



www.data.europa.eu

Metadata Quality Assurance

23.10.2026

Contents

Methodology - 3

Official INSPIRE Data and Services Catalogue (codsi) - 18

Methodology

How data.europa.eu measures the quality of all stored metadata

The Metadata Quality Assessment (MQA) is a tool developed by the consortium of the European Data Portal to assess the quality of metadata harvested by data.europa.eu. It helps data providers and data portals to identify metadata quality issues and improvement areas. The results are presented in the MQA and are also available for download. In the following we describe the functionality of the MQA and the methodology it uses.

The rationale behind the current MQA is compliance based on DCAT-AP properties putting more emphasis on the mandatory properties of the model and less on the recommended and optional ones. The methodology evaluates DCAT-AP 3 metadata properties based on concrete metrics. Such metrics are given weights based on their status as mandatory, optional or recommended. Each entity gets a score based on the weighted results. The scores are then aggregated based on the DCAT-AP data structure.

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.

This page describes the current MQA methodology for the latest version of DCAT-AP metadata.

Scope of assessment

With the MQA, we want to answer the following question: What is the metadata quality for public sector data in the pan-European region and where are the biggest hurdles to achieving better quality? With this approach we intend to provide actionable insights to data providers and data portals to increase the quality of their public data sets.

Based on this, the MQA is currently investigating the following concrete questions:

The MQA assesses the quality of metadata harvested by the European Data Portal. The current methodology evaluates metadata compliance with the latest version of DCAT-AP (currently 3.0) and its derivatives, focusing on three DCAT-AP resource classes:

- Datasets
- Distributions
- Data services

The methodology checks whether relevant metadata properties are present, valid, or use controlled vocabularies where applicable. It also checks selected URLs for technical accessibility.

The methodology is based on four FAIR-related quality dimensions:

- Findability
- Accessibility
- Interoperability
- Reusability

The scoring methodology is based on the properties listed in the tables below. Each property is mapped to a FAIR principle.

The MQA processing pipeline

With each harvesting process, metadata is evaluated by the MQA. The MQA produces metric results for datasets, distributions, and data services. These results are stored as quality metadata and used to calculate scores. The evaluation can take some time depending on the number of datasets in each catalogue.

Some checks are simple presence checks. Other checks verify URL syntax, controlled vocabulary usage, format characteristics, or HTTP accessibility. More information is provided in the tables below.

Scoring model

The current methodology uses a layered and weighted model on key DCAT-AP properties. Each evaluated metric belongs to one of three importance layers:

Importance	Layer	Weight
Mandatory	L1	1.0
Recommended	L2	0.5
Optional	L3	0.25

Each resource class has the same maximum score:

Resource class	Maximum score
Dataset	7.5
Distribution	7.5
Data service	7.5

Metadata Quality score calculation

The current methodology follows a hierarchical aggregation of the scores. The basic unit are metadata properties. The sum of such properties generates the score of entities such as Datasets, Distributions and Data Services. The entity scores are then aggregated into their parent Entities. This means that children elements such as Distributions of a Dataset are aggregated into the score of their parent Datasets. Those scores are aggregated into further parent elements they belong to, such as a Catalogue.

Property score

Each key DCAT-AP metric score (as listed in the sections “Dataset metrics”, “Distribution metrics” and “Data service metrics”) is calculated as follows.

Each metric produces a binary result:

- 1 if the metric condition is fulfilled;
- 0 if the metric condition is not fulfilled.

The score for one metric is calculated as:

$$\text{metric score} = \text{weight} \times \text{result}$$

There are no additional penalties if any property is missing. The higher importance of mandatory properties is expressed through the higher weight.

Entity score

The score of each entity (dataset, distribution or data service) is the simple sum of its property scores. Each entity receives an entity score between 0 and 7.5.

entity score = simple sum of property scores

<h3 class="western">Dataset aggregation</h3>

We calculate the final MQA score for a dataset from the dataset score and the average scores of its child resources:

dataset final score = (dataset score + average(distribution scores) + average(data service scores)) / N

Where

N = number of groups actually present (2 or 3).

Each of the three classes carries equal weight (one third each when all are present).

For example:

-

If a dataset has no distributions, the final score is the dataset score itself:

dataset final score = dataset score / 1

-

If a dataset has distributions but no data services:

dataset final score = (dataset score + average distribution score) / 2

- If a dataset has distributions and data services:

$$\text{dataset final score} = (\text{dataset score} + \text{average distribution score} + \text{average data service score}) / 3$$

Each class included in the calculation has equal weight.

Note: for Datasets, average(dataService_scores) covers the embedded data services – those gathered from inside the dataset’s distributions (dcat:accessService); in DCAT-AP 3 a data service is either standalone (catalogue root) or embedded in a distribution – never attached directly to a dataset. Therefore, data services cannot exist without distributions, so N = 2 (Dataset + distributions) or N = 3 (Dataset + distributions + data services).

Standalone data services

A data service can also exist as a standalone resource at catalogue level. A standalone data service is not aggregated into a dataset score. It is scored independently as a data service with a maximum score of 7.5.

Methodology application examples

Case	Input scores	Calculation	Final dataset score
Dataset with one distribution	Dataset: 4.25; Distribution average: 3.25	$(4.25 + 3.25) / 2$	3.75
Dataset with distributions and data services	Dataset: 5.5; Distribution average: 4.1; Data service average: 4.35	$(5.5 + 4.1 + 4.35) / 3$	4.65
Standalone data service	Data service: 5.25	Direct score	5.25

Catalogue aggregation

The catalogue score is calculated as the average of:

- all dataset final scores
- all standalone data service scores

catalogue score = average(dataset final scores + standalone data service scores)

Embedded data services are counted only as part of the dataset score. They are not counted again as standalone data services.

The catalogue score uses the same scale as distributions, data services, and dataset scores: 0 to 7.5.

Deprecated distributions

Deprecated distributions are excluded from the calculation. A distribution is considered deprecated when it has one of the following properties:

adms:status <<http://purl.org/adms/status/Deprecated>>

adms:status <<http://publications.europa.eu/resource/authority/dataset-status/DEPRECATED>>

Deprecated distributions are ignored entirely:

- They are not evaluated.
- No measurements are written for them.
- They are excluded from distribution averages.
- They are excluded from catalogue coverage calculations.

If all distributions of a dataset are deprecated, the dataset is excluded from catalogue aggregation. This differs from a metadata-only dataset with no distributions, which is still scored using its dataset score alone.

Deprecation applies to distributions only.

What the MQA does not cover

The MQA is limited to the metadata harvested by the European Data Portal. It does not assess the intrinsic quality, correctness, completeness, or legal validity of the data itself. If source metadata contains incorrect or incomplete information, this affects the MQA result.

The MQA does not verify whether the described data is correct, legally reusable in practice, semantically accurate, or complete. It evaluates metadata quality according to the defined metrics only.

Dataset metrics

The following metrics are evaluated for datasets (maximum dataset score: 7.5).

Dimension	Indicator	Property	DCAT-AP Importance	Indicator Weight	Metric definition
Findability	Title	dct:title	Mandatory	1.00	The property is set.
Findability	Description	dct:description	Mandatory	1.00	The property is set.
Findability	Keyword usage	dcat:keyword	Recommended	0.50	The property is set.
Findability	Categories	dcat:theme	Recommended	0.50	The property is set.
Findability	Geo search	dct:spatial	Recommended	0.50	The property is set.
Findability	Temporal coverage	dct:temporal	Recommended	0.50	The property is set.
Findability	Modification date	dct:modified	Optional	0.25	The property is set.
Findability	Date of issue	dct:issued	Optional	0.25	The property is set.
Findability	Related resource	dct:relation	Optional	0.25	The property is set.
Findability	Applicable legislation	dcatap:applicableLegislation	Optional	0.25	The property is set.
Accessibility	Identifier	dct:identifier	Optional	0.25	The property is set.
Accessibility	Landing page	dcat:landingPage	Optional	0.25	The property is set.
Accessibility	Identifier	adms:identifier	Optional	0.25	The property is set.
Reusability	Publisher	dct:publisher	Recommended	0.50	The property is set.
Reusability	Contact point	dcat:contactPoint	Recommended	0.50	The property is set.
Reusability	Access rights vocabulary	dct:accessRights	Optional	0.25	https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/access-right
Reusability	Access rights	dct:accessRights	Optional	0.25	The property is set.
Reusability	Provenance	dct:provenance	Optional	0.25	The property is set.

Distribution metrics

The following metrics are evaluated for distributions (maximum distribution score: 7.5).

Dimension	Indicator	Property	DCAT-AP Importance	Indicator Weight	Metric definition
Findability	Description	dct:description	Recommended	0.50	The property is set.
Findability	Title	dct:title	Optional	0.25	The property is set.
Findability	Date of issue	dct:issued	Optional	0.25	The property is set.
Findability	Modification date	dct:modified	Optional	0.25	The property is set.
Accessibility	Access URL	dcat:accessURL	Mandatory	1.00	The property is set and has a valid URL format.
Accessibility	Format	dct:format	Recommended	0.50	The property is set.
Accessibility	Access URL accessibility test	dcat:accessURL	Mandatory	0.25	The URL returns an HTTP 2xx response to a HEAD request.
Accessibility	Download URL	dcat:downloadURL	Optional	0.25	The property is set and has a valid URL format.
Accessibility	Download URL accessibility test	dcat:downloadURL	Optional	0.25	The URL returns an HTTP 2xx response to a HEAD request.
Accessibility	Media type	dcat:mediaType	Optional	0.25	The property is set.
Accessibility	File size	dcat:byteSize	Optional	0.25	The property is set.
Interoperability	Format from vocabulary	dct:format	Recommended	0.50	The controlled vocabulary for file type is used. https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/file-type
Interoperability	Non-proprietary	dct:format	Recommended	0.50	The specified format is marked as non-proprietary in the file-type vocabulary. https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/file-type

Dimension	Indicator	Property	DCAT-AP Importance	Indicator Weight	Metric definition
Interoperability	Machine readable	dct:format	Recommended	0.50	The specified format is marked as machine-readable in the file-type vocabulary. https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/file-type
Interoperability	Media type from vocabulary	dcat:mediaType	Optional	0.25	The media type is found in the IANA register. Source for data.europa.eu: https://ppe.data.europa.eu/api/hub/search/vocabularies/iana-media-types
Reusability	License information	dct:license	Recommended	0.50	The property is set.
Reusability	License vocabulary	dct:license	Recommended	0.50	The controlled vocabulary for licences is used. https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/licence
Reusability	Rights	dct:rights	Optional	0.25	The property is set.
Reusability	Documentation	foaf:page	Optional	0.25	The property is set.
Reusability	Status	adms:status	Optional	0.25	The property is set.

Data service metrics

The following metrics are evaluated for data services (maximum data service score: 7.5).

Dimension	Indicator	Property	DCAT-AP Importance	Indicator Weight	Metric definition
Findability	Title	dct:title	Mandatory	1.00	The property is set.
Findability	Keyword usage	dcat:keyword	Recommended	0.50	The property is set.
Findability	Categories	dcat:theme	Recommended	0.50	The property is set.
Accessibility	Endpoint URL	dcat:endpointURL	Mandatory	1.00	The property is set and has a valid URL format.
Accessibility	Endpoint URL accessibility test	dcat:endpointURL	Mandatory	0.25	The URL is set and returns an HTTP 2xx response to a HEAD request.
Accessibility	Landing page accessibility test	dcat:landingPage	Optional	0.25	The URL is set and returns an HTTP 2xx response to a HEAD request.
Accessibility	Format	dct:format	Optional	0.25	The property is set.
Accessibility	Landing page	dcat:landingPage	Optional	0.25	The property is set.
Interoperability	Format from vocabulary	dct:format	Optional	0.25	The controlled vocabulary for file type is used. https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/file-type
Interoperability	Conforms To	dct:conformsTo	Recommended	0.50	The property is set.
Interoperability	Endpoint description	dcat:endpointDescription	Recommended	0.50	The property is set.
Interoperability	Serves dataset	dcat:servesDataset	Recommended	0.50	The property is set.
Reusability	Publisher	dct:publisher	Recommended	0.50	The property is set.
Reusability	License information	dct:license	Optional	0.25	The property is set.
Reusability	License vocabulary	dct:license	Optional	0.25	The controlled vocabulary for licences is used. https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/licence
Reusability	Description	dct:description	Optional	0.25	The property is set.
Reusability	Access rights	dct:accessRights	Optional	0.25	The property is set.

Dimension	Indicator	Property	DCAT-AP Importance	Indicator Weight	Metric definition
Reusability	Documentation	foaf:page	Optional	0.25	The property is set.

Changes from the previous methodology

There are some differences between the current and the previous methodology. The first one is the starting point. More specifically, during the previous methodology, emphasis was put on the FAIR principles and how DCAT-AP properties are mapped to those principles. Current methodology puts more emphasis on the importance of DCAT-AP properties and their correct implementation in the European Data Portal and less on the FAIR principles.

Additionally, the new methodology takes under consideration the notion of data services as those are newly introduced in the latest version of DCAT-AP (3.0).

Another novelty of the new methodology lies on the hierarchical calculation of the overall score, but also the possibility that a data service may not be related to a dataset.

One important difference is the fact that the previous methodology didn't take under consideration the deprecated datasets in the overall score, creating distortions in the final MQA scores. This is not the case anymore.

The following table provides more insights into the differences between the two versions of the methodology.

Scoring feature	Previous methodology	Current methodology	Benefit
Weighting	Fixed absolute points per metric	Layer-based weights: 1 for mandatory, 0.5 for recommended and 0.25 for optional properties. The values of the weights were chosen to scale and highlight the importance of the property.	Consider the importance of each DCAT-AP property. Fairer evaluation.
Maximum score	405 total points	7.5 per resource class, hierarchical aggregation.	Smaller range, making it easier to grasp MQA results.
Dimensions	Five dimensions including Contextuality	Four FAIR-related dimensions; Contextuality removed	Focus on what matters most for data providers and data portals.
Data services	Not scored as first-class resources	Scored as first-class resources	Included and fair scoring of data services.
Dataset score	Used best distribution scoring logic	Dataset score is averaged with distribution and data service averages	Fairer scoring since it takes under consideration the whole annotation effort and depicts reality better.
Accessibility	HTTP 2xx and 3xx counted as accessible	Only HTTP 2xx counts as accessible	Enforcing that datasets and data services can be used.
Deprecated distributions	Not handled explicitly	Deprecated distributions are ignored; datasets with only deprecated distributions are excluded	Fairer scoring.
SHACL compliance	Included in score	Kept but not scored	Scores are based strictly on the metadata properties.
Rating bands	Bad, sufficient, good, excellent	Sufficient: 0 - 2.5; Good: 2.5 - 5; Excellent: 5.0 - 7.5	Make scores easier to understand.

Official INSPIRE Data and Services Catalogue (codsi)

Rating: Good - 3.53 / 7.5

Datasets: 942

Distributions: 2181

Data service metrics: 1754

Official INSPIRE Data and Services Catalogue (codsi)

Dataset metrics

Accessibility

Indicator	Indicator Weight	Catalogue metric coverage
ADMS Identifier	0.25	0.0%
DCT Identifier	0.25	85.0%
Landing Page	0.25	32.9%

Findability

Indicator	Indicator Weight	Catalogue metric coverage
Applicable Legislation	0.25	91.8%
Categories	0.5	100.0%
Date of issue	0.25	62.1%
Description	1.0	100.0%
Geo search	0.5	98.6%
Keyword usage	0.5	14.3%
Modification date	0.25	35.5%
Related Resource	0.25	0.0%
Time based search	0.5	33.4%
Title	1.0	100.0%

Reusability

Indicator	Indicator Weight	Catalogue metric coverage
Access Rights Vocabulary	0.25	0.0%
Access restrictions	0.25	0.0%
Contact point	0.5	75.6%
Provenance	0.25	44.3%
Publisher	0.5	6.5%

Distribution metrics

Accessibility

Indicator	Indicator Weight	Catalogue metric coverage
Access URL accessibility test	0.25	78.4%
AccessURL available	1.0	100.0%
Download URL accessibility test	0.25	-
Download URL available	0.25	0.0%
File size	0.25	0.0%
Format	0.5	55.8%
Media type	0.25	0.0%

Findability

Indicator	Indicator Weight	Catalogue metric coverage
Date of issue	0.25	66.0%
Description	0.5	24.5%
Modification date	0.25	38.7%
Title	0.25	96.6%

Interoperability

Indicator	Indicator Weight	Catalogue metric coverage
Format from Vocabulary	0.5	46.9%
Machine Interpretable	0.5	0.0%
Media Type from Vocabulary	0.25	0.0%
Non-proprietary	0.5	40.5%

Reusability

Indicator	Indicator Weight	Catalogue metric coverage
Documentation	0.25	0.0%
Known Licence	0.5	0.0%
License information	0.5	100.0%
Rights	0.25	0.2%
Status	0.25	0.0%

Official INSPIRE Data and Services Catalogue (codsi)

Data service metrics

Accessibility

Indicator	Indicator Weight	Catalogue metric coverage
Endpoint URL	1.0	100.0%
Endpoint URL Status Code	0.25	92.0%
Format	0.25	0.0%
Landing Page	0.25	0.0%
Landing Page Status Code	0.25	-

Findability

Indicator	Indicator Weight	Catalogue metric coverage
Categories	0.5	0.0%
Keyword usage	0.5	0.0%
Title	1.0	100.0%

Interoperability

Indicator	Indicator Weight	Catalogue metric coverage
Conforms To	0.5	0.0%
Endpoint Description	0.5	100.0%
Format from Vocabulary	0.25	0.0%
Serves Dataset	0.5	0.0%

Reusability

Indicator	Indicator Weight	Catalogue metric coverage
Access restrictions	0.25	0.0%
Description	0.25	0.0%
Documentation	0.25	0.2%
Known Licence	0.25	0.0%
License information	0.25	0.0%
Publisher	0.5	0.0%