



www.data.europa.eu

Metadata Quality Assurance

04.04.2026

How data.europa.eu measures the quality of harvested metadata

The Metadata Quality Assessment (MQA) is a tool developed by the consortium of data.europa.eu to study the quality of metadata harvested by data.europa.eu. It is intended to help data providers and data portals to check their metadata quality and to receive suggestions for improvements. The results are presented via the MQA and are also available as download. In the following we describe the functionality of the MQA and the methodology it uses.

If this page still does not answer all your questions, please feel free to contact us via our feedback form at the end of the page.

[Metodología de evaluación de la calidad de los metadatos](#)

Ámbito de la investigación

Con la herramienta MQA pretendemos dar respuesta a la siguiente pregunta:

¿Cuál es la calidad de los metadatos del sector público en la región paneuropea y dónde se hallan los principales escollos para mejorarla?

Partiendo de esta premisa, la herramienta MQA indaga actualmente en las siguientes cuestiones concretas:

- Conformidad con DCAT-AP y los derivados de DCAT-AP
- Divulgación de información no obligatoria para DCAT-AP
- Accesibilidad de los datos referenciados en los metadatos a través de la URL de acceso y de descarga
- Legibilidad por máquina de los datos referenciados
- Uso de licencias

Cada cuestión da lugar a su vez a exámenes individuales que se describen en detalle a continuación.

Lo que no contemplamos

The MQA is limited by the metadata it can examine. The investigation is limited exclusively to the metadata that data.europa.eu collects during the harvesting process. If there are errors in the source metadata, these can falsify the overall result. To limit this error potential, the MQA provides a validation service that can be used by data providers to validate their metadata for

valid formats and compliant DCAT-AP before integrating it into the harvesting process.

[DCAT-AP SHACL validation service web page](#)

[DCAT-AP SHACL validation service API](#)

El proceso MQA

En cada proceso de recolección la herramienta MQA también comprueba los metadatos. MQA cuantifica la calidad con arreglo a varios indicadores, los cuales se explican en las siguientes tablas. Los resultados de estas comprobaciones se almacenan como Vocabulario de calidad de datos ([DQV, Data Quality Vocabulary](#)). DQV es una especificación de W3C que sirve para describir la calidad de un conjunto de datos.

Dado que la accesibilidad puede ser volátil, es preciso llevar a cabo sucesivas comprobaciones de la URL de acceso y la URL de descarga. Por este motivo, MQA realiza comprobaciones periódicas de la accesibilidad de todas las distribuciones. A diferencia del resto de indicadores, esta operación tarda más en ejecutarse, dado que las distribuciones se comprueban a través de HTTP y el tiempo de respuesta de cada una de las URL solicitadas puede ser mayor. La herramienta MQA emplea un mecanismo que tiene en cuenta que la accesibilidad de cada URL se reevalúa pocas semanas después de la última comprobación.

Asunciones

La herramienta MQA se basa en las siguientes asunciones.

Uso de campos no obligatorios

Consideramos que el hecho de rellenar todos los campos obligatorios de DCAT-AP no es suficiente para que los metadatos sean de alta calidad. Por consiguiente, en la evaluación también se comprueban los campos no obligatorios según DCAT-AP. A continuación se recogen los campos exactos que se comprueban.

Contenido idéntico para múltiples distribuciones

Si un conjunto de datos contiene más de una distribución, todas son idénticas en contenido, solo se diferencian en la representación de los datos. Un conjunto de datos puede contar con dos distribuciones, una que presente los datos como PDF y otra en la que idénticos datos se muestran en formato RDF/XML, legible por máquina.

Dimensiones

Esta sección describe todas las dimensiones que contempla la herramienta MQA para determinar la calidad. Las dimensiones se derivan de los [principios FAIR](#).

Facilidad de localización

La siguiente tabla recoge las métricas que ayudan a personas y máquinas a encontrar conjuntos de datos. La puntuación máxima en este área es de 100 puntos.

indicador	Descripción	Métrica	Peso
Uso de palabras clave	Las palabras clave ayudan de forma directa a la búsqueda y, por tanto, aumentan la facilidad de localización del conjunto de datos.	El sistema comprueba si se han definido palabras clave. El número de palabras clave no afecta a la puntuación. Dataset dcat:keyword	30
Categorías	Las categorías ayudan a los usuarios a explorar los conjuntos de datos atendiendo a su temática.	Se comprueba si se asignan una o más categorías al conjunto de datos. El número de categorías asignadas no afecta a la puntuación. Dataset dcat:theme	30
Búsqueda geográfica	El uso de información espacial permite a los usuarios hallar conjuntos de datos con una búsqueda acotada geográficamente.	Se comprueba si se establece o no la propiedad. Dataset dcat:spatial	20
Búsqueda acotada temporalmente	El uso de información temporal permite a los usuarios realizar una búsqueda acotada temporalmente.	Se comprueba si se establece o no la propiedad.	20

Accesibilidad

La siguiente tabla describe las métricas empleadas para determinar si se garantiza el acceso a los datos a los que hacen referencia las distribuciones. La puntuación máxima en este área es de 100 puntos.

indicador	Descripción	Métrica	Peso
Accesibilidad de la URL de acceso	La URL de acceso no tiene por qué ser necesariamente un enlace directo a los datos; también puede tratarse de una URL que da acceso al conjunto de datos o que recoge información adicional sobre el conjunto de datos.	Se comprueba la accesibilidad de la URL específica a través de una solicitud HTTP HEAD Si el código de estado de la respuesta se encuentra comprendido entre 200 y 300, el recurso supera la evaluación de accesibilidad. Distribution dcat:accessURL	50
URL de descarga	La URL de descarga es un enlace directo a los datos referenciados.	Se comprueba si se establece o no la propiedad. Distribution dcat:downloadURL	20
Accesibilidad de la URL de descarga	Si existe una URL de descarga, se comprueba la accesibilidad.	Se comprueba la accesibilidad de la URL específica a través de una solicitud HTTP HEAD Si el código de estado de la respuesta se encuentra comprendido entre 200 y 300, el recurso supera la evaluación de accesibilidad. Distribution dcat:downloadURL	30

Interoperabilidad

La siguiente tabla describe las métricas que se emplean para determinar si una distribución se considera interoperable. Partiendo de la asunción «contenido idéntico con varias distribuciones», solo la distribución con más puntos se empleará para calcular la puntuación. La puntuación máxima en este área es de 110 puntos.

Indicador	Descripción	Métrica	Peso
Formato	Este campo indica el formato de archivo de la distribución.	Se comprueba si se establece o no la propiedad. Distribution dct:format	50
Tipo de medio	Este campo indica el tipo de medio de la distribución.	Se comprueba si se establece o no la propiedad. Distribution dcat:mediaType	10
Formato/tipo de medio del vocabulario	Comprueba si el formato y el tipo de medio pertenecen a un vocabulario controlado.	The format vocabulary can be found in the data.europa.eu GitLab repository . The media type is check against the IANA list Distribution dct:format dcat:mediaType	10
Sin derechos de autor	Comprueba si el formato de la distribución no está sujeto a derechos de autor.	The distribution is considered as non-proprietary if the specified format and media type is contained in the corresponding data.europa.eu GitLab repository vocabulary. Distribution dct:format	20
Legible por máquina	Comprueba si el formato de la distribución es legible por máquina.	The distribution is considered as machine-readable if the specified format and media type is contained in the corresponding data.europa.eu GitLab repository vocabulary. Distribution dct:format	20

indicador	Descripción	Métrica	Peso
Cumplimento de DCAT-AP	<p>DCAT-AP compliance is calculated across all sources and datasets available on a catalogue. This check is only performed if the metadata is originally harvested as DCAT-AP or as a valid derivate. DCAT-AP is a specification for describing linked public data in Europe. The data.europa.eu portal may also harvest metadata which does not fully comply to DCAT-AP. In order to increase conformity to DCAT-AP, the MQA checks each metadata for its DCAT-AP compliance.</p>	<p>The metadata is validated against a set of <u>SHACL shapes</u>. The metadata is not compliant, if the <u>SHACL</u> validation reports at least one issue.</p> <p>The MQA uses data.europa.eu's <u>DCAT-AP SHACL</u> validation service.</p> <p><u>SHACL</u> is a recommendation from the W3C and is used for validating RDF graphs against a set of shapes.</p>	30

Reusabilidad

La siguiente tabla recoge las métricas empleadas para comprobar la reusabilidad de los datos. La puntuación máxima en este área es de 75 puntos.

indicador	Descripción	Métrica	Peso
Información de la licencia	Una licencia constituye información valiosa para la reutilización de los datos.	Se comprueba si se establece o no la propiedad. Distribution dct:license	20
Vocabulario de la licencia	Nuestra pretensión es limitar la provisión de información incorrecta sobre las licencias. Nos encontramos, por ejemplo, con muchos casos de licencias CC en las que no se indica la versión.	This section describes all dimensions that the MQA examines in order to determine the quality. The dimensions are derived from the FAIR principles . The MQA recommends and credits the usage of controlled vocabularies. The data.europa.eu portal publishes its controlled vocabularies in GitLab . The vocabularies are derived from the EU Vocabularies . Distribution dct:license	10
Restricciones de acceso	Este campo indica si el acceso a los datos es público o, por el contrario, está restringido.	Se comprueba si se establece o no la propiedad. Dataset dct:accessRights dcat:mediaType	10
Vocabulario de las restricciones de acceso	El uso de un vocabulario controlado aumenta la usabilidad.	Se comprueba si se emplea el vocabulario controlado para los derechos de acceso . Dataset dct:accessRights	5
Punto de contacto	El punto de contacto contiene información de la persona a la que hay que dirigir cualquier pregunta relativa a los datos.	Se comprueba si se establece o no la propiedad. Dataset dct:contactPoint	20
Editor	El editor es la persona u organización que ha publicado los datos.	Se comprueba si se establece o no la propiedad. Dataset dct:publisher	10

Contextualidad

La siguiente tabla recoge algunas propiedades con poco peso que ofrecen más contexto al usuario. La puntuación máxima en esta área es de 20 puntos

indicador	Descripción	Métrica	Peso
Derechos	En algunos casos, no es posible aplicar una licencia específica a un conjunto de datos. El campo «Derechos» puede utilizarse para indicar una referencia a un recurso que informará al usuario sobre los derechos de los que dispone cuando utilice el conjunto de datos.	Se comprueba si se establece o no la propiedad. Distribution dct:rights	5
Tamaño de archivo	Especifica el tamaño del archivo en bytes.	Se comprueba si se establece o no la propiedad. Distribution dct:byteSize	5
Fecha de publicación	La fecha en la que se publicó el conjunto de datos o la distribución.	Se comprueba si se establece o no la propiedad. Dataset and Distribution dct:issued	5
Fecha de modificación	La fecha de la última modificación realizada en el conjunto de datos o la distribución	Se comprueba si se establece o no la propiedad. Dataset and Distribution dct:modified	5

Calificación

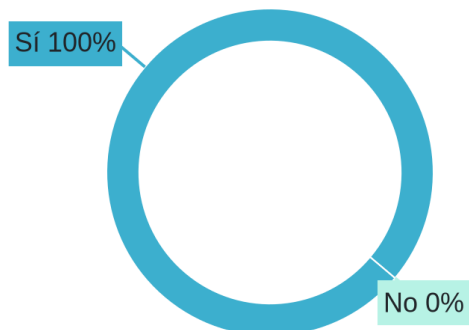
La puntuación final se calcula a partir de cuatro grupos de puntuación. En la siguiente tabla se muestra la asignación de los puntos a la categoría de puntuación. La representación de la puntuación en MQA se expresa exclusivamente a través de categorías de puntuación. Ello permite a los proveedores alcanzar la puntuación más elevada aunque se dé una ligera reducción de puntos.

Dimensión	Puntuación máxima
Facilidad de localización	100
Accesibilidad	100
Interoperabilidad	110
Reusabilidad	75
Contextualidad	20
Total	405

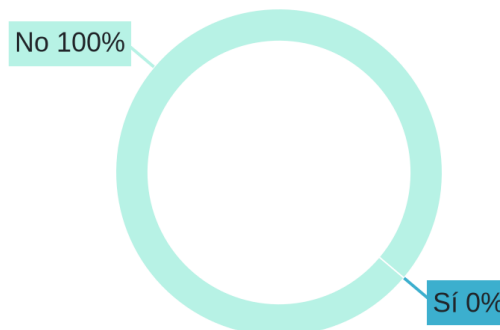
Calificación	Rango de puntos
Excelente	351 - 405
Buena	221 - 350
Suficiente	121 - 220
Mala	0 - 120

Contextualidad

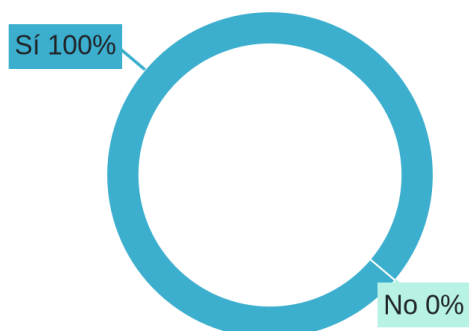
Derechos



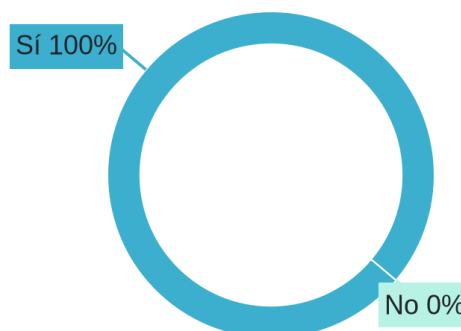
Tamaño de archivo



Fecha de publicación

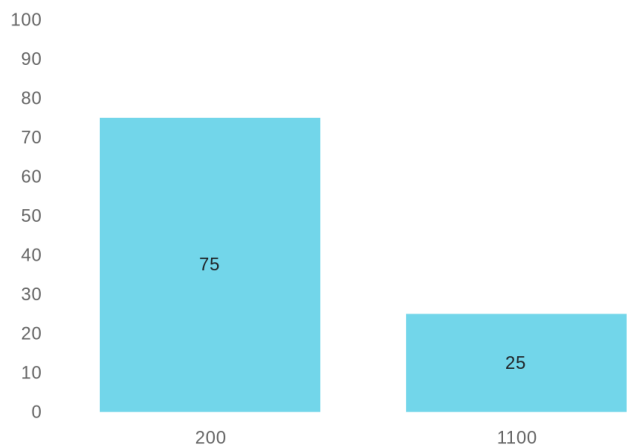


Fecha de modificación

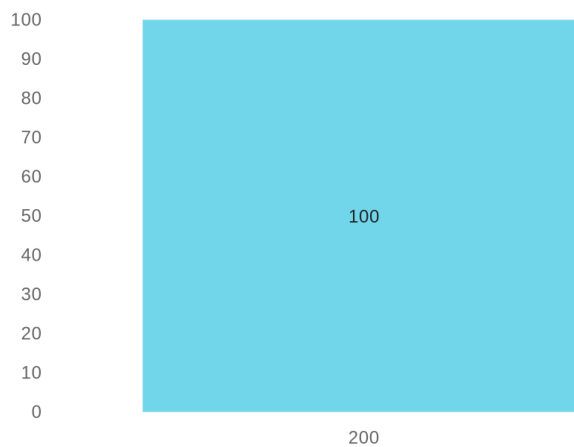


Accesibilidad

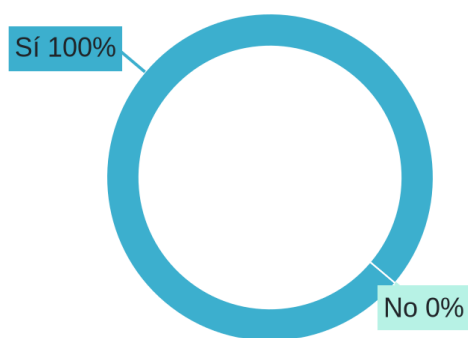
Códigos de estado de URL de acceso más frecuentes



Códigos de estado de URL de descarga más frecuentes

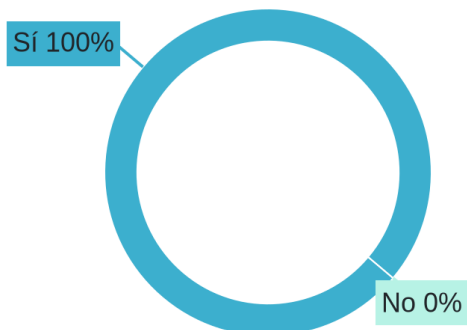


URL de descarga

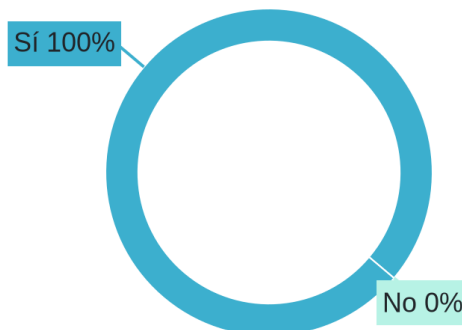


Reusabilidad

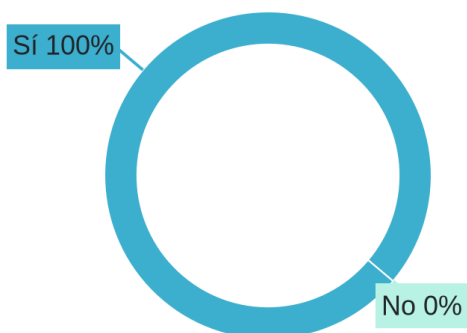
Información de la licencia



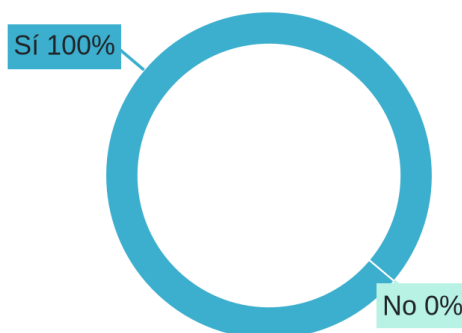
Vocabulario de la licencia



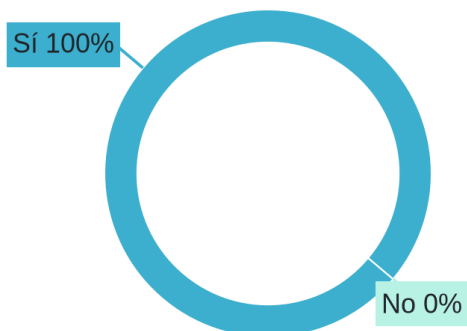
Restricciones de acceso



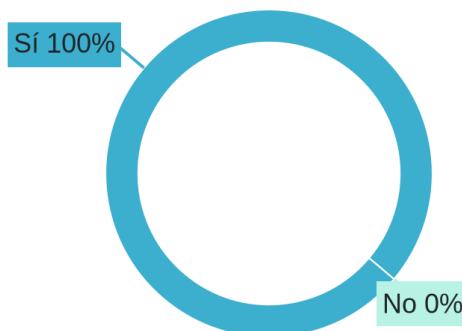
Vocabulario de las restricciones de acceso



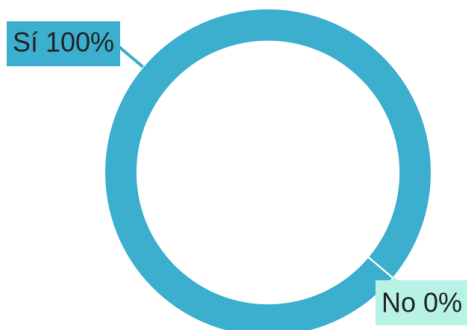
Punto de contacto



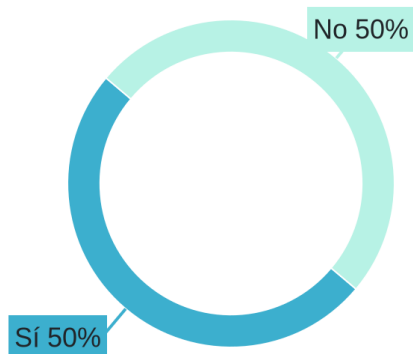
Editor



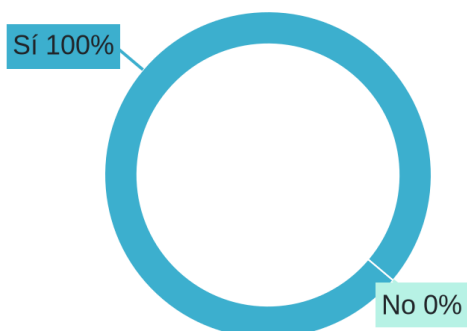
Formato



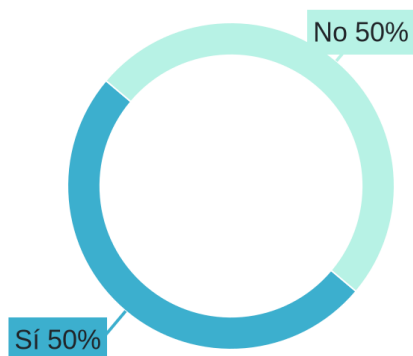
Tipo de medio



Formato/tipo de medio del vocabulario



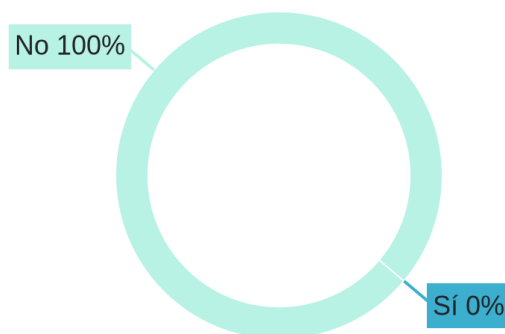
Sin derechos de autor



Legible por máquina



Cumplimento de DCAT-AP

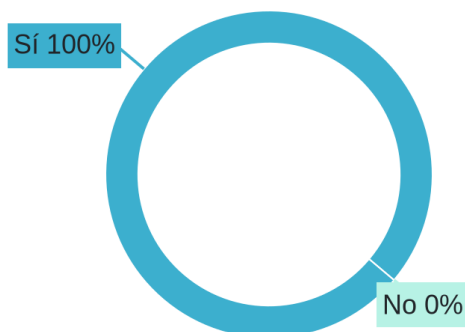


Facilidad de localización

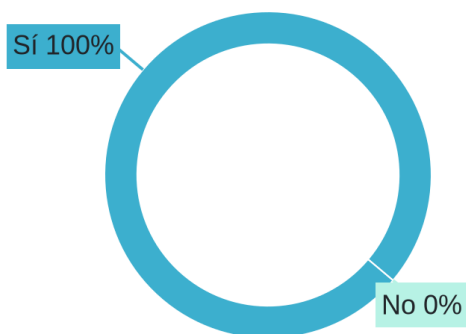
Uso de palabras clave



Categorías



Búsqueda geográfica



Búsqueda acotada temporalmente

