High-value datasets: Cross-country findability and comparability of metadata

data.europa.eu
The official portal for European data

This study has been prepared as part of data.europa.eu, an initiative of the European Commission. The Publications Office of the European Union is responsible for the management of data.europa.eu, the official portal for European data.

For more information about this study, please contact the following.

European Commission
Directorate-General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Email: CNECT-G1@ec.europa.eu

data.europa.eu
Email: info@data.europa.eu

Authors:
Leon van Eendenburg
Jolanta Suárez

Edited by:
Beatriz Fernández Nebreda
Nataliya Rozbroj Jasinskaja

Last update: 25 September 2024
https://data.europa.eu/

DISCLAIMER
The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use that may be made of the information contained herein.

Luxembourg: Publications Office of the European Union, 2024
© European Union, 2024

The Commission's reuse policy is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39, ELI: http://data.europa.eu/eli/dec/2011/833/oj). Unless otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (https://creativecommons.org/licences/by/4.0/). This means that reuse is allowed, provided appropriate credit is given and any changes are indicated.

ISBN 978-92-78-44317-7 doi: 10.2830/0033148 Catalogue number: OA-01-24-007-EN-N

Contents

Executive summary

The implementation of high-value datasets (HVDs) across European Union (EU) Member States is crucial for advancing open data initiatives and facilitating cross-country data reuse. While progress varies among Member States, a standardised approach is essential to ensure the comparability and interoperability of HVDs at the EU level. The current study addresses the challenges in harmonising metadata and improving the findability and accessibility of HVDs, guided by the FAIR principles (based on improved findability, accessibility, interoperability and reuse) and the DCAT-AP HVD guidelines.

This pilot study evaluates the findability and metadata comparability of HVDs in selected Member States (Denmark, Estonia, Latvia and Finland ([1])). It was conducted between April and May 2024, prior to the entering into force of Commission Implementing Regulation (EU) 2023/138 (hereafter referred to as 'the regulation') in June 2024.

The study aims to assess the user experience of finding and using HVDs from different Member States by evaluating a sample of 24 datasets under the regulation's six thematic categories: geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, and mobility. It also provides recommendations for enhancing metadata standardisation and dataset findability.

The key findings are as follows.

- **HVD availability.** Out of the sample of 24 assessed datasets, 20 were found, three were gated and one was temporarily unavailable.
- **Metadata language.** Six datasets had metadata solely in the national language, posing a translation risk for international users.
- **Metadata clarity.** Most datasets provided clear metadata, but some lacked sufficient detail, leading to potential user confusion.
- **Regulatory references.** Eight datasets did not reference the regulation or Directive 2007/2/EC of the European Parliament and of the Council (also known as the Inspire directive) ([2]), complicating user efforts to verify compliance.
- **Findability on data.europa.eu.** Fourteen out of 24 datasets were easily findable, while four had poor findability. Eighteen datasets were located, but only 15 included a direct download link or a link to the data.

Based on the findings, the following recommendations were made to Member States.

- **Reference the regulation.** Include clear references to the regulation or the Inspire directive in the metadata to aid the user with identification and compliance verification.

([1]) The names of Member States are listed in accordance with EU protocol order, which arranges country codes alphabetically based on their names in their national languages.

([2]) Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (OJ L 108, 25.4.2007, p. 1–14, ELI: http://data.europa.eu/eli/dir/2007/2/oj).

- **Standardise titles.** Use dataset titles that align with those in the regulation or the Inspire directive to enhance searchability and recognition.
- **Enrich metadata.** Provide detailed descriptions, including scope and parameters, and link to additional information to ensure users understand the dataset contents without needing to download it.
- **Follow DCAT-AP HVD guidelines.** Implement DCAT-AP guidelines for structured and standardised metadata descriptions.
- **Tag HVDs on the national portals.** Ensure that HVDs are correctly tagged at the national level, in order for them to appear in the HVD filter in data.europa.eu and to improve their discoverability.

The study concludes that while HVDs are generally findable on national portals, the Inspire Geoportal and data.europa.eu, inconsistencies in metadata vocabulary, structure and accessibility remain, particularly in non-Inspire categories. Improvements in metadata referencing, standardisation and dataset tagging are necessary to enhance user experience and dataset interoperability across the EU. By adopting the recommended practices, Member States can significantly aid users in finding and utilising HVDs effectively.

1. Introduction

The implementation of HVDs across Member States is a critical step towards enhancing open data initiatives. While some countries have made significant progress in this area, there is a need for a standardised approach and methodology to ensure the comparability and interoperability of HVDs at the EU level. The **main challenges** lie in harmonising metadata across different governance frameworks and legal parameters, and in facilitating the findability of the same HVD for different Member States. This is essential for the successful integration of HVDs into the official portal for European data, data.europa.eu, and for anyone who wants to reuse and compare cross-country data.

An existing resource that helps ensure the harmonisation and interoperability of metadata are the **DCAT-AP HVD guidelines**, which enable standardised descriptions of public datasets. These guidelines are at the disposal of Member States to create metadata that meets the specifications of data portals in Europe using controlled vocabularies. While DCAT-AP HVD guidelines provide a framework for the categorisation and classification of datasets into high-value categories, this study covers how comparability can be improved by standardising metadata fields such as titles, descriptions and keywords.

This research also looks at the findability and accessibility of HVDs, elemental parts of the FAIR guiding principles for scientific data management and stewardship to which the European Commission has committed itself. This paper aims to develop a methodology for a pilot findability and accessibility of metadata study. The outcome will contribute to the optimisation of dataset interoperability, facilitating data sharing and reuse across the EU. As the topic is strongly related to data homogeneity, the approach followed is similar to that of the *Report on Data Homogenisation for High-value Datasets*. However, its main focus is on assessing the user's experience and interaction with HVDs sourced from selected Member States, focusing on usability and accessibility aspects of the availability of metadata and the ease with which HVDs can be accessed. The assessment of the content or quality of the datasets falls outside the scope of this study. Insights from the *Report on High-value Datasets Best Practices in Europe* also inform this paper.

This research was conducted between April and the end of May 2024, prior to the enforcement of the regulation, which took effect on 9 June 2024. This study by no means focuses on or inspects the progress of individual Member States. Its primary objectives are to gather insights from Member States that have significantly advanced in adopting the regulation, as revealed by the *2023 Open Data Maturity Report*, and to examine HVDs from these countries from a user's perspective, identifying opportunities for enhancements to maximise the value of HVDs for users.

This study employs the methodology of the analysis of the findability and metadata comparability of available HVD data inventories in EU-27 Member States. It is a pilot study, focusing on 24 selected datasets ('the sample') for each of the six thematic categories (see section 2.1. HVDs in scope) among the most advanced Member States in HVD implementation, as per the *2023 Open Data Maturity Report*, namely Denmark, Estonia, Latvia and Finland. The requirements laid down by the regulation will serve as the basis for the metadata study.

Additionally, the study includes **an analysis** of titles and descriptions from the sample for the selected Member States, aiming to assess the consistency of the metadata descriptions across Member States and enhancing user comprehension.

The paper also investigates the presence of the selected datasets on data.europa.eu. Ideally all HVDs from the sample, along with their complete metadata, should be successfully indexed on national portals as the source, and accessible through data.europa.eu. Commonalities and differences among the Member States are considered together with any missing elements required for successful harvesting into data.europa.eu, including **recommendations** on how to close this gap.

Finally, recommendations on how Member States can advance the standardisation of the HVD metadata for these datasets to be harvested automatically into data.europa.eu are provided in line with the findings.

2. Methodology

2.1. HVDs in scope

As this is a pilot study, the research focused on **a sample of HVDs from the four selected Member States**, namely one from each of the six thematic categories. Two of the HVDs ('Companies and company ownership' and 'Transport networks') were also part of the *Report on Data Homogenisation for High-value Datasets*. Therefore, some of its insights will be reused in this report and complemented with new findings on findability and metadata comparability. The table below contains the list of HVDs that are included in this study.

Table: Sample set of HVDs included in this report

Thematic category	Sample HVD
Geospatial (Inspire)	Administrative units
Earth observation and environment (Inspire)	Environmental monitoring facilities
Meteorological (Inspire)	Climate data – Validated observations
Statistics	Poverty
Companies and company ownership	Companies and company ownership
Mobility (Inspire)	Transport Networks – Road networks

2.2. Locating HVDs on national portals

In the process of locating HVDs, the search commenced on Google as this can lead to matches on national open data portals, the Inspire geoportal and data.europa.eu. All searches were performed in English, as would likely be the case for end users who want to combine HVDs from different Member States.

Inspire-related datasets (the first three categories: geospatial, earth observation and environment and meteorological) are best found directly on the Inspire geoportal. For the non-Inspire related HVDs, the search process was as follows.

1. A Google search of key term(s) in the HVD, combined with the name of the Member State and the term 'dataset'. Optional: a reiteration of the search with different search terms.

2. Following the hit that leads to a website that we could verify as belonging to a national statistical authority. Optional: translating the webpage using a browser extension; and renavigating to the portal where open data can be downloaded (e.g. going from an article to an open data portal on the domain of the national institute).
3. Reading metadata and assessing whether it is the desired HVD.
4. Downloading and opening the dataset that best matches the HVD or downloading multiple sets if necessary.

To verify the legitimacy of the public authorities / national statistical institutes in step 2, these were cross-referenced against the list of national statistical institutes and other national authorities published by Eurostat, the statistical office of the EU.

2.3. Locating HVDs on data.europa.eu

All HVDs were searched for on data.europa.eu in a way that it is representative of the search behaviour of an international user, i.e. only using search terms similar to the title of the HVD in the regulation. For example, searching for 'weather data', 'weather observations' and 'climate data' in the effort to find the HVD on meteorological data. It was concluded that looking for a dataset on data.europa.eu without prefiltering a national data catalogue gave inadequate results, therefore a data catalogue from one of the sample Member States was always selected before making a search.

2.4. Analysing the metadata against the regulation

The sample HVDs were analysed against the technical specifications / requirements outlined in the regulation. The requirements differ per category and per dataset. Some general requirements for publication applicable to all HVDs (Article 3) include:

- available in a machine-readable format;
- available free of charge (except for libraries, museums and archives);
- available via application programming interfaces (APIs);
- available for bulk download (where indicated in the Annex);
- updated at a frequency indicated in the Annex.

As per Article 4 of the regulation, HVDs shall be made available for reuse under the conditions of the Creative Commons Public Domain Dedication , the Creative Commons BY 4.0 licence or any equivalent or less restrictive open licence, as set out in the Annex, allowing for unrestricted reuse. Some requirements specific to the thematic category concern:

- metadata requirements;
- use of vocabulary and taxonomy.

For each HVD, these requirements were checked and documented in a structured way.

2.5. A user's perspective on metadata

In addition to the aspects explicitly mentioned in the regulation, this study also focused on determining whether the metadata provides enough detail for users to clearly understand the value and potential usefulness of the dataset. This included the following:

- Checking the metadata language (English or national language).
- Checking whether the metadata gives sufficient clarity on the contents (ranked from low to high). For example, whether it is clear what will be in the dataset, what parameters are included, how recent it is, by whom it is published, etc. A low score indicates that the metadata is insufficient, requiring users to download the dataset to assess its value. A high score means that the metadata provides clear information about the dataset, allowing users to evaluate its value without needing to download it.
- Checking whether the HVD references the regulation or the Inspire directive. This could be anything in the metadata from a label or a sentence in the description, to a direct mention in the title making it clear to a user that this is an HVD compliant with either or both of the legal frameworks.
- Other comments were also documented, such as whether a dataset is comprised of one file or multiple ones, or whether a user would have to create an account before being able to download or connect via an API.

These aspects are all important to the user as they impact the accessibility of the HVDs and their interoperability in general.

2.6. A user's perspective on HVD findability on data.europa.eu

Finally, the findability of HVDs on data.europa.eu was assessed taking into account the following aspects:

- Whether the HVD is findable on the national portals without knowing its exact title.
- Whether the page contains the actual dataset, a link to the dataset or none of those.

NB: This research was conducted in May 2024, prior to the introduction of a search feature on data.europa.eu in June 2024, which allows users to specifically filter for HVDs. Naturally, this greatly improves the findability of HVD on data.europa.eu, compared to what is published in this report. However, some HVDs on the portal may not be marked as such at their source as Member States continue to implement the regulation. Therefore, the findings from this evaluation remain relevant, though to a lesser degree.

3. Results

3.1. Locating HVDs on national portals

As mentioned in the *Report on Homogenisation for High-value Datasets*, finding HVDs on national portals is a challenging and error-prone task. Locating specific HVDs frequently requires the use of multiple search terms. Therefore, aggregating these datasets on data.europa.eu and introducing a specialised advanced search feature for HVDs are crucial measures that will significantly improve accessibility and improve the user experience.

For this study, we explored a specific use case: locating and integrating poverty data across the four selected Member States, with a particular emphasis on how easily users can find and access this information. We chose to focus on the at-risk-of-poverty rate as our key metric. This choice was not arbitrary; it is a measure explicitly mentioned in the regulation's Annex 4.1, Table 13, which outlines the 'Specification for the yearly poverty rate high-value dataset' ([3]). By examining this particular dataset, we aimed to get a realistic picture of how well the system works for users trying to access important social indicators.

3.1.1. Example: Finding the poverty dataset for Denmark

Applying the search term 'Denmark at-risk-of-poverty rate dataset' does not result in relevant hits on a national statistics portal. From the Eurostat list of recognised national institutes, it can be seen that Danmark Statistik is the Danish portal to investigate. On the site of Danmark Statistik, searching for 'poverty' does not lead to datasets on poverty. The way to arrive at the dataset is to go to Statistics Bank, go to the subject 'Labour and income' and select 'Personal and family income', as seen in Figure 1: A screenshot of statbank.dk, the recognised body for statistical data in Denmark.

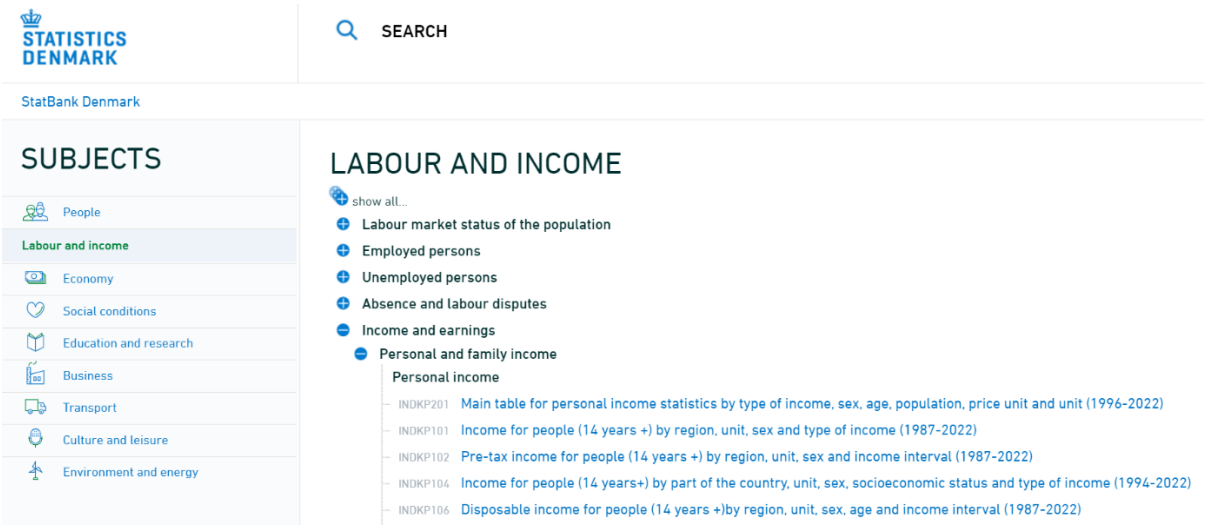

Figure 1: A screenshot of statbank.dk, the recognised body for statistical data in Denmark

There is no direct parameter for 'at risk of poverty', but there is data on disposable income, which can be used to calculate the rate.

3.1.2. Example: Finding the poverty dataset for Estonia

Searching 'Estonia at-risk-of-poverty rate dataset' on Google results in a hit at stat.ee, which was verified as the official statistical institution of Estonia. The link immediately leads to the database of stat.ee in the section 'Poverty and inequality'. The top table in the dataset 'Poverty and deprivation rate by age group and sex' contains the parameter 'at-risk-of-poverty rate %' and the breakdowns by age and sex. Other tables contain breakdowns by citizenship, labour status, education, income

([3]) Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use (OJ L 19, 20.1.2023, p. 43–75, ELI: http://data.europa.eu/eli/reg_impl/2023/138/oj).

quintile and place of residence (see Figure 2: Screenshot from stat.ee, Estonia's official statistical database.

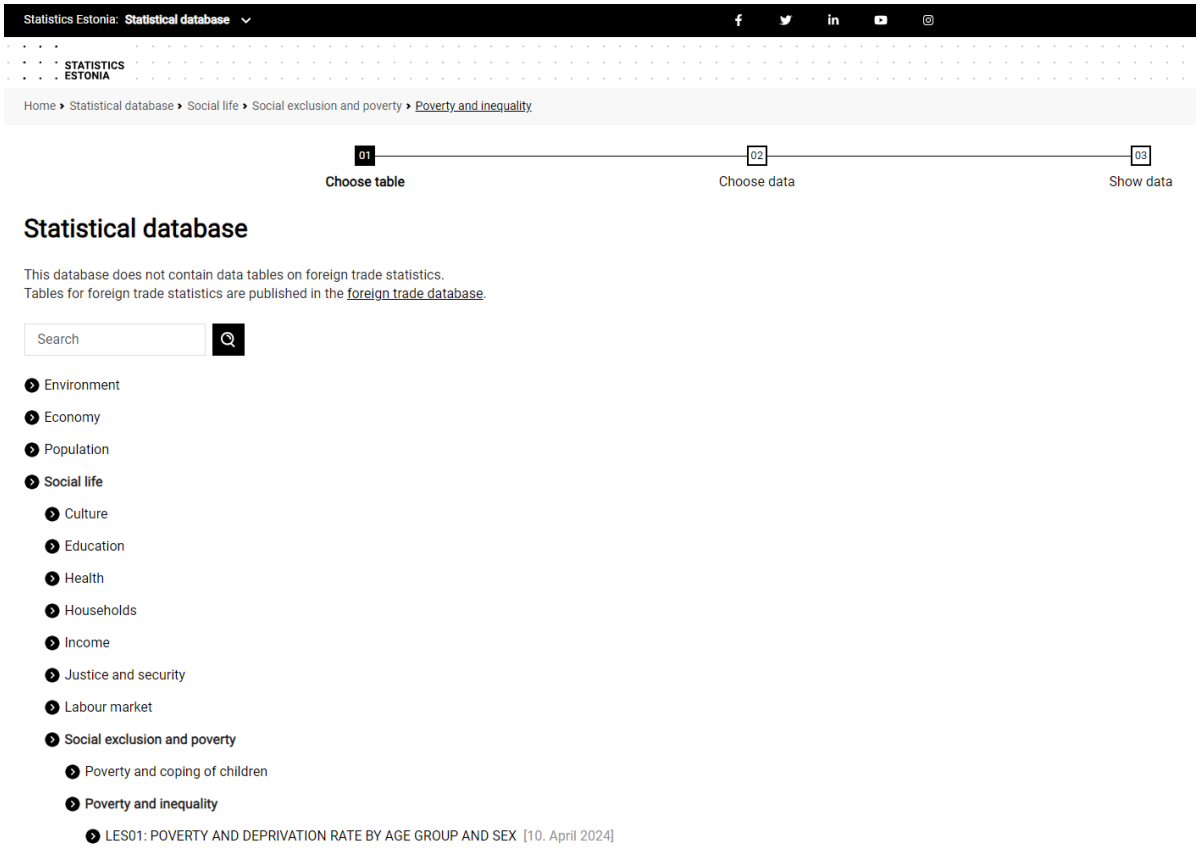

Figure 2: Screenshot from stat.ee, Estonia's official statistical database

3.1.3. Example: Finding the poverty dataset for Latvia

The search term 'Latvia at-risk-of-poverty rate dataset' has a top hit at stat.gov.lv, the recognised national statistical institute of Latvia. The hit directly leads to the Poverty and Inequality database, containing all relevant datasets and breakdowns in an intuitive interface. A screenshot of the database is shown in Figure Figure 3: Screenshot from stat.gov.lv, the recognised body for statistical data in Latvia

Figure 3: Screenshot from stat.gov.lv, the recognised body for statistical data in Latvia

3.1.4. Example: Finding the poverty dataset for Finland

Looking for the HVD of Finland by googling 'Finland at-risk-of-poverty rate dataset' leads to a hit on the site Tilastokeskus with the domain name stat.fi. In Eurostat's list of national institutes and other national authorities, it can be verified that this is the domain of Statistics Finland, a recognised body. The link leads to an article on poverty in Finland. On the page of the article there is a link to the actual database. Once there, searching for a dataset with 'poverty risk' gives the result shown in Figure 4: Screenshot from stat.fi, Finland's official statistical database.

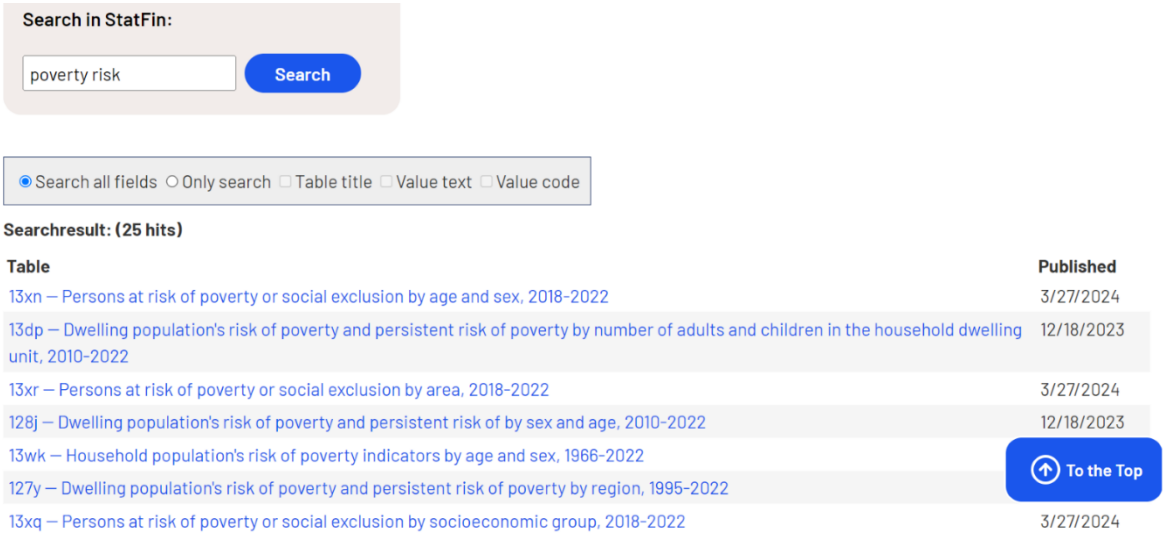

Figure 4: Screenshot from stat.fi, Finland's official statistical database

3.1.5. Findings from the example

The example above showcases the effort an HVD user would need to put into finding the same dataset from the four Member States to compare them. It demonstrates that it is possible to find the datasets for the Member States, and that each of the Member States has portals with search options for users in English and in the national language. Nevertheless, if a user wants to find, collect and compare the same data from different Member States, it would be necessary to:

- go to portals different from the national statistical institutes;
- verify the legitimacy of the portals;
- try various search terms;
- investigate definitions of parameters to verify that the data is valuable for comparison.

In other cases, the user might also have to translate web pages with the risk of mistranslating important nuances. Even when searching in English, multiple terms are often used to describe the same data:

- Denmark: 'Disposable income for people (14 years+) by region, unit, sex, age and income interval';
- Estonia: 'Poverty and deprivation rate by age group and sex';
- Latvia: 'At-risk-of-poverty rate by sex and age';
- Finland: 'Household populations risk of poverty indicators by age and sex'.

3.2. Assessment of metadata against the regulation

The search procedure as described above was performed for the sample of 24 datasets while recording the metadata aspects as described in section 2.4. Analysing the metadata against the regulation. The detailed results can be found in Annex I. Sample HVDs used for this study

The general availability of HVDs is presented in Figure 5**Error! Reference source not found.**Out of the 24 HVDs analysed, 20 were found on either national portals, the Inspire geoportal or data.europa.eu. Three HVDs were found but required the user to create an account before being able to access the actual data. These HVDs were therefore presented in the chart as 'gated'. An example of this is the 'Company and company ownership dataset' from Denmark, which was locked behind a paywall. In one instance, the HVD was not found, but this was due to a portal being temporarily unavailable.

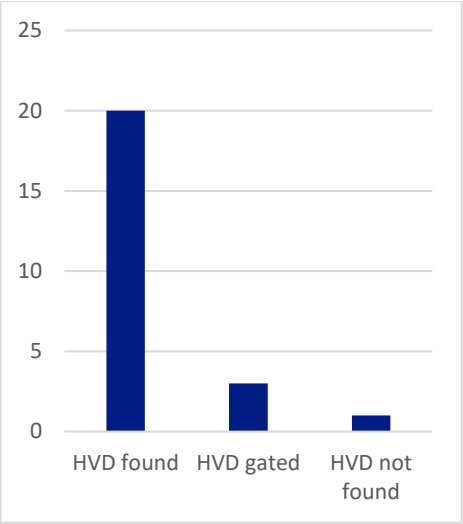

Figure 5: General HVD availability

3.3. Metadata from a user's perspective

As discussed in section 2.5. A user's perspective on metadata, the assessment of metadata in this study went beyond a review against the regulation and looked at the metadata from a user's perspective. In this regard, three aspects of the metadata were analysed.

- **Metadata language.** English or national language.
- **Metadata clarity.** Is it clear what will be in the dataset, what parameters are included, how recent it is, by whom it is published, etc.? A low score indicates that the metadata is insufficient, requiring users to download the dataset to assess its value. A high score means that the metadata provides clear information about the dataset, allowing users to evaluate its value without needing to download it.
- **Metadata references the regulation.** Is there any reference to the regulation in the metadata, making it clear to users that are looking for a specific HVD from a specific Member State that they need not look further.

FiguresFigure 6: Metadata languageFigure 7 and Figure 8 present an overview of the assessment of the metadata from a user's perspective. A more in-depth analysis can be found in ***Annex I. Sample HVDs used for this study***

3.3.1. Metadata from a user's perspective: Metadata language

Figure 6 shows that in six out of the 24 assessed datasets, the metadata was written in the national language of the Member State. For an international user, this means the metadata will have to be translated to be understood. This presents a realistic risk of mistranslation due to a wide range of HVD-specific terminology, which can lead to the user misunderstanding the data.

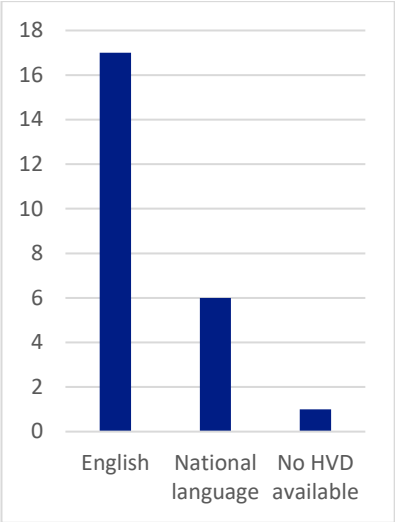

Figure 6: Metadata language

3.3.2. Metadata from a user's perspective: Metadata clarity

The analysis focused on determining whether the metadata clearly conveyed to the assessor what the dataset contains (see Figure 7). If the metadata was in the national language, it was translated first. For most of the datasets identified, the metadata provided enough clarity for users to understand the contents of the dataset. In the cases where this was not clear, the dataset contained only minimal information, leaving the user to question how the data was acquired, whether it was quality checked, etc. For a user, this could be a reason to discard the dataset as not useful.

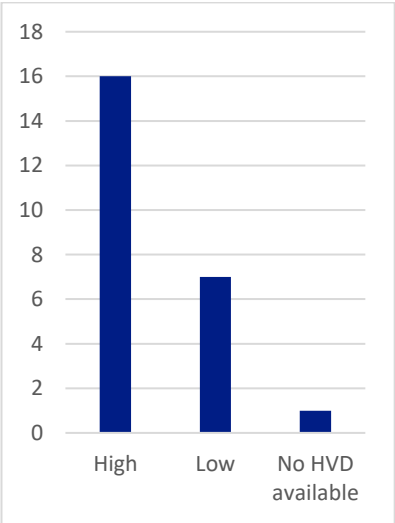

Figure 7: Metadata clarity

3.3.3. Metadata from a user's perspective: Metadata references the regulation

Figure 8 shows that eight out of the 24 datasets do not reference the regulation nor the Inspire directive in any way. For a user who is looking for a specific HVD this means having to closely inspect the metadata and compare it against the requirements of the regulation to decide whether the datasets comply with the regulation. All data found on the Inspire geoportal automatically references the Inspire directive.

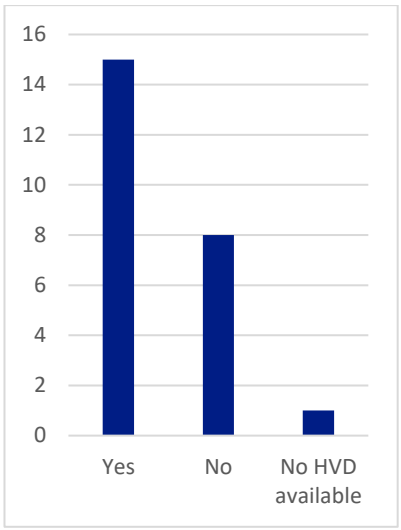

Figure 8: Metadata references the regulation

3.4. A user's perspective on HVD findability on data.europa.eu

Regarding the findability of HVDs on data.europa.eu, the following aspects of datasets were analysed.

- **HVD findable on data.europa.eu without knowing its title.** This means various relevant search terms were tried and if the HVD was not found, it was noted as not findable.
- **Metadata references the regulation.** Does the dataset metadata contain any reference to the regulation or the Inspire directive?
- **Data.europa.eu page contains the actual dataset.** Does the page contain the actual dataset, a link to the dataset or none of those?

Figures Figure 9, Figure 11 andFigure 12 provide an overview of the results of this assessment. More in-depth results can be found in ***Annex I. Sample HVDs used for this study***

NB: As mentioned in section 2.6. A user's perspective on HVD findability on data.europa.euat the time this research was conducted, data.europa.eu did not have the HVD filter feature it has now, at the time of writing. The HVD filter feature will significantly improve the findability and comparability of HVDs for its users.

3.4.1. User's perspective on HVD findability on data.europa.eu: Findability without knowing the exact title

Figure 9 illustrates the varying degrees of findability for HVDs on data.europa.eu when users are unaware of the specific dataset titles. Out of the 24 HVDs examined, 14 demonstrate good findability, indicating that users can locate these datasets relatively easily using intuitive search terms.

However, four HVDs exhibit poor findability, presenting significant challenges for users attempting to discover them without prior knowledge of their titles. This difficulty may be attributed to two main

factors. Firstly, some HVD titles are in national languages, and while data.europa.eu employs automatic machine translation for titles and descriptions, these translations are not always fully accurate or effective. Findability is further compromised when HVDs are given titles at the source which users might not intuitively associate with the content. It is worth noting that six out of the 24 HVDs were not found on data.europa.eu at all (represented as 'HVD not found' values in Figure 9), which is likely due to the dataset not having been marked as an HVD at the source. This should be resolved in due course as Member States continue to work hard at implementing the regulation.

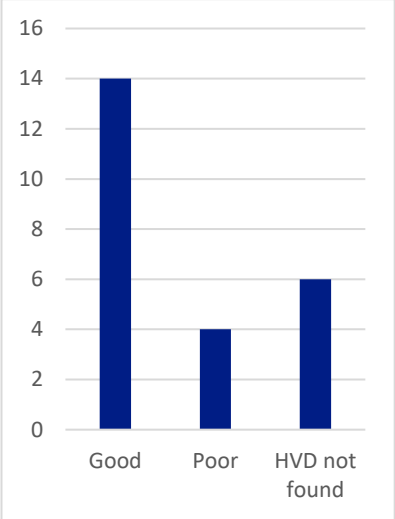

Figure 9: Findability on data.europa.eu without knowing the title

A final note regarding the findability of HVDs: data.europa.eu is organised in **catalogues**. Each Member State has one or more data catalogues to which all their datasets are linked. For the findability of datasets on data.europa.eu, it is important to search within the catalogue of the Member State that you are collecting data from. Without preselecting a data catalogue, the findability of datasets is very low. Figure 10 contains an example of preselecting the Latvian Geospatial Metadata Catalogue to increase the findability of datasets.

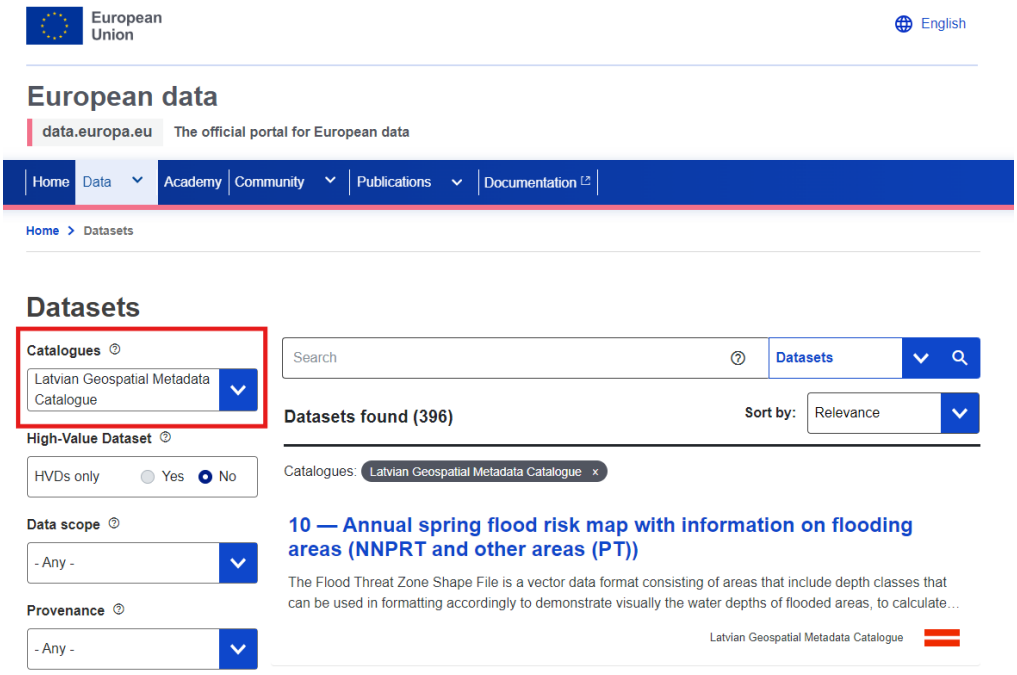

Figure 10: Visual demonstration of preselecting the Latvian Geospatial Metadata Catalogue

3.4.2. A user's perspective on HVD findability on data.europa.eu: Metadata references to the regulation

As seen in Figure 11, of the 18 HVDs that were found on data.europa.eu, nine contained a reference to the Inspire directive as being geospatial HVDs. At the time of the research, HVDs that did not fall under the Inspire categories did not yet have a reference to the regulation. As mentioned in the section above, this results in additional effort for the user, such as having to open the dataset and check its compliance with the regulation.

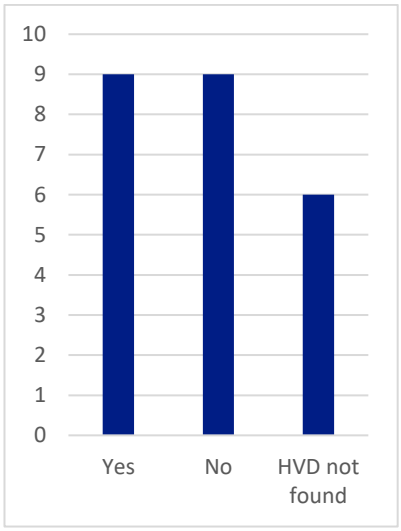

Figure 11: Metadata references the regulation on data.europa.eu

3.4.3. A user's perspective on HVD findability on data.europa.eu: data.europa.eu page contains the actual dataset

Out of the 18 HVDs that were found on data.europa.eu, 15 contained either the actual dataset as a downloadable file or a link to the data (see Figure 12). In three of the 18 HVDs, there was neither a file nor a link, making the page serve little purpose. The absence of the datasets or links to them can be explained by the datasets not being available on the national portals and hence not being able to be harvested onto data.europa.eu, an issue that should be solved as Member States move forward with the adoption of the regulation.

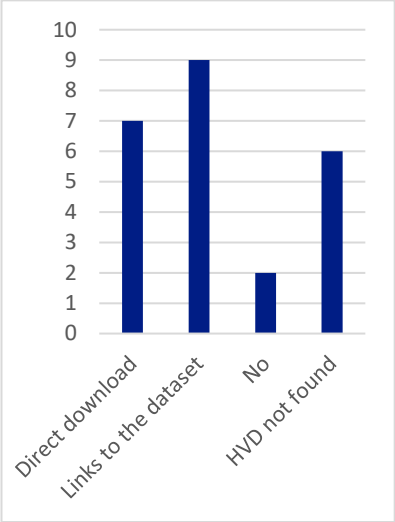

Figure 12: Page contains the actual dataset

4. Recommendations

4.1. Recommendations for Member States to improve findability and metadata accessibility

Following the research and the results presented, several recommendations were made for Member States to improve the findability and metadata accessibility of their HVDs. These recommendations are additional to what the regulation and the Inspire directive already imply in terms of dataset contents, licences, formats, scopes, etc., and are meant to help users collect and compare data from different Member States.

1. **In the metadata of a dataset, reference the regulation, together with the title of the HVD that the dataset corresponds to**.
   This would help users clearly identify and categorise the dataset, making it easier to understand its relevance and regulatory context.

   For example, Figure 13 shows part of the metadata that directly links a Latvian geospatial dataset to the matching Inspire directive. The new HVD regulation should be referenced in a similar way.

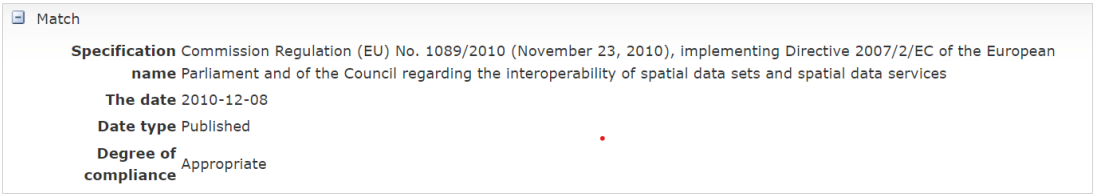

Figure 13: Screenshot of geometadati.viss.gov.lv, showing the metadata of the HVD on environmental monitoring facilities

2. **Use a dataset title that corresponds to the title mentioned in the regulation or the Inspire directive.**
   For example, a dataset with a title in a national language would be difficult to find for an international user. Having an English version of the HVD in addition to the national language in the national portals would help verify that the dataset titles are the same or at least similar to those in the regulation. The metadata in Figure 14 is an example of when this is not the case. The title in the national language makes it difficult for users to compare it with other similar HVDs across Member States.

Figure 14: Screenshot from data.europa.eu, showing the metadata of Estonia's 'Environmental monitoring facilities' HVD

3. **Enrich metadata by providing a description of the data such as its scope, parameters and where users can find additional information if necessary.**
   For example, for a meteorological dataset, the metadata could mention a list of all variables that are included in the dataset, together with a link to definitions and a link to a page that explains how the data was acquired. Figure 15 **Error! Reference source not found.**shows a best practice example, in which the metadata from the meteorological HVD links to a page with descriptions of all parameters included in the dataset.

Weather observations

Parameter name	Temporal resolution	Interpretation
Cloud cover	**Instant**	Total cloudiness given in numbers out of eight: 0 = clear skies, 8 = overcast, 9 = cloud coverage could not be determined. Takes into account all cloud layers.
Precipitation amount	Instant	Precipitation amount in previous 10 minutes. Calculated from precipitation intensity for easier usage. Observations only done every 10 minutes. Value 0 = No precipitation. Value - = No observation carried out
Wind speed	Instant	The average wind speed during the previous 10 minutes.

Figure 15: Screenshot from ilmatieteenlaitos.fi, showing interpretation guidance to the parameters of the meteorological datasets

4. **Apply DCAT-AP guidelines for metadata structuring.**
   Make use of the DCAT-AP guidelines for HVDs. DCAT-AP enables the standardised description of public sector datasets in Europe for data catalogues, facilitates content aggregation by platforms like the European Data Portal and simplifies dataset discovery for consumers through a single access point.

5. **Tag HVD datasets on data.europa.eu.**
   On data.europa.eu, the findability and comparability of HVDs is supported by the recently added HVD filter. However, when HVDs are not tagged as such on national portals, they do not appear in the search results. By checking and tagging HVDs on national portals, Member States can make sure their HVDs appear in the filter and support end users in working with the data.

5. Conclusions

The goal of this pilot study was to do a methodological assessment of the findability and metadata comparability of HVDs among Member States, to spot opportunities for improvement early in the HVD implementation process and to provide recommendations to Member States and HVD users.

Based on the research, it can be concluded that nearly all HVDs were findable on either national portals, the Inspire geoportal or data.europa.eu. In many cases there are differences in vocabulary, metadata structure and accessibility, especially in non-Inspire HVD categories, like the Statistics and Meteorological categories. For the Inspire-related HVDs, the metadata virtually always adhered to the format specified in the directive, making it straightforward for a user to identify the dataset and gather information on its contents.

Fifteen out of the 24 HVDs in the sample set were found on data.europa.eu, containing either a link to the source data or a direct download option. In four cases, the findability of the HVD on data.europa.eu was poor, meaning the dataset would likely be missed by a user specifically looking for it. There are improvements to be made in this regard, for instance through labelling HVDs at the national level and offering a click-through HVD hierarchy on data.europa.eu, where one can select

an HVD category, then a country, then a specific HVD and even sort it by legislation, like on the Inspire geoportal (see Figure 16).

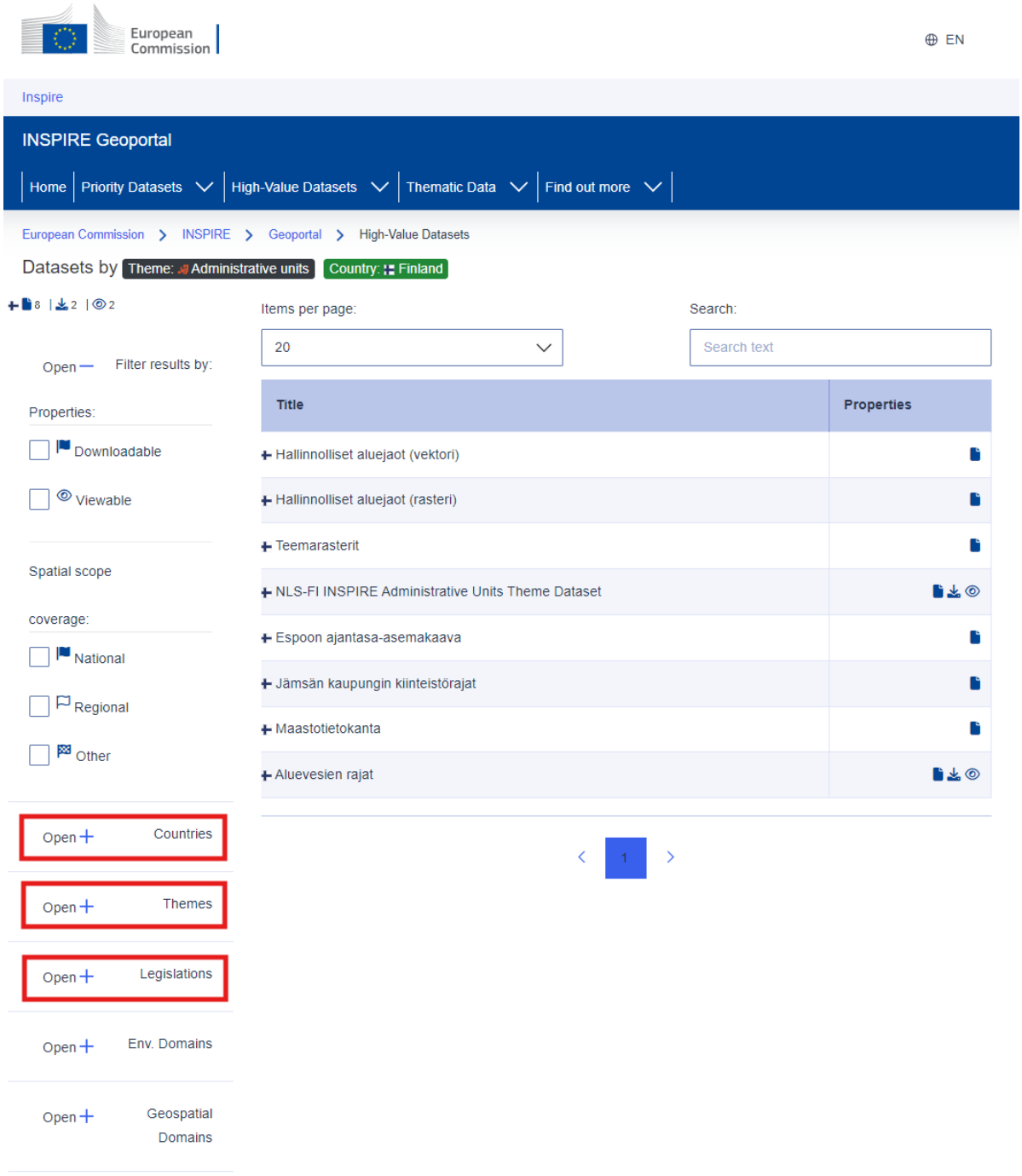

Figure 16: A visual demonstration of the search functionality of the Inspire geoportal for Finland's HVD under the geospatial category (administrative units)

For a user accessing HVDs from various Member States (particularly those outside the Inspire categories), considerable effort is required to search for and verify through the metadata that a dataset is indeed the designated HVD. Member States can help users by referencing the regulation

or the Inspire directive in the metadata of the HVD and using a dataset title that corresponds to the HVD title in the regulation.

Annex I. Sample HVDs used for this study

Below are the findings of the analysis of the 24 HVDs from the sample set, one from each of the six categories for the four selected Member States.

1. Geospatial – 'Administrative units'

Member State	HVD title	Available locations	Licence / Copyright	Latest update (approximately)	Format	Metadata according to regulation (EC) No 1205/2088	API	Bulk download	Metadata language	Clarity on content	Metadata references the regulation	Reference to IR in data.europa.eu metadata	data.europa.eu page containing actual dataset
Denmark	DK Inspire Administrative units	geodata.info.dk Inspire data.europa.eu	Danish Law	1 year ago	GML	Yes	Yes	Yes	English	High	Yes	Yes	Link
Estonia	Administrative and settlement division	geoportaal.maaamet.ee Inspire data.europa.eu	CC-BY 4.0	1 month ago	Esri SHP, MapInfo (TAB), AutoCAD (DXF), MicroStation (DGN)	Yes	Yes	Yes	English	High	No	No	Link
Latvia	Administratīvo teritoriju un teritoriālā dalījuma vienību robežas (Inspire dati)	geometadati.viss.lv Inspire data.europa.eu	No limitations	6 months ago	N/A	Yes	No	No	National language	Low	Yes	Yes	No
Finland	NLS-FI Inspire Administrative Units Theme Dataset	maanmittauslaitos.fi Inspire data.europa.eu	CC-BY 4.0	1 week ago	Geopackage, GML, ESRI shapefile	Yes	Account needed	Account needed	English	High	No	Yes	No

2. Earth observation and environment – 'Environmental monitoring facilities'

Member State	HVD title	Available locations	Licence / Copyright	Latest update (approximately)	Format	Metadata according to regulation (EC) No 1205/ 2088	API	Bulk download	Metadata language	Clarity on contents	Metadata references the regulation	Reference to IR in data.europa.eu metadata	data.europa.eu page containing actual dataset
Denmark	DK Inspire EF Monitoring Stations (MSFD)	geodata.info.dk Inspire data.europa.eu	No limitations	2 years ago	GML	Yes	Yes	Yes	English	Low	Yes	Yes	Link
Estonia	Eesti keskkonnaseire jaamade, alade ja kohtade nimistu (EELIS)	geoportal.ee Inspire data.europa.eu	CC-BY 4.0	2 months ago	GML, CSV	Yes	Yes	Yes	National language	High	Yes	No	Link
Latvia	Annex III. Inspire Dataset for Environmental monitoring Facilities Theme	geometadati.viss.lv Inspire data.europa.eu	No limitations	2 years ago	unknown	Yes	No	No	National language	Low	Yes	Yes	No
Finland	Hydrologiset havaintopaikat (multiple datasets on Inspire)	opendata.fi Inspire	CC-BY 4.0	1 month ago	unknown	Yes	No	No	National language	Low	No	N/A	N/A

3. Meteorological – 'Climate data – Validated observations'

Member State	HVD title	Available locations	Licence / Copyright	Update frequency	Format	API	Bulk download	Metadata language	Clarity on contents	Metadata references the regulation	Reference to IR in data.europa.eu metadata	data.europa.eu page containing actual dataset
Denmark	Inspire – Climate data from DMI	opendatadocs.dmi.govcloud.dk data.europa.eu	CC-BY 4.0	10 minutes	GeoJSON	Account needed	Account needed	English	High	No	No	Link
Estonia	Historical Weather Data	ilmateenistus.ee	No license provided	Yearly	XLSX	No	Yes	English	High	No	N/A	N/A
Latvia	Environmental Data Archive – Meteorology	videscentrs.lvgmc.lv	No license provided	Hourly	XLSX	No	Yes	National language	Low	No	N/A	N/A
Finland	Weather and Sea – Download Observations	ilmatieteenlaitos.fi data.europa.eu	No license provided	10 minutes	XLSX, CSV	Yes	Yes	English	High	No	No	Link

4. Statistics – 'Poverty'

Member State	HVD title	Available locations	Licence / Copyright	Update frequency	Format	API	Bulk download	Metadata language	Clarity on contents	Metadata references the regulation	Reference to IR in data.europa.eu metadata	data.europa.eu page containing actual dataset
Denmark	Income inequality	statbank.dk data.europa.eu	No license provided	Yearly	CSV, XLS, XLSX, PX	No	Yes	English	Low	No	No	Yes
Estonia	Poverty and inequality	andmed.stat.ee	No license provided	Yearly	PX, XML, CSV, XLSX, JSON	Yes	Yes	English	High	Yes	N/A	N/A
Latvia	Poverty and inequality	data.stat.gov.lv data.europa.eu	No license provided	Yearly	CSV, XLS	Yes	Yes	English	High	Yes	No	Yes
Finland	Statistics on living conditions	pxdata.stat.fi	No license provided	Daily	XML, CSV, PX, JSON	Yes	Yes	English	High	Yes	N/A	N/A

5. ‘Companies and company ownership’

Member State	HVD title	Available locations	Licence / Copyright	Update frequency	Format	API	Bulk download	Metadata language	Clarity on contents	Metadata references the regulation	Reference to IR in data.europa.eu metadata	data.europa.eu page containing actual dataset
Denmark	(no free open dataset available)	datacvr.virk.dk	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Estonia	Register of companies	avaandmed.ariregister.rik.ee data.europa.eu	CC Attribution 4.0	Daily	XML, CSV, JSON	Yes	Yes	English	High	No	No	Yes
Latvia	Public free data of the commercial register	data.gov.lv data.europa.eu	CC0 1.0 Universal	Daily	CSV	Yes	Yes	English	High	No	No	Yes
Finland	Company register	avoindata.fi data.europa.eu	CC Attribution 4.0	Monthly	CSV	No	Yes	National language	High	Yes	No	Yes

6. Mobility – 'Transport networks – Road networks'

Member State	HVD title	Available locations	Licence / Copyright	Latest update (approximately)	Format	Metadata according to regulation (EC) No 1205/ 2088	API	Bulk download	Metadata language	Clarity on contents	Metadata references the regulation	Reference to IR in data.europa.eu metadata	data.europa.eu page containing actual dataset
Denmark	DK Inspire Road transport network	du-portal-ui.dataudveksler.app.vd.dk Inspire data.europa.eu	National language	Not found	GML	Yes	Not found	Yes	English	High	Yes	Yes	Yes
Estonia	Eesti teeregister	transpordiamet.ee Inspire data.europa.eu	No limitations to public access	Not found	Unknown	Yes	Yes	Yes	National language	Low	Yes	Yes	Link
Latvia	Valsts ceļu tīkls (INPIRE dati)	geometadati.viss.gov.lv Inspire data.europa.eu	No limitations to public access	Not found	GML	Yes	Not found	Yes	English	Low	Yes	Yes	Yes
Finland	FTIA Inspire Transport Networks Theme Dataset	suomenvaylat.vayla.fi Inspire data.europa.eu	CC Attribution 4.0	Not found	N/A	Yes	Not found	Yes	English	High	Yes	Yes	Link

Publications Office
of the European Union