Analytical Report n15

Analytical Report 15

High-value datasets: understanding the perspective of data providers.

This study has been prepared by Capgemini Invent as part of the European Data Portal. The European Data Portal is an initiative of the European Commission, implemented with the support of a consortium led by Capgemini Invent, including Intrasoft International, Fraunhofer Fokus, con.terra, Sogeti, 52North, Time.Lex, the Lisbon Council, and the University of Southampton. The Publications Office of the European Union is responsible for contract management of the European Data Portal.

For more information about this paper, please contact:

European Commission
Directorate General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Daniele Rizzi – Policy Officer
Email: daniele.rizzi@ec.europa.eu

European Data Portal
Gianfranco Cecconi, European Data Portal Lead
Email: gianfranco.cecconi@capgemini.com

Written and reviewed by:
Esther Huyer
Email: esther.huyer@capgemini.com
Marit Blank
Email: marit.blank@capgemini.com

Last update: 15.07.2020
www: https://europeandataportal.eu/
@: info@europeandataportal.eu

DISCLAIMER
By the European Commission, Directorate-General of Communications Networks, Content and Technology. 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, which may be made of the information contained therein.

Luxembourg: Publications Office of the European Union, 2020

OA-BF-20-015-EN-N ISBN: 978-92-78-42092-5 ISSN: 2600-0601 doi: 10.2830/363773

Abstract

Making data available as open data across the EU Member States is vital to leverage its potential for the European society and economy, for example, to enrich research, inform decision making, or develop new products and services. The impact of open data is mainly realised through applications and depends on factors like costs, quality of the data and its documentation, or the modality of access. To further increase the impact of open data and reduce market entry barriers for start-ups and SMEs, these factors need to be addressed. To increase the impact effectively, efforts should target those datasets that have the biggest potential for society and the economy.

In the Directive on open data and the re-use of public sector information, the European Commission is tasked to adopt an implementing act specifying high-value datasets (HVDs) " '*High-value datasets' means documents the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets"* [1] Organisations in the scope of the Directive will have to make those datasets available free of charge, in machine-readable format and via APIs, and, where relevant, as a bulk download[2].

The line of thinking developed in this report runs in parallel to what the Commission is currently doing in preparation of the implementing regulation with a list of HVDs in Q1 2021. The findings and recommendations described in the report concern the specification of potential high-value datasets that can serve as input to the Commission's work on identifying the HVDs and the specification of other datasets, in addition to those defined by the Commission, that data providers may decide to focus their efforts on.

Defining the value of specific datasets, however, is very complex and the perspective and role of data providers in the Member States are instrumental. This report reviews relevant literature, political decisions and national initiatives to allow for a deeper understanding of the current status around value assessment of datasets. Findings from interviews with selected open data providers from different Member States provide insights into different perceptions and expectations around HVDs. The findings raise several vital aspects, challenges, and questions, for example:

- The value of datasets depends on the point of view, the specification and (geographical, sectoral) scope of impact. Different opinions about who should benefit from the impact created by high-value datasets are observable.
- Datasets' download statistics are often used, but not sufficient, to assess their value and potential impact. However, there is no clarity on any other standardised base for value assessment.
- Roles and responsibilities in the process of specifying, implementing, and maintaining HVDs are often not clear nor supported by a mandate or designated resources.

[1] Definition of *high-value* dataset: Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", article 2.10.

2 Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", articles 5(8) and 14.

- Reaching understanding and consistency specifying potential high-value datasets in and across the Member States while allowing for differences in local political, cultural, and ethical background.

The report concludes with six key recommendations to follow when identifying HVDs on regional, national, or European level:

1. Create intrinsic and extrinsic incentives, like additional resources, for data providers to enable and foster their active engagement in the process of specifying potential high-value datasets.
2. Set clear expectations around roles, responsibilities, and resources relevant for data providers.
3. Standardise HVDs assessment and specifications across borders.
4. Provide expert guidance that supports a consistent process and is aware of differences in language, culture, politics, perceptions of impact.
5. Work in iterative rounds to allow incremental progress and different stakeholders to reach alignment and mutual consent.
6. Beyond data providers, experts with sector / industry / subject-specific re-user experience must be involved to reach a robust definition of potential HVDs and their specifications.

Table of contents

1 Pushing open data to the next level 7

1.1 Exploring the value and the impact of open data 9

1.2 The economic impact on data providers 10

2 Relevant expertise around high-value datasets 11

2.1 Review of relevant publications and political decrees 11

2.2 Initiatives around high-value datasets 13

3 The perspective of data providers 15

3.1 Different perceptions of value, impact, roles, and responsibilities 15

3.1.1 Value depends on the point of view 15

3.1.2 General focus on download statistics 16

3.1.3 SMEs benefitting from HVDs more than large wealthy companies 17

3.1.4 HVDs across borders 17

3.1.5 Role of the Open Data and PSI Directive 17

3.1.6 Role of central governments 18

3.1.7 Role of data publishers 18

3.1.8 Legal uncertainty 19

3.2 Vital aspects when involving data providers 19

3.2.1 How to reach understanding across and in Member States 19

3.2.2 Scope of the process involving data publishers 19

3.2.3 How to derive meaningful comparable output about value and impact 20

3.2.4 How to represent local needs and constraints 20

4 Conclusion and recommendations 21

4.1 Create intrinsic and extrinsic incentives for data providers 21

4.2 Define and facilitate roles and responsibilities 21

4.3 High-value datasets must be standardised across borders 21

4.4 Guidance, transparency, and consistency is requested 21

4.5 Iterations allow for refinement and alignment 22

4.6 Data providers are vital but not sufficient in the process 22

1 Pushing open data to the next level

Open Data in the EU

Making data available as open data across the EU Member States is vital to leverage its potential for the European society and economy. When made open, the data can be re-used by anyone to create value, for example, enrich research, inform decision making, or develop new applications. Open data can reduce market entry barriers for start-ups and SMEs. Open data-based applications already range from public services to traffic and transport, tourism or health. For example, open meteorological data supports retailers to plan their offers and staffing, open statistical data helps governmental bodies in offering digital public services, and open geodata helps start-ups to innovate mobility apps like Trafikkflyt[3].

In the last years, supported by the PSI Directive[4], a magnitude of open data has been made available[5]. The European Commission, EU Member States, and data providers at local and regional level put substantial effort in selecting, preparing, publishing, and promoting open data. The ever-stronger focus lies on the impact of open data and how to increase it. Improving data and metadata quality is instrumental in this regard.

The impact of open data is mainly realised through application, which is influenced by factors like costs, quality of the data and its documentation, or the modality of access. To further increase the impact of open data, these factors need to be addressed. To do that most effectively, efforts should prioritise datasets that have the biggest potential for society and the economy. That means strategically targeting and improving availability, quality, and access for selected datasets.

High-value datasets in the EU

According to the Directive on open data and the re-use of public sector information, a *high-value* dataset is a dataset that is "associated with important benefits for society, the environment and the economy, in particular, because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets".[6]

Defining the value of specific datasets is very complex. It has different facets and depends on the point of view. Different stakeholders in the open data community already made attempts to define high-value datasets and have their methodologies to evaluate open data quality, maturity, and impact.

In the Open Data and PSI Directive, the European Commission is tasked to adopt an implementing act listing high-value datasets (HVDs) that organisations in the scope of the Open Data Directive will have

[3] https://apps.apple.com/nl/app/trafikkflyt/id559879473

[4] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information

[5] European Data Portal (2019): Open Data Maturity 2019 Report

[6] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", article 2.10.

to make available free of charge, in machine-readable format and via APIs, and, where relevant, as a bulk download[7].

The Open Data and Public Sector Information Directive

The focus on increasing the supply of high-value public data for re-use is one of the substantive changes introduced to the legal text of the latest Public Sector Information and Open Data Directive, so as to fully exploit the potential of public sector information for the European economy and society.[8] The Open Data and PSI Directive prescribes that "for the purpose of ensuring their maximum impact and to facilitate re-use, the high-value datasets should be made available for re-use with minimal legal restrictions and free of charge. (…). The Directive defines a list of six target HVD thematic categories[9]:

- Geospatial
- Earth observation and environment
- Meteorological
- Statistics
- Companies and company ownership
- Mobility

These are intended to support the identification of the HVDs, as they are areas where the important socio-economic benefits and high value for the economy and society should be more easily demonstrable. The Commission has the option to add new categories in the future, in order to reflect technological and market developments.[10] "Legal and administrative documents", for example, could also become a dedicated additional category at some point. Examples of datasets that are expected to be recognised to be of high-value according to the Directive are[11]:

- Postcodes, and national and local maps (Geospatial);
- Energy consumption and satellite images (Earth observation and environment);
- In situ data from instruments and weather forecasts (Meteorological);
- Demographic and economic indicators (Statistics);
- Business registers and registration identifiers (Companies and company ownership); and
- Road signs and inland waterways (Mobility).

[7] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", article 5.8.

[8] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", recital 4.

[9] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", Annex I.

[10] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", article 13.

[11] Examples of datasets expected to be identified as high-value, provided in the directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", recital 66.

Scope and approach of the report

The perspective on HVDs by the data providers in the Member States is vital in the process of defining and implementing HVDs. By means of desk research and qualitative interviews with selected open data providers from different Member States, we look into current initiatives and challenges around HVDs. We will explore what datasets are currently perceived as high-value and why, and what relevant initiatives already exist. We will also explore concerns and demands for the process of specifying HVDs and the implementation of the Open Data and PSI Directive.

1.1 Exploring the value and the impact of open data

In the effort to increase the impact of open data, it makes sense to focus resources on those datasets that have the most value for economy and society. Therefore, the question arises, how to define value and select those datasets that have a higher value than others.

The value of data can be perceived and defined very differently by different stakeholders. The value of a dataset for its providers might differ from the perception of businesses in different industries or from the perspective of citizens. The value of a dataset from a data provider's point of view might be the quality of the data, its metadata, its resolution or granularity or its arrangement for publication and re-use[12]. Data Providers may have several different reasons that contribute to the determination of a value of a given dataset, e.g. that it is already available at good quality.

The value of a dataset, in the HVDs context, is defined by its socio-economic impact, which in turn is very complex to measure or to define. When we look for high value, we expect high impact. However, it is important to understand that impact, i.e. the influence or effect of something on someone or something else can be perceived differently depending on the stakeholder. Often a negative and positive perception can occur at the same time by different stakeholders. In addition, most impact of HVDs is created by its application. The impact of an application, however, also depends on different factors like its user-friendliness, promotion, etc.

If we look, for example, on the impact on competition on a given market, the impact is complex. Open data that is free of charge and available via API might reduce market entry barriers, especially for SME's which, in turn, enriches competition. This can lead to increased innovation, better working conditions, etc. At the same time, increased competition can mean that existing companies have to endure more pressure and might lose their competitive advantage to the point that they have to give up or transform their business and employees could lose their jobs. This again will create multiple impacts. These dynamics make it very complex to assess impacts.

[12] https://kennisopenbaarbestuur.nl/media/254852/maatschappelijke-kosten-batenanalyses-open-data.pdf

1.2 The economic impact on data providers

Reducing or eliminating charges for data provided by governments will have important effects on the overall digital market. Governments and citizens will potentially receive benefits while incurring costs, from an economic, social, and even environmental perspective.

In the short-term, government bodies that have relied on revenues from licensing agreements – typically, the sale of the data – will need to find a new funding model and may prove resistant to the loss of direct funding opportunities. More importantly, data providers may lose the necessary incentives to not only maintain particular data quality standards but also to improve them. Analysis of energy and transportation networks, where network infrastructure has been separated from actual service delivery, have shown that infrastructure can suffer, and that regulation does not always provide the necessary discipline on those responsible for maintaining infrastructure where monetisation is not possible. In the case of high-value datasets, it can mean stagnation for innovation in data collection and improvements in maintenance and quality.

However, the Open Data Directive prescribes that "for the purpose of ensuring their maximum impact and to facilitate re-use, the high-value datasets should be made available for re-use with minimal legal restrictions and free of charge (…). However, this does not preclude public sector bodies from charging for services that they provide in relation to the high-value datasets in their exercise of public authority, in particular certifying the authenticity or veracity of documents."[13] Therefore it will need to be observed how the economic models around HVDs will evolve.

[13] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)", recital 69.

2 Relevant expertise around high-value datasets

The concept of HVDs is not original to the Open Data and PSI Directive. Previous attempts at identifying valuable datasets or data categories and highlighting their importance were made in multiple occasions. This section describes how datasets equivalent in potential to what we now call HVDs were described over the years, taking note of any methodology that was used in the past to identify the datasets and estimate their socio-economic value. The overview also shows vital steps taken by the European Commission, like the Impact Assessment of the Directive 2003/98/ED on the reuse of public sector information (PSI) and Open Data and PSI Directive, that are the base for this Impact Assessment study on the list of high-value datasets to be made available by the Member States under the new Directive.

2.1 Review of relevant publications and political decrees

In **2013**, the G8 (now G7) committed to an Open Data Charter[14]. The Charter sets out the aim to become open by default and to ensure that data is re-usable by all, in order to boost innovation and increase government transparency. Among the collective actions that were specified in the Charter, was releasing "high-value data". In doing so, the Charter attempts to define not only a revised set of 14 domains but also an indicative list of actual datasets. The domains identified by the G8 were: Companies, Crime and Justice, Earth observation, Energy and Environment, Finance and contracts, Geospatial, Global Development, Government Accountability and Democracy, Health, Science and Research, Statistics, Social mobility and welfare, and Transport and Infrastructure. However, no methodology is made explicit in the document by which the list of reference categories and datasets was created. We presume that the document simply collected the personal, qualitative considerations of the individuals who contributed to formulating the Charter.

In **2014**, the G20 emphasised the importance of open data in its anti-corruption action plan, that will further develop into a full set of open data principles[15] by the time of the 2015 gathering.
The principles re-iterate value considerations around the open data used to fight corruption, however only general, qualitative considerations are offered.

Still in **2014**, Omidyar Network commissioned to Lateral Economics the research: "Open for Business: How Open Data Can Help Achieve the G20 Growth Target".[16] The report uses the seven G20's agenda items of 2014 as their "thematic categories" of Finance, Fiscal and Monetary Policy, Anti-corruption, Employment, Energy and Infrastructure. For each category, the value of open data is backed by offering only qualitative considerations, occasionally reinforced by anecdotal information specific to one market in one country, e.g. the size of the market or the number of people it employs.

Also in **2014**, the European Commission published the notice "Guidelines on recommended standard licences, datasets and charging for the reuse of documents".[17] The notice highlighted a general

[14] See https://opendatacharter.net/g8-open-data-charter/ .

[15] See http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf .

[16] See https://lateraleconomics.com.au/wp-content/uploads/omidyar_open_business.pdf .

[17] See https://ec.europa.eu/digital-single-market/en/news/commission-notice-guidelines-recommended-standard-licences-datasets-and-charging-re-use .

"agreement on the need for the speedy release of several high-value datasets" and sets the fundaments for how HVDs will be later defined in the OD and PSI Directive. The document addresses the limitations of the 2006 MEPSIR study[18], including for example scientific data in its scope. It re-defines the thematic categories, to a list of 5 that is almost identical to the list of 6 found in the latest Directive: Geospatial, Earth observation and Environment, Transport (now "Mobility"), Statistics and Companies. Only "Meteorological" is missing. The notice associates "high-value" to data "depending on the circumstances (relevance to strategic goals, market developments, social tendencies, etc.)" in the areas of "innovation and business creation, government transparency and accountability, and improved administrative efficiency".

In **2017**, the Global Open Data Index (GODI) published its open data benchmark. GODI measured the openness of clearly defined data categories. These categories reflect key data that is relevant for civil society at large. The categories have been developed in partnership with domain experts, including organisations championing open data in their respective fields. Each dataset in each category is evaluated using a set of questions that examine the openness of the dataset. GODI reveals that the categories 'government budget' and 'national statistics' are the most open, while 'land ownership' is the least open data category.

In **2017**, the Open Data Barometer[19] analysed open data initiatives and impact in 30 specific governments that have made concrete commitments to champion open data. The study shows that fewer than 1 in 5 datasets are open. This means that – even though these 30 governments are considered to be leaders in the open data field – a vast majority of their datasets remain closed to the public, even though (some of) these datasets might be of high value. The study highlights the importance to engage with groups beyond the open data community to identify and prioritise highly demanded datasets.

In **2018**, the open data maturity report from the European Data Portal has assessed the maturity of national open data portals of EU Member States. This data reveals that the most popular data domain – i.e. the most consulted domain on the national data portals – is Government and Public Sector with 58% of the EU28+ countries mentioning it in their top 5. Following closely are the Population and Social conditions and the Energy and Environment data domain, with respectively (50%) and (46%) of the Member States mentioning it in their top 5 most popular data domains. As a pattern, the most consulted datasets are from domains that are of broad public interest, such as public spending and procurement, mobility, social-economic numbers, in particular housing and environment data.

In **2018** the European Commission launched the impact assessment of the Directive 2003/98/ED on the reuse of public sector information (PSI) to enhance the positive impact of the Directive. This impact assessment includes an evaluation of the implementation of the current version of the Directive and sets out policy options needed to address four problem areas: dynamic data, charging, scope of the Directive, and lock-in of public sector data. The preferred policy option includes the creation of a list of fundamental high-value datasets that should be freely available in all Member States. For a limited

[18] MEPSIR (2006) available at: http://ec.europa.eu/newsroom/document.cfm?doc_id=1197
[19] Open Data Barometer (2019) available at: https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf

number of fundamental high-value datasets, there will be a hard obligation to make them available through APIs.

In **2019**, the recast Open Data and PSI Directive was adopted by the EU's legislators. Its components that describe HVDs develop organically from the aforementioned 2014 EC notice, making many of its recommendations into law. The thematic categories are slightly renamed and integrate a new "Meteorological" category. The Directive's recitals highlight a perspective of value that extends across "re-users, end-users and society in general and (…) the public sector body itself".[20] They stress how value is dependent on ease, speed and versatility of access, that is "particularly important for dynamic data (including environmental, traffic, satellite, meteorological and sensor-generated data), the economic value of which depends on the immediate availability of the information and of regular updates"[21] and to be delivered through APIs were suitable. Finally, an important point is made about the value of data interoperability across the Member States, as "an Union-wide list of datasets with a particular potential to generate socioeconomic benefits together with harmonised re-use conditions constitutes an important enabler of cross-border data applications and services."[22] The Directive requires HVDs to be the first target of the effort to make this a reality.

2.2 Initiatives around high-value datasets

The topic of high-value datasets is gaining more and more attention in the Member States and beyond. Multiple initiatives have been raised to identify high-value datasets already. These initiatives often involve different stakeholders, such as the business community, the public, civil society, and the research community and are involved in order to gain a better understanding of the demand side of high-value datasets. Examples of initiatives are hackathons and roundtables, but many governments also encourage citizens to submit their data requests to public sector bodies, so that they are aware of data demand and can prioritise accordingly.

Denmark issued a digital strategy that requires government at all levels to publish high-value data, called "basic data". The high-value is recognised in how data enables public authorities to perform their tasks properly and efficiently across units, administrations and sectors. Denmark also removed the legal restrictions on the distribution of address data to third parties, as well as the fee for distribution. There are no restrictions placed on the use or redistribution of the data beyond those required to satisfy the requirements of the law, particularly in regard to personal data or product marketing. A budget of EUR 1.3 million was set aside to compensate municipalities for loss of income from sales of data for the three years after the agreement was reached.

Danish Enterprise and Construction Authority (DECA), calculated that the municipalities would realise savings from no longer having to negotiate data purchase agreements, deliver data or enforce licences. The direct financial benefits of the open address data for the period 2005-2009 totalled EUR 62 million for Denmark, through especially improved government back-end capabilities and more efficient service delivery. Another benefit is improved response accuracy for the emergency services within Denmark and increased data reuse especially by small and medium enterprises, and increased

[20] Recital 14.

[21] Recital 31.

[22] Recital 68.

economic activity, market dynamism, innovation, employment and efficiency, with little impact on the cost.

Slovakia stated that all documents related to the public procurement including receipts and contracts must be published openly. [23] These reforms had a considerable positive effect on corruption, the perception of corruption, and on the business climate. The 2014 Corruption Perception Index published by Transparency International stated that Slovakia increased its ranking by six places, to 54. This represented a jump of 12 places since 2011, making Slovakia one of the most improving countries over that period.

An example of involving multiple stakeholders in the process of identifying high-value datasets comes from **Switzerland**, where the Working Group on Digitalisation under the Ministry of Industry and Trade has identified more than 70 datasets which are necessary to boost re-use of open data. Entrepreneurs from the automobile industry and telecommunication were involved in the process of identification.

Spain issued the law "on transparency, access to public information and good governance"[24] that makes the publication of "information of legal relevance" by government authorities mandatory. The Spanish Federation of Municipalities and Provinces also specifies a list of 40 datasets to be published (as a minimum) on portals by local entities.[25]

France adopted the "Loi pour une République Numérique"[26] in 2017. The law is very ambitious and attempts to fulfil a two-fold purpose, to: "give France a head start in the digital field by promoting an open data and knowledge policy" and to "adopt a progressive digital approach, based on individuals, to strengthen their power to act and their rights in the digital world". To do so, the law is organised around three points: the circulation of data and knowledge, the protection of individuals in the digital society and access to the digital by all. Among the other things, it lists 9 datasets of reference data that become mandatory for the public sector to publish.

[23] http://odimpact.org/files/case-study-slovakia.pdf

[24] Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno

[25] http://femp.femp.es/files/3580-1617-fichero/Gu%C3%ADa%20Datos%20Abiertos.pdf .

[26] Décret n° 2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence, Article R321-5

3 The perspective of data providers

To test our findings from the literature review and gain a deeper understanding of the perception of data providers in the Member States around HVDs, we conducted a series of workshops. In those workshops, we did not only aim at understanding the current status but also the view on roles and responsibilities in the process of defining and providing HVDs. Moreover, we captured concerns and expectations and learned about vital aspects to be included or further emphasised in any methodological approach of deriving and reporting evaluations of the value of datasets.

Participants who contributed to our research in the workshops are[27]:

- Michal Kuban, national open data coordinator at the Ministry of Interior, and Jakub Klímek, linked and open data expert at the Ministry of Interior and researcher and assistant professor at Charles University in Prague, **Czech Republic**.
- Mika Honkanen, head of opendata.fi (national open data portal) at the Finnish Digital Agency (Difi), and Anssi Ahlberg, product manager at the Population Register Centre, **Finland**.
- Marcel Hopman, programme manager for the Dutch Government Data Agenda, Ministry of the Interior and Kingdom Relations (BZK), **the Netherlands**.
- Erwin van Mierlo, open data coordinator, and Leen Roosendaal, sector director at the Dutch national statistical office Statistics Netherlands (CBS), **the Netherlands**.
- Mercè Fígols i Puigbò, responsible for the Open Data BCN at the Municipal Data Office of the Barcelona City Council (MDO), and Maria Jesús Calvo, head of Statistics and Data Dissemination at the MDO, **Spain**.

The following section presents the findings from the interviews. First, we present the findings related to participants' perspectives and initiatives around HVDs, such as how value is defined, concerns, expectations, and different roles and responsibilities. Second, we present the findings related to the expectations and concerns around the process of selecting and specifying HVDs.

3.1 Different perceptions of value, impact, roles, and responsibilities

3.1.1 Value depends on the point of view

Defining the value of specific datasets is not easy and depends on point of view. From a city perspective, the Municipal Data Office (MDO) of Barcelona, emphasise that the value of data depends on the impact that is created on a local level. The participants from MDO mention that HVDs should be aligned with the challenges that the city and citizens are facing and with the policies and strategies of the city. Datasets that are re-used to address or even solve these challenges have a great impact on local level and are perceived to be of high value. The number of downloads is an important indicator

[27] The selection of research participants follows a purposive sample. It seeks a high level of variety rather than quantity, aiming at building knowledge based on qualitative data, rather than achieving population representativeness. In this report we assume that countries that already have a focus on HVDs can provide deeper insight compared to countries that did not explore the concept yet. Within this group, a geographically diverse sample is the second parameter in the selection.

the city uses to assess local re-use and impact of available data. Currently, the Open Data BCN from the MDO has identified and published 40 HVDs.

The importance of creating impact on a local level when defining value was also observed in our workshop with the Statistics Netherlands (CBS). Here it was mentioned that local and regional data, such as data on neighbourhoods and districts, create a high impact on Dutch citizens and businesses and are therefore perceived to be of high value. The participants mention that the more local information the datasets contain, the more it is downloaded. At the same time, the CBS gathers and aggregates data by orders of national and European policymakers and does this according to European standards, which enables interoperability and harmonisation of datasets, and influences national and European policies. From this perspective, CBS perceives data not only of high value when creating an impact on local level, but also on a national and European level. CBS advised the Dutch national open data portal in grouping datasets on the portal to increase discoverability and findability of datasets.

The more local and social perspective – creating an impact on citizens and city challenges – was less present in the workshop with the Czech participants; a more technical perspective to define value was observed. According to the participants, the value of datasets highly depends on the quality and technical characteristics of datasets itself, for example, the use of standard formats, the lack of errors, and the frequency of updates. Additionally, they state that it is important to create regulation that ensures harmonisation of data across borders. Harmonisation of data across borders could increase the quality and utility of data and create a unified European data market. Efforts to identify and select HVDs in the Czech Republic have already been taken. A national working party was established, consisting of a group of experts from different ministries. Re-users are involved in the process by an HVDs “wish list” in which they can express their data demands.

The importance of harmonisation and cross-border interoperability of data was also observed in our workshop with the Finnish participants. The workshop made clear that one of the most expected impacts of HVDs is the opportunity for SMEs and start-ups to expand their businesses across borders and enter new markets. Efforts around HVDs have also been taken in Finland, where amongst other things, a workshop was conducted to identify HVDs. Companies were involved in this workshop. The workshop made clear that most datasets that companies identified as high value, were already available as open data. However, for some datasets, there is a charge to receive more granularity and details.

3.1.2 General focus on download statistics

The workshops indicated that there is a general focus on the number of downloads when evaluating the impact and value of data. This means that the starting point of defining HVDs is to look at existing open datasets, then analysing the download statistics of these datasets, and then specifying the datasets with a high number of downloads as HVDs. When merely looking at these numbers, value can only be measured for datasets that are already published but miss the evaluation of datasets that are not published yet. Additionally, it might miss datasets that are not downloaded because of poor findability or quality, but with better findability or quality would be impactful. Although download numbers are an important proxy to value, high download numbers do not necessarily mean high

impact, since it is unclear if the downloaded datasets are used to create products and services, and moreover, if these products and services create an impact.

3.1.3 SMEs benefitting from HVDs more than large wealthy companies

There is general consent amongst participants about the importance of ensuring that the benefits of HVDs reach indeed the intended targets. Typically, these are Small and Medium-sized Enterprises (SMEs) in Europe that can use HVDs to create new products and services and enter new markets, due to, among other things, the obligations to provide HVDs free of charge. The concern is that HVDs may instead go to strengthen the market position of larger and wealthier companies – possibly even outside of Europe – that already had the necessary financial resources to acquire the same data. Participants, therefore, state that the re-use of open data and in particular the re-use of HVDs should be closely monitored and measured. This allows gaining insights into who creates value from HVDs and to what extent SMEs in Europe benefit from this value.

3.1.4 HVDs across borders

Another convention amongst the participants is that HVDs offer the opportunity for SMEs to scale up their businesses across borders and create a high impact. The workshops made clear that interoperability of HVDs across borders is perceived as a prerequisite for SMEs who want to upscale their business and expand across borders. For example, if geospatial data such as maps and road networks was available across countries in standard and interoperable formats and with an equivalent level of quality and detail, it would be easier for a start-up to build car navigation systems that would work internationally and could compete with the traditional market leads in the industry. Participants emphasise that standardisation of data and metadata across borders is needed to increase interoperability and enable local, national, and European application of HVDs.

Beyond interoperability, a challenge unfolded in selecting and defining HVDs across borders since some data might be recognised to be of high value in one country, but not in others, e.g. because of different strategic priorities. For example, meteorological data describing increasing sea levels could be of high value in the Netherlands since it highly impacts citizens and businesses, due to more than 1/3 of the country lying below sea level. However, the same data will be of less relevance to countries that are less affected by the phenomenon, such as Italy or Greece. To create an impact in the Netherlands, cross-border meteorological data is desired, but the data might not be perceived as high value in other countries due to the above exemplified lack of impact in these countries. The challenge in defining and selecting a list of HVDs on a European level is the question if data can truly be defined as high value when impact is only created in one or a few countries.

3.1.5 Role of the Open Data and PSI Directive

The Open Data and PSI Directive is a legislative act that sets out goals that all Member States must achieve, while it does not provide specific obligations on how to reach these goals. It is up to the Member States to transpose the Directive into national law in order to make the objectives,

requirements and deadlines directly applicable. However, the Commission will adopt an implementing act laying down a list of specific high-value datasets that Member State will be obliged to publish. The Open Data and PSI Directive itself is by some participants received as too vague, multi-interpretable, and not strict enough. The workshop also made clear that the Directive is not ambitious enough in the view of most participants. The concern raised is that the Directive will only unlock the value of HVDs in entrepreneurial countries that already have a mature open data field but does not incentivise the less mature countries to increase their efforts. This is a concern to the participants because it could damage the cross-border application of HVDs. Some participants mention that it is also up to the more mature countries to help less mature countries. Another concern raised is that the Directive might lead to improvements on existing open datasets (e.g. make them available through APIs or in a machine-readable format) but will not lead to the opening of new ones. This concern relates to the before-mentioned focus on merely looking at the number of visitors and downloads of existing open datasets when selecting and defining HVDs.

3.1.6 Role of central governments

A general attitude was observed in participants in wanting their central governments to be more ambitious and active to unlock the full potential of HVD. An example of this is the participants' scepticism about how the Open Data and PSI Directive requisite of machine readability for HVDs will be translated into the national laws. Participants state that central governments need to expand on what the Directive establishes, by providing, among other things, legal and technical guidance, a clear role division and description of stakeholders involved, and financial support to compensate public bodies for the loss of revenue and to support them in their efforts of making HVDs available according to specific requirements. Central governments should also connect data publishers, coordinate initiatives around HVDs, and function as an intermediary for discussions on EU level. In the Netherlands, where the central government has a more holistic approach to data sharing and perceives open data just as a component of the opportunities arising from data sharing in general, our participant mentioned that the central government should not only provide legal and technical frameworks for open data but for the whole spectrum of data sharing options, from closed to open.

3.1.7 Role of data publishers

Participants highlight that data publishers need to be involved as subject matter experts when it comes to identifying HVDs, since they are the most likely players to own expert knowledge about the respective area or topic. Data publishers are also relatively close to the re-users, i.e. they can observe how data is being re-used and observe demand. This enables them to share insights into trends, developments, needs, and demand with policy or decision-makers on national and European level.

At the same time, participants make clear that many data publishers will need pressure from a national or European level to provide HVDs free of charge and according to specific requirements. The general concern is that the obligations around HVDs are perceived by data publishers as extra work, that requires extra financial and human resources. Without additional support or strict obligations, it seems to not be in their own favour to increase their efforts around HVDs. They will need to understand the benefits of HVDs, and these benefits need to be concrete and tangible.

3.1.8 Legal uncertainty

A general concern raised by the participants is the current legal uncertainty. Data providers are currently scared to breach regulations around confidentiality, privacy, and personal data. This uncertainty leads to cautious behaviour towards publishing HVDs. Concerns about the impact on data privacy might keep impactful datasets closed or hinder their potential by reducing their level of detail, whether their concerns are substantiated or not. Participants emphasise the importance of clear legal frameworks in order to reduce uncertainty. Combined with the lack of additional support and the lack of understanding the benefits of HVDs, this leads to a negative balance in publishing HVDs. Clear communication is needed to address legal concerns.

3.2 Vital aspects when involving data providers

3.2.1 How to reach understanding across and in Member States

General consent among our research participants is evident when it comes to the importance of involving open data teams and data publishers at all levels in government in the Member States. However, different opinions and concerns exist around the process (how) and the content (what) of deriving and capturing insights on the evaluation of the value of datasets.

The main concern is how to reach understanding in and across the Member States. A well-prepared basis to guide and capture all discussion is suggested. However, different languages and the use of different terms that differ even when translated need to be addressed. Any document or template that will be used to capture comparable results can be understood very differently by different parties. A form of guidance, examples and support will need to explain, govern and steer the process without influencing the results in a biased way. In addition, assessment metrics will often be qualitative and “low”, “medium”, “high” needs to be defined to not be misunderstood or misused.

3.2.2 Scope of the process involving data publishers

If the task at hand is understood by all parties, another question that was raised, is the scope in terms of content and timing. A process like the evaluation of the value of open data involving a vast ecosystem of stakeholders would need to be iterative to allow for discussion, alignment and mutual understanding. However, resources are limited, and the aforementioned iterative processes would be on top of the daily business of most of the participants. It stays to be decided how much time will be needed and would be feasible without a clear mandate and additional resources to start and complete a meaningful process with all relevant parties. As documented in chapter 3.1 roles and responsibilities need to be clarified for the process to come and therefore the question of whom to involve will be part of this discussion.

This will also determine the scope of the aspired output. Among the parties the level of maturity of the output might differ and our workshop participants express different opinions and expectations around the scope and the goal of the desired output.

3.2.3 How to derive meaningful comparable output about value and impact

This challenge of **how** to do it links to the thoughts of the research participants around **what** to do. **What** will be the basis to reach meaningful comparable insights in the process? Participants express a lack of clarity around what kind of value and impact to assess and on which basis. While we observe the number of downloads as one of the few quantifiable indicators that can be a base for any evaluation, there is an awareness that downloads statistics are not sufficient to assess value.

However, if not quantifiable, the assessment of value is not clear, neither is whom the impact should benefit. The point was raised by the Open Data Portal Barcelona to prioritise citizens and SMEs over big companies when targeting impact (see 3.1.3). This specific focus would lead to different results than a focus on, for example, the next generation of re-users, or large companies, or any other choice of beneficiaries.

In addition, multiple impacts can be created by specific datasets, however, it is unclear if the benefits should apply to Europe as a whole or if an impact created by the national population and the national economy is sufficient to declare high value. The target group is just one aspect of the impact that needs to be aligned on. Lack of clarity and concern about the completeness and the weighing of different aspects is visible.

3.2.4 How to represent local needs and constraints

In line with the need for completeness, transparency and fairness, the local background should not be underrepresented, participants emphasise. Finding broad consent can only happen by taking into account local differences and not underrating or ignoring them. Existing legislation and potential alternatives to opening specific datasets should be given room, participants argue. Aspects around the political feasibility or cultural sensitivity of opening selected datasets should be indicated early on.

Overall it gets clear that different levels of enthusiasm paired with scepticism exist. To address the potential of HVDs that is clearly perceived and expressed by most participants, clear communication, expectation management and a highly individual procedure aiming for international comparability is the challenge to meet before assessing the theoretical value of datasets that will not be realised. Meaningful and feasible results across Europe can only be achieved by highly individual yet comparable procedures. The scope, time and effort should be managed and not understated.

4 Conclusion and recommendations

4.1 Create intrinsic and extrinsic incentives for data providers

HVDs can support economic growth and innovation and provide important benefits for both the public and private sector. However, HVDs are by some parties seen exclusively as a legal obligation to comply with rather than an opportunity. Low awareness and understanding of the benefits of HVDs in combination with efforts and additional resources required to define and publish HVDs, damage the extent to which data publishers are committed to unlocking the full potential of HVDs. Central governments, therefore, need to be more active and ambitious around HVDs, increasing their efforts in providing legal and technical frameworks as well as tangible and concrete benefits. Barriers as legal uncertainty and a lack of additional resources need to be tackled.

4.2 Define and facilitate roles and responsibilities

Defining, selecting, and publishing HVDs is a complex task that involves multiple parties on different levels. Important parties are for example central governments of the EU Member States and data publishers at local, regional, and national level. A clear division of roles and responsibilities is needed to clarify expectations and to eliminate the uncertainty. Clear mandates need to be given in order to understand the different duties of the different parties involved and to ensure that they have the authority to make any necessary decision. Adequate resource planning is needed to allocate tasks and budget to parties involved. Another necessary condition to decrease the current uncertainty in roles and responsibilities around HVDs is clear and transparent communication.

4.3 High-value datasets must be standardised across borders

In order to unlock the full potential of HVDs, standardisation of data and metadata across borders is needed. Standardisation enables interoperability and thereby local, national, and European application of HVDs. Although this is not a requirement by the current Open Data and PSI Directive, it is strongly advisable. In order for SMEs to expand their businesses across borders, they need to be able to re-use the same data across countries in standard and interoperable formats and with an equivalent level of quality and detail. Additionally, with European SMEs as one of the most important targeted beneficiaries of HVDs, monitoring and measurement frameworks need to be developed in order to gain insights into to what extent the benefits indeed reach SMEs.

4.4 Guidance, transparency, and consistency is requested

To ensure that the process and the outcome of deriving and documenting the evaluation around HVDs is comparable and as complete and transparent as possible, a standard process with supporting tools and templates should be the ground for the sessions. It should provide a thorough explanation and examples and supports individual aspects and background information. Moreover, it should provide multilingual guidance and be complemented by the work of facilitators that allow for contextualisation's, but also guide the process, fencing the train of thought towards impact in terms

of *for whom, where and how*, as discussed above. In addition, metrics and qualitative ratings should be defined and standardised, for example, what is called “medium” and what is called “high” impact in specific sectors.

4.5 Iterations allow for refinement and alignment

It seems inevitable that an iterative process that gradually refine results is necessary to achieve a robust and effective HVDs specification. This may start as a conceptual “inventory” of datasets and move to be a detailed specification, describing individual characteristics like the interval of updates or the granularity and detail. An impact check that discusses the expected feasibility and the expected impact (for re-users and for the data providers) if the selected datasets and specifications would be implemented as high-value datasets would help to find consent. Iteration enables taking into account local and national characteristics while ensuring that there is space for alignment and mutual consent. Thus, in light of the European Culture, the outcome in form of the implementation of the Open Data and PSI Directive can cater for all countries and stakeholder without diluting the strength and impact of the HVDs, but rather building common ground, a platform to build digital services all can benefit from.

4.6 Data providers are vital but not sufficient in the process

Data providers in the Member States are a vital part in selecting and implementing HVDs and, most importantly, can assess the feasibility of what is demanded by markets and citizens. Their insights, concerns and “buy-in” are vital for the overall process, just as understanding the national political direction around HVDs.

However, open data providers’ perspective is not sufficient in understanding where the “high value” lays. E.g. data providers might consider datasets valuable because of how often they are downloaded, or because they are already available in a very good quality, but this would neglect to exploit the value of datasets that have a very high potential but are not re-used because they are not yet published, difficult to find and download or of insufficient quality. The correlation -download equals value - might be misleading if taken alone, without the input from the demand side of businesses and citizens. Moreover, sector / industry / subject-specific re-user experience is instrumental to a robust definition of relevant HVDs, knowing that “high-value is in application” -something participants repeated often - and only expert re-users have the necessary depth of understanding to describe which data is needed for which application.

This also highlights the difficulty to assess the actual value in terms of impact, because many criteria usually used to assess datasets are of very limited use here or might even be misleading. A dataset, that has very good quality, is easily finable and very usable might still not create an impact because it is not re-used often, or the impact created with its re-use is insignificant. This explains why it makes sense to focus resources on improving the usability of the datasets that are **expected** to have the highest impact. However, this requires a high degree of speculation to determine what the value of data could be once its specifications improve—every type of use case can alter the calculation and increase the number of data points required to make a proper assessment.