High-value Datasets Best Practices Report

data.europa.eu
The official portal for European data

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

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:
Martina Ramponi
Maddalena Moro
Jolanta Suárez

Edited by:
Beatriz Fernández Nebreda
Nataliya Rozbroj Jasinskaja

Last update: 12 June 2024
https://data.europa.eu/

Luxembourg: Publications Office of the European Union, 2024

ISBN: 978-92-78-44203-3 DOI:10.2830/113131 OA-02-24-637-EN-N

Contents

Executive summary

This study delves into Commission Implementing Regulation (EU) 2023/138 , which complements the open data directive (Directive (EU) 2019/1024) by setting specific requirements for high-value datasets. These datasets provide significant social and economic benefits, enhancing public sector data accessibility. As the regulation came into force on 9 June 2024, this report aims to highlight best practices and strategies employed by EU Member States (MSs) in implementing these requirements.

A key component of the study's methodology involved conducting **interviews with representatives** from seven MSs ([1]): Czechia, Denmark, Estonia, Italy, the Netherlands, Romania and Finland. These interviews offered insights into their current compliance status, challenges faced and strategies employed in implementing the HVD requirements. The report highlights the best practices and recommendations provided by the mentioned MSs, addressing legal, organisational and technical aspects.

Additionally, the study identifies ongoing common **challenges** across MSs, including data protection concerns and issues with licencing ambiguity. It also highlights the areas where support at the EU level is necessary in areas related to data interoperability, timely access to information and clarification of the overlapping requirements with other legislation, such as the Inspire directive, as well as requirements that may sometimes be contradictory in other regulations, namely the general data protection regulation.

The study also includes a suggested roadmap for MSs, outlining steps to successfully implement the new HVD requirements. It is designed to guide MSs towards compliance with the HVD implementing regulation, promoting a more transparent and data-driven EU landscape.

Through its practical and actionable insights, the report aims to facilitate the implementation of HVD requirements across the EU, leading to improved public sector efficiency, increased economic opportunities through data reuse for all kind of users and strengthened cross-border collaboration within the EU.

The key findings are the following:

- the degree of adoption of the regulation varies among MSs;
- in specific domains, such as cadastral data, business registries and meteorological data, challenges have been identified in making HVDs available free of charge;
- the main obstacles have been encountered in the use of the EU vocabularies, APIs and bulk downloads;
- further guidance is needed with regard to the overlap between the requirements of the regulation and the Inspire directive in relation to the quality of the datasets (description, documentation and data catalogue vocabulary;
- cooperation is encouraged, among MSs to share effective practices and with the European Commission for direction to overcome existing challenges.

([1]) The MSs are listed according to EU protocol order.

1. Introduction

1.1. Context

In an era defined by digital transformation, the European Union has taken significant steps to enhance the potential of data in the public sector through the open data directive (Directive (EU) 2019/1024). This directive marked a significant milestone in promoting transparency and innovation by facilitating access to public sector data, including geographical, statistical, legal and publicly funded research data. Within this directive, the concept of **high-value datasets (HVDs)** emerged, recognising that certain datasets hold substantial social and economic benefits and that their use is particularly valuable in the roll-out of improved digital services.

On 21 December 2022, the European Commission introduced Commission Implementing Regulation (EU) 2023/138, outlining further conditions for the publication and reuse of HVDs. This regulation aims to harmonise the reuse of HVDs across EU Member States (MSs), defining conditions for their publication and outlining concrete datasets falling under the six categories of HVDs: geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, and mobility.

The availability and accessibility of HVDs play an important role in enabling effective decision-making, fostering research and driving innovation in today's interconnected world. An initial overview of the progress MSs have made in implementing the requirements of this new legislation was presented in the *2023 Open Data Maturity Report*.

1.2. Objective of the high-value datasets best practices study

The objective of this study is to assist MSs in the process of implementing the HVDs by presenting a collection of effective best practices that can be adopted to accelerate compliance. It also aims to outline the common challenges that need to be addressed at both national and EU levels to facilitate HVD implementation across the EU. Finally, this study also offers a practical **roadmap** informed by the successful practices in the leading MSs to guide and assist other MSs with compliance.

1.3. Method for data collection

The methodology employed in this study followed a three-step approach. Each step aimed at gathering comprehensive insights into the challenges, compliance needs and successful strategies related to HVD implementation across MSs.

Step 1: Analysis

We began with a content analysis of the implementing regulation on HVDs, involving textual interpretation and desktop research of existing analyses and reviews. The objective of this analysis was to identify key themes, patterns and trends that would inform the development of questionnaires on the status of HVD implementation in countries with potential for growth and in the top-performing MSs.

Step 2: Qualitative interviews and best practices research

Next, we conducted qualitative 1-hour interviews with data representatives from selected MSs, as mentioned above: Czechia, Denmark, Estonia, Italy, the Netherlands, Romania and Finland.

The interview questionnaire focused on the implementation of the requirements outlined in the regulation, exploring challenges faced by all participating countries, successful strategies employed to address them (with a special focus on the top-performing MSs) and concluding with a summary overview of each MS.

Step 3: Actionable recommendations development

Finally, we developed actionable recommendations by synthesising best practices observed in the leading MSs. These recommendations were translated into practical guidelines that can be adapted to different national contexts by other MSs, facilitating the effective implementation of the regulation. Furthermore, the report includes MSs' requests for the Commission related to unresolved issues, hence opening the dialogue for further cooperation and assistance, along with links to resources the Commission has made available to address some of the challenges.

2. Establishing the legislative framework

2.1. Short overview of the varying legislative landscape in the EU

The enforcement of the regulation is influenced by previous EU legislation, notably the open data directive and Inspire directive. These directives have laid the groundwork for HVD implementation by establishing frameworks for the reuse of public sector information and promoting the sharing of geospatial data for environmental policies and initiatives.

Previous legislation paved the way for shaping the framework for HVDs: the public sector information directives established principles for the reuse of public sector information, aligning with the objectives of the HVD regulation to enhance the accessibility and reuse of valuable datasets. Additionally, the general data protection regulation (GDPR) set standards for data protection and privacy, ensuring that HVDs adhere to stringent legal requirements to safeguard individuals' rights and personal data.

2.2. Analysis of the regulation's requirements

To begin with, it is worth noting that unlike the open data directive and Inspire, which require national transposition, the HVD implementing regulation is directly applicable in MSs. Building upon the open data directive, this regulation introduces significant changes to public data management practices encompassing various aspects, including guidance on access and charges, technical formats and standards, and arrangements for data reuse.

This regulation also introduces a specific list of HVDs, categorised into **six thematic groups** regulated in the Annex, which provides the specific description and requirements the HVDs must adhere to. This list includes geospatial data, earth observation and environment, meteorological, statistics, companies and company ownership, and mobility. By delineating these thematic groups, the implementing act provides clarity on the scope and coverage of HVDs, facilitating their identification and publication by MSs.

HVDs are required to be **downloadable in bulk** – where relevant – and through application programming interfaces (**APIs**), while also providing extensive documentation for their metadata (Article 3). In terms of **access and charges**, the regulation offers more detailed guidance compared to the open data directive. While the directive established the general rule that HVDs should be available **free of charge**, the regulation provides further clarification on any permissible **exceptions** to this

rule ([2]), ensuring transparency and consistency in data access policies. Moreover, with a few exceptions, HVDs are characterised by specific technical and legal requirements, including the open data licence, the availability of public documentation and the ensuring of machine readability (Article 4).

Additionally, the regulation clarifies any specific conditions for the reuse of HVDs, ensuring that relevant stakeholders are aware of any restrictions or limitations associated with the reuse of these datasets. By providing clarity on reusability arrangements, the regulation promotes transparency and encourages the maximum utilisation of valuable data resources.

3. High-value data – State of play at a glance

In this section, we present the state of the implementation of HVD requirements in the MSs that were interviewed for this report, as of April 2024. The MSs in this study are listed according to EU protocol order. The summary of the state of play of the participating MSs is presented in **Figure 1**: High-level overview of the implementation of the regulation in the interviewed MSs.

Czechia has successfully identified HVD providers and integrated requirements into national legislation, setting a **strong foundation for regulatory compliance**. Despite challenges slowing down adoption, ongoing efforts in resource management and technical implementation have helped Czechia make fast progress towards HVD compliance.

([2]) Recital 6 of the HVD implementing regulation: 'Pursuant to Directive (EU) 2019/1024, the requirement to make high-value datasets available free of charge shall not apply to libraries, including university libraries, museums and archives. Member States may exempt individual public sector bodies, upon their request and in line with the criteria set out in the Directive from the requirement to make high-value datasets available free of charge for a period not exceeding two years from the date of the entry into force of this Implementing Regulation.'

Denmark is well on track to HVD implementation, capitalising on its **strong digital infrastructure and culture of governmental transparency** to smoothly integrate and manage HVDs. The approach includes tailored data catalogues and direct involvement of authorities in labelling their data, ensuring precise and sustainable metadata management across the national portal.

Estonia is leading in implementing the HVD regulation in Europe. It has efficiently transitioned key datasets like its business registry to fully open models and enhanced others like meteorological data to meet stringent standards. Despite facing challenges with licencing translations and monitoring data reuse, Estonia's **robust digital infrastructure and proactive policy framework** continue to support the effective publication and accessibility of HVDs.

Italy is progressing in implementing the HVD regulation, with a working group developing a **guiding document to streamline practices** across numerous administrative bodies. Despite these efforts, challenges such as privacy concerns in company data and limited resources complicate the widespread adoption of open data standards for high-value data.

The Netherlands' approach to implementing the HVD regulation involves engaging INSPIRE data providers and adopting compliance strategies that minimally impact business models. Although the national strategy lacks specific incentives for HVD publication, it **supports the broader agenda to promote public data as open data**. Challenges include GDPR and HVD conflicts and technical limitations in API and bulk download capabilities.

Romania is putting effort into implementing the HVD regulation but faces challenges due to emerging open data culture, recent government restructuring and resource constraints. Despite the challenges, Romania is working to **develop policies, improve collaboration between agencies and improve data accessibility**, with a need for support from and collaboration with advanced Member States.

Finland leverages its strong open data practices and existing free access to datasets, with a focus on aligning Inspire and HVD requirements through integrated metadata management. Despite challenges in updating cadastral data to meet new standards, Finland's use of **established legal frameworks and active working groups** supports effective adaptation and compliance without significant resource constraints.

data.europa.eu

Figure 1: High-level overview of the implementation of the regulation in the interviewed MSs.

3.1. Czechia

As per the *2023 Open Data Maturity Report*, Czechia is on track for the timely adoption of the regulation. The quick turnaround for the implementation was made possible by an **early delineating of the relevant stakeholders** (data providers and data owners), with **clear responsibilities** in each of the regulation categories. This early engagement with the main stakeholders set the stage for streamlined data handling and resulted in successful integration of the requirements for HVD management into Czechia's national legislation, strengthening the policy framework and expediting its journey towards compliance. This measure has proven beneficial in facilitating the reporting processes to the government, where data providers are required to demonstrate compliance with both national and EU open data directives.

Metadata standards applicable to HVDs have been published and distributed among the Czech data providers, enabling appropriate and timely labelling of data. HVDs in Czechia were published on the national portal and open data catalogue in May 2024 and will be fully accessible to the public by June 2024. The portal will offer an extended set of search and filter features that are in line with EU standards and even go beyond what is required by the regulation to promote the accessibility and reuse of HVDs.

Czechia also highlighted its **successful collaboration** with Slovakia as a contributing factor to its compliance with the regulation. Both countries use the same contractor for their national data catalogues, which has facilitated easier and more efficient resolution of technical issues in both countries.

While Czechia has made significant progress, the journey has not been without **challenges**. Gaps in understanding the detailed content of various datasets, especially those under Inspire categories that are managed by different government authorities, have complicated efforts to enable interoperability at the metadata level. In addition, compliance within the meteorological category poses the biggest challenge due to the cost of handling large amounts of data and a drastic change in the national agency's funding model needing to move from a paid to an open data-sharing model. Furthermore, Czechia underscores the importance of ensuring that EU legislators consistently provide timely and clear communication. This is essential for national coordinators and data providers to adequately prepare for the implementation of the regulation. Finally, there is an ongoing need to enhance clarity regarding reporting formats and requirements, ensuring that all providers have a comprehensive understanding of their compliance obligations.

3.2. Denmark

In Denmark, the process of implementing the regulation has been rather smooth, thanks to its long tradition of digital data management and transparency in the public sector enabled by robust digital infrastructure. A holistic approach to **understanding the big picture and focusing on the value-adding aspects first** helped Denmark to assess existing solutions, map them against the regulation, identify gaps and in doing so create conditions for its successful implementation. This involved a considerable degree of **coordination with the government authorities** to address both technical challenges and their operational needs. Denmark also coordinated with and explored best practices from several Nordic neighbours and Ireland but opted for their own solution for data catalogues rather than replicating existing models in partner countries.

Denmark's strategy has focused on effective coordination and management of data across different governmental departments, ensuring that each entity knows its specific role and responsibilities concerning HVDs. While broader communications have occurred in the form of **a reference group**, the main **interaction with data providers is bilateral**. This works well because the number of authorities is reasonably small in Denmark, and a more targeted collaboration ensures a more efficient resolution of issues that differ based on each authority's unique conditions and technical solutions.

Another integral part of Denmark's strategy concerns direct responsibility of data authorities in marking their data as high-value. This method was chosen over a centralised approach to ensure that data publishers take responsibility for the metadata of their datasets. By making **individual authorities responsible for labelling their data as high-value at the source**, Denmark also ensures that the metadata management remains accurate and up to date on the national portal. Consequently, this minimises discrepancies between the national portal and the source that could potentially confuse users, reduces the burden of metadata maintenance at the central level and establishes a sustainable practice that ensures the portal's long-term value and usefulness.

Finally, Denmark's approach to handling HVDs has been a **phased process**. The focus in the beginning was on understanding the implications of the legislation and identifying which datasets are impacted by it. Currently, the effort is shifting towards ensuring the metadata is accurately attributed, grouping authorities based on how they deliver metadata to the national portal and addressing discrepancies

on a case-by-case basis. Since the start, Denmark's focus has been on data management aspects of HVDs, including documentation and accessibility of datasets, rather than promotion of their use through improved search or filtering functionalities on the national portal's front-end interface. Similarly, monitoring reuse is not among Denmark's top priorities.

In terms of challenges, similarly to other countries, Denmark has encountered difficulties in transitioning from the Inspire requirements to implementing the HVD regulation due to a lack of a common EU-wide solution. Aligning the new regulatory requirements with the pragmatic, demand-driven national approach has also presented challenges. As a standard prevailing practice, Denmark only develops new interfaces and features (such as APIs for bulk downloads) when they are necessary to meet the needs of the community.

3.3. Estonia

Estonia is currently the front runner in the implementation of the HVD regulation in Europe, as per the *2023 Open Data Maturity Report*. Its successful adoption has been made possible by Estonia's proactive and timely measures capitalising on both the country's advanced digital data management capabilities and its digital infrastructure. **Adoption of the Data Catalogue vocabulary Application Profile (DCAT-AP) as a common standard for metadata** in Estonia, with controlled vocabularies being adopted as early as 2018, is a great example of the strategic measures taken to achieve compliance. This has led to one of Estonia's biggest achievements – the rapid transition of the business registry from a paid to a fully open data service in 2020. The last category to be made regulation-compliant was meteorological data at the end of 2023, for which aspects such as granularity required enhancement to meet the new standards.

The legal and policy framework in Estonia is designed to facilitate the publication and accessibility of HVDs, although the portal itself is still pending features that label HVDs as such to promote their reuse. The framework allows for the introduction of new categories of HVDs beyond the categories currently defined by the regulation, such as language data resources.

Estonia's methodical approach involves **regular and detailed discussions within its open data working group**, which now includes nearly 600 members from various government sectors. This group is instrumental in ensuring that EU updates are efficiently communicated and that discussions about data category inclusion or expansion are timely and relevant. Estonia has previously also collaborated with the Netherlands on personal data handling, an important topic to address when opening Estonia's business registry to the public.

While progress in achieving compliance with the regulation is remarkable, Estonia's biggest **challenge** was accessing Creative Commons (CC) licensing terms in Estonian. As there was no translation available at the time, Estonia provided the translation themselves, but getting it validated by the Commission has been a challenge. Monitoring reuse and identifying data (re)users have also posed obstacles, which Estonia has been actively working on overcoming since last year. Good progress has been made in monitoring business registry data use in research.

3.4. Italy

In Italy, most HVDs are already hosted on the national open data portal but are not yet specifically labelled as such. Efforts are under way to make some datasets that were previously behind a paywall freely available from June 2024, with the objective to align Italy's practices with other MSs. Efforts to synchronise metadata formats across portals are pending updates from the Inspire directive's working

group and necessary technical upgrades to the Comprehensive Knowledge Archive Network system to accommodate HVD metadata standards.

Italy's progress in the implementation of the regulation is challenged by the distribution of responsibilities over the different categories of datasets spread out across many public administrations. To coordinate the efforts, Italy has established a **working group** that published a guiding document in December 2023 to streamline the approach towards the implementation of the regulation across the country. The second version of this document is currently under revision. The working group's efforts rest on Italy's information and communications technology plan and the transposition act of the open data directive. They form the core of the legal and policy framework in Italy and outline the need for specific technical guidelines for HVDs.

Another significant challenge for Italy has been **privacy protection** in the category of company data, hampering the sharing of existing data, such as company addresses and accounting data. The main concern is that making company data widely available as open data could potentially distort market competition in Europe if not all countries publish analogous data. Efforts are under way to tackle these issues from both legal and technical perspectives in the revised version of the guiding document ([3]).

Furthermore, Italy faces the issue of **limited resources** for the implementation of the regulation. While there is a general willingness to advance towards compliance, there is also resistance at the regional level fuelled by the scarcity of financial and human resources in places outside of the large metropolitan areas. In addition, technical challenges such as bulk downloads of large volumes of data pose issues for the optimal performance of the platform operations. This has prompted Italy to consider partial dataset downloads to mitigate negative impacts on the portal performance. Finally, the varying degrees of **technical readiness** across ministries, each responsible for datasets under the different categories of the regulation, add complexity to compliance at the national level.

Lastly, Italy also faces challenges with cadastral data, which is fee-based until the end of 2024, as per the current contract with the data owner. Similar to company data, opening cadastral data to the public from 2025 onwards will require a change in the current business model. Italy is exploring alternative revenue streams to reduce the impact on its current administrative and business practices.

3.5. The Netherlands

The Dutch approach towards implementing the regulation began by engaging with Inspire data providers to assess the additional efforts required from them to comply with the regulation, such as adapting APIs and conducting impact analyses. Most data providers selected **compliance strategies** that would have **the least impact** on their operations. This preliminary work included knowledge sharing and the development of guidelines to assist data providers in understanding and implementing HVD requirements.

However, the Dutch national strategy does not currently include specific measures to incentivise the publication of and access to HVDs beyond a general digital agenda that promotes public data as open data. This is partly because the Netherlands is still catching up with the broader open data legislation, and HVD-specific policies are relatively new. Nevertheless, efforts are under way to align HVD

([3]) Recital (8) of the HVD implementing regulation: 'Member States should make use of appropriate methods and techniques (such as generalisation, aggregation, suppression, anonymisation, differential privacy or randomisation), thus making as much data as possible available for re-use.'

implementation with existing digital agendas and sector-specific programmes and to update the governance structure.

Several additional **challenges** are impacting the Netherlands' progress in becoming compliant with the regulation. First, there is a lack of consistency in the technical set-up of APIs and bulk downloads across different data holders. Many organisations are not in favour of putting resources into the implementation of the DCAT-AP 2.0 standard while the release of the DCAT-AP 3.0 standard is forthcoming. Also, not all data is available via APIs, and not all organisations have the technical and financial capacity to support this infrastructure. Furthermore, the HVD regulation might present a potential **conflict with the GDPR** ([4]) and prioritising one can lead to non-compliance with the other. As mentioned above, following the HVD regulation, the focus is on using appropriate methods and techniques to protect personal data. However, in terms of business impacts, there is a concern and hesitation about making HVDs freely available, especially for data providers like the cadastre, which previously relied heavily on fees.

While HVDs are published through the national open data portal, these datasets are not currently identified as such on the portal and are not marked according to the DCAT-AP guidelines for HVDs. As a result, they do not yet align with legislative definitions.

3.6. Romania

In Romania, the implementation of the regulation is still in its early stages and faces several challenges. This can largely be attributed to Romania's emerging open data culture and its initial steps toward developing a national open data strategy. Efforts to map the data providers and identify responsible institutions have commenced but are challenged by recent restructuring at various governmental levels, leading to unclear responsibilities across different agencies.

Despite these difficulties, Romania has established **working groups and crafted a guiding document** for data providers to align with the HVD regulation on legal, technical and semantic aspects. **Inventory models** to streamline the approach across agencies have also been shared with the data owners as a first step in the process, although actual data publication plans are still under development.

Not all government data is aligned with the technical requirements of the regulation, though some datasets from the cadastre and trade registry are being considered for harvesting on the national portal. The granularity of these datasets, however, is not yet in line with the legislation, and the API is unavailable.

Efforts to categorise datasets based on the six groups outlined in the regulation are ongoing to advance compliance with the regulation. In the future, to improve the accessibility of HVDs, Romania has plans to enhance the search functionality on the national portal, which requires a major upgrade. This, unfortunately, is impacted by the funding constraints. An annual reporting system is also being established to monitor and measure data reuse, incorporating standardised documents for institutions to follow.

In addition, resource limitations pose significant issues, in particular those related to technical skills. The lack of infrastructure and skills impedes the ability to make datasets available in the desired data quality and through APIs or bulk downloads. Furthermore, institutions such as the cadastre, which has

([4]) Recital (8) of the HVD implementing regulation: 'Member States should make use of appropriate methods and techniques (such as generalisation, aggregation, suppression, anonymisation, differential privacy or randomisation), thus making as much data as possible available for re-use.'

historically funded as much as 90 % of its operations through data sales, are now forced to look for alternative means of financing, creating internal resistance to the change and resulting in slow progress. Government departments are helping to look for solutions collaboratively, with some data categories like meteorological data moving towards free access following internal impact assessments and exploration of alternative funding models.

3.7. Finland

Similarly to Denmark, the implementation of the HVD regulation in Finland has capitalised on the country's strong foundation in open data practices. Finland's overall compliance with HVD requirements has been relatively straightforward because most of their datasets were already available and free of charge. The primary challenge was aligning the Inspire directive with the HVD regulation, especially in managing metadata that needed to meet both Inspire and HVD requirements.

Finland's solution to the metadata challenge was to incorporate **HVD specifications within the Inspire descriptions to ensure compliance** without duplicating efforts. This approach reflects Finland's broader strategy of utilising existing frameworks and adapting them to new regulations, leveraging a cultural and institutional predisposition towards data openness and transparency. This transparency is also made possible thanks to the country being relatively small both in size and population, which facilitates easier management and dissemination of public data.

Regarding the legal and policy framework, Finland has taken significant steps in recent years to enhance the accessibility and use of government data. In 2022, the government took a crucial decision to open government data further, backed by legal frameworks and an **information management act that reinforces transparency** at the constitutional level. Two active **working groups**, one for data holders and an inter-ministerial coordinating group for legal matters, ensure that practical, technical and legal aspects of HVD implementation are continuously addressed and monitored.

While other countries mentioned resource scarcity, budgetary constraints have not hindered Finland's HVD strategy, thanks to the pre-existing financial allocation and the general government budget for these needs. However, **challenges** persist with cadastral data in some municipalities that require updates to meet the granularity standards of the regulation. Local authorities have until 2029 to comply, a timeline that reflects the degree of difficulty in upgrading certain datasets to the required digital standard.

As for the datasets themselves, apart from company data, which only became available as from 9 June 2024, most adjustments have been minimal. In Finland, a DCAT-AP format is used to ensure interoperability between HVD categories, and a comprehensive catalogue of HVDs is maintained by the coordinating working group. Finland does not actively monitor the reuse of open data, adhering to the policy that the public can use the data as they see fit.

4. Empirical findings

4.1. Best practices

In this section, we present examples of best practices gleaned from MSs, showcasing effective strategies and lessons learned. These examples highlight what has worked well for those countries and provide a valuable context for understanding the successful approaches taken by various MSs regarding the main challenges. A high-level overview of the best practices is provided in **Figure 2**: Best practices.

Figure 2: Best practices.

Policy and legal framework

Encourage a governmental culture that prioritises transparency in data management.

This best practice underscores the fundamental relationship between open data and cultural values ingrained within governmental systems. Across several MSs, such as **Finland** and **Denmark**, transparency towards citizens is not merely a contemporary trend but a deeply rooted tradition reflected in constitutional frameworks. By recognising this cultural foundation, governments can better understand the intrinsic importance of open data initiatives, which not only align with constitutional principles but also serve as a commitment to fostering transparency and accountability in governance. Therefore, efforts to promote open data should be framed within the broader context of cultural values and historical precedents, reinforcing the notion that transparency is not just a policy but a core societal value.

Embrace a wider perspective and adopt a comprehensive approach.

As a result of the ongoing process of implementing the open data directive and the Inspire directive, MSs face challenges in simultaneously complying with the HVD implementing regulation. Some MSs, such as **Denmark** and **Finland**, had pre-existing legislation in place, necessitating only minor adjustments to align with the national transposition of the open data directive. Estonia incorporated HVD provisions into its existing open data legislation, specifically the Public Information Act. As a result, the implementation of previous regulations is nearing completion, facilitating the transition to meeting the new requirements.

For MSs encountering challenges in this process, a strategic approach based on a wider perspective of the regulation, such as focusing on implementation of standards like DCAT-AP in the public sector and publishing the data, is advised. This approach prioritises sustainable solutions over quick fixes, ensuring long-term effectiveness and compliance.

Practice prioritisation and focus on achievable goals.

Prioritising and focusing on achievable goals are crucial, especially when faced with numerous tasks and limited resources. By identifying what is most important and feasible, organisations can streamline their efforts and maximise efficiency. For example, when implementing HVDs, it is prudent to prioritise categories that overlap with existing initiatives, such as Inspire. As **Czechia** and **Denmark** pointed out, since three of the HVD regulation categories also fall under Inspire, concentrating efforts on these areas allows for progress in the implementation of both pieces of legislation simultaneously, optimising resources and expediting progress. This ensures that valuable resources are allocated effectively, leading to a more efficient and timely achievement of compliance objectives.

Governance and processes

Engage in strategic partnerships and foster collaboration with key stakeholders at the national level.

Establishing collaborative networks and partnerships with key stakeholders at the national level is crucial for effective implementation of the HVD regulation. This involves coordinating efforts among various ministries and agencies responsible for different categories of HVDs, especially in MSs with decentralised governance structures.

To address challenges related to the diverse pace of progress across HVD categories, creating interdisciplinary working groups comprising relevant ministries, agencies and stakeholders is

recommended. These working groups can build upon existing structures established for previous directives, ensuring continuity and leveraging prior collaboration experiences.

For instance, **Denmark** has a reference group for the open data directive and for the Inspire directive, where they also address the HVDs. **Finland** has also established two working groups to address the implementation of HVDs. The first group, which is a coordinating group of data holders, was active until 9 June 2024 and focuses on practical and technical questions. This group shares best practices and resolves practical issues. As legal experts, they also provide guidance on understanding the HVD regulation from a legal perspective. The second group is an inter-ministerial coordinating group, a legal working group ensuring there are no conflicts or overlaps between the HVD regulation and the national legislation. Establishing such a network is crucial, as certain MSs currently lack a comprehensive overview of available datasets. Such a network can enable countries to conduct thorough stakeholder analysis, facilitating a comprehensive data inventory and clarifying which agency is responsible for each dataset.

Coordinate with other Member States.

Engaging in regular exchanges with other MSs, facilitated by events like the 2023 data.europa.eu workshop for data providers in Copenhagen, Denmark, enables the sharing of ideas and solutions to common challenges. The ability to collectively address issues not only fosters innovation but also addresses challenges such as the shortage of IT skills, as observed in countries like Czechia, Italy and Romania. Workshops like this provide a platform for the MSs with similar issues or needs to connect with each other and continue bilateral engagements beyond the event. The collaboration between **Czechia** and **Slovakia** is an inspiring success story of effective bilateral cooperation.

Promote sustainability through individual agency responsibility for its respective datasets.

Ensuring the sustainability of the national data portals means making sure that the metadata is maintained with the resources available. In order achieve that, some MSs, like **Czechia** and **Denmark**, entrust each agency with the responsibility of maintaining and managing their own datasets. In this way, the national portal only plays a monitoring role, ensuring that data publishers mark and maintain the data as agreed.

This approach is recommended for all lead authorities that seek a robust and sustainable model for maintaining metadata over time. It is crucial because data portals require regular maintenance to prevent them from becoming outdated quickly and to keep their relevance and value to businesses and society. Individual agencies' responsibility aids in inventory management, as data providers are accountable for updating the inventory when new datasets are published or existing ones updated on the portal.

Promote awareness among stakeholders, data holders and users through knowledge sharing.

Facilitating bilateral communication with data publishers is a crucial aspect of ensuring effective implementation of guidelines, particularly when certain categories may not align perfectly with established standards. For instance, **Estonia** provided awareness training sessions for stakeholders on topics such as open data and privacy.

Similarly, **Czechia** held two webinars dedicated to HVDs. The first was in cooperation with the Ministry of Environment, focusing primarily on the relationship between Inspire and HVDs. Following the publication of the DCAT-AP HVD standard, a second webinar for all data providers and users took place

to disseminate the information received from the Directorate-General for Communications Networks, Content and Technology.

By providing a platform for interactive communication and knowledge-sharing, these webinars serve as valuable tools for ensuring that stakeholders are well informed and equipped to meet their obligations effectively. Moreover, such engagement initiatives foster a sense of community and shared responsibility among stakeholders, contributing to the overall success of data governance efforts.

Use open data to address challenges and find solutions.

Consider using crowdsourcing platforms to involve citizens in solving real-world challenges such as enhancing data quality. Initiatives like **Estonia**'s Kaggle competition invite participation from those with diverse skill sets, offering an opportunity for individuals, regardless of their IT proficiency, to contribute innovative ideas and problem-solving approaches. In addition, leveraging such platforms can help bridge the gap in IT skills, enabling countries like Romania and Czechia to tap into a broader pool of talent and expertise.

Establish a robust framework for data governance.

Develop a comprehensive data governance framework by first assessing available resources, including technical expertise, data management tools and key stakeholder inputs. This assessment process enables a clear understanding of the rules, processes and responsibilities necessary for effective data governance implementation. For example, **Estonia** mapped out existing roles and identified specific actions needed across various domains, from technology to data management. This thorough mapping ensured that all aspects of data governance were addressed, facilitating a more cohesive and efficient implementation.

Technical aspects, metadata quality and new requirements

Prioritise effective data management.

Develop a comprehensive understanding of the specific requirements for HVDs. This involves identifying existing datasets to determining their compliance with the standards outlined for HVDs. For example in **Denmark**, a reference group created an inventory of datasets. The Agency for Digital Government initiated this process, which was then accepted and approved. They meticulously described each dataset, including which authority was responsible and where to harvest it, ensuring no gaps were left. Similarly, **Finland** uses a spreadsheet containing all necessary elements for a catalogue of HVDs, managed by the coordinating working group. By identifying areas where datasets fall short of meeting these requirements, organisations can establish a roadmap to address gaps and ensure full compliance over time. This process provides a systematic basis for improving data quality and accessibility and enhances the overall value of the HVDs for stakeholders and end users.

Enhance the quality of metadata directly at the data source before publishing it on the portal.

As previously mentioned, managing metadata at the source before publication on the portal offers several benefits. Firstly, it places the responsibility for maintenance on data providers rather than on the portal itself, as **Denmark** has successfully demonstrated. Additionally, this approach mitigates the risk of confusion among users caused by discrepancies in metadata between the national portal and the data source.

To enhance the quality of metadata, using the DCAT-AP high-value dataset guidelines, the controlled vocabularies for the six HVD categories and the controlled vocabulary for the subcategories, which is still ongoing, is recommended. This is also in alignment with the Inspire legislation, as the work on GeoDCAT allows Inspire metadata to be mapped to DCAT-AP. Merely relying on the Inspire team to denote metadata may not suffice, as noted by **Italy**, indicating the need for additional measures beyond the DCAT-AP document.

Top-performing MSs are integrating DCAT-AP and Inspire requirements into the metadata practices adapted to their national context. For instance, **Czechia** has developed specifications for local open data catalogues to ensure compatibility with the National Catalogue of Open Data. These specifications adhere to the DCAT-AP 3.0 standard and encompass both SPARQL endpoints and file-based record storage. Similarly, **Denmark** provides guidance within the DCAT framework, outlining essential fields to prioritise for effective metadata description.

Strengthen accessibility of high-value datasets through external funding or strategic planning.

The regulation mandates that all HVDs must be accessible free of charge; some Member States diversify funding sources by seeking financial support from external avenues such as fundraising initiatives and participation in EU projects, as exemplified by **Czechia**. In **Denmark**, the implementation of a basic data program, coupled with strategic planning outlined in the Free Data initiative, has facilitated the transition towards compliance with the regulation, particularly in sectors like meteorology.

Adapt the business model gradually to provide free data.

Monetisation strategies play a crucial role in the adoption of the regulation, particularly concerning company data and cadastral data, where fees are often mandated. A good example is **Italy**, where the current contractual obligations to keep cadastral data behind a paywall are in force until the end of 2024, although historically this has been a common business model across most MSs.

Denmark's experience transforming their meteorological data business model through the initiative called Free Data, one of the main initiatives in **Denmark**'s digital growth strategy, is a noteworthy success story. This transition for the Danish Meteorological Institute was well planned, divided into six phases between 2020 and 2023 to ensure a gradual and effective shift. In the first phase, meteorological observation data for measurements like temperature, rainfall and wind from the Danish Meteorological Institute's network of weather stations in **Denmark** and Greenland was made public. Later in 2020, data on water levels and lightning patterns followed, and since 2021, processed data in the form of climate data, radar data and forecast data has been released free of charge.

Improve the implementation of API and bulk downloads at the source.

The implementation of APIs and the facilitation of bulk downloads present significant challenges across various MSs ([5]). **Czechia** and **Romania**, for instance, struggle due to a scarcity of resources and expertise, resulting in a restricted availability of data through APIs.

Estonia delegates data responsibility to individual agencies, and if they do not want a specific API, there is a native functionality for APIs on the national portal. This solves the issue of implementing APIs

([5]) Digital government APIs: The road to value-added open API-driven services (APIS4DGOV). The Commission's Directorate-General for Communications Networks, Content and Technology together with the Joint Research Centre (JRC) launched this study with the purpose of gaining further understanding of the current use of APIs in digital government and their added value for public services.

on data holders' source portals, often associated with resource- and technical expertise-related challenges.

5. Remaining challenges

Despite significant progress in adopting the regulation, several challenges persist across MSs.

Data protection laws

The GDPR sets stringent requirements for the processing and protection of personal data within the EU. While the HVD regulation aims to promote the availability and reuse of public sector data, MSs must navigate a complex landscape where HVD requirements intersect with GDPR provisions. This intersection poses challenges, particularly in sectors where sensitive or personal data may be involved. For example, ensuring compliance with the GDPR while facilitating the reuse of health or demographic datasets presents a significant challenge for MSs. The HVD regulation proposes solving such a conflict as provided in Recital 8 ([6]); however, resolving these sector-specific challenges requires careful consideration and possibly tailored solutions to balance data accessibility with privacy protection.

For instance, **Estonia** has recognised the issue of commercially sensitive data and stressed the importance of conducting a cost–opportunity assessment to address these concerns effectively.

In **Italy**, a working group involving all public administrations was established to define guiding documents for implementing the regulation with reference to specific data categories and datasets ([7]). One issue they encountered with opening company data, namely company addresses, is that it poses a risk to the infringement of individual privacy. Italy is now working to address these barriers from both technical and legal perspectives.

Meanwhile, in the **Netherlands**, there are ongoing issues with private data protection. Some categories such as health data or company data remain behind paywalls to make sure that the most sensitive data remains publicly inaccessible and to safeguard entities' privacy.

Licence ambiguity

Another persistent challenge is the ambiguity surrounding the types of licences allowed alongside CC licences for HVDs. CC licences are widely used for open data sharing, but the specific conditions and restrictions they impose may vary. Even though CC licences, namely CC 4.0, are not mandatory, this generates confusion as MSs require clarity on whether additional or alternative licences can be used in conjunction with CC licences for HVDs ([8]). This ambiguity complicates data-sharing practices and

([6]) Recital (8) of the HVD implementing regulation: 'Member States should make use of appropriate methods and techniques (such as generalisation, aggregation, suppression, anonymisation, differential privacy or randomisation), thus making as much data as possible available for re-use.'

([7]) See Chapter 3, paragraph 4.

([8]) Recital (12) of the HVD implementing regulation: 'It is the objective of Directive (EU) 2019/1024 to promote the use of standard public licences available online for re-using public sector information. The Commission's Guidelines on recommended standard licences, datasets and charging for the re-use of documents (https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj#ntr5-L_2023019EN.01004301-E0005) identify Creative Commons ('CC') licences as an example of recommended standard public licences. CC licences are developed by a non-profit organisation and have become a leading licensing solution for public sector information, research results and cultural domain material across the world. It is therefore necessary to refer in this Implementing Regulation to the most recent version of the CC licence suite, namely CC 4.0. A licence equivalent to the CC licence suite may include additional arrangements, such as the obligation on the re-user to include updates provided by the data holder and to specify when the data were last updated, as long as they do not restrict the possibilities for re-using the data.'

introduces uncertainty for data users and providers alike. Therefore, the recommendation for MSs is to use CC licences for as many HVDs as possible ([9]).

Denmark is among the MSs impacted by the licencing issues. Although there is a common recommendation in the public sector to use the CC BY licence, a unified decision has not yet been reached. This lack of consensus complicates the establishment of clear and straightforward requirements for data sharing. **Estonia** has also faced substantial challenges in implementing the CC licences due to the absence of EU-approved Estonian translations. In order to support the MSs with this challenge, the European Commission has compiled the EU vocabularies, which includes multilingual controlled vocabularies for licences and HVD categories.

5.1. High-value dataset implementation roadmap

In moving forward, joint efforts are required to address these remaining challenges and advance the implementation of the HVD framework across MSs. Collaboration between MSs and the Commission will be essential in identifying and implementing effective solutions. As an additional aid, a suggested high-level roadmap is provided in **Figure 3:** Suggested HVD implementing regulation's compliance roadmap. and **Figure 4**: Suggested HVD implementing regulation's compliance roadmap explained. to outline specific steps to reach full compliance and successful HVD implementation.

([9]) 'Reuse policy: A study on available reuse implementing instruments and licensing considerations'. The Commission's reuse policy establishes that instruments like open licences can aid in implementing the reuse decision. It evaluates four main instruments for this purpose: the current reuse notice used by the Commission, the CC licencing suite, the open data commons licencing suite and a bespoke licence specifically created and maintained by the Commission.

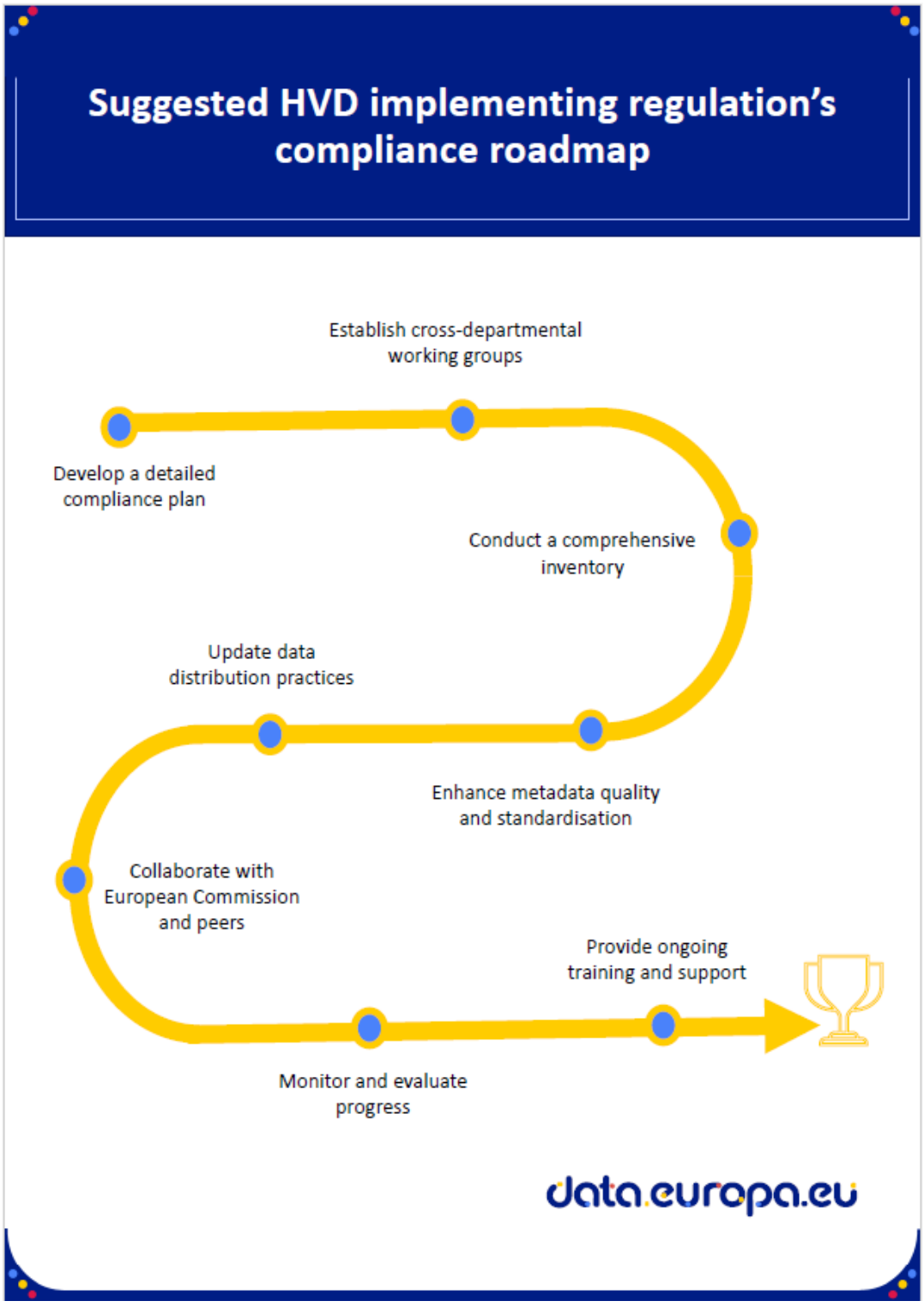

Figure 3: Suggested HVD implementing regulation's compliance roadmap.

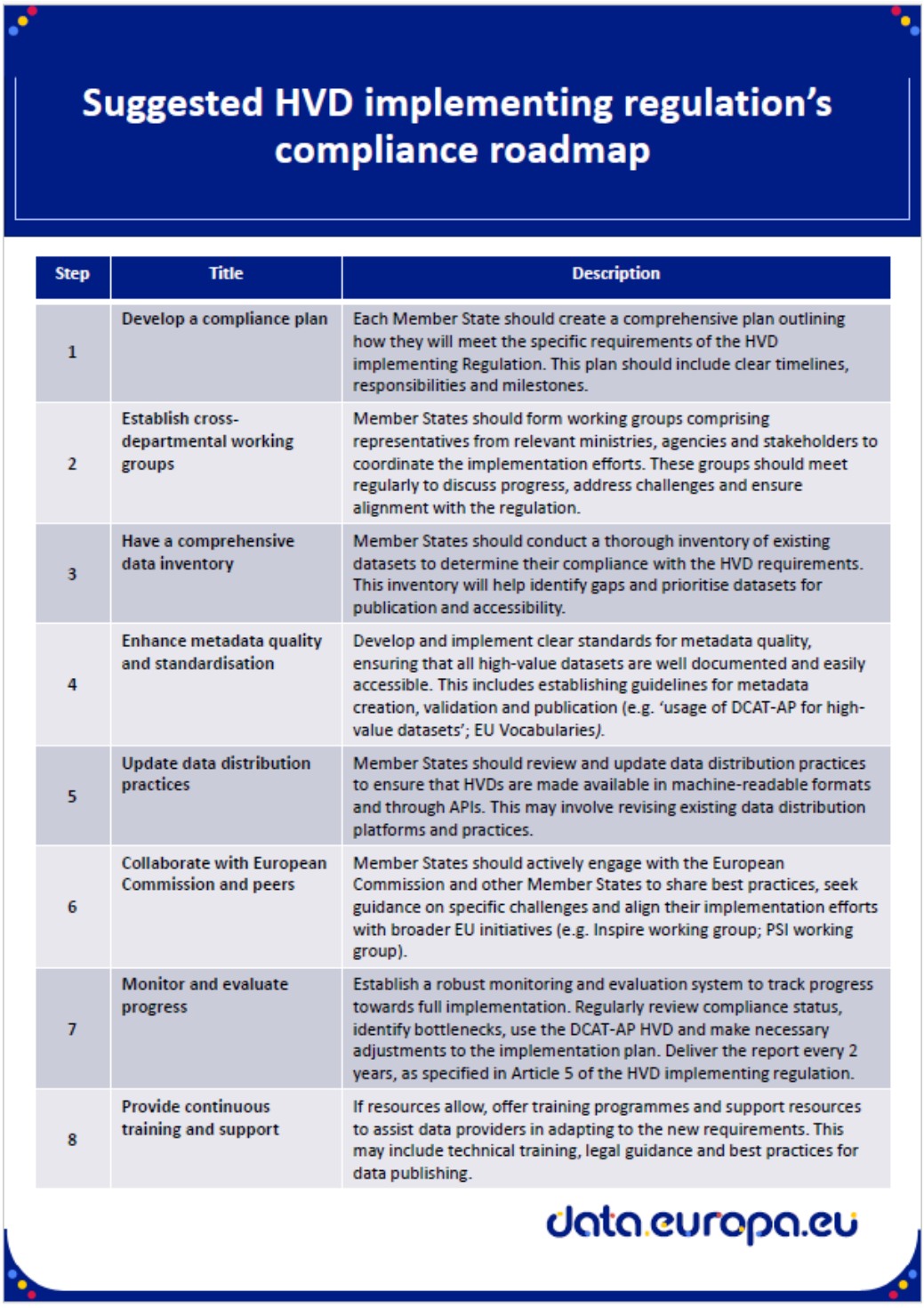

Figure 4: Suggested HVD implementing regulation's compliance roadmap explained.

6. Conclusion

As per the *2023 Open Data Maturity Report*, MSs have made significant progress in implementing the regulation. Through interviews with the selected MSs, this study highlights the progress and challenges and proposes the next steps in implementing the HVD framework across MSs. To fully realise the benefits of HVDs, continued collaboration with and support from the Commission will continue to be essential.

Several challenges persist for all the MSs. First, there is a need for consistent maintenance and availability of **EU standards and vocabularies** across all languages. Standardised terminology and specifications play a crucial role in promoting interoperability and facilitating data exchange across borders. Ensuring that EU standards are consistently updated and accessible in national languages would mitigate barriers to cross-border collaboration and data sharing. As part of the effort to enhance data interoperability and standardisation across the EU, the following guidelines and tools are at MSs' disposition:

- DCAT-AP high-value dataset guidelines. The DCAT-AP HVD guidelines provide instructions for using DCAT-AP to catalogue datasets under the regulation. These guidelines ensure that datasets meet the minimum metadata requirements for mandatory reporting, though they do not guarantee full compliance with all regulatory aspects. The DCAT-AP descriptions aid in assessment and improve metadata quality across European data portals, benefiting citizens and businesses.
- DCAT-AP 3.0. DCAT-AP is a DCAT profile for sharing information about catalogues containing datasets and data services descriptions in Europe, under maintenance by the Semantic Interoperability Community action, Interoperable Europe. It provides a minimal common basis within Europe to share datasets and data services across borders and domains.
- Mapping between Inspire and DCAT-AP – GeoDCAT. GeoDCAT-AP is an extension of the DCAT application profile for data portals in Europe (DCAT-AP) for describing geospatial datasets, dataset series, and services. Its basic use is to make spatial datasets, dataset series, and services searchable on general data portals, thereby making geospatial information more findable across borders and sectors. For this purpose, GeoDCAT-AP provides an RDF vocabulary and the corresponding RDF syntax binding for the union of metadata elements of the core profile of ISO 19115:2003 and those defined in the framework of the Inspire directive of the EU.
- EU vocabularies. It includes multilingual controlled vocabularies for licences and HVD categories.

MSs also stress that it is very important to receive **timely provision of information**. In this sense, the Commission is working on standardising the reporting process for high-value datasets, and this involves the use of SPARQL queries for data retrieval and reporting purposes. Information about this reporting process, including the relevant SPARQL queries, will be made available on data.europa.eu. Also, an official portal for European data, data.europa.eu, is ready to host and display the high-value datasets metadata from national catalogues. These datasets will be displayed on data.europa.eu in a dedicated sub catalogue titled "High-value datasets," allowing users to filter by HVD categories and other metadata across countries. To further assist the MSs, the first webinar with the MSs already took place.

Another significant obstacle for several MSs in complying with the regulation is their **budgetary constraints.** Dedicated policy implementation funding would improve MSs' ability to meet regulatory obligations, invest in necessary infrastructure and resources, and reduce resistance to change at national, regional, and local levels.

Finally, MSs seek clarification on **requirements that overlap** between the INSPIRE directive and the HVD regulation. More specifically, e.g. **Finland** is uncertain about how to express metadata that complies with both pieces of legislation. To address the issue, Finland suggests including a notation in the metadata description to indicate compliance with both INSPIRE and the HVD regulation.

To answer to this request, the European Commission has compiled the following resources:

- Mapping between Inspire and DCAT-AP: GeoDCAT. GeoDCAT-AP is an extension of the DCAT application profile for data portals in Europe (DCAT-AP) for describing geospatial datasets, dataset series, and services. Its basic use is to make spatial datasets, dataset series, and services searchable on general data portals, thereby making geospatial information more findable across borders and sectors. For this purpose, GeoDCAT-AP provides an RDF vocabulary and the corresponding RDF syntax binding for the union of metadata elements of the core profile of ISO 19115:2003 and those defined in the framework of the Inspire directive of the EU.
- List of specific priority datasets under Inspire that are all within the scope of the HVDs. The list contains 89 datasets and was drafted in collaboration with the MSs in the Inspire expert group.
- Inspire Geoportal. It showcases the progress made by each MS on both Inspire priority datasets and HVDs within the scope of Inspire.

Moving forward, collaboration between MSs and the Commission will be crucial in addressing these challenges and advancing the implementation of the HVD framework. The next opportunity for collaboration will be the HVD Best Practices in Europe webinar scheduled on 14 June 2024, aimed at explaining the findings of this study in more detail.

Publications Office
of the European Union

ISBN: 978-92-78-44203-3