BRIEFING

Requested by the AIDA committee

European Union data challenge

KEY FINDINGS

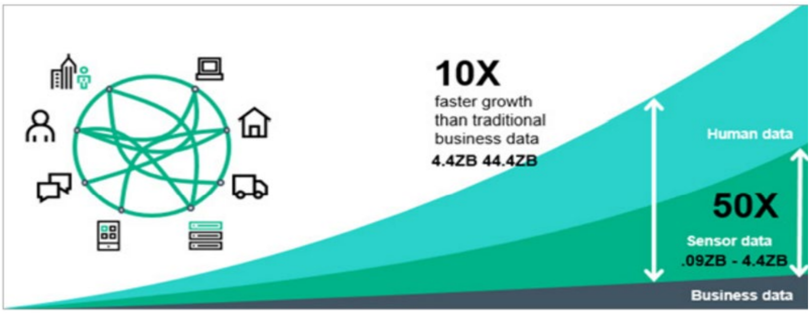
The exponential growth and importance of data generated in industrial settings have attracted the attention of policymakers aiming to create a suitable legal framework for its use. While the term 'industrial data' has no clear definition, such data possess certain distinctive characteristics: they are a subset of big data collected in a structured manner and within industrial settings; they are frequently proprietary and contain various types of sensitive data.

The General Data Protection Regulation (GDPR) rules remain of great relevance for such data, as personal data is difficult to be filtered out from mixed datasets and anonymisation techniques are not always effective. The current and planned rules relevant for business to business (B2B) sharing of industrial data exhibit many shortcomings. They lack clarity on key issues (e.g. mixed datasets), increase the administrative burden for companies, yet not always provide the data protection that businesses need. They do not provide an additional value proposition for B2B data sharing and hinder it in some cases.

While this situation warrants policy intervention, both the instrument and its content should be carefully considered. Instead of a legal instrument, soft law could clarify the existing rules; model terms and conditions could be developed and promoted and data standardisation and interoperability efforts supported.

The role of data for the economy and society at large is expected to increase in the coming years, especially as massive amounts of data are necessary to develop and deploy Artificial Intelligence (AI). The EU aims to ensure its technological and economic leadership in the data-based economy by developing a single market, where personal and non-personal data are freely shared in a secure manner that protects the rights of individuals and guarantees legitimate interests of businesses.

Exponential growth of data (incl. 50-fold growth between 2010-2020)



Source: insideBIGDATA, The intelligent use of big data on an industrial scale, 2017, p. 2.

However, benefits of the sharing and reuse of data between businesses (B2B) have not yet materialised on a sufficient scale, even though data are a non-rival

Policy Department for Economic, Scientific and Quality of Life Policies
Directorate-General for Internal Policies
Authors: Olga BATURA and Roel PEETERS
PE 662.939 - July 2021

EN

good that can be used by competitors for efficiency and innovation purposes. In particular, the potential of the so-called 'industrial data' seems to remain underused.

To gain a better understanding of industrial data and to enhance their sharing in the European Union (EU), this briefing will clarify the concept of industrial data, including its relation to personal data, the role of privacy preservation technologies, the main challenges and limitations of sharing of industrial data, and the existing and future policy interventions that impact the sharing of industrial data.

Concept of industrial data

What is industrial data

As the discussion on governance of industrial data intensifies, especially after the adoption of the proposal on the Data Governance Act and in the wake of the European Data Act, the question of what is exactly industrial data remains unanswered. The notion of industrial data is not defined in any of the legal documents or legislative proposals, and the reference to it is a fairly recent development. In the past, a more clear-cut term '**machine-generated data**' was used that potentially allowed for an easier definition and delimitation of this type of data from other data. The 2017 Communication 'Building a European data economy' defines with a high degree of precision that machine-generated data are those '**created without the direct intervention of a human by computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real**'. Therefore, machine-generated data may be created across all industrial sectors, including transport, energy, healthcare, manufacturing, ICT and others, but they go beyond data created in relation to narrowly understood industrial processes.

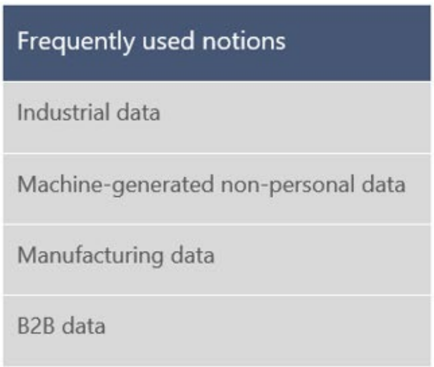
Such precise definition of machine-generated data indicates **special qualities**, which emphasise the value of such data and may justify separate rules or arrangements for its exchange. Machine-generated data are produced in **massive quantities and at high speed**, and they are **not bespoke** or one-of-a-kind datasets. Yet, by contrast to big data, these datasets are often semistructured, **structured** or even standardised, at least within one domain of production or sector.

The analysis of more recent policy publications suggests that now the expression "industrial data" is used to mean 'non-personal machine-generated data', i.e. refers to how data are generated. Additionally, the European Data Strategy refers to the context in which data are generated. It states that 'IoT data created in industrial settings' is 'industrial (manufacturing) data'. This may, however, be a restrictive understanding of what constitutes 'industry'. The official statistics and documents of the EU refer to a plethora of industry sectors beyond manufacturing (e.g. cosmetics, chemicals, gambling, tourism, food production), all of which rely on data generated without direct human involvement.

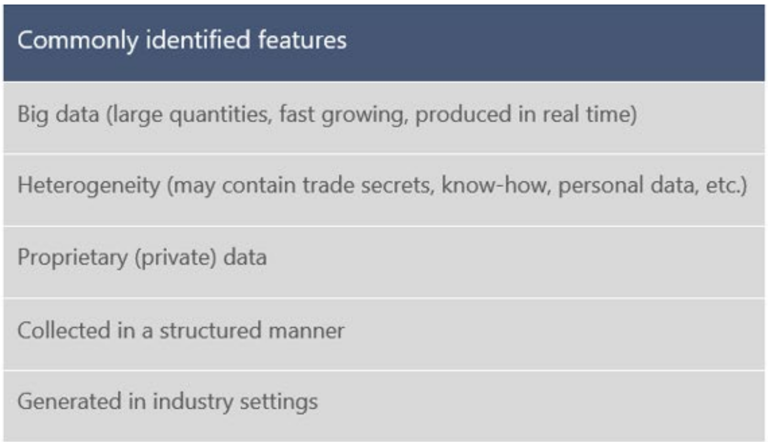
The use of the term 'industrial data' raises a question about its distinction from 'business data', 'operational data' and other types of data. The European Data Strategy indicates that 'industrial data' may be different from 'business data'. While this rings true and 'business data' seems to be a more comprehensive term, a precise distinction might be difficult to draw in practice. Especially in network industries (e.g. telecommunications) or service industries (e.g. tourism and hospitality, online retail, logistics), machine-generated 'industrial data' could simultaneously be operational and business data.

Looking closer at the contents of industrial data, the **heterogeneity** of this type of data stands out. Industrial data may contain trade secrets, reflect competition-relevant know-how or other sensitive and requiring protection information. It may contain data on products, services, networks or processes – as well as on people who are using them or otherwise involved in them. In most cases, industrial data would be proprietary data (i.e. **private**). However, at least some categories of industrial data – for instance, held by public undertakings within the meaning of the PSI Directive – would be subject to the open data regime (i.e. public).

What is industrial data



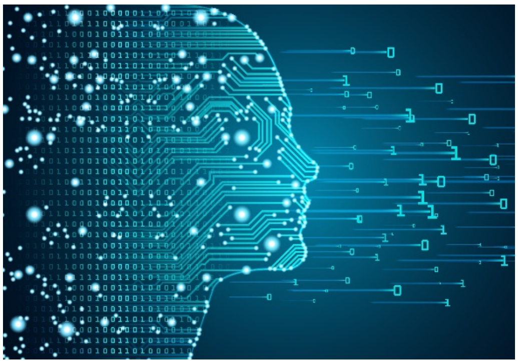
Frequently used notions
Industrial data
Machine-generated non-personal data
Manufacturing data
B2B data



Commonly identified features
Big data (large quantities, fast growing, produced in real time)
Heterogeneity (may contain trade secrets, know-how, personal data, etc.)
Proprietary (private) data
Collected in a structured manner
Generated in industry settings

Source: authors' own elaboration.

Relation to personal data



Although the current understanding of industrial data' aims to be limited to non-personal machine-generated data, in practice, **it is unlikely that no personal data is collected even in closed industrial settings**. The definition of personal data does not depend on the circumstances and manner in which such data have been created, processed, become available or for what use it is intended. The General Data Protection Regulation (GDPR) defines **personal data** on the basis of the objective criterion of identifiability (i.e. **whether information refers to identified or identifiable person**). Article 4(1) GDPR defines personal data broadly as "any information relating to an identified or identifiable natural person", while "an identifiable natural person is one who can be identified, directly or indirectly, by reference to an identifier such as a name, identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Importantly, personal data must refer to a natural person, not a legal entity. Personal data – and especially its specific subset of sensitive data (Article 9 in conjunction with Article 4 (13), (14) and (15) GDPR) – enjoy a high level of protection and can be processed only on lawful grounds with the prior consent of the data subject.

The term 'non-personal data' is defined residually: **non-personal data are electronic 'data other than personal data'** (Article 1 Regulation on the free flow of data). Examples in Recital 9 Regulation on free flow of data – currently the main legal instrument governing the circulation of non-personal data – clearly refer to the wide industrial context of data generation: data from 'automated industrial production processes', 'data on precision farming that can help to monitor and optimise the use of pesticides and water, or data on maintenance needs for industrial machines'. Such data can be exchanged and ported without the limitations imposed on personal data.

Although not all AI systems run on personal data, **the line between personal and non-personal data is becoming increasingly blurred**. Crucial causes are the lack of robustness in anonymisation techniques (see more in the next section) and the risk of re-identification (de-anonymisation) when applying data analytics even to – at first glance – non-personal data (like weather data). Particularly for AI, the use of 'a system designed to make connections and spot patterns not immediately visible to the human eye increases this risk of re-identification.' The balance between economic value and privacy is difficult to maintain over time as the possibilities for re-identification seem to be increasing. In addition, there is a fundamental trade-off between the usefulness of a dataset and the possibility to anonymise it: the more data are anonymised, the less useful data become, and the less effective data-sharing policies will be. In this context, it can be argued that most people want to unleash the power of big data because of the improvements it can bring to the economy and society. Yet, the researchers conclude '**any data that is even minutely useful can never be perfectly anonymous, and small gains in utility result in greater losses for privacy**'.

Role of privacy preservation techniques

Against this backdrop, the role of privacy preservation techniques needs to be considered as they help to turn personal data into non-personal and thereby allow for sharing of useful data in B2B context. The GDPR distinguishes between **two methods of privacy preservation: anonymisation and pseudonymisation**. Pseudonymisation entails the 'processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information' (Article 4 (5) GDPR). It means that certain identifiers that can link data to a person are removed. It is required that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable person. **Pseudonymised data remain personal data, and the GDPR applies to it in full**.

By contrast, the **GDPR does not apply to anonymous data** (Recital 26 GDPR). Personal data can be rendered anonymous in such a manner that the data subject is not or no longer identifiable (i.e. personal data become non-personal). There are two main types of anonymisation techniques: randomisation and generalisation. **Randomisation** techniques modify the veracity of data. Noise addition renders the attributes less accurate while retaining their overall distribution. Permutation swaps the attributes between different subjects (i.e. breaking the links between the data and data subjects) while exact attributes are retained. **Generalisation** means that attributes of data subjects are generalised by changing the scale or order of magnitude (e.g. from city to country level). Aggregation, K-anonymity, L-diversity, T-closeness, differential privacy are all different techniques of generalisation.

The proposal for an Artificial Intelligence Act foresees an important role for the discussed anonymisation methods. Moreover, the European Parliament has called on the European Data Protection Board to develop guidelines to harmonise the implementation of data protection requirements into practice. This should include a clear classification of legitimate and illegal use cases, as well as a list of unambiguous criteria to achieve anonymisation.

While all privacy preservation techniques enhance data protection and privacy, their **effectiveness has been repeatedly questioned**. They may become even less effective in the future with the development of AI and further data analytics. None of the anonymisation techniques guarantees full and complete protection of anonymity, especially in the big data context, because 'human traces are unique'. Experts recommend using a combination of techniques to render personal data truly anonymous. However, de-anonymisation (re-identification) is possible even then as many experiments have demonstrated. Due to possible de-anonymisation, any data could potentially become personal data, which would render the scope of application of the GDPR too broad and all-encompassing.

At the same time, re-identification is deemed illegal and – at least at this stage of technological development – requires extensive expert knowledge, (potentially) additional information and major efforts. The illegality does not necessarily preclude the applicability of the GDPR, but it is more important to consider if re-identification is probable.

In light of imperfect anonymisation techniques, the GDPR sets a **general prohibition on processing special categories of sensitive data**, such as data relating to health, race or sexual orientation, unless the data subject has provided explicit consent. Any use of AI to process such data needs to rely on one of the **specific exemptions under Article 9 GDPR** from this general prohibition. The processing of personal data of children also warrants particular care under the GDPR.

Challenges to sharing industrial data

While the challenges to sharing personal data largely stem from the nature of such data and their regulation, the challenges to sharing non-personal machine-generated data may be less obvious, yet just as systematic and inhibiting.

From a technical perspective, a key barrier to secure B2B data sharing is the **absence of industry-wide data ecosystems**. This revolves around two issues. First, in many industries, there is **a lack of industry standards and data ontologies** used by industry actors throughout the value chain. Data can only be shared if unambiguously understood and integrated by all ecosystem members. Different metrics and understandings of definitions complicate exchanges. At the same time, **negotiating common standards and ontologies is difficult**, as all parties need to agree on the ontology, and actors will be hesitant to replace their internal standards with those of a competitor. Finally, industrial environments are often a mix of older and newer IT solutions, making it even harder to reach an agreement on common standards.

Second, **sizeable investments in technical solutions and a skilled workforce are required** to develop a data-sharing infrastructure that facilitates interoperability and portability between all industry players. Companies have invested significant time and resources in developing a network of old and new internal software solutions tailored to their internal processes. Considerable commitment and investment would be required to establish a well-functioning system of interoperable company IT infrastructures. Fears of **incomplete data portability or interoperability** between the internal and ecosystem infrastructure, leading to loss of access to data, keep business users locked in with their traditional software.

Companies are not only confronted with a lack of technical capabilities. A key cause of the limited uptake of functional and trustworthy data sharing ecosystems is 'the **lack of robust legal and ethical frameworks**, as well as **governance models and trusted intermediaries** that guarantee data quality, reliability, and its fair use'. **Staff is often not equipped** to develop and implement legitimate data-sharing ecosystems that adopt ethics-by-design.

Perhaps most expressed in the context of B2B data-sharing is the issue of **limited willingness** to grant third parties (let alone competitors) access to their business information. Several reasons can explain this. There is a **perceived lack of control over data** when they leave company's premises, which is particularly pressing if a company has difficulty assessing the potential value and confidentiality of data it releases. Moreover, companies would want the **certainty of reciprocity in data-sharing**, e.g. to balance the perceived loss of value of sharing own data against the potential value of data received. Finally, there is the continuous '**risk of navigating around legal constraint** in view of potential data policies' breaches'.

An important nuance to the limited willingness to B2B data sharing comes from research that shows that the **relationship between actors is crucial for companies' incentives to share data**. In cases of horizontal competition, sharing data with producers of similar products amounts to supporting rivals, and firms are generally not eager to do so. Often, companies in traditional industries tend to observe data as a key input for their own processes rather than as a source for innovation by third parties, from which they can also benefit. However, in vertical relationships between value chain partners ('vertical data sharing'), the complementarity of benefits results in strong incentives to share data because gains are distributed across the whole value chain ecosystem. Due to the clear motives to share data vertically, a refusal to do so can be perceived as an anticompetitive expression of dominance which potentially has a detrimental impact on competition and innovation.

The collection and production of data can be **resource-intensive for companies** in terms of efforts, investments and other expenditures involved. While some existing business processes may actively or passively generate relevant data as a by-product, other data types might require deliberate and costly activities. The former category of data relies on the production of the product or provision of the service itself and can be shared at negligible costs. Nonetheless, in particular B2B contexts, 'where the data collection process can require investments in elaborate systems (e.g. technical, sensors, ICT, or administrative systems)', **sharing could harm companies' incentives to invest in the systems required to collect the data**. In the latter category of data, companies have generated data through advanced algorithms (derived data) or costly data collection activities, e.g. surveys, customer trails or experiments (produced data). Both data categories require dedicated expenditures and efforts, and if these result in substantial costs, sharing obligations may distort competition and hamper innovation efforts. Particularly for derived data, the **risk of reverse engineering further limits the incentives** of all parties to invest and innovate when implementing data sharing. As such, it is important to balance the gains for competition against the costs of reverse engineering.

Barriers to data sharing in B2B context

Technical	Regulatory	Company-level
Lack of industry standards	Unclear legal framework	Lack of skilled staff (e.g. data scientists, data compliance officers)
Lack of common data ontologies	Lack of governance models	Lack of targeted financial investments
Lack of interoperability and portability	Lack of incentives to collect and share data	Lack of efforts

Source: authors' own elaboration.

EU actions in relation to B2B data sharing

The EU has made data sharing – and data governance, more generally, - into one of the focal points of the recent legislative measures as an effective legal framework for data exchange can stimulate competition and innovation. Along with the GDPR and Regulation on free flow of data, several legislative proposals target B2B data sharing directly.

Current legal framework

The **GDPR** provides the main framework for any actions related to personal data. It has a dual aim (Article 1 GDPR) of protecting the fundamental right to data protection and stimulating the free movement of personal data under clearly defined conditions. To achieve these objectives, the GDPR gives the rights of control over personal data to data subjects and imposes requirements and duties on data controllers. Not only must sharing be subject to the prior specific consent of the data subject (Articles 4, 6-8 GDPR), it is further restricted by principles of purpose limitation and data minimisation (Article 5 (1) GDPR). At the same time, sharing and reuse of personal data are enabled by the right to data portability (Article 20 GDPR): data subjects can request and receive their data from a data controller in a structured, commonly used and machine-readable format and transmit those data to another controller without hindrance. Alternatively, data subjects can have their data transmitted directly from one controller to another on their behalf.

In the context of machine-generated data in industrial settings, data sharing under the GDPR regime raises several problems. One of the main issues is that businesses, especially SMEs, see the GDPR rules as complicated and confusing and difficult to comply. Therefore, **complex data requirements raise barriers for companies to enter the data market**, and they might refrain from collecting, sharing or exchanging data. One example of such complex rules is the data portability right that applies to the data that a data subject 'has provided to a controller'. This might indicate that data observed by IoT devices and sensors are not portable under the GDPR conditions (although Article 29 Working Party interprets this right broadly and excludes only inferred or derived data).

The **GDPR compliance might be burdensome for certain SMEs**, such as AI start-ups and scale-ups. To them, the exemptions regarding record-keeping of processing activities (Article 30 GDPR) do not apply because, at the very least, their data processing is not occasional. A 2020 survey of AI start-ups reported that, as a result of the GDPR, a large percentage of them had to delete data (over 75%) or relocate resources (over 60%). Considering the very high value of data for AI activities, such actions could seriously impact AI-based innovation and the AI ecosystem in the EU.

The 2018 **Regulation on the free flow of data** is the main legal instrument governing the exchange of non-personal data between businesses. The Regulation provides for data to move freely within the internal market by prohibiting data localisation requirements by Member States (Article 4), which used to be a serious impediment to data sharing across borders. For data porting (Article 6), the Regulation does not introduce any hard obligations but states that the Commission should encourage the development of voluntary codes of conduct and best practices (as an example, see SWIPO for the cloud services industry).

The scope of application of the Regulation on the free flow of data poses a problem as it is defined as residual, aiming to complement the GDPR regime on personal data. The broad approach of the GDPR to the definition of personal data in combination with the case-law of the Court of Justice of the EU results in situations where **non-personal data may end up being requalified as personal** if additional information can be sought from third parties to identify a data subject. This not only gives rise to the **uncertainty** as to what data may fall exclusively under the Regulation on the free flow of data but also raise legitimate **concerns of the GDPR becoming 'the law of everything', making it effectively unusable in practice**.

A particular challenge in this context is the **treatment of mixed datasets** – datasets consisting of personal and non-personal data (Article 2(2)). The Regulation on the free flow of data shall apply only to the non-personal data in a mixed dataset. In practice, the distinction or separation of datasets' parts is often difficult or impossible. Where both types of data are 'inextricably linked', the 2019 Commission Guidance advises that the GDPR data protection regime should fully apply to the whole mixed dataset, even where personal data represent only a small part of it. Considering that many datasets are likely to be mixed datasets with 'inextricably linked' personal and non-personal data, companies may be 'sitting' on a lot of data, **unable or**

unwilling to share them. Some companies (especially smaller ones) may prefer to **err on the side of caution and not share machine-collected data** where they suspect insufficient anonymisation or the presence of personal data that they could not filter out. Other companies may use the potential presence of personal data in datasets as an **excuse not to share data with competitors**.

Machine-generated data collected in industrial settings may enjoy protection under intellectual property law. Until now, the EU legal framework in this regard was limited, but this might change with the review of several instruments in the context of the development of the EU Data Act. The reach of protection of machine-generated data by intellectual property law is central for data sharing. Any rights and obligations of **data access or exchange (will) collide with and limit** the protection granted by **intellectual property** law.

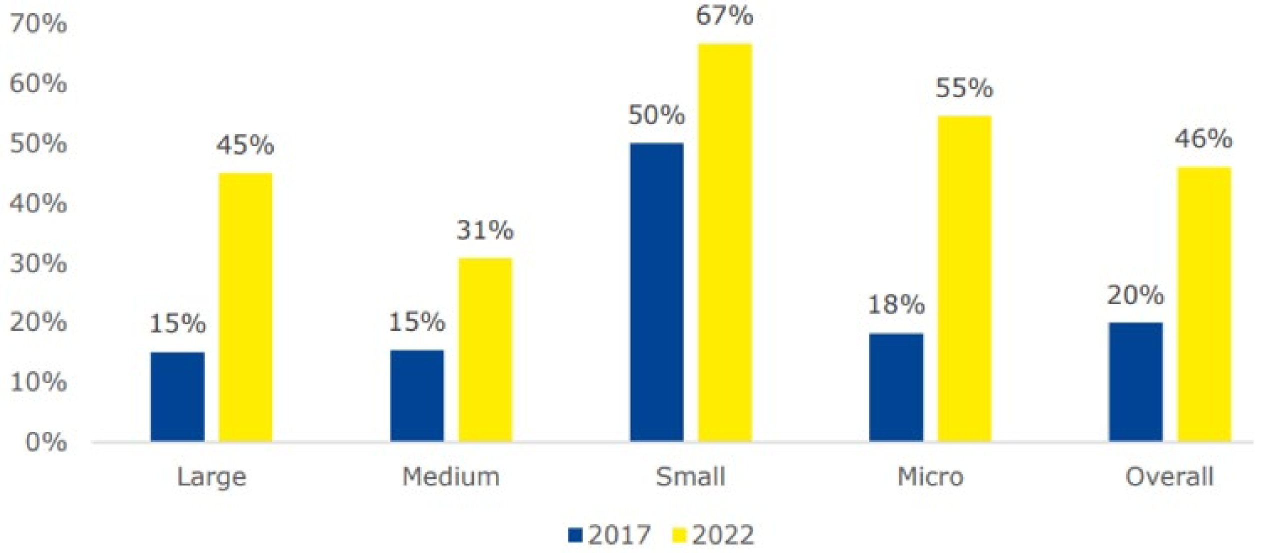
Datasets made of machine-generated data, data from IoT devices and AI, big data etc., have not yet benefitted from the **database protection** under the Database Directive. The said datasets usually do not fulfil the requirement of originality necessary to enjoy copyright protection. Nor do they qualify for the sui generis protection that requires 'investment of considerable human, technical and financial resources' (Recital 7 Database Directive). Under the current interpretation, the sui generis right does not apply to databases that are a by-product of a company's main activities but only to databases that contain data obtained from external sources. At the same time, the 2018 evaluation of the Database Directive noted examples of high investments in the content of databases, specifically in the collection, verification and maintenance of machine-generated and sensor data, and emphasised that it is 'increasingly difficult to distinguish between data creation and obtaining of data' when 'systematic categorisation of data' is performed by 'the data-collecting object (e.g. industrial robots)'.

Machine-generated data may also benefit from the **trade secrets protection**, though the relevant Trade Secrets Directive has been criticised as not accounting for big data and other data economy realities. Know-how and trade secrets enjoy protection if (1) the information is secret in the sense that 'it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question'; (2) the information has commercial value because of its secrecy; and (3) the information has been subject to reasonable steps to keep it secret (Article 2(1) Trade Secrets Directive). It has been pointed out that none of these three requirements can be easily applied to all machine-generated data. For instance, while data secrecy can be confirmed for data generated in a closed (e.g. factory) environment, it would be impossible for automated cars. Where several entities are involved in data collection across a value chain, it will be difficult to establish persons responsible for secrecy. The link between the secrecy of data and its commercial value is difficult to prove as the value usually becomes apparent through the applied data analytics (e.g. inferences, correlations with other data).

Legislative proposals

The 2020 European strategy for data provides a policy framework for future legislative and other measures in the context of sharing machine-generated data. The EU's ambition to regain data sovereignty and become an international data hub should be based on creating a Single Market for data and providing a worldwide standard for horizontal and vertical data sharing (i.e. B2B or between businesses and government (G2B and B2G)) across the EU. A Common European industrial (manufacturing) data space should be created through addressing 'issues related to the usage rights on co-generated industrial data (IoT data created in industrial settings)' as a part of the European Data Act. It is not clear whether the creation of this common space will go beyond a legal act or what other initiatives might be part of it. Next to the industrial data space, an agricultural data space, energy data space and **other sectoral data spaces** are foreseen. It follows that, even though IoT and sensors collect (at least some) data in these industry sectors, such data may not be part of industrial data. It remains **unclear how such data will be defined and distinguished from industrial data**.

Percentage of companies sharing data in 2017 vs expectations for 2022 (per company size)



Source: Everis, Study on data sharing between companies in Europe, 2018, p. 37.

The proposal for the Data Governance Act aims to create an environment of trust for data sharing in the B2B context and lower its cost. To this end, the category of data intermediaries is introduced that provide 'middleman' data sharing services to data holders and data users, e.g. by establishing data exchanges, platforms or databases. Data intermediaries must remain neutral concerning the data exchanged (e.g. not process the entrusted data for their own purposes) and ensure fair, transparent and non-discriminatory procedures for data access, including prices. Data intermediaries must notify their activities with a competent national authority who will also monitor them.

There is a lot of doubt whether the proposition of data intermediaries carries enough value-added to encourage more data sharing between businesses. The proposition is criticised as being too **vague** and creating more **legal uncertainty** for the potential data sharing services while at the same time **introducing new regulatory risks and burdens**. In the context of mixed datasets and personal data regulation, data intermediaries do not enjoy any privilege or exemption, which would put more regulatory risk on data holders sharing data via intermediaries. They cannot offer flexible pricing due to the neutrality requirement, which makes them less attractive than direct B2B data sharing. Concerns are also raised from the competition perspective, both in terms of potential anti-competitive information sharing, risks of private regulation of data exchanges and unlevel playing field between data intermediaries and tech companies that offer data sharing services. **None** of these issues create **incentives for the existing data brokers to register as data intermediaries or for the current direct contract-based data sharing to switch to using a middleman.**

Recommendations for the future

As there is not much evidence that data sharing is a significant, persistent problem across the economy, there is a need for **more empirical research into** the actual situation of **data sharing in different sectors**. Some of the questions to be examined are what data are shared, what data are desirable to be shared, what the current practices of data sharing are and what the specific problems are that an intervention at the EU level might resolve.

If legal action is taken concerning industrial data, it should be **carefully considered whether an introduction** of such a **new data category is strictly necessary** and previous research and definition of machine-generated data taken into account. If a legislative act is passed in reference to industrial data, the **definition** should be **specific** and not be mixed up with other data categories. The definition should be clear and narrow enough to allow for the proper applicability, to increase legal certainty and to build up an incentive for data sharing.

A **sectoral approach** might be warranted as different sectors might have different specifics and practices related to data sharing. In some sectors, data sharing is already taking place, and companies are willing participants. In this case, enabling measures (e.g. support for data exchange platforms, improved standards and interoperability) may suffice. In sectors where the readiness to data sharing is low, in the first step, it should be determined why this is the case, and the choice of the legal instrument and type of intervention to increase data sharing would depend on the reasons. If the sector has a high volume of personal data or data relevant to trade secrets and know-how, it may need additional safeguards for (cyber)security of data transfers (e.g. standards and certification).

To design effective measures that would encourage B2B data sharing across the economy, the value proposition to all of the industry is a decisive factor. Such measures do not necessarily need to be of legal nature but can be different types of **soft law**. For instance, the **complex or confusing elements of the GDPR** (e.g. treatment of mixed datasets) could be **clarified** in a guidance or recommendation by the European Commission. Interoperability should be ensured by **supporting the development and use of ontologies and standard formats** for data collection, exchange and processing – as these are not always coherent even within one industry sector. This would save costs, especially for SMEs, and facilitate data exchange across different industry sectors and countries. To this end, the **development of voluntary standards and sharing of best practices** could be the first step. Where a trusted and safe environment for data sharing is a problem, voluntary cybersecurity standards for data transmission between businesses could be developed. Also, safe and secure data processing European infrastructure could be promoted (e.g. following the example of GAIA-X).

Legal guides, model contracts or model terms and conditions for data sharing contracts could be developed considering the needs of different industry sectors (e.g. non-digital versus AI developers), similar to international sales contracts by the International Chamber of Commerce or model contracts for small firms doing international business by the International Trade Center. Such model contracts or legal guides would be of particular help for European businesses exchanging data with companies from third countries.

Some industry sectors are already putting significant efforts to **enable data sharing**, and the policymakers could **study and support their efforts and replicate them for other sectors**. For example, where model agreements and terms and conditions for data sharing are developed by the industry (like by the manufacturing industry), these could be collected on a specialised platform – to increase their visibility and use, including by SMEs and start-ups, and to highlight best practices in this regard could serve as a recommender system.

Main references

- Drexl, J., 2017, *Designing Competitive Markets for Industrial Data – Between Propertisation and Access*, 8 JIPITEC 257. Available at: https://www.jipitec.eu/issues/jipitec-8-4-2017/4636.
- Everis, 2018, *Study on data sharing between companies in Europe*, Study for the European Commission. Available at: https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en.
- van Gorp, N., de Bijl, P., Graef, I., Molnar, G., Peeters, R. and Regeczi, D., 2020, *Exploring data sharing obligations in the technology sector*, Study for the Dutch Ministry of Economic Affairs and Climate Policy. Available at: https://www.government.nl/ministries/ministry-of-economic-affairs-and-climate-policy/documents/reports/2020/11/30/exploring-data-sharing-obligations-in-the-technology-sector.
- Hennemann, M., Datenlizenzverträge, Recht digital 1:2, 2021, 601-70.
- Martens, B., De Streel, A., Graef, I., Tombal, T. and Duch-Brown, N., 2020, *Business-to-Business data sharing: An economic and legal analysis*, JRC Digital Economy Working Paper 2020-05. Available at: https://ec.europa.eu/jrc/sites/default/files/jrc121336.pdf.
- OECD, 2019, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*. Available at: https://www.oecd.org/publications/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm.
- Purtova, N., 2018, *The law of everything. Broad concept of personal data and future of EU data protection law*, Law, Innovation and Technology, 10:1, pp. 40-81. Available at: https://www.tandfonline.com/doi/full/10.1080/17579961.2018.1452176.

Disclaimer and copyright. The opinions expressed in this document are the sole responsibility of the authors and do not necessarily represent the official position of the European Parliament. Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

IP/A/AIDA/2021-02; Manuscript completed: July 2021; Date of publication: July 2021
Administrators responsible: Frédéric GOUARDÈRES, Matteo CIUCCI; Editorial assistant: Catherine NAAS
Contact: Poldep-Economy-Science@ep.europa.eu
This document includes hyperlinks that are only accessible via its electronic version, available on the internet at: http://www.europarl.europa.eu/supporting-analyses or via the QR code displayed on the first page.

Print ISBN 978-92-846-8408-3 | 10.2861/605827 | QA-08-21-215-EN-C
PDF ISBN 978-92-846-8407-6 | 10.2861/978972 | QA-08-21-215-EN-N