

NEW PATHS TOWARDS OPEN DATA?

IN BRIEF

- Only a few tech companies own the majority of digitally generated data. This makes the development of alternative algorithms and artificial intelligence (AI) difficult.
- With the help of legal mandates, big tech monopolies could be challenged to share their data.
- Data can be organised as a common good to promote innovation in different fields.
- Such measures are not without disadvantages and have the potential to impact individual privacy rights.

WHAT IS IT ABOUT?

Digital data play a key role in modern contemporary processes of value creation. Their accumulation and interconnectedness are of major concern. The more data are collected and connected, the better they can be used to optimize digital services and adjust them to users' needs. A crucial concern is the development of algorithms (e.g. search engines): the more data is available on the relationship between search queries and search results, the better this relationship can be adapted to the users' actual interests or requirements. Despite obvious advantages, the current monopolisation of data has problematic consequences. A small number of companies (Google, Amazon, Facebook & Apple) collect an enormous amount of data with no general public access. At the same time, large amounts of public data are also collected by governmental organisations. In both cases, the information is not accessible to third parties. This centralisation of data is creating asymmetries in power relations, with the danger of the emergence of so-called surveillance capitalism (Zuboff). People are turning into data donors for

selected corporate and governmental interests because AI and machine learning rely on large amounts of data. However, developing alternative data-based technologies is difficult when many actors are denied access to the data (e.g. non-profit search engines or social networks). How can these disadvantages be resolved or at least be minimised?



Credits: ThisIsEngineering/pexels.com

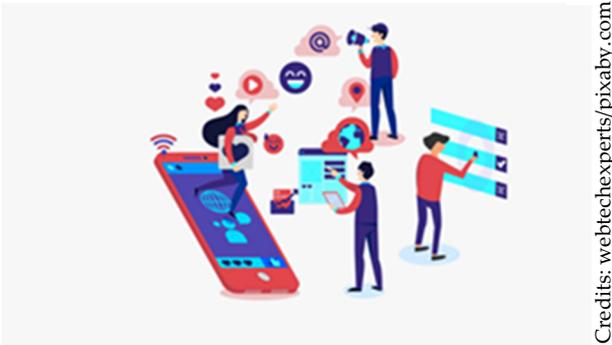
Few corporations own a lot of data about us

Diversification of data use would benefit a broad range of actors and could be accomplished through the following three approaches:

- (1) By using data sharing mandates, previously “closed” data can be shared with a variety of different organisations. Such mandates would require companies by law to release their data if it is requested by a third party. Such disclosure could be made in exchange for monetary compensations or as a free service. Both options are currently being discussed. In turn, this could create an open market for data that would open up new data uses.
- (2) “Open Data” refers to non-commercial data open to the public. The publication of data, however, requires resource-intensive infrastructure that makes the data machine-readable, contextualises it, and makes it accessible to relevant actors. The less context information is available, the more work needs to be invested in the processing of the data.
- (3) “Data Commons” poses the third path towards an alternative organization of data. The term “data commons” refers to data repositories that are generated by communities of users. These communities are responsible for the administration of the data and grant rights of use to third parties. Release and use of the data are subject to certain membership conditions. This includes, e.g., the return of new data generated from existing data back into the commons, a mechanism that allows for the expansion and quality assurance of the data pool, ensuring interoperability and inclusivity.

KEY CHALLENGES

These different strategies to counteract the increasing concentration of digital data reveal different challenges even before their implementation. A major conflict exists between the opening up of data by one of the three strategies and individual privacy rights. Indeed, only anonymised and therefore non-personal data may be published or shared, but the risk of re-identification cannot be eliminated completely when large amounts of data are being released and must be taken into consideration to avoid connecting data to specific individuals. As a result, such risks require the development of novel anonymisation techniques. Digitisation and release of analog files not intended for digitisation also pose new challenges for the opening up of data.



Opening data creates tension

A second challenge concerns the consequences of the data production itself. Data production is always purposive and takes place in a specific context. If such data are shared with third parties, they are removed from their specific context in which they were created. The more specific this context, the more difficult it is for third parties to generate value through the data.

This raises the question of who benefits the most from data sharing concepts? At present, it is feared that the main beneficiaries would be tech monopolies and other powerful actors such as states and intelligence services since they possess the relevant resources required to repurpose old and new data sets, extracting value from them. This power asymmetry increases even more when the legal conditions for the use of data become more complex and the technical infrastructure of the published data is inadequate. To counteract challenges emerging from this inequality, regulation is needed to facilitate the opening up of data.

WHAT TO DO?

**How can digital data be opened up to stimulate innovation in science, business, and civil society?
How can growing data monopolies be disrupted?
The following governance options need to be considered:**

- **Data sharing mandates** could be used to force big tech companies to share their valuable data. They would enable external organisations to legally claim access to large data sets.
- The **protection of individual privacy rights** is essential and must be considered when publishing any public data. Furthermore, (new) **anonymisation techniques** need to be further developed.
- The creation of technical **infrastructure** is necessary for the **opening up of digital data**. This includes machine-readable data formats and the harmonisation of legal licencing standards that would enable the use of data by smaller organisations with fewer resources.
- Legal restrictions regarding the use of data by third parties must be removed. The development of alternative forms of using data can help to rethink categories of private data ownership and authorship. An example of such reconfiguration of data ownership is the concept of **data commons**.

FURTHER READING

Grafenstein, M., Wernick, A., Olk, C. (2019) Data Governance: Enhancing Innovation and Protecting Against its Risks. *Intereconomics*, 54 S.228-232
intereconomics.eu/contents/year/2019/number/4/article/datagovernance-enhancing-innovation-and-protecting-against-its-risks.html

CONTACT

Astrid Mager
Email: tamail@oeaw.ac.at
Phone: +43 1 51581-6582

