Big Data, Dataism and Measurement

Dario Petri

Currently, size of data is celebrated while only their usefulness in decision-making is actually important. Principles, methodologies and techniques of metrology are crucial to ensure proper and effective collection and use of data. Unfortunately, the fundamentals of metrology are seldom offered in educational degree programs, including those in scientific and technological subjects. Consequently, we should not be surprised if dataism (i.e., unconditioned belief in data) is spreading and a tsunami of improperly used data is submerging us. In the age in which measurements permeate all areas and every level of society with the expectation of promoting inconceivable socio-economic progress, underestimating the relevance of metrology is likely to lead to many serious potentially negative consequences.

Measurement and Big Data

Measurement has fostered the evolution of human society, civilization and quality of life. Since the beginning of human history measurements have been an essential tool for trading, building, and artifact production. In the last centuries, measurement-based knowledge has been crucial for geographical discoveries and for the development of modern science, industrialization, and medicine. More generally, it has been a fundamental background for knowledge advancement and socio-economic progress [1].

In the current information society, measurements permeate all areas and every level of society. The growing opportunities to collect data at lower and lower costs offered by modern technologies encourage amassing big amounts of data under the undiscussed belief that data will ensure inconceivable progress and a deep transformation of our ways of working, living and thinking [2], [3]. The ambition is to replace judgment based on personal experience and talent with processing of information provided by ubiquitous sensors or collected through the internet. As a result, astonishing quickness and efficiency improvements are expected. Moreover, the achieved conclusions are perceived as fair and reliable because they are obtained using formal procedures and objective data instead of being the outcome of subjective judgments possibly biased by personal opinions or self-interest.

Regrettably, while according to many society observers the classic ideologies that provide foundation to modern societies (like liberalism and socialism) are weakening, a new ideology is quickly spreading, without most people being aware: *dataism* [4]. In its extreme form, the *dataist*:

- perceives the entire world as a flow of data;
- believes that data provide a fair and exhaustive representation of reality;
- has unconditioned confidence in data and bases his/her everyday judgments only on data;
- believes that artificial intelligence will overcome human intellect; and
- advocates the concept of cosmic data processing and sees living organisms as biochemical information processing systems.

Dataism is quickly spreading throughout the scientific community, too. Indeed, academic disciplines tend to be walled off from one another. Also, there is often a significant gap between academic research and real-world needs. Since *data science* seems to have the potential to break down the barriers between the different disciplines and to contribute to fill the gap between theory and practice, it is expected to become a single overarching and unifying language able to investigate and to explain the empirical world, thus promoting inter-domain extension of insights and results. According to this ideology, "Beethoven's Fifth Symphony, a stock-exchange bubble and the flu virus are just three patterns of dataflow that can be analyzed using the same basic concepts and tools" [5]. This perspective is extremely exciting: data science appears as the scientific Holy Grail that has eluded scientists for centuries.

However, confidence of data driven decision-making does not depend on the amount of data, but on their significance and usefulness for the problem at hand. Thus, we do not need "big data" but *smarter exploitation* of *useful data*.

Dark Side of Data

Unfortunately, while considering the amount of data is quite easy, recognizing data usefulness cannot be so simple and immediate. In fact, both a proper cultural background and a

1094-6969/20/$25.00©2020IEEE

Authorized licensed use limited to: IEEE Volunteer. Downloaded on August 20,2020 at 16:22:34 UTC from IEEE Xplore. Restrictions apply.

Size of data is celebrated while only their usefulness in decision-making is of value.

Principles, methodologies and techniques of metrology are crucial to ensure proper and effective use of data in decision-making.

Spreading of dataism (i.e., unconditioned belief on data) is boosted by lack of metrology culture.

suitable knowledge about the underlying problem are needed. Speaking the sharp language of the *Star Wars* movie saga, considering only the amount of data lead us to fall into the *dark side*, that is, a seemingly easy and seductive way to reach the desired objective, but also an extremely dangerous and potentially very expensive choice.

Making important and complex decisions based on data that are not critically analyzed may lead to catastrophic consequences. Let us consider, for instance, the catastrophic effect of an improper evaluation of in-flight sensor data reliability in the recent Boeing 737 Max disaster [6], [7]. Also, the assessment of research quality based only on bibliometric indicators is not devoid of possible serious consequences, even if less tragic and dreadful. Interested readers can find many emblematic situations in which the uncritical use of indicators lead to absurd and detrimental decisions in references [8] and [9].

Flaws of Uncritical Data Usage

Even if motivated by the best of intentions, a bad use of data might induce many flaws and dysfunctions that often leads to deceptive conclusions, especially when complex problems have to be addressed. Among the many flaws due to uncritical use of data, some of the most common are the following ones [9]:

Measurability Bias: due to a tendency to prefer the most easily measurable factors rather than the most relevant ones for the problem considered;

Goal Displacement: people's attention and efforts tend to be focused on measured factors, often at the expense of other, possibly more important, unmeasured factors;

Promotion of Short Termism: long-term effects are difficult, or even impossible, to be fostered from data because they depend on something that is still unknown. Consequently, data might promote short-term goals, while stifling radical innovation and creativity, valuable qualities in most settings; risk taking and long-term investments are also discouraged, possibly resulting in stagnation and work gratification decline;

Manipulation of Indicators: through a variety of improper practices, such as: avoiding situations that might have negative impact on performance (practice known as creaming), failing to report negative instances (omission), lowering standards in order to improve scores, and even altering data to fabricate false evidence (cheating);

Degradation of Cooperation and Common Purpose: data tend to promote competition rather than cooperation and common purpose, which are both based on unmeasurable—or hardly measurable—motivations.

It is also worth noticing the potential highly negative impact that wrongly used measurements might have in sensible sectors, especially when socially relevant purposes are involved, such as education or healthcare [8], [9].

Proper and Effective Data Usage

To ensure proper and effective use of data we need to master the principles, methodologies and techniques that enable:

- the discrimination between *what is relevant* for the addressed problem from what can be neglected, remembering that not everything that is important is measurable, and conversely, what can be measured is not always what is worth measuring (or, using the words of a familiar saying, "not everything that can be counted counts, and not everything that counts can be counted");
- the critical *interpretation of information acquired* with measurement, by recognizing the limits of the models to which data (explicitly or implicitly) refer; indeed, models always provide a partial description of a portion of reality that depends on the intended purposes, prejudices and knowledge limits of model designers;
- the identification of all *uncertainty sources* that affect measurements significantly, in order to evaluate trustworthiness of available information, its impact on the derived conclusions and, eventually, of the risk of wrong decision-making.

It is thus clear that, except in the simple case of single aimed, fully structured, and fully informed decision-making activities [10], generally *measurements should not replace expert judgment*. Indeed, serious problems might arise when data are uncritically used in *automatic decision criteria*. Conversely, measurements should *support expert judgment*, which needs to properly take into account the relevance of data with respect to what has not been measured. Moreover, *measurement demands judgment*: judgment about whether to measure, what and how to measure, how to assess the significance of the obtained results, and how to properly use them [9].

Principles of Core Metrology

The principles, methodologies and techniques mentioned above are at the core of *metrology*, the science of measurement and its application [10]–[12]. Unlike the dataist, the *metrologist*, leverages on the metrology body of knowledge, and:

Authorized licensed use limited to: IEEE Volunteer. Downloaded on August 20,2020 at 16:22:34 UTC from IEEE Xplore. Restrictions apply.

- is fully aware of the relevance of data for decision-making, but he/she also knows the limits of the information data provide and that data-based conclusions are not granted to be flawless;
- is capable of quantifying the confidence level that can be assigned to data and related models;
- grounds his/her conclusions not only on data, but also on experience, *a priori* knowledge, critical analysis capability, intuition, and attention to possible cognitive biases;
- is fully aware of the limits of technology he/she uses to solve practical problems and enhance (not replace) human intellect capabilities.

Thus, the basic principles and concepts of metrology should represent common knowledge and must be well known to people involved in critical decision-making processes. Unfortunately, the fundamentals of metrology are seldom included in both undergraduate and graduate degree programs, including those in the scientific and technological fields. Therefore, we should not be surprised if a tsunami of improperly used data is submerging us.

We may wonder why the scientific and educational communities have such low consideration for metrology despite its crucial relevance. Perhaps a motivation can be found in Popper's demarcation problem [13], according to which the scientific disciplines are grouped in two great classes: the empirical ones—aimed at producing theories that explain how the empirical world works—and the formal ones—that are intended to derive formal theories starting from a limited set of axioms. Metrology—whose body of knowledge consists of an organized set of concepts, principles and methods aimed at building bridges between empirical and formal disciplines—is not considered in this classification. This is probably the reason why many scholars do not acknowledge that metrology possesses the dignity of an autonomous discipline and consider it only a set of techniques and instruments for the acquisition of empirical data.

In the age in which measurements pervade all areas and every level of society—from relations between States and large corporations to the simplest daily actions—underestimating the relevance of metrology is likely to have serious and often irreversible consequences.

Author's Note

A first version of this paper (written in Italian) has been published in *Tutto_Misure*, a publication of the Italian Association *Gruppo Misure Elettriche ed Elettroniche*.

References

[1] W. Kula, *Measures and Men*. Princeton, NJ, USA: Princeton University Press, 1986.

[2] V. Mayer-Schönberger, K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London, UK: John Murray, 2014.

[3] I. Goldin and C. Kutarna, *Age of Discovery*. London, UK: Bloomsbury, 2017.

[4] Y. N. Harari, *Homo Deus: A Brief History of Tomorrow*. London, UK: Vintage, Penguin Random House, 2016.

[5] Y. N. Harari, "Big data, Google and the end of free will." [Online]. Available: https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5dd5a28c.

[6] G. Travis, "How the Boeing 737 Max disaster looks to a software developer," *IEEE Spectrum*, Apr. 2019. [Online]. Available: https://spectrum.ieee.org/aerospace/aviation/how-the-boeing-737-max-disaster-looks-to-a-software-developer.

[7] A. Ferrero, "Imputato software: assolto," Tutto Misure, (in Italian), no. 1, pp. 43-44, 2019.

[8] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY, USA: Broadway Books, Penguin Random House, 2016.

[9] J. Z. Muller, *The Tyranny of Metrics*. Princeton, NJ, USA: Princeton University Press, 2019.

[10] L. Mari and D. Petri, "The metrological culture in the context of big data: managing data-driven decision confidence," *IEEE Instrum. Meas. Mag*., pp. 4-20, 2017.

[11] "Data Science Research Area, Intelligent and Effective Use of Data," National Physical Laboratory. [Online]. Available: https://www.npl.co.uk/data-science.

[12] "Big Data at NIST," NIST Big Data, National Institute of Standards and Technology. [Online]. Available: https://bigdatawg.nist.gov.

[13] L. Mari and D. Petri, "Measurement science: constructing bridges between reality and knowledge," *IEEE Instrum. Meas. Mag*., vol. 17, no. 6, pp. 6-11, 2014.

Dario Petri is a Full Professor in Measurement Science and Electronic Instrumentation at the Department of Industrial Engineering of the University of Trento, Italy. He received the M.Sc. and Ph.D. degrees in electronics engineering from the University of Padua, Italy, in 1986 and 1990, respectively. He is an IEEE Fellow member and the recipient of the 2020 IEEE Joseph F. Keithley Award for "contributions to measurement fundamentals and signal processing techniques in instrumentation and measurement." Dr. Petri's research activities are focused on digital signal processing applied to measurement problems, data acquisition systems, and fundamentals of measurement theory.

Authorized licensed use limited to: IEEE Volunteer. Downloaded on August 20,2020 at 16:22:34 UTC from IEEE Xplore. Restrictions apply.