



EUROPEAN
COMMISSION

Brussels, 25.4.2018
SWD(2018) 127 final

COMMISSION STAFF WORKING DOCUMENT

IMPACT ASSESSMENT

Accompanying the document

**Proposal for a Directive of the European Parliament and of the Council on the re-use of
public sector information**

{COM(2018) 234 final} - {SWD(2018) 128 final}

Table of Contents

1.	INTRODUCTION: POLITICAL AND LEGAL CONTEXT.....	1
1.1.	Policy context.....	2
1.2.	Legal context.....	3
1.2.1.	Data protection legislation.....	4
1.2.2.	Other relevant legal instruments.....	5
1.3.	Economic context.....	5
2.	PROBLEM DEFINITION.....	7
2.1.	Ensuring the success of the open data policy in the evolving digital environment.....	7
2.2.	What are the problems and the problem drivers?.....	8
2.2.1.	Contextual drivers.....	8
2.2.2.	Problem area No. 1: Insufficient use of methods for access and re-use of dynamic data by machines.....	9
2.2.3.	Problem area No. 2: Market entry barriers and fragmentation.....	11
2.2.4.	Problem area No. 3: Insufficient availability of public and publicly funded data for re-use.....	13
2.2.5.	Problem area No. 4: Distortion of competition in the internal market.....	18
2.3.	Who is affected and in what ways?.....	20
2.3.1.	Public sector.....	20
2.3.2.	Private sector, in particular SMEs.....	21
2.4.	How will the problem evolve?.....	21
2.5.	REFIT considerations.....	22
2.6.	Access to private sector data for public interest purposes.....	23
3.	WHY SHOULD THE EU ACT?.....	24
3.1.	Legal basis.....	24
3.2.	Subsidiarity: Necessity of EU action.....	24
3.3.	Subsidiarity: Added value of EU action.....	25
4.	OBJECTIVES: WHAT IS TO BE ACHIEVED?.....	25
4.1.	General objectives.....	25
4.2.	Specific objectives.....	26
5.	WHAT ARE THE AVAILABLE POLICY OPTIONS?.....	27
5.1.	What is the baseline for assessing options?.....	27
5.2.	Description of the policy options.....	27
5.3.	Options discarded at an early stage.....	33
6.	WHAT ARE THE IMPACTS OF THE POLICY OPTIONS?.....	33
6.1.	Economic and data-specific impacts of the retained options.....	33

6.1.1. Baseline scenario	35
6.1.2. Policy Option 1 - Packaged solution: lower intensity of regulatory intervention .	36
6.1.3. Policy Option 2 – Packaged solution: higher intensity regulatory intervention	40
6.2. Impact on SMEs	42
6.3. Social and environmental impacts	43
6.3.1. Baseline scenario	43
6.3.2. Policy Option 1 - Packaged solution: lower intensity regulatory intervention.....	43
6.3.3. Policy Option 2 – Packaged solution: higher intensity regulatory intervention	44
6.4. Member States' and stakeholders' views	44
7. HOW DO THE OPTIONS COMPARE?	46
8. PREFERRED OPTION	48
8.1. Estimated impact of the preferred option.....	49
8.2. REFIT (simplification and improved efficiency)	50
9. HOW WILL IMPACTS BE MONITORED AND EVALUATED?.....	51
ANNEX 1: PROCEDURAL INFORMATION.....	55
ANNEX 2: CONSULTATION ON PSI DIRECTIVE REVIEW	60
ANNEX 3: VISUALISATION OF THE VIEWS EXPRESSED BY STAKEHOLDERS IN THE PUBLIC ONLINE CONSULTATION ON THE REVIEW OF THE PSI DIRECTIVE.....	69
ANNEX 4: WHO IS AFFECTED AND HOW?	74
ANNEX 5: EVALUATION REPORT	80
ANNEX 6: THE EVOLUTION OF THE DIRECTIVE ON THE RE-USE OF PUBLIC SECTOR INFORMATION.....	81
ANNEX 7: OPEN DATA IN AN INTERNATIONAL CONTEXT	83
ANNEX 8: ANALYTICAL METHODS	87
ANNEX 9: SUCCESS STORIES ON OPEN DATA.....	97

1. INTRODUCTION: POLITICAL AND LEGAL CONTEXT

This impact assessment accompanies the proposal for a review of the Directive of the European Parliament and of the Council amending Directive 2003/98/EC on the re-use of public sector information¹ (PSI Directive). The Directive was adopted by the EU legislators in order to harmonise the basic re-use conditions across the EU and to remove major barriers to re-use in the internal market, thus ensuring a competitive environment conducive to the development of a market for information-based products and services. The Directive introduced provisions on non-discrimination, charging, exclusive arrangements, transparency, licensing and practical tools facilitating the re-use of public sector documents.

The Directive was revised in 2013. The modifications introduced an obligation to allow the re-use of public data, access to which is granted under national legislation, expanded the scope of the Directive to include documents from public libraries, museums and archives, established a default charging rule limited to the marginal cost for reproduction, provision and dissemination of the information, and obliged public sector bodies to be more transparent about the charging rules and conditions they apply. The amending Directive was implemented into national legislation by all 28 EU Member States.

Article 13 of the Directive calls on the European Commission to carry out a review of the application of the Directive and to communicate the results, together with any proposal for amendments, before 18 July 2018. At the same time, this review is an important part of the initiative on accessibility and re-use of public and publicly funded data announced by the Mid-Term Review of the Digital Single Market (DSM) strategy.

In the last years, the Commission took a series of key measures to improve the framework conditions for data-driven innovation in Europe. With the General Data Protection Regulation² (GDPR) and the revision of the ePrivacy rules³, the EU has set a solid framework for digital trust which is a precondition for a competitive EU data economy. With the objective of further improving the efficient use of data across the EU, the Digital Single Market strategy put in place a series of important actions in this direction.

The Commission is now proposing a package of measures as a key step towards a common data space in the EU on a scale that will enable the emergence of new data-based products and services. The measures put forward are: 1) a proposal for a review of the Directive on the re-use of public sector information (PSI Directive)⁴, 2) a technical

¹ COM (2018) 234.

² OJ L 119, 4.5.2016, p. 1.

³ COM(2017)10 final.

⁴ COM(2018) 234.

update of the Recommendation on access to and preservation of scientific information⁵, and 3) guidance on sharing private sector data⁶. The proposed measures cover different types of data and therefore have different levels of intensity. At the same time, they all work towards the broader goal of bringing together data, as a key source of innovation and growth, from different sectors, countries and disciplines into a common data space.

The present document builds upon the evaluation of the functioning of the Directive and assesses a number of policy options aimed at updating the regulatory framework.

1.1. Policy context

The public sector in all Member States produces vast amounts of data, e.g. meteorological data, digital maps, statistics and legal information. This information is a valuable resource for the digital economy. It is not only used as valuable raw material for the production of data-based services and applications, but also brings greater efficiency to the delivery of private and public services and better informed decision-making. Therefore, the European Commission has been promoting the re-use of public sector information for several years.

Access to and re-use of public sector information together with the EU open data policy play a major role in improving transparency of public services delivery and the use of supporting technology by public sector bodies in general across the EU. There has been a general trend across the EU indicating significant improvement on cross-border availability of digital public services and accessibility of public websites from mobile devices.⁷

As an important feature, the EU open data policy is implemented through close collaboration with Member States, notably in a dedicated expert group⁸, through the European Data Portal⁹ supporting publication of datasets and their discoverability across Europe, and by research and innovation actions under the H2020 programme. The Commission has also given guidance to Member States on charging, formats and key datasets in Notice 2014/C 240/01.

Countries outside the EU have also adopted policies and legislation to unleash the power of government data, with the United States being the pioneer and others, such as Japan, Canada or Australia, following suit (see Annex 7 for a comparative overview).

The importance of openness with respect to publicly held and publicly funded data, in particular research data, is recognised globally. In particular, OECD Council

⁵ C(2018) 2375.

⁶ SWD(2018) 125.

⁷ <https://ec.europa.eu/digital-single-market/en/news/new-study-egovernment-services-europe-improving-cross-border-availability-services>

⁸ <https://ec.europa.eu/digital-single-market/en/news/public-sector-information-group-main-page>

⁹ www.europeandataportal.eu

Recommendations exist on access to and use of public sector information¹⁰ and on access to research data from public funding¹¹. Additionally, G8 leaders signed an Open Data Charter¹² which set out five strategic principles that all G8 members are to act on. These include an expectation that all government data will be published openly by default, alongside principles to increase the quality, quantity and re-use of the data that is released. G8 members have also identified fourteen high-value areas – from education to transport, and from health to crime and justice – from which they will release data.

In its commitments to implement the G8 Open Data Charter¹³, the EU also committed to the publication by the EU Member States of core and high-value datasets in line with the G8 principles. Creating a list of such datasets, which would be open and free by default, has the potential to facilitate the emergence of a new range of pan-European data products and services developed on the basis of mutually complementary cross-border datasets. However, the OECD Council Recommendations and the Open Data Charter are not compulsory for Member States to implement, therefore a stronger instrument proves to be necessary at EU level.

Apart from political commitment, there is also broad stakeholder support for the inclusion of the datasets defined in the category of 'core data sets', as defined in the G8 Open Data Charter, so as to ensure their immediate availability in Europe¹⁴, with some categories standing out as particularly relevant. These include above all geospatial data, closely followed by information in the field of transport, statistics, earth observation, environment and public finances.

1.2. Legal context

The PSI Directive is a legal instrument allowing for the implementation of a horizontal policy, with a wide field of application. A detailed discussion on the interplay between the Directive and other EU legal acts and policies (in particular the challenges and opportunities related to legislation on the protection of personal data, the Database Directive and the INSPIRE Directive) is presented in the Evaluation Report annexed to this Impact Assessment¹⁵. The evaluation indicates that the PSI Directive is overall coherent with other relevant EU legislation. Nevertheless, a technical clarification on the relation with in particular the Database Directive and the INSPIRE Directive may be useful.

¹⁰ 2008 Council Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information (C(2008)36).

¹¹ Recommendation of the Council concerning Access to Research Data from Public Funding (C(2006)184). See also the 2016 Daejeon Declaration on Science, Technology, and Innovation Policies for the Global and Digital Age calling on the OECD to assess the need to revise this Recommendation in order to move towards open science and reaping the benefits of Big Data.

¹² <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

¹³ <https://ec.europa.eu/digital-single-market/en/news/eu-implementation-g8-open-data-charter>

¹⁴ <https://ec.europa.eu/digital-single-market/en/news/consultation-guidelines-recommended-standard-licences-datasets-and-charging-re-use-public>

¹⁵ SWD(2018) 145.

1.2.1. Data protection legislation

The rules on re-use of public sector information must be applied in full compliance with data protection legislation. This is made clear in the text of the PSI Directive. The relationship with the PSI Directive is also made explicit in recital 154 of the GDPR which states 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]'*.

The Article 29 Working Party on Open Data and Public Sector Information Re-use¹⁶ stated in its opinion: *'wherever personal data are involved, data protection law must help guide the selection process of what personal data can or cannot be made available for reuse and what measures to take to safeguard personal data'*. Any processing of personal data needs to be based on one of the legal grounds foreseen in the GDPR.

As demonstrated by the results of the online public consultation, the majority of stakeholders agree that the PSI Directive is well aligned with current and new rules on the protection of personal data. This suggests that the Directive contains sufficient safeguards at the level of legislation.

However, some stakeholders voiced uncertainty over the functioning of the PSI Directive in the changing regulatory environment. From the stakeholder dialogue process, it emerged that while the principle of precedence of data protection rules over PSI re-use obligations is undisputed and well understood, public sector bodies may encounter practical implementation questions on how to facilitate re-use while ensuring compliance with the GDPR in situations where certain public registers also contain personal data (e.g. car registration databases or hospital records). This most often concerns the suitability of techniques that can be used for anonymization or ways by which purpose limitation can be ensured. In some Member States the legislation transposing the PSI Directive addresses such issues (e.g. in Belgium the legislation implementing the PSI Directive foresees that public sector bodies can seek advice from the data protection authority on the specific techniques to be used¹⁷), while in others this is tackled at the level of the recommended licensing agreements between the data holder and the re-user (e.g. Spain¹⁸).

In order to help public sector bodies with applying anonymization techniques, this review acknowledges that anonymization has a cost and that it should be possible for public sector bodies to recover such costs as part of the marginal costs or full cost recovery principles. In addition, the support towards development of privacy-preserving technologies is an important part of the Horizon 2020 innovation actions in the field of

¹⁶ Opinion 06/2013 on open data and public sector information ('PSI') reuse, Article 29 Data Protection Working Party, 5 June 2013.

¹⁷ Article 3(3) of the Belgian Federal law " Loi relatif à laréutilisation des informations du secteur public, Moniteur belge no 153 of 3 June 2016.

¹⁸ Upcoming Spanish Royal Decree for the Reuse of Public Sector Information in the Central Government

the data economy – a wider policy framework to which the review of the PSI Directive belongs.

1.2.2. *Other relevant legal instruments*

In terms of the interplay with intellectual property rights, the PSI Directive excludes from the scope content for which parties other than the public sector bodies ('third parties') hold intellectual property rights. Public sector bodies cannot give away what they do not own.

Some specific issues have been signalled in the evaluation process about the interaction between the PSI Directive and the Database Directive¹⁹. While public sector bodies have a right under Article 7 of the Database Directive to prevent extraction and/or reutilization of substantial content of databases on which *'there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents'*, such right cannot be invoked in order to prevent re-use permitted in accordance with the provisions of the PSI Directive. The ongoing reviews of both the Database²⁰ and the PSI²¹ Directives can help to clarify this issue.

Of similar importance is the relation between the PSI and the INSPIRE Directive. In 2007, the INSPIRE Directive laid down general rules aimed at the establishment of the Infrastructure for Spatial Information in the European Union, for the purposes of environmental policies and policies which may have an impact on the environment. The Directive applies to data held by or on behalf of public authorities for the performance of their public tasks. Article 2 and recital 8 of the INSPIRE Directive establishes that it is without prejudice to the PSI Directive, the objectives of which are complementary. However, it must be made sure that the INSPIRE Directive does not enter into conflict with the PSI Directive, and that any spatial information held by public bodies can be reached according to the rules laid out by the PSI Directive.

1.3. **Economic context**

The European data market is growing fast and could, with the right framework conditions in place, grow to 4% of the overall EU GDP.²² Since the public sector is one of the economy's most data-intensive sectors²³, the EU open data market²⁴ is a key building block of the overall EU data economy.

¹⁹ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31996L0009>

²⁰ SWD(2018) 146.

²¹ COM(2018) 234.

²² European Data Market study, IDC-Open Evidence, 2017, <http://datalandscape.eu/study-reports>.

²³ OECD Digital Economy, Outlook 2017, p. 220.

²⁴ The size of the EU open data market captures the aggregate value of products and services derived from open data exchanged in the European Union. The market value does not include the the direct, indirect, and induced effects of the open data market on the whole EU data economy.

Public sector information is used across the economy by a range of companies, but it is particularly important for the growth of start-ups and SMEs. An EU incubator for data start-ups discovered a positive correlation between strong and proactive open data policies in Member States and the number of successful applicants from those countries.²⁵ In a recent survey²⁶ of 450 executives from European digital start-ups and conventional businesses, 50% confirmed they had used open data to build a new product or service.

It is also a critical asset for the development of new technologies such as Artificial Intelligence (AI), which requires the processing of vast amounts of high-quality data.²⁷

Numerous studies in the last decade have attempted to quantify the economic value generated by the re-use of public sector information.²⁸ All the studies converge in concluding that the re-use of open data leads to substantial economic and societal benefits. Depending on the methodology used (e.g. the definition of 'open data', the focus on the direct impact on the data economy or wider socio-economic impacts, etc.) various figures have been put forward. The support study for this IA assesses the current direct economic value of open data²⁹ to be 52 billion EUR a year for the 28 EU Member States.³⁰

The European Data Portal 2015 study led by CapGemini³¹ (and updated in 2017³²) undertook a separate assessment of the economic benefits for open data by looking at four key indicators and how they could evolve until 2020: direct market size, number of jobs created, cost savings and efficiency gains.

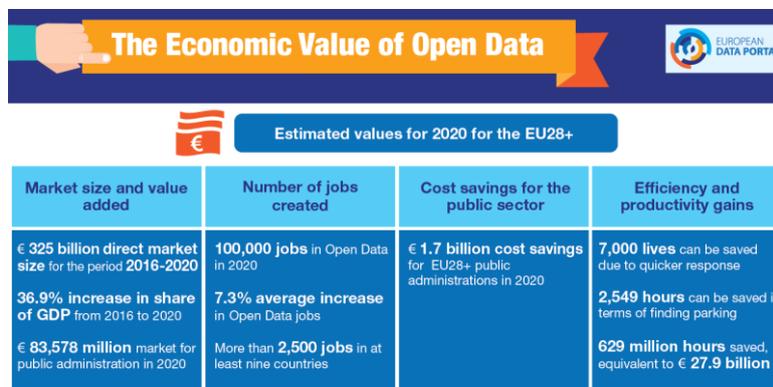


Figure 1- The economic value of open data for EU28+ in 2020
Source: European Data Portal, Analytical Report No 9: The Economic Benefits of Open Data, 2017

²⁵ Assessment of the impact of the ODINE programme, 2017, p. 12.

²⁶ <http://opendigital.economist.com/digital-economy>

²⁷ <https://www.techuk.org/insights/opinions/item/10708-guest-blog-peter-wells-odi-access-to-data-is-key-to-a-competitive-ai-market>

²⁸ See the overview of key studies in European Data Portal, Analytical Report No 9: The Economic Benefits of Open Data, 2017.

²⁹ Value that is generated by developing goods and services based on PSI, chiefly by SMEs.

³⁰ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

³¹ Creating Value through Open Data report, European Data Portal, November 2015.

³² Economic Benefits of Open Data, European Data Portal, December 2017.

Next to the considerable economic benefits, opening up government data leads to a whole range of social benefits³³ experienced by the consumers of products and services offered by direct open data re-users.

Open data has the potential to increase the efficiency of government through better policy making, including 1.7 billion EUR cost savings for the EU 28 public administrations.³⁴ Finally, it helps in bridging the gap between government and citizens in terms of information³⁵ and, in general, leads to increased social inclusion and empowerment, civic participation, and supports personal decision-making capabilities.³⁶

2. PROBLEM DEFINITION

2.1. Ensuring the success of the open data policy in the evolving digital environment

The **Evaluation Report of the PSI Directive**³⁷ (annexed to this Impact Assessment) shows that **overall the instrument works well**. The report comes to the following key conclusions:

The PSI Directive continues to contribute to the achievement of its main policy objectives, which are to stimulate the digital content market for PSI-based products and services, to stimulate cross-border exploitation of PSI and to prevent distortions of competition on the EU market. Since 2015, Member States have made substantial progress in terms of open data maturity and more datasets of better quality have become available. At the same time, the instrument has had a favourable impact on transparency, citizen empowerment and public sector efficiency.

However, there are a number of issues that would need to be addressed in order to fully exploit the potential of public sector information for the European economy and society: provision of real-time access to dynamic data via adequate technical means, increasing the supply of high-value public data for re-use (e.g. from semi-private entities executing public tasks and research establishments), the existence of new forms of exclusive arrangements, the use of exceptions to the principle of charging the marginal cost and the relationship between the PSI Directive and certain related legal instruments.

The review should help to address the remaining barriers to the re-use of PSI and address the new needs created by technical change. In addition, the review of the Directive can address the trend of some Member States embarking on legislative initiatives addressing the re-use of new categories of datasets, while other Member States are not. It is also important to remember, however, that any intervention designed to address these issues

³³ Granickas, K.(2013), *'Understanding the Impact of Releasing and re-Using Open Government Data'*. ePSI Platform Topic Report No. 2013/08, August 2013.

³⁴ Creating Value through Open Data report, European Data Portal, November 2015.

³⁵ Ibid.

³⁶ Granickas, K.(2013), *'Understanding the Impact of Releasing and re-Using Open Government Data'*. ePSI Platform Topic Report No. 2013/08, August 2013. p. 13.

³⁷ SWD(2018) 145.

will need to take into account the existing budget constraints for public authorities. While certain measures may be highly desirable from the perspective of opening up more public sector data and making it more easily re-usable, the costs of their implementation will need to be carefully weighed against the expected benefits, and their intensity may need to be modulated respectively.

2.2. What are the problems and the problem drivers?

2.2.1. Contextual drivers

The advent of new technologies, in particular in the areas of data analytics and the Internet of Things, has had a significant impact on the way data can be exploited in the economy. Two key consequences of this trend are a growing demand from businesses for dynamic data and the need to have access to a larger pool of data.

Dynamic data is captured by sensors, which become ever smaller, more performant and cheaper.³⁸ They hold valuable real-time data from industrial processes, personal activities or physical phenomena. This trend is further enhanced by increasing connectivity and sensor usage. It is estimated that 11 billion devices are currently connected to the Internet and this should almost triple to 30 billion by 2020. The public sector, in line with widespread digitisation efforts and the impact of initiatives such as Smart Cities, will also produce growing amounts of sensor-generated data. Access to such dynamic data can be greatly enhanced if public sector data holders use application programming interfaces (APIs), which allow controlled data access and exchange leading to new ways of interacting in an online environment.³⁹

Apart from the rise in commercial exploitation of dynamic data, another contextual driver fuelled by advances in data processing technologies⁴⁰ is the need for businesses to be able to combine data from different sources. There are increasing economic opportunities in accessing, using and re-using large volumes of data of different types and from different sources. Therefore it is important that different types of data are available for use in the data economy. At the same time, there is considerable heterogeneity in the types of data available from one Member State to another⁴¹, with some countries moving ahead through legislation and making more types of data freely available in new areas of public interest, while others are lagging behind.

In this context, there is an understandable tendency on the part of commercial players to seek preferential access to large public datasets. These datasets can give a competitive advantage for added-value information services and products, or be used to train artificial intelligence systems. This has led to the emergence of cases in which preferential arrangements between public sector bodies and commercial players have led to data lock-

³⁸ <https://www.theatlans.com/charts/BJsmCFAI>

³⁹ <http://nordicapis.com/tracking-the-growth-of-the-api-economy>

⁴⁰ <http://www.oecd.org/sti/ieconomy/data-driven-innovation.html>

⁴¹ See: Categories of data excluded from re-use, Impact Assessment Support Study, Deloitte, SMART 2017/0061 and findings from the INSPIRE Directive Evaluation exercise SWD(2016) 273 final, p.41.

in, thus making the prospect of the exploitation of such data by other actors commercially unviable.

Finally, the long-standing budgetary pressure on public sector bodies encourages some of them to cover a part of their operating costs from commercial activities, such as charging for the re-use of their data. Although seemingly it would appear that in an environment where public authorities are subject to budget constraints, revenue generation from data could be a viable source of financing, it has been proved that this approach leads to sub-optimal results from a macro-economic perspective.⁴² One of the widely accepted and evidence-based⁴³ principles of the amending Directive 2013/37 EU is that public sector bodies should not apply any charges or limit them to what is needed to cover the marginal costs of reproduction and dissemination of the data. Since in the digital environment such costs are minimal, the charging principle of the PSI Directive can be said to be that of zero (or at most marginal cost) pricing.

These contextual drivers are at the basis of the following four problem areas identified in the evaluation of the Directive and discussed below: 1) Insufficient use of methods for access and re-use of dynamic data by machines; 2) Market entry barriers and fragmentation; 3) Insufficient availability of public and publicly-funded data for re-use; 4) Distortion of competition in the internal market. All these problem areas prevent the full exploitation of the economic opportunities offered by public sector information and create fragmentation on the single market, albeit to different degrees. The public consultation process has confirmed that the first three are the most significant and would require swift regulatory intervention. The fourth one is a new, emerging issue, which has had limited negative impact so far, but given the growing significance of data in the economy it is reasonable to expect that its impact will grow, thus justifying intervention at this stage.

2.2.2. Problem area No. 1: Insufficient use of methods for access and re-use of dynamic data by machines

The percentage of dynamic data being made available by public sector bodies, despite its high re-use potential, is still very low.⁴⁴ The Impact Assessment study confirms that the delivery of dynamic data is an important challenge for public bodies collecting that type of information as many of those are not adequately equipped.⁴⁵ This can be due to the technical limitations and requirements that appear when the frequency to update datasets is shorter than a minute, or to the lack of resources. Additional drivers behind this problem are low awareness of the benefits of APIs and low awareness of the value that real-time data has for the creation of commercial applications. The insufficient use of

⁴² See: Models of Public Sector Information Provision via Trading Funds, Cambridge University, 02/2008, <http://www.berr.gov.uk/files/file45136.pdf>

⁴³ See: Annex 6 of the Impact Assessment accompanying the previous review of the Directive: <https://ec.europa.eu/digital-single-market/en/news/commission-staff-working-paper-impact-assessment-accompanying-document-proposal-directive>

⁴⁴ <https://www.linkedin.com/pulse/reuse-open-data-opportunity-spain-alberto-abella>

⁴⁵ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

APIs or other methods for access and the re-use of dynamic data by machines has been confirmed by stakeholders, in particular from the re-user community, who recognise this issue as one of the main barriers for data re-use. This was for instance emphasised in the Commission's 2017 public consultation on 'Building a European data economy initiative' where the summary report noted that 68% of respondents clearly support an increased use of APIs.⁴⁶

In a recent study based on a representative sample of the total volume of public sector information (20,000 datasets), it was found that while the majority of services generated from open data are based on real-time data (66%), less than 1% of the data published in open portals are updated in real time.⁴⁷ While the usability and commercial potential of dynamic data is directly related to its immediate availability, there is little awareness and competence among the public sector institutions regarding the available tools and methods, such as APIs, that could enable such availability. Given the budget pressure, there is also reluctance to invest in new technologies.

Apart from the fact that public data is not systematically provided through APIs, there are considerable differences between Member States in this area.⁴⁸ This is for example clear in the way in which their national portals can be accessed. Austria tops the list, with 71-85% of all visits to its portal deriving from machine traffic. Romania comes second, with 41-55% of its traffic coming via API calls, followed by the UK with 26-40% of visits. 22 out of 28 Member States had API traffic of less than 10%.⁴⁹ The figures are encouraging in the sense that the most costly investments on the national level have already been made. Yet, it can be safely assumed that on the lower levels of government, the provision of APIs and their actual usage are less widespread, despite clear benefits: cities produce data in many forms and from many sources, and the wide variety of formats makes it difficult to scale open data applications from city to city.⁵⁰

The results of the online public consultation indicate that stakeholders in general ask for wider availability of dynamic data and APIs from the public sector.⁵¹ 77% of stakeholders were in favour or strongly in favour of further investments to be made to encourage public sector bodies to provide dynamic data in real time and invest in technical solutions facilitating data usability, in particular through APIs. Respondents also requested dynamic data generated by the public sector for further re-use. This position was strongly voiced by the PSI re-users in particular.

⁴⁶ <https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-building-european-data-economy>

⁴⁷ http://informecotec.es/media/INFORME_REUTILIZACION-DE-DATOS.pdf

⁴⁸ See: France (<https://api.gouv.fr/apis>) and Italy (<https://developers.italia.it/en/datigov/>)

⁴⁹ <https://www.europeandataportal.eu/en/dashboard#2017>

⁵⁰ Harmonised Smart City APIs – a Cookbook for cities,

https://www.databusiness.fi/content/uploads/2017/10/20171109_HarmonisedSmartCityAPIs_WEB.pdf

⁵¹ Results of the public online consultation, December 2017.

The findings of the Evaluation Report⁵² and a dedicated external study⁵³ confirm that instead of this being just a legal requirement which is imposed on public sector bodies, additional practical support is necessary for many public sector bodies across Europe to move towards web-based functionalities and a more wide-spread use of dynamic data and APIs in general. Public sector bodies seem to need additional assistance to be able to offer data as a service and in a dynamic way. It is important to note that this approach cannot apply to all types of PSI (e.g. court decisions), but will be invaluable for others, as document-based mechanisms in various cases will not work (e.g. in the case of location apps).

While other technological trends such as the need for machine-readable data and interoperability have been integrated in the last revision of the PSI Directive, the need for timely availability of dynamic data is just mentioned once in recital 12.

As a consequence, the insufficient use of technologies for the management of dynamic data by the public sector results in more difficult access to public data repositories, lower machine-readability of dynamic data and lower suitability of data for processing (visualisation, integration into 3rd party applications). This in turn leads to a sub-optimal use of dynamic data from the public sector for the creation of value-added services.

2.2.3. Problem area No. 2: Market entry barriers and fragmentation

Public institutions holding high-value datasets (such as hydrographic or meteorological data, or digital maps) are often expected or required to maximise revenue from the re-use of this data in order to cover part of their operational costs.

This is, however, unjustified from the 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 triggers a surge in usage. It is produced as part of a public service, so its creation does not depend on market forces. Most importantly, cheap or free government data generates extra commercial activity, especially by SMEs, which translates into more jobs and revenue from taxes. In addition, the public sector is itself a key user of the data, so part of the income for one public sector body leads to costs for others, as well as to a suboptimal use of the data for public purposes. Finally, charging generates substantial costs for public sector bodies producing the data, since it presupposes, amongst other things, that an invoicing and control system is in place.

These arguments were the basis of the 2013 change from a full-cost charging regime to a regime based on (at maximum) the marginal costs of dissemination. Some of the arguments are further elaborated in section 6.1 below.

The current PSI Directive presents a number of exceptions to the rule that public sector bodies can charge at maximum the marginal cost of dissemination for making their datasets available. This is the case for public sector bodies that are required to generate

⁵² SWD(2018) 145.

⁵³ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

revenue to cover a substantial part of their costs relating to the performance of their public tasks. While there is no evidence that would indicate that there has been an excessive use of exceptions, the Impact Assessment study and the Evaluation Report⁵⁴ confirm that the use of exceptions has led to incoherent and widely varying charging practices for the re-use of similar datasets across the EU.⁵⁵ This is made worse by diverging licensing practices, as confirmed by 67% of respondents to the online consultation.

The resulting situation is one where there is market fragmentation, with higher entry barriers in some Member States than in others, as shown in the example below. Moreover, the Impact assessment study found that heterogeneous practices in terms of charging persist not only between Member States but also between public sector bodies within the same state⁵⁶. Across the EU, public organisations charge considerably different prices for a dataset with the same quality and type of data. Re-users from different countries complained about these different practices.⁵⁷ In addition to market fragmentation, the charging and licencing practices negatively affect the extent of PSI re-use by SMEs and start-ups. They result in an uneven playing field among re-users, given that large, multinational companies can easily afford the acquisition of public datasets, which are outside the reach of innovative SMEs or start-ups. This translates into an opportunity cost to be borne by the entire society as a more limited number of innovative services and products can be developed.⁵⁸

Example No 1: a Swedish company Seapilot produces digital navigation apps based on marine chart data from hydrographic offices across the EU. However, widely divergent pricing models (e.g. one off payment, royalties, fees linked to updates) and the resulting charges applied (EUR 2,745 in France to EUR 18,900 in Italy⁵⁹) make it increasingly difficult to compete on a global scale, especially given that equivalent US data is free of charge.

Example No 2: In Belgium, the company register is made available for re-use at a price of 75,000 EUR annually⁶⁰, leading to a situation in which start-ups and SMEs face a market entry barrier so that this dataset of high socio-economic interest is effectively re-used by only two companies.

As for the views of stakeholders on charging and its impact on market fragmentation, in the online public consultation most of the respondents indicated they were in favour of abolishing the current exceptions or at least clarifying the circumstances under which these exceptions could be allowed. However, there was a clear split between PSI holders and PSI re-users, with the latter clearly in favour of abolishing or at least clarifying these

⁵⁴ SWD(2018) 145.

⁵⁵ Still applied across the EU by e.g. meteorological and hydrographic institutes, business registers, legal information repositories, mapping agencies, vehicle registration databases.

⁵⁶ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

⁵⁷ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

⁵⁸ Ibid.

⁵⁹ Ibid.

⁶⁰ <https://economie.fgov.be/fr/themes/entreprises/banque-carrefour-des/services-pour-tous/banque-carrefour-des-3>. Some of the data are, however, available at low cost, in line with the relevant sectoral EU legislation

exceptions and the former mostly of the opinion that they should not be changed and need no clarification.

2.2.4. Problem area No. 3: Insufficient availability of public and publicly funded data for re-use

2.2.4.1 Data held by public undertakings in the transport and utility domains

Public sector tasks may be carried out not only by public authorities themselves but can instead be entrusted to entities with organisational or management links to the public sector, or those that lack such links but benefit from public funding (public undertakings). In some cases public sector tasks are also performed by private entities which act on the basis of special or exclusive rights or concessions from public sector bodies. This is typically the case in the transport and utility domains (provision of water, electricity, etc.). Currently, data held by public undertakings in these domains does not fall under the scope of the Directive.

Data held by entities active in the transport and utility domains is amongst the most valuable for stakeholders in the data economy and it can serve as the basis for a number of added-value services and applications. Opening up this data is of considerable value for commercial re-use. Topics such as environment, transportation, energy or housing were among the domains most often consulted in open data portals in 2016.⁶¹ National figures confirm that data from transport and traffic domains are very popular. In Spain, of all the applications with a business model behind, 47% are created with transport data. It is also important to remember that due to their link with the public sector, entities active in transport and utility domains are subject to a dedicated public procurement regime.⁶² Moreover, historically these operators have had to operate in a stricter regulatory environment compared with other businesses. Examples are to be found in the area of infrastructure provision where ex-ante or ex-post regulation has been established to enhance competition and reduce monopolistic or dominant positions in these sectors. Often governments or public companies still hold large shares in these infrastructure companies, or these companies operate under concessions or other exclusive contracts with public sector bodies.

As explained in the Impact Assessment study⁶³, access to data from entities active in transport and utility domains depends on company preferences, Member States' open data strategies and legislation, and individual re-user activities. As a result, stakeholders of the re-user community are faced with a situation in the EU where data access and re-use happens at different speeds in different countries and opportunities to develop cross-border applications are limited. It has been stressed that the potential of data-based solutions in those sectors, especially in energy and transport, is significant, but there is

⁶¹ Impact Assessment of ODINE programme. 2017, IDC.

⁶² Directive 2014/25/EU on procurement by entities operating in the water, energy, transport and postal services sectors and Regulation (EC) No 1370/2007 on public passenger transport services by rail and by road.

⁶³ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

insufficient progress in opening up data by the companies themselves. In the transport sector, for example, the study found that in all Member States data availability varies widely across the transport modes. While examples of simple travel information applications are manifold, intermodal service examples are limited, especially as regards door-to-door travel applications.

The same study⁶⁴ found that the general interest of public undertakings to generate profits to have returns on the investments related to the specific datasets represents a major barrier to sharing their data. Another barrier to sharing data with public organisations is the complexity of regulation on public undertakings in general.

Some examples of what can be done when data from entities active in transport and utility domains is opened up are provided below.⁶⁵

Example No 1: A Copenhagen-based start-up Tomorrow⁶⁶ uses real-time data from energy grid operators to act as a data broker for historical energy data and to offer dynamic APIs to track the origin of electricity in real time and to forecast consumption, enabling lower emissions, higher renewables and cheaper electricity.

Example No 2: Transport for London (TfL)⁶⁷ released via an API over 200 data sets (bus and metro arrivals, departures, status, cycle hire docking station status, etc.) which created a community of 14,400 developers with over 600 apps. London has gained around 100 million GBP direct value by technological investment and TfL's open data ecosystem has led to the creation of about 500 directly and 230 indirectly related jobs. TfL praises the effect that the re-use of its data has had in reducing the commuting time of the passengers and thereby improving TfL's efficiency.

The examples above show the positive developments in these sectors. Still, overall there is a mismatch between the expectations of the re-users and the data actually available. For example, in some Member States only 3.9% of all the open data published is from the transport area.⁶⁸ Save for a few exceptions⁶⁹, data generated as a result of activities carried out in the utilities sector is not generally made available for re-use. Only 23% of the respondents to the public online consultation agreed that data in these sectors was currently available for re-use. The figure was higher (36%) for transport, while it was considerably lower (14%) for the utilities.

The stakeholders also indicated that data from the transport and utilities sector should be more widely available. A large majority (71%) of respondents to the online consultation, in particular from the re-user community, believe that data generated in the context of the provision of a public task by publicly owned companies or by independent economic operators, irrespective of the public or private nature of the data holder, should be made available for re-use. Similarly, an overwhelming majority of replies (81%) indicate that if

⁶⁴ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

⁶⁵ For other examples, see also Impact Assessment support study, Deloitte, SMART 2017/0061.

⁶⁶ <https://www.tmrow.com/#products>

⁶⁷ <https://tfl.gov.uk/>

⁶⁸ *Open data re-use: an opportunity for Spain?*, COTEC report, 2017.

⁶⁹ <https://data.fingrid.fi/en/about> or <https://opendata.reseaux-energies.fr/pages/accueil/>

there were an obligation to make data generated in the context of the provision of a public task available, such data should be available for all interested re-users for any purpose. Although a minority, there are also more cautious opinions on these issues, in particular from businesses in the fields of transport, energy, waste and water. In particular stakeholders from the transport sector have expressed concerns. They believe that data is crucial for maintaining their competitiveness and as a result any requirements to make it more open and re-usable should be carefully balanced, as new obligations in this regard may distort competition in the sector. Stakeholders from entities active in the transport and utilities sectors have also indicated that imposing data sharing obligations on them may have an impact on ensuring the security of critical infrastructure. Moreover, some are anxious that compliance with the new requirements will mean an additional administrative burden.

The lack of availability and re-usability of data in the relevant areas has led the Commission to propose sector-specific rules. The transport sector was covered by legislation on the provision of EU-wide multimodal travel information services⁷⁰, which will lead to the wide availability of a range of relevant datasets. In the energy sector, a recent proposal for a recast of the Electricity Directive includes provisions enhancing access to consumer data⁷¹, while in the water sector the Commission has proposed provisions on the sharing of water parameters data in the context of the review of the Directive on the quality of water intended for human consumption.⁷² However, these rules are driven by sector-specific concerns and focus on selected datasets. In the case of transport data, the rules cover both public and private entities, which has lowered the ambition level with regard to the availability requirements of some of the data.

Consequently, the full innovation potential of public and publicly funded data in the areas of public transport and the utilities sector is currently not realised, and this amounts to missed opportunities, given the economic and societal potential of such data.

2.2.4.2 Research data

Scientific information produced with public funding is an invaluable resource for innovation, for the broader economy and for addressing societal issues. There is also substantial evidence that data sharing is correlated with better science.⁷³ Nevertheless, scientific information is often not readily available for re-use, despite growing acceptance of the Open Access approach to research data both globally and in the EU.⁷⁴

This is related to the following factors:

- Fragmentation of policies due to the dependence on a wide range of institutional and (national or other) funding mandates and arrangements, and weak compliance: the funding mandates are often not properly implemented.

⁷⁰ http://eur-lex.europa.eu/eli/reg_del/2017/1926/oj

⁷¹ <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016PC0864R%2801%29>

⁷² http://ec.europa.eu/environment/water/water-drink/review_en.html

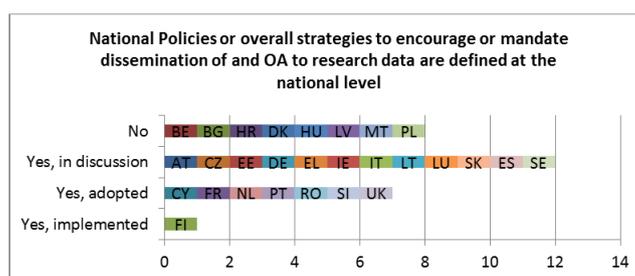
⁷³ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

⁷⁴ <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access>

- Policy framework not fully fit for purpose, given that the 2012 Commission Recommendation on access and preservation of scientific information⁷⁵ is no longer up to date. Notably, it does not take into account the recent EU policy developments in the areas of Open Science, and the spread of machine-generated and dynamic data.
- Lack of focus on re-use since open access mandates focus on access to scientific information, while re-use of this information is dealt with in disparate ways.
- Lack of incentives/rewards to allow the re-use of research data: Many researchers remain reticent on the idea of opening their results due to the lack of awareness on the potential benefits of open science. Also there are no individual rewards compensating for potential loss of competitive advantage of keeping results closed.

The Impact Assessment study⁷⁶ confirms that there is consensus in the literature about the lack of research data sharing and puts forward various examples from bibliometric analysis and surveys of researchers, ranging from only 13% up to to 79% of research data being shared. The study also confirms that there is a wide range of underlying reasons impeding the sharing of research data. Among them the most prominent causes seem to be the different data sharing cultures in the different disciplines, legal certainty regarding copyright over the datasets concerned or the reluctance of companies to share data with researchers when they know that the data will be openly shared. Many of these statements have also been expressed in a dedicated workshop on 'Open Research Data'⁷⁷ and in a public hearing on the PSI Directive review.⁷⁸

The 2012 Commission Recommendation on access and preservation of scientific information has created a framework for open access to research results at the European level. It has led to substantial progress across the EU in opening up research information.⁷⁹ At the same time Member States reports collected in 2017⁸⁰ show the limits of the non-binding approach. National policies or overall strategies to encourage or mandate Open Access to research data have been adopted in only 8 out of 28 Member States.



⁷⁵ C(2012) 4890 final.

⁷⁶ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

⁷⁷ <https://ec.europa.eu/digital-single-market/en/news/summary-report-workshop-open-research-data-within-context-directive-reuse-public-sector>

⁷⁸ <https://ec.europa.eu/digital-single-market/en/news/public-hearing-review-directive-reuse-public-sector-information>

⁷⁹ See the Commission staff working document on the review of the Recommendation.

⁸⁰ National Points of Reference reports.

Figure 2 - Member States reporting on national level strategies or policies for open access to research data

While it is true that some Member States also reported that such open access policies with respect to research data exist for at least 'some academic institutions and/or research centres' (18 Member States reported this to be the case), the FAIR principles, a community standard for findable, accessible, interoperable and re-usable research data, is supported either at national or at institutional level in only 16 Member States.⁸¹ Where mandates for open access are in place, they are often not enforced. 44% of respondents in the Figshare survey declare that they have an institutional mandate but it is not enforced.⁸²

Moreover, data resulting from publicly funded research is often not fully open or not open under harmonised and clear terms. This trend is shown by the data and analysis from the OpenAIRE Infrastructure⁸³ from October 2017, where H2020 publications account for 12,000 articles, but only 404 datasets.

In addition, there are substantial discrepancies in how open publications are available in the different research areas, and also to what extent research data is actually available. On the one hand, there is a flourishing landscape of research data repositories. Among the 1,381 research data repositories⁸⁴ examined in the project re3data⁸⁵, 95.5% were qualified as open, meaning accessible without any financial and technical barriers and 86.2% of their research data content is available in open access data at least partly. On the other hand, an analysis⁸⁶ of the URL links embedded in papers published by the American Astronomical Society over 15 years found that 44% of those links were broken 10 years after the publication. 15-20% of links pointing to curated data archives were broken, while links to project or personal websites decayed at much faster rate. Bibliometric analysis shows that data from only 13% of articles in top-level journals is available.

Finally, there is a problem of diversity and inconsistency of the licensing practices which are not oriented toward favouring re-use.⁸⁷ This issue was reported as a shortcoming of the current policy set-up by stakeholders from the research community in a workshop on Open Research Data⁸⁸. Legal solutions vary among repositories, and even among datasets. Under the FAIR principles⁸⁹ researchers retain a margin of discretion to decide on relevant re-use conditions. In many cases it is not clear what re-users are allowed to

⁸¹ National Points of Reference report (forthcoming).

⁸² https://figshare.com/blog/2015_The_year_of_open_data_mandates/143

⁸³ www.openaire.eu

⁸⁴ The project as per February 2018 lists more than 2,000 research data repositories.

⁸⁵ Kindling et al., The Landscape of Research Data Repositories in 2015: A re3data Analysis.

⁸⁶ Pepe A, Goodman A, Muench A, Crosas M, Erdmann C (2014) How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. PLoS ONE 9(8): e104798.

⁸⁷ Impact Assessment Support Study, Deloitte, chapter 6.1.2.; Kindling et al., The Landscape of Research Data Repositories in 2015: A re3data Analysis.

⁸⁸ <https://ec.europa.eu/digital-single-market/en/news/summary-report-workshop-open-research-data-within-context-directive-reuse-public-sector>

⁸⁹ <https://www.force11.org/group/fairgroup/fairprinciples>

do with the data, or there are unnecessary restrictions. For example, in two recent surveys on curated data resources, respectively over 30% and over 60% of respondents have reported to be unsure about the terms of the license to be used.⁹⁰

The large majority (81%) of respondents to the online consultation, representing different groups of stakeholders, agreed that there should be one common policy for open access in Europe, binding on all research funding organisations and academic institutions in Europe. This is currently not the case. Given the non-binding nature of the Recommendation, EU leverage to remedy the above-mentioned weaknesses is limited. As a result, the limited availability of research data for re-use makes it difficult to conduct further scientific research based on such data or to apply data analytics technologies and to mix such data with other public or private datasets to produce added-value services.

2.2.5. Problem area No. 4: Distortion of competition in the internal market

While the value of government data increases with the advances of technology, notably due to new fields of application such as the Internet of Things or artificial intelligence, the public sector often lacks expertise or sufficient funding to derive insights from the data it holds. New types of partnership agreements are emerging in which companies offer advanced data analytics services/infrastructure in exchange for preferential access to the data. With rapid technological advances and growing opportunities in the field of data analytics, these new kind of public-private partnerships will become all the more frequent.

The Impact Assessment study⁹¹ found that in practice public sector bodies are not sufficiently aware and trained to understand and spot the potential risks related to the establishment of exclusive arrangements with regard to the datasets that they hold. In certain situations, public sector bodies are financially obliged to rely on exclusive agreements in order to be able to afford the costs of new initiatives linked to, for example, digitisation or new services. In such circumstances, agreements that inadvertently lead to a situation of exclusivity on the data may pass unnoticed.

In order to create a level playing field among re-users, the PSI Directive in principle prohibits exclusive agreements for the re-use of public sector information, as well as any other form of discrimination between re-users. This means that publicly funded data cannot be given to only one or a limited set of companies, and that the conditions for re-use must be the same for all.

Taking into account that the number of public-private arrangements which involve the transfer of public sector information will be on the rise due to the new possibilities of data exploitation that the public sector itself cannot ensure, a targeted policy response appears necessary.

⁹⁰ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

⁹¹ Ibid, Annex A.

This is also motivated by the fact that public bodies and potential re-users of the data believe that exclusive exploitation of data is very rarely justified. However, businesses commonly entering into exclusive agreements tend to say that this advantage in exploitation of data is an essential incentive for the private sector to invest into projects in which otherwise it would have not invested at all.⁹²

The stakeholders' consultation confirms such conflicting views from yet another angle: while Member States generally report they are not aware of exclusive arrangements based on public sector information, other stakeholders signal that problems in this area persist. Just 38% of the respondents to the public online consultation consider that such agreements are used only exceptionally and are limited to the cases allowed under the Directive (digitisation of cultural resources, public service necessity). Additionally, among the respondents to the public consultation who indicated that certain provisions of the Directive could be clarified, many referred to the area of exclusive agreements. This request was also reiterated by several representatives of the re-use community present at the High-level Round-table discussion on Public Sector Information in January 2018.⁹³

The contradictory voices referred to above can be explained by the nature of the examples mentioned by stakeholders and experts in the context of the Impact Assessment study⁹⁴ and the broader consultation process. They mostly concern new types of agreements that are not a classical relation between a public sector body and a re-user, but involve a compensation in kind in exchange for preferential conditions to use the data. This kind of situation is not addressed by the current provisions of the Directive.

The examples mentioned by stakeholders come in the first place from the context of smart cities. They concern cases of the exclusive use of sensor data by the company installing the sensors, the exclusive use of data on street lighting, and arrangements between certain cities and service providers of navigation and localisation services involving privileged use of public data. Some local authorities are actively pursuing these kinds of partnerships, in order to limit their costs and to get hold of some of the data from private companies.⁹⁵ Another area concerns the use of health records. Examples of agreements involving public health services and companies in the UK, Belgium and Italy have recently attracted the attention of the press and the relevant authorities for reasons related in particular to data protection.⁹⁶

The new types of arrangements where in kind compensation is given in exchange for preferential use of public data have clear advantages for both the public authority and the

⁹² Ibid: p. 254.

⁹³ <https://ec.europa.eu/digital-single-market/en/news/summary-report-high-level-round-table-discussion-public-sector-information-re-use>

⁹⁴ See: examples in the Impact Assessment Support Study, Deloitte, SMART 2017/0061.

⁹⁵ Workshop on access to privately held data for public interest reasons, 26 June 2017, presentations by Amsterdam and Hamburg cities.

⁹⁶ In the field of health data, such arrangements can be particularly undesirable, as testified by the results of the Commission's public consultation 'Transformation Health and Care in the Digital Single Market' in which most respondents favour publicly-controlled, non-commercial solutions to ensure innovation based on health data,(Synopsis Report upcoming).

companies involved. At the same time, they could lead to a situation in which no other commercial re-users would have an incentive to re-use the same dataset in a similar manner, which would in practice circumvent the provisions of the PSI Directive on non-exclusivity and non-discrimination, as illustrated by the example below.

Example: One of the largest National Health Service providers in Britain provided Google-owned artificial intelligence company DeepMind with 1.6 million patient records to assist in the management of acute kidney injury by using technology to track patients' symptoms and send alerts to doctors. Although this partnership does not in principle exclude other companies from being provided access to NHS data (and re-use it), Google's lead in this regard could distort incentives for competition. Moreover, the NHS will have invested, and continues to do so, resources and human capital in this partnership.

Another concern relates to the different negotiation power and knowledge base of the partners, which could be exploited by the private partner. Due attention needs to be paid to contract clauses or situations where preferential use of the data is considered as one of the benefits for the private party.

All of the above confirms that in the context of new types of agreements between public authorities and private companies, in which the value of public datasets can be captured by one or a limited number of players, the risk of side-lining smaller entities, in particular start-ups and SMEs, is considerable. This risk will tend to grow in the near future with the expected rise of the number of public-private arrangements that will be made to meet with the growing opportunities and following demand in data analytics.

2.3. Who is affected and in what ways?

The PSI Directive broadly affects two sets of stakeholders: public sector bodies and re-users. The beneficiaries of open data policies can be found in all economic sectors.

Annex 4 to this document shows in detail how many stakeholders are affected and how.

2.3.1. Public sector

The main aim of the Directive is to ensure the availability of information held by the public sector, therefore public sector bodies across the EU are affected by its provisions. They bear the compliance costs of the legislation including for any changes to the regulatory framework. At the same time, public sector bodies use open data to improve policies and the delivery of public services. The rise of 'data for policy'⁹⁷ initiatives testifies the growing importance of data for the public sector at large. In addition, the research and educational organisations and public undertakings in the transport and utilities sector would be affected as new addressees of the Directive. However, they are also among the beneficiaries of open data policies.

⁹⁷ https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf.

2.3.2. Private sector, in particular SMEs

A myriad of companies of different sizes and expertise participate in the value chains involving the re-use of public sector information:

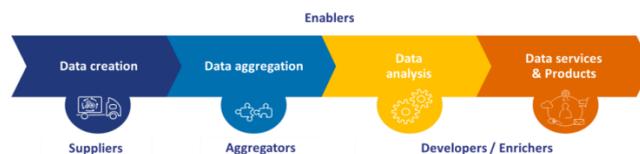


Figure 3 - The Data Value Chain and Data Value Chain Archetypes
Source: EDP Analytical Report 9: The Economic Benefits of Open Data

Companies active in various fields, (e.g. market/research, economic/financial, publishing⁹⁸) typically use open data to improve already existing services, in addition to data derived from other sources. Most can be classified as 'aggregators', with a substantial part using open data as their main input towards developing web- or mobile applications.⁹⁹ Most companies benefiting from open data are SMEs, which represent 99.8% of companies active on the European data market.¹⁰⁰

2.4. How will the problem evolve?

Overall, the effects of technological change and the trend of individual Member States to legislate on the availability of data in new sectors could lead to an increasing gap in re-use conditions in the EU.

It is expected that there will be exponential growth in the volume of real-time data without enough storage capacity for the newly created data, which will strengthen the need to release the data to those who can make use of it.¹⁰¹ Therefore the use of APIs will be increasingly important and speeding up the take-up of APIs is necessary in order to fully reap the benefits of publicly funded data.¹⁰²

The persistence of the current market access barrier linked to the cost of data is likely to lead to a strengthening of the major economic actors active in the open data markets to the detriment of local innovative SMEs and start-ups.

Data in the transport and utilities sector will continue to grow in importance as a source of innovation and new products and services. The implementation of the sectoral EU legislation will improve the availability of data, but the overall potential of opening up data in these sectors for the economy and society may not be fully realised.

⁹⁸ ASEDIE, Infomediary Sector Report, 2017.

⁹⁹ European Data Portal, Analytical Report 9: The Economic Benefits of Open Data, 2017.

¹⁰⁰ IDC European data market monitor 2017.

¹⁰¹ <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>

¹⁰² The French National Register has seen a dramatic increase in data re-use since opening of the public API (1 billion hits in 2017, with eCommerce and supermarkets being the main re-users), see: Impact Assessment Support Study, Deloitte, SMART 2017/0061.

The ongoing efforts to make open access to research information a universally accepted principle across Europe will lead to further results, but it is likely to be a lengthy process and Member States will move at different speeds. Given that researchers themselves are largely in charge of the process of implementing Open Access, the re-usability aspects, such as licensing conditions or technical usability (which are essential for ensuring the impact of research data beyond the research community) may not get sufficient attention.

The trend towards new types of arrangements between the public sector and private companies involving preferential or exclusive use of public data is likely to continue. Large companies will intensify their search for the most cost-efficient sources of high-quality data. At the same time, public sector bodies at all levels of government will look for partners who can help them to exploit their data assets, since the performance of the primary public tasks often makes it difficult to invest in data analytics capacities.

2.5. REFIT considerations

The PSI Directive currently affects two large sets of stakeholders: re-users and public sector bodies, but it only imposes obligations on the latter. When considering these obligations, it should be borne in mind that several requirements of the Directive, notably those related to the practical arrangements for making data available, are part of an overall effort towards digitising the public administration¹⁰³ rather than specific PSI Directive-related costs.

While the Evaluation Report¹⁰⁴ shows that there is an overall positive appreciation of the efficiency of the Directive so far, there are areas in which the reduction of administrative burden strictly related to the implementation of the Directive could be achieved:

Reporting obligations: The current Directive includes a reporting provision (Article 13.2) which obliges Member States to submit, every 3 years, a report to the European Commission on a number of issues that enable the Commission to measure the effects of the implementation of the Directive.

Dealing with individual re-use requests: The necessity to process re-use requests is the main administrative burden that can be attributed solely to the PSI Directive (Chapter 2 – Requests for re-use) and has been estimated to be in the range of 30 hours in terms of required time per single request.¹⁰⁵

Charging for re-use: In case the public sector body applies charges for re-use, it incurs additional administrative burden such as the processing of invoicing, but also costs

¹⁰³ See: Policy action 4 of the Tallinn Declaration, <https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration>

¹⁰⁴ SWD(2018) 145.

¹⁰⁵ See: https://vng.nl/files/vng/20171211_-_notitie_uitvoeringskosten_who_-_def.pdf

Also, the processing of similar requests (access to information) has been estimated in the UK to be in the range of 165€ per single request, <https://www.jisc.ac.uk/rd/projects/freedom-of-information-research>

related to the monitoring of compliance with the underlying license arrangements at the source of the charge.¹⁰⁶

Dealing with complaints: A corollary to the individual re-use requests, and to a lesser degree to the use of custom licensing arrangements, is the higher risk of complaints leading to administrative and judicial proceedings (redress procedures).¹⁰⁷

Clarifying the interplay between the PSI Directive and other EU acts (INSPIRE and Database Directives): The coherence with these directives makes the application of the PSI Directive easier and less-resource consuming, due to higher legal certainty.

At the same time, commercial entities (mostly SMEs) will considerably benefit from the online availability of high-quality data without cost. This will eliminate the need to make individual requests, as well as any transactional costs, thus contributing additionally to the REFIT objectives.

2.6. Access to private sector data for public interest purposes

The issue of access to private sector data for public interest purposes was part of the public online consultation on the review of the PSI Directive. While public data is already used to save costs and increase efficiencies within the public sector itself, modern policy making has become a data-intensive activity to such an extent that public institutions rely more and more on access to private sector data to carry out their tasks. This trend was acknowledged by the respondents to the public online consultation, and there was strong support for action in this area, with 81% of the respondents indicating that specific legal measures need to be put in place in order to facilitate access to private sector data for public interest purposes.

However, the wider stakeholder engagement process¹⁰⁸ showed that there is still considerable uncertainty regarding the exact objectives, justification and practical methods under which such transfer could take place. Stakeholders from the private sector (e.g. telecoms companies) and public undertakings in the transport and utilities sector indicated, notably during a recently held public hearing¹⁰⁹, that the issue is not ripe for horizontal legislative action at the EU level. This is why the subject of access to private data is not included within the policy options discussed in this Impact Assessment and will be addressed separately within the 2018 Data Package.

¹⁰⁶ The reduction of these 'transaction costs' motivates many public institutions to move towards free re-use, Pricing Of Public Sector Information Study, Deloitte, 2012.

¹⁰⁷ As shown in the Evaluation Report, redress and litigation costs are also perceived as one of the main burdens for the re-users, especially SMEs.

¹⁰⁸ As detailed in section 3 of the Evaluation Report.

¹⁰⁹ <https://ec.europa.eu/digital-single-market/en/news/public-hearing-review-directive-reuse-public-sector-information>

3. WHY SHOULD THE EU ACT?

3.1. Legal basis

The legal basis for intervention is Article 114 of the Treaty on the Functioning of the European Union. The revision will build on this legislation and any amendment to this Directive should be based on the same legal basis.

In accordance with Article 5(3) of the Treaty on European Union (TEU) any EU action should respect the principle of subsidiarity. This involves assessing two aspects: the necessity test and the EU added value test.

3.2. Subsidiarity: Necessity of EU action

The removal of the remaining obstacles to an open re-use of public sector information and simultaneously aligning the legal framework to the evolving digital socio-economic environment cannot be achieved by Member States alone. Diverging national legal solutions would likely compromise the growing tendency towards cross-border re-use, whereas the different levels of 'open data readiness' across EU Member States would persist or deepen, having a negative effect on the homogeneity and competitiveness within the Digital Single Market.

The evidence for such a diverging approach to the furthering of open data policies is already emerging, as countries such as Finland¹¹⁰, France¹¹¹ or Denmark¹¹² are enacting legislation which aims to enable the data economy to benefit from new sources of data. These efforts, while in line with the overall objectives of the DSM strategy, are not coordinated. Combined with slower or non-existent initiatives in other countries, they risk undermining the level playing field for commercial re-users, as well as the development of cross-border applications in the EU (insufficient availability of source data across the Member States makes it economically unviable to produce a cross-border service or to replicate an existing data-based service from one Member State to another).

The actions proposed are proportionate, since national intervention will not be able to achieve the same results (increase in openly re-useable PSI), whilst at the same time ensuring a competitive and non-discriminatory environment across the entire Single Market. The proposed actions, in particular restricting the use of exceptions to marginal cost charging and introducing the obligation to make certain high-value datasets freely available, can be seen as the next step towards full availability of PSI for re-use: a policy objective accepted by the Member States already in 2003 and confirmed in 2013.

¹¹⁰ Act on Transport Services, <https://www.lvm.fi/documents/20181/937315/Factsheet+57-2017+Act+on+Transport+Services.pdf/bd002762-a6a0-4867-bb49-5c1b86069380>

¹¹¹ Loi Pour une République numérique, <https://www.legifrance.gouv.fr/affichLoiPubliee.do?idDocument=JORFDOLE000031589829&type=general&legislature=14>

¹¹² Basic Data Programme, https://www.digst.dk/~media/Files/English/Fact_sheet_BasicData_pdf.pdf

The necessity of the Directive is unquestioned by stakeholders consulted in the public online consultation and in the context of the IA support study. National initiatives in the field of PSI would have not been sufficient to address possible market distortions and market failures and offer a level-playing field to all businesses.¹¹³ The necessity of coordinated action on EU level is further backed by the success of previous legislative interventions, as testified by the EDP Open Data Maturity report comparison across the years.¹¹⁴

Moreover, during the consultation process a large number of Member States expressed their support for the review of the Directive, based on the observation that the uneven implementation across Member States is creating fragmentation on the single market or bottlenecks to market development. In particular, they were supportive of extending the scope of the PSI Directive and making some of the high-value datasets obligatory for free dissemination¹¹⁵ (see also chapter 6.4).

3.3. Subsidiarity: Added value of EU action

The vast majority of respondents (87%) in the recent public online consultation, representing different groups of stakeholders, see a clear EU added value of the PSI Directive, namely that it has played a role in encouraging the national authorities to open up more public sector data across the EU. At the same time, 63% believe that it has facilitated access to PSI from countries other than the one where the person concerned lives. Similarly, almost 64% indicate that the PSI Directive has been conducive to the creation of an EU-wide market for products and services based on PSI. The EU added value of the PSI Directive was also strongly highlighted in the expert interviews conducted in the context of the support study and presented in more detail in the accompanying Evaluation Report¹¹⁶.

Action at EU level is also best suited to guarantee that public data of comparable thematic scope are available for re-use across Member State borders under similar legal and technical conditions so as 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.

4. OBJECTIVES: WHAT IS TO BE ACHIEVED?

4.1. General objectives

The overall aim of this initiative is to contribute to the strengthening of the EU's data economy by enhancing the positive effect of the re-use of public sector data on the economy and society. This will be done by increasing the amount of public sector data

¹¹³ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

¹¹⁴ <https://www.europeandataportal.eu/en/highlights/open-data-maturity-europe>

¹¹⁵ Minutes of the PSI Group meeting, 15 November 2017, <https://ec.europa.eu/digital-single-market/en/news/public-sector-information-group-main-page>

¹¹⁶ SWD(2018) 145.

available for re-use, ensuring fair competition and easy access to markets based on public sector information, and enhancing cross-border innovation based on data.

The initiative is part of the DSM Strategy that has emphasised the impact of the data economy on the growth of European businesses, modernisation of public services and the empowerment of citizens.¹¹⁷

4.2. Specific objectives

The general objectives translate into the following specific objectives:

a) Adaptation to technological changes in the field of data management and use

The production of data is no longer an issue, as data generation via computer networks and sensors has become inexpensive. It is rather the ease with which data can be found, accessed and combined with other data that defines its impact on the overall economy. This means that for the data to be searchable, understood and contextualised (especially by machines), the datasets need to conform to minimum quality standards with regard to their format, description and methods of access.

b) Lowering the existing barriers to accessing PSI re-use market and preventing the emergence of new ones

This specific objective aims at eliminating entry barriers for companies, in particular SMEs and start-ups that want to re-use public sector information. In this context, the charging rules are particularly relevant, including the mechanisms to make certain high-value datasets available free of charge, in line with the commitment made by the EU in the framework of the G8 Open Data Charter. Another issue that needs to be tackled is the risk that public sector information is locked in because of new types of agreements between public sector bodies and large companies.

c) Making more data available for re-use as raw material for innovation

Thanks to the PSI Directive, a large array of public sector datasets is available for re-use across Europe. However, data produced and held by public undertakings in utility sectors, as well as publicly funded data resulting from scientific research, is currently not covered by the Directive. Bringing the relevant data within the scope of the Directive, while respecting the particular situation of the respective sectors, will enlarge the overall pool of re-usable data produced either directly thanks to public funding or within the scope of services of general interest for the benefit of the economy and reinforce their cross-border re-use.

¹¹⁷ Mid-term review of the DSM strategy, COM(2017) 228 final.

5. WHAT ARE THE AVAILABLE POLICY OPTIONS?

5.1. What is the baseline for assessing options?

The baseline situation is described in the evaluation report of the PSI Directive¹¹⁸. It shows where Member States stand in terms of open data maturity, after having made considerable progress over the last few years.¹¹⁹ On the supply side of public sector information, the current level of open data maturity¹²⁰ is the baseline for assessing future developments. As for the impact of the policies, the baseline used as a benchmark is the current direct economic effects of open data of 52 billion EUR a year for the EU 28.¹²¹ This figure is higher than some of the other available calculations of the impact of open data policies, but considerably lower than the most optimistic ones.¹²²

5.2. Description of the policy options

(A) Baseline Scenario: Maintaining the current approach without changes

The option of 'no policy change' would mean that the current provisions of the Directive would remain applicable. The Member States would be bound by the rules set by the 2003 Directive, modified by Directive 2013/37/EU.

(B) Discontinuing existing EU action: Repeal of the PSI Directive

This option would effectively result in the removal of all the regulatory obligations currently contained in the Directive.

(C) Soft law measures only

Soft law instruments in the area of Open Data have been used in the past. After the adoption of the amending Directive in 2013 and still within the transposition period, the Commission adopted a Notice 2014/C 240/01 'Guidelines on recommended standard licences, datasets and charging for the re-use of documents'¹²³ for this purpose. Another relevant soft law instrument is the Commission Recommendation to Member States on access to and preservation of scientific information.¹²⁴ These two documents could be updated in order to reflect the technological and policy changes that have occurred since their adoption.

¹¹⁸ SWD(2018) 145.

¹¹⁹ See the 2015 and 2017 figures on open data readiness of the MS in the Evaluation Report, section 5.

¹²⁰ See section 5 of the Evaluation Report.

¹²¹ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

¹²² See the European Data Portal, Analytical Report No 9: The Economic Benefits of Open Data, 2017, giving an overview of the different analyses in this area. The highest estimate of the value of opening up public sector data worldwide came from McKinsey (2013), while other studies converged on figures in the order of 10s of billions of EUR a year for the EU28, depending on the methodology used and the types of data included.

¹²³ http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2014.240.01.0001.01.ENG

¹²⁴ C(2012) 4890 final.

(D) Packaged solution consisting of both amendments of the PSI Directive and soft law (update of the Recommendation on access to and preservation of scientific information)

For the different issues outlined in the problem description, two scenarios of varying degrees of intervention were identified. Both scenarios contain hard and soft-law elements. Despite the fact that it would be conceivable to address the policy objectives by a mix of actions scattered across the two options, it was decided to regroup the possible intervention choices into two packages along the lines of high and low intensity of regulatory intervention. Given the complexity of the problem areas, such regrouping of individual intervention areas was crucial for obtaining meaningful feedback from the stakeholders. Policy option (D) is therefore split into two policy packages having similar objectives but leading to a different level of impact in terms of costs, benefits and administrative burden. The main elements of the low and high-intensity scenarios are described below in relation to each of the intervention areas targeted. They are further mapped onto concrete legislative amendments to the text of the PSI Directive, included in the table on page 31.

Dynamic data/APIs

- The higher intensity option would oblige all public sector bodies of the Member States that already produce dynamic data, to make this data available in a timely manner and to systematically introduce APIs for that purpose.
- The lower intensity option would create an obligation for Member States to make dynamic data available in a timely manner and to introduce APIs, where possible under technical and financial circumstances. For a limited number of datasets (high value data to be defined in the Delegated Act) there would be a hard obligation to do so.

The options focus on the ways in which the availability and usability of dynamic data can be enhanced, rather than addressing the broader issue of optimal data formats and publication methods, even if respondents to the online consultation tend to view all these factors as one common 'technical barrier' to re-use. This is chiefly because the insufficient release of dynamic content is a problem in its own right (given massively rising demand for this kind of data) and because the Directive already includes provisions specific to formats (Article 5), while being silent on dynamic data and use of APIs. In addition, the Directive does not have the ambition of harmonising data or metadata standards; instead, the Commission promotes their adoption via the ongoing support actions (European Data Portal, ISA² programme) and soft law (Notice 2014/C 240/01).

Charging rules

- Under both options, the charging rules would explicitly mention the costs linked to anonymization of datasets containing personal data as costs that can be recovered by a public sector body both under the marginal costs and full cost recovery principles. This acknowledges the fact that in many re-use scenarios, datasets containing personal data need to be anonymised and that such anonymization has a cost.

- The higher intensity option would eliminate all the exceptions to the general rule that public sector bodies cannot charge more than marginal costs for dissemination.
- The lower intensity option would tighten the charging rules by eliminating the current exception of allowing charges for certain specific documents (deemed unnecessary as it does not address a wide range of real-life situations) and by enhancing transparency with a requirement for the Member States to publish a list of public sector bodies that fulfil the requirements of article 6(2)(a) and can therefore charge above marginal costs. This will only concern a small number of public sector bodies and therefore creating the list will not put a large burden on Member States. These measures would be accompanied by the adoption of a limited list of fundamental high-value datasets that should be freely available in all Member States (to be adopted through a Delegated Act).¹²⁵

Both options are realistic by taking into account the current trend of introducing a 'principle of gratuity' in national legislation (e.g. France, Slovenia), coupled with the fact that charges are typically applied not evenly across the board but rather by a small number (fewer than 10) of public institutions in each Member State.¹²⁶ An alternative option whereby charges from high-value datasets would have to be used to finance the opening up of new datasets in other areas was considered but not retained. While it could lead to a stronger data supply overall, it would risk keeping the most valuable datasets behind a paywall, and monitoring the system would be costly.

Data in the transport and utilities sector

- The higher intensity option would cover public undertakings and also private entities operating on the basis of concessions in the transport and utilities sector. The full set of rules of the Directive would apply.
- The lower intensity option would cover only public undertakings. A limited set of obligations would apply: Public undertakings could continue to charge above marginal costs for dissemination and would be under no obligation to release the data they do not want to release. 2003 rules of the PSI Directive would apply (as explained in the table on p. 31). Moreover, public undertakings would not be covered by the requirements applicable to the processing of requests for the re-use of their data.

The choice to limit the extension of scope only to public undertakings covered by Directive 2014/25/EU was motivated by the need to ensure legal certainty (reference to established legal definitions with related case law) and the fact that the services of general interest in the economic sectors referred to in the said Directive are typically carried out by public undertakings rather than solely by public sector bodies.

¹²⁵ The option of a list of freely re-usable high-value datasets is based on the existing EU commitment to enable open re-use of 14 categories of key datasets, expressed in the G8 Open Data Charter. NB.: the Delegated Act would be subject to a separate Impact Assessment and consultation process.

¹²⁶ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

Research data

- The higher intensity option would introduce in the PSI Directive a complete set of provisions to be transposed in binding national legislation mandating all researchers in the EU to make all research results – publications and research data – available in Open Access (a top-down European legislative Open Access mandate). It would define the legal re-usability of such data within the scientific community and beyond. It would thus harmonise Open Access policies that are currently defined either at national level or at the level of individual research funders. It would in particular define a common European framework on all aspects of Open Access, including embargo periods for publications, opt-outs from Open Access obligations and sanctions for researchers not complying with the Open Access mandate. The Commission Recommendation on access to and preservation of scientific information from 2012 would address accompanying measures, e.g. incentives to researchers as part of policies and legislation on scientific career, the development of relevant infrastructures and preservation.
- The lower intensity option would limit the hard law intervention through a revised PSI Directive to only cover research data (and not scientific publications) that have been made publicly available as a result of Open Access funder mandates and to ensure the legal re-usability of such research data within the scientific community and beyond. This would address the issue of quite heterogeneous licensing practices for research data, as reported also by stakeholders. It would still oblige Member States to develop policies for Open Access to research data resulting from publicly funded research, but leave flexibility on how they do this. An updated Recommendation to Member States on access to and preservation of scientific information would guide Member States on the elements ideally contained in an Open Access policy.

Non-exclusivity

- The higher intensity option would prohibit arrangements that lead to the lock-in of public sector information.
- The lower intensity option would set transparency requirements for public-private agreements involving public sector information (public and transparent scrutiny of the process leading to the conclusion of the contract and publication of the actual agreement).

In all scenarios the legislative update would be an occasion to clarify the relation between the PSI Directive on the one hand, and the Database and the INSPIRE Directives on the other hand. The intervention would clarify that the *sui generis* right included in the Database Directive cannot be exercised by public sector bodies to prevent or restrict the re-use of documents under the PSI Directive, and that the provisions of the PSI Directive, including on charging, fully apply to INSPIRE datasets.

The identified options are combined into **two main packages**, one with all elements of lower legislative intensity, and the other one with all elements of higher legislative

intensity, for the reasons described on p. 27 above. The table below gives an overview of the two packages and indicates how they would be implemented in the Directive.

In order to balance the rigid dichotomy of a 'soft' and 'hard' intervention logic and to present a more realistic choice to the stakeholders, the two policy options of different levels of intensity are accompanied by a soft-law measure in the form of an updated Recommendation on access to and preservation of scientific information and by a separate Delegated Act with a list of high value datasets. The Delegated Act (part of the lower intensity option) can be seen as a bridge between the two main policy options, since it applies the solutions assessed under the 'higher intensity' scenario, albeit to a limited set of reference data.

The 'intermediate' intervention action consisting of a Delegated Act with a list of high value datasets reflects the fact that the range of intervention measures cannot be easily split into two separate sets of actions along the intensity factor. At the same time, it is included as a part of the 'low intensity' policy option rather than a separate policy option in order to facilitate stakeholder feedback (and related cost/benefit assessment) and in the light of the requirement that the adoption of the Delegated Act should be preceded by a separate future Impact Assessment (i.e. an inclusion of a separate intermediate policy option at this stage would be premature).

**Policy options of lower and higher regulatory intervention intensity –
implementation modalities in the PSI Directive**

	Lower intensity	Higher intensity
Dynamic data	<p>Binding measures: Adding a reference in the Directive to incentivise public sector bodies to make dynamic data available for re-use immediately after collection (or at least in a timeframe that does not unduly impair the exploitation of their economic potential), and to use APIs.</p> <p>Binding measures (intermediate intensity): There would be an obligation to use APIs for the availability of high-value datasets (see below).</p>	<p>Binding measures: Strengthening the current rules on data formats (Article 5) by adding an obligation for public sector bodies that already produce dynamic data to make all such data available for re-use immediately after collection, and to systematically use APIs.</p>
Charging	<p>Binding measures: Stipulating that the default rule is zero charges or charges limited to the marginal costs. Limiting the range of situations under which exceptions to marginal cost charging would be possible while specifying (in line with Notice 2014/C 240/01) the eligibility of costs covered. Publishing the list of public sector bodies that could apply charges above marginal costs.</p> <p>Binding measures (intermediate intensity): Defining high value datasets to be released at zero charge across the EU via a Delegated Act.</p>	<p>Binding measures: While not introducing the principle of full gratuity, eliminating the current exceptions to the rule that charges can cover at maximum the marginal costs of dissemination.</p>
Scope (research data)	<p>Binding measures: Extending the scope to publicly funded research data: - Member States would be obliged to develop</p>	<p>Binding measures: Extending the scope to publicly funded research data by laying down detailed rules on Open Access (including options on how to comply,</p>

	<p>policies for Open Access to research data resulting from publicly funded research, while keeping flexibility on the specific details.</p> <p>- For research data that already are accessible through repositories, the Directive would ensure re-usability within the scientific community and beyond.</p> <p>Soft law: Update of the Recommendation on Scientific Information</p>	<p>embargo periods, rules on opt-outs from Open Access obligations, enforcement).</p> <p>Member States would be obliged to ensure that all publicly funded research data is available in open access and fully re-usable.</p> <p>Soft law: Update of the Recommendation on Scientific Information</p>
<p>Scope (data in the transport and utilities sectors):</p>	<p>Binding measures: Extending the scope to:</p> <ul style="list-style-type: none"> - public undertakings in the utilities sectors covered by Public Procurement Directive 2014/25/EU¹²⁷ and public undertakings acting as public service operators under Regulation (EC) No 1370/2007¹²⁸. <p>The default rules of the original PSI Directive (2003 rules) would apply: where data is made available for re-use, transparency, non-discrimination, maximum charges based on full-cost and reasonable return on investment, and non-exclusivity requirements would be required. Public undertakings would not be covered by the requirements applicable to the processing of requests for the re-use of their data.</p> <p>There would be a reference to data that should be openly available, in line with relevant sectoral EU legislation, to ensure that existing obligations are respected¹²⁹.</p>	<p>Binding measures: Extending the scope to:</p> <ul style="list-style-type: none"> - public undertakings in the utilities sectors covered by Public Procurement Directive 2014/25/EU and to private and public undertakings acting as public service operators under Regulation (EC) No 1370/2007; - private economic operators which have been awarded a concession, as defined in Directive 2014/23/EU¹³⁰, pursuing one of the activities in the transport and utilities sectors. <p>The default rules of the PSI Directive would apply (marginal cost charging, transparency, data formats, processing of requests, etc.).</p> <p>There would be a reference to data that should be openly available, in line with relevant sectoral EU legislation, to ensure that existing obligations are respected.</p>
<p>Non-exclusivity:</p>	<p>Binding measures: Strengthen the current article on exclusive arrangements in the Directive by introducing a procedural safeguard. This would require making public (ex ante, i.e. before taking effect) the terms of arrangements with private sector players involving the re-use of public sector data where there is a risk of the lock-in of public sector data. In addition, the final text of the agreement should also be publicly available.</p>	<p>Binding measures: Strengthen the current article on exclusive arrangements to prohibit the conclusion of agreements between the public sector bodies and private companies that may lead to the lock-in of public sector data.</p>

¹²⁷ Directive 2014/25/EU on procurement by entities operating in the water, energy, transport and postal services sectors.

¹²⁸ Regulation (EC) No 1370/2007 on public passenger transport services by rail and by road.

¹²⁹ Such as Commission Delegated Regulation (EU) 2017/1926 of 31 May 2017 in the area of transport data and other EU acts in the areas of water or energy provision mentioned in section 2.2.4.1.

¹³⁰ Directive 2014/23/EU on the award of concession contracts.

5.3. Options discarded at an early stage

1) Discontinuing existing EU action

This option would remove the safety net provided at the EU level by the established minimum PSI re-use rules. By leaving Member States free to act in the area previously subject to harmonised EU rules, it would give rise to increased legal uncertainty and divergence of national approaches, to the detriment of competition in the internal market for re-use of PSI, and to the detriment of the functioning of the Digital Single Market.

2) Soft law measures only

While the use of this option can be a non-intrusive way of addressing the problems identified, this option was discarded based on the experience gained with the two existing soft law measures. Although they have fulfilled a useful function in providing clarity and giving a general direction to Member States, due to their non-binding character they have been taken up with different intensity in the different countries. Also, they cannot be relied upon by re-users who are looking for legal certainty. Finally, soft law measures cannot fully prevent the development of regulatory divergences among Member States.

6. WHAT ARE THE IMPACTS OF THE POLICY OPTIONS?

6.1. Economic and data-specific impacts of the retained options

Based on the available evidence and the extensive earlier analytical work in this area, the Impact Assessment study calculated a baseline of the total direct economic value of public sector information for the EU28 of 52 billion EUR in 2018. The direct economic value is expected to increase until 2030 to 149 billion EUR (+185%).

This development is expected to be triggered through different supply and demand side factors. On the supply side, an increasing number of public sector bodies is expected to open up their data due to the Directive, in addition to the increase in the re-use and value of data that is already available for re-use. On the demand side, it is expected that an increasing number of stakeholders will take advantage of more ample PSI in re-usable formats and of higher availability of APIs.

As a consequence, the amount of data from various sources is expected to be higher in the future than today, while – simultaneously – its quality is expected to increase as well. This is expected to lead to an increased use of PSI by businesses, public authorities, as well as end users (through apps). Moreover, a decline of operational costs to acquire and process PSI is expected to enable businesses to develop more and better PSI-based services. This is expected to result in an increased economic value over time as the user base of PSI increases.

The graph below displays the expected development of direct economic value of PSI in the lower intensity and the higher intensity scenario, compared to the baseline scenario. The first scenario (policy option 1) is based on a combination of all the lower intensity regulatory elements described above. The second scenario (policy option 2) is based on a

combination of all the higher intensity regulatory elements. The estimates have been modelled as an S-curve. With specific regard to PSI, this means that at some point in the future (+/- year 2028) the economic value of PSI is expected to grow at a shrinking margin. Reasons for this can be, for example, that the most important data sources have been opened up and provided in a re-usable format. Moreover, in the saturation phase of the market, most consumers are expected to have purchased PSI-based goods and services. Therefore, the growth curve typically “flattens out” in the saturation phase of the market.

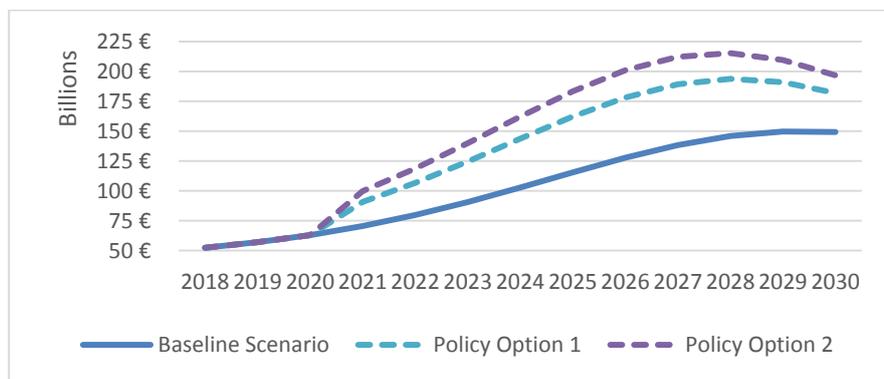


Figure 4 - Impact of the different Policy Options on the direct economic value of PSI (EU28, 2018-2030)
 Source: Impact Assessment Support Study, Deloitte, 2018

In 2028, i.e. the year in which the relative differences in economic value are expected to be largest between the policy options, the value could increase to:

- EUR 194 billion if the lower intensity regulatory intervention was introduced (+ EUR 142 billion, 273%); and
- EUR 215 billion if the higher intensity regulatory intervention was introduced (+ EUR 163 billion, 313%).

Both scenarios are considerably better than the baseline scenario, and in terms of impact the higher intensity regulatory intervention trumps the lower intensity scenario.

In terms of job creation, today 64,000 persons are employed based on PSI re-use. In 2027, i.e. the year in which the relative differences in the number of persons employed in the (data-driven) economy are expected to be largest between the policy options, the value could increase to 709,000 persons based on the lower intensity regulatory intervention and 795,000 based on the higher intensity regulatory intervention. This is considerably higher than the baseline scenario (518,000 persons employed).¹³¹ This figure may seem very high, but the example of Transport for London mentioned in section 2.2.4. (730 extra jobs created based on TfL data) indicates the potential for innovation and economic activity powered by open data.

As for the cost of making PSI available for re-use, under the baseline scenario it will be in the region of 3.8 billion EUR by 2030. The lower intensity intervention is expected to reduce it to almost 3 billion EUR, and the higher intensity intervention could lead to an

¹³¹ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

even greater reduction to 2.5 billion EUR (approx. 21% and 34% lower than the baseline, respectively).¹³²

Three main elements were taken into account in the Impact Assessment study to estimate the costs and their development:

- the number of affected stakeholders and their expected development (i.e. contracting authorities, cultural institutions, entities in the transport and utilities sectors, and research establishments);
- the expected average cost per year to open PSI per stakeholder¹³³; and
- the projected decrease of the annual costs based on the growth model applied for this study (see also Annex 8).

The costs and their development in the baseline scenario are subject to two variables: the increasing number of public authorities and businesses affected by / using PSI, as well as the decline in costs based on theory and assumptions used in the growth model (see also Annex 8).

Similar to the increase of the economic value of PSI, the decrease in costs to acquire and process PSI is driven by both its increased availability and use: through additional public sector bodies opening up their data, businesses are expected to be able to develop more and better PSI-based services. These services correspond to an increased need by end-users for such services which, in turn, is expected to translate into increased turnover for businesses, and better performance of public sector bodies.¹³⁴ This does not only improve stakeholders experience with PSI but also enables investments in the quality and re-usability of PSI by public sector bodies. Thus, data provision and service development are expected to become more efficient. This is expected to translate into lowered costs for both the provision of re-usable PSI, and its acquisition and processing by businesses.

6.1.1. Baseline scenario

Chapter 8.2.1. of the Impact Assessment study includes a detailed description of the trends underlying the overall estimated impact of the baseline scenario (treated for the purpose of the study as a separate policy option). For instance, apart from the beneficial effects of the legislative framework already in place, the number of companies using data is expected to steadily increase by 182% until 2030 while the share of employment in knowledge-intensive sectors is expected to increase until 2030 to about 52% of EU's total labour force. The baseline scenario does therefore present a growth curve, starting with the current total economic value of PSI of 236 billion Euro. It is expected that this value could increase to 672 billion Euro in 2030, or to EUR 150 billion in terms of direct

¹³² Ibid.

¹³³ Based on interviews, secondary resources, feedback received during the "Open Research Data" workshop. The estimate included costs concerning hard- and software (IT equipment, programs and licences, computers and servers), as well as related services (legal and IT services, audits and tests, APIs). In addition, the estimate contained salaries and training costs.

¹³⁴ For example, more public services can be provided with better quality and in a faster manner.

economic impact (assuming a ratio of 3.5 – 3.8 between direct and indirect value, based on earlier studies).

The following sections of the report, in line with the Impact Assessment study, assess the impact of the two policy options defined in chapter 5.2. with reference to the trends and figures presented for the baseline scenario in the current section.

Despite the overall positive trend, the lack of policy reaction to the identified problems would not allow the expected benefits of open data to fully develop. For example, a mismatch between the provision of the kind of data sought after by commercial re-users (real-time data streamed via APIs, data relevant for utility sectors such as transport and energy) and the kind of data generally offered by public sector bodies can lead to the waning of interest in public datasets on the part of re-users, especially if the methods used to allow access are outdated.¹³⁵

While the ongoing efforts to open up data on the sectoral level (mentioned in chapter 2.2.4.1) would contribute to the overall growth of data re-use, and the associated expected benefits¹³⁶, the disparities in data availability and the risk of market fragmentation across the EU would not be fully addressed, e.g. in the area of charging the *status quo* would solidify current practices with regard to high-value datasets. Larger companies with sufficient means to acquire data would fortify their market position while the smaller commercial re-users would be forced to try to innovate on other, less attractive datasets. Similar considerations apply to the exclusivity issue.

6.1.2. Policy Option 1 - Packaged solution: lower intensity of regulatory intervention

This option combines all the elements, both binding and non-binding, with the lower overall intensity of regulatory intervention.

This option is likely to **increase the availability of dynamic data** and the use of modern dissemination tools, while taking into account the technological maturity level of the public sector body in question. A soft obligation (not applicable across the board) would have an impact on the bodies holding high value data while minimising costs for smaller entities. It would thus be a proportionate and efficient way of addressing this problem area. An obligation can, however, be introduced for a limited number of datasets.

Focus on APIs

The main cost-creating element of both policy packages is the implementation and maintenance of APIs. Costs involved depend on many different factors including existing data infrastructures, operational costs due to the scale of data re-use (number of API requests), as well as the data itself (number of records, size, frequency of update, etc.).

Based on current software as a service (SAAS) offers it is possible to establish a range of costs depending on the size of the dataset and the number of API requests. It is very common for a

¹³⁵ See: Finnish open data re-users' opinions in 'Working together towards open data business', https://www.databusiness.fi/content/uploads/2017/10/20171106_WorkingTogetherTowards_WEB.pdf.

¹³⁶ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

high-value key dataset to reach several million of API calls per year, and in some cases several billion. Therefore, the price range from 2,400 EUR per year for up to 7.3 million API calls to 150,000 EUR per year for up to 1.2 billion API calls seems a good fit to estimate the cost of API for key datasets in EU.

In the G8 Open Data Charter 14 categories of high-value datasets were identified which should be opened in high quality and through API access. **A requirement to make these datasets available through APIs would imply a cost ranging from a minimum of 30,000 EUR to a maximum of 2,000,000 EUR per year for each Member State.**¹³⁷

The costs of delivering data through APIs will decrease considerably in the coming years: as the market will adapt to the demand, more and more IT systems and open data platforms will provide embedded API features by default (it is already common). Moreover, the advance of software development will make it easier and cheaper to implement an API on top of any data source. Finally, the cost of delivering data through an API is related to the cost of data infrastructures. According to the Moore's Law, those costs will continue to decrease in the next years.

As for **charging**, the principle of marginal costs for dissemination was established in 2013 based on the evidence presented at the time (Impact Assessment). All the arguments still hold, as verified in practice: in France, abandoning the cost-recovery model for the data held by the National Geographic Institute (NGI) led to an increase in the volume of data downloaded by 20 times and generated around 114 million EUR of benefits for the public, against a costs for the NGI of around 6 million EUR¹³⁸. It is clear that the direct and indirect benefits of free-of-charge re-use can be high, for all classes of stakeholders. First, a substantial lowering of charges leads to a corresponding important increase in the demand for data.¹³⁹ Secondly, low or zero-cost charges lead to more economic activity: evidence correlates cheap data with 15% faster growth.¹⁴⁰ And more economic activity leads to more tax income.¹⁴¹ It is also clear that the cost-recovery model of charging (i.e. above marginal costs) creates a market barrier¹⁴². Furthermore, prices impose unnecessary transaction costs on users.¹⁴³

Finally, free re-use is also shown to lead to efficiency gains for the public sector.¹⁴⁴ In Denmark 30% of the benefits of opening address data accrued to the public sector¹⁴⁵, while efficiency gains thanks to open geospatial data reached 22 million DK over 4

¹³⁷ This assumes that instead of using open source solutions (e.g. CKAN), MS would opt for 'open data as a service' platforms, which include additional functionalities, such as data processing and storage, making them more expensive to maintain. Open-source solutions would be less costly.

¹³⁸ Impact Assessment Support Study, Deloitte, SMART 2017/0061.

¹³⁹ See Chapter 2.2.3 on the macroeconomic considerations in favour of a zero-charge policy.

¹⁴⁰ Does Marginal Cost Pricing of Public Sector Information Spur Firms Growth ?, Heli Koski, 2011.

¹⁴¹ See Impact Assessment 2011 PSI Review.

¹⁴² Trojette, M.A. (2013), 'Rapport au Premier Ministre. Ouverture des données publiques. Les Exceptions au principe de gratuité sont-elles toutes légitimes?'

¹⁴³ Permission granted: The economic value of data assets under alternative policy regimes, A Lateral Economics report for the Open Data Institute, 2016, p.14.

¹⁴⁴ De Vries, M., Kapff, L., Negreiro Achiaga, M., Wauters, P., Osimo, D., Foley, P., Szkuta, K., O'Connor, J., and Whitehouse, D. "POPSIS – Pricing of Public Sector Information Study." (2011).

¹⁴⁵ DECA, 2010. "The value of the Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge".

years.¹⁴⁶ A shift from a cost-recovery to an open-access regime is likely to more than double the value of the re-use of the data.¹⁴⁷

The lower intensity regulatory intervention should strike a good balance between the needs of the public sector on the one hand and the impact on the data economy on the other: it would allow entities across the EU to continue charging for the re-use of PSI when really essential (e.g. where the continuation of producing the dataset depends on the income generated), while enhancing transparency and legal certainty for the re-users (by bringing the current exceptions down to one single rule and by clarifying which bodies are allowed to apply charges, as is already the case e.g. in France). This incremental change is not expected to create administrative burden or raise costs for the vast majority of public sector bodies (as they do not apply charges). This option is at the same time preferred by a majority of all respondents to the public online consultation. To attenuate the negative effect of maintaining the exceptions, a Delegated Act would identify a set of high-value datasets for free (or marginal cost) re-use, thus acting as an 'intermediate' policy option, by applying higher level requirements to a limited set of data, from a limited set of public bodies. The baseline for the Delegated Act would be constituted by the dataset categories listed in the G8 Open Data Charter¹⁴⁸. However, the final selection would be drawn up by the Commission after consulting the experts and the concerned stakeholders, and in line with the findings of a separate Impact Assessment.

With regard to **scientific information and research data**, the lower intensity option would leave it to the Member States to define the specificities of the Open Access obligations (scope, level of obligations, embargos, rules on opt-outs from Open Access obligations, enforcement), in line with the updated guidance included in the Recommendation on access to and preservation of scientific information¹⁴⁹, and extend hard law obligations only to research data (and not scientific publications) and only to such research data that have been made publicly available as a result of Open Access funder mandates through existing digital repositories.¹⁵⁰ It would be limited to ensuring the legal re-usability of such research data within the scientific community and beyond. It would limit the capacity of researchers to define restrictive terms of usage. This is in line with current Open Access mandates that, however, are implemented at the level of individual grant agreements and are thus only enforceable between the parties to the agreement. Potential re-users currently have no specific rights to re-use such data. On the other hand, it would leave flexibility to Member States and funding bodies in Member States to define the exact terms and the enforcement mechanisms, leaving room for heterogeneous funding models of research establishments and individual research

¹⁴⁶ <http://sdfe.dk/media/2917052/20170317-the-impact-of-the-open-geographical-data-management-summary-version-13-pwc-qrvkvdr.pdf>

¹⁴⁷ "Permission granted: The economic value of data assets under alternative policy regimes", A Lateral Economics report for the Open Data Institute, March 2016, p. 22.

¹⁴⁸ Full list of data categories: <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

¹⁴⁹ C(2018) 2375.

¹⁵⁰ According to a landscaping study of research data repositories, 96.6% are operated on a not-for-profit basis, cf. Kindling et al., The Landscape of Research Data Repositories in 2015: A re3data Analysis.

actions. This would be fully coherent with the updated Recommendation on access to and preservation of scientific information¹⁵¹, forming a part of the broader 2018 Data Package. It would also ensure that data which is currently accessible (but in fact rarely re-used) would not only be subject to common and harmonised rules describing the conditions under which data can be made available (including costs), but also practically searchable and discoverable, as is the case with other public datasets. Finally, the low-intensity solution would avoid any burden on research establishments or even individual researchers that otherwise would be confronted with request for access to research data that is not (yet) openly accessible through a repository. Such requests – even if to be rejected – could constitute an unnecessary administrative burden.

With regard to **data held by public undertakings in the transport and utility sectors**, the lower intensity option would primarily focus on ensuring a fair market for downstream re-use. It is expected that the economic impact of this option would be relatively limited from the point of view of data supply. It would however, lead to substantially lower compliance costs, because public undertakings would be free to decide if they want to open up their data or not. In this way it would also minimise the effect of imbalance (in terms of openness requirements) between the private companies and public undertakings in transport and utility domains active in the same markets. This solution would also be coherent with the interlinking EU legislation imposing data access requirements on the sectoral level, ensuring synergies rather than potentially conflicting norms.

With regard to **'data lock-in' effect**, the low intensity option would further discourage agreements with negative consequences for market access by ensuring transparency and scrutiny of agreements both by potential competitors and citizens (allowing them to challenge it under national redress procedures). It would also maintain a fair level of competition for public datasets, thereby increasing the probability of the most innovative re-use with the highest socio-economic benefit. At the same time, the option would not deter private entities from seeking to conclude agreements with public sector bodies. It would nevertheless increase both parties' awareness as to the potential negative impact on the market and encourage them to put in place suitable safeguards (e.g. introducing collaboration mechanisms that minimise the risk of deals having a 'data lock-in' effect).

The figure below from the Impact Assessment study indicates how the different elements of this option contribute to its impact compared to the baseline option.

¹⁵¹ C(2018) 2375.

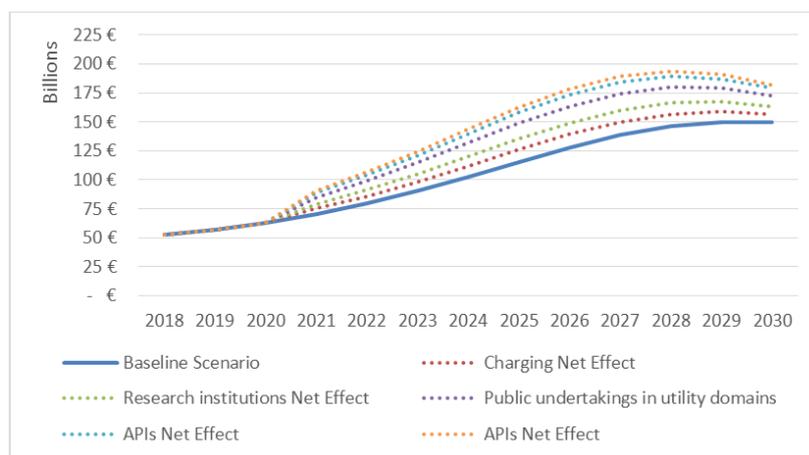


Figure 5 - Impacts of the different elements of the lower intensity regulatory intervention Policy Option on the direct economic value of PSI (EU28, 2018-2030)

Source: Impact Assessment Support Study, Deloitte, 2018

6.1.3. *Policy Option 2 – Packaged solution: higher intensity regulatory intervention*

Compulsory use of APIs and wide **release of dynamic data** would enhance the creation of commercial, value-added services based on dynamic data. Easier data access would also improve the provision of public services (by enhancing intra-government data flows, as is the case in Estonia and France). However, a strict obligation to use APIs for all public databases, applying to all levels of government and without regard to the actual technical readiness of the administration would place a large financial burden on the public sector. This could be particularly problematic for smaller public sector organisations that do not have IT expertise in-house, since they would need to outsource the implementation and maintenance of APIs in their systems, generating additional costs in the process.

As for **charging**, the immediate abolition of the possibility to generate income from data assets would have a considerable impact on institutions that cover their operating costs via the sale of PSI. Based on the estimated volume of income from PSI re-use generated from providing commercial access to geospatial reference data for 28 EU MS being in the order of 120 million EUR, and assuming that geospatial sector encompasses the largest single set of public bodies that charge for the data (next to business registers¹⁵² and meteorological institutes), we could expect the total revenue loss caused by moving to a zero-charge regime to be in the order of approximately 300 million EUR annually across the EU.¹⁵³

The overall impact across Europe would be mitigated by a) the fact that the PSI re-use revenues of public sector bodies are often small when compared to the total budget of the

¹⁵² It should be noted that under The Company Law Directive 2017/1132, the charging regime for business registers is specific. The forthcoming Company Law Package will widen the range of data to be provided free of charge.

¹⁵³ This order of magnitude is supported by national estimates, see: minutes of the last PSI Group, <https://ec.europa.eu/digital-single-market/en/news/public-sector-information-group-main-page>

public sector body concerned¹⁵⁴, b) few public sector bodies in each country (outside the cultural sector) actually charge for the re-use of data¹⁵⁵ and c) cheaper data leads to higher economic activity and indirectly – higher budget revenues in the long run.¹⁵⁶

Nevertheless, this option presents certain risks. While the evidence¹⁵⁷ demonstrates that the benefits of opening up public sector information largely outweigh the costs and the foregone revenues of individual public sector bodies, the macroeconomic benefits (including higher government revenue) appear on an aggregate budgetary level and therefore only kick in later, requiring transitory financing arrangements, e.g. in order to keep the quality of the datasets at a high level.¹⁵⁸ Public sector bodies charging above the marginal costs of dissemination are generally afraid that the economic and societal benefits as a whole will not be compensated by the Treasury in their individual budgets. Therefore, overall the higher intensity regulatory intervention as regards charging does not strike a good balance between the needs of the public sector on the one hand and the impact on the data economy on the other.

With regard to **scientific information**, the higher intensity option would, along the lines included in the updated Recommendation on access to and preservation of scientific information¹⁵⁹, harmonise Open Access policies that are currently defined either at national level or at the level of individual research funders for the entire EU. This would have positive effects within the scientific community (a level playing field as all researchers are subject to identical obligations in terms of Open Access to their research results, higher transparency of results leading to less plagiarism, more efficient peer-review and reduction of costs for data acquisition).

The Impact Assessment study lists many of examples illustrating that Open Access to research data leads to faster discovery and increased innovation. For instance, in the context of the Human Genome Project, open access to genome sequencing data led to a whole range of products and services and ultimately lowered the price of DNA sequencing from 82 million EUR to 4,074 EUR between 2001 and 2013.¹⁶⁰

While this option would be coherent with the European Commission's current policy on Open Science (as reflected in Horizon 2020 rules) and while stakeholders and Member States strongly support Open Access, the move to hard law in this area should be addressed in a gradual manner over a longer span of time. This would allow the research community and other research stakeholders to voice their opinions and to treat the question in process that includes bottom-up elements as far as possible.

¹⁵⁴ Ibid.

¹⁵⁵ For example, fewer than 10 public sector bodies charge in the UK, and only 3 in France.

¹⁵⁶ http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf, Chapter 2.

¹⁵⁷ <https://ec.europa.eu/digital-single-market/en/news/pricing-public-sector-information-study-popsis-models-supply-and-charging-public-sector>

¹⁵⁸ Marc de Vries, ABOUT GMES AND DATA : GEESE AND GOLDEN EGGS, 2012, p. 35.

¹⁵⁹ C(2018) 2375.

¹⁶⁰ E. Hayden, Technology: The \$1,000 genome". Nature 507, 294–295 (20 March 2014).

A substantial initial influx of new **data held by both public undertakings and private operators from the transport and utilities sector** could also be expected. However, obligations to allow re-use (Article 3), coupled with the obligation to apply the principle of charging marginal cost for dissemination (Article 6) would bring a noticeable burden on entities active in the transport and utility domains. In addition, an obligation to open up data could be perceived to have a negative impact on investment in sensors and data collection by entities active in the transport and utility domains, out of fear of strengthening competitors who would not be subject to such an obligation. This could affect the capacities of entities in the transport and utility domains to innovate and has often been brought as an argument against a strong horizontal intervention on EU level in this area.¹⁶¹

Finally, a full-fledged **prohibition of arrangements leading to 'data lock-in'** would discourage private companies from approaching public institutions in search of preferential treatment in terms of data access and use, with a positive impact on competition and market access by start-ups and SMEs. However, an outright ban on arrangements with a risk of a 'data lock-in' effect might also have a chilling effect on the private sector, preventing it from seeking any sort of public-private partnership for fear of breaching legislation, with an overall negative impact on data-based innovation.

6.2. Impact on SMEs

The positive effect of widely available, free of charge public data is first and foremost important for start-ups and SMEs. While open government data is also a source of innovation for large market players, including the online platforms, it is a key ingredient for many European start-ups and SMEs.¹⁶² 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.¹⁶³ Observations from other markets corroborate the importance of open data for SMEs.¹⁶⁴

In many cases SMEs struggle to get hold of the relevant data, because of lack of financial resources and because they do not have the power to negotiate with dataholders. Empirical findings show that the PSI pricing matters for firms' growth particularly from the perspective of SMEs.¹⁶⁵ Also, replacing the current system based on requesting data by proactively making data available through APIs will benefit SMEs most, since they often do not have the capacity to deal with the administrative requirements and licencing

¹⁶¹ See Annex 2.

¹⁶² Examples of SMEs and start-ups that have successfully launched businesses based on open data can be found in the evaluation report and on the European Open Data Portal.

¹⁶³ D6.3 Business models, lessons learned and success stories, ODINE, p. 40.

¹⁶⁴ Open Data – a 21st Century resource for Small and Medium Enterprises, <http://images.thegovlab.org/wordpress/wp-content/uploads/2015/08/OpenData-and-SME-Final-Aug2015.pdf>

¹⁶⁵ See for example: « Does Marginal Cost Pricing of Public Sector Information Spur Firms Growth ? », Heli Koski, The Research Institute of the Finnish Economy, 2011.

issues. Moreover, many transport and energy apps are produced by SMEs, who stand to gain most from better availability of data in these domains.

Currently, SMEs represent 99.8% of companies active on the European data-market, given the relatively modest initial capital investment necessary to launch a data-driven business. Therefore it is likely that the estimated job growth (645,000 new jobs by 2027) based on public sector information will for the large part come from employment created by SMEs.

6.3. Social and environmental impacts

Open government data brings about a whole range of societal and environmental benefits¹⁶⁶, experienced mainly by the end users, i.e. consumers of products and services based on open data. For example, it is estimated that open data has the potential to save 7,000 lives yearly thanks to quicker response in case of cardiac arrest and 1,425 lives from traffic losses (i.e. 5.5% of the European road fatalities).¹⁶⁷

Apart from having the potential to considerably increase the efficiency of government through better policy making¹⁶⁸, open data helps in reducing the distance between government and citizens as regards access to information¹⁶⁹ and generates significant benefits in the areas of social inclusion and civic participation.¹⁷⁰

6.3.1. Baseline scenario

In case of the baseline scenario, societal and environmental benefits will be continue to be realised to a certain extent, given the implementation of open data policies on EU and national level in the last decade. In fact, many countries – both in Europe and beyond – have introduced open data measures with societal objectives in mind (this is especially the case in less affluent countries where internet penetration is still low¹⁷¹).

The main drawback of this option of 'no policy change' is that the already attained societal and environmental impacts would not be amplified by additional sources of data. Also, the barriers to data market and shortcomings in terms of competition conditions would limit the activity of smaller companies, which in turn would also lead to a sub-optimal use of the available data to create social benefits presented above.

6.3.2. Policy Option 1 - Packaged solution: lower intensity regulatory intervention

The societal and environmental benefits of this option compared to the baseline option lie in the first place in the wider availability of data from the transport and utilities sector, notably energy, as well as research results. A more efficient use of available resources

¹⁶⁶ Granickas, K.(2013), 'Understanding the Impact of Releasing and re-Using Open Government Data'. ePSI Platform Topic Report No. 2013/08, August 2013.

¹⁶⁷ Creating Value through Open Data report, European Data Portal, November 2015.

¹⁶⁸ Including 1,7 billion EUR cost savings for EU28 public administrations.

¹⁶⁹ Creating Value through Open Data report, European Data Portal, November 2015.

¹⁷⁰ Granickas, K.(2013), 'Understanding the Impact of Releasing and re-Using Open Government Data'. ePSI Platform Topic Report No. 2013/08, August 2013. p. 13.

¹⁷¹ For example: <http://dataportal.opendataforafrica.org/>

(e.g. traffic infrastructure and energy grids) can reduce contamination and facilitate urban planning (e.g. transport networks or extension of heating systems). In terms of time wasted in traffic jams or waiting in delayed public transport, open data could save 629 million hours of waiting time, corresponding to some 27.9 billion EUR a year.¹⁷² As regards environmental gains, an example of the UK Windsor and Maidenhead council is striking: after having started to publish real-time data on energy use, the council has consumed 16% less gas, electricity, oil and transport fuel in its buildings and vehicles since 2009-2010.¹⁷³ Further examples are given in Annex 9.

A similar argument can be made for the role of research data, which by improving scientific processes also helps fight major societal and environmental challenges of our times (e.g. health¹⁷⁴ or climate change) and in this way increases the overall well-being of society.

In addition, establishing a list of high-value datasets to be released as open data and limiting exceptions of charging above marginal costs, will significantly expand the use of data for environmental and societal purposes. Some of the datasets concerned (in the areas of geographic and meteorological data) are particularly relevant for environmental action. As more data is actually re-used, more societal and environmental beneficial outcomes are provided, reinforcing the current trends.¹⁷⁵

6.3.3. Policy Option 2 – Packaged solution: higher intensity regulatory intervention

This option can be characterised by similar benefits to the ones presented in the packaged solution 1. However, given the higher intensity of the measures proposed and the resulting higher expected overall volume of data affected, it can be assumed that the overall social and environmental impact of this option would be higher.

6.4. Member States' and stakeholders' views

As described in Annex 2, the consultation process sought to collect the views of EU Member States and stakeholders by means of an online consultation and several workshops and meetings. The analysis contributed to an ex-post assessment of the functioning of the current legal framework, along with the exploration of the need for and scope of the review of the PSI Directive on various aspects (dynamic data/APIs, charging rules, research data, data held by entities providing services of general interest, relationship with the Database Directive, national access regimes, and barriers to making data available).

To achieve this, the consultation actions tried to reach out to various stakeholders from the public and private sectors: PSI holders and users, commercial and non-commercial

¹⁷² Ibid.

¹⁷³ Ibid.

¹⁷⁴ It is for example estimated that the time it takes to go from lab to medicine can be cut by 10-15 years with Open Notebook Science, Impact Assessment Support Study, Deloitte, SMART 2017/0061.

¹⁷⁵ <https://www.theguardian.com/sustainable-business/2014/sep/11/climate-change-open-data-apps-government-environment-agency>

re-users, educational and research establishments, public undertakings from the transport and utility sector. Member States were also consulted in dedicated workshops and 11 contributed to the online consultation.

While several workshops collected to a great extent the feedback of PSI holders, 68% of the contributors to the online questionnaire declared being primarily interested in re-using PSI. This allowed for an analysis combining sometimes diverging positions of PSI holders and re-users. The analysis of the combined results gives an indication of the support of the different stakeholder categories for the lower or higher intensity of regulatory intervention.

Most of the **Member States** expressed their views either in the online consultation or in meetings. Very few countries consider the proposal should be more ambitious; in most cases they expressed supportive or neutral positions on most of the measures of lower intensity of regulatory intervention. This primarily includes those Member States that have already enacted relevant legislation (e.g. France), but also those that do not have such provisions. For example, Lithuania, Luxembourg, Poland, Slovakia and Czech Republic all indicated that they are in favour of extending the scope of the Directive. On the charging rules, some were quite supportive (there are Member States where geospatial information is already available for free), but some others expressed concerns in the case all exceptions to charging rules were to be eliminated. The idea to fully apply PSI rules to public undertakings in the transport and utility sector also received negative reactions. Member States were also worried of the potential costs incurred by adding an obligation for public bodies to make all dynamic data available for re-use immediately after collection, or to systematically use APIs. Belgium, Italy, Luxembourg and Malta supported the idea that some essential, high-value datasets should be defined at EU level and made freely available.

PSI holders, i.e. public sector bodies, expressed their views in the online consultation and were represented in a bigger proportion in some of the workshops. They are overall in favour of a moderate evolution of the PSI Directive. In the online consultation, they were quite supportive (although less than the rest of stakeholders) of making efforts in providing dynamic data in real time, including via APIs. They clearly indicated their lack of support for the idea of a drastic modification of the charging rules, although some public bodies already issue information free of charge (geospatial data). There was strong support of PSI holders for Open Access policies and for the idea of a harmonised EU policy on access to and re-use of scientific information. Still, a few expressed concerns linked to the possible extension of the scope of the PSI Directive to research data, considered to be unnecessarily intrusive and rigid (no one-size fits all).

PSI re-users have very much used the public online consultation as a channel to express their views. The measures expected to allow for better access to and re-use of PSI received strong support, e.g. access to dynamic / standardised data as well as to data from public undertakings. This means that the limited ambition of the changes proposed may be contested.

Stakeholders and especially **public undertakings in the transport and utility sector**, in the online public consultation and in a dedicated high-level roundtable, indicated that they were in favour of a sectoral approach when it comes to defining the conditions for making data available for re-use, rather than a horizontal approach. Public undertakings, both in submitted papers and at the roundtable, indicated some concerns to be taken into account, such as commercial sensitivity, personal data, and critical infrastructure. They also raised issues related to administrative burden and their competitive position compared to private companies active in the same sector. As opposed to the application of the full range of PSI Directive rules to their sector, they are expected to be relatively open to the suggested changes as long as the necessary safeguards are put in place.

Finally, the **educational and research establishments** were a specific stakeholders category targeted by both the online consultation and specific workshops. On the possible extension of the scope to publicly funded research data, the vast majority of stakeholders were supportive of the proposed obligation to put in place Open Access policies, as this is fully in line with the 2012 Recommendation on access to and preservation of scientific information and existing policies at national level. Stakeholders considered that issues such as administrative burden, intellectual property rights (IPR) protection, protection of commercial interests, or the question of the sustainability of scientific publishing as a business model should be taken into account. Overall, universities and research establishments fully support Open Access policies and called for a harmonised policy at EU level and possibly beyond. During a workshop, representatives of universities nonetheless called for measures to be proportionate and wondered how the heterogeneity of the status of universities and of licensing practices could be captured by a hard law instrument.

7. HOW DO THE OPTIONS COMPARE?

In line with the European Commission's Better Regulation Guidelines¹⁷⁶ and its toolbox¹⁷⁷, most importantly tool 63, the Impact Assessment study has carried out a Multi-Criteria Analysis (MCA)¹⁷⁸ 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 ratings of the potential impacts of the baseline scenario and the Policy Options are derived from the analysis detailed in section 7 of the study. According to this analysis, PO1 (low intensity policy option) scores highest on the criterion of 'coherence' and 'efficiency', whereas PO2 (high intensity intervention) scores highest on 'effectiveness'. Both are then compared to the BS (baseline scenario).

¹⁷⁶ <https://ec.europa.eu/info/sites/info/files/better-regulation-guidelines.pdf>

¹⁷⁷ https://ec.europa.eu/info/sites/info/files/better-regulation-toolbox_1.pdf

¹⁷⁸ The results of the multi-criteria analysis are available in the Impact Assessment Support Study, Deloitte, SMART 2017/0061.

After performing an outranking matrix, a permutation matrix was established to allow for the selection of a final ranking of all the possible Policy Options towards each other. The results of this process can be described as PO1-PO2-BS, which means the following:

- PO1 is the preferred Policy Option as it provides the most favourable combination of coherence, effectiveness, and efficiency;
- If PO1 cannot be implemented, PO2 would be the second most favourable.
- The least favourable option is the baseline scenario.

The table below gives an overview of how the options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

	Lower intensity regulatory intervention	Higher intensity regulatory intervention
Efficiency	The option presents a favourable ratio of costs and benefits. It is expected to generate slightly lower direct and indirect economic benefits than the higher intensity option and lead to a more moderate decline in the (administrative, technical and opportunity) costs associated with the opening up of PSI.	The option also presents a favourable ratio of costs and benefits. It is expected to generate slightly higher direct and indirect economic benefits and lead to a steeper decline in the (administrative, technical and opportunity) costs associated with the opening up of PSI. This is mostly due to the stronger alignment with the policy objectives, as observed below. Given its strict regulatory approach, it is expected however to impose the costs in a rather abrupt manner, with a negative impact on political feasibility.
Effectiveness	This option is expected to address the need to adapt to technological changes and ensure growth in PSI-based services as well as in cross-border re-use of data, while addressing the emerging threats to fair competition in the Single Market. Some regulatory change is also necessary to tackle persisting market barriers linked to charging for PSI re-use and to speed up the adoption of state of the art data dissemination methods.	This option is also expected to address the need to adapt to technological changes and ensure growth in PSI-based services as well as in cross-border re-use of data, while addressing the emerging threats to fair competition in the Single Market. Given that the higher intensity regulatory intervention is based on stricter regulatory solutions, it is expected to overtake the lower intensity regulatory intervention in terms of the extent to which these policy goals are met.
Coherence	This option has the potential to minimise friction with other EU policies. The lower intensity regulatory intervention largely depends on sectoral legislation for the supply of data and focuses on enhancing downstream re-use, which should ensure full legal coherence.	The far-reaching horizontal measure proposed in the higher intensity regulatory intervention could be difficult to reconcile with a number of ongoing sectoral initiatives in the area of data access and re-use, referred to under section 2.2.4.1 above.
Legal/political feasibility	This option is both politically and legally feasible. The lower intensity regulatory intervention presents a clear advantage over the higher intensity regulatory intervention in what concerns the political buy-in from major stakeholder groups, especially	This option is legally feasible, although based on the available evidence from the stakeholder consultation process, it is less feasible politically due to a substantially lower level of stakeholder buy-in than in the case of the lower intensity regulatory option.

	Lower intensity regulatory intervention	Higher intensity regulatory intervention
	that some of the elements of that policy package have been expressly advocated by several Member States. ¹⁷⁹	
Proportionality	This option is designed to minimise negative impact of the extension of scope, by basing it on the already existing utility procurement legislation. It presents a balanced yet focused policy intervention. By targeting the new requirements in areas where change is necessary, it reduces unnecessary compliance burden in areas where change is not essential and difficult to enact. In addition, the intervention logic tested in the previous revision of the Directive (ensuring a competitive market for PSI re-use as a first step, before the application of an obligation to allow re-use) has proved to be an efficient strategy, ensuring the attainment of objectives for all groups of bodies successively brought within the scope of the Directive, while allowing for an ample adjustment period and giving space for non-legislative instruments to make impact.	This option is also designed to minimise negative impact of the extension of scope, by basing it on the already existing utility procurement legislation. However, the enforcement of the elements included in the higher intensity regulatory intervention would require substantial adjustment efforts by many stakeholders, particularly on the part of the bodies newly covered by the rules. It would also require significant investment by public sector bodies in the update of the necessary IT infrastructure and have a chilling effect on public-private collaboration projects. In addition, Member State acceptance of a radical change, relatively early after the previous regulatory revision, is low. All such risks appear in disproportion to the objectives sought.

This analysis translates into the following overview:

	Efficiency	Effectiveness	Coherence	Legal/political feasibility	Proportionality
Lower intensity regulatory intervention	+	+	+	+	+
Higher intensity regulatory intervention	++	++	-	-	-

For efficiency, effectiveness and coherence, the scores are given on the expected magnitude of impact as explained above: ++ being strongly positive, + positive, and – negative. For legal/political feasibility and proportionality, + means that the assessment is positive, and – means that it is negative.

8. PREFERRED OPTION

Based on the evidence presented above, a mixed package of lower intensity regulatory intervention combined with an update of existing soft law is the preferred option.

It allows for a targeted and proportional intervention, amounting to an incremental strengthening of the Commission's open data policy. It will lead to a significant improvement over the baseline scenario, is broadly acceptable to stakeholders and can be realistically enacted within a reasonable timeframe due to lack of notable Member State opposition.

Although the benefits of the higher intensity regulatory intervention scenario are substantial, that scenario is also generally characterised by a lower feasibility, higher

¹⁷⁹ A prime example is the list of 'open by default datasets'.

compliance costs and higher risks for legal and policy coherence. This has also been confirmed by the multi-criteria analysis performed for all the options.¹⁸⁰ Also, the higher intensity regulatory intervention scenario will not be supported by many stakeholders, including Member States.

This leads to a preferred option that is based on the following elements:

- **Dynamic data/APIs:** a ‘soft’ obligation for Member States to make dynamic data available in a timely manner and to introduce APIs. For a limited number of fundamental high-value datasets (to be adopted through a Delegated Act) there will be a hard obligation to do so.
- **Charging:** tighten the rules for Member States for invoking the exceptions to the general rule that public sector bodies cannot charge more than marginal costs for dissemination. Create a list of fundamental high-value datasets that should be freely available in all Member States (same datasets as above, to be adopted through a Delegated Act).
- **Data in the transport and utilities sector:** only public undertakings will be covered, not private companies. A limited set of obligations will apply: public undertakings can charge above marginal costs for dissemination and are under no obligation to release the data they do not want to release. 2003 rules of the PSI Directive apply (as explained in the table on p. 31).
- **Research data:** Member States will be obliged to develop policies for open access to research data resulting from publicly funded research, while keeping flexibility in implementation. An updated Recommendation to Member States on access to and preservation of scientific information would guide Member States on the elements ideally contained in an Open Access policy. The PSI Directive would be extended in scope, so as to cover research data that have already been made accessible as a result of open access mandates, focusing on re-usability aspects.
- **Non-exclusivity:** reinforced transparency requirements for public-private agreements involving public sector information

This will be combined with an update of the Recommendation on the access to and preservation of scientific information¹⁸¹ and a clarification of the interaction between the PSI Directive and the Database and INSPIRE Directives.

8.1. Estimated impact of the preferred option

The Impact Assessment study¹⁸² indicates that while in the baseline scenario the direct value of PSI is expected to grow from EUR 52 billion in 2018 to almost EUR 150 billion in 2028, implementing the preferred, politically feasible option will ensure that this growth can be 30% higher, reaching almost 194 billion EUR. It is also expected that

¹⁸⁰ The results of the multi-criteria analysis are available in the Impact Assessment Support Study, Deloitte, SMART 2017/0061.

¹⁸¹ C(2018) 2375.

¹⁸² Impact Assessment Support Study, Deloitte, SMART 2017/0061.

under the preferred option the number of jobs based on PSI will reach 709,000, i.e. 40% above the baseline scenario which foresees 518.000 jobs in that timeframe. Finally, the cost of making PSI available for re-use will be reduced to 3 billion EUR if the preferred option is chosen, which is 21% lower than the estimated cost under the baseline scenario (3.8 billion EUR).

8.2. REFIT (simplification and improved efficiency)

<i>REFIT Cost Savings – Preferred Option</i>		
<i>Description</i>	<i>Amount</i>	<i>Affected stakeholders</i>
Enhanced use of APIs and proactive publishing of dynamic data online results in a decrease of administrative burden due to a lower number of re-use requests to process and a lower risk of complaints (including litigation).	Savings due to a lower number of requests (the processing of which costs min. 375 EUR per institution per request). ¹⁸³	Reduces administrative burden and related costs of public sector bodies on local, regional and national level.
Reduced application of charges, bringing the regulatory framework in line with the recommendations included in the 'PSI Guidelines' of 2014, since what was considered 'best practice' back then would now become binding administrative procedure.	Reduced cost of processing and monitoring payments.	Reduces administrative burden for both public sector bodies and data users, SMEs. A win-win situation, especially in cases where the costs outweigh the income generated by charging for re-use.
Discontinued reporting obligation (due to a detailed and regular monitoring exercise performed by the European Data Portal, which feeds the annual Open Data Maturity report), allows the Member States to save human resources and time.	Estimated at min. 280,000 EUR (1400 man/days). ¹⁸⁴	Reduces administrative burden and related costs of public sector bodies on local, regional and national level.
Enhanced legal certainty and easier application/interpretation of the Directive thanks to clearer rules on charging for the re-use of documents and on the interplay between Database and INSPIRE Directives.	Impossible to quantify but could be considerable in case of legal counsel costs.	Reduces administrative burden and related costs of public sector bodies on local, regional and national level.
Faster and cheaper development of business models by SMEs thanks to a higher supply of high value open data (through a common list of datasets and the extension of the scope of the Directive) and its easier integration into digital services (due to the systematic use of APIs).	A key factor contributing to the increase of total economic value of PSI beyond the public sector.	Reduces investment and transactional costs, lowers barriers to market entry by SMEs, limits administrative burden linked to filing individual requests.

¹⁸³ Considering that API use might decrease the need to process the number of requests roughly by a half, an EU-wide benefit could reach 375 EUR (15h x 25 EUR) per institution per request multiplied by the number of requests which will no longer be needed.

¹⁸⁴ The direct saving for 28 Member States was estimated given that a) the elaboration of the most complete report required around 100 man/days, b) the elaboration of a medium quality report would take 50 man/days and c) the average labour costs per day are 200 EUR (25x8).

9. HOW WILL IMPACTS BE MONITORED AND EVALUATED?

- The European Data Portal¹⁸⁵ conducts the annual 'landscaping exercise' that monitors performance indicators in each EU Member State. The exercise, which is compiled in an Open Data Maturity Report, will be continued, while the definition of the performance indicators will be adjusted so as to be taken into account from the date of the adoption of the revised Directive. The European Data Portal will for example measure (both via structured feedback from national authorities and by the monitoring of the data flows through the portal) the progress in which research data and data coming from public undertakings in the transport and utility domains are published via open data portals and other repositories. It will also be able to detect the amount of dynamic data being made available and the increased use of APIs.
- The existing expert groups (PSI Group and the sub-group on portals¹⁸⁶; Expert Group on National Points of Reference on Scientific Information) will assist the Commission in evaluating the state of transposition of the revised legislation and communicate the outcome of their own, national assessment and related studies.
- PSI Request repository¹⁸⁷, an online EU-wide register of requests submitted by re-users with the description of the follow-up given by public sector bodies, is currently being set up to help the Commission gather evidence on the impact of the upcoming measures in terms of easier availability of data without the need to submit individual requests, and the ensuing reduction of administrative burden.
- Evaluation of the implementation of the Directive based on a modified review clause (Article 13), possibly six years after the adoption of the amending legislation.
- Ad hoc studies, as deemed appropriate (e.g. in line with the development of data processing technologies) and targeted surveys (e.g. to measure the reduction of the administrative burden among selected groups of stakeholders).

Monitoring indicators for specific objectives:

Specific objectives	Operational objectives	Proposed monitoring indicators
Adaptation to technological changes in the field of data management and use	Incentivising public bodies to use modern methods of access and use of dynamic data	Increase in key registries using APIs Increase in share of dynamic PSI vs total available for re-use
Lowering the existing barriers to accessing the PSI re-use market and preventing the emergence of new ones	Limiting the range of situations in which charging for PSI is allowed. Ensuring transparency around arrangements that can lead to 'data lock-in'	Reduction in the number of public sector bodies charging above marginal costs Increase in the number of data start-ups Number of complaints about unjustified exclusive data use
Making more data available for re-use as raw material for innovation	Higher supply of data from the utilities sector and research establishments	Increase in the number of datasets from these two areas available via APIs and portals

¹⁸⁵ The portal operates as a digital service infrastructure under the Connecting Europe Facility. It is funded by the European Commission: www.europeandataportal.eu.

¹⁸⁶ The expert group meets twice a year and reports to the Commission on the progress of open data policy and legislation.

¹⁸⁷ <http://www.psimonitor.eu/index.php/en/>

Glossary

<i>Term or acronym</i>	<i>Meaning or definition</i>
Artificial Intelligence (AI)	Artificial Intelligence, intelligence displayed by machines, is applied when a machine mimics cognitive functions that humans associate with other human minds, such as learning and problem solving.
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.
contracting authority	A State, regional or local authority, a body governed by public law or an association formed by one or more such authorities or one or more such bodies governed by public law.
cost recovery method	A principle which allows for the recovery of incurred eligible costs, such as costs relating to the creation of data or costs relating broadly to distribution.
Database Directive	Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, which harmonises the treatment of databases under copyright law and introduces the <i>sui generis</i> right for the creators of databases even when they do not qualify for copyright.
Public sector data lock-in arrangements	Arrangements where public sector bodies grant access to their datasets (usually free of charge) to one or a small number of economic operators, for example in exchange for in-kind compensation, leading to a situation in which re-use of such data by this economic operator would take away the incentive for commercial re-use of the same dataset by other companies.
document	Any content whatever its medium (written on paper or stored in electronic form or as a sound, visual or audio-visual recording), or any part of such content; synonymous with data.
dynamic data	Data from sensors (e.g. bus arrival times, meteorological data), whose economic value depends on its real-time availability.
European Data Portal (EDP)	A portal that harvests the metadata of Public Sector Information available on public data portals across European countries. Information regarding the provision of data and the benefits of re-using data is also included.
European Open Science Cloud (EOSC)	A virtual environment to store, share and re-use research and scientific data across disciplines and borders, which will be underpinned by the European Data Infrastructure, deploying the high-bandwidth networks, large scale storage facilities and super-computer capacity necessary to effectively access and process large datasets stored in the cloud.
FAIR (findable, accessible, interoperable and re-usable) principles	Set of principles developed within the scientific research community in order to make research data findable, accessible, interoperable and re-usable.
G8 Open Data Charter	A document from 2013 listing 5 strategic principles that all G8 members committed to act on, which include an expectation that all

	government data will be published openly by default, alongside principles to increase the quality, quantity and re-use of the data that is released.
high-value datasets	Datasets which are particularly valuable assets for the economy and society at large. Access to and the re-use of such datasets can speed up the emergence of value-added information products and services, and also encourage participatory democracy. G8 members have identified 14 high-value areas – from education to transport, and from health to crime and justice – which should help unlock the economic potential of open data, support innovation and provide greater accountability.
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.
Internet of Things (IoT)	A network of physical devices, vehicles, home appliances and other items embedded with connectivity software, which enables these objects to connect and exchange data.
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.
marginal cost method	A principle applying to all charging for public sector data re-use in the EU, with some exceptions: public sector bodies may charge no more than the marginal cost of reproducing, providing and disseminating the documents.
Open Access Infrastructure for Research in Europe (OpenAIRE)	Electronic infrastructure and supporting mechanisms for the identification, deposition, access, and monitoring of research data (publications and data) funded by H2020.
Open Access (OA)	Online, free availability of research outputs without restrictions on use commonly imposed by publisher copyright agreements. Open Access includes the outputs that scholars normally give away for free for publication; it includes peer-reviewed journal articles, conference papers and datasets of various kinds.
open data	Data that is freely available to everyone to access and re-use as they wish, without restrictions from copyright, patents or other mechanisms of control. In the context of this document, open data refers to public sector information which is openly re-usable.
open format	A file format that is platform-independent and made available to the public without any restriction that impedes the re-use of documents.
public sector body	A State, regional or local authority, a body governed by public law or an association formed by one or several such authorities or one or several such bodies governed by public law. This definition is based on that of a 'contracting authority' under EU procurement legislation (Directive 2014/24/EU, Article 2).
Public Sector Information (PSI)	Information (i.e. 'documents') in many areas of public sector activity, including social, economic, geographical, weather, tourist, business, patent and educational information, which the public sector collects or produces.

public undertaking	Any undertaking over which the contracting authorities may exercise directly or indirectly a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules which govern it.
reasonable return on investment	Amount understood as a percentage, in addition to eligible costs, allowing for the recovery of the cost of capital; and the inclusion of a real rate of return (profit).
re-use	The use by persons or legal entities of documents held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the documents were produced.
<i>sui generis</i> database right	A <i>sui generis</i> database right is a property right, comparable to but distinct from copyright, that exists to recognise the investment that is made to compile a database, even when this does not involve the creative aspect that is reflected by copyright.

ANNEX 1: PROCEDURAL INFORMATION

LEAD DG, DECIDE PLANNING/CWP REFERENCES

The proposal for an amended Public Sector Information (PSI) Directive¹⁸⁸ was prepared under the lead of the Directorate-General Communication Networks, Content and Technology. In the DECIDE Planning of the European Commission, the review process is referred to under item PLAN/2017/1391. The Commission Work Programme for 2018 includes the review of the PSI Directive among the REFIT initiatives, under the header "A Connected Digital Single Market".

ORGANISATION AND TIMING

Work on the preparation of this Directive review started in May 2017. An Inter-Service Steering Group assisted DG Communication Networks, Content and Technology in the preparation of the back-to-back exercise (evaluation and Impact Assessment) and included Commission services of 16 Directorate-Generals, together with the Commission's Legal Service and Secretariat General.

The Steering Group met on occasions between July 2017 and March 2018. At each occasion, the members of the Steering Group were given the opportunity to provide comments orally or in writing on the draft versions of the documents presented.

CONSULTATION OF THE RSB

The Impact Assessment report was reviewed by the Regulatory Scrutiny Board on 14 March 2018. Based on the Board's recommendations¹⁸⁹, the Impact Assessment has been revised in accordance with the following points:

RSB Opinion (copy of the RSB comments from the opinion)	How and where comments have been addressed
The report does not adequately reflect stakeholder views. In particular, it does not sufficiently address stakeholder concerns about personal data and database protection. It gives little indication about support from Member States and public bodies. The report should make better use of its extensive evidence base. In particular, it	The views of stakeholders have been developed with increased granularity, i.e. depicting positions by type of stakeholders (e.g. PSI holders or re-users, educational and research establishments, public undertakings in the transport and utilities sector) and relating them some specific consultations actions. The views of stakeholders are reported in

¹⁸⁸ COM(2018) 234.

¹⁸⁹ SEC(2018) 206.

<p>should report in more detail on stakeholder views. It should provide a more granular representation of the categories of stakeholders at stake. It should transparently report on whether stakeholders expressed views on the different parts of the initiative. The report should elaborate on the concerns of stakeholders and it should explain the respective mitigation measures that the initiative intends to propose.</p>	<p>the section 6.4, in a much more detailed way, but are also reinforcing the description of the problem areas in the section 2 of the IA.</p> <p>The positions of Member States have also been described in a more detailed way both in section 3.2 on subsidiarity and in section 6.4.</p> <p>An explanation was given regarding the nature of the stakeholders' view on the re-use of data in line with personal data protection requirements, as well as the possible mitigating measures in that respect.</p>
<p>The report does not sufficiently explain how the initiative avoids conflicting with the Database Directive and the General Data Protection Regulation (GDPR).</p> <p>The report should clarify the relationship and coherence of the PSI with the GDPR and the Database Directive. The report should elaborate on how the initiative addresses challenges of anonymisation of data and privacy, protection of databases and IPR. It should better address stakeholder concerns related to the necessary (specific) protection of data and databases in sensitive sectors such as the health sector, when they are open to private operators and service providers. The report should clarify the risks (of re-identification for example) and limits of the PSI. Consequently the report should better explain the safeguards against those risks. The report should elaborate on how the PSI tailors measures to tackle sui generis rights of national authorities.</p>	<p>Section 1.2 of the IA has been substantially reinforced in order to provide a more comprehensive overview of the complementarity of the PSI Directive, the GDPR, the Database Directive and the INSPIRE Directive. A specific section has been added to the text explaining how the challenges of pseudonymisation, anonymisation and data protection are to be addressed, in particular, through targeted funding activities addressing privacy enhancing technologies both under the H2020 and CEF programmes.</p>
<p>The definition of options is not sufficiently specific and their range is too narrow to provide a genuine choice of alternative solutions.</p> <p>The report should avoid including unworkable measures in option 2. The report should explain in more detail the content of the options, including the list of high-value data that potentially would</p>	<p>The choice for the distinction between the policy options has been explained in chapter 5.2., including the clarification of the interplay between the main legislative packaged solutions of 'high' and 'low' intervention intensity and the 'intermediate' option consisting of a future Delegated Act subject to a separate future Impact Assessment. The table on page 31 was also</p>

<p>have to be disseminated free of charge. Similarly, the report should explain the modalities for tightening the rules for invoking exceptions to the marginal cost of dissemination approach. The report should better clarify what issues will be dealt with in delegated acts. It could also usefully expand the range of options with intermediate solutions (possibly combining regulatory with soft-law measures) in order to provide the decision makers with a wider choice of alternatives.</p>	<p>updated to bring it in line with the descriptions of the two main packages of policy options and the inclusion of binding and non-binding elements was clarified.</p> <p>The reasons behind assessing the impact of two main policy packages, both including the soft and hard law measures, while assigning the 'intermediate' measure of high value datasets to a 'lower intensity' option, have also been explained.</p> <p>Each policy option is now described with reference to the concrete legislative changes resulting from it (see: table of policy options in chapter 5.2).</p>
<p>The report should clarify the role of the initiative in the context of the 2018 Data Package. It should also elaborate on the relationship of the PSI with international initiatives such as the G8 Open Data Charter and the recent OECD Recommendations on data openness.</p>	<p>Two paragraphs have been added to the Introduction of the report clarifying the role of the initiative in the 2018 Data Package. Additionally, Section 1.1 (Policy Context) has been reinforced in order to give a thorough overview of the relations between the PSI Directive and international initiatives like the G8 Open Data Charter and the relevant OECD Council Recommendations.</p>
<p>Based on the accompanying study, the report should present more data on the magnitude of the four problems it identifies. It should analyse them in the context of potential single market distortions. It should explain what the main drivers that underpin the four issues are.</p>	<p>All the sub-sections in Section 2.2 (on the description of the four problems have been reinforced by the respective analysis and arguments used in the PSI Impact Assessment Support Study (SMART 2017/0061) and the Evaluation Report (Annex 5). The aspect related to extent of the problems over time and in line with technological changes has also been highlighted.</p>
<p>The report should draw on the accompanying study to more extensively describe the elements underpinning the baseline scenario. This includes information on the current and projected costs of production, anonymisation, and dissemination of various PSI datasets and their potential market value. The baseline should project how the current situation will evolve in the absence of further EU action. It should serve as the comparator of</p>	<p>Chapter 6.1.1. has been complemented with new, detailed information on the factors influencing the steadily increasing impact of open data despite the lack of additional EU intervention.</p> <p>References to more detailed descriptions included in the Support Study and acting as methodological background, have also been added.</p>

the options envisaged for this initiative.	
<p>The section on impact needs also to include more information on costs and benefits. It can draw on the external study for this. It should indicate the costs that are quantified and those that are not (administrative and litigation costs). It should be more specific on how benefits and costs are distributed over time and across stakeholder categories. It should also show that public bodies can cope with additional obligations on re-use of data.</p>	<p>Additional information regarding the methodology used by the Support Study to measure costs and benefits, as well as their distribution over time and across the individual elements of the policy options, has been added to Annex 8 and Annex 5.</p> <p>Section 6.1 includes a more detailed explanation of the factors taken into account for the estimates of expected impact of the policy options retained.</p>
<p>The Board takes note of the quantification of the various costs and benefits associated to the preferred option of this initiative, as assessed in the report considered by the Board and summarised in the attached quantification tables.</p>	

EVIDENCE, SOURCES AND QUALITY

Evidence-collection process

A REFIT evaluation¹⁹⁰ (ex-post evaluation) was performed in parallel with the review and Impact Assessment. This evidence-collection process has been supported by a study (SMART 2017/0061), combining an evaluation with an Impact Assessment ("back-to-back"). Complementary elements will come from the preliminary results of the study on PSI Request repository (SMART 2016/0088) which provided a comparison of re-use requests and their outcomes across countries.

At the same time, a reporting exercise took place, as mandated by article 13.2 of the Directive, aiming to collect feedback from Member States on the availability of public sector information for re-use, the conditions under which it is made available, and redress practices. This reporting exercise was complementary to the landscaping action led by the European Data Portal on the maturity of open data across Europe. A series of indicators cover the level of development of national policies promoting open data, an assessment of the features made available on national data portals, as well as the expected impact of open data.

Stakeholders' consultation process

¹⁹⁰ SWD(2018) 145.

In addition to the feedback received on the Inception Impact Assessment (IIA), a public online consultation was conducted from September to December 2017, aiming at both evaluating the implementation and effects of the current PSI Directive, and at getting feedback on possible impacts of the different policy options. The Inception Impact Assessment served as a supporting document to this public online consultation.

In addition to these online consultations, workshops, some of which organised in the framework of the aforementioned "back-to-back" study, guaranteed further interactions with Member States, public bodies and businesses. Annex 2 presents a detailed analysis and conclusions from this consultation process.

ANNEX 2: CONSULTATION ON PSI DIRECTIVE REVIEW

INTRODUCTION

The Commission held a consultation on the review of [Directive 2003/98/EC on the re-use of public sector information \(PSI Directive\)](#) between June 2017 and late January 2018. The aim of this consultation was to assess the functioning of the Directive, consider the scope of the review, and reflect on policy options. At the same time, it explored the issue of access to private sector data which are of public interest. The consultation sought to involve PSI holders (public sector bodies) and re-users (public, private, commercial and non-commercial actors).

The feedback received from workshops (mainly from PSI holders) and an online questionnaire (68 % of respondents to which were primarily interested in re-using PSI) form the basis for the analysis and conclusions presented here.

Inception impact assessment

The [inception impact assessment](#) was available for feedback on the Better Regulation portal from 18 September to 16 October 2017. The seven [replies](#) (from associations, public organisations, national statistical offices and private individuals) addressed topics including improving the interoperability of data, maintaining current charging rules, and maintaining the exception for cultural establishments. On extending the Directive's scope to public utilities undertakings, concerns were raised as to cost and coherence with other legislation.

Online consultation

The public online consultation (19 September to 16 December 2017) asked for views on how the Directive is implemented, (including problems, objectives and possible options for the future) and addressed the issue of public access to data of public interest held by the private sector. It targeted all interested parties, including governments, public sector content-holders and users, commercial and non-commercial re-users, experts and academics, and the general public.

The 273 replies, from all Member States, included:

- almost 29 % from public organisations (general and statistical services, research, etc.);
- almost 25 % from associations (representing public or private actors); and
- 25 % from citizens.

Of the business respondents, half were very large organisations (5 000+ employees) and 40 % were SMEs.

In addition, 58 position papers and five stand-alone contributions were received, mostly on specific issues, e.g. access to scientific information and to private sector data of public interest.

Other consultation actions

- [Workshop](#) on public bodies' access to commercially held data of public interest (26 June 2017);
- [Meeting](#) of Member State representatives from the 'PSI group' (15 November 2017);
- [Meeting](#) of the National Points of Reference on access to and preservation of scientific information (5 December 2017);
- [Workshop](#) on access to scientific information and extending the scope of the Directive to research data (14 December 2017);
- [Workshop](#) for PSI holders and re-users (18 January 2018);
- [Public Hearing](#) on the PSI Directive review (19 January 2018); and
- High-level roundtable on opening up transport and utilities sector data (23 January 2018).

RESULTS OF THE CONSULTATION PROCESS

Evaluation

The implementation of the Directive was assessed against the Better Regulation criteria (effectiveness, efficiency, relevance, coherence, EU added-value and clarity). The online consultation was the main opportunity for stakeholders to make detailed statements.

Effectiveness

- ✓ 81 % of respondents to the online consultation felt that more data held by public sector bodies had become available for re-use;
- ✓ 73 % agreed that PSI was increasingly providing a basis for innovative services and products;
- ✓ over half (and most of those who provided additional feedback as free text) felt that PSI had become more affordable, including for start-ups and SMEs;
- ✗ several stakeholders felt that cross-border use of PSI was still difficult, as practices across Member States vary widely and this creates legal uncertainty;
- ✗ only 38 % of respondents felt that exclusive arrangements were used exceptionally and only in the cases set out in the Directive (while Member States representatives generally report that they are not aware of exclusive arrangements based on PSI); and
- ✗ 67 % saw the variety of licensing conditions as still hampering effective re-use.

At the 18 January workshop for PSI holders and re-users, there were several calls for EU intervention to prevent exclusive agreements.

Efficiency

- ✓ 66 % of online respondents argued that implementation costs borne by public sector bodies are offset by the socio-economic benefits of data re-use;
- ✓ 72 % argued that compliance requires public bodies to improve data management, leading to cost savings and efficiency gains; but
- ✗ 44 % criticised the redress procedures as lengthy, inefficient and costly.

Relevance

Respondents overwhelmingly agreed that the PSI Directive is still **relevant**, in particular because it ensures PSI supply (85%), fair market access for re-users (84%), the transparency and accountability of public sector bodies (81 %) and the usability (e.g. machine readability) of data (77 %).

Many said that rapid technological advances made the Directive even more relevant; these affect the possible forms of data re-use and the legislation should take account of them.

Stakeholders and especially Member State representatives were generally very supportive of the review process, based on the observation that the uneven implementation across Member States is creating fragmentation of the single market or bottlenecks to market development.

Coherence

- ✓ in general, it was felt that the Directive was coherent with other relevant EU legal acts;
- ✗ the *sui generis* database right was seen as problematic because public bodies could use it to restrict the applicability of the PSI access and re-use rules; and
- ✗ the Directive's relationship with personal data protection legislation is clear; however, the stakeholders pointed to certain practical issues. Some public sector bodies have raised the question of appropriate techniques to be used for pseudonymisation or anonymization processes.

EU added value

- ✓ 87 % of online respondents agreed that the Directive had encouraged national authorities to open up more public sector data;
- ✓ almost 63 % felt that it had facilitated access to PSI from countries other than that in which the person concerned lives;

- ✓ almost 64 % felt that it had helped create an EU-wide market for products and services based on PSI; but
- ✗ additional feedback shows that the re-use of data across borders is still burdensome, as national practices are still very different and sometimes ‘data protectionist’ (e.g. on valuable datasets such as business, address and real estate, and planning registers).

Clarity

- ✓ 57 % of online respondents felt that the Directive was easy for public bodies and re-users to understand and implement; but
- ✗ 63 % felt that some provisions could be made simpler or clearer.

Review of the Directive

Practical arrangements for document access and searches

There was general support for further standardisation of metadata, standardisation of data themselves, and greater availability and usability of real-time data. These positions were confirmed in the dozen of papers addressing these questions.

Of the 194 online respondents, 46 % strongly agreed and 25 % agreed that metadata should be made available in a mandatory open standard, e.g. DCAT-AP. Respondents acknowledged the need for metadata documenting the content of data so as to ensure their actual re-usability. The standardisation of data themselves was seen as (highly) desirable. Mandatory open standards were recommended as the way to improve data usability. Some saw funding as a critical issue.

There was an even spread of opinion as regards the availability of dynamic data (e.g. sensors, satellite data). Examples (e.g. Transport for London, Seapilot) were cited during the workshop on 18 January. Many opinions in the online consultation and position papers reported a remaining need for (more) real-time data, the cost of such availability and a need to prioritise efforts on the basis of real demand.

It was stressed that data standards should relate to specific domains (transport, geospatial, statistical, etc.). In particular for dynamic data, the cost and technical complexity of setting and implementing standards raise a need for appropriate supporting measures.

77 % of respondents confirmed the need to encourage public bodies to provide dynamic data in real time and invest in technical solutions (e.g. APIs, Application Programming Interfaces) facilitating data usability. This high demand is still only partially met. This confirms the stakeholders' opinion also expressed in a [public online consultation earlier in 2017](#), where 68% of respondents clearly supported an increased use of APIs.

Charging rules

38 % of the 206 online respondents expressed a slight preference for keeping the rules that prevent public bodies from setting excessive or arbitrary charges on re-use. However, 42 % wanted to abolish the exceptions to Article 6(1) (so that the marginal cost of dissemination would become the upper limit for charging) or at least clarify (49 %) the circumstances in which they can apply.

PSI re-users are clearly in favour of abolishing or at least clarifying exceptions, while most PSI holders consider that they should not be changed and need no clarification.

Stakeholders also expressed divergent views in their additional feedback. Some argued that the rules should promote open access by eliminating or at least minimising charges. Many others cited challenges that public administrations are facing that justify the rules, e.g.: the sustainability of investment in digitalisation (especially if more is to be done on standards, interoperability, dynamic data, etc.), and balancing the interests of PSI holders and re-users.

The variety and opacity of national charging practices was raised at the workshop for PSI holders and re-users.

Scientific information and data held by research and educational establishments

There was a consensus that data from publicly funded scientific research should be as openly accessible and re-usable as possible.

81 % of the 178 online respondents agreed that a common EU open access policy should apply to all research-funding organisations and academic institutions; only 6 % disagreed. This position was also voiced at the public hearing and the workshop on access to scientific information.

There were similar views on opening up the (currently exempt) administrative data of educational and research establishments, although at the meeting of National Points of Reference in access to an preservation of scientific information, a Member State representative expressed concerns in the context of competition among universities in Europe.

Of 159 online respondents, 90 % supported the principle of opening up scientific results (publications and research data) from public funding. In the open fields in the online consultation and at the workshop on access to scientific information, stakeholders stressed the similarities between such results and government-held information, in particular as regards data re-use value.

At the same time,

- some respondents pointed to issues for future policy-making, including the financial sustainability of publishers, the protection of intellectual property rights (IPRs), the protection of personal data and trade secrets;
- structural solutions will be required that do not add unnecessarily to the administrative burden for researchers; and
- there are other linked barrier to be addressed, including imperfect data-management capabilities, onerous licensing conditions, and problems with common (meta)data standards.

Data held by entities providing services of public interest

On extending the scope of the Directive to data generated by publicly owned companies or independent operators performing public tasks, those in favour outnumbered those against. Only 23 % of the 193 online respondents to the question agreed that such data were currently available for re-use. In a breakdown by sector, the figure is higher for transport (36 %) and considerably lower (14 %) for utilities. 71 % believe that such data should be made available for re-use, whether the data-holder is public or private. 81 % indicated that, if there were such an obligation, the data should be available for all re-users and for other purposes. However, the very few responses coming from energy, waste and water businesses expressed either doubts or strong opposition.

While stakeholders in favour emphasised the benefits (e.g. better and cheaper services for citizens), others based their opposition or caution on factors such as the commercial sensitivity of some data, the need to ensure a level playing-field for public and private actors, personal data protection considerations and the need to safeguard the security of critical infrastructure. In particular stakeholders from the transport sector have expressed concerns. They believe that data is crucial for maintaining their competitiveness and as a result any requirements to make it more open and re-usable should be carefully balanced, as new obligations in this regard may distort competition in the sector. Stakeholders from entities active in the transport and utilities sectors have also indicated that imposing data sharing obligations on them may have an impact on ensuring the security of critical infrastructure. Moreover, some are anxious that compliance with the new requirements will mean additional administrative burden. As a result any requirements to make it more open and re-usable should be carefully balanced.

At the 23 January roundtable, some argued that the issue should be addressed through sectoral legislation.

Relationship with the Database Directive

A third of the 190 online replies reported instances of public sector bodies invoking rights under the Database Directive to prevent the re-use of PSI from databases,

including transport data, legislation and case law, company registers, and public undertakings' data in areas such as healthcare, translation memories and cultural heritage.

In general, the concern was that public bodies circumvent the PSI Directive by invoking database rights, even though the data are not subject to public sector or third-party rights. The interplay between the two Directives was also raised at the public hearing.

Of the 196 online respondents to the question, 68 % were in favour of clarifying the relationship between the Directives, so as to ensure that public bodies cannot invoke database rights in order to prevent the re-use of PSI. Stakeholders noted that this was already the case in some national and regional jurisdictions (France, Flanders). Recommended approaches ranged from guidelines and clearer formulations to making the *sui generis* database right inapplicable to databases that are maintained by public bodies or with public funding. On the other hand, 17 % found the relationship sufficiently clear, thanks *inter alia* to recitals 22 and 23 of the PSI Directive, which stress that public bodies should exercise their IPRs in a way that facilitates re-use. Public transport undertakings and utilities insisted that they needed to keep their database rights over publicly funded data, in order to protect their legitimate interests in the competitive market.

National access regimes and barriers to data re-use

National access regimes do not generally seem to constitute a major obstacle to PSI re-use, but 39 % of online respondents felt that the differences between them hamper the emergence of EU-wide services and products based on PSI. Some suggested that a common definition of the types of document that could be exempted (e.g. due to national security restrictions) would help prevent the frequent misuse/abuse of exemptions. Differences between national access regimes were also raised at the 18 January workshop for PSI holders and re-users and at the public hearing.

28 % of online respondents agreed that the link between access and re-use is clear and useful, and that it prevents re-use that could harm the interests of the state, individuals or third parties. However, 23 % argued that the link is not clear and 18 % said that national rules on access (e.g. time limits for obtaining responses, administrative charges, lack of appeal options) are stricter than the Directive and hamper re-use. Some felt that certain national access regimes lack clarity and transparency.

As reasons for not making data available, public bodies cited data security rules and obligations (21 %), excessive costs (15 %), and possible misuse (13 %).

As for data held by operators under a public service contractual arrangement, the main arguments were that the arrangement did not require that they be made available and that it would be too costly.

In cases where requests for data had been granted, the obstacles reported were poor metadata (34 %), a lack of information on data management (33 %), and unclear or inconsistent terms and conditions for re-use (30 %).

Also a quarter of respondents referred to a lack of machine-to-machine interfaces (APIs), and machine-readable/standardised licences. Individual stakeholders cited the scope of IPRs, the geographical scope of licences and the unavailability of linked open data.

78 online respondents provided further input on ways to ensure data protection in the context of PSIs. Suggested solutions included:

- technical (e.g. pseudonymisation, anonymisation, encryption, separation of networks);
- legal (e.g. consent-oriented solutions or ‘privacy by design’ rules); and
- other (e.g. training of officials, appointing a person in charge of anonymisation).

Access by public sector bodies to private sector data of public interest

Acceptance of the idea of allowing public administrations to access and re-use private sector data of public interest has increased since the [‘Building a European data economy’ consultation](#) (June 2017). It was supported by 88 % of 205 respondents, across all types of contributors.

Of the 189 online respondents, 81 % felt that specific legal measures were needed. Among them, EU legislation by sector was mainly supported, followed by general principles, and then by adoption of specific rights and guarantees.

Businesses tended to support specific rights and guarantees, while public organisations favoured EU wide legislation by sectors and general principles. Online, in papers and at the public hearing, statistical offices and other public organisations expressed support for laying down principles common to all Member States and highlighted the need to legislate before wide national disparities appear in the field of statistics.

Although businesses supported that approach online, in papers and at events such as the public hearing many businesses and associations stressed the importance of voluntary measures and contractual freedom to implement them.

Stakeholders in all categories called for a clearer definition of ‘public interest’ and clarity on the objectives and scope of the initiative. This suggests a need for further discussion with stakeholders, including public organisations with a view to establishing clear areas for action.

NEXT STEPS

The consultation results contribute to the evaluation and review of the PSI Directive, as part of the data package to be adopted in April 2018. The evidence collected to support

the impact assessment was fed into an analysis of organisations' expected positions on the policy options (lighter vs. heavier regulation), according to their role in the PSI context (Member States, PSI holders and re-users, public transport and utilities undertakings, educational and research establishments). It was concluded that stakeholders are most likely to support lighter regulation.

ANNEX 3: VISUALISATION OF THE VIEWS EXPRESSED BY STAKEHOLDERS IN THE PUBLIC ONLINE CONSULTATION ON THE REVIEW OF THE PSI DIRECTIVE

Figure 6 - Coherence of the PSI Directive

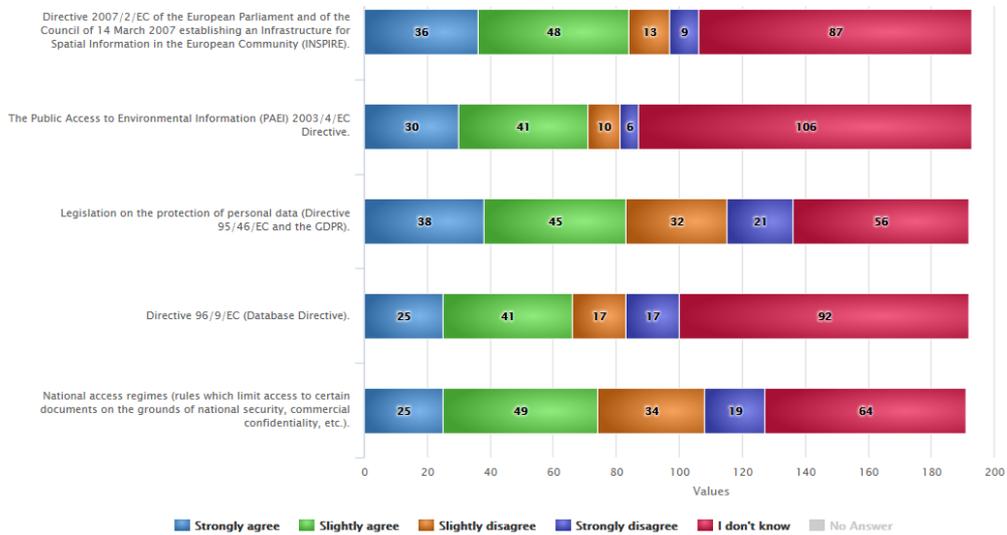


Figure 7 - Effectiveness of the PSI Directive

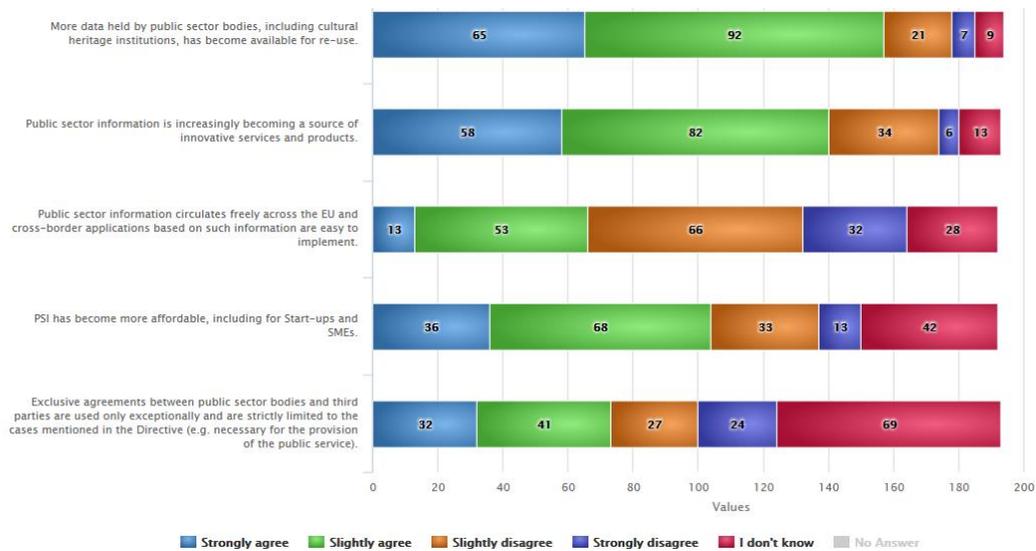


Figure 8 - Efficiency of the PSI Directive

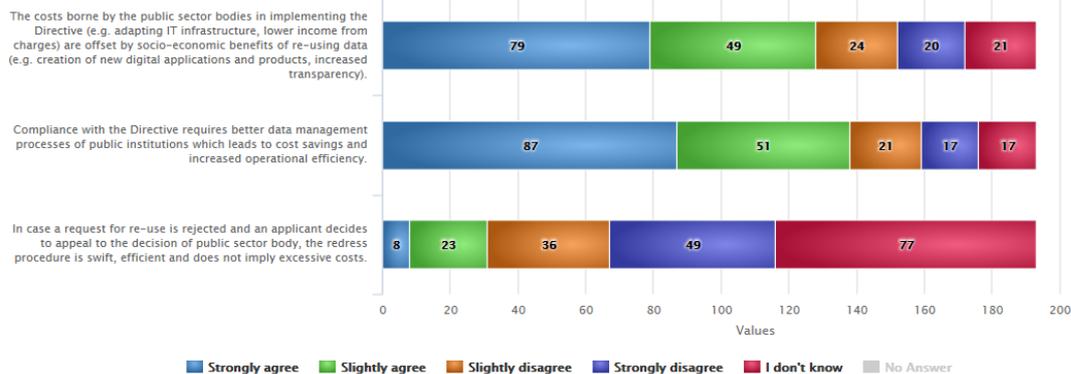


Figure 9 – Relevance of the PSI Directive

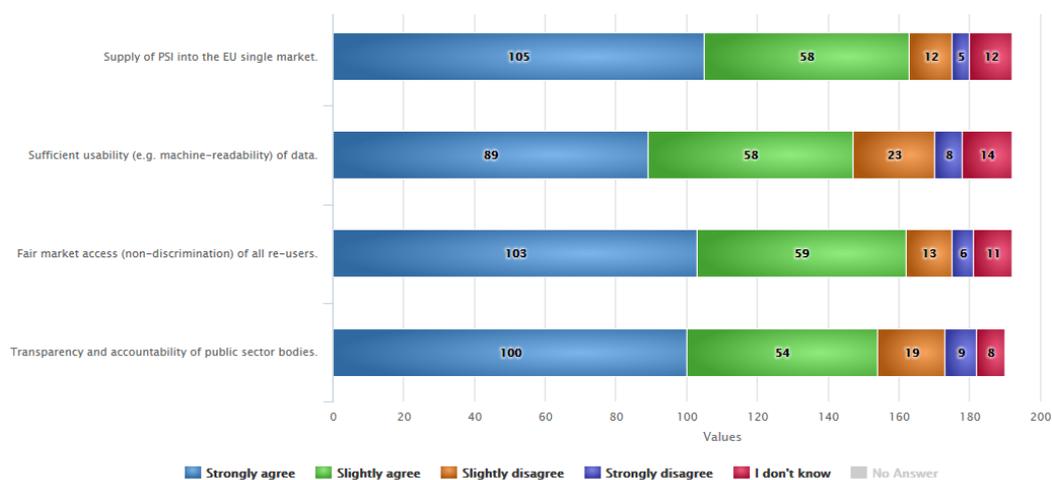


Figure 10 – Potential simplification of the PSI Directive

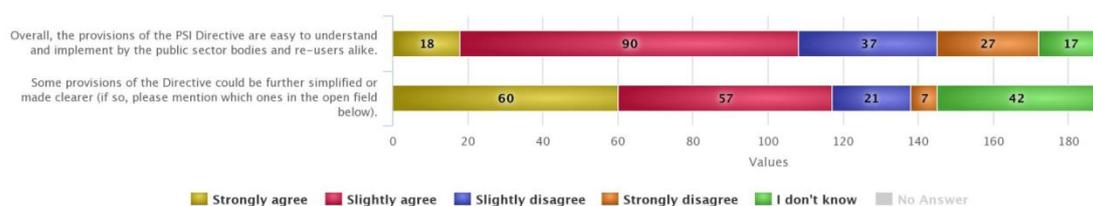


Figure 11 – Existing barriers to the re-use of data (more than one option could be chosen)

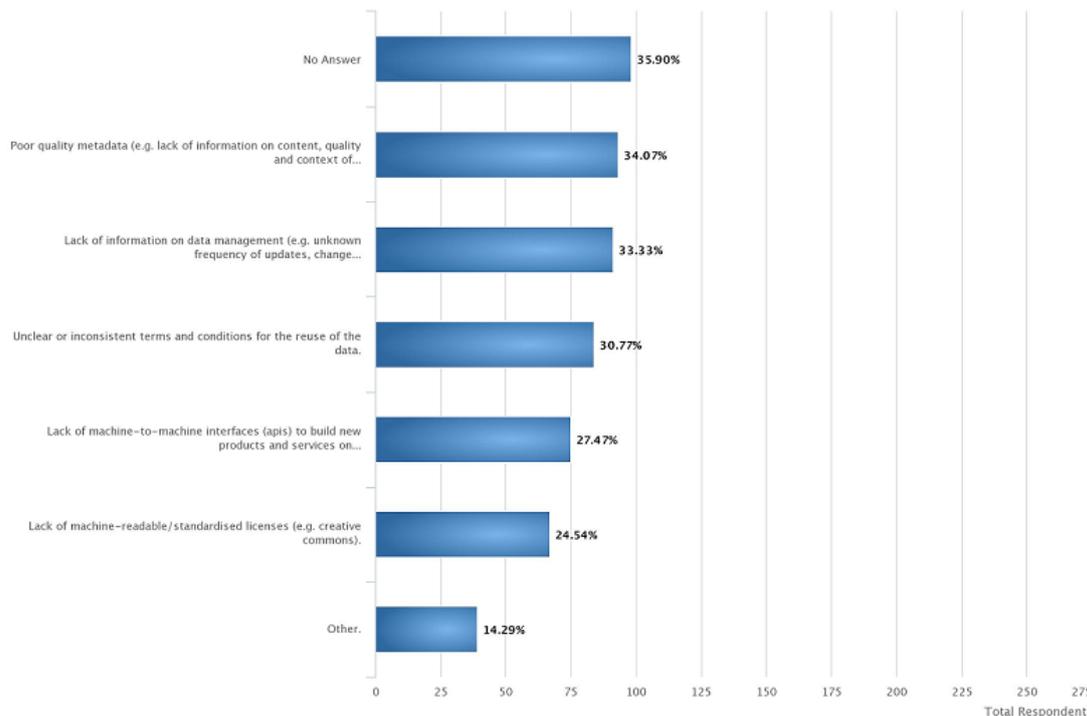


Figure 12 – Potential changes to the charging rules

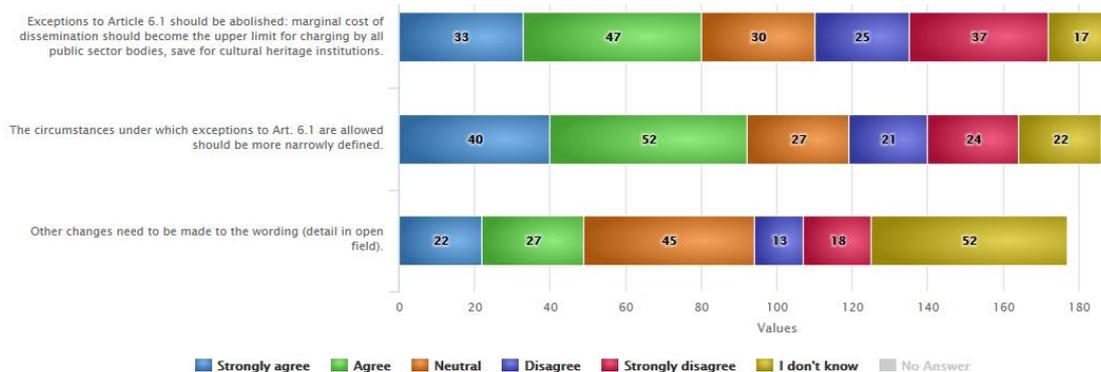


Figure 13 – Conditions to be applied for access to data to be authorised

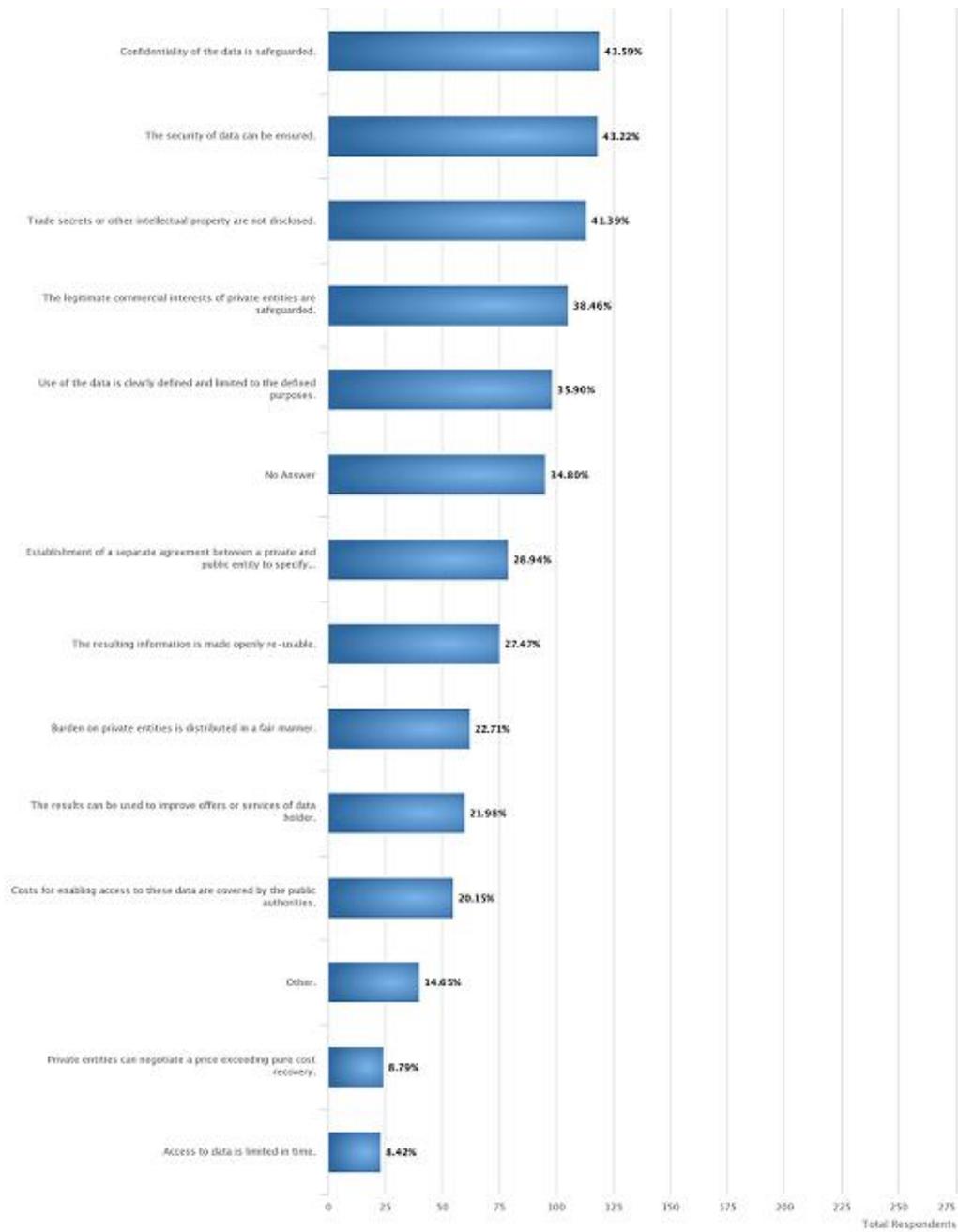


Figure 14 – Should access to data coming from private sector entities and its use by public authorities for reasons of public interest be allowed?

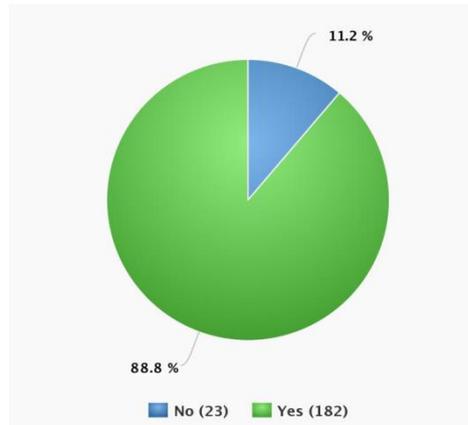
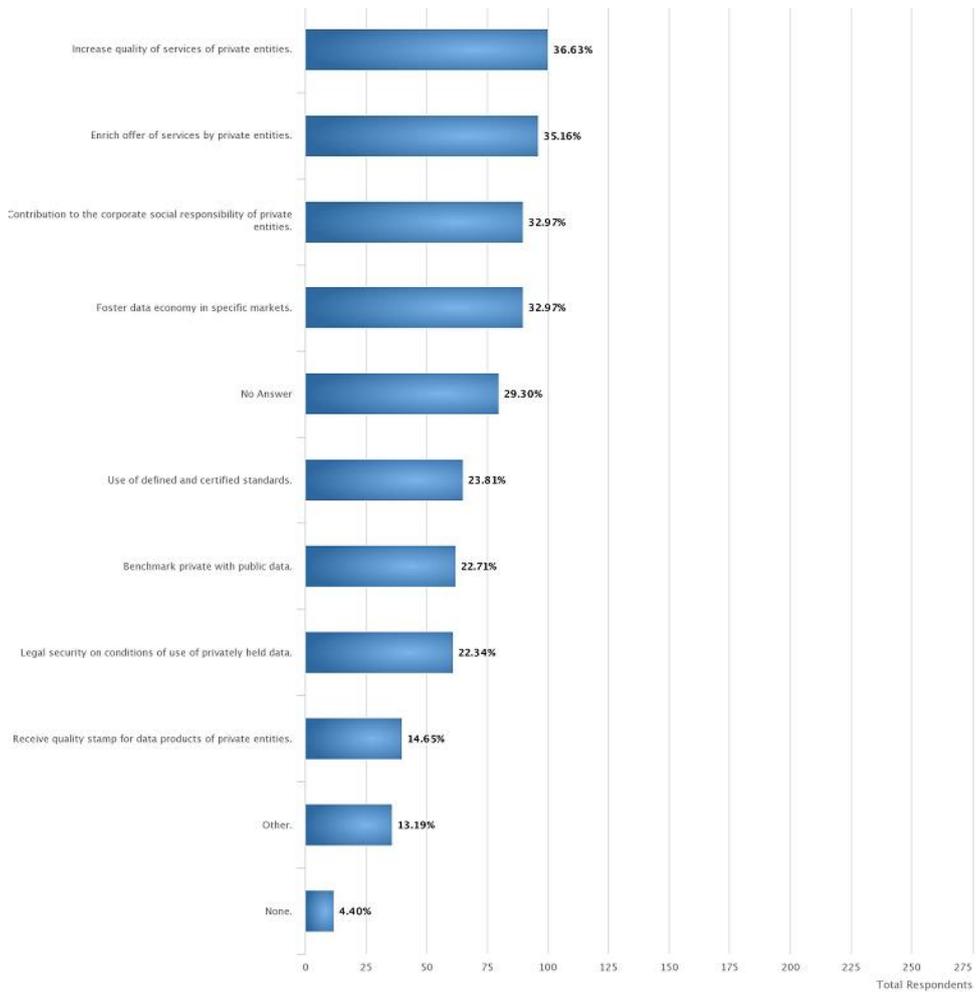


Figure 15 – What would be the possible motivation or incentives for data sharing of public interest with public authorities (more than one option could be chosen)



ANNEX 4: WHO IS AFFECTED AND HOW?

1. PRACTICAL IMPLICATIONS OF THE INITIATIVE

The proposed revision of the PSI Directive will have a range of practical implications for different groups of stakeholders.

Companies in general, and SMEs and start-ups in particular, will gain easier access to a wider pool of public sector data, which would otherwise be difficult or impossible to obtain due to a number of obstacles addressed by the proposal. They will benefit from the initiative directly by being able to reuse and recombine these different types of data and use it as the basis for a wide range of innovative services. Public data is an asset particularly sought after by innovative SMEs and start-ups due to the relatively modest initial capital investment necessary to launch a data-driven business. In 2017, around 727,000 companies are estimated to have benefitted from the PSI Directive (an increase of 56% since 2010). It is expected that by 2030 this number will increase by around 1.4 million to reach around 2.2 million.¹⁹¹

Consumers will benefit indirectly from the initiative, since they will have a wider choice of better and cheaper digital products developed on the basis of public sector data. On a more global level, they also stand to gain from a whole range of societal and environmental benefits generated by products and services based on open data, for example in such areas as public transport, healthcare and energy. Citizens will also see noticeable improvements as regards access to information, social inclusion and civic participation. At present the provisions of the PSI Directive already affect the majority of citizens in the EU in some way, whether it is through services available on their smartphones or through more accurate weather reports they get on the radio or TV. It is expected that by 2030 the impact of services based on PSI will be so pervasive that almost all citizens in the EU will be directly and indirectly impacted by the provisions of the PSI Directive.¹⁹²

Public sector bodies will be affected in two ways. On the one hand, some will be covered by new obligations, which will necessitate a certain compliance effort. On the other hand, by virtue of also being consumers of public sector data, they will be able to access and reuse the wider pool of available data to ensure a more cost-effective public service delivery and better-targeted policy making. It is important to bear in mind, however, that as far as the trade-off between the new obligations and benefits is concerned, the large majority of public sector bodies will experience only the positive effects of the initiative, and no negative impact. Mostly larger public sector data-holders (e.g. mapping agencies,

¹⁹¹ Impact Assessment support study, Deloitte, 2018, SMART 2017/0061.

¹⁹² Ibid.

meteorological institutes, and large cities) will be affected by the charging provisions, the API obligations and the non-exclusivity requirements.

For public undertakings and research organisations, practical implications of the initiative will also be twofold. In certain cases, these bodies will be expected to comply with some new requirements, but at the same time they will also benefit from the proposed changes.

Public undertakings will be able to use more data that will become available, while the products created on the basis of their own open data by third parties will improve the quality of the services they provide to their customers. As for new obligations, only those public undertakings will be affected that are trading information that is generated as a side-product of their main activity. They will have to comply with obligations on transparency, non-exclusivity and non-discrimination, and will have to respect the maximum charges for the data, based on recovering full costs plus a reasonable return on investment.

Research organisations will generally benefit from the wider reusability of research data, which is widely considered to be a positive driver for the development of science. As for obligations, the initiative will have some consequences for the operation of open science repositories that may have to adapt their licencing conditions and to check whether there are third party rights on any of the datasets. The way in which individual research institutions will be affected will depend on the way in which Member States will shape their research data policies.

2. SUMMARY OF COSTS AND BENEFITS

I. Overview of Benefits (total for all provisions) – Preferred Option		
Description	Amount	Comments
Direct benefits in 2027/2028		
Impact on the direct economic value of PSI	44 billion EUR - 30% more than in the baseline scenario (<i>Preferred option: 194 billion EUR, Baseline: 150 billion EUR</i>)	
Impact on the estimated number of persons employed in the area of PSI	191,000 - 40% more than in the baseline scenario (<i>Preferred option: 709,000, Baseline: 518,000</i>)	
Impact on the cost of opening up PSI	Cost 1.4 billion EUR (33%) lower than in the baseline scenario (<i>Preferred option: 2.7 billion EUR, Baseline: 4.1 billion EUR</i>)	Costs reduction linked to increased efficiency gains achieved through greater proactivity and higher automation of opening up PSI
Indirect benefits by 2030		
Impact on the total economic value of PSI	214 billion EUR - 32% more than in the baseline scenario (<i>Preferred option: 871 billion EUR, Baseline: 657 billion EUR</i>)	The concept of total economic value of PSI includes both direct and indirect economic value
Impact on the indirect economic value of PSI	170 billion EUR - 33% more than in the baseline scenario (<i>Preferred option: 678 billion EUR, Baseline: 511 billion EUR</i>)	

II. Overview of costs – Preferred option							
		Citizens/Consumers		Businesses		Administrations	
		One-off	Recurrent	One-off	Recurrent	One-off	Recurrent
Dynamic data/APIs	Direct costs	No costs	No costs	No costs	No costs	Approx. 50,000 EUR (per organisation / application) ¹⁹³	Max. 56 million EUR (for all EU Member States per year) for the 14 key high-value datasets ¹⁹⁴
	Indirect costs	No costs	No costs	No costs	No costs	Training costs for public sector officials: Up to 300,000 EUR per Member State	
Charging	Direct costs	No costs	No costs	No costs	No costs		
	Indirect costs	No costs	No costs	No costs	No costs		Approx. 150 million EUR (for all EU Member States per year) ¹⁹⁵
Data from public undertakings in transport and utilities sector	Direct costs	No costs	No costs	No costs	No costs		Approx. 20,000 EUR per public undertaking per year ¹⁹⁶
	Indirect costs	No costs	No costs	No costs	No costs		
Research data	Direct costs	No costs	No costs	No costs	No costs		Approx. 10,000 EUR administrative and legal cost (licencing issues, checking third party rights etc.) per open science repository, i.e. 15 million EUR for all Member States per year ¹⁹⁷
	Indirect costs	No costs	No costs	No costs	No costs		
Non-	Direct costs	No costs	No costs	No costs	No costs		Max 4.5 million

¹⁹³ An example cost of implementing an API for a medium sized application. However, this cost would be optional, since there would be no hard obligation to use APIs for most datasets. A hard obligation to use APIs would only apply to a certain number of high-value datasets.

¹⁹⁴ The maximum cost of making available 14 categories of high-value datasets in all Member States via APIs (see p. 28 of the Impact Assessment)

¹⁹⁵ It is estimated that by eliminating some of the exceptions to the marginal cost principle and making some of the high-value datasets free by default, the overall loss of income for public sector bodies will be approximately a half of the amount that would they would lose if charging was abolished entirely (300 million EUR – see. p. 30 of the Impact Assessment).

¹⁹⁶ This would only apply to those public undertakings that are trading information that is produced as a side-product of their main activity.

¹⁹⁷ There are approximately 1,500 open science repositories currently in the EU.

II. Overview of costs – Preferred option							
		Citizens/Consumers		Businesses		Administrations	
		One-off	Recurrent	One-off	Recurrent	One-off	Recurrent
exclusivity							EUR (for all Member States per year) ¹⁹⁸
	Indirect costs	No costs	No costs	No costs	No costs		

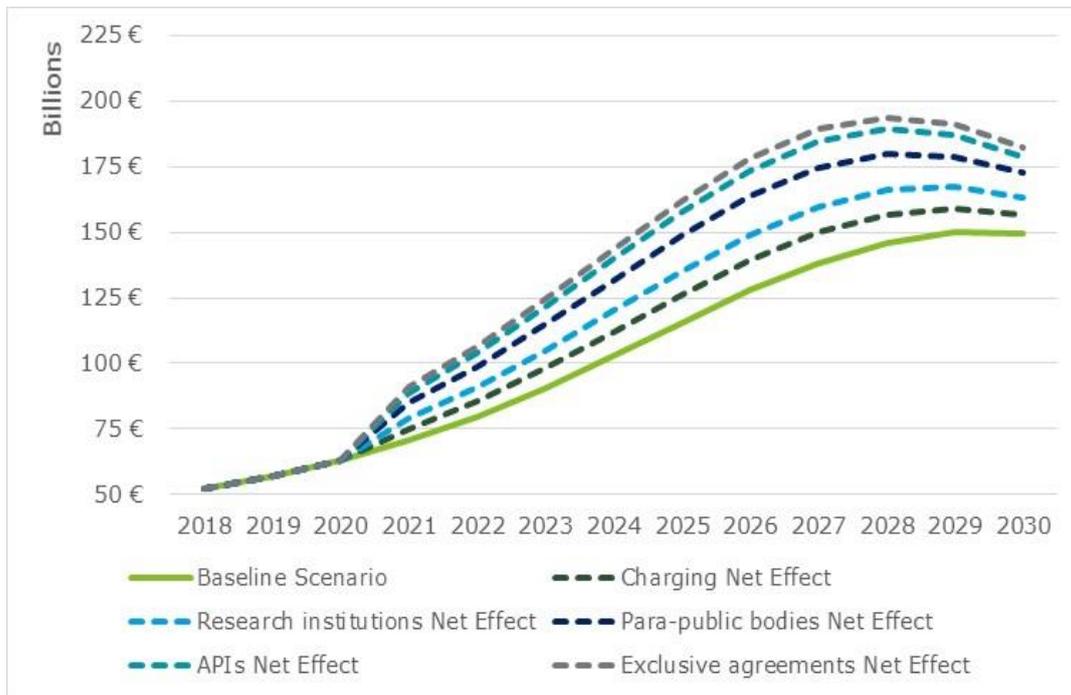
The figures cited in the table illustrate the costs under the preferred option in relation to its specific elements for different types of stakeholders. They are, however, not based on the quantitative model developed as part of the Support study by Deloitte but rather exemplify the overall costs estimated with the model and break them down into more specific aspects.

The overall methodology used by the study to estimate the baseline scenario, as well as the impacts of the policy options are provided in Annex 8.

The underlying model is based on comprehensive data from public sources multiplied with assumptions based on data from primary and secondary sources in order to provide an overall EU28 estimate (e.g. the economic value). Based on the Compound Annual Growth Rate of the Eurostat data, as well as a specific S-curve growth model (i.e. either linear growth or curve modelling depending on theory and data), the data was extrapolated until 2030. Specific assumptions have been used to break down the data by specific metrics, such as economic value by sector, or government revenue from PSI by level of government.

The following graph depicts the individual contributions of the measures under the preferred policy option to the overall direct economic value. The direct economic value should be seen as illustrative representation of the individual estimates made in the support study. The dotted lines represent the individual economic value that each element of the Policy Package adds to the baseline scenario to achieve the overall impact.

¹⁹⁸ Based on the assumption that max 20 cases per year per Member State would need to undergo an ex-ante check, which would require 40 man-days split between the public sector body concerned and the national competition authority (so max. 160,000 EUR per Member State per year). The number of such cases is moreover expected to decline over time, due to the preventive effect that the inclusion of the new legislative provision will have on the public sector bodies, as has been the case with the exclusive agreements targeted by article 11 of the Directive.



Source: Impact Assessment Support Study, Deloitte, 2018

The shares of contributions of the individual elements can be derived from the following table.

<u>Benefits:</u> Share of contributions to benefits of intermediate policy option	Policy Options	PO1	PO2
	Charging		23%
Research institutions		20%	20%
Para-public bodies		29%	36%
APIs		20%	19%
Exclusive agreements		10%	9%
<u>Costs:</u> Share of contributions to costs of intermediate policy option	Policy Options	PO1	PO2
	Charging		23%
Research institutions		20%	20%
Para-public bodies		29%	36%
APIs		20%	19%
Exclusive agreements		10%	9%

Source: Impact Assessment Support Study, Deloitte, 2018

It was not part of the model to estimate how costs and revenues are distributed among actors across the data value chain. From a qualitative perspective, in general, no single

type of stakeholders, however, is affected by costs or benefits *only*. Public sector bodies, businesses, researchers, as well as citizens all face costs and benefits alike – either monetary or with respect to more qualitative aspects such as transparency, opportunity, or freedom of choice. SMEs are of particular concern as they tend to be affected disproportionately by costs. The support study has shown the types of stakeholders profiting strongest from PSI are public authorities, as well as industry businesses and companies active in trade and transport. At the same time, however, these are also affected strongest by the costs related to PSI.

As part of the support study, estimates have not been differentiated across administrative, technical, and opportunity costs, as well as redress and litigation costs. Although the underlying data (gathered from interviews in several Member States, feedback received during the workshops, as well as expert assumptions based on the available material), provided insights into how costs are structured, the overall data based was not found to be sufficiently robust enough to provide detailed estimates per type of cost in a quantitative way.

ANNEX 5: EVALUATION REPORT

(see Staff Working Document SWD(2018) 145)

ANNEX 6: THE EVOLUTION OF THE DIRECTIVE ON THE RE-USE OF PUBLIC SECTOR INFORMATION

1. Key elements of the original PSI Directive

The 2003 Directive was built around two main pillars: ensuring the cheap availability of public sector data to third parties whilst setting a level playing field between public bodies and private competitors in the information market. It also included a set of principles on which the PSI re-use in Europe has since been based:

- *Non-discrimination*: any conditions for the re-use of public sector documents have to be non-discriminatory for comparable categories of re-use
- *No cross-subsidies*: when public institutions commercialise added-value information products based on their own data, they have to apply the same conditions and charges to their own re-use as to their private sector competitors.
- *Prohibition of exclusive arrangements*: Save for tightly-defined exceptions, re-use of public sector documents has to be open to all potential actors in the market.
- *Upper limit for charging*: to avoid the danger of public sector bodies abusing their monopoly power as the only producer of specific data, all charges for re-use of such data should not exceed the costs incurred, whilst allowing the recovery of a 'reasonable return on investment'.
- *Transparency*: conditions and standard charges for re-use should be pre-established and published. The re-user should also be informed of any available redress procedures.

2. Key elements introduced by Directive 2013/37/EU

The revision of the Directive was preceded by a consultation and an impact assessment which indicated the persistence of obstacles to wide re-use of data and the diverging practices with regard to the availability of data for re-use, the licensing conditions and the charges applied. An update of the legal framework proved therefore necessary. Its main elements are the following:

- A default rule of charging for re-use of documents based on the marginal costs of their reproduction and dissemination and subject to limited exceptions only;
- A general rule that all accessible documents are by default re-usable for any purpose, commercial or non-commercial, unless protected by third-party copyright and without prejudice to data protection legislation;

- Exclusive arrangements are generally prohibited. Exceptions can be accepted in a specific case of digitisation agreements but those should be limited in time and safeguard the public domain character of the cultural items.
- Scope of the Directive is widened so as to cover content held by museums, libraries and archives but is only re-useable if the institutions allow its re-use.
- Public sector bodies must be transparent about the licensing conditions and charges they apply and about the factors taken into account when calculating charges.

ANNEX 7: OPEN DATA IN AN INTERNATIONAL CONTEXT

Over the last few decades, a range of countries have put in place their specific open government data policies all over the world. Some of these national or cross-border open data policies have already borne their fruits in the form of measurable socio-economic benefits.

In a global context, nearly eight in 10 citizens (78%) say that open government data can help improve lives of citizens. The largest number of people who consider open government data to have an important role in improving their lives are from India (93%), followed by Taiwan (84%), Singapore (83%) and Mexico (80%). 37% of respondents think that open government data makes governments more transparent. This sentiment is particularly strong in Mexico (52%), Taiwan (51%) and Finland (47%).¹⁹⁹

A 2013 study²⁰⁰ made an attempt to estimate the annual economic potential linked to open data and calculated USD 3 trillion globally. The benefits taken into account during the calculation included increased efficiency, development of new products and services, and consumer surplus (cost savings, convenience, better-quality products) as well. The same study estimated that the potential economic benefits could be divided mainly between the United States (USD 1.1 trillion), Europe (USD 900 billion) and the rest of the world (USD 1.7 trillion).

G8 Open Data Charter and International Open Data Charter

In June 2013, G8 leaders signed an Open Data Charter²⁰¹ which set out five strategic principles that all G8 members are to act on. These include an expectation that all government data will be published openly by default, alongside principles to increase the quality, quantity and re-use of the data that is released. G8 members have also identified fourteen high-value areas – from education to transport, and from health to crime and justice – from which they will release data.

On the basis of the G8 Open Data Charter, an International Open Data Charter²⁰² has been developed, with the support of a number of international organizations including the World Bank Group, United Nations, Organization of American States (OAS), Organization for Economic Co-operation and Development (OECD), Inter-American Development Bank (IADB), African Development Bank Group, Economic Commission for Latin America and the Caribbean (ECLAC/CEPAL). The Charter has so far been adopted by 17 national and 35 local or subnational governments across the world.

¹⁹⁹ Open Government Data: Assessing demand around the world - A report from The Economist Intelligence Unit, 2017, p. 9

²⁰⁰ Open Data: Unlocking innovation and performance with liquid information, 2013, p. 6

²⁰¹ <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

²⁰² <https://opendatacharter.net/>

All G8 members have started implementing the commitment to work towards the progressive publication of datasets in accordance with the principles of “open by default” and “quality and quantity”. Other countries have made similar pledges all over the world.

Such benefits could also be realised at EU level. In its commitments to implement the G8 Open Data Charter²⁰³, the EU also committed to the publication by the EU Member States of core and high-value datasets in line with the G8 principles. Creating a list of such datasets, which would be open and free by default, has the potential to facilitate the emergence of a new range of pan-European data products and services developed on the basis of mutually complementary cross-border datasets.

OECD

In 2018, the OECD expects to submit for approval a draft Council Recommendation on the General Principles for Enhanced Access to Data. In the run-up to the draft Council Recommendation, an ad-hoc Joint Steering Group will deepen the understanding of issues such as data openness, trade-offs in data utilisation, policies to address market failures via enhanced access to data (e.g. interoperability, standards and regulations on privacy and IP). This new OECD Recommendation will combine, amongst others, the OECD Recommendation on Access to Research Data and the OECD PSI Recommendation.

Open Government Partnership (OGP)

Open data is an important component of the Open Government Partnership (OGP) initiative²⁰⁴ whose primary goal is to "increase transparency about what governments are doing – which promotes accountability and good governance, enhances public debate, helps to combat corruption, and supports the design and delivery of better public services". OGP membership increased from 8 members in 2011 to over 70 countries, including the majority of EU member states.

United States of America

In the United States, the Federal Government has been pioneering Open Data policies already under the Clinton administration placing federal information in the public domain. In 2013, President Barack Obama issued a Memorandum on "Open Data Policy – Managing Information as an Asset"²⁰⁵ and an executive order²⁰⁶ to all federal ministries and agencies to make their information easily and readily available on machine readable

²⁰³ <https://ec.europa.eu/digital-single-market/en/news/eu-implementation-g8-open-data-charter>

²⁰⁴ www.opengovpartnership.org

²⁰⁵ *Memorandum for the Heads of the Executive Departments and Agencies*, May 9, 2013:

<https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

²⁰⁶ *Executive Order – Making Open and Machine Readable the New Default for Government Information*, May 9, 2013: <https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>

form. This memorandum has been followed by concrete plans of actions²⁰⁷ and a strong political backing by "a data-driven president"²⁰⁸.

The Digital Accountability and Transparency Act of 2014 (DATA Act) advocates an opening of some governmental data. It focuses on making [federal expenditures](#) more easily accessible and transparent by, for example, establishing common standards for financial data. A related policy document²⁰⁹ instructs agencies to manage their data as an asset and, wherever possible, release it to the public in an open, discoverable and usable way. The US administration has recently started an IT modernisation effort which shifts their policy to a service-based model (i.e. rethinking how government delivers services).

Japan

In Japan, under the Basic Act on the Advancement of Utilising Public and Private Sector Data of December 2016, *"The State and local public entities are to implement necessary measures to enable citizens to easily use Public and Private Sector Data held by themselves via the Internet or any other advanced information and telecommunications network while ensuring that the rights and interests of individuals and corporations, national security, etc. are not damaged"*. In some sectors such as electric power and gas, certain measures are taken to allow access to the information they collect in their legal monopoly. The Japanese open data portal publishes data sets considered of high value such as transport timetables, education information, energy and government accountability and democracy.

Australia

The principles on open public sector information form part of a core vision for government information management in Australia.²¹⁰ These principles rest on the premise that public sector information is a national resource and a free flow of information between government, business and the community can stimulate innovation in the economy. The principles will be applied with objectives of the Freedom of Information Act 1982 (FOI Act)²¹¹.

The Australian open data portal provides data sets such as public transport timetables, international trade data, company register and government spending (in machine readable

²⁰⁷ *US Open data Action plan 2014-2015*

http://www.whitehouse.gov/sites/default/files/microsites/ostp/us_open_data_action_plan.pdf

²⁰⁸ *The Power of Open data to Solve Old Problems in New Ways*, blog post at opening session of the 3rd International Open Data Conference 2015: <http://opendatacon.org/the-power-of-open-data-to-solve-old-problems-in-new-ways/>.

²⁰⁹ <https://project-open-data.cio.gov/policy-memo/>

²¹⁰ <https://www.oaic.gov.au/information-policy/information-policy-resources/principles-on-open-public-sector-information>

²¹¹ <https://www.legislation.gov.au/Series/C2004A02562>

format). It also provides hosting for tabular, spatial and relational data with hosted APIs and the option for agencies to link data and services hosted by other government sources.

The government in Australia has committed to identifying National Interest Datasets (NIDs). According to the report "Building a reformed data system"²¹², significant improvements could come from aggregating data across the States and Territories in different domains: health, education, social welfare, child support, aged care, and linking them with datasets from other fields, such as the population census, taxation, employment, business ownership, telecommunications, private health insurance or housing. In principle, all non-sensitive datasets in fields where there are growing opportunities and capability would be opened up and released, depending on resources and sectoral demand.

Canada

The government of Canada issued the Directive on Open Government in October 2014, where it establishes an "open by default" directive for government publications and data. In 2016, the Third Biennial Plan to the Open Government Partnership was released with commitments on open data, i.e. to enhance access to information, streamline requests for personal information, expand and improve open data, provide and preserve open information, definition of an approach for measuring open government performance, development of open government skills and enhance access to culture and heritage collections.

²¹² <http://www.pc.gov.au/inquiries/completed/data-access/report>

ANNEX 8: ANALYTICAL METHODS

1. OVERALL METHODOLOGY OF THE STUDY

The study to support the review of Directive 2003/98/EC on the re-use of public sector information (SMART 2017/0061) was carried out in three Phases (inception, data collection, and analysis). With regard to the collection of data, the following key methodological and analysis tool were implemented:²¹³

- Strategic interviews;
- Desk research;
- Interviews with stakeholders at the EU and national levels;
- Legal data collection;
- Workshops with public sector practitioners and academics, as well as with PSI re-users and data economy players;
- Online surveys with public authorities, incl. public cultural institutions, education and research bodies, as well as the re-user community; and
- Analysis of the public consultation²¹⁴ carried out by the European Commission.

An overview of the all these data collection tools is provided below.

Tool	Details
Strategic interviews	Strategic interviews have been carried out with: <ul style="list-style-type: none">• European Parliament;• DG JUST;• DG CNECT;• DG GROW;• DG COMP;• DG SANTE• Publication Office;• DG ENER;• DG MOVE;• DG RTD;• JRC;• EDPS;• EuDECo²¹⁵;

²¹³ The data collected through the implementation of the above tools will be analysed through the application of the following analytical methods and processes: Legal analysis; Triangulation; Analysis of costs and benefits; and Multi-Criteria-Analysis.

²¹⁴ See: https://ec.europa.eu/info/consultations/public-consultation-review-directive-re-use-public-sector-information-psi-directive_en

Tool	Details
	<ul style="list-style-type: none"> • Europeana; • Kennisland²¹⁶; and • European University Association.
Desk research	<p>Desk research was a continuous exercise throughout the study and informed the drafting of e.g. surveys and questionnaires, as well as the draft reporting on findings. The desk research is structured around the indicators that are included in the Analytical Framework.</p> <p>For the purpose of the legal analysis of the PSI Directive in the Member States, a research questionnaire was used.</p>
Interviews	<p>The study team completed the interviews with stakeholders at the national level. Interviews were conducted with the following types of stakeholders:</p> <ul style="list-style-type: none"> • Ministries and public bodies in charge of public sector data regulation and monitoring of implementation; • Public bodies falling into the scope of the Directive; • Museums, libraries and archives; • Public undertakings in the transport and utility domains; • Universities and research centers; and • Re-users associations.
Workshops	<p>Representatives of the study team attended the PSI group meeting that was held on 15 November 2017. The team members provided a short presentation, took detailed minutes of the discussion, as well as of its key results and conclusions.</p> <p>The first workshop with research bodies was held on 14 December 2017 with more than 30 stakeholders from the area of research, publishing, open data advocacy and the private sector. As part of two panel sessions, the attendees engaged in a lively, open and fruitful discussion about the potential and challenges related to opening up research data. Members of the study team contributed with short presentations, and provided the Commission with detailed minutes of the discussion, as well as of its key results and conclusions.</p> <p>A second workshop involving data providers and re-users was held on 18 January 2018. The purpose of the workshop was to discuss the technical and non-technical barriers for opening up and reusing the public sector data and the possible ways to improve the situation. Members of the study team moderated the event and took minutes of the discussions.</p>
Online surveys	<p>The study team prepared and disseminated three online surveys. The stakeholders contacted are listed in the First Interim Report of the study.</p> <p><i>Survey on cultural data</i>²¹⁷</p>

²¹⁵ <http://data-reuse.eu/>

²¹⁶ <https://www.kl.nl/en/people/paul-keller/>

Tool	Details
	<p>This survey was focused on gathering information for the evaluation and the analysis of the effects of the new provisions that were added to the Directive in 2013 with respect to data held by museums etc.</p> <p>Survey on research data²¹⁸</p> <p>This survey investigated the effects and opportunities linked to the opening up of research data in the framework of publicly funded research. It targeted universities and research centres.</p> <p>Survey for the re-user community²¹⁹</p> <p>This survey focused on the beneficiaries of the PSI Directive and mainly concerned:</p> <ul style="list-style-type: none"> • The provisions on charging (limited to marginal costs) and the effects of possible modifications to this provision; • The possibility of opening up data held by public undertakings in the transport and utility domains or produced in relation with public contracts; • The possibility of opening up research data; • The question of APIs and dynamic data. <p>Despite significant efforts in dissemination, the response to the survey was limited mainly due to the public consultation on the PSI Directive review going on in parallel. As a consequence, the survey results could be used as indicative evidence and without the possibility to broadly generalise their insights.</p>
Public consultation analysis	<p>A public consultation on the review of the Directive on the re-use of Public Sector Information (PSI Directive)²²⁰ was carried out from 19 September 2017 to mid-December 2017.</p> <p>The study team analysed the qualitative answers to the public consultation under the instructions from the European Commission and provided a preliminary report on this analysis.</p> <p>The questions included in the public consultation were taken into account when the interview guides and online surveys were prepared in order to avoid duplication. The results of the public consultation were also used to support the analysis all through the deliverables of the study.</p>
Legal data collection	<p>The legal assessment of the implementation of the PSI Directive was performed in two steps: 1) data collection at national level by a network of legal experts and 2) legal analysis by the legal team of time.lex and Spark Legal Network.</p> <p>Firstly, data was collected through desk research, by a team of national legal experts covering the ten selected EU Member States (DE, EE, EL,</p>

²¹⁷ See: <https://ec.europa.eu/eusurvey/runner/culturaldataPSI>

²¹⁸ See: <https://ec.europa.eu/eusurvey/runner/ResearchSurvey>

²¹⁹ See: <https://ec.europa.eu/eusurvey/runner/PSIReusersSurvey>

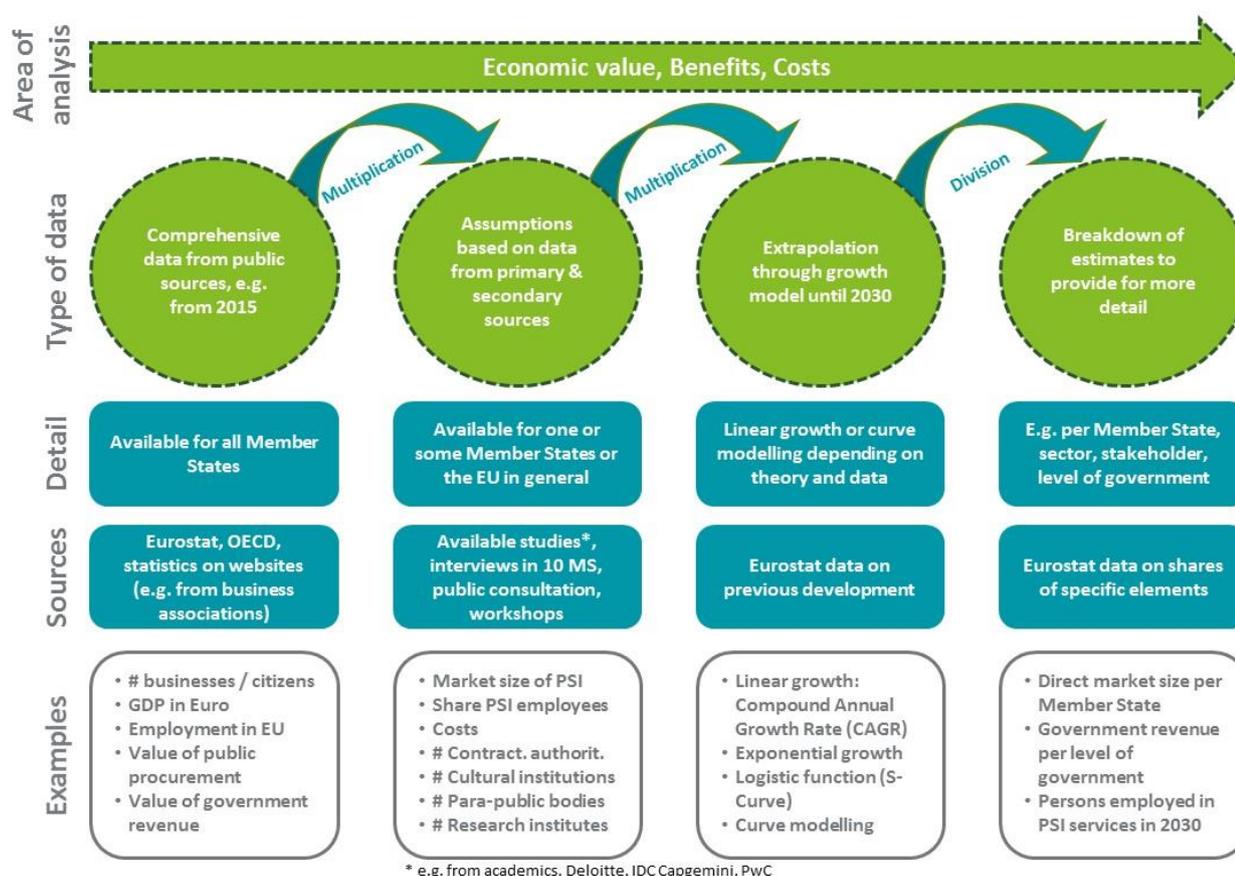
²²⁰ See: https://ec.europa.eu/info/consultations/public-consultation-review-directive-re-use-public-sector-information-psi-directive_en

Tool	Details
	<p>FR, IE, IT, NL, PL, SE, SI). This work was executed on the basis of a detailed research protocol, including a research questionnaire. The legal experts were asked to complete the research questionnaire collecting data on 1) the national transposition strategies, 2) potential legal issues arising from the implementation of the latest changes to the PSI Directive and, 3) the interplay and overlap between the PSI Directive with other relevant EU legal instruments.</p> <p>Having carried out the legal data collection exercise, the national experts submitted their reports to the legal team, which reviewed them to ensure completeness and consistent quality of the national legal reports.</p> <p>As the national legal reports were completed and quality checked, the legal team fed the results into a comparative document summarising the key issues identified during the legal data collection. This internal document facilitated the comparison between the ten Member States and the identification of similarities, problems, gaps, overlaps and diverging implementations. Thus, it allowed all results to be bundled into one document, and prepared the ground for the first comparisons to be made, which in turn fed into the legal analysis.</p> <p>After having drafted a preliminary version of the legal analysis, the legal team held a two-day internal brainstorming meeting in Brussels. The meeting focused on discussing the outcomes of the data collection at national level, identifying problems/barriers and best practices. Subsequently, and linking with the abovementioned issues, barriers and best practices, the legal team assessed the potential policy options to be considered from both a legal and practical point of view. The results of the meeting were incorporated into the legal analysis.</p>

Source: Impact Assessment Support Study, Deloitte, 2018

2. THE MODEL USED IN THE STUDY

The overall approach used to develop the estimates provided in the support study, e.g. concerning the economic value and costs related to PSI, are based on the logic visualised in the figure below.

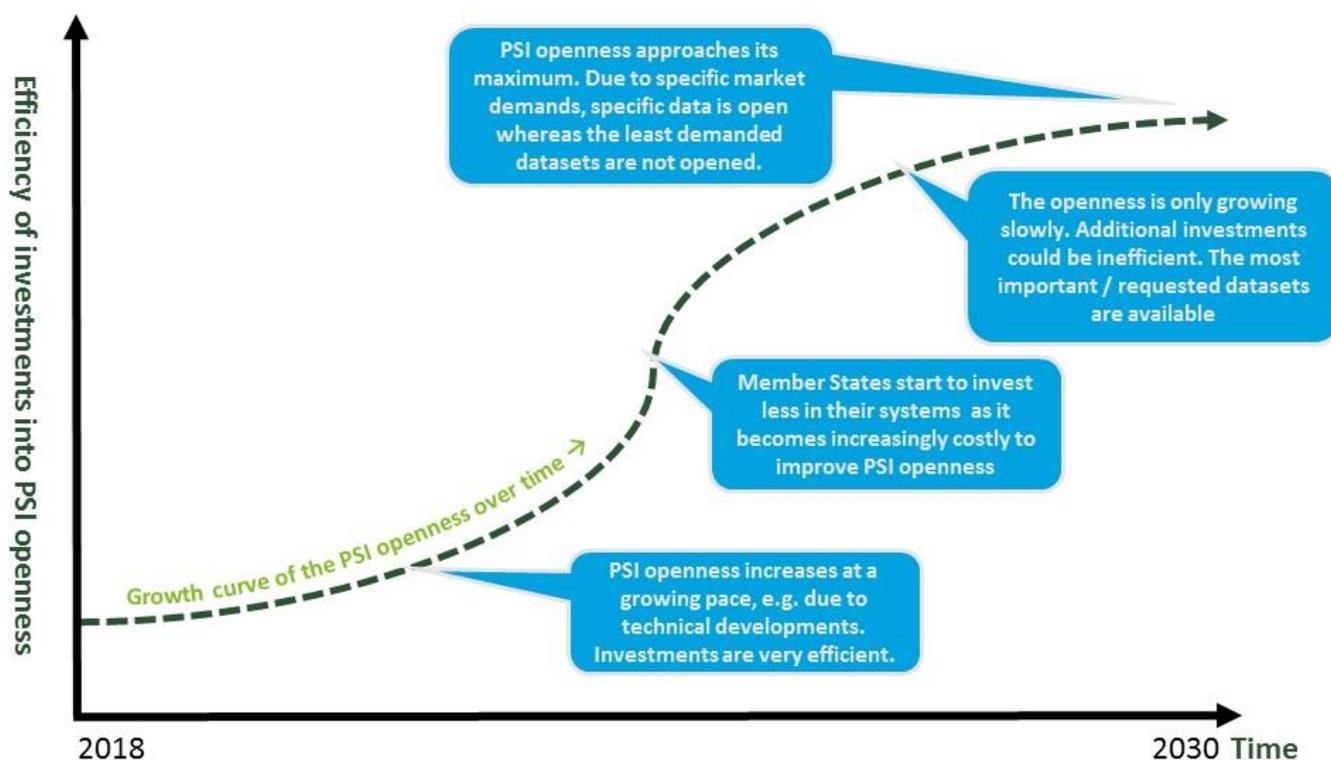


Source: Impact Assessment Support Study, Deloitte, 2018

As can be seen above, comprehensive data from public sources, e.g. the number of businesses available for multiple years via Eurostat, was multiplied with assumptions based on data from primary and secondary sources, e.g. in relation to the market size of PSI in different Member States or third countries in order to estimate the overall (direct and indirect) EU28 economic value. Based on the Compound Annual Growth Rate of the Eurostat data, as well as a specific S-curve growth model (i.e. either linear growth or curve modelling depending on theory and data, see further below), the data was the extrapolated until 2030. Finally, specific assumptions have been used to break down the data by specific metrics, such as economic value by sector, or government revenue from PSI by level of government.

The growth model has been developed to project the development of the estimates in the future, especially the expected growth in the PSI market under different option packages.

The growth model used in this study is based on the S-curve concept which is widely applied in macro-economic modelling. The concept is visualised below.



Source: Impact Assessment Support Study, Deloitte, 2018

Within this concept, the growth of a certain set of data over time, e.g. the economic value of PSI, increases over time up to a point at which the growth rate eventually declines and the data only grows marginally (the curve “flattens out”).

Such a curve can be modelled with a logistic function. The general formula used for the logistic function is the following.

$$P(t) = \frac{P_0 * e^{r-t}}{1 + P_0(e^{r-t} - 1)/P_{MAX}}$$

The formula contains the following elements:

- P represents the data point at a given time t ;
- P_0 represents the data point today;
- P_{MAX} represents the data point that can reasonably be achieved until 2030; and
- r represents the parameter by means of which the data point is expected to increase annually; and
- e is the mathematical that is the base for the natural logarithm (‘Euler’s number’).

The curve can be modelled in such a way that it does, however, not necessarily resemble the S-curve in a given period of time, e.g. by mathematically stretching its development

over a timeframe that exceeds the scope of this study (i.e. the flat part of the curve will not be reached by 2030).

Towards the end of its lifecycle, each market, industry, product, or service ultimately reaches a saturation phase. This means that marginal investments are becoming increasingly inefficient (i.e. every additional euro of investment does not increase the return on investment) with increasingly less consumers purchasing the respective products or services.

With specific regard to PSI, this means that at some point in the future the economic value of PSI is expected to grow at a shrinking margin. Reasons for this can e.g. be that the most important data sources have been opened up and provided in a re-usable format. Moreover, in the saturation phase of the market, most consumers are expected to have purchased PSI-based goods and services, meaning that innovation can reasonably be expected to become more and more difficult. Thus, consumers may look for products and services based on alternative types of data or data sources outside the EU context. Therefore, the growth curve typically “flattens out” in the saturation phase of the market.

Given the complexity of and differences between each type of industry and business, granular information on costs and benefits is usually very scarce. This is a frequent issue in Impact Assessment studies as a crucial constraint for the assessment of existing problems and the impacts of the Policy Packages. Important constraints were:

- Only few stakeholders were able to provide the information in relation to the types of evidence necessary for the Commission; and
- The specific timing of this study was a challenge for the collection and analysis of (quantitative) evidence, as the 2013 changes to the Directive did not yet have full practical impact according to numerous stakeholders.

Thus, the model is based on a pragmatic, reasonable methodology, as well as proxies and the best data available. Moreover, expert assumptions / estimates concerning the expected level of e.g. costs and benefits (based on estimates and the use of ranges) were used after a peer-review process the included the Commission.

Hence, the analyses and estimates contained in the study should rather be seen as a contribution to the general debate around PSI than exact science. Thus, the estimates should be treated with caution.

The following table provides the key assumptions used in the model. They are based on the best data available from the interviews carried out as part of this study, as well as the available secondary literature and expert opinions.

Peer review of the assumptions used

A number of assumptions used for the quantitative model are based on expert judgment. This means that the data used in the underlying formulas is based on the best data available, challenged and refined (where necessary) by the experts of the

consortium for this assignment.

The assumptions used for the estimate of the impact of the Policy Options were subject to an internal, in-depth peer review process. As part of this process, different assumptions were introduced in the model to compare the different outcomes. The result of this sensitivity analysis was that the assumptions provided in the table above seem to be, at this stage, the most reasonable and pragmatic based on the best data available in relation to this specific subject.

Area of estimates	Specific assumption / Indicator	
Number of stakeholders affected		
CAGR	Contracting authorities	0.5%
	Cultural institutions	0.5%
	Para-public bodies	-1.0%
	Research institutes	0.5%
Economic value		
Direct economic value	2010	32,000,000,000 €
	2015	44,463,765,786 €
Share of economic value per sector	Agriculture	0.5%
	Industry	13.3%
	Construction	3.6%
	Trades & Transport	13.2%
	ICT	8.9%
	Finance and Insurance	8.1%
	Real estate	11.9%
	Professional services	11.0%
	Public administration	29.2%
	Arts, entertainment, recreation	0.5%
CAGR	PSI	6.8%
Ration Direct / Indirect Market Size	Lower bound	3.5
	Upper bound	3.78
Employment		
Shares of jobs	Share of Open Data jobs in Knowledge Intensive Activities	0.078%
	Share of PSI-related jobs of all Open Data jobs in	80%

Area of estimates	Specific assumption / Indicator		
	Knowledge Intensive Activities		
	Share of PSI jobs in Knowledge Intensive Activities	0.062%	
Cost savings + revenue			
Shares	Share of spending that could be saved	0.022%	
	Share of additional revenue	0.010%	
Costs			
Data from Polish stakeholders	IT equipment, programs and licences, computers and servers	1,173,245 €	
	Preparation of project, feasibility study	205,000 €	
	IT services, audits and tests, APIs	2,069,048 €	
	Legal services, translations, consulting	140,645 €	
	Salaries (experts working on open standards, trainers, coordinating portal design, partners' costs)	2,195,714 €	
	Other salary related costs	52,724 €	
	Training	268,095 €	
	Training material	25,789 €	
	Information and promotion	150,864 €	
	Total	6,281,123 €	
	Number of registers to which these costs relate	7	
	Life time of hard- and software (i.e. number of years over which costs are split)	7	
	Share of EU funding	84.67%	
Average cost per year per register	19,651 €		
Policy options			
<u>Benefits:</u> Share of contributions to benefits of intermediate policy option	Policy Options	PO1	PO2
	Charging	23%	18%
	Research institutions	20%	20%
	Para-public bodies	29%	36%
	APIs	20%	19%
	Exclusive agreements	10%	9%
<u>Costs:</u> Share of contributions to costs of	Policy Options	PO1	PO2
	Charging	23%	18%
	Research institutions	20%	20%

Area of estimates	Specific assumption / Indicator		
intermediate policy option	Para-public bodies	29%	36%
	APIs	20%	19%
	Exclusive agreements	10%	9%

Source: Impact Assessment Support Study, Deloitte, 2018

ANNEX 9: SUCCESS STORIES ON OPEN DATA

These stories have been produced to give an account of how open data is used by European companies and how they combine open data with other kinds of data to create value for the economy and innovative services or products for their clients.

Companies are using open data to exploit gaps across markets and identify business opportunities, create new products and services and develop new business models. Start-ups using open data often combine business information with the intention to create social, economic and environmental impact.

The success stories on open data shed light over the activities of the two main actors in the open data market: data re-users and data holders.

The data re-users covered in this annex are:

- **Commoprices:** A data-based service to buy goods at the best price (France)
- **Cropti:** Efficient and environmentally-friendly farming (Spain)
- **Datasparks (Derilinx):** Inspiring change through efficient energy consumption (Ireland)
- **Go OV:** Democratising public transport services (the Netherlands)
- **Green City Solutions:** Making clean urban air profitable (Germany)
- **Viomedeo:** Accelerating patients' access to medical innovation (Germany and Austria)

The data holders covered in this annex are:

- **The Estonian Road Administration:** a digital service (Estonia)
- **Liander:** Powering household energy conservation with open data (the Netherlands)
- **Transport for London:** The creation of an open data ecosystem (United Kingdom)
 - **Open data in other transport companies**

COMMOPRICES: A DATA-BASED SERVICE TO BUY GOODS AT THE BEST PRICE

Business opportunity

Many large companies and SMEs are unable to find independent, easy to access and affordable market data regarding price references on commodities (i.e. goods such as metals, plastics, cereals, cocoa, cheese). On top of that, if a company needs to follow price references of different commodities, they needed a number of accounts from different market data providers.

Goal of the company

The goal is to **simplify the way professionals follow commodities, building on open data.** [Commoprices](#) helps procurement and supply professionals to have reliable, independent and continuous information to buy commodities at the best price, minimise risks and make the right decisions. It reduces the risks and exposure to commodity price variations, and eventually brings more transparency and less volatility to the market. It offers more than 100K price references for commodities in a very wide range of sectors: agriculture, metals, plastics, chemicals, energy, etc.

Open data used

Open data on commodities from the French customs forms the backbone of Commoprices. It was the first source of commodity prices provided to their customers. These data give authority, independence and strength to Commoprices business intelligence platform as the data were primarily collected and validated by public institutions. Commoprices has also integrated customs open data (exports and imports) from other countries (e.g. Germany or Brazil) and has created maps to visualise geographical price references.

Other data used

Commoprices has also integrated other open and non-open datasets such as datasets from the WorldBank, Eurostat, FranceAgriMer, USDA, Enterprise Resource Planning and Business Intelligence firms, media companies or data providers from regional or global markets. Currently, the platform counts with data from LME, NYSE, Singapore Exchange and other market prices and indexes.

Other data that would improve the business

Commoprices admits their business offer would significantly improve if a timely availability of macro indicators on the cost of energy, labour and transportation was assured. These datasets are held by national and European agencies but are only updated every 6 months. Clients would need a daily to monthly frequency. In addition, some operators (eg FranceAgriMer via RNM services) still charge to get access to recent open data prices. Commoprices is also interested in price reference indexes of specific industries which are only accessed against substantial payments. Examples of these price reference indexes are Energy prices (e.g. Powernext²²¹ index).

Societal impact

This start-up, founded by two young professionals and employing another four people, shows a true added value in creating more jobs and opportunities for young people in the EU. Thanks to ODINE's advice and monitoring, this start-up has tripled their monthly sales and scaled up to other EU markets.

²²¹ Powernext manages the natural gas activities and renders services to the energy sector and operates the National Registry for electricity guarantees of origin in France.

CROPTI: EFFICIENT AND ENVIRONMENTALLY-FRIENDLY FARMING

Business opportunity

With their experience in farming, Cropti founders were researching innovation opportunities as they became involved in the entrepreneurial scene. They found out that there is a low digitisation level in agriculture, and at the same time the farmers are facing with increasingly low margins and regulatory pressure. Cropti founders found an opportunity to ease compliance by bringing all data from the farm to the cloud.

Goal of the company

[Cropti](#) provides software solutions for more efficient farm management. One solution helps farmers manage their business more efficiently while complying with EU food safety regulations and industry quality standards. Cropti provides farmers with personalised recommendations based on various types of constantly updated data. .

[Crop.live](#), another solution of Cropti, allows access to satellite imagery for small and medium sized farms that can use it to identify plot variability and identify problems as quickly as possible. On top of this, Cropti is building an irrigation estimation tool that relies on satellite imagery to assess crop water demands.

Open data used

Cropti is a re-user of open data gathered by the [European Space Agency](#) as well as EU legal information relevant for the food sector. Cropti also aggregates weather data from different providers and mapping data is core to its technologies.

Other data used

Additionally, Cropti uses agronomic data, such as crops evolution and incidences and soil data, and integrates within its products a comprehensive database on agrochemicals that are authorised to be used in Spain, including all their characteristics, such as application and usage dates etc.

Other data that would improve the business

Better availability of data relevant for the mapping of plots would help the business greatly for example as bulk download or direct access via APIs as recommended by the non-binding Commission Guidelines on datasets, licensing and charging for the re-use of documents. Furthermore, these guidelines recommend that datasets coming from cadastral, topographic and other geospatial data are priorities for release. Cropti believes the administrations are holding big amounts of data that are compiled through the CAP subsidies processing (e.g. crop rotations). If these data were to be opened, it could provide a solid base for the development of AI tools that would help greatly in fields such as pest scouting and prevention, yield estimations, etc.

Societal impact

Improving agriculture is one of the most effective ways to benefit the society, as farming is closely related to the three greatest challenges of our time: environment, health and inequality. Firstly, technologies such as the ones offered by Cropti help farmers reduce their use of resources for the same outcome, bringing a more environmentally sustainable food production. Secondly, optimization of farming practices leads to healthier products, as the use of chemicals in our food can be reduced to the minimum. And last but not least, by helping farmers become more profitable, inequalities that often push them to more unsustainable practices and keep them in the ‘poverty trap’ can be reduced.

DATA SPARKS: INSPIRING CHANGE THROUGH EFFICIENT ENERGY CONSUMPTION

Company profile

[DataSparks](#) is a solution developed by Derilinx as part of a Small Business Innovation Research (SBIR) Initiative, supported by the Sustainable Energy Authority of Ireland (SEAI) and Enterprise Ireland. [Derilinx](#) is a company that helps public bodies, smart cities and NGOs publish and share data, driving decision-making, innovation and transparency. Derilinx has extensive experience in the provision of fully hosted and managed data solutions, notably managing the Irish Government's National Open Data Platform (recently ranked #1 in Europe)²²².

Business opportunity

There are many examples of stakeholders in the energy chain who are striving towards more efficient energy consumption through sustainable energy planning, provision of customised support and individual decision making. Homeowners and tenants want to live in warm and comfortable homes, while keeping energy costs down. The Irish Government has committed to achieving a 20% reduction in energy demand across Ireland by 2020. Property advertisers and realtors want to maximise property sales and rentals by selling warm homes. Energy agencies support communities and local authorities to collaboratively reduce energy consumption, but often can't evaluate the effectiveness of their campaigns and measures.

Goal of the company

DataSparks accelerates effective decision-making for these stakeholders by providing access to smart, actionable energy insights. Its Energy Data Hub provides access to key property and energy data, ensuring relevant information is easy to find and easy to use. DataSparks [WebApp BER Hero](#) helps users understand their energy usage, how it compares to their peers and how their energy efficiency can be improved to reduce energy consumption, enhance home comfort and increase property value.

²²² <https://www.europeandataportal.eu/en/dashboard#2017>

Open data and other kinds of data used

The development of the DataSparks Energy Data Hub involved direct collaboration with data publishers, such as the Sustainable Energy Agency of Ireland, the Irish Green Building Council, Central Statistics Office, Energy Agencies, etc., on the sustainable publication of high-value open data.

Datasparks data is open by default for other re-users, but it can also be restricted to internal data sharing only. The DataSparks Energy Data Hub contains over 150 datasets from 24 public and private organisations. A number of datasets were semantically enriched, improving their quality and interoperability, including the BER research data set. Some examples of datasets are: carbon emissions, energy production, energy imports and fuel poverty; Building Energy Ratings, energy efficiency, households and heating; property prices, energy consumption, fuel types and fuel prices; and energy grants and energy credits.

Societal impact

DataSparks directly accelerates the transition to a low-carbon energy future by creating more awareness and participation within users of their services. It provides them with energy insights to make their energy consumption more efficient, and thus, the tools to help fighting climate change in their day to day lives.

GOOV: DEMOCRATISING PUBLIC TRANSPORT SERVICES

Business opportunity

Many senior people or people who suffer from disabilities, brain injuries and psychiatric disorders, are unable to travel independently in their cities. As a result, they need a companion or the public services to invest in adapting public transport comprehensively. Both options are expensive for the families and for the public sector. At the same time, parents of kids with a disability believed that that their children would be able to travel by themselves, if they had step-by-step instructions and a helpline in case of emergency.

Goal of the company

[GoOV](#) is a social enterprise that offers people from 10 to 90 years old the opportunity to travel independently in their city. GoOV empowers users to travel independently and safely within their limits with the help of a mobile device and at a lower social cost than current transport solutions. It provides an app that shows the route and gives instructions to go from door to door, including walking routes and waiting time at bus stops and stations. GoOV indicates at what time the bus or train will arrive, where and when to switch and at which stop or station the traveller must get off. Their unique selling proposition of GoOv is that it sends all the supportive alerts at the right time (e.g. alerts such as "push the stop button", "walk to platform 10") to the individual users, based on their personal profile (e.g. senior, fast or slow, wheelchair).

This app comes with a help button to call GoOV helpline when the traveller does not know how to continue or in case of emergency. GoOV also offers a viewing app for parents and supervisors to follow the journey.

Open data used

GoOV uses Google maps, GPS data, Open Trip planner, Open street map and the Dutch Open Data Base on public transport with live time and place data, accessibility data of bus stops, train platforms and PT-vehicles and carriages.

Other data used

GoOV uses dynamic public transport data, a lot of which is available free of charge in the Netherlands, but these data are not fully open and complete yet. GoOV had to make deals with different entities and helped mitigating errors detected in its sources by sending feedback to their providers. Since it is impossible to get GPS signals underground, as in the metro, GoOV is turning to beacon technology made available by local Councils in Rotterdam and Amsterdam. Also, GoOV is investigating how to use crowdsourced data.

Other data that would improve the business

GoOV would need more complete open data on public transport, focusing on accessibility to train carriages or station platforms, timetables, stations, bus stops, vehicles. It would also need data on routes to and from bus stops, railway and metro stations and indoor navigation in public buildings such as railway and metro stations. In the future, they would like to access the geolocation information which is provided by the last generation of connected public LED based lampposts which are being deployed in the Netherlands in all streets –and indoors.

Societal impact

GoOV responds to the UN-treaty provisions for the rights of people with a disability, by improving personal mobility to increase and guarantee autonomy under the three pillars of physical, mental and financial accessibility. This app makes public transport accessible and adapted to the needs of senior people and people with disabilities. GoOV has found out that 84% of the users switches permanently from taxi to public transport thanks to them. Besides, it has saved up to 50% of costs in community taxi transportation.

GREEN CITY SOLUTIONS: MAKING CLEAN URBAN AIR PROFITABLE

Business opportunity

When it comes to improving quality of life, cities face significant challenges. The high levels of pollution from traffic and industry, the dense population of inner-city areas have an effect on citizens' health and cause problematic climate conditions due to heat islands. Cities are expected to promote **biodiversity** and **green areas** in public spaces. Yet they often lack not only the space for this, but also the funds. These conclusions and experiences triggered Green City Solutions' search for an ecological and economical solution for such worldwide problems. Green City Solutions (GCS) learned about the

remarkable impact specific moss cultures have on local air pollution levels and decide to use them as an air purification solution, the City Trees.

Goal of the company

[Green City Solution's vision](#) is a world in which people can live healthily in cities. Green City Solutions aims to create living conditions that allow to permanently have clean and cool air to breathe. It achieves that by linking the natural abilities of air purifying mosses with cutting-edge Internet of Things technology in a unique way. Green City Solutions has created City Trees, a mobile and intelligent biological air filter that can measure air pollution in a city. The moss protected by plant coverage binds particulate matter, nitrogen oxides and CO₂e and produces valuable oxygen. At the same time it cools the surrounding air. The measurements can help public administrations monitor environmental pollution in the cities. These trees can make a measurable and sustainable contribution to the development of intelligent and future-proof cities.

Open data used

Green City Solutions has developed an open data tool for visualisation of the City Trees' effectiveness and impact as an air purification measure. GCS has used open data from official pollution measuring stations in Europe -and Asia- to create a comprehensive map of the pollution levels in that specific location and have combined it with the data they collect from the tree sensors. The goal is to visualise the invisible air pollution and make such visualisation publicly available.

Other data used

In 2016, Green City Solutions has collaborated with Cisco Systems in a smart cities project, which provided valuable insights on the correlations of traffic and air pollution.

Other data that would improve the business

Green City Solutions would extremely benefit from accessing live data and historical data concerning the precise meteorological values. Transport data, including traffic and other data collected by road authorities, could improve the calculus of the impact of nature-based solutions. Finally, data concerning the health benefits of natural solutions (e.g. parks in the city) could be beneficial to benchmark the impact of Green City Solutions with traditional ones. These data can be provided by many actors in the health or research administrations.

Societal impact

Green City Solutions aims to mitigate climate change and fight air pollution in the cities. This is one of the challenges that cities in Europe face when aligning with the sustainable development goals. With these trees, Green City Solutions creates a smart infrastructure for sustainable and liveable cities of tomorrow, that monitors environmental pollution.

VIOMEDO: ACCELERATING PATIENTS' ACCESS TO MEDICAL INNOVATION

Business opportunity

One of the largest challenges in clinical development is finding the right patients to take part in clinical trials. This is why it takes ever more time and money to bring innovative therapies to market. Overall, about 40% of trials fail to achieve minimum patient registration. Ultimately, 85% of all trials are delayed with each day costing the organisation up to USD 8 million (EUR 7.1 million).

This leaves many EU patients without early access to clinical innovations within trials and many doctors without early experience with novel therapies. Without innovative solutions, this problem will only grow as trial sizes increase and therapies become more targeted.

Goal of the company

[Viomedo](#) develops an online platform that connects patients with medical research in order to help improve existing therapies and find cures. It uses a unique distribution network, data and technology to reach and register new patients that comply with the requirements (eligibility criteria) of the clinical trials. Together with patients, patient advocacy groups and physicians, Viomedo supports leading research institutions, life sciences companies and Clinical Research Organisations with their clinical research. At the moment, Viomedo is the leading clinical trial platform with 300 sites in Germany and Austria and with over 250.000 patients using the platform in the first half of 2017. For institutions, Viomedo provides services covering over 200 diseases.

Open data used

Viomedo uses data related to clinical trials registries in Germany (factual data, non-personal), which are usually run by public universities. These registries share research data and the type, date and location of clinical trials. Information is disclosed in an open and free manner, enhancing the transparency of the universities. Viomedo combines these data with open geographic data to connect patients with hospitals or organisations running clinical trials.

Other data used

Viomedo is building its own registry by partnering with pharmaceutical companies and doctors carrying out clinical trials. Pharmaceutical companies provide their own data on clinical trials and Viomedo combines it with publicly available data. However, the company sometimes experiences problems with the accuracy and timeliness of the open data, which ultimately impacts on the patient.

Viomedo also integrates personal data where patients opted in. It strictly follows the data privacy legislation in place in Germany. However, GDPR will offer a window opportunity by providing harmonised legislation on data protection, making scaling up easier and cheaper.

Other data that would improve the business

Research data is of high importance to Viomedo. Their future data services seek to access data from medical journals and research institutions. The interest lies in information on the kinds of clinical trials conducted, with which patients and at which sites. That research data from publicly funded research can serve as a baseline as well as provide other relevant information for future clinical trials. An initiative such as the Open Science Cloud, centralising access to public, publicly-funded and private (on a voluntary basis) research data will enhance the usability of the data.

Clear policies on usage of clinical trials data from institutions such as the European Medicines Agency would help Viomedo providing a cross-border service. Furthermore, obligations on the provision of machine readable formats and APIs would allow better use of data already available.

Beyond health statistics at a national level provided by EUROSTAT, Viomedo would need such statistics at city or regional level. This would allow them to match patients with disease incidence and provide a better service to those looking for clinical trials.

Societal impact

By connecting patients in need with clinical trial opportunities, Viomedo accelerates the process of developing new therapies. Consequently, new cures can reach all patients faster.

THE ESTONIAN ROAD ADMINISTRATION: A DIGITAL SERVICE

Profile of the organisation

The [Estonian Road Administration](#) (ERA) is a government agency operating within the administrative area of the Ministry of Economic Affairs and Communications. The ERA provides public services, which allow road users to use the road network in a safe and efficient manner, ensuring its preservation and operation. At the same time, the ERA applies requirements to road users and vehicles to ensure a safe and working traffic environment. With its public transportation management, the ERA influences the environmentally-conscious and sustainable behaviour of road users.

The decision to open data

The ERA is not obliged by law to publish its data as open data, although it has to make information publicly available. Nevertheless, the decision to make open data available was taken in view of meeting the need of users of public transportation. The ERA realised that setting APIs for providing facilitated access to its dynamic data would translate in new applications and services that would benefit the end user. Now, a big number of developers and companies make use of their open data. The only barrier that the ERA came across was to adapt the data to GTFS (General Transit Feed Specification)

standards. In comparison with other project such as the [Public Transport Information Systems](#)²²³, their investment in opening the data was not significant.

ERA's open data

The open data of the public transport register consists of data entered in the national public transport register with other kinds of data content, including descriptions, timetables and the locations of stops of domestic public transport routes. Examples of open data are: real-time train information, timetables, public transport stops, travel ticket prices, travel fare concessions; soon ferries and domestic airlines will also be included. This year the ERA plans to make real-time bus information available also as open data, which for the moment is only published under more restrictive conditions. Some information published within the PTIS is the regular services authorisations, public services contracts and service provider cards.

This open data is freely accessible to all interested parties and updated daily. When reusing the open data in publicly accessible channels (including websites, applications in smart devices), the party reusing the data commits to providing references to the original source of the data and guarantees that any data used in any publicly accessible applications is no older than 7 days from the moment the data is downloaded.

Societal and economic benefits

The social benefits can be tied to a better service to the citizen as well as to an increase of the use of public sector transportation, thus directly benefitting quality of life and the environment, in helping fighting climate change. For citizens, thanks to competition between the different information providers, citizens can make use of competitive and useful services that help to organise their mobility in different transports better.

The ERA seems to have created a community of developers and organisations that use their open data. While some of the open data users are big companies, there are other companies and developers who use this data and make competitive apps such as Ridango (a ticketing solution provider) or App Moovi (a public transit guide). Moreover, many shopping malls use ERA's open data to display public transport timetables of nearby stops.

LIANDER: POWERING HOUSEHOLD ENERGY CONSERVATION WITH OPEN DATA

Profile of the company

[Liander](#) is a Dutch company responsible for energy network distribution and administration. In the Netherlands the energy production and supply were liberalised to allow for more competition in 1998. Thus, this operator would be excluded from the

²²³ Public Transport Information System (PTIS)²²³ groups together information to plan, manage, supervise and use public transport services. It was created with the help of the European Economic Area and Norwegian Financial Mechanisms.

scope of application of the current PSI Directive (Art. 2 No. 2), even if it is a publicly regulated activity. Liander is the largest Distribution System Operator in the Netherlands. It operates in about one third of the country and it provides electricity and natural gas to 5.2 million households and businesses covering 37% of the Dutch network.

Barriers to open data

Liander believes in the power of open data for energy conservation and reduction of carbon emissions. Sharing data is considered as its responsibility to facilitate the energy transition. However, sharing data implies costs. For example, data preparation (content and format) and adaptations to optimise open data usage requires time and skills. The latter includes, amongst others, user feedback or help-desk.

Liander's open data

Since 2015, Liander²²⁴ is sharing data free of charge and accessible to everyone. The data can be downloaded at www.liander.nl/opendata and is also presented in <http://www.energieinbeeld.nl/> (web service) using a public login code. In order to ensure the privacy, data at connection level (containing personal data) are always anonymised by aggregation or otherwise.

Liander publishes openly the following data: energy consumption of small connection, gas consumption up to a capacity of 40m³/h and electricity up to a capacity of 3x80 Ampere, standard annual consumption of gas and electricity for four housing types, daily profiles showing fluctuation throughout the year of electricity and gas consumption, interruptions in energy provision due to temporary faults, detailed smart meter consumption of 80 addresses, representing typologies of user profiles and planning and realization of rolling out of smart meter.

Societal and economic benefits

Liander realised that after the substantial initial investment (around EUR 200-400K) of setting an open data policy in the company, further costs to keep the dissemination of open data operational became insignificant. Both Liander and re-users gained in efficiency and reduced costs in the long run. Open energy data in combination with other data allows for innovation that can revert in better services for their partners during the energy transition. This data use can also provide Liander with feedback on the quality e.g. flaws in the data. It also reduces the administrative burden of dealing with requests of data, as data is open for all.

Furthermore, the availability of data in the energy sector enables greater energy efficiency, with positive impact on the environment, and hence for the whole society.

²²⁴ Its business is influenced by the fact that it has a regulated public task to administer and maintain energy networks. It is important to notice that, under the Energy Act 1998, network administrators are obligated to share and exchange their own data between network administrators and the transmission systems operators. However, there is no legal requirement to disseminate data to the general public.

Thanks to its open data policy, Liander is now taking part in various Smart City projects and its reputation has improved substantially.

Examples of apps

Although they have not built a community in itself, Liander receives reactions on a weekly basis on the data that they open. GeoGraphisch is one interesting example of an app that provides a "zoom in on energy" service where it shows the energy consumption of more than 30K addresses, by combining it with other geographic data. It can show, for example, the level of insulation and energy saving of the buildings of some streets or the energy consumption rate per municipality.

TRANSPORT FOR LONDON: THE CREATION OF AN OPEN DATA ECOSYSTEM

Profile of the company

[Transport for London](#) (TfL) is an integrated transport authority responsible for delivering Mayor of London's strategy and commitments on transport. TfL runs the day-to-day operation of the Capital's public transport network and manage London's main roads.

Every day more than 31 million journeys are operated across the TfL network. The Directive on the re-use of the public sector information does not cover operators such as TfL as they have an 'industrial or commercial character'.

Open data at TfL

TfL considers that as public body funded by fare and tax-payers, transport information is rightly seen to be owned by the public. Opening their transport data increases TfL's transparency and reach dramatically and provides enough bases to create niche products, sparking innovation and high economic benefits.

Open data was the most effective way to give app services to their customers taking into account their almost 15K visits per month via mobile phones. This is almost double of the number of visits via personal computers. Thus, transport data is in great demand, particularly real-time information such as bus and train status and arrivals/departures.

Barriers to open data

TfL considers that barriers to opening up data are essentially the fear of losing control over their data. This has the risk of voluntary distortion of data and provision of misleading information. TfL has not found any real evidence that such risks have materialised and sees no other barriers to opening up data through an API to a developer community.

Data sets and licences

The data is shared via UK Open Government License with minimal additions. For example, developers need to register with TfL to receive tokens from the API portal to access the data. The data is presented in three different ways: static data files (rarely

changing), feeds refreshed at regular intervals and via APIs. APIs enable queries from applications to receive custom-made responses.

Transport for London releases over 200 data sets which include bus arrivals, tube movements, departures, status, cycle hire docking station status, river boat status and arrivals, overground, rail and trams arrival and status, journey planner and accessibility. However, the community of developers grown through TfL's open data policy combines these data with other such as air quality, detail accident geo-location data, live road disruption and planned works data, etc.

Societal benefits

Transport for London has created a community of developers with whom they engage to explain their approach and plug gaps in the market. This community enables the optimal use of the transport network. Any person needing travel information can get it wherever and whenever they wish, in any way they wish. The community consists of more than 14,400 developers registered who made over 600 apps - the vast majority of which use the dynamic API - and other products available in the form which best suit their customers. This stimulates innovation provision faster than TfL could ever have delivered themselves.

Interestingly, only 40% of the users are Londoners, meaning that these apps have national and international outreach.

Technically, feedback on the open data can improve its quality and source systems.

Economic benefits

Emerging research indicates that open data provides return on investment of up to 50 times the sums invested in terms of customer benefit. In this specific case, London has gained around £100m direct value by technological investment in London and elsewhere off the back of its data. Through saved customer time saved by journey planning, and more efficient utilisation of their resources, TfL's open data ecosystem has enabled an estimation of 1000 jobs. 175K people are now employed in the digital technology industry in London, in 45K companies with a £30bn annual turnover.

Examples of apps

TfL partners with a number of apps, who have around 10% of market share of London travel apps. These non-exclusive partnerships help TfL ensure that its customers receive more robust and consistent information across a variety of platforms. For example, combining existing data with crowd-sourced information can help TfL provide more detailed journey planning information to its customers.

This also allows the sharing of knowledge, sparks innovation and increases TfL reputation and transparency.

OPEN DATA IN OTHER TRANSPORT COMPANIES

Other transport companies in the EU and Switzerland are also starting to open up data their data:

Deutsche Bahn's Mindbox, driver of its digitisation process, has opened 27 datasets (mostly static) and 9 APIs for dynamic data. The former [German railways](#) is today a group of companies providing the physical tracks, the train stations, open access and subsidised passenger rail services as well as cargo rail services. They are now experimenting with opening up data via CC BY 4.0 and for certain datasets CC0 and see the future value this could have in creating their own developers community and better services for their customers.

SNCF (société nationale des chemins de fer français) has opened 208 datasets since it started their open data policy in 2014 and receives 20 million requests for real-time data through their APIs every month. They share data on infrastructure description (static data), maintenance and modernisation work on a weekly basis and social, environmental and financial data. The [SNCF platform](#) shares data necessary to plan and compare different journeys from station to station, to research upcoming trains (theoretical and real-time) and to consult timetables in different stations. Zac, a virtual assistant, incorporates SNCF real time data for trains via APIs. Thus, Zac can alert users of train changes and cancellations in real time.

Swiss Railways considers that transport data is also infrastructure thus belongs to the public service. The legal uncertainty regarding the status of these data in Switzerland stifles innovation. For that reason, Swiss railways has a [data platform](#) which enables access to target timetable, up-to-date data, information about stops and transport companies, rail time forecast for direct journeys via APIs. However, the data in question is not entirely open as Swiss railways has established some terms of use and limits and costs according to the intensity of the use.