Discovery testing

Milestone 4

Implementation Guidelines for National Portals and Geo-platforms

Co-financed by the Connecting Europe Facility of the European Union

02-05-2023

Revision History

Version	Status	Author	Modifications
		Add lines as needed	
0.1	Updated	Lexi Rowland	
0.9	Updated	Laura Alemany Gomez	

Distribution

Version	Status	Recipient
0.1	Draft	Antti Jakobsson, Dorus Kruse, Laura Alemany Gomez
0.9	Draft	PMG GeoE3
0,95	Draft	PMG GeoE3
1.0	Final	

Table of Contents

Revision History 2

Distribution 2

Table of Contents 3

1. Introduction 5

2. Background 6

1.1 2.1 Metadata Profiles Relevant for the European Context 6

2.1.1 INSPIRE Metadata Profile 6

2.1.2 DCAT(-AP) and GeoDCAT(-AP) 2.0 Metadata Profiles 6

2.1.3 Schema.org Metadata Profile 7

2.1.4 Mapping between INSPIRE Metadata Profiles and (Geo)DCAT 2.0 7

1.2 2.2 The European Data Portal (EDP) 7

2.2.1 National Open Data Portals 8

3. Analysis of Connection between EDP and the National Open/Spatial Data Portals 10

1.3 3.1 Spain 11

3.1.1 Illustrative Example: Dataset from the Spanish National Open Data Portal 11

3.1.2. Complete Analysis Spanish Datasets 15

1.4 3.2 Finland 17

3.2.1 Illustrative Example: Dataset from the Finnish National Open Data Portal 18

1.5 2.3 The Netherlands 22

3.3.1 Illustrative Example: Dataset from the Dutch Spatial Data Infrastructure 22

1.6 3.4 Norway 27

3.4.1 Illustrative Example: Dataset from the Norwegian Spatial Data Infrastructure 27

2.5 Estonia 32

3.4.1 Illustrative Example: Dataset from the Estonian Spatial Data Infrastructure Portal 33

1.7 3.6 Cross-Project Analysis 40

3.6.1 Title and abstract 40

3.6.2 Distribution 41

3.6.3 Spatial Extent 42

3.6.4 Identifier 42

3.6.5 Conclusions and Recommendations 44

3.7.1 Metadata from Open Data Portals 44

3.7.2 Metadata from Spatial Data Infrastructure 44

1.8 3.7 Cross-Partner Analysis 45

1.9 3.8 Interoperability Map 46

3.8.1 Spain Interoperability Map 47
3.8.2 Finland Interoperability Map 48
3.8.3 Netherlands Interoperability Map 49
3.8.1 Norway Interoperability Map 50
3.8.5 Estonia Interoperability Map 52

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.

This report provides detailed information on the investigation carried out to assess which metadata was being harvested by the EDP and whether the resulting metadata was being correctly represented in the EDP. The following sections provide detail on the process carried out and the results of this investigation. In the section ‘Analysis of EDP Connection, there is an overview of the datasets from 5 countries that are used in the project, the process carried out to assess the quality as well as the findings of this analysis. Following this, the report includes an extended analysis of the results, including comments on the progress of improving semantic interoperability across project partners. The final chapter summarises the findings.

2. Background

In this section, an overview of all standards and portals related to metadata, at both a national and European level, relevant for all project partners are discussed.

1.1 2.1 Metadata Profiles Relevant for the European Context

2.1.1 INSPIRE Metadata Profile

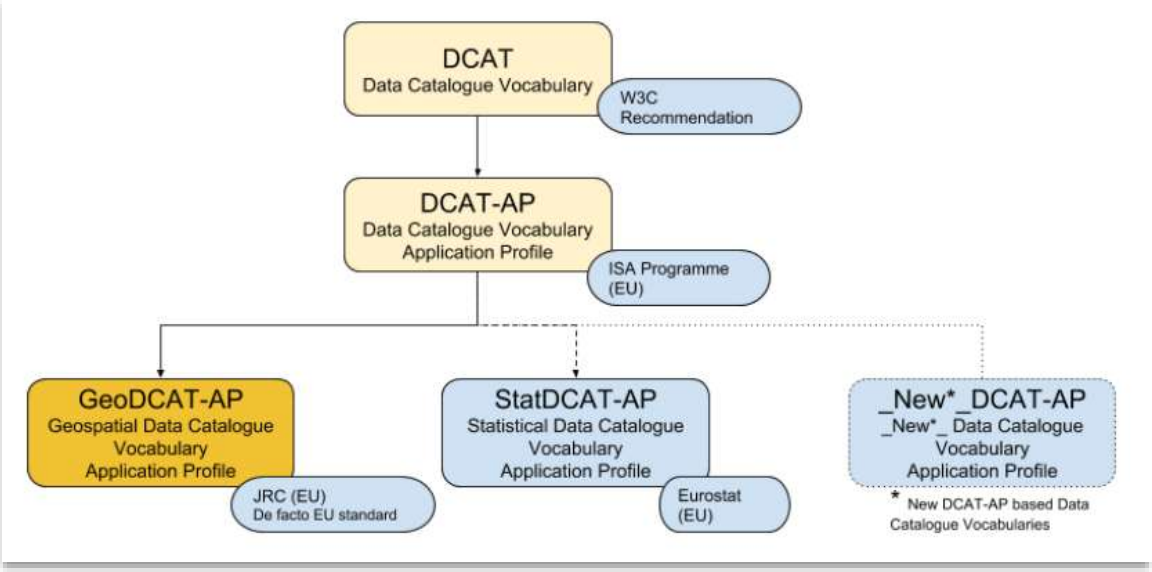
According to Article 5(1) of the INSPIRE Directive 2007/2/EC, all member states are required to create and maintain metadata for national geospatial datasets corresponding to a particular metadata profile. This profile is based on ISO 19115, 19119 and 19139 and extended slightly ensuring that being in compliance with the aforementioned standards does not necessarily ensure that metadata is compliant with an INSPIRE profile. As such, care should be taken to understand what compliance requirements are in place for INSPIRE.

Note: Not all the data and associated metadata using in the GeoE3 project is INSPIRE data.

2.1.2 DCAT(-AP) and GeoDCAT(-AP) 2.0 Metadata Profiles

The DCAT metadata profile is an RDF vocabulary that is primarily designed to better facilitate interoperability of data catalogues on the web, improve the discoverability of datasets within said catalogues, improve the harvesting of one catalogue by another and make federated searches across catalogues possible. DCAT-AP is the de facto standard for the EU, i.e., the guiding profile for application in the EU, and, as such, most open data portals within the EU make use of this application profile.

Figure 1. Relation between DCAT initiatives.



Source: OGC

GeoDCAT and the accompanying application profile for the EU was developed to further support the description through metadata of the geospatial datasets. Included in this profile is an RDF syntax binding for the union of metadata elements defined in the core ISO profiles

and those defined by the INSPIRE Directive. The official release of DCAT-AP 2.0 for European data portals can be found here in various serialisations and here for GeoDCAT.

2.1.3 Schema.org Metadata Profile

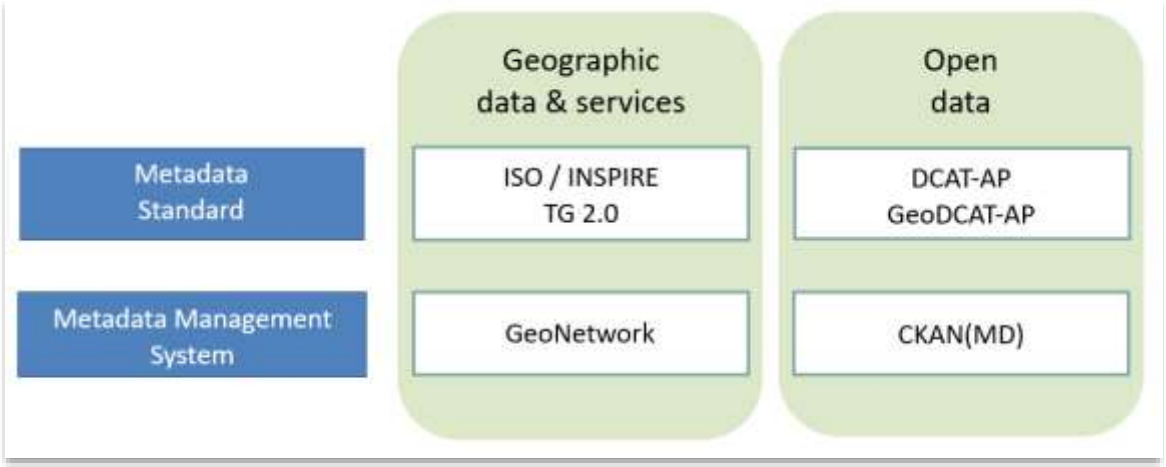
In the context of dataset description, schema.org is an extensible schema which is used by a maintainer of, for example, a data catalogue to support the findability of datasets on the web and, more specifically, in a search engine such as Google. In this context, the metadata describing a given dataset would be described using a schema.org profile and imbedding in the HTML page in the catalogue associated with said dataset. A full overview of schema.org classes and properties can be found here.

2.1.4 Mapping between INSPIRE Metadata Profiles and (Geo)DCAT 2.0

These metadata profiles have a lot of overlapping elements where one profile may be more extensive or have different mandatory fields than another metadata profile. Where it is necessary to publish metadata in various contexts using a range of metadata profiles, it might be useful to make use of a mapping between profiles in order to use the same information for the same dataset and simply present this information in compliance with a different profile. A large majority of the metadata published in a profile is reused across many profiles, making mapping necessary in order to reuse the same metadata in different contexts.

In the European context, many partner countries are already compliant with INSPIRE implementation rules and are published in a national catalogue using GeoNetwork. The European Data Portal (EDP) harvests this INSPIRE compliant metadata and automatically performs a mapping between INSPIRE and (Geo)DCAT so it is not strictly necessary for any national infrastructure to perform this mapping independently. Although an explicit mapping is not defined by the EDP, the SEMIC project does provide a tool which converts INSPIRE profiles to (Geo)DCAT. Should this be of interest, the tool can be found here and the schema here.

Figure 2. Relation between metadata standards.



1.2 2.2 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.

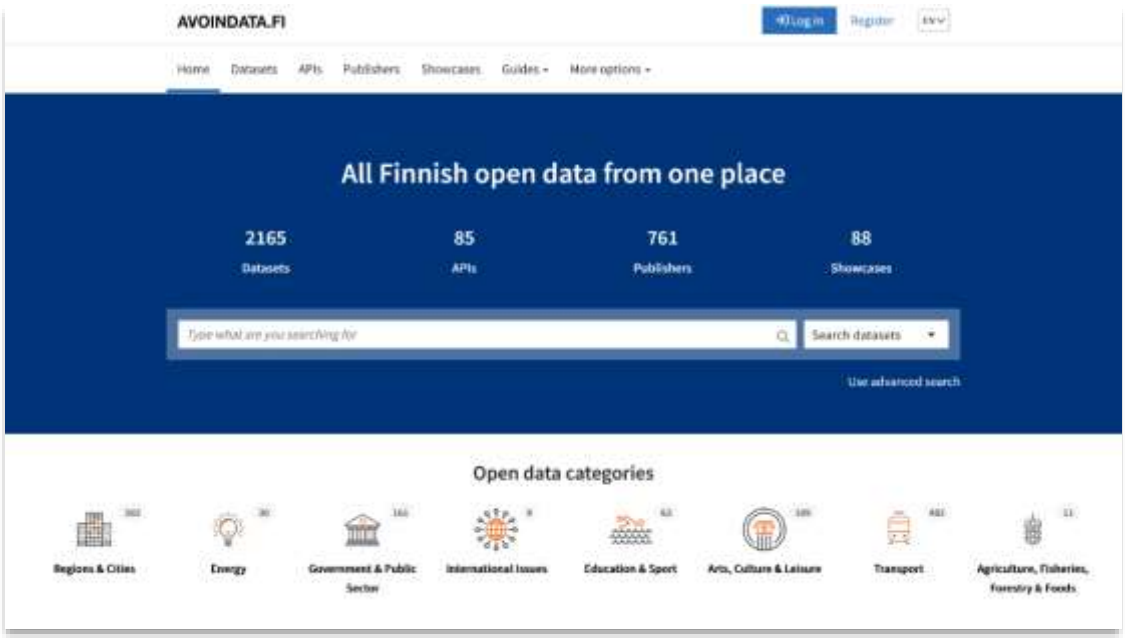
In order to foster comparability of data published across borders, the portal presents metadata references in a common metadata profile format, namely; DCAT which is a RDF vocabulary. The portal provides translations of dataset descriptions in all 24 languages using machine-translation technology.

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 GeoNetworl. 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

Geoportal[1]. 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.

[1] INSPIRE Geoportal

3. Analysis of Connection between EDP and the National Open/Spatial Data Portals

In this section, there is an overview of all datasets from 5 countries used in the use cases for the GeoE3 project. These countries include Estonia, Finland, Norway, Spain and the Netherlands. The following sections provide a detailed description of the results of the analysis for each country. In generating these results, each dataset defined for a particular project use case 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 check was performed for the following elements.

Title and abstract
Distributions
Spatial extent
identifiers
Keywords

The following sections include the results of this analysis per country. Where a quality issue or lack of completeness exists in the metadata, a comment is provided defining what issue was identified.

1.3 3.1 Spain

In order to provide an overview of the quality of the metadata across different publication outlets as well as differences in the quality of metadata following harvesting, a comparative analysis was carried out based on the elements defined above. The following sections provides an illustrative example of how such an analysis was carried out in the context of the Spanish infrastructure and then a summary table is provided highlighting the results of this analysis for all Spanish datasets included in the project.

3.1.1 Illustrative Example: Dataset from the Spanish National Open Data Portal

Dataset name: Id-29 – Digital Land Model (5 meters) for Spain

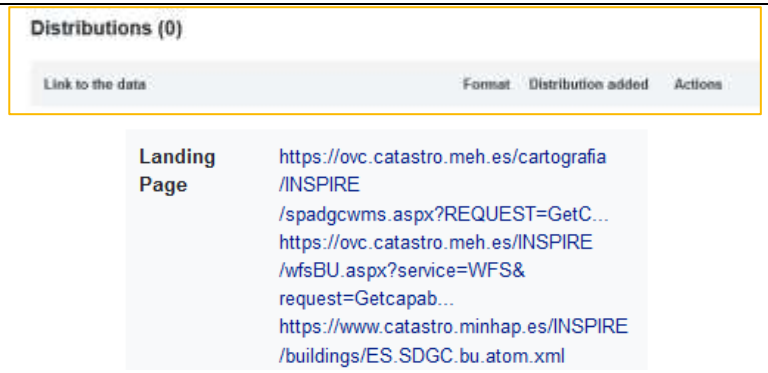
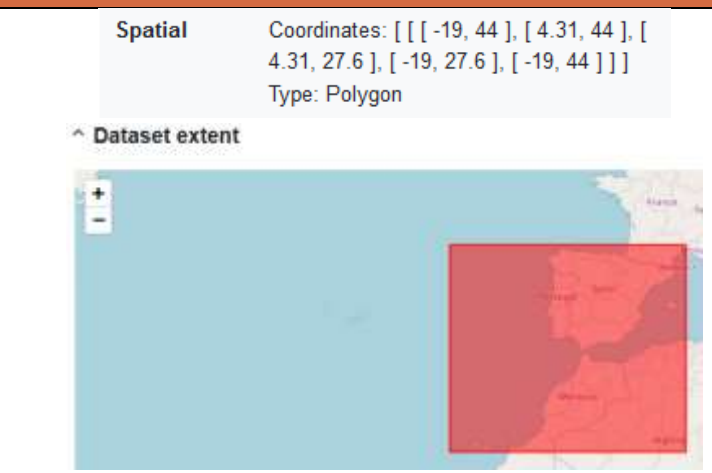
This dataset can be found in:

1. The National Open Data Portal for Spain (datos.gob.es)
   https://datos.gob.es/es/catalogo/e00125901-spaignmdt05
2. The European Data Portal
1. https://data.europa.eu/data/datasets/https-datos-gob-es-catalogo-e00125901-spaignmdt05?locale=en

The following table shows in the left column what can information can be found in the national open data portal and in the right column what can be found in the EDP. The final column provides a comment regarding whether the information is complete and, if not, what information is missing.

Table 1. Illustrative example of comparative analysis performed for Spanish datasets

No.	Description National Portal Element	Description EDP Element	Comment
1	Title and Abstract		
	Buildings of DG Cadastre. Spain Geographical location of buildings of Spain. According to Regulation (UE) Nº 1089/2010, a building is defined as all construction, above and below ground, for the purpose of accommodating people, animals and things, or production and distribution of goods or services and are permanent structures on the ground. Therefore, the building geometry model INSPIRE obtained from data of General Directorate for Cadastre is defined as: A multi-enclosure that represents the envelope of all buildings with volumes above ground level of each cadastral parcel, excluding overhangs and terraces or balconies. The Buildings INSPIRE dataset contains the official data of the DGC transformed directly to the data model defined by the INSPIRE directive, the dataset is complete including urban and rural areas for 95% of the territory under the responsibility of the DGC, (except the Basque country and Navarre). More information: http://www.catastro.meh.es/webinspire/documentos/Conjuntos%20de%20datos_en.pdf. According to the specifications of INSPIRE for the theme of buildings INSPIRE Data Specification on Buildings – Technical Guidelines. (https://inspire.ec.europa.eu/Themes/126/2892)		No comments. The same information is reflected in both.
2	Distributions		

			Apparently, this information is in the 'Landing page' in EDP instead of in the 'distributions' section
3	Spatial Extent		
			No comments. The same information is reflected in both.

4	Identifiers		
		Identifiers SDGC/BU_BuildingPart SDGC/BU SDGC/BU_Building	No comments. The same information is reflected in both.
5	Keywords		

			No comments. The same information is reflected in both.

3.1.2. Complete Analysis Spanish Datasets

The following provides an overview of the analysis performed for all Spanish datasets across the different metadata catalogues. Overall, the table highlights some significant areas of improvement with regards 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. There is not always a translation of the tile and abstract to other required languages.
3. Keywords are often harvested correctly but associated to a thesaurus rather than the keyword itself.
4. It is also common for the EDP to completely lack a geospatial extent in the resulting metadata despite this existing in the national portal metadata.

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

No.	Dataset Name	Dataset Type	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords

1	Buildings	3D building model	ok	harvested as landing page not as a distribution	ok	ok	ok except the keywords associated to a thesaurus
2	Buildings	Relevant building attributes	ok	harvested as landing page not as a distribution	ok	ok	ok except the keywords associated to a thesaurus
3	Digital Surface Model (DSM) (5 metres)	Building footprints	ok	harvested as landing page not as a distribution	ok	ok	ok except the keywords associated to a thesaurus
4	Modelo Digital de Superficies Edificación - MDSn2,5	Digital Surface Model (DSM)	ok. But EDP doesn't translate them into other languages	harvested in the "Documentation" element of EDP	no	ok	ok except the keywords associated to a thesaurus
5	Modelo Digital de Superficies normalizado de Vegetación - Primera Cobertura (2,5 metros)	Digital Surface Model Normalized (DSMn)	ok. But EDP doesn't translate them into other languages	harvested in the "Documentation" element of EDP	no	ok	ok except the keywords associated to a thesaurus
6	Digital Terrain Model of Spain (5 metres)	Digital Surface Model Normalized (DSMn)	ok. But EDP doesn't translate them into other languages	harvested in the "Documentation" element of EDP	no	ok	ok except the keywords associated to a thesaurus

7	LiDAR map of Spain (2,5 metres)	Digital Terrain Model (DTM)	ok	ok	Missing in the open data portal	no identifier in the national open data portal	ok
8	Petrol Stations of Spain	Digital Surface Model (DSM) of LiDAR	ok. But EDP doesn't translate them into other languages	harvested in the "Documentation" element of EDP	no	ok	ok except the keywords associated to a thesaurus
9	Spanish Transport Network	Petrol Stations (kml)	ok	ok	Missing in the open data portal	no identifier in the national open data portal	ok
10	Geographical Information of the Population	Transport Network (Vector)	ok. But EDP doesn't translate them into other languages	harvested in the "Documentation" element of EDP	harvested in the "Spatial" element of EDP	ok	ok except the keywords associated to a thesaurus
11	Spanish Urban Information System	Existing population data (Vector)	ok	no	ok	ok	ok except the keywords associated to a thesaurus
12	Land Occupation Information System in Spain (SIOSE), 2014	Urban planning	ok	ok	Missing in the open data portal	no identifier in the national open data portal	ok

1.4 3.2 Finland

In order to provide an overview of the quality of the metadata across different publication outlets as well as differences in the quality of metadata following harvesting, a comparative analysis was carried out based on the elements defined above. The following sections provides an illustrative example of how such an analysis was carried out in the context of the Finnish infrastructure and then a summary table is provided highlighting the results of this analysis for all Finnish datasets included in the project.

3.2.1 Illustrative Example: Dataset from the Finnish National Open Data Portal

Dataset name: Id-8 – Amount of Solar Radiant Energy on the Roofs in the Helsinki Metropolitan Area

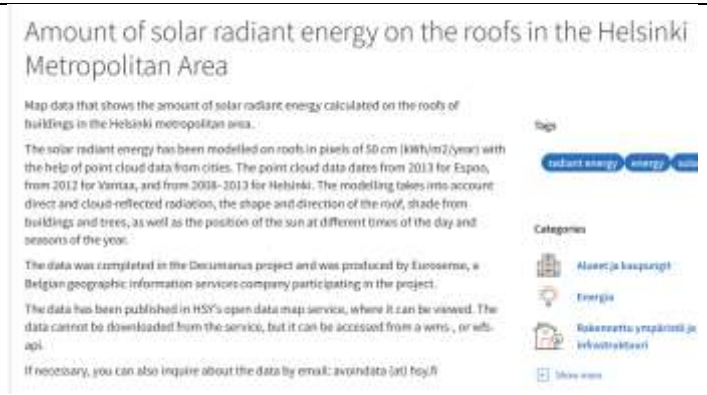
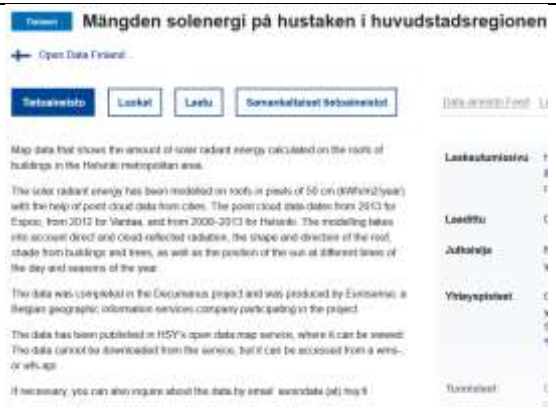
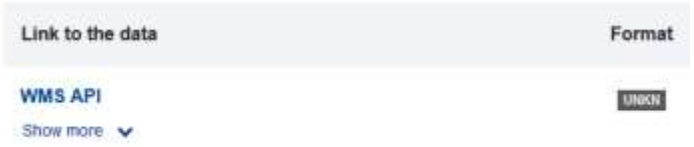
This dataset can be found in:

1. The National Open Data Portal for Finland
   https://www.avoindata.fi/data/fi/dataset/auringon-sateilyenergian-maara-katoilla-paakaupunkiseudulla
2. The European Data Portal
2. https://data.europa.eu/data/datasets/0182ded1-782d-4e47-801c-c8a736ea69b4?locale=fi

The following table shows in the left column what can information can be found in the national open data portal and in the right column what can be found in the EDP. The final column provides a comment regarding whether the information is complete and, if not, what information is missing.

Table 3. Illustrative example of comparative analysis performed for Finnish datasets

No.	Description National Portal Element	Description EDP Element	Comment
1	Title and Abstract		

			The title is not well translated into English by the EDP.
2	Distributions		
			No comments. The same information is reflected in both.
3	Spatial Extent		
			There was no information for this in the national data portal and, as such, nothing was recorded in the EDP.
4	Identifiers		

			There was no information for this in the national data portal and, as such, nothing was recorded in the EDP.
5	Keywords		
			No comments. The same information is reflected in both.

The following provides an overview of the analysis performed. Overall, the table highlights some significant areas of improvement with regards to the following:

1. Several datasets present in the national data portal is not reflected in the EDP and thus, there is a missing EDP URLs.
2. Several datasets do not include the spatial extent in the national open data portal and, therefore, this is not being correctly harvested by the EDP.
3. Several datasets do not include the identifier and/or keywords for the dataset in the national data portal and, therefore, this is not being correctly harvested by the EDP.
4. 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.

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

No.	Dataset Name	Dataset Type	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords

1	Pääkaupunkiseudun aurinkosähköpotentiaali	Solar energy potential	no	ok	Missing in the open data portal	ok	ok
2	Auringon säteilyenergian määrä katoilla pääkaupunkiseudulla	Solar energy potential	ok	ok	Missing in the open data portal	no identifier in the national open data portal	no keywords in the open data portal
3	Helsingin 3D-kaupunkimallit	3D building model LoD2	ok	ok	Missing in the open data portal	no identifier in the national open data portal	no keywords in the open data portal
4	Kuopion 3D-rakennusmalli	3D building model LoD2	ok	ok	Missing in the open data portal	no identifier in the national open data portal	no keywords in the open data portal
5	Espoon 3D-kaupunkimalli	3D building model LoD2	no	ok	Missing in the open data portal	no identifier in the national open data portal	no keywords in the open data portal
6	Kansallisen maastotietokannan rakennukset	Relevant building attributes and footprints	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing
7	Korkeusmalli 2m	Digital elevation model (DEM)	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing
8	Rinnevarjostus	Shadow index coverage	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing

9	Auringon säteilyhavainnot	Number of sunshine hours at the nearest observation station	ok	harvested as landing page not as a distribution	ok	ok	no keywords in the open data portal
10	Tuuliatlas 250m	Average wind conditions	ok	harvested as landing page not as a distribution	ok	ok	no
11	Säähavaintojen kuukausiarvot	Normal air temperature at the nearest observation station	ok	harvested as landing page not as a distribution	ok	ok	no
12	Ilmastonmuutosskenaariot	Monthly mean temperature based on climate scenarios	ok	harvested as landing page not as a distribution	ok	ok	no
13	Tilastokeskuksen aineistot	Existing population data	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing	URL EDP is missing

1.5 2.3 The Netherlands

In order to provide an overview of the quality of the metadata across different publication outlets as well as differences in the quality of metadata following harvesting, a comparative analysis was carried out based on the elements defined above. The following sections provides an illustrative example of how such an analysis was carried out in the context of the Dutch infrastructure and then a summary table is provided highlighting the results of this analysis for all Dutch datasets included in the project.

3.3.1 Illustrative Example: Dataset from the Dutch Spatial Data Infrastructure

Dataset name: Id-3 – Acad High Test File Nederland 3 (AHN3)

This dataset can be found in:

1. The Catalogue of the Spatial Data Infrastructure
   https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/41daef8b-155e-4608-b49c-c87ea45d931c

   Which shows the information in the XML metadata file

   https://www.nationaalgeoregister.nl/geonetwork/srv/api/records/41daef8b-155e-4608-b49c-c87ea45d931c/formatters/xml?approved=true

2. The European Data Portal
3. https://www.europeandataportal.eu/data/datasets/41daef8b-155e-4608-b49c-c87ea45d931c?locale=en

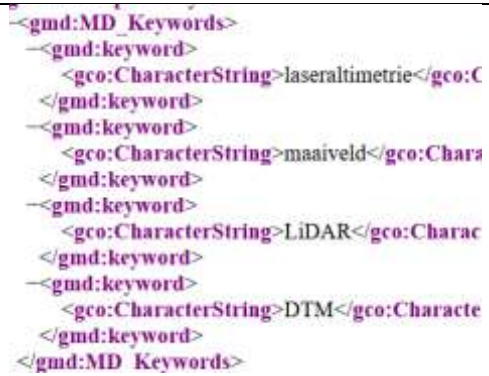
The following table shows in the left column what can information can be found in the national open data portal and in the right column what can be found in the EDP. The final column provides a comment regarding whether the information is complete and, if not, what information is missing.

Table 5. Illustrative example of comparative analysis performed for Dutch datasets

No.	Description National Portal Element	Description EDP Element	Comment
1	Title and Abstract		

			No comments. The same information is reflected in both.
2	Distributions		
			Apparently, this information is in the element 'landing page' in the EDP.

3	Spatial Extent		
	<gmd:geographicElement> <gmd:EX_GeographicBoundingBox> <gmd:westBoundLongitude> <gco:Decimal>3.118</gco:Decimal> </gmd:westBoundLongitude> <gmd:eastBoundLongitude> <gco:Decimal>7.253</gco:Decimal> </gmd:eastBoundLongitude> <gmd:southBoundLatitude> <gco:Decimal>50.67</gco:Decimal> </gmd:southBoundLatitude> <gmd:northBoundLatitude> <gco:Decimal>53.61</gco:Decimal> </gmd:northBoundLatitude> </gmd:EX_GeographicBoundingBox>	**Spatial** Coordinates: [[[3.118, 53.61], [7.253, 53.61], [7.253, 50.67], [3.118, 50.67], [3.118, 53.61]]] Type: Polygon ^ **Dataset extent** 	No comments. The same information is reflected in both.
4	Identifiers		
			No identifiers were reflected in the EDP.
5	Keywords		
		Keywords (4) 	No comments. The same information is reflected in both.

The following provides an overview of the analysis performed. Overall, the table highlights some significant areas of improvement with regards to the following:

1. Several datasets present in the national data portal is not reflected in the EDP.
2. Keywords are often harvested correctly but associated to a thesaurus rather than the keyword itself.
3. 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.
4. There is not always a translation of the tile and abstract to other required languages.

Table 6. Analysis Metadata Availability Between Dutch National Portal and EDP

No.	Dataset Name	Dataset Type	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	3D Basisvoorziening	3D, buildings and large topography	ok	ok	ok	ok	ok

2	Actueel Hoogtebestand Nederland 3 (AHN3)	Digital Terrein/Surface Model (DTM/DSM)	ok	harvested as landing page not as a distribution	ok	ok	ok except the keywords associated to a thesaurus
3	Basisregistratie Adressen en Gebouwen (BAG)	Building footprint, address data	No comparable with EDP	No comparable with EDP	No comparable with EDP	No comparable with EDP	No comparable with EDP
4	Wijk- en Buurten	Population statistics	ok	harvested as landing page not as a distribution	ok	ok	no
5	Landelijk Grondgebruik Nederland (LGN2018)	Land Use	ok	ok	ok	ok	no

1.6 3.4 Norway

In order to provide an overview of the quality of the metadata across different publication outlets as well as differences in the quality of metadata following harvesting, a comparative analysis was carried out based on the elements defined above. The following sections provides an illustrative example of how such an analysis was carried out in the context of the Norwegian infrastructure and then a summary table is provided highlighting the results of this analysis for all Norwegian datasets included in the project.

3.4.1 Illustrative Example: Dataset from the Norwegian Spatial Data Infrastructure

Dataset name: Id-36 – FKB Building

This dataset can be found in:

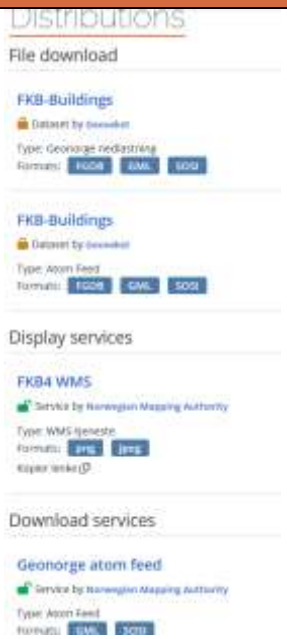
1. The National Open Data Portal
   https://kartkatalog.geonorge.no/metadata/fkb-bygning/8b4304ea-4fb0-479c-a24d-fa225e2c6e97
2. The European Data Portal

4. https://data.europa.eu/data/datasets/8b4304ea-4fb0-479c-a24d-fa225e2c6e97?locale=en

The following table shows in the left column what can information can be found in the national open data portal and in the right column what can be found in the EDP. The final column provides a comment regarding whether the information is complete and, if not, what information is missing.

Table 7. Illustrative example of comparative analysis performed for Norwegian datasets

No.	Description National Portal Element	Description EDP Element	Comment
1	Title and Abstract		
			There is no translation into English in the national open portal and no translation into Norwegian in the EDP.

2	Distributions		
			No comments. The same information is reflected in both.
3	Spatial Extent		

	Bounding box: • North: 70.85 • South: 57.47 • East: 33.67 • West: 4.56		No spatial extent was found in the EDP.
4	Identifiers		
	<gmd:fileIdentifier> <gco:CharacterString>8b4304ea-4fb0-479c-a24d-fa225e2c6e97</gco:CharacterString> </gmd:fileIdentifier>	**Identifiers** 8b4304ea-4fb0-479c-a24d-fa225e2c6e97	The identifier that is reflected in the file identifier (ISO19115)
5	Keywords		
		Keywords (16) 	No comments. The same information is reflected in both.

The following provides an overview of the analysis performed. Overall, the table highlights some significant areas of improvement with regards to the following:

1. There is not always a translation of the tile and abstract to other required languages.
2. Some of the distributions and spatial extents existing in the national portal is not reflected in the EDP.

Table 8. Analysis Metadata Availability Between Norwegian National Portal and EDP

No.	Dataset Name	Dataset Type	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	FKB-Buildings	Vector 2.5D	ok	ok	ok	ok	ok
2	Elveg 2.0	Vector, road network	ok	no	no	ok	ok

3	Elevation Laser data	Digital elevation Model (DEM)	ok. But EDP doesn't translate them into other languages	no	ok	ok	ok
4	Elevation DTM 1	Digital Terrain Model (DTM) WCS	ok	ok	ok	ok	ok
5	Elevation DSM 1	Digital Surface Model (DSM) WCS	ok	ok	ok	ok	ok
6	Population on 250 meter grid	Population	ok	ok	ok	ok	ok
7	Normal (precipitation, temperature, wind and air pressure)	Weather Data	ok	ok	ok	ok	ok

2.5 Estonia

In order to provide an overview of the quality of the metadata across different publication outlets as well as differences in the quality of metadata following harvesting, a comparative analysis was carried out based on the elements defined above. The following sections provides an illustrative example of how such an analysis was carried out in the context of the Estonian infrastructure and then a summary table is provided highlighting the results of this analysis for all Estonian datasets included in the project.

3.4.1 Illustrative Example: Dataset from the Estonian Spatial Data Infrastructure Portal

Dataset name: Id-22 – 3D Buildings, Estonia

This dataset can be found in:

1. The Catalogue of the National Spatial Data Infrastructure
   https://geoportaal.maaamet.ee/eng/Spatial-Data/Estonian-Topographic-Database-p305.html

   Which shows the information in the XML metadata file
   https://www.idee.es/csw-inspire-idee/srv/api/records/ES_SDGC_BU/formatters/xml

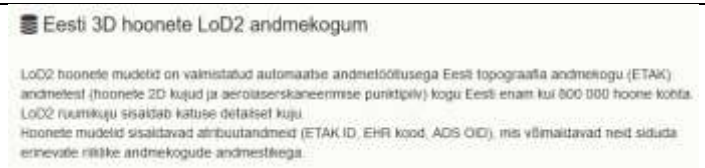
2. The European Data Portal

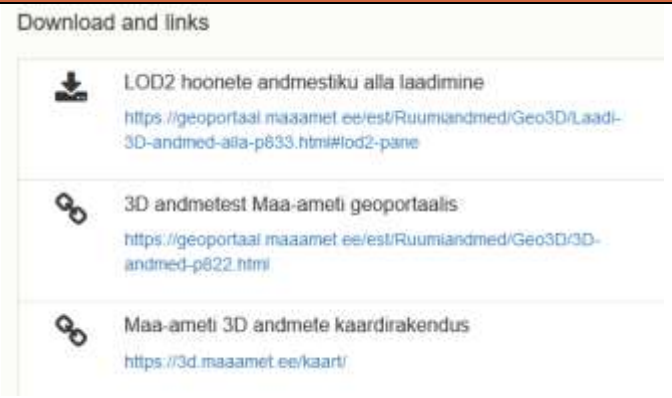
   https://data.europa.eu/data/datasets/85d8aaab-e411-4445-b142-2c4b5d18eb8b?locale=et

The following table shows in the left column what can information can be found in the national open data portal and in the right column what can be found in the EDP. The final column provides a comment regarding whether the information is complete and, if not, what information is missing.

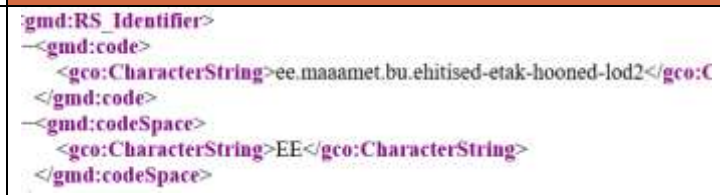
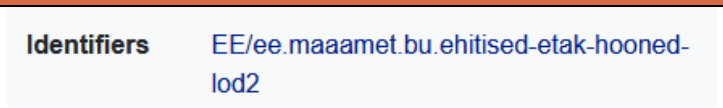
Table 9. Illustrative example of comparative analysis performed for Estonian datasets

No.	Description National Portal Element	Description EDP Element	Comment
1	Title and Abstract		

			No comments. The same information is reflected in both.

2	Distributions		
			Apparently, it works but it is not showing all the URL that are distributions.
3	Spatial Extent		

	<gmd:geographicElement> −<gmd:EX_GeographicBoundingBo −<gmd:westBoundLongitude> <gco:Decimal>21.77</gco:Dec </gmd:westBoundLongitude> −<gmd:eastBoundLongitude> <gco:Decimal>28.21</gco:Dec </gmd:eastBoundLongitude> −<gmd:southBoundLatitude> <gco:Decimal>57.51</gco:Dec </gmd:southBoundLatitude> −<gmd:northBoundLatitude> <gco:Decimal>59.68</gco:Dec </gmd:northBoundLatitude> </gmd:EX_GeographicBoundingB	**Spatial** Coordinates: [[[21.77, 59.68], [28.21, 59.68], [28.21, 57.51], [21.77, 57.51], [21.77, 59.68]]] Type: Polygon ^ Dataset extent 	No comments. The same information is reflected in both.

4	Identifiers		
			No comments. The same information is reflected in both.
5	Keywords		
			No comments. The same information is reflected in both.

The following provides an overview of the analysis performed. Overall, the table highlights some significant areas of improvement with regards to the following:

1. To a large extent, the distributions present in the national portal is not reflected in the EDP.

Table 10. Analysis Metadata Availability Between Estonian National Portal and EDP

No.	Dataset Name	Dataset Type	Title and Abstract	Distribution	Spatial Extent	Identifier	Keywords
1	LOD2 hooned	3D buildings	ok	no	ok	ok	ok
2	Eesti topograafia andmekogu - ehitised	Building roofprints	ok	ok	ok	ok	ok
3	Ehitisregister	Buildings: 2D geometry and attributes	ok	no	ok	no	ok
4	Eesti maakatte kõrgusmudel	DSM - Digital Surface Model	ok	no	ok	ok	ok
5	Eesti keskkonnaregistri ilmajaamade nimistu	Weather stations	ok	no	ok	ok	ok
6	Eesti maapinna kõrgusmudel	DTM - Digital Terrain Model	ok	no	ok	ok	ok
7	Rahvastikutiheduse ruutkaart	Population density 1x1km square map. (INSPIRE data model)	ok	no	ok	ok	ok

8	Eesti rahvastikutihedus NUTS3 piirkondades	Population density by NUTS 3 areas. (INSPIRE data model)	ok	no	ok	ok	ok
9	Aerolaserskaneerimise punktid	Lidar point cloud	ok	no	ok	ok	ok
10	Ilmajaamade aegread?	Weather stations time series	ok	no	ok	ok	ok
11	Eesti topograafia andmekogu - teed	Road network	ok	no	ok	ok	ok

1.7 3.6 Cross-Project Analysis

In total, 47 metadata files were analyzed in total. The following table provides an overview of where the metadata files originated from. Most of these originated from a spatial data infrastructure or national geoportal. The table also includes a column illustrating those metadata files for which their origin are not clear based on the EDP description.

Table 11. Origin of metadata file

Country	Open Data Portal	SDI	¿?	Total
Estonia		11		11
Finland	5	8		13
Norway			7	7
Spain	3	8		11
The Netherlands		5		5
Total	**8**	**32**	**7**	**47**

Five of the metadata files could not be compared with the data in the EDP. These are included in the table below and should be checked and/or corrected by the partner country to improve the availability of this metadata. These were subsequently not included in the following analysis.

Table 12. Missing metadata records in EDP

No.	Partner country	Dataset Name	Notes
1	The Netherlands	Basisregistratie Adressen en Gebouwen (BAG)	There does not seem to be a comparable BAG dataset on the EDP. The metadata source does not match the dataset on EDP (which is a dataset from Amsterdam rather than Kadaster).
2	Finland	Kansallisen maastotietokannan rakennukset	URL EDP is missing
3	Finland	Korkeusmalli 2m	URL EDP is missing
4	Finland	Rinnevarjostus	URL EDP is missing
5	Finland	Tilastokeskuksen aineistot	URL EDP is not working

3.6.1 Title and abstract

The following tables provide a summary of all the findings included above. With regards to the title and abstract, most metadata records are being harvested by the EDP correctly and only 2 of these metadata have not been well harvested at all; not including those for which metadata records do not match across the portals. There are six records which are not being correctly translated in the EDP.

Table 13. Summary of Analysis – Title and Abstract

Title and abstract	Estonia	Finland	Norway	Spain	The Netherlands	Total
no		2				2
ok	11	7	6	6	4	34
ok. But EDP doesn't translate them into other languages			1	5		6
Total	**11**	**9**	**7**	**11**	**4**	**42**

3.6.2 Distribution

With regards to the distributions, there is both a large proportion of records that are being harvested correctly and those which are not. There are six distributions which are being harvested as a landing page rather than a distribution and six which are being harvested as documentation.

Table 14. Summary of Analysis – Distribution

Distribution	Estonia	Finland	Norway	Spain	The Netherlands	Total
harvested as landing page not as a distribution		3		1	2	6
harvested in the "Documentation" element of EDP				5		5
most of them harvested in the "Documentation" element of EDP				1		1
no	10	2	2	1		15
ok	1	4	5	3	2	15
Total	**11**	**9**	**7**	**11**	**4**	**42**

3.6.3 Spatial Extent

With regards to the spatial extent, all metadata files that are applying DCAT and published in the national open data portal (8 out of 42) is missing the spatial extent in the original metadata file due to the application of DCAT. GeoDCAT should be used by the national portal in order for this to be correctly translated in the EDP. In summary, most of the metadata records are being harvested well, although six are not being harvested correctly; the reason for this is not clear and should be checked with the EDP. One record in particular is being harvested in as a spatial element but does not show the extent graphically as is being done by with other records.

Table 15. Summary Analysis – Spatial Extent

Spatial extent	Estonia	Finland	Norway	Spain	The Netherlands	Total
harvested in the "Spatial" element of EDP				1		1
Missing in the open data portal		5		3		8
no			1	5		6
ok	11	4	6	2	4	27
Total	**11**	**9**	**7**	**11**	**4**	**42**

3.6.4 Identifier

With regards to the identifier in the metadata record, all metadata files applying DCAT and published in the national open data portal (7 out of 42) is missing the original metadata file due to the fact that this is not a requirement for the DCAT profile. To correct this, GeoDCAT should be implemented in the national portal. In summary, most records are being harvested well by the EDP with only 1 not being harvested well. This should be addressed by the EDP.

Table 16. Summary Analysis - Identifier

Identifier	Estonia	Finland	Norway	Spain	The Netherlands	Total
no	1					1
no identifier in the national open data portal		4		3		7
ok	10	5	7	8	4	34

Total	**11**	**9**	**7**	**11**	**4**	**42**

3.6.5 Conclusions and Recommendations

Although the large extent of the metadata records are being harvested well as highlight over the course of the previous chapter, there is some room for improvement in both the publication of metadata in the national metadata portals as well as the harvested record in the EDP.

1. Regarding the translation of profiles, there seems to be some issues with some translations as seen in the example of Finland where the title is not well translated into English. Note: Norwegian is not available.
2. If the metadata is made according to DCAT, all the elements are available in EDP but it is necessary to refine the mapping when the metadata harvested is made applying INSPIRE. We should work together with EDP to improve this.
3. When harvesting INSPIRE metadata, EDP shows the spatial information that GeoDCAT allows to describe since it has a specific element, for example, *Spatial extent* and *identifiers*.
4. When harvesting INSPIRE metadata, EDP is not working properly apparently with *distributions*, which is a very important element. The information is not missing, since it is in the element "landing page", but it is not in the correct place.
5. Regarding the keywords, EDP shows all the keywords that are not defined using a thesaurus. But if the keywords are associated to a thesaurus, then the keyword is missing.

To improve the overall quality of geospatial metadata, the above should be addressed either by the EDP or by national open portals or spatial data infrastructures. Should the goal be to enrich existing metadata, it would be wise to first ensure that this base information is of high quality and enrich the metadata further based on those principles.

In order to provide some context in terms of the difference in the quality of metadata in the EDP when harvested from a national open data portal and when harvested from an SDI, the following sections provide a summary analysis.

3.7.1 Metadata from Open Data Portals

According to the datasets analysed, these are the results when the EDP harvests from a national open data portal. More specifically, the metadata harvested by the EDP is this case is already available in DCAT (and occasionally in GeoDCAT).

Table 17. Quality of metadata elements harvested from National Open Data Portals.

Open Data Portals	Spain	Finland	Norway
Title and abstract	ok	The title is not well translated into English	No translation in Norweigian
Distributions	ok	ok	ok
Spatial extent	no	no	no
identifiers	no	no	ok
Keywords	ok	ok	ok

3.7.2 Metadata from Spatial Data Infrastructure

According to the datasets analysed, these are the results when EDP harvests from a catalogue of a Spatial Data Infrastructure, i.e. the metadata applies INSPIRE and EDP harvests transforming the metadata from INSPIRE (published in the SDI catalogue) to GeoDCAT.

Table 18. Quality of metadata elements harvested from Spatial Data Infrastructure.

Spatial Data Infrastructure	Spain	Finland	The Netherlands	Estonia
Title and abstract	ok	ok	ok	ok
Distributions	*1	*1	*1	*2
Spatial extent	ok	ok	ok	ok
identifiers	ok	ok	ok	ok
Keywords	ok	*3	ok	ok

*1: The information is returned in the 'landing page' section of the EDP, not in the distributions section.

*2: Apparently, it works but it is not showing all the URLs that are distributions.

*3: EDP doesn't show the keywords if they are associated to a thesaurus.

The above sub-sections highlight that there are quality differences in the metadata depending on whether the EDP is harvesting information from an open portal or from a geoportal. What the sections do not necessarily highlight is the fact that there are many cases where the EDP is harvesting *both* the open data portal and the geoportal, resulting in duplications of the same dataset descriptions (often with differing levels of quality) in the EDP. How this should be solved is an issue that should be addressed by the EDP, perhaps considering whether the original data is geospatial in nature and, therefore, is best harvested from a native geospatial source in order to support the harvesting of rich metadata.

1.8 3.7 Cross-Partner Analysis

The previous analysis was carried out in order to assess the quality of the metadata in the EDP based on its harvesting process for only those datasets that will be used in the use cases developed by the GeoE3 project. These datasets are, of course, only a small number of all geospatial datasets published by partner countries which should be harvested by the EDP. With the goal of assessing how many of the geospatial datasets currently available through national geoportals are being harvested in some form by the EDP, an analysis was performed which simply counted the number of datasets available in the national portals and those available in the EDP for each partner country. The following table provides an overview of these results.

Table 19. Comparative count of geospatial datasets available in national catalogues and in EDP per partner country.

No.	Partner Country	No. Geospatial Datasets in National Portal	No. Datasets in National Open Data Portal	No. of Geospatial Datasets in EDP
1	Spain	14,203 (Spatial Data Infrastructure)	68,439 (Spanish Open Data Portal)	13,841 (harvested from the SDI of Spain: IDEE)
2	Finland	1,719 (Spatial Data Infrastructurehttps://www.avoindata.fi/en)	2,218 (Finnish	2,222 (harvested

			Open Data Portal)	from the Open Data Portal)
3	Netherlands	8,703 (Spatial Data Infrastructure)		7,233 (harvested from the SDI: National Georegister)
4	Norway	8,145 (Spatial Data Infrastructure)		382 (harvested from the SDI: Geonorge)
5	Estonia	273 (Spatial Data Infrastructure)		273 (harvested from the SDI: National Geoportal)

The above table highlights there does exist some (significant) differences in the number of datasets available in the national portals and those that are being harvested by the EDP. The reasons for this are not known based on the above table but should be investigated in moving forward with improving the quality of metadata.

1.9 3.8 Interoperability Map

One of the overall goals of the GeoE3 project is to improve the interoperability of datasets between partner countries, most notably those which share national borders. To support the assessment of the maturity of data and services used in GeoE3 use cases in terms of interoperability, the Interoperability Map is used as a tool. Because metadata is key to support interoperability at different levels across different datasets, the assessment of metadata plays a key role in the overall assessment of interoperability maturity levels per partner country. As such, an analysis was carried out to assess the availability and detail of metadata based on the criteria and map the results of these onto the Interoperability Map in support of the overall assessment of maturity of interoperability. The following sections and accompanying tables provide an overview of the results of each partner country. Please note, legal interoperability levels do not appear on the maps below because these do not require the availability of metadata for interoperability.

3.8.1 Spain Interoperability Map

The assessment carried out in the table below has only been done for those datasets included for use in the use cases defined for the GeoE3 project. In order to assess the full interoperability of partner countries' infrastructures, these should be applied across a wider range of datasets.

Categories	Level 0: Not interoperable and cannot be integrated	Level 1: Minimal interoperability and can be integrated with extra effort	Level 2: Intermediate interoperability and can be integrated mostly automatically	Level 3: Advanced/Optimal interoperability and can be integrated automatically
Technical Aspects/Data Access				
Metadata discoverability	No metadata available	Metadata available nationally	Metadata provided through APIs.	Metadata provided through DCAT AP 2.0 or OGC API records
Data accessibility	No data available	Data available with legacy APIs	Data available through OGC APIs.	Data available with OGC APIs
Semantic Aspects				
Vocabulary and data specifications	Vocabulary/ data descriptions not available and cannot be integrated	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Minimal definitions available and can be integrated with extra effort	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Intermediate interoperability	Vocabulary and data specifications

Data content and data quality	Data content and data quality are not described and cannot be integrated	Data content and data quality are described, but not according to any standards or in machine readable form.	Data content is sufficient for the expected usage in machine readable form.	Data content and quality are well described in machine readable form (e.g. UML).
Quality assessment	No quality assessment information available	Quality assessment done but not available through metadata	Quality goals defined and available through metadata	Quality assessment available through Data Quality Vocabulary (DQV)

3.8.2 Finland Interoperability Map

The assessment carried out in the table below has only been done for those datasets included for use in the use cases defined for the GeoE3 project. In order to assess the full interoperability of partner countries' infrastructures, these should be applied across a wider range of datasets.

Categories	Level 0: Not interoperable and cannot be integrated	Level 1: Minimal interoperability and can be integrated with extra effort	Level 2: Intermediate interoperability and can be integrated mostly automatically	Level 3: Advanced/Optimal interoperability and can be integrated automatically
Technical Aspects/Data Access				
Metadata discoverability	No metadata available	Metadata available nationally	Metadata provided through APIs.	Metadata provided through DCAT AP 2.0 or OGC API records
Data accessibility	No data available	Data available with legacy APIs	Data available through OGC APIs.	Data available with OGC APIs
Semantic Aspects				

Vocabulary and data specifications	Vocabulary/ data descriptions not available and cannot be integrated	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Minimal definitions available and can be integrated with extra effort	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Intermediate interoperability	Vocabulary and data specifications
Data content and data quality	Data content and data quality are not described and cannot be integrated	Data content and data quality are described, but not according to any standards or in machine readable form.	Data content is sufficient for the expected usage in machine readable form.	Data content and quality are well described in machine readable form (e.g. UML).
Quality assessment	No quality assessment information available	Quality assessment done but not available through metadata	Quality goals defined and available through metadata	Quality assessment available through Data Quality Vocabulary (DQV)

3.8.3 Netherlands Interoperability Map

The assessment carried out in the table below has only been done for those datasets included for use in the use cases defined for the GeoE3 project. In order to assess the full interoperability of partner countries' infrastructures, these should be applied across a wider range of datasets.

Categories	Level 0: Not interoperable and cannot be integrated	Level 1: Minimal interoperability and can be integrated with extra effort	Level 2: Intermediate interoperability and can be integrated mostly automatically	Level 3: Advanced/Optimal interoperability and can be integrated automatically
Technical Aspects/Data Access				

Metadata discoverability	No metadata available	Metadata available nationally	Metadata provided through APIs.	Metadata provided through DCAT AP 2.0 or OGC API records
Data accessibility	No data available	Data available with legacy APIs	Data available through OGC APIs.	Data available with OGC APIs
Semantic Aspects				
Vocabulary and data specifications	Vocabulary/ data descriptions not available and cannot be integrated	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Minimal definitions available and can be integrated with extra effort	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Intermediate interoperability	Vocabulary and data specifications
Data content and data quality	Data content and data quality are not described and cannot be integrated	Data content and data quality are described, but not according to any standards or in machine readable form.	Data content is sufficient for the expected usage in machine readable form.	Data content and quality are well described in machine readable form (e.g. UML).
Quality assessment	No quality assessment information available	Quality assessment done but not available through metadata	Quality goals defined and available through metadata	Quality assessment available through Data Quality Vocabulary (DQV)

3.8.1 Norway Interoperability Map

The assessment carried out in the table below has only been done for those datasets included for use in the use cases defined for the GeoE3 project. In order to assess the full interoperability of partner countries' infrastructures, these should be applied across a wider range of datasets.

Categories	Level 0: Not interoperable and cannot be integrated	Level 1: Minimal interoperability and can be integrated with extra effort	Level 2: Intermediate interoperability and can be integrated mostly automatically	Level 3: Advanced/Optimal interoperability and can be integrated automatically
Technical Aspects/Data Access				
Metadata discoverability	No metadata available	Metadata available nationally	Metadata provided through APIs.	Metadata provided through DCAT AP 2.0 or OGC API records
Data accessibility	No data available	Data available with legacy APIs	Data available through OGC APIs.	Data available with OGC APIs
Semantic Aspects				
Vocabulary and data specifications	Vocabulary/ data descriptions not available and cannot be integrated	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Minimal definitions available and can be integrated with extra effort	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Intermediate interoperability	Vocabulary and data specifications
Data content and data quality	Data content and data quality are not described and cannot be integrated	Data content and data quality are described, but not according to any standards or in machine readable form.	Data content is sufficient for the expected usage in machine readable form.	Data content and quality are well described in machine readable form (e.g. UML).

Quality assessment	No quality assessment information available	Quality assessment done but not available through metadata	Quality goals defined and available through metadata	Quality assessment available through Data Quality Vocabulary (DQV)

3.8.5 Estonia Interoperability Map

The assessment carried out in the table below has only been done for those datasets included for use in the use cases defined for the GeoE3 project. In order to assess the full interoperability of partner countries' infrastructures, these should be applied across a wider range of datasets.

Categories	Level 0: Not interoperable and cannot be integrated	Level 1: Minimal interoperability and can be integrated with extra effort	Level 2: Intermediate interoperability and can be integrated mostly automatically	Level 3: Advanced/Optimal interoperability and can be integrated automatically
Technical Aspects/Data Access				
Metadata discoverability	No metadata available	Metadata available nationally	Metadata provided through APIs.	Metadata provided through DCAT AP 2.0 or OGC API records
Data accessibility	No data available	Data available with legacy APIs	Data available through OGC APIs.	Data available with OGC APIs
Semantic Aspects				
Vocabulary and data specifications	Vocabulary/ data descriptions not available and cannot be integrated	Vocabulary and data specifications including data content and data quality are described, but not according to any standards. Minimal	Vocabulary and data specifications including data content and data quality are described, but not according	Vocabulary and data specifications

		definitions available and can be integrated with extra effort	to any standards. Intermediate interoperability	
Data content and data quality	Data content and data quality are not described and cannot be integrated	Data content and data quality are described, but not according to any standards or in machine readable form.	Data content is sufficient for the expected usage in machine readable form.	Data content and quality are well described in machine readable form (e.g. UML).
Quality assessment	No quality assessment information available	Quality assessment done but not available through metadata	Quality goals defined and available through metadata	Quality assessment available through Data Quality Vocabulary (DQV)