European Commission

JRC TECHNICAL REPORT

JRC Digital Economy Working Paper 2020-04

The economics of Business-to-Government data sharing

Martens, Bertin

Duch-Brown, Néstor

2020

Contact information
Name: Néstor Duch-Brown
Address: Inca Garcilaso 3, 41092 Seville (Spain)
Email: Nestor.DUCH-BROWN@ec.europa.eu
Tel.: (+34) 954.488.216

EU Science Hub
https://ec.europa.eu/jrc

JRC119947

Seville: European Commission, 2020

How to cite this report: Martens, B., Duch-Brown, N., *The economics of Business-to-Government data sharing*, European Commission, Seville, 2020, JRC119947.

Contents

Foreword

As part of a JRC-CNECT research programme on digital platforms and data, CNECT/G1 requested the JRC to contribute an economic input to the on-going reflections in the B2G data sharing Expert Group[1]. The Expert Group was set up to (a) identify good practices, (b) assess the legal, economic and technical obstacles preventing B2G data sharing, (c) advise on actions to promote B2G data sharing for public interest purposes and (d) give recommendations to the Commission on how to further develop its B2G data sharing policy.

The subject of B2G data sharing was first discussed in the Commission Communication (2018) "Towards a common European data space". The Staff Working Document that accompanies the Communication sets out the basic principles of B2G data sharing and discusses some legal and technical conditions that may facilitate these operations. Starting from these principles, the B2G Expert Group discussed further refinements and details and explored how they could be cast into a more operational governance framework for B2G data sharing operations. The current report complements the debates in the Expert Group with an economic perspective on several of these principles.

[1] https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing

Acknowledgements

The authors would like to thank the comments and suggestions to earlier versions of this document provided by Alberto Gago-Fernandez, Johan Bodenkamp, Julie Sors, Maria Rosaria Coduti and the rest of colleagues from DG CNECT G1. Similarly, this report has greatly benefited from inputs provided by the Business-to-Government Data Sharing High Level Expert Group, and in particular from Alberto Alemanno. In addition, Michael Lutz (JRC B6) also provided important suggestions for improving the readability of the final version. Finally, Fiammetta Rossetti provided excellent research assistance. Remaining errors and omissions are the sole responsibility of the authors.

Authors

Bertin Martens

Néstor Duch-Brown

Executive summary

Data and information are fundamental pieces for effective evidence-based policy making and provision of public services. In recent years, some private firms have been collecting large amounts of data, which, were they available to governments, could greatly improve their capacity to take better policy decisions and to increase social welfare. Business-to-Government (B2G) data sharing can result in substantial benefits for society. It can save costs to governments by allowing them to benefit from the use of data collected by businesses without having to collect the same data again. Moreover, it can support the production of new and innovative outputs based on the shared data by different users. Finally, the data available to government may give only an incomplete or even biased picture, while aggregating complementary datasets shared by different parties (including businesses) may result in improved policies with strong social welfare benefits.

The examples assembled by the High Level Expert Group on B2G data sharing show that most of the current B2G data transactions remain one-off experimental pilot projects that do not seem to be sustainable over time. Overall, the volume of B2G operations still seems to be relatively small and clearly sub-optimal from a social welfare perspective. The market does not seem to scale compared to the economic potential for welfare gains in society. There are likely to be significant potential economic benefits from additional B2G data sharing operations. These could be enabled by measures that would seek to improve their governance conditions to contribute to increase the overall number of transactions. To design such measures, it is important to understand the nature of the current barriers for B2G data sharing operations. In this paper, we focus on the more important barriers from an economic perspective: (a) monopolistic data markets, (b) high transaction costs and perceived risks in data sharing and (c) a lack of incentives for private firms to contribute to the production of public benefits. The following reflections are mainly conceptual, since there is currently little quantitative empirical evidence on the different aspects of B2G transactions.

- Monopolistic data markets. Some firms -like big tech companies for instance- may be in a privileged position as the exclusive providers of the type of data that a public body seeks to access. This position enables the firms to charge a high price for the data beyond a reasonable rate of return on costs. While a monopolistic market is still a functioning market, the resulting price may lead to some governments not being able or willing to purchase the data and therefore may cause social welfare losses. Nonetheless, monopolistic pricing may still be justified from an innovation perspective: it strengthens incentives to invest in more and better data collection systems and thereby increases the supply of data in the long run. In some cases, the data seller may be in a position to price-discriminate between commercial buyers and a public body, charging a lower price to the latter since the data would not be used for commercial purposes.

- High transaction costs and perceived risks. An important barrier for data sharing comes from the ex-ante costs related to finding a suitable data sharing partner, negotiating a contractual arrangement, re-formatting and cleaning the data, among others. Potentially interested public bodies may not be aware of available datasets or may not be in a position to handle them or understand their advantages and disadvantages. There may also be ex-post risks related to uncertainties in the quality and/or usefulness of the data, the technical implementation of the data sharing deal, ensuring compliance with the agreed conditions, the risk of data leaks to unauthorized third-parties and exposure of personal and confidential data.

- Lack of incentives. Firms may be reluctant to share data with governments because it might have a negative impact on them. This could be due to suspicions that the data delivered might be used to implement market regulations and to enforce competition rules that could negatively affect firms' profits. Moreover, if firms share data with government under preferential conditions, they may have difficulties justifying the foregone profit to shareholders, since the benefits generated by better policies or public services fuelled by the private data will occur to society as a whole and are often difficult to express in monetary terms. Finally, firms might be afraid of entering into a competitive disadvantage if they provide data to public bodies – perhaps under preferential conditions – and their competitors do not.

Several mechanisms could be designed to solve the barriers that may be holding back B2G data sharing initiatives. One would be to provide stronger incentives for the data supplier firm to engage in this type of transactions. These incentives can be direct, i.e., monetary, or indirect, i.e., reputational (e.g. as part of corporate social responsibility programmes). Another way would be to ascertain the data transfer by making the transaction mandatory, with a fair cost compensation. An intermediate way would be based on solutions that seek to facilitate voluntary B2G operations without mandating them, for example by reducing the

transaction costs and perceived risks for the provider data supplier, e.g. by setting up trusted data intermediary platforms, or appropriate contractual provisions. A possible EU governance framework for B2G data sharing operations could cover these options.

1 Introduction

Data is at the core of the digital economy and the digital transformation. Today, the rankings of the world's most valuable corporations are dominated by companies that have placed data at the centre of their business operations. Many others in almost all sectors of the economy have been collecting data for years and have just started to identify its value. The availability of these large amounts of data offer many opportunities to improve efficiency, innovation and economic prosperity. In order to be able to materialise these opportunities, the right policy frameworks have to be created. The emerging data economy poses several challenges to policymakers. Not only data markets are opaque and have reached a point where massive data collection may affect the right to privacy, but also incumbents have powerful incentives to stockpile data, limiting competition as well as the social benefits that could emerge from wider access. Hence, a range of policies could be deployed to promote data sharing to foster not only competition and innovation, but also the common good. Here we look more specifically at several issues surrounding data sharing between the private sector and the public sector. Business-to-Government (B2G) data sharing can provide a wide range of social benefits, but its effective implementation requires the identification and removal of several barriers that are holding it back.

This report is structured in three sections. Section 2 describes the potential socio-economic welfare gains for society of B2G data sharing operations. In the absence of an ex-post assessment based on actually observed impacts (due to the lack of relevant data) it illustrates the potential benefits with examples on existing B2G case studies using ex-ante assumptions about their economic impact. Section 3 discusses the economic obstacles that may hinder further growth in the volume of these operations. Three reasons are considered: monopolistic data suppliers, high transaction costs and residual uncertainties, and incentive misalignment between private data suppliers and the public benefits of B2G operations. Section 4 concludes with an exploration of possible options for a governance framework that could address these obstacles.

2 The potential socio-economic welfare gains from B2G data sharing

In this section we look first at the potential welfare gains from sharing data. Then, we provide some analysis of the economic impact of data sharing, focusing on the case of B2G. Finally, we describe some examples of transactions that suggest that there is an emerging market for B2G data sharing operations.

2.1 The potential welfare gains from sharing data

Governments need information or data in order to, among other activities, take policy decisions. The more data the more informed the decisions can be. In the digital economy, some private firms have been stockpiling large amounts of very useful information for policy making. While governments do not have direct access to that information, B2G data sharing could greatly improve a government's capacity to take better policy decisions in the public interest and to generate more social welfare for society at large (Alemanno, 2018). The B2G Expert Group provided many examples that illustrate the promise of societal welfare gains from B2G data sharing. For example, data collected by mobile phone operators about the location of their customers can be used not only for its primary purpose – keeping track of the location of mobile phones connected to a network –, but for many other purposes. Those include activities provided by the private sector such as informing drivers about traffic conditions via navigation apps, or for activities in the public interest such as allowing traffic authorities to manage traffic flows and pollution in cities, or redesigning road networks and public transport systems. Similarly, payment transactions data from banks and other payment services providers can be used for other purposes than keeping track of payments, including tracking the origin of tourist flows, expenditure patterns in cities to improve tourism promotion services, or compiling point of sales price statistics to rally estimates of consumer expenditures and price indexes by national statistical institutes.

Theoretical economics can demonstrate the benefits of data sharing. The non-rival nature of data is the fundamental economic driver of socio-economic welfare gains in data sharing operations: many parties can use the same dataset for a variety of purposes without functional loss to the original data collector (OECD, 2015; Duch-Brown et al, 2017; Jones and Tonetti, 2018). Rival goods can only be used by one party at the time. For example, a car can only be used by one driver at the time. If a car were non-rival all drivers could use the same car at the same time to drive to different destinations. The welfare gains would be enormous: it would be sufficient to invest in the production of a single car to cater to the needs of all drivers. That is precisely the promise of sharing of data as a non-rival good: if one party collects the data they can be shared with many other users, including public bodies, and all parties can benefit from the use of the data. This results in substantial cost savings for society, avoiding repeated collection of the same data. Moreover, different users may produce new and innovative outputs with the very same set of data, applications that the original data collector could have not envisaged initially. The primary data collection effort is a sunk cost that can be amortized across a large number of uses, rather than remaining confined to a single user. Jones and Tonetti (2018) use a theoretical macroeconomic model to illustrate how the social welfare gains from data sharing are vastly superior compared to a scenario without sharing.

An additional economic characteristic of data is related to the existence of economies of scope in data aggregation (Duch-Brown et al, 2017)[2]. A single dataset may give only an incomplete or even a biased picture. For example, using data from a single credit card company or payment services provider with a small market share may generate a biased picture of the number of tourists in a city or the number of cross-border payments. Combining the data from several payment services providers into an aggregated dataset may produce a more refined and accurate insight and avoid biased estimates from incomplete and non-representative data. Similarly, aggregating location data with rich socio-economic profile data from social media services may produce more valuable insights for city transport and mobility planners. Achieving the benefits of data aggregation requires data sharing. Aggregating complementarity datasets may result in strong social welfare benefits. For example, in China it is mandatory for car manufacturers to share all data generated by their electric vehicles with a government research institute. The data can be used to track the

[2] The economic benefits from non-rivalry or the ability to use data for multiple purposes is also a form of economies of scope that is known under the label of "joint production" (Panzar and Willig, 1981). Economies of scope from data aggregation as described in this paragraph are related to the improved insights that can be generated from merging several complementary datasets (Rosen, 1983; Duch-Brown et al, 2017).

mechanical and environmental performance of electric vehicles, improve the planning of battery recharging stations and put in place measures to stimulate the use of electric vehicles in dense urban traffic areas.

2.2 The economic impact of data sharing

In this section we examine the economic benefits of data sharing from an empirical perspective. There are many examples of social welfare-enhancing B2G data sharing transactions. The B2G Expert Group has collected several best-practice examples, and there are many more as the volume of B2G data sharing operations expands. There are ex-ante impact estimates for some of these, based on assumptions about the impact on citizens. These assumptions are not based on observed behaviour. We do not know to what extent these assumptions have actually been realised. There are, to the best of our knowledge, no ex-post assessments of the actual impact of data sharing operations.

What would we need to produce a quantitative economic estimate of the benefits? Some methodological issues should be born in mind:

Data have no value as such. They become valuable only when they are used as an input into the production of services that increase the welfare of citizens and/or firms. The socio-economic value of a B2G data sharing operation thus depends on the user value of the public service that has been produced with the data[3]. To estimate that value requires information on the impact of the data on the different producers and users of the public service. For example, in order to estimate the social welfare gains for society from using mobility data for the purpose of improving a public transport network, we would need information on the state of the transport network (connections, number of passengers by route, tariffs) before and after the use of mobility data. The economic value of an improved train schedule data depends on how much it increases consumer welfare (reducing search costs, ticket prices and waiting time for trains) and ticket sales revenue for the railway company. For example, a Deloitte (2017) study documents the impact of making data on public transport from the Transport-for-London company freely available to mobility apps developers. It estimates that, as a result of this data sharing and the services that were subsequently developed around these data, consumers saved between 70 and 90 million £ per year in transport time. However, the study does not cover the substitution effects and shifts in market shares between competing transport services (buses, taxis, etc.). Unfortunately, it does not cover either the economic savings in public sector subsidies to private companies so that a comprehensive and holistic transport service is ensured to citizens. A proper assessment of the economic impact of data sharing has to take into consideration all the associated benefits and costs, both direct and indirect

This example shows that it is not enough to look at the monetary value of a B2G data sharing operation. Some operations may involve a monetary transfer between a public body and the private firm while in other cases data are shared for free in the context of data philanthropy or corporate social responsibility policies. That reflects the cost side of the operation. On the benefits side, the end user of a public service may in some cases pay for the service, in others the service is a pure public good or it is offered for free to the consumer. There may be intermediary operators involved in the provision of a service and their costs and benefits have to be taken into consideration as well. Whether or not consumers pay for part of the cost, both the monetized and non-monetized part should be taken into account for the calculation of the social welfare benefits. For example, if B2G mobility data transfers are used to track the outbreak of a contagious disease and undertake public health actions to contain that outbreak, the cost-benefit analysis should take into account not only the public health budget savings thanks to less people falling ill, but also the value of working time gained by people who could have been infected but were not infected because of the countermeasures.

[3] Data are an experience good: their value can only be assessed by the user once they experience the use of the data for whatever purpose or service he intended to use them for. Conversely, once the user has experienced that value it cannot be undone and the data seller loses any leverage he has over the user to negotiate a price. The price thus needs to be paid before use. This is known as the "Arrow Information Paradox". However, in order to avoid these information asymmetries, commercial strategies such as allowing users access to sample data (for a subset of the complete data set) or for a limited period (e.g. one month) in order to evaluate the data are regularly implemented.

When an (improved) public service is excludable[4], a public service provider may decide to charge for the provision of the service, in order to be able to remunerate the private supplier of the data that contributed to the public service. Such payments would not directly[5] affect the total social welfare benefits of the public service. It would constitute a re-allocation of part of the benefits to the supplier that enabled the production of these benefits.

If no suitable impact information is available or if the impact cannot be measured in economic terms ("lives saved") then it is difficult to make an economic impact assessment. We can formulate a hypothetical impact but in the absence of effective measurement this remains speculative. The difficulty of collecting all the data may explain why we have been unable so far to find an existing economic impact assessment of a B2G data sharing operation[6]. A good example of an ex-post empirical micro-economic impact study of G2B (PSI) data sharing is the EARSC (2015) study on the impact of sharing government satellite data with private shipping companies to facilitate winter navigation in the frozen Baltic Sea. Satellite navigation reduces shipping costs for merchant companies but also ensures a steady supply of goods to consumers. Since each data sharing operation will affect different services and groups of beneficiaries in different ways it is hard to construct a macroeconomic impact assessment. A macroeconomic picture will either necessarily remain hypothetical (as in Jones and Tonetti, 2018) since it cannot collect and aggregate data on all possible specific cases, or will provide an incomplete or partial picture due to the highly aggregated perspective.

2.3 There is an emerging market for B2G data sharing operations

It is important to point out that there is an emerging market for data sharing, mostly in a B2B setting but some of it is used for B2G transactions. On the data holders' side, a distinction can be made between firms that are in the business of selling data as their main activity, and firms that are discovering the value of selling their data outside their normal line of business[7]. For the former data is their main product. They are often intermediaries between firms that generate the raw data and clients that require semi-processed data that are anonymized and preserve the commercial confidentiality of the original data producers. Examples of this first type include Bureau Van Dyck for company data, Euromonitor for industry data, Bloomberg, Euromoney and Crunchbase for financial data, Nielsen for digital media data, among others[8]. For the latter data is a by-product of their normal activity. They seek ways to generate additional revenue from the data and open new revenue streams. They do the pre-processing and anonymization and sell data directly to clients, or indirectly as a data service that sells insights but not the underlying raw data (Bergemann and Bonatti, 2018). Examples of this type include mobile phone operators that share anonymized location data to provide mobility insights to a wide range of private firms as well as government authorities, or credit card companies that share transaction and consumer profile data to client firms and to public bodies such as city tourism promotion services.

Some examples of B2G data transactions with these types of firms have been presented in the Expert Group[9]. Mobile phone operators and credit card companies operate in a competitive market; there are many operators in the market and public bodies could also participate in such markets. For example, several European Commission services, and in particular the JRC, have supply contracts with private providers of raw and semi-processed data that are used as inputs into evidence-based public policy making processes. Most of these providers are companies that sell data as their main line of business. In many other cases, public bodies can

[4] Excludability implies that the service can be made available to a specific group while other potential users can be excluded from the service. For example, a public health service can be freely accessible for all or it can be restricted to users who pay contributions to a public health insurance system.

[5] It may indirectly affect the total social welfare benefits through the pricing channel. We return to this subject in the next section.

[6] More economic research work has been done on G2B (Public Sector Information, PSI) data sharing. Even there however these are ex-ante and more speculative assessments, not ex-post assessments based on actual impact data (see European Data Portal (2015) for an overview).

[7] This distinction is based on the assumption that the costs of the joint production of a main product and a by-product are separable and hence easily identifiable from an accounting perspective. If, on the other hand, by-products are closely related to the main product, costs incurred in their production are joint costs that cannot be separated and can be allocated to the different products only in some arbitrary manner.

[8] Most -if not all- of these companies started their operations long before the relatively recent big data phenomenon, and have long been operating in the data industry.

[9] All examples discussed in the Expert Group are only described very shortly in this note. We assume that they will be made available in more detail as part of the report of the Expert Group.

also procure fully-fledged studies that may or may not incorporate data. However, these commissioned studies should not be considered as B2G data sharing initiatives.

The Expert Group also looked at examples of non-monetised B2G data sharing transactions where private firms shared their data with public service providers for free, either for corporate social responsibility (CSR) motives, data philanthropy – especially in the case of developing countries and actions that favour socially disadvantaged groups – or as part of a non-monetary exchange agreement whereby the private provider also receives some benefits in return. An example of the former type in a developing country is the use of mobility data from a private telecom service provider in Haiti to track the movement of citizens after an earthquake. The data enabled public service providers and humanitarian aid agencies to improve the provision of basic food, health and sanitation services in the wake of this disaster and thereby prevent outbreaks of diseases. An example of the latter is data sharing for cancer treatment in Romania, between a public clinic and a private firm that provides genomic analysis that enables doctors to apply treatment procedures that take into account the specific genomic defects of individual patients. Patients benefit from better treatment that increases their survival chances; the firm benefits from receiving patients' genomic data and treatment response data that contribute to improving the machine learning algorithm for the genomic treatment analysis – and may ultimately benefit many patients all over the world.

These examples show that there is a nascent market for B2G data transactions that contributes to the promise of socio-economic welfare gains from data sharing. However, many of these remain one-off experimental pilot projects that do not seem to be sustainable over time. Overall, the volume of B2G operations still seems to be relatively small and clearly sub-optimal from a social welfare perspective. The market does not seem to scale compared to the economic potential for welfare gains in society. The purpose of the B2G Expert Group was precisely to find out what is holding back these operations and what can be done to ensure that volume of transactions increases. There are likely to be significant potential economic benefits from additional B2G data sharing operations. While we cannot concretely estimate the value of these welfare gains for society we can safely assume that there are significant gains from data sharing operations. These could be enabled by a new policy initiative that would seek to improve their governance conditions, thus contributing to further growth in the overall number of transactions. We turn to that subject in the next sections.

3 Barriers to B2G data sharing and how to overcome them

The examples collected by the expert group, and many more, show that there is already an emerging B2G data sharing market with voluntary data transactions. However, there is also anecdotal evidence that supports the view that not all potentially welfare-enhancing B2G data sharing operations are actually materializing (Elkin-Koren and Gal, 2018). The current volume of B2G data transactions may be insufficient from a social welfare perspective. More data sharing could take place in the public interest. But in order to increase the number of transactions, a more active public policy stance may be required. To design such a policy initiative it is important to understand the nature of the barriers that separate currently feasible from unfeasible B2G data sharing operations. In economics these barriers are called market failures[10]. Among the reasons why barriers may occur, here we focus on three of the most frequent: (a) monopolistic data suppliers and monopolistic pricing, (b) high transaction costs and perceived risks in data sharing and (c) a lack of incentives for private firms to contribute to the production of public benefits.

3.1 Monopolistic data suppliers and data pricing

Some firms may be in a privileged position as the exclusive providers of the type of data that a public body seeks to access. It enables the firms to charge a high price for the data, i.e. a price that considerably exceeds the resources needed to produce the data, beyond a normal rate of return on costs. While a monopolistic market is still a functioning market, the market failure resides in the fact that the monopolist reduces access in order to increase the price to a level that maximizes his profit. This causes deadweight social welfare losses[11]. Monopolistic pricing may still be justified from an innovation perspective. It strengthens incentives to invest in more and better data collection systems and thereby increases the supply of data in the longer run[12].

3.1.1 Pricing in monopolistic markets

Pricing above marginal costs occur frequently in the digital economy because of the nature of the underlying digital technology. Setting up data collection, processing and distribution systems often involves high fixed set-up cost and low marginal costs: contrary to most analogue processing systems, digital data systems are easily scalable once they have been set up. If the owner of the digital system only charged the marginal cost as a price for his service he would not be able to recuperate his fixed investment cost. Mark-up pricing (above marginal costs) is therefore a standard feature in many digital economy services[13].

Competitive markets in principle operate under the marginal cost rule. Firms can only mark-up their prices above marginal costs to the extent that competition becomes more monopolistic as products from different firms are imperfect substitutes. More competition will push prices closer to marginal costs. However, pure marginal cost pricing is only sustainable in the absence of fixed costs. In a digital economy characterized by fixed costs, average cost pricing or some mark-up on top of marginal costs is necessary for the main

---

10 The concept of "market failure" is central to the economics of regulation and to the European Commission's "Better Regulation Toolbox" (EU SWD (2017)350. The toolbox identifies four reasons for regulatory intervention: market failure, regulatory failure, equity concerns and behavioural bias (p 85). The document also discusses the underlying economics of market failure, including externalities, weak incentives and incentive misalignment – similar to the arguments discussed in this note on B2G data sharing. See https://ec.europa.eu/info/sites/info/files/better-regulation-toolbox_2.pdf

11 In some B2G transactions the government may be the sole buyer of the data and therefore be in a monopsony position. If both data seller and buyer are monopolists we have a bilateral monopoly. It is not clear in that case which of the two parties will exert market power. For example, under EU Regulation 883/2013 traffic authorities may collect data on road safety and security generated by digitally connected cars. Both manufacturers and authorities are in a monopoly situation. This situation has been settled by a mandatory B2G data sharing regulation.

12 This type of reasoning underpins the economics of intellectual property rights, such as patents and copyright. In IPR, a right holder is given an exclusive monopoly right to the commercialisation of his IPR-protected product. This monopoly is inefficient from a short-term static welfare point of view but may be welfare-enhancing from a longer-term dynamic point of view because it incentivises other innovators to invest in new inventions with the prospect that they can monetize the benefits of their invention.

13 An increasing number of digital data systems are multi-sided markets where pricing on one side of the market may be used to subsidize (or tax) pricing on the other side (Rochet and Tirole, 2006). Prices may in these cases not reflect costs.

business, if not the firm will run a loss[14]. This implies that some social welfare losses are inevitable. This is illustrated in annex 1.

A B2G example of data quality discrimination concerns a public transport company that may buy a cheaper more coarse-grained mobility dataset from a telecom operator, rather than a more expensive fine-grained dataset, in order to reduce the cost of data procurement. This may result in a less optimal transport schedule for consumers. The mobility data provider may be willing to provide more fine-grained data at a higher price, thereby extracting more welfare surplus from the transport company and its consumers. Price differentiation is often linked to quality differences.

In some cases, the data seller may be in a position to price-discriminate between commercial buyers and a public body, charging a lower price to the latter. An example of this type of B2G price discrimination comes from the JRC's own experience with the procurement of a unique dataset on online digital transactions from a private market analysis firm. While these data have a high commercial value for firms operating online stores, the supplier agreed to sell it at a much lower price to the JRC because it would not be used for commercial purposes. This led to a price negotiation between the two parties, still above marginal cost however. There are also examples from our own experience where the JRC was unable to procure data at a reasonable price and had to cancel the transaction. In some instances, even if the data seller accepted a reduction in price, the price proposed was still above the available budget. In some others, after some negotiation, data holders decided not to sell the data perhaps due to the fear of reputational losses.

Pricing hinges on the assumption that the demand curve reflects true demand for the data (see annex 2 for a detailed graphical exposition). However data are often an intermediate input into a service produced by a downstream party. In this case, the willingness to pay for the data will depend on the derived demand in the downstream application. For example, mobile phone data may be an input by a public transport company to improve its bus or train schedules to better accommodate consumer demand at different places and times in a city. In that case increased demand for transport services may reflect the true welfare value of the intermediate data delivery. The cost of the data input can be recuperated via ticket prices. If the data input is competitively procured, then the transport company pays a market price to the data provider. If there is only a single provider, the transport company may have to pay a monopolistic price. This gives the upstream data provider an opportunity to extract a large share of the consumer value produced by the downstream transport service provider. In short, we cannot examine the upstream data market and downstream services market separately; we have to look at them as an integrated entity.

If there is competition among different data providers, B2G data pricing is likely to be close to marginal costs plus some mark-up or return on investment, depending on the extent of competition in the market. For example, a public body can tender for mobility data from several mobile phone operators in a country. If this were part of the firms' main line of business, fixed costs may interfere in this process and exert upward pressure on prices because in that case a firm also wants to amortize its fixed costs. When additional data processing is done for a specific user, that user may be charged the full marginal cost but also the additional data processing cost, unless competition is strong and puts downward pressure on mark-up pricing.

It is not clear to what extent monopolistic data pricing actually occurs in B2G sharing transactions since there is a lack of empirical evidence on this area and there will be high variation across sectors. One can think of examples of monopolistic data supplies in B2G data sharing situations. For example, a city authority may request access to private room reservation data from an online booking platform that operates in the city, for reasons related to tourist taxes, safety inspections and the implementation of other city council regulations with regard to private short-term rentals. Every booking platform has exclusive access to its own room reservation data and is therefore in a de facto monopoly position. If sharing the data with the city authority would be mandatory under the city regulation, that would reduce the price to zero for the public body[15],

[14] This can also be the case in markets where there are a relatively large number of firms but selling differentiated products, i.e., products that are not identical for different consumers. This feature gives firms some market power because they will be the sole producers of their variety, while the large number of competitors would suggest a market close to perfect competition. This market structure is referred to as monopolistic competition, and is characterized by the fact that firms price according to their average costs, but they retain some market power since the prices will be above marginal costs.

[15] DG COMP has a mandate under Art 101 TFEU to request data from firms, not only for specific competition cases but also for wider sector inquiries that are not necessarily case-related. It used this mandate in 2017 to collect firm data in the framework of an e-commerce sector inquiry. Attempts by DG GROW to obtain a similar mandate to access firm-level data in the context of the Single Market (the SMIT proposal) failed because there were insufficient legal grounds for such a mandate.

although the data supplier would still bear the costs of data collection, processing and transmission. If it is not mandatory, the city authority may negotiate a price with the private provider by seeking a lower price than the monopoly one[16].

3.1.2 Intervention in market pricing

The Commission SWD (2018) already suggested that pricing of B2G data transfers could be linked to the public interest purpose pursued and the social need it aims to fulfil. In other words, the compensation paid to the private data supplier should vary in function of the expected social welfare benefits. Firms supplying data to a more important, urgent and valuable public purpose would receive a lower price or compensation for their data because it would be used to tackle a highly relevant public interest matter. The B2G Expert Group has further developed this reasoning and suggested a range of public interest data sharing projects where free data, marginal cost pricing or marginal cost plus a return-on-investment mark-up would apply as pricing rules in all B2G data transactions, irrespective of the data market structure[17]. Payment of normal market prices would still be allowed in a certain number of circumstances.

The economic logic of marginal cost pricing for data was first developed in the PSI Directive as a means to facilitate G2B data sharing and wider use of government data. The underlying assumption was that (high) prices could reduce demand and the number of G2B data sharing operations. Users with a low willingness and/or ability to pay for the data would not be able to access them. Lowering the price paid for data could thus increase the number of operations. Pollock (2008, 2010) investigates the economic impact of moving from average to marginal cost pricing rules in G2B data sharing (PSI) for some specific cases of government agencies. Pollock (2010) then extrapolates the micro-findings to the macro-economic level by means of a multiplier effect. These studies do indeed show that marginal cost pricing lowers the social welfare losses compared to monopolistic pricing (See Figure A1a). These PSI studies do not consider the supply side impact of lower prices on the incentive to invest in data collection because the data were already collected by the government for other purposes, irrespective of any alternative uses. The incentive effect could therefore be ignored. A similar reasoning can be applied to data collected by firms for reasons related to their main line of business and without regard to alternative uses. The cost of collecting, storing and processing the data is already covered by the main activity of the firm. From this perspective, sharing the data with others at a marginal or zero cost would not affect the profitability of the main business of the firm.

Data holders may initially not be using their data for other purposes than those needed to sustain their main business, but they may become increasingly aware of the potential value of selling their data for other businesses. They can start developing new data-related revenue streams outside their traditional business model. For instance with the help of APIs they can launch new data-driven services in the hope that these will scale and amortize the fixed set-up costs in order to produce new revenue streams. The banking sector is a good example. Spanish bank BBVA started an API market as an intermediary platform that enables new B2B commercial data transactions related to banking operations and customers, using standardised interfaces. Selling data can therefore become a significant source of new revenues for data-intensive firms.

Most of these new data market opportunities are situated in the B2B realm, though some may also involve governments and public bodies as clients. In order to scale up the volume of B2G transactions, the Expert Group has considered the possibility of intervening in the pricing conditions for B2G transactions, i.e. a marginal cost pricing rule, possibly with a return-on-investment top-up margin. The underlying reasoning would be that lower prices would increase the purchasing power of governments to buy more data and thus expand the use of private data in the public interest. The firm would in any case continue to collect, store and use the data to pursue with its normal main business and could also start selling the data in B2B settings as

[16] Another example from the JRC's own experience: The JRC bought very detailed data on online and offline sales of electronics products in several European countries from a marketing firm for the purpose of informing an EU policy decision. The firm was in a monopoly position since the data were quite unique and not available from other firms. The JRC therefore had to negotiate a reasonable price with the monopolistic supplier and, despite the efforts, still paid a considerable amount of money. The data provider sold the data to other firms interested in understanding their niche markets. Hence, we can safely assume that the price paid was close to monopolistic levels.

[17] The concept of "marginal cost" is a theoretical economic construction that is often not so easy to calculate in practice. The additional costs generated by preparing a dataset for a new purpose may not be easy to distinguish from other applications of the dataset.

additional, new business activities. Private firms could sell this data at competitive market prices to other firms and they would in this new market activity not be deprived of additional data revenue.

Another consideration concerns competition in public procurement markets. Many government activities require the public procurement of goods, services and labour. This is usually done on market terms through tender and procurement procedures, including for data procurement. For example, some European Commission services, and particularly the JRC, buy access to private data sources for their recurrent activities through competitive tender procedures[18]. Governments may divert from these procurement procedures when there is no competitive market to supply the goods or services and they may replace it with alternative mechanisms such as framework contracts, supply arrangements, joint ventures or strategic alliances, among others. The structure of the market is thus an important issue in the choice of the appropriate mechanism, which in turn can have an impact on the final prices to be paid[19].

There is an implicit assumption that competitive procurement is the most economically efficient method, but this is not necessarily always the case. Indeed, there are many situations where the conditions for competition are not suitable or projects are complex or urgent to use public procurement. For example, if there is no active market on the type of data that governments seek to reuse, public procurement is a difficult endeavor. Companies usually tend to give a very high estimation of how much their data is worth. However, the actual value of data is not based on the data itself but on the outcomes of the potential combinations of that data with other data sources for a specific purpose. Having this in mind, the usefulness of data for the intended public interest purpose cannot be guaranteed ex-ante, which makes public procurement procedures delicate as well. If that data is finally not useful, there is a clear societal loss. Furthermore, without knowing that usefulness, governments find it difficult to set a fair maximum price in the call for tenders.

Critically, many of the public interest purposes that may be attained through B2G collaborations require access to many datasets from many different companies. This is either due to their inherent complex purpose or to the methodological need to ensure representativeness of the results. For example, if a government agency would like to understand the increased mortality of bees, officials and/or researchers would need data from various sources (e.g. data from the beekeepers, Common Agricultural Policy registrars, pesticides sellers, land owners, agricultural data, weather data, etc.). Multiple parallel procurement procedures cannot possibly ensure that all these data are accessed on time, not to mention the high cost of each individual procured dataset that in itself is not useful unless combined with the rest of the datasets. If one or several of those potential sources would reject access to its data, there is a direct impact on the real value of all procured datasets, with important societal loss.

Finally, with procurement of data, it is never clear whether a significant number of well qualified, interested contractors may participate in the tender or that the tenders will be of a sufficient size in order to justify the high transaction costs implied by procurement procedures. For these reasons, for many public interest purposes, procurement of data does not provide welfare gains for society and alternative methods should be considered.

Data are not legally protected by IPR rights, unless they would fall under the EU Database Directive that recognizes a copyright on original creatively designed databases or a sui generis right on other types of databases, subject to strict conditions[20]. However, apart from the legal argument, the economic argument of the incentive effect of protecting databases is well recognised. Indeed, this was the main motivation for setting up the EU Database Directive. The incentive argument is probably stronger for firms that sell data as their main business compared to firms who sell data as a by-product of their main activity. Indeed, the CJEU

---

[18] EU Directive on Public Procurement, EU 2014/24. It is not clear to what extent the provisions of the EU Directive on public procurement apply to data procurement. The Directive underlines that there is considerable legal uncertainty as to how far contracts concluded between entities in the public sector should be covered by public procurement rules. The Directive lists many exceptions to the rules, including for "activities of general economic interest" (Whereas 20) and for "electronic communications" (Art 8). EU commitments under the WTO Government Procurement Agreement should also be taken into account. See https://www.wto.org/english/tratop_e/gproc_e/gp_gpa_e.htm

[19] The European Commission's own procurement rules for example explicitly require a detailed market description and justification of the monopolistic position of the supplier before a direct agreement can be negotiated with a supplier, by-passing the standard procurement procedures.

[20] Governments have the possibility to request a court to grant a compulsory license for the patented product at non-monopolistic preferential prices and in the public interest. This may happen for medicines in developing countries and in medical emergencies. Normally, government health services or government-supported medical insurance schemes pay the monopolistic price and may subsidize low-income consumers to access the medicines.

has made a difference between these situations in its case law on the *sui generis* right of the EU Database Directive, and has given a narrow interpretation to this right[21]. Therefore, the 2018 Commission evaluation on the EU Database Directive has concluded that the *sui generis* right does not apply to most situations in the data economy[22]: databases which are the by-products of the main activities of an economic undertaking ('spin-off' databases) are in principle not protected by the *sui generis* right, as they would not fulfil the 'substantial investment' threshold of article 7[23].

It is sometimes claimed that the public interest of government activities is a justification for preferential pricing[24]. In fact, all government activities are meant to produce goods and services that are in the public interest and welfare enhancing for society; if not there is no reason for government to undertake that activity. The Expert Group discussed a range of interpretations of "public interest", from humanitarian emergencies and climate change to more practical applications such as statistics and ordinary government policy decisions. The public interest can be defined in economic terms as any activity that increases the welfare of society as a whole or the welfare of a disadvantaged group in society in order to address equity concerns. There are many economics handbooks that propose methods to assess and quantify the public interest (for example Auerbach and Feldstein, 1985; Diewert, 1983; World Bank, 1998). Public bodies, including the Commission, do impact assessment in order to assess the welfare impact of proposed policies.

3.2 Ex-ante transaction costs and ex-post risks

An important barrier for data sharing comes from the ex-ante costs related to finding a suitable data sharing partner, negotiating a contractual arrangement, re-formatting and cleaning the data, etc. Potentially interested public bodies may not be aware of available datasets or may not be in a position to handle them or understand their advantages and disadvantages. Given that data is an experience good, there may also be ex-post risks related to uncertainties in the quality and/or usefulness of the data, the implementation of the data sharing deal, ensuring compliance with the agreed conditions and the risk of data leaks to unauthorized third-parties and exposure of personal and confidential data. There are at least two ways to deal with these ex-ante transaction costs and ex-post risks[25]: (a) using trusted technical intermediary platforms to manage the data transfer operation and (b) upstream management of the "quality" of the data that are transferred to a downstream user.

3.2.1 Trusted technical intermediaries

Technological and institutional arrangements, and regulations, can reduce ex-ante transaction costs and ex-post risks and thereby enable more voluntary transactions to take place, without necessarily making them mandatory (IDC, 2015, p 27). The presence of an intermediary that reduces ex-ante transaction costs and ex-post risks, for example by means of standardized contractual provisions, may improve the enabling environment for data sharing. Verhulst, Young and Srinivasan (2015) discuss several means to reduce transaction costs such as APIs with standardised datasets and access protocols, data cooperatives and data

---

21 Judgment of the Court (Grand Chamber) of 9 November 2004. Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE (OPAP). Reference for a preliminary ruling: Monomeles Protodikeio Athinon - Greece. Directive 96/9/EC - Legal protection of databases - Definition of database - Scope of the sui generis right - Football fixture lists - Betting. Case C-444/02. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62002CJ0444.

22 Commission Staff Working Document, Evaluation of Directive 96/9/EC on the legal protection of databases, 2018/04/25, SWD(2018) 146 final.

23 Regarding the *sui generis right*, the investment must be qualitatively or quantitatively substantial. The CJEU has ruled that the expression «investment in the obtaining of the contents» must be understood to refer to «the resources used to seek out existing independent materials and collect them in the database», and that «it does not cover the resources used for the creation of materials which make up the contents of a database» (CJEC, Grand Chamber, 2004/11/09, The British Horseracing Board Ltd and Others v William Hill Organization Ltd, C-203/02 ; Fixtures Marketing Ltd v Oy Veikkaus Ab, C-46/02).

24 Public interest should not be confused with public goods and services. For example transport, education and health services are produced by a mix of private and public providers in most countries. They are not public goods because they are rival and excludable and can thus be monetized. This allows private service producers to step into the market and compete with government services.

25 "Ex-ante" refers to "before the B2G data transaction is concluded; "Ex-post" means after the contract has been signed, during the implementation phase.

pools that facilitate exchanges, etc. For example the creation of "data stewards"[26] on the business and government side of the market could increase the volume of voluntary market-based B2G data sharing operations. Data cooperatives and pools, such as SODA ("Scalable Oblivious Data Analytics") and OPAL ("Open Algorithms") are trusted intermediary institutions that may enable a wider range of data suppliers and users to set up a common arrangement to share data, and to share the costs and benefits of this sharing arrangement. The OPAL initiative "sends the algorithms to the data, not the other way around" to avoid misuse and exposure of personal data. SODA seeks to establish a trusted intermediary that can carry out data analytics in an environment that requires a combination of data protection and multi-party data sharing and avoids unauthorised access to commercially sensitive data. The Frauenhofer Institute's Industrial Data Space[27] offers similar trusted intermediary facilities for multiple parties with secure data processing and strictly regulated access and billing. Reducing the lack of data-analysis skills in government may also reduce transaction costs but this does not matter so much when the data are delivered indirectly as a service rather than a direct raw data delivery.

Third-party technical platforms as trusted intermediaries may facilitate data transfers, pooling, processing and management. As such they reduce transaction costs and risks for the parties involved in the B2G data transfer. However, technical platforms are not a substitute for reaching an agreement on access and use conditions between the parties (who has access to which data under which conditions) and the allocation of costs and benefits between the parties (who pays for what and who can claim which revenue or compensation). Such an agreement requires negotiations between the parties involved. Data suppliers to the pool or platform will have to find satisfactory arrangements that deal with uncertainties and potential costs associated with these uncertainties. In general, initiatives that reduce transaction costs and risks for data providers can help to promote voluntary B2G transactions.

Sometimes a platform operator may impose a standard agreement to which all parties have to adhere. There are examples however where such standard conditions for B2G data sharing arrangements failed because of insufficient attention to ex-ante transaction costs and ex-post risks. One example is the Hitachi – Copenhagen City Data Exchange. Following a public tender, the city of Copenhagen provided seed financing to Hitachi to build a data sharing platform for many types of data between many providers and users. Hitachi became the platform operator and broker between data suppliers and users. It invested in reducing transaction costs for users by searching for data, formatting and preparing them for posting on the platform, ensuring access, protecting privacy and security, and bringing the data together in a single platform that made it easier for potential users to find what they were looking for. However, the platform could not set the price and access conditions to the data. The City imposed as entry requirement that any user would get access on equal conditions. That prevented suppliers and users from negotiating specific access conditions and prices for each user. Suppliers had no leverage over who could access the data. Consequently, ex-post risks and opportunity costs were not manageable in the platform. This made data providers reluctant to make their data available. The incentives for data sharing were too weak compared to the perceived risks. As a result the platform did not scale up (IDC, 2015, p15-19).

3.2.2 Quality: management of the transformation of the transferred data

Data can be transferred in different varieties. The data can be transferred "raw", i.e. in the original format in which they were collected. They can be anonymized and/or aggregated to a more coarse-grained level from which the original raw data cannot be reconstructed. Data analytics can be applied to derive insights from the original data that can be sold as a service to the client. These variations in the data taxonomy can also be used to manage data governance for the purpose of reducing transaction costs and risk, or for reaching a negotiated agreement on data access and use conditions.

Data are usually not consumer end products. We do not search online for a train schedule because we like looking at train schedules but because we want to purchase a transport service, i.e. a train journey. We therefore say that data are usually intermediate inputs into the production of a service or a service transaction. Data transactions create a link between an upstream data market and a downstream data-driven services market. This implies that there are two modalities for sharing or trading data: direct and indirect. With direct data trade the data holder delivers data as an input to a service producer. With indirect data trade

[26] See the GovLab initiative on data stewards http://thegovlab.org/data-stewards-data-leadership-to-address-the-challenges-of-the-21st-century/

[27] See http://www.industrialdataspace.org

the data holder delivers a data-based service as an input for the downstream service provider. With direct trade, data are changing hands; with indirect trade there is no transfer or sharing of the original (raw) data.

Many B2G data sharing examples circulated in the Expert Group are examples of indirect data trade. For example, using mobile phone data to help public transport companies improve their schedules or help tourism promotion services assess the presence of foreign tourists in a city. Only derived, aggregated and processed outputs are shared with a public body in these cases, not the original raw data. The debate in the B2G Expert Group also confirms that there are other dimensions too, apart from the direct-indirect access dichotomy. For example, data sharing can include site or remote access, sending an algorithm to the supplier to extract the required insights, sharing via a trusted third party, continuous direct access via APIs, among others.

The distinction between direct and indirect data (service) trade is not always a crystal-clear black-and-white border line, as the train schedule example demonstrates. A search engine processes the underlying raw data of a train schedule to make them more readable and useful to the user who gets better insights into his travel options. Making them searchable by route, date and time and adding a ticket payment service to the search function creates an additional value-added service for consumers. The train schedule data transfer becomes a more indirect data service as more value-added features are added to the original raw data. In general, further processing of the data and adding other complementary services increases the value to the end user.

The B2G Expert Group proposes a four level data taxonomy: raw or primary data, pre-processed (cleaned and structured data), processed (for a particular purpose) and data-driven insights (no data, only data services)[28]. Higher levels in the taxonomy may sometimes coincide with a higher degree of data aggregation and anonymization of personal data. Trusted intermediary data platforms can play an important role in reducing transaction costs and risks in a B2G data transfer by doing the processing in-house and transferring only insights or data-based services to the public body that is the end user. A good example of such a set-up is the Dalberg Group[29], a firm that offers B2G data-based services and insights to public bodies in developing countries. These services are usually paid for by aid agencies.

The processing-based approach to data taxonomy or data quality may give the impression that adding value to data is a cumulative linear process. In many cases however, value is added when datasets from various sources are aggregated and result in economies of scope (Rosen, 1983, Duch-Brown et al, 2017): the insights and value derived from analysing the combined set is greater than the value that could be gained from analysing each dataset separately. For example, the value of online advertising services can only be realized by combining consumer search data with data on the supply of ads offered by advertisers. Using each dataset separately would not permit setting up a targeted advertising service. Another example are urban transport and mobility platforms that require data aggregation across several transport service providers (trains, buses, metro, taxis, etc.) in order to manage peak transport demand in dense urban areas. If economies of scope are important, the choice between a direct or an indirect data service depends on which party is in a better position to do the aggregation, the original data holder or recipient. In a B2G setting, governments may want to combine acquired data with other datasets that they already hold. For example, they may want to combine mobility data with infrastructure, land ownership, income tax or education data. In that case a trusted intermediary data platform (OPAL, SOPA etc.) may be an appropriate solution to deal with the complexity of data confidentiality, strictly regulated access conditions, etc. (see below under transaction costs).

3.3 Externalities and incentive misalignment

In this section we analyse the impacts of externalities and incentives on B2G data sharing initiatives.

[28] The Expert Group's data taxonomy specifies that data, information and insights are not synonyms and that the latter are generated by analysing information and drawing conclusions. "Insights" also consist of data points that are shared with a recipient. The value added by the service resides in the processing and analysis of the original primary data that has led to these insights. The original data can usually not be reverse engineered from the insights. From a mathematical perspective some data entropy has been lost in the processing. Extracting "insights" can therefore be seen as reducing the entropy of a dataset. See https://en.wikipedia.org/wiki/Kolmogorov_complexity

[29] See https://www.dalberg.com/

3.3.1 Negative externalities and risks

Firms may be reluctant to share data with governments because it might have a negative impact on them. This could be due to perceived risks of market regulation and competition in their upstream or downstream markets.

An example of perceived regulatory risks comes from our own experience at the JRC. When we sought to procure data from parcel delivery firms about cross-border deliveries of e-commerce packages in the EU the firms refused to provide the data because they feared they might be used by the Commission to impose new regulatory measures on the sector. Indeed the Commission was exploring these questions as part of its Digital Single Market policy package and eventually implemented new policy measures. Another example from the JRC experience concerns the procurement of data on app use from a private firm. The firm imposed strong purpose limitation conditions that would not have allowed us to use the data in a publicly available document, not even in an aggregated way. That stopped the transaction from taking place.

It is often assumed that commercial risks only occur in B2B data sharing situations, not in B2G settings because public bodies are normally not in the business of competing with private firms. However, some public bodies may run commercial services that compete with private services, for example in public transport and utilities. Public bodies may also pool data from various private and public sources and share the insights with competing parties. Mobility-as-a-Service (MaaS) urban transport platforms (Carballa, 2018) are a good example of conflicting incentives for private firms to share data with public bodies. MaaS platforms combine information about, and access to, multiple transport service providers in a single interface. This requires private and public transport operators to share data with a publicly-operated platform. In doing so, these transport operators cooperate and compete at the same time: they compete over rides and market shares in transport while they collaborate to build a common data pool and service that can bring them all more rides. Carballa (2018) shows that, although every transport service provider has incentives to share data provided that a critical mass of competitors do likewise, the cooperative-competitive dynamics of data sharing can lead to situations where some transport operators will not have an incentive to share their data because they may lose market share and revenue.

3.3.2 Positive externalities and economic incentive misalignment

While B2G data sharing may produce benefits for society at large or for particular target groups of the operation, the private firm that supplied the data may not experience any of these benefits directly. In other words, a B2G data transfer for a project that is in the public interest generates positive externalities that often cannot be directly monetized by the data supplier. If the data collection and transfer to a public sector body is the firm's main business activity, there would be a misalignment of economic incentives. It would create a gap between private incentives to supply data to public bodies and the social benefits of the operation that do not accrue to the data supplier but are dissipated among the population at large.

There is a long-standing economics research literature on the private provision of public goods[30] (see for example Bergstrom et al, 1983) that discusses in detail these incentive misalignments due to uncaptured externalities from public goods. It also explains that the private provider may have a personal preference for the provision of a public good and derive benefits from doing so. In that case it adds to his benefits and not only to his costs. In B2G data sharing operations this can be translated into Corporate Social Responsibility (CSR) and philanthropic motives for private firms to contribute data in the public interest. In that case firms may voluntarily choose to ignore the incentive gap between private costs and social benefits for CSR motives. Private contributions may be forthcoming but they are unlikely to reach the socially desirable volume of public goods production. Hence government intervention may be required to reach this socially desirable level of public good provision.

The Com SWD (2018) and the Expert Group have invoked the concept of proportionality in an attempt to contain the gap between private costs and public benefits of B2G data sharing. Proportionality compares the costs for the data provider with the benefits for society. The SWD suggests that the ratio of costs to benefits should be kept within reasonable proportions. For example, the extent of fine-graining of the data should not

[30] These are goods that are non-rival but also non-excludable. This means that the use by one individual does not reduce availability to others and that individuals cannot be excluded from use. See https://en.wikipedia.org/wiki/Public_good

exceed what is necessary for the purpose of the public project because that would make costs for the supplying firm disproportional to the benefits for society. In general, any activity in the public sector should meet the criterion that total benefit to society should be greater than the cost; if not then the activity is not social welfare enhancing for society. However, the point in B2G data transactions is not only to compare the total costs and benefits for society as a whole but also to examine who bears the costs and who enjoys the benefits. In the case of B2G data transfers, costs and benefits may be attributed to different parties: the private costs are born by the data provider while the social benefits are enjoyed by society as a whole or a selected target group. There are several mechanisms to compensate the costs incurred by private data providers that are ultimately a political decision.

4 How to overcome these market failures and boost welfare gains?

There is a nascent market for B2G data transfers that contribute to the promise of socio-economic welfare gains from data sharing. However, most of these remain one-off experimental pilot projects that do not seem to be sustainable over time and the emerging market does not seem to scale. There are cases where the market fails to realise B2G data sharing transactions that are "in the public interest" i.e. that increase social welfare outcomes for society, often in cases where firms are not in the business of directly selling their data or making them available to third-parties. Provided the social welfare benefits of the operation exceed the private costs for the data provider (data production costs and possible adverse effects in the market), it would be in the interest of society to nevertheless go ahead with the data sharing transaction. That requires finding a means to overcome the obstacles to the transaction. If there is no operational market for the data yet, how can this missing market be created? In this note we discuss three types of obstacles: monopolistic data suppliers and monopolistic pricing, high transaction costs and risks for the data supplier, and incentive misalignment between the private provider and the public interest. We also suggested some potential solutions to these obstacles, including intermediary service providers that could reduce transaction costs and risks, and managing the taxonomy or quality of the data transfers to avoid transmission of personal or commercially sensitive data.

One mechanism to solve some of the barriers that may be holding back B2G data sharing initiatives would be to provide stronger incentives for the data supplier firm to engage. These incentives can be direct (i.e., monetary) or indirect (i.e., reputational). Another way is to ascertain the data transfer by making the transaction mandatory, with cost compensation (e.g. marginal costs or marginal costs plus fair ROI). Mandatory data sharing and/or price regulations are an example of this second solution. An intermediate way would be solutions that seek to facilitate voluntary B2G operations without mandating them, for example by reducing the transaction costs and perceived risks for the provider data supplier with trusted data intermediary platforms, or appropriate contractual provisions, for instance. A possible EU governance framework for B2G data sharing operations could cover all these options. The main issue then is to set criteria for which transactions fall into which type of B2G governance box.

For situations with monopolistic data supplies and high monopolistic pricing of the data, or refusal of access to the data, a preferential pricing rule and/or mandatory access to the data could be regulated, e.g. by introducing a specific and mandatory regulation for the purpose of B2G data transfers, including a (zero or marginal cost) pricing rule. There are already mandatory B2G data transfers in place. For example, governments make it mandatory for firms to hand over some of their accounting data for tax purposes, fill in questionnaires for statistical purposes and report on their compliance with environmental regulations, and banks have to provide detailed reports on their credit portfolios to financial supervisors. In the EU, DG COMP already has powers to force firms to hand over data in the context of specific competition cases but also for wider sector enquiries. This was used for example for collecting data in the 2017 E-commerce sector enquiry[31]. DG GROW recently tried to obtain similar powers in the context of the implementation of Single Market policies (the so-called SMIT proposal) but failed. A mandatory regulatory framework of B2G data transfers under specific conditions could thus be considered. This could also cover situations where a private data supplier faces strong negative incentives to participate in a B2G data transfer, despite the fact that the public interest or social welfare impact of an operation would considerably exceed the private costs of the supplier. If these private costs would not undermine the main business model of the supplier, a case could be made to move ahead with the operation and make it mandatory. Still, it would require some legal and possibly judiciary oversight on a case-by-case basis[32].

[31] See http://ec.europa.eu/competition/antitrust/sector_inquiry_final_report_en.pdf

[32] In particular for its compatibility with rules on IP ownership and the protection of trade secrets.

References

Alemanno, A. (2018), Data for Good: Unlocking privately-held data to the benefit of the many. European Journal of Risk Regulation 2.

Auerbach, A. and M. Feldstein (1985), Handbook of public economics

Bergstrom, T., L. Blume and H. Varian (1986), On the private provision of public goods, Journal of public economics.

Carballa-Smichowski, B. (2018), Determinants of coopetition through data sharing in MaaS. Management and Data Science, vol. 2(3) 2018.

Diewert, W. (1983), Cost-benefit analysis and project evaluation: A comparison of alternative approaches, Journal of public policy.

Duch-Brown, N., B. Martens and F. Mueller-Langer (2017), Data ownership, access and trade, JRC Digital Economy Working Paper 2017-01.

Deloitte (2011), Pricing Of Public Sector Information: Models of Supply and Charging for Public Sector Information, DG CNECT, October 2011.

Drexl, J. (2016), Designing competitive markets for industrial data: Between propertisation and access, Max Planck Institute for Innovation and Competition Research Paper No. 16-13.

Elkin-Koren, N. and M. Gal (2018), The Chilling effect of governance by data on innovation. University of Chicago Law Review 86.

European Association of Remote Sensing Companies (EARSC) (2015) Copernicus Sentinels' Products Economic Value: A Case Study of Winter Navigation in the Baltic, September 2015, available at http://earsc.org/news/between-24m-and-116m-per-annum-of-economic-value-generated-thanks-to-the-use-of-satellite-radar-images

European Data Portal (2015), Creating value through open data.

International Data Corporation and the Lisbon Council (2016), Opening Up Private Data for Public Interest, a European Data market study.

Graef, I., S. Y. Wahyuningtyas and P. Valcke (2015), Assessing data access issues in online platforms, Telecommunications Policy vol 39, pp 375–387.

Jones, C. I. and C. Tonetti (2018), Nonrivalry and the Economics of Data . Stanford Graduate School of Business Working Paper 3716.

Newbury, D., L. Bently and R. Pollock (2008), Models of public sector information provision via trading funds

Martens, B. and F. Muller-Langer (2018), Access to digital car data and competition in aftersales services, JRC Digital Economy working paper 2018-06.

OECD (2015), Data-driven innovation: Big data for growth and well-being, OECD, Paris. http://dx.doi.org/10.1787/9789264229358.en

Palfrey, J. and U. Gasser (2012), Interop: The Promise and Perils of Highly Interconnected Systems.

Panzar, J. and R. D. Willig (1981), Economies of Scope, American Economic Review Vol. 71, No. 2, pp. 268-272

Pollock, R. (2007), Cumulative innovation, sampling and the hold-up problem. MPRA Paper 5022, University Library of Munich, Germany.

Rochet, J. C. and J. Tirole (2006), Two-sided markets: a progress report, Rand Journal of Economics, 37(3), 645–667.

Rosen, S. (1983), Specialization and human capital, Journal of Labor Economics, 1(1), 43–49.

Verhulst, S., A. Young and P. Srinivasan (2015), An introduction to data collaboratives, New York GovLab.

WISE DEMOS (2011), Big and open data in Europe: a growth engine or a missed opportunity?

World Bank (1998), Handbook on the economic analysis of investment operations.

Glossary of economic terms

Bilateral monopoly: a bilateral monopoly is a vertical market structure consisting of both a monopoly on the supply-side (a single seller) and a monopsony on the demand-side (a single buyer). (https://en.wikipedia.org/wiki/Bilateral_monopoly)

By product: a secondary or incidental product, as in a process of manufacture. (https://www.dictionary.com/browse/byproduct)

Competitive market: a market characterised by no barriers to the entry of new firms, so numerous producers compete to provide homogeneous goods and/or services. In this situation, consumers have plenty of choice to satisfy their needs. In general, nor single firm or consumer have the power to influence the market concerning either prices that will be set or quantities that will be produced. This creates the most advantageous framework, maximising allocative efficiency. (https://study.com/academy/lesson/competitive-market-definition-characteristics-examples.html; https://www.tutor2u.net/economics/topics/competitive-market)

Consumer surplus: it is the difference between the price that consumers pay and the price that they are willing to pay. On a supply and demand curve, it is the area between the equilibrium price and the demand curve. (https://www.economicshelp.org/blog/188/concepts/definition-of-consumer-surplus/)

Deadweight loss: It is the loss of economic efficiency in terms of utility for consumers/producers such that the optimal or allocative efficiency is not achieved. Deadweight loss can be stated as the loss of total welfare or the social surplus due to reasons like taxes or subsidies, price ceilings or floors, externalities and monopoly pricing. It is the excess burden created due to loss of benefit to the participants in trade which are individuals as consumers, producers or the government. (https://economictimes.indiatimes.com/definition/deadweight-loss).

Demand curve: represents the relationship between the price of a good or service and the quantity demanded at a given point in time. All else being equal the quantity demanded decreases as the price of a given commodity increases. (https://www.investopedia.com/terms/d/demand-curve.asp)

Differentiated products: differentiated products are the outcome of a marketing strategy that aims to distinguish a company's products or services from the competition. Successful product differentiation identifies and communicates the unique qualities of a company's offerings while highlighting the distinct differences between those offerings and others on the market. Product differentiation goes hand-in-hand with developing a strong value proposition to make a product or service attractive to a target market or audience. (https://www.investopedia.com/terms/p/product_differentiation.asp)

Economies of scale: refers to the cost advantage that firms attain when increasing their output because the per-unit fixed cost decreases as the units produced increase (i.e. the greater the quantity of output produced, the lower the per-unit fixed cost). Moreover, increasing the output also determines a drop in average variable costs because of operational efficiencies and synergies that originate from a growth in the scale of production. (https://corporatefinanceinstitute.com/resources/knowledge/economics/economies-of-scale/)

Economies of scope: while economies of scale emerge by producing a higher volume of the same product, economies of scope arise by producing 'more different types' of products using the same resources. The long-run average and marginal cost of a company decreases due to the production of some complementary goods and services. While economies of scope relate to efficiency along the dimension of variety, economies of scale relate to efficiency along the dimension of production volumes. (https://www.shopify.com/encyclopedia/economies-of-scope; https://www.investopedia.com/terms/e/economiesofscope.asp)

Excludability: this concept means that producers can prevent some people from consuming the good or service based on their ability or willingness to pay. In other words, it is possible to prevent people (consumers) who have not paid for the good/service from accessing it. (https://www.britannica.com/topic/excludability ; https://en.wikipedia.org/wiki/Excludability)

Externality: this is a cost or benefit incurred or received by a third party although the third party has no control over the creation of that cost or benefit. An externality can be either positive or negative, and it can originate from either the production or consumption of a good or service. The costs and

benefits can be both private—to an individual or an organization—or social, meaning it can affect society as a whole. (https://www.investopedia.com/terms/e/externality.asp)

Experience good: a good that can be evaluated only after it has been purchased and experienced like many personal services (e.g. restaurant, hairdresser, beauty salon, theme park, travel, and holiday). (https://en.wikipedia.org/wiki/SEC_classification_of_goods_and_services)

Fixed costs: this type of cost does not change by increasing or decreasing the amount of goods or services produced and sold. (https://www.investopedia.com/terms/f/fixedcost.asp)

Incentive effect: the result of policy design that emphasizes rewards and penalties, like stimulating a greater effort as a result of offering bonuses for reaching set goals. (http://regulationbodyofknowledge.org/glossary/i/incentive-effect/)

Intermediate input: a good or services, different from fixed assets, employed as input into the production process. An intermediate input can be either produced elsewhere in the economy or imported, and it can be either transformed or used as it is in the production process. (https://stats.oecd.org/glossary/detail.asp?ID=1395)

Joint production: joint production is a production process that yields two or more products simultaneously. (https://wiki.scn.sap.com/wiki/display/ERPFI/Joint+Production)

Marginal cost: the cost added by producing one additional unit of a product or service. The marginal cost of production is the change in total cost that comes from making or producing one additional unit. Marginal costs determine at what point a producer can achieve economies of scale to optimize production. Until the marginal cost of producing one additional unit is lower than the per-unit price the producer has room to gain a profit. (https://www.lexico.com/definition/marginal_cost ; https://www.investopedia.com/terms/m/marginalcostofproduction.asp)
For more information on how marginal costs are understood in the context of the Open Data Directive, please visit https://ec.europa.eu/digital-single-market/en/news/commission-notice-guidelines-recommended-standard-licences-datasets-and-charging-re-use

Marginal-cost pricing: the practice of setting the price of a product to equal the extra cost of producing an extra unit of output. By this policy, a producer charges, for each product unit sold, only the addition to total cost resulting from materials and direct labour. (https://www.britannica.com/topic/marginal-cost-pricing)

Market failure: refers to a situation in which the allocation of goods and services by a free market is not efficient, often leading to a net loss of economic value. Market failures are often associated with the existence of public goods, time-inconsistent preferences, information asymmetries, non-competitive markets, principal–agent problems, or externalities. (https://en.wikipedia.org/wiki/Market_failure)

Market power: the power a firm (or group of firms) to raise and maintain price above the level that would emerge if the market were competitive. Market power leads to reduced output and loss of economic welfare. (https://stats.oecd.org/glossary/detail.asp?ID=3256)

Market share: this is the percentage of total sales in an industry held by a firm. The market share is the company's sales over a period divided by the total sales of the industry in the same period. The market share convey the size of a company vis-à-vis its market and its competitors. (https://www.investopedia.com/terms/m/marketshare.asp)

Mark-up: this is the amount surcharged on the product or service marginal cost to make the selling price. (https://www.investopedia.com/ask/answers/102714/whats-difference-between-profit-margin-and-markup.asp)

Monopsony: a market structure where a single buyer controls the market as the major purchaser of goods and services offered by many sellers. (https://www.investopedia.com/terms/m/monopsony.asp ; https://en.wikipedia.org/wiki/Monopsony)

Monopoly: a market structure where a single seller controls the market as the sole producer of goods and/or services offered to many buyers. (https://www.investopedia.com/terms/m/monopoly.asp)

Non-rivalry: this concept means that one person's consumption of a product does not reduce the amount available for consumption by another. (https://en.wikipedia.org/wiki/Rivalry_(economics))

Perfect competition: pure or perfect competition is a theoretical market structure consisting of the following criteria: i) All firms sell an identical product; ii) All firms are price takers (they cannot influence the market price of their product); iii) Market share has no influence on prices; iv) Buyers have complete or "perfect" information—in the past, present and future—about the product being sold and the prices charged by each firm; v) Resources for such a labour are perfectly mobile; vi) Firms can enter or exit the market without cost. (https://www.investopedia.com/terms/p/perfectcompetition.asp)

Price discrimination: it is a selling strategy to charge different prices for the same product or service based on what the seller thinks that a specific customer is willing to pay. In pure price discrimination, the seller charges each customer the maximum price he or she will pay. (https://www.investopedia.com/terms/p/price_discrimination.asp)

Public good: a good that an individual can consume without reducing its availability to others (i.e. non-excludable good) and that can be enjoyed/consumed by more than one individual at time (i.e. non-rivalrous good). Examples of public good are law enforcement, national defence, public parks. (https://www.investopedia.com/terms/p/public-good.asp)

Public interest: refers to the welfare of the public as compared to the welfare of a private individual or company. All of society has a stake in this interest and the government recognises the promotion of and protection of the general public. (https://thelawdictionary.org/public-interest/). However, as recognised by the B2G Expert Group: "while 'public interest' broadly refers to the welfare of individuals in society, its exact boundaries remain largely undefined, being heavily dependent on socioeconomic, cultural and historical factors. While some purposes (such as public health) are generally recognised as being in the public interest, other purposes (such as improved public services) might be less clear-cut."

Profit maximization: classical economics assumes that firms seek to maximise profits (i.e. profit = total revenue – total costs). Therefore, profit maximisation occurs at the biggest gap between total revenue and total costs. (https://www.economicshelp.org/blog/3201/economics/profit-maximisation/)

Scalable: scalability is a characteristic of a system, model, or function that describes its capability to cope and perform well under an increased or expanding workload or scope. A scalable business can maintain or improve profit margins while sales volume increases. (https://www.investopedia.com/terms/s/scalability.asp)

Search costs: these costs account for the time, energy and money expended by a consumer who is researching a product or service for purchase. Search costs include the opportunity cost (i.e. effort that could have been devoted to other activities) of the time and energy spent on searching, and of the money spent to travel examining different options, purchasing research data or consult an expert for purchasing advice. (https://www.investopedia.com/terms/s/search-cost.asp)

Social welfare: the group of assistance programs designed to ensure the well-being of citizens. Therefore, social welfare systems aim to provide quality care to society participants. (https://www.myaccountingcourse.com/accounting-dictionary/social-welfare)

Socio-economic value: it seeks to incorporate elements of social value to the economic value creation. An entity creates socio-economic value by employing resources, inputs, or processes in a way that increases their economic value by generating cost savings for the society (i.e. public system or environment) of which the entity is a part. (https://garneringchange.wordpress.com/defining-social-capital/defining-social-and-socio-economic-value/)

Substitute goods: these are two alternative goods that could be used for the same purpose. (https://www.economicshelp.org/blog/glossary/substitute-goods/)

Substitution effect: this effect implies the decrease in the sales of a certain product by the moment consumers switch their preferences and purchases towards cheaper alternatives of the product itself. (https://www.investopedia.com/terms/s/substitution-effect.asp)

Transaction cost: expenses (i.e. time, money, energy, labour) experienced when buying or selling a good or service. These costs give rise to industries dedicated to facilitating exchanges. In a financial sense, transaction costs include brokers' commissions and spreads (i.e. the differences between the price the dealer paid for a security and the price the buyer pays). (https://www.investopedia.com/terms/t/transactioncosts.asp)

Willingness to pay: the maximum price at or below which a consumer will definitely buy one unit of a product. (https://en.wikipedia.org/wiki/Willingness_to_pay)

Welfare gains: these are gains generated by the impact of a government policy, or a decision by firms, on total economic welfare. Net welfare gains are gains discounted by all possible losses. (https://www.economicsonline.co.uk/Definitions/Net_welfare_gain.html)

List of figures

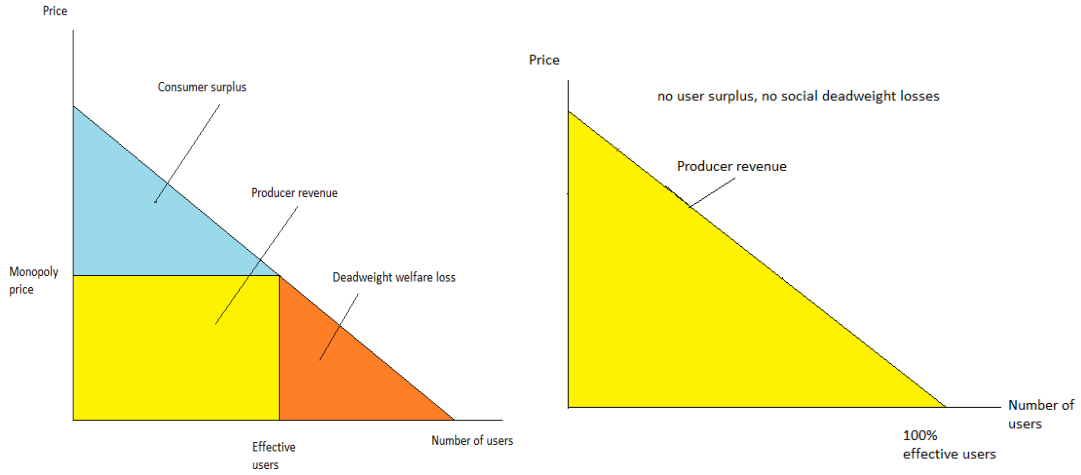
Annexes

Annex 1. Monopolistic pricing and price discrimination

In the left hand panel of Figure A1 a data seller is in a monopolistic position and charges a monopoly price. The effective number of data users is below the full potential. This results in deadweight welfare losses that benefit nobody (orange triangle). On the other hand, all data buyers pay the same monopoly price that is below their willingness to pay for the data. They therefore enjoy a consumer surplus (blue triangle) that the data seller does not capture. The seller receives only part of the value of the data as revenue (yellow area).

Compare this with the situation in the right hand panel of Figure A1 where the data seller knows the willingness to pay of each user and charges them accordingly. It results in perfect first-degree price discrimination: every user pays a different price according to his willingness to pay. This maximizes welfare from data sharing because all potential users effectively get access to the data and there is no deadweight loss (the orange triangle disappears). However, the seller appropriates the entire surplus and users do not benefit from consumer surplus (the blue triangle also disappears). Sellers may not have sufficient information about buyers to practice perfect price discrimination but they may experiment with different qualities[33] or versions of a dataset or a data service, in order to test the market and charge different prices to different groups of consumers[34] (Bergemann and Bonatti, 2018).

Figure A1: The welfare effects of monopolistic pricing and price discrimination



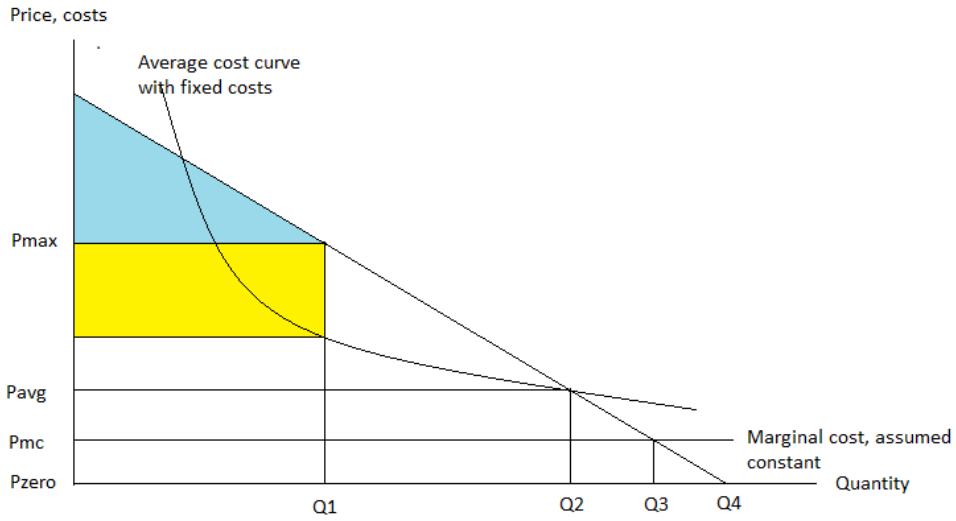
[33] "Quality" should not be interpreted as an error rate in the dataset. It refers to the degree of fine graining or aggregation of the data.

[34] Price differentiation by group rather than by individual is called third-degree price discrimination. Second-degree price discrimination occurs when prices vary by quantity.

Annex 2. Demand and data pricing regimes

How does pricing relate to production costs in a monopolistic market? This is explained in Figure A2. Typically for the digital economy, data producers face high fixed costs in setting up the data collection, storage and processing system. On the other hand, data systems often scale easily so that the marginal cost of processing an additional unit of data is very low or close to zero. It results in declining average costs. A profit-maximising firm with some degree of monopolistic market power will set prices at a profit-maximising level (Pmax where P-average cost is maximum) where it will sell a quantity Q1. In the presence of fixed costs and if a firm operates in a competitive market with little or no market power, it will price its products close to or equal to average cost (Pavg) and sell quantity Q2. In the absence of fixed costs, market pricing could go down to marginal costs (Pmc) and quantity Q3, or possibly Pmc plus a mark-up (also called a return-on-investment or ROI). The maximum quantity is reached when firms give their products away for free (Pzero). Total demand is satisfied only at Pzero, i.e. all potential users get access to the data at this point. In the presence of fixed costs, a preferential price is any price below Pmax (for a firm with market power) or below Pavg (for a firm without market power). Without fixed costs, any price below Pmc is a preferential price. The meaning of preferential pricing depends on the market structure and production costs. When the data market is monopolistic and the price is fixed at Pmax, consumer and producer welfare are defined by the blue and yellow areas under the demand curve. The difference between Q4 and Q1 represents the social deadweight loss, i.e. a number of potential data uses that are not addressed. In a competitive data market (either at Pavg or Pmc) a larger number of potential users are served (Q2, Q3 > Q1) and social welfare from the data expands, to the benefit of data users.

Figure A2: Demand for data at different price regimes



Source: Adapted from Newbury et al, 2008.

GETTING IN TOUCH WITH THE EU

In person

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

On the phone or by email

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

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

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

EU publications

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

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub