

Metadata Quality Assessments (MQA)

Quality Assessments of Implementation of Metadata Profiles for National Portals

Content

1. Introduction	3
2. Background.....	4
2.1 The European Data Portal (EDP)	4
2.2.1 National Open Data Portals.....	4
2.2.2 Catalogues in the Spatial Data Infrastructures (SDIs)	5
2.2 Metadata Quality Analysis (MQA) Tool.....	5
2.2.1 The MQA Process	5
2.2.2 Assumptions	6
2.2.3 Dimensions	7
2.3.4 Rating	12
3. Analysis of MQA Results.....	13
3.1 Spain	14
3.1.1 Illustrative Example: Dataset from the Spanish National Open Data Portal.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.1.2. Complete Analysis Spanish Datasets.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.2 Finland.....	17
3.2.1 Illustrative Example: Dataset from the Finnish National Open Data Portal.....	Virhe. Kirjanmerkkiä ei ole määritetty.
2.3 The Netherlands.....	24
3.3.1 Illustrative Example: Dataset from the Dutch Spatial Data Infrastructure	Virhe. Kirjanmerkkiä ei ole määritetty.
3.4 Norway	19
3.4.1 Illustrative Example: Dataset from the Norwegian Spatial Data Infrastructure	19
2.5 Estonia	20
3.4.1 Illustrative Example: Dataset from the Estonian Spatial Data Infrastructure Portal	Virhe. Kirjanmerkkiä ei ole määritetty.
3.6 Cross-Project Analysis	Virhe. Kirjanmerkkiä ei ole määritetty.
3.6.1 Title and abstract.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.6.2 Distribution.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.6.3 Spatial Extent.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.6.4 Identifier	Virhe. Kirjanmerkkiä ei ole määritetty.

3.6.5 Conclusions and Recommendations	36
3.7.1 Metadata from Open Data Portals.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.7.2 Metadata from Spatial Data Infrastructure	Virhe. Kirjanmerkkiä ei ole määritetty.
3.7 Cross-Partner Analysis.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.8 Interoperability Map	Virhe. Kirjanmerkkiä ei ole määritetty.
3.8.1 Spain Interoperability Map	Virhe. Kirjanmerkkiä ei ole määritetty.
3.8.2 Finland Interoperability Map	Virhe. Kirjanmerkkiä ei ole määritetty.
3.8.3 Netherlands Interoperability Map	Virhe. Kirjanmerkkiä ei ole määritetty.
3.8.1 Norway Interoperability Map.....	Virhe. Kirjanmerkkiä ei ole määritetty.
3.8.5 Estonia Interoperability Map	Virhe. Kirjanmerkkiä ei ole määritetty.

1. Introduction

Metadata is an essential component in supporting the discovery and reuse of data across time and in a variety of contexts. In the European geospatial context, metadata requirements and the standards used for metadata differ depending on the portal or environment in which metadata is being published meaning that metadata published on the national level might need to be mapped and transformed to an alternative profile for publication in a European context. It is, therefore, important that the metadata created at the national level according to the national profiles is as complete and as extensive as possible in order to ensure that the quality of this metadata is maintained at the European level.

Within the European context, various metadata profiles exist, each of which are based on a different standardisation effort or a specific initiative. Depending on the goal of the associated effort or initiative, each profile has slightly different compliancy requirements. For example, within the context of the GeoE3 project, the INSPIRE metadata profiles (which are extended versions of the ISO 19115, ISO 19119 and ISO 19139 standards) support the documentation of geospatial data at the European level. DCAT and GeoDCAT present profiles which facilitate interoperability between open data and open geospatial data and the schema.org profile represents an effort to support the findability of datasets on the web. One of the objectives of GeoE3 is making metadata more accessible and the aim of Task 2.2 is making the connection between the national infrastructure and the European Data Portal (EDP) so that the metadata of datasets can be accessible not only through the national portals but also through the EDP.

The results of Task 2.2 were made available as part of a preceding document wherein the successful connection with the EDP was analysed per partner country. As a follow up to this previous report, this report now collects the results of a whether the metadata included in the EDP meets the quality requirements of the portal based on the Metadata Quality Assessment tool (MQA) made available by the EDP. The following sections provided details on the process carried out in collecting the MQA results for each partner country. In the section, Analysis of MQA results, there is an overview of MQA results per the datasets from 5 countries that are used in the project. The final chapter summarises these findings and provides minor indications for improvements.

2. Background

In this section, an overview of the European Data Portal itself as well as the MQA tool is provided.

2.1 The European Data Portal (EDP)

The European Data Portal (EDP) was created to gather Public Sector Information from the 28 European Member States and the four EFTA states that be freely re-used for any purpose (i.e. open data on the European level). The metadata available in the portal is directly harvested from the existing portals with the aim of providing one single access point to all data.

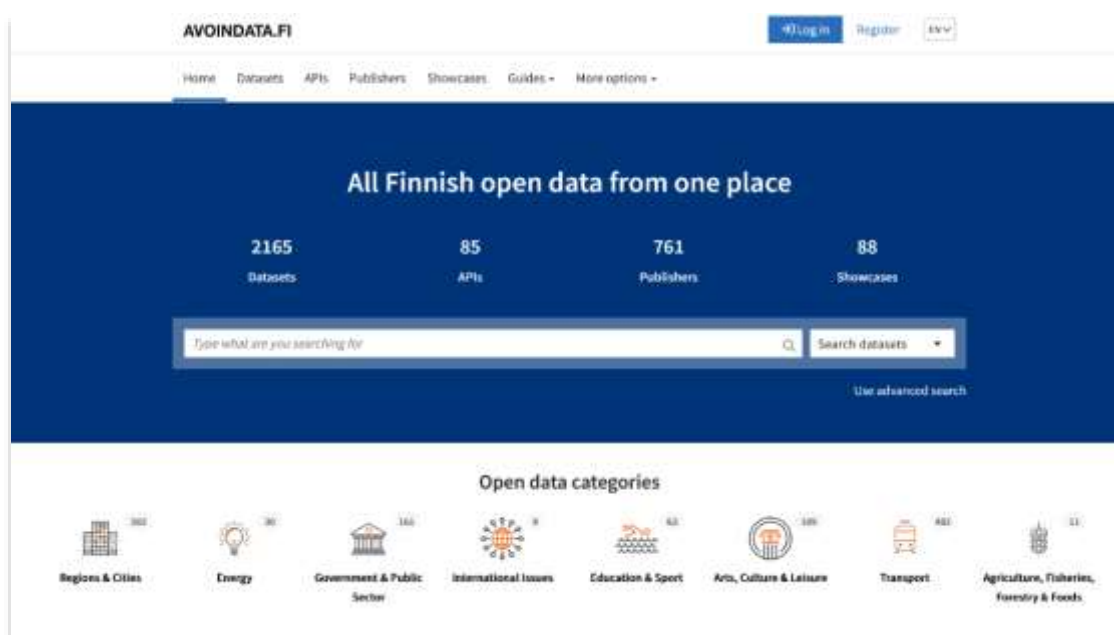
In order to foster comparability of data published across borders, the portal presents metadata references in a common metadata profile format, namely; DCAT which is a RDF vocabulary. The portal provides translations of dataset descriptions in all 24 languages using machine-translation technology.

The EDP collects not only from the national open data portals but also from other catalogues such as the Spatial Data Infrastructure (SDI) catalogues maintained by national institutions. When the EDP harvests metadata compliant with the INSPIRE metadata profile from a given portal, the metadata is transformed to GeoDCAT-AP rather than DCAT-AP due to the fact that the latter is not well equipped for the representation of geospatial components available for spatial metadata descriptions. The information in one metadata profile can be mapped into another metadata profile. EDP does the transformation automatically which means the metadata that is published in the national geoportals is directly available in the EDP.

2.2.1 National Open Data Portals

Open data portals are web-based interfaces designed to make it easier to access and reuse the information of the public sector. They offer, among other things, geographical information. Open data portals generally present metadata according to DCAT-AP or GeoDCAT-AP.

Figure 4. Example of the Finnish Open Data Portal Landing Page



2.2.2 Catalogues in the Spatial Data Infrastructures (SDIs)

Most European countries have their own spatial data infrastructure including an associated geoportal and a discovery service to publish dataset descriptions and other geospatial resources in order to make these openly available. These discovery services are based on the OGC Standard Catalogue Service for the Web (CSW) and it is very common that they are implemented using GeoNetwork. The metadata files published on an open geoportal can apply the INSPIRE metadata profile, ISO 19115, ISO 19119 and others. The central European access point to the data provided by EU Member States and several EFTA countries under the INSPIRE Directive is the INSPIRE Geoportal¹. These metadata using in this geoportal are regularly harvested from the discovery services of the EU Member States and the EFTA countries. Therefore, the metadata files will always apply the INSPIRE metadata profile in this case.

For the purpose of this report, only the Open Data Portals are assessed.

2.2 Metadata Quality Analysis (MQA) Tool

The Metadata Quality Assessment (MQA) is a tool that is developed by the consortium of data.europa.eu for the purpose of studying the quality of metadata harvested by data.europa.eu. The intention with this tool is to support data providers and portals to improve the quality of their metadata by providing suggestions for improvements. The tool itself is supported by a methodology defined to investigate the quality of metadata based concretely on the following questions. These questions also define the scope of the tool:

- Does the metadata comply with DCAT-AP and DCAT-AP Derivatives?
- Does the metadata contain information for which DCAT-AP is not extensive enough?
- Is the accessibility of data referenced in the metadata through the access and download URL?
- Is the (meta)data machine readable?
- Has a license been used in the metadata?

For the results of the tool to be accurate, it is important that the source data includes minimal errors. In order to support portals and data providers with the provision of valid metadata, the following validation services are offered. This supports provides in validating 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](#)

2.2.1 The MQA Process

With each harvesting of metadata from the source portal, the metadata is checked by the MQA tool. Here, various indicators are measured as described in the tables below (see Dimensions tables). The results of these checks are stored using a structured vocabulary Data Quality Vocabulary (DQV)².

Quality checks on metadata are not done frequently with the exception of the accessibility indicators (accessURL and downloadURL). Indeed, because accessibility can be volatile, repeated

¹ [INSPIRE Geoportal](#)

² <https://www.w3.org/TR/vocab-dqv/>



checks on these indicators are done and, in contrast to other indicators, this has a higher runtime because these distributions are checked via HTTP and each requested URL has a longer response time. Each accessibility URL is checked within a few weeks of the previous check.

2.2.2 Assumptions

The following assumptions are made by the MQA tool:

- **Use of Non-Obligatory Attributes:** It is assumed that fulfilling the requirements of the DCAT-AP profile alone is not sufficient for high quality metadata. For this reason, checks are also done on the availability of non-mandatory information in an application profile.
- **Identical content for multiple distributions:** Where a dataset contains one or more distributions, all are identical in content but are different in representation of the data. For example, a dataset can have two distributions, one offering the data PDF and the other offering the identical data as machine-readable RDF/XML.

2.2.3 Dimensions

The following are dimensions investigated by the MQA tool in determining metadata quality. These are derived from the FAIR principles³.

Findability

The following described the metrics which help humans and machines find relevant datasets. A maximum of 100 points can be scored in this area.

Table 1. Findability

Indicator	Description	Metrics	Weight
Keyword Usage	Keywords directly support the search and thus increase the findability of the dataset	The system checks whether keywords are defined. The number of keywords has no impact on the score. dcat:keyword	30
Categories	Categories help users to explore datasets thematically	It is checked whether one or more categories are assigned to the dataset. The number of categories has no impact on the score. dcat:theme	30
Geo Search	Usage of spatial information would enable users in order to find the dataset with a geo facettted search.	It is checked whether the property is set or not. dct:spatial	20
Timebased Search	Usage of temporal information would enable users for a timely based facettted search.	It is checked whether the property is set or not. dct:temporal	20

Accessibility

The following describes the metrics which determine whether access to the data is guaranteed by the referenced by the distributions. A maximum of 100 points can be scored in this area.

³ <https://www.go-fair.org/fair-principles/>

Table 2. Accessibility

Indicator	Description	Metrics	Weight
AccessURL accessibility	The AccessURL is not necessarily a direct link to the data, but also may refer to a URL that gives access to the dataset or where more information about the dataset is available.	The specified URL is checked for accessibility via a HTTP HEAD request. If the responded status code is in the 200 or 300 range, the accessibility of the resource is evaluated positively. dcat:accessURL	50
DownloadURL	The downloadURL is a direct link to the referenced data.	It is checked whether this property is set or not. dcat:downloadURL	20
DownloadURL accessibility	If a download URL exists, the accessibility is checked.	The specified URL is checked for accessibility via a HTTP HEAD request. If the responded status code is in the 200 or 300 range, the accessibility of the resource is evaluated positively. dcat:dowloadURL	30

Interoperability

The following describes the metrics used to determine whether a distributions is considered interoperable. A maximum of 110 points can be scored in this area.

Table 3. Interoperability

Indicator	Description	Metrics	Weight
Format	This field specifies the file format of the distribution.	It is checked whether the property is set or not. Distribution dct:format	20
Media Type	This field specifies the media type of the distribution.	It is checked whether the property is set or not. Distribution	10

		dcats:mediaType	
Format/Media Type from vocabulary	Checks whether format and media type belong to a controlled vocabulary.	<p>The format vocabulary can be found in the data.europa.eu GitLab repository.</p> <p>The media type is check against the IANA list</p> <p>Distribution dct:format dcats:mediaType</p>	10
Non-proprietary	Checks if the format of the distribution is non-proprietary.	<p>The distribution is considered as non-proprietary if the specified format is contained in the corresponding data.europa.eu GitLab repository vocabulary.</p> <p>Distribution dct:format</p>	20
Machine readable	Checks if the format of the distribution is machine-readable.	<p>The distribution is considered as machine-readable if the specified format is contained in the corresponding data.europa.eu GitLab repository vocabulary.</p> <p>Distribution dct:format</p>	20
DCAT-AP Compliance	<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.</p> <p>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</p>	<p>The metadata is validated against a set of SHACL shapes. The metadata is not compliant, if the SHACL validation reports at least one issue.</p> <p>The MQA uses data.europa.eu's DCAT-AP SHACL validation service.</p> <p>SHACL is a recommendation from the W3C and is used for validating RDF graphs against a set of shapes.</p>	30

	DCAT-AP. In order to increase conformity to DCAT-AP, the MQA checks each metadata for its DCAT-AP compliance.		
--	---	--	--

Reusability

The following table describes which metrics are used to check the reusability of the data. A maximum of 75 points can be scored in this area.

Table 4. Reusability

Indicator	Description	Metrics	Weight
License Information	A license is valuable information for the reuse of data.	It is checked whether the property is set or not. Distribution dct:license	20
License Vocabulary	We would like to limit the indication of incorrect license information. For example, we encounter many CC licenses that lack versioning.	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
Access Restrictions	This field indicates whether the access to the data is public or restricted.	It is checked whether the property is set or not. Dataset dct:accessRights	10
Access Restrictions Vocabulary	The use of a controlled vocabulary increases reusability.	It is checked whether the controlled vocabulary for access rights is used.	5

		Dataset dct:accessRights	
Contact Point	The contact point contains inform whom to address in case of questions regarding the data.	It is checked whether the property is set or not. Dataset dcat:contactPoint	20
Publisher	It is checked whether the property is set or not.	It is checked whether the property is set or not. Dataset dct:publisher	10

Contextuality

The following indicates some light-weight properties that provide the user with more context. A maximum of 20 points can be scored here.

Table 5. Contextuality

Indicator	Description	Metrics	Weight
Rights	In some cases, a specific license cannot be applied to a dataset. The 'Rights' field can be used to specify a reference to a resource that will inform a user about the rights he has when using the dataset.	It is checked whether the property is set or not. Distribution dct:rights	5
File size	Specifies the size of the file in bytes.	It is checked whether the property is set or not. Distribution dcat:byteSize	5
Date of Issue	The date on which the dataset or distribution was released.	It is checked whether the property is set or not. Dataset and Distribution dct:issued	5

Modification Date	The date on which the dataset or distribution was last changed.	It is checked whether the property is set or not. Dataset and Distribution dct:modified	5
--------------------------	---	---	---

2.3.4 Rating

The final rating is done based on four groups as shown in the following tables. The mapping of points to the rating category is shown in the table below. The representation of the rating in the MQA is expressed exclusively via the rating categories. This enables providers to achieve the highest rating even with a slight deduction in points.

Table 6. Rating

Dimension	Maximal Points
Findability	100
Accessibility	100
Interoperability	110
Reusability	75
Contextuality	20
Sum	405

Table 7. Range Mapping

Rating	Range of Points
Excellent	351-405
Good	221-350
Sufficient	121-220
Bad	0-120

3. Comparative Analysis: Connection between EDP and National (Geo)Portals

The first deliverable required in Activity 2 of the GeoE3 project is the analysis of whether the metadata for datasets being used in the project was being (correctly) harvested by the EDP. In order to assess this, each dataset defined for a particular use case for each country was checked in turn. The check performed assessed whether the metadata in the national open data portal or spatial data infrastructure catalogue exists, is complete and then assessed whether this metadata was being harvested by the EDP and whether the resulting metadata was done correctly and whether this was complete. The completeness and correctness of metadata is assessed based on the presence or absence of the following metrics.

Title and abstract
Distributions
Spatial extent
Identifiers
Keywords

The sections that follow will each the datasets provided by each partner country in turn and provide comments on what improvements could be made in order to improve the completeness of the metadata in both the EDP and the national portals.

3.1 Spain

The following summarises the findings of the comparative analysis performed. It should be noted that much of the metadata assessed in the following table are published by the Spatial Data Infrastructure for Spain (IDEE) with only one, the last metadata record, being published by the Open Data Portal for Spain. The following table highlights some areas of improvement with regard to the following:

1. There is a tendency for the EDP to harvest metadata from the national infrastructure and publish this using the incorrect metadata field. Indeed, distributions are often harvested as landing pages or documents rather than distributions.
2. Keywords are not always available in the EDP. This is generally a harvesting error rather than a lack of keywords in the source portal.
3. National portals do not always have a rich metadata description.

Table 1. Analysis Metadata Availability Between Spanish National Geoportal (IDEE) and EDP

No.	Dataset Name	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	Transport Network	ok	harvested as landing page and as a distribution	ok	ok	ok
2	Geographical Names	ok	harvested as landing page and as a distribution	ok	ok	ok
3	Hydrography	ok	harvested as landing page and as a distribution	ok	ok	ok
4	Administrative Units	ok	harvested as landing page and as a distribution	ok	ok	ok
5	Addresses	ok	harvested as landing page not as a distribution	ok	ok	ok
6	National Topographical Base	ok	harvested as landing page and as a distribution	ok	ok	no keywords in national portal

7	Buildings	ok	harvested as landing page and as a distribution	ok	ok	ok
8	Digital Terrein Model (MDT200)	Limited description in national portal	ok	ok	ok	ok
9	Digital Terrein Model (MDT25)	Limited description in national portal	ok	ok	ok	ok
10	Digital Terrein Model (MDT05)	Limited description in national portal	ok	ok	ok	ok
11	Digital Terrein Model (MDT1000)	Limited description in national portal	ok	ok	ok	ok
12	Digital Terrein Model (MDT500)	Limited description in national portal	ok	ok	ok	ok
13	Digital Surface Model (MDS02)	Limited description in national portal	ok	ok	ok	ok
14	Digital Surface Model (MDS05)	Limited description in national portal	ok	ok	ok	ok
15	Digital Surface Model (MDSn2.5)	Limited description in national portal	ok	ok	ok	ok
16	Digital Surface Model (MDS2.5)	Limited description in national portal	ok	ok	ok	ok
17	Maximum Temperature	ok	harvested as landing page and as a distribution. No title available for distribution	ok	ok	Keywords not harvested by EDP
18	Average Temperature Coldest Month	ok	harvested as landing page and as a	ok	ok	Keywords not harvested by EDP

			distribution. No title available for distribution			
19	Average Annual Temperature	ok	harvested as landing page and as a distribution. No title available for distribution	ok	ok	Keywords not harvested by EDP
20	Weather Stations	ok	No distributions harvested by the EDP	ok	ok	Keywords not harvested by EDP
21	Sunshine Atlas	ok	harvested as landing page and as a distribution. No title available for distribution	ok	ok	ok

3.2 Finland

The following summarises the findings of the comparative analysis performed. It should be noted that all metadata assessed in the following table are published by the Finnish Open Data Portal. The following table highlights some areas of improvement with regard to the following:

1. There is currently no metadata available for the DSM. This should be added to the national portal.
2. Keywords which exist in the national portals are not being harvested by the EDP.
3. Several metadata records do not have distributions harvested in the EDP.

Table 2. Analysis Metadata Availability Between Finnish National Portal and EDP

No.	Dataset Name	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	Background Map	ok	ok	ok	ok	ok
2	Terrain Database	ok	ok	ok	ok	ok
3	Buildings 3D	ok	ok	ok	ok	ok
4	Roads	ok	ok	ok	ok	ok
5	Height Model 2m	ok	ok	ok	ok	ok
6	DSM	No public metadata available				
7	Weather Observations	ok	No distributions harvested by EDP	ok	ok	No keywords harvested by EDP
8	Wind Atlast 250m	ok	No distributions harvested by EDP	ok	ok	No keywords harvested by EDP
9	Solar Radiation Observations	ok	No distributions harvested by EDP	ok	ok	No keywords harvested by EDP

3.3 The Netherlands

The following summarises the findings of the comparative analysis performed. It should be noted that all metadata assessed in the following table are published by the Dutch Open Data Portal. The following table highlights some areas of improvement with regard to the following:

1. Spatial extent is never present in the national portal so cannot be harvested by the EDP.

Table 3. Analysis Metadata Availability Between Dutch National Portal and EDP

No.	Dataset Name	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	BRT Background Map	ok	ok	No spatial extent recorded in the national portal	ok	ok
2	Key Register Addresses and Buildings	ok	ok	No spatial extent recorded in the national portal	ok	ok
3	DTM	ok	ok	No spatial extent recorded in the national portal	ok	ok
4	DSM	ok	ok	No spatial extent recorded in the national portal	ok	ok
5	Daily Mean Temperature	ok	ok	No spatial extent recorded in the national portal	ok	ok

3.4 Norway

The following summarises the findings of the comparative analysis performed. It should be noted that all metadata assessed in the following table are published by the Norwegian Geoportal. The following table highlights some areas of improvement with regard to the following:

1. Metadata for one dataset is not yet being harvested by the EDP
2. National portal metadata could include richer descriptions in the metadata.
3. One dataset does not yet have publicly available metadata.

Table 4. Analysis Metadata Availability Between Norwegian National Portal and EDP

No.	Dataset Name	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	Background Map	ok	ok	ok	ok	ok
2	INSPIRE Buildings WFS	National portal could include more descriptive text	ok	ok	ok	ok
3	Roads	ok	ok	ok	ok	ok
4	National Terrain Model	ok	ok	ok	ok	ok
5	National Surface Model	ok	ok	ok	ok	ok
6	Frost Climate API	ok	ok	ok	ok	ok

3.5 Estonia

The following summarises the findings of the comparative analysis performed. It should be noted that all metadata assessed in the following table are published by the Estonian Geoportal. At present, the connection between the EDP and the Estonia Geoportal is successful based on the metrics used to measure the connection.

Table 5. Analysis Metadata Availability Between Estonian National Portal and EDP

No.	Dataset Name	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	Estonian Topography Database	ok	ok	ok	ok	ok
2	Main map of Estonia 1:10000	ok	ok	ok	ok	ok
3	Orthophotos Estonia	ok	ok	ok	ok	ok
4	Shaded Relief Imagery	ok	ok	ok	ok	ok
5	Contours and Elevations Map Estonia	ok	ok	ok	ok	ok
6	Estonian Topography Buildings	ok	ok	ok	ok	ok
7	Estonian Road Register	ok	ok	ok	ok	ok
8	Estonian Topography Transport	ok	ok	ok	ok	ok
9	DTM	ok	ok	ok	ok	ok
10	DSM	ok	ok	ok	ok	ok
11	Weather and Climate Data	ok	ok	ok	ok	ok

4. Analysis of MQA Results For National Data Portals

In this section, an overview of the results of the MQA analysis is provided for all partner countries. The MQA analysis is carried out in three parts. Firstly, the MQA tool is used to assess the quality of metadata in the catalogue across a range of metrics. The results of this analysis are provided in a screenshot and discussed briefly for each country. Because compliance with DCAT-AP was an important milestone in this project, each national portal is then assessed for its level of compliance with DCAT-AP. To assess this, the number of total metadata records available in each national portal is noted and the MQA reporting is used to assess the percentage of DCAT-AP compliance for the portal as whole. In no case was this DCAT-AP compliance 100% and, as such, it is necessary to assess whether the metadata records specifically used in this project are at least compliant with DCAT-AP. This is assessed in the final part of the analysis and provided as a table. At the conclusion of each country section, comments on the interaction between this analysis and the one performed in the previous section are provided and comments on where potential for future improvement can be seen per partner country are made.

4.1 Spain

The following figure provides a screenshot of the assessed level of metadata quality for the Spanish Spatial Data Infrastructure Portal (IDEE) using the MQA Analysis tool. This provides an overview of all metadata attributes that require attention generally across the entire portal.

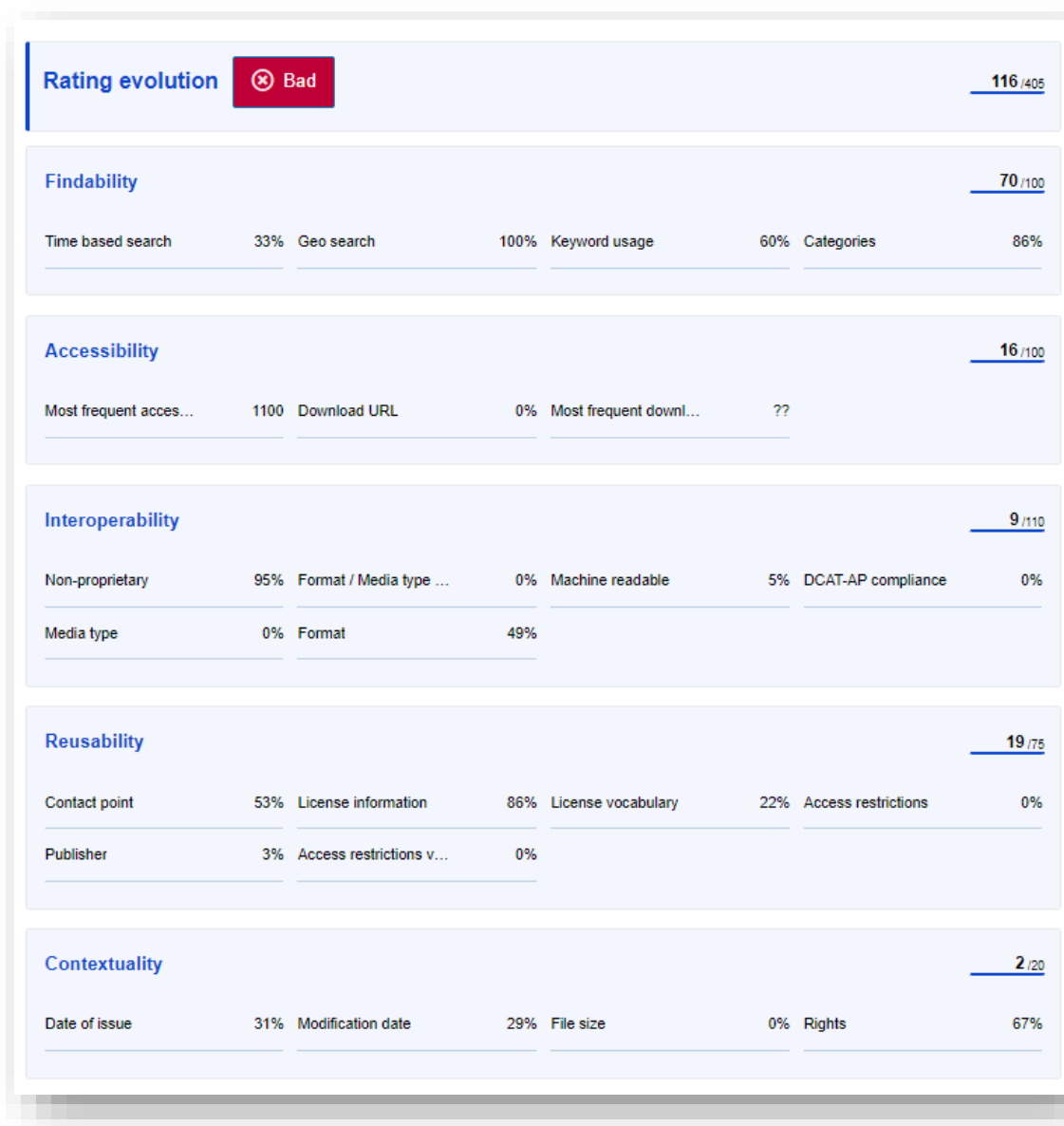
The following are the most notable areas of improvement for the publication of metadata in the national spatial infrastructure:

1. Ensure the inclusion (through validation) of access restrictions (including the lack thereof) in metadata.
2. An investigation on the availability of download URLs and whether these are being correctly defined in the national portal/harvested by the EDP should be investigated. In the previous investigation, there seemed to be a mismatch between the national portals and EDP.
3. Ensure the availability of simple attributes such as publisher and contact information as well as licensing information.

Not all errors are the result of a lack of information in the national portals. Metadata stewards should take care to assess the cause of the lack of information in the EDP based on the SHACL violations of each metadata record made available here:

<https://data.europa.eu/mqa/catalogues/idee/violations?locale=en>

Figure 1. Results of MQA Analysis – Spatial Data Infrastructure Spain (IDEE)



DCAT-AP Compliance for Spatial Infrastructure Portal Spain

In the above figure, DCAT-AP is only one of the metrics by which metadata quality is measured. As one of the key deliverables for this project, however, the lack of compliance seen here needs to be assessed more closely. The following table provides the number of metadata records currently published in the national spatial data infrastructure for Spain and the percentage of these that are compliant with DCAT-AP according to the MQA tool.

Table 6. DCAT-AP Compliance for All Metadata Records Published by Spatial Infrastructure Portal Spain

No.	Description	Results
1	Number of metadata records available	77,873
2	Percentage DCAT-AP Compliance	0%

Link: <https://data.europa.eu/api/mqa/reporter/report/en/pdf?catalogueId=idee>

The general lack of DCAT-AP compliance is likely due to the standards in which metadata is published in the national portal. Progress made by the project to improve this will be discussed in the concluding chapter of this document. The following assesses whether any of the metadata of relevance to this project are compliant with DCAT-AP.

DCAT-AP Compliance per Spanish Dataset Used in Use Cases

Each of the metadata records were assessed using the MQA tool in turn. The following table denotes whether or not a particular record was assessed to be compliant or not. To see the full analysis per record, please visit the link in the table.

Table 7. Analysis Metadata Quality

No.	Dataset Name	Link to MQA Result	DCAT-AP Compliance?
1	Transport Network	Link	No
2	Geographical Names	Link	No
3	Hydrography	Link	No
4	Administrative Units	Link	No
5	Addresses	Link	No
6	National Topographical Base	Link	No
7	Buildings	Link	No
8	Digital Terrein Model (MDT200)	Link	No
9	Digital Terrein Model (MDT25)	Link	No
10	Digital Terrein Model (MDT05)	Link	No
11	Digital Terrein Model (MDT1000)	Link	No
12	Digital Terrein Model (MDT500)	Link	No
13	Digital Surface Model (MDS02)	Link	No
14	Digital Surface Model (MDS05)	Link	No
15	Digital Surface Model (MDSn2.5)	Link	No
16	Digital Surface Model (MDS2.5)	Link	No
17	Maximum Temperature	Link	No
18	Average Temperature Coldest Month	Link	No
19	Average Annual Temperature	Link	No
20	Weather Stations	Link	No
21	Sunshine Atlas	Link	Yes

4.2 Finland

The following figure provides a screenshot of the assessed level of metadata quality for the Finnish Open Data Portal using the MQA Analysis tool. This provides an overview of all metadata attributes that require attention generally across the entire portal.

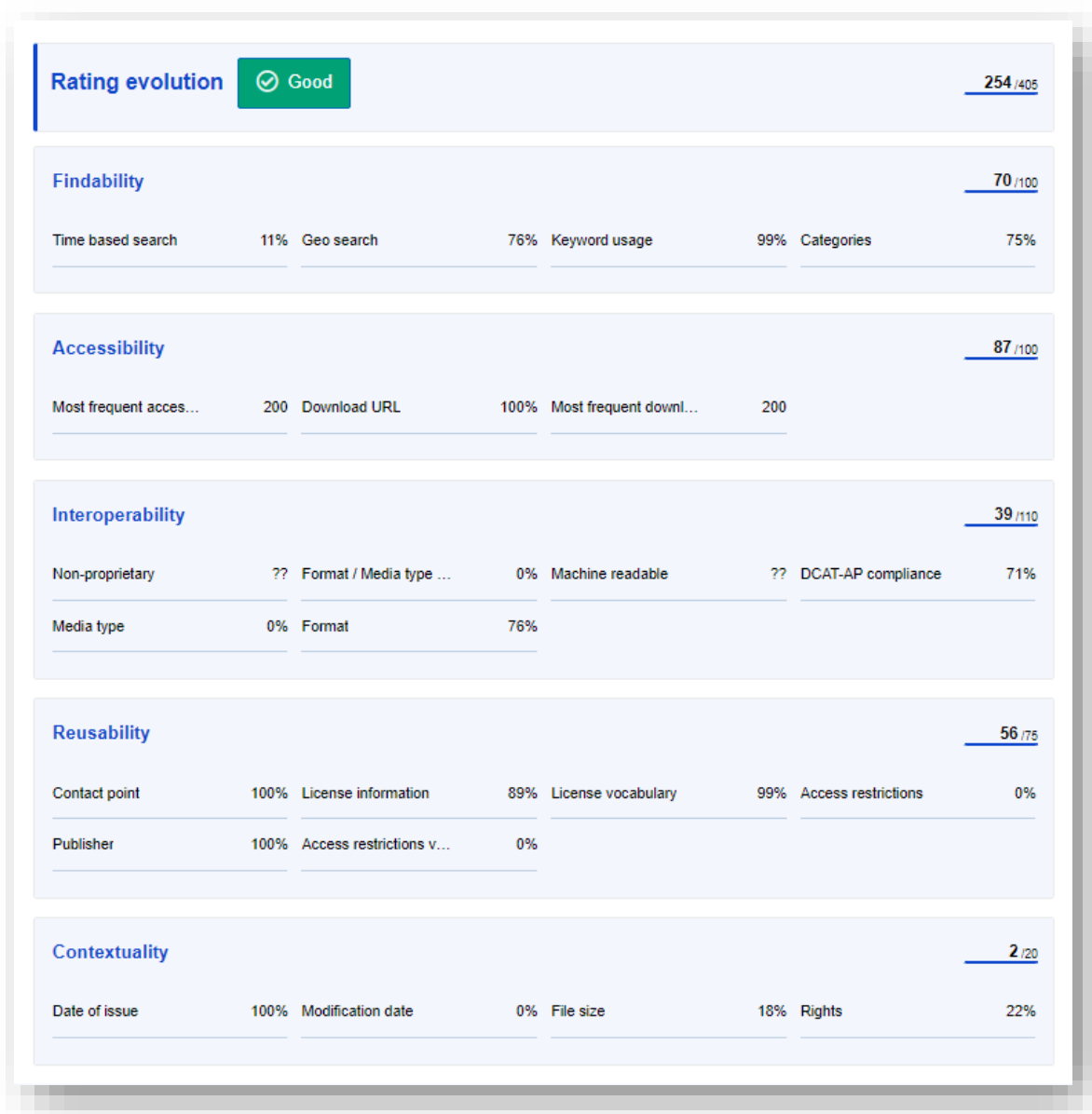
The following are the most notable areas of improvement for the publication of metadata in the national portal:

1. Ensure the inclusion of temporal and modification information in metadata records.
2. Ensure the inclusion of access restrictions (including the lack thereof) in metadata.
3. The machine readability the data is measured by the MQA tool. This is likely to be based on the lack of `dct:format` use in the metadata. With the goal of improving the MQA score, this should be universally implemented for metadata in the portal.
4. Stricter validation methods on metadata in general could improve the overall score of the quality of metadata in the catalogue.

Not all errors are the result of a lack of information in the national portals. Metadata stewards should take care to assess the cause of the lack of information in the EDP based on the SHACL violations of each metadata record made available here:

<https://data.europa.eu/mqa/catalogues/open-data-finland/violations?locale=en>

Figure 2. Results of MQA Analysis – Open Data Finland



DCAT-AP Compliance for Open Data Portal Finland

In the above figure, DCAT-AP is only one of the metrics by which metadata quality is measured. As one of the key deliverables for this project, however, the lack of compliance seen here needs to be assessed more closely. The following table provides the number of metadata records currently published in the national open data portal and the percentage of these that are compliant with DCAT-AP according to the MQA tool.

Table 8. DCAT-AP Compliance for All Metadata Records Published by Open Data Portal Finland

No.	Description	Results
1	Number of metadata records available	1,784
2	Percentage of DCAT-AP compliance	71%

Link: <https://data.europa.eu/api/mqa/reporter/report/en/pdf?catalogueId=open-data-finland>

Although the DCAT-AP compliance for Finland is good, there is still room for improvement. Progress made by the project to improve this will be discussed in the concluding chapter of this document. The following assesses whether any of the metadata of relevance to this project are compliant with DCAT-AP.

DCAT-AP Compliance per Finnish Dataset used in Use Cases

Each of the metadata records were assessed using the MQA tool in turn. The following table denotes whether or not a particular record was assessed to be compliant or not. To see the full analysis per record, please visit the link in the table.

Table 9. Analysis Metadata Quality

No.	Dataset Name	Link to MQA Result	DCAT-AP Compliance?
1	Background Map	MQA	Yes
2	Terrein Database	MQA	Yes
3	Buildings 3D	MQA	Yes
4	Roads	MQA	Yes
5	Height Model 2m	MQA	Yes
6	DSM	No metadata	
7	Weather Observations	MQA	No
8	Wind Atlas 250m	MQA	No
9	Solar Radiation Observations	MQA	No

4.3 The Netherlands

The following figure provides a screenshot of the assessed level of metadata quality for the Dutch Open Data Portal using the MQA Analysis tool. This provides an overview of all metadata attributes that require attention generally across the entire portal.

The following are the most notable areas of improvement for the publication of metadata in the national portal:

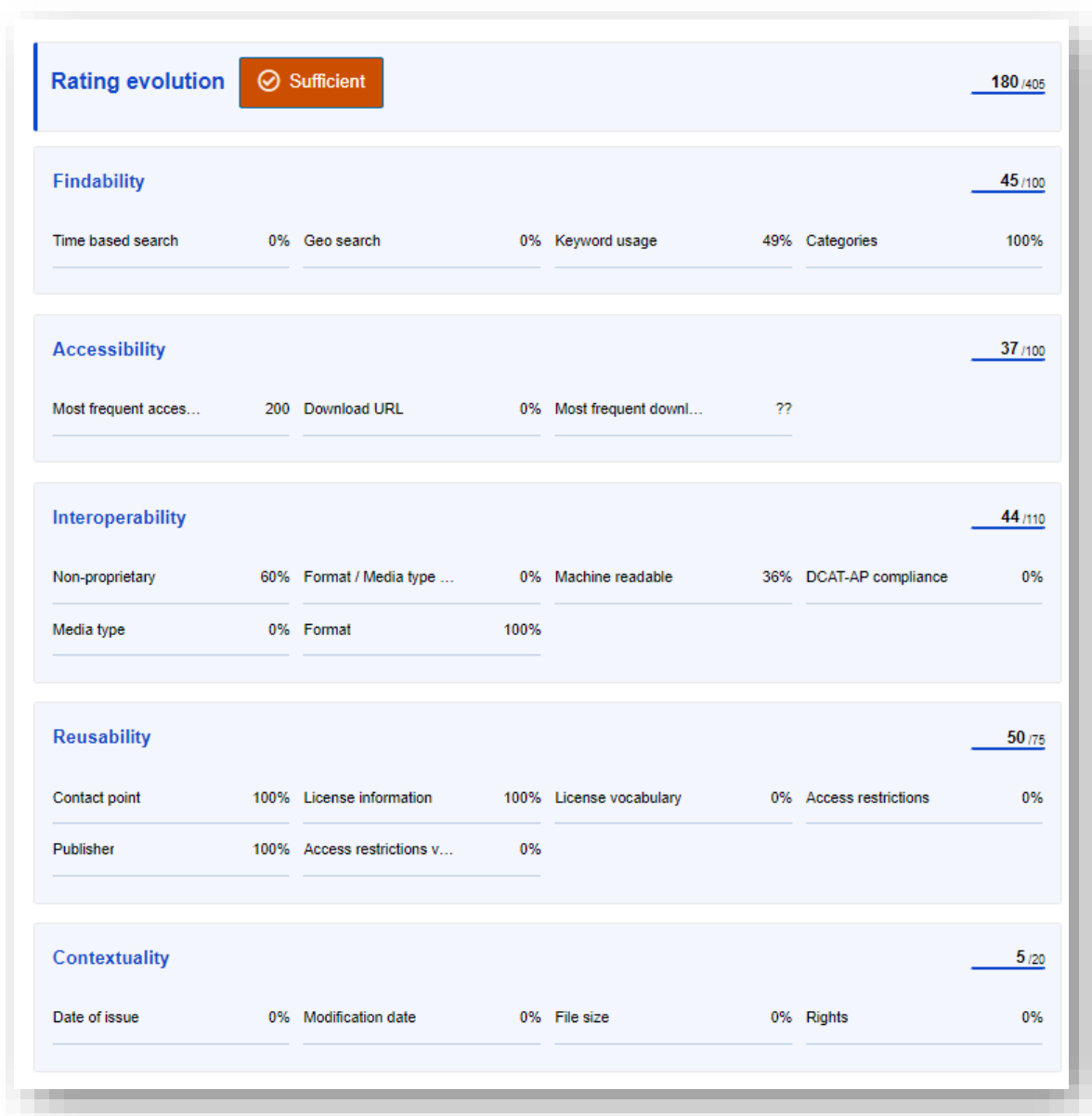
1. Compliance with DCAT-AP.
2. Ensure the inclusion of access restrictions (including the lack thereof) in metadata. Should include licensing information.
3. The machine readability of the data is measured by the MQA tool. This is likely to be based on the lack of dct:format use in the metadata. With the goal of improving the MQA score, this should be universally implemented for metadata in the portal.
4. Investigate on which factors geosearch is supported.

Not all errors are the result of a lack of information in the national portals. Metadata stewards should take care to assess the cause of the lack of information in the EDP based on the SHACL violations of each metadata record made available here:

<https://data.europa.eu/mqa/catalogues/dataportaal-van-de-nederlandse-overheid/violations?locale=en>

Note here that although DCAT-AP compliance for the Netherlands is noted as 0%, no DCAT-AP compliance violations are noted by the MQA Analysis tool (see link). This should be rectified by the EDP to ensure correct reporting.

Figure 3. Results of MQA Analysis – Data Portal for the Dutch Government (data.overheid.nl)



DCAT-AP Compliance for Open Data Portal Netherlands

In the above figure, DCAT-AP is only one of the metrics by which metadata quality is measured. As one of the key deliverables for this project, however, the lack of compliance seen here needs to be assessed more closely. The following table provides the number of metadata records currently published in the national open data portal and the percentage of these that are compliant with DCAT-AP according to the MQA tool.

Table 10. DCAT-AP Compliance for All Metadata Records Published by Open Data Portal Netherlands

No.	Description	Results
1	Number of metadata records available	22,539
2	Percentage of DCAT-AP Compliance	0%

Link: <https://data.europa.eu/api/mqa/reporter/report/en/pdf?catalogueId=dataportaal-van-de-nederlandse-overheid>

The general lack of DCAT-AP compliance is likely due to the standards in which metadata is published in the national portal. Progress made by the project to improve this will be discussed in the concluding chapter of this document. The following assesses whether any of the metadata of relevance to this project are compliant with DCAT-AP.

DCAT-AP Compliance per Dutch Dataset used in Use Cases

Each of the metadata records were assessed using the MQA tool in turn. The following table denotes whether or not a particular record was assessed to be compliant or not. To see the full analysis per record, please visit the link in the table.

Table 11. Analysis Metadata Quality

No.	Dataset Name	Link to MQA Result	DCAT-AP Compliance?
1	BRT Background Map	MQA	No
2	Basisregistratie Adressen en Gebouwen (BAG)	MQA	No
3	DTM	MQA	No
4	DSM	MQA	No
5	Daily Mean Temperature	MQA	No

4.4 Norway

The following figure provides a screenshot of the assessed level of metadata quality for the Norwegian Geoportal Portal using the MQA Analysis tool. This provides an overview of all metadata attributes that require attention generally across the entire portal.

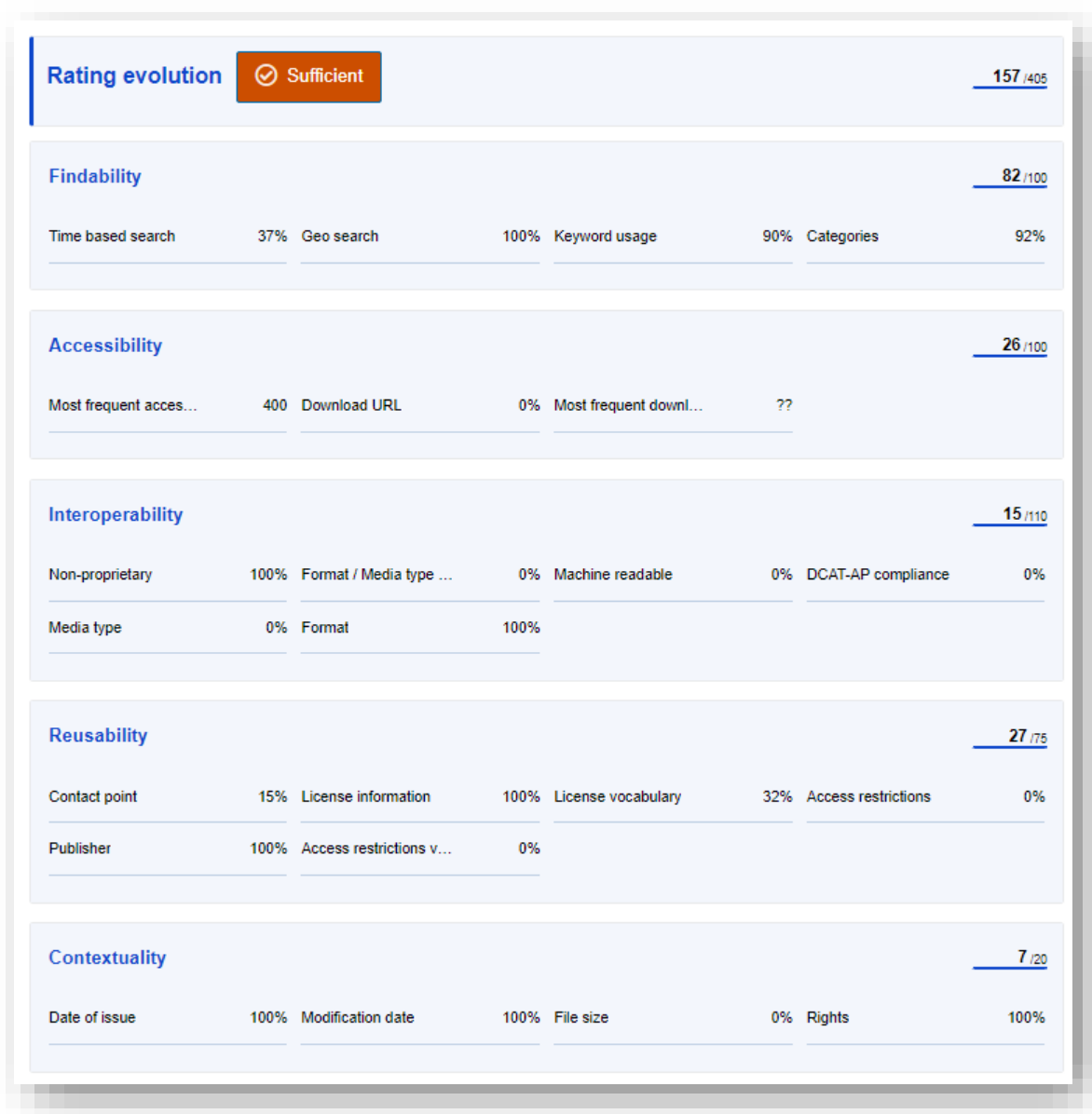
The following are the most notable areas of improvement for the publication of metadata in the national portal:

1. Ensure the inclusion (and validation) of temporal and modification information in metadata records.
2. Ensure the inclusion of access restrictions (including the lack thereof) in metadata. Should include licensing information.

Not all errors are the result of a lack of information in the national portals. Metadata stewards should take care to assess the cause of the lack of information in the EDP based on the SHACL violations of each metadata record made available here:

<https://data.europa.eu/mqa/catalogues/geonorge/violations?locale=en>

Figure 4. Results of MQA Analysis – National Geoportal of Norway



DCAT-AP Compliance for Open Data Portal Norway

In the above figure, DCAT-AP is only one of the metrics by which metadata quality is measured. As one of the key deliverables for this project, however, the lack of compliance seen here needs to be assessed more closely. The following table provides the number of metadata records currently published in the national geo portal and the percentage of these that are compliant with DCAT-AP according to the MQA tool.

Table 12. DCAT-AP Compliance for All Metadata Records Published by Open Data Portal Norway

No.	Description	Results
1	Number of metadata records available	384
2	Percentage of DCAT-AP compliance	0%

Link: <https://data.europa.eu/api/mqa/reporter/report/en/pdf?catalogueId=geonorge>

The general lack of DCAT-AP compliance is likely due to the standards in which metadata is published in the national portal. Progress made by the project to improve this will be discussed in the concluding chapter of this document. The following assesses whether any of the metadata of relevance to this project are compliant with DCAT-AP.

DCAT-AP Compliance per Norwegian Dataset used in Use Cases

Each of the metadata records were assessed using the MQA tool in turn. The following table denotes whether or not a particular record was assessed to be compliant or not. To see the full analysis per record, please visit the link in the table.

Table 13. Analysis Metadata Quality Norway

No.	Dataset Name	Link to MQA Result	DCAT-AP Compliance?
1	Background Map	MQA	No
2	INSPIRE Buildings WFS	MQA	No
3	Roads	MQA	No
4	National Terrain Model	MQA	No
5	National Surface Model	MQA	No
6	Frost Climate API	MQA	No

4.5 Estonia

The following figure provides a screenshot of the assessed level of metadata quality for the Estonian Geoportal Portal using the MQA Analysis tool. This provides an overview of all metadata attributes that require attention generally across the entire portal.

The following are the most notable areas of improvement for the publication of metadata in the national portal:

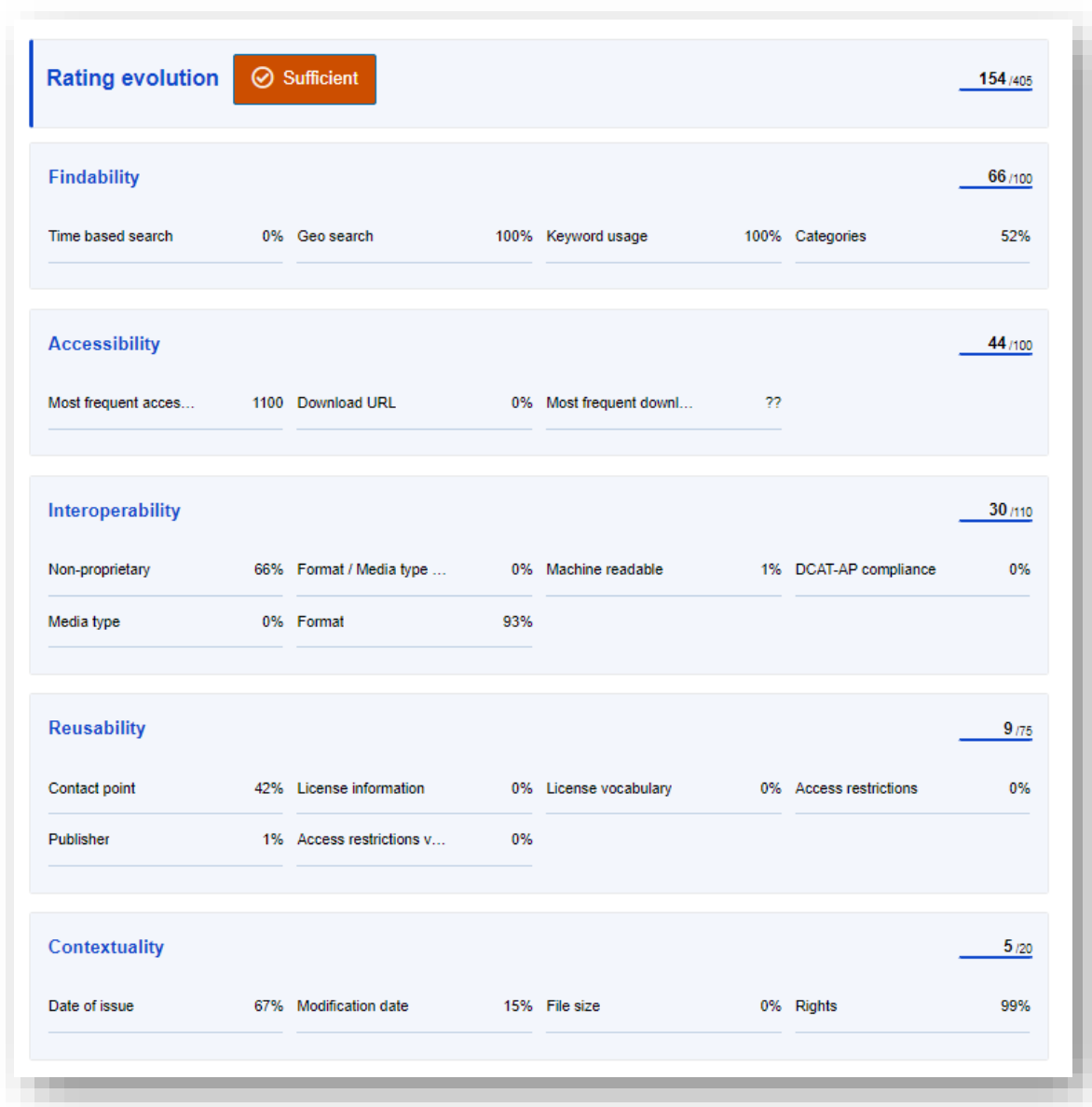
1. License and accessibility information
2. Ensure the inclusion (and validation) of access restrictions (including the lack thereof) in metadata.
3. The machine readability of the data is measured by the MQA tool. This is likely to be based on the lack of dct:format use in the metadata. With the goal of improving the MQA score, this should be universally implemented for metadata in the portal.

Not all errors are the result of a lack of information in the national portals. Metadata stewards should take care to assess the cause of the lack of information in the EDP based on the SHACL violations of each metadata record made available here:

<https://data.europa.eu/mqa/catalogues/ng-ee/violations?locale=en>

Note here that although DCAT-AP compliance for Estonia is noted as 0%, no DCAT-AP compliance violations are noted by the MQA Analysis tool (see link). This should be rectified by the EDP to ensure correct reporting.

Figure 5. Results of MQA Analysis –Geoportal Portal Estonia



DCAT-AP Compliance for Open Data Portal Estonia

In the above figure, DCAT-AP is only one of the metrics by which metadata quality is measured. As one of the key deliverables for this project, however, the lack of compliance seen here needs to be assessed more closely. The following table provides the number of metadata records currently published in the national geo portal and the percentage of these that are compliant with DCAT-AP according to the MQA tool.

Table 14. DCAT-AP Compliance for All Metadata Records Published by Open Data Portal Norway

No.	Description	Results
1	Number of metadata records available	308
2	Percentage DCAT-AP Compliance	0%

Link: <https://data.europa.eu/api/mqa/reporter/report/en/pdf?catalogueId=ng-ee>

The general lack of DCAT-AP compliance is likely due to the standards in which metadata is published in the national portal. Progress made by the project to improve this will be discussed in the concluding chapter of this document. The following assesses whether any of the metadata of relevance to this project are compliant with DCAT-AP.

DCAT-AP Compliance per Estonian Dataset used in Use Cases

Each of the metadata records were assessed using the MQA tool in turn. The following table denotes whether or not a particular record was assessed to be compliant or not. To see the full analysis per record, please visit the link in the table.

Table 15. Analysis Metadata Quality

No.	Dataset Name	Link to MQA Result	DCAT-AP Compliance?
1	Estonian Topography Database	MQA	No
2	Main map of Estonia 1:10000	MQA	No
3	Orthophotos Estonia	MQA	No
4	Shaded Relief Areas	MQA	No
5	Contours and Elevation Map Estonia	MQA	No
6	Estonian Topography Buildings	MQA	No
7	Estonian Road Register	MQA	No
8	Estonian Topography Transport	MQA	No
9	DTM	MQA	No
10	DSM	MQA	No
11	Weather and Climate Data	MQA	No

5 DCAT-AP Compliance Analysis

The level of DCAT-AP compliance currently present across all partner countries is relatively low despite efforts made throughout the project by all parties to improve this. To understand why this lack of compliance remains an issue, various metadata records and their respective SHACL violations were investigated. The following steps were taken. For the purpose of this example, one metadata record from Spain will be reference but the analysis holds across all partner countries.

Investigation Steps

1. For each of the metadata records provided in the table in Chapter 3, the RDF/XML format was downloaded from the EDP website.
Eg. https://data.europa.eu/data/datasets/spaign_igr_transporte?locale=en
2. This metadata record was then uploaded to the MQA Validation Tool following this link: <https://data.europa.eu/mqa/shacl-validator-ui/data-provision>
3. Once the metadata record was uploaded, the metadata is validated by clicking on the validate button. If the metadata is valid, the right hand panel would show a green box indicating whether the metadata record was valid and if not, SHACL violations would be returned.
4. Where SHACL violations existed, the source of each violation was investigated by a metadata expert from each partner country.

In the case of the above record, a number of violations were returned. Upon investigation, the following analysis and conclusions were made.

Issue 1

Issue in property <http://www.w3.org/ns/locn#geometry> of node [.b7](#) **maxCount[1]: Invalid cardinality: expected max 1: Got count = 2**

XML INSPIRE: Cardinality is max 1

```
<gmd:extent>
  <gmd:EX_Extent>
    <gmd:description gco:nilReason="missing"/>
    <gmd:geographicElement>
      <gmd:EX_GeographicBoundingBox>
        <gmd:westBoundLongitude>
          <gco:Decimal>-19.00</gco:Decimal>
        </gmd:westBoundLongitude>
        <gmd:eastBoundLongitude>
          <gco:Decimal>4.30</gco:Decimal>
        </gmd:eastBoundLongitude>
        <gmd:southBoundLatitude>
          <gco:Decimal>27.60</gco:Decimal>
        </gmd:southBoundLatitude>
        <gmd:northBoundLatitude>
          <gco:Decimal>44.60</gco:Decimal>
        </gmd:northBoundLatitude>
      </gmd:EX_GeographicBoundingBox>
    </gmd:geographicElement>
  </gmd:EX_Extent>
</gmd:extent>
```

RDF DCAT-AP 2.1.1 transforms the geometry in two labels, then the cardinality is 2, one is with the gml literal and the other with the wkt literal.

```
<det:spatial>
  <det:Location>
    <det:geometry rdf:type="http://replacement.io/ont/geoparq#wktLiteral">
      POLYGON((-19.00 44.60 44.60 44.60 44.60 44.60))
    </det:geometry>
    <det:geometry rdf:type="http://replacement.io/ont/geoparq#gmlLiteral">
      <gml:Envelope srsName="http://www.opengis.net/def/EPSSG/0/4326"><gml:lowerCorner>-19.00</gml:lowerCorner><gml:upperCorner>44.60 44.60</gml:upperCorner></gml:Envelope>
    </det:geometry>
  </det:Location>
</det:spatial>
```

This error has been checked with other INSPIRE metadata from other countries and the result is the same. And it is an arbitrary mistake. Two identical XML files., one gives this error ([example 1](#)) and other file does not ([example 2](#)).

Conclusion:

This error cannot be corrected in XML INSPIRE, it must be corrected in the XML to RDF transformation. This error is arbitrarily thrown for the same geometry definition in XML, there are files in XML INSPIRE that do not have this error.

Issue 2

Issue in property <http://purl.org/dc/terms/license> of node <http://data.europa.eu/88u/distribution/3aedcb3b-e116-4aa3-9251-5f58a49c5911>

maxCount[1]: Invalid cardinality: expected max 1: Got count = 2

Conclusion:

This case has been corrected in the INSPIRE XML files. This is not a bug in INSPIRE XML, as INSPIRE allows the "otherRestrictions" metadata element to have a cardinality of 1 to many, but it has been corrected in the metadata files and only one element "otherRestrictions" has been left in the metadata files.

Issue 3

Issue in property <http://purl.org/dc/terms/rights> of node <http://data.europa.eu/88u/distribution/60e57033-0625-4214-9df4-0326c64e24fe>

maxCount[1]: Invalid cardinality: expected max 1: Got count = 2

All XML files are identical and this error is arbitrary. It does not occur in all DCAT-AP files, so we believe it is a bug in the transformation between XML and DCAT-AP. Moreover, in XML the cardinality of the constraints is one, but it has metadata sub-elements and in DCAT-AP it shows it as two rights.

XML INSPIRE, pag 51 of [Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007](#)

2.3.6. Limitations on public access

The element for the limitations on public access is described in [Regulation 1205/2008], Part B, 8.2:

8.2. Limitations on public access

When Member States limit public access to spatial data sets and spatial data services under Article 13 of Directive 2007/2/EC, this metadata element shall provide information on the limitations and the reasons for them.

If there are no limitations on public access, this metadata element shall indicate that fact.

The value domain of this metadata element is free text.

```
<gmd:resourceConstraints>
  <gmd:MD_LegalConstraints>
    <gmd:accessConstraints>
      <gmd:MD_RestrictionCode codeList="http://standards.iso.org/iso/19139/resources/gmx/Codelist.xml#MD_RestrictionCode"
      </gmd:accessConstraints>
      <gmd:otherConstraints>
        <gmx:Anchor xlink:href="http://inspire.ec.europa.eu/metadata-codelist/LimitationsOnPublicAccess/noLimitations" Sin
        </gmd:MD_LegalConstraints>
      </gmd:resourceConstraints>
    <gmd:resourceConstraints>
      <gmd:MD_LegalConstraints>
        <gmd:useConstraints>
          <gmd:MD_RestrictionCode codeList="http://standards.iso.org/iso/19139/resources/gmx/Codelist.xml#MD_RestrictionCode"
          </gmd:useConstraints>
          <gmd:otherConstraints>
            <gmx:Anchor xlink:href="https://creativecommons.org/licenses/by/4.0/">CC BY 4.0</gmx:Anchor>
            </gmd:otherConstraints>
          </gmd:MD_LegalConstraints>
        </gmd:resourceConstraints>
```

RDF DCAT-AP 2.1.1 transforms the constraints in two labels, then the cardinality of rights is 2.

```
<dcct:title xml:lang="fi-t-es-t0-mtec">CNIG latauskeskus</dcct:title>
<dcct:rights rdf:resource="http://inspire.ec.europa.eu/metadata-codelist/LimitationsOnPublicAccess/noLimitations" />
<?xml:characterEncoding=UTF-8?>
```

```
<dcct:rights>
  <dcct:RightsStatement>
    <rdf:label>otherRestrictions</rdf:label>
  </dcct:RightsStatement>
</dcct:rights>
```

You can see that in this case the result is incorrect, it should not be "OtherRestrictions", instead it should be CC BY 4.0 or "OtherRestrictions like CC BY 4.0". In this case, you can see that the transformation from XML to DCAT-AP is incorrect. This error only happens in some XML files, when all XML INSPIRE files have the same license (structure, content, etc.)

Conclusion:

This error cannot be corrected in XML INSPIRE, it must be corrected in the XML to RDF transformation. DCAT-AP must comply with "Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007".

Issue in
property <http://spdx.org/rdf/terms#algorithm> and <http://spdx.org/rdf/terms#checksumValue>
of node [.b5](#)

HasValueConstraint[http://spdx.org/rdf/terms#checksumAlgorithm_sha1] : No
value http://spdx.org/rdf/terms#checksumAlgorithm_sha1 in
[http://spdx.org/rdf/terms#checksumAlgorithm_md5]

There isn't this property on XML INSPIRE.

RDF DCAT-AP 2.1.1

```
<spdx:checksum>
  <spdx:Checksum>
    <spdx:algorithm rdf:resource="http://spdx.org/rdf/terms#checksumAlgorithm_md5"/>
    <spdx:checksumValue>2480af50db37b1e9b292d5747d13bc8a</spdx:checksumValue>
  </spdx:Checksum>
</spdx:checksum>
```

Conclusion:

This error cannot be corrected in XML INSPIRE, this property is a value that allows to check the integrity of the contents of a file. We assume that if the RDF file has no errors, this error will not be raised.

The following summary is made of the above analysis:

- <http://purl.org/dc/terms/license>
The XML has been modified so that this error does not appear. But INSPIRE allows you to specify the type of licence.
- <http://www.w3.org/ns/locn#geometry>
It is an error introduced during harvesting by EDP. There are some RDF files that have this error and other files do not, and the XML INSPIRE is the same in all cases.
- <http://purl.org/dc/terms/rights>
It is an error introduced during harvesting by EDP. When there are no limitations on the restrictions, in INSPIRE is mandatory to specify the conditions.
- <http://spdx.org/rdf/terms#algorithm> and <http://spdx.org/rdf/terms#checksumValue>
This error disappears when the RDF file is free of errors.

There are two takeaway points from the above analysis. Firstly, the SHACL violations are useful when indicating errors in the national portals from which the metadata is being harvested. In the case of the above example, the licensing information was adjusted and is now improved in the EDP. Secondly, there do exist errors that are introduced during the harvesting process. From what this investigation indicates, this generally occurs when information is being harvested from geoportals. However, open data portals do also have limited DCAT-AP compliance at present, leading to the following action points for partner countries.

1. For all errors which are present in the SHACL violations and can be addressed by the partner countries in the national portals, these will be addressed in the coming quarter of this year.

2. In the case of the Netherlands, there is an ongoing national initiative to ensure metadata is published in the open data portal in DCAT-AP format in order to reduce the potential for harvesting errors. This will be completed at the end of the second quarter of this year and the DCAT-AP compliance will be tested again.

In addition to these action points, contact has already been made with the EDP in order to solve the issues introduced during the harvesting process. This is an ongoing process and partner countries will continue to work the EDP on the improvement of their metadata beyond this project.

6 Concluding Remarks

From the beginning of the GeoE3 project, there has been good cooperation with the EDP and intensive work has been carried out to improve the metadata available through the EDP. The metadata has seen an improvement over the course of the project. Indeed, most of the partner countries have achieved a sufficient level of quality in their national portals and the EDP reporting supports ongoing improvements to these metrics.

Based on the above analysis, the most significant room for improvement remains the level of DCAT-AP compliance that both national portals in general and relevant metadata records for this project achieve. Partner countries will continue to work closely with colleagues at the EDP to solve the existing issues together and ensure that metadata compliance continues to improve beyond the scope of this project.