

Sharing Data – Is It All About An “Openness” Economy?

Petra Gehring¹

1. Institute of Philosophy, Technische Universität Darmstadt, Germany.

**Date Submitted:**

2024-01-08

Date Published:

2024-02-21

DOI:doi.org/10.48694/inggrid.3932**License:**This work is licensed under [CC BY](https://creativecommons.org/licenses/by/4.0/)4.0 **Keywords:**

data sharing, research data management, open data

Corresponding Author:

Petra Gehring

petra.gehring@tu-darmstadt.de

When research became truly digital, the career of the idea that science must *share* data also began. Truly digital means: at some point, machines were not merely used for calculations (aka *computation*), but rather most stages of the cycle of methodical production of scientific knowledge were routinely performed digitally. This means that the processes for, on the one hand, data collection, gathering, transfer, archiving and, on the other hand, all forms of presentation, such as visualization or making it readable, are digitized. *Data* is synonymous with automatically processed information, but also with the automated perception of the world (aka *empiricism*) and automated methods.

Research that has undergone such far-reaching digital transformation differs in many ways from the science of previous years. It is extremely data-hungry. Yet it is also data-rich, i.e. in possession of data that is of interest to an equally data-hungry digital economy. Consequently, there are two reasons for discussing data sharing: researchers need the (raw) data of other researchers more than ever before. Data is *the* scientific resource par excellence. And: actors competing with the interests of research want to access the data obtained in science as early as possible, for example for commercial purposes. Data handling in science is supposed to be the “new oil” of a ubiquitous data economy. In this respect, global data corporations have long been conducting their own research on a large scale.

Thus, *sharing* is the order of the day. However, the term can mean very different things. Sharing can mean: I own something, but I also let others use it as long as it doesn’t take anything away from me. But it can also mean: I give something away, without price or monetary payment, but it is actually exchanged, so I receive something in return (data for data). Sharing data can also mean using data as a common resource and then sharing revenues or possible profits. And finally, sharing can also mean I give away data unconditionally, without expecting anything in return. Sharing data without loss, carefully and for common purposes in a methodically controlled manner – this is or should be common practice in research. *Opening* data sets to anyone for any purpose – that is more in the latter direction.

Good scientific practice or the sharing economy: there are definitely opposites here. Do we invest in the best possible data quality for research – or do we focus on maximum shareability, allow it to circulate and give preference to generic tools that are not optimized for scientific purposes? *Sustainability* can also be understood in very different ways: a specialist community can conduct sustainable joint research with its data (but without already donating its digital

resources to data corporations), archive it for science and prevent popularizing, perhaps falsifying or inappropriate use (*dual use*) by non-scientists – or the same specialist community can see it as its duty to feed data that has been collected on the basis of taxpayers’ money into the economic cycle as quickly as possible. Even then, of course, the question arises: exclusively share data with selected industry partners – or out with it: *open* data, because *open* is fairest?

Fortunately, by providing data sets, you can do both one and the other. However, it is important to actually reflect on this. “Sharing” requires well-founded decisions. Good sharing is not possible without asking “how?”, “with whom?” and also “with what consequences?”. And since data is a so-called non-rivalrous good (I can take it away from someone while they keep it), the question of the intermediaries that science uses for the purpose of sharing is also important. Do I give data to a publisher so that they can offer it “openly” to an undefined audience? Do I trust this publisher not to reuse it in parallel, to appropriate derivatives and sell them? Do I trust non-profit platforms – even if they can be sold at any time, like the supposedly science-owned provider *GitHub Inc.* which is now owned by *Microsoft*? Do I choose a market-neutral trustee? Do I only give data to public data centers or repositories? Or should the scientific community itself establish more collecting societies in order to exploit data produced in public research through targeted marketing for the benefit of the scientific system? Would the taxpayer thus get a piece of his “investment” back?

Commercialization can harm science. Especially in the digital world, it is surrounded and driven by data services. It is not uncommon that it has to buy back something that was created due to the outflow of its own data. Still, digital PPPs (public-private arrangements) can also benefit science. For example, in cases where methodically obtained high-quality data products cannot (or can no longer) be financed from taxpayers’ money alone, fair industry partnerships not only create economic added value, but also help the scientific system. At the interface to product development, the discussion about the sovereignty of science over its data has not yet really begun. In any case, sharing does not simply mean “everything must go”!

The German Council for Information Infrastructures (RfII) strongly supports the sharing of data within the scientific community. Researchers should share Knowledge, Information *and* Data also across disciplinary boundaries. However, the RfII recommends that the scientific community create suitable, quality-assured data products for the purpose of dissemination beyond the boundaries of science, i.e. for publication. Their use should also be subject to rules (for example, they should not be commercially re-appropriated, but must remain accessible for research). I would add here for the humanities in the era of AI: Copyright (not just labeled authorship) should also be discussed. In any case, the magic spell of “open data” alone is not enough of an answer to the complex question regarding which forms of data sharing really advance research. And science as a whole.