Statement as part of the public hearing in the Committee on Digital Affairs on the German Bundestag on November 13, 2024

Anouk Ruhaak,

Data Governance Expert,
hello@anoukruhaak.com,
anoukruhaak.com

Final version: November 12, 2024

Content:

Questions:

- I will answer questions: 3, 8, 12, 15, 16 in section 1, and questions: 1, 2, 5, 9, 13, 14 in section 2.
- I will refrain from answering questions: 4, 7, 10, 11

The shortcomings of the DGA

The core objective of the DGA is to encourage data sharing within Europe, with the aim to foster innovation and research. It is a response to the reality that data sharing by both private and public sector organizations in Europe remains limited. The DGA specifically concerns itself with what is considered sensitive data: data that sits somewhere on the spectrum between open data (data that is deemed non-sensitive) and personal data. In other words, data that cannot be published openly, but might be shared if appropriate controls were in place.

I agree that the possible sensitivity of data is an often voiced reason for not sharing it. In my own experience, there is a high perceived risk that by sharing data, organizations would open themselves up to unwanted scrutiny, or accidentally expose trade secrets. Lowering this risk by providing both technological and governance controls therefore might appear as an adequate way to reduce the barrier to data sharing.

However, what the DGA ignores is that governments and corporations often lack strong incentive to make data available for secondary use - even when doing so carries no risks. That is, they do not experience a direct benefit, the benefit is insufficient to motivate the effort, or is not core to their business model. In other words, the DGA treats the problem of sharing sensitive data primarily as a coordination and trust problem, when in reality it is often a collective action problem.

One could argue that the opportunity for data holders to monetize the re-use of their data could provide appropriate incentives. For the private sector, this strikes me as a weak argument. Either the data has value and companies have already sought ways to monetize it, or the monetization would merely cover the costs of making the data available. For the public sector this argument may hold more sway insofar as budget constraints were the primary reason for keeping data closed.

The arguments above do not apply to data altruistic organizations (DAOs), where data holders often have a strong incentive to make data available for a common good. That said, most successful examples of such initiatives heavily rely on the organization having an established reputation as a good actor in the community. It is therefore unclear to me to what extent formal recognition under the DGA would impact the proliferation of DAOs. Moreover, insofar as DAOs handle personal data, they would need to be compliant with the GDPR, which provides much stricter safeguards.

Question 3: The DGA enables the controlled sharing of sensitive data, but does not require it. Unfortunately, in light of the above, reaping the benefits from this bill would require additional access rights - which falls outside the scope of the DGA. The Data Act is a step in that direction. In addition, we might place additional burdens on data monopolies to proactively report and grant access to datasets. Another option would be the strengthening of open data regulations, as well as their enforcement. Finally, we might grant data access rights to specific categories of trusted actors, such as academics.

In addition to data access rights, we might look for public investments in specific high-value datasets that combine data from multiple data holders, who would be compensated for their efforts. However, as discussed above, it seems unlikely a monetary incentive would prove sufficient.

Question 8: Given the relative ineffectiveness of the DGA and the limited reach, I would assume those interactions are limited.

Question 12: I think the impact of the DGA on the German economy would be limited, unless accompanied by more far reaching data access laws (see question 3).

Question 15: While the impact of the DGA will likely be limited, I would recommend specific financial support for non-profits that have had success in setting up data donation campaigns.

Question 16: Where the Transparency Act focuses on access rights, the DGA primarily concerns itself with voluntary data sharing. Data standards, and specifically open and widely shared standards, are critical for all data sharing initiatives as they enable the aggregation of data from different data holders and facilitate easy processing of data.

Implementation of the DGA/DGG

Data intermediaries are defined as entities that act as the connective tissue between an undetermined number of data holders or data subjects and an undetermined number of data users. Their objective is to facilitate data sharing between those two groups and to monetize the sharing of data.

Question 13: I see data intermediaries as the potential creators of data markets. They are actors in the data ecosystem. Data trusts are a form of data intermediaries.[1]

DAOs, in contrast, are nonprofits that provide the infrastructure and governance to enable large groups of actors to volunteer their data for a good cause.

I evaluated the proposed implementation against the following four metrics:

- Ability of the intermediary to set and enforce rules
- Ability of the intermediary to monitor adherence to rules and enforce them
- Ability of the oversight body to monitor adherence to the regulation
- Ability of the oversight body to hold the intermediary accountable in case of a violation of the regulation.

In my view, the DGA, and by extension the DGG, provides sufficient room for the data intermediary and DAOs to set context-specific rules that data users need to adhere to. It also specifically puts an obligation on the data user to report any (accidental) re-identification of data subjects or trade secrets.

What could improve the effectiveness of this regulation would be to provide guidance on different kinds of governance models a data intermediary might implement. For instance, in a many-to-many data sharing system, where data from many data holders is shared with many users, the data intermediary or DAO might benefit from clear guidelines on how to engage the data holders and receive their consent, as well as guidelines on how to monitor the way data is used by data users.

The powers of the regulatory body to oversee the data intermediary seem reasonable and sufficient. And in my opinion the proposed penalties appear fair. (**Questions 5, 14)**

Question 1: The provisions seem more than adequate.

Question 2: I would highly recommend looking for international standards for publishing metadata and draw inspiration from best practices in open data (e.g. FAIR data principles).[2]

Question 9: The fee structure seems appropriate.

---

[1] https://algorithmwatch.org/en/data-trusts/

[2] https://devinit.org/wp-content/uploads/2018/01/Metadata-for-open-data-portals.pdf

Anouk Ruhaak

With a background in political economy, software development and system thinking, I'm well-positioned to provide a holistic view on data governance.

As a Senior Fellow at the Mozilla Foundation, I pioneered work on data trusts and data commons. I have spoken at numerous conferences, written extensively on data governance and collective consent and serve as an advisor to governments, nonprofits and startups.