



Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive

Executive Summary

Deloitte.

Internal identification

Contract number: SMART 2019/0025

Project number: 2020.5820

EUROPEAN COMMISSION

Directorate-General for Communications Networks, Content and Technology

Directorate G — Data

Unit G.1 — Data Policy & Innovation

Contact: cnect-g1@ec.europa.eu

European Commission

B-1049 Brussels

Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive

Executive summary

***EUROPE DIRECT is a service to help you find answers
to your questions about the European Union***

Freephone number (*):
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

**LEG
AL
NOTI
CE**
This

document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. The Commission does not guarantee the accuracy of the data included in this study. More information on the European Union is available on the Internet (<http://www.europa.eu>).

EN PDF

ISBN 978-92-76-26948-9

doi: 10.2759/464564

KK-03-20-816-EN-N

The European Commission is not liable for any consequence stemming from the reuse of this publication.

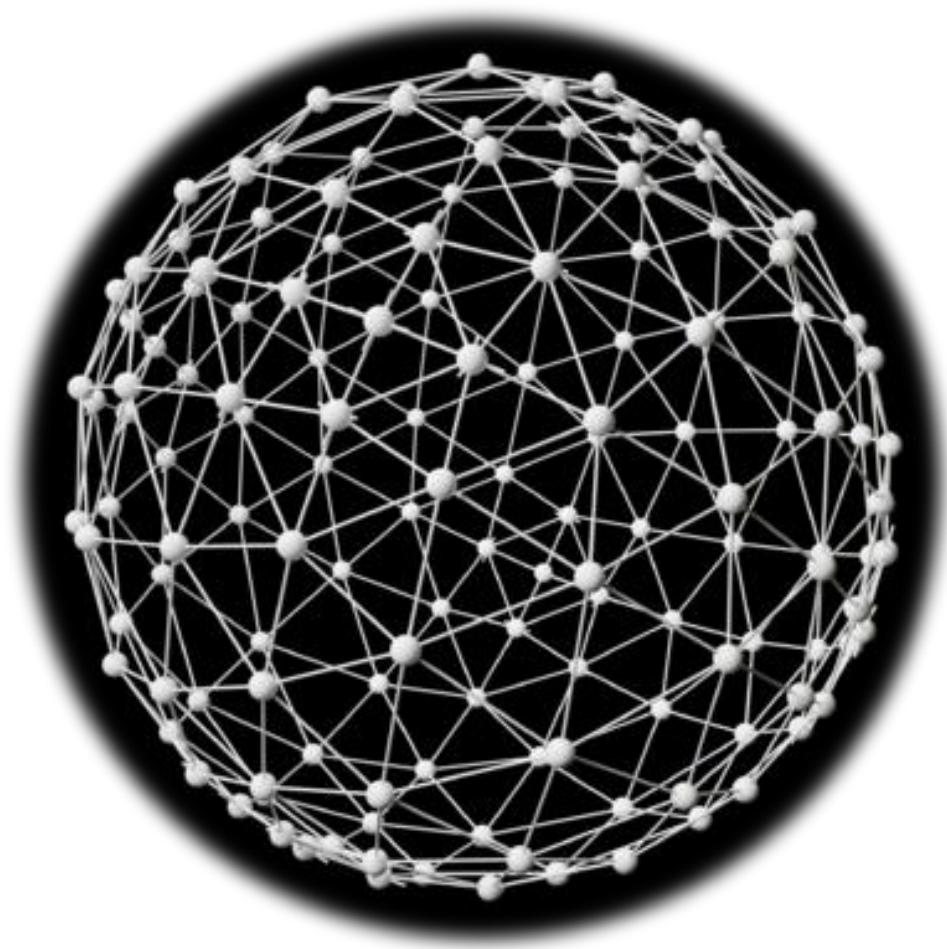
Luxembourg: Publications Office of the European Union, 2020

© European Union, 2020



The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.



SMART 2019/0025 – Executive Summary

Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive

Context of the impact assessment study

The first version of the [Public Sector Information \(PSI\) Directive](#) was adopted in 2003 and aimed to facilitate the re-use of PSI throughout the Union by harmonising the basic conditions for making PSI available to re-users, to foster Community-wide products and services based on PSI, and to avoid distortions of competition. Appreciating the further potential due to rapid technological changes and being confronted with the inherent limitations of the original version of the Directive, the Commission launched several revision initiatives over the years. The most recent version of the PSI Directive, [Directive 2019/1024](#) was adopted in June 2019 and includes the definition of “high-value datasets” (HVDs) to be provided free of charge, in a machine-readable format and through APIs and bulk-download as relevant.

The objective of the study is to support the impact assessment process of the Commission underpinning the adoption of implementing acts defining such HVDs, which shall be identified in six thematic areas: (1) Geospatial; (2) Earth Observation and Environment; (3) Meteorological; (4) Statistical; (5) Companies and company ownership; and, (6) Mobility. Within the remits of each thematic area’s scope, the study analyses from micro to macro-level the impacts of a potential inclusion of given datasets as high-value datasets under the revised PSI Directive for the EU27.

In view of establishing a methodological framework for the assessment of the value of potential HVDs considered, a literature review was conducted to collate categories of value related to the six macro characteristics of potential value derived from open data and described in the Directive: economic benefits; environmental benefits; generation of innovative services and innovation; re-use; and the improving, strengthening, and supporting of public authorities in carrying out their missions. These macro-characteristics were further defined through indicators, against which the different datasets in scope for each thematic area were assessed. Combining this approach with the assessment of potential costs to be borne by data providers at micro-level, it was possible to establish two recommended policy intervention options for each thematic area: a lower and a higher intensity intervention option encompassing the most realistic and value-adding combinations of datasets to be considered as HVDs under the revised PSI Directive. An analysis of the impacts of these options, including different combinations thereof, was then performed at macro-level in order to determine the impacts of the policy measures on the EU economy and SMEs, government revenues and employment.

Company and company ownership

The analysis of EU level legislation allows to identify a number of data fields, which all Member States should possess and are required to be made public, although not necessarily in an open data format: Basic information on company as per the [Company Law Directive](#), Company documents and accounts as per the [Accounting Directive](#), Company ownership information as per the [Anti-Money Laundering](#) and the [Transparency Directive](#) and Company insolvency status as per the [Regulation on insolvency proceedings](#).

Company, insolvency and beneficial ownership datasets are unanimously considered of high value by the literature and the reusers and many use cases have already been mapped for this type of information. From a pure cost-benefit perspective and when looking at how to improve the EU data economy, all four datasets in scope of this analysis should be considered as high-value datasets under the PSI Directive. In fact, the economic and societal benefits of such a policy choice would exceed the costs of implementation for the Member States and would bring great benefits to the data economy at the EU level. Nonetheless, our analysis also highlights that costs of implementation would be particularly concentrated on a number of countries which would then have a lower cost-benefit ratio than the others and would be more strongly affected from the changes.

The lower intensity policy option only includes basic information, company documents and accounts, and non-personal data related to company ownership. The data fields within these datasets are normally held and provided by the same data holders, hence, investments and loss of revenue would concern one

institution per country and not several. In terms of modalities for publication, this lower intervention option proposes the minimum measures to ensure the reusability of the datasets. For the higher intensity intervention, the categories of datasets on company ownership (personal data) and company insolvency status would also be included. In terms of the measures for publication, the higher intensity intervention builds on the measures previously explained for the lower intensity option, adding further requirements in terms of data formats, metadata, shared vocabularies, taxonomies and granularity.

Geospatial

Two main pieces of EU legislation allow to scope the geospatial thematic area: the [INSPIRE Directive](#) and the [Regulation on the alignment of reporting obligations in the field of legislation related to the environment](#). While no single data set or theme within scope stands out as most important, or as singularly valuable for re-use on its own it appears that Administrative units, Geographical names, Addresses, Buildings and Cadastral parcels are of particular relevance.

There is wide consensus that geospatial data are those that offer the widest set of opportunities for reuse, namely because of their combinability with other datasets. The number of use cases is almost infinite, and the recent Covid-19 crisis only served to reinforce the strategic importance of wide availability of high quality georeferenced data to track the epidemics and control the reopening. For both interventions mentioned hereafter, the loss of revenues by data providers should be regarded as the main cost driver for Member States. These costs are however seen as being outweighed by the positive impacts of opening up the data.

The lower intensity intervention implies that only limited requirements of data and minimum changes of the current available publication options will apply for datasets considered as HVDs. When it comes to the geospatial thematic area, the number of datasets in scope as HVDs are five to four, including Administrative units, Place Names, Addresses, Buildings, Cadastral Parcels. The lower-intensity intervention proposes a set of minimum adjustments' measures to ensure the reusability of the geospatial datasets in scope. The higher intensity intervention is more a far-reaching intervention in terms of the measures for publication. The higher intensity intervention builds on the lower intensity measures notably by adding requirements on granularity and timeliness.

Meteorological data

There is one main EU level piece of legislation regulating access to data in this area and this is the [INSPIRE Directive](#), which defines meteorological geographical features as one of the 34 spatial data themes needed for environmental applications and to be made available by public sector organisations. The EU legislator has only regulated spatial data sets and services within this theme. The list of HVD which was suggested for further analysis includes Observations data, Climate data, Radar data, Weather alert messages, Numerical weather prediction (NWP) model output and Lightning strike data, as this data has value to a wide variety of sectors. When looking at these data from the perspective of the HVD framework developed, it suggests that they are relevant for all the macro categories of value as they have a strong potential to bring economic, social and environmental benefits, a strong potential for reuse, they allow generation of new and innovative applications and they benefit the public sector as well.

A cost-benefit analysis conducted suggests that all datasets taken into scope should be considered high-value datasets under the PSI Directive. Existing research and cases consistently point out that the economic and societal benefits of such a policy choice would exceed the costs of implementation for the Member States and would bring great benefits to the data economy at the EU level, over time resulting in additional tax revenue outpacing the costs of provisioning.

The lower intensity intervention proposed encompasses three datasets: observations, climate data, which consists of validated observations (which may contain corrections from the original measurements or the

removal of anomalies etc, and are then entered into the permanent climate record) and digitised structured historical climate data, and weather alerts. Dealing with loss of revenue is unavoidable even at these lower intensity interventions. 13 MS currently charge for observations data. Timeliness, completeness and highest temporal/geographic resolution available drive the value of re-use. For observations reducing the number of variables made available (e.g. limited to temperature, humidity and precipitation) or the timeliness and/or temporal resolution (e.g. hourly averages, instead of 5-10 minute intervals), would immediately limit re-use value. The higher intensity intervention build on the lower intensity interventions by extending the scope of data sets involved. First, digitised heretofore unstructured historical climate data are added, introducing potential data transformation costs. Second, radar data is added. Not all MS operate radar stations from within the public sector. Third, Numerical Weather Prediction (NWP) model data is added. Data from radar stations can become voluminous, and modelling data is very voluminous, though scientific data formats are in use that can help keep such volumes down.

Earth observation and environment

The key EU legislations regulating the EO and environment thematic area are [the INSPIRE Directive](#), the [Directive public access to environmental information](#), [the Copernicus Regulation](#), [the Commission Delegated Regulation concerning GMES](#) and [the Horizon 2020 Regulation](#). As the scope of this study concerns data within scope of the PSI Directive which is held by a majority of Member States, the themes that emerged for Earth observation concern in-situ and earth based remote sensing data, concerning the state of the planet in physical, chemical or biological terms, and space based remote sensing is deemed out of scope.

Earth observation and environmental data already play an important role, but there is also much unused potential benefit still. That potential benefit does not reside in specific data sets as such, but in making the full breadth and depth of earth observation and environmental data available for re-use. It is therefore suggested to place all listed INSPIRE themes and the Environmental e-reporting priority data list within scope of the HVD requirements of the PSI Directive. That adds a layer of additional benefit on top of the existing value of, and existing infrastructure for INSPIRE data provision. Sourcing earth observation and environmental data is currently seen as high-friction and high-cost and thus the primary barrier to re-use value creation (e.g. sourcing EU wide land parcel and land use data for Copernicus services), which adding the HVD requirements to these INSPIRE themes would do away with.

The lower intensity interventions limit the scope of data involved to datasets used for reporting by Member States under the environmental acquis, as documented in the [EEA's reporting obligation database](#) and further detailed in the [INSPIRE priority data set list for environmental reporting](#). In terms of obligations, the lower intensity interventions add an open data obligation to the environmental reporting and observation data, which means mostly removing re-use restrictions and terms of use and adding minimum data provision measures. For the higher intensity interventions the scope of the data involved is broadened by adding the additional INSPIRE themes relevant to earth observation and environment to the data and themes covered in the lower intensity interventions (Environmental e-reporting priority data list, and the INSPIRE Theme EMF). This allows the type of combinations that re-use value is generally build on within this thematic area, and that were found across the varied use cases encountered in the study.

Statistics

The statistical production is often governed by the first principle of the [Fundamental Principles of Official Statistics](#) (adopted by United Nations Statistics Division) that states "Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation." At European level the statistical coherence and quality is ensure through the European Statistics Code of

Practice. It is a self-regulatory instrument and is based on 16 Principles covering the institutional environment, statistical processes and statistical outputs. Other important documents that oversee the production of official statistics and putting in place the general development guidelines are: [Regulation on European statistics](#), [Quality Assurance Framework of the European Statistical System](#), [the Commission Recommendation on reference metadata for the European Statistical System](#), [the General Data Protection Regulation \(GDPR\)](#).

According to both re-users and data holders, all official statistics should be considered as high value datasets under the PSI Directive. The three categories of datasets selected – social statistics, macroeconomic statistics and business statistics – provide relevant information of the social and economic environments within countries. In addition, majority of Member States already provide these datasets free of charge, free to use and re-use for both commercial and non-commercial purposes and in a machine-readable format, making the impact of the current recommendations remain on low-cost impact.

The lower intensity intervention implies that a set of minimum changes to the current publications options available will apply for datasets considered as HVDs. This policy option requires minimum changes to the current provisions existing for publication of statistics in the EU Member States. Due to the relatively high degree of standardisation and harmonisation of datasets publication, the current provisions have low to no impact on the National Statistical Institutes from the EU Member States. This lower intensity intervention proposes a set of adjustments' measures to ensure the reusability of the statistics' datasets in scope of this option. The higher intensity intervention, which adds new changes, in terms of measures for publication, builds on the measures of the lower intensity option. These changes target areas that might increase the burden to comply with for some of the organisation, as they will need more effort in order to fulfil the requirements related to metadata provision, the availability of APIs, data linking and the use of taxonomies and shared vocabularies.

Mobility

The following pieces of EU law should be highlighted for scoping the mobility thematic area: the [ITS Directive](#), the [Rail Interoperability Directive](#), the [VTMIS Directive](#), the [RIS Directive](#), the [ANS Regulation](#) and [EATMN Regulation](#) and the [INSPIRE Directive](#). It was concluded that information falling under the scope of the ITS Directive, with the exception of the datasets also covered by the INSPIRE Directive, would be considered out of scope in order not to interfere with ongoing initiatives related to the implementation of the ITS Directive. In addition, aviation related data prescribed by the ANS and EATMN Regulations have restricted access regimes, and thus also out of scope. The datasets which can be considered as potential high-value datasets for the purposes of this study are: transport networks data under the INSPIRE Directive, and Inland waterway and river infrastructure-related data covered by the RIS Directive.

From a cost-benefit perspective, considering inland waterway and river infrastructure data (including static data, dynamic/urgent data and ENC's) and transport networks data (including features for road, rail, waterborne, air and cable transport as per INSPIRE) as HVDs under the PSI Directive can be considered as having a positive impact on the state of the EU data economy. Overall, such policy development would contribute to reaping a number of economic, environmental and social benefits throughout the EU, while imposing a bearable amount of costs on Member States. The main added-value is the formalisation of an 'Open Data' principle in addition to the existing regulatory *aquis*.

The lower intensity intervention includes inland waterway and river infrastructure data and transport network datasets as per the INSPIRE Directive as HVDs. Starting with the INSPIRE datasets would allow to restrict the initial efforts for Member States, who would in essence "only" have to cater for infrastructural investments related to the APIs (when not already available), as these datasets are already provided in

compliance with INSPIRE data specifications which have been leveraged as part of the recommended measures for publication. The higher intensity intervention would include *all* national transport network datasets on top of the lower intensity datasets. It appears most Member States possess and publish many more datasets relating to transport networks aside from the INSPIRE-tagged ones, which could equally gain value from being considered as HVDs, and thus being provided with open licences, free of charge, in harmonised machine readable formats through APIs and bulk download as per the recommended modalities for transport networks data.

Conclusion

The assessment of impacts on the overall economy and society can only be performed at an aggregated level, by creating policy intervention packages composed of one policy option per thematic area. Based on the multi-criteria analysis performed, taking into account the interdependences between the policy options three policy packages were identified.

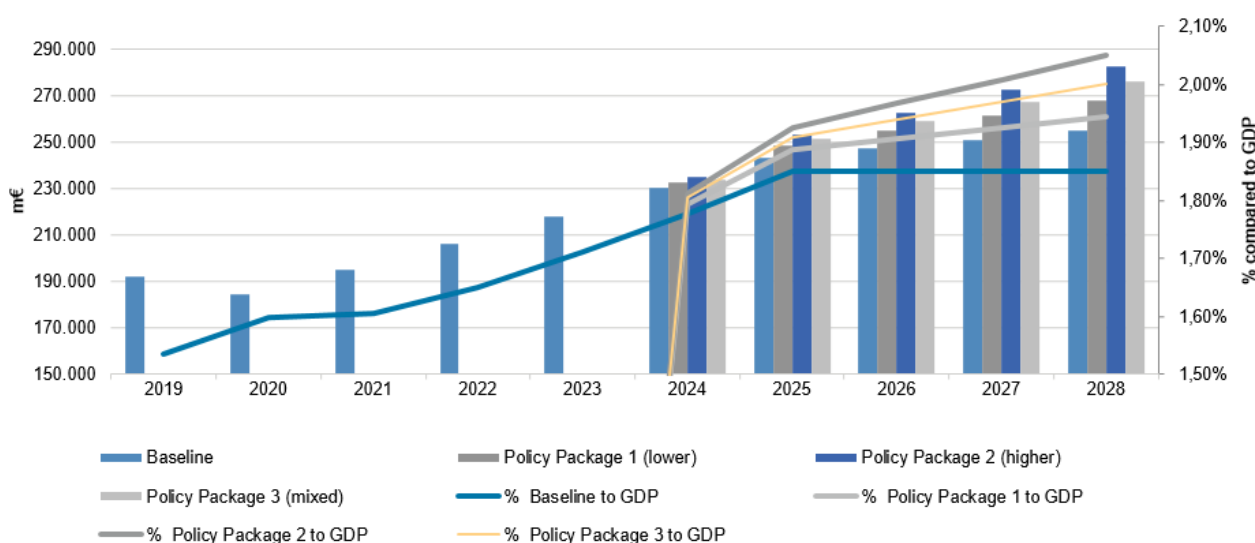
Policy Packages composition

Policy intervention option	Company & Company Ownership	Geo-spatial	Meteorological Data	Earth observation & environment	Statistics	Mobility
Package 1	Lower	Lower	Lower	Lower	Lower	Lower
Package 2	Higher	Higher	Higher	Higher	Higher	Higher
Package 3	Lower	Lower	Higher	Higher	Higher	Lower

All policy packages (lower, higher, mixed intervention) are creating positive impacts at the macroeconomic level by positively impacting the value of the economy for the EU27 from a projected autonomous growth of the direct and indirect impacts from an estimated 184 bEUR in 2020 to a projected 255 bEUR in 2028 (1.60% to 1.85% of GDP).

Economic Value of PSI/HVD per policy Package in m€ and compared to EU27 GDP

Impact on the Economic Value of the PSI/HVD Policy Packages compared to GDP



The policy package 2 (higher intervention Policy Option in all six thematic areas) creates the most impact on the total economic value of the economy. The value of the PSI/HVD related economy is estimated to

grow to bEUR 282 in 2028, compared to bEUR 255 in the baseline scenario. Compared to GDP, the ratio expected to increase from 1.85 % to 2.05 % in the year 2028.

This result is logical, as with increased scope, increased re-use is expected. It will affect more stakeholders which further create more economic value. However, the efficiency is slightly lower with this policy package as higher costs are expected.

For the mixed policy package 3, which is in accordance with the results of the multi-criteria analysis in chapter 4.1.2, the value of the PSI/HVD related economy is estimated to grow to bEUR 276 in 2028. The ratio to GDP in 2028 is expected to increase to 2.00 % compared to 1.85 % in the baseline scenario. The total value of the economy in 2028 with the mixed policy package is bEUR 276 represents an incremental impact of bEUR 21 in comparison to the baseline. This value exceeds the effect of implementing the lower policy package 1 (bEUR 268) by bEUR 8.

The total effect is composed of a direct and an indirect forward effect. In 2028, the indirect forward effect of implementing the mixed policy package is bEUR 201. This signifies an increase of bEUR 16 in comparison to the baseline (bEUR 185). The direct effect of the mixed policy package is bEUR 75, representing an incremental impact of bEUR 6 as compared to the baseline in 2028. In addition to the analyses above, which estimates the economic value of the PSI related economy and its relation to GDP for the three policy packages, the incremental impact of the policy packages was also estimated.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.



Publications Office

doi: 10.2759/464564
ISBN 978-92-76-26948-9