



Brussels, 21.12.2022
SWD(2022) 433 final

COMMISSION STAFF WORKING DOCUMENT
EXECUTIVE SUMMARY OF THE IMPACT ASSESSMENT REPORT

Accompanying the document

Commission Implementing Regulation

**laying down a list of specific high-value datasets and the arrangements for their
publication and re-use**

{C(2022) 9562 final} - {SEC(2022) 450 final} - {SWD(2022) 431 final} -
{SWD(2022) 432 final}

Executive Summary Sheet
Impact assessment on a Commission Implementing Regulation setting out a list of high value datasets and the arrangements for their publication and reuse
A. Need for action
What is the problem and why is it a problem at EU level?
<p>The public sector produces vast amounts of data (e.g. meteorological data, digital maps, statistics and legal information) that is a valuable resource for the digital economy. It is used to produce data-based services and applications, deliver private and public services more efficiently, and enable better-informed decision-making.</p> <p>The main problem is that this data – public sector information (PSI) – is not being reused, at least not to the extent that would allow it to fulfil its socio-economic potential. This is caused by two sets of obstacles. The first is a market access barrier resulting from the application of reuse charges. The second set of obstacles is low interoperability (incompatible licensing terms and low technical/semantic data standardisation) and poor machine readability of data (including access through application programming interfaces – APIs).</p> <p>The problem is common to all public sector data but is most acute for high value datasets (HVDs) in six areas: company data and company ownership, geospatial, earth observation and environmental, meteorological, mobility, and statistics. These datasets are in high demand, but reuse is lower than it should be due to persisting barriers. The problem affects all reusers: non-commercial (non-governmental organisations (NGOs), researchers and the public) and commercial – mostly start-ups and small businesses without resources to invest upfront in data acquisition.</p>
What should be achieved?
The main policy objective is to increase public sector information reuse to benefit the economy (commercial reuse) and society (non-commercial reuse, e.g. for environmental or transparency purposes).
What is the added value of EU action (subsidiarity)?
<p>Only a few Member States have introduced national measures to set out and open up the most important data categories (e.g. reference data in Denmark and Czechia, public interest data in France). However, such measures are not aimed to facilitate using data across the EU. Furthermore, the open data maturity index of the European Data Portal shows that disparities in PSI reuse in the EU still exist and that EU action is the most appropriate way to reduce them.</p> <p>The adoption of the Implementing Regulation is a legal obligation under the Open Data Directive. The Directive’s evaluation in 2018 confirmed that it brought added value to improving the use of public data across the EU. Setting out the list of HVDs and the arrangements for their publication and reuse are the next step in the long-standing EU open data policy that aims to harmonise and progressively open up public sector data. This aim is supported by Member States.</p>
B. Solutions
What are the various options to achieve the objectives? Is there a preferred option or not? If not, why?

The scope of the solution is largely set out in Article 14 of the Open Data Directive: data have to be selected from six themes and they have to be free of charge, machine readable and available through APIs. Therefore, the Commission has designed options around two other aspects: the scope of the datasets identified and the requirements for their publication and reuse.

Options of higher and lower intervention intensity for each of the six themes have been analysed and consulted with stakeholders, and their costs and benefits have been assessed. The preferred option is a mixed-intensity one based on the result of a multi-criteria analysis. It optimises the data scope and publication method for each theme. It has a lower-intensity approach for company data, geospatial data and mobility data, along with a higher-intensity approach for statistical data, environmental data and meteorological data.

What are different stakeholders' views? Who supports which option?

For all open data initiatives, reusers (small businesses creating data services, researchers, NGOs) are usually very enthusiastic while the public sector (data holders) is less so. However, this does depend on the data theme (e.g. statistical offices are very supportive despite being data holders).

All reusers favour a higher-intensity approach. Public sector bodies worry about potential costs (notably one-off costs for setting up APIs). The preferred mixed-intensity option takes both positions into account.

C. Impact of the preferred option

What are the benefits of the preferred option (if any, otherwise of main ones)?

For the preferred mixed option, the value of the PSI/HVD economy is estimated to grow to EUR 276 billion in 2028. The ratio to GDP in 2028 is expected to increase to 2.00% compared to 1.85% if no action is taken (the baseline). This means the preferred option has an incremental impact of EUR 21 billion in comparison to the baseline. It is also EUR 8 billion more than the value of implementing the lower intensity option (EUR 268 billion).

The total impact has direct and indirect effects. In 2028, the indirect forward effect of implementing the preferred option would be EUR 201 billion. This is EUR 16 billion more than the baseline (EUR 185 billion). The direct effect of the preferred option is EUR 75 billion, representing an incremental impact of EUR 6 billion compared to the baseline in 2028.

The figures above are estimates because it is difficult to quantify particular benefits corresponding to one dataset – their value is often better expressed in combination with other datasets. Using the framework developed with the impact assessment study, the datasets were associated with a set of indicators from different macroeconomic areas, along with desk research results and input from different stakeholders.

In addition to the estimated economic impact, the initiative should have social and environmental benefits. More information about companies has clear social benefits for, among others, fighting crime (including financial crime), increased public engagement and government accountability (e.g. publicly supporting businesses). Reusing environmental, spatial mobility and meteorological data is expected to contribute to, among others, researching climate change, reducing urban air pollution and improving transport infrastructure. The measures to manage the COVID-19 pandemic have also shown the importance of geolocating hospitals, relevant infrastructure and people.

What are the costs of the preferred option (if any, besides the main ones)?

The costs are different for the data themes. Assessing the costs is also complex for several reasons. It is difficult to separate HVD costs from reuse costs of other datasets because some IT infrastructure cannot be used for HVDs alone (e.g. where an infrastructure upgrade is provided for under other legislation, such as INSPIRE Directive or where it is part of a digitisation initiative). The amount of changes needed for each dataset and for each public sector body concerned would also vary a lot.

However, the general estimates available indicate the following:

- one-off costs (e.g. API set-up) range from EUR 24.9 million (low estimate) to EUR 435.9 million (high estimate), with a medium estimate of around EUR 122.3 million.
- recurrent costs (e.g. loss of revenue and resources needed to increase data quality up until 2028) are an estimated EUR 473.6 million.

Some of the costs can be offset through funding from the digital Europe programme, which is specifically designed to reduce the burden of publishing HVDs.

What is the impact on small businesses and competitiveness?

The increased availability of data will boost entrepreneurship and result in the creation of new companies, although to a different extent in each data theme (due to e.g. the current availability of data by theme and their maturity). Overall, the open availability of HVDs should lead to 537 000 new companies being created in 2028, up from 485 000 in 2024.

Besides supporting the creation of new companies, HVDs can be important enablers for start-ups to validate their business cases and attract investors. Data incubator programmes have shown a link between strong and proactive open data policies in Member States and the number of successful data start-ups from those countries. Open data can also empower small businesses indirectly: some become so skilled in handling it that data software solutions become one of their main products.

Lastly, the availability of free data helps rebalance the position of small businesses in relation to big tech companies (for whom the cost of data acquisition is not a noticeable barrier).

Will there be significant impacts on national budgets and administrations?

All costs indicated above will be borne by public administrations, but the most significant costs will be one-off (setting up APIs). For some data that reusers currently paid for (notably company, weather and cadastral parcel data), costs will be affected by this data becoming free of charge. The Directive however stipulates that Member States can benefit from a 2-year adjustment period. Moreover, the availability of free data and its machine readability should reduce data handling costs (no invoicing, no requests processing) and increase data management efficiency, thus reducing the overall costs in the medium term.

Will there be other significant impacts?

No other significant impacts were identified.

Proportionality?

The preferred option takes into account feasibility and proportionality requirements. To minimise legal difficulties (e.g. ensuring compliance with EU data protection laws) and reduce costs, the preferred option has a lower-intensity approach for the following themes: geospatial, mobility and company data.

D. Follow up**When will the policy be reviewed?**

The review of the basic act (Directive 2019/1024) can begin in 2025. However, the six themes listed in the Directive can be added by a delegated act, and the specific datasets can be amended through implementing acts at any time.