„Innovative Datenpolitik: Potenziale und Herausforderungen“
“Innovative data policy: potential and challenges”

Öffentliche Anhörung / Public Hearing

Ausschuss für Digitales/ Committee on Digital Affairs,

Testimony by Dr. Stefaan G. Verhulst

Deutscher Bundestag
26 June 2024, 14:30 – 16:30 hrs CET

Unlocking the Potential of Data: Innovative Policies for Responsible Data Reuse and Addressing Data Asymmetries

Dear Honorary Chair Tabea Rößner and distinguished members of the Committee on Digital Affairs:

I am honored to present this testimony to the Committee on Digital Affairs for the Public Hearing on "Innovative data policy: potential and challenges".

My name is Stefaan Verhulst, Phd, and I am the Co-Founder of The Governance Lab (GovLab) based in New York City and The DataTank based in Brussels. I also serve as a Research Professor at the Tandon School of Engineering at New York University.

The Potential: Responsible Data Reuse

Let me begin by highlighting the potential of data when used and reused responsibly. Although we hear much about the risks of using data–and many of the fears are indeed justified–it's also important to keep in mind the very real possibilities that data offers for advancing the public good.

We live in a datafied world, characterized by an unprecedented supply–even glut–of data. In this world, data has become a critical resource for informing policy and decision-making processes. When properly analyzed and utilized, data can play a critical role in helping policymakers–and other stakeholders–address a range of critical problems, in sectors as diverse as public health, climate, innovation and economic development, combating urban decay–and much more.

Sometimes this data is readily available. Most of the time it is not. One of the areas with the biggest potential–yet also significant challenges–is data reuse - data already collected for one purpose using it for another. Data reuse can provide invaluable insights into current phenomena, help us understand the causes of emerging trends, and guide us in developing effective solutions to pressing challenges. Moreover, analysis from data re-use can serve as a powerful tool for anticipating future developments and prescribing targeted interventions.

Over the last few years, a growing body of evidence has been established to show the potential for data reuse, and how it can ensure that it contributes to the public good while minimizing risk. To illustrate this potential, let's consider the example of climate change.

- The re-use of satellite data, for instance, can provide accurate measurements of greenhouse gas emissions such as methane and their sources.
- The re-use of social media data can create insights on people's perspectives on climate change.
- The aggregation of traditional and non-traditional data can enable the modelling of the long-term impacts of climate change on various ecosystems and vulnerable individuals and groups - such as children.
- The re-use of water utility data, for instance, can help evaluate the effectiveness of different mitigation and adaptation strategies.
- The re-use of cell phone data can help optimize resource allocation for climate resilience projects.

The Challenges: Data Asymmetries and Missed Opportunities

Despite the very potential of data and data reuse, it's undeniable we face significant challenges in realizing data's full societal value.

One of the primary obstacles is a lack of access to high-quality, timely data by the public sector, civil society, and other groups that are working toward the public good.

We live in a paradoxical situation today, marked both by the availability of an unprecedented amount of data, but also by unprecedented asymmetries in access to that data for reuse in the public interest.

I believe that the growing asymmetries between those who have data (often from the private sector) and those who are best positioned to use it for the public good, represents one of the major challenges of our era.

Data policy to date has primarily focused on preventing the misuse of data, often for valid reasons as mentioned earlier. However, this approach has inadvertently overlooked the *missed uses* of data – the opportunities we fail to capitalize on due to overly restrictive policies or lack of innovative frameworks for data sharing and utilization.

Innovative Approaches to Data Policy

Given these challenges, what can policymakers do? What steps can policymakers such as yourselves – and other stakeholders, from the private sector, academia and civil society – take to help maximize the potential of our datafied society and economy, and to

ensure that the benefits of our data age are maximized in as equitable and inclusive a manner as possible?

In the remainder of my testimony, I want to outline some of the key components of a more innovative, flexible, and responsible framework for data policy. Although the elements I outline here are not final or comprehensive, they are based on extensive research, in Europe and around the world, and years of engagement with policymakers, the private sector, civil society, academia, and others.

So let me start by outlining three specific ways in which we can help create a more balanced data ecosystem. The three policy recommendations I am going to outline revolve around addressing three specific forms of asymmetry, what I call:

- Agency Asymmetry;
- Data Asymmetry; and
- Technical and Computational Asymmetry.

1. *Tackling Agency Asymmetry* : *Advancing Digital Self Determination*

Agency Asymmetry refers to the uneven playing field that currently exists for individuals and groups to determine how their data is used. These asymmetries may reflect existing socio-economic inequalities; often, too, they exacerbate these divisions or create new ones.

Traditional data policies, including the GDPR, have relied heavily on the concept of *Consent* as the primary mechanism for individuals to control their data. For most people, consent refers to those checkboxes that pop up, asking us to assent to data use policies we probably never read. The idea behind consent is very important (giving individuals more control and choice over their data) but it is widely recognized today that the mechanism itself is highly insufficient. This is so for at least three reasons:

- First, consent is rarely truly informed due to overly complex terms and conditions.
- Second, consent is typically presented as a binary choice, thus overlooking the need for more nuance in decision-making about data use and reuse.
- And finally, existing mechanisms of consent focus on individual preferences, in the process neglecting community or group interests.

To address these limitations, I have elsewhere argued that we need to complement consent and shift towards a model of digital self-determination. Digital self-determination is a complement to consent that aims to empower individuals *and* communities in determining how their data is used and reused in a more sophisticated manner.

It encompasses three key components:

- First, process innovation: We must develop new ways to engage with people and communities regarding data use, moving beyond existing consent methods. For example, our research has shown that "data assemblies" - citizen assemblies around data re-use - can bring together diverse stakeholders to discuss and decide on data governance issues collectively. We have conducted such data assemblies to understand public sentiments of data re-use during COVID-19 and how young people feel about the re-use of data for or about them as it relates to their health and wellbeing.
- Second, digital self-determination also requires policy and governance innovation: New mechanisms are needed to express community preferences and expectations. One possible such tool–and one we have explored in our research–is the development of a "social license" for data use, which articulates agreed-upon principles and guidelines for responsible data practices within a community or sector.
- And finally, digital self-determination requires product and institutional innovation: We should explore new structures like trusted data spaces, which provide secure environments for data sharing and analysis while respecting individual and community preferences.

These three components are actively prototyped by the International Network on Digital Self Determination, steered by the Swiss Government. The Network is open to other actors and governments, including actors from Germany.

2. *Tackling Data Asymmetry: Open Data and Data Collaboratives*

A second key–and problematic–form of asymmetry is data asymmetry. Data asymmetry refers to the limited and uneven access that exists for those who are best positioned to use data for the public good, especially when that access is cross-sectoral–e.g., access by policymakers to private sector data.

In particular, we have explored the potential–and limitations–of cross-sectoral partnerships that provide access to non-traditional data sources for public interest

purposes. For instance, across the world, we have seen examples where mobile phone data can be invaluable for urban planning, disaster response, or public health interventions.

In Europe, several instruments such as the Data Act and the Data Governance Act have laid a policy foundation to tackle data asymmetries. However, to make these approaches more systematic, sustainable, and responsible, we need policy innovations in several areas:

- First, we need to work on strengthening and prioritizing data demand: too often, data initiatives focus on infrastructure, data availability; they leap for solutions without first understanding the most important questions that data could be used to address. We need innovative data policies and strategies that prioritize identifying and articulating the critical questions and use cases that would benefit from improved data access. This is one of the reasons we at the GovLab developed our 100 Questions Initiative, which seeks to catalyze a cultural shift by generating new questions and reimagining the question prioritization process across different domains. Put simply, we seek to strengthen data demand, so that it is more geared toward serving the public good. Within a European context, there have been efforts to identify high value datasets but without high value questions, identifying such data sets within the private sector is hard to do or justify.
- Second, we need new professions that can help address various forms of asymmetry. For example, we should foster the establishment of "chief data stewards": professionals who can facilitate responsible data access and re-use across sectors. These individuals would provide the human infrastructure necessary for effective data collaboratives and, more generally, promote a responsible, healthy ecology for data reuse and sharing.
- And third, we need to develop new institutional approaches: for example, we should work to support new types of organizations that can act as trusted intermediaries for data sharing. The European Data Governance Act's provisions for data intermediaries are a step in the right direction, but we should also explore other models, like data commons approaches, which could manage data assets in a more collective manner.
- Another approach to mitigate data asymmetries is the establishment of a Data Innovation Fund: a dedicated funding mechanism to support experimentation with new data-sharing models and help scale successful initiatives. This could work similarly to a Universal Service Fund, commonly used in the telecom sector to

bridge the last mile of telco provision, but focused on bringing data together for public interest purposes.

3. *Overcoming Technical and Computational Asymmetry*

Finally, to promote a more responsible, effective and equitable data ecology, we need to address the uneven distribution of technical capabilities and computational resources--what I call technical asymmetries.

Technical capacity is a requirement (if not sufficient on its own) to fully maximize the potential and minimize the harms of our data ecology. Yet it is well established that the global and national landscape is characterized by profound inequities in this capacity, especially within the area of Artificial Intelligence. There are some steps we can take, though, to address these inequities. These include:

- Democratizing Access to Computational Capacity: We need policies and initiatives that provide access to high-performance computing resources for non-commercial and public interest purposes–the very organizations typically most involved in using data for the public good. Such initiatives could, for instance, involve partnerships with cloud computing providers or the creation of public computing infrastructure.
- Developing Purpose-Led Models: We should also encourage the development of AI and machine learning models specifically designed for public interest applications. This approach ensures that advanced analytical capabilities are available for addressing societal challenges, not just commercial interests. For example, a recent report from The GovLab's Open Data Policy Lab provides a range of use cases where generative AI and open data from official sources intersect for public purposes.
- Building Data and AI Literacy: Finally, it is critical that we invest in education and training programs to increase data literacy across society. *Question Literacy* is essential to be able to define what questions should be prioritized in order to make progress in tackling societal challenges (or meet the 2030 SDG goals). Technical literacy is critical to address technical asymmetries (and, indeed, all forms of asymmetry). Enabling more individuals and organizations to effectively use and interpret data–and to do so in a responsible manner–is key to unlocking the potential of the data age. We need competencies to translate data intelligence into decision intelligence.

Recommendations

In conclusion, the potential of data to drive positive societal change is immense, but realizing this potential requires innovative approaches to data policy. By addressing agency, data, and technical asymmetries, we can create a more equitable and effective data ecosystem that serves the public interest while protecting individual and community rights.

Let me build on these observations with some specific policy recommendations for the Committee's consideration. Many of these recommendations can be operationalized and implemented through the envisioned German Data Institute, depending on how it will be structured and managed:

- Join the International Network on Digital Self Determination, to establish a framework for digital self-determination that goes beyond traditional consent models, incorporating community-level decision-making processes for data governance.
- Hold one or more data assemblies - citizen assemblies around the priorities and conditions for data-reuse (of non-traditional data) in the public interest.
- Develop a 100 Questions Initiative for Germany to identify high priority societal questions where the re-use of data, including private sector data, could be valuable.
- Toward that end, invest in Questions Literacy as the fundamental basis to inform data science and AI.
- Launch a national or EU-wide Data Innovation Fund to support access to non-traditional data and experimental approaches to data collaboration for societal benefit.
- Invest in the development of data stewardship as a professional field, including the creation of educational programs and professional certifications.
- Create the function of “Chief Data Steward” within the public and private sector to accelerate data collaboratives.
- Establish public-private partnerships to provide computational resources and technical support for non-profit organizations and researchers working on public interest projects.
- Mandate the inclusion of "data impact assessments" in major policy initiatives, ensuring that data-driven insights inform decision-making processes.
- Develop decision accelerator labs that can advance the translation of data intelligence into decision intelligence.

Considered together, these policy innovations represent a pathway towards a more balanced and productive data landscape. They can right some of the existing - and growing –imbalances in the our technical ecology, and help ensure that the fruits of the data age are distributed more equitably

However, in order for them to be effective, these approaches will require sustained commitment, investment, and collaboration across sectors.

I urge this committee to consider these recommendations as you shape the future of data policy in Germany and potentially influence broader EU-wide initiatives. By fostering innovation in how we govern, share, and utilize data, we can unlock its transformative potential to address our most pressing societal challenges.

Thank you for your attention. I look forward to answering any questions you may have.