preprint of a paper published in:
IEEE Instrumentation and Measurement Magazine, 20, 5, 4-20, 2017, doi:10.1109/MIM.2017.8036688

The Metrological Culture in the Context of Big Data: Managing Data-Driven Decision Confidence

Luca Mari and Dario Petri

In recent years, data sets are steadily growing in size because of the huge sensing, storage, computation, and transmission capabilities offered at very low cost by current Information Technology systems, which are increasingly driving ubiquitous data acquisition from the empirical world. Moreover, massive amounts of data can be collected very efficiently – sometimes at zero marginal costs – from the internet. The concept of *Big Data* (BD) epitomizes this phenomenon, among the experts and the general public (for an accessible, application-oriented introduction to BD in the broad context of data science, see, e.g., [1]). "Big Data" is in fact a vague, all-encompassing term that broadly refers to any data set, or collection of them, so large, usually heterogeneous (e.g., text, audio, images, video) and poorly structured that it becomes difficult to effectively obtain information from them by using traditional computational methods. The societal relevance of this phenomenon is related to the expectation that, in the near future, BD will positively and pervasively affect our lives, thus promoting prosperity and social evolution.

A main driver of the BD phenomenon is the importance attributed to the role of data in supporting *Decision Making* (DM). Of course, while a decision can be made independently of information available on the entity on which the decision is to be made, there is nothing new in the idea that DM takes data on such an entity into account, what is called *Data-Driven Decision Making* (D^3M). All closed-loop, negative feedback control systems are examples of D^3M, where data acquired from the field by sensors provides the information for system regulation through actuators. In this traditional situation, which here is called *weak-D^3M*, data is interpreted and processed by *a priori* established models. In the BD context, a possible novelty is the assumption that the huge amount of data can compensate for the lack of pre-existing interpretive models, up to the point that "data speaks by itself" so that decision models can be inductively built from data. In this scenario, called here *strong-D^3M*, BD promises to enable new strategies of making effective decisions in the empirical world. In the most radical vision, the novelty is not only in automatically generated models, but also in the fact that the justification of the predictive capabilities of such models entirely lies in data itself, a perspective that quite emphatically has been called "the end of theory" [2].

The distinction between weak-D^3M and strong-D^3M is not sharp, mainly because for data to produce decision models some meta-models are required (such as k-means, hierarchical clustering, support vector machines, and neural networks, which are customary techniques in data mining and machine learning), and the distinction between models and meta-models is not so sharp as a result. In any case, be the BD phenomenon a forthcoming revolution or not, weak- or strong-D^3M is going to permeate more and more our society, as the concept of "world datafication" witnesses [3].

D^3M in the Context of Metrology

In D^3M, data must somehow refer to the object of the decision, i.e., data must be data-on-something, for it to be useful. The distinction between *data-per-se* and *data-on-something* is not new. In his introductory chapter to the Shannon's seminal book [4], Weaver wrote about it in terms of "the technical problem" and "the semantic problem," respectively. In this view, data (i.e., data-per-se) is simply anything that is chosen from a predefined set that includes at least two elements, functionally identified as symbols. While some technical problems can be tackled on the basis of data-per-se, a decision expected to involve something that is not data itself must be driven by data somehow referring to the object of the decision, and then equipped with a truth value or an indicator of data quality. This is the realm of Weaver's "semantic problems," where the reference of data to something else is an aspect of that complex entity that is *meaning*. Data-on-something, and therefore data equipped with meaning, can be called *information* (see, e.g., [5]).

Of course, D^3M is possible without focusing on the distinction between data and information, if the provision of meanings to data, and the related constraints on data processing procedures, is an implicit task for human beings. For example, in a purely syntactical context one might both sum and multiply numerical data to find possible regularities in the outcomes, whereas sums of data known to refer, e.g., to masses and accelerations should be considered with suspicion, if not taken just as wrong. Conversely, in the case of strong-D^3M, with the perspective of completely computer-based decisions, the explicit distinction between data and information is essential.

A crucial point is that in D^3M the confidence in the outcomes, about the risk of making wrong decisions, depends on the quality of the available information. This is sometimes presented in terms of the so called Garbage-In-Garbage-Out (GIGO) principle: poor information quality can be the main reason of wrong decisions. Guaranteeing a sufficient quality of input information is then a critical condition.

For centuries, science and technology have developed and refined a body of knowledge aimed at characterizing the quality of information acquired from the empirical world: it is *metrology*, the "science of measurement and its application" [6]. Indeed, measurement is a process able to provide information about the quantity intended to be measured, i.e., the *measurand*, by comparing it with a reference (usually a unit). That information is returned in the form of *quantity values* for the measurand, coupled with a specification of the quality of such values, sometimes in terms of a single synthesizing parameter called *measurement uncertainty*, as shown in the block diagram of Fig.1 where the paths of both information and its quality are depicted as outputs of the measurement process.

Insert Fig. 1 here

We claim that the *metrological culture*, and in particular the capability to identify and quantify the different contributions that limit the information quality, *is crucial to assess the confidence of BD-based D^3M outcomes*.

- The actual driver of D^3M is information, not just data.
- Effective strong-D^3M requires the solution of semantic problems, not only syntactical ones.
- The metrological culture is crucial to estimate the confidence of the D^3M outcomes.

D3M analyzed in the context of metrology.

Decision Making and Big Data

DM is an information process, not a physical transformation, performed by an entity (the *decision maker*) that has identified at least two alternative options in reference to an entity (the *object of the decision*) and chooses one of them (the *decision)* as the outcome of the process. DM implies the availability of information on the object of the decision – at least on the possible options that can be decided but usually also on the current state of the object – and therefore, if the information is not already owned by the decision maker, it may require a stage of information gathering. This is D^3M.

As shown in Fig. 2, DM is guided by a possibly specified *procedure* in order to achieve a given *goal*. It is subjected to *constraints*, for example related to the resources that can be devoted to the process or to the limits in the time available to decide. In addition, an idea of the acceptable *level of risk* associated with making the decision is needed, being it a lower bound for *decision confidence*: while the acceptable level of risk is a process specification, the decision confidence depends on both the DM structure and the quality of the available information.

Insert Fig. 2 here

Main Features of Decision Making

The structurally simplest case of DM is such that:

- there exists a single best choice, implied by the well-defined purposes of the decision maker and the state of the object of the decision (*single aimed* process);
- the decision maker is completely guided toward the best choice by well-defined procedures to operate on the available information (*fully structured* process); and
- the information obtained from data suffices to make an unambiguous decision (*fully informed* process).

In this case, DM is not a matter of subjectivity and can be, in principle, completely automated. Conformity assessment may be an example of such D^3M, being based on (1) explicit conditions on the best choice (whether the item under test should be decided being in conformity with specifications of nominal values and tolerances); (2) explicit decision rules stating when the item is in conformity with specifications, or when the available information is not sufficient to discriminate (see, e.g., [7]); and usually (3) appropriate measurement processes providing the required information on the item under test.

Up to few years ago, only single aimed, fully structured, and fully informed DM could be completely automated. Today, computers not only support human decision makers in performing a lot of background work, but they have acquired the power to perform many non-single aimed, semi-structured, and partially-informed decisions. Indeed, with the advent of BD and the diffusion of the so called “cognitive computing” [8], computers can be endowed with the ability to automatically discover robust patterns from a huge number of examples, thus inferring information from them.

This scalability in DM can have strong social implications, since it is expected to exhibit strategic value for businesses as well as for knowledge development in many data-starved areas of inquiry such as health care, and social, ecological, and earth sciences. This is probably the reason why combining human knowledge with machine learning appears to be a very promising approach [9].

Decision Making in the Context of Big Data

Historically, DM in science and technology has been based on single aimed, structured, and fully informed processes, where the required information is obtained by processing data about properties of empirical objects obtained by means of specifically designed measurements, and processing procedures are based on predefined analytical models derived from well-established theories. This characterizes weak-D^3M.

In the last decades, the effectiveness of the scientific method inspired the application of D^3M in many fields of human endeavor, with the aim of minimizing the risk of wrong decisions and improving the comparability of the outcomes. However, unlike the traditional contexts of science and technology, when D^3M is performed in the context of BD (see, e.g., [10]):

- data is not acquired by well-structured and validated procedures designed for a well-defined purpose, but its availability is usually just taken for granted; and
- processing models are not based on well established theories on the phenomenon at stake, but are derived from intuition or experience, or result from observed regularities in the available data.

Thus, both data and model uncertainties are rarely reliably known or evaluable with good confidence. This situation may be expected to become more and more common with the further increase of world datafication, where the widespread diffusion of both data acquisition (from the empirical world), and information gathering (from human beings or pre-existing data sources) will make huge data sets available, generated not with a specific purpose but as generic, possibly open, "data services," in the perspective that sooner or later they might be useful somewhere to something. This is why in the BD context data is often assumed as already available, so that designing and implementing suitable acquisition procedures might not be needed, or are sometimes even possible. As a result, data quality is usually unknown except for some purely syntactical aspects such as missing data and for checking general constraints related to data types (a string is wrong where a number is expected) and ranges.

Moreover, when dealing with BD, explanatory models for data are often *a priori* unknown, and they must be extracted from data itself by adopting blind analysis methods, such as the meta-models of machine learning, which are indeed aimed at identifying patterns in the available data with no or very limited prior knowledge on the structure of the data generating phenomenon. The robustness of the identified patterns is then particularly important: a candidate pattern is expected to occur in the future so that the related information has predictive value and consequently is reliably actionable for DM [9].

Hence, it is not surprising that in the context of BD, the risk of wrong data processing, wrong data semantization, and then wrong decisions may be significant. Here, metrological principles and methods may become critical tools to (at least roughly) estimate how uncertainty of information, adopted models and procedures affect the confidence associated to the drawn conclusions.

The Metrological Culture in the Big Data Context

While the specific tools of metrology might not be always applicable in the context of BD, some lessons can be learned from it to support a methodologically correct implementation of strong-D^3M, as illustrated in this section.

Measurement and Quality of Information

According to the mentioned GIGO principle, poor information quality can be the main reason for wrong decisions. Unfortunately, the impressive proliferation of data sources and the exponential growth in data volumes that characterize BD can make it hard to assess, and even harder to assure, the quality of the available information. Hence, it is clear why *data quality management* is of crucial relevance in D^3M, and why a significant fraction of the time and budget available for BD system development is currently employed to tackle quality issues.

The dimensions of information quality include:

1. *consistency:* the condition that data is within the assumed value domain and is not duplicated;
2. *availability*: the fraction of time that data is made available by the system that stores it;
3. *currency* and *timeliness*: the degree to which data is updated and readily available for use, respectively;
4. *specificity*: a condition related to the quantity of syntactical information: stating, e.g., that a length is in the interval 10.5 ± 0.1 m is more specific, and therefore of better quality, than stating that it is in the interval 10 ±1 m; when referring to measurement results, specificity is also called *precision*; and
5. *trueness:* a condition related to the faithfulness of semantical information: were, e.g., 10.55 m the value of a length provided by the best independent method, stating that the length is 10.50 m is truer, and therefore of better quality, than stating that it is 10.40 m [6] as synthesized in terms of *accuracy* in metrology [6], [11].

The first four dimensions are related to data-per-se (where the second and the third dimensions specifically refer to data management) and are the focus of data quality management systems, mainly aimed at the detection and possibly the retrieval of erroneous or missing data in a scalable and timely manner. The fifth dimension is related to the meaning of data and, together with the fourth dimension, can be effectively managed using the principles, methods, and tools of metrology.

In this perspective, metrology can be structurally considered a *science of information quality*, since it enables one to evaluate and express the quality of information obtained by experimental means. The knowledge of metrology fundamentals is then helpful to manage decision confidence. In particular, metrology suggests that information quality is inversely related to the *uncertainty* attributed to information [12]. Indeed, while measurement can be abstractly interpreted as a selection process (a set of possible symbols is given, and the process leads to select one or more of them), it is much more [13].

First of all, a model is adopted for the measurand [6] such that the selected symbols are interpreted as values of that quantity, thus becoming data-on-something. In defining such a model, it may be admitted that the measurand cannot be completely identified and characterized, so acknowledging a non-null *definitional uncertainty*, “the practical minimum measurement uncertainty achievable in any measurement of a given measurand” [6].

Second, referring the reported values to the intended quantity is definitely not a trivial position, and it is justified by the quality of the measuring system, which is expected to generate an output that is stable in the case of repeated interactions with the object under measurement, and specifically depends on the measurand and not on other quantities, the so-called “influence quantities.” The fact that the measuring system has a limited repeatability and selectivity with respect to the measurand is acknowledged in terms of a non-null *instrumental uncertainty*.

Third, the interaction between the object under consideration and the adopted measurement system can alter the state of the object itself. This phenomenon, acknowledged in terms of a non-null *interaction uncertainty*, may occur when acquiring physical quantities – the so-called “loading effect” – and it is even more usual for non-physical quantities, as for example in most cases of interviews (see, e.g., [14]).

Non-null definitional, instrumental, and interaction uncertainties prevent the complete object-relatedness (*objectivity* for short) of the information provided by measurement.

Fourth, the reported values imply a comparison with a reference quantity, usually a unit realized by a measurement standard, so that any numerical quantity value is intended as the result of such a comparison performed according to a well-defined procedure. To ensure that the information provided by measurement is then socially available and understandable, measurement standards must be disseminated, and measuring systems must be calibrated against them. Limited quality of the calibration process prevents the complete subject-independence (*intersubjectivity* for short) of the information obtained by measurement.

A fundamental principle of the metrological culture can be then characterized as follows, as synthesized in Fig. 3: data becomes useful information when it is related to an object and its quality is evaluated in terms of objectivity and intersubjectivity, as customarily synthesized by measurement uncertainty [15], [16].

Insert Fig. 3

- The degree of measurement information objectivity is quantified by means of definitional, instrumental, and interaction uncertainties.
- The degree of measurement information intersubjectivity is quantified by means of instrument calibration uncertainty.

Quantification of the quality of information returned by measurement.

Quantification of the Quality of Measurement Information

A simplified model that synthesizes the structure of the whole process of D^3M, as analyzed in the metrological context, is schematized by the block diagram in Fig. 4. Starting from the bottom of the diagram, three different stages are identified: *information generation*, *information processing*, and *decision making*, and the main sources of uncertainty that can affect the confidence of the final decision are highlighted. The identification of such uncertainty sources and the quantification of their effect on the decision confidence can be facilitated by the knowledge of measurement fundamentals.

Information Generation: As shown in Fig. 4, in the information generation stage, data is collected through two types of procedures: measurement (that is objective and intersubjective information acquisition from the empirical world) and information gathering from existing repositories or the internet. Along the two paths, data collection activities and phenomena limiting the information quality are very different.

In most *data acquisition* processes performed in the context of BD, unlike measurements, objectivity or intersubjectivity can hardly be assessed. Even when dealing with non-physical empirical properties, the identification of uncertainty sources and the evaluation of the related uncertainty is facilitated by considering the four different sources discussed above and shown in Fig. 3: *definitional, instrumental, interaction*, and *calibration* uncertainty sources. These uncertainty sources are present in any measurement, i.e., they are physiological, and the related uncertainty should be kept within predefined bounds. However, due to uncontrolled phenomena, the acquired *information might be wrong*. These pathological situations are often related to the first three quality dimensions above: they should be detected and the related data should be rejected, or, if they are believed to convey information, they should be explained.

Also, data gathered from the information world is, in principle, affected by uncertainty which is usually unknown. However, when subsidiary information about the procedure employed to the gathered data is available, uncertainty can sometimes be evaluated using metrology methods and tools. In particular, the uncertainty sources considered above are present also when gathering information. For example, instrumental uncertainty might be related to the adopted software tools, their correct design and use. Interaction uncertainty can arise when the employed software tools may alter the stored data. Moreover, uncertainty affects the *reliability* and degree of *coverage* of data stored in repositories.

Insert Fig. 4 here

The metrological approach of identifying the various uncertainty sources is useful also to analyze the stages of information processing and decision making, as discussed in the following. Moreover, once the various uncertainty sources have been identified, the uncertainty effect on the confidence of the process outcome can be estimated using the methods and tools of metrology [12].

Information Processing: As shown in Fig. 4, in the information processing stage, quality may be limited by:

- the uncertainty of the adopted inferential or predictive processing model *(inferential uncertainty*). For example, bad model quality can occur because the selected indicators provide insignificant or wrong information for the considered DM problem. If enough data is available, modeling uncertainty can be evaluated using data itself [17]; and
- the finite amount of information conveyed by the limited amount of available data (*statistical sampling uncertainty*). In the BD context, due to the large amount of data, estimates are expected to be accurate proxies of the statistical parameters of the population, thus significantly reducing sampling uncertainty.

Also in this stage, the above uncertainty sources are present, and the related uncertainties should be kept within predefined bounds. However, it may happen that the *inferential model* *is wrong*, which reflects in a misuse of the available information (e.g., statistical correlation misinterpreted as causation). The usage of blind analysis methods, relying only on data and without *a priori* information about the phenomenon at stake, may significantly increase the risk of these pathological situations.

Decision Making: In the DM stage, the confidence in the process outcome can be limited by various causes, such as the vagueness of the goal, the ambiguity in the possible decision options, an incomplete decision procedure, conflicts in constraints, and reasoning biases. These factors reflect on reduced decision confidence even though information unaffected by uncertainty is available. Also, the significance of information is critical for ensuring effective decision support, referring to the query: "is the provided information useful to support the DM process?".

Among the reasons limiting information significance there are:

- *construct significance*, related to the query: "are we actually considering what we need to consider to effectively support DM?" When dealing with complex problems whose determinants need to be described at a high level of abstraction, it may be difficult to recognize the significance of the provided information and even more difficult to improve its quality. For example, different intelligence tests exist, but the obtained results could not be useful to decide about problem solving capabilities of a person; and
- *content significance* or *completeness,* referring to the extent to which the provided information covers the whole range of relevant aspects needed for an adequate support to DM. For example, when deciding whether to hire a metrology technician, information limited to his/her technical capabilities only could lead to wrong decisions since information about his/her process management capabilities and soft skills is also relevant.

Information significance can be improved only through a better understanding of the main determinants involved in the decision problem.

Finally, as in the previous stages, pathological DM can occur. For example, the available information could be misused due to a wrong decision procedure.

Conclusions

A main driver of the BD phenomenon is the importance attributed to information in supporting DM, both when data is interpreted and processed by *a priori* interpretive models (weak-D^3M) or when "data speaks by itself," since blind analysis methods are used (strong-D^3M). In particular, the latter situation is expected to enable new strategies of making effective decisions in the empirical world, even though the lack of predefined interpretive models of the phenomenon at stake may strongly increase the risk of wrong decisions.

This paper aimed at showing that a sound knowledge of measurement fundamentals can be helpful in managing such a risk, since it makes people aware about possible uncertainty sources and their effect on the confidence of the conclusions they draw. Specifically, a simplified model that synthesizes the structure of a D^3M process has been proposed by identifying three main process stages: information generation, information processing, and decision making. Applying approaches used in the context of metrology, the main contributions that may affect the confidence of the final decision have been identified and highlighted in a D^3M model. We think that the obtained scheme can provide useful guidelines to estimate the confidence of the D^3M outcomes and manage the risk of making wrong decisions.

Remark: One of the authors is a member of the Joint Committee for Guides in Metrology (JCGM) Working Group 2 (VIM). The opinion expressed in this paper does not necessarily represent the view of this Working Group.

References

[1] R. Schutt and C. O'Neil, *Doing Data Science*. Newton, MA, USA: O'Reilly Media, 2014.

[2] C. Anderson, "The end of theory: the data deluge makes the scientific method obsolete," Wired, 2008. [Online]. Available: https://www.wired.com/2008/06/pb-theory.

[3] K. N. Cukier and V. Mayer-Schoenberger, "The rise of big data – how it's changing the way we think about the world," 2013. [Online]. Available: https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data.

[4] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication.* Champaign, IL, USA: University of Illinois Press, 1949.

[5] L. Floridi, *The Philosophy of Information.* Oxford, England, UK: Oxford University Press, 2011.

[6] JCGM 200:2012, International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM) (2008 edition with minor corrections), Joint Committee for Guides in Metrology, 2012. [Online]. Available: http://www.bipm.org/en/publications/guides/vim.html.

[7] JCGM 106:2012, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment, Joint Committee for Guides in Metrology, 2012. [Online]. Available: http://www.bipm.org/en/publications/guides/gum.html.

[8] J. O. Gutierrez-Garcia and E. López-Neri, "Cognitive computing: a brief survey and open research challenges," in *Proc. 3rd Intern. Conf. Ap. Computing and Inform. Technology and 2nd Intern. Conf. on Computational Sci. and Intelligence*, pp. 328-333, 2015.

[9] V. Dhar, "Data science and prediction," *Comm. ACM*, vol. 56, no. 12, pp. 64-73, 2013.

[10] D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining.* Cambridge, MA, USA: MIT Press, 2001.

[11] ISO 5725-1:1994, Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions, International Organization for Standardization, 1994.

[12] JCGM 100:2008, Evaluation of measurement data – guide to the expression of uncertainty in measurement (GUM) (1995 edition with minor corrections), Joint Committee for Guides in Metrology, 2008. [Online]. Available: http://www.bipm.org/en/publications/guides/gum.html.

[13] D. Petri, L. Mari, and P. Carbone, "A structured methodology for measurement development," *IEEE Trans. Instrum. Meas.,* vol. 64, no. 9, pp. 2367-2379, 2015.

[14] O. Duncan, *Notes on Social Measurement – Historical and Critical.* New York, NY, USA: Russell Sage Foundation, 1984.

[15] 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.

[16] L. Mari, P. Carbone, and D. Petri, "Measurement fundamentals: a pragmatic view," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2107-2115, 2012.

[17] M. Gubian and D. Petri, "Uncertainty analysis of learning-from-examples algorithms," in *Proc. of the IEEE Int. Workshop on Advanced Methods for Uncertainty Estimation in Meas.,* pp. 34-39, 2008.

Luca Mari is a Full Professor of measurement science with the Cattaneo University – LIUC, Castellanza, Italy, where he teaches courses on measurement science, statistical data analysis, and system theory. He is currently the Chairman of the TC1 (Terminology) and the Secretary of the TC25 (Quantities and units) of the International Electrotechnical Commission (IEC) and an IEC expert in the WG2 (VIM) of the Joint Committee for Guides in Metrology (JCGM). He has been the Chairman of the TC7 (Measurement Science) of the International Measurement Confederation (IMEKO).

Dario Petri is a Full Professor of measurement and electronic instrumentation and the Head of the Department of Industrial Engineering, University of Trento, Italy. He is also an IEEE fellow member and has been the VP for Finance of the IEEE Instrumentation and Measurement Society since 2013 and the Chair of the IEEE Smart Cities Initiative in Trento since 2015. He was the Chair of the Italian Association of Electrical and Electronic Measurements (GMEE) from 2013 to 2016 and the chair of the IEEE Italy Section from 2012 to 2014. He received the M.Sc. degree and the Ph.D. degree in electronics engineering from the University of Padua, Padua, Italy in 1986 and 1990, respectively. His research activities are focused on data acquisition systems, embedded systems, fundamentals of measurement theory, application of digital signal processing to measurement problems.

Fig.1. Conceptual view of measurement as a process that provides information about entities of the empirical world.

Fig.2. Black box model of a DM process.

Fig.3. Conceptual view of measurement as a process that provides objective and intersubjective information.

Fig.4. Simplified model for D^3M analyzed in the context of metrology.