asignación de las URIs y la búsqueda y creación de patrones para aplicarlos a la información.
- Declarar recursos en forma de expresiones denominadas triplets que tienen la forma sujeto-predicado-objeto. El sujeto se encarga de indicar el recurso, el predicado expresa la relación entre el sujeto y el objeto apuntando sus rasgos o aspectos, y el objeto sería el valor de la relación.
- Utilizar estándares abiertos de la Web semántica para identificar y consultar recursos: RDF, SPARQL.
- Describir la semántica de los recursos siguiendo las recomendaciones definidas por la W3C, y utilizar en medida de lo posible los vocabularios definidos en Linked Open Vocabularies (ontologías).
- El dataset debe ser descrito a través de la recomendación DCAT definida por la W3C.
- Definir la procedencia de datos (confianza), por ejemplo, con la ontología PROV-O¹³ que permite expresar la información del origen de los conjuntos de datos.
<https://www.w3.org/TR/dwbp/#provenance>

¹³ <https://www.w3.org/TR/prov-o>

Quinta Estrella

Este nivel requiere haber logrado los niveles anteriores y vinculando además los datos con los que otras personas o instituciones publican, de modo que se proporcione un contexto para ellos. El formato para este nivel es el linked RDF. (Berners-Lee, 2009)

Dominio semántico (vocabulario común)

Las características de este nivel permiten lo siguiente:

- Descubrir otros datos relacionados mientras se utilizan, y aprender directamente acerca de su esquema.
- Aumentar el valor de los datos y se logra que sea posible descubrirlos de manera más sencilla
- Utilización de estándares abiertos de la Web semántica para identificar y consultar recursos: RDF, SPARQL
- Los enlaces permitirán realizar integración de datos basado en los recursos y descubrir relaciones entre fuentes de datos diferentes

Al preparar conjuntos de datos en este nivel, se debe tener las siguientes consideraciones:

- Se debe incluir URI para los recursos
- Se requiere realizar los enlaces con otros en la web con el fin de proveer contexto, así como realizar un mantenimiento óptimo de los datos, sus URIs identificativas y sus enlaces.
- Describir la semántica de los recursos siguiendo las recomendaciones definidas por la W3C, y extender en medida de lo posible los vocabularios definidos en Linked Open Vocabularies (ontologías).
- La licencia puede ser definida de las siguientes maneras:
 - Genérica para todos los recursos que se publiquen en el portal.
 - Licencias independientes por Catálogos de datos, Conjuntos de datos (datasets), Recursos, Servicio de datos, vocabularios semánticos, ontologías, y otros elementos que forman parte del dataset abierto.
 - Algunos ejemplos de licencias se pueden encontrar en: <https://www.w3.org/TR/dwbp/#licenses>
- Definir la procedencia de datos (confianza), por ejemplo, con la ontología PROV¹⁴ que permite expresar la información del origen de los conjuntos de datos.

¹⁴ <https://www.w3.org/TR/prov-o>

FASE DE CONSTRUCCIÓN

Licencias de datos abiertos

Al establecer la licencia bajo la cual se van a publicar los datos, se debe considerar que, cuando el origen del conjunto de datos es de varias instituciones, la responsable de la información de origen definirá el tipo de licencia con la que se abrirá el conjunto de datos.

Para efectos de la publicación de los conjuntos de datos, en los metadatos, en el campo Licencias, se debe señalar el nivel de utilización que puede darse a los conjuntos de datos. Para esto cada entidad deberá definir el nivel de licenciamiento de cada conjunto de datos el cual será aprobado por el Comité de Transparencia o el área Jurídica, para esto se deberá establecer los siguientes acuerdos:

- Determinar qué derechos de propiedad intelectual existen en el conjunto de datos.
- Aplicar una licencia adecuada considerando las alternativas que ofrece Creative Commons¹⁵ (definidas internacionalmente), que identifique todos esos derechos y soporte la definición de dato abierto.

Estructura de los metadatos

La importancia que tienen los metadatos en los procesos de apertura de datos está estrechamente ligada a la reutilización, ya que ésta podría mejorarse proporcionando metadatos estructurados considerables que permitan la comprensibilidad de los datos a reutilizar¹⁶. El uso adecuado de los conjuntos de datos abiertos muchas veces se ve obstaculizado porque los metadatos están mal documentados, es decir son incompletos e imprecisos, y por lo tanto la semántica de los datos puede ser ambigua¹⁷.

Los metadatos son el puente entre el conjunto de datos y el usuario (potencial reutilizador), el mismo permitirá entender la estructura del conjunto de datos, así como otros datos relacionados al manejo de la información publicada.

Basados en el Vocabulario de Catálogos de Datos (DCAT)¹⁸ y considerando importante incluir el nivel de anonimización, licencia y periodicidad del dato, por cuanto estos datos son requeridos para entender la semántica del conjunto de datos publicado, se utilizará la siguiente estructura de metadatos:

¹⁵ <https://creativecommons.org>

¹⁶ <https://www.jedem.org/index.php/jedem/article/view/138>

¹⁷ <https://dl.acm.org/doi/abs/10.1145/2463728.2463731>

¹⁸ <https://www.w3.org/TR/vocab-dcat-2/>

Campo	Descripción	Nombre del campo	Obligatorio
Identificador	Identificador único de un conjunto de datos de la organización.	met:identificador	SI
Nombre	Título del conjunto de datos.	met:nombre	SI
Descripción	Descripción textual completa del conjunto de datos.	met:descripcion	SI
Código institución	Sitio web de la organización.	met:institucion	SI
Licencia	Licencia aplicable al conjunto de datos (es decir, condiciones de reutilización).	met:licencia	SI
Fecha de creación	Fecha de creación del conjunto de datos.	met:creacion	SI
Frecuencia de actualización	Frecuencia con la que se actualizan los datos del conjunto de datos. Tiempo Real Diaria Semanal Mensual Anual	met:frecuencia	SI
Archivo maestro	La actualización se realiza a través de un archivo maestro de datos abiertos.	met:archivomaestro	OPCIONAL
Anonimización	Estado de anonimización del conjunto de datos.	met:anonimizacion	SI
Responsable	Correo de contacto del gestor institucional	met:responsable	SI
Lenguaje	Idioma de publicación.	met:lenguaje	SI
Etiquetas	Palabras clave de búsqueda para el conjunto de datos.	met:etiquetas	SI
URI	Dirección permanente del conjunto de datos	met:uri	SI

Tabla 6. Estructura de metadatos

Estructura del Diccionario de datos

El diccionario es una herramienta que contribuye al fortalecimiento de la gestión del conjunto de datos por parte de los usuarios, en la cual se encuentran las descripciones de cada uno de los campos del conjunto de datos. De la misma manera que se publica el conjunto de datos, la especificación del diccionario de datos debe encontrarse en formato abierto.

Para la creación del diccionario de datos se debe considerar lo siguiente:

- No contener tildes en el nombre de los campos.
- Ser lo más claro y detallado posible, logrando que pueda ser interpretado y usado fácilmente.
- Utilizar la codificación estándar definida por INEC (Instituto Nacional de Estadísticas y Censos) para país/provincia/cantón/parroquia, y para la identificación de los ministerios.
- Para los datos utilizar codificación UFT-8.

Aportes a la gestión de la calidad de los datos abiertos

El concepto que permite implementar la gestión de la calidad de datos es el archivo maestro de datos o MDM (de sus siglas en inglés Master Data Management)¹⁹, que considera la calidad como un proceso multidimensional. Este proceso es afectado por el origen disperso de los datos, para lo cual se propone el concepto de MDM como estrategia de creación y apropiación de los datos críticos del negocio, considerando sus diferentes orígenes²⁰.

La estructura del archivo maestro deberá contemplar las características planteadas en la ISO 8000, que son: identificación de estado, origen de los datos almacenados, así como relaciones con otros datos. Esta norma propone que existe una relación directa entre la calidad y la creación y apropiación del concepto de datos maestros dentro de una organización, por cuanto se trata de la recopilación de datos de diferentes orígenes.

La creación del archivo maestro responde a la identificación de las necesidades de los usuarios, que pueden requerir procesos de combinación, identificación y eliminación de duplicados si fuese necesario, para crear los conjuntos de datos requeridos, sin perder la información del origen de los datos.

Aplicación de criterios de calidad de los datos abiertos

La calidad de datos se puede evaluar según diversos criterios. El estándar ISO / IEC 25000: 2014, es un conjunto de normas internacionales para sistemas y requisitos de calidad y evaluación de software (SQuaRE). ISO/IEC 25012 - Data Quality model: define un modelo general aplicable a aquellos datos que se encuentran almacenados de manera estructurada y forman parte de un Sistema de Información; y la ISO/IEC 25024 - Measurement of data quality define las métricas para realizar la medición de la calidad de datos.

El modelo de Calidad de Producto de Datos definido por el estándar ISO/IEC 25012 se encuentra compuesto por las 15 características²¹:

¹⁹ Norma ISO 8000

²⁰ <https://www.researchgate.net/publication/266476656> Chapter 1 - Master Data and Master Data Management

²¹ <https://iso25000.com>

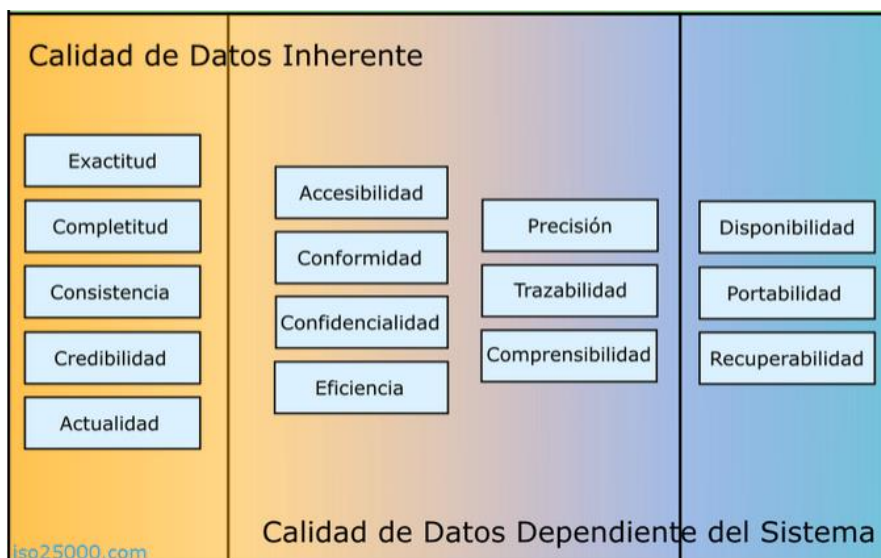


Imagen 3. Calidad de producto de datos

En este contexto, y buscando que los datos abiertos puedan ser entendidos y utilizados, consideramos los siguientes conceptos con el fin de determinar qué criterios pueden ser aplicados al concepto de datos publicados en un portal de datos abiertos.

1. Calidad de datos inherente. Se refiere al grado con que los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas. Desde el punto de vista inherente, se refiere a: valores de los datos con sus posibles restricciones, relaciones entre valores de datos y metadatos.

2. Calidad de datos dependiente del sistema. Se refiere al grado con el que la calidad de datos es alcanzada y preservada a través de un sistema informático cuando los datos son utilizados bajo condiciones específicas. Desde el punto de vista dependiente del sistema, depende del dominio tecnológico en el que los datos se utilizan, y se alcanza mediante las capacidades de los componentes del sistema informático.

Dentro de las características inherente y dependiente, se propone verificar el cumplimiento de los siguientes criterios:

No.	Característica	Descripción
1	Accesibilidad	Grado en el que los datos pueden ser accedidos en un contexto específico, particularmente por personas que necesiten tecnologías de apoyo o una configuración especial por algún tipo de discapacidad.
2	Conformidad	Grado en el que los datos tienen atributos que se adhieren a estándares, convenciones o normativas vigentes y reglas similares referentes a la calidad de datos en un contexto de uso específico.
3	Confidencialidad	Grado en el que los datos tienen atributos que aseguran que los datos son sólo accedidos e interpretados por usuarios autorizados en un contexto de uso específico. La confidencialidad es un aspecto de la seguridad de la información

		(junto con la disponibilidad y la integridad) definida como en ISO/IEC 13335-1:2004.
4	Eficiencia	Grado en el que los datos tienen atributos que pueden ser procesados y proporcionados con los niveles de rendimiento esperados mediante el uso de cantidades y tipos adecuados de recursos en un contexto de uso específico.
5	Precisión	Grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico.
6	Trazabilidad	Grado en el que los datos tienen atributos que proporcionan un camino de acceso auditado a los datos o cualquier otro cambio realizado sobre los datos en un contexto de uso específico.
7	Comprensibilidad	Grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico. Cierta información sobre la comprensibilidad puede ser expresada mediante metadatos.
8	Disponibilidad	Grado en el que los datos tienen atributos que permiten ser obtenidos por usuarios y/o aplicaciones autorizados en un contexto de uso específico.
9	Portabilidad	Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico.
10	Recuperabilidad	Grado en el que los datos tienen atributos que permiten mantener y preservar un nivel específico de operaciones y calidad, incluso en caso de fallos, en un contexto de uso específico.

Tabla 7. Criterios que aportan a la calidad de los datos abiertos

Evaluación de la calidad de los datos previa a la publicación

De acuerdo con los estándares ISO, y como se ha descrito en la sección anterior, uno de los aspectos tener en cuenta son los elementos de la calidad, relacionada con la utilización de los símbolos comprensibles, los metadatos y el diccionario de datos que deben utilizar un vocabulario común, el modelo de datos y los datos maestros vinculados.

Se debe prestar especial atención a la gestión de los conjuntos de datos, recursos y metadatos asociados que están publicando actualmente en un portal de datos abiertos, los problemas que enfrentan los portales de datos abiertos están relacionados con la falta de estándares de datos, la dificultad en el acceso a los datos y la poca capacidad de comprensión.

Basados en las normas ISO 25000 y el modelo de calidad de datos²², y en buenas prácticas para medir la calidad de los datos²³, la evaluación se realizará en función de la comprensibilidad del conjunto de datos por parte del usuario, en los siguientes aspectos:

²² <https://iso25000.com/>

²³ <https://blog.syncsort.com/2018/02/data-quality/how-to-measure-data-quality-7-metrics>

Métrica	Definición	Cómo calcular
Proporción de errores	¿Cuántos errores tiene en relación con el tamaño de su conjunto de datos?	Divida el número total de errores por el número total de elementos.
Cantidad de valores vacíos	Los valores vacíos indican que falta información en un conjunto de datos.	Cuente el número de campos que están vacíos dentro de un conjunto de datos.
Tasas de error de transformación de datos	¿Cuántos errores surgen al convertir la información a un formato diferente?	¿Con qué frecuencia los datos no se convierten con éxito?
Cantidades de datos oscuros	¿Cuánta información es inutilizable debido a problemas de calidad de datos?	Mire qué cantidad de sus datos tiene problemas de calidad de datos.
Tasas de devolución de correo electrónico	¿Qué porcentaje de destinatarios no recibió su correo electrónico porque fue a la dirección incorrecta?	Divida el número total de correos electrónicos que rebotaron por el número total de correos electrónicos enviados, luego multiplique por 100.
Tasa de campos con codificaciones definidas	¿Qué porcentaje de registros que están dentro de los rangos de las codificaciones establecidas?	Relación de los registros que están dentro de los rangos de las codificaciones establecidas en relación con el total de registros
Tasa de campos con simbología inapropiada	¿Qué porcentaje de registros tienen caracteres especiales?	Relación del número de registros con caracteres especiales en relación al total de registros

Tabla 8. Métricas para la calificación de criterios de calidad de los datos abiertos

FASE DE PUBLICACIÓN

Los recursos que serán puestos a disposición de los usuarios a través de los portales web, son:

1. El recurso digital con los datos, que sea de acceso sin restricciones
2. El recurso que contenga los metadatos
3. El recurso del diccionario de datos

Plataforma gubernamental de publicación de datos

La publicación se realiza en una plataforma de código abierto que facilita la publicación de catálogos de datos, y es el lugar donde se puede encontrar los datos generados por las entidades dependientes de la Administración Pública Central.

En el catálogo se puede realizar la búsqueda de datos por palabras clave, nombre de entidad, categoría, etiqueta o formatos. Cada conjunto de datos publicado tiene asociado una ficha de metadatos que proporciona el detalle de su contenido.

Interfaz de Programación de Aplicaciones (API)

Es el punto de entrada donde un sistema recibe peticiones para ejecutar una tarea previamente programada, facilitando el trabajo de los reutilizadores de datos que buscan crear nuevos productos o servicios basados en los datos abiertos.

Es recomendable que la existencia de las API abiertas no sustituya el acceso a través de descargas directas, sino complementarlo para mejorar el acceso a la información.

Conjuntos de datos

Los datos son publicados en unidades llamadas “datasets” o conjuntos de datos. Cuando un usuario realiza búsquedas, los resultados serán datasets individuales.

Los conjuntos de datos se construyen a partir de datos digitalizados y que se encuentran disponibles en una URL.

Recursos

Posterior a la creación del conjunto de datos con sus correspondientes metadatos, se procede a la inclusión de uno o más recursos, es decir la incorporación de los archivos que cuentan con los registros que alimentarán el conjunto de datos, mismo que puede ser un fichero de cualquier tipo, un enlace a un fichero en otro lugar de la red o un enlace a una API.

Usuarios de la plataforma y roles

La información disponible en la plataforma de publicación de datos es de libre acceso y utilización, y no es necesario registrarse para buscar y acceder a los datos; no obstante, para crear o editar conjuntos de datos es necesario disponer de una cuenta para controlar el acceso a la plataforma.

La administración de la plataforma gubernamental de datos abiertos es de responsabilidad de la Secretaría Técnica de Planificación, quien creará usuarios individuales por cada entidad de la Administración Pública Central.

Corresponde a los Gestores institucionales manejar las claves de ingreso que se refieren a las credenciales de acceso.

Se identifican los siguientes tipos de roles dentro de la institución:

1. Administrador: su rol permite añadir, eliminar y cambiar roles para los miembros de la institución.

2. Editor: su rol permite coordinar internamente la creación, edición y publicación de los datos abiertos de la entidad.

Las entidades de la Administración Pública Central, tienen a su cargo un gran conjunto de datos, y para gestionar su ingreso es necesario que se registre el login de un único responsable, mediante credenciales. El editor inicialmente creará el conjunto de datos en privado y visible; y posteriormente podrá ser publicado para su visualización y acceso.

3. Participante: su rol le permite ver los conjuntos de datos privados en la organización.

IV. PROMUEVE

Objetivo

El rol de las entidades del Estado en el uso de datos abiertos no termina con su publicación. También deben desempeñar un papel activo para apoyar una reutilización eficaz e innovadora de los mismos.

Difusión para la promoción de datos abiertos

Los datos abiertos cobran vida gracias a los usuarios, es necesario que los datos puedan ser encontrados con facilidad de manera que se promueva la generación de comunidades tales como desarrolladores de software, visualizadores, periodistas de datos, investigadores, colectivos sociales, entre otros, para que puedan reutilizarse y generar valor a partir de ellos.

Generar comunidad o una audiencia definida será una tarea central que las entidades deberán considerar en sus planes institucionales, para poder promover el uso de los datos.

Es necesario que la promoción y difusión de los datos involucre al menos:

1. Identificar sus usuarios y potenciales usuarios, para ello será importante identificarlos aliados de otros sectores (sociedad civil, academia, sector privado, agencias de cooperación, etc.) y generar vínculos con los mismo
2. Incorporar en las estrategias comunicacionales la promoción de los datos institucionales.
3. Generar contenidos tales como infografías, videos, cápsulas, artículos, sobre los datos disponibles de la entidad.
4. Vincular a distintos actores a través de acciones que permitan el encuentro e interacción, que promuevan la reutilización de los datos.
5. Promover el uso de tecnologías abiertas emergentes para generar productos o servicios a partir de los datos abiertos, en conjunto con los distintos actores.

Para lograrlo, se debe implementar estrategias lo cual incluye desarrollar al menos los siguientes pasos por parte de la entidad:

Segmentar y caracterizar a los potenciales usuarios.

1. Desarrollar un plan de comunicaciones interno que promueva la oferta, así como la demanda, facilitando la reutilización y consumo de datos; y sensibilizando sobre los beneficios.
2. Identificar los canales de difusión disponibles, de acuerdo con la realidad de la institución, buscando establecer los medios de participación y colaboración para la gestión de ideas y/o propuestas, con el fin de poder acercar a la sociedad civil y a la ciudadanía en general.
 - Hacer uso de redes sociales, en función de la población objetivo, para difundir la información disponible. Usar el perfil institucional de la entidad para la publicación.
 - Hacer uso de las bases internas de usuarios y enviar e-mailing con publicidad sobre la información que va a ser publicada por la entidad.
 - Establecer, dentro de las prioridades, procesos de capacitación dentro de la institución y hacia los infomediarios, academia, sector privado, sociedad civil, etc., sobre los datos que se están publicando.
 - Generar campañas informativas con medios de comunicación sobre los datos publicados.
 - Determinar y habilitar espacios de participación, ya sean físicos o virtuales donde se realicen iniciativas encaminadas a la utilización efectiva de los datos buscando la participación e integración entre la institución, sociedad civil, gobierno, academia, sector privado. Se deberá realizar cada año, al menos una de las siguientes opciones que se detallan a continuación:
 - Concursos colegiales, universitarios de uso y aprovechamiento de la información, que motive a los estudiantes a hacer trabajos con los datos de la institución.
 - Webinars que hablen sobre el uso de datos, y visibilice estudios de casos que han utilizado los datos de la institución.
 - Gamificación que motive la utilización y reutilización de los datos de la institución a través de incentivos y/o reconocimientos.
 - Congresos o seminarios que se centren en el uso de datos, y visibilice estudios de casos que han utilizado los datos de la institución.
 - Socialización de estudios internos de la entidad con la utilización de sus datos.

V. EVALÚA Y PERFECCIONA

Objetivo

Medir cualitativa y cuantitativamente los resultados de las acciones emprendidas y el cumplimiento versus lo planificado.

Se establece evaluar: la percepción de los usuarios si los datos publicados son adecuados para su uso, la publicación versus la planificación del portafolio institucional de datos abiertos, el nivel de descargas de los conjuntos de datos.

Percepción de los usuarios de los datos publicados

La institución deberá realizar un plan de evaluación de calidad de conjuntos de datos por parte de los usuarios con el fin de validar que los datos publicados tiene criterios de calidad, se propone los siguientes indicadores²⁴.

Indicador	Descripción
Porcentaje entre el número de valores de datos presentados por símbolos c y el número de datos que se solicita la revisión de los símbolos.	Comprensión de los símbolos. Porcentaje de uso de símbolos comprensibles.
Porcentaje del conjunto de datos que se entiende la descripción frente a los que no se entienden.	Comprensión semántica. Relación entre el vocabulario común reconocido y los términos o definiciones utilizados en el diccionario de datos.
Porcentaje de comprensión del diccionario de datos. Es una relación entre el número de valores de datos definidos en el diccionario de datos utilizando identificados con el uso de un vocabulario común y los valores de datos definidos en el diccionario de datos.	Comprensión del diccionario de datos. Relación de comprensión de los términos utilizados en un vocabulario común.
Porcentaje entre el número de metadatos entendidos frente a los metadatos de interés para los reutilizadores potenciales en un contexto específico de uso.	Comprensión de los metadatos. Relación entre la comprensión del significado de los elementos de datos.
Porcentaje de conjuntos de datos que son entendidos por la licencia, su aplicación en relación con los que no son entendidos.	Comprensión de la licencia de uso. Relación entre el conocimiento del uso que se le puede dar a una licencia y el espectro de aplicación.

²⁴ https://link.springer.com/chapter/10.1007/978-3-030-40907-4_17

Porcentaje de criterios de anonimato entendidos frente a los no entendidos.	Comprensión de las normas de anonimato y su aplicación.
---	---

Tabla 9. Indicadores para evaluar la percepción de calidad de los datos publicados

La publicación versus la planificación del portafolio institucional de datos abiertos

El Equipo de datos abiertos deberá evaluar el cumplimiento del portafolio comparando lo publicado versus lo planificado.

Se considere las solicitudes de publicaciones de datos abiertos, y el estado en que se encuentran, incluyendo aquellas que van a formar parte del nuevo portafolio especificando el estado en el que se encuentran (en proceso, incluida, no considerada).

Nivel de descarga de conjuntos de datos

Incluir indicadores numéricos de descargas realizadas en los medios donde se encuentren disponibles, como portales institucionales y portal gubernamental, establecer el ranking de conjuntos de datos más descargados, e identificar oportunidades de mejora tanto en la disponibilidad como en la calidad de los datos.

Generación de informes anuales

El Equipo de datos abiertos deberá elaborar y enviar, anualmente, al ente Rector de la Política, los primeros 15 días del año siguiente la evaluación de la implementación de la Política y de la Guía de Datos Abiertos, que incluya:

1. Equipo responsable.
2. Portafolio institucional anual de datos abiertos, y sus actualizaciones.
3. Conjuntos de datos publicados vs. los planificados
4. Resumen de acciones para la difusión y utilización de los datos publicados
5. Resumen de la evaluación de la percepción de los usuarios sobre los datos publicados.
6. Indicadores de uso y descarga de datos
 - Nivel de descarga de conjuntos de datos.
 - Solicitudes de apertura de conjunto de datos recibidas y atendidas.
 - Casos de utilización y reutilización de datos que ha registrado la institución
 - Otros estudios o reportes de terceros respecto al uso de datos abiertos de la institución
7. Conclusiones y recomendaciones para la institución y para el ente rector.

ANEXOS

HERRAMIENTA DE PLANIFICACIÓN

Portafolio institucional de datos abiertos (ejemplo sujeto a ajustes)

INSTITUCIÓN		Portafolio de datos abiertos			Versión (Actualización): 01 Código Documento: PDA-2021-v01			
Fecha de Elaboración		01/12/2020						
Período		2021						
Responsable		Juan Pérez						
E-mail Responsable		juan.perez@xx.gob.ec						
Código	Nombre	Descripción	Entidades que generan la información de origen	Fecha prevista de publicación	Formatos de publicación	Licencia	Solicitante	Categorías
A-IN-001	Censo de Información Ambiental Económica en Gobiernos Autónomos Descentralizados Municipales 2016	Contiene información de la gestión ambiental, riego y drenaje, fomento y desarrollo productivo, y gastos e inversión en gestión ambiental, de todos los Gobiernos Autónomos Descentralizados Provinciales del Ecuador	GADX, INEC	Mayo 2021	csv	CC-BY-4.0		Ambiente y vivienda
A-IN-002	Censo de Población y Vivienda Galápagos 2018	Contiene información de todas las viviendas existentes y de la población que reside habitualmente en la provincia	INEC	Septiembre 2021	csv	CC-BY-4.0		geográficos

Anexo sobre metodologías de priorización de datos

Otras metodologías y sugerencias que se pueden usar son:

- <https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=o-que-e>
- <https://drive.google.com/file/d/1hYmoTZTDgFe9E8CtxAW6qfbj5W93DL/view>,
- Ejemplo de apertura de datos con propósito son las guías que se construyen para resolver problemáticas comunes: <https://docs.google.com/document/d/182USj4La896XgMZ-LTw7A3HYbA5i7aP0p2JL8a-24/edit#heading=h.9vgybo4y3gv2>
- Es importante que se estudie lo que debería generarse y tal vez no está dentro de la demanda actual de usuarios de datos. Es importante saber qué sí se genera y lo que no se está generando.
- Consulta a las personas consumidoras de datos e información
- Buzón de sugerencias de los usuarios
- Talleres con expertos y líderes de las organizaciones de la sociedad civil, que tengan proyectos o procesos relacionados a la apertura de datos
- Open Data Guide podría ser un referente a analizar. Comparto el link: <https://opendata.guide/chapter2.html>
- Existe una experiencia muy interesante que se ha aplicado en México y se llama Rally de Datos en la Calle, la cual consiste en organizar a la ciudadanía en grupo de 3 personas aproximadamente para que puedan salir a las calles a verificar y constatar la ejecución de políticas públicas de manera directa, es una figura similar a las veedurías ciudadanas que existen en Ecuador.
- Es importante primero determinar criterios de priorización de los datos (impacto y dificultad) y ponderación de estos criterios con lo cual nos permitirá obtener los plazos de publicación de los datos si es a corto, mediano o largo plazo.
- Se podría construir una matriz de interesados, donde se incluya los interesados, la información que requieren, la urgencia, la frecuencia con que requieren los datos, para que serán usados los datos... Y la unidad responsable de entregar la información especifique la importancia y celeridad con que atiende cada solicitud (bajo, medio alto).
- Todos los temas que tengan relación con los presupuestos que administran las entidades del Estado y las privadas que reciben fondos públicos; así como otros temas relacionados con los servicios; esto es: remuneraciones, servicios que ofrecen las entidades, presupuesto institucional, contrataciones públicas, proyectos y programas institucionales, créditos internos o externos, viáticos nacionales e internacionales, entre las principales.

HERRAMIENTA DE PREPARACIÓN

Referencia de licencias para publicación de datos abiertos

<https://creativecommons.org/licenses/by-sa/4.0/legalcode.es>

Metadatos (ejemplo sujeto a ajustes)

LOGO		Tabla de Metadatos	Institución: Versión (Actualización): 01 Código Documento: TM-01-2020-V01
Código Documento Referencia:		HR-2020-V01	
No.	Tipo de Recurso	Tipo de recurso (data sets)	
1	<i>Título o Nombre del Conjunto de Datos:</i>	Directorio Institucional	
2	<i>Descripción</i>	Contiene información referente al directorio completo de la institución, con detalles de funcionarios, dirección, teléfonos, emails, exceptuando el puesto institucional.	
3	<i>Institución</i>		
4	<i>Fecha de Publicación</i>		
5	<i>Fecha de Modificación</i>		
6	<i>Frecuencia de actualización</i>	Mensual	
7	<i>Versión</i>	1	
8	<i>Formato</i>	csv	
9	<i>Idioma</i>	ESPAÑOL	
10	<i>Fuente</i>	Gerencia Coordinadora de Servicios Corporativos / Jefatura de Servicios Generales - Administración, Base de Datos Sistema de Recursos Humanos, Directorio Telefónico	
11	<i>Licencia</i>	CC-BY-4.0	
12	<i>Etiquetas</i>	sectores estratégicos, recursos humanos, personal, cargo, unidad, teléfono, email, directorio	
13	<i>Email de contacto de la Fuente</i>		
14	<i>URI</i>		

Diccionario de datos (ejemplo sujeto a ajustes)

LOGO		Diccionario de Datos Nro. 01	Institución: Versión (Actualización): 01 Código Documento: DD-01-2020-V01
Código Documento Referencia:		HR-2020-V01	
Nombre del Conjunto de Datos:		Directorio Institucional	
Nombre del Recurso:		Directorio institucional con detalles para su contacto y localización	
Descripción del Recurso:		Se puede conocer el detalle del directorio de la institución, con información de nombres de funcionarios, dirección, teléfonos, emails, excepto el puesto institucional.	
No.	Nombre del Campo (Encabezado Columna)	Descripción del Campo	
1	No.	Número de ítem o funcionario	
2	Apellidos y Nombres de los servidores y servidoras	Detalles de nombres y apellidos del funcionario o servidor público	
3	Unidad a la que pertenece	Unidad departamental a la que pertenece el funcionario dentro de la institución	
4	Dirección institucional	Dirección física de la institución	
5	Ciudad en la que labora	Ciudad en la que desempeña sus labores	
6	Teléfono institucional	Número de teléfono o PBX de la institución	
7	Extensión telefónica	Número de extensión telefónica del funcionario dentro de la institución	
8	Correo Electrónico institucional	Dirección de correo electrónico (email) del funcionario dentro de la institución	

HERRAMIENTA PARA LA PUBLICACIÓN

Categorías

Las categorías involucran a los sectores donde los datos puedan causar impacto o responden a intereses comunes. Los conjuntos de datos pueden asociar una o más categorías, entre las siguientes:

- 1 Administración Pública
- 2 Ambiente y agua
- 3 Sociedad y Bienestar
- 4 Cultura y Deporte
- 5 Comercio e Industria
- 6 Conocimiento e Investigación
- 7 Estadísticas demográficas
- 8 Energía
- 9 Economía y Finanzas
- 10 Educación
- 11 Geográficos
- 12 Salud
- 13 Seguridad y Defensa
- 14 Transporte
- 15 Vivienda
- 16 Tecnología y telecomunicaciones
- 17 Turismo

Las categorías se usan para valorar los datos abiertos en un proyecto concreto o un tema en particular. Pueden ser:

- Los Objetivos de Desarrollo Sostenible
- Objetivos del Plan Nacional de Desarrollo
- Datos para la Transparencia y ética en la administración pública
- Sectores
- Datos para la Contratación Pública
- Datos de investigación científica

REFERENCIAS

- Formatos de datos abiertos, <https://datosabiertos.jcyl.es/web/es/catalogo-datos/formatos-datos-abiertos.html>
- Formatos de archivos, <http://opendatahandbook.org/guide/es/appendices/file-formats/>
- Cinco estrellas de los datos abiertos, <https://5stardata.info/es/>
- Saquicela V, Baculima F, Lupercio L., Datos enlazados de información geográfica del Ecuador (julio-diciembre 2018), <https://pdfs.semanticscholar.org/36d2/5f8f743aabf2eb80719bb9adc0787418c85a.pdf>
- Ecuadorian Geospatial Linked Data, https://www.researchgate.net/publication/260983879_Ecuadorian_Geospatial_Linked_Data
- Agencia de Gobierno Electrónico y Sociedad de la Información de Uruguay AGESIC, <https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/guia-para-publicacion-datos-abiertos-gobierno>
- Banco Mundial, Opendat toolkit, <http://opendat toolkit.worldbank.org/es/essentials.html#:~:text=Definici%C3%B3n%20de%20Datos%20Abiertos,-El%20t%C3%A9rmino%20E2%80%9CDatos&text=Los%20Datos%20Abiertos%20tienen%20dos,uso%20y%20con%20m%C3%ADnimas%20restricciones>
- National Information Standards Organization NISO, <https://www.niso.org/publications/understanding-metadata-2017>
- Cadena S, Norberto S, Fuster A. Defining a Master Data Management Approach for Increasing Open Data Understandability, https://link.springer.com/chapter/10.1007/978-3-030-40907-4_17
- Best practices for publishing linked data, <https://www.w3.org/TR/ld-bp/>
- Research guides: semantic web and linked data, best practices, <https://guides.library.ucla.edu/semantic-web/bestpractices>

CO-CREACIÓN

Equipo MINTEL-REDAM, coordinador del proceso de co-creación

Nombre	Sector
Andrea Llerena	Sociedad Civil
Francisco Silva	Gobierno
Johanna Pazmiño	Gobierno
Juan Carlos Castillo	Gobierno
Julio López	Sociedad Civil
Margarita Yépez	Sociedad Civil
Robert Enríquez Reyes	Academia
Susana Cadena Vela	Academia