weizenbaum
institut

- Position Paper -
A European Strategy for Data

The Weizenbaum Institute for the Networked Society[1]

About the Weizenbaum Institute

The Weizenbaum Institute conducts interdisciplinary and basic research on the transformation of society through digitalisation. Its aim is to help better understand the dynamics, mechanisms, and implications of digitalisation. To this end, the Weizenbaum Institute investigates the ethical, legal, economic, and political aspects of the digital transformation and creates an empirical basis for shaping it. The Weizenbaum Institute develops options for policy, the economy, and civil society by combining interdisciplinary, problem-oriented basic research with exploring concrete solutions and opening up a dialogue with the society at large.

'Weizenbaum Institute for the Networked Society – The German Internet Institute' is a joint project funded by the Federal Ministry of Education and Research (BMBF). The consortium is composed of the four Berlin universities – Freie Universität Berlin (FU Berlin), Humboldt-Universität zu Berlin (HU Berlin), Technische Universität Berlin (TU Berlin), University of the Arts Berlin (UdK Berlin) –, the University of Potsdam (Uni Potsdam) as well as the Fraunhofer Institute for Open Communication Systems (FOKUS), and the WZB Berlin Social Science Center as coordinator.

Introduction

On 19.02.2020 the European Commission published its Communication on a European strategy for data (COM/2020/66 final). The publication was followed by public consultations, and the Weizenbaum Institute submitted its responses to the Commission's questionnaire on 31.05.2020. The purpose of this position paper is to elaborate on selected topics from the consultation process and highlight several of the observations made by the Weizenbaum Institute on these topics.

The Communication discusses several key issues related to establishing and implementing a coherent and comprehensive European data strategy. Correspondingly, the questionnaire focuses on a range of issues such as making data available for the common good, protecting individual autonomy, enhancing data literacy, developing technological and physical infrastructures, creating and implementing data governance mechanisms, standardisation, data

[1] Authors of the Position Paper are: Zohar Efroni, Martin Florian, Bianca Herlo, Sonja Schimmler, Philipp von Becker, Bennet Etsiwah, Jakob Metzger, Lena Mischau.

intermediaries as well as enabling broader and more efficient re-use and sharing of data between entities.

Strategic goal

The ultimate aim of the EU data strategy, as stated in the Communication, is to capture the benefits of a better use of data, to increase productivity in competitive markets, and, at the same time, to create public benefits in relation to healthcare, environment, governance, and public services, among others.

This overarching goal reveals what we consider to be the main challenge incorporated herein: the **need to create a better framework for economic activities** with and around data, maximising their utility and economic value, bolstering the position of the EU as a leading global player in technology markets, and, at the same time, **protecting both private and public interests** that are representative of European social and ethical values.

Regulatory intervention is needed when markets do not operate efficiently under present conditions (market failure) and when a laissez-faire approach leads to undesirable socio-political outcomes. The strategy therefore needs to consider both public and private-commercial interests while ascertaining the package of measures for achieving its goals.

Defining the strategic goal requires a strong emphasis on empowering individuals/consumers[2] with respect to their data, both in the sense of facilitating more individual control over data *via* legal rights and technological measures as well as investing in skills and data literacy.

Private-commercial and public interests as strategic goals

The basic assumption is that, indeed, the availability of more and better data can accelerate economic prosperity of private (commercial) entities while simultaneously creating public benefits in areas such as healthcare, public services, and sustainability. We note, however, that it is not always easy to neatly distinguish measures aimed primarily at the public good from their (positive or negative) effects on private-commercial actors. Some measures might collaterally benefit commercial interests and *vice versa* (e.g., where the commercial success of private companies contributes to the collective economic growth and increases societal welfare).

Such cross-effects can indeed lead to a convergence of benefits for different stakeholders. At the same time, it is important to be mindful of the actual impact of the considered measures and the interests they primarily serve - and to **make sure the measures remain, in fact, aligned with their policy objectives**. When creating new rules or processes, it is vital to ensure that a vague definition of 'the public good' is not exploited and used as a veil for promoting commercial interests above everything else. A strong focus should be on making sure that data protection laws and individual self-determination are not undermined.

[2] The term 'consumer' here and throughout the position paper is used in the broad sense of every person who uses or consumes digital services and content - synonymic with a 'user' or simply to any 'individual' who participates in the data economy.

We would also like to stress that data access for **scientific research** must be continuously developed and refined on a strategic level. This must be done in compliance with data protection principles and in line with state-of-the-art anonymisation standards, data security and organisational data protection measures.

Protecting individual rights and interests

We agree that, as an overarching goal of the EU data strategy, it is essential to place the **interests of individuals first** and in accordance with European values, fundamental rights, and rules. At the same time, as recognised in the Communication on several occasions, it is also essential to provide the **conditions for functioning data markets** to the benefit of multiple stakeholders, including commercial actors (especially start-ups and SMEs), the research and education community, and the public at large.

One of the key steps towards furthering the strategy's objectives is to facilitate **viable options** for consumers/data subjects to manage access to their data and educating them about these options, including the possible negative consequences of granting others access to their personal data. This contributes to creating the conditions for **authentic and self-determined individual choices**. Individuals should be effectively empowered to make better decisions related to the use of their personal data.

We therefore believe that the strategy should focus on the problem of information asymmetries, the dysfunctional competition in certain areas, and on the practical barriers to citizen empowerment, including the structural **difficulties to exercise valid consent** under the requirements of the GDPR (i.e., freely given, specific, informed and unambiguous), and also **to refuse to give consent** without significant personal detriment in the online space, and other factors creating **market failure** within the data economy.

That said, we recognise that educating the general public on technologically and economically complex data issues and relying on individuals to make the best possible decisions with regard to their data will reach its limit at some point. We further recognise that problems such as information asymmetries, information overload, cognitive biases, market dominance/concentration, and lock-in effects cannot be solved by feeding individuals ever more information and by opening up more options to them. The entire architecture of the data economy needs rules, tools, and mechanisms that facilitate autonomous choices in addition to providing more information and mechanical 'consent' procedures.

In addition, enhanced **trust** of consumers in existing and new products, services, organisation, technologies, and innovations premised on data utilisation and exchange is germane to achieving the strategy's goals. Safeguarding **high standards of data protection and data privacy** are essential for building and maintaining this trust. Anonymisation/pseudonymisation of personal data as well as data security are also crucial for reducing the risks for individuals as well as for rendering the data more exchangeable in strict compliance with data protection and data safety laws and standards.

Data portability

As noted, some measures might have multiple effects and serve a variety of interests simultaneously. One example are measures that enable individuals to facilitate access to their data across multiple entities (data processors, data controllers). We believe that enhancing the scope for individuals to effectively control and grant access to existing data – in the sense of **data portability**, in particular – is a necessary and important component of the strategy.

Improving data portability can not only help underpin **informational self-determination** on the individual level and reduce switching costs between providers, but also **stimulate competition** to the benefit of smaller innovative businesses that offer substitutive or complementary products and services. Prominent examples for such provisions can be already found in the data portability regimes of Art. 20 GDPR, Art. 16 Digital Content and Services Directive, and Art. 66, 67 Payment Services Directive 2. It would seem particularly useful to give consumers/data subjects the right to require the respective data processor to **directly transfer the data to the new data processor**, as laid down in Art. 20(2) GDPR.

Further measures should be carefully examined by asking whether data portability actually helps to bring about the intended affects in a given context. In particular, care should be taken that the concept of data portability cannot be misrepresented or misused and that individuals do not end up granting access to their data in an excessive or a coercive manner.

We therefore argue that, in principle, only individuals (the consumer, the user, the data subject) should actively initiate the process of granting access to their data. The **inherent implications and risks of such access should be made transparent** to the individuals before making a choice. This decision-making process will need to be helped along by technical and visual means to overcome complexities, information asymmetries, information overload, and cognitive-behavioural deficits. To fully achieve their intended purpose, data portability rules must be supplemented by enhanced data literacy, a diversity of comparable/substitutive offers as well as ICT infrastructure measures, including interoperability and technical standards.

Areas in which additional data portability mechanisms can be considered are, for instance, **access to real-time data** in certain contexts such as predictive maintenance conducted by companies for improving their offerings. In other contexts, it might be useful to consider **portability of non-personal data** as well. For instance, the Digital Content and Services Directive currently facilitates data portability for non-personal data – but only upon termination of the underlying contract. It does not grant the right to have the data directly transferred between data processors. For the purpose of multi-homing or allowing the (potential) consumer to test competing services, it is recommended to allow data portability for non-personal data before the termination of a contract. Finally, more attention should be paid to **data portability within B2B** relationships.

In any event, existing legal restrictions on data portability, such as protection of trade secrets and rights pertaining to databases, will have to be taken into account.

In addition to legal entitlements of data subjects to exercise (some) control over the use of their data, innovative technologies and data management structures should be improved to support users in managing access to their personal data. Personal Information Management Systems (PIMS) or data trusts (see below) are often highlighted in this context. These tools can provide critical assistance to users attempting to control, manage, and plan the use of their personal data in a much more self-determined manner than it is currently often the case.

Personal Information Management Systems (PIMS) and data trusts

PIMS and data trusts are still in their early stages of development, both technologically and conceptually. There are several variations to the logical, logistical, and organisational structure of these approaches (e.g., creating centralised data pools that are managed by third parties on behalf of the users vs. local data storage and self-management via user interfaces). Different architectures come with different benefits and risks. When deliberating about these issues, one should be mindful of an underlying question - **what problem are these solutions exactly meant to solve?** These tools may serve more than one purpose, for instance, improving informational self-determination of individuals, improving data and consent management of organisations, or facilitating data re-use and exchange among commercial/non-commercial actors, to name a few.

In our opinion, one central objective these tools must address is to **improve informational self-determination**. Any viable solution will have to be accompanied by measures enhancing the ability of individuals to understand their choices as well as the process and consequences of granting access to data in a given case. Specifically, users must have **sufficient knowledge** about what data are being used, by whom, and for which purpose. Additionally, they require information on possible data sharing with third parties, and, even more importantly, the potentially negative consequences of permitting access to their data.

The strategy, as aptly stated by the Communication, should mobilise the development of technological tools as well as trusted intermediaries that are supported by goal-oriented and value-driven organisational/governance structures, standards, ICT infrastructures, regulation, and enforcement. We strongly agree that tools for consent management and, more broadly, **personal information management, which involve novel technological and organisational models** implemented by neutral intermediaries (meaning administrators that do not benefit commercially from the data they manage and carry fiduciary duties vis-à-vis data subjects), bear **significant potential** for mitigating market failure in line with European legal rules and ethical values.

These technologies and institutions are potential pillars to support data altruism and public interests/public goods, including academic research and education, cultural initiatives, public health, environmentalism, and sustainability goals alongside commercial activity within the ‘European data space’. In these areas, the strategy should provide a **regulatory framework** (top-down) concerning the general structure, framework requirements, governance, quality control, providers’ duties, etc. combined with **incentives –** and not rely solely on self-regulation and the implementation of voluntary mechanisms (bottom-up).

Data literacy

The EU data strategy emphasises the **importance of data literacy**. Every individual is a participating actor in the data economy to a certain degree – as a provider or even as a processor of data. While a minority of the population might already be considered ‘data-literate’ by education or trade, the vast majority is ‘data-illiterate’, meaning that this group of agents lacks sufficient understanding of how and to what end their data might be used by third parties and whether they are likely to be personally affected by their use. The latter group is often heavily involved in data-driven innovation, while lacking the means to make informed data-related decisions. These individuals represent a relatively large group of society that would greatly benefit from even small improvements to their **data-related competences and skills**. This may

include educating data subjects on their legal rights and on ways to enforce these rights effectively.

One way of improving informed decision-making is through **education**, both on the general level of improving data literacy and on the individual level (i.e., users need to have better informational tools to assess the risk in a concrete case). Among others, **standardised privacy icons and other visual solutions** have been suggested to denote risks and consequences to individuals in a clear and intelligible way.

Public funding

It is clear that **major public investments** should be made in technologies and infrastructures that enhance data access and use. We believe that **open data** concepts should play a prominent role when setting out the priorities for EU funding. Once the principal goals of the data strategy and the balance between individual and private interests have been clearly laid out, funding should flow to those projects and initiatives that are likely to serve these goals in the best way possible.

We agree that the development of **European data spaces** should be supported in certain **strategic sectors** and domains of public interests, such as the manufacturing industry, mobility, environment, health, energy, agriculture, public administration, and education, among others. It seems to us that, in principle, **a bottom-up, diversified approach is favourable here**, meaning that the development of multiple smaller projects in each sector should be encouraged. This would stimulate competition between multiple initiative towards the best solution in that sector. The alternative of orchestrating one large, top-down data space for each sector would instead create a single point of failure and stifle competition. Additionally, **sustainability and interoperability** of the developed systems should be ensured.

Anonymisation

Anonymisation technologies can reduce the risk to privacy interests resulting from data sharing. However, a word of caution is warranted when discussing ex-post anonymisation, because what it can achieve is often exaggerated. Generally, **dataset anonymisation can neither be regarded as a panacea nor a one-size-fits-all solution**. Rather, it needs to be carefully adapted to each individual use case. In any case, applying anonymisation to datasets must always be balanced against preserving the data's explanatory power.

As shown by numerous works on the deanonymisation of datasets, **anonymisation is often hard to 'get right'** in practice, even for technology companies with extensive expertise on the topic. It should be avoided – even prohibited to some extent – to create a false impression of (absolute) privacy by offhandedly labelling a modified dataset as 'anonymised', when, in reality, the modifications do not achieve real data anonymisation and fall below acceptable standards.

Ideally, principles of anonymisation should be applied from the onset – by practicing data minimisation already during the data collection process and in this way avoiding to create datasets replete with markers enabling (re-)identification.
Proposals to **outlaw de-anonymisation** techniques to protect anonymisation as a measure to ensure privacy should be examined very carefully and critically. Anonymisation research and

the standards and technologies it creates benefits from exploring de-anonymisation, because it helps to reveal a method's weaknesses, gaps and vulnerabilities.

Data governance

The Communication and the questionnaire also inquired about the need to install cross-sectoral data governance mechanisms. We agree that **data governance** mechanisms are needed to capture the enormous potential of data, especially for **cross-sectoral** data use. Where feasible, such data governance mechanisms should be constructed, because a cross-sector perspective makes it possible to establish **overarching rules, tools and processes** with extensive unifying effects across different fields.

However, we would also like to point out that **sector-specific measures** can be equally important. It is essential to keep in mind that, under some circumstances, the specific features and idiosyncrasies of certain sectors make sector-specific approaches preferable for developing several independent solutions.

Regulation and standardisation

We agree that **standardisation** would significantly benefit the **re-use of data** in the economic sector and the society at large – and specifically for the purpose of improving interoperability (technical standards). However, standardisation measures should not be, as they occasionally are, implemented in a top-down and involuntary way.

When it comes to standardisation for interoperability, a topic that was specifically addressed by the questionnaire, we believe that strict and detailed regulation might, in fact, inhibit innovation. Instead, we recommend using coercive measures only in specific cases, subject to necessity, on a case-by-case basis, and as an exception to the rule.

To be sure, the **regulatory bodies** on the EU and Member State level can play an important role in the area of standardisation. Among other things, the regulator can ensure the **participation** of economically less powerful stakeholders in co-designing standards, channel **funding** towards improving re-usability and open-source schemes, and **coordinate and organise** commercial and non-commercial efforts, including testing and implementation. Regulators could take the lead in embracing open standards and demand they be used by contractors. They may also use funding decisions to foster the development of sustainable tools and systems.

Regulation, R&D and sensitive data

In principle, the regulator should also facilitate research and development in the public interest by making data available to R&D enterprises, including sensitive data. Again, such efforts to further **research in the public interest** should be guided by clearly defined goals and be genuinely geared towards producing collective societal benefits. That said, even broad societal benefits cannot justify an unrestricted use that compromises data protection rules and principles.

In the case of **sensitive data**, which the questionnaire specifically addressed, a rigorous discussion based on specific use cases and specific types of data is especially warranted. At the same time, data protection restrictions should neither be used to mute this discussion nor to preclude the potentially beneficial use of sensitive data *ab initio*. Access to sensitive data for scientific purposes, for instance, in the field of healthcare, can be a prerequisite for path-breaking research of critical importance.

Self-regulation

Overall, self-regulation of commercial actors in the broader sense (namely, beyond the specific area of developing technical standards discussed above) has not led to satisfactory results in the B2C sector, e.g., in the context of data protection, consumer protection, or data governance. Reasons may include a lack of commercial incentives to adopt self-regulatory measures, a lack of coordination, and a lack of organisational and technological frameworks to support the process.

At the same time, self-regulatory measures can and should be encouraged in areas where top-down regulation does not yield results, for instance, **data sharing among commercial actors** in the B2B sector (with some exceptions), and, as noted above, in the area of **developing technical standards.** We believe that, in principle, B2B data sharing should be premised on providing commercial actors with guidance and incentives to share data rather than coercing them to do so through binding, detailed regulation.

Conclusion and recommendations

We warmly welcome the efforts of the Commission to develop a comprehensive EU strategy for data. To do so with an efficient and consistent package of measures within a participatory process involving relevant stakeholders is important and necessary. We consider the following **measures and focal points** to be particularly significant and urgent in this context:

* **Collecting and processing data** should only be possible with a strong commitment to **fundamental rights**. High standards of data protection and data privacy are significant achievements that must not be undermined, even when they are competing with commercial interests or with prospects of data wealth and economic growth.

* The **right to informational self-determination** should be recognised and practiced, and supported by legal, institutional, and technological mechanisms that are aligned with European liberal-democratic principles.

* The success of the strategy critically depends on the ability to strike an **appropriate balance** between individual, commercial, and public interests concerning the use of data, and to ascertain the **right level of regulatory intervention** in data markets and technological innovation.

* Implementing **data literacy** at all levels of society, e.g., through broad-based and continual education programmes, is of fundamental importance.

* The vast **power inequality** between individuals, on the one hand, and large commercial entities, on the other hand, should be recognised and addressed by **regulatory measures** in areas such as competition, consumer protection and data protection laws, among others.

* It is important to **promote open standards and interfaces** for sharing data. Government agencies should set an example in this respect and commit to the responsible use of data in the public context and in the public interest.

* Open data standards can be promoted by imposing **certain obligations to share data** by public agencies, including in their relationship with commercial companies, for instance to achieve improved access to and (re-)use of data generated by customer-operated devices, machines, and software.

* Efforts should continue to further develop **data trustees, privacy-enhancing technologies, and PIMS** (conceptualisation, structuring, and funding) on multiple levels, including modelling, architecture, technology, standards, governance, and legislation.

* The discourse on **data rules and necessary measures** should differentiate between instances of data handling. Data use and re-use should not be regarded as an end in itself, while disregarding the **underlying policy goals** and the **actual effects** of data use and its regulation in the economic and the social sphere.

*****