



EUROPEAN
COMMISSION

Brussels, 21.12.2022
SWD(2022) 432 final

COMMISSION STAFF WORKING DOCUMENT

IMPACT ASSESSMENT

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} - {SWD(2022) 431 final} - {SWD(2022) 433 final} -
{SEC(2022) 450 final}

Table of contents

1.	INTRODUCTION: POLITICAL AND LEGAL CONTEXT	1
1.1.	Policy context	1
1.2.	Economic and societal value of open government data	4
1.3.	Legal context	5
2.	PROBLEM DEFINITION	7
2.1.	What is the problem?	7
2.2.	Problem Drivers	9
2.3.	How will the problem evolve?	14
3.	WHY SHOULD THE EU ACT?	15
3.1.	Legal basis	15
3.2.	Subsidiarity: Necessity of EU action	15
3.3.	Subsidiarity: Added value of EU action	15
4.	OBJECTIVES: WHAT IS TO BE ACHIEVED?	16
4.1.	General objective	16
4.2.	Specific objectives	17
5.	WHAT ARE THE AVAILABLE POLICY OPTIONS?	21
5.1.	What is the baseline from which options are assessed?	23
5.2.	Description of the policy options	25
5.3.	Options discarded at an early stage	29
6.	WHAT ARE THE IMPACTS OF THE POLICY OPTIONS?	30
6.1.	Economic impacts	30
6.2.	Social and environmental impacts	38
6.3.	Impact on SMEs	41
6.4.	Costs	41
6.5.	Stakeholders' views	45
7.	HOW DO THE OPTIONS COMPARE?	47
8.	PREFERRED OPTION	50
8.1.	Optimised intensity package	50
9.	HOW WILL ACTUAL IMPACTS BE MONITORED AND EVALUATED?	58
	ANNEX 1: PROCEDURAL INFORMATION	61
	ANNEX 2: STAKEHOLDER CONSULTATION	65
	ANNEX 3: WHO IS AFFECTED AND HOW?	73
	ANNEX 4: ANALYTICAL METHODS	83
	ANNEX 5: HIGH VALUE DATASETS AS SUPPORT TO START-UPS AND SMES	93
	ANNEX 6: IMPACTS PER CATEGORY	98

Glossary

<i>Term or acronym</i>	<i>Meaning or definition</i>
Application Programming Interface (API)	A set of technical protocols by means of which one piece of software asks another programme to perform a service. The service could be granting access to data or performing a specified function.
(Common European) Data Space	An arrangement composed of an IT environment for secure processing of data by an open and unlimited number of organisations, and a set of legislative, administrative and contractual rules that determine the rights of access to and processing of data.
European Data Portal (EDP)	A portal that harvests the metadata from public data portals across European countries. The portal also acts as a knowledge-sharing platform and conducts studies on the impact of Open Data, including the annual Open Data Maturity Report.
High Value Datasets (HVDs)	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;
INSPIRE Directive	Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community, which aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment.
ITS Directive	Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport
Machine-readable format	A file format structured so that software applications can easily identify, recognise and extract specific data, including individual statements of fact, and their internal structure
Open Data	Data in an open format that can be freely used, re-used and shared by anyone for any purpose. In the context of this document, open data refers to public sector information which is openly re-usable.
PSI Directive	Directive 2003/98/EC on the re-use of public sector information. It was amended by Directive 2013/37/EU and recast by Directive (EU) 2019/1024 (Open Data Directive)
Public sector information (PSI)	Information (i.e. 'documents') in areas of public sector activity, including e.g. social, economic, geographical, weather, educational information, which the public sector collects or produces.

1. INTRODUCTION: POLITICAL AND LEGAL CONTEXT

This Impact Assessment accompanies the Implementing Act laying down a list of High Value Datasets (referred to as ‘HVDs’ throughout the text), in line with the legal requirement¹ of Directive (EU) 2019/1024 EU on open data and the re-use of public sector information² (the ‘Open Data Directive’).

The Open Data Directive provides a common legal framework for a European market for public sector information.³ It aims to make the data held or funded by the public sector easily available for re-use in the economy and society. It entered into force on 16 July 2019, replacing and enhancing the 2003 Public Sector Information Directive (‘the PSI Directive’). Member States should transpose Directive (EU) 2019/1024 by 17 July 2021.

The revised Directive introduced the concept of HVDs, defined in Article 2 of the Directive as “*documents the re-use of which is associated with important benefits for society, the environment and the economy*”, which will act as an important enabler of cross-border data applications and services. The Directive prescribes that datasets identified as HVDs will be made available across the EU for re-use free of charge, in machine-readable formats, via Application Programming Interfaces (APIs) and, where relevant, by means of bulk download.

The Directive in its Annex I lays down six thematic categories of HVDs, focusing on themes in which public sector data has already demonstrated its role of a key driver of innovation and where the public sector is best positioned to act as a data provider: 1) Geospatial, 2) Earth observation and environment, 3) Meteorological, 4) Statistics, 5) Companies and company ownership, 6) Mobility.

The Directive requires the Commission to identify and subsequently adopt implementing act(s) laying down a list of specific high-value datasets belonging to the above thematic categories, and to specify arrangements for the publication and re-use of high-value datasets⁴. This Impact Assessment aims to assess different options for the identification of concrete HVDs and accompanying publishing arrangements.

1.1. Policy context

Open Data Policy

¹ ‘The Commission shall adopt implementing acts laying down a list of specific high-value datasets belonging to the categories set out in Annex I and held by public sector bodies and public undertakings among the documents to which this Directive applies.’

² https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG

³ More information about the Open Data Directive: <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>

⁴ The implementing act must be adopted in accordance with the examination procedure provided in Article 5 of Regulation (EU) No 182/2011.

The basic premise of the open data policy is that information generated with public resources should benefit the entire society. Examples are geographical information, statistics, weather data, data from publicly funded research projects and digitised books from libraries. Public sector bodies produce or collect data in order to be able to carry out their tasks, while the private sector adds value to such data by integrating it in business processes or as a basis for various digital services offered to consumers⁵.

The EU's open data policy⁶ ensures that public sector information (PSI), once produced and used for its primary purpose within the public sphere, can subsequently leave that sphere and be used and re-used for commercial or non-commercial purposes by companies, individuals, researchers or journalists. The policy ensures fair competition by imposing fair, proportionate and non-discriminatory conditions for the re-use of public sector information, and by eliminating situations where public sector bodies enter the information market using publicly funded data under preferential conditions. The principle of opening up public data under the same terms for any re-user: large or small, commercial or non-commercial⁷ was designed to help SMEs and start-ups. While multinational companies have almost unlimited resources to procure the data they need⁸, an open data approach primarily benefits entities that cannot invest capital for data acquisition or extensive 'data cleaning'⁹. A recent report has specifically recommended Open data as a tool to 'challenge the dominance of big tech'¹⁰.

The Open Data Directive treats High-Value Datasets as a special group of public sector data and subjects them to rules that make them the most re-use friendly among the public sector information. This includes a requirement to be available for re-use free of charge, in contrast to all remaining public sector datasets, which can be re-used against a charge (in specific situations foreseen by article 6 of the Directive).

EU Data Strategy

The EU's open data policy is part of the broader EU strategy on data.¹¹ The availability of data plays an important role in innovation, and is for example central to the development of Artificial Intelligence. Public sector information and HVDs in particular will feed into the

⁵ The European Data Portal classifies re-users into broad groups of data aggregators, developers and enrichers, see: European Data Portal, Analytical Report 9: The Economic Benefits of Open Data, 2017.

⁶ <https://ec.europa.eu/digital-single-market/en/open-data>

⁷ The same approach has been adopted in the US with regard to data held by the federal administration (including high value weather or satellite data) which has also been reused by EU companies.

⁸ E.g. Google initially acquired the necessary data for its Google Maps service from Navteq (now Here) and is said to be spending over \$1 billion a year on generating accurate maps and routing data (see: <https://tech.eu/features/4947/nokia-here-bidding-war-location-based-services/>).

⁹ Modifying the original data to make it suitable for a given use case. Open Data may reduce this burden by promoting common standards, formats and licences.

¹⁰ Institute for Public Policy Research, 'Creating a Digital Commons', 2020, <https://www.ippr.org/research/publications/creating-a-digital-commons>

¹¹ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

development of sectoral data spaces – a key ingredient of the Commission’s Data Strategy¹² which are complemented by a horizontal research data space (European Open Science Cloud¹³). For example, geographical information is essential for areas such as the environment, agriculture and mobility. The role of public sector data in this context has also been recognised by the Member States. The Council Conclusions of 9 June 2020¹⁴ underlined that *‘the common European data spaces¹⁵ should be based on a joint effort between the public and the private sector with the aim to deliver high quality data by all parties involved.’*

EU Economic Recovery

The availability of data, including HVDs, will play a key role in the economic recovery strategy for the EU. Among the many approaches to respond to the COVID-19 pandemic, a more intelligent use of data has been at the forefront of discussions across Europe and the world. The Recovery Plan Communication¹⁶ makes a direct reference to the HVDs as one of the measures leading to a ‘real data economy as a motor for innovation and job creation’.

European Green Deal

Safeguarding Europe’s environment and counteracting the negative impact of the climate change are among the main policy goals of the current Commission. The success of the Green Deal Strategy¹⁷ relies, inter alia, on the accessibility of interoperable data which, combined with digital infrastructure (e.g. supercomputers, cloud, ultra-fast networks) and artificial intelligence solutions, can facilitate evidence-based decisions and expand the capacity to understand and tackle environmental challenges. HVDs, notably in the geospatial, Earth observation/environment and meteorological domains¹⁸, will constitute a major data resource in this context. A Common European Green Deal data space will use the potential of the HVDs in support of the Green Deal priority actions on climate change, circular economy, zero-pollution, biodiversity, deforestation and compliance assurance. The establishment of a list of HVDs will also contribute to the “GreenData4All¹⁹” initiative which aims to simplify and modernise the key elements of the EU acquis in the area of environmental information access and provision. Finally, the ‘Destination Earth²⁰’ (digital twin of the Earth) project will

¹² Commission Communication ‘A European Strategy for Data’ of 19 February 2020, COM/2020/66 final.

¹³ <https://ec.europa.eu/research/openscience/index.cfm>

¹⁴ Shaping Europe's Digital Future, 9 June 2020.

¹⁵ The Common European Data Spaces are a concept introduced by the Commission’s 2020 Data Strategy. They can be defined as a mix of data infrastructure, data governance rules and community building whose aim is to enhance data sharing within economic sectors in the EU. The governance layer of the Data Spaces will benefit from legislation currently under preparation (Regulation on European Data Governance) while financial support to the creation of data spaces in e.g. mobility, health or agricultural sectors will be provided by the Digital Europe Programme.

¹⁶ Commission Communication ‘Europe’s moment: Repair and Prepare for the Next Generation’ of 27 May 2020, SWD(2020)98final.

¹⁷ COM(2019) 640 final.

¹⁸ Three out of six domains indicated in Annex I of the Open Data Directive.

¹⁹ https://inspire.ec.europa.eu/sites/default/files/inspire2020_greendataspace_green_deal_data_space.pdf

²⁰ <https://ec.europa.eu/digital-single-market/en/destination-earth-destine>

benefit from HVDs, as a key data source in the process which brings together European scientific and industrial excellence in the establishment of a very high precision digital model of the Earth.

1.2. Economic and societal value of open government data

Given the high share of government spending in the EU's GDP²¹ and the increasing digitalisation of administration, the public sector can be seen as a major source of data in the EU. Data from the public sector often serves as a raw material for data-driven products and services, thus stimulating economic growth and innovation. Finally, public data assets are rapidly turning into a critical driver for the development of new technologies such as artificial intelligence (AI), which require the processing of vast amounts of high-quality data²².

Several recent studies have indicated the enormous socioeconomic potential of public sector data. Examples of these studies are:

- The 2019 OECD report 'Enhancing Access to and Sharing of Data', which estimates that enhanced access and sharing of public sector data can help generate social and economic benefits worth up to 1.5% of GDP²³;
- The study supporting the impact assessment for the Open Data Directive, which predicts that the total direct economic value of public sector information in the EU will reach EUR 194 billion in 2030²⁴;
- The European Data Portal's '2020 report on the Economic Value of Open Data', which estimates that the value of the EU open data market will reach EUR 199.51 billion in 2025²⁵.

In terms of employment, the Report by the European Data Portal²⁶ considers that around 1.09 million direct and indirect open data employees²⁷ were active in the EU in 2019, forecasting an increase to 1.12 – 1.97 million open data employees by 2025.

Data is a critical resource for start-ups and small and medium-sized enterprises (SMEs)²⁸, in particular as a business can be set up with very low initial capital. Over 99% of data supplier companies and over 98.8% of data user companies in the EU are SMEs²⁹. Some 85% of new jobs created in the data economy over the last years have been created by SMEs³⁰. Easy access to high-value data can act as an important enabler for data-savvy SMEs. For example,

²¹ EU-27 general government expenditure stood at 46.7 % of EU GDP in 2018.

²² <https://www.europeandataportal.eu/en/highlights/ai-and-open-data-crucial-combination>

²³ <https://www.oecd.org/going-digital/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm>

²⁴ <https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en>

²⁵ <https://www.europeandataportal.eu/sites/default/files/the-economic-impact-of-open-data.pdf>

²⁶ The Economic Impact of Open Data, 2020, European Data Portal, www.europeandataportal.eu

²⁷ Open data employees are people in both the public and the private sector that are generating, providing, aggregating, re-using, and enriching open data.

²⁸ COM(2020) 103 final.

²⁹ European Commission (2020a). *The European data market monitoring tool*, D2.9 Final Study Report.

³⁰ European Commission, *Entrepreneurship and Small and medium-sized enterprises (SMEs)*.

the vast majority of re-users of more than 600 use-cases gathered by the European Data Portal are SMEs and start-ups³¹.

Insights generated using open government data can also help address societal challenges through the development of innovative solutions (e.g. in fields of health, climate), enhance evidence-based policymaking, increase efficiency in public administration, and facilitate democratic oversight of government activities. An overwhelming majority (91.5%) of the stakeholders contributing to the online public consultation on the European Strategy for Data³² agreed that ‘more data should be available for the common good, for example for improving mobility, delivering personalised medicine, reducing energy consumption and making our society greener.’

1.3. Legal context

As pointed out on the first page of the impacts assessment, the Open Data Directive contains a legal obligation for the Commission to adopt an Implementing Act laying down a list of specific high-value datasets belonging to the categories set out in Annex 1 of the Directive.

The Open Data Directive is a horizontal, minimum harmonisation measure³³. It creates a common EU framework under which public sector bodies and (to a limited extent) public undertakings³⁴ must allow the re-use of the data they hold. At the same time, the provisions of Union and national law that go beyond those minimum requirements (i.e. by making even more data open for re-use), in particular in cases of sectoral law, continue to apply.

An important principle underpinning the legislative framework on the re-use of public sector information is that the Directive builds on the existing access regimes in the Member States and does not intend to change the national rules for access to documents. This means that data excluded or restricted from access by virtue of national law³⁵ are out of scope of the Open Data Directive and as such, cannot be taken into account for the list of HVDs.

In terms of the interplay with intellectual property rights, the Open Data Directive excludes from its scope content for which parties other than public sector bodies ('third parties') hold intellectual property rights. Public sector bodies can publish such data with the permission of the rightholders.

³¹ <https://www.europeandataportal.eu/en/impact-studies/use-cases>

³² Summary report of the public consultation on the European strategy for data, 24 July 2020, <https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data>

³³ See Article 1(1): *this Directive establishes a set of minimum rules governing the re-use and the practical arrangements for facilitating the re-use of (...).*

³⁴ Public sector bodies include the state, regional or local authorities and bodies governed by public law whereas public undertakings can be characterised as publicly controlled companies providing services in utility sectors (legal definitions in Art. 2 of the Directive).

³⁵ Examples of such data may include sensitive personal data, national security data, confidential statistical data, etc.

Harmonisation of the rules and practices in Member States relating to the exploitation of public sector information also prevents distortions of competition on the internal market. The open data policy contributes to achieving the objectives on competition and State aid laid down in Articles 101 to 109 of the Treaty on the Functioning of the European Union.

The material scope of the Open Data Directive defined in Article 1, makes no distinction between personal and non-personal data. Nevertheless, the rules on re-use of public sector information must be applied in full compliance with data protection legislation. Article 1(4) states that the Directive is *‘without prejudice to Union and national law on the protection of personal data, in particular Regulation (EU) 2016/679 and Directive 2002/58/EC and the corresponding provisions of national law’*. Recital 154 of the GDPR clarifies this further by stating that the PSI Directive *‘leaves intact and in no way affects the level of protection of natural persons with regard to the processing of personal data under the provisions of Union and national law, and in particular does not alter the obligations and rights set out in [the GDPR]’*. In practice however, the bulk of personal data held by public sector bodies is not publicly accessible and is therefore excluded from the scope of the Open Data Directive. The Regulation on European Data Governance³⁶ sets requirements for public sector bodies in handling this type of data, to enable its re-use while fully preserving privacy and confidentiality.

In line with the recommendations of an earlier evaluation³⁷, the Open Data Directive includes new Articles 1(6) and 1(7) which clarify its relationship with the Database Directive and the INSPIRE Directive. The INSPIRE Directive is the single most relevant existing EU legal act for the HVDs initiative (especially for the Geospatial, Environment, Meteorological and Earth Observation themes). It sets the legal basis to establish an EU spatial data infrastructure for policies which may have an impact on the environment. It also lays down requirements concerning the interoperability, discoverability and public accessibility of the relevant data arranged under the ‘INSPIRE Themes’, and ranging from addresses and land cover data to species distribution³⁸. Defining datasets falling within the scope of INSPIRE as HVDs would mean adding an open data requirement to existing INSPIRE obligations: augmenting the re-use value of such data on top of the existing rules for INSPIRE data provision. EU-wide sourcing of Earth observation and environmental data is still subject to barriers that affect re-use value creation, which adding the HVD requirements to the INSPIRE themes would do away with.

Given the wide range of thematic fields within which the list of HVDs is to be identified, the corresponding sectoral legislation pertaining to the definition, collection, production and dissemination of public data resources is also of relevance. A range of EU legal acts has been

³⁶ COM/2020/767 final.

³⁷ SWD(2018) 145.

³⁸ Full list of INSPIRE Themes: <https://inspire.ec.europa.eu/Themes/Data%20Specifications/2892>

examined in the context of the initial selection of the specific High Value Datasets³⁹, especially in the field of mobility and transport. Given the close relationship between the objectives of the HVDs in the thematic area of mobility and of the ongoing Intelligent Transport Systems (ITS)⁴⁰ policy, data in scope of the ITS legislation is excluded from the scope of this Impact Assessment. The transport-related datasets that have been considered therefore all come from areas outside ITS.

Finally, the implementing act with a list of HVDs should be seen as an enabling measure (reinforcing the supply of quality data into the market) in the context of the other two upcoming legislative initiatives announced in the Commission's Data Strategy: the European Data Governance Act and the Data Act. The former should facilitate data sharing and exploitation, notably in the context of the Common Data Spaces, whilst the latter will tackle the emerging problems linked to the co-generation and control over data in Internet of Things settings or to B2G⁴¹ data sharing.

2. PROBLEM DEFINITION

2.1. What is the problem?

The public sector generates and controls an ever-growing volume of varied and constantly changing data. This trend is matched by a corresponding growing demand for large (e.g. EU-wide) datasets as a basis for cross-border digital services and for training AI algorithms⁴². Yet, existing barriers to wide and open re-use mean that the full potential of public sector information has not been reached. This negatively affects a myriad of potential re-users from all economic sectors.

A recent study⁴³ shows that an open availability of public sector information produces largest impacts in the following four sectors: public administration, scientific and technical, information and communication, transportation and storage. At the same time, it predicts a high growth in the re-use of public sector data in a number of additional domains⁴⁴ over the coming years.

All re-users (mostly companies but also NGOs, research institutes, associations and individual citizens) active in the above sectors are therefore directly concerned by any barrier created by the public institutions (entities that hold the data) which would prevent or make it harder for

³⁹ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

⁴⁰ The expected impact of ITS Directive and related delegated acts on the availability of data in the transport field should exceed that of a hypothetical inclusion in the list of High Value Datasets because the application of ITS Directive 2010/40/EU is not limited to data held by public sector bodies and public undertakings.

⁴¹ Business to Government.

⁴² <https://apifriends.com/digital-strategy/unleash-your-ai-with-apis/>

⁴³ The Economic Impact of Open Data, European Data Portal, 2020.

⁴⁴ Agriculture, Financial services and insurance, Health; Education, Wholesale retail and trade, Real estate activities.

them to exploit public sector information as input for their commercial or non-commercial activities⁴⁵. The fact that the respondents to the online public consultation were composed of (in decreasing order) companies, followed by citizens, business associations, academic / research institutions, public authorities and finally NGOs confirms this finding. The share of companies (commercial re-users) among the concerned stakeholders was even more dominant in the consultation on the Inception Impact Assessment⁴⁶.

Building on the findings of the Evaluation Report⁴⁷, the Impact Assessment accompanying the proposal for the Open Data Directive⁴⁸ put forward the need to adopt a list of HVDs as part of a broader package of measures addressing the sub-optimal use of public data resources. The HVDs initiative is thematically focused on a subset of public sector information but forms part of that larger policy intervention⁴⁹ and the key problem this initiative intends to tackle is that **public sector data are not being re-used despite their high socio-economic potential**. This problem statement should not be understood literally (i.e. that not a single public dataset is being currently re-used in Europe), but rather as depicting the crux of the matter at stake: public sector information, often of undeniable value for the economy and society, is not being re-used outside of the public sector to its full potential (by feeding into digital services and products or by informing the society in general). In other words, too few public datasets of high potential value are being published and re-used as High Value Datasets. The flip-side of the problem statement can thus be described as follows: public sector data are not widely re-used because only few of them fulfil the requirements expected from a High Value Dataset (having high re-use potential and at the same time being free of charge, machine readable and accessible via APIs). This is due to a number of barriers that inhibit the actual exploitation of HVDs (problem drivers) and which are discussed below.

The problem becomes especially acute given the rapid digitisation of the economy and society, including the rise of Artificial Intelligence or the Internet of Things. Although the problem as such is common to all public data, within the context of this impact assessment, it should be limited to the six data themes indicated by the co-legislators in the Open Data Directive.

Fig. 1. Problem Tree

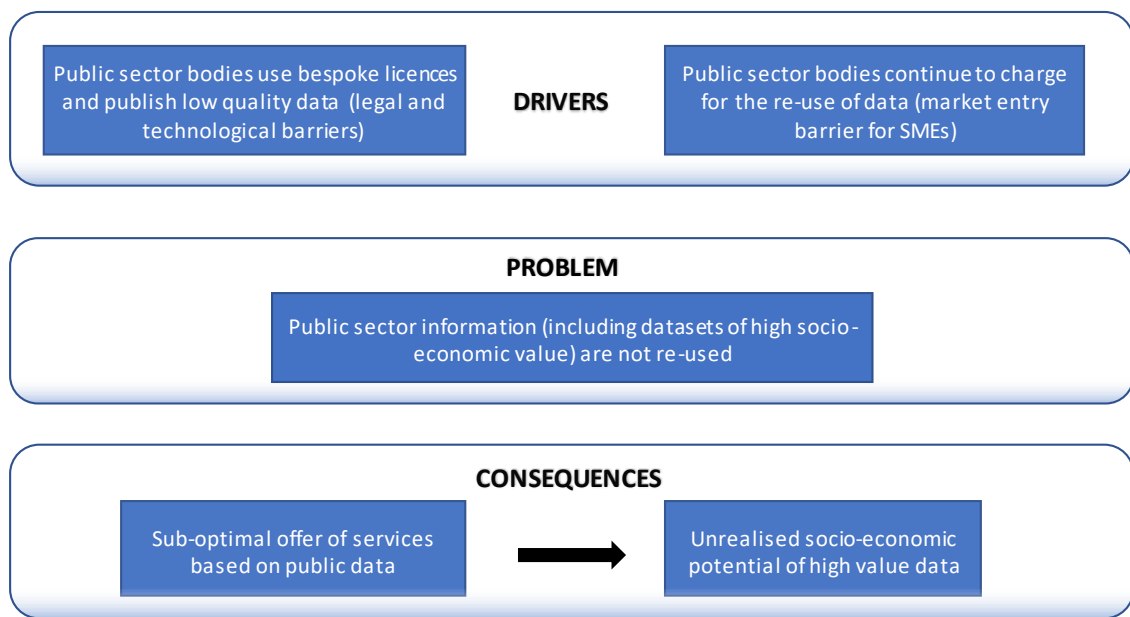
⁴⁵ Selected examples of data re-use across all sectors are showcased by the European Data Portal: <https://www.europeandataportal.eu/en/impact-studies/use-cases>

⁴⁶ See: Annex 2.

⁴⁷ SWD(2018) 145.

⁴⁸ SWD(2018) 127 final.

⁴⁹ In addition to legislation, the EU's Open Data Policy addresses the full range of problems in PSI re-use across Europe. These touch upon the issues of data skills, digitisation of the public sector or collaboration between the public and private sectors.



2.2. Problem Drivers

The problems drivers identified in Fig.1 above and detailed below should be seen in the context of the rapid digitisation of the economy and society, including the development of technologies specifically designed to interpret, transfer, store or produce data, such as Artificial Intelligence or the Internet of Things. Given that the public sector is one of the main producers and collectors of data in the EU, it is in a position to influence the development of the data economy and speed up the uptake of data-enabled technologies. This is particularly true with regard to the supply of categories of data sought after by private sector re-users (such as the HVDs covered by this Impact Assessment).

2.2.1. *Technological and legal barriers (data quality and interoperability)*

A first set of barriers negatively affecting the availability of public sector data for re-use are of a technical nature.

Sub-driver: Lack of modern data dissemination technologies

While most public sector information is nowadays born digital, public sector bodies often lack modern data dissemination technologies, such as Application Programming Interfaces (APIs).

APIs are technical protocols by means of which one piece of software asks another programme to perform a service and can greatly facilitate the access to and re-use of data⁵⁰.

A large majority (78%) of the stakeholders that contributed to the online consultation⁵¹ considered the availability of datasets via APIs to be a relevant or very relevant factor in the context of the HVDs. These supporters were mostly reusers, i.e. businesses, academic and research institutions and citizens. Of the remaining 22% only a small share (3,6%) disagreed with the statement (most being neutral or had no opinion), some of them being citizens but also a few companies or business associations. More than two-thirds (67%) of data consumers surveyed in the recent global study⁵² said that they currently receive disconnected experiences from public sector organisations in terms of data provision. A recent study indicates that fewer than half of APIs are dedicated to data access⁵³. The same study observed that new EU legislation (including the updated rules on the re-use of PSI) has stimulated the adoption of APIs, motivating the public sector to make data more universally available.

Sub-driver: low technical quality of the data

Another technical barrier is related to the actual technical quality of the data, including machine-readability as well as its interoperability. The data contained in large public datasets, such as national registers, becomes more interesting for re-users when it can be machine-processed – i.e. visualised, analysed or summarised without human intervention. Machine readability directly affects usability of data⁵⁴. Notably, it makes data more suitable as basis for a dynamic digital service and realise its commercial potential. A report by the European Data Portal confirmed that users in Europe generally consider the quality of public data to be low (poor metadata, infrequent updates, non-respect of standards, etc) , which negatively impacts the scope of re-use⁵⁵.

Sub-driver: problems to combine datasets

In addition, it is often technically difficult and costly to bring together two or more datasets coming from different data providers, located in different EU Member States⁵⁶. This is mostly due to the fact that public bodies across Europe publish data which are not mutually compatible on the level of formats, semantics or syntax. The lack of well-defined metadata and

⁵⁰ Evaluation Report SWD (2018) 145 observed that ‘the improvements in the reusability scores are driven by increased adoption of Application Programming Interfaces (APIs) and the improving availability of machine-readable datasets’.

⁵¹ <https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data>

⁵² <https://www.mulesoft.com/press-center/customer-experience-research-2019>

⁵³ Application Programming Interfaces in Governments: Why, What and How, JRC research for policy report, 2020.

⁵⁴ <https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data>

⁵⁵ Barriers in working with Open Data, EDP analytical report No 5, 2020.

⁵⁶ This can be referred to as ‘data source interoperability’, see: Publishing Data for Maximum Reuse, Pieter Colpaert, 2018, <https://phd.pietercolpaert.be/#toc>

ontologies can amplify problems with interpreting the data while inconsistent use of standards (e.g. JSON or XML) in terms of exchange formats hinders data transfers and mash-up⁵⁷.

With regard to many of the datasets in scope of this Impact Assessment the bulk of the interoperability barriers have been addressed by the INSPIRE Directive (Geospatial, Earth Observation and Environment) or international standardisation initiatives (Meteorological). For other datasets, interoperability barriers persist to various degrees. Likewise, 87% of stakeholders contributing to the open public consultation⁵⁸ considered standardised formats of data and metadata relevant for improving the re-usability of specific HVDs. In addition, making the re-use of data subject to varying licencing conditions can lead to considerable legal compliance costs.

Sub-driver: lack of licensing interoperability

The lack of licensing interoperability means that it is difficult for re-users (especially those with limited resources for legal clearance) to comply with all the applicable licensing conditions when merging data derived from different public sources as a basis for new insights or new data-based services. This legal barrier to re-use is particularly acute in the case of cross-border use of the data and the creation of pan-European products and services (with 27 different jurisdictions and slightly differing IPR regimes). This led the Commission, after a wide public consultation⁵⁹, to adopt non-binding guidance on licensing (amongst other topics)⁶⁰.

The study supporting this Impact Assessment has found⁶¹ that the usage of Creative Commons licences for the re-use of information held by public bodies which is free of third party rights (e.g. IPRs) is becoming more and more widespread for a large number of datasets (notably in the geospatial, statistical and environmental domains). However, the licensing landscape of public sector data in Europe is still very diverse, with some public bodies using CC licences directly⁶², some using licences which are designed to be compatible with CC licences⁶³ while others continuing to apply bespoke licences or opening public sector data without the use of any specific licence⁶⁴. This situation is often inherited from the times when re-use was typically based on individual requests, followed by negotiation (including on price). In the current environment in which data is published ‘open by design and default’⁶⁵, the use of bespoke licences has little justification. Current data-licensing practices for most

⁵⁷ Combination of several datasets within a larger dataset.

⁵⁸ Annex 2 of this Impact Assessment.

⁵⁹ <https://ec.europa.eu/digital-single-market/en/news/consultation-guidelines-recommended-standard-licences-datasets-and-charging-re-use-public>

⁶⁰ Commission notice — Guidelines on recommended standard licences, datasets and charging for the reuse of documents, OJ C 240, 24.7.2014.

⁶¹ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

⁶² <https://www.istat.it/it/note-legali>

⁶³ <https://data.norge.no/nlod/en/2.0/> or <https://www.etalab.gouv.fr/licence-ouverte-open-licence>

⁶⁴ <http://data.riksdagen.se/In-English/>

⁶⁵ As suggested by Recital 16 of the Open Data Directive.

public data resources severely hamper data re-use, especially at scale⁶⁶, contributing to the interoperability barrier⁶⁷.

2.2.2. *Charges for re-use as a market entry barrier for SMEs*

A second set of barriers causing unavailability for re-use of data held by public authorities are charges that create market entry barriers for data re-users, in particular SMEs. EU legislation on the re-use of public sector information has since its inception in 2003, with each amendment, gradually limited the possibility to apply fees for re-use to the extent possible (i.e. politically acceptable to MS⁶⁸). The 2019 recast of the Open Data Directive generalised the rule that public sector data should be available for free, or at maximum against the marginal costs of dissemination. However, a number of exceptions are allowed, notably in cases of public undertakings that operate in a competitive environment⁶⁹ or for situations in which public sector bodies are required to cover a substantial part of their operating costs from charges for the re-use of data⁷⁰. The latter provision is relevant for the HVDs, since the different approaches in the Member States mean that HVDs are freely available in a number of countries, while in others their re-use will only be possible against a fee. While overall the principle of free re-use is gaining ground across the public sector in the EU, central bodies that hold large data repositories are often expected to offset at least a part of their operating cost by commercialising their data in the market.

This is particularly the case⁷¹ for data within the themes of ‘Companies and company ownership’ (e.g. for data exceeding basic company info⁷² or for a larger number of searches), ‘Meteorological’ (e.g. observation data⁷³) and ‘Geospatial’ (e.g. cadastral data⁷⁴). It is important to note that charging for the re-use of public data has been found to be unjustified from a macroeconomic point of view. Public sector information is a non-rivalrous good that can be re-used multiple times and its high price elasticity means that a decrease in price

⁶⁶ See: Socio-legal Barriers to Data Reuse, National Library of Medicine, 2019, <https://nlmdirector.nlm.nih.gov/2019/06/11/socio-legal-barriers-to-data-reuse> or LAPSI License Interoperability Report, 2013, <https://ec.europa.eu/digital-single-market/en/news/lapsi-license-interoperability-report>

⁶⁷ Not least because the adoption of CC or CC-compatible licences will allow for the seamless merging of European Commission’s and Member States’ data into EU wide data services and products.

⁶⁸ The practice of charging is often linked to the funding mechanism in which public sector bodies are not fully financed by the budget.

⁶⁹ The Directive even allows for specific HVDs held by public undertaking to be charged for in cases where their free availability would lead to a distortion of competition in the relevant markets.

⁷⁰ The recast Directive makes it clear that any application of charges beyond marginal costs can only happen in exceptional cases (Recital 36) and under strict transparency requirements.

⁷¹ As evidenced in the Support Study.

⁷² E.g. company data in Belgium: <https://economie.fgov.be/en/themes/enterprises/crossroads-bank-enterprises/services-everyone/cbe-file-containing-all-public>

⁷³ E.g. weather data in Croatia: https://meteo.hr/proizvodite.php?section=proizvodite_usluge¶m=services

⁷⁴ E.g. cadastral data of the Austrian Surveying Service (BEV): https://www.bev.gv.at/portal/page?_pageid=713,3175363&_dad=portal&_schema=PORTAL#Anchor-37650

triggers a surge in usage⁷⁵. It is produced as part of a public service, so its creation does not depend on market forces. Most importantly, free government data generates extra commercial activity, especially by SMEs, which translates into more jobs and revenue from taxes. The practice of charging for public sector data mainly has a negative effect on SMEs, because they cannot afford to pay for the data or collect it by themselves. For them, it is equivalent to a market entry barrier.

In addition, the public sector itself is a key user of the data, so part of the income for one public sector body represents costs for other public bodies and leads to a suboptimal use of the data for public purposes. Also, charging generates substantial costs for public sector bodies producing the data, since it presupposes, amongst other things, that an invoicing, accounting and access control system is in place.

The available research shows that applying charges for the re-use of public sector data creates a situation in which the actual data usage is sub-optimal: in other words, the existing demand for the data and the capacity (necessary skills, technologies) to create added value on top of it are artificially thwarted by the market entry barrier described above. It also prevents sizeable efficiencies within the public sector to materialise. The anti-innovation and anti-efficiency aspects of charges are evidenced by the effects of the opening up of a dataset which was previously charged for. For instance, the opening of geospatial data in Denmark led to efficiency gains of DKK 22 million over 4 years⁷⁶ while a 2019 study⁷⁷ looked at the link between cost reduction and increase in the re-user base to assess customers' sensitivity to price. The study clearly shows the elasticity of the demand for re-use of companies' data and confirms the findings of previous studies: for each small decrease in the cost of data, the number of re-users multiplies⁷⁸. Countries such as France and Denmark, which started to provide these datasets through APIs and for free, saw a very significant surge in the number of re-users (from 12 to 1 230 full re-users for Institut national de la propriété intellectuelle (INPI),⁷⁹ and the SIRENE database of company becoming the third most accessed on the data.gouv.fr portal⁸⁰). The experience with the EU-funded data start-up incubator ODINE has also demonstrated the business potential that can be released with the right combination of

⁷⁵ According to the evaluation report SWD 2018 145: 'In the studies where specific types of data were analysed (e.g. meteorological, geographic and hydrographical data), it emerged that there is high price elasticity in these domains, that is to say that re-users are very sensitive and reactive to price'.

⁷⁶ <http://sdfe.dk/media/2917052/20170317-the-impact-of-the-open-geographical-data-management-summary-version-13-pwc-grvkvd.pdf>

⁷⁷ <https://www.gov.uk/government/publications/companies-house-data-valuing-the-user-benefits>

⁷⁸ See for instance, The cost of Geospatial Open Data, Peter A. Johnson, Renee Sieber, Teresa Scassa, Monica Stephens, Pamela Robinson, Transaction in GIS, Wiley, January 2017,

<http://onlinelibrary.wiley.com/doi/10.1111/tgis.12283/full> or See case studies on the Norwegian METNO case (meteorological data) and case study on the Dutch KNMI case (meteorological data), Study on the Pricing of Public Sector Information – POPSI Study, October 2011, Deloitte, <https://ec.europa.eu/digital-single-market/en/news/pricing-public-sector-information-study-popsis-models-supply-and-charging-public-sector>

⁷⁹ Stakeholders interviews, European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

⁸⁰ <https://www.insee.fr/fr/statistiques/4238594?sommaire=4238635>

freely available data, technological competence and seed funding together with business mentoring⁸¹.

To sum up, the available research has consistently shown that lowering the costs of data acquisition leads to an exponential increase in the re-use, both in terms of the volume of data and the number of re-users involved⁸².

2.3. How will the problem evolve?

Without intervention, the patchwork of technical, legal and market entry barriers for re-users of public sector data in the different Member States is likely to persist. Core reference datasets easily available for re-use across the EU will not emerge and their potential will not be fully realised. Individual Member States may advance their policies and legislation in order to make HVDs more open and usable⁸³, but such uncoordinated measures would hardly be perceptible at the scale of the EU data market.

This will slow down the creation of cross-border data products and services within the EU, thus holding back the socioeconomic potential of public data resources. The lack of readily usable public datasets of a sufficient scale and covering different Member States will also negatively affect the development of Artificial Intelligence applications. The barriers will continue to disproportionately affect SMEs that may not be in a position to afford the initial investment in data acquisition or in the technical means to use the data.

Deficiencies in making the relevant datasets available can lead to a paradoxical situation in which the supply of usable and commercially interesting public data does not match the pace at which the overall data supply is increasing. While the creation of public datasets will not cease or diminish (because it usually constitutes a public sector task, i.e. an obligation), the scale of its re-use will remain limited. This creates a risk of gradual irrelevance of certain types of public sector data in comparison to the datasets collected and made available via APIs by the large private sector actors of the data economy. Already in 2013, a report by the French Court of Auditors⁸⁴ observed that due to technological advances and economic barriers to data access and use, the state's monopoly in supplying reliable reference data is being eroded. Leading technological platforms are increasingly collecting and generating data themselves (e.g. by sourcing data from devices or installing their own sensors) while various collaborative initiatives (e.g. OpenStreetMap) offer innovative and attractive ways of co-creating and co-using various data directly by the users. Indeed, the report found various instances of extensive use of such data within the French administration itself. Also, non-EU

⁸¹ <https://opendataincubator.eu/wp-content/uploads/2016/01/ODINE-Final-report-by-IDC.pdf>

⁸² About GMES and Data: Geese and Golden Eggs, G Sawyer, M de Vries, 2012, Figure 3-4 'Overview of increases in demand following lowered PSI reuse charges'.

⁸³ The latest EDP's Open Data Maturity Report deplores the persisting disparities among the EU Member States and sees the High Value Datasets as an important opportunity for consolidating Europe's open data ecosystem.

⁸⁴ Trojette Report sur l'ouverture des données, 2013, Cour des Comptes, Chapter 4.1 – 'La fin du monopole public dans l'élaboration de certaines données de référence'.

countries (e.g. USA, Norway, UK, Australia) offer open data of global importance (company data, weather data, environmental information). This increasing competition in the area of data provision may in the longer run decrease the relevance of public data in the EU and call into question sustained public support towards its production.

3. WHY SHOULD THE EU ACT?

3.1. Legal basis

The legal basis for the adoption of the implementing act is the Open Data Directive, and in particular Article 14 thereof, according to which the European Commission is required to adopt a list of HVDs.

The legal basis for the adoption of the Open Data Directive is Article 114 of the Treaty on the Functioning of the European Union. The same legal basis applied to the adoption of the original Directive 2003/98/EC (recast by the Open Data Directive).

3.2. Subsidiarity: Necessity of EU action

As explained in chapter 1, EU action (both the preparation of the current Impact Assessment and the adoption of the Implementing Act) is required to fulfil a legal requirement placed on the Commission by the co-legislators.

The necessity of EU action was furthermore assessed in the Impact Assessment underlying the adoption of the Open Data Directive. It concluded that the adoption of a list of HVDs would be necessary to ensure that public data of comparable thematic scope are available for re-use across Member States under similar legal and technical conditions. This, in turn, is necessary to facilitate the offering of services based on data sourced from different EU countries or for applying a data-based business model tested in one Member State seamlessly to another. Currently, some Member States have already laid down lists of datasets for open and free re-use⁸⁵, while other Member States have made little progress in encouraging re-use beyond the introduction of legislative changes mandated by EU level rules. EU action is therefore also needed to ensure the coherence of the growing EU data market and to prevent its fragmentation.

3.3. Subsidiarity: Added value of EU action

The overall objective of the European Strategy for Data⁸⁶, in which the HVDs initiative plays a key role, is to create a single European data space – a genuine single market for data, whose attainment depends on a concerted and common implementation by all EU Member States.

⁸⁵ Denmark, France and the Czech Republic have pioneered the approach of defining ‘reference data’ for fully open re-use.

⁸⁶ COM(2020) 66 final.

The data value chains in the EU are already structured largely in a cross-border manner, with data holders, data processors, data enrichers and final data users scattered across various Member States. An increase in both the supply of public data and its exploitation therefore needs to target the single market in its entirety. In addition, the proposed measures will not operate in a legal vacuum but should be seen as the next step towards a full availability of PSI for re-use (open data): a common EU policy objective accepted by the Member States already in 2003 and confirmed in 2013 and 2019⁸⁷.

According to the vast majority of respondents (87%) in the open public consultation in the context of the review of the PSI Directive⁸⁸, EU intervention in the form of legislation on the re-use of public sector data presents a clear EU added value. This view was also shared by experts in the Evaluation Report on the functioning of the PSI Directive⁸⁹. The stakeholders generally confirm that the Directive (including subsequent amendments) has played a key role in encouraging national authorities to open up more public sector data across the EU and facilitating access to PSI from countries other than the one where the person concerned lives, and it has been conducive to the creation of an EU-wide market for products and services based on PSI.

Similarly, 82% of respondents to the open public consultation conducted in 2020 considered that the establishment of a list of HVDs, to be made available free of charge, without restrictions and via APIs, is a good way to ensure that public sector data has a positive impact on the EU's economy and society. This figure increases to 90% when considering responses from public authorities only⁹⁰. Of the remaining respondents, 10% were actually neutral or had no opinion, while 4,5% disagreed (representing various types of stakeholders, including 15 citizens). The general opinions of the stakeholders (both data producers and users) can therefore be said to validate the approach of tackling the barriers to public data re-use via EU-level instruments.

4. OBJECTIVES: WHAT IS TO BE ACHIEVED?

4.1. General objective

The general objective of the initiative is to increase the re-use of public sector data in the EU. Public sector information is an important primary material for digital products and services, and the re-use of HVDs would benefit from the scale of the Single Market. Prioritising the opening up of specific data of high re-use potential has long been advocated by open data

⁸⁷ Dates referring to the initial adoption of the Directive 2003/98/EU and its subsequent revisions.

⁸⁸ <https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-revision-directive-reuse-public-sector-information>

⁸⁹ SWD(2018) 145.

⁹⁰ Summary report of the public consultation on the European strategy for data, 24 July 2020, <https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data>

experts⁹¹ and via non-binding international initiatives⁹² as an efficient way of ensuring significant socioeconomic gains with modest public investment.

4.2. Specific objectives

The specific objectives of the planned intervention aim to ensure that a common EU-wide layer of public sector datasets characterised by their high social, economic and environmental potential is easily and freely available for re-use. Such an approach implies that the Commission should strive to maximise the number and thematic range of HVDs, because the more inclusive the final set of the datasets, the higher the overall impact of the initiative will be. The specific objectives will nevertheless need to be translated into policy options in a manner which not only maximises the intended economic impact but also takes into account in a realistic manner the legal/political feasibility of the measures and the associated costs.

The HVDs can then act as pivotal data for other (public or private sector) data and encourage the re-use of these related data (e.g. high-value public geospatial data bundled with data derived from sensors or mobile devices/cars). The specific objectives of the planned Implementing Act will tackle the corresponding problem drivers identified in chapter 2 in the specific areas identified. They ensure that the current barriers in the form of low interoperability (both legal and technical) along with the market barriers caused by charges cease to apply in relation to the re-use of HVDs selected from the wide array of public sector information that falls within the six topics listed in Annex I to the Open Data Directive. They are fully aligned to the requirements of Article 14(1) of the Directive which makes it clear that once identified and included in an Implementing Act, the HVDs will need to fulfil the following key conditions:

1. Their re-use shall be free of charge;
2. They shall be machine readable, available via APIs and, where relevant, by means of bulk download;
3. They may be subject to arrangements for publication and re-use which shall be compatible with open standard licences.

The criteria for the definition of the HVDs from among the wide set of public sector information are laid down by the Open Data Directive⁹³. The role of this Impact Assessment is therefore to inform the Commission on the possible choices regarding the content of these datasets in each of the six data themes along with the accompanying modalities for publication, with a view to maximising their beneficial impact on the society and economy. The first objective is pre-defined in the Open Data Directive in a manner which makes any gradation at the level of policy options impossible (e.g. even small charges for the re-use of

⁹¹ Report on high value datasets from European Institutions, PWC Services 2014.

⁹² Technical Annex to G8 Open Data Charter, Action 2: Release of High Value Data, <https://www.gov.uk/government/publications/open-data-charter>

⁹³ Article 14(2) mentions four separate criteria that need to be met for data to qualify as HVDs.

HVDs are forbidden) but objectives no. 2 and 3 lend themselves to a more sophisticated exercise of mapping onto policy options. They maintain the flexibility required for the formulation of various policy options leading to the optimal choice in terms of maximising the inclusiveness of datasets while respecting favourable costs-benefits ratio of the overall intervention.

Objective 1: Eliminate charges for re-use as a market entry barrier, especially for SMEs

The majority of re-users of open data consulted prior to the revision of the PSI Directive voiced their clear support for limiting exceptions to the principle of free re-use (or at maximum cap charges at marginal cost)⁹⁴. The rule for HVDs set out by the Open Data Directive is that they should be available for re-use for free. This being said, Member States may grant a maximum 2-year exemption to those public sector bodies that are required to generate revenue to cover a substantial part of their costs relating to the performance of their public tasks, where free availability would lead to a substantial impact on the budget of such bodies. In addition, the principle of free reuse could be waived for datasets held by public undertakings or by libraries, museums and archives.

The general rule on free availability of HVDs was subject to the Impact Assessment on the review of PSI Directive in 2018⁹⁵. Various studies referenced in this report have shown that oftentimes there is no valid reason to charge for the re-use of information which already exists, which has already been paid for by the taxpayer, and which has already been used by the public sector body for its primary purpose.

One of the fears sometimes raised with regard to the free availability of open data is that as a result, large multinational tech companies would benefit at the expense of European enterprises or local start-ups⁹⁶. However, as already mentioned in Chapter 1.1., the important distinction to be made in this context is that the price of data is never a barrier for the ‘Big Tech’ companies⁹⁷ while even a small charge acts as a deterrent to SMEs and start-ups⁹⁸. If the objective is to focus on supporting data use by SMEs (instead of creating a new source of revenue for the public sector), the best solution is to impose a zero-charges policy rather than

⁹⁴ <https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-revision-directive-reuse-public-sector-information>

⁹⁵ SWD(2018) 127 final.

⁹⁶ See: PASC [Public Administration Select Committee] (2014) Public Administration Committee - Tenth Report. Statistics and Open Data: Harvesting unused knowledge, empowering citizens and improving public services <http://www.publications.parliament.uk/pa/cm201314/cmselect/cmpubadm/564/56402.htm>

⁹⁷ For instance, Google has spent 29 billion dollars on its top 10 acquisitions, which include huge datasets held by Youtube, Waze or FitBit (<https://www.cbinsights.com/research/google-biggest-acquisitions-infographic/>)

⁹⁸ One third of the start-ups involved in the ODINE incubator stated that not only their business would be negatively affected but that they would simply not exist without public data being open and free. See: Impact Assessment of ODINE programme. 2017, IDC.

to differentiate charges between large and small companies. It is also worth highlighting that start-ups are among the most vocal supporters of the legislation on HVDs⁹⁹.

The rationale behind this objective is equally valid across all the data themes in the scope of the Impact Assessment, even if the actual impact will likely differ depending on the state of play for each HVD: some HVDs are already predominantly available free of charge (e.g. Earth observation data or statistical data) while others are not. In any case, the Open Data Directive clearly makes the free availability of HVDs a universal requirement, save for exceptional cases such as public undertakings exposed to competition or cultural institutions. Given the thematic scope of this initiative, such exceptional cases are not elaborated upon in this document (see also chapter 5.1 below)¹⁰⁰.

Objective 2: Ensure that HVDs are easy to process by machines

APIs that enable machine-to-machine communication ensure cost and time savings and allow public and private sector developers to build apps, widgets, websites and other tools which add value to the underlying data. They also allow the emergence of completely new use-cases. Stakeholders interviewed in the context of the Impact Assessment support study have called for the adoption of new legal instruments to speed up the adoption of API strategies.

As well as in the objective above, the provision of data via APIs is an obligation that applies to all HVDs. Nevertheless, the intervention should take into account the current practices per thematic domain, the readiness of the public sector, the expected technological evolution of APIs, the impact of the relevant legislation (e.g. INSPIRE Directive) as well as the re-users' needs (for instance, the extent to which bulk download contributes to the usability of a given HVD) and modulate the requirements applicable to APIs accordingly.

Objective 3: Ensure that HVDs are interoperable

One of the purposes of an open data policy is to ensure that technical or legal barriers do not hinder entities outside of the public sector in the re-use of public sector data. The re-use of HVDs, as a subset of public sector information, would normally be subject to the general regime of the Open Data Directive, which specifies that the re-use of public sector information should not be subject to any conditions at all, unless the licensing conditions are justified by a public interest objective. However, the Implementing Act should impose re-use conditions specific to the HVDs, allowing for the approximation of licensing practices applicable to HVDs across the EU. Such conditions would need to be compatible with open standard licences¹⁰¹. The opportunity to maximise the re-use of data by aligning with the most

⁹⁹ See: position paper on AI and data by France Digitale , a grouping of 13420 start-ups, <https://francedigitale.org/combat/digitaleu/>

¹⁰⁰ The charging exception may however become relevant in a case of future extension of the thematic range of the High Value Datasets via a delegated act, as foreseen in the Open Data Directive.

¹⁰¹ Such as Creative Commons BY or Creative Commons 0, which allow for an unrestricted re-use, including for commercial purposes.

successful global licensing standards has for instance led the Commission to adopt Creative Commons licences as the default licensing solution for the reuse of all Commission documents¹⁰². The result of the online public consultation reflects this approach, since the majority (80%) of stakeholders considered licensing and other terms applicable to re-use very relevant to improve the re-usability of specific HVDs (while less than 3% found it irrelevant and the rest was neutral). The aim is to apply compatible re-use conditions per data theme and minimise legal uncertainty for the re-users. Some flexibility might be provided by accepting the use of existing bespoke licences for those data themes where bespoke licences are still common but only if they are fully compatible with the standard licences serving as models.

Identifying the necessary minimum quality requirements applicable to HVDs, such as formats, metadata standards, update frequency, granularity or key attributes will help provide the data in a more uniform manner across the EU, facilitating their discoverability and integration.

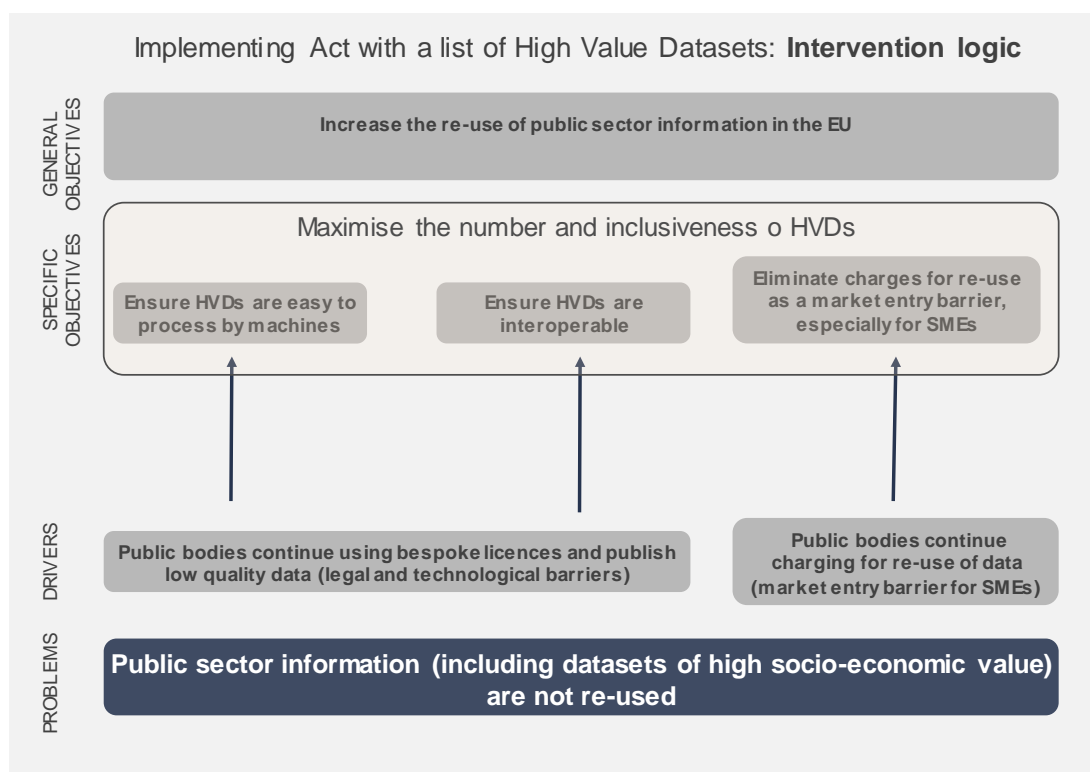
This objective should be implemented in a way which optimises the impact of standardisation investments in each theme so as to minimise the burden of its implementation by the public sector while reinforcing the application of the existing interoperability requirements. This would mean e.g. relying to a large extent on the standards set out by the applicable EU legislation (e.g. INSPIRE Directive) and the relevant non-legislative activities (e.g. the establishment of e-reporting priority list or the harmonization activities of the ESS¹⁰³) for the Geospatial, Earth observation and environment or Statistical data themes. On the contrary, a more customised solution would need to be found in the case of HVDs falling under the Company data and company ownership theme¹⁰⁴.

Fig.2. Intervention logic

¹⁰² <https://ec.europa.eu/jrc/en/news/commission-makes-it-even-easier-citizens-reuse-all-information-it-publishes-online>

¹⁰³ <https://ec.europa.eu/eurostat/web/ess/about-us>

¹⁰⁴ Apart from the case of financial reports whose formats are defined by EU ESEF Regulation 2019/815.



5. WHAT ARE THE AVAILABLE POLICY OPTIONS?

The Open Data Directive is quite specific in describing the scope of action for the Commission as regards the Implementing Act. Firstly, the adoption of the Implementing Act is an obligation. Secondly, the thematic range of the intervention is limited to the six subject areas enumerated in Annex I to the Directive. Thirdly, the basic conditions for the quality of the datasets and for their publication (machine-readable data to be freely available via APIs) are specified. Furthermore, the two main elements that the Implementing Act must contain are defined: a list of specific datasets per thematic category and the accompanying arrangements for publication and re-use¹⁰⁵. In addition, during the discussions on the Open Data Directive in 2019, the co-legislators, and in particular the European Parliament, stressed the importance of identifying HVDs in each of the categories enumerated in Annex I to the Directive.

Finally, regarding the arrangements¹⁰⁶ the Commission's room for manoeuvre is limited by the requirement for compatibility with open standard licences and the need to take into

¹⁰⁵ For the sake of completeness, it should also be noted that the process for the adoption of the Act has also been pre-defined. The proposed list of high value datasets will have to be accepted by the Member States, by way of an examination procedure.

¹⁰⁶ The Evaluation Report of the PSI Directive (SWD(2018) 145 final) when discussing its effectiveness, noted already that 'an additional area of improvement would be to provide high value datasets published with common more harmonised data models in the same level of granularity or scale in each Member State'.

account the investments made by the Member States in open data approaches, such as investments into the development and roll-out of certain standards¹⁰⁷.

For the purpose of the present Impact Assessment, the preceding remarks demonstrate the limited range of possible policy alternatives.

In line with the permissible scope of intervention set out in the Open Data Directive, **two main parameters have guided the development of the options:** the number of data fields and/or datasets to be included as high-value datasets, and the intensity of the measures for publication. By working increasingly with these two parameters, it was possible to build two options per thematic area, of varying intervention intensity: a ‘low hanging fruit’ scenario which minimises regulatory burden for the public sector, and a ‘wish list’ scenario for the re-users, maximising the re-use possibilities but also placing a substantial adjustment burden on the public data holders. These two options do not reflect the full range of real-life possibilities for publishing public sector data, since the modulation and combination of scope and re-use arrangements for any data theme are virtually endless. They are a somewhat idealised embodiments of two different approaches located on the opposite sides of the intensity scale. Their aim was to construct a workable approach for the purpose of designing one common legal act encompassing six different and often unrelated thematic fields. The two options can be seen as a pragmatic simplification allowing for a meaningful evidence collection (e.g. in terms of expected benefits and costs) and facilitating structured feedback from the stakeholders. They are designed to give the Commission a clear indication whether the preferred approach (per each theme separately) should lean towards the ‘wish list’ or towards the ‘low hanging fruit’ scenario, while leaving space for a reality check in the form of minor adjustments at the stage of preparing a draft implementing regulation and discussions in the comitology procedure (such as postponing the applicability of certain provisions beyond the date of the entry into force of the act). Both policy options were assessed for each of the themes separately and were intentionally designed not to be ‘extreme’: they are perfectly feasible and both lead to an improvement of the current status quo in terms of data re-use. The methodological approach and the content of the options were validated with the stakeholders to ensure that they are realistic and fully in line with the objectives of the initiative, as defined by the Open Data Directive. The descriptions of the two options took into account the specificities of each theme as far as possible. In each data theme the difference between the options consists in targeting the aspects that would make a real difference for re-users in practice. For themes in which a large part of the datasets are already widely available (such as statistics, mobility or geospatial data), the policy options were mostly constructed around the re-use friendliness of the measures for publication. On the other hand, the scope and inclusiveness of datasets were used as the distinctive feature of the strength of intervention in thematic areas where economic barriers to re-use persist or where specific data values are in high demand by the re-users, while the sought-after datasets are often not proactively

¹⁰⁷ Article 14(1) of the Open Data Directive.

published (company data, meteorological information). With regard to licensing, one of the two most open Creative Commons licences were chosen for each option and theme, taking into current prevailing licensing practices and re-users demands

It is important to underline that from the re-users' perspective, the scope of data and the corresponding re-use modalities are intrinsically linked and should not be decoupled. A wide dataset quickly loses value if accompanied by restrictive re-use terms, while fully open re-use conditions lose relevance if applicable to a dataset composed of a narrow set of data values. Therefore, artificially decoupling the two impact factors per category and combining them into four options (e.g. high number of datasets with low publication requirements) was not retained as a way forward.

In theory, other intermediate options could also have been introduced, for example an intermediate option for each category in terms of datasets and publication requirements. However, while complicating the Impact Assessment process and the interactions with stakeholders, this would not have added much value in terms of determining the intensity of the action in relation to the different themes. Finally, the further step of negotiating the draft implementing act with the Member States through the comitology procedure will act as a buffer and is likely to result in further adjustments.

A possibility of designing a 'gradual approach' was also discarded because the automatic application of the requirements pre-defined in the Open Data Directive made such an option very difficult to implement in practice. A unit of public sector information, once identified as a High-Value Dataset, automatically becomes freely reusable in a machine-readable format via an API.

5.1. What is the baseline from which options are assessed?

As described in detail in Chapter 1, the re-use of public sector information in the EU, including HVDs, has been the subject of horizontal legislation and forms a part of a long-standing EU policy. The relevant legal framework incorporates sectoral rules governing aspects other than re-use (data collection, data access, quality or interoperability) in each of the themes discussed in this Impact Assessment. The gradual improvement in the availability and subsequent re-use of public sector data¹⁰⁸ would therefore likely continue and influence to some extent the overall data market in the EU. Likewise, the European Data Market Monitoring Tool predicts an expansion of the EU data economy even under the most challenging economic scenarios¹⁰⁹. At the same time, without any focused initiative

¹⁰⁸ As shown in the annual Open Data Maturity Studies conducted by the European Data Portal.

¹⁰⁹ Data landscape, The European Data Market Monitoring Tool see: <http://datalandscape.eu/european-data-market-monitoring-tool-2018>

facilitating their re-use, the contribution of HVDs to the overall data market growth would remain small, limiting the potential socioeconomic benefits related to their open re-use.

HVDs that have been examined and the logic behind the design of the lower and higher intensity interventions

The Open Data Directive leaves the Commission some room for discretion (varying across thematic domains)¹¹⁰ to propose a subset of data defined as HVDs in each of the six data themes¹¹¹.

In order to choose the first layer of data or data attributes per data theme, the Commission consulted a wide range of stakeholders – representing both data holders and users¹¹². In parallel, the Impact Assessment Study mapped all the relevant legislation at the European level, in order to identify which data fields are covered by EU rules, and therefore already exist across the European Union.

Furthermore, interviews with stakeholders allowed to go one step further and develop a preliminary ‘wish list’ of datasets which are considered to be of high value from an economic, societal and reuse perspective. The study then performed a thorough evidence/literature review, generating 32 categories of value and supported by 126 quantitative and qualitative indicators (examining the social, economic and environmental value of the data)¹¹³, to assess the degree to which each dataset can indeed be considered to be of ‘high value’, in line with the assessment criteria enumerated in Article 14(2) of the Directive.

Different potential HVDs were found to have different values, for instance some bearing more potential for economic benefits (i.e. data on company and company ownership) and others for environmental benefits (e.g. mobility data). Datasets of high value either have a value in many of the categories of the framework (breadth) or a very strong value in particular categories (depth).

The initial (wide) list of HVDs per theme includes data value categories for which there is very substantial evidence supporting their high (potential) value. This wide selection

¹¹⁰ The room for manoeuvre is quite restricted for the ‘mobility’ domain. Article 14(2) of the Directive calls on the Commission to ensure complementarity with existing legal acts, such as Directive 2010/40/EU ITS Directive). This is due to the fact that the said Directive (accompanied by a series of delegated regulations) mandates the opening up of a wide scope of static and dynamic mobility data held by both public and private entities – thus exceeding the minimum level of harmonisation in the Open Data Directive. Accordingly, the Impact Assessment takes into account the data which are not in scope of the ITS Directive and its delegated regulations.

¹¹¹ The Directive gives the possibility to the Commission to also consider data held by public undertakings in several utility sectors for the purpose of identifying High Value Datasets. Given the thematic scope of the exercise, the intended objectives, the expected costs involved and the need to respect relevant sectoral law (notably the ITS Directive that covers also public undertakings) as well as clearly negative feedback from the public undertakings in the course of stakeholder consultations, no objective reasons were found to include public undertakings’ data in the scope at this stage.

¹¹² See: Annex II.

¹¹³ For full explanation of the methodology, see Annex 4.

forms the basis of the higher intensity policy option. On the other hand, the scope or granularity of datasets is reduced accordingly (where relevant) for the lower intensity option. A similar logic was applied to the accompanying arrangements for publication and easier re-usability of datasets. While the lower intensity policy option would impose practices that are in place in most Member States, the higher intensity policy option would generalise the approaches of the most re-use friendly Member States.

5.2. Description of the policy options

5.2.1. Baseline scenario (no action)

The baseline scenario supposes that the current status quo is maintained in terms of the availability of HVDs. In practice, this would mean that despite several improvements expected thanks to the transposition of the Open Data Directive in mid-2021, the actual impact of the data of highest value would remain limited, due to the persistence of the problems described in Chapter 2. The Directive, once implemented, should (inter alia) lead to better availability of dynamic data, discourage public-private arrangements that may restrict actual re-use, increase the supply of reusable data from research bodies and public undertakings as well as reinforce the policy of free re-use. It will however stop short of imposing a fully ‘open data’ approach to public sector information, even in cases of data for which an open regime would bring substantial socioeconomic gains. In other words, while the re-use of public data across the board is expected to grow, the public sector would not necessarily focus its efforts on the data most sought after by re-users or most impactful for the economy.

5.2.2. The ‘low-hanging fruit’ (lower intensity) option¹¹⁴

For this option, a relatively narrow range of datasets within each of the thematic areas, or only some data fields within a given dataset, would be considered as HVDs, in comparison to the possible range of datasets under initial consideration (composed of those public sector datasets which are already covered by EU legislation and of highest relevance for re-users).

Selecting datasets as basis for interventions of varied intensity: main steps
<ul style="list-style-type: none"> • Identifying relevant EU legislation per theme • Identifying data in scope and applicable access/availability requirements • Verification with stakeholders (wish-list) • Assessment of potential value based on 126 indicators

¹¹⁴ It should be noted that the intensities of intervention are designed and then assessed independently and separately theme by theme. It is only after the impacts of the policy options in each theme are assessed that the choice of the most appropriate policy package for the entire instrument can be made in Chapter 8.

- Assessing costs/benefits of listing the data as HVDs
- Double-checking with stakeholders (re-users/data holders/Commission)

This ‘lower intensity’ or ‘limited ambition’ approach can be described as that of collecting ‘the low-hanging fruit’, which in practice would translate into listing as HVDs mostly the datasets which are already widely available (e.g. published online), although not necessarily as open data. It would also mean that the current practices of publication in terms of completeness, openness and documentation are largely maintained. Nevertheless one should keep in mind that the very fact of being defined as a High Value Dataset entails that the dataset in question has to be available in machine-readable format, via an API and at no cost, across the entire EU. Even without substantially raising the bar in terms of data formats, use of taxonomies, accompanying documentation, frequency of updates, granularity and API setup, this is an important added value in comparison to the baseline.

The benefits of such an option would undoubtedly be lower than in the higher intensity scenario, but the compliance costs would be reduced as well (NB. costs other than those which stem directly from the basic act – the Open Data Directive – and which cannot be eliminated under any of the policy options for the current initiative, namely: free availability and the establishment of APIs). The lower intensity option, despite its lower ambition, can be seen as a first element of a more gradual course of action, paving the way for the future extension and deepening of the HVDs list.

5.2.3. *The ‘wish-list’(higher intensity) option*

The ‘wish-list’ option is constructed by including a wider range of the datasets (based on the data pre-selected in the course of the support study for which the cost-benefit analysis proved to yield positive results) and more demanding modalities of publication. This option takes into account a full range of corresponding demands from re-users in terms of data inclusiveness and publication/re-use modalities and be guided by the solutions applied in the Member States considered as ‘best practice examples’ for data re-use in a given theme. From that point of view, it can be seen as a ‘wish-list’. In addition to covering a wider subject area, such datasets would also be of higher quality (in terms of metadata richness or unified formats for example), providing more immediate value to the re-users. A higher degree of legal compatibility (i.e. a use of specific standardised open licence instead of any open licence) would enhance licensing interoperability of the data.

This option would likely yield the highest benefits but would be coupled with substantial costs, notably linked to the free re-use of currently commercialised data. High costs would also be expected in terms of maintenance of a very demanding publication model (e.g. real-time availability). It would also likely require extra efforts from the public sector in cases of datasets which can be held by different public bodies (e.g. company register vs insolvency register vs register of beneficial owners; mapping agency vs cadastral office). Such data may

fall under different legislative rules and the public bodies that hold them do not always share the same historical experience and approach towards opening up their data for re-use.

5.2.4. Description of the simplified policy options for each theme (from the low-hanging fruit towards the wish-list scenario)

Theme	Policy option 1 (low hanging fruit)	Policy option 2 (re-users' wish list)
Company and company ownership data	<p><u>Scope</u> Basic information (such as name, address, legal form, identifier etc.), company documents (financial and non-financial statements, audit reports)</p> <p><u>Publication modalities</u></p> <ul style="list-style-type: none"> • CC-BY 4.0 or equivalent bespoke open license • XML format and complete metadata as csv files (for financial reports, as mandated by EU ESEF Regulation 2019/815) • Web-available documentation (incl. structure and semantics) • Use of company codes as key attributes for disambiguation. <p>Member States could include thematically relevant personal data in accordance with applicable legislation.</p>	<p><u>Scope</u> Extended information on companies (including personal data), company documents, company ownership and company insolvency status</p> <p><u>Publication modalities</u></p> <ul style="list-style-type: none"> • Open standard licence, at most CC-BY • XML and JSON formats • Full and web-available metadata and documentation • Mandatory use of ISA2 shared vocabularies • In addition to company codes, also individual beneficiary code as key attribute
Geospatial ¹¹⁵	<p><u>Scope</u> Administrative units, Place Names, Addresses, Buildings, Geometry, identification code and location of cadastral parcels</p> <p><u>Publication modalities</u></p> <ul style="list-style-type: none"> • CC-BY 4.0 licence • Format: GeoJSON, as per INSPIRE requirements • Metadata to match the INSPIRE requirements allowing transformation to geoDCAT-AP • Partial national or national coverage in terms of granularity • Key attributes as per INSPIRE 	<p><u>Scope</u> Administrative units, Place Names, Addresses, Buildings and Full cadastral Parcels</p> <p><u>Publication modalities</u></p> <ul style="list-style-type: none"> • CC0 licence • Traceability: both national open data catalog and geodata catalog • Higher granularity of data (scale 1:5000 and beyond) and high update frequency • More stringent requirements regarding key attributes (e.g. height of buildings)
Meteorological	<p><u>Scope</u> Observations, climate data (validated observations which may contain corrections from the original measurements or the removal of anomalies and are then entered into the permanent climate record), digitised structured historical climate data, and weather alerts.</p>	<p><u>Scope</u> The same as the lower intensity scenario plus Radar data and Numerical Weather Prediction (NWP) model data are added as well as available historical digital records for observations and climate data.</p>

¹¹⁵ Maps are not included as self standing datasets, since they can be obtained by layering multiple High Value datasets (e.g. Administrative Units + Place Names + Road and Rail transport networks).

	<u>Publication modalities</u> <ul style="list-style-type: none"> • CC0 or CC-BY licence • For formats, current common practice in the field is sufficient as long as it is machine-readable • Both API and bulk download • Complete metadata (in csv or xml format) and complete documentation (web available) • Traceability for weather alerts • Specific minimum update frequency • Highest available granularity and key attributes 	<u>Publication modalities</u> In addition to the lower intensity scenario, relevant similar requirements in terms of granularity, formats, frequency and attributes are required for the extended scope of datasets (radar and NWP)
Earth observation and environment	<u>Scope</u> Environmental e-reporting priority data ¹¹⁶ and Environmental Monitoring Facilities (location of monitoring stations, the parameters measured, as well as the actual spatio-temporal observation data.) <u>Publication modalities</u> <ul style="list-style-type: none"> • CC0 or CC-BY licence • Formats, metadata completeness and other quality variables as per INSPIRE • Both API and bulk download 	<u>Scope</u> The same as the lower intensity scenario plus data under INSPIRE themes relevant to Earth observation and environment, such as hydrography, land parcels, elevation or ortho-imagery. <u>Publication modalities</u> <ul style="list-style-type: none"> • The same as under the low intensity scenario and in line with INSPIRE requirements • The currently discussed DCAT-AP2 metadata standard for INSPIRE as additional recommendation
Statistics	<u>Scope</u> Dataset categories: Business statistics (industrial production, industrial producer price index, production in construction, retail trade, trade in goods, tourism), macroeconomic statistics (consumer prices, national accounts – GDP key indicators on corporations, key indicators on households, government expenditure and revenue, government debt) and social statistics (population, fertility, mortality, healthcare expenditure, poverty, inequality, employment, unemployment, potential labour force). Each dataset specification also includes minimum required key variables (e.g. gross value added, final consumption, exports and imports) as well as the precise combination of breakdowns (e.g. regional (NUTS 2 or NUTS 3), sex, age, activity) to apply for each key variable. <u>Publication modalities</u> <ul style="list-style-type: none"> • CC-BY 4.0 licence 	<u>Scope</u> The scope is identical for both options given that their intensity, especially regarding costs, is largely related to the publication modalities, not to the thematic scope of data (most statistical data can be said to have a high reuse value and are already freely available). <u>Publication modalities</u> On top of the requirements under the low intensity scenario: <ul style="list-style-type: none"> • The requirement to use controlled vocabularies and taxonomies which are DCAT-compatible • Higher degree of granularity for the metadata files (i.e. description of the statistical data, as well as descriptions of the statistical concepts, methodologies and information on data quality)

¹¹⁶ List of data sets developed by INSPIRE Maintenance and Implementation Group (MIG) and related to environmental reporting which should be made available by Member States through the European Spatial Data Infrastructure in a stepwise manner. Full description available here: <https://webgate.ec.europa.eu/fpfis/wikis/display/InspireMIG/Action+2016.5%3A+Priority+list+of+datasets+for+e-Reporting>

	<ul style="list-style-type: none"> • CSV, XML (SDMX) and JSON formats • Both API and bulk download • Complete metadata and documentation according to official statistics baseline standard (SIMS) 	
Mobility	<p><u>Scope</u> Inland waterway and river infrastructure data based on the RIS Directive (static data, dynamic data and navigational charts - Inland ECDIS). Applies only to waterways within the scope of the RIS Directive. Also: transport network datasets as per INSPIRE (road transport, rail transport, water transport, air transport and cableways data aligned with INSPIRE specifications).</p> <p><u>Publication modalities</u></p> <ul style="list-style-type: none"> • CC-BY 4.0 licence • CSV, XML or geoJSON formats • As a rule: Formats prescribed by the RIS Directive and its implementing acts; publicly documented widely used standard format. • Bulk download and API (web service API for static data) • Clear complete documentation and use of shared vocabularies • Granularity: individual waterway, National waterway network and cross-border nodes (for charts) within the scope of RIS Directive • ISRS Location Code as the key attribute • Provision in (near) real-time in order to guarantee the accuracy of RIS services • For INSPIRE data: INSPIRE requirements regarding data and metadata quality and data provision should be applied 	<p><u>Scope</u> The same as the lower intensity scenario plus all national transport network datasets i.e. not only those covered by the INSPIRE Directive.</p> <p><u>Publication modalities</u></p> <p>As in the low intensity scenario but for the datasets not covered by EU legislation (RIS or INSPIRE Directives), the INSPIRE data quality and publication standard would only be a recommendation (non-binding).</p>

5.3. Options discarded at an early stage

The baseline scenario (no action as far as the Implementing Act is concerned), presented in point 5.2.1. above is a policy option that cannot be taken for further consideration. The co-legislators made it clear that in addition to the adoption of the Open Data Directive, the Commission is not only empowered but indeed obliged to prepare and submit for the approval of the Member States a draft act with the list of HVDs.

6. WHAT ARE THE IMPACTS OF THE POLICY OPTIONS?

6.1. Economic impacts

6.1.1. Baseline scenario

As explained in Chapter 5, the ‘do nothing’ scenario needs to be discarded for legal reasons but will be used as a benchmark and a part of a contextual framework for assessing the impacts of policy options under consideration.

Even without legislative intervention, the re-use of data across sectors would increase. This is partly due to an ongoing digitisation of the public sector, technological developments that provide new uses for data resources and the expansion of the data economy in general. The availability and quality of high-value data across the six themes would therefore likely also (slowly) increase, encouraging the use of PSI by businesses, public authorities, as well as end-users (e.g. through apps). In addition, the processing of data should gradually become cheaper, facilitating market entry to more businesses. This, however, would not sufficiently address the expected growing demand for data over the coming decade¹¹⁷ and would also fall short of the Commission’s ambition to ensure a digital recovery of Europe *‘by unleashing the re-use potential of high value public datasets to stimulate competitive innovation and providing users with greater choice of digital services’*¹¹⁸. At the same time, the costs associated with the digitisation of public service delivery, including the adoption of APIs or machine readability of main public data registries, would in any case need to be borne by the public sector due to the necessary technological updates and in line with commitments made in the scope of other initiatives, including eGovernment¹¹⁹ and eJustice¹²⁰. For instance, Europe’s Green Deal incorporates a number of data-intensive actions, such as the creation of the dedicated ‘green data space’ and the Destination Earth initiative¹²¹. Finally, persisting barriers to the re-use of key public data in the future EU data spaces would make it harder to fulfil the objectives of the SME strategy, namely to bridge the gap in the exploitation of data and data-dependent technologies (e.g. AI) between SMEs and larger companies in Europe¹²².

There is limited evidence available on the value of data in Europe today and on the economic output linked to the reuse of such datasets. Therefore it is difficult to establish a precise baseline to identify potential benefits of the inclusion of certain datasets in the HVD list. Quantifying precisely the impact of the transition remains a challenge.

¹¹⁷ See: European Data Monitor tool for the differences in the growth of the data market depending on a range of factors such as data availability. <http://datalandscape.eu/european-data-market-monitoring-tool>

¹¹⁸ Commission Communication ‘Europe’s moment – Repair and Prepare for the Next Generation’, COM(2020) 456 final.

¹¹⁹ https://ec.europa.eu/info/business-economy-euro/egovernment_en

¹²⁰ <https://e-justice.europa.eu/home.do?action=home>

¹²¹ Annex to the Commission Communication ‘A European Strategy for Data, COM/2020/66 final.

¹²² SME Strategy for a sustainable and digital Europe, COM(2020) 103 final.

In line with the observations above and using the forecasts of the European Data Market Monitoring Tool¹²³, the study supporting the Impact Assessment calculated the economic value of the baseline scenario to be at EUR 48.6 billion¹²⁴ (the baseline PSI market value), with a likely increase to EUR 69.8 billion in 2028. This increase corresponds to a conservative assumption of growth at 6.5%, whereas some literature indicates an average growth rate of 7% for the PSI market or of 9% for the overall EU data economy¹²⁵. In terms of job creation, the EDP's Economic Impact of Open Data Report¹²⁶ considers that around 230 000 direct open data employees were active in the EU in 2019. The study supporting the review of the PSI Directive¹²⁷ used a different methodology and arrived at a figure of 64 000 employees directly involved in open data re-use in 2017, with a forecasted increase to 518 000 persons employed by 2027.

When it comes to benefits, it is rather difficult to estimate particular benefits corresponding to one dataset, as often their value is better expressed in their combination with other datasets. Using the framework developed with the study, the datasets were associated with a set of indicators from different macro-economic areas, considering also desk research results and the inputs received from different stakeholders.

Open Data Employment on the national level: case study
The Spanish ASEDIE association provides detailed figures on employment in the 'infomediary sector' (grouping companies from all economic sectors that re-use public sector information). The 2020 ASEDIE report ¹²⁸ includes the figure of 22 790 employees for Spain alone. It is largely consistent with the abovementioned estimates for direct open data employment in the EU presented above.

6.1.2. *Low hanging fruit policy option (PO1)*

General observations:

While still positively affecting the size of the PSI/HVDs market in comparison to the baseline, this approach would have limited benefits from the data re-user's perspective. This is because the fewer the data attributes per dataset, the poorer the insights that can be derived from it (each data attribute augments its informative potential). In addition, the lower quality and

¹²³ Data landscape, The European Data Market Monitoring Tool see: <http://datalandscape.eu/european-data-market-monitoring-tool-2018>, NB the monitoring tool uses the "Value of the Data Market" as a proxy for the direct economic value.

¹²⁴ This value already takes into account a correction linked to the Covid-19 impact on the overall EU economy.

¹²⁵ The European Data Market Monitoring Tool (see: footnote 96 above).

¹²⁶ Economic Impact of Open Data, European Data Portal, 2020.

¹²⁷ Study to support the review of Directive 2003/98/EC on the re-use of public sector information, Deloitte, 2018.

¹²⁸ <http://www.asedie.es/assets/asedie---infomediary-sector-report-2020.pdf>

interoperability of available data would likely limit the extent of re-use, given that potential re-users would need to harmonise the data at their own expense. A recent study found that, while efforts for data harmonization and cleaning is often part of the services provided by data analytics companies, it can raise the costs of big data analytics without producing visible benefits for the clients¹²⁹. The evaluation of the INSPIRE Directive also found that only in cases where interoperability requirements had been met did the actual deployment of data services in cross-border projects show gains in effectiveness and efficiency¹³⁰. It should be noted that in the case of HVDs covered by the INSPIRE Directive or of statistical data, the interoperability and quality requirements under both policy options hardly differ (they are harmonised by the INSPIRE Directive itself or by the European Statistical System) and the lower impact of this policy option is thus attributed mostly to the reduced thematic scope or dissemination obligations (e.g. frequency of updates).

On the other hand, this policy option would also limit the compliance costs for public sector bodies. This is the case for data themes in which some of the datasets falling within this policy scenario are still charged for (meteorological, geospatial, Earth observation and company data) and where the 'low-hanging fruit' option would allow the current financing models to be kept intact.

The economic impact would also suffer from reduced re-use due to imperfect legal interoperability (licensing arrangements): the 'low hanging fruit' option decreases the legal certainty for re-users, for instance in cases where the same High Value Dataset is published under a CC0¹³¹ licence in one Member State while under a CC-BY or another similar bespoke licence in another.

It might also decrease the potential for integrating high-value data into digital services due to the less stringent requirement of API dissemination by public bodies across the EU. The effects of the lower intensity scenario on employment are directly related to the expected growth of the share of the data market affected by each theme. New business opportunities require new (qualified) jobs within the information society (data scientists, data workers), leading to an indirect creation of other types of jobs.

Theme-specific observations:

For the **company and company ownership** theme, four main sectors would benefit from the initiative: the financial sector, including in particular all entities contributing to enhance access to finance for SMEs (e.g. trade finance), the business information sector (evaluating and reviewing firms), the marketing and sales sector (market research and various business analytics services), and the business publishing sector (organisations that publish and report

¹²⁹ Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, Deloitte 2017.

¹³⁰ Report on the implementation of the INSPIRE Directive, SWD(2016) 243 final/2.

¹³¹ Creative Commons 0 (public domain dedication) and Creative Commons BY (attribution) are the most commonly used open licences worldwide, including for content held by the public sector.

on company data for the purpose of improving transparency). Many stakeholders¹³² emphasise that given the fragile economic situation affected by the Covid-19 crisis and the increasing attention to business clearance by the banks, the data-driven know your customer (KYC) activities, credit checks based on accurate and updated financial statements and audit reports (to approve trade credits, especially for SMEs), or evaluating tax compliance become essential for conducting business and to enhance access to finance in the EU.

The lower intensity intervention option for **geospatial** data includes only limited requirements of data on top of those already mandated by INSPIRE, with 4-5 HVDs in scope. Geospatial data offer the widest set of opportunities for re-use because of their combinability with other datasets. The number of use-cases is almost infinite. Location data re-use has brought significant benefits to agriculture, mining and construction¹³³, and is now also used to track the epidemic and control the reopening of borders¹³⁴. This option would therefore result in a slight to medium change in the growth rate of the geospatial sector, as the support study noted a positive trend in the growth of use-cases for data within this theme.

In the **meteorological** theme, this option includes three datasets, namely observations, climate and digitised structured historical climate data. This intervention would stay close to what is already common in terms of dissemination arrangements, but leaves room for Member States to adopt more advanced solutions. Based on the experience of countries that have opened their meteorological data¹³⁵, the following major impacts can be predicted: a strong growth in both number of re-users and volume per re-user, re-use growing outside the traditional meteorological value-added services and a surge in cross-border demand for the data.

The 'low hanging fruit' option for **Earth observation and environment** HVDs implies adding an open data obligation to the environmental reporting and observation data. It is expected that, as a result, the patterns of benefits already emerging across the EU from current re-use will strengthen. These include a strong growth in both the number of re-users and the volume per re-user across all user groups (citizens, research, commercial sector, public sector), novel uses created on top of dynamic data (e.g. environmental measurements) and overall growth in economic value creation (turnover, start-ups, employment). An important expected benefit is that reducing the barriers in sourcing Earth observation and environmental data from across all Member States (e.g. EU-wide land parcel and use data) will help expand and improve the existing Copernicus services¹³⁶. Removing re-use restrictions, while retaining

¹³² As voiced by the re-users of company data and business associations in the IIA consultation and during workshops.

¹³³ <https://www.geospatialworld.net/article/economic-value-of-geospatial-data-the-great-enabler/>

¹³⁴ For examples, see: <https://www.gsa.europa.eu/GNSS4Crisis>

¹³⁵ Such as Finland, Denmark, Ireland (EU) or Norway (EEA).

¹³⁶ <https://www.copernicus.eu/en/services>

the high interoperability levels of INSPIRE, will therefore lead to a moderate change in the growth rate in comparison to the baseline (EUR 130 million to 751 million)¹³⁷.

For the highly harmonised area of **statistics**, this policy option requires only minimal changes to the current publication options available; this is expected to have little or no impact on the institutions and stakeholders concerned. The benefits of the re-use of statistical information relate primarily to improved decision-making by the public sector, business sector and individuals, leading to more efficient policies and business processes. Despite this notable potential for generating economic value, only a slight change in the GDP growth rate can be attributed to re-use of statistical data in comparison to the baseline (growth rate of 7%)¹³⁸. This is because statistical data are already widely available at no cost and generally in bulk.

The full opening of **mobility data** should benefit the group of re-users which already exploit data covered by this theme. This means economic operators in the (public/waterborne) transport sector will face reduced costs of data acquisition and processing. The provision of new added-value services by mapping and location as well as RIS service providers can also be expected. However, the lower intervention option covers only two of the four categories of datasets under consideration. This translates into a slight change in growth rate in comparison to the baseline.

The various economic impacts of the 'low hanging fruit' option per data theme are presented below in the form of an incremental rise in GDP and in resulting economic value (in millions of euros), as well as in the number of people employed and the number of companies created¹³⁹:

Table 1 – Economic impacts of PO1

Incremental rise of GDP over the baseline due to increased re-use of HVDs (expressed as resulting economic value in m€ and as % of GDP)					
	2024	2025	2026	2027	2028
Baseline	230.417	243.632	247.389	251.262	255.239
% Baseline to GDP	1,78%	1,85%	1,85%	1,85%	1,85%
Policy Package 1 (low intervention) - Incremental rise of GDP (in m€)	3.615	7.721	11.969	16.375	20.959
% Policy Package 1 to GDP - Incremental rise of GDP (in % of GDP)	0,03%	0,06%	0,09%	0,12%	0,15%

Tables 2 and 3 – impact on employment and company creation

¹³⁷ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

¹³⁸ Ibid.

¹³⁹ Ibid.

Employment - impact	2024	2025	2026	2027	2028
Baseline					
1. Company and company ownership	151.214	159.887	162.352	164.894	167.504
2. Geospatial	829.239	876.801	890.319	904.259	918.573
3. Meteorological data	487.788	515.765	523.717	531.917	540.337
4. Earth observation and environment	360.963	381.666	387.551	393.619	399.849
5. Statistics	390.230	412.612	418.974	425.534	432.269
6. Mobility	219.504	232.094	235.673	239.363	243.152
Total	2.438.938	2.578.825	2.618.586	2.659.585	2.701.684
Policy Package 1 (low intervention) - number of jobs created					
1. Company and company ownership	710	1.505	2.267	3.018	3.766
2. Geospatial	7.786	16.543	25.358	34.310	43.461
3. Meteorological data	6.870	14.631	22.591	30.780	39.241
4. Earth observation and environment	5.084	10.827	16.717	22.777	29.039
5. Statistics	1.832	3.883	5.852	7.788	9.719
6. Mobility	2.061	4.379	6.712	9.082	11.504
Total	24.344	51.768	79.497	107.756	136.730
Enterprises number - impact					
Baseline					
1. Company and company ownership	30.079	31.804	32.294	32.800	33.319
2. Geospatial	164.949	174.409	177.098	179.871	182.718
3. Meteorological data	97.029	102.594	104.176	105.807	107.481
4. Earth observation and environment	71.801	75.919	77.090	78.297	79.536
5. Statistics	77.623	82.075	83.340	84.645	85.985
6. Mobility	43.663	46.167	46.879	47.613	48.367
Total	485.143	512.968	520.878	529.033	537.407
Policy Package 1 (low intervention) - number of enterprises created					
1. Company and company ownership	141	299	451	600	749
2. Geospatial	1.549	3.291	5.044	6.825	8.645
3. Meteorological data	1.367	2.910	4.494	6.123	7.806
4. Earth observation and environment	1.011	2.154	3.325	4.531	5.776
5. Statistics	364	772	1.164	1.549	1.933
6. Mobility	410	871	1.335	1.807	2.288
Total	4.842	10.297	15.813	21.434	27.198

6.1.3. 'Wish-list' policy option (PO2)

General observations:

The 'wish list' (or higher intensity scenario) aims to maximise the benefits to re-users by imposing high standards for data interoperability and quality as well as by harmonising technical solutions for data publishing via APIs across the Member States. This would boost the economic benefits in terms of actual re-use, resulting in an increase in the offer of services and products based on high-value data, along with the related job creation. This is due to both the scope of data considered (increasing overall data supply) and the reduced barriers to re-use (facilitating the absorption of the data and multiplying possibilities for re-use, especially

for SMEs¹⁴⁰). Similar effects were observed in the Impact Assessment for the review of the PSI Directive, in which the more rigorous intervention scenario was deemed to contribute better to the creation of commercial, value-added services in the context of the overall set of public sector data.

Theme-specific observations:

For the **company and company ownership** theme, the stronger intervention option would include a wide scope of datasets and consequently lead to a moderate to major change in the growth rate in comparison to the baseline. The benefits would be high given the well-established use-cases and the fact that some of the key datasets cannot currently be re-used free of charge or are consultation-only. The openness of all datasets in this thematic area would greatly increase their re-use and have a positive impact in terms of both direct business opportunities (wide re-use) and indirect effects related to market transparency (which encourages investment), even if the company and company ownership High Value Dataset market is still relatively small. At the same time, the implementation burden for the company registers is expected to be considerable.

For **geospatial** high-value data, the higher intervention option implies more stringent requirements regarding the licences, the APIs, the granularity and key attributes. Highly interoperable high-value geodata create a ripple effect by reducing boundaries between different categories of data, thus facilitating decision-making for public administrations (as shown in the mid-term evaluation of INSPIRE). Opening up such data would transfer these benefits to private entities as well, including beyond the geospatial sector.

In terms of the **meteorological** theme, this option also incorporates unstructured historical data, radar data and numerical weather prediction model data. As all datasets are highly voluminous, this intervention option can be considered as a major stimulus for the growth rate in comparison to the baseline. Stakeholders (including data holders) interviewed by the support study all agree that there is no doubt about the sizeable benefits to the economy to be generated via the open and free re-use of such weather-related datasets. Beyond the realm of weather forecast services, meteorological datasets play a major role in transport, logistics, construction and urban planning (e.g. buildings performance simulation) as well as in other sectors.

For **Earth observation and environment** HVDs, the more rigorous and comprehensive option includes additional INSPIRE themes relevant to Earth environment and observation. The support study¹⁴¹ has found out that this theme has the particularity of involving all groups of re-users (companies, public bodies, NGOs, companies and individuals) in equal measure, with expectations of benefits by stakeholders equally spread across them. Commercial uses

¹⁴⁰ Does Marginal Cost Pricing of Public Sector Information Spur Firms Growth?, Heli Koski, The Research Institute of the Finnish Economy, 2011.

¹⁴¹ Ibid.

are typically spread over a wide variety of sectors. The existing services aimed at consumers as well as business-to-business and business-to-government relations (to aid in decision-making) indicate a high economic impact of this policy scenario, implying a moderate to major change in comparison to the baseline.

For **statistics**, the 'wish-list' option consists of improvements in terms of measures for publication, since the vast majority of statistical data of any kind is already widely available. The level of granularity of information and increased flexibility in how re-users can work with statistical data will have a decisive impact on the benefits brought forward by re-use. These factors can also attract new users that previously have not used or re-used such datasets. For instance, a survey of Austrian small shop owners revealed their high use of statistical demographic and location-based data to optimise the commercial offer of their businesses¹⁴². Better APIs and metadata files, and development of controlled vocabularies and taxonomies are thus expected to lead to a moderate change in comparison to the baseline, also because the market of high-value data in statistics is quite advanced relative to the other thematic areas.

In the **mobility** theme, this intervention option includes all four datasets pre-selected in the course of the support study. It is expected that the formalisation of the obligation to provide these data for free, in harmonised and machine-readable formats and both through bulk download and APIs will significantly facilitate the provision of seamless, EU-wide information services. The Member States consulted in the course of the study declared receiving requests from re-users to further facilitate the access and re-use of these datasets throughout the EU, notably in order to leverage these in various mobile applications and software. A significant impact on growth can be expected.

Table 4: Economic impacts of PO2 (in terms of incremental rise in GDP and resulting economic value))¹⁴³

Incremental rise of GDP over the baseline due to increased re-use of HVDs (expressed as resulting economic value in m€ and as % of GDP)					
	2024	2025	2026	2027	2028
Baseline	230.417	243.632	247.389	251.262	255.239
% <i>Baseline to GDP</i>	1,78%	1,85%	1,85%	1,85%	1,85%
Policy Package 2 (higher intervention) - Incremental rise of GDP (in m€)	4.649	9.937	15.559	21.418	27.538
% <i>Policy Package 2 to GDP - Incremental rise of GDP (in % of GDP)</i>	0,04%	0,08%	0,12%	0,16%	0,20%

¹⁴² Barbara Huber, Alexander Kurnikowski, Stephanie Müller, Stefan Pozar, "The Economic and Political Dimension of Open Government Data in Austria," Institute for Entrepreneurship & Innovation, WU Vienna University of Economics and Business, Spring 2013.

¹⁴³ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

Tables 5 and 6 – impact on employment and company creation

Employment - impact	2024	2025	2026	2027	2028
Baseline					
1. Company and company ownership	151.214	159.887	162.352	164.894	167.504
2. Geospatial	829.239	876.801	890.319	904.259	918.573
3. Meteorological data	487.788	515.765	523.717	531.917	540.337
4. Earth observation and environment	360.963	381.666	387.551	393.619	399.849
5. Statistics	390.230	412.612	418.974	425.534	432.269
6. Mobility	219.504	232.094	235.673	239.363	243.152
Total	2.438.938	2.578.825	2.618.586	2.659.585	2.701.684
Policy Package 2 (higher intervention) - number of jobs created					
1. Company and company ownership	2.840	6.062	9.405	12.874	16.486
2. Geospatial	15.573	33.241	51.573	70.600	90.407
3. Meteorological data	13.740	29.466	46.050	63.490	81.863
4. Earth observation and environment	8.473	18.129	28.236	38.800	49.866
5. Statistics	5.496	11.705	18.072	24.624	31.393
6. Mobility	3.092	6.584	11.359	16.315	21.475
Total	49.214	105.186	164.695	226.703	291.489
Enterprises number - impact					
Baseline					
1. Company and company ownership	30.079	31.804	32.294	32.800	33.319
2. Geospatial	164.949	174.409	177.098	179.871	182.718
3. Meteorological data	97.029	102.594	104.176	105.807	107.481
4. Earth observation and environment	71.801	75.919	77.090	78.297	79.536
5. Statistics	77.623	82.075	83.340	84.645	85.985
6. Mobility	43.663	46.167	46.879	47.613	48.367
Total	485.143	512.968	520.878	529.033	537.407
Policy Package 2 (higher intervention) - number of enterprises created					
1. Company and company ownership	565	1.206	1.871	2.561	3.279
2. Geospatial	3.098	6.612	10.259	14.043	17.983
3. Meteorological data	2.733	5.861	9.160	12.629	16.284
4. Earth observation and environment	1.685	3.606	5.617	7.718	9.919
5. Statistics	1.093	2.328	3.595	4.898	6.245
6. Mobility	615	1.310	2.260	3.245	4.272
Total	9.789	20.923	32.760	45.095	57.982

6.2. Social and environmental impacts

Many countries, both in the EU and beyond, have made environmental and social gains the cornerstone of their open data strategies¹⁴⁴. Studies have shown that these kinds of benefits, while often difficult to quantify, are a stable feature of open data initiatives and can make a real contribution to society, including on the local level¹⁴⁵. In the context of the HVDs, several non-economic impacts can be expected.

¹⁴⁴ E.g. in the framework of the International Open Data Charter or Open Government Partnership.

¹⁴⁵ Pereira, G.V., Macadar, M.A., Luciano, E.M. et al. Delivering public value through open government data initiatives in a Smart City context. Inf Syst Front 19, 213–229 (2017).

For instance, a wider availability of information on companies has clear social benefits for areas such as fighting crime (incl. financial crime), increased public engagement, understanding of economic processes and government accountability (e.g. in public support to economic actors). Beneficial ownership information in particular, is crucial for ensuring transparency of business operations and for implementing anti-corruption and financial crime rules. The opening up of companies' accounts and reports would give citizens a much better overview of companies' records in terms of their real environmental impact; this would facilitate the shifting of consumer preferences towards more environmentally friendly businesses. Public sector authorities would also benefit from increased availability of information concerning companies' environmental actions, which would facilitate decisions in the context of the COVID-19 recovery plans¹⁴⁶.

There is also a clear social and environmental dimension for geospatial datasets, which act as reference data for a variety of data from other fields. The need for a spatial element is evident when planning and implementing any environmental initiative, such as reduction of air pollution in urban areas, an overhaul of transport infrastructure or assessing the pace of climate change for specific geographical units. The recent initiatives on stemming the COVID-19 epidemic have also shown the importance of geolocalising health institutions, events and populations¹⁴⁷.

Meteorological data, especially historical weather readings covered in the high intensity scenario, are instrumental in furthering scientific research on climate change and in facilitating policy response to the environmental crisis. Re-use of high value data in this thematic field will lead to significant social and environmental benefits of global and regional importance¹⁴⁸ (e.g. sustainable farming, prediction of natural disasters, transport, logistics, resource management or tourism).

Important societal gains (e.g. reduced air and water contamination, animal and plant conservation) can also be expected due to an increasing application of modern data analytics technologies for an efficient processing of large volumes of dynamic Earth observation and environmental data included in the scope of this initiative¹⁴⁹. For instance, open data tools such as Aqueduct¹⁵⁰ or JRC's flood hazard maps¹⁵¹, can help reduce the extent of water damage in the EU. Since 2000, floods in Europe have caused at least 700 deaths, the displacement of about half a million people and at least EUR 25 billion in insured economic losses¹⁵². Wider use of open data services including Denmark's Modstroem

¹⁴⁶ Some countries have established conditionalities for companies receiving funding under COVID 19 specific recovery plan, including environmental actions and CO₂ emission reduction.

¹⁴⁷ <https://blogs.worldbank.org/eastasiapacific/role-geospatial-information-confronting-covid-19-learning-korea>

¹⁴⁸ A wide range of economic benefits have already been attributed to an open re-use of data produced by the US National Oceanic and Atmospheric Administration (NOAA). <https://www.oecd.org/sti/ieconomy/40066192.pdf>

¹⁴⁹ <https://eponline.com/articles/2020/03/25/using-big-data-technology-for-environmental-protection.aspx>

¹⁵⁰ <https://www.wri.org/aqueduct>

¹⁵¹ <https://data.jrc.ec.europa.eu/collection/floods>

¹⁵² <https://www.eea.europa.eu/archived/archived-content-water-topic/water-resources/floods>

Energiberegner¹⁵³, GreenHome¹⁵⁴ in the Netherlands or Sunenergia¹⁵⁵ in Finland can also potentially lead to 5.8 million tonnes of oil equivalent in savings EU wide by helping to reduce household energy consumption¹⁵⁶. Past experience at local level (16% less gas, electricity, oil and transport fuel consumption by a local council in the UK thanks to publishing real-time data on energy use¹⁵⁷) confirm these assumptions.

Statistical data are intentionally produced to maximise their benefit for the society, ensuring equal access to information for all. Statistical information helps promote an open and empowered society and a more effective and accountable decision making.¹⁵⁸ For instance, reliable statistics on population and housing make it possible to plan and better target government services, while access to data on government spending makes politicians more accountable to the citizens and media¹⁵⁹.

The listing of mobility data within the scope of this Impact Assessment could help reduce of CO₂ emissions through a more efficient use of routes and multimodal mobility¹⁶⁰, data-driven policy (e.g.in the context of smart cities)¹⁶¹ as well as in terms of increased safety and security in the (public/waterborne) transport. Studies suggest that approximately 500 - 730 million hours could be saved each year by European drivers commuting to and from work in urban areas. In terms of safely, a better use of spatial, mapping and mobility data, enabling the emergency services to arrive at the scene of an incident just one minute faster, could save up to 54 – 202 thousand lives across the EU^{162,163}.

The differences in social and environmental impact between the lower and higher intensity policy option are difficult to quantify but it is expected that a wider supply of higher quality data combined with its better usability could play a role, notably where the use case requires the processing of wide and voluminous datasets. The differences would therefore be broadly in line with those applicable to the economic impact but of varying magnitude depending on the data theme (i.e. strongest divergence between policy options in terms of environmental and social impact are likely to be observed in the case of meteorological and Earth observation data or statistics – the typical examples of ‘big data’ datasets, the value of which is directly proportional to their size).

¹⁵³ <https://www.modstroem.dk/energiberegner/>

¹⁵⁴ www.greenhome.nl

¹⁵⁵ www.sunenergia.com/en

¹⁵⁶ European Data Portal, the Economic Impact of Open Data, 2020.

¹⁵⁷ http://www.rbwm.gov.uk/web/news_10913_energy_reduction_sustainability_strategy.htm

¹⁵⁸ UNECE, *Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics* (New York and Geneva: United Nations, 2018).

¹⁵⁹ As example: <https://www.transparency.org/en/publications/open-data-and-the-fight-against-corruption-in-germany>

¹⁶⁰ https://www.researchgate.net/publication/323940830_Big_Data_for_Sustainable_Urban_Transport

¹⁶¹ <https://www.tandfonline.com/doi/full/10.1080/09654313.2017.1294149>

¹⁶² <https://www.7wdata.be/open-data-institute/how-open-data-can-help-save-lives/>

¹⁶³ European Data Portal, the Economic Impact of Open Data, 2020.

6.3. Impact on SMEs

The initiative, under any of the policy scenarios, will have a positive effect on SMEs. This is related to the particular distribution of costs and benefits in any open data initiative (see also point 6.4 below) and has been confirmed by the stakeholders as well as the impact predictions of the support study¹⁶⁴. The initiative should not only increase the efficiency of existing SMEs and help them develop new business cases based on data but should also lead to the creation of additional enterprises. In line with the methodology of the support study¹⁶⁵, one can predict a positive trend in the creation of new companies for each of the themes under consideration. In the case of the lower intensity scenario, this would range from fewer than 1000 new companies being created in 2028 for the smaller sectors (such as the one based on the re-use of company data) to over 8 500 in the geospatial domain. The corresponding figures for the higher intensity scenario would range from 3 300 to 18 000 respectively. Apart from supporting the creation of new companies, HVDs can become an important enabler for start-ups for the validation of their business case and attracting investors¹⁶⁶. An EU incubator for data start-ups discovered¹⁶⁷ a positive correlation between strong and proactive open data policies in Member States, and the number of successful applicants from those countries. It also observed that start-ups and SMEs typically leveraged two or more types of open data, with a strong concentration of interest in geospatial/mapping and environmental data (data which BigTech companies already possess and which are therefore likely to be exploited chiefly by SMEs). Finally, open data can also empower SMEs indirectly, as it creates opportunities for EU companies that become so skilled in handling it that they start to offer home-made solutions for the management and processing of public data to others¹⁶⁸. Annex 5 to this Impact Assessment discusses the issue of impact of HVDs on SMEs in more detail, including a number of exemplary use cases.

6.4. Costs

Before assigning the cost values to individual data themes, it is useful to draw attention to the common pattern in the spread of costs and benefits in all open data initiatives: whereas the benefits associated with open data affect the re-users as well as public data holders (public bodies can be both – data suppliers and consumers), the costs fall exclusively on the side of the data holders (public sector). They are ultimately the entities that need to put in place data management processes that cater for the publishing of public sector information, maintain the technical infrastructure and ensure the necessary skills needed within the organisation.

¹⁶⁴ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

¹⁶⁵ Weighted coefficient for the EU27 ICT-service sector, representing the ratio of enterprises per GVA/GDP. For the weighted coefficient it was assumed that an average firm in the EU27 ICT-sector has 6 employees, respectively a statistical ratio of ca. 2 enterprises per 1 million € GVA.

¹⁶⁶ <https://www.sme10x.com/10x-industry/how-open-data-can-help-startups-succeed>

¹⁶⁷ Assessment of the impact of the ODINE data incubator project, 2017.

¹⁶⁸ A telling example is OpenDataSoft, a French open data start-up that since its creation has received multimillion funding from private investors and offers services to over 280 clients worldwide.

The costs associated with the publishing of a unit of public sector information nevertheless, tend to diminish over time. This trend can be explained by the maturing of recently novel technologies, the decreasing price of various data critical components and services, such as sensors and cloud storage and finally, an increasing competence of civil servants in the collecting and processing of data¹⁶⁹. In addition, an open data approach involves the adoption of more efficient data management processes which become less resource intensive (due to more frequent machine to machine interaction)¹⁷⁰. The 2018 Impact Assessment study attributed a notable drop in the costs of opening up data across the EU to the adoption of the amending Directive 2013/37/EU¹⁷¹ and the resulting efficiency gains in the public sector. The initial costs of adopting an open data approach (such as establishing an API) are therefore often ‘one-off’ costs which are not only offset by ensuing efficiency gains but can even reduce the operating costs in the long run¹⁷².

It is important to point out that the current initiative, unlike earlier legislative interventions in the area of PSI/open data, will have the possibility to offset many of the ensuing costs by the accompanying non-legislative measures, in the form of funding foreseen under the Digital Europe Programme, specifically with the purpose of facilitating both API build up and its interoperability in the context of the publishing of HVDs. It would therefore target all cost elements which are not directly related to the extent of the thematic scope of the intervention or the requirement of free availability. Other flanking measures likely to mitigate costs include the actions on promoting interoperability of APIs in the public sector under the European Interoperability Framework¹⁷³ and the EU-supported FIWARE platform which offers free and open source Standard APIs for data management and exchange, as well as harmonised data models¹⁷⁴.

Costs per theme:

In order to estimate the range of costs, the number of datasets in scope for each intervention option was determined to allow for an estimate of the number of data holders involved. For the majority of cases considered, the available information does not permit to ascertain the costs on individual institutional level. The support study shows that such information is often missing or only partly available (e.g. when the data holder has not performed such a cost assessment itself) or the costs related to the re-use of high value datasets are only reported as

¹⁶⁹ As testified by the improving DESI indicators: <https://ec.europa.eu/digital-single-market/en/desi>.

¹⁷⁰ For instance, the Impact Assessment Support Study of 2018 associated the lowest costs for the public sector with the policy option of highest intensity (which was not retained).

¹⁷¹ Impact Assessment Support Study, Deloitte, SMART 2017/0061, figure 42.

¹⁷² Examples of costs reductions due to increased efficiency have been discussed e.g. in the following study:

https://www.w3.org/2013/share-psi/wiki/images/6/67/Impact_of_Open_Data_in_the_Public_Sector_Koski_2015.pdf

¹⁷³ COM(2017) 134 final. Even if not specific for APIs, the EIF provides public administrations with a set of recommendations to improve governance of their interoperability activities, establish cross-organisational relationships and streamline processes supporting end-to-end digital services.

¹⁷⁴ <https://www.fiware.org/about-us/>

part of a more general cost item (e.g. IT development, data management). In such cases, a categorisation of impact is proposed (low means losses up to 5% of the overall budget of the data holder, medium losses are 5-25% respectively, high losses consist in 25-45% of the overall budget for the provision of HVD of the data holder and very high losses are losses above 45% of the overall budget).

The main cost drivers for the thematic area **company and company ownership** can be categorised into infrastructural costs, data transformation costs, operational costs, other costs and lost income for data suppliers. According to the information available, the highest infrastructural costs (one-off costs) would be due to setting up an API. Costs estimation regarding the set-up of an API range between EUR 30 000 and EUR 2.5 million (for very sophisticated API solutions). Data from company registers are however not dynamic (i.e. subject to very frequent or real-time updates), which makes the use of the most expensive API solutions unnecessary in most cases. The average cost of a one-off setting up of an API is estimated at EUR 50 000¹⁷⁵. Given that 3/4 of the EU Member States would need to set-up an API under policy option 1, the indicative cost for the 75% out of the EU 27 Member States covered by the support study, based on the average costs for an API (EUR 50 000), amounts to EUR 1 million.

In addition, annual operational costs mainly related to data updates, replies to user requests, corrections of errors in the datasets and similar can also be quantified to 4 to 10 FTEs for both intervention options (amounting to between EUR 3.2 million and EUR 16 million for the remaining (75%) EU 27)¹⁷⁶. For the 'low hanging fruit' policy option, the lower figure is more likely. Admittedly, revenues from charging are substantial for several Member States (e.g. as much as EUR 50 million per year in case of the Netherlands¹⁷⁷). However, under the lower intensity option, the vast majority of the revenue would be retained as the data values in scope (such as name, address, legal form, identifier etc.) are not the ones usually charged for by the data holders (such as information on directors and shareholders)¹⁷⁸. Furthermore, the retained option will not impact on the significant revenues from the paid services consisting in provision of official documents having a legal value, since only company registers are authorised to provide such services. Only a fraction of the total revenue would therefore be lost as not all datasets in scope would be considered as HVDs. Company registers would still be able to charge for some of the data points, and collect revenue.

With regard to **geospatial datasets** the loss of revenues is a key cost driver but this is specific to some Member States only. As an example, revenues from the distribution of official basic

¹⁷⁵ Study to support the review of Directive 2003/98/EC on the re-use of public sector information, Deloitte, 2018, page 409.

¹⁷⁶ Many of these costs are however not specific to the High Value Datasets but rather to any type of online dissemination of data by the company registers, already happening.

¹⁷⁷ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

¹⁷⁸ Ibid, Section 3.1.1.

geodata in Bavaria (Germany) amount to approximately EUR 14 million annually (approx. EUR 4 million from datasets of the National Survey, approx. EUR 10 million from datasets of the Real Estate Cadastre¹⁷⁹). With regard to cadastral parcels, the German Laender Hamburg and Berlin are in the range of EUR 1 million in terms of loss of revenue. The estimated revenues in Ireland are in the range of EUR 4-5.5 million per annum, of these approximately EUR 1.5 million for opening addresses. For Sweden, costs related to opening up of geospatial data are estimated at SEK 553 million (EUR 50 million). Out of these SEK 553 million, SEK 511 million are losses of income from selling datasets and 42 million are investments and other costs that are needed to provide the data according to HVD regulations. Data transformation costs or costs linked to improving the infrastructure might be applicable but are difficult to quantify. On the other hand, for countries such as Spain, Netherlands, Italy, Portugal Denmark or Estonia the current high level of data publication means that very moderate costs would be necessary to bring data dissemination to the level intended for the HVDs.

Regarding **meteorological data**, there is one data holder per Member State. For the lower intensity option, the costs are mostly determined by the loss of revenue, whereas the higher intensity also adds costs related to e.g. cloud storage for voluminous data or extra labour for improving data quality). Currently, 13 Member States charge for observations data (with Denmark in transition period), while 9 have an open data policy. The loss of revenue is in relative terms very high for Hungary and Austria as their budget depends heavily on it. In the higher intervention option, the costs grow due to APIs and potentially also due to labour intensiveness of quality controlled data. As radar and numerical weather prediction data can be quite voluminous, they may require heavier infrastructure. In consequence, the costs for the following Member States are expected to rise: Croatia will have low-medium instead of low costs, France medium-high instead of medium costs, Finland, Ireland, Luxembourg and Greece will face high instead of medium costs, and Poland will face medium-high instead of medium costs.

For the theme **Earth observation and environment**, loss of revenue will be an important cost factor as earth observation data sets like land parcels, land use, ortho-imagery, elevation models are often not free to re-use. The higher intervention option encompasses a wider scope of data across INSPIRE themes, with at least 3-4 national entities involved in each Member State and two cost drivers: firstly infrastructure and API related cost and secondly, revenue loss. Of the 11 Member States where information about costs was available, five countries (Cyprus, Denmark, Estonia, Finland and The Netherlands) would have low costs and six countries (Austria, Germany, Lithuania, Malta, Poland and Sweden) would have medium costs. Member States generally report that costs would be important if HVD added requirements for data provisioning, data structure, and metadata significantly differing from

¹⁷⁹ Ibid.

INSPIRE requirements. However, the policy options build upon the common practices that have been informed by INSPIRE.

For **statistics**, the main stakeholder affected in the Member States will be the National Statistical Institutes. It is assumed that the provision of API and bulk download is associated with medium to high costs in both intervention options. Almost negligible as little to no costs occur for both intervention options are licence and terms of use for CC BY, the provision of an open format and the metadata and documentation.

Total 11 of the 27 EU Member States would need to set-up an API in both policy options whereas around one –third of the national statistical offices are fully ready to disseminate HVDs. Based on the average cost of API mentioned above, the total costs of at least EUR 550 000 might be envisaged.

For the **mobility data**, the main cost driver is setting up an API. For inland waterways, 8 out of 21 Member States would be affected. Based on the average costs (EUR 50 000 per API set-up), costs of EUR 400 000 could be expected.

6.5. Stakeholders' views

As presented in more detail in Annex 2, the open public consultation confirmed the overall support of all groups of stakeholders for the main elements of the intervention. The majority of the 761 respondents (82%) across all stakeholder groups strongly agreed or somewhat agreed with the statement: “The establishment of a list of high-value datasets, to be made available free of charge, without restrictions and via APIs, is a good way to ensure that public sector data has a positive impact on the EU's economy and society.” Furthermore, more than half of the respondents strongly agreed that the above mentioned elements have positive impact on the EU's economy, and only approx. Close to 7% of the respondents disagreed (strongly disagree or somewhat disagree) with this statement, representing a mix of different stakeholders categories.

Survey respondents were also consulted about other possible technical and legal arrangements to improve the re-usability of high-value data. In this area, the importance of ensuring the legal compatibility of data gathered overwhelming support, since 80% of the respondents found licensing and other terms applicable to re-use very relevant or relevant for enhancing re-use.

Other considerations often voiced in the replies (e.g. by Member States) included the need to reconcile personal data protection requirements (data anonymisation and pseudonymisation, GDPR rules.) with the re-use obligations for high value data and the need to ensure data quality aspects (e.g. granularity, completeness, accuracy, and timeliness).

The separately-held consultation on the Inception Impact Assessment¹⁸⁰ yielded comparable results. As predicted, it presented a split of opinions between the two main groups of stakeholders which are affected differently by the HVDs initiative (and indeed, any open data initiative, given that re-users are their main beneficiaries while public authorities bear the costs). As a result, the public sector was cautious with regard to the requirement to provide datasets of high value for free, often arguing that high-quality comes at a price. The discussions within the focus groups convened by the IA support study and the results of stakeholder consultations show that the re-users' 'wish-lists' as regards the inclusiveness of the datasets are wider than what the public sector bodies are ready to concede, perhaps with the exception of statistical information. At the same time, even for the themes where some high value data continues to be charged for (meteorology or geospatial), the study reports a consensus among both data holders and data users as to the substantial socio-economics to be gained from a fully open re-use. The discussion seems to no longer be about 'why' data should be open but how to ensure that this is done in a sustainable manner for the affected data holders¹⁸¹.

The public authorities were also sceptical about the possibility of aligning the regime of open re-use with that of personal data protection. The latter was very often mentioned in relation to the opening up of high value data from company registers. Finally, while the public data holders generally do not question the important socio-economic potential linked to the opening of High Value Data, they often ask for increased EU-level efforts (in terms of financial and organisational support) in addition to legislation. Re-users on the other hand emphasised the deficiencies of the current open data regime in Europe such as sub-optimal levels of re-use due to low harmonisation of data both in terms of their original quality and dissemination modalities. Associations representing smaller businesses and start-ups (commercial re-users) called for both doing away with existing barriers and possibly extending the scope of the HVDs in the future (e.g. legal information from courts).

The replies provided by stakeholder workshops organised in the context of the IA support study echoed these findings and further clarified the preferences of various stakeholder groups for the options of each data theme. Accordingly, the stakeholders generally agreed with the description of the policy options and the modulation between the low and high intensity intervention for each theme (over 70% in case of company and company ownership data to 90% in the area of geospatial data).

¹⁸⁰ 50 replies, gathered between 28 July 2020 and 25 August 2020.

¹⁸¹ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

7. HOW DO THE OPTIONS COMPARE?

In line with the European Commission's Better Regulation Guidelines, a Multi-Criteria Analysis (MCA) has been carried out in full detail in order to take full account of the complexity of the subject matter and the level of granularity of the analyses carried out.

The Commission services were supported by an external contractor to develop the ratings of the potential impacts of the baseline scenario and the Policy Options¹⁸². They were, to the necessary extent, also based on the Impact Assessment for the review of the PSI Directive, performed in 2018 (especially on the effects of the baseline scenario). According to this analysis, PO1 (lower intensity option) scores across thematic categories typically highest on the criterion of feasibility, PO2 (higher intensity option) scores across thematic categories typically highest on effectiveness. On the remaining criteria, neither PO1 nor PO2 scored consistently better across all the data themes. As a result, for three thematic categories PO1 was rated the best option, while for the other three thematic categories it was PO2. This shows the necessity to compose an optimised (mixed) intensity package, which consists of the preferred option in each of the thematic categories.

The table below gives an illustrative overview of how the options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality across all themes. The factors which were decisive in the choice of a given policy option for each theme are highlighted.

	PO1 (low-hanging fruit)	PO2 (re-users wish-list)	Optimised intensity package
EFFICIENCY	The lower intensity option will bring positive benefits compared to the baseline. In categories companies and company ownership and geospatial, their formal inclusion as open data will unleash significant benefits thanks to their better reusability, even though a limited scope of datasets is chosen. This implies that economic benefits, innovation, social benefits, and environmental and climate change benefits can be expected, albeit lower than in the case of	The benefits under the higher intensity intervention are expected to be higher , as all the datasets in consideration in individual thematic areas are in scope. On the other hand, this option is expected to generate considerable costs, particularly related to infrastructural investments (i.e. setting up APIs), and the update frequency and timeliness of publication. In categories such as companies and company ownership, but also for geospatial or mobility data, this level of policy would require a great deal of	This intervention presents a favourable ratio of costs and benefits. The mixed intensity intervention is designed to avoid situations where the expected costs to the Member States would create an obstacle in terms of feasibility. Building on existing infrastructure and practice, such as INSPIRE, plus expanding the download facilities and removing restrictions is an efficient step. The changes proposed will significantly increase the benefits as discoverability and re-usability of the

¹⁸² See: Annex 4.

	the higher intensity option. In categories such as statistics or Earth observation and environment, the proposed changes to the current situation are minimal, therefore the increase of benefits will remain limited.	reorganisation of the data publication and transformation processes.	datasets will be improved.
EFFECTIVENESS	Limited scope of datasets or their level of granularity and/or moderate measures for publication of data will result in a limited effectiveness of this option in reaching the policy objectives of this initiative, in comparison to the higher intensity option.	The higher intensity option is typically broader in terms of scope and imposes more demanding measures for publication, which will considerably increase effectiveness, i.e. facilitate re-use of a relatively wide array of quality public data.	The mixed intensity intervention will ensure a good level of effectiveness, at the same time reducing the necessity to adjust current rules in Member States on management of datasets and not putting at risk the economic sustainability of bodies that currently charge for the datasets.
COHERENCE	The lower intensity option is in principle coherent with the current legal framework. It includes data required to be made public (although not necessarily in an open data format and in a re-use friendly way) by the current framework. There may still be inconsistencies with the national laws governing the cadastral data.	Some incoherence issues might arise at the national level, in particular in the categories companies and company ownership, geospatial or mobility , in cases when the selected data are controlled by different entities that follow diverging EU level sectoral rules concerning accessibility and publication of data.	The mixed intensity intervention will ensure coherence between following the objectives of this initiative set out in the Open Data Directive and respecting objectives set out by sectoral rules applicable to different categories of high-value datasets.
LEGAL/POLITICAL FEASIBILITY	The policy option is regarded as feasible to a great extent, as it encompasses datasets which are largely already provided across Member States. Moreover, this option aims at minimising personal data in its scope, which is a sensitive point for some Member States. Data publication obligations such as implementing APIs is considered legally and politically feasible.	This option may meet serious obstacles, since it may be necessary to change some national rules concerning accessibility and publication of data. Furthermore, in particular in the category companies and company ownership, some Member States oppose making personal data available as open data. This option would also mean that some public sector bodies would lose revenues from charges for data, which may be politically difficult to carry through. More demanding measures for publication would also require substantial investments and	This intervention takes into consideration the most politically and legally sensitive issues, such as the question of including datasets containing personal data. In this option, Member States could include thematically relevant personal data in accordance with applicable legislation. National government and EU level support will be needed in some cases to strengthen technical capabilities and resources needed for infrastructure investments and for replacing revenues from

		policy changes by Member States in the categories of geospatial or mobility data.	charges currently applied for making some datasets available for re-use.
PROPORTIONALITY	The lower intensity intervention is largely proportional to the objectives, bearing in mind that in some categories such as meteorological or statistics the benefits of this option will be modest. It includes the most necessary datasets to facilitate their reuse, while achieving the objectives of this initiative in most categories of data, albeit only partially.	The higher intensity regulatory intervention is considered as disproportionate in view of the objectives of this initiative for several categories of data (in particular data within the company data, geospatial and mobility data themes) , where the intervention would require substantial adjustment efforts by a number of public sector bodies. It would also require sometimes significant investment by many public sector bodies in the update of their IT infrastructure. The risk of refusal of more substantial changes in some thematic categories in some Member State appears in disproportion to the objectives sought.	This intervention is designed to minimise negative impact of the initiative perceived by some Member States. It presents a balanced yet focused policy intervention with an intensity proportional to its objectives. By targeting the new requirements to the areas where change is beneficial, but still feasible, it will avoid disproportionate compliance burden.

This analysis translates into the following overview:

	Efficiency	Effectiveness	Coherence	Legal/political feasibility	Proportionality
Regulatory intervention with lower intensity PO1	+-	+-	+-	+	+-
Regulatory intervention with higher intensity PO2	+-	++	+-	-	+-
Regulatory intervention with mixed intensity	+	+	+	+	+

8. PREFERRED OPTION

8.1. Optimised intensity package

Based on the evidence presented above, an optimised package, combining elements of the two different degrees of regulatory interventions that are best suited for each data category, is the preferred approach.

The optimised (mixed) intensity intervention is preferred, as it provides the most favourable combination of the five criteria applied, depending on the thematic category. It is best suited for a regulatory initiative that affects six different thematic fields governed by different sets of EU and national rules, as well as different political approaches and traditions across the EU Member States, with corresponding varying degrees of harmonisation and interoperability.

Preferred PO per theme (Equal Weight 0.20)	
Company & company ownership	PO1
Geospatial data	PO1
Meteorological data	PO2
Earth observation & Environment	PO2
Statistics	PO2
Mobility	PO1

The concrete content of the preferred intervention is described in the table below. The final description of the datasets and the corresponding publication measures has, in addition to the multi criteria analysis, been subject to a verification process (involving various Commission services) allowing for slight adjustments to e.g. align the wording with the applicable legislation or practices as well as to avoid a clash of conflicting legal provisions or unrealistic expectations with regard to the short-term adoption of certain data formats.

Table 7 – HVDs and re-use arrangements for the preferred option

Thematic field	Scope of the datasets	Formats, metadata	API and bulk download	Licensing
Companies and Company ownership PO1	Basic information (such as name, address, legal form, identifier etc.), company documents	XML format and complete metadata in csv format (where feasible, i.e. excluding documents kept	Both API and bulk download. Open, standard-based documented APIs.	CC-BY 4.0 or equivalent or less restrictive open licence. Terms of use concerning personal data

Thematic field	Scope of the datasets	Formats, metadata	API and bulk download	Licensing
	(financial and non-financial statements and accounts, audit reports); limited inclusion of identifiable personal data, taking into account data protection rules.	in formats as filed by companies); web-available documentation (incl. structure and semantics), use of formats for financial reports as mandated by EU ESEF Regulation 2019/815		and registration.
Geospatial PO1	Administrative units, Geographical names, Addresses, Buildings, cadastral parcels (geometry, identification code and location) Partial national or national (all generalisation levels available) coverage in terms of granularity. Key attributes as per INSPIRE.	Publicly documented widely used standard format. Metadata to match the INSPIRE requirements allowing transformation to geoDCAT-AP.	Bulk download; INSPIRE distribution services; Open, standard-based documented APIs.	CC-BY 4.0 or equivalent or less restrictive open licence.
Meteorological PO2	Observations, climate data (validated observations, digitised structured and unstructured historical climate data), weather alerts,	Formats: current common practice in the field as long as it is machine-readable and if it is a publicly documented widely used standard format.	Both API and bulk download. Open, standard-based documented APIs.	CC0 or CC-BY4.0 or equivalent or less restrictive open licence.

Thematic field	Scope of the datasets	Formats, metadata	API and bulk download	Licensing
	Radar data and Numerical Weather Prediction (NWP) model data.	Complete metadata (in csv or xml format) and complete documentation (web available).		
Earth Observation and Environment PO2	Environmental e-reporting priority data and Environmental Monitoring Facilities (location of monitoring stations, the parameters measured, as well as the actual spatio-temporal observation data). Data under INSPIRE themes relevant to earth observation and environment, such as hydrography, land parcels, elevation or ortho-imagery.	Formats, metadata completeness and other quality variables as per INSPIRE. As a rule: publicly documented widely used standard format.	Both API and bulk download. Open, standard-based documented APIs.	CC0 or CC-BY 4.0 or equivalent or less restrictive open licence.
Statistics PO2	Dataset categories: Dataset categories: Business statistics (industrial production, industrial producer price index,	CSV, XML (SDMX) and JSON formats. Complete metadata and documentation according to official statistics baseline standard (SIMS).	Both API and bulk download. Open, standard-based documented APIs.	CC-BY 4.0 licence or equivalent or less restrictive open licence.

Thematic field	Scope of the datasets	Formats, metadata	API and bulk download	Licensing
	production in construction, retail trade, trade in goods, tourism), macroeconomic statistics (consumer prices, national accounts – GDP key indicators on corporations, key indicators on households, government expenditure and revenue, government debt) and social statistics (population, fertility, mortality, healthcare expenditure, poverty, inequality, employment, unemployment, potential labour force). Each dataset specification also includes minimum required key variables (e.g. gross value added, final consumption, exports and imports) as well as the precise combination of breakdowns (e.g. regional	Use of DCAT-compatible controlled vocabularies and taxonomies.		

Thematic field	Scope of the datasets	Formats, metadata	API and bulk download	Licensing
	(NUTS 2 or NUTS 3), sex, age, activity) to apply for each key variable.			
Mobility PO1	Inland waterways and river infrastructure data based on the RIS Directive (static data, dynamic data and navigational charts -Inland ECDIS). Applies only to waterways within the scope of the RIS Directive. Transport networks datasets as per INSPIRE (road transport, rail transport, water transport, air transport and cableways data aligned with INSPIRE specifications).	As defined in the RIS Directive. GML, GeoPackage, geoJSON. As a rule: publicly documented widely used standard format. Clear complete documentation and use of shared vocabularies. For INSPIRE data: INSPIRE requirements regarding data and metadata quality and data provision. For inland waterways: formats prescribed by the RIS Directive and its implementing acts;	Bulk download and API (web service API for static data). Open, standard-based documented APIs.	CC-BY 4.0 licence or equivalent or less restrictive open licence.

The preferred option allows for a targeted and proportional intervention, taking into account the state of play in terms of openness, digitisation and policy/legal framework across the Member States. The package chosen assigns the most suitable intensity level of regulatory intervention based on different impacts of the assessed policy options in the thematic areas. The policy option is also acceptable to both major groups of stakeholders and can be implemented within a reasonable timeframe, without the need to substantially defer the date of application of the measure.

It is worth indicating that considering all themes at once, and from a purely economic perspective, the benefits of the higher intensity regulatory intervention scenario are prevailing over the lower intensity approach. However, the adoption of this policy option for all thematic categories would not sufficiently take into account the complexities of the matter at hand. The high intensity scenario is burdened by a lower feasibility and higher adjustment costs of the public sector bodies. This includes potential friction (notably for the Company data theme) with the personal data protection rules¹⁸³. The costs involved may also negatively influence the effectiveness of the measure, as it would require a longer adjustment period (later date of applicability of all or of selected provisions of the Implementing Act). This could lead to negative repercussions in the actual short-term re-use, given that the success of the data economy depends on recent data and technologies. In addition, the full ‘higher intensity’ scenario would not be welcome by the data holders, while the ‘lower intensity’ scenario would meet considerable opposition from the re-users.

Economic value of the mixed intensity intervention:

Commercial data activities, including those based on the re-use of HVDs, are usually positioned at an early stage of value chain. The indirect (forward) effect on downstream industries can be significant. This is because value-added products and services based on PSI/HVD data are deployed across all industry sectors (e.g. services based on weather data are used in agriculture but also in the energy sector, those based on Earth observation data, in construction and transport sectors). Taking this into account, the support study calculated the indirect economic impact as a magnitude between ca. 2.5 and 3.0 (understood as multipliers of the direct economic impact¹⁸⁴). In order to estimate the total economic impact, the direct impact and indirect effects presented above can be aggregated and presented as a difference of the economic value against the baseline.

Accordingly, for the mixed policy package (preferred) the value of the PSI/HVD related 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 % in the baseline scenario.

The total value of the economy in 2028 with the mixed policy package is EUR 276 billion and represents an incremental impact of EUR 21 billion in comparison to the baseline.

The total economic effect is composed of a direct and an indirect forward effect. In 2028, the indirect forward effect of implementing the mixed policy package is EUR 201 billion. This signifies an increase of EUR 16 billion in comparison to the baseline (EUR 185 billion). The direct effect of the mixed policy package is EUR 76 billion, representing an incremental impact of EUR 6 billion as compared to the baseline in 2028. Other economic effects include impact on SME creation and employment (See figures below).

¹⁸³ The Directive is ‘without prejudice’ to EU and related national data protection provisions.

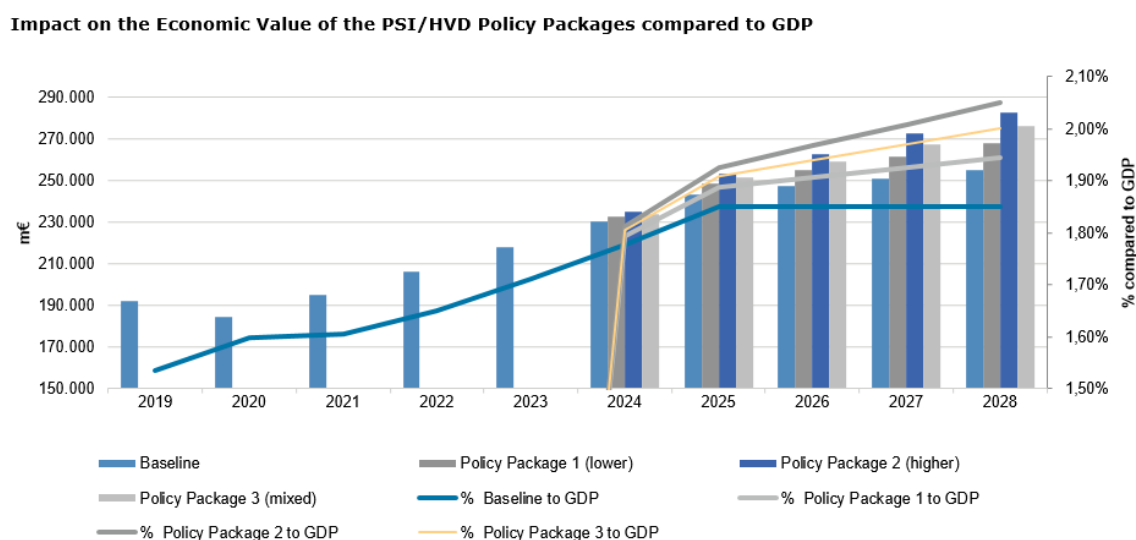
¹⁸⁴ European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

It should also be noted that creative use of high value data included in the preferred policy option will generate benefits well beyond the realm of the economy, as discussed in chapter 6.2. The effects of open data on the transparency of public life, good governance but also on the capacity to tackle great social and environmental challenges, are, at least, as impactful as those related to the economy. The use cases recorded by the Global Open Data Impact Map¹⁸⁵ confirm this observation.

While the total impact may seem modest in comparison to that of other EU initiatives in the field of the data economy, it should be kept in mind that the Implementing Act will only concern a small subset of public sector information held by public sector bodies in the EU. Not only is it restricted by the reference to the six data themes indicated by the co-legislators but it is also a fruit of a further selection based on the suitability of the specific datasets to be defined as having ‘high value’. Furthermore, the final selection does not include data held by public undertakings, research data of research institutes or data of cultural establishments which otherwise fall within the scope of open data provisions¹⁸⁶.

As far as costs are concerned, the exact values are hard to establish due to the difficulty in decoupling general data or IT infrastructure costs from the cost specifically linked to the reuse of HVDs. Nevertheless, the available generic estimates point to the following figures: one off costs (such as API set up) range from EUR 24.9 million, (low estimate) to EUR 435.9 million (high estimate), with a medium costs expected to be around EUR 122.3 million. The average recurrent costs (such as loss of revenue and resources needed to increase data quality) up until 2028 are of the order of EUR 473.6 million.

Fig. 3 summarises the economic impact (in terms of GDP and total economic value) of the chosen policy option against other options considered in the IA as well as against the baseline.



¹⁸⁵ <https://opendataimpactmap.org/regions>

¹⁸⁶ These limitations stem from the requirements of article 14 of the Open Data Directive.

Table 8 below shows the incremental rise of GDP over the baseline for the chosen intervention option.

Incremental rise of GDP over the baseline due to increased re-use of HVDs (expressed as resulting economic value in m€ and as % of GDP)

	2024	2025	2026	2027	2028
Baseline	230.417	243.632	247.389	251.262	255.239
% Baseline to GDP	1,78%	1,85%	1,85%	1,85%	1,85%
Policy Package 3 (mixed intervention) - Incremental rise of GDP (in m€)	3.615	7.721	11.969	16.375	20.959
% Policy Package 3 to GDP - Incremental rise of GDP (in % of GDP)	0,03%	0,06%	0,09%	0,12%	0,15%

Tables 9 and 10 below translate the impact of an increased GDP in the chosen intervention option on the creation of new enterprises as well as employment impact, defined as incremental rise in the number of persons employed by thematic area. Both types of impacts are measured against the baseline scenario, for the period 2024-2028¹⁸⁷.

Employment - impact	2024	2025	2026	2027	2028
Baseline					
1. Company and company ownership	151.214	159.887	162.352	164.894	167.504
2. Geospatial	829.239	876.801	890.319	904.259	918.573
3. Meteorological data	487.788	515.765	523.717	531.917	540.337
4. Earth observation and environment	360.963	381.666	387.551	393.619	399.849
5. Statistics	390.230	412.612	418.974	425.534	432.269
6. Mobility	219.504	232.094	235.673	239.363	243.152
Total	2.438.938	2.578.825	2.618.586	2.659.585	2.701.684
Policy Package 3 (mixed intervention) - number of jobs created					
1. Company and company ownership	710	1.505	2.267	3.018	3.766
2. Geospatial	7.786	16.543	25.358	34.310	43.461
3. Meteorological data	13.740	29.466	46.050	63.490	81.863
4. Earth observation and environment	8.473	18.129	28.236	38.800	49.866
5. Statistics	5.496	11.705	18.072	24.624	31.393
6. Mobility	2.061	4.379	6.712	9.082	11.504
Total	38.267	81.727	126.696	173.324	221.853
Enterprises number - impact	2024	2025	2026	2027	2028
Baseline					
1. Company and company ownership	30.079	31.804	32.294	32.800	33.319
2. Geospatial	164.949	174.409	177.098	179.871	182.718
3. Meteorological data	97.029	102.594	104.176	105.807	107.481
4. Earth observation and environment	71.801	75.919	77.090	78.297	79.536
5. Statistics	77.623	82.075	83.340	84.645	85.985
6. Mobility	43.663	46.167	46.879	47.613	48.367
Total	485.143	512.968	520.878	529.033	537.407
Policy Package 3 (mixed intervention) - number of enterprises created					
1. Company and company ownership	141	299	451	600	749
2. Geospatial	1.549	3.291	5.044	6.825	8.645
3. Meteorological data	2.733	5.861	9.160	12.629	16.284
4. Earth observation and environment	1.685	3.606	5.617	7.718	9.919
5. Statistics	1.093	2.328	3.595	4.898	6.245
6. Mobility	410	871	1.335	1.807	2.288
Total	7.612	16.257	25.202	34.477	44.130

¹⁸⁷ Given the adoption date and the expected timeline of the actual applicability of several provisions of the implementing act (e.g. on charging or APIs), impacts are measured as of 2024.

Employment - impact	2024	2025	2026	2027	2028
---------------------	------	------	------	------	------

Source: Impact Assessment Support Study, Deloitte.

9. HOW WILL ACTUAL IMPACTS BE MONITORED AND EVALUATED?

Monitoring of the impact of open data initiatives is one of the main factors to ensure the long-term viability of the open data policy. Recent studies have found that much of the potential value of open data remains untapped because data holders do not understand who is using open data or, more importantly, who is not using open data but could benefit from the insights it may generate¹⁸⁸.

The European Data Portal conducts an annual 'landscaping exercise' that monitors performance indicators in each EU Member State. The exercise, which is compiled in an Open Data Maturity Report¹⁸⁹, will be continued. The definition of the performance indicators has already been adjusted so as to take into account the changes introduced by the revised Open Data Directive and will be further adjusted to enable the monitoring of the impact of the Implementing Act. The European Data Portal will for example measure (via structured feedback from national authorities and the monitoring of the data flows through the portal on the supply side as well as via reported and collected use cases on the demand side) the progress in which HVDs are published through APIs integrated in open data portals and other repositories.

The use of API keys and analytics in the publishing of HVDs allows for more control than the standard use of an open data portal. It also helps understand who is using the public content and how. This inherent characteristic of API-enabled access will notably enhance the efficiency of usage monitoring and its subsequent reporting on national and EU levels.

The existing expert group (PSI Group¹⁹⁰) will assist the Commission in evaluating the state of transposition of the revised legislation (Open Data Directive) and the Implementing Act based on it and communicate the outcome of their own, national assessments and related studies.

Evaluation of the implementation of the Open Data Directive based on a modified review clause (Article 18), will be a key milestone allowing the Commission to assess the impact of the Implementing Act and decide on the possible revision of the Directive or the extension

¹⁸⁸ <http://www.thegovlab.org/open-data-demand.html>

¹⁸⁹ <https://www.europeandataportal.eu/en/impact-studies/open-data-maturity>

¹⁹⁰ <https://ec.europa.eu/digital-single-market/en/news/public-sector-information-group-main-page>

(via delegated or implementing acts) of the list of specific HVDs. Such evaluation will be undertaken no sooner than 17 July 2025.

Ad hoc studies as deemed appropriate (e.g. in line with the development of data processing technologies), meetings with stakeholders' associations (e.g. PSI Alliance¹⁹¹) and targeted surveys or consultations will be undertaken to e.g. to measure the reduction of the administrative burden among selected groups of stakeholders and to provide evidence for the actual take-up of the data for the development of new services and products by SMEs on the one hand and by big tech companies on the other..

Finally, the mid-term review of the Digital Europe Programme will allow the Commission to observe the changes brought about in the field of the data economy in Europe by the exploitation of the HVDs, alongside other public and private sector data, in the framework of the sectoral European data spaces.

Monitoring indicators for specific objectives:

Specific objectives	Operational objectives	Proposed indicators
Ensure the datasets are easy to process by machines.	Impose an obligation to publish HVDs via APIs (incl. bulk download where available) and indicate the minimum common quality requirements/principles for APIs.	The increase in machine-to-machine requests for the use of HVDs in the period of 2 years following the establishment of APIs for a given dataset: to be derived regularly from the EDP maturity report and via www.api-dashboard.com
Eliminate charges as market entry barriers to re-use.	Impose a prohibition to apply any charges in relation to the re-use of the HVDs.	The increase in the number of start-ups and SMEs as commercial re-users of HVDs, as well as the increase of non-commercial reuse of datasets (researchers, students, journalists). The number of services developed by large tech companies based on HVDs vs the increase in re-use by Start-ups and SMEs. The information will be gathered from national data

¹⁹¹ <http://psialliance.info/>

		portals (many of which publish the ‘success stories’ of re-use) and via reporting on voluntary basis by the private sector (with the help of EU and national re-users’ associations). It should be collected not earlier than 2 years after adoption (to ensure that all relevant datasets are free by then ¹⁹²)
Ensure the datasets are interoperable.	<p>Impose an obligation to use one of the two most open Creative Commons licences or an equivalent bespoke licence.</p> <p>Describe the HVDs in a way which allows for the publication of the same content across all EU MS (where the data exists), by referring to existing data specifications (e.g. based on INSPIRE).</p>	<p>The increase in number of data-driven services of a regional (e.g. Nordic countries, Mediterranean region) or pan-EU scope using HVDs.</p> <p>The degree of usage of HVDs in the common European data spaces.</p> <p>Information on licences can be derived e.g. from the national ‘chapters’ of Creative Commons and from the re-users associations that monitor the availability of data in practice. The remaining information will be gathered from the European Data Portal and via regular reporting by the MS.</p>

¹⁹² This is the maximum delay given by the OD Directive for the MS to discontinue all charges on HVDs.

ANNEX 1: PROCEDURAL INFORMATION

1. LEAD DG, DECIDE PLANNING/CWP REFERENCES

The proposal for a Commission Implementing Regulation laying down a list of specific High-Value Datasets (HVDs) belonging to the categories set out in Annex I of the Directive on open data and re-use of public sector information (Directive (EU) 2019/1024) was prepared under the lead of the Directorate-General Communication Networks, Content and Technology. In the DECIDE Planning of the European Commission, the process is referred to under item PLAN/2019/5761. The Commission Work Programme for 2021 includes the adoption of the Implementing Act [*reference to be added after Work Programme 2021 adoption*].

2. ORGANISATION AND TIMING

Work on the preparation of this Implementing Act started in July 2019 with the political validation of the initiative and the subsequent setting-up of the Committee on open data and re-use of public sector information (the Open Data Committee). An Inter-service Steering Group assisted DG Communication Networks, Content and Technology in the preparation of the Impact Assessment and included Commission services of 16 Directorate-Generals, together with the Commission's Legal Service and Secretariat General.

The Inter-service Steering Group discussed the consultation strategy and the Inception Impact Assessment on 23 March 2020, and the draft Impact Assessment on 10 November 2020.

The Open Data Committee met on 30 October 2019 and on 25 June 2020.

[Timing ISC and information on the Committee's vote to be inserted]

3. CONSULTATION OF THE RSB

The Impact Assessment report was reviewed by the Regulatory Scrutiny Board on 16 December 2020, which delivered a positive opinion with comments. Based on the Board's recommendations, the Impact Assessment has been revised in accordance with the following points:

<i>Comments of the RSB</i>	<i>How and where comments have been addressed</i>
(B) Summary of findings and (C) What to improve	
<p>(B.1) The report's intervention logic is not sufficiently coherent. The problem description and objectives are not sufficiently linked to the options.</p> <p>(C.1) The intervention logic should cover both the number and type of HVDs and the way the public sector shares them. The current problems and objectives neglect the number and type of HVDs, which does not allow a proper link between the problem description, objectives and options. The report could be clearer on what defines the high value of data.</p>	<p>The link between the re-usability of datasets and the socio-economic impact was clarified in the problem definition and the specific objectives, and linked to the definition of options. In Chapter 4, we recalled in the objectives the thematic scoping of the initiative (as defined in Annex 1 to the Directive) as well as the decisive criteria for selecting the High-Value Datasets (as prescribed in Article 14(2) of the Directive).</p> <p>We also introduced the number of datasets as part of the problem definition and objectives, while also explaining the limitations imposed by the Open Data Directive in terms of data scoping and selection.</p> <p>We also clarified upfront in the Impact Assessment (Chapter 1) the fact that the Open Data Directive treats High-Value Datasets as a special group of public sector data and subjects them to rules that significantly differ from those applicable to other types of public sector information. This includes a requirement for a free-of-charge re-use for all High-Value Datasets. This was contextualised with an explanation of how this relates to the more general charging regime.</p>
<p>(B.2) The report does not sufficiently justify the set of options.</p> <p>(C.2) The report should better explain why the options consist only of a higher intensity and lower intensity intervention, applying to both the scope and the publication modalities. It should justify why it does not consider any intermediate options, or options with gradual elements, such as a longer implementation period for some datasets.</p>	<p>We clarified the arguments that led to the choice of two policy options per theme instead of three or four, especially in the methodological part of Chapter 5.</p> <p>We also ensured that the definition of the policy options in Chapter 5.2 would get a clearer justification for the decision to keep the scope of the data and the related publication requirements bound together in the design of the intensity of the intervention.</p> <p>Still in Chapter 5, we highlighted the impact of the compulsory free availability of High-Value Datasets on the definition of the policy options (e.g. making it impossible to use charges as an element of a 'gradual approach').</p>
(C.3) The report should be clearer on the limitations and degree of uncertainty of the	We expanded Chapter 6 on the limitations and degree of uncertainty of the quantitative estimates

quantitative estimates of the costs and impacts.	accordingly. Under subtitle “Costs per theme”, some relevant limitations and subsequently the proposed categorisation of impact were already described. These explanations were expanded in order to comprehensively address the matter of limitations and uncertainty. Regarding the impacts, key explanatory information was added on the basis of the support study.
--	---

4. EVIDENCE, SOURCES AND QUALITY

Evidence-collection process

A study to support the impact assessment (SMART 2019/0025) was the main source of the necessary evidence. The evidence collection process, conducted under difficult circumstances of the COVID-19 crisis, lasted from December 2019 to July 2020. One of the key tasks of the contractor was to reach out to all relevant stakeholder groups in order to gather the evidence on the possible use cases as well as the costs and benefits associated with the re-use of the final High Value Dataset. The study targeted individual data holders in the public sector in each of the six thematic domains mentioned in the Annex I of the Directive. It also liaised with associations of data re-users as well as separate companies, NGOs and researchers of relevance to the thematic range of the data covered by the assignment. To this end, it conducted direct interviews, targeted online surveys, and convening webinars, focus groups, a workshop and a public hearing in September 2020.

Given that the adoption of HVDs was part of the chosen policy option presented in the Impact Assessment SWD(2018) 127 final, the evidence collected for the purpose of the ("back-to-back") exercise combining an evaluation with an Impact Assessment leading to the adoption of the Open Data Directive was also used. This included the results of a support study (SMART 2017/0061 – Study to support the review of Directive 2003/98/EC on the re-use of public sector information).¹⁹³

Further important sources of evidence include the input received both in writing and orally during the meetings of the Commission Expert Group (PSI Group) and the Open Data Committee, as well as by the results of the public online consultation on the EU Data Strategy¹⁹⁴.

Finally, the evidence base was complemented by reports and studies conducted in the course of 2019 and 2020 by the European Data Portal, notably the Open Data Maturity Report 2019, which details the state of openness of public data resources of each Member State of the EU¹⁹⁵ as well as the 2020 Report on the Economic Value of Open Data¹⁹⁶.

¹⁹³ <https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1>

¹⁹⁴ <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12271-European-Strategy-for-data/public-consultation>

¹⁹⁵ <https://www.europeandataportal.eu/en/dashboard/2019>

Stakeholders' consultation process

The consultation of stakeholders aimed to contribute to identifying public sector datasets that have strong potential to generate important socio-economic or environmental benefits, innovative services, a high number of users or use cases¹⁹⁷, the revenues they may help generate, and their potential for being combined with other datasets. With regard to HVDs held by public undertakings in the scope of the Open Data Directive, the consultation actions paid special attention to the role of public undertakings in a competitive economic environment.

For issues relating to open data, the European Commission usually consults all interested parties. In view of the implementing act, both HVD holders or providers (i.e. public sector bodies and public undertakings, in the different thematic domains) and HVD re-users (i.e. commercial and non-commercial re-users, as well as all public sector bodies acting as data re-users) were targeted as they will be directly impacted by or directly benefitting from the initiative.

The Commission consulted all stakeholders via different actions, including an open public consultation from February to May 2020, in the framework of the consultation on the European strategy for data. It also included the consultation of the Inception Impact Assessment in July-August 2020. These actions promoted in several webinars / workshops gathering the so-called open data community (HVD holders and re-users).

The Commission also held more targeted actions, starting with a series of sectoral workshops on the future Common EU data spaces in second half of 2019, gathering public and private stakeholders, and aiming to assess the data needs within major economic sectors where common European data spaces could be created, and where HVDs could have a role.

Finally, closer to the finalisation of this Impact Assessment, the European Commission organised two workshops aiming to present, discuss and validate the different policy options jointly with the contractor of the support study. They took place on 28 July and 4 September.

¹⁹⁶ CapGemini Invent, 2020, <https://www.europeandataportal.eu/en/impact-studies/open-data-impact>

¹⁹⁷ for which the datasets are relevant, in particular SMEs.

ANNEX 2: STAKEHOLDER CONSULTATION

INTRODUCTION

The consultation of stakeholders aimed to contribute to the identification of public sector datasets that have strong potential to generate important socio-economic or environmental benefits, innovative services, a high number of users or use cases for which the datasets are relevant, in particular SMEs, the revenues they may help generate, and their potential for being combined with other datasets. With regard to high-value datasets held by public undertakings in the scope of the Open Data Directive, the consultation actions considered the role of public undertakings in a competitive economic environment.

For issues relating to open data, the European Commission usually consults all interested parties. In view of the implementing act, the consultation targeted two specific categories of stakeholders that will be directly impacted by or directly benefitting from the decisions on which HVDs are covered by the implementing act:

- HVD holders or providers, meaning public sector bodies and public undertakings, in the different thematic domains.
- HVD re-users and re-users organisations, commercial and non-commercial re-users, as well as all public sector bodies acting as data re-users.

Consulting the national representatives of the public sector

The *Committee* on open data and the re-use of public sector information¹⁹⁸ was created in September 2019. Its role is to provide a formal opinion (vote) on the Commission's proposed implementing act, but it also contributes to the work in preparation of the Impact Assessment. The meetings¹⁹⁹ of the Open Data Committee took place on 30 October 2019 and on 25 June 2020.

The existing *expert group*, called “the PSI group (Public Sector Information group)”, played a key role in the process. The PSI group is convened at a more technical level, at regular intervals. Its scope goes beyond HVDs and covers open data in general. PSI Group meetings of relevance²⁰⁰ for the process of identifying HVDs took place on 29 October 2019 and 28 May 2020. Notably, the PSI Group members contributed to the process by sharing with the Commission their suggestions for an initial wide list of datasets of high value in their respective Member States.

¹⁹⁸ Committee on open data and the re-use of public sector information, Committee code: C51600.

¹⁹⁹ The meeting reports are available in the Commission's Comitology Register:

<https://ec.europa.eu/transparency/regcomitology/index.cfm>

²⁰⁰ The meeting minutes are available on a dedicated Europa.eu page: <https://ec.europa.eu/digital-single-market/en/news/public-sector-information-group-main-page>

Consulting stakeholders, including HVD providers and re-users

Public consultations

- *Open public consultation*: As foreseen by the Better Regulation rules, the Commission conducted an online consultation, starting with the adoption of the European Strategy on data on 19 February 2020 and running until 31 May 2020. It was accessible in all languages and targeted all types of stakeholders.

In total, 806 contributions were received²⁰¹, of which 219 were on behalf of companies, 119 from business associations, 201 from EU citizens, 98 on behalf of academic / research institutions, and 57 from public authorities. 7 respondents represented consumers, and 54 respondents were non-governmental organisations. Amongst the 219 companies/business organisations, 43.4% were SMEs. Overall, 92.2% of the replies came from the EU-27.

During the public consultation, 230 position papers were submitted, among which around 60 dealt with the topic of high-value datasets.

- *Inception Impact Assessment feedback*: As foreseen by the Better Regulation rules, the Inception Impact Assessment (IIA) was published on the Better Regulation portal and made available for feedback for 4 weeks (from 28 July 2020 to 25 August 2020). In total, the Commission received 50 contributions, including 48 through the Better Regulation Portal²⁰². This consultation action benefitted from an important participation from the business sector, with half of the contributors being businesses, and a third being associations representing businesses. Public authorities represented close to 10% of the participants. The Commission also received a few contributions from non-governmental organisations, EU citizens, or other types of stakeholders.

In their feedback to the IIA, stakeholders made general comments on the HVDs as a concept, but addressed also arrangements and their potential impacts (e.g. competition distortion, overlap with existing legislation, data protection). Some stakeholders made more suggestions on specific datasets that should or should not be part of the list. Some of them also shared existing and good practices, especially in the field of mobility.

- *Other events*: usually in the form of online workshops or webinars because of COVID-19 outbreak, these events were instrumental in promoting the various consultation actions and interacting with the so-called open data community on the topic of High Value Datasets. They helped generating grassroots pressure on the national level towards a successful and impactful conclusion of the discussion on the list of High Value Datasets. This includes a workshop on High Value Datasets organised by the Konrad Adenauer Foundation (KAS) in February

²⁰¹ [Outcome of the online consultation on the European strategy for data.](#)

²⁰² <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12111-Implementing-act-on-a-list-of-High-Value-Datasets>

2020²⁰³, and a webinar on the economic impact of open data in Europe, conducted by the European Data Portal in March 2020²⁰⁴. A workshop on the lessons learned from INSPIRE²⁰⁵ and a workshop specifically on HVDs from the European Data Portal²⁰⁶ both took place in the framework of the INSPIRE Conference 2020 in June 2020.

Targeted consultations

- *Sectoral workshops on the future Common EU data spaces*: after the adoption of the Open Data Directive and in preparation to the European Data Strategy Communication of February 2020, the Commission conducted a series of workshops. Gathering public and private stakeholders, they aimed to assess the data needs within major economic sectors where common European data spaces could be created, and where High Value Datasets could have a role²⁰⁷. Although not specifically aligned with the data themes in Annex 1 of the OD Directive, the workshops managed to discuss data needs and opportunities of setting up data spaces in the areas relevant to the HVDs exercise, such as mobility or environment. In addition, HVDs were presented and discussed during two workshops on law enforcement (10 July and 24 October 2019), and the INSPIRE Maintenance and Implementation Group (Committee) on 3 June 2020.

- *Workshop on different policy options* (28 July 2020): organised by study contractor who presented the preliminary results of the study, the workshop collected feedback from both data holders and re-users, regarding the policy options and impact modelling. The workshop gathered around 100 participants, both from private and public sectors.

- *Workshop/Webinar on final results of the support study* (4 September 2020): Organised by the Commission with the assistance of the study contractor who presented the final results of the support study, it aimed to present the datasets in scope, the policy options and their benefits and costs, for each category. Similarly to the July workshop, this event gathered around 250 participants, both from private and public sectors.

- *Bilateral contacts*, involving Commission services and the main private and public sector stakeholders have been taking place regularly after the adoption of the Open Data Directive and in the context of the preparation of the HVDs list and the rollout of the common European data spaces under the Digital Europe Programme.

²⁰³ <https://www.kas.de/documents/259586/0/Summary+-+Final+Remarks.pdf>


²⁰⁴ <https://www.europeandataportal.eu/en/news/webinar-economic-impact-open-data-europe>

²⁰⁵ <https://www.europeandataportal.eu/en/highlights/inspire-2020-european-data-portal-web-session-high-value-datasets>

²⁰⁶ <https://www.europeandataportal.eu/en/news/european-data-portal-and-high-value-datasets>

²⁰⁷ <https://ec.europa.eu/digital-single-market/en/news/report-european-commissions-workshops-common-european-data-spaces>

Summary of consultation actions

Purpose	Consultation actions	Stakeholders targeted	Language regime
Evidence gathering in view of the Impact Assessment  Testing results of the study and finalising the Impact Assessment	Sectoral workshops on the future Common EU data spaces (2 nd half 2019)	Both HVD providers and re-users	EN
	Open Data Committee and PSI Expert Group (<i>meetings until June 2020</i>)	National Ministries	EN
	Open public consultation (19 February – 31 May 2020)	All stakeholders	All EU languages
	Inception Impact Assessment feedback (<i>from 28 July to 25 August 2020</i>)	All stakeholders	IIA in EN, Feedback in all EU languages
	Workshop on policy options (28 July 2020)	Both HVD providers and re-users	EN
	Bilateral contacts (<i>all along 2020</i>)	Member States representatives; both HVD providers and re-users	EN or other languages
	Open Data Committee and PSI Expert Group (<i>meetings after June 2020</i>)	National Ministries	EN
	Workshop organised by study contractor (4 September 2020)	Both HVD providers and re-users	EN

RESULTS OF THE CONSULTATION PROCESS

High-value datasets as a concept

The consultation process supporting the design of the legislative framework on the governance of common European data spaces²⁰⁸ contributed to highlight the relevance of the initiative on high-value datasets. These workshops agreed that datasets and the conditions under which they will be reusable will contribute to the data pools under the common European data spaces in different areas. This role of HVDs was especially highlighted in the

²⁰⁸ PLAN/2020/7446

workshop on ‘Creating a common European data space for environmental and climate-related data’ that took place in Brussels on 23 September 2019.

In the online consultation on the European strategy for data (19 February – 31 May 2020), 761 stakeholders contributed to the section on HVDs. A proportion of 82.2% considered that the establishment of a list of HVDs a good way to ensure that public sector data has a positive impact on the EU’s economy and society. This figure increases to 90% when considering public authorities only. Of the remaining respondents, 10% were actually neutral or had no opinion, while 4,5% disagreed (representing various types of stakeholders, including 15 citizens). In open questions and papers, several respondents expressed their views about governance of datasets, considering that the selection of high Value Datasets should be based on use cases with potential societal benefits. Therefore, the issue of who can benefit from the opening of databases needs to be considered.

Stakeholders could share additional insights in papers attached to their online questionnaire reply, and through their feedback on the Inception Impact Assessment. Putting the initiative in a broader picture, many of them considered that opening up the data economy by the use of high-value would be an accelerator, and would foster data use for innovative businesses and for the public good. Overall, data protection appeared to be the most divisive issue. Input from the position papers shows that several Member States raise the challenge of conformity with personal data protection, and urge caution notably with regard to company data. At the same time, associations of re-users and NGOs encourage the Commission to ensure a wide scope of data to be considered as HVDs. Same diverging views have been expressed in stakeholder workshops (see below). The Commission took this issue into account and, also based on the evidence collected in the support study, chose the low-intensity option for the category of company and company ownership.

In this context, several data providers highlighted the increasingly relevant role of anonymisation and pseudonymisation as well as the importance to support the development of new techniques and solutions.

Some stakeholders also called for striking the right balance between making data reusable in the public interest and competition aspects. Representatives of public undertakings called for a cautious approach as regards HVDs that would fall in their remit and highlight a risk of competitive distortion. They also called for an entry into force of the Implementing Act only after the transposition of the Open Data Directive (i.e. mid-2021). More specifically in the mobility sector, several public sector organisations highlighted existing legislation in the field, and called for addressing only gaps and inconsistencies.

Although the Open Data Directive prohibits any charges on the re-use of HVDs, some public sector stakeholders questioned the free of charge principle, raising concerns on the potential impact in terms of lower data quality on the end-customer.

Factors, arrangements and their impacts

In the 2020 online consultation, stakeholders considered the *availability of datasets under uniform conditions across the entire EU* as a relevant or very relevant factor to be taken into account when defining high-value datasets (almost 84% of respondents to the question). Out of the remaining respondents, most were actually neutral and only 6% found it irrelevant (these 6 respondents were a varied sample of types of stakeholders).

Stakeholders also considered the *dataset's availability via an API* (78%) and the *availability free of charge* (75%) as relevant or very relevant factors. In the same way, the majority of remaining respondents were actually neutral or had no opinion. The least appealing factor was *the dataset belonging to a thematic area in which there are few EU-level requirements for opening up data*, which was indicated as very relevant or relevant by 52%.

Some stakeholders indicated additional relevant factors, such as data quality (from different angles: up-to-date, consistency, and an appropriate granularity), as well as data protection and the creation of taxonomies.

The survey also covered the relevance of arrangements that could improve the re-usability of high-value datasets. Stakeholders considered the provision of *standardised formats of data and metadata* as a relevant arrangement (84%), underlying the importance of open standards and of the FAIR principles²⁰⁹. *Licensing and other terms applicable to re-use* also revealed to be relevant (80%): according to several opinions it is key to avoid that global players monopolise the added value of these datasets, and standardised data license agreements can facilitate new collaborative approaches for sharing & reusing data resources. On this question from the online questionnaire, less than 3% found this irrelevant, while the rest was neutral to the question. In papers and feedback to the Inception Impact Assessment, some HVD potential re-users highlighted the importance of open data licences.

The least relevant arrangement was *specific technical arrangements for dissemination*, which 53% of respondents found relevant or very relevant. However, these respondents share the views of a need for user-friendly tools to download datasets with standardised (or common) data exchange protocols and well-documented APIs. The arrangements should also include multiple access points and ensure 24/7 access for users via thematic portals or dedicated websites at the European level (e.g. Linked Open Data according to W3C standards). Some respondents consider that the dissemination solution could even re-use some of INSPIRE principles and standards and ensure Machine-to-Machine access and include the availability of high performance data analytics platform.

²⁰⁹ Principles along which data should be findable, accessible, interoperable and reusable.

The various feedback to the IIA and position papers received show that, while the public data holders are generally in favour of taking into account to the extent possible the already existing publication arrangements, some countries have on the contrary asked the Commission to be more ambitious in the definition of dissemination arrangements and data quality.

Identifying the datasets

The 2020 online consultation could not yet gather feedback on specific datasets. The survey opened in February 2020, together with the adoption of the European strategy for data, while the Commission services were still designing the methodology to define the specific datasets, together with the contractors of the support study. Therefore, requesting input on specific datasets would have been premature.

However, the online consultation offered all stakeholders the possibility to propose specific High-Value Datasets along the various categories foreseen in the Open Data Directive. Overall, the respondents suggested:

- 20 datasets in the field earth observation and environment (e.g. hydrographic data);
- 17 datasets in the field of geospatial (e.g. EU land and coastlines);
- 12 datasets in the field of mobility (e.g. accidents data);
- 10 datasets in the field of companies and company ownership (e.g. life cycle assessment indicators);
- 9 datasets in the field of meteorological data (e.g. climate model data); and
- 5 datasets in the field of statistics (e.g. demographic data).

As an example, 33 respondents from the category of academic and research institutions gave dataset examples throughout the different categories. In the field of meteorological data, the mentioned very broad types of data (e.g. ‘weather’) or more precise ones (e.g. ‘rainfall extremes of short duration’, ‘high resolution maps about air quality’).

In various documents added to the online consultation, as well as in the feedback on the Inception Impact Assessment, some stakeholders made comments on specific datasets that, to their views, should or should not be part of the list of HVDs. For instance, as regards business registers, some stakeholders claim it should not be part of the list (due to possible privacy risks), while others consider it a critical data category for transparency, business integrity, fight against corruption, etc.

The views of stakeholders on the different policy options

The workshops on 28 July and 4 September 2020 were an opportunity for the stakeholders to express their views, based on the conclusions of the study, including on the policy options and their direct and indirect impacts. As both HVDs providers and re-users attended, this was an opportunity to confront diverging views. For instance, on the company and company

ownership category, several stakeholders (including Company Registers) raised the issue of the possible presence of personal data in such datasets, considering it as an obstacle to re-use. On the other hand, the community of re-users in this thematic domain is supportive of a wide opening of company and beneficial ownership data, in the public interest.

During these workshops, the support study contractor presented the methodology that led to the design of policy options and their estimated impacts. The multi-criteria analysis (MCA) revealed to be complex to understand for the stakeholders, but overall, they expressed support to the mix elaborated for the preferred option. During the 28 July workshop specifically (where half of the 94 participants were data holders, 20% data re-users and the rest ‘others’), the contractor of the support study held a survey. Overall, the stakeholders generally agreed with the description of the policy options and the modulation between the low and high intensity intervention, for each theme (over 70% in case of Company data to over 90% in the area of Geospatial data).

The role of EU Programmes

In the 2020 online consultation, the vast majority of respondents supported the idea that EU programmes may provide funding to enhance the availability and re-use of high-value datasets across Europe, especially those *to improve the quality (e.g. machine-readability) and interoperability of the data /metadata* (87%), and to a lesser extent funding aiming *to engage with re-users* (70%).

As additional activities, respondents highlighted the relevance of creating pilot project prototypes, pilots with stakeholders for testing, validation and self-assessment at the preliminary stage of the opening of databases. It looks important to stakeholders to set conditions and obligations for secondary data users and data aggregators and to engage with potential re-users and discovering their needs. This cooperation fostering amongst users and providers would allow the alignment of availability and demand.

ANNEX 3: WHO IS AFFECTED AND HOW?

10. SUMMARY OF COSTS AND BENEFITS

<i>I. Overview of Benefits (total for all provisions) – Preferred Option (policy package 3)</i>		
<i>Description</i>	<i>Amount</i>	<i>Comments</i>
Direct benefits (range per annum 2024-2028)		
Economic impact	982 - 5734 million EUR	Incremental direct economic impact (GVA/GDP) is estimated as differential to baseline. Annual values for the years 2024 and 2028 are presented.
Employment	10 - 61 thousand persons employed	Incremental employment impact for the preferred policy package. Total number of persons employed, estimated based on the employment coefficient/ratio to GVA for the EU27 ICT sector.
Governmental revenues	452 - 2638 million EUR	According to the definition of Eurostat, the governmental revenue is the sum market output, of taxes, net social contributions, sales, other current revenues and capital transfer revenues. Combining these categories of governmental revenue, a weighted coefficient of EU27 by GDP is obtained. Following the calculations of Eurostat, this coefficient has an approximately value of 46% of GDP for the EU27. Ratio is applied to the direct economic impact. Hence it should be noted, that the indicator presented reflects the statistical revenues for governments in general, induced by the stimulation of GDP due to the policy interventions.
Indirect benefits (range 2024-2028)		
Economic impact	2633 - 15225 million EUR	The analysis focuses on the indirect (forward) impact on downstream industries, which is considered to be the major indirect impact. To measure these impacts with regard to the PSI/HVD economic activity, results of the EU Data Monitoring have been analysed. As a result, for the indirect (forward) impact a magnitude between ca. 2.5 and 3.0 has been identified. To estimate the indirect (forward) economic impact, multipliers at the lower bound between 2.6 and 2.8 have been applied to the direct impacts.
Employment	28 - 161 thousand persons employed	
Governmental revenues	1211 - 7004 million EUR	

(1) Estimates are relative to the baseline for the preferred option as a whole (i.e. the impact of individual actions/obligations of the preferred option are aggregated together); (2) Please indicate which stakeholder group is the main recipient of the benefit in the comment section; (3) For reductions in regulatory costs, please describe details as to how the saving arises (e.g. reductions in compliance costs, administrative costs, regulatory charges, enforcement costs, etc.; see section 6 of the attached guidance).

II. Overview of costs – Preferred option (policy package 3)

		Citizens/Consumers		Businesses		Administrations					
		One-off	Recurrent	One-off	Recurrent	One-off (low estimate)	One-off (high estimate)	One-off (medium/average)	Recurrent (low estimate)	Recurrent (high estimate)	Recurrent (medium/average)
Defining list of HVDs free from barriers inhibiting their reuse	Direct costs	n/a	n/a	n/a	n/a	24.9m€	435.9m€	122.3m€	153.1m€	1 316.9m€	473.6m€
	Indirect costs	n/a	n/a	n/a	n/a	n/a			n/a		

(1) Estimates to be provided with respect to the baseline; (2) costs are provided for each identifiable action/obligation of the preferred option otherwise for all retained options when no preferred option is specified; (3) If relevant and available, please present information on costs according to the standard typology of costs (compliance costs, regulatory charges, hassle costs, administrative costs, enforcement costs, indirect costs; see section 6 of the attached guidance).

Methodological note for estimation of benefits and costs

1) Benefits

The benefits above are presented for the preferred policy package. The calculation is based on the results presented in the Final Study Report in chapter 4.4. All values are derived from the results in the Final Study Report.

The Policy package 3 (mixed intervention) consist of the following policy options:

Preferred PO	
Company & Company Ownership	PO1
Geospatial Data	PO1
Meteorological Data	PO2
Earth Observation & Environment	PO2
Statistics	PO2
Mobility	PO1

- Incremental direct economic impact (GVA/GDP) is estimated as differential to baseline. Annual values for the years 2024 and 2028 are presented.

- Incremental employment impact for the preferred policy package. Total number of persons employed, estimated based on the employment coefficient/ratio to GVA for the EU27 ICT sector.
- According to the definition of Eurostat, the governmental revenue is the sum market output, of taxes, net social contributions, sales, other current revenues and capital transfer revenues. Combining these categories of governmental revenue, a weighted coefficient of EU27 by GDP is obtained. Following the calculations of Eurostat, this coefficient has an approximately value of 46% of GDP for the EU27. Ratio is applied to the direct economic impact. Hence it should be noted, that the indicator presented reflects the statistical revenues for governments in general, induced by the stimulation of GDP due to the policy interventions.
- The analysis focuses on the indirect (forward) impact on downstream industries, which is considered to be the major indirect impact. To measure these impacts with regard to the PSI/HVD economic activity, results of the EU Data Monitoring have been analysed. As a result, for the indirect (forward) impact a magnitude between ca. 2.5 and 3.0 has been identified. To estimate the indirect (forward) economic impact, multipliers at the lower bound between 2.6 and 2.8 have been applied to the direct impacts.

2) Costs

The costs presented in the table above are estimated for the preferred policy package. The calculation is based on a bottom-up calculation for each thematic area (with regard to the Policy Option in the preferred Policy package 3). The calculation includes estimations for major cost categories, including:

- CAPEX and/or API costs as initial one-off costs. Costs for API development have been based on findings in the 2018 PSI study, including a low and a high estimate as well as an average. The costs are calculated based on the average value, multiplied with the expected number of APIs to be set-up in the Member State, e.g.: according to the information available, the most important infrastructural costs (one-off costs) in the thematic area Company & company ownership would be setting up an API. Costs estimation regarding the set-up of an API range between EUR 30 000 and EUR 2.5 million. The average was estimated at EUR 50 000. Based on the information, that approx. three quarter of the EU Member States would need to set-up an API under policy option 1 and policy option 2, the cost range is estimated between approx. EUR 600 000 (lower bound) and approx. EUR 50,6 million (upper bound) for the remaining (75%) EU 27 Member States in total. Based on the average costs for an API (EUR 50 000) total costs of EUR 1 million could be expected. The logic has been followed in all thematic areas, except Earth observation and environment, Geospatial and Mobility (transport network). For Earth observation and environment, one-off costs include estimation of CAPEX as provided during the interviews, which include API costs. The findings in the study are used to extrapolate to the EU27. In Geospatial and Mobility (transport network), it was assumed that most of the costs would be expected even in the baseline (BAU) scenario, since the development in this thematic area would follow a similar path as in the preferred policy option, mainly because of INSPIRE. For the thematic area statistics, only limited data regarding loss in revenues

will be available respectively only limited losses will be expected. Hence, no estimations are included.

- Regarding recurring costs, mainly operational cost and lost revenues for administrations have been assessed. The estimations are based on the finding in the study. For each category, the low and high estimates have been compiled from the final study. Furthermore, an average has been calculated. The estimation costs and lost revenues are based on an extrapolation of the average costs, multiplying the average value with the number of MS for which incremental costs/loss in revenues would be expected. The calculation can be described exemplary for the sector Company & company ownership, again: annual operational costs mainly related to data updates, replies to user requests, and corrections of errors in the datasets etc. can be quantified to 4 to 10 FTEs. Taking into account average hourly labour costs in the EU27, between EUR 20 for administrative and support service activities and approx. EUR 40 per hour in the ICT sector, total additional annual costs between EUR 3.2 million and EUR 16 million for the remaining (75%) EU 27 Member States in total could be expected for both policy options according to Eurostat. In case no data was available, or only insignificant costs would be expected, no estimations are made.
- For the thematic areas Geospatial and Mobility (transport networks), again only limited incremental costs would be expected, since the evolution in baseline is expected to be similar due to the INSPIRE obligations applicable in any case.
- The details of the estimations are presented in the tables below.

II. Overview of costs – Preferred option (policy package 3): 1. Company and company ownership											
		Citizens/Consumers		Businesses		Administrations					
		One-off	Recurrent	One-off	Recurrent	One-off (low estimate)	One-off (high estimate)	One-off (medium/ average)	Recurrent (low estimate)	Recurrent (high estimate)	Recurrent (medium/ average)
Defining list of HVDs free from barriers inhibiting	Direct costs	n/a	n/a	n/a	n/a	5,4m€	70,0m€	13,4m€	151,6m€	1.183,9m€	408,9m€
	Indirect costs	n/a	n/a	n/a	n/a	n/a			n/a		

II. Overview of costs – Preferred option (policy package 3): 2. Geospatial											
		Citizens/Consumers		Businesses		Administrations					
		One-off	Recurrent	One-off	Recurrent	One-off (low estimate)	One-off (high estimate)	One-off (medium/ average)	Recurrent (low estimate)	Recurrent (high estimate)	Recurrent (medium/ average)
Defining list of HVDs free from barriers inhibiting	Direct costs	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Indirect costs	n/a	n/a	n/a	n/a	n/a			n/a		

[illegible]

II. Overview of costs – Preferred option (policy package 3): 4. Earth observation and environment											
		Citizens/Consumers		Businesses		Administrations					
		One-off	Recurrent	One-off	Recurrent	One-off (low estimate)	One-off (high estimate)	One-off (medium/ average)	Recurrent (low estimate)	Recurrent (high estimate)	Recurrent (medium/ average)
Defining list of HVDs free from barriers inhibiting	Direct costs	n/a	n/a	n/a	n/a	9,2m€	230,0m€	83,2m€	0,0m€	110,2m€	55,1m€
	Indirect costs	n/a	n/a	n/a	n/a	n/a			n/a		

II. Overview of costs – Preferred option (policy package 3): 5. Statistics											
		Citizens/Consumers		Businesses		Administrations					
		One-off	Recurrent	One-off	Recurrent	One-off (low estimate)	One-off (high estimate)	One-off (medium/ average)	Recurrent (low estimate)	Recurrent (high estimate)	Recurrent (medium/ average)
Defining list of HVDs free from barriers inhibiting	Direct costs	n/a	n/a	n/a	n/a	5,8m€	60,5m€	19,8m€	1,8m€	3,5m€	2,6m€
	Indirect costs	n/a	n/a	n/a	n/a	n/a			n/a		

II. Overview of costs – Preferred option (policy package 3): 6. Mobility (inland waterway)											
		Citizens/Consumers		Businesses		Administrations					
		One-off	Recurrent	One-off	Recurrent	One-off (low estimate)	One-off (high estimate)	One-off (medium/ average)	Recurrent (low estimate)	Recurrent (high estimate)	Recurrent (medium/ average)
Defining list of HVDs free from barriers inhibiting	Direct costs	n/a	n/a	n/a	n/a	2,1m€	21,8m€	2,2m€	0,0m€	0,0m€	2,4m€
	Indirect costs	n/a	n/a	n/a	n/a	n/a			n/a		

Cost estimation administrations I 1. Company and company ownership

	One-off costs								
	CAPEX (initial invest)			API costs			data transformation costs		
	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)
costs/unit (k€)	API estimates as main costs			30	2.500	50	240	1.000	620
no. of MS: 20 (75%)	provided			20	20	20	20	20	20
total costs (k€)				600	50.000	1.000	4.800	20.000	12.400

Cost estimation administrations I 2. Geospatial

	One-off costs								
	CAPEX (initial invest)			API costs			data transformation costs		
	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)
costs/unit (k€)	250	3.000	1.625	30	2.500	50	100	200	150
no. of MS	-	-	-	-	-	-	-	-	-
total costs (k€)	-	-	-	-	-	-	-	-	-

Cost estimation administrations I 3. Meteorological data

	One-off costs								
	CAPEX (initial invest)			API costs			data transformation costs		
	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)
costs/unit (k€)	API estimates as main costs			30	2.500	50	100	200	150
no. of MS	provided			20	20	20	18	18	18
total costs (k€)				600	50.000	1.000	1.800	3.600	2.700

Cost estimation administrations I 4. Earth observation and environment

	One-off costs								
	CAPEX (initial invest)			API costs			data transformation costs		
	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)
costs/unit (k€)	400	10.000	3.617	estimates included in CAPEX			n/a or insignificant		
no. of MS	23	23	23						
total costs (k€)	9.200	230.000	83.183						

Cost estimation administrations I 5. Statistics

	One-off costs								
	CAPEX (initial invest)			API costs			data transformation costs		
	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)
costs/unit (k€)	500	3.000	1.750	30	2.500	50	n/a or insignificant		
no. of MS	11	11	11	11	11	11			
total costs (k€)	5.500	33.000	19.250	330	27.500	550			

Cost estimation administrations I 6. Mobility (inland waterway)

	One-off costs								
	CAPEX (initial invest)			API costs			data transformation costs		
	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)	Low (k€)	high (k€)	Ø (k€)
costs/unit (k€)	231	231	231	30	2.500	50	n/a or insignificant		
no. of MS	8	8	8	8	8	8			
total costs (k€)	1.846	1.846	1.846	240	20.000	400			

Cost estimation administrations | 1. Company and company ownership

	Recurrent costs									
	Lost revenues (for 10 MS data available, 27 MS expected)							OPEX		
	Data available	low (k€)	high (k€)	Ø (k€)	total - low (k€)	total - high (k€)	total - Ø (k€)	low (k€)	high (k€)	Ø (k€)
costs/MS or unit (k€)	survey total	30	60.000	14.788				160	800	480
no. of MS: 20 (75%)	10	17	17	17	27	27	27	20	20	20
total costs (k€)	147.880	510	1.020.000	251.396	148.390	1.167.880	399.276	3.200	16.000	9.600

Cost estimation administrations | 2. Geospatial

	Recurrent costs									
	Lost revenues			OPEX			data transformation costs			
		low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)
costs/unit (k€)					150	350	250	100	200	150
no. of MS	n/a				API coming anyway			API coming up anyway		
total costs (k€)										

Cost estimation administrations | 3. Meteorological data

current costs						
	Lost revenues (13 MS currently charge)			OPEX		
	low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)
costs/unit (k€)	20	2.300	776	150	300	225
no. of MS	13	13	13	18	18	18
total costs (k€)	260	29.900	10.085	2.700	5.400	4.050

Cost estimation administrations | 4. Earth observation and environment

current costs						
	Lost revenues (19 MS loss expected)			OPEX		
	low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)
	costs/unit (k€)	-	5.800	2.900		
no. of MS	19	19	19	n/a		
total costs (k€)	-	110.200	55.100			

Cost estimation administrations | 5. Statistics

	Recurrent costs															
	Lost revenues (10 countries data available)				OPEX			other costs			maintenance			data transformation costs		
		low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)	low (k€)	high (k€)	Ø (k€)
costs/unit (k€)					100	200	150	20	50	35	10	20	15	30	50	40
no. of MS	na/ or insignificant				11	11	11	11	11	11	11	11	11	11	11	11
total costs (k€)					1.100	2.200	1.650	220	550	385	110	220	165	330	550	440

Cost estimation administrations | 6. Mobility (inland waterway)

	Recurrent costs								
	Lost revenues			OPEX		other costs		staff costs	
		low (k€)	high (k€)	Ø (k€)	Ø (k€)		Ø (k€)		Ø (k€)
costs/unit (k€)					62		8		231
no. of MS	na/ or insignificant				8		8		8
total costs (k€)					492		62		1.846

ANNEX 4: ANALYTICAL METHODS

1. Methodological framework to identify the HVDs within each of the themes mentioned in Annex 1 to the Open Data Directive

The development of a suitable methodological framework was the first task of the support study. The paramount importance of this step can be explained by the wide thematic scope of the exercise (six data themes with their own stakeholders, sectoral legislation and national policies), the specific nature of data as an asset and the fact that no existing and ready methodological approach could be re-used for the purpose of this assignment.

The framework allowed the study team to identify potential HVDs that could be included in the scope of the analysis, and discuss those with stakeholders based on commonly understood value drivers and criteria. To this end, the study performed an initial analysis of the **relevant legislation** (and the data which are covered by EU level obligations, i.e. under the INSPIRE Directive or ITS Directive). Mapping all the relevant legislation at the European level was useful to identify which datasets and particular data values must already be made available by Member States, and therefore exist all across the European Union. Building on the identification of the relevant legislation and by performing a preliminary desk research, the team has established a **preliminary list of datasets to be covered** by the analysis. This very initial was not granular and meant as a starting point for discussion with key stakeholders during the **strategic interviews**. Where possible (and especially for the statistics and company and company ownership thematic areas) the team took into account the specific input shared by Member States when developing this list²¹⁰. Furthermore, the team assessed the possible number and categories of the public **data providers** for the datasets considered initially in scope. Finally, in this initial phase, the study team identified the key **characteristics of each thematic area** to be investigated in more detail.

As a second step, the study performed a literature review to assign categories of value related to the macro characteristics of potential value described in the Open Data Directive²¹¹. These include the economic benefits; environmental and climate-related benefits; generation of innovative services and innovation (innovation and artificial intelligence); reuse; and the improving, strengthening, and supporting of public authorities in carrying out their missions (public services and public administration,

²¹⁰ Documents provided by the members of the PSI Group of national experts.

²¹¹ See Official Journal of the European Union (2019) 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), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>

social).²¹² The review generated **32 categories of value, supported by 126 quantitative and qualitative indicators, within the six characteristics.**

The third step of the assignment was to assign (via both desk research and numerous interactions with all the relevant stakeholders, including the European Commission services, public sector data holders, NGOs and commercial re-users) the expected values related to the re-use of potential HVDs. This translated into the possible benefits associated with the use of such data. Furthermore, the study team was able to investigate how the said datasets are currently made available within all of the EU Member States as well as who and is using them and how (the state of play). Finally, after gathering information on the costs related to the production and dissemination of the data, the team was able to describe the expected impact of the changes related to the inclusion of the data in the list of HVDs.

2. Policy options, comparison of the policy options – Multi Criteria Analysis.

In line with the scope of intervention set out in the Open Data Directive, two main parameters have guided the development of the options: the number of data fields and/or datasets to be included as high value datasets, and the intensity of the measures for publication. By working incrementally with these two parameters, it was possible to build two options per thematic area, of varying intervention intensity: a ‘low hanging fruit’ and the re-users’ ‘wish-list’ policy options (alternatively: lower and a higher intensity intervention).

The description of the lower and higher policy option in turn was adjusted to the specificities of each theme. For themes in which a large part of the datasets are already widely available (such as statistics or environmental data), the policy options were mostly constructed around the intensity of the measures for publication. On the other hand, the scope and inclusiveness of datasets were used as the distinctive feature of the intensity of intervention in thematic areas where economic barriers to re-use persist or where specific data values are in high demand by the re-users (company data, meteorological information).

This methodological approach was chosen as a good compromise between the granularity and the cost-efficiency of examining six wide thematic themes within one single policy initiative. It has the benefit of presenting a clear variation in the intensity of policy approach per theme, which facilitates discussion with both the public data holders and the re-users. It is also helpful in assessing the expected benefits and costs, as the main

²¹² See Official Journal of the European Union (2019) 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), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>

factors for both are spread evenly across the two policy options. It is at the same time sufficiently flexible to allow for a targeted approach at the level of the overall legislative action (allowing for a mixed intervention option). Finally, it allows for micro-adjustments at the stage of the negotiations with the Member States in the Open Data Committee prior to the final vote.

The comparison of the policy options against the assessment criteria was based on a **Multi Criteria Analysis** (MCA). The MCA is a largely qualitative analysis of the policy options, based on ratings and rankings with quantitative data supporting the assessment. The MCA was performed in line with the European Commission's *Better Regulation Guidelines*²¹³ and its toolbox²¹⁴, most importantly tool 63²¹⁵. The assessment was built on the prior analysis of each individual option. The MCA was used for two reasons: First, it is an alternative to the Cost-Benefit Analyses (CBA) performed previously per each theme.²¹⁶ Secondly, it is particularly relevant at the stage of assessing the economic, social, and environmental dimensions of each of the two intervention options, and for comparing the policy options against the main criteria effectiveness, efficiency and coherence as well as the criteria proportionality and (legal and political) feasibility.²¹⁷

The scores/values assigned to each of the two policy options for each of the five criteria range from -3 (negative impact) to +3 (positive impact). The table below summarises the performance value for the intervention options.

Criteria	Performance range: -3 / + 3	Lower intervention (PO1)	Higher intervention (PO2)
Company and company ownership			
Effectiveness		1.50	2.50
Efficiency		2.50	1.50
Coherence		2.00	2.50
Proportionality		2.00	1.50
Feasibility		2.50	-1.00
Geospatial			
Effectiveness		1.75	2.75

²¹³ http://ec.europa.eu/smart-regulation/guidelines/toc_guide_en.htm

²¹⁴ http://ec.europa.eu/smart-regulation/guidelines/toc_tool_en.htm

²¹⁵ https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-63_en_0.pdf

²¹⁶ Due to a lack of data, costs and benefits have been quantified only to the extent possible. Costs and benefits have been assessed in a qualitative manner whenever quantitative data is lacking.

²¹⁷ http://ec.europa.eu/smart-regulation/guidelines/toc_guide_en.htm

Criteria	Performance range: -3 / + 3	Lower intervention (PO1)	Higher intervention (PO2)
Efficiency		3.00	1.75
Coherence		3.00	1.75
Proportionality		2.50	3.00
Feasibility		3.00	2.00
Meteorological Data			
Effectiveness		2.00	3.00
Efficiency		3.00	2.00
Coherence		2.00	3.00
Proportionality		3.00	3.00
Feasibility		1.00	1.00
Earth observation and environment			
Effectiveness		0.50	3.00
Efficiency		3.00	2.00
Coherence		2.00	3.00
Proportionality		2.50	2.00
Feasibility		2.50	2.50
Statistics			
Effectiveness		1.50	2.50
Efficiency		1.75	2.90
Coherence		2.50	2.50
Proportionality		1.00	2.00
Feasibility		2.00	1.75
Mobility			
Effectiveness		1.00	2.50
Efficiency		1.50	1.50
Coherence		3.00	2.50
Proportionality		3.00	2.50
Feasibility		3.00	1.50

Source: European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

Conducting the MCA with the performance values assessed based on the input matrix, the support study obtained the following outcomes:

Outcome of MCA: preferred policy options per theme

	Preferred PO (Equal Weight 0.20)
Company & Company Ownership	PO1
Geospatial Data	PO1
Meteorological Data	PO2
Earth Observation & Environment	PO2
Statistics	PO2
Mobility	PO1

Source: European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

Applying equal weights to all five criteria within the algorithm used, PO1 (lower intensity intervention) is identified as preferred option in the thematic areas of Company & Company Ownership, Geospatial data and Mobility. In the thematic areas Meteorological Data, Statistics, Earth Observation & Environment Policy Option 2 (higher intensity intervention) is the preferred option.

3. Impact modelling

Finally, the study performed impact modelling understood as comparing the expected macro-economic impacts of the lower and the higher intensity intervention and the preferred association of policy options on the overall economy and society compared to the baseline scenario.

The market size of PSI data is defined as the market size of products, services, and content improved or enabled by PSI Data.²¹⁸ However, which share of this value is attributed to open data can only be estimated. There exist several measures and methodological approaches to estimate the market value of PSI although many studies after 2011, (including the Study Supporting the Review of the PSI Directive by Deloitte, 2018), refer to or are based on approaches or results from the Vickery 2011 study²¹⁹ in order to determine the value of PSI. The baseline for the current study was therefore calculated according to the extrapolation of that study and the forecasts of the European Data Market Monitoring Tool²²⁰. The impact modelling assumptions were based on relevant literature and the findings of the support study. The table below provides an

²¹⁸ See EC (2020), The economic impact of open data, pp. 18ff.

²¹⁹ Vickery 2011, Review of Recent Studies on PSI Re-Use and Related Market Developments.

²²⁰ Data landscape, The European Data Market Monitoring Tool see: <http://datalandscape.eu/european-data-market-monitoring-tool-2018>

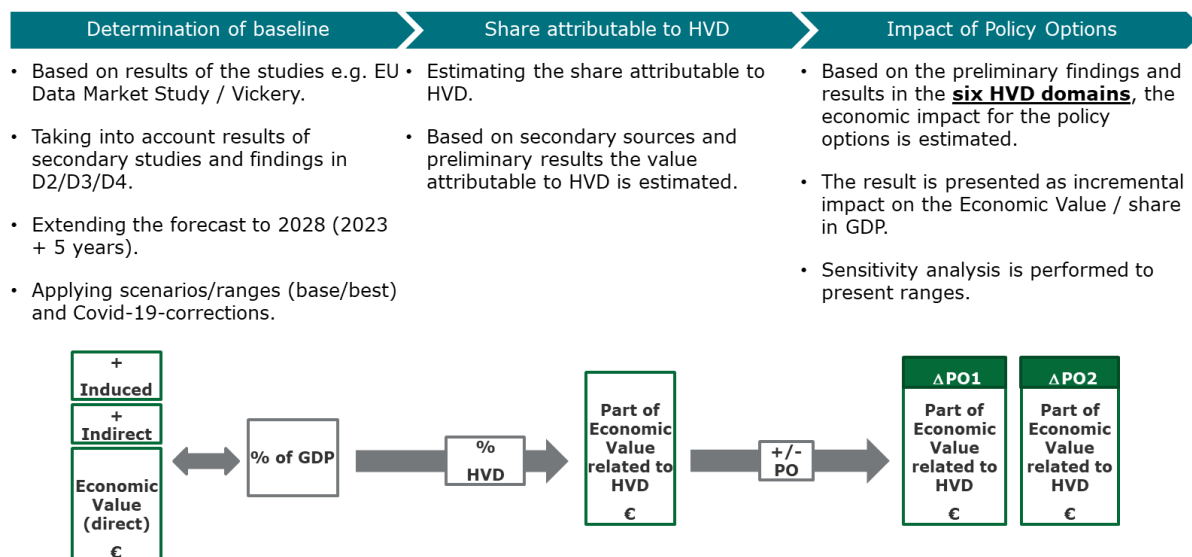
overview of the different studies that were considered as most relevant for the estimation of the baseline values.

Study	Context	Assumptions about data economy/ open data
Vickery 2011 , Review of Recent Studies on PSI Re-Use and Relate Market Developments	Part of the impact assessment of the first EU PSI Directive. Looks at the impact of the adoption of the PSI Directive in 2006, extrapolates sectoral + national impact research for the entire EU, based on GDP and ICT	Predicts a direct economic benefit of open data of 40 bio. EUR in the EU Predicts an indirect economic benefit of 140 bio. EUR in the EU. Predicts socio-economic benefits of 40 bio. EUR in the EU. Predicts EU market for government data in 2010 at about 32 bio. EUR.
McKinsey 2013 , Open Data	Large scale report about the economic potential of data and open data with macro-economic estimates.	Estimates the total economic potential in the EU at 900 bio. Dollar, applying a bottom-up approach.
EU 2015 , Creating Value through open data	Macro-economic research into economic potential in the EU for 2016-20, building on Vickery 2011.	Predicts a direct economic value of open data of 75.7 bio. EUR in 2020. Predicts public sector savings in the EU of 1.7 bio. In 2020. Predicts an indirect value of open data in the EU from 265-286 bio. EUR in 2020.
EU 2017 , Open Data Maturity in Europe	Yearly EU Data Portal Study	Predicted 325 bio. EUR of potential contribution of open data for 2016-2020 with 30.000 new jobs created in 2020.
Deloitte 2018 , Reuse of Public Sector Information	Evaluation assessed the performance of the PSI Directive, whether it still responds to the stakeholders' needs and expectations and whether it fits the purpose of the next years.	Study based on the Vickery Study 2011 and assumptions of Eurostat and the Data Monitoring Tool. Assessment identified problem areas of re-use of PSI Data, among them costs of data re-use, availability, exclusive agreements

Source: European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

The direct market size refers to the monetized benefits that are realized in market transactions in revenues and gross value added (GVA). The impact modelling of the study combined the direct and the indirect impact (understood as impact on downstream sectors) and referred to it as the total market size.

The following top-down approach was applied to obtain the economic impact of the PSI/HVD economy, in relation to its contribution to GDP:



Source: European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

The data from the European Data Monitoring Tool provides a baseline for the Economic value of the data economy and relates it to GDP. This data was used to calculate the baseline (including Covid-19 outbreak macroeconomic impact adjustments). In order to define the relevant market size, two conservative assumptions were made: The share attributable to high value data was based on the renowned Vickery study and this value was extrapolated to the year 2020 and then adjusted for corrections with the baseline values of the data economy from the European Data Monitoring Tool.

For each of the six thematic areas, further assumptions have been made to understand in more details the magnitude of impact of the specific policy options on this potential gap. The experts' assumptions are based on the findings of the interviews carried out and the literature studies. Based on existing literature and on the study research, the support study estimated the market share of the six thematic areas in % of the PSI market for the baseline scenario and the direct impact in EUR million of each thematic area for the years 2023-2028.

Baseline estimates – market shares

HVD | Baseline and Economic Impact

M€	2020	2021	2022	2023	2024	2025	2026	2027	2028
Market size PSI - Vickery	48 649	51 811	55 179	58 765	62 585	66 653	67 681	68 740	69 828
Assumption: baseline growth rate: 7%	(7.1%)	6.5%	6.5%	6.5%	6.5%	6.5%	1.5%	1.6%	1.6%
% PSI on data market value	90%	89%	87%	86%	85%	83%	83%	83%	83%
forecast based on EU Data Monitoring Tool						OECD GDP forecast			
Market share [% of PSI Market]									
1. Company and company ownership	6%	6%	6%	6%	6%	6%	6%	6%	6%
2. Geospatial	34%	34%	34%	34%	34%	34%	34%	34%	34%
3. Meteorological data	20%	20%	20%	20%	20%	20%	20%	20%	20%
4. Earth observation and environment	15%	15%	15%	15%	15%	15%	15%	15%	15%
5. Statistics	16%	16%	16%	16%	16%	16%	16%	16%	16%
6. Mobility	9%	9%	9%	9%	9%	9%	9%	9%	9%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Market share									
1. Company and company ownership	3 016	3 212	3 421	3 643	3 880	4 132	4 196	4 262	4 329
2. Geospatial	16 541	17 616	18 761	19 980	21 279	22 662	23 011	23 372	23 742
3. Meteorological data	9 730	10 362	11 036	11 753	12 517	13 331	13 536	13 748	13 966
4. Earth observation and environment	7 200	7 668	8 166	8 697	9 263	9 865	10 017	10 174	10 335
5. Statistics	7 784	8 290	8 829	9 402	10 014	10 664	10 829	10 998	11 173
6. Mobility	4 378	4 663	4 966	5 289	5 633	5 999	6 091	6 187	6 285
Total	48 649	51 811	55 179	58 765	62 585	66 653	67 681	68 740	69 828

Source: European Commission (2020, forthcoming). Support Study to this Impact Assessment, SMART 2019/0025, prepared by Deloitte.

The growth rates of the baseline scenario for the PSI market are conservative: A growth rate of 6.5% for the years from 2021-2025 is applied. The literature indicates an average growth rate of 7% of PSI Data and for the data economy the EU Data Monitoring Tool has estimated a CAGR of app. 9% in general from 2002-2025.²²¹

The macroeconomic impacts of the policy packages are based on the respective growth rates for the different thematic areas. Policy intervention options cover a different data scope, different policy areas and industries and are therefore expected to have distinct growth rates. The study estimated the growth rates for the lower intervention option and the high intervention option for each thematic area separately and then summed up the impact of each of the six thematic areas to obtain the overall impact of the policy packages. This allowed differentiating the growth between different thematic areas and account for already mature areas such as statistics, where lower growth rates are expected.

The baseline growth rates for each thematic area correspond to the baseline growth rate of the PSI market value of 6.5% annually. For the lower and the higher intervention option, the growth rates are, depending on the intensity and scope of the intervention options expected to be higher than the baseline growth. Therefore, three changes in growth rates were categorised: a slight change in the growth rate, a medium and a major change in the growth rate. The changes in the growth rates are a result of the scope and nature of the respective intervention option. For a slight change in the growth rate a differential of + 0.5-1.5% was added to the baseline growth rate. A major change in growth rate was asserted with a differential of + 2.0-3.0%-points as compared to the baseline growth rate. In line with OECD estimates for the growth rate in years 2026-

²²¹ See e.g. European Commission, European Data Portal (2020), The Economic Impact of Open Data: Opportunities for value creation in Europe. Study conducted by Capgemini.

2028, the study accordingly assumed that the growth rates for each thematic area would be lower for the years 2026-2028.

In order to fully reflect on the overall impact on the economy, indirect impacts on downstream industries (forward effect) of the policy options per each theme were also calculated. They are expressed as multipliers (of magnitude between ca. 2.5 and 3.0) applicable to the previously identified direct impact.

HVD Baseline and Economic Impact									
M€	2020	2021	2022	2023	2024	2025	2026	2027	2028
EU Data Monitoring Tool Multipliers (% of direct impact)									
<u>Baseline</u>									
Direct Impact	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Indirect Backward Impact	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05
Indirect Forward Impact	2.79	2.76	2.74	2.71	2.68	2.66	2.66	2.66	2.66
Induced Impact	1.83	1.97	2.12	2.29	2.47	2.66	2.66	2.66	2.66
Total Impact	5.68	5.79	5.91	6.05	6.20	6.37	6.37	6.37	6.37
<u>High Growth</u>									
Direct Impact	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Indirect Backward Impact	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Indirect Forward Impact	2.79	2.84	2.89	2.94	2.99	3.04	3.04	3.04	3.04
Induced Impact	1.83	2.09	2.39	2.74	3.13	3.58	3.58	3.58	3.58
Total Impact	5.68	5.99	6.34	6.73	7.18	7.68	7.68	7.68	7.68

In addition, the study provided an employment indicator on the total number of persons additionally employed, directly and indirectly including part-time and self-employed per theme. To calculate that figure, the coefficient of employment as per EUR million gross value added was determined. The employment coefficient was calculated as a weighted average number of persons employed per millions of Euro Gross Value Added in the ICT sector in the EU27 2018²²². The employment coefficient was then presented as a per-ratio increase in the number of persons employed per theme, which result from an increase in GDP.

Finally, the study attempted to calculate the **benefits for the public sector itself**, stemming mainly from increased general revenues and taxation of the increased overall economic activities.

The following methodology for this estimate was employed. Firstly, combining different categories of governmental revenue as defined by Eurostat²²³, a weighted coefficient of EU27 by GDP was obtained (approximately value of 46% of GDP for the EU27). The indicator reflects the revenues for governments in general, induced by the stimulation of GDP due to the policy interventions.²²⁴ Secondly, the incremental impact with regard to

²²² With regard to the indirect effects included however, the employment coefficient of the ICT sector can only serve as a proxy.

²²³ Eurostat 2020, Statistics Explained, Glossary: government revenue and expenditure. https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Government_revenue_and_expenditure

²²⁴ It must be noted, however, that this total governmental revenue includes – as defined in the European System of Accounts 2010 – also the market output, output for own final use and payments for non-market production. As this definition is a rather broad concept and as the macroeconomic effect of the introduction of the Policy Packages depends on a lot yet unknown factors, market output, output for own final use and payments for non-market production cannot be predicted as precisely as the other variables of governmental revenues. Excluding the categories mentioned, the adjusted governmental revenues would

governmental revenues was calculated, applying the average total governmental revenues as % of GDP. This allowed the study to present the individual contributions of the thematic areas by Policy Options on the development of the additional (incremental) governmental revenues in the 27 EU Member States.

lower to approximately 38% of GDP according to OECD estimates. OECD, 2020, Comparative Statistics: Governmental Revenue. <https://stats.oecd.org/Index.aspx?DataSetCode=REV>

ANNEX 5: HIGH VALUE DATASETS AS SUPPORT TO START-UPS AND SMES

This Annex addresses how and why HVDs will have an impact on start-ups and SMEs, in terms of 1) the benefits from access to HVDs, 2) benefits compared to large enterprises, and 3) the challenges that SMEs face today regarding the use of open data.

SMEs are decisive drivers of the data economy. Businesses with 100 or fewer employees report business intelligence and big data re-use rates that are as much as triple that of their larger corporate counterparts^{225,226}. SMEs represent 99.8% of companies active on the European data-market (including public sector data), given the relatively modest initial capital investment necessary to launch a data-driven business²²⁷. There are many reasons why big data is proving to be an excellent asset for SMEs, and the most important is reported as being the natural advantage that SMEs have over larger competitors: agility. Big data and analytics are useful for detecting trends and gleaning insights in order to adapt to a changing business environment. However, it is the ability to react quickly that makes big data such a natural fit for SMEs²²⁸. They create a significant economic impact, including creation of attractive job opportunities.

Today, data sharing in the economy is associated with high transaction costs. This relates to costs companies incur in finding a suitable data-sharing partner; negotiating, drafting and monitoring the contract and developing interoperability solutions for transferring, transforming and clearing the data²²⁹. This has been highlighted by stakeholders (especially SMEs)²³⁰. The OECD has confirmed that these high transaction costs might heavily affect those in a weaker position, notably consumers and SMEs.

HVDs will constitute a common EU-wide data layer specifically designed to minimise transaction costs and market entry barriers so often encountered by SMEs. Wider availability of such valuable and voluminous datasets should facilitate the uptake of big data analytics by SMEs, improving their performance, productivity, efficiency, new market entries, financial stability and control and innovation, thus achieving a

²²⁵ <https://www.smartdatacollective.com/big-data-and-the-sme-prepare-to-succeed/>

²²⁶ <https://www.forbes.com/sites/louiscolumbus/2017/10/08/small-businesses-are-the-real-mvps-of-analytics-and-bi-growth/#4af291526ca4>

²²⁷ Deloitte (2018), Study supporting the review of the PSI Directive.

²²⁸ <https://www.smartdatacollective.com/big-data-and-the-sme-prepare-to-succeed/>

²²⁹ Deloitte (2018). Realising the economic potential of machine-generated, non-personal data in the EU, Report for Vodafone Group.

OECD (2019). Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies, OECD Publishing, Paris.

²³⁰ European Commission (2017). Synopsis report consultation on the ‘building a European data economy’ initiative.

measurable effect in cost, profits, revenues, growth and agility²³¹. In addition, the analysis of large datasets adds significant value when used to carry out strategic and business decision-making processes²³².

As one of the main barriers to a wide re-use of public data is the financial barrier (high cost of data acquisition), making HVDs available free of charge is expected to play a crucial and enabling role for SMEs. It should foster agility and development, create cross-domain synergies and enable a rapid development of new products and services. It can also create a counterbalance for SMEs to large corporate entities that are much less dependent on HVDs. Better availability of HVDs also contributes to the objectives of the SME strategy²³³ and enhances its actions in the areas of capacity-building and support for the transition to sustainability and digitalisation, reducing regulatory burden and improving market access. Finally, free availability of HVDs is expected to result in an increase of the number of SMEs employing digital technologies.

On the other hand, the findings of a recent study²³⁴ indicate that the adaptation to a data-driven business by SMEs can be complex and takes time. The challenges that cause SMEs to lag behind in the adoption of big data analytics include data complexity, computational complexity and system complexity²³⁵. The wide availability of high value data of high quality and degree of interoperability can greatly help SMEs in overcoming many of these challenges. SMEs active in the data market may also encounter ethical concerns related to data mining²³⁶ and those linked to the use of de-centralised computing (privacy, performance, reliability, and data security²³⁷). Such concerns can be better addressed, if interoperability is well defined and established and when the data supplier guarantees the veracity and respect of the datasets with the applicable law, including personal data protection. This would be the case of HVDs, e.g. from public registers.

The lowering of barriers to reuse and higher availability of large datasets encourage the development of data analytics technologies and have a positive impact on competition. In France for example, larger re-user companies saw an increase in competition in their

²³¹ Soroka, A. et al. (2017) 'Big Data Driven Customer Insights for SMEs in Redistributed Manufacturing', *Procedia CIRP*, 63(1), pp. 692–697.

²³² Provost, F. and Fawcett, T. (2013) 'Data Science and its Relationship to Big Data and Data-Driven Decision Making', *Big Data*, 1(1), pp. 51–59.

²³³ COM(2020) 103 final. An SME Strategy for a sustainable and digital Europe.

²³⁴ Britzelmaier-B, Sterk-M, Graue-C: Big data in SMEs – findings of an empirical study. (2020) *Global Business and Economics Review* 22(1/2):115.

²³⁵ Jin, X. et al. (2015) 'Significance and Challenges of Big Data Research', *Big Data Research*, 2(2), pp. 59–64.

²³⁶ Boyd, D. and Crawford, K. (2012) 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon', *Information Communication and Society*, 15(5), pp. 662–679.

²³⁷ Depeige, A. and Doyencourt, D. (2015) 'Actionable Knowledge As A Service (AKAAS): Leveraging big data analytics in cloud computing environments', *Journal of Big Data*, 2(1), pp. 1–16.

markets, due to smaller re-users entering into business thanks to the INSEE and INPI databases being made available for free and through APIs²³⁸.

The UK's Company House study²³⁹ states that the effects of charging for (Company House) information and data are likely to be disproportionate across users. When charges apply, less regular users drop out of the market and these users tend to be SMEs. Building on this evidence, it would seem logical to conclude that including the datasets analysed by the study in the list of HVDs would open up the market to new players and allow smaller companies, in particular, to grow, thus leading to direct economic benefits.

The current stakeholder survey of the German National Weather Service (DWD) indicates that a rising number of start-ups and SMEs are involved in the re-use of meteorological data. Initial results of stakeholder research by the DWD indicate that there is a growing and dynamic market of re-users, with an increasing number of start-ups emerging in parallel to the established providers of meteorological services²⁴⁰.

USE-CASES:

Some selected use-cases for different thematic categories of HVDs, as referred to in Article 13(1) of the Directive, are presented below.

Geospatial and Statistics:

According to a study by the Aalto School of Engineering²⁴¹ on the effects of the opening of topographic data by the National Land Survey in Finland, SMEs are the user group with largest growth after the data were made open: 29% of SMEs and 26% of large companies use the data to refine products and services, the rest being mostly for internal use²⁴².

An example of an SME active in the geospatial domain is QMAP srl which develops products based on geolocalisation by following “IoT” logic and makes them available with a licence, on premise or as a service. QMap developed a platform for the analysis and management of big data. The platform allows for the merging of raw geospatial data with multiple other data to give them context, generating information with high added value. The ability to represent structured information geographically makes the data accessible and easy to understand. Geospatial data are for example correlated with population open data and other open data provided by ISTAT, the Italian National

²³⁸ Stakeholders interviews.

²³⁹ Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015.

²⁴⁰ Impact Assessment study on the list of high-value datasets to be made available by the Member States under the PSI Directive, 2020.

²⁴¹ Jaana Mäkelä, Paula Ahonen-Rainio and Kirsi Virrantaus “Effects of open topographic data in Finland, A user study one and half years after the opening” Dept of Real Estate, Planning and Geoinformatics, Aalto School of Engineering, 2014.

²⁴² Impact Assessment study on the list of high-value datasets to be made available by the Member States under the PSI Directive, 2020.

Institute of Statistics. The possibility to easily correlate anonymized data from different categories among themselves and with other data, regarding population, social, cultural or economic levels allows a powerful analysis and information service for a range of use-cases. The company also facilitates the availability of anonymized datasets about the geographical distribution of several categories of sensitive data. Research institutes, the scientific community and service providers can then provide analyses, information services and resource optimization. This allows them, for example to better respond to emergencies²⁴³.

Meteorological:

The Viennese start-up SharedMobility.ai has come up with ingenious ways of re-using historical weather data. SharedMobility.ai provides capacity predictions for the Vienna City bike-sharing service. Citybike Vienna is a station-based shared mobility service with fixed mobility points. Enhancing the rental experience with predictable lendings and returns avoids customer frustration and improves the rental frequency in the long term. Since the overall figure of the bike rentals are highly influenced by the local weather conditions, SharedMobility.ai improved the quality of its service by re-using weather information obtained from ZAMG, the national meteorological service of Austria²⁴⁴.

Earth Observation and Environment:

Slovenian company Synergise provides access to quadrillions of bits of information coming from Copernicus, USGS and other satellites. Such information is among the most valuable and voluminous datasets in Europe but is also very complex. The technological solution offered by Synergise makes such data easily accessible to end-users and application developers around the world. By doing this, Synergise also enables other start-ups to focus on creating added-value products and services instead of data cleaning and management.

Chloe Irrigation Systems is a start-up that launched a mobile app which re-uses both agricultural data, such as field's soil and crop types, and weather conditions to optimise irrigations and reduce water and costs. It is an AI-powered platform that offers personalised and optimised irrigation scheduling to create a better future with less water waste and increased crop yield.

Companies and company ownership:

Spanish company Dato Capital created a web application that combines business information to offer a real vision of companies. It matches information from different countries which is of great interest for its clients. The tool explores and analyses information from the Spanish Commercial Register to collect business information and to

²⁴³ <http://www.qmap.it/index.php/en/products/>

²⁴⁴ <https://sharedmobility.ai/2020/06/why-long-term-meteorological-data-is-a-key-factor-for-urban-mobility-services/>

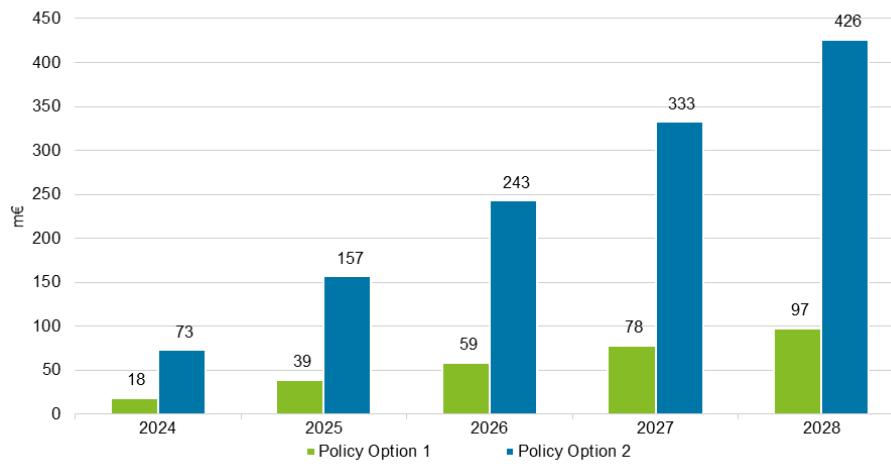
find links between data. It obtains information from APIs, downloading dataset and searching in web pages from the registers. The future is to keep the information updated in an international database with information about companies and administrations, and to maintain this tool for accessing it²⁴⁵.

²⁴⁵ https://www.europeandataportal.eu/sites/default/files/use-cases/use_case_spain_-_dato_capital.pdf

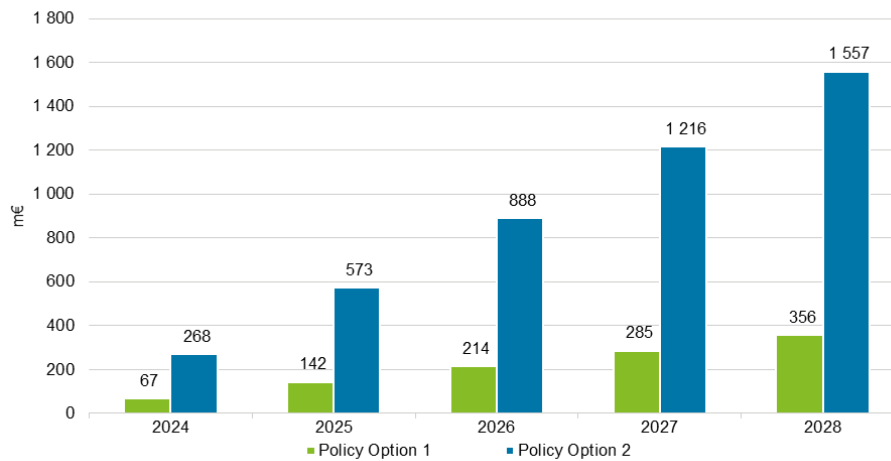
ANNEX 6: IMPACTS PER CATEGORY

Source: Deloitte estimation

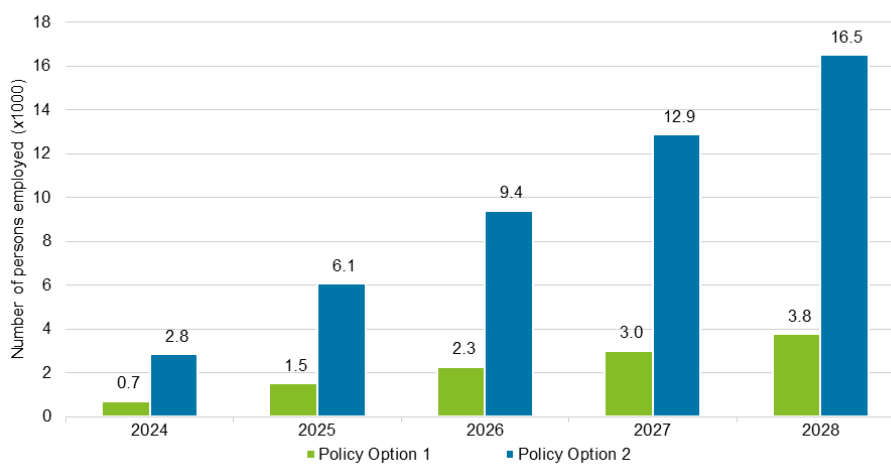
Economic Impact direct - 1. Company and company ownership



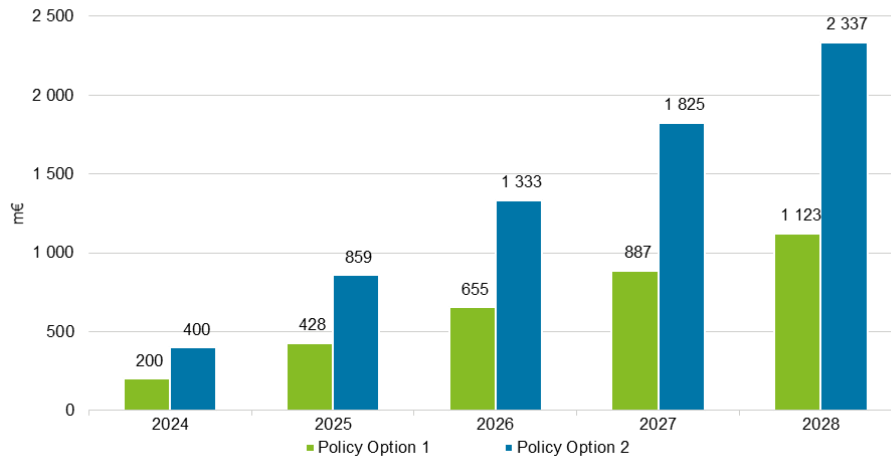
Economic Impact indirect - 1. Company and company ownership



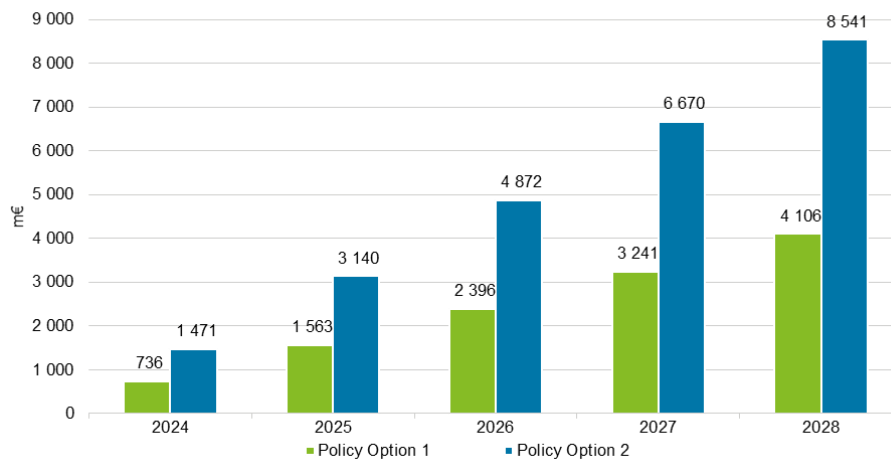
Employment Impact - 1. Company and company ownership



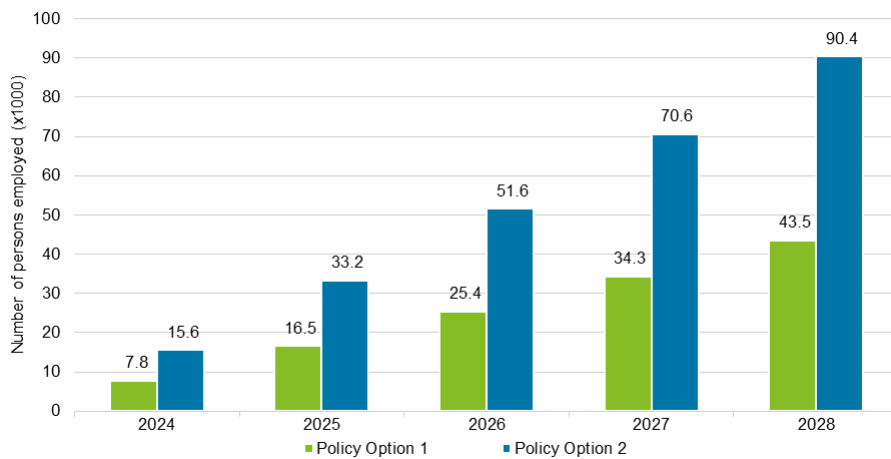
Economic Impact direct - 2. Geospatial



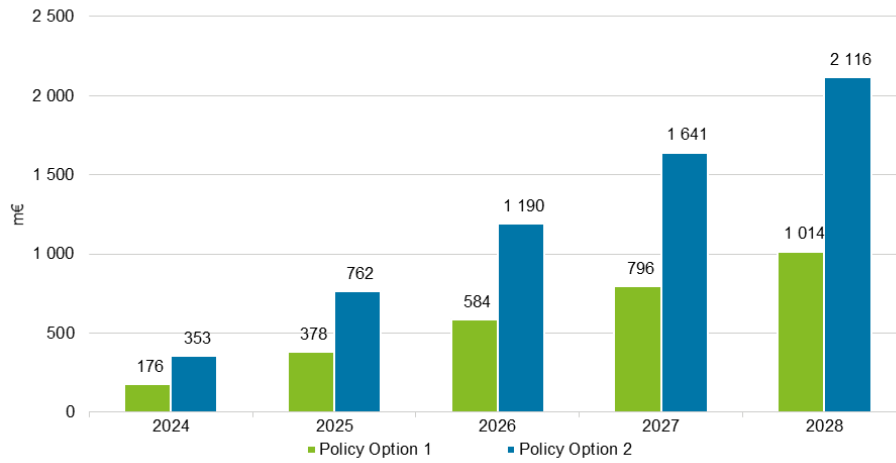
Economic Impact indirect - 2. Geospatial



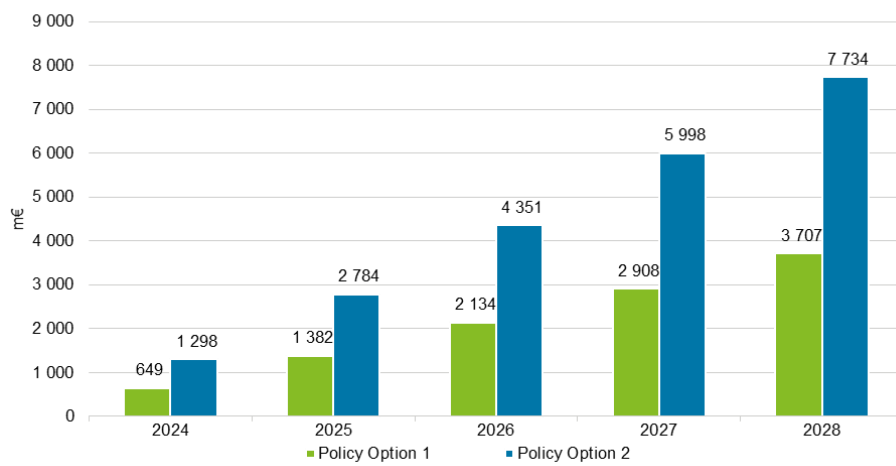
Employment Impact - 2. Geospatial



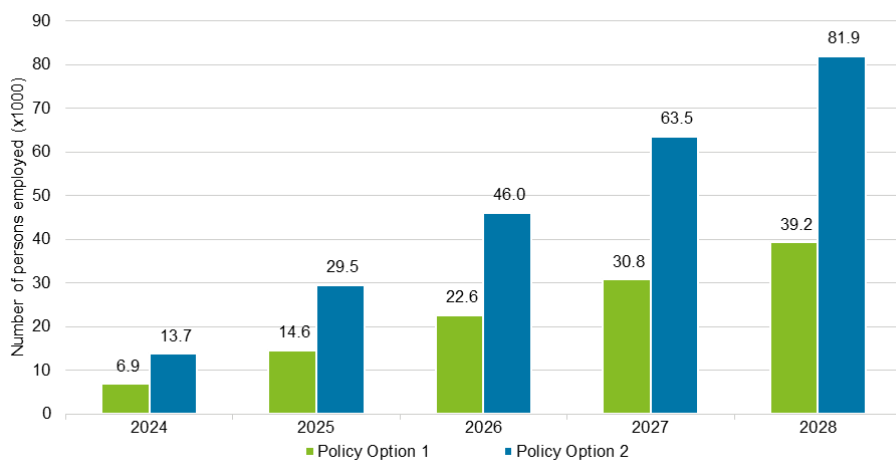
Economic Impact direct - 3. Meteorological data



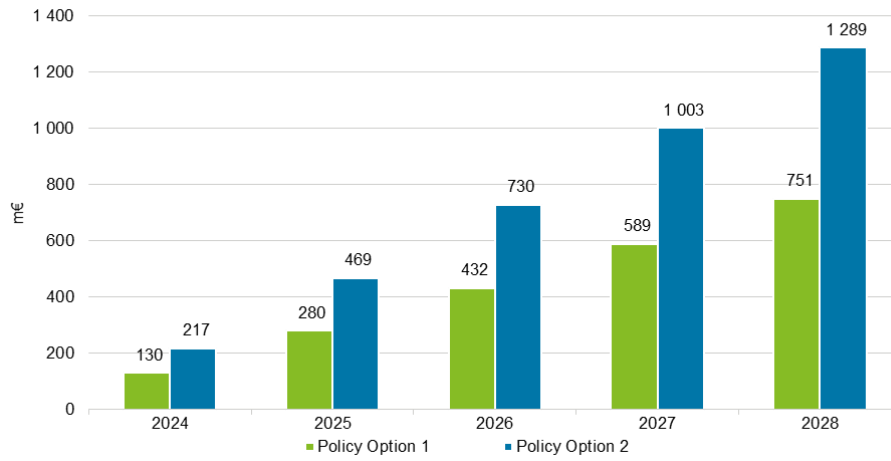
Economic Impact indirect - 3. Meteorological data



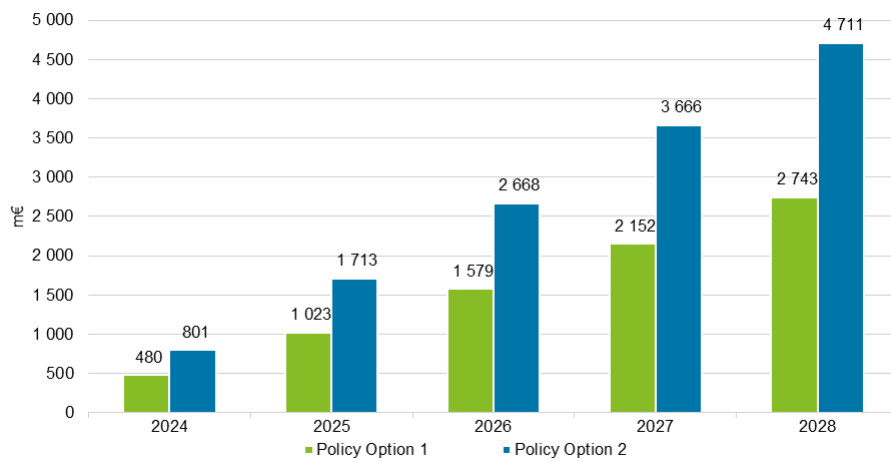
Employment Impact - 3. Meteorological data



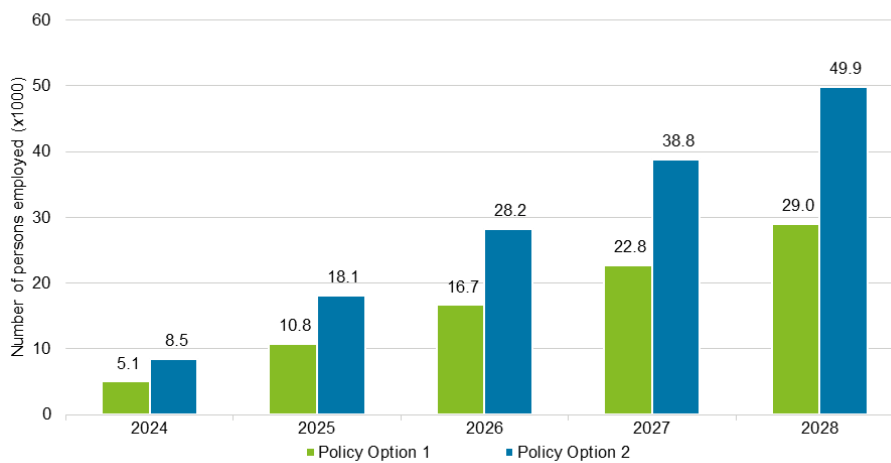
Economic Impact direct - 4. Earth observation and environment



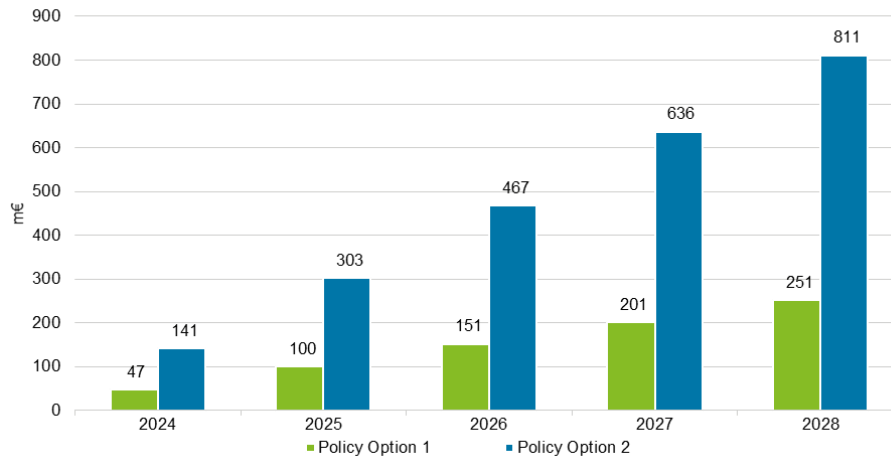
Economic Impact indirect - 4. Earth observation and environment



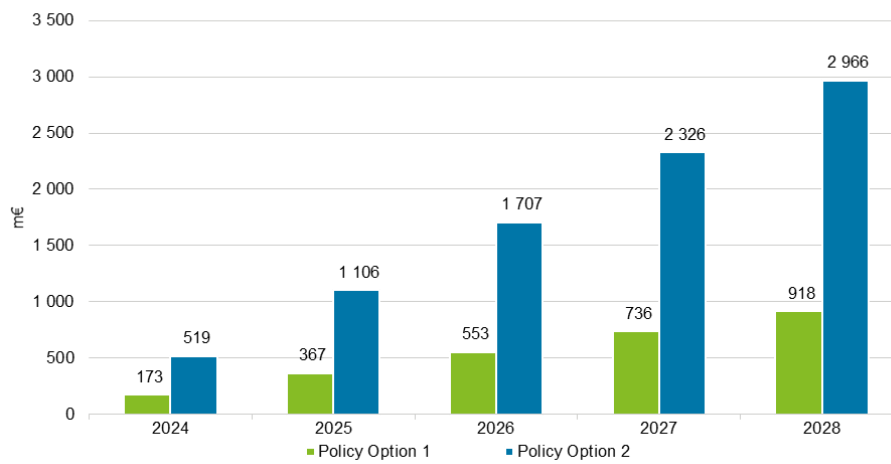
Employment Impact - 4. Earth observation and environment



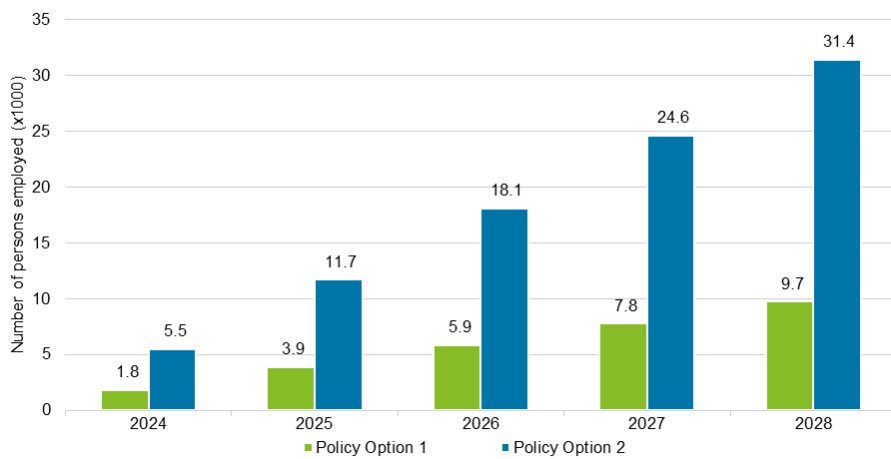
Economic Impact direct - 5. Statistics



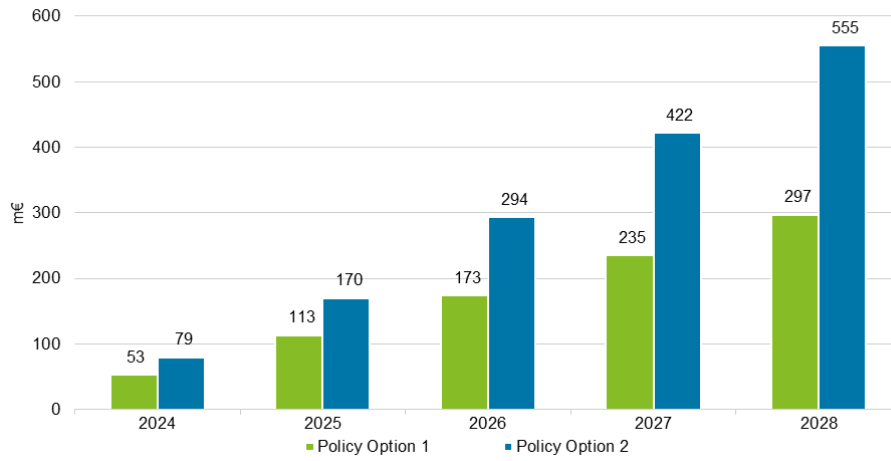
Economic Impact indirect - 5. Statistics



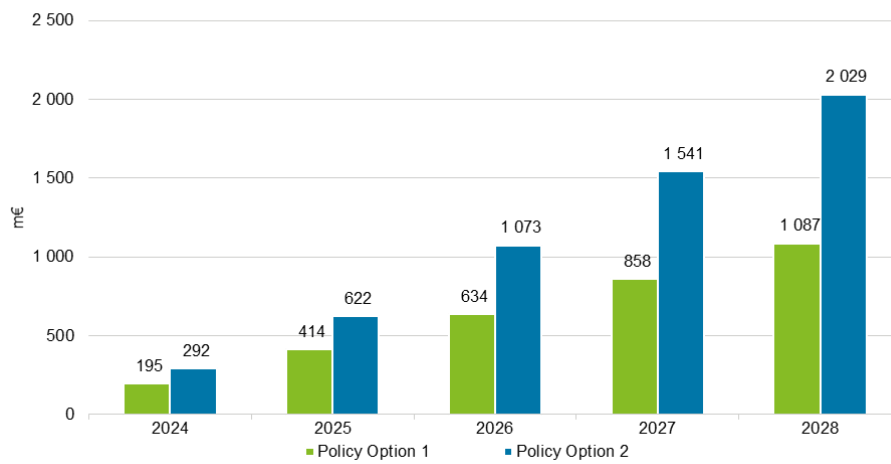
Employment Impact - 5. Statistics



Economic Impact direct - 6. Mobility



Economic Impact indirect - 6. Mobility



Employment Impact - 6. Mobility

