The widespread collection of data from all sorts of digital systems is mostly associated with the notion that it will help us do things smarter, better or more efficiently. The emphasis on the idea that various actors will use data knowingly and deliberately, seems to ignore the fact that data in and by itself can produce effects as well. That is, data, once it is collected, can and will have an impact simply because we cannot unknow what we have learned from it.

Our observations

- **Smart governments, e.g. smart cities**, strive to quantify the public domain in order to deliver higher quality services against lower costs. Their efforts will produce vast amounts of data and possibly some unexpected insights. One could imagine that such insights have profound political implications, for instance, when they uncover corruption or extreme levels of water or air pollution.

- Self-driving cars will be collecting massive amounts of data regarding their own activity, but they will also measure and film their surroundings, including other vehicles, pedestrians and other subjects. It seems inevitable that this data, once it is available, will eventually be used for surveillance and other purposes beyond its intended use.

- Pre-crime, elaborate statistical models that predict when and where various crimes may be committed, has a positive effect on overall crime rates, e.g. in Chicago. In practice, however, this means that local police focus on specific neighborhoods, which may lead to skewed crime statistics of those areas and, indirectly, to racial profiling.

- Researchers have demonstrated how old data from the Bitcoin’s public ledger can be combined with additional (public) datasets to trace people involved in drug deals or other (illegal) transactions that were once thought anonymous.

- Bruno Latour, among other thinkers, ascribes agency to non-human entities; things do things. Such entities, or actants in Latour’s words, are part of larger networks of humans (actors) and non-humans who together form the structure in which agency takes place. To illustrate, a fence forces people to make a detour (or to climb or cut the fence) and air quality measurements trigger people to consider the effects of air pollution and think about preventive action.
Data can be understood as a resource, somewhat similar to natural resources such as oil. Thus, it is no wonder that data is treated as the single-most important input in today’s economy. However, the dominant perspective on data is overly simplistic. It seems to assume that we have full control over the kind of data we collect, how it is analyzed and how subsequent outcomes can be interpreted and used. Two factors call for a more reflexive perspective on data. First, aside from oil and other material resources, the process from extraction to final product is quite predictable and manageable. Data, by contrast, is much more complex since the process of collecting and processing data is crucial to the kinds of output that are produced and these may include rather unexpected results. In the near future, AI systems are bound to generate even more unexpected and unplanned insights, i.e. correlations between seemingly unrelated data sets.

Second, for better or worse, once the new collected insights are “out there”, they will start to lead a life of their own and have an impact on everyday life beyond the original (human) intentions of collecting the data. To illustrate, data from self-driving cars will most likely lead to an increase in surveillance and data collection from garbage collection may originally be used to streamline processes of recycling, but might lead to monitoring consumption patterns of individual households.

The Dutch expression *meten is weten* (i.e. to measure is to know) implicitly underpins current efforts to quantify every dimension of society, business and everyday life. In that vein, data is presented as an objective set of information on the basis of which smarter decisions can be made. However, this perspective ignores the fact that many choices are made in terms of what is measured, how it is measured and how raw data is analyzed. This means that data only represents a limited cross-section of what it is supposed to measure and presents matters from a specific angle. Nevertheless, data will be overwhelming and outpower other (non-quantified) insights or opinions and push for ever more technocratic decision-making. To be sure, this is not necessarily a bad development, data may help societies disrupt existing power structures and facilitate progress, but the question is whether society will be able to control such data and its impact in a meaningful manner. Beyond the level of specific practices or concrete political decisions, the rise of data will shift our perspective on life and the societies we live in to a macro-level. The same happened in the past when new tools allowed us to measure natural phenomena more precisely, make more elaborate analyses and communicate findings more easily and rapidly. Eventually, societies will develop new data-driven worldviews similar to those tools of past inspired modernism.

### Implications

- The old saying that *knowledge is power* assumes that someone possesses exclusive knowledge and uses that knowledge to attain some goal. Public knowledge also exerts power, cf. old-fashioned propaganda and digital fake news, and control over data (what is made public, how is it framed and presented?) will be ever more important in any power struggle.

- Similar to ideas about “Privacy by Design” (i.e. designing digital systems so that no sensitive data is ever collected), a broader principle for data collection may be that only the most basic data necessary is collected for a specific purpose. This would exclude efforts to collect as much data as possible, from some system, without prior plans for concrete analyses.

- Databases around the world contain far more raw, unmined, data than data that is actually processed. In other words, there’s an enormous pool of as yet meaningless data that may produce concrete insights in the future. This may bring solutions to societal problems, but it may also have repercussions on a personal level; from tax fraud to drug deals on the dark web.