

Insideview

Data Sharing: Don't Forget the Costs

INTERVIEW WITH STEFAN BENDER

The discussion about data and its widespread use for the benefit of society continues to be widely discussed. Data sharing is on everyone's agenda. What needs to be considered?

As the German government stated at its Digital Summit in 2022, data continues to be at the heart of digital change. Whether you like it or not, data influence our daily lives, such as production processes and supply chains or our consumer behavior. Therefore, the use of growing amounts of data is a key to innovation and sustainable growth. However, the cost/benefit calculation of data sharing should not be overlooked in the discussion.

Cost/benefit analysis of data sharing? What does this mean?

Data sharing only works if the benefits exceed the costs for all parties involved. The parties involved here are those from whom the data

originates, those who use the data and then us as a society. A different cost/benefit calculation has to be determined for each of these parties. For example, a data provider must comply with legal regulations – such as data protection – and technical challenges to fulfill data security and runs the risk of data misuse. These are clearly cost factors of data sharing. The benefits – e.g., through better data quality because of better documentation, increased use of the data or more relevant results from these data – must exceed the costs.

Can you give a concrete example of this?

Of course: through its Research Data and Service Center, the Bundesbank offers access to sensitive microdata for independent research. For the access, some conditions have to be fulfilled by the researchers, which may vary through the different data sets. If you want to communicate this information to the researchers in a systematic and machine-read-



Prof. Stefan Bender
Head of Data Service Center
Deutsche Bundesbank

able way, you quickly realize: the existing metadata schema describes the content of the data sets, but does not make it possible to systematically document data access rules. Therefore, we – together with colleagues from other central banks in our data sharing network "INEXDA" – have developed the so-called Annodata schema. It not only facilitates data access, but also increases the reproducibility of research results.

Does such a schema also have further practical uses?

Indeed, the machine-readability of Annodata schema was central. Therefore, we were able to develop our automated contract generator on the basis of this schema. That is why we could switch from the manual preparation of our contracts for data access to a software solution. That is digitalization in practice. However, the Annodata schema is only one measure among many by which we reduce the

costs of data access. For example, we also have tools for output control, because only absolutely anonymous results may be published.

All in all, are we on the right track when it comes to data sharing?

The EU Commission adopted the European Data Strategy in February 2020. This is intended to establish an internal European market for data. The EU wants to develop into one of the most secure, agile and attractive data areas in international comparison. On the one hand, data literacy should be broadly increased, but – while maintaining data protection and data security – the provision and access to data should also be made easier. This is the only way to shift the cost/benefit ratio of data sharing to greater benefits for all actors and to use the high potential that lies in the data for the benefit of society.

Thank you for this interesting conversation.