Data Governance Act Proposal

A position paper by the research groups “Frameworks for Data Markets”, “Work and Cooperation in the Sharing Economy”, “Trust in Distributed Environments”, “Responsibility and the Internet of Things”, and “Reorganizing Knowledge Practices” of the Weizenbaum Institute for the Networked Society
Abstract

This Position Paper contains statements drafted by several Research Groups at the Weizenbaum Institute concerning the Data Governance Act (DGA) Proposal. Each statement is followed by a short explanation. The purpose of this Paper is to highlight a number of important aspects of the DGA Proposal and stimulate the debate around it with a special emphasis on the part that concerns regulation of data sharing services (Chapter III, DGA Proposal). The Paper touches upon a number of selected matters without the ambition to cover all the important issues the DGA legislation raises. The statements address the potential risks in creating a centralized architecture for data intermediaries, the problem of imposing a duty on data sharing services to offer data on a non-discriminatory basis, the role and expertise supervision authorities will need to assume and exercise and questions regarding the interface between the anticipated DGA and existing data protection law in the EU. The Paper includes a number of specific recommendations regarding the formulation of several DGA provisions, specifically in connection with its intersection points with the GDPR.
Table of Contents

1 Architecture and Governance of Data Intermediaries 5

1.1 Centralized architecture that amasses large quantities of personal and confidential data in the hands of data intermediaries might increase the potential for privacy breaches and power asymmetry risks. 5

1.2 Ensuring non-discriminatory data access via data sharing services is not sufficiently guaranteed. In particular, there is a lack of precise specifications for the required design in practice. 5

1.3 Supervisory bodies and consultation services in connection with data sharing via intermediaries must be equipped with significant in-house expertise to ensure the effective application of privacy-enhancing measures such as anonymization by the data intermediary. 6

2 The Interface between Data Protection Law and the DGA 7

2.1 The DGA should build on the strong data protection laws in the EU already in place. It should not be conceived as altering or limiting the rights and duties stipulated in existing data protection laws. 7

2.2 The DGA should apply to the sharing of personal data. Excluding such data would lead to a lower level of protection and security for personal data. 8

2.3 Including personal data further normalizes the commercialization of personal data. This is a political decision and it should be a conscious one. 8

2.4 The DGA should prominently recognize the general applicability of the GDPR. Isolated references to the GDPR should be avoided. 9

2.5 The DGA should differentiate between the legal consequences for personal and non-personal data only where such differentiation is absolutely necessary. 9

2.6 Instead of introducing new legal terminology in relation to personal data, the DGA should use the terminology established under the GDPR to the extent possible. 10
3 Specific Formulation Proposals

3.1 The DGA should exclude „data subjects“ from the definition of „data holder“. “Data subjects” should be defined by a dynamic reference to the GDPR.

3.2 The term „access“ in Art. 2(5), (6), (8) DGA Proposal should be replaced by the term „processing“.
1 Architecture and Governance of Data Intermediaries

1.1 Centralized architecture that amasses large quantities of personal and confidential data in the hands of data intermediaries might increase the potential for privacy breaches and power asymmetry risks.

Data intermediaries such as data sharing services are a central pillar of the current DGA concept. They are expected “to increase trust in sharing personal and non-personal data and lower transaction costs linked to B2B and C2B data sharing” (DGA Proposal Explanatory Memorandum, Section 5). The current proposal sets out to outline requirements and a notification regime for data sharing providers so that they gain the trust of data producers. The specifics of how data sharing services can be designed and operated have so far largely been left open, raising concerns that the encouragement of such novel data intermediaries might lead to new and exaggerated privacy and confidentiality risks.

Data intermediaries designed as data storage and redistribution platforms necessarily have full access over the data they store. Additionally, popular data intermediaries will likely mediate data from a large number of original data sources, making them an especially attractive target for cyberattacks, among other things. The possibility that data intermediaries could reach near-monopoly status in a given data sharing context might actually be considered a desirable effect in the context of reducing data sharing costs, as such intermediaries would become “one-stop shops”.

The centralized storage of large amounts of sensitive data is to be scrutinized in the face of possible threats from cyber criminals and foreign intelligence services. Additional risks result from the fact that while the misuse of stored data can be constrained through appropriate legal framing and organizational measures, there is no guarantee that protection standards and control mechanisms will retain their quality throughout the whole duration of data storage. If legal and organizational conditions change, the collected data might lead to the affected data subjects and enterprises being harmed in ways considered unacceptable at the time of initial data collection.

Risks resulting from excessive and excessively centralized data collection can be mitigated by disincentivizing data intermediaries from storing sensitive data in the first place. From an IT security standpoint, each additional data storage location implies an additional point at which data can be leaked, stolen or (unlawfully) monitored.

Depending on the anticipated data usage context, it might be fully sufficient for data sharing services to store only rigorously anonymized data. Notably, this approach does not prevent data sharing services from supporting data producers with the anonymization of their data. Data sharing services can offer consultation services or even perform the anonymization themselves, as long as they can be obliged to delete their copies of the original data immediately after anonymization. Lastly, concepts for data sharing services are conceivable in which intermediaries are only the middlemen between the data providers and data users. Data can then flow directly between the providers and users, thereby minimizing the number of additional data copies.

1.2 Ensuring non-discriminatory data access via data sharing services is not sufficiently guaranteed. In particular, there is a lack of precise specifications for the required design in practice.

The DGA emphasizes the importance of „fair, transparent and non-discriminatory“ access to data not only for the re-use of protected data held by public sector bodies but also in particular for the
requirements for data access via data sharing services. According to Art. 11(3) of the DGA Proposal, a data sharing service provider has to ensure that the process for access to its service is fair, transparent, and non-discriminatory for both data holders and data users – with regard to the conditions of access as well as to prices. However, aside from this vague objective and with regard to data sharing services, the legal text of the DGA Proposal does not provide precise specifications of how non-discriminatory access is to be ensured in practice. Also, the recitals remain silent on this point, offering no further guidance for interpretation. In this respect, it is left to the Member States, through the competent authorities with supervision powers under Article 13 of the DGA Proposal, to set more precise instructions. More specific guidelines in the DGA would be preferable, however, especially with the aim of supporting small and medium enterprises (SMEs) and ensuring a more systematic harmonization across EU member states.

Particularly concerning the prices to be paid for the data provided, there is a risk of disadvantaging smaller and/or financially weaker companies (such as non-profit organizations). Even if prices and access conditions are equal for all data users (e.g., while applying uniform prices), the price might be prohibitively high for smaller actors and might therefore inhibit participation in data sharing services of SMEs or research institutions. Due to the DGA’s notification regime of data sharing services and the associated special status of data sharing service providers, risks of indirect discrimination should be addressed in the DGA.

Even though the risk of discrimination is generally taken into consideration in Art. 11(3), unlike the provision on data re-use by public sector bodies (see Chapter II of the DGA Proposal), the provisions on data access via data sharing services (Chapter III of the DGA Proposal) do not provide any practical guidance beyond this point. It is important that SMEs and non-profit organizations without large capital resources can also participate in the new opportunities for shared data use. Desired innovations often happen at the level of startups or research institutions.

In particular, there is a misalignment between data re-use from public sector bodies and data sharing via data sharing services. For comparison: In Art. 6(4) of the DGA Proposal, it is expressly stipulated that public sector bodies shall financially incentivize the re-use of data for non-commercial purposes and by SMEs in line with state aid rules. Comparable guidelines should also be established with regard to data access via data sharing services. For instance, the DGA could contain provisions mandating basic public funding possibilities in line with state aid rules, thus facilitating participation in data sharing services for specific categories of companies and/or types of data uses. Without any such provision, participation of SMEs in data sharing services cannot be sufficiently ensured.

1.3 Supervisory bodies and consultation services in connection with data sharing via intermediaries must be equipped with significant in-house expertise to ensure the effective application of privacy-enhancing measures such as anonymization by the data intermediary.

The choice of suitable technical and organizational measures for balancing data usage with associated privacy risks as well as the actual potential and limits of such measures are decisively influenced by the specific context of the desired data usage. Approaches such as anonymization, pseudonymization or the generation of synthetic data are not universal solutions that achieve satisfactory results by themselves. Each data usage scenario is associated with an individual notion of utility, and different types of data imply different types of privacy threats and linkage-based deanonymization possibilities. For example, location data requires inherently different approaches to anonymization than tabular health data.
A wide range of successful de-anonymization attacks on (in hindsight) inadequately anonymized data sets furthermore demonstrate that effective, utility-preserving anonymization is a non-trivial task. As recent results show, even promising new approaches such as the generation of synthetic data from privacy-relevant data sets are demonstrably not better than anonymization in terms of the data utility / privacy-preservation trade-offs they imply. It must also be noted that the state of the art of both privacy enhancement and privacy attacks is constantly evolving. Even in cases where the particularities of individual specific contexts are taken into account when developing a data protection strategy, deployed systems might need to be reexamined periodically.

One consequence to be derived from the above observations is that privacy-relevant data should be handled with extreme care, even if it is reportedly anonymized. Control bodies must be established to continuously monitor key actors of the data sharing economy, including large data sharing intermediaries, evaluating their operation in light of the current state of research. Consultation bodies should provide objective expertise with respect to privacy risks and the choice and implementation of specific technical and organizational data protection approaches. The difficulty of maintaining privacy standards while achieving adequate data utility makes the availability of adequate expertise a necessity. To ensure the objectivity of control and consultation bodies and reduce their dependence on large private-sector actors, control and consultation bodies should be equipped with sufficient organizational and financial flexibility to attract and maintain relevant technical expertise in-house. Significant in-house expertise on technical matters is a key requirement for any agency tasked with overseeing the correct realization of data protection regulation and other data-related regulation.

2 The Interface between Data Protection Law and the DGA

2.1 The DGA should build on the strong data protection laws in the EU already in place. It should not be conceived as altering or limiting the rights and duties stipulated in existing data protection laws.

The DGA Proposal is rather clear when it comes to the intended relationship to the GDPR, as it repeatedly states that its rules should “be without prejudice” (Art. 3(3)(2), 9(2), Recitals 3, 28 DGA Proposal) to the GDPR, and specific norms such as the European Data Altruism Consent Form (Art. 22 DGA Proposal) should be “in full compliance with the data protection rules” (Recital 39 DGA Proposal). Furthermore, the DGA even aims to increase “the control that natural persons have over the data they generate”. [1] In conclusion, from a data protection perspective, the DGA builds on existing data protection legislation and is not intended to alter said legislation.

This approach is a good choice as current data protection legislation guarantees a thorough level of self-determination, security and trust for data subjects. While the economic value of data is undisputed, lowering the safety net is not the right approach. This does not only hold true from a fundamental rights perspective, but also from an economic point of view: While the main purpose of data protection laws should be the protection of the individual’s personality rights, European data protection legislation and especially the GDPR have a positive auxiliary effect. The European Union has managed to create a worldwide leading level of trust that can be used by companies as a brand selling point (cf. the certification mechanisms in Art. 42 GDPR) and could drive business within the European Union. This long-term advantage should not be dismissed in favor of short-term economic considerations.
2.2 The DGA should apply to the sharing of personal data. Excluding such data would lead to a lower level of protection and security for personal data.

It would not be appropriate to exclude personal data if the GDPR and DGA are supposed to complement each other, as stated above. The sharing of personal data represents an important part of the ambition to support the economy through simplified data traffic. The European Data Strategy already states the intention to establish a single European data space.[2] It is noteworthy that the majority of economically relevant data is personal data. Therefore, the DGA would lose a large part of its scope if personal data were excluded. The objective of promoting data trading can only be achieved if personal and non-personal data are regulated holistically.

The DGA is intended to support the enforcement of a high level of data protection, which was established by the GDPR. However, without the inclusion of personal data, the additional assurance of the level of protection would be omitted. A regulation that complements the protection of the GDPR can be found, for instance, in Art. 11(10) DGA Proposal, which intends to assist data subjects in exercising their rights, in particular by advising them.

Finally, from a practical perspective, it is difficult to distinguish between personal and non-personal data. As noted in the EDPB-EDPS joint opinion, this problem exists especially when large collections of data sets are involved.[3] It must be taken into account that the bundling of data is a certain consequence of the DGA.

2.3 Including personal data further normalizes the commercialization of personal data. This is a political decision and it should be a conscious one.

While the only viable solution is to include personal data within the Data Governance Act, doing so adds to the ongoing legal acknowledgement of the idea of a commercialization of personal data. A first step in this direction was taken with the adoption of the idea of data as a counter-performance in the Directive on Digital Content and Digital Services (DCDS)[4], where the situation was similar: Leaving out personal data would have lowered the level of protection for customers paying with their data even further.[5] However, through such developments, the idea of monetizing personal data is subtly furthered. Although there have been some critical voices in the past, there seems to have been little public discussion about the question of whether we – as a society – actually want to advance the commercialization of personal data and, if so, how this can be achieved while at the same time upholding the individual’s right to self-determination. Once again, it seems this discussion is being avoided in the current debate. On a similar note, the EDPB-EDPS conclude „that this policy trend toward a data-driven economy framework without a sufficient consideration of personal data protection aspects raises serious concerns from a fundamental rights viewpoint“, and specify, „The clear incentive to ‘monetize’ personal data also increases the importance of checks on data protection compliance. Regrettably, in this regard, as well as in relation to the other chapters of the Proposal, the impact assessment does not take the data protection risks into account.”[6]

When compared to the commercialization of other personal rights, especially the right to one’s own image, the lack of discussion becomes even more apparent. As early as the invention of photography, questions surrounding the commercialization of one’s own image have sparked a broad discussion covering judicial disputes, legal and economic articles as well as general public discussions.[7] In a similar manner, the possibility of a commercialization of moral rights has been, and still is, at the center of legal and political discussions, focusing especially on the limits of such a commercialization, established in some jurisdictions through inalienable parts of the droit moral.[8] In contrast, there seems to be little to no substantial debate or...
legislative efforts concerning the monetization of personal data. While the GDPR mandates general rights and duties concerning the processing of personal data, it does not explicitly regulate the aspect of a commercialization of such data. Although the so-called „prohibition of coupling“ (Art. 7(4) GDPR) can be seen as proof of an engagement with this question, its final wording turned out to be so vague that it has had little effect on business-models of personal data as a counter-performance.

The intention here is neither to make an argument in favor or against the commercialization of personal data, and it is not proposed to either explicitly acknowledge or prohibit such business models. The aim is merely to shed light on this discussion that, considering its relevance, seems underdeveloped.

2.4 The DGA should prominently recognize the general applicability of the GDPR. Isolated references to the GDPR should be avoided.

Since the delineation between personal and non-personal data is essential for implementing data protection, the relationship between the GDPR and the DGA should be clarified explicitly at a prominent spot within the DGA’s main body. The regulation of a horizontal applicability of the GDPR, as proposed in Art. 1(3) DGA-Council (Presidency compromise text of 22 February 2021 (2020/0340(COD))), would prevent the risk of creating an additional legal basis for the processing of personal data and would integrate the principles of the GDPR.

In addition, the DGA should have no additional specific references to established rights and duties under the GDPR, unless strictly necessary. If the DGA is clear about the unconditional applicability of the GDPR, such additional links can lead to legal uncertainty. If specific rights or duties are referenced in specific contexts, the omission of other rights or duties or of the same rights or duties in different contexts of the DGA would lead to uncertainty as to why a specific right or duty was (not) mentioned in a specific context of the DGA.

For example, Art. 22(3) DGA Proposal states that „the European data altruism consent form shall ensure that data subjects are able to give (...) and withdraw consent“ in compliance with the GDPR. This raises the question of why the possibility of giving and withdrawing consent was mentioned and whether, as an argumentum e contrario, other duties of the GDPR, especially informational duties, should not apply. Similarly, Art. 9(2) DGA Proposal states that the „Chapter [concerning Data Sharing Services] shall be without prejudice to the application of other Union and national law“, including such on „the protection of personal data“. Again, this leads to uncertainty as it might lead to the conclusion that, for example, Chapter IV on data altruism should indeed be with prejudice to data protection legislation. In conjunction with the intended general reference, such references appear unnecessary, if not misleading, and should be removed.

2.5 The DGA should differentiate between the legal consequences for personal and non-personal data only where such differentiation is absolutely necessary.

As the demarcation of personal and non-personal data is exceedingly difficult, from a compliance perspective it is recommendable not to mandate differing legal consequences for each type of data, unless unavoidable. At the same time, it can be useful to explicitly mention both types of data for clarification purposes, even when they are treated the same.

This idea seems to be included in the current DGA draft and the proposed amendments thereto. Especially in Chapter II, which complements the Open Data Directive with respect to data protected by third parties’ rights, it is necessary and sensible to only refer to personal data and to implement separate rules where there is a specific need for it.[9] In
other parts, there seem to be no differing legal consequences depending on the type of data.

As noted, explicitly mentioning personal and non-personal data can sometimes make sense when the same legal consequences are mandated for both types of data. Such a distinction could actually prevent inconsistencies with data protection legislation and improve the clarity of the DGA.

Contrary to the statement of the EDPB and EDPS in their joint opinion, the DGA does not blur the distinction between personal and non-personal data by creating a parallel set of rules for personal data outside the GDPR.[10] This problem does not exist if the GDPR applies without exceptions to all personal data within the DGA, as this prevents a contradiction between the two regimes. The primacy of the GDPR also for mixed data sets is noted in the compromise proposal.[11] The foregoing is underscored by the fact that data sets often cannot be easily separated into the categories of person-related and non-person-related. A division of the regulatory text would therefore also be unsuitable from a practical point of view.

2.6 Instead of introducing new legal terminology in relation to personal data, the DGA should use the terminology established under the GDPR to the extent possible.

In order to avoid friction between the two legal frameworks, the DGA should adopt the legal terminology in relation to personal data as set out in the GDPR. The resulting clearer terminological delimitation would lead to more legal certainty and thus to a simpler application of the law. New terminology should only be introduced in relation to non-personal data. Hereby, the high level of protection of personal data under the GDPR regime is ensured and the inseparable link between the right to informational self-determination and the data subject, who should have autonomy of choice over his or her personal data, is highlighted.

3 Specific Formulation Proposals

3.1 The DGA should exclude „data subjects“ from the definition of „data holder““. „Data subjects“ should be defined by a dynamic reference to the GDPR.

The DGA introduces the new term “data holder”. It is defined as a “legal person or data subject who (...) has the right to grant access to or to share certain personal or non-personal data under its control”. The DGA hereby conflates the already existing definition of „data subject“, which is defined in Art. 4(1) GDPR as an identified or identifiable natural person, with the new terminology.

However, for a simpler application of the law and a clearer linguistic delimitation, it is desirable to define the “data holder” and “data subject” separately.[12] By excluding “data subjects” from the definition of „data holder“ and defining it with reference to Art. 4(1) GDPR, a parallel to the GDPR is created, emphasizing the continued standard of protection for personal data in the DGA.

The DGA should, whenever it refers to personal data granted access to (cf. These III. 2.) or shared by „data subjects“, use the terms „data subject“ and „consent“ in the sense of the GDPR, in order to clearly demonstrate the lawfulness of the processing of personal data in accordance with the GDPR. For the incorporation of the term „consent“ in the definitions listed in Art. 2 DGA, a dynamic reference to Art. 4(11) GDPR is also conceivable. Since only a
“data subject” can consent to the processing of his or her data according to this definition, the term „consent“ should not be used for “data holders”. While “data holders” can receive “consent” from a “data subject” - including the right to have third parties process the data – they can never consent themselves on behalf of the data subject. Therefore, whenever data is being granted access to or shared by data holders, the terms „data holder“ and „permission“ should be used. If „data subjects“ and „data holders“ are to be regulated, both shall be mentioned. In addition to adapting the definitions in Art. 2 DGA, the terms should be adjusted throughout the DGA. The clear differentiation between „consent“ and „permission“ ensures that any authorization to process data is always traceable to the data subject.

The exclusion of “data subjects” from the scope of Art. 2(5) DGA has the effect that only legal persons are covered by the definition. However, a natural person can also be a „data holder“ without being a „data subject“. Therefore, natural persons should be included in the definition. In deviation from the Council proposal \[13\], instead of using the term “legal entity”, the term “natural or legal person” ought to be used in accordance with the wording in Art. 2(2), (4), (6) DGA.[14]

### 3.2 The term „access“ in Art. 2(5), (6), (8) DGA Proposal should be replaced by the term „processing“.

Under the current proposal, the term „access“ is defined in Art. 2(8) DGA-Proposal as „processing by a data user of data that has been provided by a data holder, in accordance with specific technical, legal, or organizational requirements, without necessarily implying the transmission or downloading of such data“. The term „access“ can further be found in the definitions of „data holder“ in Art. 2(5) DGA Proposal - a person who has the right to grant access to data - and „data user“ in Art. 2(6) DGA Proposal - a person who has lawful access to data and is authorized to use it.

In the definition, „access“ is described as „processing“. „Processing“ is a term that is already defined in Art. 4(2) GDPR as „any operation or set of operations which is performed on personal-data or on sets of personal-data in electronic format, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.“.

The literal meaning of the term „access“ inadequately describes the practice of processing data. A dynamic reference to the GDPR as a definition of „processing“ in the DGA would be insufficient though, as the definition of „processing“ in Art. 2(4) GDPR only refers to personal data while the DGA captures both personal and non-personal data. The definition of „processing“ in the DGA should therefore be identical in wording, except that it should encompass personal and non-personal data alike. This modified definition would only change the scope but not the concept of „processing“, which technically functions in the same way for all kinds of data.

In addition to defining „access“ as „processing“, Art. 2(8) DGA Proposal further requires this processing to be „in accordance with specific technical, legal, or organizational requirements, without necessarily implying the transmission or downloading of such data“. This requirement arises either from the GDPR or from other legal instruments which apply alongside the DGA. Therefore, the wording is not needed and does not need to be adopted under the definition of “processing”.

---

\[13\]: Council proposal.

\[14\]: DGA proposed.
Endnotes


[9] For example, the possibility to impose an obligation to anonymize or pseudonymize the data in Art. 5(3) DGA.


