

datos.gob.es
**Manual práctico para mejorar la
calidad de los datos abiertos**

Iniciativa
aporta



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

ÍNDICE

Introducción

1. ¿Cómo se mide la calidad de los datos abiertos?

- 1.1. Definiendo la calidad de los datos abiertos
- 1.2. Fallos frecuentes de calidad en los datos abiertos
 - 1.2.1. Buscadores de datos excesivamente guiados o acotados
 - 1.2.2. Datos incompletos o limitados
 - 1.2.3. Datos fragmentados y de difícil acceso
 - 1.2.4. Datos obsoletos y sin actualizar
 - 1.2.5. Metadatos actualizados para datos desactualizados
 - 1.2.6. Formatos no reutilizables o no estandarizados
 - 1.2.7. Modelos *freemium* para el acceso a los datos
 - 1.2.8. Falta de claridad con las licencias

2. La calidad de los contenidos

- 2.1. Datos por la rendición de cuentas
 - 2.1.1. Presupuesto nacional
 - 2.1.2. Gasto público
 - 2.1.3. Propiedad inmobiliaria y de las tierras
 - 2.1.4. Resultados electorales
 - 2.1.5. Legislación
- 2.2. Datos para las políticas sociales
 - 2.2.1. Censo y estadísticas nacionales
 - 2.2.2. Contaminación medioambiental
 - 2.2.3. Funcionamiento de los servicios públicos básicos
- 2.3. Datos para la innovación
 - 2.3.1. Mapa nacional
 - 2.3.2. Límites administrativos y localizaciones
 - 2.3.3. Contratación pública
 - 2.3.4. Registro de empresas

3. Referencias

4. Gráficos del informe

INTRODUCCIÓN

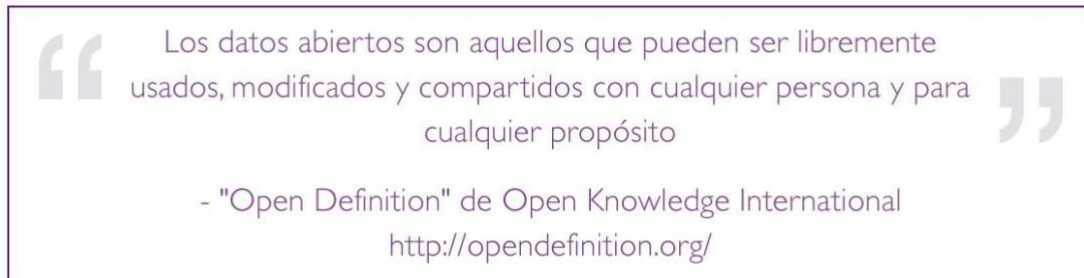
Hoy en día disponemos cada vez de más fuentes de datos a nuestro alcance. Sin embargo, paradójicamente, aun cuando los datos son más asequibles que nunca esto no quiere decir que sean necesariamente **más fáciles de usar**. Los potenciales usuarios de esos datos tienen que hacer frente muchas veces a **múltiples barreras** que dificultan el acceso y uso de esos datos.

Desde **datos incorrectos a formatos inadecuados** pasando por ficheros de difícil acceso, conjuntos de datos divididos entre múltiples localizaciones u otros imposibles de descifrar, e incluso en ocasiones contradictorios entre sí. Este es el panorama a veces desolador que nos muestran **algunas de las conclusiones** de varios estudios de referencia como el [Open Data Barometer](#) o el [Global Open Data Index](#), y que, en parte, ha estado también influenciado por el convencimiento inicial de que lo importante era **publicar la mayor cantidad de información** cuanto antes sin importar su estado real.

En este informe estableceremos una **definición y requisitos mínimos** de lo que podemos considerar como *calidad de los datos abiertos* para facilitar su posterior medida. Asimismo, haremos un recorrido por los errores más comunes a la hora de publicar datos abiertos de calidad, ofreciendo también **recomendaciones prácticas** para evitarlos. Y finalmente nos centraremos en cómo mejorar los contenidos de algunos de los conjuntos de datos más relevantes que se pueden publicar desde los gobiernos, incluyendo detalles sobre los **puntos fuertes y débiles** de los datos publicados actualmente, así como recomendaciones sobre los estándares de referencia a seguir junto a ejemplos de conjuntos de alta calidad publicados por otros países.

1. ¿CÓMO SE MIDE LA CALIDAD DE LOS DATOS ABIERTOS?

Todos conocemos ya las características básicas con las que deben contar los datos cuando los publicamos para poder ser considerados **datos abiertos**, y que según la última versión de la [open definition](#) se resumen como:



Si analizamos estas características más en detalle podremos comprobar cómo se han de cumplir al menos tres requisitos básicos ya ampliamente conocidos, como se muestra en el siguiente gráfico¹:



Gráfico 1: Los tres requisitos básicos de los datos abiertos

Como vemos, estas características están orientadas principalmente a facilitar el acceso a los datos y su posterior utilización. Sin embargo, debemos ir un paso más allá y esforzarnos no sólo en **publicar grandes cantidades de datos**, sino también en

¹ Las recomendaciones de este informe siguen las pautas indicadas por la *open definition*, aunque en las normativas de reutilización del sector público Europea y Española se permiten algunos cobros marginales.

publicar datos de gran calidad. Si no somos capaces de asegurar una calidad al menos aceptable en los datos que publicamos estaremos añadiendo una nueva barrera que complicaría el uso de esos datos por parte de terceros, que al final es lo que pretendemos fomentar con la apertura de datos.

1.1. Definiendo la calidad de los datos abiertos

Dado que no existe una definición única de a qué nos referimos cuando hablamos de la **calidad de los datos** en general, nos fijaremos en algunas referencias clave para intentar crear nuestra propia definición en el caso de los datos abiertos en particular:

1. Los [principios originales de los datos abiertos gubernamentales](#) definidos en la denominada *reunión de Sebastopol*.
2. Los [principios definidos por la Carta Internacional de los Datos Abiertos](#).
3. Las [dimensiones de la calidad de los datos abiertos](#) definidas por la iniciativa *Open Data Support* de la Comisión Europea.
4. Los [atributos de calidad inherente a los productos de datos](#) definidos por la *norma ISO/IEC 25012*.

Si observamos las distintas características que figuran en cada una de las referencias anteriores podemos obtener un patrón común que, sumándose a las características ya indicadas anteriormente por la *open definition*, nos servirá a la hora de establecer nuestra propia definición básica de lo que significa la calidad en el caso de los datos abiertos a través de los siguientes indicadores:

- **Exactitud** de los datos respecto a cada una de las entidades a las que representan en el mundo real.
- **Consistencia** de los datos y ausencia de contradicciones, siendo coherentes respecto a los otros datos existentes en el mismo contexto de uso.
- Garantía de **disponibilidad**, tanto en un momento puntual, como a largo de amplios periodos de tiempo y de forma indefinida.

- **Compleitud** de los datos en cuanto a todos los atributos esperados para la entidad que está siendo representada.
- **Conformidad** respecto a los estándares, reglas, convenciones y normativas de referencia establecidos para la captura y publicación de los datos.
- **Credibilidad** de las fuentes de información utilizadas, garantizando además la veracidad respecto al origen de los datos y su trazabilidad.
- **Precisión** de los datos disponibles con los niveles de detalle y granularidad adecuados para ser relevantes en el área de conocimiento de la que tratan.
- **Actualidad** de los datos, reflejando el estado actual de los mismos y estando disponibles a tiempo y sin retrasos que afecten a su relevancia.
- **Comprensibilidad**, expresando los datos de forma que se puedan interpretar inequívocamente a través de los metadatos y documentación disponibles.

Indicadores de la calidad de los datos



Gráfico 2: Indicadores de la calidad de los datos

1.2 Fallos frecuentes de calidad en los datos abiertos

Una vez establecidas nuestras métricas de calidad para los datos abiertos, a continuación, repasamos algunos de los errores más frecuentes que se cometen para cada uno de los atributos de apertura y calidad anteriormente mencionados, así como algunas recomendaciones que nos ayudarán a evitarlos de ahora en adelante. Estos consejos serán aplicables no sólo a cualquier administración, sino también a otras empresas públicas y contratadas a cargo de servicios públicos, y que por tanto gestionan también datos gubernamentales.

Fallos frecuentes de calidad en los datos abiertos



Gráfico 3: Fallos frecuentes de calidad en los datos abierto

1.2.1. Buscadores de datos excesivamente guiados o acotados

Muchas veces cuando intentamos acceder a unos datos la única opción que nos encontramos es un **simple campo de búsqueda**, sobre todo para ciertos tipos de datos gestionados por registros públicos. Esto generalmente no es suficiente para garantizar el acceso a los datos, independientemente de si contamos con más o menos opciones de filtrado.

La búsqueda puede facilitar el acceso cuando se tienen ya ciertas nociones de lo que se está buscando, sin embargo, **los buscadores muchas veces pueden suponer una barrera en lugar de una ayuda** cuando lo que se quiere es acceder a todo el conjunto de datos o a una parte substancial de él. Esto es debido a que generalmente ofrecen sólo datos parciales y muchas veces es necesario también tener información previa sobre lo que queremos buscar (*identificadores, números de registro, nombres, códigos, titulares...*) para conseguir realizar una búsqueda efectiva.

Un problema relacionado es también el uso de formularios guiados a modo de “*asistente*”, en el que el usuario tiene que ir interactuando a lo largo de varios pasos y seleccionar una serie de filtros u opciones para poder explorar y acceder a las distintas partes de los datos que nos puedan interesan (opción bastante común por ejemplo en los sistemas que gestionan datos estadísticos). El problema aquí nuevamente es que no se concede suficiente libertad al usuario para poder acceder

a los datos según sus necesidades y preferencias, **teniendo que adaptarse forzosamente a lo que le ofrezca el sistema.**



Características de calidad afectadas: Accesibilidad y Disponibilidad.



Recomendaciones: No es en absoluto necesario, ni siquiera recomendable, prescindir totalmente de los buscadores, pero éstos se deben complementar con una opción fácilmente localizable que nos permite hacer una **búsqueda completa** que incluya todos los posibles valores (no se debe utilizar una búsqueda sin haber introducido ningún campo únicamente como sucede muchas veces en la práctica, ya que la mayoría de los usuarios ni siquiera se dará cuenta de que esa opción existe).

Otra alternativa complementaria sería añadir una opción adicional para acceder directamente a toda la información en bruto y/o poder descargar.

1.2.2. Datos incompletos o limitados

Es bastante frecuente que existan datos públicos en alguna materia determinada, pero que al mismo tiempo estos estén disponibles de una forma tan limitada que no tiene sentido considerar que dicha información sea realmente abierta.

Por ejemplo, los datos podrían estar **disponibles únicamente para un periodo muy limitado de tiempo** debido a que se realiza una única publicación aislada en el tiempo que nunca más se repitió (por ejemplo, datos que cubren únicamente un periodo de dos meses durante los dos últimos años)

En otros casos también sucede que los datos se han venido **publicando de una forma muy irregular** a lo largo del tiempo, dando lugar a que algunos periodos estén disponibles y otros no (por ejemplo, sólo están disponibles los datos sobre cinco trimestres diferentes en los últimos tres años). Los datos podrían también estar siendo publicados de forma más continua pero únicamente con **información muy básica que aporta poco valor** y no cumple con los mínimos necesarios para obtener

una representación completa de las entidades que buscamos (ver sección 2 del informe acerca de la calidad de los contenidos)



Características de calidad afectadas: Completitud, Disponibilidad, Actualidad, Credibilidad y Precisión.



Recomendaciones: Una vez se toma la decisión de publicar un conjunto de datos debe prepararse también un **plan de publicación** para garantizar que ese conjunto seguirá actualizándose y estando disponible a lo largo del tiempo.

Por otra parte, debemos también mejorar las rutinas de publicación de los datos y asegurarnos de que cuando publicamos datos los hacemos siempre siguiendo los estándares y referencias existentes en cuanto a los **modelos de datos** a seguir, para así poder garantizar que la información publicada será completa y útil.

Además, **publicar siempre los datos en crudo** desagregando la información al mayor nivel de detalle posible y asegurarse de incluir la desagregación también en el propio proceso de recogida de datos.

1.2.3. Datos fragmentados y de difícil acceso

No es poco frecuente que los datos que buscamos en realidad existan y estén disponibles en alguna parte, pero sean muy difíciles de encontrar.

Sucede con frecuencia que los datos están **divididos** y **distribuidos** a lo largo de distintas secciones o páginas dentro de una web institucional, o incluso de distintos sitios web que corresponden a las distintas agencias que trabajan con ellos. En otras ocasiones es la propia estructura del sitio la que nos dificulta el acceso a aquello que buscamos, debido al uso de índices incompletos, etiquetado poco intuitivo o enlaces rotos con demasiada frecuencia.

A esto se añade un problema adicional que es la existencia de **múltiples versiones de los datos replicadas en distintos sitios oficiales** y que cuentan además con distintas características, lo que puede fácilmente confundir a los usuarios.

Otra variante de este problema sucede cuando hay ciertos datos disponibles online pero **el acceso se restringe a un grupo determinado de personas**, por ejemplo, datos que sólo estarán disponibles para cierto personal del gobierno o a los que sólo es posible acceder si eres habitante de una determinada región o ciudad y te identificas como tal.



Características de calidad afectadas: Accesibilidad, Consistencia, Completitud y Credibilidad.



Recomendaciones: Mejorar la *usabilidad* de los sitios en general y la organización de los contenidos y el etiquetado en particular, establecer inventarios y catálogos centralizados para facilitar el acceso a los activos de datos disponibles y establecer conexiones entre los distintos conjuntos de datos para visibilizar las conexiones existentes entre ellos.

Utilizar los **metadatos** siempre, de forma completa y adecuada y también proporcionándolos en un formato legible por las máquinas, como por ejemplo usando [DCAT](#) o [DCAT-AP](#) y [GeoDCAT-AP](#), según lo descrito en la [guía de aplicación de la Norma Técnica de Interoperabilidad](#) en su apartado 6.2 sobre *descripción de información reutilizable*. Esto servirá para facilitar la *encontrabilidad* de los datos automáticamente a través de motores de búsqueda, *agregadores* u otras herramientas.

Eliminar además cualquier restricción de acceso a los datos que no venga impuesta por motivos de **privacidad o seguridad** de la información.

1.2.4. Datos obsoletos y sin actualizar

En algunos casos los datos que se publican están claramente obsoletos bien por tratarse de **una única publicación puntual que nunca más se actualiza** o bien por haberse abandonado las actualizaciones después de haberse mantenido los conjuntos de datos de forma regular por un tiempo inicial, ya sea como consecuencia de un cambio de estrategia o por falta de recursos (por ejemplo, cuando los datos están disponibles para una década entera, pero cuya última actualización cuenta con cinco o más años de antigüedad).



Características de calidad afectadas: Disponibilidad, Completitud y Actualidad.



Recomendaciones: Nuevamente en este caso la solución pasa por preparar un plan de publicación que garantice la actualización futura de los datos, indicando claramente en que periodos y cómo se llevarán a cabo las distintas actualizaciones.

Por otro lado, **la gestión y el mantenimiento de los datos** (no necesariamente la publicación) debería estar también **descentralizada y distribuida** por los distintos departamentos para no tener que depender de un único proyecto o equipo que la haga extremadamente vulnerable a cualquier cambio en el mismo.

1.2.5. Metadatos actualizados para datos desactualizados

Un problema también muy frecuente consiste en que los metadatos puedan mostrar que la fecha de última actualización es bastante reciente, sin embargo, cuando accedemos a los datos nos encontramos con que son **mucho más antiguos de lo que se indica en esos metadatos** y que no se han actualizado hace bastante tiempo.

Esto generalmente es un problema asociado a las herramientas que utilicemos para gestionar el catálogo y sucede porque los metadatos acerca de la última actualización están configurados para que se actualicen automáticamente cada vez que se edita un

registro determinado (o incluso en ocasiones simplemente la propia página) aunque los datos no hayan variado en absoluto.



Características de calidad afectadas: Exactitud, Consistencia y Actualidad.



Recomendaciones: Los metadatos de un conjunto de datos deben hacer siempre referencia a las características de los datos a los que acompañan, **nunca a las de los registros** que los contienen. En caso de que se necesario proporcionar metadatos también sobre los registros se debería hacer de manera separada y claramente diferenciada para evitar cualquier tipo de confusión en los usuarios de los datos.

1.2.6. Formatos no reutilizables o no estandarizados

Es muy frecuente encontrarnos dentro de la administración con una gran cantidad de informes de todo tipo en formato textual que a su vez están basados en un análisis previos de los datos disponibles. Sin embargo, muy pocas veces se publican esos datos junto al informe y cuando se hace suele ser simplemente como parte del texto del propio informe junto al resto de información, lo que hace muy difícil que puedan ser reutilizados de alguna forma.

Por otra parte, **es también frecuente que se utilicen algunos formatos no estandarizados, no abiertos** o simplemente **poco comunes** para distribuir la información o los datos, sin dar tampoco otras alternativas a los usuarios, lo que en la práctica reduce claramente el público objetivo al que nos podemos dirigir.

Finalmente, podemos encontrar errores *físicos* en los propios ficheros que a veces están *corruptos* y no se pueden abrir, o en la codificación de caracteres que se utiliza para darles formato, que bien no se ha especificado o se ha hecho de forma incorrecta y no coincide con la codificación física del fichero. Esto hace que sea más difícil el **procesamiento automático** de los mismos para poder encontrar errores o simplemente analizar la información, lo que a la larga redundaría nuevamente en una reducción de la calidad.



Características de calidad afectadas: Accesibilidad, Reutilización y Conformidad.



Recomendaciones: Cualquier informe o similar debería ir siempre acompañado de una serie de **ficheros reutilizables que faciliten el acceso a los datos en los que se basa** el informe o a los que se hace referencia desde el mismo.

Debemos además dar siempre mayor preferencia a aquellos formatos con mayor grado de **apertura y compatibilidad** (por ejemplo, csv frente a xls/xlsx), pero sin marginar tampoco a otros que sean muy populares entre los usuarios o que puedan suponer un *estándar de facto* en ciertos nichos de conocimiento específicos (por ejemplo, shp para datos espaciales).

Proporcionar además *interfaces* para la programación de aplicaciones (APIs) que cumplan con las buenas prácticas para la transferencia de datos a través de los protocolos de la web ([RESTful](#)) y usando formatos de intercambio a los que los desarrolladores estén acostumbrados (como XML o JSON). Con ello conseguiremos proporcionar una serie de funciones que podrán ser utilizadas por otras aplicaciones para facilitar la creación de nuevos servicios sobre los datos, como hacen la mayoría de aplicaciones más populares (como por ejemplo [Twitter](#) o [Google](#)).

Finalmente, adoptar el [estándar UTF](#) en la codificación de todos los archivos y documentos para garantizar la máxima compatibilidad, y asegurarse también de que las herramientas de gestión que se vayan a utilizar son compatibles con ese estándar.

1.2.7. Modelos freemium para el acceso a los datos

En este caso el problema consiste en que se ofrece una *“prueba gratuita”* de los datos facilitando el **acceso a una versión reducida o básica** de los mismos, pero luego se requiere realizar un pago adicional para poder acceder a la versión completa (cuando

no son directamente de pago por completo o requieren comprar un formato o soporte físico específico no disponible de otra manera).

Este problema es particularmente preocupante porque puede dar lugar al desarrollo de una nueva **“brecha digital de datos”** que fomente nuevas **desigualdades** en la sociedad dado el valor actual que tienen los datos en múltiples aspectos de nuestras vidas.



Características de calidad afectadas: Disponibilidad y Completitud.



Recomendaciones: El acceso online a **todos los datos** gubernamentales debe ser **gratuito para los usuarios sin excepciones ni filtros ni distintas categorías o niveles de acceso**, ya que su coste real ya está cubierto por la propia actividad de la administración que los recopila y gestiona como parte de su trabajo diario.

En aquellos casos excepcionales en los que se estuviese incurriendo en algún tipo de coste adicional por proporcionar los datos en un formato o soporte físico específico (por ejemplo, un *USB pen-drive*, un *DVD* o un *disco duro*), los cargos por dicho soporte deben estar claramente **justificados de forma transparente**. Además, esto nunca podrá reemplazar la obligación de **proporcionar los mismos datos de forma totalmente gratuita online**, sino que sería únicamente como servicio complementario y voluntario (por ejemplo, para aquellos usuarios que quieran acceder a grandes volúmenes de información pero cuenten únicamente con un ancho de banda limitado).

1.2.8. Falta de claridad con las licencias

Existe mucha información pública ya disponible online que simplemente no se puede reutilizar porque **las condiciones para su reutilización son totalmente desconocidas**, bien sea por la **ambigüedad** con la que se expresan o porque simplemente no se indican de ninguna manera.

Ese desconocimiento provoca una incertidumbre legal que a la larga da lugar a que se rehúya su uso para **evitar potenciales problemas legales**, sobre todo cuando se trata de un posible uso comercial.

Es frecuente también la falta de información sobre licencias o que ésta esté *oculta* en alguna otra parte del sitio u en otro sitio sin clara relación con los datos a los que hace referencia. En otros casos la licencia puede ser fácilmente localizable pero está **descrita de una forma tan vaga o críptica** que es casi imposible saber qué es lo que se puede hacer y qué no.

También puede pasar que los usuarios se encuentren con la contradicción de encontrar licencias abiertas y *copyrights* aplicados sobre las mismas páginas o secciones donde se encuentran los datos, no sabiendo cuál será aplicable al final.



Características de calidad afectadas: Apertura y Credibilidad.



Recomendaciones: Usar una **licencia-tipo fácilmente reconocible** (como por ejemplo [Creative Commons](#)) y claramente asociada a todos y cada uno de los conjuntos de datos a los que se aplica (a través de metadatos, notas a pie de página, etc.)

Indicar siempre y de forma clara y explícita la licencia aplicable. Incluso en los casos en los que pueda haber una legislación de mayor rango aplicable es necesario indicar la posibilidad de reutilización para mayor garantía de los usuarios. Acompañar las licencias con ejemplos claros de qué se puede o no se puede hacer con los datos para los casos de uso más comunes.

Es necesario además hacer una distinción más clara entre los términos aplicables al portal en sí (por ejemplo, el *copyright* del diseño del portal) y los aplicables a los conjuntos de datos publicados en él en particular.

2. CALIDAD DE LOS CONTENIDOS

Hasta el momento hemos hecho un completo repaso por los **fallos más comunes** que nos podemos encontrar respecto a la **calidad de los datos publicados por los gobiernos**, sin embargo, la calidad no solo se define a través de los atributos de los datos, sino que una parte muy importante también se define a través del propio **contenido** de los mismos.

En esta sección nos centraremos en esos contenidos haciendo hincapié en **cómo se deberían publicar algunos conjuntos de datos de referencia** según las recomendaciones elaboradas por la *Web Foundation* y *Open Knowledge Internacional* y siguiendo varios estándares de referencia para cada una de las materias que trataremos a continuación para tres áreas de conocimiento generales: rendición de cuentas, políticas sociales e innovación (ver gráfico inferior).



Gráfico 4: Áreas de conocimiento para publicar datasets

Además, para cada uno de los conjuntos de datos explorados haremos también un breve análisis sobre cuáles son los datos disponibles a nivel nacional en esa materia, así como sus principales fortalezas y debilidades, proporcionando también ejemplos de conjuntos de alta calidad publicados por otros países.

2.1. Datos por la rendición de cuentas

Es fundamental que los gobiernos ofrezcan información sobre el uso de los recursos públicos. Esto incluye datos sobre **en qué se invierte los impuestos**, cómo se conceden los **contratos del gobierno** o **cómo se canaliza el dinero en las campañas políticas**.

Para ello, será necesario publicar aquellos datos que son vitales para la transparencia, como los datos como los *presupuestos públicos, contratación y adjudicaciones, propiedad inmobiliaria y de la tierra, registros de empresas, legislación o resultados electorales.*



2.1.1. Presupuesto nacional

Presupuesto Nacional

Proyección del presupuesto nacional de alto nivel para el año entrante (no la ejecución presupuestaria del año en curso).

Los datos sobre el presupuesto deben incluir al menos los siguientes mínimos básicos:

- Actualización mínima anual.
- Presupuestos para cada departamento, ministerio o agencia gubernamentales incluyendo gastos e ingresos.
- Detalles de los presupuestos desglosados a nivel de sub-departamento, programa y tipo de gasto o ingreso.
- Descripciones de las distintas partidas presupuestarias.



Estándares de referencia: [Paquete de datos de la iniciativa de Open Spending](#), el [código de buenas prácticas en transparencia fiscal](#) del Fondo Monetario Internacional, las [buenas prácticas definidas por la International Budget Partnership](#) y el *framework* definido por PEFA para la [evaluación de las prácticas de gestión financiera pública](#).



Datos nacionales:

<http://www.sepg.pap.minhap.gob.es/Presup/PGE2016Ley/MaestroDocumentos/PGE-ROM/MnSerieRoja.htm>



Puntos fuertes: El nivel de detalle es adecuado, la frecuencia de actualización es correcta (y garantizada por ley) y existe también un amplio histórico disponible con los datos de ejercicios pasados.



Debilidades: Los datos están disponibles únicamente en un PDF y no se sigue ningún formato estándar de publicación, lo cual disminuye severamente sus posibilidades de reutilización. La estructura de los datos es bastante complicada y difícil de entender para aquellos no expertos en la materia. No cuenta con ningún tipo de metadatos.



Datos de alta calidad publicados por otros países:

Australia - <https://data.gov.au/dataset/budget-2015-16-tables-and-data>

Alemania - <https://www.bundeshaushalt-info.de/download.html>

México - http://www.transparenciapresupuestaria.gob.mx/es/PTP/Datos_Abiertos



Gasto público

2.1.2. Gasto público

Histórico de los datos de gasto real por cada transacción para al menos cualquier gasto mayor de 100.000€. Debe corresponderse con el gasto real a nivel transaccional, una base de datos de contratación pública o similar no será considerada suficiente.

Los datos sobre gasto público deben incluir al menos los siguientes mínimos básicos:

- Actualización mínima mensual.
- Organismo público que realiza la transacción.
- Fecha de la transacción.
- Identificación del proveedor.
- Cantidad nominal de la transacción.
- Registros individuales por cada transacción unitaria.



Estándares de referencia: [Paquete de datos de la iniciativa de Open Spending](#), el [código de buenas prácticas en transparencia fiscal](#) del Fondo Monetario Internacional y el *framework* definido por PEFA para la [evaluación de las prácticas de gestión financiera pública](#).



Datos nacionales:

<http://www.igae.pap.meh.es/sitios/igae/esES/EjecucionPresupuestaria/Paginas/EjecucionPresupuestaria.aspx>



Puntos fuertes: La publicación se hace con frecuencia mensual.



Debilidades: Los datos no están disponibles en ninguna fuente a nivel de transacción y lo único que se publica son los datos de ejecución presupuestaria que cuentan con un nivel de agregación mucho mayor.



Datos de alta calidad publicados por otros países:

Reino Unido - <https://www.gov.uk/government/publications/oscar-publishing-data-from-the-database-march-2016>

Uruguay

http://agev.opp.gub.uy/advni/documentos/uy_credito_presupuestal_detalle.zip

Grecia - <https://diavgeia.gov.gr/>



Propiedad inmobiliaria
y de las tierras

2.1.3. Propiedad inmobiliaria y de las tierras

Información sobre los propietarios de las tierras y los bienes inmuebles, así como la geo-localización de las parcelas e información sobre las transacciones y los precios de las mismas, según consten en la agencia de registro de la propiedad o el catastro nacional, donde generalmente se gestiona el registro nacional de propiedad de terrenos.

- El registro de propiedad debe incluir al menos los siguientes datos mínimos básicos:
- Localización de las parcelas
- Identificación de las parcelas
- Límites de las parcelas
- Valoración de la propiedad (precio tasado o de última transacción)
- Tipo de propiedad (pública, privada, comunitaria...)



Estándar de referencia: [Property Rights Data](#) de la Fundación Catasta por la información abierta sobre la propiedad y los derechos de la tierra.



Datos nacionales:

<https://www.sedecatastro.gob.es/OVCFrames.aspx?TIPO=TIT&a=masiv%BB>



Puntos fuertes: Existe información muy completa tanto en formato gráfico como textual, incluyendo detalles que no se encuentran muy frecuentemente a nivel global, como las divisiones de parcelas y valoraciones. Hay datos disponibles en formatos reutilizables. Las actualizaciones son bastante frecuentes.



Debilidades: Es necesaria certificación electrónica o registro previo para poder acceder a los datos. Algunos datos como los relativos a la propiedad requieren además un pago de tasas para poder acceder a ellos. Cuenta con su propia normativa para el uso de los datos que no es abierta y es además bastante compleja. Los datos históricos no están disponibles.



Datos de alta calidad publicados por otros países:

Canadá

<https://www.nrcan.gc.ca/earth-sciences/geomatics/canada-lands-surveys/11092#CLdata>

Nueva Zelanda

<https://data.linz.govt.nz/data/category/property-ownership-boundaries/>

Uruguay <https://catalogodatos.gub.uy/dataset/padrones-urbanos-y-rurales>



Resultados electorales

2.1.4. Resultados electorales

Resultados de las elecciones nacionales, incluyendo también información sobre el proceso (votos registrados, no válidos, en blanco...).

Los datos sobre resultados electorales deben incluir al menos los siguientes mínimos básicos:

- Actualización mínima con cada nueva elección.
- Información sobre los candidatos.
- Número de votos registrados.
- Número de votos no válidos.
- Número de votos en blanco.
- Datos a nivel de mesa electoral.



Estándares de referencia: [Election Markup Language](#) definido por OASIS y la [guía de datos electorales de la Open Election Data Initiative](#).



Datos nacionales: <http://www.infoelectoral.mir.es/min/>



Puntos fuertes: Buen histórico de datos y algunas visualizaciones interactivas disponibles. Datos reutilizables y posibilidad de descargas en bruto.



Debilidades: El nivel de desagregación no es todo lo detallado que sería recomendable. Es necesario contar con algunos *plugins* adicionales para poder beneficiarse de las visualizaciones. No se utiliza el estándar de referencia para el modelo de datos. La información sobre candidaturas no está disponible en un formato reutilizable.



Datos de alta calidad publicados por otros países:

Austria

<https://www.data.gv.at/katalog/dataset/09716341-2bea-4298-9525-e936d8247d19>

Eslovaquia <http://volby.statistics.sk/nrsr/nrsr2016/sk/download.html>

Suecia <http://www.val.se/val/val2014/statistik/index.html>



Legislación

2.1.5 Legislación

Referencia legislativa del país, incluyendo al menos la Constitución y el conjunto de todas las leyes y estatutos en vigor, así como aquellos que se encuentren en trámite en el parlamento.

Los datos legislativos deben incluir al menos los siguientes mínimos básicos:

- Actualización mínima trimestral.
- Contenido de la ley o estatuto correspondiente.
- Histórico de todas las enmiendas aplicadas.
- Fecha de la publicación original y de la última actualización o enmienda.
- Transcripciones de los debates parlamentarios sobre las propuestas y proyectos de ley.
- Detalle de las votaciones sobre los proyectos de ley por miembro del parlamento.



Estándar de referencia: [Declaración sobre la transparencia parlamentaria](#) de OpeningParliament.org



Datos nacionales: <https://www.boe.es/legislacion/legislacion.php>



Puntos fuertes: Actualización diaria, formatos reutilizables, sistema de publicación muy trabajado y datos completamente gratuitos. Gran histórico de datos disponible.



Debilidades: Dificultad para acceder a grandes cantidades de datos en bruto.



Datos de alta calidad publicados por otros países:

En este caso cabe destacar que España puede ser considerado referente de buenas prácticas para este sector. Otros países con buenas prácticas serían también:

Reino Unido – <http://www.legislation.gov.uk/developer>

Nueva Zelanda - <http://legislation.govt.nz/subscribe/>

2.2. Datos para las políticas sociales

Los datos abiertos tienen el potencial de hacer que los servicios públicos esenciales — como la sanidad, la educación y la gestión medioambiental — sean más efectivos e inclusivos, lo que a su vez contribuye a fortalecer a todos los ciudadanos. Los datos abiertos pueden mejorar directamente la distribución de servicios, dado que ofrecen a los ciudadanos más herramientas para la elección — o indirectamente, ayudando a quienes establecen las políticas públicas a identificar y abordar problemas sociales de todo el sistema.



Censo y estadísticas nacionales

1.2.1. Censo y estadísticas nacionales

Incluyendo indicadores clave en el ámbito social (ej. desempleo), demográfico (ej. población) y económico (ej. producto interior bruto) generalmente ofrecidos por la Agencia Nacional de Estadística. Los datos sobre censo y estadísticas deben incluir al

menos los siguientes mínimos básicos:

- Actualización mínima trimestral (excepto para el censo)
- Desagregado por sexos.
- Censo nacional de población.
- Estadísticas vitales (nacimientos y defunciones)
- Estadísticas de empleo y actividades (incluyendo desempleo)
- Producto interior bruto y renta per cápita.



Estándares de referencia: vocabulario del [Statistical Data and Metadata eXchange](#) y las [categorías de datos estadísticos analizadas por el Open Data Inventory](#).



Datos nacionales: http://www.ine.es/inebaseDYN/cp30321/cp_inicio.htm



Puntos fuertes: Existe una gran variedad de datos disponibles y generalmente en múltiples formatos reutilizables. Los datos están actualizados y con un amplio histórico disponible. En ocasiones se ofrece incluso documentación adicional.



Datos de alta calidad publicados por otros países:

Italia

<http://dati.istat.it/>

Japón

http://www.estat.go.jp/SG1/estat/GL38020103.do?_toGL38020103_&tclassID=000001077438&cycleCode=0

Bélgica

<http://statbel.fgov.be/fr/statistiques/opendata/>



Contaminación
medioambiental

1.2.2. Contaminación medioambiental

Datos sobre la concentración de elementos contaminantes en aire y agua potable, particularmente aquellos que son considerados nocivos para el ser humano.

Los datos sobre contaminación ambiental deben incluir al menos los siguientes mínimos básicos:

- Actualización mínima semanal.
- Disponible por cada estación de medición.
- Partículas en suspensión (PM 10 y PM2.5).
- Gases contaminantes (SOx, NOx y CO)
- Bacterias coliformes fecales.



Estándares de referencia: Pautas para la [calidad del aire](#) y para la [calidad del agua potable](#) de la *World Health Organisation*.



Datos nacionales:

<http://www.mapama.gob.es/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei/>



Puntos fuertes: Existen datos disponibles sobre emisiones nocivas y

contaminantes, calidad del agua y deforestación. Los datos se actualizan regularmente y están acompañados información cartográfica. Se autoriza la reutilización. Se proporciona también alguna documentación adicional.



Debilidades: La información disponible no es completa y los datos no están totalmente actualizados, y en el caso de la calidad del agua incluso se podría afirmar que están claramente desactualizados. No hay ningún tipo de metadatos acompañando a los datos.



Datos de alta calidad publicados por otros países:

Dinamarca

<http://www.statbank.dk/statbank5a/SelectVarVal/Define.asp?MainTable=MRU1>

Finlandia

[http://www.syke.fi/fiFI/Avoim_tieto/Avoimet_rajapinnat/Avoimen_ymparistotiedon_rajapinnat\(37459\)](http://www.syke.fi/fiFI/Avoim_tieto/Avoimet_rajapinnat/Avoimen_ymparistotiedon_rajapinnat(37459))

Suecia

<http://utslappisiffror.naturvardsverket.se/>

Rusia

<http://data.gov.ru/opendata/7704221753-monthlyairpollutiondata>



Funcionamiento de los servicios públicos básicos

2.2.3. Funcionamiento de los servicios públicos básicos Indicadores sobre el rendimiento de los servicios públicos de los sistemas sanitario y educativo.

Los datos sobre servicios públicos deben incluir al menos los siguientes mínimos básicos:

- Actualización mínima trimestral (excepto para las calificaciones)
- Desagregado por sexos
- Nivel de acceso a la sanidad pública
- Niveles de vacunación en la población
- Listas de espera para ser atendidos por los servicios sanitarios

- Gasto medio por cada uno de los servicios sanitarios
- Niveles de mortalidad y supervivencia por tipo de dolencia
- Tasas de escolarización en todos los niveles de educación
- Ratios de absentismo escolar
- Calificaciones por nivel y exámenes oficiales
- Número de graduados por titulaciones



Estándares de referencia: Indicadores de [calidad del sistema sanitario](#) y los indicadores de educación recogidos en el informe [Education at a Glance](#) de la OECD.



Datos nacionales:

http://www.ine.es/ss/Satellite?L=es_ES&c=INEPublicacion_P&cid=1259924822559&idp=1259924822559&p=1254735110606&pagename=ProductosYServicios%2FPYSLayout&titema=Salud

<https://www.mecd.gob.es/servicios-al-ciudadano-mecd/estadisticas/portada.html;jsessionid=AF671774661C77784E1E4C611310FFF2>



Puntos fuertes: Una buena parte de los datos que están disponibles lo están en formatos reutilizables y están desagregados por sexos. Además, en general se han publicado con licencias compatibles con el uso por parte de terceros.



Debilidades: La información en esta área es generalmente bastante limitada y no está completamente actualizada. Los datos disponibles para esta materia se encuentran dispersos entre múltiples agencias y en diferentes tipos de bases de datos y formatos, haciendo que sea bastante encontrarla y reutilizarla. No existe una fórmula consistente para la gestión de estos datos. Algunos datos están restringidos para uso exclusivamente personal.



Datos de alta calidad publicados por otros países:

Suiza <https://opendata.swiss/en/dataset/quality-indicators-of-the-swiss-acute-care-hospitals-2014>

EE UU <https://www.healthdata.gov/search/type/dataset>

Brasil <http://portal.inep.gov.br/indicadores-educacionais>

Francia <http://www.data.gouv.fr/fr/datasets/indicateurs-de-resultat-des-lycees-denseignement-general-et-technologique/>

2.3. Datos para la innovación

Los datos abiertos cuentan con un gran potencial a la hora de fomentar la innovación, tanto pública como privada. Los datos son usados por los emprendedores en sus aplicaciones y pueden también desbloquear un gran valor para las empresas desarrollando nuevos servicios. Además, también pueden contribuir a aumentar la eficiencia y productividad de los servicios públicos actuales. Por ejemplo, se puede crear innovación y valor económico usando conjuntos de datos como los datos cartográficos, horarios de transporte público y datos sobre comercio internacional o criminalidad.



Mapa Nacional

2.3.1 Mapa Nacional

Mapa(s) digital(es) del país publicado por alguna agencia nacional.

El mapa debe incluir al menos los siguientes datos mínimos básicos:

- Actualización mínima anual.
- Resolución mínima de 1:250.000 (1cm = 2.5 Km.)
- Fronteras administrativas.
- Topografía y figuras de relieve.
- Cuencas hidrográficas.
- Localización de las principales infraestructuras públicas de comunicación (carreteras y ferrocarril).
- Coordenadas.
- Proyecciones geográficas para habilitar la interpretación de las coordenadas.



Estándar de referencia: [Comité de expertos de Naciones Unidas para la Gestión de Información Geoespacial Global](#).



Datos nacionales:

<http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do#selectedSerie>



Puntos fuertes: Existe una amplia cantidad y variedad de datos cartográficos y topográficos disponibles en alta resolución. Se utilizan licencias compatibles con la apertura y libre reutilización de la información. Se proporcionan también metadatos básicos.



Debilidades: Los formatos utilizados son principalmente gráficos y no facilitan la reutilización. No es posible hacer descargas masivas de datos. No hay datos históricos disponibles.



Datos de alta calidad publicados por otros países:

Holanda <https://www.pdok.nl/nl/ahn3-downloads>

Reino Unido <https://www.ordnancesurvey.co.uk/opendatadownload/userDetails.html>

Dinamarca - <https://download.kortforsyningen.dk/>

2.3.2. Límites administrativos y localizaciones

Unidades o áreas definidas para el país en los distintos niveles administrativos, incluyendo las divisiones de códigos postales. Los datos sobre límites administrativos deben incluir al menos los siguientes mínimos básicos:

- Límites administrativos para los distintos niveles (nacional, regional, local)
- Direcciones de códigos postales.
- Coordenadas de los límites administrativos y los códigos postales.
- Polígonos con las proyecciones geográficas de los límites administrativos y los códigos postales.
- Etiquetas o nombres aplicables a los distintos polígonos (provincia, municipio, barrio...).



Estándares de referencia: Las [Global Administrative Unit Layers](#) de la UN-FAO, la [base de datos geográfica](#) y de [códigos postales](#) mantenida por *GeoNames* y el [Manual de direcciones y códigos postales](#) de la *Universal Postal Union*.



Datos nacionales:

<http://www.correos.es/ss/Satellite/site/pagina-1349169615030/info>



Puntos fuertes: Existen datos que además están enriquecidos con otros indicadores socioeconómicos y territoriales. Datos bien estructurados. Hay documentación básica disponible. Existe la posibilidad de visualizar los datos sobre mapas digitales y acceder a los datos vectoriales. Actualizaciones trimestrales.



Debilidades: Sólo es posible acceder a los datos mediante pago previo. No hay licencia abierta y se otorga acceso bajo contrato específico con licencia de uso anual renovable. Los datos no tienen geoinformación asociada.



Datos de alta calidad publicados por otros países:

Singapur <https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-no-sea>

Paraguay <http://geo.stp.gov.py/user/dgeec/datasets>

Eslovaquia <https://data.gov.sk/dataset?tags=register+adries>



2.3.3. Contratación pública

Todos los concursos y adjudicaciones realizados por cada oficina del gobierno nacional. Los datos sobre contratación pública deben

Contratación pública

incluir al menos los siguientes mínimos básicos:

- Actualización mínima semanal.
- Concursos y adjudicaciones para cada oficina gubernamental.
- Descripciones de los concursos y de las adjudicaciones.
- Estado actualizado del concurso.
- Valor final adjudicado.

- Identificación sobre el adjudicatario.



Estándares de referencia: Los [principios globales de contratación abierta](#) y el [esquema de datos del Estándar de datos de Contratación](#) del *Open Contracting Partnership*.



Datos nacionales: <https://contrataciondelestado.es/wps/portal/licRecientes>



Puntos fuertes: Existe información completa sobre concursos, ganadores, cantidades, fechas, organizaciones, ofertas, etc. La información disponible se actualiza a diario y la actualización está automatizada. Se permite la reutilización.



Debilidades: Es necesario registrarse para acceder a toda la información y en formato reutilizable, aunque ni siquiera así es posible acceder a la información en bruto. El sistema de búsqueda está algo restringido. Los formatos no son reutilizables. No se ha adoptado el estándar internacional para los datos de contratación. La licencia es propia en lugar de usar una licencia-tipo.



Datos de alta calidad publicados por otros países:

Filipinas <http://www.ps-philgeps.gov.ph/egp/data.html>

Canadá <https://buyandsell.gc.ca/>



Registro de empresas

2.3.4. Registro de empresas

Información básica sobre las empresas constituidas en el país sin necesidad de entrar en detalles financieros como los balances.

Los datos sobre empresas deben incluir al menos los siguientes mínimos

básicos:

- Actualización mínima semanal.
- Nombre de la compañía.
- Identificador de la compañía.
- Dirección fiscal.

- Responsables y beneficiarios.
- Actividades registradas.



Estándar de referencia: [Modelo de datos de Open Corporates](#).



Datos nacionales: <http://www.rmc.es>



Puntos fuertes: Datos completos y actualización diaria.



Debilidades: Los datos sólo están disponibles bajo pago y con registro obligatorio. No queda claro si los datos están en formatos reutilizables o si es posible descargarlos en bruto.



Datos de alta calidad publicados por otros países:

Bulgaria

<https://opendata.government.bg/dataset/tbprobckn-pernctbp>

Canadá

<http://open.canada.ca/data/en/dataset/0032ce54-c5dd-4b66-99a0-320a7b5e99f2>

3. REFERENCIAS

Web Foundation, Open Data Barometer 4ª Edición – Informe Global

<http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport-ES.pdf>

Open Knowledge, The State of Open Government Data in 2017.

<https://blog.okfn.org/files/2017/06/FinalreportTheStateofOpenGovernmentDatain2017.pdf>

Open Knowledge, Open Definition 2.1.

<http://opendefinition.org/od/2.1/en/>

Joshua Tauberer, The Annotated 8 Principles of Open Government Data.

<https://opengovdata.org/>

Carta Internacional de los Datos Abiertos, Principios.

<https://opendatacharter.net/principles-es/>

European Commission, Joinup Initiative – Introduction to Open Data Quality.

<https://joinup.ec.europa.eu/document/tm22-open-data-metadata-quality-en>

ISO/IEC 25012, Modelo de Calidad del Producto de Datos

<http://iso25000.com/index.php/normas-iso-25000/iso-25012>

International Monetary Fund, The Fiscal Transparency Code.

<http://blog-pfm.imf.org/files/ft-code.pdf>

International Budget Partnership, Open Budget Survey Guide and Questionnaire.

<https://www.internationalbudget.org/publications/open-budget-survey-2017-guide-questionnaire-english/>

PEFA, Framework for assessing public finance management.

<https://pefa.org/user-guidance>

CADASTA, An overview of property rights data.

<http://cadasta.org/open-data/overview-of-property-rights-data/>

OASIS, Election Markup Language (EML) Specification Version 7.0.

<http://docs.oasis-open.org/election/eml/v7.0/eml-v7.0.html>

Open Election Initiative, Election Data Guide.

<https://openelectiondata.net/en/guide/>

OpeningParliament.org, Declaración sobre la transparencia parlamentaria.

<https://www.openingparliament.org/static/pdfs/spanish.pdf>

Open Data Watch, The Open Data Inventory 2016 methods report.

<http://odin.opendatawatch.com/Downloads/otherFiles/ODIN-2016-Methodology.pdf>

Statistical Data and Metadata eXchange, SDMX Technical Specifications.

https://sdmx.org/?page_id=5008

Organización Mundial de la Salud, Guías de calidad del aire de la OMS.

http://apps.who.int/iris/bitstream/10665/69478/1/WHO_SDE_PHE_OEH_06.02_spa.pdf

World Health Organization, Guidelines for drinking water quality.

http://www.who.int/water_sanitation_health/publications/drinking-water-quality-guidelines-4-including-1st-addendum/en/

OECD, Data for Measuring Health Care Quality and Outcomes.

<http://www.oecd.org/health/health-systems/health-care-quality-indicators.htm>

OECD, Education at a Glance 2017.

<http://www.oecd.org/education/education-at-a-glance-19991487.htm>

UNGGIM – ISPRS, The status of topographic mapping in the world.

http://www.isprs.org/documents/reports/The_Status_of_Topographic_Mapping_in_the_World.pdf

FAO, Global administrative unit layers (GAUL).

<http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691&currTab=simple>

GeoNames, Geographical Database.

<http://www.geonames.org/countries/>

Universal Postal Union, Addressing and Postcode Manual.

<http://www.upu.int/fileadmin/documentsFiles/activities/addressingAssistance/manualAddressingAddressingAndPostcodeManualEn.pdf>

Open Contracting Partnership, Principios globales.

<https://www.open-contracting.org/comenzar/principios-globales/?lang=es>

Open Contracting Data Standard, Schema Reference.

<http://standard.open-contracting.org/latest/en/schema/>

Open Corporates, API Reference.

<http://api.opencorporates.com/documentation/API-Reference>

4.GRÁFICOS

Gráfico 1. Los tres requisitos básicos de los datos abiertos.

Gráfico 2. Indicadores de calidad de los datos.

Gráfico 3. Fallos frecuentes en la calidad de los datos abiertos.

Gráfico 4. Áreas de conocimiento de la calidad de los datasets.