EUROPEAN
COMMISSION

Brussels, 25.4.2018
SWD(2018) 125 final

COMMISSION STAFF WORKING DOCUMENT

Guidance on sharing private sector data in the European data economy

Accompanying the document

Communication from the Commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the Regions

"Towards a common European data space"

{COM(2018) 232 final}

1. Introduction

Data-driven innovation is a key enabler of growth and jobs in Europe. The importance of data collected online, the growing importance of data generated by objects connected to the Internet of Things (IoT), the increasing availability of Big Data analytics tools and the emergence of broad availability of certain Artificial Intelligence applications are key technical drivers. The non-rivalrous nature of data, making it possible for the same data to support a range of new products or services or new methods of production, suggests that it can become efficient for companies to share more data they hold with other companies so that the value resulting from the data can be exploited to the maximum.

New data-driven business models building on these technical drivers are an opportunity not only for big companies, but also for SMEs and start-ups in Europe. Similarly, the public sector is beginning to seize the opportunities of data-driven innovation. Businesses already benefit from access to public sector information available as Open Data[1] as well as from sharing data between themselves. However, SMEs and start-ups still face obstacles when either making available their data or re-using data from other companies. This is in particular the case when it comes to machine-generated, non-personal data. Likewise, public sector bodies need to modernise the way they function and exploit the potential of new data sources in order to become more data-driven and cost-efficient. Citizens and business, in particular SMEs, are expected to benefit from this. While in certain cases relevant services based on data can be acquired on the market, in other cases it may be necessary for the public sector to directly analyse data held by a private company or set up regular data acquisition, e.g. for the purpose of official statistics. These data might not always be accessible to the public sector as a result of concerns over data confidentiality or perceived risks to companies' commercial interests. This suggests that questions of data supply and (re-)use ("data sharing") need to be addressed in two situations: business-to-business (B2B) and business-to-government/public sector (B2G).

The Commission has already proposed measures to improve the availability of data for business. With the General Data Protection Regulation (GDPR) and the ePrivacy Directive[2] the EU has set in place a solid framework for the processing of personal data and of electronic communications data, intended to create digital trust which is a key precondition for any data sharing. This framework lays the foundations for a future competitive advantage for European business actors to make the most of data technologies. Moreover, the proposal for a Regulation on the free flow of non-personal data[3] will make it easier to transfer such data within the EU.

---

[1] Including via the European Data Portal https://www.europeandataportal.eu.

[2] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) (OJ L 201, 31.7.2002, p.37). See also: Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), COM(2017) 10 final of 10.1.2017.

[3] COM(2017) 495 final.

With the Communication 'Building a European Data Economy' of 10 January 2017[4], the Commission put forward a first description of potential issues of data access in particular with respect to machine-generated data and with respect to platform-to-business relations. It also mentioned the importance of access to private sector data for public interest purposes.

A broad stakeholder dialogue was conducted on the basis of that Communication. It concluded that the issue at stake did not justify horizontal legislative intervention at this stage and that guidance would be more appropriate.[5]

In the Communication that this Staff Working Document accompanies[6], the Commission defines a series of key principles to be considered so as to make data interactions in business-to-business and business-to-government situations a success for all parties involved.

Additionally, this Staff Working Document aims to provide a toolbox for companies that are data holders, data users, or both at the same time. For this purpose, it contains a "How to" guide on legal, business and technical aspects of data sharing that can be used in practice when considering and preparing data transfers between companies coming from the same or different sectors.

The guidance provided in this document is designed for across all sectors of the economy. As a result of the difference in structure of individual markets, it may need to be complemented by sector specific measures.

Finally, this document does not constitute a statement of the law and is without prejudice to the interpretation of EU law by the Court of Justice of the European Union (CJEU). It does not bind the Commission as regards the application of EU law, in particular with regard to the competition rules in Articles 101 and 102 of the Treaty on the Functioning of the European Union (TFEU).

[4] COM(2017) 9 final.
[5] https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-building-european-data-economy
[6] COM(2018) 232.

2. Principles for business-to-business (B2B) and business-to-government (B2G) data sharing

The Communication which this Staff Working Document accompanies[7] defines the following principles in order to ensure **fair markets for IoT objects and for products and services relying on data created by such objects**:

a) **Transparency**: The relevant contractual agreements should identify in a transparent and understandable manner (i) the persons or entities that will have access to the data that the product or service generates, the type of such data, and at which level of detail; and (ii) the purposes for using such data.

b) **Shared value creation**: The relevant contractual agreements should recognise that, where data is generated as a by-product of using a product or service, several parties have contributed to creating the data.

c) **Respect for each other's commercial interests**: The relevant contractual agreements should address the need to protect both the commercial interests and secrets of data holders and data users.

d) **Ensure undistorted competition**: The relevant contractual agreements should address the need to ensure undistorted competition when exchanging commercially sensitive data.

e) **Minimised data lock-in**: Companies offering a product or service that generates data as a by-product should allow and enable data portability as much as possible[8]. They should also consider, where possible and in line with the characteristics of the market they operate on, offering the same product or service without or with only limited data transfers alongside products or services that include such data transfers.

The Communication also states that the respect of the following principles could support the **supply of private sector data to public sector bodies** under preferential conditions for re-use:

a) **Proportionality in the use of private sector data:** Requests for supply of private sector data under preferential conditions for re-use should be justified by clear and demonstrable public interest. The request for private sector data should be adequate and relevant to the intended public interest purpose and be proportionate in terms of details, relevance and data protection. The cost and effort required for the supply and re-use of private sector data should be reasonable compared with the expected public benefits.

b) **Purpose limitation**: The use of private sector data should be clearly limited for one or several purposes to be specified as clearly as possible in the contractual provisions that establish the business-to-government collaboration. These may include a limitation of

[7] COM(2018) 232.

[8] E.g. data produced by robots in the context of industrial processes, relevant for provision of after-sales services (e.g. repair and maintenance), or data on the rating of service providers.

duration for the use of these data. The private sector company should receive specific assurances that the data obtained will not be used for unrelated administrative or judicial procedures; the strict legal and ethical provisions governing statistical confidentiality in the European Statistical System could serve as a model in this regard.

c) **'Do no harm':** Business-to-government data collaboration must ensure that legitimate interests, notably the protection of trade secrets and other commercially sensitive information, are respected. Business-to-government data collaboration should allow companies to continue being able to monetise the insights derived from the data in question with respect to other interested parties.

d) **Conditions for data re-use:** business-to-government data collaboration agreements should seek to be mutually beneficial while acknowledging the public interest goal by giving the public sector body preferential treatment over other customers.

This should be reflected in particular in the level of compensation agreed, the level of which could be linked to the public interest purpose pursued.

Business-to-government data collaboration agreements that involve the same public authorities performing the same functions should be treated in a non-discriminatory way.

Business-to-government data collaboration agreements should reduce the need for other types of data collection such as surveys. This should reduce the overall burden on citizens and companies.

e) **Mitigate limitations of private sector data:** To address the potential limitations of private sector data, including potential inherent bias, companies supplying the data should offer reasonable and proportionate support to help assess the quality of the data for the stated purposes, including through the possibility to audit or otherwise verify the data wherever appropriate. Companies should not be required to improve the quality of the data in question. Public bodies, in turn, should ensure that data coming from different sources is processed in such a way to avoid possible 'selection bias'.

f) **Transparency and societal participation:** business-to-government collaboration should be transparent about the parties to the agreement and their objectives. Public bodies' insights and best practices of business-to-government collaboration should be made publicly available as long as they do not compromise the confidentiality of the data.

3. Business-to-business (B2B) data sharing – a 'How to' guide

The supply and the re-use of data in B2B relations can take many forms in terms of technical mechanisms, underlying business models and the legal vehicle that supports the B2B data sharing arrangement. This section describes some of them in more detail.

3.1. Models of B2B data sharing

The underlying business models of data sharing can differ quite substantially and it strongly depends on the type of data in question and the strategic business interest. They can range from an Open Data approach to exclusive data partnerships with only one party:

a) **An Open Data approach**: An Open Data approach, whereby the data in question are made available by the data supplier to an in principle open range of (re-)users with as few restrictions as possible and against either no or very limited remuneration, can be chosen when the data supplier has a strong interest in the data re-use. Examples are providers of services that would like to make use of an ecosystem of third party application developers in order to reach the final customers.

b) **Data monetisation on a data marketplace**: Data monetisation or trading can take place through a data marketplace as an intermediary on the basis of bilateral contracts against remuneration. This can be interesting for companies that do not know potential re-users for their data and aim at engaging in one-off data monetisation efforts. This mechanism appears suitable when either (1) there are limited risks of illicit use of the data in question, (2) the data supplier has grounds to trusts the (re-)user, or (3) the data supplier has technical mechanisms to prevent or identify illicit use. Model contract terms can lower the costs of drawing up data usage agreements.

c) **Data exchange in a closed platform**: Data exchange may take place in a closed platform, either set up by one core player in a data sharing environment or by an independent intermediary. The data in this case may be supplied against monetary remuneration or against added-value services, provided e.g. inside the platform. This solution allows offering added-value services and thus provides for a more comprehensive solution for more stable data partnerships and allows for more mechanisms of control on the usage made of the data; model contract terms can lower the costs of drawing up data usage agreements. Where the data sharing is exclusive, it would need to comply with the competition rules.[9]

Variations and combinations of these models are possible and need to be adapted to each concrete business need. The term ‘data sharing’ is used in order to describe all possible forms and models underpinning B2B data access or transfer.

[9] See e.g. the Commission's Guidelines on Vertical Restraints, OJ C 130, 19.05.2010, p. 1, and Guidance on the Commission's enforcement priorities in applying Article 82 of the EC Treaty [now Article 102 TFEU] to abusive exclusionary conduct by dominant undertakings, OJ C 45, 24.2.2009, p. 7.

3.2. The legal aspects of data sharing: data usage or licensing agreements

B2B data sharing is typically implemented on the basis of contracts. In data usage or licensing agreements parties agree on the subject and value of the contract as well as on all other modalities put down in contract terms. Data monetisation agreements may not only be bilateral in nature but may also be concluded between multiple parties.

The design of the relevant contract terms for data usage or licensing agreements requires special attention so as to both comply with existing legislation, in particular legislation that would prevent data sharing or make it subject to specific conditions, and to ensure that the strategic interests of each party and competition are being preserved.

Model contract terms for different types of data sharing agreements and for some sectors or types of data sharing are already being developed. The Commission, through a Support Centre for data sharing which will become operational in early 2019, is planning to collect best practices, existing model contract terms and checklists.[10]

The following considerations may help companies in the preparation and/or negotiation of data usage agreements:

a) What data shall be made available?

   - Describe data which you wish to share as concretely and precisely as possible (e.g. R&D data, customer data, diagnostic data), including the levels of updates to be expected in the future. When interpretative resources that make analytics possible (e.g. methods, models) are shared together with datasets, they should be described.
   - What quality levels can be assured for the data, also over time? Shared data needs to be of good quality, i.e. accurate, reliable and when necessary up-to-date. Ensure that data are not missing, duplicate, unstructured. Specify the source/origin of data and how it was collected/constructed. A mechanism for reporting error in the data may be set up.
   - Is the data sharing about a data set or a data stream?
   - Ensure compliance with legal obligations that may prohibit the access or transfer of the data in question to others. Ensure respect of rights that others may have on the data. Verify rights on content represented by the data (intellectual and industrial property rights).
   - Ensure respect of data protection legislation. Among others verify that there is a legal basis for the processing of personal data in line with the General Data Protection Regulation.

b) Who can access and (re-)use the data in question?

[10] Cf. Annex to the Commission implementing decision on the adoption of the work programme for 2018 and on the financing of Connecting Europe Facility (CEF) - Telecommunications Sector, p. 42.

- Ensure that the contract defines in a transparent, clear and understandable way who has a right to access, right to (re-)use, and right to distribute data and under which conditions. Specify if and how data may be licensed for re-use. Be clear when explaining the conditions of licences for data re-use and distribution. Sub-licensing needs also to be considered: either it should be specifically excluded or the conditions under which it is allowed and for what types of data should be specified
- The right to access and (re-)use of data does not need to be unlimited. The agreement may for instance limit the right to access: for example, only to members of specific professional groups (e.g. farmers) or link it to certain purposes of use of data (e.g. for a limited commercial use).

c) What can the (re-)user do with the data?

- In the contract negotiations, the (re-)user should be as open and as clear as possible about how the data is going to be used, including by parties downstream. This will ensure transparency and increase the trust of the supplier of the data.
- Specify the exact usage that can be made of the data, including rights on derivatives of the data (analytics).
- Define non-disclosure rules regarding downstream parties.

d) Define the technical means for the data access and/or exchange, including

- Frequency of data access and maximum loads;
- IT security requirements;
- Service levels for support.

e) What data do I need to protect and how do I protect it?

- Ensure that proper measures to protect your data are in place. These measures should apply to data sharing transactions and to data storage as data can be subject to theft or misuse by organised crime groups and individual hackers. Data can also be released accidentally, for example through human error or because of a technical problem. Data can also be subject to unauthorised access or disclosure or can be lost.
- Ensure the protection of trade secrets, sensitive commercial information, licenses, patents, intellectual property rights. Neither party shall aim at retrieving sensitive information from the other side as a result of the data exchanges.

f) Include rules on liability provisions for supply of erroneous data, disruptions in the data transmission, low quality interpretative work, if shared with datasets, or for destruction/loss or alteration of data (if it is unlawful or accidental) that may potentially cause damages.

g) Define rights of both parties to perform audits on the respect of the mutual obligations.

h) What is the intended duration of the contract? What rights to terminate the contract? What notice to be given to your partners?

i) Agree on applicable law and dispute settlement mechanisms.

3.3. The technical aspects of data sharing

There are a number of technical mechanisms for sharing data in the B2B context. Some of the technical mechanisms can provide data usage rules while offering a trusted and secure environment for the exchange of datasets.[11]

Three types of mechanisms can be distinguished: (a) the data holder makes available selected data directly to a larger number of re-users e.g. via an Application Programming Interface; (b) the data holder makes available selected data via an intermediary (a data marketplace) to one or several re-users with limited control over the subsequent use; (c) the data holder makes available selected data via an intermediary (a data space or platform) to one or several re-users in an environment that allows stronger control and traceability of the subsequent use.

a) **One-to-many data sharing via an Application Programming Interface (API) or an Industrial Data Platform**: Some companies that are engaging in data interactions with others are using unilateral mechanisms for technically enabling data access such as APIs or specific platforms they have set up for data storage, processing and exchange.

   Opening access for third parties to data through public APIs, i.e. an API accessible to a wider public and not only to parties inside the same organisation, is becoming more and more common. The number of APIs has increased dramatically since 2010 and continues to do so.[12]

   APIs can allow smaller firms in particular to easily use or re-use business data. User-friendly and well-designed APIs help to create and expand ecosystems with new and innovative products using data that have already been gathered.

   APIs have the potential to facilitate interoperability, allowing software applications to exchange datasets and data streams.[13] Inherently, APIs can include specifications of the datasets themselves and can offer at a technical level the management of access rights.

   On this basis, the Commission encourages[14] companies all over Europe to consider using open, standardised and well-documented APIs more broadly. This could include

[11] The mechanisms described and the examples are taken from a Study on data-sharing between companies in Europe, performed by Everis on behalf of the Commission (study report forthcoming).
[12] http://www.programmableweb.com/api-research
[13] See the details of the guidance document on APIs developed by the Share PSI network co-funded by the European Commission under the Competitiveness and Innovation Framework Programme: http://www.w3.org/TR/dwbp/#useanAPI.
[14] COM(2017) 9 final.

making data available in machine-readable formats and the provision of associated metadata.

TomTom is a Dutch company that produces traffic, navigation and mapping products. According to findings recorded by a Commission-funded study[15], most revenue from the company's activities comes from the data (maps and online services) licensed to other companies.

TomTom offers Application Programming Interfaces[16] for developers as a means of data access.

According to TomTom this has the following advantages compared to other technical means to share data:

- Easy and swift access to data
- Monitoring the use of data
- Verification of breaches of contract
- Rapid action on cases of data misuse (i.e. terminate or suspend access to data)

Companies, in particular larger companies, also develop dedicated **data platforms** in order to manage regular data interactions with third parties. They offer additional functionalities when it comes to data exchange, in particular for two-way data exchange, for storage inside the platform and for additional services to be provided on top of the data (based on data analytics).

Airbus is a European multinational corporation that designs, manufactures, and sells civil and military aeronautical products.

Having used various ways of making available data to authorities and business partners, in June 2017 Airbus launched Skywise[17] – an "open digital platform for aviation".

Client companies make data available in return for services based on data analytics.

Based on Hadoop software, the main advantage of this technical approach is the seamless integration with airlines' existing IT infrastructures, thereby making it easy for participants to make their data available on the platform. Airbus can work on the basis of the original file format and return insights through the platform using common tables and visualisation tools.

[15] Everis, Study on data-sharing between companies in Europe (forthcoming)
[16] https://developer.tomtom.com/tomtom-maps-apis-developers
[17] https://services.airbus.com/maintenance/expertise-and-other-services/skywise/skywise

b) **Data monetisation via a many-to-many data marketplace**: The term "data marketplace" is employed here in order to designate a specific type of intermediary that may have three essential functions: (1) match-making between potential data supplier and data buyer; this can include specific settings in which potential supplier and potential buyer can remain anonymous in a first part of the establishment of the data transfer preparation as the intent to supply or to buy can reveal already secret business information (future business strategies); (2) the actual transfer of the data (and the agreed compensation), notably the creation of trust that the object of the negotiation will not be altered during the course of the negotiations; (3) a certification function that the transaction has actually happened, interesting potentially for reporting in the corporate balance sheet. Furthermore, such intermediaries can provide additional services such as model contract clauses or anonymization services (if personal or confidential data are exchanged). The role of this type of intermediary ends once the data have been transferred.

DAWEX[18]is a French company describing itself as a "global data marketplace" that was founded in 2015.

Dawex does not purchase or sell data. Instead, Dawex brings together companies interested in monetising and re-using data, and fosters transparency between data suppliers and users by ensuring that they communicate and conduct the transaction directly on the platform.

Dawex developed a series of tools to help both data suppliers and users to understand, assess and communicate about the data. Visualisation tools (e.g. heat maps, tree maps) provide data users with different information about a complete dataset that can be securely shared before a transaction is completed. Sampling tools automatically generate representative data samples based on algorithms to avoid any bias. Data users and data suppliers communicate using a messaging tool embedded in the platform. Additionally, DAWEX supports the negotiation of the contractual agreement by model terms that can be automatically generated.

[18] https://www.dawex.com/en/

c) **Data sharing via a technical enabler**: Different from the type of intermediary discussed above, such technical enablers have a strong focus on providing services in addition to a data exchange such as processing the relevant data in view of responding to certain business needs or questions. Most importantly, such type of intermediary would offer additional features allowing the data supplier to control the use made of the data, in particular the respect of the provisions of the data transfer agreement. This can include forms of track-and-trace of data usage made, e.g. the logging of all data access and processing actions – potentially using distributed ledger technology (blockchain), or developing forms of digital watermarking. The intermediary may also develop instances of self-regulation within the community of users of the data space or platform, possibly including a set of sanctions for data users in violation of individual data transfer agreements.

Nallian[19] has developed a cloud-based platform that enables real-time data sharing and supports process synchronisation. The company works with a basic data sharing technology layer that can be customised to meet the needs of data users in a particular community or domain. The platform is based on cloud technology combined with a community management tool.

The current users of Nallian's technical solution are companies working in logistics, in vertical supply chains and in multimodal transport networks. For these companies, the ability to overcome fragmentation issues and share data in a seamless way appears to be key.

The platform accepts a wide range of options to inject data to the cloud: from simple file uploads to API-based integrations. The platform is enriched with value-added APIs and apps that use a common data model in order to take advantage of all data stored on the platform and provide valuable insights to users. Finally, the platform also accepts data pushed through connected devices or B2B messages exchanged through electronic data interchange (EDI).

The platform allows data suppliers to maintain a granular control over who has access to which data and for what purpose. This control is enabled by a rights-granting engine embedded in the platform that allows data suppliers to define roles and sharing rules for the different community members down to field-level, including for app providers. In addition, the platform facilitates data anonymization and aggregation to meet the necessary privacy requirements.

[19] https://www.nallian.com/

4. Making B2G data collaboration a success – a 'How to' check-list

The supply and the re-use of data in B2G relations can take many forms both in terms of the underlying mechanisms and the legal vehicle that supports them. This section describes some of them in more detail.

4.1. Models of B2G data sharing

a) **Data donorship**: B2G data supply could take the form of data donorship. It can be considered as a form of corporate social responsibility. One of the potential effects would be that such a data donorship programme is backed up by a dedicated team supporting any potential party interested in using the data.

Mastercard's Data Philanthropy[20]

Mastercard considers that organisations on a mission to alleviate human suffering — regardless of size and influence — should have the necessary tools and resources to access and use data to solve problems. The Mastercard Center for Inclusive Growth is committed to closing the gap through data philanthropy by:

- sharing data, e.g. granting access to their proprietary data — in a way that fully protects consumer privacy — to assist research;
- sharing data knowledge, e.g. leveraging in-house expertise to conduct an analysis and release the findings for broader use;
- leveraging expertise, e.g. working with its partners to provide additional expertise and capacity.

b) **Prizes**: B2G collaboration can also set up prizes that would encourage individuals and companies specialised in data analytics to find solutions to a particular public interest challenge. For example, a public organisation could set up a challenge in collaboration with a company that would provide the private sector data needed to solve such a challenge.

Horizon Prize Big Data Technologies[21]

Under the EU's Horizon 2020 funding programme, a prize in the area of Big Data technologies has been announced to find ways to optimise the use of energy grids through a more precise forecasting system. The winning solutions will need to demonstrate the ability to analyse extremely large collections of structured geospatial temporal datasets, time recordings of weather conditions and other data with different parameters used in the operation of energy grid management.

[20] https://mastercardcenter.org/action/call-action-data-philanthropy/
[21] http://ec.europa.eu/research/horizonprize/index.cfm?prize=bigdata

c) **B2G data partnerships**: B2G collaboration can take the form of data partnerships. Public sector bodies can enter into arrangements with private companies, which include the mutual sharing of data, in compliance with the PSI Directive[22] as regards public sector information shared with the private sector. This can bring benefits also for the private company as it will be able to draw insights from the correlation of the private and public sector data.

'Assessing the Quality of Mobile Phone Data as a Source of Statistics' - study by the Statistical Agency of Belgium and Eurostat[23]

A study jointly conducted by Statistics Belgium and Eurostat demonstrated the potential of mobile network data for estimating population density. It aimed at assessing the quality of Belgian mobile phone data (from the major network operator, Proximus) focusing on actual present population. The mobile network data were tested for internal consistency and contrasted with the results of the Belgian population census of 2011, which is constantly updated as part of the population register. Both data sets were aggregated for privacy reasons.[24]
The results of the study were beneficial to both parties. On the one hand, it could be shown that mobile network data provide valid and accurate information which may serve as a complement to traditional statistics. On the other, mobile network operators could for example benefit from resident population data to improve estimations of persons' mobility for new applications offered by the mobile network operator.

d) **Intermediaries**: In cases when there is no previous relationship between a company and a public sector body and trust between the two is absent, an intermediary can be tasked to obtain insights necessary for public interest purposes.

Consumer Data Research Centre (CDCR UK)[25]

Vast amounts of UK consumer data are generated each day, providing valuable insight to help organisations operate more efficiently. The aim of CDRC is to work with organisations to open up their data to trusted researchers so they can provide solutions that drive economic growth and improve society.

[22] Directive 2003/98/EC of the European Parliament and the Council on the re-use of public sector information (OJ L 345, 31.12.2003, p. 90).
[23] De Meersman et al (2016): Assessing the Quality of Mobile Phone Data as a Source of Statistics, https://ec.europa.eu/eurostat/cros/system/files/assessing_the_quality_of_mobile_phone_data_as_a_source_of_statistics_q2016.pdf
[24] Statistical offices maintain registers containing personal and companies' data but these registers cannot be shared with other parties due to protection of personal data and statistical confidentiality restrictions. However, data from private sector can be linked with register data while ensuring the security of the data. Aggregated statistical results that cannot be traced back to the data subject may be published as a result of this analysis.
[25] https://www.cdrc.ac.uk/

e) **"Civic data sharing":** Individuals may be encouraged to authorise public sector bodies to process their personal data which were previously processed by a private company. It should be highlighted that in this case public authorities would also need to comply with data protection legislation. The processing must be in accordance with an appropriate legal basis (e.g. consent under Article 6(1)(a) or the performance of a task carried out in the public interest under Article 6(1)(e)[26]). Such "civic data sharing" is most likely to work in situations where there is either a sufficiently strong link between the citizen and the public sector body in question (e.g. the municipality where they live) or when the public interest purpose is particularly convincing from a citizen's perspective (fighting certain diseases, make travel flows around popular events work etc.).

4.2. Legal and practical considerations in B2G data sharing collaboration

The following considerations may help public bodies and companies in the preparation and/or negotiation of data usage agreements:

a) Public bodies should identify a public interest purpose, the private sector data and the level of granularity needed. Some examples of private sector data serving public interest purposes can be social media data, transaction data or retailer data. Companies may also reflect on their data can contribute to a public interest purpose and start the negotiation process.

b) The parties should identify the internal challenges and constraints related to data sharing.

- Public bodies and companies might need to invest in knowledge management and data governance.
- Companies that establish corporate departments responsible for data sharing including data monetisation in B2B contexts will find B2G data sharing to be less costly and challenging in terms of data governance, infrastructure, and legal drafting. As data sharing gains importance for more companies, the cost and burden per individual collaboration are likely to go down.
- Companies and public sector bodies need to ensure compliance with the provisions of the GDPR and ePrivacy legislation (ensure lawfulness of processing, including reliance on a legal basis, such as consent, proper use of anonymization techniques, confidentiality respect of the principle of data protection by design and by default, use of privacy-preserving analytics methods, data protection impact assessments where required).

[26] In case public authorities rely on Article 6(1)(e) of the GDPR ("processing is necessary for the performance of a task carried out in the public interest") such legal basis must be laid down by the Union or the Member State law. Moreover, in case of such "civic data sharing", data subjects would have to be clearly informed, including on the right to withdraw consent and about any possible further processing of their personal data by public authorities.

- In order to ensure representativeness of the insights, avoiding selection bias, public sector bodies need to make a careful analysis of potential data sources and ascertain limitations of one specific data provider. They should carefully consider data triangulation, constant observation and recalibration of models, and a combination with e.g. public consultation and tools to gather evidence and stakeholders' views to mitigate risks and possible methodological limitations of the private sector data sources.

c) Parties need to choose the technical or practical modalities for data sharing that are best suited to their internal challenges and data governance.
   - Public bodies need to safeguard the protection of legitimate commercial interests (e.g. business confidential information, trade secrets) and ensure the security of the technical modality to access the private sector data. Private sector data that are transferred to a public sector body should be treated as confidential data. It needs to be made explicit in the relevant data processing infrastructures through annotation and access restrictions that they are covered by established exemptions if the public sector body is subject to access to documents legislation. Adequate measures to ensure network and information systems security need to be in place.
   - Public bodies might need to expand their technical and staff capabilities to exploit the possibilities to use private sector data.

d) The contract should include the conditions for implementation, the time limitations and the specific data sets that would be used.
   - Public bodies should ensure that their request for specific private data complies with the principle of proportionality and is necessary to achieve the public interest purpose defined. The agreement should specify that after the purpose has been achieved or the limit of the duration has been reached, the data transmitted are to be erased. The usage of the same data for a different purpose should be subject to a new or amended collaboration agreement.
   - The parties should define the conditions at operational level for the transfer of data: format of the data and metadata, quality, granularity and duration of access and mode of access.
   - The parties should determine the compensation. In this regard different options exist, namely limiting the remuneration to a pro rata recovery of the costs incurred in the production, preservation and dissemination of the data – only exceptionally combined with allowing a fair return on investment – and limiting the remuneration to, at maximum, the costs related to the dissemination of the data, considering that the costs of production and preservation of the data depending on the instant case may have already been covered by other revenue streams. The choice of the option could be linked to the public interest purpose pursued and the specificities of the social need it aims to fulfil.

- In order to allow public bodies to make the necessary quality assessment in order to ascertain the presence of potential selection biases or other quality limitations that may only become apparent after the conclusion of the agreement, companies supplying the data should offer, to the best of their abilities, reasonable and proportionate support so as to enable the assessment of the quality of the data for the stated purposes, including through the possibility to audit or otherwise verify the data wherever appropriate.

e) The parties should agree on common guiding principles for the monitoring of the implementation of the contract:
   - They may agree on a code of conduct or use existing ethical rules such as the European Statistics Code of Practice[27], install a coordination committee or appoint an independent auditor to oversee the data use.
   - Public bodies put in place the necessary safeguards preventing the misuse of the accessed data for other purposes than the ones defined in the contract.

f) The contract should include rules on liability for supply of erroneous data, disruptions in the data transmission, low quality interpretative work, if shared with datasets, or for destruction/loss or alteration of data (if it is unlawful or accidental) that may cause damages.

g) The contract should establish the applicable law and the dispute settlement mechanisms.
   Any party should have the freedom to terminate the contract when there is a legal or technical risk as regards the treatment or the use of the data shared.

h) Public bodies should disseminate the results/insights of the B2G collaboration and ensure mechanisms for public feedback, whenever necessary or relevant, without compromising the confidentiality of the private sector data.

[27] In the case of agreements with statistical offices, this could be the European Statistics Code of Practice, http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955

4.3. Technical means to establish B2G collaboration

In any B2G collaboration a decision has to be made about how the insights from the private sector data are to be derived for public interest purposes. This may mean an actual transfer of private sector data to the IT environment of the public body in question. However, this is not the only possibility and other mechanisms can be considered. This section offers an overview of technical means that are alternative to the transfer of private sector data to the IT environment of the public body. These technical mechanisms can provide access and data usage rules while offering a trusted and secure environment for the exchange of datasets.

a) **Data platforms**: The creation of data platforms can offer a secure environment to store and exchange data between companies and public bodies. Such platforms can provide public bodies with standardised data to create shared data resources or insights, in collaboration with companies.

Centre for Big Data Statistics, the Netherlands[28]

The Centre for Big Data Statistics (CBS) partners with a variety of organizations from the private sector to gather the necessary private sector data to create high quality data visualisations. Since CBS is a public sector organization, it also has access to the Netherland's large repository of government and sensor data, which they are able to pool with these new data sources to provide new insights.

b) **Algorithm-to-the-data**: Bringing the algorithm to the data can be a solution to the security, data protection and privacy challenges of data. It would respect one of the main considerations for ensuring protection of personal data and privacy, which is to move data as little as possible. Using this solution means that the algorithm is installed within the IT environment of the private company and the analysis takes place there. Only the anonymous insights derived by the algorithm are transferred back to the public sector body. The data query interface and analytics possibilities could be co-designed by the company and/or the public organisation in question (or by a trusted intermediary).

The Open Algorithms (OPAL)[29]

The project is a socio-technological innovation developed by Data-Pop Alliance, Imperial College London, MIT Media Lab, Orange and the World Economic Forum to leverage private sector data for public good purposes by "sending the code to the data" in a privacy preserving, predictable, participatory, scalable and sustainable manner. The design of the algorithm has the input from the local advisory Committees for the Orientation of Development and Ethics (CODE) so that these algorithms serve local needs and respect local standards, instead of imposing external perspectives and expertise.

[28] https://www.cbs.nl/en-gb/our-services/innovation/big-data
[29] http://www.opalproject.org/about-us/

c) **Privacy-preserving computation:** In the last years several computation models were developed that allow performing operations on data that need to remain confidential. Such models allow extracting the desired output information without disclosing the input data. Therefore, data computation can take place collaboratively across different administrative domains (public or private) with no need to move data outside the company. Such models imply a fundamental paradigm shift from "sharing data" to "sharing computation". Among the existing privacy-preserving computation methods, the class of secure multi-party computation seems to be particularly suited in the context of B2G data collaboration. Some simple secure multi-party computation techniques are very scalable and powerful. A few companies already provide the technology and relevant platforms. Studies have been conducted that made use of this technique in B2G collaboration.

Secure multi-party computation[30]

Secure multi-party computation is a practical cryptographic method for processing confidential data. Research progress has led to its use in privacy-preserving statistical analysis. In 2015, statisticians from the Estonian Centre of Applied Research (CentAR) conducted a big data study to look for correlations between working during university studies and failing to graduate in time. The study was conducted by linking the database of individual tax payments from the Estonian Tax and Customs Board and the database of higher education events from the Ministry of Education and Research. Data collection, preparation and analysis were conducted using the Share-mind secure multi-party computation system that provided end-to-end cryptographic protection to the analysis. Using ten million tax records and half a million education records in the analysis, this is the largest cryptographically private statistical study ever conducted on real data.

[30] Bogdanov (et al.), Students and Taxes: a Privacy-Preserving Social Study Using Secure Computation. In Proceedings on Privacy Enhancing Technologies, PoPETs, 2016 (3), pp 117–135, 2016. (Extended version, PDF).