ACTA IMEKO
ISSN: 2221-870X
December 2024, Volume 13, Number 4, 1 - 6

Big-Data handling as a metrological challenge beyond digitalisation: some reflections

Franco Pavese[1]

[1] *Independent Scientist, Torino, Italy*

ABSTRACT
Digitalisation is considered one of the new technical tools in metrology as well; it is a general strategic issue concerning the future treatment of human knowledge in progress. Digitalisation has reached a dominant status also in data science, inducing a collation of an amount of "data", i.e. of potential information, that is by now by far quantitatively enormously wider than ever before in the science history. It is called Big Data and it is widely recognized. However, it is also considered to be able to replace some of the traditional handling tools in the relevant scientific information, namely the metrological ones. It is called "dataism".
This short paper is proposing some reflections on Big Data handling as an additional challenge also for metrology, with suggestions on possible improvement. In fact, Its regulatory field is considered to need taking into account and providing guidelines to the new concepts.

Section: COMMENTS

Keywords: digitisation; digitalisation; Big Data; data scales; "dataism"; data quality; metrological guiding tools

Citation: , Big-Data handling as a metrological challenge beyond digitalisation: some reflections, Acta IMEKO, vol. 13 (2024) no. 4, pp. 1-6. DOI: 10.21014/actaimeko.v13i4.1880

Section Editor: Francesco Lamonaca, University of Calabria, Italy

Received May 17, 2024; **In final form** November 10, 2024; **Published** December 2024

Corresponding author: Franco Pavese, e-mail: frpavese@gmail.com

1. INTRODUCTION

Digitalisation is today's "mantra" of science, being one of the sources of the quick technological developments of the recent times. In metrology, digitalisation is considered one of the technical needs that, to date, have become necessary. There are several definitions/interpretations of digitalisation, from the action taken by simple digitising to its concept, consequences, and use. [1]-[7] In fact, it does not only mean a tool to handle data but additionally includes the need for developing other related topics, such as digital representation, ontology, and definitions.

The following definition, among many, will be regarded as adequate, appropriate and sufficient: *digitalisation is the process of converting information into a digital* (i.e., computer-readable) *format*. That is often indicated as the definition of "digitisation", [8] while "digitalisation" is assumed to rather refer to further advantages deriving from digitisation.

A most general strategic issue deriving as an important *consequence* of digitisation to digitalisation is considered in this Comment, involving the future treatment of human knowledge. It consists of the fact that digital means, having reached a dominant status also in data science, induced a collection of an amount of data, i.e. of potential information, which is by now enormously far wider in quantity than ever in human history.

It is common belief, in science as well, that greater the information, the higher is the reliability − at any given level of confidence. In that sense, the use of digital means is favoured also in science-positive developments.

These large datasets are today called Big Data, whose *definition* is extremely varied [9]-[12]. Here the focus will be limited to the fact that they also induced a new position, considering a different way of thinking about data, called "dataism" [13]-[16] – it was found difficult to find references in mainstream literature of philosophy of science and also of measurement science, apart in some social fields.

Since each datum is considered a piece of information, the use of an "accumulation" method might in principle be considered sufficient in itself. Such a position assumes that a set of data increased in quantity will also increase the amount of information, i.e. that more data necessarily equals more reliable knowledge, and even of better quality.

However, an increased amount of information not necessarily does always mean that the overall obtained information remains consistent with itself − so implying an effective univocal increase of knowledge, the necessary condition for reliable decisions.

In addition, it is also common sense in science that different positions almost always arise in the evaluation of the currently available knowledge. [17] These positions are initially all considered potentially valid for discussion, until the Community is able to converge onto a single position (sometimes after Centuries, such as the fact of Earth roundness), generally the prevalent one among them, i.e. based on *consensus*. Thus, only a "probabilistic" nature is assigned to most of the newer changes in the status of knowledge.

Moreover, scientists are accustomed to considering the information carried out by "data" the safest way to increase knowledge, with their add-ups allowing it to increase in quantity/quality. In other words, they may consider knowledge to become in that way more reliable or wider − without that meaning closer-to-truth as one does not know where the truth is [18].

In the case of Big Data, on the one hand, one can appreciate advantages deriving from the digital frame making possible the new advancement; on the other hand, one should also realise that possible *defects* associated with any new technology/frame of thinking are initially ignored until their evidence may arise.

However, while digitalisation is today already a matter of discussion and documents also in the metrological Community, it is still rare to find in it concern about the need for a targeted treatment/handling of Big Data: see Conclusions for a proposal concerning this issue.

The focus of this Comment, which is not intended to be a review, is neither to bring new light to the recent evolution of scientific data treatments. Its main aim is *to alert* about possible missing sufficient consideration, within the *metrological frame*, of some consequences that the new development of knowledge tools may induce, namely according to the dataist vision, today quite widespread in the literature (e.g. in [13]-[16] and their references).

The comments intend to analyse and stress the need to keep the basic current metrological data–handling rules typical of measurement science, namely those related to data uncertainty, also in Big Data "handling" obtained via digitalisation, i.e. intend to provide a stimulus to metrology and suggest its positive action toward new provisions and rules for also handling Big Data. Obviously, that would require some adjustments to match the new frame.

2. BIG DATA VS DIGITALISATION

Concerning digitalisation Big Data is a sub-topic while for science Big Data is instead the topic, today unavoidable.

In fact, the massive increase of available data provided by digitalisation is leading part of the Scientific Community to believe that an accelerated knowledge increase is now possible and feasible − including an increase in its quality. [19]

A specific field where the impact of digital means is of special importance, apart from the economic ones outside of interest here, is the general scientific aim of *prediction*, e.g. by *extrapolation*. Most often it concerns the observed trend of a set of human/instrumental observations. Data from observations form the necessary bases and constraints for anyone trying to predict a future trend, i.e., to accurately anticipate future observations. There is a vast amount of literature about the new requirements and tools for such an expectation to become more reliable and credible: some examples concerning data "uncertainty" consideration and treatment can be found in [20]-[30]: notice an almost lack among them of authors from core Countries/Journals. Also notice, as a real fact the lack of a basic tool in metrological analysis, the "uncertainty budget" (UB).

In the new rapidly growing tendency related to Big Data called "dataism" data are considered as *evidence and objective* to such a confidence level to be assumed to intrinsically have the dominant importance. The extremely large number of them now available is thought to compensate for any "traditional" check of their quality, consistency, uncertainty, etc.. According to that position Big Data does not require further "handling", e.g. *scientific analyses*, but data can simply be treated with simple mathematical means for dataset analyses. That excludes, e.g., the need to analyse also the influence of *systematic effects*, which depends instead on the *process* − see later).

In such a new frame, digitalisation may look to be basically limited to providing tools for a digital representation of metrological rules and symbols (e.g., *by incorporating the units as a data attribute* [31]-[33]).

3. BIG DATA FROM THE METROLOGICAL VIEWPOINT

This Comment is basically referring to *numerical* Big Data representing the results of measurements, though the reflections are also valid for all the *types of database* (e.g., qualitative, see later in this Section). The fact that the dataset is much larger than in the past does not change data nature (see Section 3.2) and the consequent need to treat them according to the rules of the measurement science frame, i.e. overlapping metrology, *at least for a very large portion of its total contents*.

3.1. *Repeated data*

Basically, Big Data can be considered as *repeated* (or *replicated*, see [34]) measurements. That definition is more complex than it may appear to be − and will also reveal limitations about the Big Data usefulness. In fact, the recognised need for repeated measurements has to be confronted with two basic different situations, depending on whether the measurand is *stable* (during a repetition time period) or it shows a *trend*.

For a *single* series of data:

(A) Data in a series with a measurand *stable in time*. They are "repeated" in the strict sense, though never perfectly − identical values could occur only if the "sensitivity" of the measurement is too low, or if we consider the values "perfectly repeated" when their dispersion is within the measurement uncertainty. The dispersion of the data values is the common situation, without trend in this case, so basically random and evaluated statistically. Obviously, the dispersion of the values does not provide an estimate of the overall uncertainty, since only a full analysis of the *process generating the data* and of its influence factors can evaluate the systematic effects affecting the repeated data.

(B) Data series of a measurand showing a *trend in time*. It means that the measurand in itself is the (main) reason for the observed trend of the data (and of the possible correlation among them). This new occurrence does *never* exclude also the previous random components: it means that a mathematical function (deterministic in nature) cannot fully model the process. It may even happen that the uncertainty of the data is such that it is not possible to infer a sufficiently univocal trend. In addition, again the systematic effects are not evaluated, but only the quality of the fit, its reproducibility, and not the accuracy.

Also, the *time scale* of the data acquisition is an important factor, as it has to be related to the "time constant" of the process. For repeated series:

(A) More data are collected in a short time period with respect to the total time scale of full collation relevant to the measured phenomenon. In this case, it is more probable that all influence factors will remain sufficiently stable with respect to the typical time constant of variation of the phenomenon. In such a situation, the data-acquisition repetition brings significant advantages only if largely exceeding that constant: big data may basically look *useless*.

(B) More data are collected and distributed during (almost) the full-time period of interest. In this case, in first approximation the observed trend is providing the most important information: however, its evaluation does not normally require an exceedingly large number of samplings, while details already come from a short-time sampling. Also, in this case, the advantage of big data may be limited.

In conclusion, not always a higher *data density* does ensure better information in the short term while for an extended period of observation some advantages can arise.

What is much more important is the *repetition of the series* over a (long) period of time. Series repetitions may improve the repeatability of the mean value of a *stable* measurand, while in the case of a *trend* they allow to estimate it for longer periods. In both cases, without a specific analysis of the full *process*, that type of analysis of the data does *not allow to assign accuracy* to the set of observations.

3.2. *Data types*

Another metrological issue that looks basically absent from the interest of dataism is the fact that *not all data are numerical*, and *not all scales* in which they are traditionally ordered are *quantitative*, according to their usual classification in several scale types, shown in Figure 1.

Scales would need to be introduced also concerning Big Data analysis to show the different types involved in science, as also Big Data handling depends on them. All scale features can be adopted by digitalised means, concerning all types of Big Data – this fact already places in itself limitations to some aspects of the "dataist" vision.

Any handling and analysis of Big Data should start just from the differences in properties between the different types of data-representing scales. Instead, even for the *numerical* categories, Big Data handling seems not always aware of the fact that there is *more than one* numerical type. On the other hand, the *qualitative types* are often the main kind of information that is currently available in non-scientific fields also covered by Big Data, like market statistics or analyses of customers' preferences for commercial products.

Incidentally, there are no reasons to exclude from Big Data treatment the use of their representation of the geometrical types (like already done, e.g., for the "spatial Big Data", with the concurrent use of maps), [35], [36] a frame whose advantages are generally not yet enough taken in due consideration.

It is not the scope of this short analysis of Big Data to go further into deeper detail. However, it seems useful to also recall that their *volatility*, i.e., the rate of change and lifetime of the data, is basically limited to data (typically of *nominal scales*) related, e.g., to social media analyses, where sentiments and trending topics change quickly and often. Still, diverse information that arrives in increasing volumes and with ever-higher velocity in Big Data can be either structured (often numeric, they are easily formatted and stored) or unstructured (more free-form, less quantifiable).

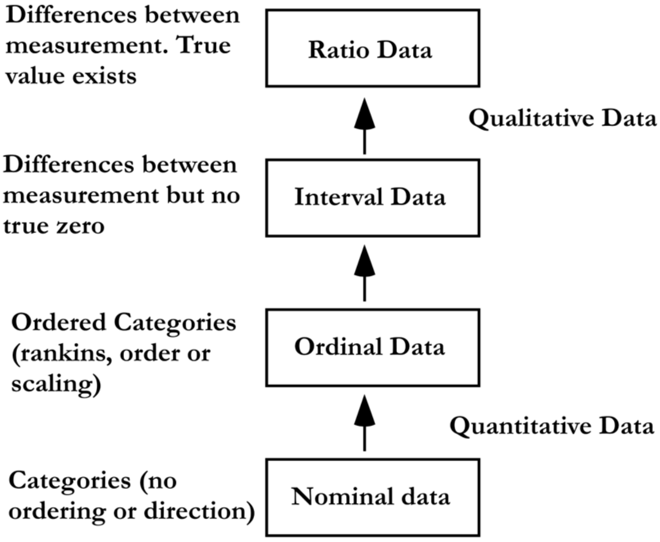

Figure 1. Data scale types.

3.3. *Reproducibility vs accuracy of a dataset*

The Author's interest in Big Data mainly derived from his interest to analyse several large databases on Earth's parameters, and to look at their associated *uncertainty* treatment, including the resulting UB. He found significant differences in the adopted procedures with respect to the ones commonly used in metrology and a full lack of the UB. This issue is treated in Section 4.

4. A FURTHER ROLE ALSO FOR METROLOGY?

A recent Comment appeared in the Journal Nature [37], concerning the need for the metrological Community to deal with digitalisation, recalling the fact that "CODATA (Committee on Data of the International Science Council) formed the Task Group on Digital Representation of Units of Measurement (DRUM)".

Similarly, dataism, already active in scientific literature, represents *another challenge to metrology*, since it is assumed to disregard the part of measurement science concerning the scientific analysis of the obtained data, by assuming instead that sufficient/full evidence intrinsically arises from the data in themselves [13]-[16]. This attitude looks reducing metrology to a kind of bureaucratic or, at the best, technical frame, sometimes to the function of the standardization bodies. In such a situation, measurement science, and particularly metrology, risks to be put aside, as can be deduced from the consultation and analysis of the vast literature on the Big Data subject, [9]-[12].

Because of that, the (scientific?) frame developing at present looks paralleling metrology, and its influence is strongly accelerating and using a different language idiom, with scarce connections to the relevant traditional field of data analysis and its idioms. Most of the authors seem not to be "on the same page" of measurement science, as they are considering new and different paradigms sufficient for the Big Data treatment. At least for them, such a position looks like implying that *data quantity sufficiently replaces data quality* [38] – and even that an apparent increase of information can compensate for the traditional consideration of terms like uncertainty and accuracy.

Instead, the information content of data, the only important issue for science and metrology, should first be evaluated in terms of its *quality*, a property that is not intrinsic in any data but that has to be verified.

For example, according to [38] with whom the Author agrees, the dimensions of *information quality* should include (and that is valid for any discipline):

“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;
5. *Trueness*: a condition related to the faithfulness of semantic 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 as synthesized in terms of accuracy in metrology”. (*italics* added)

There is instead a strong tendency to *reduce the analysis of the dataset to its reproducibility*, its evaluation being obtained by checking the trend of the dataset − most often in time, using a mathematical model fitting data: the Author does not feel necessary to report here “real facts”, because they are quite abundant in the literature.

At least in the field of Climate, deeply explored by the Author, he never found a specific analysis in *any* Report of the main international Organisation in the field, like IPCC, HadCRUT, NOOA, etc. Even, in the specific IPCC Report about the treatment of uncertainty [39], words are used instead of numbers, risk concept is preferred to uncertainty and in general a language is used more common to the one used in the field of economic analyses or in decision-taking strategies.

The only parameter considered is, in fact, the standard deviation of the fit of databases covering the whole Earth surface, as checked in several cases by the Author. One typical case is the series of annual Surface Average Mean Temperature, SAMT (also called GMST, Ground Mean Surface Temperature) values for the period 1850-2023 reported in GAWSIS [40]. When fitted by the Author on a 4th order polynomial, he found a s.d.(1850-1945) = 0.12 °C, s.d.(1945-2023) = 0.10 °C and s.d.(total) = 0.11 °C. Similarly, in NOAA [41] the SAMT s.d. (1980-2023) = 0.21 °C and s.d.(1910-2023) = 0.17 °C.

The annual value is obtained from daily meteorological observations, which are naturally quite variable. A typical dispersion is shown in Figure 2, [43].

Notice from the decrease in dispersion the relatively recent increase in precision of the temperature measurement.

It is difficult to imagine that a very large number of stations in the World (up to 35.000) [44] can provide a big dataset with a lower dispersion (possibly except locally). In fact, in 2013 the standard WMO precision was still ± 1 °C, while more recently only the *best* meteorological stations (classes 1 and 2, certainly not the majority of the whole set) claim a precision of ± 0.2 °C [45].

In this way no evaluation of the accuracy of the dataset is obtained, which does not depend only on the observed *dispersion* of the data, but also, and possibly mainly, on the *systematic* effects affecting the influence factors and related to the measuring *process* − in one word the (further) components affecting the UB, the latter being a concept generally absent in Big Data literature.

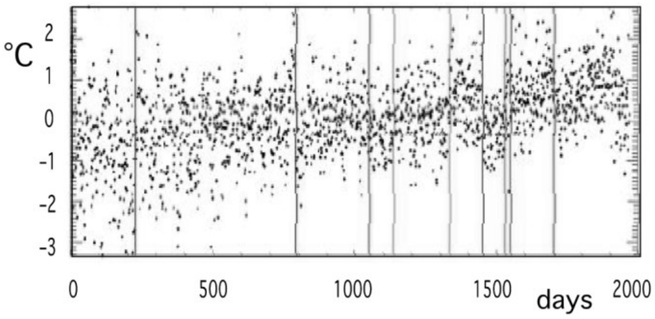

Figure 2. Daily local temperatures between years 1851 and 2018 [42].

In fact, in most of the Big Data literature most scientific terms are lacking for decorating the terms “uncertainty” and “quality”, though still commonly used − but often related only to decisions to be taken [39] − or used with a different meaning.

Several questions arise (R. for *reply*):

Is it true that a larger database ensures wider/better information? [R. *no*]

Are there quantitative limits to the truth of that statement? [R. *yes*]

Does the evidence carried by a large number of data sufficiently replace the need for understanding the process behind those results? [R. *no*]

Does that technique cancel the (hidden) effect of (non-random) causes being at the origin of the data uncertainty? [R. *no*]

To which level a large number of results does mitigate the dataset uncertainty beyond the normal statistical laws? [R. *depends*]

Does the informatics treatment, made necessary for the analysis of a very large dataset, compensate for the lack of “traditional” analysis of data, i.e. following the discipline of metrology? [R. *no*]

Does such analysis possibly benefit from AI as a replacement for “traditional” handling? [R. *possibly*]

However, all the shown replies require *metrological awareness and competence*.

Metrology should certainly have to enter the field by creating new tools suitable to avoid unnecessary dispersion of confusion in the scientific Communities.

Since language is always a critical issue in human communication, the first step for metrology could consist of the integration into VIM terminology of the appropriate terms for the new field of Big Data. Possibly, some of the terms should simply extend their validity with some adaptations to the new field, as recently has been already considered necessary for other frames, like that of data expressed via integer numerical values (namely, e.g., for counting [45]).

A further useful way would be the setup of a *metrological Project* making proposals for further solutions, e.g. by preparing a version of BIPM/ISO terminology documents compatible with the format necessary to include, e.g., also into the AI training specific terms from the metrological literature, [46].

Then, informatics could become an integral part of metrological science, not merely as a technique but also by exploiting new possibilities for the analysis of the data (i.e., for being more informative in general).

5. CONCLUSIONS

In conclusion of the above analyses and reflections, the Author does think that contrast should be easily avoided between metrology and the Big Data frame by supplying the due tools. Only by this way dataism could not lead to two possibly separated frames, as apparently is its present goal.

In addition to digitalisation, today already a matter of discussion and documents also in the metrological Community [47], [48], also the Big Data frame should find the way, through a certain number of adjournments from the metrology side (see, e.g., [49]), to be integrated into the metrological frame and possibly become, after deep analysis, the object of specific treatments within the traditional metrological frame concerning scientific data analysis. Metrology could, e.g., collaborate with some existing International frames like the International Surface Temperature Initiative (ISTI), [42].

Further to numerical data, a useful metrological contribution should also concern more than now the qualitative classes of data (e.g., also in the form of maps), where the variety of interpretations and treatment looks most sparse, and often the term “measurement” is vague or inappropriate, especially in non-scientific communications – the ones being largely more influent than scientific communications.

REFERENCES

[1] L. Wright, Digital metrology: benefits, challenges and open questions, ITS10 Conference Digest 73, NIST SP 2100-05 April 2023, in press.

[2] V. Pleskach, A. Semenchenko, D. Gololobov, O. Kalmykov, V. Krasnoshchok, Digitalization of Science as a Modern Trend of the Information Society Development, CEUR-WS.org 2020 Vol.1-2740/20200146. Use permitted under Creative Commons License Attribution 3.0.

[3] A. V. Denikin, A. V. Sablukov, V. L. Primakov, V. A. Lapshov, P. S. Shikhgafizov, Digitalization in the system of methods of scientific knowledge. SHS Web of Conferences 125, 05008 (2021).
DOI: 10.1051/shsconf/202112505008

[4] OECD, The Digitalisation of Science, Technology and Innovation. Key Developments and Policies, Extended Summary OECD (2020), pp. 5–23.
DOI: 10.1787/b9e4a2c0-en

[5] P. Hellstena, A. Paunu, Digitalization: A Concept Easier to Talk about than to Understand, Proc. of the 12th Int. Joint Conf. on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020) – Vol. 3: KMIS, pp. 226-233.
DOI: 10.5220/0010145302260233

[6] J. Reis, M. Amorim, N. Melão, Y. Cohen, M. Rodrigues, Digitalization: A Literature Review and Research Agenda. Z. Anisic et al. (editors). IJCIEOM (2019), LNMUINEN (2020), pp. 443–456.
DOI: 10.1007/978-3-030-43616-2_47

[7] E. V. Katrin, Approaches to the definition of the term "Digitalization", Amur. Con. (2021). eISSN: 2357-1330 (2022). Use permitted under Creative Commons License Attribution 3.0.
DOI: 10.15405/epsbs.2022.06.49

[8] Wikipedia, Digitization. Online [Accessed 2 July 2024]
https://en.wikipedia.org/wiki/Digitization

[9] S. Sagiroglu, D. Sinanc, Big Data: A Review, IEEE (2013), pp. 42–47.
DOI: 978-1-4673-6404-1/13

[10] A. De Mauro, M. Greco, M. Grimaldi, What is Big Data? A Consensual Definition and a Review of Key Research Topics (2014) AIP Proceedings, 4th Int. Conf. on Integrated Information, Madrid, Spain, 5-8 September 2014, pp. 1–28.
DOI: 10.13140/2.1.2341.5048

[11] A. Mediratta, Big Data: Terms. Definitions, and Applications, KS (2015).

[12] J. M. Cavanillas, E. Curry, W. Wahlster, New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, Book (2016) Big Data Public Private Forum, Springer Open Publ.
DOI: 10.1007/978-3-319-21569-3

[13] D. Petri, Big Data, Dataism and Measurement, IEEE Instrumentation & Measurement Magazine, (2020) 1094-6969/20.
DOI: 10.1109/MIM.2020.9082796

[14] I. Syllaidopoulos, Dataism: The rise of a data-driven world? A Guide for Data-Oriented Policy and Management, Int. J. Multidisciplinary Research and Analysis (2023) 6 (3), pp. 977-982.
DOI: 10.47191/ijmra/v6-i3-16

[15] D. L. Kobelieva, N. M. Nikolaienko, From Information Search to the Loss of Personality: The Phenomenon of Dataism, UDC 141:316.77:004.04, Anthropological Measurements of Philosophical Research (2021) n° 20, pp. 100–112.
DOI: 10.15802/ampr.v0i20.249591

[16] Wikipedia, Dataism. Online [Accessed 2 July, 2024]
https://en.wikipedia.org/wiki/Dataism

[17] F. Pavese, P. De Bièvre, Fostering diversity of thought in measurement. In: Advanced mathematical and computational tools in metrology and testing X. F. Pavese, W. Bremser, A. G. Chunovkina, N. Fischer, A. B. Forbes (editors.). Series on advances in mathematics for applied sciences 86, World Scientific, Singapore (2015), pp. 1–8.

[18] T. S. Kuhn, The structure of scientific revolutions, Book (1962) University of Chicago Press.

[19] R. Kitchin, Big Data, new epistemologies and paradigm shifts, Big Data & Society (2014) April–June 2014 pp. 1–12.
DOI: 10.1177/2053951714528481

[20] Yanlei Diao, Boduo Li, Anna Liu, Liping Peng, Ch. Sutton, Thanh Tran, M. Zink, Capturing Data Uncertainty in High-Volume Stream Processing, 4th Biennial Conf. on Innovative Data Systems Research (CIDR), Asilomar, California, USA, 4-7 January 2009, pp. 4–7.
DOI: 10.48550/arXiv.0909.1777

[21] C. A. de Braganca Pereira, A. Polpo, A. S. Rodrigues, Data Science: Measuring Uncertainties, Editorial, Entropy (2020) 22, 1438.
DOI: 10.3390/e22121438

[22] Guowei Cai, S. Mahadevan, Big Data Analytics in Uncertainty Quantification: Application to Structural Diagnosis and Prognosis, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part A Civil Engineering, March 2018.
DOI: 10.1061/AJRUA6.0000949

[23] G. Kapetanios, M. Marcellino, F. Papailas, Filtering techniques for big data and big data based uncertainty indexes, EUROSTAT Working papers (2017) pp. 1–70. Luxembourg: Publications Office of the European Union.
DOI: 10.2785/880943

[24] P. York, M. Bamberger, Measuring results and impact in the age of big data: The nexus of evaluation, analytics, and digital technology, March 2020. The Rockfeller Foundation, pp. 1–72.

[25] A. Lazzari, J.-M. Pou, Ch. Dubois, L. Leblond, Smart Metrology: The Importance of Metrology of Decisions in the Big Data Era, IEEE Instrumentation & Measurement Magazine, December 2017, 1094-5962.

[26] B. J. Dorr, C. S. Greenberg, P. Fontana, M. Przybocki, M. Le Bras, C. Ploehn, O. Aulov, M. Michel, E. J. Golden, Wo Chang, A new data science research program: evaluation, metrology, standards, and community outreach, Int. J. Data Sci. Anal. (2016), pp. 1–21.
DOI: 10.1007/s41060-016-0016-z

[27] E. Upodhyay, A critical evaluation of handling uncertainty in Big Data processing, Advances in Engineering Software, Vol. 173, Nov 2022.
DOI: 10.1016/j.advengsoft.2022.103246

[28] S. Strauss, Datification and the seductive power of uncertainty - A critical exploration of Big Data enthusiasm, Information (2015) 6,

pp. 836–847.
DOI: 10.3390/info6040836

[29] V. Polužanski, U. Kovacevic, N. Bacanin, T.A. Rashid, S. Stojanovic, B. Nikolic, Application of Machine Learning to Express Measurement Uncertainty, Appl. Sci. (2022) 12, 8581.
DOI: 10.3390/app12178581

[30] R. H. Hariri, E. M. Fredericks, K. M. Bowers, Uncertainty in bid data analytics: survey, opportunities, and challenges, Journal of Big Data (2019) 6, pp. 44.
DOI: 10.1186/s40537-019-0206-3

[31] B. D. Hall, Software for calculation with physical quantities, IEEE Int. Workshop on Metrology for Industry 4.0 & IoT, Remote Workshop, 3-5 June 2020.
DOI: 10.1109/MetroInd4.0IoT48571.2020.9138281

[32] C. McBride, F. Nordvall-Forsberg, Type systems for programs respecting dimensions. In: AMCTM XII. F. Pavese, A.B. Forbes, N. F. Zhang, A. G. Chunovkina, (editors.), 12. Series on Advances in Mathematics for Applied Sciences 90 (2021). World Scientific Publishing Company, pp. 331–346.
DOI: 10.1142/9789811242380_0020

[33] N. Gehani, Units of measure as a data attribute, in Computer languages, Pergamon Press (1977), pp. 93–111.

[34] F. Pavese, Replicated observations in metrology and testing: modelling of repeated and non-repeated measurements, ACQUAL (2007) 12, pp. 525–534.
DOI: 10.1007/s00769-007-0303-4

[35] S. N. Li, S. Dragicevic, F. A. Castro, M. Sester, S. Winter, A. Coltekin, Ch. Éettit, B. Jiang. J. Haworth, A. Stein, T. Cheng, Geospatial big data handling theory and methods: A review and research challenges, ISPRS Journal of Photogrammetry and Remote Sensing (2016) 115, pp. 119–133.
DOI: 10.1016/j.isprsjprs.2015.10.012

[36] F. Pavese, The choice of cartographic system on the calculation of Earth's surface parameters from maps, namely of the GMST (2023) ESIN.
DOI: 10.1007/s12145-023-01061-0

[37] R. Hanisch, S. Chalk, R. Coulon, S. Cox, S. Emmerson, F. J. Flamenco Sandoval, A. Forbes, J. Frey, B. Hall, R. Hartshorn, P. Heus, S. Hodson, K. Hosaka, D. Hutzschenreuter, Chu-Shik Kang, S. Picard, R. White, Unclear units stymie science, Nature Comment Vol. 605, 12 May 2022.

[38] L. Mari, D. Petri, The metrological culture in the context of Big Data: Managing data-driven decision confidence, IEEE Instrumentation and Measurement Mag. (2017) 20 (5), pp. 4–20.
DOI: 10.1109/MIM.2017.8036688

[39] Integrated Risk and Uncertainty Assessment of Climate Change Response Policies. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[40] Federal Office of Meteorology and Climatology MeteoSwiss, GAWSIS – Global Atmosphere Watch – Station Information System. Online [Accessed 30 November 2024]
https://gawsis.meteoswiss.ch/GAWSIS/

[41] National Centers for Environment, Global Climate Report - Annual 2020. Online [Accessed 30 November 2024]
https://www.ncdc.noaa.gov/sotc/global/202013

[42] I. Gillespie, Investigating the potential use of sparse-input reanalyses to homogenise long-term land surface air temperature records. Thesis 2021. Manynooth University, National University of Ireland.

[43] F. Pavese, Decisions at the time of Big Data, Global Journal of Science Frontier Research: A Physics and Space Science (2023) 23 (7), pp. 1–7. Online [Accessed 30 November 2024]
https://journalofscience.org/index.php/GJSFR/article/view/102760

[44] WMO, Guide to Instruments and Methods of Observation. Volume I –Measurement of Meteorological Variables. N° 8. (2018) ISBN 978-92-63-10008-5.
http://public.wmo.int/en/resources/

[45] F. Pavese, From VIM3 toward the next edition of the International Vocabulary of Metrology, Special Issue in Memory of Paul de Bièvre, Accreditation and Quality Assurance (2023) 28 (4), pp.83–88.
DOI: 10.1007/s00769-023-01538-x

[46] L. Leffer, New Training Method Helps AI Generalize like People Do, Scientific American Journal, October 26, 2023. Online [Accessed 12 December 2023]
https://www.scientificamerican.com/article/new-training-method-helps-ai-generalize-like-people-do/

[47] BIPM, On the global digital transformation and the International System of Units, Resolution 2 of the 27th CGPM, November 2022.
DOI: 10.59161/CGPM2022RES2E

[48] BIPM, Strategic Plan for the BIPM Work Programme (2022).

[49] NPL, Teddington UK, “Applying HPC, Big Data and cognitive systems in science and engineering”, in Workshop on Data Metrology and Standards, p. 201.